Introduction to Linear Algebra

5th Edition

Gillbert Strang

Summary

Six great theorem in LA:

Dimension Th. All bases for a vec sp. have the same number of vecs.

Counting Th. Dimension of row sp. + dimension of nullspace = numbers of cols.

Rank Th. Dimension of row sp. = dimension of col sp. This is the rank.

Fundamental Th. The row sp. and nullspace of A are otho. complements in $\mathbb{R}^n$.

SVD There are ortho. bases (v's and u's for r and c sp.) so that $Av_i = \sigma_i u_i$.

Spectral Th. If $A^T = A$ there are ortho. q's so that $Aq_i = \lambda q_i$ and $A = Q\Lambda Q^T$.

LA in a nutshell:

nonsingular	singular
A is inv. rows/cols are ind.	A is not inv. rows/cols are dependent.
det A is not zero.	det A is zero.
$Ax=0$ has one solution $x=0$.	$Ax=0$ has inf. many solutions.
$Ax=b$ has one solution $x=A^{-1}b$.	$Ax=b$ has no solution or inf. many.
A has n pivots. full rank $r=n$.	A has $r<n$ pivots. rank $r<n$.
rref $R=I$. row/col sp. is all of $\mathbb{R}^n$.	rref R has at least one zero row. row/col sp. $\dim_r < n$.
A's all eigenvalues are nonzero.	zero is an eigenvalue of A.
A^TA is sym. pos. has n (pos.) singular values.	A^TA is only semipos. has $r<n$ singular values.

Matrix factorizations:

$A=LU$. $A=LDU$ ($S=LDL^T$). $PA=LU$ $EA=R$ $A=QR$ $A=X\Lambda X^{-1}$. $S=Q\Lambda Q^T$.

$A=U\Sigma V^T$ $=C^TC$). $A^+=V\Sigma^+U^T$. $A=QS$ $A=U\Lambda U^{-1}$ ($U\Lambda U^H$) $A=BJB^{-1}$ $A=QTQ^H$

Alphabet

A any matrix

L lower tri. matrix

U upper tri. matrix . left singular matrix

E elimination matrix (. eye matrix)

P permutation matrix . pascal matrix . projection matrix

I identity matrix

D diagnonal matrix

D_n ($\overline{D_n}$ / $\underline{D_n}$) lower/upper difference matrix
(forward/backward)

R rref matrix . upper tri. matrix . reflection matrix . rotation matrix
(reduced row echelon form)

Q orthogonal matrix

C cofactor matrix (companion matrix) . correlation matrix

S symmetric matrix

X eigenvector matrix

Λ eigenvalue matrix

V right singular matrix

Σ singular value matrix

J Jordan matrix

B basis matrix

1. Introduction to Vectors

1.1 vectors and linear combinations

lin. comb: $c\vec{v}+d\vec{w}$ fill a plane? solution: if right side is on the plane.

1.2 lengths and dot products

dot p : $\vec{v}\cdot\vec{w}$ perpendicular : $\vec{v}\cdot\vec{w}=0$, length $\|\vec{v}\|=\sqrt{\vec{v}\cdot\vec{v}}$.

angle $\theta=\cos^{-1}\frac{\vec{v}\cdot\vec{w}}{\|\vec{v}\|\|\vec{w}\|}$,

all vectors have $|\vec{v}\cdot\vec{w}|\leq\|\vec{v}\|\|\vec{w}\|$

$\|\vec{i}\|=1$ unit vector

1.3 matrices

$A\vec{x}$: comb of the cols. components of $A\vec{x}$ are the dots of A's rows and $\vec{x}$.

equation: $A\vec{x}=\vec{b}$ solution could write $\vec{x}=A^{-1}\vec{b}$

and some matrices don't have A^{-1}!

2. Solving Linear Equations

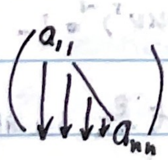
2.1 vectors and linear combinations $A\vec{x}=\vec{b}$

col picture: comb of n cols of A produces the vec $\vec{b}$.

row picture: m eqs from m rows give m planes meeting at $\vec{x}$.

$\vec{b}=0$, at least one possibility is $\vec{x}=0$.

2.2 the idea of elimination



pivot a_{ii}, multiplier $l_{ij}=\frac{a_{ji}}{a_{ii}}$ and then substract it from j.

elimination (Gauss~) $A\rightarrow U$. elimination doesn't change solutions.

$A\vec{x}=\vec{b}$: forward elimination + back substitution

elimination breaks down if zero appears in pivot.

exchanging two eqs may solve it. (zero is not allowed in pivot ← nonsingular)

when breakdown is permanent, $A\vec{x}=\vec{b}$ has no solution or inf. many.

◎ $A^2=I.A.0$ $A^3=I.A.0$

幂等矩阵的性质:

(零/一) · $A^2=0$: $C(A)$ is contained in $N(A)$, so $r \le n-r$. (见后)

· $A^2=A$: if idem. sym. mt, we call proj.

(idempotent mt.) if idem. mt. full rank, it's I

① $tr(A) = rank(A)$ ② $A \sim B$, B idem.

$\det(A)=0$ or 1. $\lambda=0.1$. $A \sim \begin{pmatrix} I_r & \\ & 0 \end{pmatrix}$

($A^3=A$ odd idem. $\lambda=0.\pm1$)

vice versa!

◦ 引入LA?

1. 线性空间 $\vec{x} = \alpha\vec{u}+\beta\vec{v}+\gamma\vec{w}$

$\vec{x} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$ $\vec{u}=\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ 向量

基底、线性变换与变基 $A\vec{x}=\vec{b}$ inv. 满秩

$\downarrow$ 一一对应

2. 分块视角. $\begin{pmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \end{pmatrix}$ 多角度AB 矩阵乘法

◦ (课) 求行列式
- 直算 / 余子式 / 降阶 Big Formular △递推 △全排列
- 消元 → 上下三角
- 大分析 Big Partition △剪子阵 shear matrix
- /加边
- 特殊阵
  - Vandermonde $V_n = \prod_{i>j} (x_i-x_j)$
  - cycle (冲项 even)
  - tridiagonal (band) triD det: $D_n = aD_{n-1} - bcD_{n-2}$ (见后 more properties.)

求逆
- 直算 / A^*
- 解方程 $AA^{-1}=I$
- 初等变换 / 加边 (消元) $E \longleftrightarrow E^{-1}(L)$; $(A\,I) \longleftrightarrow (I\,A^{-1})$ — G-J elimination
- 分块 $\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ (见2.4) △ Sherman-Morrison公式: $(A+uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1+v^TA^{-1}u}$
  esp. $(I+uv^T)^{-1} = I - \frac{uv^T}{1+v^Tu}$
- 等式变形 $A^n = I.f(A)$

◦ $\boxed{\text{左行右列}}$ $\boxed{\text{行}\overset{T}{\leftrightarrow}\text{列 伪证}}$ ◦ 矩阵等式 (张量代数) 证明

形式、公式结论 行列、向量 元 (展开、下标)

◦ 非方阵不同时存在左右逆: A m×n

(广义逆) left inv., right inv. n×m

comb 出 $\begin{pmatrix} \vdots \end{pmatrix} \cdots \begin{pmatrix} \vdots \end{pmatrix} \Rightarrow rank A \ge n.\ rank A \ge m$

而 $rank A \le n.m$ 故 $n=m$.

◦ 左逆 = 右逆

left
- $A\vec{x}=\vec{b} \Rightarrow \vec{x}=A_l^{-1}\vec{b}$ 唯一解
- $A\vec{x}=(AA_l^{-1})\vec{b}=\vec{b}$
- $\forall\vec{b} \Rightarrow AA_l^{-1}=I.\ A_l^{-1}=A_r^{-1}$

right (或T同理)
- $\vec{x}_0=A_r^{-1}\vec{b}$ 为一解
- $A_r^{-1}A\vec{x}_0=\vec{x}_0$ $\forall\vec{b} \Rightarrow \forall\vec{x}_0$
- $A_r^{-1}A=I.\ A_r^{-1}=A_l^{-1}$

如果非∀, 则无法有 $A\vec{x}\Rightarrow\vec{b}$ ∀!

inv. uniqueness: $A_1^{-1}=A_1^{-1}\cdot I = A_1^{-1}AA_2^{-1} = A_2^{-1}$.

2.3 elimination using matrices

views of matrix multiplications : $A = (\vec{a_1}, \cdots, \vec{a_n})$, $A\vec{x} = x_1\vec{a_1} + \cdots + x_n\vec{a_n}$

$E(A\ \vec{b}) = (EA\ ,\ E\vec{b})$ augmented matrix $AB = (A\vec{b_1}, \cdots, A\vec{b_n})$

elimination matrix / elemontary transform (ET) : $E_{ij} = \begin{pmatrix} 1 & & \\ & \ddots & \\ -l_{ij} & & 1 \end{pmatrix}$

permutation matrix $P_{ij} = \begin{pmatrix} 1 & & & \\ & 0 \cdots 1 & \\ & 1 \cdots 0 & \\ & & & 1 \end{pmatrix}$ (may used) → could see from ① multiply / act vecs. (rows/cols)

② do same step onto I.

one matrix both steps : $\prod PE \ldots = E$

times then.

to create $U = EA$

2.4 rules for matrix operations

$(m\times n)(n\times p) = (m\times p)$ $\underline{mnp}$ seperated muls
- → A rows · B cols (mp entries)
- → A cols · B rows (n adds)
- → A rows · B (vec · mt)
- → A · B cols ☆ (mt · vec)

$(AB)C = A(BC)$. $AB \neq BA$ usually

block multiplication, if their shapes permit. → B rows recomb (comb)

→ A cols recomb (comb)

block elimination . Schur complement: $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ $D - CA^{-1}B$

▵ left row, right col!

so Ax → recomb of A cols

→ x after a linear transform (bases (dual) transform)

2.5 inverse matrices

invertible: $AA^{-1} = A^{-1}A = I$. sq. A (means $n\times n$)

(not sq. mt do not have inv., for transpass over diff $\mathbb{R}$ $n \leftrightarrow m$) ▵ log: inv.

if all invertible, $(AB)^{-1} = B^{-1}A^{-1}$

test invertibility:

① A must have n (nonzero) pivots.

② det A must not be zero.

③ $A\vec{x} = 0$, $\vec{x} = 0$ must be the only solution.

(③' $A\vec{x} = \vec{b}$, $\vec{x} = A^{-1}\vec{b}$ be the only solution, not none or inf.)

④ rows/cols of A are independent.

(④' row/col space are $\mathbb{R}^n$. ④'' rank $A = n$)

$③ \Leftrightarrow ③' \Leftrightarrow ① \Leftrightarrow ②$; $\Leftrightarrow ④, ④', ④''$

Gauss-Jordan elimination: $O_{(G\text{-}J)} \approx n^3$

$(A\ I) \to (I\ A^{-1})$ $(A\ \vec{b}) \to (I\ \vec{x})$

(Jor means back elimination, the same essence)

□ $\begin{pmatrix} 1 & & 0 \\ -5 & 1 & \\ 0 & -4 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & 0 \\ 5 & 1 & \\ 0 & 4 & 1 \end{pmatrix}$ —— so: how you view a tri mt ?

○ $\vec{x} = A^{-1}\vec{b}$'s a math trans!

☆ elim^{-1} is no interception (coupling, rolling)

elim is rolling.

1. acts on row or col ? / how it produced ?
2. view as E or directly E^{-1} ?
3. start from tip or top ? one can be rolling.

to get an inv., must go through rolling once.

(turn an unrolling way back)

矩阵相乘 (A=B·inv.) 与行/列空间的基 (re.comb.)

- useful justify dependent/uninv.: r/c s adds to zero, or one has combs of others.

blockize: better sq. in diag.

- 矩阵收集仓 (-):

sth. Ⓐ sth. dual (-1, T..) is a good thing. for A pop out can merge into (sth.·B).

  - adjacency matrix $S = \begin{pmatrix} 0 & 1 & \cdots \\ 1 & 0 & \cdots \\ \vdots & & \ddots \end{pmatrix}$ (无向图为 sym.) $S^2 \ldots S^n$: length n, node $i \to j$
  - sum matrix $S = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 1 & & 1 \end{pmatrix}$
  - difference matrix $\overline{D} = \begin{pmatrix} 1 & -1 & \\ & \ddots & \ddots \\ 0 & & 1 \end{pmatrix}$ $\underline{D} = \begin{pmatrix} 1 & & 0 \\ -1 & \ddots & \\ 0 & -1 & 1 \end{pmatrix}$. $\underline{D}^{-1} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 1 & & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & & 1 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$
    n.th diff matrix D_n $\quad D_n^{-1} = S^n$
  - cycle matrix (zero-cycle matrix means adds equal to zero, $C = \begin{pmatrix} a & b & \cdots \\ b & c & \cdots \\ c & a & \cdots \end{pmatrix}$ it's uninv.) (九口)
  - Pascal matrix / binomial coff. matrix

$P_L = \begin{pmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 2 & 1 & \\ 1 & 3 & 3 & 1 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix}$ $\quad D = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \\ & & & -1 \\ & & & & \ddots \end{pmatrix}$ $\quad \underline{P_L D}\,\underline{P_L D} = I$

$DP_LD = \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ 1 & -2 & 1 & & \\ -1 & 3 & -3 & 1 & \\ 1 & -4 & 6 & -4 & 1 \end{pmatrix}$ $\quad$ sym. pascal matrix $P_{sym.} = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots \\ 1 & 2 & 3 & 4 & \\ 1 & 3 & 6 & & \\ 1 & 4 & & \cdots & \end{pmatrix} = P_L P_U$ (so $|P_{sym}| = 1$)

  - diagonally dominant matrix $a_{ii} > \sum_{j \neq i} a_{ij}$ is inv. for $A\vec{x} = 0$ only can be zero.

- cyclic matrix 循环矩阵 (见后) $\quad C = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & & \cdots & a_{n-1} \\ a_{n-1} & a_n & & \ddots & \\ \vdots & \vdots & & & \vdots \\ a_2 & a_3 & \cdots & & a_1 \end{pmatrix}$

  - basic cyclic matrix $J = \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & 0 & \ddots & \\ 0 & 0 & & 1 \\ 1 & 0 & \cdots & 0 \end{pmatrix}$, $J^k = \begin{pmatrix} 0 & I_{n-k} \\ I_k & 0 \end{pmatrix}$ (k mod n)

$C = a_1 I_n + a_2 J + \cdots + a_n J^{n-1} = g(J)$ eigenfunction.

  - eigen value (J) $|\lambda I_n - J| = \lambda^n - 1$. $\quad$ ($\{I_n, J, \ldots J^{n-1}\}$ a base in $\mathbb{C}_n$)

$1 \le k \le n-1$ (uniroot) $\omega_k = \cos\frac{2k\pi}{n} + i \sin\frac{2k\pi}{n}$ $\quad$ eigenvector $\alpha_k = (1, \omega_k, \omega_k^2, \ldots \omega_k^{n-1})^T$

$X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega_1 & \cdots & \omega_{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \omega_1^{n-1} & \cdots & \omega_{n-1}^{n-1} \end{pmatrix}$ then $\Lambda = \mathrm{diag}(1, \omega_1, \ldots, \omega_{n-1})$

$J = X \Lambda X^{-1}$,

<可参考 FFT. DFT> $\quad C = X g(\Lambda) X^{-1}$, so C is diagonizable.

$|C| = \prod g(1) \cdots g(\omega_{n-1})$. (non-zero cyclic means $g(1) \neq 0$)

  - some properties: ① n dim complex matrix B is diagonizable <=> B is similar to a cyclic matrix.

② if C is a cyclic matrix, so does C^*, C^{-1}. (use dignonalize)

LDU LDU

(triangular / Doolite / Crout decompsition)

2.6 elimination = factorization : $A = LU$ $\qquad$ A_k upperleft submt ✓ $\rightarrow PA = LU$. (or $A = L'P'U$)

the whole forward elimination process (with no row exchanges) is inverted by L, which still lower triangular, every multiplier l_{ij} stands explicitly at $r_i c_j$.

and to unitize, pops D (a diag) out. $\qquad$ uniqueness: $L_1 D_1 U_1 = L_2 D_2 U_2$

(lu itself isn't unique) leads to $L_1^{-1} L_2 D_2 = D_1 U_1 U_2^{-1}$

$L = U$, must be diag!

so $L_1^{-1} L_2 = I$, $L_1 = L_2$ ...

solving a tri system : $O(\frac{n^2}{2})$ (backward)

elimination / solving $A\vec{x} = \vec{b}$: $O(\frac{n^3}{3})$

° again: echelon! no interact.

(先简后) row 3 of U = row 3 of A − l_{31} row 1 of U − l_{32} row 2 of U

that's row 3 of A = $(l_{31}\ l_{32}\ 1) \cdot \begin{pmatrix} \text{row 1 of } U \\ \text{row 2 of } U \\ \text{row 3 of } U \end{pmatrix}$

o' there also can be writtn $A = LU'D$, depends on how you pop D or how U' acts (r/c?)

the algorithm of $A\vec{x} = \vec{b}$: 1. factor into $A = LU$ (muler stored in L) $\quad \frac{n^3}{3} \quad nw^2$

2. solve $Lc = b$ ($b \xrightarrow{L^{-1}} c$) $\quad n^2 \quad 2nw$

$Ux = c$ (backward substitution).

sparse : faster. for a random / full mt. $n = 1000$ on a PC takes 1 sec, then n^3 increases.

