1.

Math 403 Spring 2025

Advanced Linear Algebra

Tue / Thu 11:45 - 13:00

Physics 205

Office Hours: Tue 13:00-14:00 right after class } outside if possible
Thu 13:00-14:00 Physics 209 if not
Zoom if necessary

Safety:
- ~~masks, hand wash/sanitize~~ politely point out noncompliance — even me!
- ~~distance if possible~~
- know where the exits are from the room and the building

Policies
- covered on Tue ⇒ fair game for HW due Sat
- collaboration / academic honesty
  - Yes on HW write your solutions yourself
  - No on exams I have • brought numerous cases to the Office of Student Conduct
    • never lost

Index cards

1. Ezra Miller
2. he / him
3. 46th grade
4. Major or potential major: Math, Music
5. What you hope to get out of this course

   students who know how to use linear algebra

6. The most important thing you've learned about how you learn

   not to take notes!

7. Hobbies: frisbee, gardening, photography, beer
8. Something unique about yourself

✗ "I'm from MA"

✓ "I'm from HI —
but I'm allergic to pineapple!"

Fields

Def: A group is a set G with an $\underbrace{\text{associative}}$ binary operation

$$*: G\times G \to G \qquad (g*g')*g'' = g*(g'*g'')$$
$$(g,g')\mapsto g*g'$$

that has
- an identity e with $e*g = g*e = g \ \forall\, g\in G$
- an inverse g^{-1} for each $g\in G$, so $g^{-1}g=e$.

G is abelian if $*$ is commutative: $g*h = h*g \ \forall\, g,h\in G$.

E.g. $(\mathbb{R},+)$ $\qquad$ $(\mathbb{R}\setminus\{0\},\cdot)$ $\qquad$ $C^0(\mathbb{R}^n\to\mathbb{R},+)$

$\mathbb{C}$ $\qquad$ $\mathbb{C}$ $\qquad$ $\mathrm{Fun}(S\to A,+)$ $\quad$ A = any abelian group!

$\mathbb{Q}$ $\qquad$ $\mathbb{Q}$

$\mathbb{Z}$ $\qquad$ $\cancel{\mathbb{Z}}$

$\left(\text{m}\times\text{n matrices with any of these coefficients}, +\right)$ $\qquad$ $\cancel{\left(\text{m}\times\text{n matrices with any of these coefficients}, \cdot\right)}$ $\qquad$ non-abelian: $\{A\in\mathbb{R}^{2\times2} \mid \det A = 1\}$

Def: A field is an abelian group $(F,+)$ with additive identity $0\in F$ such that
- $F^* = F\setminus\{0\}$ is an abelian group $(F^*,\cdot)$ and
- multiplication $\cdot$ distributes over addition $+$: $a\cdot(b+c) = a\cdot b + a\cdot c$.

E.g. $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{F}_2=\{0,1\}$, $\mathbb{F}_3=\{-1,0,1\}$, $\mathbb{F}_p=\{0,1,\dots,p-1\}$ for $p\in\mathbb{Z}$ prime

$\quad\ \|$

$\mathbb{R}(i)$, $\mathbb{Q}(i)$

Math 221 works verbatim with any F in place of $\mathbb{R}$,
except for notions of length, angle, order $(a<b)$

$\Downarrow$
closeness (topology)

Def: A vector space over F is ... review from 221.

$(V,+)$ abelian group with an action of F:

$$F\times V\to V \qquad (\alpha,v)\mapsto \alpha v$$

- distributes over $+$ on both sides
- associative: $\alpha(\beta v) = (\alpha\beta)v$
- $1v = v \ \forall\, v\in V$.

Def: A homomorphism of vector spaces is a linear map.

E.g. $\mathcal{B}$ is a basis for V

$$\Leftrightarrow \{\text{functions } \mathcal{B}\xrightarrow{f} W\} \overset{\leftrightarrow}{\cancel{\Rightarrow}} \cancel{\exists !}\{\text{homomorphisms } \varphi: V\to W \text{ with } \varphi|_{\mathcal{B}} = f\}$$

Def: A homomorphism $\varphi: V \to W$ has

- kernel $\ker \varphi = \{ v \in V \mid \varphi(v) = 0 \} \subseteq V$
- image $\operatorname{im} \varphi = \{ \varphi(v) \mid v \in V \} \subseteq W$.

subspaces

Thm (rank-nullity): $\dim(\ker\varphi) + \dim(\operatorname{im}\varphi) = \dim V$.

rank

nullity

Pf: Pick $\mathcal{B}' =$ basis of $\ker\varphi$

$\mathcal{B}'' \subseteq V$ with $\mathcal{B}'' \hookrightarrow \varphi(\mathcal{B}'') =$ basis of $\operatorname{im}\varphi$.

Then $\mathcal{B} = \mathcal{B}' \cup \mathcal{B}''$ is a basis of V. □

$\# = \# + \#$

requires proof, but it's a straightforward exercise:

$\sum_{v \in \mathcal{B}'} \alpha_v v + \sum_{w \in \mathcal{B}''} \beta_w w = 0$

$\Rightarrow \sum_{w \in \mathcal{B}''} \beta_w \varphi(w) = 0$ since $\varphi(\mathcal{B}') = 0$

$\Rightarrow \beta_w = 0 \ \forall w \in \mathcal{B}''$ since $\varphi(\mathcal{B}'')$ indep.

$\Rightarrow \alpha_v = 0 \ \forall v \in \mathcal{B}'$ " $\mathcal{B}''$ ".

But $v \in V \Rightarrow \varphi(v) \in \operatorname{span}(\varphi(\mathcal{B}''))$, say $\varphi(v) = \varphi(b)$ with $b \in \operatorname{span}(\mathcal{B}'')$. Then $v - b \in \ker\varphi = \operatorname{span}(\mathcal{B}')$, so $v \in b + \operatorname{span}(\mathcal{B}') \subseteq \operatorname{span}(\mathcal{B}'' \cup \mathcal{B}')$.

Def: A homomorphism $\varphi: V \to W$ is an isomorphism if φ is (injective and surjective) = bijective

$\Leftrightarrow \ker\varphi = 0 \qquad \operatorname{im}\varphi = W$

E.g. $\dim V = n \Rightarrow V \cong F^n$.

Pf: Basis $v_1, \dots, v_n$ of $V \Rightarrow v_i \mapsto e_i$ induces $\cong$: $\ker = 0$ because $e_1, \dots, e_n$ indep, and $\operatorname{im} = F^n$ because $e_1, \dots, e_n$ span F^n. □

E.g. $\operatorname{sols}_{\mathbb{R}}(f'' + f = 0) \cong \mathbb{R}^2 = \operatorname{span}(\sin, \cos)$ $\updownarrow$?

$\qquad\ \ \mathbb{C} \qquad\qquad\qquad\quad \mathbb{C}^2 = \operatorname{span}(e^{ix}, e^{-ix})$

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i}, \quad \cos x = \frac{e^{ix} + e^{-ix}}{2}$$

Def: An affine subspace is a translate of a subspace = sols(inhomogeneous linear system)

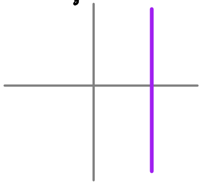
E.g. Set™ has one rule: find an affine line in $\mathbb{F}_3^4$.

card $= u \in \mathbb{F}_3^4$

$= (a, b, c, d)$. $\quad v = (a', b', c', d') \Rightarrow$ affine line $\overline{uv}$ is

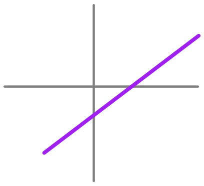
$$L = v + \{ \lambda(u - v) \mid \lambda \in \mathbb{F}_3 \} = \{ v, u, -u-v \}.$$

Two possibilities:



(i) $L \subseteq x_1 = a$ (if $a' = a$):

(ii) x_1-coordinates of all points in L are distinct (if $a' \neq a$)



Hence the rule: "If two are and one isn't then it's not a set."

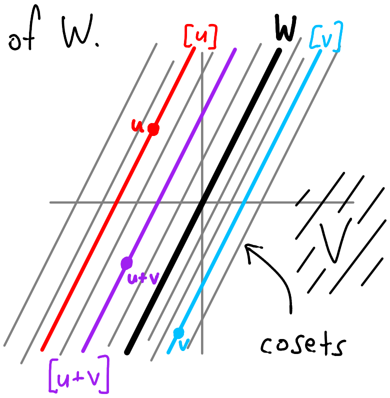
Q. What makes $\mathbb{F}_3$ so special?

A. Each pair of points (cards) yields a unique third in L.

2.

Def: Fix subspace $W \subseteq V$.

1. A coset of W is an affine subspace $[v] = v + W$ for some $v \in V$.
2. The quotient V/W is the set of cosets of W.
   "V mod W"



Prop: V/W is a vector space with

$$[u] + [v] = [u+v] \quad \text{and} \quad \lambda[v] = [\lambda v].$$

Pf: $(u+W) + (v+W) \underset{\uparrow}{=} (u+v) + W.$

"add ~~lines~~ affine subspaces" since $W + W = W$

$$\lambda(v+W) = \lambda v + \lambda W = \lambda v + W \text{ if } \lambda \neq 0, \text{ and}$$

$$0[v] = 0(v+W) = \{0\} \subseteq W = [0] = [0v]. \quad \square$$

Cor: $\dim V = \dim W + \dim V/W$.

Pf: $V \to V/W$ is a homomorphism (by Prop) with $\ker = W$ and $\operatorname{im} = V/W$. $\square$

Universal property of quotients

A homomorphism $\underbrace{V \xrightarrow{\varphi} U \text{ is } 0 \text{ on } W}_{W \subseteq \ker \varphi} \Leftrightarrow \varphi$ factors through V/W:

$$\begin{tikzcd} V \arrow[r] \arrow[rr, bend left, "\varphi"] & V/W \arrow[r] & U. \end{tikzcd}$$

Pf: φ induces well defined function — forget algebraic properties, like "homomorphism"

$V/W \to U \Leftrightarrow$ each coset of $W \to$ single point in U

$\underset{\varphi \text{ linear}}{\Leftrightarrow} W \to$ single point in U.

$\Leftrightarrow W \subseteq \ker \varphi$. $\square$

Def: Arbitrary homomorphism $V \to W$ has
- kernel $K \subseteq V$
- image $I \subseteq W$
- cokernel $W \twoheadrightarrow W/I$
- coimage ~~??~~ $V \twoheadrightarrow V/K$

standard abuse of notation: "$W/I = \operatorname{coker}(V \to W)$."

$$K \overset{\ker}{\subseteq} V \overset{\text{coim}}{\twoheadrightarrow} V/K \cong I \overset{\text{im}}{\subseteq} W \overset{\text{coker}}{\twoheadrightarrow} W/I$$

First Isomorphism Thm (requires proof!)

Pf: $V \twoheadrightarrow I$ by def of im, so $V \twoheadrightarrow V/K \twoheadrightarrow I$ by universal property of coker.

$\ker(V/K \twoheadrightarrow I) = \{[v] \in V/K \mid v \mapsto 0\} = [K]$ by def of ker,

$= 0 \in V/K \Rightarrow V/K \hookrightarrow I$. $\square$

<u>Def</u>: A sequence $V_0 \xrightarrow{\varphi_1} V_1 \longrightarrow \cdots \xrightarrow{\varphi_r} V_r$ is <u>exact</u> if $\ker \varphi_{i+1} = \operatorname{im} \varphi_i \;\forall i$.

<u>E.g.</u> $0 \to V \to V' \to 0$ exact $\cancel{\Leftrightarrow ?}$ $V \cong V'$

$0 \to V \to V'$ exact $\cancel{\Leftrightarrow ?}$ $V \hookrightarrow V'$

$V \to V' \to 0$ exact $\cancel{\Leftrightarrow ?}$ $V \twoheadrightarrow V'$

$0 \to K \to V \to W \to W/I \to 0$ is exact

↖ exact here by FIT

$$\begin{tikzcd} 0 \arrow[r] & A \arrow[r] \arrow[d, phantom, "\parallel"] & B \arrow[r] \arrow[d, phantom, "\parallel"] & C \arrow[r] & 0 \\ 0 \arrow[r] & A \arrow[r] & B \arrow[r] & B/A \arrow[r] & 0 \end{tikzcd}$$

$\cancel{\Leftrightarrow ?}$ $A \subseteq B$ and $C \cong B/A$

<u>Def</u>: $V_\bullet$: $\cdots \longrightarrow V_{i-1} \longrightarrow V_i \longrightarrow V_{i+1} \longrightarrow \cdots$ is a <u>complex</u> if $V_{i-1} \to V_i \to V_{i+1}$ is 0 $\forall i$.

$\underbrace{\hphantom{V_{i-1} \to V_i \to V_{i+1} \text{ is } 0}}$

$V_\bullet$ has <u>homology</u> $H_i V_\bullet = \ker \varphi_{i+1} / \operatorname{im} \varphi_i$.

<u>Lemma</u>: $\Leftrightarrow$ $\operatorname{im}(V_{i-1} \xrightarrow{\varphi_i} V_i) \subseteq \ker(V_i \xrightarrow{\varphi_{i+1}} V_{i+1})$

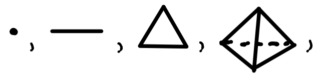
measures how far a complex is from being exact.

Problem: you don't know many complexes yet. (Do you?)

(poll)

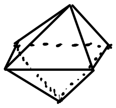
<u>E.g.</u> algebraic topology

<u>simplices</u>



, …

of dim 0, 1, 2, 3, …

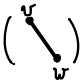
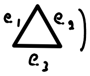
e.g. octahedron  ⇝ vector spaces $/\mathbb{F}_2$

V: basis = vertices

E: basis = edges

F: basis = faces

$C_\bullet$: $0 \xleftarrow{\partial_V} V \xleftarrow{\partial_E} E \xleftarrow{\partial_F} F \leftarrow 0$ <u>chain complex</u>

$\partial_E($  $) = v + w$ $\quad \partial_F($  $) = e_1 + e_2 + e_3$

<u>Prop</u>: $C_\bullet$ is a complex!

<u>Pf</u>: 2 (= 0) ways to get from a simplex of dim i to a simplex of dim $i-2$. □

Compute: $H_0 C_\bullet = \ker \partial_V / \operatorname{im} \partial_E = V/B$ $\quad B = \operatorname{span}(v+w \mid \text{vertices } v, w)$ $\quad \dim H_0 = \cancel{?}\; 1$

$H_1 C_\bullet = \ker \partial_E / \operatorname{im} \partial_F = \cancel{?}\; 0$ exercise (not assigned)

$H_2 C_\bullet = \ker \partial_F / 0 = \cancel{?}\; \operatorname{span}(f_1 + \cdots + f_8)$ $\quad \dim H = \cancel{?}\; 1$

<u>Thm</u> (rank-nullity): $\sum_i (-1)^i \dim H_i = \sum_i (-1)^i \dim V_i$.

$= $ <u>Euler characteristic</u> of $V_\bullet$

$0 \to K \to V \to I \to 0$

exact $\Rightarrow \dim K - \dim V + \dim I = 0$

$\Rightarrow \dim K + \dim I = \dim V$

<u>Cor</u>: exercise! : $6 - 12 + 8 = 1 - ? + 1 \Rightarrow ? = 0$

<u>Pf of Thm</u>: HW

3.

Duality

Def: Fix vector space V over F. The dual of V is

$$V^* = \{\text{homomorphisms } V \to F\}$$
$$= \mathrm{Hom}(V, F)$$

Lemma: V^* is a vector space$/F$.

Pf:
$$\left.\begin{matrix}\left.\begin{matrix}\ell \in V^* \\ \alpha \in F\end{matrix}\right\} \Rightarrow \alpha\ell : v \mapsto \alpha\ell(v) = \ell(\alpha v), \text{ so } \alpha\ell \in V^*. \\ \ell_1, \ell_2 \in V^* \Rightarrow \ell_1 + \ell_2 : v \mapsto \ell_1(v) + \ell_2(v), \text{ so } \ell_1 + \ell_2 \in V^*.\end{matrix}\right\} \Rightarrow \text{subspace of } \mathrm{Fun}(V,F). \quad \square$$

E.g. 1. $V = F^n_{col} \Rightarrow V^* = F^n_{row}$ $\quad [y_1 \cdots y_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = y_1x_1 + \cdots + y_nx_n \in F.$

2. $V = C^0([0,1]) \Rightarrow \int_0^1 \cdot\, dt \in V^* : f \mapsto \int_0^1 f(t)dt$ is linear.

$ev_{1/2} \in V^* : f \mapsto ev_{1/2} f = f(1/2)$ is linear by def.

3. $V = C^1(\mathbb{R}) \Rightarrow ev_{1/2} \circ \frac{d}{dt} \in V^* : f \mapsto f'(1/2)$ is linear.

Notation: write $\ell(v) = \langle \ell, v\rangle$. Why? Place V^* and V on equal footing, and:

Suppose
$$\begin{tikzcd} W \arrow[d, "\ell"'] & V \arrow[l, "\varphi"'] \arrow[dl, "\ell\circ\varphi"] \\ F & \end{tikzcd}$$
Then $\ell \in W^* \Rightarrow \ell \circ \varphi \in V^*$

So $\varphi^* : \ell \mapsto \ell\circ\varphi \Rightarrow \langle \varphi^*\ell, v\rangle = \langle \ell\circ\varphi, v\rangle$
$\quad W^* \to V^*$
$$= \langle \ell, \varphi v\rangle$$

E.g. 1. $\langle \overset{\in (F^m)^*}{y}, \overset{\in F^m}{\lambda_A(x)}\rangle = \langle y, Ax\rangle$

$= yAx$

$= \langle yA, x\rangle$

$= \langle \underset{\in (F^n)^*}{\rho_A(y)}, \underset{\in F^n}{x}\rangle \Rightarrow$

$F^m_{col} \xleftarrow{\lambda_A} F^n_{col}$ left multiplication by $A \in F^{m\times n}$ on F^n_{col}

$$\boxed{\lambda_A^* = \rho_A}$$

$F^m_{row} \xrightarrow{\rho_A} F^n_{row}$ right " " " " F^m_{row}

left " " A^T " F^m_{col}

right way $\quad$ usual way

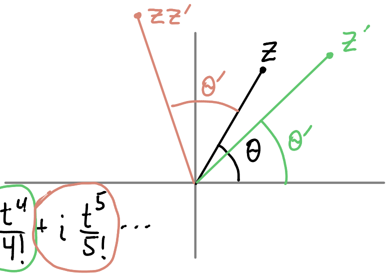
Note: also used notation $\langle u, v\rangle$ for inner products; compatible?

Let's review, but over $\mathbb{C}$, not just $\mathbb{R}$.

Geometry of $\mathbb{C}$ Who has seen it? (poll)

- $\mathbb{C} = \{a+bi \mid a,b \in \mathbb{R}\}$ vector space$/\mathbb{R}$ with basis $1, i$
- conjugate of $z = a+bi$ is $\bar{z} = a-bi$; conjugation $z \mapsto \bar{z}$ automorphism of $\mathbb{C}$
- $z\bar{z} = |z|^2 = a^2+b^2 = (a+bi)(a-bi)$
- $z = re^{i\theta}$, where $r = |z| =$ modulus of z

  $\theta =$ argument of z
- $zz' = (re^{i\theta})(r'e^{i\theta'}) = (rr')e^{i(\theta+\theta')}$

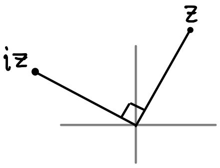


- $e^{i\theta} = \cos\theta + i\sin\theta$ Why?

1. Taylor series: $e^{it} = 1 + it - \frac{t^2}{2!} - i\frac{t^3}{3!} + \frac{t^4}{4!} + i\frac{t^5}{5!} \cdots$

   $= \cos t + i \sin t$

2. ODE for e^{it}: $\frac{d}{dt} e^{it} = ie^{it}$



You'll recognize this if you took Math 221 with me.

$z = a+bi \Rightarrow iz = -b+ai$

What curve has tangents $\perp$ vectors in $\mathbb{R}^2$ along curve? circle!

Which circle? $e^{i0} = e^0 = 1 = 1+0i \Rightarrow$ unit circle

E.g. $z = -1 \Rightarrow \left.\begin{array}{l} |z| = 1 \\ \arg z = \pi \end{array}\right\} \Rightarrow \boxed{e^{i\pi} = -1}$ Euler's formula (alt: $e^{i\pi}+1=0$)

Def: A (hermitian) inner product on a vector space $V/\underbrace{\mathbb{R} \text{ or } \mathbb{C}}_{\mathbb{F}}$ is a function $V\times V \to \mathbb{F}$, $(v,w) \mapsto \langle v,w\rangle$

that is

1. linear: $\langle \alpha x + \beta y, z\rangle = \alpha\langle x,z\rangle + \beta\langle y,z\rangle$
2. (conjugate) symmetric: $\langle x,y\rangle = \overline{\langle y,x\rangle}$ $\forall\, x,y \in V \Rightarrow \langle x, \alpha y + \beta z\rangle = \bar{\alpha}\langle x,y\rangle + \bar{\beta}\langle x,z\rangle$
3. nonnegative: $\langle x,x\rangle \in \mathbb{R}_{\geq 0}$ $\forall\, x \in V$
4. nondegenerate: $\langle x,x\rangle = 0 \Leftrightarrow x = 0$.

sesquilinear: linear in first argument, conjugate-linear in second argument;

bilinear if $\mathbb{F} = \mathbb{R}$ since then $\begin{array}{l}\bar{\alpha} = \alpha \\ \bar{\beta} = \beta\end{array}$.

E.g. $V = \mathbb{C}^n$ $\langle z,w\rangle = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n$

$= \bar{w}^T z$.

Def: For $A \in F^{m\times n}$, its adjoint is $A^* = \bar{A}^T$ (conjugate transpose).

Note: $\mathbb{F} = \mathbb{R} \Rightarrow A^* = A^T$.

inner product on W

Check: $W \xleftarrow{A} V \Rightarrow \langle w, Av\rangle = (Av)^* w$

$= v^*(A^*w)$ inner product on V

$= \langle A^*w, v\rangle$.

All still works: triangle inequality
Cauchy - Schwarz $|\langle x,y\rangle| \leq \|x\|\,\|y\|$
Gram-Schmidt
least squares
spectral thm
relations among fundamental subspaces: $\ker A = (\operatorname{im} A^*)^\perp, \ldots$

$$\Rightarrow \quad W \xleftarrow{A} V \quad \text{factors as} \quad W \underset{\text{injection}}{\hookleftarrow} \operatorname{im} A \underset{\text{isomorphism}}{\xleftarrow{\sim}} \operatorname{im} A^* \underset{\text{surjection}}{\overset{\text{orthogonal projection}}{\twoheadleftarrow}} V$$

Prop: Fix V with $\langle\cdot,\cdot\rangle$. If $\dim V < \infty$ then $v \mapsto \langle\cdot, v\rangle$ induces a conjugate-linear isomorphism $V \xrightarrow{\sim} V^*$: $\alpha v \mapsto \langle\cdot, \alpha v\rangle = \bar{\alpha}\langle\cdot, v\rangle$.

