Δ L.A. revisit

Δ determinants

- $\det(A) = 0$ squeeze original state space, to lower dimension
- $\det(A) = c$

Area $= c$

if: $\mathrm{sign}(c) = (-)$

then: "flipped"

$Ax = b$

- if $\det(A) = 0$:
  - A transform x to "lower dimension" cannot unpack. i.e., ~~if~~ transform back
  - Sometimes can! already "live" on the new "lower dim."

  call!

- if $\det(A) \neq 0$:
  - then can reverse $\det$!

Δ Rank

- $\mathrm{rank}(A) = n$

  then n is the dimension of the spanned space of the output

  $Ax = b$ DoF (allowing dimension)

  column space ≜ all possible outputs Ax

- full rank

  $\mathrm{rank}(A) = \min(m, n)$

Δ Nullspace (kernel)

this term can be understood as "transformation" spatially, unexplaining space. SVMs CNNs

$Ax = 0 \quad A \in R^{m \times n}$

↳ a set of vectors, which form a space, whose DoF(s) get "squeeze" to the origin;

note that all A has trivial null space $\{0\}$;

null space $\in R^{n-m}$ if $n > m$; reduced rank

Δ Range, Rank, Nullspace (summary)

$A \in R^{m \times n}$

⇒ RANGE(A)
- the set of all vecs in R^m
- linear combinations of col. vec.
- $\{Ax \in R^m \mid x \in R^n\}$
- span of col. vec.

⇒ RANK(A)
- the no. of linearly independent col.
- dimension of span of col. vec.

dimension is

⇒ NULLSPACE(A)
- the set of vecs get mapped to 0, i.e., null.
- $\{x \in R^n \mid Ax = 0\}$

Δ orthogonal decomposition

- $U \in R^n \rightarrow U^{\perp}$ orthogonal complement

$U^{\perp} = \{x \mid z^T x \ \ \forall z \in R^n\}$

$\therefore U^{\perp\perp} = U$

- $\mathrm{NULLSPACE}(A) = \mathrm{RANGE}(A^T)^{\perp}$

  $\{x \in R^n \mid Ax = 0\}$ = $\{A^T x \in R^n \mid x \in R^m\}^{\perp}$ ($R^{n \times m}$)

  $= \{x' \mid [A^T x]^T x' = 0\}$

  $= \{x' \mid \underset{0}{x^T A x'} = 0\}$

  $\Updownarrow$

  $\{x' \mid Ax' = 0\}$

- $R(A) = N(A^T)^{\perp}$

  $R(A) = \{Ax \in R^m \mid x \in R^n\}$

  $N(A^T)^{\perp} = \{x \in R^m \mid A^T x = 0\}^{\perp}$ ($R^{n \times m}$ $R^{m \times 1}$)

  $= \{x' \in R^m \mid x^T x' = 0, A^T x = 0\}$

  $= \{x' \in R^m \mid x'^T x = 0, A^T x = 0\}$ this means span of cols is ⊥ to x; this is the set of vecs ⊥ to x

  $\Leftrightarrow \{A^T v \in R^m \mid v \in R^n\}$

$\Rightarrow N(A) \overset{\perp}{\oplus} R(A^T) = R^n$

orthogonal direct sum; span ⇒ R^n

Δ matrix norm

$[A] = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

row sum w/ abs.

$\|A\|_\infty = \max\{|1|+|2|, |0|+|1|\}$

$= \max\{3, 1\} = 3$

$\|A\|_1$ = sum of row 1 = 3

$\|A\|_2$ = sum of row 2 = 1

e.g. recall $\mathrm{cond}[A] = \|A\|_\infty \|A^{-1}\|_\infty$

$A = \begin{bmatrix} 400 & -201 \\ -800 & 401 \end{bmatrix}$

$A^{-1} = \begin{bmatrix} -1.002 & -5.01 \\ -2 & 1 \end{bmatrix}$

$\Rightarrow \|A\|_\infty = |-801| + |400| = 1201$

$\|A^{-1}\|_\infty = |-2| + |-1| = 3$

$\Rightarrow \mathrm{cond}(A) = 3603$

Δ more on matrix norm

- recall vector norms

  L_1-norm $\quad \|u\|_1 = |u_1| + |u_2|$

  L_2-norm. $\quad \|u\|_2 = \sqrt{u_1^2 + u_2^2}$

- matrix norms "based on" vector norm

  matrix ⟶ linear transformation

  ∴ matrix norm:

  "HOW BIG MY MATRIX CAN GET MY OUTPUT VECTOR".