$\begin{pmatrix} 1&1&0&0 \\ 1&2&1&0 \\ 0&1&2&1 \\ 0&0&1&2 \end{pmatrix} = \begin{pmatrix} 1&&& \\ 1&1&0& \\ 0&1&1& \\ 0&0&1&1 \end{pmatrix} \begin{pmatrix} 1&1&0&0 \\ &1&1&0 \\ 0&&1&1 \\ &&&1 \end{pmatrix}$ the zeros at start (r or c) will still be reserved in L or U.

w is the band width.

□ Cholesky factorization (sq. root factorization) $A = LL^T$, when A is sym. and positive.

(lemma : B is triang., $A = BB^T$ is sym. pos. for $x^T A x = (xB)^T (xB) > 0$) (B's row axes)

$A = LDU = A^T = U^T D L^T \longrightarrow L = U^T$.

$A = (L\sqrt{D})(L\sqrt{D})^T$ $\qquad$ (lu chooses the largest pivot in col in algorithm.)

compare: Lu $\quad n^2 + (n-1)^2 + \cdots \quad \frac{n^3}{3}$

Chol $\quad \frac{n^2}{2} + \frac{(n-1)^2}{2} + \cdots \quad \frac{n^3}{6}$ half ← sym.'s credit.

to reduce error. since $LUx = b$ not stable (big swallow small)

algo : (recur) $A = \begin{pmatrix} a_{11} & A_1^T \\ A_1 & A' \end{pmatrix} = \begin{pmatrix} l_{11} & \\ L_1 & L' \end{pmatrix} \begin{pmatrix} l_{11} & L_1^T \\ & L'^T \end{pmatrix}$, $l_{11} = \sqrt{a_{11}}$, $L_1 = \frac{1}{l_{11}} A_1$.

$\underbrace{L' L'^T = A' - L_1 L_1^T}_{O(\frac{n^2}{2})}$ —— recursion.

(end: $\sqrt{A'}$. that's $\sqrt{a_{nn}}$)

└ sym.

○ 矩阵收集仓(2):

· Toeplitz matrix $T = \begin{pmatrix} t_0 & t_{-1} & \cdots & t_{-n+1} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ t_{n-1} & \cdots & & \end{pmatrix}$ persymmetric: $A^T = JAJ$ 中心对称 (旋 π)

次对称

· reversal matrix $J = \begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix}$ persym.$^{-1}$ is also persym.

cyclic matrix is a esp. Toeplitz matrix.

□ full-rank decompsition / CR factorization

满秩分解 · $A = CR$ C contains r independent cols of A. (a base of $\mathcal{C}(A)$)

$R = (I\ F)P$, $A = (C\ CF)P$ F expresses the remaining n−r dependent cols

(P permutes them according to A)

iow, the rows of R can be a base of row space of A ($\mathcal{R}(A)$) (expressed by C)

row R is independent, for it contains I.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

In space view, A: $\mathbb{R}^n \to \underset{\mathcal{R}(A)}{\mathbb{R}^r} \xrightarrow{W^{-1}} \underset{\mathcal{C}(A)}{\mathbb{R}^r} \to \mathbb{R}^m$

· $A = CW^{-1}B$ B also contains r actual rows of A.

if $P = I$, $A = \begin{pmatrix} W & H \\ J & K \end{pmatrix} = \begin{pmatrix} W \\ J \end{pmatrix} W^{-1}(W\ H)$ (so $R = W^{-1}B = (I, W^{-1}H)$. $B = WR$)

$K = JW^{-1}H$

W is inv., where B meets C at r×r positions (base), which called 'r-c intersection' or 'kernel'.

(for a sq. and inv. A, it reduced to $A = WW^{-1}W$ or $A = C = B = W$)

(considering P is nowhere any diff, you can AP rather than A, or $CW^{-1}B$ not influenced.)

(to get (I F) kind like) CR

· Uniqness: C is one basis you choose, then R is unique, but W, B aren't unique.

This is consistent with the uniqueness of the rref R_0, which has non-uniqueness steps from A to R_0.

□' dual $\vec{w} = a\vec{u} + b\vec{v}$ $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}^{-1}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} u^*(\vec{w}) \\ v^*(\vec{w}) \end{pmatrix} = \begin{pmatrix} u^* \\ v^* \end{pmatrix}\vec{w}$

对偶浅说

dual. stand ortho. basis (mutual) $u^*(\vec{u}) = 1$, $u^*(\vec{v}) = 0$, $u^*(\vec{w}) = a$

linear functional ~ row vec. $v^*(\vec{u}) = 0$, $v^*(\vec{v}) = 1$, $v^*(\vec{w}) = b$

(act: dot)

$\Rightarrow \vec{u}^* = \begin{pmatrix} \frac{v_2}{|u,v|} \\ -\frac{v_1}{|u,v|} \end{pmatrix}^T$ $\vec{v}^* = \begin{pmatrix} -\frac{u_2}{|u,v|} \\ \frac{u_1}{|u,v|} \end{pmatrix}^T$

$$\begin{tikzcd} V \arrow[r, shift left, "A"] \arrow[d, dash] & W \arrow[l, shift left, "A^{-1}"] \arrow[d, dash] \\ V^* & W^* \arrow[l, "A^T"'] \end{tikzcd}$$

$\vec{u}, \vec{v}$; $\vec{w}$; $\vec{u}^*, \vec{v}^*$; $\vec{w}^*$; $\xrightarrow{(A^{-1})^T}$

· $(\varphi u, v) = (u, \varphi^{adj.} v)$ $\varphi: \overset{u}{V} \to W^{v}$ · pull-back:

on the dual basis, φ, $\varphi^{adj.}$ $\varphi^{adj.}: V^* \leftarrow W^*$ $A: V \to W$. f on W.

induced mt. are transpose. $(A^T f)(x) = f(A(x))$

2.7 transposes and permutations

$(Ax)^T = x^T A^T$, $(AB)^T = B^T A^T$, $(A^{-1})^T = (A^T)^{-1}$ note as A^{-T}

inner product. $x \cdot y = x^T y = y^T x$ the idea behind A^T is $Ax \cdot y = x \cdot A^T y$

symmetric matrix $S^T = S$. antisymmetric matrix $S^T = -S$ (skew-sym.)

and $A^T A$ always sym. orthogonal matrix $Q^T = Q^{-1}$ ($QQ^T = I$)

(and lu is $S = LDL^T$) it's standard (unit/orthonormal)

cols of Q are ortho. unitvectors. (orthoaxes)

permu. mt. P puts $x_1, \dots, x_n$ in new order, and $P^T = P^{-1}$, and can be combed by several P_{ij}.

($n!$ kinds, half odd and half even)

$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ ① 312 ② 231 ($P_{ij} = P_{ij}^T = P_{ij}^{-1}$)

3. Vector Spaces and Subspaces

3.1 spaces of vectors Vector Space S ← vec can be matrices even funcs of x.

$\mathbb{R}^n$ contains all real col.vec with n components. if v, w in S, every comb $cv + dw$ must be in S.

Subspace of $\mathbb{R}^n$ is a vecspace inside $\mathbb{R}^n$, (closure to adds and muls) must contain zero.

(Z: one-point space consists of $x = 0$)

col space of A contains all combs of A's cols, which is a subspace of $\mathbb{R}^n$.

col space contains all vecs Ax. so $Ax = b$ is solvable when b is in $C(A)$

we call it's spanned by A's cols. $C(A)$: $\mathrm{Im} A$ (image) $N(A)$: $\mathrm{Ker} A$ (kernel)

3.2 the nullspace of A: solving $Ax = 0$ and $Rx = 0$

null space $N(A)$ (a subspace in $\mathbb{R}^n$) contains all solution x to $Ax = 0$. and it must contain Z (zero).

elimination from A to U to R does not change the nullspace $N(A) = N(U) = N(R)$

$R = \mathrm{rref}(A)$ means reduced row echelon form, has all pivots $= 1$, with zeros above and below.

numbers of pivots = numbers of nonzero rows in R = rank rows (also rank of cols)

every matrix with $m < n$ has nonzero solutions in its N.

There are $n - r$ free cols (no pivot in). their combs is the N (the comb to produce itself, or $x_j = 1$) complete solution of $Ax = 0$ or $N(A)$ ($Rx = 0$)

so we have $\dim C(A) + \dim N(A) = n$.

(for R (pivot rows and cols) contains I, Z → $\begin{pmatrix} 1 & | & | \\ & 1 & | \end{pmatrix}$)

The rank of A is the true size of A or a linear system.

A and U and R have r ind. rows/cols.

'special solution' s = − free colvec with $x_j = 1$ and zeros.

□ full-rank factor: $A = $ (pivot cols of A)(first r rows of A) $m \times n = (m \times r)(r \times n)$

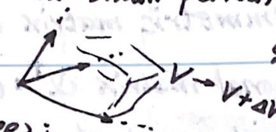
- 几何想象LA?

  1. basic: 基、变换、特征向量、行列式、Cramer's rule   4. SVD (见后)

  2. det and tr: $\det = \prod\lambda$, $tr = \sum\lambda$, $tr = \det'$

① $y' = Ay$, $y = e^{tA}y_0$ in small period, trace is the velocity of volume change.

$|e^{tA}| = 1 + t \cdot trA + O(t^2)$



small changes mainly in the edges, on its own direction.

② intrinsic (coordinate free):
mt. is depicted wholely by traces. $tr(A) + \cdots + tr(A^n) = \lambda_1^n + \cdots + \lambda_n^n$.

③ fix-point: $tr\,\phi = \sum_i \langle \phi(e_i), e_i \rangle$ (like 'how much remained?')

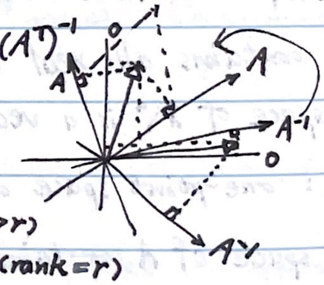
esp. $A^2 = A$. the rank of A = A's dim of fix-point spanned space = trA

for full fix, $\langle \phi, e_i \rangle = 1$.

3. transpose: $A^T x$: 在向列向量基上投影(实为点乘), 注意与逆区分.

用此解释: ① $\langle x, Ay \rangle = \langle A^T x, y \rangle$ ② $(AB)^T = B^T A^T$ ③ $Q^T = Q^{-1}$, $Q^T Q = I$ (not unit: D^2)

④ $(A^{-1})^T = (A^T)^{-1}$

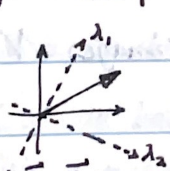


- 几个证明与理解:

  1. $r_{row} = r_{col}$ → full-rank decomp.
     → rref. pivots
     → $\boxed{W}^r$ □ suppose $r_{row} < r_{col}$,
     then the remaining cols (>r) could be expressed by W (rank=r)
     → $A = C \cdot B$ ($m\times n$, $m\times r$, $r\times n$)   取r_{col}组成C, 用B表达。
     反过来r行被C组为A;
     $r_{row}(A) \le c = r_{col}(A)$.
     同理 $r_{col}(A) \le r_{row}(A)$. 故equal.

  2. $|A| = |A^T|$ → geometry analogy; LU/rref 理解; SVD 理解; 全排列定义

symmetric:



since $\langle \vec{x_1}, A\vec{x_2} \rangle = \langle A\vec{x_1}, \vec{x_2} \rangle$, if $\vec{x_2}$ is eigen direction, then $\vec{x_1} \cdot \vec{x_2} = 0 \Rightarrow A\vec{x_1} \cdot A\vec{x_2} = \lambda^2 \vec{x_1} \cdot \vec{x_2} = 0$.

so A maintains the ortho of eigen direction. sym.mt's eigen directions are ortho basis, ($\vec{x_1}$ also is)

(won't keep the origin direc.)   its diagonalize just rotate the axes.

(if $\lambda_1 = \lambda_2$, it's true that the whole plane is eigen.)   so $S = Q\Lambda Q^T$ (orthonormal)

(another algebraic proof 见后) and 2 for λ inequal.

- 更多关于R的:

  R could be write as $\begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix} P$.

  'row-col reduced form' $R' = (\mathrm{rref}(R^T))^T = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$

  - $A^T A$ always has the same N with A. ($A^TAx = 0$ 与 $Ax=0$ 为同解方程组)

  proof: ① $Ax_0 = 0 \Rightarrow A^T(Ax_0) = 0$
  $A^TAx_0 = 0 \Rightarrow x_0^T A^T A x_0 = 0$, $(Ax_0)^T(Ax_0) = 0$. $Ax_0 = 0$.

  ② 见后. 正交性或写出R观察. ($C(A) \perp N(A^T)$, so $A^TAx = 0$ only when $Ax = 0$)

  ▷ Generably, $r(A^TA) = r(A)$ ($= r(AA^T)$) ($= r(\Lambda) = r(\Sigma)$)

3.3 the complete solution to $Ax=b$

- again: If a row/col contains a pivot, it's not a comb of previous rows/cols, otherwise is.

(以前文) U tells us which row/col are combs of earlier ones; R tells us what those combs are and the special solution to $Ax=0$.

R is always the same (determined by A).

Three fudmental subspaces: $\mathcal{C}(A)$ — choose the pivot cols of $\underset{\Delta}{A}$ as a basis.

$\mathcal{R}(A)$ — choose the nonzero rows of R as a basis.

$\mathcal{N}(A)$ — choose the special solution to $Rx=0$ ($Ax=0$)

Counting Theorem:

$$\underset{}{n} = \underset{\text{pivot/ind. vars}}{r} + \underset{\text{free vars}}{n-r}$$

$(A\, I) \longrightarrow (R, E)$ will virtually tell you everything about A, when A is sq. and inv., R is I and E is A^{-1}.

pivot vars are determined after free vars are chosen.

Complete solution to $Ax=b$: x = (one particular solution x_p) + (any x_n in nullspace) (thus comb of special solution basis)

$$(A\,\vec{b}) \longrightarrow (R\,\vec{d})$$

$Ax=b$ / $Rx=d$ is solvable only when all zero rows of R have zeros in $\vec{d}$.

full row rank $r=m$ when its col space $\mathcal{C}(A)$ is $\mathbb{R}^m$, $Ax=b$ is always solvable, but may many.

full col rank $r=n$ when its nullspace $\mathcal{N}(A)=Z$, no free vars, one solution or none.

When is solvable, one particular x_p can be all free vars equal to zero, pivot vars from $\vec{d}$.

To conclude, there are four cases:

ranks	$r=m=n$	$r=m<n$	$r=n<m$	$r<m, r<n$
types of R	(I)	$(I\ F)$	$\begin{pmatrix} I \\ 0 \end{pmatrix}$	$\begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$
conclusion	A is inv.	every $Ax=b$ is solvable, but inf.	$Ax=b$ has 1 or 0 solutions	0 or ∞ solutions

o' if $Ax=b$ and $Cx=b$ have same solutions for every b, that means $A=C$.

(proof: let $x=(1,0,0\cdots)$ to get colvec $1_{A=C}$)

3.4 independence, basis and dimension

linear independence: $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = 0$ only happens when all x's are zero.

Any set of n vecs in $\mathbb{R}^m$ must be dependent if $n>m$.

basis: (linearly) independent vectors that span the space, like pivot cols of A for $\mathcal{C}(A)$.

every vector in the space is a unique comb of the basis vecs. (empty set for Z)

dimension: All bases for a space have the same number of vecs called dim, like r_A for $\mathcal{C}(A)$.

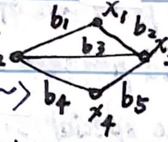
Basis of $\vec{v}_i$'s from basis of $\vec{w}_i$'s when the change of basis matrix $\underset{B}{}$ is inv. $V=WB$

dim of outputs + dim of nullspace = dim of inputs. Δ $\dim(V+W) + \dim(V\cap W) = \dim V + \dim W$

- ° $\vec{x}=(x_1,x_2,\cdots,x_n)$ 的全排列 ($n!$个) 张成的 S, dim S 有四种: ① 0 ② 1 ③ $n-1$ ④ n
  (③: find a vec that dot to zero / always perpendicular to $\vec{x}$.)
- ° AB 's rows are combs of B's rows. AB 's cols are combs of A's cols.
  (so its row/col space contain in B/A's) $\text{rank}(AB)\le\text{rank}(A).\ \text{rank}(B)$. (矩阵乘法的秩)
  If multiply by an inv. mt, the rank will not change, because rank can't jump back simply through acts. (mul an inverse mt.)
  so if $AB=0$, we have $C(B)\subset N(A)$ / $C(A^T)\subset N(B^T)$, $r_A+r_B\le n$. ($A: m\times n,\ B: n\times k$)

□ Kirchhoff 's Law · graphs and incidence matrix:

<可参考 基尔霍夫陵 关联矩阵、邻接~>



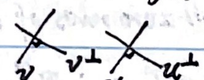
$$Ax=b,\quad A=\begin{pmatrix}-1&1&&\\-1&&1&\\&-1&1&\\&-1&&1\\&&-1&1\end{pmatrix}$$

since $\vec{b}=0 \Rightarrow x_1=x_2=x_3=x_4$
thus $\vec{x}=(c,c,c,c)$ in $N(A)$,
the rank must be 3.

since $b_1-b_2-b_3$ is a loop, rows 1.2.3 of A
are not lin. ind. but $(1,-1,1,0,0)$ will be in A's left nullspace.
(two ind. loops, $m-r=5-3=2$.)

- Kirchhoff 's Voltage Law $Ax=b$
  Kirchhoff 's Current Law $A^Ty=0$ → 'balance equation'. most important eq. in applied maths.
  (y_i flows on b_i edge)

- ° $A=uv^T$'s four subspaces are



. If B have those same subspaces, then $B=cA$.
Generally, mt. A and B have same four subspaces means their rref equals.
(proof: same row space is the key. every row in rref(A) must be comb of rref(B), but notice I in these rref, it means each row should be equal in rref.)