Pf: Any orthonormal basis $e_1, \ldots, e_n$ maps to the dual basis $e_1^*, \ldots, e_n^*$ of V^* such that $e_j^*(e_i) = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases} = \delta_{ij}$. □

More generally: nondegenerate bilinear forms
$\longleftrightarrow$ isomorphisms $V \to V^*$

$$\begin{tikzcd}
Ax \arrow[ddd, mapsto] & & & & x \arrow[llll, mapsto] \arrow[r, mapsto] & A^*Ax \\
& y \arrow[d, mapsto] & Y \arrow[d] & X \arrow[l, "\lambda_A"'] & \varphi^* & \\
& \bar{y}^T & Y^* \arrow[r, "\rho_A"'] & X^* \arrow[u] & \varphi \arrow[u, mapsto] & \\
x^*A^* \arrow[rrrrr, mapsto] & & & & & x^*A^*A \arrow[uuu, mapsto]
\end{tikzcd}$$

Jordan form

REVIEW

Fix
- vector space V/field F, $\dim V = n < \infty$
- $\varphi : V \longrightarrow V$ linear with matrix $A = [\varphi]_{\mathcal{E}} \in F^{n \times n}$ with respect to basis $\mathcal{E} = e_1, \ldots, e_n$ of V
- characteristic polynomial $p_\varphi(t) = \det(\varphi - tI) = \det(A - tI)$

  λ eigenvalue $\Leftrightarrow$ λ root of p_φ $\Leftrightarrow$ $p_\varphi(\lambda) = 0$
- geometric multiplicity $g(\lambda) = \dim \ker(\varphi - \lambda I)$

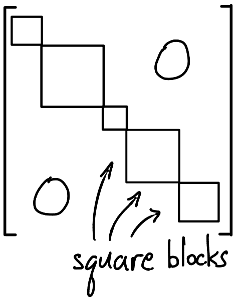
  $= \#$ linearly independendent eigenvectors with eigenvalue λ
- algebraic multiplicity $a(\lambda) = \#$ times $(t-\lambda)$ divides $p_\varphi(t)$

  $g(\lambda) \leqslant a(\lambda) \ \forall \lambda$

  φ diagonalizable $\Leftrightarrow$ $g(\lambda) = a(\lambda) \ \forall \lambda$

Thm (Jordan form) or "... canonical..." or "...normal..."

Assume all $\lambda \in F$. V has a basis $\mathcal{B}$ such that $[\varphi]_{\mathcal{B}}$ is block diagonal



with each block being a Jordan block

$$\begin{array}{c c} & d \\ d & \begin{bmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & \ddots & & \\ & & & \ddots & 1 & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{bmatrix} \end{array} \quad \begin{array}{l} g(\lambda) = 1 \\ a(\lambda) = d \end{array}$$

of some size d for some eigenvalue λ.

The multiset of blocks — i.e., the d's and λ's — depends only on φ, not on $\mathcal{B}$.

4. Today: what does Jordan form mean, and how does it lead to proof?

Def: F is algebraically closed if every polynomial with coeffs in F has a root in F.

$\Rightarrow p(t)=\alpha(t-\alpha_1)^{n_1}\cdots(t-\alpha_r)^{n_r} \Rightarrow$ hypothesis of JF thm. How to guarantee all $\lambda\in F$

Fundamental Thm of Algebra: $\mathbb{C}$ is algebraically closed.

What block diagonal means:

Def (direct sum): $V=V_1\oplus V_2$ means $V=V_1+V_2$ and $V_1\cap V_2=0$.

WARNING: does not mean $V_2=V_1^\perp$, although that suffices with $\langle\cdot,\cdot\rangle$

Prop: $\Leftrightarrow$ V has basis $\mathcal{B}=\underset{\text{partition}}{\mathcal{B}_1\sqcup\mathcal{B}_2}$ with $V_i=\operatorname{span}\mathcal{B}_i$ for $i=1,2$.

E.g. $F^n=F^m\oplus F^{n-m}=\operatorname{span}(e_1,\dots,e_m)\oplus\operatorname{span}(e_{m+1},\dots,e_n)$.

Def: $V=V_1\oplus\cdots\oplus V_r$ if $V=V_1+\cdots+V_r$ and $V_i\cap\sum_{j\neq i}V_j=0\ \forall i$.

Prop: $\Leftrightarrow$ V has basis $\mathcal{B}=\underset{\text{partition}}{\mathcal{B}_1\sqcup\cdots\sqcup\mathcal{B}_2}$ with $V_i=\operatorname{span}\mathcal{B}_i\ \forall i$.

E.g. V has basis $v_1,\dots,v_n \Leftrightarrow V=\operatorname{span}(v_1)\oplus\cdots\oplus\operatorname{span}(v_n)$.

Prop: φ block diagonal $\Leftrightarrow V=V_1\oplus\cdots\oplus V_r$ with $\underbrace{\varphi(V_i)\subseteq V_i}$ $\forall i$.

What Jordan blocks mean; needs: Def: V_i is φ-invariant

Cayley-Hamilton Thm: $p_\varphi(\varphi)=0$.

entries are elements of V

Pf: Fix basis $e_1,\dots,e_n$ of V, so $\varphi e_i=a_{i1}e_1+\cdots+a_{in}e_n=A_i\begin{bmatrix}e_1\\ \vdots\\ e_n\end{bmatrix}\Rightarrow\begin{bmatrix}\varphi e_1\\ \vdots\\ \varphi e_n\end{bmatrix}=A\begin{bmatrix}e_1\\ \vdots\\ e_n\end{bmatrix}$. (∗)

By def, $p_\varphi(t)=\det(A-tI)$. Need $p_\varphi(\varphi)e_i=0\ \forall i$.

Equivalently, $\underbrace{\begin{bmatrix}p_\varphi(t)&&\\&\ddots&\\&&p_\varphi(t)\end{bmatrix}}_{\det(A-tI)I\,=\,C^T(A-tI)}\begin{bmatrix}e_1\\ \vdots\\ e_n\end{bmatrix}=\begin{bmatrix}p_\varphi(t)\,e_1\\ \vdots\\ p_\varphi(t)\,e_n\end{bmatrix}$ becomes $\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$ when evaluated at $t=\varphi$.

where $C=$ cofactor matrix of $A-tI$.

But $(A-\varphi I)\begin{bmatrix}e_1\\ \vdots\\ e_n\end{bmatrix}=A\begin{bmatrix}e_1\\ \vdots\\ e_n\end{bmatrix}-\begin{bmatrix}\varphi e_1\\ \vdots\\ \varphi e_n\end{bmatrix}=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$ by (∗). Now multiply by $C^T|_{t=\varphi}$ on the left. □

Def: A minimal polynomial of φ (or A) is a monic $\overset{m\neq 0}{}$ polynomial $m(t)$ of minimal degree satisfying $m(\varphi)=0$ (or $m(A)=0$).

Prop: 1. $\exists!\ m(t)$.

2. $f(\varphi)=0\Rightarrow m\mid f$.

subtract monomial multiples of m from f to cancel leading terms recursively until you can't anymore

Pf: 1. follows from 2: $m\mid f$ and $\deg m=\deg f\Rightarrow f=\alpha m\overset{\text{monic}}{\Rightarrow}f=m$.

2. Assume $f(\varphi)=0$. Write $f=qm+r$ with $\deg r<\deg m$. (∗)

Then $0=f(\varphi)-q(\varphi)m(\varphi)=r(\varphi)\overset{(*)}{\Rightarrow}r=0$.

Jordan block: $A \in F^{d\times d}$ or $\varphi: V \longrightarrow V$ with $\dim V = d$ whose minimal polynomial is $(t-\lambda)^d$

$d=1$: $(\varphi-\lambda)v = 0 \iff v \in E(\lambda) \Rightarrow \mathcal{B}=\{v\}$ has $[\varphi]_{\mathcal{B}} = [\lambda]$

$d=2$: $(\varphi-\lambda)V \neq V$ or else $(\varphi-\lambda)^2 V = (\varphi-\lambda)\big((\varphi-\lambda)V\big) = (\varphi-\lambda)V = V$, but

$(\varphi-\lambda)V \neq 0$ by def of minimal polynomial

$\Rightarrow \dim(\varphi-\lambda)V = 1 \Rightarrow (\varphi-\lambda)V$ is in the $d=1$ case

$\vdots$

d arbitrary (prove by easy induction): $(\varphi-\lambda)V$ has dim $d-1$

$$\begin{array}{rll} & (\varphi-\lambda)^2V & d-2 \\ & \vdots & \vdots \\ V_{d-k} = & (\varphi-\lambda)^kV & d-k \quad \Rightarrow \dim V_k = k \\ & \vdots & \vdots \end{array}$$

Choose $v = v_d \in V \setminus V_{d-1}$. Then $(\varphi-\lambda)^d v_d = 0$ but $(\varphi-\lambda)^{d-1}v_d \neq 0$, so

$(\varphi-\lambda)v = v_{d-1} \in V_{d-1} \setminus V_{d-2}$ ...

$\vdots$

$(\varphi-\lambda)^{d-k}v = v_k \in V_k \setminus V_{k-1} \Rightarrow (\varphi-\lambda)v_k = v_{k-1} \Rightarrow \varphi v_k = \lambda v_k + v_{k-1}$ when $k \geq 2$, and

$\vdots$

$\varphi v_1 = \lambda v_1$.

So $\mathcal{B} = v_d, v_{d-1}, \dots, v_1 \Rightarrow [\varphi]_{\mathcal{B}} = \begin{array}{c|c} & d \\ \hline d & \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} \end{array}$.

<u>Pf of Jordan form thm</u>: Need $V = V_1 \oplus \cdots \oplus V_r$ with
- V_ℓ φ-invariant $\forall \ell$
- (min. poly. of $\varphi|_{V_\ell}$) $= (t-\lambda_\ell)^{d_\ell}$ for $d_\ell = \dim V_\ell$.

Not so hard to do directly, but best done using rings and modules.

<u>Def</u>: A <u>~~commutative~~ ring</u> satisfies all field axioms except
- multiplicative inverses need not exist
- multiplication need not be commutative

<u>E.g.</u> field, $\mathbb{Z}$, $F^{n\times n}$, $\mathbb{Z}^{n\times n}$, $R^{n\times n}$ for any commutative ring R

$F[t]$, $\mathbb{Z}[t]$, $R[t]$ for any ring R and any set t of variables

<u>Def</u>: A <u>module</u> over a ring R satisfies the same axioms as a vector space$/F$ but with scalars R.

<u>E.g.</u> vector space V/F with $\varphi: V \longrightarrow V$ is a module$/F[t]$ with $tv = \varphi(v)$.

JF thm follows by classifying all $F[t]$-modules of $\dim_F < \infty$: Look up "module over PID"

principal ideal domain

all are $\bigoplus$ invariant submodules $\langle v \rangle$ with $p^d(v) = 0$ for some irreducible p.

<u>Note</u>: same classification describes all finitely generated abelian groups:
$\mathbb{Z}$ and $F[t]$ are both PIDs

5.

Banach spaces: complete normed vector spaces over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$

Cauchy sequences converge

Def: A **norm** on $V/\mathbb{F}$ is $\nu : V \to \mathbb{R}_+$ with

nonnegative

- positive-definite • $\nu(x) = 0 \Leftrightarrow x = 0$
- homogeneous • $\nu(\alpha x) = |\alpha|\nu(x)$
- subadditive • $\nu(x+y) \leq \nu(x) + \nu(y)$ triangle inequality

E.g. Manhattan

$p=1$ • $\|x\|_1 = |x_1| + \cdots + |x_n|$

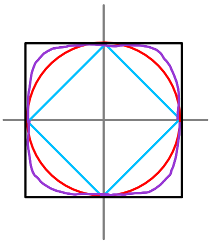
Euclidean

$p=2$ • $\|x\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{1/2}$

$\lim_{p\to\infty}$ • $\|x\|_\infty = \max_{i=1}^{n} |x_i|$

• $\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$

$B_\varepsilon^{\|\cdot\|_1}(0)$
$\subseteq$
$B_\varepsilon^{\|\cdot\|_2}(0)$
$\subseteq$
$B_\varepsilon^{\|\cdot\|_\infty}(0)$



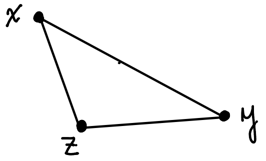
...or put any convex set here...

$B_\varepsilon^\nu(x) = \{y \in V \mid \nu(x-y) < \varepsilon\}$ next lecture

default: $B_\varepsilon = B_\varepsilon^{\|\cdot\|_2}$

Def: A **metric space** is a set X with a **distance** $d : X \times X \to \mathbb{R}_+$ such that $\forall x, y \in X$

- $d(x,y) = 0 \Leftrightarrow x = y$ separates
- $d(x,y) = d(y,x)$ symmetric
- $d(x,y) \leq d(x,z) + d(z,y)$ $\forall z \in X$ triangle inequality



E.g. norm ν induces distance $d_\nu(x,y) = \nu(x-y)$, such as **Euclidean metric** $\|x-y\|_2$ on $\mathbb{F}^n$

Manhattan metric $\|x-y\|_1$ on $\mathbb{F}^n$

All norms "pretty much feel the same". In what sense?

Def: A **topology** on a set S is a collection $\mathcal{U}$ of subsets called **open sets** such that

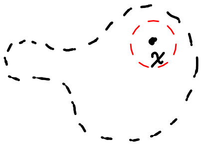
- any union of open sets is open $\quad \cup\, \mathcal{U} \in \mathcal{U}$
- any finite intersection of open sets is open $\quad \bigcap_{<\infty} \mathcal{U} \in \mathcal{U}$
- S and $\emptyset$ are open

Note: metric induces norm if homogeneous

E.g. usual topology on $\mathbb{F}^n$: $U \in \mathcal{U} \Leftrightarrow B_\varepsilon(x) \subseteq U$ $\forall x \in U$ and $\varepsilon = \varepsilon_x \ll 1$.

More generally: metric d on X ⇝ topology on X with

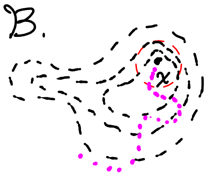
$$U \text{ open} \Leftrightarrow B_\varepsilon^d(x) \subseteq U \quad \forall x \in U \text{ and } \varepsilon = \varepsilon_x \ll 1.$$



Def: $\mathcal{B} \subseteq \mathcal{U}$ is a **base** for the topology if $U \in \mathcal{U} \Rightarrow U = \bigcup_{B \in \mathcal{B}'} B$ for some $\mathcal{B}' \subseteq \mathcal{B}$.

E.g. $\{B_\varepsilon^\nu(x) \mid x \in V \text{ and } \varepsilon \in \mathbb{R}_{>0}\}$

or $\varepsilon \ll 1$ or $\varepsilon \ll \varepsilon_0$



parti~~tot~~ally ordered

Def: $\{x_k\}_{k\in K} \to x$ if $\{x_k\}$ is **eventually in** U $\forall$ open $U \ni x$, meaning $\exists N_U \in K$ with $x_k \in U$ $\forall k \overset{\not<}{\cancel{\geq}} N_U$.

$X \subseteq S$ is **closed** if $x \in X$ whenever $\{x_k\} \to x$ in S with $\{x_k\} \subseteq X$. "X contains its limit points"

Prop: $X \subseteq S$ closed $\Leftrightarrow$ $S \setminus X$ open.

$\{x_k\}$ is never in $S\setminus X$, eventually or otherwise

Pf: $S\setminus X$ open and $\{x_k\} \subseteq X \Rightarrow \lim x_k$ (if $\exists$) can't lie in $S\setminus X$, so it must lie in X.

$S\setminus X$ not open $\Rightarrow \exists y \in S\setminus X$ such that $\forall$ open $U \ni y$ $\exists x_U \in U \cap X$. Then $\{x_U\}_{y\in U}^{\text{parti}\,\not<} \to y \notin X$. $\square$

Prop: Any norm ν on $\mathbb{F}^n$ is continuous in the Euclidean metric.

Pf: Given $\varepsilon>0$, need δ so that $|\nu(x)-\nu(y)|<\varepsilon$ whenever $\|x-y\|<\delta$.

Subadditivity $\Rightarrow \nu(x) \leq \nu(x-y)+\nu(y)$ and $\nu(y) \leq \nu(y-x)+\nu(x)$

$\Rightarrow \nu(x)-\nu(y) \leq \nu(x-y) \qquad \nu(y)-\nu(x) \leq \nu(y-x)$, so

$$|\nu(x)-\nu(y)| \leq \nu(x-y) = \nu\Big(\sum_{i=1}^{n}(x_i-y_i)e_i\Big) \leq \sum_{i=1}^{n}|x_i-y_i|\,\nu(e_i)$$

$$(*) \leq \|x-y\|_2\,\|v\|_2, \text{ where } v=(\nu(e_1),\dots,\nu(e_n)).$$

Pick $\delta=\frac{\varepsilon}{\|v\|_2}$. □

Q. why? A. Cauchy-Schwarz!

Def: Norms ν and μ on $V=\mathbb{F}^n$ are (topologically) equivalent, written $\nu\sim\mu$, if

$$\exists\ \alpha,\beta\in\mathbb{R}_{>0} \text{ with } \alpha\nu(x)\leq\mu(x)\leq\beta\nu(x) \quad \forall\, x\in V.$$

Interpretation: $\nu\sim\mu \Leftrightarrow B^{\nu}_{\varepsilon/\beta}(x)\subseteq B^{\mu}_{\varepsilon}(x)\subseteq B^{\nu}_{\varepsilon/\alpha}(x) \quad \forall x\in V$

$\alpha\nu(x-y)\leq\mu(x-y)<\varepsilon \Rightarrow \nu(x-y)<\varepsilon/\alpha$

$y\in B^{\mu}_{\varepsilon}(x) \subseteq y\in B^{\nu}_{\varepsilon/\alpha}(x)$

$\Leftrightarrow$ every ε-ball base for the μ-topology is a base for the ν-topology

E.g. $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2$

$\frac{1}{\sqrt{n}}\|x\|_2 \leq \|x\|_\infty \leq \|x\|_2$ **Pf:** exercise (not assigned)

$\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty$

Lemma: $\sim$ is an equivalence relation.

Pf: symmetric: $\frac{1}{\beta}\mu(x)\leq\nu(x)\leq\frac{1}{\alpha}\mu(x)$.

transitive: exercise.

reflexive: $\alpha=\beta=1$. □

Thm: μ,ν norms on $V=\mathbb{F}^n \Rightarrow \nu\sim\mu$.

Pf: By Lemma, need only check $\nu=\|\cdot\|_2$. Can assume $x\neq0$.

$(*)$ with $y=0$ and μ instead of $\nu \Rightarrow \mu(x)\leq\|x\|_2\|v\|_2$ for $v=(\mu(e_1),\dots,\mu(e_n))$

$\Rightarrow$ take $\beta=\|v\|_2$.

Set $\alpha=\min\{\mu(x)\mid \|x\|_2=1\}$, which exists by Prop because sphere S^{n-1} is closed and bounded.

and $\neq 0$ since $\mu(x)=0 \Rightarrow x=0$

Then $\mu(x)=\mu\Big(\|x\|_2\cdot\frac{x}{\|x\|_2}\Big)=\|x\|_2\,\mu\Big(\frac{x}{\|x\|_2}\Big)$

$\geq \|x\|_2\,\alpha$. □

CAN OMIT:

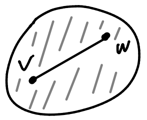
Def: norm on V^* dual to ν on V is $\nu^*(\varphi)=\max\limits_{\nu(x)=1}|\varphi(x)|$.

well defined since $S_\nu=\{x\in V\mid \nu(x)=1\}$ is closed and bounded by Thm.

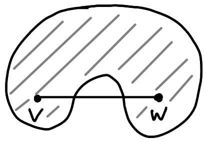
6. Convexity

Fix $\mathbb{R}$-vector space V.

Def: $X \subseteq V$ is convex if $v, w \in X \Rightarrow$ line segment $\overline{vw} \subseteq X$.



convex

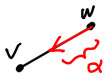


not convex

$$\overline{vw} = \{\alpha v + (1-\alpha)w \mid \alpha \in [0,1]\} = \left\{\alpha v + \beta w \,\middle|\, \begin{matrix}\alpha+\beta=1\\ \alpha\geq 0,\ \beta\geq 0\end{matrix}\right\}$$ = all weighted averages of v and w

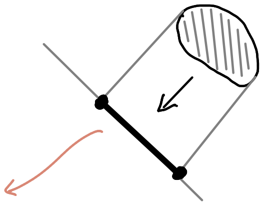
$$= \{w + \alpha(v-w) \mid \alpha \in [0,1]\}$$ = parametrized line segment from w to v

E.g. any interval $\subseteq \mathbb{R}$ $\quad \underbrace{[a,b] \quad (a,b] \quad (a,b) \quad [a,b)}$ $\quad a = -\infty$ or $b = \infty$ allowed



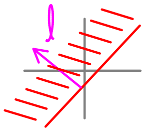
Pf: $a \overset{<}{\leq} v, w \overset{<}{\leq} b \Rightarrow [v,w] \subseteq$ " . ☐

Properties



1. $\varphi: V \to W$ linear $\Rightarrow \varphi(\text{convex})$ is convex. Pf: $\varphi(\overline{vw}) = \overline{\varphi(v)\varphi(w)}$.
2. " $\Rightarrow \varphi^{-1}(\text{convex})$ is convex. Pf: same! $v,w \in \varphi^{-1}(Y) \Rightarrow \overline{\varphi(v)\varphi(w)} \subseteq Y \Rightarrow \overline{vw} \subseteq \varphi^{-1}(Y)$.
3. X convex $\Rightarrow a + X$ convex $\forall a \in V$. Pf: $a+v, a+w \in a+X \Leftrightarrow v, w \in X \Leftrightarrow \overline{vw} \subseteq X \Leftrightarrow \underbrace{a + \overline{vw}}_{\overline{(a+v)(a+w)}} \subseteq a + X$.
4. $\{X_i\}_{i\in I}$ all convex $\Rightarrow \bigcap_{i\in I} X_i$ convex. Pf: $\overline{vw} \subseteq X_i\ \forall i \Leftrightarrow \overline{vw} \subseteq \bigcap_{i\in I} X_i$.

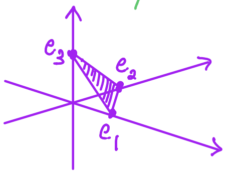
Cor: 5. Every affine subspace is convex. Pf: 1.+3.



6. Every open halfspace ~~closed~~ $\overline{H}^{\pm} = \{x \in V \mid \ell(x) \overset{\leq}{\gtrless} c\}$ for some $\ell \in V^*$ and $c \in \mathbb{R}$. Pf: $H^+ = \ell^{-1}(\text{ray})$; use 2.

closed

sometimes $\mathring{H}^{\pm}$ to emphasize "open"



E.g. standard $(n-1)$-simplex $\sigma = \{x \in \mathbb{R}^n \mid x_i \geq 0\ \forall i \text{ and } x_1 + \cdots + x_n = 1\}$ is convex.

