

线性代数
Linear Algebra

Dait

目 录

第一章	向量和矩阵	1
1.1	向量	1
1.2	矩阵	3
1.3	矩阵的逆	5
1.3.1*	Woodbury 矩阵恒等式	7
1.3.2*	分块矩阵的逆	8
1.4	矩阵的迹	9
第二章	线性方程组	11
2.1	消元法	11
2.2	矩阵的行变换	11
2.3	LU 分解	14
第三章	线性空间	16
3.1	线性空间	16
3.2	线性独立、基和维度	17
3.3	矩阵 A 的四个子空间	18
3.4	矩阵的秩、线性代数基本定理	19
第四章	正交性	22
4.1	正交性	22
4.2	投影	24
4.3	最小二乘法	25
4.4	正交归一基	26
4.4.1	Gram-Schmidt 法则	27
4.4.2	QR 分解	27
第五章	行列式	28
5.1	行列式	28
5.2	行列式的性质	29
5.3	行列式的运算	30
5.4	Cramer 法则、伴随矩阵	32

第六章 特征值和特征向量	35
6.1 特征值和特征向量	35
6.2 特征多项式	36
6.3 矩阵对角化	36
6.4 * Jordan 标准型	38
6.5 对称矩阵	41
6.6 正定矩阵	42
第七章 奇异值分解	45
7.1 奇异值分解	46
7.2 矩阵的模	47
7.3 伪逆	49
7.4 主成分分析	49
第八章 线性映射	51
8.1 线性映射和矩阵	52
8.2 线性映射的性质	53
8.3 基的变换	55
8.4 对偶空间	56
8.5 直和、直积	58
8.6 张量	59
第九章 复线性空间	62
9.1 内积和内积空间	62
9.2 Hermite 矩阵	64
9.3 么正矩阵	65
第十章 群、环、域	66
10.1 二元运算	66
10.2 群与子群	66
10.3 群同态	69
10.4 群同构	71
10.5 等价关系	71

第一章 向量和矩阵

1.1 向量

定义 1.1.1: 数域

给定复数集的子集 $\mathbb{F} \subset \mathbb{C}$, 若满足:

- 非平凡: $0, 1 \in \mathbb{F}$;
- 封闭性: $\forall a, b \in \mathbb{F}$, 有 $a \pm b \in \mathbb{F}$, $ab \in \mathbb{F}$, 且 $a/b \in \mathbb{F}$ (当 $b \neq 0$ 时).

则称 \mathbb{F} 是一个数域 (number field).

例 1.1.1: 数域的例子

- 最小的数域: 有理数域 \mathbb{Q} ;
- 二次数域 $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$;
- 实数域 \mathbb{R} 、复数域 \mathbb{C} .

注. 直到第九章复线性空间之前, 均只考虑实数域 $\mathbb{F} = \mathbb{R}$.

定义 1.1.2: 向量

一个 n 维向量 (vector) v 由 n 个标量 (scalar) $v_1, \dots, v_n \in \mathbb{F}$ 组成, 记作:

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix},$$

组成向量 v 的标量 v_1, \dots, v_n 称为 v 的分量 (component).

所有 n 维向量构成的集合记作 \mathbb{F}^n .

注. 如无特别说明, 向量均默认为列向量, 即 n 行 1 列.

定义 1.1.3: 零向量、反向量

零向量 (zero vector) 是所有分量均为 0 的向量, 记作 0 ;

一个向量 v 的反向量 (opposite vector) 对应每个分量取相反数, 记作 $-v$.

定义 1.1.4: 向量的加法

向量的加法 (addition) 即对应分量相加.

推论. 向量加法的性质: $\forall v, u, w \in \mathbb{F}^n$,

- 交换律: $v + u = u + v$;
- 结合律: $v + (u + w) = (v + u) + w \equiv v + u + w$;
- 零向量: $0 + v = v + 0 = v$;
- 反向量: $v + (-v) = 0$.

定义 1.1.5: 向量的数乘

向量与标量的数乘 (scalar product) 即每个分量乘标量.

推论. 向量数乘的性质: $\forall v, u \in \mathbb{F}^n, c, d \in \mathbb{F}$,

- $1v = v, (-1)v = -v, 0v = 0$.
- 结合律: $c(dv) = (cd)v \equiv cdv$;
- 对标量的分配律: $(c + d)v = cv + dv$;
- 对向量的分配律: $c(v + u) = cv + cu$.

定义 1.1.6: 线性组合

向量 v 和 u 的线性组合 (linear combination) 形如

$$cv + du, \quad \forall c, d \in \mathbb{F}.$$

一般地, n 个向量 v_1, v_2, \dots, v_n 的线性组合形如

$$c_1 v_1 + c_2 v_2 + \dots + c_n v_n, \quad \forall c_i \in \mathbb{F}.$$

注. 线性组合是线性代数中最重要的概念之一.

定义 1.1.7: (实) 向量的内积

两个向量的内积 (inner product) 结果是一个实数 $\langle \cdot, \cdot \rangle : \mathbb{F}^n \times \mathbb{F}^n \rightarrow \mathbb{R}$. 对于实向量

$$\langle v, u \rangle := \sum_{i=1}^n v_i u_i \equiv v_1 u_1 + v_2 u_2 + \dots + v_n u_n. \quad (1.1)$$

有时可以写成点乘形式 $\langle v, u \rangle = v \cdot u$; 特别地, 简记 $v^2 \equiv \langle v, v \rangle$.

推论. 实向量内积的性质: $\forall v, u \in \mathbb{F}^n, c \in \mathbb{F}$,

- 交换律: $\langle v, u \rangle = \langle u, v \rangle$;
- 与数乘的结合律: $\langle cv, u \rangle = \langle v, cu \rangle = c \langle v, u \rangle$;
- 分配律: $\langle v + u, w \rangle = \langle v, w \rangle + \langle u, w \rangle$;
- 正定性: $v^2 \geq 0$, 且 $v^2 = 0 \iff v = 0$.

定义 1.1.8: 向量的长度

向量的长度 (或范数, norm) 可通过内积定义:

$$\|v\| := \sqrt{v^2} = (v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2}. \quad (1.2)$$

长度为 1 的向量是单位向量 (unit vector).

推论. 与向量 v 同向的单位向量是

$$\hat{v} := \frac{v}{\|v\|}, \quad (1.3)$$

1.2 矩阵

定义 1.2.1: 矩阵

一个 m 行 n 列的矩阵 (matrix) A 形如

$$A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \quad A_{ij} \in \mathbb{F}.$$

其中 A_{ij} 是矩阵第 i 行第 j 列的元素. 矩阵 A 的行数和列数分别记作 $\text{row}(A) = m$, $\text{col}(A) = n$. 特别地, 当 $m = n$ 时, 称矩阵是 n 阶方阵 (square matrix).

所有 m 行 n 列矩阵构成的集合记作 $\mathbb{F}^{m \times n}$.

矩阵的加法和数乘与向量的运算规律相同, 是平凡的 (trivial).

定义 1.2.2: 矩阵和向量的乘法 (左乘)

$m \times n$ 矩阵 A 左乘 n 维向量 x , 结果 $b = Ax$ 是一个 m 维向量, 其各分量为:

$$b_i = \sum_{j=1}^n A_{ij}x_j, \quad i = 1, \dots, m. \quad (1.4)$$

注. Ax 可以看成 A 所有列的线性组合, 或者说 A 的各行与 x 分别内积. 这样线性方程组就可以等价地写成 $Ax = b$ 的形式, 即:

$$\begin{cases} A_{11}x_1 + \cdots + A_{1n}x_n = b_1, \\ \cdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n = b_m, \end{cases} \iff \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

这种思想有助于我们掌握线性代数的理念.

定义 1.2.3: 矩阵的乘法

$m \times n$ 矩阵 A 左乘 $n \times p$ 矩阵 B , 结果 AB 是一个 $m \times p$ 的矩阵, 其分量为

$$(AB)_{ij} = \sum_{k=1}^n A_{ik} B_{kj}. \quad (1.5)$$

因此若 A, B 可以相乘, 要求 $\text{col}(A) = \text{row}(B)$; 特别地, 当 A 是方阵时, 简记

$$A^n \equiv \underbrace{A \cdots A}_n.$$

推论. 矩阵乘法的性质: $\forall A, B, C$ 可运算的,

- 左分配律: $(A + B)C = AC + BC$;
- 右分配律: $A(B + C) = AB + AC$.
- 结合律: $(AB)C = A(BC) \equiv ABC$;

证明. 仅给出矩阵乘法结合律的证明: $\forall A \in \mathbb{F}^{m \times n}, B \in \mathbb{F}^{n \times p}, C \in \mathbb{F}^{p \times q}$

$$\begin{aligned} [A(BC)]_{ij} &= \sum_{k=1}^n A_{ik} (BC)_{kj} = \sum_{k=1}^n A_{ik} \sum_{\ell=1}^p B_{k\ell} C_{\ell j} \\ &= \sum_{\ell=1}^p \sum_{k=1}^n A_{ik} B_{k\ell} C_{\ell j} = \sum_{\ell=1}^p (AB)_{i\ell} C_{\ell j} = [(AB)C]_{ij}. \end{aligned}$$

对应元素均相等, 即证. □

注. 矩阵乘法一般不满足交换律, 即一般 $AB \neq BA$, 比如

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

定义 1.2.4: 单位矩阵

n 阶单位矩阵 (unit matrix) 是 n 阶方阵, 对角项均为 1, 其余项均为 0:

$$I_n = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}.$$

推论. $\forall A \in \mathbb{F}^{m \times n}$, 均有

$$I_m A = A I_n = A.$$

定理 1.2.1

若方阵 A 与任意方阵可交换 (commutable), 则 $A = cI$, $c \in \mathbb{F}$

证明. 定义 e_{ij} 表示仅 i 行 j 列为 1, 其余项均为 0. 则

$$(Ae_{ij})_{k\ell} = \sum_{p=1}^n A_{kp}(e_{ij})_{p\ell} = A_{ki}\delta_{j\ell};$$

$$(e_{ij}A)_{k\ell} = \sum_{p=1}^n (e_{ij})_{kp}A_{p\ell} = \delta_{ki}A_{j\ell},$$

当 $k \neq i = j = \ell$ 时, $A_{ki} = 0$; 当 $k = i \neq j = \ell$ 时, $A_{ii} = A_{jj}$. □

定义 1.2.5: 分块矩阵

可以将矩阵分块, 每一块 (block) 是一个小矩阵, 比如

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \end{array} \right] \equiv \begin{bmatrix} A & B \\ & D \end{bmatrix}.$$

注. 分块矩阵乘法: 每个块当作矩阵的元素, 块之间使用矩阵乘法.

定义 1.2.6: 矩阵的转置

矩阵 $A \in \mathbb{F}^{m \times n}$ 的转置 (transpose) 记作 A^\top , $A^\top \in \mathbb{F}^{n \times m}$ 且元素由 A 给出

$$(A^\top)_{ij} = A_{ji}.$$

推论. 矩阵转置的性质: $\forall A, B$ 可运算的,

- $(A^\top)^\top = A$;
- $(A + B)^\top = A^\top + B^\top$, $(cA)^\top = cA^\top$;
- $(AB)^\top = B^\top A^\top$.

1.3 矩阵的逆

定义 1.3.1: 矩阵的逆

方阵 $A \in \mathbb{F}^{n \times n}$ 的逆矩阵 (inverse) $A^{-1} \in \mathbb{F}^{n \times n}$ 满足

$$AA^{-1} = A^{-1}A = I_n.$$

若存在 A^{-1} , 则称矩阵 A 是可逆的; 否则称为不可逆的, 也称奇异的 (singular).

推论. 矩阵逆的性质: $\forall A, B \in \mathbb{F}^{n \times n}$ 可逆,

- $(A^{-1})^{-1} = A$;
- $(cA)^{-1} = c^{-1}A^{-1}$;
- $(AB)^{-1} = B^{-1}A^{-1}$;
- $(A^{-1})^\top = (A^\top)^{-1}$.

注. 一般地, A, B 均可逆 $\not\Rightarrow (A+B)$ 可逆. 除非满足一些额外性质, 见定理 1.3.2.

例 1.3.1: 二阶方阵的逆

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \quad (1.6)$$

例 1.3.2: 对角矩阵

非对角项均为 0 的方阵称为对角矩阵 (diagonal matrix), 对角项为 a_1, \dots, a_n 记作

$$\text{diag}(a_1, \dots, a_n).$$

对角矩阵有很多简单的性质, 比如对角矩阵的乘法很简单

$$\text{diag}(a_1, \dots, a_n) \text{diag}(b_1, \dots, b_n) = \text{diag}(a_1 b_1, \dots, a_n b_n). \quad (1.7)$$

显然, 对角矩阵的逆为

$$\text{diag}(a_1, \dots, a_n)^{-1} = \text{diag}(a_1^{-1}, \dots, a_n^{-1}). \quad (1.8)$$

这对于分块对角矩阵也成立, 即式中 a_i 可以为可逆矩阵.

定理 1.3.1: 左逆和右逆

若 A, B, C 为方阵, 且 $BA = AC = I$, 则 $B = C$.

证明. 考察恒等式 $B(AC) = (BA)C$ 即得 $B = C$. □

注. 事实上, 可以有更强的命题: 若 A, B 为方阵, 则 $BA = I \iff AB = I$.

例 1.3.3: 幂零矩阵

若 A 是幂零矩阵 (nilpotent matrix), 即 $\exists n \in \mathbb{N}$ 使 $A^n = O$, 则 $I + A$ 可逆, 且

$$(I + A)^{-1} = I - A + A^2 - \dots + (-A)^{n-1}.$$

因为

$$(I + A)(I - A + \dots + (-A)^{n-1}) = I + A^n = I.$$

1.3.1* Woodbury 矩阵恒等式

定理 1.3.2: Woodbury 矩阵恒等式

给定 $A \in \mathbb{F}^{n \times n}, U \in \mathbb{F}^{n \times k}, C \in \mathbb{F}^{k \times k}, V \in \mathbb{F}^{k \times n}$ 且 $A, C, (C^{-1} + VA^{-1}U)$ 可逆, 则 $(A + UCV)$ 可逆, 且

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \quad (1.9)$$

证明. 可以直接验证. 也可以对比定理 1.3.4 中的对应块矩阵得到. \square

推论.

- 取 $A = I_n, C = I_k$, 有

$$(I_n + UV)^{-1} = I_n - U(I_k + VU)^{-1}V. \quad (1.10)$$

特别地, 当 $k = 1$ 时, U, V 都是向量:

$$(I - uv^\top)^{-1} = I - \frac{uv^\top}{1 + v^\top u}. \quad (1.11)$$

- 取 $n = k$ 且 $U = V = I_n$, 有

$$(A + B)^{-1} = A^{-1} - A^{-1}(A^{-1} + B^{-1})^{-1}A^{-1} \quad (1.12a)$$

$$= A^{-1} - (A + AB^{-1}A)^{-1}. \quad (1.12b)$$

式 (1.12b) 称为华罗庚恒等式 (Hua's identity).

定理 1.3.3: Sherman-Morrison 公式

给定可逆矩阵 $A \in \mathbb{F}^{n \times n}$ 和向量 $u, v \in \mathbb{F}^n$, 则 $(A + uv^\top)$ 可逆 $\iff 1 + v^\top A^{-1}u \neq 0$, 此时

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}. \quad (1.13)$$

证明. 将式 (1.11) 中的 u 替换为 $A^{-1}u$ 即可. \square

例 1.3.4: 应用

求逆矩阵:

$$A = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{bmatrix}$$

设 $v = [1 \ \cdots \ 1]^\top$, 不难看出 $A = -I + vv^\top$, 故

$$A^{-1} = -I - \frac{vv^\top}{1-n} = \frac{1}{n-1} \begin{bmatrix} 2-n & 1 & \cdots & 1 \\ 1 & 2-n & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2-n \end{bmatrix}.$$

1.3.2* 分块矩阵的逆

给定分块矩阵

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

当 M 满足某些条件时, 其逆可以被表达出来.

定义 1.3.2: Schur 补

当 A 可逆时, 记 $M/A := D - CA^{-1}B$ 是 A 在矩阵 M 中的 Schur 补 (Schur complement); 当 D 可逆时, D 的 Schur 补是 $M/D = A - BD^{-1}C$.

定理 1.3.4: 分块矩阵的逆

当 $A, M/A$ 可逆时, M 可逆, 且

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & -A^{-1}B \\ & I \end{bmatrix} \begin{bmatrix} A^{-1} & \\ & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & \\ -CA^{-1} & I \end{bmatrix} \quad (1.14a)$$

$$= \begin{bmatrix} A^{-1} + A^{-1}B(M/A)^{-1}CA^{-1} & -A^{-1}B(M/A)^{-1} \\ -(M/A)^{-1}CA^{-1} & (M/A)^{-1} \end{bmatrix}; \quad (1.14b)$$

当 $D, M/D$ 可逆时, M 可逆, 且

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} I & \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A - BD^{-1}C)^{-1} & \\ & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ & I \end{bmatrix} \quad (1.15a)$$

$$= \begin{bmatrix} (M/D)^{-1} & -(M/D)^{-1}BD^{-1} \\ -D^{-1}C(M/D)^{-1} & D^{-1} + D^{-1}C(M/D)^{-1}BD^{-1} \end{bmatrix}. \quad (1.15b)$$

当 $A, D, M/A, M/D$ 均可逆时, M 可逆, 且逆可以写成简单的分解:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & \\ & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ -CA^{-1} & I \end{bmatrix}. \quad (1.16)$$

证明. 由于分块对角矩阵的逆是方便求的. 当 A 可逆时, 可以利用 Gauss-Jordan 消元法 (见

第 2.1 节) 将 M 化为分块对角的, 即 LDU 分解 (见定理 2.3.2)

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & \\ & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ & I \end{bmatrix},$$

上式的分块对角矩阵中出现了 A 的 Schur 补 M/A ; 当 D 可逆时, 同理有

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & \\ & D \end{bmatrix} \begin{bmatrix} I & \\ D^{-1}C & I \end{bmatrix},$$

对上两式求逆即得. □

推论. 对比 (1.14b) 和 (1.15b), 可得

$$A^{-1} + A^{-1}B(M/A)^{-1}CA^{-1} = (M/D)^{-1} \quad (1.17a)$$

$$-A^{-1}B(M/A)^{-1} = -(M/D)^{-1}BD^{-1} \quad (1.17b)$$

$$-(M/A)^{-1}CA^{-1} = -D^{-1}C(M/D)^{-1} \quad (1.17c)$$

$$(M/A)^{-1} = D^{-1} + D^{-1}C(M/D)^{-1}BD^{-1} \quad (1.17d)$$

可得 Woodbury 矩阵恒等式 (定理 1.3.2).

