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: $\mathbb{R}^n, \mathbb{C}^n, \mathbb{R}^{m\times n}$
- 线性相关与线性无关, 秩, 基, 维数
- 线性子空间 (简称子空间)
- 张成子空间: $\mathrm{span}\{x_1, x_2, \ldots, x_k\}$
- 像空间 (值域, 列空间) $\mathrm{Ran}(A)$, 零空间 (核) $\mathrm{Ker}(A)$

直和

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

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

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

定理 设 $\mathcal{S}_1$ 是 $\mathcal{S}$ 的子空间, 则存在另一个子空间 $\mathcal{S}_2$, 使得

$$\mathcal{S} = \mathcal{S}_1 \oplus \mathcal{S}_2$$

例 设 $A \in \mathbb{C}^{m\times n}$, 则

$$\mathbb{C}^n = \mathrm{Ker}(A) \oplus \mathrm{Ran}(A^*), \quad \mathbb{C}^m = \mathrm{Ker}(A^*) \oplus \mathrm{Ran}(A)$$

内积空间

- 内积
- 内积空间 (欧氏空间, 酉空间)

例 常见内积空间:

- $\mathbb{C}^n : (x, y) = y^* x, \quad \mathbb{R}^n : (x, y) = y^\intercal x$
- $\mathbb{R}^{m\times n} : (A, B) = \mathrm{tr}(B^\intercal A)$
- $C[a, b] : (f, g) = \displaystyle\int_a^b f(x)g(x)\,\mathrm{d}x$

正交与正交补

- 正交: 向量之间的正交, 向量与子空间正交, 子空间之间的正交
- 正交补空间: $\mathcal{S} = \mathcal{S}_1 \oplus \mathcal{S}_1^{\perp}$

2 向量范数与矩阵范数

向量范数

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

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

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

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

则称 $f(x)$ 为 $\mathbb{C}^n$ 上的范数, 通常记作 $\|\cdot\|$

✍ 相类似地, 我们可以定义实数空间 $\mathbb{R}^n$ 上的向量范数.

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

- 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: $$\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$$
- p-范数: $$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, \quad 1 \leq p < \infty$$

向量范数的等价性

定义 (范数的等价性)

$\mathbb{C}^n$ 上的向量范数 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 等价: 存在正常数 c_1, c_2, 使得

$$c_1\|x\|_\alpha \le \|x\|_\beta \le c_2\|x\|_\alpha, \quad \forall\, x \in \mathbb{C}^n$$

✍ 该定义可推广到任意赋范线性空间

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

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

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

$$\|x\|_\infty \le \|x\|_1 \le n\,\|x\|_\infty.$$

(证明留作练习)

✍ 事实上, 任意有限维赋范线性空间上的所有范数都是等价的

Cauchy-Schwartz 不等式

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

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

(证明留作练习)

内积与范数

推论 设 $(\cdot,\cdot)$ 是 $\mathbb{C}^n$ 上的内积, 则 $\boxed{\|x\| \triangleq \sqrt{(x,x)}}$ 是一个向量范数.

(证明留作练习)

范数的连续性

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

矩阵范数

定义 (矩阵范数) 若函数 $f:\mathbb{C}^{n\times n}\to\mathbb{R}$ 满足

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

(2) $f(\alpha A)=|\alpha|\cdot f(A), \forall\, A\in\mathbb{C}^{n\times n}, \alpha\in\mathbb{C}$; **(正齐次性)**

(3) $f(A+B)\leq f(A)+f(B), \forall A,B\in\mathbb{C}^{n\times n}$; **(三角不等式)**

则称 $f(x)$ 为 $\mathbb{C}^{n\times n}$ 上的范数, 通常记作 $\|\cdot\|$.

相容的矩阵范数: (4) $f(AB)\leq f(A)f(B), \quad \forall\, A,B\in\mathbb{C}^{n\times n}$

若未明确指出, 讲本课程所涉及矩阵范数都是指 **相容矩阵范数**

算子范数

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

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

是 $\mathbb{C}^{n\times n}$ 上的范数, 称为**算子范数**, 或**诱导范数**, **导出范数**.

