

MATH 115A - LINEAR ALGEBRA

MATTHEW GHERMAN

These notes are based on “Linear Algebra (5th Edition)” by Stephen Friedberg, Arnold Insel, and Lawrence Spense. They are adapted from notes by Joseph Breen.

CONTENTS

1. Introduction	2
2. Sets	3
3. Induction	5
4. Fields	7
5. Vector spaces	10
6. Subspaces	13
7. Direct sums of subspaces	15
8. Linear combinations and span	17
9. Linear independence	18
10. Bases and dimension	20
11. Linear transformations	26
12. Coordinates and matrix representations	33
13. Composition of linear maps	36
14. Invertibility and isomorphisms	39
15. Change of coordinates	43
16. Diagonalization, eigenvalues, and eigenvectors	46
17. Determinant and the characteristic polynomial	53
18. Invariant subspaces and Cayley-Hamilton Theorem	56
19. Inner products	59
20. Cauchy-Schwarz inequality and angles	62
21. Orthonormal bases and orthogonal complements	64
22. Adjoints	69
23. Normal operators, self-adjoint operators, and the Spectral Theorem	73

1. INTRODUCTION

Math 115A is a course in linear algebra. You might be thinking to yourself: *Wait a second — I already took a course in linear algebra, Math 33A. Why am I taking another one?* This is a fair question, and in this introduction I hope to give you a satisfying answer.

Broadly speaking, there are two main goals of 115A.

- (1) Develop linear algebra from scratch in an abstract setting.
- (2) Improve logical thinking and technical communications skills.

I'll discuss each of these goals separately, and you should keep them in the back of your mind throughout the quarter.

(1) Develop linear algebra from scratch in an abstract setting.

In a lower division linear algebra class like 33A, the subject is usually presented as the study of matrices, or at the least it tends to come off in this way. In reality, you should think about linear algebra at the 115A-level as the study of *vector spaces* and their transformations.

I haven't told you what a *vector space* is yet, so currently this sentence should mean very little to you. To continue saying meaningless things, a *vector space* is simply a universe in which one can do linear algebra. We'll talk about this carefully soon enough, but for now I'll tell you about a vector space that you're already familiar with: \mathbb{R}^n , the set of all n -tuples of real numbers. This is baby's first vector space, and in a linear algebra class like 33A it's usually the *only* vector space that you encounter. In 115A we will develop the theory of linear algebra in other vector spaces, which turns out to be a useful thing to do. Here are some examples to convince you that this is a worthwhile pursuit. I'll repeat: I haven't told you what a vector space is, so all of these examples are only supposed to be interesting stories.

Example 1.1. It turns out that *infinite dimensional vector spaces* are important and come up in math all the time. The one vector space you have seen before, \mathbb{R}^n , is definitely not infinite dimensional. For example, consider the partial differential equation called the *Laplace equation*:

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0.$$

Don't worry if you don't know anything about partial differential equations — you can just trust me that they are important. It turns out that the above equation, and many other differential equations, can be presented as a transformation of an infinite dimensional vector space! In particular, the elements of the vector space are the functions $f(x, y)$.

Example 1.2. In a similar vein, you may have heard of the *Fourier transform*. Here is the Fourier transform of a function $f(x)$:

$$\mathcal{F}(f)(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x} dx.$$

Again, don't worry if this means nothing to you; just trust me that the Fourier transform is important. Looking at the above formula — with an integral, an exponential, and imaginary numbers — it may seem like the Fourier transform is as far from “linear algebra” as possible. In reality, the Fourier transform is just another transformation of an infinite dimensional vector space of functions!

Example 1.3. Infinite dimensional vector spaces arise naturally in physics as well. For example, in quantum mechanics, the set of possible states of a quantum mechanical system forms an infinite dimensional vector space. An *observable* in quantum mechanics is just a transformation of that infinite dimensional vector space. By the way, don't ask me too many questions about this — I don't know anything about quantum physics!

Example 1.4. Finite dimensional vector spaces that are not \mathbb{R}^n are also important. There are a number of simple examples I could give, but I'll describe something a little more out there. In geometry and topology, mathematicians are usually interested in detecting when two complicated shapes are either the same or different. One fancy way of doing this is with something called *homology*. You can think of homology as a complicated machine that eats in a shape and spits out a bunch of data. Oftentimes, that data is a list of vector spaces. In other words, if S_1 and S_2 are two complicated mathematical shapes, and H is a homology machine, you can feed S_1 and S_2 to H to get:

$$\begin{aligned} H(S_1) &= \{V_1, \dots, V_n\} \\ H(S_2) &= \{W_1, \dots, W_n\}. \end{aligned}$$

Here, V_1, \dots, V_n and W_1, \dots, W_n are all vector spaces (and they aren't just copies of \mathbb{R}^n). If the homology machine spits out different lists for the two shapes, then those shapes must have been different! This might sound ridiculous (because it is ridiculous) but if your shapes live in, like, 345033420 dimensions then it's usually easier to distinguish them by comparing the vector spaces output by a homology machine, rather than trying to distinguish them in some geometric way.

My point is that vector spaces of all sizes and shapes are extremely common in math, physics, statistics, engineering, and life in general, so it is important to develop a theory of linear algebra that applies to all of these, rather than just \mathbb{R}^n . We will approach the subject by starting from square one. A healthy perspective to take is to forget almost all math you've ever done and treat 115A like a foundational axiomatic course to develop a particular field of math. This is the first goal of 115A.

The last remark about goal (1) that I'll make is the following. You might be thinking: *Wow, linear algebra in vector spaces other than \mathbb{R}^n must be wild and different from what I'm used to! I can't wait to learn all of the new interesting theory that Joe is hyping up!* If you are thinking this, then I'm going to burst your bubble and spoil the punchline of 115A: *Abstract linear algebra in general vector spaces is basically the same as linear algebra in \mathbb{R}^n . Nothing new or interesting happens.* We will talk about linear independence, linear transformations, kernels and images, eigenvectors and diagonalization, all topics that you are familiar with in the context of \mathbb{R}^n , and everything will work the same way in 115A.

(2) Improve logical thinking and technical communication skills.

At some level, this goal is a flowery way of referring to "proof-writing", but I don't like boiling it down to something as simple as that. Upper division math (and real math in general) is different than lower division math because of the focus on *discovering and communicating truth*, rather than computation. As such, you should treat every solution you write in 115A (and any other math class, ever) as a mini technical essay. Long gone are the days where you do scratch work to figure out the answer to some problem and then just submit that. High level math is all about polished, logical, and clear communication of truth.

This is difficult to learn to do well and it takes a lot of time and practice!

2. SETS

Before we discuss vector spaces, we need to take care of a few boring preliminaries. The basic building block of a vector space is something called a *field*, which is what we will discuss in the next section. But before even that, I want to introduce you to some notation and basic concepts that will be central to the entire course. Hopefully you are already familiar with the basic notions of sets to some extent, so this first subsection will be a brief overview.

The most fundamental object of interest in all of math is a *set*. A set is just a collection — possibly infinite — of things. For example,

$$S_1 = \{1, 2, -400\}$$

is a set consisting of the numbers 1, 2, and -400 . As another example,

$$S_2 = \{\text{blue}, :, \pi\}$$

is a set consisting of the word blue, a smiley face, and the number π .

For larger sets, we will sometimes use the following notation:

$$S_3 = \{n : n \text{ is a positive, even number}\}.$$

This notation is read: S_3 is the set consisting of elements of the form n such that n is any positive, even number. The colon is read as “such that,” and the stuff after the colon is a collection of conditions that all elements of the set must satisfy. Unraveling the set S_3 above,

$$S_3 = \{n : n \text{ is a positive, even number}\} = \{2, 4, 6, 8, \dots\}.$$

We use the symbol \in to indicate if something is an element of a set. For example, recall the set $S_1 = \{1, 2, -400\}$ from above. We could write

$$2 \in S_1$$

because 2 is an element of S_1 . We could also write

$$3 \notin S_1$$

because 3 is *not* an element of S_1 .

We can define operations on sets. For example, if A and B are sets, then we define

$$\begin{aligned} A \cup B &:= \{x : x \in A \text{ or } x \in B\} \\ A \cap B &:= \{x : x \in A \text{ and } x \in B\} \end{aligned}$$

The first is the *union* of the sets A and B , and the second is the *intersection*. For example, using S_1 and S_3 from above,

$$\begin{aligned} S_1 \cup S_3 &= \{-400, 1, 2, 4, 6, 8, \dots\} \\ S_1 \cap S_3 &= \{2, -400\}. \end{aligned}$$

The *empty set*, denoted \emptyset , is the set consisting of no elements. That is, $\emptyset := \{\}$. We could write

$$S_1 \cap S_2 = \emptyset.$$

When two sets have empty intersection, we say that they are *disjoint*.

We can also discuss *subsets*. In particular, if A and B are two sets, then we say $A \subset B$ (or $A \subseteq B$, both notations confusingly mean the same thing) if every element of A is also an element of B . For example,

$$\{4, 6, e\} \subset \{4, 6, e, 10, 24\}.$$

One thing that you will have to do often in this class, and in life, is show that two sets are the same. To show $A = B$, you should show that $A \subset B$ and $B \subset A$.

The following are some important sets for this course.

$$\mathbb{N} := \{x : x \text{ is a natural number}\} = \{0, 1, 2, 3, \dots\}$$

$$\mathbb{Z} := \{x : x \text{ is an integer}\} = \{\dots, -2, -1, 0, 1, 2, \dots\}$$

$$\mathbb{R} := \{x : x \text{ is a real number}\}$$

$$\mathbb{Q} := \{x : x \text{ is a rational number}\} = \left\{\frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0\right\}$$

$$\mathbb{C} := \{x : x \text{ is a complex number}\} = \{a + bi : a, b \in \mathbb{R}\}$$

3. INDUCTION

One of the most useful proof techniques in all of mathematics is induction. Although you might already know how induction works, many students are not aware when it should be used. We will introduce induction as a proof technique and focus on identifying characteristics of an induction problem.

Induction is typically a way of proving a statement we already know to be true that can be indexed by the natural numbers. At first glance, that might seem restrictive, but sometimes intuitive claims that are easy to state are extremely difficult to prove (e.g. Goldbach conjecture or Collatz conjecture). The general idea is to prove the claim for a simple case (the base case). Then we prove that if the claim is true at some point, it must be true in the next iteration (the inductive step). The statement is true in each case because it was true in the previous case. We are able to lump infinitely many proofs (one for each natural number) into a two-step process. We illustrate the process with a basic example.

Proposition 3.1. For each natural number n , the sum of the first n positive integers is $\frac{n(n+1)}{2}$.

Proof. As a base case, we inspect the claim when $n = 1$. Clearly, $1 = \frac{1 \cdot 2}{2}$.

We now attempt the inductive step. Assume that the claim is true for n so $1 + 2 + \dots + n = \frac{n(n+1)}{2}$. We will prove that $1 + 2 + \dots + n + (n + 1) = \frac{(n+1)(n+2)}{2}$. We start with the assumption and add $n + 1$ to both sides to obtain the left hand side of the equality we want to prove. We find a common denominator and factor the numerator on the right hand side to obtain the desired result.

$$\begin{aligned} 1 + 2 + \dots + n &= \frac{n(n+1)}{2} \\ 1 + 2 + \dots + n + (n+1) &= \frac{n(n+1)}{2} + (n+1) \\ 1 + 2 + \dots + n + (n+1) &= \frac{n^2 + n + 2(n+1)}{2} \\ 1 + 2 + \dots + n + (n+1) &= \frac{n^2 + 3n + 2}{2} \\ 1 + 2 + \dots + n + (n+1) &= \frac{(n+1)(n+2)}{2} \end{aligned}$$

□

Proposition 3.1 is clearly true for small cases of n . In the inductive step, we show that the next formula will be true based purely on the previous formula. The process continues forever, proving the claim for all instances of n .

The discussion thus far has been about *weak induction* since, in the inductive step, we only use the previous value for n to prove the next. With *strong induction*, we actually assume that each value for n up to the desired value is true. We will illustrate strong induction with the following.

Proposition 3.2. The Fibonacci sequence is defined by $F_{n+2} = F_{n+1} + F_n$ for all integers $n \geq 0$ with starting values $F_1 = 1$ and $F_2 = 1$. A formula for F_n is given by

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right).$$

Proof. We proceed via strong induction on n . As a base case, take $n = 1$. We have $F_1 = 1$ and

$$\frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^1 - \left(\frac{1-\sqrt{5}}{2} \right)^1 \right) = \frac{1}{\sqrt{5}} \cdot \frac{2\sqrt{5}}{2} = 1.$$

Now for the inductive step. Assume that

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right)$$

for all $n \leq k$. We will show that

$$F_{k+1} = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \right).$$

The formula for the Fibonacci sequence states that $F_{k+1} = F_k + F_{k-1}$. We also note that

$$\left(\frac{1 \pm \sqrt{5}}{2} \right)^2 = \frac{6 \pm 2\sqrt{5}}{4} = \frac{3 \pm \sqrt{5}}{2}.$$

Thus

$$\begin{aligned} F_{k+1} &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^k - \left(\frac{1-\sqrt{5}}{2} \right)^k \right) + \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k-1} - \left(\frac{1-\sqrt{5}}{2} \right)^{k-1} \right) \\ &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^k + \left(\frac{1+\sqrt{5}}{2} \right)^{k-1} \right) - \frac{1}{\sqrt{5}} \left(\left(\frac{1-\sqrt{5}}{2} \right)^k + \left(\frac{1-\sqrt{5}}{2} \right)^{k-1} \right) \\ &= \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{k-1} \left(\frac{3+\sqrt{5}}{2} \right) - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^{k-1} \left(\frac{3-\sqrt{5}}{2} \right) \\ &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \right) \end{aligned}$$

as desired. □

One reason the proof went smoothly is we were able to assume the previous two steps of the process during the inductive step. Weak induction would not have been enough to prove the statement in this way.

In a surprising turn of events, strong induction and weak induction are equivalent. We can prove strong induction works using weak induction and visa versa. As a result, most people use strong induction and weak induction interchangeably.

End of lecture 1

4. FIELDS

Some sets are just simple collections of elements with no extra structure. Other sets naturally admit an extra amount of structure and interaction (i.e., *algebra*.) For example, in the set of integers, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$, we are familiar with the algebraic operations of addition (+) and multiplication (\cdot). That is, given two elements $n, m \in \mathbb{Z}$, we can construct a third element $n + m \in \mathbb{Z}$ by adding them together, and likewise a fourth element $n \cdot m \in \mathbb{Z}$. Furthermore, these algebraic operations obey a handful of rules (commutativity, distribution) that you learned when you were a kid.

In contrast, consider the set $S_2 = \{\text{blue}, :, \pi\}$ from before. We don't have any familiar algebraic structure on this set, so for now it will be just be an unstructured collection of random elements.

In Math 115A there is a particular type of set called a *field* that will be of utmost importance. It is a set with two operations that satisfy a bunch of rules. I'll give you the formal definition, and then we'll look at some examples.

Definition 4.1. A **field** is a set \mathcal{F} with two operations, addition (+) and multiplication (\cdot), that take a pair of elements $x, y \in \mathcal{F}$ and produce new elements $x + y, x \cdot y \in \mathcal{F}$. Furthermore, these operations satisfy the following properties.

- (1) For all $x, y \in \mathcal{F}$,

$$x + y = y + x$$

$$x \cdot y = y \cdot x.$$

We refer to this property as commutativity of addition and multiplication respectively.

- (2) For all $x, y, z \in \mathcal{F}$,

$$(x + y) + z = x + (y + z)$$

$$(x \cdot y) \cdot z = x \cdot (y \cdot z)$$

We refer to this property as associativity of addition and multiplication respectively.

- (3) For all $x, y, z \in \mathcal{F}$,

$$x \cdot (y + z) = x \cdot y + x \cdot z.$$

We refer to this property as distributivity of multiplication over addition.

- (4) There are elements $0, 1 \in \mathcal{F}$ such that, for all $x \in \mathcal{F}$,

$$0 + x = x$$

$$1 \cdot x = x.$$

The element 0 is an **additive identity** and the element 1 is a **multiplicative identity**.

- (5) For each $x \in \mathcal{F}$, there is an element $x' \in \mathcal{F}$, called an **additive inverse**, such that $x + x' = 0$. Similarly, for every $y \neq 0 \in \mathcal{F}$, there is an element $y' \in \mathcal{F}$, called a **multiplicative inverse**, such that $y \cdot y' = 1$.

Example 4.2. The main example of a field is \mathbb{R} , the set of real numbers, with the usual operations of addition and multiplication. All of the above properties should look familiar to you, precisely because they are modeled after the behavior of \mathbb{R} . Throughout Math 115A we will work with abstract fields \mathcal{F} , but usually you can secretly think about \mathbb{R} in your head.

Example 4.3. Other familiar examples of field are \mathbb{Q} and \mathbb{C} .

Example 4.4. The set of integers, \mathbb{Z} , with the usual operations of addition and multiplication, is *not* a field. Almost all of the field properties are satisfied, except for the multiplicative inverse property. In particular, it is *not* the case that for any $y \neq 0 \in \mathbb{Z}$, there is a $y' \in \mathbb{Z}$ such that

$y \cdot y' = 1$. For example, the element $2 \in \mathbb{Z}$ does not have a multiplicative inverse; we know in our minds that such a number would have to be $\frac{1}{2}$, but that number doesn't exist in \mathbb{Z} .

Similarly, \mathbb{N} is not a field. Not only does \mathbb{N} not have multiplicative inverses, but it also doesn't have additive inverses!

Example 4.5. Here is an example of a field that you may not have seen before. Let $\mathcal{F}_2 := \{0, 1\}$ be the set consisting of 2 elements, 0 and 1. Define addition as

$$\begin{aligned} 0 + 0 &:= 0 \\ 0 + 1 &:= 1 \\ 1 + 1 &:= 0, \end{aligned}$$

and define multiplication as

$$\begin{aligned} 0 \cdot 0 &:= 0 \\ 0 \cdot 1 &:= 0 \\ 1 \cdot 1 &:= 1. \end{aligned}$$

We claim that \mathcal{F}_2 is a field! We won't verify all of the properties, but each element has an additive inverse (the additive inverse of 0 is 0, and the additive inverse of 1 is 1), and each non-zero element has a multiplicative inverse (the multiplicative inverse of 1 is 1).

We will now prove some basic properties of fields. Of course, we want to establish some of these properties, but the main purpose here is to practice writing proofs and to get you in the correct mindset for the course. There are a lot of algebraic operations that you take for granted in a field, and we need to prove them using the defining properties.

Proposition 4.6 (Cancellation laws). Let \mathcal{F} be a field. Let $x, y, z \in \mathcal{F}$.

- (i) If $x + y = x + z$, then $y = z$.
- (ii) If $x \cdot y = x \cdot z$ and $x \neq 0$, then $y = z$.

Proof. First, we prove (i). Suppose that $x + y = x + z$. By Definition 4.1(5), there exists an element x' such that $x + x' = 0$. Adding x' to both sides of the assumed equality gives

$$x' + (x + y) = x' + (x + z).$$

By associativity of addition in a field, this is equivalent to

$$(x' + x) + y = (x' + x) + z.$$

Using the fact that $x' + x = 0$ gives

$$0 + y = 0 + z \quad \Rightarrow \quad y = z.$$

Next, we prove (ii). Suppose that $x \cdot y = x \cdot z$ and $x \neq 0$. By Definition 4.1(5), there is an element x' such that $x' \cdot x = 1$. Multiplying both sides of the assumed equality,

$$x' \cdot (x \cdot y) = x' \cdot (x \cdot z).$$

By associativity of multiplication, it follows that

$$(x' \cdot x) \cdot y = (x' \cdot x) \cdot z$$

so

$$1 \cdot y = 1 \cdot z \quad \Rightarrow \quad y = z.$$

□

As a corollary, we get another fact that you have also taken for granted (and is not directly stated in the definition of a field).

Corollary 4.6.1. The elements 0 and 1 in a field are unique.

Proof. Suppose that $0' \in \mathcal{F}$ is another additive identity, so that $0' + x = x$ for all $x \in \mathcal{F}$. Then, since $0 + x = x$, we have

$$0' + x = 0 + x$$

for all $x \in \mathcal{F}$. By Proposition 4.6(i), it follows that $0' = 0$.

Similarly, suppose that there is an element $1' \in \mathcal{F}$ such that $1' \cdot x = x$ for all $x \in \mathcal{F}$. Then since $1 \cdot x = x$, we have

$$1' \cdot x = 1 \cdot x$$

for all $x \in \mathcal{F}$. In particular, we may choose $x = 1$. By the Proposition 4.6(ii), it follows that $1 = 1'$. \square

A similar statement is the uniqueness of multiplicative and additive inverses.

Corollary 4.6.2. For each $x \in \mathcal{F}$, the element x' satisfying $x + x' = 0$ is unique. If $x \neq 0$, the element x' satisfying $x' \cdot x = 1$ is unique.

Proof. Assume that x' and x'' are elements of \mathcal{F} such that $x + x' = 0$ and $x + x'' = 0$. In particular, $x + x' = x + x''$ so Proposition 4.6(i) proves $x' = x''$.

Assume that x' and x'' are elements of \mathcal{F} such that $x \cdot x' = 1$ and $x \cdot x'' = 1$. In particular, $x \cdot x' = x \cdot x''$ so Proposition 4.6(ii) proves $x' = x''$. \square

These corollaries allows us to talk about *the* additive identity, *the* multiplicative identity, and *the* additive inverse of an element. Furthermore, we can make the following notational definition.

Definition 4.7. Let \mathcal{F} be a field, and let $x \in \mathcal{F}$. The additive inverse is also denoted $-x$, and then multiplicative inverse (if $x \neq 0$) is denoted x^{-1} or $\frac{1}{x}$.

Here are some more familiar properties of real numbers that are true in all fields.

Proposition 4.8. Let \mathcal{F} be a field, and let $x, y \in \mathcal{F}$.

- (i) $0 \cdot x = 0$,
- (ii) $-(-x) = x$,
- (iii) $(-x) \cdot y = x \cdot (-y) = -(x \cdot y)$,
- (iv) $(-x) \cdot (-y) = x \cdot y$,
- (v) If \mathcal{F} has more than one element, then 0 has no multiplicative inverse.

Proof. (i)

$$\begin{aligned}
 0 \cdot x &= 0 \cdot x + 0 && \text{additive identity} \\
 &= 0 \cdot x + (x + (-x)) && \text{additive inverse} \\
 &= (0 \cdot x + x) + (-x) && \text{associativity of addition} \\
 &= (x \cdot 0 + x \cdot 1) + (-x) && \text{commutativity of multiplication} \\
 &= x \cdot (0 + 1) + (-x) && \text{distributivity of multiplication over addition} \\
 &= x + (-x) && \text{multiplicative identity} \\
 &= 0
 \end{aligned}$$

- (ii) We want to show that x is the additive inverse of $-x$. By commutativity of addition, $0 = x + (-x) = (-x) + x$. By uniqueness of additive inverses, Corollary 4.6.2, $-(-x) = x$.

- (iii) In order to prove $x \cdot (-y) = -(x \cdot y)$, we need to show that $x \cdot (-y)$ is the additive inverse of $x \cdot y$. We have

$$\begin{aligned}
 x \cdot y + x \cdot (-y) &= x \cdot (y + (-y)) && \text{distributivity of multiplication over addition} \\
 &= x \cdot 0 && \text{additive inverse} \\
 &= 0 \cdot x && \text{commutativity of multiplication} \\
 &= 0 && \text{(i)}
 \end{aligned}$$

- (iv) Follow a similar argument to show that $(-x) \cdot y = -(x \cdot y)$.

$$\begin{aligned}
 (-x) \cdot (-y) &= -(x \cdot (-y)) && \text{(iii)} \\
 &= -(-(x \cdot y)) && \text{(iii)} \\
 &= x \cdot y && \text{(ii)}
 \end{aligned}$$

- (v) We will prove the contrapositive. Assume there is some $x \in \mathcal{F}$ such that $0 \cdot x = 1$. By (i), $0 = 0 \cdot x = 1$. For any $y \in \mathcal{F}$, we have $y = 1 \cdot y = 0 \cdot y = 0$. The field contains only the zero element. □

We can also now define the notions of subtraction and division in a field, more things that you've taken for granted!

Definition 4.9. Let \mathcal{F} be a field. For $x, y \in \mathcal{F}$, define

$$x - y := x + (-y).$$

Similarly, if $y \neq 0$, define

$$\frac{x}{y} := x \cdot \frac{1}{y}.$$

5. VECTOR SPACES

Just as a field is an abstraction of \mathbb{R} , a vector space will be an abstraction of our understanding of \mathbb{R}^n . Vector spaces are one the main objects of interest in linear algebra.

Definition 5.1. A **vector space over a field \mathcal{F}** , also referred to as an **\mathcal{F} -vector space**, is a set V with two operations, addition (+) and scalar multiplication (\cdot), the first of which takes a pair of elements $v, w \in V$ and produces a new element $v + w \in V$, and the second of which takes an element $\lambda \in \mathcal{F}$ and an element $v \in V$ and produces a new element $\lambda \cdot v \in V$. Moreover, these operations satisfy the following properties:

- (1) For all $v, w \in V$, $v + w = w + v$.
- (2) For all $u, v, w \in V$, $(u + v) + w = u + (v + w)$.
- (3) There exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$.
- (4) For each $v \in V$, there is an element $v' \in V$ such that $v + v' = 0$.
- (5) For all $v \in V$, $1 \cdot v = v$.
- (6) For all $\lambda, \mu \in \mathcal{F}$ and $v \in V$, $(\lambda \cdot \mu) \cdot v = \lambda \cdot (\mu \cdot v)$.
- (7) For all $\lambda \in \mathcal{F}$ and $v, w \in V$, $\lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$.
- (8) For all $\lambda, \mu \in \mathcal{F}$ and $v \in V$, $(\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v$.

Definition 5.2. A **vector** is an element of a vector space.

Remark 5.3. An important but subtle point is that the operations that define a vector space are *distinct* from those that define a field. For example, in Definition 5.1(6) there are two completely different types of multiplication happening on each side of the equation. In the expression

$$(\lambda \cdot \mu) \cdot v$$

the multiplication in the parentheses is the *multiplication operation in the field \mathcal{F}* . The second \cdot represents *scalar multiplication in the vector space V* , which is a completely different operation. In contrast, the expression

$$\lambda \cdot (\mu \cdot v)$$

has only scalar multiplication in the vector space V .

Example 5.4. The set

$$\mathbb{R}^n := \{(x_1, \dots, x_n) : x_j \in \mathbb{R}\}$$

is a vector space over \mathbb{R} with addition defined as

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

and scalar multiplication defined as

$$\lambda \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \lambda \cdot x_1 \\ \vdots \\ \lambda \cdot x_n \end{pmatrix}.$$

All of the above vector space axioms are the usual familiar algebraic rules in \mathbb{R}^n .

Example 5.5. More generally, we can consider

$$\mathcal{F}^n := \{(x_1, \dots, x_n) : x_j \in \mathcal{F}\}$$

where \mathcal{F} is any field. Then \mathcal{F}^n is a vector space over \mathcal{F} with addition defined as

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

and scalar multiplication defined as

$$\lambda \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \lambda \cdot x_1 \\ \vdots \\ \lambda \cdot x_n \end{pmatrix}.$$

Example 5.6. Let $M_{m \times n}(\mathcal{F}) := \{A : A \text{ is an } m \times n \text{ matrix with entries in } \mathcal{F}\}$. Define addition in the usual way. If $A, B \in M_{m \times n}(\mathcal{F})$, then

$$(A + B)_{ij} := A_{ij} + B_{ij}.$$

Here, A_{ij} is the (i, j) th entry of the matrix A . Likewise, define scalar multiplication for $\lambda \in \mathcal{F}$ and $A \in M_{m \times n}(\mathcal{F})$ as

$$(\lambda A)_{ij} := \lambda A_{ij}.$$

Then $M_{m \times n}(\mathcal{F})$ is a vector space over \mathcal{F} . Again, we don't verify all of the properties here, but these are the usual algebraic operations on matrices. The zero matrix $0 \in M_{m \times n}(\mathcal{F})$ satisfies Definition 5.1(3).

Example 5.7. We can endow \mathbb{C} with the structure of a vector space in a few ways.

- (1) With the usual operations of complex addition and multiplication, \mathbb{C} is a vector space over \mathbb{C} . (In this case, both field multiplication *and* scalar multiplication in the vector space are given by the usual multiplication of complex numbers).
- (2) We can also endow \mathbb{C} with the structure of a vector space over \mathbb{R} . This time, scalar multiplication for $\lambda \in \mathbb{R}$ and $a + bi \in \mathbb{C}$ is defined as

$$\lambda \cdot (a + bi) = \lambda a + \lambda bi.$$

This is a *different* vector space structure than (1)!

Example 5.8. In general, if \mathcal{F} is a field, then \mathcal{F} is a vector space over \mathcal{F} .

Example 5.9. Let S be a set and \mathcal{F} a field. Define

$$F(S, \mathcal{F}) := \{f : S \rightarrow \mathcal{F}\}.$$

The notation $f : S \rightarrow \mathcal{F}$ is a function f whose domain is S and whose codomain is \mathcal{F} . Define addition as

$$(f + g)(s) = f(s) + g(s)$$

and scalar multiplication as

$$(\lambda f)(s) = \lambda f(s).$$

Then $F(S, \mathcal{F})$ is a vector space over \mathcal{F} .

End of lecture 2

Proposition 5.10 (Cancellation law for vector addition). Let V be a vector space, and let $u, v, w \in V$. Suppose that $u + v = u + w$. Then $v = w$.

Proof. Let u' be an additive inverse of u .

$$\begin{aligned} u + v &= u + w \\ u' + (u + v) &= u' + (u + w) \\ (u' + u) + v &= (u' + u) + w && \text{associativity of addition} \\ (u + u') + v &= (u + u') + w && \text{commutativity of addition} \\ 0 + v &= 0 + w && \text{additive inverse} \\ v &= w && \text{additive identity} \end{aligned}$$

□

Corollary 5.10.1. In a vector space, the element 0 is unique. Likewise, for each $v \in V$, the element $v' \in V$ satisfying $v + v' = 0$ is unique.

Proof. We will first prove that 0 is unique. Let $0' \in V$ be such that $0' + v = v$ for all $v \in V$. Then $0' = 0' + 0 = 0$ and the additive identity is unique.

Let v' and v'' be elements of V such that $v + v' = 0$ and $v + v'' = 0$. Then $v + v' = v + v''$ so $v' = v''$ by Proposition 5.10. □

Definition 5.11. Let $v \in V$. Define $-v \in V$ to be the unique element satisfying $v + (-v) = 0$.

