目 录

This page intentionally left blank.

1 线性代数基础

1.1 矩阵基础知识

- 行列式及其性质
- 线性相关与线性无关
- Gram-Schmidt 正交化过程
- 矩阵的秩
  设 $A \in \mathbb{R}^{m\times n}$, 则称 A 的列向量组的秩为 A 的**列秩**, 称 A 的行向量组的秩为 A 的**行秩**. 可以验证, 矩阵 A 的列秩与行秩是相等的. 因此我们统一称它们为矩阵 A 的秩, 记为 $\text{rank}(A)$.

定理 1.1

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

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

定理 1.2

设 $A \in \mathbb{R}^{m\times k}$, $B \in \mathbb{R}^{k\times n}$, 则
$$\text{rank}(A) + \text{rank}(B) - k \le \text{rank}(AB) \le \min\{\text{rank}(A),\ \text{rank}(B)\}.$$

- 特征值与特征向量

定义 1.1: 特征多项式

设 $A \in \mathbb{R}^{n\times n}$, 称 $p_A(\lambda) = \det(A - \lambda I)$ 为 A 的**特征多项式**, 其零点就是 A 的**特征值**.

- 特征多项式 $p_A(\lambda)$ 是 n 次多项式, 其零点就是 A 的特征值, 记为 $\lambda_1, \lambda_2, \ldots, \lambda_n$, 即 $p_A(\lambda) = \prod_{i=1}^{n}(\lambda - \lambda_i)$
- 本讲义中, 如果没有特别指出, 我们通常假定

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$$

定义 1.2: 特征值与特征向量

设 $A \in \mathbb{R}^{n\times n}$. 若存在 $\lambda \in \mathbb{C}$ 和非零向量 $x, y \in \mathbb{C}^n$, 满足

$$Ax = \lambda x, \qquad y^* A = \lambda y^* \ (\text{或 } A^* y = \bar{\lambda} y),$$

则称 λ 为 A 的**特征值**, x 为 A 对应于 λ 的 (右) **特征向量**, y 为 A 对应于 λ 的**左特征向量**, 并称 (λ, x) 为 A 的一个**特征对** (**eigenpair**).

关于特征值的几点说明

- 只有当 A 是方阵时, 才具有特征值与特征向量;
- 实矩阵的特征值与特征向量也有可能是复的, 如果实矩阵存在复特征值, 则一定成对出现 (相互共轭);
- n 阶矩阵总是存在 n 个特征值 (其中可能有相等的);
- 特征值有代数重数和几何重数, 几何重数不超过代数重数.

思考: 设 $A \in \mathbb{R}^{n\times n}$, 则 $A^{\intercal}$ 和 A 的特征值和特征向量是什么关系?
A^{-1} 和 A 的特征值和特征向量是什么关系?
更一般地, 设 $p(t)$ 是一个多项式, 则 $p(A)$ 和 A 的特征值与特征向量是什么关系?

- 两类等价关系

定义 1.3: 矩阵相似

设 $A, B \in \mathbb{R}^{n\times n}$, 若存在非奇异矩阵 $X \in \mathbb{R}^{n\times n}$ 使得 $A = XBX^{-1}$, 则称 A 与 B **相似**.

定义 1.4: 矩阵合同

设 $A, B \in \mathbb{R}^{n\times n}$ 都是对称矩阵, 若存在非奇异矩阵 $X \in \mathbb{R}^{n\times n}$ 使得 $A = XBX^{\intercal}$, 则称 A

与 B **合同**.

性质

- 若 A 与 B 相似, 则 A 与 B 有相同的特征值
- 若 A 与 B 合同, 则 A 与 B 有相同的惯性.

- 可对角化

若存在非奇异矩阵 $X \in \mathbb{R}^{n\times n}$, 使得

$$\boxed{A = X\Lambda X^{-1}, \quad \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)} \tag{1.1}$$

则称 A 是**可对角化**的. 若 A 可对角化, 则称 (1.1) 为 A 的**特征值分解**或**谱分解**.

定理 1.3

设 $A \in \mathbb{R}^{n\times n}$.

(1) A 可对角化的充要条件是 A 具有 n 个线性无关的特征向量;

(2) A 可对角化的充要条件是 A 的每个特征值的代数重数和几何重数相等;

(3) 若 A 的特征值互不相同, 则 A 可对角化.

例 1.1 若 $A \in \mathbb{R}^{n\times n}$ 对称, 则 A 可对角化.

- 矩阵的**迹**

定义 1.5: 矩阵的迹

设 $A \in \mathbb{R}^{n\times n}$, 称 $\mathrm{tr}(A) \triangleq a_{11} + a_{22} + \cdots + a_{nn}$ 为矩阵 A 的**迹** (trace).

定理 1.4

设 $A \in \mathbb{R}^{n\times n}$, 则

$$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n, \quad \det(A) = \lambda_1\lambda_2\cdots\lambda_n.$$

- 矩阵的谱与谱半径

所有特征值组成的集合 $\sigma(A)$ 称为 A 的**谱**, 称 $\rho(A)$ 为 A 的**谱半径**, 其中

$$\boxed{\rho(A) \triangleq \max_{\lambda\in\sigma(A)} \{|\lambda|\}, \quad \sigma(A) \triangleq \{\lambda_1, \lambda_2, \ldots, \lambda_n\}}$$

- 特征值的连续性

根据多项式零点关于多项式系数的连续性, 我们可以得到下面的结论.

定理 1.5

矩阵的特征值关于矩阵元素是连续的. 即当矩阵的元素发生变化时, 其特征值的变化是连续的.

- 特征值的估计

定义 1.6: Gerschgorin 圆盘

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

$$\mathcal{D}_i \triangleq \left\{ z \in \mathbb{C} \;:\; |z - a_{ii}| \le \sum_{j=1, j\neq i}^{n} |a_{ij}| \right\}, \quad i = 1, 2, \ldots, n. \tag{1.2}$$

我们称 $\mathcal{D}_i$ 为 A 的 n 个 **Gerschgorin 圆盘**.

定理 1.6: Gerschgorin 圆盘定理

设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$, 则 A 的所有特征值都包含在 A 的 Gerschgorin 圆盘的并集中, 即

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

证明. 设 λ 是 A 的特征值, 对应的非零特征向量为 $x = [x_1, x_2, \ldots, x_n]^\intercal \in \mathbb{C}^n$, 即 $Ax = \lambda x$. 不失一般性, 设 $\|x\|_\infty = |x_i|$, 则 $|x_i| > 0$. 考察 $Ax = \lambda x$ 的第 i 个方程可得

$$\lambda x_i - a_{ii} x_i = \sum_{j=1, j\neq i}^{n} a_{ij} x_j.$$

因此

$$|\lambda - a_{ii}| = \frac{1}{|x_i|} \cdot \left| \sum_{j=1, j\neq i}^{n} a_{ij} x_j \right| \le \sum_{j=1, j\neq i}^{n} |a_{ij}| \cdot \frac{|x_j|}{|x_i|} \le \sum_{j=1, j\neq i}^{n} |a_{ij}|.$$

所以 $\lambda \in \mathcal{D}_i$. □

将 A 的非对角线元素换成 τa_{ij}, 其中 $0 \le \tau \le 1$, 并利用特征值关于矩阵元素的连续性, 我们就可以得到下面的结论.

定理 1.7

设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$, 如果 $\bigcup_{i=1}^{n} \mathcal{D}_i$ 可分解成两个不相交的子集 S 和 T, 即

$$\bigcup_{i=1}^{n} \mathcal{D}_i = \mathsf{S} \bigcup \mathsf{T} \quad 且 \quad \mathsf{S} \bigcap \mathsf{T} = \emptyset.$$

假定 S 由 k 的圆盘组成, 而 T 由其它 $n-k$ 个圆盘组成, 则 S 中恰好包含 A 的 k 个特征

值 (重特征值按重数计算), 而 T 中则包含 A 的其它 $n-k$ 个特征值.

- 矩阵的直和

矩阵 $A_i \in \mathbb{R}^{n_i \times n_i}$ $(i=1,2,\ldots,k)$ 的**直和**定义为

$$A_1 \oplus A_2 \oplus \cdots \oplus A_k \triangleq \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix}, \quad \text{即块对角矩阵}$$

直和的简单性质

- $\det(A \oplus B) = \det(A) \cdot \det(B)$
- $(A \oplus B)^{-1} = A^{-1} \oplus B^{-1}$
- $\alpha(A \oplus B) = (\alpha A) \oplus (\alpha B)$

- 特殊矩阵
  - 单位矩阵, 数量矩阵, 对角矩阵, 三对角矩阵, 上 (下) 三角矩阵, 上 (下) Hessenberg 矩阵, 带状矩阵
  - 置换 (排列) 矩阵, 对称矩阵 (Hermitian 矩阵), 反向单位矩阵, 反向对称矩阵, 斜对称矩阵
  - Vandermonde 矩阵, Toeplitz 矩阵, 循环矩阵, Hankel 矩阵
  - 正定矩阵, 半正定矩阵, 对角占优矩阵, 不可约矩阵
  - 正交矩阵 (酉矩阵), 对合矩阵 ($A^2 = I$), 幂等矩阵 ($A^2 = A$, 也称投影矩阵), 幂零矩阵 ($A^k = 0$)
  - 分块矩阵 (块对角, 块三角, ...)

1.2 对称正定矩阵

定义 1.7

设 $A \in \mathbb{R}^{n\times n}$.

- 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 $A \in \mathbb{R}^{n\times n}$ 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的.

定理 1.8

设 $A \in \mathbb{R}^{n\times n}$, 则

- A 正定的充要条件是对任意非零向量 $x \in \mathbb{R}^n$ 有 $x^\intercal Ax > 0$;
- A 半正定的充要条件是对任意非零向量 $x \in \mathbb{R}^n$ 有 $x^\intercal Ax \geq 0$.

