第一章 向量与复数

1.1 高维数组向量

定义 1.1. 一个 n **维数组向量** $\boldsymbol{a}$ 是一个有序 n 元数组 $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$. 其中, a_i 是 $\boldsymbol{a}$ 的第 i 个分量. 这是向量写成行的形式, 称为一个**行向量**. 我们也可以写成列的形式, 得到一个列向量 $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ 或 $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$.

关于数组向量 (以下简称为**向量**), 我们有如下的**性质与运算**:

(1) 两个 n 维**向量相等**当且仅当对应的分量相等.

(2) **零向量**: 所有分量为 0 的向量, 有时会用 $\mathbf{0}$ 来表示.

(3) 两个**向量的加法**为对应分量相加, 即,

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) \coloneqq (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

这是一个数组向量. 而这儿的符号 $\coloneqq$ 用语言表述就是 "定义为". 在二维平面或者三位立体空间中, 我们可以用大家熟知的平行四边形法则或者等价的三角形法则来理解向量的加法.

(4) 对一个**向量的数乘** (**向量的标量乘法**) 为对其对应分量的乘法, 即, 若 λ 是一个数, 则

$$\lambda(a_1, a_2, \ldots, a_n) \coloneqq (\lambda a_1, \lambda a_2, \ldots, \lambda a_n).$$

这是一个数组向量. 对于 $\boldsymbol{a} = (a_1, \ldots, a_n)$, 我们会将 $(-1)\boldsymbol{a} = (-a_1, \ldots, -a_n)$ 称为 $\boldsymbol{a}$ 的**负向量**, 并且简记为 $-\boldsymbol{a}$.

(5) 提醒一下, 如果 $\boldsymbol{a}, \boldsymbol{b}$ 和 $\boldsymbol{c}$ 都是同维数组向量, 而 λ, μ 都是标量的话, 我们很容易验证如下的简单性质:

- $\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$,
- $\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})=(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}$,
- $\boldsymbol{a}+\boldsymbol{0}=\boldsymbol{a}$,
- $\boldsymbol{a}+(-\boldsymbol{a})=\boldsymbol{0}$,
- $1\boldsymbol{a}=\boldsymbol{a}$,
- $\lambda(\mu\boldsymbol{a})=(\lambda\mu)\boldsymbol{a}$,
- $(\lambda+\mu)\boldsymbol{a}=\lambda\boldsymbol{a}+\mu\boldsymbol{a}$,
- $\lambda(\boldsymbol{a}+\boldsymbol{b})=\lambda\boldsymbol{a}+\lambda\boldsymbol{b}$.

(6) 两个**向量的点乘(内积)**:

$$(a_1,a_2,\ldots,a_n)\cdot(b_1,b_2,\ldots,b_n)\coloneqq a_1b_1+a_2b_2+\cdots+a_nb_n.$$

这仅仅是一个数 (标量).

(7) **基本向量** $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$, 其中 $\boldsymbol{e}_i$ 是 n 维向量 $\boldsymbol{e}_i=(0,\ldots,0,1,0,\ldots,0)$, 该向量仅仅第 i 个分量为 1, 其它分量都是 0.

(8) 形如 $\lambda_1\boldsymbol{a}_1+\cdots+\lambda_m\boldsymbol{a}_m$ 的数组向量称为 n 维数组向量向量 $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m$ 的一个**线性组合**, 其中 $\lambda_1,\ldots,\lambda_m$ 为数.

例 1.2. $c\begin{pmatrix}1\\1\end{pmatrix}+d\begin{pmatrix}2\\3\end{pmatrix}=\begin{pmatrix}c+2d\\c+3d\end{pmatrix}$ 是 2 维数组向量 $\begin{pmatrix}1\\1\end{pmatrix}$ 与 $\begin{pmatrix}2\\3\end{pmatrix}$ 的一个线性组合.

例 1.3. 任何一个 n 维数组向量 $\boldsymbol{a}=(a_1,\ldots,a_n)$ 都可以写成

$$\boldsymbol{a}=a_1\boldsymbol{e}_1+a_2\boldsymbol{e}_2+\cdots+a_n\boldsymbol{e}_n,$$

其中 $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$ 为上面提到的基本向量. 即, 任何 n 维数组向量都是基本向量 $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$ 的线性组合. 另外, 不难看出这样的线性表示是唯一的, 即, 若

$$\boldsymbol{a}=a'_1\boldsymbol{e}_1+a'_2\boldsymbol{e}_2+\cdots+a'_n\boldsymbol{e}_n,$$

则对于每个 i 皆有 $a_i=a'_i$.

例 1.4 (向量内积的几何意义)**.** 对于的 n 维数组向量 $\boldsymbol{a}=(a_1,\ldots,a_n)$ 与 $\boldsymbol{b}=(b_1,\ldots,b_n)$, 我们首先注意到 $\boldsymbol{a}\cdot\boldsymbol{a}=a_1^2+a_2^2+\cdots+a_n^2\geq 0$, 于是可以定义 $|\boldsymbol{a}|:==\sqrt{\boldsymbol{a}\cdot\boldsymbol{a}}\geq 0$, 并称之为 $\boldsymbol{a}$ 的**长度**或者**模长**. 显然, $\boldsymbol{a}=\boldsymbol{0}$ 当且仅当 $|\boldsymbol{a}|=0$. 另外, 利用勾股定理可知, 若 $n=2$ 或者 3, 我们确实看到 $|\boldsymbol{a}|$ 就是直角坐标系中向量 $\boldsymbol{a}$ 我们所熟知的长度.

接下来, 我们假定 $\boldsymbol{a}\neq\boldsymbol{0}\neq\boldsymbol{b}$, 即 $|\boldsymbol{a}|\neq 0\neq|\boldsymbol{b}|$. 由著名的 *Cauchy* 不等式 (证明留作作业) 可知,

$$(a_1b_1+a_2b_2+\cdots+a_nb_n)^2\leq(a_1^2+a_2^2+\cdots+a_n^2)(b_1^2+b_2^2+\cdots+b_n^2).$$

这说明

$$(\boldsymbol{a}\cdot\boldsymbol{b})^2 \leq |\boldsymbol{a}|^2|\boldsymbol{b}|^2.$$

此时,

$$-1 \leq \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} \leq 1,$$

于是存在唯一的 $\theta \in [0,\pi]$, 使得 $\cos(\theta) = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}$. 利用解析几何的方法, 可以证明 θ 恰好是向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 之间的夹角. 我们这儿在 $n=2$ 的情形下简要地证明这一结果. 不妨设以原点为中心, 从 x 轴正方向开始作逆时针旋转, 需要 α 角到 $\boldsymbol{a}$, 然后再需要 θ 角到 $\boldsymbol{b}$. 若 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的向量长度分别为 A 与 B, 于是 $\boldsymbol{a} = (A\cos(\alpha), A\sin(\alpha))$, 而 $\boldsymbol{b} = (B(\cos(\alpha+\theta), B\sin(\alpha+\theta)))$. 此时, 我们可以直接看到确实有

$$\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} = \frac{A\cos(\alpha)B\cos(\alpha+\theta)+A\sin(\alpha)B\sin(\alpha+\theta))}{AB} = \cos(\theta).$$

$n=3$ 的情形也是类似的, 不过证明稍微复杂一些.

习题 1.5. *阅读教材第一章的内容, 重点学习第 1、2、3、6、7 小节, 第 5 小节可以略去.*

习题 1.6. *证明如下形式的 Cauchy 不等式: 若 $a_1,\ldots,a_n,b_1,\ldots,b_n \in \mathbb{R}$, 则必有*

$$(a_1b_1+a_2b_2+\cdots+a_nb_n)^2 \leq (a_1^2+a_2^2+\cdots+a_n^2)(b_1^2+b_2^2+\cdots+b_n^2).$$

习题 1.7. *设三维空间中有向量 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$, 它们的长度相同, 两两之间的夹角相等. 若 $\boldsymbol{a}=(1,1,0)$, $\boldsymbol{b}=(0,1,1)$, 求出 $\boldsymbol{c}$.*

第三章　线性方程组

线性代数来源于解线性方程组.

什么是线性方程? 形如

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

的方程称为含有 n 个未知元的 **(线性) 方程**, 其中 $a_1, a_2, \ldots, a_n$ 和 b (暂时) 为实数, $x_1, x_2, \ldots, x_n$ 称为**变量**或**未知元**.

含有 m 个方程和 n 个未知元的方程组定义为

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}x_m + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{3.1}$$

其中 a_{ij} 与 b_i 皆为实数. 若 $b_1, b_2, \ldots, b_m$ 不全为 0, 这样的方程组 (3.1) 通常被称为 $m \times n$ **的非齐次线性方程组**; 否则, 称之为$m \times n$ **的齐次线性方程组**.

若方程组 (3.1) 为非齐次的线性方程组, 我们也称齐次方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases} \tag{3.2}$$

为原方程组 (3.1) 的**导出组**.

在中学里我们已经学过简单的线性方程组的求解. 那么, 在大学里, 当我们重新涉及线性方程组的求解时, 我们到底在干什么? 有什么新的认识角度?

先看一个简单的例子.

例 3.1. 考察二元一次方程组

$$\begin{cases} 2x - \ \ y = 0 \\ -x + 2y = 3 \end{cases}$$

请学生先自行解方程. 答案为 $x = 1$, $y = 2$.

(1) **矩阵的看法**. 之后对此类方程的方程的标准处理是将其写成

$$\underbrace{\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} x \\ y \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} 0 \\ 3 \end{pmatrix}}_{\boldsymbol{b}},$$

即

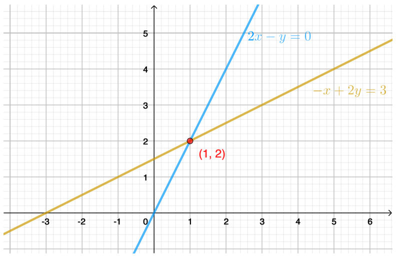
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

其中 $\boldsymbol{A}$ 称为方程组的**系数矩阵**. 矩阵可以视作数组向量的推广, 是以二维点阵的方式来有序地描述一些数 (数学量). 若是类比于求解如下的简单方程:

$$5x = 6 \Rightarrow 5^{-1} \cdot (5x) = 5^{-1} \cdot 6 \Rightarrow (5^{-1} \cdot 5)x = \frac{6}{5} \Rightarrow 1x = \frac{6}{5} \Rightarrow x = \frac{6}{5},$$

我们会希望得到一个代数对象 $\boldsymbol{A}^{-1}$, 使得 $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$. 在之后的学习里, 我们会看到如何定义矩阵的乘法, 以及什么时候矩阵 $\boldsymbol{A}$ 会存在相应的 $\boldsymbol{A}^{-1}$.

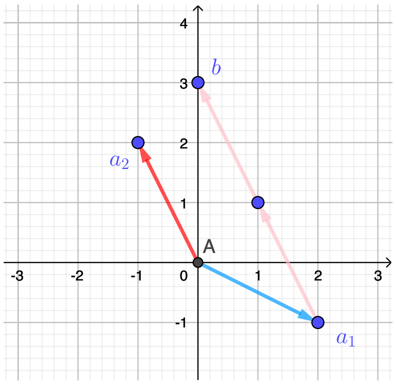
(2) **行的看法**. 线性方程组的每一行都是一个线性方程. 在这个例子里, 它们分别代表着不同的直线. 求解方程组则对应于考察这些直线的交点. 这提供了一个以几何的观点来解线性方程组的视角.



(3) **列的看法**. 原线性方程组可以写成 2 维列向量的线性组合的形式:

$$x\cdot\begin{pmatrix}2\\-1\end{pmatrix}+y\cdot\begin{pmatrix}-1\\2\end{pmatrix}=\begin{pmatrix}0\\3\end{pmatrix}.$$

将上式中出现的 3 个列向量依次记作 $\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{b}$. 则上面的线性组合的写法意味着寻找合适的系数 x 与 y.



从这个观察点出发, 我们以后的讨论会问类似于下面的问题:

(i) 这样的线性组合是否存在 (对应于原线性方程组是否有解)? 这样的线性组合是否唯一 (原线性方程组的解是否唯一)?

(ii) 通过调整线性组合的系数, $\boldsymbol{a}_1$ 与 $\boldsymbol{a}_2$ 的线性组合能否得到平面上的所有向量?

在大学里处理线性方程组问题时一般会借用较为整体更加抽象的方法, 但它不是空中楼阁, 是建立在我们中学已有的知识上的. **第一步**, 我们将方程组视为一个整体, 一个系统, 用 **Gauss 消元法***来求解. 其核心是先消元, 然后利用回代逐步求解.

*类似的消去法大约在公元前 250 年由中国数学家应用. 这种方法在西方直到 19 世纪才由著名德国数学家 C.F. 高斯发现. 德国工程师 W. 若尔当把它写在 1988 年的一本测地学著作中, 这种方法才被普及.

例 3.2. 对前面例 *3.1* 中的线性方程组求解.

$$\xrightarrow[\text{首先消去 } x]{\text{开始消元}} \begin{cases} 2x - \ \ y = 0 \\ -x + 2y = 3 \end{cases} \xrightarrow{(1)\leftrightarrow(2)} \begin{cases} -x + 2y = 3 \\ 2x - \ \ y = 0 \end{cases} \xrightarrow{2(1)\to(2)} \begin{cases} -x + 2y = 3 \\ \qquad 3y = 6 \end{cases}$$

$$\xrightarrow{\frac{1}{3}(2)} \begin{cases} -x + 2y = 3 \\ \qquad\ \ y = 2 \end{cases} \xrightarrow[-(2)\to(1)]{\text{开始回代}} \begin{cases} -x \qquad = -1 \\ \qquad\ \ y = 2 \end{cases} \xrightarrow{-(1)} \begin{cases} x = 1 \\ y = 2 \end{cases}$$

在上面的例子里, 我们本质上进行了如下 3 类基本操作. 为了讨论的方便, 我们**暂时**采用如下更为简洁的记号来表示相应的操作.

(1) 交换两个线性方程, 记作 $(i) \leftrightarrow (j)$, 表示将第 i 个方程与第 j 个方程互换.

(2) 某个方程乘以一个非零的常数, 记作 $\lambda(i)$, 表示将第 i 个方程乘以非零的常数 λ.

(3) 某个方程乘以常数后加到另一个方程, 记作 $\lambda(i) \to (j)$, 表示将第 i 个方程乘以常数 λ 后加到第 j 个方程.

它们被称为解线性方程组的**初等变换**.

注 3.3. 需要指出的是, 这些变换是**可逆的**. 若两个方程被对换, 则再次对换它们就会还原为原来的状态. 若某个方程乘以非零常数 λ, 则将所得的方程乘以 $1/\lambda$ 就得出原来的方程. 最后, 若把第 i 个方程的 λ 倍加到第 j 个方程得到新的第 j 个方程, 那么 "逆" 变换就是把第 i 个方程的 $-\lambda$ 倍加到新的第 j 个方程上就得到原来的第 j 个方程.

解线性方程组时, 该方程组的解的全体称为该方程组的**解集**. 若方程组的解集为空集, 我们称该方程组是**不相容的**, 否则, 称之为**相容的**.

在求解线性方程组时, 方程组的具体形式会改变. 对于每次操作, 我们有理由希望相邻的两个方程组的解具有相同的解集, 即该求解的操作会得到**同解的**方程组 (也称为**等价的**方程组). 不难验证 (请学生自行确认这一点), 上面提到的每个初等变换都能保证得到同解的方程组, 从而最初的方程组与经过一系列初等变换后化简得到的方程组具有相同的解集.

利用上面的记号, 我们再来看一个稍微复杂一点的例子.

例 3.4. 教材例题 3.1.1, P57

$$\begin{cases} 3x_1 + 2x_2 - \ \ x_3 = 6 \\ \ \ x_1 + 3x_2 + 2x_3 = 9 \\ 2x_1 - \ \ x_2 + 3x_3 = 3 \end{cases} \xrightarrow[(1)\leftrightarrow(2)]{\text{消元. 先消去 } x_1} \begin{cases} \ \ x_1 + 3x_2 + 2x_3 = 9 \\ 3x_1 + 2x_2 - \ \ x_3 = 6 \\ 2x_1 - \ \ x_2 + 3x_3 = 3 \end{cases} \xrightarrow[-2(1)\to(3)]{-3(1)\to(2)} \begin{cases} x_1 + 3x_2 + 2x_3 = 9 \\ \quad - 7x_2 - 7x_3 = -21 \\ \quad - 7x_2 - \ \ x_3 = -15 \end{cases}$$

$$\xrightarrow[-(2)\to(3)]{\substack{\text{暂时无法消去 } x_1,\\ \text{但是可以开始考虑消去 } x_2}} \begin{cases} x_1+3x_2+2x_3=9 \\ \quad -7x_2-7x_3=-21 \\ \qquad\qquad 6x_3=6 \end{cases} \xrightarrow[\text{得到严格三角形的形式}]{-\frac{1}{7}(2),\quad \frac{1}{6}(3)} \begin{cases} x_1+3x_2+2x_3=9 \\ \qquad x_2+\ \ x_3=3 \\ \qquad\qquad x_3=1 \end{cases}$$

$$\xrightarrow[\text{回代}]{\substack{-2(3)\to(1)\\ -(3)\to(2)}} \begin{cases} x_1+3x_2 \qquad =7 \\ \qquad x_2 \qquad =2 \\ \qquad\qquad x_3=1 \end{cases} \xrightarrow{-3(2)\to(1)} \begin{cases} x_1=1 \\ x_2=2 \\ x_3=1 \end{cases}$$

可以看出, 在如上的 Guass 消元法里, 我们利用初等变换依次地在部分的方程里部分地消去未知元, 以期得到严格三角形的形式. 然后再回代, 反序地逐次求解各个未知元.

较为准确地说, 对于关于变量 $x_1,\ldots,x_n$ 的 $n\times n$ 的线性方程组而言, 理想状态下, 在求解过程中希望得到的**严格三角形的方程组**满足: 对于 $1\le k\le n$, 第 k 个方程的前 $k-1$ 个变量前的系数均为 0, 且 x_k 前的系数均不为 0. 若 $n\times n$ 的线性方程组通过化简确实能得到严格三角形的形式, 则我们会得到唯一解. 这个解的存在与唯一性都是由回代法保证的. 大致来说, 我们的回代是通过第 n 个方程解出 x_n, 将其代入第 $n-1$ 个方程, 解得 x_{n-1}; 将 x_n 与 x_{n-1} 的值代入第 $n-2$ 个方程解得 x_{n-2}, 并以此类推.

我们有必要练习一下回代的过程.

课堂练习 3.5. 用刚才介绍的记号求解

$$\begin{cases} 2x_1-x_2+3x_3-2x_4=1 \\ \qquad x_2-2x_3+3x_4=2 \\ \qquad\qquad 4x_3+3x_4=3 \\ \qquad\qquad\qquad 4x_4=4 \end{cases}$$

答案*:* $x_1=1$, $x_2=-1$, $x_3=0$, $x_4=1$.

再做一个稍微复杂的练习.

课堂练习 3.6. 用刚才介绍的记号求解

$$\begin{cases} x_1+2x_2+\ \ x_3=3 \\ 3x_1-\ \ x_2-3x_3=-1 \\ 2x_1+3x_2+\ \ x_3=4 \end{cases}$$

答案*:* $x_1=3$, $x_2=-2$, $x_3=4$.

作为抽象化、规范化处理解线性方程组问题的**第二步**, 我们需要将线性方程组改写成矩阵的形式, 即, 将

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}x_m + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

改写成对应的**增广矩阵**

$$\underbrace{\begin{pmatrix} a_{11} & \cdots\cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots\cdots & a_{mn} \end{pmatrix}}_{\boldsymbol{A}} \vdots \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}}_{\boldsymbol{b}}.$$

该矩阵由两个子矩阵构成, 分别为**系数矩阵** $\boldsymbol{A}$ 与**常数项向量** $\boldsymbol{b}$.

我们将用 r_i 来表示增广矩阵的第 i 行 (row). 对应于矩阵求解的三个初等变换, 我们有矩阵的三个**初等行变换**:

(对换行) 互换矩阵的两行, 记作 $r_i \leftrightarrow r_j$, 表示将矩阵的第 i 行和第 j 行互换.

(倍乘行) 某一行乘以一个非零的常数, 记作 λr_i, 表示将矩阵的第 i 行乘以非零的常数 λ.

(倍加行) 将某一行乘以一个常数后加到零一行去, 记作 $\lambda r_i \to r_j$, 表示将矩阵的第 i 行乘以 λ 后加到矩阵的第 j 行去.

例 3.7. *下面将例 3.4 中的求解过程用矩阵的方式表示出来.*

$$\left(\begin{array}{ccc:c} 3 & 2 & -1 & 6 \\ 1 & 3 & 2 & 9 \\ 2 & -1 & 3 & 3 \end{array}\right) \xrightarrow{r_1 \leftrightarrow r_2} \left(\begin{array}{ccc:c} 1 & 3 & 2 & 9 \\ 3 & 2 & -1 & 6 \\ 2 & -1 & 3 & 3 \end{array}\right) \xrightarrow[-2r_1 \to r_3]{-3r_1 \to r_2} \left(\begin{array}{ccc:c} 1 & 3 & 2 & 9 \\ 0 & -7 & -7 & -21 \\ 0 & -7 & -1 & -15 \end{array}\right) \xrightarrow{-r_2 \to r_3}$$

$$\left(\begin{array}{ccc:c} 1 & 3 & 2 & 9 \\ 0 & -7 & -7 & -21 \\ 0 & 0 & 6 & 6 \end{array}\right) \xrightarrow[-\frac{1}{7}r_2]{\frac{1}{6}r_3} \left(\begin{array}{ccc:c} 1 & 3 & 2 & 9 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow[-2r_3 \to r_1]{-r_3 \to r_2} \left(\begin{array}{ccc:c} 1 & 3 & 0 & 7 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right) \xrightarrow{-3r_2 \to r_1}$$

$$\left(\begin{array}{ccc:c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right) \quad \Leftrightarrow \quad \begin{cases} x_1 = 1 \\ x_2 = 2 \\ x_3 = 1 \end{cases}$$

简要小结一下. 对于 $n \times n$ 的线性方程组, 在逐步消元后若得到严格三角形的形式, 则方程组有唯一的解, 这个解通过回代法可以得到. 而得到严格三角形的逐步消元的化简过程大致上是采取了如下的步骤. 第一步, (正常情况下) 增广矩阵的第一列不会为零, 在这一列中选取一个不为零的 (最为简单的) 元素令为**主元** (pivot), 将这一列令为**主行** (pivotal row), 并通过交换行的方法, 把主行移到第一行的位置. 接下来, 进行初等行变换, 用主元将其它行的第一个元素 (即该主元下面的元素) 消去. 第二步, 在第二列中从第二行到第 n 行选取一个不为零的 (最为简单的) 元素令为新的主元, 将其所在行令为新的主行, 通过交换行的方法, 把新的主行移到第二行的位置. 接下来, 进行初等行变换, 用主元将其下面的元素消去. 对后面的列作类似的操作. 理想状态下, 如果在每一步中都能在该列找到相应的主元, 则最后必然得到一个严格 (上) 三角形的形式.

并不是所有的 $n \times n$ 的线性方程组都能通过化简得到严格三角形的形式的.

例 3.8. *解方程组*

$$\begin{cases} x_1 - 2x_2 + 3x_3 - 4x_4 = 4 \\ \qquad x_2 - x_3 + x_4 = -3 \\ x_1 + 3x_2 \qquad - 3x_4 = 1 \\ \quad - 7x_2 + 3x_3 + x_4 = -3 \end{cases}$$

解.

$$\xrightarrow{\text{矩阵形式}} \left(\begin{array}{cccc:c} 1 & -2 & 3 & -4 & 4 \\ 0 & 1 & -1 & 1 & -3 \\ 1 & 3 & 0 & -3 & 1 \\ 0 & -7 & 3 & 1 & -3 \end{array}\right) \xrightarrow[\text{消元}]{\substack{-r_1 \to r_3 \\ 7r_2 \to r_4}} \left(\begin{array}{cccc:c} 1 & -2 & 3 & -4 & 4 \\ 0 & 1 & -1 & 1 & -3 \\ 0 & 5 & -3 & 1 & -3 \\ 0 & 0 & -4 & 8 & -24 \end{array}\right)$$

$$\xrightarrow[-\frac{1}{4}r_4]{-5r_2 \to r_3} \left(\begin{array}{cccc:c} 1 & -2 & 3 & -4 & 4 \\ 0 & 1 & -1 & 1 & -3 \\ 0 & 0 & 2 & -4 & 12 \\ 0 & 0 & 1 & -2 & 6 \end{array}\right) \xrightarrow{r_3 \leftrightarrow r_4} \left(\begin{array}{cccc:c} 1 & -2 & 3 & -4 & 4 \\ 0 & 1 & -1 & 1 & -3 \\ 0 & 0 & 1 & -2 & 6 \\ 0 & 0 & 2 & -4 & 12 \end{array}\right)$$

$$\xrightarrow{-2r_3 \to r_4} \left(\begin{array}{cccc:c} \boxed{1} & -2 & 3 & -4 & 4 \\ 0 & \boxed{1} & -1 & 1 & -3 \\ 0 & 0 & \boxed{1} & -2 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow[\text{回代}]{\substack{r_3 \to r_2 \\ -r_3 \to r_1}} \left(\begin{array}{cccc:c} \boxed{1} & -2 & 0 & 2 & -14 \\ 0 & \boxed{1} & 0 & -1 & 3 \\ 0 & 0 & \boxed{1} & -2 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{2r_2\to r_1}\left(\begin{array}{cccc:c}\boxed{1}&0&0&0&-8\\0&\boxed{1}&0&-1&3\\0&0&\boxed{1}&-2&6\\0&0&0&0&0\end{array}\right)\Leftrightarrow\begin{cases}x_1=-8\\x_2-x_4=3\\x_3-2x_4=6\end{cases}\xrightarrow{\text{令 } x_4 \text{ 为自由变量 } t}$$

$$\begin{cases}x_1=-8\\x_2=3+t\\x_3=6+2t\\x_4=t\end{cases}\quad\text{其中 } t \text{ 任意}\quad\Leftrightarrow\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}-8\\3\\6\\0\end{pmatrix}+t\begin{pmatrix}0\\1\\2\\1\end{pmatrix}.$$

上面的计算结果给出了线性方程组的**通解**, 而 $(x_1,x_2,x_3,x_4)=(-8,3,6,0)$ 是方程组的一个**特解**. □

例 3.9. 解方程组

教材例题 3.2.3

$$\begin{cases}3x_1-2x_2+5x_3+4x_4=2\\6x_1-7x_2+4x_3+3x_4=3\\9x_1-9x_2+9x_3+7x_4=-1\end{cases}$$

解.

$$\xrightarrow{\text{矩阵形式}}\left(\begin{array}{cccc:c}3&-2&5&4&2\\6&-7&4&3&3\\9&-9&9&7&-1\end{array}\right)\xrightarrow[\text{消元}]{\substack{-2r_1\to r_2\\-3r_1\to r_3}}\left(\begin{array}{cccc:c}3&-2&5&4&2\\0&-3&-6&-5&-1\\0&-3&-6&-5&-7\end{array}\right)$$

$$\xrightarrow{-r_2\to r_3}\left(\begin{array}{cccc:c}3&-2&5&4&2\\0&-3&-6&-5&-1\\0&0&0&0&-6\end{array}\right)\Leftrightarrow\begin{cases}3x_1-2x_2+5x_3+4x_4=2\\ \quad -3x_2-6x_3-5x_4=-1\\ \quad 0=-6\end{cases}$$

从最后一个方程出发可以看出, 原线性方程组无 (实数) 解. □

注 3.10. 由前面的例子出发可以看出, 一个线性方程组可能 (1) 有唯一解, (2) 有 (无穷) 多解, 或者 (3) 无解.

下面描述线性方程组的解法.

$$\begin{array}{ccc}\text{线性方程组}&\longleftrightarrow&\text{相应的矩阵表示}\\ \Big\downarrow{\scriptstyle\text{方程组的初等变换}}&&\Big\downarrow{\scriptstyle\text{矩阵的初等行变换}}\\ \text{阶梯形线性方程组}&\longleftrightarrow&\text{矩阵的标准形式 (Row Echelon Form)}\\ \Big\downarrow{\scriptstyle\text{回代}}&&\Big\downarrow{\scriptstyle\text{矩阵的初等行变换}}\\ \text{方程组的通解}&\longleftrightarrow&\text{矩阵的约化标准形}\end{array}$$

下面描述线性方程组的解法.

$$\begin{tikzcd}
\text{线性方程组} \arrow[r, leftrightarrow] \arrow[d, "\text{方程组的初等变换}"] & \text{相应的矩阵表示} \arrow[d, "\text{矩阵的初等行变换}"] \\
\text{阶梯形线性方程组} \arrow[r, leftrightarrow] \arrow[d, "\text{回代}"] & \text{矩阵的标准形式 (Row Echelon Form)} \arrow[d, "\text{矩阵的初等行变换}"] \\
\text{方程组的通解} \arrow[r, leftrightarrow] & \text{矩阵的约化标准形}
\end{tikzcd}$$

接下来介绍上面出现的一些术语. 关于非零矩阵的**标准形式** (**阶梯形矩阵**), 我们要求

(1) 所有**非零行** (该行至少有一个非零元素) 在所有**全零行**的上面 (从而全零行都在矩阵的底部).

(2) 任何一个非零行的**主元** (该行最左边的首个非零元素) 所在的列严格地比上面行 (若存在, 则必然也是非零行) 的主元所在的列更靠右. 有的教材里会进一步要求主元为 1, 这儿我们暂时不作此要求.

(3) 从而, 主元所在列中, 在主元下面的元素皆为零.

注 3.11. (1) *非零矩阵总是可以通过初等行变换化为标准形式. 例如, 我们可以采用如下的算法.*

第一步 *找到最左的非零列, 将其视作***主元列***.*

第二步 *在主元列中选取一个非零的元素作为主元. 如有必要, 通过对换行的方法, 把这个元素移到这一列最顶端的位置.*

第三步 *用 "倍加行" 的方法将主元下面的元素变为 0.*

第四步 *暂时不用管该主元所在的行以及它上面的各行, 对剩下的子矩阵重复使用上面的三个步骤, 直到没有非零行需要处理为止.*

需要提醒的是, 通过一系列初等行变换所能得到的标准形式并不唯一.

(2) *但是, 若进一步要求所有主元为 1, 且在主元所在列, 在主元上面的元素为 0, 则对于给定的矩阵而言, 其通过一系列初等行变换所能得到的符合条件的标准形式是存在且唯一的. 这样的标准形式称为原矩阵的***约化标准形** *(reduced row echelon form). 关于约化标准形的存在性是不难看出的. 其唯一性用本节稍后关于解集的讨论可以证明出来; 大家也可以参考这个超链接所指向的论文, 或者是 [2] 的附录.*

用于得到约化标准形的算法可以描述如下. 在前述的四步的基础上, 我们只需进行如下操作.

第五步 对于每个主元, 若它不是 1, 利用“倍乘行”的方法将它变为 1.

第六步 由最右边的主元开始, 利用“倍加行”的方法把每个主元上方的元素都变为 0.

在上面的第六步里, 从最左边开始当然也是可以的, 但是计算量会稍微大一点.

回到我们处理线性方程组时遇到的增广矩阵, 它的标准形式一般形如

$$\left(\begin{array}{l:c}
0\cdots\cdots 0 \quad \boxed{c_{1j_1}} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots c_{1n} & d_1 \\
0\cdots\cdots\cdots\cdots\cdots 0 \quad \boxed{c_{2j_2}} \cdots\cdots\cdots\cdots c_{2n} & d_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots \\
0\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 \quad \boxed{c_{rj_r}} \cdots\cdots c_{rn} & d_r \\
0\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 & \boxed{d_{r+1}} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots \\
0\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & 0
\end{array}\right)$$

其中 $c_{1j_1}, c_{2j_2}, \ldots, c_{rj_r}$ 皆非零 (一般情形下倾向于选为 1), $1 \le j_1 < j_2 < \cdots < j_r \le n$, 而 d_{r+1} 可能为 0. 并且一般情况下, 我们会有 $j_1 = 1$.

定理 3.12. *线性方程组通过逐步消元, 总可以化为阶梯形方程组. 等价地, 其增广矩阵总可以利用初等变换化为如上的标准形. 此时, 我们有如下的性质.*

(1) *若 $d_{r+1} \neq 0$, 则原方程组无解.*

(此时的等价方程组里会出现形如 $0x_1 + \cdots + 0x_n = d_{r+1} \neq 0$ 的无解方程)

(2) *若 $d_{r+1} = 0$, 且 $r = n$, 则方程组有唯一解.*

(此时的方程组本质上为 $n \times n$ 的严格三角形的方程组; 对方程组进行回代的操作即可看到这一点)

(3) *若 $d_{r+1} = 0$, 而 $r < n$ (即有效方程的个数小于未知元的个数), 则原方程组有多解.*

(此时, 通常会选取 x_i ($i \notin \{j_1, j_2, \ldots, j_r\}$) 这 $n - r$ 个变元为自由元, 将含有这些自由元的项移到方程组的右边, 则得到了关于 $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$ 这些主元为未知元的严格三角形的方程组. 对于每组确定的自由元的取值, 这些主元所对应的未知元都有唯一的解, 从而原方程组有无穷多解. 需要进一步说明两点. 其一, 自由元的选取方式并不是唯一的, 但是自由元的个数总是固定的. 其二, 虽然有可能有其它的自由元的选取方式, 但是其选取方式并不好统一描述, 而这儿的选取方式总是简单易行的.)

定理的证明在课堂上不详细给出, 由学生课后自行阅读. 关于上面的第 (3) 点, 我们举一个简单的例子.

例 3.13. 假定 3×4 的方程组的增广矩阵可以化简为如下的 (约化) 标准形:

$$\left(\begin{array}{cccc:c} \boxed{1} & 2 & 0 & 3 & 4 \\ 0 & 0 & \boxed{1} & 5 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right),$$

则主元在第 1 与第 3 列出现, 而我们将以 x_2 和 x_4 为自由元. 上面的矩阵表示等价于

$$x_1\begin{pmatrix}1\\0\\0\end{pmatrix}+x_2\begin{pmatrix}2\\0\\0\end{pmatrix}+x_3\begin{pmatrix}0\\1\\0\end{pmatrix}+x_4\begin{pmatrix}3\\5\\0\end{pmatrix}=\begin{pmatrix}4\\6\\0\end{pmatrix}.$$

从而我们的线性方程组的通解的列向量表示为

$$x_1\begin{pmatrix}1\\0\\0\end{pmatrix}+x_3\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}4\\6\\0\end{pmatrix}-x_2\begin{pmatrix}2\\0\\0\end{pmatrix}-x_4\begin{pmatrix}3\\5\\0\end{pmatrix},$$

即

$$\begin{pmatrix}x_1\\x_3\end{pmatrix}=\begin{pmatrix}4-2x_2-3x_4\\6-5x_4\end{pmatrix}.$$

(这儿体现了将增广矩阵写成约化标准形的优越性) 我们一般会将通解写成

$$x_1=4-2t_1-3t_2,\quad x_2=t_1,\quad x_3=6-5t_2,\quad x_4=t_2,$$

或等价地,

$$\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}4\\0\\6\\0\end{pmatrix}+t_1\begin{pmatrix}-2\\1\\0\\0\end{pmatrix}+t_2\begin{pmatrix}-3\\0\\-5\\1\end{pmatrix},$$

其中参数 $t_1,t_2\in\mathbb{R}$ 任意.

习题 3.14. 在上面的例题中, 我们还可以选取哪些未知元来组成自由元?

推论 3.15. 齐次线性方程组 (即 $b_1=b_2=\cdots=b_n=0$ 的情形) 总是有零解 (平凡解) 的. 它有非零解的充要条件是 $r<n$. 特别地, 若 $m<n$ (由于 $r\le m$, 故有 $r<n$), 齐次线性方程组必有非零解.

对于增广矩阵里面便于理解的竖线, 一般情况下并不一定要写出.

例 3.16. *是否存在二次函数 $f(x)=ax^2+bx+c$, 其图像经过 $(0,2),(-4,1),(-1,3),(1,2)$ 这四个点?*

解. 图像经过这四个点的充要条件是方程组有解:

$$\begin{cases} 0a+0b+c=2 \\ 16a-4b+c=1 \\ a-\ \ b+c=3 \\ a+\ \ b+c=2 \end{cases} \Leftrightarrow \begin{pmatrix} 0 & 0 & 1 & 2 \\ 16 & -4 & 1 & 1 \\ 1 & -1 & 1 & 3 \\ 1 & 1 & 1 & 2 \end{pmatrix} \xrightarrow{r_1 \leftrightarrow r_4} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 16 & -4 & 1 & 1 \\ 1 & -1 & 1 & 3 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow[-r_1 \to r_3]{-16r_1 \to r_2} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -20 & -15 & -31 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \xrightarrow{r_2 \leftrightarrow r_3} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -2 & 0 & 1 \\ 0 & -20 & -15 & -31 \\ 0 & 0 & 1 & 2 \end{pmatrix} \xrightarrow{-10r_2 \to r_3} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & -15 & -41 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{r_3 \leftrightarrow r_4} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & -15 & -41 \end{pmatrix} \xrightarrow{15r_3 \to r_4} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & -11 \end{pmatrix}.$$

由于相应的阶梯形方程组出现了方程 $0=-11$, 原方程组无解. □

从教学的角度来看, 同学们在初步掌握了之前关于方程组求解的基本理论 (定理 3.12) 后, 需要加强如下类型题目的训练.

例 3.17. *对于参数 a 的取值, 讨论如下方程组的解的情况:*

$$\begin{cases} x_1+\ \ x_2+\ \ x_3=3 \\ 2x_1+\ \ x_2-ax_3=9 \\ x_1-2x_2-3x_3=-6 \end{cases} \Leftrightarrow \begin{pmatrix} 1 & 1 & 1 & 3 \\ 2 & 1 & -a & 9 \\ 1 & -2 & -3 & -6 \end{pmatrix}$$

$$\xrightarrow[-r_1 \to r_3]{-2r_1 \to r_2} \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & -a-2 & 3 \\ 0 & -3 & -4 & -9 \end{pmatrix} \xrightarrow{-3r_2 \to r_3} \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & -a-2 & 3 \\ 0 & 0 & 3a+2 & -18 \end{pmatrix}.$$

故方程组无解的充要条件是 $3a+2=0$, 即 $a=-\frac{2}{3}$.

当 $a \neq -\frac{2}{3}$ 时, 方程组的有效方程个数 = 未知元个数, 从而方程组有唯一解. 接下来, 利用回代, 我们可以解出

$$x_1=-\frac{6}{3a+2}, \quad x_2=\frac{9a+30}{3a+2}, \quad x_3=-\frac{18}{3a+2}.$$

例 3.18. 当参数 λ 为何值时，下列方程组有解？求出其通解.

教材例题 3.3.1

$$\begin{cases} 2x_1 - \ \ x_2 + \ \ x_3 + \ \ \ x_4 = 1 \\ x_1 + 2x_2 - \ \ x_3 + \ 4x_4 = 2 \\ x_1 + 7x_2 - 4x_3 + 11x_4 = \lambda \end{cases} \Leftrightarrow \begin{pmatrix} 2 & -1 & 1 & 1 & 1 \\ 1 & 2 & -1 & 4 & 2 \\ 1 & 7 & -4 & 11 & \lambda \end{pmatrix}$$

$$\xrightarrow{r_1 \leftrightarrow r_2} \begin{pmatrix} 1 & 2 & -1 & 4 & 2 \\ 2 & -1 & 1 & 1 & 1 \\ 1 & 7 & -4 & 11 & \lambda \end{pmatrix} \xrightarrow[-r_1 \to r_3]{-2r_1 \to r_2} \begin{pmatrix} 1 & 2 & -1 & 4 & 2 \\ 0 & -5 & 3 & -7 & -3 \\ 0 & 5 & -3 & 7 & \lambda - 2 \end{pmatrix}$$

$$\xrightarrow{r_2 \to r_3} \begin{pmatrix} 1 & 2 & -1 & 4 & 2 \\ 0 & -5 & 3 & -7 & -3 \\ 0 & 0 & 0 & 0 & \lambda - 5 \end{pmatrix}.$$

由此可以看出，方程组无解的充要条件是 $\lambda - 5 \neq 0$，即 $\lambda \neq 5$.

