Introduction/Summary

ERIC PETERSON, http://math.harvard.edu/~ecp/teaching/Fall2016/25a/
ecp@math.harvard.edu.

Office hours: 321h SC, T1-2pm, W11-12pm.

CAs: Thayer Anderson, Davis Lazowski, Handong Park, Rohil Prasad.

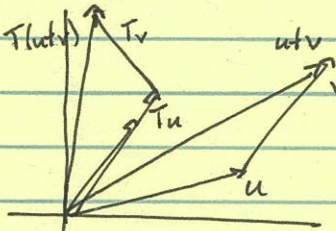
Grades: · Homeworks due Wednesday morning, before class begins, separated by CA. (25%) → LaTeX
- A midterm: 10/26, in class (25%).
- Final exam: 12/10, 9am (50%).

CAs have office hours. Weekly problem night: M8pm, Leverett.

A function is "linear" if $T(c \cdot x) = c \cdot T(x)$ and $T(x+y) = T(x) + T(y)$.

Ex: The only such f$^{\underline{ns}}$ $\mathbb{R} \longrightarrow \mathbb{R}$ are $T(x) = k \cdot x$ for some $k \in \mathbb{R}$.
↳ But there are more with other domains + codomains.

Exs: A rotation of $\mathbb{R}^2$:



Evaluation of polynomials:
$T(f) = f(1)$.

Derivatives: $\frac{d}{dx}(f(x)+g(x)) = \frac{df}{dx} + \frac{dg}{dx}$, and $\frac{d}{dx}(c \cdot f(x)) = c \cdot \frac{df}{dx}$.

Linear algebra is about studying these T's + the equations they appear in: $T(x) = y$ ⟵ how can we solve this eq$^{\underline{n}}$ for fixed y?
$T(x) = x$ ⟵ how about this, with x on both sides?
$T(U) \subseteq U$

↙ Gaussian elimination + matrix representations, basic structure of vector spaces ①

⟶ eigenvectors and eigenvalues, the Spectral Theorems + SVD, quadratic f$^{\underline{ns}}$. ②

⟶ generalized eigenvalues, Jordan normal form. ③

Main goals for this class:

Linear algebra in its own right: tangible, successful mathematics.
Linear algebra for math. apps.: calculus next semester.
Proof-writing. Manipulating def$^{\underline{ns}}$. Math as simulation + substrate.

Proof techniques I

Major goal of this class: learning to write proofs. Proofs divide into two main camps: algebraic + analytic. There are a/b.

To start, we will consider some basic, universal techniques. Mathematics is about designing models + then arguing about their behaviors. It is important to become a good + flexible debater, and to be eager to consider all points of view — mathematics is very rigid, but mathematicians are highly fallible.

Ex: A ~~number~~ ~~integer~~ whole # is divisible by 9 exactly if the sum of its decimal digits is.

First, examples: $81 = 9\cdot 9$, and $8+1 = 9 = 9\cdot 1$.

$693 = 9\cdot 77$, and $6+9+3 = 18 = 9\cdot 2$.

Meanwhile, $500 = 4\cdot 5^3$, and $5+0+0 = 5$.

convert to symbols.

not allowed to pick an example

Pf: Suppose that n is ~~an~~ a pos. integer. Its decimal expansion is

$$n = 10^k a_k + 10^{k-1} a_{k-1} + \cdots + 10\cdot a_1 + a_0 \quad \text{for } 0 \le a_j \le 10.$$

The digital sum is then

$$s(n) = a_k + a_{k-1} + \cdots + a_1 + a_0.$$

Explore!

These are very similar expressions, so — on a lark — we subtract them:

$$\Delta := n - s(n) = (10^k-1)a_k + (10^{k-1}-1)a_{k-1} + \cdots + (10-1)a_1 + (1-1)a_0.$$

This is divisible by 9, since each summand has a factor like $99\ldots99$, no matter what n was

Break into cases: the two "directions" of proof.

We then have $n = \Delta - s(n)$ and $s(n) = n - \Delta$. So if

$s(n)$ is div. by 9 or n is div. by 9

then n is div. by 9 $s(n)$ is div. by 9

because the sum/difference of two things divisible by 9 is again so. □

Thought experiment (Wason selection task):

You are told: every card has a number on one side and a color on the other. If the number is even, then the color must be red.

$\boxed{7}$ $\boxed{18}$ 🟩 🟥. How many cards do you need to check to verify the claim (underlined)?

× ✓ ✓ ×

~~Drunk driving is illegal~~. Underage drinking is illegal. If you're drunk, you must be ≥ 21.

$\boxed{28}$ $\boxed{14}$ $\boxed{\text{Drunk}}$ $\boxed{\text{Sober}}$

This is meant to illustrate the contrapositive:

logically equivalent statements {
If you are drunk, then you must be ≥21.
If you are not ≥21, then you must not be drunk.

However, sometimes one is easier than the other.

~~Ex: Show that there exist two irrational #s with a^b rational.~~

Ex: Show that if xy and $x+y$ are even, then x and y are even.

HERE IS THE CONTRAPOSITIVE [Pf: We instead show that if x and y are not both even, then xy and $x+y$ will not both be even.

Case 1: x even, y odd means $x+y$ is odd.
Case 2: x odd, y even means $x+y$ is odd.

} Formally identical, since $x+y=y+x$. This is sometimes phrased as "Assume one of the two is odd and the other even. Without loss, we may take x odd + y even."

Case 3: x and y both odd means $x \cdot y$ is odd.

Also do a direct version of this.
+ ⇒ same parity

Interstitial examples:

x	y	x+y	xy
2	1	3	2
3	4	7	12
5	7	12	35
(8	6	14	48

odd!

← even!)

Another powerful proving tool for statements indexed by $\mathbb{N}$ is induction.

Ex: $1+4+9+\cdots+n^2 = n(n+1)(2n+1)/6$ for all n.

Pf: For $n=1$, $1 = 1\cdot(1+1)(2\cdot 1+1)/6$. ✓

Assume $1+4+9+\cdots+j^2 = j(j+1)(2j+1)/6$ for some j.

Then $1+4+9+\cdots+j^2+(j+1)^2 = \frac{j(j+1)(2j+1)}{6}+(j+1)^2$

$= \frac{(j+1)}{6}\left(j(2j+1)+6(j+1)\right) = \frac{j+1}{6}\left(2j^2+7j+6\right) = \frac{(j+1)(j+2)(2j+3)}{6}$ ✓

Yihang: Fri 11am-12pm

Proof techniques II

Today we talk about two more complicated aspects of proofwriting: quantification and contradiction.

We saw quantification yesterday: when we showed that all integers n were div$^{\text{ble}}$ by 9 exactly when their digital sums are, the "all" is a quantifier. There is a second kind of quantifier, also of interest:

These are claims about general behavior.

"There is a solution x to the equation $x^2 + x = 0$."

called an existential quantifier. These are claims about examples, (Pf: Pick $x = 0$ or $x = -1$.) and they are often short.

These are interrelated:

If not all x satisfy P, then there must exist an x not satisfying P.

If ~~no~~ there does not exist an x satisfying P, then all x must not satisfy P.

Moving "not" past the quantifier changes it! This is the font of proof by counterexample: if you want to show that not all x have property P, then you need exhibit only one such x.

Constructive

Ex: Falsify the statement that for any $y \in \mathbb{R}$ there is an $x \in \mathbb{R}$ with $y = x^2$.

Pf: We need to show that there is a y which for any x, $y \neq x^2$. If we select $y = -1$, then any x has x^2 nonnegative, hence $y \neq x^2$. □

It is also possible to prove existence statements without actually exhibiting a particular value.

Ex: There exist irrational a and b with a^b rational.

nonconstructive

Pf: Consider $(\sqrt{2})^{\sqrt{2}}$. If it is rational, we are done. If it is irrational, set $a = (\sqrt{2})^{\sqrt{2}}$ and $b = \sqrt{2}$, so that

$$\left((\sqrt{2})^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{2})^{\sqrt{2}\cdot\sqrt{2}} = (\sqrt{2})^{2} = 2. \quad \Box$$

Buried in here is an idea that P is either true for x or it is false, and there is no third course. (This is different from demonstrating either of these, which is quite subtle.) This is usually summarized as saying that if P is not-false, then it is true, and conversely.

This leads to a different kind of proof technique: <u>contradiction</u>. The idea is that if some premise leads you to say that something else must be both <u>true</u> and <u>false</u>, then your premise itself must have been unsound.

<u>Ex:</u> There are infinitely many prime numbers.

<u>Pf</u>: Suppose otherwise, that there are just finitely many, named $p_1, p_2, \dots, p_k$. We then form the number $N = (p_1 \cdot p_2 \cdot \cdots \cdot p_k) + 1$, which is not divisible by p_j for any j.

This means that either N is prime (and not on the list) or that N decomposes into primes not on the list. In either case, we have shown our complete list of primes to be incomplete — a contradiction. Our initial assumption must have been wrong: it must instead be the case that there are ∞^{ly} many prime numbers. □

Functions, properties, cardinalities

We have one more foundational issue to address before we begin linear algebra in earnest.

We will avoid actually saying what a set is. Suffice it to say that it is a collection of elements for which membership can be tested, e.g., $2 \in \{n \in \mathbb{N} \mid n \text{ is even}\} \subseteq \mathbb{N}$.
$3 \notin$

A function $f : A \longrightarrow B$ is an assignment of elements of A to those of B. That is, for any element $a \in A$ there is a single corresponding element $f(a) \in B$, called its image.

Functions tend to serve two purposes: operation and transmogrification.

Ex: The operation "$+$" on $\mathbb{R}$ can be thought of as a function $+ : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$, where "$\mathbb{R} \times \mathbb{R}$" indicates the set of pairs of real numbers. You can also specialize this to get a function $s(x) = x + 1$, by setting $y = 1$. There's also $p(x) = x \cdot x$, which comes from specializing $\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ to $x = y$.

Ex: The function $\cos : \mathbb{R} \longrightarrow \mathbb{R}$ plays more of the second role: it takes in an angle value (thought of as a real number) and gives out a ratio of lengths (thought of as a real number).

These functions have certain properties which tell you interesting information about them.

- Injectivity: A function is injective if no two inputs give the same output. For transmogrifications, this is a losslessness, that you can recover the input uniquely from the output. ($\cos$ is not injective, because $\cos(0) = \cos(2\pi)$.) For operations, this is about solutions: s is injective, so $x + 1 = s(x) = y$ has a single solution. p is not, so $x^2 = p(x) = y$ may have many.

- <u>Surjectivity</u>: This is the statement that every output has at least one input realizing it. For transmogrifications, this is a kind of <u>efficiency</u>: there's no "wasted space" in the codomain of impossible values. For operations, this is again about solutions: $x+1=s(x)=y$ <u>is</u> surjective, so it is always possible to solve this eq^{n} for x, no matter what y is. ~~p~~ $x^2=p(x)=y$ is <u>not</u>, so there are y with no solution in x.
- <u>Bijectivity</u>: Simultaneously injective and surjective. These are "perfect dictionary" transmogrifications, or equations with exactly 1 solution for any choice of y.

Every function can be broken into these parts in the following way:

$$\begin{tikzcd} \mathbb{Z} \arrow[r, "f", "⓪"'] \arrow[d, two heads, "②"'] \arrow[dr, "①"] & \mathbb{Z} \\ \{\{n,-n\} \mid n\in\mathbb{Z}\} \arrow[r, "\cong", "③"'] & \mathbb{Z}_{\geq 0} \arrow[u, hook, "①"'] \end{tikzcd}$$

$\mathbb{R} \xrightarrow{f} \{y\in\mathbb{R} \mid y\geq 0\}$

① If the function is not surjective, then we can restrict its codomain to just the values it does take on. In turn, this subset injects into the original codomain.

② If the function is not injective, we can collect together all the elements of the domain that give the same value into different subsets. These sets are said to <u>partition</u> the domain, meaning they do not overlap + yet their union is all of the domain. There is a surjective map assigning each element to the subset it belongs to.

③ Finally the original function defines a new function as at the bottom: given a subset, the f^{n} takes on the same value on any of its members, so gives an element of the restricted codomain. This map is surjective and injective, hence bijective.

This is a good rep^{n} of what functions "do". They forget a little information, ← eg, the sign then they represent what's left inside of the codomain according to some rule.

Ex: ~~[illegible]~~ For cos, ② tells you it's 2π-periodic + ③ tells you it lies in $[-1,1]$.

First, a stray example from last time:

$$\begin{tikzcd} \mathbb{R} \arrow[r, "\tan"] \arrow[d, shift left, "j"] \arrow[dr, "q"] & \mathbb{R} \\ \{\{\theta+\pi k \mid k\in\mathbb{Z}, \theta\neq\frac{\pi}{2}\}\} \arrow[u, shift left] \arrow[r, "r"'] & \mathbb{R} \arrow[u, "i"'] \end{tikzcd}$$

$$\begin{tikzcd} \mathbb{R} \arrow[r, "\cos"] \arrow[d, "j"'] \arrow[dr, "q"] & \mathbb{R} \\ \{\{\theta+2\pi k, -\theta+2\pi k \mid k\in\mathbb{Z}\}\} \arrow[r, "r"] & {[-1,1]} \arrow[u, "i"'] \end{tikzcd}$$

Vector Spaces

Remember that we are interested in functions $T: V \longrightarrow W$ satisfying equations like $T(k\cdot u) = k\cdot T(u)$ and $T(u+v) = T(u)+T(v)$, where V and W are fancy co/domains. We need to make sense of "+" and "·" inside of V and W.

Ex: Represent a point in the plane by a coordinate pair (x,y). Then rotation by 90° is specified by $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}$.

same objects

Defining + and · componentwise $\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} x+w \\ y+z \end{pmatrix}$, $k\cdot\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} kx \\ ky \end{pmatrix}$,

we find that T is a linear map. However, you can see that the op^ns + and · are kind of complicated! different!

Def: A field k is a set with +, −, ·, and / ⟵ defined on nonzero elements satisfying comm., assoc., + distributivity.

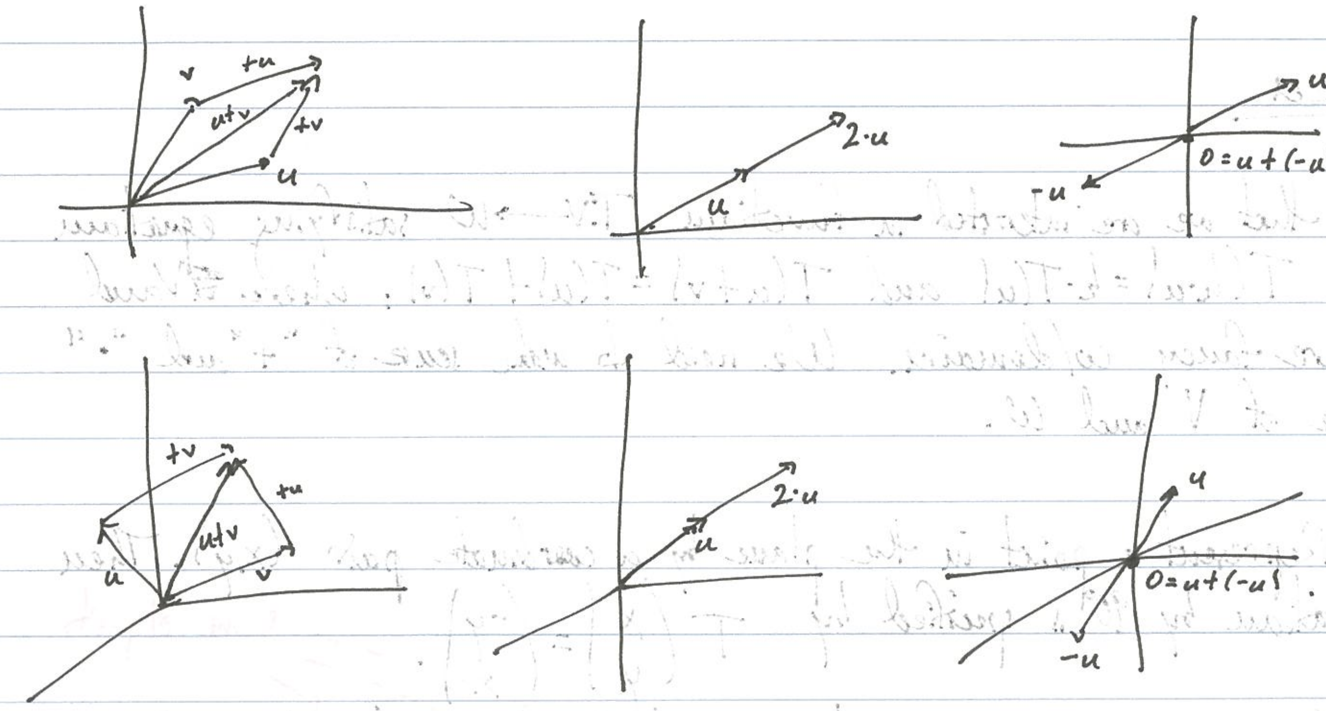
Ex: $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/p$, $\mathbb{Q}$. Non-ex: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}/4$.