(证明留作练习)

定理 1.9

设 $A \in \mathbb{R}^{n\times n}$ 对称, 则

- A (半) 正定的充要条件是其所有特征值都是 (非负) 正实数;
- A (半) 正定的充要条件是 A 的合同矩阵也 (半) 正定.

对任意 $A \in \mathbb{R}^{n\times n}$, 我们都可以将其写为

$$A = H + S, \quad \text{其中} \quad H = \frac{1}{2}(A + A^\intercal), \quad S = \frac{1}{2}(A - A^\intercal).$$

容易验证 $H^\intercal = H$, $S^\intercal = -S$, 即 H 是对称矩阵, S 是斜对称矩阵.

定理 1.10

设 $A \in \mathbb{R}^{n\times n}$, 则 A 正定 (或半正定) 的充要条件是矩阵 $H = \dfrac{1}{2}(A + A^\intercal)$ 正定 (或半正定).

(板书)

如果 A 是对称 (半) 正定矩阵, 则可以定义其平方根, 即存在唯一的对称 (半) 正定矩阵 B, 使得 $B^2 = A$. 事实上, 我们有下面更一般的性质.

定理 1.11

设 $A \in \mathbb{R}^{n\times n}$ 是一个对称半正定矩阵, k 是一个给定的正整数. 则存在一个唯一的对称半正定矩阵 $B \in \mathbb{C}^{n\times n}$ 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.

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

1.3 向量范数与矩阵范数

1.3.1 向量范数

定义 1.8: 向量范数

若函数 $f(x) : \mathbb{R}^n \to \mathbb{R}$ 满足

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

(2) $f(\alpha x) = |\alpha| \cdot f(x), \forall\, x \in \mathbb{R}^n, \alpha \in \mathbb{R}$ (正齐次性, homogeneity)

(3) $f(x + y) \leq f(x) + f(y), \forall x, y \in \mathbb{R}^n$; (三角不等式, triangular inequality)

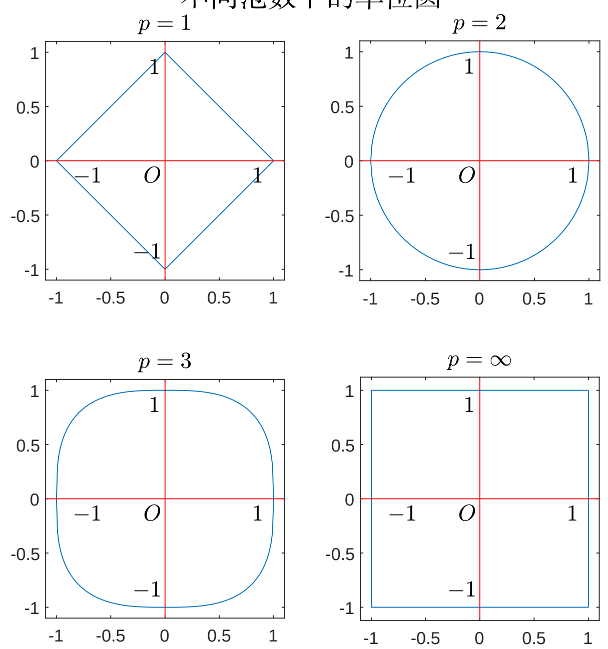
则称 $f(x)$ 为 $\mathbb{R}^n$ 上的 **范数** (**norm**), 通常记作 $\|x\|$.

✍ 为简单起见, 我们这里只考虑实数域情形, 复数域情形完全类似.

例 1.2 $\mathbb{R}^n$ 上的常用范数: 设 $x = [x_1, x_2, \ldots, x_n]^\intercal \in \mathbb{R}^n$,

- 1-范数: $\|x\|_1 = \sum\limits_{i=1}^{n} |x_i|$;
- 2-范数: $\|x\|_2 = \left(\sum\limits_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}$;
- p-范数: $\|x\|_p = \left(\sum\limits_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$; $\quad (p \geq 1)$
- ∞-范数: $\|x\|_\infty = \max\limits_{1 \leq i \leq n} |x_i|$.

不同范数下的单位圆



✍ 1-范数, 2-范数, ∞-范数可以看作是 p-范数在 $p=1,2,\infty$ 时的特殊情形.

✍ 验证 $\|x\|_p$ 满足条件 (3) 时可以利用 Minkowski 不等式

$$\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{1/p} \le \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n}|y_i|^p\right)^{1/p}.$$

下面给出向量范数的一些基本性质.

定理 1.12: 向量范数的连续性

设 $\|\cdot\|$ 是 $\mathbb{R}^n$ 上的一个向量范数, 则 $f(x)\triangleq\|x\|$ 关于 x 的每个分量是连续的.

(证明留作课外练习, 利用范数的三角不等式)

定义 1.9: 范数的等价性

设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 $\mathbb{R}^n$ 空间上的两个向量范数, 若存在正常数 c_1, c_2, 使得

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

对任意 $x\in\mathbb{R}^n$ 都成立, 则称 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是**等价**的.

定理 1.13: 向量范数的等价性

$\mathbb{R}^n$ 上的所有向量范数都是等价的, 特别地, 有

$$\begin{aligned}
&\|x\|_2 \le \|x\|_1 \le \sqrt{n}\,\|x\|_2,\\
&\|x\|_\infty \le \|x\|_1 \le n\,\|x\|_\infty,\\
&\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\,\|x\|_\infty.
\end{aligned}$$

(板书, 以第一个为例, 其余留作练习)

1.3.2 矩阵范数

定义 1.10: 矩阵范数

若函数 $f(X):\mathbb{R}^{n\times n}\to\mathbb{R}$ 满足

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

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

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

(4) $f(AB)\le f(A)f(B), \forall\, A,B\in\mathbb{R}^{n\times n}$;

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

✍ 在矩阵范数的定义中, 条件 (4) 称为**相容性条件**. 在有的教材中, 矩阵范数只需满足条件 (1), (2), (3), 并称满足条件 (4) 的矩阵范数为**相容矩阵范数**. 这样的定义与普通范数保持一致. 但在实际使用中, 有用的矩阵范数基本上都要满足相容性, 所以也通常把矩阵范数直接定义成相容矩阵范数. 如果没有特别说明, 本讲义中矩阵范数就是指相容矩阵范数.

✍ 类似地, 我们可以定义 $\mathbb{R}^{m\times n}$ 上的矩阵范数 (相容性条件会复杂一些).

一类常用的矩阵范数是由向量范数导出的算子范数.

引理 1.14: 算子范数

设 $\|\cdot\|$ 是 $\mathbb{R}^n$ 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 $\mathbb{R}^{n\times n}$ 上的范数, 称为**算子范数**, 也称为**诱导范数**或**导出范数**. (板书)

✍ 对于算子范数, 我们可以立即得到下面的性质: 设 $A \in \mathbb{R}^{n\times n}$, $x \in \mathbb{R}^n$, 则

$$\|Ax\| \leq \|A\| \cdot \|x\|. \tag{1.3}$$

如果没有特别指出哪类范数, 对向量和矩阵使用同样的范数时, 默认是指算子范数.

✍ 如果矩阵范数和向量范数满足 (1.3), 则称它们是**相容**的.

例 1.3 $\mathbb{R}^{n\times n}$ 上常见的矩阵范数:

- Frobenius 范数, 简称 F-范数

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

- p-范数 (算子范数)

$$\|A\|_p \triangleq \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p}, \quad p = 1, 2, \infty.$$

✍ 可以证明: $\|A\|_F^2 = \mathrm{tr}(A^\intercal A)$, 留作练习.

例 1.4 对任意算子范数, 我们都有 $\|I\| = 1$. 但对一般矩阵范数, 该结论不一定成立, 比如 F-范数.

定理 1.15

设 $A \in \mathbb{R}^{n\times n}$, 则

(1) 1-范数 (也称为 **列范数**): $\|A\|_1 = \max_{1\le j\le n}\left(\sum_{i=1}^{n}|a_{ij}|\right)$;

(2) ∞-范数 (也称为 **行范数**): $\|A\|_\infty = \max_{1\le i\le n}\left(\sum_{j=1}^{n}|a_{ij}|\right)$;

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

(板书, 以 ∞-范数和 2-范数为例, 1-范数留作练习)

例 1.5 设 $A = \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix}$, 计算 $\|A\|_1, \|A\|_2, \|A\|_\infty, \|A\|_F$.

(板书)

定理 1.16

若 A 是对称矩阵, 则 $\|A\|_2 = \rho(A)$.

(证明留作练习)

✍ 计算 2-范数时需要求谱半径, 因此通常比计算 1-范数和 ∞-范数更困难. 但在某些情况下可以用下面的范数等价性来估计一个矩阵的 2-范数.

定理 1.17: 矩阵范数的等价性

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

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

(板书)

定理 1.18: 矩阵范数的连续性

设 $\|\cdot\|$ 是 $\mathbb{R}^{n\times n}$ 上的一个矩阵范数, 则 $f(A) \triangleq \|A\|$ 关于 A 的每个分量是连续的.

(证明留作课外自习)

定理 1.19

设 $\|\cdot\|$ 是 $\mathbb{R}^{n\times n}$ 上的任一算子范数, 若 $\|B\|<1$, 则 $I\pm B$ 非奇异, 且

$$\|(I\pm B)^{-1}\| \leq \frac{1}{1-\|B\|}.$$

证明. 这里仅证明减号情形, 加号情形的证明类似.

先证明 $I-B$ 非奇异. 反证法, 假定 $I-B$ 奇异, 则 $(I-B)x=0$ 有非零解, 设为 $\tilde{x}$, 即 $B\tilde{x}=\tilde{x}$. 所以 $\|B\| \geq \dfrac{\|B\tilde{x}\|}{\|\tilde{x}\|}=1$, 与条件 $\|B\|<1$ 矛盾. 因此 $I-B$ 非奇异.