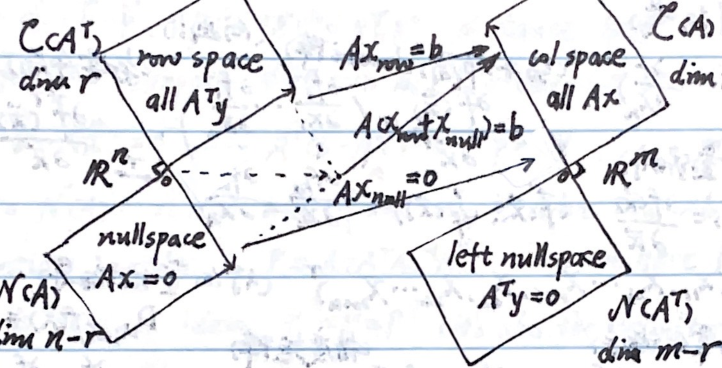
- ° More Advanced Level: Matrix Spaces and Function Spaces
  - dim of the subspace of sym. mts. is $\frac{1}{2}n^2+\frac{1}{2}n$.
  - second derivative equtions: solution space
    - $y''=-y$ has two basis funcs: $\sin x$ and $\cos x$.
    - $y''=y$: e^x and e^{-x}.
    - $y''=0$: x and 1.
    - ...

    but $y''=2$ don't form a subspace, a particular solution $y_p=x^2$. then the complete solution is $y=x^2+cx+d$.
  - Ex. $A=\begin{pmatrix}1&0&-1\\-1&1&0\\0&-1&1\end{pmatrix}$, notice $A\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$. $AX=B$, $X\in 3\times3$ mt. space. The counting theorem writes $\underset{C}{6}+\underset{N}{3}=9$.

- ° Fredholm's Alternative: Exactly one of these problems has a solution $\begin{cases}Ax=b\\A^Ty=0 \text{ with } y^Tb\neq0\end{cases}$

3.5 dimensions of the four subspaces

The Big Picture:



1. A has the same row space as R. same dim r and same basis.
2. The col space of A has dim r. col rank equals row rank. (Ranking Theorem) ▷ $\mathcal{C}(A) \neq \mathcal{C}(R)$! but dim ✓. (for same combs of col $=0$ for A and R, thus $Ax=0 \Leftrightarrow Rx=0$) ($\mathcal{R}(A)=\mathcal{R}(R)$)
3. A has same nullspace as R. same dim $n-r$ and same basis. (again, doesn't change solutions)
4. The left nullspace of A has dim $m-r$. (Couting Theorem or $R^Ty=0$ / $y^TR=0$, $R=EA$)

Fundamental Theorem of Linear Algebra:

Part 1. The col space and row space both have dim r;
The nullspaces have dim $n-r$ and $m-r$.

rank-one matrix: $A=uv^T=$ col times row. $\mathcal{C}(A)$ has basis u, $\mathcal{C}(A^T)$ has basis v.

rank-one's sum: $EA=R$, $A=E^{-1}R=CR$ Ex. $A=\begin{pmatrix}1&0&3\\1&1&7\\4&2&20\end{pmatrix}=\begin{pmatrix}1&0&0\\1&1&0\\4&2&1\end{pmatrix}\begin{pmatrix}1&0&3\\0&1&4\\0&0&0\end{pmatrix}=CR$.

(pre-read: ▫ full-rank decomp.) Every rank r mt. is a sum of r rank-one mts:
∘ left eigenvecs.

(rank-one expansion) $A=(\vec{u_1}\ \vec{u_2}\ \vec{u_3})\begin{pmatrix}\vec{v_1}^T\\ \vec{v_2}^T\\ \text{zero row(s)}\end{pmatrix}=\vec{u_1}\vec{v_1}^T+\vec{u_2}\vec{v_2}^T=\text{rank}1+\text{rank}1$

4. Orthogonality

4.1 orthogonality of the four subspaces

The orthogonal complement of a subspace V contains every vector that is perpendicular to V. denoted by $V^{\perp}$

orthogonal vectors $v^Tw=0$ and have $\|v\|^2+\|w\|^2=\|v+w\|^2=\|v-w\|^2$.
orthogonal spaces $v^Tw=0$ for all v in V and all w in W.
Orthogonality is impossible when $\dim V+\dim W > \dim$ (whole space) $(V+W)$ (since it only can be zero to be contained in both ortho spaces.)

Part 2. $\mathcal{N}(A)$ is the ortho complement of $\mathcal{C}(A^T)$; (in $\mathbb{R}^n$)
$\mathcal{N}(A^T)$ is the ortho complement of $\mathcal{C}(A)$. (in $\mathbb{R}^m$)

This means every x can be split into x_r+x_n. When A multiplies, nullspace component x_n goes to zero $Ax_n=0$, rowspace component x_r goes to colspace $Ax_r=Ax$.

$v^Tv=0 \Rightarrow v=0$, a row of A can't be in $\mathcal{N}(A)$.

Every b in colspace comes from one and only one vec x_r in rowspace.

There is an $r\times r$ inv. mt. hiding inside A, if throw away two nullspaces.
From the rowspace to the colspace, A is inv. ('pseudoinverse', see 7.4)

∘' the definition of basis has two properties, but one implies the other: ind. $\Leftrightarrow$ span $\mathbb{R}^n$ (n vecs)

○' 更多关于 tri.mt 的: L/U 的 inv. 仍为 L/U; L_1L_2 / U_1U_2 积仍为 L/U; $LX=L'$/$UX=U'$ 则 X 一定也为 L/U.
(且对角线 $d=d_1d_2$)

□ matrix derivative (向量值函数)

矩阵求导 · 定义概念. vec.func. $\vec F(x)$ $f(\vec x)$ $\vec f(X)$ mt.func. $F(X)=(F_{ij}(X_{m\times n}))$

(标函:) (默认列向量)

向元: $\frac{\partial f(\vec x)}{\partial \vec x} = \begin{pmatrix}\frac{\partial f}{\partial x_1}\\ \vdots \\ \frac{\partial f}{\partial x_n}\end{pmatrix} \xrightarrow{T} \frac{\partial f(\vec x)}{\partial \vec x^T} = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right)$.

(向函:) (分子布局. Jacobi 阵) $\frac{\partial \vec f(\vec x)}{\partial \vec x^T} = \begin{pmatrix}\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n}\\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n}\end{pmatrix}$. $\xleftrightarrow{T}$ (分母布局) $\frac{\partial \vec f^T(\vec x)}{\partial \vec x} = \begin{pmatrix}\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1}\\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_m} & \cdots & \frac{\partial f_n}{\partial x_m}\end{pmatrix}$

(行偏导) (梯度/列向量偏导)

(向量) $D_x f(\vec x) = \frac{\partial f(\vec x)}{\partial \vec x^T} \xleftrightarrow{T} \nabla_x f(\vec x) = \frac{\partial f(\vec x)}{\partial \vec x}$ ($\vec f(\vec x)$、$J_f(\vec x)$)

矩元: vec(X): (列堆栈) 向量化 $=(x_{11}, x_{21} \cdots x_{m1}, x_{12}, x_{22} \cdots x_{mn})^T$ (行向量化偏导)

$D_{vecX} f(X) = \left(\frac{\partial f}{\partial x_{11}}, \frac{\partial f}{\partial x_{21}}, \dots, \frac{\partial f}{\partial x_{m1}}, \frac{\partial f}{\partial x_{12}}, \dots, \frac{\partial f}{\partial x_{mn}}\right)$

(Jacobi 阵) (列向量化偏导) (梯度矩阵)

$D_X f(X) = \begin{pmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{m1}}\\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{1n}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{pmatrix}_{n\times m}$. $\nabla_{vecX} f(X) = \frac{\partial f(X)}{\partial vecX} = \left(\frac{\partial f}{\partial x_{11}} \cdots \frac{\partial f}{\partial x_{mn}}\right)^T$ $\nabla_X f(X) = \begin{pmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{pmatrix}_{m\times n}$

同理矩函也有 $D_X F(X) = \frac{\partial vec(F(X))}{\partial vecX}$. $\nabla_X F(X) = \frac{\partial vec^T(F(X))}{\partial vecX}$ △ Transpose 的 T 位置可以变动.

· 向元标函: $\frac{\partial c}{\partial \vec x}=0$ $\frac{\partial f(\vec x)g(\vec x)}{\partial \vec x} = \frac{\partial f(\vec x)}{\partial \vec x} g(\vec x) + f(\vec x)\frac{\partial g(\vec x)}{\partial \vec x}$ 等其他老规律. 常用新规律:

1. $\frac{\partial (x^Ta)}{\partial x} = \frac{\partial (a^Tx)}{\partial x} = a$ 2. $\frac{\partial (x^Tx)}{\partial x} = 2x$ 3. $\frac{\partial (x^TAx)}{\partial x} = Ax + A^Tx$ 4. $\frac{\partial (a^Txx^Tb)}{\partial x} = (ab^T+ba^T)x$

矩元标函: 5. $\frac{\partial (a^TXb)}{\partial X} = ab^T$, $\frac{\partial (a^TX^Tb)}{\partial X} = ba^T$ 7. $\frac{\partial (a^TXX^Tb)}{\partial X} = (ab^T+ba^T)X$, $\frac{\partial (a^TX^TXb)}{\partial X} = X(ab^T+ba^T)$

· 矩元标函 $df(X) = \frac{\partial f}{\partial x_{11}}dx_{11} + \dots + \frac{\partial f}{\partial x_{mn}}dx_{mn} \triangleq tr\left(\frac{\partial f(X)}{\partial X^T}dX\right)$ 常用新规律:

1. $d(AXB) = A\,dX\,B$ 2. $d|X| = |X|tr(X^{-1}dX) = tr(X^*dX)$ 3. $dX^{-1} = -X^{-1}dX X^{-1}$ ($XX^{-1}=I$ 作微分)

△ $df(X) = d(tr f(X)) = tr(df(X))$ (标量迹不变)

($d(X^T)=(dX)^T$)

· 例: 1. $\frac{\partial (a^TXX^Tb)}{\partial X}$: $d(a^TXX^Tb) = tr(a^Td(XX^T)b) = tr(a^TdXX^Tb) + tr(a^TXdX^Tb)$

$\xrightarrow[tr(A)=tr(A^T)]{tr(AB)=tr(BA)} = tr(X^Tba^TdX) + tr(X^Tab^TdX)$

故 $\frac{\partial (a^TXX^Tb)}{\partial X^T} = X^T(ba^T+ab^T)$,

$\frac{\partial (a^TXX^Tb)}{\partial X} = (ab^T+ba^T)X$ 2. $\frac{\partial tr(X^TX)}{\partial X} = 2X$: $d(tr(X^TX)) = tr(dX^TX) + tr(X^TdX) = 2tr(X^TdX)$

故 $\frac{\partial tr(X^TX)}{\partial X^T} = 2X^T$, $\frac{\partial tr(X^TX)}{\partial X} = 2X$

3. $\frac{\partial \log|X|}{\partial X} = X^{-T}$: $d(\ln|X|) = tr\left(\frac{1}{|X|}d|X|\right) = tr(X^{-1}dX)$

4. $\frac{\partial tr(X+A)^{-1}}{\partial X} = -((X+A)^{-2})^T$

5. $\frac{\partial |X^3|}{\partial X} = 3|X^3|X^{-T}$ 6. 概统估计: $x \sim N(\mu, \Sigma)$

$f(x) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

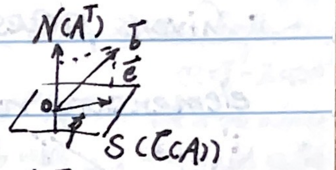
极大似然 $\ln L(\mu, \Sigma) = \sum_i \ln f(x_i) = -\frac{p}{2}n\ln 2\pi - \frac{1}{2}n\ln|\Sigma| - \frac{1}{2}\sum_i (x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$

有 $\left\{\begin{aligned} &\frac{\partial \ln L(\mu, \Sigma)}{\partial \mu} = \Sigma^{-1}\sum_i (x_i-\mu) = 0 \\ &\frac{\partial \ln L(\mu, \Sigma)}{\partial \Sigma} = 0 \end{aligned}\right.$ (见后)

1. 4.2 projections

The projection of $\vec{b}$ onto a subspace S (m dim.) is the closest vector $\vec{p}$ in S with $\vec{b}-\vec{p}$ is ortho to S.

error: $\vec{e}=\vec{b}-\vec{p}$. $\|\vec{p}\|^2+\|\vec{e}\|^2=\|\vec{b}\|^2$. distance: $\|e\|$



A's cols span S, suppose they are ind. (otherwise use those r ind. cols)

$\vec{a}_i\cdot(\vec{b}-A\hat{x})=0 \longrightarrow A^T(b-A\hat{x})=0$, $p=A\hat{x}$. $A^Tb=A^TAx$

$N(A^TA)=N(A)$: A^TA is inv. if and only if A has ind. cols. so $p=A(A^TA)^{-1}A^Tb$.

The projection matrix is $P=A(A^TA)^{-1}A^T$ △ we have $P_CAP_R=A(=P_CA=AP_R)$. $P_{C/R}$: proj. to col/row space.

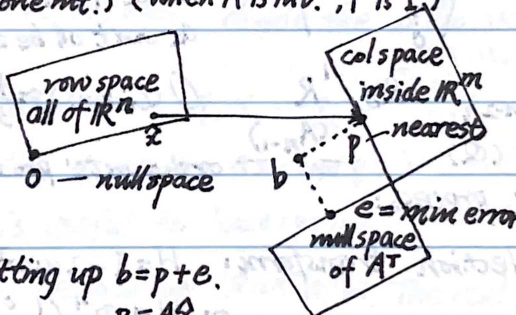
proj. mt is sym. sq. idem. $P^2=P=P^T$ (it's also the definition of proj. mt.). rank$(P)=$ rank$(A)=r$ $(P\sim\begin{pmatrix}I_r & \\ & 0\end{pmatrix})$

(for $\vec{a}$ vec, $\hat{x}=\frac{a^Tb}{a^Ta}$, $p=\frac{aa^T}{a^Ta}$ a rank-one mt.) (When A is inv., P is I.)

4.3 least squares approximations

$A=\begin{pmatrix}1 & t_1\\ \vdots & \vdots\\ 1 & t_m\end{pmatrix}$ $x=\begin{pmatrix}C\\D\end{pmatrix}$ $b=\begin{pmatrix}b_1\\ \vdots\\ b_m\end{pmatrix}$

$(b=C+Dt$
$m>2, n=2)$



△ since first col of A is $(1,1\cdots,1)$, $\Sigma e_i=0$ or t mean will go to b mean. $(\bar{b}=C+D\bar{t})$

There are no solutions to $Ax=b$, we splitting up $b=p+e$.

$\hat{x}$ is the least square solution, for the square length of $Ax-b$ is minimized: $p=A\hat{x}$ $E=\|Ax-b\|^2$ min. $(=\Sigma e_i^2)$

By geometry: the nearest point is the projection. solve $A^TAx=A^Tb$.

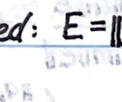
By algebra/calculus: the partial derivatives of E are zero. $\frac{\partial E}{\partial C}=0$, $\frac{\partial E}{\partial D}=0$ or $\frac{\partial E}{\partial x}=0$ as a whole, we get $(Ax-b)^TA=0$.

$A^TA=\begin{pmatrix}m & \Sigma t_i\\ \Sigma t_i & \Sigma t_i^2\end{pmatrix}$ $A^Tb=\begin{pmatrix}\Sigma b_i\\ \Sigma t_ib_i\end{pmatrix}$, $\hat{x}=\begin{pmatrix}C\\D\end{pmatrix}=(A^TA)^{-1}A^Tb$

If $\Sigma t_i=0$ then A has ortho cols and A^TA is diag. It's worth shifting times t_i by substacting the average $\hat{t}=\frac{\Sigma t_i}{m}$, $T=t-\hat{t}$ or make cols ortho. in advance by Gram-Schmit. $b=C+D(t-\hat{t})$

Like fitting a straight line, the unknowns $C+Dt+Et^2$ still appear linearly in fitting a parabola. Then $\hat{x}=(C,D,E)$, nothing diff.

(if A has dependent cols, we have many solution lines, pseudoinv. of A will choose the shortest solution to $A\hat{x}=p$. when A is ind. cols the pseudoinv. is usual left inv. $L=(A^TA)^{-1}A^T$.)



(A is Vandermonde mt.)

(or fitting a plane, $C+Dx+Ey$, and so on.)

∘ $A = QR = QHR'$ (diag R all 1's. with heights h_i in H)
h_i tells the height col i above the plane of cols 1 to $i-1$.

$O(G\text{-}S) \approx n^3$

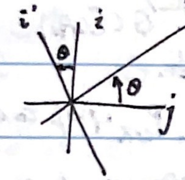
□ ways of QR decomp.: Householder decomp. and Givens decomp.

QR分解之豪斯霍尔德分解. 吉文斯分解

(A has ind. cols!)

