

矩阵论笔记

Cheanus

2026 年 1 月 18 日

目录

1 微积分与矩阵求导	3
1.1 分子布局与分母变量	3
1.2 链式法则中的转置约定	4
2 矩阵分解	4
2.1 方阵的分解	4
2.1.1 对角化	5
2.1.2 Jordan 矩阵	5
2.2 对称/正定矩阵的分解	5
2.3 任意实矩阵的分解	6
2.3.1 SVD 与奇异值性质	6
2.3.2 Ky-Fan 范数不等式	7
2.4 秩分解	7
3 矩阵概念	7
3.1 幂等、幂单与幂零矩阵	7
3.2 Hermitian 矩阵 (复对称矩阵)	8
3.3 反 Hermitian 矩阵	8
3.4 正定与半正定矩阵	9
3.5 协方差矩阵	9
3.6 随机矩阵	10
3.7 投影矩阵	10
3.8 Householder 变换	10
3.9 初等变换矩阵	11
3.10 初等矩阵	11
3.11酉矩阵 (复正交矩阵)	11
3.12 正规矩阵	12
3.13 Vandermonde 矩阵	12
3.14 Toeplitz 矩阵	12
3.15 循环矩阵	13

3.16 Fourier 矩阵	13
3.17 Hardmard 矩阵	13
3.18 稀疏矩阵	14
3.19 Markov 矩阵	14
3.20 上/下三角矩阵	14
3.21 旋转矩阵	15
4 矩阵关系	15
4.1 相似	15
4.2 合同	15
5 运算	16
5.1 直和	16
5.2 Hardmard 积	16
5.3 Kronecker 积	16
5.4 迹	17
5.5 伪逆 (Moore-Penrose 逆)	17
5.6 行列式与可逆性	18
5.7 矩阵指数	21
6 定理	21
6.1 Cayley-Hamilton 定理与特征/最小多项式	21
6.2 Frobenius 不等式及常见秩不等式	22
6.3 秩-零化度定理	23
6.4 维数定理	23
6.5 Rayleigh 商与极值	23
6.6 Gersgrin 圆	24
6.7 Courant-Fischer 极小极大定理	24
6.8 Weyl 定理	24
6.9 Cauchy 交错定理	24
6.10 RIP	25
6.11 Schur 补定理	25
6.12 Sylvester 方程	25
6.13 Weinstain-Aronszajn 恒等式	26
7 研究	26
7.1 方程有无解	26
7.2 0-1 矩阵满秩的概率	26
7.3 矩阵的条件数	27
7.4 特征值与特征向量	27
7.5 代数重数与几何重数	27

7.6	$AB = BA$	28
7.7	行列式因子、不变因子和初等因子	28
7.8	施密特正交化	28
7.9	核/零空间	29
8	范数	29
8.1	向量范数	29
8.2	向量内积	30
8.3	矩阵范数	30
8.4	矩阵内积	31
9	向量空间	31
9.1	定义	31
9.2	子空间	32
9.3	两子空间的主角	32
10	代数不等式	33
10.1	均值不等式	33
10.2	Cauchy-Schwarz 不等式	33
10.3	Jensen 不等式	33
11	概率论	33
11.1	期望和方差	33
11.2	常见分布	34

1 微积分与矩阵求导

本节总结在向量和矩阵情形下常用的求导记号与链式法则要点，采用“分子布局”的记号约定。

1.1 分子布局与分母变量

- **分子布局：**将待求导的标量或向量函数放在“分子”，自变量放在“分母”，例如

$$\frac{\partial f}{\partial \mathbf{x}^T}, \quad \frac{\partial f}{\partial X},$$

其中 \mathbf{x} 为向量， X 为矩阵。

- 对**向量变量**求导时，假设 f 是列向量，常见记号为 $\frac{\partial f}{\partial \mathbf{x}^T}$ ，则结果矩阵的第 i 行、第 j 列元素为 $\frac{\partial f_i}{\partial x_j}$ 。
- 对**矩阵变量** X 求导时，假设 f 是标量函数，常直接写成 $\frac{\partial f}{\partial X}$ ，不再对 X 取转置，则结果矩阵的第 i 行、第 j 列元素为 $\frac{\partial f}{\partial X_{ij}}$ 。

- 链式法则中，若分母变量为向量 \mathbf{x} ，则所有涉及 \mathbf{x} 的分母均写成 \mathbf{x}^T 的形式；若分母变量为矩阵 X ，则所有涉及 X 的分母均写成 X 的形式。

分子布局下，对向量 $\mathbf{x} \in \mathbb{R}^n$ 和矩阵 $A \in \mathbb{R}^{m \times n}$ ，有

$$\frac{\partial \mathbf{x}}{\partial \mathbf{x}^T} = I_n, \frac{\partial A\mathbf{x}}{\partial \mathbf{x}^T} = A, \frac{\partial \mathbf{x}^T A}{\partial \mathbf{x}^T} = A^T, \frac{\partial \mathbf{x}^T \mathbf{x}}{\partial \mathbf{x}^T} = 2\mathbf{x}, \frac{\partial \mathbf{x}^T A\mathbf{x}}{\partial \mathbf{x}^T} = (A + A^T)\mathbf{x}.$$

以及

$$\frac{\partial \text{tr}(AX)}{\partial X} = A^T, \frac{\partial \text{tr}(X^T AX)}{\partial X} = AX + A^T X.$$

1.2 链式法则中的转置约定

设 $f = f(\mathbf{y})$, $\mathbf{y} = \mathbf{y}(\mathbf{x})$, 其中 $f : \mathbb{R}^m \rightarrow \mathbb{R}$, $\mathbf{y} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 。采用分子布局并在分母写成 \mathbf{x}^T 的形式时，有

$$\frac{\partial f}{\partial \mathbf{x}^T} = \frac{\partial f}{\partial \mathbf{y}^T} \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T}.$$

要点：一旦在左端分母中写成 \mathbf{x}^T ，则右端所有关于中间变量的分母也都统一写成 “ $.^T$ ” 的形式，以保证维度与布局的一致性。

2 矩阵分解

本节按矩阵类型总结常见的矩阵分解形式及其应用场景。

2.1 方阵的分解

主要针对可逆或正规矩阵。

名称	数学形式	适用条件	说明	应用
特征值分解 (EVD)	$A = PDP^{-1}$	$A \in \mathbb{R}^{n \times n}$, 可对角化	D 为对角矩阵, 对角线为特征值, P 的列为特征向量	动力系统、PCA、谱分析
Schur 分解	$A = QTQ^H$	任意方阵	Q 为酉(正交)矩阵, T 为上三角矩阵(对角线为 A 特征值)	理论分析、稳定性判断
Jordan 分解	$A = PJP^{-1}$	任意方阵	J 为 Jordan 标准形, 用于不可对角化情形	不可对角化矩阵的结构分析
LU 分解	$A = LU$	A 可逆	L 为单位下三角矩阵, U 为上三角矩阵	线性方程组求解、数值计算
LDU 分解	$A = LDU$	A 可逆	L 为单位下三角矩阵, D 为对角矩阵, U 为单位上三角矩阵	数值稳定分解

2.1.1 对角化

性质：

- 若矩阵 A 有 n 个线性无关的特征向量，则 A 可对角化。
- 若矩阵 A 有 n 个不同的特征值，则 A 可对角化。
- 矩阵 A 可对角化的充分必要条件是其特征多项式的每个特征值的代数重数等于几何重数。
- 实对称矩阵、Hermitian 矩阵、酉矩阵和正规矩阵均可对角化。
- 如果 $AB = BA$, 且 A 可对角化, 则 A 与 B 可同时对角化。

2.1.2 Jordan 矩阵

Jordan 矩阵 J 由若干 Jordan 块 $J_k(\lambda)$ 沿对角线排列组成, 每个 Jordan 块对应一个特征值 λ , 形式为

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}_{k \times k}.$$

性质：

- 矩阵 A 的 Jordan 标准形唯一确定, 且与 A 相似。
- Jordan 块的大小的最大值决定了对应特征值在最小多项式中的重数。
- Jordan 块的块个数等于对应特征值的几何重数, 特征值个数等于代数重数。
- 设 λ_i 的 Jordan 块 J_k 大小为 s , 则 $(J_k - \lambda_i I)^x = 0$ 当且仅当 $x \geq s$ 。

2.2 对称/正定矩阵的分解

名称	数学形式	适用条件	说明	应用
谱分解 (对称矩阵的 EVD)	$A = Q\Lambda Q^T$	A 为实对称或正规矩阵	Q 正交, Λ 为实对角矩阵, 含特征值	PCA、协方差矩阵分析
Cholesky 分解	$A = LL^T$	A 为对称正定矩阵	L 为下三角矩阵	数值优化、蒙特卡洛采样
LDL^T 分解	$A = LDL^T$	A 为对称矩阵	L 为单位下三角, D 为对角矩阵, 避免开方	数值稳定分解

2.3 任意实矩阵的分解

名称	数学形式	适用条件	说明	应用
奇异值分解 (SVD)	$A = U\Sigma V^T$	任意 $A \in \mathbb{R}^{m \times n}$	U, V 正交, Σ 为对角且非负实数	降维、推荐系统、图像压缩、伪逆
QR 分解	$A = QR$	任意 $A \in \mathbb{R}^{m \times n}$	Q 正交, R 上三角, 源于施密特正交化	最小二乘、Gram–Schmidt、数值稳定算法
极分解	$A = UP$ 或 $A = PV$	任意实矩阵	U 正交, P 半正定对称, 或反过来	类似复数极坐标的分解, 用于几何变换分析
秩分解	$A = BC$	任意 $A \in \mathbb{R}^{m \times n}$	$B \in \mathbb{R}^{m \times r}$, $C \in \mathbb{R}^{r \times n}$, $r = \text{rank}(A)$	数据压缩、低秩近似

2.3.1 SVD 与奇异值性质

公式:

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

其中 $U \in \mathbb{R}^{m \times m}$ 、 $V \in \mathbb{R}^{n \times n}$ 为正交矩阵, $\Sigma \in \mathbb{R}^{m \times n}$ 为对角矩阵。