由 $(I-B)^{-1}(I-B)=I$ 可知

$$(I-B)^{-1}=I+(I-B)^{-1}B.$$

从而

$$\|(I-B)^{-1}\| = \|I+(I-B)^{-1}B\| \leq \|I\|+\|(I-B)^{-1}\|\cdot\|B\|.$$

解得

$$\|(I-B)^{-1}\| \leq \frac{\|I\|}{1-\|B\|} = \frac{1}{1-\|B\|}.$$

□

矩阵范数的更多性质

(1) 对任意算子范数 $\|\cdot\|$, 有 $\|A^k\| \leq \|A\|^k$;

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

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

(4) F-范数不是算子范数;

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

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

(6) $\|A^{\intercal}\|_2=\|A\|_2,\quad \|A^{\intercal}\|_1=\|A\|_\infty$.

(证明留作课外自习)

定理 1.20

设 $A\in\mathbb{R}^{n\times n}$, 则对任意算子范数, 有 $\rho(A)\leq\|A\|$. (板书)

✍ 事实上, 上述定理中的结论对任意矩阵范数都成立.

1.4 Kronecker 积

定义 1.11

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

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

✍ Kronecker 积也称为**直积** (direct product) 或**张量积** (tensor product).

✍ 任意两个矩阵都存在 Kronecker 积, 且 $A \otimes B$ 和 $B \otimes A$ 是同阶矩阵, 但通常 $A \otimes B \neq B \otimes A$.

定理 1.21

矩阵的 Kronecker 积有以下性质:

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

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

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

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

(9) $\mathrm{rank}(A \otimes B) = \mathrm{rank}(A)\,\mathrm{rank}(B)$;

定理 1.22

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

定理 1.23

设 $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times n}$, 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) $A \otimes I_n + I_m \otimes B$ 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;

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

推论 1.24

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

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

$A \otimes B$ 与 $B \otimes A$ 通常是不相等的, 但它们之间存在下面的关系式.

定理 1.25

设 $A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$, 则存在 $m + n$ 阶置换矩阵 P 使得

$$P^\intercal(A \otimes B)P = B \otimes A.$$

定理 1.26

设矩阵 $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m\times n}$, 记 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

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

Kronecker 积一个重要应用是可以将某些矩阵方程转化成一般的代数方程.

定理 1.27

矩阵方程

$$AX + XB = D$$

等价于代数方程

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

1.5 课后练习

练习 1.1 (定理 1.13) 证明:

(1) $\|x\|_\infty \leq \|x\|_1 \leq n\,\|x\|_\infty$,

(2) $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\,\|x\|_\infty$.

练习 1.2 设 $A \in \mathbb{R}^{n\times n}$, 证明: $\|A\|_1 = \max\limits_{1\le j\le n}\left(\sum\limits_{i=1}^{n}|a_{ij}|\right)$.

练习 1.3 设 $A \in \mathbb{R}^{n\times n}$, 证明: $\|A\|_F^2 = \mathrm{tr}(A^\intercal A)$.

练习 1.4 设 $A \in \mathbb{R}^{n\times n}$ 是对称矩阵, 证明: $\|A\|_2 = \rho(A)$.

练习 1.5 设矩阵 $A = \begin{bmatrix} 5 & -4 & 2 \\ -4 & 5 & -2 \\ 2 & -2 & 1 \end{bmatrix}$. 计算 $\|A\|_1, \|A\|_2, \|A\|_\infty, \|A\|_F$.

练习 1.6 证明:

(1) $\dfrac{1}{\sqrt{n}}\|A\|_F \le \|A\|_1 \le \sqrt{n}\|A\|_F$;

(2) $\|A\|_2 \le \|A\|_F \le \sqrt{n}\|A\|_2$.

练习 1.7 设 $A \in \mathbb{R}^{n\times n}$, 对任意正交矩阵 $U, V \in \mathbb{R}^{n\times n}$, 证明:

$$\|UAV\|_2 = \|A\|_2, \quad \|UAV\|_F = \|A\|_F.$$

练习 1.8 设 $\|\cdot\|$ 是 $\mathbb{R}^m$ 空间上的一个向量范数, $A \in \mathbb{R}^{m\times n}$, 且 $\mathrm{rank}(A) = n$.
证明: $\|x\|_A \triangleq \|Ax\|$ 是一个向量范数.

练习 1.9 设 $A \in \mathbb{R}^{n\times n}$, 证明:

(1) $\|A\|_2^2 \le \|A\|_1 \cdot \|A\|_\infty$;

(2) $\max\limits_{1\le i,j\le n}\{|a_{ij}|\} \le \|A\|_2 \le n \max\limits_{1\le i,j\le n}\{|a_{ij}|\}$.

练习 1.10 (定理 1.8) 证明: $A \in \mathbb{R}^{n\times n}$ 正定的充要条件是对任意非零向量 $x \in \mathbb{R}^n$ 有 $x^\intercal Ax > 0$.

练习 1.11 设 $A \in \mathbb{R}^{n\times n}$ 正定, 证明: A^{-1} 也正定.

.... **以下为可选题** ....

练习 1.12 设 $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}$, 证明: $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

练习 1.13 设 $A \in \mathbb{R}^{n\times n}$ 是正交矩阵. 证明: $\det(A) = \pm 1$.

练习 1.14 设 $A, B \in \mathbb{R}^{n\times n}$ 都是正交矩阵, 且 $\det(A) = -\det(B)$. 证明: $A + B$ 奇异.

练习 1.15 设 $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ 对称正定, 证明: 当 $i \ne j$ 时有 $a_{ij}^2 < a_{ii}a_{jj}$.

练习 1.16 (Sherman-Morrison 公式) 设 $A \in \mathbb{R}^{n\times n}$ 非奇异, $x, y \in \mathbb{R}^n$.
证明: 若 $y^\intercal A^{-1}x \ne 1$, 则 $A - xy^\intercal$ 可逆, 且

$$(A - xy^\intercal)^{-1} = A^{-1} - \frac{A^{-1}xy^\intercal A^{-1}}{y^\intercal A^{-1}x - 1}.$$

练习 1.17 设 $A \in \mathbb{R}^{n\times n}$ 非奇异, $X, Y \in \mathbb{R}^{n\times m}$ $(m \ge n)$.
证明: 若 $Y^\intercal A^{-1}X - I$ 非奇异, 则 $A - XY^\intercal$ 可逆, 且

$$(A - XY^\intercal)^{-1} = A^{-1} - A^{-1}X(Y^\intercal A^{-1}X - I)^{-1}Y^\intercal A^{-1}.$$

(提示：可利用公式 $B^{-1} = A^{-1} - A^{-1}(B - A)B^{-1}$)

练习 1.18 设 $a \in \mathbb{R}$, 证明:

$$\begin{bmatrix} 1 & -a & & & \\ & 1 & -a & & \\ & & \ddots & \ddots & \\ & & & \ddots & -a \\ & & & & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & a & a^2 & \cdots & a^{n-1} \\ & 1 & a & \ddots & \vdots \\ & & \ddots & \ddots & a^2 \\ & & & \ddots & a \\ & & & & 1 \end{bmatrix}.$$

练习 1.19* 设 $\|\cdot\|$ 是 $\mathbb{R}^n$ 空间上的一个向量范数. 证明: $\|A^{-1}\|^{-1} = \min\limits_{\|x\|=1} \|Ax\|$.

练习 1.20* 设 $A \in \mathbb{R}^{n\times n}$. 试证明对任意矩阵范数都有 $\rho(A) \leq \|A\|$.

练习 1.21* 设 $\|\cdot\|_\alpha$ 是定义在 $\mathbb{C}^{n\times n}$ 上的一个矩阵范数. 证明: 存在 $\mathbb{C}^n$ 上的向量范数 $\|\cdot\|_\beta$, 该范数与 $\|\cdot\|_\alpha$ 相容, 即

$$\|Ax\|_\beta \leq \|A\|_\alpha \|x\|_\beta, \quad \forall\, A \in \mathbb{C}^{n\times n}, x \in \mathbb{C}^n.$$

2 矩阵与线性变换

2.1 线性空间

2.1.1 数域与线性空间

线性空间是线性代数最基本的概念之一, 它是定义在某个数域上并满足一定条件的集合. 我们首先给出数域的概念.

定义 2.1: 数域

设 $\mathbb{F}$ 是包含 0 和 1 的一个数集, 如果 $\mathbb{F}$ 中的任意两个数的和, 差, 积, 商 (除数不为 0) 仍然在 $\mathbb{F}$ 中, 则称 $\mathbb{F}$ 为一个**数域**.

例 2.1 常见的数域有: 有理数域 $\mathbb{Q}$, 实数域 $\mathbb{R}$ 和复数域 $\mathbb{C}$.

✍ 本讲义只考虑实数域和复数域.