  $\|A\|_2 = \max_{\|u\|_2 = 1} \|Au\|_2$

- $\|AB\|_p \leq \|A\|_p \|B\|_p$
- $\|A^k\|_p = \max_{\|u\|_p = 1} \|A^k u\|_p$

  $A[u_1 \cdots u_i] = \lambda_i [u_1 \cdots u_i]$

  $AU = \lambda U$

  $\Rightarrow \|Au_i\|_p = |\lambda_i| \|u_i\|_p \leq \|A\|_p \|u_i\|_p$

  $\Rightarrow \|\lambda_i\| \leq \|A\|_p$

  $\Rightarrow \max |\lambda_i| \leq \|A\|_p$

  $\Rightarrow \max |\lambda_i|^k \leq \|A^k\|_p$

Δ ill/well - condition of matrix

$Ax = c$

ill conditioned? well conditioned? } small change in $[A]$ small/large change in $[c]$?

vice-versa ⟶ small change in $[c]$ small/large change in $[A]$?

Δ small ΔA ⟶ small Δc "well-conditioned"

Δ small ΔA ⟶ large Δc "ill-conditioned"

e.g. $\begin{bmatrix} 400 & -201 \\ -800 & 401 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 200 \\ -200 \end{bmatrix}$ $x = \begin{bmatrix} -100 \\ -200 \end{bmatrix}$

$\begin{bmatrix} 401 & -201 \\ -800 & 401 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 200 \\ -200 \end{bmatrix}$ $x = \begin{bmatrix} 40{,}000 \\ 79{,}800 \end{bmatrix}$

Δ condition number

$\mathrm{cond}[A] \approx 1.0$ well

$\mathrm{cond}[A] > 1.0$ ill

$\mathrm{cond}[A] = \|A\|_\infty \|A^{-1}\|_\infty$

Rule of thumb $\log[\mathrm{cond}[A]]$ = decimals of accuracy lost

Δ On ill-conditioned & Eigenvalues.

- usually: ill-condition matrix sometimes contains column vecs. that are colinear
- e.g. $A = \begin{bmatrix} 1.01 & 3 \\ 1 & 3 \end{bmatrix}$

  $\Rightarrow \begin{bmatrix} 1.01-\lambda & 3 \\ 1 & 3-\lambda \end{bmatrix}$

  $\Rightarrow (3-\lambda)(1.01-\lambda) - 3$

  $= 3.03 - 4.01\lambda + \lambda^2 - 3$

  $= \lambda^2 - 4.01\lambda + 0.03$

  $\Rightarrow \lambda = 4$ or 0.005

  Δ $\mathrm{COND}(A) = \|A\|_\infty \|A^{-1}\|_\infty$

  $A = \begin{bmatrix} 1.01 & 3 \\ 1 & 3 \end{bmatrix}$

  $A^{-1} = \frac{1}{0.03} \begin{bmatrix} 3 & -3 \\ -1 & 1.01 \end{bmatrix}$

  $\|A\|_\infty \|A^{-1}\|_\infty = 4.01 + 200$

  $= 204.01$

- observation:
  - small eigen value: implies that some eigenvectors span is "squeezed" very small.
  - small eigen value: bad condition matrix
  - small eigenvalue: colinearity, approximately

Δ Eigen-decomposition

Δ revision of eigenvectors & eigenvalues

- let's say

  $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} v = w$

- v will hv a span

  linear transformation! (this is what matrix does)

  span of v

- usually, w is knocked off from the span
- some, w remains on the span

  THIS IS A EIGENVECTOR

- mathematically

  $Av = \lambda v \qquad \lambda v \in \mathrm{SPAN}(v)$

  eigenvector, eigenvalue

  "new span can be understood as λ-times SPAN(v)"

Δ calculation of eigen-value vector

$Av = \lambda v$

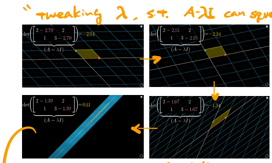
$\Rightarrow (A - \lambda I)v = 0$

$v \in \mathrm{Null}(A - \lambda I)$, non-trivial nullspace

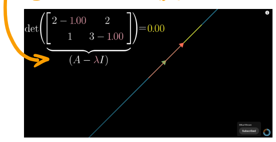
hv to be rank deficient!