- 范数关系:

$$\|A\|_2 = \sigma_1, \quad \|A\|_F = \|\Sigma\|_F = \sqrt{\sum_i \sigma_i^2}.$$

- 行列式:

$$|\det(A)| = |\det(\Sigma)| = \prod_{i=1}^r \sigma_i,$$

- 伪逆: A 的 Moore–Penrose 伪逆为

$$A^+ = V\Sigma^+U^T,$$

其中 Σ^+ 由将 Σ 中非零奇异值取倒数后转置得到。

- 对于 AA^T 的特征值分解:

$$AA^T = U\Lambda U^T$$

其中 $\Lambda = \Sigma\Sigma^T$, 即奇异值的平方构成 AA^T 特征值。

- 矩阵半正定当且仅当奇异值等于其特征值 (HW9)。
- 酉相似不改变矩阵奇异值。
- U 的前 r 列 ($r = \text{rank}(A)$) 构成 A 的列空间的标准正交基, V 的前 r 列构成 A 的行空间的标准正交基; U 的后 $m-r$ 列构成 A 的左零空间 ($A^Tx = 0$) 的标准正交基, V 的后 $n-r$ 列构成 A 的零空间的标准正交基。且 $\text{Null}(A) = (\text{Row}(A))^\perp$, $\text{Null}(A^T) = (\text{Col}(A))^\perp$ 。

- $\sigma_1 = \max_{\|x\|_2=1} \|Ax\|_2, \sigma_n = \min_{\|x\|_2=1} \|Ax\|_2, \sigma_n \leq |\lambda_i| = \frac{\|Av_i\|_2}{\|v_i\|_2} \leq \sigma_1$ 。
- $\sum_k \sigma_k^2 \geq \sum_k \lambda_k^2$ (用 schur 分解证), $\sum_k \sigma_k \geq |\sum_k \lambda_k|$ (SVD+ 迹性质)。(HW9)
- 其他形式:

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T, Av_i = \sigma_i \mathbf{u}_i,$$

其中 \mathbf{u}_i 和 \mathbf{v}_i 分别为 U 和 V 的第 i 列。

- 最佳低秩近似: 对于任意秩为 $k < r$ 的矩阵 B , 有

$$\|A - B\|_F \geq \|A - A_k\|_F = \sqrt{\sum_{i=k+1}^r \sigma_i^2},$$

其中 $A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ 。

2.3.2 Ky-Fan 范数不等式

:

设 $A, B \in \mathbb{R}^{m \times n}$, 则对于任意 $k \leq \min(m, n)$, 有 (HW5)

$$\sum_{i=1}^k \sigma_i(A + B) \leq \sum_{i=1}^k \sigma_i(A) + \sum_{i=1}^k \sigma_i(B).$$

2.4 秩分解

构造方法:

- 通过 SVD: 设 $A = U\Sigma V^T$ 为 A 的 SVD 分解, 取 $B = U[:, 1:r] \Sigma_{[1:r, 1:r]}^{1/2}, C = \Sigma_{[1:r, 1:r]}^{1/2} V_{[:, 1:r]}^T$ 。
- 通过列空间和行空间基底: 选取 A 的列空间的一个基底组成矩阵 B , 选取 A 的行空间的一个基底组成矩阵 C , 则 $A = BC$ 。
- 最简形: 例如先把 A 化成行最简形 C , 然后取 A 的前 r 个列 B , 得 $A = BC$

3 矩阵概念

3.1 幂等、幂单与幂零矩阵

- **幂等矩阵:** 若 $A^2 = A$, 则称 A 为幂等矩阵。幂等矩阵的特征值只可能为 0 或 1, 因此在其 Jordan 标准形中, 对角线元只有 0 和 1, $\text{Rank}(A) = \text{tr}(A)$ 。
- **幂单矩阵:** 若存在正整数 k 使得 $A^k = I$, 则称 A 为幂单矩阵。若 $k = 2$, 则 $A^2 = I$, 此时特征值只能是 1 或 -1 , 在某些语境下也可称为“对合”矩阵。
- **幂零矩阵:** 若存在正整数 k 使得 $A^k = 0$, 则称 A 为幂零矩阵。幂零矩阵的所有特征值均为 0。

性质：

- 幂等矩阵 A 满足 $A(A - I) = 0$, 其最小多项式为 $m_A(x) = x(x - 1)$ 。
- 幂单矩阵 A 的特征值均为单位根, 且满足 $x^k - 1 = 0$ 的最小多项式。
- 幂零矩阵 A 满足 $A^k = 0$, 其最小多项式为 $m_A(x) = x^r$, 其中 $r \leq k$ 为 A 的指数。

3.2 Hermitian 矩阵（复对称矩阵）

复矩阵 $A \in \mathbb{C}^{n \times n}$ 若满足

$$A = A^H,$$

则称 A 为 **Hermitian 矩阵** (厄米矩阵)。在实数情形下, Hermitian 矩阵即对称矩阵。

Hermitian 矩阵具有以下重要性质：

- 所有特征值均为实数。
- 存在酉矩阵 Q 使 $A = Q\Lambda Q^H$, 其中 Λ 为实对角矩阵 (谱定理)。
- 若对任意复向量 \mathbf{x} 有 $\mathbf{x}^H A \mathbf{x} \in \mathbb{R}$, 则 A 为 Hermitian 矩阵。
- 一切半正定 (或正定) 矩阵都是 Hermitian 矩阵: 由定义 $\mathbf{x}^H A \mathbf{x}$ 为实数且 ≥ 0 (或 > 0) 可推出 $A = A^H$ 。
- 线性组合: 若 A, B 为 Hermitian 矩阵, 且 $c_1, c_2 \in \mathbb{R}$, 则 $c_1 A + c_2 B$ 亦为 Hermitian 矩阵。
- 不同特征值对应的特征向量正交, 所有特征向量可以选取为正交的。

3.3 反 Hermitian 矩阵

设 $A \in \mathbb{R}^{n \times n}$ 为实矩阵, 若满足

$$A^H = -A,$$

则称 A 为 **反 Hermitian 矩阵**。

性质：

- 对任意向量 $\mathbf{x} \in \mathbb{R}^n$, 矩阵反 Hermitian 当且仅当 $\mathbf{x}^H A \mathbf{x} = 0$ 。
- 反 Hermitian 矩阵的对角线元素均为零。
- 反 Hermitian 矩阵的特征值要么为零, 要么成纯虚数共轭对出现。
- 反 Hermitian 矩阵的秩为偶数。
- 任意矩阵均可唯一分解为 Hermitian 矩阵与反 Hermitian 矩阵之和: $A = \frac{A+A^H}{2} + \frac{A-A^H}{2}$ 。
- 设 A 为实反对称矩阵, 则 e^A 为正交矩阵且 $\det(e^A) = 1$ 。

3.4 正定与半正定矩阵

设 $A \in \mathbb{R}^{n \times n}$ 为对称矩阵, 则称 A 为

- **正定矩阵**, 若对任意非零向量 $\mathbf{x} \in \mathbb{R}^n$ 有

$$\mathbf{x}^T A \mathbf{x} > 0.$$

记作 $A \succ 0$ 。

- **半正定矩阵**, 若对任意向量 $\mathbf{x} \in \mathbb{R}^n$ 有

$$\mathbf{x}^T A \mathbf{x} \geq 0.$$

记作 $A \succeq 0$ 。

性质:

- 特征值: A 为正定矩阵的充分必要条件是其所有特征值均为正实数; A 为半正定矩阵的充分必要条件是其所有特征值均为非负实数。
- 主子矩阵: A 为正定矩阵的充分必要条件是其所有主子矩阵的行列式均为正; A 为半正定矩阵的充分必要条件是其所有主子矩阵的行列式均为非负。
- 逆与和: 若 A 为正定矩阵, 则 A^{-1} 亦为正定矩阵; 若 A, B 均为半正定矩阵, 则 $A + B$ 亦为半正定矩阵。
- 平方根: 若 A 为半正定矩阵, 则存在唯一的半正定矩阵 B 使得 $B^2 = A$, 记为 $B = A^{1/2} = U \Lambda^{1/2} U^T$, 其中 $A = U \Lambda U^T$ 。
- 矩阵半正定当且仅当 $\lambda_i = \sigma_i$ (schur 分解)。
- $P^H A P$ 的正定性与 A 相同, 若 P 可逆。
- 若 $AB = BA$, 且 A, B 为正定矩阵, 则 AB 亦为正定矩阵。
- A, B 为正定矩阵, 则 AB 的特征值非负 (因为 $AB \sim B^{1/2} A B^{1/2}$)。

3.5 协方差矩阵

设随机向量 $\mathbf{X} \in \mathbb{R}^n$, 其协方差矩阵为

$$\Sigma = \text{Cov}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E} \mathbf{X})(\mathbf{X} - \mathbb{E} \mathbf{X})^T].$$

对任意向量 $\mathbf{a} \in \mathbb{R}^n$, 有

$$\mathbf{a}^T \Sigma \mathbf{a} = \mathbb{E}[\mathbf{a}^T (\mathbf{X} - \mathbb{E} \mathbf{X})(\mathbf{X} - \mathbb{E} \mathbf{X})^T \mathbf{a}] = \mathbb{E}\left[\left(\mathbf{a}^T (\mathbf{X} - \mathbb{E} \mathbf{X})\right)^2\right] \geq 0.$$

因此 Σ 为 **半正定矩阵**。若 \mathbf{X} 的分量线性无关且具有非退化方差, 则 Σ 进一步是正定矩阵。

3.6 随机矩阵

高斯零均值随机矩阵 $G \in \mathbb{R}^{n \times n}$ 的每个元素均独立同分布于正态分布 $\mathcal{N}(0, \sigma^2)$, 则 $\mathbb{E}(G^T G) \approx n\sigma^2 I_n, \mathbb{E}(Gx x^T G^T) \approx \sigma^2 \|x\|_2 I_n$, Gx 的每个元素近似独立同分布于 $\mathcal{N}(0, \sigma^2 \|x\|_2^2)$ 。

