第一讲　预备知识：线性代数基础

1 线性空间与内积空间

- 数域, 如: $\mathbb{Q}, \mathbb{R}, \mathbb{C}$
- 线性空间, 如: $\mathbb{R}^n, \mathbb{C}^n, \mathbb{R}^{n\times n}, \mathbb{R}^{m\times m}, \mathbb{R}^{m\times n}, C[a,b]$
- 线性相关与线性无关, 秩, 基, 维数
- 子空间: 判别, 和, 维数公式
- 张成子空间: $\mathrm{span}\{x_1, x_2, \ldots, x_k\}, \quad \mathrm{span}(A)$
- 像空间 (值域, 列空间) $\mathrm{Ran}(A)$, 零空间 (核) $\mathrm{Ker}(A)$, 行空间 $\mathrm{Ran}(A^\intercal)$, 左零空间 $\mathrm{Ker}(A^\intercal)$

为了描述方便, 线性空间中的元素通常称为向量.

直和

设 $\mathbb{S}_1, \mathbb{S}_2$ 是子空间, 若 $\mathbb{S}_1 + \mathbb{S}_2$ 中的任一元素都可唯一表示成

$$x = x_1 + x_2, \quad x_1 \in \mathbb{S}_1, x_2 \in \mathbb{S}_2,$$

则称 $\mathbb{S}_1 + \mathbb{S}_2$ 为**直和**, 记为 $\mathbb{S}_1 \oplus \mathbb{S}_2$.

说明

本讲义只考虑有限维线性空间.

直和的判别与重要性质

定理 (判别定理) 设 $\mathbb{S}_1, \mathbb{S}_2$ 是线性空间 $\mathbb{S}$ 的两个子空间, 则下面的论述等价:

(1) $\mathbb{S}_1 + \mathbb{S}_2$ 是直和;

(2) $\mathbb{S}_1 + \mathbb{S}_2$ 中的零元素表示方法唯一;

(3) $\mathbb{S}_1 \cap \mathbb{S}_2 = \{0\}$;

(4) $\dim(\mathbb{S}_1) + \dim(\mathbb{S}_2) = \dim(\mathbb{S}_1 + \mathbb{S}_2)$;

定理 (重要性质) 设 $\mathbb{S}_1$ 是线性空间 $\mathbb{S}$ 的一个子空间, 则存在 $\mathbb{S}$ 的另一个子空间 $\mathbb{S}_2$, 使得

$$\mathbb{S} = \mathbb{S}_1 \oplus \mathbb{S}_2.$$

内积空间

内积与内积空间的定义与基本性质 (见讲义或线性代数教材)

例 常用内积:

- $\mathbb{R}^n/\mathbb{C}^n$: $(x,y)=y^{\mathsf{T}}x \,/\, (x,y)=y^*x$ (标准内积, 点积)
- $\mathbb{R}^{m\times n}$: $(A,B)=\mathrm{tr}(B^{\mathsf{T}}A)$
- $C[a,b]$: $(f,g)=\displaystyle\int_a^b f(x)g(x)\,\mathrm{d}x$

以上是常用内积, 但并不是唯一内积.

正交与正交补

正交: 向量与向量, 向量与子空间, 子空间与子空间 (见讲义或线代教材)

定理 设 $\mathbb{S}_1, \mathbb{S}_2$ 是内积空间 $\mathbb{S}$ 的两个子空间, 如果 $\mathbb{S}_1 \perp \mathbb{S}_2$, 则 $\mathbb{S}_1 + \mathbb{S}_2$ 是直和.

(留作课外自习)

例 (正交基) 设 $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$. 若 $x_1, x_2, \ldots, x_n$ 线性无关, 则 $\{x_1, x_2, \ldots, x_n\}$ 构成 $\mathbb{R}^n$ 的一组基. 进一步, 如果 $x_1, x_2, \ldots, x_n$ 相互正交, 即

$$(x_i, x_j) = x_j^{\mathsf{T}} x_i = 0, \quad i, j = 1, 2, \ldots, n,$$

则称它们是一组正交基. 如果还满足

$$(x_i, x_i) = x_i^{\mathsf{T}} x_i = 1, \quad i = 1, 2, \ldots, n,$$

则称它们是一组标准正交基或规范正交基.

特别地, 记 e_i 为单位矩阵的第 i 列, 则 $\{e_1, e_2, \ldots, e_n\}$ 构成 $\mathbb{R}^n$ 的一组标准正交基, 这组基通常称为自然基.

任何一组基都可以通过 Gram-Schmidt 正交化过程构造标准正交基.

正交补

定义 (正交补) 设 $\mathbb{S}_1$ 是内积空间 $\mathbb{S}$ 的一个子空间, 则 $\mathbb{S}_1$ 的正交补定义为

$$\mathbb{S}_1^{\perp} \triangleq \{\, x \in \mathbb{S} \,:\, x \perp \mathbb{S}_1 \,\},$$

即 $\mathbb{S}$ 中所有与 $\mathbb{S}_1$ 正交的元素组成的集合.

可以证明 $\mathbb{S}_1^{\perp}$ 也是子空间, 且

定理 设 $\mathbb{S}_1$ 是内积空间 $\mathbb{S}$ 的一个有限维子空间, 则 $\mathbb{S}_1^{\perp}$ 存在唯一, 且

$$\mathbb{S} = \mathbb{S}_1 \oplus \mathbb{S}_1^{\perp}.$$

2 矩阵与投影

矩阵的秩、特征值、特征向量

- 矩阵的秩
- 特征多项式, 特征值, 特征向量, 左特征向量, 特征对
- 代数重数和几何重数, 特征空间
- 零化多项式, 最小多项式
- 可对角化, 特征值分解/谱分解, 可对角化的充要条件

为了讨论方便, 本节仅考虑实数情形, 对于复数情形, 结论类似.

一些基本性质

定理 设 $A, B \in \mathbb{R}^{m\times n}$, 则

- $\mathrm{rank}(A) \le \min\{m, n\}$;
- $\mathrm{rank}(A^{\intercal}) = \mathrm{rank}(A)$;
- $\mathrm{rank}(A^{\intercal}A) = \mathrm{rank}(AA^{\intercal}) = \mathrm{rank}(A)$;
- $\mathrm{rank}(A + B) \le \mathrm{rank}(A) + \mathrm{rank}(B)$;
- 对任意非奇异矩阵 $P \in \mathbb{R}^{m\times m}$ 和 $Q \in \mathbb{R}^{n\times n}$, 有

$$\mathrm{rank}(PA) = \mathrm{rank}(AQ) = \mathrm{rank}(PAQ) = \mathrm{rank}(A).$$

矩阵秩分解

定理 设 $\mathrm{rank}(A)=\ell$, 则存在非奇异矩阵 $P\in\mathbb{R}^{m\times m}$ 和 $Q\in\mathbb{R}^{n\times n}$ 使得

$$A=P\begin{bmatrix} I_\ell & 0 \\ 0 & 0 \end{bmatrix}Q.$$

进一步, $\mathrm{rank}(A)=\mathrm{rank}(B)$ 当且仅当存在非奇异矩阵 $P\in\mathbb{R}^{m\times m}$ 和 $Q\in\mathbb{R}^{n\times n}$ 使得 $A=PBQ$.

