COURSE NOTES FOR MATH 540, HONORS LINEAR ALGEBRA

MELODY CHAN

Contents

Date: August 5, 2021.

These are evolving course notes for Math 540, Honors linear algebra, at Brown University.

Part 1. Sets, functions, and $\mathbb{R}^n$.

This course weaves the topic of sets, logic, and proof into the course, studying them through the lens of linear algebra. One could teach a whole course on elements of proof, and fortunately there is one: Math 1001, offered in 2021-2022 by Professor Jordan Kostiuk! For now, a great reference is [Ham18, Chapter 1].

1. Sets

Mathematics as it is studied and communicated today is built on set theory.

Definition 1.1. A set is collection of things. The things in the set are called its *elements* or *members*.

That's the best description we can give at the moment.

Sets are sometimes written with curly brackets enclosing a list, or at least a partial list, of elements, separated by commas. For example: the integers

Definition 1.2. We let

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\} \quad \text{and} \quad \mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, \ldots\},$$

called the *integers*, respectively *nonnegative integers*.

Here the $\mathbb{Z}$ is for the German *Zahlen*, numbers.

Remark 1.3. Discuss the grammar of the above *notation-setting* sentence.

Definition 1.4.

(1) Let $\mathbb{Q} =$

(2) Let $\mathbb{R} =$

(3) Let $\mathbb{C} =$

Definition 1.5. Define the notation for sets and elements $\in, \ni, \subseteq, \supseteq, =$: Let X, Y be sets.

(1) $x \in X$ means:

(2) $x \notin X$ means:

(3) $X \subseteq Y$ means:

(4) $Y \supseteq X$ means:

(5) $X = Y$ means:

2. New sets from old

We follow [Ham18, p.5] closely. A common way to describe sets is what Hammack calls **set-builder notation**. Basically, you specify a set by giving

$$X = \{\text{expression} : \text{rule}\}.$$

For example, the set of even integers $E = \{\ldots, -6, -4, -2, 0, 2, 4, 6, \ldots\}$ can be written in set-builder notation in many ways:

$$\begin{aligned} E &= \{2n : n \in \mathbb{Z}\} \\ E &= \{n : n \in \mathbb{Z} \text{ is even}\} = \{n \in \mathbb{Z} : n \text{ is even}\} \\ E &= \{n \in \mathbb{Z} : \text{there exists } k \in \mathbb{Z} \text{ such that } 2k = n\} \end{aligned}$$

The curly brackets are pronounced "the set of all" and the colon is pronounced "such that." (Sometimes people write | instead of the colon. They are synonymous. I use them indiscriminately, depending on my mood.)

It is important to practice translating, both ways, between set-builder notation and natural language descriptions of sets. One of the more unfamiliar feelings you may need to get used to is the need to *name things*, i.e., assign symbols to quantities.

Example 1.6.

(1) Set-builder notation to English: Describe in words, as naturally as possible, the set
$$\{a \in \mathbb{Q} : 2a \in \mathbb{Z}\}.$$

(2) English to set-builder notation: Describe in set-builder notation the set of real numbers whose squares are rational numbers.

Definition 1.7. Let A and B be sets. We let

(1) the *intersection* of A and B to be
$$A \cap B =$$

(2) the *union* of A and B to be
$$A \cup B =$$

(3) the *set-theoretic difference* of A and B to be
$$A \setminus B =$$

3. Cartesian product

Definition 1.8. An **ordered pair** is a list of two things (x, y) separated by a comma, inside parentheses.

An **ordered n-tuple** is a list of n things $(x_1, \ldots, x_n)$ separated by a comma, inside parentheses.

Definition 1.9. Let A, B be sets. The product, or **Cartesian product**, of A and B, denoted $A \times B$, is the set

$$\{(a, b) : a \in A \text{ and } b \in B\}.$$

Definition 1.10. Similarly, define the Cartesian product of n sets $A_1, \ldots, A_n$:

If A is a set and n is a positive integer, then A^n denotes the n-fold Cartesian product

$$A \times \cdots \times A.$$

Example 1.11. Let's pause to draw $\mathbb{R}^2$, the familiar Cartesian plane, and $\mathbb{R}^3$.

4. $\mathbb{R}^n$ as a vector space

We are going to carry $\mathbb{R}^2$ and $\mathbb{R}^3$ through this class as running examples of **vector spaces**, so let's get started! Source: [Axl15, §1, pp. 7-10]

Let $n \geq 0$ be a positive integer; $n = 2$ or $n = 3$ are the best examples to keep in mind to start with. (The case $n = 0$ is permitted, but is mind-bending the first time you think it through!)

Definition 1.12.

(1) (Addition in $\mathbb{R}^n$) We define an addition operation on $\mathbb{R}^n$ by adding in each coordinate:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) =$$

(2) (Scalar multiplication in $\mathbb{R}^n$) We define, for each $\lambda \in \mathbb{R}$ and $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\lambda \cdot (x_1, \ldots, x_n) =$$

Often we omit the $\cdot$ for short.

Remark 1.13. Discuss how to add and scale vectors in $\mathbb{R}^2$; draw pictures.

5. Functions

Let A and B be sets. What is a function $f\colon A \to B$? Informally it is a rule that determininstically associates some $b \in B$ to each $a \in A$. We write $f(a) = b$ in that situation, or $a \mapsto b$. In this situation A is called the **domain** and B is called the **codomain** or **target**.

Definition 1.14. Formally, a function $f\colon A \to B$ is a subset I'll denote G_f[1] of $A \times B$ with the property that for every $a \in A$, there is a *unique* $b \in B$ with $(a, b) \in G_f$. In this situation, for $a \in A$, we write

$$f(a) = b$$

or

$$a \mapsto b$$

for this unique $b \in B$.

Definition 1.15. Let $f\colon A \to B$ be a function.

(1) We say f is **injective** or is an **injection** if:

(2) We say f is **surjective** or is a **surjection** if:

(3) We say f is **bijective** if or is a **bijection** if:

[1]Visualize G_f as the "graph of f."

Part 2. Fields and vector spaces

Reference throughout: [Axl15].

6. Fields

Reference: [Axl15, page 3]. The real numbers $\mathbb{R}$ with their usual notions of addition and multiplication are an example of a *field*.

Definition 2.1. A field is a set $\mathbb{F}$ containing elements named 0 and 1, together with binary operations $+$ and $\cdot$ on $\mathbb{F}$, satisfying:

- commutativity:
- associativity:
- identities:
- additive inverse:
- multiplicative inverse of nonzero elements:
- distributive property:

A good second example of a field is the field of **complex numbers** $\mathbb{C}$.

Definition 2.2. The field of **complex numbers** $\mathbb{C}$:

Let's verify that $\mathbb{C}$ is a field. Let's do the most interesting (to me) of the verifications, namely the existence of multiplicative inverses.

Proposition 2.3. For every $\alpha \in \mathbb{C}$ with $\alpha \neq 0$, there exists $\beta \in \mathbb{C}$ with $\alpha\beta = 1$.

Let's do scratchwork before doing the proof: we have

$$(a+bi)(a-bi) = a^2 - b^2, \qquad \text{so } (a+bi)(a-bi)/(a^2-b^2) = 1.$$

Proof.

□

Example 2.4. A different kind of example, just so you know they're out there: let p be a prime number, say $p = 2$ or $p = 5$. Let's define the field $\mathbb{F}_p$ of p elements, and write out its addition and multiplication tables (think "clock arithmetic".)

Finite fields—that is, fields with only finitely many elements—are a big deal in cryptography/number theory these days.

Remark 2.5. A lovely digression. What goes wrong if you try the above definition for p a composite number, say $p = 4$? Try it.

So that multiplication table doesn't work, but remarkably it turns out that there *is* a field with exactly four elements, it's just not given by clock arithmetic on the numbers $0, 1, 2, 3$. The version I have in mind has 4 elements, called $0, 1, x, x+1$, with addition and multiplication tables:

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

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

I remember that as "polynomials in x with coefficients in the field $\mathbb{F}_2$, where $1 + 1 = 0$, subject to the rule $x^2 = x + 1$."

How did I figure this out? Please take Math 1530/1540 to find out. Remarkably *for an integer $n > 1$, there is a field with n elements precisely when n is a power of a prime!*

Now that you know how rich and different fields can be, it should feel more empowering to prove innocent-sounding statements like "zero times anything is always zero" in *any* field. Okay, that's "obvious" in $\mathbb{R}$, say, but why does it definitely hold for *all* fields?

Proposition 2.6. In any field $\mathbb{F}$, $0 \cdot x = 0$ for any $x \in \mathbb{F}$.

Proof.

□

7. The definition of a vector space

See [Axl15, pp. 12-13]. Fix a field $\mathbb{F}$.

Definition 2.7. A **vector space** is a set V, together with an addition operation $+$ on V and a scalar multiplication on V such that:

- commutativity:
- associativity of addition and of scalar multiplication:
- additive identity:
- additive inverse:
- multiplicative identity:
- distributive properties:

Definition 2.8. Elements of a vector space are called **vectors**, or sometimes, when thinking geometrically, **points.**

Example 2.9. Examples of vector spaces over $\mathbb{F}$.

(1) $\mathbb{F}^n$, for some $n \geq 0$. (We interpret $\mathbb{F}^0$ to be the vector space $\{0\}$, with a single vector 0.[2])
(2) $\mathbb{F}^\infty = \{(x_1, x_2, \ldots) : x_j \in \mathbb{F} \text{ for } j = 1, 2, \ldots\}$. Addition and scalar multiplication are defined as expected:
(3) More generally, let S be any set. Then $\mathbb{F}^S$ denotes the set of functions from S to $\mathbb{F}$. Define addition and scalar multiplication on $\mathbb{F}^S$, and discuss why this is a vector space.

[2]This actually makes sense per item (3) below—this takes some careful thought about the empty set. But it's fine just to remember this as a convention, or, frankly, not to remember it at all.

Remark 2.10. Discuss why we said "more generally" above. For example, if $S = \{1, 2, 3\}$ then $\mathbb{R}^S = \mathbb{R}^3$. In other words, a function $\{1, 2, 3\} \to \mathbb{R}$ is the same thing[3] as an ordered triple (x_1, x_2, x_3) of real numbers.

Example 2.11. Other examples: continuous functions $\mathbb{R} \to \mathbb{R}$. Polynomials in x over $\mathbb{R}$, denoted $\mathbb{R}[x]$.

Let V be a vector space over $\mathbb{F}$. Let's establish some basic properties of vector spaces. This is both mundane and kinda remarkable at the same time. After all, these properties might seem obvious for examples of vector spaces, like $\mathbb{R}^n$, that you already know. But these follow from the definitions of vector spaces, so they apply to *all* vector spaces.

Proposition 2.12. V has a *unique* additive identity.

Proof. □

Proposition 2.13. Every vector $v \in V$ has a *unique* additive inverse. We will denote this additive inverse $-v$.

Proof. □

Proposition 2.14. For all $v \in V$,

$$0 \cdot v = 0.$$

Proof. □

[3]OK, not quite the same thing. But given a function $\{1, 2, 3\} \to \mathbb{R}$, how would you extract an ordered triple (x_1, x_2, x_3) of real numbers? And conversely? So to be perfectly precise, we have constructed a bijection between the elements of the vector space of functions $\{1, 2, 3\} \to \mathbb{R}$, and the elements of the vector space $\mathbb{R}^3$ of ordered triples of reals. This bijection preserves the operations of addition and scalar multiplication (make this precise).

Proposition 2.15. For all $a \in \mathbb{F}$,

$$a \cdot \mathbf{0} = \mathbf{0}.$$

(Here $\mathbf{0}$ denotes the zero vector. Above, 0 denoted the zero *scalar* in $\mathbb{F}$.)

Proof. □

Proposition 2.16. $(-1) \cdot v = -v$ for all $v \in V$.

Parse this one slowly!

Proof. □

8. Subspaces

Let V be a vector space over $\mathbb{F}$.

Definition 2.17. A subset U of V is a **linear subspace** or **subspace** of V if

(1) $\mathbf{0} \in U$;
(2) $u, w \in U$ implies $u + w \in U$,
(3) $a \in \mathbb{F}$ and $u \in U$ implies $au \in U$.

Proposition 2.18. A subspace of a vector space is itself a vector space, using the same addition and scalar multiplication as on V.

Proof. Let U be a subspace of a vector space V. The point is that if U satisfies the above conditions (2) and (3), then U is *closed* under addition and scalar multiplication. Still, there are other things to check to check that U is a vector space:

- Commutativity of $+$ and associativity of $+$ and $\cdot$ hold because they hold in V.
- The existence of an additive identity holds by (1).
- The existence of additive inverses holds: given $u \in U$, we know $(-1)u \in U$ by (3), and $u + (-1)u = \mathbf{0}$.
- The property that $1u = u$ for all $u \in U$ holds because $1v = v$ holds for all $v \in V$.
- The distributive laws hold because they hold in V.

□

Example 2.19. Let V be the space of functions $\mathbb{R} \to \mathbb{R}$. Then the subset U of *continuous* functions $\mathbb{R} \to \mathbb{R}$ is a subspace.

