线性代数

主要整理自颜文斌老师讲义

by Dait at THU

2021/12/28 - 2023/9/5

目录

1 向量和矩阵

1.1 向量

定义 1.1.1: 向量

在数域 (number feild) $\mathbb{F}$[I] 内，一个 n 维向量 (vector) v 由 n 个标量 (scalar) $v_1, \ldots, v_n$ 组成，记作

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \quad v_i \in \mathbb{F}.$$

组成向量的标量 $v_1, \ldots, v_n$ 称为向量的分量 (component).

[I] 直到第 9 章复线性空间之前，均只考虑实数域 $\mathbb{R}$，即 $\mathbb{F} = \mathbb{R}$.

定义 1.1.2: 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；
v 的反向量 (reverse vector) 对应每个分量取相反数，记作 $-v$.

定义 1.1.3: 向量的加法

向量的加法 (addition) 即对应分量相加，

$$v + u = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} v_1 + u_1 \\ \vdots \\ v_n + u_n \end{bmatrix}.$$

因此只有分量数相同的向量之间才可以相加.

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

定义 1.1.4: 向量的数乘

向量与标量的数乘 (scalar product) 即每个分量乘标量

$$cv = c\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ \vdots \\ cv_n \end{bmatrix}.$$

- $1v = v,\ (-1)v = -v,\ 0v = 0.$
- 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 1.1.5: 线性组合

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

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

以此类推，n 个向量 $v_1, v_2, \ldots, v_n$ 的线性组合形如

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n, \quad c_i \in \mathbb{F}.$$

定义 1.1.6: 向量的内积

向量 v 和 u 的内积 (inner product) 结果是一个标量，其值为

$$v \cdot u := \sum_{i=1}^{n} v_iu_i = v_1u_1 + v_2u_2 + \cdots + v_nu_n. \tag{1.1}$$

特别的，$v^2 := v \cdot v.$

- 交换律：$v \cdot u = u \cdot v;$
- 与数乘的结合律：$(cv) \cdot u = c(v \cdot u) \equiv cv \cdot u;$
- 分配律：$(v+u) \cdot w = v \cdot w + u \cdot w;$
- $v^2 \geqslant 0$，取等号当且仅当 $v = 0.$

定义 1.1.7: 向量的长度

通过内积我们可以定义向量的长度 (norm)

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

单位向量 (unit vector) 就是长度为 1 的向量.

显然，和 v 同向的单位向量是 $\hat{v} := v / \|v\|$.

1.2 矩阵

形式上看，标量 c 是 1×1 的，向量 v 是 $m \times 1$ 的[I]，继而可定义 $m \times n$ 的矩阵 (matrix) 形如

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

A_{ij} 是矩阵第 i 行第 j 列的元素.

矩阵的行数和列数分别为 $\mathrm{row}(A) = m$，$\mathrm{col}(A) = n$. 当 $m = n$ 时，矩阵是 n 阶方阵 (square matrix).

矩阵的加法和数乘与向量的运算规律相同，是平凡的.

定义 1.2.1: 矩阵和向量的乘法

$m \times n$ 矩阵 A 乘 n 维向量 x，结果是一个 m 维向量 $b = Ax$，其分量为:

$$b_i = \sum_{j=1}^{n} A_{ij} x_j, \quad i = 1, \ldots, m. \tag{1.3}$$

因此 Ax 是看成 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写成矩阵的形式:

$$Ax = b.$$

这种思想有助于我们掌握线性代数的理念.

[I]如无特别说明，向量均默认为列向量.

定义 1.2.2: 矩阵的乘法

$m \times n$ 矩阵 A 乘 $n \times p$ 矩阵 B，结果是一个 $m \times p$ 的矩阵 $C = AB$，其分量为

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}. \tag{1.4}$$

因此若 A, B 可以相乘，要求 $\mathrm{col}(A) = \mathrm{row}(B)$.

- 一般不满足交换律，即 $AB \neq BA$，比如
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$
- 左分配律：$(A+B)C = AC + BC$;
- 右分配律：$A(B+C) = AB + AC$.

定理 1.2.1: 矩阵乘法的结合律

$$(AB)C = A(BC) \equiv ABC. \tag{1.5}$$

证明： 对于 $A_{m\times n}, B_{n\times p}, C_{p\times q}$

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

定义 1.2.3: 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：

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

易证，$\forall m \times n$ 的矩阵 A，均有

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

定理 1.2.2: 交换性

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

证明：　取 $B = e_{ij}$，表示仅 i 行 j 列为 1，其余项均为 0

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

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

定义 1.2.4: 分块矩阵

可以将矩阵分块，每一块 (block) 是一个小矩阵，比如

$$\left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \equiv \begin{bmatrix} I & I \\ I & I \end{bmatrix}.$$

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.

1.3　矩阵的逆和转置

定义 1.3.1: 矩阵的逆

方阵 A 的逆矩阵 (inverse matrix) A^{-1} 满足

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

称奇异矩阵 (singular matrix) 是不可逆的矩阵.

若 A, B 均可逆，则 $(AB)^{-1} = B^{-1}A^{-1}$. 但 $A + B$ 不一定可逆.

例 1.3.1: 二阶方阵的逆

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

例 1.3.2: 对角矩阵的逆

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

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

对角矩阵有很多简单的性质，比如对角矩阵的乘法很简单

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

显然，对角矩阵的逆为

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

定理 1.3.1: 左逆和右逆

若存在左逆 (left inverse)，则其也是右逆 (right inverse).

证明： 设 B 是 A 的左逆，即 $BA = I$. 构造映射 f, g

$$f(X) := AXB, \quad g(X) := BXA,$$

则 $g \circ f(X) = BAXBA = X$，g 是双射．又

$$g(AB) = BABA = I = g(I),$$

故 $AB = I$. □

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

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

因为

$$I = I^n + A^n = (I + A)\left[I^{n-1} - I^{n-2}A + \cdots + (-A)^{n-1}\right]. \qquad \square$$

定义 1.3.2: 矩阵的转置

$m \times n$ 矩阵 A 的转置 (transpose) $A^\top$ 是 $n \times m$ 矩阵，且

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

特别的，若 $S^\top = S$，则 S 是对称矩阵 (sysmmetric matrix).

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

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

$$\operatorname{tr}(A) := \sum_{i=1}^{n} A_{ii}. \tag{1.9}$$

- $\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$，$\operatorname{tr}(cA) = c\operatorname{tr}(A)$;
- $\operatorname{tr}(A^\top) = \operatorname{tr}(A)$;
- 两个 n 维列向量 u, v，有 $\operatorname{tr}(uv^\top) = v^\top u$
$$\operatorname{tr}(uv^\top) = \sum_{i=1}^{n} u_i v_i = u^\top v = v^\top u.$$
- $\operatorname{tr}(AB) \neq \operatorname{tr}(A)\operatorname{tr}(B)$.

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

$$\operatorname{tr}(AB) = \operatorname{tr}(BA). \tag{1.10}$$

证明：

$$\operatorname{tr}(AB) = \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ji}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji} A_{ij} = \sum_{j=1}^{n} (BA)_{jj} = \mathrm{tr}(BA). \qquad \square$$

除此之外，迹还有一些重要性质，将在后面讲.

2 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)

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

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: 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

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

而主元 (pivot element) 就是每个方程第一个非 0 系数.

