MATH 25A

Professor: Bena Tshishiku

Michele Tienni
Lowell House
Cambridge, MA 02138
micheletienni@college.harvard.edu

Please note that these notes are not official and they are not to be considered as a substitute for your notes. Specifically, you should not cite these notes in any of the work submitted for the class (problem sets, exams). The author does not guarantee the accuracy of the content of this document. If you find a typo feel free to email me.

Contents

Date: December 1, 2017.

1. 8-30

Linear algebra is the study of **linear maps** and vector spaces. Linear maps are maps satisfying the identity

$$\begin{aligned} T(cx) =& cT(x) \\ T(x+y) =& T(x)+T(y). \end{aligned}$$

Examples of linear maps include:

- Geometry: rotations of vectors in $\mathbb{R}^2$ (draw the picture!)
- Algebra: evaluation of polynomials, i.e.

$$T(p) = p(0).$$

- Analysis: derivative, i.e.

$$T = \frac{d}{dx}.$$

In fact,

$$\begin{aligned} \frac{d}{dx}(f+g) =& \frac{df}{dx} + \frac{dg}{dx} \\ \frac{d}{dx}(cf) =& c\frac{df}{dx}. \end{aligned}$$

In this course we will study linear finctions and the equations they appear in. For example, we will be asking if we can solve equations of the form $Tx = y$ (which will lead us into Gaussian elimination, matrix representations, rank-nullity), $Tx = x$ (which will lead us into eigenvectors, eigenvalues and the spectral theorem), and $T(U) \in U$ (generalized eigenvectors, Jordan normal form).

This course on inear algebra will be important in three ways:

- in its own right, as a mathematical theory;
- for its applications, e.g. in analysis;
- as an intro to abstraction an proof writing.

1.1. Sets.

The first topic of this course will be sets.

Definition 1.1. A **set** is a collection of things.

Example 1.2. Things like

$$\{\text{blue pandas}\}, \qquad \emptyset,$$

the latter being the **empty set** (the set with no elements).

Example 1.3. $\{1, 2, 4\}$ has three elements, {Millard Fillmore, the 13th US president} has one element (Fillmore was the 13th president).

The former example indicates that there are two typical ways to describe a set: either list its elements, or give a property that the elements satisfy (e.g. $\{1, -1\} = \{x \in \mathbb{R} \mid x^2 = 1\}$). It is not always easy to tell when two sets are equal (that is to say, they have the same elements). An example of sets that are *not* equal, we have

$$\begin{aligned} \{\{1,2\}\} \neq& \{1,2\} \\ \{\text{my cat}\} \neq& \{\text{cells in my cat}\}. \end{aligned}$$

Definition 1.4. A set is called **normal** if $X \notin X$. If $X \in X$ it is called abnormal.

The set of all things that are not squirrels is abnormal, because the set of all things that are not squirrels is not a squirrel. On the other hand the set of all things that are a squirrel is normal. Is $\{1, \{1\}\}$ normal? No, because its only elements are 1 and $\{1\}$. Let us define N as the set of normal sets. Is N normal? If N is normal, then it is contained in the set of all normal sets, and thus it is abnormal. If N is abnormal then it is contained in itself and thus it is normal. This inconsistency (N being neithe normal nor abnormal) is also known as **Russel's paradox** (1901), and its implications are that phrases like "the set of sets" cause some trouble. More precisely this indicates that our definition of what a set is is unsatisfactory. To avoid this issue we can fix a universe set U and only consider sets whose elements are in U. In this course we can just consider something like $\mathbb{R}^n$ for some n.

Definition 1.5. A is a subset of B (written $A \subset B$) if every element of A is an element of B.

Definition 1.6. For a set X, the power set $\mathcal{P}(X)$ is the set of subsets of X. Namely,

$$\mathcal{P}(X) = \{A \mid A \subset X\}.$$

Example 1.7.

$$\mathcal{P}(\{1,2\}) = \{\{1\}, \{2\}, \{1,2\}, \emptyset\}.$$

Remark. If a statement p is false, then the statement $p \implies q$ is true (**vacuously true**).

1.2. **Next time.** Next time we will explore the similarities between $\mathbb{R}$ and $\mathcal{P}(X)$.

2.1. **Algebraic Properties of Sets & Cardinality.**

2.1.1. *Basic Features of* $\mathbb{R}$.

(1) $\mathbb{R}$ is a field, namely we have operations

$$\begin{aligned} + &: \mathbb{R} \times \mathbb{R} \to \mathbb{R} \\ \cdot &: \mathbb{R} \times \mathbb{R} \to \mathbb{R} \end{aligned}$$

that satisfy the properties

$$\begin{aligned} (x+y)+z &= x+(y+z) \\ (xy)z &= x(yz) && \text{associativity} \\ x+y &= y+z \\ xy &= yx && \text{commutativity} \\ x+0 &= 0 \\ x\cdot 1 &= x && \text{existence of identity} \\ x+(-x) &= 0 \\ x(1/x) &= 1 && \text{existence of inverse} \\ x(y+z) &= xy+xz && \text{distributivity.} \end{aligned}$$

Other fields include $\mathbb{Q}$, but not $\mathbb{Z}$.

(2) $\mathbb{R}$ has a **total order relation** $\leq$, i.e. a relation that satisfies
- totality (either $x \leq y$ or $y \leq x$)
- antisymmetry ($x \leq y$ and $y \leq x$ implies $x = y$)
- reflexivity ($x \leq x$)
- transitivity ($x \leq y$ and $y \leq z$ implies $x \leq z$).

At this point we can ask ourselves the question: for a set X, to what extent does $\mathcal{P}(X)$ have similar properties?

2.1.2. *Unions and Intersections.* To answer this question we introduce the analogue of "addition" and "multiplication" of sets. Fix a set X. For subsets $A \subset X$ and $B \subset X$ we define the **union** to be

$$A \cup B = \{x \in X \mid x \in A \text{ or } x \in B\}$$

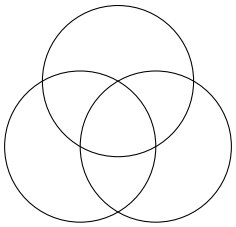
and the **intersection**

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

We can view union and intersection as operations on $\mathcal{P}(X)$ (namely functions $\mathcal{P}(X) \times \mathcal{P}(X) \to \mathcal{P}(X)$).

Lemma 2.0.1 (Distributive Law)**.** *For subsets $A, B, C \in X$ we have*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$



Proof. For an informal proof, fill the areas corresponding to the sets in the statement. For a formal proof:

$$\begin{aligned} x \in A \cap (B \cup C) \iff & x \in A \text{ and } x \in B \cup C \\ \iff & \text{ either } x \in A \text{ and } x \in B \\ & \quad \text{or } x \in A \text{ and } x \in C \\ \iff & x \in A \cap B \text{ or } x \in A \cap C \\ \iff & x \in (A \cap B) \cup (A \cap C). \end{aligned}$$

□

Corollary 2.0.1.

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Proof. We apply distributivity to the right hand side.

$$\begin{aligned} (A \cup B) \cap (A \cup C) =& (A \cap A) \cup (A \cap B) \cup (A \cap C) \cup (B \cap C) \\ =& A \cup (B \cap C) \qquad \text{because } A \cap B, A \cap C \subset A \cap A = A. \end{aligned}$$

□

2.1.3. *Subsets.* We list some properties of "$\subset$":

- $A \subset A$ for every set A;
- $A \subset B$ and $B \subset A$ then $A = B$;
- $A \subset B$ and $B \subset C$ implies $A \subset C$.

This allows us to define a **partial order** (i.e. a total order without totality).

2.1.4. *Other Useful Definitions.*

Definition 2.1. For $A \subset X$, the **complement** A^c is defined as

$$A^c = \{x \in X \mid x \notin A\}.$$

For sets X, Y the **product of sets** $X \times Y$ is defined as

$$X \times Y = \{(x, y) \mid x \in X, y \in Y\}.$$

2.1.5. *Cardinality.*

Definition 2.2. A **mapping** between sets $f : X \to Y$ is an assignment to each $x \in X$ of an element $f(x) \in Y$.

Some special properties f may have:

- **injectivity**: f is injective if $f(x) = f(x')$ implies $x = x'$;
- **surjectivity**: f is surjective if for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$;

- **bijectivity**: f is bijective if it is injective and surjective.

The existence of an injection or a surjection (or both) between two sets tells us something about the size of said sets.

Theorem 2.3 (Fundamental Theorem of Caveman Mathematics)**.** *If X, Y have finitely many elements then there exists a bijection $f : X \to Y$ if and only if X and Y have the same number of elements.*

It looks like this should be downright obvious, but from the point of view of rigorous mathematics this is not completely obvious, so the proof is left as an exercise in the homework. We will need the following definition:

Definition 2.4. Two sets X, Y have the same **cardinality** if there is a bijection $f : X \to Y$.

Theorem 2.5 (Cantor)**.** *Fix any set X. Then* $\mathrm{X} \neq \mathrm{card}(\mathcal{P}(X))$*, i.e. there is no bijection $f : X \to \mathcal{P}(X)$.*

Example 2.6. Suppose $X = \{1, \ldots, n\}$ is finite. Then the number of elements in $\mathcal{P}(X)$ is 2^n (to choose $A \subset X$ choose for each j if $j \in A$ or $j \notin A$). Can n and 2^n be equal? We can start wiriting the values in a table, and this might suggest us that it is not the case. But that is not quite a proof, so how can we (im)prove this? We are going to do so by induction.

Lemma 2.6.1.

$$n < 2^n$$

Proof. By induction. The base case is just $1 < 2$. We now go to the induction hypothesis, namely we show that if the lemma is true for n then it is true for $n + 1$ as well. Let us assume then that $n < 2^n$. Then

$$\begin{aligned} 2^{n+1} =& 2 \cdot 2^n \\ >& 2 \cdot n \\ >& n + 1 \text{ for } n > 1. \end{aligned}$$

□

3.1. **Cardinality and Countability.**

3.1.1. *Countable Sets.*

Definition 3.1. A set X is **countably infinite** if there is a bijection $f : \mathbb{N} \to X$ (where $\mathbb{N} = \{1, 2, 3, \dots\}$ are the natural numbers).

In the terminology from last class, X has the same cardinality as $\mathbb{N}$, i.e. $\text{card}(\mathbb{N}) = \text{card}(X)$.

Definition 3.2. A set is **countable** if it is either finite or countably infinite. Otherwise it is called **uncountable**.

Example 3.3. (1) $\mathbb{Z}$ is countably infinite. To show this we need a bijection $f : \mathbb{N} \to \mathbb{Z}$, and we can just define

$$f(n) = \begin{cases} n/2 & n \text{ even} \\ -\frac{n-1}{2} & n \text{ odd} \end{cases}.$$

(2) Let

$$P = \{p \in \mathbb{N} : p \text{ is prime}\}$$

(where $p \in \mathbb{N}$ is prime if it can't be written as $p = ab$ with $a, b \in \mathbb{N} - \{1\}$). A basic fact about numbers is that every nonprime number has a prime divisor. We claim that P is countably infinite. Let us consider a flawed proof first.

Flawed proof. Define $f : \mathbb{N} \to P$ by $f(n) = n$th smallest prime number.

- f is a surjection: given p, $n =$ number of primes $\leq p$, $f(n) = p$;
- f is injective: nth¡$n + 1$st primes are different.

□

The proof here fails at the injective step, because if there are finitely many primes our proof for the injection is false. Thus we need to prove that there are infinitely many primes.

Correct proof. We now show that P is infinite. This proof goes back to Euclid in 300 BC. Suppose $P = \{p_1, \dots, p_n\}$ is finite. Consider

$$q = p_1 \times \cdot \times p_n + 1.$$

We see that $q > p_i$ for all i so $q \notin P$. Then we must have that p_i divides q for some i. But then since p_i divides $p_1 \cdots p_n$ we must have that p_i divides $q - p_1 \cdots p_n = 1$, and this is a contradiction. □

Remark. We see that to say that X is countable means that we can list the elements of X as $X = \{x_1, x_2, \dots\}$. In practice to prove that X is countable it is often easier to list elements than to define the bijection $\mathbb{N} \to X$ (which is equivalent).

(3) Consider

$$\mathbb{Q}_+ = \left\{ \frac{a}{b} \middle| a, b \in \mathbb{N} \right\}.$$

We want to prove it is countable. We can fill the lattice $(n, m) \in \mathbb{N} \times \mathbb{N}$ where a/b is sent to (a, b) and list the diagonals so as to get a list

$$\mathbb{Q}_+ = \Big\{ \underbrace{\frac{1}{1}}_{a+b=2}, \underbrace{\frac{2}{1}, \frac{1}{2}}_{a+b=3}, \underbrace{\frac{3}{1}, \frac{2}{2}, \frac{1}{3}}_{a+b=4}, \dots \Big\}$$

3.1.2. *Uncountable sets.* Recall

Theorem 3.4 (Cantor)**.** *Fix any set X. Then* $\mathrm{X} \neq \mathrm{card}(\mathcal{P}(X))$*, i.e. there is no bijection $f : X \to \mathcal{P}(X)$.*

Corollary 3.4.1. *$\mathcal{P}(\mathbb{N})$ is uncountable.*

Corollary 3.4.2. *None of the sets $\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\mathcal{P}(\mathbb{N}))$ have the same cardinality.*

Remark. There is an injection

$$\begin{aligned} X &\to \mathcal{P}(X) \\ a &\mapsto \{a\} \end{aligned}$$

so the main content of the theorem is that there is no surjection $X \to \mathcal{P}(X)$.

There are several cases for this theorem. The first is the case where X is finite. Then we have the countable case, which we will prove here, and last you will prove the general case in the homework.

Cantor's theorem for countably infinite sets. Let assume $X = \{x_1, \dots, \}$ is countably infinite. We want to prove there is no surjection $X \to \mathcal{P}(X)$. First we observe that a subset $A \subset X$ can be encoded as a sequence of $0'$s and $1'$s. For example the subset $\{x_2, x_3, x_6\}$ corresponds to the sequence

$$01100100 \cdots$$

and X corresponds to

$$1111 \cdots$$

To show that there is no surjection we will show that for every $f : X \to \mathcal{P}(X)$ there exists $B \in \mathcal{P}(X)$ so that $f(x) \neq B$ for every $x \in X$ (what is called a **diagonal argument**). Given $f : X \to \mathcal{P}(X)$ write $f(x_i)$ as a sequence (as previously described), so that

$$f(x_i) = w_{i1} w_{i2} \cdots w_{ij} \cdots$$

where $w_{ij} = 0, 1$. Consider now the sequence $z = z_1 z_2 \cdots$ where

$$z_i = \begin{cases} 1 & \text{if } w_{ii} = 0 \\ 0 & \text{if } w_{ii} = 1 \end{cases}.$$

Then z differs from $f(x_i)$ in the ith coordinate, and thus the subset B corresponding to z satisfies $f(x_i) \neq B$ for every $i \in \mathbb{N}$. □

Remark. A variance of Cantor's diagonalization argument shows that there is no surjection $\mathbb{N} \to [0, 1] = \{x \in \mathbb{R} \colon 0 \leq x \leq 1\}$. You can do this by expanding $x \in [0, 1]$ as a decimal with binary digits. However, the same argument does *not* show that there is no surjection $\mathcal{P}(X) \to \mathcal{P}(\mathcal{P}(X))$ since you can't list elements of $\mathcal{P}(X)$.

4.1. **Equivalence Relations.**

4.1.1. *Partitions and Equivalence Relations.*

Example 4.1. Consider the set

$$P = \{\text{people in Math 25a}\}$$

together with the subsets

$$P_i = \{\text{people in } P \text{ born in the } i\text{th month}\}.$$

These sets are *disjoint* (i.e. $P_i \cap P_j = \emptyset$ if $i \neq j$), they are nonempty and every person in P is in some P_i.

Definition 4.2. A **partition** of a set X is a collections of subsets $X_i \subset X$ such that

(a) $X_i \neq \emptyset$ for all i
(b) $i \neq j$ implies $X_i \cap X_j = \emptyset$
(c) $X = \bigcup_i X_i$.

Example 4.3. $\{2\}, \{1, 3\}$ is a partition of $\{1, 2, 3\}$, but not $\{1\}, \{3\}$ or $\{1, 3\}, \{2, 3\}$.

We can see this from another point of view. For $x, y \in P$ write $x \sim y$ if x and y are born in the same month. This relation is

(a) reflexive: $x \sim x$ for every x;
(b) symmetric: $x \sim y$ implies $y \sim x$;
(c) transitive: $x \sim y$ and $y \sim z$ implies $x \sim z$.

Definition 4.4. A relation $\sim$ on X that satisfies (a),(b), and (c) above is called an **equivalence relation**.

Remark. A relation is a subset $R \subset X \times X$. However instead of writing $(x, y) \in R$ one usually writes $x \sim y$.

Definition 4.5. For a relation $\sim$ on X and $y \in X$ we define the **equivalence class** of y as

$$[y] = \{x \in X \colon x \sim y\}.$$

The equivalence classes give a collection of subsets of X.

Example 4.6. We have that

$$[\text{Bena}] = \{\text{people born in May}\}$$

Lemma 4.6.1. *A relation $\sim$ on X is an equivalence relation if and only if the equivalence classes $\{[x] \colon x \in X\}$ form a partition of X.*

Thus equivalence relations and partitionis are the same idea, just packaged in two different ways.

As a fun fact, in a room of 60 people there is a 99.4 chance that 2 people have the same birthday.

4.1.2. *More Examples.*

Example 4.7. *Modular arithmetic.* $X = \mathbb{Z}$. define $x \sim y$ is $x - y$ is even. Then there are two equivalence classes: $[0] = \{\text{even numbers}\}$ and $[1] = \{\text{odd numbers}\}$. These two sets form a partition so $\sim$ is an equivalence relation. Note that there is no difference in writing $X = [0] \cup [1]$ or $X = [8] \cup [3]$. We see that addition on the integers defines an addition on $\{[0], [1]\}$ according to the rule

$$[a] + [b] := [a + b].$$

For example

$$\begin{aligned}[8] + [10] =& [18] = [0]\\ [40] + [11] =& [51] = [1].\end{aligned}$$

However there is an issue of well-definedness. Why does the definition of addition not depend on how we write our equivalence classes? In fact we do not want the rule for addition to depend on whether we write [0] or [8] or anything else. Fortunately this is not the case (addition does not the choice of representatives), since

$$\begin{aligned}\text{even} + \text{even} &= \text{even}\\ \text{odd} + \text{odd} &= \text{even}\\ \text{odd} + \text{even} &= \text{odd}\,.\end{aligned}$$

We can also multiply equivalence classes in the same way, i.e.

$$[a] \cdot [b] = [a \cdot b],$$

and these operations make $\{[0], [1]\}$ into a field! This will become useful later on when we talk about vector spaces.

Example 4.8. *Rational numbers.* Let

$$X = \{(a, b) \colon a, b, \in \mathbb{Z}, b \neq 0\}$$

and define

$$(a, b) \sim (a', b') \iff ab' = a'b.$$

Then we can write

$$\frac{a}{b} := [(a, b)].$$

Then the set of equivalence classes is simply

$$\mathbb{Q} = \left\{ \frac{a}{b} \colon (a, b) \in X \right\}$$

and you can (should!) check that

$$\begin{aligned}\text{addition:} \quad & \frac{a}{b} + \frac{a'}{b'} = \frac{ab' + a'b}{bb'}\\ \text{multiplication:} \quad & \frac{a}{b} \cdot \frac{a'}{b'} = \frac{aa'}{bb'}\end{aligned}$$

are well-defined operations on $\mathbb{Q}$ and turn it into a field.

Example 4.9. *Cardinality equivalence.* Fix a set X. This works for any set, but for this time we consider $X = \mathbb{R}$. Define $\sim$ on $\mathcal{P}(X)$ as

$$A \sim B \iff \text{card}(A) = \text{card}(B)$$

(that is to say, if there exists a bijection $f : A \to B$). For example, consider $A = \{1, 2, 3\}$. The equivalence class of A is then the collection of $3-$element subsets of $\mathbb{R}$. On the other hand, $[\mathbb{N}]$ is the collection of all countably infinite subsets of $\mathbb{R}$, and we know for example that $\mathbb{Z}, \mathbb{Q} \in [\mathbb{N}]$ and $[0, 1] \notin [\mathbb{N}]$ (where the brackets on the left hand side mean a different thing, i.e. closed interval).

A couple questions for next time. For $x, y \in \mathbb{R}$ let

$$(x, y) = \{z \in \mathbb{R} \colon x < z < y\}.$$

The questions are:

(1) $\text{card}(-1, 1) = \text{card}(\mathbb{R})$?
(2) $\text{card}((0, 1) \times (0, 1)) = \text{card}(0, 1)$?

The answer to the first question is yes, and we can define the bijection as

$$x \mapsto \frac{x}{1 - x^2}.$$

5.1. **Cardinality.**

5.1.1. *Cardinalities of Uncountable Sets.*

Example 5.1. Let

$$\begin{aligned} B =& \{x \in \mathbb{R} \colon -1 < x < 1\} \\ A =& B \cup \{1, 2, 3, 4\}. \end{aligned}$$

Does $\text{card}(A)$ equal $\text{card}(B)$?

Remark. If B was countably infinite, this would follow from homework 2, problem 2.

We will see that $\text{card}(A) = \text{card}(B)$. But how do we define the bijection $f : A \to B$?

Remark. So far the only way to show two sets have the same cardinality is to construct an explicit bijection between them. The following theorem is a nonconstructive tool for showing two sets have the same cardinality.

Theorem 5.2 (Schroeder-Bernstein)**.** *Suppose X, Y are sets and there exists injections*

$$\begin{aligned} f : X \to& Y \\ g : Y \to& X. \end{aligned}$$

Then there exists a bijection $F : X \to Y$.

Remark. If we define $\text{card}(X) \leq \text{card}(Y)$ to mean the existence of an injection from X to Y we can rephrase this theorem as

$$\text{card}(X) \leq \text{card}(Y) \text{ and } \text{card}(Y) \leq \text{card}(X) \implies \text{card}(X) = \text{card}(Y).$$

Example 5.3. Take A, B as above. Since $B \subset A$ there is an injecction $g : B \to A$ defined by $b \mapsto a$. define $f : A \to B$ by $a \mapsto a/5$. If f, g are injective it follows from the above theorem that there exists a bijection $F : A \to B$.

Example 5.4. Let $Y = \{y \in \mathbb{R} \colon 0 < y < 1\}$ and $X = Y \times Y$. We claim that $\text{card}(X) = \text{card}(Y)$. It is easy to find an injection $g : Y \to X$, for example $y \mapsto (y, 1/2)$. For the other direction we can use decimal expansion (since every number $y \in Y$ has a unique decimal expansion, $y = 0.y_1y_2 \ldots$ where $0 \leq y_i \leq 9$ for $i \in \mathbb{N}$ that does not end in repeating $9'$s so that we have uniqueness). Then given $(a, b) \in X$ we write $a = 0.a_1a_2 \ldots, b = 0.b_1b_2 \ldots$ and define $f : X \to Y$ with

$$(a, b) \mapsto 0.a_1b_1a_2b_2 \ldots.$$

To prove it is injective, consider $(a, b), (a', b') \in A \times B$ such that $f(a, b) = f(a', b')$. This means that they have the same decimal expansion, and since this is unique we have that $(a, b) = (a', b')$. So f, g are injective, and by Schroeder-Bernstein we have an injection $F : X \to Y$.

Remark. The function f defined above is not surjective, since $0.9090909 \ldots$ is not the image of any (a, b) since a would need to end in repeated $9'$s.

Proof. We are given X, Y and the respective injections $f : X \to Y$ and $g : Y \to X$. As a simplifying assumption we assume $Y \subset X$ and $g(y) = y$. This is a harmless assumption since any injection defines a bijections between its range and domain. We are now going to define a bijection $F : X \to Y$.

Define $f(X) = \{f(a) \colon a \in X\}$. By definition $f(X) \subset Y$ but by definition $f(X) \subset X$. We similarly define

$$f^k(X) = \{f^k(a) \colon a \in X\}.$$

for $k \in \mathbb{N}$. Observe that $f^{k+1}(X) \subset f(X)$. This is because

$$\begin{aligned} b \in f^{k+1}(X) &\Longrightarrow b = f^{k+1}(a) = f^k(f(a)) \text{ for some } a \in X \\ &\Longrightarrow b \in f^k(a). \end{aligned}$$

For each $z \in X$ there are 3 possibilities:

(1) $z \in f^k(X)$ for all $k \in \mathbb{N}$.
(2) There is some $k \in \mathbb{N}$ such that $z \in f^k(X) - f^{k+1}(X)$. In this case
   (a) $z \in f^k(X) - f^k(Y)$, or
   (b) $z \in f^k(Y) - f^{k+1}(X)$.

These three cases are mutually exclusive. Thus they give a descrption of X as a disjoint union

$$X = X_1 \cup X_{2(\mathrm{a})} \cup X_{2(\mathrm{b})}.$$

where sets are defined according to their respective case, e.g.

$$X_{2(\mathrm{a})} = \left\{z \in X \colon z \in f^k(X) - f^k(Y) \text{ for some } k \geq 0\right\}$$

with the convention $f^0(X) = X$. Check that

$$Y = f(X_1) \cup f(X_{2(\mathrm{a})} \cup X_{2(\mathrm{b})}).$$

From this we can define the bijection $F : X \to Y$ by

$$F(z) = \begin{cases} f(z) & \text{if } z \in X_1 \cup X_{2(\mathrm{a})} \\ z & \text{if } z \in X_{2(\mathrm{b})} \end{cases}.$$

To show F is bijective, we start by proving F is injective. Suppose $F(z_1) = F(z_2)$. There are several cases:

- $z_1, z_2 \in X_1 \cup X_{2(\mathrm{a})}$. Then
$$f(z_1) = F(z_1) = F(z_2) = f(z_2)$$
and therefore $z_1 = z_2$ since f is injective.

Fill in the remaining cases. To prove F is surjective we see that

$$\begin{aligned} F(X_1 \cup X_{2(\mathrm{a})} \cup X_{2(\mathrm{b})}) &= F(X_1) \cup F(X_{2(\mathrm{a})}) \cup F(X_{2(\mathrm{b})}) \\ &= f(X_1) \cup f(X_{2(\mathrm{a})}) \cup X_{2(\mathrm{a})} \\ &= Y. \end{aligned}$$

□

A very good exercise to understand this proof:

(1) Use SB to prove there is a bijection $[0, 1) \to (0, 1)$;

(2) Follow the proof to actually construct the bijection.