Example 2.20. The subset

$$\{f \in \mathbb{R}[x] : f(5) = 0\}$$

forms a subspace of $\mathbb{R}[x]$. But not

$$\{f \in \mathbb{R}[x] : f(5) = 1\}$$

Example 2.21. What are all of the subspaces of $\mathbb{R}^2$?

9. Span

The following is conceptually very important. Suppose I have a subspace W of $\mathbb{R}^2$ in mind but I'm keeping it a secret. But you figure out that the vectors $(2, 1)$ and $(1, 2)$ are in my subspace. What else can you conclude is in my subspace?

Let V be a vector space over $\mathbb{F}$.

Definition 2.22. A **linear combination** of a list $v_1, \ldots, v_m$ of vectors in V is a vector of the form:

Example 2.23. Do an example.

Suppose $v_1, \ldots, v_m \in V$.

Definition 2.24. The **linear span** or **span** of $v_1, \ldots, v_m$, denoted $\text{span}(v_1, \ldots, v_m)$, is the set of all linear combinations of $v_1, \ldots, v_m$.

$$\{a_1 v_1 + \cdots + a_m v_m : a_1, \ldots, a_m \in \mathbb{F}\}.$$

The span of *no vectors at all* is defined to be $\{0\}$. What? That convention fits conceptually with the following proposition:

Proposition 2.25. The span of $v_1, \ldots, v_m$ is the smallest subspace of V containing $v_1, \ldots, v_m$.

We mean smallest in terms of *containment.* In other words, the proposition consists of two statements:

(1) $\text{span}(v_1, \ldots, v_m)$ is a subspace of V.

(2) If W is a subspace of V and $v_1, \ldots, v_m \in W$, then also $\operatorname{span}(v_1, \ldots, v_m) \subseteq W$.

Proof. □

Definition 2.26.

(1) If $\operatorname{span}(v_1, \ldots, v_m) = V$ then say that $v_1, \ldots, v_m$ *spans* V.
(2) We call V *finite-dimensional* if there is a list (by which we mean a finite list) of vectors that span V.

Example 2.27. (An example of an infinite-dimensional vector space) If time permits: discuss [Axl15, p.31] of the space of polynomials $\mathcal{P}(\mathbb{F})$ in one variable.

Part 3. Linear maps

10. The definition of a linear map

Let $\mathbb{F}$ denote a field, and let V, W be vector spaces over $\mathbb{F}$. What *should* a linear map be? Discuss this.

Definition 3.1. A function $T\colon V \to W$ is called a *linear map* if it satisfies:

The set of linear maps from $V \to W$ is denoted $\mathcal{L}(V, W)$.

So it isn't a requirement that a linear map $V \to W$ send $\mathbf{0} \in V$ to $\mathbf{0} \in W$. Rather, it follows from the definition, as we now show.

Proposition 3.2. Let $T\colon V \to W$ be a linear map. Then $T(\mathbf{0}) = \mathbf{0}$.

Proof. We have $T(\mathbf{0}) = T(\mathbf{0}+\mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$, by linearity. Now add the additive inverse of $T(\mathbf{0})$ to both sides; we obtain $\mathbf{0} = T(\mathbf{0})$. □

Example 3.3. Let $T\colon \mathbb{R} \to \mathbb{R}$ be given by $T(x) = 5x$. Is T a linear map? What are all linear maps from $\mathbb{R} \to \mathbb{R}$?

Example 3.4. Important building block examples: the zero map $\mathbf{0}\colon V \to W$. Or, in the case of domain=codomain, the identity map $I\colon V \to V$.

Example 3.5. Differentiation is a linear map on polynomial functions:

The following is conceptually crucial:

Example 3.6. (Conceptually important!) Suppose I promise you that I have a linear map $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ in mind, but I'm reluctant to tell it to you. But I do let on that

$$T(1,0) = (2,1), \qquad T(0,1) = (1,-1).$$

What else can you figure out about T?

Remark 3.7. Also, it's hard, but how you might *draw* the linear map T?

11. Null space and range

Let $T\colon V \to W$ be a linear map.

Definition 3.8. The **nullspace** or **kernel** of T, written $\operatorname{null} T$, is "the stuff that is sent to 0":

Definition 3.9. The **range** of T is, well, the range:

Proposition 3.10. $\operatorname{null} T$ and $\operatorname{range} T$ are subspaces (of V and W respectively.)

Proof.

□

The words injective, surjective, and bijective apply to linear maps too. Injectivity deserves more mention because it turns out there is a convenient way to check injectivity of linear maps. Note that if a linear map T is injective, then in particular it doesn't send two distinct elements of V both to $\mathbf{0} \in W$. In particular, since $T(\mathbf{0}) = \mathbf{0}$, we conclude that $\operatorname{null} T = \{0\}$ for an injective linear map T.

Actually, the converse holds too:

Proposition 3.11. Let $T\colon V \to W$ be a linear map. If $\operatorname{null}(T) = \{\mathbf{0}\}$ then T is injective.

Proof. Suppose $\operatorname{null} T = \{\mathbf{0}\}$; we wish to show that T is injective. Given $v, v' \in V$ such that $Tv = Tv'$, we wish to deduce $v = v'$. Indeed,

$$\mathbf{0} = Tv' + (-Tv) = Tv' + T(-v) = T(v' + (-v)),$$

so $v' + (-v) = \mathbf{0}$ by the assumption on the nullspace. So $v' = v$ by adding v to both sides. □

Part 4. Linear independence, bases, and dimension

12. Linear independence

Let V be a vector space.

Definition 4.1. A list $v_1, \dots, v_m$ of vectors in V is called *linearly independent* if for any $a_1, \dots, a_m \in \mathbb{F}$,

$$a_1 v_1 + \cdots + a_m v_m = 0 \text{ implies } a_1 = \cdots = a_m = 0.$$

The empty list of vectors is also declared to be linearly independent.

The definition of linear independence above is the most useful for proving things. But conceptually I find it easier (at first) to understand the *negation*, which is called linear dependence:

Definition 4.2. A list $v_1, \dots, v_m$ of vectors in V is called *linearly dependent* if they are not linearly independent.

Exercise 4.3. Write this explicitly using quantifiers. At this point it's good to nail down how to negate a statement P implies Q.

Thus, observe that $v_1, \dots, v_m$ are linearly *dependent* iff one of the vectors is in the span of the others. (Do you agree?) This is the easiest criterion for me to use for my intuition, because I have an easier time picturing *spans*.

In fact we can make the more precise claim below, which will be instrumental in the next proofs.

Lemma 4.4. (Linear dependence lemma)[4] If $v_1, \dots, v_m \in V$ are linearly dependent, there is some v_j such that

(1) $v_j \in \operatorname{span}(v_1, \dots, v_{j-1})$, and
(2) the span of the remaining vectors $v_1, \dots, \hat{v_j}, \dots, v_m$ equals $\operatorname{span}(v_1, \dots, v_m)$.[5]

Proof. Proved in class. See [Axl15, p. 34]. □

[4] A *lemma* is a little thing that you prove en route to a cooler thing.

[5] Discuss hat notation! It means "with v_j removed."

Proposition 4.5. Let V be a finite-dimensional vector space. The length of every list of linearly independent vectors is at most the length of every spanning set.

Proof. Proved in class. See [Axl15, p. 35]. □

Corollary 4.6. Every subspace of a finite-dimensional vector space is finite-dimensional.

Proof. Proved in class. See [Axl15, p. 36]. □

13. Bases

Proposition 4.7. Vectors $v_1, \ldots, v_m \in V$ are linearly independent iff for any vector v in their span, there exists a *unique* choice of $a_1, \ldots, a_m \in \mathbb{F}$ such that

$$v = a_1 v_1 + \cdots + a_m v_m.$$

Proof. Omitted.

The best situation is when you have a spanning set that is also linearly independent.

Definition 4.8. A *basis* of V is a list of linearly independent vectors that spans V.

Example 4.9. A good example is the *standard basis* of $\mathbb{R}^n$:

$$e_1 = (1, 0, \ldots, 0),\ e_2 = (0, 1, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1).$$

But there are *lots* of other bases. What are all the bases of $\mathbb{R}^2$? What do they look like? Work this out.

Bases are completely crucial to working with vector spaces! Why? If $v_1, \ldots, v_m$ is a basis of V, then

(1) the fact that $v_1, \ldots, v_m$ span V means that each vector can be written in the form
$$v = a_1 v_1 + \cdots + a_m v_m, \qquad \text{for some } a_1, \ldots, a_m \in \mathbb{F},$$
(2) and the fact that $v_1, \ldots, v_m$ are LI implies that those $a_1, \ldots, a_m$ are determined uniquely. Indeed, if
$$a_1 v_1 + \cdots + a_m v_m = b_1 v_1 + \cdots + b_m v_m,$$
then
$$(a_1 - b_1)v_1 + \cdots + (a_m - b_m)v_m = \mathbf{0},$$
so by linear independence, $a_1 - b_1 = \cdots = a_m - b_m = 0$.

The converse holds too. (Argue this.) Putting together the statement above and its converse, we conclude:

Proposition 4.10. A list of vectors $v_1, \ldots, v_m$ is a basis of V iff for every $v \in V$, there exist *unique* $a_1, \ldots, a_m \in \mathbb{F}$ such that
$$v = a_1 v_1 + \cdots + a_m v_m.$$

Now we prove that every spanning list can be whittled down to a basis, and that every linearly independent list can be beefed up into a basis.

Let V be a vector space.

Proposition 4.11. Every spanning list of V contains a basis of V.

Proof. See [Axl15, p.40].

□

Corollary 4.12. Every finite-dimensional vector space has a basis.

Proposition 4.13. Every linearly independent list of vectors in a finite-dimensional vector space V can be extended to a basis of V.

Proof. See [Axl15, p.41].

□

Both of these propositions are proven using a "greedy algorithm." This is an umbrella term for any kind of procedure to work towards a goal in which *you can never paint yourself into a corner.* As long as you're making progress towards your goal, you're good. The "greedy" nature of both of these procedures (whittling down spanning lists, beefing up linearly independent lists) is conceptually quite deep.

14. Dimension

We have everything we need to define the *dimension* of a vector space. This is the most important number you can associate to a finite-dimensional vector space: it is the length of any basis. Indeed:

Proposition 4.14. Any two bases of a finite-dimensional vector space V have the same length.

By "length" we just mean number of elements.

Proof.

□

Definition 4.15. The dimension of a finite-dimensional vector space V is the length of any basis. It is denoted $\dim V$.

Example 4.16. The dimension of $\mathbb{F}^n$ is n. What is the dimension of $\{0\}$?

Proposition 4.17. Let V be a vector space of dimension n.

(1) Every linearly independent list of vectors in V of length n is a basis.
(2) Every spanning list of vectors in V of length n is a basis.

15. The rank-nullity theorem

Axler calls this the fundamental theorem of linear maps [Axl15, p. 63]. Let V and W be vector spaces over $\mathbb{F}$.

Theorem 4.18. Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$ is a linear map to W. Then $\operatorname{range} T$ is finite-dimensional and

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Proof.

□

This theorem is *powerful*. Here is an example:

Consider 5 homogeneous linear equations in 6 variables. For example

$$a_1x_1 + \cdots + a_6x_6 = 0$$

is one such equation (for scalars $a_1, \ldots, a_6$). *Homogeneous* means the right hand side of the equation is 0.

Does this system of equations have a solution? Well yes, here's one: $x_1 = \cdots = x_6 = 0$. Does it have any others? Discuss, and convince ourselves of:

Proposition 4.19. A homogeneous system of linear equations with more variables than equations has a nonzero solution.

Amazing!

Part 5. Telling your friend a linear map

How do you tell your friend a linear map? In a text message, as efficiently as possible?

16. Linear maps are freely and uniquely determined by what they do to a basis

Axler slips this in without much fanfare, but this is for me the absolute most important result in the book (in my opinion). See [Axl15, p. 54]. Let V and W be vector spaces over $\mathbb{F}$.

Theorem 5.1. Suppose $v_1, \ldots, v_n$ is a basis for V, and $w_1, \ldots, w_n \in W$ are any vectors. Then there exists a unique linear map $T\colon V \to W$ such that

$$T(v_1) = w_1, \ldots, T(v_n) = w_n.$$

My slogan for this theorem is: Linear maps are *freely and uniquely* determined by what they do to a basis. This theorem was the challenging second problem during the Proof Clinic.

By "freely" I mean that once you pick a basis $v_1, \ldots, v_n$ of V, a three-year-old could pick vectors w_i completely according to their whim and there would be a linear map taking v_1 to w_1, v_2 to w_2, etc. *There are no constraints on the w_i needed to ensure that such a linear map exists.* By "uniquely" I mean that there is a unique such linear map.

Proof.

□

17. The matrix associated to a linear map

Textbook reference: [Axl15, p. 70]. First of all, an $m \times n$ *matrix* over a field $\mathbb{F}$ is just a rectangular array, m rows and n columns, filled with elements of $\mathbb{F}$. The entry of a matrix A in row j, column k is denoted $A_{j,k}$.