推论. 对于比较简单的分块矩阵, 其逆的形式更简单:

- 当 $C = O$ 时, M 可逆 $\iff A, D$ 均可逆, 且

$$\begin{bmatrix} A & B \\ & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ & D^{-1} \end{bmatrix}; \quad (1.18)$$

- 当 $A, D = O$ 时, M 可逆 $\iff B, C$ 均可逆, 且

$$\begin{bmatrix} & B \\ C & \end{bmatrix}^{-1} = \begin{bmatrix} & C^{-1} \\ B^{-1} & \end{bmatrix}. \quad (1.19)$$

1.4 矩阵的迹

定义 1.4.1: 矩阵的迹

n 阶方阵的迹 (trace) 是对角元的和

$$\text{tr}(A) := \sum_{i=1}^n A_{ii}. \quad (1.20)$$

推论. 矩阵迹的性质: $\forall A, B \in \mathbb{F}^{n \times n}$

- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$, $\text{tr}(cA) = c \text{tr}(A)$;
- $\text{tr}(A^\top) = \text{tr}(A)$;
- 两个 n 维列向量 u, v , 有 $\text{tr}(uv^\top) = v^\top u$

$$\text{tr}(uv^\top) = \sum_{i=1}^n u_i v_i = u^\top v = v^\top u.$$

注. 一般地, $\text{tr}(AB) \neq \text{tr}(A) \text{tr}(B)$.

定理 1.4.1: 交换矩阵乘法的迹

$\forall m \times n$ 的矩阵 A 和 $n \times m$ 的矩阵 B , 都有

$$\operatorname{tr}(AB) = \operatorname{tr}(BA). \quad (1.21)$$

证明. 直接展开计算:

$$\begin{aligned} \operatorname{tr}(AB) &= \sum_{i=1}^m (AB)_{ii} = \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ji} \\ &= \sum_{j=1}^n \sum_{i=1}^m B_{ji} A_{ij} = \sum_{j=1}^n (BA)_{jj} = \operatorname{tr}(BA). \end{aligned} \quad \square$$

注. 除此之外, 迹还有一些重要性质, 将在后面讲.

第二章 线性方程组

线性方程 (linear equation) 是未知数最高次数为 1 的方程. 考虑 m 个 n 元线性方程构成的线性方程组 (linear equation set)

$$\sum_{j=1}^n A_{ij}x_j = b_i, \quad i = 1, 2, \dots, m.$$

可以把系数写成系数矩阵 A , 即 $Ax = b$.

2.1 消元法

定义 2.1.1: Gauss 消元法

消元法 (elimination) 就是通过对方程之间倍加消元, 得到一个上三角方程组, 比如

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases} \implies \begin{cases} x - 2y = 1 \\ 8y = 8 \end{cases}$$

而主元 (pivot element) 就是每个方程的第一个非 0 系数.

注. 当主元数目 $<$ 未知数时, 消元法失效, 若最终结果:

- 得到 $0 \neq 0$, 则方程组无解;
- 得到 $0 = 0$, 则方程组有无穷多解.

因此消元法要求: 方程个数与未知数个数相同.

方法 2.1.1: 消元法算法

1. 找到第 1 个 x_1 系数不为 0 的方程并移到最上面;
2. 通过将方程 1 倍加, 从第 2 个到第 n 个方程中消去 x_1 ;
3. 得到的第 2 个到第 n 个方程构成 $(n-1)$ 元的线性方程组, 重复步骤 1.
4. 最后结果如果是一个上三角方程组, 便可从最后一个方程开始解出全部未知数. 否则消元法失效.

2.2 矩阵的行变换

方程中, 置换、倍加、倍乘同时作用在系数矩阵 A 和 b 上, 因此可以写成增广矩阵 (augmented matrix) $[A \ b]$ 并对其消元得到 $[I \ b']$. 类似的, 可以考虑对一般矩阵进行置换、倍加、

倍乘的操作.

定义 2.2.1: 矩阵的初等行变换

- 倍加 (row addition): 一行乘系数加到另一行
- 对换 (row switching): 交换两行
- 倍乘 (row multiplication): 一行乘以一个非零系数

注. 利用 Gauss-Jordan 消元法对增广矩阵 $[A \ I]$ 做消元操作:

$$[A \ I] \rightarrow [I \ A^{-1}].$$

就可以得到 A 的逆 A^{-1} .

定义 2.2.2: 行等价

如果一个矩阵可以经过行变换得到另一个矩阵, 则它们行等价 (row-equivalent).

定义 2.2.3: 行阶梯矩阵

矩阵 M 若其满足以下性质:

- 如果 M 的第 i 行是 0 行, 则下面的所有行的都是 0 行;
- 如果 M 的第 i 行不全是 0, 则从左数第一个非 0 元素叫做主元. 每个主元所在的列都在其上一行的主元的所在列的右边;
- 同一列中在主元下面的元素都是 0.

则称矩阵 M 为行阶梯型 (row echelon form).

比如, 形如 (其中主元 $*$ 非 0, \cdot 任意)

$$\begin{bmatrix} 0 & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & * & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & * & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

就是一个行阶梯矩阵, 其中 $*$ 是主元.

注. 消元法就是把增广矩阵变成行阶梯矩阵的过程.

显然, 行阶梯矩阵并不唯一, 还可以进一步化简.

定义 2.2.4: 约化行阶梯矩阵

约化行阶梯型 (reduced row echelon form) 矩阵还满足以下额外性质:

- 每个主元都是 1;
- 主元所在列只有主元非 0, 称为主列, 其他列称为自由列.

若 A 与约化行阶梯矩阵 U 行等价, 记作 $U = \text{rref}(A)$.

按上面的例子，其约化行阶梯矩阵为

$$\begin{bmatrix} 0 & 1 & \cdot & 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & 1 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 1 & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

第 2,4,5 列为主列，其余为自由列。

定理 2.2.1

一个矩阵的约化行阶梯矩阵是唯一的。

证明. 唯一性证明见 Lay 书的附录 A. □

注. 借助约化行阶梯矩阵的概念，我们可归纳出解方程组 $Ax = b$ 的方法：

- 将增广矩阵 $[A \ b]$ 约化为 $[\text{rref}(A) \ b']$;
- 解的存在性：若 $\text{rref}(A)$ 有 0 行，且 b' 对应行元素非 0，则无解；反之有解；
- 解的唯一性：若 $\text{rref}(A)$ 没有自由列，则解唯一。

定义 2.2.5: 初等矩阵

对 $m \times n$ 矩阵 A 行变换 \iff 用 $m \times m$ 初等矩阵 (elementary matrix) 左乘 A .
初等矩阵有以下三种类型：

- 倍加： A 的第 i 行的 a 倍加到第 j 行 ($r'_j = ar_i + r_j$)

$$E_{ij}(a) := \begin{bmatrix} \ddots & & & a \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} = I + ae_{ij},$$

逆： $E_{ij}(a)^{-1} = E_{ij}(-a)$;

- 置换：置换 A 的第 i 行和第 j 行 ($r_i \rightleftharpoons r_j$)

$$E_{ij} := \begin{bmatrix} \ddots & & & \\ & 0 & & 1 \\ & & \ddots & \\ & 1 & & 0 \\ & & & \ddots \end{bmatrix} = I + e_{ij} + e_{ji} - e_{ii} - e_{jj},$$

逆： $E_{ij}^{-1} = E_{ij}$;

- 倍乘: A 的第 i 行乘一个非 0 常数 c ($r'_i = cr_i$)

$$E_i(c) := \begin{bmatrix} \ddots & & \\ & c & \\ & & \ddots \end{bmatrix} = I + (c-1)e_{ii}.$$

逆: $E_i(c)^{-1} = E_i(c^{-1})$.

消元法: 用一系列初等矩阵 E_1, \dots, E_k 左乘 A , 把 A 化简成行阶梯矩阵.

2.3 LU 分解

定义 2.3.1: 上/下三角矩阵

上三角矩阵 (upper triangular matrix) U 是主对角线以下元素都是 0 的方阵

$$U_{ij} = 0, \quad \forall i > j,$$

同理可定义下三角矩阵 (lower ...) L 满足 $L_{ij} = 0, \forall i < j$.

不难注意到, 倍加矩阵和逆矩阵都同时是上/下三角矩阵. 这是 LU 分解的基础.

定理 2.3.1: LU 分解

对于方阵 A 来说, A 的 LU 分解 (lower-upper decomposition/factorization) 是将 A 分解成一个下三角矩阵 L 和一个上三角矩阵 U 的乘积:

$$A = LU, \tag{2.1}$$

有时需要再乘上一个置换矩阵 $PA = LU$, 称为 LUP 分解 (LU factorization with partial pivoting).

证明. U 是和 A 等价的行阶梯矩阵, U 是上三角的. 如果 A 化成行阶梯矩阵 U 的过程中没有置换, 则从 A 到 U 的过程中, 我们只需消去主元下面的元素:

$$E_k \cdots E_2 E_1 A = U.$$

由于 E_i 及其逆 E_i^{-1} 都是下三角的, 故

$$L = E_1^{-1} E_2^{-1} \cdots E_k^{-1},$$

也是下三角的, $A = LU$; 反之, 则存在一个置换矩阵 P , 使得 $PA = LU$. \square

注. LU 分解可以被视为 Gauss 消元法的矩阵形式. 在数值计算上, LU 分解经常被用来解线性方程组, 且在求逆矩阵和计算行列式中都是一个关键的步骤.

定理 2.3.2: LDU 分解

LDU 分解 (lower-diagonal-upper decomposition) 是将矩阵写成

$$A = LDU,$$

其中 D 是对角矩阵, L, U 是单位下/上三角的 (unitriangular), 即其对角元均为 1.

第三章 线性空间

3.1 线性空间

定义 3.1.1: 线性空间

定义域 \mathbb{F} 上的线性空间 (linear space) V 是具有加法 $+: V \times V \rightarrow V$ 和数乘 $\cdot: \mathbb{F} \times V \rightarrow V$ 运算且满足以下公理的集合.

- | | |
|--------------|---------------------------------|
| 1. 加法交换律 | $x + y = y + x;$ |
| 2. 加法结合律 | $x + (y + z) = (x + y) + z;$ |
| 3. 加法零元 | $x + 0 = x;$ |
| 4. 加法逆元 | $x + (-x) = 0;$ |
| 5. 数乘单位元 | $1x = x;$ |
| 6. 数乘结合律 | $(c_1 c_2)x = c_1(c_2 x);$ |
| 7. 数乘对向量的分配律 | $c(x + y) = cx + cy;$ |
| 8. 数乘对标量的分配律 | $(c_1 + c_2)x = c_1 x + c_2 x.$ |

例 3.1.1

\mathbb{R}^n 和 \mathbb{C}^n 都是线性空间.

定义 3.1.2: 子空间

给定线性空间 V 的子集 $V_s \subset V$, 若其对于加法和数乘封闭: $\forall v, w \in V_s, \forall c \in \mathbb{F}$

$$v + w \in V_s, \quad cv \in V_s,$$

则 V_s 为 V 的子空间 (subspace). 子空间中元素的线性组合都在同一个子空间,

推论. 子空间必然包含零向量. 因为若 $v \in V_s$, 则 $v + (-v) = 0 \in V_s$.

注. 线性空间 V 的一个子集一般不是子空间, 但我们可以通过其构造出子空间.

定义 3.1.3: 线性扩张

S 的线性扩张 (linear span) 是 S 中向量的所有线性组合的集合, 记作 $\text{span}(S)$.

推论. V 的子集的线性扩展是 V 的子空间.

3.2 线性独立、基和维度

定义 3.2.1: 线性独立

n 个向量 $\{v_i\}$ 是线性独立的 (linear independent), 当且仅当

$$\sum_{i=1}^n x_i v_i = 0,$$

只在 $x_i = 0$ 时成立, 即只有零解. n 个向量 $\{v_i\}$ 不是线性独立, 那么他们是线性相关的 (linear correlate).

等价描述: 集合中每一个向量都不能写成其它向量的线性组合.

注. 向量是否线性独立同数域的选择有关.

定义 3.2.2: 线性空间的基

线性空间 V 的基 (base) 是一组线性无关的向量 $\{v_i\}$, 并且他们张成整个线性空间 V .

例 3.2.1

$\{e_i\}$ 构成 \mathbb{R}^n 的一组基.

定义 3.2.3: 线性空间的维度

线性空间的维度 (dimension) 是一组基中向量的个数, 记作 $\dim(V)$.

定理 3.2.1: 维度的确定性

线性空间的维度和基的选取无关.

证明. 若线性空间 V 存在两组基 $\{v_1, \dots, v_m\}, \{w_1, \dots, w_n\}$ 元素个数不等, 不妨设 $n > m$. 因为 $\{w_i\}$ 是基, $\{v_i\}$ 可以被表示为其线性组合

$$v_i = \sum_{j=1}^n w_j a_{ji}, \quad \forall i.$$

考虑线性组合

$$\sum_{i=1}^m x_i v_i = \sum_{i=1}^m \sum_{j=1}^n x_i w_j a_{ji} = 0.$$

因为 $\{w_i\}$ 线性无关, 故

$$\sum_{i=1}^m a_{ji} x_i = 0, \quad \forall j,$$

但是其未知数的个数 $m >$ 方程的个数 n , 系数矩阵 A 一定有自由列, 所以 x 有非零解. 这与 $\{v_i\}$ 线性无关矛盾! 故 $m = n$. \square

定理 3.2.2

若 $\forall v \in V_1$ 可以写成 V_2 中向量的线性组合, 则 $\dim(V_1) \leq \dim(V_2)$.

证明. 这个定理是 trivial 的, 证明留给读者. \square

定义 3.2.4: 变换矩阵

不同基之间的变换相应的矩阵称为变换矩阵 (transformation matrix).

3.3 矩阵 A 的四个子空间

给定矩阵 $A \in \mathbb{F}^{m \times n}$, 可由其得到四个子空间: 列空间、行空间、零空间和左零空间.

定义 3.3.1: 列空间和行空间

矩阵 A 的列空间 (column space) 是 A 的所有列的线性组合的集合, 记作 $C(A)$.

矩阵 A 的行空间 (row space) 是 A 的所有行的线性组合的集合, 可用 $C(A^\top)$ 表示.

推论. $C(A)$ 是 \mathbb{F}^m 的子空间, $C(A^\top)$ 是 \mathbb{F}^n 的子空间.

注. 线性方程组 $Ax = b$ 有解 $\iff b \in C(A)$.

定义 3.3.2: 零空间和左零空间

矩阵 A 的零空间 (null space) 是所有满足 $Ax = 0$ 的 x 的集合, 记作 $N(A)$.

矩阵 A 的左零空间 (left null space) 是所有满足 $x^\top A = 0$ 的 x 的集合, 可用 $N(A^\top)$ 表示.

推论. $N(A)$ 是 \mathbb{F}^n 的子空间, $N(A^\top)$ 是 \mathbb{F}^m 的子空间.

例 3.3.1: 子空间的基

$N(A)$ 的基: 显然 $N(A) = N(\text{rref}(A))$, 而 $\text{rref}(A)$ 的每个自由列各给出一个 $N(\text{rref}(A))$ 的基. 若第 j 列为自由列, 则其对应的向量 x 的分量为:

$$x_i = \begin{cases} \delta_{ij}, & \text{第 } i \text{ 列为自由列} \\ -\text{rref}(A)_{ij}, & \text{第 } i \text{ 列为主列} \end{cases}$$

$C(A)$ 的基: $\text{rref}(A)$ 的所有主列构成 $C(A)$ 的一组基.

3.4 矩阵的秩、线性代数基本定理

定义 3.4.1: 矩阵的秩

矩阵 A 的秩 (rank) 定义为行空间的维数:

$$\text{rank}(A) := \dim(C(A)).$$

进而可定义行满秩矩阵 (full row rank matrix) 满足

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

列满秩矩阵 (full column rank matrix) 定义类似.

定理 3.4.1: 线性方程组 $Ax = 0$ 的完整解

线性方程组 $Ax = b$ 的通解可以分解为特解 x_p 和零解 x_n (即满足 $Ax_n = 0$) 的和:

$$x = x_p + x_n,$$

特解 x_p 的各个分量可以由约化的增广矩阵 $[\text{rref}(A) \ b']$ 中得到

$$x_{pi} = \begin{cases} 0, & \text{第 } i \text{ 列为自由列} \\ b'_j, & \text{第 } i \text{ 列为主列, 其主元 } 1 \text{ 在第 } j \text{ 行} \end{cases}$$

推论.

- 若 A 列满秩, 则 $\text{rref}(A)$ 没有自由列, $N(A) = \{0\}$. 当 $b \in C(A)$ 时有唯一解, 否则无解;
- 若 A 行满秩, 则 $\text{rref}(A)$ 没有零行, $C(A) = \mathbb{R}^m$, $\forall b$ 都有解, 有唯一解或无穷多解.

四个子空间的维度 已经知道, 初等行变换就是用初等矩阵 E 左乘 A , 相应的, 列变换就是 E 右乘 A .