As in a field, we can then define *subtraction* in a vector space as $v - w := v + (-w)$.

Proposition 5.12. Let V be a vector space over a field \mathcal{F} .

- (i) For each $v \in V$, $0 \cdot v = 0$.
- (ii) For each $v \in V$ and $\lambda \in \mathcal{F}$, $(-\lambda)v = \lambda(-v) = -(\lambda v)$.
- (iii) For each $\lambda \in \mathcal{F}$, $\lambda \cdot 0 = 0$.

Proof. (i)

$$\begin{aligned}
 0 \cdot v &= 0 \cdot v + (v + (-v)) && \text{additive inverse} \\
 &= (0 \cdot v + v) + (-v) && \text{associativity of addition} \\
 &= (0 + 1) \cdot v + (-v) && \text{distributivity} \\
 &= v + (-v) && \text{multiplication by 1} \\
 &= 0 && \text{additive inverse}
 \end{aligned}$$

(ii) Recall that $-(\lambda v)$ is the unique element in V such that $\lambda v + (-\lambda v) = 0$. Then

$$\lambda v + (-\lambda)v = (\lambda + (-\lambda))v$$

by Definition 5.1(8). Since $\lambda + (-\lambda) = 0$ in \mathcal{F} ,

$$\lambda v + (-\lambda)v = 0v = 0$$

where the last equality follows from (i). By Corollary 5.10.1, $(-\lambda)v = -(\lambda v)$.

We, likewise, show that

$$\lambda v + \lambda(-v) = \lambda(v + (-v)) = \lambda 0 = 0.$$

Thus $\lambda(-v)$ is the additive inverse of λv and $\lambda(-v) = -(\lambda v)$.

(iii)

$$\begin{aligned}
 \lambda \cdot 0 &= \lambda \cdot (v + (-v)) && \text{additive inverse} \\
 &= \lambda v + \lambda(-v) && \text{distributivity} \\
 &= \lambda v + (-\lambda v) && \text{(ii)} \\
 &= 0 && \text{additive inverse}
 \end{aligned}$$

□

6. SUBSPACES

Definition 6.1. Let V be a vector space over a field \mathcal{F} . A **subspace** of V is a subset $W \subset V$ such that W is a vector space over \mathcal{F} with the operations inherited from V .

Example 6.2. Let V be a vector space. Then $\{0\} \subset V$ and $V \subset V$ are both subspaces of V .

Note that, for *any* subset of a vector space, the axioms (1), (2), (5), (6), (7), and (8) in Definition 5.1 automatically hold because the subset inherits the operations of V . Thus, to determine if a given subset W of V is a subspace, one needs to only verify that addition and scalar multiplication are well-defined when restricted to the subset along with following axioms.

(3) There exists an element $0 \in W$ such that $w + 0 = w$ for all $w \in W$.

(4) For each $w \in W$, there is an element $w' \in W$ such that $w + w' = 0$.

In fact, we can identify subspaces in the following efficient manner.

Proposition 6.3. Let V be a vector space over \mathcal{F} . A subset $W \subset V$ is a subspace if and only if the following three properties hold for the operations defined in V .

(1) $0 \in W$

(2) If $v, w \in W$, then $v + w \in W$.

(3) If $\lambda \in \mathcal{F}$ and $v \in W$, then $\lambda v \in W$.

Proof. (\Rightarrow) Assume that $W \subset V$ is a subspace. Since W is a vector space with the operations inherited from V , (2) and (3) automatically hold. Furthermore, since W is a vector space, there exists an element $0' \in W$ such that $0' + w = w$ for all $w \in W$. The zero element from V satisfies $0 + v = v$ for all $v \in V$. By the cancellation law, Proposition 5.10, $0 = 0'$. Therefore, $0 \in W$, and (1) holds.

(\Leftarrow) Assume that (1), (2), and (3) hold. We wish to show that W is a subspace of V . Because W inherits the operations of V , it follows that (1), (2), (3), (5), (6), (7), and (8) from Definition 5.1 automatically hold in W . We only need to verify the existence of additive inverses. Fix $w \in W$. By closure under scalar multiplication, $(-1)w = -w \in W$ so w has an additive inverse in W . Therefore, W is a subspace. \square

We call the second property closure under addition and the third property closure under scalar multiplication. Using these, we can easily identify more examples and non-examples of subspaces.

Example 6.4. The intuitive notion of a subspace should be familiar in the setting of \mathbb{R}^n . Namely, any hyperplane passing through the origin is a subspace. A hyperplane *not* passing through the origin is *not* a subspace, in part because it doesn't contain 0. A set like

$$S = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$$

is not a subspace of \mathbb{R}^2 since it isn't closed under addition or scalar multiplication (though it does contain the origin).

End of lecture 3

Example 6.5. Consider the vector space $M_{n \times n}(\mathcal{F})$ of $n \times n$ matrices with entries in a field \mathcal{F} . Let $W \subset M_{n \times n}(\mathcal{F})$ be the set of all *symmetric* matrices,

$$W := \{A \in M_{n \times n}(\mathcal{F}) : A_{ij} = A_{ji} \text{ for all } 1 \leq i, j \leq n\}.$$

Note that the 0 matrix is symmetric. If A and B are symmetric, then $A + B$ and λA are symmetric. Thus, W is a subspace of $M_{n \times n}(\mathcal{F})$.

End of lecture 4

We can also define the set of *skew-symmetric* matrices,

$$Z := \{A \in M_{n \times n}(\mathcal{F}) : A_{ij} = -A_{ji} \text{ for all } 1 \leq i, j \leq n\}.$$

Note that the 0 matrix is skew-symmetric. If A and B are skew-symmetric, then $A + B$ and λA are skew-symmetric. Thus Z is a subspace of $M_{n \times n}(\mathcal{F})$.

Example 6.6. Consider the vector space $M_{n \times n}(\mathbb{R})$ of $n \times n$ matrices with entries in \mathbb{R} . Let $S \subset M_{n \times n}(\mathbb{R})$ be the set of matrices with non-negative entries,

$$S := \{A \in M_{n \times n}(\mathbb{R}) : A_{ij} \geq 0\}.$$

The set S is not a subspace since it isn't closed under scalar multiplication. For example, if $A \in M_{n \times n}(\mathbb{R})$ is any non-zero matrix, then $(-1)A \notin S$.

Proposition 6.7. Let V be a vector space, and let $W_1, W_2 \subset V$ be subspaces. Then $W_1 \cap W_2$ is a subspace.

Proof. Since W_1 and W_2 are subspaces, $0 \in W_1$ and $0 \in W_2$. Thus, $0 \in W_1 \cap W_2$.

Let $v, w \in W_1 \cap W_2$. Since $v, w \in W_1$ and W_1 is a subspace, we have $v + w \in W_1$. Likewise, since $v, w \in W_2$ and W_2 is a subspace, we have $v + w \in W_2$. Thus, $v + w \in W_1 \cap W_2$.

Let $v \in W_1 \cap W_2$ and $\lambda \in \mathcal{F}$. Since W_1 is a subspace and $v \in W_1$, we have $\lambda v \in W_1$. Likewise, since W_2 is a subspace and $v \in W_2$, we have $\lambda v \in W_2$.

By Proposition 6.3, $W_1 \cap W_2$ is a subspace. \square

In general, the union of two subspaces is not a subspace!

Example 6.8. Consider \mathbb{R}^2 with W_1 the x -axis and W_2 the y -axis. Then $W_1 \cup W_2$ is not closed under addition. For example, $(1, 0) + (0, 1) = (1, 1)$ is not on either axis. Thus $W_1 \cup W_2$ is not a subspace of \mathbb{R}^2 .

7. DIRECT SUMS OF SUBSPACES

Even though the union of two subspaces is not a subspace in general, it would be nice to have a way to “combine” two subspaces to form a new one.

Definition 7.1. Let $S_1, S_2 \subset V$ be subsets of a vector space V . The **sum** of S_1 and S_2 is

$$S_1 + S_2 := \{s_1 + s_2 : s_1 \in S_1, s_2 \in S_2\}.$$

Definition 7.2. A vector space V is the **direct sum** of two subspaces $W_1, W_2 \subset V$, written $V = W_1 \oplus W_2$, if $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$.

Example 7.3. Consider the vector space \mathbb{R}^2 . Let W_1 and W_2 be the x -axis and y -axis respectively as in Example 6.8. Fix any element $(x, y) \in \mathbb{R}^2$. Then

$$(x, y) = (x, 0) + (0, y).$$

Since $(x, 0) \in W_1$ and $(0, y) \in W_2$, we have $\mathbb{R}^2 = W_1 + W_2$. Further, $W_1 \cap W_2 = \{(0, 0)\}$ so $\mathbb{R}^2 = W_1 \oplus W_2$.

More generally, if W_1 and W_2 are any two non-parallel lines passing through the origin in \mathbb{R}^2 , then $\mathbb{R}^2 = W_1 \oplus W_2$.

End of lecture 5

Example 7.4. Consider the vector space \mathbb{R}^3 . Let $W_1 = \{(x, y, 0) : x, y \in \mathbb{R}\}$ be the xy -plane, and let $W_2 = \{(x, 0, z) : x, z \in \mathbb{R}\}$ the xz -plane. We can show that $W_1 + W_2 = \mathbb{R}^3$. However, $W_1 \cap W_2 = \{(x, 0, 0) : x \in \mathbb{R}\}$ is the x -axis. As a result, \mathbb{R}^3 is not the direct sum of W_1 and W_2 .

We note that a vector like $(1, 1, 1) \in \mathbb{R}^3$ can be written as two different sums of an element in W_1 and an element in W_2 .

$$(1, 1, 1) = (1, 1, 0) + (0, 0, 1)$$

$$(1, 1, 1) = (0, 1, 0) + (1, 0, 1)$$

The lack of a unique representation for $(1, 1, 1)$ is an important indicator that a vector space is not a direct sum of the given subspaces. We will revisit this concept in Proposition 7.6.

Example 7.5. Consider the vector space $F(\mathbb{R}, \mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R}\}$ of functions from \mathbb{R} to \mathbb{R} . Define the following subspaces of even functions W_1 and odd functions W_2 .

$$W_1 = \{f \in F(\mathbb{R}, \mathbb{R}) : f(-x) = f(x)\}$$

$$W_2 = \{f \in F(\mathbb{R}, \mathbb{R}) : f(-x) = -f(x)\}$$

We claim that $F(\mathbb{R}, \mathbb{R}) = W_1 + W_2$. Let $f \in F(\mathbb{R}, \mathbb{R})$. Define $g, h \in F(\mathbb{R}, \mathbb{R})$ as follows.

$$g(x) := \frac{f(x) + f(-x)}{2}$$

$$h(x) := \frac{f(x) - f(-x)}{2}$$

Then

$$\begin{aligned} g(x) + h(x) &= \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2} \\ &= \frac{f(x)}{2} + \frac{f(x)}{2} \\ &= f(x). \end{aligned}$$

Note that

$$g(-x) = \frac{f(-x) + f(-(-x))}{2} = \frac{f(-x) + f(x)}{2} = g(x)$$

so $g \in W_1$. Similarly,

$$h(-x) = \frac{f(-x) - f(-(-x))}{2} = -\frac{f(x) - f(-x)}{2} = -h(x)$$

so $h \in W_2$. Since $f = g + h$, this shows that $F(\mathbb{R}, \mathbb{R}) = W_1 + W_2$.

We claim that $F(\mathbb{R}, \mathbb{R}) = W_1 \oplus W_2$. It remains to show that $W_1 \cap W_2 = \{0\}$ where $0 : \mathbb{R} \rightarrow \mathbb{R}$ is the zero function. Suppose that $f \in W_1 \cap W_2$. Then $f(x) = f(-x)$ and $f(x) = -f(-x)$ for all $x \in \mathbb{R}$. Thus $f(-x) = -f(-x)$ for all $x \in \mathbb{R}$. The only number satisfying this property is 0 so $f(x) = f(-x) = 0$ for all $x \in \mathbb{R}$. Therefore, $W_1 \cap W_2 = \{0\}$, and $F(\mathbb{R}, \mathbb{R}) = W_1 \oplus W_2$.

The following proposition gives another perspective on the meaning of a direct sum.

Proposition 7.6. Let V be a vector space and let $W_1, W_2 \subset V$ be subspaces. Then $V = W_1 \oplus W_2$ if and only if each element of v can be uniquely written as $w_1 + w_2$ for some $w_1 \in W_1, w_2 \in W_2$.

Proof. (\Rightarrow) Suppose that $V = W_1 \oplus W_2$. Fix $v \in V$. Since $V = W_1 \oplus W_2$, $V = W_1 + W_2$ so there exist $w_1 \in W_1$ and $w_2 \in W_2$ such that $v = w_1 + w_2$. We need to show that this decomposition is unique. Suppose that there is another decomposition $v = w'_1 + w'_2$ such that $w'_1 \in W_1$ and $w'_2 \in W_2$. Then $w_1 + w_2 = w'_1 + w'_2$, which implies $w_1 - w'_1 = w'_2 - w_2$. Because W_1 and W_2 are closed under addition, the left side of the equality is an element in W_1 and the right side is an element in W_2 . Thus the equality represents an element of $W_1 \cap W_2$. Since $V = W_1 \oplus W_2$, $W_1 \cap W_2 = \{0\}$ so $w_1 - w'_1 = 0 = w'_2 - w_2$. Therefore, $w_1 = w'_1$ and $w_2 = w'_2$. The decomposition $v = w_1 + w_2$ is unique.

(\Leftarrow) Suppose that $W_1, W_2 \subset V$ are subspaces such that every $v \in V$ has a unique decomposition $v = w_1 + w_2$ for some $w_1 \in W_1$ and $w_2 \in W_2$. We need to show that $V = W_1 \oplus W_2$. Since every element of v admits a decomposition $v = w_1 + w_2$, it follows that $V = W_1 + W_2$. It remains to show that $W_1 \cap W_2 = \{0\}$. Suppose that $v \in W_1 \cap W_2$. Note that

$$v = v + 0$$

is a decomposition of v since $v \in W_1$ and $0 \in W_2$. On the other hand, we can view

$$v = 0 + v$$

as a decomposition with $w_1 = 0 \in W_1$ and $w_2 = v \in W_2$. By assumption, the decomposition is unique so $v = 0$. Thus $W_1 \cap W_2 = \{0\}$, and $V = W_1 \oplus W_2$. \square

We can generalize the notion of direct sum to multiple subspaces.

Definition 7.7. Let W_1, \dots, W_k be subspaces of a vector space V . Define

$$W_1 + \dots + W_k := \{w_1 + \dots + w_k : w_j \in W_j\}.$$

We say that V is the **direct sum** of W_1, \dots, W_k , written

$$V = W_1 \oplus \dots \oplus W_k,$$

if $V = W_1 + \dots + W_k$ and $W_i \cap (W_1 + \dots + W_{i-1} + W_{i+1} + \dots + W_k) = \{0\}$ for all $1 \leq i \leq k$.

8. LINEAR COMBINATIONS AND SPAN

Definition 8.1. Let V be a vector space over a field \mathcal{F} . A **linear combination** of vectors $v_1, \dots, v_k \in V$ is an element of the form

$$a_1 v_1 + \dots + a_k v_k \in V$$

where $a_1, \dots, a_n \in \mathcal{F}$.

Example 8.2. Consider the real vector space \mathbb{R}^2 . The element $(1, 6) \in \mathbb{R}^2$ is a linear combination of $(1, 2)$ and $(-1, 0)$ as follows.

$$3 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \end{pmatrix}$$

Example 8.3. Consider the vector space $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \rightarrow \mathbb{R}$. The element $2x + 3 \sin x$ is a linear combination of $x \in F(\mathbb{R}, \mathbb{R})$ and $\sin x \in F(\mathbb{R}, \mathbb{R})$.

Definition 8.4. Let V be a vector space, and let $S \subset V$ a non-empty subset. The **span** of S , denoted $\text{Span}(S)$, is the set of linear combinations of all elements of S ,

$$\text{Span}(S) := \{a_1 v_1 + \dots + a_k v_k : a_j \in \mathcal{F}, v_j \in S\}.$$

The the span of the empty set, $\text{Span}(\emptyset)$, is defined to be $\{0\}$.

End of lecture 6

Example 8.5. Consider the vector space \mathbb{R}^3 and the set $S = \{(1, 0, 0), (0, 1, 0)\}$. The span of S is geometrically described as the xy -plane. Formally,

$$\begin{aligned} \text{Span}(S) &= \left\{ a_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} : a_1, a_2 \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} : a_1, a_2 \in \mathbb{R} \right\}. \end{aligned}$$

Taking the span of a set of vectors is an easy way to build a subspace in a vector space.

Proposition 8.6. Let V be a subspace, and let $S \subset V$ be a subset. Then $\text{Span}(S)$ is a subspace of V . Moreover, it is the smallest subspace containing S .

Proof. First, we show that $\text{Span}(S)$ is a subspace. Let $s \in S$ be any element. Then

$$0 = 0 \cdot s \in \text{Span}(S).$$

Let $v, w \in \text{Span}(S)$. By Definition 8.4, there exist $s_1, \dots, s_k \in S$ and $a_1, \dots, a_k \in \mathcal{F}$ such that

$$v = a_1 s_1 + \dots + a_k s_k.$$

Likewise, there exist $s'_1, \dots, s'_n \in S$ and $a'_1, \dots, a'_n \in \mathcal{F}$ such that

$$w = a'_1 s'_1 + \dots + a'_n s'_n.$$

Then

$$v + w = a_1 s_1 + \dots + a_k s_k + a'_1 s'_1 + \dots + a'_n s'_n$$

so $v + w$ is a linear combination of elements in S . Thus $v + w \in \text{Span}(S)$. Let $v \in \text{Span}(S)$ and $\lambda \in \mathcal{F}$. As before, there exist $s_1, \dots, s_k \in S$ and $a_1, \dots, a_k \in \mathcal{F}$ such that

$$v = a_1 s_1 + \dots + a_k s_k.$$

Then

$$\lambda v = (\lambda a_1)s_1 + \cdots + (\lambda a_k)s_k.$$

Thus λv is a linear combination of elements of S so $\lambda v \in \text{Span}(S)$. By Proposition 6.3, it follows that $\text{Span}(S)$ is a subspace.

Next, we prove that $\text{Span}(S)$ is the smallest subspace containing S . Suppose that W is a subspace such that $S \subset W$. We will show that $\text{Span}(S) \subset W$. Let $v \in \text{Span}(S)$. There exist $s_1, \dots, s_k \in S$ and $a_1, \dots, a_k \in \mathcal{F}$ such that

$$v = a_1 s_1 + \cdots + a_k s_k.$$

Note that W is closed under addition and scalar multiplication. Since each $s_j \in S \subset W$, the linear combination $v = a_1 s_1 + \cdots + a_k s_k \in W$. Thus $\text{Span}(S) \subset W$. \square

Corollary 8.6.1. Let W be a subspace of a vector space V that contains set S . Then $\text{Span}(S) \subset W$.

Definition 8.7. A subset S of a vector space V **generates** (or **spans**) V if $\text{Span}(S) = V$. In this case, we also say that the vectors of S generate V .

We can start to describe vector spaces with infinitely many elements in terms of linear combinations of finitely many elements. In the case of \mathbb{R}^2 , we often describe each vector as a linear combination of $(1, 0)$ and $(0, 1)$. The entire xy -plane is encapsulated by only two vectors.

Example 8.8. We will show that the matrices

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

generate $M_{2 \times 2}(\mathbb{R})$. In other words, an arbitrary matrix $A \in M_{2 \times 2}(\mathbb{R})$ can be written as a linear combination of the four matrices. We have

$$\begin{aligned} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= \left(\frac{1}{3}a_{11} + \frac{1}{3}a_{12} + \frac{1}{3}a_{21} - \frac{2}{3}a_{22} \right) \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \\ &+ \left(\frac{1}{3}a_{11} + \frac{1}{3}a_{12} - \frac{2}{3}a_{21} + \frac{1}{3}a_{22} \right) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ &+ \left(\frac{1}{3}a_{11} - \frac{2}{3}a_{12} + \frac{1}{3}a_{21} + \frac{1}{3}a_{22} \right) \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \\ &+ \left(-\frac{2}{3}a_{11} + \frac{1}{3}a_{12} + \frac{1}{3}a_{21} + \frac{1}{3}a_{22} \right) \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}. \end{aligned}$$

However, the matrices

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

do not generate $M_{2 \times 2}(\mathbb{R})$ because any linear combination will have equal diagonal entries.

9. LINEAR INDEPENDENCE

Next, we discuss linear independence, a concept that should be familiar from linear algebra in \mathbb{R}^n . Our goal is to describe sets of vectors for which there are no redundancies. Each vector in a linearly independent set provides information about a “new” direction in the vector space.

Recall from Example 8.5 that if $S = \{(1, 0, 0), (0, 1, 0)\} \subset \mathbb{R}^3$, then

$$\text{Span}(S) = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} : a_1, a_2 \in \mathbb{R} \right\}.$$

On the other hand, let $S' = \{(1, 0, 0), (0, 1, 0), (1, 1, 0)\}$ where we include the vector $(1, 1, 0)$ to S . The span does not change. That is,

$$\text{Span}(S') = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} : a_1, a_2 \in \mathbb{R} \right\}.$$

Roughly this is because the third vector is *already* a linear combination of the first two so it doesn't add "new information." Equivalently, there is a *non-trivial* linear combination of the three vectors that produces the zero vector.

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Definition 9.1. Let S be a subset of a vector space V . We say that S is or the elements of S are **linearly dependent** if there are vectors $v_1, \dots, v_n \in S$ and scalars $a_1, \dots, a_n \in \mathcal{F}$ such that

$$a_1 v_1 + \dots + a_n v_n = 0$$

for some a_i non-zero. A set S is **linearly independent** if it is not linearly dependent.

Note that we have made no assumption about S being finite! The definition holds for sets of infinite vectors.

Example 9.2. Consider the vector space $P(\mathbb{R})$ of polynomials with coefficients in \mathbb{R} . Define

$$S = \{1 + x, x - 2x^2, 1 + 2x^2\}.$$

Then S is linearly dependent because

$$1(1 + x) - 1(x - 2x^2) - 1(1 + 2x^2) = 0.$$

On the other hand, the set $S' = \{1, x, x^2\}$ is linearly independent. Suppose there were a linear combination of these vectors that produce the zero vector. That is, there are constants a_0, a_1, a_2 such that

$$a_0 + a_1 x + a_2 x^2 = 0.$$

By an elementary fact about polynomials, the only way this is possible is if $a_i = 0$ for $0 \leq i \leq 2$.

End of lecture 7

Remark 9.3. We will also write $P_k(\mathbb{R})$ to be the vector space of polynomials of degree less than or equal to k with coefficients in \mathbb{R} . Example 9.2 would also work in the vector space $P_2(\mathbb{R})$.

Example 9.4. Suppose that $0 \in S$. We have the following non-trivial linear combination of elements in S that produces the zero vector.

$$1 \cdot 0 = 0$$

Thus S is linearly dependent. Any set containing the zero vector is linearly dependent.

Example 9.2 leads to an important equivalent way to think about linear independence, which is essentially a reformulation of Definition 9.1.

Proposition 9.5. A set $S \subset V$ is linearly independent if and only if, for any $v_1, \dots, v_n \in S$,

$$a_1 v_1 + \dots + a_n v_n = 0 \quad \Rightarrow \quad a_1 = \dots = a_n = 0.$$

Example 9.6. Let $v \in V$ be a non-zero vector. Then $\{v\}$ is a linearly independent set since $av = 0$ implies $a = 0$.

Example 9.7. Consider the vector space $F(\mathbb{R}, \mathbb{R})$. The vectors $\sin x$ and $\cos x$ are linearly independent. Indeed, suppose

$$a \sin x + b \cos x = 0$$

for some constants a, b . Note that this is an equality of *functions*, not of numbers! Thus

$$a \sin x + b \cos x = 0$$

for all $x \in \mathbb{R}$. Plugging in $x = 0$ gives $b = 0$, and plugging in $x = \frac{\pi}{2}$ gives $a = 0$. Therefore, $\sin x$ and $\cos x$ are linearly independent.

Proposition 9.8. Let V be a vector space, and let $S_1 \subset S_2 \subset V$. If S_1 is linearly dependent, then S_2 is linearly dependent.

Proof. Since S_1 is linearly dependent, there is some linear combination $a_1 v_1 + \cdots + a_k v_k = 0$ for $v_i \in S$ and non-zero scalars a_i . Each v_i is also in S_2 so the linear combination $a_1 v_1 + \cdots + a_k v_k = 0$ is a non-trivial representation of 0 with vectors in S_2 . Thus S_2 is linearly dependent. \square

Corollary 9.8.1. Let V be a vector space, and let $S_1 \subset S_2 \subset V$. If S_2 is linearly independent, then S_1 is linearly independent.

Proposition 9.9. Let S be a linearly independent subset of a vector space V . Let v be a vector in V that is not in S . Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \text{Span}(S)$.

Proof. (\Rightarrow) Assume that $S \cup \{v\}$ is linearly dependent. Then there are vectors u_1, u_2, \dots, u_k in $S \cup \{v\}$ such that $a_1 u_1 + \cdots + a_k u_k = 0$ for some non-zero scalars a_i . If each $u_i \in S$, then $a_i = 0$ for all $1 \leq i \leq k$ by S linearly independent. Therefore, one of the u_i , say u_1 , is equal to v . We can write

$$v = -a_1^{-1}(a_2 u_2 + \cdots + a_k u_k)$$

so $v \in \text{Span}(S)$ as desired.

(\Leftarrow) Assume $v \in \text{Span}(S)$. Then $v = a_1 u_1 + \cdots + a_k u_k$ for $u_i \in S$ and scalars a_i . We can write $0 = a_1 u_1 + \cdots + a_k u_k - v$. The set $\{u_1, \dots, u_k, v\}$ is linearly dependent so $S \cup \{v\}$ is linearly dependent by Proposition 9.8. \square

End of lecture 8

10. BASES AND DIMENSION

The idea of linear independence was motivated by the desire to identify a “minimal spanning set.” In words, given a vector space or subspace, could we find a set S with as few elements as possible that span the space?

Definition 10.1. Let V be a vector space. A **basis** of V is a linearly independent set \mathcal{B} such that $\text{Span}(\mathcal{B}) = V$.

Proposition 10.2. Let V be a vector space, and let $\mathcal{B} = \{u_1, \dots, u_n\}$ be a subset of V . Then \mathcal{B} is a basis of V if and only if every element $v \in V$ admits a unique decomposition of the form

$$v = a_1 u_1 + \cdots + a_n u_n$$

for some $a_1, \dots, a_n \in \mathcal{F}$.

Proof. (\Rightarrow) Suppose that \mathcal{B} is a basis. Let $v \in V$. Since $\text{Span}(\mathcal{B}) = V$, there exist scalars a_1, \dots, a_n in \mathcal{F} such that

$$v = a_1 u_1 + \cdots + a_n u_n.$$

Suppose that $v = a'_1 u_1 + \cdots + a'_n u_n$ is another such decomposition. Then since

$$a_1 u_1 + \cdots + a_n u_n = a'_1 u_1 + \cdots + a'_n u_n,$$

we have

$$(a_1 - a'_1)u_1 + \cdots + (a_n - a'_n)u_n = 0.$$

Since $\{u_1, \dots, u_n\}$ is linearly independent, it follows that $a_1 - a'_1 = 0, \dots, a_n - a'_n = 0$. Thus $a'_1 = a_1, \dots, a'_n = a_n$ so the above decomposition is unique.

(\Leftarrow) Assume that every element $v \in V$ has a unique decomposition of the form $a_1u_1 + \cdots + a_nu_n$. Since such a representation exists for each $v \in V$, we have $V = \text{Span}(\mathcal{B})$. For the zero vector, we always have the representation $0u_1 + \cdots + 0u_n$. By assumption, the representation is unique. Proposition 9.5 proves that \mathcal{B} is linearly independent and, thus, a basis for V . \square

Example 10.3. Let $e_i \in \mathcal{F}^n$ be the vector $(0, \dots, 0, 1, 0, \dots, 0)$ where the 1 is in the i th position. The *standard basis* for the vector space \mathcal{F}^n is the set $\mathcal{B} = \{e_i : 1 \leq i \leq n\}$. Clearly, each vector in \mathcal{F}^n can be written as a linear combination of the standard basis vectors. We need only check that \mathcal{B} is linearly independent. For the linear combination

$$a_1e_1 + \cdots + a_ne_n = 0,$$

we want to show that each a_i has to be 0. Writing both sides as vectors,

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The only linear combination of the vectors in \mathcal{B} that produces the 0 vector requires all 0 coefficients. We can check that each vector in \mathcal{F}^n can be written uniquely as a linear combination of \mathcal{B} .

Example 10.4. Every polynomial in $P(\mathbb{R})$ can be written as a linear combination of the elements of $\mathcal{B} = \{1, x, x^2, \dots\}$. The only way that $\sum_{i=0}^k a_ix^i = 0$ is if $a_i = 0$ for each i . Thus \mathcal{B} is linearly independent. We call \mathcal{B} the standard basis for $P(\mathbb{R})$.

Theorem 10.5 (Replacement Theorem). Let V be a vector space, and suppose that $S \subset V$ is a finite subset with n elements such that $\text{Span}(S) = V$. Let $L \subset V$ be a set of linearly independent vectors with m elements.

- (i) $m \leq n$
- (ii) There is a subset $S' \subset S$ of $n - m$ vectors such that

$$\text{Span}(L \cup S') = V.$$

Proof. ¹ Fix $n \geq 0$. We proceed by induction on m .

As a base case, consider $m = 0$. A subset $L \subset V$ with $m = 0$ elements is necessarily the empty set. Since $0 \leq n$, (i) holds. Further, $\text{Span}(L \cup S) = \text{Span}(S) = V$ so (ii) holds with $S' = S$.

Now we perform the inductive step. Suppose the claim is true for some $m \geq 0$. Let L be a set of linearly independent vectors with $m + 1$ elements. We wish to show that $m + 1 \leq n$ and there is a subset $S' \subset S$ with $n - (m + 1)$ elements such that $\text{Span}(L \cup S') = V$.

Write $L = \{v_1, \dots, v_{m+1}\}$. L is linearly independent so Corollary 9.8.1 implies $\{v_1, \dots, v_m\} \subset L$ is linearly independent. By the inductive hypothesis, we have $m \leq n$. Also by the inductive hypothesis, there is a subset $S' := \{u_1, \dots, u_{n-m}\} \subset S$ such that

$$\text{Span}(v_1, \dots, v_m, u_1, \dots, u_{n-m}) = V.$$

It may still be the case that $m = n$. Since $v_{m+1} \in V$, it follows that

$$v_{m+1} = a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_{n-m}u_{n-m}$$

¹This proof is optional.

for some scalars $a_j, b_j \in \mathcal{F}$. Suppose that $m = n$ or $n - m = 0$. Then v_{m+1} is a linear combination of v_1, \dots, v_m , which contradicts that fact that L is linearly independent. Thus $n - m > 0$ and $m + 1 \leq n$. This proves (i) in the statement of the theorem.