Definition 5.2. Let $T\colon V \to W$ be a lienar map, $v_1, \ldots, v_n$ a basis for V, and $w_1, \ldots, w_m$ a basis for W. The matrix of T with respect to these bases is the $m \times n$ matrix, denoted $A = \mathcal{M}(T)$, whose entries $A_{j,k}$ are defined by

$$T(v_k) = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

Remark 5.3. Important remark: Note that since $w_1, \ldots, w_m$ is a basis, the scalars $A_{j,k}$ are determined uniquely. That is, there's always a unique way to write each $T(v_k)$ as a linear combinations of $w_1, \ldots, w_m$. In this way, $\mathcal{M}(T)$ is determined *uniquely* by T.

The indexing gets finnicky here, admittedly. The notational conventions are perhaps best illustrated by example:

Example 5.4. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the unique linear map such that

$$T(1,0) = (1,2,7), \qquad T(0,1) = (3,5,9).$$

Then with respect to the standard bases of $\mathbb{R}^2$ and $\mathbb{R}^3$,

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}.$$

I remember the notation by: the k^{th} column of the matrix records "where v_k goes." The point is that, *once bases of V and W respectively are agreed upon,* a matrix $\mathcal{M}(T)$ **encodes** T without losing information; it's like shorthand for T.

18. $\mathcal{L}(V, W)$ as a vector space.

Source: [Axl15, pp. 52-55] Let V and W be vector spaces over a field $\mathbb{F}$.

Definition 5.5. The set of linear maps from V to W is denoted $\mathcal{L}(V, W)$.

In fact $\mathcal{L}(V, W)$ can be given the structure of a vector space! By this, I mean that we can, and will, define addition and scalar multiplication on $\mathcal{L}(V, W)$, in such a way that the properties of a vector space are satisfied. As follows:

Definition 5.6. (Definition of addition and scalar multiplication on $\mathcal{L}(V, W)$): Let $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$.

- We define $S + T$ to be the function from V to W given by

$$(S + T)(v) = S(v) + T(v) \quad \text{for all } v \in V.$$

- We define λT to be function from V to W given by

$$(\lambda T)(v) = \lambda \cdot T(v) \quad \text{for all } v \in V.$$

In fact, the functions $S + T$ and λT are again linear maps, as we next check.

Proposition 5.7. $S + T$ and λT, defined above, are linear maps.

Proof.

□

With addition and scalar multiplication defined in this way on $\mathcal{L}(V, W)$, we assert that $\mathcal{L}(V, W)$ is itself a vector space over $\mathbb{F}$. There are various properties to check are satisfied, like associativity, distributivity, etc. We won't.

Remark 5.8. What is the additive identity of the vector space $\mathcal{L}(V, W)$?

Let $\mathbb{F}^{m,n}$ denote the set of $m \times n$ matrices over $\mathbb{F}$.

Definition 5.9. Define addition and scalar multiplication on $\mathbb{F}^{m,n}$:

With these operations, $\mathbb{F}^{m,n}$ is a $\mathbb{F}$-vector space. Now we come to the main proposition, which states, intuitively, that "the vector space structure on $\mathbb{F}^{m,n}$ agrees with the vector space structure on $\mathcal{L}(V, W)$."

Proposition 5.10. Let V and W be finite-dimensional vector spaces over $\mathbb{F}$, and fix bases $v_1, \ldots, v_n$ for V and $w_1, \ldots, w_m$ for W. Then for any $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$,

$$\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T) \quad \text{and} \quad \mathcal{M}(\lambda T) = \lambda \mathcal{M}(T).$$

Proof. To ease notation, let $A = \mathcal{M}(S)$ and $B = \mathcal{M}(T)$. By definition of $\mathcal{M}(S)$ and $\mathcal{M}(T)$, for each $i = 1, \ldots, n$, we have

$$\begin{aligned} S(v_i) &= A_{1,i} w_1 + \cdots + A_{m,i} w_m \\ T(v_i) &= B_{1,i} w_1 + \cdots + B_{m,i} w_m. \end{aligned}$$

Now, $(S+T)$ is the unique linear map from V to W such that for each $i = 1, \ldots, n$,

$$\begin{aligned}(S+T)(v_i) &= S(v_i) + T(v_i) \\ &= (A_{1,i}w_1 + \cdots + A_{m,i}w_m) + (B_{1,i}w_1 + \cdots + B_{m,i}w_m) \\ &= (A_{1,i} + B_{1,i})w_1 + \cdots + (A_{m,i} + B_{m,i})w_m\end{aligned}$$

So $\mathcal{M}(S+T) = A + B = \mathcal{M}(S) + \mathcal{M}(T)$ as desired.

The proof that $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$ proceeds similarly; depending on how you felt about the above proof that $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$, you can fill it in as an exercise. □

In other words: $\mathcal{M}(\cdot)$ is itself a linear map from $\mathcal{L}(V, W)$ to $\mathbb{F}^{m,n}$. And actually it is bijective.

Proposition 5.11. [Axl15, p.83] Let V and W be finite-dimensional vector spaces over $\mathbb{F}$ and choose bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ respectively. Then the map

$$\mathcal{M}(\cdot) \colon \mathcal{L}(V, W) \to \mathbb{F}^{m,n}$$

sending $T \mapsto \mathcal{M}(T)$ is a bijective linear map.

Proof. The previous proposition proves that it is linear. Why is it bijective? Given $A \in \mathbb{F}^{m,n}$, we must show that there exists a unique linear map $T \colon V \to W$ such that $\mathcal{M}(T) = A$.

Indeed, by definition of the matrix of a linear map, we wish to show that there exists a unique linear map $T \colon V \to W$ such that for each $i = 1, \ldots, n$,

$$T(v_i) = A_{1,i}w_1 + \cdots + A_{m,i}w_m.$$

But this is true by the theorem we proved that linear maps are freely and uniquely determined by what they do to a basis! □

Remark 5.12. We will return to the following important definition: a *invertible linear map* or *isomorphism* is a linear map that is a bijection. Thus the proposition above can be summarized by saying that we have an isomorphism between $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$.

Corollary 5.13. When V and W are finite-dimensional, $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$.

19. Composition of linear maps and product of matrices

See [Axl15, p.55].

Definition 5.14. If $T \colon U \to V$ and $S \colon V \to W$ are linear maps, then their **product** or **composition** is the linear map $ST \colon U \to W$ defined by

$$(ST)(u) = S(T(u))$$

for all $u \in U$.

Definition 5.15. A linear map $T\colon V \to V$ *from a vector space to itself* is called a *linear operator*.

If $T \in \mathcal{L}(V, V)$ is a linear operator, then we write T^2 for TT, etc. Linear operators are great because you can apply them repeatedly, and it's often interesting to consider such iterations.

The notation for composition of functions $S \circ T$ is sometimes used. The order is a perennial source of confusion here. ST means "first apply T, then apply S."

Let's check the claim implicit in the definition that the composition of linear maps is again linear: for all $a_1, a_2 \in \mathbb{F}, u_1, u_2 \in U$,

$$S(T((a_1u_1 + a_2u_2)) = S(a_1T(u_1) + a_2T(u_2)) = a_1S(T(u_1)) + a_2S(T(u_2)).$$

Proposition 5.16. Associativity, identity, and distributive properties of composition of linear maps:

Important caution: If $S, T\colon V \to V$ are linear operators on V, **it is generally not the case that $ST = TS$.** Can you think of examples? Another important caution is that $ST = 0$ generally does not imply that $S = 0$ or $T = 0$. Can you think of examples?

How about a product structure on matrices?

Definition 5.17. (See [Axl15, p.75]) Definition of matrix multiplication:

The definition is best explained in examples. For example, let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the unique linear map given by

$$T(1, 0) = (2, 1), \quad T(0, 1) = (1, 1),$$

and let $S\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the unique linear map given by

$$S(1, 0) = (1, 2, 7), \quad S(0, 1) = (3, 5, 9).$$

Example 5.18. Calculate $\mathcal{M}(S), \mathcal{M}(T)$, $\mathcal{M}(S)\mathcal{M}(T)$, and $\mathcal{M}(ST)$.

Proposition 5.19. [Axl15, p. 75] Suppose $T\colon U \to V$ and $S\colon V \to W$ are linear maps between finite-dimensional vector spaces, and fix bases

$$u_1, \dots, u_p, \qquad v_1, \dots, v_n, \qquad w_1, \dots, w_m$$

for U, V, and W respectively. Then with respect to these bases,

$$\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T).$$

Proof. □

Example 5.20. Let $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ be rotation counterclockwise by angle θ about the origin. Convince yourself that R_θ is linear; consider $(R_\theta)^2$ to deduce the double angle formulas.

20. Invertible linear maps

Definition 5.21. An *invertible linear map*, also called an *isomorphism*, is a bijective linear map.

Example 5.22. The rotation-by-θ map $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a bijective linear map, so it is invertible. Also, discuss an example in which the domain and codomain are not the same vector space.

But the following equivalent formulation of invertibility is conceptually more important.[6]

Proposition 5.23. A linear map $T \in \mathcal{L}(V, W)$ is *invertible* if and only if there exists $S \in \mathcal{L}(W, V)$ such that $ST = I_V$ and $TS = I_W$.

If so, we write $S = T^{-1}$ and say that T has an inverse, namely S. You will see shortly in the proof below that if T is invertible, then its inverse S is unique. We say that V and W are *isomorphic*, and write $V \cong W$, if there exists an isomorphism from V to W.

Example 5.24. What is the inverse of R_θ?

Proof. We'll sketch the proof of the proposition. Suppose T is a linear map.

First, we'll convince ourselves that if T fails to be injective, then there cannot possibly be an $S \in \mathcal{L}(W, V)$ such that $ST = I_V$. Do you agree?

Next, we'll convince ourselves that if T fails to be surjective, then there cannot possibly be an $S \in \mathcal{L}(W, V)$ such that $TS = I_W$. Do you agree?

These two statements together imply that if T has an inverse, then T is bijective. Now it remains to prove the converse. Say T is a bijective linear map, and let $S\colon W \to V$ be the functino that sends $w \in W$ to the unique $v \in V$ such that $T(v) = w$. Then $S \circ T = I_V$ and $T \circ S = I_W$. (Convince ourselves that this is the only possibility for an inverse of T.)

Are we done? Not yet: it remains to prove that S is a *linear* map. We wish to show that

$$S(a_1 w_1 + a_2 w_2) = a_1 S(w_1) + a_2 S(w_2)$$

for all $a_1, a_2 \in \mathbb{F}$, $w_1, w_2 \in W$. Since T is injective, we win if we can show that

$$T(\text{right hand side}) = T(\text{left hand side}).$$

Do this!

[6] In the textbook, in fact, invertibility is defined as below and then proved to be equivalent to being a bijective linear map.

□

Proposition 5.25. Let V and W be finite-dimensional vector spaces. Then $V \cong W$ iff $\dim V = \dim W$.

Proof. We have already seen, from the rank-nullity theorem, that if $T\colon V \to W$ is a bijective linear map between finite-dimensional vector spaces, then $\dim V = \dim W$. This proves $\Rightarrow$. Conversely, suppose $\dim V = \dim W$, and choose bases $v_1, \dots, v_n$ and $w_1, \dots, w_n$ for V and W respectively. Let T be the unique linear map such that $T(v_i) = w_i$ for $i = 1, \dots, n$.

We claim that T is bijective. There are many equally good ways to proceed at this point. For example, first note $\operatorname{range}(T) = W$ since $w_1, \dots, w_n$ span $\operatorname{range}(T)$, so T is surjective. To prove injectivity of T, either prove it directly or invoke the rank-nullity theorem. □

Corollary 5.26. Every finite-dimensional vector space over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$ for some $n \geq 0$.

It's worth discussing slowly *why* we don't study the vector spaces $\mathbb{F}^n$ at this point, given the corollary above.

We end with the following useful proposition:

Proposition 5.27. Suppose V and W are finite-dimensional vector spaces with $\dim V = \dim W$, and let $T \in \mathcal{L}(V, W)$. The following are equivalent ("TFAE"):[7]

(1) T is invertible.
(2) T is injective.
(3) T is surjective.

Proof. Recall that

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

So we have $\dim \operatorname{null} T = 0$ iff $\dim \operatorname{range} T = \dim V$ iff $\dim \operatorname{null} T = 0$ AND $\dim \operatorname{range} T = \dim V$. So T is injective iff T is surjective iff T is bijective. □

[7] Discuss what this means! I.e., each pair of statements is related by iff.

Part 6. Sums and direct sums

21. Sums

Let's motivate the concept of the *sum* of subspaces, before officially defining it. Consider the following subspaces of $\mathbb{R}^3$:

$$U_1 = \{(x, 0, 0) : x \in \mathbb{R}\}, \qquad U_2 = \{(0, y, 0) : y \in \mathbb{R}\}.$$

Note that $U_1 \cup U_2$ is not a subspace. What's the *smallest* subspace containing both U_1 and U_2? Discuss.

Definition 6.1. Let $U_1, \ldots, U_m$ be subspaces of a vector space V. Their *sum* is defined to be the subset of V

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Proposition 6.2. $U_1 + \cdots + U_m$ is the *smallest* subspace of V containing each of $U_1, \ldots, U_m$.