Def: A vector space V (over k) is a set with $+, -: V\times V \longrightarrow V$ and $\cdot: k\times V \longrightarrow V$,

also satisfying comm., assoc., + distributivity.

<u>Ex:</u> $\mathbb{R}^n$ and $\mathbb{C}^n$. $\mathbb{R}^\infty$. Polynomials. Polynomials of degree n. Polynomials vanishing at 0. Polynomials vanishing at 1. Functions $[0,1] \longrightarrow \mathbb{R}$. Functions valued in a vector space. $\mathbb{C}$ over $\mathbb{R}$ and $\mathbb{R}$ over $\mathbb{Q}$.

Pictoral rep$^{\underline{n}}$ of vector arithmetic in $\mathbb{R}^2$ and $\mathbb{R}^3$.



Subspaces (1.C)

Implicit in our discussion thus far has been a notion of a subset:

a subset $Y \subseteq X$ is a set s.t. each element $y \in Y$ is already also an elt. $y \in X$.

This is a statement about size: X is at least as large as Y.

They're often described by properties: $\{n \in \mathbb{N} \mid n \text{ is div. by } 2\} \subseteq \mathbb{N}$, the subset of even natural numbers.

There's a corresponding notion for vector spaces: $U \subseteq V$ is a subspace of V if U is a subset and itself a vector space with the same op$^{\underline{ns}}$ as on V.

Ex: $\{(x,y,z) \in \mathbb{R}^3 \mid x = y\} \subseteq \mathbb{R}^3$.

Polynomials vanishing at $0 \subseteq$ all polynomials.

Non-ex: $\{(x,y,z) \in \mathbb{R}^3 \mid x = 5\} \subseteq \mathbb{R}^3$ is a subset but not a subspace.

This is because $(5,0,0)$ and $(5,10,32)$ are elements of the subset, but $(5,0,0)+(5,10,32) = (\underline{\underline{10}},10,32)$ is not.

Sets have various interesting op$^{\underline{ns}}$ on them, like intersection, union, + complement. These have analogues in vector spaces, but their behavior is more complex.

Intersection: The intersection of 2 subspaces is a subspace.

Union: The union of 2 subspaces U_1, U_2 is a subspace iff one contains the other. (This is homework.) This has a replacement though: the sum is $U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\}$.

Lem: This is the smallest subspace containing U_1 and U_2.

Pf:

Pf: It is a subspace: $(u_1+u_2)+(u_1'+u_2') = \overbrace{(u_1+u_1')}^{\in\mathcal{U}_1} + \overbrace{(u_2+u_2')}^{\in\mathcal{U}_2}$, and

: $k\cdot(u_1+u_2) = ku_1 + ku_2$. "Smallest" means that any other subspace $\mathcal{W}$ with $\mathcal{U}_1, \mathcal{U}_2 \subseteq \mathcal{W}$ has $\mathcal{U}_1+\mathcal{U}_2 \subseteq \mathcal{W}$. This is clear too: for $u_1+u_2 \in \mathcal{U}_1+\mathcal{U}_2$, $u_1 \in \mathcal{U}_1$ and $u_2 \in \mathcal{U}_2$ and hence $u_1, u_2 \in \mathcal{W}$. ~~an~~ Then $u_1+u_2 \in \mathcal{W}$ b/c $\mathcal{W}$ is a subspace, and $\mathcal{U}_1+\mathcal{U}_2 \subseteq \mathcal{W}$. □

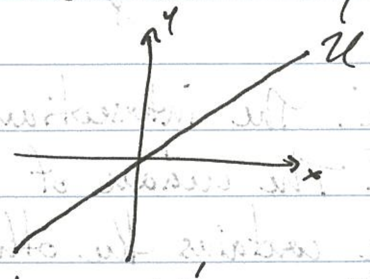
Direct sum: A particularly nice kind of sum of subspaces is when $\mathcal{U}_1 \cap \mathcal{U}_2 = 0$. In this case, any $v \in \mathcal{U}_1+\mathcal{U}_2$ has a unique representation as $v = u_1+u_2$.

Pf: If $v = u_1+u_2$ and $v = u_1'+u_2'$, then $v - v = u_1+u_2-u_1'-u_2' = 0$, and $\underbrace{u_1-u_1'}_{\in\mathcal{U}_1,\neq 0.} = \underbrace{u_2-u_2'}_{\in\mathcal{U}_2,\neq 0.}$. This violates the intersection condition. □

This is a kind of "disjoint union" condition: $\mathcal{U}_1$ + $\mathcal{U}_2$ have no overlap.

Complementation: For $A \subseteq X$ a subset, there is another unique subset $X \setminus A$ such that A + $X \setminus A$ are disjoint and $A \cup (X\setminus A) = X$.

This is kind of true for vectorspaces — what fails is unicity. Proving ~~the~~ existence in generality is more trouble than it's worth — you need the Axiom of Choice. Instead, let's look at how unicity fails.



Ex: $\mathcal{U} := \{(x,y)\in\mathbb{R}^2 \mid x=y\} \subseteq \mathbb{R}^2$.

One complement: $\mathcal{W} = \{(x,y)\in\mathbb{R}^2 \mid x=-y\}$.

Pf: If $x=y$ and $x=-y$, then $x=y=0$, so $\mathcal{U}\cap\mathcal{W}=0$. Given $(s,t)\in\mathbb{R}^2$, we solve $x+\overset{x'}{\cancel{y}}=s$, $x-\overset{x'}{\cancel{y}}=t$ to get $x=\frac{s+t}{2}$, $x'=\frac{s-t}{2}$.

Ex': $\mathcal{W} = \{(\overset{x}{\cancel{y}}, y)\in\mathbb{R}^2 \mid x=0\}$. Pf: Again, $\mathcal{U}\cap\mathcal{W}=0$. For (s,t), we find

$$\begin{pmatrix} s \\ t \end{pmatrix} = \underset{\mathcal{U}}{\underset{\uparrow}{\begin{pmatrix} s \\ s \end{pmatrix}}} + \underset{\mathcal{W}}{\underset{\uparrow}{\begin{pmatrix} 0 \\ t-s \end{pmatrix}}}.$$

Finite-dimensional vector spaces (2.*)

Last time we skirted around the existence of complements of vector subspaces. Here's a naive approach to constructing a complement which will be of interest to us:

Procedure: ⓪ Start with a subspace $\mathcal{U} \subseteq V$ and $W = 0_V$.

① If $\mathcal{U} + W \neq V$, there is some missing vector $v \in V \smallsetminus (\mathcal{U} + W)$.

② Replace W by $W + \langle v \rangle$, where $\langle v \rangle = \{ cv \mid c \in k \}$ is the smallest subspace containing v.

③ Go back to ①.

→ If not, then we're done: $\mathcal{U} \cap W = 0$ and $\mathcal{U} + W = V$.

WARNING: THIS MAY NOT TERMINATE IF V IS "TOO LARGE"!!

The subspace W we construct has a very particular form:

$$W = \langle v_1 \rangle + \langle v_2 \rangle + \cdots + \langle v_n \rangle = \{ c_1 v_1 + \cdots + c_n v_n \mid c_j \in k \},$$

where v_j is the vector picked on the j^{th} time through the loop.

Def: W is called the span of $(v_1, \ldots, v_n)$. A particular element $w = c_1 v_1 + \cdots + c_n v_n$ is called a linear combination of $(v_1, \ldots, v_n)$.

There is an interesting edge case of this algorithm: if $\mathcal{U} = 0$, then its complement should be all of V. However, the algorithm presents V in a special form: $V = \langle v_1 \rangle + \langle v_2 \rangle + \cdots + \langle v_n \rangle$. If the algorithm terminates, V is called finite dimensional (of dimension n).

Ex: $\mathbb{R}^2 \cong \langle (1,0) \rangle + \langle (0,1) \rangle$.

{polynomials} is not finite dimensional. $(1, x, x^2, x^3, \ldots)$

{polynomials of degree $\leq n$} $= \langle 1 \rangle + \langle x \rangle + \cdots + \langle x^n \rangle$ has dimension $(n+1)$.

There is a flaw in using this algorithm as a definition: it is non-deterministic, meaning that it may behave differently based on what v_j is chosen at each step. This is worrying: does it sometimes terminate & sometimes not? Is the concluding number n always the same? We will have to work for a while to see this.

Lem: A linear dependence is a nonzero linear combination $c_1w_1 + \cdots + c_dw_d = 0$. Suppose c_j is nonzero. Then $\underset{A}{\operatorname{span}(w_1, \dots, w_d)} = \underset{B}{\operatorname{span}(w_1, \dots, w_{j-1}, w_{j+1}, \dots w_d)}$, with w_j removed.

Pf: Automatically, $B \subseteq A$. To see $A \subseteq B$, note that

$$w_j = \frac{1}{c_j}\left(-c_1w_1 - \cdots - c_{j-1}w_{j-1} - c_{j+1}w_{j+1} - \cdots - c_dw_d\right),$$

which lets us write any element $a \in A$ as

$$\begin{aligned} a &= k_1w_1 + \cdots + k_dw_d \\ &= k_1w_1 + \cdots + k_{j-1}w_{j-1} + k_j(\text{——}\,//\,\text{——}) + k_{j+1}w_{j+1} + \cdots + k_dw_d. \end{aligned}$$

This does not involve w_j. □

Rem: Being linearly independent is the same as $\langle w_1\rangle + \cdots \langle w_d\rangle$ being a direct sum.

Car: The length of any linearly independent list $\leq$ ← v's.
the length of any spanning list. ← w's

Pf: ① Start with $(w_1, \dots, w_d)$ and $(v_1, \dots, v_n)$.

② Prepend the first v-vector to the w list.

③ There is a dependence in the w-list, not involving the v's. Use this to eliminate a w-vector. (↑ any finite set of them is still lin. ind.)

④ Repeat.

Eventually you'll run out of v's, before you run out of w's. That means $n \leq d$. □

Car: The algorithm gives the same n no matter what.

Pf: If you have $v_1, \dots, v_n$ and $v_1', \dots, v_{n'}'$, then $n \leq n'$ and $n' \leq n$. □

More on the dimension algorithm: (2.*)

We can squeeze some more out of the ideas from last time:

Cor.: If $\mathcal{U} \subseteq V$ is a subspace + V is finite dim'l, then so is $\mathcal{U}$.

Pf: Run the algorithm on $\mathcal{U}$ and on V. The list resulting from $\mathcal{U}$ is linearly independent (in V) and the list from V spans V. The Lemma from last time says length span $\geq$ length l.i. □

(In fact, $\dim \mathcal{U} \leq \dim V$.)

I've been obtuse and avoided giving you some useful vocabulary: a basis for V is a set that is both linearly independent and spans V. (The list resulting from the algorithm is a basis for the complementary subspace.)

Reinforcing ex: $\{(1,0),(0,1)\}$ is a basis for $\mathbb{R}^2$. So is $\{(3,5),(2,1)\}$. However, none of $\{(1,1)\}$, $\{(1,0),(0,1),(1,1)\}$, and $\{(1,0),(2,0)\}$ are bases.

doesn't span

linearly dependent

Lem: Any spanning list can be shortened to a basis.

Pf: ① Start with $j = 1$.

② If v_j is in span$\{v_1, \dots, v_{j-1}\}$, then discard v_j. Otherwise keep it and continue to the next j.

At the end, the rest of the list will still span V. It's now linearly independent: if there were a dependence, then there would be a last nonzero coefficient in the dependence. That would violate step ② at that stage. □

~~Cor:~~

Lem: Every l.i. list of vectors in a finite dim$^{\underline{l}}$ V extends to a basis of V.

Pf: Take $\mathcal{U} = \operatorname{span}\{u_1, \dots, u_d\}$ to be the span of the l.i. list. Use the complementary algorithm to find $\mathcal{W} = \langle v_1 \rangle + \cdots + \langle v_n \rangle$. Then $V = \mathcal{U} + \mathcal{W} = \langle u_1 \rangle + \cdots + \langle u_d \rangle + \langle v_1 \rangle + \cdots + \langle v_n \rangle$, and this is a direct sum because the lists are l.i. and $\mathcal{U} \cap \mathcal{W} = 0$. □

~~Rem:~~

Lem: If $\{v_1, \dots, v_n\}$ is l.i. and $\dim V = n$, then $\{v_*\}$ is a basis.

Pf: Extend it to a basis — but it's already length n! So no new vectors are added. □

Lem: If $\{w_1, \dots, w_d\}$ is spanning and $\dim V = d$, then $\{w_*\}$ is a basis.

Pf: It can be reduced to a basis — but it's already length d! So, no vectors can be erased. □

Axler 2.43

Lem: For $\mathcal{U}_1, \mathcal{U}_2 \subseteq \mathcal{U}$, we have $\dim(\mathcal{U}_1 + \mathcal{U}_2) = \dim \mathcal{U}_1 + \dim \mathcal{U}_2 - \dim(\mathcal{U}_1 \cap \mathcal{U}_2)$

Pf: Use the algorithm to build a $\overset{\{u_*\}}{\text{basis}}$ for $\mathcal{U}_1 \cap \mathcal{U}_2$. Consider it as a l.i. set in $\mathcal{U}_1$ and $\mathcal{U}_2$ separately, and extend it to a basis $\{u_1, \dots, u_d, v_1, \dots, v_n\}$ of $\mathcal{U}_1$ and $\{u_1, \dots, u_d, w_1, \dots, w_m\}$ of $\mathcal{U}_2$. We claim that the combined list $\{\underbrace{u_1, \dots, u_d}_{\mathcal{U}_1 \cap \mathcal{U}_2}, \underbrace{v_1, \dots, v_n}_{\text{new to } \mathcal{U}_1}, \underbrace{w_1, \dots, w_m}_{\text{new to } \mathcal{U}_2}\}$ is a basis for $\mathcal{U}_1 + \mathcal{U}_2$.

It clearly spans. Suppose there were a linear dependence:

$$a_1 u_1 + \cdots + a_d u_d + b_1 v_1 + \cdots + b_n v_n + c_1 w_1 + \cdots + c_m w_m = 0.$$

$$-(\underbrace{a_1 u_1 + \cdots + a_d u_d}_{\in \mathcal{U}_1 \cap \mathcal{U}_2} + \underbrace{b_1 v_1 + \cdots + b_n v_n}_{\in \mathcal{U}_1}) = \underbrace{c_1 w_1 + \cdots + c_m w_m}_{\in \mathcal{U}_2 \leftarrow \text{actually } \mathcal{U}_1 \cap \mathcal{U}_2.}.$$

Hence $c_1 w_1 + \cdots + c_m w_m = d_1 u_1 + \cdots + d_d u_d$. Substituting this back:

$a_1' u_1 + \cdots + a_d' u_d + b_1 v_1 + \cdots + b_n v_n = 0$, but this list is l.i.. □

~~Linear maps + Factorization~~

Linear maps + Kernels (3.A-B)

Finally, we turn our attention to how vector spaces relate to one another through linear maps. One more time:

Def$^{\underline{n}}$: A f$^{\underline{n}}$ $T: V \longrightarrow W$ is linear (T, W vector spaces) when $T(v_1+v_2) = T(v_1) + T(v_2)$ and $T(k \cdot v) = k \cdot T(v)$.

Ex: ① $T: \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by $T(x,y) = x-y$, or
$T': \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by $T(x,y) = y$.
② $T: \{\text{polynomials}\} \longrightarrow \mathbb{R}$ given by $T(f) = f(0)$, or
$T': \{\text{polynomials}\} \longrightarrow \mathbb{R}$ given by $T(f) = f(1)$.

The basic op$^{\underline{ns}}$ on linear functions are:
① Addition: given $T_1, T_2 : V \longrightarrow W$, we can form $(T_1+T_2)(v) = T_1(v) + T_2(v)$, which is also linear.
② Scaling: given $T: V \to W$ and $k \in K$, we can form $(k \cdot T)(v) = k \cdot (T(v))$, which is also linear.
③ Composition: given $V \xrightarrow{T} W \xrightarrow{T'} W'$, we can compose $(T' \circ T)(v) = T'(T(v)) \in W'$ to get a linear map.
These play nicely with each other. For instance, $\circ$ distributes over $+$.

There ~~is~~ are also ~~a~~ 2 natural subspaces associated to T:
Def: The kernel of $T: V \to W$ is $\{v \in V \mid T(v) = 0\} \subseteq V$.
It is a subspace. The image of T is $\{w \in W \mid \exists v \in V \text{ with } Tv = w\}$.
It, too, is a subspace.

dim ker	dim V	dim im
$1 \hookrightarrow$	$2 \longrightarrow$	1
line $\hookrightarrow$	$\mathbb{R}^2 \longrightarrow$	line
line $\hookrightarrow$	$\mathbb{R}^2 \longrightarrow$	line

Ex: ① $\ker T = \{(x,y) \in \mathbb{R}^2 \mid x - y = 0\} \subseteq \mathbb{R}^2$,
$\ker T' = \{(x,y) \in \mathbb{R}^2 \mid y = 0\} \subseteq \mathbb{R}^2$.