✍ 算子范数都是相容的, 且

$$\|Ax\| \leq \|A\| \cdot \|x\|, \quad A \in \mathbb{C}^{n\times n}, x \in \mathbb{C}^n$$

✍ 类似地, 我们可以定义 $\mathbb{R}^{n\times n}, \mathbb{C}^{m\times n}, \mathbb{R}^{m\times n}$ 上的矩阵范数.

常见算子范数及计算公式

引理 可以证明:

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

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

(证明留作练习)

✍ 另一个常用范数 F-范数:

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

矩阵范数的等价性

定理 定义在 $\mathbb{R}^{n\times n}$ 上的所有范数都是等价的, 特别地, 有

$$\frac{1}{\sqrt{n}}\|A\|_2 \le \|A\|_1 \le \sqrt{n}\,\|A\|_2,$$

$$\frac{1}{\sqrt{n}}\|A\|_2 \le \|A\|_\infty \le \sqrt{n}\,\|A\|_2,$$

$$\frac{1}{n}\|A\|_\infty \le \|A\|_1 \le n\,\|A\|_\infty,$$

$$\frac{1}{\sqrt{n}}\|A\|_1 \le \|A\|_F \le \sqrt{n}\,\|A\|_2.$$

(证明留作练习)

矩阵范数的一基本些性质

- 对任意的算子范数 $\|\cdot\|$, 有 $\|I\| = 1$
- 对任意的相容范数 $\|\cdot\|$, 有 $\|I\| \geq 1$
- F-范数不是算子范数
- $\|A^*\|_2 = \|A\|_2$, $\|A^*\|_1 = \|A\|_\infty$
- 若 A 是正规矩阵, 则 $\|A\|_2 = \rho(A)$
- $\|\cdot\|_2$ 和 $\|\cdot\|_F$ 是 **酉不变范数**, 即

$$\|UA\|_2 = \|A\|_2 = \|AV\|_2, \quad \|UA\|_F = \|A\|_F = \|AV\|_F,$$

其中 U, V 为任意酉矩阵.

向量序列的收敛

设 $\{x^{(k)}\}_{k=1}^{\infty}$ 是 $\mathbb{C}^n$ 中的一个向量序列, 如果存在 $x \in \mathbb{C}^n$, 使得

$$\lim_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \ldots, n,$$

则称 $\{x^{(k)}\}$ **(按分量) 收敛到** x, 记为 $\lim\limits_{k\to\infty} x^{(k)} = x$.

向量序列收敛性判断

定理 设 $\|\cdot\|$ 是 $\mathbb{C}^n$ 上的任意一个向量范数, 则 $\lim\limits_{k\to\infty} x^{(k)} = x$ 的充要条件是

$$\lim_{k\to\infty} \|x^{(k)} - x\| = 0.$$

✍ 类似地, 我们可以给出**矩阵序列**的收敛性和判别方法.

收敛速度

设点列 $\{\varepsilon_k\}_{k=1}^{\infty}$ 收敛, 且 $\lim\limits_{k=\infty} \varepsilon_k = 0$. 若存在一个有界常数 $0 < c < \infty$, 使得

$$\lim_{k\to\infty} \frac{|\varepsilon_{k+1}|}{|\varepsilon_k|^p} = c,$$

则称点列 $\{\varepsilon_k\}$ 是 p **次 (渐进) 收敛**的. 若 $1 < p < 2$ 或 $p = 1$ 且 $c = 0$, 则称点列是**超线性收敛**的.

向量序列的收敛速度

令 $\varepsilon_k \triangleq \|x^{(k)} - x\|$, 则 $\lim\limits_{k\to\infty} x^{(k)} = x$ 等价于 $\lim\limits_{k=\infty} \varepsilon_k = 0$.

因此, 向量序列 $\{x^{(k)}\}$ 的收敛速度可以通过 $\{\varepsilon_k\}$ 来定义.