What do we mean by *smallest subspace*? It means two claims rolled into one:

(1) $U_1 + \cdots + U_m$ is, itself, a subspace of V.
(2) It is the *smallest* such in the sense that any subspace of V that contains each $U_1, \ldots, U_m$ contains $U_1 + \cdots + U_m$.

So to prove the proposition, we need to prove the two claims.

Proof. □

It's good to meditate on the following: Let U be a subspace of V. Then $U = U + U$. In fact

$$U = U + U = U + U + U = \cdots.$$

This brings us to: how to prove two sets are equal? See 11.

22. Direct sums

Sometimes sums are "wasteful" in the sense that a given element $v \in U_1 + \cdots + U_m$ can be expressed *more than one way* in the form $u_1 + \cdots + u_m$ with each $u_j \in U_j$.

Example 6.3. Discuss an example.

Definition 6.4. Let $U_1, \dots, U_m$ be subspaces of V, and let $W = U_1 + \cdots + U_m$ denote their sum. Then W is called a **direct sum** if each element $v \in W$ can be written *uniquely* as

$$v = u_1 + \cdots + u_m, \quad u_j \in U_j \text{ for each } j = 1, \dots, m.$$

We emphasize that W is a direct sum by writing

$$W = U_1 \oplus \cdots \oplus U_m.$$

Example 6.5. Discuss an example.

That definition seems clunky because it looks like to check that $U_1 + \cdots + U_m$ is a direct sum, you'd have to check that for *each* element v in the sum, it has a unique expression as $v = u_1 + \cdots + u_m, \quad u_j \in U_j$ for each $j = 1, \dots, m$. But it turns out there's that *you only have to check that condition for* $v = \mathbf{0}$, as we now show.

Proposition 6.6. Suppose $U_1, \dots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if the only way to write $\mathbf{0}$ as a sum $u_1 + \cdots + u_m$, with each $u_j \in U_j$, is

$$\mathbf{0} = \mathbf{0} + \cdots + \mathbf{0}.$$

Proof.

□

Part 7. Linear operators: eigenvalues, eigenvectors, and invariant subspaces

Here's a rough account of what we've studied so far. We studied **vector spaces** and structures associated to them (bases, subspaces). And we studied **linear maps** between vector spaces, and structures associated to them (nullspace, range). Also, we studied **matrix representations** of linear maps: i.e., how to communicate a linear map from $V \to W$, having fixed choices of bases of V and W.

Next, we begin a more in-depth study of *linear operators* on finite-dimensional vector spaces. (Remember, *linear operator* is just a special term for a linear map from a vector space V to itself.) You have seen many examples of operators that perhaps convince you of their importance in the study of linear algebra, e.g.,:

- "physical" transformations of familiar vector spaces like $\mathbb{R}^2$, such as "dilating by a factor of 2," "projecting to the line $x = y$, "rotating by angle θ";
- operators on polynomials (more generally, on infinitely-differentiable functions) such as differentiation;
- operators that encode transition probabilities between states in a Markov chain, such as in the hot potato problem on the problem set.

So, now it is time to study structures associated to operators. In particular, we'll define invariant subspaces, eigenvalues, eigenvectors, eigenspaces. We'll see some structures that all operators on finite-dimensional *complex* vector spaces (vector spaces over $\mathbb{C}$) enjoy.

23. Eigenvalues, eigenvectors, and eigenspaces

Source: [Axl15, Ch. 5]. Notation: We write $\mathcal{L}(V)$ for $\mathcal{L}(V, V)$ for short; i.e., $\mathcal{L}(V)$ denotes the set of linear operators on V. Given $T \in \mathcal{L}(V)$, we may start to write Tv for $T(v)$, just for short.

Definition 7.1. Let V be any vector space over $\mathbb{F}$ and let $T\colon V \to V$ be a linear operator. An element $\lambda \in \mathbb{F}$ is called an *eigenvalue* for T if $Tv = \lambda v$ for some nonzero vector $v \in V$.

Definition 7.2. Such a nonzero vector $v \in V$ is called an eigenvector of T corresponding to λ.. In other words, a vector $v \in V$ is an *eigenvector with eigenvalue* λ if $v \neq \mathbf{0}$ and $Tv = \lambda v$.

Example 7.3. Do some familiar running examples of operators $\mathbb{R}^2 \to \mathbb{R}^2$: horizontal stretch by a factor of 2, projection to the x-axis.

The definition of an eigenvector associated to λ is a little annoying because the set of eigenvectors associated to λ don't actually form a subspace; we specifically said that 0 is never an eigenvector. Much more natural is to study *the eigenvectors associated to* λ *together with* $\mathbf{0}$.

Definition 7.4. (See [Axl15, p. 55].) Let $T \in \mathcal{L}(V)$ and let $\lambda \in \mathbb{F}$. The *eigenspace* of T corresponding to λ, denoted $E(\lambda, T)$, is:

Rather than studying the eigenvectors associated to λ, it's much better to throw $\mathbf{0}$ in and study the eigenspace of T corresponding to λ, since eigenspaces are actually subspaces!

Proposition 7.5. ([Axl15, p. 134]) Let V be a finite-dimensional vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{F}$. The following are equivalent:

(1) λ is an eigenvalue of T.
(2) $T - \lambda I$ is not injective.
(3) $T - \lambda I$ is not surjective.
(4) $T - \lambda I$ is not bijective.

Here $I = 1_V \colon V \to V$ denotes the identity linear map on V.

Proposition 7.6. [Axl15, p. 136] Let $T \in \mathcal{L}(V)$ and let $v_1, \ldots, v_m$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ respectively. Then $v_1, \ldots, v_m$ are linearly independent.

Example 7.7. Make sure your answers to the review sheet questions were consistent with this proposition. For example, this says that you can't have three different 1-dimensional eigenspaces in $\mathbb{R}^2$, although you can in $\mathbb{R}^3$.

Proof.

□

Corollary 7.8. Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

24. Polynomials applied to operators

Let V be a vector space over $\mathbb{F}$.

Definition 7.9. Let $T \in \mathcal{L}(V)$, and let $p \in \mathcal{P}(\mathbb{F})$, say

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m.$$

We define

$$p(T) = a_0 I + a_1 T + \cdots + a_m T^m.$$

What is $p(T)$? It is itself a linear operator on V.

Polynomials have some remarkable properties, many of which are beyond the scope of the course. For this reason, it is useful to study polynomials applied to operators—so that we can borrow some of the magic of polynomials.

For example, let's state (without proof) the Fundamental Theorem of Algebra:

Theorem 7.10. Every nonconstant polynomial over $\mathbb{C}$ has a root. In fact, every polynomial over $\mathbb{C}$ can be factored into linear factors

$$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m).$$

Proof. See [Axl15, p. 124-125] □

Then, we are ready to prove an important and not at all obvious theorem:

Theorem 7.11. [Axl15, p. 145] Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

Proof. □

25. Upper triangular matrix representations

Note: I ultimately decided not to cover Theorem 7.12 in this class.

A big theme here is *finding good bases*. Given a linear operator $T \in \mathcal{L}(V)$, can we find a basis $v_1, \ldots, v_n$ of V with respect to which $\mathcal{M}(T)$ is in some convenient form? The theorem we just proved says that every linear operator T on a complex finite-dimensional vector space has a basis with respect to which the first column of $\mathcal{M}(T)$ is

$$\begin{pmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The *best* kinds of matrices are diagonal matrices. Discuss the definition of diagonal matrix, and why they are great when it comes to computations. For example, it's hard to compute A^n for some big n in general, but easy to do so if A is diagonal?

Unfortunately, it is not true that every operator T on a finite-dimensional complex vector space V can be represented by a diagonal matrix with respect to some basis of V. But it's worth stopping to note that T can always be given a *upper triangular* matrix:

Theorem 7.12. Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.

Proof. Either do the proof as in [Axl15, p. 149], or omit it. After all, the conceptually most enlightening proof would invoke linear operators on quotient spaces, which is done in the book as "Proof 2" on p. 150. So this is a proof that is best done after learning about quotient spaces, but that is a conceptual jump that is best done in a second course in algebra or linear algebra. (But see Section 3.E if you want.) □

26. Diagonalizability

It's also worth studying diagonalizability, in particular knowing equivalent conditions under which T is diagonalizable.

Theorem 7.13. [Axl15, p. 157] Given a finite-dimensional vector space V and $T \in \mathcal{L}(V)$, the following are equivalent:[8]

(1) T is diagonalizable,
(2) V has a basis consisting of eigenvectors of T;

[8]Axler's version also throws in a couple other conditions that are equivalent to diagonalizability.

(3) $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.

Example 7.14. Before proving the theorem, let's illustrate it in the example of $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by $(x, y, z) \mapsto (x, 2y, 3z)$.

Proof. □

Optional, if time: the existence of Jordan canonical Forms for complex operators, without proof.

Definition 7.15. A matrix is in *Jordan canonical Form* if:

Theorem 7.16. Every operator on a complex finite-dimensional vector space V can be written in Jordan canonical form, with respect to some basis of V.

Final remark: how do you actually compute eigenvalues of operators? I.e., given a finite-dimensional vector space V and $T \in \mathcal{L}(V)$, how do you find its eigenvalues?

For example, how would you calculate the eigenvalues of $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x, y) = (x + y, x + y)$? A number $\lambda \in \mathbb{R}$ is an eigenvalue of T iff $T - \lambda I$ is not invertible. Now, with respect to the standard basis of $\mathbb{R}^2$, the matrix $\mathcal{M}(T - \lambda I)$ is

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

Say a matrix $A \in \mathbb{F}^{n,n}$ is *invertible* iff there is a matrix $B \in \mathbb{F}^{n,n}$ such that $AB = BA = I$ where I denotes the $n \times n$*identity matrix*.

Remark 7.17. Notice that $\mathcal{M}(I_V) = I$, *no matter what basis of* V *we choose*. Do you agree?

Then

Proposition 7.18. Let V be a finite-dimensional vector space, $S \in \mathcal{L}(V)$ a linear operator. Fix a basis $v_1, \dots, v_n$ of V. Then S is invertible iff $\mathcal{M}(S)$ is invertible.

Fill in the proof yourself; it follows directly from the fact that matrix product corresponds to product of linear operators.

So the question becomes, *for which real numbers λ is the matrix above invertible?* It is time to learn determinants, which help us answer this question, and more!

Part 8. Determinants

Following Treil's LADW book.

27. Desiderata for the determinant

Motivation for deterimants: suppose you have a linear operator $T\colon \mathbb{R}^n \to \mathbb{R}^n$. Then T transforms the unit n-dimensional cube (easiest to picture for $n = 1, 2, 3$ to a parallelopiped in $\mathbb{R}^n$ of some volume. The determinant is a measure of that volume. More precisely it is a *signed*, or *oriented*, measure of volume, in a way that we shall see.

The parallelopiped in question is the one associated to the vectors $Te_1, \dots, Te_n$. Recall that these vectors are written as the columns of the matrix $\mathcal{M}(T)$ of T (with respect to the standard basis $e_1, \dots, e_n$ of $\mathbb{R}^n$. So what we seek is *a way to associate a number to an $n \times n$ matrix,*or equally well n vectors in $\mathbb{R}^n$, in a way that captures our intuition for oriented/signed volumes.

Wish list. Let $\mathbb{F}$ be a field. We wish for a function $D\colon \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ that satisfies:

(1) D is *multilinear*, i.e., linear in each copy of $\mathbb{F}^n$ separately. That is, for each $k = 1, \dots, n$,

$$D(v_1, \dots, av_k + a'v_k', \dots, v_n) = aD(v_1, \dots, v_k, \dots, v_n) + a'D(v_1, \dots, v_k', \dots, v_n)$$

(2) D is *alternating*. That is

$$D(v_1, \dots, v_j, \dots, v_k, \dots, v_n) = 0 \qquad \text{if } v_j = v_k.$$

(3) D is *normalized*:

$$D(e_1, \dots, e_n) = 1.$$

Remark 8.1. Explain *why* we wish for these properties!

Remark 8.2. Explain *why* alternating is called alternating. Namely, show that *if* D satisfies properties 1, 2, 3 above, then

$$D(v_1, \dots, v_j, \dots, v_k, \dots, v_n) = -D(v_1, \dots, v_k, \dots, v_j, \dots, v_n).$$

Theorem 8.3. (Theorem-Definition) There exists a unique multilinear, alternating, normalized function $D\colon \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$. We'll call it the *determinant.*

We are going to postpone the proof of this theorem. Instead, we are going to derive some basic properties of it, assuming it does exist.

28. First properties of the determinant, assuming it exists

Proposition 8.4. Let A be a square matrix. If the column vectors $v_1, \ldots, v_n \in \mathbb{F}^n$ are linearly dependent then $\det A = 0$.

Proof. □

Remark 8.5. The converse will be proved later: if $\det A = 0$ then the column vectors $v_1, \ldots, v_n \in \mathbb{F}^n$ are linearly dependent.

Proposition 8.6.

(1) The determinant of a diagonal matrix is the product of its diagonal entries.
(2) The determinant of an upper triangular (resp. lower triangular) matrix is the product of its diagonal entries.