对于随机向量 x , 其每个元素服从独立同分布于 $\mathcal{N}(0, \sigma^2)$, 则 $w^T x$ 服从 $\mathcal{N}(0, \sigma^2 \|w\|_2^2)$ 。

3.7 投影矩阵

设 $P \in \mathbb{R}^{n \times n}$ 满足

$$P^2 = P,$$

则称 P 为 **投影矩阵** (有时特指正交投影矩阵)。

正交投影矩阵将向量投影到某子空间上, 满足

$$\langle x - Px, Px \rangle = 0, \quad \forall x \in \mathbb{R}^n,$$

◦

投影矩阵具有以下性质:

- 投影矩阵的特征值仅为 0 或 1。
- 若 P 为对称矩阵, 则为**正交投影矩阵**, 此时 P 可表示为某子空间上的正交投影。
- 投影矩阵在数据分析、信号处理等领域有广泛应用, 用于将数据投影到特定子空间以实现降维或特征提取。

正交投影矩阵 P 可表示为

$$P = A(A^T A)^{-1} A^T,$$

其中 A 的列空间为投影的目标子空间。正交投影矩阵会使向量 2 范数不变大。

3.8 Householder 变换

设 $w \in \mathbb{R}^n$ 为非零向量, $\sigma \in \mathbb{R}$, 则称矩阵

$$H = I - 2ww^H, w^H w = 1$$

为 **Householder 矩阵**。

性质 (HW6):

- Householder 矩阵 H 为对称矩阵且满足 $H^2 = I$, 因此 $H^{-1} = H$ 。
- Householder 矩阵 H 为正交矩阵, 其行列式为 $\det(H) = -1$ 。
- Householder 变换可用于将任意向量 x 反射到与给定向量 y 共线的方向上, 具体通过选择适当的 w 实现。

3.9 初等变换矩阵

初等行变换相当于左乘一个初等矩阵，初等列变换相当于右乘一个初等矩阵。常见的初等变换矩阵包括：

- 交换两行（列）的矩阵 E_{ij} 。
- 将某行（列）乘以非零常数的矩阵 $D_i(k)$ 。
- 将某行（列）加上另一行（列）的倍数的矩阵 $L_{ij}(k)$ 。

3.10 初等矩阵

初等矩阵公式： $E(u, v, \sigma) = I + \sigma uv^H$ 。初等矩阵 E 通过左乘作用于矩阵 A ，实现对 A 的行变换；通过右乘作用于矩阵 A ，实现对 A 的列变换。

性质：

- 可逆性：初等矩阵 E 总是可逆的，其逆矩阵为 $E(u, v, \sigma)^{-1} = E(u, v, \frac{\sigma}{\sigma v^H u - 1})$ ，前提是 $\sigma v^H u - 1 \neq 0$ 。
- 行列式： $\det(E) = 1 - \sigma v^H u$ ，证明可研究其特征值得到。
- 对任意 2 个非零向量 a, b ，存在初等矩阵 E 使得 $Ea = b$ 。

3.11酉矩阵（复正交矩阵）

设 $Q \in \mathbb{C}^{n \times n}$ 满足

$$Q^H Q = Q Q^H = I,$$

则称 Q 为 酉矩阵。在实数情形下，酉矩阵即正交矩阵。

酉矩阵具有以下重要性质：

- 保距性：对任意向量 $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ ，有 $\|Q\mathbf{x} - Q\mathbf{y}\|_2 = \|\mathbf{x} - \mathbf{y}\|_2$ 。
- 保内积性：对任意向量 $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ ，有 $(Q\mathbf{x})^H(Q\mathbf{y}) = \mathbf{x}^H \mathbf{y}$ 。
- 行列式的模为 1： $|\det(Q)| = 1$ 。
- 特征值位于单位圆上：酉矩阵的所有特征值均满足 $|\lambda| = 1$ 。
- 酉矩阵的逆等于其共轭转置： $Q^{-1} = Q^H$ 。
- Kronecker 积封闭性：若 Q_1, Q_2 均为酉矩阵，则 $Q_1 \otimes Q_2$ 亦为酉矩阵。

3.12 正规矩阵

设 $A \in \mathbb{C}^{n \times n}$ 满足

$$AA^H = A^H A,$$

则称 A 为 **正规矩阵**。

正规矩阵具有以下重要性质：

- 存在酉矩阵 Q 使 $A = Q\Lambda Q^H$, 其中 Λ 为对角矩阵 (谱定理的推广)。
- 所有 Hermitian 矩阵、酉矩阵和对称矩阵均为正规矩阵。
- 正规矩阵的特征向量可选取为正交的。
- 矩阵正规当且仅当 $|\lambda_i| = \sigma_i$ (schur 分解)。

3.13 Vandermonde 矩阵

给定数列 x_1, x_2, \dots, x_n , 对应的 **Vandermonde 矩阵**可写为

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}.$$

Vandermonde 矩阵在插值、多项式拟合等领域有重要应用。其行列式为

$$\det(V) = \prod_{1 \leq i < j \leq n} (x_j - x_i),$$

因此当且仅当 x_1, x_2, \dots, x_n 两两不同 (互异) 时, V 为满秩矩阵。

3.14 Toeplitz 矩阵

给定序列 c_0, c_1, \dots, c_{n-1} , 对应的 $n \times n$ **Toeplitz 矩阵**定义为

$$T = \begin{bmatrix} c_0 & c_{-1} & c_{-2} & \cdots & c_{-(n-1)} \\ c_1 & c_0 & c_{-1} & \cdots & c_{-(n-2)} \\ c_2 & c_1 & c_0 & \cdots & c_{-(n-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \cdots & c_0 \end{bmatrix},$$

其中 $c_{-k} = c_k$ 可视为对称 Toeplitz 矩阵。

3.15 循环矩阵

给定序列 c_0, c_1, \dots, c_{n-1} , 对应的 $n \times n$ **循环矩阵**定义为

$$C = \begin{bmatrix} c_0 & c_{n-1} & c_{n-2} & \cdots & c_1 \\ c_1 & c_0 & c_{n-1} & \cdots & c_2 \\ c_2 & c_1 & c_0 & \cdots & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \cdots & c_0 \end{bmatrix}.$$

循环矩阵在信号处理、图像处理等领域有重要应用。其特征值可通过离散傅里叶变换 (DFT) 计算, 具体为

$$\lambda_k = \sum_{j=0}^{n-1} c_j e^{-2\pi i j k / n}, \quad k = 0, 1, \dots, n-1.$$

3.16 Fourier 矩阵

给定正整数 N , 对应的 $N \times N$ **Fourier 矩阵**定义为

$$F_N = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \cdots & \omega^{(N-1)^2} \end{bmatrix},$$

其中 $\omega = e^{-2\pi i / N}$ 为 N 次单位根。

反向 Fourier 矩阵为 $F_N^{-1} = \frac{1}{N} F_N^*$ 。

3.17 Hardmard 矩阵

给定正整数 n , 对应的 $2^n \times 2^n$ **Hadamard 矩阵**定义为

$$H_{2^n} = \begin{bmatrix} H_{2^{n-1}} & H_{2^{n-1}} \\ H_{2^{n-1}} & -H_{2^{n-1}} \end{bmatrix},$$

其中 $H_1 = [1]$ 。

也可用 Kronecker 积表示为

$$H_n = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \otimes H_{n/2}, \quad n = 2^k, k \in \mathbb{N}.$$

快速计算 Hadamard 变换类似 FFT 的快速算法, 时间复杂度为 $O(n \log n)$, 核心公式:

$$H_n \mathbf{x} = \begin{bmatrix} H_{n/2} & H_{n/2} \\ H_{n/2} & -H_{n/2} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{[1:n/2]} \\ \mathbf{x}_{[n/2+1:n]} \end{bmatrix}$$

性质:

- 正交矩阵: $H_n H_n^T = nI_n$, 即 H_n 的列向量两两正交。
- 行列式: $|\det(H_n)|^2 = \det(H_n H_n^T) = n^n \implies |\det(H_n)| = n^{n/2}$ 。
- 最大行列式: Hadamard 矩阵的行列式在所有元素绝对值为 1 的矩阵中达到最大值。

3.18 稀疏矩阵

稀疏矩阵是指大部分元素为零的矩阵。

压缩感知中, $y = Ax$, 其中 A 为测量矩阵, x 为稀疏向量, 已知 x 的稀疏度 $S = \|x\|_0$ 时:

- 若 $S = 0$, 非凸优化, 启发式算法求次优解
- 若 $S = 1$, 凸优化, 高复杂度, 不可微, 解集属于 $S = 0$ 的解集
- 若 $S = 2$, 凸优化, 非零元素较多

关于 L_1 解集属于 L_2 解集的充要条件的讨论见 HW4, 稀疏信号 L_1 重构见 6.10 RIP。

3.19 Markov 矩阵

设 $P \in \mathbb{R}^{n \times n}$ 满足

$$P_{ij} \geq 0, \quad \sum_{j=1}^n P_{ij} = 1, \quad \forall i = 1, 2, \dots, n,$$

则称 P 为 **Markov 矩阵**。

Markov 矩阵具有以下性质:

- 所有特征值的模均不大于 1, 且至少存在一个特征值等于 1, 其特征向量为 $\mathbf{1}$ (全 1 向量), 其代数重数等于几何重数。
- 幂仍然为 Markov 矩阵。
- $A^p \rightarrow ex^T$ 当且仅当 A 是不可约且非周期的 Markov 矩阵, 其中 e 为某个概率向量, x 为左特征向量对应于特征值 1, 即 $A^T x = x, x^T e = 1$ 。

3.20 上/下三角矩阵

矩阵 $A \in \mathbb{R}^{n \times n}$ 若满足

- $A_{ij} = 0$, 当 $i > j$, 则称 A 为 **上三角矩阵**。
- $A_{ij} = 0$, 当 $i < j$, 则称 A 为 **下三角矩阵**。

性质:

- 上/下三角矩阵的乘积仍为上/下三角矩阵。
- 上/下三角矩阵的逆 (若存在) 仍为上/下三角矩阵。
- 上/下三角矩阵的行列式为其对角线元素的乘积。
- 上/下三角矩阵的特征值为其对角线元素。