We need to identify a subset $S' \subset S$ with $n - (m + 1)$ elements such that $L \cup S'$ spans V . Note that at least one of the b_j coefficients must be non-zero. Possibly after reordering elements, assume that $b_1 \neq 0$. Let $S' = \{u_2, \dots, u_{n-m}\}$ so S' has $n - (m + 1)$ elements. We claim that

$$\text{Span}(L \cup S') = V.$$

Note that

$$\{v_1, \dots, v_m, u_2, \dots, u_{n-m}\} \subset \text{Span}(L \cup S').$$

Since

$$u_1 = \frac{1}{b_1}v_{m+1} - \frac{a_1}{b_1}v_1 - \dots - \frac{a_m}{b_1}v_m - \frac{b_2}{b_1}u_2 - \dots - \frac{b_{n-m}}{b_1}u_{n-m},$$

$u_1 \in \text{Span}(L \cup S')$. Therefore,

$$\{v_1, \dots, v_m, u_1, \dots, u_{n-m}\} \subset \text{Span}(L \cup S').$$

The span of the set on the left is V by the inductive hypothesis. Corollary 8.6.1 implies that V is a subset $\text{Span}(L \cup S')$. \square

Corollary 10.5.1. Let V be a vector space with a finite basis. Then every basis of V contains the same number of elements.

*Proof.*² Let \mathcal{B} be a basis of V with n elements. Let \mathcal{B}' be another basis. We wish to show that \mathcal{B}' also has n elements. Suppose that \mathcal{B}' has more than n elements (possibly infinitely many). Let S be a subset of \mathcal{B}' with $n + 1$ elements. Since \mathcal{B}' is linearly independent, S is linearly independent by Corollary 9.8.1. Since \mathcal{B} is a spanning set of V , Replacement Theorem implies $n + 1 \leq |\mathcal{B}| = n$, a contradiction. Thus \mathcal{B}' cannot have more than n elements.

Now let m denote the (finite) number of elements of \mathcal{B}' . By above, $m \leq n$. Reverse the roles of \mathcal{B} and \mathcal{B}' . Since $\text{Span}(\mathcal{B}') = V$ and \mathcal{B} is linearly independent, Replacement Theorem implies that $n \leq m$. We conclude $n = m$. \square

Using Corollary 10.5.1, we can now formally define the notion of *dimension*. Dimension assigns to each vector space a number. The number indicates the “size” of the vector space or, more specifically, how many vectors we need to uniquely describe each element.

End of lecture 9

Definition 10.6. A vector space V is **finite dimensional** if it has a basis with finitely many elements. The number of vectors in any basis for a finite dimensional vector space V over \mathcal{F} is the **dimension** of V , denoted $\dim_{\mathcal{F}}(V)$. If a vector space does not have a finite basis, it is **infinite dimensional**.

Remark 10.7. When it is clear to which field we are referring, we will drop the subscript part of the dimension notation.

Example 10.8. The vector space $\{0\}$ has dimension 0 because $\text{Span}(\emptyset) = \{0\}$ by definition.

Example 10.9. The dimension of \mathcal{F}^n is n since the standard basis from Example 10.3 has n elements.

Example 10.10. The vector space $P(\mathcal{F})$ is infinite dimensional because the set $\{1, x, x^2, \dots\}$ is a basis by Example 10.4.

²This proof is also optional.

Example 10.11. Recall that \mathbb{C} can be given a vector space structure in two different ways: as a real vector space (over \mathbb{R}) and as a complex vector space (over \mathbb{C}).

- (1) View \mathbb{C} as a vector space over \mathbb{C} . According to Example 10.9 with $\mathcal{F} = \mathbb{C}$ and $n = 1$, the vector space has dimension 1 with possible basis $\{1\}$. Thus $\dim_{\mathbb{C}}(\mathbb{C}) = 1$.
- (2) View \mathbb{C} as a vector space over \mathbb{R} . Then $\{1, i\}$ is a basis for \mathbb{C} so \mathbb{C} has dimension 2. Indeed, any complex number $a + bi \in \mathbb{C}$ is a linear combination of 1 and i via

$$a + bi = (a)1 + (b)i.$$

We now show that $\{1, i\}$ is linearly independent over \mathbb{R} . Suppose that $a_1, a_2 \in \mathbb{R}$ are real numbers such that

$$a_1 \cdot 1 + a_2 \cdot i = 0.$$

Then $a_1 + a_2 i = 0$ implies $a_1 = a_2 = 0$, and $\{1, i\}$ is linearly independent. Therefore, $\{1, i\}$ is a basis for \mathbb{C} as a vector space over \mathbb{R} . We have $\dim_{\mathbb{R}}(\mathbb{C}) = 2$.

The next two results prove that, in a finite dimensional vector space, a spanning set can be reduced to a basis and a linearly independent set can be extended to a basis. If we have a linearly independent or spanning set, we need only count the number of elements to check whether the set is a basis for a finite dimensional vector space.

Proposition 10.12. If a vector space V is generated by a finite set S , then some subset of S is a basis for V . Hence V has a finite basis.

Proof. If $S = \emptyset$ or $S = \{0\}$, then $V = \{0\}$. We note that $\emptyset \subset S$ is a basis for V . Otherwise S contains a non-zero vector u_1 . By Example 9.6, $\{u_1\}$ is a linearly independent set. Continue, if possible, choosing vectors u_2, \dots, u_k in S such that $\{u_1, u_2, \dots, u_k\}$ is a linearly independent set of k vectors. Since S is finite, the process must terminate with a linearly independent set $\mathcal{B} = \{u_1, u_2, \dots, u_k\}$. There are two ways this could happen.

(1) The set $\mathcal{B} = S$. In this case, S is both a linearly independent set and a generating set for V . That is, S is a basis for V .

(2) The set \mathcal{B} is a proper linearly independent subset of S such that adjoining to \mathcal{B} any new vector in S produces a linearly dependent set. We claim that $\mathcal{B} \subset S$ is a basis for V . Since \mathcal{B} is linearly independent by construction, it suffices to show that \mathcal{B} spans V . If we can show that $S \subset \text{Span}(\mathcal{B})$, then $V = \text{Span}(S) \subset \text{Span}(\mathcal{B})$ by Corollary 8.6.1. Let $v \in S$. If $v \in \mathcal{B}$, then clearly $v \in \text{Span}(\mathcal{B})$. Otherwise, $v \notin \mathcal{B}$, and $\mathcal{B} \cup \{v\}$ is linearly dependent by assumption. Thus $v \in \text{Span}(\mathcal{B})$ by Proposition 9.9. We conclude that $S \subset \text{Span}(\mathcal{B})$. \square

Proposition 10.13. Let V be a vector space with dimension n .

- (i) If $S \subset V$ is a finite spanning set of V , then $|S| \geq n$. If $|S| = n$, then S is a basis.
- (ii) If L is a linearly independent subset of V such that $|L| = n$, then L is a basis.
- (iii) If L is a linearly independent subset of V , then L can be extended to a basis (that is, there is a basis \mathcal{B} such that $L \subset \mathcal{B}$).

Proof. Let \mathcal{B} be a basis of V , which will have n elements by Corollary 10.5.1.

- (i) Let S be a finite generating set for V . By Proposition 10.12, some subset H of S is a basis for V . Corollary 10.5.1 implies that H contains exactly n vectors. Therefore, S must contain at least n vectors. If S contains exactly n vectors, then we must have $H = S$ and S is a basis for V .
- (ii) Let L be a linearly independent subset of V containing exactly n vectors. Replacement Theorem proves there is a subset H of \mathcal{B} containing $n - n = 0$ vectors such that $L \cup H$ generates V . Thus $H = \emptyset$, and L generates V . Since L is also linearly independent, L is a basis for V .

- (iii) If L is a linearly independent subset of V containing m vectors, then Replacement Theorem asserts that there is a subset H of \mathcal{B} containing exactly $n - m$ vectors such that $L \cup H$ generates V . Thus $L \cup H$ contains at most n vectors. By (i), the spanning set $L \cup H$ contains at least n vectors. Therefore, $L \cup H$ contains exactly n vectors so $L \cup H$ is a basis for V by (i).

□

We can briefly summarize the contents of the Replacement Theorem, Proposition 10.12, and Proposition 10.13. Let V be a vector space of dimension n . Then linearly independent sets have size $\leq n$, generating sets have size $\geq n$, and bases have size n . Any linearly independent set can be made into a basis by strategically adjoining elements until we have size n . Any spanning set can be made into a basis by strategically removing elements until we have size n .

As a point of caution, not all subsets of V of size $\leq n$ are linearly independent. Likewise, not all subsets of V of size $\geq n$ are spanning sets. We always need to show at least one of linear independence or span.

The following is an example of how we can reduce a spanning set to basis.

Example 10.14. Consider the vector space $P(\mathbb{R})$ of polynomials (of any degree) with coefficients in \mathbb{R} . Define the following vectors.

$$\begin{aligned} p_1(x) &= x + 2x^3 \\ p_2(x) &= 1 + x + x^2 \\ p_3(x) &= 3 + 4x + 3x^2 + 2x^3 \\ p_4(x) &= 0 \end{aligned}$$

Let $W := \text{Span}(\{p_1(x), p_2(x), p_3(x), p_4(x)\})$ so $S = \{p_1(x), p_2(x), p_3(x), p_4(x)\}$ is a generating set for W . By Example 9.4, $p_4(x) = 0 \in S$ implies that S is not linearly independent. We want to reduce the spanning set S to a basis for W . We will need to remove $p_4(x)$. Note further that

$$\begin{aligned} p_3(x) &= 3 + 4x + 3x^2 + 2x^3 \\ &= (x + 2x^3) + 3(1 + x + x^2) \\ &= p_1(x) + 3p_2(x). \end{aligned}$$

We guess that $\{p_1(x), p_2(x)\}$ is a basis for W .

Suppose $a_1, a_2 \in \mathbb{R}$ such that

$$\begin{aligned} a_1 p_1(x) + a_2 p_2(x) &= 0 \\ a_1 (x + 2x^3) + a_2 (1 + x + x^2) &= 0 \\ a_2 + (a_1 + a_2)x + a_2 x^2 + 2a_1 x^3 &= 0. \end{aligned}$$

For a polynomial to be the zero polynomial, all of its coefficients must be 0. We obtain the following system of linear equations.

$$\begin{aligned} a_2 &= 0 \\ a_1 + a_2 &= 0 \\ a_2 &= 0 \\ 2a_1 &= 0 \end{aligned}$$

The first and fourth equations imply that $a_1 = a_2 = 0$. Therefore, $\{p_1(x), p_2(x)\}$ is linearly independent.

Next, we show that $\text{Span}(\{p_1(x), p_2(x)\}) = W$. Let $v(x) \in W$. Then

$$v(x) = a_1p_1(x) + a_2p_2(x) + a_3p_3(x) + a_4p_4(x)$$

for some $a_1, a_2, a_3, a_4 \in \mathbb{R}$. By above,

$$\begin{aligned} v(x) &= a_1p_1(x) + a_2p_2(x) + a_3(p_1(x) + 3p_2(x)) + a_4 \cdot 0 \\ &= (a_1 + a_3)p_1(x) + (a_2 + 3a_3)p_2(x). \end{aligned}$$

Thus $v(x) \in \text{Span}(\{p_1(x), p_2(x)\})$, and $W \subset \text{Span}(\{p_1(x), p_2(x)\})$. Since $p_1(x), p_2(x) \in W$, Corollary 8.6.1 implies $\text{Span}(\{p_1(x), p_2(x)\})$ is a subset of W . Therefore, $W = \text{Span}(\{p_1(x), p_2(x)\})$. We conclude that $\{p_1(x), p_2(x)\}$ is a basis and $\dim(W) = 2$.

End of lecture 10

The following is an example of how we can extend a linearly independent set to a basis.

Example 10.15. Consider the vector space $P_3(\mathbb{R})$ of polynomials of degree less than or equal to 3 with coefficients in \mathbb{R} . Define $S = \{1, 1 + x, 1 + x^2\}$. We immediately know that S is not a basis for $P_3(\mathbb{R})$ since $|S| = 3 < 4 = \dim(P_3(\mathbb{R}))$. However, we can show that S is linearly independent as follows. Suppose $a_1, a_2, a_3 \in \mathbb{R}$ are such that

$$a_1 \cdot 1 + a_2 \cdot (1 + x) + a_3 \cdot (1 + x^2) = 0.$$

We group the terms based on degree to obtain

$$(a_1 + a_2 + a_3) \cdot 1 + a_2 \cdot x + a_3 \cdot x^2 = 0$$

and the system of equations

$$\begin{aligned} a_1 + a_2 + a_3 &= 0 \\ a_2 &= 0 \\ a_3 &= 0. \end{aligned}$$

Thus $a_1 = a_2 = a_3 = 0$ is the only possible solution. We conclude that S is linearly independent.

We will now extend S to a basis for $P_3(\mathbb{R})$. We want to adjoin some element of $P_3(\mathbb{R})$ that is not in $\text{Span}(S)$. Since the degree of each term of S is less than or equal to 2, an element of $P_3(\mathbb{R})$ of degree 3 should be a good choice. Define $S' = S \cup \{x^3\}$. Since $|S'| = 4 = \dim(P_3(\mathbb{R}))$, it is sufficient to show S' is linearly independent in order to show S' is a basis for $P_3(\mathbb{R})$ by Proposition 10.13(ii). Suppose $a_1, a_2, a_3, a_4 \in \mathbb{R}$ are such that

$$a_1 \cdot 1 + a_2 \cdot (1 + x) + a_3 \cdot (1 + x^2) + a_4 \cdot x^3 = 0.$$

We derive a system of linear equations whose only solution is $a_1 = a_2 = a_3 = a_4 = 0$. Therefore, S' is linearly independent and a basis for $P_3(\mathbb{R})$.

We can prove that subspaces of vector spaces cannot have larger dimension than the overall vector space. Further, a subspace of the same dimension as the overall vector space must be the whole thing. In other words, dimension is a well-behaved tool for judging the size of vector spaces.

Proposition 10.16. Let V be a finite dimensional vector space, and let $W \subset V$ be a subspace. Then W is finite dimensional, and $\dim(W) \leq \dim(V)$. Moreover, if $\dim(W) = \dim(V)$, then $W = V$.

Proof. Let $\dim(V) = n$. If $W = \{0\}$, then $\dim(W) \leq \dim(V)$. Assume $W \neq \{0\}$ so there is some non-zero vector $v_1 \in W$. Continue choosing vectors $v_i \in W$ so that $\{v_1, \dots, v_k\}$ is linearly independent. Since no set of linearly independent vectors can have more than n vectors, the process terminates at a point at which $k \leq n$. Further, adjoining any new vector from W produces

a linearly dependent set. By Proposition 9.9, $\{v_1, \dots, v_k\}$ spans W . Therefore, $\{v_1, \dots, v_k\}$ is a basis for W so $\dim(W) = k \leq n$.

If $\dim(W) = \dim(V)$, then \mathcal{B}_W is a linearly independent set of V with $\dim(V)$ vectors. Proposition 10.13(ii) implies that \mathcal{B}_W is a basis for V . Therefore, $V \subset \text{Span}(\mathcal{B}_W) = W$, and $V = W$. \square

We can also connect the concepts of basis and dimension to direct sums of subspaces.

Proposition 10.17. Assume that a vector space V can be written as $W_1 \oplus \dots \oplus W_k$ for finite dimensional subspaces W_1, \dots, W_k . Let $\mathcal{B}_1, \dots, \mathcal{B}_k$ be bases for W_1, \dots, W_k respectively. Then $\mathcal{B} := \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$ is a basis for $W_1 \oplus \dots \oplus W_k$.

Proof. We will proceed by induction on k . The result is trivial for $k = 1$. Assume the claim holds for k . We will prove it for $k + 1$. Note that $W_1 \oplus \dots \oplus W_{k+1} = (W_1 \oplus \dots \oplus W_k) \oplus W_{k+1}$. Then $\mathcal{C} := \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$ be a basis for $W_1 \oplus \dots \oplus W_k$ by the inductive hypothesis.

Define $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_{k+1}$. Let $v \in V$. Then $v = a_1 w_1 + \dots + a_{k+1} w_{k+1}$ for $w_i \in W_i$ and scalars a_i . We can write $a_1 w_1 + \dots + a_k w_k$ as a linear combination of vectors in \mathcal{C} and $a_{k+1} w_{k+1}$ as a linear combination of \mathcal{B}_{k+1} . Thus \mathcal{B} spans V .

Let $\mathcal{B}_i = \{v_1^i, \dots, v_{m_i}^i\}$. Assume $\sum_{i=1}^{k+1} \sum_{j=1}^{m_i} a_{i,j} v_j^i = 0$ for scalars $a_{i,j}$. We can rearrange the equation to obtain $\sum_{i=1}^k \sum_{j=1}^{m_i} a_{i,j} v_j^i = -(\sum_{j=1}^{m_{k+1}} a_{k+1,j} v_j^{k+1})$. Since $W_{k+1} \cap (W_1 + \dots + W_k) = \{0\}$,

$$\begin{aligned} \sum_{i=1}^k \sum_{j=1}^{m_i} a_{i,j} v_j^i &= 0 \\ \sum_{j=1}^{m_{k+1}} a_{k+1,j} v_j^{k+1} &= 0. \end{aligned}$$

The set \mathcal{C} is linearly independent so $a_{i,j} = 0$ for $1 \leq i \leq k$ and $1 \leq j \leq m_i$. Further, \mathcal{B}_{k+1} is linearly independent so $a_{k+1,j} = 0$. Therefore, \mathcal{B} is linearly independent, and \mathcal{B} is a basis for V . \square

Corollary 10.17.1. Let W_1, \dots, W_k be finite dimensional vector spaces. Then

$$\dim(W_1 \oplus \dots \oplus W_k) = \sum_{i=1}^k \dim(W_i).$$

11. LINEAR TRANSFORMATIONS

So far, we have studied linear algebra in the context of a single vector space. We want to expand our horizons and incorporate communication between two vector spaces. Perhaps we want to understand their similarities and differences. To do so, we need to study functions $f : V \rightarrow W$ whose domain is some vector space V and whose codomain is another vector space W . In order for this to be useful, the function needs to *respect the vector space structure*. We develop the definition of a *linear transformation*.

Definition 11.1. Let V, W be vector spaces over the field \mathcal{F} . A function $T : V \rightarrow W$ is a **linear transformation**, **linear map**, **linear function**, or simply **linear** if the following hold.

(i) For all $v, v' \in V$,

$$T(v + v') = T(v) + T(v').$$

(ii) For all $v \in V$ and $\lambda \in \mathcal{F}$,

$$T(\lambda v) = \lambda T(v).$$

Remark 11.2. Note that the addition and multiplication on the left side of each equality above occurs in V while the addition and multiplication on the right side of each equality occurs in W .

Remark 11.3. We will rarely mention the field \mathcal{F} . Whenever we discuss a linear transformation, it is assumed that the domain and codomain vector spaces are with respect to the same field.

Example 11.4. Consider the map $T : \mathbb{R} \rightarrow \mathbb{R}$ defined by $T(x) = 3x$. We can show that T is linear as follows.

$$\begin{aligned} T(x + y) &= 3(x + y) \\ &= 3x + 3y \\ &= T(x) + T(y) \\ T(\lambda x) &= 3(\lambda x) \\ &= \lambda \cdot 3x \\ &= \lambda T(x) \end{aligned}$$

for all $x, y \in \mathbb{R}$ and $\lambda \in \mathbb{R}$.

On the other hand, consider $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x) = 3x + 1$. The function f is *not* linear.

$$\begin{aligned} f(2 \cdot 1) &= 3(2 \cdot 1) + 1 = 7 \\ 2 \cdot f(1) &= 2(3 \cdot 1 + 1) = 8 \end{aligned}$$

We will consider more examples soon, but one important and identifying feature of a linear transformation is that it always takes the zero vector to the zero vector. This provides a way to immediately recognize that f in Example 11.4 is not linear.

Proposition 11.5. Suppose that $T : V \rightarrow W$ is linear. Then $T(0) = 0$.

Proof. We have $T(0) = T(0 \cdot 0) = 0 \cdot T(0) = 0$. □

Example 11.6. Let $A \in M_{m \times n}(\mathcal{F})$. Define $T : \mathcal{F}^n \rightarrow \mathcal{F}^m$ by $T(x) = Ax$ where Ax is the usual matrix-vector multiplication of an $m \times n$ matrix by an $n \times 1$ vector. Then T is a linear transformation.

Example 11.7. Let $T : P(\mathbb{R}) \rightarrow P(\mathbb{R})$ be the map given by $T(p)(x) = p'(x)$ for all $p \in P(\mathcal{F})$. In words, T is the map that eats a polynomial and spits out the derivative. We claim T is linear. Let $p, q \in P(\mathbb{R})$ and $\lambda \in \mathbb{R}$. By properties of the derivative from calculus, we have

$$\begin{aligned} T(p + q)(x) &= (p + q)'(x) \\ &= p'(x) + q'(x) \\ &= T(p)(x) + T(q)(x) \\ T(\lambda p)(x) &= (\lambda p)'(x) \\ &= \lambda p'(x) \\ &= \lambda T(p)(x). \end{aligned}$$

Therefore, T is linear.

End of lecture 11

Example 11.8. Let $0 : V \rightarrow W$ be the zero transformation. In other words, $0(v) = 0$ for all $v \in V$. Then 0 is linear.

Example 11.9. Let $I : V \rightarrow V$ be the identity map so $I(v) = v$ for all $v \in V$. Then I is linear.

Two important features of a linear transformation are its *kernel* (also called the *null space*) and *image* (also called the *range*).

Definition 11.10. Let $T : V \rightarrow W$ be a linear transformation. The **kernel** of T is the subset $\ker(T) \subset V$ defined by

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

The **image** of T is the set $\text{im}(T) \subset W$ defined by

$$\text{im}(T) := \{w \in W : w = T(v) \text{ for some } v \in V\}.$$

In words, the kernel is the set of elements in the domain that gets mapped to the 0 element of W . The image is the set of all outputs of T in W .

Remark 11.11. The textbook uses $N(T)$ in place of $\ker(T)$ and $R(T)$ in place of $\text{im}(T)$. The notation $N(T)$ stands for null space and $R(T)$ stands for range.

Proposition 11.12. Let $T : V \rightarrow W$ be linear. Then $\ker(T)$ is a subspace of V , and $\text{im}(T)$ is a subspace of W .

Proof. First, we show that $\ker(T)$ is a subspace of V . Note that $T(0_V) = 0_W$ by Proposition 11.5 so $0_V \in \ker(T)$. Next, suppose that $v, v' \in \ker(T)$. Then

$$T(v + v') = T(v) + T(v') = 0 + 0 = 0$$

so $v + v' \in \ker(T)$. Similarly, if $v \in \ker(T)$ and $\lambda \in \mathcal{F}$, then

$$T(\lambda v) = \lambda T(v) = \lambda \cdot 0 = 0$$

so $\lambda v \in \ker(T)$. Thus $\ker(T)$ is a subspace of V .

Next, we show that $\text{im}(T)$ is a subspace of W . Since $0_W = T(0_V)$ by Proposition 11.5, we have $0_W \in \text{im}(T)$. Next, suppose that $w, w' \in \text{im}(T)$. Then there are $v, v' \in V$ such that $w = T(v)$ and $w' = T(v')$. Thus

$$w + w' = T(v) + T(v') = T(v + v')$$

so $w + w' \in \text{im}(T)$. Similarly, if $w \in \text{im}(T)$ and $\lambda \in \mathcal{F}$, then $w = T(v)$ for some $v \in V$. We have

$$\lambda w = \lambda T(v) = T(\lambda v)$$

so $\lambda w \in \text{im}(T)$. Therefore, $\text{im}(T)$ is a subspace of W . □

Example 11.13. Let $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ be the linear transformation defined by

$$T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3).$$

Let $(a_1, a_2, a_3) \in \ker(T)$. Then $T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3) = (0, 0)$. We have

$$\begin{aligned} a_1 &= a_2 \\ 2a_3 &= 0 \end{aligned}$$

so $\ker(T) = \{(a, a, 0) : a \in \mathbb{R}\}$.

For a possible element of the image, $(b_1, b_2) \in \mathbb{R}^2$, we obtain the following system of equations.

$$\begin{aligned} a_1 - a_2 &= b_1 \\ 2a_3 &= b_2 \end{aligned}$$

A possible solution to the system is $a_1 = 1$, $a_2 = 1 - b_1$ and $a_3 = \frac{b_2}{2}$. Thus every element of \mathbb{R}^2 is in the image of T or $\text{im}(T) = \mathbb{R}^2$.

The next result simplifies the process of describing the image of a linear transformation.

Proposition 11.14. Let V and W be vector spaces with linear transformation $T : V \rightarrow W$. If $\mathcal{B} = \{v_1, \dots, v_n\}$ is a basis for V , then $\text{im}(T) = \text{Span}(\{T(v_1), \dots, T(v_n)\})$.

Proof. Clearly, $T(v_i) \in \text{im}(T)$ for each $1 \leq i \leq n$. By Corollary 8.6.1,

$$\text{Span}(\{T(v_1), \dots, T(v_n)\}) \subset \text{im}(T).$$

Suppose $w \in \text{im}(T)$. Then $w = T(v)$ for some $v \in V$. Since \mathcal{B} is a basis for V , we have $v = \sum_{i=1}^n a_i v_i$ for scalars a_i . By linearity of T ,

$$w = T(v) = \sum_{i=1}^n a_i T(v_i),$$

and $w \in \text{Span}(\{T(v_1), \dots, T(v_n)\})$. Thus $\text{im}(T) \subset \text{Span}(\{T(v_1), \dots, T(v_n)\})$. \square

Example 11.15. Return to Example 11.13. Let $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ be the linear transformation

$$T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3).$$

We can find a spanning set for $\text{im}(T)$ easily using Proposition 11.14. We pick the standard basis $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ for the domain \mathbb{R}^3 . Then

$$\begin{aligned} \text{im}(T) &= \text{Span}(\{T(1, 0, 0), T(0, 1, 0), T(0, 0, 1)\}) \\ &= \text{Span}(\{(1, 0), (-1, 0), (0, 2)\}) \\ &= \text{Span}(\{(1, 0), (0, 2)\}) \end{aligned}$$

by noticing the redundancy $(-1, 0) = -1 \cdot (1, 0)$. We can show that $\{(1, 0), (0, 2)\}$ is a basis for \mathbb{R}^2 so $\text{im}(T) = \mathbb{R}^2$.

End of lecture 12

Since $\ker(T)$ and $\text{im}(T)$ are subspaces, we can make the following definition.

Definition 11.16. Let V and W be vector spaces with linear transformation $T : V \rightarrow W$. Suppose that $\ker(T)$ and $\text{im}(T)$ are finite dimensional. The **nullity** of T is

$$\text{null}(T) := \dim(\ker(T)),$$

and the **rank** of T is

$$\text{rank}(T) := \dim(\text{im}(T)).$$

Theorem 11.17 (Rank-Nullity Theorem). Let V and W be vector spaces with linear transformation $T : V \rightarrow W$. Suppose that V is finite dimensional. Then

$$\dim(\ker(T)) + \dim(\text{im}(T)) = \dim(V).$$

Proof. Let $n = \dim(V)$ and $k = \dim(\ker(T))$.

Let $\{u_1, \dots, u_k\}$ be a basis for $\ker(T)$. Since $\{u_1, \dots, u_n\}$ is a basis of $\ker(T)$, it is a linearly independent set in V . We can extend the set to a basis $\{u_1, \dots, u_k, \tilde{u}_{k+1}, \dots, \tilde{u}_n\}$ of V by Proposition 10.13(iii). We claim that $\{T(\tilde{u}_{k+1}), \dots, T(\tilde{u}_n)\}$ is a basis for $\text{im}(T)$.

First, we prove linear independence. Suppose that

$$a_{k+1}T(\tilde{u}_{k+1}) + \dots + a_nT(\tilde{u}_n) = 0.$$

Then by linearity of T ,

$$T(a_{k+1}\tilde{u}_{k+1} + \dots + a_n\tilde{u}_n) = 0.$$

Thus $a_{k+1}\tilde{u}_{k+1} + \dots + a_n\tilde{u}_n \in \ker(T)$. Since $\{u_1, \dots, u_k\}$ is a basis for $\ker(T)$, there exists $a_1, \dots, a_k \in \mathcal{F}$ such that

$$a_{k+1}\tilde{u}_{k+1} + \dots + a_n\tilde{u}_n = a_1u_1 + \dots + a_ku_k,$$

which implies

$$a_1u_1 + \dots + a_ku_k - a_{k+1}\tilde{u}_{k+1} - \dots - a_n\tilde{u}_n = 0.$$

Since $\{u_1, \dots, u_k, \tilde{u}_{k+1}, \dots, \tilde{u}_n\}$ is linearly independent, $a_1 = \dots = a_k = -a_{k+1} = \dots = -a_n = 0$. In particular, $a_{k+1} = \dots = a_n = 0$ so $\{T(\tilde{u}_{k+1}), \dots, T(\tilde{u}_n)\}$ is linearly independent.