定义 2.2: 线性空间

设 $\mathbb{S}$ 是一个非空集合, $\mathbb{F}$ 是一个数域. 在 $\mathbb{S}$ 上定义一种代数运算, 称为**加法**, 记为 "+" (即对任意 $x, y \in \mathbb{S}$, 都存在唯一的 $z \in \mathbb{S}$, 使得 $z = x + y$), 并定义一个从 $\mathbb{F} \times \mathbb{S}$ 到 $\mathbb{S}$ 的代数运算, 称为**数乘**, 记为 "$\cdot$" (即对任意 $\alpha \in \mathbb{F}$ 和任意 $x \in \mathbb{S}$, 都存在唯一的 $y \in \mathbb{S}$, 使得 $y = \alpha \cdot x$). 如果这两个运算满足下面的规则, 则称 $(\mathbb{S}, +, \cdot)$ 是数域 $\mathbb{F}$ 上的一个**线性空间** (通常简称 $\mathbb{S}$ 是数域 $\mathbb{F}$ 上的一个线性空间):

- 加法四条规则
  (1) **交换律**: $x + y = y + x, \quad \forall\, x, y \in \mathbb{S}$;
  (2) **结合律**: $(x + y) + z = x + (y + z), \quad \forall\, x, y, z \in \mathbb{S}$;
  (3) **零元素**: 存在一个元素 0, 使得 $x + 0 = x, \quad \forall\, x \in \mathbb{S}$;
  (4) **逆运算**: 对任意 $x \in \mathbb{S}$, 都存在**负元素** $y \in \mathbb{S}$, 使得 $x + y = 0$, 记 $y = -x$;
- 数乘四条规则
  (1) **单位元**: $1 \cdot x = x, \quad 1 \in \mathbb{F}, \forall\, x \in \mathbb{S}$;
  (2) **结合律**: $\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x, \quad \forall\, \alpha, \beta \in \mathbb{F},\ x \in \mathbb{S}$;

(3) **分配律**: $(\alpha+\beta)\cdot x = \alpha\cdot x + \beta\cdot x, \quad \forall\, \alpha,\beta\in\mathbb{F},\ x\in\mathbb{S}$;

(4) **分配律**: $\alpha\cdot(x+y) = \alpha\cdot x + \alpha\cdot y, \quad \forall\, \alpha\in\mathbb{F},\ x,y\in\mathbb{S}$.

为了表示方便, 通常省略数乘符号, 即将 $\alpha\cdot x$ 写成 αx.

例 2.2 常见的线性空间有:

- $\mathbb{R}^n \to$ 所有 n 维实向量组成的集合, 是 $\mathbb{R}$ 上的线性空间.
- $\mathbb{C}^n \to$ 所有 n 维复向量组成的集合, 是 $\mathbb{C}$ 上的线性空间.
- $\mathbb{R}^{m\times n} \to$ 所有 $m\times n$ 阶实矩阵组成的集合, 是 $\mathbb{R}$ 上的线性空间.
- $\mathbb{C}^{m\times n} \to$ 所有 $m\times n$ 阶复矩阵组成的集合, 是 $\mathbb{C}$ 上的线性空间.
- $C[a,b] \to$ 区间 $[a,b]$ 上所有连续函数组成的集合.

基和维数

设 $x_1,x_2,\ldots,x_n$ 是 $\mathbb{S}$ 中的一组线性无关的向量, 如果 $\mathbb{S}$ 中的任意一个向量都可以由 $x_1,x_2,\ldots,x_n$ 线性表示, 则称 $x_1,x_2,\ldots,x_n$ 是 $\mathbb{S}$ 的一组**基**, 并称 $\mathbb{S}$ 是 n 维的, 即 $\mathbb{S}$ 的**维数**为 n, 记为

$$\dim(\mathbb{S}) = n.$$

如果 $\mathbb{S}$ 中可以找到任意多个线性无关向量, 则称 $\mathbb{S}$ 是**无限维**的.

例 2.3 $\mathbb{R}^n$ 和 $\mathbb{R}^{n\times n}$ 都是有限维的, 维数分别是 n 和 n^2, 而 $C[a,b]$ 是无限维的.

注记

为了讨论方便, 我们假定本讲所讨论的线性空间都是有限维的.

定理 2.1

设 $\mathbb{S}$ 是 n 维的线性空间, $x_1,x_2,\ldots,x_k$ $(1\le k<n)$ 是 $\mathbb{S}$ 中的一组线性无关向量, 则 $x_1,x_2,\ldots,x_k$ 可扩充成 $\mathbb{S}$ 的一组基.

2.1.2 子空间

设 $\mathbb{S}$ 是一个线性空间, $\mathbb{U}$ 是 $\mathbb{S}$ 的一个非空子集合. 如果 $\mathbb{U}$ 关于 $\mathbb{S}$ 上的加法和数乘也构成一个线性空间, 则称 $\mathbb{U}$ 为 $\mathbb{S}$ 的一个**线性子空间**, 简称**子空间**.

例 2.4 设 $\mathbb{S}$ 是一个线性空间, 则由零向量组成的子集 $\{0\}$ 是 $\mathbb{S}$ 的一个子空间, 称为零子空间. 另外, $\mathbb{S}$ 本身也是 $\mathbb{S}$ 的子空间. 这两个特殊的子空间称为 $\mathbb{S}$ 的**平凡子空间**, 其他子空间都是**非平凡子空间**.

下面给出子空间的判别定理.

定理 2.2: 子空间判别定理

设 $\mathbb{S}$ 是数域 $\mathbb{F}$ 上的一个线性空间, $\mathbb{U}$ 是 $\mathbb{S}$ 的一个非空子集合, 则 $\mathbb{U}$ 是 $\mathbb{S}$ 的一个子空间的充要条件是 $\mathbb{U}$ 关于加法和数乘封闭, 即

(1) 对任意 $x, y \in \mathbb{U}$, 有 $x + y \in \mathbb{U}$;

(2) 对任意 $\alpha \in \mathbb{F}$ 和任意 $x \in \mathbb{U}$, 有 $\alpha x \in \mathbb{U}$.

例 2.5 设 $\mathbb{S}$ 是数域 $\mathbb{F}$ 上的一个线性空间, $x_1, x_2, \ldots, x_k \in \mathbb{S}$, 记

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

即由 $x_1, x_2, \ldots, x_k$ 的所有线性组合构成的集合. 可以验证, $\mathrm{span}\{x_1, x_2, \ldots, x_k\}$ 是 $\mathbb{S}$ 的一个线性子空间, 称为由 $x_1, x_2, \ldots, x_k$ **张成的线性空间**.

例 2.6 设 $\mathbb{U}, \mathbb{V}$ 是线性空间 $\mathbb{S}$ 的两个子空间, 则 $\mathbb{U} + \mathbb{V}$ 和 $\mathbb{U} \cap \mathbb{V}$ 也是子空间.

(证明留作课外自习)

下面是关于子空间的维数的一个重要性质.

定理 2.3: 维数公式

设 $\mathbb{U}, \mathbb{V}$ 是线性空间 $\mathbb{S}$ 的两个有限维子空间, 则 $\mathbb{U} + \mathbb{V}$ 和 $\mathbb{U} \cap \mathbb{V}$ 也都是 $\mathbb{S}$ 的子空间, 且

$$\dim(\mathbb{U} + \mathbb{V}) + \dim(\mathbb{U} \cap \mathbb{V}) = \dim(\mathbb{U}) + \dim(\mathbb{V}).$$

(板书)

2.1.3 直和与补空间

设 $\mathbb{U}, \mathbb{V}$ 是线性空间 $\mathbb{S}$ 的两个子空间, 如果 $\mathbb{U} + \mathbb{V}$ 中的任一元素 x 都可以唯一表示成

$$x = x_1 + x_2, \quad x_1 \in \mathbb{U}, x_2 \in \mathbb{V},$$

则称 $\mathbb{U} + \mathbb{V}$ 为**直和**, 记为 $\mathbb{U} \oplus \mathbb{V}$.

关于子空间的直和的判定, 有下面的结论.

定理 2.4: 直和的充要条件

设 $\mathbb{U}, \mathbb{V}$ 是线性空间 $\mathbb{S}$ 的两个子空间, 则下面的论述等价:

(1) $\mathbb{U} + \mathbb{V}$ 是直和;

(2) $\mathbb{U} + \mathbb{V}$ 中的零元素表示方法唯一, 即若 $0 = x_1 + x_2$, $x_1 \in \mathbb{U}$, $x_2 \in \mathbb{V}$, 则 $x_1 = x_2 = 0$;

(3) $\mathbb{U} \cap \mathbb{V} = \{0\}$;

(4) $\dim(\mathbb{U} + \mathbb{V}) = \dim(\mathbb{U}) + \dim(\mathbb{V})$.

(板书)

定理 2.5: 补空间

设 $\mathbb{U}$ 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 $\mathbb{V}$, 使得

$$\mathbb{S} = \mathbb{U} \oplus \mathbb{V}.$$

我们称 $\mathbb{V}$ 为 $\mathbb{U}$ 的**补子空间**, 简称**补空间**. 易知, $\mathbb{U}$ 和 $\mathbb{V}$ 互为补空间. (证明留作课外自习, 将 $\mathbb{U}$ 的一组基扩充成 $\mathbb{S}$ 的一组基即可)

思考: 如何计算补空间? 补空间是否唯一?

类似地, 可以定义多个子空间的直和:

$$\mathbb{U}_1 \oplus \mathbb{U}_2 \oplus \cdots \oplus \mathbb{U}_k.$$

2.2 线性变换

在本节中, 我们总是假定 $\mathbb{U}$ 和 $\mathbb{V}$ 是同一个数域 $\mathbb{F}$ 上的两个线性空间.

2.2.1 线性变换的定义

定义 2.3

设 $f : \mathbb{U} \to \mathbb{V}$ 是一个映射, 如果 f 满足:

(1) $f(x + y) = f(x) + f(y), \quad \forall\, x, y \in \mathbb{U};$ (可加性)

(2) $f(\alpha x) = \alpha f(x), \quad \forall\, x \in \mathbb{U},\ \alpha \in \mathbb{F}.$ (齐次性)

则称 f 是一个**线性映射**或**线性变换**.

以上两个性质 (可加性和齐次性) 可统一写为

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y), \quad \forall\, x, y \in \mathbb{U},\ \alpha, \beta \in \mathbb{F}.$$

有的教材中当 $\mathbb{V} = \mathbb{U}$ 时才将 f 称为**线性变换**或**线性算子**, 本讲义不做这方面的区分.

例 2.7 设 $\mathbb{U} = \mathbb{V} = \mathbb{R}^n$, 线性变换 $f : \mathbb{R}^n \to \mathbb{R}^n$.