消元法失效：主元数目 $<$ 未知数

- 得到 $0 \neq 0$，无解；
- 得到 $0 = 0$，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: 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. 从第 2 个到第 n 个方程中消去 x_1 (方程 $i - \ell_{i1}\times$方程 1).
3. 得到第 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. 最后结果要么是一个上三角方程组，要么失效.
5. 上三角的情况，从最后一个方程开始解出全部未知数.

利用消元法，可以求矩阵的逆.

方法 2.2: Gauss-Jordan 消元法

对增广矩阵 (augmented matrix) (A, I) 做消元操作.

$$(A, I) \quad \to \quad (I, A^{-1}).$$

就可以得到 A 的逆 A^{-1}.

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 (A, b) 并对其消元. 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

- 对换：交换两行
- 倍加：一行乘系数加到另一行
- 倍乘：一行乘以一个非零系数

如果一个矩阵可以行变换成另一个矩阵，则它们行等价 (equivalence).

定义 2.2.2: 行阶梯矩阵

行阶梯矩阵 (row echelon matrix) M 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. 每个主元都在它上面行的主元的右边的列；
- 同一列中在主元下面的元素都是 0.

比如，形如 ($*$ 表示非 0 项，$\cdot$ 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.

若 A 与行阶梯矩阵 U 行等价，记作 $U \in \text{ref}(A)$. 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 2.2.3: 约化行阶梯矩阵

约化行阶梯矩阵 (reduced row echelon matrix) 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.

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.

若 A 与约化行阶梯矩阵 U 行等价，记作 $U = \text{rref}(A)$. 可以证明，约化行阶梯矩阵是唯一的. 唯一性证明见 Lay 书的附录 A. 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 $\big(\text{rref}(A), b'\big)$；
- 解的存在性：若 $\text{rref}(A)$ 有 0 行，但 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 2.2.4: 初等矩阵

对 $m \times n$ 的矩阵 A 行变换，等价于用 $m \times m$ 初等矩阵 (ementary matrix) 左乘 A，初等矩阵有以下三种类型：

- 倍加：A 的第 i 行乘一个非 0 常数 a 再加到第 j 行

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

- 置换：置换 A 的第 i 行和第 j 行

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

- 倍乘：A 的第 i 行乘一个非 0 常数 c

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

初等矩阵的性质是均可逆：

- 倍加矩阵：$E = I + ae_{ij}$，$E^{-1} = I - ae_{ij}$;
- 置换矩阵与自己互逆；
- 倍乘矩阵：$E = I + (c-1)e_{ii}$，$E^{-1} = I + (c^{-1}-1)e_{ii}$.

消元法：用一系列初等矩阵 $\{E_i\}$ 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.

U 是和 A 等价的行阶梯矩阵，U 是上三角的

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

从 A 到 U 的过程中我们只需消去主元下面的元素，$E_1, E_2, \ldots, E_k$ 及他们的逆 $E_1^{-1}, E_2^{-1}, \ldots, E_k^{-1}$ 都是下三角的，故

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

也是下三角的．进而

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

如果 A 化成行阶梯矩阵 U 的过程中没有置换，则 A 有一个 LU 分解；反之，则存在一个置换矩阵 P，使得 PA 有一个 LU 分解．

3 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 $+: V \times V \to V$ 和数乘 $\cdot: \mathbb{F} \times V \to V$ 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 $(c_1c_2)x = c_1(c_2x)$;
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 $(c_1 + c_2)x = c_1x + c_2x$.

比如 $\mathbb{R}^n$ 和 $\mathbb{C}^n$ 都是线性空间.

定义 3.1.2: 子空间

线性空间 V 的子空间 (subspace) $V_s \subset V$，且对于加法和数乘封闭：$\forall v, w \in V_s, \forall c \in \mathbb{F}$

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

即子空间中元素的线性组合都在同一个子空间.

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

一般来说，线性空间 V 的子集 S 不是子空间，但我们可以从 S 中构造出子空间.

定义 3.1.3: 线性扩张

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

向量是否线性独立同数域的选择密切相关．

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．

$\{e_i\}$ 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) $\dim(V)$ 等于任一组基中向量的个数．

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，系数矩阵一定有自由列，所以有非零解．这与 $\{v_i\}$ 线性无关矛盾！故 $m = n$. □

若 $\forall v \in V_1$ 可以写成 V_2 中向量的线性组合，则 $\dim(V_1) \leqslant \dim(V_2)$．这个定理是 trivial 的，证明留给读者．

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

3.3 矩阵 A 的四个子空间

对于 $m \times n$ 矩阵 A，可以由其得到四个子空间：列空间 $\mathrm{C}(A)$、行空间 $\mathrm{C}(A^\top)$、零空间 $\mathrm{N}(A)$ 和左零空间 $\mathrm{N}(A^\top)$．

定义 3.3.1: 列空间

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

类似的，行空间 (row space) 是所有行的线性组合的集合，由于转置并不影响性质，行空间可以用 $\mathrm{C}(A^\top)$ 表示．不难验证，$\mathrm{C}(A)$ 是 $\mathbb{R}^n$ 的子空间，$\mathrm{C}(A^\top)$ 是 $\mathbb{R}^m$ 的子空间．

线性方程组 $Ax = b$ 有解等价于 $b \in \mathrm{C}(A)$．

定义 3.3.2: 零空间

矩阵 A 的零空间 (null space) $\mathrm{N}(A)$ 是 $Ax = 0$ 所有解 x 构成的线性空间．

类似的，左零空间 (left null space) 是 $x^\top A = 0$ 所有解 x 构成的线性空间，可用 $\mathrm{N}(A^\top)$ 表示．$\mathrm{N}(A)$ 是 $\mathbb{R}^m$ 的子空间，$\mathrm{N}(A^\top)$ 是 $\mathbb{R}^n$ 的子空间．

子空间的基　显然 $\mathrm{N}(A) = \mathrm{N}\big(\mathrm{rref}(A)\big)$，$\mathrm{N}\big(\mathrm{rref}(A)\big)$ 中的基可以由这样给出：

- 每个自由列给出一个向量，因此 $\dim\big(\mathrm{N}(A)\big) = \mathrm{rref}(A)$ 中自由列的数量；

- 自由列 j 对应向量 x 中，$x_j = 1$，x 对应其他自由列分量为 0，对应主列 i 分量 $x_i = -\operatorname{rref}(A)_{ij}$

$\operatorname{rref}(A)$ 所有主列构成 $\mathrm{C}(A)$ 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) $\operatorname{rank}(A)$ 定义为行空间或列空间的维数[I]：

$$\operatorname{rank}(A) := \dim\big(\mathrm{C}(A)\big) = \dim\big(\mathrm{C}(A^\top)\big).$$

进而可定义行满秩矩阵 (full row rank matrix) 满足

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

列满秩矩阵 (full column rank matrix) 定义类似.

[I]后面很快会证明行空间列空间维数相等.

线性方程组 $Ax = 0$ 的完整解 考虑线性方程组 $Ax = b$，通解

$$x = x_p + x_n,$$

其中 $x_n \in \mathrm{N}(A)$，特解 x_p 可以从约化的增广矩阵 $\big(\operatorname{rref}(A), b'\big)$ 中选取：自由列 i 对应 $x_{pi} = 0$，若主元列 i 中的 1 在第 j 行，则 $x_{pi} = b'_j$.