推论 (满秩分解) 设 $\mathrm{rank}(A)=\ell$, 则存在非奇异矩阵 $F\in\mathbb{R}^{m\times\ell}$ 和 $G\in\mathbb{R}^{\ell\times n}$ 使得

$$A=FG.$$

秩的更多性质

定理 设 $A \in \mathbb{R}^{m\times k}$, $B \in \mathbb{R}^{k\times n}$, 则

$$\operatorname{rank}(A) + \operatorname{rank}(B) - k \le \operatorname{rank}(AB) \le \min\{\operatorname{rank}(A),\ \operatorname{rank}(B)\}.$$

推论 设 $A \in \mathbb{R}^{m\times k}$, $B \in \mathbb{R}^{k\times m}$, $k \le m$. 若 A 和 B 都是满秩矩阵, 则

$$\operatorname{rank}(AB) = \operatorname{rank}(BA) = \operatorname{rank}(A) = \operatorname{rank}(B) = k.$$

张成的线性空间

设 $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$, 记

$$\mathrm{span}\{x_1, x_2, \ldots, x_k\} \triangleq \{\, \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k \ : \ \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}\},$$

为由 $x_1, x_2, \ldots, x_k$ **张成的线性空间**

$A \in \mathbb{R}^{m\times n}$ 相关的四个子空间

- **像空间, 列空间, 值域**: $\mathrm{Ran}(A) = \mathrm{span}\{A\}$
- **零空间, 核**: $\mathrm{Ker}(A) \triangleq \{\, x \in \mathbb{R}^n \ : \ Ax = 0 \,\}$
- **行空间**: $\mathrm{Ran}(A^\intercal) = \mathrm{span}\{A^\intercal\}$
- **左零空间**: $\mathrm{Ker}(A^\intercal)$

定理 设 $A \in \mathbb{R}^{m\times n}$, 则有

- $\dim(\mathrm{Ran}(A)) = \dim(\mathrm{Ran}(A^\intercal)) = \mathrm{rank}(A)$;
- $\dim(\mathrm{Ker}(A)) + \dim(\mathrm{Ran}(A^\intercal)) = n$;
- $\mathrm{Ran}(A^\intercal A) = \mathrm{Ran}(A^\intercal), \quad \mathrm{Ker}(A^\intercal A) = \mathrm{Ker}(A)$.

例 设 $A \in \mathbb{R}^{m\times n}$, 则

$$\mathrm{Ran}(A)^\perp = \mathrm{Ker}(A^\intercal).$$

(留作课外自习, 证明见讲义)

例 设 $A \in \mathbb{R}^{m\times n}$, 则由上例可知

$$\mathrm{Ker}(A) \oplus \mathrm{Ran}(A^\intercal) = \mathbb{R}^n, \qquad \mathrm{Ker}(A^\intercal) \oplus \mathrm{Ran}(A) = \mathbb{R}^m.$$

特征值与特征向量

特征值相关性质: $A \in \mathbb{R}^{n\times n}$

- $\lambda_1\lambda_2\cdots\lambda_n = \det(A)$
- $\lambda_1 + \lambda_2 + \cdots + \lambda_n = \mathrm{tr}(A)$ (**矩阵的迹**: $\mathrm{tr}(A) \triangleq a_{11} + a_{22} + \cdots + a_{nn}$)
- **矩阵的谱**: $\sigma(A) \triangleq \{\, \lambda_1, \lambda_2, \ldots, \lambda_n \,\}$
- **谱分解/特征值分解 (可对角化)**: $X^{-1}AX = \Lambda$ **或** $A = X\Lambda X^{-1}$
- **矩阵的谱半径**: $\rho(A) \triangleq \max_{\lambda\in\sigma(A)} |\lambda|$

定理 矩阵的特征值关于矩阵元素是连续的.

思考:

- 设 $A \in \mathbb{R}^{n\times n}$, 则 A^{T} 与 A 的特征值和特征向量是什么关系?
- A^{-1} 与 A 的特征值和特征向量是什么关系?
- 设 $p(t)$ 是多项式或有理函数, 则 $p(A)$ 与 A 的特征值和特征向量是什么关系?

特征值估计: Bendixson 定理

定理 设 $A \in \mathbb{C}^{n\times n}$, H 和 S 分别为 A 的 Hermite 部分和 Skew-Hermite 部分:

$$H \triangleq \frac{1}{2}(A + A^*), \quad S \triangleq \frac{1}{2}(A - A^*),$$

则

$$\lambda_{\min}(H) \le \mathrm{Re}(\lambda(A)) \le \lambda_{\max}(H),$$

$$\lambda_{\min}(-\mathbf{i}\, S) \le \mathrm{Im}(\lambda(A)) \le \lambda_{\max}(-\mathbf{i}\, S).$$

其中 $\mathrm{Re}(\cdot)$ 和 $\mathrm{Im}(\cdot)$ 分别表示实部和虚部, $\mathbf{i}$ 是虚部单位. 即: 特征值的实部的取值范围由其 **Hermite** 部分确定, 而虚部则由其 **Skew-Hermite** 部分确定.

特征值估计: Gerschgorin 圆盘

设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$, 则 A 的 **Gerschgorin 圆盘** (盖尔圆盘) 定义为

$$\mathcal{D}_i \triangleq \left\{ z \in \mathbb{C} \ : \ |z - a_{ii}| \leq \sum_{j=1, j\neq i}^{n} |a_{ij}| \right\}, \quad i = 1, 2, \ldots, n.$$

定理 (Gerschgorin 圆盘定理) 设 $A \in \mathbb{C}^{n\times n}$, 则 A 的所有特征值都包含在 A 的 Gerschgorin 圆盘的并集中, 即 $\sigma(A) \subset \bigcup_{i=1}^{n} \mathcal{D}_i$. (留作课外自习, 证明见讲义)

定理 如果 $\bigcup_{i=1}^{n} \mathcal{D}_i$ 可分解成两个不相交的子集 S 和 T, 其中 S 由 k 个圆盘组成, 而 T 由其它 $n-k$ 个圆盘组成, 则 S 恰好包含 A 的 k 个特征值, 而 T 则包含其它 $n-k$ 个特征值.

不变子空间

设 $A \in \mathbb{R}^{n\times n}$, 子空间 $\mathbb{S} \subseteq \mathbb{R}^n$. 若 $A\mathbb{S} \subseteq \mathbb{S}$, 即对任意 $x \in \mathbb{S}$, 都有 $Ax \in \mathbb{S}$, 则称 $\mathbb{S}$ 为 A 的一个 不变子空间 .

一类特殊的不变子空间

设 $x_1, x_2, \ldots, x_m$ 是 A 的一组线性无关的特征向量, 则 $\mathrm{span}\{x_1, x_2, \ldots, x_m\}$ 是 A 的一个 m 维不变子空间.