Pf: $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0\ \forall i\} = \bigcap_{i=1}^{n} \{e_i^*(x) \geq 0\}$ convex by 6.+4.

$\{x \in \mathbb{R}^n \mid x_1 + \cdots + x_n = 1\}$ convex by 5. or 2. Now apply 4. ☐

Prop: $x_1, \dots, x_n \in V \Rightarrow$ 7. $P = \{\alpha_1 x_1 + \cdots + \alpha_n x_n \mid \alpha_i \geq 0\ \forall i \text{ and } \alpha_1 + \cdots + \alpha_n = 1\}$ is convex.

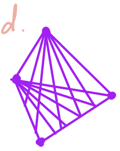
8. $X \supseteq \{x_1, \dots, x_n\}$ convex $\Rightarrow X \supseteq P$.

convex combinations of $x_1, \dots, x_n$

Def: P = convex hull = $\operatorname{conv}\{x_1, \dots, x_n\}$ is a polytope.

Pf: 7. $P = \varphi(\sigma)$ for $\varphi: \mathbb{R}^n \to V$, $e_i \mapsto x_i$. Use 1.

8. Use induction on n. Trivial if $n=1$. And by definition if $n=2$, though that's not needed.



$n \geq 2$: need $\alpha_1 x_1 + \cdots + \alpha_n x_n \in X$. Assume $\alpha_n \neq 1$. else trivial: $x_n \in X$

Then $\alpha_1 x_1 + \cdots + \alpha_n x_n = \alpha(\underbrace{\beta_1 x_1 + \cdots + \beta_{n-1}x_{n-1}}_{\text{by induction } \in X \text{ since } \sum_{i=1}^{n-1}\beta_i = 1}) + (1-\alpha)x_n$ for $\alpha = 1 - \alpha_n = \alpha_1 + \cdots + \alpha_{n-1}$

$\beta_i = \frac{\alpha_i}{\alpha}$ for $i = 1, \dots, n-1$. ☐

Further examples of convex sets:

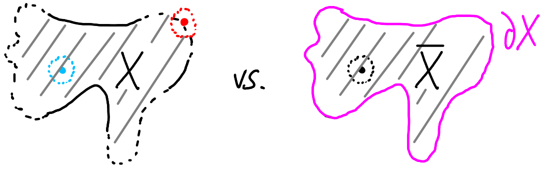
E.g. • $X = \{ f \in \mathbb{R}[t] \mid f(\alpha) > 0 \ \forall \alpha \in (0,1) \}$

• $X = \{ A \in \underset{\mathbb{C}?}{\mathbb{R}}^{n\times n} \mid A = A^* \text{ and } \lambda > 0 \ \forall \text{ eigenvalues } \lambda \text{ of } A \}$ see Lecture 9

Def: $v \in X \subseteq V$ is an interior point if $v \in U \subseteq X$ for some open $U \subseteq V$.

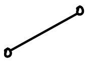
$v \in \partial X$ boundary point if $U \cap X$ and $U \cap (V \setminus X)$ both nonempty $\forall$ open $v \in U \subseteq V$.

$\overline{X} = X \cup \partial X$ is the closure of X.



Prop: $\overline{X} = \{ \lim x_k \mid \{x_k\} \subseteq X \text{ converges} \}$.

Pf: Prop, p. ⑪.

Q. Is  open?

A. In $\mathbb{R}^1$: yes.

In $\mathbb{R}^2$: no. relative interior can fix this, but let's not get into it

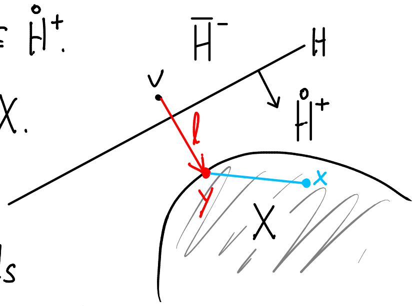
Prop: $X \subseteq V$ convex $\Rightarrow \overline{X}$ convex.

Pf: $\begin{matrix} v_k \to v \\ w_k \to w \end{matrix} \Rightarrow \alpha v_k + \beta w_k \to \alpha v + \beta w$. But $v_k \in X$ and $w_k \in X \ \forall k \overset{\alpha+\beta=1}{\Rightarrow} \alpha v_k + \beta w_k \in X \ \forall k$ by convexity

$\Rightarrow \alpha v + \beta w \in \overline{X}$ by Prop. □

Def: A hyperplane $H \subseteq V$ separates v from X if $v \in \overline{H}^-$ and $X \subseteq \mathring{H}^+$.

linear function $\ell \in V^*$ " " " " $\ell(v) < \ell(x) \ \forall x \in X$.



Thm: $X = \overline{X}$ convex and $v \notin X \Rightarrow \exists H$ separating v from X.

Pf: For any choice of inner product, the norm $\|\cdot\| = \langle \cdot , \cdot \rangle^{1/2}$ yields $f(x) = \|x - v\| : X \to \mathbb{R}$ continuous, bounded below (by 0), and proper (f^{-1}(bounded set) is bounded) by topological equivalence: $f^{-1}([0,r]) = B_r^{\|\cdot\|}(v)$ is bounded. Since X is closed, f attains a minimum on X, say at $y \in X$. Set $\ell = \langle y - v, \cdot \rangle$. Note $\ell \neq 0$ since $v \notin X$ (so $y \neq v$).

$\forall \alpha \in (0,1]$ and $x \in X$, $\cancel{\|y-v\|^2} \underset{\text{min}}{\leq} \|y + \alpha(x-y) - v\|^2 = \|y - v + \alpha(x-y)\|^2$

$$= \cancel{\|y-v\|^2} + \underbrace{2\cancel{\alpha}\langle y-v, x-y\rangle}_{-2\alpha\langle v-y,\, x-y\rangle} + \alpha^{\cancel{2}}\|x-y\|^2$$

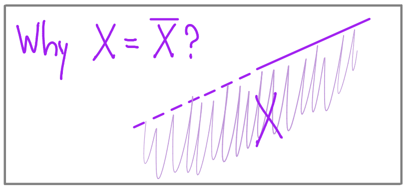
law of cosines

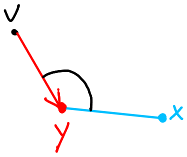
$\Rightarrow 0 \leq 2\ell(x-y) + \alpha\|x-y\|^2$

$\overset{\alpha \to 0}{\Rightarrow} \ell(y) \leq \ell(x)$.

But $\ell(v) < \ell(y)$ because $\ell(y-v) = \|y-v\|^2 > 0$,

so $\ell(v) < \ell(x) \ \forall x \in X$. □


Why $X = \overline{X}$?



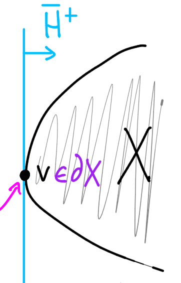
Applications: optimization (linear programming), analysis (bounds away from 0), CS (support vector machines)

7.

Def: H is a **support hyperplane** of X at $v\in\overline{X}$ if $v\in H$ and $X\subseteq\overline{H}^+$ or $\overline{H}^-$

compare: separating has $X\subseteq\mathring{H}^+$, $v\in\overline{H}^-$, so supporting is "closer" to X

think $v\in\partial X$, though not necessary by def



extreme point

$v\in X$ or $v\notin X$ both work

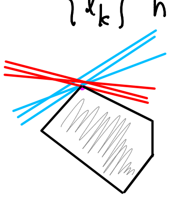
Thm: X convex and $v\in\partial X$ ⇒ ∃ support H of X at v.

Pf: $X\subseteq\overline{H}^+\Rightarrow\overline{X}\subseteq\overline{H}^+$ so assume $X=\overline{X}$. Fix a norm $\|\cdot\|$ on V^*.

For $k\in\mathbb{N}$ pick $v_k\notin X$ and $\ell_k\in S^{n-1}=\{w\in V^*\mid\|w\|=1\}\subseteq V^*$ separating v_k from X.

Assume $v_k\longrightarrow v$ as $k\longrightarrow\infty$. S^{n-1} is compact (closed + bounded) so $\{\ell_k\}$ has a convergent subsequence; replace $\{\ell_k\}$ with that to get $\{\ell_k\}\longrightarrow\ell$.

If you weren't convinced that V^* should be thought about separately from V, then reconsider now.



$$\forall x\in X,\ \{\underbrace{\ell_k(x)-\ell_k(v_k)}\}_{k\in\mathbb{N}}\longrightarrow\underbrace{\ell(x)-\ell(v)}$$

$$\geq 0 \Longrightarrow \geq 0 \Rightarrow H=\{y\in V\mid\ell(y)=\ell(v)\}$$ suffices. □

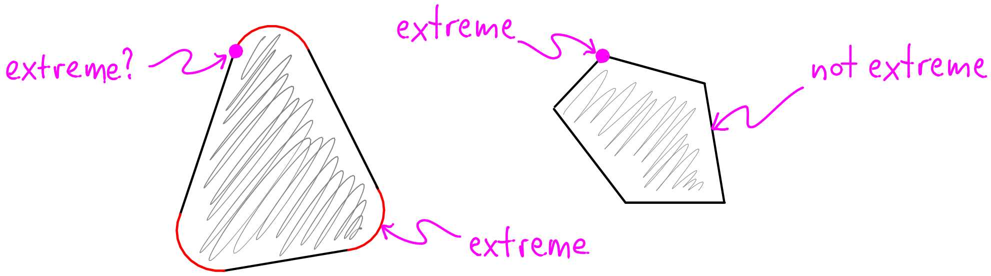
Def: $S\subseteq V$ has **convex hull** $\text{conv}(S)=\{$convex combinations of points in $S\}$

Prop: $=$ smallest convex set $\supseteq S$.

Pf: A convex combination of convex combinations is a convex combination. Now use Property 8. □

Def: For X convex, $v\in X$ is an **extreme point** if $v\notin\text{conv}(X\setminus\{v\})$.

Lemma: $\Leftrightarrow v\notin\overline{xy}$ whenever $v\neq x\in X$ and $v\neq y\in X$. i.e. two points suffice to witness



Thm: $X=\overline{X}$ bounded + convex ⇒ each supporting H contains an extreme point of X.

Pf (assuming $\dim V<\infty$): Fix supporting H. Set $Y=X\cap H$: bounded, closed, convex.

Claim: $v\in Y$ extreme in Y ⇒ v extreme in X.

dim V = 1 elementary

⇒ Thm by induction on dim V (as $\dim H=\dim V-1$) via previous Thm.

All that's really needed for Thm is for Y to contain any extreme point of X.

Pf of Claim: Pick $\ell\perp H$ with $\ell(X)\geq\ell(v)$.

$v=\alpha x+\beta y$ with $x,y\in X$ and $\alpha+\beta=1\Rightarrow\ell(v)=\alpha\ell(x)+\beta\ell(y)$

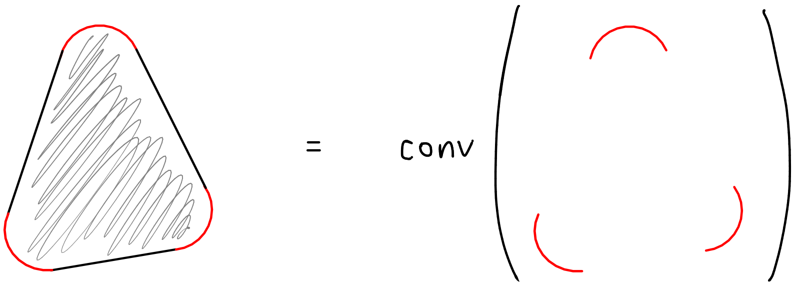
$\Rightarrow\ell(x)=\ell(v)=\ell(y)$ since $\ell(x)\geq\ell(v)$ and $\ell(y)\geq\ell(v)$

$\Rightarrow x\in H$ and $y\in H$. □

no inner product is being used here

Thm (Krein–Milman): $X=\overline{X}$ bounded + convex $\Rightarrow$ $X = \mathrm{conv}(\text{extreme points of } X)$.

E.g.



Pf: Suppose $X \supseteq Y$ with $Y=\overline{Y}$ convex. think: $Y = \mathrm{conv}(\text{extreme points of } X)$

$X \not\subseteq Y \Rightarrow \exists\, v \in X\setminus Y$ with ℓ separating v from Y.

X closed + bounded (i.e. compact)

$\Rightarrow$ ℓ attains min. at some $x \in X$;

note that $\ell(x) < \ell(Y)$ by separation.

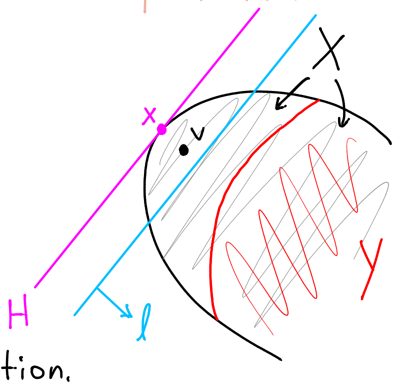
$H = \{w \in V \mid \ell(w) = \ell(x)\}$ supports X at x by construction.

Previous thm $\Rightarrow$ $\exists$ extreme point of X in H

$\Rightarrow$ Y can't contain all extreme points of X.

Thus $X \subseteq \mathrm{conv}(\text{extreme points of } X)$.

$\supseteq$ by Prop. $\square$



8. Grassmannians

Def: A k-plane in a vector space / F is a subspace of dimension k.

The grassmannian is $G_k(W) = \{k\text{-planes in } W\}$

- What does $G_k(F^n)$ "look like"?
- How close is one k-plane to another?
- Find "best" k-plane for given purpose ⟷ optimization on $G_k(F^n)$

data points

e.g. least squares ⟷ on $\mathbb{P}(\mathbb{R}^n) = G_1(\mathbb{R}^n)$ for cost function $f(V) = \sum_{i=1}^{r} d(x_i, V)^2$

Q. $V \subseteq F^n$ specified by ... ~~?~~ A. basis $v_1, \dots, v_k$

$\Rightarrow F^{k\times n} \overset{?}{\longleftrightarrow} G_k(F^n)$ no: need rank = k $F_*^{k\times n}$

$F_*^{k\times n} \twoheadrightarrow G_k(F^n)$

$A \mapsto \text{span}(\text{rows of } A)$ $\twoheadrightarrow$ but not $\hookrightarrow$

Note: $F^n = F^n_{row}$ here; could just as easily do everything with columns

E.g. $A = \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 1 \end{bmatrix} \mapsto V$

$\Rightarrow A' = \begin{bmatrix} 2 & 0 & 2 & 6 \\ 1 & 1 & 0 & 4 \end{bmatrix} \mapsto V$ $A' = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} A$

Prop: $A, A' \in F_*^{k\times n}$ yield same $V \Leftrightarrow$ rows of A are linear combinations of rows of A'

$\Leftrightarrow A' = gA$ for some $g \in F^{k\times k}$

(*) $\Leftrightarrow$ " $g \in F_*^{k\times k} \overset{\text{Def}}{=\!=} GL_k(F)$, the group of invertible k×k matrices

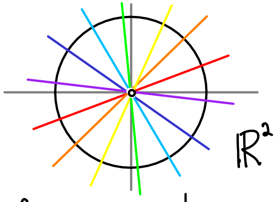
Cor: $G_k(F^n) = GL_k(F)\backslash F_*^{k\times n}$, the quotient of $F_*^{k\times n}$ modulo the action of GL_k on the left $GL_k \times F^{k\times n} \to F^{k\times n}, \dots$

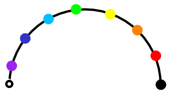
$= \{[A] \subseteq F_*^{k\times n} \mid A \in F_*^{k\times n}\}$, where $A' \in [A] \Leftrightarrow$ (*)

$[A] = GL_k(F)A$, the orbit of A

E.g. n=2, k=1

$[\alpha\ \beta] \mapsto [[\alpha\ \beta]]$





$\mathbb{R}^2 \setminus \{0\} \to G_1(\mathbb{R}^2) = \mathbb{RP}^1 = \mathbb{P}^1(\mathbb{R})$

Compare: $V/W = \{[v] \subseteq V \mid v \in V\}$, where $v' \in [v] \Leftrightarrow v' = v + w$ for some $w \in W$

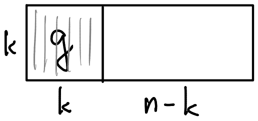
$[v] = W + v$, the coset of v

key point

Note: $F^n_{row} \rightsquigarrow F^n_{col} \Rightarrow G_k(F^n_{col}) = F_*^{\boxed{n\times k}} / GL_k(F) = \{A \cdot GL_k \mid A \in F_*^{n\times k}\}$

How does this quotient business help? What kind of structure does it induce on $G_k(F^n)$?

Suppose left k cols of $A \in F^{k\times n}$ are independent.



Let $g = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_k \\ | & & | \end{bmatrix} \in GL_k$ i.e. invertible

so $\hat{A} = g^{-1}A = \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & \text{...stuff...} \\ & & 1 & \end{array}\right]$ = reduced echelon form of A!

Here is a reason to be using row vectors.

Then $A \mapsto V$

$\Rightarrow \hat{A} \mapsto V$

<u>Lemma</u>: $A' \mapsto V \Rightarrow \hat{A}' = \hat{A}$. <u>Pf</u>: same row span, in echelon form, which is unique. □

$\Updownarrow$

<u>Lemma'</u>: $\{ A \in F^{k\times n} \mid \text{left } k \text{ cols} = I_k \} \hookrightarrow G_k(F^n)$.

Q. $\twoheadrightarrow$? A. No.

rank $A = k \Rightarrow$ some set σ of k cols is independent.

Write $F^{k\times n}_\sigma = \{ \hat{A} \in F^{k\times n} \mid \hat{A} \text{ has } I_k \text{ in cols from } \sigma \}$.

<u>E.g.</u> $F^{k\times n}_{[k]}$ with $[k] = \{1,\dots,k\}$.

<u>Lemma''</u>: $F^{k\times n}_\sigma \overset{\pi_\sigma}{\hookrightarrow} G_k(F^n) \quad \forall\ \sigma \in \binom{[n]}{k} = \binom{\{1,\dots,n\}}{k}$. Same proof.

$\cap\!|$ ⟵ affine subspace

$F^{k\times n}$ ⟵ vector space

Set $G_k^\sigma = \text{im}(\pi_\sigma) \cong F^{k\times(n-k)}$

$G_k(F^n)$ locally looks like $F^{k\times(n-k)}$

Summary:

<u>Prop</u>: $G_k(F^n)$ covered by $\binom{n}{k}$ affine spaces $F^{k\times n}_\sigma \overset{\pi_\sigma}{\xrightarrow{\sim}} G_k(F^n) = \bigcup_{\sigma \in \binom{[n]}{k}} G_k^\sigma$.

<u>Def</u>: A <u>manifold</u> is a topological space X with an <u>atlas</u>: a set of maps $U_\alpha \overset{\pi_\alpha}{\hookrightarrow} X$ <u>charts</u>

such that • $X = \bigcup_\alpha X_\alpha$, where $X_\alpha = \text{im}(\pi_\alpha)$ is open in X

and $\forall \alpha$ • U_α is open in an affine space of dim d $\simeq F^d$

• π_α is a <u>homeomorphism</u> $U_\alpha \xrightarrow{\sim} X_\alpha$ $\quad W \subseteq U_\alpha$ open $\Leftrightarrow \pi_\alpha(W) \subseteq X_\alpha$ open

and $\forall \alpha,\beta$ • $\pi_\beta^{-1} \circ \pi_\alpha : U_{\alpha\beta} \xrightarrow{\sim} U_{\beta\alpha}$ is $\quad$ e.g. $U_{\alpha\beta} = \pi_\alpha^{-1}(X_{\alpha\beta}) \subseteq U_\alpha$ open

$$X_{\alpha\beta} = X_\alpha \cap X_\beta$$

Which one(s) is $G_k(F^n)$?

$F = \mathbb{R}$ - continuous $\Rightarrow$ <u>topological manifold</u>

$\Uparrow$

- differentiable $\Rightarrow$ <u>differentiable manifold</u>

$\Uparrow$

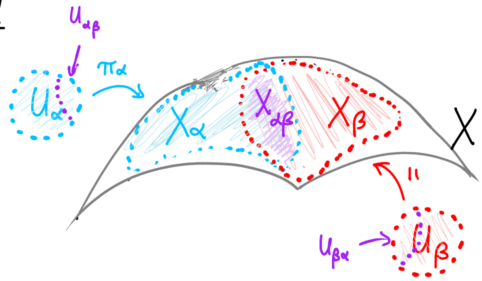
- smooth (C^∞) $\Rightarrow$ <u>smooth manifold</u>

$\Uparrow$

$F = \mathbb{C}$ or $\mathbb{R}$ - analytic $\Rightarrow$ <u>analytic manifold</u>

$\Uparrow$

F arbitrary - ratio of polynomials $\Rightarrow$ <u>rational algebraic variety</u>

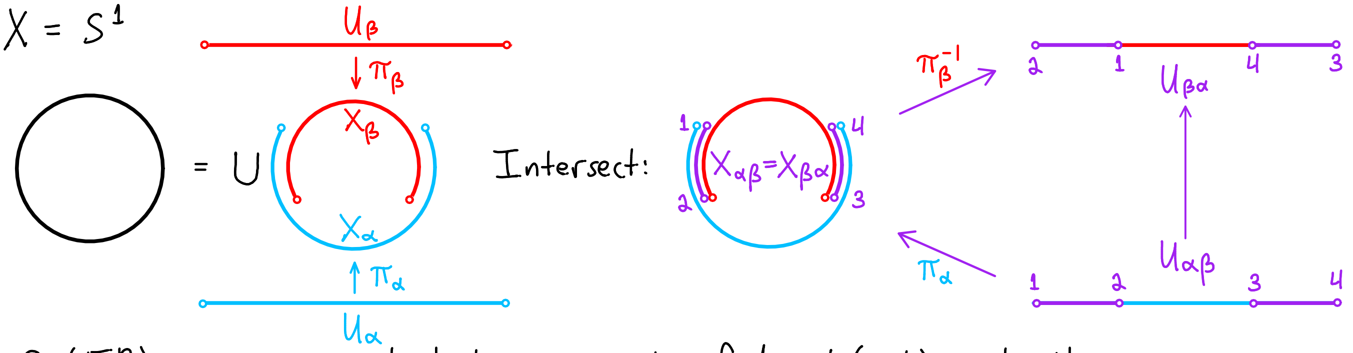


<u>Thm</u>: $G_k(F^n)$ is a rational algebraic variety of dim $k(n-k)$ with atlas

$$\left\{ \pi_\sigma : F^{k\times n}_\sigma \to G_k(F^n) \,\middle|\, \sigma \in \binom{[n]}{k} \right\}.$$

9. Review def of manifold; e.g. X = surface of Earth, atlas = actual "rectangular" maps

E.g. $X = S^1$



Thm: $G_k(F^n)$ is a rational algebraic variety of dim $k(n-k)$ with atlas

$$\left\{\pi_\sigma : F_\sigma^{k\times n} \longrightarrow G_k(F^n) \;\middle|\; \sigma \in \binom{[n]}{k}\right\}.$$

Pf: • Prop $\Rightarrow X = \bigcup_\alpha X_\alpha$: $G_k(F^n) = \bigcup_\sigma G_k^\sigma$

• and $G_k^\sigma \simeq F_\sigma^{k\times n} \simeq F^d$ for $d = k(n-k)$

• Declare $U \subseteq G_k(F^n)$ to be open $\Leftrightarrow$ $U \cap G_k^\sigma$ is open $\forall\, \sigma \in \binom{[n]}{k}$. HW 2.16: well defined

• Set $F_{\sigma,\tau}^{k\times n} = \{A \in F_\sigma^{k\times n} \mid A_\tau = [\text{cols of } A \text{ indexed by } \tau] \text{ is invertible}\}$

$= \pi_\sigma^{-1}(G_k^\sigma \cap G_k^\tau)$. recall: [cols indexed by σ] is I_k for $A \in F_\sigma^{k\times n}$

Then $\pi_\tau^{-1} \circ \pi_\sigma : F_{\sigma,\tau}^{k\times n} \longrightarrow F_{\tau,\sigma}^{k\times n}$

$A \mapsto ?$ Find matrix A' with $[A'] = [A]$ and $A'_\tau = I_k$

easy: $A' = \underbrace{A_\tau^{-1} A}_{\text{entries are rational functions of entries of } A.}$ □

So that's how grassmannians work. Let's all do an exercise together to see these methods in action.