若 A 列满秩，$\operatorname{rref}(A)$ 没有自由列，$\mathrm{N}(A) = \{0\}$，有唯一解 ($b \in \mathrm{C}(A)$) 或无解；若 A 行满秩，$\operatorname{rref}(A)$ 没有零行，$\mathrm{C}(A) = \mathbb{R}^m$，$\forall b$ 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 3.4.1: 初等变换和子空间

- $\mathrm{N}(A) = \mathrm{N}(EA)$，因为

$$Ax = 0 \quad \Leftrightarrow \quad EAx = 0.$$

- $\dim\big(\mathrm{C}(A)\big) = \dim\big(\mathrm{C}(EA)\big)$，因为

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

- $\mathrm{C}(A) = \mathrm{C}(AE)$，因为

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

- $\dim\big(\mathrm{N}(A)\big) = \dim\big(\mathrm{N}(AE)\big)$，因为可由

$$Ax = 0 \quad \Leftrightarrow \quad AE(E^{-1}x) = 0,$$

推出

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

即，矩阵 A 在初等变换下，$\dim\big(\mathrm{C}(A)\big)$ 和 $\dim\big(\mathrm{N}(A)\big)$ 均不变，而行变换下 $\mathrm{N}(A)$ 不变，列变换下 $\mathrm{C}(A)$ 不变.

因此，可以将 A 先由行变换为 $\mathrm{rref}(A)$，再列变换为

$$\tilde{I} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

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

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

以上的这些量在初等变化下都不变，故

定理 3.4.2: 线性代数基本定理 · 一

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

推论是方阵 A 可逆 $\;\Leftrightarrow\;$ A 满秩.

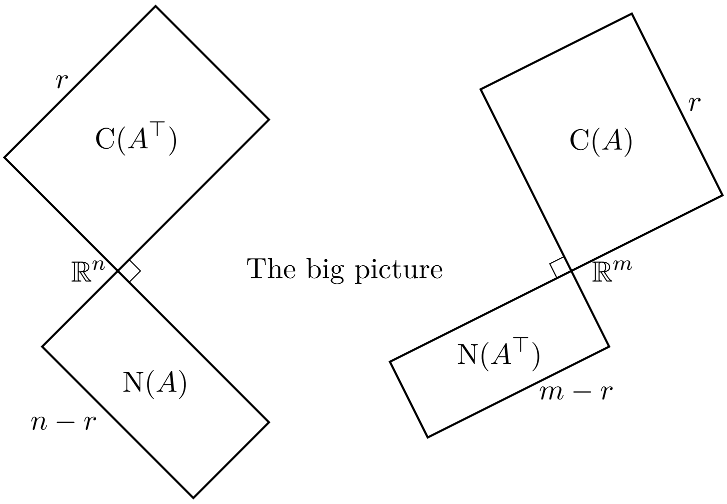

图 1　A 的四个子空间

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

定理 3.4.3

4 正交性

将列向量看做矩阵，则向量 v 和 w 的内积可以看做 $v \cdot w = v^\top w$.

定义 4.0.1: 向量的正交

若 $v^\top w = 0$，则称 v 和 w 正交 (orthogonal).

依定义，0 和所有向量正交.

定理 4.0.1: 正交的模

若 v, w 正交，则

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2. \tag{4.1}$$

4.1 正交性

定义 4.1.1: 子空间的正交

定义线性空间 L 的两个子空间 V, W 正交，若 V 中的每一个向量均和 W 中每一个向量正交.

显然，若 L 的子空间 V, W 正交，则

$$\dim(L) \geqslant \dim(V) + \dim(W). \tag{4.2}$$

定义 4.1.2: 正交补

子空间 V 的正交补 (orthogonal complement) $V^\perp$ 由所有同 V 正交的向量组成.

只有 0 同时属于 V 和 $V^\perp$.

定理 4.1.1: 线性代数基本定理 · 二

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

定理 4.1.2: 分解

$\forall x \in \mathbb{R}^n$ 均可以分解成

$$x = x_r + x_n,$$

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

证明： 由

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

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

若存在 $x_r, x_r' \in \mathrm{C}(A^\top)$ 满足 $Ax_r = Ax_r'$，则 $x_r - x_r'$ 同时在 $\mathrm{C}(A^\top)$ 和 $N(A)$ 中，故 $x_r - x_r' = 0$. □

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

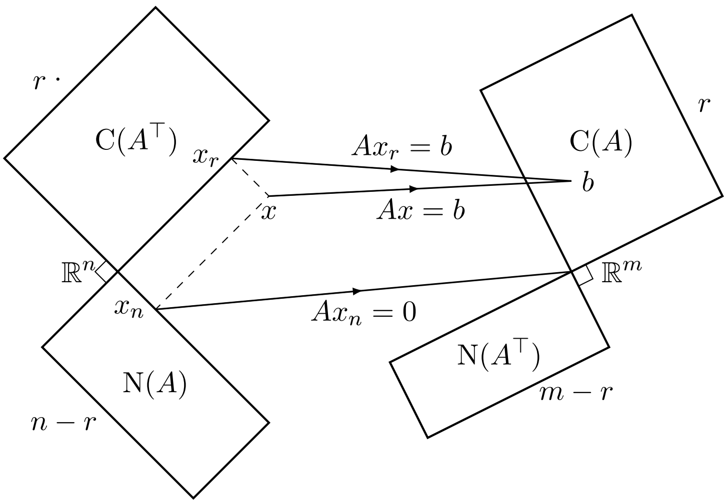

图 2　big picture 升级版

4.2　投影

考虑向量 b 在向量 a 上的投影 p

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

考虑 $\mathbb{R}^m$ 中线性无关的 n 个向量 $(a_1, \dots, a_n) =: A$ 张成的子空间，找到向量 p 在上面的投影

$$p = Ax = x_1 a_1 + \cdots + x_n a_n \in \mathrm{C}(A).$$

考虑投影的性质：p 的终点在子空间中距离 b 的端点最近，则 $(b-p)$ 同子空间垂直：

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

相当于求解线性方程组

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

若 $A^\top A$ 可逆，则 $x=(A^\top A)^{-1}A^\top b$，由 $p=Ax$ 可定义投影矩阵 (project matrix)

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

投影矩阵有性质：$P^2=P$，这也是符合投影性质的.

定理 4.2.1: $A^\top A$ 的可逆性

$A^\top A$ 可逆 $\Leftrightarrow$ A 的列之间线性无关.

证明： 只需证明 $\mathrm{N}(A^\top A)=\mathrm{N}(A)$ 即可.

$\forall x\in\mathrm{N}(A)$，$Ax=0$，左乘 $A^\top$ 得：$A^\top Ax=0$，$x\in\mathrm{N}(A^\top A)$；

$\forall x\in\mathrm{N}(A^\top A)$，$A^\top Ax=0$，左乘 $x^\top$ 得：$x^\top A^\top Ax=\|Ax\|^2=0$，即 $Ax=0$，$x\in\mathrm{N}(A)$. □

推论：

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

4.3 最小二乘法

考虑线性方程组 $Ax=b$，A 的行列 $m>n$ 甚至 $m\gg n$，一般来说无解. 但仍可找到 x 使得 $\|b-Ax\|$ 最短.

