Math 121: Linear Algebra and Applications

Course Taught by Elden Elmanto
Notes Taken by Forrest Flesher

Fall 2019

Welcome to Math 121: Linear Algebra and Applications. I'm Forrest, your course assistant for the semester.

Some important information:

- The course webpage is:
  https://scholar.harvard.edu/elden/classes/math-121-fall-2019
- Office hour times, course times and more information is located in the syllabus, which is at:
  https://scholar.harvard.edu/files/elden/files/syllabus-121_fall_2019-final.pdf
  Office hours are at the following times in Science Center 231:
  - Sundays: 2:00-3:00 pm
  - Thursdays: 2:00-3:00 pm
- I will hold two office hours per week. The first Friday office hour will be an interactive LaTeX introduction session. My office hours are at the following times:
  - Mondays: 8:00-10:00 pm, Leverett Dining hall
  - Fridays: 3:00-4:15 pm, Science Center 309A
- The text for the course is Axler's "Linear Algebra Done Right."
- I will by typing up notes in class, since there will be times when the course deviates from the book.
- My email is forrestflesher@college.harvard.edu. Email with any questions, comments, or concerns, especially if you find a mistake in these notes.
- We will use the Canvas site for submitting/grading problem sets.
- For submitting your problem sets, please use LaTeX. This is strongly encouraged for the first few psets, and required after the fifth pset. If you are new tex and need help getting started, come to office hours. I'd recommend using overleaf: https://www.overleaf.com

Contents

§1 September 3, 2019

§1.1 Introduction

The class will have two learning outcomes:

1. Introduction to proofs. You will not be expected to multiply massive matrices, but will be expected to write good, understandable proofs.
2. Understand linear algebra, and applications. We will cover theoretical aspects, and some applications. Prior exposure to linear algebra is helpful.

Some introductory information:

- **Homework**: Worth 50% of the grade. Due before class on Tuesday, and assigned after class on Tuesday. The lowest homework will be dropped.
- **Exams**: Worth 10% of the grade each (2 exams)
  The exams are September 26 and October 31.
- **LaTeX**: You are encouraged to tex problem sets. The first week, you get +10 points for texing your problem set, the second week, +8, ..., the fifth week +0 and the 6th week -2,....
- **Survey**: Please fill out the survey.
- **Office Hours**: Email: elmanto@math.harvard.edu Science Center 231
  Tuesdays: 9:30am-10:30am
  Thursdays: 2:00pm-3:00pm
- **Final Project**: Worth 30% of the grade. An expository paper on a linear algebra topic.

§1.2 What is a proof?

- Logic.
- Build on accepted theorems (other proofs).
- Axioms, rules of inference.
- Rigorous and airtight.

These are all correct. We now demonstrate an example of a (stupid) proof:
Consider the following multiplication table on the set $\{0, 1\}$:

$$0 + 0 = 0$$
$$0 + 1 = 1$$
$$1 + 0 = 1$$
$$1 + 1 = 0.$$

We make a proposition about this multiplication table:

Proposition 1.1
Given any element $x \in S$, there exists a unique $y \in S$ such that $x + y = 0$.

Remark 1.2. For this, we need to prove two things: uniqueness and existence. What we need to show is that for any x, there is a unique element which we will call '$-x$.'

Proof. Proof by inspection. We examine every element, and conclude the result. If $x = 0$, we can use 0. If $x = 1$, then we can use 1. Since it is true for 0 and 1, the proposition is proved. □

Now we will examine an example of a famous, nontrivial proof, due to Euclid. First, a couple of definitions:

Definition 1.3 — We say that a **divides** b if there exists a c such that $b = a \cdot c$. We write $a \mid b$. A number $p \geq 2$ is **prime** if the only positive divisors of p are 1 and p. For example: 2,3,5,7,11.

Theorem 1.4

There are infinitely many prime numbers.

To prove this theorem we will use the following lemma, the proof of which will be in your homework (left as an exercise to the reader):

Lemma 1.5 (Fundamental Theorem of Arithmetic)

Every positive integer n can be written as a product of $n = p_1^{\alpha_1} \cdot \cdots \cdot p_k^{\alpha_k}$, where each p_i is prime, and α_i is a positive integer. This is known as prime decomposition.

Now for the proof of the theorem:

Proof. Suppose that this is not true, i.e. the there are only *finitely* many primes. Since there are only finitely many, we can list them all: $p_1, p_2, \ldots, p_n$. Define the number $N := p_1 \cdot p_2 \cdot p_n + 1$, the product of all the primes, plus 1. To complete the proof, we use the lemma above. By lemma 1.5, we can write N as a product of prime numbers. But this cannot be any of the prime numbers in the list, since there will always be a remainder of 1 when N is divided by $p_1, \ldots, p_n$. Hence N has a prime factor which was not in the list. But this means that the list of primes does not contain all the prime numbers, which is a contradiction. Thus, the assertion that there are only finitely many primes is invalid, and there are infinitely many primes. □

This is a nonconstructive proof: Euclid was able to avoid listing all primes, but still prove their infinitude. Now that we've seen an introductory example of a proof, we move on to the linear algebra.

§1.3 What is linear algebra?

So what is linear algebra? Some concepts:

- Matrices
- Vectors
- Dimensions

- $\otimes$

Linear algebra is extremely useful in research math, applications, and everywhere. Let's look at an example of an application in coding theory

Example 1.6 (Coding Theory)

Suppose that Sandy wants to send Randy the code 0101. Sandy puts this code into a coder, which transforms the code into 0101100. In this case, the message is just the first four letters. Randy has a decoder, which takes off the first four letters, and Randy sees 0101. But what if there was a thunderstorm, which corrupted the code. In this case, Sandy will need to send the code three times in order for Randy to know whether or not the code was corrupted.

To make this more efficient, we'll use an idea of Hamming. Write the code 0101 as a column vector:

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

Sandy sends an additional three digits at the end of the message x_1, x_2, x_3, x_4. He sends $x_1, x_2, x_3, x_4, e_1, e_2, e_3$, where $e_1 = (x_1 + x_2 + x_4), e_2 = (x_2 + x_3 + x_4), e_3 = (x_1 + x_3 + x_4)$, where the addition is modulo 2 (the addition system from above where $1 + 1 = 0$). Randy receives some code $c_1, c_2, c_3, c_4, c_5, c_6, c_7$. He checks that $c_1 + c_2 + c_4 + c_5 = c_2 + c_3 + c_4 + c_6 = c_1 + c_3 + c_4 + c_7 = 0$. This system corrects codes. Hamming codes are widely used in ECC memory (e.g. in Intel Xeon processors). This is linear algebra, since we can do this using matrices (we will talk more about this later in the course). It might help to draw the Venn diagram to visualize adding the numbers, as we did in class.

§1.4 Proof Techniques

Proof by contraposition: To prove $A \implies B$, it suffices to prove $\neg B \implies \neg A$. For example, we can prove the following proposition:

Proposition 1.7

If n^2 is odd, then n is odd.

Proof. It suffices to prove that n is even implies that n^2 is even. If n is even then $n = 2k$ for some k. Then $n^2 = (2k)^2 = 4k^2 = 2(2k^2)$, so n^2 is even, and we are done. □

Proof by contradiction: Suppose that the result is false, and show that this leads to a contradiction. WARNING: not everything is a contradiction: there is no problem with using contradiction, but sometimes it doesn't work better. For example, we can prove the following proposition:

Proposition 1.8

The number $\sqrt{2}$ is irrational.

Proof. Suppose that $\sqrt{2}$ is rational. Then $\sqrt{2} = \frac{p}{q}$ in reduced form, where $p, q \in \mathbf{Z}$, and p and q are relatively prime, i.e. one does not divide the other. Square both sides, and we have $2 = \frac{p^2}{q^2}$. Then $2q^2 = p^2$. This means that p^2 is even, so p is even, so $p = 2x$ for some x. We can substitute and write $2q^2 = (2x^2) = 4x^2$. Cancel on both sides to get $q^2 = 2x^2$. This implies that q^2 is even, which implies that q is even. But then 2 divides both p and q, and they are not relatively prime, which is a contradiction. Thus, we must have that $\sqrt{2}$ is irrational. □

§2 September 5, 2019

As a reminder, your first midterm exam will be on September 26th, and the homework is due Tuesday at the beginning of class (submit on Canvas).

§2.1 Proof Techniques (continued)

Last class, we talked about proof by **contraposition** and proof by **contradiction**. Today, we begin with proof by **induction**. The idea behind induction is that if we know something is true for 1, and that being true for any number implies it is true for the next number, then it is true for all number, in other words: we show that a statement is true for 1, and then we show that (true for 1) implies (true for 2) implies (true for 3) .... We give an example of proof by induction.

Proposition 2.1

For positive integers n, the following formula is true:

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Proof. We proceed by induction. For the base case, we see that $1 = \frac{1(1+1)}{2}$. Now, we assume that the statement is true for n, and our goal is to show that the result is true for $n+1$. In other words, we need to show that

$$1 + \cdots + (n+1) = \frac{(n+1)(n+2)}{2}.$$

By the induction hypothesis (since it is true for n), we have that

$$1 + 2 + \cdots + n + (n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{(n+1)(n+2)}{2}.$$

Thus, the result is true for $n+1$, and therefore for all integers. □

As another example, the following proposition can be proven by induction:

Proposition 2.2

The number of ways of arranging the set $\{1, \ldots, n\}$ is $n!$.

Now that we've covered these proof techniques, we move on to some ideas in set theory: Cartesian products, equivalence relations, and others. These will be very important for everything we do in linear algebra.

§2.2 Basic Operations on Sets

Let's begin with some definitions. We won't state the definition of a set.

Definition 2.3 — We say that A is a **subset** of B, and we write $A \subseteq B$, if for any $a \in A$, a is also in B. We say that A is a **proper** subset of B, and we write $A \subset B$ if there exists $b \in B$ which is not in A.

Proposition 2.4

If A, B, C are sets, then

1. $A \subseteq A$
2. $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. If we want to prove $A = B$ for some sets, we can break this down to the two steps of showing $A \subseteq B$ and $B \subseteq A$ (this is very useful).
3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.
4. The empty set is a subset of every set: $\varnothing \subseteq A$.

The proof of this proposition is left as an exercise to the reader. We now cover some important definitions in set theory.

Definition 2.5 — If A, B are sets, then the **union** of A and B is denoted $A \cup B$, and defined as

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

The **intersection** of A and B is denoted $A \cap B$, and defined as

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Recall set builder notation: for example the set X of all tropical fruits would be denoted $X = \{x : x \text{ is a tropical fruit}\}$, read "the set of all x such that x is a tropical fruit."

Proposition 2.6 (Boolean Rules)

If A, B, C are sets, then

1. $(A \cup B) \cup C = A \cup (B \cup C)$
2. $(A \cap B) \cap C = A \cap (B \cap C)$
3. $A \cup \varnothing = A$
4. $A \cup B = \varnothing$ if and only if $A = \varnothing$ and $B = \varnothing$.
5. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Again, the proof is left as an exercise to the reader.

Definition 2.7 — If X is a set, and $A \subseteq X$, then the **complement** of A in X is defined as $A^c = \{x \in X : x \notin A\}$.

Proposition 2.8 (Demorgan's rules)
If A, B, X are sets, and $A, B \subseteq X$, then

1. $(A \cup B)^c = A^c \cap B^c$
2. $(A \cap B)^c = A^c \cup B^c$

Proof. 1. To prove this statement, we need to prove that

(a)
$$(A \cup B)^c \subseteq A^c \cap B^c$$

(b)
$$A^c \cap B^c \subseteq (A \cup B)^c$$

. To show (a), suppose that $X \in (A \cup B)^c$, which is the same as saying that $x \notin A \cup B$. Then by definition, then $x \notin A$ and $x \notin B$. Since $x \notin A$ then $x \in A^c$, and since $x \notin B$ then $x \in B^c$. Since $x \in A^c$ and $x \in B^c$, then $x \in A^c \cap B^c$, by definition of intersection. Thus, $x \in (A \cup B)^c$ implies that $x \in A^c \cap B^c$, which means that $(A \cup B)^c \subseteq A^c \cap B^c$. Use a similar method to show (b).

2. Left as an exercise to the reader.

□

§2.3 Cartesian Products

Our punchline for Cartesian products will be "dimension boosting." Think about why this punchline makes sense. We start with the definition of the Cartesian product:

Definition 2.9 — If A and B are sets, then the **Cartesian product** of A and B is defined as
$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$
It is import to remember that a and b are treated as *independent* elements in $A \times B$

Example 2.10
If $A = B = \mathbf{R}$, then $A \times B$ is the familiar xy-plane. You can think about the xy-plane as a family of lines parameterized by another line. That is, we take the x-axis, and we move it along the y axis to make the entire xy-plane. Don't think too hard about what the word 'parameterized' means, but do try to think about this in terms of our punchline of "dimension boosting."

Proposition 2.11
If A and B are sets, and B has n elements (n is finite), and B has m elements (m also finite). Then $A \times B$ has $n \cdot m$ elements.

Why is this true? Write out the elements of A and B on a grid, where the horizontal is A and the vertical is B:

$$\begin{pmatrix} m & \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 3 & \cdot & \cdot & \cdots & \cdot \\ 2 & \cdot & \cdot & \cdots & \cdot \\ 1 & 2 & 3 & \cdots & n \end{pmatrix}.$$

Think about why this pictorial representation tells us about the number of elements in $A \times B$. To prove the proposition formally (without pictures), you should use induction.

§2.4 Equivalence Relations

Our punchline for equivalence relations will be "gluing." To begin, we look at the example of the torus (donut shape). Take the xy-plane, and draw a 1×1 square with the corner at $(0, 0)$. Now, we set the bottom edge equal to the top edge, and we get a cylinder (try with a piece of paper as the square). Then, we set the left edge equal to the right edge, and the cylinder becomes a torus (we have "glued" edges of a square together to get a torus). To make this notion of "gluing edges" more precise mathematically, we need equivalence relations.

Definition 2.12 — Let X be a set. Then an **equivalence relation** on X is a subset $R \subseteq X \times X$, subject to the following rules:

1. **Reflexive**: For all $a \in X$ the element $(a, a) \in R$. That is, everything on the "diagonal" is in the set R.
2. **Symmetric**: If $(a, b) \in R$, then $(b, a) \in R$.
3. **Transitive**: If $(a, b) \in R$ and $(b, c) \in R$, then $(a, c) \in R$.

Usually, instead of writing $(a, b) \in R$, we write $a \sim b$ (read "a is equivalent to b" or "a is related to b"). We can rewrite the properties above using this notation:

1. **Reflexive**: $a \sim a$
2. **Symmetric**: If $a \sim b$ then $b \sim a$.
3. **Transitive**: If $a \sim b$ and $b \sim c$ then $a \sim c$.

This definition is rather abstract, so looking at examples will be useful. We will go over a very important example now:

Example 2.13

Let $X = \mathbf{Z}$, and $n \in \mathbf{Z}$, with $n \geq 2$. Define the following equivalence relation: $a \sim b$ if and only if $a - b$ is divisible by n. If we use the set notation, then we write the relation as $R \subseteq \mathbf{Z} \times \mathbf{Z}$, and we have $(a, b) \in R$ if and only if $a - b$ is divisible by n.

Proof. We need to prove that the relation satisfies the three properties of an equivalence relation.

1. To show $a \sim a$, we need to show that $a - a$ is divisible by n. But $a - a = 0$, and 0 is divisible by n for any n.

2. Suppose that $a \sim b$. We need to show that $b \sim a$. Since $a \sim b$, then $a - b$ is divisible by n. To show $b \sim a$, we need to show that $b - a$ is divisible by n. But $b - a = -(a - b)$. Since $a - b$ is divisible by n, then $-(a - b)$ is also divisible by n, which means that $b - a$ is divisible by n, so $b \sim a$.

3. Suppose that $a \sim b$ and $b \sim c$. This means that $a - b$ is divisible by n and $b - c$ is divisible by n. We need to show that $a - c$ is divisible by n. But $a - c = a - b + b - c$, which we can write as $(a - b) + (b - c)$. Since $a - b$ and $b - c$ are divisible by n then their sum is also divisible by n, so $(a - b) + (b - c) = a - c$ is divisible by n.

□

§3 September 10, 2019

If you are interested, there is a talk tomorrow at 4:30pm in SC507 at the Open Neighborhood Seminar by Joe Harris, entitled "How many lines meet each of 4 given lines in 3-space, and why it matters." Snacks and refreshments will be served in the common room afterwards.

Also, there will be a new office hours on Sundays at 2:00pm.

§3.1 Descending Operations

Our goal will be to define what are called **finite fields**, but first we discuss descending operations. The format for descending operations is that we have some operation on a "bigger" set, and we want to induce an operation on a "smaller" set. We now discuss a definition.

Definition 3.1 — The set $\mathbf{Z}/n$ is defined as $\mathbf{Z}/\sim$, which is the set $\{C(a) : a \in \mathbf{Z}\}$, where the set $C(a)$ is the one defined in your first homework. The $\sim$ is the equivalence relation $a \sim b$ if $a - b$ is divisible by n, for $n \geq 2$. This is a rather abstract definition, so let's look at an example.

Example 3.2

$$\mathbf{Z}/5 = \{C(0), C(1), C(2), C(3), C(4)\}.$$

The set $C(0)$ is $C(0) = \{0, 5, 9, \ldots\}$, the set $C(1)$ is $C(1) = \{1, 6, 10, \ldots\}$. We read $\mathbf{Z}/5$ as "z mod 5." We also write it as the set $\{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}\}$, where the numbers with lines over them represent classes of numbers. That is, $\mathbf{Z}$ is a set with 5 elements, which we think of as $\{0, 1, 2, 3, 4\}$. We now want to "naturally endow" the set $\mathbf{Z}/5$ with an operation of addition - that is, we would like to be able to add things together in the set. This "addition" is modular addition, which you might be familiar with from computer science. To do this formally, we define $\bar{i} + \bar{j}$ (the "sum" of $\bar{i}$ and $\bar{j}$ inside this new set $\mathbf{Z}/5$), as $\overline{i + j}$. In other words $\bar{0} + \bar{1} = \overline{0 + 1} = \bar{1}$, and $\bar{3} + \bar{2} = \overline{3 + 2} = \bar{5} = \bar{0}$. It's important to note that this is our *definition* of addition. We want to show that this operation is well defined (i.e. using $\bar{6}$ versus $\bar{1}$ doesn't matter when we add things).

Proposition 3.3
The operation + defined in the example above is well defined.

Proof. Suppose that $i' \in \bar{i}$ and $j' \in \bar{j}$. Then we need to show that $i' + j' \in \overline{i+j}$, or in other words, that $\overline{i'+j'} = \overline{i+j}$. This means that the sum of the equivalence classes $\bar{i} + \bar{j}$ doesn't depend on which element of the class we use for addition. That is, we want to show that it doesn't matter whether we use i or i' to represent the equivalence class. To do this, notice that $i + j - (i' + j') = (i - i') + (j - j')$, which is divisible by n since $n \mid (i - i')$ and $n \mid (j - j')$. This means that $i + j$ and $i' + j'$ are in the same equivalence class, so $\overline{i+j} = \overline{i'+j'}$. □

Proposition 3.4
Define $\bar{i} \cdot \bar{j} = \overline{ij}$. This operation is also well defined.

The proof is in your homework. The phrase "well defined" might be confusing at first - it just means that our results don't depend on the representations we use. For example, if we have a map $f(x)$, and we plug in $x = 6$, then $f(6)$ should be the same whether we write $f(6)$ or $f(2 \cdot 3)$.

We will now see why for the rest of the class, we will restrict our attention on $\mathbf{Z}/n$ to $\mathbf{Z}/p$, where p is a prime number. The reason is that we want to have **multiplicative inverses**. If n is not prime, we always do this. For example, in $\mathbf{Z}/6$, try to think of a multiplicative inverse of $\bar{2}$ (hint: there isn't one). In addition, if n is not prime, then we have **zero divisors**, which are things that multiply to zero. For example, in $\mathbf{Z}/6$, we have $\bar{2} \cdot \bar{3} = 0$, which is bad. If p is prime, then we always have multiplicative inverses, which is the result of the next theorem.

Theorem 3.5
If $\bar{0} \neq \bar{i} \in \mathbf{Z}/p$, then there exists $\bar{0} \neq \bar{j} \in \mathbf{Z}/p$ such that $\bar{i} \cdot \bar{j} = 1$.

Why do we care about having multiplicative inverses? For example, we would like to be able to invert matrices:

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

Proof. Begin by listing the nonzero equivalence classes in a set $A = \{\bar{1}, \ldots, \overline{p-1}\}$. Then multiply through by $\bar{i}$, to get the set $B = \{\bar{1} \cdot \bar{i}, \ldots, \overline{p-1} \cdot \bar{i}\}$. We need to show that

1. B has $p - 1$ elements, and
2. $0 \notin B$.

This is enough because if B has $p - 1$ elements and $0 \notin B$, then $\bar{1}$ must be in B, since there are only $p - 1$ options total for nonzero numbers, and B would have all of them.

The idea behind (1) is that multiplying through by $\bar{i}$ does not send two elements to the same thing. We need to show that if $\bar{k} \cdot \bar{i} = \bar{k'} \cdot \bar{i}$, then $\bar{k} = \bar{k'}$. To prove this, note that $\bar{k} \cdot \bar{i} = \bar{k'} \cdot \bar{i}$ implies that $\overline{k - k'} \cdot \bar{i} = \bar{0}$, which means that $(k - k') \cdot i$ is divisible by p. But p

does not divide i, since $\bar{i} \neq \bar{0}$, and since p is prime, then p divides $k - k'$. This means that $\overline{k - k'} = \bar{0}$, which means that $\bar{k} = \bar{k'}$, so we are done proving (1).

We now need to prove (2). To do this, notice that if $\bar{k} \cdot \bar{i} = \bar{0}$, then $\bar{k} = \bar{0}$. That is, we can't have zero divisors. To see why this is true, if $\bar{k} \cdot \bar{i} = \bar{0}$, then p divides $k \cdot i$, and since p does not divide i, then p divides k, so $\bar{k} = 0$. However, we assumed that $\bar{k}$ was nonzero to begin with, which means that $\bar{0} \notin B$. □

Below is an important fact about prime numbers which we used in the proof above:

Lemma 3.6

If $m, n \in \mathbf{Z}$, and mn is divisible by p, then $p \mid m$ or $p \mid n$.

We have shown that if $n = p$ is a prime, then for any $\bar{i} \in \mathbf{Z}/p$, $\bar{i} \neq \bar{0}$, so that every element in $\mathbf{Z}/p$ has a **multiplicative inverse**, if p is prime. This is sometimes called Fermat's Little Theorem.

§3.2 Fields

As motivation for why we need fields, lets look at the inverse of a matrix:

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

Notice that in order to do matrix multiplication, we need to be able to add, subtract, multiply, and divide (we divide 1 by $ad - bc$). These properties are used to define a field.

Definition 3.7 — A **field** is is a set K with two operations: $+ : K \times K \to K$ and $\cdot : K \times K \to K$, which follow these axioms

1. **Associativity**: For all $x, y, z \in K$, then $(x + y) + z = x + (y + z)$, and $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.
2. **Commutativity**: For all $x, y \in K$, then $x + y = y + x$, and $x \cdot y = y \cdot x$.
3. **Distributivity**: For all $x, y, z \in K$, then $x \cdot (y + z) = x \cdot y + x \cdot z$.
4. **Additive Identity**: There exists an element $0 \in K$ such that for all $x \in K$, $x + 0 = x$.
5. **Additive Inverse**: For any $x \in K$, there exists some element $y \in K$ such that $x + y = 0$.
6. **Multiplicative Identity**: There exists an element $1 \in K$ such that for all $x \in K$, $x\dot{1} = x$.
7. **Multiplicative Inverse**: For any $x \in K$, there exists some element $z \in K$ such that $x \cdot z = 1$.

Let's go over some examples of fields.

Example 3.8

- The real numbers is a field: $(\mathbf{R}, +, \cdot, 0, 1)$.
- The complex numbers is a field: $(\mathbf{C}, +, \cdot, 0, 1)$.

- The rational numbers is a field: $(\mathbf{Q}, +, \cdot, 0, 1)$.
- The set $\mathbf{Z}/p$ with the operations $+, \cdot$, discussed above with p prime is a field: $(\mathbf{Z}/p, +, \cdot, \bar{0}, \bar{1})$.

The integers $(\mathbf{Z}, +, \cdot, 0, 1)$ is *not* a field, since there are no multiplicative inverses (since $\mathbf{Z}$ does not contain $1/2$ for example). The *positive* real numbers $\mathbf{R}_{>0}$ is *not* a field, since there are no additive inverses.

§4 September 12, 2019

§4.1 Functions

We start with the definition of a function. We will see again the notion of 'well defined' maps. Here's the (formal) definition of a function:

Definition 4.1 — A **function** between two sets X and Y, written $f : X \to Y$, is a subset $\Gamma_F \subseteq X \times Y$ of the form

$$\Gamma_f = \{(x, y) : \forall x, \exists! y \in Y\}.$$

The ! in front of the $\exists$ means there exists a *unique* y. We would read this line as "the set of all (x, y) such that for all x, there exists a unique y in Y."

A classic example of a function is $f(x) = x^2$, whose graph is a parabola. Notice that for element on the x-axis, there is only one element on the y-axis corresponding to it (it is okay for multiple points on the x-axis to correspond to the same point on the y-axis, but not vice-versa). You might be familiar with the vertical line test - if a vertical line goes through multiple points on the graph, then it is *not* a function. This is a good way to remember what well defined means. For example, the horizontal parabola is not a well defined function.

Another important property of functions that we will need is injective/surjective properties.

Definition 4.2 — A function $f : X \to Y$ is **injective** if for all $x, y \in X$, $f(x) = f(y)$ implies $x = y$. This is also known as **one-to-one**.