· Givens process:

elementary rotation transform:



$$R_{ij} = \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array}\begin{array}{c} \begin{array}{ccc} 1 & i & j \end{array} \\ \begin{pmatrix} 1 & & & & \\ & \cos\theta & & \sin\theta & \\ & & \ddots & & \\ & -\sin\theta & & \cos\theta & \\ & & & & 1 \end{pmatrix} \end{array} \text{ (left mul)}$$

∘ QR's uniqueness: $Q_1 R_1 = Q_2 R_2$

$Q_1 = Q_2 R_2 R_1^{-1} = Q_2 U$ leads to

$Q_1^T Q_1 = I = U^T U$ must be diag!

since we make R positive diag, so $U = I$. $R_1 = R_2$...

choose $\cos\theta = \dfrac{x_i}{\sqrt{x_i^2 + x_j^2}}$ rotate x in i-j plane to $x_j' = 0$. $x_i' = \sqrt{x_i^2 + x_j^2}$.

notice that R_{ij} is orthogonal.

$$A_1 = R_{1n} R_{1n-1} \cdots R_{12} A = \begin{pmatrix} \|a_1\| & \cdots \\ 0 & \\ 0 & \\ \vdots & \\ 0 & \end{pmatrix}$$

then rotate a_2.
(since A ind. cols, a_2 can't all be zeros.)

$$A_2 = R_{2n} R_{2n-1} \cdots R_{23} A_1 = \begin{pmatrix} \|a_1\| & a_{21} & \\ 0 & \|a_2 - a_{21}\| & \cdots \\ 0 & 0 & \\ \vdots & \vdots & \\ 0 & 0 & \end{pmatrix}$$

so $A = \underbrace{(R_{n-1\,n} R_{n-2\,n} \cdots R_{12})^{-1}}_{(Q)} \underbrace{R}_{(A_{n-1})}$ $O(\text{Givens}) \approx \frac{4}{3} n^3$ it's better for sparse matrix.

ortho. mts' product is also ortho.

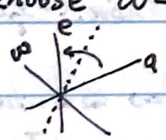
· Householder process:

elementary reflection transform: $H = I - 2\omega\omega^T$
ω is mirrornormal unitvec.
notice that H is orthogonal and sym.

or $H = \underset{(Q^T)}{Q^{-1}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} Q$ where $Q\omega = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $(|H| = -1, |R_{ij}| = 1)$

$(H = 2P - I$ ano. way)

choose $\omega = \dfrac{a - \|a\| e}{\|a - \|a\| e\|}$ to reflect a to e



$$H_1 A = \begin{pmatrix} \|a_1\| & \cdots \\ 0 & \\ \vdots & A_2 \\ 0 & \end{pmatrix}$$

choose $e_1 = (1, 0 \cdots 0)^T$

then preserve col 1 and row 1.
find $H_2' A_2$ to make $\begin{pmatrix} a_{22} & \cdots \\ 0 & \\ \vdots & A_3 \\ 0 & \end{pmatrix}$. $H_2 = \begin{pmatrix} 1 & 0 \\ 0 & H_2' \end{pmatrix}$

so $A = \underbrace{(H_{n-1} \cdots H_1)^{-1}}_{(Q)} \underbrace{R}_{(A_{n-1})}$ $O(\text{House}) \approx \frac{2}{3} n^3$

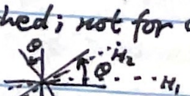
To conclude:

Gram-Smidt ✓ small mt.; generate q in every step
X not for sparse mt.; not stable; save all mt. thus cost large space.

Householder ✓ suitable for dense mt.; not calculate Q explicitly (if only need R)
X not generate q's until finished; not for sparse mt.

Givens ✓ suitable for sparse mt.
X not generate q's until finished; not for dense mt.

· a rotation mt. = two reflection mts: $R_{ij} = H_1 H_2$



$H_1 = I - 2\omega_1\omega_1^T$, $\omega_1 = (0 \cdots 0, \sin\frac{\theta}{4}, 0 \cdots 0, \cos\frac{\theta}{4}, 0 \cdots 0)^T$; $H_2 = I - 2\omega_2\omega_2^T$, $\omega_2 = (0 \cdots 0, \sin\frac{3}{4}\theta, 0 \cdots 0, \cos\frac{3}{4}\theta, 0 \cdots 0)^T$

4.4 orthonormal bases and Gram-Schmidt

not require sq.!

A matrix Q with orthonormal cols satisfies $Q^TQ=I$. It leaves lengths unchanged: $\|Qx\| = \|x\|$, and also preserves dot.p $(Qx)^T(Qy)=x^Ty$. Some Q's examples: Rotation. Permutation. Reflection $Q=I-2uu^T$ ($\vec{u}$ is the mirror normal with $\|\vec{u}\|^2=1$) $\vec{u}\to-\vec{u}$

The least squares solution to $Qx=b$ is $\hat{x}=Q^Tb$, the proj: $p=QQ^Tb=Pb$. thus dot.p of b with $q_1, q_2 \cdots q_n$, then use Q comb/proj.

(If Q is sq./orthogonal, it spans all space so $P=I$. $Q^T=Q^{-1}$) The one-dim projs are uncoupled. $b=\sum q_i(q_i^Tb)$ △

Gram-Schmidt process:

Substract from every new vec its projs in the directions already set.

$A=a$ (first) $\quad C=c-\frac{A^Tc}{A^TA}A-\frac{B^Tc}{B^TB}B \,/\|C\|$

$B=b-\frac{A^Tb}{A^TA}A \,/\|B\|$

At the end, divide the ortho. vecs by their length.

factorization: $A=QR$ (orthotri. decomp.) $\quad (q_1\ q_2 \cdots q_n)\begin{pmatrix} q_1^Ta_1 & q_1^Ta_2 & \cdots & q_1^Ta_n \\ & q_2^Ta_2 & & \\ & & & q_n^Ta_n \end{pmatrix}$ (A. B. C..)

$a_1 \cdots a_n \to q_1 \cdots q_n$ later q's are ortho. to earlier a's (and q's.) $R=Q^TA$ to recognize the upper tri.

Any $m\times n$ ind.cols A can be factored. It's useful for least squares: (R's diag is $\|A\|, \|B\| \dots$)

$A^TA=R^TR$. we solve $R\hat{x}=Q^Tb$ by back substitution ($\hat{x}=R^{-1}Q^Tb$). The real cost is G-S: $O(mn^2)$.

(modified G-S: more stable than original which substracts all projs at once. it substracts one proj at a time as in C^*, C: $\quad C^*=a_3-(q_1^Ta_3)q_1 \quad C=C^*-(q_2^TC^*)q_2 \quad q_3=\frac{C}{\|C\|}$.)

5. Determinants

5.1 the properties of determinants

ortho. $|Q|=$

$|\det A|=$ volume of box whose sides are A's rows/cols. (sign is the direction)

① the det of $I_{n\times n}$ is 1. ② the det changes sign when two row exchanged.

△③ the det is a linear func of each row separately. $\begin{vmatrix} a+ta' & b+tb' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + t\begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$

then there are some important corollaries: ④ if two rows of A are equal, then $\det A=0$.

⑤ substracting a multiple of one row from ano. row leaves A unchanged. $\begin{vmatrix} a & b \\ c-la & d-lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$

⑥ a mt. with a row of zeros has $\det=0$. ⑦ if A is triangular then $\det A = a_{11} \cdots a_{nn}$ ← diag.

⑧ singular A's $\det=0$, inv. A's $\det\neq0$. △⑨ the det AB is $\det A$ times $\det B$. esp. $\det A^{-1}=\frac{1}{\det A}$.

(proof: $\frac{|AB|}{|B|}$ has the same three properties that define $|A|$ (①②③). and the case $|B|=0$ is easy for AB is singular.)

△⑩ the transpose A^T has the same det as A.
(use $PA=LU$, $A^TP^T=U^TL^T$ and ⑨ proof)

△ try to image those in geometry!

o' if A is singular, $AC^T=0$, then each col of C^T is in the nullspace of A.
($A_{n\times n}$ of rank $n-2$ or smaller, all cofactors are zero and we only find $\vec{x}=0$.)

o $AC^T=(\det A)I \longrightarrow (\det A)^{n-1}=\det C$. so if you know all cofactors of A (with position) you could get A.

□ Cauchy-Binet formula : 柯西-比内公式

It gives the det of a sq. mt. AB (esp. A^TA) when the factors A, B are rectangular.

$A_{s\times n}\cdot B_{n\times s}$: $s>n$ leads to $\det AB=0$;
$s\leq n$ then $\det AB=$ sum of ($s\times s$ dets in A)($s\times s$ dets in B).

$$\begin{pmatrix}\underline{a}&\underline{b}&c\\\underline{d}&\underline{e}&f\end{pmatrix}\begin{pmatrix}\underline{g}&\underline{j}\\\underline{h}&\underline{k}\\i&l\end{pmatrix}\cdot\begin{pmatrix}\underline{a}&b&\underline{c}\\\underline{d}&e&\underline{f}\end{pmatrix}\begin{pmatrix}\underline{g}&\underline{j}\\h&k\\\underline{i}&\underline{l}\end{pmatrix}\cdot\begin{pmatrix}a&\underline{b}&\underline{c}\\d&\underline{e}&\underline{f}\end{pmatrix}\begin{pmatrix}g&j\\\underline{h}&\underline{k}\\\underline{i}&\underline{l}\end{pmatrix}$$

(s's cols of A correspond with s's rows of B)
(calculate the 2×2 dets underlined)

o Jacobi formula: The cofactor formula can be generalized. (or 'Laplace expansion')
For $n\times n$ mt., we can choose $k\times k$ det times $(n-k)\times(n-k)$ det. Then add them with signs.
The sign depends on the permutation of $a_1,\cdots a_k, b_1,\cdots b_{n-k}$ (a and b is the column you choose)

$$\begin{vmatrix}2&-1&0&0&0\\-1&2&-1&0&0\\0&-1&2&-1&0\\0&0&-1&2&-1\\0&0&0&-1&2\end{vmatrix}=\begin{vmatrix}2&-1\\-1&2\end{vmatrix}\cdot\begin{vmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{vmatrix}\cdot P_{(1,2,3,4,5)}+\begin{vmatrix}2&0\\-1&-1\end{vmatrix}\cdot\begin{vmatrix}-1&-1&0\\0&2&-1\\0&-1&2\end{vmatrix}\cdot P_{(1,3,2,4,5)}=6$$

o $\boxed{\text{rank-one mt.}}$ $A=uv^T$ has an eigenvector u. its $\lambda=v^Tu$. Other eigenvalues are zeros. (eigenvecs on the plane $v^{\perp}$)
$\sigma=\|u\|\|v\|$. Check $\sum\lambda=v^Tu$ $(=u_1v_1+\cdots+u_nv_n)$

更多关于特征值: permutation mt. P have $P^k=I \Rightarrow \lambda^k=1$ Then we find P's eigens. (maybe complex)
$\sum_i\lambda_i^k=\mathrm{tr}A^k$. (every perm. mt. leaves $\vec{x}=(1,1,\cdots,1)^T$ unchanged so $\lambda=1$)
rotation mt. can do the same. $\lambda^{\frac{2\pi}{\theta}}=1$ we find $e^{\pm i\theta}$ (for higher dim, $\lambda=1,1,\cdots,e^{\pm i\theta_{ij}}$)

o' When a mt. is diagl, the number of nonzero eigenvalues is its rank. $\sim\begin{pmatrix}\Lambda_r&\\&0\end{pmatrix}$

o $A+B$. AB might not have $\lambda=\lambda_a+\lambda_b$. $\lambda_a\lambda_b$! It only happens when A, B have same eigenvecs.
△ esp. $B=I$, $A+cI$ can have $\lambda=\lambda_a+c$ for A's eigenvecs must in I's eigen.
(when λ has multiplicity, the whole plane spanned by its eigenvess are eigen comb; for I it's whole space)

o no interconnection between invertibility ······ diagonalizability !
$\lambda\neq0$?　　　　$X\exists$?

5.2 permutations and cofactors

The determinant of any $n \times n$ matrix can be found in three ways: (more: Big Partition...) (大前)

1. pivot formula: $PA = LU \longrightarrow \det A = \pm \det U = \pm$ (product of pivots)

(we don't need row exchanges when all the upper left submts. have $\det A_k \neq 0$. The kth pivot is $d_k = \frac{\det A_k}{\det A_{k-1}}$.)

2. big formula: $\det A$ = sum over all $n!$ col permutations $P = (\alpha, \beta, \cdots, \omega)$

$= \sum \det P \cdot a_{1\alpha} a_{2\beta} \cdots a_{n\omega}$ choose one entry from every row and col.

$$\begin{vmatrix} a_{11} & \cdots & \\ \vdots & \ddots & \\ & & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33}\begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31}\begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{21}a_{32}\begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix} + a_{11}a_{23}a_{32}\begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33}\begin{vmatrix} & 1 & \\ 1 & & \\ & & 1 \end{vmatrix} + a_{13}a_{22}a_{31}\begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}$$

$n!$ ways to order, half odd and half even.

3. cofactor formula: $\det A = a_{i1}C_{i1} + \cdots + a_{in}C_{in}$. each cofactor C_{ij} includes its correct sign:

since a_{ij} removes row i and col j, we can linearly expand det by r_i/c_j. $C_{ij} = (-1)^{i+j} \underset{\det}{M_{ij}}$

$$\begin{vmatrix} a_{11} & \cdots & \\ \vdots & \ddots & \\ & & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & a_{23} \\ & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & a_{23} \\ a_{31} & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{33} \\ a_{21} & a_{22} & \\ a_{31} & a_{32} & \end{vmatrix}$$

to permutate a_{ij} to a_{11} need $(-1)^{i+j-2}$ reverse, then M_{ij}'s det permutates as itself.

$= a_{11}(a_{22}a_{13} - a_{32}a_{23}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{33}(a_{21}a_{32} - a_{22}a_{31})$ ($\det A^T = \det A$, so did col j.)

5.3 Cramer's rule, inverse, and volumes

Cramer's rule solves $Ax = b$. we construct $A\begin{pmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} b & a_2 & a_3 \end{pmatrix} = B_1$. $x_1 = \frac{\det B_1}{\det A}$.

then $x_n = \frac{\det B_n}{\det A}$, where mt. B_n has the nth col of replaced by b.

Then we find the cols of A^{-1} by solving $AA^{-1} = I$. $B_n = \begin{pmatrix} \cdots & \begin{smallmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{smallmatrix} & \cdots \end{pmatrix}$. so

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $(A^{-1})_{ij} = \frac{C_{ji}}{\det A}$ and $A^{-1} = \frac{C^T}{\det A}$

A direct proof of $AC^T = (\det A)I$: (C_{ij} go into the cofactor matrix C)

The cofactor formula yields $\det A$ on the diagonal. multiplying cofactors from diff rows is the cofactor formula for a new matrix when the second row is copied into first row, so yields zero.

The triangle with corners $(x_1, y_1)(x_2, y_2)(x_3, y_3)$ has area $= \frac{\det}{2} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$,

We can proof that the parallelogram's area when $(x_3, y_3) = (0, 0)$, area $= \frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$.

or the box's volume obeys the determinant properties ①②③.

cross p.: $\vec{u} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$ is perpendicular to $\vec{u}$ and $\vec{v}$.

triple p.: $(\vec{u} \times \vec{v}) \cdot \vec{w} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$ = determinant = volume.

$\circ$ Heisenberg's uncertainty principle: (inf. mt.) position mt. P, momentum mt. Q. (sym.) (antisym.)

$QP - PQ = I$ (To have $Px=0$ and $Qx=0$ would require $x=0$)

$x^T x = x^T QPx - x^T PQx \leq 2\|Px\|\|Qx\|$ Explain that last step by using Schwarz ineq. $\|u^T v\| \leq \|u\|\|v\|$.

so $\frac{\|Px\|\|Qx\|}{\|x\|^2} \geq \frac{1}{2}$. (Impossible to get position and momentum error all very small.)

$\circ$ 1. A and B share the same n ind. eigenvecs if and only if $AB = BA$. (commutable) (proof见后)

2. If A is $m\times n$ and B is $n\times m$, then AB and BA have some nonzero eigenvalues. (and $|n-m|$ zeros.)

(proof: $\begin{pmatrix} I & -A \\ & I \end{pmatrix}\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}\begin{pmatrix} I & A \\ & I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$ $\begin{pmatrix} m\times m & m\times n \\ n\times m & n\times n \end{pmatrix}$

so $E = \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \sim F = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$, they have same $m+n$ values: E has m eivalues of AB + n zeros, F has n eivalues of BA + m zeros)

3. Suppose A_1, A_2 $n\times n$ inv., then A_1A_2 is similar to A_2A_1, same eigenvalues. (B choose A_2)

$\circ$ Gershgorin circles: Every λ is in the circle around one or more diagonal entries a_{ii}: $|a_{ii} - \lambda| \leq R_i = \sum_{j\neq i} |a_{ij}|$

(reason: $A - \lambda I$ is singular, it cannot be diag dominant.)

$\circ$ eigenvalues of A equals eigenvalues of A^T. (since $\det(A-\lambda I) = \det(A^T - \lambda I)$)

The eigenvectors of A and A^T from different λ's are perpendicular. (esp. sym. mt's eigenvecs are ortho.)

since $A\vec{x}_1 \cdot \vec{x}_2 = \vec{x}_1 \cdot A^T \vec{x}_2 \Rightarrow \lambda_1 \vec{x}_1 \cdot \vec{x}_2 = \lambda_2 \vec{x}_1 \cdot \vec{x}_2$, $\lambda_1 \neq \lambda_2$ leads to $\vec{x}_1 \perp \vec{x}_2$.

(or $A = X\Lambda X^{-1}$, $A^T = X^{-T}\Lambda X^T$, the eigenvecs of A^T are X^{-T}, $X^{-1}X = I$ leads to $x_i \cdot x_j$ $(i\neq j) = 0$.) (见前)

→ left eigenvectors of A ($y^T A = \lambda y^T$). rank-one's sum: $A = \lambda_1 x_1 y_1^T + \cdots + \lambda_n x_n y_n^T$. $(n = r(A))$

$\circ$ 矩阵收集仓(2):

· Hadamard matrix $H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ $H_{ij} = \pm 1$ and $H^T H = nI$, every cols or rows in H are orthogonal. (so n cannot be odd) $H = H^T = H^{-1}$

(Sylvester theorem) $H \checkmark \longrightarrow \begin{pmatrix} H & H \\ H & -H \end{pmatrix} \checkmark$

(Hadamard conjuncture: $n|4$, $H_n \checkmark$)

Hadamard Theorem: $A = (a_{ij})_{n\times n}$ with $-1 \leq a_{ij} \leq 1$, then $|\det A| \leq n^{\frac{n}{2}}$. equality achieves when A is H ($a_{ij} = \pm 1$ and rows are ortho.)