由投影的性质，若 Ax 是 b 在 $\mathrm{C}(A)$ 上的投影，则 $\|b-Ax\|$ 最短，此时

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

方法 4.1: 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

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

$\operatorname{rank}(A) = 2$，故 $A^\top A$ 可逆，

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

在此处 (x_i) 特指对 x_i 求和，得到

$$\begin{cases} a = \dfrac{(x_i^2)(y_i) - (x_i)(x_i y_i)}{m(x_i^2) - (x_i)^2} \\ b = \dfrac{-(x_i)(y_i) - m(x_i y_i)}{m(x_i^2) - (x_i)^2} \end{cases}$$

方法 4.2: 多项式拟合

m 组数据 (x_i, y_i)，确定多项式关系

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

中的系数 $a_0, \ldots, a_n$，即

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

从而 $x = (A^\top A)^{-1} A^\top b$.

一般最小二乘拟合：……

两组数据相关不一定代表有因果.

4.4　正交基、Gram-Schmidt 法则、QR 分解

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

基 $\{q_1, \ldots, q_n\}$ 是正交归一的 (orthonomal) 的，若

$$q_i \cdot q_j = \delta_{ij}$$

将一组正交归一基 $\{q_1, \ldots, q_n\}$ 按列排成矩阵 Q，则 $Q^\top Q = I$，称为正交矩阵 (orthogonal matrix)．特别的，若 Q 是方阵，则

$$Q^\top = Q^{-1},$$

则 $Q^\top Q = QQ^\top = I$，便得到正交归一基的完备性 (complete)

$$\sum_{i=1}^{n} q_i q_i^\top = I.$$

例 4.4.1: Fourier 级数

定义函数 f, g 在 $[-\pi, \pi]$ 上的内积

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

则三角函数系列

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

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

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

Fourier 系数

$$\begin{cases} a_n := \dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\cos nx\,\mathrm{d}x \\ b_n := \dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\sin nx\,\mathrm{d}x \end{cases}$$

给定一组基 $\{a_1, \ldots, a_n\}$，如何构造一组正交归一基 $\{q_1, \ldots, q_n\}$?

方法 4.3: Gram-Schmidt 法则

1. 选取 $b_1 := a_1$;
2. 从 a_2 减去沿着 b_1 方向的分量，作为 b_2
$$b_2 := a_2 - \frac{b_1^\top a_2}{b_1^\top b_1} b_1;$$
3. 从 a_i 减去沿着 $b_1, \dots, b_{i-1}$ 方向的分量，作为 b_i
$$b_i := a_i - \sum_{j=1}^{i-1} \frac{b_j^\top a_i}{b_j^\top b_j} b_j.$$

再归一化 $\{b_1, \dots, b_n\}$
$$q_i := \frac{b_i}{\|b_i\|}.$$

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

$$R := Q^\top A = \begin{bmatrix} q_1^\top a_1 & q_1^\top a_2 & \cdots & q_1^\top a_n \\ 0 & q_2^\top a_2 & \cdots & q_2^\top a_n \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & q_n^\top a_n \end{bmatrix}$$

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：

$$A = QR. \tag{4.4}$$

在最小二乘法等应用中，$A^\top A = R^\top R$

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

效率更高.

5 行列式

在例 1.3.1 中，2 阶方阵的逆为

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

定义 2 阶方阵的行列式 (determinant)

$$\det(A) \equiv |A| \equiv \begin{vmatrix} a & b \\ c & d \end{vmatrix} := ad - bc. \tag{5.1}$$

特别的，给出 3 阶方阵的行列式

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - (a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}). \end{aligned} \tag{5.2}$$

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

对于 n 阶方阵 A，记 $A_{\neq ij}$ 为去掉第 i 行第 j 列得到的 $(n-1)$ 阶方阵．则 A 的余子式 (minor) 定义为

$$M_{ij} := \det(A_{\neq ij}); \tag{5.3}$$

代数余子式 (cofactor) 定义为

$$C_{ij} := (-)^{i+j} M_{ij}. \tag{5.4}$$

代数余子式前的正负号与矩阵中元素所在位置的关系：

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{5.5}$$

定义 5.1.2: 行列式

A 的行列式定义为：

1. $n=1$ 时，$\det(A)=a_{11}$;
2. $n>1$ 时，行展开或列展开 (Laplace 展开)

$$\det(A)=\sum_{i=1}^{n} a_{ij}C_{ij}=\sum_{\ell=1}^{n} a_{k\ell}C_{k\ell}. \tag{5.6}$$

易得，单位矩阵行列式 $\det I=1$，对角矩阵行列式

$$\det(\operatorname{diag}(a_1,\dots,a_n))=a_1\cdots a_n. \tag{5.7}$$

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.

按行列式的定义进行计算即证.

5.2　行列式的性质

行列式可看做 n 个 n 维向量到数域 $\mathbb{F}$ 的映射：

$$\det(A)=T(a_1,\dots,a_n),$$

由定义，行列式是线性的：

1. $T(a_1,\dots,ka_i,\dots,a_n)=kT(a_1,\dots,a_i,\dots,a_n)$;
2. $T(\dots,a_i+b_i,\dots)=T(\dots,a_i,\dots)+T(\dots,b_i,\dots)$.

除了线性，行列式还满足初等变换相关的性质：

3. 交换 A 任意两行或两列得到 B，则 $\det(B)=-\det(A)$;
   其推论是，若 A 中有任意两行或两列相同，则 $\det(A)=0$.
4. 将 A 的第 i 行乘一个常数加到第 j 行得到 B，则 $\det(A)=\det(B)$.
   其推论是，若 A 的行/列之间线性相关，或者说秩小于阶，则 $\det(A)=0$.

因此 A 可逆 $\Leftrightarrow$ $\det(A)\neq 0$.

定义 5.2.1: 全反对称张量

定义全反对称张量 (Levi-Civita symbol) $\epsilon_{i_1\cdots i_n}$，其中 $i_1,\dots,i_n$ 取值范围 $1,\dots,n$，满足

- $\epsilon_{1\cdots n}=1$;
- $\epsilon_{\cdots i_p\cdots i_q\cdots}=-\epsilon_{\cdots i_q\cdots i_p\cdots}$;
- 任两个指标相同则 $\epsilon_{\cdots i_p\cdots i_p\cdots}=0$.

如果 $i_1,\dots,i_n$ 是 $1,\dots,n$ 的一个全排列，其从 $1,\dots,n$ 变换而来需要两两交换的次数为 p，则

$$\epsilon_{i_1\cdots i_n}=(-1)^p.$$

利用全反对称张量，行列式也可以定义为

$$\det(A)=\sum_{i_1,\dots,i_n=1}^{n}\epsilon_{i_1\cdots i_n}a_{1i_i}\cdots a_{ni_n}. \tag{5.8}$$

或者进行一个抽象的定义：行列式 $\det(A)=T(a_1,\dots,a_n)$ 是 n 个 n 维向量到数域 $\mathbb{F}$ 的映射，且满足以下三个性质：

1. $\det I=1$;
2. 任意交换两列，行列式反号；
3. 线性 ...

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

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

定理 5.3.1: 行列式与矩阵乘法

$$\det(AB)=\det(A)\det(B). \tag{5.9}$$

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

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

其一个推论是

$$\det(A^{-1}) = \det(A)^{-1}. \tag{5.10}$$

但是行列式与矩阵加法之间并无必然联系：

$$\det(A + B) \neq \det(A) + \det(B). \tag{5.11}$$

但是当 A, B 满足特殊的条件时，$\det(A+B)$ 可以化简，见后文的定理 5.3.4.