6. 9-13

6.1. **Vector Spaces.**

6.1.1. *Formal Definition.*

Definition 6.1. A **field** is a set F with operations

$$\dagger : F \times F \to F \qquad \star : F \times F \to F$$

that satisfy

$$\begin{aligned} x \dagger y &= y \dagger x & (x \dagger y) \dagger z &= x \dagger (y \dagger z) \\ x \dagger 0 &= 0 \dagger x = x & x \dagger (-x) &= 0 \\ x \star y &= y \star x & (x \star y) \star z &= x \star (y \star z) \\ x \star 1 &= 2 \star x = x & x \star (x^{-1}) &= 1 \\ x \star (y \dagger z) &= x \star y \dagger x \star z \end{aligned}$$

Example 6.2. $\mathbb{R}$ with $\dagger = +, \star = \cdot$.

Example 6.3. $\mathbb{Q}$ with the same operations as $\mathbb{R}$.

Example 6.4. $\mathbb{C}$ with the same operations.

Example 6.5. $\mathbb{Z}/p\mathbb{Z}$ with p prime, with operations

$$\begin{aligned} [a] \dagger [b] &= [a + b] \\ [a] \star [b] &= [ab] \end{aligned}$$

There are some nonexamples, for example $\mathbb{Z}$ and $\mathbb{Z}/4\mathbb{Z}$. These have no inverses for some elements.

Remark. We usually write all field operations addition $+$ and multiplication $\cdot$.

Definition 6.6. Let F be a field. A **vector space over** F is a set V with

(a) an addition $+ : V \times V \to V$ that is commutative, associative, has an identity and has inverses;
(b) a scalar multiplication $\cdot : F \times V \to V$ that is associative, namely $(a \cdot b) \cdot v = a \cdot (b \cdot v)$ for all $a, b \in F$ and $v \in V$, distributive, namely

$$\begin{aligned} a \cdot (v + w) &= a \cdot v + a \cdot w \\ (a + b) \cdot v &= a \cdot v + b \cdot v \end{aligned}$$

and satisfies $1 \cdot v = v$ for all $v \in V$.

Remark.
- The elements of V are called **vectors**, and the elements of F are called **scalars**.
- Property (i) means that V is an **abelian group**.

6.1.2. *Examples.*

Example 6.7. Let F be a field. Then

$$F^n = \{(x_1, \ldots, x_n) \mid x_i \in F\}$$

with operations

$$\begin{aligned}(x_1, \ldots, x_n) + (y_1, \ldots, y_n) &= (x_1 + y_1, \ldots, x_n + y_n)\\ a(x_1, \ldots, x_n) &= (ax_1, \ldots, ax_n)\end{aligned}$$

is a vector space because F is a field (check this). In particular, $\mathbb{R}^2$ (vectors in a plane) and $\mathbb{R}^3$ (vectors in space) are vector field. For $F = \mathbb{Z}/2\mathbb{Z}$ we can visualize F^3 as the vertices of a cube.

Example 6.8. If F is a field, the set $\mathrm{Poly}(F)$ of polynomials on F is a vector space over F with the usual addition and scalar multiplication. Specifically so is $\mathrm{Poly}_d(F)$, the space of all polynomials of degree at most d.

Remark. The mapping

$$\begin{aligned}F^{d+1} &\to \mathrm{Poly}_d(F)\\ (a_0, \ldots, a_d) &\mapsto a_d x^d + \cdots + a_1 x + a_0\end{aligned}$$

is a bijection.

Example 6.9. For a set S, define $\mathrm{Fun}(S, F)$ to be the set of all functions $f : S \to F$. This is a vector space with operations $(f + g)(s) = f(s) + g(s)$ and $(cf)(s) = c \cdot f(s)$ for all $c \in F, f, g \in \mathrm{Fun(S)}, s \in S$. For example let $S = [0, 1], F = \mathbb{R}$. For another case let $S = \{1, \ldots, n\}$. The mapping

$$\begin{aligned}\mathrm{Fun}(S, F) &\to F^n\\ f &\mapsto (f(1), \ldots, f(n))\end{aligned}$$

is a bijection.

6.1.3. *Subspaces.* Let V be a vector space over F.

Definition 6.10. $W \subset V$ is a **subspace** if

- W is closed under vector addition, namely

$$w_1, w_2 \in W \Rightarrow w_1 + w_2 \in W$$

- W is closed under scalar multiplication, i.e.

$$a \in F, w \in W \Rightarrow aw \in W.$$

If W is a subspace of V the operations on V make W a vector space.

Example 6.11. $V = \mathbb{R}^2$.Let us check whether or not

$$\begin{aligned}W_1 &= \{(x, y) \colon y = 10x\}\\ W_2 &= \{(x, y) \colon xy \leq 0\}\end{aligned}$$

are subspaces. For W_1 we see that

$$\begin{aligned}(x_1, y_1), (x_2, y_2) \in W_1 \Rightarrow & 10(x_1 + x_2) = 10x_1 + 10x_2 = y_1 + y_2 \\ \Rightarrow & w_1 + w_2 \in W\end{aligned}$$

and similarly W is closed under scalar multiplication. Therefore W_1 is a subspace. On the other hand, pick $(0, 1), (1, 0) \in W_2$. Then

$$(0, 1) + (1, 0) = (1, 1) \notin W_2.$$

This shows that W_2 is *not* closed under addition, but it is closed under multiplication since

$$c(x, y) = (cx, cy)$$

and $(cx)(cy) = c^2xy \leq 0$ since $c^2 \geq 0$.

Lemma 6.11.1. *If $W \subset V$ is a subspace of V then $0 \in W$.*

Proof. This follows from the following lemma.

Lemma 6.11.2. *For every $v \in V$ we have $0 \cdot v = 0$. The first 0 is the additive identity on F and the second 0 is the additive identity on V.*

Proof. We have

$$\begin{aligned}0v = & (0 + 0)v \\ = & 0v + 0v\end{aligned}$$

and thus

$$\begin{aligned}0 = & 0v + (-0v) \\ = & 0v + 0v + (-0v) \\ = & 0v.\end{aligned}$$

□

As a warm-up exercise, prove $-1v = -v$.

□

7.1. **Basic Properties of Vector Spaces.**

- In a field F we always have an additive identity $0_F \in F$ and a multiplicative identity $1_F \in F$ and its additive inverse $-1_F \in F$. We usually drop the subscript when it is clear from the context.
- In a vector space V we always have an additive identity $0_V \in V$ and for every $v \in V$ there is an additive inverse $-v \in V$.

Lemma 7.0.1. *Let V be a vector space over F. Then for every $v \in V$ and $c \in F$*

(1) $0_F \cdot v = 0_V$*;*
(2) $(-1)_F \cdot v = -v$*;*
(3) $c \cdot 0_V = 0_V$*;*
(4) $c \cdot v$ *implies that either* $c = 0$ *or* $v = 0$*.*

Proof. (1) We already proved (a) last class.
(2) We write

$$\begin{aligned} 0_V =& 0_F \cdot v \\ =& (1 + (-1)) \cdot v \\ =& 1 \cdot v + (-1) \cdot v \end{aligned}$$

(3) Exercise.
(4) If $c = 0$ we are done. If $c \neq 0$ then

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

□

7.1.1. *Operations on Subspaces.* Recall that a subspace $U \subset V$ is a subset that is closed under addition and scalar multiplication. Today we are going to exhibit operations on subspaces analogous to $\cup, \cap, ^c$ on sets.

Let $U_1, U_2 \subset V$ be subspaces. Then $U_1 \cap U_2$ is a subspace of V (this is a homework assignment). The union $U_1 \cup U_2$ is typically not a subspace.

Definition 7.1. For $v \in V$ define the **span** of a vector as

$$\operatorname{span}(v) = \{a \cdot v \colon a \in F\}.$$

This is a subspace of V. In fact

$$\begin{aligned} av + a'v =& (a + a')v \\ a'(av) =& (a'a)v. \end{aligned}$$

Example 7.2. For an example for which the union is not a subset, let $V = \mathbb{R}^2$. Let

$$U_1 = \operatorname{span}((1,0)) \qquad U_2 = \operatorname{span}((1,1)).$$

Then for example $(1,0) + (1,1) = (2,1) \notin U_1 \cup U_2$.

We can see how unions, despite being closed under scalar multiplication, are not closed under addition.

Definition 7.3. Define the **sum** as

$$U_1 + U_2 = \{u_1 + u_2 \colon u_1 \in U_1, u_2 \in U_2\}.$$

The sum is a subspace since

$$(u_1 + u_2) + (u_1' + u_2') = (u_1 + u_1') + (u_2 + u_2') \in U_1 + U_2.$$

Note that $U_1, U_2 \subset U_1 + U_2$.

We can similarly define the sum of multiple subspaces $U_1, \ldots, U_k$ as

$$U_1 + \cdots + U_k = \{u_1 + \cdots + u_k : u_1 \in U_1, \cdots, u_k \in U_k\}.$$

This is a subspace.

Example 7.4. For $v_1, \ldots, v_k \in V$ we have

$$\operatorname{span}(v_1) + \cdots + \operatorname{span}(v_k) = \{a_1 v_1 + \cdots + a_k v_k \colon a_1, \ldots, a_k \in F\}.$$

A sum of the form $a_1 v_1 + \cdots + a_k v_k \colon a_1, \ldots, a_k$ is called a linear combination of $v_1, \ldots, v_k$. If

$$V = \operatorname{span}(v_1) + \cdots + \operatorname{span}(v_k)$$

we say that $v_1, \ldots, v_k$ span V, and V is finite dimensional.

Definition 7.5. If $V = U_1 + U_2$ and $U_1 \cap U_2 = \{0\}$ then we write

$$V = U_1 \oplus U_2$$

and say that V is the **direct sum** of U_1 and U_2.

Remark. This is the analogue of disjoint union for sets.

Example 7.6. (1) $V = \mathbb{R}^3$. Then

$$\operatorname{span}((1,0)) + \operatorname{span}((0,0,1)) = \{(a,0,b) \colon a, b \in \mathbb{R}\}.$$

(2) We have that

$$\mathbb{R}^3 = \{(x,y,0)\} + \{(0,y,z)\}.$$

However this is not a direct sum since $(0,1,0)$ is in both subspaces.

(3) Every polynomial can be written uniquely as the sum of a polynomial with only odd exponents and one with only even exponents. For example

$$a_5 x^5 + \cdot + a_0 = (a_5 x^5 + a_3 x^3 + a_1 x) + (a_4 x^4 + a_2 x^2 + a_0).$$

Therefore

$$\operatorname{Poly}(F) = U_{odd} \oplus U_{even}.$$

This is because U_{odd} and U_{even} are subspaces.

Remark. If $V = U_1 + U_2$ then every $v \in V$ can be written $v = u_1 + u_2$ with $u_1 \in U_1, u_2 \in U_2$ (this is just the definition).

Lemma 7.6.1. *If $V = U_1 \oplus U_2$ then every $v \in V$ can be written* uniquely *as $v = u_1 + u_2$ with $u_1 \in U_1, u_2 \in U_2$.*

Example 7.7. If we look at the previous example of a sum which is *not* direct,

$$\mathbb{R}^3 = \{(x, y, 0)\} + \{(0, y, z)\}$$

we see that we can write

$$(1, 1, 1) = (1, 1, 0) + (0, 0, 1) = (1, -1, 0) + (0, 2, 1).$$

Proof of Lemma. Suppose that

$$v = u_1 + u_2 = u_1' + u_2'.$$

Then

$$u_1 - u_1' = u_2' - u_2$$

and since the left hand side is in U_1 and the right hand side is in U_2 (and by definition of direct sum we have $U_1 \cap U_2 = \{0\}$) it follows

$$u_1 - u_1' = u_2' - u_2 = 0.$$

□

We now look at the analogue of set complements when it comes to vector spaces. Recall that for a set $A \subset X$ we defined

$$A^c = \{x \in X \colon x \notin A\}.$$

This is the unique set so that

$$A \cap A^c = \subset, \quad A \cup A^c = X.$$

An immediate fact is that the complement of a subspace is not a subspace since it does not contain 0. But we can still hope that for a subspace $U \subset V$ there exists a subspace $W \subset V$ so that

$$U + W = V \quad U \cap W = \{0\}.$$

The second condition implies that if W exists then $V = U \oplus W$. We will call such a W a **complement** of U. We will see that even though such a subspace exists, it is not unique.

Example 7.8. $U = \{(x, y) \colon x = y\} \subset \mathbb{R}^2$. This has at least two complements, the set $W = \{(x, y) \colon x = -y\}$ and $W' = \{(x, y) \colon x = 0\}$. In fact you can prove that for any $v \in \mathbb{R}^2 - U$ we have that

$$U \oplus \operatorname{span}(v) = V.$$

8.1. **Existence of complements for finite dimensional V.**

8.1.1. *The complement algorithm.* The idea is to try to build a complement inductively. We are given a vector space V over F, and a subspace $U \subsetneq V$. We want to get a subspace W such that $V = W \oplus U$. Since $V \neq U$ we can choose $w \in V \backslash U$ (where $\backslash$ indicates the set theoretic complement). Set

$$W = \operatorname{span}(w) = \{aw \colon a \in F\}.$$

Note that $W \cap U = \{0\}$, since if $aw \in U$ for some $a \neq 0$ then $a^{-1}(aw) = w \in U$. Now, if $V = U + W$, then $V = U \oplus W$ and we are done. If that is not the case then there is some $v \in V \backslash (U + W)$, and we can replace W with $W + \operatorname{span}(v)$ and repeat the previous step.

This algrithm has some potential issues. In fact, what one wants from an algorithm is that it will stop. This algorithm might *not* stop, and that will happen if V is "too large" (loosely defined – more on this in the future). The second problem is that even if the algorithm does stop, it is not clear that the stopping time is well-defined. This comes from the nonuniqueness issue, namely: we may have many choices for v at each step.

8.1.2. *Finite dimensionality.* In this section we will explain the notion of "too large" in more detail.

Definition 8.1. A **linear combination** of vectors $v_1, \ldots, v_k \in V$ is a vector of the form

$$v = a_1 v_1 + \cdots + a_k v_k, \qquad a_1, \ldots, a_k \in F.$$

Definition 8.2. For $v_1, \ldots, v_k \in V$, the **span** is defined as

$$\begin{aligned} \operatorname{span}(v_1, \ldots, v_k) &= \operatorname{span}(v_1) + \cdots + \operatorname{span}(v_k) \\ &= \{a_1 v_1 + \cdots + a_k v_k \colon a_i \in F\}. \end{aligned}$$

If $\mathrm{v}_1, \ldots, \mathrm{v}_\mathrm{k}$ we say that $v_1, \ldots, v_k$ span V.

Definition 8.3. A vector space is **finite dimensional** if

$$V = \operatorname{span}(v_1, \ldots, v_n), \qquad k \in \mathbb{N}.$$

We see that the subspace W constructed in the complement lagorith has the form $W = \operatorname{span}(v_1, \ldots, v_k)$. For example, if $U = \{0\}$ then the algorithm stops if and only if V is finite dimensional.

Example 8.4. For a field F, we have that

$$F^n = \{(x_1, \ldots, x_n) \colon x_i \in F\}$$

is finite dimensional. In fact, let us define

$$e_i = (0, \ldots, 0, \underbrace{1}_{i\text{th digit}}, 0, \ldots, 0).$$

Then

$$F^n = \operatorname{span}(e_1, \ldots, e_n).$$

Example 8.5. Consider

$$\mathrm{Poly}_d(F) = \left\{ a_d x^d + \cdots + a_0 \colon a_i \in F \right\}.$$

This is finite dimensional since

$$\mathrm{Poly}_d(F) = \mathrm{span}(x^d, \ldots, x, 1).$$

We can write

$$\mathrm{Poly}(F) = \bigcup_{d \geq 0} \mathrm{Poly}_d(F).$$

We claim that $\mathrm{Poly}(F)$ is not finite dimensional.

Proof. We will show that no finite collection spans this space. Let $p_1, \ldots, p_n \in \mathrm{Poly}(F)$. Let D be the largest degree among them. Then

$$p_1, \ldots, p_n \in \mathrm{Poly}_D(F)$$

and therefore

$$\mathrm{span}(p_1, \ldots, p_n) \subset \mathrm{Poly}_D(F) \neq \mathrm{Poly}(F).$$

□

Other non-example (examples of non finite dimensional) include F^∞ as well as $\mathbb{R}$ is not finite dimensional as a vector space over $\mathbb{Q}$.

8.1.3. *Linear independence.*

Definition 8.6. A collection of vectors $v_1, \ldots, v_k \in V$ is **linearly independent** if

$$0 = a_1 v_1 + \cdots + a_k v_k \iff a_1 = \cdots = a_k = 0.$$

Otherwise if there is nonzero $a_1, \ldots, a_k$ such that

$$0 = a_1 v_1 + \cdots + a_k v_k$$

we say that $v_1, \ldots, v_k$ are **linearly dependent**.

Assume now V is finite dimensional and suppose we run the complement algorithm with $U = \{0\}$. Does the algorithm always stop? If

$$V = W = \mathrm{span}(v_1, \ldots, v_k)$$
$$V = W' = \mathrm{span}(e_1, \ldots, e_n)$$

does this imply $k = n$?

Remark. At the kth step the algorithm produces

$$W = \mathrm{span}(v_1, \ldots, v_k)$$

where $v_1, \ldots, v_k$ are linearly independent. In fact, suppose they are linearly dependent, Then there exist $a_1, \ldots, a_k \in F$ such that

$$a_1 v_1 + \cdots + a_k v_k = 0.$$

Let j be the largest integer such that $a_j \neq 0$. Then

$$0 = a_1 v_1 + \cdot + a_j v_j, \qquad a_j \neq 0.$$

This means that we can write

$$v_j = (-a_j)^{-1}(a_1 v_1 + \cdots + a_{j-1} v_{j-1}).$$

This means that

$$v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$$

which contradicts the choice of v_j in the algorithm. Therefore $v_1, \ldots, v_k$ are linearly independent.

Theorem 8.7 (Linear independence theorem)**.** *Let $V = \operatorname{span}(v_1, \ldots, v_n)$ be a finite dimensional space. If $u_1, \ldots, u_k$ are linearly independent then $k \leq n$.*

Corollary 8.7.1. *For V finite dimensional, if the complement algorithm gives*

$$V = \operatorname{span}(v_1, \ldots, v_k) \text{ and } V = \operatorname{span}(u_1, \ldots, u_\ell)$$

then $k = \ell$.

Proof of corollary. We just use the theorem both ways to get

$$k \leq \ell, \quad \ell \leq k \Longrightarrow k = \ell.$$

□

Recall that

Definition 9.1. A collection of vectors $v_1, \dots, v_k \in V$ is **linearly independent** if

$$0 = a_1 v_1 + \cdots + a_k v_k \iff a_1 = \cdots = a_k = 0.$$

Otherwise if there is nonzero $a_1, \dots, a_k$ such that

$$0 = a_1 v_1 + \cdots + a_k v_k$$

we say that $v_1, \dots, v_k$ are **linearly dependent**.

Now we need to prove that

Theorem 9.2 (Linear independence theorem)**.** *Let $V = \operatorname{span}(v_1, \dots, v_n)$ be a finite dimensional space. If $u_1, \dots, u_k$ are linearly independent then $k \leq n$.*

Example 9.3. $\mathbb{R}^3 = \operatorname{span}((1,0,0),(0,1,0),(0,0,1))$ and therefore we cannot find 4 vectors in $\mathbb{R}^3$ that are linearly independent.

In order to prove the linear independence theorem we will prove the following lemma

Lemma 9.3.1. Linear dependence lemma. *Suppose $w_1, \dots, w_\ell$ are linearly dependent. Then there exist $a_1, \dots, a_\ell$ with $a_i \neq 0$ such that*

$$0 = a_1 w_1 + \cdots + a_\ell w_\ell.$$

Then

$$A := \operatorname{span}(w_1, \dots, w_\ell) = \operatorname{span}(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_\ell) =: B.$$

Proof. We will show $A \subset B$ and $B \subset A$. *Step 1.* $A \subset B$. This follows because every linear combination of $w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_\ell$ is a linear combination of $w_1, \dots, w_{i-1}, w_i, w_{i+1}, \dots, w_\ell$ where the coefficient of w_i is zero. *Step 2.* $B \subset A$. We first introduce some notation.

$$\sum_{1 \leq j \leq \ell} b_j w_j = b_1 w_1 + \cdots + b_\ell w_\ell$$

$$\sum_{\substack{1 \leq j \leq \ell \\ j \neq i}} b_j w_j = b_1 w_1 + \cdots + b_{i-1} w_{i-1} + b_{i+1} w_{i+1} + \cdots + b_\ell w_\ell.$$

To show $B \subset A$ we need to show that if

$$(\dagger) \qquad v = \sum_{1 \leq j \leq \ell} b_j w_j$$

then we can write

$$v = \sum_{\substack{1 \leq j \leq \ell \\ j \neq i}} b_j w_j.$$

By assumption,

$$-a_i w_i = \sum_{\substack{1 \leq j \leq \ell \\ j \neq i}} a_j w_j$$

and therefore since F is a field

$$w_i = \sum_{\substack{1 \leq j \leq \ell \\ j \neq i}} -\frac{a_j}{a_i} w_j.$$

We can therefore substitute this expression for w_i in equation (†) to get

$$\begin{aligned} v &= \sum_{\substack{1 \leq j \leq \ell \\ i \neq j}} b_j w_j + b_i w_i \\ &= \sum_{\substack{1 \leq j \leq \ell \\ i \neq j}} b_j w_j + b_i \left(\sum_{\substack{1 \leq j \leq \ell \\ i \neq j}} -\frac{a_j}{a_i} w_j \right) \\ &= \sum_{\substack{1 \leq j \leq \ell \\ i \neq j}} \left(b_j - \frac{a_j b_i}{a_i} \right) w_j \\ &\in \operatorname{span}(w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_\ell). \end{aligned}$$

□

The idea of the proof of the linear independence theorem is as follows: start with the lists

$$(u_1, \ldots, u_k) \qquad \text{and} \qquad (v_1, \ldots, v_,)$$

and merge these lists by replacing hte elements from the v list with elements from the u list preserving the fact that the v list spans. To show that $k \leq m$ we show that we run out of u's before we run out of v's.

Proof of linear independence theorem. Step 1. Since

$$V = \operatorname{span}(v_1, \ldots, v_m)$$

we can write

$$u_1 = a_1 v_1 + \cdots + a_m v_m$$

where $a_i \neq 0$ for some $1 \leq i \leq m$. Up to relabeling we can assume $a_1 \neq 0$. Then by the linear dependence lemma this implies that

$$\begin{aligned} \operatorname{span}(u_1, v_1, \ldots, v_m) &= \operatorname{span}(u_1, v_2, \ldots, v_m) \\ &= V. \end{aligned}$$

Step 2. Repeat the argument with the lists

$$(u_2, \ldots, u_k) \qquad \text{and} \qquad (u_1, v_2, \ldots, v_m).$$

We can always do so because if $(u_1, \ldots, u_k)$ is linearly independent, then so is $(u_2, \ldots, u_k)$. Therefore

$$u_2 = b_1 u_1 + b_2 v_2 + \cdots + b_m v_m$$

implies that $b_i \neq 0$ from some $\underline{i \geq 2}$. After this step we conclude analogously to step 1 that

$$\operatorname{span}(u_1, u_2, v_3, \ldots, v_m) = \operatorname{span}(u_1, v_2, v_3, \ldots, v_m).$$

Step 3. We continue until either

(i) the u-list becomes empty, in which case

$$V = \text{span}(u_1, \ldots, u_k, v_{k+1}, \ldots, v_m).$$

This means that the v-list contains m elements and contains the u-list as well, and therefore $k \leq m$;

(ii) v-list only contains elements of the u-list but not all of them. In this case

$$V = \text{span}(u_1, \ldots, u_\ell), \qquad \ell < k$$

and therefore

$$u_{\ell+1} \in \text{span}(u_1, \ldots, u_\ell)$$

which contradicts the hypothesis that the u-list is linearly independent.

Therefore only scenario (i) is possible, proving the theorem. □

9.1. **Bases.**

Definition 9.4. A collection of vectors $v_1, \ldots, v_m \in V$ is a **basis** if it is linearly independent and spans V.

Example 9.5. Consider the set of functions

$$V = \text{Fun}(S, F)$$

where $S = \{a, b\}$. Consider the vectors

$$f_1 : \begin{cases} a \mapsto 1 \\ b \mapsto 0 \end{cases}$$

$$f_2 : \begin{cases} a \mapsto 0 \\ b \mapsto 1 \end{cases}$$

$$f_3 : \begin{cases} a \mapsto 1 \\ b \mapsto 1 \end{cases} .$$

Which is a basis?

(i) f_1, f_2, f_3
(ii) f_1
(iii) f_1, f_2
(iv) f_1, f_3

Answer: (iii),(iv).

Remark. If $u_1, \ldots, u_k$ is a basis then any $v \in V$ can be written uniquely as a linear combination of these vectors, i.e.

$$v = a_1 u_1 + \cdots + a_k u_k$$

uniquely. The proof of this is an exercise. This means that

$$u_1, \ldots, u_k \text{ is a basis} \iff V = \text{span}(u_1) \oplus \cdots \oplus \text{span}(u_k).$$

The definition of a basis leads to three questions: do bases exist? How do we find a basis? Is the size of a basis well-defined?

Theorem 9.6. *Every finite dimensional vector space V has a basis. Furthermore any tow bases for V have the same finite cardinality.*

Definition 9.7. The **dimension** of a vector space is the cardinality of its basis.

The above definition makes sense because of the theorem.

10.1. **Bases.** Last time we left with three unanswered questions:
Do bases exist?
Is the size of a basis well-defined?
How do we find a basis?
We are going to address all of these questions at once.

Theorem 10.1. *Let V be a finite dimensional vector space over F. Then*
(a) V has a basis;
(b) any two bases have the same cardinality.

To prove these theorem we will make use of some results of the last few lectures: the linear independence theorem (theorem 8.7 and its corollary 8.7.1) and the linear dependence lemma (lemma 9.3.1).

Proof. We use the complement algorithm (section 8.1.1) applied to $U = \{0\}$. We keep going until

$$V = \text{span}(w_1, \ldots, w_n).$$

for some $w_1, \ldots, w_n \in V$. We see that at each step the vectors $w_1, \ldots, w_k$ are linearly independent because otherwise we would have

$$w_i \in \text{span}(w_1, \ldots, w_{i-1})$$

which is false by construction. Since V is finite dimensional, we can write

$$V = \text{span}(v_1, \ldots, v_m)$$

and by the linear dependence theorem we have

$$V = \text{span}(w_1, \ldots, w_n).$$

for $n \leq m$ and so $w_1, \ldots, w_n$. □

Addendum. (finding bases under constraints).
(a) *Extending to a basis.* If $u_1, \ldots, u_k \in V$ are linearly independent, then there exists a basis for V that contains $u_1, \ldots, u_k$.
(b) *Restricting to a basis,* If $v_1, \ldots, v_m \in V$ span V then some subset of $\{v_1, \ldots, v_m\}$ is a basis for V.