Def: A (complete) flag in V is a chain

$$0 = V_0 \subset V_1 \subset \cdots \subset V_{n-1} \subset V_n = V$$

of subspaces of V with $\dim V_i = i$ $\forall i$. Set

$$\mathcal{F}\ell_n(F) = \{\text{complete flags in } F^n\}.$$

Ex. Express $\mathcal{F}\ell_n$ as a quotient.

• of what? How do you write down a flag? Do it, then quotient modulo choices.
This a very general principle in math.

$V_1 = \langle v_1 \rangle$

$V_2 = \langle v_1, v_2 \rangle$ $\qquad A = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \in GL_n(F)$ Why?

$\vdots$ (Yes we're using columns now.)

$V_i = \langle v_1, \ldots, v_i \rangle$

$\vdots$

- by what?

$\langle v_1\rangle = \langle \alpha v_1\rangle$ for any $\alpha \in F^*$

$\langle v_1, v_2\rangle = \underbrace{\langle \alpha v_1\rangle + \langle \beta_1 v_1 + \beta_2 v_2\rangle}$ for any $\alpha \in F^*$, $\beta_1 \in F$, and $\beta_2 \in F^*$

$\langle v_1, v_2, v_3\rangle = \qquad\qquad + \langle \underbrace{\text{some replacement}}_{\substack{\in \operatorname{span}(v_1,v_2,v_3)\setminus \operatorname{span}(v_1,v_2), \\ \text{so } \gamma_1 v_1 + \gamma_2 v_2 + \gamma_3 v_3 \text{ with } \gamma_3 \neq 0}} \text{ for } v_3\rangle$

$\vdots$

$$A = AB \text{ for } B = \begin{bmatrix} F^* & F & F & \cdots & F \\ 0 & F^* & F & & \vdots \\ \vdots & 0 & F^* & & \\ \vdots & \vdots & & \ddots & F \\ 0 & 0 & \cdots & 0 & F^* \end{bmatrix} = \begin{bmatrix} \diagdown & * \\ 0 & \diagdown \end{bmatrix} \subseteq \underset{\substack{\Leftrightarrow \text{ diagonal entries all} \neq 0, \\ \text{given upper-triangular}}}{GL_n(F)}$$

<u>Def</u>: $B_n^+ = \left\{ \begin{bmatrix} \diagdown & * \\ 0 & \diagdown \end{bmatrix} \right\} \subseteq GL_n$ is the <u>Borel subgroup</u>.

<u>Prop</u>: $\mathcal{F}\ell_n = GL_n / B_n^+$. □

Is it a manifold? A variety? Find

- "big" subset that is an open subset of a vector space
- enough copies to cover.

Assume A "generic". What does that mean? Don't know yet. Try; see what's needed.

Use • b_{11} to make $a_{11} = 1$ needs "generic"

then • $b_{12}, \dots, b_{1n}$ to cancel $a_{12}, \dots, a_{1n}$

then • b_{22} to make new $a_{22} = 1$

then • $b_{23}, \dots, b_{2n}$ to cancel $a_{23}, \dots, a_{2n}$

$\vdots$

$$B_1 = \begin{bmatrix} a_{11}^{-1} & -a_{11}^{-1}a_{12} \cdots -a_{11}^{-1}a_{1n} \\ 0 & \\ \vdots & I_{n-1} \\ 0 & \end{bmatrix}$$

column reduce A without swapping any columns

$$AB_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ * & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & * \end{bmatrix}$$

$$AB_1B_2 = \begin{bmatrix} 1 & 0 \cdots 0 \\ * & 1 \; 0 \cdots 0 \\ \hline \multicolumn{2}{c}{*} \end{bmatrix}$$

$$AB_1 \cdots B_n = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ * & \ddots & 1 \end{bmatrix} \in U_n^- \quad \underline{\text{unipotent subgroup}}$$

<u>Prop</u>: $U_n^- \hookrightarrow \mathcal{F}\ell_n = GL_n / B_n^+$.

<u>Pf</u>: HW.

<u>Thm</u>: $\mathcal{F}\ell_n$ is a rational algebraic variety with atlas

$\{ wU_n^- \longrightarrow GL_n / B_n^+ \mid w \text{ is a permutation matrix} \}$.

<u>Pf</u>: HW.

10. Orthogonal groups

intuitively: pick up S^2 and put it down on top of itself

unit sphere

Q. What is {symmetries of $S^2 \subseteq \mathbb{R}^3$}?

Def: An isometry of a metric space X is a bijection $\varphi: X \to X$ with
$$d(\varphi x, \varphi y) = d(x,y) \ \forall\, x,y \in X.$$

In $S^{\textcircled{2}\, n-1}$, $d(x,y) = \angle(x,y) \in [0,\pi]$

Note: only need $\varphi: X \twoheadrightarrow X$, since $x \neq y \Rightarrow d(\varphi x, \varphi y) = d(x,y) \neq 0 \Rightarrow \varphi x \neq \varphi y$.

E.g. Find (X,d) and φ satisfying everything except $\twoheadrightarrow$. e.g. $X = \mathbb{R}_+$, $\varphi: x \mapsto x+1$

Q. Is every isometry of S^2 rotation about some axis?

A. No, but if assume it preserves orientation then

- φ isom of $S^2 \Rightarrow$ preserves orthonormality of any basis of $\mathbb{R}^3$
- (v_1, v_2, v_3) right-handed if $v_1 \times v_2 = v_3$ — in $\mathbb{R}^n$: positively oriented if $\det\begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} > 0$
- Def: φ preserves orientation if φ preserves handedness

E.g.
- $\varphi = -I_3 \Rightarrow$ not
- $\varphi =$ reflection $\Rightarrow$ not
- $\varphi =$ rotation $\Rightarrow$ ✓

Lemma: $\varphi: S^{n-1} \to S^{n-1}$ isometry $\Rightarrow \{\alpha\varphi \mid \alpha \in \mathbb{R}_+\}$ is an isometry of $\mathbb{R}^n = \bigcup_{\alpha \geq 0} \alpha S^{n-1}$

i.e. φ extends to an isometry of $\mathbb{R}^n$ $\Rightarrow$ need only study isometries of $\mathbb{R}^n$.

Pf: $x, y \in \mathbb{R}^n \Rightarrow \varphi$ preserves
- $\|x\|$ and $\|y\|$ by construction
- $\angle(x,y) \overset{\text{def}}{=} \angle\left(\frac{x}{\|x\|}, \frac{y}{\|y\|}\right)$ by isometry of S^{n-1}

$\Rightarrow$
- $\|x-y\|$ by law of cosines. ☐

$$a^2 + b^2 = c^2 + 2ab\cos C$$
$\|x\|^2$ ✓ $\|y\|^2$ ✓ $\|x-y\|^2$ $\|x\|$ ✓ $\|y\|$ ✓ $\angle(x,y)$ ✓

Thm: Every isometry of $\mathbb{R}^n$ has the form $A + T_v$ for some
- $A \in O_n(\mathbb{R}) = \{Q \in \mathbb{R}^{n\times n} \mid Q^{-1} = Q^T\}$ orthogonal group
- $T_v =$ translation by $v \in \mathbb{R}^n$.

Pf: Assume $\varphi \in \operatorname{isom}(\mathbb{R}^n)$ with $\varphi(0) = v$. Replace φ with $\varphi - T_v$ to assume $v = 0$.

Need $\varphi \in O_n(\mathbb{R})$; proof essentially as in Lemma + (preserves inner products $\Rightarrow$ linear). ☐

$(Ax)_i = \langle Ax, e_i \rangle$

as a set, not pointwise

Note: $S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ fixed by isometry φ of $\mathbb{R}^n \Leftrightarrow \varphi(0) = 0$.

Cor: $\operatorname{Isom}(S^{n-1}) \longleftrightarrow O_n(\mathbb{R})$.

Answer to Q.

- $A \in O_3 \Rightarrow A$ has a real eigenvalue Why? deg(char. poly) = 3 odd

  $\Rightarrow$ at least one eigenvalue $\lambda = \pm 1$, since $|\lambda| = 1$ HW

  $\Rightarrow$ eigenline = axis fixed pointwise by A or $-A$. One of these preserves orientation since 3 is odd

But • $A \in O_3 \Rightarrow A$ takes orthonormal basis of $P = \text{axis}^{\perp}$

to " " " "

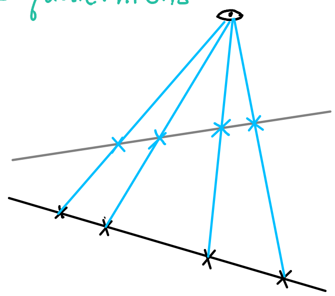
$\Rightarrow A|_P$ is rotation of P, possibly $\underbrace{\text{followed by reflection}}_{\text{but not if } A \text{ preserves orientation.}}$

Pf: $(e_1, e_2, e_3) \overset{A}{\mapsto} (v_1, v_2, v_3) \Rightarrow \det A = v_1 \cdot (v_2 \times v_3) = v_1 \cdot (\pm v_1)$.

But $\text{rot}_{\text{axis}}(\theta)$ has eigenvalues $1, \lambda, \bar{\lambda}$ for some $\lambda \in \mathbb{C}$ with $|\lambda|^2 = \lambda\bar{\lambda} = 1$,

so $\det(\ \uparrow\ ) = 1 \cdot \lambda \cdot \bar{\lambda} = 1 \Rightarrow v_1 \cdot (\pm v_1) = 1 \Rightarrow +$. □

Aside: Other symmetry groups G
- rotation ($\mathbb{R}$ or $\mathbb{C}$ or $\mathbb{H}$) ← Hamilton's quaternions
- scaling ($\mathbb{R}$, $\mathbb{C}$, arbitrary F)
- translation
- affine or projective transformations



Preserve: angle, distance, collinearity, …

Applications: computer vision, rendering, 3D image reconstruction, morphometrics

E.g. face recognition ($n=2$) $n\,\underset{\nwarrow\nwarrow\text{landmarks}\nearrow\nearrow}{\overset{d}{\boxed{\;||||\qquad A \qquad ||||\;}}} \in \mathbb{R}^{n \times d}$ $A \sim A'$ same data point if $A' = gA$ $g \in G$

$X = G \backslash \mathbb{R}^{n \times d}$ algebraic variety, metric space: $d([A],[B])$ = something from linear algebra of A and B

Back to O_n…

Prop: Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and $\langle \cdot , \cdot \rangle$ standard hermitian form on $\mathbb{F}^n$. TFAE. I never want to see this written in TEX!

1. $\langle xA, yA \rangle = \langle x, y \rangle \ \forall$ rows $x, y \in \mathbb{F}^n$ Def: $A \in O_n(\mathbb{F})$
2. right multiplication ρ_A preserves orthonormal bases: $v_1, \dots, v_k$ orthonormal
3. rows of A are orthonormal basis of $\mathbb{F}^n$ $\Rightarrow v_1 A, \dots, v_k A$ orthonormal
4. $AA^* = I_n$
5. cols of A are orthonormal basis of $\mathbb{F}^n$

Pf: Exercise (not assigned). Discuss orally if time permits.

Def: $U_n = O_n(\mathbb{C})$ unitary group What does it "look like"?

11. Unitary matrices

$\mathbb{C}^n \cong \mathbb{R}^{2n}$ via $(a_1+b_1 i, \dots, a_n+b_n i) \mapsto (a_1, b_1, \dots, a_n, b_n)$

double the size of each entry

$$\underset{\overset{\|}{\mathbb{C}^{n\times n}}}{M_n\mathbb{C}} \qquad \underset{\overset{\|}{\mathbb{R}^{2n\times 2n}}}{M_{2n}\mathbb{R}}$$

induces $d_n : \mathbb{C}^{n\times n} \to \mathbb{R}^{2n\times 2n}$ by viewing each $\mathbb{C}$-entry as a 2x2 block

$$a+bi \mapsto \begin{bmatrix} a & \\ & a \end{bmatrix} + \begin{bmatrix} & -b \\ b & \end{bmatrix} \quad \text{or} \quad 1 \mapsto \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$$

$$i \mapsto \begin{bmatrix} & -1 \\ 1 & \end{bmatrix} \Rightarrow -i \mapsto \begin{bmatrix} & 1 \\ -1 & \end{bmatrix} = \begin{bmatrix} & -1 \\ 1 & \end{bmatrix}^T$$

$M_n\mathbb{C}$ $M_{2n}\mathbb{R}$ $\mathbb{C}$-

Def: $A \in M_{2n}\mathbb{R}$ is complex-linear if $A \in \text{image}(d_n)$.

Prop: $A \in U_n \Leftrightarrow d_n(A) \in O_{2n}(\mathbb{R}) \cap d_n(M_n\mathbb{C})$.

Pf: $d_n(A^*) = d_n(A)^*$ $\qquad -i \mapsto \begin{bmatrix} & 1 \\ -1 & \end{bmatrix} = \begin{bmatrix} & -1 \\ 1 & \end{bmatrix}^T$

adjoint — ordinary transpose

Check: $d_n(AB) = d_n(A)d_n(B)$. $\qquad d_n$ is a ring homomorphism.

- view A, B as $\mathbb{C}$-linear on $\mathbb{C}^n$
- compose
- view $\mathbb{C}^n$ as v.s./$\mathbb{R}$

- view A, B as operators on $\mathbb{R}^{2n}$
- compose

The two sides are equal as functions $\mathbb{C}^n \to \mathbb{C}^n$, regardless of whether $\mathbb{C}^n$ is viewed as v.s./$\mathbb{C}$ or v.s./$\mathbb{R}$ or merely as a set.

Thus $d_n(A)d_n(A)^* = d_n(A)d_n(A^*) = d_n(AA^*)$.

$$d_n(A) \in O_{2n}(\mathbb{R}) \Leftrightarrow \overset{\|}{I_{2n}} \qquad \overset{\|}{I^{\mathbb{R}}_{2n}} \Leftrightarrow AA^* = I^{\mathbb{C}}_n \Leftrightarrow A \in U_n. \quad \square$$

Def: For v.s. $V/\mathbb{F}$ with $\langle \cdot, \cdot \rangle$, $O(V) = \{\varphi : V \to V \mid \langle \varphi x, \varphi y \rangle = \langle x, y \rangle \ \forall x, y \in V\}$.

Prop: $O(V) = \{\varphi \in GL(V) \mid \|\varphi x\| = \|x\| \ \forall x \in V\}$. $\quad \Rightarrow O(\mathbb{F}^n) = O_n(\mathbb{F})$

Pf: $\|x-y\|^2 = \langle x-y, x-y \rangle_{\mathbb{R}} = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle_{\mathbb{R}}$ $\qquad \langle x, y \rangle_{\mathbb{C}} = \langle d_n x, d_n y \rangle_{\mathbb{R}} + i\langle d_n x, d_n(iy) \rangle_{\mathbb{R}}$

$\Rightarrow \langle x, y \rangle_{\mathbb{R}} = \frac{1}{2}\left(\|x\|^2 + \|y\|^2 - \|x-y\|^2\right)$

so φ preserves norm $\Leftrightarrow$ preserves inner product$_{\mathbb{R}}$.

For $\mathbb{F} = \mathbb{C}$, $\varphi : \mathbb{C}^n \to \mathbb{C}^n$ preserves norms $\Rightarrow d_n(\varphi)$ preserves norms $\quad$ check!

For general $V/\mathbb{C}$, choose $\underbrace{\text{orthonormal basis}}_{(V, \langle \cdot, \cdot \rangle) \cong (\mathbb{C}^n, \|\cdot\|_2)}$ $\qquad \Rightarrow d_n(\varphi) \in O(\mathbb{R}^{2n})$

$\Rightarrow \varphi \in O(\mathbb{C}^n)$. $\square$

Prop: $A \in O_n(\mathbb{F}) \Rightarrow |\det A| = 1$.

Pf: $\det(AA^*) = (\det A)\overline{(\det A)} = |\det A|^2 = 1$ if $AA^* = I$. $\square$

Def: $SO_n(\mathbb{R}) = \{A \in O_n(\mathbb{R}) \mid \det A = 1\}$. $\quad$ could have been ± 1

$SU_n = \{A \in U_n \mid \det A = 1\}$. $\quad$ " " " $e^{i\theta}$ for some $\theta \in \mathbb{R}$

Running assumption: V with $\langle\cdot,\cdot\rangle$, $\dim_{\mathbb{C}} V = n$, $\varphi: V \to V$ $\mathbb{C}$-linear

Thm: V has orthonormal basis $\mathcal{B}$ with $[\varphi]_{\mathcal{B}}$ upper-$\triangle$.

$\Updownarrow$ $A \in M_n\mathbb{C} \Rightarrow U^*AU$ upper-$\triangle$ for some $U \in \mathcal{U}_n$.

Pf: $n=1$ ✓

$n \geq 2$: pick unit eigenvector u_1 and $\perp$normal basis $v_2, \dots, v_n$ for $u_1^\perp$ with $Q^*AQ = \begin{bmatrix} \lambda_1 & * \\ \begin{matrix}0\\ \vdots \\ 0\end{matrix} & \boxed{A'} \end{bmatrix}$

where $Q = \begin{bmatrix} | & | & & | \\ u_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$. By induction, choose $W = \begin{bmatrix} | & & | \\ w_2 & \cdots & w_n \\ | & & | \end{bmatrix} \in \mathcal{U}_{n-1}$ so $W^*A'W$ upper-$\triangle$.

Set $U = \begin{bmatrix} 1 & \\ & \boxed{W} \end{bmatrix} Q$. Then $U^*AU = \begin{bmatrix} 1 & \\ & \boxed{W^*} \end{bmatrix} Q^*AQ \begin{bmatrix} 1 & \\ & \boxed{W} \end{bmatrix} = \begin{bmatrix} 1 & \\ & \boxed{W^*} \end{bmatrix} \begin{bmatrix} 1 & * \\ & \boxed{A'} \end{bmatrix} \begin{bmatrix} 1 & \\ & \boxed{W} \end{bmatrix} = \begin{bmatrix} 1 & * \\ & \boxed{\diagdown *} \end{bmatrix}$. □

Note: Thm + pf work/$\mathbb{R}$ if all eigenvalues assumed $\in \mathbb{R}$.

Cor (Spectral thm): $\varphi = \varphi^* \Rightarrow V$ has orthonormal basis of eigenvectors and all eigenvalues $\in \mathbb{R}$.

$\Updownarrow$ $A = A^* \Rightarrow A = UDU^*$ for some unitary U and real diagonal D.

unitarily similar to real diagonal

Pf: Thm $\Rightarrow$ $A = UBU^*$ with B upper-$\triangle$. But

$\|$

$A^* = UB^*U^* \Rightarrow B = B^* \Rightarrow B$ diagonal and real. □

Def: φ is normal if $\varphi\varphi^* = \varphi^*\varphi$. φ commutes with its adjoint. E.g. $\varphi \in \mathcal{U}_n$

Cor 2: φ normal $\Rightarrow V$ has orthonormal basis of eigenvectors.

$\Updownarrow$ A normal $\Rightarrow A = UDU^*$ for some $U \in \mathcal{U}_n$ and diagonal D. ~~real~~

Pf: Suffices by Thm: normal upper-$\triangle$ is diagonal. Same induction as for Thm.

Assume N normal upper-$\triangle$. Check

- $(N^*N)_{11} = \bar{a}_{11}a_{11} = |a_{11}|^2$
- $(NN^*)_{11} = |a_{11}|^2 + \underbrace{|a_{12}|^2 + \cdots + |a_{1n}|^2}_{=0}$

$$N = \begin{bmatrix} a_{11} & \overset{0}{\cancel{a_{12} \cdots a_{1n}}} \\ \begin{matrix}0\\ \vdots \\ 0\end{matrix} & \boxed{\begin{matrix} \ddots & * \\ 0 & \ddots \end{matrix}} \end{bmatrix}$$

upper-$\triangle$

$\Rightarrow$

$\Rightarrow$ $\overset{0}{\cancel{a_{12} \cdots a_{1n}}}$

$\Rightarrow$ N^*N and NN^* computed block by block

$\Rightarrow$ upper-$\triangle$ is normal

$\Rightarrow$ " " diagonal by induction. □

Prop: φ normal $\Leftrightarrow \|\varphi x\| = \|\varphi^* x\|$ $\forall x \in V$.

Pf: $\Rightarrow$: $\|\varphi x\|^2 = \langle \varphi x, \varphi x\rangle = \langle \varphi^*\varphi x, x\rangle = \langle \varphi\varphi^* x, x\rangle = \langle \varphi^* x, \varphi^* x\rangle = \|\varphi^* x\|^2$.

omitted $\Leftarrow$: Compare $\langle \varphi^*\varphi x, y\rangle = \langle \varphi x, \varphi y\rangle$ to $\langle \varphi\varphi^* x, y\rangle = \langle \varphi^* x, \varphi^* y\rangle$ by expressing $\langle\cdot,\cdot\rangle$ in terms of $\|\cdot\|$. □

12.

Positive (semi)definite matrices and singular values

Def: V with $\langle \cdot , \cdot \rangle$ and $\dim_{\mathbb{C}} V = n$. $\mathbb{C}$-linear $\varphi : V \to V$ is self-adjoint if $\varphi = \varphi^*$ and is further

- positive semidefinite if $\langle \varphi x, x \rangle \geq 0 \ \forall x \in V$ "$\varphi \geq 0$" e.g. diagonal and ≥ 0
- positive definite $>$ "$\varphi > 0$" e.g. > 0

Thm: Let $\varphi = \varphi^*$. Then $\varphi \underset{>}{\geq} 0 \Leftrightarrow$ all eigenvalues of φ are $\underset{>}{\geq} 0$.