Next, we show that $\{T(\tilde{u}_{k+1}), \dots, T(\tilde{u}_n)\}$ spans $\text{im}(T)$. Let $w \in \text{im}(T)$. Then there exists some $v \in V$ such that $w = T(v)$. Since $\{u_1, \dots, u_k, \tilde{u}_{k+1}, \dots, \tilde{u}_n\}$ is a basis for V , there exist $b_1, \dots, b_n \in \mathcal{F}$ such that

$$v = b_1 u_1 + \dots + b_k u_k + b_{k+1} \tilde{u}_{k+1} + \dots + b_n \tilde{u}_n.$$

Since $u_j \in \ker(T)$ for $1 \leq j \leq k$,

$$\begin{aligned} w &= T(v) \\ &= T(b_1 u_1 + \dots + b_k u_k + b_{k+1} \tilde{u}_{k+1} + \dots + b_n \tilde{u}_n) \\ &= b_1 T(u_1) + \dots + b_k T(u_k) + b_{k+1} T(\tilde{u}_{k+1}) + \dots + b_n T(\tilde{u}_n) \\ &= b_{k+1} T(\tilde{u}_{k+1}) + \dots + b_n T(\tilde{u}_n). \end{aligned}$$

Thus $w \in \text{Span}(\{T(\tilde{u}_{k+1}), \dots, T(\tilde{u}_n)\})$, and $\text{im}(T) \subset \text{Span}(\{T(\tilde{u}_{k+1}), \dots, T(\tilde{u}_n)\})$. The opposite containment is clear. We conclude that $\{T(\tilde{u}_{k+1}), \dots, T(\tilde{u}_n)\}$ is a basis for $\text{im}(T)$ so

$$\dim(\text{im}(T)) = n - k.$$

Therefore,

$$\dim(\ker(T)) + \dim(\text{im}(T)) = k + (n - k) = n = \dim(V).$$

□

Example 11.18. Define the linear transformation $T : P_2(\mathbb{R}) \rightarrow M_{2 \times 2}(\mathbb{R})$ by

$$T(f(x)) = \begin{pmatrix} f(1) - f(2) & 0 \\ 0 & f(0) \end{pmatrix}.$$

Since $\mathcal{B} = \{1, x, x^2\}$ is a basis for $P_2(\mathbb{R})$, we have

$$\begin{aligned} \text{im}(T) &= \text{Span}(\{T(1), T(x), T(x^2)\}) \\ &= \text{Span}\left(\left\{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -3 & 0 \\ 0 & 0 \end{pmatrix}\right\}\right) \\ &= \text{Span}\left(\left\{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}\right\}\right) \end{aligned}$$

by Proposition 11.14. We can show that the above set is linearly independent so $\dim(\text{im}(T)) = 2$.

An element $p(x) \in \ker(T)$ is a polynomial for which $T(p(x))$ is the zero matrix. If we let $p(x) = a_0 + a_1 x + a_2 x^2$, then

$$\begin{aligned} T(p(x)) &= \begin{pmatrix} (a_0 + a_1 + a_2) - (a_0 + 2a_1 + 4a_2) & 0 \\ 0 & a_0 \end{pmatrix} \\ &= \begin{pmatrix} -a_1 - 3a_2 & 0 \\ 0 & a_0 \end{pmatrix}. \end{aligned}$$

We obtain the following system of linear equations by setting the matrix equal to the zero matrix.

$$\begin{aligned} -a_1 - 3a_2 &= 0 \\ a_0 &= 0 \end{aligned}$$

Therefore, $\ker(T) = \{3ax - ax^2 : a \in \mathbb{R}\}$ with possible basis $\{3x - x^2\}$. As Rank-Nullity Theorem predicts, $\dim(\ker(T)) = \dim(\mathbb{R}^3) - \dim(\text{im}(T)) = 3 - 2 = 1$.

Next, we recall the notions of *injectivity* (*one-to-one*) and *surjectivity* (*onto*). The following definition is very general, and holds for any function between sets.

Definition 11.19. Let X, Y be sets, and let $f : X \rightarrow Y$.

- (i) The function f is **injective** if $f(x) = f(x')$ implies $x = x'$.
- (ii) The function f is **surjective** if, for all $y \in Y$, there is an $x \in X$ such that $f(x) = y$.

Remark 11.20. We can unravel the definitions of injective and surjective to get the following equivalent notions.

- (i) A function $f : X \rightarrow Y$ is *not* injective if and only if there exist $x, x' \in X$ with $x \neq x'$ such that $f(x) = f(x')$.
- (ii) A function $f : X \rightarrow Y$ is surjective if and only if $\text{im}(f) = Y$.

End of lecture 13

If we know that a function is linear between two vector spaces, we can say even more about the injectivity relation.

Proposition 11.21. A linear map $T : V \rightarrow W$ is injective if and only if $\ker(T) = \{0\}$.

Proof. (\Rightarrow) Suppose that T is injective. Let $v \in \ker(T)$ so $T(v) = 0 = T(0)$. By injectivity of f , it follows that $v = 0$. Thus $\ker(T) = \{0\}$.

(\Leftarrow) Suppose that $\ker(T) = \{0\}$. Take $v, v' \in V$ such that $T(v) = T(v')$. By the linearity of T ,

$$\begin{aligned} T(v) - T(v') &= 0 \\ T(v - v') &= 0. \end{aligned}$$

Thus $v - v' \in \ker(T)$. The only element in $\ker(T)$ is the zero vector so

$$\begin{aligned} v - v' &= 0 \\ v &= v'. \end{aligned}$$

We conclude that T is injective. □

If we know that the domain and codomain are finite dimensional of the same dimension, we can find an even stronger relation between injectivity and surjectivity.

Proposition 11.22. Suppose V, W are finite dimensional vector spaces with $\dim(V) = \dim(W)$. Let $T : V \rightarrow W$ be linear. Then T is injective if and only if T is surjective.

Proof. (\Rightarrow) Suppose that T is injective. Then $\dim(\ker(T)) = 0$. By Rank-Nullity Theorem, it follows that $\dim(\text{im}(T)) = \dim(V)$. Since $\dim(W) = \dim(V)$, we have $\dim(\text{im}(T)) = \dim(W)$. Proposition 10.16 implies that the subspace $\text{im}(T)$ is all of W . Therefore, T is surjective.

(\Leftarrow) Suppose that T is surjective. Then $\text{im}(T) = W$ so $\dim(\text{im}(T)) = \dim(W) = \dim(V)$. Rank-Nullity Theorem implies $\dim(\ker(T)) = \dim(V) - \dim(\text{im}(T)) = 0$. Thus $\ker(T) = \{0\}$ so T is injective by Proposition 11.21. □

Finally, we prove a result that, intuitively, states the action of a linear transformation on a vector space is *entirely determined by what it does to a basis*.

Proposition 11.23. Let $T : V \rightarrow W$ be linear. Suppose that V is finite dimensional with a basis $\{u_1, \dots, u_n\}$. Let $S : V \rightarrow W$ be another linear transformation such that $S(u_j) = T(u_j)$ for all $1 \leq j \leq n$. Then $S = T$.

Proof. Let $v \in V$. Because $\{u_1, \dots, u_n\}$ is a basis for V , we can write $v = a_1u_1 + \dots + a_nu_n$ for some $a_1, \dots, a_n \in \mathcal{F}$. By linearity of S and T ,

$$\begin{aligned} S(v) &= S(a_1u_1 + \dots + a_nu_n) \\ &= a_1S(u_1) + \dots + a_nS(u_n) \\ &= a_1T(u_1) + \dots + a_nT(u_n) \\ &= T(a_1u_1 + \dots + a_nu_n) \\ &= T(v). \end{aligned}$$

Since $S(v) = T(v)$ for all $v \in V$, it follows that $S = T$. \square

Corollary 11.23.1. Let V and W be vector spaces, and suppose $\{v_1, \dots, v_n\}$ is a basis for V . For w_1, \dots, w_n in W , there exists exactly one linear transformation $T : V \rightarrow W$ such that $T(v_i) = w_i$ for $1 \leq i \leq n$.

Proof. Let $x \in V$. Then $x = \sum_{i=1}^n a_i v_i$ for unique $a_1, \dots, a_n \in \mathcal{F}$ by Proposition 10.2. Define

$$T(x) = \sum_{i=1}^n a_i w_i$$

so $T(v_i) = w_i$ for all $1 \leq i \leq n$. We will first prove that T is linear. Let $u, v \in V$ and $d \in \mathcal{F}$. We can write $u = \sum_{i=1}^n b_i v_i$ and $v = \sum_{i=1}^n c_i v_i$. Then $du + v = \sum_{i=1}^n (db_i + c_i) v_i$ so

$$\begin{aligned} T(du + v) &= T\left(\sum_{i=1}^n (db_i + c_i) v_i\right) \\ &= \sum_{i=1}^n (db_i + c_i) w_i \\ &= d \sum_{i=1}^n b_i w_i + \sum_{i=1}^n c_i w_i \\ &= dT(u) + T(v). \end{aligned}$$

Uniqueness follows from Proposition 11.23. \square

Example 11.24. Let $T : P_2(\mathbb{R}) \rightarrow P_3(\mathbb{R})$ be the linear transformation defined as

$$T(p)(x) = 2p'(x) + \int_0^x 3p(t)dt.$$

Since $\{1, x, x^2\}$ is a basis for $P_2(\mathbb{R})$,

$$\begin{aligned} \text{im}(T) &= \text{Span}(\{T(1), T(x), T(x^2)\}) \\ &= \text{Span}\left(\left\{3x, 2 + \frac{3}{2}x^2, 4x + x^3\right\}\right) \end{aligned}$$

by Proposition 11.14. We can show that the above set is linearly independent so

$$\text{rank}(T) = \dim(\text{im}(T)) = 3.$$

Since $\text{rank}(T) < 4 = \dim(P_3(\mathbb{R}))$, T is not surjective. By Rank-Nullity Theorem,

$$\text{null}(T) = \dim(P_2(\mathbb{R})) - \text{rank}(T) = 3 - 3 = 0.$$

Thus $\ker(T) = \{0\}$, and T is injective by Proposition 11.21.

End of lecture 14

End of midterm material

12. COORDINATES AND MATRIX REPRESENTATIONS

We have seen a number of vector spaces thus far. There are familiar vector spaces like

$$\mathbb{R}^3 = \left\{ \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} : a_j \in \mathbb{R} \right\}.$$

There are some vector spaces that, perhaps initially, were less familiar, like

$$P_2(\mathbb{R}) = \{a_0 + a_1x + a_2x^2 : a_j \in \mathbb{R}\}.$$

If you've thought to yourself *these two vector spaces basically feel like the same thing*, then you're correct! The next few topics we discuss will formalize this notion.

Definition 12.1. Let $\mathcal{B} = (u_1, \dots, u_n)$ be an (ordered) basis for a vector space V . Let $v \in V$, and let $v = a_1u_1 + \dots + a_nu_n$ be the unique expression of v as a linear combination of elements of \mathcal{B} . The **coordinate vector of v with respect to \mathcal{B}** is

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathcal{F}^n.$$

Remark 12.2. Note that the order of the elements in the basis matters! When we talk about coordinates, we need to keep track of this order, which is why the above definition refers to the basis as an “ordered basis.” We will use the parenthesis notation instead of the curly brace notation to indicate an ordered basis.

Example 12.3. Let $p(x) = 1 - 2x^2 \in P_2(\mathbb{R})$. Let $\mathcal{B} = (1, x, x^2)$. Then

$$\begin{aligned} p(x) &= 1 \cdot 1 + 0 \cdot x + (-2) \cdot x^2 \\ [p(x)]_{\mathcal{B}} &= \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}. \end{aligned}$$

We could rearrange the basis like $\mathcal{B}' = (1, x^2, x)$. Then

$$\begin{aligned} p(x) &= 1 \cdot 1 + (-2) \cdot x^2 + 0 \cdot x \\ [p(x)]_{\mathcal{B}'} &= \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}. \end{aligned}$$

If we pick the basis $\mathcal{C} = (1, x + x^2, x - x^2)$ for $P_2(\mathbb{R})$, we obtain

$$\begin{aligned} p(x) &= 1 \cdot 1 - 1 \cdot (x + x^2) + 1 \cdot (x - x^2) \\ [p(x)]_{\mathcal{C}} &= \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}. \end{aligned}$$

The point here is that, given an abstract (finite dimensional) vector space V , if we pick an ordered basis we can represent the elements of V as numerical vectors. This is an extremely powerful idea! We can go even further and find numerical representations (or matrices) of linear transformations.

Let V and W be finite dimensional vector spaces. Let $\mathcal{B} = (u_1, \dots, u_n)$ be a basis for V , and let $\mathcal{C} = (w_1, \dots, w_m)$ be a basis for W . For each $1 \leq j \leq n$, we can consider the vector $T(u_j) \in W$. Since \mathcal{C} is a basis for W , we have a unique decomposition

$$T(u_j) = a_{1j}w_1 + \dots + a_{mj}w_m$$

for some scalars a_{1j}, \dots, a_{mj} . We keep track of the j index in this list of coefficients because the scalars will depend on which vector u_j we feed to T .

Definition 12.4. The **coordinate matrix of T with respect to \mathcal{B} and \mathcal{C}** , denoted $[T]_{\mathcal{B}}^{\mathcal{C}}$, is the $m \times n$ matrix defined by

$$([T]_{\mathcal{B}}^{\mathcal{C}})_{ij} := a_{ij}.$$

Observe that the j th column of $[T]_{\mathcal{B}}^{\mathcal{C}}$ is $[T(u_j)]_{\mathcal{C}}$. The \mathcal{B} subscript is the basis corresponding to input coordinates while the \mathcal{C} superscript is the basis corresponding to output coordinates.

Example 12.5. Let $T : P_3(\mathbb{R}) \rightarrow P_2(\mathbb{R})$ be the derivative operator so $T(p)(x) = p'(x)$. Let $\mathcal{B} = (1, x, x^2, x^3)$ and $\mathcal{C} = (1, x, x^2)$ be (ordered) bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$, respectively. We will find $[T]_{\mathcal{B}}^{\mathcal{C}}$. We compute the output of T for each element of the input basis \mathcal{B} .

$$\begin{aligned} T(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x^2) &= 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ T(x^3) &= 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 \end{aligned}$$

Then we find the coordinates of each output with respect to the output basis \mathcal{C} .

$$\begin{aligned} [T(1)]_{\mathcal{C}} &= (0, 0, 0) \\ [T(x)]_{\mathcal{C}} &= (1, 0, 0) \\ [T(x^2)]_{\mathcal{C}} &= (0, 2, 0) \\ [T(x^3)]_{\mathcal{C}} &= (0, 0, 3) \end{aligned}$$

The order of the input basis \mathcal{B} determines the column order.

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Before we continue exploring coordinate representation of linear transformations, we have the following important fact.

Proposition 12.6. Let V and W be vector spaces with linear transformations $T : V \rightarrow W$ and $S : V \rightarrow W$. Let $\lambda \in F$. Then $T + S$ and λT are linear as well.

Proof. Let $v, v' \in V$ and $c \in F$. Then

$$\begin{aligned} (T + S)(v + v') &= T(v + v') + S(v + v') \\ &= T(v) + T(v') + S(v) + S(v') \\ &= [T(v) + S(v)] + [T(v') + S(v')] \\ &= (T + S)(v) + (T + S)(v') \\ (T + S)(cv) &= T(cv) + S(cv) \\ &= cT(v) + cS(v) \\ &= c(T + S)(v) \end{aligned}$$

Thus $T + S$ is linear.

Let $v, v' \in V$ and $c \in \mathcal{F}$. Then

$$\begin{aligned} (\lambda T)(v + v') &= \lambda T(v + v') \\ &= \lambda(T(v) + T(v')) \\ &= \lambda T(v) + \lambda T(v') \\ &= (\lambda T)(v) + (\lambda T)(v') \\ (\lambda T)(cv) &= \lambda T(cv) \\ &= \lambda(cT(v)) \\ &= (\lambda c)T(v) \\ &= c(\lambda T)(v) \end{aligned}$$

Thus λT is linear. □

Definition 12.7. We denote the set of linear transformations from V to W by $\mathcal{L}(V, W)$.

Corollary 12.7.1. The set $\mathcal{L}(V, W)$ is a vector space.

In the next two results, we prove that the process of taking a coordinate representation respects the linear structure of $\mathcal{L}(V, W)$.

Proposition 12.8. Let V and W be finite dimensional vector spaces with (ordered) bases \mathcal{B} and \mathcal{C} respectively. Let $T, S \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} [T + S]_{\mathcal{B}}^{\mathcal{C}} &= [T]_{\mathcal{B}}^{\mathcal{C}} + [S]_{\mathcal{B}}^{\mathcal{C}} \\ [\lambda T]_{\mathcal{B}}^{\mathcal{C}} &= \lambda [T]_{\mathcal{B}}^{\mathcal{C}} \end{aligned}$$

for all $\lambda \in \mathcal{F}$.

Proof. Let $\mathcal{B} = (u_1, \dots, u_n)$ and $\mathcal{C} = (w_1, \dots, w_m)$. Then

$$\begin{aligned} T(u_j) &= a_{1j}w_1 + \dots + a_{mj}w_m \\ S(u_j) &= b_{1j}w_1 + \dots + b_{mj}w_m \end{aligned}$$

for some scalars $a_{ij}, b_{ij} \in \mathcal{F}$. We write

$$\begin{aligned} (T + S)(u_j) &= (a_{1j} + b_{1j})w_1 + \dots + (a_{mj} + b_{mj})w_m \\ ([T + S]_{\mathcal{B}}^{\mathcal{C}})_{ij} &= a_{ij} + b_{ij} \\ &= ([T]_{\mathcal{B}}^{\mathcal{C}})_{ij} + ([S]_{\mathcal{B}}^{\mathcal{C}})_{ij} \end{aligned}$$

so $[T + S]_{\mathcal{B}}^{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}} + [S]_{\mathcal{B}}^{\mathcal{C}}$.

We have

$$T(u_j) = a_{1j}w_1 + \dots + a_{mj}w_m$$

for some scalars $a_{ij} \in \mathcal{F}$. Then

$$\begin{aligned} (\lambda T)(u_j) &= (\lambda a_{1j})w_1 + \dots + (\lambda a_{mj})w_m \\ ([\lambda T]_{\mathcal{B}}^{\mathcal{C}})_{ij} &= \lambda a_{ij} \\ &= \lambda ([T]_{\mathcal{B}}^{\mathcal{C}})_{ij} \end{aligned}$$

so $[\lambda T]_{\mathcal{B}}^{\mathcal{C}} = \lambda [T]_{\mathcal{B}}^{\mathcal{C}}$. □

End of lecture 15

We can rephrase the result in a more abstract way as follows.

Corollary 12.8.1. Let V and W be finite dimensional vector spaces. Fix ordered bases \mathcal{B} and \mathcal{C} of V and W respectively. Define $\phi : \mathcal{L}(V, W) \rightarrow M_{m \times n}(\mathcal{F})$ by $\phi(T) = [T]_{\mathcal{B}}^{\mathcal{C}}$. Then ϕ is a linear transformation.

When we pick a basis for the domain and a basis for the codomain, every linear transformation is identified with a matrix. Example 11.6 shows that each matrix produces a linear transformation. Soon we will be able to say that this construction proves ϕ is an *isomorphism*. Basically, $\mathcal{L}(V, W)$ and $M_{m \times n}(\mathcal{F})$ are the “same” as vector spaces!

13. COMPOSITION OF LINEAR MAPS

Definition 13.1. Let $T : V \rightarrow W$ and $S : W \rightarrow U$ be functions. The **composition of S and T** is the function $S \circ T : V \rightarrow U$ defined by $(S \circ T)(v) = S(T(v))$ for all $v \in V$.

Proposition 13.2. Let $T : V \rightarrow W$ and $S : W \rightarrow U$ be linear. Then $S \circ T : V \rightarrow U$ is linear.

Proof. Let $v, v' \in V$ and $c \in \mathcal{F}$. Then

$$\begin{aligned} (S \circ T)(v + v') &= S(T(v + v')) \\ &= S(T(v) + T(v')) \\ &= S(T(v)) + S(T(v')) \\ &= (S \circ T)(v) + (S \circ T)(v') \\ (S \circ T)(cv) &= S(T(cv)) \\ &= S(cT(v)) \\ &= cS(T(v)) \\ &= c(S \circ T)(v). \end{aligned}$$

Therefore, $S \circ T$ is linear. □

Proposition 13.3. Let T, S_1, S_2 be linear transformations. Assume that the respective domains and codomains are such that each of the compositions below makes sense.

- (i) $T \circ (S_1 + S_2) = T \circ S_1 + T \circ S_2$ and $(S_1 + S_2) \circ T = S_1 \circ T + S_2 \circ T$
- (ii) $(T \circ S_1) \circ S_2 = T \circ (S_1 \circ S_2)$
- (iii) $\lambda(S_1 \circ S_2) = (\lambda S_1) \circ S_2 = S_1 \circ (\lambda S_2)$

Proof. (i) We will prove the first statement. Let $S_1, S_2 : U \rightarrow V$ and $T : V \rightarrow W$. Let $u \in U$. By linearity of T ,

$$\begin{aligned} (T \circ (S_1 + S_2))(u) &= T(S_1(u) + S_2(u)) \\ &= T(S_1(u)) + T(S_2(u)) \\ &= (T \circ S_1)(u) + (T \circ S_2)(u). \end{aligned}$$

Thus $T \circ (S_1 + S_2) = T \circ S_1 + T \circ S_2$.

(ii) Let $S_2 : U \rightarrow V$, $S_1 : V \rightarrow W$, and $T : W \rightarrow X$. Let $u \in U$. Then

$$\begin{aligned} ((T \circ S_1) \circ S_2)(u) &= (T \circ S_1)(S_2(u)) \\ &= T(S_1(S_2(u))) \\ &= T((S_1 \circ S_2)(u)) \end{aligned}$$

so $(T \circ S_1) \circ S_2 = T \circ (S_1 \circ S_2)$.

- (iii) We will prove the first equality. Let $S_2 : U \rightarrow V$ and $S_1 : V \rightarrow W$. Let $u \in U$. By linearity of S_1 ,

$$\begin{aligned}\lambda(S_1 \circ S_2)(u) &= \lambda S_1(S_2(u)) \\ &= (\lambda S_1)(S_2(u)).\end{aligned}$$

Thus $\lambda(S_1 \circ S_2) = (\lambda S_1) \circ S_2$.

□

Remark 13.4. Note that composition is not commutative. If $S \circ T$ is defined, the other composition $T \circ S$ may not even be well-defined, let alone equal to $S \circ T$!

If we have two linear maps T and S between finite dimensional vector spaces, can we determine a relationship between the coordinate matrices of T , S , and $S \circ T$? Indeed, we have the following result, which is the underlying motivation for the familiar and, initially, odd definition of matrix multiplication.

Proposition 13.5. Let V , W , and U be finite dimensional vector spaces with (ordered) bases \mathcal{B} , \mathcal{C} , and \mathcal{D} respectively. Let $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$. Then

$$[S \circ T]_{\mathcal{B}}^{\mathcal{D}} = [S]_{\mathcal{C}}^{\mathcal{D}} [T]_{\mathcal{B}}^{\mathcal{C}}.$$

Proof. Recall that the (i, j) th entry of a matrix product AB is given

$$(AB)_{ij} = \sum_k A_{ik} B_{kj}.$$

Enumerate the elements of the bases as follows.

$$\begin{aligned}\mathcal{B} &= (v_1, \dots, v_n) \\ \mathcal{C} &= (w_1, \dots, w_m) \\ \mathcal{D} &= (u_1, \dots, u_\ell)\end{aligned}$$

Let $a_{ij} := ([S]_{\mathcal{C}}^{\mathcal{D}})_{ij}$ and let $b_{ij} := ([T]_{\mathcal{B}}^{\mathcal{C}})_{ij}$. Note that

$$\begin{aligned}(S \circ T)(v_j) &= S(T(v_j)) = S\left(\sum_{k=1}^m b_{kj} w_k\right) = \sum_{k=1}^m b_{kj} S(w_k) \\ &= \sum_{k=1}^m b_{kj} \sum_{i=1}^{\ell} a_{ik} u_i = \sum_{k=1}^m \sum_{i=1}^{\ell} b_{kj} a_{ik} u_i.\end{aligned}$$

By commutativity and associativity of addition, we can reverse the order of the summations to get

$$(S \circ T)(v_j) = \sum_{i=1}^{\ell} \left(\sum_{k=1}^m a_{ik} b_{kj} \right) u_j.$$

Thus, the (i, j) -th entry of $[S \circ T]_{\mathcal{B}}^{\mathcal{D}}$ is $\sum_{k=1}^m a_{ik} b_{kj}$. By definition of matrix multiplication, this is precisely $([S]_{\mathcal{C}}^{\mathcal{D}} [T]_{\mathcal{B}}^{\mathcal{C}})_{ij}$. Thus

$$[S \circ T]_{\mathcal{B}}^{\mathcal{D}} = [S]_{\mathcal{C}}^{\mathcal{D}} [T]_{\mathcal{B}}^{\mathcal{C}}.$$

□

Example 13.6. Let $T : P_3(\mathbb{R}) \rightarrow P_2(\mathbb{R})$ and $S : P_2(\mathbb{R}) \rightarrow P_3(\mathbb{R})$ be defined as

$$\begin{aligned}T(p)(x) &= p'(x) \\ S(p)(x) &= \int_0^x p(t) dt\end{aligned}$$

respectively. Let $\mathcal{B} = (1, x, x^2, x^3)$ and $\mathcal{C} = (1, x, x^2)$. Recall from Example 12.5 that

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

We can compute $[S]_{\mathcal{C}}^{\mathcal{B}}$ as follows.

$$\begin{aligned} S(1) &= x, \quad [S(1)]_{\mathcal{B}} = (0, 1, 0, 0) \\ S(x) &= \frac{1}{2}x^2, \quad [S(x)]_{\mathcal{B}} = (0, 0, 1/2, 0) \\ S(x^2) &= \frac{1}{3}x^3, \quad [S(x^2)]_{\mathcal{B}} = (0, 0, 0, 1/3) \end{aligned}$$

Thus

$$[S]_{\mathcal{C}}^{\mathcal{B}} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}.$$

Recall that the Fundamental Theorem of Calculus states

$$\frac{d}{dx} \int_0^x p(t) dt = p(x).$$

Phrased differently, $(T \circ S)(p)(x) = p(x)$ so $T \circ S = I_{P_2(\mathbb{R})}$. We can verify that

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

Since $[I_{P_2(\mathbb{R})}]_{\mathcal{C}}^{\mathcal{C}} = I_3$, we have $[T \circ S]_{\mathcal{C}}^{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}}[S]_{\mathcal{C}}^{\mathcal{B}}$.

Finally, we state a result which in some sense captures the utility of passing to a coordinate matrix. If you want to compute the coordinates of $T(v)$, then multiply the coordinate matrix of T by the coordinate vector of v .

Proposition 13.7. Let V and W be finite dimensional vector spaces with (ordered) bases \mathcal{B} and \mathcal{C} respectively. Let $T : V \rightarrow W$ be linear, and let $v \in V$. Then

$$[T(v)]_{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}}[v]_{\mathcal{B}}.$$

Proof. Let $\mathcal{B} = (v_1, \dots, v_n)$, $\mathcal{C} = (w_1, \dots, w_m)$. Write $v = a_1v_1 + \dots + a_nv_n$ and $b_{ij} := ([T]_{\mathcal{B}}^{\mathcal{C}})_{ij}$. Then

$$\begin{aligned} T(v) &= T\left(\sum_{j=1}^n a_j v_j\right) \\ &= \sum_{j=1}^n a_j T(v_j) \\ &= \sum_{j=1}^n a_j \sum_{i=1}^m b_{ij} w_i \\ &= \sum_{i=1}^m \left(\sum_{j=1}^n b_{ij} a_j\right) w_i. \end{aligned}$$

The i th entry of the coordinate vector $[T(v)]_{\mathcal{C}}$ is $\sum_{j=1}^n b_{ij}a_j$, which is precisely the i th entry of the matrix vector product $[T]_{\mathcal{B}}^{\mathcal{C}}[v]_{\mathcal{B}}$. Thus

$$[T(v)]_{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}}[v]_{\mathcal{B}}.$$

□

Example 13.8. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be rotation by $\frac{\pi}{2}$ counterclockwise. Let $\mathcal{B} = \{e_1, e_2\}$ be the standard basis for \mathbb{R}^2 and $\mathcal{C} = \{(1, 1), (1, -1)\}$. To determine the matrix $[T]_{\mathcal{B}}^{\mathcal{C}}$, we need to know how T maps the elements of the input basis \mathcal{B} . Then write the output in terms of the basis \mathcal{C} .

$$T(1, 0) = (0, 1) = \frac{1}{2}(1, 1) - \frac{1}{2}(1, -1)$$

$$T(0, 1) = (-1, 0) = -\frac{1}{2}(1, 1) - \frac{1}{2}(1, -1)$$

The first column of $[T]_{\mathcal{B}}^{\mathcal{C}}$ is $[T(1, 0)]_{\mathcal{C}}$ and the second column of $[T]_{\mathcal{B}}^{\mathcal{C}}$ is $[T(0, 1)]_{\mathcal{C}}$. We have

$$[T]_{\mathcal{B}}^{\mathcal{C}} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & -1/2 \end{pmatrix}$$

In order to determine how T maps a vector $v \in \mathbb{R}^2$ in \mathcal{C} coordinates, we can multiply the column vector $[v]_{\mathcal{B}}$ by the matrix $[T]_{\mathcal{B}}^{\mathcal{C}}$. Let $v = 2e_1 - e_2$. We obtain

$$[T(v)]_{\mathcal{C}} = [T]_{\mathcal{B}}^{\mathcal{C}}[v]_{\mathcal{B}} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & -1/2 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3/2 \\ -1/2 \end{pmatrix}.$$

Once we compute the matrix, applying the transformation is a quick computation.

End of lecture 16

14. INVERTIBILITY AND ISOMORPHISMS

Definition 14.1. Let $T : V \rightarrow W$. We say that T is **invertible** if there is another function $S : W \rightarrow V$ such that $S \circ T = I_V$ and $T \circ S = I_W$.

Example 14.2. Let $f : [0, \infty) \rightarrow [0, \infty)$ be given by $f(x) = x^2$. Let $g : [0, \infty) \rightarrow [0, \infty)$ be defined by $g(x) = \sqrt{x}$. Then

$$\begin{aligned} (f \circ g)(x) &= f(g(x)) \\ &= (\sqrt{x})^2 \\ &= x. \end{aligned}$$

Since $(f \circ g)(x) = x$ for all $x \in [0, \infty)$, $f \circ g = I_{[0, \infty)}$. Similarly, one can check that $(g \circ f)(x) = x$ for all $x \in [0, \infty)$. Thus f is invertible.

Proposition 14.3. Let $T : V \rightarrow W$ be linear. Suppose that T is invertible. Then the map $S : W \rightarrow V$ satisfying $S \circ T = I_V$ and $T \circ S = I_W$ is unique and linear.

Proof. Suppose that $S' : W \rightarrow V$ also satisfies $S' \circ T = I_V$ and $T \circ S' = I_W$. Then

$$\begin{aligned} S' &= S' \circ I_W \\ &= S' \circ (T \circ S) \\ &= (S' \circ T) \circ S \\ &= I_V \circ S \\ &= S. \end{aligned}$$

Thus $S' = S$ so S is unique.