- **零变换**: $f(x) = 0, \forall\, x \in \mathbb{R}^n$;
- **数乘变换**: $f(x) = \alpha x, \forall\, x \in \mathbb{R}^n$, 其中 α 为给定的非零实数;

- **恒等变换**: $f(x) = x, \forall\, x \in \mathbb{R}^n$.

线性变换基本性质

设 $f : \mathbb{U} \to \mathbb{V}$ 是线性变换, 则

(1) $f(0) = 0$;

(2) $f(-x) = -f(x), \forall\, x \in \mathbb{U}$;

(3) $f(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k) = \alpha_1 f(x_1) + \alpha_2 f(x_2) + \cdots + \alpha_k f(x_k), \forall\, x_i \in \mathbb{U}, \alpha_i \in \mathbb{F}$.

✍ 设 f 是线性变换, 若 $x_1, x_2, \ldots, x_k$ 线性相关, 则 $f(x_1), f(x_2), \ldots, f(x_k)$ 也线性相关.

设 f 和 g 是从 $\mathbb{U}$ 到 $\mathbb{V}$ 的两个线性变换, 则它们的**和**与**乘积**定义为

$$(f+g)(x) = f(x) + g(x), \quad \forall x \in \mathbb{U},$$
$$(fg)(x) = f(g(x)), \quad \forall x \in \mathbb{U}.$$

定理 2.6

设 f 和 g 是从 $\mathbb{U}$ 到 $\mathbb{V}$ 的两个线性变换, 则 $f+g$ 和 fg 也是线性变换.

(证明留作课外自习)

2.2.2 线性变换与矩阵

定理 2.7: 线性变换的构造

设 $u_1, u_2, \ldots, u_n$ 是线性空间 $\mathbb{U}$ 的一组基, $y_1, y_2, \ldots, y_n$ 是 $\mathbb{V}$ 中任意一组向量, 则存在唯一的线性变换 f, 使得 $f(u_i) = y_i, i = 1, 2, \ldots, n$. (板书)

证明. 首先将线性变换 f 构造出来. 设映射 $f : \mathbb{U} \to \mathbb{V}$ 满足 (直接定义即可)

$$f(u_i) = y_i, \quad i = 1, 2, \ldots, n.$$

对任意 $x \in \mathbb{U}$, 由于 $u_1, u_2, \ldots, u_n$ 是 $\mathbb{U}$ 的一组基, 因此可设

$$x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n,$$

定义映射 f 如下:

$$f(x) = f(\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n) = \alpha_1 f(u_1) + \alpha_2 f(u_2) + \cdots + \alpha_n f(u_n).$$

容易验证, f 满足可加性和齐次性, 因此是一个线性映射.

对于**唯一性**, 如果存在另外一个线性变换 g 满足 $g(u_i) = y_i,\ i = 1, 2, \ldots, n$, 则可以证明对任意 $x \in \mathbb{U}$, 都有 $g(x) = f(x)$, 即 $g = f$. □

✍ 由该定理可知, 要确定一个线性变换, 只需知道其在一组基下的像即可.

下面讨论如何用矩阵来表示线性变换. 设 $u_1, u_2, \dots, u_n$ 和 $v_1, v_2, \dots, v_m$ 分别是 $\mathbb{U}$ 和 $\mathbb{V}$ 的一组基, f 是从 $\mathbb{U}$ 到 $\mathbb{V}$ 的线性变换, 且

$$\begin{aligned} f(u_1) &= a_{11}v_1 + a_{21}v_2 + \cdots + a_{m1}v_m, \\ f(u_2) &= a_{12}v_1 + a_{22}v_2 + \cdots + a_{m2}v_m, \\ &\cdots \quad \cdots \quad \cdots \\ f(u_n) &= a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{mn}v_m, \end{aligned}$$

写成矩阵形式为

$$[f(u_1), f(u_2), \dots, f(u_n)] = [v_1, v_2, \dots, v_m]A, \quad \text{其中} \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

我们称 A 为线性变换 f 关于基 $u_1, u_2, \dots, u_n$ 和 $v_1, v_2, \dots, v_m$ 的矩阵.

线性变换与矩阵

由上面的结论可知, 取定 $\mathbb{U}$ 和 $\mathbb{V}$ 的一组基后, 任意线性变换 f 都存在唯一的矩阵 A 与之对应, 因此我们可以直接用矩阵 A 来表示线性变换 f.

为了便于理解, 教材中常常会用花写字体表示线性变换, 直立字体表示对应的矩阵. 在后面的表述中, 我们就采用这种写法, 即

$$\big[\mathcal{A}(u_1), \mathcal{A}(u_2), \dots, \mathcal{A}(u_n)\big] = [v_1, v_2, \dots, v_m]A, \tag{2.1}$$

或

$$\mathcal{A}(u_1, u_2, \dots, u_n) = [v_1, v_2, \dots, v_m]A.$$

我们称 A 为线性变换 $\mathcal{A}$ 在基 $u_1, u_2, \dots, u_n$ 和 $v_1, v_2, \dots, v_m$ 下的矩阵.

如果 $\mathbb{V} = \mathbb{U}$ 且 $v_i = u_i$, $i = 1, 2, \dots, n$, 则称 A 为 $\mathcal{A}$ 在基 $u_1, u_2, \dots, u_n$ 下的矩阵.

坐标

设 $u_1, u_2, \dots, u_n$ 是 $\mathbb{U}$ 的一组基, 对任意 $x \in \mathbb{U}$, 我们可将其表示为

$$x = x_1u_1 + x_2u_2 + \cdots + x_nu_n,$$

我们称 $x_1, x_2, \dots, x_n$ 为 x 在基 $u_1, u_2, \dots, u_n$ 下的坐标.

为方便起见, 我们通常记为

$$x = [x_1, x_2, \dots, x_n]^\intercal.$$

矩阵向量乘法的几何意义

设 $v_1, v_2, \ldots, v_m$ 是 $\mathbb{V}$ 的一组基, $\mathcal{A}$ 是从 $\mathbb{U}$ 到 $\mathbb{V}$ 的线性变换, 则由 (2.1) 可得

$$\mathcal{A}(x) = \left[\mathcal{A}(u_1), \mathcal{A}(u_2), \ldots, \mathcal{A}(u_n)\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [v_1, v_2, \ldots, v_m]Ax,$$

因此 $\mathcal{A}(x)$ 在 $v_1, v_2, \ldots, v_m$ 下的坐标为 Ax. 我们可以简写为

$$\boxed{\mathcal{A}(x) = Ax}$$

这就是矩阵乘法的几何意义.

✍ 需要指出的是, x 和 $\mathcal{A}(x)$ 在不同基下面的坐标可能是不一样的. 事实上, 线性变换 $\mathcal{A}$ 在不同基下的矩阵也不一样.

✍ 若 $\mathbb{V} = \mathbb{U}$, 则 x 和 Ax 就是同一个元素在不同基下的坐标.

推论 2.8

设 $u_1, u_2, \ldots, u_n$ 和 $v_1, v_2, \ldots, v_m$ 分别是 $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 一组基, 则所有从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性变换与所有 $\mathbb{R}^{m\times n}$ 中的矩阵一一对应, 即

$$\mathcal{A}(u_1, u_2, \ldots, u_n) = [v_1, v_2, \ldots, v_m]A.$$

定理 2.9

在相同的基下面, 线性变换的和对应矩阵的和, 线性变换的乘积对应矩阵的乘积, 即

$$(\mathcal{A}+\mathcal{B})(u_1, u_2, \ldots, u_n) = [v_1, v_2, \ldots, v_m](A+B),$$
$$(\mathcal{A}\mathcal{B})(u_1, u_2, \ldots, u_n) = [v_1, v_2, \ldots, v_m](AB).$$

2.2.3 过渡矩阵

设 $x \in \mathbb{U}$, 下面我们讨论 x 在不同基下的坐标之间的关系. 设 $u_1, u_2, \ldots, u_n$ 和 $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$ 是 $\mathbb{U}$ 的两组基, 并设

$$\begin{aligned}
\tilde{u}_1 &= s_{11}u_1 + s_{21}u_2 + \cdots + s_{n1}u_n, \\
\tilde{u}_2 &= s_{12}u_1 + s_{22}u_2 + \cdots + s_{n2}u_n, \\
&\cdots \quad \cdots \quad \cdots \\
\tilde{u}_n &= s_{1n}u_1 + s_{2n}u_2 + \cdots + s_{nn}u_n,
\end{aligned}$$

即

$$[\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n] = [u_1, u_2, \ldots, u_n]S \quad \text{其中} \quad S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & S_{n2} & \cdots & s_{nn} \end{bmatrix},$$

或

$$[\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n]S^{-1} = [u_1, u_2, \ldots, u_n].$$

我们称 S 为从基 $u_1, u_2, \ldots, u_n$ 到基 $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$ 的**过渡矩阵**.

设 x 在基 $u_1, u_2, \ldots, u_n$ 和 $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$ 下的坐标分别为 $[x_1, x_2, \ldots, x_n]^\intercal$ 和 $[\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n]^\intercal$, 即

$$x = [u_1, u_2, \ldots, u_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n]\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix} = [u_1, u_2, \ldots, u_n]S\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix}.$$

所以有

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = S\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix},$$

即 x 在不同基下的坐标相差一个过渡矩阵.