A function $f : X \to Y$ is **surjective** if for all $y \in Y$, there exists an $x \in X$ such that $f(x) = y$. This is also known as **onto**.

A **bijective** function is a function which is both injective and surjective. This is also known as **one-to-one and onto**.

As we move through the course, it will help for remembering these definitions to think of "well defined" as passing the *vertical* line test, and as "injective" as passing the *horizontal* line test. The function $f(x) = x^2$ is well defined, since it passes the vertical line test, but not injective, since it doesn't pass the horizontal line test.

A note about arrows: if f is a function from the set X to the set Y, we write $f : X \to Y$. We write $f(x) = y$ as $f : x \mapsto y$ (where x and y are *elements* of X and Y). If f is injective, we write $f : X \hookrightarrow Y$, and if f is surjective, we write $f : X \twoheadrightarrow Y$.

One last note: we can define equivalence classes by looking at the function $\mathbf{Z} \to \mathbf{Z}/n$, defined by $i \mapsto C(i)$. This function is a surjection

Proposition 4.3

A function $f : A \to B$ is bijective if and only if there exists $f^{-1} : B \to A$ such that $f \circ f^{-1}(x) = x$ and $f^{-1} \circ f(y) = y$. That is, a function is bijective if and only if it has an inverse.

Proof. Suppose that f is bijective (injective and surjective). For $y \in B$, let x be the unique element of A such that $f(x) = y$. It exists because the function is surjective, and unique because f is injective. Define $f^{-1} : B \to A$ by $y \mapsto x$. This inverse function is well defined because f is injective (since our choice of x was unique). Then $f \circ f^{-1} : y \mapsto x \mapsto y$, and $f^{-1} \circ f : x \mapsto y \mapsto x$, so we are done with this direction of the proof.

Now, suppose that the two composition conditions are satisfied. We need to show that if $x, y \in A$ are such that $f(x) = f(y)$, then we must have $x = y$. To do this, we apply the inverse to both sides of this equation: $f^{-1}(f(x)) = f^{-1}(f(y))$. By definition of the inverse, this simplifies to $x = y$, and we are done proving that the function is injective. We still need to prove that the function is surjective. To do this, suppose that $y \in B$. We need to show that there exists some x with $f(x) = y$. Apply the function f to the element $x = f^{-1}(y) \in A$, to get $f(f^{-1}(y)) = y$. We have found an element with $f(x) = y$, and the function is surjective. □

§4.2 More on Fields

First, we should all recall definition 3.7, the definition of a field. Some examples of fields are the real numbers $\mathbf{R}$ and the complex numbers $\mathbf{C}$ (these are the fields that are the main focus in Axler). Some more fields are $\mathbf{Z}/p$ where p is prime, the rationals $\mathbf{Q}$, the constructible numbers, and the p-adic numbers. Some non examples are the natural numbers $(\mathbf{N}, +, \cdot, 0, 1)$ and the integers $(\mathbf{Z}, +, \cdot, 0, 1)$. Now we move on to a new proposition:

Proposition 4.4

Let K be a field, and $x, y \in K$. Then if $x \cdot y = 0$, the $x = 0$ or $y = 0$ ('or' is inclusive - in general unless explicitly noted, 'or' will always include the 'and' case as well). This proposition says that fields do not have zero divisors.

Proof. If $x = 0$ we are done. Thus, we should suppose that $x \neq 0$. In this case, multiply $x \cdot y = 0$ by x^{-1} on both sides, we're in a field, so x^{-1} exists. Then we have that

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

We've shown that $y = 0$, so we're done. We've used the fact that for any $z \in K$, then $z \cdot 0 = 0$. To see why this is true, notice that

$$\begin{aligned} z \cdot 0 &= z \cdot (0 + 0) \\ &= z \cdot 0 + z \cdot 0 \\ &= 2 \cdot (z \cdot 0), \end{aligned}$$

which implies that $0 = z \cdot 0$. □

Before we move on, we'll examine one last fact about fields. As a remark, note that for any field K, there exists a (canonical) function can such that

$$\begin{aligned} \text{can} : &\mathbf{Z} \to K \\ &1 \mapsto 1 \\ &\underbrace{1 + \cdots + 1}_{n \text{ times}} \mapsto \underbrace{1 + \cdots + 1}_{n \text{ times}} \\ &0 \mapsto 0. \end{aligned}$$

We can ask if this map is injective or surjective. It's not always injective, for example because we can have the map $\mathbf{Z} \twoheadrightarrow \mathbf{Z}/p$ (which is surjective but not injective). It's also not always surjective, for example because we can consider $\mathbf{Z} \hookrightarrow \mathbf{Q}$ (which is injective but not surjective). We are lead to the following definition:

Definition 4.5 — The **characteristic** of a field K is the *smallest* n such that $\text{can} : n \mapsto 0$ in the canonical map from $\mathbf{Z}$ to K If n is not finite then we say that the field K has **characteristic zero**.

Example 4.6

- The characteristic of $\mathbf{Z}/p$ is p, since $\underbrace{1 + \cdots + 1}_{p \text{ times}} \mapsto p = 0$, under the map $\mathbf{Z} \twoheadrightarrow \mathbf{Z}/p$.
- The characteristic of $\mathbf{Q}$ is zero, under the map $\mathbf{Z} \hookrightarrow \mathbf{Q}$.

An important note about fields is that $0 \neq 1$ in *all* fields. This means that the additive identity cannot be the same thing as the multiplicative identity, which means there is no field with one element. There are however, fields with two elements (which are often useful in problem sets for counterexamples). Now, we have the following proposition about characteristic:

Proposition 4.7

The characteristic of a field K is either zero or prime.

Proof. Suppose that the characteristic of K is $n \neq 0_{\mathbf{Z}}$, and further suppose that n is not prime. Then n is a composite of two numbers $n = f \cdot g$. Since n is the characteristic of the field, then the image of n in the field is zero (n itself is not zero, but $\text{can}(n) = 0_K$ in the field). That is, $1 + \cdots + 1 = 0_K$ (n times) in K. To be clear, there is a difference between n, which is an integer, and the image of n, which is in the field. In the integers, $n > 0_{\mathbf{Z}}$, but the image $\text{can}(n)$ is zero (by definition of characteristic). For example, in $\mathbf{Z} \to \mathbf{Z}/5$, the integer 5 maps to the element $[5] = [0]$ of $\mathbf{Z}/5$, which is an equivalence class. Here, 0_K is the zero element of the field K, and $0_{\mathbf{Z}}$ is the zero element of the integers.

To summarize, if the characteristic n of the field K is nonzero, then $n > 0_{\mathbf{Z}}$ and $\text{can}(n) = 0_K$. Since $n = f \cdot g$, then when we map into the field, we have $0_K = \text{can}(n) = \text{can}(f \cdot g) = \text{can}(f) \cdot \text{can}(g)$. (Think about why the last equality is true in terms of the definition of the canonical map.) Since $\text{can}(f) \cdot \text{can}(g) = 0_K$, then by proposition 4.4, either $\text{can}(f)$ or $\text{can}(g)$ must be zero. But $n = f \cdot g$ in the integers, which means $f, g < n$. This contradicts the definition of the characteristic as the *smallest* number with the property $\text{can}(n) = 0$, which finishes the proof. □

The above proof might seem confusing. As an example, let's think about $\mathbf{Z}/6$, which looks like it has "characterstic" 6. We showed in proposition 4.4 that in a field, $xy = 0$ implies $x = 0$ or $y = 0$. In $\mathbf{Z}/6$, we have $[2] \cdot [3] = [6] = [0]$, and neither $[2]$ nor $[3]$ is zero. This means that $\mathbf{Z}/6$ is not a field, which means "characteristic 6" doesn't really exist.

§5 September 17, 2019

Remember, there is a midterm on Thursday September 26th, during class.

§5.1 Vector Spaces: Definitions and Examples

Vector spaces are the central objects of linear algebra. Let's begin with the definition of a vector space, and then look at some examples.

Definition 5.1 — Let k be a field. A **vector space** over k is a set V, with two functions, **addition**: $+ : V \times V \to V$, and **scalar multiplication**: $\cdot : k \times V \to V$, which satisfy the following properties.

- **Commutativity**: For all $v, w, \in V$, then $v + w = w + v$.
- **Associativity**: For all $u, v, w \in V$, then $(u + v) + w = u + (v + w)$.
- **Additive Identity**: There exists an element in V which we denote 0, such that for any $v \in V$, $0 + v = v$.
- **Additive Inverses**: For any $v \in V$, there exists some $v' \in V$ such that $v + v' = 0$
- **Multiplicative Identity**: For any $v \in V$, then for $1 \in k$, we have that $1 \cdot v = v$.
- **Distributivity**: For all $v, w, \in V$, and for all $\alpha, \beta \in k$, then $\alpha \cdot (v + w) = \alpha \cdot v + \alpha \cdot w$, and $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$.

It's important that we are very clear on the distinction between a field and a vector space.

Remark 5.2.
- In a field, we are allowed to multiply elements together: if x and y are elements of the field, we can multiply $x \cdot y$. We can't do this in a vector space: we can only multiply by *scalars*, which are elements of the field: if $v \in V$ and $a \in k$, we can multiply $a \cdot k$.
- You might be familiar with the dot product or the cross product, but please forget about this completely for the time being - nothing we're doing now has anything to do with dot products or cross products.
- We can't add elements of the field to elements of the vector space. Keep this in mind when working with the distributive property.

This definition is abstract, and it probably won't make sense until we've looked at a lot of examples. To this end, let's look at some examples.

Example 5.3
- Our favorite field is the real numbers $\mathbf{R}$, and our favorite vector space will be $V = \mathbf{R} \times \cdots \times \mathbf{R}$, which we write as $\mathbf{R}^n$. This is the set of n-tuples of real numbers: an element $v \in \mathbf{R}^n$ looks like $v = (x_1, \ldots, x_n)$. The vector space structure on this set is as follows:

  - Addition: if $v = (x_1, \ldots, x_n)$ and $w = (y_1, \ldots, y_n)$, then $v + w = (x_1 + y_1, \ldots, x_n + y_n)$.
  - Scalar multiplication: if $v = (x_1, \ldots, x_n) \in \mathbf{R}^n$, and $a \in \mathbf{R}$, then $a \cdot v = (ax_1, \ldots, ax_n)$ (element-wise scalar multiplication).

- Pictures: In section 5.1 is a visual demonstration of the commutativity of vector addition. You can draw similar pictures for the other properties.
- For any field k, then $k^n = k \times \cdots \times k$ is a vector space over k.
- As an example from quantum mechanics: in order to represent "observables" in quantum mechanics, we use vectors. If $\vec{a}$ and $\vec{b}$ are states of a quantum mechanical system, then $\frac{1}{\sqrt{2}}\vec{a} + \frac{1}{\sqrt{2}}\vec{b}$ is also a quantum state of the system. We include the factor of $1/\sqrt{2}$ in the addition so that the state is "normalized," which essentially means that the sum of probabilities of observing the system in a different states is 1.
- If our base field is $\mathbf{R}$, then the set of functions $\text{func}(\mathbf{R}, \mathbf{R}) = \{f : \mathbf{R} \to \mathbf{R}\}$ is a vector space. If f and g are functions, then $f + g : \mathbf{R} \to \mathbf{R}$ is defined to be the function $f + g : x \mapsto f(x) + g(x)$ (the $\mapsto$ means $(f + g)(x) = f(x) + g(x)$). If α is a scalar, then scalar multiplication by α is defined as $\alpha \cdot f : x \mapsto \alpha \cdot f(x)$.

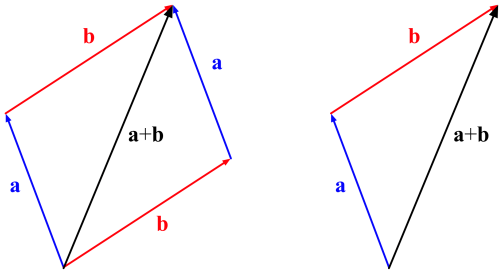


Figure 1: Demonstration of the commutativity of vector addition.

§5.2 Vector Spaces: Basic Properties

We've seen some definitions and examples pertaining to vector spaces. We want to understand vector spaces more, so now we'll talk about some basic properties that will be very useful later.

Proposition 5.4

Let V be a vector space over the field k, and let $0_V \in V$ be the additive identity of V, and $0_k \in k$ be the additive identity of the field k. then

(a) The additive identity 0_V in the vector space is unique (the proof is similar to the proofs so far in your homework).

(b) If we multiply the additive inverse of the identity in the field by an element of V, we get the additive inverse of the element, that is: $(-1) \cdot v = -v$.

(c) The zero element of the field times a vector is the zero element of the vector space, that is: $0_k \cdot v = 0_V$.

(d) Any scalar times the additive identity of the vector space is the additive identity of the vector space, that is: for any $\alpha \in k$, then $\alpha \cdot 0_V = 0_V$.

A very important warning: the additive identity of the field is *not* a vector - more generally, nothing in the field is a vector. Even if the vector space is just the field as a vector space over itself, we regard elements of the field and elements of the vector space as separate.

§5.3 Subspaces

Similarly to how we have subsets of sets, we can have sub*spaces* of vector spaces. Keep in mind though that subspaces and subsets are not the same thing. A subspace is always a subset, but not the other way around. In short, a subspace is a subset which is also a vector space.

Definition 5.5 — If V is a vector space over a field k, then a **subspace** W of V is a subset $W \subseteq V$ such that W is a vector space over k with the same addition and scalar multiplication as V.

Let's unpack this definition a little bit. The subset W is a vector space: it must be closed under addition - if $w_1, w_2 \in W$, then $w_1 + w_2$ must be in W. It must also be closed under scalar multiplication - if $\alpha \in k$ and $w \in W$, then $\alpha \cdot w \in W$. Continuing on this note, we prove a lemma, which will make it much easier for us to check when a subset is actually a subspace.

Proposition 5.6

If V is a vector space over a field k, and W is a sub*set* of V, then W is a sub*space* of V if and only if

(a) $0 \in W$

(b) For any $w, w' \in W$, then $w + w' \in W$.

(c) For any $\alpha \in k$ and $w \in W$, then $\alpha \cdot w \in W$.

Proof. First, if W is a subspace, then $+ : (w \times w) \subseteq W$, and $\cdot : (k \times w) \subseteq W$, since it is a vector space under the same addition and multiplication. This proves the first direction

for (b) and (c). To see (a), W must have an additive inverse, since $0_V = 0_k \cdot w$, for any $w \in W$, by the properties of W.

To prove the other direction, suppose that (2) and (3) hold. We endow W with the structure of a vector space, by using the $+$ and $\cdot$ from V. Then (2) and (3) tell us that the results of these operations land in W. We see that the endowed vector space structure on W comes from V. □

We can use the lemma to check that the following examples are indeed subspaces.

Example 5.7 The set $\{0\} \subseteq V$ is always a subspace of V.

The set $V \subseteq V$ is always a subspace of V.

Remark 5.8 (Structure vs. Property)**.** If X is a set, we can speak of the *structure* of a field on the set, or the *structure* of a vector space on the set, or any other object. That is, fields and vector spaces are sets, except we require them to have certain structure. If V is a vector space, and $W \subseteq V$, then W being a vector space is now a *property* of W.

Here the point is: we *impose* a structure on a set. We *check* that a set satisfies properties.

§6 September 19, 2019

§6.1 Linear Combinations and Span

We begin today with some definitions.

Definition 6.1 — Let V be a vector space over a field k. If $v_1, \dots, v_n$ are vectors in V, then $v \in V$ is a **linear combination** of $v_1, \dots, v_n$ if there exist scalars $\alpha_1, \dots, \alpha_n$, such that $v = \alpha_1 \cdot v_1 + \cdots + \alpha_n \cdot v_n$.

Definition 6.2 — If $v_1, \dots, v_n$ are vectors in V, then the **span** of $v_1, \dots, v_n$ is defined as

$$\operatorname{Span}(v_1, \dots, v_n) = \{v : v \text{ is a linear combination of } v_1, \dots, v_n\}.$$

That is, the span of a set of vectors is the set of all linear combinations of those vectors.

Lemma 6.3

If $v_1, \dots, v_n$ are vectors in V, then $\operatorname{Span}(v_1, \dots, v_n) \subseteq V$ is a sub*space* of V. Furthermore if $W \subseteq V$ is a subspace such that for all i we have $v_i \in W$, then $\operatorname{Span}(v_1, \dots, v_n) \subseteq W$ is a subspace of W.

Proof. We want to show that $\operatorname{Span}(v_1, \dots, v_n)$ is a subspace of V, so we need to show three things

(a) $0 \in$ span: This is true since $0 = \sum_i 0 \cdot v_i$.

(b) If $w, w' \in \operatorname{Span}(v_1, \dots, v_n)$, then $w + w' \in \operatorname{Span}(v_1, \dots, v_n)$: We have $w = \sum \alpha_i v_i$, and $w' = \sum \beta_i v_i$, then $w + w' = \sum(\alpha_i + \beta_i)v_i$.

(c) If $\alpha \in k, w \in \text{Span}(v_1, \ldots, v_n)$, then $\alpha \cdot w \in \text{Span}(v_1, \ldots, v_n)$: Write $w = \sum \alpha_i v_i$, then $\alpha\left(\sum \alpha_i v_i\right) = \sum \alpha\alpha_i v_i$. To prove the part after "furthermore," we first prove that the span is a sub*set* of W. If $w = \sum \alpha_i v_i$, then we know that for all i, $\alpha_i v_i \in W$, since W is a subspace and therefore closed under scalar multiplication. Then $\sum \alpha_i v_i \in W$, because W is closed under addition.

□

Remark 6.4. The span of $v_1, \ldots, v_n \in V$ is the smallest subspace of V containing all the v_i's.

Definition 6.5 — If U and W are subspaces of V, then the **sum** of U and W is

$$U + W = \{u + w : u \in U, w \in W\} \subseteq V.$$

Proposition 6.6

The sum $U + W$ is in fact a vector space. The proof is left as an exercise to the reader (in your pset).

Lemma 6.7

The spans of vectors "add." That is, using the vector space sum from above, then $\text{span}(v_1) + \cdots + \text{span}(v_n) = \text{span}(v_1, \ldots, v_n)$.

Proof. First, note that $\text{Span}(v_1) + \cdots + \text{Span}(v_n) \subseteq \text{Span}(v_1, \ldots, v_n)$, since $\alpha_1 v_1 + \cdots + \alpha_n v_n \in \text{Span}(v_1, \ldots, v_n)$, by definition of Span. Next, we have $\text{Span}(v_1, \ldots, v_n) \subseteq \text{Span}(v_1) + \cdots + \text{Span}(v_n)$, by the previous lemma and proposition - the proposition tells us that the right hand side contains all v_i. □

Definition 6.8 — If $W \subseteq V$ is a subspace, then we say $v_1, \ldots, v_n$ **spans** W if $\text{Span}(v_1, \ldots, v_n) = W$.

Definition 6.9 — We say that a vector space V is **finite dimensional** if there exists $v_1, \ldots, v_n \in V$ such that $\text{Span}(v_1, \ldots, v_n) = V$. If not, we say that V is infinite dimensional. Note, this definition only covers what it means to be finite dimensional, and it doesn't say anything about what it means to have dimension n - we'll get to that later.

We've gone over a lot of abstract definitions, so let's look at some examples.

Example 6.10 • $\mathbf{R}^n$ is finite dimensional over $\mathbf{R}$. To see this, we need to find a finite list of vectors which space $\mathbf{R}^n$. We can use the set of vectors $e_1, \ldots, e_n$,

defined by

$$\begin{aligned} e_1 &= (1, 0, \ldots, 0) \\ e_2 &= (0, 1, \ldots, 0) \\ &\vdots \\ e_n &= (0, 0, \ldots, 1). \end{aligned}$$

We can write $v = (\alpha_1, \ldots, \alpha_n)$ as

$$(\alpha_1, \ldots, \alpha_n) = \alpha_1(1, 0, \ldots, 0) + \alpha_2(0, 1, \ldots, 0) + \cdots + \alpha_n(0, 0, \ldots, 1).$$

The set $\{e_1, \ldots, e_n\}$ is called that **standard basis** for $\mathbf{R}^n$. Notice: if we add another vector to this set, it still spans $\mathbf{R}^n$ (we can just use 0 as the coefficient for it, and nothing changes).

- The set of polynomials of degree less than or equal to d for any d:

$$\text{Poly}_{\leq d}(k) = \{\alpha_d t^n + \alpha_{d-1} t^{d-1} + \cdots + \alpha_1 t + \alpha_0 : \alpha_i \in k\}.$$

This is just the set of polynomials with coefficients in k, with degree less than or equal to d. Polynomials are also functions: an element $p \in \text{Poly}_{\leq d}(k)$ is a function from $k \to k$, where we just evaluate the polynomial p at a point.

Proposition 6.11

The space $\text{Poly}_{\leq d}(k)$ is finite dimensional.

Proof. The polynomials of degree less than or equal d are equal to $\text{Span}(1, t, t^2, \ldots, t^d)$. □

Proposition 6.12

The vector space $\text{Poly}(k)$, (the polynomials without restriction on degree), is infinite dimensional.

Our goal going forward is to relate the notion of span to the notion of linear combination. To do this, we first define linear dependence.

Definition 6.13 — A list of vectors $v_1, \ldots, v_n$ is **linearly dependent** if $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ if there exist $\alpha_1, \ldots, \alpha_n \in k$, not all zero, such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$.

Example 6.14

In $\mathbf{R}^2$, the vectors $(0, 1)$ and $(0, 2)$ are linearly dependent: take $\alpha_1 = -2$ and $\alpha_2 = 1$, and then $-2(0, 1) + 1(0, 2) = 0$.

Definition 6.15 — If $v_1, \dots, v_n$ are *not* linearly dependent, then they are **linearly independent**. We can also define this as follows: $v_1, \dots, v_n$ are linearly independent if $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ implies $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

One remark about sets that we probably should have covered on the first day: sets do not have repeated elements. That is, $\{1, 1\}$ is the same thing as the set $\{1\}$. The "set" structure only keeps track of which elements are in the set.

Proposition 6.16

In k^n, the set $\{e_1, \dots, e_n\}$ is linearly independent.

Proof. If $\sum \alpha_i e_i = 0$, then $(\alpha_1, 0, \dots, 0) + (0, \alpha_2, \dots, 0) + \cdots + (0, 0, \dots, \alpha_n) = (0, 0, \dots, 0)$. Since addition is elementwise, then $\alpha_1 = 0$, then $\alpha_2 = 0$, all the way up to $\alpha_n = 0$. □

A helpful note on proving linear independence: (almost) every proof about linear independence starts with "take α_i's such that $\sum \alpha_i v_i = 0$." You should take a minute to go back over the definitions of linear independence, linear dependence, span, linear combination, and make sure you are very clear on the differences between them, and how you go about writing proofs using these. We now relate the notion of span and linear independence, and we will arrive at the notion of a basis.

Idea: let V be a vector space, and suppose that it is finite dimensional. Then it has a spanning set $v_1, \dots, v_n$, but we might be able to eliminate some of these vectors, and still have a spanning set - that is, we can have "too many" vectors in this set. The idea of "too many" vectors is the notion of linear dependence. What we want is a spanning set which is also linearly *independent*.

Example 6.17

In $\mathrm{Poly}_{\leq d}(k)$, then $\{1, \dots, t^d, t^{d+1}\}$ is a spanning set, but we can delete the t^{d+1}, and we still have a spanning set.

If we have a spanning set, we would like to be able to delete vectors. That is, we would like to be able to "reduce" our spanning set to a smaller size. To do so, we need the following lemma.

Lemma 6.18

Let $v_1, \dots, v_n$ be linearly *dependent*. Then there exists $j \in \{2, \dots, n\}$ such that

1. $v_j \in \mathrm{Span}(v_1, \dots, v_{j-1})$, and
2. $\mathrm{Span}(v_1, \dots, \hat{v}_j, \dots, v_n) = \mathrm{Span}(v_1, \dots, v_n)$. The $\hat{v}_j$ means that we remove the vector v_j from the set.

That is, we can remove v_j without changing the span of the set.

§7 September 24, 2019

§7.1 Span and Linear Independence

Recall the definition of linear (in)dependence from last time:

Definition 7.1 — A collection of vectors $\{v_1, \ldots, v_n\}$ is said to be **linearly dependent** if there exist $\alpha_1, \ldots, \alpha_n$ such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$.

A collection of vectors $\{v_1, \ldots, v_n\}$ is said to be **linearly independent** if it is not linearly dependent - that is, if we have $\alpha_1, \ldots, \alpha_n \in k$ such that

$$\sum_{i=1}^{n} \alpha_i v_i = 0$$

implies $\alpha_1 = \cdots = \alpha_n = 0$.

Now, we will prove some results. We use a new proof technique, **proof by algorithm**. This will be useful for many of our proofs throughout the course.

Lemma 7.2 (Reduction)

Suppose that $v_1, \ldots, v_n$ are linearly dependent. Then there exists $j \in \{2, \ldots, n\}$, such that

1. $v_j \in \text{Span}(v_1, \ldots, vj-1)$
2. $\text{Span}(v_1, \ldots, v_{j-1}, \hat{v}_j, v_{j+1}, \ldots, v_n) = \text{Span}(v_1, \ldots, v_n)$. Here, the $\hat{v}_j$ notation means that the v_j vector is deleted from the list.

The point of this lemma is that if we have a spanning set which is linearly dependent, then we can delete some vector(s), and it will still be a spanning set.

Proof. Since $v_1, \ldots, v_n$ are linearly dependent, then we can find $\alpha_1, \ldots, \alpha_n$, not all zero, such that

$$\sum_{i=1}^{n} \alpha_i v_i = 0.$$

That is just the definition of linear dependence.