Proof. (a) Apply complement algorithm to $U = \text{span}(u_1, \ldots, u_k)$ to get

$$W = \text{span}(w_1, \ldots, w_\ell)$$

where $w_1, \ldots, w_\ell$ are linearly independent and $V = U \oplus W$. We claim that $u_1, \ldots, u_k, w_1, \ldots, w_\ell$ are a basis for V. They span because

$$\begin{aligned} V &= U + W \\ &= \text{span}(u_1, \ldots, u_k) + \text{span}(w_1, \ldots, w_\ell). \end{aligned}$$

They are linearly independent because

$$0 = \underbrace{a_1 u_1 + \cdots + a_k u_k}_{:= u \in U} + \underbrace{b_1 w_1 + \cdots + b_\ell w_\ell}_{:= w \in W}$$

implies $w = -u$ and since subspaces are closed under scalar multiplication we have that $u \in U \cap W$ and therefore $u = 0$ by definition of direct sum. Then $a_1 = \cdots = a_k = 0$ by linear independence of the basis of U and therefore $b_1 = \cdots = b_\ell = 0$ by linear independence of the basis of W.

(b) Inductively we can consider $v_1, \ldots, v_j$ starting with $j = 2$. If

$$v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$$

we discard v_j, if not we keep it and continue to v_{j+1}. At the end we will be left with a list

$$\{v_{i_1}, \ldots, v_{i_n}\}, \qquad \{i_1, \ldots, i_n\} \subset \{1, \ldots, m\}$$

which still spans V. We claim that $\{v_{i_1}, \ldots, v_{i_n}\}$ is also linearly independent. In fact, if we were not we would have that

$$v_{i_j} \in \operatorname{span}(v_{i_1}, \ldots, v_{i_{j-1}})$$

which is false by construction.

□

10.2. **Counting bases of $V = (\mathbb{Z}/p\mathbb{Z})^n$.** . Let p be a prime number. We have seen that

$$\mathbb{Z}/p\mathbb{Z} = \{0, 1, \ldots, p-1\}$$

is a field. How many bases does $(\mathbb{Z}/p\mathbb{Z})^n$ have? For example, let $p = 2, n = 3$. To warm up, we count the subspace supspaces of $(\mathbb{Z}/n\mathbb{Z})^2$. We claim that $p+3$ subspaces. As an exercise you should prove that

(i) a subpace $U \subset V$ has $\dim U \leq \dim V$;
(ii) if $\dim U = \dim V$ then $U = V$.

Then any subspace has

$$\dim U = \begin{cases} 0 & \Rightarrow U = \{0\}; \\ 1 & \Rightarrow U = \operatorname{span}(u),\ u \neq 0 \\ 2 & \Rightarrow U = V. \end{cases}$$

We see that $(\mathbb{Z}/n\mathbb{Z})^2$ has $p^2 - 1$ vectors, but two vectors might be in the same subspace, i.e. $u' = cu$ for some $c \in \mathbb{Z}/p\mathbb{Z}$. Since there are $p-1$ nonzero scalars in $\mathbb{Z}/p\mathbb{Z}$ there are

$$\frac{p^2 - 1}{p - 1} = p + 1$$

one dimensional subspaces of $(\mathbb{Z}/p\mathbb{Z})^2$. To go back to $(\mathbb{Z}/p\mathbb{Z})^n$ we see that there are $p^n - 1$ nonzero vectors to choose v_1 from. For v_2 there are

$$p^n - 1 - (p-1) = p^n - n$$

choices and in general for v_j there are

$$p^n - p^{j-1}$$

choices and therefore there are

$$(p^n - p^{n-1}) \cdots (p^n - 1)$$

ordered bases.

11.1. **Linear Maps. Motivating Example.** Let V be finite dimensional with basis $v_1, \ldots, v_n$. This means that any $v \in V$ can be written uniquely as

$$v = \sum a_i v_i.$$

We define a map of sets

$$\begin{aligned} T : F^n &\to V \\ (a_1, \ldots, a_n) &\mapsto a_1 v_1 + \cdots + a_n v_n. \end{aligned}$$

T is a bijection of sets. It is surjective since $v_1, \ldots, v_n$ span V and injective because $v_1, \ldots, v_n$ are linearly independent. Moreover, T is better than just a bijection of *sets*. In fact, T transforms the vector space operations on F^n to those of V. Let

$$\begin{aligned} u &= (a_1, \ldots, a_n) \in F \\ u' &= (a_1', \ldots, a_n') \in F. \end{aligned}$$

Then

$$\begin{aligned} T(u + u') &= T(a_1 + a_1', \ldots, a_n + a_n') \\ &= \sum (a_i + a_i') v_i \\ &= \sum a_i v_i + \sum a_i' v_i \\ &= T(u) + T(u') \end{aligned}$$

and for all $c \in F$

$$\begin{aligned} T(cu) &= T(ca_1, \ldots, ca_n) \\ &= \sum c a_i v_i \\ &= c \sum a_i v_i \\ &= cT(u). \end{aligned}$$

Definition 11.1. Let V, W be vector spaces over F. We say that $T : V \to W$ is a **linear map** (or **linear transformation**) if for each $v_1, v_2 \in V, c_1, c_2 \in F$ we have

$$T(c_1 v_1 + c_2 v_2) = c_1 T(v_1) + c_2 T(v_2).$$

If $T : V \to W$ is linear and a bijection, T is called a **linear isomorphism**.

Example 11.2. There is a linear isomorphism

$$\begin{aligned} F^{m+1} &\to \mathrm{Poly}_m(F) \\ (a_0, \ldots, a_m) &\mapsto a_m x^m + \cdots + a_1 x + a_0. \end{aligned}$$

Example 11.3. Let $V = \mathrm{Fun}(S, F)$ with $S = \{1, \dots, m\}$. There is a linear isomorphism

$$\begin{aligned} F^m \to & V \\ (a_1, \dots, a_m) \mapsto & \sum a_i f_i \\ f_i(x) = & \begin{cases} 1 & x = i \\ 0 & \text{else} \end{cases}. \end{aligned}$$

Remark. Not every linear map is an isomorphism.

Example 11.4. For example we have the zero map,

$$\begin{aligned} Z : V \to & W \\ v \mapsto & 0_W. \end{aligned}$$

This map is surjective if and only if $\dim W = 0$ ahd it is injective if and only if $\dim V = 0$.

Example 11.5. Evaluating polynomials. Let $V = \mathrm{Poly}(\mathbb{C})$, namely polynomials with coefficients in $\mathbb{C}$ of arbitrary degree. Define

$$\begin{aligned} E : V \to & \mathbb{C} \\ p \mapsto & p(0), \end{aligned}$$

that is to say, if

$$p = a_d x^d + \cdots a_1 x_1 + a_0$$

then

$$\begin{aligned} E(p) = & p(0) \\ = & a_d (0)^d + \cdots a_1 \cdot 0 + a_0 \\ = & a_0. \end{aligned}$$

E is linear because if a_0 is the constant term of $p \in V$ and b_0 is the constant term of $q \in V$ then the constant coefficient of $cp + dq$ is $ca_0 + db_0$ and therefore

$$E(cp + dq) = cE(p) + dE(q).$$

Note that the map is surjective (because $E(a_0) = a_0$ for a_0 a degree 0 polynomial) but it is not injective. Mor generally for $\lambda \in \mathbb{C}$ we can define

$$\begin{aligned} E_\lambda : V \to & \mathbb{C} \\ p \mapsto & p(\lambda) \end{aligned}$$

and this is also going to be a linear map.

11.2. **Subspaces associated to linear maps.** We are going to see some subspaces that measure the failure of a linear map to be an isomorphism.

Definition 11.6. For a linear map $T : V \to W$ we define

- the **kernel** of T (also called the **nullspace**)

$$\ker T = \{v \in V : Tv = 0\}$$

- the **image** of T (also called the **range**)

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

These subsets are in fact subspaces!

Proposition 11.1. $\ker T$ *is a subspace.*

Proof. For $v_1, v_2 \in \ker T$ and $c_1, c_2 \in F$ we have

$$\begin{aligned} T(c_1v_1 + \cdots c_2v_2) =& c_1Tv_1 + c_2Tv_2 \\ =& c_1 \cdot 0 + c_2 \cdot 0 \\ =& 0 \end{aligned}$$

and therefore $c_1v_1 + c_2v_2 \in \ker T$. □

Proposition 11.2. $\operatorname{im} T$ *is a subspace.*

Proof. Let $w_1, w_2 \in \operatorname{im} T$ and $c_1, c_2 \in F$. □

This means that there are some $v_1, v_2 \in V$ such that $Tv_1 = w_1, Tv_2 = w_2$. By linearity,

$$\begin{aligned} c_1w_1 + c_2w_2 =& c_1T(v_1) + c_2T(v_2) \\ &+T(c_1v_1 + c_2v_2) \end{aligned}$$

and therefore $c_1w_1 + c_2w_2 \in \operatorname{im} T$.

Remark. For any linear map $T : V \to W$ we have $T(0) = 0$. If fact $T(0) = T(0 + 0) = T(0) + T(0)$ and so $T(0) = 0$.

Example 11.7. In the case of example 11.5 we see that $\operatorname{im} E = \mathbb{C}$ and $\ker E$ is equal to the set of polynomial with zero constant term.

Lemma 11.7.1. *Let $T : V \to W$ be linear. Then T is injective if and only if* $\ker T = \{0\}$.

Proof. $\Rightarrow$: Let $v \in \ker T$. We know

$$T(0) = 0 = T(v)$$

and since T is injective it follows that $v = 0$. $\Leftarrow$: Suppose $v_1, v_2 \in V$ and $T(v_1) = T(v_2)$. This means that

$$\begin{aligned} T(v_1 - v_2) =& T(v_1) - T(v_2) \\ =& 0. \end{aligned}$$

Then $v_1 - v_2 \in \ker T$ and thus $v_1 - v_2 = 0$. □

11.3. **Matrices.**

Example 11.8. Let $V = F^2$ and $W = F^3$. Choose bases v_1, v_2 for V and w_1, w_2, w_3 for W. **Obserbation.** Any linear map $T : V \to W$ is determined by its values $T(v_1)$ and $T(v_2)$. In fact, since for all $v \in V$ we have $v = a_1v_1 + a_2v_2$ uniquely, we have that

$$\begin{aligned} T(v) =& T(a_1v_1 + a_2v_2) \\ =& a_1T(v_1) + a_2T(v_2). \end{aligned}$$

We can write

$$\begin{aligned} T(v_1) =& a_{11}w_1 + a_{21}w_2 + a_{31}w_3 \\ T(v_2) =& a_{12}w_1 + a_{22}w_2 + a_{32}w_3 \end{aligned}$$

and thus any linear map $T : F^2 \to F^3$ is determined by 6 scalars.

Example 11.9. What are the 6 numbers that determine the map

$$\begin{aligned} T : \mathbb{R}^2 \to& \mathbb{R}^3 \\ (x, y) \mapsto& (2x + y, x, y - x)? \end{aligned}$$

First of all it is important to not that it does not make any sense to talk about "the" 6 numbers, since our answer depends on the choice of a basis, which is in no way unique. For example, let $v_1 = (1, 0)$ and $v_2 = (0, 1)$, and $w_1 = (1, 0, 0), w_2 = (0, 1, 0), w_3 = (0, 0, 1)$. Then we see that

$$\begin{aligned} T(v_1) =& (2, 1, -1) = 2w_1 + w_2 - w_3 \\ T(v_2) =& (1, 0, 1) = w_1 + w_3. \end{aligned}$$

12.1. **Linear Maps – Continued.** Recall that a map $T : V \to W$ between vector spaces over a field F is linear if it preserves addition, i.e.

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

and it preserves scalar multiplication,i.e.

$$T(cv) = cT(v).$$

Example 12.1. When is $f : \mathbb{R} + \mathbb{R}^2$ linear? Say that

$$f(x) = (g(x), h(x)),$$

with $g, h : \mathbb{R} \to \mathbb{R}$. If we want f to be linear we need

$$\begin{aligned}(g(ax + by), h(ax + by)) &= f(ax + by)\\ &= af(x) + bf(y)\\ &= a(g(x), h(x)) + b(g(y), h(y))\\ &= (ag(x) + bg(y), ah(x) + bh(y)).\end{aligned}$$

Therefore we see that f is linear if and only if g and h are. In homework 5 you will prove that $k : \mathbb{R} \to \mathbb{R}$ if and only if $k(x) = \lambda x$ for some $\lambda \in \mathbb{R}$. Therefore in this case

$$f(x) = (\lambda x, \eta x)$$

for some $\lambda, \eta \in \mathbb{R}$. Therefore we see that

$$x \mapsto (x, 2x)$$

is linear, whereas

$$x \mapsto (x, x^2)$$

is not, since for example $(1 + 1)^2 \neq 1^2 + 1^2$.

We can also use an alternate approach for this example. Recall from last time that there are two subspaces associated to $f : \mathbb{R} \to \mathbb{R}^2$, namely the kernel $\ker(f)$ and the image $\mathrm{im}(f)$. Since the subspaces of $\mathbb{R}$ are $\{0\}$ and $\mathbb{R}$. The possibilities for $\ker(f)$ are therefore $\{0\}$ and $\mathbb{R}$, because the kernel is a subspace. If $\ker(f) = \mathbb{R}$ we get that f is the zero map, i.e. $f(x) = 0$ for all x. In the case $\ker(f) = \{0\}$ we get from a lemma we proved last time that f is injective. Therefore we now need to look at its image. Again $\mathrm{im}(f)$ is a subspace, and the subspaces of $\mathbb{R}^2$ are $\{0\}, \mathrm{span}(w)$ for some nonzero $w \in \mathbb{R}^2$ and $\mathbb{R}^2$ itself. We already saw that $\mathrm{im}(f) = 0$ when $f = 0$. Say $w = (a, b)$. Then $\mathrm{im}(f) = (ax, bx)$ when

$$\begin{aligned}f(x) &= (ax, bx)\\ &= x \cdot (a, b).\end{aligned}$$

Can $\mathbb{R}^2 = \mathrm{im}(f)$ for some $f : \mathbb{R} \to \mathbb{R}^2$? We have already shown in the first derivation that

$$\begin{aligned}f(x) &= (x\lambda, x\eta)\\ &= x(\lambda, \eta)\end{aligned}$$

and therefore $\mathrm{im}(f) \neq \mathbb{R}^2$ for any linear f.

Example 12.2. Is

$$\begin{aligned} f : F &\to F^2 \\ x &\mapsto (x, x^2) \end{aligned}$$

linear when $F = \mathbb{Z}/2\mathbb{Z}$? The answer is yes, because for all $x \in F$ we have $x = x^2$ (we can check this directly: $0^2 = 0$ and $1^2 = 1$).

Example 12.3. Consider

$$\begin{aligned} D : \mathrm{Poly}(F) &\to \mathrm{Poly}(F) \\ x^n &\mapsto n x^{n-1}. \end{aligned}$$

This is a linear map. What is $\ker(D)$

- when $F = \mathbb{C}$?
- when $F = \mathbb{Z}/2\mathbb{Z}$?

In the first case we see that

$$p := a_m x^m + \cdots + a_1 x + a_0 \in \ker(D)$$

if

$$m a_m x^{m-1} + \cdots + a_1 = 0,$$

so that we see that we must have

$$a_m = \cdots = a_1 = 0.$$

Therefore

$$\ker(D) = \{ \text{ constant polynomials } p = a_0 \}.$$

In the case $\mathbb{Z}/2\mathbb{Z}$ we see that in the example

$$p = a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$$

we get

$$\begin{aligned} D(p) &= 4a_4 x^3 + 3a_3 x^2 + 2a_2 x + a_1 \\ &= 3a_3 x^2 + a_1 \end{aligned}$$

because $2n = 0$ in $\mathbb{Z}/2\mathbb{Z}$. Therefore

$$\ker(D) = \mathrm{span}(1, x^2, x^4, \dots).$$

Remark. An element $p \in \mathrm{Poly}(F)$ is different from the function it defines. For example, $x^2 + x \in \mathrm{Poly}(\mathbb{Z}/2\mathbb{Z})$ is nonzero, although the function

$$\begin{aligned} f : \mathbb{Z}/2\mathbb{Z} &\to \mathbb{Z}/2\mathbb{Z} \\ x &\mapsto x^2 + x \end{aligned}$$

is the zero function.

12.2. **Rank-Nullity.** We start with a question: is there a surjective linear map $F^{10} \to F^{11}$? If $F = \mathbb{Z}/2\mathbb{Z}$ we can see that there is none, since F^{10} has 2^{10} elements and F^{11} has 2^{11} elements. However, for infinite fields this changes, since for example there is a surjection $f : \mathbb{R}^{10} \to \mathbb{R}^{11}$, in the same way there is one from $[0, 1]$ to $[0, 1]^2$. However, we are requiring our map to be not only a surjection, but a surjective linear map. This theorem will prove useful.

Theorem 12.4 (Rank-Nullity)**.** *Let V be finite dimensional, and $T : V \to W$ linear. Then* $\operatorname{im}(T) \subset W$ *is finite dimensional and*

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

Corollary 12.4.1. *There is no surjective linear map $F^{10} \to F^{11}$, since*

$$\begin{aligned} T \text{ is surjective } &\iff \operatorname{im}(T) = F^{11} \\ &\iff \dim \operatorname{im}(T) = 11 \end{aligned}$$

but by the rank-nullity theorem

$$\dim \operatorname{im}(T) \leq \dim F^{10} = 10.$$

Proof. Since $\ker(T) \subset V$ is finite dimensional we choose a basis $u_1, \ldots, u_k$ for $\ker(T)$. Then we extend it to a basis $u_1, \ldots, u_k, v_1, \ldots, v_\ell$.

Claim: $T(v_1), \ldots, T(v_\ell)$ is a basis for $\operatorname{im}(T)$.

–to be continued □

13. 10-2

13.1. **Rank-nullity.** Last time we stated the rank-nullity theorem, namely

Theorem 13.1 (Rank-Nullity)**.** *Let V be finite dimensional, and $T : V \to W$ linear. Then* $\operatorname{im}(T) \subset W$ *is finite dimensional and*

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

Example 13.2. Let us consider a linear map

$$\begin{aligned} S : \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto y - x. \end{aligned}$$

Then

$$\ker S = \{(x, y) \mid x = y\}.$$

We can choose a complement W such that $\mathbb{R}^2 = W \oplus \ker(S)$, for example

$$W = \{(x, y) \mid x = 0\}.$$

We see that

$$\begin{aligned} S\big|_W : W &\to \mathbb{R} \\ (0, y) &\mapsto y \end{aligned}$$

is a linear isomorphism. Therefore

$$\dim W = \dim \operatorname{im} S$$

and

$$\begin{aligned} \dim V &= \dim \ker S + \dim W \\ &= \dim \ker S + \dim \operatorname{im} S. \end{aligned}$$

Proof of the theorem. Choose a complement U of $\ker T$ such that $V = \ker T \oplus U$, and choose bases

$$\overbrace{\underbrace{u_1, \dots, u_k}_{\text{basis for } \ker T}, \underbrace{w_1, \dots, w_\ell}_{\text{basis for } U}}^{\text{basis for } V}.$$

Claim. $T(w_1), \dots, T(w_\ell)$ *form a basis for* $\operatorname{im} T$. Given this we can just write

$$\begin{aligned} \dim V &= k + \ell \\ &= \dim \ker T + \dim U \\ &= \dim \ker T + \dim \operatorname{im} T. \end{aligned}$$

Proof of the claim. We first prove that $T(w_1), \dots, T(w_\ell)$ span $\operatorname{im} T$. For $w \in \operatorname{im} T$, we can write $w = T(v)$ for some $v \in V$. Since

$$v = \sum b_i u_i + \sum c_j w_j$$

we have that

$$\begin{aligned} w =& T(v) \\ =& T\left(\sum b_i u_i + \sum c_j w_j\right) \\ =& \sum b_i \underbrace{T(u_i)}_{=0} + \sum c_j T(w_j) \\ =& \sum c_j T(w_j) \end{aligned}$$

and so $T(w_1), \ldots, T(w_\ell)$ span $\operatorname{im} T$. We now need to show that they are linearly independent, namely we need to show that if

$$0 = \sum a_i T(w_i)$$

then $a_1 = \cdots = a_\ell = 0$. Bty linearity,

$$\begin{aligned} 0 =& \sum a_i T(w_i) \\ =& T\left(\sum a_i w_i\right) \end{aligned}$$

and therefore

$$\sum a_i w_i \in \ker T.$$

However, we also know that

$$\sum a_i w_i \in U$$

because $w_1, \ldots, w_\ell$ form a basis for U. Since $U \cap \ker T = 0$ we have

$$\sum a_i w_i = 0$$

and since $w_1, \ldots, w_\ell$ is linearly independent it follows that $a_1 = \cdots = a_\ell = 0$. □

As we stated at the beginning of the claim, this proves the theorem. □

Corollary 13.2.1. *There is no injection $f : V \to W$ where $\dim W < \dim V$ and no surjection the other way (e.g. no injection $F^{12} \to F^7$ and no surjection $F^7 \to F^{11}$).*

13.2. **Matrices and linear maps.** Let V, W be finite dimensional and let $T : V \to W$ be a linear map. Choose bases

$$v_1, \ldots, v_n \in V \qquad w_1, \ldots, w_m \in W.$$

For $j = 1, \ldots, n$ we can write

$$\begin{aligned} T(v_j) =& a_{1j} w_1 + \cdots + a_{mj} w_m \\ =& \sum_{i=1}^{m} a_{ij} w_i \end{aligned}$$

with $a_{ij} \in F$. This allows us to create a **matrix** with the jth column containing the coefficients of $T(v_j)$, as in the following:

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

Example 13.3. Let

$$\begin{aligned} T : \mathbb{R}^2 &\to \mathbb{R}^3 \\ (x, y) &\mapsto (2x + y, x, y - x) \\ (1, 0) &\mapsto (2, 1, -1) \\ (0, 1) &\mapsto (1, 0, 1) \end{aligned}$$

so that the matrix of T with the usual bases for $\mathbb{R}^n$ is equal to

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

However, if we use the basis $(1, 0), (1, 1)$ for $\mathbb{R}^2$ we get that since

$$(1, 1) \mapsto (3, 1, 0)$$

our matrix becomes

$$\text{matrix of } T = \begin{bmatrix} 2 & 3 \\ 1 & 1 \\ -1 & 0 \end{bmatrix}.$$

Remark. The matrix for T depends on the choice of basis.

Thus we saw that matrices are not a canonical way of representing a linear map (they do depend on the choice of a basis); however, they are a useful computational tool.

Example 13.4. We can compute $T(v)$ using matrix multiplication. For $v = \sum x_j v_j$ we have

$$\begin{aligned} T(v) &= T\left(\sum x_j v_j\right) \\ &= \sum x_j T(v_j) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} w_i \\ &= \sum_j \sum_i x_j a_{ij} w_i \\ &= \sum_j \left(\sum_i x_j a_{ij}\right) w_i. \end{aligned}$$

We can write this using matrices in the following way:

$$(\star) \qquad \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1} + \cdots + a_{mn} \end{bmatrix}$$
$$= T(v) \in W.$$

Here we write elements of V, W as column vectors with n and m entries respectively, and use a product defined entry-wise (the dot product).

Remark. Equation $(\star)$ reduces the problem of computing $\ker T$ to solving a system of equations. In fact, if we want $T(v) = 0$ then this is equivalent to solving the equations

$$a_{11}x_1 + \cdots + a_{1n}x_n = 0$$
$$\vdots$$
$$a_{m1} + \cdots + a_{mn} = 0$$

Example 13.5. Does the system of equation

$$(\star\star) \qquad \begin{cases} 3x + 4y - z = 0 \\ -2x + y + 2z = 0 \end{cases}$$

have a solution? First of all we see that $(0, 0, 0)$ solves it. But does it have a nontrivial (e.g. nonzero) solution? What if $F = \mathbb{Z}/5\mathbb{Z}$? The answer is to consider the linear map

$$T : F^3 \to F^2$$
$$(x, y, z) \mapsto (3x + 4y - z, -2x + y + 2z).$$

Then (x, y, z) is a solution to $(\star\star)$ if and only if $(x, y, z) \in \ker T$. By rank-nullity we have that

$$\dim F^3 = \dim \ker T + \dim \operatorname{im} T$$

and therefore in this case

$$\dim \ker T = 3 - \dim \operatorname{im} T$$
$$\geq 1$$

and so $(\star\star)$ has nontrivial solutions.

14.1. **Linear maps and bases.**

Lemma 14.0.1. *let V, W be vector spaces, with V finite dimensional and $v_1, \dots, v_n$ a basis for V. For* any *$z_1, \dots, z_n \in W$ there exists a unique linear map $T : V \to W$ such that $T(v_i) = w_i$.*

Remark. This lemma is useful to define a linear map by its values on a basis, e.g. on homework 4, problems $2, 7, 8, 10$, as well as problem 3a in homework 5.

Proof. For $v \in V$ we can write uniquely

$$v = a_1 v_1 + \cdots + a_n v_n$$

and define

$$T(v) = \sum a_j z_j.$$

We claim that T is linear. Let

$$v = \sum a_i v_i \qquad v' = \sum a_i' v_i.$$

Then

$$\begin{aligned} T(cv + c'v') =& T\left(c \sum a_i v_i + c' \sum a_i' v_i\right) \\ =& T\left(\sum (ca_i + c'a_i') v_i\right) \\ =& \sum c(a + a') z_i \\ =& c \sum a_i v_i + c' \sum a_i v_i \\ =& cT(v) + c'T(v'). \end{aligned}$$

We now prove it is unique. Suppose $S : V \to W$ is also linear and $S(v_i) = z_i$. We want to show that $S = T$. Let's fix

$$v = a_1 v_1 + \cdots + a_n v_n \in V.$$

Since S is linear,

$$\begin{aligned} S(v) =& S\left(\sum a_i v_i\right) = \qquad\qquad\qquad\qquad a_i \sum S(v_i) \\ =& a_i \sum z_i \\ =& a_i T(v_i) \\ =& T(v). \end{aligned}$$

□

14.2. **Operations on linear maps.** Let

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

There is an addition

$$+ : L(V,W) \times L(V,W) \to L(V,W)$$

given by

$$(S+T)(v) = S(v) + T(v).$$

We actually need to check that $T+S$ thus defined is linear (for all we know it is just a map from V to W). By definition

$$\begin{aligned}(S+T)(av+a'v') =& S(av+a'v') + T(av+a'v') \\ =& aT(v) + a'T(v') + aS(v) + a'T(v') \\ =& aT(v) + aS(v) + a'T(v') + a'S(v') \\ =& a(T+S)(v) + a'(T+S)(v).\end{aligned}$$

$L(V,W)$ also has scalar multiplication

$$\cdot F \times L(V,W) \to L(V,W)$$

which you can checkin a similar way that is also linear. These operations make $L(V,W)$ into a vector space.