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

$$\begin{vmatrix} A & B \\ 0 & D \end{vmatrix} = |A|\,|D|\,. \tag{5.12}$$

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

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \begin{bmatrix} L_A & \\ & L_D \end{bmatrix} \begin{bmatrix} U_A & L_A^{-1} B \\ & U_D \end{bmatrix}$$

前者为下三角矩阵，后者为上三角矩阵，故

$$\begin{vmatrix} A & B \\ 0 & D \end{vmatrix} = |L_A|\,|L_D|\,|U_A|\,|U_D| = |A|\,|D|\,.$$ □

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，B 是 $m \times n$ 矩阵，C 是 $n \times m$ 矩阵，则

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A|\,\left|D - CA^{-1}B\right| = |D|\,\left|A - BD^{-1}C\right| \tag{5.13}$$

证明： 注意到

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & -CA^{-1}B + D \end{bmatrix};$$

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A - BD^{-1}D & 0 \\ C & D \end{bmatrix}.$$ □

定理 5.3.4: 矩阵行列式引理

若 A 可逆且 u, v 均为 n 维列向量，则

$$\det(A + uv^\top) = (1 + v^\top A^{-1}u)\det(A). \tag{5.14}$$

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

$$\det(I + uv^\top) = 1 + v^\top u \tag{5.15}$$

进而

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

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

$$Ax = b,$$

Ae_i 是 A 的第 i 列，定义矩阵 B_i 为把 A 中第 i 列换为 b 的矩阵

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

左右取行列式，则

$$x_i = \frac{\det(B_i)}{\det(A)}. \tag{5.16}$$

Cramer 法则的应用之一是矩阵求逆，利用 $AA^{-1} = I$，把 A^{-1} 的元素看做未知数，解线性方程组

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

定义 5.4.1: 伴随矩阵

n 阶方阵 A 的伴随矩阵 (adjoint matrix) $\operatorname{adj}(A)$

$$(\operatorname{adj}(A))_{ij} := C_{ji}. \tag{5.17}$$

根据 Cramer 法则，

$$A^{-1} = \frac{\operatorname{adj}(A)}{\det(A)}. \tag{5.18}$$

特别注意，即使 A 不可逆，$\operatorname{adj} A$ 依然存在.

定理 5.4.2: 伴随矩阵的性质

伴随矩阵与原矩阵可交换，其乘积为

$$A \operatorname{adj}(A) = \operatorname{adj}(A) A = \det(A) I. \tag{5.19}$$

证明： 根据 Laplace 展开

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

进而有

$$\det(A) \det\bigl(\operatorname{adj}(A)\bigr) = \det\bigl(\det(A) I\bigr) = \bigl(\det(A)\bigr)^n.$$

故伴随矩阵的行列式

$$\det\bigl(\operatorname{adj}(A)\bigr) = \det(A)^{n-1}. \tag{5.20}$$

上面的性质并不要求 A 可逆. 特别的，若 A 可逆，则伴随矩阵也可逆

$$\operatorname{adj}(A)^{-1} = \bigl(\det(A) A^{-1}\bigr)^{-1} = \frac{A}{\det(A)}, \tag{5.21}$$

伴随矩阵的伴随矩阵

$$\operatorname{adj}\bigl(\operatorname{adj}(A)\bigr) = \bigl(\operatorname{adj}(A)\bigr)^{-1} \det\bigl(\operatorname{adj}(A)\bigr) = \bigl(\det(A)\bigr)^{n-2} A. \tag{5.22}$$

最后评价一句 Cramer 法则：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[II]，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．

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

6 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 $\Leftrightarrow$ 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 $\mathrm{N}(A - \lambda I)$.

定理 6.1.2: 特征值

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

$$x_m = c_1 x_1 + \cdots + c_{m-1} x_{m-1}, \tag{$*$}$$

两边同时左乘 A 得

$$\lambda_m x_m = c_1 \lambda_1 x_1 + \cdots + c_{m-1} \lambda_{m-1} x_{m-1}, \tag{$**$}$$

$\lambda_m(*) - (**)$ 得，

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

由 $x_1, \ldots, x_{m-1}$ 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)

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

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

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

由 Vieta 定理，在考虑重根的情况下，

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

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

$$a_2 = \frac{1}{2}\big(\operatorname{tr}^2(A) - \operatorname{tr}(A^2)\big).$$

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： 对下式两边取行列式即证

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

直接推论是：相似变换与原矩阵的迹相同

$$\operatorname{tr}(B^{-1}AB) = \operatorname{tr}(A). \tag{6.3}$$

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 $\Leftrightarrow$ 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

$$X = (x_1, \dots, x_n), \quad \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n). \tag{6.4}$$

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. □

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 $\dim\big(\mathrm{N}(\lambda I - A)\big)$.
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

$\mathrm{GM} \leqslant \mathrm{AM}$.

证明： 考虑 n 阶矩阵 A，假设特征值 λ_1 的 $\mathrm{GM} = \dim(\lambda_1 I - A) = m$，
取 $\{x_1, \dots, x_m\}$ 为 $\mathrm{C}(\lambda_1 I - A)$ 的一组正交归一基.
取 $\{b_1, \dots, b_{n-m}\}$ 为 $\mathrm{N}(\lambda_1 I - A)^{\perp}$ 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 $\mathrm{AM}\geqslant\mathrm{GM}$. □

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 $\mathrm{AM}=\mathrm{GM}$.

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 $\{\lambda_1,\dots,\lambda_s\}$，$\lambda_i$ 对应特征子空间为 V_i，几何重数 $m_i=\dim(V_i)$，记 $n_i:=m_1+\cdots+m_{i-1}$，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 $B_1, \ldots, B_s$ 其中至少一个的特征值，不妨考虑 ξ_j 是 B_i 的特征值[III]，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？

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．

证明： 其证明是线性代数的核心．其中一些概念需要等到第 8 章线性映射才会提及．先给出几个概念证明引理．

[III] 无需考虑 $\mathrm{N}(B_i)$

定义 6.4.1: 广义特征向量

线性映射 $T: V \to V$ 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 $I: V \to V$ 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. □

定理 6.4.3

令 $u_i := (T-\lambda I)^i v$，则 $B = \{u_0, \ldots, u_{d-1}\}$ 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 $(T-\lambda I)^{d-2}, (T-\lambda I)^{d-3}, \ldots$ 可得到所有系数为 0，从而 $u_0, \ldots, u_{d-1}$ 线性无关. □

定理 6.4.4

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

证明： $1 \leqslant j < d-1$ 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

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

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

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

因为 X 是 T 的不变子空间，我们可以把 T 看成是 $X \to X$ 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 $v_1, \ldots, v_r$ 是 T 的广义特征向量，且相应的幂指数是 d_i，设

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

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 $v_1, \ldots, v_r$ 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 $K_i = \ker(T^i)$, $U_i = \mathrm{Im}(T^i)$，则

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

证明：

6.5　对称矩阵

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 $\lambda = \bar{\lambda}$. □