定理 3.4.2: 初等变换和子空间

- $N(A) = N(EA)$, 因为

$$Ax = 0 \iff EAx = 0.$$

- $\dim(C(A)) = \dim(C(EA))$, 因为

$$\{v_i\} \text{ 是 } C(A) \text{ 一组基} \iff \{Ev_i\} \text{ 是 } C(EA) \text{ 一组基}.$$

- $C(A) = C(AE)$, 因为

$$AE \text{ 的每一列} \in C(A) \text{ 且 } A = (AE)E^{-1} \text{ 的每一列} \in C(AE)$$

- $\dim(N(A)) = \dim(N(AE))$, 因为可由 $Ax = 0 \iff AE(E^{-1}x) = 0$ 推出

$$\{v_i\} \text{ 是 } N(A) \text{ 一组基} \iff \{E^{-1}v_i\} \text{ 是 } N(AE) \text{ 一组基}$$

推论. 矩阵 A 在初等变换下, $\dim(C(A))$ 和 $\dim(N(A))$ 均不变, 而行变换下 $N(A)$ 不变, 列变换下 $C(A)$ 不变. 因此, 可以将 A 先由行变换为 $\text{rref}(A)$, 再列变换为

$$\tilde{I} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{F}^{m \times n}.$$

显然, \tilde{I} 的行秩 $\dim(C(\tilde{I}^\top)) =$ 列秩 $\dim(C(\tilde{I})) = r$, 且

$$\dim(C(\tilde{I}^\top)) + \dim(N(\tilde{I})) = n;$$

$$\dim(C(\tilde{I})) + \dim(N(\tilde{I}^\top)) = m.$$

定理 3.4.3: 线性代数基本定理 · 一

1. 行秩 = 列秩: $\text{rank}(A) = \dim(C(A)) = \dim(C(A^\top))$;
2. $\dim(C(A^\top)) + \dim(N(A)) = n$;
3. $\dim(C(A)) + \dim(N(A^\top)) = m$.

推论. 方阵 A 可逆 $\iff A$ 满秩.

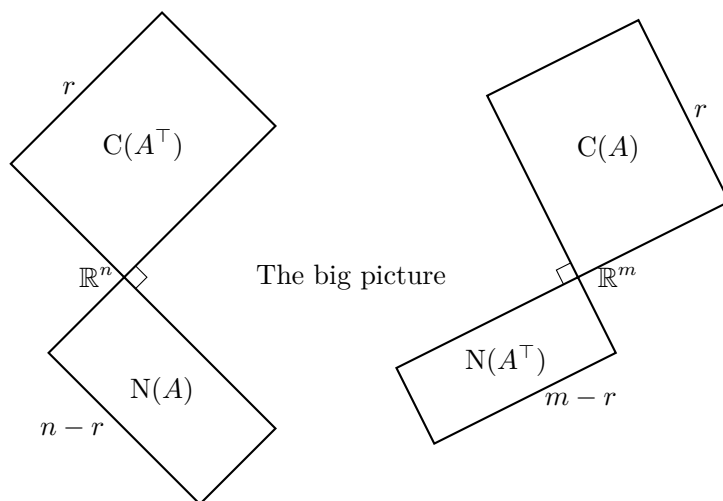


图 3.1: A 的四个子空间

矩阵的秩的不等式 参考: <https://zhuanlan.zhihu.com/p/55206421>

定理 3.4.4

对于矩阵 A, B 有

$$\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B). \quad (3.1)$$

证明. 设 $C(A), C(B)$ 的一组基分别为 $\{a_1, \dots, a_{\text{rank}(A)}\}$ 和 $\{b_1, \dots, b_{\text{rank}(B)}\}$, 则 $A + B$ 中的列必然可以由 $\{a_1, \dots, a_{\text{rank}(A)}, b_1, \dots, b_{\text{rank}(B)}\}$ 的线性组合表示出, 故不等式成立. \square

定理 3.4.5: Sylvester 不等式

对 n 阶方阵 A, B 有

$$\text{rank}(AB) \geq \text{rank}(A) + \text{rank}(B) - n. \quad (3.2)$$

证明. 注意到

$$\text{rank}(AB) + n = \text{rank}\left(\begin{bmatrix} I_n & \\ & AB \end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix} I_n & -B \\ A & \end{bmatrix}\right) \geq \text{rank}(A) + \text{rank}(B). \quad \square$$

定理 3.4.6: Frobenius 秩不等式

对于矩阵 A, B, C , 有

$$\text{rank}(ABC) \geq \text{rank}(AB) + \text{rank}(BC) - \text{rank}(B). \quad (3.3)$$

证明. 注意到

$$\begin{aligned} \text{rank}(ABC) + \text{rank}(B) &= \text{rank}\left(\begin{bmatrix} B & \\ & ABC \end{bmatrix}\right) \\ &= \text{rank}\left(\begin{bmatrix} BC & B \\ & AB \end{bmatrix}\right) \geq \text{rank}(AB) + \text{rank}(BC). \quad \square \end{aligned}$$

第四章 正交性

将列向量看做矩阵，则实向量 v 和 w 的内积可以看做 $\langle v, w \rangle = v^\top w$.

定义 4.0.1: 向量的正交

若 $\langle v, w \rangle = 0$ ，则称 v 和 w 正交 (orthogonal)，记作 $v \perp w$.

注. 依定义， 0 和所有向量正交.

定理 4.0.1: 正交的模

若向量 $v \perp w$ ，则

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2. \quad (4.1)$$

4.1 正交性

定义 4.1.1: 向量和子空间的正交

给定线性空间 L 及其子空间 W ，给定 $v \in L$ ，若 $\forall w \in W$ 均有 $v \perp w$ ，则称 v 和 W 正交，记作 $v \perp W$.

定义 4.1.2: 子空间的正交

给定线性空间 L 的两个子空间 V, W ，若 $\forall v \in V, w \in W$ 均有 $v \perp w$ ，则称 V 和 W 正交，记作 $V \perp W$.

推论. 若 L 的子空间 V, W 正交，则

$$\dim(L) \geq \dim(V) + \dim(W). \quad (4.2)$$

定义 4.1.3: 正交补

线性空间 L 的子空间 V 的正交补 (orthogonal complement) 由 L 中所有同 V 正交的向量组成，记作 V^\perp .

推论. $V \cap V^\perp = \{0\}$.

定理 4.1.1: 线性代数基本定理 · 二

在 \mathbb{F}^n 中, $N(A) = C(A^\top)^\perp$.

在 \mathbb{F}^m 中, $N(A^\top) = C(A)^\perp$.

定理 4.1.2: 分解

给定矩阵 $A \in \mathbb{F}^{m \times n}$, 则 $\forall x \in \mathbb{F}^n$ 均可以分解成

$$x = x_r + x_n,$$

其中 $x_r \in C(A^\top)$, $x_n \in N(A)$, 且这种分解是唯一的.

证明. 由

$$Ax = A(x_r + x_n) = Ax_r \in C(A^\top)$$

知, 只需证明: $\forall b \in C(A^\top)$, 存在唯一的 $x_r \in C(A^\top)$ 使得 $Ax_r = b$.

若存在 $x_r, x'_r \in C(A^\top)$ 满足 $Ax_r = Ax'_r$, 则 $x_r - x'_r$ 同时在 $C(A^\top)$ 和 $N(A)$ 中, 故 $x_r - x'_r = 0$. \square

定理 4.1.3: 矩阵的可逆部分

对于矩阵 A , 把 $N(A)$ 和 $N(A^\top)$ 对应的行和列去掉之后总是一个 r 阶可逆矩阵.

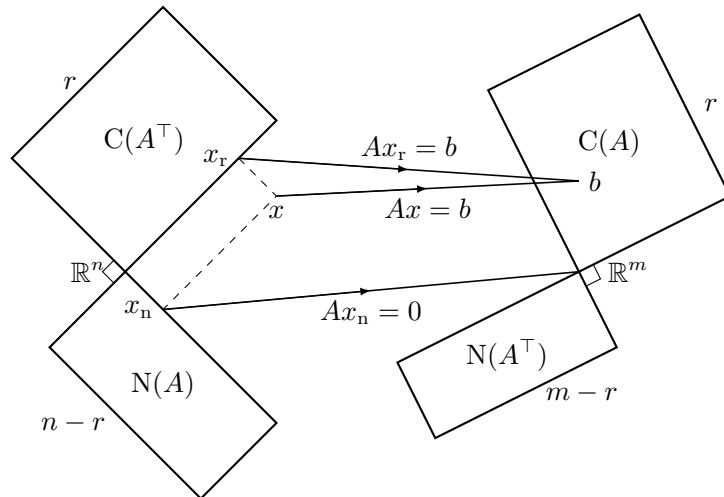


图 4.1: big picture 升级版

4.2 投影

例 4.2.1: 向量在向量上的投影

考虑向量 b 在向量 a 上的投影 (projection) p , 依内积的性质:

$$p = (\hat{a} \cdot b)\hat{a} = \frac{a^\top b}{a^\top a}a = \frac{aa^\top}{a^\top a}b.$$

那么对于一个矩阵, 投影的概念应该如何定义?

考虑 \mathbb{R}^m 中 n 个线性无关的向量 a_1, \dots, a_n 张成的子空间 $\text{span}(a_1, \dots, a_n) = C(A)$, 其中 $A = [a_1 \ \cdots \ a_n]$. 向量 $b \in \mathbb{R}^m$ 在 $C(A)$ 上的投影为 $p \in C(A)$

$$p = Ax = x_1 a_1 + \cdots + x_n a_n.$$

投影的性质要求: p 的终点在 $C(A)$ 中, 且距离 b 的终点最近, 即 $(b - p) \perp C(A)$, 等价于

$$A^\top(b - Ax) = 0,$$

相当于求解线性方程组

$$A^\top Ax = A^\top b.$$

若 $A^\top A$ 可逆, 则 $x = (A^\top A)^{-1}A^\top b$, 故 b 在 $C(A)$ 上的投影为 $p = Ax = A(A^\top A)^{-1}A^\top b$.

定义 4.2.1: 投影矩阵

一个矩阵 A 的投影矩阵 (projection matrix) 为

$$P = A(A^\top A)^{-1}A^\top. \quad (4.3)$$

推论. $P^2 = P$, 这也是符合投影性质的.

注. 只有当 A 可逆时, 才能拆开 $(A^\top A)^{-1} = A^{-1}(A^\top)^{-1}$, 此时 $P = AA^{-1}(A^\top)^{-1}A^\top = I$.

定理 4.2.1: $A^\top A$ 的可逆性

$A^\top A$ 可逆 $\iff A$ 的列之间线性无关.

证明. 只需证明 $N(A^\top A) = N(A)$ 即可.

一方面: $\forall x \in N(A)$, 对 $Ax = 0$ 左乘 A^\top 得 $A^\top Ax = 0$, 故 $x \in N(A^\top A)$;

另一方面: $\forall x \in N(A^\top A)$, 对 $A^\top Ax = 0$ 左乘 x^\top 得

$$x^\top A^\top Ax = \|Ax\|^2 = 0, \implies Ax = 0, \implies x \in N(A).$$

综上, $N(A^\top A) = N(A)$. □

推论.

$$\text{rank}(A) = \text{rank}(A^\top) = \text{rank}(A^\top A) = \text{rank}(AA^\top).$$

4.3 最小二乘法

问题背景 考虑线性方程组 $Ax = b$. 当 A 的行数 $m >$ 列数 n 甚至 $m \gg n$ 时, 一般地 A 不可逆, x 无解, 即 $\forall x \in \mathbb{R}^n$ 均有 $\|b - Ax\| > 0$. 但可以找到一个解 x' 使得 $\|b - Ax'\|$ 最小:

$$x' = \arg \min_x \|b - Ax\|,$$

由投影的性质, Ax' 是 b 在 $C(A)$ 上的投影, 即

$$x' = (A^\top A)^{-1} A^\top b. \quad (4.4)$$

其一个典型应用就是最小二乘法.

例 4.3.1: 直线拟合 (最小二乘法)

m 组数据 (x_i, y_i) , 确定线性关系 $y = a + bx$ 中的系数 a, b , 即

$$\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \iff Ax = b.$$

A 列满秩 $\text{rank}(A) = 2$, 故 $A^\top A$ 可逆,

$$A^\top A = \begin{bmatrix} m & \langle x_i \rangle \\ \langle x_i \rangle & \langle x_i^2 \rangle \end{bmatrix}, \quad (A^\top A)^{-1} = \frac{1}{m\langle x_i^2 \rangle - \langle x_i \rangle^2} \begin{bmatrix} \langle x_i^2 \rangle & -\langle x_i \rangle \\ -\langle x_i \rangle & m \end{bmatrix}.$$

为了简化符号, 此处且仅在此处 $\langle \cdot \rangle$ 特指对 $i = 1, \dots, m$ 求和, 得到

$$\begin{cases} a = \frac{\langle x_i^2 \rangle \langle y_i \rangle - \langle x_i \rangle \langle x_i y_i \rangle}{m\langle x_i^2 \rangle - \langle x_i \rangle^2} \\ b = \frac{-\langle x_i \rangle \langle y_i \rangle + m\langle x_i y_i \rangle}{m\langle x_i^2 \rangle - \langle x_i \rangle^2} \end{cases} \quad (4.5)$$

例 4.3.2: 多项式拟合

m 组数据 (x_i, y_i) , 使用 n 次多项式拟合:

$$y = a_0 + a_1 x + \cdots + a_n x^n,$$

求 $n+1$ 个系数 a_0, \dots, a_n , 即

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \iff Ax = b.$$

从而 $x' = (A^\top A)^{-1} A^\top b$.

注. 两组数据相关不一定代表有因果.

4.4 正交归一基

基是一组线性无关的向量并且张成整个线性空间，我们对基之间的夹角和长度并没有要求。为了方便，我们可以要求基具备一些额外的性质。

定义 4.4.1: 正交归一基

n 维线性空间中的一组基 $\{q_1, \dots, q_n\}$ 若满足，

$$\langle q_i, q_j \rangle = \delta_{ij}, \quad \forall i, j.$$

则称这组基是正交归一的 (orthonormal)。

定义 4.4.2: 正交矩阵

将一组正交归一基 $\{q_1, \dots, q_n\}$ 按列排成矩阵

$$Q = [q_1 \ \cdots \ q_n],$$

则 Q 是方阵且 $Q^\top = Q^{-1}$ ，称 Q 为正交矩阵 (orthogonal matrix)。

定理 4.4.1: 正交归一基的完备性

由 $Q^\top Q = QQ^\top = I$ ，可得正交归一基的完备性 (completeness)

$$\sum_{i=1}^n q_i q_i^\top = I. \quad (4.6)$$

例 4.4.1: Fourier 级数

定义函数 f, g 在 $[-\pi, \pi]$ 上的内积

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(x)g(x) \, dx,$$

则三角函数系列

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\sin \theta}{\sqrt{\pi}}, \frac{\cos \theta}{\sqrt{\pi}}, \frac{\sin 2\theta}{\sqrt{\pi}}, \frac{\cos 2\theta}{\sqrt{\pi}}, \dots \right\},$$

构成一组正交归一基，任意 $[-\pi, \pi]$ 上平方可积的函数 $f(x)$ 均可被展开为 Fourier 级数 (Fourier series):

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

其中 Fourier 系数

$$\begin{cases} a_n := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \\ b_n := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx. \end{cases}$$

4.4.1 Gram-Schmidt 法则

给定一组基 $\{a_1, \dots, a_n\}$, 如何构造一组正交归一基?

方法 4.4.1: Gram-Schmidt 法则

1. 选取 $b_1 := a_1$;
2. 从 a_i 减去沿着 b_1, \dots, b_{i-1} 方向的分量, 作为 b_i

$$b_i := a_i - \sum_{j=1}^{i-1} \frac{\langle b_j, a_i \rangle}{\langle b_j, b_j \rangle} b_j.$$

得到一组正交基 $\{b_1, \dots, b_n\}$, 再归一化 $q_i = \hat{b}_i$, 便得到一组正交归一基 $\{q_1, \dots, q_n\}$.

4.4.2 QR 分解

定理 4.4.2: QR 分解

若 $m \times n$ 矩阵 $A = (a_1, \dots, a_n)$ 的列之间线性无关, 可用 Gram-Schmidt 法则构造一组正交归一基 $\{q_1, \dots, q_n\}$, q_i 同 a_1, \dots, a_{i-1} 正交, 定义

$$R := Q^T A = \begin{bmatrix} q_1^T a_1 & q_1^T a_2 & \cdots & q_1^T a_n \\ & q_2^T a_2 & \cdots & q_2^T a_n \\ & & \ddots & \vdots \\ & & & q_n^T a_n \end{bmatrix}$$

R 是个上三角矩阵, 故 A 可以写成正交矩阵和上三角矩阵的乘积:

$$A = QR. \quad (4.7)$$

注. 在最小二乘法等应用中, $A^T A = R^T R$

$$x = (A^T A)^{-1} A^T b = R^{-1} Q^T b.$$

效率更高.

第五章 行列式

5.1 行列式

下面我们用递归的方式给出行列式的定义。首先引入代数余子式。

定义 5.1.1: 代数余子式

给定 n 阶方阵 A ，记 $A_{\neq ij}$ 为去掉第 i 行和第 j 列得到的 $(n-1)$ 阶方阵，其行列式 $\det(A_{\neq ij})$ 定义为 A_{ij} 的余子式 (minor)； A_{ij} 的代数余子式 (cofactor) 定义为

$$\text{cof}(A)_{ij} := (-1)^{i+j} \det(A_{\neq ij}). \quad (5.1)$$

代数余子式前的正负号与位置的关系：

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (5.2)$$

定义 5.1.2: 行列式 · 一

n 阶方阵 $A \in \mathbb{F}^{n \times n}$ 的行列式 (determinant) 记作 $\det(A)$ ，定义为：

1. $n = 1$ 时， $\det(A) = A_{11}$ ；
2. $n > 1$ 时，任取第 i 行展开或第 j 列展开 (Laplace 展开)

$$\det(A) = \sum_{i=1}^n A_{ij} \text{cof}(A)_{ij} = \sum_{j=1}^n A_{ij} \text{cof}(A)_{ij}. \quad (5.3)$$

推论. 单位矩阵行列式 $\det(I) = 1$ ；对角矩阵行列式就是对角元的积

$$\det(\text{diag}(a_1, \dots, a_n)) = a_1 \cdots a_n. \quad (5.4)$$

定理 5.1.1: 三角矩阵的行列式

三角矩阵的行列式等于对角元的乘积。

证明。按行列式的定义进行计算即证。 □

例 5.1.1: 二阶行列式和三阶行列式

二阶行列式:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \quad (5.5)$$

三阶行列式:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \quad (5.6)$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - (a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}).$$

5.2 行列式的性质

定义 5.2.1: 行列式 · 二

行列式还可定义为 n 个 n 维向量到数域 \mathbb{F} 的映射:

$$\det : \underbrace{\mathbb{F}^n \times \cdots \times \mathbb{F}^n}_n \rightarrow \mathbb{F}$$

且满足:

- 多线性:

$$\det(\dots, ka_i + k'a'_i, \dots) = k \det(\dots, a_i, \dots) + k' \det(\dots, a'_i, \dots); \quad (5.7)$$

- 反对称:

$$\det(\dots, a_i, \dots, a_j, \dots) = -\det(\dots, a_j, \dots, a_i, \dots); \quad (5.8)$$

- 单位化: $\det(I) = 1$.