Remark. The additive identity in $L(V,W)$ is

$$\begin{aligned}Z : T \to& V \\ v \mapsto& 0.\end{aligned}$$

We can also look at another, closely related vector space,

$$M_{m\times n}(F) = \{m \times n \text{ matrices } A = (a_{ij}) \text{ with } a_{ij} \in F\}.$$

with operations

$$\begin{aligned}(A+A')_{ij} =& a_{ij} + a'_{ij} \\ (cA)_{ij} =& ca_{ij}.\end{aligned}$$

Why are these closely related spaces? We can choose bases $v_1, \ldots, v_n \in V$ and $w_1, \ldots, w_m \in W$ and define

$$\begin{aligned}\phi : L(V,W) \to& M_{m\times n}(F) \\ T \mapsto& A = (a_{ij})\end{aligned}$$

where the a_{ij} are defined by

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Claim. ϕ is a bijection, i.e. ϕ is a linear isomorphism. This is important because it tells us that we can go back and forth between linear maps and matrices depending on what is easier.

Proof of linearity of ϕ. If $A = (a_{ij})$ is the matrix of T and $A' = (a'_{ij})$ is the matrix of T' we want to show that $A + A'$ is the matrix of $T + T'$. By linearity

$$\begin{aligned}(T + T')(v_j) =& T(v_j) + T'(v_j)\\ =& \sum a_{ij}w_i + \sum a'_{ij}w_i\\ =& \sum (a_{ij} + a'_{ij})w_i\end{aligned}$$

and the coefficients $(a_{ij} + a'_{ij})$ are the cefficients of $A + A'$. □

Exercise. How many elements does $M_{2\times 3}(F)$ have for $F = \mathbb{Z}/2\mathbb{Z}$? And what is the dimension of $M_{2\times 3}(F)$?

14.3. **Isomorphisms and invertibility. Recall.** Let $f : X \to Y$ be a map of sets. We say f is invetible if there exists $g : Y \to X$ such that $g \circ f = id_X$ and $f \circ g = id_Y$. These equalities (co)imply injectivity and surjectivity respectively.

Definition 14.1. A linear map $T : v \to W$ is **invertible (as a linear map)** if there exists $S : W \to V$ such that S is linear and $ST = id_V, TS = id_W$.

When is $T : V \to W$ invertible (as a linear map)? In some homework problem you learned that if $T = W = F$ then every linear $T : V \to W$ is given by $T(v) = \lambda v$ for some $\lambda \in F$. As long as $\lambda \neq 0$ we have that T is invertible and its inverse is $S : W \to V$ with $S(w) = \lambda^{-1}w$. What about $\lambda \neq 0$? In this case T is not surjective, so T is not injective. In conclusion, in this restricted case

$$T \text{ is invertible } \iff \lambda \neq 0 \iff T \text{ is a bijection .}$$

Proposition 14.1. *$T : V \to W$ is invertible if and only if T is a linear isomorphism.*

Proof. For the "only if" direction, if T is invertible then by definition there exists S linear such that $ST = id_V$ and $TS = id_W$. On the homework you showed that this implies that T is injective and surjective respectively, thus T is a linear isomorphism. For the "if" direction, suppose T is a linear isomorphism. Then T is a bijection and thus it has a set theoretic inverse $S : W \to V$ such that $ST = id_V, TS = id_W$. We need to show that S is linear. Take $v, v' \in V$ such that $T(v) = w, T(v') = w'$. Therefore

$$\begin{aligned}S(aw + a'w') =& av + av'\\ =& aS(v) + a'S(v').\end{aligned}$$

□

Example 14.2. An application of this proposition is the following. Let us return to

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

In order to show that it is a linear isomorphism it suffices to define an inverse. Define

$$\begin{aligned}\psi : M_{m\times n}(F) \to& L(V, W)\\ A = (a_{ij}) \mapsto& \psi(A)\end{aligned}$$

so that

$$\psi(A)(v_j) = a_{ij}w_i.$$

Then $\psi(A)$ is linear by Lemma 14.0.1. Need to check that $\psi\phi = id_{L(V,W)}$ and that $\phi\psi = id_{M_{m\times n}(F)}$.

15.1. **Summary of linear maps so far.**

- Associated to a linear map $T : V \to W$ we get two subspaces, $\ker T \subset V$ and $\operatorname{Im} T \subset W$.
- The rank-nullity theorem states that if V if finite dimensional,
$$\dim V = \dim \ker T + \dim \Im T.$$
We say T is a linear isomorphism if
  - T is injective ($\ker T = 0$);
  - T is surjective ($\Im T = W$).
- $$L(V, W) = \{T : V \to W : T \text{ is linear } \}$$
is a vector space over F.
- $T \in L(V, W)$ is determined by its values on a basis.
- A choice of bases for V, W gives a linear isomorphism
$$L(V, W) \to M_{m\times n}(F)$$
where $m = \dim W$ and $n = \dim V$.

15.2. **Linear operators.** A linear map $T : V \to V$ is called a **linear operator**. We write $L(V)$ instead of $L(V, V)$ and $M_n(F)$ instead of $M_{n\times n}$.
Observation. Given $S, T \in L(V)$ then the composite $S \circ T : V \in V$ is also linear and $S \circ T \in L(V)$ as well. This means that $L(V)$ is not only a vector space but it also has a "product"

$$\begin{aligned} \circ : L(V) \times L(V) &\to L(V) \\ (S, T) &\mapsto S \circ T. \end{aligned}$$

Moreover, if $T \in L(V)$ is invertible then T^{-1} is also linear (as we saw last time), i.e. $T^{-1} \in V$.
Questions.

(1) If we now matrices for S and T, what is the matrix for $S \circ T$?
(2) If T is invertible, what is the matrix for T^{-1}?

15.3. **Matrix multiplication.** The answer to the first question is **matrix multiplilcation**. Let

$$\begin{aligned} A &= (a_{ij}) \in M_{m\times n}(F) \\ B &= (b_{ij}) \in M_{\ell\times m}(F). \end{aligned}$$

The **product** $BA \in M_{\ell\times n}$ is the matrix whose (i, j) entry is the dot product of the ith row of B with the jth column of A, namely

$$\begin{aligned} (BA)_{ij} &= b_{i1}a_{1j} + \cdot + b_{im}a_{mj} \\ &= \sum_{k=1}^{m} b_{ik}a_{kj}. \end{aligned}$$

Example 15.1.

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$
$$= \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} & b_{11}a_{13} + b_{12}a_{23} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} & b_{21}a_{13} + b_{22}a_{23} \end{pmatrix}$$

Remark. In order to multiply B and A we need that the number of columns of B be equal to the number of rows of A. If this does not happen the above definition does not make sense.

Properties of matrix mulitplication.

(i) Matrix multiplication is non-commutative. When $m = \ell = n$ we often get that $AB \neq BA$.

Example 15.2. The matrices

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

do not commute. In fact on the basis $(1,0),(0,1)$ we have that

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

and therefore

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

$$BA \begin{pmatrix} 1 \\ 0 \end{pmatrix} = B \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

(ii) There are "zero divisors", i.e. $AB = 0$ does not imply that eithe A or B is equal to 0. For example,

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

Remark. $M_2(\mathbb{R})$ has addition and multiplication but is *not* a field (since multiplication has to be commutative).

Proposition 15.1. *Matrix multiplication corresponds to composition of linear maps, i.e. if A is the matrix for $T : U \to V$ and B is the matrix for $S : V \to W$ then AB is the matrix for $ST : U \to W$.*

Proof. Assume $U = V = W$. Fix a basis $v_1, \ldots, v_n$ for V and let A, B, C be the matrices associated with T, S, ST. On the one hand,

$$ST(v_j) = \sum_{i=1}^{n} c_{ij} v_i.$$

On the other hand,

$$T(v_j) = \sum a_{ij} v_i \qquad S(v_k) = \qquad \sum b_{ik} v_i$$

and so

$$\begin{aligned} S(T(v_j)) =& S\left(\sum a_{kj} v_k\right) \\ =& \sum a_{kj} S(v_k) \\ =& \sum_{k=1}^{n} a_{kj} \sum_{i=1}^{n} b_{ij} v_i \\ =& \sum_{i=1}^{n} \left(\sum_{k=1}^{n} b_{ij} a_{jk} \right) v_i \end{aligned}$$

and therefore

$$c_{ij} = \sum_{k=1}^{n} b_{ij} a_{jk}$$

which matches the definition of matrix product.

□

Remark. As a consequence matrix multiplication is associative, i.e. $(A_1 A_2) A_3 = A_1 (A_2 A_3)$, since composition of linear maps is associative.

Exercise. Consider $\mathbb{R}^2$ with the standard basis $v_1 = (1, 0), v_2 = (0, 1)$. The linear map

$$\begin{aligned} T : \mathbb{R}^2 \to& \mathbb{R}^2 \\ (1, 0) \mapsto& (a, b) = a v_1 + b v_2 \\ (0, 1) \mapsto& (c, d) = c v_1 + d v_2. \end{aligned}$$

Consider the maps

$$\begin{aligned} T_1 : \mathbb{R}^2 &\to \mathbb{R}^2 & \qquad T_2 : \mathbb{R}^2 &\to \mathbb{R}^2 \\ (1, 0) &\mapsto (1, 1) & (1, 1) &\mapsto (0, 2) \\ (0, 1) &\mapsto (-1, 1) & (1, 0) &\mapsto (1, 1) \\ T_3 : \mathbb{R}^2 &\to \mathbb{R}^2 & T_4 : \mathbb{R}^2 &\to \mathbb{R}^2 \\ (1, 1) &\mapsto (1, 1) & (1, 0) &\mapsto (1, 1) \\ (-1, 1) &\mapsto (-1, 1) & (0, 1) &\mapsto (-1, 1). \end{aligned}$$

Which ones are the same?

Answer: $T_1 = T_2 = T_4$.

Remark. We already saw T_1 has matrix

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

with respect to the basis $(1,0),(0,1)$ on the domain and codomoain and T_3 has the same matrix *with respect to the basis* $(1,1),(-1,1)$.

16.1. **Invertibility.**

Theorem 16.1. *Let V be a finite dimensional vector space over F with basis $v_1, \ldots, v_n$ and $T \in \mathcal{L}(V)$, with $A \in M_n(F)$ the matrix of T with respect to this basis. The following are equivalent:*

(i) T is invertible (i.e. there exists $S \in \mathcal{L}(V)$ such that $TS = ST = \mathrm{id}_V$);
(ii) T is bijective;
(iii) T is injective;
(iv) T is surjective;
(v) there exists $B \in M_n(F)$ such that $AB = BA = I$ where

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

*is the **identity matrix**.*

Proof. (i)$\Longleftrightarrow$(ii) was proved last week. Rank-nullity will tell you that

$$\dim V = \dim \ker T + \dim \operatorname{im} T$$

so that injective means that $\dim V = \dim \operatorname{im} T$ so that $\operatorname{im} T = V$ given that the image is a subspace. An alternate solution is to note that if T is injective then $T(v_1), \ldots, T(v_n)$ are linearly independent and therefore a basis, and if a map sends a basis to a basis it is an isomorphism. So this proves that (ii)$\Longleftrightarrow$(iii) and similarly one proves equivalence of (ii),(iii), and (iv). To prove (i)$\Longleftrightarrow$(v) we take B to be the matrix for S in this basis. Since composition of linear maps corresponds to multiplication of matrices we have that $TS = ST = \mathrm{id}_V$ then $AB = BA = I$ and vice versa. This way we proved (v)$\Longleftrightarrow$(i)$\Longleftrightarrow$(ii)$\Longleftrightarrow$(iii)$\Longleftrightarrow$(iv) and this proves the theorem. □

Definition 16.2. Define

$$GL(V) = \{T \in \mathcal{L}(V) \colon T \text{ is invertible } \}.$$

- $GL(V)$ is not a subspace (since the zero map is not in $GL(V)$).
- However, is S, T are invertible then ST is also invertible and thus $GL(V)$ has a multiplication given by composition, i.e.

$$\begin{aligned} \circ : GL(V) \times GL(V) &\to GL(V) \\ (S, T) &\mapsto S \circ T \end{aligned}$$

  which is associative (since composition of functions is associative), has an identity element id_V and every element $T \in GL(V)$ has an inverse $T^{-1} \in GL(V)$. This makes $GL(V)$ into a group, the **general linear group**.

In some sense this is the most important group in linear algebra, and we will come back to it soon.

16.2. **Eigenvectors and Eigenvalues.** We are going to continue to study linear maps, but now with a different perspective. Let $T \in \mathcal{L}(V)$.

Definition 16.3. We say that $v \neq 0$ is an **eigenvector** of T if $T(v) = \lambda v$ for some $\lambda \in F$. λ is called an **eigenvalue**.

Remark. Note that if v is an eigenvector then v preserves the line spanned by v (i.e. T acts on v by simply "stretching" it).

Question. Does T have an eigenvector? Does T have a *basis* of eigenvectors?

Note that if $v_1, \ldots, v_n \in V$ are a basis of eigenvectors (i.e. $T(v_i) = \lambda_i v_i$) then the matrix has the form

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

Such a matrix is called a **diagonal matrix**.

Example 16.4. Let

$$\begin{aligned} T : \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x, y) &\mapsto (3x - y, 3y - x). \end{aligned}$$

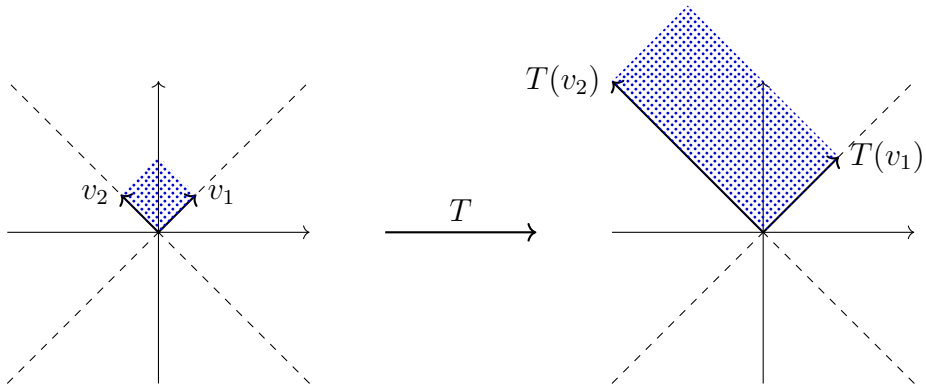
With respect to the standard basis the matrix is given by

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

We claim that T has eigenvalues $\binom{1}{1}, \binom{-1}{1}$. In fact

$$\begin{aligned} A \begin{pmatrix} 1 \\ 1 \end{pmatrix} &= \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ A \begin{pmatrix} -1 \\ 1 \end{pmatrix} &= \begin{pmatrix} -4 \\ 4 \end{pmatrix} = 4 \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \end{aligned}$$

We can represent this graphically in the following way:



Example 16.5. Let$T$ be such that

$$(1,0) \mapsto (0,0)$$
$$(0,1) \mapsto (1,0).$$

Then the matrix for T is

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

We saw last time that A^2. Does A have a basis of eigenvectors? If so, the resulting matrix is

$$B = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

and therefore we must have that

$$B^2 = \begin{pmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

which implies $\lambda_1 = \lambda_2 = 0$, which would imply that T is the zero map. But this is not the case, and therefore A does not admit a basis of eigenvectors.

Questions:

(a) When does T have a basis of eigenvalues?
(b) When does T have any eigenvector?
(c) How do we find eigenvectors/eigenvalue?

The surprising answer is that the math of eigenvectors/eigenvalues is based on the algebra of polynomials. A naive approach to question (c) would be to transform the eigenvalue equation into a system of linear equation. With the matrix from the previous example,

$$\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

yields the system

$$3x - y = \lambda x$$
$$3y - x = \lambda y$$

which implies $x = (3 - \lambda)y$ and $(8 - 6\lambda + \lambda^2) = 0$, and since y cannot be 0 (otherwise x would be zero) we get that $\lambda = 2, 4$, hence the answer.

16.3. **Polynomials and roots.** Let $p \in \text{Poly}(F)$ with $p = a_n x^n + \cdots + a_1 x + a_0$. Then for all $\lambda \in F$ we have that

$$p(\lambda) = a_n \lambda^n + \cdots + a_1 \lambda + a_0 \in F.$$

Definition 16.6. We say that λ is a **root** of p if $p(\lambda) = 0$.

Theorem 16.7 (Euler's root theorem)**.** *If λ is a root of p then*

$$p = (x - \lambda)q$$

for some $q \in \text{Poly}(F)$.

Corollary 16.7.1. *A nonzero polynomial of degree n has at most n roots.*

Proof. By induction. The base case has that for $\deg p = 1$ we can write $p = ax + b$ and therefore there is exactly one root $\lambda = -a^{-1}b$. By induction if the degree of p is n and p has a root λ we can write $p = (x - \lambda)q$ and since q has degree $n - 1$, by induction it has at most $n - 1$ roots. $\square$

17. 10-13

"In mathematics you don't understand things. You just get used to them."
–John von Neumann

Last time we left with some questions:

- When does $T : V \to V$ have an eigenvector?
- How many/few eigenvectors can T have?

The main tool to answer these questions turns out to be polynomials.

How many/few roots can a polynomial $p \in \text{Poly}(F)$ have?

17.1. **Division algortihm.** Poly(F) and $\mathbb{Z}$ have a lot in common. They admit factoring, they have addition, they have a multiplication which is associative and commutative. Yet, their multiplication is in general not invertible (for instance, $1/x$ is not a polynomial). Nevertheless, both have *division with remainder*.

17.1.1. *Division algorithm.*

Theorem 17.1 (Division algorithm)**.** *Given $p, q \in \text{Poly}(F)$ there exists $s, r \in \text{Poly(F)}$ such that $\deg(r) < \deg(s)$ and $p = sq + r$.*

Example 17.2. Let $p = 3x^2 + 2x - 1$ and $q = 2x + 8$. Then $p = (3x/2 - 5)q + 39$.

General proof. Write

$$\begin{aligned} p =& a_n x^n + \cdots + a_1 x + a_0 \\ q =& b_m x^m + \cdots + b_1 x + b_0. \end{aligned}$$

If $m > n$ then $p = 0 \cdot q + p$ and there's nothing left to prove. Otherwise, try

$$s_1 = \frac{a_n}{b_m} x^{n-m}.$$

We get that

$$\begin{aligned} r_1 =& p - q s_1 \\ =& \left(\cancel{a_n x^n} + \cdots + a_1 x + a_0\right) - \frac{a_n}{b_m} x^{n-m} \left(\cancel{b_m x^m} + \cdots + b_1 x + b_0\right). \end{aligned}$$

Note that $\deg(r_1) < \deg(p)$. Now, if $\deg(r_1) < \deg(q)$, we stop. Otherwise repeat with r_1 and q, namely write

$$\begin{aligned} r_1 =& s_2 q + r_2 \\ r_2 =& s_3 q + r_3 \\ & \vdots \\ r_k =& s_{k+1} q + r_{k+1} \end{aligned}$$

with $\deg(r_k) < \deg(q)$ for all k. At the end we substitute

$$p = s_1 q + r_1 = s_1 q + (s_2 q + r_2) = \cdots = (s_1 + s_2 + \cdot + s_{k+1})\, q + r_{k+1}.$$

□

More fun proof (using linear algebra). Let $\mathrm{Poly}_k \subset \mathrm{Poly}(F)$ be the subspace of polynomails with degree k. Given p, q with $\deg p > \deg q$ we definie a map

$$\begin{aligned} T : \mathrm{Poly}_{n-m} \times \mathrm{Poly}_{m-1} &\to \mathrm{Poly}_n \\ (s, r) &\mapsto sq + r. \end{aligned}$$

We want to show that T is surjective. We note that

$$\begin{aligned} \dim \mathrm{Poly}_{n-m} \times \mathrm{Poly}_{m-1} =& (n - m + 1) + (m - 1 + 1) \\ =& n + 1 \\ =& \dim \mathrm{Poly}_n . \end{aligned}$$

Since T is linear, it then suffices to show that T is injective (from exercise proved last time). Namely, we want to show that $\ker(T) = 0$. Suppose

$$0 = T(s, r) = sq + r.$$

This implies that $sq = -r$. If s, r are nonzero this implies that

$$m = \deg q \leq \deg(sq) = \deg(r) \leq m - 1$$

by assumption, which is a contradiction. Therefore $s = r = 0$, that is to say T is injective. □

Corollary 17.2.1 (Euler's root theorem)**.** *If λ is a root of $p \in \mathrm{Poly}(F)$ then $p = (x - \lambda)s$ for some $s \in \mathrm{Poly}(F)$.*

Proof. By the division algorith,

$$p = s(x - \lambda) + r.$$

We want to show that $r = 0$. We know from the definition of remainder that $\deg r < \deg(x - \lambda)$, and therefore r is a constant. From the above expression we get that

$$0 = p(\lambda) = s(\lambda - \lambda) + r(\lambda) = r$$

and therefore $r = 0$. □

Corollary 17.2.2. *A polynomial of degree n has at most n roots.*

Example 17.3. Consider $p = x^2 - 1$ as an element of $\mathrm{Poly}(F)$ where $F = \mathbb{Z}/8\mathbb{Z}$. We see that the roots are $1, 3, 5, 7$. Why are there 4 roots despite the fact that the polynomial has degree 2? The answer is that F in this case is not a field. In fact, the division algorith *required* F to be a field, since we wrote things like a_n/b_m, but in $\mathbb{Z}/8\mathbb{Z}$ not all b_m's have inverses.

18. 10-16

The main question we are trying to answer is how many/few eigenvectors can a linear operator have? The answer to this question is related to the number of roots of a polynomial, and we saw last time that a polynomial of degree m has at most m roots. Can we be more precise?

18.1. **Fundamental theorem of algebra.**

Example 18.1. The polynomial x^m has only 1 root with multiplicity m (the root is 0). The polynomial $x^2+1 \in Poly(\mathbb{R})$ has no roots, however $x^2+1 \in Poly(\mathbb{C})$ has two roots, namely $\pm i$.

Theorem 18.2 (Fundamental theorem of algebra)**.** *Every polynomial $p \in Poly(\mathbb{C})$ with $\deg(p) \geq 1$ has a root.*

Proof. Next semester (we are going to need analysis). □

Corollary 18.2.1 (Polynomial factorization theorem)**.** .

(a) *Every $p \in Poly(\mathbb{C})$ is (uniquely) a product of linear factors, i.e.*

$$p = (x-\lambda_1)\cdots(x-\lambda_m)$$

with $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$.

(b) *Every $p \in Poly(\mathbb{R})$ is uniquely a product of linear and irreducible quadratic factors (x^2+ax+b is* irreducible *if it has no real roots).*

The proof combines several basic facts.

Definition 18.3. For $z = a+ib \in \mathbb{C}$ the **complex conjugate** $\overline{z}$ is equal to $\overline{z} = a-ib$.

Easy properties.

(i) z is real if and only if $\overline{z} = z$ (since $\overline{z} = z \iff a+ib = a-ib \iff b = 0$)
(ii) $\overline{z+w} = \overline{z} + \overline{w}$
(iii) $\overline{zw} = \overline{z} \cdot \overline{w}$
(iv) $\overline{\overline{z}} = z$

Definition 18.4. For $p = a_n x^n + \cdots a_1 x + a_0 \in Poly(\mathbb{C})$ the complex conjugate $\overline{p}$ is defined as

$$\overline{p} = \overline{a_n} x^n + \cdots + \overline{a_1} x + \overline{a_0}$$

and $p = \overline{p}$ if and only if $p \in Poly(\mathbb{R})$.

Lemma 18.4.1. *If $p \in Poly(\mathbb{R}) \subset Poly(\mathbb{C})$ and if $\lambda \in \mathbb{C}$ is a root of p, then $\overline{\lambda}$ is also a root.*

Proof. We know that $p(\lambda) = 0$, and we want to show that $p(\overline{\lambda}) = 0$ as well. First, let

$$p = a_n x^n + \cdots + a_1 x + a_0.$$

By assumption, $a_i \in \mathbb{R}$ for all $0 \leq i \leq n$. Then

$$\begin{aligned}
0 =& \overline{0} \\
=& \overline{a_x \lambda^n + \cdots + a_1 x + a_0} \\
=& \overline{a_n}\overline{\lambda^n} + \cdots + \overline{a_1}\overline{\lambda} + \overline{a_0} \\
=& a_n \overline{\lambda}^n + \cdots + a_1 \overline{\lambda} + a_0 \quad \text{since } p = \overline{p} \\
=& p(\overline{\lambda}).
\end{aligned}$$

□

Observe that for $\mu \in \mathbb{C}$ we have

$$(x - \mu)(x - \overline{\mu}) = x^2 - (\mu + \overline{\mu})\, x + \mu\overline{\mu} \in Poly(\mathbb{R})$$

since

$$\begin{aligned}
\overline{\mu + \overline{\mu}} =& \overline{\mu} + \overline{\overline{\mu}} = \overline{\mu} + \mu \\
\overline{\mu\overline{\mu}} =& \overline{\mu} \cdot \overline{\overline{\mu}} = \overline{\mu}\mu.
\end{aligned}$$

Proof of corollary. We prove part (a) by induction. For $\deg p = 1$ the fundamental theorem of algebra tells us that p has one root, and $\deg p = 1$. Suppose this works for all polynomials with degree $< n$. Then for $\deg p = n$ we can write

$$p = (x - \lambda)q$$

with $\deg q = n - 1$ and therefore q itself is uniquely a product of linear factors. To prove part (b), we know that p has roots $\lambda_1, \cdots, \lambda_n$. Suppose $\lambda_1, \ldots, \lambda_k$ are real, and $\lambda_{k+1}, \ldots, \lambda_n$ are not. Then by the lemma we can write

$$\{\lambda_{k+1}, \ldots, \lambda_m\} = \{\mu_1, \overline{\mu}_1, \ldots, \mu_m, \overline{\mu}_m\}.$$

Then

$$\begin{aligned}
p =& \prod_{i=1}^{n} (x - \lambda_i) \\
=& \prod_{i=1}^{k} (x - \lambda_i) \prod_{j=1}^{m} (x^2 - (\mu + \overline{\mu})x + \mu\overline{\mu}).
\end{aligned}$$

□

18.2. **Eigenvectors existence.** First of all, eigenvectors don't always exist!

Example 18.5. The map

$$\begin{aligned}
T : \mathbb{R}^2 \to& \mathbb{R}^2 \\
(x, y) \mapsto& (-y, x)
\end{aligned}$$

is a rotation counterclockwise by 90° and therefore it does not preserve any lines.

Theorem 18.6. *Let V be a complex vector space of finite dimension. Every $T \in \mathcal{L}(V)$ has an eigenvector.*

To prove this, we want to use polynomials. For $p = a_n x^n + \cdots + a_0 \in Poly(\mathbb{C})$ and $T \in \mathcal{L}(V)$ we can define

$$p(T) := a_n T^n + \cdots + a_1 T + a_0 I \in \mathcal{L}(V).$$

Moreover, if $p = (x - \lambda_1) \cdots (x - \lambda_n)$ then we can also write $p(T)$ as a composition, i.e.

$$p(T) = (T - \lambda_1 I) \cdots (T - \lambda_n I).$$

Idea of proof. Given $T \in \mathcal{L}(V)$ we try to find a polynomial $q \in Poly(\mathbb{C})$ so that $q(T)$ is *not* injective. Then

$$q(T) = (T - \mu_1 I) \cdots (T - \mu_n I)$$

is not injective, and so $(T - \mu_i I)$ is not injective for some i, namely there exists $v \in V - \{0\}$ such that $(T - \mu_i I)v = 0$, i.e. $Tv = \mu_i v$.