$\det(A - \lambda I) = 0$

"tweaking λ, s.t. $A - \lambda I$ can squeeze



someone to null!"



Δ $\mathrm{Eig}(A) \in \mathbb{C}$ (complex number)

⇒ hv some rotation!

e.g. $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \rightarrow \pm i$

Δ double-rooted $\mathrm{Eig}(A)$

⇒ shear

e.g. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

Δ Eigen-basis ⟹ Eigendecomposition

A → express w/ new basis w/ eigen vectors

$A = T^{-1} D T$

Δ change of basis

$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$ [a, b] in Jennifer's coordinate

- Jennifer's basis
- a different basis

★ our grid ⟶ Jennifer grid

$[A]$

our lang. ⟵ Jennifer's lang.

★ $A^{-1} \ T \ A \begin{bmatrix} a \\ b \end{bmatrix}$

Jennifer's lang.

our lang.

some transf. that we understand still in our lang.

back to Jennifer!

equivalent transformation in different coordinate!

Δ Non-square matrix

$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ -2 & 1 \end{bmatrix} x \in R^3$

$R^{3 \times 2} \quad R^{2 \times 1}$

This is the span of 2 col vecs. in 3D space!

$A = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \end{bmatrix} x \in R^2$

$R^{2 \times 3} \quad R^{3 \times 1}$

shrink down to lower space!

SUM

1. Think of skinny matrix as taking the span of some vecs in high dimension

   $\begin{bmatrix} A_1 \cdots A_n \end{bmatrix}$

2. Think of fat matrix as shrinking down to lower space

   $\begin{bmatrix} A \\ 0 \cdots 0 \end{bmatrix} \quad \det(\text{ghost } A) = 0$

Δ Eigen-Decomposition (cont'd)

- $\begin{cases} Au_1 = \lambda_1 u_1 \\ Au_2 = \lambda_2 u_2 \end{cases}$

$\Rightarrow A[u_1 \ u_2] = [u_1 \ u_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$

$\Rightarrow AU = U\Lambda$

$\Rightarrow A = U \Lambda U^{-1}$

- if $A = A^T$ i.e., symmetric

$\Rightarrow A = Q \Lambda Q^T$, $QQ^T = I$ $Q^{-1} = Q^T$

(spectral decomposition)

$\Rightarrow \det A = \prod_{i=1}^{n} \lambda_i \quad \mathrm{tr} A = \sum_{i=1}^{n} \lambda_i$

$\Rightarrow$ spectral norm & Frobenius norm

$\|A\|_2 = \max_{\lambda \in \Lambda} |\lambda_i| \qquad \|A\|_F = (\sum \lambda_i^2)^{1/2}$

$\Rightarrow \lambda_{min}(A) x^T x \leq x^T A x \leq \lambda_{max}(A) x^T x$

$\forall x$

Δ SVD

- $\forall R^{m \times n}$ $p = \mathrm{Rank}(A) \leq n$ linearly independent!
- $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_p u_p v_p^T$

  $R^{m \times 1} R^{1 \times n}$ a rank

  $= \begin{bmatrix} u_1 \cdots u_p \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_p \end{bmatrix} \begin{bmatrix} -v_1^T- \\ -v_2^T- \\ \vdots \\ -v_p^T- \end{bmatrix}$

  $R^{m \times p} \quad R^{p \times p} \quad R^{p \times n}$

  U ($U^T U = I$) $\quad \Sigma$ DIAG $\quad V^T$ ($V^T V = I$)

$\Rightarrow A = U \Sigma V^T$

$\Rightarrow AV = U \Sigma V^T V = U \Sigma I$

$\Rightarrow U^T A = U^T U \Sigma V^T = I \Sigma V^T$

$\Rightarrow AV = U\Sigma \qquad U^T A = \Sigma V^T$

sing. values; right sing. vectors; left sing. vectors

- why?