· Hessenberg matrix $H_3 = \begin{pmatrix} 2 & 1 & \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$
(a tri. with one extra diag.)

$|H_n| = |H_{n-1}| + |H_{n-2}|$ (Fibonacci F_{n+2})

· Markov matrix largest eigenvalue $\lambda_m = 1$. this eigenvec is the steady state. (A^∞)

Mar. mt. each column adds up to 1. ('possiblity transfer')

(proof: since $\lambda_A = \lambda_{A^T}$, thus λ can be 1; (见后)

and Gershogrin circle tells $\lambda - a_{ii} \leq \sum_{j\neq i} a_{ji}$, $\lambda \leq 1$)

6. Eigenvalues and Eigenvectors

6.1 introduction to eigenvalues

An eigen vector $\vec{x}$ lies along the same line as Ax : $Ax=\lambda x$. The eigenvalue is λ.

Then $(A-\lambda I)x=0$ and $A-\lambda I$ is singular and $\det(A-\lambda I)=0$. So 1. Compute the det of $A-\lambda I$.
2. Find the n roots of this polynomial of degree n. 3. For each λ, solve $(A-\lambda I)x=0$ to find an eigenvector x.

The eigenvalues of A^2 and A^{-1} are λ^2 and λ^{-1}, with the same eigenves.

(Vieta's formula or geometry) The sum of the λ's equals the sum down the main diag of A or called the trace. $\sum_{i=1}^{n}\lambda_i = a_{11}+a_{22}+\cdots+a_{nn}=\mathrm{tr}(A)$. The products of the λ's equals the det of A. $\prod_{i=1}^{n}\lambda_i=\det A$.

Special mts. (properties) lead to special eigenvalues and eigenvectors:

Proj. P have $\lambda=1$ and 0. Reflection R have $\lambda=1$ and -1. Rotation have (Only) $\lambda=e^{i\theta}$ and $e^{-i\theta}$ (complex!).

Singular mt. has $\lambda=0$. Triangular mt. has λ's on its diagonal. Identity mt. has all n $\lambda=1$.

sym. mt. can be compared to a real number, antisym. mt. can be compared to an imaginary number.

Orthogonal mt. corresponds to a complex number with $|\lambda|=1$ (conjugated). For these three special mts, their eigenvecs are perpendicular/orthonormal (Normal matrix) $N^TN=NN^T$

6.2 diagonalizing a matrix

$Ax_i=\lambda x_i$ are the cols of $AX=X\Lambda$. The eigenvalue matrix Λ is diagonal and correspond with the order of x_i in X.

n ind. eigenvecs diagonalize A : $A=X\Lambda X^{-1}$. $\Lambda=X^{-1}AX$ (also diagonalize all powers $A^k=X\Lambda^kX^{-1}$)

No equal eigenvalues : ind. x from diff. λ. so eigenvecs are ind. X is inv. and A is diagonalizable.

Equal eigenvalues : A might have too few ind. eigenvecs. X^{-1} fails.

we have $GM \leq AM$ (geometric multiplicity $\leq$ algebraic multiplicity)

when $GM<AM$, A is not diagl. $\dim(N(A-\lambda I))$ roots of $\det(A-\lambda I)=0$

(△ because maybe two cases during $A-\lambda I \to U_\lambda$:

1. pivots here (black is zero when λ_i)
2. $\begin{vmatrix} 1 & \\ & (\lambda-\lambda_i)^2 \end{vmatrix}$ compressed in one pivot.)

proof: multiply A several times - multiply λ_j to $c_1x_1+\cdots+c_jx_j=0$ $\Rightarrow \prod_j(\lambda_1-\lambda_j)c_1x_1=0$ so $c_1=0$. all other $c_j=0$. ind.

Similar matrices:

all $A=BCB^{-1}$. $A\sim C$. (only I itself similar I). they all share the eigenvalues of C.

Classical examples:

① Markov $\begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} \xrightarrow{\infty} \begin{pmatrix} .6 & .6 \\ .4 & .4 \end{pmatrix}$ ② I. $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ (Jordan form) ($\lambda=1$) ③ $A^2=A$. $\frac{xy^T}{y^Tx}$. ($\lambda=1,0$) ④ Fibnums. $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ $\vec{u}_{k+1}=A\vec{u}_k$. ($u_k=A^ku_0$) (difference eq.)

(ano2. way: $Sx=\lambda x$. $Sy=\beta y$. its colspace $C \perp$ nullspace N!
$S-\beta I$ is sym. so $x \perp y$ $(\lambda_1 \neq \lambda_2)$;

o' we can call R is S (sym.)'s sq.root if $S=R^TR$. so from diaglize we find R can be $=\Lambda^{\frac{1}{2}}Q$ (not unique 见右)

□ 对称阵性质证明

a proof of sym. mts has ① ortho. eigenvecs in algebraic way:

$S=X\Lambda X^{-1}$. $S^T=X^{-T}\Lambda X^T \Rightarrow X^TX\Lambda=\Lambda X^T\Lambda$ or ΛX^TX is sym.

we know if A, B sym. (AB) sym. needs $AB=BA$. so Λ and X^TX commutable and have same eigenvecs

since Λ is diag, X^TX must be like I or Λ, it's Q. (ano. way: $\lambda_1 x^Ty=(Sx)^Ty=x^TSy=\lambda_2 x^Ty$ $(x:\lambda_1, y:\lambda_2)$ so $x^Ty=0$ $(\lambda_1 \neq \lambda_2)$ (与上))

② above has a premise that sym. mts can be diagonalized:

$\boxed{\text{Schur's Theorem}}$: Every sq. A factors into QTQ^{-1} where T is upper tri. and $\bar{Q}^T=Q^{-1}$.
If A has real eivalues then Q and T can be chosen real: $Q^TQ=I$.
When $S^T=S$, T must be diagonal: $S=Q\Lambda Q^T$.

(informal: if S has repeated λ's, slightly change S in diag$(c, 2c, \cdots, nc)$. with $c\to 0$, orthonormal eivecs remained.)

③ real eigenvalues: $Sx=\lambda x \longrightarrow S\bar{x}=\bar{\lambda}\bar{x}$, $\bar{x}^TS=\bar{x}^T\bar{\lambda}$.

$x^TSx=x^T\lambda x$ $\qquad$ $\bar{x}^TSx=\bar{x}^T\bar{\lambda}x$

(all mts have $\prod d=\prod\lambda$) $\qquad$ $\bar{x}^Tx$ is the squared length >0. so $\lambda=\bar{\lambda}$, it's real

Δ $|A_{nti}|$ = positive or zero.
(for antisym., $\lambda=-\bar{\lambda}$, it's pure imag.)

④ same signs with pivots:

tri. fac. $S=LDL^T=\begin{pmatrix}1&\\ \bigcirc&1\end{pmatrix}\begin{pmatrix}d&\\&\ddots\end{pmatrix}\begin{pmatrix}1&\bigcirc\\&1\end{pmatrix}$

when L moves to I $(\bigcirc\to 0)$, IDI^T has eigenvalues = pivots d's. (eivalue $\to$ pivot)

But to change sign, a real eivalue would cross zero $\longrightarrow$ matrix singular at a moment ✗

but we always have pivots ds so no sign change in moving

o 幂零阵 nilpotent matrix: the following propositions are equivalent:

① A is nilpotent $(\exists m, s.t. A^m=0)$ ② $A^n=0 \leftarrow A^{r(A)+1}=0$ ③ all λ's of $A=0$

④ (from ③) $\det(A+kI)=k^n$, $\det(kA+I)=1$ $\forall k$ ⑤ $\forall k$ $tr(A^k)=0$

(幂一阵 $A^2=I$: nothing new. $\lambda=\pm1$. esp. sym.+ortho.)

(o $\boxed{\text{Cayley-Hamilton's Theorem}}$: eigenpolynomial: $\det(\lambda I-A)=f(\lambda)$
for nilp. A it's λ^n,
then we have $f(A)=0$.)

o 更多关于矩阵指数: $\cdot$ e^{At_1} times e^{At_2} must be $e^{A(t_1+t_2)}$, but e^Ae^B, e^Be^A, e^{A+B} can be all diff! $(e^A \cdot e^A=e^{2A}$ ✓)

$\cdot$ when A is tri., X, X^{-1}, e^{At} are also tri. (since $A-\lambda I$: ..., X is tri.)

$\cdot$ repeated λ: diaglize is impossible so compute e^{At} directly:

$A=\begin{pmatrix}0&1\\-1&2\end{pmatrix}$. $\lambda=1,1$. $\quad e^{At}=e^{It}e^{(A-I)t}=e^t(I+(A-I)t) \longleftarrow (A-I)^2$ is zero. (见上)

some mt.'s series stops early. so we find te^t here.

$\cdot$ antisym. $A=\begin{pmatrix}0&1\\-1&0\end{pmatrix}$, then $e^{At}=\begin{pmatrix}\cos t&\sin t\\-\sin t&\cos t\end{pmatrix}$, its eigenvalues are e^{it}, e^{-it}.
$(A^4=I)$

$\cdot$ extension: Cosine form: solve $\frac{d^2u}{dt^2}=-A^2u$ $\qquad$ $\cos At=I-\frac{1}{2!}A^2+\frac{1}{4!}A^4-\ldots$
$(\cos At)''=-A^2\cos At$ which corresponds.

Short form $u_{(t)}=\cos(At)u_{(0)}$ / specific form $u_{(0)}=c_1x_1+c_2x_2$,
$u_{(t)}=c_1\cos\lambda_1 t\,x_1+c_2\cos\lambda_2 t\,x_2$

6.3 systems of differential equations

$\frac{du}{dt}=Au$. if $Ax=\lambda x$ then $u(t)=e^{\lambda t}x$ will solve it. each λ and x gives a ind. solution.

start from $u(0)$. it can be written as comb $c_1x_1+\cdots+c_nx_n$ or $Xc=u(0)$, $c=X^{-1}u(0)$.

then each eigenvec is muled by $e^{\lambda_i t}$, the solution is $u(t)=c_1e^{\lambda_1 t}x_1+\cdots+c_ne^{\lambda_n t}x_n$.

(at past $\frac{d}{dt}\binom{y}{z}=\begin{pmatrix}0&1\\1&0\end{pmatrix}\binom{y}{z}$ you may use $\frac{d}{dt}(y\pm z)=\pm(y\pm z)$ instead of $u_1(t)=e^{t}\binom{1}{1}$. $u_2(t)=e^{-t}\binom{1}{-1}$.)

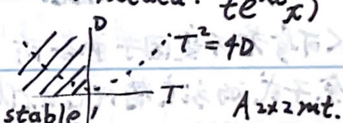
Matrix exponential form: $e^{At}=I+At+\frac{1}{2!}(At)^2+\cdots$ 1. $(e^{At})'=Ae^{At}$.

2. e^{At}'s eigenvalues are $e^{\lambda t}$ (sub $A=X\Lambda X^{-1}$ into exponential, $e^{At}=Xe^{\Lambda t}X^{-1}$) 3. e^{At} has inverse e^{-At}. When A is antisym, ($A^T=-A$) e^{At} is orthogonal.

so $u(t)=\sum_i c_ie^{\lambda_i t}x_i=Xe^{\Lambda t}X^{-1}u(0)=e^{At}u(0)$.

(we call A is stable when $e^{At}\to 0$ thus all eigenvalues of A have real part <0. imaginary part ω gives oscillation.)

(Not included: if two λ's are equal, with only one eigenvec, ano. solution is needed: $te^{\lambda t}x$)



$A_{2\times 2}$ mt. is stable when tr <0, det >0.

6.4 symmetric matrices

Every sym. matrix S can be diagonalized, and has n real eigenvalues λ_i and n orthonormal eigenvectors $q_1,\cdots,q_n$.

(Principle Axis / Spectral Theorem) $S=Q\Lambda Q^T=\sum_i\lambda_iq_iq_i^T$

The number of positive eivalues of S = The number of positive pivots of S (proof 见后) (signs matched)

(Antisym. matrix $A^T=-A$ have imag. λ's and ortho. complex q's.)

For real matrices (nonsym.), complex λ's and x's come in conjugate pairs: $Ax=\lambda x$, $A\bar{x}=\bar{\lambda}\bar{x}$

(every ortho. Q have $|\lambda|=1$ since then)

6.5 positive definite matrices

positive definite (sym.) S: 1. all eivalues >0 ↔ 2. all pivots >0 ↔ 3. all upper left dets >0.

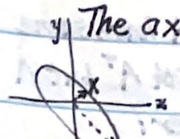
4. ('Energy-based def/test') $x^TSx>0$ for all vecs $x\neq 0$. (so obviously S pos. T pos. → $S+T$ pos.)

5. A col ind. (to rule out zero), $S=A^TA$. ($A=\text{chol}(S)=L\sqrt{D}$ / $A=Q\sqrt{\Lambda}Q^T$ etc.) ($x^TSx=\|Ax\|^2>0$)

positive semidefinite: allows $\lambda=0$. pivot $=0$. det $=0$. $x^TSx=0$ ($x\in N(S)=N(A)$). A has dep. cols.

geometry: The graph of $x^TSx=1$ is an ellipse: $(x\ y)Q\Lambda Q^T\binom{x}{y}=(X\ Y)\Lambda\binom{X}{Y}=\lambda_1X^2+\lambda_2Y^2=1$. (standard form)

The axes point along eivecs of S, half-lengths are $\frac{1}{\sqrt{\lambda}}$. ($X=q_1^Tx$, $Y=q_2^Tx$)

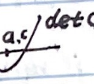


(if one λ is neg. ellipse changes to hyperbola.)

o' S pos. T pos. $\rightarrow ST$ might not sym, but pos. eivalues: $STx = \lambda x$. $\lambda = \frac{(Tx)^T S T x}{\lambda x^T T x} > 0$

(Normal includes sym. artisym. ortho.)

- N has n orthonormal eigenvectors ($N = Q\Lambda Q^H$) if and only if N is normal. ($N^H N = N N^H$)
  (proof: ① $N = Q\Lambda Q^H$ subin. ② Schur T: $N = QTQ^H$ and T is Λ.)

o' sym. S: ① no diag entry can be larger than λ_{max} (use $S = \lambda_i q_i q_i^T$)
② diag entries fall in between the λ's. ($S_{2\times 2}$:  a.c det(A-λI))

□ Congruence 合同矩阵

$x^T A x$

$x^T \underbrace{C^T A C}_{B} x$

- similarity — same transform on diff. basis.
  congruence — same bilinear / (0,2) tensor / metric on diff. basis.
  $T = T_{ij} e_i e_j^T$, if $e_i = R_{ij} e_j'$ then $T'_{ij} = T_{kl} R_{ki} R_{jl}$, mt. form: $T' = R^T T R$.
  same quadratic surface on diff. basis: ⊕ → ⊕⊕
- In real, congruence ⟷ equal inertia index. keep the pos./neg. inertia index.
  In complex, congruence ⟷ equal rank.

□ Jordan canonical form 约当标准形 (观后)

- sq. $A \sim J = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots J_{gm} \end{pmatrix}$, $J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}$ <可参考 J_i^k: Jordan pot.>

<可参考不变因子和各阶余子式的方法替代 GM.AM> gm = number of Jordan blocks $\leq am$ = scale/cols of Jordan blocks. call λ_i adds)

- $A = XJX^{-1}$. $AX = XJ$ solve x_i one by one. $\begin{array}{c} x_1 x_2 x_3 \cdots \\ \begin{pmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \ddots \end{pmatrix} \end{array}$ (each Jor. block) $(A-\lambda I)x_1 = 0$, $(A-\lambda I)x_2 = x_1$, ⋮

- a proof of nilpot. A has $A^{r(A)+1} = 0$:
  since $\lambda = 0$, $A \sim J = \begin{pmatrix} J(0,k_1) & & \\ & \ddots & \\ & & J(0,k_s) \end{pmatrix}$
  $r(A) = r(J) \geq r(J(0,k_i)) = k_i - 1$ so $k_i \leq r(A)+1$,
  $J(0,k_i)^{k_i} = 0 \longrightarrow A^{r(A)+1} = 0$

- ' a more classical way of Jordan: $(A-\lambda I)x_j = x_{j-1}$
  ① Primary Decomp. 准素分解 $f(\lambda) = (\lambda-\lambda_1)^{n_1} \cdots (\lambda-\lambda_m)^{n_m}$
  $V_i = \ker(A-\lambda_i I)^{n_i}$ then $V = \bigoplus_{i=1}^{t} V_i$ (根子空间)
  ② Cyclic Decomp. 循环分解 $V_i = \bigoplus_{j=1}^{t} \mathrm{span}\{N^l \omega_j : 0 \leq l \leq l_j\}$. ($t = \dim(\ker N)$. $N = A - \lambda I$) (强循环子空间)
  <可参考Young图>

- rank-one decomp. / Spectral decomp.: a synthesis concept of decomp. $A = \sum_i \lambda_i G_i = \sum_i \lambda_i u_i v_i^T$
  秩-1分解 谱分解 ($G_i G_j = \delta_{ij} G_i$ if S)
  — G_i is idempot. λ_i is the eigenvalue. ($G_i = q_i q_i^T$ or $u_i v_i^T$)
  — I generalize it: all n rank-one mts $u_i v_i^T$ adds up to A. Obviously it's not unique.