3.21 旋转矩阵

二维空间中的旋转矩阵表示绕原点逆时针旋转角度 θ 的线性变换，定义为

$$R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

4 矩阵关系

4.1 相似

设 $A, B \in \mathbb{R}^{n \times n}$ ，若存在可逆矩阵 P 使得

$$B = P^{-1}AP,$$

则称 A 与 B **相似**。

相似矩阵具有以下性质：

- 相似关系为等价关系：自反性、对称性、传递性均成立。
- 相似矩阵具有相同的特征值（含重数）、行列式、迹、秩等不变量。如果是酉相似，则矩阵拥有相同的奇异值。
- 相似矩阵的特征向量通过变换矩阵 P 相互关联。
- 相似变换常用于矩阵的对角化、Jordan 标准形等分解形式。

4.2 合同

设 $A, B \in \mathbb{R}^{n \times n}$ 为对称矩阵，若存在可逆矩阵 P 使得

$$B = P^TAP,$$

则称 A 与 B **合同**。

合同矩阵具有以下性质：

- 合同关系为等价关系：自反性 ($A = I^T A I$)、对称性、传递性均成立。
- 合同矩阵具有相同的惯性指数（正惯性指数、负惯性指数、零惯性指数）。
- 合同变换常用于二次型的规范形、矩阵的正定性分析等。
- 合同变换不改变特征值的符号。

5 运算

5.1 直和

给定 $A \in \mathbb{R}^{m \times n}$ 、 $B \in \mathbb{R}^{p \times q}$, 其直和 (direct sum) 定义为

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \in \mathbb{R}^{(m+p) \times (n+q)}.$$

性质:

- 交换、结合: $A \oplus B = B \oplus A$, $(A \oplus B) \oplus C = A \oplus (B \oplus C)$ 。
- 分配律: $(A \oplus B)(C \oplus D) = AC \oplus BD$, $(A \oplus B) + (C \oplus D) = (A + C) \oplus (B + D)$ 。
- 逆与转置: 若 A, B 可逆, 则 $(A \oplus B)^{-1} = A^{-1} \oplus B^{-1}$; $(A \oplus B)^T = A^T \oplus B^T$ 。
- 尺寸相加: $\dim(A \oplus B) = (m+p) \times (n+q)$ 。
- 秩相加: $\text{rank}(A \oplus B) = \text{rank}(A) + \text{rank}(B)$ 。
- 行列式与迹: 若 A, B 均为方阵, 则 $\det(A \oplus B) = \det(A) \det(B)$, $\text{tr}(A \oplus B) = \text{tr}(A) + \text{tr}(B)$ 。
- 特征值为并集: 若 A, B 方阵, 则 $A \oplus B$ 的特征值集合为 A 与 B 特征值的并集 (重数相加)。

5.2 Hardmard 积

给定同型矩阵 $A, B \in \mathbb{R}^{m \times n}$, Hardmard 积 (Schur 乘积) 逐元素相乘:

$$(A * B)_{ij} = A_{ij} B_{ij}.$$

性质:

- 交换、结合、对加法分配: $A * B = B * A$, $(A * B) * C = A * (B * C)$, $(A + B) * C = A * C + B * C$ 。
- 如果 A, B, D 都是 $m \times m$ 矩阵, 并且 D 是对角阵, 则 $(DA) * (BD) = D(A * B)D$ (HW1)。
- 与 Frobenius 内积的关系: $\langle A, B \rangle_F = \text{tr}(A^T B) = \mathbf{1}^T (A * B) \mathbf{1}$, 其中 $\mathbf{1}$ 为全 1 向量。
- $\text{rank}(A * B) \leq \text{rank}(A) \text{rank}(B)$ (HW1)。
- Schur 乘积定理: 若 A, B 为对称半正定矩阵, 则 $A * B$ 亦为半正定矩阵 (HW8)。

5.3 Kronecker 积

给定 $A \in \mathbb{R}^{m \times n}$ 、 $B \in \mathbb{R}^{p \times q}$, 其 Kronecker 积定义为

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{(mp) \times (nq)}.$$

常用性质 (假设尺寸可乘):

- 结合律: $(A \otimes B) \otimes C = A \otimes (B \otimes C)$ (HW1)。
- 分配律: $(A + B) \otimes C = A \otimes C + B \otimes C$, $A \otimes (B + C) = A \otimes B + A \otimes C$ 。
- 混合乘法: $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ 。
- 转置与逆: $(A \otimes B)^T = A^T \otimes B^T$; 若 A, B 可逆, 则 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ 。
- 秩、迹、行列式: $\text{rank}(A \otimes B) = \text{rank}(A) \text{rank}(B)$ (HW1); 若 A, B 方阵, $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$, $\det(A \otimes B) = \det(A)^p \det(B)^m$ 。
- 特征值: 若 λ 为 A 的特征值、 μ 为 B 的特征值, 则 $\lambda\mu$ 为 $A \otimes B$ 的特征值。
- 向量化恒等式: $\text{vec}(AXB) = (B^T \otimes A) \text{vec}(X)$, 便于将矩阵方程转化为线性方程组。

5.4 迹

给定方阵 $A \in \mathbb{R}^{n \times n}$, 其迹定义为主对角线元素之和:

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}.$$

性质:

- 线性: $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$, $\text{tr}(kA) = k \text{tr}(A)$ 。
- 循环不变性: $\text{tr}(AB) = \text{tr}(BA)$, 但一般 $\text{tr}(AB) \neq \text{tr}(A) \text{tr}(B)$ 。
- 与特征值的关系: $\text{tr}(A) = \sum_{i=1}^n \lambda_i$, 其中 λ_i 为 A 的特征值 (含重数)。
- 与 Frobenius 内积的关系: $\langle A, B \rangle_F = \text{tr}(A^T B)$ 。
- $\text{tr}(A^2) \leq \text{tr}(A^H A)$, 等号当且仅当 A 为 hermitian 矩阵时成立。
- $\text{tr}(AB) = \text{vec}(B)^T \text{vec}(A^T)$, $\text{tr}(A^T) = \text{tr}(A)$
- $\text{tr}(ABCD) = \text{vec}(D)^T (A \otimes C^T) \text{vec}(B^T)$ (HW2)

5.5 伪逆 (Moore-Penrose 逆)

设 $A \in \mathbb{R}^{m \times n}$, 其 Moore-Penrose 伪逆 $A^+ \in \mathbb{R}^{n \times m}$ 满足以下四个条件:

- $AA^+A = A$,
- $A^+AA^+ = A^+$,
- $(AA^+)^T = AA^+$,
- $(A^+A)^T = A^+A$ 。

伪逆是唯一的, 计算方法:

- 一般情形下，可通过 SVD 分解 $A = U\Sigma V^T$ ，则 $A^+ = V\Sigma^+U^T$ ，其中 Σ^+ 通过将 Σ 中非零奇异值取倒数并转置得到。
- 若 A 为满列秩矩阵，则左伪逆 $A^+ = (A^T A)^{-1} A^T$ ，对应最小二乘解 $\min_x \|Ax - b\|_2$ 。
- 若 A 为满行秩矩阵，则右伪逆 $A^+ = A^T (AA^T)^{-1}$ ，对应最小范数解 $\min_x \|x\|_2$ ，使 $Ax = b$ 。
- 秩分解得 $A = BC$ ，则 $A^+ = C^+B^+$

性质：

- 若 A 可逆，则 $A^+ = A^{-1}$ 。
- $(A^T)^+ = (A^+)^T$ 。
- $A^+ = (A^T A)^+ A^T = A^T (AA^T)^+, A^T = A^T AA^+, \text{Im}(A^+) = \text{Row}(A)$ 。
- $(kA)^+ = \frac{1}{k}A^+$ ，其中 $k \neq 0$ 。
- $(AB)^+ \neq B^+A^+$ ，但若 A 为满列秩且 B 为满行秩，则 $(AB)^+ = B^+A^+$ 。
- AA^+ 为 A 的列空间的正交投影矩阵， A^+A 为 A^T 的列空间的正交投影矩阵。
- 零向量的伪逆为零： $\mathbf{0}^+ = \mathbf{0}$ 。
- $\text{Col}(A^+) = \text{Row}(A)$, $\text{Row}(A^+) = \text{Col}(A)$ 。
- 令 $x_0 = A^+b$ ，则 $\|Ax_0 - y\|_2 \leq \|Ax - y\|_2$ ，等号成立当且仅当 $A(x - x_0) = 0$ ；且在所有等号成立的解中， x_0 的 2 范数唯一最小 (HW3)。

5.6 行列式与可逆性

给定方阵 $A \in \mathbb{R}^{n \times n}$ ，其行列式定义为

$$|A| = \det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)},$$

其中 S_n 为所有 n 阶排列的集合， $\text{sgn}(\sigma)$ 为排列 σ 的符号。

伴随矩阵对行列式的计算也很有用，定义为

$$\text{adj}(A) = [(-1)^{i+j} \det(M_{ji})]_{n \times n},$$

其中 M_{ji} 为去掉 A 的第 j 行第 i 列后的 $(n-1) \times (n-1)$ 子矩阵。则有

$$A \cdot \text{adj}(A) = \text{adj}(A) \cdot A = \det(A)I.$$

性质：

- 线性： $\det(kA) = k^n \det(A)$ 。
- 乘积： $\det(AB) = \det(A) \det(B)$ ，假设 A, B 都为方阵。

- 可交换: $\det(AB) = \det(BA)$ 。
- 逆与转置: 若 A 可逆, 则 $\det(A^{-1}) = 1/\det(A)$; $\det(A^T) = \det(A)$ 。
- 与特征值的关系: $\det(A) = \prod_{i=1}^n \lambda_i$, 其中 λ_i 为 A 的特征值 (含重数)。
- 与奇异值的关系: $|\det(A)| = \prod_i \sigma_i$
- 初等变换对行列式的影响:
 - 交换两行 (列): 行列式变号。
 - 某行 (列) 乘以非零常数 k : 行列式乘以 k 。
 - 某行 (列) 加上另一行 (列) 的倍数: 行列式不变。
- $\det(A^H) = \text{conj}(\det(A))$, 其中 A^H 为 A 的共轭转置。
- 对于 ABCD 型分块矩阵, 若 A 可逆, 则