② $\ker T = \{f \text{ a poly}^{\underline{l}} \mid f(0) = 0\} \subseteq \{\text{all polynomials}\}$,
$\ker T' = \{f \text{ a poly}^{\underline{l}} \mid f(1) = 0\} \subseteq \{\text{all polynomials}\}$.

These are all subspaces we've thought about before! This is interesting: what exactly is the relationship between $f^{\underline{ns}}$ $T : V \to W$ and subspaces $\mathcal{U}$? Can we get all $\mathcal{U}$? How many Ts give the same $\mathcal{U}$? What does W have to do with it? (Consider $W = 0$.)

~~For the moment, we're going to think about something we've seen before: factorizations. The image subspace plays the role of "Step 1":~~

$$\begin{tikzcd}
\ker T \arrow[d, hook] & & \\
V \arrow[rr, "T"] \arrow[rrd] & & W \\
& & \operatorname{im} T \arrow[u, phantom, "\cup"]
\end{tikzcd}$$

These are all interesting questions. For the moment, we're going to produce a relation between $\ker T$ and $\operatorname{im} T$:

Lem: $\dim(\ker T) + \dim(\operatorname{im} T) = \dim V$, for f.d. V.

Pf: Extend a basis of $\ker T$ to one of V. The image of the extension in $\operatorname{im} T$ is a basis there. □

This is interesting: the proof says that $\operatorname{im} V$ has the same dimension as a complement of $\ker T$ — but $V \to \operatorname{im} T \hookrightarrow W$ is canonical, whereas a choice of complement $(\ker T)^c \leq V$ is not unique. We will think about this next time in the context of factorizations.

Factorizations for Linear Maps:

Last time we talked about subspaces associated to

$$\begin{tikzcd}
\ker T \arrow[d, phantom, "\cap|" description] & & \\
V \arrow[rr, "T"] \arrow[drr, "\textcircled{1}"'] & & W \arrow[d, phantom, "\cup|" description] \\
 & & \operatorname{im} T
\end{tikzcd}$$

We basically did Step ① by factoring T through $\operatorname{im} T$, which injects into W.

~~We're missing an analogue of Step ②: a way to build a surjection with kernel ker T.~~

To start, this picture suggests an interesting lemma:

Lem: A map $T: V \longrightarrow W$ is injective if and only if $\ker T = 0$.

Pf: If ~~ker~~ T is injective, then $\{v \in V \mid T(v) = 0\} = 0$. If T is not injective, then there are $v_1 \neq v_2$ with $T(v_1) = T(v_2)$. But then $T(v_1 - v_2) = T(v_1) - T(v_2) = 0$ exhibits $v_1 - v_2 \in \ker T$. □

To complete the picture, we're missing an analogue of Step ②: a way to build a surjection with kernel the subspace $\ker T \leq V$. This actually looks a lot like what we did for sets:

Def^n: Given $\mathcal{U} \leq V$, we define $V/\mathcal{U}$ by
$V/\mathcal{U} = \{v + \mathcal{U} \mid v \in V\}$, a collection of subsets of V.

Lem: There is a map $f: V \longrightarrow V/\mathcal{U}$ given by $f(v) = v + \mathcal{U}$ which is surjective with kernel $\mathcal{U}$. □

This construction fills in the 2nd step:

3.91 in Axler.

$$\begin{tikzcd}
\ker T \arrow[d, phantom, sloped, "\subseteq"] & & \\
V \arrow[rr, "T"] \arrow[d, "②"'] \arrow[drr, "①"] & & W \arrow[d, phantom, sloped, "\supseteq"] \\
V/\ker T \arrow[rr, "\simeq", "③"'] & & \operatorname{im} T
\end{tikzcd}$$

; and like last time we can fill in the linear map ③ with a bijection.

Let's think about the members of V/U some more.

<u>Rem</u>: U itself is one member, since $0+U=U$.

<u>Rem</u>: The other members of V/U look like <u>translates</u> of U off of the origin. We know these are not subspaces, but they are useful enough to earn a name: they are <u>affine subspaces</u> (or <u>translates</u>).

These show up when considering the soln sets to equations like $T(v) = ~~a~~\omega$.

<u>Lem</u>: The sol^n set ~~to~~ $\{v \in V \mid T(v) = \omega\}$ is empty or a translate of $\ker T$.

Pf: <u>If</u> the sol^n set is empty, we are done. <u>If</u> it's nonempty, pick a $v \in V$ with $T(v) = \omega$. Then $v + \ker T$ is exactly the sol^n set:

① For $k \in \ker T$, $T(v+k) = T(v) + T(k) = ~~\mathbb{F}~~\, \omega + 0 = \omega$ gives another sol^n.

② For another sol^n v', $T(v-v') = T(v) - T(v') = \omega - \omega = 0$, so $v - v' \in \ker T$. ☐

<u>Ex</u>:

$$\begin{tikzcd}
y\text{-axis} \arrow[d, phantom, sloped, "\subseteq"] & \\
\mathbb{R}^2 \arrow[r, "\operatorname{proj}_x"] \arrow[d] & \mathbb{R}^2 \arrow[d, phantom, sloped, "\supseteq"] \\
\text{space of vert. lines} \arrow[r, "x\text{-coord}"] & \mathbb{R} = \{(x,0) \mid x \in \mathbb{R}\}.
\end{tikzcd}$$

<u>Degeneracy</u>: If $\ker T = 0$, then $V \to V/\ker T$ is already bijective. If T is surjective, then $\operatorname{im} T = W$. Now take $\dim V = \dim W < \infty$.

- ~~If V is f.d., then~~ T surj $\Rightarrow \dim \ker T = 0 \Rightarrow T$ inj.
- ~~If V is f.d. then~~ T inj $\Rightarrow$ a basis of $\operatorname{im} T$ gives a l.i. set in W of size $= \dim W$.

Bases as presentations

A few times in this class we've drawn some picture like

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "j"'] \arrow[r, bend right, "i \circ \tilde{f} \circ j"'] & W \\ V/\ker(f) \arrow[r, "\tilde{f}"] & \operatorname{im} f \arrow[u, "i"'] \end{tikzcd}$$

to communicate the identity $f = i \circ \tilde{f} \circ j$.

These pictures are called diagrams, their nodes are labeled by vector spaces, their arrows by linear maps, and they encode how different paths with the same start + end are the same.

A useful puzzle piece when drawing these pictures is the isomorphism, which is a bijective (or invertible) linear map.

These look like
$$\begin{tikzcd} V \arrow[loop left, "f^{-1} \circ f"] \arrow[r, bend left, "f"] & W \arrow[l, bend left, "f^{-1}"] \arrow[loop right, "f \circ f^{-1}"] \end{tikzcd}.$$

Note that going $V \xrightarrow{f} W \xrightarrow{f^{-1}} V$ is the same as staying stationary at V — i.e., the identity map $1 : V \longrightarrow V$.

Lem: A basis S of length n for a vector space V gives an iso^{m} $K^n \xrightarrow{\varphi} V$, and conversely.

Pf: Given $S = (v_1, \ldots, v_n)$, we define $\varphi(k_1, \ldots, k_n) = k_1 v_1 + \cdots + k_n v_n$. This is surjective because S spans, and it's injective because S is linearly independent. If we're instead given φ, we set $v_j = \varphi\underbrace{(0, \ldots, 0, 1, 0, \ldots, 0)}_{e_j}$, in the j^{th} position. Again, surjectivity gives spanning + injectivity gives linear independence. □

This means that K^n are the "standard" vector spaces, — and because these are such concrete spaces, we can say a lot about them.

~~Lem: Linear maps $V \to W$ are determined by their rows columns~~

~~Lem: Linear maps $K^n \to K^m$ are encoded by $m \times n$ matrices.~~

~~Pf~~

Lem: Linear maps $K^n \xrightarrow{f} K^m$ are encoded by $m \times n$ matrices.

Pf: Any vector $v = (k_1, \ldots, k_n) \in K^n$ can be decomposed as $v = k_1 e_1 + \cdots + k_n e_n$, where e_j is as before. So, we only need to evaluate $f(e_j)$, which itself has a decomposition $f(e_j) = a_{1j} \cdot f_1 + a_{2j} f_2 + \cdots + a_{mj} f_m$, f_j the jth std basis vector in K^m. Arranging these numbers into a grid (a_{ij}), we uncover a matrix. ~~Conversely,~~ Linear indep. of (e_j) shows that ~~the~~ a matrix specifies a function, which is checked to be linear. Linear indep. of (f_j) shows that no two matrices rep^t the same map. □

Thm: Under a choice of basis on the domain and codomain, linear maps and matrices correspond.

Pf:

$$\begin{tikzcd}
V \arrow[r, "f"] \arrow[d, shift left] \arrow[dr, phantom, "\Downarrow"] & W \arrow[d, shift left] \\
K^n \arrow[u, shift left] \arrow[r, "\text{matrix}"] & K^m \arrow[u, shift left]
\end{tikzcd}
,
\begin{tikzcd}
V \arrow[r, bend right, "\text{linear map.}"'] \arrow[d, shift left] & W \arrow[d, shift left] \\
K^n \arrow[u, shift left] \arrow[r] \arrow[ur, phantom, "\Uparrow"] & K^m \arrow[u, shift left]
\end{tikzcd}
. \quad \square$$

Lem: Matrix multiplication encodes function composition.

Pf: Represent f by (a_{ij}) and g by (b_{kl}). Then

$$f(g(e_l)) = f\left(\sum_{k=1}^{m} b_{kl} f_k\right) = \sum_{k=1}^{m} b_{kl} f(f_k) = \sum_{k=1}^{m} b_{kl} \cdot \left(\sum_{i=1}^{m'} a_{ik} g_i\right)$$

$$= \sum_{i=1}^{m'} \underbrace{\left(\sum_{k=1}^{m} a_{ik} b_{kl}\right)} \cdot g_i \quad = (c_{il}). \ \square$$

Duality (3.F)

One of the consequences of last time is:

Cor: $\dim \mathcal{L}(V,W) = \dim V \cdot \dim W$.

"the dual"

In particular, $\dim \mathcal{L}(V,K) \cong \dim V$, yet $V^* := \mathcal{L}(V,K)$ has interesting properties not exactly like those of V. The most basic such property is that it is backwards or contravariant.

Def: A map $f: V \to W$ induces a map $f^*: W^* \to V^*$ defined by $(f^*\varphi)(v) = \varphi(f(v))$ (or precomposition).

Ex: There is an iso$^{\underline{m}}$ $\mathcal{L}(K^n, K) \cong K^n$ by $\varphi \mapsto (\varphi(e_j))_j$. The induced matrix is $(a_{ij})^* = (a_{ji})$, called the transpose. You might enjoy checking the identity $(AB)^* = B^*A^*$.

We have two tasks to take care of today.

I. Pairings: A map $V \times W \xrightarrow{\langle -,- \rangle} K$ is called a pairing, and the pairing is moreover perfect if $\forall\, w \overset{\neq 0}{\in} W\ \exists v \in V$ with $\langle v, w \rangle \neq 0$.

Lem: A perfect pairing determines an injection $W \xhookrightarrow{\tau} V^*$.

Pf: This is just "currying". $\tau(w)(v) = \langle v, w \rangle$, and the perfection condition shows that $\tau(w) \neq 0$ so that τ is injective. □

Rem: There is a "natural" iso$^{\underline{m}}$ $V \cong V^{**}$. There is an evaluation perfect pairing $V \times V^* \to K$, and flipping this around gives an injection $V \hookrightarrow (V^*)^*$. Since these are equidimensional, it's an isomorphism. (In the ∞-dimensional case, we at least get an injection.)

II. Subspaces associated to dual maps

Continuing our obsession with subspaces, it would be nice to understand $\ker(f^*)$ and $\operatorname{im}(f^*)$ in terms of f.

Toward this, we make the following interrelating def$^{\underline{n}}$:

Def: Given $\mathcal{U} \subseteq V$, we define the annihilator $\mathcal{U}^\circ \subseteq V^*$ by $\{\varphi \in V \mid \mathcal{U} \subseteq \ker(\varphi)$, or $\varphi(\mathcal{U}) = 0\}$. This is a subspace.

Lem: $\dim \mathcal{U} + \dim \mathcal{U}^\circ = \dim V$.

Pf: Consider $i : \mathcal{U} \hookrightarrow V$ and its dual $i^* : V^* \longrightarrow \mathcal{U}^*$.

We have
$$\dim V^* = \dim \ker(i^*) + \dim \operatorname{im}(i^*)$$
$$\dim V^* = \dim \mathcal{U}^\circ + \dim \mathcal{U}^*$$
$$\dim V = \dim \mathcal{U}^\circ + \dim \mathcal{U}. \quad \square$$

The annihilator also give the desired relations between f + f^*:

Lem: $\ker(f^*) = (\operatorname{im} f)^\circ$, and $\operatorname{im}(f^*) = (\ker f)^\circ$.

Pf: The first equality is a matter of definitions. In the second case, only $\operatorname{im}(f^*) \subseteq (\ker f)^\circ$ is obvious from the def$^{\underline{n}}$. However, $\dim \operatorname{im} f^* = \dim W^* - \dim \ker f^*$
$= \dim W - \dim (\operatorname{im} f)^\circ = \dim \operatorname{im} f,$
$= \dim V - \dim \ker f = \dim (\ker f)^\circ$. So, $\operatorname{im}(f^*)$ is a top-dim$^{\underline{l}}$ subspace of $(\ker f)^\circ$ and hence they're equal. $\square$

Cor: $\dim \operatorname{im} f^* = \dim W^* - \dim \ker f^*$
$= \dim W^* - \dim (\operatorname{im} f)^\circ$
$= \dim \operatorname{im} f. \quad \square$

Polynomials over $\mathbb{R}$ and $\mathbb{C}$ (Ch. 4)

Soon, we will move on to the second major goal of this course: understanding eq.^{ns} of the form $f(v) = k \cdot v$ for f a linear f^{n} $f: V \longrightarrow V$ from a vector space to itself. We will find out that analysis of this situation ~~go~~ involve polynomials, which have particularly nice properties $/\mathbb{C}$ + $/\mathbb{R}$.

Lem: The only zero function is the zero polynomial (over $\mathbb{R}$ or $\mathbb{C}$).

Pf: Suppose $f(x)$ is a nonzero polynomial $f(x) = a_m x^m + \cdots + a_1 x + a_0$.
May as well take $a_m = 1$, and set $z = |a_0| + |a_1| + \cdots + |a_{m-1}| + 1$.
We must have $z > 1$, so $z^{<m-1} \leq z^{m-1}$, so

$$|a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}| \leq |a_0| + |a_1| z + \cdots + |a_{m-1}| z^{m-1}$$
$$= (|a_0| + |a_1| + \cdots + |a_{m-1}|) z^{m-1}$$
$$< (|a_0| + |a_1| + \cdots + |a_{m-1}| + 1) z^{m-1} = z^m.$$

Hence, $(a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}) + z^m \neq 0$. □

Lem: For $p, s \in P$ ~~$\mathbb{R}$ or $\mathbb{C}$~~ with $s \neq 0$ there are $\overset{\text{unique!}}{\vee}$ polynomials $q, r \in P$ ~~($\mathbb{R}$ or $\mathbb{C}$)~~ with $p = sq + r$ and $\deg r < \deg s$.

Pf: $T(q,r): P_{n-m} \times P_{m-1} \xrightarrow{T} P_n$
$(q, r) \longmapsto sq + r$ is a linear map.
For $\overset{n}{\overset{\|}{\deg p}} \geq \deg s^{=m}$, T is injective, since otherwise $sq = -r$ for nonzero poly^{s} of degree $\geq m$ and $\leq m-1$.
But $\dim(P_{n-m} \times P_{m-1}) = (n-m+1) + (m-1+1) = n+1 = \dim P_n$, so it is also surjective. □

A special case of this is when $\deg s = 1$.

Def^{n}: A root of a polynomial p is a value α with $p(\alpha) = 0$.

Cor: α is a root of p iff $(z-\alpha)$ divides $p(z)$.

Pf: If $p(z) = (z-\alpha)q(z)$, then $p(\alpha) = 0$. Otherwise, $p(z) = (z-\alpha)q(z) + r$ for some r with $\deg r = 0$, i.e., a constant. So, $p(\alpha) = r \neq 0$ and α is not a root. □

Cor: A nonzero polynomial of deg n has at most n roots.

Pf: In degree zero, $f(z) = a_0 \neq 0$ has no roots. In degree 1, $f(z) = a_0 + a_1 z$ has a unique root $z = {-a_0}/{a_1}$. Otherwise, induct: in degree n $f(z)$ either has no roots (✓) or at least one root. Pick one α and divide it out: $f(z) = q(z)(z-\alpha)$. The zero-product property reduces to q, with $\deg q = n-1$. □

Important fact: Every $\deg \geq 1$ poly^{l} over $\mathbb{C}$ has a root. "Fund. Th^{m} of Algebra"

Cor: Every $f \in P(\mathbb{C})$ has a unique (up to order) factorization as

$$f(z) = c \cdot (z-\alpha_1) \cdots (z-\alpha_n).$$

Pf: The existence of fact^{zn} follows from the Fact. If we had two such, we could pair one root α by ZPP. The two quotients agree except maybe at α — but they must agree here too by the previous Cor. Induct. □

Lem: Real polynomials factor into $(z-\alpha)$ and $((z-h)^2 + \beta^2)$, $\beta > 0$.