Pf: Pick orthonormal $\mathcal{B}$ so $[\varphi]_{\mathcal{B}}$ diagonal. Now $[\varphi]_{\mathcal{B}} \underset{>}{\geq} 0 \Leftrightarrow$. □

Cor: $A = A^* \geq 0 \Rightarrow \exists ! \ B \geq 0$ with $B^2 = A$. **Def:** $B = \sqrt{A}$.

Pf: $\exists$: $A = UDU^* \Rightarrow \sqrt{A} = U\sqrt{D}\,U^*$.

$!$: exercise. □

Prop: $A > 0 \Rightarrow \nu(x) = \sqrt{x^*Ax}$ is a norm.

Pf: $\nu(x) = \| A^{1/2} x \|_2$; apply HW2 #3: $\nu = \mu \circ \varphi$ for $\mu = \|\cdot\|_2$ and $\varphi = A^{1/2}$. □

Def: $A \in \mathbb{C}^{m \times n}$ has modulus $|A| = \sqrt{A^*A} \geq 0$. $|A| \in \mathbb{C}^{n \times n}$

Needs: $(A^*A)^* = A^*A$ ✓

$\langle A^*Ax, x \rangle = \langle Ax, Ax \rangle = \|Ax\|^2 \geq 0 \ \forall x \in \mathbb{C}^n$ ✓

Prop: $\| |A| x \| = \|Ax\| \ \forall x \in \mathbb{F}^n$.

Pf: $\| |A|x \|^2 = \langle |A|x, |A|x \rangle = \langle |A|^* |A| x, x \rangle$

$= \langle |A|^2 x, x \rangle$ since $|A|$ hermitian

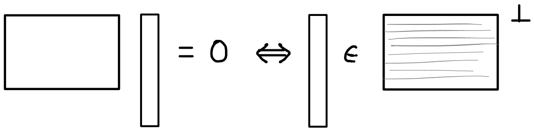
$= \langle A^*A x, x \rangle$ by def. of $|A|$

$= \langle Ax, Ax \rangle = \|Ax\|^2$. □

$$\begin{tikzcd}
Ax \arrow[ddd, mapsto] & & & & x \arrow[llll, mapsto] \arrow[r, mapsto] & A^*Ax \\
& y \arrow[d, mapsto] & Y \arrow[d] & X \arrow[l, "\lambda_A"'] & \varphi^* & \\
& \bar{y}^T & Y^* \arrow[r, "\rho_A"'] & X^* \arrow[u] & \varphi \arrow[u, mapsto] & \\
x^*A^* \arrow[rrrrr, mapsto] & & & & & x^*A^*A \arrow[uuu, mapsto]
\end{tikzcd}$$

$\lambda_{A^*} \leftrightarrow \rho_A$

Cor: $\ker A = \ker |A| = \operatorname{im}(|A|)^{\perp}$.



Pf: $=_1$: $\|Ax\| = 0 \Leftrightarrow \| |A|x \| = 0$.

$=_2$: $\ker T = (\operatorname{im} T^*)^{\perp}$, and $|A|^* = |A|$. □

Goal: geometry of A in terms of $|A|$

Note: Cor $\Rightarrow$

- $|A|$ = orthogonal projection followed by ... some $\cong$ of $\operatorname{im}|A|$ ...
- $\operatorname{im} A \overset{|A|}{\xrightarrow{\sim}} (\ker A)^{\perp}$.

by universal property of quotients:
- $|A|$ is 0 on $\ker A$.
- $(\ker A)^{\perp} = \operatorname{im} |A|$

Def: The singular values of A are the eigenvalues of $|A|$

$\Updownarrow$ $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$, where A^*A has eigenvalues $\lambda_1, \dots, \lambda_n$

Running assumptions $Y \xleftarrow{A} X$ homomorphism of hermitian v.s. / $\mathbb{F}$

- singular values $\underbrace{\sigma_1, \dots, \sigma_r}_{\neq 0}, \underbrace{\sigma_{r+1}, \dots, \sigma_n}_{=0}$ of A $r = \cancel{?}$ rank A
- $v_1, \dots, v_n$ orthonormal basis of X so that $|A|$ scales v_i by σ_i

$\Rightarrow v_{r+1}, \dots, v_n$ " " " $\ker A$.

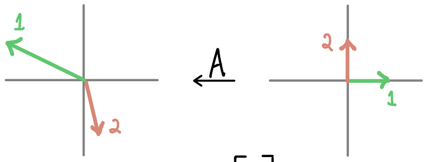
Geometry of $|A|$: kill $\ker A$, rescale $v_1, \dots, v_r$ by positive $\sigma_1, \dots, \sigma_r$.

Thm (polar decomposition) $A \in \mathbb{F}^{n \times n} \Rightarrow A = U|A|$ for some $U \in O_n(\mathbb{F})$ unique if $A \in GL_n$

Pf: Follows from ^to come^ SVD. ☐ $z \in \mathbb{C} \Rightarrow z = e^{i\theta}|z|$ $\theta \in \mathbb{R}$ polar coordinates

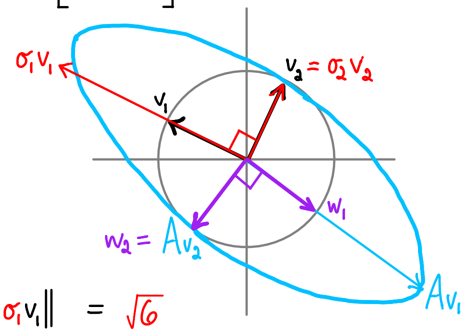
E.g. $A = \begin{bmatrix} -2 & \frac{1}{5}(4-\sqrt{6}) \\ 1 & \frac{1}{5}(-2-2\sqrt{6}) \end{bmatrix} \approx \begin{bmatrix} -2 & .31 \\ 1 & -1.38 \end{bmatrix}$

$$A^*A = \frac{1}{5}\begin{bmatrix} -2 & 1 \\ 4-\sqrt{6} & -2-2\sqrt{6} \end{bmatrix} \frac{1}{5}\begin{bmatrix} -2 & 4-\sqrt{6} \\ 1 & -2-2\sqrt{6} \end{bmatrix} = \begin{bmatrix} 4+1 & -2 \\ \frac{-8+\cancel{2\sqrt{6}}-2-\cancel{2\sqrt{6}}}{5} & \frac{16-\cancel{8\sqrt{6}}+6+4+\cancel{8\sqrt{6}}+24}{25} \end{bmatrix}$$



$$p_{A^*A}(t) = (5-t)(2-t)-4$$
$$= t^2 - 7t + 6 \quad \Leftarrow \quad = \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}$$
$$= (t-6)(t-1)$$

$\sigma_1^2 = 6$: $v_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ $\sigma_2^2 = 1$: $v_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$



$$Av_1 = \frac{1}{\sqrt{5}}\left(\begin{bmatrix} 4 \\ -2 \end{bmatrix} + \frac{1}{5}\begin{bmatrix} 4-\sqrt{6} \\ -2-2\sqrt{6} \end{bmatrix}\right) \quad Av_2 = \frac{1}{\sqrt{5}}\left(\begin{bmatrix} -2 \\ 1 \end{bmatrix} + \frac{2}{5}\begin{bmatrix} 4-\sqrt{6} \\ -2-2\sqrt{6} \end{bmatrix}\right)$$

$$= \frac{1}{5\sqrt{5}}\begin{bmatrix} 24-\sqrt{6} \\ -12-2\sqrt{6} \end{bmatrix} \approx \begin{bmatrix} 1.93 \\ -1.51 \end{bmatrix} \quad = \frac{1}{5\sqrt{5}}\begin{bmatrix} -2-2\sqrt{6} \\ 1-4\sqrt{6} \end{bmatrix} \approx \begin{bmatrix} -.62 \\ -.79 \end{bmatrix}$$

$\|Av_1\| = \|\sigma_1 v_1\| = \sqrt{6}$

$\|Av_2\| = \|\sigma_2 v_2\| = 1$

Lemma: $w_k = \frac{1}{\sigma_k} Av_k \Rightarrow w_1, \dots, w_r$ orthonormal in Y.

Pf: $\langle \sigma_i w_i, \sigma_j w_j \rangle = \langle Av_i, Av_j \rangle = \langle A^*Av_i, v_j \rangle = \langle \sigma_i^2 v_i, v_j \rangle = \sigma_i^2 \langle v_i, v_j \rangle = \begin{cases} 0 & i \neq j \\ \sigma_i^2 & i = j \end{cases}$. ☐

Thm: A has Schmidt decomposition: $A = \sum_{k=1}^{r} \sigma_k w_k v_k^*$ $v_k^*: x \mapsto \langle x, v_k \rangle$, so coeff. on w_k in Ax is $\sigma_k \langle x, v_k \rangle$

Pf: R.H.S. applied to v_i is $\sigma_i w_i = Av_i$ if $i \leq r$ and 0 if $i > r$. ☐

Lemma: $A = \sum_{k=1}^{r} \sigma_k w_k v_k^*$ for orthonormal $v_1, \dots, v_r$ and $w_1, \dots, w_r \Rightarrow v_1, \dots, v_r$ eigenvectors of A^*A.

Pf: $A^* = \sum_{k=1}^{r} \sigma_k v_k w_k^*$, so in A^*A all terms vanish except $\sigma_i^2 v_i \overset{=1}{(w_i^* w_i)} v_i^*$ so this is a Schmidt decomposition

since $\langle w_i^*, w_j \rangle = \delta_{ij}$. But $\boxed{A^*A = \sum_{i=1}^{r} \sigma_i^2 v_i v_i^*} \Leftrightarrow A^*A v_i = \sigma_i^2 v_i \; \forall i$. ☐

Thm: A has a reduced singular value decomposition $A = \tilde{W}\tilde{\Sigma}\tilde{V}^*$ with

(S V D)

- $\tilde{V}^*$ $r \times n$ orthonormal rows; row i = (eigenvector of A^*A with eigenvalue σ_i^2)*
- $\tilde{\Sigma}$ $r \times r$ diagonal, entries $\sigma_1, \dots, \sigma_r$
- $\tilde{W}$ $m \times r$ orthonormal cols

$$\begin{bmatrix} | & & | \\ w_1 & \cdots & w_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} -v_1^*- \\ \vdots \\ -v_r^*- \end{bmatrix}$$

Pf: Schmidt decomposition; rows of $\tilde{V}^*$ by Lemma. ☐

13.

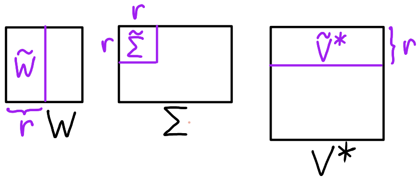
Thm: $A \in \mathbb{F}^{m\times n}$ has SVD $A = W\Sigma V^*$ with

- $V^* \in O_n(\mathbb{F})$
- $\Sigma \in \mathbb{F}^{m\times n}$ all 0 except $\sigma_1, \dots, \sigma_r$ on main diagonal
- $W \in O_m(\mathbb{F})$.

$\Leftrightarrow$ $[\lambda_A]_{\mathcal{V},\mathcal{W}} = \Sigma$ for orthonormal bases $\mathcal{V} = v_1, \dots, v_n$ of $\mathbb{F}^n$ and $\mathcal{W} = w_1, \dots, w_m$ of $\mathbb{F}^m$.

emphasize: simple geometry

Pf: Complete bases in Schmidt decomposition to orthonormal bases of $\ker A$ (for V^*) and $\ker A^*$ (for W). □



Cor: $A \in \mathbb{F}^{n\times n} \Rightarrow A = U|A|$ for some $U \in O_n(\mathbb{F})$.

Pf: $A = W\Sigma V^* = \underbrace{WV^*}\underbrace{V\Sigma V^*}$

$= U \quad |A|$. □

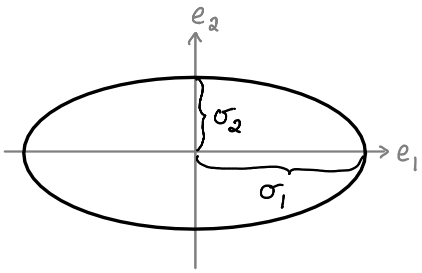
$\boxed{A^*A = \sum_{i=1}^{r} \sigma_i^2 v_i v_i^*} = V\Sigma^2 V^* = \overset{A^*}{V\Sigma \cancel{W^*}}\overset{A}{\cancel{W}\Sigma V^*}$

Note: SVD efficient numerically: fast + accurate

Q. How big can $\|Ax\|$ be, given that $\|x\| = 1$?

$B = \{x \in \mathbb{F}^n \mid \|x\| \le 1\}$ has image = ?

A. If $\Sigma = \operatorname{diag}(\sigma_1 \ge \cdots \ge \sigma_n)$ with $\sigma_{r+1} = \cdots = \sigma_n = 0$ then



$y = \Sigma x$ for $x \in B \Leftrightarrow y_i = \sigma_i x_i$ and $x \in B$

$\Leftrightarrow y_i = \sigma_i x_i$ and $|x_1|^2 + \cdots + |x_n|^2 \le 1$

$\Leftrightarrow y_i = 0$ for $i > r$ and $\left|\frac{y_1}{\sigma_1}\right|^2 + \cdots + \left|\frac{y_r}{\sigma_r}\right| \le 1$ ellipsoid

- $\mathbb{F} = \mathbb{R}$: principal axes of length $2\sigma_1, \dots, 2\sigma_r$ along $e_1, \dots, e_r$
- $\mathbb{F} = \mathbb{C}$: i^{th} real and imaginary principal axes of length $2\sigma_i$.

A arbitrary $\Rightarrow [\lambda_A]_{\mathcal{V},\mathcal{W}} = \Sigma$

$A \quad = W\Sigma V^*$ unitary $\Rightarrow$ do not alter $\|\cdot\|$

Thm: A(unit ball) = ellipsoid in im(A) with principal half-axes along $w_1, \dots, w_r$ of lengths $\sigma_1 \ge \cdots \ge \sigma_r$. □

⊥normal basis

Cor: A has operator norm $\|A\| = \max_{x\in B} \|Ax\| = \sigma_1$. $\Leftrightarrow \|Ax\| \le C\|x\|\ \forall x$, and $C = \|A\|$ is smallest such.

Lemma: $A \mapsto \|A\|$ is a norm on $\mathbb{F}^{m\times n}$.

Pf:
- $\|\alpha A\| = |\alpha|\|A\|$
- $\|A + B\| \le \|A\| + \|B\|$
- $\|A\| \ge 0 \quad \forall A$
- $\|A\| = 0 \Leftrightarrow A = 0$.

Compare Frobenius norm $\|A\|_2 = \sqrt{tr(A^*A)}$
(or Hilbert-Schmidt)

← huh?

$tr(A^*A) = \sum_{i,j} \underbrace{\bar{a}_{ij} a_{ij}}_{|a_{ij}|^2}$

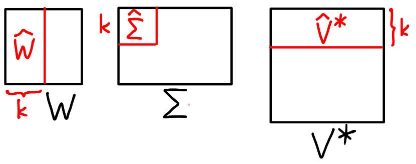
$= $ "$\langle A, A \rangle$"

Prop: $\|A\| \le \|A\|_2$.

Pf: $\|A\| = \sigma_1 \le \sqrt{\sigma_1^2 + \cdots + \sigma_n^2} = \|A\|_2$ since $tr(A^*A) = \sum$ eigenvalues(A^*A). □

application: in computation, statistics, economics, ... often better to have low-rank approximation of A

Eckart-Young Thm: Given $A \in \mathbb{F}^{m\times n}$, the $\hat{A}$ of rank $\leqslant k$ minimizing $\|A-\hat{A}\|_2$ is $\hat{W}\hat{\Sigma}\hat{V}^*$ where



Pf: omitted for time, though we could totally do it

Principal Component Analysis (PCA)

$$A = \begin{array}{c|c|} & n \\ \hline & -A_1- \\ m & \vdots \\ & -A_m- \\ \hline \end{array} \longleftrightarrow m \text{ points in } \mathbb{F}^n$$

sample

number of features

sample size

weight
height
temperature
death rate
response rate to drug
or other stimulus
⋮

unit vector

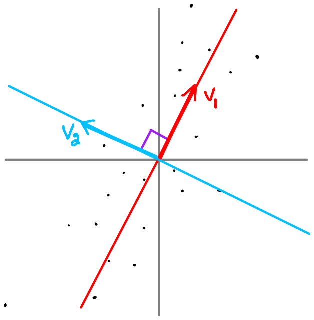
PC 1 = direction $v_1 \in \mathbb{F}^n_{col}$ maximizing sample variance: $|A_1v_1|^2+\cdots+|A_mv_1|^2$

$\hat{A}_1$ = project rows of A orthogonally to $v_1^\perp$

PC 2 = direction $v_2 \in v_1^\perp \subseteq \mathbb{F}^n_{col}$ maximizing $\hat{A}_1$-sample variance

$\hat{A}_2 = \hat{A}_1 / v_2$

⋮



score matrix

Def: The PC decomposition of A is $T = AV$, where V has columns $v_1,\ldots,v_n$.

ij entry is score of sample i along PC j.

Interpretation: $\text{cols}(V) \longleftrightarrow$ alternative features

- linear combinations of original features
- explain variance in uncorrelated ($\perp$) way

Thm: $A = W\Sigma V^* \Rightarrow$ $v_1,\ldots,v_n$ are the columns of V and

$T = W\Sigma$ is polar decomposed

Pf: (sample variance in direction v) $= \|Av\|^2$ for $v \in \mathbb{F}^n_{col}$.

v maximizes $\|Av\|^2 \Leftrightarrow v \overset{A}{\mapsto}$ longest principal axis! (by Cor: $\|A\| = \sigma_1$)

$\Rightarrow V$ is SVD: $\text{cols}(V)$ $\perp$normal basis of eigenvectors of A^*A (by induction)

$\Rightarrow T = AV = W\Sigma V^*V = W\Sigma$. □

PCA $\rightsquigarrow$ low-rank projection of data: use only PC 1, ..., PC k

$$\mathbb{F}^n \xrightarrow{\perp} \mathbb{F}^k$$

variation in directions PC k+1, ..., PC n is small

Note: PC 1, PC 2, ..., PC n $\rightsquigarrow$ flag of best approximating subspaces.

14.

Perturbation theory

Vague:
- has >1 meaning
- different classes of A ⇝ different behavior
- looking for continuity, diff'ability, bounds, ...

Given $A \in \mathbb{C}^{n\times n}$, how do $\Lambda(A) = \underline{\text{spectrum}}$ of $A = \overset{\text{multiset}}{\{\text{roots of } p_A\}}$ and $\Lambda(\tilde{A})$ relate if $\tilde{A} = A + E$ with $\|E\| < \varepsilon$?

E.g. $A = 0 \Rightarrow |\lambda| \leq \|E\|$ for $\lambda \in \Lambda(\tilde{A}) = \Lambda(E)$.

Works for operator norm $\|\cdot\|$. What about other choices ?

Def: $V: \mathbb{C}^{m\times k} \quad \mathbb{C}^{k\times n} \quad \mathbb{C}^{m\times n}$

norms $\mu \quad \nu \quad \rho$ are consistent if $\rho(AB) \leq \mu(A)\nu(B) \quad \forall A, B$

$\mu = \nu = \rho$ on $\mathbb{C}^{n\times n}$: ν is consistent

E.g.
- $\|\cdot\|_2 = \sqrt{tr(A^*A)}$ is consistent (HW 4)
- $\nu_\infty(A) = \max\limits_{i,j} |a_{ij}|$ norm, but $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \Rightarrow \nu_\infty(A^2) = 2 > 1 = \nu_\infty(A)\nu_\infty(A)$ not consistent

Prop: $\|\cdot\|$ consistent on $\mathbb{C}^{n\times n} \Rightarrow \exists$ norm ν on $\mathbb{C}^n$ consistent with $\|\cdot\|$: $\nu(Ax) \leq \|A\|\nu(x)$

Pf: Fix $v \in \mathbb{C}^n \setminus \{0\}$. Set $\nu(x) = \|\underset{n\times n \text{ matrix}}{xv^T}\|$.

- ν is a norm by HW 2.3 : $\mathbb{C}^n \hookrightarrow \mathbb{C}^{n\times n} \xrightarrow{\|\cdot\|} \mathbb{C}$ via $x \mapsto xv^T \mapsto \|xv^T\|$.
- ν consistent with $\|\cdot\|$: $\nu(Ax) = \|Axv^T\| \leq \|A\|\,\|xv^T\| = \|A\|\nu(x)$. □

Def: $A \in \mathbb{C}^{n\times n}$ has spectral radius $\rho(A) = \max\{|\lambda| \mid \lambda \in \Lambda(A)\}$.

Thm: $\|\cdot\|$ consistent on $\mathbb{C}^{n\times n} \Rightarrow \rho(A) \leq \|A\| \quad \forall A \in \mathbb{C}^{n\times n}$.

Pf: Pick ν consistent with $\|\cdot\|$ by Prop. If $\lambda \in \Lambda(A)$ and $Ax = \lambda x$ then

$|\lambda|\cancel{\nu(x)} = \nu(\lambda x) = \nu(Ax) \leq \|A\|\cancel{\nu(x)} \quad x \neq 0$ so $|\lambda| \leq \|A\|$. □

E.g. $A = 0 \Rightarrow |\lambda| \leq \|E\|$ for $\lambda \in \Lambda(\tilde{0})$

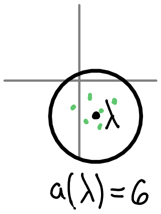
$\tilde{0} = 0 + E = E$, so $\|E\| \sim 10^{-8}$ (say) $\Rightarrow \rho(\tilde{0}) < \sim 10^{-8}$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ & 0 & 1 & 0 \\ & & 0 & 1 \\ & & & 0 \end{bmatrix} \quad E = \begin{bmatrix} & & & \\ & & 0 & \\ & & & \\ \varepsilon & & & \end{bmatrix} \quad \text{so} \quad \tilde{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \varepsilon & 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{array}{l} \Lambda(\tilde{A}) = \{\pm\varepsilon^{1/4}, \pm i\varepsilon^{1/4}\} \\ \varepsilon \sim 10^{-8} \Rightarrow \rho(\tilde{A}) \sim 10^{-2} \end{array}$$

different behavior. Nonetheless:

Thm: Locations of eigenvalues are continuous under perturbation:

if $\lambda \in \Lambda(A)$ has algebraic multiplicity $a(\lambda) = m$, $\|\cdot\|$ any norm, and $\varepsilon \ll 1$, then $\exists\, \delta > 0$ such that $\|E\| < \delta \Rightarrow B_\varepsilon(\lambda) \supseteq$ exactly m eigenvalues of $\tilde{A} = A + E$.