19. 10-18

19.1. **Eigenvector existence.** Recall that last time we stated the following theorem:

Theorem 19.1. *Let V be a complex vector space of finite dimension. Every $T \in \mathcal{L}(V)$ has an eigenvector.*

The idea of the proof is that T has an eigenvector if and only if $\ker(T - \lambda I) \neq \{0\}$ for some λ. The proof is divided in three steps:

Step 1. Find $p = a_0 + a_1 x + \cdots + a_n x^n \in \text{Poly}(\mathbb{C})$ such that $p(T)$ is not injective.

Step 2. Factor $p = (x - \lambda_1) \cdots (x - \lambda_n)$ by the fundamental theorem of algebra.

Step 3. Conclude that

$$\{0\} \neq \ker p(T) = \ker(T - \lambda_1 I) \cdots (T - \lambda_n I)$$

and therefore $\ker(T - \lambda_j I) \neq 0$ for some $1 \leq j \leq n$.

The only part of the proof that we actually need to develop is step 1, since the others follow.

Proof. Fix nonzero vector $u \in V$. There's not a lot that we can do with u, therefore we can just consider $u, Tu, T^2u, \ldots$ and note that since $\dim V = n$ we have that the list $u, Tu, \ldots, T^n u$ is linearly dependent. Thus there exist scalars, not all of them zero, such that

$$\begin{aligned} 0 =& b_0 u + b_1 Tu + \cdots + b_n T^n u \\ =& (b_0 I + b_1 T + \cdots + b_n T^n) u. \end{aligned}$$

This finishes step 1 and therefore proves the theorem. □

Remark. This says that eigenvectors exist, but doesn't tell us how to find them.

Example 19.2. Last time we saw that the map

$$\begin{aligned} S : \mathbb{R}^2 \to& \mathbb{R}^2 \\ (x, y) \mapsto& (-y, x) \end{aligned}$$

has no eigenvector. However the theorem says that

$$\begin{aligned} T : \mathbb{C}^2 \to& \mathbb{C}^2 \\ (x, y) \mapsto& (-y, x). \end{aligned}$$

We can follow the steps of the proof so that we can find an eigenvector. Fix a vector $u = (1, 0)$. We see that

$$Tu = (0, 1) \qquad T^2 u = (-1 = 0)$$

so that the linear dependence relation becomes

$$\begin{aligned} 0 =& T^2 u + u \\ =& (T^2 + I) u \end{aligned}$$

so that $p = x^2 + 1 = (x + i)(x - i)$. Therefore either $\ker(T - iI) = 0$ or $\ker(T + iI) = 0$. By solving the system of equations for

$$T(x, y) = i(x, y)$$

we find $T(1, -i) = (i, 1) = i(1, -i)$ and $T(1, i) = (-i, 1) = -i(1, i)$.

Example 19.3. The theorem does not hold for the infinite dimensional case. For example, the operator

$$T(z_1, z_2, \dots) = (0, z_1, z_2, \dots)$$

has no eigenvector.

Question: Does every $T \in \mathcal{L}(V)$ for $\dim V \geq 2$ have two eigenvectors?

The silly answer is that v is an eigenvector then any scalar multiple is as well. But does T have two linearly independent eigenvectors?

Example 19.4. It is possible for $T \in \mathcal{L}(V)$ to have *only one* eigenvector (up to scaling), i.e.

$$\begin{aligned} T : \mathbb{C}^2 &\to \mathbb{C}^2 \\ (x, y) &\mapsto (x + y, y). \end{aligned}$$

The matrix for T with respect to the standard basis is

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

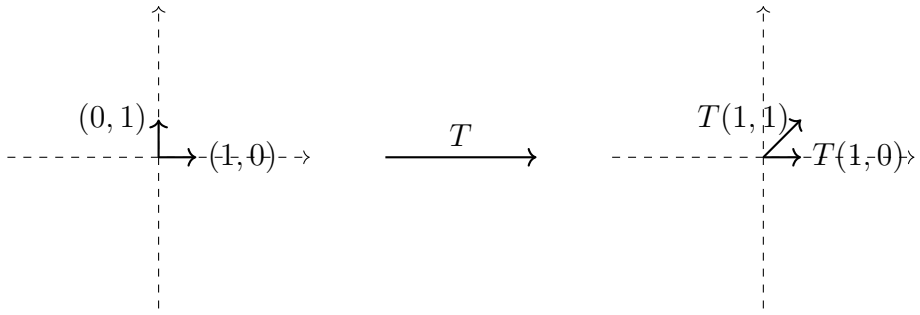
What are the possible eigenvalues? For an eigenvalue λ we need $T - \lambda I$ to be not injective, i.e. not invertible. Since

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

and we showed in the homework that a 2×2 matrix is invertible if and only if $ad - bc \neq 0$, in this case we have that $ad - bc = (1 - \lambda)^2$ so that it is zero only for $\lambda = 1$. This means that 1 is the only eigenvalue. If T has 2 linearly independent eigenvectors then there is a basis of $\mathbb{C}^2$ on which the matrix of T is

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

But this is the identity and T is not the identity (by definition), therefore T cannot have two linearly independent eigenvectors. We can draw the map restricted to $\mathbb{R}^2$: so that we



see that T fixes the x axis but skews the rest.

Question: What is the most eigenvector that $T \in \mathcal{L}(V)$ can have?

Example 19.5. For $T = I_V$ every nonzero vector is an eigenvector with eigenvalue 1.

How about *distinct eigenvalues*?

Proposition 19.1. *Let V be a vector space over F and $T \in \mathcal{L}(V)$. If $v_1, \ldots, v_m \in V$ are eigenvectors with distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ then $v_1, \ldots, v_m$.*

Proof. Assume $v_1, \ldots, v_m$ are linearly dependent. By the linear dependence lemma there is some $k \leq m$ so that $v_1, \ldots, v_{k-1}$ are linearly independent and $v_k \in \text{span}(v_1, \ldots, v_{k-1})$, i.e.

$$v_k = a_1 v_v + \cdots a_{k-1} v_{k-1}.$$

Then

$$\begin{aligned} \lambda_k V =& T(v_k) \\ =& T(a_1 v_v + \cdots a_{k-1} v_{k-1}) \\ =& a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1} \end{aligned}$$

and upon rearranging we find that

$$0 = a_1(\lambda_k - \lambda_1)v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1})$$

and since $v_1, \ldots, v_{k-1}$ are linearly independent we have that

$$a_1(\lambda_k - \lambda_1) = \cdots = a_{k-1}(\lambda_k - \lambda_{k-1}) = 0$$

and since at least one of the a_i is nonzero we have that $\lambda_k = \lambda_i$, contradicting our hypothesis. □

19.2. **Eigenvectors for $T \in \mathcal{L}(V)$ for V real.** Let V be a real vector space. The goal is to find an eigenvector of T, and if we fail (which we will), try to find something else.

Going back to the proof of theorem 19.1 we see that the steps become

Step 1. Find $p = a_0 + a_1 x + \cdots + a_n x^n \in \text{Poly}(\mathbb{C})$ such that $p(T)$ is not injective.

Step 2. Factor

$$p = \prod_{i=1}^{k} (x - \lambda_i) \prod_{j=1}^{m} (x^2 - (\mu_j + \overline{\mu})x + \mu\overline{\mu})$$

by the fundamental theorem of algebra.

Step 3. Conclude that one of the factors of $p(T)$ is not injective, i.e. if T has no eigenvector it has a 2-dimensional invariant subspace (more about this next time).

20. 10-20

20.1. **Last time.**

Theorem 20.1 (Eigenvector existence)**.** *Let $T \in \mathcal{L}(V)$. Then*

(a) if V is complex, T has an eigenvector;
(b) if V is real, T either has an eigenvector or has a 2 dimensional invariant subspace.

Definition 20.2. A subspace $W \subset V$ is **invariant** under T if $w \in W$ implies $Tw \in W$.

Example 20.3. An eigenvector is equivalent to a 1-dimensional invariant subspace.

An invariant subspace is therefore a generalization of the notion of eigenvector.

Lemma 20.3.1. *Suppose $T \in \mathcal{L}(V)$ and $\ker(T^2 + aT + b) \neq 0$. Then T has a 2-dimensional invariant subspace.*

Proof. Take $u \neq 0$ such that

$$T^2u + aTu + bu = 0.$$

We will show that $W = \operatorname{span}(u, Tu)$ is invariant under T. For $w \in W$ we can write $w = cu + dTu$ for some $c, d \in F$. Then

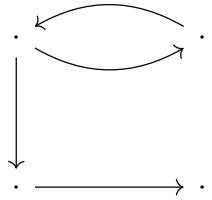
$$\begin{aligned} Tw =& T(cu + dTu) \\ =& cTu + dT^2u \\ =& cTu + d(-aTu - bu) \\ =& (c - ad)Tu - bdu \in W \end{aligned}$$

and therefore W is invariant.

□

20.2. **Google's page-rank algorithm.** (or, how to use eigenvectors to make a fortune)
The problem. You are given a colleciton of webpages (e.g. pages that contain some search words), and want to rank these pages in a "reasonable way". How we can think of such a collection in a mathematical way?

- A collection of webpages can be thought of as a **directed graph**, namely a collection of vertices with some edges between them which have a direction. In this case the vertices are the pages, and the edges are links.



- We want to define an importance function

$$I : \{\text{pages}\} \to [0, \infty) = \{x \in \mathbb{R} \colon x \geq 0\}$$

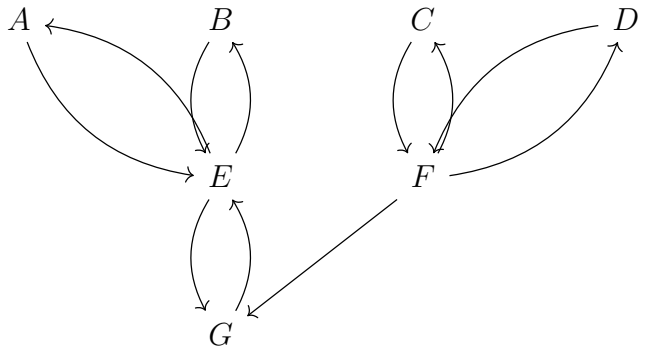
Defining I.

- *Idea/principle 1.* The importance of a page p should depend on how many pages link to p – "the web is a democracy where pages vote for other pages by linking to them".

- *Naive definition #1.* Define

$$\begin{aligned} I(p) &:= \#\text{ pages that link to } p \\ &= \sum_{\substack{q \text{ links} \\ \text{to } p}} 1 \end{aligned}$$

Example 20.4. Consider the case



In this case we have that $G > E = F > A = B = C = D$. However F might be more important because F links to more pages. Maybe G is more important than F because the pages that link to G are more important.

Example 20.5. Consider the case

$$\cdot \longrightarrow A$$

$$\cdot \longrightarrow B$$

$$\cdot \longrightarrow C$$

$$\cdot \longrightarrow D$$

where a the node to the left of A is isolated, the node to the left of B is linked to by many pages but does not link to any, the node to the left of C links to many pages and is linked to by many pages, and the note to the left of D links to many pages but is not linked to by any. In the case of A, noone likes it, so who cares what it thinks. For D, the pages likes D, but it likes everyone, so it makes the fact that it likes D less significant. B is very popular, so you think that the most popular page likes me and only me.

- *Idea #2.* $I(p)$ should depend not only on how many pages link to p, but also how important they are.

Definition 20.6. We define

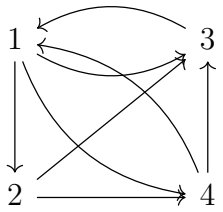
$$I(p) := \sum_{\substack{q \text{ links} \\ \text{to } p}} I(q).$$

Remark. In this "democracy" if q links to many pages, q has many votes. Maybe this is unfair. We will define $N(q)$ as the number of pages that q links to. Therefore we get to our final definiton.

Definition 20.7. Define

$$I(p) = \sum_{\substack{q \text{ links} \\ \text{to } p}} \frac{I(q)}{N(q)}$$

Example 20.8. In the case



We have that

$$\begin{aligned} I(1) =& \frac{I(3)}{1} + \frac{I(2)}{2} \\ I(2) =& \frac{I(1)}{3} \\ I(3) =& \frac{I(1)}{3} + \frac{I(2)}{2} + \frac{I(4)}{2} \\ I(4) =& \frac{I(1)}{3} + \frac{I(2)}{2} \end{aligned}$$

and therefore we can write the left hand side as a linear map

$$\begin{aligned} T : F^4 \to& F^4 \\ (x_1, x_2, x_3, x_4) \mapsto& \left(\frac{x_3}{1} + \frac{x_4}{2}, \frac{x_1}{3}, \frac{x_1}{3} + \frac{x_2}{2} + \frac{x_4}{2}, \frac{x_1}{3} + \frac{x_2}{2} \right), \end{aligned}$$

or, in matrices,

$$\begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix} \begin{pmatrix} I(1) \\ I(2) \\ I(3) \\ I(4) \end{pmatrix} = \begin{pmatrix} I(1) \\ I(2) \\ I(3) \\ I(4) \end{pmatrix}.$$

This means that computing I reduces to finding an eigenvector with eigenvalue 1. We say that A is the **weighted adjacency matrix**: the ith column is the votes of page i weigthed by the total number of votes. However, what if the matrix has no eigenvalues? The answer to this question is that the sum of the entries of each column sum to 1.

Definition 20.9. A matrix where columns sum to 1 is called a **stochastic matrix** (also **probability matrix**, **transition matrix**, **Markov matrix**).

By last week's homework, this matrix has an eigenvector with eigenvalue 1.

21.1. **Satisfied polynomials.**

Definition 21.1. For $T \in \mathcal{L}(V)$ and $p \in \text{Poly}(F)$ we saw T satisfies p if $p(T) = 0 \in \mathcal{L}(V)$. Similarly, if $A \in M_n(F)$ we say that A satisfies p if $p(A) = 0 \in M_n(F)$.

Example 21.2. The matrix

$$A = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

satisfies $p = x - 5$.

Example 21.3. The matrix

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

satisfies $p = x^2$.

Example 21.4. The matrix

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

satisfies $p = (x - 2)(x - 3)$.

Example 21.5. The matrix

$$A = \begin{pmatrix} 0 & 2 \\ 3 & 0 \end{pmatrix}$$

satisfies $p = (x^2 - 6)$.

Example 21.6. The matrix

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

satisfies $p = x^2 - 3x + 1$.

Question. Does every linear operator satisfy *some* nonzero polynomial? If yes, how do we find such polynomial? And what does a satisfied polynomial p tell us about T?

Proposition 21.1. *Suppose $T \in \mathcal{L}(V)$ satisfies $p \in \text{Poly}(F)$. If v is an eigenvector for T with eigenvalue λ then λ is a root of p, i.e. $p = (\lambda) = 0$.*

Corollary 21.6.1. *The matrix*

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

has no eigenvectors.

Proof of the corollary. If A had an eigenvector, its eigenvalue would be a real root of $p = x^2 + 1$ (which is satisfied by A), but this polynomial has no real roots. □

Proof of the proposition. We know that $p(T) = 0$ and $Tv = \lambda v$. Let

$$p = a_0 + a_1 x + \cdots + a_n x^n.$$

Then

$$\begin{aligned} 0 =& p(T)(v) \\ =& \left(a_0 I + \cdots + a_n T^n\right) v \\ =& a_0 I v + \cdots + a_n T^n v \\ =& a_0 I v + a_1 \lambda v + \cdots + a_n \lambda^n v \\ =& \left(a_0 + a_1 \lambda + \cdots + a_n \lambda^n\right) v \\ =& p(\lambda) v \end{aligned}$$

and since $v \neq 0$ we have that $p(\lambda) = 0$. □

21.1.1. *Finding satisfied polynomials.* Assume $\dim V = n$ and $T \in \mathcal{L}(V)$. We want to find p so that $p(T) = 0$.

Remark. Recall that given $v \in V$ we found $p \in \mathrm{Poly}(F)$ so that $p(T)v = 0$ by considering the vectors $v, Tv, \ldots, T^n v$ and observing they were linearly dependent.

A similar trick will work for finding satisfied polynomials.
Consider

$$I, T, T^2, \ldots, \in \mathcal{L}(V).$$

Since

$$\dim \mathcal{L}(V) = \dim M_n(F) = n^2$$

we conclude that there exist $a_0, \ldots, a_{n^2} \in F$ not all of them 0 such that

$$0 = a_0 I + a_1 T + \cdots + a_{n^2} T^{n^2}$$

and therefore T satisfies

$$p = a_0 + a_1 x + \cdots + a_{n^2} T^{n^2}.$$

This proves the following:

Proposition 21.2. *If V is finite dimensional then every $T \in \mathcal{L}(V)$ satisfies some polynomial.*

Alternate proof. For $T \in \mathcal{L}(V)$ we can define a linear map

$$\begin{aligned} \phi : \mathrm{Poly}(F) \to& \mathcal{L}(V) \\ p \mapsto& p(T) \end{aligned}$$

(this is simlar to $p \mapsto p(a) : \mathrm{Poly}(F) \to F$ for fixed a). Thus p satisfies T if and only if $p \in \ker(\phi)$. If we restrict ϕ to $\mathrm{Poly}_{n^2}(F)$ then

$$\dim \mathrm{Poly}_{n^2}(F) = n^2 + 1 > n^2 = \dim \mathcal{L}(V)$$

and therefore $\ker \phi \neq 0$ by rank-nullity. □

Remark. Although we prove the existence of satisfied polynomials, these proofs are not constructive, i.e. they don't tell you how to find p.

Theorem 21.7 (Cayley-Hamilton)**.** *For $A \in M_n(F)$, A satisfies $p_A = \det(xI - A)$.*

Here, det indicates the determinant, which we won't define until later. The proof is therefore postponed.

Example 21.8. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R}).$$

Then A satisfies

$$\begin{aligned} p_A =& \det(xI - A) \\ =& \det \begin{pmatrix} x-a & -b \\ -c & x-d \end{pmatrix} \\ =& (x-a)(x-c) - bc \\ =& x^2 - (a+d)x + ad - bc. \end{aligned}$$

For example

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

satisfies

$$p_A = x^2 - (2+1)x + 2 - 1 = x^2 - 3x - 1$$

as seen in a previous example.

Remark. We see that A has eigenvectors if and only if p_A has real roots. In fact, if p_A has real roots then $p_A = (x-\lambda)(x-\mu)$ and therefore either of the two factors of $p(T)$ is not injective.

The roots of p_A are given by

$$\frac{(a+d) \pm \sqrt{(a+d)^2 - 4(ad-bc)}}{2}$$

Remark. If A is symmetric, i.e.

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

then $(a+d)^2 - 4(ad - b^2) = (a-d)^2 + 4b^2 > 0$ and so p has real roots, meaning A has an eigenvector.

Definition 21.9. A matrix $A = (a_{ij}) \in M_n(F)$ is **symmetric** if $a_{ij} = a_{ji}$.

Theorem 21.10 (Spectral Theorem)**.** *If $A \in M_n(\mathbb{R})$ is symmetric then A has a basis of eigenvectors.*

To understand and prove this theorem we will need some vector geometry.

Example 21.11. In $\mathbb{R}^2$ we can define the norm (or length) of $v = (x, y) \in \mathbb{R}^2$ by

$$|v| = \sqrt{x^2 + y^2},$$

which is the distance from (x, y) to the origin. More generally, for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we have

$$|x| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

We notice that the norm has several interesting properties, such as $|x| \geq 0$ for all $x \in \mathbb{R}^n$ and that $|x| = 0$ if and only if $x = 0$. To introduce the norm we will actually introduce a more general structure, namely the dot product (also inner product), defined as

$$x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

such that $x \cdot x = |x|^2$. This also has interesting properties, such as the Cauchy-Schwartz inequality (more on this next time).

22.1. **Inner products.**

22.1.1. *Definition of an inner product.* Last time we saw an example of a dot product $x \cdot y$ (which will henceforth be denoted $\langle x, y\rangle$) defined by

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i.$$

Properties.

(i) The dot product is bilinear:

$$\begin{aligned}\langle x, y + y'\rangle &= \sum x_i(y_i + y_i') \\ &= \sum x_i y_i + \sum x_i y + i' \\ &= \langle x, y\rangle + \langle x, y'\rangle\end{aligned}$$

and similarly $\langle x + x', y\rangle = \langle x, y\rangle + \langle x', y\rangle$ and for $c \in \mathbb{R}$ we have $\langle cx, y\rangle = \langle x, cy =\rangle = c\langle x, y\rangle$.

(ii) Symmetry:

$$\langle x, y\rangle = \langle y, x\rangle$$

(iii) Positivity:

$$\langle x, x\rangle \geq 0$$

and

$$\langle x, x\rangle = 0 \iff x = 0$$

Definition 22.1. If V is a finite dimensional *real* vector space, a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is an **inner product** if it satisfies properties (i) through (iii). We will call a vector space with an inner product an **inner product space**.

Example 22.2. The dot product on $\mathbb{R}^n$ is an inner product.

Example 22.3. For $a < b \in \mathbb{R}$ we can define an inner product on Poly($\mathbb{R}$) by

$$\langle p, q\rangle := \int_a^b p(x)q(x)dx$$

Example 22.4. On $\mathbb{R}^2$ we can define

$$\langle x, y\rangle := 5x_1y_1 + 3x_1y_2 + 3x_2y_1 + 2x_2y_2.$$

A nonexample would be

$$\langle x, y\rangle = x_1y_1 - x_2y_2$$

since

$$\langle (0, 1), (0, 1)\rangle = -1$$

and thus is not positive. Also,

$$\langle x, y\rangle = x_1y_2 - x_2y_1$$

is not only not positive but also not symmetric.

Remark. For vector spaces over other fields $F \neq \mathbb{R}$ this definition does not work – we don't always have a notion of being ≥ 0. For example we don't have a notion of ≥ 0 on $\mathbb{Z}/p\mathbb{Z}$ or $\mathbb{C}$.

As with linear maps, to get a handle on inner products we will choose a basis.

Proposition 22.1. *Suppose $v_1, \ldots, v_n$ is a basis on V. Then an inner product is determined by the values $\langle v_i, v_j \rangle$.*

Proof. Suppose $v = c_1 v_1 + \cdots + c_n v_n$ and $v' = c_1' v_1 + \cdots + c_n' v_n$. By linearity,

$$\begin{aligned}\langle v, v' \rangle &= \left\langle \sum_{i=1}^{n} c_i v_i, v' \right\rangle \\ &= \sum_{i=1}^{n} c_i \langle v_i, v' \rangle \\ &= \sum_{i=1}^{n} c_i \left\langle v_i, \sum_{j=1}^{n} c_j' v_j \right\rangle \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j' \langle v_i, v_j \rangle.\end{aligned}$$

□

We can record the values $\langle v_i, v_j \rangle$ in a matrix $A = a_{ij}$, and this matrix will be symmetric.

Example 22.5. On $\mathbb{R}^2$, the inner product

$$\langle x, y \rangle := 5x_1y_1 + 3x_1y_2 + 3x_2y_1 + 2x_2y_2$$

with respect to the standard basis is given by the matrix

$$A = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}$$

since

$$\langle e_1, e_1 \rangle = 5, \quad \langle e_2, e_1 \rangle = \langle e_1, e_2 \rangle = 3, \quad \langle e_2, e_2 \rangle = 2.$$

Moreover we can express $\langle x, y \rangle$ using scalar multiplication.

Example 22.6. In our previous example,

$$\langle x, y \rangle = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

(check this!).

Remark. Not all bases are created equal. For example, consider the basis

$$\begin{aligned}v_1 &= \frac{1}{\sqrt{5}} e_1 \\ v_2 &= \frac{1}{\sqrt{5}} \left(e_2 - \frac{3}{5} e_1 \right).\end{aligned}$$

In this basis the inner product of example 22.5 has matrix

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

From this is clear that the $\langle \cdot, \cdot \rangle$ is positive, since if $v = av_1 + bv_2$ we have

$$\langle v, v \rangle = a^2 + b^2 \geq 0$$

and in particular $\langle v, v \rangle = 0$ if and only if $a = b = 0$.

Definition 22.7. For V inner product space, we say that $v, w \in V$ are **orthogonal** if $\langle v, w \rangle = 0$. A basis $v_1, \ldots, v_n$ is called **orthonormal** if

$$\langle v_i, v_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Remark. We see that with respect to an orthonormal basis the matrix for $\langle \cdot, \cdot \rangle$ is the identity.

Theorem 22.8. *Every real finite dimensional inner product space has an orthonormal basis.*

We will talk about this next time. The proof will also tell us the procedure to find such a basis.

22.1.2. *Three geometric theorems.*

Definition 22.9. For $v \in V$ the **norm** of v is defined as

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

We take the square root so that $||c \cdot v|| = c||v||$.

Theorem 22.10 (Pythagorean theorem)**.** *If u, v are orthognoal we have*

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

Proof. By expanding,

$$\begin{aligned} ||u + v|| &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + \cancel{2\langle u, v \rangle} + \langle v, v \rangle \\ &= ||u||^2 + ||v||^2. \end{aligned}$$

□

Theorem 22.11 (Cauchy-Schwartz inequality)**.** *For all $u, w \in V$ we have*

$$|\langle u, w \rangle| \leq ||u|| \cdot ||w||.$$

Remark. For all $0 \neq w, u \in V$ we can write $u = u_1 + u_2$ so that $u_1 = \lambda w$ and $\langle u_2, w \rangle = 0$. To see this, write

$$u = \lambda w + (u - \lambda w)$$

and solve for λ in the expression

$$0 = \langle u - \lambda w, w \rangle = \langle u, w \rangle - \lambda \langle w, w \rangle$$

so that

$$\lambda = \frac{\langle u, w \rangle}{||w||^2}.$$

Proof of Cauchy-Schwartz. If $w = 0$ then both sides of the inequality are 0. If $w \neq 0$ we can write

$$u = \lambda w + (u - \lambda w)$$

with

$$\lambda = \frac{\langle u, w\rangle}{||w||^2}.$$

Then

$$\begin{aligned} ||u||^2 =& ||\lambda w + (u - \lambda w)||^2 \\ =& ||\lambda w||^2 + ||u - \lambda w||^2 \\ \geq& ||\lambda w||^2 \\ =& \lambda^2 ||w||^2 \\ =& \frac{|\langle u, w\rangle|^2}{||w||^2} \end{aligned}$$

and therefore

$$|\langle u, w\rangle| \leq ||u|| \cdot ||w||.$$

□

On thing you can se from the Cauchy-Schwartz inequality is that

$$-||u||||w|| \leq \langle u, w\rangle \leq ||u||||w||$$

and therefore

$$-1 \leq \frac{\langle u, w\rangle}{||u||||w||} \leq 1$$

so that we the following definition makes sense.