1. Now, let j be the largest index such that $\alpha_j \neq 0$ (this exists because we know not all the α_i are zero). We want to show that $v_j \in \text{Span}(v_1, \ldots, v_{j-1})$. Since j is the *largest* index such that $\alpha_j \neq 0$, then

   $$\alpha_1 v_1 + \cdots + \alpha_{j-1} v_{j-1} + \alpha_j v_j = 0.$$

   We can rearrange this to get

   $$v_j = \frac{1}{\alpha_j}(-\alpha_1 v_1) + \cdots + \frac{1}{\alpha_j}(-\alpha_{j-1} v_{j-1}).$$

   We have written v_j as a linear combination of $v_1, \ldots, v_{j-1}$, so we are done proving the first part of the lemma. Note: we needed α_j to be nonzero so we could divide by α_j.

2. First, we note that $\text{Span}(v_1, \ldots, \hat{v}_j, \ldots, v_n) \subseteq \text{Span}(v_1, \ldots, v_n)$. We need to show that $\text{Span}(v_1, \ldots, \hat{v}_j, \ldots, v_n) \supseteq \text{Span}(v_1, \ldots, v_n)$. Let $v \in \text{Span}(v_1, \ldots, v_n)$, and write

   $$v = \sum_{i=1}^{n} \beta_i v_i.$$

Using our expression from part 1 above, we can write this as

$$v = \beta_1 v_1 + \cdots + \beta_{j-1} v_{j-1} + \beta_j \left(\frac{1}{\alpha_j}(-\alpha_1 v_1) + \cdots + \frac{1}{\alpha_j}(-\alpha_{j-1} v_{j-1}) \right) + \cdots + \beta_n v_n.$$

Since we have written v as a linear combination of vectors in $\{v_1, \ldots, \hat{v}_j, \ldots, v_n\}$, then $v \in \text{Span}(v_1, \ldots, \hat{v}_j, \ldots, v_n)$. Since arbitrary $v \in \text{Span}(v_1, \ldots, v_n)$ is in $\text{Span}(v_1, \ldots, \hat{v}_j, \ldots, v_n)$, then $\text{Span}(v_1, \ldots, v_n) \subseteq \text{Span}(v_1, \ldots, \hat{v}_j, \ldots, v_n)$, and therefore the two sets are equal.

□

Now, another important theorem, called the Steinitz exchange lemma. This will be very useful throughout the course.

Theorem 7.3 (Steinitz Exchange Lemma)
Let V be a finite dimensional vector space over a field k. Suppose that $\{u_1, \ldots, u_m\}$ is a linearly independent set and $\{v_1, \ldots, v_n\}$ spans V. Then $m \leq n$. That is, linearly independent sets are smaller than (or the same size as) spanning sets.

Proof. If $S_1 = \{u_1, v_1, \ldots, v_n\}$, then either $u_1 \in \{v_1, \ldots, v_n\}$ or S_1 is linearly dependent, since $u_1 = \sum_{i=1}^{n} \alpha_i v_i$. Then, apply the reduction lemma, so there exists some v_{j_1} so that $\text{Span}(u_1, v_1, \ldots, v_n) = \text{Span}(u_1, v_1, \ldots, \hat{v}_{j_1}, \ldots, v_n)$. Then throw v_{j_1} out of the set. Now, we have a new linearly independent set $\{u_2, \ldots, u_m\}$, and a new spanning set $\{u_1, v_1, \ldots, \hat{v}_{j_1}, \ldots, v_n\}$, that is, we've moved u_1 to the spanning set and deleted something from the spanning set. Continue this process by moving u_2, etc. Note that at each step, the spanning set stays the same size, and the linearly independent set decreases in size by one. We can never delete any of the u_i from the spanning set (since they are linearly dependent). This process ends when all the u_i are gone. Now, $m \leq n$, since otherwise we will run out of v_j's before u_i, and then there is some proper subset of $\{u_1, \ldots, u_m\}$ which spans the space. But that can't happen, since $\{u_1, \ldots, u_m\}$ are linearly independent. □

The main point of this is that the number of linearly independent vectors is less than or equal the number of spanning vectors. When these numbers are equal, we get something special, which we call a basis.

§7.2 Bases and Dimension

Definition 7.4 — Let V be a vector space over k. A collection of vectors $\{v_1, \ldots, v_n\}$ is called a **basis** if it is both linearly independent and spanning.

Informally, you can think of a basis as a choice of coordinates (but try to avoid this thought).

Example 7.5

- The set $\{(1,0,0), (0,1,0), (0,0,1)\}$ is a basis for $\mathbf{R}^3$. There are many bases for $\mathbf{R}^3$. The fact that $\mathbf{R}^n$ has very many bases is what makes matrix based linear algebra very difficult. In this course, try to avoid choosing bases as much as possible.
- If $V = k^n$, then $\{e_1, \ldots, e_n\}$ is a basis, where the e_i are the standard basis vectors.

- If $V = \text{Poly}_{\leq d}(k)$, then a basis is $\{1, x, \ldots, x^d\}$.

Now, we will relate "finite dimensional" to bases in the way that you would expect.

Theorem 7.6 1. Every nonzero finite dimensional vector space V has a basis of finite length.

2. If $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_m\}$ are two bases for V, then $m = n$.

Proof. 1. Since V is finite dimensional, then $V = \text{Span}(w_1, \ldots, w_r)$, for some $w_1, \ldots, w_r$. Then apply the reduction lemma until the w_i are linearly independent.

2. Since $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_m\}$ are bases, then they are both linearly independent and spanning. Since $\{u_1, \ldots, u_n\}$ is linearly independent and $\{v_1, \ldots, v_m\}$ is a spanning set, then $n \leq m$, by the exchange lemma. Since $\{v_1, \ldots, v_m\}$ is linearly independent and $\{u_1, \ldots, u_n\}$ is spanning, then $m \leq n$, again by the exchange lemma. Since $n \leq m$ and $m \leq n$, then $n = m$.

□

Now, we are finally in a position to define the dimension of a vector space.

Definition 7.7 — A vector space V has **dimension n** if there exists a basis $\{v_1, \ldots, v_n\}$ for V of length n.

Proposition 7.8

A vector space V is of dimension n if and only if all bases are of length n.

Proof. Any two bases have the same length. This tells us that the notion of "dimension" is well defined. □

NB: We say that something is "basis independent" if it does not depend on any choice of basis. For example, "dimension" is basis independent, while "coordinates" for a given vector are not basis independent. In general, basis independent notions are much nicer to work with.

Example 7.9

- $\dim(k^n) = n$.
- $\dim(\text{Poly}_{\leq d}(k)) = d + 1$.
- $\dim(k^S) = \#S$, if S is a finite set.

Proposition 7.10

If $W \subseteq V$ is a subspace of V, then $\dim W \leq \dim V$.

Proof. The basis for V also spans W, so its dimension is at most that of V. □

Lemma 7.11 (Extension)
If $v_1, \dots, v_n \in V$ are linearly independent, then we can extend the set $\{v_1, \dots, v_n\}$ to a basis of V

Proof. Let $\{w_1, \dots, w_m\}$ be a spanning set for V. We can extend $\{v_1, \dots, v_n, w_1, \dots, w_m\}$, which spans V. Then use the reduction lemma to make the list linearly independent. Notice that we will only remove the w's, since the $v_1, \dots v_n$ are linearly independent. You should look back at the proof of the reduction lemma to make sure you are clear on all the details of why this works. □

To recap everything from today: say we have a finite dimensional vector space V. Since V is finite dimensional, $V = \text{Span}(v_1, \dots, v_m)$ for some $v_1, \dots, v_m$. If it's linearly dependent, toss out some vectors to get a basis. Also, if you have a linearly independent set, you can always extend it to spanning set. We can always to the following:

$$\{\text{spanning set}\} \xrightarrow{\text{reduce}} \{\text{linearly independent set}\}$$
$$\{\text{linearly independent set}\} \xrightarrow{\text{extend}} \{\text{spanning set}\}.$$

§8 October 1, 2019

§8.1 Linear Transformations

We begin today with the definition of a linear transformation, which is the most important type of function in linear algebra.

Definition 8.1 — A **linear transformation** or **linear map** between two vector spaces V and W both over a field k is a function $f : V \to W$ such that

- for all $v, w \in V$, then $f(v + w) = f(v) + f(w)$, and
- for all $\alpha \in k$ and $v \in V$, $f(\alpha v) = \alpha f(v)$.

Example 8.2
- The map $\mathbf{R} \xrightarrow{\iota} \mathbf{C}$, the inclusion map, which takes $r \mapsto r$ is an $\mathbf{R}$-linear transformation.
- For any vector space V, the identity map $V \xrightarrow{\text{id}} V$, defined as $v \mapsto v$ for all v, is a linear transformation.
- The map $v \mapsto 0$ for all v is a linear transformation.
- The map $m_\alpha : V \to V$, defined by $m_\alpha : v \mapsto \alpha \cdot v$ (familiar from the last problem set), is a linear transformation. To verify this, note that $m_\alpha(v+w) = \alpha \cdot (v+w) = \alpha v + \alpha w = m_\alpha(v) + m_\alpha(w)$, and also for $\beta \in k$, then $m_\alpha(\beta v) = \alpha \cdot (\beta v) = \beta \alpha v = \beta m_\alpha(v)$.
- If we work in the polynomial space $\frac{d}{dt} : \text{Poly}_{\leq d}(k) \to \text{Poly}_{\leq d-1}(k)$, the derivative map, is a linear transformation. Recall the definition of a derivative on

polynomials:

$$\frac{d}{dt}\left(\alpha_n t^n + \alpha_{n-1}t^{n-1} + \cdots + \alpha_0\right) = n\alpha_n t^{n-1} + (n-1)\alpha_{n-1}t^{n-2} + \cdots + \alpha_1 + 0.$$

Also recall that the derivative is linear: for any (differentiable) functions, we have $\frac{d}{dt}(p(t)+q(t)) = \frac{d}{dt}p(t) + \frac{d}{dt}q(t)$ and $\frac{d}{dt}(\alpha p(t)) = \alpha\frac{d}{dt}p(t)$.

- You might be used to seeing linear transformations defined as matrices. You should forget about this, and try to avoid using matrices as much as you can. We'll come back to matrices later.

Now for some basic properties of linear transformations, which will be very important throughout the rest of the course.

Proposition 8.3

If f is a linear map, then $f(0) = 0$, and $f(-v) = -f(v)$ for all $v \in V$. The proof of this proposition is in your homework.

Definition 8.4 — If V, W are vector spaces, then we denote the set of all linear transformations from V to W as $\mathcal{L}(V,W)$. That is,

$$\mathcal{L}(V,W) = \{f : V \to W : f \text{ is linear}\}.$$

Proposition 8.5

The set $\mathcal{L}(V,W)$ is a vector space with the following operations:

$$\begin{aligned} &+ : \mathcal{L}(V,W) \times \mathcal{L}(V,W) \to \mathcal{L}(V,W) \\ &+ : (f,g) \mapsto f+g \\ &\cdot : k \times \mathcal{L}(V,W) \to \mathcal{L}(V,W) \\ &\cdot : (\alpha, f) \mapsto \alpha f. \end{aligned}$$

Recall that $f+g$ is the function $f+g : v \mapsto f(v)+g(v)$, and αf is the function $\alpha f : v \mapsto \alpha f(v)$.

Proof. For addition, we know that f and g are linear, and we need to show that $f+g$ is as well. By definition,

$$\begin{aligned} (f+g)(v+v') &= f(v+v') + g(v+v') \\ &= f(v) + f(v') + g(v) + g(v') \\ &= f(v) + g(v) + f(v') + g(v') \\ &= (f+g)(v) + (f+g)(v'). \end{aligned}$$

Also, we have that

$$\begin{aligned} (\alpha f)(v+v') &= \alpha(f(v+v')) \\ &= \alpha(f(v) + f(v')) \\ &= \alpha f(v) + \alpha f(v'). \end{aligned}$$

Since $(f+g)(v+v') = (f+g)(v) + (f+g)(v')$ and $(\alpha f)(v+v') = (\alpha f)(v) + (\alpha f)(v')$, then the map is linear. The additive identity in this vector space is $0 : V \to W$ defined as $0 : v \mapsto 0$ for all $v \in V$. The rest of the proof is left as an exercise to the reader. □

In addition to having a vector space of linear transformations, we can also compose linear transformations. Let's say that U is some other vector space, in addition to V and W. Then we have a map $\mathcal{L}(V,W) \times \mathcal{L}(W,U) \to \mathcal{L}(V,U)$ defined as

$$(f : V \to W, g : W \to U) \mapsto (g \circ f)V \to U.$$

In fact, this map is a linear transformation between $\mathcal{L}(V,W)$ and $\mathcal{L}(W,U)$:

Lemma 8.6

If $f : V \to W$ and $g : W \to U$ are linear, then so is $(g \circ f) : V \to U$.

Proof. If $v, v' \in V$, then

$$\begin{aligned}(g \circ f)(v+v') &= g(f(v+v')) \\ &= g(f(v) + f(v')) \\ &= g(f(v)) + g(f(v')),\end{aligned}$$

where the first second equality is by linearity of f and the third is by linearity of g. We leave the proof for the scalar multiplication part to you. □

Now, let's suppose that $U = V = W$. Then using the above, we have $\mathcal{L}(V,V) \times \mathcal{L}(V,V) \to \mathcal{L}(V,V)$. This case, where all the vector spaces are equal, is probably the case that we care about most. This function defined a multiplication (*not* scalar multiplication) on the set $\mathcal{L}(V,V)$. In the special case in which V is a field, this gives $k \times k \to k$. Be careful though: not all the maps in $\mathcal{L}(V,V) \times \mathcal{L}(V,V) \to \mathcal{L}(V,V)$ have inverses, so the multiplication isn't quite the same as in a field.

Proposition 8.7

Let V, W, U be vector spaces, and define the maps $V \xrightarrow{f} W \xrightarrow{g} U$.

(a) Suppose that $Y \xrightarrow{h} V$ is a linear map. Then $(g \circ f) \circ h = g \circ (f \circ h)$. That is, composition is associative.

(b) If id is the identity map, then $g \circ \mathrm{id}_W = g = \mathrm{id}_U \circ g$.

(c) $(f_1 + f_2) \circ g = f_1 \circ g + f_2 \circ g$, and $g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2$.

Proof. The properties are inherited from function composition and linearity. Parts (a) and (b) are general properties of function composition, and (3) follows from linearity. □

Now, using this multiplication and addition on the set $\mathcal{L}(V,V)$, we *don't* get a field, but instead, we get a k-**algebra**, or an **algebra over the field** k. It's not a field since it doesn't have multiplicative inverses.

§8.2 Matrices (eww)

We will now discuss how to construct linear maps using **matrices**. We can ask the question: what sort of data specifies a linear map $f : V \to W$. That is, we want a "cooking recipe" for making linear transformations. First, let's choose a basis $\{v_1, \ldots, v_n\}$ for V. Then, we choose a collection of vectors $\{w_1, \ldots, w_n\}$ in W, (any collection, not necessarily a basis). Then we set $f(v_i) = w_i$. This completely determines a linear transformation f at every single vector in V, since if $v \in V$, then we can write $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$, and then $f(v) = f(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha f(v_1) + \cdots + \alpha_n f(v_n)$. Since we know $f(v_i)$ for each i, then we know $f(v)$. Notice that if the $\{v_i\}$ are not a basis, then f might not be well defined. If it is a basis , then this defines f uniquely. Now we have the following recipe for cooking up linear transformations:

1. Choose a basis, $\{v_1, \ldots, v_n\}$, for V.
2. Choose n vectors, $\{w_1, \ldots, w_n\}$ in W.
3. Let $f(v_i) = w_i$ for each i.
4. To find $f(v)$ for *any* v, just write $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$, and compute $\alpha_1 f(v_1) + \cdots + \alpha_n f(v_n)$
5. Congratulations, you have cooked up a tasty linear transformation!

Thus we see that the "data" we spoke of above is just: a basis for V, n elements of W, and a choice of w_i for each v_i. Now, we need to know that this actually works. To do this, we use the following lemma.

Lemma 8.8

The above procedure specifies the unique linear map that sends $v_i \mapsto w_i$. To be more precise, if $w_1, \ldots, w_m$ is a basis for W, then

$$\sum_{j=1}^{n} \alpha_{ji} w_j = \alpha_{1i} w_1 + \alpha_{2i} w_2 + \cdots + \alpha_{mi} w_m.$$

Now, we will see our first matrix:

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{pmatrix}.$$

We write this matrix as (a_{ij}). This matrix specifies a linear transformation as follows

$$\begin{aligned} f(v) &= f\left(\sum_{i=1}^{n} \beta_i v_i\right) \\ &= \sum_{i=1}^{n} \beta_i f(v_i) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_i \alpha_{ji} w_j \\ &= \sum_{j=1}^{m} \sum_{i=1}^{n} \beta_i \alpha_{ji} w_j. \end{aligned}$$

Let's do an example to see this in action.

Example 8.9

Let $\frac{d}{dt} : \text{Poly}_{\leq d}(k) \to \text{Poly}_{\leq d-1}(k)$ be the derivative map. We can choose a basis $\{1, t, \ldots, t^d\}$ for $\text{Poly}_{\leq d}(k)$, and also a basis $\{1, t, \ldots, t^{d-1}\}$ for $\text{Poly}_{\leq d-1}(k)$. We know that $\frac{d}{dt} : d^j \mapsto jt^{j-1}$. We have $\frac{d}{dt}(1) = 0$, so the first column of our matrix will be zero. Since $\frac{d}{dt} = 1 \cdot 1$, and $\frac{d}{dt}t^2 = 2t$, so the second column will have a 1 in the first slot and zeros in the other slots, and the third column will have a 2 in the second slot and zeros elsewhere, etc. We can write out our matrix as

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & n-1 \end{pmatrix}$$

As you may have notices, all these indices are really obnoxious, difficult to move around, and also very non-canonical. This is why we don't like matrices very much. In general, on your problem sets, try not to use matrices (unless you really have to). Instead, try to use fundamental properties of linear transformations - this will also make your LaTeXing much easier. Keep in mind the following quote:

> "It is my experience that proofs involving matrices can be shortened by 50% if one throws the matrices out." -Emil Artin, *Geometric Algebra* .

§9 October 3, 2019

§9.1 Kernel and Image

Today, we are going to work towards the rank-nullity theorem: $\dim V = \dim(\text{im} f) + \dim(\ker f)$. First, we need to understand the kernel, $\ker f$, and the image, $\text{im} f$.

Definition 9.1 — If $f : V \to W$ is a linear map, then the **image** (or *range*) of f is

$$\text{im}(f) = \{f(v) : v \in V\}.$$

Notice that the image is a subset of W (we'll prove that it's also a subspace). Colloquially, you can think of this as "the set of things that f maps to."

Definition 9.2 — If $f : V \to W$ is a linear map, then the **kernel** or **nullspace** of f is

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

Notice that the kernel is a subset of V (again, we'll prove that it's also a subspace). Colloquially, this is "the set of things that f sends to zero."

The kernel and image are extremely important in linear algebra, so you should make sure you are very comfortable with these definitions. In addition, the kernel and image are important for other objects outside of just vector spaces (such as groups, rings, fields, and almost everything else in algebra). Rank-nullity is one of the fundamental theorems in linear algebra, and is one of the most useful theorems in applications of linear algebra.

Example 9.3

Let $f_v : k \to V$ be the function $f : \alpha \mapsto \alpha \cdot v$. Then $\text{im}(f) = \text{Span}(v)$, since $\text{im}(f) = \{f(\alpha) = \alpha \cdot v\}$, and the set of $\{\alpha \cdot v\}$ is just the span of v (by definition).

Example 9.4

If $U \subseteq V$ is a subspace, then we can think of the subspace inclusion as a function $\iota : U \hookrightarrow V$. If $u \in U$, then $\iota : u \mapsto u$. This is an injective map. The image of this map is $U \subseteq V$ (since ι maps U to itself). The kernel of ι is zero: $\ker(\iota) = \{0\}$. Notice that ι is injective, and its kernel is zero (these two ideas are related, as we'll see later).

Now, as mentioned above, the kernel and the image are actually sub*spaces*, not just subsets. We will now prove this, so that we can apply our theory about subspaces to the kernel and image of a map.

Lemma 9.5

IF $f : V \to W$, then

(a) $\ker(f) \subseteq V$ is a subspace of V.

(b) $\text{im}(f) \subseteq W$ is a subspace of W.

Proof. (a) First, $0 \in \ker(f)$, since linear maps always send zero to zero. If $v, v' \in \ker(f)$, then $f(v+v') = f(v)+f(v') = 0+0 = 0$. That is, $v+v'$ is also in the kernel, so the space is closed under addition. Also, if $\alpha \in k$ and $v \in \ker(f)$, then $f(\alpha v) = \alpha f(v) = \alpha 0 = 0$, so αv is also in the kernel, and the kernel is closed under scalar multiplication and therefore a subspace.

(b) First $0 = f(0)$ for all linear maps, so $0 \in \text{im}(f)$. If $w = f(v)$ and $w' = f(v')$, then $w + w' = f(v) + f(v') = f(v + v')$, so $w + w'$ is also in the image of f. If $\alpha \in k$ and $w = f(v)$, then $\alpha w = \alpha f(v) = f(\alpha v)$, so αw is also in the image. Thus the image is closed under addition and scalar multiplication, and is therefore a subspace. □

Lemma 9.6

Let $f : V \to W$ be a linear map. Then

(a) f is surjective if and only if $\text{im}(f) = W$.

(b) f is injective if and only if $\ker(f) = \{0\}$.

Proof. (a) Left as an exercise to the reader. There isn't much to prove, just think about it for a while.

(b) Suppose that $f : V \to W$ is injective. We know that $f(0) = 0$. Since f is injective, then if $f(v) = 0$, then $f(v) = f(0)$ implies $v = 0$.

Now, suppose that $\ker(f) = \{0\}$. We want to show that f is injective. Suppose that $f(u) = f(v)$. We need to show that $u = v$. Subtracting $f(v)$ from both sides, then

we have $f(u) - f(v) = 0$, so $f(u - v) = 0$, which means that $u - v \in \ker(f)$. Since the kernel is zero, then $u - v = 0$, which means $u = v$.

□

We've see the definition of injective, surjective and bijective. We'll now go over the definition of an isomorphism, which is a bijective map, that *also* preserves the vector space structure. By "preserves vector space structure," we mean that the map is linear.

Definition 9.7 — An **isomorphism** between vector spaces V and W is a linear map which is bijective. If there exists an isomorphism between V and W, we say that V and W are **isomorphic**

Lemma 9.8

Suppose that $f : V \to W$ and $g : W \to V$ are functions such that $f \circ g = \mathrm{id}_W$ and $g \circ f = \mathrm{id}_V$. Then g is linear.

The main idea here is that function inverses can be "promoted" to linear maps.

Proof. First, we show that $f(g(w + w')) = f(g(w) + g(w'))$. We don't know that g is linear yet, so this isn't obvious. It follows since $f(g(w + w')) = w + w'$, since f and g are inverses. Now, we know that $f(g(w) + g(w')) = f(g(w)) + f(g(w')) = w + w'$, since f is linear, and f and g are inverses. Thus, we know that $f(g(w + w')) = f(g(w) + g(w'))$.

We now need to show that $g(w + w') = g(w) + g(w')$. Let's look at $g(w + w')$. This is equal to $g((f \circ g)(w + w'))$, since $f \circ g$ is the identity. But then this is $g(f(g(w) + g(w'))$, from what we showed above. But we can write this as $(g \circ f)(g(w) + g(w'))$, which is $g(w) + g(w')$, since $g \circ f$ is the identity. In summary, we have

$$\begin{aligned} g(w + w') &= g((f \circ g)(w + w')) \\ &= g(f(g(w) + g(w')) \\ &= g(w) + g(w'). \end{aligned}$$

You also need to show that scalar multiplication works with g. This is left as an exercise to the reader. □

Now, we can state a few equivalent definitions of an isomorphism.

Proposition 9.9

If $f : V \to W$, then the following are equivalent:

(a) f is an isomorphism

(b) $\mathrm{im}(f) = W$ and $\ker(f) = 0$.

(c) There exists an inverse function for f.

Everything you need for the proof was shown above, and you can work out the details on your own.

We will now deviate from Axler, so you should pay attention.

§9.2 Exact Sequences

Definition 9.10 — Suppose that V, V', V'' are vector spaces. Then a sequence of linear maps (note the order)

$$V' \xrightarrow{f} V \xrightarrow{g} V''$$

is called **exact** if $\operatorname{im}(f) = \ker(g)$.

Example 9.11

The sequence $0 \to V \xrightarrow{f} W$ is exact if and only if f is injective. The sequence is exact if the image of $0 \to V$ is the kernel of f. But the image of $0 \to V$ is just 0, and we know that f is injective if and only if the kernel of f is zero.

Example 9.12

The sequence $V \xrightarrow{g} W \to 0$ is exact if and only if g is surjective. The sequence is exact if the image of g is the kernel of $W \to 0$. But the kernel of W is all of W, since $W \to 0$ sends everything to zero. So g is surjective if and only if the image of g is W which is the kernel of $W \to 0$.

In order to give more examples of exact sequences, we need quotients.

Definition 9.13 — Suppose that $U \subseteq V$ is a subspace. We define V/U to be the equivalence classes of the following equivalence relation on V: $v \sim w$ if and only if $v - w \in U$.

Definition 9.14 — We define $v + U = \{v + u : u \in U\}$.

Lemma 9.15

The class of v, $C(v)$, under this equivalence relation, is equal to $v + U$.

Proof. If $v' \sim v$, then by definition $v' - v = u$ for some $u \in U$. This means that $v' = v + u$, so $v' \in v + U$, so $C(v) \subseteq v + U$. Now, if $x \in v + U$, then $x = v + u$ for some $u \in U$. But then $x - v = u \in U$, so $x \sim v$, so $v + U \subseteq C(v)$, and thus $C(v) = v + U$. □

Geometrically, in $\mathbf{R}^2$, we can think of the subspace U as a line through the origin, the red line, and $v + U$ as a translation of U, the blue line. The quotient V/U is the set of all the translates of U, which is the set of all blue lines.