当 $\lambda = 5$ 时，方程组的有效方程个数 $= 2 <$ 未知元个数 $= 5$，因此方程组有多解. 我们可以以 x_1 和 x_2 为主元，以 x_3, x_4 为自由元. 将上面的矩阵作进一步的简化，可以得到

$$\xrightarrow{-\frac{1}{5}r_2} \begin{pmatrix} 1 & 2 & -1 & 4 & 2 \\ 0 & 1 & -\frac{3}{5} & \frac{7}{5} & \frac{3}{5} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{-2r_2 \to r_1} \begin{pmatrix} 1 & 0 & \frac{1}{5} & \frac{6}{5} & \frac{4}{5} \\ 0 & 1 & -\frac{3}{5} & \frac{7}{5} & \frac{3}{5} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

从而（这一步如果没有反应过来，请回顾一下例 *3.13* 中的解答过程），方程组的通解为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \\ 0 \\ 0 \end{pmatrix} + t_1 \begin{pmatrix} -\frac{1}{5} \\ \frac{3}{5} \\ 1 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} -\frac{6}{5} \\ -\frac{7}{5} \\ 0 \\ 1 \end{pmatrix} \quad \Leftrightarrow \quad \begin{cases} x_1 = \dfrac{4}{5} - \dfrac{1}{5}t_1 - \dfrac{6}{5}t_2 \\ x_2 = \dfrac{3}{5} + \dfrac{3}{5}t_1 - \dfrac{7}{5}t_2 \\ x_3 = t_1 \\ x_4 = t_2 \end{cases}$$

其中参数 $t_1, t_2 \in \mathbb{R}$ 任意.

作业. *P68: #1(1)(5), #2, #3, #4.*

习题 3.19. (1) 给出一个由 3 个方程构成的关于未知元 x_1, x_2, x_3 的齐次线性方程组，使得 $x_1 = 3$, $x_2 = -2$, $x_3 = 1$ 是它的一个解，并且方程组的增广矩阵的约化标准形的主元在前两列.

(2) *是否存在一个由 3 个方程构成的关于未知元 x_1, x_2, x_3 的齐次线性方程组, 使得 $x_1 = 3$, $x_2 = -2$, $x_3 = 1$ 仍然是它的一个解, 但是方程组的增广矩阵的约化标准形主元在第一列和第三列?*

本章的讨论初步解决了线性方程组的求解问题, 其内容与方法本质上还是初等的. 细致看来, 还有些相关的理论问题可以探讨:

(1) 增广矩阵的标准形里的 r 是否唯一?

(2) 这个 r 反映了该线性方程组的什么几何、代数特性?

(3) 方程组在可解时的解集有什么几何性质?

(4) 在特殊情形下, 方程组有怎样的公式解?

(5) 等等.

这些问题的解决, 需要我们在后续章节中对矩阵、行列式、线性空间等代数对象作进一步地讨论.

第四章　矩阵与行列式

4.1　矩阵的定义

在前面一章里, 我们已经见到了矩阵的身影. 从这一章开始, 我们来正式学习与矩阵相关的系列知识.

矩阵记号　简单地说, **矩阵** (matrix) 就是一个矩形的数字阵列. 一个 m 行和 n 列的矩阵称为 $m \times n$ **矩阵**. 矩阵中出现的数称为该矩阵的**元素** (entry) 或**标量** (scalar). 当我们引用矩阵时, 通常用大写字母 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ 等来表示矩阵. 一般地, 习惯上会用 a_{ij} 或者 $a_{i,j}$ 表示矩阵 $\boldsymbol{A}$ 的第 i 行第 j 列的元素, 并称之为 $\boldsymbol{A}$ 的**第 (i, j) 元素**. 因此, 若 $\boldsymbol{A}$ 是一个 $m \times n$ 矩阵, 则一般会写成

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{或} \quad \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

有时候还会将上面的写法简记为 $\boldsymbol{A} = (a_{ij})_{m \times n}$ 或 $\boldsymbol{A} = (a_{ij})$. 类似地, 矩阵 $\boldsymbol{B}$ 可表示为 (b_{ij}), 矩阵 $\boldsymbol{C}$ 可表示为 (c_{ij}) 等等.

在我们的讨论中, 矩阵的元素通常取值于 $F = \mathbb{R}$ 或 $\mathbb{C}$. 此时, 我们也称该矩阵是 F **上的矩阵**. F 上的 $m \times n$ 矩阵的全体记作 $F^{m \times n}$.

F 中的元素也可以自然地被视为一个 1×1 矩阵.

相等　两个**矩阵相等**是指它们的维数以及它们的对应元素相等.

向量　仅有一行的矩阵称为一个**行向量** (row vector), 仅有一列的矩阵称为一个**列向量** (column vector). 一个 n 元数组既可以表示成行向量的的形式, 也可以表示成列向量的

形式. 在解方程时, 一般习惯于将解写成列向量的形式. 在我们的课堂里, 一般习惯上用 (黑斜体) 小写字母 (主要为拉丁字母和希腊字母) 表示向量.

对于 $m \times n$ 矩阵 $\boldsymbol{A} = (a_{ij})$, 它的第 i 个行向量是 $(a_{i1}, a_{i2}, \ldots, a_{in})$ (暂记为 $\boldsymbol{\alpha}_i$), 是 $\boldsymbol{A}$ 的第 i 行元素构成的行向量. $\boldsymbol{A}$ 的第 j 个列向量是

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

(暂记为 $\boldsymbol{\beta}_j$), 是 $\boldsymbol{A}$ 的第 j 列元素构成的的列向量. 此时,

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_n \end{pmatrix} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n),$$

从而 $\boldsymbol{A}$ 即可以表示成其行向量按列依次排列, 也可以表示成其列向量按行依次排列.

本节更多的定义, 需要学生课后自行阅读.

4.2 矩阵的运算

标量乘法 矩阵的**标量乘法** (数乘) 是向量的标量乘法的推广. 设 $\boldsymbol{A} = (a_{ij})$ 是一个矩阵, $\lambda \in F$ 是一个标量 (常数), 则 $\lambda\boldsymbol{A}$ 是将 $\boldsymbol{A}$ 的每个元素都乘以 λ 后得到矩阵: $\lambda\boldsymbol{A} \coloneqq (\lambda a_{ij})$.

例 4.1. 设 $\boldsymbol{A} = \begin{pmatrix} 4 & 8 & 2 \\ 6 & 8 & 10 \end{pmatrix}$, 则 $\frac{1}{2}\boldsymbol{A} = \begin{pmatrix} 2 & 4 & 1 \\ 3 & 4 & 5 \end{pmatrix}$, 而 $3\boldsymbol{A} = \begin{pmatrix} 12 & 24 & 6 \\ 18 & 24 & 30 \end{pmatrix}$.

矩阵加法 矩阵的加法是向量的加法的推广: 两个相同维数的矩阵的加法通过对应元素的加法得到. 具体来说, 若 $\boldsymbol{A} = (a_{ij})$ 和 $\boldsymbol{B} = (b_{ij})$ 都是 $m \times n$ 矩阵, 则 $\boldsymbol{A} + \boldsymbol{B}$ 也是一个 $m \times n$ 矩阵, 其第 (i, j) 元素为 $a_{ij} + b_{ij}$, 即 $\boldsymbol{A} + \boldsymbol{B} \coloneqq (a_{ij} + b_{ij})$.

例 4.2.

$$\begin{pmatrix} 3 & 2 & 1 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 2 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 5 & 4 & 3 \\ 5 & 7 & 9 \end{pmatrix},$$
$$\begin{pmatrix} 2 \\ 1 \\ 8 \end{pmatrix} + \begin{pmatrix} -8 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -6 \\ 4 \\ 10 \end{pmatrix}.$$

相同维数的矩阵的减法 $\boldsymbol{A} - \boldsymbol{B}$ 定义为 $\boldsymbol{A} + (-1)\boldsymbol{B}$, 从而新得到的矩阵的元素为对应元素的减法, 即 $\boldsymbol{A} - \boldsymbol{B} \coloneqq (a_{ij} - b_{ij})$.

例 4.3.

$$\begin{pmatrix} 2 & 4 \\ 3 & 1 \end{pmatrix} - \begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} -2 & -1 \\ 1 & -2 \end{pmatrix}.$$

我们习惯上会将 $m \times n$ 的所有元素为 0 的矩阵记作 $\boldsymbol{O}_{m\times n}$ 或 $\boldsymbol{O}$, 称之为 ($m \times n$ 的) **零矩阵**. 不难验证,

$$\boldsymbol{A} + \boldsymbol{O} = \boldsymbol{O} + \boldsymbol{A} = \boldsymbol{A}.$$

矩阵的乘法及线性方程组 以上定义的矩阵的数乘与加法明显脱胎于数组向量的相应运算. 接下来要介绍的矩阵乘法则略有不同, 某种意义上来说, 矩阵的乘法来源于线性方程组的简化表示. 需要提及的是, 在数学里引入新的概念或者工具时, 我们经常会问: 它与之前已知的数学对象之间有怎样的联系, 以及引入这样的新的数学对象有什么意义, 或者说我们为什么要引入这一的新的对象?

行向量 $\times$ 列向量 设

$$\boldsymbol{\alpha} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

分别是一个 $1 \times n$ 的行向量和一个 $n \times 1$ 的列向量, 即分别为一个 n 维行向量和一个 n 维列向量. 定义矩阵乘法

$$\boldsymbol{\alpha\beta} := a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

这是一个标量, 即 1×1 矩阵. 它可以视作之前提到的数组向量的点乘的一个变形.

例 4.4.

$$\begin{pmatrix} 2 & 1 & -3 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \\ -2 \end{pmatrix} = 2 \cdot 3 + 1 \cdot 2 + (-3) \cdot 1 + 4 \cdot (-2) = -3.$$

例 4.5. 线性方程 $3x_1 + 2x_2 + 5x_3 = 4$ 可以表示成

$$\begin{pmatrix} 3 & 2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 4$$

的形式.

矩阵 $\times$ 列向量 给定一个 $m \times n$ 矩阵 $\boldsymbol{A} = (a_{ij})$ 和一个 n 维列向量 $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$, 并设 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix}$ 是由其行向量 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_m$ 按列有序排列得到的. 注意到对于 “行向量 $\times$ 列向量” 的乘法, 我们已有定义, 因此, 对于矩阵乘积 $\boldsymbol{Ab}$, 我们可以将其定义为

一个 m 维列向量, 由这些 $\boldsymbol{\alpha}_i\boldsymbol{b}$ 按列有序排列得到:

$$\boldsymbol{A}\boldsymbol{b}=\begin{pmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\\ \vdots\\ \boldsymbol{\alpha}_m\end{pmatrix}\boldsymbol{b}:=\begin{pmatrix}\boldsymbol{\alpha}_1\boldsymbol{b}\\ \boldsymbol{\alpha}_2\boldsymbol{b}\\ \vdots\\ \boldsymbol{\alpha}_m\boldsymbol{b}\end{pmatrix}.$$

具体来说, m 维列向量 $\boldsymbol{A}\boldsymbol{b}$ 的第 i 个元素为 $a_{i1}b_1+a_{i2}b_2+\cdots+a_{in}b_n$. 由于一个 n 维行向量可以视作一个 $1\times n$ 的矩阵, 这儿定义的乘法可以视作前述的 “行向量 $\times$ 列向量” 的乘法的推广. 需要注意的是, 这儿的乘法要求矩阵 $\boldsymbol{A}$ 的列数等于列向量 $\boldsymbol{b}$ 的行数.

例 4.6.

$$\begin{pmatrix}-3 & 1\\ 2 & 5\\ 4 & 2\end{pmatrix}\begin{pmatrix}2\\ 4\end{pmatrix}=\begin{pmatrix}-3\cdot 2+1\cdot 4\\ 2\cdot 2+5\cdot 4\\ 4\cdot 2+2\cdot 4\end{pmatrix}=\begin{pmatrix}-2\\ 24\\ 16\end{pmatrix}.$$

例 4.7. 线性方程组

$$\begin{cases}3x_1+2x_2+\ \ x_3=5\\ \ \ x_1-2x_2+5x_3=-2\\ 2x_1+\ \ x_2-3x_3=1\end{cases}$$

可以表示成

$$\begin{pmatrix}3 & 2 & 1\\ 1 & -2 & 5\\ 2 & 1 & -3\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}=\begin{pmatrix}5\\ -2\\ 1\end{pmatrix}$$

的形式.

注 4.8. 在接下来的讲义中, 我们希望能反映科学家和工程师实际应用线性代数的方式, 而这意味着在定义和证明中处理的更多是矩阵的列, 而不是矩阵的元素. 其中的一个核心课题就是用另外一个观点来看待矩阵与向量的乘积 $\boldsymbol{A}\boldsymbol{b}$, 将其作为关于 $\boldsymbol{A}$ 的列的一个线性组合.

若令 $\boldsymbol{A}$ 的列向量依次为 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_n$, 则 $b_1\boldsymbol{\beta}_1+b_2\boldsymbol{\beta}_2+\cdots+b_n\boldsymbol{\beta}_n$ 是 n 个 m 维列向量的线性组合, 从而仍然是一个 m 维列向量. 研究其第 i 个分量, 不难看出相应的元素为 $b_1a_{i1}+b_2a_{i2}+\cdots+b_na_{in}$, 即 m 维列向量 $\boldsymbol{A}\boldsymbol{b}$ 的第 i 个分量. 这说明

$$\boldsymbol{A}\boldsymbol{b}=b_1\boldsymbol{\beta}_1+b_2\boldsymbol{\beta}_2+\cdots+b_n\boldsymbol{\beta}_n.$$

特别地, 这意味着 $\boldsymbol{A}\boldsymbol{b}$ 是 $\boldsymbol{A}$ 的列向量的一个线性组合.

这种现代方法简化了许多论述，且将向量空间思想和线性方程组的研究联系在一起.

例 4.9. 对于 $m \times n$ 的线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \qquad \vdots \\ a_{m1}x_m + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \end{cases}$$

若令

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

则利用上面引入的矩阵的乘法，原线性方程组可以写成 $\boldsymbol{Ax} = \boldsymbol{b}$ 的形式.

若是结合注 *4.8* 中的观察，可以看出，线性方程组 $\boldsymbol{Ax} = \boldsymbol{b}$ 有解的充要条件是常数项列向量 $\boldsymbol{b}$ 是系数矩阵 $\boldsymbol{A}$ 的列向量的一个线性组合.

矩阵 × 矩阵　更一般地，给定两个矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，并假定 $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_p \end{pmatrix}$ 由其列向量 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p$ 按行依次排列得到. 这些列向量的行数即为 $\boldsymbol{B}$ 的行数. 若这些行数与 $\boldsymbol{A}$ 的列数相等，则乘法 $\boldsymbol{A\beta}_j$ 是有定义的，而且得到的是一个列向量. 于是，我们将定义矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的乘法，规定其为一个新的矩阵，记作 $\boldsymbol{AB}$，由这些列向量 $\boldsymbol{A\beta}_1, \boldsymbol{A\beta}_2, \ldots, \boldsymbol{A\beta}_p$ 按行依次排列得到:

$$\boldsymbol{AB} \coloneqq \begin{pmatrix} \boldsymbol{A\beta}_1 & \boldsymbol{A\beta}_2 & \cdots & \boldsymbol{A\beta}_p \end{pmatrix}.$$

因此，若 $\boldsymbol{A} = (a_{ij}) \in F^{m\times n}$, $\boldsymbol{B} = (b_{ij}) \in F^{n\times p}$，则 $\boldsymbol{AB} = (c_{ij}) \in F^{m\times p}$，其中 $\boldsymbol{AB}$ 的第 (i, j) 元素为

$$\begin{aligned} c_{ij} &= \boldsymbol{A\beta}_j \text{ 的第 } i \text{ 个分量} = \boldsymbol{A} \text{ 的第 } i \text{ 个行向量乘以 } \boldsymbol{B} \text{ 的第 } j \text{ 个列向量} \\ &= \sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}. \end{aligned} \tag{4.1}$$

由于一个 p 维列向量可以视作一个 $p \times 1$ 的矩阵，这儿定义的乘法可以视作前述的“矩阵 × 列向量”的乘法的推广.

若结合注 4.8 来看，易知 $\boldsymbol{AB}$ 的每个列向量都是 $\boldsymbol{A}$ 的列向量组的一个线性组合.

习题 4.10. 验证: 矩阵乘法 $\boldsymbol{AB}$ 中的每个行向量都是 $\boldsymbol{B}$ 的行向量组的一个线性组合.

例 4.11.

$$\begin{pmatrix} 3 & -2 \\ 2 & 4 \\ 1 & -3 \end{pmatrix}\begin{pmatrix} -2 & 1 & 3 \\ 4 & 1 & 6 \end{pmatrix} = \begin{pmatrix} 3\cdot(-2)+2\cdot 4 & 3\cdot 1-2\cdot 1 & 3\cdot 3-2\cdot 6 \\ 2\cdot(-2)+4\cdot 4 & 2\cdot 1+4\cdot 1 & 2\cdot 3+4\cdot 6 \\ 1\cdot(-2)-3\cdot(4) & 1\cdot 1-3\cdot 1 & 1\cdot 3-3\cdot 6 \end{pmatrix}$$
$$= \begin{pmatrix} -14 & 1 & -3 \\ 12 & 6 & 30 \\ -14 & -2 & -15 \end{pmatrix}.$$

注 4.12. (1) 矩阵的乘法一定要遵守 "第一个矩阵的列数 $=$ 第二个矩阵的行数" 这一要求.

(2) 矩阵的乘法一般不满足交换性. <u>学生直接验证</u>:

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}.$$

因此, 在我们谈两个矩阵的乘法时需要注意其顺序. 我们称 $\boldsymbol{AB}$ 为 $\boldsymbol{A}$ 左乘 $\boldsymbol{B}$, 也称之为 $\boldsymbol{B}$ 右乘 $\boldsymbol{A}$.

(3) 消去律对矩阵乘法不成立, 即若 $\boldsymbol{AB} = \boldsymbol{AC}$, 即便 $\boldsymbol{A}$ 不是零矩阵, 仍然有可能 $\boldsymbol{B} \neq \boldsymbol{C}$. 例如, 对于 $\boldsymbol{A} = \begin{pmatrix} 2 & -3 \\ -4 & 6 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} 8 & 4 \\ 5 & 5 \end{pmatrix}$, 以及 $\boldsymbol{C} = \begin{pmatrix} 5 & -2 \\ 3 & 1 \end{pmatrix}$, 我们有 $\boldsymbol{AB} = \begin{pmatrix} 1 & -7 \\ -2 & 14 \end{pmatrix} = \boldsymbol{AC}$, 但是 $\boldsymbol{B} \neq \boldsymbol{C}$. 事实上, 为了让消去律成立, 我们一般会需要这儿的 $\boldsymbol{A}$ 在作矩阵乘法时有**逆元**存在. 这一概念将在稍后介绍.

(4) 等价地, 若乘积 $\boldsymbol{AB}$ 为零矩阵, 我们一般也得不到 $\boldsymbol{A}$ 或 $\boldsymbol{B}$ 为零矩阵. 例如, 可直接验证, 在上面一条的例子里, $\boldsymbol{A}(\boldsymbol{B}-\boldsymbol{C}) = \boldsymbol{O}$, 但是 $\boldsymbol{A} \neq \boldsymbol{O} \neq \boldsymbol{B}-\boldsymbol{C}$.

(5) 若矩阵 $\boldsymbol{A}$ 是 $n \times n$ 的, 则称 $\boldsymbol{A}$ 是 n **阶方阵**. 对于 n 阶方阵 $\boldsymbol{A} = (a_{ij})$, 称 $a_{11}, a_{22}, \ldots, a_{nn}$ 为 $\boldsymbol{A}$ 的**对角元**, 称这些元素在 $\boldsymbol{A}$ 的 (主) **对角线**上 (*diagonal*). 若方阵 $\boldsymbol{A}$ 的所有非对角元都是 0, 则称 $\boldsymbol{A}$ 是一个**对角矩阵**; 此时, 可以将该对角矩

阵记作

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{pmatrix} =: \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn}).$$

若 n 阶对角矩阵的所有对角元都是 1, 则称之为 (n 阶) **单位矩阵** (*identity matrix*), 记作 $\boldsymbol{I}_n$ 或 $\boldsymbol{I}$. 不难看出, $\boldsymbol{I} = (\delta_{ij})$, 其中

$$\delta_{ij} = \begin{cases} 1 & \text{如果 } i = j, \\ 0 & \text{如果 } i \neq j \end{cases}$$

是 ***Kronecker* 符号** (*Kronecker delta*). 更一般地, 若 (n 阶) 对角矩阵的所有对角元都是 $\lambda \in F$, 则称之为一个**数量矩阵**, 记作 $\lambda\boldsymbol{I}$; 显然, 它是单位矩阵 $\boldsymbol{I}$ 关于标量 λ 所作的标量乘法.

(6) 若 $\boldsymbol{A} \in F^{m\times n}$, 则可直接验证, $\boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n$. 另外, 在矩阵相乘允许的条件下, 零矩阵与任何矩阵左乘或右乘都得到零矩阵 (维数可能会改变).

(7) 有关矩阵的加法、乘法的运算的基本性质, 由学生自行阅读教材的定理 *4.2.1* 和定理 *4.2.2*. 仅仅强调一下矩阵乘法的结合律: $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$. 从而一长串矩阵的乘法, 在乘法运算允许的条件下, 与计算乘法时 (打括号) 的顺序无关. 刚刚的那个式子可以简记为 $\boldsymbol{ABC}$.

为了验证该结合律, 由乘法的定义, 不妨假定 $\boldsymbol{A} \in F^{m\times n}$, $\boldsymbol{B} \in F^{n\times p}$, $\boldsymbol{C} \in F^{p\times q}$. 为了说明的方便, 在这儿, 对于一个矩阵右下角的双重下标 ij, 我们是在表示该矩阵的第 (i,j) 元素. 反复利用公式 (4.1), 我们有

$$\begin{aligned} ((\boldsymbol{AB})\boldsymbol{C})_{ij} &= \sum_{k=1}^{p} (\boldsymbol{AB})_{ik}\boldsymbol{C}_{kj} = \sum_{k=1}^{p}\left(\sum_{l=1}^{n}\boldsymbol{A}_{il}\boldsymbol{B}_{lk}\right)\boldsymbol{C}_{kj}, \\ (\boldsymbol{A}(\boldsymbol{BC}))_{ij} &= \sum_{l=1}^{n}\boldsymbol{A}_{il}(\boldsymbol{BC})_{lj} = \sum_{l=1}^{n}\boldsymbol{A}_{il}\left(\sum_{k=1}^{p}\boldsymbol{B}_{lk}\boldsymbol{C}_{kj}\right). \end{aligned}$$

利用交换求和顺序的技巧, 可以看出上面的两式相等. 由于 $(\boldsymbol{AB})\boldsymbol{C}, \boldsymbol{A}(\boldsymbol{BC}) \in F^{m\times q}$, 而这两个矩阵的元素对应相等, 故有 $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$.

例 4.13. 下面解释一下为何要引入或者说定义如此 "复杂" 的矩阵乘法运算. 接下来, 我们将所有数组空间的元素都视为列向量. 设矩阵 $\boldsymbol{A} = (a_{ij}) \in F^{m\times n}$, 则我们有一个映

射

$$\mathscr{A}: F^n \to F^m, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} \sum_{k=1}^n a_{1k}x_k \\ \sum_{k=1}^n a_{2k}x_k \\ \vdots \\ \sum_{k=1}^n a_{mk}x_k \end{pmatrix}. \tag{†}$$

这样的映射之后将被称作是从 F^n 到 F^m 的一个**线性映射**, 因为它满足线性要求: 对任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in F^n$ 以及 $\lambda_1, \lambda_2 \in F$ 都有 $\mathscr{A}(\lambda_1\boldsymbol{x}_1 + \lambda_2\boldsymbol{x}_2) = \lambda_1\mathscr{A}(\boldsymbol{x}_1) + \lambda_2\mathscr{A}(\boldsymbol{x}_2)$. 于是, 从一个 $m \times n$ 矩阵出发, 我们就会得到一个从 n 维数组空间到 m 维数组空间的线性映射.

这反过来也是正确的, 即若 $\mathscr{A}: F^n \to F^m$ 是一个线性映射, 则存在一个 $m \times n$ 矩阵 $\boldsymbol{A} = (a_{ij})$, 使得 $\mathscr{A}$ 恰由公式 (†) 给出. 关于这一点, 请大家借用数组空间的基本向量组来验证. 于是, 数组空间之间的线性映射与矩阵就对应起来了.

接下来, 我们再引入一个新的矩阵 $\boldsymbol{B} = (b_{ij}) \in F^{n\times p}$, 于是它对应于某个线性映射

$$\mathscr{B}: F^p \to F^n, \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} \mapsto \begin{pmatrix} \sum_{k=1}^p b_{1k}y_k \\ \sum_{k=1}^p b_{2k}y_k \\ \vdots \\ \sum_{k=1}^p b_{nk}y_k \end{pmatrix}.$$

然后, 我们再来看复合映射

$$\mathscr{C} := \mathscr{A} \circ \mathscr{B}: F^p \to F^n \to F^m.$$

容易验证, $\mathscr{C}$ 仍然是一个线性映射, 从而存在矩阵 $\boldsymbol{C} = (c_{ij}) \in F^{m\times p}$ 使得

$$\mathscr{C}: F^p \to F^m, \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} \mapsto \begin{pmatrix} \sum_{k=1}^p c_{1k}y_k \\ \sum_{k=1}^p c_{2k}y_k \\ \vdots \\ \sum_{k=1}^p c_{mk}y_k \end{pmatrix}.$$

可以直接验证

$$c_{ij} = \sum_{k=1} a_{ik}b_{kj}.$$

换言之, $\boldsymbol{C}$ 就是我们之前引入的矩阵乘法 $\boldsymbol{AB}$.

线性映射是线性代数这门课程里最重要的研究对象之一, 而映射的复合是自然定义的, 不需要加什么额外的条件. 在处理线性映射的复合时, 我们将其视作线性映射的 "乘

法". 基于线性映射与矩阵的对应关系, 我们将上面引入的 $\boldsymbol{C}$ 称作 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的乘法, 记作 $\boldsymbol{AB}$. 这意味着, 我们考虑如此"复杂"的矩阵的乘积, 是因为线性映射的复合迫使我们不得不这么处理.

习题 4.14. 计算

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix}.$$

例 4.15. 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$, $\boldsymbol{B}=\operatorname{diag}(b_1,\ldots,b_m)$, $\boldsymbol{C}=\operatorname{diag}(c_1,\ldots,c_n)$. 由学生来计算 教材例题 4.2.4

$$\boldsymbol{BA}=\begin{pmatrix} b_1a_{11} & b_1a_{12} & \cdots & b_1a_{1n} \\ b_2a_{21} & b_2a_{22} & \cdots & b_2a_{1n} \\ \vdots & \vdots & & \vdots \\ b_ma_{m1} & b_ma_{m2} & \cdots & b_ma_{mn} \end{pmatrix}, \qquad \boldsymbol{AC}=\begin{pmatrix} a_{11}c_1 & a_{12}c_2 & \cdots & a_{1n}c_n \\ a_{21}c_2 & a_{22}c_2 & \cdots & a_{2n}c_n \\ \vdots & \vdots & & \vdots \\ a_{m1}c_m & a_{m2}c_2 & \cdots & a_{mn}c_n \end{pmatrix}.$$

借用这个例子, 希望学生能体会对于一个矩阵分别左乘和右乘一个特殊矩阵的意义.

按照矩阵的乘法规则, 矩阵 $\boldsymbol{A}\in F^{m\times n}$ 能与自身相乘的充要条件是 $m=n$. 若 $\boldsymbol{A}$ 为 n 阶方阵, 则它可以与自身反复相乘. 对于一个正整数 k, 我们自然地定义

$$\boldsymbol{A}^k := \underbrace{\boldsymbol{A}\boldsymbol{A}\cdots\boldsymbol{A}}_{k\text{ 个}}.$$

我们也约定 $\boldsymbol{A}^0=\boldsymbol{I}_n$ (n 阶单位矩阵). 更一般地, 若有一个多项式

$$f(x)=c_0+c_1x+c_2x^2+\cdots+c_kx^k,$$

则我们定义**矩阵多项式**

$$f(\boldsymbol{A}) := c_0\boldsymbol{I}+c_1\boldsymbol{A}+c_2\boldsymbol{A}^2+\cdots+c_k\boldsymbol{A}^k.$$

注 4.16. (1) 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为给定的 n 阶方阵. 一般情况下, $\boldsymbol{AB}\neq\boldsymbol{BA}$. 但是若有 $\boldsymbol{AB}=\boldsymbol{BA}$, 则可以验证, 对于任意的非负整数 k 皆有 $(\boldsymbol{AB})^k=(\boldsymbol{BA})^k=\boldsymbol{A}^k\boldsymbol{B}^k$.

(2) 若 $f(x)$ 和 $g(x)$ 都是多项式, $\boldsymbol{A}$ 为方阵, 则

$$f(\boldsymbol{A})g(\boldsymbol{A})=g(\boldsymbol{A})f(\boldsymbol{A})=\underbrace{fg}_{\text{乘积多项式}}(\boldsymbol{A}).$$

(3) 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为同阶方阵, 并满足 $\boldsymbol{AB}=\boldsymbol{BA}$, 则有二项式展开公式

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

这儿的组合数 C_n^i 为熟知的二项式展开的系数.

例 4.17. 求 *Fibonacci* 数列 $F_1=F_2=1$, $F_n=F_{n-1}+F_{n-2}$ $(n\geq 3)$ 的通项公式. 教材例题 4.2.6

解. **(解法一)** 解相应的特征方程 $x^2=x+1$ (一般地, 若迭代方程为 $aF_n=bF_{n-1}+cF_{n-2}$, 则相应的特征方程为 $ax^2=bx+c$), 得到 $x=\frac{1\pm\sqrt{5}}{2}$. 令 $x_1=\frac{1+\sqrt{5}}{2}$, $x_2=\frac{1-\sqrt{5}}{2}$. 此时, 解方程组

$$\begin{cases} F_n=k_1x_1^n+k_2x_2^n \\ F_1=F_2=1 \end{cases} \Rightarrow \boxed{\begin{cases} k_1\cdot\dfrac{1+\sqrt{5}}{2}+k_2\cdot\dfrac{1-\sqrt{5}}{2}=1 \\ k_1\left(\dfrac{1+\sqrt{5}}{2}\right)^2+k_2\left(\dfrac{1-\sqrt{5}}{2}\right)^2=1 \end{cases}} \Rightarrow \begin{cases} k_1=\dfrac{1}{\sqrt{5}} \\ k_2=-\dfrac{1}{\sqrt{5}} \end{cases}$$

学生课前
解该方程

故 $F_n=\frac{1}{\sqrt{5}}\cdot\left(\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n\right)$.

(解法二) 考虑辅助方程组

$$\begin{cases} F_n=F_{n-1}+F_{n-2} \\ F_{n-1}=F_{n-1} \end{cases} \Leftrightarrow \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}=\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix} \quad (n\geq 3).$$

反复利用上面的公式 (或者说用归纳法), 我们可以得到

$$\begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}=\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-2}\begin{pmatrix} F_2 \\ F_1 \end{pmatrix}=\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-2}\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

从而

$$F_n=\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}=\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

于是, 问题归结到求解矩阵的幂 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-2}$. 在教材第六章中, 我们会利用矩阵的特征值将该矩阵设法相似对角化, 化为对角阵的幂来求解. 而解法一中出现的 x_1 和 x_2 就是该矩阵的特征值. □

注 4.18. 对于矩阵方幂的求解, 除了上面提到的利用相似对角化的方法外, 通常还可以

(1) 借用 *Caylay–Hamilton* 定理 (也要涉及到矩阵的特征值问题);

(2) 对于一些特殊的矩阵作专门的分解, 再利用其特殊结构来计算;

(3) 对于一些特殊地矩阵先进行初步地计算总结规律, 再利用归纳法.

例 4.19. 对于矩阵 $\boldsymbol{A}=\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}$, 计算 $\boldsymbol{A}^m$, 其中 m 为正整数.

解. **(解法一)** 我们观察到

$$\boldsymbol{A}=a\begin{pmatrix}1&0\\0&1\end{pmatrix}+b\begin{pmatrix}0&1\\0&0\end{pmatrix}=:a\boldsymbol{I}+b\boldsymbol{J},$$

其中 $\boldsymbol{IJ}=\boldsymbol{JI}$ (乘法可交换), 且 $\boldsymbol{J}^2=\boldsymbol{O}$. 此时, 利用二项式公式, 我们有

$$\begin{aligned}\boldsymbol{A}^m&=(a\boldsymbol{I}+b\boldsymbol{J})^m=\sum_{k=0}^{m}C_m^k(a\boldsymbol{I})^k(b\boldsymbol{J})^{m-k}\\&=\sum_{k=0}^{m}C_m^ka^kb^{m-k}\boldsymbol{I}^k\boldsymbol{J}^{m-k}=a^m\boldsymbol{I}+C_m^1a^{m-1}b\boldsymbol{J}\\&=\begin{pmatrix}a^m&ma^{m-1}b\\0&a^m\end{pmatrix}.\end{aligned}$$

(解法二) 直接计算有

$$\boldsymbol{A}^2=\begin{pmatrix}a^2&2ab\\0&a^2\end{pmatrix},\qquad\boldsymbol{A}^3=\begin{pmatrix}a^3&3a^2b\\0&a^3\end{pmatrix},$$

于是猜测并用归纳法可证 $\boldsymbol{A}=\begin{pmatrix}a^m&ma^{m-1}b\\0&a^m\end{pmatrix}$. □

例 4.20. 对于矩阵 $\boldsymbol{A}=\begin{pmatrix}a&b\\0&c\end{pmatrix}$, 计算 $\boldsymbol{A}^m$, 其中 m 为正整数.

在这儿, $\boldsymbol{A}=\begin{pmatrix}a&0\\0&c\end{pmatrix}+\begin{pmatrix}0&b\\0&0\end{pmatrix}\triangleq\boldsymbol{B}+\boldsymbol{C}$. $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 的高阶方幂都很好求, 可是一般而言, $\boldsymbol{BC}\neq\boldsymbol{CB}$, 故无法利用二项式展开的公式来求解.

解. 直接计算, 有

$$\begin{aligned}\boldsymbol{A}^2&=\begin{pmatrix}a^2&(a+c)b\\0&c^2\end{pmatrix},\qquad\boldsymbol{A}^3=\begin{pmatrix}a^3&(a^2+ac+c^2)b\\0&c^3\end{pmatrix},\\\boldsymbol{A}^4&=\begin{pmatrix}a^4&(a^3+a^2c+ac^2+c^3)b\\0&c^4\end{pmatrix}.\end{aligned}$$

因此, 可以猜测并通过归纳法来证明

$$\boldsymbol{A}^m=\begin{pmatrix}a^m&(a^{m-1}+a^{m-2}c+a^{m-3}c^2+\cdots+ac^{m-2}+c^{m-1})b\\0&c^m\end{pmatrix}.$$ □

例 4.21. 求与 $\boldsymbol{A}=\begin{pmatrix}3&1&0\\0&3&1\\0&0&3\end{pmatrix}$ 乘法可交换的所有矩阵.

解. 设 $\boldsymbol{X}$ 是满足条件的矩阵, 则它必为一个 3 阶方阵. 若设 $\boldsymbol{J}=\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}$, 则 $\boldsymbol{A}=3\boldsymbol{I}+\boldsymbol{J}$, 于是 $\boldsymbol{AX}=\boldsymbol{XA}\Leftrightarrow\boldsymbol{JX}=\boldsymbol{XJ}$. 设 $\boldsymbol{X}=(x_{ij})_{3\times3}$, 直接计算可得

$$\boldsymbol{JX}=\begin{pmatrix}x_{21}&x_{22}&x_{23}\\x_{31}&x_{32}&x_{33}\\0&0&0\end{pmatrix},\qquad \boldsymbol{XJ}=\begin{pmatrix}0&x_{11}&x_{12}\\0&x_{21}&x_{22}\\0&x_{31}&x_{32}\end{pmatrix}.$$

于是前面的矩阵方程等价于

$$x_{21}=x_{31}=x_{32}=0,\qquad x_{11}=x_{22}=x_{33},\qquad x_{23}=x_{12}.$$

这说明所求的矩阵的一般形式为

$$\boldsymbol{X}=\begin{pmatrix}x_{11}&x_{12}&x_{13}\\0&x_{11}&x_{12}\\0&0&x_{11}\end{pmatrix}=x_{11}\boldsymbol{I}+x_{12}\boldsymbol{J}+x_{13}\boldsymbol{J}^2,$$

其中 $x_{11},x_{12},x_{13}\in F$ 任意. □

习题 4.22. 计算

$$\boldsymbol{J}=\begin{pmatrix}0&1&0&\cdots&0\\\vdots&\ddots&\ddots&\ddots&\vdots\\&&&\ddots&0\\&&&\ddots&1\\0&\cdots&\cdots&\cdots&0\end{pmatrix}_{n\times n}$$

的幂 $\boldsymbol{J}^t$ $(t\geq1)$, 并找出与 $\boldsymbol{J}$ 乘法可交换的所有矩阵.

逆矩阵 先回顾一下, 对一个实数 a, 如果存在一个数 b 使得 $ab=1$, 则称 a 关于乘法有逆元 (倒数). 任何非零的实数 a 都存在唯一的一个逆元 $b=\frac{1}{a}$. 我们可以将这个概念推广到一般方阵的逆.

类似地, 设矩阵 $\boldsymbol{A}\in F^{n\times n}$, 如果存在 $\boldsymbol{B}\in F^{n\times n}$ 使得 $\boldsymbol{BA}=\boldsymbol{AB}=\boldsymbol{I}_n$, 则称 $\boldsymbol{A}$ 是**可逆的**或**非奇异的**. 满足条件的矩阵 $\boldsymbol{B}$ 称为 $\boldsymbol{A}$ (在 $\mathbb{F}^{n\times n}$ 中) 的**逆矩阵**.

注意一下, 1 在实数乘法里为单位元, 即对任意的实数 a 皆有 $1 \cdot a = a$. 而在矩阵乘法里, 对任意的方阵 $\boldsymbol{A}$, 皆有 $\boldsymbol{IA} = \boldsymbol{A} = \boldsymbol{AI}$. 因此, 在涉及到矩阵乘法时, 我们经常将单位方阵 $\boldsymbol{I}$ 类比于实数乘法中的 1.

注 4.23. (1) 对于方阵 $\boldsymbol{A}$, 它的逆矩阵若存在必唯一, 即, $\boldsymbol{A}$ 至多有一个逆矩阵. 为了看到这一点, 假设 $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 都是 $\boldsymbol{A}$ 的逆矩阵. 此时,

$$\boldsymbol{B} = \boldsymbol{BI} = \boldsymbol{B}(\boldsymbol{AC}) = (\boldsymbol{BA})\boldsymbol{C} = \boldsymbol{IC} = \boldsymbol{C}.$$

因此, 在 $\boldsymbol{A}$ 的逆矩阵存在的条件下, 我们将该逆矩阵记作 $\boldsymbol{A}^{-1}$.

(2) 若 $\boldsymbol{B}$ 是 $\boldsymbol{A}$ 的逆矩阵, 则 $\boldsymbol{A}$ 也是 $\boldsymbol{B}$ 的逆矩阵. 因此, 我们可以称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 互为逆矩阵. 用公式表示有 $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.

(3) 在之后利用行列式的方法我们可以验证, 若同阶方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 满足 $\boldsymbol{AB} = \boldsymbol{I}$, 则必有 $\boldsymbol{BA} = \boldsymbol{I}$, 从而 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 互为逆矩阵.

(4) 并不是所有的方阵都存在逆矩阵的. 显然, 零矩阵 $\boldsymbol{O}_n$ 不存在逆矩阵. 由于

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ 0 & 0 \end{pmatrix},$$

而该等式右边的方阵永远都不可能成为单位矩阵 $\boldsymbol{I}_2$, 因此 $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ 也不是可逆的.