不变子空间的重要性质

定理 设 $A \in \mathbb{R}^{n\times n}$, $X \in \mathbb{R}^{n\times k}$ 且 $\mathrm{rank}(X) = k$, 则 $\mathrm{span}(X)$ 是 A 的不变子空间的充要条件是存在一个矩阵 $B \in \mathbb{R}^{k\times k}$ 使得

$$AX = XB,$$

此时, B 的特征值都是 A 的特征值. (板书)

推论 设 $A \in \mathbb{R}^{n\times n}$, $X \in \mathbb{R}^{n\times k}$ 且 $\mathrm{rank}(X) = k$. 若存在 $B \in \mathbb{R}^{k\times k}$ 使得 $AX = XB$, 则 (λ, v) 是 B 的特征对当且仅当 (λ, Xv) 是 A 的特征对.

(留作课外自习)

投影变换

设 $\mathbb{S} = \mathbb{S}_1 \oplus \mathbb{S}_2$, 则 $\mathbb{S}$ 中的任意向量 x 都可唯一表示为

$$x = x_1 + x_2, \quad x_1 \in \mathbb{S}_1,\ x_2 \in \mathbb{S}_2.$$

我们称 x_1 为 x **沿 $\mathbb{S}_2$ 到 $\mathbb{S}_1$ 上的投影**, 记为 $x|_{\mathbb{S}_1}$.

由于 $\mathbb{S}_1$ 的补空间不唯一, 因此在讨论投影时要明确给定 $\mathbb{S}_2$.

例 设 $\mathbb{S}_1 = \mathrm{span}\{e_1\}$, $\mathbb{S}_2 = \mathrm{span}\{e_2\}$, $\tilde{\mathbb{S}}_2 = \mathrm{span}\{e\}$, 其中 $e_1 = [1,0]^\mathsf{T}$, $e_2 = [0,1]^\mathsf{T}$, $e = [1,1]^\mathsf{T}$. 于是有

$$\mathbb{R}^2 = \mathbb{S}_1 \oplus \mathbb{S}_2 = \mathbb{S}_1 \oplus \tilde{\mathbb{S}}_2.$$

$x = [2,3]^\mathsf{T}$ 沿 $\mathbb{S}_2$ 到 $\mathbb{S}_1$ 上的投影是 $[2,0]^\mathsf{T}$, 沿 $\tilde{\mathbb{S}}_2$ 到 $\mathbb{S}_1$ 上的投影是 $[-1,0]^\mathsf{T}$.

投影变换与投影矩阵

定义线性变换 $P:\mathbb{S}\to\mathbb{S}$ 如下:

$$Px = x|_{\mathbb{S}_1},$$

则称 P 是从 $\mathbb{S}$ 沿 $\mathbb{S}_2$ 到 $\mathbb{S}_1$ 上的 **投影变换** 或 **投影算子**

对应的变换矩阵称为 **投影矩阵**

几点注记

- 线性变换在不同的基下对应不同的变换矩阵. 在不加特别指出时, 本讲义中如果线性空间是 $\mathbb{R}^n$ 或 $\mathbb{R}^{n\times n}$, 我们采用自然基.
- 为书写方便, 我们使用 P 既表示投影变换也表示其对应的投影矩阵.

投影矩阵由其像空间和零空间唯一确定

设 P 是从 $\mathbb{S}$ 沿 $\mathbb{S}_2$ 到 $\mathbb{S}_1$ 上的投影变换, 则容易验证

$$\boxed{\mathrm{Ran}(P)=\mathbb{S}_1,\ \mathrm{Ker}(P)=\mathbb{S}_2} \Longrightarrow \boxed{\mathbb{R}^n=\mathrm{Ran}(P)\oplus\mathrm{Ker}(P)}$$

注: 对于一般矩阵有 $\mathbb{R}^n=\mathrm{Ran}(A)\oplus\mathrm{Ker}(A^{\intercal})$

定理 若 $\mathbb{S}_1$ 和 $\mathbb{S}_2$ 是 $\mathbb{R}^n$ 的两个子空间, 且 $\mathbb{R}^n=\mathbb{S}_1\oplus\mathbb{S}_2$, 则存在唯一的投影矩阵 P, 使得

$$\mathrm{Ran}(P)=\mathbb{S}_1,\quad \mathrm{Ker}(P)=\mathbb{S}_2.$$

- 若 $\mathbb{S}_1=\mathbb{R}^n$, 则 $\mathbb{S}_2=\{0\}$, 所对应的投影矩阵即为单位矩阵 I.
- 若 $\mathbb{S}_1=\{0\}$, 则 $\mathbb{S}_2=\mathbb{R}^n$, 此时所对应的投影矩阵即为零矩阵.

投影矩阵的一个简单性质

引理 设 $P \in \mathbb{R}^{n\times n}$ 是一个投影矩阵, 则

(1) $I - P$ 也是一个投影矩阵, 且 $\mathrm{Ker}(P) = \mathrm{Ran}(I - P)$;

(2) P^{T} 也是一个投影矩阵.

投影矩阵的判别

定理 矩阵 $P \in \mathbb{R}^{n\times n}$ 是投影矩阵的充要条件是 $P^2 = P$.

(留作课外自习, 证明见讲义)

投影变换的矩阵表示

设 $\mathbb{S}_1$ 和 $\mathbb{S}_2$ 是 $\mathbb{R}^n$ 的两个 m 维子空间, 且 $\mathbb{R}^n = \mathbb{S}_1 \oplus \mathbb{S}_2^{\perp}$, 则存在唯一的投影变换 P, 使得

$$\mathrm{Ran}(P) = \mathbb{S}_1, \quad \mathrm{Ker}(P) = \mathbb{S}_2^{\perp}.$$

此时, 我们称 P 是 $\mathbb{S}_1$ 上与 $\mathbb{S}_2$ 正交的投影矩阵.

定理 设 $P \in \mathbb{R}^{n\times n}$ 是 $\mathbb{S}_1$ 上与 $\mathbb{S}_2$ 正交的投影变换, $v_1, v_2, \ldots, v_m$ 和 $w_1, w_2, \ldots, w_m$ 分别构成 $\mathbb{S}_1$ 和 $\mathbb{S}_2$ 的一组基, 则

$$P = V(W^{\intercal}V)^{-1}W^{\intercal},$$

其中 $V = [v_1, v_2, \ldots, v_m], W = [w_1, w_2, \ldots, w_m]$. (留作练习)

正交投影变换

设 $\mathbb{S}_1$ 是内积空间 $\mathbb{S}$ 的一个子空间, $x \in \mathbb{S}$, 则 x 可唯一分解成

$$x = x_1 + x_2, \quad x_1 \in \mathbb{S}_1,\ x_2 \in \mathbb{S}_1^{\perp},$$

其中 x_1 称为 x **在 $\mathbb{S}_1$ 上的正交投影**.