设 $\mathcal{A}$ 是 $\mathbb{U}$ 到 $\mathbb{V}$ 的线性变换, 则

$$\mathcal{A}(\tilde{u}_i) = s_{11}\mathcal{A}(u_1) + s_{21}\mathcal{A}(u_2) + \cdots + s_{n1}\mathcal{A}(u_n), \quad i = 1, 2, \ldots, n.$$

所以

$$\mathcal{A}(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n) = \mathcal{A}(u_1, u_2, \ldots, u_n)S. \tag{2.2}$$

2.2.4 线性变换在不同基下的矩阵

设 $u_1, u_2, \ldots, u_n$ 和 $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$ 是 $\mathbb{U}$ 的两组基, 从 $u_1, u_2, \ldots, u_n$ 到 $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$ 的过渡矩阵为 $S \in \mathbb{R}^{n\times n}$, 即

$$[\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n] = [u_1, u_2, \ldots, u_n]S. \tag{2.3}$$

设 $v_1, v_2, \ldots, v_m$ 和 $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_m$ 是 $\mathbb{V}$ 的两组基, 从 $v_1, v_2, \ldots, v_m$ 到 $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_m$ 的过渡矩阵为 $T \in \mathbb{R}^{m\times m}$, 即

$$[\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_m] = [v_1, v_2, \ldots, v_m]T \quad \text{或} \quad [\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_m]T^{-1} = [v_1, v_2, \ldots, v_m]. \tag{2.4}$$

设线性变换 $\mathcal{A}$ 在基 $u_1, u_2, \ldots, u_n$ 和 $v_1, v_2, \ldots, v_m$ 下的矩阵为 A, 在基 $\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n$ 和 $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_m$ 下的矩阵为 $\tilde{A}$, 即

$$\mathcal{A}(u_1, u_2, \ldots, u_n) = [v_1, v_2, \ldots, v_m]A, \quad \mathcal{A}(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n) = [\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_m]\tilde{A}. \tag{2.5}$$

于是由 (2.2)-(2.5) 可知

$$\mathcal{A}(\tilde{u}_1,\tilde{u}_2,\ldots,\tilde{u}_n)=\mathcal{A}(u_1,u_2,\ldots,u_n)S=[v_1,v_2,\ldots,v_m]AS=[\tilde{v}_1,\tilde{v}_2,\ldots,\tilde{v}_m]T^{-1}AS.$$

由线性变换与矩阵之间对应关系的唯一性可知

$$\boxed{\tilde{A}=T^{-1}AS}.$$

如果 $\mathbb{U}=\mathbb{V}=\mathbb{R}^n$, 并取 $v_i=u_i$, $\tilde{v}_i=\tilde{u}_i$, $i=1,2,\ldots,n$, 则 $T=S$, 且

$$\boxed{\tilde{A}=S^{-1}AS},$$

即 $\mathcal{A}$ 在两组不同的基下的矩阵相似.

2.3 内积空间

2.3.1 内积与内积空间

内积空间就是带有内积运算的线性空间.

定义 2.4: 内积与内积空间

设 $\mathbb{S}$ 是数域 $\mathbb{F}$ ($\mathbb{C}$ 或 $\mathbb{R}$) 上的一个线性空间, 定义一个从 $\mathbb{S}\times\mathbb{S}$ 到 $\mathbb{F}$ 的代数运算, 记为 “$(\cdot,\cdot)$”, 即对任意 $x,y\in\mathbb{S}$, 都存在唯一的 $f\in\mathbb{F}$, 使得 $f=(x,y)$. 如果该运算满足

(1) $(y,x)=\overline{(x,y)}$, $\quad\forall\, x,y\in\mathbb{S}$;

(2) $(x+y,z)=(x,z)+(y,z)$, $\quad\forall\, x,y,z\in\mathbb{S}$;

(3) $(\alpha x,y)=\alpha(x,y)$, $\quad\forall\,\alpha\in\mathbb{F},\ x,y\in\mathbb{S}$;

(4) $(x,x)\geq 0$, 等号当且仅当 $x=0$ 时成立;

则称 $(\cdot,\cdot)$ 为 $\mathbb{S}$ 上的一个 内积 (**inner product**), 定义了内积的线性空间称为 内积空间.

✍ 内积有时也称为 标量积 (**scalar product**).

✍ 定义在实数域 $\mathbb{R}$ 上的内积空间称为欧氏空间, 定义在复数域 $\mathbb{C}$ 上的内积空间称为酉空间.

✍ $\overline{(x,y)}$ 表示 (x,y) 的共轭. 当 $\mathbb{F}=\mathbb{R}$ 时, 条件 (1) 即为 $(y,x)=(x,y)$.

例 2.8 在线性空间 $\mathbb{R}^n$ 上定义内积

$$(x,y)\triangleq y^{\intercal}x=\sum_{i=1}^{n}x_iy_i,$$

则 $\mathbb{R}^n$ 构成一个内积空间. 这种方式定义的内积有时称为 欧拉内积 (**Euclidean innner product**),

或 **点积** (**dot product**), 或 **标准内积** (**standard inner product**).

✍ 显然, $\mathbb{R}^n$ 上的内积不唯一, 事实上, 任意一个对称正定矩阵都可以定义一个内积, 见练习 2.19.

例 2.9 对任意 $A, B \in \mathbb{R}^{m\times n}$, 定义

$$(A, B) = \mathrm{tr}(B^{\intercal}A),$$

则可以证明 (A, B) 是一个内积, 因此 $\mathbb{R}^{m\times n}$ 构成一个欧氏空间. (证明留作练习)

2.3.2 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 2.10 $\mathbb{R}^n$ 上由标准内积导出的范数为

$$\|x\| = (x, x)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}.$$

这就是 2-范数.

2.3.3 正交与正交补

有了内积以后, 我们就可以定义正交.

定义 2.5: 正交

设 $\mathbb{S}$ 是内积空间, $x, y \in \mathbb{S}$, 如果 $(x, y) = 0$, 则称 x 与 y **正交**, 记为 $x\perp y$;
设 $\mathbb{U}$ 是 $\mathbb{S}$ 的子空间, $x \in \mathbb{S}$, 如果对任意 $y \in \mathbb{U}$ 都有 $(x, y) = 0$, 则称 x 与 $\mathbb{U}$ 正交, 记为 $x\perp\mathbb{U}$;
设 $\mathbb{U}, \mathbb{V}$ 是 $\mathbb{S}$ 的两个子空间, 如果对任意 $x \in \mathbb{U}$, 都有 $x\perp\mathbb{V}$, 则称 $\mathbb{U}$ 与 $\mathbb{V}$ 正交, 记为 $\mathbb{U}\perp\mathbb{V}$.

例 2.11 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

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

则称它们是一组正交基 (orthogonal basis). 如果还满足

$$(x_i, x_i) = x_i^\intercal x_i = 1, \quad i = 1, 2, \ldots, n,$$

则称它们是一组标准正交基或规范正交基 (orthonormal basis).

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 (canonical/standard basis).

✍ 任何一组基都可以通过 Gram-Schmidt 正交化过程构造出一组标准正交基.

定理 2.10

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

(证明留作课外自习)

定义 2.6: 正交补

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

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

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

容易验证, $\mathbb{U}^\perp$ 也是 $\mathbb{S}$ 的一个子空间. 另外, 我们还可以得到下面的结论.

定理 2.11

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

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

(证明留作练习)

例 2.12 设 $A \in \mathbb{R}^{m\times n}$, 则与 A 相关的四个子空间:

- $\mathrm{Ran}(A) \triangleq \{y = Ax \,:\, x \in \mathbb{R}^n\} \to A$ 的像空间, 值域或列空间;
- $\mathrm{Ker}(A) \triangleq \{x \in \mathbb{R}^n \,:\, Ax = 0\} \to A$ 的 零空间或核;
- $\mathrm{Ran}(A^\intercal) \triangleq \{x = A^\intercal y \,:\, y \in \mathbb{R}^m\} \to A$ 的 行空间;
- $\mathrm{Ker}(A^\intercal) \triangleq \{y \in \mathbb{R}^m \,:\, A^\intercal y = 0\} \to A$ 的 左零空间.

其中 $\mathrm{Ran}(A)$ 和 $\mathrm{Ker}(A^\intercal)$ 是 $\mathbb{R}^m$ 的子空间, $\mathrm{Ker}(A)$ 和 $\mathrm{Ran}(A^\intercal)$ 是 $\mathbb{R}^n$ 的子空间. 另外, $\mathrm{Ran}(A) = \mathrm{span}(A)$, 即由 A 的列向量所张成的线性空间.

引理 2.12

设 $A \in \mathbb{R}^{m\times n}$, 则有

- $\dim(\mathrm{Ran}(A)) = \mathrm{rank}(A)$
- $\dim(\mathrm{Ran}(A^\intercal A)) = \dim(\mathrm{Ran}(A^\intercal))$
- $\mathrm{Ran}(A^\intercal A) = \mathrm{Ran}(A^\intercal)$
- $\mathrm{Ker}(A^\intercal A) = \mathrm{Ker}(A)$

(证明留作课外自习)

例 2.13 设 $A \in \mathbb{R}^{m\times n}$, 则

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

证明. 首先证明 $\mathrm{Ker}(A^\intercal) \subseteq \mathrm{Ran}(A)^\perp$. 设 $y \in \mathrm{Ker}(A^\intercal)$, 即 $A^\intercal y = 0$. 设 z 是 $\mathrm{Ran}(A)$ 中的任意一个向量, 则存在 $x \in \mathbb{C}^n$, 使得 $z = Ax$. 于是

$$z^\intercal y = (Ax)^\intercal y = x^\intercal (A^\intercal y) = 0, \quad \forall\, z \in \mathrm{Ran}(A),$$

即 $y \in \mathrm{Ran}(A)^\perp$. 所以 $\mathrm{Ker}(A^\intercal) \subseteq \mathrm{Ran}(A)^\perp$.