Lemma 9.16

The quotient V/U has a unique vector space structure such that the function $V \to V/U$ is linear. Think of this as an analogy to $\mathbf{Z} \to \mathbf{Z}/n$ from before (although remember that these aren't vector spaces).

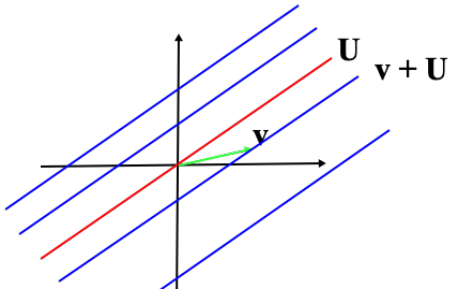


Figure 2: Geometric visualization of quotient spaces

Proof. We have $(v+U)+(v'+U) = (v+v'+U)$, and $\alpha(v+U) = \alpha v+U$, as definitions. As an exercise, you should check that these definitions are well defined. Now, $f : V \to V/U$ defined by $v \mapsto v+U$ and $v+v' \mapsto v+v'+U$ is linear. You should fill in the details of this proof on your own. (It's similar to what we did with $\mathbf{Z}$ and $\mathbf{Z}/n$.) □

Lemma 9.17

If U is a subspace of V, then the following sequences are exact:

$$0 \to U \to V$$
$$U \to V \to V/U$$
$$V \to V/ \to 0.$$

In this case, we write $0 \to U \to V \to V/U \to 0$.

By analogy, we can also write the exact sequence (although not for vector spaces) $0 \to \mathbf{Z} \xrightarrow{\cdot n} \mathbf{Z} \to \mathbf{Z}/(n) \to 0$.

Next week, we'll discuss the rank nullity theorem, in its more general form:

Theorem 9.18 (Euler Exactness)

If $0 \to V' \to V \to V'' \to 0$ is exact, then $\dim(V') + \dim(V'') = \dim(V)$.

If you want a fun example of exact sequences that you can write down to show off your math skills to your friends, read about the **snake lemma**, here: Wikipedia Snake Lemma. It looks something like this:

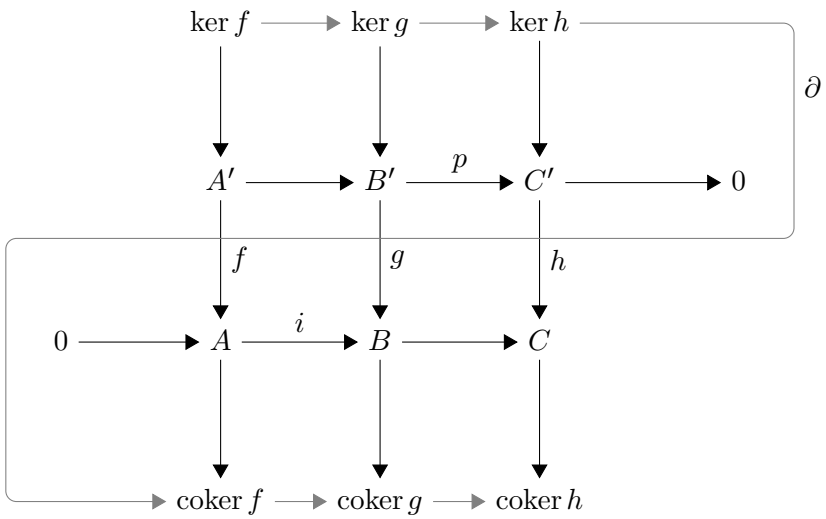


Figure 3: The snake lemma

§10 October 8, 2019

§10.1 Rank Nullity and More Exact Sequences

Before we begin, let's remember the definition of an exact sequence. A sequence of maps $V_1 \to V_2 \to \cdots \to V_n$ is called **exact** if the image of each map is the kernel of the next map. That is, $\operatorname{im}(V_1 \to V_2) = \ker(V_2 \to V_3)$, etc. Recall the following theorem:

Theorem 10.1

If $0 \to V' \to V \to V'' \to 0$ is an exact sequence of vector spaces, then $\dim V = \dim V' + \dim V''$.

Proof. We can identify V' with a subspace of V, since the map $V' \to V$ is injective (by exactness). Now, choose a basis $v_1', \ldots, v_n'$ for V'. By the extension lemma, we can extend $f(v_1'), \ldots, f(v_n')$ to a basis of V, which we call $f(v_1'), \ldots, f(v_n'), w_1, \ldots, w_m$. The $f(v_i')$ are linearly independent since the map f is injective (remember from your problem set), which tells us that $\dim V' + m = \dim V$. In order to prove the thoerem, we need to show that $m = \dim V''$. To do this, it suffices to show that $\{g(w_1), \ldots, g(w_m)\}$ is a basis for V''. That is, we need to show that $\{g(w_1), \ldots, g(w_m)\}$ spans and is linearly independent.

By exactness of the sequence, $\operatorname{im}(V \to V'') = \ker(V'' \to 0)$. This means that the map $V \to V''$ is surjective onto V''. We claim that $\operatorname{Span}(g(w_1), \ldots, g(w_m)) = V''$. We need to see why this is true. For any $v'' \in V''$, we can find a $v \in V$ such that $g(v) = v''$. Using the basis for V, there exist $\alpha_1, \ldots, \alpha_{n+m}$ such that

$$g(\alpha_1 v_1' + \alpha_2 v_2' + \cdots + \alpha_n v_n' + \alpha_{n+1} w_1 + \cdots + \alpha_{n+m} w_m) = v''.$$

That is, we know that $\operatorname{Span}(g(v_1'), \ldots, g(v_n'), g(w_1), \ldots, g(w_m)) = V''$. By exactness, all of the $\alpha_i v_i'$ are in the kernel of g, this expression is equal to

$$0 + \cdots + 0 + \alpha_{n+1} g(w_1) + \cdots + \alpha_{m+n} g(w_m).$$

That is, anything in V'' is in the span of $g(w_1), \dots, g(w_m)$, so $\text{Span}(g(w_1), \dots, g(w_m)) = V''$.

We still need to see that $\{g(w_1), \dots, g(w_m)\}$ is linearly independent. To do this, suppose that there exist $\alpha_1, \dots, \alpha_m$ such tat $\alpha_1 g(w_1) + \cdots + \alpha_m g(w_m) = 0$. We need to show that $\alpha_i = 0$ for all i. By linearity, we can write this as

$$g\left(\sum_{i=1}^{m} \alpha_i w_i\right) = 0.$$

This tells us that $\sum \alpha_i w_i \in \ker(g)$. Since the image of $V' \to V$ is equal to the kernel of $V \to V''$, then we can write anything in the kernel of $V \to V''$ as a linear combination of vectors in the image of f. Thus we can write

$$\sum_{j=1}^{n} \beta_j f(v_j') = \sum_{i=1}^{m} \alpha_i w_i.$$

To see why this implies the result, recall the definition of $w_1, \dots, w_m$ as the extension of the $f(v_1'), \dots, f(v_n')$, which means that the $w_1, \dots, w_m$ are linearly independent to $f(v_1'), \dots, f(v_n')$. We can move around the above sum to get

$$\beta_1 f(v_1') + \cdots + \beta_n f(v_n') - \alpha_1 w_1 - \cdots - \alpha_m w_m = 0.$$

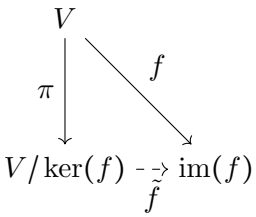
Since the $w_1, \dots, w_m, f(v_1'), \dots, f(v_n')$ are linearly independent, then each of the α_i, β_j must be zero. Thus, we have $\alpha_1 g(w_1) + \cdots + \alpha_m g(w_m) = 0$ implies $\alpha_i = 0$ for all i, which means that $\{g(w_1), \dots, g(w_m)\}$ is a linearly independent set.

Since $\{g(w_1), \dots, g(w_m)\}$ is linearly independent and spans V'', it is a basis for V''. Thus, we know that $\dim V' + m = \dim V$ from above, as desired. □

Now, in order to prove the rank nullity theorem, we will need the following lemma, which is also known as the first isomorphism theorem:

Lemma 10.2 (First Isomorphism Theorem)
If $f : V \to W$ is a linear map, then $\text{im}(f) \simeq V/\ker(f)$.

Proof. We construct a linear map $\tilde{f}$ from $V/\ker(f)$ to $\text{im}(f)$, defined by $\tilde{f} : v + \ker(f) \mapsto f(v)$. This of this map using the following diagram:



This map is well defined since $v + \ker(f) = w + \ker(f)$ if and only if $v - w \in \ker(f)$, by definition of quotient spaces. Then $f(v) = f(w)$ if and only if $f(v - w) = 0$, if and only if $v - w \in \ker(f)$, so the map is well defined. This maps is linear since f is linear. Why is is surjective? Since $V \to \text{im}(f)$ is surjective, and since $\pi : V \to V/\ker(f)$ is surjective, then $\tilde{f}$ is also surjective, since $\tilde{f} \circ \pi = f$. The map $\tilde{f}$ is injective since if $\tilde{f}(v + \ker(f)) = 0$, then $f(v) = 0$ implies $v \in \ker(f)$. Then $v + \ker(f) = 0 + \ker(f)$, which is zero in $V/\ker(f)$. That is, the kernel of $\tilde{f}$ is zero, so it is injective. □

Corollary 10.3

If $f : V \to W$ is a linear map, then $\dim V = \dim \operatorname{im}(f) + \dim \ker(f)$.

Proof. Let's write the exact sequence $0 \to \ker(f) \to V \to V/\ker(f) \to 0$. Now, by the theorem 10.1, we know $\dim V = \dim \ker(f) + \dim V/\ker(f)$, and by lemma 10.2, this is equal to $\dim \ker(f) + \dim \operatorname{im}(f)$, since $\operatorname{im}(f) \simeq V/\ker(f)$. Thus, we have $\dim V = \dim \ker(f) + \dim \operatorname{im}(f)$, as desired. □

Now that we've proven rank nullity, let's go over some examples to see how useful it is

Example 10.4

- Suppose that we have a system of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0. \end{aligned}$$

  We know that$x_1 = x_2 = \cdots = x_n = 0$ is a solution to this, but we want to know if there are other solutions. Using rank nullity, we can prove that if there are more variables that equations, then there exist non-trivial solutions. To do this, let's suppose that $f : V \to W$ is such that $\dim V > \dim W$. Then $\ker f$ contains nonzero elements. By rank nullity, then $\dim V = \dim \operatorname{im}(f) + \dim \ker(f)$, which we can rewrite as $\dim \ker(f) = \dim V - \dim \operatorname{im}(f) \geq \dim V - \dim W$, since $\dim \operatorname{im}(f) \leq \dim(W)$. By assumption, $\dim V - \dim W > 0$, so we have $\dim \ker(f) > 0$. Thinking of this map as a system of linear equations, then $\dim V > \dim W$ means that there are more variables than equations, and $\dim \ker(f) > 0$ means that there is something in the kernel aside from zero: that is, there is a solution to the system of equations other that $x_1 = \cdots = x_n = 0$.

- Let's define direct sums. Suppose that $U_1, U_2 \subseteq V$. We've seen the sum of subspaces, defined as $U_1 + U_2 := \{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}$. Now, we define the **direct sum** as $U_1 \oplus U_2$ as the vector space with the underlying set $U_1 \times U_2$ (the Cartesian product). That is, we write $(u_1, u_2) \in U_1 \oplus U_2$, and addition is defined pointwise. We then have the following exact sequence:

$$0 \to \ker(a) \to U_1 \oplus U_2 \xrightarrow{a} U_1 + U_2 \to 0,$$

  where a is the map $U_1 \oplus U_2 \to U_1 + U_2$, defined by $(u_1, u_2) \to u_1 + u_2$ The following proposition will be useful for working with direct sums.

Proposition 10.5

The following are equivalent:

(a) $U_1 + U_2 \simeq U_1 \oplus U_2$

(b) $\dim U_1 + \dim U_2 = \dim(U_1 + U_2)$

(c) $\ker(a) = 0$, where a is the map $U_1 \oplus U_2 \to U_1 + U_2$, defined by $(u_1, u_2) \to u_1 + u_2$.

Proof. Left as an exercise to the reader. □

In summary, we have classified all finite dimensional vector spaces. We can build larger vector spaces from one dimensional vector spaces by taking a direct sum n times. That is, we start with a field k, and do $k \oplus \cdot \oplus k$, adding n times. In addition, every exact sequence splits. That is, every exact sequence is isomorphic to

$$0 \to V' \to V' \oplus V'' \to V'' \to 0.$$

The good news is that vector spaces are easy: any vector space is completely determined by its dimension, and all vector spaces of the same dimension are isomorphic. The bad news is that linear transformations are more difficult, and we'll be studying these for the rest of the course.

October 10, 2019

Suppose we have an exact sequence $0 \to V' \to V \to V'' \to 0$. Last time, we showed that $\dim V = \dim V' + \dim V''$. This tells us that if $f : V \to W$ is a map, then $\dim V = \dim \operatorname{im}(f) + \dim \ker(f)$, since the sequence $0 \to \ker f \to V \to \operatorname{im} f \to 0$ is exact, and $\operatorname{im} f \simeq V/\ker f$. We also know that any two vector spaces are isomorphic if and only if they have the same dimension. That is, we classified vector spaces completely. Now, we would like classify linear transformations.

§10.2 Functionals and Duals

Suppose we have a linear map $f : k \to V$ that is linear. What determines this map? The map is completely determined by $f(1)$, since 1 is the only basis element of k, and f is determined by its action on the basis.

What if we have a linear map $f : V \to k$ that is linear? If $\{v_1, \dots, v_n\}$ is a basis for V, then this map is determined by $f(v_1), \dots, f(v_n)$. That is, f picks out exactly n scalars.

How are the maps $f : k \to V$ and $f : V \to k$ related? To start, let's look at a 2x2 matrix:

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

How do we "pick out" the entry a? To do this, we can use column and row vectors:

$$\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = a.$$

Multiplying these vectors and matrices, we get a. Notice that in order to get a, we need both a **column vector** and a **row vector**.

Let's think about this more abstractly. The *row vector* is really a map $V \to k$, the *column vector* is a map $k \to V$, and the matrix is a map $V \to V$. To see why this is true, we recall precisely the rules for matrix multiplication.

Remark 10.6. We can multiply a *column* vector by a scalar:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} a = \begin{pmatrix} av_1 \\ \vdots \\ av_n \end{pmatrix}.$$

That is, our *column* vector takes the scalar a, to a vector. So *column* vectors are maps $k \to V$, scalar to vector.

A *row* vector can be multiplied by a vector to give a number:

$$\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + \cdots + v_n w_n.$$

The last sum is just a scalar, so we see that our *row* vector takes a vector to a scalar, so *row* vectors are maps $V \to k$, vector to scalar.

Recall our notation $\mathcal{L}(V, W)$ as the set of linear maps from V to W. The above matrix/vector multiplication can be thought of as

$$\begin{aligned} \mathcal{L}(k, V) \times \mathcal{L}(V, W) \times \mathcal{L}(W, k) &\to , \\ (g, f, h) &\mapsto h \circ f \circ g : k \to V \to W \to k. \end{aligned}$$

We have a special name for linear maps from $W \to k$.

Definition 10.7 — Let V be a vector space. A **linear functional** f is a linear map from the vector space to the field, $f : V \to k$. The **dual space** of V is the vector space of all linear functionals on V, which is $\mathcal{L}(V, k)$. We denote the dual space of V as $V^\vee$. Another way of writing this is

$$V^\vee = \{ f : V \to k | f \text{ is linear } \}.$$

Notice that, by the above remark, a linear function $A : V \to k$ will act on a vector v by

$$Av = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = a_1 v_1 + \cdots + a_n v_n.$$

That is, we can think of a linear functional as a row vector. Since row vectors are just column vectors turned on their sides, then we are led to the following lemma:

Lemma 10.8

If V is a vector space, then the dual $V^\vee$ satisfies

(a) $\dim V^\vee = \dim V$, so $V^\vee \simeq V$.

(b) If $\{v_1, \ldots, v_n\}$ is a basis for V, then the set of functionals $\{\delta_1, \ldots, \delta_n\}$ forms a basis for $V^\vee$, where the functional δ_i is defined as

$$\delta_i(v_j) = \begin{cases} 0 & i \neq j \\ 1 & j = i \end{cases}.$$

You might have seen these delta functions before, called **Kronecker delta** functions.

Proof. If we prove part (b), then we will have a basis $\{\delta_1, \ldots, \delta_n\}$ for $V^\vee$ with the same number of elements as the basis of V, so $\dim V^\vee = \dim V$, which proves part (A). So we really only need to prove the second part.

To do this, we first need to show that the δ_i are linearly independent. Let's take $\alpha_1, \ldots, \alpha_n$ scalars. Then suppose that $\alpha_1 \delta_1 + \cdots + \alpha_n \delta_n = 0$. Since the δ_i are functions

$V \to k$, then so is this linear combination. So we can apply $(\alpha_1\delta_1 + \cdots + \alpha_n\delta_n)$ as a function to v_1, and we see

$$(\alpha_1\delta_1 + \cdots + \alpha_n\delta_n)(v_1) = \alpha_1,$$

by the definition of the delta functions. But we already said that $\alpha_1\delta_1 + \cdots + \alpha_n\delta_n = 0$, so this means that $\alpha_1 = 0$. We can repeat this for each α_i, and we see that $\alpha_i = 0$ for all i, which means that the set $\{\delta_1, \ldots, \delta_n\}$ is linearly independent.

Now, we need to show that $\{\delta_1, \ldots, \delta_n\}$ spans the dual space. Let's take any linear functional φ. We need to show that we can write φ as a linear combination of the δ_i. Now, φ is a linear functional, which is just a special type of linear map, so it is completely determined by what it does to the basis $v_1, \ldots, v_n$. Since $\varphi : V \to k$, then $\varphi(v_i) = \alpha_i \in k$ for each i. Then we can write $\varphi = \alpha_1\delta_1 + \cdots + \alpha_n\delta_n$. Notice that $(\alpha_1\delta_1 + \cdots + \alpha_n\delta_n)(v_i) = \alpha_i$ for each i. That is, this linear combination act in the same way as φ on the basis, so it is the same map as φ. Thus, the δ_i span the dual space, and so $\{\delta_1, \ldots, \delta_n\}$ is a basis for the dual space. □

To recap, the dual is defined as $V^\vee = \{f : V \to k\}$. If we pick a basis for V, then a functional is determined by n scalars. Any functional can be written as

$$\varphi = (\alpha_1, \ldots, \alpha_n).$$

when we apply the functional to a vector, we get

$$\varphi v = \begin{pmatrix} \alpha_1 & \cdots & \alpha_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

We can think of the δ_i functions in terms of the row vectors as well: δ_i is $(0, \ldots, 1, \ldots, 0)$, with the 1 in the i'th spot.

Let's now define the dual of a function:

Definition 10.9 (Dual Map) — If $f : V \to W$ is a linear map, then

$$f^\vee : W^\vee \to V^\vee$$

is defined as the map $\varphi \mapsto \varphi \circ f$. This map is called the **dual map** to f.

Using this, our goal is to work towards the following theorem:

Theorem 10.10 (Duality)

If $f : V \to W$ is a map of finite dimensional vector spaces, then

(a) We can associate to f a dual

$$f^\vee : W^\vee \to V^\vee.$$

(b) The function f is surjective if and only if $f^\vee$ is injective.

(c) The function f is injective if and only if $f^\vee$ is surjective.

Now, let's cover one more definition, just for fun.

Definition 10.11 — If $U \subseteq V$ is a subspace, then the **annihilator** of U, written $\text{ann}(U)$, is the set of functionals φ which send everything in U to zero. That is,

$$\text{ann}(U) = \{\varphi | \varphi(u) = 0 \text{ for all } u \in U\}.$$

Using the above, we also obtain the following proposition:

Proposition 10.12

If V, W are vector spaces over a field k, then

$$\begin{aligned}\mathcal{L}(V,W) &\simeq \mathcal{L}(V,k) \oplus \cdots \oplus \mathcal{L}(V,k) \\ &\simeq V^\vee \oplus \cdots \oplus V^\vee,\end{aligned}$$

where the direct sum is done $\dim W$ times.

§11 October 15, 2019

As a reminder, you have another midterm on Halloween. Also, it might be time to start thinking about your final project, which is expected to be an 8-10 page paper on some topic in linear algebra or its applications.

§11.1 Duals Part 2

Recall the definition of the dual of a vector space:

Definition 11.1 — If V is a vector space, then the **dual** of V is the vector space $V^\vee = \mathcal{L}(V,k)$, where k is the base field of V.

We also have the notion of a dual map

Definition 11.2 — If $f : V \to W$ is a linear map of vector spaces, then the **dual map** is defined as $f^\vee : W^\vee \to V^\vee$ defined by $f^\vee(\varphi) = \varphi \circ f$.

Also, recall from your problem set that if the sequence

$$V' \to V \to V''$$

is exact, then so is the sequence

$$(V'')^\vee \to V^\vee \to (V')^\vee,$$

and note the switching of the order of arrows.

Recall also the definition of the annihilator:

Definition 11.3 — If $U \subseteq V$ is a subspace, then the **annihilator** of U is defined as

$$\text{ann}(U) = \{\varphi : \varphi(U) = 0\}.$$

We also have the following lemma.

Lemma 11.4
Let $f : V \to W$ be a map of vector spaces. Then $\ker(f^\vee) = \operatorname{ann}(\operatorname{im}(f))$.

Proof. First, we show that $\ker(f^\vee) \subseteq \operatorname{ann}(\operatorname{im}(f))$. Suppose $\varphi \in \ker(f^\vee)$. We want to show that for all $w = f(v)$, then $\varphi(w) = 0$. But $\varphi(w) = \varphi(f(v) = (f^\vee(\varphi))(v) = 0$, since $v \in \ker(f^\vee)$, so $\ker(f^\vee) \subseteq \operatorname{ann}(\operatorname{im}(f))$.

Now, we show that $\ker(f^\vee) \supseteq \operatorname{ann}(\operatorname{im}(f))$. To this end, suppose that $\varphi \in \operatorname{ann}(\operatorname{im}(f))$. We want to show that $f^\vee(\varphi) = 0$, that is, that φ is in the kernel of $f^\vee$. Suppose that $v \in V$. Recall the definition of the dual map $f^\vee(\varphi) = \varphi \circ f$. Thus, $(f^\vee\varphi)(v) = (\varphi \circ f)(v) = \varphi(f(v))$. But $\varphi(f(v)) = 0$, since $\varphi \in \operatorname{ann}(\operatorname{im}(f))$. □

This lemma tells us that there is some interchange between kernels and images under duality. Before we proceed, we prove the lemma that you used on the last homework:

Lemma 11.5
Suppose that $U \subseteq V$ is a subspace. Then $\dim(V) = \dim(U) + \dim(\operatorname{ann}(U))$.

Proof. We know that the sequence

$$0 \to U \to V \to V/U \to 0$$

is exact. By your homework, the sequence

$$0 \to (V/U)^\vee \to V^\vee \to U^\vee \to 0$$

is also exact. Now, we have that $(V/U)^\vee \simeq \ker(v^\vee)$ by exactness, and also $(V/U)^\vee \simeq \operatorname{ann}(U)$, by our previous lemma. Then

$$\begin{aligned} \dim V &= \dim V^\vee \\ &= \dim(V/U)^\vee + \dim U^\vee \\ &= \dim \operatorname{ann}(U) + \dim U, \end{aligned}$$

as desired. □

We also show the following lemma, which will help us to prove the duality theorem.

Lemma 11.6 (a) $\operatorname{im}(f^\vee) = \operatorname{ann}(\ker(f))$

(b) $\operatorname{ann}(\operatorname{im}(f)) = \ker(f^\vee)$

(c) $\operatorname{ann}(\ker(f)) = \operatorname{im}(f^\vee)$

Proof. We prove (a), since we've already proved (b) and (c). You proved $\operatorname{im}(f^\vee) \subseteq \operatorname{ann}(\ker(f))$ in the homework. To show that the vector spaces are equal, we only need show that $\dim(\operatorname{im}(f^\vee)) = \dim(\operatorname{ann}(\ker(f)))$.

To see this, notice first that $\dim(\operatorname{im}(f^\vee)) = \dim(\operatorname{im}(f))$. To see why this is true, we note that

$$\begin{aligned} \dim(\operatorname{im}(f^\vee)) &= \dim V^\vee - \dim(\ker(f^\vee)) \\ &= \dim V - \dim \operatorname{ann}(\operatorname{im}(f)) \\ &= \dim(\operatorname{im}(f)), \end{aligned}$$

where we have used the facts about dimensions from parts (b) and (c) of this lemma. By rank nullity, we have $\dim(\mathrm{im}(f)) = \dim V - \dim\ker(f)$, and by the previous lemma, this is equal to $\dim\mathrm{ann}(\ker(f))$.

In summary, we have

$$\begin{aligned}\dim(\mathrm{im}(f^\vee)) &= \dim V^\vee - \dim(\ker(f^\vee)) \\ &= \dim V - \dim\mathrm{ann}(\mathrm{im}(f)) \\ &= \dim(\mathrm{im}(f)) \\ &= \dim V - \dim\ker(f) \\ &= \dim\mathrm{ann}(\ker(f)),\end{aligned}$$

so we have $\dim(\mathrm{im}(f^\vee)) = \dim\mathrm{ann}(\ker(f))$, as desired.

□

The power of what we are doing here is that we can reduce questions about linear maps to questions about dimensions, which is just numerics. We now prove an important theorem about duality.