(5) 当然, 可逆矩阵是存在的, $\boldsymbol{I}_n$ 就是可逆矩阵. 可以直接验算, $\begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix}$ 和 $\begin{pmatrix} 7 & -3 \\ -2 & 1 \end{pmatrix}$ 互为逆矩阵. 一般情形下, 如何求解形如 $\begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix}$ 的方阵的逆矩阵? 由定义, 假定其逆矩阵存在, 可以设为 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. 于是有矩阵方程

$$\begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

这等价于同时解两个线性方程组

$$\begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix} \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{和} \quad \begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

之前提到, 可以对各自的增广矩阵利用 *Gauss* 消元法来分别求解.

$$\left(\begin{array}{cc|c} 1 & 3 & 1 \\ 2 & 7 & 0 \end{array}\right) \xrightarrow{-2r_1 \to r_2} \left(\begin{array}{cc|c} 1 & 3 & 1 \\ 0 & 1 & -2 \end{array}\right) \xrightarrow{-3r_2 \to r_1} \left(\begin{array}{cc|c} 1 & 0 & 7 \\ 0 & 1 & -2 \end{array}\right)$$

$$\left(\begin{array}{cc|c} 1 & 3 & 0 \\ 2 & 7 & 1 \end{array}\right) \xrightarrow{-2r_1 \to r_2} \left(\begin{array}{cc|c} 1 & 3 & 0 \\ 0 & 1 & 1 \end{array}\right) \xrightarrow{-3r_2 \to r_1} \left(\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & 1 \end{array}\right)$$

Gauss 消元法的重点在于对其系数矩阵化简成 (约化) 标准形. 显然, 这两个方程组的系数矩阵一致, 因此我们采取了相同的化简策略. 但是我们显然没有必要进行重复的操作, 它们可以同时进行:

$$\left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 2 & 7 & 0 & 1 \end{array}\right) \xrightarrow{-2r_1 \to r_2} \left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{array}\right) \xrightarrow{-3r_2 \to r_1} \left(\begin{array}{cc|cc} 1 & 0 & 7 & -3 \\ 0 & 1 & -2 & 1 \end{array}\right)$$

由解方程组的意义知所求的逆矩阵为

$$\begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix}^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 7 & -3 \\ -2 & 1 \end{pmatrix}.$$

这样的方法是求解可逆矩阵的逆矩阵的标准方法, 之后会重新讨论的.

用定义, 我们可以很容易地证得以下结果.

定理 4.24. 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为同阶的可逆矩阵. 教材定理 4.2.3

(1) $\boldsymbol{A}^{-1}$ 可逆, 且 $(\boldsymbol{A}^{-1})^{-1}=\boldsymbol{A}$.

(2) 若 $\lambda\neq 0$, 则 $\lambda\boldsymbol{A}$ 可逆, 且 $(\lambda\boldsymbol{A})^{-1}=\lambda^{-1}\boldsymbol{A}^{-1}$.

(3) $\boldsymbol{AB}$ 可逆, 且 $(\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$. (形象的说法是, 脱鞋的逆是穿鞋, 脱袜子的逆是穿袜子. 我们睡觉前是先脱鞋再脱袜子, 起床时是先穿袜子再穿鞋.)

推论 4.25. (1) 设 $\boldsymbol{A}_1,\boldsymbol{A}_2,\ldots,\boldsymbol{A}_k$ 是同阶的可逆矩阵, 则

$$(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_{k-1}\boldsymbol{A}_k)^{-1}=\boldsymbol{A}_k^{-1}\boldsymbol{A}_{k-1}^{-1}\cdots\boldsymbol{A}_2^{-1}\boldsymbol{A}_1^{-1}.$$

(2) 设 $\boldsymbol{A}$ 为可逆矩阵. 若 k 为正整数, 则 $(\boldsymbol{A}^k)^{-1}=(\boldsymbol{A}^{-1})^k$; 我们将该矩阵记作 $\boldsymbol{A}^{-k}$. 此时, 若 m,n 为整数, 则 $\boldsymbol{A}^m\boldsymbol{A}^n=\boldsymbol{A}^{m+n}$.

注 4.26. 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为 n 阶方阵, 且 $\boldsymbol{AB}$ 可逆, 则 $\boldsymbol{A}(\boldsymbol{B}(\boldsymbol{AB})^{-1})=(\boldsymbol{AB})(\boldsymbol{AB})^{-1}=\boldsymbol{I}$, 这说明 $\boldsymbol{A}$ 可逆, 并且 $\boldsymbol{B}=\boldsymbol{A}^{-1}(\boldsymbol{AB})$ 也可逆.

例 4.27. 设 $\boldsymbol{A}$ 为方阵, k 为某个正整数, 满足 $\boldsymbol{A}^k=\boldsymbol{O}$. 由多项式恒等式

$$x^k-1=(x-1)(x^{k-1}+x^{k-2}+\cdots+x+1)$$

可知,

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

这说明 $\boldsymbol{A}-\boldsymbol{I}$ 可逆, 且

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

习题 4.28. 设 $\boldsymbol{A}$ 为方阵, k 为某个正整数, 满足 $\boldsymbol{A}^k=\boldsymbol{O}$. 若 $\lambda\in F=\mathbb{R}$, $\boldsymbol{A}+\lambda\boldsymbol{I}$ 是否可逆? 若可逆, 求出其逆矩阵; 若不可逆, 解释原因.

例 4.29. 假设方阵 $\boldsymbol{A}$ 满足方程 $\boldsymbol{A}^2+\boldsymbol{A}-2\boldsymbol{I}=\boldsymbol{O}$, 我们验证 $\boldsymbol{A}-2\boldsymbol{I}$ 可逆, 并求其逆. 为此, 对多项式 x^2+x-2 关于 $x-2$ 作带余除法:

$$x^2+x-2=(x-2)(x+3)+4.$$

这说明

$$\boldsymbol{O}=\boldsymbol{A}^2+\boldsymbol{A}-2\boldsymbol{I}=(\boldsymbol{A}-2\boldsymbol{I})(\boldsymbol{A}+3\boldsymbol{I})+4\boldsymbol{I}.$$

从而

$$(\boldsymbol{A}-2\boldsymbol{I})\left(-\frac{1}{4}(\boldsymbol{A}+3\boldsymbol{I})\right)=\boldsymbol{I}.$$

这说明 $\boldsymbol{A}-2\boldsymbol{I}$ 可逆, 且其逆矩阵为 $-\frac{1}{4}(\boldsymbol{A}+3\boldsymbol{I})$.

矩阵的转置 从一个 $m \times n$ 矩阵 $\boldsymbol{A} = (a_{ij})$ 出发, 我们有时候需要构造一个各列是 $\boldsymbol{A}$ 的各行的 $n \times m$ 矩阵. 这样的矩阵称为 $\boldsymbol{A}$ 的**转置矩阵** (transpose), 记作

$$\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{t} = \boldsymbol{A}^{\tau} = \boldsymbol{A}' := \begin{pmatrix} a_{11} & a_{21} \cdots\cdots a_{m1} \\ a_{12} & a_{22} \cdots\cdots a_{m2} \\ \vdots & \vdots \qquad\qquad \vdots \\ a_{1n} & a_{2n} \cdots\cdots a_{mn} \end{pmatrix}.$$

例 4.30. 对于 $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, 我们有转置矩阵 $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

定理 4.31. 矩阵的转置运算有如下的基本性质 (在运算允许的条件下). 教材定理 4.2.4

(1) $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}$.

(2) $(\boldsymbol{A} + \boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}$.

(3) $(\lambda \boldsymbol{A})^{\mathrm{T}} = \lambda \boldsymbol{A}^{\mathrm{T}}$, 其中 λ 为标量.

(4) $(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}}$.

(5) 若 $\boldsymbol{A}$ 为可逆方阵, 则 $\boldsymbol{A}^{\mathrm{T}}$ 可逆, 且 $(\boldsymbol{A}^{\mathrm{T}})^{-1} = (\boldsymbol{A}^{-1})^{\mathrm{T}}$.

证明. (4) 的证明. 设 $\boldsymbol{A} \in F^{m\times n}$, $\boldsymbol{B} \in F^{n\times p}$, 则 $\boldsymbol{A}^{\mathrm{T}} \in F^{n\times m}$, $\boldsymbol{B}^{\mathrm{T}} \in F^{p\times n}$. 为了说明的方便, 在这儿, 对于一个矩阵右下角的双重下标 ij, 我们是在表示该矩阵的第 (i,j) 元素. 此时

$$\begin{array}{ccccc} \left((\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}}\right)_{ij} & = & (\boldsymbol{A}\boldsymbol{B})_{ji} & = & \sum_{k=1}^{n} \boldsymbol{A}_{jk}\boldsymbol{B}_{ki} \\ & & & & \| \\ (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}})_{ij} & = & \sum_{k=1}^{n} (\boldsymbol{B}^{\mathrm{T}})_{ik}(\boldsymbol{A}^{\mathrm{T}})_{kj} & = & \sum_{k=1}^{n} \boldsymbol{B}_{ki}\boldsymbol{A}_{jk} \end{array}$$

(5) 的证明. 只需验证 $(\boldsymbol{A}^{-1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$. 这一点可直接验证:

$$(\boldsymbol{A}^{-1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} = (\boldsymbol{A}\boldsymbol{A}^{-1})^{\mathrm{T}} = \boldsymbol{I}^{\mathrm{T}} = \boldsymbol{I}.$$ □

推论 4.32. 在矩阵乘法有意义的条件下,

$$(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_{k-1}\boldsymbol{A}_k)^{\mathrm{T}} = \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_{k-1}^{\mathrm{T}}\cdots\boldsymbol{A}_2^{\mathrm{T}}\boldsymbol{A}_1^{\mathrm{T}}.$$

从转置运算出发, 我们可以引入更多的概念.

(1) 方阵 $\boldsymbol{A}$ 称为**对称的** (symmetric) 是指其满足 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$.

(2) 方阵 $\boldsymbol{A}$ 称为**反对称的**或者**斜对称的** (skew-symmetric) 是指其满足 $\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$. 这样的矩阵的主对角线上的元素全为 0.

(3) 实数方阵 $\boldsymbol{A}$ 称为是**正交的** (orthogonal) 是指其满足 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{I}$ (或等价地, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}$). 不难看出, 正交矩阵 $\boldsymbol{A}$ 可逆, 且它的逆 $\boldsymbol{A}^{-1}$ 正好为 $\boldsymbol{A}^{\mathrm{T}}$.

习题 4.33. (1) 给定列向量 $\boldsymbol{u}\in\mathbb{R}^n$, 假定 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1$. 对于 $\boldsymbol{P}=\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ 以及 $\boldsymbol{Q}=\boldsymbol{I}_n-2\boldsymbol{P}$, 证明:

(i) $\boldsymbol{P}^2=\boldsymbol{P}$, (ii) $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{P}$, (iii) $\boldsymbol{Q}^2=\boldsymbol{I}_n$.

(2) 在上面一小问中, 变换 $\boldsymbol{x}\mapsto\boldsymbol{P}\boldsymbol{x}$ 被称作一个**投影**, 而 $\boldsymbol{x}\mapsto\boldsymbol{Q}\boldsymbol{x}$ 被称作一个**豪斯霍尔德反射** (*Householder reflection*). 为了理解这一点, 对于 $\boldsymbol{u}=\begin{pmatrix}0\\0\\1\end{pmatrix}$ 和 $\boldsymbol{x}=\begin{pmatrix}1\\5\\3\end{pmatrix}$, 分别计算 $\boldsymbol{P}\boldsymbol{x}$ 以及 $\boldsymbol{Q}\boldsymbol{x}$. 用几何的语言来解释 $\boldsymbol{Q}\boldsymbol{x}$ 与 $\boldsymbol{x}$ 的关系.

复矩阵的共轭 对于复数 $z=a+b\mathrm{i}\in\mathbb{C}$ $(a,b\in\mathbb{R})$, 它的 (复) 共轭是 $\overline{z}=a-b\mathrm{i}\in\mathbb{C}$. 类似地, 对于复矩阵 $\boldsymbol{A}=(a_{ij})_{m\times n}$, 我们可以定义其**共轭** (conjugate) 为

$$\overline{\boldsymbol{A}}\coloneqq(\overline{a_{ij}})=\begin{pmatrix}\overline{a_{11}} & \overline{a_{12}} & \cdots\cdots & \overline{a_{1n}}\\ \overline{a_{21}} & \overline{a_{22}} & \cdots\cdots & \overline{a_{2n}}\\ \vdots & \vdots & & \vdots\\ \overline{a_{m1}} & \overline{a_{m2}} & \cdots\cdots & \overline{a_{mn}}\end{pmatrix}.$$

例 4.34. 对于 $\boldsymbol{A}=\begin{pmatrix}1 & 3+5i & 7+8i\\ 13-6i & 2 & 3\end{pmatrix}$, 我们有 $\overline{\boldsymbol{A}}=\begin{pmatrix}1 & 3-5i & 7-8i\\ 13+6i & 2 & 3\end{pmatrix}$.

注 4.35. (1) 作为正交矩阵的推广, 复数方阵 $\boldsymbol{A}$ 称为一个**酉矩阵**是指其满足 $\boldsymbol{A}\overline{\boldsymbol{A}}^{\mathrm{T}}=\boldsymbol{I}$. 于是, 一个实方阵是正交矩阵当且仅当它是一个酉矩阵.

(2) 对于一般的复矩阵 $\boldsymbol{A}$ 而言, 有 $(\overline{\boldsymbol{A}})^{\mathrm{T}}=\overline{\boldsymbol{A}^{\mathrm{T}}}$; 我们可以将其称为 $\boldsymbol{A}$ 的共轭转置 (*conjugate transpose*). 有不少教材会将其记作 $\boldsymbol{A}^*$, 但是这与我们教材里稍后要介绍的伴随矩阵的记号相冲突, 所以我们暂时不采用这样的记号. 当然, 也有不少教材会将其记作 $\boldsymbol{A}^{\mathrm{H}}$.

方阵的迹　方阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 的对角线元素的和称为 $\boldsymbol{A}$ 的**迹** (trace):

$$\operatorname{tr}(\boldsymbol{A}) := a_{11}+a_{22}+\cdots+a_{nn}=\sum_{i=1}^{n} a_{ii}.$$

例 4.36. (1) 对于 $\boldsymbol{A}=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}$, 我们有 $\operatorname{tr}(\boldsymbol{A})=1+5+9=15$.

(2) $\operatorname{tr}(\boldsymbol{I}_n)=n$.

(3) 若 $\boldsymbol{A}$ 为反对称矩阵, 则 $\operatorname{tr}(\boldsymbol{A})=0$.

定理 4.37. 矩阵的迹有如下的性质 (在运算允许的条件下). 教材定理 4.2.5

(1) $\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B})=\operatorname{tr}(\boldsymbol{A})+\operatorname{tr}(\boldsymbol{B})$.

(2) $\operatorname{tr}(\lambda\boldsymbol{A})=\lambda\operatorname{tr}(\boldsymbol{A})$, 其中 λ 为标量.

(3) $\operatorname{tr}(\boldsymbol{A})=\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}})$, $\operatorname{tr}(\overline{\boldsymbol{A}})=\overline{\operatorname{tr}(\boldsymbol{A})}$.

(4) $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B})=\operatorname{tr}(\boldsymbol{B}\boldsymbol{A})$. 这儿的 $\boldsymbol{A}\in F^{m\times n}$, $\boldsymbol{B}\in F^{n\times m}$, 不要求矩阵为方阵, 只要求相应的维数相反.

证明. (4) 的证明. 为了说明的方便, 在这儿, 对于一个矩阵右下角的双重下标 ij, 我们是在表示该矩阵的第 (i,j) 元素. 此时,

$$\begin{array}{ccccc}
\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) & = & \sum_{i=1}^{m}(\boldsymbol{A}\boldsymbol{B})_{ii} & = & \sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{A}_{ij}\boldsymbol{B}_{ji}\\
 & & & & \parallel\\
\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) & = & \sum_{j=1}^{n}(\boldsymbol{B}\boldsymbol{A})_{jj} & = & \sum_{j=1}^{n}\sum_{i=1}^{m}\boldsymbol{B}_{ji}\boldsymbol{A}_{ij}
\end{array}$$

□

例 4.38. 对于 $\boldsymbol{A}\in\mathbb{C}^{m\times n}$, 证明:

$$\boldsymbol{A}=\boldsymbol{O}\Leftrightarrow\operatorname{tr}(\boldsymbol{A}\overline{\boldsymbol{A}}^{\mathrm{T}})=0\Leftrightarrow\operatorname{tr}(\overline{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{A})=0.$$

证明. 由于有定理 4.37 (4), 我们只需证明 "$\operatorname{tr}(\boldsymbol{A}\overline{\boldsymbol{A}}^{\mathrm{T}})=0\Rightarrow\boldsymbol{A}=\boldsymbol{O}$". 为此, 设 $\boldsymbol{A}=(a_{ij})$, 则由上面的计算可知 $0=\operatorname{tr}(\boldsymbol{A}\overline{\boldsymbol{A}}^{\mathrm{T}})=\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2$. 右式的求和是一些非负实数的和, 故对每个 (i,j), 必有 $|a_{ij}|^2=0$, 即 $a_{ij}=0$. 这说明 $\boldsymbol{A}=\boldsymbol{O}$. □

习题 4.39. 对于 n 阶实方阵 $\boldsymbol{A}$, 我们可以定义函数 $\psi_{\boldsymbol{A}}:\mathbb{R}^{n\times n}\to\mathbb{R}$, $\boldsymbol{X}\mapsto\operatorname{tr}(\boldsymbol{A}\boldsymbol{X})$. 若 n 阶实方阵 $\boldsymbol{A},\boldsymbol{B}$ 满足 $\psi_{\boldsymbol{A}}=\psi_{\boldsymbol{B}}$, 证明: $\boldsymbol{A}=\boldsymbol{B}$.

分块矩阵　若把矩阵 $\boldsymbol{A}$ 按行分成若干组, 按列分成若干组, 则 $\boldsymbol{A}$ 可以视为由若干个子矩阵有序排列构成的数表:

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\cdots\cdots & \boldsymbol{A}_{1s}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\cdots\cdots & \boldsymbol{A}_{2s}\\ \vdots & \vdots & \vdots\\ \boldsymbol{A}_{r1} & \boldsymbol{A}_{r2}\cdots\cdots & \boldsymbol{A}_{rs}\end{pmatrix}.$$

这是将 $\boldsymbol{A}$ 写成**分块矩阵**的形式, 并可简写为 $\boldsymbol{A}=(\boldsymbol{A}_{ij})_{r\times s}$. 所有的这些 $\boldsymbol{A}_{ij}$ 称为 $\boldsymbol{A}$ 的**子块** (block). 显然, 上面的表示中, 任何一行的子块具有相同的行数, 任何一列的子块具有相同的列数.

例 4.40. 对于矩阵

$$\boldsymbol{C}=\begin{pmatrix}1 & -2 & 4 & 1 & 3\\ 2 & 1 & 1 & 1 & 1\\ 3 & 3 & 2 & -1 & 2\\ 4 & 6 & 2 & 2 & 4\end{pmatrix},$$

我们可以将一、二行视为一组, 三、四行视为一组, 从而在第二、三行之间画一条横线, 我们再将一、二、三列视为一组, 四、五列视为一组, 从而在第三、四列之间画一条竖线, 于是将 $\boldsymbol{C}$ 分块得到了如下的表示形式:

$$\boldsymbol{C}=\begin{pmatrix}\boldsymbol{C}_{11} & \boldsymbol{C}_{12}\\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22}\end{pmatrix}=\left(\begin{array}{ccc:cc}1 & -2 & 4 & 1 & 3\\ 2 & 1 & 1 & 1 & 1\\ \hdashline 3 & 3 & 2 & -1 & 2\\ 4 & 6 & 2 & 2 & 4\end{array}\right),$$

其中

$$\boldsymbol{C}_{11}=\begin{pmatrix}1 & -2 & 4\\ 2 & 1 & 1\end{pmatrix}\quad \boldsymbol{C}_{12}=\begin{pmatrix}1 & 3\\ 1 & 1\end{pmatrix}\quad \boldsymbol{C}_{21}=\begin{pmatrix}3 & 3 & 2\\ 4 & 6 & 2\end{pmatrix}\quad \boldsymbol{C}_{22}=\begin{pmatrix}-1 & 2\\ 2 & 4\end{pmatrix}.$$

注 4.41. 本章一开始提到的, 将矩阵写成将其列向量按行排列的形式, 或者写成将其行向量按列排列的形式, 都是将矩阵写成分块矩阵的特殊情形.

分块矩阵的运算只是矩阵运算的分块表示而言. 具体来说, 我们有如下的操作; 它们的证明只需直接运用定义即可.

(1) **加法**. 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为同型 (具有相同的维数) 的矩阵, 并且采用相同的方法分块:

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\cdots\cdots\boldsymbol{A}_{1s}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\cdots\cdots\boldsymbol{A}_{2s}\\ \vdots & \vdots \qquad\qquad \vdots\\ \boldsymbol{A}_{r1} & \boldsymbol{A}_{r2}\cdots\cdots\boldsymbol{A}_{rs}\end{pmatrix},\qquad \boldsymbol{B}=\begin{pmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\cdots\cdots\boldsymbol{B}_{1s}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\cdots\cdots\boldsymbol{B}_{2s}\\ \vdots & \vdots \qquad\qquad \vdots\\ \boldsymbol{B}_{r1} & \boldsymbol{B}_{r2}\cdots\cdots\boldsymbol{B}_{rs}\end{pmatrix},$$

其中 $\boldsymbol{A}_{ij}$ 与 $\boldsymbol{B}_{ij}$ 是同型矩阵, 则

$$\boldsymbol{A}+\boldsymbol{B}=\begin{pmatrix}\boldsymbol{A}_{11}+\boldsymbol{B}_{11} & \boldsymbol{A}_{12}+\boldsymbol{B}_{12}\cdots\cdots\boldsymbol{A}_{1s}+\boldsymbol{B}_{1s}\\ \boldsymbol{A}_{21}+\boldsymbol{B}_{21} & \boldsymbol{A}_{22}+\boldsymbol{B}_{22}\cdots\cdots\boldsymbol{A}_{2s}+\boldsymbol{B}_{2s}\\ \vdots & \vdots \qquad\qquad \vdots\\ \boldsymbol{A}_{r1}+\boldsymbol{B}_{r1} & \boldsymbol{A}_{r2}+\boldsymbol{B}_{r2}\cdots\cdots\boldsymbol{A}_{rs}+\boldsymbol{B}_{rs}\end{pmatrix}.$$

(2) **数乘**. 假设 $\boldsymbol{A}$ 分块如上, 则其对于标量 k 的数乘为

$$k\boldsymbol{A}=\begin{pmatrix}k\boldsymbol{A}_{11} & k\boldsymbol{A}_{12}\cdots\cdots k\boldsymbol{A}_{1s}\\ k\boldsymbol{A}_{21} & k\boldsymbol{A}_{22}\cdots\cdots k\boldsymbol{A}_{2s}\\ \vdots & \vdots \qquad\qquad \vdots\\ k\boldsymbol{A}_{r1} & k\boldsymbol{A}_{r2}\cdots\cdots k\boldsymbol{A}_{rs}\end{pmatrix}.$$

(3) **矩阵乘法**. 假设 $\boldsymbol{A}$ 分块如上, 而矩阵 $\boldsymbol{B}$ 分块为

$$\boldsymbol{B}=\begin{pmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\cdots\cdots\boldsymbol{B}_{1t}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\cdots\cdots\boldsymbol{B}_{2t}\\ \vdots & \vdots \qquad\qquad \vdots\\ \boldsymbol{B}_{s1} & \boldsymbol{B}_{s2}\cdots\cdots\boldsymbol{B}_{st}\end{pmatrix},$$

并进一步假设, $\boldsymbol{A}_{ij}$ 是 $m_i\times l_j$ 矩阵 $(1\le i\le r,\ 1\le j\le s)$, $\boldsymbol{B}_{jk}$ 是 $l_j\times n_k$ 矩阵 $(1\le j\le s,\ 1\le k\le t)$, 即 $\boldsymbol{A}$ 分块时列的组数等于 $\boldsymbol{B}$ 分块时的行的组数, 且 $\boldsymbol{A}$ 的每个列组的列数等于 $\boldsymbol{B}$ 的相应行组的行数. 此时,

$$\boldsymbol{AB}=\begin{pmatrix}\boldsymbol{C}_{11} & \boldsymbol{C}_{12}\cdots\cdots\boldsymbol{C}_{1t}\\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22}\cdots\cdots\boldsymbol{C}_{2t}\\ \vdots & \vdots \qquad\qquad \vdots\\ \boldsymbol{C}_{r1} & \boldsymbol{C}_{r2}\cdots\cdots\boldsymbol{C}_{rt}\end{pmatrix},$$

其中 $\boldsymbol{C}_{ik}=\sum_{j=1}^{s}\boldsymbol{A}_{ij}\boldsymbol{B}_{jk}$ 是 $m_i\times n_k$ 矩阵. 这与将分块矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 中的每个块**视为**标量时矩阵乘法的规则具有相同的**形式**.

(4) **转置**. 假设 $\boldsymbol{A}$ 分块如上, 则

$$\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{A}_{11}^{\mathrm{T}} & \boldsymbol{A}_{21}^{\mathrm{T}} \cdots\cdots \boldsymbol{A}_{r1}^{\mathrm{T}} \\ \boldsymbol{A}_{12}^{\mathrm{T}} & \boldsymbol{A}_{22}^{\mathrm{T}} \cdots\cdots \boldsymbol{A}_{r2}^{\mathrm{T}} \\ \vdots & \vdots \qquad\qquad \vdots \\ \boldsymbol{A}_{1s}^{\mathrm{T}} & \boldsymbol{A}_{2s}^{\mathrm{T}} \cdots\cdots \boldsymbol{A}_{rs}^{\mathrm{T}} \end{pmatrix}$$

(5) **等等**.

注 4.42. 矩阵分块时的分组一般需要依照具体运算的要求, 对于矩阵乘法, 一般会尽可能地凑出零矩阵或者单位矩阵.

例 4.43. 对于

$$\boldsymbol{A} = \left(\begin{array}{cc:c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hdashline -1 & 2 & 1 \\ 1 & 1 & 0 \end{array}\right) = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \qquad \boldsymbol{B} = \left(\begin{array}{cc:cc} 1 & 0 & 3 & 2 \\ -1 & 2 & 0 & 1 \\ \hdashline 1 & 0 & 4 & 1 \end{array}\right) = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix},$$

我们有

$$\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} & \boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} \end{pmatrix},$$

其中

$$\boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} = \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 4 \\ 0 & 2 \end{pmatrix}$$

$$\boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} = \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & 3 \end{pmatrix}.$$

由此可知,

$$\boldsymbol{AB} = \left(\begin{array}{cc:cc} 1 & 0 & 3 & 2 \\ -1 & 2 & 0 & 1 \\ \hdashline -2 & 4 & 1 & 1 \\ 0 & 2 & 3 & 3 \end{array}\right).$$

例 4.44. 设有分块矩阵 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{O} & \boldsymbol{A}_{22} \end{pmatrix}$, 其中 $\boldsymbol{A}_{11} \in F^{p\times p}$, $\boldsymbol{A}_{22} \in F^{q\times q}$. 证明: $\boldsymbol{A}$ 可逆的充要条件是 $\boldsymbol{A}_{11}$ 和 $\boldsymbol{A}_{22}$ 都是可逆的. 并在 $\boldsymbol{A}$ 可逆的条件下, 求出 $\boldsymbol{A}^{-1}$.

证明. 先设 $\boldsymbol{A}$ 有逆, 从而矩阵方程 $\boldsymbol{AX} = \boldsymbol{I}_{p+q}$ 有解. 我们可以对 $\boldsymbol{X}$ 做相同的分块: $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22} \end{pmatrix}$, 其中 $\boldsymbol{X}_{11} \in F^{p\times p}$, $\boldsymbol{X}_{22} \in F^{q\times q}$. 于是方程化为

$$\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{O} & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_{11} & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}_q \end{pmatrix},$$

即

$$\begin{aligned} \boldsymbol{A}_{11}\boldsymbol{X}_{11} + \boldsymbol{A}_{12}\boldsymbol{X}_{21} &= \boldsymbol{I}_p, \\ \boldsymbol{A}_{11}\boldsymbol{X}_{12} + \boldsymbol{A}_{12}\boldsymbol{X}_{22} &= \boldsymbol{O}, \\ \boldsymbol{A}_{22}\boldsymbol{X}_{21} &= \boldsymbol{O}, \\ \boldsymbol{A}_{22}\boldsymbol{X}_{22} &= \boldsymbol{I}_q. \end{aligned}$$

由此不难依次看出, $\boldsymbol{A}_{22}$ 可逆, 且 $\boldsymbol{X}_{22} = \boldsymbol{A}_{22}^{-1}$; $\boldsymbol{X}_{21} = \boldsymbol{O}$; $\boldsymbol{A}_{11}$ 可逆, 且 $\boldsymbol{X}_{11} = \boldsymbol{A}_{11}^{-1}$; $\boldsymbol{X}_{12} = -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$. 从而 $\boldsymbol{A}$ 的逆 $\boldsymbol{X}$ 具有指定的形式.

反之, 若 $\boldsymbol{A}_{11}$ 和 $\boldsymbol{A}_{22}$ 可逆, 则上面给出的 $\boldsymbol{X}$ 显然满足 $\boldsymbol{AX} = \boldsymbol{I}$, 从而 $\boldsymbol{A}$ 可逆. □

习题 4.45. 对于分块矩阵 $\boldsymbol{A} = (\boldsymbol{A}_1 \ \boldsymbol{A}_2) \in F^{m\times(m+n)}$, 若方阵 $\boldsymbol{A}_1$ 可逆, 求出所有的 $\boldsymbol{B} \in F^{(m+n)\times m}$ 使得 $\boldsymbol{AB} = \boldsymbol{I}_m$.

4.3 行列式

行列式的定义 每一个 n 阶方阵 $\boldsymbol{A}$ 都可以和一个称为矩阵的的行列式的标量相对应. 这个数值是否为零, 将告诉我们该矩阵是否为可逆矩阵. 另外, 行列式的几何意义是这个矩阵的 n 个列向量 (或等价地, 它的 n 个行向量) 在 n 维空间中张成的平行多面体的有向体积.

对于方阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$, 它的**行列式** (determinant) 将被记作 $\det(\boldsymbol{A})$ 或 $|\boldsymbol{A}|$ 或

$$\begin{vmatrix} a_{11} & a_{12}\cdots\cdots a_{1n} \\ a_{21} & a_{22}\cdots\cdots a_{2n} \\ \vdots & \vdots \qquad\quad \vdots \\ a_{n1} & a_{n2}\cdots\cdots a_{nn} \end{vmatrix}.$$

行列式的概念是由解方程组而引入的.

(1) 考虑方程 $ax=b$, 显然当 $a\neq 0$ 时, 方程有唯一解 $x=\frac{b}{a}$. 由此, 对于 1×1 的矩阵 $A=a$ (标量), 我们定义 $\det(\boldsymbol{A})=a$. 此时为了避免混淆, 我们一般不写成 $|a|$, 但是允许写成 $|\boldsymbol{A}|$.

(2) 考虑含两个未知元 x_1 与 x_2 的两个线性方程构成的方程组:

$$\begin{cases} a_{11}x_1+a_{12}x_2=b_1, \\ a_{21}x_1+a_{22}x_2=b_2. \end{cases}$$

当 $a_{11}a_{22}-a_{12}a_{21}\neq 0$ 时, 不难看出该方程组有唯一的解:

$$x_1=\frac{b_1a_{22}-b_2a_{12}}{a_{11}a_{22}-a_{12}a_{21}}, \qquad x_2=\frac{a_{11}b_2-a_{21}b_1}{a_{11}a_{22}-a_{12}a_{21}}. \tag{4.2}$$

考虑其系数矩阵 $\boldsymbol{A}=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{21} \end{pmatrix}$. 对于这样的 2 阶方阵, 我们定义其**行列式**为

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22}-a_{12}a_{21}.$$

于是, (4.2) 中的解可以用较为规律的形式表示出来:

$$x_1=\frac{\begin{vmatrix} \color{red}{b_1} & a_{12} \\ \color{red}{b_2} & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}, \qquad x_2=\frac{\begin{vmatrix} a_{11} & \color{red}{b_1} \\ a_{21} & \color{red}{b_2} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}. \tag{4.3}$$

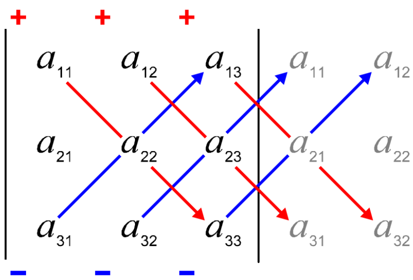
(3) 类似地, 我们可以考虑含三个未知元 x_1, x_2 与 x_3 的三个线性方程构成的方程组:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2, \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3. \end{cases} \tag{4.4}$$

考虑其系数矩阵 $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. 对于这样的 3 阶方阵, 我们定义其**行列式**为

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &:= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22}. \end{aligned}$$

下面的图 (来自维基百科) 在一定程度上可以帮助记忆, 方便直接写出三阶行列式的值.



在系数矩阵的行列式不为零的条件下, 不难验证 (留作练习), 线性方程组 (4.4) 有唯一解:

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \quad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}, \quad x_3 = \frac{\begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}} \tag{4.5}$$

注 4.46. 对于由含 n 个未定元的 n 个线性方程构成的方程组, 在其系数矩阵的行列式不为零的条件下, 我们可以证明其存在唯一的解, 而其解可以用类似于公式 (4.3) 和

(4.5) 的方法轻松写出; 这是克莱姆 (*Cramer*) 法则的主要内容. 在本节稍后的位置会学习到.

上面从 $n=2$ 到 $n=3$ 的讨论, 实际上反映了一种递归定义的方法. 下面来详细阐述.

对于方阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$, 若 $n>1$, 对于 $1\leq i,j\leq n$, 我们暂时记 M_{ij} 为由 $\boldsymbol{A}$ 删去包含 a_{ij} 所在的行和列后得到的 $(n-1)\times(n-1)$ 方阵的行列式 (由于我们在递归地考虑阶数为 n 的情形, 因此此时可以假定 $n-1$ 阶方阵的行列式已经有了恰当的定义). 例如, 在上面 $n=3$ 的情形中, 我们有

$$M_{11}=\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \qquad M_{12}=\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, \qquad M_{23}=\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}.$$

M_{ij} 将被称为行列式 $\det(\boldsymbol{A})$ 关于元素 a_{ij} (强调的是其位置 (i,j)) 的**余子式**, 而 $A_{ij}:=(-1)^{i+j}M_{ij}$ 将被称为行列式 $\det(\boldsymbol{A})$ 关于元素 a_{ij} 的**代数余子式**. 此时, 对于 $n\geq 2$, 我们递归地定义:

我们这儿的定义与教材上略有不同, 但是不用担心.

$$\det(\boldsymbol{A}):=\sum_{r=1}^{n}(-1)^{r+1}a_{r1}M_{r1}=\sum_{r=1}^{n}a_{r1}A_{r1}. \tag{4.6}$$

该定义中的求和式在之后将被称为**按第一列展开**来求行列式.

习题 4.47. 令 $\boldsymbol{u}=\begin{pmatrix} a \\ b \end{pmatrix}$, $\boldsymbol{v}=\begin{pmatrix} c \\ 0 \end{pmatrix}$, 其中为了讨论的方便, 可以假定 a,b,c 皆为正数. 利用初等的方法, 计算由 $\boldsymbol{u},\boldsymbol{v},\boldsymbol{u}+\boldsymbol{v}$ 和 $\boldsymbol{0}$ 为顶点构成的平行四边形的面积. 再利用定义分别计算 2 阶方阵 $(\boldsymbol{u}\ \boldsymbol{v})$ 和 $(\boldsymbol{v}\ \boldsymbol{u})$ 的行列式. 解释一下你的 "发现".

例 4.48. 设 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 为**上三角方阵**, 即对任意的 $1\leq j<i\leq n$, 有 $a_{ij}=0$. 我们来验证 $|\boldsymbol{A}|$ 为 $\boldsymbol{A}$ 的对角线元素的乘积. 这是因为

$$\begin{aligned}\det(\boldsymbol{A}) &= \sum_{k=1}^{n}(-1)^{k+1}a_{k1}M_{k1} \qquad (\text{其中对于 } k\geq 2,\ \text{有 } a_{k1}=0) \\ &= a_{11}M_{11}=a_{11}\underbrace{\begin{vmatrix} a_{22} & \cdots & a_{2n} \\ 0 & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots\ 0 & a_{nn} \end{vmatrix}}_{\text{仍为上三角方阵的行列式}} \overset{\text{归纳法}}{=\!=\!=} a_{11}(a_{22}a_{33}\cdots a_{nn}).\end{aligned}$$

行列式的常用基本性质 这儿, 我们先简单地列举一些行列式的常用性质. 其证明留到稍后的位置再展开.

定理 4.49. *n 阶方阵 $\boldsymbol{A}$ 的行列式可以按任意的行或列展开来计算:*

$$\det(\boldsymbol{A}) \xlongequal{\text{按第 } i \text{ 行展开}} \sum_{k=1}^{n} a_{ik}A_{ik} \xlongequal{\text{按第 } j \text{ 列展开}} \sum_{k=1}^{n} a_{kj}A_{kj}.$$

定理 4.50 (行列式的基本性质)**.** 教材定理 4.3.2

(1) 交换方阵 $\boldsymbol{A}$ 的两行得到方阵 $\boldsymbol{B}$, 则 $\det(\boldsymbol{B}) = -\det(\boldsymbol{A})$.

(2) 将方阵 $\boldsymbol{A}$ 的第 k 行乘以标量 λ 后得到方阵 $\boldsymbol{B}$, 则 $\det(\boldsymbol{B}) = \lambda\det(\boldsymbol{A})$. 特别地, $\det(\lambda\boldsymbol{A}) = \lambda^n\det(\boldsymbol{A})$.

(3) 若 $\boldsymbol{A}$ 的某一个行向量是两个向量的和, 则 $\det(\boldsymbol{A})$ 可以拆成两个行列式之和. 例如
$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a'_{21}+a''_{21} & a'_{22}+a''_{22} & a'_{23}+a''_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a''_{21} & a''_{22} & a''_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

(4) 若方阵 $\boldsymbol{A}$ 的两行相同, 则 $\det(\boldsymbol{A}) = 0$. 等价地, 若方阵 $\boldsymbol{A}$ 的两行成比例, 则 $\det(\boldsymbol{A}) = 0$.

(5) 将方阵 $\boldsymbol{A}$ 的某一行的常数倍加到该矩阵的另一行后得到方阵 $\boldsymbol{B}$, 则 $\det(\boldsymbol{B}) = \det(\boldsymbol{A})$.

定理 4.51. *转置运算不改变方阵的行列式: $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$.* 教材定理 4.3.4

推论 4.52. *定理 4.50 中的行操作也可以改为相应的列操作.*

定理 4.53. *设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为同阶的方阵, 则 $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$. 特别地, $\det(\boldsymbol{AB}) = \det(\boldsymbol{BA})$.* 教材定理 4.3.6

行列式的计算 关于一般行列式的计算, 我们最基本的技巧是用 “高斯消元法” 将方阵化为上三角或下三角的形式.

这儿需要非常熟悉三种行 (列) 初等变换操作带来的行列式的变化.

$$r_i \leftrightarrow r_j, \qquad \lambda r_i, \qquad \lambda r_i \to r_j,$$
$$c_i \leftrightarrow c_j, \qquad \lambda c_i, \qquad \lambda c_i \to c_j.$$

例 4.54. 求解行列式

$$D = \begin{vmatrix} -1 & 1 & 1 & 1 \\ 2 & -1 & 3 & -2 \\ 4 & 1 & -1 & -2 \\ -3 & 1 & 5 & -1 \end{vmatrix}.$$

解. 这儿可以用 $(1,1)$ 位置的 -1 消去第一行或第一列的其它元素, 但是不要混合行与列的操作.

$$D \xlongequal[\substack{c_1 \to c_3 \\ c_1 \to c_4}]{c_1 \to c_2} \begin{vmatrix} -1 & 0 & 0 & 0 \\ 2 & 1 & 5 & 0 \\ 4 & 5 & 3 & 2 \\ -3 & -2 & 2 & -4 \end{vmatrix} \xlongequal{\text{按第 1 行来展开}} (-1) \begin{vmatrix} 1 & 5 & 0 \\ 5 & 3 & 2 \\ -2 & 2 & -4 \end{vmatrix}$$

$$\xlongequal[\text{或直接展开}]{\text{可以对该 3 阶行列式继续作类似操作}} (-1) \cdot 64 = -64.$$

□

指明教材中例题 4.3.10 中的行列式为

$$M_{kn} = \left|\begin{array}{cccc:cccc} x & & & & & & & \\ -1 & \ddots & & & & & & \\ & \ddots & \ddots & & & & & \\ & & -1 & x & & & & \\ \hdashline & & & & -1 & x & & \\ & & & & & \ddots & \ddots & \\ & & & & & & \ddots & x \\ & & & & & & & -1 \end{array}\right|,$$

其左上角为 $k-1$ 阶方阵, 右下角为 $n-k$ 阶方阵.