- 若 P 是沿 $\mathbb{S}_1^{\perp}$ 到 $\mathbb{S}_1$ 上的投影变换, 则称 P 为 **$\mathbb{S}_1$ 上的正交投影变换**, 对应的矩阵为 **正交投影矩阵**, 记为 $P_{\mathbb{S}_1}$
- 非正交投影变换称为 **斜投影变换**

正交投影变换的矩阵表示

设 P 是 $\mathbb{S}_1$ 上的正交投影变换, 则

$$\boxed{P = VV^{\intercal},}$$

其中 $V = [v_1, v_2, \ldots, v_m]$ 的列向量组构成 $\mathbb{S}_1$ 的一组 **标准正交基**.

正交投影的判别

定理 投影矩阵 $P \in \mathbb{R}^{n\times n}$ 是正交投影矩阵的充要条件 $P^{\intercal} = P$.

(留作练习)

3 向量范数与矩阵范数

向量范数

定义 (向量范数) 若函数 $f:\mathbb{R}^n \to \mathbb{R}$ (或 $f:\mathbb{C}^n \to \mathbb{R}$) 满足

(1) $f(x) \geq 0, \forall\, x \in \mathbb{R}^n$ (或 $\mathbb{C}^n$), 等号当且仅当 $x = 0$ 时成立;

(2) $f(\alpha x) = |\alpha| \cdot f(x), \forall\, x \in \mathbb{R}^n$ (或 $\mathbb{C}^n$), $\alpha \in \mathbb{R}$ (或 $\mathbb{C}$);

(3) $f(x+y) \leq f(x) + f(y), \forall\, x, y \in \mathbb{R}^n$ (或 $\mathbb{C}^n$);

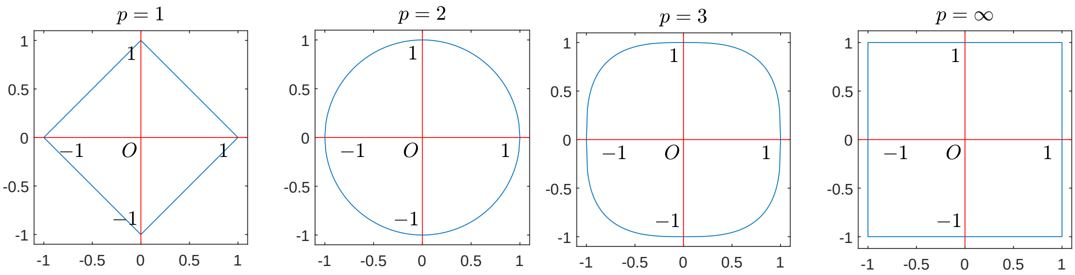
则称 $f(x)$ 为 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 上的 **范数**, 通常记作 $\|x\|$.

如果 f 只满足 $f(x) \geq 0$, 以及 (2) 和 (3), 则称为 **半范数**.

$\mathbb{R}^n$ 和 $\mathbb{C}^n$ 上常见的向量范数

- 1-范数: $\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$;
- 2-范数: $\|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$;
- ∞-范数: $\|x\|_\infty = \max\limits_{1\leq i\leq n} |x_i|$;
- p-范数: $\|x\|_p = \left(\sum\limits_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad p \geq 1.$

例 (p-范数举例)



p-范数也称为 Hölder 范数或 ℓ_p 范数.

范数的连续性

定理 设 $\|\cdot\|$ 是 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 上的一个向量范数, 则 $f(x) \triangleq \|x\|$ 是 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 上的连续函数.

(留作课外自习, 利用范数的三角不等式)

范数的等价性

定义 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 上向量范数, 若存在正常数 c_1, c_2, 使得

$$c_1\|x\|_\alpha \leq \|x\|_\beta \leq c_2\|x\|_\alpha$$

对任意 $x \in \mathbb{R}^n$ (或 $\mathbb{C}^n$) 都成立, 则称 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是等价的.

定理 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 上的所有向量范数都是等价的, 特别地, 有

$$\|x\|_2 \le \|x\|_1 \le \sqrt{n}\,\|x\|_2,\ \|x\|_\infty \le \|x\|_1 \le n\,\|x\|_\infty,\ \|x\|_\infty \le \|x\|_2 \le \sqrt{n}\,\|x\|_\infty$$

(板书, 以 $\mathbb{R}^n$ 为例, 只证明等价性, 三个不等式留作练习)

事实上, 有限维赋范线性空间上的所有范数都是等价的.

事实上, 我们有

$$\|x\|_p \ge \|x\|_q, \quad \forall\, 1 \le p \le q.$$

(留作课外自习, 证明见讲义)

定理 (Cauchy-Schwarz 不等式) 设 $(\cdot,\cdot)$ 是 $\mathbb{C}^n$ (或 $\mathbb{R}^n$) 上的内积, 则对任意 $x, y \in \mathbb{C}^n$ (或 $\mathbb{R}^n$), 有

$$|(x,y)|^2 \le (x,x)\cdot(y,y)$$

且等号成立的充要条件是 x 与 y 线性相关.

(板书)

中学数学中的 Cauchy 不等式

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \le \left(a_1^2 + a_2^2 + \cdots + a_n^2\right)\left(b_1^2 + b_2^2 + \cdots + b_n^2\right),$$

是 Cauchy-Schwarz 不等式的特例.

更一般地, 我们有下面的 Holder 不等式.

定理 (Holder 不等式) 设 $(\cdot,\cdot)$ 是 $\mathbb{C}^n$ (或 $\mathbb{R}^n$) 上的内积, 则对任意 $x, y \in \mathbb{C}^n$ (或 $\mathbb{R}^n$), 有

$$|(x,y)| \leq \|x\|_p \cdot \|y\|_q,$$

其中 $p, q > 0$, 且 $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

(留作课外自习)

内积导出范数

设 $\mathbb{S}$ 是内积空间, 对任意 $x \in \mathbb{S}$, 定义

$$\|x\| \triangleq (x, x)^{\frac{1}{2}},$$

则可以验证, $\|x\|$ 是 $\mathbb{S}$ 上的范数. 这就是 **由内积导出的范数**

任意一个内积都可以导出一个相应的范数.

例 $\mathbb{R}^n/\mathbb{C}^n$ 上由标准内积导出的范数为 $\|x\| = (x, x)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$.

矩阵范数

定义 若函数 $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ (或 $f : \mathbb{C}^{m\times n} \to \mathbb{R}$) 满足

(1) $f(A) \geq 0, \forall\, A \in \mathbb{R}^{m\times n}$ (或 $\mathbb{C}^{m\times n}$) 且等号当且仅当 $A = 0$ 时成立;

(2) $f(\alpha A) = |\alpha|\, f(A), \forall\, A \in \mathbb{R}^{m\times n}$ (或 $\mathbb{C}^{m\times n}$), $\alpha \in \mathbb{R}$ (或 $\mathbb{C}$);

(3) $f(A+B) \leq f(A) + f(B), \forall A, B \in \mathbb{R}^{m\times n}$ (或 $\mathbb{C}^{m\times n}$);

则称 $f(X)$ 为 $\mathbb{R}^{m\times n}$ (或 $\mathbb{C}^{m\times n}$) 上的矩阵范数, 通常记作 $\|X\|$.