3 矩阵与投影

特征值与特征向量

- 特征多项式, 特征值, 特征向量, 左特征向量, 特征对
- n 阶矩阵 A 的谱: $\sigma(A) \triangleq \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$
- 代数重数和几何重数, 特征空间
- 最小多项式
- 可对角化, 特征值分解
- 可对角化的充要条件

特征值估计：Bendixson 定理

设 $A \in \mathbb{C}^{n\times n}$, 令 $H = \frac{1}{2}(A + A^*)$, $S = \frac{1}{2}(A - A^*)$. 则有

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

$$\lambda_{\min}(iS) \leq \mathrm{Im}(\lambda(A)) \leq \lambda_{\max}(iS),$$

其中 $\mathrm{Re}(\cdot)$ 和 $\mathrm{Im}(\cdot)$ 分别表示实部和虚部.

✍ 一个矩阵的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：圆盘定理

设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$, 定义集合

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

这就是 A 的 n 个 **Gerschgorin 圆盘**.

定理 (Gerschgorin 圆盘定理) 设 $A = [a_{ij}] \in \mathbb{C}^{n\times n}$. 则 A 的所有特征值都包含在 A 的 Gerschgorin 圆盘的并集中, 即 $\sigma(A) \subset \bigcup_{i=1}^{n} \mathcal{D}_i$.

投影

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

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

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

投影变换与投影矩阵

设线性变换 $P:\mathcal{S}\to\mathcal{S}$. 如果对任意 $x\in\mathcal{S}$, 都有

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

则称 P 是从 $\mathcal{S}$ 沿 $\mathcal{S}_2$ 到 $\mathcal{S}_1$ 上的 **投影变换** (或 **投影算子**), 对应的变换矩阵称为 **投影矩阵**.

投影矩阵基本性质

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

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

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

$$\mathbb{R}^n = \mathrm{Ran}(P) \oplus \mathrm{Ker}(P). \tag{1.1}$$

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

投影矩阵由其像空间和零空间唯一确定

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

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

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 $P^2 = P$.

(证明留作练习)

投影算子的矩阵表示

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

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

此时, 我们称 **P 是 $\mathcal{S}_1$ 上与 $\mathcal{S}_2$ 正交**的投影矩阵, 且有

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

其中 $V = [v_1, v_2, \ldots, v_m]$ 和 $W = [w_1, w_2, \ldots, w_m]$ 分别由 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 的基向量组构成.

思考: 设 P 是从 $\mathcal{S}$ 沿 $\mathcal{S}_2$ 到 $\mathcal{S}_1$ 上的投影变换, $V = [v_1, v_2, \ldots, v_m]$ 和 $W = [w_1, w_2, \ldots, w_{n-m}]$ 分别是 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 的一组基, 如何表示 P?

正交投影

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

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

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

- 若 P 是沿 $\mathcal{S}_1^{\perp}$ 到 $\mathcal{S}_1$ 上的投影变换, 则称 P 为 **$\mathcal{S}_1$ 上的正交投影变换** (简称 **正交投影**), 对应的矩阵为 **正交投影矩阵**, 记为 $P_{\mathcal{S}_1}$
- 如果 P 不是正交投影变换, 则称其为 **斜投影变换**

正交投影的矩阵表示

设 P 是 $\mathcal{S}_1$ 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 $\mathcal{S}_1$ 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 $P^\intercal = P$.

(证明留作练习)

正交投影重要性质

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

$$\|P\|_2 = 1,$$

且对 $\forall\, x \in \mathbb{R}^n$, 有

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

正交投影的一个重要应用

定理 设 $\mathcal{S}_1$ 是 $\mathbb{R}^n$ 的子空间, $z \in \mathbb{R}^n$ 是一个向量. 则最佳逼近问题

$$\min_{x \in \mathcal{S}_1} \|x - z\|_2$$

的唯一解为

$$x_* = P_{\mathcal{S}_1} z.$$

即 $\mathcal{S}_1$ 中距离 z 最近 (2-范数意义下) 的向量是 z 在 $\mathcal{S}_1$ 上的正交投影.