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A) \det(D - CA^{-1}B) = \det(D) \det(A - BD^{-1}C).$$

- 当 A, B 列正交时, $\det(A^H B) \leq 1$ 。因为 AA^H 是正交投影矩阵, $\|A^H B\|_2 \leq 1$ 。
- $|\det(A^H B)|^2 \leq \det(A^H A) \det(B^H B)$, 等号当且仅当 A 与 B 线性相关时成立, 即存在酉矩阵 Q 使得 $A = BQ$ (HW5, QR 分解)。
- 正定矩阵的行列式为正, 半正定矩阵的行列式非负。
- Hardmard 不等式: 若 A 为半正定矩阵, 则 (HW4)

$$\det(A) \leq \prod_{i=1}^n a_{ii},$$

等号当且仅当 A 为对角矩阵时成立 (用数学归纳法和初等变换证, 也可用 QR 分解 $|A| < \prod_i |r_{ii}| \leq \prod_i \|r_i\|_2 = \prod_i \|a_i\|_2$, 还可用 LDL 分解证) (行列式题灵活使用分解相似合同)。

- 若 A, B 均为半正定矩阵, 则

$$\det(A + B)^{1/m} \geq \det(A)^{1/m} + \det(B)^{1/m}.$$

证明见图 1 或 HW2, 若 $m=1$ 则可用 $A^{1/2}$ 来证。

- $\log \det X$ 当 X 正定为凸函数, 即 $\det(\theta X + (1-\theta)Y) \geq (\det X)^\theta (\det Y)^{1-\theta}$, 拆分 $X^{1/2}$ 来证; 也可求导证: $g'(t) = \text{tr}((X + tV)^{-1}V)$, $g''(t) = -\text{tr}((X + tV)^{-1}V(X + tV)^{-1}V)$
- 严格对角占优矩阵 (矩阵的每一行主对角线元素的模都大于该行非主对角线元素模之和) 是可逆的。

(2). A, B 中存在正定阵，不妨设为 A

A 为同单位阵 I ($A \succeq I$)

$$\Rightarrow \exists C, 可逆 \quad C^H A C = I \Rightarrow A = C^H C^{-1}$$

$$\text{def: } B' = C^H B C \Rightarrow B = C^{-H} B' C^{-1}$$

$$1) \det(A+B) = \det(C^H C^{-1} + C^H B' C^{-1}) = \det(C^{-1})^2 \det(I+B')$$

$$2) \det(A) = \det(C^{-1})^2$$

$$3) \det(B) = \det(C')^2 \cdot \det(B')$$

$$\Rightarrow \text{原不等式} \Leftrightarrow \det(I+B')^{\frac{1}{m}} \geq 1 + \det(B')^{\frac{1}{m}}$$

$B' = C^H B C$ 半正定，设其特征值 $\lambda_1, \dots, \lambda_m$

$$\Leftrightarrow \det(I+B') = \det(P P^T \oplus \cancel{P \sum_i P^T})$$

$$= \det(I + \sum_{i=1}^m \lambda_i) = \prod_{i=1}^m (1 + \lambda_i)$$

$$\text{原式} \Leftrightarrow \prod_{i=1}^m (1 + \lambda_i)^{\frac{1}{m}} \geq 1 + \left(\prod_{i=1}^m \lambda_i \right)^{\frac{1}{m}} \#.$$

且 $f(t) = \ln(1+e^t)$ 凸性 \Rightarrow Jensen

或 $(LHS)^m \leq (RHS)^m$ = 原展开，按入阶数逐阶用基本不等式

图 1: 行列式不等式证明

5.7 矩阵指数

给定方阵 $A \in \mathbb{R}^{n \times n}$, 其矩阵指数定义为

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

由 Cayley-Hamilton 定理可以计算矩阵指数, 若 A 的特征多项式为 $p(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$, 则存在唯一的多项式 $q(\lambda)$, 使得

$$e^{At} = \sum_{k=0}^{\infty} \frac{(At)^k}{k!} = q(At)$$

且 $\deg(q) < n$ 。

存在等式: $e^{\lambda_i t} = q(\lambda_i t)$, $\frac{d^k}{dt^k} e^{\lambda_i t} = \frac{d^k}{dt^k} q(\lambda_i t)$, 其中 λ_i 为 A 的特征值。使用待定系数法即可求出 $q(\lambda)$ 的系数。

例题:

(3) 由 教案-哈密顿定理的推论知

$$e^{At} = \sum_{k=0}^{\infty} a_k(t) A^k = a_0(t)I + a_1(t)A + a_2(t)A^2$$

其中代入 特征值求 $a_0(t)$, $a_1(t)$, $a_2(t)$ 。有: $\begin{cases} e^t = a_0(t) + a_1(t) + a_2(t) \\ e^{2t} = a_0(t) + 2a_1(t) + 4a_2(t) \\ te^{2t} = a_1(t) + 4a_2(t) \end{cases}$

解得: $\begin{cases} a_0(t) = 2te^{2t} - 3e^{2t} + 4e^t \\ a_1(t) = -3te^{2t} + 4e^{2t} - 4e^t \\ a_2(t) = te^{2t} - e^{2t} + e^t \end{cases}$

$$\therefore A^2 = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 1 & -1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 8 & 4 & -4 \\ 3 & 4 & -3 \\ 7 & 4 & -3 \end{bmatrix},$$

$$\therefore e^{At} = a_0 I + a_1 A + a_2 A^2$$

$$= \begin{bmatrix} a_0 + 3a_1 + 8a_2 & a_1 + 4a_2 & -a_1 - 4a_2 \\ a_1 + 3a_2 & a_0 + 2a_1 + 4a_2 & -a_1 - 3a_2 \\ 2a_1 + 7a_2 & a_1 + 4a_2 & a_0 - 3a_2 \end{bmatrix}$$

$$= \begin{bmatrix} te^{2t} + e^{2t} & te^{2t} & -te^{2t} \\ e^{2t} - e^t & e^{2t} - e^t & -e^{2t} + e^t \\ te^{2t} + e^{2t} - et & e^{2t} & -te^{2t} + e^t \end{bmatrix}$$

图 2: 矩阵指数例题

6 定理

6.1 Cayley-Hamilton 定理与特征/最小多项式

设 $A \in \mathbb{R}^{n \times n}$, 其特征多项式为

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

则 Cayley-Hamilton 定理断言, 矩阵 A 满足其特征多项式 (schur 分解证) (HW7):

$$p(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0I = 0.$$

性质:

- 特征多项式的根即为 A 的所有特征值 (含重数)。
- $c_0 = (-1)^n \det(A)$, $c_{n-1} = -\text{tr}(A)$ 。
- 若 A 可逆, 则 $A^{-1} = -\frac{1}{c_0}(A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1I)$ 。

此外, 零化多项式 (annihilating polynomial) 是指满足 $q(A) = 0$ 的非零多项式 $q(\lambda)$ 。Cayley-Hamilton 定理表明, 特征多项式 $p(\lambda)$ 是 A 的一个零化多项式, 性质:

- 最小多项式 $m(\lambda)$ 是所有零化多项式中次数最低的一个, 且 $m(\lambda)$ 整除 $p(\lambda)$ 。
- 若 A 可对角化, 则 $m(\lambda)$ 的根即为 A 的所有不同特征值。
- 若 A 的特征值均不相等, 则 $m(\lambda) = p(\lambda)$ 。
- 每个特征值都是最小多项式的根。
- 若 A 的 Jordan 标准形中, 某特征值对应的最大 Jordan 块的大小为 k , 则该特征值在最小多项式中的重数为 k 。

6.2 Frobenius 不等式及常见秩不等式

设 $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, 则 **Frobenius 秩不等式** 为

$$\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB) \leq \min \{ \text{rank}(A), \text{rank}(B) \}.$$

常见的秩不等式还包括:

- 对任意可加同型映射 A, B , 有

$$\text{rank}(A + B) \leq \text{rank}([A \ B]) \leq \text{rank}(A) + \text{rank}(B).$$

- 若 A 可逆, 则 $\text{rank}(AB) = \text{rank}(B)$, $\text{rank}(BA) = \text{rank}(B)$ 。
- $\text{rank}(A + B) \leq \text{rank}([A \ B]) \leq \text{rank}(A) + \text{rank}(B)$
- 伴随矩阵 $\text{adj}(A)$ 满足

$$\text{rank}(\text{adj}(A)) = \begin{cases} n, & \text{若 } \text{rank}(A) = n; \\ 1, & \text{若 } \text{rank}(A) = n - 1; \\ 0, & \text{若 } \text{rank}(A) < n - 1. \end{cases}$$

- 设 $X \in \mathbb{R}^{m \times p}$ 为满列秩矩阵, $Y \in \mathbb{R}^{q \times n}$ 为满行秩矩阵, 则

$$\text{rank}(A) = \text{rank}(XAY).$$

- $\text{rank}(A) = \text{rank}(A^T A) = \text{rank}(AA^T)$ 。

6.3 秩—零化度定理

设线性映射 $T : V \rightarrow W$, 其矩阵表示为 A 。当 V 是有限维线性空间 ($\dim V < \infty$) 时, 有著名的 **秩—零化度定理**:

$$\dim \text{Im}(T) + \dim \ker(T) = \dim V.$$

其中 $\ker(T) \equiv \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$, 注意 T 的核空间是属于 V 的子空间。

或者以矩阵形式表示为

$$\text{rank}(A) + \text{nullity}(A) = n,$$

其中 $A \in \mathbb{R}^{m \times n}$, n 为列数。

在线性方程组 $A\mathbf{x} = \mathbf{b}$ 的语境下, $\text{nullity}(A)$ 给出齐次解空间的维数, 而 $\text{rank}(A)$ 控制非齐次方程是否有解及解的维数。