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：

$$(Sw)^\top v = w^\top S^\top v = w^\top Sv = \lambda w^\top v = 0,$$ □

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：

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

则 $W^\perp$ 在 S 的作用下稳定：

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

证明： $\forall w \in W, u \in W^\perp$

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

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 $\mathrm{C}(q_1)^\perp$ 上也是稳定的，假设 $\mathrm{C}(q_1)^\perp$ 上有一组正交归一基为 $\{a_1, \ldots, a_{n-1}\}$，构造矩阵 $X_1 = [q_1, a_1, \ldots, a_{n-1}]$，且 X 是正交的 $X_1^\top X_1 = I$，

$$X_1^\top S X_1 = X_1^\top [\lambda q_1, Sa_1, \ldots, Sa_{n-1}] = \begin{bmatrix} \lambda_1 & \\ & S_1 \end{bmatrix}.$$

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

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

$Q_2 := X_1 \operatorname{diag}(1, X_2)$ 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 $Q = [q_1, \ldots, q_n]$；
- 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 $x^\top Sx$ 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 $x^\top Sx > 0$，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： $1 \Rightarrow 2$：S 对称，则 $\Lambda = Q^\top SQ$

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

$2 \Rightarrow 1$：S 的所有 n 个特征值都是正的，故

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

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

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

$1 \Rightarrow 5$：S 正定，故

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

$3 \Rightarrow 4$：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.

$4 \Rightarrow 3$：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.

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

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

$5 \Rightarrow 3$：A 列之间线性无关，故 $A = QR$

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

□

定义 6.6.3: 半正定矩阵

如果 $\forall x \neq 0$，二次型 $x^\top Sx \geqslant 0$，则称 S 是半正定的 (positive semi-definite).

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.

7 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 $\forall x$

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

$AA^\top$ 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 $\lambda_i \geqslant 0$，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：

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

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 $\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0$ 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 $x = c_1v_1 + \cdots + c_nv_n$，从而 $\forall y \in \text{C}(A)$，$y$ 可以被写为 $\{Av_1, \dots, Av_r\}$ 的线性组合：

$$y = Ax = c_1Av_1 + \cdots + c_rAv_r + 0 + \cdots + 0,$$

故 $\{Av_1, \dots, Av_r\}$ 是 $\text{C}(A)$ 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

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

$m \times m$ 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.

$A^\top A$ 是对称矩阵，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 $\forall i \neq j$，

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

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 $\mathrm{C}(A)$ 的一组正交归一基.

再设 $\{u_{r+1}, \dots, u_m\}$ 是 $\mathrm{N}(A^\top)$ 中的一组正交归一基，因为 $\mathrm{N}(A^\top) = \mathrm{C}(A)^\perp$，则 $\{u_1, \dots, u_m\}$ 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

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

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 $x_i^\top A^\top A x_i = \lambda_i x_i^\top x_i > 0$，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.

例 7.1.1: 四个子空间的正交归一基

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

例 7.1.2: 数据压缩

假设 $\mathrm{rank}(A) < \min(m, n)$，则

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

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

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

分量误差的绝对值

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

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)

$$\|x\| := \sqrt{x^\top x}.$$

定理 7.2.1

$$\|Ax\| \leqslant \sigma_1 \|x\| . \tag{7.1}$$

证明：

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

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

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

由矩阵模的定义可直接导出 $\forall x \neq 0$，

$$\|Ax\| \leqslant \|A\| \|x\| .$$

定理 7.2.2: 三角不等式

$$\|A + B\| \leqslant \|A\| + \|B\| . \tag{7.3}$$

证明：

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

故

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

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

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

证明: 只需证 $\forall B$ 秩为 k，都有

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

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

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

□

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

$$A^+ := V\Sigma^+ U^\top. \tag{7.5}$$

其中 Σ^+ 是一个 $n \times m$ 的矩阵

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

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 $\mathrm{C}(A^\top)$ 的矩阵；

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

是投影到 $\mathrm{C}(A)$ 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：

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

7.4 主成分分析

一组数据 $\mu = (\mu_1, \ldots, \mu_n)$ 来源于 n 个样本，其平均值 (mean) 和方差 (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)

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

对角线上 S_{ii} 是样本方差；S_{ij} 是样本协方差．

方法 7.1: 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.

这种变换总方差是不变的：

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

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

8 线性映射

定义 8.0.1: 映射

S, S' 是两个集合，如果 $\forall x \in S$，均有一个 $f(x) = x' \in S'$ 与之对应，这种对应关系 $f : S \to S'$ 便叫映射 (mapping)．
S 为定义域 (domain of definition)，S' 为陪域 (codomain)．$f(x)$ 叫做 x 在映射 f 下的像 (image)，$f(S)$ 为值域 (domain of function)．

定义 8.0.2: 映射的复合

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

$$g \circ f(x) = g\big(f(x)\big).$$

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: 映射有关的概念

映射 $f : S \to S'$，y 的原像 (preimage)

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

单射 (injection)：$\forall x, y \in S$，$x \neq y$ 均有 $f(x) \neq f(y)$．
满射 (surjection)：$f(S) = S'$．
双射 (bijection)：既是单射又是满射
恒等映射 (identity map)：$\mathrm{id} : S \to S$，$\forall x \in S,\ \mathrm{id}(x) = x$．
逆映射 (inverse map)：若存在 $g : S' \to S$ 使得

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

则称映射 $f : S \to S'$ 可逆，$g = f^{-1}$ 为 f 的逆．

定理 8.0.1: 可逆映射

$$\text{映射 } f : S \to S' \text{ 可逆} \quad \Leftrightarrow \quad f \text{ 是双射}.$$

证明： 若 f 可逆，$g: S' \to S$ 为 f 的逆．若 $x, y \in S$，满足 $f(x) = f(y)$，则

$$x = g\big(f(x)\big) = g\big(f(y)\big) = y.$$

故 f 为单射；又 $\forall z \in S'$，取 $x = g(z) \in S$，可得 $f(x) = f\big(g(z)\big) = z$，故 f 为满射．

若 f 为双射，因为 f 为满射，$\forall z \in S'$，有 $x \in S$ 使得 $f(x) = z$，又因 f 为单射，因此 x 是唯一的，我们可以定义 $g(z) = x$，故 g 是 f 的逆映射． □

8.1　线性映射和矩阵

定义 8.1.1: 线性映射

V, W 是两个线性空间，映射 $T: V \to W$ 是线性映射 (linear mapping) 若 T 满足：

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

线性映射也被称为线性变换；特别的，$V \to \mathbb{R}$ 的称为线性函数．

推论：$T(0) = 0$.

定理 8.1.1: 线性映射与基

V, W 是线性空间，$\{v_1, \dots, v_n\}$ 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 $T: V \to W$ 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 $v = c_1v_1 + \cdots + c_nv_n$，定义映射 $T: V \to W$

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

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

$$\begin{aligned} T(v+u) &= T\big((c_1+d_1)v_1 + \cdots + (c_n+d_n)v_n\big) \\ &= (c_1+d_1)w_1 + \cdots + (c_n+d_n)w_n = T(v) + T(u); \end{aligned}$$

$$T(cv) = T(cc_1v_1 + \cdots + cc_nv_n) = cc_1w_1 + \cdots + cc_nw_n = cT(v).$$

(唯一性) 假设存在另一个线性映射 $F : V \to W$ 满足

$$F(v_1) = w_1, \quad \ldots, \quad F(v_n) = w_n,$$

则

$$F(v) = c_1F(v_1) + \cdots + c_nF(v_n) = c_1w_1 + \cdots + c_nw_n = T(v).$$ □

因此只要知道一个线性映射在基上的值，就唯一决定了整个线性映射.