(证明留作练习)

✍ 该定理在构造 Krylov 子空间迭代算法时很有用.

推论 设矩阵 $A \in \mathbb{R}^{n\times n}$ 对称正定, 向量 $x_* \in \mathcal{S}_1 \subseteq \mathbb{R}^n$, 则 x_* 是最佳逼近问题

$$\min_{x\in\mathcal{S}_1} \|x - z\|_A$$

的解的充要条件是

$$A(x_* - z) \perp \mathcal{S}_1.$$

这里 $\|x - z\|_A \triangleq \|A^{\frac{1}{2}}(x - z)\|_2$. (证明留作练习)

✍ 该结论可用于证明共轭梯度法所求得的解的最优性.

不变子空间

设 $A \in \mathbb{R}^{n\times n}$, $\mathcal{S}$ 是 $\mathbb{R}^n$ 的一个子空间, 记

$$A\mathcal{S} \triangleq \{Ax : x \in \mathcal{S}\},$$

即 A 关于子空间 $\mathcal{S}$ 的像空间, .

定义 若 $A\mathcal{S} \subseteq \mathcal{S}$, 则称 $\mathcal{S}$ 为 A 的一个不变子空间.

一类重要的不变子空间

定理 设 $x_1, x_2, \ldots, x_m$ 是 A 的一组线性无关特征向量, 则

$$\mathrm{span}\{x_1, x_2, \ldots, x_m\}$$

是 A 的一个 m 维不变子空间. (证明留作练习)

不变子空间的一个重要性质

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

$$AX = XB,$$

此时, B 的特征值都是 A 的特征值. (板书)

✍ 这是设计大规模矩阵特征值计算的算法的数学基础.

推论 设 $A \in \mathbb{R}^{n\times n}$, $X \in \mathbb{R}^{n\times k}$ 且 $\operatorname{rank}(X) = k$. 若存在一个矩阵 $B \in \mathbb{R}^{k\times k}$ 使得 $AX = XB$, 则 (λ, v) 是 B 的一个特征对当且仅当 (λ, Xv) 是 A 的一个特征对.

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

$$X^{-1}AX = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_p \end{bmatrix} \triangleq J,$$

其中 J_i 的维数等于 λ_i 的代数重数, 且具有下面的结构

$$J_i = \begin{bmatrix} J_{i1} & & & \\ & J_{i2} & & \\ & & \ddots & \\ & & & J_{i\nu_i} \end{bmatrix}, \quad J_{ik} = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix},$$

这里 ν_i 为 λ_i 的几何重数, J_{ik} 称为 **Jordan 块**, 每个 Jordan 块对应一个特征向量.

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

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

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 标准型在理论研究中非常有用, 但数值计算比较困难, 目前还没有找到稳定的数值算法.

基于 Jordan 标准型, 我们可以得到下面的非常重要的结论.

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.

Schur 标准型

定理 设 $A \in \mathbb{C}^{n\times n}$, 则存在一个酉矩阵 $U \in \mathbb{C}^{n\times n}$ 使得

$$U^*AU = \begin{bmatrix} \lambda_1 & r_{12} & \cdots & r_{1n} \\ 0 & \lambda_2 & \cdots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \triangleq R \quad \text{或} \quad A = URU^*, \tag{1.2}$$

其中 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 是 A 的特征值 (排序任意). (板书)

关于 Schur 标准型的几点说明

- Schur 标准型可以说是酉相似变化下的最简形式
- U 和 R 不唯一, R 的对角线元素可按任意顺序排列
- A 是正规矩阵当且仅当 (1.2) 中的 R 是对角矩阵;
- A 是 Hermite 矩阵当且仅当 (1.2) 中的 R 是实对角矩阵.

实 Schur 标准型

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

$$Q^\intercal AQ = T,$$

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)

对称正定矩阵

设 $A \in \mathbb{C}^{n\times n}$.