6.4 维数定理

设 V_1, V_2 为线性空间 V 的子空间, 则有以下**维数定理**:

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2)$$

6.5 Rayleigh 商与极值

设 $A \in \mathbb{R}^{n \times n}$ 为对称矩阵, $\mathbf{x} \in \mathbb{R}^n$ 为非零向量。定义 **Rayleigh 商** 为

$$R(\mathbf{x}, A) = \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}.$$

Rayleigh 商的极值与 A 的特征值密切相关:

- 最大特征值 λ_{\max} 满足

$$\lambda_{\max} = \max_{\mathbf{x} \neq 0} R(A, \mathbf{x}).$$

- 最小特征值 λ_{\min} 满足

$$\lambda_{\min} = \min_{\mathbf{x} \neq 0} R(A, \mathbf{x}).$$

如果令特征向量空间 $U_m = \text{span}\{u_1, \dots, u_m\}$, $V_m = \text{span}\{u_{N-m+1}, \dots, u_N\}$, 则 (HW6/8)

$$\lambda_k = \min_{\mathbf{x} \in V_{k-1}^\perp, \mathbf{x} \neq 0} R(A, \mathbf{x}) = \max_{\mathbf{x} \in U_k, \mathbf{x} \neq 0} R(A, \mathbf{x})$$

$$\lambda_{N-k+1} = \max_{\mathbf{x} \in U_{k-1}^\perp, \mathbf{x} \neq 0} R(A, \mathbf{x}) = \min_{\mathbf{x} \in V_k, \mathbf{x} \neq 0} R(A, \mathbf{x})$$

其他性质:

- $R(\alpha x, \beta A) = \beta R(x, A)$
- $R(x, A - \alpha I) = R(x, A) - \alpha$
- $x \perp (A - R(x, A)I)x$
- $\|(A - R(x, A)I)x\| = \|(A - \mu I)x\|$
- 设 $W(A) = \{R(x, A) : x \neq 0\}$, 则 $W(A+B) \subseteq W(A) + W(B)$, $W(A+\beta I) = W(A) + \beta$, $W(\alpha A) = \alpha W(A)$

6.6 Gersgrin 圆

设 $A = [a_{ij}] \in \mathbb{R}^{n \times n}$, 定义第 i 行的 Gersgrin 圆心与半径为

$$C_i = a_{ii}, \quad R_i = \sum_{j \neq i} |a_{ij}|.$$

则 A 的所有特征值均位于这些 Gersgrin 圆的并集中, 即

$$\lambda \in \bigcup_{i=1}^n \{z \in \mathbb{C} : |z - C_i| \leq R_i\}.$$

6.7 Courant-Fischer 极小极大定理

设 $A \in \mathbb{R}^{n \times n}$ 为对称矩阵, 按非递增顺序排列其特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 。则第 k 个特征值可通过以下极小极大原理表述 (HW8):

$$\lambda_k = \min_{\substack{S \subset \mathbb{R}^n \\ \dim(S)=n-k+1}} \max_{\substack{x \in S \\ x \neq 0}} R(A, x) = \max_{\substack{S \subset \mathbb{R}^n \\ \dim(S)=k}} \min_{\substack{x \in S \\ x \neq 0}} R(A, x)$$

$$\lambda_{N-k+1} = \max_{\substack{S \subset \mathbb{R}^n \\ \dim(S)=n-k+1}} \min_{\substack{x \in S \\ x \neq 0}} R(A, x) = \min_{\substack{S \subset \mathbb{R}^n \\ \dim(S)=k}} \max_{\substack{x \in S \\ x \neq 0}} R(A, x).$$

6.8 Weyl 定理

设 $A, B \in \mathbb{R}^{n \times n}$ 为对称矩阵, 按非递增顺序排列其特征值分别为 $\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_n(A)$ 和 $\lambda_1(B) \geq \lambda_2(B) \geq \dots \geq \lambda_n(B)$ 。则 $A+B$ 的特征值 $\lambda_k(A+B)$ 满足以下不等式 (HW6):

$$\lambda_i(A) + \lambda_j(B) \geq \lambda_{i+j-1}(A+B), \quad \text{若 } i+j-1 \leq n,$$

$$\lambda_i(A) + \lambda_j(B) \leq \lambda_{i+j-n}(A+B), \quad \text{若 } i+j-n \geq 1.$$

推论 (HW6):

$$\lambda_k(A) + \lambda_n(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_1(B), \quad \text{其中 } 1 \leq k \leq n.$$

技巧: 最大特征值和最小特征值的关系往往可用倒数联系。

6.9 Cauchy 交错定理

设 $A \in \mathbb{R}^{n \times n}$ 为对称矩阵, 按非递增顺序排列其特征值为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 。设 $B \in \mathbb{R}^{(n-1) \times (n-1)}$ 为 A 的任一主子矩阵, 按非递增顺序排列其特征值为 $\mu_1 \geq \mu_2 \geq \dots \geq \mu_{n-1}$ 。则有以下交错不等式 (HW8):

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \dots \geq \mu_{n-1} \geq \lambda_n.$$

更一般地, 设 $B \in \mathbb{R}^{k \times k}$ 为 A 的任一主子矩阵, 按非递增顺序排列其特征值为 $\mu_1 \geq \mu_2 \geq \dots \geq \mu_k$ 。则有以下交错不等式:

$$\lambda_i \geq \mu_i \geq \lambda_{n-k+i}, \quad \text{其中 } i = 1, 2, \dots, k.$$

可从 Courant-Fischer 极小极大定理出发证明。

6.10 RIP

设测量矩阵 $A \in \mathbb{R}^{m \times n}$, 则 A 满足**限制等距性质** (Restricted Isometry Property, RIP) 如果存在常数 $\delta_S \in (0, 1)$ 使得对于所有 S -稀疏向量 $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_0 \leq S$, 都有

$$(1 - \delta_S)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_S)\|\mathbf{x}\|_2^2.$$

通过求解 L_1 优化问题可以重构稀疏信号:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1, \quad \text{s.t. } A\mathbf{x} = \mathbf{b}.$$

, 需满足以下任一条件:

- $\delta_S + \delta_{2S} + \delta_{3S} < 1$
- $\delta_{2S} < \sqrt{2} - 1$
- $\delta_S < 0.307$

6.11 Schur 补定理

设矩阵 M 分块表示为

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

其中 $A \in \mathbb{R}^{k \times k}$, $D \in \mathbb{R}^{(n-k) \times (n-k)}$ 。则称

$$S = D - CA^{-1}B$$

为 M 关于 A 的 **Schur 补**。

若 A 可逆, 则以下命题等价:

- M 为正定矩阵;
- A 为正定矩阵, 且 S 为正定矩阵;
- D 为正定矩阵, 且 $A - BD^{-1}C$ 为正定矩阵。

6.12 Sylvester 方程

设 $A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{m \times n}$ 。则 **Sylvester 方程**

$$AX - XB = C$$

存在唯一解 $X \in \mathbb{R}^{m \times n}$ 的充分必要条件为 A 和 B 没有共同的特征值。

简短证明:

- $A^k X = X B^k$, $p(A)X = X p(B)$, 其中 $p(\lambda)$ 为 A 的特征多项式, 得到 $p(A) = 0$, 则 $p(A)X = X p(B) = 0$ 。

- 若 A 和 B 没有共同特征值，则 $p(B)$ 可逆，得到 $X = 0$ ，则齐次方程只有零解，非齐次方程有唯一解。
- 反之，若 A 和 B 有共同特征值 λ ，则存在非零向量 u, v 使得 $Au = \lambda u, Bv = \lambda v$ ，则 $X = uv^T$ 为齐次方程的非零解，非齐次方程无唯一解。

6.13 Weinstain-Aronszajn 恒等式

设 $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times m}$, 则有以下恒等式:

$$\det(I_m + AB) = \det(I_n + BA).$$

简短证明: 构造分块矩阵

$$M = \begin{bmatrix} I_m & A \\ -B & I_n \end{bmatrix}.$$

, 两种初等变换即证。

AB 和 BA 的非零特征值和代数重数相同, 证明可令 $A = A, B = -B/\lambda$, 得 $p_{AB}(\lambda) = \lambda^{m-n} p_{BA}$ 。非零特征值的几何重数也相同: $BA(BX) = B(AX) = \lambda(BX), BX \neq 0$ 。由上可知, 零特征值的代数重数相差 $m - n$ 。

7 研究

7.1 方程有无解

设线性方程组 $Ax = b$, 其中 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ 。则该方程组:

- 有解的充分必要条件为 $\text{rank}(A) = \text{rank}([A \ b])$ 。
- 唯一解的充分必要条件为 $\text{rank}(A) = \text{rank}([A \ b]) = n$ 。
- 无穷多解的充分必要条件为 $\text{rank}(A) = \text{rank}([A \ b]) < n$ 。
- 无解的充分必要条件为 $\text{rank}(A) < \text{rank}([A \ b])$ 。

7.2 0-1 矩阵满秩的概率

随机生成的 0-1 矩阵 (各元素独立同分布) 满秩的概率:

$$P(\text{满秩}) = \prod_{i=0}^{n-1} (1 - 2^{i-n})$$

其中 n 为矩阵的行数或列数 (假设为方阵)。当 n 较大时, 满秩概率趋近于 1。

7.3 矩阵的条件数

对于 $Ax = b$ 线性方程组，条件数衡量输入扰动对解的影响程度：

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \kappa(A) \frac{\|\delta A\|}{\|A\|},$$