可以证明, 这个满足此定义的映射存在且唯一, 并且与定义 5.1.2 是等价的.

引理.

证明. 思路在于通过将矩阵 A 表示为一系列初等矩阵 E_1, \dots, E_n 和 $\text{rref}(A)$ 的乘积得到, 过程略. \square

推论.

- 若 A 中有任何两行或两列相同, 则 $\det(A) = 0$;
- 若 A 的行/列之间线性相关, 或者说秩小于阶, 则 $\det(A) = 0$.

定理 5.2.1

$$A \text{ 可逆} \iff \det(A) \neq 0.$$

注. 因此任意给定一个方阵, 其不可逆的可能是很小的. 因为不可逆 ($\det(A) = 0$) 是一个额外的约束.

推论. A 不可逆 $\iff \exists x \neq 0$ 使得 $Ax = 0$.

定义 5.2.2: 排列

给定集合 $S = \{1, \dots, n\}$ 和 n 个数 $i_1, \dots, i_n \in S$, 若映射 σ :

$$\sigma: S \rightarrow S, \sigma(k) \mapsto i_k,$$

是一个一一映射, 则称 σ 是 S 的一个排列 (permutation).

定义 5.2.3: 逆序

给定排列 σ , 若 $i < j$ 且 $\sigma(i) > \sigma(j)$, 则称 (i, j) 是 σ 的一个逆序 (inversion).

σ 中所有逆序组成的集合为逆序集, 逆序集中元素的个数称为逆序数 $\text{inv}(\sigma)$. 排序的奇偶性 (parity) 便定义为逆序数的奇偶性:

$$\text{sgn}(\sigma) = (-1)^{\text{inv}(\sigma)}. \quad (5.9)$$

注. 有的地方将逆序定义为 $(\sigma(i), \sigma(j))$, 二者是等价的.

定义 5.2.4: 全反对称张量

给定 n 个数 $i_1, \dots, i_n \in S = \{1, \dots, n\}$, 定义全反对称张量 (Levi-Civita symbol)

$$\epsilon_{i_1 \dots i_n} = \begin{cases} \text{sgn}(\sigma), & i_1, \dots, i_n \text{ 是 } S \text{ 的一个排列 } \sigma \\ 0, & \text{otherwise} \end{cases} \quad (5.10)$$

定义 5.2.5: 行列式 · 三

利用全反对称张量, 行列式也可以定义为

$$\det(A) = \sum_{i_1, \dots, i_n \in S} \epsilon_{i_1 \dots i_n} a_{1i_1} \cdots a_{ni_n}. \quad (5.11)$$

5.3 行列式的运算

行列式与矩阵运算 行列式在矩阵转置下不变

$$\det(A^T) = \det(A),$$

定理 5.3.1: 行列式与矩阵乘法

$$\det(AB) = \det(A) \det(B). \quad (5.12)$$

证明. 若 A 不可逆, 则 AB 不可逆, 等式成立:

$$\det(AB) = \det(A) \det(B) = 0;$$

若 A 可逆, 则可表示为一系列初等矩阵的乘积 $A = E_k \cdots E_1$, 从而

$$\begin{aligned} \det(AB) &= \det(E_k \cdots E_1 B) = \det(E_k) \det(E_{k-1} \cdots E_1 B) \\ &= \det(E_k) \cdots \det(E_1) \det(B) = \det(E_k \cdots E_1) \det(B) = \det(A) \det(B). \quad \square \end{aligned}$$

推论. 逆的行列式为原行列式的倒数:

$$\det(A^{-1}) = \det(A)^{-1}. \quad (5.13)$$

注. 行列式与矩阵加法之间并无必然联系, 即一般地, $\det(A+B) \neq \det(A) + \det(B)$. 但是当 A, B 满足一些特定条件时, $\det(A+B)$ 可以化简, 见定理 5.3.5.

定理 5.3.2: 分块矩阵的行列式 · 一

若 A 是 m 阶方阵, D 是 n 阶方阵, B 是 $m \times n$ 矩阵, 则

$$\det \begin{pmatrix} A & B \\ D & \end{pmatrix} = \det(A) \det(D). \quad (5.14)$$

证明. 对 A, D 进行 LU 分解, $A = L_A U_A, D = L_D U_D$, 则

$$\begin{bmatrix} A & B \\ D & \end{bmatrix} = \begin{bmatrix} L_A & \\ & L_D \end{bmatrix} \begin{bmatrix} U_A & L_A^{-1} B \\ & U_D \end{bmatrix}$$

前者为下三角矩阵, 后者为上三角矩阵, 故

$$\det \begin{pmatrix} A & B \\ D & \end{pmatrix} = \det(L_A) \det(L_D) \det(U_A) \det(U_D) = \det(A) \det(D). \quad \square$$

定理 5.3.3: 分块矩阵的行列式 · 二

若 A 是 m 阶方阵, D 是 n 阶方阵, 且 A, D 至少一个可逆, 则

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{cases} \det(A) \det(D - CA^{-1}B), & A \text{ 可逆} \\ \det(D) \det(A - BD^{-1}C), & D \text{ 可逆} \end{cases}. \quad (5.15)$$

证明. 对定理 1.3.4 的证明中的矩阵取行列式即证. □

定理 5.3.4: Weinstein-Aronszajn 恒等式

给定 $A \in \mathbb{F}^{m \times n}, B \in \mathbb{F}^{n \times m}$, 则

$$\det(I_m + AB) = \det(I_n + BA). \quad (5.16)$$

证明. 考察分块矩阵

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

显然 I_m, I_n 都是可逆的, 由定理 5.3.3, 即得

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

定理 5.3.5: 矩阵行列式引理

若 A 可逆且 u, v 均为 n 维列向量, 则

$$\det(A + uv^\top) = (1 + v^\top A^{-1}u) \det(A). \quad (5.17)$$

证明. 先证明命题对于 $A = I$ 成立, 事实上

$$\begin{bmatrix} I & \\ v^\top & 1 \end{bmatrix} \begin{bmatrix} I + uv^\top & u \\ & 1 \end{bmatrix} \begin{bmatrix} I & \\ -v^\top & 1 \end{bmatrix} = \begin{bmatrix} I & u \\ & 1 + v^\top u \end{bmatrix}.$$

故

$$\det(I + uv^\top) = 1 + v^\top u \quad (5.18)$$

进而

$$\det(A + uv^\top) = \det(A) \det(I + A^{-1}uv^\top) = (1 + v^\top A^{-1}u) \det(A). \quad \square$$

注. 行列式的运算技巧不宜写得过多, 因为这远非线性代数的精髓.

5.4 Cramer 法则、伴随矩阵

定理 5.4.1: Cramer 法则

考虑线性方程组

$$Ax = b,$$

记 A 的第 i 列为 $a_i = Ae_i$, 定义矩阵 $B_i := [a_1 \cdots a_{i-1} \ b \ a_{i+1} \cdots a_n]$ 为把 A 中第 i 列换为 b 的矩阵, 则

$$B_i = A[e_1 \cdots e_{i-1} \ x \ e_{i+1} \cdots e_n],$$

左右取行列式, 得到 Cramer 法则

$$x_i = \frac{\det(B_i)}{\det(A)}. \quad (5.19)$$

推论. Cramer 法则的应用之一是矩阵求逆, 利用 $AA^{-1} = I$, 把 A 左乘 A^{-1} 的每一列看做一个线性方程组, 解得

$$(A^{-1})_{ij} = \frac{\text{cof}(A)_{ji}}{\det(A)}.$$

定义 5.4.1: 伴随矩阵

方阵 A 的伴随矩阵 (adjugate matrix) 定义为其代数余子式矩阵的转置:

$$\text{adj}(A) := \text{cof}(A)^\top. \quad (5.20)$$

推论. 伴随矩阵的性质:

- $\text{adj}(I) = I$; $\text{adj}(O) = O$, 除了 $\text{adj}([0]) = [1]$;
- $\text{adj}(cA) = c^{n-1} \text{adj}(A)$;
- $\text{adj}(A^\top) = \text{adj}(A)^\top$;
- $\text{adj}(AB) = \text{adj}(B) \text{adj}(A)$, 故 $\text{adj}(A^k) = \text{adj}(A)^k$.

定理 5.4.2: 伴随矩阵与逆的关系

根据 Cramer 法则,

$$A^{-1} = \frac{\text{adj}(A)}{\det(A)}. \quad (5.21)$$

注. 即使 A 不可逆, $\text{adj}(A)$ 依然存在.

定理 5.4.3: 伴随矩阵与原矩阵的乘积

伴随矩阵与原矩阵可交换, 其乘积为

$$A \text{adj}(A) = \text{adj}(A)A = \det(A)I. \quad (5.22)$$

证明. 根据 Laplace 展开,

$$\begin{aligned} [A \text{adj}(A)]_{ij} &= \sum_k A_{ik} \text{adj}(A)_{kj} = \sum_k (-)^{j+k} A_{ik} \det(A_{\neq jk}) = \delta_{ij} \det(A); \\ [\text{adj}(A)A]_{ij} &= \sum_k \text{adj}(A)_{ik} A_{kj} = \sum_k (-)^{i+k} \det(A_{\neq ki}) A_{kj} = \delta_{ij} \det(A). \end{aligned}$$

对应元素相等, 即证. □

推论.

- 伴随矩阵的行列式 (并不要求 A 可逆)

$$\det(\text{adj}(A)) = \det(A)^{n-1}; \quad (5.23)$$

- 伴随矩阵的伴随矩阵

$$\text{adj}(\text{adj}(A)) = \det(A)^{n-2} A. \quad (5.24)$$

- A 可逆 $\iff \text{adj}(A)$ 可逆

$$\text{adj}(A)^{-1} = \text{adj}(A^{-1}) = \frac{A}{\det(A)}; \quad (5.25)$$

定理 5.4.4: 伴随矩阵的秩

矩阵 A 的伴随矩阵 $\text{adj}(A)$ 的秩为

$$\text{rank}(\text{adj}(A)) = \begin{cases} n, & \text{rank}(A) = n \\ 1, & \text{rank}(A) = n - 1 \\ 0, & \text{rank}(A) \leq n - 2 \end{cases} \quad (5.26)$$

证明. $\text{rank}(A) = n$ 时, A 可逆, $\text{adj}(A)$ 也可逆;

$\text{rank}(A) = n - 1$ 时, $A \text{adj}(A) = O$, 故 $C(\text{adj}(A)) \subset N(A)$, 即

$$\text{rank}(\text{adj}(A)) \leq n - \text{rank}(A) = 1,$$

又 A 存在非 0 的余子式, 故 $\text{rank}(\text{adj}(A)) \geq 1$, 即 $\text{rank}(\text{adj}(A)) = 1$;

$\text{rank}(A) \leq n - 2$ 时, A 的余子式均为 0, 故 $\text{adj}(A) = O$. □

注. 计算机运用 Cramer 法则解 n 元线性方程组的时间复杂度是 $\mathcal{O}(n \cdot n!)$, 且在数值上不稳定^I, 这是不可接受的. 与其在计算方面的作用相比, 其理论价值更为重大, 即: 研究了方程组的系数与方程组解的存在性与唯一性关系.

^ICramer, Gabriel (1750). "Introduction à l'Analyse des lignes Courbes algébriques" (in French). Geneva: Europeana. pp. 656-659. Retrieved 2012-05-18.

第六章 特征值和特征向量

6.1 特征值和特征向量

定义 6.1.1: 特征值和特征向量

方阵 A 的特征向量 (eigenvector) $x \neq 0$ 满足:

$$Ax = \lambda x, \quad \lambda \in \mathbb{F}.$$

其中 λ 称为 A 的特征值 (eigenvalue).

定理 6.1.1: 特征值的性质

1. A 的特征向量 x 也是 A^n 的特征向量, 特征值是 λ^n ;
2. 若 A 可逆, 则 x 也是 A^{-1} 的特征向量, 特征值是 λ^{-1} ;
3. 三角矩阵的特征值就是对角元;
4. A 可逆 $\iff A$ 所有特征值非 0.

只需注意到 A 可逆时 $Ax = 0$ 只有零解即可.

定义 6.1.2: 特征子空间

A 的所有特征值为 λ 的特征向量再加上 0 构成 \mathbb{R}^n 的一个线性子空间, 这个子空间就是 $N(A - \lambda I)$.

定理 6.1.2: 特征值

假设 n 阶矩阵 A 有特征向量 x_1, \dots, x_r , 对应特征值为 $\lambda_1, \dots, \lambda_r$, 且这些特征值两两不等, 则 x_1, \dots, x_r 线性无关.

证明. 运用数学归纳法证明. $r = 1$ 时, 定理自动成立;

假设 $r = m - 1$ 定理成立, 若定理在 $r = m$ 时不成立, 则

$$x_m = c_1 x_1 + \dots + c_{m-1} x_{m-1}, \quad (*)$$

两边同时左乘 A 得

$$\lambda_m x_m = c_1 \lambda_1 x_1 + \dots + c_{m-1} \lambda_{m-1} x_{m-1}, \quad (**)$$

$\lambda_m(*) - (**) \text{ 得,}$

$$0 = c_1(\lambda_m - \lambda_1)x_1 + \cdots + c_{m-1}(\lambda_m - \lambda_{m-1})x_{m-1},$$

由 x_1, \dots, x_{m-1} 线性无关可得所有的 $c_i = 0$, 矛盾!

□

6.2 特征多项式

求特征值 λ 需要解如下的特征方程 (eigenfunction)

$$\det(\lambda I - A) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n = 0,$$

特别地, 特征多项式 (eigen-polynomial) 的系数有如下性质:

$$a_1 = -\operatorname{tr}(A), \quad a_n = (-1)^n \det(A) \quad (6.1)$$

由 Vieta 定理, 在考虑重根的情况下,

$$\operatorname{tr}(A) = \sum_{i=1}^n \lambda_i, \quad \det(A) = \prod_{i=1}^n \lambda_i. \quad (6.2)$$

更一般地, Cayley-Hamilton 定理给出了其他项的系数表达式, 比如

$$a_2 = \frac{1}{2}(\operatorname{tr}(A)^2 - \operatorname{tr}(A^2)), \quad (6.3a)$$

$$a_3 = \frac{1}{6}(\operatorname{tr}(A)^3 - 3\operatorname{tr}(A)\operatorname{tr}(A^2) + 2\operatorname{tr}(A^3)), \quad (6.3b)$$

定理 6.2.1: 相似变换的特征多项式

A 的相似变换 $B^{-1}AB$ 和 A 有相同的特征多项式.

证明.

$$\lambda I - B^{-1}AB = B^{-1}(\lambda I - A)B.$$

对两边取行列式即证.

□

推论. 相似变换与原矩阵的迹相同

$$\operatorname{tr}(B^{-1}AB) = \operatorname{tr}(A). \quad (6.4)$$

6.3 矩阵对角化

定理 6.3.1: 可对角化判定

n 阶矩阵 A 可对角化 $\iff A$ 有 n 个线性无关的特征向量 x_1, \dots, x_n ,
此时 $A = X\Lambda X^{-1}$, 且

$$X = (x_1, \dots, x_n), \quad \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n). \quad (6.5)$$

证明. 假设 A 有 n 个线性无关的特征向量 x_1, \dots, x_n ,

$$AX = (Ax_1, \dots, Ax_n) = (\lambda_1 x_1, \dots, \lambda_n x_n) = X\Lambda,$$

故 A 可对角化; 反过来说, 若 A 可对角化为 $X\Lambda X^{-1}$, 则 $AX = X\Lambda$, 即

$$(Ax_1, \dots, Ax_n) = (\lambda_1 x_1, \dots, \lambda_n x_n),$$

故 x_1, \dots, x_n 是 A 的特征向量, 又 X 可逆, 故 x_1, \dots, x_n 线性无关. \square

推论. 有 n 个互不相同特征值的 n 阶矩阵 A 可对角化.

当特征值重复时, 引入两个概念

定义 6.3.1: 几何重数和代数重数

几何重数 (geometric multiplicity, GM): 特征值 λ 对应的最大线性无关的特征向量的个数, 即 $\dim(N(\lambda I - A))$.

代数重数 (algebraic multiplicity, AM): 特征值 λ 作为特征方程 $\det(\lambda I - A) = 0$ 的根的重复次数.

特征方程可以写成

$$\prod_{i=1}^r (\lambda - \lambda_i)^{m_i} = 0,$$

其中 λ_i 是互不相同的根, m_i 是 λ_i 的代数重数.

定理 6.3.2

$GM \leq AM$.

证明. 考虑 n 阶矩阵 A , 假设特征值 λ_1 的 $GM = \dim(\lambda_1 I - A) = m$,

取 $\{x_1, \dots, x_m\}$ 为 $C(\lambda_1 I - A)$ 的一组正交归一基.

取 $\{b_1, \dots, b_{n-m}\}$ 为 $N(\lambda_1 I - A)^\perp$ 的一组正交归一基.

设 $n \times n$ 矩阵

$$P = (x_1, \dots, x_m, b_1, \dots, b_{n-m}) = (X, B),$$

P 是可逆的且 $P^{-1} = P^\top$, 且 $X^\top B = 0$, 则

$$P^{-1}AP = \begin{bmatrix} \lambda_1 I_m & X^\top AB \\ 0 & B^\top AB \end{bmatrix},$$

分块三角矩阵

$$\det(\lambda I - P^{-1}AP) = (\lambda - \lambda_1)^m \det(\lambda I - B^\top AB),$$

A 和 $P^{-1}AP$ 有相同的特征方程, 故 λ_1 必然是 A 的特征方程的根, 且其 $AM \geq GM$. \square

推论. n 阶矩阵 A 的全部特征值为 $\{\lambda_1, \dots, \lambda_r\}$, A 可对角化当且仅当

$$\sum_{i=1}^r \dim(\lambda_i I - A) = n,$$

即所有特征值的 $AM = GM$.

定理 6.3.3: 同时对角化

若 A, B 可对角化, 则他们可以同时对角化当且仅当 $AB = BA$.