Theorem 11.7 (Duality)

Suppose $f : V \to W$ is a linear map. Then f is a surjection if and only if $f^\vee$ is an injection. Also, f is an injection if and only if $f^\vee$ is a surjection.

Proof. Let $f : V \to W$ be a linear map, and $f^\vee : W^\vee \to V^\vee$ its dual. Then $f^\vee$ is surjective if and only if $\mathrm{im}(f^\vee) = V^\vee$, which is true if and only if $\mathrm{ann}(\ker(f)) = V^\vee$, which is true if and only if $\ker(f) = 0$, which is true if and only if f is injective. That is, $f^\vee$ is surjective if and only if f is injective, as desired.

The proof of $f^\vee$ injective if and only if f is surjective is similar, and we leave that proof to the reader. □

Let's see how all of this works in terms of matrices. First, lets recall the definition of the **transpose** of a matrix. This is just the matrix we get by flipping a matrix on its side. Let's take the matrix A to be

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

then the transpose of A is

$$A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Suppose that $f : V \to W$ is a linear map, and let $\{v_1, \ldots, v_n\}$ be a basis for V and $\{w_1, \ldots, w_m\}$ a basis for W. Recall that the linear map f is completely determined by its action on the basis of V, and we can describe f by a matrix:

$$\begin{pmatrix} \cdot & \cdot & \alpha_{i1} & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \alpha_{ij} & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdot & \cdot & \alpha_{im} & \cdot & \cdot \end{pmatrix},$$

where the action of the map f on the basis element v_i is

$$f(v_i) = \sum_{j=1}^{m} \alpha_{ij} w_j.$$

Let's say we want to "extract" α_{ij} from this matrix. To do this, we examine the dual bases $\{v_1^\vee, \dots, v_n^\vee\}$ and $\{w_1^\vee, \dots, w_m^\vee\}$. We then apply $w_j^\vee$ to $f(v_i)$. We get

$$\begin{aligned} w_j^\vee(f(v_i)) &= w_j^\vee(\alpha_{i1} w_1 + \cdots + \alpha_{im} w_m) \\ &= \alpha_{i1} w_j^\vee w_1 + \cdots + \alpha_{ij} w_j^\vee w_j + \cdots + \alpha_{im} w_j^\vee w_m \\ &= \alpha_{ij}, \end{aligned}$$

where in the last step, we have just used the definition of the dual basis functionals. That is, we have $w_j^\vee f(v_i) = \alpha_{ij}$.

We summarize the above in the following theorem.

Theorem 11.8

Let $f : V \to W$ be a linear map. Pick a basis $\{v_1, \dots, v_n\}$ for V and a basis $\{w_1, \dots, w_m\}$ for W, and let (α_{ij}) be a matrix for f. Then in the dual basis $\{v_1^\vee, \dots, v_n^\vee\}$, $\{w_1^\vee, \dots, w_m^\vee\}$, the matrix for $f^\vee$ is the transpose (α_{ji}) of the matrix for f. In other words, the transpose of a matrix gives the dual map.

Corollary 11.9

If A is an invertible matrix, then A^t is also invertible.

Proof. If f is linear and an isomorphism, then $f^\vee$ is linear and also an isomorphism, by theorem 11.7. □

§12 October 17, 2019

As a reminder, your midterm will be on Halloween. There will be extra credit awarded for "interesting" costumes. The lecture before this will be purely review.

§12.1 Eigenstuff Part I

Recall our goal for this part of the course: we are trying to understand $\mathcal{L}(V, W)$. In the previous lectures on duals, we looked at $\mathcal{L}(V, k)$, which is the dual space. If W is finite dimensional, then from our previous results, we have

$$\begin{aligned} \mathcal{L}(V, W) &\simeq \mathcal{L}(V, k \oplus \cdots \oplus k) \\ &\simeq \bigoplus \mathcal{L}(V, k) \\ &\simeq \bigoplus V^\vee. \end{aligned}$$

However, the above involves making non-canonical choices, which we would like to avoid. In order to decompose spaces in a more nice way, we use the following lemma about eigenvalues/eigenvectors, which will allow us to decompose our space in a more useful way.

Lemma 12.1

If $f : V \to V$ is a linear map with eigenvectors $v_1, \ldots, v_n$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\{v_1, \ldots, v_n\}$ is linearly independent.

Today, we will work towards proving this result. To start, let's recall the definition of eigenvalues/eigenvectors.

Definition 12.2 — Suppose that $f : V \to V$ is a linear map, and that v is a nonzero vector such that $f(v) = \lambda v$, where λ is a scalar. Then we say that v is an **eigenvector** of f, and that λ is the corresponding **eigenvalue**.

Let's look at a basic example.

Example 12.3

Consider the matrix

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

The eigenvectors of the matrix are $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, corresponding to eigenvalues 2 and 1/2 respectively. To see this, note that

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

Now we prove a very useful lemma about maps from vector spaces to themselves.

Lemma 12.4 1. If $f : V \to V$, and f is injective, then f is an isomorphism.

2. If $f : V \to V$, and f is surjective, then f is an isomorphism.

Proof. 1. We know that $\dim(V) = \dim \operatorname{im}(f) + \dim \ker(f)$. Since f is injective, then $\ker(f) = 0$, so $\dim(V) = \dim \operatorname{im}(f)$, so f is surjective. Since f is surjective and injective, it is an isomorphism.

2. Since the map is surjective, then $\dim \operatorname{im}(f) = \dim(V)$. Thus, $\dim \ker(f) = 0$, which implies f is injective, which means f is an isomorphism.

□

Corollary 12.5

Suppose that f is a linear map, and $\lambda \in k \setminus \{0\}$. Then λ is an eigenvalue if and only if $(f - \lambda \cdot \mathrm{id}) : V \to V$ is *not* injective, if and only if it's *not* surjective, if and only if it's *not* an isomorphism.

Proof. Notice that v is an eigenvector with eigenvalue λ if and only if v is in the kernel of $(f - \lambda \cdot \mathrm{id}) : V \to V$, which is defined by $w \mapsto f(w) - \lambda w$. Saying it's in the kernel of this map just means that $f(v) - \lambda v = 0$, or $f(v) = \lambda v$. Then the corollary follows from the lemma, since there will be some eigenvalue λ if and only if there is some v in the kernel of $(f - \lambda \cdot \mathrm{id})$, which means the kernel is nonzero, which means the map is not surjective/injective/isomorphism. □

Lemma 12.6

If $v_1, \ldots, v_k$ are eigenvectors of $f : V \to V$ with *distinct* eigenvalues $\lambda_1, \ldots, \lambda_k$, then the set $\{v_1, \ldots, v_k\}$ is linearly independent. In particular, if $k = \dim(V) <$ then the set $\{v_1, \ldots, v_k\}$ forms a basis for V.

We prove this lemma by contradiction. As a side note, most of our proofs involving linearly independence involve contradiction. This is because we are proving something like "if $\alpha_1 v_1 \cdots \alpha_n v_n = 0$, then *not* all of the α_i are zero." That is, we are proving a negative statement. Usually it's easier to prove positive statements, which we can do by using contradiction.

Proof. Suppose for contradiction that the set $\{v_1, \ldots, v_k\}$ is linearly dependent. Let $\ell \leq k$ be the smallest integer such that $\alpha_1 v_1 + \cdots + \alpha_\ell v_\ell$, such that $\alpha_\ell \neq 0$. Since ℓ is the smallest such integer, then $\{v_1, \ldots, v_{\ell-1}\}$ is linearly independent. Now, we apply f, to get

$$\begin{aligned} 0 &= f(0) \\ &= f(\alpha_1 v_1 + \cdots + \alpha_\ell v_\ell) \\ &= \alpha_1 \lambda_1 v_1 + \cdots + \alpha_\ell \lambda_\ell v_\ell. \end{aligned}$$

On the other hand, notice that

$$\begin{aligned} 0 &= \lambda_\ell(\alpha_1 v_1 + \cdots + \alpha_\ell v_\ell) \\ &= \lambda_\ell \alpha_1 v_1 + \cdots + \lambda_\ell \alpha_\ell v_\ell. \end{aligned}$$

So, we have that

$$\alpha_1 \lambda_1 v_1 + \cdots + \alpha_\ell \lambda_\ell v_\ell = \lambda_\ell \alpha_1 v_1 + \cdots + \lambda_\ell \alpha_\ell v_\ell.$$

Canceling, we have

$$\alpha_1 \lambda_1 v_1 + \cdots + \alpha_\ell \lambda_{\ell-1} v_{\ell-1} = \lambda_\ell \alpha_1 v_1 + \cdots + \lambda_\ell \alpha_{\ell-1} v_{\ell-1}.$$

Subtracting one side from the other, we see that

$$\alpha_1(\lambda_1 - \lambda_\ell) v_1 + \cdots + \alpha_{\ell-1}(\lambda_{\ell-1} - \lambda_\ell) v_{\ell-1} = 0.$$

But we know that the set $\{v_1, \ldots, v_{\ell-1}\}$ is linearly independent. Thus $\alpha_i(\lambda_i - \lambda_\ell) = 0$ for all $i = 1, \ldots, \ell - 1$. But we assumed that the λ_i are distinct, so this implies that $\alpha_i = 0$ for each i. Recall that we had

$$\lambda_\ell \alpha_1 v_1 + \cdots + \lambda_\ell \alpha_\ell v_\ell = 0.$$

But $\alpha_i = 0$ for $i = 1, \ldots, \ell$. Thus,

$$0 + \cdots + \lambda_\ell \alpha_\ell v_\ell = 0,$$

or

$$\lambda_\ell \alpha_\ell v_\ell = 0.$$

But this implies that $\alpha_\ell v_\ell = 0$. But α_ℓ was assumed to be nonzero, so this means $v_\ell = 0$. But v_ℓ cannot be zero, since it is an eigenvector, and eigenvectors are nonzero. Thus, we have reached our contradiction, and we see that $\{v_1, \ldots, v_k\}$ must be linearly independent, as desired. □

Corollary 12.7

A linear map $f : V \to V$ has *at most* $\dim(V)$ eigenvalues.

Proof. Suppose that f has $\dim(V) + 1$ eigenvalues. Then each one must correspond to some eigenvector. But this gives us a linearly independent set with $\dim(V) + 1$ elements, which is a contradiction to the definition of the dimension of V. □

An important note about eigenstuffs: things tend to be field dependent. That is, we can ask: if $f : V \to V$, then does there exists $v \neq 0$ such that $f(v) = \lambda v$. To see this, let's do a fun interactive!

At this point in the class, everyone was asked to stand up. You cannot sit down until you answer a question correctly.

Example 12.8

Let

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

Let's find the eigenvalues! Let's apply this to the vector $\begin{pmatrix} x \\ y \end{pmatrix}$.

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

Thus, we must have

$$\lambda x = -y$$
$$\lambda y = x.$$

We can solve this system by dividing the top equation by the bottom, and we get

$$\frac{x}{y} = \frac{-y}{x},$$

or $x^2 = -y^2$. This implies that either $x = y = 0$, or x, y are imaginary. Thus, over the field $\mathbf{R}$, A has no eigenvalues/eigenvectors, and over $\mathbf{C}$ it has eigenvalues i and $-i$, corresponding to eigenvectors

$$\begin{pmatrix} 1 \\ -i \end{pmatrix}, \begin{pmatrix} 1 \\ i \end{pmatrix},$$

respectively. At this point everybody is sitting down, which means you are very smart. Yay.

Let's do another example. We won't make you stand up this time.

Example 12.9

Let's look at the same matrix

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

except instead of being in $\mathbf{R}$ or $\mathbf{C}$, we'll work in our most favorite field: $\mathbf{F}_2$. Remember that in $\mathbf{F}_2$, $-1 = 1$. Thus, we have

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

This has one eigenvalue: 1. The eigenvalue has eigenvector

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Now, let's cover a proposition

Proposition 12.10

If $f : V \to V$ has $\dim(V) = n$ distinct eigenvectors $\{v_1, \ldots, v_n\}$ with $\dim(V)$ eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, then the matrix of f in the basis $\{v_1, \ldots, v_n\}$ is diagonal, and looks like

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

These facts have many implications about matrices. Let's look at another example.

Example 12.11

Take the matrix

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}.$$

This matrix is upper triangular. The matrix is invertible if and only if the determinant is nonzero, which happens if and only if $ad \neq 0$. That it, we only need to look at the diagonal elements. In addition, the eigenvalues are determined by the diagonal elements. This is true in general, as we'll see soon.

Lemma 12.12

IF A is an upper triangular matrix, then A is invertible if and only if *none* of the diagonal elements are zero.

Proposition 12.13

If A is an upper triangular matrix, then all the diagonal entries are eigenvalues.

Theorem 12.14

If $f : V \to V$ is a linear map, then there exists a basis for V such that the matrix for f is upper triangular.

We'll prove all of this stuff next time.

§13 October 22, 2019

§13.1 Upper Triangular Matrices

What we're going to try to prove today is that if a matrix A is upper triangular, then all the diagonal entries are eigenvalues. If we have a linear map f, and it corresponds to a matrix which is upper triangular, then this will allows us to easily extract its eigenvalues.

To show this, we want to see that if λ_i, one of the eigenvalues, appears on the diagonal of the matrix, then

$$A - \lambda \mathrm{id} = \begin{pmatrix} \lambda_1 & & * \\ & 0 & \\ & & 0 \\ 0 & & \lambda_n \end{pmatrix}$$

To see why this gives us the result, recall that a linear map f has an eigenvalue λ if and only if $f - \lambda\mathrm{id}$ is not invertible. That is, it has nontrivial kernel, which means there's some v with $(f - \lambda\mathrm{id})(v) = 0$, then $fv = \lambda\mathrm{id}v$, and λ is an eigenvalue.

In order to show this, we use the following lemma:

Lemma 13.1

If A is an upper triangular matrix, then A is invertible if and only if *none* of the diagonal elements are zero.

Proof. We write λ_i for the diagonal elements of our upper triangular matrix A. First, we show the $\Leftarrow$ direction.

We want to show that if all of the elements on the diagonal of a matrix are nonzero, then it is invertible. Recall that a linear map $f : V \to V$ is surjective if and only if it is an isomorphism (and thus invertible). Thus, all we need to show is that it is surjective. We know that $\mathrm{im}(f)$ is spanned by $\lambda_i v_i + u_i$, where $u_i \in \mathrm{Span}(v_1, \dots, v_{i-1})$. You can see this by looking at what the upper triangular matrix does to the basis vectors. Now, we claim that for all $1 \le j \le n$,

$$\mathrm{Span}(\lambda_1 v_1, \lambda_2 v_2 + u_2, \dots, \lambda_j v_j + u_j) = \mathrm{Span}(v_1, \dots, v_j).$$

We prove this by induction on j. Since none of the λ_i are zero, we can scale all of the λ_i so that they are all 1. We have $\mathrm{Span}(v_1) = \mathrm{Span}(\lambda_1 v_1)$. By induction hypothesis, we have that $\mathrm{Span}(v_1, \dots, v_j + u_j) = \mathrm{Span}(v_1, \dots, v_j)$. For purposes of induction, let's look at $\mathrm{Span}(v_1, v_2 + u_2, \dots, v_j + u_j, v_{j+1} + u_{j+1})$. By the induction hypothesis, this is equal to $\mathrm{Span}(v_1, \dots, v_j, v_{j+1} + u_{j+1})$. But remember that we have $u_{j+1} \in \mathrm{Span}(v_1, \dots, v_j)$, by assumption. This means that $\mathrm{Span}(v_1, \dots, v_{j+1} + u_{j+1}) = \mathrm{Span}(v_1, \dots, v_{j+1})$, which proves our inductive step, and therefore the first direction of the proof.

Now, we show that $\Rightarrow$ direction. We want to show that if A is invertible, then none of the elements on the diagonal are zero. For sake of contradiction, suppose that there is

some diagonal element which is zero. Let j be the smallest index such that $\lambda_j = 0$. Then, similarly to above, we have $\text{Span}(f(v_1), \ldots, f(v_j), \ldots, f(v_n))$ has dimension less than or equal $n-1$, since when we get to the λ_{j+1} term, we'll already be in the span of $v_1, \ldots, v_j$, and we don't get a new dimension. Since the dimension of the span of the $f(v_i)$ is not the dimension of V, then f cannot be surjective, and is thus not invertible. □

The idea of upper-triangularism is very powerful. If we can upper-triangularize a matrix, then we learn very much about the underlying map. We would like to upper-triangularize every matrix, because then life would be easy. Unfortunately, we can't make every matrix upper triangular. As a side note, when we say "upper-triangularize a matrix A," what we mean is that if f is the underlying map for the matrix A, then we find another matrix representing f which is upper triangular.

Now, let's prove a very important theorem. We will call this theorem the "Fundamental Theorem of Algebra," because it is very important. It says that we can upper-triangularize any map from a complex vector space to itself.

Theorem 13.2 (Fundamental Theorem of Algebra)

Let V be a vector space over $\mathbf{C}$. If $f : V \to V$ is a linear map, then there exists a basis $\{v_1, \ldots, v_n\}$ such that f is upper triangular.

Let's recall the more familiar Fundamental Theorem of Algebra, which says that if $p(x) \in \text{Poly}(\mathbf{C})$, then $p(x)$ can be factored as $p(x) = c(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$. By analogy, the Fundamental Theorem of *Arithmetic* says that we can factor any integer n into primes. So the Fundamental Theorem of Algebra sort of tells us that we can treat complex polynomials as integers.

Before we proceed, let's prove a lemma, that will be useful in the future.

Lemma 13.3

Every linear map has an eigenvalue.

Proof. Suppose that V is a finite dimensional vector space (over $\mathbf{C}$), and let $n = \dim V$. Let us examine the $n+1$ vectors $\{v, f(v), f^2(v), \ldots, f^n(v)\}$. These vectors are linearly dependent, since there are $n+1$ of them. Thus, there exist α_i with

$$0 = \alpha_0 v + \alpha_1 f v + \cdots + \alpha_n f^n v.$$

By the FToA, we can factor this as $c(f-\lambda_1 \text{id})(f-\lambda_2 \text{id})\cdots(f-\lambda_n \text{id})v$. Let's get rid of the zero. Then 0 is now the zero map, and we have

$$0 = c(f-\lambda_1 \text{id})\cdots(f-\lambda_n \text{id}),$$

which means $f - \lambda_i \text{id}$ is not injective for some λ_i, which means one of the λ_i is an eigenvalue. □

Recall from your problem set that if $f^2 = f$, then $V \simeq \text{im}(f) \oplus \text{im}(\text{id} - f)$. We can write $f^2 = f$ as $f^2 - f = 0$, which we can then write as $f(f-1) = 0$. These are just symbols, like $x^2 - x = x(x-1)$. That is, we can treat $f^2 - f$ as a polynomial. As another example, recall

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots.$$

If a function f is nilpotent, meaning $f^N = 0$ for some N, then we can write $1 + f + \cdots + f^{N-1} + 0 + 0 + \cdots$. Then we have

$$\frac{1}{1-f} = 1 + f + \cdots + f^{N-1}.$$

The whole $1/(1-f)$ thing seems crazy, but it's actually not, as long as f is nilpotent. We need f to be nilpotent since infinite sums of matrices don't make sense. In general, the moral of the story is that if we have some object, we can put it into a polynomial, and that's fine.

Now, let's prove our version of the Fundamental theorem of Algebra.

Theorem 13.4 (Fundamental Theorem of Algebra)

Let V be a vector space over $\mathbf{C}$. If $f : V \to V$ is a linear map, then there exists a basis $\{v_1, \ldots, v_n\}$ such that f is upper triangular.

Proof. Let us proceed by induction on the dimension of the vector space. For 1 dimensional vector spaces, the matrices are 1x1 matrices. All 1x1 matrices are upper triangular. This proves the base case.

Now, by the lemma we just proved, there exists some $v_1 \in V$ such that $v_1 \neq 0$ and v_1 is an eigenvalue of f. Then we can draw the exact sequence

$$0 \to \mathrm{Span}(v_1) \to V \to V/\mathrm{Span}(v_1) \to 0.$$

We can then draw the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Span}(v_1) & \longrightarrow & V & \longrightarrow & V/\mathrm{Span}(v_1) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f|_{\mathrm{Span}(v_1)}} & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle \tilde{f}} & & \\ 0 & \longrightarrow & \mathrm{Span}(v_1) & \longrightarrow & V & \longrightarrow & V/\mathrm{Span}(v_1) & \longrightarrow & 0 \end{array},$$

where $f_{\mathrm{Span}(v_1)}$ is the restriction of f to the span of v_1, and $\tilde{f}$ is the reduction of f modulo $\mathrm{Span}(v_1)$. By the induction hypothesis, and the fact that $\dim(V/\mathrm{Span}(v_1)) < n = \dim V$, then there exists a basis for $V/\mathrm{Span}(v_1)$ such that $\tilde{f}$ is upper triangular. This basis is $\{v_2 + \mathrm{Span}(v_1), v_3 + \mathrm{Span}(v_1), \ldots, v_n + \mathrm{Span}(v_1)\}$. You should think about why this is a basis for $V/\mathrm{Span}(v_1)$, by recalling the definitions of quotient spaces. We claim that $\{v_1, v_2, \ldots, v_n\}$ is a basis for V for which f is upper triangular. We will finish this proof next lecture. □

§14 October 24, 2019

§14.1 More Upper Triangular Matrices

Recall the lemma from last time.

Lemma 14.1

If $f : V \to V$ is a linear map and $\{v_1, \ldots, v_n\}$ is a basis for V. Suppose that A is a matrix for f in this basis. Then A is upper triangular if and only if for all j, we have $f(\mathrm{Span}(v_1, \ldots, v_j)) \subseteq \mathrm{Span}(v_1, \ldots, v_j)$.

Proof. Recall from last time, we have the following commutative diagram between exact sequences:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathrm{Span}(v_1) & \longrightarrow & V & \longrightarrow & V/\mathrm{Span}(v_1) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f|_{\mathrm{Span}(v_1)}} & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle \tilde{f}} & & \\ 0 & \longrightarrow & \mathrm{Span}(v_1) & \longrightarrow & V & \longrightarrow & V/\mathrm{Span}(v_1) & \longrightarrow & 0 \end{array},$$

and by the induction hypothesis, there exists a basis for which $\tilde{f}$ is upper triangular. We can then find vectors $v_2, \dots, v_n$ so that $\{v_2 + \mathrm{Span}(v_1), \dots, v_n + \mathrm{Span}(v_1)\}$ is a basis for $V/\mathrm{Span}(v_1)$, such that $\tilde{f}$ is upper triangular. In other words, for each $2 \le j \le n$, then

$$\tilde{f}(\mathrm{Span}(v_2 + \mathrm{Span}(v_1), \dots, v_j + \mathrm{Span}(v_1)) \subseteq \mathrm{Span}(v_2 + \mathrm{Span}(v_1), \dots, v_j + \mathrm{Span}(v_1)) \quad (*).$$

This is just what it means for $\tilde{f}$ to be upper triangular.

Now, we need to show that $v_1, v_2, \dots, v_n \in V$ is a basis for V in which f is upper triangular. To write this explicitly, w want that for $1 \le j \le n$, then $f(\mathrm{Span}(v_1, \dots, v_j)) \subseteq \mathrm{Span}(v_1, \dots, v_j)$. Now, $f(\mathrm{Span}(v_1)) = \mathrm{Span}(\lambda_1 v_1) = \mathrm{Span}(v_1)$. But by $(*)$ above, we know that $f(\mathrm{Span}(v_2, \dots, v_j) + \mathrm{Span}(v_1)) \subseteq \mathrm{Span}(v_2, \dots, v_j) + \mathrm{Span}(v_1)$. But $f(\mathrm{Span}(v_2, \dots, v_j) + \mathrm{Span}(v_1)) = f(\mathrm{Span}(v_1, \dots, v_j))$. □

Let's do a little summary of eigenstuff and upper triangular stuff:

(a) We want to understand linear maps.

(b) An eigenvector v with eigenvalue λ satisfies $fv = \lambda v$.

(c) If $v_1, \dots, v_m$ are eigenvectors with *distinct* eigenvalues, then $v_1, \dots, v_m$ are linearly independent.

(d) In order to understand linear maps, we would like matrices to be diagonalizable. But alas, this is not always possible :(

(e) The next best thing to diagaonalizability it upper-triangularizibility. Why do we like upper triangular matrices?

  (i) The diagonals of an upper triangular matrix are eigenvalues.

  (ii) An upper triangular matrix is invertible if and only if all diagonal entries are nonzero.

  (iii) The part above the diagonal is junk that we don't care about

  (iv) Over the complex numbers (or any algebraically closed field), all matrices are upper triangularizable.

§14.2 Generalized Eigenstuff

Definition 14.2 — Let V be a vector space over a field k. Let $\lambda \in k \setminus \{0\}$. The λ-**eigenspace** of f is

$$E(\lambda, f) = \{v : fv = \lambda v\} \subseteq V.$$

That is, it is the space of all vectors $v \in V$ which have eigenvalue λ.

We called this the "eigenspace." The following lemma tells us that its actually a subspace.

Lemma 14.3
$E(\lambda, f)$ is a subspace.

Lemma 14.4 (a) $E(\lambda, f)$ is f invariant. That is, $f\big|_{E(\lambda,f)}(E(\lambda, f))$ is equal to $E(\lambda, f)$. The notation $f\big|_{E(\lambda,f)}$ means the map f with its domain restricted to $E(\lambda, f)$.

(b) On $E(\lambda, f)$, then f can be written as a diagonal matrix:

$$\begin{pmatrix} \lambda & & 0 \\ & \ddots & \\ 0 & & \lambda \end{pmatrix}$$

(c) If $\lambda_1 \neq \lambda_2$, then $E(\lambda_1, f) \cap E(\lambda_2, f) = 0$. Hence $E(\lambda_1, f) + E(\lambda_2, f) = E(\lambda_1, f) \oplus E(\lambda_2, f)$. So $V \supseteq E(\lambda_1, f) \oplus + \cdots \oplus E(\lambda_n, f)$. If the $\supseteq$ is an equality, then we can write f as a diagonal matrix.