△ a rank r mt. can be expanded by r rank-one mts and at least r. ($r = r - k + k$, general.)
proof: A can be elementary transformed to $P\begin{pmatrix} I_r & \\ & 0 \end{pmatrix}Q$. so $A = \sum_{i=1}^{r} p_i \cdot 1 \cdot q_i^T$.
and since $r(A+B) \leq r(A) + r(B)$, at least should r rank-one mts.

□ Table of Eigens [Matrix/λ/x] P362 特殊特征值/向量表

△ Q has all singular values $= 1$ as well as I.

o a proof of commutable mts have common eigenvecs:
A has a λ and its eigensubspace V_λ. $v \in V_\lambda$.
$ABv = BAv = \lambda Bv$
so V_λ also a invariant subspace of B. There must be a μ that s.t. $Bu = \mu u$ with $Au = \lambda u$.
Then u is the common eivec.

△ esp. normal mt A have same X and Λ with A^T but Λ has diff. order! ('a paradox')

7. The Singular Value Decomposition (SVD)

7.1 image processing by linear algebra

The SVD seperates any matrix A into rank-one pieces $A=\sum_i \underbrace{\sigma_i}_{\text{singular value}} \underbrace{u_i}_{\text{left sin. vecs.}} \underbrace{v_i^T}_{\text{right sin. vecs.}}$. in order of importance $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$.

The choices of u's and v's are orthon. $u_i^T u_j = \delta_{ij}$, $v_i^T v_j = \delta_{ij}$. (the dogma of image process.)

◺ has infinite rank, 吕 回 has low finite rank. Discard small σ's.

7.2 bases and matrices in the SVD $A = U\Sigma V^T = u_1\sigma_1 v_1^T + \cdots + u_r \sigma_r v_r^T$.

$\Uparrow$

one way to deduce: $A^TAv_i = \sigma_i^2 v_i \longrightarrow Av_i = \sigma_i u_i \longrightarrow A^T u_i = \sigma_i v_i \longrightarrow AA^T u_i = \sigma_i^2 u_i$

1. both A^TA, AA^T sym. pos. v_i chosen to be eigenvecs of A^TA with $\sigma_i > 0$. (r nonzero) so $v_i \perp v_j$. $v_1 \cdots v_r \in R(A)$ and left $v_{r+1} \cdots v_n \in N(A)$.

2. choose $u_i \in C(A)$. $(u_1 \cdots u_r)$ and we can proof: $u_i^T u_j = \frac{v_i^T A^T A v_j}{\sigma_i \sigma_j} = \frac{\sigma_j}{\sigma_i} v_i^T v_j = 0$. ($u_i^T u_i = v_i^T v_i = 1$ also.)

3. dual to 2.

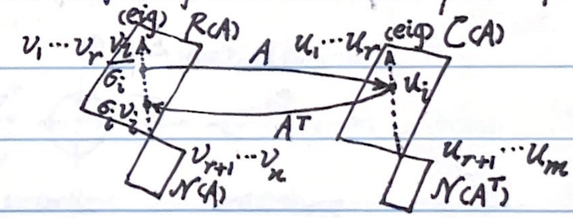
4. u_i are eigenvecs of AA^T. so again $u_i \perp u_j$. $u_i \in R(A^T) = C(A)$ and left $u_{r+1} \cdots u_m \in N(A^T)$.

To conclude: $\sigma_1^2 \cdots \sigma_r^2$ are the nonzero eigenvalues of A^TA and AA^T. (不同)

the orthon. cols of U and V are the eigenvectors of A^TA and AA^T.

Also those cols hold orthon. bases for the four fundamental subspaces of A.

then diagonalize A: $AV = U\Sigma$.



(reduced SVD: $\underset{m\times r}{(u_1 \cdots u_r)} \underset{r\times r}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}} \underset{r\times n}{\begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix}}$;

full SVD: $\underset{m\times m}{(u_1 \cdots u_m)} \underset{m\times n}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}} \underset{n\times n}{\begin{pmatrix} v_1^T \\ \vdots \\ v_n^T \end{pmatrix}}$.)

△ esp. When A is pos. (semipos.) sym. mt and only when can $A = X\Lambda X^{-1} = Q\Lambda Q^T = U\Sigma V^T$ ($X=U=V$; $\lambda \geq 0$, $\Lambda = \Sigma$)

Singular versus Eigen:

$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \frac{1}{60000} & 0 & 0 & 0 \end{pmatrix}$ △ $\lambda = 0.0.0.0$ jump to $\frac{\pm 1}{10} \cdot \frac{\pm i}{10}$. This shows serious instablity of eigen values when AA^T is far from A^TA.

The singularvalues are always stable. σ_4 emerges as $\frac{1}{60000}$ from 0. $\sigma = 3.2.1.0$

(At the other extreme, $AA^T = A^TA$ 'normal' means A has ortho. eigenvecs and stable eigenvalues.)

△ since $\sigma = \sqrt{\lambda}$, we have (for sq.) $\prod_i \lambda_{i(A)} = \prod_i \sigma_{i(A)} = \det A$ $(= \prod d_{i(A)} = \prod h_{i(A)})$

one way one-at-a-time instead of all-at-once: m. $S = A^TA (= V\Sigma^T\Sigma V^T)$

$S = \lambda_1 v_1 v_1^T + \cdots + \lambda_r v_r v_r^T$; $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$.

$$\begin{cases} \lambda_1 = \max \frac{x^TSx}{x^Tx} & \text{when } x = q_1 = v_1 \\ \sigma_1 = \max \frac{\|Ax\|}{\|x\|} & \text{when } x = v_1 \quad \left(\overset{1.}{\max} \frac{x^TA^TAx}{x^Tx}\right) \end{cases}$$

$$\begin{cases} \lambda_2 = \max \frac{x^TSx}{x^Tx} \text{ with } q_1^Tx = 0 & \text{when } x = q_2 = v_2 \\ \sigma_2 = \max \frac{\|Ax\|}{\|x\|} \text{ with } v_1^Tx = 0 & \text{when } x = v_2 \end{cases}$$

$\vdots$

n

(Rayleigh quotient $r(x) = \frac{x^TSx}{x^Tx}$. 1. you can see Q^Tx as x. the $x^TQ^T\Lambda Qx = \sum_i \lambda_i \cos\theta_i^2$. chose λ_{max}.

2. $\frac{\partial r}{\partial x} = 0$. For next step, Q^Tx has first component $= 0$ or $Q_1^TSQ_1 = \begin{pmatrix} \lambda_1 & \\ & S_{n-1} \end{pmatrix}$.)

(2. $\|Ax\| = \|U\Sigma V^Tx\| = \|\Sigma V^Tx\| \leq \sigma_1 \|V^Tx\| = \sigma_1 \|x\|$.)

o' 补充说明：when $q_1 \cdots q_n$ ortho., we have $q_1q_1^T + \cdots + q_nq_n^T = I$. vice versa.(?)

- $Q_1AQ_2^T$ has the same σ's as A and QSQ^T has the same λ's as S.

Computing of singular: $Q^TSQ \rightarrow$ triang. $Q_1AQ_2^T \rightarrow$ bidiag.

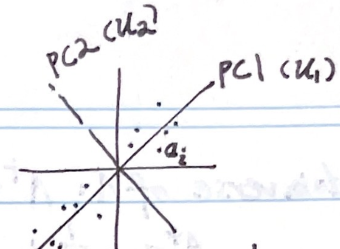
- $\sigma_{max} \geq |\lambda|_{max}$; $\sigma_{min} \leq |\lambda|_{min}$. (以前) $\|Ax\| \leq \sigma_1\|x\|$; $\|Ax\| \geq \sigma_{min}\|x\|$ or inverse to A^{-1}: $\frac{1}{\sigma_{min}} \geq \left|\frac{1}{\lambda_{min}}\right|$
$= |\lambda|\|x\|$

□ Norm of matrices: 矩阵范数

- 向量 p 范数 $\|x\|_p = \left(\sum_j |x_j|^p\right)^{\frac{1}{p}}$ $(1 \leq p)$ · 常见 1-norm. 2-norm (Euclidean). ∞-norm ($\max_j |x_j|$)
- 范数满足 ① $\|x\| \geq 0$ ② $\|\alpha x\| = |\alpha|\|x\|$ ③ $\|x+y\| \leq \|x\| + \|y\|$.
- inducing norm/operator norm: $\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$. 常见 $\|A\|_1 = \max_j \sum_i |a_{ij}|$ (max col sum)
- Frobenius norm (F-norm): $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{tr(A^TA)}$ $= \sqrt{\sum_i \sigma_i^2}$
 $\|A\|_2 = \sqrt{\lambda_{max}(A^TA)} = \sigma_1$ (Euc./spectrum)
 $\|A\|_\infty = \max_i \sum_j |a_{ij}| = \|A^T\|_1$
- theorem 1: $\rho(A) \leq \|A\|$ (ρ is spetrum radius, $\rho(A) = \max_i |\lambda_i|$)
 ($\forall \varepsilon > 0$, $\exists \|\cdot\|_p$ s.t. $\|A\|_p \leq \rho(A) + \varepsilon$)
 eq. when $\|\cdot\|$ is 2-norm and A is sym.
- theorem 2 (Gelfand formula): $\rho(A) = \lim_{k \to \infty} \|A^k\|^{\frac{1}{k}}$

o $\boxed{\text{rank-one mt.}}$ $A = \sigma uv^T$ (unitvec u and v), $A^+ = \frac{vu^T}{\sigma}$. $AA^+ = uu^T$, $A^+A = vv^T$.
 (C) (R)

o' SVD of A is equivalent to the diagonlization of sym. $M = \begin{pmatrix} 0 & A^T \\ A & 0 \end{pmatrix}$, with eivec $= \begin{pmatrix} v \\ u \end{pmatrix}$.



7.3 principal component analysis (PCA by the SVD)

Data comes in a matrix: n samples and m measurements per sample.

Substracting the mean of each row to center A. Then SVD finds the ($\vec{u}$, $\vec{v}$)

(sample) covariance matrix: $S=\frac{AA^T}{n-1}$ 'eigen-' or 'comb-' that contains the most info. or explains the most variance.

Then we find λ and σ. u and the leading direction in the scatter plot.

Total Variance $T = s_1^2+\cdots+s_m^2 = \sigma_1^2+\cdots+\sigma_m^2 = tr(S)$. PC1 accounts for $\frac{\sigma_1^2}{T}$ of total var.

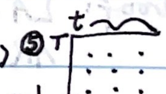
PCA = perpendicular least squares (orthogonal regression): (not OLS/vertical least squares!) $A^TAx=A^Tb$

$\sum_{j=1}^{n} \|a_j\|^2 = \|A^Tu_1\|^2 + \|A^Tu_2\|^2 = \sum_{j=1}^{n} |a_j^Tu_1|^2 + \sum_{j=1}^{n} |a_j^Tu_2|^2$. When $u_1^TAA^Tu_1$ maximized, the squared distances sum minimize (and u_2 the second direction in ortho)

Correlation Matrix. $C=\frac{DA^TAD}{n-1} = DSD$.

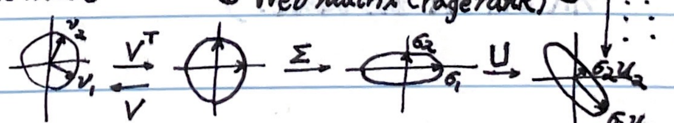
measurements may have diff units and original scaling is not meaningful.

DA makes row have same length $\sqrt{n-1}$. or makes S has 1's on diag by dividing $s_1 \cdots s_n$.

App. Examples: ① Genetic variation in Europe.  ② Eigenfaces. ③ Model order reduction, POD. ④ Web matrix (Pagerank) ⑤  yield curve.

(△ When A is centered and $m>n$, σ's have at most $n-1$ nonzeros.)

7.4 the geometry of the SVD



rotation — stretching — rotation (circle x → ellipse axes u.)

The norm $\|A\| = \max \frac{\|Ax\|}{\|x\|} = \sigma_1$.

tri. ineq. $\|A+B\| \le \|A\| + \|B\|$, product ineq. $\|AB\| \le \|A\|\|B\|$ (proof: $\|Ax\| \le \|A\|\|x\|$, times x)

Eckart-Young (-Mirsky) Theorem: The closest rank k mt to A is $A_k = \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$.
($\|A-B\| \ge \|A-A_k\| = \sigma_{k+1}$, $\forall B$ of $r(B)=k$.)

Polar Decomp.: like $e^{i\theta}r$, $A=QS$ where Q is orthogonal and S is sym. pos. semidefinite.
rotation · stretching $Q=UV^T$, $S=V\Sigma V^T$. if A is inv., S is pos. definite.

▲ $Q=UV^T$ is the nearest ortho. mt to A. ($\|Q-A\|$ min.) (or S is the sq. root of $A^TA = V\Sigma^2V^T$.) (also $A=KQ$ with $K=U\Sigma U^T$, and Q the same)

A_0 is the nearst sin. mt to A by changing σ_{min} to zero.

pseudoinverse of A: $A^+ = V\Sigma^+U^T$ $\quad \underbrace{(v_1 \cdots v_n)}_{n\times n} \overbrace{\begin{pmatrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \end{pmatrix}}^{n\times m} \underbrace{(u_1 \cdots u_m)^T}_{m\times m}$

$A^+u_i = \frac{1}{\sigma_i}v_i$ $(i \le r)$ and $A^+u_i = 0$. $(i > r)$ $\xrightarrow{}$ always exists.

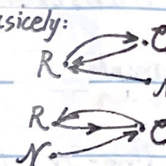
If A^{-1} exists, A^+ is the same as A^{-1}. in that case $m=n=r$ and $V\Sigma^{-1}U^T$.

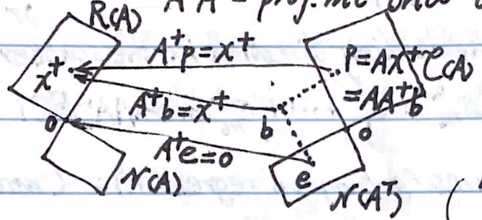
$N(A^{+T}) = N(A)$
$N(A^+) = N(A^T)$
$C(A^+) = R(A)$
$R(A^+) = C(A)$

since $\Sigma^+\Sigma = \begin{pmatrix} I & \\ & 0 \end{pmatrix}$, we get

$AA^{-1} = A^{-1}A = I$ (proj. itself) $\longrightarrow$ $AA^+ =$ proj. mt onto the col space of A.
$A^+A =$ proj. mt onto the row space of A.

consicely:





A^+ can get Ax^+ in col space back to $A^+Ax^+ = x^+$ in row space

$\left(\begin{array}{l} A^+A = \begin{pmatrix} I & \\ & 0 \end{pmatrix}_{n\times n} \begin{array}{l} \text{— row space} \\ \text{— nullspace} \end{array} = \sum_i v_iv_i^T \\ AA^+ = \begin{pmatrix} I & \\ & 0 \end{pmatrix}_{m\times m} \begin{array}{l} \text{— col space} \\ \text{— left nullspace} \end{array} = \sum_i u_iu_i^T \end{array} \right)$ an ex!

least squares with dependent cols: $A^TAx = A^Tb$ has inf. many solutions when A is singular.

pseudoinv. gives a way to choose a best solution: $x^+ = A^+b$ which is the shortest ($\|x^+\|_{min}$, no nullspace part) solution to $A^TA\hat{x} = A^Tb$ and $A\hat{x} = p$.

If A is full col rank ($r=n$) then it has left inv. $L = (A^TA)^{-1}A^T$, $LA = I$. $A^+ = L$ in this case;
If A is full row rank ($r=m$) then it has right inv. $R = A^T(AA^T)^{-1}$, $AR = I$. $A^+ = R$ in this case.
(proof: LA and AL are projs onto row space ($\mathbb{R}^n$) and col space (sub of $\mathbb{R}^m$, the origin P); AR and RA the same.)

8. Linear Transformations

8.1 the idea of a linear transformation

A transformation assigns an output $T(v)$ to each input vector $v \in V$.

Linearity requires $T(cv+dw) = cT(v) + dT(w)$. Note $T(0)=0$, so affine trans. isn't linear. ($T(v) = Av + u_0$. ex.)

Lines → Lines, Triangles → Triangles. Comb → Comb, so basis is important.

$A = \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & & 2 & \\ 0 & 0 & & n \end{pmatrix}$ = mt. form of the derivative $T = \frac{d}{dx}$, with basis $v_i = x^{i-1}$. nullspace a line in func. sp.

and integral give the pseudoinv. $T^+ = \int_0^x dx$. (mt. form $A^+ = \begin{pmatrix} 0 & \cdots & 0 \\ 1 & & 0 \\ & \frac{1}{2} & \\ 0 & & \frac{1}{n} \end{pmatrix}$) $A^+A = \begin{pmatrix} 0 & \\ & I \end{pmatrix}$, $AA^+ = I$.

The product of two trans. $(ST)(v) = S(T(v))$ still linear. Inverse trans. T^{-1} brings every vec $T(v)$ back v.

$T: V \to W$ Range of T = set of outputs. (Image) Kernel of T = set of all inputs for which $T(v)=0$.

Ex. $T(v) = Av$ $\qquad \mathcal{C}(A) \qquad \mathcal{N}(A)$

8.2 the matrix of a linear tansformation

Every lin. trans. from V to W can be converted to a matrix. This mt. depends on bases.

Cols of 1 to n of the matrix will contain those outputs $T(v_1)$ to $T(v_n)$.

$T(v_i)$ = comb of output basis vectors $= a_{1i}w_1 + \cdots + a_{mi}w_i$.

For every $v \in V$ that $v = c_1v_1 + \cdots + c_nv_n$, linearity gives $T(v) = c_1T(v_1) + \cdots + c_nT(v_n)$, as Ac.