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 $L_A : \mathbb{R}^n \to \mathbb{R}^m$，

$$\forall x \in \mathbb{R}^n, \quad L_A(x) = Ax.$$

定理 8.1.2: 线性映射和矩阵

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

证明: 设 $\{e_1, \ldots, e_n\}$ 是 $\mathbb{R}^n$ 的标准基，$\{f_1, \ldots, f_m\}$ 是 $\mathbb{R}^m$ 的标准基，$\forall x \in \mathbb{R}^n$，有 $x = x_1e_1 + \cdots + x_ne_n$，则

$$L(x) = x_1L(e_1) + \cdots + x_nL(e_n).$$

$L(e_i) \in \mathbb{R}^m$，故可写成基的线性组合

$$L(e_i) = a_{1i}f_1 + \cdots + a_{mi}f_m.$$

故

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

便唯一确定了一个矩阵 A. □

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

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

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

定理 8.2.1: 核和单射

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

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 $u - v = 0$. □

定理 8.2.2: 核的性质

线性映射 $F: V \to W$ 的核 $\ker(F) = \{0\}$，若 $v_1, \dots, v_n \in V$ 线性无关，则 $F(v_1), \dots, F(v_n)$ 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 $x_1 F(v_1) + \cdots + x_n F(v_n) = 0$，则

$$F(x_1 v_1 + \cdots + x_n v_n) = 0, \quad \Rightarrow \quad x_1 v_1 + \cdots + x_n v_n = 0.$$

$v_1, \dots, v_n$ 线性无关，故只有零解. □

定义 8.2.2: 线性映射的像

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

$$\mathrm{Im}(F) := \{F(v) \in W \mid \forall v \in V\}.$$

$\mathrm{Im}(F)$ 是 W 的线性子空间．$\mathrm{Im}(L_A)=\mathrm{C}(A)$．

定理 8.2.3: 核和像的关系

V 是线性空间，$L:V\to W$ 是线性映射

$$\dim(V)=\dim\big(\ker(L)\big)+\dim\big(\mathrm{Im}(L)\big). \tag{8.1}$$

证明：　(矩阵版本)

$$\dim(V)=\dim\big(\mathrm{N}(A)\big)+\dim\big(\mathrm{C}(A^\top)\big)=\dim\big(\mathrm{N}(A)\big)+\dim\big(\mathrm{C}(A)\big).$$

(抽象版本) 略 □

定理 8.2.4: 核、像和双射

线性映射 $L:V\to W$，且 $\dim(V)=\dim(W)$，则

$$\ker(F)=\{0\}\quad\Leftrightarrow\quad\mathrm{Im}(F)=W\quad\Leftrightarrow\quad L\text{ 是双射}.$$

证明：　略 □

8.3　基的变换

设 $B=\{v_1,\dots,v_n\}$ 是线性空间 V 上的一组基，$\forall v\in V$ 均可唯一写成 $v=x_1v_1+\cdots+x_nv_n$

定义 8.3.1: 坐标向量

向量 v 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 $x_B:V\to\mathbb{R}^n$ 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：

$$(u_1,\dots,u_n)=(v_1,\dots,v_n)M, \tag{$*$}$$

v 也可以写成 $v=y_1u_1+\cdots+y_nu_n$，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

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

定理 8.3.1: 换基矩阵

$L: V \to W$ 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 $\forall v \in V$，

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

证明： $\forall v \in V$，有

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

$L(v_i) \in W$，所以

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

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

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

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 $L: V \to W$ 到 $\dim(W) \times \dim(V)$ 矩阵的线性映射，并且是一个双射．

特别的，当 $L \equiv \mathrm{id}: V \to V$ 时，

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

定理 8.3.2: 线性变换的复合与矩阵乘法

线性映射 $L_1: U \to V,\ L_2: V \to W$，$B, B', B''$ 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: $M^B_{B'}(\mathrm{id})$ 可逆

$$M^B_{B'}(\mathrm{id}) = M^{B'}_B(\mathrm{id})^{-1}.$$

定理 8.3.4

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

$$M^{B'}_{C'}(L) = M^C_{C'}(\mathrm{id}) M^B_C(L) M^{B'}_B(\mathrm{id}) = M^{C'}_C(\mathrm{id})^{-1} M^B_C(L) M^{B'}_B(\mathrm{id})$$

证明： 利用 $L = \mathrm{id}_W \circ L \circ \mathrm{id}_V$. □

推论：$L: V \to V$，B, B' 是 V 上的两组基，则

$$M^{B'}_{B'}(L) = M^{B'}_B(\mathrm{id})^{-1} M^B_B(L) M^{B'}_B(\mathrm{id}).$$

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 $L: V \to \mathbb{R}$ 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 $\{v_1, \ldots, v_n\}$ 可构造 V^* 的基 $\{v^{*1}, \ldots, v^{*n}\}$，满足

$$v^{*i}(v_j) = \delta^i{}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{8.2}$$

证明： (完备性) $\forall L \in V^*$，由定理 8.1.1，L 可由 $\{L(v_1), \ldots, L(v_n)\}$ 唯一决定，又 $L(v_1)v^{*1} + \cdots + L(v_n)v^{*n}$ 和 L 在 $v_1, \ldots, v_n$ 上取到了相同的值，故二者相等，即 L 可被写成 $\{v^{*1}, \ldots, v^{*n}\}$ 的线性组合：

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

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

例 8.4.1: Fourier 变换

Fourier 变换

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

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 $F:V^*\to\mathbb{R}$ 构成的线性空间. $\forall v\in V$，可定义映射 $u^{**}\in V^{**}$，使得 $\forall L\in V^*$

$$u^{**}(L)=L(u).$$

因此 V^{**} 和 V 是自然同构的 (natural isomorphism). 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. 由此我们可以将 V^{**} 和 V 视为同一个线性空间. 从而 $V^{**},V^{***},\ldots$ 也就再没有研究价值了.

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

$$v^{*i}(v_j)=\delta^i{}_j\qquad u^{*i}(u_j)=\delta^i{}_j$$

若 $(v_1,\ldots,v_n)=(u_1,\ldots,u_n)A$，$(v^{*1},\ldots,v^{*n})^\top=B(u^{*1},\ldots,u^{*n})^\top$，即

$$v_i=\sum_{j=1}^n u_jA^j{}_i,\qquad v^{*i}=\sum_{j=1}^n B^i{}_ju^{*j}.$$

则 $B=A^{-1}$：

$$\begin{aligned}v^{*i}(v_j)&=\sum_{k=1}^n B^i{}_ku^{*k}\left(\sum_{\ell=1}^n u_\ell A^\ell{}_j\right)=\sum_{k,\ell}B^i{}_kA^\ell{}_ju^{*k}(u_\ell)\\&=\sum_{k,\ell}B^i{}_kA^\ell{}_j\,\delta^k{}_\ell=\sum_{k=1}^n B^i{}_kA^k{}_j=\delta^i{}_j.\end{aligned}$$

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：

$$V+W := \{v+w \,|\, v\in V, w\in W\}.$$

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 $W=\mathrm{span}(w_1,\dots,w_m)$ 即可. □

推论：

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

定义 8.5.3: 线性空间的直积

给定两个线性空间 V, W，其直积 (direct product) $V\times W$ 是所有形如 (v,w) $v\in V, w\in W$ 的元素的集合：

$$V\times W := \{(v,w) \,|\, v\in V, w\in W\}.$$

$V \times W$ 是一个线性空间．且

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

8.6　张量

定义 8.6.1: 多重线性映射

映射 $L: V_1 \times \cdots \times V_r \to W$ 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：

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

定义 8.6.2: 张量空间 $V^* \otimes V^*$

考虑所有多重线性函数 $L: V \times V \to \mathbb{R}$ 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：$O(u, v) \equiv 0$.