另一方面, 设 $y \in \mathrm{Ran}(A)^\perp$, 则对任意向量 $z \in \mathrm{Ran}(A)$, 都有 $y^\intercal z = 0$. 又 $AA^\intercal y \in \mathrm{Ran}(A)$, 所以

$$(A^\intercal y)^\intercal (A^\intercal y) = y^\intercal (AA^\intercal y) = 0.$$

因此 $A^\intercal y = 0$, 即 $y \in \mathrm{Ker}(A^\intercal)$. 所以 $\mathrm{Ran}(A)^\perp \subseteq \mathrm{Ker}(A^\intercal)$. 由此可知, 结论成立. □

✍ 同理可得

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

因此有

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

该结论可用来计算正交补.

定理 2.13: Cauchy-Schwartz 不等式

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

$$|(x,y)|^2 \le (x,x)\cdot(y,y) \quad 或 \quad |(x,y)| \le \|x\|_2 \cdot \|y\|_2,$$

且等号成立的充要条件是 x 与 y 线性相关. (板书)

证明. 若 $y = 0$, 则结论显然成立.

假设 $y \neq 0$, 则对任意 $\alpha \in \mathbb{R}$ 有

$$0 \le (x - \alpha y, x - \alpha y) = (x,x) - 2\alpha(x,y) + \alpha^2(y,y).$$

由于 $y \neq 0$, 所以 $(y,y) > 0$. 取 $\alpha = \dfrac{(x,y)}{(y,y)}$, 代入上式可得

$$0 \leq (x,x) - \frac{(x,y)^2}{(y,y)}, \quad \text{即} \quad |(x,y)|^2 \leq (x,x) \cdot (y,y).$$

下面考虑等号成立的条件.

充分性: 如果 x 与 y 线性相关, 则通过直接验证即可知等号成立.

必要性: 假设等号成立. 如果 $y = 0$, 则显然 x 与 y 线性相关. 现假定 $y \neq 0$. 取 $\alpha = \dfrac{(x,y)}{(y,y)}$, 则

$$(x - \alpha y, x - \alpha y) = (x,x) - \frac{|(x,y)|^2}{(y,y)} = 0,$$

即 $x - \alpha y = 0$. 所以 x 与 y 线性相关. □

更一般地, 我们有下面的 Holder 不等式.

定理 2.14: Holder 不等式

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

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

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

2.3.4 * 对称变换与正交变换

如果 $\mathcal{A}$ 是从 $\mathbb{U}$ 到 $\mathbb{U}$ 的线性变换, 我们简称 $\mathcal{A}$ 是 $\mathbb{U}$ 上的线性变换.

定义 2.7: 对称变换

设 $\mathcal{A}$ 是欧氏空间 $\mathbb{U}$ 上的线性变换, 若 $\mathcal{A}$ 满足

$$(\mathcal{A}(x), y) = (x, \mathcal{A}(y)), \quad \forall x, y \in \mathbb{U},$$

则称 $\mathcal{A}$ 为对称变换.

定理 2.15: 对称变换与对称矩阵

设 $u_1, u_2, \ldots, u_n$ 是欧氏空间 $\mathbb{U}$ 的一组标准正交基, 则 $\mathbb{U}$ 上的线性变换 $\mathcal{A}$ 是对称变换的充要条件是 $\mathcal{A}$ 在这组基下的矩阵 A 为对称矩阵.

定义 2.8: 正交变换

设 $\mathcal{A}$ 是欧氏空间 $\mathbb{U}$ 上的线性变换, 若 $\mathcal{A}$ 满足

$$(\mathcal{A}(x), \mathcal{A}(y)) = (x, y), \quad \forall x, y \in \mathbb{U},$$

则称 $\mathcal{A}$ 为**正交变换**.

定理 2.16

欧氏空间 $\mathbb{U}$ 上的线性变换 $\mathcal{A}$ 是正交变换的充要条件是 $\mathcal{A}$ 是**保距变换**, 即

$$\|\mathcal{A}(x)\| = \|x\|, \quad \forall x \in \mathbb{U},$$

其中 $\|x\| \triangleq \sqrt{(x,x)}$.

定理 2.17: 正交变换与正交矩阵

设 $u_1, u_2, \ldots, u_n$ 是欧氏空间 $\mathbb{U}$ 的一组标准正交基, 则 $\mathbb{U}$ 上的线性变换 $\mathcal{A}$ 是正交变换的充要条件是 $\mathcal{A}$ 在这组基下的矩阵 A 为正交矩阵.

2.4 投影变换

2.4.1 投影变换与投影矩阵

设 $\mathbb{U}$ 和 $\mathbb{V}$ 是内积空间 $\mathbb{S}$ 的两个子空间, 且 $\mathbb{S} = \mathbb{U} \oplus \mathbb{V}$, 则 $\mathbb{S}$ 中的任意向量 x 都可唯一表示为

$$x = x_1 + x_2, \quad x_1 \in \mathbb{U},\ x_2 \in \mathbb{V}.$$

我们称 x_1 为 x 沿 $\mathbb{V}$ 到 $\mathbb{U}$ 上的 **投影**, 记为 $x|_{\mathbb{U}}$.

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

$$Px = x|_{\mathbb{U}}, \quad \forall x \in \mathbb{S}.$$

称 P 是从 $\mathbb{S}$ 沿子空间 $\mathbb{V}$ 到子空间 $\mathbb{U}$ 上的 **投影变换** (也称 **投影算子**), 对应的变换矩阵称为 **投影矩阵**.

- 对于给定的子空间 $\mathbb{U}$ 和 $\mathbb{V}$, 上面定义的投影变换是唯一的.
- 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 $\mathbb{R}^{m\times n}$, 我们采用自然基, 如 $\{e_1, e_2, \ldots, e_n\}$.

设 P 是从 $\mathbb{S}$ 沿 $\mathbb{V}$ 到 $\mathbb{U}$ 上的投影变换, 则对任意 $x \in \mathbb{U}$ 都有 $Px = x$. 因此, $\mathbb{U} \subseteq \mathrm{Ran}(P)$. 又由定义可知 $\mathrm{Ran}(P) \subseteq \mathbb{U}$, 所以

$$\mathbb{U} = \mathrm{Ran}(P).$$

类似地, 我们也可以验证

$$\mathbb{V} = \mathrm{Ker}(P).$$

于是存在直和分解

$$\mathbb{S} = \mathrm{Ran}(P) \oplus \mathrm{Ker}(P).$$

若 $\mathbb{S} = \mathbb{R}^n$, 则立即可以得到下面的结论.

引理 2.18

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

$$\mathbb{R}^n = \mathrm{Ran}(P) \oplus \mathrm{Ker}(P). \tag{2.6}$$

下面的性质表明, 投影矩阵由其像空间和零空间所唯一确定.

引理 2.19

设 $\mathbb{R}^n = \mathbb{U} \oplus \mathbb{V}$, 则存在唯一的投影矩阵 P, 使得

$$\mathrm{Ran}(P) = \mathbb{U}, \quad \mathrm{Ker}(P) = \mathbb{V},$$

即对任意向量 $x \in \mathbb{R}^n$, 有

$$Px \in \mathbb{U}, \quad x - Px \in \mathbb{V}.$$

例 2.14 若 $\mathbb{U} = \mathbb{R}^n$, 则 $\mathbb{V} = \{0\}$, 所对应的唯一投影矩阵即为单位矩阵 I.

反之, 若 $\mathbb{U} = \{0\}$, 则 $\mathbb{V} = \mathbb{R}^n$, 此时所对应的唯一投影矩阵即为零矩阵.

下面给出一个投影矩阵的判别定理.

定理 2.20

矩阵 $P \in \mathbb{R}^{n\times n}$ 是投影矩阵的充要条件是 $P^2 = P$, 即 P 是**幂等矩阵** (**Idempotence**).

(板书)

证明. **必要性**: 设 P 是投影矩阵, 则对任意 $x \in \mathbb{R}^n$, 都有

$$P^2x = P(Px) = Px.$$

因此 $P^2 = P$.

充分性: 设 $P^2 = P$. 我们只需证明 $\mathrm{Ran}(P) + \mathrm{Ker}(P) = \mathbb{R}^n$. 显然 $\mathrm{Ran}(P) + \mathrm{Ker}(P) \subseteq \mathbb{R}^n$, 因此只要证明 $\mathbb{R}^n \subseteq \mathrm{Ran}(P) + \mathrm{Ker}(P)$. 对任意 $x \in \mathbb{R}^n$, 有 $x = Px + (x - Px)$. 由 $P(x - Px) = Px - P^2x = 0$ 可知 $x - Px \in \mathrm{Ker}(P)$. 因此 $\mathbb{R}^n \subseteq \mathrm{Ran}(P) + \mathrm{Ker}(P)$. 所以结论 $\mathrm{Ran}(P) + \mathrm{Ker}(P) = \mathbb{R}^n$ 成立. □

思考: 在证明充分性时, 为什么只需证明 $\mathrm{Ran}(P) + \mathrm{Ker}(P) = \mathbb{R}^n$?

引理 2.21

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) $P^\intercal$ 也是一个投影矩阵.

设 $\mathbb{U}$ 和 $\mathbb{V}$ 是 $\mathbb{R}^n$ 的两个 m 维子空间. 如果 $\mathbb{U}\cap\mathbb{V}^\perp=\{0\}$ (或 $\mathbb{R}^n=\mathbb{U}\oplus\mathbb{V}^\perp$), 则存在唯一的投影矩阵 P, 使得

$$\mathrm{Ran}(P)=\mathbb{U},\quad \mathrm{Ker}(P)=\mathbb{V}^\perp.$$

此时, 我们称 **P 是 $\mathbb{U}$ 上与 $\mathbb{V}$ 正交**的投影矩阵. 令 $v_1,v_2,\ldots,v_m$ 和 $w_1,w_2,\ldots,w_m$ 分别是 $\mathbb{U}$ 和 $\mathbb{V}$ 的一组基, 则 P 可以由这两组基来表示.