Definition 22.12. The **angle** $\theta \in [0, \pi]$ between u and w is defined as

$$\cos\theta = \frac{\langle u, w\rangle}{||u||||w||}.$$

For example, if $u = \lambda w$ we have

$$\begin{aligned} \cos\theta =& \frac{\lambda\langle w, w\rangle}{|\lambda|||w||||w||} \\ =& \frac{\lambda}{|\lambda|} \\ =& \pm 1 \end{aligned}$$

and therefore $\theta = 0$ or π. This matches our intuition.

Theorem 22.13 (Triangle inequality)**.** *For $u, w \in V$ we have*

$$||u + w|| \leq ||u|| + ||w||.$$

23.1. **Orthonormal bases.**

Theorem 23.1. *Let V be a finite dimensional vector space over $\mathbb{R}$ with an inner product $\langle \cdot, \cdot \rangle$. Then there exists an orthonormal basis for $\langle \cdot, \cdot \rangle$.*

Recall. The following are equivalent:

(a) $\langle \cdot, \cdot \rangle$ is symmetric
(b) for some basis e_j the matrix A_{ij} is symmetric
(c) for any basis e_j the matrix A_{ij} is symmetric

Remark. The content of an inner prooduct is really the positivity!

We can restate the theorem as follows:

Theorem 23.2 (Restated)**.** *Given a symmetric bilinear form $\langle \cdot, \cdot \rangle$ the following are equivalent:*

(1) $\langle \cdot, \cdot \rangle$ is positive;
(2) for some basis v_i the matrix is $A_{ij} = I$, i.e. the identity.

We loos at some examples of forms in $\mathbb{R}^2$, all with respect to the stadard basis.

Example 23.3. The form

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

is clearly positive. However, the form

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

is not, and neither is

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

For the first, $\langle v_1, v_1 \rangle = 0$, and for the second $\langle v_1 - v_2, v_1 - v_2 \rangle = -2$.

Is there a way to easily check if a matrix is positive? It turns out that it is quite hard in general, but there is a simple answer for 2×2 matrices.

Proposition 23.1. *A symmetric form*

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

is positive if and only if $ac - b^2 > 0$ and $a > 0$.

Application. Let $z = f(x, y)$. We can define a bilinear form on $\mathbb{R}^2$ by

$$\langle v, w \rangle := D_v D_w f(\cdot, \cdot).$$

In standard basis,

$$A_{ij} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy}. \end{pmatrix}$$

The form is positive if and only if $f_{xx}f_{yy} - f_{xy}^2 > 0$ and $f_{xx} > 0$, i.e. this forms detects local extrema.

Proof of the theorem. Induction on dimension of V. For $\dim V = 0$ the basis is the empty set which is orthonormal. For the induction step, we take a subspace $W \subset V$ with $\dim W = n-1, \dim V = n$. We construct an orthonormal basis $e_1, \dots, e_{n-1}$ on W (by the induction hypothesis). Take now f_n independent of $e_1, \dots, e_{n-1}$. Consider now

$$\tilde{f}_n = f_n - \sum_{i=1}^{n-1} \langle f_n, e_i \rangle e_i.$$

This way, $\tilde{f}_n$ is orthogonal to each e_i, i.e. $\langle \tilde{f}_n, e_i \rangle = 0$ for all $1 \leq i \leq n-1$. Thus we only need to normalize it, and we define

$$e_n = \frac{\tilde{f}_n}{||\tilde{f}_n||}.$$

In conclusion, $e_1, \dots, e_n$ form an orthonormal basis. □

Remark. This process for finding an orthonormal basis goes by the name of **Gram-Schmidt process**.

Example 23.4. $\mathbb{R}^2$ with standard basis f_1, f_2. If we have a positive symmetric bilinear form

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

we wish to find an orthonormal basis for this form. The first step is just to find a vector and normalize it. We take

$$e_1 = \frac{f_1}{||f_1||} = \frac{f_1}{\sqrt{2}}.$$

We can then define

$$\begin{aligned} \tilde{f}_2 =& f_2 - \langle f_2, e_1 \rangle e_1 \\ =& (0,1) \langle (0,1), \frac{(1,0)}{\sqrt{2}} \rangle \frac{(1,0)}{\sqrt{2}} \\ =& (0,1) - \frac{1}{2}(1,0) \\ =& (-1/2, 1) \end{aligned}$$

and therefore

$$\begin{aligned} e_2 =& \frac{\tilde{f}_2}{||\tilde{f}_2||} \\ =& (-1/2,1) \frac{1}{\sqrt{\frac{1}{2} - 1 + 2}} \\ =& \sqrt{\frac{2}{3}} \left(-\frac{1}{2}, 1 \right). \end{aligned}$$

This theorem really needs finite dimensions (there exists a large class of uncountable dimension vector spaces for which no orthonormal set can be a basis).
Variations on this theme. Let $W \subset V$ be a subspace.

Definition 23.5. We define $W^\perp$ to be the set of all $v \in V$ such that $\langle v, w, \rangle = 0$ for all $w \in W$.

Basic propetries.

(a) $W \cap W^\perp = 0$. In fact, suppose $w \in W \cap W^\perp$. Then $\langle w, w \rangle = 0$ and thus $w = 0$.
(b) If $W_1 \subset W_2$ then $W_1^\perp \supset W_2^\perp$.
(c) $W = \left(W^\perp\right)^\perp$. The inclusion $W \subset \left(W^\perp\right)^\perp$ follows from the fact that if $w \in W$ then for all $v \in W^\perp$ then $\langle w, v \rangle$. But then $\langle v, w \rangle = 0$ and therefore $w \in \left(W^\perp\right)^\perp$. To check the other inclusion, we prove the following.

Lemma 23.5.1. *If $W \subset V$ is a finite dimensional subspace, then*

(a)

$$W \oplus W^\perp = V$$

(b)

$$W = \left(W^\perp\right)^\perp.$$

Proof. (a) Take any $v \in V$. Consider an orthonormal basis $e_1, \ldots, e_n$ for W. Then we write

$$v = \sum \langle v, e_i \rangle e_i + \left(v - \sum \langle v, e_i \rangle e_i \right).$$

We claim that

$$v - \sum \langle v, e_i \rangle e_i \in W^\perp.$$

In fact,

$$\begin{aligned} \left\langle v - \sum \langle v, e_i \rangle e_i, e_j \right\rangle =& \langle v, e_j \rangle - \sum \langle v, e_i \rangle \langle e_i, e_j \rangle \\ =& \langle v, e_j \rangle - \langle v, e_j \rangle \\ =& 0. \end{aligned}$$

(b) Exercise.

□

Corollary 23.5.1. *Let V be a finite dimensional real vector space with inner product $\langle \cdot, \cdot \rangle$. Then any orthonormal set can be extended to an orthonormal basis.*

24. 10-30

24.1. **Polynomial approximation. Problem.** Given a function $f\mathbb{R} \to \mathbb{R}$ the problem is to find a "good" polynomial approximation. For example, consider $f(x) = \sin(x)$. We would like to find some polynomial whose graph is close to that of $\sin(x)$. In fact, we are never going to be able to get *exactly* $\sin(x)$, seeing as it has infinitely many roots. Let's try find "the" polynomial $q \in \mathrm{Poly}_5(F)$ which approximates $\sin(x)$ in the interval $[-\pi, \pi]$.

Remark. You might ask why we are looking specifically at polynomials. The answer is that they are very easy to evaluate and compute with. For example, to estimate $\sin(2)$ without any external resource polynomials come in handy.

This problem has 2 answers:
Answer 1. (Calculus/Taylor series) We interpret "best" to mean that q and $\sin(x)$ have the same value at 0 and have the same first 5 derivatives so that

$$q = f(0) + f'(0)x + \cdots + \frac{f^{(5)}(0)}{5!}x^5.$$

In particular for $\sin(x)$ we get

$$\sin(x) \approx x - \frac{x^3}{6} + \frac{x^5}{120}$$

so $\sin(2) \approx 0.9333$ which is a pretty good approximation compared to $\sin(2) = 0.9092$.
Answer 2. (Linear algebra) We interpret "best" as distance minimizing with respect to the inner product. To be more precise, our vector space is $\mathrm{ContFun}([-\pi, \pi], \mathbb{R})$ (continuous functions) and $\mathrm{Poly}(F)$ is a subspace, so that we want to find the point in $\mathrm{Poly}(F)$ which is closest to $\sin(x)$.

24.2. **Orthogonal projections.** Let V be an inner product space (not necessarily finite dimensional), and $U \subset V$ a finite dimensional subspace. Last time we defined

$$U^\perp = \{w \in V : \langle u, w\rangle = 0 \forall u \in U\}.$$

Last time we saw that $V = U \oplus U^\perp$.

Definition 24.1. Define the **orthogonal projection**

$$\pi : V \to U$$
$$v = u + w \mapsto u \qquad (u \in U, w \in U^\perp).$$

Explicitly, if $e_1, \ldots, e_k$ is an orthonormal basis of U then

$$\pi(v) = \langle e_1, v\rangle e_1 + \cdots + \langle e_k, v\rangle e_k.$$

Example 24.2. In the case $k = 1$ we have

$$v = \langle e_1, v\rangle e_1 + (v - \langle v, e_1\rangle e_1)$$

as we already saw in the proof of the Cauchy-Schwartz inequality.

Properties of π.

(i) $\ker \pi = U^\perp$;
(ii) the restriction of π to U is the identity.

Proof. (i) Using the notation of the definition, if $v \in U^\perp$ then $u = 0$ and $\pi(v) = 0$. If $\pi(v) = 0$, then $u = 0$ and $v \in U^\perp$.

(ii) Writing $v = a_1 e_1 + \cdots + v_k e_k$ we have

$$\begin{aligned}\pi(v) =& \sum \langle v, e_i \rangle e_i \\ =& \sum_i \left\langle \sum_j a_j e_j, e_i \right\rangle \\ =& \sum_i \sum_j a_j \langle e_j, e_i \rangle e_i \\ =& \sum_i a_i e_i \\ =& v\end{aligned}$$

□

Theorem 24.3. *Let V be an inner product space and $U \subset V$ a finite dimensional subspace with orthogonal projection $\pi : V \to U$. Then*

(i) $||\pi(v)|| \leq ||v||$ for all $v \in V$;
(ii) for each $v \in V$ and $x \in U$

$$||x - \pi(v)|| \leq ||x - v||.$$

Remark. Part (i) of the theorem says that applying π never makes vectors longer. Part (ii) says that $\pi(v)$ minimizes the distance from v to U.

Proof. (i) Write $v = u + w$ with $u \in U$, $w \in U^\perp$. Then $\pi(v) = u$ and

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

and therefore $||\pi(v)|| \leq ||v||$.

(ii) Similarly we write

$$v - x = \underbrace{(v - \pi(v))}_{\in U^\perp} + \underbrace{(\pi(v) - x)}_{\in U}$$

and use the pythagorean theorem as before.

□

24.3. **Application to approximation.** Consider the vector space of continuous functions $V = \text{ContFun}([-\pi, \pi], \mathbb{R})$ with inner product

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

In this vector space we have a subspace $U = \mathrm{Poly}_5(\mathbb{R})$ and therefore we have an orhtogonal projection

$$\begin{aligned} \pi : V &\to U \\ f &\mapsto \pi(f) \end{aligned}$$

where $\pi(f)$ minimizes the distance from f to U, i.e. it minimizes

$$||f - q||^2 = \left(\int_{-\pi}^{\pi} (f(x) - q(x))^2 dx \right)^2 .$$

We thus interpret $\pi(f)$ as the polynomial that best approximates f. How do we compute it?

Step 1. We start by computing an orthonormal basis $e_0, e_1, \ldots, e_5$ for U.

Step 2. We write $\pi(f) = \langle f, e_0 \rangle e_0 + \cdots + \langle f, e_5 \rangle e_5$.

None of this is fun to compute, but in principle the way to do it is to start from the standard basis $1, \ldots, x^5$ and apply the Gram-Schmidt algorithm together with the fact

$$\int_{-\pi}^{\pi} x^k dx = \begin{cases} 0 & k \text{ odd} \\ \frac{2\pi^{k+1}}{k+1} & k \text{ even} \end{cases} .$$

In practice it is easier to use a computer (for example Mathematica).

25.1. **Dual spaces and inner products.**

Definition 25.1. For a vector space V the **dual space** V^* is defined as

$$V^* := L(V, F) = \{\phi : V \to F \text{ linear } \}$$

and an element $\phi \in V$ is called a **linear funcitonal**.

25.1.1. *Duality.* The duality appears in many properties.

- For a basis $v_1, \ldots, v_n$ of V there is a nice choice of basis $\phi_1, \ldots, \phi_n$ for V^* (called the **dual basis**) defined by

$$\phi_i(v_j) = \begin{cases} 1 & i = j \\ 0 & \text{else} \end{cases}.$$

- You can think of these basis vectors ϕ_i as picking out the v_i-coordinate of $v \in V$. For example, if we have $v = a_1 v_1, \ldots, a_n v_n$ then

$$\phi_i(v) = a_i$$

  so that we can write

$$v = \phi_1 v_1 + \cdots + \phi_n v_n.$$

- Similarly, given any linear functional, it is determined by its values on $v_1, \ldots, v_n$, and these values give the coordinates of ϕ with respect to the basis $\phi_1, \ldots, \phi_n$. Thus for $v = a_1 v_1, \ldots, a_n v_n$ we can write

$$\begin{aligned} \phi(v) &= \sum a_j \phi(v_j) \\ &= \sum \phi_j(v)\phi(v_j) \\ &= \left(\sum \phi(v_j)\phi_j\right)(v) \end{aligned}$$

  so that since this holds for all $v \in V$ we can write

$$\phi = \sum \phi(v_j)\phi_j.$$

Example 25.2. On $V = \mathbb{R}^2$ we have standard basis $e_1 = (1, 0), e_2 = (0, 1)$ and corresponding dual basis. Thus

$$\phi_1(x, y) = \phi_1(xe_1 + ye_2) = \qquad\qquad x$$

and in general

$$(x, y) = xe_1 + ye_2 = \phi_1(x, y)e_1 + \phi_2(x, y)e_2.$$

Similarly for $\phi \in (\mathbb{R}^2)^*$ we have $\phi(x, y) = ax + by$ for some $a, b \in \mathbb{R}$ and you can check that

$$\phi = \phi(e_1)\phi_1 + \phi(e_2)\phi_2 = a\phi_1 + b\phi_2.$$

25.1.2. *Inner products.* If V has an inner product (in particular, V is real) we can use it to define functionals as follows: for $u \in V$ define

$$\begin{aligned}\langle \cdot, u\rangle : V &\to \mathbb{R}\\ v &\mapsto \langle v, u\rangle.\end{aligned}$$

This map is linear since $\langle \cdot, \cdot\rangle$ is *bi*linear.
IfV is finite dimensional and $e_1, \ldots, e_n$ is an orthonormal basis for V, then $\langle \cdot, e_1\rangle, \ldots, \langle \cdot, e_n\rangle$ form a dual basis. Then for $v \in V$ we can write

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

as we have previously seen (the vector $\langle v, e_i\rangle$ is the projection of v on the subspace $\operatorname{span}(e_i)$). Since $\langle \cdot, e_i\rangle \in V^*$ are a basis, every linear functional $\phi \in V^*$ is a linear combination of them.

Proposition 25.1 (The representation proposition)**.** *If V is a finite inner product space, for every $\phi \in V^*$ there exists a unique $u \in V$ such that $\phi = \langle \cdot, u\rangle$, i.e. $\phi(v) = \langle v, u\rangle$.*

Example 25.3. Consider

$$\begin{aligned}\phi : \mathbb{R}^2 &\to \mathbb{R}\\ (x, y) &\mapsto 2x - 4y\end{aligned}$$

where $\mathbb{R}^2$ is endowed with the usual inner product. In this case, we see that

$$2x - 4y = (x, y) \cdot (2, -4)$$

so that $u = (2, -4)$. In general, for

$$\begin{aligned}\phi : \mathbb{R}^2 &\to \mathbb{R}\\ (x, y) &\mapsto ax + by\end{aligned}$$

we have $u = (a, b)$.

Proof. Given $\phi \in V^*$ choose an orthonormal basis $e_1, \ldots, e_n$ for V and let $b_j = \phi(e_j)$. Write $v = \langle v, e_j\rangle e_j$. Then

$$\begin{aligned}\phi(v) &= \sum \langle v, e_j\rangle \phi(e_j)\\ &= \sum \langle v, b_j e_j\rangle\\ &= \left\langle v, \sum b_j e_j \right\rangle\\ &= \langle v, u\rangle\end{aligned}$$

where $u = b_1 e_1 + \cdots + b_n e_n$. To prove it is unique, suppose $\phi = \langle \cdot, u\rangle = \langle \cdot, u'\rangle$. Then $\langle v, u\rangle = \langle v, u'\rangle$ for all $v \in V$ and therefore

$$\langle v, u - u'\rangle = 0$$

for all $v \in V$. In particular

$$\langle u - u', u - u'\rangle = 0$$

and by positivity this implies $u - u' = 0$. □

25.2. **Adjoints.** The motivation for introducind adjoints is the following theorem:

Theorem 25.4 (Spectral theorem)**.** *If V is an inner product space and $T \in L(V)$, if T is self-adjoint it has an orthonormal basis of eigenvectors.*

25.2.1. *Defining the adjoint.* Fix $T \in L(V)$. We will define a new linear map $T^* \in L(V)$. Fix $u \in V$. Recall that $\langle \cdot, u \rangle$ is a linear functional. Note that $\langle T(\cdot), u \rangle$ is also a linear functional (by linearity of T, and since the composition of two linear maps is linear). By the representation proposition, this linear functional can be written as $\langle T(\cdot), u \rangle = \langle \cdot, u' \rangle$. We will define

$$\begin{aligned} T^* : V &\to V \\ u &\mapsto u', \end{aligned}$$

that is to say, $T^*(u)$ is defined by

$$\langle Tv, u \rangle = \langle v, T^* u \rangle.$$

Remark. This is a weird way to define a linear map. In particular, it is a bt nonintuitive to see what T^* "is".

Proposition 25.2. *T^* is linear.*

Proof. We compute $T^*(cu + c'u')$. By definition we have that for all $v \in V$

$$\begin{aligned} \langle v, T^*(cu + c'u') \rangle &= \langle T(v), cu + c'u' \rangle \\ &= c\langle Tv, u \rangle + c' \langle Tv, u' \rangle \\ &= c\langle v, T^* u \rangle + c' \langle v, T^* u' \rangle \\ &= \langle v, cT^* u + c' T^* u' \rangle. \end{aligned}$$

Since this is true for all $v \in V$, we have $T^*(cu + c'u') = cT^* u + c' T^* u'$. □

Remark. Swapping T^* for T in $\langle \cdot, \cdot \rangle$ as per the definition is a useful trick for working with T^*.

Example 25.5. Let

$$\begin{aligned} T : \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x, y) &\mapsto (ax + by, cx + dy) \end{aligned}$$

so that the corresponding matrix with respect to the standard basis is

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

To calculate the matrix for T^* we need to look at the image of the basis vectors. In particular,

$$\begin{aligned} T^* e_1 &= \langle T^* e_1, e_1 \rangle e_1 + \langle T^* e_1, e_2 \rangle e_2 \\ &= \langle e_1, T e_1 \rangle e_1 + \langle e_1, T e_2 \rangle e_2 \\ &= a e_1 + b e_2 \end{aligned}$$

so that in con conclusion the matrix for T^* is

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

26.1. **More adjoints.** Let V be a finite dimensional inner product space (over $\mathbb{R}$). From last time, recall that for every $T \in L(V)$ we deifned is adjoint $T^* \in L(V)$ by

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

In particular, we say that T is self-adjoint if $T = T^*$. for all $v, w \in V$.

Proposition 26.1. *Let $e_1, \dots, e_n \in V$ be an orthonormal basis. If $A = (a_{ij})$ is the matrix for T in this basis and $B = (b_{ij})$ is the matrix for T^* then $B = A^*$ is the transpose, i.e. $a_{ij} = b_{ji}$.*

Corollary 26.0.1. *$T = T^*$ implies that $A = A^*$, i.e. the matrix is symmetric.*

Proof. We write

$$Te_j = \sum a_{ij}e_i$$

so $a_{ij} = \langle Te_j, e_i\rangle$. If we repeat the same procedure with the adjoint, we find that

$$T^*e_j = \langle b_{ij}e_i$$

and so $b_{ij} = \langle T^*e_j, e_i\rangle$. In conclusion

$$\begin{aligned} b_{ij} =& \langle T^*e_j, e_i\rangle \\ =& \langle e_j, Te_i\rangle \\ =& a_{ji} \end{aligned}$$

where the last equality comes from the symmetry of the inner product. □

Proposition 26.2. *(a)* $\ker T^* = (\operatorname{Im} T)^\perp$
(b) $\operatorname{Im} T^* = (\ker T)^\perp$
(c) $\ker T = (\operatorname{Im} T^*)^\perp$
(d) $\operatorname{Im} T = (\ker T^*)^\perp$.

Proof. The proof of (a) is as follows. Let $u \in \ker T^*$. We want to show that $\langle Tv, u\rangle = 0$ for all $v \in V$. As usual, we need to follow the definition of the adjoint and swap the two maps inside the inner product. We know that $T^*u = 0$ if and only if $\langle T^*u, v\rangle = 0$ for all $v \in V$. But this means that$\langle u, Tv\rangle = 0$ for all $v \in V$, i.e. $u \in (\operatorname{Im} T)^\perp$.

The remaining propositions can be proved similarly. At the same time, since $W = \left(W^\perp\right)^\perp$ we see that (a) implies (d) and so on (since we just take the orthogonal complement of both sides). □

An alternate point of view is the following. Let A be the matrix for T under some orhtonormal basis. Then $\operatorname{Im} T$ is spanned by the columns of A (thus it also goes by the name of "column space"). At the same time, the kernel of T is the orthogonal complement of the row space (i.e. the subspace generated by the rows). In matrix multiplication notation, this means

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \qquad\qquad \begin{pmatrix} (a,b)\cdot(x,y) \\ (c,d)\cdot(x,y) \end{pmatrix}.$$

We can prove (c) this way.

Alternate proof of (c). Let $R \subset V$ be the row space. We know that $R^\perp = \ker T$. But R is a subspace spanned by the columns of A^*, i.e. $R = \operatorname{Im} T$. Therefore $\ker T = (\operatorname{Im} T^*)^\perp$. $\square$

26.2. **Spectral Theorem.** Proviously we have seen that for $T \in L(V)$ there exists a basis for V consisting of eigenvecors if and only if the matrix of T is diagonal with respect to some basis (it is "diagonalizable"). The spectral theorem gives some condition for this to happen.

Theorem 26.1 (Spectral Theorem)**.** *Let V be a finite dimensional inner product space. Then the following are equivalent:*

(i) T is self-adjoint
(ii) V has an orthonormal basis of eigenvectors
(iii) The matrix for T is diagonalizable with respect to some orthonormal basis.

Remark. (ii)$\Longleftrightarrow$(iii) follows from what we said before stating the theorem. (ii)$\Longrightarrow$(i) follows since a diagonal matrix is symmetric. Therefore the hard part is (i)$\Longrightarrow$(ii).

Remark. The reason for the name of the theorem is that the set of eigenvalues of T counting multiplicities is called the **spectrum** of T.

Outline of (i)$\Longrightarrow$(ii). The first step is to note that T has an eigenvector, and the second step uses induction to conclude the proof. In fact, let's assume step 1 and show how step 2 works. As we said, we are going to use induction. The base case, $\dim V = 1$, is just scalar multiplication, in which every matrix is self-adjoint (any unit vector in V is an orthonormal basis). Suppose now that the theorem holds for $\dim U < n = \dim V$, and let $T \in L(V)$ be self adjoint. By step 1 there exists an eigenvector with $Tv = \lambda v$. Consider $U = v^\perp = \{u \in V : \langle u, v\rangle = 0\}$. To apply the induction hypothesis to U we need to show that U is invariant under T. Let's fix a vector $u \in U$. We want to show that $Tu \in U$. But this is equivalent to showing that $\langle v, Tu\rangle = 0$, which follows from

$$\begin{aligned}\langle v, Tu\rangle =& \langle T^*v, u\rangle \\ =& \langle Tv, u\rangle \\ =& \lambda\langle v, u\rangle \\ =& 0.\end{aligned}$$

Thus U is invariant. Consider the restriction $T' : U \to U$. We need to show that T' is self adjoint. But this means that

$$\langle Tv, w\rangle = \langle v, Tw\rangle$$

for all $v, w \in V$ which in particular is true for all $v, w \in U$. Therefore by induction hypothesis there exists an orthonormal basis $e_1, \ldots, e_n$ of U made by eigenvectors of T. Together with $e_1 := v/||v||$ this forms an orthonormal basis of eiegenvectors of V. This reduces the proof of (i)$\Longrightarrow$(ii) to proving that T has an eigenvector.

Recall that the strategy for finding eigenvectors was to first find some $p \in \operatorname{Poly}(\mathbb{R})$ so that $p(T) = 0$. Then we factor p so that $p(T)$ is a composition of the form

$$0 = p(T) = (T - \lambda_1 I) \cdot (T - \lambda_k I)(T^2 - b_1 T + c_1 I) \cdots (T^2 - b_\ell T + c_\ell I).$$

We see that the problem might arise if there are no linear factors, and that is what we are going to address.

27.1. **Ingredients for spectral theorem.** We will assume that V is a finite dimensional real inner product vector space throughout.