Multiplication corresponds. $TS: \underset{\mathbb{R}^p}{U} \to \underset{\mathbb{R}^n}{V} \to \underset{\mathbb{R}^m}{W}$. $AB: (m\times n)(n\times p) = (m\times p)$.

Change of basis: $Vc = V'c'$. (same vec written in diff basis) (B contains new basis vecs b in cols written in the standard/old basis.)

1. $V'B = V$, $B = V'^{-1}V$; $c' = Bc$.
2. $A' = B_{out}^{-1}AB_{in}$ (of course it's inv.)

8.3 the search for a good basis

vec.sp.

1. $B_{in} = B_{out}$ = eigenvec mt. X, $X^{-1}AX$ = eigenvalues in Λ. (sq. A and has n ind. eivecs.)
2. $B_{in} = V$ and $B_{out} = U$: singular vecs of A, $U^{-1}AV = \mathrm{diag}\,\Sigma$. — 'isometric' (def: $C = Q_1^{-1}AQ_2$ when Q_1, Q_2 are ortho.)
3. $B_{in} = B_{out}$ = generalized eigenvecs of A. $B^{-1}AB$ = Jordan form J. (sq. A but only $s<n$ ind. vecs.)

$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{pmatrix}$, Jordan block: $J_i = \begin{pmatrix} \lambda_i & 1 & \\ & \ddots & 1 \\ & & \lambda_i \end{pmatrix}$ $\quad (A - \lambda_i I)b_{j+1} = b_j$ (at one block)

$(J - \lambda_i I)x_{j+1} = x_j$ instead of 0, $x_j = (0; \cdots 0 \overset{j}{1}, 0 \cdots 0)$. and $Bx_j = b_j$.)

Matrices are similar if they share the same Jordan form, not otherwise.

In app. we take J^k and e^{Jt} (involves $e^{\lambda t}$ times powers $1, t, \cdots t^{s-1}$)

(Jordan form is unstable!)

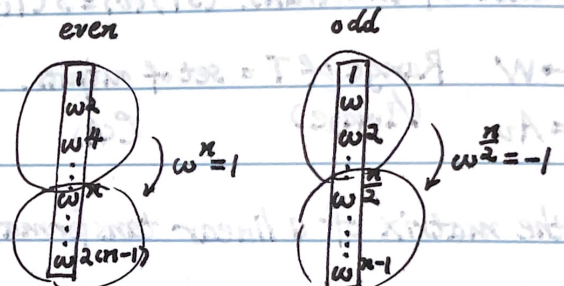
○ 矩阵收集仓（四）：

· Hilbert matrix $H=\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \\ \frac{1}{3} & \frac{1}{4} & & \cdots \\ \frac{1}{4} & & & \end{pmatrix}$, $H_{ij}=\frac{1}{i+j-1}$ $(\int_0^1 x^{i+j-2}dx)$ if use powers $1, x, x^2, x^3 \ldots$ as a func basis, B^TB is not ortho. and equals H : bad! H is ill-conditioned.

○ FFT 形式化补充：

· 递归，$F_n=\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)^2} \end{pmatrix}$

$\blacktriangle \ \triangledown \ \blacktriangle \cdots\cdots \triangledown$



$\begin{matrix}1&1&1\\1&\omega^2&\omega^4\end{matrix} = \begin{matrix}1&1&1\\1&\omega'&\omega'^2\end{matrix}$ so $\begin{pmatrix} I \\ I \end{pmatrix}$.

$\begin{matrix}1&1&1\\\omega&\omega^3&\omega^5\\\omega^2&\omega^6&\omega^{10}\end{matrix} = \begin{matrix}1&1&1\\1&\omega'&\omega'^2\\1&\omega'^2&\omega'^4\end{matrix}$ —$\cdot 1$ —$\cdot\omega$ —$\cdot\omega^2$

so $\begin{pmatrix} D \\ -D \end{pmatrix} = \begin{pmatrix} 1 & & \\ & \omega & \\ & & \omega^2 \\ -1 & & \\ & -\omega & \\ & & -\omega^2 \end{pmatrix}$.

○ proof of Schur Theorem: $U^HAU=T$. (Schur decomp.) ($A\in\mathbb{C}^n$. U utilary, T tri.)

induction: find A's λ_1 a eigenvector ε_1, expand it to a basis. (ortho.)

$U_1=(\varepsilon_1,\eta_2,\cdots,\eta_n)$ $U_1^HAU_1=\begin{pmatrix}\lambda_1 & \cdots \\ 0 & A_1\end{pmatrix}$ then do the same thing on A_1, till A_{n-1}:

finally $(U_1U_2'\cdots U_{n-1}')^HA(U_1U_2'\cdots U_{n-1}')=\begin{pmatrix}\lambda_1 & & \cdots \\ & \ddots & \\ & & \lambda_n\end{pmatrix}$. $U_2^HA_1U_2=\begin{pmatrix}\lambda_2 & \cdots \\ 0 & A_2\end{pmatrix}$, $U_2'=\begin{pmatrix}1 & 0 \\ 0 & U_2\end{pmatrix}$.

Schur inequality: $\sum_i|\lambda_i|^2\le\sum_i\sum_j|a_{ij}|^2$. (proof: $\sum_i|\lambda_i|^2=\sum_i|t_{ii}|^2\le\sum_i|t_{ii}|^2+\sum_{i\ne j}|t_{ij}|^2=tr(TT^T)$ $= tr(AA^T) = \sum_i\sum_j|a_{ij}|^2$.

(eq. when A is normal i.e. unitary similar diag.)

○ change in A^{-1} from a change in A: Sherman-Woodbury-Morrison formula:

1. $M=I-uv^T$, $M^{-1}=I+\frac{uv^T}{1-v^Tu}$ (rank-one change) not inv. if $v^Tu=1$.
2. $M=A-uv^T$, $M^{-1}=A^{-1}+\frac{A^{-1}uv^TA^{-1}}{1-v^TA^{-1}u}$ ▲ $Ax=b\to My=b$: $(M=A-uv^T)$
   ① solve $Az=u$, compute $c=1-v^Tz$
   ② $c\ne0$ then $y=x+\frac{v^Tx}{c}z$.
3. $M=I-UV$, $M^{-1}=I_n+U(I_m-VU)^{-1}V$
4. (matrix inversion lemma) $M=A-UW^{-1}V$, $M^{-1}=A^{-1}+A^{-1}U(W-VA^{-1}U)^{-1}VA^{-1}$

○ Hamilton-Cayley Theorem.（见前）$f(\lambda)=\det(\lambda I-A)$ then we have $f(A)=0$.

diagnolizable mts. are dense.

corollary: · A^k can be simplified by $f(A)=0$ (downgrade)

· A^{-1} can be expaned by a $g(A)$ ($f(A)$ times A^{-1} to get)

4. $B_{in} = B_{out}$ = Fourier matrix F. Fx is a Discrete Fourier Transformation of x. (DFT)

the eigenvector mt. F diagonlizes the permutation mt. P: ex. $\begin{pmatrix} & 1 & & \\ & & 1 & \\ & & & 1 \\ 1 & & & \end{pmatrix}$ $\Lambda = \begin{pmatrix} 1 & & & \\ & i & & \\ & & -1 & \\ & & & -i \end{pmatrix}$ $F = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$

F is sym. ortho? (unitary) Circulant matrix: $C = \begin{pmatrix} c_0 & c_1 & c_2 & c_3 \\ c_3 & c_0 & c_1 & c_2 \\ c_2 & c_3 & c_0 & c_1 \\ c_1 & c_2 & c_3 & c_0 \end{pmatrix}$ eigenvalues $c_0 + c_1\lambda + c_2\lambda^2 + c_3\lambda^3$ ($\lambda = \pm 1, \pm i$)

func.sp. its four eigenvalues are given by $Fc = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix}$

5. The Fourier basis. $1, \sin x, \cos x, \sin 2x, \cos 2x \cdots$ periodic and orthogonal. so Fourier coeff.

6. The Legendre basis. $1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x \cdots$ ex. $a_i = \dfrac{\int f(x)\cos ix\,dx}{\int \cos ix \cos ix\,dx}$.

come from applying the G-S idea to orthogonalize $1, x, x^2, x^3 \ldots$ ex. $\dfrac{(x^3, x)}{(x, x)} = \dfrac{3}{5}$, so $x^3 - \frac{3}{5}x \perp 1, x, x^2$

7. The Chebyshev basis. $1, x, 2x^2 - 1, 4x^3 - 3x \cdots$ $\cos n\theta \to \cos\theta$. (remind even $\perp$ odd (x).)

9. Complex Vectors and Matrices

9.1 complex numbers   complex plane. polar form. nth roots of 1. (unity) $|z|^2 = r^2 = z\bar{z}$.

$\omega = e^{\frac{2\pi i}{n}}$. $\frac{1}{z} = \frac{\bar{z}}{|z|^2}$, $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$.

9.2 hermitian and unitary matrices

Real versus Complex in a nutshell: When you transpose vec z or mt A, take the complex conjugate too.

$\bar{z}^T = z^H$ called Hermitian or adjoint. (For truly useful $u^H u = \|u\|^2$ not $u^T u$.)

dot.p / inner.p: $u^H v$ the order is now important $u^H v \neq v^H u$. $(Au)^H v = u^H (A^H v)$

$(AB)^H = B^H A^H$

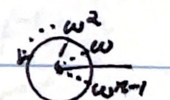
$S = S^H$ called Hermitian matrix. $z^H S z$ is real for $\forall z$; all eigenvalues of S is real; eigenvecs from diff eigenvalues are ortho. $q_1^H q_2 = 0$

$Q^H Q = I$, $Q^H = Q^{-1}$ (sq.) called unitary.

$\|Qz\| = \|z\|$ and then $Qz = \lambda z$ leads to $|\lambda| = 1$.

▲ good row space is no longer $C(A^T)$ but $C(A^H)$.

9.3 the Fast Fourier Transform (FFT) $F_n c = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \cdots & \cdots & \omega^{(n-1)^2} \end{pmatrix}\begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{pmatrix} = y = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{pmatrix}$



$\omega = e^{\frac{2\pi i}{n}}$.

$F_n^H = \bar{F}_n$. $F_n \bar{F}_n = nI$, $F_n^{-1} = \frac{1}{n}\bar{F}_n$.

frequency space c $\underset{\bar{F}\,(F^{-1})}{\overset{F}{\rightleftarrows}}$ physical space y

(with $\bar{w} = \bar{\omega} = e^{-\frac{2\pi i}{n}}$, or permute row $i \leftrightarrow N-i$)

▲ y_j = the Fourier series $\sum_k c_k e^{ikx_j}$. those n points x (angles) equally spaced around.

▲ entry in row j col k is ω^{jk}, zeroth row and col contains all $\omega^0 = 1$.

▲ Fr.mt. is the Vandermonde mt. for interpolation at n. ($p(z) = f$)

recursion: $F_n \to F_{\frac{n}{2}}$

$F_n = \begin{pmatrix} I_{\frac{n}{2}} & D_{\frac{n}{2}} \\ I_{\frac{n}{2}} & -D_{\frac{n}{2}} \end{pmatrix}\begin{pmatrix} F_{\frac{n}{2}} & \\ & F_{\frac{n}{2}} \end{pmatrix}\begin{pmatrix} 1 & & & & \\ & & 1 & & \\ \cdots & \cdots & \cdots & 1 & \cdots \\ & 1 & & & \\ & & & & 1 \end{pmatrix}$ even-odd perm.

$y_j = \sum_0^{n-1} \omega^{jk} c_k = \sum_0^{\frac{n}{2}-1} \omega^{j2k} c_{2k} + \sum_0^{\frac{n}{2}-1} \omega^{j(2k+1)} c_{2k+1}$

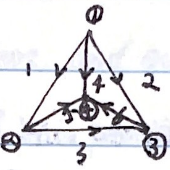
$(\omega' = \omega^2)$ $= \sum_0^{\frac{n}{2}-1} \omega'^{jk} c_{2k} + \sum_0^{\frac{n}{2}-1} \omega'^{jk} c_{2k+1} \cdot \omega^j = y'_{j,even} + \omega^j y'_{j,odd}$

$D = \mathrm{diag}(1, \omega, \cdots, \omega^{\frac{n}{2}-1})$

$O(n^2) \to O(\frac{1}{2} n \log_2 n)$ (C's final permutation: $\overleftarrow{0101}$ bit-reversed order)

10. Applications

10.1 graphs and networks



$$A = \begin{array}{c} \begin{array}{cccc} ① & ② & ③ & ④ \end{array} \\ \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \end{array}$$

node

edge 1 5 6

The incidence matrix A comes from a connected graph with n nodes and m edges.

Elimination reduces every graph to a tree. Rows' dependcy $\longleftrightarrow$ edges form a loop.

$N(A)$: The constant vec $(c, c \cdots, c)$ make up the nullspace, $\dim = 1$.

$C(A^T)$: The edges of any tree give r ind. rows, $r = n-1$.

△ Use orthogonality to judge a ver whether in a space of A.

$C(A)$: Ax gives voltage diffs. Voltage Law: The components of Ax add to zero around all loops. $\dim = n-1$

$N(A^T)$: Current Law: $A^Ty = 0 =$ (flowin) $-$ (flowout) is solved by loop currents, there are $m - r = m-n+1$ ind. (small) loops.

Euler's Formula: Nodes $-$ Edges $+$ Loops $= 1$. (small)

since Ohm's Law $y = -CAx$. C is the conductance mt. (diag)

if all $c = 1$, we get the (graph) Laplacian matrix A^TA. (With A and C, we call it a Network.)

$A^TCAx = f$ (batteries / current sources ..)

△ A^TA is not inv. one node has to be grounded. $(x_j = 0)$ Thus remove one row and col.

10.2 matrices in engineering

Δx, $0 \; \Delta x \; 2\Delta x \cdots n\Delta x$, $\Delta x = \frac{1}{n+1}$

$-\frac{d}{dx}\left(c(x)\frac{d}{dx}\right) = f(x)$ with boundary conditions: $u(0) = 0$ and ($u(1) = 0$ or $\frac{du}{dx}(1) = 0$) ('fixed-free' or 'fixed-fixed')

divide the bar into n pieces of Δx;

replace $\frac{d}{dx}$ by A and $-\frac{d(\,)}{dx}$ by A^T; (they include $\frac{1}{\Delta x}$)

end conditions are $u_0 = 0$ and ($u_N = 0$ or $y_N = 0$); $(Ku = f)$

$c(x)$ corresponds to C, thus: $f = A^Ty$, $y = Ce$, $e = Au$ give $A^TCAu = f$.

fix, y, u, y, u, y, u, $\downarrow g$, free

we have three choices in A replacing $\frac{d}{dx}$: $\frac{u(x+\Delta x) - u(x)}{\Delta x}$, $\frac{u(x) - u(x-\Delta x)}{\Delta x}$, $\frac{u(x+\Delta x) - u(x-\Delta x)}{2\Delta x}$.

forward, backward, centered

(forward diff)$^T = -$ backward diff.

fixed-fixed $\begin{pmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & 1 & -1 \end{pmatrix}$ $\begin{pmatrix} -1 & 1 & \\ & -1 & 1 \\ & & -1 \end{pmatrix}$. fixed-free $\begin{pmatrix} 1 & -1 & \\ & 1 & -1 \\ & & 1 \end{pmatrix}$ $\begin{pmatrix} 1 & & \\ -1 & 1 & \\ & -1 & 1 \end{pmatrix}$, free-free $\begin{pmatrix} -1 & & \\ 1 & -1 & \\ & 1 & \end{pmatrix}$ $\begin{pmatrix} -1 & 1 & \\ & -1 & 1 \end{pmatrix}$.

K's properties: $\begin{pmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{pmatrix}$

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

$\begin{pmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 1 \end{pmatrix}$

△ sin. semipos. a $\lambda = d = 0$. $x_N = (1,1,1)$, thus f have $f_1 + f_2 + f_3 = 0$

$\begin{pmatrix} 1 & -1 & \\ -1 & 2 & -1 \\ & -1 & 1 \end{pmatrix}$

① sym. ② tridiag.

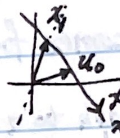
③ pos. def. ($c_i > 0$ and A has ind. cols)

④ K^{-1} is full (not sparse) with all entries > 0.

$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$ also $x_N = (1,1,1)$ in nullspace.

- some proof of Markormt.'s properties:

A's cols adds to 1 $\rightarrow (1,1,\cdots)^T$ an A^T's eigenvec with $\lambda=1$, also A's λ. $\rightarrow$ other eigenvecs will perpendicular to A^T's $(1,1,\cdots)^T$, which means



(so all u_0 would finally go to x_1)

$\longrightarrow$ we need to proof that all other $|\lambda|<1$:

□ Perron-Frobenius Theorem:

- spectrum radius $\rho(A)=\max|\lambda(A)|$, here are some equal conditions: ① $\rho(A)<1$ ② $\lim_{k\to\infty}A^k=0$ ③ $\sum_{k=0}^{\infty}A^k$ converges. $(=(I-A)^{-1})$ Neumann Series. with $\rho(A)<1$ it can converge.