Pf: Real roots occur in conjugate pairs: $p(\bar{\alpha}) = \overline{p(\alpha)} = \bar{0} = 0$. For a complex root, translate it to the origin to get $(z-(h+i\beta))(z-(h-i\beta))$. □

Invariant Subspaces (5.A)

When we talked about matrices, we noticed that it was easier to understand

$$\begin{tikzcd} K^n \arrow[r, "f"] \arrow[d, phantom, "\cong"] & K^m \\ K_1 \oplus \cdots \oplus K_n \arrow[ur, "f|_{K_1} \oplus \cdots \oplus f|_{K_n}"'] & \end{tikzcd}$$

, where $K_j = \langle v_j \rangle$ was the span of v_j.

This idea holds in more generality: we can let the domain be V and take any sum decomposition of V. The idea, again, is that understanding $f|_{V_j}$ should be an easier problem than understanding f itself. Things are complicated by studying maps of the form $f : V \longrightarrow V$ with the same domain + codomain.

Consider:

$$\begin{tikzcd} \mathbb{R}^2 \arrow[rrr, "\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}"] \arrow[d, phantom, "\cong"] & & & \mathbb{R}^2 \arrow[d, phantom, "\cong"] \\ \langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle \arrow[urrr, "\begin{pmatrix} 1 \\ 0 \end{pmatrix}"] \arrow[rr, bend right=40, "\cdot 1"'] & \oplus \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle \arrow[urr, "\begin{pmatrix} 3 \\ 2 \end{pmatrix}"'] \arrow[r, bend right, "\cdot 2"] \arrow[rr, bend right=50, "\cdot 3"'] & \langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle & \oplus \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle \end{tikzcd}$$

. The linear map in this example does not respect the subspace decomposition: the second factor of the domain is blurred across both in the source.

Def$^{\underline{n}}$: A subspace $U \leq V$ is invariant if $\underset{\text{im}'' f|_U}{f(U)} \subseteq U$.

~~General pro~~

General problem: How can we find invariant subspaces? How finely can we find them (to avoid the trivial sol$^{\underline{n}}$ $U = V$)?

We'll start as finely as possible:

Def$^{\underline{n}}$: A vector $v \in V$ satisfying $f(v) = k \cdot v$ (i.e., $f(\langle v \rangle) \subseteq \langle v \rangle$) is called an eigenvector of eigenvalue k.

Ex: In the example above, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ are eigenvectors of eigenvalue 1 and 2 respectively. Moreover, $\mathbb{R}^2 \cong \begin{pmatrix} 1 \\ 0 \end{pmatrix} \oplus \begin{pmatrix} 3 \\ 1 \end{pmatrix}$, so this is as fine as possible.

$$\begin{tikzcd} \mathbb{R}^2 \arrow[r] \arrow[d, phantom, "\cong"] & \mathbb{R}^2 \arrow[d, phantom, "\cong"] \\ (1 \ 0) \arrow[r, "\cdot 1"] & (1 \ 0) \\ \oplus & \oplus \\ (3 \ 1) \arrow[r, "\cdot 2"] & (3 \ 1) \end{tikzcd}$$

Ex: Recall the 90° rotation operator $\mathbb{R}^2 \xrightarrow{\begin{pmatrix}0&-1\\1&0\end{pmatrix}} \mathbb{R}^2$. This has no eigenvectors over $\mathbb{R}$.

Over $\mathbb{C}$, it does: $\begin{pmatrix}y\\-x\end{pmatrix} = \begin{pmatrix}0&-1\\1&0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \lambda\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}\lambda x\\ \lambda y\end{pmatrix}$ has nonzero solutions by setting $y = \lambda x$ and $-x = \lambda y = \lambda(\lambda x)$, or $\lambda^2 + 1 = 0$, or $\lambda = \pm i$. Some corresponding eigenvectors are $(x, \cancel{x} -ix)$ for i and (x, ix) for $-i$.

Lem: If $v_1, \dots, v_m$ are eigenvectors for distinct eigenvalues then they are l.i.

Pf: Let $a_1 v_1 + \cdots + a_k v_k = 0$ be the earliest dependence. Then

$\overset{f}{\rightsquigarrow} a_1\lambda_1 v_1 + \cdots + a_k \lambda_k v_k = 0$, and

$\overset{\cdot \lambda_k}{\rightsquigarrow} a_1 \lambda_k v_1 + \cdots + a_k \lambda_k v_k = 0$, whose difference is

$a_1(\lambda_1 - \lambda_k)v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = 0$. This is ~~a shorter~~ an earlier dependence. □

Cor: $f: V \to V$ has at most $\dim V$ distinct eigenvalues. □

Def^n: If $\mathcal{U} \le V$ is invariant for f, then we can build two operators:

$$\begin{tikzcd}
\mathcal{U} \arrow[r, hook] \arrow[d, "f|_{\mathcal{U}}"] & V \arrow[r] \arrow[d, "f"] & V/\mathcal{U} \arrow[d, "f/\mathcal{U}"] \\
\mathcal{U} \arrow[r, hook] & V \arrow[r] & V/\mathcal{U}
\end{tikzcd}.$$

the behavior of f on $\mathcal{U}$ $\rightsquigarrow$ $f|_{\mathcal{U}}$

$f/\mathcal{U}$ $\leftsquigarrow$ intuitively, what's left over ignoring $\mathcal{U}$.

Warning: Ignoring $\mathcal{U}$ can get you into trouble, if there's no invariant complement to $\mathcal{U}$. Typical example:

$$y \longmapsto \begin{pmatrix}0\\y\end{pmatrix} \qquad \begin{pmatrix}x\\y\end{pmatrix} \longmapsto x$$

$$\begin{tikzcd}
\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right) \arrow[r, hook] \arrow[d, "0"] & \mathbb{R}^2 \arrow[r] \arrow[d, "\begin{pmatrix}0&0\\1&0\end{pmatrix}"] & \mathbb{R} \arrow[d, "0"] \\
\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right) \arrow[r, hook] & \mathbb{R}^2 \arrow[r] & \mathbb{R}
\end{tikzcd}$$

has zero $f|_{\mathcal{U}}$ and $f/\mathcal{U}$, but f is nonzero.

Eigenvectors + U-T matrices (5.B)

Q: How are we supposed to find eigenvectors? How do we even know that they exist?

Thm: $f : V \longrightarrow V$ for V an n-dim'l ~~f.d.~~ vector space $/ \mathbb{C}$ has an eigenvector.

Pf: Pick $v \neq 0$ and consider $\{v, fv, ffv, \ldots, f^n v\}$. This must have a dependence: $a_0 v + a_1 f v + \cdots + a_n f^n v = 0$, for nonzero coeff^s a_j. The polynomial $p(f) = a_0 + a_1 \cdot f + \cdots + a_n f^n$ factors as $p(f) = c(f - \lambda_1) \cdots (f - \lambda_n)$, and we substitute this in: $c(f - \lambda_1) \cdots (f - \lambda_n) v = 0$. One of these maps $f - \lambda_j$ must fail to be injective, i.e., $\exists\, \omega$ with $f\omega = \lambda_j \cdot \omega$. □

Some days ago, we worked through the example $\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}$, which had eigenvalues 1 and 2. This behavior is actually generic: the eigenvalues of an upper-triangular matrix lie on its diagonal. Additionally, every matrix admits ($/\mathbb{C}$) an upper-triangular presentation.

Rem: In terms of invariant subspaces, the matrix of f in a basis $(v_1, \ldots, v_n)$ is upper triangular when $\operatorname{span}(v_1, \ldots, v_j)$ is invariant for each j.

Cor: Over $\mathbb{C}$, every operator f admits an upper-triangular presentation.

Pf: We will induct on n, as the result is trivially true at $n = 1$. By the Theorem, let λ be an eigenvalue for f, and set $\mathcal{U} = \operatorname{im}(f - \lambda)$. This is a <u>proper</u> subspace which is <u>invariant under f</u>: for $u \in \mathcal{U}$, $f(u) = (f - \lambda)u + \lambda u$ decomposes as two things in $\mathcal{U}$. Hence, we can find an upper-triangular basis for $f|_{\mathcal{U}}$, $(u_1, \ldots, u_m)$, which we extend to a basis $(u_1, \ldots, u_m, v_1, \ldots, v_n)$ of V. By hypothesis, $u_j \in \operatorname{span}(u_1, \ldots, u_j)$. For v_j, $f(v_j) = (f - \lambda)(v_j) + \lambda v_j \in$ ~~span~~ $\mathcal{U} + \operatorname{span}(v_j)$

$\subseteq \operatorname{span}(u_1, \ldots, u_m, v_1, \ldots, v_j)$. □

(<u>Alternatively</u>: An eigenvector $v_1 \in V$ gives a matrix $\left(\begin{array}{c|c} \lambda_1 & \text{stuff from lifting.} \\ \hline 0 & \text{study } f/(v_1) \\ 0 & \text{and induct.} \\ \vdots & \\ 0 & \text{This is } \nabla. \end{array}\right)$.)

<u>Cor</u>: An upper-triangular matrix is invertible iff its diagonal entries are nonzero.

Pf: If λ_j are all nonzero, we can back-substitute to get $v_j \in \operatorname{im} f$ for all v_j in the basis. But then $\dim \operatorname{im} f = \dim V$. Conversely, if $\lambda_j = 0$ for some j, then $\operatorname{im} f|_{v_1, \ldots, v_j} \subseteq \operatorname{span}(v_1, \ldots, v_{j-1})$. This forces f not to be injective. □

<u>Cor</u>: The eigenvalues of an upper-triangular matrix appear in its diagonal.

<u>Pf</u>: $\begin{pmatrix} \lambda_1 - \lambda & * & \cdots & * \\ 0 & \ddots & & \vdots \\ \vdots & & & * \\ 0 & \cdots & 0 & \lambda_n - \lambda \end{pmatrix} = M - \lambda_j$ is non-invertible if and only if $\lambda = \lambda_j$ for some j. □

<u>Ex</u>: $M = \begin{pmatrix} -2 & 3 \\ -4 & 5 \end{pmatrix}$. $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $Mv = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$, $M^2 v = \begin{pmatrix} -8 \\ -12 \end{pmatrix} = 3Mv - 2v$. $M^2 - 3M + 2 = 0$
$(M-1)(M-2) = 0$.

$\lambda = 1$, as a guess. Has $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as a witness, so it's an eigenvalue.

Then $M - \lambda = \begin{pmatrix} -3 & 3 \\ -4 & 4 \end{pmatrix}$ has image $\mathcal{U} = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} = u_1 \right\}$.

Extend this to a basis $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} = u_1, \begin{pmatrix} 0 \\ 1 \end{pmatrix} = v_1 \right\}$. Then

$$\begin{tikzcd}
\mathbb{R}^2 \arrow[r, "\begin{pmatrix} -2 & 3 \\ -4 & 5 \end{pmatrix}"] \arrow[d, bend left, "\varphi^{-1}"] & \mathbb{R}^2 \arrow[d, bend left, "\varphi^{-1}"] \\
\mathbb{R}^2 \arrow[r, "\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix}"'] \arrow[u, bend left, "\varphi"] & \mathbb{R}^2 \arrow[u, bend left, "\varphi"]
\end{tikzcd}$$

$\begin{pmatrix} 1 & 3 \\ 0 & 2 \end{pmatrix} \leftarrow$ upper-triangular!

$$\begin{pmatrix} -2 & 3 \\ -4 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \textcircled{1} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \textcircled{0}.$$

$$\begin{pmatrix} -2 & 3 \\ -4 & 5 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \textcircled{3} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \textcircled{2}$$

Diagonalizability: a special case (5.C)

Previously we've discussed eigenvalues and eigenvectors.

Rem: If v_1 and v_2 are eigenvectors for the same eigenvalue λ, then so is any linear combination $k_1v_1 + k_2v_2$.

Def: The eigenspace associated to an eigenvalue λ is $E(\lambda, f) = \ker(f - \lambda)$.

Lem: For $\lambda \neq \lambda'$, $E(\lambda) \cap E(\lambda') = 0$. □

Cor: For $\{\lambda_j\}$, the sum $E(\lambda_1) + \cdots + E(\lambda_n)$ is direct, and $\dim(E(\lambda_1) + \cdots E(\lambda_n)) = \sum_j \dim E(\lambda_j)$. □

Rem: If f is diagonalizable, then its eigenvalues are the diagonal entries, and $V = \bigoplus_j E(\lambda_j)$.

Lem: Take V f.d., $f: V \to V$ linear with eigenvalues $\lambda_1, \dots, \lambda_m$. TFAE:

a) f is diagonalizable. b) V has a basis of eigenvectors.

c) There exist 1-dim'l invariant subspaces $U_j \leq V$ with $V = \bigoplus_j U_j$.

d) $V = E(\lambda_1) \oplus \cdots \oplus E(\lambda_m)$. e) $\dim V = \dim E(\lambda_1) + \cdots + \dim E(\lambda_m)$.

Pf: $a \Leftrightarrow b \Rightarrow c \Rightarrow b$ are all easy. $b \Rightarrow d$ by collecting invariant subspaces of like eigenvalue. $d \Rightarrow e$ by directness. To get $e \Rightarrow b$, union bases for the individual subspaces together. □

Cor: If f has $n = \dim V$ distinct eigenvalues, then f is diagonalizable.

Ex: $\begin{pmatrix} 1 & 1 & -1 \\ -6 & 8 & -3 \\ -4 & 4 & 1 \end{pmatrix}$ has eigenvalues 3, 2, and 5. eigenvectors $\begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, and $\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$

Ex: $\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ has eigenvalues 0 alone. eigenvectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ alone.

Ex: $\begin{pmatrix} -5 & -6 & 3 \\ 3 & 4 & -3 \\ 0 & 0 & -2 \end{pmatrix}$ has eigenvalues 1 and -2
eigenvectors $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

Ex: $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has eigenvalues 1 alone.
eigenvectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ alone.

Lingering questions:
- How can we effectively compute eigenvalues and eigenvectors?
- How can we recognize (special classes of) diagonal matrices?
- How can we compute (bounds on) $\dim E(\lambda)$?

Orthogonality: (6.A-B)

angle **Def**: An inner product on V is a bilinear f^n $\langle -,- \rangle : V \times V \longrightarrow$ ~~ℝ~~ $\mathbb{R}$ or $\mathbb{C}$ such that: (i) $\langle v,v \rangle \geq 0$ for all $v \in V$ (requires "$\geq$" on $\langle v,v \rangle$...), (ii) $\langle v,v \rangle = 0$ iff $v = 0$, (iii) $\langle u+v, w \rangle = \langle u,w \rangle + \langle v,w \rangle$, (iv) $\langle cu,v \rangle = c\langle u,v \rangle$, and (v) $\langle u,v \rangle = \overline{\langle v,u \rangle}$.

Ex: ① The dot products on $\mathbb{R}^n$ and $\mathbb{C}^n$.

② For numbers $c_i \geq 0$, the modified dot product
$$u \underset{c}{\cdot} w = c_1 u_1 \overline{w_1} + \cdots + c_n u_n \overline{w_n}.$$

③ $\langle f,g \rangle = \int_{-1}^{1} f(x) \cdot g(x)\,dx$ on $V = \{\text{integrable } f^{ns}\ [-1,1] \longrightarrow \mathbb{R}\}$.

length **Def**: The norm is defined by $\|v\| = \sqrt{\langle v,v \rangle}$. It satisfies $\|v\| = 0$ iff $v = 0$ and $\|\lambda v\| = |\lambda| \cdot \|v\|$.

Def: u and v are orthogonal when $\langle u,v \rangle = 0$.

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

Pf: $\langle u+v, u+v \rangle = \langle u,u \rangle + \cancel{\langle u,v \rangle}^{0} + \cancel{\langle v,u \rangle}^{0} + \langle v,v \rangle$. □

Thm (Cauchy-Schwarz): $|\langle u,v \rangle| \leq \|u\| \cdot \|v\|$, maximized only when $u = k \cdot v$ for a scalar k.

Pf: Write $u = \underbrace{\frac{\langle u,v \rangle}{\|v\|^2} \cdot v}_{\text{collinear with } v} + \underbrace{\left(u - \frac{\langle u,v \rangle}{\|v\|^2} \cdot v\right)}_{\text{orthogonal to } v \ =: w}$.