Proposition 27.1. *Let $S, T \in L(V)$. Then $(S+T)^* = S^* + T^*$.*

Proof. By definition of adjoint

$$\begin{aligned}\langle v, (T+S)^* u\rangle =& \langle (T+S)v, u\rangle \\ =& \langle Tv, u\rangle + \langle Sv, u\rangle \\ =& \langle v, T^* u\rangle + \langle v, S^* u\rangle \\ =& \langle v, (T^* + S^*)u\rangle\end{aligned}$$

and scince this holds for all $u, v \in V$ we have $(T+S)^* = T^* + S^*$. □

Lemma 27.0.1. *Let $S \in L(V)$ be self adjoint, and $\lambda > 0 \in \mathbb{R}$. THen $S^2 + \lambda I$ is injective.*

Proof. We will show that if $u \neq 0$ then $(S^2 + \lambda I)u \neq 0$. To show this it suffices to show $w \in V$ such that $\langle w, (S^2 + \lambda I)v\rangle \neq 0$. Take $w = u$. Then

$$\begin{aligned}\langle u, (S^2 + \lambda I)u\rangle =& \langle u, S^2 u\rangle + \langle u, \lambda u\rangle \\ =& \langle Su, Su\rangle + \lambda \langle u, u\rangle \\ >& 0\end{aligned}$$

since $\lambda > 0, u \neq 0$ and inner products are positive definite. □

Lemma 27.0.2. *Let $T \in L(V)$ be self-adjoint. Then if $q = x^2 + bx + c \in \mathrm{Poly}(\mathbb{R})$ has no real roots $q(T)$ is injective.*

Proof. We start by completing the square:

$$\begin{aligned}q =& x^2 + bx_c \\ =& x^2 + bx + \left(\frac{b}{2}\right)^2 - \left(\frac{b}{2}\right)^2 + c \\ =& \left(x^2 + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right).\end{aligned}$$

Thus we can write

$$q(T) = \left(T + \frac{b}{2} I\right)^2 + \left(c + \frac{b^2}{4}\right) I$$

and moreover we are done if we can apply Lemma 27.0.1. This means that we need to check that $T + bI/2$ is self-adjoint and $c - b^2/4 > 0$. The latter is true since by the quadratic formula the roots of q are

$$\frac{b \pm \sqrt{b^2 - 4c}}{2}$$

which are not real if and only if $b^2 - 4c < 0$, namely $c - b^2/4 > 0$. Moreover, since T is self-adjoint and any multiple of I is as well it follows from Proposition 27.1 that $T + b/2 \cdot I$ is self-adjoint. □

27.2. **Proof of spectral theorem.** We are now ready to prove the spectral theorem, or at least we can check the last condition for Theorem 26.1. Namely, we are going to prove

Theorem 27.1. *If T is self-adjoint, then there exists an orthonormal basis of V consisting of eigenvectors of T.*

We only need to prove that T has an eigenvector, since last time we proved that we can conlcude the theorem from here by induction.

Proof. Let $p \in \text{Poly}(\mathbb{R})$ such that $p(T) = 0$. By the fundamental theorem of algebra we can factor p as

$$p = (x - \lambda_1) \cdots (x - \lambda_k)(x^2 + b_1 x + c_1) \cdots (x^2 + b_m x + c_m)$$

and therefore

$$0 = p(T) = (T - \lambda_1 I) \cdots (T - \lambda_k I)(T^2 + b_1 T + c_1) \cdots (T^2 + b_m T + c_m).$$

But since all the quadratic terms are all injective it follows that $T - \lambda_i I$ is not injective for some $1 \leq i \leq k$, namely λ_i is an eigenvalue. □

27.3. **Positive operators and isometries.**

Definition 27.2. $T \in L(V)$ is called **positive** if $\langle Tv, v \rangle \geq 0$ for all $v \in V$. T is called an **isometry** (a.k.a. **orthogonal operator**) if $||Tv|| = ||v||$ for all $v \in V$.

Example 27.3. An example of an isometry is a rotation in $\mathbb{R}^2$, i.e.

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Example 27.4. The matrix

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

is positive, as you can check.

Remark. It is easy to chekc if T is postive if T has an orthonormal basis of eigenvectors. In fact, in this case T is positive if and only if all of the eigenvalues are nonnegative.

A nonexample is

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

which is not positive. In fact, its eigenvectors are $(1, 1)$ and $(-1, 1)$ with eigenvalues 1 and -1 respectively. Therefore

$$\left\langle \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\rangle = -2$$

and thus A is not positive.

Remark. Positive operators and isometries are particularly simple. We will see that any operator can be expressed in terms of these (via the so called polar decomposition and singular value decomposition).

Proposition 27.2 (Characterization of isometries)**.** *The following are equivalent:*

(a) T *is an isometry, i.e.* $||Tv|| = ||v||$*;*
(b) T *preserves the inner product, i.e.* $\langle Tv, Tu\rangle = \langle v, u\rangle$ *(*T *preserves angles);*
(c) $T^*T = I$*;*
(d) T *is invertible and* $T^{-1} = T^*$*.*

28.1. **Square roots in $L(V)$.** As usual let V be a finite dimensional real vector space.

Definition 28.1. $S \in L(V)$ is a **square root** of T if $S^2 = T$. In this case we write $\sqrt{T} = S$.

Tihs definition brings up some question, such as: when does T have a square root? And how many can it have? How do we find square roots?

Example 28.2. Consider the case $V = \mathbb{R}$. In this case $L(V) \cong M_1(\mathbb{R}) \cong \mathbb{R}$, namely every operator is of the form $T : x \mapsto \lambda x$. When does such an operator have a square root? Since in this case composition corresponds to multiplication, T has a square root if and only if $\lambda \geq 0$. In particular, if $\lambda > 0$ then T has two distinct square roots, namely $x \mapsto \sqrt{\lambda}x, x \mapsto -\sqrt{\lambda}x$. You might still wonder as to how we actually compute $\sqrt{\lambda}$, but we'll see this later.

For a general V, however, the answer is not as simple.

Example 28.3. The matrix

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

does not have a square root, as you might intuitively expect from its negative coefficients. However, intuition should not be followed too much since

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

and so the matrix on the left has a square root even though all of its entries are negative. Then why does $\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$ not have a square root? Suppose it did. Then

$$\begin{aligned} \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} a^2 + bc & (a+d)b \\ (a+d)c & d^2bc \end{pmatrix} \end{aligned}$$

so that we want $a^2 + bd = 1$ and all the other entries are 0. So either $(a + d) \neq 0$, in which case we would have $a^2 = -1$ or $a + d = 0$ in which case we would have $a^2 + bc = d^2 + bc = 0$ (last entry).

Some properties

- Square roots (when they exist) are never unique. If $S^2 = T$ then $(-S)^2 = T$.
- The number of square roots depends on the operator. For example,

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

  has two square roots, namely

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

while

$$\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$$

has four square roots, namely

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

and the zero matrix has infinitely many square roots of the form

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

with $a \in \mathbb{R}$. The identity also has infinitely many square roots (exercise).

There are cases in which it is easy to say whether a matrix has a square root.

Lemma 28.3.1. *Suppose T is diagonalizable, i.e. there exists a basis of eigenvectors of T. If the eigenvalues are nonnegative then T has a square root.*

Proof. Let $v_1, \ldots, v_n$ be a basis of eigenvectors, i.e. $Tv_i = \lambda_i v_i$ for some $\lambda_i \in \mathbb{R}_{\geq 0}$. Then the map $S : v_i \mapsto \sqrt{\lambda_i} v_i$ has the property $S^2 = T$. □

Remark. In general the number of square roots will depend on the multiplicity of the eigenvalues. Recall theat for an eigenvalue λ, the **multiplicity** of λ is the number of eigenvectors with eigenvalue λ. Tis can also be interpreted as the dimension of the space $\{v \in V : Tv = \lambda v\}$.

28.2. **Positive operators.** Let V be a finite dimensional inner product sapce.

Definition 28.4. We say $T \in L(V)$ is positive if T is self adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$.

Theorem 28.5. *A postivie operator has a unique postive square root.*

Remark. The proof will be an application of the spectral theorem.

Proposition 28.1 (Characterization of positive operators)**.** *For $T \in L(V)$ the following are equivalent:*

(a) T is positive
(b) T is self-adjoint and its eigenvalues are ≥ 0
(c) T has a positive square root
(d) T has a self adjoint square root
(e) $T = S^*S$ for some $S \in L(V)$.

Proof of Proposition 28.1.

- (a)$\Longrightarrow$(b). Let λ be an eigenvalue. Take $v \in V$ such that $Tv = \lambda v$. T is positive so

$$\begin{aligned} 0 \leq & \langle Tv, v\rangle \\ = & \langle \lambda v, v\rangle \end{aligned}$$

  and so $\lambda \geq 0$ since $\langle v, v\rangle = ||v||^2 > 0$.
- (b)$\Longrightarrow$(c). T has an orthonormal basis of eigenvectors, so that the argument follows from Lemma 28.3.1.

- (c)$\Longrightarrow$(d). This follows from the definition of positive, since positive implies self-adjiont.
- (d)$\Longrightarrow$(e). If $T = S^2$ then since $S^* = S$ we have $T = S^*S$.
- (a)$\Longrightarrow$(a). Given $T = S^*S$ take $v \in V$. Then

$$\begin{aligned}\langle Tv, v\rangle =& \langle S^*Sv, v\rangle \\ =& \langle Sv, Sv\rangle \\ =& ||Sv|| \\ \geq& 0.\end{aligned}$$

  Thus we now need to show that T is self-adjoint. But $T^* = (S^*S)^* = S^*(S^*)^* = S^*S = T$. Thus T is positive.

□

29. 11-13

29.1. **Inner products, revisited.** Today we will see that there is a bijection

$$\{\text{inner products on } \mathbb{R}^n\} \leftrightarrow \{\text{positive operators on } \mathbb{R}^n \text{ with positive eigenvalues}\}.$$

Fix a standard basis $e_1, \ldots, e_n \in \mathbb{R}^n$ with the standard inner product $\langle -, - \rangle_0$. Then we can write

$$\langle u, v \rangle = u^T v = \sum_{i=1}^{n} u_i v_i.$$

We can also use this basis to identify

$$L(\mathbb{R}^n) \cong M_n(\mathbb{R})$$

Definition 29.1. We say that a matrix $C_n(\mathbb{R}^n)$ is **positive** if C is symmetric and $\langle Cv, v \rangle_0 \geq 0$ for all $v \in V$.

We now fix another inner product $\langle -, - \rangle$ and represent it as a matrix $A = (a_{ij})$, with $a_{ij} = \langle e_i, e_j \rangle$. We know that $\langle -, - \rangle$ being symmetric implies that A is symmetric, but we also know that not all symmetric matrices gives an inner product, for example the zero matrix. At the same time, $\langle -, - \rangle$ being positive implies that A is positive (since $\langle Av, v \rangle_0 = \langle v, v \rangle \geq 0$), but not all positive matrices give an inner product (the zero matrix is another example).

Last time we saw that A is positive if and only if it can be expressed as a product $A = B^T B$ for some $B \in M_n(\mathbb{R}^n)$. Then

$$\begin{aligned} \langle v, v \rangle =& \langle Av, v \rangle_0 \\ =& \langle B^T B v, v \rangle_0 \\ =& \langle Bv, Bv \rangle_0. \end{aligned}$$

Since $\langle -, - \rangle$ is postive definite, B must be injective, hence invertible.

Conclusion. Let $A \in M_n(\mathbb{R}^n)$ be positive. Then the following are equivalent:

(i) $\langle u, v \rangle = u^T A v$ is an inner product
(ii) $A = B^T B$ for some invertible $B \in M_n(\mathbb{R}^n)$
(iii) A has positive eigenvalues (not just nonnegative).

This indicates the bijection

$$\{\text{inner products on } \mathbb{R}^n\} \cong \{\text{positive operators on } \mathbb{R}^n \text{ with positive eigenvalues}\}.$$

In particular we have inclusions

$$\begin{aligned} M_n(\mathbb{R}^n) \supset& \{\text{symmetric matrices } A = A^T\} \\ \supset& \{\text{positive matrices } A = B^T B\} \\ \supset& \{\text{positive matrices with positive eigenvalues}\}. \end{aligned}$$

29.2. **Operator/matrix decomposition theorems.** In practice it's often helpful to be able to express $T \in L(\mathbb{R}^n)$ (or equivalently $A \in M_n(\mathbb{R}^n)$) in terms of simpler operators/matrices with respect to the standard basis.

Theorem 29.2 (Eigendecomposition)**.** *Let $A \in M_n\mathbb{R}$ and suppose there exists a basis $v_1, \ldots, v_n \in \mathbb{R}^n$ of eigenvectors such that $Av_i = \lambda_i v_i$. Let*

$$V = \left(v_1 \middle| \cdots \middle| v_n \right), \qquad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Then

$$A = VDV^{-1}.$$

Example 29.3. Let

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

The eigenvalues are 3 (with eigenvector $(0, 1)$) and 2 with eigenvector $(-2, 1)$. Therefore

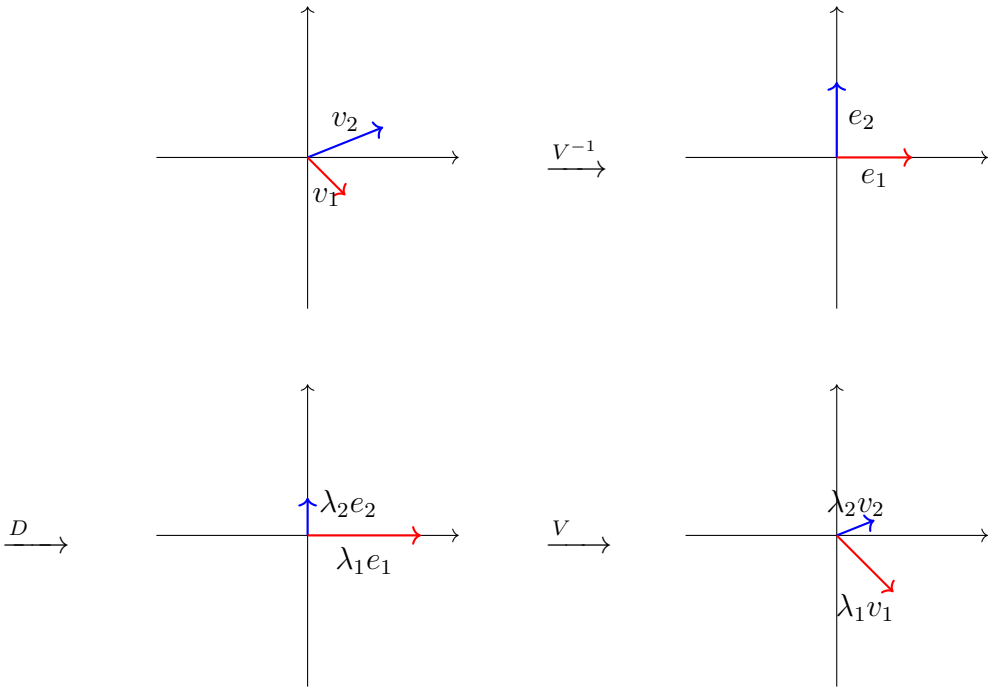
$$\begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 0 & -2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/2 & 1 \\ -1/2 & 0 \end{pmatrix}.$$

Proof of Theorem 29.2. Check $A = VDV^{-1}$ on the basis $v_1, \ldots, v_n$. We have that

$$\begin{aligned} VDV^{-1}(v_i) =& VD(e_i) \\ =& V(\lambda_i e_i) \\ =& \lambda_i v_i \\ =& Av_i. \end{aligned}$$

□

We can informally visualize the proof of this theorem through the following illustration:



Remark. Conversely if $A = VDV^{-1}$ with D diagonal, then the proof shows that the columns of V are eigenvectors entries of D are eigenvalues.

Application. We can apply this to the polynomials of a matrix. If $A = VDV^{-1}$ then $A^2 = VDV^{-1}VDV^{-1} = VD^2V^{-1}$ and so on. This illustrates that eigenvectors for A^n are the eigenvectors for A, and its eigenvalues are the n-th power of those of A. So for $p \in \mathrm{Poly}(\mathbb{R})$ we have $p = \sum_{i=0}^m a_i x^i$ and

$$\begin{aligned} p(A) =& \sum_{i=0}^m a_i \left(VDV^{-1}\right)^i \\ =& \sum_{i=0}^m a_i V D^i V^{-1} \\ =& V \left(\sum_{i=0}^m a_i D^i \right) V^{-1} \\ =& V p(D) V^{-1}. \end{aligned}$$

In particular, if

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

then

$$p(D) = \begin{pmatrix} p(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & p(\lambda_n) \end{pmatrix}.$$

Remark. This gives a computational tool, since D^k is easier to deal with than A^k. Extending the argument, this gives a way to calculate square roots etc., since

$$A^{1/2} = VD^{1/2}V^{-1}$$

(if A has a square root).

Next time we will see polar decomposition.

30. 11-15

30.1. **Matrix decomposition theorems.** Last time we saw eigendecomposition in Theorem 29.2.

Theorem 30.1. *Let $A \in M_n\mathbb{R}$ and suppose there exists a basis $v_1, \ldots, v_n \in \mathbb{R}^n$ of eigenvectors such that $Av_i = \lambda_i v_i$. Let*

$$V = \left(v_1 \middle| \cdots \middle| v_n \right), \qquad D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Then

$$A = VDV^{-1}.$$

A special case happens when A is symmetric. In this case the spectral theorem tells us that there exists an orthonormal basis of eigenvectors with $Av_i = \lambda_i v_i$. We claim that in this case the matrix V is an isometry, namely $V^TV = I$. In fact,

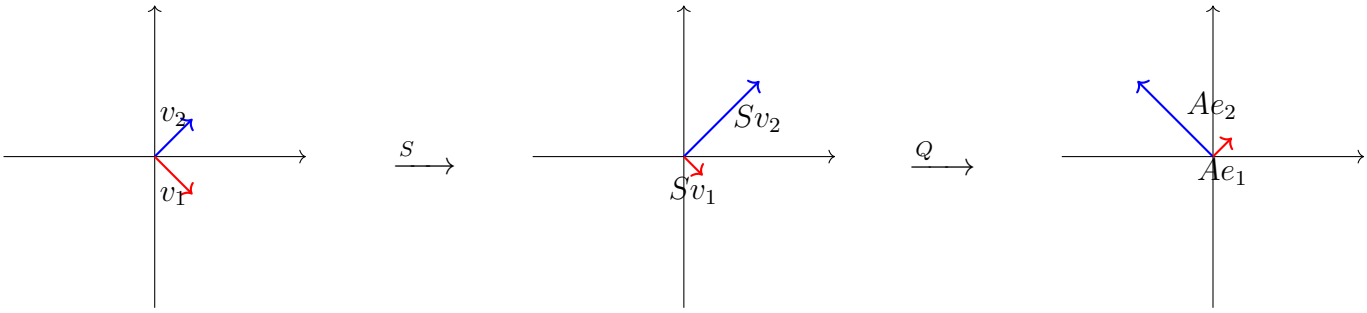
$$\left(\begin{array}{c} v_1 \\ \hline \vdots \\ \hline v_n \end{array}\right)\left(\begin{array}{c|c|c} v_1 & \cdots & v_n \end{array}\right) = \begin{pmatrix} v_1\cdot v_1 & \cdots & v_1\cdot v_n \\ \vdots & \ddots & \vdots \\ v_n\cdot v_1 & \cdots & v_n\cdot v_n \end{pmatrix} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

Thus $A = VDV^{-1} = VDV^T$.

30.1.1. *Polar decomposition.*

Theorem 30.2 (Polar decomposition theorem)**.** *Let $A \in M_n(\mathbb{R})$. Then there exists and isometry Q and a positive S such that $A = QS$. If A is invertible then Q, S are unique.*

For a geometrical picture, we know that S has an orthonormal basis of eigenvectors, so we can visualize the theorem as the following:



Remark. Polar decomposition works for any matrix A, not just for diagonalizable. Polar composition is a generalization of the polar decomposition of complex numbers, namely for $z = x + iy$ we can write $z = re^{i\theta}$ with $r = \sqrt{z\bar{z}} = \sqrt{x^2+y^2}$ and $e^{i\theta} = \cos\theta + i\sin\theta$. Viewing $\mathbb{C} = M_1(\mathbb{C})$ this is just polar decomposition.

We are going to see an incorrect proof at first:

Incorrect proof. Suppose we can write $A = QS$ with $QQ^T = I$ and S positive. Let's now try to deduce something about S and Q. First of all we see that

$$A^TA = (QS)^T(QS) = S^TQ^TQS = S^TS = S^2.$$

Then take $S = \left(A^TA\right)^{1/2}$ (in particular S is unique). Now for $A = QS$ we need $Q = AS^{-1}$. We need to show Q is an isometry. In fact,

$$\begin{aligned} Q^TQ =& (AS^{-1})^T(AS^{-1}) \\ =& (S^{-1})^TA^TAS^{-1} \\ =& S^{-1}S^2S^{-1} \\ =& S^{-1}SSS^{-1} \\ =& I. \end{aligned}$$

□

This proof has several problems. For example, square roots are not necessarily unique. S might not be invertible either. Moreover, why is it that $(S^{-1})^T = S^{-1}$? This latter point is the main one. In fact, if we define S as the square root of AA^T, then since AA^T is positive by construction it is fine to define a unique positive square root. Moreover,

$$S^T(S^{-1})^T = (S^{-1}S)^T = I^T = I$$

and therefore $(S^{-1})^T = (S^T)^{-1}$. Thus we only need to address the fact that S might not be invertible.

Example 30.3. Let us consider

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

We see that

$$A^T A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \left(\frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \right)^2$$

so that in this case

$$S = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$$

is perfectly invertible. This we can write $Q = AS^{-1}$, namely

$$Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

which we can check is unitary.

Going back to our proof, we know that if $A = QS$ then $S = \sqrt{A^T A}$. We want to define an isometry Q such that $A = QS$. Note that Q should take $\operatorname{Im} S$ to $\operatorname{Im} A$. Let's naively define

$$\begin{aligned} Q : \operatorname{Im} S &\to \operatorname{Im} A \\ Sv &\mapsto Av. \end{aligned}$$

There are two issues about this. First of all, it is not well defined, since if $Sv_1 = Sv_2$ then Av_1 is not defined. More importantly, we need Q to be an isometry, and Q is not defined on the whole of V. Q is an isometry if and only if $||Av|| = ||Sv||$ for all $v \in V$. Recall that $S^2 = A^T A$. Therefore

$$\begin{aligned} ||Sv||^2 &= \langle Sv, Sv \rangle \\ &= \langle S^2 v, v \rangle \\ &= \langle A^T A v, v \rangle \\ &= \langle Av, Av \rangle \\ &= ||Av||^2. \end{aligned}$$

For Q being defined on all of $V = \mathbb{R}^n$, write

$$\begin{aligned} \mathbb{R}^n &= \operatorname{Im} S \oplus (\operatorname{Im} S)^\perp \\ \mathbb{R}^n &= \operatorname{Im} A \oplus (\operatorname{Im} A)^\perp. \end{aligned}$$

In particualr, $Q : \operatorname{Im} S \to \operatorname{Im} A$ is an isomorphism, since it's surjective and it's injective (because if $0 = QSv$ then $Av = 0$ and therefore $||Sv|| = 0$), and so we note that

$$\dim \operatorname{Im} S = \dim \operatorname{Im} A.$$

We define $Q : (\operatorname{Im} S)^{\perp} \to (\operatorname{Im} A)^{\perp}$ as any isometry. Therefore for $v = u + w$ with $u \in \operatorname{Im} S, w \in (\operatorname{Im} S)^{-1}$ (this decomposition is unique) we can define $Qv = Qu + Qw$. We can check that Q is linear. We claim that Q is an isometry. In fact, since Qw and Qu are orthogonal (by construction) we can use Pythagoras theorem we can write

$$\begin{aligned} ||Qv||^2 =& ||Qu + Qw||^2 \\ =& ||Qu||^2 + ||Qw||^2 \\ =& ||u||^2 + ||w||^2 \\ =& ||u + w||^2 \\ =& ||v||^2. \end{aligned}$$

31. 11-17

31.1. **Low-rank approximation.** Given $A \in M_{1000}(\mathbb{R})$ (that is, a matrix with a million entries), computation becomes a very expensie process. Therefore we want a matrix $A' \in M_{100}(\mathbb{R})$ which can serve as a good approximatoin. Another option is to look for a matrix $A'' \in M_{1000}(\mathbb{R})$ which is still 1000×1000 but such that the dimension of its image is at most 100. A good idea would be to equip our vector space $M_{1000}(\mathbb{R})$ with a norm, so as to measure the "distance" between two norms. Such norm exists and it is called the Frobenius norm. For $B = (b_{ij}) \in M_n(\mathbb{R})$ we define the frobenius norm as

$$||B||_F = \left(\sum_{ij,=1}^{n} b_{ij}^2 \right)^{1/2}$$

so that this way we see our space of operators as $M_n\mathbb{R} \cong \mathbb{R}^{n^2}$.
Problem. Given $A \in M_n(\mathbb{R})$ and $1 \leq k \leq n$ find $A' \in M_n(\mathbb{R})$ such that $\dim \operatorname{Im} A' = k$ and $||A - A'||$ is as small as possible.

Remark. Previously we saw that if $x \in V$ and $U \subset V$ is a subspaces, then $P_U(x) \in U$ for P_U the orthogonal projection of U is the point on U which is closest to x.

We might want to apply the above ot $V = M_n\mathbb{R}, x = A$, with

$$U = \{A' \colon \dim \operatorname{Im} A' \leq k\}.$$

The problem with this approach is that $\underline{U \text{ is not a subspace}}$. In fact, for $n = 2, k = 1$ we see

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = I$$

and the identity has rank (dimension of image) 2. The solution to this problem can be found in what is known as singular value decomposition.

31.2. **Singular value decomposition. Recap.**

- Eigendecomposition.
  (i) If $S \in M_n\mathbb{R}$ is diagonalizable then
  $$S = VDV^{-1}$$
  where
  $$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}, \qquad V = \left(v_1 \middle| \cdots \middle| v_n \right)$$
  such that $Sv_i = \lambda_i v_i$.
  (ii) if $S = S^T$ is symmetric then
  $$S = VDV^T$$
  with V an isometry and $v_1, \ldots, v_n$ an orthonormal basis.
- Polar decoposition.
  For $\underline{\text{any}}$ $A \in M_n\mathbb{R}$ we can write $A = QS$ with Q an isometry and $S = \sqrt{A^T A}$ positive.

Now fix $A \in M_n\mathbb{R}$ and let $A = QS$.Apply eigendecomposition to S, namely $S = VDV^T$. Then

$$A = QS = QVDV^{-1}.$$

Since Q, V are both isometries, so is QV.

Theorem 31.1 (Singular value decompositon)**.** *For any $A_{\in}M_n\mathbb{R}$ there exist isometries U, V ad diagonal D such that $A = UDV^T$. Moreover the entries of D are the eigenvalues of $\sqrt{A^TA}$.*

Definition 31.2. The eigenvalues of $\sqrt{A^TA}$ are called the **singular values** of A. A may not have eigenvalues but A always has n singular values and they are all ≥ 0.