Next, we prove that S is linear. Let $w, w' \in W$ and $\lambda \in \mathcal{F}$. Then

$$\begin{aligned}
 S(w + w') &= S(I_W(w) + I_W(w')) \\
 &= S((T \circ S)(w) + (T \circ S)(w')) \\
 &= S(T(S(w)) + T(S(w'))) \\
 &= S(T(S(w) + S(w'))) \\
 &= (S \circ T)(S(w) + S(w')) \\
 &= I_V(S(w) + S(w')) \\
 &= S(w) + S(w') \\
 S(\lambda w) &= S(\lambda I_W(w)) \\
 &= S(\lambda(T \circ S)(w)) \\
 &= (S \circ T)((\lambda S)(w)) \\
 &= (S \circ T)(\lambda S(w)) \\
 &= I_V(\lambda S(w)) \\
 &= \lambda S(w).
 \end{aligned}$$

Thus S is linear. □

Definition 14.4. Let $T : V \rightarrow W$ be an invertible linear map. The **inverse** of T , denoted T^{-1} , is the unique linear map $T^{-1} : W \rightarrow V$ such that $T^{-1} \circ T = I_V$ and $T \circ T^{-1} = I_W$.

Proposition 14.5. Let $T : V \rightarrow W$ be linear. Then T is invertible if and only if T is both injective and surjective.

Proof. (\Rightarrow) First, suppose that T is invertible. Let $T(v) = 0$ for some $v \in V$. Apply T^{-1} to both sides of the equation to obtain $T^{-1}(T(v)) = T^{-1}(0)$. Since T^{-1} is linear $T^{-1}(0) = 0$ and $v = 0$. Thus $\ker(T) = \{0\}$, and T is injective. Next, let $w \in W$. Since $T(T^{-1}(w)) = w$, T is surjective.

(\Leftarrow) Assume that T is both injective and surjective. For each $w \in W$, there is a unique $v \in V$ such that $T(v) = w$. Define the function $T^{-1} : W \rightarrow V$ as $T^{-1}(w) = v$. Then $(T^{-1} \circ T)(v) = v$ and $(T \circ T^{-1})(w) = w$ so T is invertible. □

Corollary 14.5.1. Suppose that V and W are finite dimensional vector spaces. If $T : V \rightarrow W$ is an invertible linear transformation, then $\dim(V) = \dim(W)$.

Proof. Since T is invertible, it is injective. Thus $\ker(T) = \{0\}$ so $\dim(\ker(T)) = 0$. Since T is also surjective, $\text{im}(T) = W$ and $\dim(\text{im}(T)) = \dim(W)$. By Rank-Nullity Theorem, it follows that $\dim(V) = \dim(W)$. □

End of extra lecture on February 20

Definition 14.6. Let V and W be vector spaces. We say that V and W are **isomorphic** if there is an invertible linear map $T : V \rightarrow W$. In this case, we say that T is an **isomorphism** from V to W . We also sometimes write $V \cong W$.

Proposition 14.7. Let V and W be finite dimensional vector spaces with (ordered) bases \mathcal{B}, \mathcal{C} respectively. A linear map $T : V \rightarrow W$ is invertible if and only if the coordinate matrix $[T]_{\mathcal{B}}^{\mathcal{C}}$ is invertible. Moreover,

$$[T^{-1}]_{\mathcal{C}}^{\mathcal{B}} = ([T]_{\mathcal{B}}^{\mathcal{C}})^{-1}.$$

Proof. (\Rightarrow) Suppose that T is invertible. Then, by definition, $T^{-1} \circ T = I_V$ and $T \circ T^{-1} = I_W$. Taking coordinate matrices gives

$$\begin{aligned} [T^{-1} \circ T]_{\mathcal{B}}^{\mathcal{B}} &= [I_V]_{\mathcal{B}}^{\mathcal{B}} \\ [T \circ T^{-1}]_{\mathcal{C}}^{\mathcal{C}} &= [I_W]_{\mathcal{C}}^{\mathcal{C}}. \end{aligned}$$

By Proposition 13.5 and $[I_V]_{\mathcal{B}}^{\mathcal{B}} = [I_W]_{\mathcal{C}}^{\mathcal{C}} = I_n$, where $n = \dim(V) = \dim(W)$, we have

$$\begin{aligned} [T^{-1}]_{\mathcal{C}}^{\mathcal{B}} [T]_{\mathcal{B}}^{\mathcal{C}} &= I_n \\ [T]_{\mathcal{B}}^{\mathcal{C}} [T^{-1}]_{\mathcal{C}}^{\mathcal{B}} &= I_n. \end{aligned}$$

Thus $[T]_{\mathcal{B}}^{\mathcal{C}}$ is invertible and $([T]_{\mathcal{B}}^{\mathcal{C}})^{-1} = [T^{-1}]_{\mathcal{C}}^{\mathcal{B}}$.

(\Leftarrow) Suppose that $[T]_{\mathcal{B}}^{\mathcal{C}}$ is invertible. Let $b_{ij} := ([T]_{\mathcal{B}}^{\mathcal{C}})^{-1}_{ij}$. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_n\}$. Define $S : W \rightarrow V$ on \mathcal{C} as

$$S(w_j) := b_{1j}v_1 + \dots + b_{nj}v_n$$

and extend the definition of S to W by linearity. By definition, $[S]_{\mathcal{C}}^{\mathcal{B}} = ([T]_{\mathcal{B}}^{\mathcal{C}})^{-1}$. Then

$$[S \circ T]_{\mathcal{B}}^{\mathcal{B}} = [S]_{\mathcal{C}}^{\mathcal{B}} [T]_{\mathcal{B}}^{\mathcal{C}} = ([T]_{\mathcal{B}}^{\mathcal{C}})^{-1} [T]_{\mathcal{B}}^{\mathcal{C}} = I_n$$

so $(S \circ T)(v_j) = v_j$ for all $1 \leq j \leq n$. By linearity, it follows that $S \circ T = I_V$. By a similar argument, one can show that $T \circ S = I_W$. We conclude that T is invertible. \square

Theorem 14.8. Let V, W be finite dimensional vector spaces. Then V and W are isomorphic if and only if $\dim(V) = \dim(W)$.

Proof. (\Rightarrow) Suppose that $V \cong W$. By definition, there exists an invertible linear map $T : V \rightarrow W$. By Proposition 14.5.1, $\dim(V) = \dim(W)$.

(\Leftarrow) Suppose that $\dim(V) = \dim(W)$. Let

$$\begin{aligned} \mathcal{B} &:= \{v_1, \dots, v_n\} \\ \mathcal{C} &:= \{w_1, \dots, w_n\} \end{aligned}$$

be bases for V and W respectively. Define $T : V \rightarrow W$ as follows. For $a_1v_1 + \dots + a_nv_n \in V$, define

$$T(a_1v_1 + \dots + a_nv_n) = a_1w_1 + \dots + a_nw_n.$$

In other words, we are defining T by $T(v_j) := w_j$ and “extending by linearity.” Recall that the behavior of a linear map is completely determined by its behavior on a basis by Proposition 11.14.

First, note that T is clearly linear. Next, we claim that T is injective. Suppose that

$$T(a_1v_1 + \dots + a_nv_n) = 0.$$

Then $a_1w_1 + \dots + a_nw_n = 0$. Since \mathcal{C} is a basis, it is linearly independent, so $a_1 = \dots = a_n = 0$. Thus $a_1v_1 + \dots + a_nv_n = 0$, and $\ker(T) = \{0\}$. We conclude that T is injective by Proposition 11.21. Since $\dim(V) = \dim(W)$, T is also surjective by Proposition 11.22. Thus T is invertible and is an isomorphism from V to W . \square

Corollary 14.8.1. Let V be an n -dimensional vector space over \mathcal{F} . Then $V \cong \mathcal{F}^n$.

Remark 14.9. There are many transformations $T : V \rightarrow \mathcal{F}^n$ that provide the isomorphism in Corollary 14.8.1. Let \mathcal{B} be a basis for the n -dimensional vector space V . The most useful isomorphism is often $T : V \rightarrow \mathcal{F}^n$ that sends a vector $v \in V$ to the coordinate vector $[v]_{\mathcal{B}}$.

Example 14.10. Let $T : P_3(\mathbb{R}) \rightarrow P_3(\mathbb{R})$ be the derivative operator $T(p)(x) = p'(x)$. Take the basis $\mathcal{B} = \{1, x, x^2, x^3\}$ for $P_3(\mathbb{R})$. Example 13.6 shows that

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We note that $[T]_{\mathcal{B}}^{\mathcal{B}}$ is not full rank and, hence, not an invertible matrix. Proposition 14.7 implies T is not an invertible linear transformation.

Example 14.11. Let V be an n -dimensional vector space with a basis $\mathcal{B} = (v_1, \dots, v_n)$, and let W be an m -dimensional vector space with a basis $\mathcal{C} = (w_1, \dots, w_m)$. Recall the linear map

$$\phi : \mathcal{L}(V, W) \rightarrow M_{m \times n}(\mathcal{F})$$

given by $\phi(T) = [T]_{\mathcal{B}}^{\mathcal{C}}$ from Corollary 12.8.1. We claim that ϕ is an isomorphism so

$$\mathcal{L}(V, W) \cong M_{m \times n}(\mathcal{F}).$$

In other words, the vector spaces are essentially the same.

Indeed, first suppose that $\phi(T) = 0_{m \times n}$. Let $v \in V$. By Proposition 13.7,

$$[T(v)]_{\mathcal{C}} = 0_{m \times n}[v]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathcal{F}^m.$$

Thus $T(v) = 0w_1 + \dots + 0w_m = 0 \in W$, and $T = 0 \in \mathcal{L}(V, W)$. We conclude that $\ker(\phi) = \{0\}$ so ϕ is injective. Next, we show that ϕ is surjective. Let $A \in M_{m \times n}(\mathcal{F})$. Define $T : V \rightarrow W$ on \mathcal{B} by

$$T(v_j) := a_{1j}w_1 + \dots + a_{mj}w_m.$$

Extend the definition of T to V by linearity. By construction, $\phi(T) = A$. Thus ϕ is surjective and, hence, ϕ is an isomorphism.

Since $\dim(M_{m \times n}(\mathcal{F})) = mn$, we have $\dim(\mathcal{L}(V, W)) = mn$. By Corollary 14.8.1,

$$\mathcal{L}(V, W) \cong M_{m \times n}(\mathcal{F}) \cong \mathcal{F}^{mn}.$$

In lower division linear algebra, we basically treat every linear transformation like matrix multiplication. This statement justifies the approach!

Example 14.12. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be projection onto the line $y = x$. Let $\mathcal{B} = \{e_1, e_2\}$ be the standard basis for \mathbb{R}^2 . To construct $[T]_{\mathcal{B}}^{\mathcal{B}}$, we need to project e_1 and e_2 onto the line $y = x$. Take the vector $v = e_1 + e_2$ on the line. Recall the following projection formula from Math 33A.

$$\begin{aligned} \text{proj}_v e_1 &= \left(\frac{e_1 \cdot v}{v \cdot v} \right) v = \frac{1}{2}(1, 1) \\ \text{proj}_v e_2 &= \left(\frac{e_2 \cdot v}{v \cdot v} \right) v = \frac{1}{2}(1, 1) \end{aligned}$$

The first and second columns of $[T]_{\mathcal{B}}^{\mathcal{B}}$ are $\text{proj}_v e_1$ and $\text{proj}_v e_2$ respectively so

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Subtract the first row from the second row to obtain

$$[T]_{\mathcal{B}}^{\mathcal{B}} \rightarrow \begin{pmatrix} 1/2 & 1/2 \\ 0 & 0 \end{pmatrix}.$$

As a result, $[T]_{\mathcal{B}}^{\mathcal{B}}$ is not full rank and, hence, not an invertible matrix. Proposition 14.7 implies that T is not an invertible linear transformation.

By Proposition 14.5, we could also study the kernel and image of T to determine invertibility. Assume that $T(v) = 0$ for some $v \in \mathbb{R}^2$. Thinking geometrically, v must be perpendicular to the line $y = x$. We find $\ker(T) = \text{Span}(\{e_1 - e_2\})$ and $\dim(\ker(T)) = 1$. We conclude that T is not injective. By Rank-Nullity Theorem, $\dim(\text{im}(T)) = \dim(\mathbb{R}^2) - \dim(\ker(T)) = 1$. Since $\dim(\text{im}(T)) = 1 < 2 = \dim(\mathbb{R}^2)$, T is not surjective either.

15. CHANGE OF COORDINATES

We will begin with a motivating example that resembles Math 33A.

Example 15.1. Let $V = \mathbb{R}^2$ and $T : V \rightarrow V$ reflection about the line $y = 2x$ in the xy -plane. Let $v = e_1 + 2e_2$ be a vector on the line. The reflection formula from Math 33A is $\text{ref}_v u = 2\text{proj}_v u - u$. With respect to the standard basis $\mathcal{B} = \{e_1, e_2\}$,

$$\begin{aligned} T(e_1) &= 2 \left(\frac{e_1 \cdot v}{v \cdot v} \right) v - e_1 = -\frac{3}{5}e_1 + \frac{4}{5}e_2 \\ T(e_2) &= 2 \left(\frac{e_2 \cdot v}{v \cdot v} \right) v - e_2 = \frac{4}{5}e_1 + \frac{3}{5}e_2 \\ [T]_{\mathcal{B}}^{\mathcal{B}} &= \begin{pmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{pmatrix}. \end{aligned}$$

If we were only given the basis and the matrix, it would not be immediately clear what is occurring geometrically.

Instead, pick the basis $\mathcal{C} = \{e_1 + 2e_2, -2e_1 + e_2\}$ so

$$\begin{aligned} T(e_1 + 2e_2) &= e_1 + 2e_2 \\ T(-2e_1 + e_2) &= -(-2e_1 + e_2) \\ [T]_{\mathcal{C}}^{\mathcal{C}} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}$$

We see that the first basis element is fixed and the second basis element, orthogonal to the first, is flipped. Thus T represents reflection about the line in the direction of the first basis vector. The geometry is more apparent from the basis \mathcal{C} than from the standard basis \mathcal{B} . A change of reference can sometimes uncover information about the linear transformation.

End of lecture 17

For the time being, we will focus on linear maps T to and from a single vector space V . In this case, we will call T a *linear operator* on V . We write

$$\mathcal{L}(V) := \mathcal{L}(V, V)$$

to denote the space of all linear operators on V . If \mathcal{B} is a basis for a finite dimensional space V then we will write

$$[T]_{\mathcal{B}} := [T]_{\mathcal{B}}^{\mathcal{B}}$$

for the coordinate matrix with respect to \mathcal{B} in both the domain and codomain.

Definition 15.2. Let V be a finite dimensional vector space. Let $\mathcal{B}, \mathcal{B}'$ be two different bases of V . The matrix $Q := [I_V]_{\mathcal{B}}^{\mathcal{B}'}$ is the **change of coordinate matrix from \mathcal{B} to \mathcal{B}'** .

Intuitively, the change of coordinate matrix does not alter vectors or the geometry of transformations. The change of coordinate matrix merely switches from one reference frame to another. The coordinates, or the way we describe a vector, will change, but the underlying vector remains the same.

Example 15.3. Let $V = P_1(\mathbb{R})$ with $p(x) = 1 + x \in P_1(\mathbb{R})$. Let $\mathcal{B} = (1, x)$ be the standard ordered basis while $\mathcal{B}' = (1 - x, 1 + x)$. We can write

$$\begin{aligned} p(x) &= 1 \cdot 1 + 1 \cdot x \\ p(x) &= 0 \cdot (1 - x) + 1 \cdot (1 + x). \end{aligned}$$

Thus

$$\begin{aligned} [p(x)]_{\mathcal{B}} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ [p(x)]_{\mathcal{B}'} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Let $Q = [I_V]_{\mathcal{B}}^{\mathcal{B}'}$ be the change of coordinate matrix from \mathcal{B} to \mathcal{B}' . Then $[p(x)]_{\mathcal{B}'} = Q[p(x)]_{\mathcal{B}}$. The change of coordinate matrix does not change the underlying vector, but it does change how the vector is represented.

Since Q is a matrix representation for I_V , we can construct Q by applying I_V to the basis elements of \mathcal{B} , the input basis. We find

$$\begin{aligned} I_V(1) &= 1 \\ I_V(x) &= x. \end{aligned}$$

The columns of Q are these outputs written in \mathcal{B}' coordinates. Then

$$\begin{aligned} I_V(1) &= 1 = \frac{1}{2} \cdot (1 - x) + \frac{1}{2} \cdot (1 + x) \\ I_V(x) &= x = -\frac{1}{2} \cdot (1 - x) + \frac{1}{2} \cdot (1 + x) \end{aligned}$$

implies

$$Q = ([I_V(1)]_{\mathcal{B}'} \quad [I_V(x)]_{\mathcal{B}'}) = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

We can confirm $[p(x)]_{\mathcal{B}'} = Q[p(x)]_{\mathcal{B}}$ since

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

Theorem 15.4. Let V be a finite dimensional vector space, and let $\mathcal{B}, \mathcal{B}'$ be two different bases of V . Let $Q = [I_V]_{\mathcal{B}}^{\mathcal{B}'}$ be the change of coordinate matrix from \mathcal{B} to \mathcal{B}' .

- (i) The matrix Q is invertible.
- (ii) For any $v \in V$,

$$[v]_{\mathcal{B}'} = Q[v]_{\mathcal{B}}.$$

- (iii) For any $T \in \mathcal{L}(V)$,

$$[T]_{\mathcal{B}} = Q^{-1}[T]_{\mathcal{B}'}Q.$$

Proof.

- (i) Since $I_V : V \rightarrow V$ is clearly invertible, Proposition 14.7 implies that Q is invertible.
- (ii) By Proposition 13.7,

$$[v]_{\mathcal{B}'} = [I_V(v)]_{\mathcal{B}'} = [I_V]_{\mathcal{B}}^{\mathcal{B}'}[v]_{\mathcal{B}} = Q[v]_{\mathcal{B}}.$$

- (iii) By Proposition 13.5,

$$[T]_{\mathcal{B}} = [I_V \circ T \circ I_V]_{\mathcal{B}} = [I_V]_{\mathcal{B}}^{\mathcal{B}'}[T]_{\mathcal{B}'}[I_V]_{\mathcal{B}}^{\mathcal{B}'}$$

Since $Q^{-1} = ([I_V]_{\mathcal{B}}^{\mathcal{B}'})^{-1} = [I_V^{-1}]_{\mathcal{B}'}^{\mathcal{B}} = [I_V]_{\mathcal{B}'}^{\mathcal{B}}$, it follows that

$$[T]_{\mathcal{B}} = Q^{-1}[T]_{\mathcal{B}'}Q.$$

□

Here is a possibly useful way to visualize all of this nonsense that we've talked about with coordinate representations and coordinate changes. Suppose we have a linear operator $T \in \mathcal{L}(V)$ on a finite dimensional vector space. This is represented by the following diagram.

$$V \xrightarrow{T} V$$

Next, choose a basis \mathcal{B} for V , and let $f_{\mathcal{B}} : V \rightarrow \mathcal{F}^n$ be the map that sends a vector $v \in V$ to its coordinate vector $[v]_{\mathcal{B}}$. That is, $f_{\mathcal{B}}(v) = [v]_{\mathcal{B}}$. These coordinate changes allow us to pass to a new “equivalent” row of the diagram. Corresponding to the map T is the map $\mathcal{F}^n \rightarrow \mathcal{F}^n$ given by multiplication by $[T]_{\mathcal{B}}$.

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ \downarrow f_{\mathcal{B}} & & \downarrow f_{\mathcal{B}} \\ \mathcal{F}^n & \xrightarrow{[T]_{\mathcal{B}}} & \mathcal{F}^n \end{array}$$

On the other hand, we could pick a *different* basis \mathcal{B}' and get a different, yet equivalent, row!

$$\begin{array}{ccc} \mathcal{F}^n & \xrightarrow{[T]_{\mathcal{B}'}} & \mathcal{F}^n \\ \uparrow f_{\mathcal{B}'} & & \uparrow f_{\mathcal{B}'} \\ V & \xrightarrow{T} & V \\ \downarrow f_{\mathcal{B}} & & \downarrow f_{\mathcal{B}} \\ \mathcal{F}^n & \xrightarrow{[T]_{\mathcal{B}}} & \mathcal{F}^n \end{array}$$

The change of coordinate matrix is the direct translation between the bottom and top row.

$$Q \left(\begin{array}{ccc} \mathcal{F}^n & \xrightarrow{[T]_{\mathcal{B}'}} & \mathcal{F}^n \\ \uparrow f_{\mathcal{B}'} & & \uparrow f_{\mathcal{B}'} \\ V & \xrightarrow{T} & V \\ \downarrow f_{\mathcal{B}} & & \downarrow f_{\mathcal{B}} \\ \mathcal{F}^n & \xrightarrow{[T]_{\mathcal{B}}} & \mathcal{F}^n \end{array} \right) Q$$

Definition 15.5. Two operators $T \in \mathcal{L}(V)$ and $S \in \mathcal{L}(W)$ are **similar** if there is an isomorphism $\phi : V \rightarrow W$ such that $T = \phi^{-1} \circ S \circ \phi$. Likewise, two matrices $A, B \in M_{n \times n}(\mathcal{F})$ are **similar** if there is an invertible matrix $Q \in M_{n \times n}(\mathcal{F})$ such that $A = Q^{-1}BQ$.

You should think about similar operators or similar matrices as objects that represent the “same” transformation, up to a change in coordinates or change in perspective in a different vector space.

Example 15.6. Let $P(\mathbb{R})$ be the vector space of polynomials in any degree with coefficients in \mathbb{R} . Let \mathbb{R}^∞ be the vector space of sequences with finitely many non-zero entries. That is,

$$\mathbb{R}^\infty := \{(a_0, a_1, a_2, \dots) : a_j \neq 0 \text{ for finitely many } j\}.$$

First, we claim that $P(\mathbb{R}) \cong \mathbb{R}^\infty$. Let $\phi : P(\mathbb{R}) \rightarrow \mathbb{R}^\infty$ be given by

$$\phi(a_0 + a_1x + \dots + a_nx^n) = (a_0, a_1, \dots, a_n, 0, \dots).$$

It is straightforward to check that ϕ is linear, injective, and surjective.

Next, we define two operators, one on $P(\mathbb{R})$ and one on \mathbb{R}^∞ . Let $T : P(\mathbb{R}) \rightarrow P(\mathbb{R})$ be given by $T(p)(x) := xp(x)$. Let $S : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$ be the “right-shift” operator given by

$$S(a_0, a_1, a_2, \dots) := (0, a_0, a_1, \dots).$$

We claim that T and S are similar or, in other words, fundamentally represent the “same” transformation. Indeed, note that

$$\begin{aligned} (\phi^{-1} \circ S \circ \phi)(a_0 + a_1x + \dots + a_nx^n) &= \phi^{-1}(S(\phi(a_0 + a_1x + \dots + a_nx^n))) \\ &= \phi^{-1}(S(a_0, a_1, \dots, a_n, 0, 0, \dots)) \\ &= \phi^{-1}(0, a_0, \dots, a_{n-1}, a_n, 0, \dots) \\ &= a_0x + a_1x^2 + \dots + a_nx^{n+1} \\ &= x(a_0 + a_1x + \dots + a_nx^n) \\ &= T(a_0 + a_1x + \dots + a_nx^n). \end{aligned}$$

Thus $T = \phi^{-1} \circ S \circ \phi$ so T and S are similar. The following diagram represents this situation.

$$\begin{array}{ccc} \mathbb{R}^\infty & \xrightarrow{S} & \mathbb{R}^\infty \\ \phi \uparrow & & \uparrow \phi \\ P(\mathbb{R}) & \xrightarrow{T} & P(\mathbb{R}) \end{array}$$

End of lecture 18

16. DIAGONALIZATION, EIGENVALUES, AND EIGENVECTORS

We once again look to Example 15.1 for motivation. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be reflection about the line $y = 2x$ in the xy -plane with $\mathcal{C} = \{e_1 + 2e_2, -2e_1 + e_2\}$. The matrix $[T]_{\mathcal{C}}$ is easy to understand geometrically because it is diagonal. Each basis vector is only multiplied by a scalar via T so the information of the linear operator is encoded in the finite basis and a finite list of multiplying factors. The basis vectors of \mathcal{C} in some sense provide the most “natural” reference frame for T . In the following sections, we will develop tools to find the most “natural” basis for a given linear operator.

Definition 16.1. Let V be a finite dimensional vector space, and let $T \in \mathcal{L}(V)$. The operator T is **diagonalizable** if there is an (ordered) basis \mathcal{B} of V such that $[T]_{\mathcal{B}}$ is a diagonal matrix.

Example 16.2. Suppose we have a diagonalizable operator $T \in \mathcal{L}(V)$ and a basis $\mathcal{B} = (v_1, \dots, v_n)$ such that $[T]_{\mathcal{B}}$ is diagonal. Let a_{ij} denote the (i, j) th entry of $[T]_{\mathcal{B}}$. Then, as usual, we have

$$T(v_j) = a_{1j}v_1 + \dots + a_{nj}v_n.$$

Since $[T]_{\mathcal{B}}$ is diagonal, $a_{ij} = 0$ if $i \neq j$. In the above expression, all terms are 0 except for the one corresponding to a_{jj} . Thus

$$T(v_j) = a_{jj}v_j.$$

This suggests that an operator will be diagonalizable if we can find a basis $\{v_1, \dots, v_n\}$ of V such that $T(v_j) = \lambda_j v_j$ for some scalars λ_j . The scaling factors a_{jj} would ultimately be the diagonal entries of the resulting coordinate matrix.

Definition 16.3. Let V be a vector space, and let $T \in \mathcal{L}(V)$. A non-zero vector $v \in V$ is an **eigenvector** of T with associated **eigenvalue** λ if $T(v) = \lambda v$ for some $\lambda \in \mathcal{F}$.

Theorem 16.4. Let V be finite dimensional, and let $T \in \mathcal{L}(V)$. Then T is diagonalizable if and only if there is a basis \mathcal{B} of V consisting of eigenvectors of T .

Proof. (\Rightarrow) Suppose that T is diagonalizable. Then there is a basis $\mathcal{B} = (v_1, \dots, v_n)$ such that $[T]_{\mathcal{B}}$ is diagonal. The computation in Example 16.2 shows that $T(v_j) = a_{jj}v_j$, where a_{ij} is the (i, j) th entry of $[T]_{\mathcal{B}}$. Thus \mathcal{B} is a basis consisting of eigenvectors of T .

(\Leftarrow) Suppose that $\mathcal{B} = (v_1, \dots, v_n)$ is a basis consisting of eigenvectors of T with associated eigenvalues $\lambda_1, \dots, \lambda_n \in \mathcal{F}$. Since $T(v_j) = \lambda_j v_j$, it follows that $[T(v_j)]_{\mathcal{B}}$ is the column vector with λ_j in the j th row and 0 elsewhere. Thus

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

and T is diagonalizable. □

Example 16.5. Consider the transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Let $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Note that

$$\begin{aligned} T \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ T \begin{pmatrix} 1 \\ -1 \end{pmatrix} &= \begin{pmatrix} -1 \\ 1 \end{pmatrix} = -1 \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \end{aligned}$$

Thus v_1 is an eigenvector of T with eigenvalue 3, and v_2 is an eigenvector of T with eigenvalue -1 . It is straightforward to verify that $\mathcal{B} = (v_1, v_2)$ is a basis of V . Theorem 16.4 implies that T is diagonalizable, and

$$[T]_{\mathcal{B}} = ([T(v_1)]_{\mathcal{B}} \quad [T(v_2)]_{\mathcal{B}}) = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}.$$

Example 16.6. Let $C^\infty(\mathbb{R})$ denote the vector space of infinitely differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ (that is, functions whose derivatives of all orders exist). Let $T : C^\infty(\mathbb{R}) \rightarrow C^\infty(\mathbb{R})$ be given by $T(f)(x) = f'(x)$. Fix $\lambda \in \mathbb{R}$, and let $f(x) = e^{\lambda x}$. Then

$$T(f)(x) = f'(x) = \lambda e^{\lambda x} = \lambda f(x).$$

Since $T(f) = \lambda f$, it follows that f is an eigenvector of T with eigenvalue λ . Thus every single real number is an eigenvalue of T ! However, it does not make sense to ask if T is diagonalizable. The vector space $C^\infty(\mathbb{R})$ is infinite dimensional, and, as a result, there are no matrix representations of linear operators on $C^\infty(\mathbb{R})$.

At this point, we do not have a good method for identifying eigenvalues or eigenvectors of a linear operator. If we try to solve for eigenvalues and eigenvectors simultaneously, we produce a non-linear system of equations since the eigenvector components are multiplied by the eigenvalue. The following result provides conditions for finding an eigenvalue. With an eigenvalue, we can find eigenvectors by solving a system of linear equations.

Proposition 16.7. Let V be finite dimensional, and let $T \in \mathcal{L}(V)$. Fix $\lambda \in \mathcal{F}$. The following four statements are equivalent.

- (i) The element $\lambda \in \mathcal{F}$ is an eigenvalue of T .
- (ii) The operator $T - \lambda I$ is not injective.
- (iii) The operator $T - \lambda I$ is not surjective.
- (iv) The operator $T - \lambda I$ is not invertible.

Proof. Assume (i) and prove (ii). Since λ is an eigenvalue of T , there is a non-zero vector $v \in V$ such that $T(v) = \lambda v$. Thus $(T - \lambda I)(v) = 0$ and $v \in \ker(T - \lambda I)$. Since $v \neq 0$, it follows that $T - \lambda I$ is not injective.

Assume (ii) and prove (iii). Since $T - \lambda I$ is not injective and the dimensions of the domain and codomain are equal, Proposition 11.22 implies $T - \lambda I$ is not surjective.

Assume (iii) and prove (iv). Since $T - \lambda I$ is not surjective, Proposition 14.5 implies $T - \lambda I$ is not invertible.

Assume (iv) and prove (i). Since $T - \lambda I$ is not invertible and the dimensions of the domain and codomain are equal, Propositions 14.5 and 11.22 imply $T - \lambda I$ is not injective. Thus there is a non-zero vector $v \in \ker(T - \lambda I)$ so $(T - \lambda I)(v) = 0$. Rearranging, we have $T(v) = \lambda v$ so λ is an eigenvalue of T . \square

Example 16.8. Consider the transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Fix $\lambda \in \mathbb{R}$ and consider the operator $T - \lambda I$. To identify eigenvalues of T , we want to understand when $T - \lambda I$ is not injective. Let \mathcal{B} be the standard basis of \mathbb{R}^2 so

$$[T - \lambda I]_{\mathcal{B}} = [T]_{\mathcal{B}} - \lambda [I]_{\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \lambda & 1 \\ 0 & 1 - \lambda \end{pmatrix}.$$

We can immediately see that if $\lambda = 1$, the operator $T - \lambda I$ will not be invertible. Indeed,

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

so $T - I$ is not injective, and $\lambda = 1$ is an eigenvalue of T . Suppose $\lambda \neq 1$ and $(T - \lambda I) \begin{pmatrix} x \\ y \end{pmatrix}$ is the zero vector. Then

$$\begin{aligned} \begin{pmatrix} 0 \\ 0 \end{pmatrix} &= \begin{pmatrix} x + y \\ y \end{pmatrix} - \lambda \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} (1 - \lambda)x + y \\ (1 - \lambda)y \end{pmatrix}. \end{aligned}$$

Since $1 - \lambda \neq 0$, the second component implies $y = 0$. Then $(1 - \lambda)x = 0$ so, similarly, $x = 0$. We conclude that $T - \lambda I$ is injective when $\lambda \neq 1$ so $\lambda \neq 1$ is not an eigenvalue of T .

