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) $\mathrm{Ker}(A)$

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

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

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

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

(证明留作练习)

✍ 任意有限维赋范线性空间上的所有范数都是等价的

内积与范数

定理 (Cauchy-Schwartz 不等式)

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

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

(板书)

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

(证明留作练习)

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)**

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

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

✍ 若未明确指出, 本讲义所涉及矩阵范数都指 **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**.

✍ 对任意算子范数有

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

✍ 类似地, 我们可以定义 $\mathbb{C}^{m\times n}$, $\mathbb{R}^{n\times n}$, $\mathbb{R}^{m\times n}$ 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): $\|A\|_1 = \max\limits_{1\leq j\leq n}\left(\sum\limits_{i=1}^{n}|a_{ij}|\right)$

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

(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

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\|A\|_2 &\le \|A\|_1 \le \sqrt{n}\,\|A\|_2,\\
\frac{1}{\sqrt{n}}\|A\|_2 &\le \|A\|_\infty \le \sqrt{n}\,\|A\|_2,\\
\frac{1}{n}\|A\|_\infty &\le \|A\|_1 \le n\,\|A\|_\infty,\\
\frac{1}{\sqrt{n}}\|A\|_1 &\le \|A\|_F \le \sqrt{n}\,\|A\|_2.
\end{aligned}
$$

(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.

3 矩阵特征值

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 $H = \dfrac{1}{2}(A + A^*)$, $S = \dfrac{1}{2}(A - A^*)$. 则有

$$
\begin{aligned}
\lambda_{\min}(H) \leq \mathrm{Re}(\lambda(A)) \leq \lambda_{\max}(H), \\
\lambda_{\min}(iS) \leq \mathrm{Im}(\lambda(A)) \leq \lambda_{\max}(iS),
\end{aligned}
$$

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 $\sigma(A) \subset \bigcup_{i=1}^{n} \mathcal{D}_i$.

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

$$U^*AU = \begin{bmatrix} \lambda_1 & r_{12} & \cdots & r_{1n} \\ 0 & \lambda_2 & \cdots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \triangleq R \quad \text{或} \quad A = URU^*, \tag{1.4}$$

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 (1.4) 中的 R 是对角矩阵;
- A 是 Hermite 矩阵当且仅当 (1.4) 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

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

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.

5 几类特殊矩阵

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 Hermitian

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) $\operatorname{rank}(B) = \operatorname{rank}(A)$, 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**, 通常记为 $A^{\frac{1}{2}}$.

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 $A_{11} \in \mathbb{R}^{k\times k}$ $(1 \le k < n)$, 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.

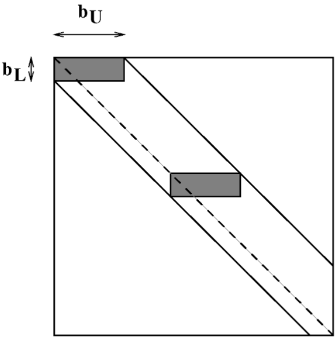
严格对角占优与不可约弱对角占优矩阵的基本性质

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

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

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