31.2.1. *Interpretation of U, V.* We see that

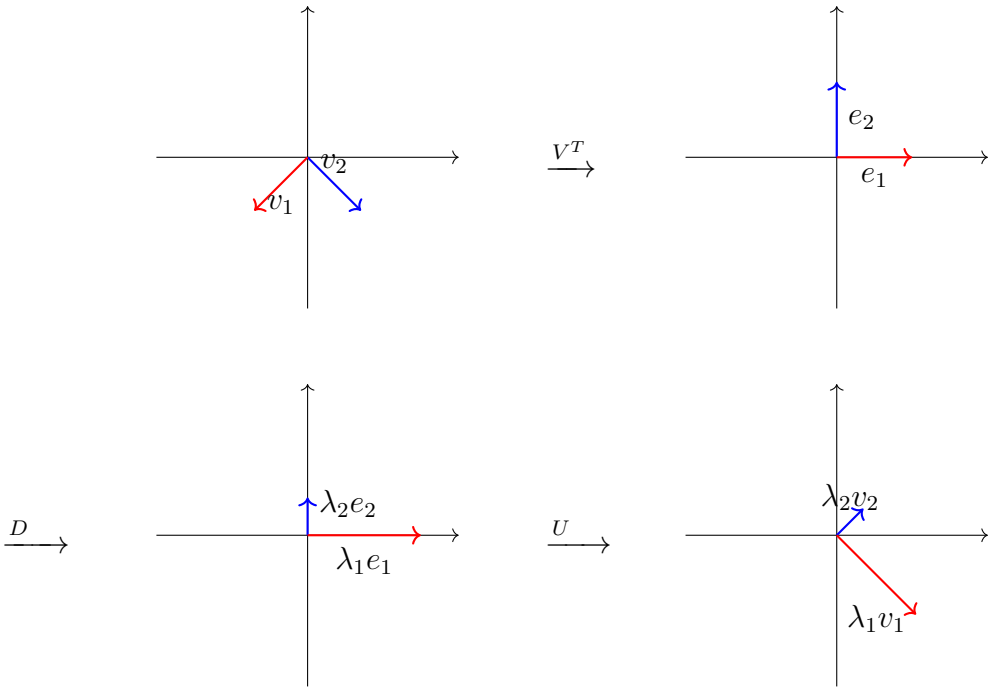
$$A^TA = \left(UDV^T\right)^T \left(UDV^T\right) = VD^T \underbrace{U^TU}_{=I} DV^T = VS^TSV^T = VS^2V^T.$$

This is the eigendecomposition of A^TA, and the columns of V are the eigenvectors of A^TA. Similarly

$$AA^T = UD^2U^T$$

so columns of U are eigenvectors of AA^T.

31.2.2. *Picture of SVD.*



Corollary 31.2.1. *A map $T \in L(\mathbb{R}^n)$ sends a sphere*

$$\{x_1^2 + \cdots + x_n^2 = d^2\}$$

centered at the origin to an ellipse

$$\left\{ \left(\frac{x_1}{a_1}\right)^2 + \cdots + \left(\frac{x_n}{a_n}\right)^2 = d'^2 \right\}$$

centered at the origin.

Alternate statement (cf. Axler). Given$A \in M_n\mathbb{R}$ there exist orthonormal bases $v_1, \ldots, v_n$ and $u_1, \ldots, u_n$ with $\sigma_1, \ldots, \sigma_n \geq 0$ such that

$$Ax = \sigma_1 \langle x, v_1 \rangle u_1 + \cdots + \sigma_n \langle x, b_n \rangle u_n.$$

You can check that the two statements are equivalent, with

$$U = \left(u_1 \middle| \cdots \middle| u_n \right), V = \left(v_1 \middle| \cdots \middle| v_n \right), S = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$$

31.3. **SVD and approximation. Claim.** We can write

$$\begin{aligned} A = UDV^T &= \left(u_1 \middle| \cdots \middle| u_n \right) \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix} \left(\begin{array}{c} v_1 \\ \hline \vdots \\ \hline v_n \end{array} \right) \\ &= \left(\sigma_1 u_1 \middle| \cdots \middle| \sigma_n u_n \right) \left(\begin{array}{c} v_1 \\ \hline \vdots \\ \hline v_n \end{array} \right) \\ &= \sigma_1 u_1 v_1^T + \cdots + \sigma_n u_n v_n^T. \end{aligned}$$

Aside. (useful viewpoints on matrix multiplication). The following will be based on 2×2 matrices but is perfectly valid for larger matrices. Given

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad B = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

we see that

$$AB = \begin{pmatrix} az + bw & ay + bw \\ cz + dw & cy + dw \end{pmatrix}.$$

Therefore we can write the product of matrices in terms of inner product of rows and columns, i.e.

$$\left(\begin{array}{c} v_1 \\ \hline v_2 \end{array} \right) \left(u_1 \middle| u_2 \right) = \begin{pmatrix} v_1 \cdot u_1 & v_1 \cdot u_2 \\ v_2 \cdot u_1 & v_2 \cdot u_2 \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \left(\begin{array}{c} v_1 \\ \hline v_2 \end{array} \right) = \left(\begin{array}{c} av_1 + bv_2 \\ \hline cv_1 + dv_2 \end{array} \right)$$

$$\left(u_1 \middle| u_2 \right) \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \left(xu_1 + zu_2 \middle| yu_1 + wu_2 \right)$$

and in particular

$$\left(\begin{array}{c|c} u_1 & u_2 \end{array}\right)\begin{pmatrix} x & 0 \\ 0 & w \end{pmatrix} = \left(\begin{array}{c|c} xu_1 & wu_2 \end{array}\right).$$

Lastly,

$$\left(\begin{array}{c|c} v_1 & v_2 \end{array}\right)\left(\begin{array}{c} u_1 \\ \hline u_2 \end{array}\right) = v_1u_1 + v_2u_2 = \begin{pmatrix} a \\ c \end{pmatrix}\begin{pmatrix} x & y \end{pmatrix} + \begin{pmatrix} b \\ d \end{pmatrix}\begin{pmatrix} z & w \end{pmatrix}.$$

Given this, we see that if we relabel the singular values such that $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n$ we can set

$$A_k = \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$$

so that A_k solves our initial approximation problem.

32. 11-20

32.1. **Determinants, case study.** Let $V = \mathbb{R}^2$ equipped with the standard inner product. for $w_1, w_2 \in V$ let $P(w_1, w_2)$ be the parallelogram spanned by w_1 and w_2. What follows is a temporary definition of the determinant bases on parallelograms.

Definition 32.1. For a basis v_1, v_2 of V and $T \in L(V)$ we define the determinant of T as

$$\det T = \frac{\operatorname{Area} P(Tv_1, Tv_2)}{\operatorname{Area} P(v_1, v_2)}.$$

Remark. From this definition it is not clear if the determinant is independent of the choice of basis.

Goals.

(1) Define $\det : L(V) \to F$ for any V and any F.
(2) Give an algebraic way to compute it.

We are going to use the above definition to investigate the properties of determinants in general.

(i) If $S, T \in L(V)$ then

$$\det(S \circ T) = \det S \det T.$$

In fact, starting from the left hand side (assuming T is injective, which we'll address later in more generality)

$$\begin{aligned}\det(S \circ T) =& \frac{\operatorname{Area} P(STv_1, STv_2)}{\operatorname{Area} P(v_1, v_2)} \\ =& \frac{\operatorname{Area} P(STv_1, STv_2)}{\operatorname{Area} P(Tv_1, Tv_2)} \frac{\operatorname{Area} P(Tv_1, Tv_2)}{\operatorname{Area} P(v_1, v_2)} \\ =& \det S \det T\end{aligned}$$

since the determinant does not depend on the choice of basis.

(ii) The determinant of the identity is clearly

$$\det I = 1.$$

(iii) For invertibility we see that if T is invertible then

$$\det T \det T^{-1} = \det(TT^{-1}) = \det I = 1$$

so that

$$\det T^{-1} = \frac{1}{\det T}.$$

On the other hand, if T is not invetible then the area of the parallelogram is 0 (since the dimension of the image is at most 1, and the parallelogram gets compressed to a 1-dimensional object). Thus T is invertible if and only if $\det T \neq 0$ (determinants are a useful test for invertibility).

(iv) The determinant can be seen as a function of the columns of a matrix. Fix the standard basis $e_1, e_2 \in \mathbb{R}^2$. Identify $L(\mathbb{R}^2) \cong M_2(\mathbb{R})$, and in turn we identify $M_2(\mathbb{R}) \cong \mathbb{R}^2 \times \mathbb{R}^2$ via the bijection

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \to \left(\begin{pmatrix} a \\ c \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} \right).$$

Therefore, we obtain an expression of the determinant as a map

$$\det : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}.$$

Since the images of the standard basis vectors under T are (a, c) and (b, d), respectively, and the area of the unit square is 1, we get that under this map

$$\left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right) \mapsto \operatorname{Area} P \left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right).$$

By drawing the corresponding parallelograms, we see that

$$\det \left(\begin{pmatrix} a \\ c \end{pmatrix} + \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right) = \det \left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right) + \det \left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right).$$

Since the matrix

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

is not invertible, we get that its determinant is 0. Therefore

$$\begin{aligned} 0 &= \det \left(\begin{pmatrix} a+b \\ c+d \end{pmatrix}, \begin{pmatrix} a+b \\ c+d \end{pmatrix} \right) \\ &= \det \left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right) + \left(\begin{pmatrix} b \\ d \end{pmatrix}, \begin{pmatrix} a \\ c \end{pmatrix} \right) + \det \left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} a \\ c \end{pmatrix} \right) + \det \left(\begin{pmatrix} b \\ d \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right) \end{aligned}$$

and so

$$\det \left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right) = - \left(\begin{pmatrix} b \\ d \end{pmatrix}, \begin{pmatrix} a \\ c \end{pmatrix} \right)$$

Therefore swapping the columns changes the determinant by -1. In particular, this means that we need to allow negative values. Geometrically, we can see this as the determinant giving information about the orientation. As an example, consider the determinants

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

Then the latter map preserves area but reverses "orientation", since it swaps e_1 and e_2 on the unit square.

We can also check geometrically that

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

The takeaway from this analysis is that $\det : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is bilinear form and is alternating (in this case, this means that swapping two columns we get a minus sign). In particular,

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

In fact, the area of a parallelogram is base $\cdot$ height. In this case by letting $(a, c) = v_1, (b, d) = v_2$ we have that

$$\begin{aligned} \text{area}^2 =& ||v_1||^2 \left(||v_2||^2 - \frac{\langle v_1, v_2 \rangle^2}{||v_1||^2} \right) \\ =& (a^2 + c^2) \left((b^2 + d^2) - \frac{(ab + cd)^2}{a^2 + c^2} \right) \\ =& ad - bc. \end{aligned}$$

An alternating proof is by drawing the corresponding parallelogram inscribed in the rectangle with opposite vertices $(0, 0)$ and $(a + b, c + d)$. Thus

$$\begin{aligned} \text{area} =& (a + b)(c + d) - ac - bd - 2bc \\ =& ac + ad + bc + bd - ac - bd - 2bc \\ =& ad - bc. \end{aligned}$$

One last thing that we can see from this is the relation to eigenvalues. In fact, suppose $A \in M_2(\mathbb{R})$ is diagonalizable. Then

$$\det A = \det \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \lambda_1 \lambda_2.$$

32.2. **Determinants abstractly.** Let's start with V a vector space over F.

Definition 32.2. A k-**multilinear form** is a function

$$f : \underbrace{V \times \cdots \times V}_{k \text{ times}} \to F$$

which is linear in each argument when the other coordinates are fixed. For example,

$$f(av_1 + a'v_1', v_2, \ldots, v_n) = af(v_1, v_2, \ldots, v_n) + a'f(v_1', v_2, \ldots, v_n).$$

Definition 32.3. A k-multilinear form is **alternating** if

$$f(v_1, \ldots, v_n) = 0 \qquad \text{if } v_i = v_j \text{ for some } i \neq j.$$

Theorem 32.4 (Determinant Theorem)**.** *There is a unique function*

$$D : \underbrace{V \times \cdots \times V}_{n \text{ times}} \to F$$

such that

(1) D is n-multilinear
(2) D is alternating
(3) $D(I) = 1$

33. 11-27

33.1. **Determinants algebraically.** Last time we saw that we can define the determinant through the following properties (guaranteed by the Determinant Theorem):

Theorem 33.1 (Determinant Theorem)**.** *There is a unique function*

$$D : \underbrace{V \times \cdots \times V}_{n \text{ times}} \to F$$

such that

(1) D is n-multilinear
(2) D is alternating
(3) $D(I) = 1$

Other ways of considering the determinant are:

- As a function

$$\det : M_n(F) \to F$$

  such that
  (i) $\det(AB) = \det A \cdot \det B$
  (ii) $\det A \neq 0 \Leftrightarrow A$ is invertible
  (iii) $\det A^T = \det A$
- As a map

$$\det : L(V) \to F$$

  which is independent of the choice of basis.

For today we will focus on the Determinant theorem.

33.2. **The determinant theorem.** Let V be a vector space over F of dimension n.

Definition 33.2. A k-**multilinear form** is a function

$$f : \underbrace{V \times \cdots \times V}_{k \text{ times}} \to F$$

which is linear in each argument when the other coordinates are fixed. For example,

$$f(av_1 + a'v_1', v_2, \ldots, v_n) = af(v_1, v_2, \ldots, v_n) + a'f(v_1', v_2, \ldots, v_n).$$

Example 33.3. A multilinear $1-$form is a linear functional $\varphi : V \to F$, and a multilinear 2-form is a bilinear function $\varphi : V \times V \to F$.

Definition 33.4. A k-multilinear form is **alternating** if

$$f(v_1, \ldots, v_n) = 0 \qquad \text{if } v_i = v_j \text{ for some } i \neq j.$$

Some immediate consequences. If φ is alternating then the value of φ changes by -1 if we swap two entries, e.g.

$$\varphi(v_1, v_2) = -\varphi(v_2, v_1).$$

This is because

$$\begin{aligned}0 =&\varphi(v_1 + v_2, v_1 + v_2)\\ =&\varphi(v_1, v_1) + \varphi(v_1, v_2) + \varphi(v_2, v_1) + \varphi(v_2, v_2)\\ =&\varphi(v_1, v_2) + \varphi(v_2, v_2).\end{aligned}$$

We denote the set of multilinear k-forms on V by $L^(V)$ and that of alternating k-forms by $A^k(V)$. Then $A^k \subset L^k$.

Useful observation. L^k is a vector space over F and A^k is a subspace.

Question: What's the dimension of $L^k(V), A^k(V)$?

Set $V = F^n$ with standard basis $e_1, \ldots, e_n$. We see that if $k = 1$ then $\varphi \in L^1(V) = L(V, F)$ is determined by its value on the basis through

$$\varphi\left(\sum x_i e_i\right) = \sum x_i \varphi(e_i)$$

and so $\dim L^1(V) = n$. For $k = 2$ we have that $\varphi \in L^2(V)$ is determined by the values on (pairs of) basis elements, by

$$\varphi\left(\sum x_i e_i, \sum y_i e_i\right) = \sum_{i,j} x_i y_i \varphi(e_i, e_j)$$

so that $\dim L^2(V) = n^2$. We can generalize this process to see that $\varphi \in L^k(V)$ is determined by its value on k-tuples of basis elements, and therefore $\dim L^k(V) = n^k$.

Let's now turn our attention to the alternating case.

For $k = 1$ we have that A^1 is defined by a vacuous condition, and therefore $A^1 = L^1$. For $k = 2$ we have that $\varphi \in A^2(V)$ satisfies

$$\begin{aligned}\varphi(e_i, e_i) =&0\\ \varphi(e_i, e_j) =&-\varphi(e_j, e_i)\end{aligned}$$

and therefore φ is determined by its values $\varphi(e_i, e_j)$ where $j < i$, and therefore

$$\dim A^2(V) = \frac{n(n-1)}{2}.$$

In general $\varphi \in A^k(V)$ is determined by its values $\varphi(e_{i_1}, \ldots, e_{i_k})$ such that $1 \leq i_1 < \cdots < i_n \leq n$ and therefore

$$\dim A^k(V) = \binom{n}{k}.$$

In particular, $\dim A^k(V) = 0$ for all $k > n$.

Proof of determinant theorem, part 1. $A^n(V)$ is a vector space of dimension $\binom{n}{n} = 1$. In particular, the map

$$\begin{aligned}A^n(F^n) \to&F\\ \varphi \mapsto&\varphi(e_1, \ldots, e_n)\end{aligned}$$

is a linear isomorphism. In particular there exists a unique element $D \in A^n(F^n)$ such that $D(e_1, \ldots e_n) = 1$. $\square$

We now want to find a formula for D. Let's start from the case $n = 2$. We see that

$$\begin{aligned}\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} =& D(ae_1 + ce_2, be_1 + de_2) \\ =& abD(e_1, e_1) + adD(e_1, e_2) - bcD(e_2, e_1) + cdD(e_2, e_2) \\ =& ad - bc.\end{aligned}$$

For a general formula we start by introducing some notation. First of all, let $\mathrm{Perm}(n)$ denote the set of bijections $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ (that is, the permutations on the set of n elements). We say that $\tau \in \mathrm{Perm}(n)$ is called a **transposition** if it interchanges exactly two elements of $\{1, \ldots, n\}$, e.g. $(12), (23), (13)$, but not (123).

Fact. $\mathrm{Perm}(n)$ is a group under composition and any $\sigma \in \mathrm{Perm}(n)$ can be written as a composition of transpositions, e.g. $(123) = (23) \circ (13)$.

Definition 33.5. The **sign** map is define as

$$\mathrm{sign} : \mathrm{Perm}(n) \to \{\pm 1\} \, \sigma \mapsto \begin{cases} 1 & \sigma \text{ is the product of an even number of transpositions} \\ -1 & \sigma \text{ is the product of an odd number of transpositions} \end{cases}$$

Now consider $A = (a_{ij}) \in M_n(F)$. We have that

$$\begin{aligned}\det A =& D\left(\sum_{i_1=1}^{n} a_{i_1,1} e_{i_1}, \ldots, \sum_{i_n=1}^{n} a_{i_n,n} e_{i_n} \right) \\ =& \sum_{i_1, \ldots, i_n} a_{i_1,1} \cdots a_{i_n,n} D(e_{i_1}, \ldots, e_{i_n}).\end{aligned}$$

The individual terms in the last line are 0 unless the i_j's are a permutation of $\{1, \ldots, n\}$. Therefore since in this case

$$D(e_{i_1}, \ldots, e_{i_n}) = \mathrm{sign}\, \sigma$$

where σ is the corresponding permutation, we have that

$$\det A = \sum_{\sigma \in \mathrm{Perm}(n)} \mathrm{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n}.$$

Example 33.6. For

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(F)$$

we have that

$$\det A = ad - bc.$$

For

$$A = \begin{pmatrix} a & b & c \\ i & j & k \\ x & y & z \end{pmatrix} \in M_3(F)$$

we have that

$$\det A = ajz + bkx + ciy - cjx - biz - aky.$$

Remark. The explicit formulas for vector spaces of dimension ≥ 4 are quite complex and virtually never used.

34.1. **Counting spanning trees. Problem.** You're Delta Airlines and you have a bunch of flights between a bunch of cities. You want to

- connect airports by flights;
- have any two airports connected by a path;
- do this cheaply.

This is a problem of finding spanning trees.

Definition 34.1. For a graph G, a subgraph $T \subset G$ is called a **spanning tree** if the following hold:

(i) T is a tree (connected, no loops)
(ii) T mees every vertex.

If G has finitely many vertices and edges, define $s(G)$ to be the number of spanning trees. How can we compute $s(G)$? We can use linear algebra to do so.

34.2. **Laplace matrix & the Matrix-Tree theorem.**

Definition 34.2. Let G be a finite graph and label vertices from 1 to n. The Laplace matrix of G is defined as $L = (\ell_{ij})$ where

$$\ell_{ij} = \begin{cases} -1 & i \neq j, i \text{ and } j \text{ share an edge} \\ 0 & i \neq j, i \text{ and } j \text{ don't share an edge} \\ \deg(i) & i = j \end{cases}$$

where the degree at i is the number of edges that meet at that vertex.

Theorem 34.3 (Matrix-Tree theorem)**.** *Let G be a graph with Laplace matrix L, and let L' be the matrix obtained by L by removing the last row and column. Then*$\det(G) = \det(L')$.

We don't have enough time to give a full proof of this theorem, but we can get an idea of how to do so.

Recall. Lat time we saw that

$$(\star) \qquad \det(L') = \sum_{\sigma \in \mathrm{Perm}(n-1)} \mathrm{sign}(\sigma) \ell_{1,\sigma(1)} \cdots \ell_{n,\sigma(n)}.$$

Each summand in the above expression cerresponds to a choice of an entry from each row and column. In our case, ℓ_{ij} is either $0, -1$ or a positive integer in the case $i = j$, and we can write $\ell_{ii} = 1 + \cdots + 1$ and distribute in equation $(\star)$ to get a larger sum that also computes $\det(L')$. This is called the **superexpansion**. In our case every term in the superexpansion is either a product of ± 1 (or zero). The idea is that you can associate a directed graph to each summand in the superexpansion. Then we can interpret the determinant as a weighted count of directed graphs. Some of these graphs will be spanning trees, some not. But the point will be that the total sum of all the non-spanning tree graphs will be 0. The directed graph is obtained as follows:

- if -1 is selected in row i, column j, then we draw a vertex from i to j (*with* direction);

- if the k-th 1 is selected in the diagonal entry l_{ii} we draw and edge from i to the k-th smallest neighbor (with ordering coming from the labeling). In the resulting expansion, all of the spanning trees will have sign $+1$ and all of the other ones will come in pairs with opposite signs, and will therefore cancel out.

35. 12-1 – Last Class!

35.1. **Distance geometry. Problem (from chemistry)**. We want to study a protein with a large number N of atoms. Using NMR spectroscopy one can determine the distance between atoms. Once we have this information, the question is: what is the protein shape?

35.1.1. *Math formulation.* This what is calleed a *distance geometry problem*: fix $N \geq 2$. Given distances $m_{ij} \geq 0$ for $1 \leq i, j \leq N$ such that $m_{ij} = m_{ji}$ and $m_{ij} = 0$ whenever $i = j$, the problem is:

(a) Does there exist $x_1, \ldots, x_n \in \mathbb{R}^3$ such that $||x_i - x_j|| = m_{ij}$?
(b) Find $x_1, \ldots, x_n$.
(c) Is there multiple configurations that work?

Warmup. Find $x, y, z \in \mathbb{R}^2$ such that

(a) $||x - y|| = ||y - z|| = ||z - x|| = 1$
(b) $||x - y|| = ||y - z|| = 1, ||x - z|| = 3$
(c) Find $u, v, x, y, z \in \mathbb{R}^2$ such that the distances between the pairs are $1, 1, 1, 1, \sqrt{2}, \sqrt{2}, 2, 2, \sqrt{5}, \sqrt{5}$.

For (a) we have an equilateral triangle, for (b) we have no solution (because of the triangle inequality, we would have $||x - y|| + ||y - z|| \leq ||x - z||$). For (c), multiple configurations work, such as an isosceles triangle with base 2 and height 2 with one point at the midpoint of its height.

35.1.2. *Triangle inequality & metric spaces.*

Theorem 35.1. *Let $a \geq b \geq c \geq 0$. Then there exist $x, y, z \in \mathbb{R}^2$ such that $||x - y|| = a$, $||y - z|| = b$, $||x - z|| = c$ if and only if $a \leq b + c$.*

Proof. For the forward direction, if x, y, z exist, then we just apply the triangle inequality as

$$a = ||x - y|| \leq ||x - z|| + ||z - x|| = b + c.$$

For the backward direction, we can draw x, y and the circles of radius b and c centered at x, y respectively. Then these two circles intersect since a (which is the distance between x and y) is less than or equal to $y + z$. □

Definition 35.2. For a set X, a **metric** on X is a function

$$d : X \times X \to [0, \infty)$$

such that

(i) $d(x, y) = 0$ if and only if $x = y$
(ii) $d(x, y) = d(y, x)$
(iii) $d(x, z) = d(x, y) + d(y, z)$ for all $x, y, z \in X$. The pair (X, d) is called a **metric space**. In our setting, if m_{ij} satisfies $m_{ij} = m_{ji} > 0$, $m_{ii} = 0$ and $m_{ik} = m_{ij} + m_{jk}$ then $X = \{1, \ldots, N\}$ is a metric space with $d(i, j) = m_{ij}$ our metric.

Remark. The central example is $\mathbb{R}^n$ with $d(x, y) = ||x - y||$, and this will be crucial to our study of analysis next semester.

With the tools we introduced we are now able to rephrase the distance geometry problem in the following way:

Problem: Given a metric space (X, d) does there exist a distance preserving injection $f : X \to \mathbb{R}^n$ for some n? (such a function is also known as an isometric embedding.)

Remark. The triangle inequality, though necessary, is not sufficient to get a solution. In fact, consider $X = \{1, 2, 3, 4\}$ with the possible distances given by $2, 2, 2, 2, 3, 3$. This is a metric space, but it cannot be realized as the vertices of a tetrahedron in $\mathbb{R}^3$.

Now fix $\{0, \dots, n\}$ with metric $d(i, j) = m_{ij}$ and suppose there exist $x_0, \dots, x_n$ such that $||x_i - x_j|| = m_{ij}$. We want to find a constraint on the m_{ij}. We start with an easy fact: for $u, v \in \mathbb{R}^n$ we have that

$$\langle u, v\rangle = \frac{||u||^2 + ||v||^2 - ||u - v||^2}{2}$$

(we can prove this by expanding it as in homework 9, problem 8). Consider now $a_i = x_i - x_0$. Then

$$\begin{aligned}\langle a_i, a_j\rangle =& \frac{||x_i - x_0||^2 + ||x_j - x_0||^2 - ||x_i - x_j||^2}{2}\\ =& \frac{1}{2}\left(m_{i0}^2 + m_{j0}^2 - m_{ij}^2\right).\end{aligned}$$

Consequently, if we define

$$A = \left(a_1 \middle| \cdots \middle| a_n\right)$$

and $G = (g_{ij})$ defined by

$$g_{ij} = \frac{1}{2}\left(m_{i0}^2 + m_{j0}^2 - m_{ij}^2\right).$$

we have that $G = A^T A$.

Summary: if $x_0, \dots, x_n$ exist, then the matrix g_{ij} is positive. Amazingly, we can reverse this process! In fact, given g_{ij} (defined as above), we can form the matrix G. If G is positive, then define x_i to be the ith column of A (where A is the matrix such that $G = A^T A$) with the convention $x_0 = 0$. By definition,

$$||x_i - x_0||^2 = ||x_i||^2 = \langle x_i, x_i\rangle = g_{ii} = m_{i0}^2.$$

We can also show that $||x_i - x_j|| = x_{ij}$. In fact,

$$2\langle x_i, x_j\rangle = ||x_i||^2 + ||x_j||^2 - ||x_i - x_j||^2 = m_{0i}^2 + m_{0j}^2 - ||x_i - x_j||^2$$

and on the other hand, by definition

$$2\langle x_i, x_j\rangle = 2g_{ij}$$

so that by definition of g_{ij} we have $||x_i - x_j||^2 = m_{ij}^2$.

Theorem 35.3. *Fix $n \geq 1$ and fix $X = \{0, \dots, n\}$ with metric $d(i, j) = m_{ij}$. Let $G = (g_{ij})$ such that*

$$g_{ij} = \frac{1}{2}\left(m_{i0}^2 + m_{j0} - m_{ij}^2\right).$$

Then there exist $x_0, \dots, x_n \in \mathbb{R}^n$ such that $||x_i - x_j|| = m_{ij}$ if and only if G is positive.

Index