($\approx$ 2-norm) (Gelfand Formula: $\rho(A)=\lim_{k\to+\infty}|||A^k|||^{\frac{1}{k}}\leq|||A|||$)

- ($A\geq0$) specr's monotocity: $B\in\mathbb{C}$ s.t. $|b_{ij}|<a_{ij}\ \forall i,j$. then $\rho(B)\leq\rho(|B|)\leq\rho(A)$.

specr with sums: $\min_{i\text{ or }j}\sum a_{ij}\leq\rho(A)\leq\max_{i\text{ or }j}\sum a_{ij}$ (row/col)

Collatz-Wielandt Formula: $\min_{(1\leq i\leq n)}\frac{(Ax)_i}{x_i}\leq\rho(A)\leq\max\frac{(Ax)_i}{x_i}$ ('=' when x is eivec)

- Perron-Frobenius: $A\geq0$ and irreducible:

then 1. $\rho(A)>0$ and has GM=AM=1. 2. all nonneg. eigenvec correspond to eigenvalue $\rho(A)$ (all other λ's vec has entries <0)

3. $\exists$ only one $v\in\mathbb{R}^n$ s.t. $v>0$, $\sum_i v_i=1$ and $Av=\rho(A)v$; (left eivec) ..... $v^Tw=1$ and $w^TA=\rho(A)w^T$ (or $A^Tw=\rho(A)w$)

4. Perron Proj.: $\lim_{k\to\infty}\left(\frac{A}{\rho(A)}\right)^k=vw^T$ (reducible: $\exists P$ s.t. $P^TAP=\begin{pmatrix}B&C\\&D\end{pmatrix}$ (blockized upper tri.) then A's graph X strongly connected.)

- sensitivity of eigenvalues: A makes small change ΔA, which results in $\Delta\Lambda$:

$$\Delta\Lambda=(X+\Delta X)^{-1}(A+\Delta A)(X+\Delta X)-X^{-1}AX=(X^{-1}-X^{-1}\Delta XX^{-1})A(X+\Delta X)+X^{-1}\Delta AX-X^{-1}AX$$

we use $X^{-1}A=\Lambda X^{-1}$ $\longrightarrow$ $=\Lambda X^{-1}\Delta X-X^{-1}\Delta X\Lambda+X^{-1}\Delta AX$. (omit the second term)

now we'll show that $\Lambda X^{-1}\Delta X$ and $X^{-1}\Delta X\Lambda$ cancel on the diag: $\Lambda(X^{-1}\Delta X)$ with $(X^{-1}\Delta X)\Lambda$.

so $\Delta\Lambda\approx X^{-1}\Delta AX$,

or $\Delta\lambda=y^T\Delta Ax$. ($Ax=\lambda x$, $A^Ty=\lambda y$)

- proof of $(I-A)^{-1}$: $\rho(A)<1$ vice above; $\rho(A)=1$ means a $\lambda=0$, singular; $\rho(A)>1$ means a $\lambda=\frac{1}{1-\rho}<0$, since its eigenvec $u>0$, $(I-A)^{-1}u=\lambda u<0$ shows there must be entries <0.

- matrix series: $\sum_i c_i z^i$ has convergence radius R, then $(A\in\mathbb{C}^{n\times n})\sum_i c_iA^i$: ① $\rho(A)<R$, absolute converge; ② $\rho(A)>R$, diverge.

10.3 Markov matrices, population, and economics

Markov mt. : every entry of A is positive $a_{ij}>0$ $(A>0)$; every cols of A adds to 1.

Then $\lambda_1 = 1$ is larger than any other values, its vec x_1 is the steady state: $u_k = x_1 + \sum_{i=2}^{n} c_i (\lambda_i)^k x_i$

$u_\infty \to x_1$.

- Perron-Frobenius Theorem (正定) $(\rho(M)=1)$: for $A>0$, All numbers in $Ax = \lambda_{max} x$ are positive.

Leslie mt. : populatin: $\begin{pmatrix} F_1 & F_2 & F_3 \\ P_1 & & \\ & P_2 & \end{pmatrix}$ with $\lambda_{max} > 1$ maybe. F's and P's change in ΔA makes $\Delta\lambda$ (不定)

Consumption mt. : $p - Ap = y$, $p = (I-A)^{-1} y$ Neumann/geometric Series:
(Leontief) product. demand. $(I-A)^{-1} = I + A + A^2 + \cdots$

it converges if $\rho(A) < 1$; $\rho(A) = 1$ makes $(I-A^{-1})$ fail to exist; $\rho(A) > 1$ then $(I-A^{-1})$ has neg. entries. (cannot meet demand)

10.4 linear programming

(A has $m<n$ usually)

Primal problem : minimize $c^T x$ with $Ax = b$ $(Ax \geq b)$ and $x \geq 0$

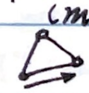
Dual problem : maximize $y^T b$ with $A^T y \leq c$ and $y \geq 0$

$x \geq 0$ and the slack $s = c - A^T y \geq 0$ gave $x^T s \geq 0$ which means $y^T b \leq x^T c$, when $y^T b$ max. $c^T x$ min.
equality means $x^T s = 0$ thus either $x_j = 0$ or $s_j = 0$, i.e. y solves m eqs. $A^T y = c$ in the m components that are nonzero in x.

geometry: a feasible set ('triangle'), maybe unbounded.

The optimal solution will be one of corners. (m nonzeros and $n-m$ zeros)

① list them and compare. ② the simplex method: start from one corner,
enter one variable $0 \to 1$, others have to adjust to keep $Ax=b$.
compute $c \cdot x$ and choose that gives the most neg. change.
then dicide how much it can enter and drop one var. to $0 \downarrow$.
repeat until all changes are pos. (move to neighbor)



③ the interior point method: remove the constrains $x_j \geq 0$ by minimize $c^T x - \theta \sum \log x_i$ with $Ax=b$.

Langrange $L(x,y,\theta) = c^T x - \theta \sum \log x_i - y^T (Ax - b)$

$\frac{\partial L}{\partial x_j} = c_j - \frac{\theta}{x_j} - (A^T y)_j = 0$ which is $x_j s_j = \theta \to 0$. actually solved by Newton's method (iteration)

10.5 Fourier series : linear algebra for functions

Hilbert space : if and only if their length are finite: $\|v\|^2 = v_1^2 + v_2^2 + v_3^2 + \cdots$ $\|f\|^2 = \int_0^{2\pi} f(x)^2 dx$ — $(f,g) = \int_0^{2\pi} f(x) g(x) dx$ (Schwarz ineq. ✓ $|v \cdot w| \leq \|v\| \|w\|$)

$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots$ (orthon. basis. means $\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \cdots$)

$\|f\|^2 = 2\pi a_0^2 + \pi (a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots)$

since basis are ortho. (orthon.) $a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx$, $b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx$ $\left(a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx = \overline{f(x)}\right)$

ex. $\delta(x)$;

10.6 computer graphics

we use homogeneous coordinates. note to seperate 'point' with 'vector'. $(x,y,z,1)$ (x,y,z)

Here are some operations: translation $T = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ x_0 & y_0 & z_0 & 1 \end{pmatrix}$; (cx, y, z, c) like to rescale the space by $\frac{1}{c}$) (add two points like to find middle point.)

rescaling $S = \begin{pmatrix} c_1 & & \\ & c_2 & \\ & & c_3 & \\ & & & 1 \end{pmatrix}$; rotation (affine) $R = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}$; projection $\begin{pmatrix} I - uu^T & 0 \\ 0 & 1 \end{pmatrix}$ △ we use row vec!

(any affine matrix $A = \begin{pmatrix} T(1,0,0) & 0 \\ T(0,1,0) & 0 \\ T(0,0,1) & 0 \\ T(0,0,0) & 1 \end{pmatrix}$ (mirror $I - 2uu^T$)

sometimes there are more steps: ex. proj. to a flat. $\begin{pmatrix} I & \\ -v_0 & 1 \end{pmatrix} \begin{pmatrix} I - nn^T & \\ & 1 \end{pmatrix} \begin{pmatrix} I & \\ v_0 & 1 \end{pmatrix}$

10.7 linear algebra for cryptography

modular arithmetic : Inversion of every y $(0<y<p)$ will be possible if and only if p is prime. △

Hill cipher: $y, 2y, \cdots py$ have diff reminders, there must be 1. $((m-n)y \equiv 0 \mod p$, since $m-n$ or $y<p$, cannot.)

divide message into blocks $\vec{v}_1, \vec{v}_2, \cdots$ of size n.

multiply each by encryption mt. $E \pmod p$. decryption mt. will be E^{-1} (p is prime and this exists.)

finite field $F_p = \{0,1,2,\cdots,p-1\}$, or p^k members. define its $+, \times$ rules. (ex. mod) (p's factors won't have inverse, $\equiv 1$)

F_2^3 have 8 vecs. 2^9 mts. over F_2 (3×3), maybe sin. (It can be difficult when p is large or done many times.)

11. Numerical Linear Algebra speed. accuracy. stability

11.1 Gaussian elimination in practice

roundoff error ⟶ partial pivoting: choose the largest num in row k or below, exchange. (may also exchange cols or rescale)

for pos. definite mt, row exchanges aren't required. (no improve)

operation counts of full mt.: Gaussian elim. has two advantages over $A^{-1}b$:

① $\frac{1}{3}n^3$ mul-sub compared to n^3 in A^{-1} (next n^2 equal)

orthoganalization: $\frac{2}{3}n^3$ (Householder) (Givens) (先消)

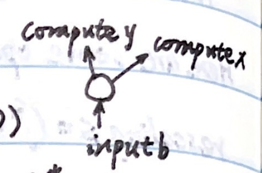
② if A is banded so are L, U, but A^{-1} is full. $(w^2n + 2wn)$

△Euler angles: $Q_{32}Q_{31}Q_{21}A = R$, where A is ortho so $R = I$! $A = Q_{21}^{-1}Q_{31}^{-1}Q_{32}^{-1}$

□ LA in probablility and statistics:

线代在概统中(简)

· Monte Carlo method: $E(x) \leftarrow \frac{\sum_i x_i}{N}$, $\sigma \sim \frac{1}{\sqrt{N}}$



multilevel MC: 2-level $E(x) = \frac{1}{N}\sum_{i=1}^{N} y(b_i) + \frac{1}{N^*}\sum_{i=1}^{N^*}(x(b_i) - y(b_i))$

(for same accuracy, since $x-y$ has a smaller σ^*, N^* can smaller than N)

fixed cost $NC + N^*C^* = T$, optimal ratio $\frac{N^*}{N} = \frac{\sigma^*}{\sigma}\sqrt{\frac{C}{C^*}}$

3-level $E(x) \approx \frac{\sum z}{N} + \frac{\sum(y-z)}{N^*} + \frac{\sum(x-y)}{N^{**}}$ (σ_{min}) ... and optimize those N.

· Covariance matrix: $V = E((X-\bar{X})(X-\bar{X})^T)$ ($\sum\sum$ or $\int dX$)

V is sym. positive semidefinite (most pos., unless the experiments are dependent) (sin. $|V|=0$)

$\hookrightarrow u^TVu = E(\|u^T(X-\bar{X})\|^2) \geq 0.$

correlation mt. $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \in [-1,1]$ and we have $\sigma_i^2\sigma_j^2 \geq \sigma_{ij}^2$ transform for $Z = AX$:

$R = \begin{pmatrix} 1 & \rho_{xy} \\ \rho_{yx} & 1 \end{pmatrix} = DVD$ where $D = \mathrm{diag}(\frac{1}{\sigma_i})$. $V_Z = AV_XA^T$

· (multivariate) Gaussian $p(x) = \frac{1}{(\sqrt{2\pi})^m|V|} e^{-\frac{(x-m)^TV^{-1}(x-m)}{2}}$, m = means of x.

we can diagonalize the cov. mt $V^{-1} = Q\Lambda^{-1}Q^T$, $Y = Q^TX$, $X = QY$, which means finding combs. that are ind. (experi.)

verify: $\int p'(x)dX = \int e^{-\frac{Y^T\Lambda^{-1}Y}{2}}dY = \prod_i \sqrt{2\pi\lambda_i} = \sqrt{2\pi}^m|V|$ $(dX = dY)$

$\int Xp(X)dX = \int Xp(X)dX + m\int p(x)dx = m.$

$\int(x-m)p(x)(x-m)^Tdx = Q\int YY^Te^{-\frac{Y^T\Lambda^{-1}Y}{2}}dY \cdot Q^T = Q\Lambda Q^T = V$

$\sqrt{2\pi}^m|V|$ ($\hookrightarrow Z = \frac{1}{\sqrt{\Lambda}}Y$, $dY = \sqrt{\Lambda}dZ$.

$\int ZZ^Te^{-\frac{Z^TZ}{2}}dZ = \frac{1}{\sqrt{2\pi}^m}I$)

· weighted least squares:

previously we choose $\hat{x}$ minimize $\|b-Ax\|^2$, get $A^TA\hat{x} = A^Tb$. ← normal dis. $\sigma^2=1, m=0$.

now we minimize $E = (b-Ax)^TV^{-1}(b-Ax)$

(if b's errors are not ind. or variances not equal)

'whitening noise' —— $A^TV^{-1}Ax = A^TV^{-1}b$ $(\min \sum_{i=1}^{m} \frac{(b-Ax)_i^2}{\sigma_i^2})$

(means $\frac{b_i}{\sigma_i}$ to get $N(0,1)$. then $V^{-\frac{1}{2}}Ax = V^{-\frac{1}{2}}b$

substitute by $A \rightarrow V^{-\frac{1}{2}}A$ and $b \rightarrow V^{-\frac{1}{2}}b$)

variance of $\hat{x}$: since $\hat{x} = Lb = (A^TV^{-1}A)^{-1}A^TV^{-1}$ use trans. formula to get $V(\hat{x}) = (A^TV^{-1}A)^{-1}$)

· recursive least squares: static / dynamic: Kalman filter:

$A_0\hat{x}_0 = b_0$ (without $x_k = F_{k-1}x_{k-1} \cdot C_k$ changing state here)

update: $\begin{pmatrix} A_0 \\ A_1 \end{pmatrix}\hat{x}_1 = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$ $(A_0^T A_1^T)\begin{pmatrix} V_0^{-1} & \\ & V_1^{-1} \end{pmatrix}\begin{pmatrix} A_0 \\ A_1 \end{pmatrix}\hat{x}_1 = (A_0^T A_1^T)\begin{pmatrix} V_0^{-1} & \\ & V_1^{-1} \end{pmatrix}\begin{pmatrix} b_0 \\ b_1 \end{pmatrix}$

→ $\hat{x}_1 = \hat{x}_0 + K_1(b_1 - A\hat{x}_0)$. cov. mt. $W_1^{-1} = W_0^{-1} + A_1^TV_1^{-1}A_1$ (the same form)

Kalman gain mt. $K_1 = W_1A_1^TV_1^{-1}$

11.2 norms and condition numbers

we choose ℓ^2 norm here. sym. A : $\|A\| = |\lambda|_{max}(A)$ ($=\rho(A)$).

sym. or unsym. A: $\|A\| = \sqrt{\lambda_{max}(A^TA)}$ ($=\sigma_{max}(A)$).

sensitivity to error:

$Ax = b$ — $A(x+\Delta x) = b + \Delta b$ $\quad \frac{\|\Delta x\|}{\|x\|} \leq c \frac{\|\Delta b\|}{\|b\|}$ condition number $c = \|A\|\|A^{-1}\| \geq 1$

— $(A+\Delta A)(x+\Delta x) = b$ $\quad \frac{\|\Delta x\|}{\|x+\Delta x\|} \leq c \frac{\|\Delta A\|}{\|A\|}$ (for pos. mt, $c = \frac{\lambda_{max}}{\lambda_{min}}$)

$\frac{\|\Delta x\|}{\|x\|} \approx$

equal when b along largest eigen, Δb along smallest eigen, amplified by c.

11.3 iterative methods and preconditioners

split A into $S-T$. iteration $Sx_{k+1} = Tx_k + b$ — two goals: speed per step and fast converg.

1. Jacobi method: keep the diag(A) as S, off-diag part is T. $\qquad \rho < 1$ (diag) or not.)

2. Gauss-Seidel method: keep the lower tri. part as S. $\qquad$ error eq. $e_{k+1} = S^{-1}Te_k$

(cut storage and usually speed up iter.) $\qquad$ ($e_i = x_\infty - x_i$)

& SOR (successive overrelaxation): S has diag(A) and below diag of ωA.

$\omega Ax = \omega b$

choose ω to make the $\rho(S^{-1}T)$ small asp.

3. Elimination (exact LU) (triang. very fast.)

& incomplete LU: ('fill-in'. sparse mt. that has nonzeros far from diag.)

$L_0U_0x_{k+1} = (L_0U_0 - A)x_k + b$ set small nonzeros to 0 in L.U.

4. multigrid 5. conjugate gradients & preconditioned CG.

& GMRES. MINRES. $\qquad$ 7. <krylov> $\qquad$ (A sym. H tridiag.) $AQ = QH$

for eigens: 6. power method & inverse power method: <Arnodi iter. ; Lanczos iter>

$u_k = A^k u_0$ the largest eivalue dominates. (u_k gradually points towards x_1);

or by A^{-1}, solve $Au_{k+1} = u_k$ (by saving L.U), λ_{min} dominates. For speed we shifted:

$A - \lambda^* I$ (λ^* can choose a Rayleigh quo. $\frac{x^TAx}{x^Tx}$) to make λ_{min} even smaller.

8. QR method: $A = QR$, $A_1 = RQ = Q^{-1}AQ$ thus same eivalue with A.

$A_1 = Q_1R_1$, $A_2 = R_1Q_1$ .... eivalues begin to show up in diag of A_n. (last entry $(0 0 \cdots 0 \lambda_n)$, then remove this r.c.)

two extra ideas:

① shift. $A_k - c_kI$ into Q_kR_k. $A_{k+1} = R_kQ_k + c_kI$ back. (c_k choose to near an (unknown) eivalue through diag of A_k.)

② elimination. obtain zeros before. EAE^{-1} (or Givens) (move) leave nonzeros along subdiag, which QR factor. drops from (Hessenberg mt.) $O(n^3) \to O(n^2)$.