学生课后自学教材 §4.3.3 中的例题, 其中例 4.3.11 中关于 Vandermonde 行列式的公式必须牢记.[*]

---

[*]克莱因所著的《古今数学思想》里介绍: 范德蒙德是第一个对行列式理论做出连贯的逻辑的阐述 (即把行列式理论与线性方程组求解相分离) 的人, 虽然他也把它应用于解线性方程组. 他还给出了一条法则, 用二阶子式和它们的余子式来展开行列式. 从集中到对行列式本身进行研究这一点来说，他是这门理论的奠基人.

习题 4.55. 设 $p_i(x)$ 是关于变元 x 的 i 次多项式, 其 x^i 前的系数为 c_i. 证明:

$$\begin{vmatrix} p_0(x_1) & p_1(x_1) \cdots\cdots p_{n-1}(x_1) \\ p_0(x_2) & p_1(x_2) \cdots\cdots p_{n-1}(x_2) \\ \vdots & \vdots \qquad\qquad \vdots \\ p_0(x_n) & p_1(x_n) \cdots\cdots p_{n-1}(x_n) \end{vmatrix} = c_0c_1\cdots c_{n-1}\prod_{1\le i<j\le n}(x_j-x_i).$$

例 4.56. 设 $A=\begin{vmatrix} 1 & -5 & 1 & 3 \\ 1 & 1 & 3 & 4 \\ 1 & 1 & 2 & 3 \\ 2 & 2 & 3 & 4 \end{vmatrix}$, 而 M_{ij} 是其中 (i,j) 元素所对应的余子式, 求 $-M_{41}+M_{42}-M_{43}+M_{44}$.

解. **(解法一)** 可以直接计算 $-M_{41}$, M_{42}, M_{43}, M_{44}.

(解法二) 若 A_{ij} 为对应的代数余子式, 则

$$\text{所求} = A_{41}+A_{42}+A_{43}+A_{44} = 1\cdot A_{41}+1\cdot A_{42}+1\cdot A_{43}+1\cdot A_{44}$$

$$\xlongequal[\text{沿第四行}]{\text{反用行列式的展开}} \begin{vmatrix} 1 & -5 & 1 & 3 \\ 1 & 1 & 3 & 4 \\ 1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{vmatrix} \xlongequal[\substack{-r_4\to r_2 \\ -r_4\to r_3}]{-r_4\to r_1} \begin{vmatrix} 0 & -6 & 0 & 2 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

$$\xlongequal{\text{按第一列展开}} 1\cdot(-1)^{4+1}\cdot\begin{vmatrix} -6 & 0 & 2 \\ 0 & 2 & 3 \\ 0 & 1 & 2 \end{vmatrix} = (-1)(-6)\begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix} = 6\cdot 1 = 6.$$

□

例 4.57. 计算行列式

$$D_n=\begin{vmatrix} a_1+b_1 & a_1+b_2 \cdots\cdots a_1+b_n \\ a_2+b_1 & a_2+b_2 \cdots\cdots a_2+b_n \\ \vdots & \vdots \qquad\qquad \vdots \\ a_n+b_1 & a_n+b_2 \cdots\cdots a_n+b_n \end{vmatrix}.$$

解.

$$\text{原式} \xlongequal{\text{按第一列分开}} \begin{vmatrix} a_1 & a_1+b_2 \cdots\cdots a_1+b_n \\ a_2 & a_2+b_2 \cdots\cdots a_2+b_n \\ \vdots & \vdots \qquad\qquad \vdots \\ a_n & a_n+b_2 \cdots\cdots a_n+b_n \end{vmatrix} + \begin{vmatrix} b_1 & a_1+b_2 \cdots\cdots a_1+b_n \\ b_1 & a_2+b_2 \cdots\cdots a_2+b_n \\ \vdots & \vdots \qquad\qquad \vdots \\ b_1 & a_n+b_2 \cdots\cdots a_n+b_n \end{vmatrix}$$

$$
\xlongequal[\substack{\text{第二个行列式}\\ \text{第一列提出 } b_1}]{\substack{\text{第一个行列式}\\ -c_1\to c_2\\ -c_1\to c_3\\ \vdots\\ -c_1\to c_n}} \begin{vmatrix} a_1 & b_2 \cdots\cdots b_n \\ a_2 & b_2 \cdots\cdots b_n \\ \vdots & \vdots \quad\quad \vdots \\ a_n & b_2 \cdots\cdots b_n \end{vmatrix} + b_1 \begin{vmatrix} 1 & a_1+b_2 \cdots\cdots a_1+b_n \\ 1 & a_2+b_2 \cdots\cdots a_2+b_n \\ \vdots & \vdots \quad\quad \vdots \\ 1 & a_n+b_2 \cdots\cdots a_n+b_n \end{vmatrix}
$$

$$
\xlongequal[\substack{\text{第二个行列式}\\ (-b_2)c_1\to c_2\\ \vdots\\ (-b_n)c_1\to c_n}]{\substack{\text{第一个行列式}\\ \text{第 2 列提出 } b_2\\ \vdots\\ \text{第 } n \text{ 列提出 } b_n}} b_2b_3\cdots b_n \begin{vmatrix} a_1 & 1 \cdots\cdots 1 \\ a_2 & 1 \cdots\cdots 1 \\ \vdots & \vdots \quad\quad \vdots \\ a_n & 1 \cdots\cdots 1 \end{vmatrix} + b_1 \begin{vmatrix} 1 & a_1 \cdots\cdots a_1 \\ 1 & a_2 \cdots\cdots a_2 \\ \vdots & \vdots \quad\quad \vdots \\ 1 & a_n \cdots\cdots a_n \end{vmatrix}
$$

$$
\xlongequal[n\geq 3?]{\substack{\text{需要依据 } n \text{ 判断}\\ \text{是否有相同的列}}} \begin{cases} a_1+b_1, & \text{若 } n=1, \\ b_2\begin{vmatrix} a_1 & 1 \\ a_2 & 1 \end{vmatrix} + b_2\begin{vmatrix} 1 & a_1 \\ 1 & a_2 \end{vmatrix} = b_2a_1 - b_2a_2 + b_1a_2 - b_1a_1, & \text{若 } n=2, \\ 0, & \text{若 } n\geq 3. \end{cases} \qquad \square
$$

例 4.58 (利用加边来计算行列式)**.** 计算

$$
D_n = \begin{vmatrix} a_1+\lambda_1 & a_2 \cdots\cdots\cdots\cdots & a_n \\ a_1 & a_2+\lambda_2 \cdots\cdots\cdots & a_n \\ \vdots & \vdots \quad \ddots & \vdots \\ a_1 & a_2 \cdots\cdots\cdots & a_n+\lambda_n \end{vmatrix}_{n\times n},
$$

其中每个 $\lambda_i \neq 0$.

解. **(解法一)**

$$
D_n = \left|\begin{array}{c:cccc} 1 & a_1 & a_2 & \cdots\cdots & a_n \\ \hdashline 0 & a_1+\lambda_1 & a_2 & \cdots\cdots & a_n \\ 0 & a_1 & a_2+\lambda_2 & \cdots\cdots & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_1 & a_2 & \cdots\cdots & a_n+\lambda_n \end{array}\right|_{(n+1)\times(n+1)} \xlongequal{\substack{-r_1\to r_2\\ -r_1\to r_3\\ \vdots\\ 1\to r_{n+1}}} \begin{vmatrix} 1 & a_1 & a_2 \cdots\cdots a_n \\ -1 & \lambda_1 & \\ -1 & & \lambda_2 \\ \vdots & & & \ddots \\ -1 & & & & \lambda_n \end{vmatrix}
$$

$$
\xlongequal{\substack{-\frac{1}{\lambda_1}c_2\to c_1\\ \vdots\\ \frac{1}{\lambda_n}c_{n+1}\to c_1}} \begin{vmatrix} 1+\sum_{i=1}^n \frac{a_i}{\lambda_i} & a_1 \cdots\cdots a_n \\ 0 & \lambda_1 \\ \vdots & & \ddots \\ 0 & & & \lambda_n \end{vmatrix} \xlongequal{\text{上三角矩阵}} \lambda_1\lambda_2\cdots\lambda_n\left(1+\sum_{i=1}^n \frac{a_i}{\lambda_i}\right).
$$

若将上面的计算结果展开, 可以得到一个不显式含分母项的多项式. 它是 D_n 在允许 $\lambda_i = 0$ 时的计算结果; 这一点很容易直接验证.

(解法二)

$$D_n=\begin{vmatrix}1 & 0 & 0\cdots\cdots\cdots\cdots 0\\ \hline 1 & a_1+\lambda_1 & a_2\cdots\cdots\cdots\cdots a_n\\ 1 & a_1 & a_2+\lambda_2\cdots\cdots\cdots a_n\\ \vdots & \vdots & \vdots\quad\ddots\quad\vdots\\ 1 & a_1 & a_2\cdots\cdots\cdots a_n+\lambda_n\end{vmatrix}_{(n+1)\times(n+1)}\xlongequal[-a_nc_1\to c_{n+1}]{\substack{-a_1c_1\to c_2\\ -a_2c_1\to c_3\\ \vdots}}\cdots.\qquad\square$$

例 4.57 也可以用加边的方法帮助快速讨论.

习题 4.59. 计算

$$D_n=\begin{vmatrix}\lambda-a_1^2 & -a_1a_2\cdots\cdots-a_1a_n\\ -a_2a_1 & \lambda-a_2^2\cdots\cdots-a_2a_n\\ \vdots & \vdots\quad\ddots\quad\vdots\\ -a_na_1 & -a_na_2\cdots\cdots\lambda-a_n^2\end{vmatrix}.$$

例 4.60 (递归公式法)**.** 计算行列式

$$D_n = \begin{vmatrix} 3 & 2 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 2 \\ & & 1 & 3 \end{vmatrix}_{n\times n}$$

解. 按第一行展开, 我们得到递推公式

$$D_n = 3D_{n-1} - 2D_{n-2}, \qquad n \geq 3.$$

此时, 可以用 (高中竞赛常用的) 特征方程法来求解, 也可以直接计算,

$$D_n - 2D_{n-1} = D_{n-1} - 2D_{n-2} = \cdots = D_2 - 2D_1 = \begin{vmatrix} 3 & 2 \\ 1 & 3 \end{vmatrix} - 2\cdot 3 = 7 - 6 = 1.$$

故

$$D_n + 1 = 2(D_{n-1}+1) = 2^2(D_{n-2}+1) = \cdots = 2^{n-1}(D_1+1) = 2^{n-1}\cdot 4.$$

这说明 $D_n = 2^{n+1} - 1$. □

习题 4.61. 计算行列式

$$D_n = \begin{vmatrix} a+b & a & & \\ b & \ddots & \ddots & \\ & \ddots & \ddots & a \\ & & b & a+b \end{vmatrix}_{n\times n}.$$

例 4.62 (Vandermonde 型)**.**

$$D_{n+1} = \begin{vmatrix} a_1^n & a_1^{n-1}b_1 & a_1^{n-2}b_1^2 & \cdots\cdots & a_1 b_1^{n-1} & b_1^n \\ a_2^n & a_2^{n-1}b_2 & a_2^{n-2}b_2^2 & \cdots\cdots & a_2 b_2^{n-1} & b_2^n \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ a_{n+1}^n & a_{n+1}^{n-1}b_{n+1} & a_{n+1}^{n-2}b_{n+1}^2 & \cdots\cdots & a_{n+1} b_{n+1}^{n-1} & b_{n+1}^n \end{vmatrix}_{(n+1)\times(n+1)},$$

其中 a_i 全不为 0 或 b_j 全不为 0.

解. 不妨假定所有的 a_i 不为 0. 对于每个 i, 从第 i 行提出 a_i^n, 得到

$$
\begin{aligned}
D_{n+1} &= a_1^n a_2^n \cdots a_{n+1}^n \begin{vmatrix} 1 & b_1/a_1 & (b_1/a_1)^2 \cdots\cdots\cdots\cdots (b_1/a_1)^n \\ 1 & b_2/a_2 & (b_2/a_2)^2 \cdots\cdots\cdots\cdots (b_2/a_2)^n \\ \vdots & \vdots & \vdots \qquad\qquad\qquad \vdots \\ 1 & b_{n+1}/a_{n+1} & (b_{n+1}/a_{n+1})^2 \cdots\cdots (b_{n+1}/a_{n+1})^n \end{vmatrix} \\
&= a_1^n a_2^n \cdots a_{n+1}^n \prod_{1\le i<j\le n+1} (b_j/a_j - b_i/a_i) = \prod_{1\le i<j\le n+1} (a_i b_j - a_j b_i).
\end{aligned}
$$

最后的计算结果在 D_n 允许某些 $a_i = 0$ 和某些 $b_j = 0$ 时也成立: 这一点可以通过直接计算轻松验证. □

习题 4.63 (Cauchy 行列式)**.** 设 $a_1, \cdots, a_n$ 和 $b_1, \cdots, b_n$ 是一些复数, 满足对任意的 $i, j = 1, \cdots, n$ 都有 $a_i + b_j \neq 0$. 设 $\boldsymbol{C} = (c_{ij})$ 为 n 阶方阵, 其元素 $c_{ij} := \frac{1}{a_i + b_j}$. 证明:

$$
\det(\boldsymbol{C}) = \frac{\prod_{1\le i<j\le n}(a_i - a_j)(b_i - b_j)}{\prod_{1\le i,j\le n}(a_i + b_j)}.
$$

(提示: 第一步, 先将用最后一列来减前面 $n-1$ 列, 并通过提出各行各列的公因式来化简. 第二步, 用前面的 $n-1$ 行来减最后一列, 继续化简, 并用归纳法.)

习题 4.64. (1) 设 $\boldsymbol{A}$ 是一个 n 阶反对称的矩阵, 其中 n 为奇数. 证明: $\det(\boldsymbol{A}) = 0$.

(2) 设 $\boldsymbol{A}$ 是一个 n 阶反对称的矩阵, 其中 n 为偶数. 若给 $\boldsymbol{A}$ 的每个元素都加上相同的一个数 λ, 从而得到新矩阵 $\boldsymbol{B}$. 证明: $\det(\boldsymbol{A}) = \det(\boldsymbol{B})$. (提示: 若将 $\boldsymbol{B}$ 视作 $\boldsymbol{A}$ 的每个列向量都加上了一个新的列向量所得到的矩阵. 将其行列式按列全部拆开, 你会得到 2^n 个行列式. 接下来考虑 $n+1$ 阶反对称阵 $\begin{pmatrix} 0 & 1 \cdots\cdots 1 \\ -1 & \\ \vdots & \boldsymbol{A} \\ -1 & \end{pmatrix}$, 将其行列式按第一行展开.)

(3) 设 $\boldsymbol{A}$ 是一个 n 阶反对称的矩阵, 其中 n 为偶数, 并且 $\boldsymbol{A}$ 的对角线右上方的所有元素都是 1. 求 $\det(\boldsymbol{A})$.

习题 4.65. 设矩阵 $\boldsymbol{A}$ 形如

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ a_0 & a_1 & a_2 \cdots\cdots & a_{n-2} & a_{n-1} \end{pmatrix}_{n\times n}.$$

求 $\det(\lambda \boldsymbol{I}_n - \boldsymbol{A})$, 其中 $\lambda \in F$ 为标量.

习题 4.66. 设 $\boldsymbol{A} = (a_{ij})_{n\times n}$, 而 A_{ij} 是 $\boldsymbol{A}$ 相对于 (i,j) 位置的代数余子式. 证明:

$$\begin{vmatrix} a_{11} & \cdots\cdots & a_{1n} & x_1 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots\cdots & a_{nn} & x_n \\ y_1 & \cdots\cdots & y_n & 0 \end{vmatrix} = -\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j A_{ij}.$$

行列式的性质及其证明 在这一部分里, 我们将证明之前提到的关于行列式的一些性质, 并介绍更多的与之相关的内容.

回忆. 对于矩阵, 我们有初等变换操作:

- 三种初等行变换: $r_i \leftrightarrow r_j$, λr_i, $\lambda r_i \to r_j$;
- 三种初等列变换: $c_i \leftrightarrow c_j$, λc_i, $\lambda c_i \to c_j$.

接下来, 从单位矩阵出发, 我们引入相应的三种**初等矩阵**:

$$\boldsymbol{I}_n \xrightarrow{r_i\leftrightarrow r_j} \boldsymbol{S}_{ij} \xleftarrow{c_i\leftrightarrow c_j} \boldsymbol{I}_n$$

$$\boldsymbol{I}_n \xrightarrow{\lambda r_i} \boldsymbol{D}_i(\lambda) \xleftarrow{\lambda c_i} \boldsymbol{I}_n$$

$$\boldsymbol{I}_i \xrightarrow{\lambda r_j \to r_i} \boldsymbol{T}_{ij}(\lambda) \xleftarrow{\lambda c_i \to c_j} \boldsymbol{I}_n$$

由学生写出这三种初等矩阵的具体形式. 其中注意到 $\boldsymbol{T}_{ij}(\lambda)$ 中的 λ 出现在方阵的 (i,j) 位置上. 需要注意的是, 不同教材里对这些初等矩阵的记法是不一致的.

定理 4.67. 教材定理 4.4.1

(1) 矩阵的初等行变换对应于初等矩阵的**左乘**, 即, 对于矩阵 $\boldsymbol{A}$, $r_i \leftrightarrow r_j$ 等同于 $\boldsymbol{S}_{ij}\boldsymbol{A}$, λr_i 等同于 $\boldsymbol{D}_i(\lambda)\boldsymbol{A}$, $\lambda r_j \to r_i$ 等同于 $\boldsymbol{T}_{ij}(\lambda)\boldsymbol{A}$.

(2) 矩阵的初等列变换对应于初等矩阵的**右乘**, 即, 对于矩阵 $\boldsymbol{A}$, $c_i \leftrightarrow c_j$ 等同于 $\boldsymbol{A}\boldsymbol{S}_{ij}$, λc_i 等同于 $\boldsymbol{A}\boldsymbol{D}_i(\lambda)$, $\lambda c_i \to c_j$ 等同于 $\boldsymbol{A}\boldsymbol{T}_{ij}(\lambda)$.

证明. 直接验证. □

定理 4.68. 初等矩阵都是可逆矩阵, 且其逆矩阵仍然是初等矩阵, 并满足 教材定理 4.4.2

$$\boldsymbol{S}_{ij}^2 = \boldsymbol{D}_i(\lambda)\boldsymbol{D}_i(1/\lambda) = \boldsymbol{T}_{ij}(\lambda)\boldsymbol{T}_{ij}(-\lambda) = \boldsymbol{I}.$$

证明. 可以直接计算验证; 也可以用初等行变换的观点来看, 将这些初等矩阵通过左乘作用在单位阵 $\boldsymbol{I}$ 上. □

讨论. 给定一个 n 阶方阵 $\boldsymbol{A}$, 我们可以通过一系列初等行变换将其化为约化标准形; 我们将后者记作rref(A). rref(A) 的主元个数显然至多为 n 个.

(1) 若 rref(A) 恰有 n 个, 那么由约化标准形的定义可知, $\mathrm{rref}(A) = \boldsymbol{I}_n$.

(2) 若 rref(A) 的主元的个数小于 n, 那么 rref$(\boldsymbol{A})$ 的最后一行全为 0. 对任意的 n 阶方阵 $\boldsymbol{X}$, 容易看出矩阵乘法 $\mathrm{rref}(\boldsymbol{A})\boldsymbol{X}$ 的最后一行也全为 0, 从而不可能为 $\boldsymbol{I}_n$. 这说明此时, rref$(\boldsymbol{A})$ 不是可逆方阵.

定理 4.69. 设 $\boldsymbol{A}$ 是 n 阶方阵, 则以下三条等价:

(a) $\boldsymbol{A}$ 可以通过一系列初等行变换化为单位阵 $\boldsymbol{I}_n$;

(b) $\boldsymbol{A}$ 是一些初等矩阵的乘积;

(c) $\boldsymbol{A}$ 是可逆的.

证明. (a)⇒(b): 利用定理 4.67 中的等价, 我们可以得到一系列初等矩阵 $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_s$, 使得 $\boldsymbol{E}_1 \cdots \boldsymbol{E}_s \boldsymbol{A} = \mathrm{rref}(\boldsymbol{A}) = \boldsymbol{I}$. 此时, $\boldsymbol{A} = \boldsymbol{E}_s^{-1} \cdots \boldsymbol{E}_1^{-1}$, 是一些初等矩阵的乘积. 对 $\boldsymbol{A}$ 做的第一个初等行变化是由 $\boldsymbol{E}_s$ 给出的

(b)⇒(c): 因为初等矩阵都是可逆的, 而可逆阵的矩阵乘法仍然是可逆的.

(c)⇒(a): 由上面的讨论及定理 4.67 中的等价可知, 存在一系列初等矩阵 $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_s$, 使得 $\boldsymbol{E}_1 \cdots \boldsymbol{E}_s \boldsymbol{A} = \mathrm{rref}(\boldsymbol{A})$. 由于 $\boldsymbol{A}$ 是可逆的, 由此看出 rref$(\boldsymbol{A})$ 也是可逆的. 若 rref$(\boldsymbol{A})$ 的主元个数小于 n, 则 rref$(\boldsymbol{A})$ 不可逆. 这说明 rref$(\boldsymbol{A})$ 的主元个数为 n, 即 $\mathrm{rref}(\boldsymbol{A}) = \boldsymbol{I}_n$. □

注 4.70. 由于方阵 $\boldsymbol{A}$ 可逆当且仅当 $\boldsymbol{A}^{\mathrm{T}}$ 可逆, 故上面的几条也等价于说 $\boldsymbol{A}$ 可以通过一系列初等列变换化为单位阵 $\boldsymbol{I}_n$.

接下来继续讨论 n 阶方阵 $\boldsymbol{A}$ 的行列式 $\det(\boldsymbol{A})$. 由于 $\boldsymbol{A}$ 是它的 n 个行向量 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 的有序排列, 故 $\det(\boldsymbol{A})$ 也可以视作这些行向量的函数: $\det(\boldsymbol{A}) = \det(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$.

定理 4.71. *行列式函数满足以下的性质.*

(a) $\det(\boldsymbol{I}_n) = 1$.

(b) $\det(\boldsymbol{A})$ *关于* $\boldsymbol{A}$ *的每个行向量都是线性的 (在固定其它行向量的条件下): 若第* i_0 *个行向量* $\boldsymbol{\alpha}_{i_0} = k_1\boldsymbol{\beta} + k_2\boldsymbol{\gamma}$, *其中* $k_1, k_2 \in F$, *那么*

$$\begin{aligned}\det(\boldsymbol{A}) &= \det(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i_0-1}, \boldsymbol{\alpha}_{i_0}, \boldsymbol{\alpha}_{i_0+1}, \ldots, \boldsymbol{\alpha}_n) \\ &= k_1 \det(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i_0-1}, \boldsymbol{\beta}, \boldsymbol{\alpha}_{i_0+1}, \ldots, \boldsymbol{\alpha}_n) + k_2 \det(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i_0-1}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{i_0+1}, \ldots, \boldsymbol{\alpha}_n).\end{aligned}$$

特别地, 若 $\boldsymbol{A}$ *有一行全为* 0, *则* $\det(\boldsymbol{A}) = 0$.

(c) *若* $\boldsymbol{A}$ *有相邻的两行相等, 则* $\det(\boldsymbol{A}) = 0$.

证明. 不难看出, 这些结果当 $n = 1$ 时是成立的. 以下不妨设 $n \geq 2$.

(a) 这是例 4.48 的一个特例.

(b) 设将 $\boldsymbol{A} = (a_{ij})$ 的第 i_0 个行向量分别换为 $\boldsymbol{\beta}$ 和 $\boldsymbol{\gamma}$ 后得到的矩阵为 $\boldsymbol{B} = (b_{ij})$ 与 $\boldsymbol{C} = (c_{ij})$, 并记相应的代数余子式分别为 B_{ij} 与 C_{ij}. 从行列式的归纳定义 (4.6) 出发, 我们只需证明对任意的 $r = 1, 2, \ldots, n$, 我们都有

$$a_{r1}A_{r1} = k_1 b_{r1} B_{r1} + k_2 c_{r1} C_{r1}, \tag{†}$$

其中 A_{r1} 是 $\boldsymbol{A}$ 相应位置的代数余子式.

- 设 $r = i_0$. 此时, 我们所处理的行不在这些子式里, 这说明 $A_{r1} = B_{r1} = C_{r1}$. 另一方面, $a_{r1} = k_1 b_{r1} + k_2 c_{r1}$. 从而等式 (†) 得证.
- 设 $r \neq i_0$. 此时, 我们所处理的行在这些 $(n-1)$ 阶子式里, 由归纳法可知 $A_{r1} = k_1 B_{r1} + k_2 C_{r1}$. 另一方面, $a_{r1} = b_{r1} = c_{r1}$. 从而等式 (†) 得证.

特别地, 当 $\boldsymbol{A}$ 的某一行 (比如说 $\boldsymbol{\alpha}_{i_0}$) 为零时, 我们有 $\boldsymbol{\alpha}_{i_0} = 0\boldsymbol{\beta} + 0\boldsymbol{\gamma}$, 此时的公式告诉我们

$$\det(\boldsymbol{A}) = 0 \det(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i_0-1}, \boldsymbol{\beta}, \boldsymbol{\alpha}_{i_0+1}, \ldots, \boldsymbol{\alpha}_n) + 0 \det(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i_0-1}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{i_0+1}, \ldots, \boldsymbol{\alpha}_n) = 0.$$

(c) 设 $\boldsymbol{A}$ 的第 k 个行向量等于其第 $k+1$ 个行向量. 那么对于 $r=1,2,\ldots,n$, 若 $r\neq k,k+1$, 则代数余子式 A_{r1} 中有两行相等, 从而由归纳法可知为零. 于是, 从行列式的递归定义 (4.6) 出发, 我们有

$$\det(\boldsymbol{A})=a_{k,1}A_{k,1}+a_{k+1,1}A_{k+1,1}.$$

由于 $a_{k,1}=a_{k+1,1}$, 而 $A_{k,1}$ 与 $A_{k+1,1}$ 恰好相差一个正负号, 上式为 0. □

定理 4.72. *行列式函数满足以下的性质.*

(a) *设* $\boldsymbol{B}$ *由方阵* $\boldsymbol{A}$ *通过倍乘行的初等行变换得到:* $\boldsymbol{A}\xrightarrow{\lambda r_i}\boldsymbol{B}$, *即* $\boldsymbol{B}=\boldsymbol{D}_i(\lambda)\boldsymbol{A}$, *则* $\det(\boldsymbol{B})=\lambda\det(\boldsymbol{A})$.

(b) *设* $\boldsymbol{B}$ *由方阵* $\boldsymbol{A}$ *通过倍加的初等行变换得到:* $\boldsymbol{A}\xrightarrow{\lambda r_j\to r_i}\boldsymbol{B}$, *即* $\boldsymbol{B}=\boldsymbol{T}_{ij}(\lambda)\boldsymbol{A}$, *则* $\det(\boldsymbol{B})=\det(\boldsymbol{A})$.

(c) *设* $\boldsymbol{B}$ *由方阵* $\boldsymbol{A}$ *通过交换行的初等行变换得到:* $\boldsymbol{A}\xrightarrow{r_j\leftrightarrow r_i}\boldsymbol{B}$, *即* $\boldsymbol{B}=\boldsymbol{S}_{ij}\boldsymbol{A}$, *则* $\det(\boldsymbol{B})=-\det(\boldsymbol{A})$.

(d) *若* $\boldsymbol{A}$ *的不同两行* $\boldsymbol{\alpha}_i$ *与* $\boldsymbol{\alpha}_j$ *对应成比例, 则* $\det(\boldsymbol{A})=0$.

证明. (a) 可以从定理 4.71的(b). 接下来只需验证 (b), (c) 和 (d).

先考虑下标 i,j 是相邻的情形, 例如 $j=i+1$. 对于 (b), 我们有

$$\begin{aligned}\det(\boldsymbol{B})&=\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\boldsymbol{\alpha}_i+\lambda\boldsymbol{\alpha}_{i+1},\boldsymbol{\alpha}_{i+1},\ldots,\boldsymbol{\alpha}_n)\\&\xlongequal{\text{定理 4.71 (b)}}\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+1},\ldots,\boldsymbol{\alpha}_n)+\lambda\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\underbrace{\boldsymbol{\alpha}_{i+1},\boldsymbol{\alpha}_{i+1}}_{\text{相等的相邻行}},\ldots,\boldsymbol{\alpha}_n)\\&\xlongequal{\text{定理 4.71 (c)}}\det(\boldsymbol{A})+\lambda 0=\det(\boldsymbol{A}).\end{aligned}$$

该证明显然也适用于 $j=i-1$ 的情形. 为了证明 (c), 我们反复运用 (b):

$$\begin{aligned}\det(\boldsymbol{B})&=\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\boldsymbol{\alpha}_{i+1},\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+2},\ldots,\boldsymbol{\alpha}_n)\\&\xlongequal{-r_{i+1}\to r_i}\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\boldsymbol{\alpha}_{i+1}-\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+2},\ldots,\boldsymbol{\alpha}_n)\\&\xlongequal{r_i\to r_{i+1}}\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\boldsymbol{\alpha}_{i+1}-\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i+(\boldsymbol{\alpha}_{i+1}-\boldsymbol{\alpha}_i),\boldsymbol{\alpha}_{i+2},\ldots,\boldsymbol{\alpha}_n)\\&=\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\boldsymbol{\alpha}_{i+1}-\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+1},\boldsymbol{\alpha}_{i+2},\ldots,\boldsymbol{\alpha}_n)\\&\xlongequal{-r_{i+1}\to r_i}\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},-\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+1},\boldsymbol{\alpha}_{i+2},\ldots,\boldsymbol{\alpha}_n)\\&\xlongequal{\text{定理 4.71 (b)}}-\det(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_{i-1},\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+1},\boldsymbol{\alpha}_{i+2},\ldots,\boldsymbol{\alpha}_n)\\&=-\det(\boldsymbol{A}).\end{aligned}$$

为了证明 (d), 我们只需运用 (a) 和 定理 4.71 (c).

接下来对于一般的指标 $i \neq j$, 我们来考虑(b), (c) 和(d). 为此, 我们反复进行 "交换相邻两行" 的操作, 再利用上面的讨论即可. 我们给出几个简单 4 阶方阵的例子来说明, 具体的证明留给同学们自己来补充.

- 在进行了 $4r_1 \to r_4$ 后, $\det(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \lambda\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_4) = -\det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_3, \lambda\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_4) = \det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_1, \lambda\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_4) = \det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_4) = -\det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4) = \det(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)$. 更一般地, 若要考虑第 i 行和第 j 行的倍加行变换, 可以通过 $2|j-i|-2$ 次交换相邻行以及一个相邻行的倍加行变换的方法来实现.

- 在进行了 $r_1 \leftrightarrow r_4$ 后, $\det(\boldsymbol{\alpha}_4, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_1) = -\det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_4, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_1) = \det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4, \boldsymbol{\alpha}_1) = -\det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_4) = \det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4) = -\det(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)$. 更一般地, 若要交换第 i 行和第 j 行, 可以通过 $2|j-i|-1$ 次交换相邻行的方法来实现.

- 若 $\boldsymbol{\alpha}_4 = \lambda\boldsymbol{\alpha}_1$, $\det(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \lambda\boldsymbol{\alpha}_1) = -\det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_3, \lambda\boldsymbol{\alpha}_1) = \det(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_1, \lambda\boldsymbol{\alpha}_1) = 0$. 更一般地, 若第 i 行和第 j 行成比例, 可以先通过 $|j-i|-1$ 个交换相邻行的方法, 将指定的两个行向量置于新矩阵的相邻行位置, 从而得到其行列式为 0 的事实. □

推论 4.73. (1) $\det(\boldsymbol{S}_{ij}) = -1$, $\det(\boldsymbol{D}_i(\lambda)) = \lambda$, $\det(\boldsymbol{T}_{ij}(\lambda)) = 1$.

(2) *若 $\boldsymbol{E}$ 是一个初等矩阵, 则* $\det(\boldsymbol{E}\boldsymbol{A}) = \det(\boldsymbol{E})\det(\boldsymbol{A})$.

证明. (1) 由于初等矩阵可以从 $\boldsymbol{I}_n$ 出发利用初等行变换得到, 利用定理 4.67 (1), 定理 4.71 (a), 和定理 4.72 的 (a), (b) 和 (c) 即可.

(2) 同样地, 利用定理 4.67 (1), 以及定理 4.72 的 (a), (b) 和 (c), 还有上面的计算结果. □

定理 4.74. *设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为同阶的方阵, 则* $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$. *特别地,* $\det(\boldsymbol{AB}) = \det(\boldsymbol{BA})$.

证明. 若 $\boldsymbol{AB}$ 可逆, 在注 4.26 中我们已经见到, $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 同时可逆. 由定理 4.69 可知, 存在初等方阵 $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_r$ 和 $\boldsymbol{E}'_1, \ldots, \boldsymbol{E}'_s$ 使得 $\boldsymbol{A} = \boldsymbol{E}_1 \cdots \boldsymbol{E}_r$, 而 $\boldsymbol{B} = \boldsymbol{E}'_1 \cdots \boldsymbol{E}'_s$. 此时, 反复运用推论 4.73 (2), 我们有

$$\det(\boldsymbol{A}) = \det(\boldsymbol{E}_1 \cdots \boldsymbol{E}_r) = \det(\boldsymbol{E}_1)\det(\boldsymbol{E}_2 \cdots \boldsymbol{E}_r) = \cdots = \det(\boldsymbol{E}_1)\cdots\det(\boldsymbol{E}_r),$$

以及

$$\det(\boldsymbol{B}) = \det(\boldsymbol{E}'_1 \cdots \boldsymbol{E}'_s),$$

和

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{E}_1 \cdots \boldsymbol{E}_r \boldsymbol{E}_1' \cdots \boldsymbol{E}_s') = \det(\boldsymbol{E}_1) \cdots \det(\boldsymbol{E}_r) \det(\boldsymbol{E}_1') \cdots \det(\boldsymbol{E}_s').$$

由此可以迅速看出, $\det(\boldsymbol{AB}) = \det(\boldsymbol{A}) \det(\boldsymbol{B})$.

另一方面, 若 $\boldsymbol{AB}$ 不可逆, 则 $\boldsymbol{A}$ 或 $\boldsymbol{B}$ 不可逆. 接下来只需验证 "若 $\boldsymbol{A}$ 不可逆, 则 $\det(\boldsymbol{A}) = 0$". 对于该断言, 若 $\boldsymbol{A}$ 不可逆, 由定理 4.69可知, $\mathrm{rref}(\boldsymbol{A})$ 不是单位阵, 即它的最后一行全为 0. 此时, 我们可以找到初等矩阵 $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_r$, 使得 $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_r \boldsymbol{A} = \mathrm{rref}(\boldsymbol{A})$. 于是由定理 4.71 (b) 和推论 4.73(2) 可知,

$$0 = \det(\mathrm{rref}(\boldsymbol{A})) = \det(\boldsymbol{E}_1 \cdots \boldsymbol{E}_r \boldsymbol{A}) = \det(\boldsymbol{E}_1) \cdots \det(\boldsymbol{E}_r) \det(\boldsymbol{A}).$$

由于初等方阵的行列式非零, 这迫使 $\det(\boldsymbol{A}) = 0$. $\square$

推论 4.75. (a) 方阵 $\boldsymbol{A}$ 可逆的充要条件是 $\det(\boldsymbol{A}) \neq 0$. 若 $\boldsymbol{A}$ 可逆, 则 $\det(\boldsymbol{A}^{-1}) = \det(\boldsymbol{A})^{-1}$.

(b) 对于同阶方阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$, 我们有 $\boldsymbol{AB} = \boldsymbol{I} \Leftrightarrow \boldsymbol{BA} = \boldsymbol{I}$.

(c) $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$.

(d) 定理 *4.71* 和定理 *4.72* 中的性质对于列向量也成立. 另外, 行列式可以按照任意行展开, 也可以按照任意列展开.

证明. (a) 若 $\boldsymbol{A}$ 可逆, 则存在 $\boldsymbol{A}^{-1}$, 此时 $\det(\boldsymbol{A})\det(\boldsymbol{A}^{-1}) = \det(\boldsymbol{AA}^{-1}) = \det(\boldsymbol{I}) = 1$. 这说明 $\det(\boldsymbol{A}) \neq 0$. 若 $\boldsymbol{A}$ 不可逆, 定理 4.74 的证明里我们已经证得了 $\det(\boldsymbol{A}) = 0$.

(b) 由对称性, 我们只需证明 $\boldsymbol{AB} = \boldsymbol{I} \Rightarrow \boldsymbol{BA} = \boldsymbol{I}$. 若 $\boldsymbol{AB} = \boldsymbol{I}$, 则 $\det(\boldsymbol{A})\det(\boldsymbol{B}) = \det(\boldsymbol{I}) = 1$, 故 $\det(\boldsymbol{A}) \neq 0$, 从而 $\boldsymbol{A}$ 可逆. 此时, 对 $\boldsymbol{AB} = \boldsymbol{I}$ 的两边同时左乘 $\boldsymbol{A}^{-1}$, 可得 $\boldsymbol{B} = \boldsymbol{A}^{-1}$, 从而亦有 $\boldsymbol{BA} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

(c) 对于任意的初等方阵 $\boldsymbol{E}$, 我们容易看出其转置矩阵为同类型初等方阵:

$$\boldsymbol{S}_{ij}^{\mathrm{T}} = \boldsymbol{S}_{ij}, \qquad \boldsymbol{D}_i(\lambda)^{\mathrm{T}} = \boldsymbol{D}_i(\lambda), \qquad \boldsymbol{T}_{ij}(\lambda)^{\mathrm{T}} = \boldsymbol{T}_{ji}(\lambda).$$

从而, 由推论 4.73(1)可知, 我们总有 $\det(\boldsymbol{E}) = \det(\boldsymbol{E}^{\mathrm{T}})$. 接下来对于 $\boldsymbol{A}$, 若 $\boldsymbol{A}$ 可逆, 我们可以找到一列初等矩阵 $\boldsymbol{E}_1 \cdots \boldsymbol{E}_r$, 使得 $\boldsymbol{A} = \boldsymbol{E}_1, \dots, \boldsymbol{E}_r$. 此时, $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{E}_r^{\mathrm{T}} \cdots \boldsymbol{E}_1^{\mathrm{T}}$. 利用定理 4.74 或者推论 4.73 (2), 我们不难验证 $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$.

若 $\boldsymbol{A}$ 不可逆, 由定理 4.31 (5) 可知, $\boldsymbol{A}^{\mathrm{T}}$ 也不可逆, 从而 $\det(\boldsymbol{A}) = 0 = \det(\boldsymbol{A}^{\mathrm{T}})$.

(d) 第一句话中的论断利用前面一条性质即可. 关于第二句话中的论断, 我们引入的行列式的定义是说行列式可以按照第一列展开, 利用转置运算, 这说明行列式可以按照第一行展开; 再通过交换行的方法, 可以看出行列式可以通过任意行展开; 最后, 再次利用转置运算, 可以看出行列式可以按照任意列展开. □

对于方阵 $\boldsymbol{A} = (a_{ij})_{n\times n}$, 从中取出第 $i_1 < i_2 < \cdots < i_p$ 行与 $j_1 < j_2 < \cdots < j_p$ 列的交叉位置的元素. 这些有序排列的元素构成了 $\boldsymbol{A}$ 的一个 p **阶子阵**:

$$\boldsymbol{A}\begin{pmatrix} i_1 & i_2 & \cdots & i_p \\ j_1 & j_2 & \cdots & j_p \end{pmatrix} := \begin{pmatrix} a_{i_1j_1} & a_{i_1j_2} \cdots\cdots & a_{i_1j_p} \\ a_{i_2j_1} & a_{i_2j_2} \cdots\cdots & a_{i_2j_p} \\ \vdots & \vdots & \vdots \\ a_{i_pj_1} & a_{i_pj_2} \cdots\cdots & a_{i_pj_p} \end{pmatrix}.$$

之前提到, 行列式可以按照任何一行展开来求解. 这其实可视作下面结果的一个特例.