Let's do an example

Example 14.5
Let f be represented by the matrix

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

The only eigenvalue of this matrix is 2, and the eigenvector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then the space $E(2, f)$ is $E(2, f) = \text{Span}(e_1)$. But $E(2, f) \subsetneq k^2$.

Now, let's look at a new definition to help us with these things.

Definition 14.6 — An element $\lambda \in k \setminus \{0\}$ is a **generalized eigenvalue** with **generalized eigenvector** v if there exists $j \geq 1$ such that

$$(f - \lambda \text{id})^j v = 0.$$

The **generalized eigenspace** of λ is

$$G(\lambda, f) = \{v : (f - \lambda \text{id})^j v = 0 \text{ for some j}\}.$$

Remember that for normal eigenstuff, we just had $(f - \lambda \text{id})v = 0$. Here we've just added the j. What exactly is going on here, and why do we care. Before, the eigenspace was $E(\lambda, f) = \ker(f - \lambda \text{id}) \subseteq \ker(f - \lambda \text{id})^2$, where $(f - \lambda \text{id})^2 = (f - \lambda \text{id})(f - \lambda \text{id})$. The inclusion is true since if $(f - \lambda \text{id})v = 0$, then $(f - \lambda \text{id})(f - \lambda \text{id})v = (f - \lambda \text{id})(0) = 0$. We can keep going with this kernel thing, and we get

$$E(\lambda, f) = \ker(f - \lambda \text{id}) \subseteq \ker(f - \lambda)^2 \subseteq \ker(f - \lambda \text{id})^3 \subseteq \cdots \subseteq \bigcup_{n \geq 1} \ker(f - \lambda \text{id})^n = G(\lambda, f).$$

One more definition:

Definition 14.7 — The **geometric multiplicity** of λ is the dimension $\dim(E(\lambda, f))$. The **algebraic multiplicity** of λ is $\dim(G(\lambda, f))$. From the above inclusions, the geometric multiplicity is always less than or equal to the algebraic multiplicity.

Lemma 14.8

The generalized eigenspace $G(\lambda, f)$ is in fact a vector space. Moreover, $G(\lambda, f)$ is f-invariant. That is, $f\big|_{G(\lambda,f)} : G(\lambda, f) \to G(\lambda, f)$.

Example 14.9

Let's go back to the matrix

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

Recall that $E(2, \lambda) = \mathrm{Span}(e_1) \subsetneq V$. Now, look at $f(e_2)$. We have

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

But then look at

$$\begin{aligned}(f-2)(f-2)e_2 &= (f-2)\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 0 \\ 2 \end{pmatrix}\right) \\ &= (f-2)\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}.\end{aligned}$$

Thus, we see that $(f - 2\mathrm{id})^2 e_2 = 0$. So e_2 is in the generalized eigenspace $G(2, f)$. Thus, $E(2, \lambda) \subsetneq G(2, \lambda) = V$.

Let's prove some lemmas about generalized eigenstuff. We'll see from these that generalized eigenspaces are very nice.

Lemma 14.10

If, for some j, we have $\ker(f - \lambda\mathrm{id})^j = \ker(f - \lambda\mathrm{id})^{j+1}$, then for all $k \geq 0$, we have $\ker(f - \lambda\mathrm{id})^j = \ker(f - \lambda\mathrm{id})^{j+k}$.

Proof. We proceed by induction on k. We know that $\ker(f - \lambda\mathrm{id})^{j+k} \subseteq \ker(f - \lambda\mathrm{id})^{j+k+1}$. If v satisfies $(f - \lambda\mathrm{id})^{j+k+1}(v) = 0$, then we have $(f - \lambda\mathrm{id})^{j+k}((f - \lambda\mathrm{id})(v)) = 0$. By the induction hypothesis, we have $(f - \lambda\mathrm{id})v \in \ker(f - \lambda\mathrm{id})^{j+k-1}$. □

Lemma 14.11

For all $\lambda \in k \setminus \{0\}$, we have $\ker(f - \lambda\mathrm{id})^{\dim V} = G(\lambda, f)$, if V is finite dimensional.

Proof. This was on your midterm! We construct the flag

$$\{0\} \subseteq \ker(f-\lambda\mathrm{id}) \subseteq \ker(f-\lambda\mathrm{id})^2 \subseteq \cdots \ker(f-\lambda\mathrm{id})^{\dim V} = \ker(f-\lambda\mathrm{id})^{\dim V+1} = \cdots = G(\lambda, f).$$

To see that this stops at $\dim V$, note the lemma we proved above. If, for any of the $\ker(f-\lambda\mathrm{id})^j \subseteq \ker(f-\lambda\mathrm{id})^{j+1}$ is actually an equality, then we know the sequence must stabilize at that point. If $j < \dim V$, then it stops there and we are done. If the sequence of flags is strictly increasing, so all the $\subseteq$ are $\subsetneq$, then it must stop at $\dim V$, since that is the greatest length of a flag in a dimension V vector space, from your midterm. □

The point of all of this is that we can't always write $E(\lambda, f) \oplus \cdots \oplus E(\lambda_n, f) = V$. We can do this if the matrix is diagonalizable. However, we can do this with the generalized eigenvalues.

Theorem 14.12

If V is a vector space over $\mathbf{C}$, and $f : V \to V$ is a linear map, let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of f. Then $G(\lambda_1, f) \oplus G(\lambda_2, f) \oplus G(\lambda_m, f) = V$.

An immediate result of this theorem is the following corollary.

Corollary 14.13

The following equation holds:

$$\sum_{\lambda \in k \setminus \{0\}} \mathrm{algmult}(\lambda) = \dim V.$$

November 5, 2019

Tensor Products

We begin with a quote:

Whereof one cannot speak, thereof one must be silent

> "Whereof one cannot speak, thereof one must be silent"
>
> -Ludwig Wittgenstein

Now, let's talk about products. The products you've seen so far include

- Cross products $\times : \mathbf{R}^3 \times \mathbf{R}^3 \to \mathbf{R}$
- Dot product $\cdot : k^n \times k^n \to k$.
- Matrix multiplication $M_{m\times l}(k) \times M_{l\times n}(k) \to M_{m\times n}(k)$.
- Determinants.

Definition 14.14 — Suppose that V, W, Z are vector spaces over a field. Then a map $\varphi : V \times W \to Z$ is **bilinear** if

(a) $\varphi(\alpha v, w) = \alpha\varphi(v, w) = \varphi(v, \alpha w)$

(b) $\varphi(v_1 + v_2, w) = \varphi(v_1, w) + \varphi(v_2, w)$,

(c) Scalar multiplication $\cdot : k \times V \to V$.

(d) and $\varphi(v, w_1 + w_2) = \varphi(v, w_1) + \varphi(v, w_2)$.

This means that if we hold v constant, then we get a *linear* map $\varphi(v, \cdot) : W \to Z$, which takes $w \mapsto \varphi(v, w)$, and a *linear* map $\varphi(\cdot, w) : V \to Z$, taking $v \to \varphi(v, w)$. The familiar examples of products above are all bilinear, except for the determinant, which is multilinear.

Now, let's define a slight generalization of the bilinear map, called a multilinear map.

Definition 14.15 — Suppose $V_1, \dots, V_k$ are vector spaces, and Z is a vector space. Then a map $\varphi : V_1 \times \cdots \times V_k \to Z$ is k-**multilinear** if for all $i \in \{1, \dots, k\}$,

(a) $\varphi(v_1, \dots, \alpha v_i, \dots, v_k) = \alpha\varphi(v_1, \dots, v_i, \dots, v_k)$, and

(b) $\varphi(v_1, \dots, v_i + v_i', \dots, v_k) = \varphi(v_1, \dots, v_i, \dots, v_k) + \varphi(v_1, \dots, v_i', \dots, v_n)$.

Let's check that scalar multiplication on vector spaces is bilinear.

Example 14.16

The map $\cdot : k \times V \to V$ of scalar multiplication is bilinear. We have $\cdot(\alpha+\beta, v) = (\alpha+\beta)\cdot v = \alpha v + \beta v = \cdot(\alpha, v) + \cdot(\beta, v)$. Also, $\cdot(\alpha, v + w) = \alpha(v + w) = \alpha v + \alpha w = \cdot(\alpha, v) + \cdot(\alpha, w)$. Thus, the map is bilinear.

A lot of more advanced math is based on the fact that things are multilinear. Much of what you have seen in your life is multilinear, you just haven't realized it. Now that we're moving on to some more abstract topics, we use the "theorem/definition" idea that is common in more advanced math. That it, we define an object by its properties. Since the object has to be well defined, this definition also includes a "theorem" along with it, which says that the object being defined actually makes sense (e.g. is unique or exists).

Definition 14.17 — The **determinant** is the unique multilinear map

$$\det : k^n \times \cdots \times k^n \to k,$$

which satisfies

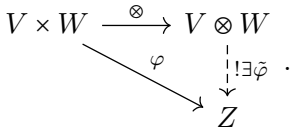
1. $\det(\mathrm{id}) = 1$, and
2. $\det(\text{matrix with two columns equal}) = 0$, known as the **alternating** property.

Example 14.18

$$\begin{aligned}
\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) &= \det(ae_1 + be_2, ce_1 + de_2) \\
&= a\det(e_1, ce_1 + d_2) + b\det(e_2, ce_1 + de_2) \\
&= ac\det(e_1, e_1) + ad\det(e_1, e_2) + bc\det(e_2, e_1) + bd\det(e_2, e_2) \\
&= ad - bc,
\end{aligned}$$

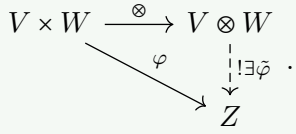
where the last equality follows from the alternating property of the determinant, and the others follow from the multilinearity of the determinant.

Before we examine the tensor product more, we need to understand multilinear algebra. The key object in multilinear algebra is the tensor product. The basic idea of tensor products is that if we have a bilinear map $\varphi : V \times W \to Z$, then we can "factor through" the tensor product, $V \times W \to V \otimes W \to Z$, so we get the same map, and such that the map $V \otimes W \to Z$ is unique. This is summarized by the following diagram:



We say that this diagram **commutes**, meaning whichever path you take along the arrows gives you the same map. The $!\exists$ means that there exists a *unique* map $\tilde{\varphi} : V \otimes W \to Z$. The idea here is that instead of using the bilinear map $\varphi : V \times W \to Z$, we can use a simpler unique *lienar* map $\tilde{\varphi}$, by first going to the tensor product.

Definition 14.19 — Given vector spaces V and W, the **tensor product**, and its associated bilinear map $\otimes$ is defined as the object such that for any vector space Z and for any bilinear map $\varphi : V \times W \to Z$, there exists a unique map $\tilde{\varphi} : V \otimes W \to Z$ such that $\varphi = \tilde{\varphi} \circ \otimes$. That is, there exists a unique $\tilde{\varphi}$ such that the following diagram commutes:



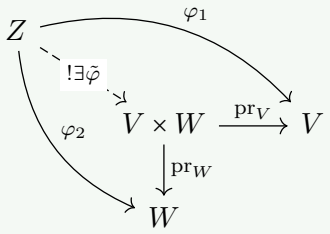
Objects defined this way are said to be defined by **universal property**. Let's go over a more familiar example of a universal property, since it's a rather abstract concept.

Example 14.20

Suppose V, W are vector spaces. We have projection maps $V \times W \to V$, and $V \times W \to W$. Consider Z, a third vector space, and linear maps from Z to $V \times W$. The set of linear maps from $Z \to V \times W$ is the same as the set of linear maps $Z \to V$ to $Z \to W$, in the following way.

Definition 14.21 — If V, W are vector spaces, then the **product** of V and W is the vector space $V \times W$, equipped with projection maps $\mathrm{pr}_V : V \times W \to V$ and $\mathrm{pr}_W : V \times W \to W$ satisfying the universal property: given any vector space Z, and maps $\varphi_1 : Z \to V$ and $\varphi_1 : Z \to W$, there exists a unique map $\tilde{\varphi} : Z \to V \times W$ such

that the following diagram commutes:



This is the "universal property of the product." Now, let's talk about category theory.

§14.3 Category Theory

Let's start with the definition of a category.

Definition 14.22 — A **category** $\mathbf{C}$ consists of a collection of objects $\mathrm{Obj}(\mathcal{C}) = \{X, Y, Z, \ldots\}$ and for any $X, Y \in \mathrm{Obj}(\mathbf{C})$, a set $\mathrm{Maps}(X, Y)$ with a map $\mathrm{Maps}(X, Y) \times \mathrm{Maps}(Y, Z) \to \mathrm{Maps}(X, Z)$, such that there is an element $\mathrm{id}_X \in \mathrm{Maps}(X, X)$, and such that the usual rules of composition are satisfied.

Some examples of categories:

Example 14.23 1. The category of sets: **Sets**, with objects $\mathrm{Obj} = \{S\}$, and maps $\mathrm{Maps}(S, T) = \{f : S \to T\}$.

2. The category of integers: the objects are the elements of the integers, and the maps are maps sending one integer to another, for example the map $+2 : 3 \mapsto 5$.

3. The category of vector spaces: $\mathbf{Vect}_k$. The objects are k-dimensional vector spaces, and the maps are the linear transformations between these.

Definition 14.24 — A **functor** is a type of map between categories $F : \mathbf{C} \to \mathbf{D}$. A functor takes objects to objects, and maps to maps.

Example 14.25
There is a functor $\cdot \to \mathbf{Vect}_k$, from the point category to the category of vector spaces, which picks out a single vector space.

§15 November 7, 2019

§15.1 Inner Product Spaces

Recall the theorem we proved about vector spaces over the complex numbers.

Theorem 15.1
If V is a vector space over $\mathbf{C}$, and $f : V \to V$ is a linear transformation, then there exists a basis for V in which f is upper triangular.

We can rewrite the theorem in the following way. Take $\lambda_1, \dots, \lambda_k$ which are eigenvalues of f, and consider $G(\lambda_1, f) \oplus G(\lambda_2, f) \oplus \cdots \oplus G(\lambda_k, f) \subseteq V$. The $\subseteq$ is an equality if f is upper triangular. A matrix is diagonalizable if and only if $E(\lambda_1, f) \oplus \cdots \oplus E(\lambda_k, f) = V$. In particular, if f has $n = \dim V$ distinct eigenvalues, then there exists a basis for V in which f is diagonal. This is because each $E(\lambda_i, f)$ is one-dimensional, and there are n of them, which means $E(\lambda_1, f) \oplus \cdots \oplus E(\lambda_k, f) = V$. But we can do a little better than this, using what's called the spectral theorem.

Theorem 15.2 (Spectral Theorem)
Suppose V is a vector space over $\mathbf{C}$. Then if f is **normal**, then f is diagonalizable.

We'll work towards proving this theorem. For now, don't worry about what the word "normal" means - we'll cover this in a bit. To prove this theorem, we'll need to know something about inner product spaces.

Definition 15.3 — An **inner product** on a vector space V, over a field k, is a function

$$(\cdot, \cdot) : V \times V \to k,$$

such that

(a) $|(v, v)| \geq 0$ for all $v \in V$, with equality if and only if $v = 0$,

(b) $(v, w) = \overline{(w, v)}$. Where the $\overline{(w, v)}$ is complex conjugation. Recall that if $(w, v) \in \mathbf{R}$, then $\overline{(w, v)} = (w, v)$, since the imaginary part a real number is zero.

(c) For all $w \in V$, then map $v \mapsto (w, v)$ is a linear functional.

Example 15.4 • The **dot product** on real vector spaces $\cdot : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$. Let's check the properties of the inner product to make sure:

(a) The inner product of a vector with itself is $(x_1, \dots, x_n) \cdot (x_1, \dots, x_n) = x_1^2 + \cdots + x_n^2$. Since the x_i are real numbers, then $x_i^2 \geq 0$, so the sum is also nonnegative. If any of the x_i are nonzero, then the sum will also be nonzero.

(b) $(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = x_1 y_1 + \cdots + x_n y_n = y_1 x_1 + \cdots + y_n x_n = (y_1, \dots, y_n) \cdot (x_1, \dots, x_n)$.

(c) We want to know that the map is linear in each component. We fix $\vec{w} = (w_1, \dots, w_n)$, and lets look at $(\alpha\vec{x} + \vec{y}, \vec{w})$. We want to show this is equal to $\alpha(\vec{x}, \vec{w}) + (\vec{y}, \vec{w})$. We can expand this as

$$\begin{aligned}
&(\alpha(x_1, \dots, x_n) + (y_1, \dots, y_n)) \cdot (w_1, \dots, w_n) \\
&= (\alpha x_1 + y_1, \alpha x_n + y_2, \dots, \alpha x_n + y_n) \cdot (w_1, \dots, w_n) \\
&= w_1(\alpha x_1 + y_1) + \cdots + w_n(\alpha x_n + y_n).
\end{aligned}$$

Expanding and rearranging gives the desired result.

• Now let's take our vector space to be over the complex numbers $\mathbf{C}$. This inner

product is given as

$$\vec{z} \cdot \vec{w} = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n.$$

Notice that $\vec{w} \cdot \vec{z} = \overline{\vec{z} \cdot \vec{w}}$. The proof that this is an inner product is left to the reader.

- Let's take our vector space $V = \text{Cont}([0,1], \mathbf{C})$, the set of continuous functions $[0,1] \to \mathbf{C}$. Recall that our vector addition is pointwise addition of functions $(f+g)(x) = f(x) + g(x)$, and scalar multiplication is $(\alpha f)(x) = \alpha(f(x))$. The inner product on this space is

$$(f, g) = \int_0^1 f(x)\overline{g(x)}dx.$$

This is a definition. To get some intuition about this inner product, you can do the physics trick of thinking of integrals as infinite extensions of sums. The inner product of two vectors in a finite dimensional vector space over $\mathbf{C}$, $\vec{z}$ and $\vec{w}$, is $\vec{z} \cdot \vec{w} = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n$. We can write this as

$$\sum_{i=1}^{n} z_i\bar{w}_i.$$

Now let's promote $\vec{z}$ and $\vec{w}$ to "infinite dimensional vectors," which we view as functions. The sum becomes an integral, and instead of the index i, we use the "continuous index" x, and we get

$$\int z(x)\overline{w(x)}dx,$$

which is what we had above. This isn't very rigorous (unless you're a physicist), so just think of it as intuition for this function inner product.

Now, let's prove something about inner products.

Proposition 15.5

For any inner product $(\cdot, \cdot)$ on V, the following are true:

(a) $(0, v) = 0 = (v, 0)$ for all $v \in V$.

(b) $(u, \lambda v + w) = \overline{\lambda}(u, v) + (u, w)$. That is, if we take a constant out of the second component of the inner product, we need to conjugate it .

Proof. Homework □

Now, let's go over some geometric intuition behind the inner product, which is one of the biggest reasons we care about inner products. You've probably seen a lot of these fact before, but we'll prove them rigorously here.

Definition 15.6 — The **length** of a vector v is $\|v\| = \sqrt{|(v, v)|}$. Two vectors u, v are said to be **orthogonal**, or **perpendicular** if $(u, v) = 0$.

Why does this make sense geometrically? There is a very nice relationship between inner products and angles, given in the following lemma.

Lemma 15.7

If $(\cdot,\cdot) := \cdot$ in $\mathbf{R}^2$ (the dot product in $\mathbf{R}^2$), then

$$u \cdot v = \|u\| \, \|v\| \cos\theta,$$

where θ is the angle between the two vectors.

Proof. Homework □

We will now generalize a theorem you probably first saw when you were two years old: the Pythagorean Theorem.

Theorem 15.8 (Generalized Pythagorean Theorem)

If u and v are orthogonal vectors, then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.

Proof. Using the definition of the norm and expanding, we have

$$\begin{aligned}\|u+v\|^2 &= (u+v, u+v)\\ &= (u,u) + (u,v) + (v,u) + (u,u)\\ &= (u,u) + (v,v)\\ &= \|u\|^2 + \|v\|^2,\end{aligned}$$

where we've used the fact that $(u,v) = (v,u) = 0$ since u and v are orthogonal. □

We will generalize a fundamental fact of life: the square of averages is less than or equal the average of squares.

Proposition 15.9

If $x_1, \ldots, x_n$ are real numbers, then

$$\left(\frac{x_1 + \cdots + x_n}{n}\right)^2 \leq \frac{x_1^2 + \cdots + x_n^2}{n}.$$

The generalization of this is called the Cauchy Schwartz Inequality. Before we prove this theorem, we prove a helpful lemma.

Lemma 15.10

Let u and v be vectors, with $v \neq 0$. Then we can write u as $u = cv + w$, where $c \in k$ (the field), and w is orthogonal to v.

Proof. Let's use for c and w:

$$c = \frac{(u,v)}{\|v\|^2}$$
$$w = u - \frac{(u,v)}{\|v\|^2} v.$$

Note that $c \in k$. We need to check that w is orthogonal to v. Distributing gives

$$\begin{aligned} w \cdot v &= \left(u - \frac{(u,v)}{\|v\|^2} v \right) \cdot v \\ &= (u,v) - \frac{(u,v)(v,v)}{\|v\|^2} \\ &= (u,v) - \frac{(u,v)\|v\|^2}{\|v\|^2} \\ &= (u,v) - (u,v) \\ &= 0. \end{aligned}$$

So we indeed have w orthogonal to v. Now, we have

$$\begin{aligned} cv + w &= \frac{(u,v)}{\|v\|^2} v + u - \frac{(u,v)}{\|v\|^2} v \\ &= \left(\frac{(u,v)}{\|v\|^2} v - \frac{(u,v)}{\|v\|^2} v \right) + u \\ &= u, \end{aligned}$$

so we have written the vector u in the desired form. □

Now, let's use this to prove the Cauchy-Schwartz inequality.

Theorem 15.11 (Cauchy-Schwartz Inequality)
Let u and v be vectors. Then $|(u,v)| \leq \|u\| \|v\|$.

Proof. If $v = 0$, then $(u,v) = (u,0) = 0 = \|u\| \|0\| = \|u\| \|v\|$. It's true in the case $v = 0$. Thus, to proceed further, we can assume $v \neq 0$ (which will be very important, since we will divide by the norm of v). By the previous lemma, we can write $u = cv + w$, where w is orthogonal to v. Now, all we need to prove

$$\|u\|^2 \geq \frac{|(u,v)|^2}{\|v\|^2}.$$

We have just square the Cauchy-Schwartz inequality, and moved things around (which we're allowed to do since everything is greater than zero). Now, let's examine $\|u\|^2$. We have

$$\begin{aligned} \|u\|^2 &= |c|^2 \|v\|^2 + \|w\|^2 \\ &\geq |c|^2 \|v\|^2. \end{aligned}$$

Recall that in proof of the previous lemma, we had

$$c = \frac{|(u,v)|}{\|v\|^2},$$

so the above becomes

$$\begin{aligned} \|u\|^2 &\geq \frac{|(u,v)|^2}{\|v\|^4} \|v\|^2 \\ &= \frac{|(u,v)|^2}{\|v\|^2}, \end{aligned}$$

which is our desired result. □

Let's apply this to get the life fact from above: take $x = (x_1, \dots, x_n)$ and $y = (1, \dots, 1)$. Then we have

$$(x_1 + \cdots + x_n)^2 = |xy|^2 \le \|x\|^2 \|y\|^2 = (x_1^2 + \cdots + x_n^2) \cdot n^2.$$

The middle inequality is the Cauchy-Schwartz inequality, and the last inequality is from the fact that $\|y\|^2 = n^2$, since $(y, y) = 1 + 1 + \cdots + 1$, n times.

§15.2 Orthonormality

Using orthonormal bases in general can greatly improve your life. If you ever do any math-heavy computer programming, you'll likely run into orthonormal bases. We can use them for matrix decompositions, which are very important.

Definition 15.12 — A set of vectors $\{v_1, \dots, v_n\}$ is said to be **orthonormal** if

(a) For all i, j with $i \neq j$, then $(v_i, v_j) = 0$.

(b) If $i = j$, then $\|v_i\|^2 = 1$.

The whole "$i \neq j$ implies 0 and $i = j$ implies 1" might seem familiar: let's relate this to dual bases. Recall that if $\{v_1, \dots, v_n\}$ is a basis for V, then there is a dual basis of vectors

$$\delta_i(v_j) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}.$$

The proofs for orthonormality will be very similar to many of the proofs about duals. This can be made precise by the Riesz Representation Theorem, which we won't cover right now. Now, let's do one proof that will seem very similar to a proof from our times working with duals.

Proposition 15.13
An orthonormal set is linearly independent.

Proof. Suppose that $\alpha_1, \dots, \alpha_k$ are scalars such that $0 = \alpha_1 v_1 + \cdots + \alpha_k v_k$. We consider the v_i as an inner product, with $0 = (0, v_j) = (\alpha_j v_j, v_j) = \alpha_j \|v_j\|^2 = \alpha_j$, which implies $\alpha_j = 0$. □

This leads us to define an **orthonormal basis** as an orthonormal set which is also a basis. As an application of this definition, we have the following lemma:

Lemma 15.14
If $\{v_1, \dots, v_n\}$ is an orthonormal basis, then $v = (v, v_1)v_1 + \cdots + (v, v_n)v_n$. That is, we can write v as a sum of its projections onto the axes of the orthonormal basis.