证明. 若 A, B 可以同时对角化, 故

$$A = X\Lambda_A X^{-1}, \quad B = X\Lambda_B X^{-1},$$

故

$$AB - BA = X(\Lambda_A \Lambda_B - \Lambda_B \Lambda_A)X^{-1} = 0;$$

若 $AB = BA$, 下证 A, B 可同时对角化.

设 A 的特征值为 $\{\lambda_1, \dots, \lambda_s\}$, λ_i 对应特征子空间为 V_i , 几何重数 $m_i = \dim(V_i)$, 记 $n_i := m_1 + \dots + m_{i-1}$, 取 $\{v_{n_i+1}, \dots, v_{n_i+m_i}\}$ 表示 V_i 的一组基, 记 $X = (v_1, \dots, v_n)$, 则 X 可对角化 A

$$X^{-1}AX = \begin{bmatrix} \lambda_1 I_{m_1} & & \\ & \ddots & \\ & & \lambda_s I_{m_s} \end{bmatrix},$$

$\forall x \in V_i$,

$$(AB - BA)x = (A - \lambda_i I)Bx = 0,$$

故 $Bx \in V_i$, 从而 $X^{-1}BX$ 和 $X^{-1}AX$ 一样是分块对角的:

$$X^{-1}BX = \begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_s \end{bmatrix},$$

其中 B_i 是 $m_i \times m_i$ 的. 给定 B 的特征值 ξ_j , 其必是 B_1, \dots, B_s 其中至少一个的特征值, 不妨考虑 ξ_j 是 B_i 的特征值¹, 若 ξ_j 的 AM $>$ B_i 的 GM, 则 ξ_j 的 AM $>$ B 的 GM, B 便不能被对角化, 与前提矛盾! 故 B_i 均可被特定的 Y_i 对角化, 即 $Y_i^{-1}B_iY_i = \Lambda_i$, 构造

$$Y = \begin{bmatrix} Y_1 & & \\ & \ddots & \\ & & Y_s \end{bmatrix}$$

则 $Y^{-1}X^{-1}BXY$ 是对角化的, 同时 $Y^{-1}X^{-1}AXY$ 也是对角化的, 取 $Z = XY$, 便可同时对角化 A, B . □

6.4* Jordan 标准型

不是所有方阵都可以对角化, 如果 n 阶矩阵 A 有 $r < n$ 个线性独立的特征向量, 怎么把 A 变成最接近对角矩阵的形式?

¹无需考虑 $N(B_i)$

定理 6.4.1: Jordan 标准型

n 阶矩阵 A 有 r 个特征值, 则存在 B , 使得

$$B^{-1}AB = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{bmatrix}, \quad J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

其中 J_i 称为 Jordan 块, λ_i 是 A 的第 i 个特征值.

其证明是线性代数的核心. 其中一些概念需要等到第八章线性映射才会提及. 先给出几个概念证明引理.

定义 6.4.1: 广义特征向量

线性映射 $T: V \rightarrow V$ 的广义特征向量 (general eigenvector) $v \in V$ 且 $v \neq 0$, 使得 $(T - \lambda I)^k v = 0$ 对某个正整数 k 成立. 这里 $I: V \rightarrow V$ 是恒等映射. 使得 $(T - \lambda I)^d v = 0$ 成立的最小正整数 d 称为 v 的幂指数 (exponent).

定理 6.4.2

给定正整数 k , 广义特征方程 $(T - \lambda I)^k v = 0$ 有解当且仅当 λ 是 T 的特征值.

证明. 若 λ 是 T 的特征值, 则 $(T - \lambda I)v = 0$, 左乘 $(T - \lambda I)^{k-1}$ 即可.

若 $(T - \lambda I)^k v = 0$ 有解, 则 $w = (T - \lambda I)^{k-1}v$ 满足 $(T - \lambda I)w = 0$, λ 是 T 的特征值. \square

定理 6.4.3

令 $u_i := (T - \lambda I)^i v$, 则 $B = \{u_0, \dots, u_{d-1}\}$ 是一组线性无关的向量.

证明. 设

$$\sum_{i=0}^{d-1} a_i u_i = \sum_{i=0}^{d-1} a_i (T - \lambda I)^i v = 0,$$

左乘 $(T - \lambda I)^{d-1}$, 左边只剩 $a_0 (T - \lambda I)^{d-1} v$, 故 $a_0 = 0$;

递推地左乘 $(T - \lambda I)^{d-2}, (T - \lambda I)^{d-3}, \dots$ 可得到所有系数为 0, 从而 u_0, \dots, u_{d-1} 线性无关. \square

定理 6.4.4

$$Tu_j = \begin{cases} \lambda u_j + u_{j+1}, & 1 \leq j < d-1 \\ \lambda u_j, & j = d-1 \\ 0, & j > d-1 \end{cases}$$

证明. $1 \leq j < d-1$ 时, $(T - \lambda I)u_j = u_{j+1}$, 即 $Tu_j = \lambda u_j + u_{j+1}$.

$j = d-1$ 时, $(T - \lambda I)u_j = 0$, 即 $Tu_j = \lambda u_j$;

$j > d-1$ 时, $u_j = 0$. □

定理 6.4.5

$X = \text{span}(B)$ 是 T 的不变子空间, 即 $T(X) \subset X$

证明. 由上式, $\forall u = a_0 u_0 + \cdots + a_{d-1} u_{d-1} \in X$

$$Tu = \sum_{i=0}^{d-2} a_i (\lambda u_i + u_{i+1}) + a_{d-1} \lambda u_{d-1} \in X.$$

因为 X 是 T 的不变子空间, 我们可以把 T 看成是 $X \rightarrow X$ 的线性映射. 取 B 作为 X 的一组基, 则 T 在 B 下的表示矩阵为

$$T = \begin{bmatrix} \lambda & & & & \\ 1 & \lambda & & & \\ & \ddots & \ddots & & \\ & & 1 & \lambda \end{bmatrix}$$

这与 Jordan 块在定理 6.4.1 中的定义仅仅是转置的差别. □

接下来我们将证明 V 中存在一组基, T 在这组基上的表示矩阵是分块对角的, 而且每一块都是 Jordan 块的形式.

定理 6.4.6

若 v_1, \dots, v_r 是 T 的广义特征向量, 且相应的幂指数是 d_i , 设

$$u_{ij} := (T - \lambda_i I)^j v_i, \quad V_i := \text{span}(u_{i0}, \dots, u_{id-1}).$$

之前证明了 V_i 是 T 的不变子空间, 且 T 在 V_i 上的表示矩阵是 Jordan 块. 故 T 在 $V_1 \oplus \cdots \oplus V_r$ 上的表示矩阵是分块对角的, 且每一块都是 Jordan 块的形式.

所以我们只要证明存在这样一组广义特征向量 v_1, \dots, v_r 使得 $V = V_1 \oplus \cdots \oplus V_r$ 就可以证明 Jordan 标准型的定理.

假设 λ 是 T 的某个特征值. 如果 $T - \lambda I$ 可以写成 Jordan 块的形式, 则 T 也可以写成 Jordan 块的形式. 所以以下我们用 $T - \lambda I$ 代替 T , 或者说, 考虑有一个特征值是 0 的线性映射 T .

定理 6.4.7

设 $K_i = \ker(T^i)$, $U_i = \text{Im}(T^i)$, 则

$$K_1 \subset K_2 \subset \cdots, \quad U_1 \supset U_2 \supset \cdots$$

证明. 待补 □

6.5 对称矩阵

定义 6.5.1: 对称矩阵

若 $S^\top = S$, 则称 S 是对称矩阵 (symmetric matrix);

若 $A^\top = -A$, 则称 A 是反对称矩阵 (skew-symmetric/anti-symmetric).

定理 6.5.1: 对称矩阵的性质 · 一

若 S 是一个 n 阶实对称矩阵, 则 S 至少有一个实特征值 λ

证明. 由代数基本定理, 对任何矩阵, S 的特征方程至少会得到一个复特征值 λ , 其对应的特征向量为 z (一般也是复的), 则 $\bar{z}^\top z > 0$.

$$Sz = \lambda z, \quad S\bar{z} = \bar{S}\bar{z} = \overline{Sz} = \bar{\lambda}\bar{z},$$

由 S 的对称的性质, 注意到

$$\bar{z}^\top Sz = \lambda \bar{z}^\top z = \lambda (\bar{z}^\top z)^\top = \lambda z^\top \bar{z} = (Sz)^\top \bar{z} = z^\top S\bar{z} = \bar{\lambda} z^\top \bar{z}.$$

故 $\lambda = \bar{\lambda}$. □

推论. 由代数基本定理的递归性, 可推知 S 的所有特征值都是实数.

定理 6.5.2: 对称矩阵的性质 · 二

v 是 S 的特征向量, 若 $w \perp v$, 则 $Sw \perp v$.

证明.

$$(Sw)^\top v = w^\top S^\top v = w^\top Sv = \lambda w^\top v = 0. \quad \square$$

定理 6.5.3: 对称矩阵的性质 · 三

若 W 是 \mathbb{R}^n 的一个线性子空间, 且在 S 的作用下稳定, 即:

$$\forall w \in W, Sw \in W,$$

则 W^\perp 也在 S 的作用下稳定:

$$\forall u \in W^\perp, Su \in W^\perp.$$

证明. $\forall w \in W, u \in W^\perp$

$$(Su)^\top w = u^\top S^\top w = u^\top (Sw) = 0. \quad \square$$

定理 6.5.4: 谱定理

对称矩阵 S 总可以被一个正交矩阵 Q 对角化.

证明. 由定理 6.5.1 和推论可知 S 至少有一个实特征值 λ_1 和实特征向量 q_1 且 $q_1^\top q_1 = 1$, S 在 q_1 张成的一维线性空间上是稳定的.

由定理 6.5.3 可知 S 作用在 $C(q_1)^\perp$ 上也是稳定的, 假设 $C(q_1)^\perp$ 上有一组正交归一基为 $\{a_1, \dots, a_{n-1}\}$, 构造矩阵 $X_1 = [q_1, a_1, \dots, a_{n-1}]$, 且 X_1 是正交的 $X_1^\top X_1 = I$,

$$X_1^\top S X_1 = X_1^\top [\lambda q_1, S a_1, \dots, S a_{n-1}] = \begin{bmatrix} \lambda_1 & \\ & S_1 \end{bmatrix}.$$

S_1 是一个 $(n-1)$ 阶方阵, 且 $(S_1)_{ij} = a_i^\top S a_j$ 的, 显然它也是对称的.

重复上述步骤, 直到用 S 的特征向量构造出 \mathbb{R}^n 的一组正交归一基: 对 S_1 可构造 $(n-1)$ 阶的正交矩阵 X_2 , 使得

$$X_2^\top S_1 X_2 = \begin{bmatrix} \lambda_2 & \\ & S_2 \end{bmatrix},$$

其中 S_2 是一个 $(n-2)$ 阶对称方阵. 从而

$$\begin{bmatrix} 1 & \\ & X_2^\top \end{bmatrix} X_1^\top S X_1 \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & S_2 \end{bmatrix}.$$

$Q_2 := X_1 \text{diag}(1, X_2)$ 也是正交的……最终有

$$Q_n^\top S Q_n = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Q_n 也是正交的. □

对角化对称矩阵 S 的正交矩阵 Q 可被构造:

- 若 S 的特征值互不相同, 对应的归一特征向量 q_i 两两正交, 可选 $Q = [q_1, \dots, q_n]$;
- 若 S 的特征值有重复, 取 $\{q_i\}$ 为相应特征子空间的正交基.

6.6 正定矩阵

定义 6.6.1: 二次型

二次型 (quadratic form) 是形如 $x^\top S x$ 的二次多项式, 其中 S 是实对称矩阵.

定义 6.6.2: 正定矩阵

给定对称矩阵 S , 如果 $\forall x \neq 0$, 二次型 $x^\top S x > 0$, 则称 S 是正定的 (positive definite).

定理 6.6.1: 正定矩阵的判定

对于对称矩阵 S , 下述命题是等价的:

1. $\forall x \neq 0$, 二次型 $x^\top Sx > 0$;
2. S 的所有 n 个特征值都是正的;
3. S 可以只通过换行和倍加后得到 n 个正的主元;
4. S 的所有左上行列式 (前 i 行 i 列子矩阵的行列式) 均 > 0 ;
5. 存在 A 列之间线性无关, 使得 $S = A^\top A$.

证明. $1 \implies 2$: S 对称, 则 $\Lambda = Q^\top S Q$

$$\lambda_i = e_i^\top \Lambda e_i = e_i^\top Q^\top S Q e_i = (Q e_i)^\top S (Q e_i) > 0.$$

$2 \implies 1$: S 的所有 n 个特征值都是正的, 故

$$x^\top Sx = x^\top Q \Lambda Q^\top x = \sum_{i=1}^n \lambda_i (Q^\top x)_i^2 > 0.$$

$5 \implies 1$: A 列之间线性无关, 故 $\forall x \neq 0, Ax \neq 0$

$$x^\top Sx = x^\top A^\top A x = (Ax)^\top (Ax) > 0.$$

$1 \implies 5$: S 正定, 故

$$S = Q \Lambda Q^\top = Q \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix} Q^\top =: A^\top A.$$

$3 \implies 4$: 行倍加不改变所有左上行列式, S 做行倍加得到上三角矩阵 U , 其 $i \times i$ 的左上行列式就是前 i 个主元的乘积, 所以 > 0 .

$4 \implies 3$: U 的左上行列式都 > 0 , 所以前 i 个主元乘积都 > 0 , 所以主元全正.

$3 \implies 5$: $S = LDU$, 由 S 对称且 LDU 分解唯一可知 $L = U^\top$, 又主元全正, 故

$$S = U^\top D U = U^\top \begin{bmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_n} \end{bmatrix} \begin{bmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_n} \end{bmatrix} U =: A^\top A.$$

$5 \implies 3$: A 列之间线性无关, 故 $A = QR$

$$A^\top A = R^\top Q^\top Q R = R^\top R = LDU.$$

□

定义 6.6.3: 半正定矩阵

如果 $\forall x \neq 0$, 二次型 $x^\top Sx \geq 0$, 则称 S 是半正定的 (positive semidefinite).

定理 6.6.2: 半正定但非正定矩阵的判定

1. S 的最小特征值是 0;
2. 存在 A 列之间线性相关, 使得 $S = A^{\top}A$.

半正定但非正定矩阵的行列式为 0.

第七章 奇异值分解

特征值和特征向量只适用于方阵，对于一般的 $m \times n$ 矩阵 A ，有没有类似的操作？考虑 $A^\top A$ 和 AA^\top ，他们都是半正定的，因为 $\forall x$

$$x^\top A^\top A x = \|Ax\|^2 \geq 0,$$

AA^\top 同理，因此 $A^\top A$ 和 AA^\top 都可以对角化。

定义 7.0.1: 奇异值

$A^\top A$ 是半正定的，因此所有特征值 $\lambda_i \geq 0$ ，矩阵 A 的奇异值 (singular value) 便定义为 $A^\top A$ 特征值的平方根： $\sigma_i := \sqrt{\lambda_i}$ 。

为了后续方便，我们将所有奇异值从大到小排列：

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

由定理 4.2.1， $\text{rank}(A^\top A) = \text{rank}(A)$

定理 7.0.1: 非零奇异值的数量

A 的非零奇异值的数量 $r = \text{rank}(A)$ 。

证明. 设 $\{v_1, \dots, v_n\}$ 为 \mathbb{R}^n 中可以把 $A^\top A$ 对角化的一组正交归一基， $\{\lambda_1, \dots, \lambda_n\}$ 为对应的特征值，则 $\{Av_1, \dots, Av_n\}$ 是一个正交向量集合，即 $\forall i \neq j$,

$$(Av_i)^\top Av_j = v_i^\top A^\top Av_j = v_i^\top (\lambda_j v_j) = 0.$$

假设 $\lambda_1 \geq \cdots \geq \lambda_r > 0$ 是所有的正特征值，则 $Av_{r+1}, \dots, Av_n = 0$ 。

$\forall x \in \mathbb{R}^n$, x 可以写成 $x = c_1 v_1 + \cdots + c_n v_n$ ，从而 $\forall y \in C(A)$, y 可以被写为 $\{Av_1, \dots, Av_r\}$ 的线性组合：

$$y = Ax = c_1 Av_1 + \cdots + c_r Av_r + 0 + \cdots + 0,$$

故 $\{Av_1, \dots, Av_r\}$ 是 $C(A)$ 的一组正交基， $r = \text{rank}(A)$ 。□

7.1 奇异值分解

定理 7.1.1: 奇异值分解

$m \times n$ 矩阵 A 秩为 r , 则存在一个 $m \times n$ 的矩阵 Σ

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \quad D = \text{diag}(\sigma_1, \dots, \sigma_r).$$

$m \times m$ 的正交矩阵 U 和 $n \times n$ 的正交矩阵 V , 且

$$A = U\Sigma V^\top.$$

证明. 直接构造出 U, Σ, V .

由 $A^\top A$ 是对称矩阵, 故存在一组 \mathbb{R}^n 中的正交归一基 $\{v_1, \dots, v_n\}$ 可将 $A^\top A$ 对角化, $\{\lambda_1, \dots, \lambda_n\}$ 为对应的特征值, 且 $\lambda_1, \dots, \lambda_r > 0$, $\lambda_{r+1}, \dots, \lambda_n = 0$.

因为 $\{v_1, \dots, v_n\}$ 之间是正交的, 则 $\forall i \neq j$,

$$(Av_i)^\top (Av_j) = v_i^\top A^\top Av_j = \lambda_j v_i^\top v_j = 0.$$

所以 $\{Av_1, \dots, Av_r\}$ 之间也是正交的, $Av_{r+1}, \dots, Av_n = 0$. 令

$$u_i = \frac{Av_i}{\|Av_i\|} = \frac{Av_i}{\sqrt{\lambda_i}} = \frac{Av_i}{\sigma_i}, \quad i = 1, \dots, r.$$

则 $\{u_1, \dots, u_r\}$ 是 $C(A)$ 的一组正交归一基.

再设 $\{u_{r+1}, \dots, u_m\}$ 是 $N(A^\top)$ 中的一组正交归一基, 由 $N(A^\top) = C(A)^\perp$, $\{u_1, \dots, u_m\}$ 是 \mathbb{R}^m 的一组正交归一基.