  $A = R^{100 \times 10} \longrightarrow$ #1000

  suppose $\sigma_1 = 3$

  $\sigma_2 = 2$

  $\sigma_i \approx 0 \ \ \forall i > 2$

  $\therefore A \approx \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$

  $= [u_1 \ u_2] \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} -v_1^T- \\ -v_2^T- \end{bmatrix}$

  $R^{100 \times 2} \quad R^{2 \times 2} \quad R^{2 \times 10}$

  #200 #2 #20

  ⇒ #222 < 25% !

Δ SVD v Eigen.

$A = U \Sigma V^T \quad \in R^{m \times n}$

$\begin{cases} A^T A \in R^{n \times n} \\ A A^T \in R^{m \times m} \end{cases}$

$\Rightarrow A^T A V$

$= V \Sigma^T U^T U \Sigma V^T V$

$= V \Sigma^T \Sigma$

$= V \Sigma^2$

$A^T A \begin{bmatrix} v_1 \cdots v_p \end{bmatrix} = \begin{bmatrix} \sigma_1^2 v_1 \cdots \sigma_p^2 v_p \end{bmatrix}$

Eigen vec. Eigen val. of $A^T A$

$\Rightarrow A A^T U$

$= U \Sigma V^T V \Sigma^T U^T U$

$= U \Sigma \Sigma^T$

$= U \Sigma^2$

$\Rightarrow AA^T \begin{bmatrix} u_1 \cdots u_p \end{bmatrix} = \begin{bmatrix} \sigma_1^2 u_1 \cdots \sigma_p^2 u_p \end{bmatrix}$

Eigen vec Eigen val of AA^T

Δ SVD & Pseudo-inverse

- $A = U \Sigma V^T$ (P linear independent cols)

  $m \times n \quad m \times p \quad p \times p \quad p \times n$

  $A^+ \equiv V \Sigma^{-1} U^T$

$\Leftrightarrow A^+ = \lim_{\epsilon \to 0} (A^T A + \epsilon I)^{-1} A^T$

$= \lim_{\epsilon \to 0} A^T (AA^T + \epsilon I)^{-1}$ damped to ensure inverse exists

- if $\mathrm{rank} A = n$

  $A^+ (A^T A)^{-1} A^T$

- if $\mathrm{rank} A = m$

  $A^+ = A^T (AA^T)^{-1}$

Δ Principal Component Analysis

Δ Intuition

- in this graph, we can actually use 2 axis to represent the data in 3D.
- reduce dimension & feature extract better VIZ!

Δ ALGORITHMS

input: $\{x_n\} \in R^D$

output: W s.t. $W^T x = y \in R^K$, $K < D$

1. $\{x_n\} \quad n = 1, \ldots, N$

   u_1, u_2: candidates of principle components

2. projection of x on u_i

   a. $|u_i|_2 = 1$

   b. $\mathrm{PROJ}_{u_i}(x_n) = u_i^T x_n u_i$

   - $u_i^T x_n = |u_i||x_n| \cos\theta$

     $= |x_n| \cos\theta$ (length of projection on u_i)

   - $(u_i^T x_n) u_i$ scalar; vector norm = 1, give direction

   c. ∴ mean of projection

   $u_i^T \bar{x} u_i$

3. variance of projection

   $\frac{1}{N} \sum_{n=1}^{N} (u_i^T x_n - u_i^T \bar{x})^2$

   $= \frac{1}{N} \sum_{n=1}^{N} [u_i^T (x_n - \bar{x})]^2$

   $= \frac{1}{N} u_i^T \left[\sum_{n=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T\right] u_i$

   $= u_i^T S u_i$ ↳ covariance matrix

4. max $u_i^T S u_i$

   s.t. $u_i^T u_i = 1$

   $\Updownarrow$

   min $u_i^T S u_i + \lambda_i (1 - u_i^T u_i)$

   $\Updownarrow$

   $2 S u_i - \lambda_i 2 u_i = 0$

   $\Updownarrow$

   $S u_i = \lambda u_i$ ⟹ our old friend!

   $\Updownarrow$

   $u_i^T S u_i = \lambda_i u_i^T u_i$

   $\Updownarrow$

   $u_i^T S u_i = \lambda_i$

5. get max K λ_i's

   then hv K u_i's!

   ⇒ data in R^K!