设 f 是 $\mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$) 上的矩阵范数, 如果 f 还满足

(4) $f(AB) \leq f(A)f(B), \forall\, A, B \in \mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$),

则称 f 是 **相容的矩阵范数** (本讲义中所涉及的矩阵范数都是指相容的矩阵范数)

算子范数

引理 设 $\|\cdot\|$ 是 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 上的向量范数, 则

$$\|A\| \triangleq \sup_{x\in\mathbb{R}^n,\ x\neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$$

是 $\mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$) 上的矩阵范数, 称为算子范数, 有时也称为诱导范数或导出范数.

(板书)

相应地, 可以定义 $\mathbb{R}^{m\times n}$ (或 $\mathbb{C}^{m\times n}$) 上的算子范数, 此时涉及 $\mathbb{R}^m$ 和 $\mathbb{R}^n$ (或 $\mathbb{C}^m$ 和 $\mathbb{C}^n$) 上的向量范数.

例 设 $A \in \mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$), 常见的矩阵范数有:

- p-范数 (算子范数)

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p}, \quad p \geq 1.$$

- Frobenius 范数, 简称 F-范数

$$\|A\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2}\ ;$$

引理 可以证明:

(1) 矩阵 1-范数 (列范数): $\|A\|_1 = \max\limits_{1\le j\le n}\left(\sum\limits_{i=1}^{n}|a_{ij}|\right)$;

(2) 矩阵 ∞-范数 (行范数): $\|A\|_\infty = \max\limits_{1\le i\le n}\left(\sum\limits_{j=1}^{n}|a_{ij}|\right)$;

(3) 矩阵 2-范数: $\|A\|_2 = \sqrt{\rho(A^*A)}$;

(4) 矩阵 F-范数: $\|A\|_F = \sqrt{\mathrm{tr}(A^*A)}$

(板书, 以 ∞-范数和 2-范数为例, 1-范数和 F-范数留作练习)

定理 (矩阵范数的等价性) $\mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$) 上的所有范数都是等价的, 特别地, 有

$$\frac{1}{\sqrt{n}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\,\|A\|_1, \qquad \frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\,\|A\|_\infty.$$

(留作练习)

矩阵范数的更多性质

(1) 对任意矩阵范数 $\|\cdot\|$, 有 $\|A^k\| \le \|A\|^k$;

(2) 对任意算子范数 $\|\cdot\|$, 有 $\|Ax\| \le \|A\|\cdot\|x\|$, $\|AB\| \le \|A\|\cdot\|B\|$;

(3) $\|Ax\|_2 \le \|A\|_F \cdot \|x\|_2$, $\|AB\|_F \le \|A\|_F \cdot \|B\|_F$;

(4) F-范数不是算子范数;

(5) $\|\cdot\|_2$ 和 $\|\cdot\|_F$ 是**酉不变范数**, 即对任意酉矩阵 (或正交矩阵) U, V, 有

$$\|UA\|_2 = \|AV\|_2 = \|UAV\|_2 = \|A\|_2 \ ,$$

$$\|UA\|_F = \|AV\|_F = \|UAV\|_F = \|A\|_F$$

(6) $\|A^*\|_2 = \|A\|_2$, $\|A^*\|_1 = \|A\|_\infty$;

(7) 若 A 是正规矩阵, 则 $\|A\|_2 = \rho(A)$. (留作课外自习)

其他范数

在数据处理和机器学习等学科中经常会用到下面的范数:

- 向量 ℓ_0 范数

$$\|x\|_0 \triangleq x \text{ 中非零元素的个数}, \quad x \in \mathbb{R}^n \ (\text{或 } \mathbb{C}^n).$$

需要指出的是, 上式定义的 ℓ_0 范数并不满足向量范数定义中的条件 (2). 该范数主要用于衡量向量的稀疏性, 是压缩感知和稀疏优化中的研究对象.

- 矩阵核范数

$$\|A\|_* \triangleq \sum \sigma_i, \quad \text{其中 } \sigma_i \text{ 为 } A \text{ 的所有奇异值}, A \in \mathbb{R}^{m\times n} \ (\text{或 } \mathbb{C}^{m\times n}).$$

(矩阵奇异值将在第三讲中介绍) 根据奇异值的性质, 核范数也可以定义为

$$\|A\|_* \triangleq \operatorname{tr}\left(\sqrt{A^{\mathsf{T}} A}\right).$$

矩阵直和

设 $A_i \in \mathbb{R}^{n_i\times n_i}$ (或 $\mathbb{C}^{n_i\times n_i}$), $i=1,2,\ldots,k$, 定义 **直和**

$$\bigoplus_{i=1}^{k} A_i = A_1 \oplus A_2 \oplus \cdots \oplus A_k \triangleq \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix},$$

即以 A_i 为对角块的块对角矩阵. 可以验证

$$\left\| \bigoplus_{i=1}^{k} A_i \right\|_p = \max_{1\le i\le k} \|A_i\|_p, \quad p=1,2,\infty.$$

谱半径与范数

定理 (谱半径与范数的关系) 设 $A \in \mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$), 则

(1) 对任意算子范数, 有 $\rho(A) \le \|A\|$;

(2) 反之, 对任意 $\varepsilon > 0$, 都存在一个算子范数 $\|\cdot\|_\varepsilon$, 使得 $\|A\|_\varepsilon \le \rho(A) + \varepsilon$, 其中范数 $\|\cdot\|_\varepsilon$ 依赖于 A 和 ε. 所以, 若 $\rho(A) < 1$, 则存在算子范数 $\|\cdot\|_\varepsilon$, 使得 $\|A\|_\varepsilon < 1$;

(板书)

事实上, 定理 3.8 中的结论 (1) 对任意矩阵范数都成立, 留作练习.

推论 设 $A \in \mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$), 则 $\|A\|_2^2 \leq \|A\|_1 \|A\|_\infty$, 且

$$\max_{1\leq i,j\leq n} \{|a_{ij}|\} \leq \|A\|_2 \leq n \max_{1\leq i,j\leq n} \{|a_{ij}|\}.$$

(留作练习, 利用矩阵特征值与迹的关系)

最佳逼近与正交投影

正交投影变换的一个常用性质

定理 设 $P \in \mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$) 是正交投影矩阵, 则 $\|P\|_2 = 1$, 且对 $\forall x \in \mathbb{R}^n$ (或 $\mathbb{C}^n$), 有

$$\|x\|_2^2 = \|Px\|_2^2 + \|(I-P)x\|_2^2.$$

用正交投影表示最佳逼近问题的解

定理 设 $\mathbb{S}$ 是 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 的子空间, $z \in \mathbb{R}^n$ (或 $\mathbb{C}^n$) 是一个给定的向量, 则最佳逼近问题

$$\min_{x \in \mathbb{S}} \|x - z\|_2$$

的唯一解为

$$x_* = P_{\mathbb{S}} z.$$

即 $\mathbb{S}$ 中距离 z 最近 (在 2-范数意义下) 的向量是 z 在 $\mathbb{S}$ 中的正交投影.