Pythag: $\|u\|^2 = \frac{|\langle u,v \rangle|^2}{\|v\|^4} \cdot \|v\|^2 + \|w\|^2 \geq \frac{|\langle u,v \rangle|^2}{\|v\|^2}$. □

Cor (Triangle Ineq): $\|u+v\| \leq \|u\| + \|v\|$ for all u, v. ■

Pf: $\langle u+v, u+v \rangle = \langle u,u \rangle + \langle v,v \rangle + \langle u,v \rangle + \overline{\langle u,v \rangle}$

$= $ —//— $+ \underline{2\mathrm{Re}\,\langle u,v \rangle} \leq 2|\langle u,v \rangle| \leq 2\|u\| \cdot \|v\|$

$= (\|u\| + \|v\|)^2$. □

<u>Def</u>: An <u>orthonormal basis</u> is a basis $(v_1, \ldots, v_n)$ with $\|v_j\| = 1$ and $\langle v_i, v_j \rangle \overset{i \neq j}{=} 0$.

<u>Lem</u>: In an orthonormal basis, $v = \langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n$. ~~□~~

Pf: Certainly $v = k_1 v_1 + \cdots + k_n v_n$. We can calculate k_j by applying $\langle -, v_j \rangle$. □

<u>Thm</u> (Gram-Schmidt): Every basis can be made orthonormal.

Pf: We induct. Make v_j orthogonal to the ones before it by replacing it by $v_j - \langle v_j, v_1 \rangle v_1 - \cdots - \langle v_j, v_{j-1} \rangle v_{j-1}$, and normalize that by replacing it with $v_j / \|v_j\|$. The span is preserved. □

<u>Rem</u>: This procedure preserves upper-triangularity.

Recall that we have a diagram

$$\begin{tikzcd}
(u,v) \arrow[rr, mapsto] \arrow[ddd, bend right, mapsto] & & \langle u,v \rangle \arrow[dd, equal] \\
V \times V \arrow[r, "{\langle -,- \rangle}"] \arrow[d, "\mathrm{id} \times \tau"'] & \mathbb{C} & \\
V \times V^* \arrow[ur, "\mathrm{ev}"] & & \langle u,v \rangle . \\
(u, \langle -, v \rangle) \arrow[urr, mapsto] & &
\end{tikzcd}$$

<u>Thm</u> (Riesz): For V finite dim^{l} and $\varphi \in V^*$, there is a unique $\omega \in V$ such that $\varphi = \tau(\omega)$. (That is, $\langle -, - \rangle$ induces an iso$^{\underline{m}}$ $V \xrightarrow{\tau} V^*$.)

<u>Pf</u>:
$$\begin{aligned}
\varphi(v) &= \varphi(\langle v, v_1 \rangle v_1 + \cdots + \langle v, v_n \rangle v_n) \\
&= \langle v, v_1 \rangle \varphi(v_1) + \cdots + \langle v, v_n \rangle \varphi(v_n) \\
&= \langle v, \underbrace{\overline{\varphi(v_1)} v_1 + \cdots + \overline{\varphi(v_n)} v_n}_{\omega} \rangle .
\end{aligned}$$

If ω_1 and ω_2 both do the job, then $\underset{\langle v, \omega_1 - \omega_2 \rangle .}{\langle v, \omega_1 \rangle - \langle v, \omega_2 \rangle} = \varphi(v) - \varphi(v) = 0$

So, pick $v = \omega_1 - \omega_2$ and use nondegeneracy. □

(<u>Rem</u>: We already knew τ was injective by a past Lemma.)

Minimization: (6.C)

Def: For an inner-productspace, the annihilator gives rise to the orthogonal subspace: $\mathcal{U}^\perp = \{ v \in V \mid \langle v, u \rangle = 0 \text{ for all } u \in \mathcal{U} \}$

Lem: this has a number of properties, some immediate from the connection to the annihilator:

(a) $V = \mathcal{U} \oplus \mathcal{U}^\perp$ for f.d. $\mathcal{U}$.

(b) $\dim \mathcal{U}^\perp = \dim V - \dim \mathcal{U}$ for f.d. V.

(c) $(\mathcal{U}^\perp)^\perp = \mathcal{U}$ for f.d. $\mathcal{U}$. □

Def: Part (a) gives rise to the projection operator: $P_\mathcal{U}(v) = P_\mathcal{U}(u + u^\perp) = u$, which discards the $\mathcal{U}^\perp$ component of v. For an orthonormal basis $e_1, \dots, e_n$ of $\mathcal{U}$, we have $P_\mathcal{U}(v) = \sum_j \langle e_j, v \rangle \cdot e_j$.

Lem: (a) $\ker P_\mathcal{U} = \mathcal{U}^\perp$.

(b) $\operatorname{im} P_\mathcal{U} = \mathcal{U}$, and $P_\mathcal{U}|_\mathcal{U} = \mathrm{id}$.

(c) $P_\mathcal{U}^2 = P_\mathcal{U}$.

(d) $\|P_\mathcal{U}(v)\| \leq \|v\|$. □

All of these are easy to verify. The real utility of $P_\mathcal{U}$ is the following:

Thm: Take $v \in V$, $\mathcal{U} \leq V$ f.d., and $u \in \mathcal{U}$. Then $\|v - P_\mathcal{U} v\| \leq \|v - u\|$ (with equality only at $u = P_\mathcal{U}(v)$).

Pf: $\|v - P_{\mathcal{U}}(v)\|^2 \leq \|v - P_{\mathcal{U}}(v)\|^2 + \|P_{\mathcal{U}}(v) - u\|^2$

$= \|v - P_{\mathcal{U}}(v) + P_{\mathcal{U}}(v) - u\|^2$ (Pythag.)

$= \|v - u\|^2$,

Equality happens iff $\|P_{\mathcal{U}}(v) - u\|^2 = 0$, or $P_{\mathcal{U}}(v) = u$. □

"$P_{\mathcal{U}}(v)$ is the closest point to v in $\mathcal{U}$."

Ex: As an example, we can use this to build approximations inside of function spaces.

Set $V = C[-\pi, \pi]$

$\mathcal{U} = \text{span}\{1, x, x^2, x^3, x^4, x^5\}$,

$f = \sin(x) \in V$.

Step ①: Do Gram-Schmidt to the basis of $\mathcal{U}$.

Step ②: Do orthogonal projection of f to $\mathcal{U}$ using the orthonormal basis.

(Tinkered with Mathematica examples)

Self-adjoint + normal operators (7.A)

$$\begin{tikzcd}
V \arrow[d, "f"] & & V^* \arrow[r, shift left, "\cong"] & V \arrow[l, shift left] \\
W \arrow[rr, phantom, "\rightsquigarrow"] & & W^* \arrow[u, "f^*"] \arrow[r, shift left, "\cong"] & W \arrow[l, shift left] \arrow[u, "f^*"]
\end{tikzcd}$$

using the inner-products on $V + W$.

So: f^* can also be considered as a map $f^* : W \longrightarrow V$ in the presence of inner-product. This map is called the adjoint of f, and it satisfies $\langle f(v), w\rangle_W = \langle v, f^*(w)\rangle_V$.

Lem: Again, we can mix annihilators + inner-products to produce:

(a) $\ker f^* = (\operatorname{im} f)^\perp$,

(b) $\operatorname{im} f^* = (\ker f)^\perp$,

(c) $\ker f = (\operatorname{im} f)^\perp$,

(d) $\operatorname{im} f = (\ker f)^\perp$, for $f : V \longrightarrow W$. $\square$

both f.d.

Lem: For $(e_1, \ldots, e_n)$ and $(f_1, \ldots, f_m)$ orthonormal bases of V and W, the matrix M^* representing $f^* : W \longrightarrow V$ is the conjugate transpose of M rep$^{\text{ning}}$ $f : V \longrightarrow W$. $\square$

Def: $f : V \longrightarrow V$ is self-adjoint when $f = f^*$.
(Cor: In an orthonormal basis, f is conjugate-symmetric.)

These operators have particularly nice properties. Here are some:

Lem: Every eigenvalue of a self-adjoint operator is real.

Pf: $\langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \bar{\lambda}\|v\|^2$

$\overset{\shortparallel}{\langle \lambda v, v\rangle} = \lambda \|v\|^2$. $\square$

Lem: Suppose V is complex and $f : V \longrightarrow V$ is linear $f^{\underline{n}}$. f is self-adjoint if + only if $\langle Tv, v\rangle \in \mathbb{R}$ for each $v \in V$. $\square$

Def: A slightly weaker property is for f to be normal: $f^* \circ f = f \circ f^*$.

Ex: $\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$ is normal but not self-adjoint (i.e., not symmetric).

Lem: T is normal iff $\|Tv\| = \|T^*v\|$ for all v.

Pf: Consider $\langle (TT^* - T^*T)v, v\rangle = 0$. □

Cor: T and T^* have the same eigenvectors w/ conjugate eigenvalues.

Pf: $0 = \|(T-\lambda)v\| = \|(T-\lambda)^*v\| = \|(T^* - \bar{\lambda})v\|$. □

Lem: If f is normal, then $v \in E(\lambda_1)$ and $w \in E(\lambda_2)$ are orthogonal.

Pf: $0 = \langle Tu, v\rangle - \langle u, T^*v\rangle = \langle \alpha u, v\rangle - \langle u, \bar{\beta} v\rangle$

$= (\alpha - \beta)\langle u, v\rangle$. Since $\alpha - \beta \neq 0$,

we must have $\langle u, v\rangle = 0$. □

We are going to prove the following theorem:

Thm: Let $f: V \to V$ be a linear f^n.

(a) If V is complex, then f is normal iff f is diagonalizable.

(b) If V is real, then f is self-adjoint iff f is diagonalizable.

... in an orthonormal basis.

The Spectral Theorem (7.B)

Last time, we announced two diagonalization theorems. Today, we prove them.

Thm: V a f.d. $\mathbb{C}$-vector space, $f: V \to V$ linear.
f is normal iff f admits an orthonormal diagonalization.

Pf: ($\Longleftarrow$) If f admits an orthonormal diagonalization, then f^* is diagonal for the same basis. Diagonal matrices commute.

($\Longrightarrow$) Start by finding an orthonormal basis in which f is upper-triangular, using Schur's theorem. We want to conclude that normality + U.T. $\Rightarrow$ diagonal.

Write $M = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & 0 & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix}$, so that $M^* = \begin{pmatrix} \overline{a_{11}} & & & \\ \overline{a_{12}} & \overline{a_{22}} & & 0 \\ \vdots & \vdots & \ddots & \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdots & \overline{a_{nn}} \end{pmatrix}$.

We proved last time that $\|Tv\| = \|T^*v\|$ for normal operators T, so we learn $\|a_{11}\|^2 = \|Te_1\|^2 = \|T^*e_1\|^2 = \|a_{11}\|^2 + \|a_{12}\|^2 + \cdots + \|a_{1n}\|^2$.
This forces $a_{12} = \cdots = a_{1n} = 0$. We can repeat this on $e_2, \ldots, e_n$. $\square$

In the real case, we are much worse off: we don't even know that real operators admit U.T. presentations (or eigenvectors).

Lem: Self-adjoint real operators have eigenvectors.

Pf: Begin the same as before: starting with $v \neq 0$, find a linear dependence in $(v, fv, f^2v, \ldots, f^n v)$, guaranteed by $\dim V = n$. From the dependence $a_n f^n v + \cdots + a_1 fv + a_0 v = 0$, extract a polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$, and factor it as $p(x) = c((x-h_1)^2 + k_1^2) \cdots ((x-h_m)^2 + k_m^2)(x-r_1)\cdots(x-r_\ell)$.
We want to show that $((f-h_j)^2 + k_j^2)$ is invertible, $k_j > 0$.

$$\text{We just do it: } \langle \overbrace{((f-h)^2+k^2)}^{f^2-2hf+h^2+k^2} v,v\rangle = \overset{\substack{\text{self-adj.}\\ \downarrow}}{\langle f^2 v,v\rangle} - 2h\overset{\substack{\text{Cauchy-Schwarz}\\ \downarrow}}{\langle fv,v\rangle} + (h^2+k^2)\langle v,v\rangle$$

$$\geq \|fv\|^2 + |2h|\|fv\|\|v\| + (h^2+k^2)\|v\|^2$$

$$= (\|fv\| + h\|v\|)^2 + (k\|v\|)^2 > 0 \text{ for } \|v\|>0.$$

With these factors eliminated, we proceed as in the $\mathbb{C}$ case. □

Thm: ~~[illegible]~~ V a f.d. $\mathbb{R}$-vector space with inner product, $f: V \to V$ linear. f is self-adjoint iff f admits an orthonormal diagonalization.

Pf: We induct on the dimension of V, since it is trivial for $\dim V = 1$. By the Lemma, f admits an eigenvector v, spanning an invariant 1-dim'l subspace $\mathcal{U}$. First, note that $\mathcal{U}^\perp$ is also invariant under f: for any $u \in \mathcal{U}$ and $v \in \mathcal{U}^\perp$, we have $\langle u, fv\rangle = \langle fu, v\rangle = 0$, so $fv \in \mathcal{U}^\perp$. Additionally, $f|_{\mathcal{U}^\perp}$ is still self-adjoint: $\langle f|_{\mathcal{U}^\perp} v, w\rangle = \langle fv, w\rangle = \langle v, fw\rangle = \langle v, f|_{\mathcal{U}^\perp} w\rangle$. Hence, we can induct on $f|_{\mathcal{U}^\perp}: \mathcal{U}^\perp \to \mathcal{U}^\perp$ to complete the proof. □

That's enough for one day. To summarize:

- $\mathbb{C}$ and normal ≡ diagonalizable orthonormally.
- $\mathbb{C}$ and self-adjoint ≡ ——〃—— + real eigenvalues.
- $\mathbb{R}$ and self-adjoint ≡ diagonalizable orthonormally.
- $\mathbb{R}$ and normal ⇜ §9.B.

Square roots and geometry (7.C)

One consequence of the spectral theorem is that, because diagonalized operators have very easy arithmetic, so do self-adjoint operators. Consider the following:

Def: An operator $f: V \to V$ is *positive* when it is self-adjoint and when $\langle fv, v\rangle \geq 0$ for all v. (If V is complex, we just ask for the inequality + drop the adjointness.)

Lem: For $f: V \longrightarrow V$, TFAE:

(a) f is positive. (b) f is self-adjoint + all the e.values are ≥ 0.

(c) f has a positive square root. (d) f has a self-adjoint square root.

(e) There is a second operator $g: V \to V$ with $f = g^* \circ g$.

Pf: $a \Rightarrow b$ by positivity on the e.vectors. $b \Rightarrow c$ by taking an entry-wise square root of the diagonal. $c \Rightarrow d$ trivially, as does $d \Rightarrow e$. To get $e \Rightarrow a$, $T^* = (R^*R)^* = R^*R^{**} = R^*R = T$. □

In fact, if we fully restrict attention to positive operators,

Lem: ... the positive square root of a positive operator is unique.

Pf: For g a root of f and $fv = \lambda v$ an eigenvector of f, we want to show $gv = \sqrt{\lambda}\, v$, for g any positive square root. We know g admits a diagonalization and that its square has eigenvalues the squares of those of g. The l.i. lemma for eigenvectors forces $gv = \sqrt{\lambda}\, v$. □

Lingering question: How many other square roots are there?

Non-obvious def^{n}: $f: V \to V$ is an *isometry* if $\|fv\| = \|v\|$ for all $v \in V$. These are the "geometry-preserving f^{ns}".

Lem: For $f: V \to V$, TFAE:

(a) f is an isometry. (b) $\langle fu, fv \rangle = \langle u, v \rangle$ for all u, v.

(c) $fe_1, \dots, fe_n$ is orthonormal for each orthonormal list $e_1, \dots, e_n$.

(d) there exists any orthonormal list $e_1, \dots, e_n$ s.t. $fe_1, \dots, fe_n$ is too.

(e) $f^* \circ f = id$. (f) $f \circ f^* = id$. (g) f^* is an isometry.

(h) f is invertible and $f^{-1} = f^*$.

Pf: (a$\Rightarrow$b) This was homework: inner products can be computed from norms.

(b$\Rightarrow$c) Being orthonormal is an inner product condition.

(c$\Rightarrow$d) Trivial: pick any orthonormal basis.

(d$\Rightarrow$e) We have $\langle e_i, e_j \rangle = \langle fe_i, fe_j \rangle = \langle f^* f e_i, e_j \rangle$. Since (e_j) forms a basis, this gives $\langle f^* f u, v \rangle$ for all $u, v \in V$. This forces $f^* f = id$.

(e$\Rightarrow$f) Since V is finite dimensional, $f^* f = id$ forces $f f^* = id$.

(f$\Rightarrow$g) $\| f^* v \|^2 = \langle f^* v, f^* v \rangle = \langle f f^* v, v \rangle = \langle v, v \rangle = \| v \|^2$.

(g$\Rightarrow$h) Apply a$\Rightarrow$e$\wedge$f for f^*.

(h$\Rightarrow$a) $\| f v \|^2 = \langle fv, fv \rangle = \langle f^* f v, v \rangle = \langle v, v \rangle = \| v \|^2$. □

Rem: (e) is supposed to mean that there are lots of square roots of the identity, connected to the various isometries

Polar decomposition + SVD (7.D)

Today we use our study of square roots to tackle presentations of arbitrary operators.

"Polar decomp."

Thm: For $f: V \to V$, there is an isometry g with $f = g \circ \underbrace{\sqrt{f^* \circ f}}$.