定理 2.22

设 $P\in\mathbb{R}^{n\times n}$ 是 $\mathbb{U}$ 上与 $\mathbb{V}$ 正交的投影矩阵, 则

$$P=V(W^\intercal V)^{-1}W^\intercal, \tag{2.7}$$

其中 $V=[v_1,v_2,\ldots,v_m]$, $W=[w_1,w_2,\ldots,w_m]$.

(证明留作练习, 根据投影的唯一性, 只需证明 (2.7) 是 $\mathbb{U}$ 上与 $\mathbb{V}$ 正交的投影)

✍ 虽然投影矩阵 P 由子空间 $\mathbb{U}$ 和 $\mathbb{V}$ 唯一确定, 但其矩阵表示形式 (2.7) 并不唯一.

定理 2.23: 正交投影

设 $\mathbb{U}$ 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

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

其中 x_1 称为 x 在 $\mathbb{U}$ 中的**正交投影**.

2.4.2 正交投影

若 P 是从 $\mathbb{S}$ 沿子空间 $\mathbb{U}^\perp$ 到子空间 $\mathbb{U}$ 上的投影变换, 则称 P 为子空间 $\mathbb{U}$ 上的 **正交投影变换** (也称 **正交投影算子**, 对应的矩阵称为 **正交投影矩阵**), 记为 $P_{\mathbb{U}}$. 如果 P 不是正交投影变换, 则称为**斜投影变换** (**oblique projector**).

由定理 2.22 可立即得到下面的结论.

推论 2.24

设 P 是子空间 $\mathbb{U}$ 上的 **正交投影变换**. 令 $\{v_1,v_2,\ldots,v_m\}$ 是 $\mathbb{U}$ 的一组标准正交基, 则

$$P=VV^\intercal.$$

定理 2.25

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)

下面是关于正交投影变换的一个常用性质, 可以直接通过 2-范数的定义证明.

定理 2.26

设 $P \in \mathbb{R}^{n\times n}$ 是一个正交投影矩阵, 则对 $\forall x \in \mathbb{R}^n$, 有

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

且

$$\|P\|_2 = 1.$$

(证明留作课外自习)

2.4.3 最佳逼近与正交投影

下面是关于正交投影矩阵的一个很重要的应用.

定理 2.27

设 $\mathbb{U}$ 是 $\mathbb{R}^n$ 的一个子空间, $z \in \mathbb{R}^n$ 是一个给定的向量, 则最佳逼近问题

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

的唯一解为

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

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

上述定理中的 2-范数可以推广到一般的能量范数.

推论 2.28

设矩阵 $A \in \mathbb{R}^{n\times n}$ 对称正定, $\mathbb{U}$ 是 $\mathbb{R}^n$ 的一个子空间, 则 x_* 是最佳逼近问题

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

的解的充要条件是

$$A(x_* - z) \perp \mathbb{U}.$$

此处**能量范数** $\|\cdot\|_A$ 的定义为: $\|x-z\|_A \triangleq \sqrt{(x-z)^\intercal A(x-z)}$. (证明留作练习)

2.5 课后练习

练习 2.1 设 $\mathbb{U}$ 和 $\mathbb{V}$ 都是线性空间 $\mathbb{S}$ 的非平凡子空间, 则 $\mathbb{U}\cup\mathbb{V}$ 是否构成 $\mathbb{S}$ 的子空间? 如果结论成立, 则给出证明, 否则给出反例.

练习 2.2 设 $A\in\mathbb{R}^{n\times n}$. 证明: $\mathrm{rank}(A)=1$ 的充要条件是存在非零向量 $a,b\in\mathbb{R}^n$ 使得 $A=ab^\intercal$.

练习 2.3 已知 $x_1=[1,0,1]^\intercal$, $x_2=[1,1,3]^\intercal$, 计算 $\mathrm{span}\{x_1,x_2\}$ 在 $\mathbb{R}^3$ 中的正交补.

练习 2.4 设 $A\in\mathbb{R}^{n\times n}$, 证明 A 可唯一表示为一个对称矩阵和一个**斜对称矩阵**之和, 即

$$A=H+S,\quad 其中\quad H,S\in\mathbb{R}^{n\times n}\quad 满足\quad H^\intercal=H,\quad S^\intercal=-S.$$

练习 2.5 设 $A\in\mathbb{R}^{m\times n}$, 其中 $m\geq n$. 试证明: 矩阵 $\begin{bmatrix} I & A \\ A^\intercal & 0\end{bmatrix}$ 非奇异的充要条件是 $\mathrm{rank}(A)=n$.

练习 2.6 (定理 2.11) 设 $\mathbb{U}$ 是 $\mathbb{S}$ 的子空间, 证明: $\mathbb{U}^\perp$ 也是 $\mathbb{S}$ 的子空间, 且 $\mathbb{S}=\mathbb{U}\oplus\mathbb{V}$.

练习 2.7 (例 2.9) 证明 $(A,B)=\mathrm{tr}(B^\intercal A)$ 是 $\mathbb{R}^{m\times n}$ 上的内积.

练习 2.8 设 $P\in\mathbb{R}^{n\times n}$ 是一个投影矩阵, 且秩为 r, 计算 P 的所有特征值.

练习 2.9 设 P 是沿 $\mathbb{V}$ 到 $\mathbb{U}$ 上的投影变换, $V=[v_1,v_2,\ldots,v_m]$ 构成 $\mathbb{U}$ 的一组基, $W=[w_1,w_2,\ldots,w_{n-m}]$ 构成 $\mathbb{V}$ 的一组基. 证明:

$$P=\begin{bmatrix} W & 0\end{bmatrix}\begin{bmatrix} W & V\end{bmatrix}^{-1}.$$

练习 2.10 (定理 2.22) 设 $P\in\mathbb{R}^{n\times n}$ 是 $\mathbb{U}$ 上与 $\mathbb{V}$ 正交的投影矩阵. 证明:

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

其中 $V=[v_1,v_2,\ldots,v_m]$ 和 $W=[w_1,w_2,\ldots,w_m]$ 的列向量组分别构成 $\mathbb{U}$ 和 $\mathbb{V}$ 的一组基.

练习 2.11 (定理 2.25) 证明: 投影矩阵 $P\in\mathbb{R}^{n\times n}$ 是正交投影矩阵的充要条件 $P^\intercal=P$.

练习 2.12 (推论 2.28) 设矩阵 $A\in\mathbb{R}^{n\times n}$ 对称正定, 向量 $x_*\in\mathbb{U}\subseteq\mathbb{R}^n$. 证明: x_* 是最佳逼近问题

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

的解的充要条件是

$$A(x_*-z)\perp\mathbb{U}.$$

这里 $\|x-z\|_A\triangleq\sqrt{(x-z)^\intercal A(x-z)}$.

以下为可选题

练习 2.13 设 $\mathbb{U}$ 是欧氏空间, $x\in\mathbb{U}$, 定义映射 $\mathcal{A}:\mathbb{U}\to\mathbb{U}$ 如下:

$$\mathcal{A}(y)=y-2(y,x)x,\quad \forall y\in\mathbb{U}.$$

证明: (1) $\mathcal{A}$ 是线性变换; (2) $\mathcal{A}$ 是正交变换.

练习 2.14 设 $\mathcal{A}$ 是欧氏空间 $\mathbb{U}$ 上的线性变换, 若 $\mathcal{A}$ 满足

$$(\mathcal{A}(x), y) = -(x, \mathcal{A}(y)), \quad \forall x, y \in \mathbb{U},$$

则称 $\mathcal{A}$ 为**斜对称变换**. 证明: $\mathcal{A}$ 是斜对称变换的充要条件是 $\mathcal{A}$ 在一组标准正交基下的矩阵 A 为斜对称矩阵.

练习 2.15* 设 $B \in \mathbb{R}^{m\times m}$ 对称半正定, $A \in \mathbb{R}^{m\times n}$, 其中 $m \geq n$. 试证明: 矩阵 $\begin{bmatrix} B & A \\ A^{\intercal} & 0 \end{bmatrix}$ 非奇异的充要条件是 A 列满秩且矩阵 $[B, A]$ 行满秩 (即 A 列满秩且 $\mathrm{Ker}(B) \cap \mathrm{Ker}(A^{\intercal}) = \{0\}$).

练习 2.16* 设 $A \in \mathbb{R}^{n\times n}$ 非奇异, $B \in \mathbb{R}^{n\times n}$ 是幂零矩阵, 则 $A + B$ 是否可逆?

练习 2.17* 设 $A, D \in \mathbb{R}^{n\times n}$ 非奇异, $B, C \in \mathbb{R}^{n\times n}$ 是幂零矩阵, 则 $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ 是否可逆?

练习 2.18* 设 $B \in \mathbb{R}^{m\times n}$ $(m \leq n)$ 是满秩矩阵, $C \in \mathbb{R}^{m\times m}$ 是对称半正定矩阵. 证明:

$$B^{\intercal}(BB^{\intercal})^{-1}B - B^{\intercal}(C + BB^{\intercal})^{-1}B$$

是对称半正定的.

练习 2.19* 设 $A \in \mathbb{R}^{n\times n}$ 是对称正定矩阵, 证明

$$f(x, y) \triangleq y^{\intercal}Ax$$

是 $\mathbb{R}^n$ 上的一个内积. 反之, 设 $(\cdot, \cdot)$ 是 $\mathbb{R}^n$ 上的一个内积, 证明: 存在一个对称正定矩阵 $A \in \mathbb{R}^{n\times n}$ 使得

$$(x, y) = y^{\intercal}Ax.$$

练习 2.20* 设 $J = \begin{bmatrix} \lambda & \varepsilon & & \\ & \ddots & \ddots & \\ & & \ddots & \varepsilon \\ & & & \lambda \end{bmatrix}$, 其中 $\lambda \geq 0, \varepsilon \geq 0$. 证明 $\|J\|_2 \leq \lambda + \varepsilon$.