注 4.76 (Laplace 展开定理*). 若取定行指标 $i_1 < i_2 < \cdots < i_p$, 则

$$\det(\boldsymbol{A}) = \sum_{j_1<j_2<\cdots<j_p} \det\left(\boldsymbol{A}\left(\begin{smallmatrix} i_1 & i_2 & \cdots & i_p \\ j_1 & j_2 & \cdots & j_p \end{smallmatrix}\right)\right)\left((-1)^{i_1+i_2+\cdots+i_p+j_1+j_2+\cdots+j_p} \det\left(\boldsymbol{A}\left(\begin{smallmatrix} i_{p+1} & i_{p+2} & \cdots & i_n \\ j_{p+1} & j_{p+2} & \cdots & j_n \end{smallmatrix}\right)\right)\right),$$

其中

$$\begin{aligned}\{1,2,\ldots,n\} &= \{i_1 < i_2 < \cdots < i_p\} \cup \{i_{p+1} < i_{p+1} < \cdots < i_n\} \\ &= \{j_1 < j_2 < \cdots < j_p\} \cup \{j_{p+1} < j_{p+1} < \cdots < j_n\}.\end{aligned}$$

我们不给出该公式的证明. 感兴趣的同学可以自己研究.

例如, 对于

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 1 & 2 & -1 & -1 \\ 0 & 1 & -2 & 3 \end{pmatrix},$$

若我们选定 $i_1 = 1$, $i_2 = 2$, 则需要计算以下行列式:

$$\det\left(\boldsymbol{A}\left(\begin{smallmatrix} 1 & 2 \\ 1 & 2 \end{smallmatrix}\right)\right) = \begin{vmatrix} 1 & -1 \\ -1 & 0 \end{vmatrix} = -1, \quad (-1)^{(1+2)+(1+2)} \det\left(\boldsymbol{A}\left(\begin{smallmatrix} 3 & 4 \\ 3 & 4 \end{smallmatrix}\right)\right) = (-1)^{(1+2)+(1+2)} \begin{vmatrix} -1 & -1 \\ -2 & 3 \end{vmatrix} = -5,$$

$$\det\left(\boldsymbol{A}\left(\begin{smallmatrix} 1 & 2 \\ 1 & 3 \end{smallmatrix}\right)\right) = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1, \quad (-1)^{(1+2)+(1+3)} \det\left(\boldsymbol{A}\left(\begin{smallmatrix} 3 & 4 \\ 2 & 4 \end{smallmatrix}\right)\right) = (-1)^{(1+2)+(1+3)} \begin{vmatrix} 2 & -1 \\ 1 & 3 \end{vmatrix} = -7,$$

$$\det\left(\boldsymbol{A}\left(\begin{smallmatrix} 1 & 2 \\ 1 & 4 \end{smallmatrix}\right)\right) = \begin{vmatrix} 1 & 1 \\ -1 & 0 \end{vmatrix} = 1, \quad (-1)^{(1+2)+(1+4)} \det\left(\boldsymbol{A}\left(\begin{smallmatrix} 3 & 4 \\ 2 & 3 \end{smallmatrix}\right)\right) = (-1)^{(1+2)+(1+4)} \begin{vmatrix} 2 & -1 \\ 1 & -2 \end{vmatrix} = -3,$$

$$\det\left(\boldsymbol{A}\left(\begin{smallmatrix} 1 & 2 \\ 2 & 3 \end{smallmatrix}\right)\right) = \begin{vmatrix} -1 & 0 \\ 0 & 1 \end{vmatrix} = -1, \quad (-1)^{(1+2)+(2+3)} \det\left(\boldsymbol{A}\left(\begin{smallmatrix} 3 & 4 \\ 1 & 4 \end{smallmatrix}\right)\right) = (-1)^{(1+2)+(2+3)} \begin{vmatrix} 1 & -1 \\ 0 & 3 \end{vmatrix} = 3,$$

$$\det\left(\boldsymbol{A}\left(\begin{smallmatrix} 1 & 2 \\ 2 & 4 \end{smallmatrix}\right)\right) = \begin{vmatrix} -1 & 1 \\ 0 & 0 \end{vmatrix} = 0, \quad (-1)^{(1+2)+(2+4)} \det\left(\boldsymbol{A}\left(\begin{smallmatrix} 3 & 4 \\ 1 & 3 \end{smallmatrix}\right)\right) = (-1)^{(1+2)+(2+4)} \begin{vmatrix} 1 & -1 \\ 0 & -2 \end{vmatrix} = 2,$$

---

*克莱因所著的《古今数学思想》里介绍: 参照克莱姆和贝祖的工作, 拉普拉斯在 1772 年的论文《对积分和世界体系的探讨》(Recherches sur le calcul intégral et sur le système du monde) 中, 证明了范德蒙德的一些规则, 并推广了他的展开行列式的方法, 用 r 行中所含的子式和它们的余子式的集合来展开行列式, 这个方法现在仍然以他的名字命名.

$$\det\left(\boldsymbol{A}\left(\begin{smallmatrix}1&2\\3&4\end{smallmatrix}\right)\right)=\begin{vmatrix}0&1\\1&0\end{vmatrix}=-1,\quad (-1)^{(1+2)+(2+3)}\det\left(\boldsymbol{A}\left(\begin{smallmatrix}3&4\\1&2\end{smallmatrix}\right)\right)=(-1)^{(1+2)+(2+3)}\begin{vmatrix}1&2\\0&1\end{vmatrix}=1.$$

于是, 由 *Laplace* 展开定理可知

$$\det(\boldsymbol{A})=(-1)\cdot(-5)+1\cdot(-7)+1\cdot(-3)+(-1)\cdot 3+0\cdot 2+(-1)\cdot 1=-9.$$

习题 4.77. 在这里, 验证 *Laplace* 展开定理的一个特殊形式 (有困难的同学可以参考教材上的定理 *4.3.1*). 设 $\boldsymbol{A}$ 为 n 阶方阵, $1<k\le n$. 对任意的 $1\le i<j\le n$, 记 D_{ij} 为 $\boldsymbol{A}$ 删去第 $1,k$ 行, 并删去第 i,j 列后得到的 $n-2$ 阶方阵的行列式. 证明:

$$\det(\boldsymbol{A})=\sum_{1\le i<j\le n}(-1)^{1+k+i+j}\begin{vmatrix}a_{1i}&a_{1j}\\a_{ki}&a_{kj}\end{vmatrix}D_{ij}.\tag{4.7}$$

习题 4.78. 在这里, 再考虑 *Laplace* 展开定理的另外一种特殊形式. 事实上, *Laplace* 展开定理的 (复杂) 证明就是以讨论形如这样的结果开始的. 对于下面的分块矩阵, 证明相应的行列式公式, 其中 $\boldsymbol{A}\in F^{m\times m}$, $\boldsymbol{D}\in F^{n\times n}$, $\boldsymbol{C}\in F^{n\times m}$, $\boldsymbol{B}\in F^{m\times n}$.

(i) $\det\begin{pmatrix}\boldsymbol{A}&\boldsymbol{O}\\\boldsymbol{O}&\boldsymbol{I}_n\end{pmatrix}=\det(\boldsymbol{A})$. (ii) $\det\begin{pmatrix}\boldsymbol{I}_m&\boldsymbol{O}\\\boldsymbol{C}&\boldsymbol{D}\end{pmatrix}=\det(\boldsymbol{D})$.

(iii) $\det\begin{pmatrix}\boldsymbol{A}&\boldsymbol{O}\\\boldsymbol{C}&\boldsymbol{D}\end{pmatrix}=\det(\boldsymbol{A})\det(\boldsymbol{D})=\det\begin{pmatrix}\boldsymbol{A}&\boldsymbol{B}\\\boldsymbol{O}&\boldsymbol{D}\end{pmatrix}$.

提示: 在前两小问里分别对最后一行和第一行作展开, 在第三小问里可以利用前两小问的矩阵的乘法. 第三小问的公式要求熟记.

关于矩阵乘积的行列式 (参见定理 4.74), 我们还有如下的进一步推广.

注 4.79 (关于矩阵乘积的 **Binet–Cauchy 公式**). 对于矩阵 $\boldsymbol{A}$, 我们用 $\boldsymbol{A}\left(\begin{smallmatrix}i_1&i_2&\cdots&i_r\\j_1&j_2&\cdots&j_r\end{smallmatrix}\right)$ 来表示由 $\boldsymbol{A}$ 的第 $i_1<i_2<\cdots<i_r$ 行第 $j_1<j_2<\cdots<j_r$ 列元素构成的 r 阶子方阵. 假定 $\boldsymbol{A}\in F^{p\times q}$, $\boldsymbol{B}\in F^{q\times s}$, 则对于 $\boldsymbol{C}=\boldsymbol{AB}$, 其 r 阶子阵的行列式为

$$\left|\boldsymbol{C}\left(\begin{smallmatrix}i_1&i_2&\cdots&i_r\\j_1&j_2&\cdots&j_r\end{smallmatrix}\right)\right|=\begin{cases}0, & \text{当 } r>q,\\ \displaystyle\sum_{1\le k_1<k_2<\cdots<k_r\le q}\left|\boldsymbol{A}\left(\begin{smallmatrix}i_1&i_2&\cdots&i_r\\k_1&k_2&\cdots&k_r\end{smallmatrix}\right)\right|\cdot\left|\boldsymbol{B}\left(\begin{smallmatrix}k_1&k_2&\cdots&k_r\\j_1&j_2&\cdots&j_r\end{smallmatrix}\right)\right|, & \text{当 } r\le q.\end{cases}$$

我们不给出该公式的证明. 感兴趣的同学可以自己研究.

行列式的展开 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n$ 是 $\boldsymbol{A}$ 的 n 个行向量. 在前面的讨论里我们已经提到, 由于这些行向量的有序排列唯一地决定了方阵 $\boldsymbol{A}$, 行列式 $\det(\boldsymbol{A})$ 也可以视为关于行向量 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n$ 的函数: $\det(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n)$. 该多元函数满足:

(1) **反对称性**: 交换两个行向量的位置, 则行列式变号.

(2) **多重线性**: 行列式关于每个行向量都是线性的 (在固定其它行向量的条件下).

(3) **规范性**: 若 $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ 是 n 维数组中的基本向量, 则 $\det(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n) = 1$. 注意, 这组基本向量对应的方阵为 $\boldsymbol{I}_n$.

值得一提的是, 我们在定理 4.72, 推论 4.73, 定理 4.74 和推论 4.75 的证明中, 我们仅用到了行列式函数 det 满足定理 4.71 所给出的性质. 由于在多重线性的条件下, 反对称性与定理 4.71 中的(c)是等价的, 这说明, 若关于 n 个 n 维行向量的多元函数满足以上三条, 则它必然是相应矩阵的行列式.

设 $\boldsymbol{A} = (a_{ij})_{n\times n}$, 而 $\boldsymbol{\alpha}_i$ 是 $\boldsymbol{A}$ 的第 i 个行向量 $(1 \le i \le n)$. 设 $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ 是 n 维数组中的基本向量, 写成行向量的形式. 因此, $\boldsymbol{\alpha}_i = \sum_{j=1}^n a_{ij}\boldsymbol{e}_j$. 从而

$$\begin{aligned}
\det(\boldsymbol{A}) &= \det(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n) \\
&= \det\left(\sum_{j_1=1}^n a_{1j_1}\boldsymbol{e}_{j_1}, \sum_{j_2=1}^n a_{2j_2}\boldsymbol{e}_{j_2}, \ldots, \sum_{j_n=1}^n a_{nj_n}\boldsymbol{e}_{j_n}\right) \\
&\xlongequal{\text{多重线性}} \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_n=1}^n a_{1j_1}a_{2j_2}\cdots a_{nj_n} \det(\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}, \ldots, \boldsymbol{e}_{j_n})
\end{aligned}$$

注意到若 $j_p = j_q$ 而 $p \neq q$, 则对应的行列式为 0. 于是

$$\text{上式} = \sum_{\substack{j_1,j_2,\ldots,j_n\in\{1,2,\ldots,n\}\\ \text{且互不相等}}} a_{1j_1}a_{2j_2}\cdots a_{nj_n} \det(\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}, \ldots, \boldsymbol{e}_{j_n})$$

其中的求和的条件也称为 $j_1, j_2, \ldots, j_n$ 构成 $1, 2, \ldots, n$ 的一个**排列**. 为了表示的方便, 我们将其记作 $(j_1, j_2, \ldots, j_n) \in \mathcal{S}_n$, 而 $\mathcal{S}_n$ 表示所有的 $1, 2, \ldots, n$ 的排列的集合. 有的教材里也将这个集合记作 S_n 或者 $\mathfrak{S}_n$. 故

$$\text{上式} = \sum_{(j_1,j_2,\ldots,j_n)\in\mathcal{S}_n} a_{1j_1}a_{2j_2}\cdots a_{nj_n} \det(\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}, \ldots, \boldsymbol{e}_{j_n}).$$

因此, 关于行列式 $\det(\boldsymbol{A})$ 的计算化归成关于 $\det(\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}, \ldots, \boldsymbol{e}_{j_n})$ 的计算. 后者是单位矩阵 $\boldsymbol{I}_n$ 打乱行的顺序后得到的新矩阵的行列式. 由于其可以通过若干次交换行的操作得到 (参见接下来的讨论), 因此该行列式为 $(-1)^? \det(\boldsymbol{I}_n) = \pm 1$.

接下来具体讨论其取值. 不妨设 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{e}_{j_1} \\ \boldsymbol{e}_{j_2} \\ \vdots \\ \boldsymbol{e}_{j_n} \end{pmatrix}$ 是由单位矩阵 $\boldsymbol{I}_n$ 通过 N 次的交换行得到的. 很明显, 不同人来操作, 所选用的 N 可能不同, 但是其奇偶性必然一致, 这是因为 $\det(\boldsymbol{A}) = (-1)^N$ 由矩阵 $\boldsymbol{A}$ 唯一决定. 下面关注 N 的奇偶性.

更一般地, 考察由两两不同的实数构成的一个有序数组 $s = (a_1, a_2, \ldots, a_n)$, 并称之为一个**排列**. 对于该 s, 若 $i < j$ 而 $a_i > a_j$, 则数对 (a_i, a_j) 称为 (相对于 s 的) 一个**逆序**. s 的逆序的个数称为 s 的**逆序数**, 记作 $\tau(s)$. 若 $\tau(s) = 0$, 即 $a_1 < a_2 < \cdots < a_n$, 则 s 称为一个**顺序排列**. 若 $\tau(s)$ 为奇数 (对应地, 偶数), 则称 s 为一个**奇排列** (对应地, **偶排列**). 对于一个排列, 若交换其中的两个元素的位置, 则称对其做了一次**对换**.

例 4.80.

$$\begin{array}{ccc} s = (1,3,4,2) & \text{逆序为: } (3,2),\ (4,2) & \tau(s) = 2 \\ \big\downarrow{\scriptstyle \text{对换 } 2,4} & & \\ s' = (1,3,2,4) & \text{逆序为: } (3,2) & \tau(s') = 1 \\ \big\downarrow{\scriptstyle \text{对换 } 2,3} & & \\ s'' = (1,2,3,4) & \text{这是一个顺序排列} & \tau(s'') = 0 \end{array}$$

引理 4.81. *每个排列 s 都可以经过 $\tau(s)$ 次相邻位置的对换变成一个顺序排列.* 教材引理 4.3.1

证明. 设 s 是关于 $a_1 < a_2 < \cdots < a_n$ 的一个排列. 设在 s 中形如 $(*, a_i)$ 的逆序的个数为 m_i, 则显然 $\tau(s) = m_1 + m_2 + \cdots + m_n$. 此时, a_1 必然在 (从左数的) 第 $m_1 + 1$ 个位置上. 该 a_1 可以通过 m_1 次与相邻靠前的元素的对换, 移到第 1 个位置上. 每次对换时, 相应的 $m_2, m_3, \ldots, m_n$ 都不改变.

再考虑此时的后 $n-1$ 数构成的有序数组, 它是 $a_2 < a_3 < \cdots < a_n$ 的一个排列. 其逆序数为 $m_2 + m_3 + \cdots + m_n$. 由归纳法, 我们可以通过这么多个对换, 形成一个顺序排列.

由于 a_1 在这些数之前, 比这些数都小, 连接在一起后, 我们得到一个顺序排列. 这说明原来的有序数组 s 可以通过 $\tau(s)$ 个相邻对换得到一个顺序排列. □

推论 4.82. *$\mathcal{S}_n$ 中的排列 $(j_1, j_2, \ldots, j_n)$ 可以通过 $\tau(j_1, j_2, \ldots, j_n)$ 个对换得到顺序排列 $(1, 2, \ldots, n)$, 从而 $\det(\boldsymbol{e}_{j_1}, \boldsymbol{e}_{j_2}, \ldots, \boldsymbol{e}_{j_n}) = (-1)^{\tau(j_1, j_2, \ldots, j_n)} \det(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n) = (-1)^{\tau(j_1, j_2, \ldots, j_n)}$.*

综上, 我们证明了行列式有如下的展开式.

定理 4.83. 设方阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$, 则 教材定理 4.3.3

$$\det(\boldsymbol{A})=\sum_{(j_1,j_2,\ldots,j_n)\in\mathcal{S}_n}(-1)^{\tau(j_1,j_2,\ldots,j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}. \tag{4.8}$$

注 4.84. (1) 若 $\mathcal{S}_n$ 中的排列 $(j_1,j_2,\ldots,j_n)$ 可以通过 k 个对换得到顺序排列 $(1,2,\ldots,n)$, 则 $\det(\boldsymbol{e}_{j_1},\boldsymbol{e}_{j_2},\ldots,\boldsymbol{e}_{j_n})=(-1)^k$. 这说明 k 与 $\tau(j_1,j_2,\ldots,j_n)$ 同奇偶, 即 $(-1)^k=(-1)^{\tau(j_1,j_2,\ldots,j_n)}$.

(2) 前面提到一个排列可以通过若干次对换得到顺序排列. 这样的操作若反过来, 则可以用来说明这样一个简单事实: 任何一个 $\mathcal{S}_n$ 中的排列总可以通过若干次的对换得到 $\mathcal{S}_n$ 中事先指定的排列.

(3) 若一个排列 σ 在经过一次对换后得到新的排列 σ', 则 σ' 在经过相同的对换后可以还原到 σ. 这说明 σ' 可以通过 $\tau(\sigma)+1$ 次对换得到顺序排列. 特别地, 这说明了 σ 与 σ' 的奇偶性相反.

注 4.85. 对于展开式 (4.8) 中的项 $a_{1j_1}a_{2j_2}\cdots a_{nj_n}$, 若任意调换其中因子的顺序得到了 $a_{i_1k_1}a_{i_2k_2}\cdots a_{i_nk_n}$, 则其行指标和列指标分别构成了排列 $(i_1,i_2,\ldots,i_n)$ 和 $(k_1,k_2,\ldots,k_n)$. 它们的逆序数有如下的关系:

$$(-1)^{\tau(j_1,j_2,\ldots,j_n)}=(-1)^{\tau(i_1,i_2,\ldots,i_n)+\tau(k_1,k_2,\ldots,k_n)},$$

或者等价地, $\tau(j_1,j_2,\ldots,j_n)$ 与 $\tau(i_1,i_2,\ldots,i_n)+\tau(k_1,k_2,\ldots,k_n)$ 有相同的奇偶性.

为了看出这一点, 我们注意到任意对调 $a_{i_1k_1}a_{i_2k_2}\cdots a_{i_nk_n}$ 中的两个因子时, 排列 $(i_1,i_2,\ldots,i_n)$ 和 $(k_1,k_2,\ldots,k_n)$ 同时做了一次对换, 于是 $\tau(i_1,i_2,\ldots,i_n)$ 和 $\tau(k_1,k_2,\ldots,k_n)$ 同时改变了奇偶性. 因而 $\tau(i_1,i_2,\ldots,i_n)+\tau(k_1,k_2,\ldots,k_n)$ 的奇偶性不变. 又因为 $a_{1j_1}a_{2j_2}\ldots a_{nj_n}$ 可由 $a_{i_1k_1}a_{i_2k_2}\cdots a_{i_nk_n}$ 经过有限次对调因子得到, 所以 $\tau(j_1,j_2,\ldots,j_n)$ 与 $\tau(i_1,i_2,\ldots,i_n)+\tau(k_1,k_2,\ldots,k_n)$ 有相同的奇偶性.

作为上面注解的一个重要应用, 我们还有如下的公式.

定理 4.86. 方阵 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 的行列式还可以对称地按列的自然顺序展开成

$$\det(\boldsymbol{A})=\sum_{(i_1,i_2,\ldots,i_n)\in\mathcal{S}_n}(-1)^{\tau(i_1,i_2,\ldots,i_n)}a_{i_11}a_{i_22}\cdots a_{i_nn}.$$

证明. **(思路一)** 我们只需证明

$$\sum_{(j_1,j_2,\ldots,j_n)\in\mathcal{S}_n}(-1)^{\tau(j_1,j_2,\ldots,j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}=\sum_{(i_1,i_2,\ldots,i_n)\in\mathcal{S}_n}(-1)^{\tau(i_1,i_2,\ldots,i_n)}a_{i_11}a_{i_22}\cdots a_{i_nn}.$$

对于左边求和式中的通项

$$(-1)^{\tau(j_1,j_2,\ldots,j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n},$$

假定在适当打乱因子的顺序后假定我们有

$$a_{1j_1}a_{2j_2}\cdots a_{nj_n}=a_{i_11}a_{i_2i}\cdots a_{i_nn}.$$

由注 4.85 可知, 我们有

$$(-1)^{\tau(j_1,j_2,\ldots,j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}=(-1)^{\tau(i_1,i_2,\ldots,i_n)}a_{i_11}a_{i_2i}\cdots a_{i_nn}.$$

由于不同的排列 $(j_1,j_2,\ldots,j_n)$ 给出了不同的排列 $(i_1,i_2,\ldots,i_n)$, 当 $(j_1,j_2,\ldots,j_n)$ 遍历 $\mathcal{S}_n$ 时, 对应的 $(i_1,i_2,\ldots,i_n)$ 也遍历 $\mathcal{S}_n$. 故结论成立.

(思路二) 我们已经有 $\det(\boldsymbol{A})=\det(\boldsymbol{A}^{\mathrm{T}})$. 接下来只需将公式 (4.8) 用于 $\boldsymbol{A}^{\mathrm{T}}$ 即可. □

之前我们已经引入了方阵的代数余子式 A_{ij} 的概念. 对于方阵 $\boldsymbol{A}$, 我们定义 $\boldsymbol{A}$ 的**伴随矩阵** (adjoint matrix) 为

$$\operatorname{adj}(\boldsymbol{A})=\boldsymbol{A}^*:=\begin{pmatrix} A_{11} & A_{21}\cdots\cdots & A_{n1}\\ A_{12} & A_{22}\cdots\cdots & A_{n2}\\ \vdots & \vdots & \vdots\\ A_{1n} & A_{2n}\cdots\cdots & A_{nn}\end{pmatrix}.$$

请特别注意其下标的排列顺序. 特别地, $\boldsymbol{A}^*$ 的第 (i,j) 元素为 $\boldsymbol{A}$ 的第 (j,i) 元素对应的代数余子式 A_{ji}. 不难看出

$$(\boldsymbol{A}^{\mathrm{T}})^*=(\boldsymbol{A}^*)^{\mathrm{T}}.$$

定理 4.87. *对于方阵 $\boldsymbol{A}$, 我们有 $\boldsymbol{A}\boldsymbol{A}^*=\boldsymbol{A}^*\boldsymbol{A}=\det(\boldsymbol{A})\cdot\boldsymbol{I}_n$.* 教材定理 4.3.7

之前我们已经引入了方阵的代数余子式 A_{ij} 的概念. 对于方阵 $\boldsymbol{A}$, 我们定义 $\boldsymbol{A}$ 的**伴随矩阵** (adjoint matrix) 为

$$\operatorname{adj}(\boldsymbol{A}) = \boldsymbol{A}^* := \begin{pmatrix} A_{11} & A_{21} \cdots\cdots A_{n1} \\ A_{12} & A_{22} \cdots\cdots A_{n2} \\ \vdots & \vdots \qquad\quad \vdots \\ A_{1n} & A_{2n} \cdots\cdots A_{nn} \end{pmatrix}.$$

请特别注意其下标的排列顺序. 特别地, $\boldsymbol{A}^*$ 的第 (i,j) 元素为 $\boldsymbol{A}$ 的第 (j,i) 元素对应的代数余子式 A_{ji}. 不难看出

$$(\boldsymbol{A}^{\mathrm{T}})^* = (\boldsymbol{A}^*)^{\mathrm{T}}.$$

定理 4.3.43. 对于方阵 $\boldsymbol{A}$, 我们有 $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A} = \det(\boldsymbol{A}) \cdot \boldsymbol{I}_n$. 教材定理 4.3.7

证明. 设 $\boldsymbol{A}$ 的行向量依次为 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n$, 而 $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{P} = (p_{ij})_{n\times n}$.

(1) 对 $\boldsymbol{A}$ 的行列式, 我们可以按第 i 行展开, 从而得到

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} a_{ik}A_{ik}.$$

按照上面的记号, 这恰为 p_{ii}.

(2) 若 $1 \le i \ne j \le n$, 令

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_{i-1} \\ \boldsymbol{\alpha}_j \\ \boldsymbol{\alpha}_{i+1} \\ \vdots \\ \boldsymbol{\alpha}_n \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow \text{第 } i \text{ 行}. \\ \\ \\ \\ \end{matrix}$$

则 $\boldsymbol{B}$ 的第 i 行与第 j 行完全相同, 从而 $\det(\boldsymbol{B}) = 0$. 另外, 我们还注意到, $\boldsymbol{B}$ 的第 i 行的元素的代数余子式与 $\boldsymbol{A}$ 的第 i 行的元素的代数余子式对应相等. 此时, 对于 $\boldsymbol{B}$ 的行列式按第 i 行展开, 我们类似得到

$$0 = \det(\boldsymbol{B}) = \sum_{k=1}^{n} a_{jk}A_{ik} = p_{ji}.$$

综上,

$$p_{ij}=\begin{cases}\det(\boldsymbol{A}), & 若\ i=j,\\ 0, & 若\ i\neq j.\end{cases}$$

这说明 $\boldsymbol{AA}^*=\det(\boldsymbol{A})\cdot\boldsymbol{I}_n$. 类似可证明 $\boldsymbol{A}^*\boldsymbol{A}=\det(\boldsymbol{A})\cdot\boldsymbol{I}_n$. □

上面的定理有个非常重要的推论.

定理 4.3.44. 若方阵 $\boldsymbol{A}$ 可逆, 则 $\boldsymbol{A}^{-1}=\frac{1}{\det(\boldsymbol{A})}\cdot\boldsymbol{A}^*$. 教材定理 4.3.8

证明. 若 $\boldsymbol{A}$ 可逆, 则 $\det(\boldsymbol{A})\neq 0$, 由定理 4.3.43 可知:

$$\boldsymbol{A}\left(\frac{1}{\det(\boldsymbol{A})}\cdot\boldsymbol{A}^*\right)=\left(\frac{1}{\det(\boldsymbol{A})}\cdot\boldsymbol{A}^*\right)\boldsymbol{A}=\boldsymbol{I}_n.$$

这说明 $\boldsymbol{A}^{-1}$ 存在, 且为 $\frac{1}{\det(\boldsymbol{A})}\cdot\boldsymbol{A}^*$. □

注 4.3.45. 上面的定理给出了求可逆矩阵的逆矩阵的一个理论解法. 不过由于它需要求解 n^2 个 $n-1$ 阶方阵的行列式, 计算量大, 并不实用. 当然, 在 $n=2$ 时, 该方法是非常合适的.

命题 4.3.46. 对于 $n\geq 2$, 若 $\boldsymbol{A}$ 为 n 阶方阵, 则 $\det(\boldsymbol{A}^*)=\det(\boldsymbol{A})^{n-1}$.

证明. 由 $\boldsymbol{AA}^*=\det(\boldsymbol{A})\boldsymbol{I}_n$ 出发, 我们得到 $\det(\boldsymbol{A})\det(\boldsymbol{A}^*)=\det(\boldsymbol{A})^n$.

若 $\det(\boldsymbol{A})\neq 0$, 从而可以得到 $\det(\boldsymbol{A}^*)=\det(\boldsymbol{A})^{n-1}$.

若 $\det(\boldsymbol{A})=0$, 我们只需证明 $\det(\boldsymbol{A}^*)=0$. 若否, 则 $\boldsymbol{A}^*$ 可逆, 再由 $\boldsymbol{AA}^*=\det(\boldsymbol{A})\boldsymbol{I}_n=\boldsymbol{O}$ 出发可推知, $\boldsymbol{A}=\boldsymbol{O}$, 从而 $\boldsymbol{A}^*=\boldsymbol{O}$. 这与 $\det(\boldsymbol{A}^*)\neq 0$ 的假设相矛盾. □

习题 4.3.47. 设 $\boldsymbol{A}$ 为一个 n 阶反对称方阵. 若 n 为奇数, 验证 $\boldsymbol{A}^*$ 是一个对称矩阵; 若 n 为偶数, 验证 $\boldsymbol{A}^*$ 是一个反对称矩阵.

习题 4.3.48. 设 $\boldsymbol{A}$ 为一个 n 阶反对称方阵, 其主对角线的右上角的元素全是 1. 计算 $\boldsymbol{A}^*$.

习题 4.3.49. 若 $\boldsymbol{V}=\begin{pmatrix}1 & x_1 & x_1^2\cdots\cdots x_1^{n-1}\\ \vdots & \vdots & \vdots \qquad\qquad \vdots\\ 1 & x_n & x_n^2\cdots\cdots x_n^{n-1}\end{pmatrix}$, 其中 $x_1,\dots,x_n$ 互不相等. 计算 $\boldsymbol{V}^{-1}$.

习题 4.3.50. 设 $\boldsymbol{A}=\begin{vmatrix}1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{vmatrix}$, 求 $\boldsymbol{A}$ 的所有 2 阶代数余子式之和.

Cramer 法则

定理 4.3.51. 若 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 为可逆矩阵, $\boldsymbol{x}=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$, $\boldsymbol{b}=\begin{pmatrix}b_1\\ \vdots\\ b_n\end{pmatrix}$, 则线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 有唯一解. 该解满足 $x_i=|\boldsymbol{A}_i|/|\boldsymbol{A}|$, 其中 $\boldsymbol{A}_i$ 为 $\boldsymbol{A}$ 的第 i 列换成 $\boldsymbol{b}$ 后得到的新矩阵.* 教材定理 4.3.9

证明. 若 $\boldsymbol{A}$ 可逆, 则方程组显然有唯一解 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{b}$. 此时, 设 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$ 分别为 $\boldsymbol{A}$ 的各个列向量, 则 $\boldsymbol{b}=\sum_j x_j\boldsymbol{a}_j$, 其中 $\boldsymbol{x}=(x_1,\dots,x_n)^{\mathrm{T}}$. 于是,

$$\begin{aligned}\det(\boldsymbol{A}_i)&=\det(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{i-1},\boldsymbol{b},\boldsymbol{a}_{i+1},\dots,\boldsymbol{a}_n)=\det(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{i-1},\sum_j x_j\boldsymbol{a}_j,\boldsymbol{a}_{i+1},\dots,\boldsymbol{a}_n)\\&=\sum_j x_j\det(\boldsymbol{a}_1,\dots,\boldsymbol{a}_{i-1},\boldsymbol{a}_j,\boldsymbol{a}_{i+1},\dots,\boldsymbol{a}_n)=x_i\det(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)=x_i\det(\boldsymbol{A}),\end{aligned}$$

其中求和式里, 若 $j\neq i$, 则相应的行列式为 0. 从而, $x_i=\det(\boldsymbol{A}_i)/\det(\boldsymbol{A})$. $\square$

例 4.3.52. 设 $a_1,a_2,\dots,a_n$ 是互不相同的数. 解方程组

$$\begin{cases}x_1+a_1x_2+\cdots+a_1^{n-1}x_n=1,\\ x_1+a_2x_2+\cdots+a_2^{n-1}x_n=1,\\ \qquad\vdots\\ x_1+a_nx_2+\cdots+a_n^{n-1}x_n=1.\end{cases}$$

解. 该线性方程组的系数方阵的行列式为

$$\begin{vmatrix}1 & a_1\cdots\cdots a_1^{n-1}\\ 1 & a_2\cdots\cdots a_2^{n-1}\\ \vdots & \vdots \qquad\quad \vdots\\ 1 & a_n\cdots\cdots a_n^{n-1}\end{vmatrix}=\prod_{1\leq i<j\leq n}(a_j-a_i)\neq 0.$$

*克莱因所著的《古今数学思想》里介绍: 用行列式的方法解含有两个、三个和四个未知量的联立线性方程, 可能在 1729 年, 是由**麦克劳林**开创的, 并发表在他的遗作《代数论著》(Treatise of Algebra, 1748) 中. 虽然书中的记法不太好, 但是他的法则是我们今天所使用的法则. **克莱姆**把它发表在他的《线性代数分析导言》(Introduction a l'analyse des lignes courbes algebriques, 1750) 中. 克莱姆给出了一条法则, 用于确定经过五个点的般的二次曲线 $A+By+Cx+Dy^2+Exy+x^2=0$ 的系数. 他的行列式和现在一样, 是这样一些乘积的和, 这些乘积是在每一行和每一列中取一个且只取一个元素组成: 每一个乘积的符号是这样确定的, 即从标准次序出发, 得到这些元素的排列所需的重排数, 如果这个数是偶数, 则符号是正的, 否则就是负的. 1764 年, **贝祖**把确定行列式每一项的符号的手续系统化了. 给定了含 n 个未知量的 n 个齐次线性方程, 贝祖证明: 系数行列式等于零 (结式等于零) 是这方程组有非零解的条件.

故该方程组有唯一解. 显然,

$$x_1 = 1, \quad x_2 = x_3 = \cdots = x_n = 0$$

是它的 (唯一) 解.

若使用定理 4.3.51 中的记号, 则 $j = 1$ 时, $\boldsymbol{A} = \boldsymbol{A}_1$ 可逆, 从而 $x_1 = 1$. 而对于 $j = 2, 3, \ldots, n$, A_j 中的第 1 列和第 j 列相等, 从而 $\det(\boldsymbol{A}_j) = 0$, 进而有 $x_j = 0$. □

习题 4.3.53. 设 $\boldsymbol{A} = \boldsymbol{A}(t) = (a_{ij}(t))$ 是一个可逆方阵, 其每个元素都是关于实变量 t 的可微函数. 记 $\boldsymbol{A}' = \boldsymbol{A}'(t) = (a'_{ij}(t))$ 为对每个元素都求导后得到的方阵. 证明: 行列式 $\det(\boldsymbol{A})$ 作为 t 的函数的导数满足

$$\frac{\mathrm{d}}{\mathrm{d}t}(\det(\boldsymbol{A})) = \det(\boldsymbol{A}) \cdot \operatorname{tr}(\boldsymbol{A}' \cdot \boldsymbol{A}^{-1}).$$

4.4 初等变换

定理 4.4.1. 对于任意矩阵 $\boldsymbol{A} = (a_{ij})_{m \times n}$, 存在一系列的 m 阶初等矩阵 $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_s$ 和 n 阶初等矩阵 $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_t$ 使得 教材定理 4.4.3

$$\boldsymbol{P}_s \cdots \boldsymbol{P}_1 \boldsymbol{A} \boldsymbol{Q}_1 \cdots \boldsymbol{Q}_t = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}.$$

证明. 我们只需证明存在一系列的行和列的初等变换, 将 $\boldsymbol{A}$ 化为 $\left(\begin{smallmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{smallmatrix}\right)$ 的形式.

$$\boldsymbol{A} \xrightarrow[\text{Gauss 消元法}]{\text{一系列的初等行变换}} \text{约化标准形} \xrightarrow[\text{其次, 适当交换列}]{\text{首先, 用各行主元 1 将该行其它元素消为 0}} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}. \qquad \square$$

很明显, 上面定理中的 r 是 $\boldsymbol{A}$ 的行标准形 (阶梯形矩阵) 的主元的个数 (非零行数). 另外, 将上面定理中的矩阵乘积 $\boldsymbol{P}_s \cdots \boldsymbol{P}_1$ 记作 $\boldsymbol{P}$, 矩阵乘积 $\boldsymbol{Q}_1 \cdots \boldsymbol{Q}_t$ 记作 $\boldsymbol{Q}$, 则 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 分别为 m 阶和 n 阶可逆矩阵. 此时有

定理 4.4.2. 对于任意矩阵 $\boldsymbol{A} = (a_{ij})_{m \times n}$, 存在 m 阶可逆矩阵 $\boldsymbol{P}$ 和 n 阶可逆矩阵 $\boldsymbol{Q}$, 使得 $\boldsymbol{PAQ} = \left(\begin{smallmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{smallmatrix}\right)$. 教材定理 4.4.4

求逆矩阵的初等变换法 考虑分块矩阵

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{I}_n \end{pmatrix} \xrightarrow[\substack{\text{等同于依次左乘了 } \boldsymbol{P}_1, \ldots, \boldsymbol{P}_s \\ \text{这些 } \boldsymbol{P}_i \text{ 是初等矩阵}}]{\text{假定存在一系列的初等行变换}} \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{X} \end{pmatrix},$$

其中 $\boldsymbol{A}$ 为 n 阶方阵. 若记 $\boldsymbol{P}=\boldsymbol{P}_s\cdots\boldsymbol{P}_1$, 则上面的操作说明

$$\begin{pmatrix}\boldsymbol{I}_n & \boldsymbol{X}\end{pmatrix}=\boldsymbol{P}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{I}_n\end{pmatrix}=\begin{pmatrix}\boldsymbol{PA} & \boldsymbol{PI}_n\end{pmatrix}.$$

即

$$\boldsymbol{PA}=\boldsymbol{I}_n,\qquad \boldsymbol{PI}_n=\boldsymbol{X}.$$

从而 $\boldsymbol{X}=\boldsymbol{P}=\boldsymbol{A}^{-1}$ 为可逆矩阵 $\boldsymbol{A}$ 的逆. 这是我们求方阵的逆的标准方法, 大家需要熟练掌握. 当然, 我们需要说明, 若 $\boldsymbol{A}$ 可逆, 则这个方法一定可以求出 $\boldsymbol{A}^{-1}$. 这是因为此时 $\boldsymbol{A}^{-1}$ 也可逆, 从而可以表示成初等方阵的乘积: $\boldsymbol{A}^{-1}=\boldsymbol{P}_s\cdots\boldsymbol{P}_1$. 对 $\begin{pmatrix}\boldsymbol{A} & \boldsymbol{I}_n\end{pmatrix}$ 依次施行 $\boldsymbol{P}_1,\ldots,\boldsymbol{P}_s$ 所对应的初等行变换即可.

另外, 对称地, 我们也可以有

$$\begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{I}_n\end{pmatrix}\xrightarrow{\text{一系列的初等列变换}}\begin{pmatrix}\boldsymbol{I}_n\\ \boldsymbol{A}^{-1}\end{pmatrix}.$$

例 4.4.3. 求可逆方阵 $\boldsymbol{A}=\begin{pmatrix}1 & -1 & 2\\ 3 & 0 & 1\\ 2 & -1 & 0\end{pmatrix}$ 的逆矩阵.