Pf: Start by noting $\|fv\|^2 = \langle f^* f v, v\rangle = \langle \sqrt{f^*f}\, v, \sqrt{f^*f}\, v\rangle$ $\underline{\|}$ positive.
$= \|\sqrt{f^*f}\, v\|^2$. We "define" a function $g : \operatorname{im}\sqrt{f^*f} \to \operatorname{im} f$ by $g(\sqrt{f^*f}(v)) = f(v)$ — now we need to check ① that this def$^{\underline{n}}$ is sound, ② that it extends to V, and ③ that we get an isometry in the end.

①: $\|fv_1 - fv_2\| = \|f(v_1 - v_2)\| = \|\sqrt{f^*f}(v_1 - v_2)\| = 0$, so that $\ker\sqrt{f^*f} = \ker f$.

②: We also learn that $\dim \operatorname{im} f = \dim \operatorname{im}\sqrt{f^*f}$ and that $\dim(\operatorname{im} f)^\perp = \dim(\operatorname{im}\sqrt{f^*f})^\perp$. We use this to extend g: g acts as above on $\operatorname{im}\sqrt{f^*f}$ and by any isometry carrying an orthonormal basis of $(\operatorname{im}\sqrt{f^*f})^\perp$ to $(\operatorname{im} f)^\perp$ one of.

③ So extended, g is an isometry: g's two definitions are individually isometric, and the Pythagorean Theorem extends this over the orthogonal sum. ☐

Rem: Even though g and $\sqrt{f^*f}$ are diagonalizable, this may require different orthonormal bases for each.

In fact, this isn't such a problem, because isometries are nice enough in any basis. Favoring the orthonormal basis diagonalizing $\sqrt{f^*f}$ leads to the Singular Value Decomposition.

Def: the singular values of f are the eigenvalues of $\sqrt{f^*f}$, with each eigenvalue λ repeated $\dim E(\lambda, \sqrt{f^*f})$ times. (these are the diagonal entries of an orthonormal diagonal presentation of $\sqrt{f^*f}$.)

Thm (SVD): There exist orthonormal bases (e_j) and (d_j) of V such that $f(v) = s_1 \langle v, e_1\rangle d_1 + \cdots + s_n \langle v, e_n\rangle d_n$, for $(s_1, \ldots, s_n)$ the singular values of f.

Pf: Let $(e_1, \ldots, e_n)$ present an orthonormal diagonalization of $\sqrt{f^*f}$, so that $(\sqrt{f^*f})(v) = s_1 \langle v, e_1\rangle e_1 + \cdots + s_n \langle v, e_n\rangle e_n$. Then the columns of the isometry g appearing in the polar decomposition of f give an orthonormal set $(d_j = g(e_j))$, and

$$f(v) = g\left(s_1 \langle v, e_1\rangle e_1 + \cdots + s_n \langle v, e_n\rangle e_n\right)$$

$$= s_1 \langle v, e_1\rangle d_1 + \cdots + s_n \langle v, e_n\rangle d_n. \qquad \square$$

Rem: This is a slick, useful upgrade from Gaussian elimination, which also cleverly picked bases that diagonalized a matrix.

Rem: The e.values of $\sqrt{f^*f}$ are the nonnegative roots of the e.values of f^*f.

Ex: $f(x_1, x_2, x_3, x_4) = (0, 3x_1, 2x_2, -3x_4)$ has $f^*f(x_1, x_2, x_3, x_4) = (9x_1, 4x_2, 0, 9x_4)$, so the s.values of f are $(3, 3, 2, 0)$, whereas the e.values of f are merely -3 + 0, which is not enough to recover f (since f is not normal, hence not diagonalizable).

Generalized Eigenvectors (8.A)

In this chapter, we are aiming to correct a deficiency in our discussion of eigenspaces and diagonalization: the only operators f admitting diagonal presentations are those with $V = \bigoplus_j E(\lambda_j, f)$, but in general $\bigoplus_j E(\lambda_j, f)$ may be a proper subspace of V, as with $\begin{pmatrix}0&1\\0&0\end{pmatrix}$, which has $E(0) = \langle e_1 \rangle \lneq \mathbb{R}^2$. This example is instructive: the behavior of this operator stabilizes after two applications, and $E(0, f^2) = \mathbb{R}^2$ is not a proper subspace.

Lem: For any $f: V \to V$, $0 = \ker f^0 \leq \ker f^1 \leq \ker f^2 \leq \cdots$ □

Lem: For $N \geq \dim V$, $\ker f^N = \ker f^{N+1}$.

Pf: First, note that if $\ker f^N = \ker f^{N+1}$ is ever satisfied then $\ker f^{N+M} = \ker f^{N+M+1}$ for any $M \geq 0$: for $v \in \ker f^{N+M+1}$ we have $0 = f^{N+M+1}(v) = f^{N+1}(f^M v) \Rightarrow f^N(f^M v) = 0$. Second, we can't have an ascending chain of subspaces of length $> \dim V$. □

Cor: For $N \geq \dim V$, $V = \ker f^N \oplus \operatorname{im} f^N$.

Pf: First check directness: $v \in (\ker f^N) \cap (\operatorname{im} f^N)$ satisfies $f^N v = 0$ and $v = f^N w$, but then $f^{2N} w = 0$ implies $f^N w = 0 = v$.
From here, apply the FToLA to $f^N : V \to V$. □

The extreme case of this gets a special name:

Def: f is called nilpotent if $N \gg 0$ gives $\ker f^N = V$ (or $f^N = 0$).

Rem: Build a basis of $\ker f$, extend to one of $\ker f^2$, ..., etc.
f is upper-triangular with a 0 diagonal for this basis.

This also leads us to consider the "stable" behavior of eigenvectors.

Def: The generalized eigenspace is $G(\lambda, f) = \ker (f - \lambda \cdot I)^N$, $N \geq \dim V$.

Generalized eigenvectors have properties akin to classical eigenvectors.

<u>Lem:</u> If $v_1, \ldots, v_m$ are generalized eigenvectors for distinct e.values $\lambda_1, \ldots, \lambda_m$, then $(v_1, \ldots, v_m)$ is a linearly independent list.

<u>Pf:</u> Consider a dependence $0 = a_1 v_1 + \cdots + a_m v_m$. Let $k \geq 0$ be the largest value with $w = (f - \lambda_1)^k v_1 \neq 0$, so that $(f - \lambda_1) w = 0$, witnesses w as an eigenvector. Hence, we calculate

$$\begin{aligned} 0 &= (f-\lambda_1)^k (f-\lambda_2)^n \cdots (f-\lambda_m)^n (a_1 v_1 + \cdots + a_m v_m) \\ &= (f-\lambda_1)^k (f-\lambda_2)^n \cdots (f-\lambda_m)^n (a_1 v_1) \\ &= a_1 (f-\lambda_2)^n \cdots (f-\lambda_m)^n w \\ &= a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n w, \text{ which forces } a_1 = 0. \end{aligned}$$

Repeating this with other a_j gives $a_j = 0$ for each j. □

<u>Cor:</u> This gives us an extension

$$\bigoplus_j E(\lambda_j, f) \leq \bigoplus_j G(\lambda_j, f) \leq V.$$

<u>Next time</u>: This is always an equality.

Decomposition of an operator (8.B):

Thm: For $V/\mathbb{C}$ finite dim'l, and $f: V \to V$ a linear operator, let $\lambda_1, \dots, \lambda_m$ be the eigenvalues of f. Then:

(a) $V = \bigoplus_j G(\lambda_j, f)$.

(b) Each $G(\lambda_j, f)$ is invariant under f.

(c) $(f - \lambda_j)|_{G(\lambda_j, f)}$ is nilpotent.

Pf: (b) Note that $\operatorname{im} p(f)$ and $\ker p(f)$ are invariant under f for any poly'l p. Then, $G(\lambda_j, f) = \ker (f - \lambda_j)^N$ is such a subspace.

(c) Follows from the def'n, since $G(\lambda_j, f) = \ker (f - \lambda_j)^N$.

(a) We induct on dim V. Start by extracting an eigenvalue λ_1 of f, which decomposes V as $G(\lambda_1, f) \oplus \mathcal{U}$, $\mathcal{U} := \operatorname{im}(f - \lambda_1)^N$, which is also an invariant subspace. We induct to get $\mathcal{U} = G(\lambda_2, f|_\mathcal{U}) \oplus \dots \oplus G(\lambda$ and we want to show $G(\lambda_j, f|_\mathcal{U}) = G(\lambda_j, f)$. "$\subseteq$" is immediate. To get "$\supseteq$", take $v \in G(\lambda_j, f)$, which we write as $v = v_1 + u$, and decompose $u = \underset{G(\lambda_2, f|_\mathcal{U})}{v_2} + \dots + \underset{G(\lambda_m, f|_\mathcal{U})}{v_m}$ to get $v = \underset{G(\lambda_1, f)}{v_1} + v_2 + \dots + v_m$. || The linear independence lemma then forces $v_2 = 0$ except for v_j. In particular, $v_1 = 0$, so $v = u$, but then $v \in G(\lambda_j, f|_\mathcal{U})$. $\square$

So, if you are willing to tolerate generalized eigenvectors, you can exhaust V. Our question is then: what good is this?

Def: the algebraic multiplicity of λ is $\dim G(\lambda, f)$. The geometric multiplicity of λ is $\dim E(\lambda, f)$. (Axler just calls the former "multiplicity"

Def: Block matrices are matrices built by sewing smaller matrices together. A matrix is block diagonal if it's diagonal as a block matrix.

Cor: Every $\mathbb{C}$-operator admits a basis s.t. its presentation is block-diagonal with U.T. blocks.

Pf: Break V up into $\bigoplus_j G(\lambda_j, f)$. Then $f - \lambda_j|_{G(\lambda_j, f)}$ is nilpotent, so admits an U.T. presentation w/ zeroes on the diagonal. The same basis makes $f|_{G(\lambda_j, f)}$ U.T. w/ λ_j's on the diagonal. □

In 8.D we will do even better than this. Right now, though, we can already find a neat application:

Lem: If $M = I + N$ for N nilpotent, there exists $\sqrt{M}$.

Pf: Taylor expand $\sqrt{1-x}$. Because N is nilpotent, we only need finitely many terms + don't care about convergence. □

Cor: Any invertible operator $f / \mathbb{C}$ has a square root.

Pf: Decompose f into block diagonal U.T. form. Each block can be written as $\lambda \cdot I + N$ for $\lambda \neq 0$ and N nilpotent, hence each block has a square root. Reassembling the blocks gives a square root for f. □

Characteristic + Minimal Polynomials (8.C)

Def: The minimal polynomial of an operator f is the monic polynomial p of minimal degree such that $p(f) = 0$.

Lem: Such a polynomial exists.

Pf: Take $n = \dim V$. Then $(1, f, f^2, \ldots, f^{n^2})$ is dependent in $\mathcal{L}(V, V)$, and we take m to be the smallest index with $(1, f, \ldots, f^m)$ dependent. The dependence gives a candidate polynomial. To see uniqueness, note that the difference of two candidate poly^s is another poly^l with lower degree. □

Rem: $\deg(\text{minpoly}(f)) \le (\dim V)^2$, by this proof.

Def: For $f / \mathbb{C}$, the characteristic polynomial of f is

$$\text{charpoly}(f) = \prod_{\text{eigenvalues } \lambda} (z - \lambda)^{\dim G(\lambda, f)}.$$

~~Rem~~ Thm (Cayley-Hamilton): ~~Then~~ The characteristic polynomial of f evaluated at f gives zero.

Pf: Decompose $V = \bigoplus_j G(\lambda_j, f)$, and shuffle the factors of $\text{charpoly}_f(z)$ so that $(z - \lambda_j)^{\dim G(\lambda_j, f)}$ appears last. This kills the vectors in $G(\lambda_j, f)$ by definition. □

Cor: $\text{minpoly}(f) \mid \text{charpoly}(f)$, and in particular $\deg \text{minpoly} \le \overset{\overset{n}{\shortparallel}}{\deg \text{charpoly}}$.

Pf: In fact, the minimal poly divides any poly q with $q(f) = 0$. The division algorithm gives $q = \min \cdot s + r$ with $\deg r < \deg \min$ and $r(f) = r(f) + \min(f) \cdot s(f) = q(f) = 0$, which forces $r = 0$. □

Our last result is that the minimality of the minimal polynomial does not remove from it the basic features of the characteristic poly.

Lem: Write p for the minimal polynomial. For λ a zero of p,
$p(z) = (z-\lambda)\cdot q(z)$, and $p(f)(v) = (f-\lambda)\circ q(f)(v) = 0$.
By minimality, $q(f)(v) \neq 0$ for some v, so this is an e.vector of f with weight λ. In the other direction, if λ is an e.value of f w/ e.vector v, then $0 = p(f)(v) = p(\lambda)v$, hence $p(\lambda) = 0$. □

Ex: $M = \begin{pmatrix} 0&0&0&0&-3 \\ 1&0&0&0&6 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \end{pmatrix}$, $M^2 = \begin{pmatrix} 0&0&0&-3&0 \\ 0&0&0&6&-3 \\ 1&0&0&0&6 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \end{pmatrix}$, $M^3 = \begin{pmatrix} 0&0&-3&0&0 \\ 0&0&6&-3&0 \\ 0&0&0&6&-3 \\ 1&0&0&0&6 \\ 0&1&0&0&0 \end{pmatrix}$

$M^4 = \begin{pmatrix} 0&-3&0&0&0 \\ 0&6&-3&0&0 \\ 0&0&6&-3&0 \\ 0&0&0&6&-3 \\ 1&0&0&0&6 \end{pmatrix}$, $M^5 = \begin{pmatrix} -3&0&0&0&-18 \\ 6&-3&0&0&36 \\ 0&6&-3&0&0 \\ 0&0&6&-3&0 \\ 0&0&0&6&-3 \end{pmatrix}$. $\rightsquigarrow M^5 - 6M$ is diagonal
$\rightsquigarrow M^5 - 6M + 3 = 0$

gives the minimal polynomial.

i.e., e.values!

Fact from Math 123: There are 5 roots of this polynomial, all distinct, none expressable in terms of radical expressions.

Cor: Computing eigenvalues exactly is not a solvable problem.

Jordan Form (8.D)

Previously we've shown the nilpotent operators admit bases in which their matrix representative is U.T. with vanishing main diagonal. Our goal today is to improve this: we will show that we can find a matrix that is nonzero only on the superdiagonal, and there has only 0's + 1's.

Lem: For N nilpotent, there are vectors $v_1, \dots, v_m$ and indices $k_1, \dots, k_m$ such that

(a) $N^{k_1} v_1, \dots, v_1, \dots, N^{k_m} v_m, \dots, v_m$ is a basis for V.

(b) $N^{k_1+1} v_1 = \cdots = N^{k_m+1} v_m = 0$.

Pf: We induct on $\dim V$, since $\dim V = 1 \Rightarrow N = 0$. Since N is nilpotent, N is neither injective nor surjective, and we can form $N|_{\operatorname{im} N}$. Applying the inductive hypothesis, we get vectors $v_1, \dots, v_m \in \operatorname{im} N$ and indices $k_1, \dots, k_m$ satisfying (a) and (b). Preimage each v_j to $N(u_j) = v_j$ and trade k_j for $k_j + 1$. We claim this is at least l.i.: a dependence would image to a dependence in $\operatorname{im} N$, leaving just $N^{k_1+1} u_1, \dots, N^{k_m+1} u_m$ unaccounted for — but there too are l.i.. Extending to a basis gives other vectors $w_1, \dots, w_j$ with $Nw_1, \dots, Nw_j \in \operatorname{im} N$, hence these can be perturbed to have the property $Nw_1 = \cdots = Nw_j = 0$. □

Thm: Every $\mathbb{C}$-operator $f: V \to V$, $\dim V < \infty$, admits a Jordan basis, where f has a block-diagonal expression by blocks of the form $\begin{pmatrix} \lambda_j & 1 & 0 \\ & \ddots & 1 \\ 0 & & \lambda_j \end{pmatrix}$.

Pf: Nilpotent operators were handled by the previous Lemma. In general, break up $V = \bigoplus_j G(\lambda_j, f)$, and consider $(f - \lambda_j)|_{G(\lambda_j, f)}$, which is nilpotent. A Jordan basis for $f - \lambda_j$ is also a Jordan basis for f, hence we can take the union over j. □

Complexification (9.A)

Jordan normal form is about as much as anyone knows about nice presentations of complex operators. We now turn to real operators, where most of our theorems fail b/c we cannot assume the existence of an e.vector. Our strategy will be to replace $f : V \to V$ over $\mathbb{R}$ with a complex operator that retains much of the information of f.

<u>Def</u>: For $V/\mathbb{R}$, we define $V_\mathbb{C}/\mathbb{C}$ by $V_\mathbb{C} = $ "$V \oplus i \cdot V$".
For $f : V \to V$, we define $f_\mathbb{C} : V_\mathbb{C} \to V_\mathbb{C}$ by $f(v + iv') = f(v) + i f(v')$.