(留作练习)

上述定理中的 2-范数可以推广到一般的能量范数.

推论 设 $A \in \mathbb{R}^{n\times n}$ (或 $\mathbb{C}^{n\times n}$) 对称正定 (或 Hermite 正定), $\mathbb{S}$ 是 $\mathbb{R}^n$ (或 $\mathbb{C}^n$) 的子空间, 给定 $z \in \mathbb{R}^n$ (或 $\mathbb{C}^n$), 则 x_* 是最佳逼近问题

$$\min_{x\in\mathbb{S}} \|x - z\|_A$$

的解的充要条件是

$$x_* \in \mathbb{S} \quad \text{且} \quad A(x_* - z) \perp \mathbb{S}.$$

此处能量范数 $\|\cdot\|_A$ 的定义为: $\|x - z\|_A \triangleq \sqrt{(x-z)^* A(x-z)}$.

(留作练习)

4 矩阵标准型

计算矩阵特征值的一个基本思想

通过相似变换, 将其转化成一个形式尽可能简单的矩阵, 使得其特征值更易于计算.

两个非常有用的特殊矩阵是 Jordan 标准型 和 Schur 标准型

Jordan 标准型

定理 设 $A \in \mathbb{C}^{n\times n}$ 有 p 个不同特征值, 则存在非奇异矩阵 $X \in \mathbb{C}^{n\times n}$, 使得

$$X^{-1}AX = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_p \end{bmatrix} \triangleq J, \qquad (\text{Jordan 标准型})$$

其中 J_i 的维数等于 λ_i 的代数重数, 且具有下面的结构

$$J_i = \begin{bmatrix} J_i^{(1)} & & & \\ & J_i^{(2)} & & \\ & & \ddots & \\ & & & J_i^{(\nu_i)} \end{bmatrix}, \quad J_i^{(k)} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}.$$

这里的 ν_i 为 λ_i 的几何重数, $J_i^{(k)}$ 称为 (对应于 λ_i 的) **Jordan 块**.

另一种描述

设 $A \in \mathbb{C}^{n\times n}$, 则存在非奇异矩阵 $X \in \mathbb{C}^{n\times n}$ 使得

$$X^{-1}AX = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_\ell \end{bmatrix},$$

其中

$$J_k = \begin{bmatrix} \lambda_k & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_k & 1 \\ & & & \lambda_k \end{bmatrix}_{n_k\times n_k}, \quad n_k \geq 1,\ k = 1, 2, \ldots, \ell,\ 且 \sum_{k=1}^{\ell} n_k = n.$$

矩阵 J_k 就称为 **Jordan 块**, $\lambda_1, \lambda_2, \ldots, \lambda_\ell$ 是 A 的特征值 (不要求互异).

Jordan 标准型的基本性质

- 除了 Jordan 块的排列次序外, Jordan 标准型是唯一确定的.
- Jordan 块的个数等于 A 的线性无关的特征向量的个数;
- A 可对角化的充要条件是每个 Jordan 块都是 1×1 的, 此时 X 的列向量就是 A 的特征向量.
- 可以证明, 对于每一个 Jordan 块 $J_i^{(k)}$, 都存在一个列满秩矩阵 $X_i^{(k)}$ 使得

$$AX_i^{(k)} = X_i^{(k)} J_i^{(k)}.$$

Jordan 标准型的应用

推论 所有可对角化矩阵组成的集合在所有矩阵组成的集合中是稠密的, 即任何一个矩阵都可以通过可对角化矩阵来逼近.

Jordan 标准型的另一个重要应用是可以用来计算矩阵的最小多项式.

定理 设 $\lambda_1, \lambda_2, \ldots, \lambda_p$ 为 $A \in \mathbb{C}^{n\times n}$ 的互不相等的特征值, 则 A 的最小多项式为

$$p(\lambda) = \prod_{i=1}^{q} (\lambda - \lambda_i)^{r_i},$$

其中 r_i 是与 λ_i 所对应的最大 Jordan 块的维数.

Schur 分解

定理 设 $A \in \mathbb{C}^{n\times n}$ (或 $\mathbb{R}^{n\times n}$), 则存在酉矩阵 $U \in \mathbb{C}^{n\times n}$ 使得

$$U^*AU = \begin{bmatrix} \lambda_1 & r_{12} & \cdots & r_{1n} \\ 0 & \lambda_2 & \cdots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \triangleq R \quad \text{或} \quad A = URU^*,$$

其中 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 是 A 的特征值 (可以按任意顺序排列). (板书)

关于 Schur 分解的几点说明

- 该结论告诉我们, 任意一个矩阵都可以**酉三角化**.
- 三角矩阵可以说是一般矩阵在酉相似变化下的最简形式.
- 定理中的 U 和 R 不是唯一的.
- R 的对角线元素可以按任意顺序排列, 特别地, 可以按模从大到小排列.

推论 设 $A \in \mathbb{C}^{n\times n}$, 则

(1) A 是正规矩阵当且仅当 R 是对角矩阵,

即 A 可酉对角化当且仅当 A 是正规矩阵;

(2) A 是 Hermite 矩阵当且仅当 R 是实对角矩阵.

实 Schur 分解 (或拟 Schur 分解)

定理 设 $A \in \mathbb{R}^{n\times n}$, 则存在正交矩阵 $Q \in \mathbb{R}^{n\times n}$, 使得

$$Q^{\intercal} A Q = T,$$

其中 $T \in \mathbb{R}^{n\times n}$ 是拟上三角矩阵, 即 T 是块上三角的, 且对角块为 1×1 或 2×2 的块矩阵. 若对角块是 1×1 的, 则其就是 A 的一个特征值, 若对角块是 2×2 的, 则其特征值是 A 的一对共轭复特征值.

(板书)

5 几类特殊矩阵

对称正定矩阵

设 $A \in \mathbb{C}^{n\times n}$.

- 若对所有向量 $x \in \mathbb{C}^n$ 有 $\mathrm{Re}(x^*Ax) \geq 0$, 则称 A 是**半正定**的;
- 若对所有非零向量 $x \in \mathbb{C}^n$ 有 $\mathrm{Re}(x^*Ax) > 0$, 则称 A 是**正定**的;
- 若 A 是 Hermite 的且半正定, 则称 A 为 **Hermite 半正定**;
- 若 A 是 Hermite 的且正定, 则称 A 为 **Hermite 正定**;
- 若 $A \in \mathbb{R}^{n\times n}$ 是对称的且半正定, 则称 A 为**对称半正定**;
- 若 $A \in \mathbb{R}^{n\times n}$ 是对称的且正定, 则称 A 为**对称正定**.