其计算可以留作课前热身题

解.

$$\begin{aligned}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{I}_3\end{pmatrix}&=\begin{pmatrix}1 & -1 & 2 & 1 & 0 & 0\\ 3 & 0 & 1 & 0 & 1 & 0\\ 2 & -1 & 0 & 0 & 0 & 1\end{pmatrix}\xrightarrow[-2r_1\to r_3]{-3r_1\to r_2}\begin{pmatrix}1 & -1 & 2 & 1 & 0 & 0\\ 0 & 3 & -5 & -3 & 1 & 0\\ 0 & 1 & -4 & -2 & 0 & 1\end{pmatrix}\\ &\xrightarrow{r_2\leftrightarrow r_3}\begin{pmatrix}1 & -1 & 2 & 1 & 0 & 0\\ 0 & 1 & -4 & -2 & 0 & 1\\ 0 & 3 & -5 & -3 & 1 & 0\end{pmatrix}\xrightarrow[-3r_2\to r_3]{r_2\to r_1}\begin{pmatrix}1 & 0 & -2 & -1 & 0 & 1\\ 0 & 1 & -4 & -2 & 0 & 1\\ 0 & 0 & 7 & 3 & 1 & -3\end{pmatrix}\\ &\xrightarrow{\frac{1}{7}r_3}\begin{pmatrix}1 & 0 & -2 & -1 & 0 & 1\\ 0 & 1 & -4 & -2 & 0 & 1\\ 0 & 0 & 1 & 3/7 & 1/7 & -3/7\end{pmatrix}\xrightarrow[4r_3\to r_2]{2r_3\to r_1}\begin{pmatrix}1 & 0 & 0 & -1/7 & 2/7 & 1/7\\ 0 & 1 & 0 & -2/7 & 4/7 & -5/7\\ 0 & 0 & 1 & 3/7 & 1/7 & -3/7\end{pmatrix}.\end{aligned}$$

这说明

$$\boldsymbol{A}^{-1}=\begin{pmatrix}-1/7 & 2/7 & 1/7\\ -2/7 & 4/7 & -5/7\\ 3/7 & 1/7 & -3/7\end{pmatrix}.$$

□

例 4.4.4. 设 $\boldsymbol{A}=\begin{pmatrix}0&1&2\\1&1&-1\\2&4&2\end{pmatrix}$, 用上面的方法, 在做过一系列的初等行变换后, 我们有

$$\begin{pmatrix}\boldsymbol{A}&\boldsymbol{I}\end{pmatrix}\to\begin{pmatrix}1&1&-1&0&1&0\\0&1&2&1&0&0\\0&0&0&1&1&-\frac{1}{2}\end{pmatrix}\to\operatorname{rref}(\boldsymbol{A},\boldsymbol{I})=\begin{pmatrix}1&0&-3&0&2&-\frac{1}{2}\\0&1&2&0&-1&\frac{1}{2}\\0&0&0&1&1&-\frac{1}{2}\end{pmatrix}.$$

其计算可以留作课前热身题

从中间的阶梯标准形就可以看出, $\boldsymbol{A}$ 不是可逆矩阵; 因为有可逆矩阵左乘它得到一个不可逆矩阵. 当然, 由于 $\det(\boldsymbol{A})=0$, 我们也可以很容易地看到这一点.

分块矩阵的初等行变换

(1) 两个块行互换位置;

(2) 用一个可逆矩阵左乘某一块行;

(3) 用一个矩阵 $\boldsymbol{P}$ (不要求可逆) 左乘某一块行加到另一块行上. 例如

$$\begin{pmatrix}\boldsymbol{A}_{11}&\boldsymbol{A}_{12}\\\boldsymbol{A}_{21}&\boldsymbol{A}_{22}\end{pmatrix}\xrightarrow{Pr_1\to r_2}\begin{pmatrix}\boldsymbol{A}_{11}&\boldsymbol{A}_{12}\\\boldsymbol{P}\boldsymbol{A}_{11}+\boldsymbol{A}_{21}&\boldsymbol{P}\boldsymbol{A}_{12}+\boldsymbol{A}_{22}\end{pmatrix}.$$

注意, 这里的 r_1 表示的分块矩阵的第一个块行, 不是元素的第一行. 类似地, 还有分块矩阵的初等列变化. 但是对应的 (2) 和 (3) 需要考虑的是矩阵的右乘.

类似于初等矩阵的定义, 将**分块单位阵** (单位阵分块后得到的分块矩阵) 经过一次初等行 (列) 变换后得到的分块矩阵称为**分块初等矩阵**.

类似于之前的结果 (定理 4.3.22), 对于分块矩阵作初等行 (列) 变换等同于用一个相应的分块初等矩阵左乘 (右乘).

例 4.4.5. 设 $\boldsymbol{A},\boldsymbol{B}$ 为 n 阶方阵, 证明:

$$\begin{vmatrix}\boldsymbol{A}&\boldsymbol{B}\\\boldsymbol{B}&\boldsymbol{A}\end{vmatrix}=|\boldsymbol{A}+\boldsymbol{B}|\cdot|\boldsymbol{A}-\boldsymbol{B}|.$$

证明.

$$\begin{vmatrix}\boldsymbol{A}&\boldsymbol{B}\\\boldsymbol{B}&\boldsymbol{A}\end{vmatrix}\xlongequal{r_2\to r_1}\begin{vmatrix}\boldsymbol{A}+\boldsymbol{B}&\boldsymbol{A}+\boldsymbol{B}\\\boldsymbol{B}&\boldsymbol{A}\end{vmatrix}\xlongequal{-c_1\to c_2}\begin{vmatrix}\boldsymbol{A}+\boldsymbol{B}&\boldsymbol{O}\\\boldsymbol{B}&\boldsymbol{A}-\boldsymbol{B}\end{vmatrix}=|\boldsymbol{A}+\boldsymbol{B}|\cdot|\boldsymbol{A}-\boldsymbol{B}|.\quad\square$$

例 4.4.6.

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{P} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{P}\boldsymbol{A}_{11}+\boldsymbol{A}_{21} & \boldsymbol{P}\boldsymbol{A}_{12}+\boldsymbol{A}_{22} \end{pmatrix}.$$

若对该式两边同取行列式, 由于 $\det \begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{P} & \boldsymbol{I} \end{pmatrix} = 1$ (参见习题 *4.3.33*), 这步计算表明对于分块矩阵的倍加行 (列) 操作不改变相应矩阵的行列式.*

进一步地, 若其中的 $\boldsymbol{A}_{11}$ 为方阵, $\boldsymbol{A}_{22}$ 为方阵, 且 $\boldsymbol{A}_{11}$ 可逆, 那么我们可以选取 $\boldsymbol{P} = -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}$ (为了用 $\boldsymbol{A}_{11}$ 消去 $\boldsymbol{A}_{21}$), 从而得到

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I} \end{pmatrix} \underbrace{\begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}}_{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{O} & -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}+\boldsymbol{A}_{22} \end{pmatrix}.$$

其中右下角的 $-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}+\boldsymbol{A}_{22}$ 称为 $\boldsymbol{A}_{11}$ 在 $\boldsymbol{A}$ 中的 ***Schur* 补**. 若取行列式 (参见习题 *4.3.33*), 则得到 ***Schur* 行列式公式**

$$\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{vmatrix} = |\boldsymbol{A}_{11}| \cdot \left|-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}+\boldsymbol{A}_{22}\right|.$$

例 4.4.7. 设 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{D} \\ \boldsymbol{O} & \boldsymbol{C} \end{pmatrix}$, 其中 $\boldsymbol{B}, \boldsymbol{C}$ 为方阵. 故 $\det(\boldsymbol{A}) = \det(\boldsymbol{B})\det(\boldsymbol{C})$ (参见习题 *4.3.33*), 从而 $\boldsymbol{A}$ 可逆当且仅当 $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 皆为可逆矩阵. 假设这一条件得到满足, 我们通过初等行变换来求 $\boldsymbol{A}$ 的逆矩阵:

$$\begin{aligned} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{I} \end{pmatrix} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{D} & \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{C} & \boldsymbol{O} & \boldsymbol{I} \end{pmatrix} &\xrightarrow{-\boldsymbol{D}\boldsymbol{C}^{-1}r_2 \to r_1} \begin{pmatrix} \boldsymbol{B} & \boldsymbol{O} & \boldsymbol{I} & -\boldsymbol{D}\boldsymbol{C}^{-1} \\ \boldsymbol{O} & \boldsymbol{C} & \boldsymbol{O} & \boldsymbol{I} \end{pmatrix} \\ &\xrightarrow[-\boldsymbol{C}^{-1}r_2]{-\boldsymbol{B}^{-1}r_1} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} & \boldsymbol{B}^{-1} & -\boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{C}^{-1} \\ \boldsymbol{O} & \boldsymbol{I} & \boldsymbol{O} & \boldsymbol{C}^{-1} \end{pmatrix}. \end{aligned}$$

这说明

$$\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{B}^{-1} & -\boldsymbol{B}^{-1}\boldsymbol{D}\boldsymbol{C}^{-1} \\ \boldsymbol{O} & \boldsymbol{C}^{-1} \end{pmatrix}.$$

习题 4.4.8. 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是 n 阶的可逆方阵, 求下面分块矩阵 $\begin{pmatrix} \boldsymbol{I} & \boldsymbol{A} & \boldsymbol{C} \\ \boldsymbol{O} & \boldsymbol{I} & \boldsymbol{B} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{I} \end{pmatrix}$ 的逆.

---

*对于分块矩阵的另外两个类型的初等变换, 其对行列式的改变就没有这么简单, 计算时需要慎重.

习题 4.4.9. (1) 设 $\boldsymbol{A}, \boldsymbol{B}$ 为 n 阶方阵, 且 $\boldsymbol{I}-\boldsymbol{BA}$ 可逆. 求矩阵 $\boldsymbol{M}=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{A}\\ \boldsymbol{B} & \boldsymbol{I}\end{pmatrix}$ 的逆矩阵.

(2) 在 $\boldsymbol{I}-\boldsymbol{AB}$ 可逆的条件下重新求上面的 $\boldsymbol{M}$ 的逆矩阵.

(3) 证明: $\boldsymbol{I}-\boldsymbol{AB}$ 可逆的充要条件是 $\boldsymbol{I}-\boldsymbol{BA}$ 可逆 (参见教材 *P115#25*, 其中 *#25* 的公式称为 ***Schur* 公式** (的特殊形式) 或 ***Sylvester* 行列式公式**, 需要牢记).

习题 4.4.10. 设 $\boldsymbol{B}\in F^{n\times m}$, $\boldsymbol{A}\in F^{m\times n}$. 证明 $\begin{pmatrix}\boldsymbol{B} & \boldsymbol{I}_n\\ \boldsymbol{O} & \boldsymbol{A}\end{pmatrix}$ 可逆的充要条件是 $\boldsymbol{AB}$ 可逆.

下面的例子展示了微扰法的技巧; 其本质是 n 阶可逆方阵的全体在 $F^{n\times n}$ 中稠密.

例 4.4.11. 设 $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}$ 为 4 个 n 阶复方阵, 满足 $\boldsymbol{AC}=\boldsymbol{CA}$. 证明:

$$\begin{vmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{vmatrix}=\left|\boldsymbol{AD}-\boldsymbol{CB}\right|.$$

证明. (1) 先考虑矩阵 $\boldsymbol{A}$ 可逆的情形. 此时

$$\begin{vmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{vmatrix}\xlongequal[\substack{\text{用第一行的 } \boldsymbol{A}\\ \text{消去第二行的 } \boldsymbol{C}}]{-(\boldsymbol{CA}^{-1})r_1\to r_2}\begin{vmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C}-(\boldsymbol{CA}^{-1})\boldsymbol{A} & \boldsymbol{D}-(\boldsymbol{CA}^{-1})\boldsymbol{B}\end{vmatrix}=\begin{vmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{O} & \boldsymbol{D}-\boldsymbol{CA}^{-1}\boldsymbol{B}\end{vmatrix}$$
$$=\left|\boldsymbol{A}\right|\left|\boldsymbol{D}-\boldsymbol{CA}^{-1}\boldsymbol{B}\right|=\left|\boldsymbol{AD}-\boldsymbol{ACA}^{-1}\boldsymbol{B}\right|.$$

由于 $\boldsymbol{AC}=\boldsymbol{CA}$, 可以推出 $\boldsymbol{ACA}^{-1}=\boldsymbol{C}$. 从而上式为 $|\boldsymbol{AD}-\boldsymbol{CB}|$.

(2) 假设 $\boldsymbol{A}$ 不可逆, 即 $|\boldsymbol{A}|=0$. 考虑辅助矩阵 $\boldsymbol{A}_t=\boldsymbol{A}+t\boldsymbol{I}$, 其中 $t\in\mathbb{C}$. 从而 $\boldsymbol{A}_0=\boldsymbol{A}$. 一个事实是:

$$|\boldsymbol{A}_t|=|\boldsymbol{A}+t\boldsymbol{I}|=t^n+\operatorname{tr}(\boldsymbol{A})t^{n-1}+\cdots+|\boldsymbol{A}|,$$

是一个关于 t 的首项系数为 1 的 n 次复系数多项式, 其系数由矩阵 $\boldsymbol{A}$ 决定, 中间的其它系数比较复杂. 这个多项式最多有 n 个不同的复根, 将其全体记作 $T=\{t_1,\dots,t_m\}$, 其中 $m\le n$. 故对任意的 $t\in\mathbb{C}\setminus T$, 有 $|\boldsymbol{A}_t|\neq 0$, 从而 $\boldsymbol{A}_t$ 可逆. 注意到 $\boldsymbol{A}_t\boldsymbol{C}=\boldsymbol{CA}_t$, 由上面的讨论, 知道

$$\begin{vmatrix}\boldsymbol{A}_t & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{vmatrix}=\left|\boldsymbol{A}_t\boldsymbol{D}-\boldsymbol{CB}\right|.$$

上面等式两边都是关于 t 的次数不超过 n 的多项式, 是连续函数. 对它们同时取极限 $\lim\limits_{\substack{t\to 0\\ t\notin T}}$, 左式收敛于 $\left|\begin{smallmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{smallmatrix}\right|$, 而右式收敛于 $|\boldsymbol{AD}-\boldsymbol{CB}|$, 故仍有等式

$$\begin{vmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{vmatrix}=\left|\boldsymbol{AD}-\boldsymbol{CB}\right|. \qquad \square$$

习题 4.4.12. 设 $\boldsymbol{A}=(a_{ij})\in F^{n\times n}$. 若 $n=2$ 或 3, 分别计算 $\det(\boldsymbol{A}+\lambda\boldsymbol{I}_n)$, 并将结果表示成关于 λ 的多项式.

习题 4.4.13. 设 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 为 n 阶复方阵, 证明以下关于伴随矩阵的等式:

(1) $(\boldsymbol{AB})^*=\boldsymbol{B}^*\boldsymbol{A}^*$;

(2) $(\boldsymbol{XAX}^{-1})^*=\boldsymbol{XA}^*\boldsymbol{X}^{-1}$, 其中 $\boldsymbol{X}$ 为可逆的 n 阶方阵;

(3) 若 $\boldsymbol{AB}=\boldsymbol{BA}$, 则 $\boldsymbol{A}^*\boldsymbol{B}=\boldsymbol{BA}^*$.

(提示: 先在 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 皆可逆的条件下证明这些结果. 若它们不同时可逆, 将其分别用 $\boldsymbol{A}+t\boldsymbol{I}_n$ 和 $\boldsymbol{B}+t\boldsymbol{I}_n$ 代替, 再取极限.)

习题 4.4.14. 设 $\boldsymbol{x},\boldsymbol{y}$ 为 n 维列向量, 证明以下关于伴随矩阵的等式:

$$(\boldsymbol{I}-\boldsymbol{xy}^{\mathrm{T}})^*=\boldsymbol{xy}^{\mathrm{T}}+(1-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x})\boldsymbol{I}.$$

(提示: 先在 $\boldsymbol{I}-\boldsymbol{xy}^{\mathrm{T}}$ 可逆的条件下证明该公式)

习题 4.4.15. 设 $\boldsymbol{A}$ 为 n 阶方阵, $\boldsymbol{v}$ 为 n 维列向量, $\boldsymbol{u}$ 为 n 维行向量, a 是一个数. 证明:

$$\begin{vmatrix}\boldsymbol{A} & \boldsymbol{v}\\ \boldsymbol{u} & a\end{vmatrix}=a|\boldsymbol{A}|-u\boldsymbol{A}^*\boldsymbol{v}.$$

(提示: 先在 $\boldsymbol{A}$ 可逆的条件下证明该公式)

习题 4.4.16. 设 $\boldsymbol{A}$ 为 n 阶方阵, $\boldsymbol{u},\boldsymbol{v}$ 为 n 维行向量. 证明:

$$|\boldsymbol{A}+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}|=|\boldsymbol{A}|+\boldsymbol{vA}^*\boldsymbol{u}^{\mathrm{T}}.$$

4.5 秩与相抵

在本节中, 对于矩阵 $\boldsymbol{A}$, 我们仍然用 $\boldsymbol{A}\binom{i_1\ i_2\ \cdots\ i_r}{j_1\ j_2\ \cdots\ j_r}$ 来表示由 $\boldsymbol{A}$ 的第 $i_1 < i_2 < \cdots < i_r$ 行、第 $j_1 < j_2 < \cdots < j_r$ 列元素构成的 r **阶子阵**, 其行列式称为 $\boldsymbol{A}$ 的一个 r **阶子式**.

例 4.5.1. 对于矩阵

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 5 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

而言, 它有 $C_4^3 C_3^3 = 4$ 个不同的 3 阶子阵, $C_4^2 C_3^2 = 18$ 个不同的 2 阶子阵.

定义 4.5.2. $\boldsymbol{A}$ 的非零子式的最高阶数 r 称为 $\boldsymbol{A}$ 的**秩** (*rank*)*, 记作 $r(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}) = r$. 约定: 零矩阵的秩为 0.

例 4.5.3. 上面例子中的 $\boldsymbol{A}$ 的所有 3 阶子式都为 0, 而其第 1, 2 行和 2, 3 列的元素所组成的 $\boldsymbol{A}$ 的 2 阶子式

$$\left|\boldsymbol{A}\binom{1\ 2}{2\ 3}\right| = \begin{vmatrix} 2 & 3 \\ 0 & 5 \end{vmatrix} = 10 \neq 0.$$

故 $\operatorname{rank}(\boldsymbol{A}) = 2$.

注 4.5.4. (1) 若 $\boldsymbol{A} \in F^{m\times n}$, 则 $\operatorname{rank}(\boldsymbol{A}) \leq \min(m, n)$.

(2) 由行列式的展开定义可知, 若 $\boldsymbol{A}$ 的所有 k 阶子式为 0, 那么 $\boldsymbol{A}$ 的所有高于 k 阶的子式 (若存在) 也必然为 0. 故 $\operatorname{rank}(\boldsymbol{A}) = r$ 的充要条件是 $\boldsymbol{A}$ 有 r 阶非零子式, 且所有 $r+1$ 阶子式为 0 (若 $r = \min(m, n)$, 最后一条视为自动成立).

(3) 由于转置不改变行列式, 而 $\boldsymbol{A}^{\mathrm{T}}$ 的每个子式都是由 $\boldsymbol{A}$ 的某个子阵转置后的行列式, 故 $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}) = \operatorname{rank}(\boldsymbol{A})$.

(4) 类似地, 对于 $\lambda \neq 0$, 有 $\operatorname{rank}(\lambda\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$.

(5) $\operatorname{rank}(\boldsymbol{A}) = 0$ 当且仅当 $\boldsymbol{A} = \boldsymbol{O}$. 若 $\operatorname{rank}(\boldsymbol{A}) = 1$, 则 $\boldsymbol{A}$ 的各个行向量互成比例, 其各个列向量也互成比例.

---

*其实很多人都认为, 用矩阵的行向量张成的线性空间的维数 (即下一章引入的矩阵的**行秩**这一概念) 来定义矩阵的秩, 是更佳的方式.

例 4.5.5. 设有阶梯形矩阵

$$\begin{pmatrix} 0\cdots\cdots 0 & a_{1j_1} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots a_{1n} \\ 0\cdots\cdots\cdots\cdots\cdots 0 & a_{2j_2}\cdots\cdots\cdots\cdots\cdots\cdots a_{2n} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 0\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 & a_{rj_r}\cdots\cdots a_{rn} \\ 0\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 \\ \vdots \qquad\qquad\qquad\qquad\qquad \vdots \\ 0\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots 0 \end{pmatrix}$$

其中 $j_1 < j_2 < \cdots < j_r$, 并且对于 $i = 1, 2, \ldots, r$ 有 $a_{ij_i} \neq 0$. 很明显, $\boldsymbol{A}$ 的所有 $r+1$ 阶子式全为 0, 并且 $\boldsymbol{A}$ 有 r 阶子式

$$\det\left(\boldsymbol{A}\begin{pmatrix} 1 & 2 & \ldots & r \\ j_1 & j_2 & \ldots & j_r \end{pmatrix}\right) = a_{1j_1}a_{2j_2}\cdots a_{rj_r} \neq 0,$$

这说明 $\operatorname{rank}(\boldsymbol{A}) = r$. 特别地, $\left(\begin{smallmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{smallmatrix}\right)$ 的秩为 r.

定理 4.5.6. 初等变换不改变矩阵的秩.

证明. 由对称性, 仅考虑初等行变换, 即证明初等行变换不改变矩阵的秩. 为此, 我们断言, 对于任意的矩阵 $\boldsymbol{A}$ 和初等矩阵 $\boldsymbol{E}$, 则必有

$$\operatorname{rank}(\boldsymbol{EA}) \leq \operatorname{rank}(\boldsymbol{A}). \tag{4.9}$$

由于反过来 $\boldsymbol{A} = \boldsymbol{E}^{-1}(\boldsymbol{EA})$, 而 $\boldsymbol{E}^{-1}$ 也是初等矩阵, 故对称地, 我们也会有 $\operatorname{rank}(\boldsymbol{A}) \leq \operatorname{rank}(\boldsymbol{EA})$. 从而得到 $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{EA})$. 由于初等行变换与左乘初等矩阵是等同的, 这说明了初等行变换不改变矩阵的秩.

下面来证明不等式 (4.9). 下面设 $\operatorname{rank}(\boldsymbol{A}) = r$. 此时只需证明 $\boldsymbol{EA}$ 的所有 $r+1$ 阶子式为 0. 由于 $\boldsymbol{A}$ 的所有 $r+1$ 阶子式为 0, 由 Binet–Cauchy 公式 (由于这儿的 $\boldsymbol{E}$ 非常简单, 也可以分情况直接计算验证; 参见教材的引理 4.5.1) 可知, 该论断成立. □

推论 4.5.7. 设 $\boldsymbol{A}$ 为 $m \times n$ 矩阵, $\boldsymbol{P}$ 为 m 阶可逆矩阵, $\boldsymbol{Q}$ 为 n 阶可逆矩阵, 则 $\operatorname{rank}(\boldsymbol{PAQ}) = \operatorname{rank}(\boldsymbol{A})$.

证明. 仅需注意以下事实: (1) 可逆矩阵是初等矩阵的乘积; (2) 初等行变换等同于左乘初等矩阵; (3) 初等列变换等同于右乘初等矩阵. □

求矩阵的秩的常用方法 对于矩阵 $\boldsymbol{A}$, 在 $\boldsymbol{A}$ 的维数比较大时, 若用定义来求解 $\boldsymbol{A}$ 的秩, 则计算量非常大, 从而并不实用. 由于我们已有定理 4.5.6 和例 4.5.5, 一般地, 我们会通过初等行变换, 将 $\boldsymbol{A}$ 化为阶梯形矩阵 $\boldsymbol{B}$, 则 $\boldsymbol{B}$ 的非零行的行数 r 就是 $\boldsymbol{A}$ 的秩. 另外, 在求秩的时候, 没有必要计算其约化标准形, 阶梯形标准形就够用了.

例 4.5.8. *设有 4×5 矩阵 $\boldsymbol{A}$ 通过一系列初等行变换后得到:*

$$\boldsymbol{A}=\begin{pmatrix}1&-1&0&1&1\\2&-2&0&2&2\\1&1&1&0&0\\2&0&1&1&1\end{pmatrix}\xrightarrow[\substack{-r_1\to r_3\\-2r_1\to r_4}]{-2r_1\to r_2}\begin{pmatrix}1&-1&0&1&1\\0&0&0&0&0\\0&2&1&-1&-1\\0&2&1&-1&-1\end{pmatrix}\to\begin{pmatrix}1&-1&0&1&1\\0&2&1&-1&-1\\0&0&0&0&0\\0&0&0&0&0\end{pmatrix}\triangleq\boldsymbol{B}.$$

故 $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{B})=2$.

定义 4.5.9. *若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 同为 $m\times n$ 矩阵, 且存在可逆的 m 阶矩阵 $\boldsymbol{P}$ 和可逆的 n 阶矩阵 $\boldsymbol{Q}$ 使得 $\boldsymbol{B}=\boldsymbol{PAQ}$, 则称 $\boldsymbol{A}$ 和 $\boldsymbol{B}$* **相抵** *(也称为***等价***, equivalent). 不难验证, 相抵关系是一个***等价关系***, 即满足*

(**反身性**) *$\boldsymbol{A}$ 与自身相抵;*

(**对称性**) *若 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相抵, 则 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 也相抵;*

(**传递性**) *$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相抵, 且 $\boldsymbol{B}$ 与 $\boldsymbol{C}$ 相抵, 则 $\boldsymbol{A}$ 与 $\boldsymbol{C}$ 也相抵.*

*所有的 $m\times n$ 矩阵依照相抵关系分为不同的***相抵等价类** *(同一相抵等价类的矩阵互相相抵, 两个不同相抵等价类的矩阵互不相抵).*

前面的推论 4.5.7 说明: 若 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相抵, 则 $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{B})$.

另一方面, 定理 4.4.2 说明, 矩阵 $\boldsymbol{A}$ 必与某个 $\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O})$ 相抵, 从而 $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O}))=r$. 这说明 $\boldsymbol{A}$ 所在的相抵等价类里存在唯一的形如 $\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O})$ 的矩阵, 其中的 r 必为 $\operatorname{rank}(\boldsymbol{A})$. 该矩阵 $\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O})$ 称为 $\boldsymbol{A}$ 的**相抵标准形**.

由此, 对于同型的矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$, 若 $\operatorname{rank}(\boldsymbol{A})=r=\operatorname{rank}(\boldsymbol{B})$, 则 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 同时相抵于 $\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O})$, 从而 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相抵.

综上, 我们证明了

定理 4.5.10. *同型的矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相抵的充要条件是 $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{B})$.* 教材定理 4.5.2

例 4.5.11. *检验矩阵*

$$\boldsymbol{A}=\begin{pmatrix}-1&0&1&1\\1&-2&-1&-3\\2&-1&-2&-3\end{pmatrix}\quad\text{和}\quad\boldsymbol{B}=\begin{pmatrix}2&-1&7&0\\0&-1&1&2\\1&-1&4&1\end{pmatrix}$$

是否相抵.

解. 考虑 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的阶梯标准形, 我们通过初等行变换可以得到

$$\boldsymbol{A} \to \begin{pmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \boldsymbol{B} \to \begin{pmatrix} 1 & -\frac{1}{2} & \frac{7}{2} & 0 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

这说明 $\operatorname{rank}(\boldsymbol{A}) = 2 = \operatorname{rank}(\boldsymbol{B})$, 从而 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相抵. □

例 4.5.12. 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为任意矩阵, 证明:

教材例题 4.5.2, 公式需要牢记

$$\operatorname{rank}\begin{pmatrix} \boldsymbol{A} & \\ & \boldsymbol{B} \end{pmatrix} = \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}).$$

证明. 设 $\operatorname{rank}(\boldsymbol{A}) = r$, $\operatorname{rank}(\boldsymbol{B}) = s$, 则存在可逆矩阵 $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{Q}_1, \boldsymbol{Q}_2$ (阶数可能不一致) 使得 $\boldsymbol{P}_1\boldsymbol{A}\boldsymbol{Q}_1 = \operatorname{diag}(\boldsymbol{I}_r, \boldsymbol{O})$, $\boldsymbol{P}_2\boldsymbol{B}\boldsymbol{Q}_2 = \operatorname{diag}(\boldsymbol{I}_s, \boldsymbol{O})$. 此时,

$$\begin{pmatrix} \boldsymbol{P}_1 & \\ & \boldsymbol{P}_2 \end{pmatrix}\begin{pmatrix} \boldsymbol{A} & \\ & \boldsymbol{B} \end{pmatrix}\begin{pmatrix} \boldsymbol{Q}_1 & \\ & \boldsymbol{Q}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{P}_1\boldsymbol{A}\boldsymbol{Q}_1 & \\ & \boldsymbol{P}_2\boldsymbol{B}\boldsymbol{Q}_2 \end{pmatrix} = \operatorname{diag}(\boldsymbol{I}_r, \boldsymbol{O}, \boldsymbol{I}_s, \boldsymbol{O}).$$

通过初等变换 (行变换和列变换都需要), 不难看出由 $\operatorname{diag}(\boldsymbol{I}_r, \boldsymbol{O}, \boldsymbol{I}_s, \boldsymbol{O})$ 可以得到 $\operatorname{diag}(\boldsymbol{I}_r, \boldsymbol{I}_s, \boldsymbol{O}, \boldsymbol{O})$, 而后者即为 $\operatorname{diag}(\boldsymbol{I}_{r+s}, \boldsymbol{O})$, 它的秩恰为 $r+s$. □

例 4.5.13. 对于矩阵

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & b \\ 2 & 3 & a & 4 \\ 3 & 5 & 1 & 7 \end{pmatrix},$$

若已知 $\operatorname{rank}(\boldsymbol{A}) = 3$, 求出 a 和 b.

解. 做初等变换

$$\boldsymbol{A} \xrightarrow[-3r_1 \to r_4]{-2r_1 \to r_3} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & b \\ 0 & 1 & a-2 & 2 \\ 0 & 2 & -2 & 4 \end{pmatrix} \xrightarrow[-2r_2 \to r_4]{-r_2 \to r_3} \left(\begin{array}{cc:cc} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & b \\ \hdashline 0 & 0 & a-1 & 2-b \\ 0 & 0 & 0 & 4-2b \end{array}\right)$$

$$\xrightarrow{-c_1\boldsymbol{P} \to c_2} \left(\begin{array}{cc:cc} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hdashline 0 & 0 & a-1 & 2-b \\ 0 & 0 & 0 & 4-2b \end{array}\right) \xlongequal{\text{记作}} \boldsymbol{B},$$

其中的 $\boldsymbol{P}=\begin{pmatrix}1&1\\0&1\end{pmatrix}^{-1}\begin{pmatrix}1&1\\-1&b\end{pmatrix}$. 这说明

$$3=\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{B})=\operatorname{rank}\begin{pmatrix}1&1\\0&1\end{pmatrix}+\operatorname{rank}\begin{pmatrix}a-1&2-b\\0&4-2b\end{pmatrix}.$$

由于 $\operatorname{rank}\begin{pmatrix}1&1\\0&1\end{pmatrix}=2$, 这说明

$$1=\operatorname{rank}\begin{pmatrix}a-1&2-b\\0&4-2b\end{pmatrix}\xlongequal{\text{初等行变换}}\operatorname{rank}\begin{pmatrix}a-1&0\\0&4-2b\end{pmatrix}=\operatorname{rank}(a-1)+\operatorname{rank}(4-2b),$$

(其中后面的两个是一阶方阵的秩) 即 $a=1$, $b\neq 2$ 或者 $a\neq 1$, $b=2$. □

例 4.5.14. *对于矩阵 $\boldsymbol{A}\in F^{m\times n}$, $\boldsymbol{B}\in F^{n\times p}$, 证明:*

教材例题 4.5.7, 公式需要牢记

$$\operatorname{rank}(\boldsymbol{AB})\leq\min(\operatorname{rank}(\boldsymbol{A}),\operatorname{rank}(\boldsymbol{B})).$$

证明. **(思路一)** 教材上的解法, 利用分块 + 相抵标准形的方法. 要求学生体会其中的分块技巧, 及分块所带来的便利性. 设 $\operatorname{rank}(\boldsymbol{A})=r$, $\operatorname{rank}(\boldsymbol{B})=s$, 则存在可逆矩阵 $\boldsymbol{P}_1,\boldsymbol{P}_2,\boldsymbol{Q}_1,\boldsymbol{Q}_2$ (阶数可能不一致) 使得 $\boldsymbol{A}=\boldsymbol{P}_1\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O})\boldsymbol{Q}_1$, $\boldsymbol{B}=\boldsymbol{P}_2\operatorname{diag}(\boldsymbol{I}_s,\boldsymbol{O})\boldsymbol{Q}_2$. 将 $\boldsymbol{Q}_1\boldsymbol{P}_2$ 写成分块矩阵 $(\boldsymbol{R}_{ij})_{2\times 2}$, 其中 $\boldsymbol{R}_{11}$ 为 $r\times s$ 矩阵. 此时有

$$\boldsymbol{AB}=\boldsymbol{P}_1\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O})\boldsymbol{Q}_1\boldsymbol{P}_2\operatorname{diag}(\boldsymbol{I}_s,\boldsymbol{O})\boldsymbol{Q}_2=\boldsymbol{P}_1\operatorname{diag}(\boldsymbol{R}_{11},\boldsymbol{O})\boldsymbol{Q}_2.$$

于是

$$\operatorname{rank}(\boldsymbol{AB})=\operatorname{rank}(\operatorname{diag}(\boldsymbol{R}_{11},\boldsymbol{O}))=\operatorname{rank}(\boldsymbol{R}_{11})\leq\min(r,s).$$

(思路二) 利用 Binet–Cauchy 公式. □

例 4.5.15.

教材 P116, 作业题 #41, 公式需要牢记

(1) (i) $\operatorname{rank}(\boldsymbol{A}\ \boldsymbol{B})\geq\operatorname{rank}\left((\boldsymbol{A}\ \boldsymbol{B})\begin{pmatrix}\boldsymbol{I}\\\boldsymbol{O}\end{pmatrix}\right)=\operatorname{rank}(\boldsymbol{A})$.

(ii) $\operatorname{rank}(\boldsymbol{A}\ \boldsymbol{B})\geq\operatorname{rank}\left((\boldsymbol{A}\ \boldsymbol{B})\begin{pmatrix}\boldsymbol{O}\\\boldsymbol{I}\end{pmatrix}\right)=\operatorname{rank}(\boldsymbol{B})$.

(iii) $\operatorname{rank}(\boldsymbol{A}\ \boldsymbol{B})\geq\operatorname{rank}\left((\boldsymbol{A}\ \boldsymbol{B})\begin{pmatrix}\boldsymbol{I}\\\boldsymbol{I}\end{pmatrix}\right)=\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B})$.

(2)

$$\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})=\operatorname{rank}\begin{pmatrix}\boldsymbol{A} & \\ & \boldsymbol{B}\end{pmatrix}\geq\operatorname{rank}\left((\boldsymbol{I}\ \boldsymbol{I})\begin{pmatrix}\boldsymbol{A} & \\ & \boldsymbol{B}\end{pmatrix}\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{I}\end{pmatrix}\right)$$
$$=\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B}).$$

(3) $\operatorname{rank}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{C}\\ \boldsymbol{O} & \boldsymbol{B}\end{pmatrix}\geq\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})$. 为了证明该不等式, 我们利用定义. 假定 $\operatorname{rank}(\boldsymbol{A})=r$, $\operatorname{rank}(\boldsymbol{B})=s$. 从而存在子式

$$\det\left(\boldsymbol{A}\begin{pmatrix}i_1' & i_2' & \dots & i_r'\\ j_1' & j_2' & \dots & j_r'\end{pmatrix}\right)\neq 0,\qquad \det\left(\boldsymbol{B}\begin{pmatrix}i_1'' & i_2'' & \dots & i_s''\\ j_1'' & j_2'' & \dots & j_s''\end{pmatrix}\right)\neq 0.$$

若设 $\boldsymbol{A}\in F^{p\times q}$, 那么由此可以看出 $\left(\begin{smallmatrix}\boldsymbol{A} & \boldsymbol{C}\\ \boldsymbol{O} & \boldsymbol{B}\end{smallmatrix}\right)$ 的子式

$$\det\left(\begin{pmatrix}\boldsymbol{A} & \boldsymbol{C}\\ \boldsymbol{O} & \boldsymbol{B}\end{pmatrix}\begin{pmatrix}i_1' & i_2' & \dots & i_r' & p+i_1'' & p+i_2'' & \dots & p+i_s''\\ j_1' & j_2' & \dots & j_r' & q+j_1'' & q+j_2'' & \dots & q+j_s''\end{pmatrix}\right)$$
$$\xlongequal{\text{是准上三角分块矩阵的行列式}}\det\left(\boldsymbol{A}\begin{pmatrix}i_1' & i_2' & \dots & i_r'\\ j_1' & j_2' & \dots & j_r'\end{pmatrix}\right)\det\left(\boldsymbol{B}\begin{pmatrix}i_1'' & i_2'' & \dots & i_s''\\ j_1'' & j_2'' & \dots & j_s''\end{pmatrix}\right)\neq 0.$$

这说明 $\operatorname{rank}\left(\begin{smallmatrix}\boldsymbol{A} & \boldsymbol{C}\\ \boldsymbol{O} & \boldsymbol{B}\end{smallmatrix}\right)\geq r+s=\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})$.

习题 4.5.16. 设 $\boldsymbol{A}\in F^{n\times m}$, $\boldsymbol{B}\in F^{m\times p}$. 若 $m=\operatorname{rank}(\boldsymbol{AB})$, 证明: $m=\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{B})$.

例 4.5.17 (**Sylvester 不等式**). 对于矩阵 $\boldsymbol{A}\in F^{m\times n}$, $\boldsymbol{B}\in F^{n\times p}$, 我们有不等式

公式需要牢记

$$\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})-n\leq\operatorname{rank}(\boldsymbol{AB}).$$

证明.

$$\begin{aligned}
n+\operatorname{rank}(\boldsymbol{AB}) &= \operatorname{rank}(\boldsymbol{I}_n)+\operatorname{rank}(\boldsymbol{AB})=\operatorname{rank}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{AB}\end{pmatrix}\\
&\xlongequal{-c_1\boldsymbol{B}\to c_2}\operatorname{rank}\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{B}\\ \boldsymbol{O} & \boldsymbol{AB}\end{pmatrix}\xlongequal{\boldsymbol{A}r_1\to r_2}\operatorname{rank}\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{B}\\ \boldsymbol{A} & \boldsymbol{O}\end{pmatrix}\\
&\underbrace{\geq}_{\text{教材作业 }\#41(3)}\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B}).
\end{aligned}$$

□

更一般地, 我们有如下的不等式. 其证明方法是类似的.

习题 4.5.18 (**Frobenius 不等式**). *假定 $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$ 分别为 $n\times m$, $m\times p$, $p\times q$ 矩阵. 试证: $\operatorname{rank}(\boldsymbol{AB})+\operatorname{rank}(\boldsymbol{BC})\leq\operatorname{rank}(\boldsymbol{B})+\operatorname{rank}(\boldsymbol{ABC})$. 特别地, $\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})\leq m+\operatorname{rank}(\boldsymbol{AB})$, 以及当 $\operatorname{rank}(\boldsymbol{AB})=\operatorname{rank}(\boldsymbol{B})$ 时, 有 $\operatorname{rank}(\boldsymbol{BC})=\operatorname{rank}(\boldsymbol{ABC})$. (提示: 研究分块矩阵 $\operatorname{diag}(\boldsymbol{B},\boldsymbol{ABC})$)*

作为应用, 我们有

习题 4.5.19. *设 $\boldsymbol{A}$ 为 n 阶方阵, 并存在正整数 N 使得 $\operatorname{rank}(\boldsymbol{A}^N)=\operatorname{rank}(\boldsymbol{A}^{N+1})$. 证明: $\operatorname{rank}(\boldsymbol{A}^N)=\operatorname{rank}(\boldsymbol{A}^{N+1})=\operatorname{rank}(\boldsymbol{A}^{N+2})=\cdots$. (提示: Frobenius 不等式)*

例 4.5.20. *设 n 阶方阵 $\boldsymbol{A}$ 为*幂等矩阵*, 即 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=\boldsymbol{A}$. 证明: $\operatorname{tr}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A})$.*

证明. 设 $\operatorname{rank}(\boldsymbol{A})=r$. 由相抵标准形可知, 存在可逆的同阶方阵 $\boldsymbol{P},\boldsymbol{Q}$ 使得 $\boldsymbol{PAQ}=\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O}_{n-r})$. 此时条件 $\boldsymbol{A}^2=\boldsymbol{A}$ 等价于

$$\boldsymbol{P}^{-1}\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}\boldsymbol{Q}^{-1}\boldsymbol{P}^{-1}\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}\boldsymbol{Q}^{-1}=\boldsymbol{P}^{-1}\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}\boldsymbol{Q}^{-1},$$

即

$$\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}\boldsymbol{Q}^{-1}\boldsymbol{P}^{-1}\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}.$$

设方阵 $\boldsymbol{Q}^{-1}\boldsymbol{P}^{-1}=\begin{pmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{pmatrix}$ 为分块的形式, 其中 $\boldsymbol{R}_{11}$ 为 r 阶方阵. 代入上面的等式, 我们解得 $\boldsymbol{R}_{11}=\boldsymbol{I}_r$. 此时,

$$\operatorname{tr}(\boldsymbol{A})=\operatorname{tr}\left(\boldsymbol{P}^{-1}\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}\boldsymbol{Q}^{-1}\right)=\operatorname{tr}\left(\begin{pmatrix}\boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r}\end{pmatrix}\boldsymbol{Q}^{-1}\boldsymbol{P}^{-1}\right)$$

$$= \operatorname{tr}\left(\begin{pmatrix} \boldsymbol{I}_r & \\ & \boldsymbol{O}_{n-r} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{pmatrix}\right) = \operatorname{tr}\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{R}_{12} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} = r = \operatorname{rank}(\boldsymbol{A}). \qquad \square$$

例 4.5.21. 对于 n 阶方阵 $\boldsymbol{A}$, 证明: 教材例题 4.5.8

$$\boldsymbol{A}^2 = \boldsymbol{A} \qquad \Longleftrightarrow \qquad \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I} - \boldsymbol{A}) = n.$$

证明. **(思路一)** 作分块矩阵的初等变换, 我们有

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I}-\boldsymbol{A} \end{pmatrix} \xrightarrow{r_1 \to r_2} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{O} \\ \boldsymbol{A} & \boldsymbol{I}-\boldsymbol{A} \end{pmatrix} \xrightarrow{c_1 \to c_2} \begin{pmatrix} \boldsymbol{A} & \boldsymbol{A} \\ \boldsymbol{A} & \boldsymbol{I} \end{pmatrix}$$
$$\xrightarrow{-\boldsymbol{A}r_2 \to r_1} \begin{pmatrix} \boldsymbol{A}-\boldsymbol{A}^2 & \boldsymbol{O} \\ \boldsymbol{A} & \boldsymbol{I} \end{pmatrix} \xrightarrow{-c_2\boldsymbol{A} \to c_1} \begin{pmatrix} \boldsymbol{A}-\boldsymbol{A}^2 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{I} \end{pmatrix}.$$

这说明 $\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I}-\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}-\boldsymbol{A}^2) + \operatorname{rank}(\boldsymbol{I}) = \operatorname{rank}(\boldsymbol{A}-\boldsymbol{A}^2) + n$. 从而 $\boldsymbol{A} = \boldsymbol{A}^2$ 当且仅当 $\operatorname{rank}(\boldsymbol{A}-\boldsymbol{A}^2) = 0$, 当且仅当 $\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I}-\boldsymbol{A}) = n$.