<u>Ex</u>: $\mathbb{R}^n_\mathbb{C} \cong \mathbb{C}^n$, and this preserves matrices:

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, phantom, "\cong" rotate=90] & W \arrow[d, phantom, "\cong" rotate=90] & & V_\mathbb{C} \arrow[r, phantom, "\cong"] \arrow[d, "f_\mathbb{C}"] & \mathbb{R}^n_\mathbb{C} \arrow[r, phantom, "\cong"] \arrow[d] & \mathbb{C}^n \arrow[d, "M_\mathbb{C}"] \\ \mathbb{R}^n \arrow[r, "M"] & \mathbb{R}^m \arrow[r, phantom, "\rightsquigarrow"] & & W_\mathbb{C} \arrow[r, phantom, "\cong"] & \mathbb{R}^m_\mathbb{C} \arrow[r, phantom, "\cong"] & \mathbb{C}^m \end{tikzcd}$$

has the same entries as M.

<u>Lem</u>: Real operators admit invariant subspaces of dim. 1 or 2.
Pf: $f_\mathbb{C}$ has an eigenvector: $f(u+iv) = (a+bi)(u+iv)$
$= (au - bv) + i(bu + av)$. So, take $U = \operatorname{span}\{u, v\}$. □

<u>Lem</u>: The minimal poly$^{\underline{s}}$ of f and $f_\mathbb{C}$ agree.
Pf: For p the min. poly. of f, $p(f_\mathbb{C}) = p(f))_\mathbb{C} = 0$.
Conversely, if $q \in \mathbb{C}[x]$ satisfies $q(f_\mathbb{C}) = 0$, then $(\operatorname{Re} q)(f) = 0$, so comparing degrees forces $f_\mathbb{C}$'s min poly $= p_\mathbb{C}$. □

<u>Cor</u>: For $\lambda \in \mathbb{R}$, λ is an e.value of f iff it's an e.value of $f_\mathbb{C}$. □

<u>Lem</u>: $(f_\mathbb{C} - \lambda)^j (u + iv) = 0$ iff $(f_\mathbb{C} - \bar{\lambda})^j (u - iv) = 0$. □

<u>Cor</u>: $\lambda \in \mathbb{C}$ is an e.value of $f_\mathbb{C}$ iff $\bar{\lambda}$ is too, <u>and</u> their multiplicities agree. □

<u>Cor</u>: Every real operator on an odd-dim$^{\underline{l}}$ space has an e.value. □

Cor: The characteristic polynomial of $f_{\mathbb{C}}$ is actually real.

Pf: Remember the formula $\text{char poly}_f(z) = \prod_{\lambda} (z-\lambda)^{\dim G(\lambda, f)}$.

Our previous Cor says $\lambda \in \mathbb{C} \setminus \mathbb{R}$ and $\bar{\lambda}$ come in equal weights, and these factors collect to give

$$(z-\lambda)^{\dim G(\lambda,f)}(z-\bar{\lambda})^{\dim G(\bar{\lambda},f)} = (z - 2\mathrm{Re}(\lambda)z + |\lambda|^2)^{\dim G(\lambda,f)}. \quad \square$$

Def: The characteristic polynomial of a real operator f is the characteristic polynomial of the complexification $f_{\mathbb{C}}$.

Consequences: $f: V \to V$ a real operator.

(a) $\deg \text{char poly}(f)(z) = \dim V$.

(b) real zeroes of char poly(f) are real e.values of f.

(c) [Cayley-Hamilton] $\text{charpoly}_f(f) = 0$.

(d) The minimum polynomial divides the char. polynomial.

(e) deg min. poly $\leq$ deg char. poly. $\square$

Operators on real inner product spaces (9.B)

Goal: Understand normal real operators, the last case of the spectral-type results.

Start just with $\dim V = 2$:

Lem: For $f : V \to V$, $\dim V = 2$, TFAE:

(a) f normal but not self-adjoint.

(b) All orthonormal bases present f as $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ for $b \neq 0$.

(c) For some ortho-basis, f is presented as $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ for $b > 0$.

Pf: ($a \Rightarrow b$) Start with an ortho-basis e_1, e_2, presenting f as $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$. Then $\|Te_1\|^2 = \|T^* e_1\|^2$ implies $a^2 + b^2 = a^2 + c^2$. If $b = c$, we're self-adjoint, so $-b = c$ instead. Then,

$$f^* f = \begin{pmatrix} a^2 + b^2 & -ab + bd \\ -ab + bd & b^2 + d^2 \end{pmatrix} \text{ and } ff^* = \begin{pmatrix} a^2 + b^2 & ab - bd \\ ab - bd & b^2 + d^2 \end{pmatrix},$$

and $f^* f = ff^*$ forces $2ab = 2bd$, or $a = d$.

($b \Rightarrow c$) Either (e_1, e_2) works or $(e_1, -e_2)$ does.

($c \Rightarrow a$) Actually do the matrix mult. □

Now, we will soon want an inductive decomposition of V. The following lemma assures us that this is a sane thing to do.

Lem: V a f.d. inner-product space, $f : V \to V$ normal, $U \leq V$ invariant.

(a) $U^\perp$ is invariant under f.

(b) U is invariant under f^*.

(c) $(f|_U)^* = f^*|_U$.

(d) $f|_U$ and $f|_{U^\perp}$ are normal operators.

Pf: Begin by extending an orthonormal basis of U to one of V:
$(e_1, \ldots, e_m, f_1, \ldots, f_n)$.

Inside of this basis, f presents as $\left(\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right)$, since $\mathcal{U}$ is invariant.

But: $\sum_{i,j} |A_{ij}|^2 = \sum_j \|f e_j\|^2 = \sum \|f^* e_j\| = \sum_{ij} |A_{ij}|^2 + \sum_{ij} |B_{ij}|^2$,

so B is the zero matrix. Invariance of $\mathcal{U}^\perp$ follows. For (b) + (c), the conjugate transpose of f's matrix is again block-diagonal, which gives invariance of $\mathcal{U}$ under f^* and a calculation of $f^*|_{\mathcal{U}}$. □

Thm: For V a f.d. $\mathbb{R}$-inner product space, $f: V \to V$ is normal iff V has an orthonormal basis where f presents as block-diagonal with 1×1 scaling blocks and 2×2 scale + rotate blocks.

Pf: ($\Leftarrow$) Scaling + rotations all commute.

($\Rightarrow$) Induct on dim V; f has an invariant subspace $\mathcal{U}$ of dim 1 or 2, and $\mathcal{U}^\perp$ is invariant under f. We did the 2×2 case at the beginning. □

Cor: For V as above $f: V \to V$ is an isometry iff f admits an orthonormal presentation as a block diagonal matrix as above without scaling (i.e., the 1×1-blocks are $[1]$ or $[-1]$ + 2×2's are $\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}$.).

Pf: Isometries are normal, so the Thm applies. Because f is an isometry, it can't scale anything. □

Thm w/o proof: Real operators also admit Jordan decompositions. Each Jordan block is either (i) a complex Jordan block $\begin{pmatrix}\lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda\end{pmatrix}$ or (ii) a block diagonal matrix itself with identical 2×2 blocks $\begin{bmatrix}a_i & b_i \\ -b_i & a_i\end{bmatrix}$ (describing mult. by λ) on the diagonal + 2×2 identity blocks on the block-superdiagonal. □

Trace (10.A):

As briefly advertised earlier in the semester, some of the coefficients of the characteristic polynomial deserve special attention: the trace and the determinant. The trace is the less interesting of the two, so we treat it first.

Def: In the expansion $(z-\lambda_1)\cdots(z-\lambda_n) = z^n - (\lambda_1 + \cdots + \lambda_n)z^{n-1} + \cdots$, the coeff^t of $-z^{n-1}$ is called the trace of the operator f. (It is the sum of the e.values, repeated by algebraic mult.)

Our main goal today is to show that this value is actually computable — unlike any particular e.value alone.

Def: Given a square matrix M, the trace of the matrix is the sum $\sum_{j=1}^{n} M_{jj}$ of its diagonal entries.

Thm: The two definitions of the trace agree when expanding f in a basis.

Lem: If A and B are matrices of the same size, then $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

Pf: The j^{th} term on the diagonal of AB is expressed by $(AB)_{jj} = \sum_{k=1}^{n} A_{jk}B_{kj}$. Summing over j, we have

$$\mathrm{tr}(AB) = \sum_{j=1}^{n} (AB)_{jj} = \sum_{j=1}^{n}\sum_{k=1}^{n} A_{jk}B_{kj} = \sum_{k=1}^{n}\sum_{j=1}^{n} B_{kj}A_{jk} = \sum_{k=1}^{n} (BA)_{kk} = \mathrm{tr}(BA). \quad \square$$

Cor: The trace of a matrix is invariant under change of basis.

Pf:

$$\begin{tikzcd}
 & V \arrow[rr, "f"] \arrow[d, shift left] \arrow[d, phantom, "\cong"] \arrow[ddl, bend right] & & V \arrow[d, shift left] \arrow[d, phantom, "\cong"] \arrow[ddr, bend left] & \\
 & \mathbb{R}^n \arrow[u, shift left] \arrow[rr, "M"] \arrow[dl, shift left, "B^{-1}"] \arrow[dl, phantom, "\cong"] & & \mathbb{R}^n \arrow[u, shift left] \arrow[dr, shift left, "B^{-1}"] \arrow[dr, phantom, "\cong"] & \\
\mathbb{R}^n \arrow[ur, shift left, "B"] \arrow[uur, bend left] \arrow[rrrr, "N=B^{-1}MB."] & & & & \mathbb{R}^n \arrow[ul, shift left, "B"] \arrow[uul, bend right]
\end{tikzcd}$$

encodes a change of basis.

Hence, $\mathrm{tr}(N) = \mathrm{tr}(B^{-1}(MB)) = \mathrm{tr}((MB)B^{-1}) = \mathrm{tr}(M)$. □

Pf of Thm: Put f (or $f_{\mathbb{C}}$) into upper-triangular form. There, the two definitions of trace clearly agree. Coupling that to Cor, we are done. □

This has surprising corollaries of its own:

Cor: tr is additive: $\mathrm{tr}(M+N) = \mathrm{tr}(M) + \mathrm{tr}(N)$. □

Cor: There do not exist operators f, g with $fg - gf = \mathrm{id}$.

Pf: $\mathrm{tr}(fg - gf) = \mathrm{tr}(fg) - \mathrm{tr}(gf) = \mathrm{tr}(fg) - \mathrm{tr}(fg) = 0$. Meanwhile, $\mathrm{tr}(\mathrm{id}) = \dim V \neq 0$. □

Determinants (10.B)

Def: The *determinant* of an *operator* f is $(-1)^{\dim V}$ times the constant coeff^t of its characteristic poly^l:

$$\operatorname{charpoly}(z) = z^n - \operatorname{tr}(f) z^{n-1} + \cdots + (-1)^n \det(f).$$

Cor: (From homework): f is invertible iff $\det(f) \neq 0$. □

Cor: The characteristic poly^l of f is $\det(z - f)$.

Pf: Note that λ is an e.value of f iff $(z-\lambda)$ is an e.value of $z-f$: $-(f-\lambda) = (z-f) - (z-\lambda)$. Raising both sides to $\dim V$ and taking nullspaces also shows the algebraic multiplicities match. The characteristic poly^l of f and the determinant of $z-f$ thus match factor-wise. □

Warning: Above, we slyly traded our k-linear map $f: V \to V$ for a $k[z]$-linear map $f_{k[z]}: V_{k[z]} \to V_{k[z]}$ à la complexification. **However**, $k[z]$ is *not* a *field*! You can make this legal either by inventing *modules* or by using the field $k(z)$ of rat^l poly^s. You need to build gen. e.space decomposition either way, though...

As last time, we now want to start computing the determinant of operators presented as matrices. In the diagonal case,

$$\det \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} = a_{11} \cdot a_{22} \cdot \cdots \cdot a_{nn},$$

which is also *multiplicative*. However, this doesn't work for $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Motivating example: $\begin{pmatrix} 0 & 0 & \cdots & 0 & a_n \\ a_1 & & & & 0 \\ & \ddots & & & \vdots \\ 0 & & & a_{n-1} & 0 \end{pmatrix} = M$. Starting with v_1, we compute $(v_1, Mv_1, M^2 v_1, \ldots, M^{n-1} v_1) = (v_1, a_1 v_2, a_1 a_2 v_3, \ldots, a_1 \cdots a_{n-1} v_n)$, so that the l.i. of this list $\Rightarrow$ deg min poly $\geq n$. This forces char $=$ min $= z^n - a_1 \cdots a_n$.

So, the determinant seems to care about <u>all</u> diagonals, not just the main one. We'll take this a step further:

<u>Def</u>: M an $n \times n$-matrix, $\det M = \sum_{(m_1,\dots,m_n) \in \operatorname{perm} \underline{n}} \operatorname{sign}(m_1,\dots,m_n) M_{m_1,1} \cdots M_{m_n,n}$, where $\operatorname{sign}(m_1,\dots,m_n)$ is -1 raised to the # of disorders in $(m_1,\dots,m_n)$.

<u>Ex</u>: $\det(a_{11}) = a_{11}$. $\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \overset{1<2}{a_{11}a_{22}} - \overset{2>1}{a_{21}a_{12}}$.

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \overset{1<2<3}{a_{11}a_{22}a_{33}} - \overset{2>1<3}{a_{21}a_{12}a_{33}} + \overset{3>2>1}{a_{31}a_{22}a_{13}} - \cdots \leftarrow 3 \text{ more terms.}$$

<u>Cor</u>: Interchanging two columns reverses the sign of det. Hence, if two columns are equal, $\det = 0$. □

<u>Lem</u>: For column vectors $A_{\cdot 1}, \dots, A_{\cdot (j-1)}, A_{\cdot (j+1)}, \dots, A_{\cdot n}$, the map

$$\left(A_{\cdot j} \in k^n\right) \longmapsto \det\left(A_{\cdot 1} \middle| A_{\cdot 2} \middle| \cdots \middle| A_{\cdot n}\right) \text{ is } \underline{\text{linear}}.$$

(That is, the determinant is a "multilinear, alternating" map.) □

<u>Cor</u>: $\det(AB) = \det(AB_{\cdot 1} | AB_{\cdot 2} | \cdots | AB_{\cdot n})$ (det is <u>multiplicative</u>.)

$$= \det\left(A \sum_{m_1=1}^{n} B_{m_1,1} \cdot e_{m_1} \middle| \cdots \middle| A \sum_{m_n=1}^{n} B_{m_n,n}\, e_{m_n}\right)$$

$$= \sum_{m_1} \cdots \sum_{m_n} B_{m_1,1} \cdots B_{m_n,n} \cdot \det\left(Ae_{m_1} | \cdots | Ae_{m_n}\right)$$

$$= \sum_{(m_1,\dots,m_n) \in \operatorname{perm} \underline{n}} \text{———}/\!/\text{———}.$$

$$= \sum_{/\!/} \operatorname{sign}(m_1,\dots,m_n) \det A \cdot B_{m_1,1} \cdots B_{m_n,n} = \det A \cdot \det B. \quad □$$

<u>Cor</u>: det of a matrix is invariant under change of basis, and the two notions of det agree. □

Determinant and Volume (10.B):

Today we will investigate an important geometric aspect of determinants: their connection to volumetric properties of linear maps.

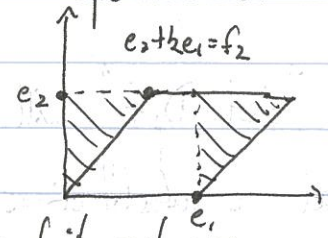
Here's the slogan for today:

Thm: For $f: V \to V$ a real operator on a f.d. inner product space, $\det(f)$ computes the volume of a unit cube imaged by f.

We will prove this in two ways, according to the two ways we have developed to present linear operators:

"Pf" using Gaussian elimination: We showed in a sequence of homework exercises that f can be expressed as a sequence of row and column operations applied to the identity matrix. So, we can compute $\det(f)$ by understanding the determinants of these operations.

- Scale-and-add:
  These all have determinant 1.
  They translate the endpoints of the parallelopiped, + thus does not disturb its volume.



- Scale: These have determinant the scalar. They scale one axis of the parallelopiped, and thus scales its volume by the scalar.
- Swap: These have determinant -1 using the matrix formula.