其中 $\kappa(A)$ 为矩阵 A 的条件数，定义为

$$\kappa(A) = \|A\| \|A^{-1}\|.$$

也等于最大奇异值与最小奇异值之比。

7.4 特征值与特征向量

设 $A \in \mathbb{R}^{n \times n}$ ，若存在非零向量 $x \in \mathbb{R}^n$ 和标量 $\lambda \in \mathbb{R}$ 使得

$$Ax = \lambda x,$$

则称 λ 为 A 的特征值， x 为对应的特征向量。

也存在左特征值与左特征向量的概念，满足

$$y^T A = \lambda y^T,$$

其中 $y \in \mathbb{R}^n$ 为非零向量。

绝对值最大的特征值称为谱半径，记为 $\rho(A) = \max_i |\lambda_i|$ 。

性质：

- 特征值可通过求解特征多项式 $\det(\lambda I - A) = 0$ 得到。
- 不同特征值对应的特征向量线性无关。
- 若 A 可对角化，则存在可逆矩阵 P 使得 $A = P\Lambda P^{-1}$ ，其中 Λ 为对角矩阵，其对角线元素为 A 的特征值。相似也不改变矩阵特征值。
- 特征值与矩阵的迹和行列式的关系： $\text{tr}(A) = \sum_{i=1}^n \lambda_i$, $\det(A) = \prod_{i=1}^n \lambda_i$ 。
- 若 A 为实对称矩阵，则其特征值均为实数，且存在正交矩阵 Q 使得 $A = Q\Lambda Q^T$ 。
- 左特征值与右特征值相同，左特征向量与右特征向量不一定相同。
- AB 与 BA 具有相同的非零特征值（含重数），证明见 6.13 Weinstein-Aronszajn 恒等式。

7.5 代数重数与几何重数

设 $A \in \mathbb{R}^{n \times n}$, λ 为 A 的特征值。则：

- λ 的代数重数 (algebraic multiplicity) 定义为 λ 作为特征多项式 $p(\lambda) = \det(\lambda I - A)$ 的根的重数，记为 $m_a(\lambda)$ 。

- λ 的几何重数 (geometric multiplicity) 定义为对应特征空间的维数, 即 $\ker(A - \lambda I)$ 的维数, 记为 $m_g(\lambda)$ 。

性质:

- 对任一特征值 λ , 都有 $1 \leq m_g(\lambda) \leq m_a(\lambda)$ 。
- 矩阵 A 可对角化或其特征向量组成一组基的充分必要条件是对每个特征值 λ , 均有 $m_g(\lambda) = m_a(\lambda)$ 。
- 代数重数反映了特征值在特征多项式中的出现频率, 而几何重数反映了对应特征向量的线性无关数量。

7.6 $AB = BA$

设 $A, B \in \mathbb{R}^{n \times n}$, 若 $AB = BA$, 则称 A 与 B 可交换。可交换矩阵具有以下性质:

- 共同特征向量: 若 A 可对角化且 $AB = BA$, 则 A 与 B 具有一组共同的特征向量和特征值。
- 多项式闭包: 若 A 与 B 可交换, 则任意多项式 $p(A, B)$ 与 $q(A, B)$ 也可交换, 即 $p(A, B)q(A, B) = q(A, B)p(A, B)$ 。
- 对称矩阵的可交换性: 若 A 与 B 均为对称矩阵且可交换, 则它们可以被同一个正交矩阵对角化。
- 迹的可交换性: $\text{tr}(AB) = \text{tr}(BA)$, 因此可交换矩阵的迹具有对称性。

7.7 行列式因子、不变因子和初等因子

设 $A \in \mathbb{R}^{m \times n}$, 则其行列式因子 (determinantal divisors) 定义为

$$d_k(A) = \gcd\{\det(M) : M \text{ 为 } A \text{ 的 } k \times k \text{ 子矩阵}\},$$

对于 $k = 1, 2, \dots, r$, 其中 $r = \text{rank}(A)$, \gcd 表示最大公约数。行列式因子满足 $d_1(A) | d_2(A) | \dots | d_r(A)$ 。

设 $s_1(A) = d_1(A)$, $s_k(A) = d_k(A)/d_{k-1}(A)$, 则称 $s_k(A)$ 为 A 的不变因子 (invariant factors)。不变因子满足 $s_1(A) | s_2(A) | \dots | s_r(A)$ 。

进一步, 将所有不变因子分解为不可约多项式的幂次乘积, 称这些幂次项为 A 的初等因子 (elementary divisors)。

7.8 施密特正交化

设 V 为内积空间, $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ 为 V 中的一组线性无关向量。通过施密特正交化过程, 可以构造出一组正交 (或正交归一化) 的向量 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, 其生成的子空间与原向量组相同。具体步骤如下:

1. 设 $\mathbf{u}_1 = \mathbf{v}_1$ 。

2. 对于 $k = 2, 3, \dots, n$, 定义

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \frac{\langle \mathbf{v}_k, \mathbf{u}_j \rangle}{\langle \mathbf{u}_j, \mathbf{u}_j \rangle} \mathbf{u}_j.$$

7.9 核/零空间

设线性映射 $T : V \rightarrow W$, 其矩阵表示为 A 。则 T 的核 (kernel) 或零空间 (null space) 定义为

$$\ker(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\} = \{\mathbf{v} \in V : A\mathbf{v} = \mathbf{0}\}.$$

性质:

- $\ker(T)$ 是 V 的子空间。
- 矩阵 A 的零空间 $\ker(A)$ 与 $\text{Row}(A)^\perp$ 相同。
- $\dim(\ker(T))$ 称为 T 的零化度 (nullity)。
- 秩-零化度定理: $\dim(\text{Im}(T)) + \dim(\ker(T)) = \dim(V)$ 。
- 若 A 可逆, 则 $\ker(T) = \{0\}$ 。

8 范数

8.1 向量范数

设 V 为数域 P 上的向量空间, 映射 $\|\cdot\| : V \rightarrow \mathbb{R}$ 若满足下列条件, 则称其为 V 上的范数:

1. 正定性: 对任意 $\alpha \in V$, 有 $\|\alpha\| \geq 0$, 且当且仅当 $\alpha = 0$ 时取等号。
2. 齐次性: 对任意 $\alpha \in V$ 与 $k \in P$, 有 $\|k\alpha\| = |k|\|\alpha\|$ 。
3. 三角不等式: 对任意 $\alpha, \beta \in V$, 有 $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$ 。

常见向量范数包括:

- p -范数: 对 $p \geq 1$, 定义

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

特殊情形包括 $p = 1$ (曼哈顿范数)、 $p = 2$ (欧几里得范数)、 $p = \infty$ (最大范数)。

- 最大范数: 定义

$$\|\mathbf{x}\|_{\max} = \max_{1 \leq i \leq n} |x_i|.$$

酉矩阵或列正交矩阵不改变向量的 2-范数: $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2, \forall Q^T Q = I$ 。

8.2 向量内积

设 V 为数域 P 上的向量空间, 定义映射 $\langle \cdot, \cdot \rangle : V \times V \rightarrow P$, 若对任意 $\alpha, \beta, \gamma \in V$ 与任意 $k \in P$, 下列条件都成立, 则称 $\langle \cdot, \cdot \rangle$ 为 V 上的内积, $(V, \langle \cdot, \cdot \rangle)$ 为内积空间:

1. 正定性: $\langle \alpha, \alpha \rangle \geq 0$, 且当且仅当 $\alpha = 0$ 时取等号。
2. 共轭对称性: $\langle \alpha, \beta \rangle = \overline{\langle \beta, \alpha \rangle}$ 。
3. 线性性: $\langle k\alpha + \beta, \gamma \rangle = k\langle \alpha, \gamma \rangle + \langle \beta, \gamma \rangle$ 。

常见性质:

- **柯西-施瓦茨不等式:** 对任意 $\alpha, \beta \in V$, 有

$$|\langle \alpha, \beta \rangle|^2 \leq \langle \alpha, \alpha \rangle \langle \beta, \beta \rangle.$$

- **三角不等式:** 对任意 $\alpha, \beta \in V$, 有

$$\sqrt{\langle \alpha + \beta, \alpha + \beta \rangle} \leq \sqrt{\langle \alpha, \alpha \rangle} + \sqrt{\langle \beta, \beta \rangle}.$$

- **平行四边形恒等式:** 对任意 $\alpha, \beta \in V$, 有

$$\langle \alpha + \beta, \alpha + \beta \rangle + \langle \alpha - \beta, \alpha - \beta \rangle = 2\langle \alpha, \alpha \rangle + 2\langle \beta, \beta \rangle.$$

内积还可以写成范数的形式:

$$\langle \alpha, \beta \rangle = \frac{1}{4}(\|\alpha + \beta\|^2 - \|\alpha - \beta\|^2).$$

相关(不)等式证明见 HW2。

8.3 矩阵范数

设 V 为数域 P 上的矩阵空间, 映射 $\|\cdot\| : V \rightarrow \mathbb{R}$ 若满足下列条件, 则称其为 V 上的矩阵范数:

1. 正定性: 对任意 $A \in V$, 有 $\|A\| \geq 0$, 且当且仅当 $A = 0$ 时取等号。
2. 齐次性: 对任意 $A \in V$ 与 $k \in P$, 有 $\|kA\| = |k|\|A\|$ 。
3. 三角不等式: 对任意 $A, B \in V$, 有 $\|A + B\| \leq \|A\| + \|B\|$ 。
4. 一致性: 对任意可乘矩阵 $A, B \in V$, 有 $\|AB\| \leq \|A\|\|B\|$ 。

常见矩阵范数包括:

- p -范数: 对 $p \geq 1$, 定义

$$\|A\|_p = \sup_{\mathbf{x} \neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}.$$

特殊情形包括: $\|A\|_0 = \max_{1 \leq j \leq n} \|a_n\|_0$ (列 0 范数)、 $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$ (列和范数)、 $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$ (行和范数)、 $\|A\|_2 = \sigma_{\max}$ (谱范数, σ_{\max} 为 A 的最大奇异值)。(HW2)

- Frobenius 范数: 定义

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{tr}(A^T A)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}.$$

- 核范数: 定义

$$\|A\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i,$$

其中 σ_i 为 A 的奇异值。

性质:

- 谱半径小于任意矩阵范数, 即 $\rho(A) \leq \|A\|$ 。
- $\|A + B\|^2 + \|A - B\|^2 = 2\|A\|^2 + 2\|B\|^2$
- $\|A + B\| \cdot \|A - B\| \leq \|A\|^2 + \|B\|^2$
- $\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$

8.4 矩阵内积

表达式: $\langle A, B \rangle = \text{tr}(A^T B) = \text{vec}(A)^T \text{vec}(B)$, 性质:

- 正定性: $\langle A, A \rangle \geq 0$, 且当且仅当 $A = 0$ 时取等号。
- 对称性: $\langle A, B \rangle = \langle B, A \rangle$ 。
- 线性性: $\langle kA + B, C \rangle = k\langle A, C \rangle + \langle B, C \rangle$ 。

9 向量空间

9.1 定义

设 V 是数域 P (例如 \mathbb{R} 或 \mathbb{C}) 上的非空集合, 集合上定义了封闭的向量加法 $+ : V \times V \rightarrow V$ 与数乘 $\cdot : P \times V \rightarrow V$ 。若对任意 $\alpha, \beta, \gamma \in V$ 与任意 $k, l \in P$, 下列公理都成立, 则称 $(V, +, \cdot)$ 为 P 上的向量空间:

1. 交换律: $\alpha + \beta = \beta + \alpha$ 。
2. 结合律: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ 。
3. 存在零元 $0 \in V$, 使得对一切 $\alpha \in V$ 有 $\alpha + 0 = \alpha$ 。
4. 对每个 $\alpha \in V$, 存在逆元 $-\alpha \in V$, 使得 $\alpha + (-\alpha) = 0$ 。
5. 数乘单位元: 对域中单位元 $1 \in P$, 有 $1\alpha = \alpha$ 。
6. 数乘结合律: $(kl)\alpha = k(l\alpha)$ 。

7. 数乘对标量加法的分配律: $(k + l)\alpha = k\alpha + l\alpha$ 。

8. 数乘对向量加法的分配律: $k(\alpha + \beta) = k\alpha + k\beta$ 。

9.2 子空间

设 V 是数域 P 上的向量空间, $W \subseteq V$ 。若 W 本身在 V 的加法与数乘下也构成向量空间, 则称 W 为 V 的子空间。

定义 2 种子空间运算:

- 和空间: U, W 为 V 的子空间, 则其和空间定义为

$$U + W = \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U, \mathbf{w} \in W\}.$$

- 交空间: U, W 为 V 的子空间, 则其交空间定义为

$$U \cap W = \{\mathbf{v} \mid \mathbf{v} \in U \text{ 且 } \mathbf{v} \in W\}.$$

和空间是包含 U, W 的最小子空间, 交空间是包含在 U, W 中的最大子空间。且有维度性质:

$$\dim(U + W) + \dim(U \cap W) = \dim(U) + \dim(W).$$

9.3 两子空间的主角

设 V 为数域 P 上的向量空间, $U, W \subseteq V$ 为其子空间。则存在一组主角 (principal angles) $0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_k \leq \frac{\pi}{2}$, 其中 $k = \min\{\dim(U), \dim(W)\}$, 使得:

- 对每个 $i = 1, 2, \dots, k$, 存在单位向量 $\mathbf{u}_i \in U$ 与 $\mathbf{w}_i \in W$, 使得

$$\cos(\theta_i) = \langle \mathbf{u}_i, \mathbf{w}_i \rangle.$$

- 向量组 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ 在 U 中正交归一, 向量组 $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$ 在 W 中正交归一。
- 对任意 $i < j$, 有 $\langle \mathbf{u}_i, \mathbf{w}_j \rangle = 0$ 。

性质:

- $\theta_i = 0$ 当且仅当 U 与 W 在该方向上有非零交集。
- $\theta_i = \frac{\pi}{2}$ 当且仅当 U 与 W 在该方向上正交。
- 主角可以通过奇异值分解计算得到: 设 U, W 的列正交基分别为矩阵 A, B , $A^H A = B^H B = I$, 则 $A^T B$ 的奇异值即为 $\cos(\theta_i)$ 。可推出 $\|AA^T - BB^T\|_2 = \sin(\theta_{max})$ 。(HW9)

10 代数不等式

10.1 均值不等式

设 a_1, a_2, \dots, a_n 为非负实数, 则有以下均值不等式:

$$\text{算术平均数} \geq \text{几何平均数} \geq \text{调和平均数},$$

即

$$\frac{a_1 + a_2 + \dots + a_n}{n} \geq \sqrt[n]{a_1 a_2 \cdots a_n} \geq \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \dots + \frac{1}{a_n}}.$$

变体:

- 对任意实数 a_1, a_2, \dots, a_n , 有

$$\frac{a_1 + a_2 + \dots + a_n}{n} \geq \sqrt[n]{|a_1 a_2 \cdots a_n|}.$$

- 对任意非负实数 a_1, a_2, \dots, a_n 与 $p \geq q > 0$, 有

$$\left(\frac{a_1^p + a_2^p + \dots + a_n^p}{n} \right)^{\frac{1}{p}} \geq \left(\frac{a_1^q + a_2^q + \dots + a_n^q}{n} \right)^{\frac{1}{q}}.$$

10.2 Cauchy-Schwarz 不等式

设 a_1, a_2, \dots, a_n 与 b_1, b_2, \dots, b_n 为实数, 则有以下柯西不等式:

$$(a_1^2 + a_2^2 + \dots + a_n^2)(b_1^2 + b_2^2 + \dots + b_n^2) \geq (a_1 b_1 + a_2 b_2 + \dots + a_n b_n)^2.$$

10.3 Jensen 不等式

设 $I \subseteq \mathbb{R}$ 为区间, $f: I \rightarrow \mathbb{R}$ 为凸函数, 即对任意 $x, y \in I$ 与 $\theta \in [0, 1]$, 有

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

则对任意 $x_1, x_2, \dots, x_n \in I$ 与非负权重 $\alpha_1, \alpha_2, \dots, \alpha_n$, 满足 $\sum_{i=1}^n \alpha_i = 1$, 都有

$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i f(x_i).$$

例如信息论中有: $kx + (1 - k)y \leq x^k y^{1-k}$, 其中 $x, y > 0$, $k \in [0, 1]$ 。

11 概率论

11.1 期望和方差

设随机变量 X 的概率密度函数为 $f(x)$, 则其期望 (数学期望) 定义为

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

方差定义为

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

性质:

- 线性性质：对任意常数 a, b , 有 $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ 。
- 方差的缩放性质：对任意常数 a , 有 $\text{Var}(aX) = a^2 \text{Var}(X)$ 。
- 独立性：若 X 与 Y 独立，则 $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ 。
- $D[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

11.2 常见分布

- 正态分布： $X \sim \mathcal{N}(\mu, \sigma^2)$, 概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

期望为 $\mathbb{E}[X] = \mu$, 方差为 $\text{Var}(X) = \sigma^2$ 。

- 复正态分布： $X \sim \mathcal{CN}(\mu, \sigma^2)$, 概率密度函数为

$$f(x) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{|x-\mu|^2}{\sigma^2}\right).$$

期望为 $\mathbb{E}[X] = \mu$, 方差为 $\text{Var}(X) = \sigma^2$ 。

- 均匀分布： $X \sim U(a, b)$, 概率密度函数为

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{否则.} \end{cases}$$

期望为 $\mathbb{E}[X] = \frac{a+b}{2}$, 方差为 $\text{Var}(X) = \frac{(b-a)^2}{12}$ 。

- 指数分布： $X \sim \text{Exp}(\lambda)$, 概率密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

期望为 $\mathbb{E}[X] = \frac{1}{\lambda}$, 方差为 $\text{Var}(X) = \frac{1}{\lambda^2}$ 。

- 伯努利分布： $X \sim \text{Bern}(p)$, 概率质量函数为

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

期望为 $\mathbb{E}[X] = p$, 方差为 $\text{Var}(X) = p(1-p)$ 。

- 二项分布： $X \sim \text{Bin}(n, p)$, 概率质量函数为

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

期望为 $\mathbb{E}[X] = np$, 方差为 $\text{Var}(X) = np(1-p)$ 。

- **泊松分布:** $X \sim \text{Poisson}(\lambda)$, 概率质量函数为

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

期望为 $\mathbb{E}[X] = \lambda$, 方差为 $\text{Var}(X) = \lambda$ 。

- **几何分布:** $X \sim \text{Geom}(p)$, 概率质量函数为

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

期望为 $\mathbb{E}[X] = \frac{1}{p}$, 方差为 $\text{Var}(X) = \frac{1-p}{p^2}$ 。

- **卡方分布:** $X \sim \chi^2(k)$, 概率密度函数为

$$f(x) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

期望为 $\mathbb{E}[X] = k$, 方差为 $\text{Var}(X) = 2k$ 。

- **t 分布:** $X \sim t(\nu)$, 概率密度函数为

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}.$$

期望为 $\mathbb{E}[X] = 0$ (当 $\nu > 1$ 时), 方差为 $\text{Var}(X) = \frac{\nu}{\nu-2}$ (当 $\nu > 2$ 时)。

- **F 分布:** $X \sim F(d_1, d_2)$, 概率密度函数为

$$f(x) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}, \quad x > 0.$$

期望为 $\mathbb{E}[X] = \frac{d_2}{d_2-2}$ (当 $d_2 > 2$ 时), 方差为 $\text{Var}(X) = \frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ (当 $d_2 > 4$ 时)。

性质:

- 正态分布 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的平方的分布与非中心卡方分布相关, 其期望为 $\mathbb{E}[X^2] = \mu^2 + \sigma^2$, 方差为 $\text{Var}(X^2) = 4\mu^2\sigma^2 + 2\sigma^4$ 。
- 独立正态变量的平方和服从卡方分布: 若 $X_i \sim \mathcal{N}(0, 1)$ 独立, 则 $\sum_{i=1}^k X_i^2 \sim \chi^2(k)$ 。
- t 分布可表示为正态分布与卡方分布的比值: 若 $Z \sim \mathcal{N}(0, 1)$, $V \sim \chi^2(\nu)$ 独立, 则 $T = \frac{Z}{\sqrt{V/\nu}} \sim t(\nu)$ 。
- F 分布可表示为两个独立卡方分布的比值: 若 $U_1 \sim \chi^2(d_1)$, $U_2 \sim \chi^2(d_2)$ 独立, 则 $F = \frac{(U_1/d_1)}{(U_2/d_2)} \sim F(d_1, d_2)$ 。