因此所有多重线性函数 $L: V \times V \to \mathbb{R}$ 的集合构成一个线性空间，我们把这个空间叫做张量空间 (tensor space) $V^* \otimes V^*$．$V^* \otimes V^*$ 中的每一个元素 L 是二阶协变张量 (covariant tensor)，即 $(0, 2)$ 张量．

若 V 的一组基为 $\{v_1, \ldots, v_n\}$，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

对偶空间 V^* 的基 $\{v^{*1}, \ldots, v^{*n}\}$ 满足 $v^{*i}(v_j) = \delta^i{}_j$．继而定义张量 (tensor) $v^{*i} \otimes v^{*j}$ 满足

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

从而

$$v^{*i}\otimes v^{*j}(v_k,v_\ell)=v^{*i}(v_k)v^{*j}(v_\ell)=\delta^i{}_k\,\delta^j{}_\ell.$$

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 $\{u_1,\ldots,u_n\},\{u^{*1},\ldots,u^{*n}\}$，有变换

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

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

$$w'_{ij}=w(u_i,u_j)=w\left(\sum_{k=1}^n v_kA^k{}_i,\sum_{\ell=1}^n v_\ell A^\ell{}_j\right)=\sum_{k,\ell}w_{k\ell}A^k{}_iA^\ell{}_j.$$

因此这也是协变 (covariant) 的含义：分量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：$(au)\otimes v=u\otimes(av)=a(u\otimes v)$.

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

U,V 是两个线性空间，各自有一组基 $\{u_1,\ldots,u_m\},\{v_1,\ldots,v_n\}$，定义新的基 $\{u_i\otimes v_j\,|\,1\leqslant i\leqslant m,1\leqslant j\leqslant n\}$ 张成的线性空间为 U,V 的

张量积，记为 $U \otimes V$.

由定义

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

例 8.6.2

$\forall u \in U, v \in V$，$u = x_1u_1 + \cdots + x_mu_m$，$v = y_1v_1 + \cdots + y_nv_n$，则

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

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 $V^* \times V^* \to \mathbb{R}$ 的双线性函数构成的线性空间，但我们也可以用张量积定义. $V \otimes V$ 中的元素

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

称作二阶逆变张量 (contravariant tensor)，即 $(2,0)$ 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 $V \otimes V^*$

$V \otimes V^*$ 中的元素是 $(1,1)$ 张量

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

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

$$\left\{v_{i_1}\otimes\cdots\otimes v_{i_k}\otimes v^{*j_1}\otimes\cdots\otimes v^{*j_\ell}\,\middle|\,1\leqslant i_1,\ldots,i_k,j_1,\ldots,j_\ell\leqslant n\right\}$$

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

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

是 (k,ℓ) 阶张量．基变换

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

9 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 $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$：

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

注意：对第一个变量不是简单的线性，而是多一个复共轭：

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

定义了内积的空间叫做内积空间 (inner product space)．

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： $v_1, \ldots, v_n$ 是 V 上的一组基，$g_{ij} := \langle v_i, v_j\rangle$，则 $g_{ji} = \bar{g}_{ij}$．

$\forall u, w \in V$, $u = u^1 v_1 + \cdots + u^n v_n$, $w = w^1 v_1 + \cdots + w^n v_n$

$$\langle u, w\rangle = \left\langle u^1 v_1 + \cdots + u^n v_n, w^1 v_1 + \cdots + w^n v_n \right\rangle = \sum_{i,j} \bar{u}^i w^j g_{ij}. \qquad \square$$

例 9.1.1: 内积与对偶空间

V 的对偶空间 V^* 是所有 $V \to \mathbb{C}$ 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

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n,$$

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

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

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

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

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

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

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

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

这与实际 Hermite 多项式的定义也只是系数的差别.

9.2 Hermite 矩阵和幺正矩阵

所有实矩阵相关的内容可以复制到复矩阵.

定义 9.2.1: Hermite 矩阵

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

Hermite 矩阵其实是对称矩阵在复空间的推广.

例 9.2.1: Pauli 矩阵

给出三个 Pauli 矩阵

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

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

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

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： $Hz=\lambda z$，左乘 $z^\dagger$ 得 $z^\dagger Hz=\lambda z^\dagger z$，由 $z^\dagger Hz, z^\dagger z$ 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： $Hz_1=\lambda_1 z_1$，$Hz_2=\lambda_2 z_2$， $\lambda_1\neq\lambda_2$

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

故 $z_2^\dagger z_1=0$. □

定理 9.2.4: 谱定理

Hermite 矩阵的特征向量构构成 $\mathbb{C}^n$ 中的一组幺正基.

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

证明： 略. □

定义 9.2.2: 幺正矩阵

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

幺正矩阵也是正交矩阵在复空间的推广.

定理 9.2.5: 幺正变换

幺正变换保持复向量的模不变.

证明：

$$\|Uz\|^2 = z^\dagger U^\dagger U z = z^\dagger z = \|z\|^2.$$ □

定理 9.2.6: 幺正矩阵的行列式

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

证明：

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

10 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) 是映射 $\circ : S \times S \to S$. 其中 $S \times S \equiv S^2$ 是笛卡尔积 (Cartesian product)，

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

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 (inversible)，若 $\exists a^{-1} \in S,\ \ a \circ a^{-1} = a^{-1} \circ a = e$.

特别的，简记

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

10.2 群与子群

定义 10.2.1: 群

群 (group) 是有二元运算 $\circ$ 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：$e \circ a = a \circ e = a$;
- 逆：$\forall a \in G,\ \ \exists a^{-1},\ \ a^{-1} \circ a = a \circ a^{-1} = e$.

若还满足交换律，则称为交换群或 Abel 群.

群的阶 (order) $\mathrm{ord}(G)$ 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, f，则

$$e = e \circ f = f;$$

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 $(\mathbb{Z}, +)$；
- 非零实数乘法群 $(\mathbb{R}\backslash\{0\}, \times)$；
- 一般线性 (general linear) 群 $\mathrm{GL}(n)$：所有 n 阶可逆矩阵集合.

定理 10.2.2: 消去律

- $a \circ b = a \circ c, \quad \Rightarrow \quad b = c;$
- $b \circ a = c \circ a, \quad \Rightarrow \quad b = c;$
- $b \circ a = a$ 或 $a \circ b = a, \quad \Rightarrow \quad b = e.$

证明： 左乘/右乘 a^{-1}. □

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幺半群 (monoid).

定义 10.2.2: 对称群

给定有限集合 T，所有 $f : T \to T$ 的双射在映射的复合下构成一个群 $\mathrm{sym}(T)$，称做对称群 (symmetric group).

例 10.2.2: 群的例子 (续)

置换群：T