For the only eigenvalue $\lambda = 1$, we find $\ker(T - I) = \text{Span}\{(1, 0)\}$. Since $\dim(\mathbb{R}^2) = 2$, there is no way to produce a basis of \mathbb{R}^2 with only eigenvectors of T . Therefore, T should not be diagonalizable. We formalize this argument later in Theorem 16.14. For now, we can take a geometric approach to this problem. The linear transformation described by T is a horizontal shear. The only direction that is fixed by T is the x -axis. Intuitively, this should indicate that there is no way to form a basis for \mathbb{R}^2 out of eigenvectors.

End of lecture 19

Example 16.9. Let V be a vector space with $T \in \mathcal{L}(V)$. Suppose v_1 and v_2 are eigenvectors with corresponding eigenvalues λ_1 and λ_2 . If $\lambda_1 \neq \lambda_2$, we will show that $\{v_1, v_2\}$ is linearly independent. Let $a_1, a_2 \in \mathcal{F}$ such that

$$a_1 v_1 + a_2 v_2 = 0.$$

Apply T to both sides to obtain

$$\begin{aligned} T(a_1v_1 + a_2v_2) &= T(0) \\ a_1T(v_1) + a_2T(v_2) &= 0 \\ a_1\lambda_1v_1 + a_2\lambda_2v_2 &= 0. \end{aligned}$$

By multiplying the original equation by λ_2 , we have

$$a_1\lambda_2v_1 + a_2\lambda_2v_2 = 0.$$

Subtracting the two equations,

$$a_1(\lambda_2 - \lambda_1)v_1 = 0.$$

Since $v_1 \neq 0$ and $\lambda_2 \neq \lambda_1$, we conclude that $a_1 = 0$. Returning to the original equation, $v_2 \neq 0$ implies $a_2 = 0$. Therefore, $\{v_1, v_2\}$ is linearly independent.

We can extend the argument of Example 16.9 using induction.

Proposition 16.10. Let $T \in \mathcal{L}(V)$, and suppose that $v_1, \dots, v_k \in V$ are eigenvectors of T with associated eigenvalues λ_j . If all λ_j 's are distinct (that is, if $\lambda_i \neq \lambda_j$ for all i, j) then $\{v_1, \dots, v_k\}$ is linearly independent.

Proof. We will proceed via induction on the number of vectors k . As a base case, suppose $k = 1$. Proposition 9.6 implies that $\{v_1\}$ is linearly independent since $v_1 \neq 0$.

Assume the statement holds for k . We will prove the claim for $k + 1$. Suppose that v_1, \dots, v_{k+1} are eigenvectors of T with associated distinct eigenvalues $\lambda_1, \dots, \lambda_{k+1}$. Suppose further that

$$a_1v_1 + \dots + a_{k+1}v_{k+1} = 0$$

for some scalars $a_1, \dots, a_{k+1} \in \mathcal{F}$. Apply T to both sides to obtain

$$\begin{aligned} T(a_1v_1 + \dots + a_{k+1}v_{k+1}) &= T(0) \\ a_1T(v_1) + \dots + a_{k+1}T(v_{k+1}) &= 0 \\ a_1\lambda_1v_1 + \dots + a_{k+1}\lambda_{k+1}v_{k+1} &= 0. \end{aligned}$$

On the other hand, multiply both sides of the original equation by λ_{k+1} to obtain

$$a_1\lambda_{k+1}v_1 + \dots + a_{k+1}\lambda_{k+1}v_{k+1} = 0.$$

By subtracting the two equations, we have

$$a_1(\lambda_{k+1} - \lambda_1)v_1 + \dots + a_k(\lambda_{k+1} - \lambda_k)v_k = 0.$$

By the inductive hypothesis, $\{v_1, \dots, v_k\}$ is linearly independent so $a_j(\lambda_{k+1} - \lambda_j) = 0$ for each $1 \leq j \leq k$. Since $\lambda_{k+1} - \lambda_j \neq 0$ for $1 \leq j \leq k$, we find $a_j = 0$ for $1 \leq j \leq k$. Returning to the original equation,

$$a_{k+1}v_{k+1} = 0.$$

Since $v_{k+1} \neq 0$, we have $a_{k+1} = 0$. Therefore, $\{v_1, \dots, v_{k+1}\}$ is linearly independent. \square

Corollary 16.10.1. Let V be finite dimensional, and let $T \in \mathcal{L}(V)$. Then T has at most $\dim(V)$ distinct eigenvalues.

Proof. Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues. Let v_j be an eigenvector associated to λ_j . By Proposition 16.10, $\{v_1, \dots, v_k\}$ is linearly independent. Proposition 10.13(iii) proves $k \leq \dim(V)$. \square

Corollary 16.10.2. Let V be finite dimensional, and suppose that $T \in \mathcal{L}(V)$ has $\dim(V)$ distinct eigenvalues. Then T is diagonalizable.

Proof. Let $n = \dim(V)$ and $\lambda_1, \dots, \lambda_n$ be the distinct eigenvalues of T . Let v_j be an eigenvector associated to λ_j . Then $S = \{v_1, \dots, v_n\}$ is a linearly independent set in V . Since $n = \dim(V)$, S is a basis for V by Proposition 10.13(ii). Thus V has a basis consisting of eigenvectors of T so T is diagonalizable by Theorem 16.4. \square

Example 16.11. Let $T : P_1(\mathbb{R}) \rightarrow P_1(\mathbb{R})$ be $T(p)(x) = xp'(x)$. An eigenvector of T would be a polynomial for which $xp'(x)$ is a constant multiple of $p(x)$. By guessing, we find that the polynomials $p(x) = 1$ and $q(x) = x$ satisfy $T(p)(x) = 0$ and $T(q)(x) = x$. Therefore, p is an eigenvector of T with eigenvalue 0 and q is an eigenvector of T with eigenvalue 1. Proposition 16.10 implies that $\mathcal{B} = \{p, q\}$ is linearly independent. Since $\dim(P_1(\mathbb{R})) = 2$, the set \mathcal{B} is a basis for $P_1(\mathbb{R})$ and T is diagonalizable. We find

$$[T]_{\mathcal{B}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Note that $P_1(\mathbb{R}) \cong \mathbb{R}^2$. If we were instead looking at the vector space \mathbb{R}^2 , a matrix like $[T]_{\mathcal{B}}$ would represent projection onto the line spanned by the second basis vector of \mathcal{B} along the span of the first basis vector of \mathcal{B} . The linear operator T should be similar to a projection transformation operator on \mathbb{R}^2 .

Proposition 16.12. Suppose $T, S \in \mathcal{L}(V)$ are similar. Then T and S have the same eigenvalues.

Proof. Let $\phi \in \mathcal{L}(V)$ be an isomorphism such that $T = \phi^{-1} \circ S \circ \phi$. Let $\lambda \in \mathcal{F}$ be an eigenvalue of S . Then $S - \lambda I$ is not injective. We claim that $T - \lambda I$ is not injective as well. Note that

$$\begin{aligned} T - \lambda I &= (\phi^{-1} \circ S \circ \phi) - \lambda I \\ &= (\phi^{-1} \circ S \circ \phi) - \lambda(\phi^{-1} \circ I \circ \phi) \\ &= \phi^{-1} \circ (S - \lambda I) \circ \phi. \end{aligned}$$

Let $v \in \ker(S - \lambda I)$ be non-zero. Define $w := \phi^{-1}(v) \in V$. Since ϕ^{-1} is injective, $w \neq 0$. We have

$$\begin{aligned} (T - \lambda I)(w) &= (\phi^{-1} \circ (S - \lambda I) \circ \phi)(w) \\ &= \phi^{-1}((S - \lambda I)(v)) \\ &= \phi^{-1}(0) \\ &= 0. \end{aligned}$$

Since $w \neq 0$, the operator $T - \lambda I$ is not injective, and λ is an eigenvalue of T .

By a symmetric argument, if λ is an eigenvalue of T , then λ is also an eigenvalue of S . Therefore, T and S have the same eigenvalues. \square

The following definition and result provide a useful equivalent condition for diagonalizability.

Definition 16.13. Let $\lambda \in \mathcal{F}$ be an eigenvalue of $T \in \mathcal{L}(V)$. The **eigenspace** of λ is the subspace of all eigenvectors associated to λ , that is,

$$E_{\lambda}(T) := \ker(T - \lambda I).$$

The **geometric multiplicity** of λ , when defined, is $\dim(E_{\lambda}(T))$.

Theorem 16.14. Let V be a finite dimensional vector space. Then $T \in \mathcal{L}(V)$ is diagonalizable if and only if

$$V = E_{\lambda_1}(T) \oplus \cdots \oplus E_{\lambda_k}(T)$$

where $\lambda_1, \dots, \lambda_k$ are the (distinct) eigenvalues of T .

Proof. (\Rightarrow) Suppose that T is diagonalizable. Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T . Theorem 16.4 guarantees a basis

$$(v_1^1, \dots, v_{j_1}^1, \dots, v_1^k, \dots, v_{j_k}^k)$$

of V consisting of eigenvectors of T . The eigenvectors $\{v_1^i, \dots, v_{j_i}^i\}$ correspond to the eigenvalue λ_i . We have to show that $V = E_{\lambda_1}(T) + \dots + E_{\lambda_k}(T)$. Fix $v \in V$. Write

$$v = (a_1^1 v_1^1 + \dots + a_{j_1}^1 v_{j_1}^1) + \dots + (a_1^k v_1^k + \dots + a_{j_k}^k v_{j_k}^k).$$

Thus $v \in E_{\lambda_1}(T) + \dots + E_{\lambda_k}(T)$.

We want to show that $E_{\lambda_i}(T) \cap (E_{\lambda_1}(T) + \dots + E_{\lambda_{i-1}}(T) + E_{\lambda_{i+1}}(T) + \dots + E_{\lambda_k}(T)) = \{0\}$. Let $v \in E_{\lambda_i}(T) \cap (E_{\lambda_1}(T) + \dots + E_{\lambda_{i-1}}(T) + E_{\lambda_{i+1}}(T) + \dots + E_{\lambda_k}(T))$. Then

$$\begin{aligned} v &= v_i \\ v &= v_1 + \dots + v_{i-1} + v_{i+1} + \dots + v_k \end{aligned}$$

for $v_j \in E_{\lambda_j}(T)$. We have

$$\begin{aligned} v_i &= v_1 + \dots + v_{i-1} + v_{i+1} + \dots + v_k \\ 0 &= v_1 + \dots + v_{i-1} - v_i + v_{i+1} + \dots + v_k. \end{aligned}$$

Thus $\{v_1, \dots, v_k\}$ is not linearly independent. The contrapositive to Proposition 16.10 implies that $v_j = 0$ for some $1 \leq j \leq k$. Repeated use of this argument proves that $v = v_i = 0$.

(\Leftarrow) Suppose that $V = E_{\lambda_1}(T) \oplus \dots \oplus E_{\lambda_k}(T)$. Let $\mathcal{B}_i = (v_1^i, \dots, v_{j_i}^i)$ be a basis for the subspace $E_{\lambda_i}(T)$ for each $1 \leq i \leq k$. Note that each element of \mathcal{B}_i is an eigenvector of T with eigenvalue λ_i . We claim that $\mathcal{B} := \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$ is a basis of V . Fix $v \in V$. Since $V = E_{\lambda_1}(T) + \dots + E_{\lambda_k}(T)$, we can write $v = v_1 + \dots + v_k$ where each $v_i \in E_{\lambda_i}$. Further, $v_i = a_1^i v_1^i + \dots + a_{j_i}^i v_{j_i}^i$ so

$$v = (a_1^1 v_1^1 + \dots + a_{j_1}^1 v_{j_1}^1) + \dots + (a_1^k v_1^k + \dots + a_{j_k}^k v_{j_k}^k).$$

Thus $\text{Span}(\mathcal{B}) = V$. By assumption, $\{v_1^i, \dots, v_{j_i}^i\}$ is linearly independent.

We want to show that \mathcal{B} is linearly independent. Let $a_m^\ell \in \mathcal{F}$ be such that

$$(a_1^1 v_1^1 + \dots + a_{j_1}^1 v_{j_1}^1) + \dots + (a_1^k v_1^k + \dots + a_{j_k}^k v_{j_k}^k) = 0.$$

We can rearrange the equation to obtain

$$(a_1^1 v_1^1 + \dots + a_{j_1}^1 v_{j_1}^1) + \dots + (a_1^{k-1} v_1^{k-1} + \dots + a_{j_{k-1}}^{k-1} v_{j_{k-1}}^{k-1}) = -(a_1^k v_1^k + \dots + a_{j_k}^k v_{j_k}^k).$$

The vector on the left is an element of $E_{\lambda_1}(T) + \dots + E_{\lambda_{k-1}}(T)$ while the vector on the right is an element of $E_{\lambda_k}(T)$. Since the vectors are equal, it is an element of $E_{\lambda_k}(T) \cap E_{\lambda_1}(T) + \dots + E_{\lambda_{k-1}}(T)$. By the definition of the direct sum, the vector must be the zero vector. In other words,

$$a_1^k v_1^k + \dots + a_{j_k}^k v_{j_k}^k = 0.$$

Since \mathcal{B}_i is linearly independent, $a_m^k = 0$ for all $1 \leq m \leq j_k$. Repeat the argument for each $1 \leq \ell \leq k$ to prove that $a_m^\ell = 0$ for all $1 \leq \ell \leq k$ and $1 \leq m \leq j_\ell$. Therefore, \mathcal{B} is linearly independent and, hence, a basis for V . We conclude that T is diagonalizable by Theorem 16.4. \square

Corollary 16.14.1. Let V be a finite dimensional vector space. Then $T \in \mathcal{L}(V)$ is diagonalizable if and only if the sum of the geometric multiplicities of all eigenvalues of T is $\dim(V)$. That is,

$$\sum_{j=1}^k \dim(E_{\lambda_j}(T)) = \dim(V).$$

Proof. Apply Proposition 10.17.1 to the result of Theorem 16.14. \square

End of lecture 20

We introduce another example to demonstrate where diagonalizability can fail.

Example 16.15. Consider the map $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $T(x, y) = (-y, x)$. Geometrically, this map rotates every vector counterclockwise by 90 degree. From this description, it should be clear that there are no eigenvalues, but we will prove this carefully. Fix $\lambda \in \mathbb{R}$ and consider the operator $T - \lambda I$. Note that

$$(T - \lambda I)(x, y) = T(x, y) - \lambda(x, y) = (-y - \lambda x, x - \lambda y).$$

We claim that $T - \lambda I$ is injective. Suppose $(T - \lambda I)(x, y) = (0, 0)$. We obtain the following system of equations.

$$\begin{cases} y + \lambda x = 0 \\ x - \lambda y = 0 \end{cases}$$

The second equation implies $x = \lambda y$. Substitute into the first equation to obtain $y + \lambda(\lambda y) = 0$. Factor out y so $y(1 + \lambda^2) = 0$. Since $1 + \lambda^2 > 0$, we have $y = 0$. Then $x = \lambda y$ so $x = 0$. We conclude that $\ker(T - \lambda I) = \{0\}$, and $T - \lambda I$ is injective. Since $T - \lambda I$ is injective for all $\lambda \in \mathbb{R}$, there are no eigenvalues of T over \mathbb{R} .

We should contrast Example 16.15 to that of Example 16.8. In Example 16.8, the horizontal shear operator is not diagonalizable over any field. There is at most one eigenvalue for the operator, and the eigenvalue produces no more than one corresponding eigenvector. However, the rotation operator in Example 16.15, while not diagonalizable over \mathbb{R} , is diagonalizable over \mathbb{C} . The following results indicate that the choice of field can matter.

Theorem 16.16. Let V be a finite dimensional vector space over \mathbb{C} . Let $T \in \mathcal{L}(V)$. Then T has an eigenvalue.

The proof is a classic argument that uses the Fundamental Theorem of Algebra below.

Theorem 16.17 (Fundamental Theorem of Algebra). Any polynomial $p \in P(\mathbb{C})$ can be completely factored into linear parts.

Proof of Theorem 16.16. Let $n = \dim(V)$. Fix a non-zero vector $v \in V$. Consider the following list of vectors

$$v, T(v), T^2(v), \dots, T^n(v)$$

where by T^k we mean T composed with itself k times. Since the list has $n + 1$ vectors, either there will be repeats among the vectors or there will be a linear dependence among the vectors by Proposition 10.13(iii). In either case, there exist scalars $a_0, \dots, a_n \in \mathbb{C}$, not all 0, such that

$$\begin{aligned} a_0 v + a_1 T(v) + a_2 T^2(v) + \dots + a_n T^n(v) &= 0 \\ (a_0 I + a_1 T + a_2 T^2 + \dots + T^n)(v) &= 0. \end{aligned}$$

By the Fundamental Theorem of Algebra, the left-hand side can be factored as

$$c[(T - \lambda_1 I) \circ (T - \lambda_2 I) \circ \dots \circ (T - \lambda_n I)](v) = 0$$

where $\lambda_1, \dots, \lambda_n$ are the (possibly repeated) complex roots of the polynomial $a_0 + a_1 z + \dots + a_n z^n$. Since $v \neq 0$, the operator $(T - \lambda_1 I) \circ (T - \lambda_2 I) \circ \dots \circ (T - \lambda_n I)$ is not injective and, hence, not invertible. Thus at least one of the factors, say $(T - \lambda_j I)$, is not invertible so λ_j is an eigenvalue of T . \square

Even over \mathbb{C} , there will be linear operators that are not diagonalizable as in Example 16.8. However, working over \mathbb{C} , as opposed to \mathbb{R} , provides a better framework for checking diagonalizability since every linear operator has at least one eigenvalue.

Example 16.18. We will revisit Example 16.15 over \mathbb{C} . Consider the map $T : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ defined by $T(x, y) = (-y, x)$. We find the eigenvalue eigenvector pairings below.

$$T(i, 1) = (-1, i) = i(i, 1)$$

$$T(-i, 1) = (-1, -i) = -i(-i, 1)$$

We can show $((i, 1), (-i, 1))$ is a basis for \mathbb{C}^2 . Therefore, T is diagonalizable by Theorem 16.4.

17. DETERMINANT AND THE CHARACTERISTIC POLYNOMIAL

Thus far, we have powerful results about when a linear operator is diagonalizable. However, we do not have a simple computational method for finding eigenvalues and eigenvectors. In order to develop such a method, we will first review some facts about the determinant. We will use the determinant to define the characteristic polynomial, an algebraic tool for finding eigenvalues.

Proposition 17.1. There exists a map $\det : M_{n \times n}(\mathcal{F}) \rightarrow \mathcal{F}$ called the **determinant** that satisfies the following properties.

- (i) $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$
- (ii) $\det(AB) = (\det A)(\det B)$
- (iii) $\det(A^{-1}) = (\det A)^{-1}$
- (iv) A is invertible if and only if $\det A \neq 0$

Example 17.2. We might need to take determinants of larger square matrices. For this, we can proceed by cofactor expansion. Let A be an $n \times n$ matrix. Define the minor \widehat{A}_{ij} to be the $(n-1) \times (n-1)$ matrix where the i th row and j th column of A have been removed. The determinant of A can be computed in terms of the determinants of these minors as

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} A_{ij} \det(\widehat{A}_{ij}).$$

We call the process expanding along the i th row since each minor was chosen by removing the i th row. We build the computation for $\det(A)$ out of the determinants of smaller matrices. Thus the definition of the determinant of a 2×2 matrix is sufficient for computing the determinant of any $n \times n$ matrix. However, this process becomes extremely tedious for large choices of n . We will predominantly work with 2×2 and 3×3 matrices in this course.

To practice, we will compute the determinant of

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

Expanding along the first row,

$$\begin{aligned} \det(A) &= (-1)^{1+1} \cdot 0 \cdot \det \begin{pmatrix} -1 & 3 \\ 3 & 2 \end{pmatrix} + (-1)^{1+2} \cdot 1 \cdot \det \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} + (-1)^{1+3} \cdot 2 \cdot \det \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix} \\ &= -(2 \cdot 2 - 3 \cdot 1) + 2(2 \cdot 3 - (-1) \cdot 1) \\ &= -1 + 14 \\ &= 13. \end{aligned}$$

Since $\det(A) \neq 0$, we conclude that A is an invertible matrix.

End of lecture 21

Definition 17.3. Let V be an n -dimensional vector space. Let $T \in \mathcal{L}(V)$ and \mathcal{B} be a basis of V . The **characteristic polynomial** of T is

$$c_T(t) := \det([T]_{\mathcal{B}} - tI_n)$$

Proposition 17.4. Let V be a finite dimensional vector space, and let $T \in \mathcal{L}(V)$.

- (i) The characteristic polynomial does not depend on the choice of basis in Definition 17.3.
- (ii) The eigenvalues of T are precisely the zeroes of $c_T(t)$, viewed as a polynomial over \mathcal{F} .

Proof.

- (i) Let \mathcal{B}' be another basis of V with Q the change of coordinate matrix from \mathcal{B} to \mathcal{B}' . Then

$$\begin{aligned} \det([T]_{\mathcal{B}} - tI_n) &= \det(Q^{-1}[T]_{\mathcal{B}'}Q - tQ^{-1}I_nQ) \\ &= \det(Q^{-1}([T]_{\mathcal{B}'} - tI_n)Q) \\ &= \det(Q^{-1})\det([T]_{\mathcal{B}'} - tI_n)(\det Q) \\ &= (\det Q)^{-1}(\det Q)\det([T]_{\mathcal{B}'} - tI_n) \\ &= \det([T]_{\mathcal{B}'} - tI_n). \end{aligned}$$

- (ii) Note that $T - tI$ is invertible if and only if $[T - tI]_{\mathcal{B}} = [T]_{\mathcal{B}} - tI_n$ is invertible. Equivalently, the matrix is invertible if and only if

$$c_T(t) = \det([T]_{\mathcal{B}} - tI_n) \neq 0.$$

Thus the eigenvalues of T are precisely the scalars $\lambda \in \mathcal{F}$ such that $c_T(\lambda) = 0$. □

For the remainder of this section, we will assume that $\mathcal{F} = \mathbb{C}$ so that every characteristic polynomial factors completely into linear terms.

Definition 17.5. Let V be a finite dimensional vector space with $T \in \mathcal{L}(V)$. Let λ be an eigenvalue of T . The **algebraic multiplicity** of λ is the largest possible k for which $(t - \lambda)^k$ is a factor of the characteristic polynomial $c_T(t)$. In other words, the algebraic multiplicity of λ is how many times it appears as a root of the characteristic polynomial.

Proposition 17.6. Let V be a finite dimensional vector space with $T \in \mathcal{L}(V)$. Let λ be an eigenvalue of T with algebraic multiplicity m . Then $1 \leq \dim(E_{\lambda}(T)) \leq m$.

Proof. Choose an ordered basis (v_1, \dots, v_p) for $E_{\lambda}(T)$ and extend to an ordered basis

$$\mathcal{B} = (v_1, \dots, v_p, v_{p+1}, \dots, v_n)$$

for V by Proposition 10.13. Therefore,

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda I_p & B \\ 0 & C \end{pmatrix}$$

where B is a $p \times (n - p)$ matrix, 0 is an $(n - p) \times p$ matrix, and C is an $(n - p) \times (n - p)$ matrix. We can compute the characteristic polynomial

$$\begin{aligned} c_T(t) &= \det([T]_{\mathcal{B}} - tI_n) \\ &= \det \begin{pmatrix} (\lambda - t)I_p & B \\ 0 & C - tI_{n-p} \end{pmatrix} \\ &= \det((\lambda - t)I_p) \det(C - tI_{n-p}) \\ &= (\lambda - t)^p \det(C - tI_{n-p}). \end{aligned}$$

Thus $(\lambda - t)^p$ is a factor of $c_T(t)$, and the algebraic multiplicity of λ is at least p . □

Proposition 17.7. Let V be a finite dimensional vector space with $T \in \mathcal{L}(V)$. Then T is diagonalizable if and only if the geometric multiplicity of each eigenvalue equals its algebraic multiplicity.

Proof. Let $n = \dim(V)$. Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of T with respective algebraic multiplicities m_i . Let $d_i = \dim(E_{\lambda_i}(T))$ be the corresponding geometric multiplicities.

(\Rightarrow) Assume that T is diagonalizable. Let \mathcal{B} be a basis for V made up of eigenvectors of T . For each $1 \leq i \leq k$, define $\mathcal{B}_i := \mathcal{B} \cap E_{\lambda_i}(T)$. Denote by n_i the number of vectors in \mathcal{B}_i . Then $n_i \leq d_i$ by Proposition 10.16 and $d_i \leq m_i$ by Proposition 17.6. Since \mathcal{B} has n vectors and the $E_{\lambda_i}(T)$ has trivial intersection with the sum of the other eigenspaces, $\sum_{i=1}^k n_i = n$. The sum of the algebraic multiplicities is the degree of the characteristic polynomial so $\sum_{i=1}^k m_i = n$ as well. We have

$$n = \sum_{i=1}^k n_i \leq \sum_{i=1}^k d_i \leq \sum_{i=1}^k m_i = n.$$

It follows that $\sum_{i=1}^k (m_i - d_i) = 0$ and $d_i = m_i$ for each $1 \leq i \leq k$ since $0 \leq m_i - d_i$.

(\Leftarrow) Assume that $d_i = m_i$ for each $1 \leq i \leq k$. For each $1 \leq i \leq k$, let \mathcal{B}_i be an ordered basis for $E_{\lambda_i}(T)$ and $\mathcal{B} := \mathcal{B}_1 \cup \dots \cup \mathcal{B}_k$. By a similar argument to that of Theorem 16.14, \mathcal{B} is linearly independent. Since $d_i = m_i$ for each $1 \leq i \leq k$, \mathcal{B} contains $\sum_{i=1}^k d_i = \sum_{i=1}^k m_i = n$ vectors. Therefore, \mathcal{B} is an ordered basis for V consisting of eigenvectors of T so T is diagonalizable. \square

Example 17.8. We return to Example 16.15. Let $T : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ be defined by $T(x, y) = (-y, x)$. Pick \mathcal{B} to be the standard basis. The characteristic polynomial is

$$\begin{aligned} c_T(t) &= \det([T]_{\mathcal{B}} - tI_2) \\ &= \det \begin{pmatrix} -t & -1 \\ 1 & -t \end{pmatrix} \\ &= t^2 + 1. \end{aligned}$$

Over \mathbb{C} , we factor $c_T(t) = (t - i)(t + i)$. The eigenvalues of T are $\pm i$. At this point, we have found two distinct eigenvectors in the two-dimensional vector space \mathbb{C}^2 . Corollary 16.10.2 implies that T is diagonalizable. However, we find the corresponding eigenvectors anyway.

Let $\lambda = i$. Then $[T]_{\mathcal{B}} - iI_2 = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}$. We find $E_i(T) = \ker([T]_{\mathcal{B}} - iI) = \text{Span}\{(i, 1)\}$.

Let $\lambda = -i$. Then $[T]_{\mathcal{B}} + iI_2 = \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix}$. We find $E_{-i}(T) = \ker([T]_{\mathcal{B}} + iI) = \text{Span}\{(-i, 1)\}$.

End of lecture 22

Example 17.9. Let $T : P_2(\mathbb{C}) \rightarrow P_2(\mathbb{C})$ be defined by $T(p(x)) = p'(x)$. We want to figure out if T is diagonalizable. Pick the standard basis $\mathcal{B} = (1, x, x^2)$ for $P_2(\mathbb{C})$. We have

$$\begin{aligned} T(1) &= 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x) &= 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x^2) &= 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \end{aligned}$$

so

$$[T]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Compute the characteristic polynomial by expanding the determinant along the first row as follows.

$$\begin{aligned}
 c_T(t) &= \det([T]_{\mathcal{B}} - tI_3) \\
 &= \det \begin{pmatrix} -t & 1 & 0 \\ 0 & -t & 2 \\ 0 & 0 & -t \end{pmatrix} \\
 &= (-1)^{1+1}(-t) \det \begin{pmatrix} -t & 2 \\ 0 & -t \end{pmatrix} + (-1)^{1+2} \det \begin{pmatrix} 0 & 2 \\ 0 & -t \end{pmatrix} \\
 &= -t(t^2) \\
 &= -t^3
 \end{aligned}$$

Thus the only eigenvalue of T is 0 with algebraic multiplicity 3. We want to find the geometric multiplicity of the eigenvalue 0, which is equivalent to computing the dimension of the kernel of $[T]_{\mathcal{B}} - 0I_3 = [T]_{\mathcal{B}}$. The matrix

$$[T]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

is rank 2 so $\dim(\ker(T)) = \dim(P_2(\mathbb{C})) - \text{rank}(T) = 1$ by Rank-Nullity Theorem. Since the geometric multiplicity of the eigenvalue 0 does not equal its algebraic multiplicity, the operator T is not diagonalizable.

18. INVARIANT SUBSPACES AND CAYLEY-HAMILTON THEOREM

We are skipping Section 18 in lecture. The material will not be tested.

If $v \in V$ is an eigenvector of a linear operator $T : V \rightarrow V$, then T sends v to an element of $\text{Span}(\{v\})$. As a result, the subspace $\text{Span}(\{v\})$ of V is mapped into itself via T . We can generalize this notion with the following definition.

Definition 18.1. Let V be a vector space and $T : V \rightarrow V$ a linear operator. A subspace W of V is a **T -invariant subspace** of V if $T(W) \subset W$. In other words, for each $w \in W$, $T(w) \in W$.

Example 18.2. We have already encountered many examples of T -invariant subspaces.

- (1) $\{0\}$
- (2) V
- (3) $\text{im}(T)$
- (4) $\ker(T)$
- (5) E_{λ} for any eigenvalue λ of T

Example 18.3. Let $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be the operator defined as $T(x, y, z) = (x + y, y + z, 0)$. Let W be the xy -plane. For any $(x, y, 0) \in W$, we have $T(x, y, 0) = (x + y, y, 0) \in W$. Thus W is T -invariant. Let X be the x -axis. Then for $(x, 0, 0) \in \mathbb{R}^3$, we have $T(x, 0, 0) = (x, 0, 0) \in X$ so X is T -invariant as well. In fact, the vectors on the x -axis are eigenvectors of T corresponding to the eigenvalue 1.