$a(\lambda) = 6$

□

Pf uses:

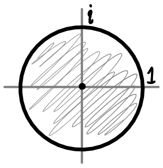
Rouché's thm: Suppose $\Omega \subseteq \mathbb{C}$ open and $\phi, f : \overline{\Omega} \to \mathbb{C}$. Assume

Math 333

- f analytic on $\overline{\Omega}$ (Taylor series at $z \to f(z)$ $\forall z \in \overline{\Omega}$) as is ϕ e.g. f, ϕ polynomials
- $\partial\Omega$ is a simple closed curve (from Math 222: $\simeq S^1$)
- $|\phi(z)| < |f(z)|$ $\forall z \in \partial\Omega$.

Then f and $f+\phi$ have the same #roots in Ω, counted with multiplicity.

E.g. $f(z) = z^n$ on $\overline{\Omega} = B_1(0)$



$\phi(z) = \varepsilon z^j$ for any $j \in \mathbb{N}$

$\Rightarrow z^n + \varepsilon z^j$ has n roots in Ω whenever $\varepsilon < 1$.

General: $z^n + \varepsilon_j z^j + \cdots + \varepsilon_1 z + \varepsilon_0$ has $\underbrace{\text{exactly } n}_{\text{all, if } j \leq n}$ roots in Ω if $\sum |\varepsilon_k| < 1$.

Can be used to prove $\mathbb{C} = \overline{\mathbb{C}}$

Lemma: $A \mapsto p_A$ (char poly) is continuous function $\mathbb{C}^{n \times n} \to P_n = \{\text{polynomials of degree} \leq n\}$.

Pf: coeffs of p_A are polynomial functions of the entries a_{ij}. ☐

Pf of Continuity Thm: Choose ε so $\overline{\Omega} = \overline{B_\varepsilon}(\lambda)$ has no eigenvalues of A other than λ.

Lemma $\Rightarrow p_{\tilde{A}} \to p_A$ as $\tilde{A} \to A$

$\Rightarrow \underbrace{p_{\tilde{A}} - p_A}_{\phi_E(z)} \to 0$ for $z \in \overline{\Omega} \supseteq \partial\Omega$ as $E \to 0$.

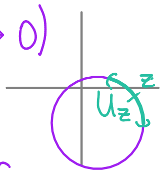
$\partial\Omega$ compact $\underset{\substack{f(z)=p_A(z) \\ \text{not } 0 \text{ on } \partial\Omega}}{\Longrightarrow}$ $|f(z)|$ bounded away from 0: achieves min $\alpha \neq 0$ at $z_0 \in \partial\Omega$.

$\partial\Omega$ compact (closed + bounded) $\Rightarrow$ $\exists\, \delta > 0$ with $|\phi_E(z)| < \alpha$ whenever $\|E\| < \delta$.

Rouché's thm $\Rightarrow f + \phi_E = p_{\tilde{A}}$ has same #roots in Ω as f does. ☐

Detail: compact $\Leftrightarrow$ every open cover has finite subcover [Heine-Borel]

- for each $z \in \partial\Omega$ pick δ_z with $|\phi_E(z)| < \alpha$ whenever $\|E\| < \delta_z$ (since $\phi_E(z) \to 0$)
- ϕ_E continuous $\Rightarrow$ $\phi_E(w) < \alpha$ $\forall w \in$ open nbd U_z of z whenever $\|E\| < \delta_z$
- $\partial\Omega$ compact $\Rightarrow$ finitely many U_z cover; take δ = min of corresponding δ_z.



Maybe the δ_z accumulate around 0, but:

15.

<u>Def</u>: $d(\tilde\lambda, \Lambda)$ = distance from $\tilde\lambda \in \mathbb{C}$ to closed (here, finite) Λ: $\min_{\lambda\in\Lambda} |\tilde\lambda - \lambda|$.

Fix $A \in \mathbb{C}^{n\times n}$ and $\tilde A = A + E$.

1. <u>spectral variation</u> $sv_A(\tilde A) = \max_{\tilde\lambda \in \Lambda(\tilde A)} d(\tilde\lambda, \Lambda(A))$ each $\tilde\lambda$ has a closest $\lambda(\tilde\lambda) \in \Lambda(A)$; $sv_A(\tilde A) = \max_{\tilde\lambda} |\tilde\lambda - \lambda(\tilde\lambda)|$

not symmetric $\Rightarrow$ not a metric: $n = 2$, $\lambda_1 = \tilde\lambda_1 = \tilde\lambda_2 = 0$ and $\lambda_2 = 1 \Rightarrow sv_A(\tilde A) = 0$ but $A \neq \tilde A$.

geometric interpretation: $\Lambda(\tilde A) \subseteq \bigcup_{i=1}^{n} D_i$ where $D_i = \{z \in \mathbb{C} \mid |z - \lambda_i| \le sv_A(\tilde A)\}$

$\uparrow$ so symmetrize:

How much must Λ be fattened to swallow $\tilde\Lambda$?

2. <u>Hausdorff distance</u> $hd(A, \tilde A) = \max\{sv_A(\tilde A), sv_{\tilde A}(A)\}$.

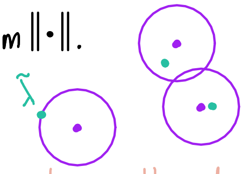
3. <u>matching distance</u> $md(A, \tilde A) = \min_{\text{permutations } \pi\in S_n} \{\max_i |\tilde\lambda_{\pi(i)} - \lambda_i|\}$ = min length(longest edge in perfect matching)

<u>Lemma</u> (Hadamard's inequality): $|\det A| \le \prod_{i=1}^{n} \|\text{col}_i \text{ of } A\|_2$. vol(parallelepiped) $\le \prod \|\text{edges}\|_2$

<u>Pf</u>: True for A upper-Δ, and both sides unchanged by $A \mapsto UA$ for unitary U.

Use Schur decomposition. ☐

<u>Elsner's Thm</u>: $hd(A, \tilde A) \le \left(\|A\| + \|\tilde A\|\right)^{1-1/n} \|E\|^{1/n} = \beta$ for operator norm $\|\cdot\|$.



<u>Pf</u>: β symmetric in A and $\tilde A$, so need only $sv_A(\tilde A) \le \beta$.

Suppose $sv_A(\tilde A) = d(\tilde\lambda, \Lambda)$. So $\tilde\lambda$ is the last eigenvalue of $\tilde A$ to be swallowed.

Pick $\perp$normal basis $x_1, \dots, x_n$ with $\tilde A x_1 = \tilde\lambda x_1$. Then

$$sv_A(\tilde A)^n \le \prod_{\lambda\in\Lambda} |\tilde\lambda - \lambda| \qquad d(\tilde\lambda, \Lambda) \le |\tilde\lambda - \lambda| \quad \forall\, \lambda \in \Lambda$$

$$= |\det(A - \tilde\lambda I)| \qquad \det = \prod \text{eigenvalues and } \lambda \in \Lambda(A) \Leftrightarrow \tilde\lambda - \lambda \in \Lambda(A - \tilde\lambda I)$$

$$\le \prod_{i=1}^{n} \|(A - \tilde\lambda I) x_i\|_2 \quad \text{by Lemma}$$

$$(A - \tilde A)x_1 \qquad = \|\underbrace{(A - \tilde\lambda I)x_1}\|_2 \prod_{i=2}^{n} \|(A - \tilde\lambda I) x_i\|_2$$

$$\le \|E x_1\|_2 \prod_{i=2}^{n} \left(\|A x_i\|_2 + \|\tilde\lambda x_i\|_2\right)$$

$$\le \|E\| \left(\|A\| + \|\tilde A\|\right)^{n-1}. \quad \text{Take } n^{\text{th}} \text{ root.} \ ☐$$

<u>E.g.</u> (**)

$$\begin{bmatrix} 1 & 10^{-4} \\ 10^{-4} & 2 \end{bmatrix} = \begin{bmatrix} 1 & \\ & 2 \end{bmatrix} + \begin{bmatrix} & 10^{-4} \\ 10^{-4} & \end{bmatrix} \Rightarrow hd(A, \tilde A) \le (2 + \underbrace{2 + 10^{-8}}_{\sqrt{\max \Lambda(\tilde A^T \tilde A)}})^{1/2} (10^{-4})^{1/2} < (2 + 10^{-7}) \cdot 10^{-2} < .021$$

$$p_{\tilde A}(z) = z^2 - 3z + 2 - 10^{-8} \qquad \Rightarrow \tilde\Lambda \subseteq [.979, 1.021] \cup [1.979, 2.021]$$

$$\tfrac{3}{2} \pm \tfrac{1}{2}\sqrt{1 + 4\cdot 10^{-8}} = \tfrac{3}{2} \pm \tfrac{1}{2} \pm 10^{-8} \mp O(10^{-16}) \qquad \tilde\Lambda = \{1 - \varepsilon, 2 + \varepsilon\} \text{ for } \varepsilon = 10^{-8} + O(10^{-16})$$

$\Rightarrow$ pretty bad bound

<u>Thm</u> (Ostrowski, Elsner): $md(A, \tilde A) \le n\beta$. Pf omitted.

used in Pf of Gerschgorin

Bauer–Fike Thm: $\|\cdot\|$ consistent on $\mathbb{C}^{n\times n}$ and $\tilde\lambda \in \tilde\Lambda \setminus \Lambda \Rightarrow \underbrace{\underbrace{\|\underbrace{(A-\tilde\lambda I)}_{\sim 0}^{-1}\|}_{\text{big}}^{-1}}_{\sim 0 \quad \text{how small?}} \leq \|E\|$.

Pf: $\underbrace{\tilde A - \tilde\lambda I}_{\text{singular}} = A - \tilde\lambda I + E = \underbrace{(A-\tilde\lambda I)}_{\text{invertible}}\underbrace{(I + (A-\tilde\lambda I)^{-1}E)}_{\Rightarrow\ \text{singular}}$

$$\Rightarrow 1 \leq \|(A-\tilde\lambda I)^{-1}E\| \qquad (*)$$

$$\leq \|(A-\tilde\lambda I)^{-1}\|\,\|E\|. \quad \square$$

"$\tilde\lambda$ is close to some $\lambda\in\Lambda$, bounded by norm of E."

E.g. $A = \operatorname{diag}(\lambda_1,\dots,\lambda_n) \Rightarrow (A-\tilde\lambda I)^{-1}$ has "max entry" $(\lambda-\tilde\lambda)^{-1}$, so thm $\Rightarrow \lambda - \tilde\lambda$ can't be too big.

Gerschgorin's Thm: For $A\in\mathbb{C}^{n\times n}$, let $\alpha_i = \sum_{j\neq i} |a_{ij}|$ and $G_i(A) = B_{\alpha_i}(a_{ii}) = \{z\in\mathbb{C} \mid |z-a_{ii}| \leq \alpha_i\}$.

Then $\Lambda \subseteq \bigcup_{i=1}^{n} G_i(A)$. Equivalently, $\lambda\in\Lambda \Rightarrow \lambda\in G_i(A)$ for some i.

Pf: Fix $\lambda\in\Lambda$. $\lambda = a_{ii} \Rightarrow \lambda\in G_i$, so assume $\lambda\neq a_{ii}$ $\forall i$.

In Bauer–Fike thm view A as perturbation of $D = \begin{bmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn}\end{bmatrix}$, so $A = D + E$ $(E = A - D)$:

$$(*) \Rightarrow \underbrace{1 \leq \|(D-\lambda I)^{-1}E\|_\infty} = \max_i |a_{ii}-\lambda|^{-1}\sum_{j\neq i}|a_{ij}|, \quad \text{where } \|C\|_\infty = \max_i \sum_{j=1}^{n}|c_{ij}|.$$

$$\Rightarrow |a_{ii}-\lambda| \leq \sum_{j\neq i}|a_{ij}| \text{ for some } i. \quad \square$$

Def: $G_i(A) = i^{th}$ Gerschgorin disk.

Note: More is true: $G_{i_1}\cup\cdots\cup G_{i_m}$ disjoint from the other $n-m$

$\Rightarrow \#\Lambda\cap(G_{i_1}\cup\cdots\cup G_{i_m}) = m$. HW4 (apply continuity)

E.g. (**) $\begin{bmatrix} 1 & 10^{-4} \\ 10^{-4} & 2\end{bmatrix} = \begin{bmatrix} 1 & \\ & 2\end{bmatrix} + \begin{bmatrix} & 10^{-4} \\ 10^{-4} & \end{bmatrix} \Rightarrow \tilde\Lambda \subseteq [.9999, 1.0001] \cup [1.9999, 2.0001]$

better, but still 4 orders of magnitude off.

Trick: $\tilde A \sim \begin{bmatrix} \gamma & \\ & 1\end{bmatrix} \tilde A \begin{bmatrix} \gamma^{-1} & \\ & 1\end{bmatrix} = \begin{bmatrix} 1 & 10^{-4}\gamma \\ 10^{-4}\gamma^{-1} & 2\end{bmatrix}$

has same $\tilde\Lambda$ but different $G_i(\tilde A)$!

Choose γ small but with $10^{-4}\gamma + 10^{-4}\gamma^{-1} < 1$ so $\tilde G_1 \cap \tilde G_2 = \emptyset$.

Suffices: $\gamma^{\pm 1} = 10^{-4} + 10^{-11} \Rightarrow \tilde\Lambda \subseteq [1-\delta, 1+\delta] \cup [2-\delta, 2+\delta]$

for $\delta = 10^{-8} + 10^{-15}$ Quite sharp!

Check: $10^{-4}(1+10^{-7})^{-1} = 10^{-4}(1 - 10^{-7} + 10^{-14} - \cdots)$

$\Rightarrow 10^{-4}\gamma + 10^{-4}\gamma^{-1} = (10^{-8} + 10^{-15}) + (1 - 10^{-7} + 10^{-14} - \cdots)$

$< 1 - 10^{-7} + 2\cdot 10^{-8}$. $\square$

16. Lie algebras

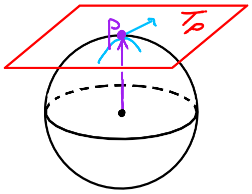
Def: For $X \subseteq$ vector space $V/\mathbb{R}$, the tangent space at $p \in X$ is

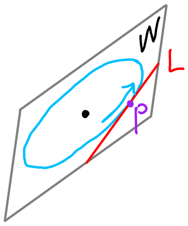
$$T_pX = \{\, \gamma'(0) \mid \gamma : (-\varepsilon, \varepsilon) \to X \text{ differentiable with } \gamma(0) = p \,\}.$$

= initial velocities of differentiable paths in X through p

E.g. $X = S^{n-1} \subseteq \mathbb{R}^n$. $\gamma(t) \subseteq S^{n-1} \Leftrightarrow \gamma(t)\cdot\gamma(t) = 1$

$$\begin{aligned} \overset{\parallel}{\gamma_1(t)^2 + \cdots + \gamma_n(t)^2} &\Rightarrow 2\gamma_1(t)\gamma_1'(t) + \cdots + 2\gamma_n(t)\gamma_n'(t) = 0 \\ &\Leftrightarrow \gamma(t)\cdot\gamma'(t) = 0 \\ &\Leftrightarrow \gamma'(t) \in \gamma(t)^\perp . \end{aligned}$$





Thus $T_pS^{n-1} \subseteq p^\perp$. But $L \perp p \Rightarrow L = T_p(W \cap X)$ for $W = \mathrm{span}(L, p)$.

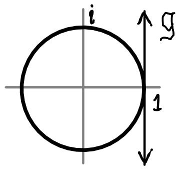
So $T_pS^{n-1} = p^\perp$.

Def: Subgroup $G \subseteq GL_n\mathbb{F}$ that is a manifold has Lie algebra $\mathfrak{g} = T_I G$.

"Lee"

\mathfrak{g}

E.g. • $G = U_1 \cong O_2(\mathbb{R})$



$\Rightarrow \mathfrak{g} = \mathrm{span}_{\mathbb{R}}(i) = i\mathbb{R} = \mathfrak{u}_1$ $\quad \frac{d}{dt}e^{kit}\Big|_{t=0} = ki \quad k \in \mathbb{R}$

$\cong \mathrm{span}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \mathfrak{o}_2$

• $G = GL_n(\mathbb{F}) \Rightarrow \mathfrak{g} = \mathfrak{gl}_n(\mathbb{F}) = M_n(\mathbb{F})$

Pf: $GL_n = M_n \setminus \underbrace{\{\det = 0\}}_{\text{closed}} \Rightarrow GL_n$ open $\Rightarrow T_pGL_n = T_pM_n$.

Def: $A \in M_n(\mathbb{F})$ has (matrix) exponential $\exp(A) = e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots + \frac{A^k}{k!} + \cdots$

convergence issues: sequences in $M_n(\mathbb{F}) \cong \mathbb{R}^m$ for some m, often using norm $\|A\|_2$

Prop: $\gamma : \mathbb{R} \to M_n(\mathbb{F})$ via $\gamma(t) = e^{tA}$ is differentiable with $\gamma'(t) = A\gamma(t) = \gamma(t)A$.

Pf: Termwise and entrywise differentiate

$$e^{tA} = I + tA + t^2\frac{A^2}{2!} + \cdots + t^k\frac{A^k}{k!} + \cdots \text{ to get}$$

$$(e^{tA})' = 0 + A + tA^2 + t^2\frac{A^3}{2!} + \cdots + t^k\frac{A^{k+1}}{k!} + \cdots = Ae^{tA} = e^{tA}A. \ \square$$

Prop: $AB = BA \Rightarrow e^Ae^B = e^{A+B}$.

Pf: Binomial thm $\Rightarrow \frac{(A+B)^k}{k!} = \sum_{i+j=k} \frac{A^i}{i!}\frac{B^j}{j!}$.

These are the deg k terms in e^Ae^B and e^{A+B}. $\square$

Cor: $\exp : \mathfrak{gl}_n(\mathbb{F}) \to GL_n(\mathbb{F})$.

Pf: $e^Ae^{-A} = e^{A-A} = e^0 = I$. $\square$

To compute more examples of $\mathfrak{g}$:

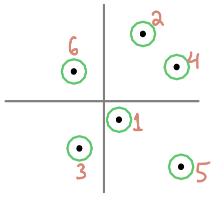
Thm: tr = det′: $\frac{d}{dt}\Big|_{t=0}(\det\gamma(t)) = \mathrm{tr}(\gamma'(0))$ if $\gamma(0) = I$.

Pf: Suppose $\gamma(t)$ is a path in $M_n\mathbb{F}$ with distinct eigenvalues $\lambda_1(t), \dots, \lambda_n(t)$ in $\mathbb{C}$ for $t \neq 0$.

Q. How can consistent choices of numbering $\lambda_1, \dots, \lambda_n$ be made for varying t?

How do we know which eigenvalue of $\gamma(\tilde{t})$ is supposed to correspond to (say) $\lambda_1(t)$?

A. Continuity thm!



$\Lambda(t)$

$\Lambda(\tilde{t}) \subseteq \bigcup \odot$

number however you like

Then $\det\gamma(t) = \lambda_1(t)\cdots\lambda_n(t)$

$$\Rightarrow (\det\gamma(t))' = \lambda_1'(t)\frac{\det\gamma(t)}{\lambda_1(t)} + \cdots + \lambda_n'(t)\frac{\det\gamma(t)}{\lambda_n(t)}$$

$$\Rightarrow (\det'\gamma)(0) = \lambda_1'(0)\cdot 1 + \cdots + \lambda_n'(0)\cdot 1 \quad \text{if } \gamma(0) = I.$$

Even if $|\Lambda(\gamma(t))| < n$, same argument works by replacing $\gamma(t)$ with

$$\gamma_\varepsilon(t) = \gamma(t) + \varepsilon D \text{ for } D = \begin{bmatrix} 1 & & \\ & 2 & \\ & & \ddots_n \end{bmatrix} \text{ and taking } \lim_{\varepsilon\to 0}. \text{ Uses}$$

- $|\Lambda(A)| = n \Rightarrow |\Lambda(A+E)| = n \ \forall \ \|E\| \ll 1$ ← why perturbation theory before Lie algebras

  holds by continuity thm $\Rightarrow |\Lambda(\gamma_\varepsilon(t))| = n \ \forall \ \varepsilon \neq 0$ and $t \ll 1$: $A = I + \varepsilon D$

  $E = \gamma(t) - I$

- $(\det\gamma_\varepsilon)'(0) \xrightarrow{\varepsilon\to 0} (\det\gamma)'(0)$

  holds because $M_n\mathbb{F} \times \mathbb{R} \to \mathbb{F}$

  $A, \varepsilon \mapsto \det(A + \varepsilon D)$ is differentiable (it is polynomial)

- $\gamma_\varepsilon'(t) = \gamma'(t)$

  holds because $\gamma_\varepsilon'(t) = (\gamma(t) + \varepsilon D)' = \gamma'(t) + \cancelto{0}{(\varepsilon D)'} \ \forall \varepsilon, t.$ □

17.

<u>E.g.</u> $G = SL_n(\mathbb{F}) = \{A \in GL_n(\mathbb{F}) \mid \det A = 1\}$

$\Rightarrow \mathfrak{g} = \mathfrak{sl}_n(\mathbb{F}) := \{A \in M_n(\mathbb{F}) \mid \operatorname{tr} A = 0\}$.

<u>Pf</u>: $\det \gamma(t) \equiv 1$ if $\gamma \subseteq SL_n \Rightarrow \mathfrak{g} \subseteq \mathfrak{sl}_n(\mathbb{F})$ by $\operatorname{tr} = \det'$. exercise with exp

For $\supseteq$, let $A \in \mathfrak{sl}_n(\mathbb{F})$. Then $\det e^{tA} = e^{\operatorname{tr}(tA)} = e^0 = 1 \Rightarrow e^{tA} \subseteq SL_n$.

Thus e^{tA} realizes A as $(e^{tA})'\big|_{t=0}$. □

<u>Lemma</u>: $(\gamma(t)^*)' = \gamma'(t)^*$. Transpose and conjugation both commute with derivative.

<u>Prop</u>: $\beta, \gamma : (-\varepsilon, \varepsilon) \to M_n$ differentiable $\Rightarrow (\beta\gamma)' = \beta'\gamma + \beta\gamma'$.

<u>Pf</u>: Entry by entry, sum by sum, this is the usual product rule. □

<u>E.g.</u>
$$\left(\underset{\beta(t)}{\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}}\underset{\gamma(t)}{\begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}}\right)' = \begin{bmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{bmatrix}' = \begin{bmatrix} a_1'a_2 + a_1a_2' + b_1'c_2 + b_1c_2' & \cdots \\ \vdots & \ddots \end{bmatrix}$$
$$= \begin{bmatrix} a_1'a_2 + b_1'c_2 & \cdots \\ \vdots & \ddots \end{bmatrix} + \begin{bmatrix} b_1c_2' + a_1a_2' & \cdots \\ \vdots & \ddots \end{bmatrix}$$

<u>E.g.</u> $G = O_n(\mathbb{F}) \Rightarrow \mathfrak{g} = \mathfrak{o}_n(\mathbb{F}) := \{A \in M_n(\mathbb{F}) \mid A^* = -A\}$.