Proof. □

29. Construction of the determinant

Let's warm up with 2×2 matrices. Suppose there *were* a determinant function, i.e., a function

$$D\colon \mathbb{F}^2 \times \mathbb{F}^2 \to \mathbb{F}$$

that is *multilinear, alternating, and normalized.* Write $e_1 = (1, 0)$ and $e_2 = (0, 1)$ for the standard basis of $\mathbb{F}^2$. Then for all $a, b, c, d \in \mathbb{F}$,

$$\det \begin{vmatrix} a & b \\ c & d \end{vmatrix} = D(ae_1 + ce_2, be_1 + de_2)$$

$= D(ae_1, be_1)+D(ae_1, de_2)+D(ce_2, be_1)+D(ce_2, de_2) = adD(e_1, e_2)+bcD(e_2, e_1) = ad-bc.$
Now we check that the function defined by $D((a, b), (c, d)) = ad-bc$ *actually is* multilinear, alternating, and normalized. That shows that the 2×2 determinant exists and is unique!

Remark 8.7. Check that, and do the analogous calculation for determinants of 3×3 matrices.

Definition 8.8. A *permutation* is a bijection $\sigma\colon \{1, \dots, n\} \to \{1, \dots, n\}$.

Remark 8.9. There are various ways to write down a permutation, e.g., in one-line or two-line notation. Discuss these. (In real life it is often more useful to use *cycle notation*, which we will not cover in this class.)

Definition 8.10. The *product*, of two permutations is their composition: $\sigma\tau = \sigma \circ \tau$. Write S_n for the set of permutations $\{1, \dots, n\} \to \{1, \dots, n\}$. How many elements of S_n are there?

Definition 8.11. A *transposition* is:

Remark 8.12. Every permutation is a product of transpositions.

Do you believe that? Do a "physical proof."

Definition 8.13. (Important preliminary definition) The *sign* of a permutation σ, denoted $\operatorname{sgn}(\sigma)$, is:

Proposition 8.14. Definition of determinant. If there exists an $n \times n$ determinant—that is, if there exists a multilinear, alternating, and normalized function

$$D\colon \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F},$$

then it *must* be:

Proof. Imitate the proof of the 2×2 and 3×3 cases. □

Proposition 8.15. The function

$$D \colon \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$$

defined above really is multilinear, alternating, and normalized. Therefore, the determinant D exists and is unique.

We're done, other than justifying our preliminary definition of the *sign* of a transposition. We have to prove that the sign is a well-defined function. In other words, we have to prove:

Proposition 8.16. Suppose σ is expressible as a product of k transpositions and is also a product of ℓ transpositions. Then the numbers k and ℓ have the same *parity* (both even or both odd.)

Proof. We'd win if we can show that σ "wears" its sign, in other words, we win if we can show that just by looking at σ, you can tell the parity of the number of transpositions τ_i in any sequence of transpositions whose product is σ.

Definition 8.17. A *descent* in a permutation is:

Proposition 8.18. If σ is a product of an odd number of transpositions, then the number of descents in σ is odd. If σ is a product of an even number of transpositions, then the number of descents in σ is even.

Proof. If $\sigma = \tau_1 \cdots \tau_k$ where τ_i are transpositions, then the number of descents in σ is $a_1 + \cdots + a_k$, for odd numbers a_i. This sum is odd if k is odd and even if k is even. □

Therefore the sign of σ is well-defined—and so is the determinant!

□

30. The transpose of a matrix

Now that we've constructed it, let's derive some more important properties of the determinant.

Definition 8.19. Let $A \in \mathbb{F}^{m,n}$ be a matrix. The *transpose* of A, written A^T, is the matrix in $\mathbb{F}^{n,m}$ whose entries are $(A^T)_{i,j} = A_{j,i}$.

So the transpose of a square matrix is another square matrix of the same size. The deeper conceptual significance of transposes is somewhat beyond the scope of this course. Briefly, it is as follows: if A is the matrix associated to a linear map $T\colon V \to W$, then A^T is the matrix associated to the *dual* of T. What is the dual of T? You can read this in [Axl15, Ch. 3].

Proposition 8.20. Let $A \in \mathbb{F}^{n,n}$ be any square matrix. Then $\det A^T = \det A$.

Proof.

□

This tells us that the *entire story of determinants* can equally well be told with rows replacing columns everywhere. The determinant is the unique alternating, multilinear, normalized function on the n *rows* of an $n \times n$ matrix. If there is a repeated *row* then the determinant is 0. Etc.

31. Multiplicativity of the determinant

Here is an important proposition that we will use freely in the class. (I originally thought we could prove in a slick way, but I had one missing link, unfortunately. So it's one of the few statements that we will accept henceforth without proving. Sorry about that!)

Proposition 8.21. Let A and B be $n \times n$ matrices. Then

$$\boxed{\det(AB) = \det(A)\det(B).}$$

Proof. Several proofs are possible; the one presented in Treil's book involves decompositng arbitrary matrices into products of elementary matrices, which we don't have time to cover.. You can see the textbook: Theorem 3.5 on page 82 in Treil's book. □

Sanity check: this multiplicativity should agree with our notion of the determinant being an oriented volume scaling factor.

32. Cofactors

Definition 8.22. Given an $n \times n$ matrix A, let $A_{j,k}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row j and column k.

Definition 8.23. The numbers $C_{j,k} = (-1)^{j+k} \det A_{j,k}$ are called *cofactors* of A.

Discuss the checkerboard sign pattern of $(-1)^{j+k}$.

Theorem 8.24. (Cofactor expansion of the determinant) For each $1 \leq j \leq n$,

$$\det A = a_{j,1}C_{j,1} + \cdots + a_{j,n}C_{j,n}.$$

An example will help clarify the theorem statement:

Example 8.25.

How would we prove the theorem? We will show that the right hand side is an alternating, multilinear, and normalized function $\mathbb{F}^{n,n} \to \mathbb{F}$!

Proof. □

33. Cofactors and inverses of square matrices

How do you compute the inverse of a square matrix? First let's say what we mean by inverse: if A is an $n \times n$ matrix, then an inverse of A is a matrix B such that $AB = BA = I$. Here I denotes the $n \times n$ identity matrix (the diagonal matrix with all 1s on the diagonal.)

Proposition 8.26. Let A be a square matrix. Let C be its matrix of cofactors: $C_{j,k} = (-1)^{j+k} \det A_{j,k}$. Then, $AC^T = (\det A)I$.

Do a 2×2 example.

Proof. The proof consists of calculations of the entries of AC^T. Namely, we wish to show that the diagonal entries of AC^T are $\det A$ and the off-diagonal entries are 0.

We have

$$(AC^T)_{j,j} = a_{j,1}C_{j,1} + \cdots + a_{j,n}C_{j,n}$$

by definition of matrix multiplication. But the right hand side is the cofactor expansion of $\det A$ along the j^{th} row, so it is $\det A$.

Next, for $j \neq k$, we have

$$(AC^T)_{k,j} = a_{k,1}C_{j,1} + \cdots + a_{k,n}C_{j,n}$$

Let A' be the matrix obtained from A by replacing row j with row k. Then the right hand side is the cofactor expansion of $\det A'$. (Think about this). But $\det A' = 0$. Done. □

Remark 8.27. A similar calculation will show $C^T A = I$. Actually, for arbitrary $n \times n$ matrices A and B, it is a general fact that $AB = I$ iff $BA = I$.

The proposition above gives a formula for the inverse of a matrix!

Corollary 8.28. A is invertible iff $\det A \neq 0$. If A is invertible, then $A^{-1} = \frac{1}{\det A} C^T$.

Proof. First of all, if A is invertible then $1 = \det(AA^{-1}) = \det(A)\det(A^{-1})$ implies $\det A \neq 0$. Conversely, if $\det A \neq 0$ then $A(\frac{1}{\det A} C^T) = I$. □

We now have a lot of equivalent ways to characterize invertibility.

Another way of saying the same thing: the columns of A are *dependent* iff $\det A = 0$.

Proof. We already proved that if the columns of A are dependent, then $\det A = 0$. Conversely, suppose $\det A = 0$; we wish to show that the columns of A are dependent. Indeed, if $\det A = 0$, we showed $AC^T = 0$ where C is the cofactor matrix. That implies that A isn't invertible, which (by the line of reasoning above) implies that the columns of A are dependent. □

Summarizing,

Proposition 8.29. The following are equivalent for an $n \times n$ matrix A.

(1) A is invertible,
(2) $\det A \neq 0$.
(3) The columns of A are linearly independent,
(4) The rows of A are linearly independent,
(5) T is invertible, where $T\colon \mathbb{F}^n \to \mathbb{F}^n$ is the linear operator whose matrix (with respect to the standard basis, say) is A.

Proof. (Sketch) We just showed (1)⇔(2), above.

Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be the unique linear map whose matrix, with respect to the standard basis, is A. Then A is invertible if and only if T is invertible. (Think about this.)

Next, T is invertible if and only if $Te_1, \dots Te_n$ are linearly independent. In other words, T is invertible if and only if the columns are A are linearly independent. (Think about this.)

If so, these arguments together show that (1), (3), and (5) are equivalent. Now applying the same arguments to A^T, and noting that $\det A = \det A^T$, yields that (1), (4), and (5) are equivalent. □

34. Postscript: using determinants to compute eigenvalues

Let's return to the following situation. Say you have a linear operator $T\colon V \to V$. You want to calculate the eigenvalues and eigenvectors (if any) of T. How?

Let A be the matrix of T with respect to any basis of V. Then a scalar $\lambda \in \mathbb{F}$ is an eigenvalue of T iff $T - \lambda I$ is not invertible. What is the matrix of $T - \lambda I$? It is $A - \lambda I$. In other words:

Proposition 8.30. $\lambda \in \mathbb{F}$ is an eigenvalue of T iff $\det(A - \lambda I) = 0$.

This is how you solve for eigenvalues: you write out $\det(A - \lambda I) = 0$, which is a polynomial equation in λ. Then you solve for values of λ. Then, once an eigenvalue λ is obtained, you can solve a system of linear equations to obtain all eigenvectors associated to λ.

Example 8.31. Do the example of reflection across the line $y = x$. What are the eigenvalues and eigenvectors of T?

Definition 8.32. Let $T\colon V \to V$ be any linear operator on a finite-dimensional vector space V. Let A be any matrix representation of T (i.e., with respect to any basis of V). The polynomial

$$p(\lambda) = \det(A - \lambda I)$$

is called the *characteristic polynomial* of T.

It looks at first as if $p(\lambda)$ depends on the choice of A, i.e., the choice of basis for V. But it is a fact, not too hard to prove, that $p(\lambda)$ is actually independent of choice of A.

Then, it is a remarkable fact that T *satisfies its own characteristic polynomial.* This is the *Cayley-Hamilton theorem*, but is beyond the scope of this course.

Part 9. Inner product spaces

35. Motivation; complex numbers review

So far the main objects of study have been vector spaces, and linear maps between them. What we are missing so far is a notion of *length* and *angle* of vectors in a vector space. (These shall be seen to be closely related). This is structure that we are used to seeing in familiar vector spaces like $\mathbb{R}^n$.

It will be possible to define the structure we want, an *inner product* structure, on vector spaces over both $\mathbb{R}$ and $\mathbb{C}$. For this reason, we first briefly recall some basic properties of complex numbers. In this entire section, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Definition 9.1. (Reference: [Axl15, p. 119]) Let $z = a + bi$ be a complex number; here $a, b \in \mathbb{R}$. Then the *complex conjugate* and *absolute value* of z are defined as

$$\overline{z} = a - bi, \qquad |z| = \sqrt{a^2 + b^2} = \sqrt{z\overline{z}}.$$

Note that complex conjugation respects addition and multiplication. Absolute value respects multiplication, but not addition; rather we have a triangle inequality

$$|w + z| \leq |w| + |z|$$

for $w, z \in \mathbb{C}$.

36. Definition of an inner products

Before giving the general definition of an inner product, let's start with an important special case: the Euclidean inner product.

Definition 9.2. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. The Euclidean inner product on $\mathbb{F}^n$ is the function $\mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ defined by:

$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

Definition 9.3. The Euclidean norm of $v \in \mathbb{F}^n$ is

$$||v|| = \sqrt{\langle v, v \rangle}.$$

Discuss this example slowly; does it accord with our usual notion of norm or length?

Definition 9.4. (See [Axl15, p. 166] Let V be a vector space over $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. An inner product on V is a function $\langle \cdot, \cdot \rangle \colon \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$, taking ordered pairs (u, v) of elements in V to a number $\langle u, v \rangle \in \mathbb{F}$, satisfying:

(1) conjugate symmetry:

(Note that when $\mathbb{F} = \mathbb{R}$, "conjugate symmetry" is just "symmetry.") Note for $\mathbb{F} = \mathbb{C}$, this implies $\langle v, v \rangle \in \mathbb{R}$ for all $v \in V$.)

(2) positive definiteness:

(3) linearity in the first slot:

Definition 9.5. An *inner product space* henceforth means a vector space V over $\mathbb{R}$ or $\mathbb{C}$ together with an inner product on V. In particular, if we talk about an inner product space, we *always* mean a vector space over $\mathbb{R}$ or $\mathbb{C}$.