§16 November 12, 2019

§16.1 Iwasawa Decomposition and Gram-Schmidt

First, we will cover Iwasawa decomposition for $SL_2(\mathbf{R})$. The **special linear group** $SL_2(\mathbf{R})$ is a subset of $M_2(\mathbf{R}) = \{2 \times 2 \text{ matrices}\}$. It is in fact the subset defined as

$$\mathrm{SL}_2(\mathbf{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \det = 1, (ad - bc = 1) \right\}.$$

We use matrices with determinant 1 so that the product of two of these is still in the same set, since $\det(AB) = \det(A)\det(B)$, and if $\det(A) = \det(B) = 1$, then $\det(AB) = 1$. You can think of this as a hypersurface in four dimensional space, defined by the equation $ad - bc = 1$, or alternatively as the zero locus of $F(a, b, c, d) = ad - bc - 1$, the map from $\mathbf{R}^4 \to \mathbf{R}$. This is similar to how the circle in $\mathbf{R}^2$ is defined as the zero locus of $x^2 + y^2 - 1$.

Now, let's visualize $SL_2(\mathbf{R})$. To do this, let's look at the subsets of $SL_2(\mathbf{R})$ defined by

$$K = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \right\}.$$

The two columns of this matrix will form an orthonormal basis for $\mathbf{R}^2$. Let's now look at some more subsets of $SL_2(\mathbf{R})$, given by

$$A = \left\{ \begin{pmatrix} r & 0 \\ 0 & 1/r \end{pmatrix} : r > 0 \right\} \simeq \mathbf{R}_{>0}$$
$$U = \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \right\} \simeq \mathbf{R}.$$

The set of the matrices A are the scaling matrices - notice that the x and y axes are scaled by inverse factors, so a scaled parallelogram will still have the same area. The matrices U are the shearing matrices.

Theorem 16.1 (Iwasawa)
Any matrix in $SL_2(\mathbf{R})$ can be written uniquely as a product of a rotation, a shear, and a scale. In other words, we have, in the above notation,

$$SL_2(\mathbf{R}) \simeq K \times A \times U \simeq S^1 \times \mathbf{R}_{>0} \times \mathbf{R}.$$

Notice, S^1 is the unit circle (not filled in), and $\mathbf{R}_{>0}$ is the positive real numbers. Then the product $S^1 \times \mathbf{R}_{>0}$ is a cylinder (we attach $\mathbf{R}_{>0}$ at each point of the circle). Then, when we do $(S^1 \times \mathbf{R}_{>0}) \times \mathbf{R}$, we attach a cylinder to each point of the real line. That is, we can think of $SL_2(\mathbf{R})$ as a cylinder moving along (parameterized by) the real numbers. Before we prove this theorem, we'll need the definition of an **orientation**

Definition 16.2 — Given an ordered pair of vectors (v_1, v_2), we say that (v_1, v_2) has **positive orientation** if $\det(v_1, v_2) > 0$, where (v_1, v_2) here is the matrix with the v_i as columns. If $\det(v_1, v_2) < 0$, then (v_1, v_2) has **negative orientation**.

Please note, when you read through this proof, please try to draw pictures of everything, which will make things make more sense.

Proof. Given

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

with $ad - bc = 1$, we consider its action on the standard basis $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then

$$Ae_1 = \begin{pmatrix} a \\ c \end{pmatrix}$$
$$Ae_2 = \begin{pmatrix} b \\ d \end{pmatrix}.$$

Also, let θ be the angle between the positive x-axis and Ae_1. Define $\rho_{-\theta} : \mathbf{R}^2 \to \mathbf{R}^2$ as the clockwise rotation of the plane by θ, so $\rho_{-\theta}(Ae_1) \in \mathbf{R}_{x,>0}$. Now, $\det A = 1 > 0$, so Ae_1 and Ae_2 have the same orientation as (e_1, e_2). Hence, $\rho_{-\theta}(Ae_2)$ is in the upper half plane. That is, $\rho_{-\theta}$ takes Ae_1 to the x-axis, and Ae_2 to the upper half plane (meaning it points above the x-axis). Now, $\rho_{-\theta} = re_1$ for some $r > 0$. But who is r? This is just $\|\rho_{-\theta}(Ae_1)\| = \|Ae_1\|$, where the equality holds since $\rho_{-\theta}$ is a rotation and doesn't disturb the length. That is, r is just the length of Ae_1. Since $\rho_{-\theta}$ moves Ae_1 to the x-axis, then we can write $\rho_{-\theta}(Ae_1)$ as re_1. Recalling the column vector definition of Ae_1, then $r = \sqrt{a^2 + c^2}$.

Now, let's look at

$$B = \begin{pmatrix} 1/r & 0 \\ 0 & 1/r \end{pmatrix},$$

where $r = \sqrt{a^2 + c^2}$ from above. Notice that $B(\rho_{-\theta}(Ae_1)) = e_1$, since it scales the x-axis as $1/r$, and $\rho_{-\theta}(Ae_1) = re_1$. What does B do to $\rho_{-\theta}(Ae_2)$? If we call

$$\rho_{-\theta}(Ae_2) = \begin{pmatrix} x \\ y \end{pmatrix},$$

then $B(\rho_{-\theta}(Ae_2)) = \begin{pmatrix} (1/r)x \\ ry \end{pmatrix}$. Since $B(\rho_{-\theta}(Ae_1)) = e_1$, then $B(\rho_{-\theta}(Ae_2))$ must be the vector in the upper half plane which forms a *unit* parallelogram with e_1. This is because everything has determinant 1, and therefore the parallelogram formed by the vectors must have unit area. There is only one choice for this vector that forms a unit parallelogram. This means that $B(\rho_{-\theta}(Ae_2))$ is a shear of e_2 (draw a picture, and this will make sense). That is, $B\rho_{-\theta}A$ is a shear matrix

$$B\rho_{-\theta}A = \begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix},$$

for some p. Then to get $B(\rho_\theta Ae_2)$ to the y-axis (or equivalently to e_2), we multiply by

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

This matrix does not change $B(\rho_{-\theta}(Ae_1)) = e_1$. Thus, we have

$$CB\rho_{-\theta}A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now, using this, we can write

$$A = \rho_\theta \begin{pmatrix} r & 0 \\ 0 & 1/r \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix},$$

which is a matrix in K times a matrix in A times a matrix in U, as desired. Thus, $SL_2(\mathbf{R}) \simeq K \times A \times U$. You can also prove uniqueness, but this is boring, so we won't do it here. □

So how does this relate to Gram-Schmidt? Notice that we can write a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} r & 0 \\ 0 & 1/r \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}.$$

Keep this decomposition in mind for the following theorem.

Theorem 16.3 (Gram-Schmidt)
Every finite dimensional vector space V (over $\mathbf{R}$ or $\mathbf{C}$) has an orthonormal basis.

Proof. Take a basis. Apply Gram-Schmidt algorithm. The basis is now orthonormal. Done. □

Now, let's do this in terms of the Iwasawa decomposition above. Define

$$K = \{(v_1, \dots, v_n) : \{v_i\} \text{ orthonormal}\}$$
$$A = \left\{ \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \right\}$$
$$U = \left\{ \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \right\}$$

These are just generalizations of the 2×2 matrices from the discussion above. Then we have the generalized theorem:

Theorem 16.4 (Iwasawa)
We have $SL_n(\mathbf{R}) = \{A : \det A = 1\} \simeq K \times A \times U$.

Proof. Gram-Schmidt. □

Next, we want to prove the spectral theorem. For this, we need the following definition.

Theorem 16.5
If $f : V \to W$ is a linear map, and V, W are inner product spaces, then the conjugate transpose of f is a linear map $f^* : W \to V$, such that $v \in V, w \in W$, then $(fv, w)_W = (v, f^*w)_V$.

Now, recall from last time, the Riesz representation theorem, which said that any functional α can be written as $\alpha(v) = (v, w_\alpha)$, where w_α is unique. Think about $f^\vee$ and f^*, and how they are related.

§17 November 14, 2019

§17.1 Riesz Representation Theorem

Theorem 17.1 (Riesz Representation Theorem)
If V is a finite dimensional vector space, and $\alpha \in V^\vee$, then there exists a unique w_α such that

$$\alpha(\cdot) = (\cdot, w_\alpha) : V \to k.$$

The content of this is that linear functionals can all be thought of as inner products.

Proof. By Gram-Schmidt, V has an orthonormal basis $\{e_1, \dots, e_n\}$. We want to find a w_α that works. We choose

$$w_\alpha = \overline{\alpha(e_1)}e_1 + \cdots + \overline{\alpha(e_n)}e_n,$$

and show that this works. To see this, note that

$$\begin{aligned}(v, w_\alpha) &= (v, \overline{\alpha(e_1)}e_1 + \cdots + \overline{\alpha(e_n)}e_n) \\ &= \alpha(e_1)(v, e_1) + \cdots + \alpha(e_n)(v, e_n) \\ &= \alpha((v, e_1)e_1 + \cdots + (v, e_n)e_n) \\ &= \alpha(v).\end{aligned}$$

Thus, we indeed have $(v, w_\alpha) = \alpha(v)$, so $(\cdot, w_\alpha) = \alpha(v)$. □

Corollary 17.2
Consider the map $\varphi : V \to V^\vee$ defined by

$$w \mapsto (\cdot, w),$$

where $(\cdot, w)$ is the map $V \to k$ taking v to (v, w). The map φ is an isomorphism.

§17.2 Normal Maps

Given a map $f : V \to W$, consider the functional $\beta_w = (\cdot, w) : W \to k$. Consider $\beta_w(f(\cdot)) : V \to k$. By Riesz, there exists a unique u such that $\beta_w(f(\cdot)) = (\cdot, u)$ on V. That is, $(f(v), w) = beta_w(f(v)) = (v, u)$. Renaming $u = f^* w$, we get a unique map $f^* : W \to V$, which takes $w \mapsto u = f^* w$. Note then that $(v, u) = (v, f^* w) = (fv, u)$.

Definition 17.3 — If $f : V \to W$ is a linear map, where V, W have inner products, then the **conjugate** of f is the unique linear map $f^* : W \to V$ such that for all $v \in V$ and $w \in W$, then $(fv, w) = (v, f^* w)$. You should check that f^* exists.

Lemma 17.4
The map f^* is linear.

Proof. We have the following equalities

$$\begin{aligned}(v, f^*(w+w')) &= (fv, w+w') \\ &= (fv, w) + (fv, w') \\ &= (v, f^*w) + (v, f^*w') \\ &= (v, f^*w + f^*w').\end{aligned}$$

Now, this implies that $f^*(w+w') = f^*w + f^*w'$, by the uniqueness part of the Riesz representation theorem. Note, we can apply uniqueness since the equality holds for *all* w, w'. The scalar multiplication part of the linearity is left to the reader. □

Some basic properties of f^*, which we leave to the reader:

Lemma 17.5

- $(f+\lambda g)^* = f^* + \lambda g^*$,
- $(f^*)^* = f$,
- $(\mathrm{id})^* = \mathrm{id}$,
- $(f \circ g)^* = g^* \circ f^*$.

Consider a map $f : V \to W$. We have a map $f^* : W \to V$. Note that this does not introduce a dual. We also have $f^\vee : W^\vee \to V^\vee$. The former is "genuine duality," and the latter is "strong duality." We also have a map $W \to W^\vee$, from the Riesz representation theorem, and similarly a map $V \to V^\vee$.

$$\begin{array}{ccc} W & \longrightarrow & W^\vee \\ \downarrow{\scriptstyle f^*} & & \downarrow{\scriptstyle f^\vee} \\ V & \longrightarrow & V^\vee \end{array}$$

The above diagram commutes. You can check this by drawing a commutative diagram. The proof is left to the reader.

Now, let's talk about normality.

Definition 17.6 — A map $f : V \to V$ is **normal** if $f^*f = ff^*$. A map $f : V \to V$ is **self-adjoint**, or **Hermitian** if $f = f^*$. All self adjoint maps are normal.

The property of being normal is very nice. We will show that normal implies diagonalizable. Before we prove this, we cover the following lemma.

Lemma 17.7

A map f is self-adjoint if and only if $(fv, w) = (v, fw)$.

The proof is left to the reader.

Example 17.8

Let $V = \mathbf{R}^n$. The inner product on this space is the dot product. Let $\{e_1, \dots, e_n\}$ be an orthonormal basis, and let A be a matrix corresponding to f under this basis, where f is self-adjoint. Then

$$\begin{aligned} A_{ij} &= (Ae_i, e_j) \\ &= (e_i, Ae_j) \\ &= (Ae_j, e_i) \\ &= A_{ji}. \end{aligned}$$

This tells us that the matrix is symmetric ($\alpha_{ij} = \alpha_{ji}$). Over the complex numbers, we find instead that $\alpha_{ij} = \overline{\alpha_{ji}}$.

As a remark, the condition of being normal is a *weakening* of the condition of self-adjointness, where diagonalizability still holds.

Lemma 17.9

A self-adjoint map $f : V \to V$ has real eigenvalues, and eigenvectors for distinct eigenvalues are orthogonal.

Proof. Suppose λ is an eigenvalue of f corresponding to some eigenvector v. Then

$$\begin{aligned} \lambda \|v\|^2 &= (\lambda v, v) \\ &= (fv, v) \\ &= (v, fv) \\ &= (v, \lambda v) \\ &= \bar{\lambda} \|v\|^2 . \end{aligned}$$

Since v is an eigenvector, it is nonzero, so we can divide by $\|v\|^2$ to get $\lambda = \bar{\lambda}$, which implies λ is real.

Now, suppose that $fv = \alpha v$, and $fw = \beta w$, where $\alpha \neq \beta$. We know that v, w are linearly independent, but we'd like to show now that they're orthogonal, which is better. We have

$$\begin{aligned} (\alpha - \beta)(v, w) &= (fv, w) - (v, fw) \\ &= (fv, w) - (v, f^* w) \\ &= (fv, w) - (fv, w) \\ &= 0, \end{aligned}$$

and since $\alpha - \beta \neq 0$, then $(v, w) = 0$. □

We will now work towards the spectral theorem. First, we need the following lemma:

Lemma 17.10

If $f : V \to V$ is a linear map, then $\|fv\| = \|f^* v\|$ if and only if f is normal.

Once we have this lemma we have the spectral theorem. Any matrix over $\mathbf{C}$ can be made upper triangular. Then we know that if (α_{ij}) is the upper triangular matrix for f, then $\|fe_1\|^2 = |\alpha_{11}|^2$, and $\|f^*e_1\|^2 = |\alpha_{11}|^2 + \cdots + |\alpha_{1n}|^2$. Since each $|\alpha_{ij}|^2 \geq 0$, this tells us that $|\alpha_{12}|^2 = |\alpha_{13}|^2 = \cdots = |\alpha_{1n}|^2 = 0$. We can do this for the other rows/columns as well, and we get that the matrix is diagonal. Thus, in order to prove the spectral theorem, we only need to prove the above lemma. In order to prove this lemma, we need some more lemmas.

Lemma 17.11 (Symmetryzing maps)
For any $f : V \to V<$ then $f^*f - ff^*$ is self-adjoint.

Proof. We have

$$\begin{aligned}(f^*f - ff^*)^* &= (f^*f)^* - (ff^*)^* \\ &= (f)^*(f^*)^* - (f^*)^*(f)^* \\ &= f^*f - ff^*,\end{aligned}$$

so the map is self-adjoint. □

Lemma 17.12 (Zero detection)
If f is self-adjoint, then $f = 0$ if and only if for all $v \in V$, then $(fv, v) = 0$. The proof is left to the reader.

Now we prove the main lemma.

Lemma 17.13
A map f is normal if and only if $\|fv\| = \|f^*v\|$.

Proof. The map f is normal if and only if $f^*f - ff^* = 0$, by definition. By the zero detection lemma, this is true if and only if $((f^*f - ff^*)v, v) = 0$, which is true if and only if $(f^*f, v) = (ff^*v, v)$, which is true if and only if $\|fv\|^2 = \|f^*v\|^2$, which is true if and only if $\|fv\| = \|f^*v\|$, as desired. □

§18 November 19, 2019

§18.1 Spectral Theory

Suppose that V is a finite dimensional vector space over $\mathbf{C}$, and $f : V \to V$ is linear. Since V is a complex vector space, then f is upper triangularizable. In this case, we can define the determinant: write the matrix of f as an upper triangular matrix, where λ_i are the elements of the diagonal. Then the **determinant** of f is the product

$$\det(f) = \prod_{\lambda_i \in \text{ diagonal}} \lambda_i.$$

We will prove that the determinant is independent of the basis chosen for V (later). Let's also define the **characteristic polynomial**, as

$$\operatorname{char}(f)(z) = \prod_{\lambda_i \in \text{ diagonal}} (z - \lambda_i).$$

That is, it's a polynomial $(z-\lambda_1)^k \cdots (z-\lambda_j)^k$, where the λ_i are the diagonal elements, and the k_i are their multiplicity.

Further, we define the **spectrum** of a matrix, as the multiset of eigenvalues of the matrix. It's the multiset since we also count multiplicity.

Let's ask a question: can you guess the shape of a drone, just from listening? That is, can we recover f just from it's spectrum? Let's try to answer this question.

Theorem 18.1 (Spectral Theorem)

If f is normal, then f is diagonalizable.

We want to prove this theorem. To do this, we'll need the following lemma.

Lemma 18.2 (Zero Detection)

If f is *any* linear map over $\mathbf{C}$ (we don't require normality here), then f is zero if and only if $(fv, v) = 0$ for all v.

Proof. One direction of the proof is easy: if f is zero, then $(f(v), v) = (0, v)$ for all v. For the other direction, we use a trick. Suppose that $(fu, u) = 0$ for all $u \in V$. We know that if $(fv, w) = 0$ for all v, w, then v is zero (here the v, w can be different). Now, for all $v, w \in V$, we have

$$(fv, w) = \frac{(f(v+w), v+w) - (f(v-w), v-w))}{4} + \frac{(f(v+iw), v+iw) - (f(v-iw), v-iw)}{4}.$$

Now all four terms are of the form (fu, u), which are zero. Thus $(fv, w) = 0$ for all v, w, therefore f is zero. This sort of trick is fairly common, so it might be worth remembering. □

Unfortunately, a similar trick doesn't work for the real numbers. Now, we prove another lemma we'll need for proving the spectral theorem.

Lemma 18.3

A map f is normal if and only if $\|fv\| = \|f^* v\|$.

Proof. First, note that $ff^* - f^* f$ is self adjoint. Then f is normal if and only if $ff^* - f^* f = 0$, which by the lemma is true if and only if $((ff^* - f^* f)v, v) = 0$ for all f, which is true if and only if $(f^* fv, v) = (ff^* v, v)$, which is true if and only if $\|fv\|^2 = \|f^* v\|^2$. □

Now, we cover a theorem of Schur.

Theorem 18.4 (Schur)

If f is upper triangular with respect to τ_0, a basis for V, then f is upper triangular with respect to some orthognormal basis.

Proof. Recall that in order to be upper triangular, we need to find a basis $\{v_1, \ldots, v_n\}$, such that for all $1 \le j \le n$, then

$$f(\text{Span}(v_1, \ldots, v_j)) \subseteq \text{Span}(v_1, \ldots, v_j).$$

To prove the theorem, we just do Gram-Schmidt on this basis to orthonormalize. □

Now, we have all the tools to prove the spectral theorem.

Theorem 18.5 (Spectral Theorem)
If f is normal, then f is diagonalizable.

Proof. By the corollary, there is a basis such that f is upper triangular. Write the matrix of f in this basis as (a_{ij}), where $a_{ij} = 0$ for $i \not\geq j$. Now, notice that

$$\begin{aligned} \|fe_1\|^2 &= |a_{11}|^2 \\ \|f^* e_1\|^2 &= |a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2 . \end{aligned}$$

Since f is normal, then $\|fe_1\|^2 = \|f^* e_1\|^2$, and each of the $|a_{ij}| \geq 0$, then $a_{12} = a_{13} = \cdots = a_{1n} = 0$. Similarly, we have

$$\begin{aligned} \|fe_2\|^2 &= |a_{12}|^2 + |a_{22}|^2 = |a_{22}|^2 \\ \|f^* e_2\|^2 &= |a_{22}|^2 + |a_{23}|^2 + \cdots + |a_{2n}|^2 . \end{aligned}$$

Similarly to above, we then have that $a_{23} = a_{24} = \cdots = a_{2n} = 0$. We continue this process, and eventually we get $a_{ij} = 0$ for all $i \neq j$. □

That's the spectral theorem over the complex numbers. Now let's do the spectral theorem over the real numbers. First, let's define the determinant. Suppose $f : V \to V$ is a linear map, where V is a real vector space. We can write $f_{\mathbf{C}}$ for the map

$$f_{\mathbf{C}} : V \otimes_{\mathbf{R}} \mathbf{C} \to V \otimes_{\mathbf{R}} \mathbf{C}.$$

Now, $f_{\mathbf{C}}$ is a linear map over $\mathbf{C}$. Thus, it has an upper triangular form over $\mathbf{C}$, so the determinant is defined over $\mathbf{C}$. Are we allowed to actually do this? We'll examine this in more detail later. For now, let's use the following definition of the determinant:

Definition 18.6 — The **determinant** of a matrix is the unique function $\det : \mathcal{M}(n,n) \to k$ such that

- $\det(I) = 1$
- det is linear in the rows of a matrix
- If two adjacent rows of a matrix A the same, then $\det(A) = 0$.

Before we prove the spectral theorem for the reals, let's prove a lemma.

Lemma 18.7
Let $f : V \to V$ be a linear map over a real vector space V. Suppose $b, c \in \mathbf{C}$, and $b^2 < 4c$. Then $f^2 + bf + cI$ is invertible.

Proof. The inner product

$$\begin{aligned} (f^2 + bf + cI)v, v) &= (f^2 v, v) + (bfv, v) + (cv, v) \\ &= (fv, fv) + b(fv, v) + c(v, v) \\ &\geq \|fv\|^2 - |b| \, \|f(v)\| \, \|v\| + c\|v\|^2 , \end{aligned}$$

where the last inequality follows from Cauchy-Schwartz. Then this is equal to

$$\left(\|fv\|^2 - \frac{|b|\,\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{v}\right)\|v\|^2.$$

This last quantity is strictly greater than zero. Thus, we know that $(f^2 + bf + cI)v \neq 0$ for any $v \neq 0$. This implies that $f^2 + bf + cI$ is invertible. □

Now, let's do the real spectral theorem.

Theorem 18.8 (Real Spectral Theorem)
If $V \neq \{0\}$ is a vector space over $\mathbf{R}$, then if f is self-adjoint, then f is diagonalizable.

Proof. First, we find an eigenvalue for $\mathbf{R}$-self-adjoint matrices. We proceed similarly to the $\mathbf{C}$ case. Consider

$$v, fv, f^2v, \ldots, f^{\dim V}v,$$

which are linearly dependent, since there are $\dim V + 1$ of them. Then we can find $\alpha_0, \ldots, \alpha_n$ such that

$$0 = \alpha_0 v + \alpha_1 fv + \cdots + \alpha_n f^n,$$

where n is $\dim V$. In the complex case, we could factor this into linear terms. However, in the real case, the best we can do is to factor this as

$$0 = c(f^2 + b_1 f + c_1 I)\cdots(f^2 + b_m f + c_m I)(f - \lambda_1 I)\cdots(f - \lambda_k I)v,$$

such that $4b_j < c_j$ for each j. Then by the lemma above, the quadratic terms are invertible, so we have that

$$0 = (f - \lambda_1 I)\cdots(f - \lambda_k I)v.$$

Thus, we have shown that there exists an eigenvalue for f, an arbitrary self-adjoint matrix.

Now, we proceed by induction. If $\dim V = 1$, then we're done. Suppose that $\dim V = n > 1$, and that the result holds for matrices of size $(n-1) \times (n-1)$, then find an eigenvector v with $\|v\| = 1$, by the above. We can assume its norm is 1 since we can just divide by a scalar and it has the same norm. Now, let $U = \operatorname{Span}(v)$. Consider $U^\perp$. Then $f\big|_{U^\perp}$ is self-adjoint (in your homework). Since $U^\perp$ is an $n-1$ dimensional matrix, then it is diagonalizable by the induction hypothesis. When we combine the matrices for $f\big|_{U^\perp}$ and the $f\big|_{U}$, we get a diagonal matrix, so the full $n \times n$ matrix is diagonalizable, as desired. □

§19 November 21, 2019

Today's lecture is given by Ana Balibanu.

§19.1 Isometries, Positivity, and Polar Decomposition

For today, all of our vector spaces will be over $\mathbf{C}$ or $\mathbf{R}$, and they will all be inner product spaces. Let's begin with some motivation, by comparing our linear transformations to the complex numbers. Say we have $z \in \mathbf{C}$ a complex number, and $T : V \to V$ a linear transformation. The transformation T has an adjoint T^*, such that $(Tu, v) = (u, T^*v)$.

What is the complex number analog of the adjoint? The complex number analog of this is the complex conjugate of z, which is $\bar{z}$. We can see this by noting that $(\lambda u, v) = (u, \bar{\lambda} v)$. What is the complex number analog of a self-adjoint matrix? Recall that self-adjoint transformations are those for which $T = T^*$. Well, the adjoint corresponds to complex conjugation, so the self-adjoint condition corresponds to $\lambda = \bar{\lambda}$ for complex lambda. The complex numbers for which $\lambda = \bar{\lambda}$ are just the real numbers. That is, we can make the following table:

$z \in \mathbf{C}$	$T : V \to V$
conjugation $(z \mapsto \bar{z})$	taking adjoint $(T \mapsto T^*)$
real $(z \in \mathbf{R}, z = \bar{z})$	self adjoint $(T = T^*)$

That is, we can compare linear transformations to the complex numbers. Today, we'll fill in some more entries of this table. First, let's go over some definitions and examples.

Definition 19.1 — A transformation T is **positive** if it is self-adjoint and $(T(v), v) \geq 0$ for all $v \in V$.