<u>Pf</u>: $\subseteq$: product rule + Lemma: $\gamma(t) \subseteq O_n(\mathbb{F}) \Rightarrow$

$\gamma(t)\gamma(t)^* \equiv I \Rightarrow 0 = \gamma'(t)\gamma(t)^* + \gamma(t)\gamma'(t)^* \overset{t=0}{=} \gamma'(0)I + I\gamma'(0)^*$.

$\supseteq$: $\Leftrightarrow \exists\, \gamma(t) \subseteq O_n(\mathbb{F})$ with $\gamma'(0) = A$ whenever $A^* = -A$.

Again use $\gamma(t) = e^{tA}$, which has

- $\gamma'(0) = A$ by Prop from last time
- $e^{tA}(e^{tA})^* = e^{tA}e^{tA^*} = e^{tA}e^{-tA} = e^0 = I$. □

<u>Prop</u>: $\dim X = \dim T_pX$ for any p in manifold X.

<u>Pf</u>: $\{\gamma : (-\varepsilon, \varepsilon) \to X_\alpha\} \leftrightarrow \{\gamma : (-\varepsilon, \varepsilon) \to U_\alpha\}$. Use that (diffeomorphism to image)$'$ is injective. □

<u>Cor</u>: $\dim O_n(\mathbb{F}) = \cancel{?}$

- $\mathbb{F} = \mathbb{R}$: $A^T = -A \Rightarrow \boxed{\begin{matrix} 0 & & d \\ & \ddots & \\ O & & 0 \end{matrix}} \Rightarrow n^2 = 2d + n$

  $\Rightarrow d = \frac{1}{2}(n^2 - n) = \binom{n}{2}$.

- $\mathbb{F} = \mathbb{C}$: $A^* = -A \Rightarrow \boxed{\begin{matrix} i\mathbb{R} & & 2d \\ & \ddots & \\ O & & i\mathbb{R} \end{matrix}} \Rightarrow \dim = 2d + n = n^2$.

<u>Crucial question</u>: Why are these subgroups of GL_n manifolds in the first place?

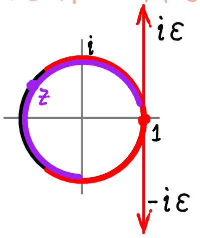
Thm: Fix closed subgroup $G \subseteq GL_n(\mathbb{F})$ with Lie algebra $\mathfrak{g}$. Then

- $A \in \mathfrak{g} \Rightarrow e^A \in G$
- $B_\varepsilon = \{ A \in M_n \mid \|A\|_2 < \varepsilon \} \Rightarrow \exp_\varepsilon : \mathfrak{g} \cap B_\varepsilon \hookrightarrow G$ if $\varepsilon \ll 1$

  neighborhood of I in G
- G is a manifold with atlas $\{ g \cdot \exp_\varepsilon \mid g \in G \}$.

Pf: omitted, though we could do it with enough time. See Math 421, 603, 620

E.g. $G = U_1 = S^1$ $\quad i\varepsilon \in \mathfrak{u}_1$



$$\varepsilon < \pi \Rightarrow \exp_\varepsilon : (-\varepsilon, \varepsilon) \hookrightarrow S^1 \text{ nbd of } 1$$
$$t \mapsto e^{it}$$

$$|z| = 1 \Rightarrow z \cdot \exp_\varepsilon : (-\varepsilon, \varepsilon) \hookrightarrow S^1 \text{ nbd of } z$$
$$t \mapsto ze^{it}$$

Q. Why closed?

A. $\mathbb{Q} \subseteq \mathbb{R}$.

But that's not in GL_1, you complain? O.K. But $(\mathbb{R}, +) = \begin{bmatrix} 1 & \mathbb{R} \\ 0 & 1 \end{bmatrix} \subseteq GL_2\mathbb{R}$ closed subgroup.

$x_{11} = 1$ intersection
$x_{22} = 1$ of polynomial
$x_{21} = 0$ level sets

To connect with previous units:

Thm: closed subgroup $H \subseteq G \Rightarrow G/H$ and $H\backslash G$ are manifolds.

Pf: omitted, and this would take more work; still doable, but not as elementary.

E.g.
- $\mathcal{Fl}_n$ for $G = GL_n$ and $H = B_n^+ =$ upper-$\triangle$
- $G_k(\mathbb{F}^n)$ for $G = GL_n$ and $H =$ block upper-$\triangle$ with block sizes k and $n-k$
- chains $\{V_d \subseteq V_e\}$ for $G = GL_n$ and $H =$ block upper-$\triangle$ with block sizes $d, e-d, n-e$

No need to fiddle with explicit charts.

18.

Perron-Frobenius theory

Def: $A \geq B$ for real A, B of same size if $a_{ij} \geq b_{ij}\ \forall i,j$.
$\succ$ $\qquad$ $A \geq B$ and $a_{ij} > b_{ij}$ for some i,j.

E.g. $P \geq 0 \Leftrightarrow$ entrywise nonnegative
$>$ $\qquad$ positive

Perron's Thm: $P \in \mathbb{R}^{n \times n}$ and $P > 0 \Rightarrow P$ has dominant eigenvalue $\lambda(P)$:

1. $\lambda(P) > 0$ and $Pv = \lambda(P)v$ for some $v > 0$;
2. $a(\lambda(P)) = 1$; algebraic multiplicity 1 and

for $\kappa \in \Lambda(P) \setminus \{\lambda(P)\}$:

3. $|\kappa| < \lambda(P)$ and
4. $Py = \kappa y$ and $y \neq 0 \Rightarrow y \not\geq 0$.

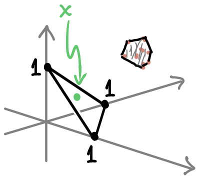
today: proof
next time: consequences

Pf: Set $L(P) = \{\lambda \geq 0 \mid Px \geq \lambda x$ for some $x \overset{\geq}{\cancel{>}} 0$ and $\mathbb{1}x = 1\}$.

$P\sigma$ is some polytope in $\mathbb{R}^n_+$
$\lambda \gg 0 \Rightarrow P\sigma$ entirely under $\lambda\sigma$
$\lambda \ll 1 \Rightarrow \lambda\sigma$ sufficiently under $P\sigma$

Let $\mathbb{1} = [1 \cdots 1]$ so $\mathbb{1}x = \|x\|_1 = x_1 + \cdots + x_n$.

$Px \geq \lambda x \Rightarrow$ same for $\frac{x}{\|x\|_1} \Rightarrow$ assume $\mathbb{1}x = 1$.



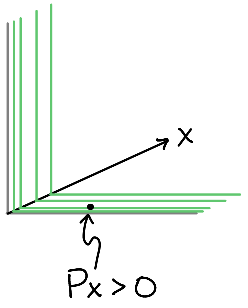
Lemma: $L(P)$ compact and has some $\lambda > 0$.

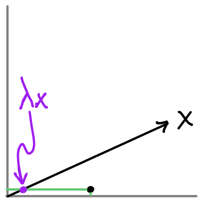
Pf: $x \in \mathbb{R}^n$ and $0 \neq x \geq 0 \Rightarrow Px > 0$.

$\lambda \to 0_+ \Rightarrow \lambda x \to 0 \Rightarrow \lambda x < \varepsilon \mathbb{1}^T$ eventually

$\Rightarrow \lambda x < Px$ $\qquad$ "

$\Rightarrow \lambda \in L(P)$ $\qquad$ " $\qquad$.





bounded: $b = b\mathbb{1}x \geq \mathbb{1}Px \geq \mathbb{1}\lambda x = \lambda\mathbb{1}x = \lambda$ as $\mathbb{1}x = 1$; $b = \|\mathbb{1}P\|_\infty \Rightarrow b = b\mathbb{1}x \geq$
$= $ max entry of $\underbrace{\mathbb{1}P}_{\Sigma \text{rows of } P}$

closed: $\lambda_k \to \lambda$ with $\lambda_k \in L(P)\ \forall k \in \mathbb{N}$

$\Rightarrow \exists x_k$ with $Px_k \geq \lambda_k x_k\ \forall k$; may as well assume $\mathbb{1}x_k = 1$.

$\{x_k\}_{k \in \mathbb{N}}$ has convergent subsequence since σ_{n-1} compact

$\sigma_{n-1} = \{x \in \mathbb{R}^n_{\geq 0} \mid \mathbb{1}x = 1\}$
simplex

$\Rightarrow$ can replace $\{\lambda_k\}_{k \in \mathbb{N}}$ and $\{x_k\}_{k \in \mathbb{N}}$ with subsequences to assume

$\lambda_k \to \lambda$ and $x_k \to x$

$\Rightarrow \lim\limits_{k \to \infty} (Px_k \geq \lambda_k x_k)$ is $(Px \geq \lambda x) \Rightarrow \lambda \in L(P)$. $\square$

As $\lambda\sigma$ grows, its λx's lie beneath fewer Px's. $\lambda(P)$ is when the last Px works.

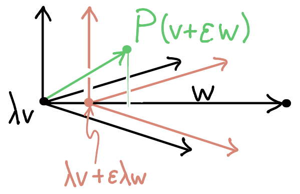
1. Set $\lambda(P) = \max L(P)$. Lemma $\Rightarrow \lambda(P) > 0$.

Claim: $\lambda(P) \in \Lambda(P)$. In fact, $Pv \geq \lambda(P)v$ for $v \geq 0 \Rightarrow Pv = \lambda(P)v$.

<u>Pf</u>: Suppose $\lambda \in L(P)$, so $Pv \geq \lambda v$ for some $v \geq 0$. P moves $v+\varepsilon w$ to interior of $\mathbb{R}^n_{\geq \lambda v}$

Want: this $\rightsquigarrow$ $Pv \neq \lambda v \Rightarrow 0 \neq Pv - \lambda v =: w$

$\Rightarrow \lambda \neq \lambda(P)$



$Pv > 0 \Rightarrow$ everything along $\xrightarrow{w}$ is > 0

$$\Rightarrow \varepsilon P w > 0 \ \forall \varepsilon > 0, \text{ since } P > 0$$

$$\Rightarrow P(v+\varepsilon w) = Pv + \varepsilon P w > Pv = \lambda v + w$$

$$\geq \lambda v + \varepsilon \lambda w = \lambda(v+\varepsilon w) \text{ for } \varepsilon \leq \frac{1}{\lambda}.$$

So $x = v + \varepsilon w \Rightarrow Px > \lambda x$

$$\Rightarrow Px \geq \lambda' x \text{ for any } \lambda' > \lambda \text{ with } \lambda' - \lambda \ll 1$$

$$\Rightarrow \lambda \neq \lambda(P) \text{ by maximality. } \square$$

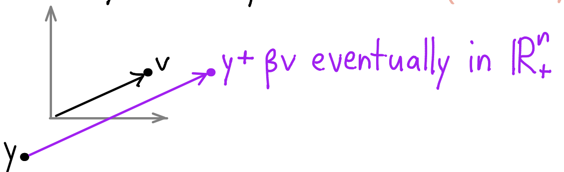
$\lambda(P) \in \Lambda(P) \Rightarrow Pv = \lambda(P)v$ for nonzero $v \geq 0 \overset{P>0}{\Rightarrow} \lambda(P)v > 0 \overset{\lambda(P)>0}{\Rightarrow} \underbrace{v > 0}.$

2. $g(\lambda(P)) = 1$: $w \in E(\lambda(P))$ indep. of $v \Rightarrow$ line $\overleftrightarrow{vw}$ exits $\mathbb{R}^n_+ \Rightarrow E(\lambda(P)) \cap \partial\mathbb{R}^n_+ \setminus \{0\} \neq \emptyset \ \rightarrow\leftarrow$

For $a(\lambda(P)) = 1$ need: no $y \in \mathbb{R}^n$ with $Py = \lambda(P)y + \alpha v$ (∗) $\quad (P - \lambda(P))y \in \operatorname{span}(v)$

By $y \rightsquigarrow -y$ assume $\alpha > 0$

$y \rightsquigarrow y + \beta v$ assume $y > 0$.



(∗) and $v > 0 \Rightarrow Py > \lambda(P)y \Rightarrow Py > \lambda' y$ for any $\lambda' > \lambda(P)$ with $\lambda' - \lambda(P) \ll 1 \ \overset{\lambda(P)\text{ maximal}}{\rightarrow\leftarrow}$

3. $\kappa \in \Lambda(P)$ with $Py = \kappa y$, both $/\mathbb{C}$

$$\Rightarrow p_{i1}y_1 + \cdots + p_{in}y_n = \kappa y_i \Rightarrow p_{i1}|y_1| + \cdots + p_{in}|y_n| \overset{(**)}{\geq} |p_{i1}y_1 + \cdots + p_{in}y_n| = |\kappa||y_i|$$

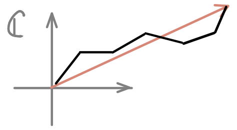
$$\Rightarrow |\kappa| \in L(P) \Rightarrow |\kappa| \leq \lambda(P). \text{ But}$$

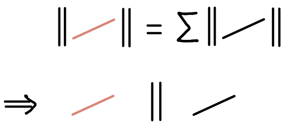
$|\kappa| = \lambda(P) \Rightarrow \begin{bmatrix} |y_1| \\ \vdots \\ |y_n| \end{bmatrix} \in E(\lambda(P))$ by Claim $\Rightarrow \ = \alpha v$ and "=" in (∗∗)

$\Rightarrow y_1, \dots, y_n$ lie along a ray in $\mathbb{C}$

$y_i = \omega|y_i| \ \forall i$ for some $\omega \in U_1$

$\Rightarrow \omega\alpha v \in E(\lambda(P)) \Rightarrow \kappa = \lambda(P).$





4. $P > 0 \Rightarrow P^T > 0 \Rightarrow \exists\, \varphi^T > 0$ in $E(\lambda(P^T))$ want this for $Py = \kappa y$

$$\Rightarrow y \ngtr 0 \text{ if } \varphi y = 0.$$

$Py = \kappa y$ and $P^T\varphi^T = \lambda\varphi^T \Rightarrow \varphi P = \lambda\varphi \Rightarrow \lambda\varphi y = \varphi P y = \varphi(\kappa y) = \kappa\varphi y$

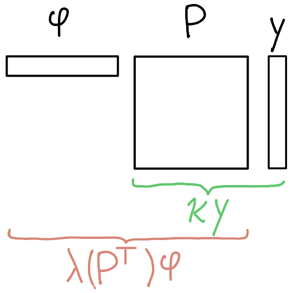
$$\Rightarrow \varphi y = 0 \text{ if } \lambda \neq \kappa$$

Take $\lambda = \lambda(P^T)$

<u>Lemma</u>: $= \lambda(P)$.

<u>Pf</u>: $(P - \lambda I)^T = P^T - \lambda I$. $\square$

$\lambda(P^T)\varphi y =$



$= \varphi(\kappa y)$

$= \kappa\varphi y$

$\square$

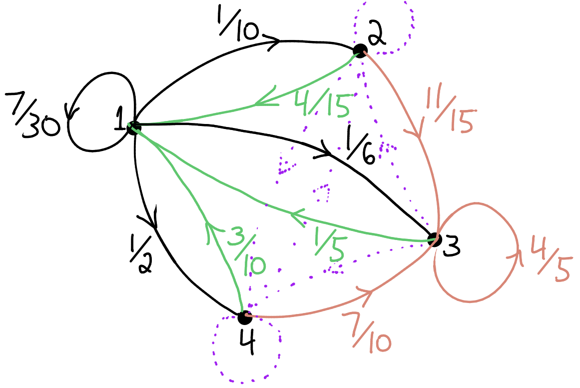
19.

General: $A \in F^{n\times n} \longleftrightarrow$ labeled directed graph

$e_j \mapsto \sum_{i=1}^{n} a_{ij} e_i \longleftrightarrow$ labels $a_{1j}, \dots, a_{nj}$ on edges exiting vertex j

$\nwarrow$ j^{th} column of A

E.g.



$$\begin{bmatrix} 7/30 & 4/15 & 1/5 & 3/10 \\ 1/10 & 0 & 0 & 0 \\ 1/6 & 11/15 & 4/5 & 7/10 \\ 1/2 & 0 & 0 & 0 \end{bmatrix}$$

Def: $P \geq 0$ stochastic if all col sums $= 1$ $(\mathbb{1}P = \mathbb{1})$ $(p_{1j} + \cdots + p_{nj} = 1 \ \forall j)$ $11/15 \ 4/5 \ 7/10$

edge labels $\longleftrightarrow$ transition probabilities or fractions:

really the same
- How much of the stuff at j moves to i ?
- What chance does the thing at j have of moving to i ? finite Markov chain

instead of just one item, think of huge # of trials: place token and play

vector $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $\longleftrightarrow$ where the tokens got placed in your trials

$\longleftrightarrow$ stuff

Q. Iterate $x \mapsto Px \mapsto PPx \mapsto \cdots$

$x = e_j$: where does token sit after k iterations? $\mathbb{P}(\text{token at } i \mid \text{started at } j)$

$x = \text{arbitrary} \geq 0$: how much stuff is at i after k iterations?

E.g. PageRank : n webpages,

$p_{ij} = .85(\text{fraction of links } j \to i)$ $\Rightarrow$ where are you likely to be

$+15\%$ go somewhere random after k iterations? Rank by $\mathbb{P}$.

$\Rightarrow P > 0$

Thm: $P > 0$ stochastic $\Rightarrow$
- $\lambda(P) = 1$ and
- $\forall\, 0 \neq x \geqslant 0$, $P^k x \to \alpha v$ for some $\alpha > 0$, where $Pv = v > 0$.

Pf: $P > 0 \Rightarrow P^T > 0 \Rightarrow$ same $\overset{\text{dominant}}{\text{eigenvalues}}$ by Lemma. But

$\mathbb{1}P = \mathbb{1} \Rightarrow \mathbb{1}$ is (unique!) dominant eigenvector of P^T by Perron's Thm.

$\Rightarrow \lambda(P^T) = 1 = \lambda(P).$

For P^k, first assume P diagonalizable, so

$$x = \sum_{i=1}^{n} \alpha_i v_i \quad \text{for} \quad P v_i = \lambda_i v_i$$

$$\Rightarrow \quad P^k x = \sum_{i=1}^{n} \alpha_i \lambda_i^k v_i . \qquad \lambda(P) \text{ dominant} \Rightarrow \text{all other } |\lambda_i| < 1$$

$$\Rightarrow \text{all summands} \to 0 \text{ except } \lambda(P) \text{ term}$$

$\Rightarrow$ $P^k x \to \alpha v$ for some α.

Why $\alpha > 0$?

$\mathbb{1}P = \mathbb{1} \Rightarrow \mathbb{1}P^k = \mathbb{1}$

$$\Rightarrow \mathbb{1}x = \mathbb{1}P^k x \to \mathbb{1}\alpha v = \alpha \mathbb{1} v$$

$$\Rightarrow \alpha \leftarrow \frac{\mathbb{1}P^k x}{\mathbb{1} v} > 0 \quad \text{because } 0 \neq x \geqslant 0 \text{ and } v > 0.$$

For P not diagonalizable: HW 5. □

Cor: Markov chain transition probabilities all > 0

$\Rightarrow$ convergence to unique stationary distribution.

computable by iterative methods from any $x > 0$!

Frobenius Thm: $M \in \mathbb{R}^{n \times n}$ and $M \geqslant 0$ $\Rightarrow$ $\exists\ \lambda(M) \in \Lambda(M)$ satisfying

1. $\lambda(M) \geqslant 0$, and $\exists\ 0 \neq v \geqslant 0$ $(v \succ 0)$ with $Mv = \lambda(M) v$;
2. $\kappa \in \Lambda(M)$ and $|\kappa| = \lambda(M)$ $\Rightarrow$ $\kappa = e^{2\pi i k/m} \lambda(M)$ for some $k, m \in \mathbb{N}$ with $m \leqslant n$;
3. $\kappa \in \Lambda(M) \Rightarrow |\kappa| \leqslant \lambda(M)$.

Pf summary: Express M as lim of $P > 0$. □

20. Multilinear algebra

Def: Fix vector spaces $V_1, \dots, V_r$ and W over F.

$f: V_1 \times \cdots \times V_r \to W$ multilinear if

$$f(\dots, v_{i-1}, \alpha v_i + v_i', v_{i+1}, \dots) = \alpha f(\dots, v_{i-1}, v_i, v_{i+1}, \dots) + f(\dots, v_{i-1}, v_i', v_{i+1}, \dots)$$

$\forall i \quad \forall v_i, v_i' \in V_i \quad \forall \alpha \in F$ with v_j fixed for $j \neq i$.

- $F \times \cdots \times F \to F$
  $(\alpha_1, \dots, \alpha_r) \mapsto \alpha_1 \cdots \alpha_r$

E.g.
- $A \in F^{m \times n} \Rightarrow (v, w) \mapsto vAw$ for $v \in F^m_{\text{row}}$ and $w \in F^n_{\text{col}}$ bilinear
- $V_i = F^n \ \forall i = 1, \dots, n$ and $(v_1, \dots, v_n) \mapsto \det[v_1 \ \cdots \ v_n]$

Lemma: $V_1 \times \cdots \times V_r = \prod V_i$

$$\begin{tikzcd}
\prod V_i \arrow[r, "\text{multilinear}"] \arrow[dr, "\Rightarrow \text{multilinear}"'] & W \arrow[d, "\text{linear}"] \\
& W'
\end{tikzcd}$$

Interpretation: $\{$vector spaces W with multilinear $\prod V_i \to W\}$ forms a category: Obj + Mor

Def: The tensor product of $V_1, \dots, V_r$ is a universal such thing:

multilinear $t: V_1 \times \cdots \times V_r \to T$ such that $\forall f \ \exists! \downarrow \tilde{f}$ with $f = \tilde{f} \circ t$

$$\begin{tikzcd}
V_1 \times \cdots \times V_r \arrow[r, "t"] \arrow[dr, "f"'] & T \arrow[d, "\Rightarrow \exists! \tilde{f}"] \\
& W
\end{tikzcd}$$

"the most general multilinear map from $\prod V_i$"

Thm: T exists. Notation: $T = \bigotimes_{i=1}^{r} V_i = V_1 \otimes \cdots \otimes V_r$

$$t(v_1, \dots, v_r) = v_1 \otimes \cdots \otimes v_r$$

Q. Does every element of $V_1 \otimes \cdots \otimes V_r$ have the form $v_1 \otimes \cdots \otimes v_r$?

A. No. Key example: $w \in W = F^m_{\text{col}}$, $\varphi \in V^* = F^n_{\text{row}}$ $\Rightarrow w \otimes \varphi = w\varphi$ has rank 1

$$\underset{w}{i\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix}} \otimes \underset{\varphi}{\begin{bmatrix}0 & \cdots & 0 & \overset{j}{1} & 0 & \cdots & 0\end{bmatrix}} = i\begin{bmatrix} & & \\ & \overset{j}{1} & \\ & & \end{bmatrix}$$

But $W \otimes V^* \overset{\text{very general}}{=} \operatorname{Hom}(V, W) = F^{m \times n}$, and lots of matrices have rank > 1.

$$\left(x \mapsto x_j e_i = i\begin{bmatrix}0\\ \vdots\\ 0\\ x_j\\ 0\\ \vdots\\ 0\end{bmatrix}\right) \in \text{basis } B.$$

$$\begin{tikzcd}
(w, \varphi) \arrow[r, mapsto] \arrow[dr, "\text{multilinear!}"'] & w \otimes \varphi \arrow[r, mapsto] \arrow[d, "\Rightarrow \exists!"] & (x \mapsto \varphi(x) w) \\
& w\varphi &
\end{tikzcd}$$

Pf: Construct T by "freely" multiplying elements of $V_1, \dots, V_r$ because
" " " " is multilinear.