A 是 **半正定** $\Longleftrightarrow \mathrm{Re}(x^*Ax) \geq 0, \forall\, x \in \mathbb{C}^n$

A 是 **正定** $\Longleftrightarrow \mathrm{Re}(x^*Ax) > 0, \forall\, x \in \mathbb{C}^n, x \neq 0$

A 是 **Hermite 半正定** $\Longleftrightarrow$ A Hermite 且半正定

A 是 **Hermite 正定** $\Longleftrightarrow$ A Hermite 且正定

✍ 正定和半正定矩阵不要求是对称或 Hermite 的

对称正定矩阵的判别

定理 设 $A \in \mathbb{C}^{n\times n}$, 则 A 正定 (半正定) 的充要条件是矩阵

$$H = \frac{1}{2}(A + A^*)$$

正定 (半正定). (证明留作练习)

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

$$x^\intercal Ax > 0 \quad (\text{或 } x^\intercal Ax \geq 0),$$

即实矩阵只要针对实向量成立即可。 (证明留作练习)

矩阵平方根

定理 设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 半正定, k 是正整数. 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) $BA = AB$, 且存在一个多项式 $p(t)$ 使得 $B = p(A)$;

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

(3) 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**, 通常记为 $A^{\frac{1}{2}}$.

Hermite 正定矩阵与内积之间的关系

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

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

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

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

是 $\mathbb{C}^n$ 上的一个内积. (证明留作练习)

✍ 该结论说明内积与 Hermite 正定矩阵是一一对应的.

✍ 该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**. 若对所有 $i = 1, 2, \ldots, n$ 不等式都严格成立, 则称 A 是**严格行对角占优**. 通常简称为**弱对角占优**和**严格对角占优**.

✍ 类似地, 可以定义**弱列对角占优** 和 **严格列对角占优**.

可约与不可约

设 $A \in \mathbb{R}^{n\times n}$, 若存在置换矩阵 P, 使得 $PAP^\intercal$ 为块上三角, 即

$$PAP^\intercal = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

其中 $A_{11} \in \mathbb{R}^{k\times k}$ $(1 \leq k < n)$, 则称 A 为**可约**, 否则**不可约**.

✍ 若 A 可约, 则 $Ax = b$ 和 $Ax = \lambda x$ 都可以转化为更小规模问题的求解.

可约与不可约矩阵的判别

定理 设 $A \in \mathbb{R}^{n\times n}$, 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 $J \subset \mathbb{Z}_n$ 且 $J \neq \mathbb{Z}_n$, 使得

$$a_{ij} = 0, \quad i \in J \text{ 且 } j \in \mathbb{Z}_n \setminus J.$$

这里 $\mathbb{Z}_n \setminus J$ 表示 J 在 $\mathbb{Z}_n$ 中的补集.

严格对角占优与不可约弱对角占优矩阵的基本性质

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

(证明见讲义, 留作自习)

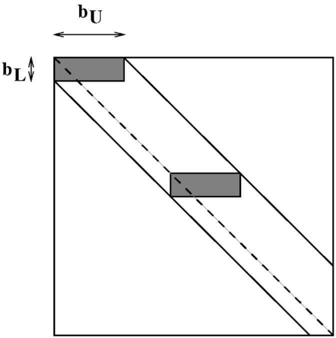
定理 设 $A \in \mathbb{C}^{n\times n}$ 不可约弱对角占优, 则 A 非奇异.

(证明见讲义, 留作自习)

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 $A \otimes B \neq B \otimes A$.

思考: 是否存在置换矩阵 P, 使得 $P^{\mathsf{T}}(A \otimes B)P = B \otimes A$?

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

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

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

特征值

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 $A = Q_1\Lambda_1 Q_1^{-1}$, $B = Q_2\Lambda_2 Q_2^{-1}$, 则

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

定理 设 $A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$, 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 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$;

(4) $A \otimes I_n + I_m \otimes B$ 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;

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

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