When we identify a T -invariant subspace W of a vector space V . We can restrict the linear operator T to a linear operator on W . We denote $T_W : W \rightarrow W$ the function $T_W(w) = T(w)$ for all $w \in W$. Since T is a linear transformation, we can show that T_W is a linear transformation. As the following result illustrates, there are some properties of T that T_W inherits.

Proposition 18.4. Let V be a finite dimensional vector space with linear operator $T : V \rightarrow V$. Let W be a T -invariant subspace of V . Then the characteristic polynomial of T_W divides the characteristic polynomial of T .

Proof. Choose a basis $\mathcal{B}' = (v_1, \dots, v_k)$ for W and extend it to a basis $\mathcal{B} = (v_1, \dots, v_k, v_{k+1}, \dots, v_n)$ for V via Proposition 10.13. Let $v_i \in \mathcal{B}'$. Since W is T -invariant, $T(v_i) = T_W(v_i) \in \text{Span}(\mathcal{B}')$ so

$$[T]_{\mathcal{B}} = \begin{pmatrix} [T_W]_{\mathcal{B}'} & A \\ 0 & B \end{pmatrix}$$

Let $f(t)$ be the characteristic polynomial of T and $g(t)$ the characteristic polynomial of T_W . Then

$$f(t) = \det([T]_{\mathcal{B}} - tI_n) = \det \begin{pmatrix} [T_W]_{\mathcal{B}'} - tI_k & A \\ 0 & B - tI_{n-k} \end{pmatrix} = g(t) \det(B - tI_{n-k}).$$

□

The following example is a common type of invariant subspace.

Example 18.5. Let V be a vector space and $T : V \rightarrow V$ a linear operator. Let $v \in V$. Then $W := \text{Span}(\{v, T(v), T^2(v), \dots\})$ is a T -invariant subspace of V . In fact, W is the smallest T invariant subspace containing v . We call W the *T -cyclic subspace of V generated by v* .

Example 18.6. Let $T : P_2(\mathbb{R}) \rightarrow P_2(\mathbb{R})$ be the linear operator defined by $T(p(x)) = p'(x)$. The T -cyclic subspace of $P_2(\mathbb{R})$ generated by x^2 is $\text{Span}(\{x^2, 2x, 2\}) = P_2(\mathbb{R})$. We conclude that there is no proper T -invariant subspace of $P_2(\mathbb{R})$ that contains x^2 .

As the next result proves, there is an easy way to produce a basis for and the characteristic polynomial corresponding to a cyclic subspace.

Proposition 18.7. Let V be a finite dimensional vector space with linear operator $T : V \rightarrow V$. Let W be a T -cyclic subspace of V generated by a non-zero vector $v \in V$. Let $k = \dim(W)$.

- (i) $\{v, T(v), T^2(v), \dots, T^{k-1}(v)\}$ is a basis for W .
- (ii) If $a_0v + a_1T(v) + \dots + a_{k-1}T^{k-1}(v) + T^k(v) = 0$, then the characteristic polynomial of T_W is $f(t) = (-1)^k(a_0 + a_1t + \dots + a_{k-1}t^{k-1} + t^k)$.

Proof.

- (i) Since $v \neq 0$, $\{v\}$ is linearly independent by Example 9.6. Let j be the largest positive integer for which $\mathcal{B} = \{v, T(v), \dots, T^{j-1}(v)\}$ is linearly independent. Such a j must exist since V is finite dimensional. Equivalently, $\mathcal{B} \cup \{T^j(v)\}$ is linearly dependent. Proposition 9.9 implies that $T^j(v)$ is a linear combination of the elements of \mathcal{B} . Let $Z = \text{Span}(\mathcal{B})$ so \mathcal{B} is a basis for Z . We will show that Z is T -invariant. Let $z \in Z$. Then $z = b_0v + \dots + b_{j-1}T^{j-1}(v)$ and, by linearity of T ,

$$T(z) = T(b_0v + \dots + b_{j-1}T^{j-1}(v)) = b_0T(v) + \dots + b_{j-1}T^j(v).$$

By above, $T^j(v)$ is a linear combination of the elements of \mathcal{B} so $T(z)$ is a linear combination of elements of \mathcal{B} . Thus $T(z) \in Z$, and Z is T -invariant. Since W is the smallest T -invariant subspace that contains v , $W \subset Z$. Clearly, $Z \subset W$ so $W = Z$. It follows that \mathcal{B} is a basis for W and $j = \dim(W) = k$.

- (ii) View \mathcal{B} from (i) as an ordered basis for W . Let a_0, \dots, a_{k-1} be scalars such that

$$a_0v + a_1T(v) + \dots + a_{k-1}T^{k-1}(v) + T^k(v) = 0.$$

Rearrange to obtain

$$T^k(v) = -a_0 - a_1T(v) - \dots - a_{k-1}T^{k-1}(v).$$

Then

$$[T_W]_{\mathcal{B}} = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -a_{k-1} \end{pmatrix}.$$

A computation reveals that the characteristic polynomial is

$$f(t) = (-1)^k(a_0 + a_1t + \dots + a_{k-1}t^{k-1} + t^k).$$

□

Cayley-Hamilton Theorem is an important result that links the algebra of characteristic polynomials with the geometry of linear operators. Although we won't see applications of the result in this class, we will run into Cayley-Hamilton many times in the course of our mathematical journey.

Theorem 18.8 (Cayley-Hamilton Theorem). Let V be a finite dimensional vector space with linear operator $T : V \rightarrow V$. Let $f(t)$ be the characteristic polynomial of T . Then $f(T)$ is the zero transformation. In other words, T “satisfies” its characteristic polynomial.

Proof. We will show that $f(T)(v) = 0$ for all $v \in V$. Since the claim is clear when $v = 0$, we may assume that $v \neq 0$. Let W be the T -cyclic subspace generated by v . Suppose $\dim(W) = k$. By Proposition 18.7(i), there are scalars a_0, \dots, a_{k-1} for which

$$a_0v + a_1T(v) + \dots + a_{k-1}T^{k-1}(v) + T^k(v) = 0.$$

Thus Proposition 18.7(ii) implies that

$$g(t) = (-1)^k(a_0 + a_1t + \dots + a_{k-1}t^{k-1} + t^k)$$

is the characteristic polynomial of T_W . As a result, $g(T)(v) = 0$. By Proposition 18.4, $g(t)$ divides $f(t)$ so $f(T)(v) = 0$. □

Example 18.9. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the linear operator defined by

$$T(x, y) = (x + 2y, -2x + y).$$

Choose the standard basis $\mathcal{B} = (e_1, e_2)$. In order to find a matrix representation of the transformation, apply T to the input basis elements and write the result in terms of the output basis.

$$T(e_1) = (1, -2) = e_1 - 2e_2$$

$$T(e_2) = (2, 1) = 2e_1 + e_2$$

Then

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}.$$

The characteristic polynomial of T is

$$\begin{aligned} \det([T]_{\mathcal{B}} - tI_2) &= \det \begin{pmatrix} 1-t & 2 \\ -2 & 1-t \end{pmatrix} \\ &= (1-t)(1-t) + 4 \\ &= t^2 - 2t + 5. \end{aligned}$$

Cayley-Hamilton Theorem implies that $T^2 - 2T + 5I$ is the zero transformation. We can verify the claim on the matrix representation of T as follows.

$$\begin{aligned} [T]_{\mathcal{B}}^2 - 2[T]_{\mathcal{B}} + 5I_2 &= \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} - 2 \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} + 5 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -3 & 4 \\ -4 & -3 \end{pmatrix} + \begin{pmatrix} -2 & -4 \\ 4 & -2 \end{pmatrix} + \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

19. INNER PRODUCTS

In an abstract vector space, there is no inherent notion of angle, length, or distance. We might think there is in \mathbb{R}^n , but that comes from our previous bias about Euclidean space. We want to introduce geometry into vector spaces so we define the *inner product*.

For now, we will consider vector spaces over either \mathbb{R} or \mathbb{C} . We briefly recall some notions about complex numbers. The *conjugate* of $a + bi \in \mathbb{C}$ is

$$\overline{a + bi} := a - bi \in \mathbb{C}$$

and the *modulus* of a complex number $a + bi \in \mathbb{C}$ is

$$|a + bi| := \sqrt{a^2 + b^2}.$$

Note that if $\lambda \in \mathbb{R}$ is a real number, we have $\bar{\lambda} = \lambda$. We can prove that if $z, w \in \mathbb{C}$, then

$$\begin{aligned} \overline{zw} &= \bar{z} \bar{w} \\ z\bar{z} &= |z|^2. \end{aligned}$$

Definition 19.1. Let V be a (real or complex) vector space. An **inner product** is a function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathcal{F}$ satisfying the following properties.

- (1) $\langle v + v', w \rangle = \langle v, w \rangle + \langle v', w \rangle$
- (2) $\langle \lambda v, w \rangle = \lambda \langle v, w \rangle$ for all $v, v', w \in V$ and $\lambda \in \mathcal{F}$
- (3) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for all $v, w \in V$
- (4) If $v \in V$ is non-zero, then $\langle v, v \rangle > 0$

Example 19.2. Take the vector space \mathbb{C}^n . We can define an inner product as follows. Let $v = (v_1, \dots, v_n)$ and $w = (w_1, \dots, w_n)$ for $v_i, w_j \in \mathbb{C}$. Then $\langle v, w \rangle := \sum_{i=1}^n v_i \bar{w}_i$. We refer to this inner product as the *standard inner product* on \mathbb{C}^n .

For the real vector space \mathbb{R}^n , the standard inner product is $\langle v, w \rangle := \sum_{i=1}^n v_i w_i$ since the complex conjugate does not affect real numbers. In multivariable calculus, we call this inner product the dot product.

Example 19.3. We can build a non-standard inner product on \mathbb{R}^2 as follows.

$$\begin{aligned} \left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle_2 &:= \begin{pmatrix} x_2 & y_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \\ &= \begin{pmatrix} x_2 & y_2 \end{pmatrix} \begin{pmatrix} 2x_1 + y_1 \\ x_1 + 2y_1 \end{pmatrix} \\ &= 2x_1x_2 + x_2y_1 + x_1y_2 + 2y_1y_2 \end{aligned}$$

Check that the properties of Definition 19.1 are satisfied.

Example 19.4. The vector space $P(\mathbb{R})$ can be endowed with an inner product

$$\langle p, q \rangle = \int_{-1}^1 p(t)q(t)dt.$$

Check that the properties of Definition 19.1 are satisfied.

Example 19.5. Take the vector space $\mathbb{C}([0, 1])$ of continuous \mathbb{C} -valued continuous functions on $[0, 1]$. Until we define an inner product, there is no notion of angle or distance between complex-valued continuous functions. For $f, g \in \mathbb{C}([0, 1])$, define $\langle f, g \rangle := \int_0^1 f(t)\overline{g(t)}dt$. Check that the properties of Definition 19.1 are satisfied.

Definition 19.6. An **inner product space** is a vector space with a choice of inner product.

Proposition 19.7. Let V be an inner product space. Then

- (i) $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$
- (ii) $\langle v, \lambda w \rangle = \overline{\lambda} \langle v, w \rangle$
- (iii) $\langle v, 0 \rangle = \langle 0, v \rangle = 0$
- (iv) $\langle v, v \rangle = 0$ if and only if $v = 0$.

Proof.

- (i) Definition 19.1 (1) and (3) imply

$$\begin{aligned} \langle v, w + w' \rangle &= \overline{\langle w + w', v \rangle} \\ &= \overline{\langle w, v \rangle + \langle w', v \rangle} \\ &= \overline{\langle w, v \rangle} + \overline{\langle w', v \rangle} \\ &= \langle v, w \rangle + \langle v, w' \rangle. \end{aligned}$$

- (ii) Definition 19.1 (2) and (3) imply

$$\begin{aligned} \langle v, \lambda w \rangle &= \overline{\langle \lambda w, v \rangle} \\ &= \overline{\lambda \langle w, v \rangle} \\ &= \overline{\lambda} \overline{\langle w, v \rangle} \\ &= \overline{\lambda} \langle v, w \rangle. \end{aligned}$$

- (iii) By (ii),

$$\begin{aligned} 2\langle v, 0 \rangle &= \langle v, 2 \cdot 0 \rangle \\ &= \langle v, 0 \rangle. \end{aligned}$$

Over \mathbb{R} or \mathbb{C} , $\langle v, 0 \rangle = 0$. Apply Definition 19.1(3) to obtain the result for $\langle 0, v \rangle$.

- (iv) (\Rightarrow) Assume that $\langle v, v \rangle = 0$. The contrapositive of Definition 19.1(4) proves the result.
 (\Leftarrow) Assume that $v = 0$. Then $\langle v, v \rangle = 0$ by (iii).

□

End of lecture 23

Remark 19.8. We say that an inner product is *conjugate linear* in the second component. When $\mathcal{F} = \mathbb{R}$, Definition 19.1 and Proposition 19.7 imply that an inner product is bilinear.

We can use an inner product to check equality of two vectors.

Corollary 19.8.1. If $\langle u, v \rangle = \langle u, w \rangle$ for all $u \in V$, then $v = w$.

Proof. We can rewrite the equation as $\langle u, v - w \rangle = 0$ for all $u \in V$. Pick $u = v - w$. Proposition 19.7(iv) implies that $v - w = 0$ or $v = w$. \square

Now we can introduce the first notions of geometry into an inner product space. Again, none of these notions exist a priori — they are *defined* using an inner product.

Definition 19.9. Let V be an inner product space. The **length** or **norm** or **magnitude** of a vector $v \in V$ is $\|v\| := \sqrt{\langle v, v \rangle}$. A **unit vector** is a vector such that $\|v\| = 1$.

Here we state some basic properties of the norm associated to an inner product. These should be familiar from multivariable calculus.

Proposition 19.10. Let V be an inner product space with $v \in V$ and $\lambda \in \mathbb{C}$.

- (i) $\|\lambda v\| = |\lambda| \cdot \|v\|$ where $|\lambda|$ is the magnitude of a complex number
- (ii) $v = 0$ if and only if $\|v\| = 0$

Proof.

- (i) We can write $\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v \rangle} = \sqrt{\lambda \overline{\lambda} \langle v, v \rangle} = \sqrt{|\lambda|^2 \langle v, v \rangle} = |\lambda| \cdot \|v\|$.
- (ii) The statement is equivalent to Proposition 19.7(iv). \square

Definition 19.11. Let V be an inner product space. Two vectors $v, w \in V$ are **orthogonal**, sometimes written $v \perp w$, if $\langle v, w \rangle = 0$.

Example 19.12. Recall the inner product from Example 19.3. We can compute

$$\begin{aligned} \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\rangle_2 &= \begin{pmatrix} 1 & -2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= 2 + 0 - 2 + 0 \\ &= 0. \end{aligned}$$

Thus $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ are orthogonal with respect to this non-standard inner product. With respect to the standard inner product, the dot product, these vectors are not orthogonal. The geometry of \mathbb{R}^2 changes completely when we choose a different inner product.

Example 19.13. Let $V = \mathbb{R}([0, 1])$ be the continuous \mathbb{R} -valued functions on $[0, 1]$. Recall the inner product from Example 19.5. The functions $f(t) = \sin(2\pi t)$ and $g(t) = \cos(2\pi t)$ are orthogonal with respect to the inner product since

$$\begin{aligned} \langle f, g \rangle &= \int_0^1 \sin(2\pi t) \cos(2\pi t) dt \\ &= \int_0^1 \frac{1}{2} \sin(4\pi t) dt \\ &= \left[-\frac{1}{8\pi} \cos(4\pi t) \right]_0^1 \\ &= 0. \end{aligned}$$

The next two definitions will help define an inner product on the vector space $M_{n \times n}(\mathbb{C})$.

Definition 19.14. The **trace** of a square matrix is the sum of the diagonal entries. For a matrix A in $M_{n \times n}(\mathbb{C})$, we write $\text{tr}(A) := \sum_{i=1}^n A_{ii}$.

Definition 19.15. Let $A \in M_{m \times n}(\mathbb{C})$. The **conjugate transpose** or **adjoint** of A is the $n \times m$ matrix A^* such that $(A^*)_{ij} = \overline{A_{ji}}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$.

Example 19.16. We can define an inner product on the vector space $M_{m \times n}(\mathbb{C})$ as follows. For $A, B \in M_{m \times n}(\mathbb{C})$, we have

$$\langle A, B \rangle := \text{tr}(B^* A).$$

The inner product is known as the *Frobenius inner product*.

The matrices $A = \begin{pmatrix} 1 & i \\ 0 & -1 \end{pmatrix}$ and $\begin{pmatrix} -i & 0 \\ 1 & -i \end{pmatrix}$ are orthogonal with respect to the Frobenius inner product since

$$\begin{aligned} \langle A, B \rangle &= \text{tr}(B^* A) \\ &= \text{tr} \left(\begin{pmatrix} i & 1 \\ 0 & i \end{pmatrix} \begin{pmatrix} 1 & i \\ 0 & -1 \end{pmatrix} \right) \\ &= \text{tr} \begin{pmatrix} i & -2 \\ 0 & -i \end{pmatrix} \\ &= i - i \\ &= 0. \end{aligned}$$

End of lecture 24

20. CAUCHY-SCHWARZ INEQUALITY AND ANGLES

We are skipping Section 20 in lecture. The material will not be tested.

In the inner product space \mathbb{R}^n endowed with the dot product, we have angles and famous inequalities with inner products and magnitudes. We will derive analogous statements for some general inner product spaces. The next theorem is one of the most important inequalities in all of math! There are deep interpretations of the result in physics as well.

Theorem 20.1 (Cauchy-Schwarz inequality). Let V be an inner product space. Then for all $v, w \in V$,

$$|\langle v, w \rangle| \leq \|v\| \|w\|.$$

Furthermore, equality holds if and only if v and w are parallel.

Proof. Assume that $\mathcal{F} = \mathbb{R}$. The proof in the complex case is essentially identical, with some annoying technical differences.

Fix $v, w \in V$. The result is immediate if either v or w is 0 so assume v and w are non-zero. Define a function $f : \mathbb{R} \rightarrow \mathbb{R}$ as follows. For $t \in \mathbb{R}$,

$$f(t) := \langle v + tw, v + tw \rangle.$$

Note that $f(t) = \|v + tw\|^2 \geq 0$. On the other hand, note that

$$\begin{aligned} f(t) &= \langle v + tw, v + tw \rangle \\ &= \langle v, v \rangle + \langle tw, v \rangle + \langle v, tw \rangle + \langle tw, tw \rangle \\ &= \|v\|^2 + 2\langle v, w \rangle t + \|w\|^2 t^2. \end{aligned}$$

We are viewing $f(t)$ as a quadratic polynomial expression

$$f(t) = (\|w\|^2) t^2 + (2\langle v, w \rangle) t + \|v\|^2.$$

If the discriminant of $f(t)$ is positive, there would be two distinct real roots of $f(t)$. In that case, $f(t)$ would have to be negative for some values of t , contradicting the construction $f(t) \geq 0$. Thus

$$(2\langle v, w \rangle)^2 - 4(\|w\|^2)\|v\|^2 \leq 0.$$

Rearranging the inequality gives

$$|\langle v, w \rangle| \leq \|v\|\|w\|.$$

We will now prove $\langle v, w \rangle = \|v\|\|w\|$ if and only if v and w are parallel. (\Rightarrow) Assume that $|\langle v, w \rangle| = \|v\|\|w\|$. Follow the argument above with the substitution $|\langle v, w \rangle| = \|v\|\|w\|$ to obtain $f(t) = (\|w\|t + \|v\|)^2$. If $\|w\| = 0$, then $w = 0$ by Proposition 19.10 so v and w are parallel. Assume $\|w\| \neq 0$. The repeated root of $f(t)$ is $-\frac{\|v\|}{\|w\|}$. In other words,

$$f\left(-\frac{\|v\|}{\|w\|}\right) = \left\langle v - \frac{\|v\|}{\|w\|}w, v - \frac{\|v\|}{\|w\|}w \right\rangle = 0.$$

By Proposition 19.7(iv), $u - \frac{\|v\|}{\|w\|}w = 0$ or $u = \frac{\|v\|}{\|w\|}w$.

(\Leftarrow) Assume that v and w are parallel. Then $w = cv$ for some $c \in \mathcal{F}$. We have

$$|\langle v, w \rangle| = |\langle v, cv \rangle| = |\bar{c}\langle v, v \rangle| = |c| \cdot \|v\|^2 = \|v\|\|w\|$$

since $\|w\| = \|cv\| = |c| \cdot \|v\|$. □

The triangle inequality, one of the most important results in analysis, follows easily from Cauchy-Schwarz inequality.

Corollary 20.1.1 (Triangle inequality). Let V be an inner product space with $v, w \in V$. Then

$$\|v + w\| \leq \|v\| + \|w\|.$$

Proof. Again, we prove the triangle inequality in the real case. The complex version is similar with technical differences.

Fix $v, w \in V$. Note that

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w \rangle \\ &= \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle \\ &= \|v\|^2 + 2\langle v, w \rangle + \|w\|^2. \end{aligned}$$

By the Cauchy-Schwarz inequality, $\langle v, w \rangle \leq \|v\|\|w\|$. Thus,

$$\begin{aligned} \|v + w\|^2 &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\ &= (\|v\| + \|w\|)^2. \end{aligned}$$

Taking square roots of both sides gives

$$\|v + w\| \leq \|v\| + \|w\|.$$

□

Now, we can discuss the notion of *angle* in an abstract inner product space. The notion of angle and orthogonality is *defined* via the inner product. It wasn't already there. Note that when we define angle, we have to specialize to an inner product space over the real numbers.

Definition 20.2. Let $v, w \in V$ be vectors in a real inner product space. The **angle** between v and w is

$$\theta_{v,w} := \arccos\left(\frac{\langle v, w \rangle}{\|v\|\|w\|}\right).$$

Note that the domain of $\arccos x$ is $[-1, 1]$ so $\theta_{v,w}$ is well-defined because of Cauchy-Schwarz!

Proposition 20.3. For $v, w \in V$ in an inner product space, $\langle v, w \rangle = \|v\| \|w\| \cos(\theta_{v,w})$. The formulas surrounding angles are reminiscent of those used in Math 33A.

Remark 20.4. We aren't typically interested in angles. We recommend avoiding the use of angles in proofs. It's solely interesting that we can define the notion of angle in more generality!

21. ORTHONORMAL BASES AND ORTHOGONAL COMPLEMENTS

Suppose that V is a finite dimensional inner product space. We can look for a particularly useful kind of basis called an *orthonormal basis*.

Definition 21.1. Let V be an inner product space. A set of vectors $S \subset V$ is **orthonormal** if $\|v\| = 1$ for each $v \in S$ and $\langle v, w \rangle = 0$ for $v, w \in S$ and $v \neq w$. An **orthonormal basis** is an ordered basis of V that is orthonormal.

Example 21.2. Consider \mathbb{R}^2 with the standard inner product (the dot product). Then the standard basis is an example of an orthonormal bases.

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

If we want to find the standard coordinates of a vector $e_1 + 2e_2$, we can write

$$e_1 + 2e_2 = \langle e_1 + 2e_2, e_1 \rangle e_1 + \langle e_1 + 2e_2, e_2 \rangle e_2.$$

The following is not an orthonormal basis even though v_1 and v_2 are orthogonal.

$$\left(v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)$$

The same technique to write $e_1 + 2e_2$ as a linear combination of v_1 and v_2 will not work since

$$e_1 + 2e_2 \neq \langle e_1 + 2e_2, v_1 \rangle v_1 + \langle e_1 + 2e_2, v_2 \rangle v_2 = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

Divide each of v_1 by its magnitude and divide v_2 by its magnitude to obtain the following orthonormal basis for \mathbb{R}^2 with respect to the dot product.

$$\left(\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right)$$

The original technique we used with the standard basis will once again work for the new orthonormal basis since

$$e_1 + 2e_2 = \langle e_1 + 2e_2, u_1 \rangle u_1 + \langle e_1 + 2e_2, u_2 \rangle u_2 = \frac{3}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} - \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

The following result describes a method of constructing an orthonormal basis from any basis. Further, in order to find the coefficients of a linear combination of an orthonormal basis, we need only compute inner products of elements. Before, we had to solve a system of linear equations.

Theorem 21.3. Let V be a finite dimensional inner product space.

- (i) There is an orthonormal basis of V .
- (ii) If (u_1, \dots, u_n) is an orthonormal basis of V , then for all $v \in V$,

$$v = \sum_{j=1}^n \langle v, u_j \rangle u_j.$$

Proof.

- (i) Pick a basis (v_1, \dots, v_n) for V . We build an orthonormal basis $\mathcal{B} := (u_1, \dots, u_n)$ recursively. Let $u_1 = \frac{1}{\|v_1\|} v_1$. Given $\{u_1, \dots, u_k\}$, define

$$w_{k+1} := v_{k+1} - \sum_{i=1}^k \langle v_{k+1}, u_i \rangle u_i$$

$$u_{k+1} := \frac{1}{\|w_{k+1}\|} w_{k+1}.$$

Let $v_1 = w_1$ for ease of notation. We want to prove that \mathcal{B} is orthonormal inductively. Each vector in \mathcal{B} is a unit vector via

$$\langle u_i, u_i \rangle = \left\langle \frac{1}{\|w_i\|} w_i, \frac{1}{\|w_i\|} w_i \right\rangle = \frac{1}{\|w_i\| \|w_i\|} \langle w_i, w_i \rangle = \frac{\|w_i\|^2}{\|w_i\|^2} = 1.$$

As a result, $\{u_1\}$ is orthonormal, and the base case holds. Assume that $\{u_1, \dots, u_k\}$ is orthonormal. We want to show that $\{u_1, \dots, u_{k+1}\}$ is orthonormal. By the inductive hypothesis,

$$\begin{aligned} \langle u_{k+1}, u_j \rangle &= \left\langle \frac{1}{\|w_{k+1}\|} w_{k+1}, u_j \right\rangle \\ &= \frac{1}{\|w_{k+1}\|} \langle w_{k+1}, u_j \rangle \\ &= \frac{1}{\|w_{k+1}\|} \left\langle v_{k+1} - \sum_{i=1}^k \langle v_{k+1}, u_i \rangle u_i, u_j \right\rangle \\ &= \frac{1}{\|w_{k+1}\|} \left(\langle v_{k+1}, u_j \rangle - \sum_{i=1}^k \langle v_{k+1}, u_i \rangle \langle u_i, u_j \rangle \right) \\ &= \frac{1}{\|w_{k+1}\|} (\langle v_{k+1}, u_j \rangle - \langle v_{k+1}, u_j \rangle \langle u_j, u_j \rangle) \\ &= \frac{1}{\|w_{k+1}\|} (\langle v_{k+1}, u_j \rangle - \langle v_{k+1}, u_j \rangle) \\ &= 0. \end{aligned}$$

Thus u_{k+1} is orthogonal to each u_j for $1 \leq j \leq k$, and \mathcal{B} is orthonormal.

We need to show that \mathcal{B} is a basis of V . By construction,

$$v_j = w_j + \sum_{i=1}^{j-1} \langle v_j, u_i \rangle u_i = \|w_j\| u_j + \sum_{i=1}^{j-1} \langle v_j, u_i \rangle u_i.$$

Thus $v_j \in \text{Span}(\mathcal{B})$ for each $1 \leq j \leq n$. By Corollary 8.6.1,

$$V = \text{Span}(\{v_1, \dots, v_n\}) \subset \text{Span}(\mathcal{B}) \subset V$$

so \mathcal{B} spans V . There are n vectors in \mathcal{B} so \mathcal{B} is a basis by Proposition 10.13(i).

- (ii) Let $v \in V$. Since (u_1, \dots, u_n) is a basis, we can write

$$v = \sum_{j=1}^n a_j u_j$$

for some constants $a_j \in \mathcal{F}$. Since $\{u_1, \dots, u_n\}$ is orthonormal,

$$\begin{aligned} \langle v, u_i \rangle &= \left\langle \sum_{j=1}^n a_j u_j, u_i \right\rangle \\ &= \sum_{j=1}^n \langle a_j u_j, u_i \rangle \\ &= \sum_{j=1}^n a_j \langle u_j, u_i \rangle \\ &= a_i \langle u_i, u_i \rangle \\ &= a_i. \end{aligned}$$

Therefore, $v = \sum_{i=1}^n \langle v, u_i \rangle u_i$.

□

Remark 21.4. The proof of Theorem 21.3(i) is the Gram-Schmidt orthonormalization process. Given any linearly independent set of a vector space, the Gram-Schmidt process will output an orthonormal set. We usually use Gram-Schmidt on a basis for the vector space in which case the process outputs an orthonormal basis for the vector space.

Definition 21.5. Let S be an orthonormal subset of an inner product space V . Let $v \in V$. Then the **Fourier coefficients** of v relative to S are the scalars $\langle v, u_i \rangle$ for $u_i \in S$.

Example 21.6. Let \mathbb{R}^4 be endowed with the standard inner product, the dot product. Let W be the subspace of \mathbb{R}^4 with basis $((1, 0, 1, 0), (1, 1, 1, 1), (0, 1, 2, 1))$. We will apply Gram-Schmidt process to find an orthonormal basis for W . Denote

$$\begin{aligned} v_1 &= (1, 0, 1, 0) \\ v_2 &= (1, 1, 1, 1) \\ v_3 &= (0, 1, 2, 1). \end{aligned}$$

Then $\|v_1\|^2 = \langle v_1, v_1 \rangle = 2$. We have $u_1 = \frac{1}{\|v_1\|} v_1 = \frac{1}{\sqrt{2}}(1, 0, 1, 0)$. Next,

$$\begin{aligned} w_2 &= v_2 - \langle v_2, u_1 \rangle u_1 \\ &= (0, 1, 0, 1) \end{aligned}$$

with $\|w_2\|^2 = \langle w_2, w_2 \rangle = 2$. Thus $u_2 = \frac{1}{\|w_2\|} w_2 = \frac{1}{\sqrt{2}}(0, 1, 0, 1)$. Finally,

$$\begin{aligned} w_3 &= v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2 \\ &= (-1, 0, 1, 0) \end{aligned}$$

with $\|w_3\|^2 = \langle w_3, w_3 \rangle = 2$. Therefore, $u_3 = \frac{1}{\|w_3\|} w_3 = \frac{1}{\sqrt{2}}(-1, 0, 1, 0)$. The ordered basis

$$\left(\frac{1}{\sqrt{2}}(1, 0, 1, 0), \frac{1}{\sqrt{2}}(0, 1, 0, 1), \frac{1}{\sqrt{2}}(-1, 0, 1, 0) \right)$$

is an orthonormal basis for W .

We can find the orthogonal projection of any vector $v \in V$ onto W as follows.

$$\text{proj}_W(v) = \langle v, u_1 \rangle u_1 + \langle v, u_2 \rangle u_2 + \langle v, u_3 \rangle u_3$$

The vector $\text{proj}_W(v)$ is the vector closest to v that lies in the subspace W where distance is determined by the inner product.