Want:
$$\cdots \otimes (v_i + v_i') \otimes \cdots = (\cdots \otimes v_i \otimes \cdots) + (\cdots \otimes v_i' \otimes \cdots)$$
$$\cdots \otimes (\alpha v_i) \otimes \cdots = \alpha (\cdots \otimes v_i \otimes \cdots)$$

Do with $e_{(v_1, \dots, v_r)}$ instead of $(v_1, \dots, v_r)$

Set $T = M/N$ for $M = \operatorname{span}((v_1, \dots, v_r) \mid v_i \in V_i \ \forall i)$

$$N = \operatorname{span}((\dots, \alpha v_i + v_i', \dots) - \alpha(\dots, v_i, \dots) - (\dots, v_i', \dots))$$

Compare: $\mathbb{R}^2 \hookrightarrow M$, $(a,b) \mapsto e_{(a,b)}$ $\quad N: e_{(a,2b)} - 2e_{(a,b)}$

$\prod V_i \hookrightarrow M$ map of sets — not multilinear, but

$\prod V_i \hookrightarrow M \twoheadrightarrow M/N$ multilinear by construction.

Suppose

$$\begin{tikzcd}
\prod V_i \arrow[r] \arrow[dr, "f"'] & M \arrow[d] \\
& W
\end{tikzcd}$$

Then $\downarrow$ because $\prod V_i$ is a basis of M (!)

But f multilinear $\Rightarrow N \subseteq \ker(\downarrow)$.

Universal property of quotients $\Rightarrow$ $\begin{tikzcd} M \arrow[d] \\ W \end{tikzcd}$ induces unique $\begin{tikzcd} M/N \arrow[d] \\ W \end{tikzcd}$ $= T$. □

How to compute?

Lemma: $B \subseteq V$ independent iff $\exists$ linear $\{\varphi_b : V \to F\}_{b \in B}$ with $\varphi_b(b') = \delta_{b,b'}$. f true but not needed

Pf: $\sum_{b \in B} \alpha_b b = 0 \Rightarrow 0 = \varphi_{b'}\left(\sum_{b \in B} \alpha_b b\right) = \alpha_{b'}\ \forall\, b' \in B.$ □

Thm: B_i basis for $V_i \Rightarrow B_1 \times \cdots \times B_r \xrightarrow{\sim}$ basis B for T.

$$(b_1, \dots, b_r) \mapsto b_1 \otimes \cdots \otimes b_r$$

E.g. $\mathbb{R}^2 \otimes \mathbb{R}^3$ has basis $e_1 \otimes e_1,\ e_1 \otimes e_2,\ e_1 \otimes e_3,\ e_2 \otimes e_1,\ e_2 \otimes e_2,\ e_2 \otimes e_3$

Pf: Multilinear map on $\prod V_i$ determined by values on $\prod B_i$:

$v_1 = \sum_j \alpha_j b_1^j \Rightarrow f(v_1, \dots\text{stuff}\dots) = \sum_j \alpha_j f(b_1^j, \dots\text{stuff}\dots)$ and similarly for $i > 1$.

Thus $B = \{b_1 \otimes \cdots \otimes b_r \mid b_i \in B_i\ \forall i\}$ spans T. Need independence.

Suppose $f_i : V_i \to F\ \forall i$. Set $f = f_1 \otimes \cdots \otimes f_r : \prod V_i \to F$ product in F

$$\begin{tikzcd}
(v_1, \dots, v_r) \arrow[r, mapsto] \arrow[d, mapsto, "t"'] & f_1(v_1) \cdots f_r(v_r) \\
v_1 \otimes \cdots \otimes v_r \arrow[ur, mapsto, "\tilde{f}"'] &
\end{tikzcd}$$

f multilinear $\Rightarrow$ induces $\tilde{f} : T \to F$.

Take $f_i = b_i^* \in B_i^*$ dual basis, so

$$b_i^*(b_i) = 1 \quad \text{but} \quad b_i^*(b_i') = 0 \quad \text{when } b_i' \in B_i \setminus \{b_i\}.$$

Then $b \in B \rightsquigarrow \tilde{f}_b : T \to F$ satisfying

$$b \mapsto 1$$

$$b' \mapsto 0 \quad \text{for } b' \in B \setminus \{b\}.$$ Use Lemma. □

Cor: $\dim V_1 \otimes \cdots \otimes V_r = (\dim V_1) \cdots (\dim V_r)$. □

E.g. $v \in \mathbb{R}^4$ $w \in \mathbb{R}^2$ (get from class) $\Rightarrow v \otimes w =$

Universal property redux: $\{\text{multilinear } \prod_{i=1}^{r} V_i \to W\} \leftrightarrow \{\text{linear } \bigotimes_{i=1}^{r} V_i \to W\}$.

E.g. $\exists!$ isomorphism $V \otimes W \to W \otimes V$
$v \otimes w \mapsto w \otimes v$.

Pf: HW5; use $V \times W \to W \otimes V$
$(v, w) \mapsto w \otimes v$. □

21.

Exterior algebra

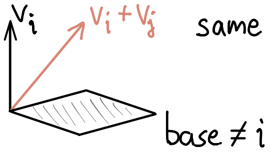
Def: An alternating operator $\varphi: \underbrace{V \times \cdots \times V}_{r} = V^{\times r} \to W$ is a multilinear map such that

$v_1, \dots, v_r$ linearly dependent $\Rightarrow \varphi(v_1, \dots, v_r) = 0$.

E.g. volume of parallelepiped on $v_1, \dots, v_n \in \mathbb{R}^n$: vol $= 0$ if $v_1, \dots, v_n$ linearly dependent

$\overset{221}{\Rightarrow}$ vol = det $\left\{ \begin{array}{l} \bullet\ v_i \rightsquigarrow \alpha v_i \Rightarrow \text{vol} \rightsquigarrow \alpha\,\text{vol}\ \ \forall \alpha \\ \bullet\ v_i \rightsquigarrow v_i + v_j \text{ for } j \neq i \Rightarrow \text{vol unchanged} \\ \bullet\ \text{vol}(e_1, \dots, e_n) = 1. \end{array} \right.$

including $\alpha < 0$: signed or oriented volume



same height $\Rightarrow$ same vol

Def: r^{th} exterior power of V is a universal alternating operator:

alternating map $\Lambda^r : V^{\times r} \to U$ such that $\forall\ \varphi$ alternating $\exists!\ \tilde{\varphi}$ with $\varphi = \tilde{\varphi} \circ \Lambda^r$.

$$\begin{tikzcd} V^{\times r} \arrow[r, "\Lambda^r"] \arrow[dr, "\text{alternating } \varphi"'] & U \arrow[d, "\Rightarrow \exists!\ \tilde{\varphi}"'] \\ & W \end{tikzcd}$$

Thm: Λ^r exists.

Pf: Set $U = \Lambda^r V = \overset{V \otimes \cdots \otimes V}{\overset{\shortparallel}{V^{\otimes r}}} \Big/ \overset{Y}{\boxed{\text{span}(v_1 \otimes \cdots \otimes v_r \mid \text{two of the } v\text{'s are equal})}}$ r-forms

$V^{\times r} \to \Lambda^r V$

$(v_1, \dots, v_r) \mapsto v_1 \wedge \cdots \wedge v_r$ multilinear because factors through $V^{\otimes r}$

\wedge

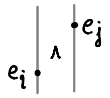
alternating because $v_1 = \sum_{j>1} \alpha_j v_j \Rightarrow v_1 \wedge \cdots \wedge v_r = \sum_{j>1} \alpha_j v_j \wedge (v_2 \wedge \cdots \wedge v_r)$

$= 0$, and same for $i > 1$.

$\varphi: V^{\times r} \to W$ multilinear $\Rightarrow \varphi$ factors through $V^{\otimes r}$...

alternating $\Rightarrow$ ... and kills $Y \Rightarrow$ factors through $V^{\otimes r}/Y$. ☐

E.g. $v, w \in \mathbb{R}^4_{\text{col}}$ (get from class) $\Rightarrow v \wedge w =$



Prop: $V \xrightarrow{\varphi} W$ linear induces canonical linear map $\Lambda^r V \xrightarrow{\Lambda^r \varphi} \Lambda^r V$. Λ^r is a functor.

$v_1 \wedge \cdots \wedge v_r \mapsto \varphi(v_1) \wedge \cdots \wedge \varphi(v_r)$

Pf: HW5, including entries of matrix if A is given. ☐

Quintessential E.g. $V = W$ and $r = n = \dim V$: determinant of $\varphi: V \to V$ is $\det \varphi = \Lambda^n \varphi$.

Note: $\det \varphi = \Lambda^{top} \varphi$ since $\Lambda^r V = 0$ for $r \geq n+1$.

$$\varphi(e_j) = v_j = \sum_{i=1}^{n} a_{ij} e_i \Rightarrow \Lambda^n \varphi(e_1 \wedge \cdots \wedge e_n) = v_1 \wedge \cdots \wedge v_n = \left(\sum_{i=1}^{n} a_{i1} e_i\right) \wedge \cdots \wedge \left(\sum_{i=1}^{n} a_{in} e_i\right) = \underline{\qquad} e_1 \wedge \cdots \wedge e_n$$

$$= \sum_{i_1, \dots, i_n} a_{i_1 1} e_{i_1} \wedge \cdots \wedge a_{i_n n} e_{i_n}.$$

Terms are 0 unless $i_1, \dots, i_n$ distinct, so $i_j = \pi(j)$ for some permutation $\pi \in S_n$. Thus

$$v_1\wedge\cdots\wedge v_n = \sum_{\pi\in S_n} a_{\pi(1)1}e_{\pi(1)}\wedge\cdots\wedge a_{\pi(n)n}e_{\pi(n)}$$

det A_π for permutation matrix A_π

$$= \underbrace{\sum_{\pi\in S_n}(-1)^\pi a_{\pi(1)1}\cdots a_{\pi(n)n}}_{\det A}\underbrace{e_1\wedge\cdots\wedge e_n}_{e_{[n]}}$$

<u>Notation</u>: For $\sigma = \{\sigma_1<\cdots<\sigma_r\}\subseteq[n]$ and $E = e_1,\ldots,e_n \in V$ set

$$e_\sigma = e_{\sigma_1}\wedge\cdots\wedge e_{\sigma_r}$$

$$\text{and } \Lambda^r E = \{e_\sigma \mid \sigma\in\tbinom{[n]}{r}\}.$$

<u>Prop</u>: E is a basis for $V \Rightarrow \Lambda^r E$ spans $\Lambda^r V$.

<u>Pf</u>: $\Lambda^r E = \{$squarefree elements of basis $E^{\otimes r}$ of $V^{\otimes r}\}$

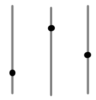
and nonsquarefree elements $\mapsto 0$ in $\Lambda^r V$. □

Q. coeff. on e_σ in $v_1\wedge\cdots\wedge v_r = ?$ i.e. How can $v_1\wedge\cdots\wedge v_n$ be expressed as a linear combination of elements e_σ?

A. $\det A_\sigma$, where $A = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{bmatrix}$ and A_σ takes rows indexed by σ.

<u>Pf</u>: In rows from σ, get one e_i from each column j with coeff a_{ij}:

$\sigma = \begin{Bmatrix}\sigma_1=\sigma_{\pi(2)}\\ \sigma_2=\sigma_{\pi(3)}\\ \sigma_3=\sigma_{\pi(1)}\end{Bmatrix} \begin{matrix}\rightarrow\\ \rightarrow\\ \rightarrow\end{matrix}$



$$v_1\wedge\cdots\wedge v_r = \sum_{\sigma\in\binom{[n]}{r}}\sum_{\pi\in S_r} a_{\sigma_{\pi(1)}1}e_{\sigma_{\pi(1)}}\wedge\cdots\wedge a_{\sigma_{\pi(r)}r}e_{\sigma_{\pi(r)}}$$

$$= \sum_{\sigma\in\binom{[n]}{r}}\sum_{\pi\in S_r} a_{\sigma_{\pi(1)}1}\cdots a_{\sigma_{\pi(r)}r}\, e_{\sigma_{\pi(1)}}\wedge\cdots\wedge e_{\sigma_{\pi(r)}}$$

$$= \sum_{\sigma\in\binom{[n]}{r}}\underbrace{\sum_{\pi\in S_r} a_{\sigma_{\pi(1)}1}\cdots a_{\sigma_{\pi(r)}r}(-1)^\pi}_{\det A_\sigma} e_\sigma. \quad \square$$

<u>Thm</u>: E is a basis for $V \Rightarrow \Lambda^r E$ is a basis for $\Lambda^r V$.

<u>Pf</u>: spans by Prop.

independent: existence of determinants $\Rightarrow (v_1,\ldots,v_r)\mapsto \det([v_1\cdots v_r]_\sigma)$ is alternating, so

induces $e^*_\sigma : \Lambda^r V\to F$ with $e^*_\sigma(e_\tau) = \delta_{\sigma,\tau}$; Lemma $\Rightarrow$ independent. □

<u>Cor</u>: $\dim V = n \Rightarrow \dim\Lambda^r V = \binom{n}{r}$ if $r\leqslant n$ and $\Lambda^r V = 0$ if $r>n$. □

<u>Prop</u>: $v\in V \Rightarrow v\wedge : \Lambda^r V\to\Lambda^{r+1}V$ linear

$\omega\in\Lambda^j V \Rightarrow \omega\wedge : \Lambda^r V\to\Lambda^{r+j}V$ linear

$(\omega_1\wedge)\circ(\omega_2\wedge) = (\omega_1\wedge\omega_2)\wedge : \Lambda^r V\to\Lambda^{r+j+k}V$ if $\omega_1\in\Lambda^j V$ and $\omega_2\in\Lambda^k V$ (associativity)

<u>Remark</u>: $\Rightarrow \Lambda^* V = \bigoplus_r \Lambda^r V$ is a ring (Lec. ④).

22.

Def: The cross product of two vectors $v, w \in \mathbb{R}^3$ is $u = v \times w$ whose entries are the coeffs on i, j, k in $\det \begin{bmatrix} i & j & k \\ - & v & - \\ - & w & - \end{bmatrix}$.

Prop: $u \cdot v = u \cdot w = 0$ and $\|u\| =$ area of parallelogram spanned by v and w. $\square$ $\mathbb{R}^n$?

Def: $\mathrm{vol}(v_1, \dots, v_r) = |\mathrm{vol}(v_1, \dots, v_r, u_{r+1}, \dots, u_n)|$ for any $\perp$normal basis $u_{r+1}, \dots, u_n$ of $\{v_1, \dots, v_r\}^\perp$.

Def: cross product u of $v_2, \dots, v_n \in \mathbb{R}^n$ satisfies $v_j \cdot u = 0$ $\forall j = 2, \dots, n$ and $\|u\| = \mathrm{vol}(v_2, \dots, v_n)$.

Thm: $u = \sum_{i=1}^n u_i e_i \Rightarrow -v_2 \wedge \cdots \wedge v_n$ has coeff. $(-1)^i u_i$ on $e_{\bar{i}}$, where $\bar{i} = [n] \setminus \{i\}$.

Pf: $v_2 \wedge \cdots \wedge v_n \in \mathrm{span}(e_{\bar{1}}, \dots, e_{\bar{n}}) \Rightarrow$ it is $-\sum_{i=1}^n (-1)^i u_i e_{\bar{i}}$ for some u. For any $w \in \mathbb{R}^n$,

$$w \wedge v_2 \wedge \cdots \wedge v_n = \left(\sum_{i=1}^n w_i e_i\right) \wedge \left(-\sum_{i=1}^n (-1)^i u_i e_{\bar{i}}\right)$$

$$= \sum_{i=1}^n - w_i u_i (-1)^i e_i \wedge e_{\bar{i}}$$

$$= w \cdot u \; e_1 \wedge \cdots \wedge e_n .$$

$$w = v_j \Rightarrow v_j \cdot u \, e_{[n]} = v_j \wedge v_2 \wedge \cdots \wedge v_n = 0 \overset{e_{[n]} \neq 0}{\Rightarrow} v_j \cdot u = 0 \;\; \forall j \geq 2. \quad (*)$$

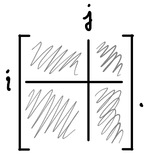
Let $w \in \{v_2, \dots, v_n\}^\perp$ and $\|w\| = 1$. Then

$\det(w, v_2, \dots, v_n)$

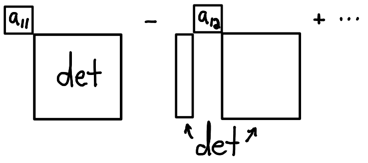
$$\mathrm{vol}(v_2, \dots, v_n) = |\mathrm{vol}(w, v_2, \dots, v_n)| = |\text{coeff. on } e_{[n]} \text{ in } w \wedge v_2 \wedge \cdots \wedge v_n|$$

$$= |w \cdot u| \overset{(*)}{=} \frac{u}{\|u\|} \cdot u = \frac{\|u\|^2}{\|u\|} = \|u\|. \;\square$$

Recall: Laplace expansion of det along a row: $\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A_{ij}$ for $A_{ij} =$



E.g. $i = 1$:



Along (top) r rows: $A[\leq r]$ = top r rows $\qquad A[\leq r]^\sigma = r \times r$ submatrix with cols from $\sigma \in \binom{[n]}{r}$

$A[>r]$ = bottom $n-r$ rows $\qquad A[>r]^{\bar\sigma} \qquad \bar\sigma = [n] \setminus \sigma$

Thm: $\det A = \sum_{\sigma \in \binom{[n]}{r}} (-1)^\sigma \det A[\leq r]^\sigma \det A[>r]^{\bar\sigma}$ $\qquad (-1)^\sigma = (-1)^{\#\text{swaps needed to put } \sigma, \bar\sigma \text{ in order}}$

Pf: $\det A = $ coeff. on $e_{[n]}$ in $v_1 \wedge \cdots \wedge v_n$, where $A = \begin{bmatrix} - & v_1 & - \\ & \vdots & \\ - & v_n & - \end{bmatrix}$.

Factor $v_1 \wedge \cdots \wedge v_n$: (coeff. on e_σ in $v_1 \wedge \cdots \wedge v_r$) $= \det A[\leq r]^\sigma$

$e_{\bar\sigma} \qquad v_{r+1} \wedge \cdots \wedge v_n \; = \det A[>r]^{\bar\sigma}$

$$\Rightarrow v_1 \wedge \cdots \wedge v_n = (v_1 \wedge \cdots \wedge v_r) \wedge (v_{r+1} \wedge \cdots \wedge v_n) = \sum_\sigma \det A[\leq r]^\sigma \det A[>r]^{\bar\sigma} \underbrace{e_\sigma \wedge e_{\bar\sigma}}_{(-1)^\sigma e_{[n]}}. \;\square$$

Pf very simple and explains
exactly where the sign comes from

Thm: rank $A < r \iff$ all $\underbrace{\text{minors of size } r}_{\det(r\times r \text{ submatrix})}$ vanish.

entries are the r-minors of A (HW5)

Pf: $A \in F^{m\times n}$ represents $\varphi: F^n \to F^m$. rank $\varphi < r \iff \dim(\operatorname{im}\varphi) < r \iff \Lambda^r\varphi = 0$. □

Intuition: r-dim vol in dim n needs to specify
- an r-dim subspace V
- a full-dim (i.e. dim r) volume in V

Thm: $v_1\wedge\cdots\wedge v_r$ identifies $V = \operatorname{span}(v_1,\ldots,v_r)$ up to a scalar if $v_1,\ldots,v_r$ independent:

$$w_1\wedge\cdots\wedge w_r = \alpha v_1\wedge\cdots\wedge v_r \text{ for some } \alpha \neq 0 \iff \underset{\substack{\shortparallel \\ W}}{\operatorname{span}(w_1,\ldots,w_r)} = V.$$

Geometric interpretation: $F = \mathbb{R}$ and $u_1,\ldots,u_r$ $\perp$normal in V

$$\Rightarrow v_1\wedge\cdots\wedge v_r = \alpha u_1\wedge\cdots\wedge u_r \text{ with } |\alpha| = \operatorname{vol}(v_1,\ldots,v_r).$$

Pf of Thm: $W = V \Rightarrow \underset{\mathcal{W}}{\begin{bmatrix} - w_1 - \\ \vdots \\ - w_r - \end{bmatrix}} = A \underset{\mathcal{V}}{\begin{bmatrix} - v_1 - \\ \vdots \\ - v_r - \end{bmatrix}}$ for some $A \in GL_r$

multiplies all r-minors by same scalar, namely $\det A$

$\Rightarrow w_1\wedge\cdots\wedge w_r = \det A\; v_1\wedge\cdots\wedge v_r$, because coeff on e_σ in $w_1\wedge\cdots\wedge w_r$ is

$$e_\sigma^*(w_1\wedge\cdots\wedge w_r) = \det\left(A\begin{bmatrix} - v_1 - \\ \vdots \\ - v_r - \end{bmatrix}^\sigma\right) = \det A\; e_\sigma^*(v_1\wedge\cdots\wedge v_r).$$

Now suppose $w_1\wedge\cdots\wedge w_r = \alpha v_1\wedge\cdots\wedge v_r \neq 0$ for some $\alpha \in F$.

Then $w_i\wedge v_1\wedge\cdots\wedge v_r = \alpha^{-1} w_i\wedge w_1\wedge\cdots\wedge w_r = 0 \quad \forall i$.

Lemma: $w\wedge v_1\wedge\cdots\wedge v_r = 0 \iff w \in \operatorname{span}(v_1,\ldots,v_r)$.

Pf: $\Leftarrow$: def. of alternating.

$\Rightarrow$: $w \notin \operatorname{span}(v_1,\ldots,v_r) \Rightarrow$ can complete $w, v_1,\ldots,v_r$ to a basis of F^n whose Λ is nonzero. □

Def: $\left\{ e_\sigma^*(v_1\wedge\cdots\wedge v_r) \,\middle|\, \sigma \in \binom{[n]}{r} \right\}$ = Plücker coordinates of V.

Cor: $G_r(F^n) \longleftrightarrow \{\text{decomposable forms in } \mathbb{P}\Lambda^r F^n = G_1(\Lambda^r F^n)\}$.