设矩阵 $U = (u_1, \dots, u_m)$, $V = (v_1, \dots, v_n)$, U 和 V 都是正交矩阵, 且

$$AV = (Av_1, \dots, Av_n) = (\sigma_1 u_1, \dots, \sigma_r u_r, 0, \dots, 0) = U\Sigma. \quad \square$$

定理 7.1.2

$A^\top A$ 和 AA^\top 的非零特征值相同.

证明. 假设 x_i 是 $A^\top A$ 的特征值为 $\lambda_i \neq 0$ 的特征向量,

$$(\lambda_i I - A^\top A)x_i = 0,$$

左乘 A ,

$$A(\lambda_i I - A^\top A)x_i = (\lambda_i I - AA^\top)(Ax_i) = 0,$$

又 $x_i^\top A^\top Ax_i = \lambda_i x_i^\top x_i > 0$, 所以 $Ax_i \neq 0$, 所以 Ax_i 是 AA^\top 的特征值为 λ_i 的特征向量.

同理, 如果 x_i 是 AA^\top 的特征值为 $\lambda_i \neq 0$ 的特征向量, $A^\top x_i$ 是 $A^\top A$ 的特征值为 λ_i 的特征向量.

从而 $A^\top A$ 和 AA^\top 非零特征值对应的特征向量一一对应. \square

例 7.1.1: 四个子空间的正交归一基

$\{v_1, \dots, v_r\}$ 是 $C(A^\top)$ 的正交归一基, $V_r = (v_1, \dots, v_r)$
 $\{v_{r+1}, \dots, v_n\}$ 是 $N(A)$ 的正交归一基, $V_{n-r} = (v_{r+1}, \dots, v_n)$
 $\{u_1, \dots, u_r\}$ 是 $C(A)$ 的正交归一基, $U_r = (u_1, \dots, u_r)$
 $\{v_{r+1}, \dots, v_n\}$ 是 $N(A^\top)$ 的正交归一基, $U_{n-r} = (u_{r+1}, \dots, u_m)$

例 7.1.2: 数据压缩

若 $r = \text{rank}(A) < \min(m, n)$, 则

$$A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix} = U_r D V_r^\top = \sum_{i=1}^r \sigma_i u_i v_i^\top.$$

可以用 U_r, D, V_r 这三个矩阵的 $r(m+1+n)$ 个分量完全决定 A 原来的 mn 个分量. (无损)

甚至可以把很小的奇异值当成 0, 进一步压缩图片. (有损) 由此带来的误差

$$\delta A = \sum_{\ell=k+1}^r \sigma_\ell u_\ell v_\ell^\top.$$

误差分量的绝对值

$$|\delta A_{ij}| = \left| \sum_{\ell=k+1}^r \sigma_\ell (u_\ell)_i (v_\ell)_j \right| \leq \sum_{\ell=k+1}^r \sigma_\ell.$$

因此误差由忽略的奇异值控制, 忽略的越少误差越小

7.2 矩阵的模

我们用内积定义了向量的模 (即长度)

$$\|x\| := \sqrt{\langle x, x \rangle}.$$

定理 7.2.1

$$\|Ax\| \leq \sigma_1 \|x\|. \quad (7.1)$$

证明.

$$\begin{aligned}
 \|Ax\|^2 &= x^\top A^\top A x = x^\top V \Sigma^\top \Sigma V^\top x \\
 &= \sum_{k=1}^r x^\top v_k \sigma_k^2 v_k^\top x \leq \sigma_1^2 \sum_{k=1}^n x^\top v_k v_k^\top x \\
 &= \sigma_1^2 x^\top V V^\top x = \sigma_1^2 x^\top x,
 \end{aligned}$$

等号可在 $x = cv_1$ 时成立. □

定义 7.2.1: 矩阵的模

矩阵的模 (norm) 定义为

$$\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_1. \quad (7.2)$$

由矩阵模的定义可直接导出 $\forall x \neq 0$,

$$\|Ax\| \leq \|A\| \|x\|.$$

定理 7.2.2: 三角不等式

$$\|A + B\| \leq \|A\| + \|B\|. \quad (7.3)$$

证明.

$$\|(A + B)x\| = \|Ax + Bx\| \leq \|Ax\| + \|Bx\| \leq \|A\| \|x\| + \|B\| \|x\|.$$

故

$$\|A + B\| = \max_{x \neq 0} \frac{\|(A + B)x\|}{\|x\|} \leq \|A\| + \|B\|.$$

□

定理 7.2.3: Eckart-Young-Mirsky 定理

同矩阵 A 最接近的秩为 k 的矩阵为

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^\top. \quad (7.4)$$

证明. 只需证 $\forall B$ 秩为 k , 都有

$$\|A - B\| \geq \|A - A_k\| = \sigma_{k+1}.$$

设 $w = c_1 v_1 + \cdots + c_{k+1} v_{k+1}$, 因为 $\text{rank}(B) = k$, 故 Bv_1, \dots, Bv_{k+1} 必然线性相关, 继而存在非零的 c_1, \dots, c_{k+1} 使得 $Bw = 0$, 在此基础上再归一化 w , 从而

$$\begin{aligned} \|A - B\|^2 &\geq \|(A - B)w\|^2 = \|Aw\|^2 \\ &= \sigma_1^2 c_1^2 + \cdots + \sigma_{k+1}^2 c_{k+1}^2 \geq \sigma_{k+1}^2 (c_1^2 + \cdots + c_{k+1}^2) = \sigma_{k+1}^2. \end{aligned}$$

□

7.3 伪逆

定义 7.3.1: 伪逆

$m \times n$ 矩阵 $A = U\Sigma V^\top$, 定义伪逆 (pseudoinverse) 是一个 $n \times m$ 的矩阵

$$A^+ := V\Sigma^+U^\top. \quad (7.5)$$

其中 Σ^+ 是一个 $n \times m$ 的矩阵

$$\Sigma^+ := \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad D^{-1} = \text{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1}).$$

注. 伪逆与原矩阵的乘积并不是单位矩阵:

$$A^+A = V \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V^\top,$$

是投影到 $C(A^\top)$ 的矩阵;

$$AA^+ = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^\top,$$

是投影到 $C(A)$ 的矩阵.

但额外的满足:

$$AA^+A = A, \quad A^+AA^+ = A^+.$$

定理 7.3.1: 伪逆与最小二乘法

最小二乘法

$$A^\top Ax = A^\top b,$$

的解为 $x^+ = A^+b$.

证明.

$$A^\top Ax^+ = V\Sigma^\top U^\top U\Sigma V^\top V\Sigma^+U^\top b = V\Sigma^\top U^\top b = A^\top b. \quad \square$$

7.4 主成分分析

一组数据 $\mu = (\mu_1, \dots, \mu_n)$ 来源于 n 个样本, 其样本均值 (mean) 和样本方差 (variance) 分别为

$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^n \mu_i, \quad \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (\mu_i - \bar{\mu})^2.$$

将数据存在一个 $m \times n$ 的矩阵 A 中, 每一行对应一种数据, 每一列代表一个样本. 将每个元素减去其所在行的平均值

$$A_{ij} := (A_0)_{ij} - \frac{1}{n} \sum_{k=1}^n (A_0)_{ik}.$$

由此得到矩阵 A , 每一行都是以 0 为中心的分布.

定义 7.4.1: 协方差矩阵

定义协方差矩阵 (covariance matrix)

$$S := \frac{AA^\top}{n-1}. \quad (7.6)$$

对角线上 S_{ii} 是样本方差; S_{ij} 是样本协方差.

方法 7.4.1: 主成分分析

主成分分析 (principal component analysis, PCA): 找到原有数据的一系列线性组合作为新的数据, 新数据之间的协方差为 0.

利用 $A = U\Sigma V^\top$, 定义新的数据矩阵 $B := U^\top A = \Sigma V$, B 的协方差矩阵

$$\frac{BB^\top}{n-1} = \frac{\Sigma\Sigma^\top}{n-1}$$

是对角的, 故 B 之间协方差为 0.

总方差在这种变换下是不变的:

$$\text{tr}\left(\frac{BB^\top}{n-1}\right) = \frac{\text{tr}(U^\top AA^\top U)}{n-1} = \frac{\text{tr}(UU^\top AA^\top)}{n-1} = \text{tr}\left(\frac{AA^\top}{n-1}\right).$$

所有数据点分布在 $\{u_1, \dots, u_r\}$ 张成的 $C(A)$ 上, u_1 是所有数据变化最大的方向 (方差最大)、 u_2 次之……因此 $\{u_1, \dots, u_r\}$ 称作主成分 (principal component).

第八章 线性映射

定义 8.0.1: 映射

给定两个集合 S, S' , 如果 $\forall x \in S$, 均有一个对应的 $f(x) \in S'$, 这种对应关系 f 便称为映射 (mapping), 记作

$$f: S \rightarrow S', \quad x \mapsto f(x). \quad (8.1)$$

其中 S 称为定义域 (domain), S' 称为陪域 (codomain). $f(x)$ 称为 x 的像 (image), 所有像的集合 $f(S) = \{f(x) | x \in S\}$ 称为值域 (image of S), 有 $f(S) \subset S'$.

给定 $y \in S'$, 所有满足 $f(x) = y$ 的 x 的集合称为 y 的原像 (preimage), 记作

$$f^{-1}(y) := \{x \in S | f(x) = y\}. \quad (8.2)$$

注. 符号 $f(\cdot)$ 便代表映射 $x \mapsto f(x)$.

定义 8.0.2: 映射的复合

映射 $f: U \rightarrow V$, $g: V \rightarrow W$ 的复合 (composition) 构成一个新的映射:

$$g \circ f: U \rightarrow W, \quad x \mapsto g(f(x)).$$

推论. 映射的复合满足结合律

$$(h \circ g) \circ f = h \circ (g \circ f) \equiv h \circ g \circ f.$$

定义 8.0.3: 单射、满射和双射

- 单射 (injection): $\forall x, y \in S$ 且 $x \neq y$, 均有 $f(x) \neq f(y)$.
- 满射 (surjection): $f(S) = S'$.
- 双射 (bijection): 既是单射又是满射.

定义 8.0.4: 恒等映射

定义 S 上的恒等映射 (identity) 为

$$\text{id}_S: S \rightarrow S, \quad x \mapsto x. \quad (8.3)$$

推论. $\forall f: S \rightarrow S'$, 有 $f \circ \text{id}_S = \text{id}_{S'} \circ f = f$.

定义 8.0.5: 逆映射

给定映射 $f: S \rightarrow S'$, 若存在 $g: S' \rightarrow S$ 使得

$$g \circ f = \text{id}_S, \quad f \circ g = \text{id}_{S'},$$

则称映射 $f: S \rightarrow S'$ 可逆, $g = f^{-1}$ 为 f 的逆映射 (inverse).

定理 8.0.1: 有关映射的等价描述

1. 映射 f 为单射 \iff 存在映射 g 使得 $g \circ f = \text{id}$;
2. 映射 f 为满射 \iff 存在映射 g 使得 $f \circ g = \text{id}$;
3. 映射 f 为双射 $\iff f$ 可逆.

证明. 显然 (1), (2) \implies (3), 故只需证明 (1), (2). 证明留作习题. \square

8.1 线性映射和矩阵

定义 8.1.1: 线性映射

给定两个线性空间 V, W , 若映射 $T: V \rightarrow W$ 满足:

1. $\forall u, v \in V, T(u + v) = T(u) + T(v)$;
2. $\forall c \in \mathbb{F}, T(cu) = cT(u)$.

则称映射 T 是线性映射 (或线性变换, linear mapping). 特别地, $V \rightarrow \mathbb{F}$ 的称为线性函数.

推论. $T(0) = 0$.

定理 8.1.1: 线性映射与基

给定两个线性空间 V, W , $\{v_1, \dots, v_n\}$ 是 V 中的一组基, $\{w_1, \dots, w_n\}$ 是 W 中任意 n 个元素, 则存在唯一的线性映射 $T: V \rightarrow W$ 使得

$$T(v_1) = w_1, \dots, T(v_n) = w_n.$$

证明. (存在性) $\forall v \in V$ 均可唯一写成基的线性组合 $v = c_1v_1 + \dots + c_nv_n$, 定义映射 $T: V \rightarrow W$

$$T(v) = c_1w_1 + \dots + c_nw_n,$$

下面证明 T 是线性映射, 再任取 $u = d_1v_1 + \dots + d_nv_n \in V$

$$\begin{aligned} T(v + u) &= T((c_1 + d_1)v_1 + \dots + (c_n + d_n)v_n) \\ &= (c_1 + d_1)w_1 + \dots + (c_n + d_n)w_n = T(v) + T(u); \\ T(cv) &= T(cc_1v_1 + \dots + cc_nv_n) = cc_1w_1 + \dots + cc_nw_n = cT(v). \end{aligned}$$

(唯一性) 假设存在另一个线性映射 $F : V \rightarrow W$ 满足

$$F(v_1) = w_1, \dots, F(v_n) = w_n,$$

则

$$F(v) = c_1 F(v_1) + \dots + c_n F(v_n) = c_1 w_1 + \dots + c_n w_n = T(v).$$

综上, 存在唯一的线性映射. □

注. 只要知道线性映射在基上的值, 就唯一决定了这个线性映射.

例 8.1.1: 矩阵定义线性映射

$m \times n$ 的矩阵 A 可定义一个线性映射:

$$L_A : \mathbb{F}^n \rightarrow \mathbb{F}^m, x \mapsto Ax. \quad (8.4)$$

定理 8.1.2: 线性映射和矩阵

设 $L : \mathbb{F}^n \rightarrow \mathbb{F}^m$ 是线性映射, 则存在唯一的矩阵 A 使得 $L = L_A$.

证明. 设 $\{e_1, \dots, e_n\}$ 是 \mathbb{F}^n 的标准基, $\{f_1, \dots, f_m\}$ 是 \mathbb{F}^m 的标准基, $\forall x \in \mathbb{F}^n$, 有 $x = x_1 e_1 + \dots + x_n e_n$, 则

$$L(x) = x_1 L(e_1) + \dots + x_n L(e_n).$$

$L(e_i) \in \mathbb{F}^m$, 故可写成基的线性组合

$$L(e_i) = a_{i1} f_1 + \dots + a_{im} f_m.$$

故

$$\begin{aligned} L(x) &= x_1(a_{11} f_1 + \dots + a_{m1} f_m) + \dots + x_n(a_{n1} f_1 + \dots + a_{nm} f_m) \\ &= (a_{11} x_1 + \dots + a_{1n} x_n) f_1 + \dots + (a_{m1} x_1 + \dots + a_{mn} x_n) f_m \\ &= \begin{bmatrix} a_{11} x_1 + \dots + a_{1n} x_n \\ \vdots \\ a_{m1} x_1 + \dots + a_{mn} x_n \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} =: Ax. \end{aligned}$$

便唯一确定了一个矩阵 A . □

注. 线性映射给出了矩阵和向量乘法的自然定义.

8.2 线性映射的性质

利用线性映射和矩阵的对应, 线性映射的加法和数乘等价于矩阵的加法和数乘, 零映射对应零矩阵, 这些都是平凡的.

给出加法、数乘、零映射的定义后, 所有 $V \rightarrow W$ 的线性映射的集合 $\{T\}$ 便构成一个线性空间, 可验证满足 8 条公理.

定义 8.2.1: 线性映射的核

线性映射 $F: V \rightarrow W$ 的核 (kernel) 是所有满足 $F(v) = 0$ 的向量 v 的集合

$$\ker(F) \equiv \{v \in V \mid F(v) = 0\}.$$

推论. $\ker(F)$ 是 V 的线性子空间. $\ker(L_A) = N(A)$.

定理 8.2.1: 核和单射

$$\ker(F) = \{0\} \iff F \text{ 是单射}.$$

证明. (矩阵版本) 对应矩阵零空间为 $\{0\}$, $Av = b$ 若有解则解必唯一.

(抽象版本) 若 $u, v \in V$ 满足 $F(u) = F(v)$, 则 $F(u - v) = F(u) - F(v) = 0$, 从而 $u - v = 0$. \square

定理 8.2.2: 核的性质

若线性映射 $F: V \rightarrow W$ 的核 $\ker(F) = \{0\}$, 则对于线性无关的一组 $v_1, \dots, v_n \in V$, 有 $F(v_1), \dots, F(v_n)$ 线性无关.

证明. (矩阵版本) 对应矩阵零空间为 $\{0\}$, 则列满秩, 列之间线性无关.

(抽象版本) 假设 $x_1 F(v_1) + \dots + x_n F(v_n) = 0$, 则

$$F(x_1 v_1 + \dots + x_n v_n) = 0, \implies x_1 v_1 + \dots + x_n v_n = 0.$$

v_1, \dots, v_n 线性无关, 故只有零解. \square

定义 8.2.2: 线性映射的像

线性映射 $F: V \rightarrow W$ 的像 (image) 是所有 $F(v)$ 的集合

$$\operatorname{im}(F) \equiv \{F(v) \in W \mid \forall v \in V\}.$$

推论. $\operatorname{im}(F)$ 是 W 的线性子空间. $\operatorname{im}(L_A) = C(A)$.

定理 8.2.3: 核和像的关系

V 是线性空间, $L: V \rightarrow W$ 是线性映射

$$\dim(V) = \dim(\ker(L)) + \dim(\operatorname{im}(L)). \quad (8.5)$$

证明. (矩阵版本)

$$\dim(V) = \dim(N(A)) + \dim(C(A^\top)) = \dim(N(A)) + \dim(C(A)).$$

(抽象版本) 略 \square

定理 8.2.4: 核、像和双射

线性映射 $L: V \rightarrow W$, 且 $\dim(V) = \dim(W)$, 则

$$\ker(F) = \{0\} \iff \operatorname{im}(F) = W \iff L \text{ 是双射}.$$

证明. 略

□

8.3 基的变换

定义 8.3.1: 坐标向量

设 $B = \{v_1, \dots, v_n\}$ 是线性空间 V 上的一组基, V 中的向量 $v \in V$ 可唯一写成 $v = x_1 v_1 + \dots + x_n v_n$, 其在基 B 下的坐标向量 (coordinate vector) 为

$$x_B(v) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

显然 $x_B: V \rightarrow \mathbb{F}^n$ 是线性映射, 且是一个双射.