Remark 9.6. Note right away that

- If $\mathbb{F} = \mathbb{R}$, then the inner product is also linear in the second slot.
- If $\mathbb{F} = \mathbb{C}$, then *the inner product is not in general linear in the second slot.* In fact, it is additive in the second slot, but we have

$$\langle u, \lambda v\rangle = \overline{\langle \lambda v, u\rangle} = \overline{\lambda} \cdot \overline{\langle v, u\rangle} = \overline{\lambda}\langle v, u\rangle.$$

Example 9.7.

(1) The Euclidean inner product deserves its name: it is an inner product. Check this.
(2) Let V be the real vector space of continuous functions $[-1, 1] \to \mathbb{R}$. Then

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)dx$$

defines an inner product on V.

37. Norms

The structure of an inner product allows us to recapture a notion of *length*, or *norm*:

Definition 9.8. The *norm* of a vector v in an inner product space V is defined to be

$$||v|| = \sqrt{\langle v, v\rangle}.$$

Check some basic properties:

Proposition 9.9. Let v be a vector in an inner product space V over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

(1) $||v| = 0$ iff $v = \mathbf{0}$.
(2) For any $\lambda \in \mathbb{F}$,

$$||\lambda v|| = |\lambda|||v||.$$

Proof. (1) follows from the fact that $\langle v, v\rangle = 0$ iff $v = 0$.

For (2), we calculate

$$||\lambda v|| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\lambda\overline{\lambda}\langle v, v\rangle} = \sqrt{\lambda\overline{\lambda}}\sqrt{\langle v, v\rangle} = |\lambda|||v||.$$

□

38. Orthogonality, orthogonal decomposition.

Let V be an inner product space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Two vectors $u, v \in V$ are called *orthogonal* if $\langle u, v\rangle = 0$.

So "orthogonal" means perpendicular. Note that $\langle u, v\rangle = 0$ iff $\langle v, u\rangle = 0$ by conjugate symmetry. Draw some pictures.

Proposition 9.10. (Pythagorean theorem in an inner product space)

$$||u + v||^2 = ||u||^2 + ||v||^2$$

for any orthogonal u, v in an inner product space V.

Proof.

$$||u + v||^2 = \langle u + v, u + v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle = ||u||^2 + ||v||^2,$$

using that $\langle u, v\rangle = \langle v, u\rangle = 0$ by orthogonality of u and v. □

Orthogonality makes up about 5% of my mathematical intuition! See [Axl15, p. 174] for a pretty great use of the word "orthogonal" in oral arguments at the U.S. Supreme Court. It is a great way to be subtly dismissive of something (not that I encourage you to be dismissive of anyone in any way...) Let's say V is an inner product space, v is a (nonzero) vector you single-mindedly care about—for any vector $u \in V$, you are single-mindedly interested in isolating the "part of u in the direction of v," as distinct from the "part of u orthogonal to v." (Draw a picture—the geometric intuition is the most important thing here!) In other words, our goal is the following.

Proposition 9.11. Let V be an inner product space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and fix a nonzero vector $v \in V$. Given a any vector $u \in V$, there exists a unique choice of number $c \in \mathbb{F}$ and vector $w \in V$ orthogonal to v such that $u = cv + w$.

Proof. Proof of uniqueness: well, if such a choice of a number c and a vector w existed, then $w = u - cv$. So

$$0 = \langle u - cv, v\rangle = \langle u, v\rangle - c\langle v, v\rangle$$

which implies c *must be* $\frac{\langle u,v\rangle}{\langle v,v\rangle}$. Then w *must be* $u - cv$ for that particular value of c.

Existence: now check that this value of c and w really do satisfy that $cv + w = u$ and $\langle w, v\rangle = 0$; this amounts to redoing the calculation in the line above. □

39. Cauchy-Schwarz

Equipped with the previous result, we can prove probably a surprisingly useful inequality that holds for a general inner product space. See [Axl15, p. 172]

Proposition 9.12. (Cauchy-Schwarz Inequality) Let V be an inner product space, and let $u, v \in V$. Then

$$|\langle u, v\rangle| \leq ||u||||v||.$$

Moreover this inequality is an equality iff one of u, v is a scalar multiple of the other.

Note the inequality is an inequality of *real numbers*. Make sure you agree.

Proof. If $v = \mathbf{0}$ then both sides are 0 and we win. Otherwise, consider the orthogonal decomposition of u with respect to v:

$$u = cv + w, \qquad \text{where } c = \frac{\langle u, v\rangle}{\langle v, v\rangle}.$$

and $\langle w, v\rangle = 0$. The Pythagorean theorem then implies

$$\begin{aligned} \|u\|^2 &= \|cv\|^2 + \|w\|^2 \\ &= \left|\frac{\langle u, v\rangle}{\langle v, v\rangle}\right|^2 \cdot \|v\|^2 + \|w\|^2 \\ &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \\ &\leq \frac{|\langle u, v\rangle|^2}{\|v\|^2}. \end{aligned}$$

Now clear denominators and take the square root of both sides. Moreover, equality holds iff $w = \mathbf{0}$, i.e., $u = cv$ is a scalar multiple of v. □

The Cauchy-Schwarz inequality is a favorite of college-level math puzzle writers, because it often hides in problems that, at first glance, have nothing to do with inner products. For example, prove:

Proposition 9.13. For any real numbers $x_1, \ldots, x_n, y_1, \ldots, y_n$,

$$(x_1y_1 + \cdots + x_ny_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

Proof. It is certainly possible to prove this directly, without knowing Cauchy-Schwarz. But better to simply note that this is the Cauchy-Schwartz inequality applied to the vectors $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ with its Euclidean inner product. □

Finally, one deduces a *triangle inequality* from Cauchy-Schwarz:

Proposition 9.14. (Triangle Inequality) Let u, v be vectors in an inner product space V. Then

$$\|u + v\| \leq \|u\| + \|v\|.$$

Draw the picture! For the proof, see [Axl15, p. 173].

40. Orthonormal bases

Definition 9.15. A list of vectors $v_1, \ldots, v_n$ in an inner product space V is called *orthonormal* if:

Orthonormal vectors are useful for so many reasons. For instance:

Proposition 9.16. Let $e_1, \ldots, e_m$ be an orthonormal list of vectors in V. Then for any numbers $a_1, \ldots, a_m$,

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2.$$

Proof.

□

We deduce that orthonormal lists are linearly independent. (Argue this). Orthogonal lists are thus also linearly independent.

Orthonormal *bases* are so useful that we want a procedure for making them. Fortunately, such a procedure exists: it is called the Gram-Schmidt Algorithm. First, we need a helpful lemma:

Lemma 9.17. Say $e_1, \ldots, e_n$ is an orthonormal basis of V, and $v \in V$ is any vector. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n.$$

Proposition 9.18. [Axl15, p. 183], Gram-Schmidt Algorithm. Suppose $v_1, \ldots, v_m$ is a linearly independent list. Let $e_1 = v_1/\|v_1\|$, and for $j = 2, \ldots, m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\|v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}\|}.$$

Then $e_1, \ldots, e_m$ is an orthonormal list of vectors and moreover

$$\operatorname{span}(e_1, \ldots e_j) = \operatorname{span}(v_1, \ldots, v_j).$$

In particular, if $v_1, \ldots, v_m$ was a basis, then $e_1, \ldots, e_m$ is an orthonormal basis.

Thus every finite-dimensional inner product space has an orthonormal basis! Indeed, pick any basis, and then use Gram-Schmidt to transform it into an orthonormal basis.

41. Orthogonal complements

Let U be a subset of an inner product space V. Usually, we will be interested in the case that U is a subspace of V.

Definition 9.19. The *orthogonal complement* of U is

$$U^\perp = \{v \in V : \langle v, u\rangle = 0 \text{ for every } u \in U.\}$$

In other words, the orthogonal complement is the set of vectors that are orthogonal to *everyone* in U.

See [Axl15, p. 193] for basis properties of the orthogonal complement. The most important one is:

Proposition 9.20. $U^\perp$ is a subspace of V.

Proof.

□

Why are orthogonal complements important? Because they allow you to *project away the stuff you don't care about.* Namely:

Proposition 9.21. Let U be a finite-dimensional subspace of an inner product space. Then

$$V = U \oplus U^\perp.$$

Proof. (See [Axl15, p. 194])

□

Now in complete generality, if $V = U \oplus W$, then there is a natural *projection* map to U along W: it is the linear map $P_{U,W} \colon V \to V$ sending

$$P_{U,W}(u + w) = u \quad \text{for all } u \in U, w \in W.$$

Check that this really is linear, and check that $T^2 = T$, so you really have a projection operator, whose range is U. In the case that V is an inner product space and $W = U^\perp$, then this projection operator gets a special name: orthogonal projection.

Definition 9.22. Let U be a finite-dimensional subspace of an inner product space V. The orthogonal projection of V onto U is the operator

$$P_U \colon V \to V$$

defined as

$$P_U(u + w) = u$$

whenever $u \in U$, $w \in U^\perp$.

Page 196 of [Axl15] lists many properties of orthogonal projections. Many of these you know from your Problem Set problem on projections. Here are two properties that are special to *orthogonal* projections in an inner product space:

Proposition 9.23.
(1) $\|P_U(v)\| \leq \|v\|$ for $v \in V$;
(2) $P_U(v) = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$ for any orthonormal basis $e_1, \ldots, e_m \in U$.

Proof.

□

42. Riesz representation theorem

See [Axl15, p. 188]. First of all, there's a special name for a linear map of vector spaces over $\mathbb{F}$ in which the codomain is $\mathbb{F}$:

Definition 9.24. Let V be a vector space over $\mathbb{F}$. A linear map $V \to \mathbb{F}$ is called a *linear functional.*

Now suppose V is an inner product space, and $u \in V$ Then "inner product with u" is a function $\langle \cdot, v \rangle \colon V \to \mathbb{F}$. Namely, it sends $v \mapsto \langle v, u \rangle$. In fact, $\langle \cdot, u \rangle$ is a linear functional: this amounts to the fact that the inner product is linear in the first coordinate.

Theorem 9.25. (Riesz representation theorem) Let V be a finite-dimensional inner product space, and let $\phi \colon V \to \mathbb{F}$ be a linear functional on V. Then there exists a unique vector $u \in V$ such that

$$\phi(v) = \langle v, u \rangle$$

for every $v \in V$.

In other words: there is a unique $u \in V$ such that $\phi = \langle \cdot, u \rangle$.

Example 9.26. Consider $\mathbb{R}^2$ with its Euclidean inner product. Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be the linear functional given by

$$\phi(x, y) = x + y.$$

Then $\phi = \langle \cdot, u \rangle$ for a (unique) vector $u \in \mathbb{R}^2$. Which one?

Proof. (Proof of Riesz representation) Let $e_1, \ldots, e_n$ be an orthonormal basis for V. Given a linear functional ϕ on V, let

$$u = \overline{\phi(e_1)} e_1 + \cdots + \overline{\phi(e_n)} e_n.$$

We claim that $\phi(v) = \langle v, u \rangle$ for all $v \in V$. Indeed,

$$\begin{aligned} \phi(v) &= \phi(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n) \\ &= \langle v, e_1 \rangle \phi(e_1) + \cdots + \langle v, e_n \rangle \phi(e_n) = \langle v, u \rangle. \end{aligned}$$

Now prove uniqueness: say $u_1, u_2 \in V$ are such that $\langle v, u_1 \rangle = \langle v, u_2 \rangle$ for all $v \in V$. Then $\langle v, u_1 - u_2 \rangle = 0$ for all $v \in V$. Then in particular $\langle u_1 - u_2, u_1 - u_2 \rangle = 0$, so $u_1 - u_2 = \mathbf{0}$. □

Rephrasing the last part slightly: we found a vector, $u_1 - u_2$, which is orthogonal to all vectors. Such a vector must be $\mathbf{0}$, because only $\mathbf{0}$ is orthogonal to itself.

Notice that the uniqueness statement in Riesz Representation implies in particular that the choice of u is actually independent of choice of orthonormal basis $e_1, \ldots, e_n$.

Let's take the broad view. The Riesz representation theorem gives us a function

$$R \colon \mathcal{L}(V, \mathbb{F}) \to V.$$

(I'm temporarily calling it R for "Riesz"; this is not standard notation.) Now R is a linear map if $\mathbb{F} = \mathbb{R}$. If $\mathbb{F} = \mathbb{C}$, it not linear, but rather *conjugate linear.* Do you see why?

Part 10. Operators on inner product spaces

43. The adjoint of a linear operator

Reference: Chapter 7 of [Axl15].

Throughout, let V and W be **finite-dimensional** inner product spaces over the same field $\mathbb{F}$.

Definition 10.1. Let $T\colon V \to W$ be a linear map. The adjoint of T is the function $T^*\colon W \to V$ such that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for every $v \in V, w \in W$.

Note that the inner product on the left is the one on W; the one on the right is the one on V.