Example 19.2

- The identity transformation id is positive.
- Projection operators are positive
- The map $T : \mathbf{R}^2 \to \mathbf{R}^2$ given by $T(x, y) = (-x, -y)$ is *not* positive. Take $((-x, -y), (x, y)) = -x^2 - y^2 < 0$ if x or y is nonzero.
- If $T : V \to V$ is any linear transformation, then T^*T is positive. To see this, note that $(T^*T(v), v) = (T(v), T(v)) = \|T(v)\|^2 \geq 0$, since norms are always greater than or equal zero.
- Suppose that T is orthogonally diagonalizable, with nonegative eigenvalues. Then T is positive. To see why this is true, note that by assumption, T has an orthogonal basis of eigenvectors $e_1, \ldots, e_n \in V$, with eigenvalues $\lambda_1, \ldots, \lambda_n \geq 0$. Let's take $v \in V$, and check that the condition of positivity is satisfied. We can write $v = a_1 e_1 + \cdots + a_n v_n$. Then

  $$(Tv, v) = (\lambda_1 a_1 e_1 + \cdots + \lambda_n a_n e_n, a_1 e_1 + \cdots + a_n e_n).$$

  Since the e_i are orthogonal, this simplifies to $(\lambda_1 a_1 e_1, a_1 e_1) + \cdots + (\lambda_n a_n e_n, a_n e_n)$. But this is

  $$\lambda_1 |a_1|^2 + \cdots + \lambda_n |a_n|^2 .$$

  Notice that that $|a_i|^2$ comes from $a_i \bar{a_i}$ (we need the conjugate), since a_i is complex. But since all the $\lambda_i \geq 0$, then $(Tv, v) \geq 0$ for all v.

Now, let's cover another definition motivated by the complex numbers.

Definition 19.3 — Let $T : V \to V$ be a linear transformation. A map $R : V \to V$ is said to be a **square root** of T if $R^2 = T$.

We expect that a linear transformation T will have multiple square roots. As an example, consider

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

The following are square roots of T:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$
$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 3/5 & -4/5 \\ -4/5 & 3/5 \end{pmatrix}.$$

In fact, there are an infinite number of square roots, since we can replace the $3, 4, 5$ in the last example with any Pythagorean triple. So far, it doesn't seem like the notion of a square root is extremely useful, since there are an infinite number of them and we don't quite know how these behave like our familiar square roots in the complex numbers. Recall the following property of square roots in $\mathbf{R}$: a number $r \in \mathbf{R}$ is nonnegative if and only if r has a nonnegative square root. This motivates the following proposition:

Theorem 19.4

The following are equivalent, for a linear map T:

i) T is positive.

ii) T is self-adjoint with nonnegative eigenvalues.

iii) There exists a positive $R : V \to V$ such that $R^2 T$.

iv) There exists $R : V \to V$ such that $R^* R = T$.

Proof.

- i) $\Rightarrow$ ii): By definition, T positive implies that T is self adjoint by definition. Suppose λ is an eigenvalue of T, so that $Tv = \lambda v$. Since T is positive, then $0 \leq (Tv, v) = (\lambda v, v) = \lambda \|v\|^2$. Since $\lambda \|v\|^2 \geq 0$, then $\lambda \geq 0$, so the eigenvalues are nonnegative.

- ii) $\Rightarrow$ iii): To prove this, we use the spectral theorem. Recall from last lecture that the spectral theorem tells us that T has an orthonormal eigenbasis $e_1, \ldots, e_n \in V$, with eigenvalues $\lambda_1, \ldots, \lambda_n$. By assumption, for each i, $\lambda_i \geq 0$. Define the transformation $R : V \to V$ by $R(e_i) = \sqrt{\lambda_i} e_i$. We take $\sqrt{\lambda_i}$ to be the positive square root of λ_i, which exists since $\lambda_i \geq 0$. This completely determines R, and we also have that $R^2(e_i) = \lambda_i e_i$. Since R^2 and T agree on a basis, the $R^2 = T$. This tells us that R has an orthonogmal eigenbasis, with nonegative eigenvalues. This implies, by the example above, that R is positive.

- iii) $\Rightarrow$ iv): Suppose that there exists positive $R : V \to V$ such that $R^2 = T$. This implies R is self adjoint, so $R = R^*$. Since $R^2 = T$, then $RR^* = T$.

- iv) $\Rightarrow$ i): Recall from the above example that any linear transformation of the form $R^* R$ is positive. But $R^* R = T$ so T is positive.

□

Now, we can add some more analogies between linear transformations and complex numbers to our table from above. The idea of positivity $R^* R = T$ for linear transformations corresponds to $s\overline{s} = z$ for some complex s, where z is our complex number. But this is the same as $z \geq 0$, so the positive linear transformations correspond to nonnegative complex numbers.

$z \in \mathbf{C}$	$T : V \to V$
conjugation $(z \mapsto \bar{z})$	taking adjoint $(T \mapsto T^*)$
real $(z \in \mathbf{R}, z = \bar{z})$	self adjoint $(T = T^*)$
nonnegative $(z \geq 0, z = \bar{s} \cdot s)$	positive $(T = R^* R)$

Now, another important definition, which has a strong geometric interpretation:

Definition 19.5 — A map $S : V \to V$ is an **isomoetry** if $\|Sv\| = \|v\|$ for all $v \in V$. That is, isomoetries are linear maps which preserve norms of vectors.

Rotations and reflections are examples of isomoetries, while projections, shears, and dilations are not in general isometries.

Theorem 19.6

The following are equivalent for a linear map S:

i) S is an isometry

ii) $(Su, Sv) = (u, v)$ for all $u, v \in V$.

iii) If $e_1, \dots, e_n \in V$ are orthonormal, then $Se_1, \dots, Se_n$ is also orthonormal.

iv) S is invertible and $S^{-1} = S^*$.

Proof.

- i) $\Rightarrow$ ii): We have that

$$\begin{aligned}(u, v) &= \frac{1}{4}\left(\|u+v\|^2 - \|u-v\|^2\right) \\ &= \frac{1}{4}\left(\|Su+Sv\|^2 - \|Su-Sv\|^2\right) \\ &= (Su, Sv),\end{aligned}$$

  so $(Su, Sv) = (u, v)$ as desired.

- ii) $\Rightarrow$ ii): We have that $e_1, \dots, e_n$ is orthonormal if and only if $(e_i, e_j) = \delta_{ij}$, the Kronecker delta. Then $(Se_i, Se_j) = (e_i, e_j) = \delta_{ij}$ as well, since $(Su, Sv) = (u, v)$ for any vectors u, v. This implies $Se_1, \dots, Se_n$ is orthonormal.

- iii) $\Rightarrow$ iv): We have that

$$\begin{aligned}(S^* Se_i, e_j) &= (Se_i, Se_j) \\ &= (e_i, e_j).\end{aligned}$$

  This implies that $(S^* Su, v) = (u, v)$ for any $u, v \in V$, which follows by writing u, v as linear combinations of the e_i. But then $(S^* S(u) - u, v) = 0$ for all $u, v \in V$. This means that $((S^* S - I)u, v) = 0$ for all $u, v \in V$. This implies that $S^* S - I$ is zero, which follows from zero detection from last lecture. But then $S^* S = I$, so $S^* = S^{-1}$.

- iv) $\Rightarrow$ i): We have that

$$\begin{aligned}\|S(v)\|^2 &= (Sv, Sv) \\ &= (S^* S, v) \\ &= (v, v) \\ &= \|v\|^2,\end{aligned}$$

as desired.

□

Now, the concept of an isomoetry allows us to add our final row to the table from the beginning. Isometries preserve norm. This corresponds to complex numbers with norm one, which lie on the unit circle.

$z \in \mathbf{C}$	$T : V \to V$
conjugation $(z \mapsto \bar{z})$	taking adjoint $(T \mapsto T^*)$
real $(z \in \mathbf{R}, z = \bar{z})$	self adjoint $(T = T^*)$
nonnegative $(z \geq 0, z = \bar{s} \cdot s)$	positive $(T = R^* R)$
norm one $(\lvert z \rvert = 1, \bar{b} \cdot b = 1)$	isometry $(S^* S = I)$

Now, recall that for any complex number z, we have a polar decomposition $z = re^{i\theta}$. Also, remember that $e^{i\theta} = \cos\theta + i \sin\theta$. The norm of this is $\sqrt{\cos^2\theta + \sin^2\theta} = 1$, so $e^{i\theta}$ is a complex number with norm 1. Our vector r is a nonnegative real number, so we can express $r = |z| = \sqrt{z \cdot \bar{z}}$. That is, we can write z as something like (nonnegative) $\cdot$ (norm one). Using the table of analogies above, we can ask whether we can do the same thing with linear transformations. In fact we can, which is the content of the polar decomposition theorem.

Theorem 19.7 (Polar Decomposition Theorem)

For any linear transformation $T : V \to V$, there exists an isometry $S : V \to V$ such that

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

where $\sqrt{T^* T}$ is the positive square root of $T^* T$.

§20 November, 26, 2019

§20.1 Singular Value Decomposition

Recall from last time that a linear map f is **positive** if and only if f has a self-adjoint square root, if and only if there exists g such that $f = g^* g = g \cdot g$. Also, recall that f is an isometry if it preserves norms: $\|fv\| = \|v\|$. On V, we can define $d(v, w) = \|v - w\|$, the **distance** between the two vectors. Then f is an isometry if $d(fv, fw) = d(v, w)$. Next, we'll cover singular value decomposition, which is somewhat analogous to the Iwasawa decomposition we saw earlier in the class: $\mathrm{SL}_2(\mathbf{R}) = K \times A \times U$. The idea behind singular value decomposition is that we can write a *complex number* as

$$a = \frac{z}{|z|} |z| = \frac{z}{|z|} \sqrt{\bar{z} \cdot z}.$$

We will use the following result in proving singular value decomposition.

Proposition 20.1

If $f : V \to V$ is a nonzero linear map, then there exists an isometry $g : V \to V$ such that $f = g\sqrt{f^* f}$, where for h positive, then $\sqrt{h}$ is a linear map with $(\sqrt{h})^2 = h$.

Proof. We first show that $\|fv\|^2 = \left\|\sqrt{f^*f}v\right\|^2$. We have

$$\begin{aligned}\|fv\|^2 &= (fv, fv)\\ &= (f^*fv, v)\\ &= (\sqrt{f^*f}\sqrt{f^*f}v, v)\\ &= (\sqrt{f^*f}v, \sqrt{f^*f}v)\\ &= \left\|\sqrt{f^*f}v\right\|^2,\end{aligned}$$

thus $\|fv\| = \left\|\sqrt{f^*f}v\right\|^2$. Then we define g' by $g'(\sqrt{f^*f}v) = fv$, where $g : \mathrm{im}(\sqrt{f^*f}) \to \mathrm{im}(f)$. Then we can extend g' to our desired map g by

$$\begin{array}{ccc}\mathrm{im}(\sqrt{f^*f}) & \xrightarrow{g'} & \mathrm{im}(f)\\ \downarrow{\scriptstyle \iota_1} & & \downarrow{\scriptstyle \iota_2}\\ V & \xrightarrow{\quad g \quad} & V\end{array}$$

Here the ι_i are embeddings. □

Definition 20.2 — If f is a linear map, then the **singular values** of f are the eigenvalues of $\sqrt{f^*f}$ counted with multiplicities. Denote this by $\mathrm{Sing}(f)$. Since $\sqrt{f^*f}$ is a *positive* linear map, then $\mathrm{Sing}(f) \subseteq \mathbf{R}_{\geq 0}$.

Example 20.3

Let f be defined by $f(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4)$. The eigenvalues are 3 and 2. But $f^*f(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4)$. Then $\sqrt{f^*f}(z_1, z_2, z_3, z_4) = (3z_1, 2z_2, 0, 3z_4)$. The eigenvalues of this are $\mathrm{Sing}(f) = \{3, 3, 2, 0\}$.

Now, we are ready to state the singular value decomposition.

Theorem 20.4 (Singular Value Decomposition (SVD))

Suppose that f is a map with $\mathrm{Sing}(f) = \{\lambda_1, \ldots, \lambda_n\}$. Then we can write

$$fv = \lambda_1(v, e_1)f_1 + \cdots + \lambda_n(v, e_n)f_n,$$

where $\{f_1, \ldots, f_n\}$ is an orthonormal basis.

Proof. First, note that $\sqrt{f^*f}e_j = \lambda_j e_j$, by definition of the singular values. Write $v = (v, e_1)e_1 + \cdots + (v, e_n)e_n$. Then

$$\sqrt{f^*f}v = \lambda_1(\vec{v}, \vec{e}_1)\vec{e}_1 + \cdots + \lambda_n(\vec{v}, \vec{e}_n)\vec{e}_n,$$

where we've used the $\vec{}$ notation to make it more clear what's a vector. Then

$$g\sqrt{f^*f} = \lambda_1(\vec{v}, \vec{e}_1)g(\vec{e}_1) + \cdots + \lambda_n(\vec{v}, \vec{e}_n)g(\vec{e}_n),$$

where g is an isometry, using the previous theorem. Since g is an isometry, we also have that $\{ge_i\}$ is an orthonormal basis. □

§20.2 Trace

We begin with the important definition of the characteristic polynomial. First recall that if $f : V \to V$ is a map over a complex vector space, then f is upper triangularizable. Then we define the **characteristic polynomial** $p_f(z)$ as

$$p_f(z) = (z - \lambda_1)^{m_1} \cdot \cdots \cdot (z - \lambda_k)^{m_k}.$$

This sits inside $\text{Poly}(\mathbf{C})$. Notice that p_f is independent of the upper triangularization.

Now, over the reals, we do not in general have an upper triangular presentation. We do this by what is called **descent**. We'll need a few definitions to do this.

Definition 20.5 — If V is a vector space over $\mathbf{R}$, then its **complexification** is the $\mathbf{C}$-vector space given by $V \otimes_{\mathbf{R}} \mathbf{C}$.

As a set, the complexification is just $V \times V$, and we can write $(v, w) = v + iw$. We have $\mathbf{R} \otimes_R \mathbf{C} \simeq \mathbf{C} \simeq \mathbf{R} \times \mathbf{R}$, as a set. Why is this true? You can think of the tensor product with the subscript as allowing us to move things over the tensor product, as long as they're in the subscript set. For example, in $\mathbf{R} \otimes_{\mathbf{R}} \mathbf{C}$, we can move a real number across the tensor product, but not an imaginary number. For example, we have $1 \otimes 3 = 3(1 \otimes 1) = 3 \otimes 1$, but we can't take the i out of $1 \otimes i$, since we'd need to scalar multiply by i, which we can't do since $\mathbf{R} \otimes_{\mathbf{R}} \mathbf{C}$ is a *real* vector space. Since we can move over anything in $\mathbf{R}$ through the tensor product $\otimes_{\mathbf{R}}$, $\mathbf{R} \otimes_{\mathbf{R}} V \simeq V$ for any vector space V.

The addition rules in the complexification are $(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$. The scalar multiplication rules are $(a + bi)(u + iv) = (au - bv) + i(av + bu)$. That is, we have

$$\begin{aligned}(u_1, v_1) + (u_2, v_2) &= (u_1 + u_2, v_1 + v_2) \\ (a + bi)(u, v) &= (au - bv, av + bu).\end{aligned}$$

Now, some more useful facts about tensor products to make things more concrete.

Proposition 20.6 (a) $V \otimes_{\mathbf{R}} \mathbf{C}$ is a $\mathbf{C}$-vector space and a $\mathbf{R}$-vector space.

(b) $\dim_{\mathbf{C}} V \otimes_{\mathbf{R}} \mathbf{C} = \dim_{\mathbf{R}} V$

(c) $\dim_{\mathbf{R}} V \otimes_{\mathbf{R}} \mathbf{C} = 2 \dim_{\mathbf{R}} V$.

Definition 20.7 — Let f be a linear map $V \to V$, where V is a real vector space. Then the **complexification** of f is defined as

$$f_{\mathbf{C}} := (f \otimes_{\mathbf{R}} \mathbf{C}) = fu + ifv.$$

Lemma 20.8

The complexification $f_{\mathbf{C}}$ is linear, and if λ is an eigenvalue for f, then λ is also an eigenvalue for $f_{\mathbf{C}}$.

The above definition will allow us to define the characteristic polynomial of f over a real vector space.

Definition 20.9 — The characteristic polynomial $p_f(x)$ of $f : V \to V$ over $\mathbf{R}$ is defined as

$$p_f(x) = p_{f_{\mathbf{C}}}(x),$$

where $f_{\mathbf{C}}$ is the complexification of f, and the $p_{f_{\mathbf{C}}}$ is the characteristic polynomial over the complex vector space as defined above.

In fact, the above definition works for any field. Since the characteristic polynomial is defined in terms of the complexification, we need to check that the coefficients are actually in $\mathbf{R}$. To this end, we have the following theorem.

Theorem 20.10

If $f : V \to V$, where V is a real vector space, then $p_f(x) \in \mathrm{Poly}(\mathbf{R})$. That is $p_f(x) = \alpha_n xn + \alpha_{n-1}x^{n-1} + \cdots + \alpha_0$, where the $\alpha_i \in \mathbf{R}$.

Proof. Suppose that λ is a nonreal eigenvalue of $f_{\mathbf{C}}$, then observe that $(z-\lambda)^m(z-\bar{\lambda})^m = (z^2 - 2\Re(\lambda)z + |\lambda|^2)^m$, for any m. It suffices to prove that eigenvalues of $f_{\mathbf{C}}$ come in pairs. That is, if λ is an eigenvalue of $f_{\mathbf{C}}$, then so is $\bar{\lambda}$, and they have the same multiplicity. If so, then

$$p_f(x) = (x-\lambda_1)^{m_1}\cdots(x-\lambda_k)^{m_k}(x-\mu_1)^{n_1}(x-\bar{\mu}_1)^{n_1}\cdots(x-\mu_l)^{n_l}(x-\mu_l)^{n_l},$$

where the λ_i are in $\mathbf{R}$, and the μ_l are in $\mathbf{C}$. Since the μ_l occur in conjugate pairs, then the polynomial has coefficients in $\mathbf{R}$, as desired. □

Now, let's talk about trace and determinant.

Definition 20.11 — The **trace** of a map f is defined as

$$\mathrm{Tr}(f) = \sum_{i=1}^{n} \lambda_i.$$

We will define the determinant $\det(f)$ next time. Whatever definition we use for $\det(f)$, note that we will have $\det(f - \lambda\mathrm{id}) = p_f(\lambda)$.

§21 December 3, 2019

§21.1 Trace (cont.)

Suppose that $f : V \to V$ is a linear map, and $\{e_i\}$ and $\{f_i\}$ are bases for V. We write $\mathcal{M}(f, \{e_i\}, \{f_i\})$ for the matrix of f expressed in $\{e_i\}$, $\{f_i\}$, meaning we have chosen the basis $\{e_i\}$ for the domain and $\{f_i\}$ for the codomain. We have that $\mathcal{M}(f, \{f_i\}, \{f_i\}) = \mathcal{M}(f, \{f_i\}, \{e_i\})$, and $\mathcal{M}(f, \{e_i\}, \{e_i\}) = \mathcal{M}(f, \{e_i\}, \{f_i\})$.

Now, some more facts about trace.

Proposition 21.1

For matrices A, B, we have $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

Proof. We have

$$(A_B)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj},$$

then we have $\mathrm{Tr}(A) = \sum_i A_{ii}$, so

$$\begin{aligned}\mathrm{Tr}(AB) &= \sum_i \sum_k A_{ik}B_{ki} \\ &= \sum_i \sum_i B_{ik}A_{ki} \\ &= \mathrm{Tr}(BA),\end{aligned}$$

by switching indices. □

Using this, suppose that $B = S^{-1}AS$. Then

$$\begin{aligned}\mathrm{Tr}(B) &= \mathrm{Tr}(S^{-1}AS) \\ &= \mathrm{Tr}((S^{-1}A)S) \\ &= \mathrm{Tr}(S(S^{-1}A)) \\ &= \mathrm{Tr}(A),\end{aligned}$$

which tells us that the trace of similar matrices are the same, i.e. the trace doesn't depend on basis.

As a corollary, if $f : V \to V$ is a linear map over a *complex* vector space, then $\mathrm{Tr}(f) = \sum k_i\lambda_i$, where λ_i are the eigenvalues, and k_i are the geometric multiplicities. This is because we can upper triangularize the matrix with the λ_i on the diagonal.

§21.2 Determinant

Recall from the first few days of class: a **permutation** on n letters is a bijection $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$. We can define the sign of a permutation as follows

Definition 21.2 — The **sign** of a permutation is defined as

$$\mathrm{sgn}(\pi) = (-1)^{v(\pi)},$$

where $v(\pi)$ is the number of pairs i, j such that $i < j$ and $\pi(i) > \pi(j)$. If $\mathrm{sgn}(\pi) = 1$, then we say that π is an **even** permutation, and if $\mathrm{sgn}(\pi) = -1$, we say that π is an **odd** permutation.

For example, if $\pi : \{1, 2\} \to \{1, 2\}$ with $\pi(1) = 2$ and $\pi(2) = 1$, then $\mathrm{sgn}(\pi) = -1$, since $1 < 2$ and $\pi(1) = 2 > 1 = \pi(2)$. What is the sign of the permutation $\pi : \{1, 2, 3\} \to \{1, 2, 3\}$ with $\pi(1) = 2, \pi(2) = 3, \pi(3) = 1$?

Now that we've defined sgn, we're ready for determinants.

Definition 21.3 — The **determinant** of an $n \times n$ matrix A is defined as

$$\det(A) = \sum_{\sigma \in \Sigma_n} \mathrm{sgn}(\sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)},$$

here Σ_n is the set of all permutations:

$$\Sigma_n = \{\pi : \{1, \dots, n\} \to \{1, \dots, n\} | \pi \text{ is a permutation}\}.$$

Then $A_{i,\sigma(i)}$ is the $(i,\sigma(i))$ component of the matrix A.

This is a strange way to think about the determinant, so let's do an example.

Example 21.4

Let A be the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Since A is 2×2, we use Σ_2 as the set of all permutations $\{1,2\} \to \{1,2\}$. There are only two of these permutations, σ_1 and σ_2, with

$$\sigma_1(1) = 1, \sigma_1(2) = 2$$
$$\sigma_2(1) = 2, \sigma_2(2) = 1.$$

We have $\operatorname{sgn}(\sigma_1) = 1$ and $\operatorname{sgn}(\sigma_2) = -1$. Then

$$\sum_{\sigma \in \Sigma_2} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} = (1) A_{11} A_{22} + (-1) A_{12} A_{21},$$

which tells us that the determinant of the matrix is $A_{11}A_{22} - A_{12}A_{21} = ad - bc$, which confirms the familiar identity.

Proposition 21.5

The determinant is multiplicative: $\det(AB) = \det(A)\det(B) = \det(B)\det(A) = \det(BA)$.

Proposition 21.6

If A is upper triangular, then $\det(A) = \lambda_1^{k_1} \cdots \lambda_m^{k_m}$, where the λ_i are the diagonal entries and the k_i are the multiplicities.

Proof. If σ is a permutation $\sigma \neq \mathrm{id}$, then there exists i such that $\sigma(i) < i$. But this means that for each of the $\sigma \neq \mathrm{id}$, then in the product

$$A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}$$

will be zero, since one of the $(i,\sigma(i))$ pairs will have $\sigma(i) < i$, which means $A_{i\sigma(i)}$ is in the lower triangle of the matrix, and therefore zero. Thus, the product is zero.

Now, the only nonzero contribution in the sum over permutations in the determinant is with $\sigma = \mathrm{id}$, which gives us

$$\det(A) = \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{n\sigma(n)} = A_{11} \cdots A_{nn} = \lambda_1^{k_1} \cdots \lambda_m^{k_m}.$$

□

As a corollary, we have the following

Corollary 21.7
The characteristic polynomial for a map f, $p_f(z)$, is $p_f(z) = \det(A - zI)$.

This is the end of the course material.

§21.3 Extras

Please excuse any mistake in these notes. Forrest is not extremely familiar with K-theory.

We will now discuss a very difficult question in linear algebra, who's solution is due to Voevodsky. Suppose that k is a field. Let Q_n be the number of isomorphism classes of finite dimensional vector spaces over k. The size of Q_n is $\aleph_0$, the size of the natural numbers (countable infinity).

Now, recall bilinear forms: $(V, (\cdot,\cdot))$. How many nondegenerate bilinear forms $(V, (\cdot,\cdot))$ are there? This is very hard.

Definition 21.8 — We define

$$K_0(k) = \{V/k : V \text{ finite dim}\}^{\text{gp}}/(\sim_{\text{iso}}),$$

where we mod out by the isomorphism class relation, and the superscript gp is the group completion, which means we formally add negatives. This is called an **algebraic K-theory**.

Definition 21.9 — The **Grothendieck-Witt** group $GW(k)$ is defined as

$$GW(k) = (\{(V, (\cdot,\cdot))\})^{\text{gp}}/(\sim_{\text{iso}}).$$

Example 21.10
- $GW(\mathbf{C}) \simeq \mathbf{Z}$, by the spectral theorem
- $GW(\mathbf{R}) \simeq \mathbf{Z} \oplus \mathbf{Z}$. Sylvester proved this in the 1800's
- $GW(\mathbb{F}_q) \simeq \mathbf{Z} \oplus \mathbf{Z}/2$, which follows from quadratic reciprocity.

Every field has a bilinear form given by

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

by using

$$\begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \big((\alpha, \beta)\big).$$

Now, we have something called the **fundamental square**.

$$\begin{array}{ccccccccc} 0 & \longrightarrow & I & \longrightarrow & GW(k) & \xrightarrow{\text{rank}} & K_0(k) \simeq \mathbf{Z} & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \sim} & & \downarrow{\scriptstyle /h} & & \downarrow & & \\ 0 & \longrightarrow & I & \longrightarrow & W(k) & \longrightarrow & \mathbf{Z}/2 & \longrightarrow & 0 \end{array},$$

where I is the kernel of the rank map. Now, there is a conjecture due to Milnor, which says

$$I^{\otimes n+1}(k)/I^{\otimes n}(k).$$

Voevodsky computes this explicitly using geometry.

Index