本讲义中, 正定和半正定矩阵不要求是对称或 Hermite.

对称正定矩阵的判别

定理 设 $A \in \mathbb{C}^{n\times n}$, 则

(1) A 正定的充要条件是 $H = \dfrac{1}{2}(A + A^*)$ 正定;

(2) A 半正定的充要条件是 $H = \dfrac{1}{2}(A + A^*)$ 半正定.

(留作练习)

定理 设 $A \in \mathbb{R}^{n\times n}$, 则 A 正定 (或半正定) 的充要条件是对任意非零向量 $x \in \mathbb{R}^n$ 有 $x^\intercal A x > 0$ (或 $x^\intercal A x \geq 0$).

(留作练习)

该结论表明: 如果 A 是实矩阵, 则只需在实数域中考虑即可.

对称正定矩阵的基本性质

(1) A Hermite 正定当且仅当 A Hermite 且所有特征值都是正的;

(2) A Hermite 正定当且仅当存在酉矩阵 U 使得 $A = U\Lambda U^*$, 其中 Λ 为对角矩阵且对角线均为正实数;

(3) A Hermite 正定当且仅当 S^*AS 对称正定, 其中 $S \in \mathbb{C}^{n\times n}$ 是一个任意的非奇异矩阵;

(4) 若 A Hermite 正定, 则 A 的任意主子矩阵都 Hermite 正定;

(5) 若 A Hermite 正定, 则 A 的所有对角线元素都是正的, 且 $\max\limits_{i\neq j}\{|a_{ij}|\} < \max\limits_{i}\{a_{ii}\}$, 即绝对值最大的元素出现在对角线上.

平方根

定理 设 $A \in \mathbb{C}^{n\times n}$ 是一个 Hermite 半正定矩阵, k 是一个给定的正整数. 则存在唯一的 Hermite 半正定矩阵 $B \in \mathbb{C}^{n\times n}$ 使得

$$B^k = A.$$

同时, 我们还有下面的性质:

(1) $\operatorname{rank}(B) = \operatorname{rank}(A)$, 因此, 若 A 是正定的, 则 B 也正定;

(2) 如果 A 是实矩阵的, 则 B 也是实矩阵.

特别地, 当 $k = 2$ 时, 称 B 为 A 的平方根, 记为 $A^{\frac{1}{2}}$.

(留作课外自习)

Hermite 正定矩阵与内积

定理 设 $(\cdot,\cdot)$ 是 $\mathbb{C}^n$ 上的一个内积, 则存在一个 Hermite 正定矩阵 $A \in \mathbb{C}^{n\times n}$ 使得

$$(x,y) = y^* A x.$$

反之, 若 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 正定矩阵, 则

$$f(x,y) \triangleq y^* A x$$

是 $\mathbb{C}^n$ 上的一个内积.

(留作练习)

定理的结论在实数域中也成立.

对角占优矩阵

定义 设 $A \in \mathbb{C}^{n\times n}$, 若

$$|a_{ii}| \geq \sum_{j\neq i} |a_{ij}|$$

对所有 $i = 1, 2, \ldots, n$ 都成立, 且至少有一个不等式严格成立, 则称 A 为弱行对角占优, 简称弱对角占优或对角占优. 若所有不等式都严格成立, 则称 A 是严格行对角占优, 简称严格对角占优.

类似地, 可以定义列对角占优 和 严格列对角占优.

严格对角占优矩阵的非奇异性

定理 若 $A \in \mathbb{C}^{n\times n}$ 是严格对角占优矩阵, 则 A 非奇异.

(板书)

如果 A 是严格对角占优的, 则可以给出 $\|A^{-1}\|_1$ 的一个估计.

定理 设 $A \in \mathbb{R}^{n\times n}$ 严格列对角占优, 记

$$\delta \triangleq \min_{1\leq j\leq n} \left(|a_{jj}| - \sum_{i=1}^{n} |a_{ij}| \right) > 0,$$

则有 $\|A^{-1}\|_1 \leq \delta^{-1}$.

(留作练习)

可约矩阵与不可约矩阵

定义 设 $A \in \mathbb{C}^{n\times n}$, 如果存在置换矩阵 P 使得 $PAP^\intercal$ 为块上三角矩阵, 即

$$PAP^\intercal = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \quad A_{11} \in \mathbb{C}^{k\times k}, \quad A_{22} \in \mathbb{C}^{(n-k)\times(n-k)}$$

其中 $1 \leq k < n$, 则称 A 是可约的, 否则就称 A 为不可约的.

不可约的意义

若 A 可约, 即存在置换矩阵 P, 使得 $PAP^{\intercal} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$,

则 $Ax = b$ 等价于 $(PAP^{\intercal})(Px) = Pb$. 记 $y \triangleq Px$, $f \triangleq Pb$, 则

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}.$$

- 因此, 原方程组就转化为两个更小规模的子方程组.
- 对于特征值问题, A 的特征值就是 A_{11} 和 A_{22} 的特征值的并.

如果 A_{22} 或 A_{11} 仍然是可约的, 则可以转化为更小规模的子问题.

可约与对角线元素无关

由于 $PAP^\intercal$ 保持 A 的对角线元素仍然在对角线上, 因此主对角线元素是否为零并不影响矩阵的可约性.

推论 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 不可约, 则 $A+D$ 也不可约, 其中 D 是任意对角矩阵. 特别地, $B\triangleq A-\mathrm{diag}(a_{11},a_{22},\ldots,a_{nn})$ 不可约.

可约矩阵的幂

如果 A 可约, 则对任意正整数 k, 有

$$PA^kP^\intercal = (PAP^\intercal)^k = \begin{bmatrix} A_{11}^k & \tilde{A}_{12}^{(k)} \\ 0 & A_{22}^k \end{bmatrix}.$$

定理 设 $A \in \mathbb{C}^{n\times n}$. 若 A 可约, 则 A^k 也可约. 反之, 若存在一个正整数 k, 使得 A^k 是不可约的, 则 A 也不可约.

不可约矩阵的幂不一定不可约. 如 $A = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$, 但 $A^2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ 可约.

矩阵不可约的充要条件（一）

定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$, 指标集 $\mathbb{Z}_n=\{1,2,\ldots,...,n\}$, 则 A 可约的充要条件是存在非空指标集 $\mathcal{S},\mathcal{T}\subset\mathbb{Z}_n$ 满足 $\mathcal{S}\oplus\mathcal{T}=\mathbb{Z}_n$, 使得

$$a_{ij}=0,\quad i\in\mathcal{S},\quad j\in\mathcal{T}.$$

(留作课外自习, 证明见讲义)

矩阵不可约的充要条件（二）

定理 设 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$, 指标集 $\mathbb{Z}_n=\{1,2,\ldots,n\}$, 则 A 可约的充要条件是存在两个相异的正整数 $k,l\in\mathbb{Z}_n$, 使得对任意指标序列 $\{i_1,i_2,\ldots,i_r\}\subseteq\mathbb{Z}_n$, 都有

$$a_{ki_1}a_{i_1i_2}\cdots a_{i_rl}=0.$$

这里 $r>0$ 是任意正整数. (留作课外自习)

等价描述

矩阵 $A=[a_{ij}]\in\mathbb{C}^{n\times n}$ 不可约的充要条件是: 对任意两个相异正整数 $k,l\in\mathbb{Z}_n$, 都存在一个指标序列 $\{i_1,i_2,\ldots,i_m\}\subseteq\mathbb{Z}_n$ 使得

$$a_{ki_1}a_{i_1i_2}\cdots a_{i_ml}\neq 0.$$

不可约对角占优矩阵

定理 设 $A \in \mathbb{C}^{n\times n}$ 是不可约的弱对角占优矩阵, 则 A 非奇异.