(思路二) 先证明 "$\Rightarrow$". 由于 $\boldsymbol{A}^2 = \boldsymbol{A}$, 我们有等式 $\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A}) = \boldsymbol{O}$. 由此, 我们有

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I}-\boldsymbol{A}) - n \underbrace{\leq}_{\text{例 4.5.17}} \operatorname{rank}(\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{O}) = 0,$$

这说明 $\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I}-\boldsymbol{A}) \leq n$. 另一方面,

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I}-\boldsymbol{A}) \underbrace{\geq}_{\text{作业 \#41(1)(2)}} \operatorname{rank}(\boldsymbol{A} + (\boldsymbol{I}-\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{I}) = n.$$

综上即有 $\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I}-\boldsymbol{A}) = n$.

再证明 "$\Leftarrow$". 设 $\operatorname{rank}(\boldsymbol{A}) = r$. 由相抵标准形可知, 存在可逆的同阶方阵 $\boldsymbol{P}, \boldsymbol{Q}$ 使得 $\boldsymbol{PAQ} = \operatorname{diag}(\boldsymbol{I}_r, \boldsymbol{O}_{n-r})$. 那么,

$$\boldsymbol{I} - \boldsymbol{A} = \boldsymbol{I} - \boldsymbol{P}^{-1}\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}\boldsymbol{Q}^{-1} = \boldsymbol{Q}\left(\boldsymbol{I} - \boldsymbol{Q}^{-1}\boldsymbol{P}^{-1}\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}\right)\boldsymbol{Q}^{-1}.$$

若令 $\boldsymbol{H} = \boldsymbol{Q}^{-1}\boldsymbol{P}^{-1} = \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} \\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22} \end{pmatrix}$, 其中 $\boldsymbol{H}_{11}$ 为 r 阶方阵, 则有

$$n - r = \operatorname{rank}(\boldsymbol{I}-\boldsymbol{A}) = \operatorname{rank}\left(\boldsymbol{I} - \boldsymbol{H}\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}\right) = \operatorname{rank}\left(\boldsymbol{I} - \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{O} \\ \boldsymbol{H}_{21} & \boldsymbol{O} \end{pmatrix}\right)$$

$$= \operatorname{rank}\begin{pmatrix} I_r - H_{11} & O \\ -H_{21} & I_{n-r} \end{pmatrix} \xrightarrow{c_2 H_{21} \to c_1} \operatorname{rank}\begin{pmatrix} I_r - H_{11} & O \\ O & I_{n-r} \end{pmatrix}.$$

由于 $\operatorname{rank}(I_{n-r}) = n - r$, 这说明 $\operatorname{rank}(I_r - H_{11}) = 0$, 即 $H_{11} = I_r$. 此时,

$$\begin{aligned} A^2 &= P^{-1}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Q^{-1} \\ &= P^{-1}\begin{pmatrix} H_{11} & O \\ O & O \end{pmatrix} Q^{-1} = P^{-1}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} Q^{-1} = A. \end{aligned}$$

□

例 4.5.22. *设 A, B 为 n 阶方阵, 满足 $A^2 = A$, $B^2 = B$, 且 $I - A - B$ 可逆. 证明: $\operatorname{rank}(A) = \operatorname{rank}(B)$.*

证明. **(思路一)** 由于 $I - A - B$ 可逆, 我们有

$$n = \operatorname{rank}(I - A - B) \le \operatorname{rank}(I - A) + \operatorname{rank}(-B) = \operatorname{rank}(I - A) + \operatorname{rank}(B).$$

由前面的例 4.5.21, 我们已经知道了 $\operatorname{rank}(A) + \operatorname{rank}(I - A) = n$, 从而 $\operatorname{rank}(A) \le \operatorname{rank}(B)$. 但是很明显, 由对称性, 我们可以得到 $\operatorname{rank}(B) \le \operatorname{rank}(A)$. 从而得到 $\operatorname{rank}(A) = \operatorname{rank}(B)$.

(思路二) 通过直接计算, 我们有等式

$$(I - A - B)B = -AB = A(I - A - B).$$

由于乘以可逆矩阵不改变矩阵的秩, 我们立刻得到 $\operatorname{rank}(A) = \operatorname{rank}(B)$. □

例 4.5.23. *设 n 阶方阵 A 为***对合矩阵***, 即 A 满足 $A^2 = I$. 求方阵 $\operatorname{diag}(I + A, I - A)$ 的相抵标准形.* P116#38, #43

解. 问题归结于求矩阵的秩 $r := \operatorname{rank}(\operatorname{diag}(I + A, I - A))$. 由于 $A = \pm I$ 都满足 $A^2 = I$, 且对应的矩阵的秩 $r = n$, 因此, 我们有理由猜测 $r = n$ 对于所有的 $A^2 = I$ 都成立. 下面来证明这一结论.

一方面, 我们有

$$\operatorname{rank}(\operatorname{diag}(I + A, I - A)) \ge \operatorname{rank}((I + A) + (I - A)) = \operatorname{rank}(2I) = n,$$

另一方面, 我们有

$$\begin{aligned} \operatorname{rank}(\operatorname{diag}(I + A, I - A)) &= \operatorname{rank}(I + A) + \operatorname{rank}(I - A) \\ &\le \operatorname{rank}((I + A)(I - A)) + n = \operatorname{rank}(O) + n = n, \end{aligned}$$

这说明确实有 $r = n$, 从而欲求的相抵标准形为 $\operatorname{diag}(I_n, O_n)$. □

定义 4.5.24. (1) 设矩阵 $\boldsymbol{A} \in F^{m\times n}$. 若 $\operatorname{rank}(\boldsymbol{A}) = m$, 则称 $\boldsymbol{A}$ 是**行满秩**的. 若 $\operatorname{rank}(\boldsymbol{A}) = n$, 则称 $\boldsymbol{A}$ 是**列满秩**的. (注意: 对于 $m \times n$ 矩阵 $\boldsymbol{A}$, 我们总有 $\operatorname{rank}(\boldsymbol{A}) \le \min(m, n)$)

(2) 可逆的方阵又称为**满秩矩阵**, 不可逆的方阵又称为**降秩矩阵**.

例 4.5.25. 设矩阵 $\boldsymbol{A} \in F^{m\times n}$ 是行满秩的, 即 $m = \operatorname{rank}(\boldsymbol{A}) \le n$. 证明: 存在矩阵 $\boldsymbol{B} \in F^{n\times m}$, 使得 $\boldsymbol{AB} = \boldsymbol{I}_m$.

证明. 设可逆矩阵 $\boldsymbol{P}, \boldsymbol{Q}$ 满足 $\boldsymbol{PAQ} = (\boldsymbol{I}_m, \boldsymbol{O})$, 即 $\boldsymbol{A} = \boldsymbol{P}^{-1}(\boldsymbol{I}_m, \boldsymbol{O})\boldsymbol{Q}^{-1}$, 那么可以取 $\boldsymbol{B} = \boldsymbol{Q}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{O} \end{pmatrix}\boldsymbol{P}$ 即可. □

例 4.5.26. 设 $\boldsymbol{A} \in F^{m\times k}$, $\boldsymbol{B} \in F^{k\times n}$ 都是列满秩的矩阵. 证明: $\boldsymbol{AB}$ 也是列满秩的矩阵.

证明. 由于列满秩的条件, 我们有 $\operatorname{rank}(\boldsymbol{A}) = k$, $\operatorname{rank}(\boldsymbol{B}) = n$. 另外, 我们有不等式

$$k + n - k = \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) - k \le \operatorname{rank}(\boldsymbol{AB}) \le \operatorname{rank}(\boldsymbol{B}) = n,$$

从而 $\operatorname{rank}(\boldsymbol{AB}) = n$, 即 $\boldsymbol{AB}$ 列满秩. □

例 4.5.27. (1) (**矩阵的满秩分解定理**) 对于矩阵 $\boldsymbol{A} \in F^{m\times n}$, 若 $\operatorname{rank}(\boldsymbol{A}) = r > 0$, 则存在列满秩的矩阵 $\boldsymbol{G} \in F^{m\times r}$ 和行满秩的矩阵 $\boldsymbol{H} \in F^{r\times n}$, 使得 $\boldsymbol{A} = \boldsymbol{GH}$.

(2) 上面的分解定理反过来也成立, 即对于列满秩的矩阵 $\boldsymbol{G} \in F^{m\times r}$ 和行满秩的矩阵 $\boldsymbol{H} \in F^{r\times n}$, 我们有 $\operatorname{rank}(\boldsymbol{GH}) = r$.

证明. (1) 由 $\boldsymbol{A}$ 的相抵标准形知存在可逆方阵 $\boldsymbol{P} \in F^{m\times m}$ 和 $\boldsymbol{Q} \in F^{n\times n}$, 使得

$$\boldsymbol{PAQ} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_r \\ \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \end{pmatrix}.$$

此时,

$$\boldsymbol{A} = \underbrace{\left(\boldsymbol{P}^{-1}\begin{pmatrix} \boldsymbol{I}_r \\ \boldsymbol{O} \end{pmatrix}\right)}_{\boldsymbol{G}} \underbrace{\left(\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \end{pmatrix}\boldsymbol{Q}^{-1}\right)}_{\boldsymbol{H}}.$$

其中, 由于乘一个可逆矩阵不改变矩阵的秩, 我们有

$$\operatorname{rank}(\boldsymbol{G}) = \operatorname{rank}\left(\boldsymbol{P}^{-1}\begin{pmatrix} \boldsymbol{I}_r \\ \boldsymbol{O} \end{pmatrix}\right) = \operatorname{rank}\begin{pmatrix} \boldsymbol{I}_r \\ \boldsymbol{O} \end{pmatrix} = r.$$

类似地, $\operatorname{rank}(\boldsymbol{H}) = r$.

(2) 一方面, 我们有 $\operatorname{rank}(\boldsymbol{GH}) \le \operatorname{rank}(\boldsymbol{H}) = r$. 另一方面, 我们有

$$r = r + r - r = \operatorname{rank}(\boldsymbol{G}) + \operatorname{rank}(\boldsymbol{H}) - r \le \operatorname{rank}(\boldsymbol{GH}). \qquad \square$$

注 4.5.28. *学生课后自学教材例题 4.5.5, 并注意到: 每个秩为 r 的矩阵不能表示为少于 r 个秩为 1 的矩阵的和.*

习题 4.5.29. *设方阵 $\boldsymbol{A} = (a_{ij})_{n\times n}$, 而 $\operatorname{rank}(\boldsymbol{A}) \le 1$.*

(1) *对于正整数 s, 求 $\boldsymbol{A}^s$.* (提示: 用满秩分解定理)

(2) *求行列式 $\det(\boldsymbol{I}+\boldsymbol{A})$.* (提示: 用 Sylvester 行列式公式, 即教材习题 #25)

习题 4.5.30. *设矩阵 $\boldsymbol{A} \in F^{3\times 2}$ 和 $\boldsymbol{B} \in F^{2\times 3}$ 满足 $\boldsymbol{AB} = \begin{pmatrix} 8 & 2 & -2 \\ 2 & 5 & 4 \\ -2 & 4 & 5 \end{pmatrix}$. 证明: $\boldsymbol{BA} = \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$.* (提示: 教材习题 #42)

第五章 线性空间

5.1 数组空间及其子空间

在这一章中 F 总是一个数域, 例如 $F=\mathbb{Q}$ (有理数域), $\mathbb{R}$ (实数域) 或 $\mathbb{C}$ (复数域). 正如在第一章中所约定的, $F^n \coloneqq \{\, (a_1,a_2,\ldots,a_n) \mid a_i \in F \,\}$ 是 F 上的一个 n 维数组空间. F^n 中的元素依照具体情形可以写成行向量或列向量的形式 (如果没有具体要求, 建议将之视为列向量). F^n 中的加法与数乘是对应坐标的操作. 若 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m \in F^m$, $\lambda_1,\lambda_2,\ldots,\lambda_m \in F$, 则称 $\boldsymbol{\alpha} \coloneqq \sum_{i=1}^m \lambda_i\boldsymbol{\alpha}_i$ 是 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 的一个**线性组合** (**线性表示**). 在本章中, 我们也会用 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m\}$ 来不规范地表示向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m$. 相应地, 这儿的 m 称为该向量组的**长度**; 注意这并不是指单个向量的长度 n.

更一般地, 我们引入如下的抽象概念.

定义 5.1.1. *设 V 是 F^n 的一个非空子集, 满足*

(1) *对任意的 $\boldsymbol{\alpha},\boldsymbol{\beta} \in V$, 有 $\boldsymbol{\alpha}+\boldsymbol{\beta} \in V$,*

(2) *对任意的 $\boldsymbol{\alpha} \in V$ 和任意的 $\lambda \in F$ 有 $\lambda\boldsymbol{\alpha} \in V$,*

则称 V 是 F^n 的一个 (**线性**)**子空间**. *等价地, F^n 的非空子集 V 是一个***子空间***当且仅当对任意正整数 m, 对任意数组向量 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 和任意标量 $\lambda_1,\ldots,\lambda_m \in F$, 有线性组合 $\sum_{i=1}^m \lambda_i\boldsymbol{\alpha}_i \in V$.*

例 5.1.2. **零子空间**$\{\mathbf{0}=(0,0,\ldots,0)\}$ *与***全空间** *F^n 是 F^n 的两个***平凡子空间**.

从定义出发, 我们很容易验证以下的重要事实.

例 5.1.3. *给定 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m \in F^n$, 则*

$$\operatorname{Span}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m) = \langle \boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m \rangle \coloneqq \left\{ \sum_{i=1}^m \lambda_i\boldsymbol{\alpha}_i \,\middle|\, \lambda_i \in F \right\}$$

是 F^n 的一个线性子空间. 我们将称该子空间为由 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ **生成的子空间***或***张成的子空间** *(spanning subspace), 并称 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 为该子空间的一组***生成元**.

例 5.1.4. 在三维空间 $\mathbb{R}^3$ 中, 坐标轴和坐标平面都是子空间. 其中, xy 坐标平面

$$Oxy \coloneqq \{\, (x, y, 0) \mid x, y \in \mathbb{R} \,\}$$

是由 $(1,0,0)$ 和 $(0,1,0)$ 生成的子空间.

注 5.1.5. (1) $\langle a_1, a_2, \ldots, a_m \rangle$ 是可由 $a_1, a_2, \ldots, a_m$ 线性表示的向量的全体.

(2) 由向量组生成子空间的生成元一般并不唯一. 例如, Oxy 平面也可以由 $(1,0,0)$ 和 $(1,1,0)$ 生成. 另一方面, 可以证明, 任意的 F^n 的子空间都可以由有限多个向量生成.

(3) 可以证明, 由生成元生成的子空间其实是包含这些生成元的最小的子空间, 并且任何子空间都包含零向量 $\mathbf{0}$. 由此, 我们约定: 零空间 $\{\mathbf{0}\}$ 是由零向量生成的, 也可以视作例 *5.1.3* 中 $n = 0$ 时的退化情形, 即零空间是由空集生成的子空间.

例 5.1.6. 考虑三维空间 $\mathbb{R}^3$ 中满足 $x + y + z = 1$ 的点的集合 V. 几何上, 这是一个过 $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ 这三个点的平面. 由于 $\mathbf{0} \notin V$, V 不是一个子空间. 换一个角度, 由于向量的加法 $(1,0,0) + (0,1,0) \notin V$, 我们同样也看出 V 不是一个子空间.

例 5.1.7. 在三维空间 $\mathbb{R}^3$ 中的直线和平面是子空间的充要条件是它们过原点. 平凡子空间 (0 维和 3 维) 和过原点的直线 (1 维)、平面 (2 维) 是 $\mathbb{R}^3$ 中的所有子空间.

5.2 线性相关与线性无关

定义 5.2.1. 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m \in F^n$. 若存在不全为零的标量 $\lambda_1,\dots,\lambda_m \in F$, 使得 $\sum_{i=1}^m \lambda_i\boldsymbol{\alpha}_i = \mathbf{0}$, 则称向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ **线性相关**; 否则, 称它们**线性无关**.

注 5.2.2. (1) 若 $m=1$, 则 $\boldsymbol{\alpha}_1$ 线性无关的充要条件是 $\boldsymbol{\alpha}_1 \neq \mathbf{0}$.

(2) 若 $m \geq 2$, 向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m \in F^n$ 线性相关的充要条件是向量组中某个向量可以被剩下的向量组线性表示. 下面给出简要证明.

设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 线性相关, $\sum_{i=1}^m \lambda_i\boldsymbol{\alpha}_i = \mathbf{0}$, 其中 $\lambda_{i_0} \neq 0$, 则 $\boldsymbol{\alpha}_{i_0} = -\sum_{i\neq i_0}\frac{\lambda_i}{\lambda_{i_0}}\boldsymbol{\alpha}_i$, 即 $\boldsymbol{\alpha}_{i_0}$ 可以由其它向量线性表示.

反之, 在向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 中, 若某个 $\boldsymbol{\alpha}_{i_0}$ 可以由其它的向量线性表示: $\boldsymbol{\alpha}_{i_0} = \sum_{i\neq i_0}\lambda_i\boldsymbol{\alpha}_i$, 则 $\sum_{i=1}^m \lambda_i\boldsymbol{\alpha}_i = \mathbf{0}$, 其中补充定义 $\lambda_{i_0} = -1$. 故, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 线性相关.

例 5.2.3. 包含零向量的向量组一定线性相关, 这是因为向量组长度至少为 2 时该零向量是其余向量的以零为系数的线性组合即可, 而向量组长度为 1 时也是显然的.

定义 5.2.4. 线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 中的方程称为**线性相关** (或**线性无关**) 当且仅当其对应的增广矩阵 $(\boldsymbol{A}\ \boldsymbol{b})$ 中对应的行向量组线性相关 (或线性无关).

故线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 中的方程线性相关的充要条件是该系统中有冗余的方程.

定理 5.2.5. 设向量组 $S_1 = \{\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\dots,\boldsymbol{\alpha}_{i_k}\}$ 是向量组 $S_2 = \{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_m\}$ 的一个子集. 若 S_1 线性相关, 则 S_2 也线性相关; 等价地, 若 S_2 线性无关, 则 S_1 也线性无关. 教材定理 5.2.3

证明. 设 S_1 线性相关, 则存在不全为零的标量 $\lambda_{i_1},\lambda_{i_2},\dots,\lambda_{i_k}$ 使得 $\sum_{j=1}^k \lambda_{i_j}\boldsymbol{\alpha}_{i_j} = \mathbf{0}$. 此时, 对于 $i \in \{1,2,\dots,m\}\setminus\{i_1,i_2,\dots,i_k\}$, 补充定义 $\lambda_i = 0$. 于是 $\sum_{i=1}^m \lambda_i\boldsymbol{\alpha}_i = \sum_{j=1}^k \lambda_{i_j}\boldsymbol{\alpha}_{i_j} = \mathbf{0}$. 而 $\lambda_1,\lambda_2,\dots,\lambda_m$ 不全为零, 从而 S_2 线性相关. □

定义 5.2.6. 若向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 中的每一个向量都可以用向量组 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 线性表示, 则称 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ **线性表示**.

定理 5.2.7. (1) 向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由向量组 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 线性表示的充要条件是

$$\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle \quad\subseteq\quad \langle\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell\rangle. \tag{5.1}$$

(2) 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由 $\boldsymbol{b}_1,\dots,\boldsymbol{b}_\ell$ 线性表示, $\boldsymbol{b}_1,\dots,\boldsymbol{b}_\ell$ 可以由 $\gamma_1,\dots,\gamma_t$ 线性表示, 则 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由 $\gamma_1,\dots,\gamma_t$ 线性表示, 即, 向量组的线性表示具有传递性.

证明. (1) 假定包含关系 (5.1) 成立, 由于 $\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\}\subseteq\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$, 而 $\langle\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell\rangle$ 是可由 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 线性表示的向量的全体, 故 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 线性表示.

反之, 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 线性表示, 从而存在 $\lambda'_{ij}\in F$ 使得 $\boldsymbol{\alpha}_i=\sum_{j=1}^{\ell}\lambda'_{ij}\boldsymbol{\beta}_j$. 若 $\boldsymbol{x}\in\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$, 则 $\boldsymbol{x}=\sum_{i=1}^{m}\lambda''_i,\boldsymbol{\alpha}_i$ 可以由 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 线性表示, 其中 $\lambda''_i\in F$. 此时, $\boldsymbol{x}=\sum_{i=1}^{m}\lambda''_i\sum_{j=1}^{\ell}\lambda'_{ij}\boldsymbol{\beta}_j=\sum_{j=1}^{\ell}(\sum_{i=1}^{m}\lambda''_i\lambda'_{ij})\boldsymbol{\beta}_j$, 可由 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 线性表示, 等价地, $\boldsymbol{x}\in\langle\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell\rangle$. 由 $\boldsymbol{x}\in\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$ 的任意性知包含关系 (5.1) 成立.

(2) 假设设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 线性表示, 从而等价地有包含关系 (5.1) 成立. 再假设 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 可以由 $\boldsymbol{\gamma}_1,\dots,\boldsymbol{\gamma}_t$ 线性表示, 从而等价地有包含关系

$$\langle\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell\rangle\subseteq\langle\boldsymbol{\gamma}_1,\dots,\boldsymbol{\gamma}_t\rangle,$$

于是有

$$\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle\subseteq\langle\boldsymbol{\gamma}_1,\dots,\boldsymbol{\gamma}_t\rangle,$$

从而等价地可知, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 可以由 $\boldsymbol{\gamma}_1,\dots,\boldsymbol{\gamma}_t$ 线性表示. □

定理 5.2.8. 设 $m\geq 2$. 则向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\in F^n$ 线性相关的充要条件是存在 $i_0\in\{1,2,\dots,m\}$, 使得 教材定理 5.2.1

$$\langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_m\rangle=\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_{i_0-1},\boldsymbol{\alpha}_{i_0+1},\dots,\boldsymbol{\alpha}_m\rangle,\tag{5.2}$$

即向量 $\boldsymbol{\alpha}_{i_0}$ 对于生成子空间 $\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$ 是多余的.

证明. 设向量组 $S_1\coloneqq\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\}$ 线性相关, 因此存在某个向量 $\boldsymbol{\alpha}_{i_0}$, 它可以由剩下的向量组 $S_2\coloneqq\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_{i_0-1},\boldsymbol{\alpha}_{i_0+1},\dots,\boldsymbol{\alpha}_m\}$ 线性表示. 于是由定义, S_1 可以被 S_2 线性表示, 即

$$\langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_m\rangle\subseteq\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_{i_0-1},\boldsymbol{\alpha}_{i_0+1},\dots,\boldsymbol{\alpha}_m\rangle.$$

但是, S_2 作为子集很明显可以被 S_1 线性表示, 即

$$\langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_m\rangle\supseteq\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_{i_0-1},\boldsymbol{\alpha}_{i_0+1},\dots,\boldsymbol{\alpha}_m\rangle.$$

综上可知, 等式 (5.2) 成立.

反过来, 设存在 $i_0\in\{1,2,\dots,m\}$, 使得等式 (5.2) 成立, 特别地, 包含关系 "$\subseteq$" 成立. 利用上面定义的 S_1,S_2, 我们看到 S_1 可以被 S_2 线性表示, 特别地, 向量 $\boldsymbol{\alpha}_{i_0}$ 可以被 S_2 线性表示. 这说明 S_1 线性相关. □

定理 5.2.9. 设 $\boldsymbol{A}$ 为 $n\times m$ 矩阵, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 为 $\boldsymbol{A}$ 的列向量. 则以下三条等价: 教材定理 5.2.4

(1) 向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 线性相关;

(2) (关于 λ 的) 齐次线性方程组 $\boldsymbol{A}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_m\end{pmatrix}=\boldsymbol{0}$ 有非零解;

(3) 矩阵 $\boldsymbol{A}$ 不是列满秩的, 即, $\operatorname{rank}(\boldsymbol{A})<m$.

证明. 由定义可以看出, (1)$\Leftrightarrow$(2) 是显然的.

记 $\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_m)^{\mathrm{T}}$. 若 (2) 成立, 则 $\boldsymbol{\lambda}\neq\boldsymbol{0}$, 从而等价地, $\operatorname{rank}(\boldsymbol{\lambda})=1$. 由于 $\boldsymbol{A\lambda}=\boldsymbol{0}$, 由例 4.5.17 中的不等式可知,

$$\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{\lambda})-m\leq\operatorname{rank}(\boldsymbol{0}),$$

即 $\operatorname{rank}(\boldsymbol{A})\leq m-1$, 即 $\operatorname{rank}(\boldsymbol{A})<m$. 从而 (3) 成立.

反之, 若 (3) 成立, 我们记 $r=\operatorname{rank}(\boldsymbol{A})<m$. 此时, 存在可逆矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 使得 $\boldsymbol{PAQ}=\left(\begin{smallmatrix}\boldsymbol{I}_r&\boldsymbol{O}\\ \boldsymbol{O}&\boldsymbol{O}\end{smallmatrix}\right)$ 为 $\boldsymbol{A}$ 的相抵标准形. 从而,

$$\boldsymbol{AQ}=\left(\boldsymbol{P}^{-1}\begin{pmatrix}\boldsymbol{I}_r\\ \boldsymbol{O}\end{pmatrix},\boldsymbol{O}_{n\times(m-r)}\right).$$

若我们设 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\dots,\boldsymbol{\beta}_m$ 是 $\boldsymbol{Q}$ 的列向量, 由于 $m>r$, 比较该等式的最后一列后我们得到 $\boldsymbol{A\beta}_m=\boldsymbol{0}$. 另一方面, 由于 $\boldsymbol{Q}$ 可逆, $\boldsymbol{\beta}_m\neq\boldsymbol{0}$. 这说明 $\boldsymbol{\lambda}=\boldsymbol{\beta}_m$ 是方程组 $\boldsymbol{A\lambda}=\boldsymbol{0}$ 的一个非零解, 从而 (2) 成立. □

注意到一组数组向量在将其写成列向量时是线性相关的当且仅当将其写成行向量时是线性相关的.

推论 5.2.10. 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\in F^n$. 设 $\boldsymbol{A}$ 是以 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 为行向量的矩阵. 教材推论 5.2.1

(1) 若 $m>n$, 则 $\boldsymbol{A}$ 满足 $\operatorname{rank}(\boldsymbol{A})\leq\min(m,n)=n<m$, 故 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 线性相关.

(2) 若 $m=n$, 则 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 线性相关 $\Leftrightarrow$ m 阶方阵 $\boldsymbol{A}$ 满足 $\operatorname{rank}(\boldsymbol{A})<m$ $\Leftrightarrow$ 方阵 $\boldsymbol{A}$ 的行列式 $\det(\boldsymbol{A})=0$.

例 5.2.11. 判断向量组 $\boldsymbol{\alpha}_1=(3,4,-2,5)$, $\boldsymbol{\alpha}_2=(2,-5,0,-3)$, $\boldsymbol{\alpha}_3=(5,0,-1,2)$, $\boldsymbol{\alpha}_4=(3,3,-3,5)$ 是否线性无关. 教材例题 5.2.4(4)

解. 将这些向量视作行向量, 由于这些向量的长度都是 4, 而向量组的长度也是 4, 我们将其视作一个 4 阶方阵的行向量, 并计算其行列式:

$$\begin{vmatrix}\boldsymbol{\alpha}_1\\\boldsymbol{\alpha}_2\\\boldsymbol{\alpha}_3\\\boldsymbol{\alpha}_4\end{vmatrix}=\begin{vmatrix}3&4&-2&5\\2&-5&0&-3\\5&0&-1&2\\3&3&-3&5\end{vmatrix}\xlongequal[-r_2\to r_3]{-r_1\to r_4}\begin{vmatrix}3&4&-2&5\\2&-5&0&-3\\3&5&-1&5\\0&-1&-1&0\end{vmatrix}\xlongequal{-r_1\to r_3}\begin{vmatrix}3&4&-2&5\\2&-5&0&-3\\0&1&1&0\\0&-1&-1&0\end{vmatrix}=0.$$

由推论 5.2.10 可知, 这说明向量组 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4$ 线性相关. □

例 5.2.12. *已知列向量组 $\boldsymbol{\alpha}_1=(1,2,-1,4)^{\mathrm{T}}$, $\boldsymbol{\alpha}_2=(0,-1,a,3)^{\mathrm{T}}$, $\boldsymbol{\alpha}_3=(2,5,3,5)^{\mathrm{T}}$ 线性无关, 求参数 a.*

证明. 设 $\boldsymbol{A}$ 是以这些向量为列向量所构成的矩阵. 那么由推论 5.2.10 的等价变形可知, $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 线性无关的充要条件是 $\boldsymbol{A}$ 的秩为 3, 即 $\operatorname{rank}(\boldsymbol{A})=3$. 利用初等行变换, 我们有

$$\boldsymbol{A}^{\mathrm{T}}=\begin{pmatrix}1&0&2\\2&-1&5\\-1&a&3\\4&3&5\end{pmatrix}\xrightarrow[\substack{r_1\to r_3\\-4r_1\to r_4}]{-2r_1\to r_2}\begin{pmatrix}1&0&2\\0&-1&1\\0&a&5\\0&3&-3\end{pmatrix}\xrightarrow[3r_2\to r_4]{ar_2\to r_3}\begin{pmatrix}1&0&2\\0&-1&1\\0&0&5+a\\0&0&0\end{pmatrix}.$$

由于 $\boldsymbol{A}$ 的秩为 3, 故上述的阶梯标准形里必恰有三个非零行出现, 即 $a\neq-5$, □

学生自习教材 P124 定理 5.2.5.

习题 5.2.13. *设 $\boldsymbol{A}$ 是 n 阶可逆矩阵. 证明: 列向量 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\in F^n$ 是线性无关的当且仅当 $\boldsymbol{A}\boldsymbol{\alpha}_1,\ldots,\boldsymbol{A}\boldsymbol{\alpha}_m$ 是线性无关的.*

5.3 极大无关组与秩

给定一组向量 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\in F^n$. 若这组向量线性相关, 则用它们来生成子空间 $\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\rangle$ 时, 就存在冗余的向量. 这节的一个目标是: 通过反复移去多余的向量, 使得剩下的向量组不再存在冗余, 从而线性无关.

定义 5.3.1. *设 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\in F^n$ 为一组向量. 若有*

(1) *存在向量组 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性无关, 并且*

(2) *任意添加一个其它的向量 $\boldsymbol{\alpha}_{i_{r+1}}$ 后, 向量组 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r},\boldsymbol{\alpha}_{i_{r+1}}$ 线性相关,*

则称 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 构成了向量组 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 的**一个极大无关组**. 简言之, 它是原向量组 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 的所有线性无关的子向量组依照集合的包含关系所能取到的极大元.

由上面定义之前的讨论不难看出, 当 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m$ 中包含非零向量时, 通过反复移去多余的向量, 一定可以找到它的一个极大无关组.

注 5.3.2. 在上面的定义中, 由于 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r},\boldsymbol{\alpha}_{i_{r+1}}$ 线性相关, 存在不全为零的标量 $\lambda_j\in F$, 使得 $\sum_{j=1}^{r+1}\lambda_j\boldsymbol{\alpha}_{i_j}=\mathbf{0}$. 但是又由于 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性无关, 故 $\lambda_{r+1}\neq 0$, 从而 $\boldsymbol{\alpha}_{i_{r+1}}$ 可以由 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性表示 (进一步地, 不难看成, 这样的线性表示也是唯一的). 由 $\boldsymbol{\alpha}_{i_{r+1}}$ 的任意性可知, $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 中的所有向量都可以由 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性表示, 由此可以推出 $\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\rangle\subseteq\langle\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}\rangle$. 当然, 在这儿包含关系 "$\supseteq$" 是自动成立的, 从而两个线性子空间相等.

学生自学教材 P125 的例题 5.3.1.

定理 5.3.3. 设 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\in F^n$ 为一组列向量, $\boldsymbol{A}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m)$ 是相应的 $n\times m$ 矩阵. 假定 $\boldsymbol{A}$ 经过一系列的初等行变换后得到矩阵 $\boldsymbol{B}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_m)$, 则 教材定理 5.3.1

(1) $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性相关 (线性无关) $\Leftrightarrow$ $\boldsymbol{\beta}_{i_1},\boldsymbol{\beta}_{i_2},\ldots,\boldsymbol{\beta}_{i_r}$ 线性相关 (线性无关);

(2) $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 为向量组 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 的极大无关组 $\Leftrightarrow$ $\boldsymbol{\beta}_{i_1},\boldsymbol{\beta}_{i_2},\ldots,\boldsymbol{\beta}_{i_r}$ 为向量组 $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m$ 的极大无关组.

证明. (2) 可以由 (1) 直接推出. 而为了证明 (1), 我们不妨假设 $r=m$, $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 依次为 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m$. 则由定理 5.2.9 可知, $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 线性相关的充要条件是 $\operatorname{rank}(\boldsymbol{A})<m$. 由于 $\boldsymbol{B}$ 可以由 $\boldsymbol{A}$ 经过一系列的初等行变换得到, 从而 $\operatorname{rank}(\boldsymbol{B})=\operatorname{rank}(\boldsymbol{A})$. 另一方面, $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_m$ 线性相关的充要条件是 $\operatorname{rank}(\boldsymbol{B})<m$. 综上可以看出, (1) 成立. □

例 5.3.4. 对于向量组 $\boldsymbol{\alpha}_1=(-1,5,3,-2)^{\mathrm{T}}$, $\boldsymbol{\alpha}_2=(4,1,-2,9)^{\mathrm{T}}$, $\boldsymbol{\alpha}_3=(2,0,-1,4)^{\mathrm{T}}$, $\boldsymbol{\alpha}_4=(1,15,7,2)^{\mathrm{T}}$ (仅有第四个向量与教材的例 *5.3.2* 的相应向量不同), 我们有

$$(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4)=\begin{pmatrix}-1&4&2&1\\5&1&0&15\\3&-2&-1&7\\-2&9&4&2\end{pmatrix}\xrightarrow[\text{约化标准形}]{\text{(计算留作热身题)}}\begin{pmatrix}1&0&0&3\\0&1&0&0\\0&0&1&2\\0&0&0&0\end{pmatrix}\triangleq(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4).$$

通过检验相应子矩阵的秩, 我们不难看出, $\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3\}$, $\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_4\}$,$\{\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4\}$ 是向量组 $\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4\}$ 的所有极大无关组, 而 $\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4\}$ 线性相关. 从而对应地, $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\}$,

$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_4\}$,$\{\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4\}$ 是向量组 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4\}$ 的所有极大无关组, 而 $\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4\}$ 线性相关.

注 5.3.5. (1) 用约化标准形可以相对比较容易地找出所有的极大无关组. 不过一般情形下, 也没有必要找出所有的极大无关组.

(2) 设矩阵 $\boldsymbol{A}\in F^{n\times m}$ 用初等行变换可以得到阶梯形矩阵 $\boldsymbol{B}=(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m)$. 设主元依次在 $i_1<i_2<\cdots<i_r$ 列上, 则我们断言 $\boldsymbol{\beta}_{i_1},\boldsymbol{\beta}_{i_2},\ldots,\boldsymbol{\beta}_{i_r}$ 为 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_m$ 的一个极大无关组, 从而 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r}$ 为 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_m$ 的一个极大无关组.

为了证明该断言, 我们注意到子矩阵 $\boldsymbol{B}':=(\boldsymbol{\beta}_{i_1},\boldsymbol{\beta}_{i_2},\ldots,\boldsymbol{\beta}_{i_r})$ 虽然不是方阵, 但是从 $(1,1)$ 出发的主对角线上的元素全部非零, 而其之下的元素皆为 0, 这说明 $\operatorname{rank}(\boldsymbol{B}')=r$, 从而 $\boldsymbol{\beta}_{i_1},\boldsymbol{\beta}_{i_2},\ldots,\boldsymbol{\beta}_{i_r}$ 线性无关. 另一方面, $\boldsymbol{B}$ 的任意一个 $r+1$ 列的子矩阵的秩不超过 $\boldsymbol{B}$ 的秩, 即至多为 r. 这说明 $\boldsymbol{B}$ 的列向量的长度为 $r+1$ 的子组必定是线性相关的. 故由定义可知, $\boldsymbol{\beta}_{i_1},\boldsymbol{\beta}_{i_2},\ldots,\boldsymbol{\beta}_{i_r}$ 为 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\ldots,\boldsymbol{\beta}_m$ 的一个极大无关组.

例 5.3.6. 求向量组 $\boldsymbol{\alpha}_1=(1,2,2,0)^{\mathrm{T}}$, $\boldsymbol{\alpha}_2=(2,4,4,0)^{\mathrm{T}}$, $\boldsymbol{\alpha}_3=(1,0,3,1)^{\mathrm{T}}$ 和 $\boldsymbol{\alpha}_4=(0,4,-2,-2)^{\mathrm{T}}$ 的一个极大无关组, 并求出其余向量该相对于该极大无关组对的线性表示式.

解. 将这组列向量按行排列组成矩阵 $\boldsymbol{A}$, 对其进行初等行变换操作, 以得到约化标准形

$$\boldsymbol{A}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4)=\begin{pmatrix}1&2&1&0\\2&4&0&4\\2&4&3&-2\\0&0&1&-2\end{pmatrix}\xrightarrow{\text{(计算留作热身题)}}\begin{pmatrix}1&2&0&2\\0&0&1&-2\\0&0&0&0\\0&0&0&0\end{pmatrix}=\boldsymbol{B}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4).$$

由此看出, $\boldsymbol{\beta}_1,\boldsymbol{\beta}_3$ 是 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4$ 的一个极大无关组, 而 $\boldsymbol{\beta}_2=2\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_4=2\boldsymbol{\beta}_1-2\boldsymbol{\beta}_3$ (这是约化标准形所带来的便利). 由于 $\boldsymbol{B}$ 是由 $\boldsymbol{A}$ 通过一系列初等行变换得到的, 故存在可逆矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{B}=\boldsymbol{P}\boldsymbol{A}$, 即 $\boldsymbol{\beta}_i=\boldsymbol{P}\boldsymbol{\alpha}_i$ $(1\le i\le 4)$. 从而, $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_3$ 是 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\alpha}_4$ 的一个极大无关组, 而 $\boldsymbol{\alpha}_2=2\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_4=2\boldsymbol{\alpha}_1-2\boldsymbol{\alpha}_3$. □

现在有一个很自然的问题: 一个向量组的不同的极大无关组的长度是否相等?