These observations collect to give a description in terms of G. Elimination:

$$\text{identity matrix} \overset{\text{row ops}}{\rightsquigarrow} \cdots \overset{\text{row ops}}{\rightsquigarrow} M$$

$$\text{unit volume } 1 \overset{\text{row ops}}{\rightsquigarrow} \cdots \overset{\text{row ops}}{\rightsquigarrow} \text{volume of the parallelopiped determined by } M.$$

□

Pf using Polar Decomposition: Every f factors as $f = g \circ \sqrt{f^*f}$ for some isometry g. First, note that $|\det g| = 1$, since the only eigenvalues of g satisfy $|\lambda| = 1$. Second, we know that the positive operator $\sqrt{f^*f}$ admits orthonormal expression as a diagonal matrix —— whose behavior on a unit cube is easy to understand. Hence, $\det f = \det g \circ \sqrt{f^*f}| = \det \sqrt{f^*f}$ = vol. of unit cube under f. □

Our main application of this will arise next semester.

$$\begin{tikzcd} M \arrow[r, "f"] & M' \\ p \arrow[u, phantom, sloped, "\in"] \arrow[r, mapsto] & f(p) \arrow[u, phantom, sloped, "\in"] \end{tikzcd}$$

a smooth map $\rightsquigarrow D_pf : T_pM \longrightarrow T_{f(p)}M'$.

derivative of f at p

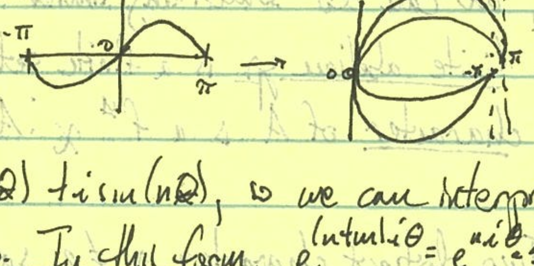
space of tangent directions to M at p.

The object "$\det D_pf$" will play a role analogous to $u'(t)$ in the classical u-substitution formula $\int_{u(\alpha)}^{u(\beta)} f(u)\,du = \int_{\alpha}^{\beta} f(u(t)) \cdot u'(t)\,dt$.

Finite Fourier Analysis

You proved a bunch of results about $\sin nx + \cos nx$ on your hwk, considered as $f^{\underline{ns}}$ $[-\pi,\pi] \longrightarrow \mathbb{R}$. A summary of Fourier analysis is:

① These can be interpreted as $f^{\underline{ns}}$ on the circle by gluing $-\pi$ to π:

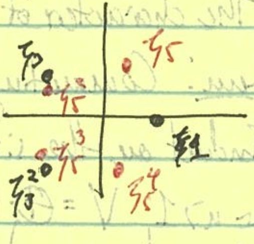


② Early on, we used $e^{ni\theta} = \cos(n\theta) + i\sin(n\theta)$, so we can interpret these results about $\mathbb{C}$-valued $f^{\underline{ns}}$ too. In this form, $e^{(n+m)i\theta} = e^{ni\theta} \cdot e^{mi\theta}$.

③ The main Thm: the subspace spanned by $\{e^{in\theta}\}$ is dense in all $f^{\underline{ns}}$, meaning any $f^{\underline{n}}$ can be approximated arbitrarily well using these sums.

This last theorem is beyond our reach. Today we will prove some finite analogues of it, beginning with the following:

Def: Let μ_n be the set of $n^{\underline{th}}$ roots of 1 in $\mathbb{C}$,
i.e., $\mu_n = \{e^{2\pi i k/n} \in \mathbb{C} \mid 0 \leq k < n\}$,
and let V_n be the set of $f^{\underline{ns}}$ $\{\mu_n \longrightarrow \mathbb{C}\}$.



We would like analogues of the special $f^{\underline{ns}}$ $e^{in\theta}$ from above, whose main property seems to be $e^{in\theta} \cdot e^{im\theta} = e^{i(n+m)\theta}$.

Def: Let $e_\ell \in V_n$ be the function $e_\ell(\zeta^k) = \zeta^{k\ell}$.

These satisfy $e_{n+m}(\zeta^k) = \zeta^{k(n+m)} = \zeta^{kn} \cdot \zeta^{km} = e_n(\zeta^k)\, e_m(\zeta^k)$.

Lem: Under the inner product $\langle f,g \rangle = \sum_{\zeta^k \in \mu_n} f(\zeta^k)\overline{g(\zeta^k)}$, the e_ℓ are orthogonal.

Pf: $\langle e_i, e_j \rangle = \sum_{\zeta^k \in \mu_n} \zeta^{(i+j)k} = 0$, the sum of a full set of roots of unity. □

Cor: These form a basis, as they're of the right length.

Def: Given $f: \mu_n \to \mathbb{C}$, its Fourier transform $\hat{f}$ is
$$\hat{f}(\zeta^m) = \frac{1}{n} \cdot \sum_{\zeta^k \in \mu_n} f(\zeta^k) \cdot \zeta^{-mk} = \frac{1}{n} \cdot \langle f, e_m \rangle.$$

Cor: Fourier inversion states $f(\zeta^k) = \sum_{\zeta^m \in \mu_n} \hat{f}(\zeta^m) \cdot \zeta^{mk}$.

Pf: $f = \sum_{\zeta^m} \hat{f}(\zeta^m) e_m$, so $f(\zeta^k) = \sum_{\zeta^m} \hat{f}(\zeta^m)\cdot\left(\zeta^{mk}\right)$. Just evaluate. □

In fact, we can do something similar for $\mathbb{C}$-valued fns on any finite abelian gp.

<u>Def</u>: A <u>finite abelian gp</u> is a finite set A equipped w/ a comm., unital sum w/ inverses.

<u>Def</u>: A <u>character</u> of A is a f^{n} $\chi : A \to \mathbb{C}$ satisfying $\chi(a+b) = \chi(a)\chi(b)$.

<u>Lem</u>: Two distinct characters $\chi \neq \rho$ satisfy $\langle \chi, \rho\rangle = 0$.

Pf: Recall $\langle \chi, \rho\rangle = \frac{1}{|A|}\sum_{a\in A}\chi(a)\overline{\rho(a)}$, $= \frac{1}{|A|}\sum_{a\in A}\chi(a)\cdot\rho^{-1}(a) = \frac{1}{|A|}\sum_{a\in A}(\chi\cdot\rho^{-1})(a)$.

We will show this sum is zero for any $\chi \neq \rho$, so that $\chi\rho^{-1} = \eta \neq 0$.

Choose a $b \in A$ with $\eta(b) \neq 1$. Then $\eta(b)\sum_a \eta(a) = \sum_a \eta(a+b) = \sum_a \eta(a)$, so $=0$. □

<u>Thm</u>: The characters of A form a basis for $V_A = \{A \to \mathbb{C}\}$.

<u>Lem</u>: Commuting families of unitary transformations are simultaneously diag^{ble}.

Pf: Induct on the size n of the family $(f_1, \ldots, f_n)$. $n=1$ is the spectral theorem.

For $n>1$, $V = \bigoplus_j E(\lambda_j, f_n)$. On each eigenspace, we have $f_n f_i(v_j) = f_i f_n(v_j) = f_i(\lambda_j v_j) = \lambda_j f_i(v_j)$, so $f_i(v_j) \in E(\lambda_j, f_n)$. On $E(\lambda_j, f_n)$, <u>any</u> basis ~~for~~ presents $f_n|_{E(\lambda_j, f_n)}$ as $\lambda_j \cdot \mathrm{Id}$, which commutes w/ everything. □

<u>Pf of Thm</u>: Set $T_a : V_A \to V_A$ by $(T_a f)(x) = f(a+x)$. These commute, so diag^{ize} them, in a basis $(v_b) \in V_A$. Pick any such v, — then $v(1) \neq 0$, since otherwise $v(a) = (T_a v)(1) = \lambda_a v(1)$, but $\lambda_a \neq 0$. Define $\omega(x) = \lambda_x = v(x)/v(1)$.

We claim ω is a character: $\omega(a+b) = (T_a\omega)b = \lambda_a\omega(b) = \lambda_a\lambda_b = \omega(a)\omega(b)$.

Since there are $|G|$ many v giving rise to $|G|$ many $\perp$ ω's, we are done. □

<u>Cor</u>: Set $\hat{f}(e) = \frac{1}{|A|}\sum_{a\in A} f(a)\overline{e(a)}$, for $e : A \to \mathbb{C}^\times$ a character. Then

$$f = \sum_{\substack{\text{characters} \\ e}} \hat{f}(e)\cdot e,$$

the Fourier inversion formula.

The Fast Fourier Transform + Complexity

Classical problem in complexity: sorting an unsorted list.

"Insertion sort": Form a new list by inserting old list elements one-by-one into a new sorted list.

$(1, \dots, n) \rightsquigarrow (2, \dots, n) + (1)$

Q: How long does this take? ~~$(n, n-1, \dots, 1)$~~ $\rightsquigarrow (3, \dots, n) + (1, 2)$ 1 swap

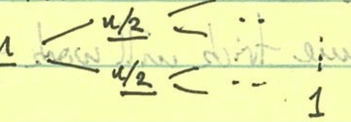
$\vdots$

$1 + \dots + (n-1) = n(n-1) \cdot \frac{1}{2}$. $(1) + (1, 2, \dots, n-1, n)$ n-1 sw

Merge sort: Take a list, divide it into 2 halves, mergesort them, union them.

Q: How long does this take? Call it $f(n)$. Then $f(n) = 2f(n/2) + n$.

Picture this like:



$n + n + \dots + n = n \cdot \lg_2(n)$. So, mergesort is faster.

Another problem: multiplication. $\begin{array}{r} 2642 \\ \times 5821 \\ \hline \end{array}$ $= 15{,}379{,}082$ requires 16 ·'s, and some additions. In general, this algorithm takes $\approx n^2 + 2n$ steps.

Today we will use the material from last time to improve this...

Observation 1: Multiplication of integers is close to multiplication of polynomials — just with a carrying step. Namely, setting ~~$p = 2 + 4x + 6x^2 + 2x^3$~~ $p = 2x^3 + 6x^2 + 4x + 2$ and $q = 5x^3 + 8x^2 + 2x + 1$, we have $p(10) = 2642$, $q(10) = 5821$, and $(p \cdot q)(10) =$ the product.

Observation 2: Polynomials of degree $n-1$ are determined by their values on any n points. In particular, picking the points μ_n from last time, $P_{n-1} \xrightarrow{a} \mu_n$ is a linear isomorphism. It's also multiplicative!: $p(\zeta^k) q(\zeta^k) = (p \cdot q)(\zeta^k)$. So, if we want to multiply two poly's, we just multiply their points.

Observation 3: The map a is a form of the Fourier transform:

$$a(p)(\zeta^k) = p(\zeta^k) = \sum_{j=0}^{n-1} p_j \zeta^{jk} = (p_j, e_{-k}) = \hat{p}(k).$$

Observation 4: Computing $\hat{p}$ is more efficient than you might think.

$$\hat{p}(k) = \sum_{j=0}^{n-1} p_j \zeta^{jk} = \sum_{j=0}^{\frac{n}{2}-1} p_{2j}(\zeta^{2k})^j + \zeta \sum_{j=0}^{\frac{n}{2}-1} p_{2j+1}(\zeta^{2k})^j$$ split into even and odd parts, which reappear when computing different k. This organizes into a scheme: $p_{k|-} = p_k$, $p_{*|s} = p_{0*|s'} + \zeta^{2^{\ell(*)}\cdot s} p_{1*|s'}$; $p_{-|k} = \hat{p}(k)$. This organizes into $n \lg(n)$ ~~operations~~ multiplications.

Observation 5: The Fourier inversion formula is so similar to the formula for $\hat{p}(k)$ that the same trick will work for it.

Ex: Take $p(x) = 2x^3 + 6x^2 + 4x + 2$ and $q(x) = 5x^3 + 8x^2 + 2x + 1$.

	000	001	010	011	100	101	110	111
$p_{***\|}$	2	4	6	2	0	0	0	0
$p_{**\|*}$	2	2	4	4	6	6	2	2
$p_{*\|**}$	8	$2+6i$	-4	$2-6i$	6	$4+2i$	2	$4-2i$
$p_{\|***}$	14	$2+6i+\zeta(4+2i)$	$-4+2i$	$2-6i-\bar{\zeta}(4-2i)$	2	$2+6i-\zeta(4+2i)$	$-4-2i$	$2-6i+\bar{\zeta}(4+2i)$
$q_{\|***}$	16	$1+8i+\zeta(2+5i)$	$-7-3i$	$1-8i-\bar{\zeta}(2-5i)$	2	$1+8i-\zeta(2+5i)$	$-7+3i$	$1-8i+\bar{\zeta}(2-5i)$
$pq_{\|***}$	224	$-70+20i+\zeta(-38+56)$	$34-2i$	$-70-20i+\bar{\zeta}(38+56i)$	4	$-70+20i+\zeta(38-56i)$	$34+2i$	$-70-20i+\bar{\zeta}(-38-56i)$
$pq_{***\|}$	2	8	30	56	72	46	10	0

$\rightsquigarrow$ 15379082.

Observation 6: In a computer implementation, we can work mod $2^N \pm 1$, so that 2 is an N^{th} or $2N^{th}$ root of unity ($= \zeta$), so that mult. by ζ is also fast. We can also recurse on the "$pq_{|***}$" step. In all, this runs in $n \lg n \cdot \lg\lg n$ time.

Dirichlet's theorem

At the start of this class, we proved an ancient theorem:

Thm: There are infinitely many prime numbers. □

It is easy to ask for more information than this. For instance,

Q: Are there ∞ly many primes $\equiv 1 \bmod 4$? $\equiv 3 \bmod 4$?

Pf of $\equiv 3 \bmod 4$: Assume there are finitely many, and let $\{3, p_1, \ldots, p_n\}$ be an enumeration. Set $N = 4p_1 \cdots p_n + 3$. Since two primes $\equiv 1 \bmod 4$ multiply to $\equiv 1 \bmod 4$, there must be a prime $\equiv 3 \bmod 4$ dividing N. Can't be 3 or p_j for any j. □

There is no known elementary proof of the other case. There is an analytic proof, which today we will describe. The jumping-off point is:

Thm (Euler): There is a factorization $\zeta(s) = \sum_n \frac{1}{n^s} = \prod_p (1 - p^{-s})^{-1}$.

Pf idea: $\frac{1}{1-p^{-s}}$ expands as $1 + \frac{1}{p^s} + \frac{1}{(p^2)^s} + \cdots$. Each term $\frac{1}{n^s}$ corresponds to exactly one product term by factorization into primes. □

Thm. The series $\sum_p \frac{1}{p}$ diverges. (Note this $\Rightarrow$ ancient theorem.)

Pf: $\log \zeta(s) = \log \sum_n \frac{1}{n^s} = \log \prod_p (1-p^{-s})^{-1} = -\sum_p \log(1-p^{-s})$. The Taylor formula for log gives $-\sum_p \log(1-p^{-s}) = -\sum_p \left(\frac{-1}{p^s} + O(1/p^{2s})\right)$ $= \sum_p 1/p^s + O(1)$. Finally, let $s \longrightarrow 1^+$. □

It turns out that this is the style of argument that generalizes to handle the case $p \equiv 1 \bmod 4$, — and the modification is through finite Fourier analysis. Consider the f^n $\chi : (\mathbb{Z}/4)^\times \longrightarrow \mathbb{C}$ defined by [illegible] $\chi(1) = 1$, $\chi(-1) = -1$. This "extends" to all of $\mathbb{Z}$ by $\chi(n) = \begin{cases} 0 & \text{if } n \text{ even}, \\ 1 & \text{if } n \equiv 1 \bmod 4, \\ -1 & \text{if } n \equiv 3 \bmod 4. \end{cases}$ Define $L_\chi(s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = 1 - \frac{3}{n^s} + \frac{5}{n^s} - \frac{7}{n^s} + \cdots$, and $L_\chi(1) = \pi/4$. The same "pf idea" gives $L_\chi(s) = \prod_p \frac{1}{1 - \frac{\chi(p)}{p^s}}$.

Taking logs gives $\log L_\chi(s) = \sum_p \chi(p)\cdot p^{-s} + O(1)$, and the fact that $L_\chi(s) \longrightarrow \pi/4 \neq 0$ or ∞ as $s \longrightarrow 1^+$ means $\sum_p \chi(p)\cdot p^{-s}$ is convergent as $s \longrightarrow 1^+$. We break it into pieces:

$$\sum_p \chi(p)\cdot p^{-s} = \sum_{p\equiv 1 \bmod 4} \frac{1}{p^s} + \sum_{p\equiv 3 \bmod 4} \frac{-1}{p^s}.$$ We know $\sum_p \frac{1}{p^s} \longrightarrow \infty$ as $s \longrightarrow 1^+$,

so adding these gives $2\cdot \sum_{p\equiv 1} \frac{1}{p^s} \longrightarrow \infty$ as $s \longrightarrow 1^+$. □

The general Theorem is:

<u>Thm</u> (Dirichlet): For ℓ and q coprime, there exist ∞^{ly} many primes of the form $p = \ell + q\cdot k$, $k \in \mathbb{Z}$.

We're not going to prove this, but it feels a lot like the proof just given. The main point is that $\delta_\ell(n) = \begin{cases} 1 & \text{if } n \equiv \ell \bmod q \\ 0 & \text{o/w} \end{cases}$ admits a finite Fourier expansion in terms of characters $\chi : (\mathbb{Z}/q)^* \longrightarrow \mathbb{Z}$, and each nontrivial such χ gives rise to a cts function $L_\chi(s) \longrightarrow \neq 0, \neq \infty$ for $s \longrightarrow 1^+$. Once you've made it this far, you can mimic the rest of the proof above. The real meat is in the convergence of $L_\chi(1)$; we could manually calculate it, but in general this is <u>not</u> possible.