What is going on with this definition, First, why would such a function exist? Well, fix $w \in W$, and consider the linear map $V \to \mathbb{F}$ sending $v \mapsto \langle Tv, w\rangle$. This is a linear map: it is the composition of T and $\langle \cdot, w\rangle$.

Then Riesz Representation implies that there is a unique vector in V—let's henceforth call it T^*w—that represents the linear functional by taking inner product with it. In other words, there is a unique vector T^*w such that $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for every $v \in V, w \in W$.

Pretty soon we will see that $T^*\colon W \to V$ is a *linear* map. For now it is just a function.

The book's example (page 204) is very instructive.

Example 10.2. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

Find a formula for T^*.

Solution: For each $(y_1, y_2) \in \mathbb{R}^2$, we wish to calculate the vector $T^*(y_1, y_2)$. Whatever this vector is, it satisfies, for all $(x_1, x_2, x_3) \in \mathbb{R}^3$,

$$\langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle = \langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle$$

$$= x_2y_1 + 3x_3y_1 + 2x_1y_2 = \langle ((x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle.$$

So it must be that $T^*(y_1, y_2) = (2y_2, y_1, 3y_1)$.

Let's notice two things in the example. The first thing is that in this example, the function T^* is actually linear. The second thing is about matrices: with respect to the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$—which, by the way, are orthonormal bases—the matrices $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$ are transposes. Neither of these observations is a coincidence:

Proposition 10.3. Let $T\colon V \to W$ be a linear map of inner product spaces. The adjoint $T^*\colon W \to V$ is a linear map.

Proof. Let $\lambda_1, \lambda_2 \in \mathbb{F}$, $w_1, w_2 \in W$. We claim that for every $v \in V$,

$$\langle v, T^*(\lambda_1 w_1 + \lambda_2 w_2)\rangle = \langle v, \lambda_1 T^* w_1 + \lambda_2 T^* w_2\rangle.$$

If so, then $T^*(\lambda_1 w_1 + \lambda_2 w_2) = \lambda_1 T^* w_1 + \lambda_2 T^* w_2$. (Why?) Now prove the claim.

□

Remark 10.4. Here's a high-brow explanation of what the adjoint really is, including for free the fact that it really is linear, in case you are interested. (You may find this enlightening, or you may find this maddening. If the latter, just forget about it.) Remember that the Riesz representation theorem gives conjugate-linear bijections $W \to \mathcal{L}(W, \mathbb{F})$ and $V \to \mathcal{L}(V, \mathbb{F})$. (These are bijections which are conjugate-linear; their inverses are also conjugate-linear). Now given a linear map $T\colon V \to W$, I claim (or define) the adjoint $T^*\colon W \to V$ to be the composition

$$W \xrightarrow{\cong} \mathcal{L}(W, \mathbb{F}) \to \mathcal{L}(V, \mathbb{F}) \xrightarrow{\cong} V.$$

What is the middle map $W^* \to V^*$? It is the *dual* of T, namely, composing by T on the left. This composition maps $w \in W$ to

$$w \mapsto \langle \cdot, w\rangle \mapsto \langle T\cdot, w\rangle \mapsto \langle \cdot, T^* w\rangle.$$

In other words, the vector T^*w is the unique vector (whose existence, and uniqueness, is guaranteed by Riesz) such that we have an equality of linear functionals $\langle T\cdot, w\rangle = \langle \cdot, T^* w\rangle$.

Proposition 10.5. [Axl15, p. 208] Let $T\colon \mathcal{L}(V, W)$, let $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ be any orthonormal bases of V and W respectively. With respect to these orthonormal bases,

$$\boxed{\mathcal{M}(T^*) \text{ is the conjugate transpose of } \mathcal{M}(T).}$$

Proof. Let's calculate $\mathcal{M}(T)_{j,k}$ and $\mathcal{M}(T^*)_{k,j}$ and show that these two numbers are conjugates. If so, we are done.

Writing

$$Te_k = \langle Te_k, f_1\rangle f_1 + \cdots + \langle Te_k, f_m\rangle f_m$$

shows that $\mathcal{M}(T)_{j,k} = \langle Te_k, f_j\rangle = \langle e_k, T^* f_j\rangle$. Similarly, writing

$$T^* f_j = \langle T^* f_j, e_1\rangle e_1 + \cdots + \langle T^* f_j, e_n\rangle e_n$$

shows that $\mathcal{M}(T)_{k,j} = \langle T^* f_j, e_k\rangle = \overline{\langle e_k, T^* f_j\rangle}$. We are done. □

44. Self-adjoint operators

Definition 10.6. Let $T\colon V \to V$ be a linear operator. Then say that T is *self-adjoint* if, well, it is the adjoint of itself. That is, T is called self-adjoint if $T = T^*$.

How can you tell if an operator T is self-adjoint? Well, write down its matrix $\mathcal{M}(T)$ with respect to any orthonormal basis. Then T is self-adjoint iff $\mathcal{M}(T)$ is equal to its conjugate transpose. That is, T is self-adjoint if

$$\mathcal{M}(T)^t = \overline{\mathcal{M}(T)}.$$

Definition 10.7. Given an $n \times n$ matrix A, define its adjoint $A^* = \overline{A}^t$. So the matrix of the adjoint is the adjoint of the matrix (with respect to an orthonormal basis.)

It's a bit hard to build intuition from scratch. My own intuition works best in the real case $\mathbb{F} = \mathbb{R}$. In this case, a real operator T is self-adjoint if it has a *symmetric* matrix with respect to an orthonormal basis. In the complex case, you can get some intuition from the case that V is 1-dimensional. In this case, the space of operators $\mathcal{L}(V)$ may be identified with 1×1 complex matrices—in other words, the complex numbers $\mathbb{C}$. Then an operator is self-adjoint iff its matrix is *real*. In general, a rough intuition for self-adjoint operators, among all operators, is that they are an analogue of real numbers among all complex numbers.

45. The Spectral theorem

One of the cornerstones of linear algebra is the Spectral theorem.

Theorem 10.8. (**Spectral Theorem** for self-adjoint operators on a finite-dimensional inner product space) Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be a self-adjoint linear operator. Then V has an orthonormal basis consisting of eigenvectors of T. Equivalently, T has a diagonal matrix with respect to some orthonormal basis of V.

We now build up to a proof of the Spectral theorem.

Proposition 10.9. Let $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_0, \ldots, a_n \in \mathbb{R}$. Then $p(x)$ factors, uniquely up to rearranging the order of the factors, into a product of the form

$$c(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_m)(x - \lambda_1) \cdots (x - \lambda_m),$$

where all the numbers c, b_i, c_i, λ_j are real, and $b_i^2 - 4c_i < 0$. Thus, the real roots of $p(x)$ are exactly $\lambda_1, \ldots, \lambda_m$.[9]

[9]We aren't actually claiming that $p(x)$ necessarily has any real roots. For example, consider $p(x) = x^2 + 1$. The case that $p(x)$ has no real roots should be interpreted as the case $m = 0$, when there are no factors of the form $(x - \lambda_i)$.

Proof. Prove this; see [Axl15, p. 128].

□

Remember the heuristic that self-adjoint operators play the role of *real* numbers? The heuristic makes the next proposition plausible:

Proposition 10.10. Let $b, c \in \mathbb{R}$ with $b^2 - 4c < 0$. Say T is a self-adjoint operator on a finite-dimensional real inner product space. Then $T^2 + bT + cI$ is invertible.

Proof. We claim that for all $v \in V \setminus \{\mathbf{0}\}$, that

$$\langle (T^2 + bT + cI)v, v\rangle \neq 0.$$

If so, then $(T^2 + bT + cI)v \neq 0$ for all nonzero v, so $T^2 + bT + cI$ is injective and hence invertible.

To prove the claim, first recall that the polynomial $p(x) = x^2 + bx + c$ can be written

$$p(x) = (x + \frac{b}{2})^2 + (c - \frac{b^2}{4})$$

by "completing the square." The same reasoning shows that

$$T^2 + bT + cI = (T + \frac{b}{2}I)^2 + (c - \frac{b^2}{4})I.$$

Moreover, note that $T + \frac{b}{2}I$ is self-adjoint since T is (think about its matrix with respect to an orthonormal basis.)

Now, we calculate

$$\langle (T^2 + bT + cI)v, v\rangle = \langle (T + \frac{b}{2}I)^2 v, v\rangle + \langle (c - \frac{b^2}{4})v, v\rangle$$

$$= \langle (T + \frac{b}{2}I)v, (T + \frac{b}{2}I)v\rangle + (c - \frac{b^2}{4})\langle v, v\rangle > 0,$$

and we are done. □

Proposition 10.11. Let T be a self-adjoint operator on a (nonzero) finite-dimensional *real* inner product space. Then T has an eigenvector.

Remember the context for this lemma: every operator on a finite-dimensional *complex* vector space has an eigenvector. In general, an operator on a finite-dimensional *real* vector space may fail to have an eigenvector. (For example, 60 degree rotation of $\mathbb{R}^2$ about the origin.) So the lemma says something special about self-adjoint operators on real inner product spaces amongst all operators on real inner product spaces.

Proof. Let $n = \dim V$. Let $v \in V$ be any nonzero vector and consider the $n+1$ vectors

$$v, Tv, \ldots, T^n v.$$

These must be linearly independent (why?) so there exist real numbers $a_0, \ldots, a_n$, not all zero, such that

$$a_0 v + a_1 Tv + \cdots + a_n T^n v = \mathbf{0}.$$

Rewriting: let $p(z) = a_0 + a_1 z + \cdots + a_n z^n$. Then $p(T)(v) = \mathbf{0}$.

Factoring,

$$\mathbf{0} = c(T^2 + b_1 T + c_1 I) \cdots (T^2 + b_M T + c_M I)(T - \lambda_1 I) \cdots (T - \lambda_m I)v,$$

for real numbers c, b_i, c_i, λ_j, with $b_i^2 - 4c_i < 0$ for each $i = 1, \ldots, M$.

But all the operators $T^2 + b_i T + c_i I$ are invertible. So it must be that some $T - \lambda_j I$ is not invertible. So λ_j is an eigenvalue of T. □

Now we prove Theorem 10.8, the spectral theorem for self-adjoint operators.

Proof. We'll prove it by induction on $\dim V$. (Discuss induction!) If $\dim V = 0$, then V *vacuously* has a basis consisting of eigenvectors of T. (Every single element in a basis of V is an eigenvector of T!)

Now suppose the theorem holds for all self-adjoint operators on a inner product spaces of dimension $n-1$. We will show that the theorem holds for all self-adjoint operators on inner product spaces of dimension n. Let V be an n-dimensional inner product space, for $n \geq 1$, and $T \in \mathcal{L}(V)$ a self-adjoint operator.

The only thing we need is that T has an eigenvector v, say with eigenvalue λ. (This is true over $\mathbb{R}$ by Proposition 10.11. And it is true over $\mathbb{C}$ for *every single operator*, as we showed earlier in the course. By rescaling, we may pick v to have norm 1. Let $U = \operatorname{span}(v)$.

Claim. We claim that T restricts to an operator on $U^\perp$. By this we mean that $T(w) \in U^\perp$ for all $w \in U^\perp$, so that we may regard T as giving rise to a linear operator on the vector space $U^\perp$. We write $T|_{U^\perp} \colon U^\perp \to U^\perp$ for this restriction.

To prove the claim, suppose $w \in U^\perp$; we wish to show that $Tw \in U^\perp$. Given any $u \in U$, we have $\langle Tw, u\rangle = \langle w, Tu\rangle = \overline{\lambda}\langle w, u\rangle = 0$ by self-adjointness and the fact that $Tu = \lambda u$. This proves the claim.

Finally, notice that $\dim U^\perp = n - 1$ (why?) So $U^\perp$ has an orthonormal basis of eigenvectors of $T|_{U^\perp}$. (These vectors are thus eigenvectors of T.) Adjoining this orthonormal basis of $U^\perp$ with the basis vector v for U, we obtain an orthonormal basis of V consisting of eigenvectors of T. □

Corollary 10.12. The eigenvalues of a self-adjoint operator on a finite-dimensional inner product space are always real.

Proof. This can be proved directly (see [Axl15, p. 210]), but it also follows for free from the Spectral theorem. Just think: how can a diagonal matrix be equal to its conjugate transpose? Only if the diagonal entries are all real. □

Remark 10.13. Discuss that the reverse implication in the Spectral Theorem holds if $\mathbb{F} = \mathbb{R}$. If $\mathbb{F} = \mathbb{C}$, then there is a more general class of operators for which the conclusion of the Spectral Theorem holds, namely *normal operators*.

46. Positive operators

To end the class, we're going to do the following. We'll first study two classes of operators: positive operators and isometries. Each of these classes of operators may seem quite special, so the next step may seem a surprise: we'll prove a polar decomposition theorem that says that *all* operators are a product of an isometry and a positive operator. Finally, consequence of polar decomposition theorem is Singular Value Decomposition, which is quite useful in applications.

Reference: [Axl15, p. 225] Throughout, V denotes a finite-dimensional inner product space over $\mathbb{F}$. First, a preliminary definition:

Definition 10.14. Let $T \in \mathcal{L}(V)$. By a *square root* of T we just mean any linear operator R such that $R^2 = T$.