End of lecture 25

Example 21.7. Take the vector space $P_2(\mathbb{R})$ with inner product

$$\langle p, q \rangle = \int_{-1}^1 p(t)q(t)dt$$

for $p, q \in P_2(\mathbb{R})$ as in Example 19.4. Let $\mathcal{B} = (1, x, x^2)$ be the standard basis for $P_2(\mathbb{R})$. We will find a corresponding orthonormal basis using Gram-Schmidt.

We have $v_1 = 1$ so $\|v_1\|^2 = \langle v_1, v_1 \rangle = \int_{-1}^1 dt = 2$. Then $u_1 = \frac{1}{\sqrt{2}}$.

We have $v_2 = x$ so

$$\begin{aligned} w_2 &= v_2 - \langle v_2, u_1 \rangle u_1 \\ &= x - \frac{1}{\sqrt{2}} \left(\int_{-1}^1 \frac{1}{\sqrt{2}} t dt \right) \\ &= x - \frac{1}{2} \left[\frac{1}{2} t^2 \right]_{-1}^1 \\ &= x. \end{aligned}$$

Find the norm of w_2 ,

$$\begin{aligned} \|w_2\|^2 &= \langle w_2, w_2 \rangle \\ &= \int_{-1}^1 t^2 dt \\ &= \left[\frac{1}{3} t^3 \right]_{-1}^1 \\ &= \frac{2}{3} \end{aligned}$$

We can then make w_2 a unit vector via

$$\begin{aligned} u_2 &= \frac{1}{\|w_2\|} w_2 \\ &= \sqrt{\frac{3}{2}} x. \end{aligned}$$

Finally, $v_3 = x^2$ so

$$\begin{aligned} w_3 &= v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2 \\ &= x^2 - \frac{1}{2} \int_{-1}^1 t^2 dt - \frac{3}{2} x \int_{-1}^1 t^3 dt \\ &= x^2 - \frac{1}{2} \left[\frac{1}{3} t^3 \right]_{-1}^1 - \frac{3}{2} x \left[\frac{1}{4} t^4 \right]_{-1}^1 \\ &= x^2 - \frac{1}{3}. \end{aligned}$$

Find the norm of w_3 ,

$$\begin{aligned}
 \|w_3\| &= \langle w_3, w_3 \rangle \\
 &= \int_{-1}^1 \left(t^2 - \frac{1}{3} \right)^2 dt \\
 &= \int_{-1}^1 \left(t^4 - \frac{2}{3}t^2 + \frac{1}{9} \right) dt \\
 &= \frac{2}{5} - \frac{4}{9} + \frac{2}{9} \\
 &= \frac{8}{45}.
 \end{aligned}$$

We can then make w_3 a unit vector via

$$\begin{aligned}
 u_3 &= \frac{1}{\|w_3\|} w_3 \\
 &= \sqrt{\frac{5}{8}} (3x^2 - 1).
 \end{aligned}$$

An orthonormal basis for $P_2(\mathbb{R})$ with respect to the given inner product is (u_1, u_2, u_3) .

End of lecture 26

End of final material

Definition 21.8. Let $W \subset V$ be a subspace of an inner product space. The **orthogonal complement** of W is the set

$$W^\perp := \{v \in V : \langle v, w \rangle = 0 \text{ for all } w \in W\}.$$

Proposition 21.9. Let $W \subset V$ be a subspace of an inner product space V . Then W^\perp is a subspace of V .

Proof.

- (1) By Proposition 19.7(iii), $\langle 0, w \rangle = 0$ for all $w \in W$ so $0 \in W^\perp$.
- (2) Let $u, v \in W^\perp$. Then $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle = 0 + 0 = 0$ for all $w \in W$ so $u + v \in W^\perp$.
- (3) Let $v \in W^\perp$ and $c \in \mathcal{F}$. Then $\langle cv, w \rangle = c\langle v, w \rangle = 0$ for all $w \in W$ so $cv \in W^\perp$.

□

The next result proves that it is sufficient to check elements of the orthogonal complement on a spanning set.

Lemma 21.10. Let $W \subset V$ be a subspace of an inner product space. If $v \in V$ is orthogonal to each element of a spanning set S of W , then $v \in W^\perp$.

Proof. Let $w \in W$. Then $w = \sum_{i=1}^k a_i w_i$ for $w_i \in S$ and $a_i \in \mathcal{F}$. We have

$$\langle v, w \rangle = \left\langle v, \sum_{i=1}^k a_i w_i \right\rangle = \sum_{i=1}^k \overline{a_i} \langle v, w_i \rangle = 0.$$

Thus $v \in W^\perp$.

□

Example 21.11. In the vector space \mathbb{C}^3 under the standard inner product, take the subspace $W = \text{Span}(\{e_1, e_2\})$. We want to find W^\perp . Let $v \in \mathbb{C}^3$ be a vector such that $\langle v, e_1 \rangle = 0$ and $\langle v, e_2 \rangle = 0$. By Lemma 21.10, this is sufficient to say that $v \in W^\perp$. We can write $v = a_1 e_1 + a_2 e_2 + a_3 e_3$. The assumptions imply that $a_1 = a_2 = 0$. Therefore, $W^\perp = \text{Span}(\{e_3\})$. Using careless language, the orthogonal complement to the xy -plane is the z -axis.

Proposition 21.12. Let $W \subset V$ be a subspace of a finite dimensional inner product space. Then

$$V = W \oplus W^\perp.$$

Proof. Let $\mathcal{B} = (w_1, \dots, w_k)$ be an orthonormal basis for W , which exists by Theorem 21.3(i). Let $v \in V$. Define an element of W as $w := \sum_{i=1}^k \langle v, w_i \rangle w_i$. Since \mathcal{B} is orthonormal,

$$\begin{aligned} \langle v - w, w_j \rangle &= \langle v, w_j \rangle - \langle w, w_j \rangle \\ &= \langle v, w_j \rangle - \left\langle \sum_{i=1}^k \langle v, w_i \rangle w_i, w_j \right\rangle \\ &= \langle v, w_j \rangle - \sum_{i=1}^k \langle v, w_i \rangle \langle w_i, w_j \rangle \\ &= \langle v, w_j \rangle - \langle v, w_j \rangle \\ &= 0. \end{aligned}$$

Lemma 21.10 implies $v - w$ is in W^\perp and $v = w + (v - w) \in W + W^\perp$. In other words, $V = W + W^\perp$.

Let $v \in W \cap W^\perp$. Then $\langle v, v \rangle = 0$ so $v = 0$ by Proposition 19.10. Thus $W \cap W^\perp = \{0\}$. \square

Remark 21.13. In the proof of Proposition 21.12, the definition of $w = \sum_{i=1}^k \langle v, w_i \rangle w_i$ is the projection of v onto the subspace W . We can prove that w is the vector in W that minimizes $\|v - w\|$. In other words, w is the vector closest to v that lies in W . The main takeaway from the section is that orthonormal bases make finding linear combinations and projections onto subspaces easy.

The projection technique works whenever we have an orthonormal basis for a subspace. However, if we try to project onto a subspace without first finding an orthonormal basis, we might get a completely wrong result as the next result illustrates.

Example 21.14. Find the projection of $v = (1, 2, -1)$ onto the subspace $W = \text{Span}(\{e_1 + e_2, e_2\})$. Note that $\{e_1 + e_2, e_2\}$ is a basis for W that is not orthonormal. We observe that $W = \text{Span}(\{e_1, e_2\})$ so we expect the projection of v onto W to be $(1, 2, 0)$. However,

$$\begin{aligned} \langle v, e_1 + e_2 \rangle &= 1 + 2 = 3 \\ \langle v, e_2 \rangle &= 2 \end{aligned}$$

so, if we misuse the formula from the proof of Proposition 21.12, the projection of v onto W is $3(e_1 + e_2) + 2e_2 = (3, 5, 0) \neq (1, 2, 0)$. The simple projection formula requires first finding an orthonormal basis for the subspace.

Corollary 21.14.1. Let $W \subset V$ be a subspace of a finite dimensional inner product space. Then

$$\dim(V) = \dim(W) + \dim(W^\perp).$$

Proof. Apply Corollary 10.17.1 to the result of Proposition 21.12. \square

22. ADJOINTS

For a matrix, we introduced the conjugate transpose in Definition 19.15. In the presence of an inner product, we can often generalize this notion to adjoints of linear operators. However, an adjoint to a linear operator does not always exist. We will prove that adjoints exist and are unique in a finite dimensional vector space. Pairing an operator with an adjoint is a common technique in algebra. Information about one can provide information about the other.

Lemma 22.1. Let V be a finite dimensional inner product space over \mathcal{F} with $g : V \rightarrow \mathcal{F}$ linear. Then there exists a unique $w \in V$ such that $g(v) = \langle v, w \rangle$ for all $v \in V$.

Proof. Let $\mathcal{B} = (u_1, \dots, u_n)$ be an orthonormal basis for V , which exists by Theorem 21.3(i). Let $w = \sum_{i=1}^n \overline{g(u_i)} u_i$. Define the linear function $h : V \rightarrow \mathcal{F}$ by $h(v) = \langle v, w \rangle$. For $1 \leq j \leq n$,

$$\begin{aligned} h(u_j) &= \left\langle u_j, \sum_{i=1}^n \overline{g(u_i)} u_i \right\rangle \\ &= \sum_{i=1}^n g(u_i) \langle u_j, u_i \rangle \\ &= g(u_j). \end{aligned}$$

Since g and h agree on \mathcal{B} , linearity proves that g and h agree on any linear combination of \mathcal{B} . Every vector $v \in V$ is an element of $\text{Span}(\mathcal{B})$ so $g = h$.

In order to show that w is unique, suppose that $g(v) = \langle v, w' \rangle$ for all $v \in V$. Then $\langle v, w' \rangle = \langle v, w \rangle$ for all $v \in V$ so Corollary 19.8.1 implies $w' = w$. \square

Proposition 22.2. Let V be a finite dimensional inner product space with $T \in \mathcal{L}(V)$. Then there exists a unique function $T^* : V \rightarrow V$ such that $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$ for all $v, w \in V$. Furthermore, T^* is linear.

Proof. Let $w \in V$. Define $g : V \rightarrow \mathcal{F}$ as $g(v) = \langle T(v), w \rangle$ for all $v \in V$. We first show that g is linear. Let $v_1, v_2 \in V$ and $c \in \mathcal{F}$. Then

$$\begin{aligned} g(v_1 + v_2) &= \langle T(v_1 + v_2), w \rangle \\ &= \langle T(v_1) + T(v_2), w \rangle \\ &= \langle T(v_1), w \rangle + \langle T(v_2), w \rangle \\ &= g(v_1) + g(v_2) \\ g(cv_1) &= \langle T(cv_1), w \rangle \\ &= \langle cT(v_1), w \rangle \\ &= c\langle T(v_1), w \rangle \\ &= cg(v_1). \end{aligned}$$

Lemma 22.1 provides a unique w' such that $g(v) = \langle v, w' \rangle$ for all $v \in V$. Define $T^*(w) = w'$ so $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$.

Next we prove linearity of T^* . Let $w, w' \in V$ and $c \in \mathcal{F}$. Then, for all $v \in V$,

$$\begin{aligned} \langle v, T^*(w + w') \rangle &= \langle T(v), w + w' \rangle \\ &= \langle T(v), w \rangle + \langle T(v), w' \rangle \\ &= \langle v, T^*(w) \rangle + \langle v, T^*(w') \rangle \\ &= \langle v, T^*(w) + T^*(w') \rangle \\ \langle v, T^*(cw) \rangle &= \langle T(v), cw \rangle \\ &= \overline{c} \langle T(v), w \rangle \\ &= \overline{c} \langle v, T^*(w) \rangle \\ &= \langle v, cT^*(w) \rangle. \end{aligned}$$

Corollary 19.8.1 implies $T^*(w + w') = T^*(w) + T^*(w')$ and $T^*(cw) = cT^*(w)$.

We finally prove uniqueness. Assume that there is another map $S : V \rightarrow V$ such that

$$\langle T(v), w \rangle = \langle v, S(w) \rangle$$

for all $v, w \in V$. Then $\langle v, S(w) \rangle = \langle v, T^*(w) \rangle$ for all $v, w \in V$. By Corollary 19.8.1, $S(w) = T^*(w)$ for all $w \in V$. Thus $S = T^*$, and T^* is unique. \square

Definition 22.3. Let V be an inner product space with $T \in \mathcal{L}(V)$. The **adjoint** of T , when it exists, is the unique, linear map $T^* : V \rightarrow V$ such that

$$\langle T(v), w \rangle = \langle v, T^*(w) \rangle$$

for all $v, w \in V$.

Remark 22.4. Adjoints do not always exist for operators on infinite dimensional vector spaces. When an adjoint does exist the proof of Proposition 22.2 shows that it is unique and linear.

Whenever we state a proposition or problem involving an adjoint, we will assume implicitly that we are working with a transformation T for which the adjoint exists.

Example 22.5. Take the inner product space \mathbb{R}^2 with the dot product. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3y \\ x - y \end{pmatrix}.$$

We claim that the adjoint $T^* : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is defined by

$$T^* \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ 3x - y \end{pmatrix}.$$

Let $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \in \mathbb{R}^2$. Then

$$\begin{aligned} \left\langle T \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} 3y_1 \\ x_1 - y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle = 3x_2y_1 + x_1y_2 - y_1y_2 \\ \left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, T^* \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ 3x_2 - y_2 \end{pmatrix} \right\rangle = x_1y_2 + 3x_2y_1 - y_1y_2. \end{aligned}$$

Since the two dot products coincide for any choice of vectors in \mathbb{R}^2 , we have defined T^* correctly.

Example 22.6. Let \mathbb{C}^2 be an inner product space with the standard inner product. Define $T : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ as

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2i & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2ix + 3y \\ x - y \end{pmatrix}.$$

We claim that the adjoint $T^* : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ is defined by

$$T^* \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2i & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2ix + y \\ 3x - y \end{pmatrix}.$$

Let $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \in \mathbb{C}^2$. Then

$$\begin{aligned} \left\langle T \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} 2ix_1 + 3y_1 \\ x_1 - y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle = 2ix_1\overline{x_2} + 3\overline{x_2}y_1 + x_1\overline{y_2} - y_1\overline{y_2} \\ \left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, T^* \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} -2ix_2 + y_2 \\ 3x_2 - y_2 \end{pmatrix} \right\rangle = x_1(\overline{-2ix_2}) + x_1\overline{y_2} + 3\overline{x_2}y_1 - y_1\overline{y_2} \\ &= 2ix_1\overline{x_2} + x_1\overline{y_2} + 3\overline{x_2}y_1 - y_1\overline{y_2}. \end{aligned}$$

Since the two inner products coincide for any choice of vectors in \mathbb{C}^2 , we have defined T^* correctly.

Examples 22.5 and 22.6 capture the motivation for the adjoint. The adjoint is an abstraction of the conjugate-transpose operation (which for real vector spaces is the transpose operation). Proposition 22.8 makes this precise and explains why the conjugate transpose of Definition 19.15 is often called the matrix adjoint. First, however, we need some useful properties of the adjoint

Proposition 22.7. Let V be an inner product space with linear operators $S : V \rightarrow V$ and $T : V \rightarrow V$ for which adjoints exist.

- (i) $S + T$ has an adjoint, and $(S + T)^* = S^* + T^*$.
- (ii) cT has an adjoint, and $(cT)^* = \bar{c}T^*$ for any $c \in \mathcal{F}$.
- (iii) $S \circ T$ has an adjoint, and $(S \circ T)^* = T^* \circ S^*$.
- (iv) T^* has an adjoint, and $(T^*)^* = T$.
- (v) The identity operator is its own adjoint.

Proof. Let $v, w \in V$ and $c \in \mathcal{F}$.

(i)

$$\begin{aligned} \langle (S + T)(v), w \rangle &= \langle S(v) + T(v), w \rangle = \langle S(v), w \rangle + \langle T(v), w \rangle = \langle v, S^*(w) \rangle + \langle v, T^*(w) \rangle \\ &= \langle v, S^*(w) + T^*(w) \rangle = \langle v, (S^* + T^*)(w) \rangle \end{aligned}$$

(ii)

$$\langle (cT)(v), w \rangle = \langle cT(v), w \rangle = c\langle T(v), w \rangle = c\langle v, T^*(w) \rangle = \langle v, \bar{c}T^*(w) \rangle$$

(iii)

$$\langle (S \circ T)(v), w \rangle = \langle S(T(v)), w \rangle = \langle T(v), S^*(w) \rangle = \langle v, T^*(S^*(w)) \rangle = \langle v, (T^* \circ S^*)(w) \rangle$$

(iv)

$$\langle T^*(v), w \rangle = \overline{\langle w, T^*(v) \rangle} = \overline{\langle T(w), v \rangle} = \langle v, T(w) \rangle$$

(v) Let $I : V \rightarrow V$ be the identity operator.

$$\langle I(v), w \rangle = \langle v, w \rangle = \langle v, I(w) \rangle$$

□

Proposition 22.8. Let V be a finite dimensional inner product space with $T \in \mathcal{L}(V)$. Let $\mathcal{B} = (u_1, \dots, u_n)$ be an orthonormal basis of V . Then

$$[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*.$$

Proof. Let $A = [T]_{\mathcal{B}}$ and $B = [T^*]_{\mathcal{B}}$. Since \mathcal{B} is orthonormal, Theorem 21.3(ii) shows

$$T^*(u_j) = \langle T^*(u_j), u_1 \rangle u_1 + \dots + \langle T^*(u_j), u_n \rangle u_n.$$

Thus $B_{ij} = \langle T^*(u_j), u_i \rangle$ and $A_{ij} = \langle T(u_j), u_i \rangle$. Note that by Proposition 22.7(iv),

$$\overline{A_{ji}} = \overline{\langle T(u_i), u_j \rangle} = \langle u_j, T(u_i) \rangle = \langle T^*(u_j), u_i \rangle = B_{ij}.$$

Since $\overline{A_{ji}} = B_{ij}$ it follows that $B = \bar{A}^T$.

□

Example 22.9. Recall from Example 15.6 the vector space \mathbb{R}^∞ of sequences of real numbers with finitely many non-zero entries. Since only finitely many terms in each sequence are non-zero, an inner product $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathcal{F}$ can be defined like the usual dot product. Recall the right shift operator $S : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$ that satisfies

$$S(a_0, a_1, a_2, \dots) = (0, a_0, a_1, \dots).$$

Define the “left-shift” operator $T : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$ as

$$T(a_0, a_1, a_2, \dots) = (a_1, a_2, a_3, \dots).$$

Let $(a_0, a_1, a_2, \dots), (b_0, b_1, b_2, \dots) \in \mathbb{R}^\infty$ be two suitable sequences. Then

$$\begin{aligned}\langle S(a_0, a_1, a_2, \dots), (b_0, b_1, b_2, \dots) \rangle &= \sum_{i=0}^{\infty} a_i b_{i+1} \\ \langle (a_0, a_1, a_2, \dots), T(b_0, b_1, b_2, \dots) \rangle &= \sum_{i=0}^{\infty} a_i b_{i+1}.\end{aligned}$$

Therefore, $T = S^*$.

23. NORMAL OPERATORS, SELF-ADJOINT OPERATORS, AND THE SPECTRAL THEOREM

Lemma 23.1. Let $T : V \rightarrow V$ be a linear operator on a finite dimensional inner product space V . If T has an eigenvector, then so does T^* .

Proof. Suppose that v is an eigenvector of T with corresponding eigenvalue λ . Then for any $w \in V$,

$$0 = \langle 0, w \rangle = \langle (T - \lambda I)(v), w \rangle = \langle v, (T - \lambda I)^*(w) \rangle = \langle v, (T^* - \bar{\lambda}I)(w) \rangle.$$

Thus v is orthogonal to the image of $T^* - \bar{\lambda}I$ so $T^* - \bar{\lambda}I$ is not surjective. By Proposition 16.7, T^* has eigenvalue $\bar{\lambda}$. \square

Theorem 23.2 (Schur's Theorem). Let T be a linear operator on a finite dimensional inner product space V . Suppose that the characteristic polynomial of T factors into linear polynomials. Then there exists an orthonormal basis \mathcal{D} for V such that the matrix $[T]_{\mathcal{D}}$ is upper triangular.

Proof. We will first find an ordered basis \mathcal{C} for which $[T]_{\mathcal{C}}$ is upper triangular. Since the characteristic polynomial of T factors into linear polynomials, T always has at least one eigenvalue λ with corresponding eigenvector v_1 . Then $\{v_1\}$ can be extended to an ordered basis \mathcal{C} of V by Proposition 10.13 for which

$$[T]_{\mathcal{C}} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where A is 1×1 , 0 is $(n-1) \times 1$, B is $1 \times (n-1)$, and C is $(n-1) \times (n-1)$.

Assume that there exists an ordered basis $\mathcal{B} = (v_1, \dots, v_n)$ for which

$$[T]_{\mathcal{B}} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where A is $k \times k$ upper triangular, 0 is $(n-k) \times k$, B is $k \times (n-k)$, and C is $(n-k) \times (n-k)$. Note that $W = \text{Span}(\{v_1, \dots, v_k\})$ is T -invariant. As a result, we can define the linear operator $T_W : W \rightarrow W$ by $T_W(w) = T(w)$ for all $w \in W$. The characteristic polynomial of $[T]_{\mathcal{B}}$ is the product of the characteristic polynomials of A and C . Since the characteristic polynomial of T factors into linear terms, we can find an eigenvalue μ of the matrix C over \mathcal{F} . Let w be a corresponding eigenvector. Then

$$T \begin{pmatrix} 0 \\ w \end{pmatrix} = \begin{pmatrix} Bw \\ Cw \end{pmatrix} = \begin{pmatrix} Bw \\ \mu w \end{pmatrix}.$$

By Proposition 10.13, the linearly independent set $\mathcal{C}' = \{v_1, \dots, v_k, w\}$ can be extended to an ordered basis \mathcal{C} for which

$$[T]_{\mathcal{C}} = \begin{pmatrix} A' & B' \\ 0 & C' \end{pmatrix}$$

where A' is $(k+1) \times (k+1)$ upper triangular, 0 is $(n-(k+1)) \times (k+1)$, B' is $(k+1) \times (n-(k+1))$, and C' is $(n-(k+1)) \times (n-(k+1))$. We have proven recursively that there is an ordered basis $\mathcal{C} = \{v_1, \dots, v_n\}$ for which $[T]_{\mathcal{C}}$ is upper triangular.

Apply Gram-Schmidt to \mathcal{C} to obtain an ordered orthonormal basis $\mathcal{D} = (u_1, \dots, u_n)$. We let $S_k = \{u_1, \dots, u_k\}$ and $S'_k = \{v_1, \dots, v_k\}$. Then $\text{Span}(S_k) = \text{Span}(S'_k)$ for all $1 \leq k \leq n$. Further, $T(u_k) \in \text{Span}(S_k)$ and $T(v_k) \in \text{Span}(S'_k)$ for all $1 \leq k \leq n$. We conclude $[T]_{\mathcal{D}}$ remains upper triangular. \square

Definition 23.3. Let V be an inner product space with linear operator $T : V \rightarrow V$ for which the adjoint exists. Then T is **normal** if $TT^* = T^*T$.

Proposition 23.4. Let V be an inner product space with normal linear operator $T : V \rightarrow V$.

- (i) $\|T(v)\| = \|T^*(v)\|$ for all $v \in V$.
- (ii) $T - cI$ is normal for all $c \in \mathcal{F}$.
- (iii) If $v \in V$ is an eigenvector of T with corresponding eigenvalue $\lambda \in \mathcal{F}$, then v is an eigenvector of T^* with corresponding eigenvalue $\bar{\lambda}$.
- (iv) If $\lambda_1, \lambda_2 \in \mathcal{F}$ are distinct eigenvalues of T with corresponding eigenvectors $v_1, v_2 \in V$, then v_1 and v_2 are orthogonal.

Proof.

- (i) For all $v \in V$, we have

$$\|T(v)\|^2 = \langle T(v), T(v) \rangle = \langle T^*T(v), v \rangle = \langle TT^*(v), v \rangle = \langle T^*(v), T^*(v) \rangle = \|T^*(v)\|^2.$$

- (ii) By Proposition 22.7 and the normality of T ,

$$\begin{aligned} (T - cI)(T - cI)^* &= (T - cI)(T^* - \bar{c}I) = TT^* - \bar{c}T - cT^* + |c|^2I = T^*T - \bar{c}T - cT^* + |c|^2I \\ &= (T^* - \bar{c}I)(T - cI) = (T - cI)^*(T - cI). \end{aligned}$$

- (iii) Suppose that $T(v) = \lambda v$ for some $v \in V$. Let $U = T - \lambda I$ so $U(v) = 0$ and U is normal by (ii). Thus (i) and Proposition 22.7 imply

$$0 = \|U(v)\| = \|U^*(v)\| = \|(T^* - \bar{\lambda}I)(v)\| = \|T^*(v) - \bar{\lambda}v\|.$$

By Proposition 19.10(ii), $T^*(v) = \bar{\lambda}v$.

- (iv) Let λ_1 and λ_2 be distinct eigenvalues of T with corresponding eigenvectors v_1 and v_2 respectively. By (iii),

$$\lambda_1 \langle v_1, v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \langle T(v_1), v_2 \rangle = \langle v_1, T^*(v_2) \rangle = \langle v_1, \bar{\lambda}_2 v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle.$$

Since $\lambda_1 \neq \lambda_2$, $\langle v_1, v_2 \rangle = 0$.

\square

Theorem 23.5 (Spectral Theorem for Normal Operators). Let V be a finite dimensional complex inner product space with $T : V \rightarrow V$ a linear operator. Then T is normal if and only if there exists an orthonormal basis for V consisting of eigenvectors of T .

Proof. (\Rightarrow) Suppose that T is normal. By the Fundamental Theorem of Algebra, the characteristic polynomial of T factors into linear terms. Apply Schur's Theorem to obtain an orthonormal basis $\mathcal{B} = (u_1, \dots, u_n)$ for V such that $[T]_{\mathcal{B}}$ is upper triangular.

We will prove that each u_i is an eigenvector of T via induction. Note that u_1 is an eigenvector of T by construction. Assume that u_1, \dots, u_{k-1} are eigenvectors of T . We will show that u_k is also an eigenvector of T . For $j < k$, let $\lambda_j \in \mathcal{F}$ be the eigenvalue corresponding to u_j . Since $[T]_{\mathcal{B}}$ is upper triangular, $T(u_k) = \sum_{i=1}^k A_{ik} u_i$. By Theorem 21.3(ii), Proposition 23.4(iii), and the orthonormality of \mathcal{B} ,

$$A_{ik} = \langle T(u_k), u_i \rangle = \langle u_k, T^*(u_i) \rangle = \langle u_k, \bar{\lambda}_i u_i \rangle = \lambda_i \langle u_k, u_i \rangle = 0.$$

for $1 \leq i \leq k-1$. Thus $T(u_k) = A_{kk} u_k$, and u_k is an eigenvector of T .

(\Leftarrow) Assume that there exists an orthonormal basis \mathcal{B} for V consisting of eigenvectors of T . Then $[T]_{\mathcal{B}}$ is diagonal. By Proposition 22.8, $[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*$ is the conjugate transpose of a diagonal matrix. Thus $[T^*]_{\mathcal{B}}$ is also diagonal. We conclude that T and T^* commute. \square

Definition 23.6. Let V be an inner product space with linear operator $T : V \rightarrow V$ for which the adjoint exists. Then T is **self-adjoint** if $T = T^*$.

Remark 23.7. A self-adjoint linear operator T is automatically normal since $TT^* = T^2 = T^*T$.

Lemma 23.8. Let V be a finite dimensional inner product space with self-adjoint linear operator $T : V \rightarrow V$.

- (i) Every eigenvalue of T is real.
- (ii) Suppose V is a real inner product space. Then the characteristic polynomial of T factors into linear terms.

Proof.

- (i) Suppose that $T(v) = \lambda v$ for non-zero $v \in V$ and $\lambda \in \mathbb{C}$. Since T is normal, Proposition 23.4(iii)

$$\lambda v = T(v) = T^*(v) = \bar{\lambda}v.$$

Thus $\lambda = \bar{\lambda}$ and λ is real.

- (ii) Let $n = \dim(V)$ with \mathcal{B} an orthonormal basis for V . Denote $A = [T]_{\mathcal{B}}$, which is a real symmetric matrix. Define $T_A : \mathbb{C}^n \rightarrow \mathbb{C}^n$ by $T_A(v) = Av$ for $v \in \mathbb{C}^n$. Since T_A is represented by a real symmetric matrix, T_A is self-adjoint over the complex inner product space \mathbb{C}^n . Let $f(t)$ be the characteristic polynomial of T_A . By the Fundamental Theorem of Algebra, $f(t)$ factors as a product of terms of the form $(t - \lambda)$. Each λ is a root of f and, thus, an eigenvalue of T_A . By (i), λ is real. Therefore, $f(t)$ factors as a product of linear terms over the real numbers. The characteristic polynomial of T is the same as the characteristic polynomial of T_A .

\square

Theorem 23.9 (Spectral Theorem for Self-Adjoint Operators). Let V be a finite dimensional real inner product space with linear operator $T : V \rightarrow V$. Then T is self-adjoint if and only if there exists an orthonormal basis for V consisting of eigenvectors of T .

Proof. (\Rightarrow) Suppose T is self-adjoint. By Lemma 23.8(ii), the characteristic polynomial of T factors into linear terms. Apply Schur's Theorem to obtain an orthonormal basis \mathcal{B} for V such that $[T]_{\mathcal{B}}$ is upper triangular. We have $[T]_{\mathcal{B}}^* = [T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}$ so $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}}^*$ are upper triangular. Since $[T]_{\mathcal{B}}^*$ is the conjugate transpose of $[T]_{\mathcal{B}}$, we conclude that $[T]_{\mathcal{B}}$ is diagonal. Therefore, each vector in the basis \mathcal{B} is an eigenvector.

(\Leftarrow) Assume there exists an orthonormal basis $\mathcal{B} = (u_1, \dots, u_n)$ for V consisting of eigenvectors of T . Let λ_i be the eigenvalue corresponding to the eigenvector u_i . Then $[T]_{\mathcal{B}}$ is a diagonal matrix with λ_i in the i th diagonal position. By Proposition 22.8, $[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*$. The entry in the i th diagonal of $[T^*]_{\mathcal{B}}$ is $\bar{\lambda}_i$ by Proposition 23.4(iii). However, λ_i is real so $[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}$. We conclude that $T = T^*$, and T is self-adjoint. \square