Δ Usage

- let $W = D \begin{bmatrix} u_1 & u_2 & \cdots & u_K \end{bmatrix} \in R^{D \times K}$ (K columns)
- recall projection:

  $\mathrm{PROJ}_{u_i}(x_n) = u_i^T x_n u_i$ (scalar on i axis (principal components))

- new data x_{n+1}

  $\Rightarrow x'_{n+1} = x_{n+1} - \bar{x}$ (PCA is built around mean!)

  $\Rightarrow y = \begin{bmatrix} u_1^T x'_{n+1} \\ u_2^T x'_{n+1} \\ \vdots \\ u_K^T x'_{n+1} \end{bmatrix}$

  $= \begin{bmatrix} -u_1^T- \\ -u_2^T- \\ \vdots \\ -u_K^T- \end{bmatrix} x'_{n+1}$

  $= W^T x'_{n+1}$

- w/ y → you can do:

  a. throw in ML lower Dim.
  - classification
  - clustering
  - anomaly detection

  b. data VIZ.

  c. noise reduction

  d. feature engineering

Δ PCA v SVD

- statistical interp. of EVD
- proposal coordinate system (based on data)

1. $\bar{x} = \begin{bmatrix} - \bar{x}^T - \end{bmatrix}$
2. $\bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_j \qquad \bar{x}_{avg} = [\;]$
3. $B = X - \bar{X}$ (mean centered data) ↳ $N(0, \Sigma)$
4. $\mathrm{Cov}\ S = B^T B$
5. Eigen: $SV = VD$ principal component; loading

$T = BV$

$B = U \Sigma V^T \qquad \lambda = \sigma^2 \qquad y \quad W^T x_{\text{new data}}$

$\Rightarrow T = U \Sigma V^T V$

$= U\Sigma$

△ Cholesky Decomposition

$A = LL^T$ Cholesky Decomp.

solving

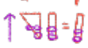
$A\hat{x} = b$

- first decompose this
- $Ax = b$

  $LL^Tx = b$

$\Rightarrow Ly = b$

- $L^Tx = y$



- solve x !

△ QR decomposition

- first assume

  A square, invertible

  $A = QR$

  orthogonal

  unitary matrix

  in Real $Q^T = Q^{-1}$

  in Complex $Q^* = Q^{-1}$

  cols of Q form orthonormal basis

- Gram-Schmidt process

  $A = [a_1\ a_2\ a_3] \in \mathbb{R}^{3\times 3}$

  transform to ONB

  $Q = [q_1\ q_2\ q_3]$ & R

1. $q_1 := \frac{a_1}{\|a_1\|}$

   $\therefore a_1 = \|a_1\| q_1$

2. $\langle a_2, q_1 \rangle q_1$

   $a_2^\perp = a_2 - \langle a_2, q_1 \rangle q_1$

   $q_2 := \frac{a_2^\perp}{\|a_2^\perp\|}$

   $\therefore a_2 = \langle a_2, q_1 \rangle q_1 + a_2^\perp$

   $= \langle a_2, q_1 \rangle q_1 + \|a_2^\perp\| q_2$

3. $a_3^\perp = a_3 - \langle a_3, q_1 \rangle q_1 - \langle a_3, q_2 \rangle q_2$

   $q_3 = \frac{a_3^\perp}{\|a_3^\perp\|}$

   $\therefore$

   $a_3 = \langle a_3, q_1 \rangle q_1 + \langle a_3, q_2 \rangle q_2 + \|a_3^\perp\| q_3$

4. $\therefore Q = [q_1\ q_2\ q_3]$

$$R = \begin{bmatrix} \|a_1\| & \langle a_2, q_1 \rangle & \langle a_3, q_1 \rangle \\ 0 & \|a_2^\perp\| & \langle a_3, q_2 \rangle \\ 0 & 0 & \|a_3^\perp\| \end{bmatrix}$$

△ Cross Product

$$\begin{cases} \vec{v} \times \vec{w} = \text{area of} \\ \vec{v} \times \vec{w} = \det([\vec{v}\ \vec{w}]) \end{cases}$$

(+)

(−)

NOT THE CROSS PRODUCT

THIS IS THE PRODUCT $\vec{v} \times \vec{w}$

length = area

$$\vec{v} \times \vec{w} = \det\begin{bmatrix} i & | & | \\ j & \vec{v} & \vec{w} \\ k & | & | \end{bmatrix}$$

△ outer product

inner product

u^Tv

outer product

uv^T