Remark 10.15. Beware that an operator T can have infinitely many square roots, in general. For example, can you think of some square roots of $I \colon \mathbb{R}^2 \to \mathbb{R}^2$?

Definition 10.16. An operator T on V is called *positive* if it is self-adjoint and all eigenvalues are nonnegative.

Since we know all about self-adjoint operators now, we can give the following rephrasing of positivity:

T is positive iff it has a diagonal matrix with nonnegative numbers on the diagonal, with respect to some orthonormal basis.

Positive operators are also called *positive semidefinite* in the literature. (T is called *positive definite* if it is self-adjoint and all eigenvalues are strictly positive.) In retrospect, it would probably make more sense to call positive operators "nonnegative operators," but we are unfortunately stuck with the terminology.

Draw some pictures. The book uses a different definition which is equivalent to this one:

Proposition 10.17. The following are equivalent, for $T \in \mathcal{L}(V)$:

(1) T is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$. (This is how the book *defines* positive operator.)
(2) T is self-adjoint and all the eigenvalues of T are nonnegative;
(3) T has a square root that is positive.
(4) T has a square root that is self-adjoint;
(5) There exists an operator $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Proof. Let's prove most but not all of the proposition: we'll prove (2) implies (3) implies (4) implies (5) implies (2). (See [Axl15, p. 226] for the full proof.)

Suppose (2) holds. Then with respect to some orthonormal basis, it has a diagonal matrix with nonnegative diagonal entries. Take the square root down the diagonal: this new matrix is the matrix of an operator R with $R^2 = T$, and R is again positive. So (2) implies (3).

(3) implies (4) for the simple reason that positive operators are self-adjoint, by definition.

(4) implies (5): Let R be a self-adjoint square root of T. In other words, $R^2 = T$ and $R = R^*$. In other words, $T = R^*R$.

Finally, (5) implies (2) was achieved on the homework. □

Proposition 10.18. Every positive operator T on V has a *unique* positive square root, henceforth written $\sqrt{T}$.

Proof. I'm omitting this proof. See [Axl15, p. 227]. □

Example 10.19. For example, what's the unique positive square root of $I\colon \mathbb{R}^2 \to \mathbb{R}^2$?

47. Isometries

The idea of an isometry is simple: it is a distance-preserving, equivalently norm-preserving, linear map.

Definition 10.20. An operator $S \in \mathcal{L}(V)$ is called an isometry if

$$\|Sv\| = \|v\|$$

for all $v \in V$.

For your reference in the future, the following terminology is more commonly used:

> In the literature, isometries are called *orthogonal* operators if they are over $\mathbb{R}$, and *unitary* operators if they are over $\mathbb{C}$.

Example 10.21. What are some isometries of $\mathbb{R}^2$? $\mathbb{R}^3$?

It may not be apparent at first, but isometries actually preserve angles as well as lengths:

Proposition 10.22. If S is an isometry, then $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in V$.

Proof. This follows from two calculations presented in the book as Exercises 6.A.19 and 6.A.20. The calculations show that over either $\mathbb{R}$ or $\mathbb{C}$, "inner products can be rederived from norms." Therefore, if S preserves norms, then it also preserves inner products.

The calculations are as follows:

(1) In a real inner product space,
$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}.$$
(2) In a complex inner product space,
$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}.$$

□

Proposition 10.23. (Shortened version of [Axl15, p. 229]) The following are equivalent, for $S \in \mathcal{L}(V)$:

(1) S is an isometry;
(2) $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in V$;
(3) S sends any given orthonormal basis to another orthonormal basis;
(4) $S^*S = I$.

Proof. We just proved (1) implies (2).

Suppose (2) holds, so S preserves inner products. Then S preserves norms *and* the relationship of orthogonality, so (3) holds.

For (3) implies (4), suppose S sends an orthonormal basis $e_1, \dots, e_n$ to $f_1, \dots, f_n$ respectively. We claim that $S^*Se_j = e_j$ for all j; if so, we win. Indeed, for all $k = 1, \dots, n$, $\langle S^*Se_j, e_k\rangle = \langle Se_j, Se_k\rangle = \langle e_j, e_k\rangle$. The only way this can happen is if $S^*Se_j = e_j$. For (4) implies (1), we have, for all $v \in V$, $\langle Sv, Sv\rangle = \langle S^*Sv, v\rangle = \langle v, v\rangle$, so $\|Sv\| = \|v\|$.

□

48. Polar decomposition and singular value decomposition

Here are some analogies:

$\mathcal{L}(V)$	$\mathbb{C}$
complex conjugation	adjoint
self-adjoint operators	$\mathbb{R}$
positive operators	$\mathbb{R}_{\geq 0}$
isometries	$\{z \in \mathbb{C} : \lvert z\rvert = 1\}$.

These are analogies for the sake of intuition. But they aren't all fiction: they are also the literal truth in the case that V is a 1-dimensional complex inner product space. The point of this section is to add one more analogy to this table, namely an analogue of polar decomposition.

First, notice that any complex number $z \in \mathbb{C}$ can be written as the product of a complex number of norm 1 and a nonnegative real number. Precisely:

$$z = \left(\frac{z}{|z|}\right)|z| = \left(\frac{z}{|z|}\right)\sqrt{\overline{z}z}.$$

(Draw a picture).

Theorem 10.24. (Polar decomposition) Let $T \in \mathcal{L}(V)$. Then there exists an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

This is remarkable: any linear operator is the product of an isometry and a positive operator.

Definition 10.25. The *singular values* of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ repeated $\dim E(\lambda, \sqrt{T^*T})$ times.

In other words, if you express $\sqrt{T^*T}$ using a diagonal matrix with respect to an orthonormal basis, then the singular values are the numbers you see down the diagonal.

Draw some pictures. The singular values give a useful way to understand a completely general linear operator T.

Lemma 10.26. We have $\|Tv\| = \|\sqrt{T^*T}v\|$ for all $v \in V$.

Proof.

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle v, T^*Tv\rangle = \langle v, \sqrt{T^*T}\sqrt{T^*T}v\rangle = \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle = \|\sqrt{T^*T}\|^2.$$

□

Now we outline the proof of the Polar Decomposition theorem. First, some scratchwork: if such an isometry S existed, it would have the property that, for all $v \in V$, $S(\sqrt{T^*T}v) = Tv$.

Proof.

(1) Let us first define a linear map

$$S_1 \colon \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$$

by sending $S_1(\sqrt{T^*T}v) = Tv$. We have to prove first that S_1 is a well-defined function: given $v_1, v_2 \in V$ such that $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$, we need to show that $Tv_1 = Tv_2$. Indeed, using the Lemma above,

$$\|Tv_1 - Tv_2\| = \|\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2\| = 0,$$

so $Tv_1 - Tv_2 = 0$.

(2) Next, check that S_1 is indeed linear. Notice also that the Lemma says that S_1 is an isometry.

(3) Finally, simply extend S_1 arbitrarily to an isometry S on all of V by sending an orthonormal basis of $(\text{range}\sqrt{T^*T})^\perp$ to an orthonormal basis of $(\text{range}\, T)^\perp$.

□

Corollary 10.27. (Singular value decomposition) Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \ldots, s_n$. Then there exist orthonormal bases $e_1, \ldots, e_n$, $f_1, \ldots, f_n$ for V such that

$$Te_1 = s_1 f_1, \ldots, Te_n = s_n f_n.$$

In other words, any operator has a diagonal matrix with its singular values along the diagonal... as long as you permit *two different* orthonormal bases!

Remark 10.28. The general formula for T can be read off as follows: given an arbitrary $v \in V$, we obtain

$$\begin{aligned} Tv &= T(\langle v, e_1\rangle e_1 + \cdots \langle v, e_n\rangle e_n) \\ &= \langle v, e_1\rangle s_1 f_1 + \cdots + \langle v, e_n\rangle s_n f_n. \end{aligned}$$

Proof. We have $T = S\sqrt{T^*T}$ for some isometry S, by Polar decomposition. Let $e_1, \ldots, e_n$ e an orthonormal basis of eigenvectors of $\sqrt{T^*T}$, with eigenvalues $s_1, \ldots, s_n$. Let

$$f_1 = Se_1, \ldots, f_n = Se_n,$$

so $f_1, \ldots, f_n$ is also an orthonormal basis for V since S is an isometry. Then for each $i = 1, \ldots, n$,

$$Te_i = S(\sqrt{T^*T}e_i) = S(\lambda_i e_i) = \lambda_i f_i,$$

as desired. □

Part 11. Appendix: Proving things

49. Proof Clinic

Here's one way to get started writing a proof, in five steps.

Step 1: Rephrase if needed. (This is a "training wheels step." Can skip later.) Write

We wish to prove that...

Then rephrase the statement, if needed using the quantifiers

For all... *and* There exists... such that...

and by breaking down the statement into a "to-do list" if needed.

Step 2: Unpack the quantifiers one at a time from left to right, as follows.

(1) To respond to a universal quantifier

For every X satisfying [blah], statement P is true, write

Given X satisfying [blah], **we wish to show that** P is true.

(2) To respond to an existential quantifier

"There exists an X satisfying [blah] such that P is true," write

Let X **be... We claim that** P **is true.**

Can't find the right X? Try a proof by contradiction instead; see below.

Step 3: Keep unpacking definitions/givens. For example, if you wrote in step 2 that

"Given linearly independent vectors $v_1, \dots, v_m$," you could write

Since $v_1, \dots, v_m$ are linearly independent, **we know that** for all $a_1, \dots, a_m \in \mathbb{F}$, $a_1v_1 + \cdots + a_mv_m = 0$ implies $a_1 = 0, \dots, a_m = 0$.

Step 4: Prove it! This is the truly mathematical part. If you're trying to prove P, try to logically deduce a sequence of statements one by one, ending at P.

Stuck? Try a proof by contradiction: assume P is false and try to logically deduce a statement that is patently false or that contradicts something that is given in the problem as true.

Step 5: Summarize and conclude. Summarize the last thing you showed, and (if needed) how it implies what you were trying to prove.

In-class exercises for the proof clinic.

Example 11.1. Prove that for any linear map $T\colon V \to W$, if $\operatorname{null} T = \{\mathbf{0}\}$ then T is injective.

Proof.

□

For the next example, we should first agree on when two linear maps are equal:

Definition 11.2. Let $S, T\colon V \to W$ be linear maps. Then $S = T$ if and only if:

Example 11.3. Let V and W be vector spaces, let $v_1, \ldots, v_m$ be vectors in V, and let $w_1, \ldots w_m$ be vectors in W.

Suppose $v_1, \ldots, v_m$ span V. Prove that there is at most one linear map $T\colon V \to W$ such that

$$T(v_1) = w_1, \ldots, T(v_m) = w_m.$$

Proof.

□

50. How to prove it

(1) **Set containment:**

For sets A and B, **prove that** $A \subseteq B$.

Almost always, you should immediately write "Given $a \in A \ldots$" Then prove that a is in fact in B. In other words, $A \subseteq B$ means *for all $a \in A$, we have that $a \in B$.*
The *contrapositive* can sometimes be used instead: show that any element not in B is in fact not in A either.

(2) **Set equality:**

For sets A and B, **prove that** $A = B$.

What does $A = B$ mean? It means that $A \subseteq B$ and, simultaneously, $B \subseteq A$. So usually you do two proofs of set containment.
Sometimes, one of the containments is already given to you. For example, suppose you are asked to prove that

$$\mathbb{Q} = \{a \in \mathbb{Q} : a = 2b \text{ for some } b \in \mathbb{Q}\}.$$

Here, *by its construction*, the set on the left hand side is evidently a subset of $\mathbb{Q}$. So you need only prove the containment $\subseteq$, namely, that *every* rational number a is of the form $2b$ for some rational number b. (Do you believe it?)

(3) **Uniqueness:**

Prove that there is **at most one** blah such that [some condition holds].

In this situation, you suppose that there are two blahs such that the condition holds; name them x and y. Then endeavor to show that $x = y$.
This may feel unnatural at first. It's not really the way we argue things in real life—is it? Say you want to argue that you have *at most one* sister. That's a weird thing to argue. Think of it as arguing *against* a contrarian who insists you have two sisters: "there's the sister x I went to middle school with, and the sister y I saw at the grocery store the other day." You have to argue that *those two people are one and the same*: you have to argue $x = y$.

(4) **Existence and uniqueness:**

Prove that there **exists a unique** blah such that [some condition holds].

In this situation, there are two claims: there's at least one blah, and there's at most one blah. You have to prove them both.
It is counterintuitive, but it is often useful to **prove uniqueness first, then existence.** Basically, when proving uniqueness, you tend to narrow down what the object blah could be, which then makes it more direct to actually exhibit the thing.

References

[Axl15] Sheldon Axler, *Linear algebra done right*, 3rd ed., Undergraduate Texts in Mathematics, Springer, 2015.

[Ham18] Richard Hammack, *Book of proof*, 3.2 ed., Richard Hammack, 2018.

Department of Mathematics, Brown University, Box 1917, Providence, RI 02912
E-mail address: melody_chan@brown.edu