(留作练习, 仿照严格对角占优, 再利用不可约充要条件（一）)

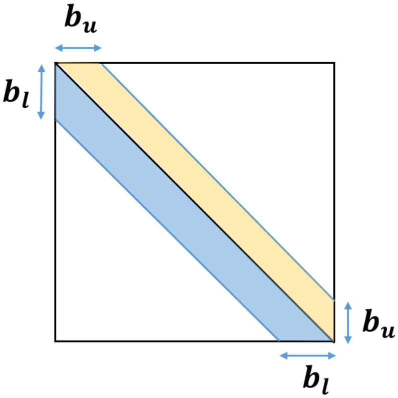
一个有意思的现象

通常, 如果某个元素全不为零的矩阵具有某种性质, 则这个性质往往能够推广到不可约矩阵.

其他常见特殊矩阵

- 带状矩阵:

  $a_{ij} \neq 0$ only if $-b_u \leq i-j \leq b_l$, 其中 b_u 和 b_l 为非负整数, 分别称为下带宽和上带宽, $b_u + b_l + 1$ 称为 A 的带宽



- 上 Hessenberg 矩阵: $a_{ij} = 0$ for $i - j > 1$,

$$
\begin{bmatrix}
* & * & * & \cdots & * \\
* & * & * & \cdots & * \\
 & * & * & \cdots & * \\
 & & \ddots & \ddots & \vdots \\
 & & & * & *
\end{bmatrix}
$$

- 下 Hessenberg 矩阵

$$
\begin{bmatrix}
* & * & & & \\
* & * & * & & \\
\vdots & \vdots & \ddots & \ddots & \\
* & * & \cdots & * & * \\
* & * & \cdots & * & *
\end{bmatrix}
$$

- Toeplitz 矩阵:

$$T = \begin{bmatrix} t_0 & t_{-1} & \cdots & t_{-n+1} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_{n-1} & \cdots & t_1 & t_0 \end{bmatrix}$$

- 循环矩阵 (circulant): 一类特殊的 Toeplitz 矩阵

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

- Hankel 矩阵:

$$H = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n-2} & h_{n-1} \\ h_1 & \iddots & \iddots & \iddots & h_n \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ h_{n-2} & \iddots & \iddots & \iddots & h_{2n-2} \\ h_{n-1} & h_n & \cdots & h_{2n-2} & h_{2n-1} \end{bmatrix}$$

Hankel 矩阵与 Toeplitz 矩阵

设 $J = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}$, 则 JH 是 Toeplitz 矩阵.

6 Kronecker 积

定义 设 $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$, 则 A 与 B 的 **Kronecker 积**定义为

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \in \mathbb{C}^{mp\times nq}.$$

Kronecker 积也称为直积, 或张量积.

任意两个矩阵都存在 Kronecker 积, 但通常 $A \otimes B \neq B \otimes A$. ($A \otimes B$ 和 $B \otimes A$ 有什么关系?)

Kronecker 积基本性质

(1) $(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha(A \otimes B), \quad \forall\, \alpha \in \mathbb{C}$

(2) $(A \otimes B)^{\intercal} = A^{\intercal} \otimes B^{\intercal}, \quad (A \otimes B)^* = A^* \otimes B^*$

(3) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$

(4) $(A + B) \otimes C = A \otimes C + B \otimes C$

(5) $A \otimes (B + C) = A \otimes B + A \otimes C$

(6) $\operatorname{rank}(A \otimes B) = \operatorname{rank}(A) \operatorname{rank}(B)$

(7) **混合积**: $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$

$$(A_1 \otimes A_2 \otimes \cdots \otimes A_k)(B_1 \otimes B_2 \otimes \cdots \otimes B_k) = (A_1B_1) \otimes (A_2B_2) \otimes \cdots \otimes (A_kB_k)$$

$$(A_1 \otimes B_1)(A_2 \otimes B_2) \cdots (A_k \otimes B_k) = (A_1A_2 \cdots A_k) \otimes (B_1B_2 \cdots B_k)$$

Kronecker 积的特征值

推论 设 $A = Q_1\Lambda_1Q_1^{-1} \in \mathbb{C}^{m\times m}$, $B = Q_2\Lambda_2Q_2^{-1} \in \mathbb{C}^{n\times n}$, 则

$$A \otimes B = (Q_1 \otimes Q_2)(\Lambda_1 \otimes \Lambda_2)(Q_1 \otimes Q_2)^{-1}.$$

定理 设 $A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$, 并设 (λ, x) 和 (μ, y) 分别是 A 和 B 的一个特征对, 则 $(\lambda\mu, x \otimes y)$ 是 $A \otimes B$ 的一个特征对. 由此可知, $B \otimes A$ 与 $A \otimes B$ 具有相同的特征值.

推论 设 $A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$, 则 $A \otimes I_n + I_m \otimes B$ 的特征值为 $\lambda_i + \mu_j$, 其中 λ_i 和 μ_j 分别为 A 和 B 的特征值.

Kronecker 积的迹和逆

定理 设 $A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$, 则

(1) $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$;

(2) $\det(A \otimes B) = \det(A)^n \det(B)^m$;

(3) 若 A 和 B 都非奇异, 则 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

Kronecker 积与向量的乘积

定理 设矩阵 $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m\times n}$, 记 $\mathrm{vec}(X)$ 为 X 按列拉成的 mn 维列向量, 即

$$\mathrm{vec}(X) = [x_1^{\intercal}, x_2^{\intercal}, \ldots, x_N^{\intercal}]^{\intercal},$$

则有

$$\mathrm{vec}(AX) = (I \otimes A)\mathrm{vec}(X), \quad \mathrm{vec}(XB) = (B^{\intercal} \otimes I)\mathrm{vec}(X),$$

以及

$$\boxed{(A \otimes B)\mathrm{vec}(X) = \mathrm{vec}(BXA^{\intercal}).}$$

$\mathrm{vec}(X)$ 称为向量化算子或拉直算子. 该结论对节省运算量和存储量都有好处.

定理 矩阵方程

$$AX + XB = D$$

等价于代数方程

$$(I \otimes A + B^{\mathsf{T}} \otimes I)\mathrm{vec}(X) = \mathrm{vec}(D).$$

谢谢

THANK YOU