我们可以选取 V 上的另一组基 $B' = \{u_1, \dots, u_n\}$, 基变换矩阵:

$$(u_1, \dots, u_n) = (v_1, \dots, v_n)M, \quad (*)$$

v 也可以写成 $v = y_1 u_1 + \dots + y_n u_n$, 由于向量在基的变换下保持不变, 故

$$v = (v_1, \dots, v_n)(x_1, \dots, x_n)^\top = (u_1, \dots, u_n)(y_1, \dots, y_n)^\top.$$

可以推出

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = M^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \quad (**)$$

定理 8.3.1: 换基矩阵

$L: V \rightarrow W$ 是一个线性映射, $B = \{v_1, \dots, v_n\}$ 是 V 上的一组基, $B' = \{w_1, \dots, w_m\}$ 是 W 上的一组基. 则存在唯一的 $m \times n$ 矩阵 $M_{B'}^B(L)$, 使得 $\forall v \in V$,

$$x_{B'}(L(v)) = M_{B'}^B(L)x_B(v).$$

证明. $\forall v \in V$, 有

$$v = x_1 v_1 + \dots + x_n v_n, \quad L(v) = x_1 L(v_1) + \dots + x_n L(v_n).$$

$L(v_i) \in W$, 所以

$$L(v_i) = m_{1i} w_1 + \dots + m_{mi} w_m.$$

写成矩阵的形式即 $(L(v_1), \dots, L(v_n)) = (w_1, \dots, w_m)M$, 从而

$$L(v) = (L(v_1), \dots, L(v_n))(x_1, \dots, x_n)^\top = (w_1, \dots, w_m)M(x_1, \dots, x_n)^\top.$$

故 $L(v)$ 在 B' 上的坐标为 $M(x_1, \dots, x_n)^\top$. \square

$M_{B'}^B(L)$ 是所有线性变换 $L: V \rightarrow W$ 到 $\dim(W) \times \dim(V)$ 矩阵的线性映射, 并且是一个双射.

特别地, 当 $L \equiv \text{id}: V \rightarrow V$ 时,

$$x_{B'}(v) = M_{B'}^B(\text{id})x_B(v).$$

定理 8.3.2: 线性变换的复合与矩阵乘法

线性映射 $L_1: U \rightarrow V$, $L_2: V \rightarrow W$, B, B', B'' 分别是 U, V, W 上的一组基, 则

$$M_{B''}^B(L_2 \circ L_1) = M_{B''}^{B'}(L_2)M_{B'}^B(L_1).$$

线性映射的复合等价于对应矩阵的乘法, 由此可自然得到矩阵乘法的规则.

定理 8.3.3: $M_{B'}^B(\text{id})$ 可逆

$$M_{B'}^B(\text{id}) = M_B^{B'}(\text{id})^{-1}.$$

定理 8.3.4

线性映射 $L: V \rightarrow W$, B, B' 是 V 上的两组基, C, C' 是 W 上的两组基, 则

$$M_{C'}^{B'}(L) = M_{C'}^C(\text{id})M_C^B(L)M_B^{B'}(\text{id}) = M_{C'}^{C'}(\text{id})^{-1}M_C^B(L)M_B^{B'}(\text{id})$$

证明. 利用 $L = \text{id}_W \circ L \circ \text{id}_V$. \square

推论. $L: V \rightarrow V$, B, B' 是 V 上的两组基, 则

$$M_{B'}^{B'}(L) = M_{B'}^B(\text{id})^{-1}M_B^B(L)M_B^{B'}(\text{id}).$$

因此相似变换就是换基, 矩阵对角化就是找到描述线性变换的最好的基.

8.4 对偶空间

如何从已知的线性空间构造新的线性空间?

定义 8.4.1: 对偶空间

线性空间 V 的对偶空间 (dual space) V^* 是所有线性映射 $L: V \rightarrow \mathbb{R}$ 构成的线性空间.

定理 8.4.1: 对偶空间的基

通过 V 的一组基 $\{v_1, \dots, v_n\}$ 可构造 V^* 的基 $\{v^{*1}, \dots, v^{*n}\}$, 满足

$$v^{*i}(v_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (8.6)$$

证明. (完备性) $\forall L \in V^*$, 由定理 8.1.1, L 可由其在基上的取值 $\{L(v_1), \dots, L(v_n)\}$ 唯一决定, 又 $L(v_1)v^{*1} + \dots + L(v_n)v^{*n}$ 和 L 在基 v_1, \dots, v_n 上取到了相同的值, 故二者相等:

$$L = L(v_1)v^{*1} + \dots + L(v_n)v^{*n}.$$

即 L 可被写成 $\{v^{*1}, \dots, v^{*n}\}$ 的线性组合.

(线性无关) 若基的线性组合是零映射 $x_1v^{*1} + \dots + x_nv^{*n} = O$, 则 $O(v_i) = x_i = 0$, 即只有零解. \square

例 8.4.1: Fourier 变换

Fourier 变换

$$\hat{f}(k) = \int f(x) e^{ik \cdot x} d^3x,$$

就是将 \mathbb{R}^3 的函数变成 $(\mathbb{R}^3)^*$ 的函数.

例 8.4.2: 对偶的对偶

依定义, 对偶空间 V^* 的对偶空间 V^{**} 是所有线性映射 $F: V^* \rightarrow \mathbb{R}$ 构成的线性空间. $\forall v \in V$, 可定义映射 $u^{**} \in V^{**}$, 使得 $\forall L \in V^*$

$$u^{**}(L) = L(v). \quad (8.7)$$

因此 V^{**} 和 V 是自然同构的 (natural isomorphism): $V^{**} \cong V$. 这是范畴论 (category theory) 的概念, 粗糙地说就是这种同构关系不依赖于基的选取, 而 V 和 V^* 的同构是依赖于基的. 因此我们可以将 V^{**} 和 V 视为同一个线性空间, 从而 V^{**}, V^{***}, \dots 也就再没有研究价值了.

定理 8.4.2: 对偶空间的基变换

给定线性空间 V 及其中的两组基 $\{v_1, \dots, v_n\}$ 和 $\{u_1, \dots, u_n\}$, 我们可以给出对偶空间 V^* 的基 $\{v^{*1}, \dots, v^{*n}\}$ 和 $\{u^{*1}, \dots, u^{*n}\}$, 满足

$$v^{*i}(v_j) = \delta_{ij}, \quad u^{*i}(u_j) = \delta_{ij},$$

若 $(v_1, \dots, v_n) = (u_1, \dots, u_n)A$, 则 $(v^{*1}, \dots, v^{*n})^\top = A^{-1}(u^{*1}, \dots, u^{*n})^\top$

证明. 若 $(v^{*1}, \dots, v^{*n})^\top = B(u^{*1}, \dots, u^{*n})^\top$, 即

$$v_i = \sum_{j=1}^n u_j A_{ij}^j, \quad v^{*i} = \sum_{j=1}^n B_{ij}^i u^{*j}.$$

则

$$\begin{aligned} v^{*i}(v_j) &= \sum_{k=1}^n B_{ik}^i u^{*k} \left(\sum_{\ell=1}^n u_\ell A_{j\ell}^\ell \right) = \sum_{k,\ell} B_{ik}^i A_{j\ell}^\ell u^{*k}(u_\ell) \\ &= \sum_{k,\ell} B_{ik}^i A_{j\ell}^\ell \delta_{k\ell} = \sum_{k=1}^n B_{ik}^i A_{jk}^k = \delta_{ij}^i. \end{aligned}$$

故 $B = A^{-1}$. □

8.5 直和、直积

定义 8.5.1: 线性空间的和

线性空间 U 的两个子空间 V, W 的和 (sum) $V + W$ 定义为所有 $v + w, v \in V, w \in W$ 的集合:

$$V + W \equiv \{v + w \mid v \in V, w \in W\}. \quad (8.8)$$

显然, $V + W$ 也是 U 的子空间.

定义 8.5.2: 线性空间的直和

线性空间 U 是 V 和 W 的直和 (direct sum) $U = V \oplus W$, 若 $\forall u \in U$, 存在唯一的 $v \in V, w \in W$ 使得 $u = v + w$.

定理 8.5.1: 和与直和

若 $U = V + W$ 且 $V \cap W = \{0\}$, 则 $U = V \oplus W$.

证明. 假设 $u \in U$ 可以写成 $u = v + w = v' + w'$, 则 $v - v' = w - w'$,

又 $v - v' \in V, w - w' \in W$ 且 $V \cap W = \{0\}$, 所以 $v - v' = w - w' = 0$. 故分解是唯一的. □

定理 8.5.2: 直和的存在

U 是一个有限维线性空间, V 是 U 的子空间, 则存在 U 的子空间 W 使得 $U = V \oplus W$.

证明. 取 V 的一组基 $\{v_1, \dots, v_r\}$, 可将其扩张成 U 的一组基 $\{v_1, \dots, v_r, w_1, \dots, w_m\}$, 取 $W = \text{span}(w_1, \dots, w_m)$ 即可. □

推论.

$$\dim(V \oplus W) = \dim(V) + \dim(W). \quad (8.9)$$

定义 8.5.3: 线性空间的直积

两个线性空间 V, W 的直积 (direct product) $V \times W$ 是所有形如 $(v, w), v \in V, w \in W$ 的元素的集合:

$$V \times W \equiv \{(v, w) \mid v \in V, w \in W\}. \quad (8.10)$$

定理 8.5.3: 直基的维度

$V \times W$ 是一个线性空间. 且

$$\dim(V \times W) = \dim(V) + \dim(W). \quad (8.11)$$

8.6 张量

定义 8.6.1: 多重线性映射

映射 $L: V_1 \times \cdots \times V_r \rightarrow W$ 是一个多重线性映射 (multiple linear mapping), 若其对于每一个变量都是线性的:

$$L(\dots, au + bw, \dots) = aL(\dots, u, \dots) + bL(\dots, w, \dots).$$

定义 8.6.2: 张量空间 $V^* \otimes V^*$

考虑所有多重线性函数 $L: V \times V \rightarrow \mathbb{R}$ 的集合, 我们可以在这个集合上定义加法和数乘:

- 加法: $(L_1 + L_2)(u, v) = L_1(u, v) + L_2(u, v);$
- 数乘: $(cL)(u, v) = cL(u, v);$
- 零元: $O(u, v) \equiv 0.$

因此这个集合构成一个线性空间, 称作张量空间 (tensor space) $V^* \otimes V^*$. 其中的每一个元素 L 是二阶协变张量 (covariant tensor), 记作 $(0, 2)$ 张量.

若 V 的一组基为 $\{v_1, \dots, v_n\}$, 则 $\forall L \in V^* \otimes V^*$

$$L(u, v) = L\left(\sum_{i=1}^n a_i v_i, \sum_{j=1}^n b_j v_j\right) = \sum_{i,j} a_i b_j L(v_i, v_j).$$

n^2 个函数值 $L(v_i, v_j)$ 便可唯一确定函数 L .

例 8.6.1: $V^* \otimes V^*$ 的基

给定对偶空间 V^* 的一组基 $\{v^{*1}, \dots, v^{*n}\}$ 满足 $v^{*i}(v_j) = \delta^i_j$. 继而定义张量 $v^{*i} \otimes v^{*j} \in V^* \otimes V^*$ 满足

$$v^{*i} \otimes v^{*j}(u, v) = v^{*i}(u) v^{*j}(v).$$

从而

$$v^{*i} \otimes v^{*j}(v_k, v_\ell) = v^{*i}(v_k) v^{*j}(v_\ell) = \delta_k^i \delta_\ell^j.$$

n^2 个张量 $v^{*i} \otimes v^{*j}$ 构成 $V^* \otimes V^*$ 的一组基.

张量 $\forall w \in V^* \otimes V^*$,

$$w = \sum_{i,j} w_{ij} v^{*i} \otimes v^{*j}, \quad w_{ij} = w(v_i, v_j).$$

给出 V, V^* 的另一组基 $\{u_1, \dots, u_n\}, \{u^{*1}, \dots, u^{*n}\}$, 有变换

$$\begin{aligned} (u_1, \dots, u_n) &= (v_1, \dots, v_n)A, \\ (u^{*1}, \dots, u^{*n})^\top &= (v^{*1}, \dots, v^{*n})^\top A^{-1}. \end{aligned}$$

张量 w 在基 $\{u^{*i} \otimes u^{*j}\}$ 下的分量

$$w'_{ij} = w(u_i, u_j) = w\left(\sum_{k=1}^n v_k A^k_i, \sum_{\ell=1}^n v_\ell A^\ell_j\right) = \sum_{k,\ell} w_{k\ell} A^k_i A^\ell_j.$$

因此这也是协变 (covariant) 的含义: 分量在坐标变换下同基的变换规律一致.

定义 8.6.3: 张量积

U, V 是两个线性空间, 定义 $u \in U, v \in V$ 的张量积 (tensor product) 是一个新的元素 $u \otimes v$, 且满足以下性质:

- 结合律: $(u \otimes v) \otimes w = u \otimes (v \otimes w) \equiv u \otimes v \otimes w$;
- 左分配律: $(u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v$;
- 右分配律: $u \otimes (v_1 + v_2) = u \otimes v_1 + u \otimes v_2$;
- 数乘: $(au) \otimes v = u \otimes (av) = a(u \otimes v)$.

注. 张量积并不满足交换律, 即 $u \otimes v \neq v \otimes u$ 是两个不同的张量.

定义 8.6.4: 线性空间的张量积

若 U, V 是两个线性空间, 各自有一组基 $\{u_1, \dots, u_m\}, \{v_1, \dots, v_n\}$, 定义基的张量积 $\{u_i \otimes v_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ 张成的线性空间为 U, V 的张量积, 记为 $U \otimes V$.

推论. 由定义

$$\dim(U \otimes V) = \dim(U) \dim(V).$$

例 8.6.2

$\forall u \in U, v \in V, u = x_1 u_1 + \cdots + x_m u_m, v = y_1 v_1 + \cdots + y_n v_n$, 则

$$u \otimes v = \sum_{i=1}^m \sum_{j=1}^n x_i y_j u_i \otimes v_j \in U \otimes V.$$

但并不是所有 $U \otimes V$ 的元素都能写成 $u \otimes v$ 的形式.

例 8.6.3: 张量空间 $V \otimes V$

$V \otimes V$ 是所有 $V^* \times V^* \rightarrow \mathbb{R}$ 的双线性函数构成的线性空间, $\forall v \in V \otimes V$ 都可以写为

$$v = \sum_{i,j} v^{ij} v_i \otimes v_j.$$

称作二阶逆变张量 (contravariant tensor), 记作 $(2,0)$ 张量.

换基时,

$$v^{k\ell} = \sum_{i,j} A^k_i A^\ell_j v^{ij}, \quad v^{ij} = \sum_{k,\ell} (A^{-1})^i_k (A^{-1})^j_\ell v^{k\ell}.$$

逆变 (contravariant) 的含义: 换基时分量每个指标对应的变换矩阵是基的变换矩阵的逆矩阵.

例 8.6.4: 混合张量 $V \otimes V^*$

$(1,1)$ 张量 $v \in V \otimes V^*$ 可以写成

$$v = \sum_{i,j} v^i_j v_i \otimes v^{*j}.$$

例 8.6.5: $V \otimes \cdots \otimes V \otimes V^* \otimes \cdots \otimes V^* = V^{\otimes k} \otimes V^{*\otimes \ell}$

$V \otimes \cdots \otimes V \otimes V^* \otimes \cdots \otimes V^* = V^{\otimes k} \otimes V^{*\otimes \ell}$ 的基

$$\{v_{i_1} \otimes \cdots \otimes v_{i_k} \otimes v^{*j_1} \otimes \cdots \otimes v^{*j_\ell} \mid 1 \leq i_1, \dots, i_k, j_1, \dots, j_\ell \leq n\}$$

$V^{\otimes k} \otimes V^{*\otimes \ell}$ 的元素

$$v = \sum_{\substack{i_1, \dots, i_k \\ j_1, \dots, j_\ell}} (v^{i_1 \cdots i_k}_{j_1 \cdots j_\ell}) v_{i_1} \otimes \cdots \otimes v_{i_k} \otimes v^{*j_1} \otimes \cdots \otimes v^{*j_\ell}.$$

是 (k, ℓ) 阶张量. 基变换

$$v^{i_1 \cdots i_k}_{j_1 \cdots j_\ell} = \sum_{\substack{p_1, \dots, p_k \\ q_1, \dots, q_\ell}} (A^{-1})^{i_1}_{p_1} \cdots (A^{-1})^{i_k}_{p_k} (v^{p_1 \cdots p_k}_{q_1 \cdots q_\ell}) A^{q_1}_{j_1} \cdots A^{q_\ell}_{j_\ell}.$$

第九章 复线性空间

这一章我们将数域由实数域 \mathbb{R} 扩展至复数域 \mathbb{C} ，复数的定义和运算高中已经讲过，也可参见复变函数的笔记。在此略。

复数构成的向量 z 的共轭即将其中所有元素取共轭，记作 \bar{z} 。共轭转置记作 $z^\dagger := \bar{z}^\top$ 。

所有实线性空间的知识都可以推广到复线性空间，只需要把原来是实数的地方换成复数。

9.1 内积和内积空间

定义 9.1.1: \mathbb{C}^n 标准内积

复向量 u, v 的内积

$$u^\dagger v = \sum_{i=1}^n \bar{u}_i v_i = \bar{u}_1 v_1 + \cdots + \bar{u}_n v_n.$$

一般复线性空间 V 的内积 $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$:

- 交换共轭: $\langle u, v \rangle = \overline{\langle v, u \rangle}$;
- 对第二个变量线性: $\langle u, cv \rangle = c \langle u, v \rangle$, $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$;
- 正定: $\langle u, u \rangle \geq 0$ 当且仅当 $u = 0$ 时取等号。

注. 对第一个变量不是简单的线性，而是多一个复共轭:

$$\langle cu, v \rangle = \bar{c} \langle u, v \rangle, \quad \langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle.$$

定理 9.1.1

只需要知道基之间的内积就可以算出任意向量之间的内积。

证明. v_1, \dots, v_n 是 V 上的一组基, $\forall u, w \in V$, $u = u^1 v_1 + \cdots + u^n v_n$, $w = w^1 v_1 + \cdots + w^n v_n$

$$\langle u, w \rangle = \langle u^1 v_1 + \cdots + u^n v_n, w^1 v_1 + \cdots + w^n v_n \rangle = \sum_{i,j} \bar{u}^i w^j \langle v_i, v_j \rangle. \quad \square$$

例 9.1.1: 内积与对偶空间

V 的对偶空间 V^* 是所有 $V \rightarrow \mathbb{C}$ 的线性函数的集合。通过内积可以建立 V, V^* 的一一映射

$$\forall v \in V, g_v \in V^*, \quad g_v(w) := \langle v, w \rangle.$$

例 9.1.2: Legendre 多项式

所有不高于 n 的实系数多项式

$$f(x) = a_0 + a_1x + \cdots + a_nx^n,$$

构成线性空间 $\mathscr{P}^n(\mathbb{R})$, 显然 $\{1, x, x^2, \dots, x^n\}$ 构成 $\mathscr{P}^n(\mathbb{R})$ 的一组基. 定义内积

$$\langle f, g \rangle := \int_{-1}^1 f(x)g(x) dx,$$

用 Gram-Schmidt 法则将 $\{1, x, x^2, \dots, x^n\}$ 变成一组正交基

$$\begin{aligned} P_0 &= 1, \\ P_1 &= x - \frac{\langle P_0, x \rangle}{\langle P_0, P_0 \rangle} P_0 = x, \\ P_2 &= x^2 - \frac{\langle P_1, x^2 \rangle}{\langle P_1, P_1 \rangle} P_1 - \frac{\langle P_0, x^2 \rangle}{\langle P_0, P_0 \rangle} P_0 = x^2 - \frac{2}{3}, \\ P_3 &= x^3 - \frac{\langle P_2, x^3 \rangle}{\langle P_2, P_2 \rangle} P_2 - \frac{\langle P_1, x^3 \rangle}{\langle P_1, P_1 \rangle} P_1 - \frac{\langle P_0, x^3 \rangle}{\langle P_0, P_0 \rangle} P_0 = x^3 - \frac{3}{5}x, \\ &\dots \end{aligned}$$

这与实际 Legendre 多项式的定义只是系数的差别.

例 9.1.3: Hermite 多项式

在 $\mathscr{P}^n(\mathbb{R})$ 内定义内积

$$\langle f, g \rangle := \int_{-\infty}^{+\infty} f(x)g(x) e^{-x^2/2} dx,$$

用 Gram-Schmidt 法则将 $\{1, x, x^2, \dots, x^n\}$ 变成一组正交基

$$\begin{aligned} H_0 &= 1, \\ H_1 &= x - \frac{\langle H_0, x \rangle}{\langle H_0, H_0 \rangle} H_0 = x, \\ H_2 &= x^2 - \frac{\langle H_1, x^2 \rangle}{\langle H_1, H_1 \rangle} H_1 - \frac{\langle H_0, x^2 \rangle}{\langle H_0, H_0 \rangle} H_0 = x^2 - 1, \\ H_3 &= x^3 - \frac{\langle H_2, x^3 \rangle}{\langle H_2, H_2 \rangle} H_2 - \frac{\langle H_1, x^3 \rangle}{\langle H_1, H_1 \rangle} H_1 - \frac{\langle H_0, x^3 \rangle}{\langle H_0, H_0 \rangle} H_0 = x^3 - 3x, \\ &\dots \end{aligned}$$

这与实际 Hermite 多项式的定义也只是系数的差别.

9.2 Hermite 矩阵

定义 9.2.1: Hermite 矩阵

方阵 H 是厄米 (Hermite) 矩阵若 $H^\dagger = H$.

Hermite 矩阵是对称矩阵 (见定义 6.5.1) 在复空间的推广.

例 9.2.1: Pauli 矩阵

给出三个 Pauli 矩阵

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

$\sigma_1, \sigma_2, \sigma_3$ 都是 Hermite 的, 且

$$\sigma_i \sigma_j = i \sigma_k, \quad (ijk) = (123).$$

定理 9.2.1: Hermite 矩阵的二次型

$\forall z \in \mathbb{C}, z^\dagger H z$ 是实数.

证明. $(z^\dagger H z)^\dagger = z^\dagger H^\dagger z = z^\dagger H z$. □

定理 9.2.2: Hermite 矩阵的特征值

Hermite 矩阵 H 的特征值都是实数.

证明. $H z = \lambda z$, 左乘 z^\dagger 得 $z^\dagger H z = \lambda z^\dagger z$, 由 $z^\dagger H z, z^\dagger z$ 均是实数知, λ 也是实数. □

定理 9.2.3: Hermite 矩阵的特征向量

Hermite 矩阵 H 不同特征值对应的特征向量正交.

证明. $H z_1 = \lambda_1 z_1, H z_2 = \lambda_2 z_2, \lambda_1 \neq \lambda_2$

$$\lambda_1 z_2^\dagger z_1 = z_2^\dagger H z_1 = (z_1^\dagger H z_2)^\dagger = (\lambda_2 z_1^\dagger z_2)^\dagger = \lambda_2 z_2^\dagger z_1.$$

故 $z_2^\dagger z_1 = 0$. □

定理 9.2.4: 谱定理

Hermite 矩阵的特征向量构成 \mathbb{C}^n 中的一组么正基.

$$H = Q \Lambda Q^\dagger.$$

证明. 略. □

9.3 幺正矩阵

定义 9.3.1: 幺正矩阵

矩阵 U 是幺正的 (unitary) 若 $U^\dagger U = I$.

幺正矩阵是正交矩阵 (见定义 4.4.2) 在复空间的推广.

定理 9.3.1: 幺正变换

幺正变换保持复向量的模不变.

证明.

$$\|Uz\|^2 = z^\dagger U^\dagger U z = z^\dagger z = \|z\|^2. \quad \square$$

定理 9.3.2: 幺正矩阵的行列式

$|\det(U)| = 1$.

证明.

$$1 = \det(U^\dagger U) = \det(U^\dagger) \det(U) = \overline{\det(U)} \det(U) = |\det(U)|^2. \quad \square$$

第十章 群、环、域

10.1 二元运算

定义 10.1.1: 二元运算

集合 S 上的一个二元运算 (binary operation) 形如映射 $\circ: S \times S \rightarrow S$.
其中 $S \times S \equiv S^2$ 是笛卡尔积 (Cartesian product),

$$A \times B := \{(a, b) \mid a \in A, b \in B\}.$$

二元运算在 S 上是封闭的 (property of closure).

定义 10.1.2: 恒等元

$e \in S$ 是恒等元 (identity element), 若 $\forall a \in S, e \circ a = a \circ e = a$.

定义 10.1.3: 可逆

$a \in S$ 是可逆的 (invertible), 若 $\exists a^{-1} \in S, a \circ a^{-1} = a^{-1} \circ a = e$.

特别地, 简记

$$a^m \equiv a \circ \cdots \circ a, \quad a^{-m} \equiv a^{-1} \circ \cdots \circ a^{-1}.$$

10.2 群与子群

定义 10.2.1: 群

群 (group) 是有二元运算 $\circ: G \times G \rightarrow G$ 和集合 G 并满足下列性质的组合 (G, \circ) :

- 结合律: $(a \circ b) \circ c = a \circ (b \circ c) \equiv a \circ b \circ c$;
- 单位元: $\exists e \in G$ 使得 $e \circ a = a \circ e = a$;
- 逆: $\forall a \in G, \exists a^{-1}$ 使得 $a^{-1} \circ a = a \circ a^{-1} = e$.

若还满足交换律, 则称为交换群或 Abel 群.

群的阶 (order) $\text{ord}(G)$ 表示其元素的个数. 群可分为有限群和无限群.

定理 10.2.1: 单位元和逆元的唯一性

在群中只能有一个单位元，而群中的每个元素都正好有一个逆元素。

证明. 若一个群存在两个单位元 e, e' ，则

$$e = e \circ e' = e';$$

若一个元素 a 存在两个逆 b, c ，则

$$b = b \circ e = b \circ (a \circ c) = (b \circ a) \circ c = e \circ c = c.$$

□

例 10.2.1: 群的例子

- 整数加群 $(\mathbb{Z}, +)$: 单位元 0;
- 非零实数乘法群 $(\mathbb{R} \setminus \{0\}, \times)$: 单位元 1;
- 一般线性 (general linear) 群 $GL(n)$: 所有 n 阶可逆矩阵集合，单位元 I_n .

定理 10.2.2: 消去律

$\forall a, b, c \in G$ ，有

$$a \circ b = a \circ c \implies b = c; \quad (10.1a)$$

$$b \circ a = c \circ a \implies b = c; \quad (10.1b)$$

$$b \circ a = a \text{ 或 } a \circ b = a \implies b = e. \quad (10.1c)$$

证明. 左乘/右乘 a^{-1} .

□

注. 逆 a^{-1} 的存在很关键，如果 G 上的运算只是结合的，则 (G, \circ) 是一个半群 (semigroup)，有单位元的半群又叫么半群 (monoid).

定义 10.2.2: 置换群和对称群

给定有限集合 T ，所有可逆映射 $f: T \rightarrow T$ 构成一个群 $\text{sym}(T)$ ，运算是映射的复合，称做置换群 (permutation group)

当 $T = \{1, 2, \dots, n\}$ 时，对应的置换群称为对称群 (symmetric group) S_n .

例 10.2.2: S_2

$S_2 = \{1, p\}$ ，其中

$$1 = \text{id}: \{1, 2\} \rightarrow \{1, 2\},$$

$$p: \{1, 2\} \rightarrow \{1, 2\}, \quad p(1) = 2, \quad p(2) = 1.$$

S_2 是交换群. 可列出 Cayley 表

	1	p
1	1	p
p	p	1

例 10.2.3: S_3

S_3 : 定义生成元 x, y 满足:

$$x(1) = 2, \quad x(2) = 3, \quad x(3) = 1;$$

$$y(1) = 2, \quad y(2) = 1, \quad y(3) = 3.$$

可以证明生成元之间的关系: $x^3 = 1, y^2 = 1, x^2y = yx$, 故 S_3 中所有元素都能写成生成元的积:

$$S_3 = \{1, x, x^2, y, xy, x^2y\},$$

易知 S_3 不交换. 根据生成元, S_3 还可写为 $S_3 = \{x, y \mid x^3 = 1, y^2 = 1, x^2y = yx\}$. 生成元及其关系称作一个群的表现 (presentation), 一个群的表现不唯一.

定义 10.2.3: 子群

$H \subset G$ 是 G 的子群 (subgroup), 若 H 满足

- 封闭性: $\forall a, b \in H, a \circ b \in H$;
- 单位元: $e \in H$;
- 逆元: $\forall a \in H, \exists a^{-1} \in H$.

$\{e\}$ 和 G 都是平凡子群, 其他子群称为真子群 (proper subgroup).

例 10.2.4: 子群的例子

- 圆群: $(\{z \in \mathbb{C} \mid |z| = 1\}, \times) \subset (\mathbb{C} \setminus \{0\}, \times)$;
- 特殊线性群 $SL(n) \subset GL(n)$: 所有行列式为 1 的 n 阶方阵;

定义 10.2.4: 循环群

循环群 (cyclic group) 是

$$Z_n \equiv \{1, x, \dots, x^{n-1} \mid x^n = 1\}, \quad (10.2)$$

其生成元为 x .

例 10.2.5

S_3 有两个子群是循环群: $\{x^k \mid x^3 = 1\} = Z_3$ 和 $\{y^k \mid y^2 = 1\} = Z_2$.

10.3 群同态

定义 10.3.1: 群同态

$(G, \circ), (G', \circ')$ 是群, 映射 $\phi: G \rightarrow G'$ 是群同态 (group homomorphism) 若 $\forall a, b \in G$

$$\phi(a \circ b) = \phi(a) \circ' \phi(b). \quad (10.3)$$

也称映射 ϕ 和群上的乘法相容 (compatible).

定理 10.3.1: 群同态下的单位元和逆元

若 $\phi: G \rightarrow G'$ 是群同态, G, G' 的单位元分别为 $1, 1'$, $a \in G$, 则

$$\phi(1) = 1', \quad \phi(a^{-1}) = \phi(a)^{-1}. \quad (10.4)$$

证明. (1) 由 $\phi(1) = \phi(1 \circ 1) = \phi(1) \circ' \phi(1)$, 再运用消去律可得 $1' = \phi(1)$;

(2) 由 $\phi(a^{-1}) \circ' \phi(a) = \phi(a^{-1} \circ a) = \phi(1) = 1'$ 可得 $\phi(a^{-1}) = \phi(a)^{-1}$. \square

例 10.3.1: 线

空间和 $+$ 构成一个群, 线性映射都是群同态.

定义 10.3.2: 群同态的像

群同态 $\phi: G \rightarrow G'$ 的像 (image) $\text{im } \phi$ 定义为

$$\text{im } \phi \equiv \{x \in G' \mid \exists a \in G, \phi(a) = x\}. \quad (10.5)$$

定义 10.3.3: 群同态的核

群同态 $\phi: G \rightarrow G'$ 的核 (kernel) $\ker \phi$ 定义为

$$\ker \phi \equiv \{a \in G \mid \phi(a) = 1'\}. \quad (10.6)$$

定理 10.3.2

若 $\phi: G \rightarrow G'$ 是群同态, 则 $\text{im } \phi$ 是 G' 的子群, $\ker \phi$ 是 G 的子群.

证明. 考虑线性空间在 $+$ 下构成的群, 此时线性映射作为群同态的像与核同之前线性映射的像与核相同. \square

定义 10.3.4: 左陪集

H 是 G 的子群, $a \in G$, 则

$$a \circ H \equiv \{a \circ h \mid h \in H\} \quad (10.7)$$

是 H 在 G 下的一个左陪集 (left coset). 同理可定义右陪集.

定理 10.3.3

群同态 $\phi: G \rightarrow G'$, $a, b \in G$, 则以下命题等价:

1. $\phi(a) = \phi(b)$;
2. $a^{-1} \circ b \in \ker \phi$;
3. $b \in a \circ \ker \phi$;
4. $b \circ \ker \phi = a \circ \ker \phi$.

证明. (1) \Rightarrow (2):

$$\phi(a) = \phi(b) \Rightarrow 1' = \phi(a)^{-1} \circ' \phi(b) = \phi(a^{-1} \circ b) \Rightarrow a^{-1} \circ b \in \ker \phi;$$

(2) \Rightarrow (3):

$$a^{-1} \circ b = h \in \ker \phi \Rightarrow b = a \circ h \in a \circ \ker \phi.$$

(3) \Rightarrow (4): 由 $b \in a \circ \ker \phi$, $\exists h \in \ker \phi$ 使得 $b = a \circ h$; $\forall b' \in b \circ \ker \phi$, $\exists h' \in \ker \phi$ 使得 $b' = b \circ h' = (a \circ h) \circ h' = a \circ (h \circ h') \in a \circ \ker \phi$, 故 $b \circ \ker \phi \subset a \circ \ker \phi$; 同理由 $a = b \circ h^{-1}$ 可以证明 $a \circ \ker \phi \subset b \circ \ker \phi$, 故 $a \circ \ker \phi = b \circ \ker \phi$.

(4) \Rightarrow (1): $\forall h \in \ker \phi$, $\exists h' \in \ker \phi$ 使得 $a \circ h = b \circ h'$

$$\Rightarrow \phi(a \circ h) = \phi(b \circ h') \Rightarrow \phi(a) \circ' \phi(h) = \phi(b) \circ' \phi(h')$$

$$\Rightarrow \phi(a) \circ' 1' = \phi(b) \circ' 1' \Rightarrow \phi(a) = \phi(b).$$

故以上 4 个命题等价. □

注.

- 群同态的核不仅告诉我们 G 中的哪些元素映射到 1, 也告诉我们哪些元素的像相同;
- 上面的命题在线性方程组中的应用就是 $Ax = b$ 的通解 = 特解 + $\{Ax = 0\}$ 的通解.

推论. 群同态 $\phi: G \rightarrow G'$ 是单射 $\iff \ker \phi = \{1\}$.

定义 10.3.5: 正规子群

$N \subset G$ 是 G 的正规子群 (normal subgroup) 若 $\forall a \in N, \forall g \in G$, 共轭 $g \circ a \circ g^{-1} \in N$.

定理 10.3.4

$\phi: G \rightarrow G'$ 是群同态, $\ker \phi$ 是 G 的正规子群.

定义 10.3.6: 中心

群 G 的中心 (center) 是

$$Z_G \equiv \{z \in G \mid z \circ x = x \circ z, \forall x \in G\}, \quad (10.8)$$

中心 Z_G 总是 G 的正规子群.

例 10.3.2

- 行列式是 GL 的群同态:

$$\det : GL(n) \rightarrow (\mathbb{R} \setminus \{0\}, \times)$$

核 $\ker \det = SL(n)$ 是 $GL(n)$ 的正规子群.

- $Z_{SL(2)} = \{I, -I\}$;
- $Z_{S_n} = \{1\}, n \geq 3$.

10.4 群同构

定义 10.4.1: 群同构

若群同态 $\phi : G \rightarrow G'$ 是双射, 则称 ϕ 为群同构 (isomorphism), 称 G, G' 是同构的 (isomorphic), 记作 $G \simeq G'$.

群到自己的同构 $\phi : G \rightarrow G$ 也叫自同构. (automorphism)

恒等映射 $\text{id} : G \rightarrow G$ 是自同构.

例 10.4.1: 群同构的例子

- 指数函数

$$\exp : (\mathbb{R}, +) \rightarrow (\mathbb{R}_{>0}, \times), x \mapsto e^x.$$

- P 是投影矩阵: $P^2 = P$

$$S_2 \rightarrow \{I, I - 2P\}.$$

10.5 等价关系

定义 10.5.1: 等价关系

集合 S 上的等价关系 \sim 是 S 中两个元素 a, b 之间的关系, 记作 $a \sim b$, 满足:

- 传递性: $a \sim b, b \sim c \implies a \sim c$;
- 对称性: $a \sim b \implies b \sim a$;
- 自反性: $\forall a, a \sim a$.

注.

- 等价关系可以看成 $=$ 的抽象;
- 等价关系可以理解为映射 $f: S \times S \rightarrow \{0, 1\}$ 满足

$$a \sim b \iff f(a, b) = 1.$$

- 未竟