定义 5.3.7. 若向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 与 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell$ 可以相互线性表示, 则称它们**等价**, 并记作

$$\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\}\quad\sim\quad\{\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell\}.$$

由定理 *5.2.7* 不难看出, 这等价于

$$\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle=\langle\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_\ell\rangle. \qquad (\text{教材定理 } 5.3.2)$$

进一步地, 不难看出, “$\sim$” 满足等价关系的三条公理 (反身性, 对称性, 传递性).

定理 5.3.8. 一个向量组与它的任何一个极大无关组等价. 教材定理 5.3.3

证明. 利用前面的注 5.3.2. □

推论 5.3.9. 向量组的任何两个极大无关组彼此等价. 教材推论 5.3.1

证明. 利用向量组等价关系的传递性. □

定理 5.3.10. 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\in F^n$. 则 $\boldsymbol{\alpha}_{i_1},\dots,\boldsymbol{\alpha}_{i_r}$ 是 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 的一个极大无关组当且仅当 $\boldsymbol{\alpha}_{i_1},\dots,\boldsymbol{\alpha}_{i_r}$ 线性无关且 教材定理 5.3.4

$$\langle\boldsymbol{\alpha}_{i_1},\dots,\boldsymbol{\alpha}_{i_r}\rangle=\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle. \tag{5.3}$$

证明. 假设 $\boldsymbol{\alpha}_{i_1},\ldots,\boldsymbol{\alpha}_{i_r}$ 是 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 的一个极大无关组, 从而由定义出发, 我们知道 $\boldsymbol{\alpha}_{i_1},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性无关, 并且由定理 5.3.8 知等式 (5.3) 成立.

反之, 设 $\boldsymbol{\alpha}_{i_1},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性无关, 且 $\langle\boldsymbol{\alpha}_{i_1},\ldots,\boldsymbol{\alpha}_{i_r}\rangle=\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\rangle$ 成立. 为了证明 $\boldsymbol{\alpha}_{i_1},\ldots,\boldsymbol{\alpha}_{i_r}$ 是 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 的一个极大无关组, 由定义出发, 只需证明, 对于任意的向量 $\boldsymbol{\alpha}_{i_{r+1}}$, 向量组 $\boldsymbol{\alpha}_{i_1},\boldsymbol{\alpha}_{i_2},\ldots,\boldsymbol{\alpha}_{i_r},\boldsymbol{\alpha}_{i_{r+1}}$ 线性相关. 为此, 只需证明 $\boldsymbol{\alpha}_{i_{r+1}}$ 可由向量组 $\boldsymbol{\alpha}_{i_1},\ldots,\boldsymbol{\alpha}_{i_r}$ 线性表示. 由于 $\boldsymbol{\alpha}_{i_{r+1}}\in\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\rangle=\langle\boldsymbol{\alpha}_{i_1},\ldots,\boldsymbol{\alpha}_{i_r}\rangle$, 这是显然的. □

定理 5.3.11. 若两个线性无关的向量组 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\}$ 和 $\{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_s\}$ 等价, 则 $r=s$. 教材定理 5.3.5

证明. 由于 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\}$ 可以由 $\{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_s\}$ 线性表示, 存在系数 $\mu_{ij}\in F$ 使得 $\boldsymbol{\alpha}_i=\sum_{j=1}^{s}\mu_{ij}\boldsymbol{\beta}_j$. 将这些向量视为列向量, 并按行依次排列组成矩阵, 则有

$$\underbrace{\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots\cdots & \boldsymbol{\alpha}_r\end{pmatrix}}_{\boldsymbol{A}}=\underbrace{\begin{pmatrix}\boldsymbol{\beta}_1 & \cdots\cdots & \boldsymbol{\beta}_s\end{pmatrix}}_{\boldsymbol{B}}\underbrace{\begin{pmatrix}\mu_{11} & \cdots\cdots & \mu_{r1}\\ \vdots & & \vdots\\ \mu_{1s} & \cdots\cdots & \mu_{rs}\end{pmatrix}}_{\boldsymbol{C}}.$$

将其记为 $\boldsymbol{A}=\boldsymbol{B}\boldsymbol{C}$, 并注意 $\boldsymbol{C}$ 中元素的下标.

由于 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\}$ 线性无关, 由定理 5.2.9 可知, $\boldsymbol{A}$ 是列满秩的: $\operatorname{rank}(\boldsymbol{A})=r$. 类似地, $\operatorname{rank}(\boldsymbol{B})=s$. 此时,

$$r=\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\boldsymbol{B}\boldsymbol{C})\le\operatorname{rank}(\boldsymbol{B})=s.$$

其中的不等式用到了例 4.5.14 的结果.

由对称性, 我们也可以得到 $s\le r$. 故 $r=s$. □

推论 5.3.12. 一个向量组的任何两个极大无关组的长度相等. 教材推论 5.3.2

定义 5.3.13. 一个向量组 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m\}$ 的极大无关组的长度称为该向量组的**秩**, 记为 $r(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m)$ 或 $\operatorname{rank}(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m)$. 我们约定: 只含零向量的向量组以及空的向量组 (即此处 $m=0$ 的情形, 可类比于空集) 的秩为 0.

注 5.3.14. 在上面的定义中, 若将这些向量视为列向量, 并按行依次排列得到矩阵

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots\cdots & \boldsymbol{\alpha}_m\end{pmatrix},$$

则向量组的秩 $\operatorname{rank}(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m)$ 等于矩阵的秩 $\operatorname{rank}(\boldsymbol{A})$. 这是因为注 *5.3.5* 中我们已经证明了向量组 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m$ 有一个极大无关组, 其长度就是 $\boldsymbol{A}$ 的阶梯标准形的非零行数, 即 $\operatorname{rank}(\boldsymbol{A})$.

定义 5.3.15. 给定 $m \times n$ 矩阵 $\boldsymbol{A}$, 则

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_n \end{pmatrix}$$

将 $\boldsymbol{A}$ 写成了行向量按列排列以及列向量按行排列的形式. 行向量 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ 组成的 F^n 中的向量组的秩称为矩阵 $\boldsymbol{A}$ 的**行秩**, 列向量 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ 组成的 F^m 中的向量组的秩称为矩阵 $\boldsymbol{A}$ 的**列秩**. 与之相关地, 行向量 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ 生成的 F^n 的子空间称为矩阵 $\boldsymbol{A}$ 的**行空间** (记作 $\mathrm{Row}(\boldsymbol{A})$), 列向量 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ 生成的 F^m 的子空间称为矩阵 $\boldsymbol{A}$ 的**列空间** (记作 $\mathrm{Col}(\boldsymbol{A})$).

定理 5.3.16. 任何矩阵的秩等于它的行秩和列秩. 教材定理 5.3.7

证明. 在注 5.3.14 中, 我们证明了 $\boldsymbol{A}$ 的列秩 $\mathrm{rank}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n) = \mathrm{rank}(\boldsymbol{A})$.

对称地, $\boldsymbol{A}$ 的行秩 $\mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}) = \mathrm{rank}(\boldsymbol{A})$. □

再结合定理 5.2.9, 我们容易得到如下的结果.

推论 5.3.17. 对于 n 阶方阵 $\boldsymbol{A}$, 以下几条等价: 教材推论 5.3.3

(1) $\boldsymbol{A}$ 可逆;

(2) $\mathrm{rank}(\boldsymbol{A}) = n$;

(3) $\boldsymbol{A}$ 的行向量线性无关;

(4) $\boldsymbol{A}$ 的列向量线性无关.

定理 5.3.18. 设向量 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s \in F^n$, 则有: 教材定理 5.3.6

(1) $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 线性无关当且仅当 $\mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r) = r$;

(2) $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 线性相关当且仅当 $\mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r) < r$;

(3) 若 $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s\}$ 可以由 $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r\}$ 线性表示, 则 $\mathrm{rank}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s) \le \mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r)$;

(4) 若 $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s\}$ 与 $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r\}$ 等价, 则 $\mathrm{rank}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s) = \mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r)$;

(5) 向量 $\boldsymbol{b} \in F^n$ 可以表示成 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 的线性组合当且仅当 $\mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r) = \mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r, \boldsymbol{b})$.

证明. 将这些向量视为列向量, 并且记矩阵 $\boldsymbol{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r)$, $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s)$.

利用定理 5.2.9 和定理 5.3.16, 我们知道 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 线性相关的充要条件是 $\mathrm{rank}(\boldsymbol{A}) < r$, 而这又等价于向量组的秩 $\mathrm{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r) < r$. 从而 (1) 与 (2) 成立.

关于 (3), 若 $\{\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_s\}$ 可以由 $\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r\}$ 线性表示, 那么由定义, 这等价于说存在矩阵 $\boldsymbol{C}\in F^{r\times s}$ 使得 $\boldsymbol{B}=\boldsymbol{AC}$. 于是 (3) 等价于证明 $\operatorname{rank}(\boldsymbol{AC})\le\operatorname{rank}(\boldsymbol{A})$. 由例 4.5.14 知该不等式成立.

关于 (4), 由定义可知, $\{\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_s\}$ 与 $\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r\}$ 等价意味着这两组向量可以相互线性表示, 从而 (4) 可由 (3) 推出.

关于 (5), 先设向量 $\boldsymbol{b}\in F^n$ 可以表示成 $S_1=\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r\}$ 的线性组合, 而这等价于向量组 $S_2=\{\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r,\boldsymbol{b}\}$ 可由向量组 S_1 线性表示. 但是 S_1 总是可以由 S_2 线性表示, 从而这等价于说 S_1 与 S_2 可以相互线性表示, 即 S_1 与 S_2 等价. 由 (4) 推出 $\operatorname{rank}(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r)=\operatorname{rank}(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r,\boldsymbol{b})$.

反之, 设 $\operatorname{rank}(S_1)=\operatorname{rank}(S_2)=t$. 不失一般性, 假定 $S_3=\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_t\}$ 是 S_1 的一个极大无关组. 由于 S_2 的子向量组 $S_4=\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_t,\boldsymbol{b}\}$ 的长度为 $t+1$, 大于 S_2 的秩, 因而 S_4 线性相关. 这说明了 $\boldsymbol{b}$ 可以由 S_3 线性表示 (教材习题 #13), 从而等价地, $\boldsymbol{b}$ 可以由 S_1 线性表示. □

推论 5.3.19. *若矩阵 $\boldsymbol{A}$ 满足 $\operatorname{rank}(\boldsymbol{A})=r$, 则 $\boldsymbol{A}$ 的不等于零的 r 阶子式所在的行 (列) 构成 $\boldsymbol{A}$ 的行 (列) 向量组的极大无关组.* 教材推论 5.3.4

证明. 不妨设 $\left|\boldsymbol{A}\binom{1,\cdots,r}{1,\cdots,r}\right|\ne 0$. 将 $\boldsymbol{A}$ 写成行向量按列排列的形式: $\boldsymbol{A}=\begin{pmatrix}\boldsymbol{\alpha}_1\\ \vdots\\ \boldsymbol{\alpha}_m\end{pmatrix}$, 并记 $\boldsymbol{B}=\begin{pmatrix}\boldsymbol{\alpha}_1\\ \vdots\\ \boldsymbol{\alpha}_r\end{pmatrix}$. 则, $\left|\boldsymbol{B}\binom{1,\cdots,r}{1,\cdots,r}\right|\ne 0$. 从而 $\operatorname{rank}(\boldsymbol{B})=r$, 从而 $\boldsymbol{B}$ 是行满秩的, 从而 $\operatorname{rank}(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r)=r$, 从而由定理 5.3.18(1) 知 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r$ 线性无关.

另一方面, 对于任意给定的 t, 满足 $r+1\le t\le m$, 令 $\boldsymbol{C}=\begin{pmatrix}\boldsymbol{\alpha}_1\\ \vdots\\ \boldsymbol{\alpha}_r\\ \boldsymbol{\alpha}_t\end{pmatrix}$. 则作为子阵, $\operatorname{rank}(\boldsymbol{C})\le\operatorname{rank}(\boldsymbol{A})=r$. 从而 $\operatorname{rank}(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r,\boldsymbol{\alpha}_t)\le r$, 从而由定理 5.3.18(2) 知 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r,\boldsymbol{\alpha}_t$ 线性相关. (这儿也可以利用 $\boldsymbol{C}$ 的 $r+1$ 阶子式是 $\boldsymbol{A}$ 的 $r+1$ 阶子式从而为零来证明, 或者引用教材作业 #22)

由定义, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r$ 是 $\boldsymbol{A}$ 的行向量的极大无关组. □

习题 5.3.20. (1) *设 $m\times n$ 矩阵 $\boldsymbol{A}$ 的秩为 r. 任取 $\boldsymbol{A}$ 的 r 个线性无关的行和 r 个线性无关的列. 证明: 这 r 行和这 r 列交叉处的 r 阶子阵可逆.*

(2) 若上题中的 $r < \operatorname{rank}(\boldsymbol{A})$, 则其结论不成立. 试给出反例.

(3) 证明: 矩阵 $\boldsymbol{A}$ 的非零子式所在的行向量组和列向量组都是线性无关的.

(4) 上题的逆命题不成立, 即位于线性无关的行向量组和线性无关的列向量组交叉处的子式不一定非零. 试给出反例.

学生课后比较教材例题 5.3.4(P131) 与例题 4.5.7(P112) 的证明.

补充例题

例 5.3.21. 已知向量组 $(S_1): \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$, $(S_2): \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4$, $(S_3): \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_5$, 且向量组的秩 $\operatorname{rank}(S_1) = \operatorname{rank}(S_2) = 3$, $\operatorname{rank}(S_3) = 4$. 证明: 向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_5 - \boldsymbol{\alpha}_4$ 的秩为 4.

证明. 由于 $\operatorname{rank}(S_1) = \operatorname{rank}(S_2) = 3$, 我们知道向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 是线性无关的, 而 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4$ 是线性相关的, 从而 $\boldsymbol{\alpha}_4 = \lambda_1\boldsymbol{\alpha}_1 + \lambda_2\boldsymbol{\alpha}_2 + \lambda_3\boldsymbol{\alpha}_3$ 是 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 的线性组合. 由此, 不难验证, 向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_5 - \boldsymbol{\alpha}_4$ 与向量组 (S_3) 是等价的, 从而该向量组的秩为 $\operatorname{rank}(S_3) = 4$. □

例 5.3.22. 在 F^n 中, 如果向量组 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$ 可以由向量组 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t$ 线性表示, 而且 $s > t$, 那么 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$ 线性相关.

证明. 将 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t$ 视为列向量, 考虑矩阵

$$\boldsymbol{A} = \Big(\boldsymbol{\alpha}_1 \cdots\cdots \boldsymbol{\alpha}_s\Big)_{n\times s} \qquad \text{和} \qquad \boldsymbol{B} = \Big(\boldsymbol{\beta}_1 \cdots\cdots \boldsymbol{\beta}_t\Big)_{n\times t}.$$

由于 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$ 可以由 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t$ 线性表示, 存在矩阵 $\boldsymbol{C} \in F^{t\times s}$ 使得 $\boldsymbol{A} = \boldsymbol{BC}$. 此时, $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{BC}) \le \operatorname{rank}(\boldsymbol{C}) \le t < s$. 这说明矩阵 $\boldsymbol{A}$ 不是列满秩的, 从而 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$ 线性相关. □

例 5.3.23. 设 $\boldsymbol{A}$ 为 n 阶方阵, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n \in F^n$ 为 n 个线性无关的向量. 证明: 矩阵的秩 $\operatorname{rank}(\boldsymbol{A})$ 等于向量组的秩 $\operatorname{rank}(\boldsymbol{A}\boldsymbol{\alpha}_1, \ldots, \boldsymbol{A}\boldsymbol{\alpha}_n)$. 特别地, $\operatorname{rank}(\boldsymbol{A}) = n$ 当且仅当 $\boldsymbol{A}\boldsymbol{\alpha}_1, \boldsymbol{A}\boldsymbol{\alpha}_2, \ldots, \boldsymbol{A}\boldsymbol{\alpha}_n$ 线性无关.

证明. 由定理 5.3.16, 向量组的秩 $\operatorname{rank}(\boldsymbol{A}\boldsymbol{\alpha}_1, \ldots, \boldsymbol{A}\boldsymbol{\alpha}_n)$ 等于矩阵的秩

$$\operatorname{rank}((\boldsymbol{A}\boldsymbol{\alpha}_1, \ldots, \boldsymbol{A}\boldsymbol{\alpha}_n)) = \operatorname{rank}(\boldsymbol{A}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)) = \operatorname{rank}(\boldsymbol{A}).$$

最后的等式成立是因为 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 是线性无关的 n 为列向量, 从而矩阵 $(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$ 是可逆矩阵.

特别地, $\operatorname{rank}(\boldsymbol{A})=n$ 当且仅当 $\operatorname{rank}(\boldsymbol{A}\boldsymbol{\alpha}_1,\ldots,\boldsymbol{A}\boldsymbol{\alpha}_n)=n$, 当且仅当向量组 $\boldsymbol{A}\boldsymbol{\alpha}_1,\ldots,\boldsymbol{A}\boldsymbol{\alpha}_n$ 线性无关. □

例 5.3.24. 设 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 和 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_r$ 是两个向量组, 满足 $r\geq 2$, $\boldsymbol{\beta}_i=\sum_{j\neq i}\boldsymbol{\alpha}_j$. 试证明向量组的秩 $\operatorname{rank}(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r)=\operatorname{rank}(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_r)$.

证明. 我们只需证明 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 和 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_r$ 是等价的向量组.

(思路一) 容易看出 $(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_r)=(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r)\boldsymbol{T}$, 其中矩阵

$$\boldsymbol{T}=\begin{pmatrix}0 & 1 & \cdots & 1\\ 1 & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 1\\ 1 & \cdots & 1 & 0\end{pmatrix}.$$

下面来证明矩阵 $\boldsymbol{T}$ 的行列式非零, 即 $\boldsymbol{T}$ 为可逆矩阵. 于是 $(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r)=(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_r)\boldsymbol{T}^{-1}$. 从而 $\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_r$ 和 $\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_r$ 是等价的向量组.

(a) **(加边法)**

$$\begin{vmatrix}0 & 1 & \cdots & 1\\ 1 & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 1\\ 1 & \cdots & 1 & 0\end{vmatrix}_{r\times r} \xlongequal{\text{加边}} \begin{vmatrix}1 & 1 & 1 & \cdots & 1\\ 0 & 0 & 1 & \cdots & 1\\ \vdots & 1 & \ddots & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & 1\\ 0 & 1 & \cdots & 1 & 0\end{vmatrix}_{(r+1)\times(r+1)} \xlongequal{-r_1\to r_i,\forall i\geq 2} \begin{vmatrix}1 & 1 & \cdots & 1\\ -1 & -1 & & \\ \vdots & & \ddots & \\ -1 & & & -1\end{vmatrix}$$

$$\xlongequal{-c_i\to c_1,\forall i\geq 2} \begin{vmatrix}1-r & 1 & \cdots & 1\\ & -1 & & \\ & & \ddots & \\ & & & -1\end{vmatrix} = (-1)^r(1-r)\neq 0.$$

(b) **(利用 Schur 公式)**, 即教材作业 P115 #25.

$$\begin{aligned}\det(\boldsymbol{T}) &= \det(-\boldsymbol{I}_r+(1,1,\ldots,1)^{\mathrm{T}}(1,1,\ldots,1))\\ &= (-1)^r\det(\boldsymbol{I}_r-(1,1,\ldots,1)^{\mathrm{T}}(1,1,\ldots,1))\\ &= (-1)^r\det(\boldsymbol{I}_1-(1,1,\ldots,1)(1,1,\ldots,1)^{\mathrm{T}})\\ &= (-1)^r\det(1-r)=(-1)^r(1-r).\end{aligned}$$

(思路二) 我们只需证明 $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_r$ 可以由 $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_r$ 线性表示即可. 注意到 $\boldsymbol{\beta}_1 + \cdots + \boldsymbol{\beta}_r = (r-1)(\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_r)$. 因此, 对任意的 i, 我们有 $\boldsymbol{\alpha}_i = (\boldsymbol{\alpha}_1 + \cdots + \boldsymbol{\alpha}_r) - \boldsymbol{\beta}_i = \frac{1}{r-1}(\boldsymbol{\beta}_1 + \cdots + \boldsymbol{\beta}_r) - \boldsymbol{\beta}_i$, 是 $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_r$ 的一个线性组合. $\square$

习题 5.3.25. (1) 设 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是矩阵 $\boldsymbol{A} \in F^{m\times n}$ 的某个阶梯标准形的所有非零行的行向量, 证明这些向量生成了行空间 $\mathrm{Row}(\boldsymbol{A})$.

(2) 证明: 向量组

$$\boldsymbol{\alpha}_1 = (1, 2, -1, 3), \quad \boldsymbol{\alpha}_2 = (2, 4, 1, -2), \quad \boldsymbol{\alpha}_3 = (3, 6, 3, -7)$$

与

$$\boldsymbol{\beta}_1 = (1, 2, -4, 11), \quad \boldsymbol{\beta}_2 = (2, 4, -5, 14)$$

等价.

5.4 基与维数

回忆. (i) F^n 的非空子集 V 是一个**子空间**当且仅当对任意正整数 m, 对任意数组向量 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\in V$ 和任意标量 $\lambda_1,\dots,\lambda_m\in F$, 都有线性组合 $\sum_{i=1}^m\lambda_i\boldsymbol{\alpha}_i\in V$.

(ii) F^n 的子空间 V 是由向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\in F^n$ **生成**的当且仅当 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\in V$ 并且任意的数组向量 $\boldsymbol{a}\in V$ 都是 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 的线性组合, 即, 存在标量 $\lambda_1,\dots,\lambda_m\in F$ 使得 $\boldsymbol{a}=\sum_{i=1}^m\lambda_i\boldsymbol{\alpha}_i$. 此时记作 $V=\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$.

引理 5.4.1. 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i$ 是 F^n 中线性无关的一组向量, 而 $\boldsymbol{\alpha}_{i+1}\in F^n\setminus\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i\rangle$, 则 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+1}$ 也是一组线性无关的向量.

证明. 用反证法和教材课后习题 #13. □

下面的定理说明 F^n 的任何子空间都是某一组向量生成的.

定理 5.4.2. 设非空集合 V 是 F^n 的子空间, 则存在线性无关的向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r$ 使得

$$V=\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r\rangle\,.$$

证明. 若 V 是零空间, 上面的 $r=0$.

若 V 不是零空间, 先任意选取非零向量 $\boldsymbol{\alpha}_1\in V$. 考虑如下递归的选取方法: 假设已经选出了线性无关的向量 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i\in V$, 其中 $i\geq 1$. 若 $\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i\rangle\subsetneq V$, 则任意选取 $\boldsymbol{\alpha}_{i+1}\in V\setminus\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i\rangle$. 由上面的引理 5.4.1 可知, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_{i+1}$ 是线性无关的向量组.

另一方面, 由推论 5.2.10(1), F^n 中线性无关的向量组的长度最多为 n. 这说明上述的递归选取的方法无法无限地继续下去. 从而存在非负整数 $r\leq n$, 使得 $V=\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r\rangle$. □

注 5.4.3. (i) 在上面的定理中, V 的任意向量 $\boldsymbol{x}$ 都可以由 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r$ 唯一地线性表示出来: $\boldsymbol{x}=\sum_{i=1}^r\lambda_i\boldsymbol{\alpha}_i$, 其中 $\lambda_i\in F$. 表示的存在性是由于 V 是由 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r$ 生成的. 表示的唯一性在于: 若另外存在 $\lambda_i'\in F$ 使得 $\boldsymbol{x}=\sum_{i=1}^r\lambda_i'\boldsymbol{\alpha}_i$, 则 $\sum_i(\lambda_i-\lambda_i')\boldsymbol{\alpha}_i=\boldsymbol{a}-\boldsymbol{a}=\boldsymbol{0}$. 由于 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_i$ 线性无关, 对于任意 i, 系数 $\lambda_i-\lambda_i'=0$, 即 $\lambda_i=\lambda_i'$.

(ii) 定理中的线性无关的有序向量组 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_r$ 的选取一般并不唯一.

定义 5.4.4. 对于 F^n 的子空间 V，若存在线性无关的向量组 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r \in V$ 满足 $V=\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\rangle$，则称该向量组构成 V 的一组**基**. 由上面的注可知，V 中的任意向量 $\boldsymbol{x}$ 此时都可以唯一地由 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r$ 线性表示出来：

$$\boldsymbol{x}=\sum_{i=1}^{r}\lambda_i\boldsymbol{\alpha}_i.$$

我们称数组 $(\lambda_1,\ldots,\lambda_r)\in F^r$ 是 $\boldsymbol{x}$ 在基 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\}$ 下的**坐标向量** (或简称为**坐标**).

注 5.4.5. 在上述定义中，

(1) 可以证明：向量组 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\}$ 构成向量空间 V 的一组基的充要条件是 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\in V$，且 V 中的所有向量都可以由这组向量唯一地线性表示出来. 留作练习

(2) 进一步地，由于 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r$ 又生成了 V，基 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\}$ 是 V (无限多个向量构成的向量组) 的一个极大无关组 (这意味着 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r$ 是线性无关的，而任取 $\boldsymbol{\alpha}_{r+1}\in V$ 后，$\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r,\boldsymbol{\alpha}_{r+1}$ 是线性相关的).

(3) 若 $\{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_t\}$ 是 V 的另外一组基，则 $t\le n$，且 $\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\rangle=\langle\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_t\rangle$，从而向量组 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_t$ 等价. 而它们都是线性无关的向量组，从而由定理 *5.3.12* 知 $r=t$. 这说明该线性子空间的基的长度不依赖于基的具体选取. 这个公共的长度 r 称为子空间 V 的**维数**，记作 $\dim(V)=r$. 不难看出，$\dim(V)\le n$. (按照前面推广后的定义，这是无限长度的向量组 V 的极大无关组的长度).

(4) 更一般地，可以证明：F^n 中一组向量 $\{\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\}$ 的秩 (其极大无关组的公共长度) 等于它们生成的子空间 $\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_r\rangle$ 的维数，而该向量组的任一极大无关组都构成了该子空间的一组基. 留作练习

(5) 由此，若 $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_r\in F^n$ 也生成了线性空间 V，其中 $r=\dim(V)$，则 $\operatorname{rank}(\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_r)=r$. 由此，由定理 *5.3.19*(1)可知，$\boldsymbol{c}_1,\boldsymbol{c}_2,\ldots,\boldsymbol{c}_r$ 线性无关，从而是 V 的一组基.

注 5.4.6. (1) 用一 (有限长度的) 向量组来生成线性子空间 V 时，从生成集中删除冗余向量的操作，在余下的集合变成线性无关时必须停止. 如果再多删一个向量，该向量将不是剩下向量的线性组合，从而这个较小的集合将不再生成 V. 所以，基是一个尽可能小的生成集.

(2) 若 S 是 V 的一组基，在 S 中再添加进一个新的向量，比如是从 V 中取的一个 $\boldsymbol{w}$，则新的集合不再是线性无关了，这是因为 S 生成 V，因此 $\boldsymbol{w}$ 是 S 中元素的线性组合. 所以，基还是尽可能大的线性无关集.

例 5.4.7. 回忆: F^n 有一组**基本向量** $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, 其中 $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, 而 1 在 $\boldsymbol{e}_i$ 的第 i 个位置上. 容易验证: $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ 构成了 F^n 的一组基, 称为**自然基**或**标准基**. 对于任何的 $\boldsymbol{b} = (b_1, \ldots, b_n) \in F^n$, 有唯一的线性表示: $\boldsymbol{b} = \sum_{i=1}^n b_i \boldsymbol{e}_i$. 特别地, F^n 作为 F 上的线性空间的维数为 n.

例 5.4.8. (1) 在 $\mathbb{R}^2$ 中, 向量 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ 不共线 $\Leftrightarrow$ $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ 线性无关 $\Leftrightarrow$ $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ 是 $\mathbb{R}^2$ 的一组基. $\mathbb{R}^2$ 作为自身的平凡的子空间是 2 维的.

(2) 在 $\mathbb{R}^3$ 中, 向量 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 不共面 $\Leftrightarrow$ $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 线性无关 $\Leftrightarrow$ $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 是 $\mathbb{R}^3$ 的一组基. $\mathbb{R}^3$ 作为自身的平凡的子空间是 3 维的. 例如, 由于

$$\begin{vmatrix} 1 & -2 & 1 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{vmatrix} = 2 \neq 0,$$

向量组 $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$ 构成了 $\mathbb{R}^3$ 的一组基.

例 5.4.9. 矩阵 $\boldsymbol{A}$ 的行向量组的任何一个极大无关组都构成了行空间 $\mathrm{Row}(\boldsymbol{A})$ 的一组基. 若 $\boldsymbol{A}$ 通过一系列初等行变换后得到矩阵 $\boldsymbol{B}$, 不难看出行空间 $\mathrm{Row}(\boldsymbol{A}) = \mathrm{Row}(\boldsymbol{B})$. 此时, 若 $\boldsymbol{B}$ 是阶梯形矩阵, 则它的非零行的向量也显然构成了这个线性子空间的一组基.

很多应用问题可以通过从一个坐标系转化为另一坐标系, 使得问题简化. 在一个向量空间里, 转换坐标系其实和从一组基转换为另一组基在本质上是相同的. 因此, 接下来, 我们来讨论与基的转换相关的问题.

若 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是 F^n 中子空间 V 的一组基, $\boldsymbol{b} \in V$, 则存在唯一的一组标量 $\lambda_i \in F$, $1 \le i \le r$, 使得 $\boldsymbol{b} = \sum_{i=1}^r \lambda_i \boldsymbol{\alpha}_i$. 若将所有的数组向量看成列向量, 则该表达式可以写成

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_r \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_r \end{pmatrix}.$$

从而, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_r)^{\mathrm{T}}$ 是 $\boldsymbol{b}$ 在 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r$ 下的坐标向量.

在接下来的讨论中, 除非特别说明, 我们总把数组向量看成列向量.

设 $V \subseteq F^n$ 有两组不同的基: $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 和 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$. 接下来讨论相应的坐标的变换. 由于 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 可以由 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 线性表示, 存在标量 $t_{ji} \in F$, 使得 $\boldsymbol{\beta}_i = \sum_{j=1}^r t_{ij}\boldsymbol{\alpha}_j$, $1 \le i \le r$. 注意 t 的下标的顺序! 用矩阵的形式写出, 即有

$$\begin{pmatrix}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_r\end{pmatrix} = \begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r\end{pmatrix} \underbrace{\begin{pmatrix} t_{11} & \cdots & t_{r1} \\ \vdots & & \vdots \\ t_{1r} & \cdots & t_{rr}\end{pmatrix}}_{\boldsymbol{T}},$$

其中 $\boldsymbol{T} \in F^{r\times r}$ 称为从 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 到 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 的**过渡矩阵**或**转移矩阵** (transition matrix).

注 5.4.10. (1) 反之, 设有矩阵 $\boldsymbol{T}' = (t'_{ij}) \in F^{r\times r}$ 满足

$$\begin{pmatrix}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_r\end{pmatrix} = \begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r\end{pmatrix} \boldsymbol{T}',$$

则 $\boldsymbol{\beta}_j = \sum_{i=1}^r \boldsymbol{\alpha}_i t'_{ij}$, 即 $\boldsymbol{T}'$ 的第 j 个列向量是向量 $\boldsymbol{\beta}_j$ 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 下的坐标向量, 即 $\boldsymbol{T}$ 的第 j 个列向量. 这说明 $\boldsymbol{T}' = \boldsymbol{T}$. 我们可以将这一事实简述为“过渡矩阵 $\boldsymbol{T}$ 由 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 唯一确定”.

(2) 容易看出, $\boldsymbol{I}_r$ 是 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 到自身的过渡矩阵.

(3) 若从 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 到 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 的过渡矩阵为 $\boldsymbol{S}$, 即

$$\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_r\end{pmatrix} \boldsymbol{S}.$$

此时, 可以推出

$$\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_r\end{pmatrix} \boldsymbol{S} = \left(\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r\end{pmatrix} \boldsymbol{T}\right) \boldsymbol{S} = \begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r\end{pmatrix} (\boldsymbol{T}\boldsymbol{S}).$$

这说明 $\boldsymbol{T}\boldsymbol{S}$ 即为 $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_r$ 到自身的过渡矩阵 $\boldsymbol{I}_r$. 由此可以看出, $\boldsymbol{S}$ 与 $\boldsymbol{T}$ 为互逆矩阵. 特别地, 它们都是可逆矩阵.

(4) 在有些国外的教材里, 从 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 到 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 的**转移矩阵**指的是矩阵 $\boldsymbol{T}^{-1}$; 其出发点是下面马上要提到的坐标转换公式. 请在具体阅读时仔细辨别.

接着上面的的讨论. 任取向量 $\boldsymbol{v} \in V$, 假设它在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 下的坐标分别为 $\boldsymbol{X} = (x_1, \ldots, x_r)^{\mathrm{T}}$ 与 $\boldsymbol{Y} = (y_1, \ldots, y_r)^{\mathrm{T}}$, 即有

$$\boldsymbol{v} = \begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r\end{pmatrix} \boldsymbol{X}$$

$$= \begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_r \end{pmatrix} \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r \end{pmatrix} \boldsymbol{TY}.$$

从而 $\boldsymbol{X}$ 与 $\boldsymbol{TY}$ 同为 $\boldsymbol{v}$ 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 下的坐标. 由坐标表示的唯一性知 $\boldsymbol{X} = \boldsymbol{TY}$. 从而在从旧的基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 到新的基 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 的变换过程中, 从旧的坐标 $\boldsymbol{X}$ 到新的坐标 $\boldsymbol{Y}$ 满足**坐标变换公式**:

$$\boldsymbol{Y} = \boldsymbol{T}^{-1}\boldsymbol{X}.$$

例 5.4.11. *已知 $V = \mathbb{R}^3$ 的两组基 $\boldsymbol{\alpha}_1 = (1,1,1)^{\mathrm{T}}$, $\boldsymbol{\alpha}_2 = (1,0,-1)^{\mathrm{T}}$, $\boldsymbol{\alpha}_3 = (1,0,1)^{\mathrm{T}}$, 和 $\boldsymbol{\beta}_1 = (1,2,1)^{\mathrm{T}}$, $\boldsymbol{\beta}_2 = (2,3,4)^{\mathrm{T}}$, $\boldsymbol{\beta}_3 = (3,4,3)^{\mathrm{T}}$.*

(1) *求 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 到 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ 的过渡矩阵 $\boldsymbol{T}$.*

(2) *对于 $\boldsymbol{u} = 3\boldsymbol{\alpha}_1 + 2\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_3$, 求 $\boldsymbol{u}$ 在 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ 下的坐标.*

解. (1) 由定义有

$$\begin{pmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \boldsymbol{\beta}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \boldsymbol{\alpha}_3 \end{pmatrix} \boldsymbol{T}.$$

记 $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \boldsymbol{\beta}_3 \end{pmatrix}$, $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \boldsymbol{\alpha}_3 \end{pmatrix}$, 则有 $\boldsymbol{B} = \boldsymbol{AT}$. 由于 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 和 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ 都是 $\mathbb{R}^3$ 的基, $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是列满秩的方阵, 从而都是可逆矩阵. 此时, 可以推出, $\boldsymbol{T} = \boldsymbol{A}^{-1}\boldsymbol{B}$.

考虑分块矩阵的行运算,

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \end{pmatrix} \xrightarrow{\text{经过一系列初等行变换}} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{X} \end{pmatrix},$$

由于进行一系列行变换相当于左乘可逆矩阵, 我们必然有 $\boldsymbol{X} = \boldsymbol{A}^{-1}\boldsymbol{B}$, 从而可以求出 $\boldsymbol{T}$.

用这个办法, 我们看到

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \end{pmatrix} = \left(\begin{array}{ccc:ccc} 1 & 1 & 1 & 1 & 2 & 3 \\ 1 & 0 & 0 & 2 & 3 & 4 \\ 1 & -1 & 1 & 1 & 4 & 3 \end{array}\right) \xrightarrow{\text{(计算留作热身题)}} \left(\begin{array}{ccc:ccc} 1 & 0 & 0 & 2 & 3 & 4 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 \end{array}\right)$$

从而过渡矩阵为

$$\boldsymbol{T} = \boldsymbol{A}^{-1}\boldsymbol{B} = \begin{pmatrix} 2 & 3 & 4 \\ 0 & -1 & 0 \\ -1 & 0 & -1 \end{pmatrix}.$$

(2) (a) 显然, $\boldsymbol{x}$ 在 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 下的坐标为 $\boldsymbol{X} = (3, 2, -1)^{\mathrm{T}}$. 利用初等行变换, 我们有

$$\begin{pmatrix} \boldsymbol{T} & \boldsymbol{X} \end{pmatrix} = \left(\begin{array}{ccc:c} 2 & 3 & 4 & 3 \\ 0 & -1 & 0 & 2 \\ -1 & 0 & -1 & -1 \end{array}\right) \xrightarrow{\text{(计算留作热身题)}} \left(\begin{array}{ccc:c} 1 & 0 & 0 & -5/2 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 7/2 \end{array}\right)$$

因此, $\boldsymbol{u}$ 在 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ 下的坐标为 $\boldsymbol{Y} = \boldsymbol{T}^{-1}\boldsymbol{X} = (-5/2, -2, 7/2)^{\mathrm{T}}$.

(b) 换一个思路来考虑. 可直接算出 $\boldsymbol{u} = (4, 3, 0)^{\mathrm{T}}$. 为了计算 $\boldsymbol{u}$ 在 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ 下的坐标 $\boldsymbol{Y}$, 由定义, 我们有

$$\boldsymbol{u} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Y}.$$

而 $\boldsymbol{B}$ 是一个可逆方阵. 因此, $\boldsymbol{Y} = \boldsymbol{B}^{-1}\boldsymbol{u}$. 用初等行变换, 我们看到

$$\begin{pmatrix} \boldsymbol{B} & \boldsymbol{u} \end{pmatrix} = \left(\begin{array}{ccc:c} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 3 \\ 1 & 4 & 3 & 0 \end{array}\right) \xrightarrow{\text{(计算留作热身题)}} \left(\begin{array}{ccc:c} 1 & 0 & 0 & -5/2 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 7/2 \end{array}\right).$$

由此看出, $\boldsymbol{u}$ 在 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ 下的坐标为 $\boldsymbol{Y} = \boldsymbol{B}^{-1}\boldsymbol{u} = (-5/2, -2, 7/2)^{\mathrm{T}}$. □

设 V 是 F^n 的一个 r 维子空间, 则 V 的基为它的一个极大无关组 (长度必为 r). 若 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$ 是 V 中线性无关的向量, 则仿照定理 5.4.2 的证明, 可以陆续添加 $\boldsymbol{\alpha}_{s+1}, \ldots$, 使得 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s, \boldsymbol{\alpha}_{s+1}, \ldots$ 仍然线性无关, 并且生成了 V, 从而成为 V 的一组基. 显然, 新添加的向量的个数必为 $r - s$. 此时的 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 称为 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$ 的一组**扩充基**. 若 $s < r$, 一般而言, $\boldsymbol{\alpha}_{s+1}, \ldots, \boldsymbol{\alpha}_r$ 的选取不唯一.

例 5.4.12. 对于给定的行向量 $\boldsymbol{\alpha}_1 = (1, 0, 1, 0)$ 和 $\boldsymbol{\alpha}_2 = (1, -1, 2, 0)$, 求包含 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ 的 $\mathbb{R}^4$ 的一组基.

解. 我们只需找到 $\boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4 \in \mathbb{R}^4$, 使得 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_4$ 线性无关即可. 选法不唯一. 由于 $\boldsymbol{\alpha}_1$ 和 $\boldsymbol{\alpha}_2$ 写成了行向量的形式, $\boldsymbol{\alpha}_3$ 和 $\boldsymbol{\alpha}_4$ 也需要写成行向量.

容易看出, 若取 $\boldsymbol{\alpha}_3 = (0, 0, 1, 0)$, $\boldsymbol{\alpha}_4 = (0, 0, 0, 1)$, 则矩阵

$$\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \\ \boldsymbol{\alpha}_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

的行列式不为 0, 从而这四个向量线性无关. □

定理 5.4.13. 对于数组空间 F^n 有:

教材定理 5.4.2

(1) 若 V 是 F^n 的 r 维子空间, 则 V 中的任意 $r+1$ 个向量线性相关;

(2) 若 V 是 F^n 的 r 维子空间, 则 V 中的任意 r 个线性无关的向量构成 V 的一组基;

(3) 设 $U \subseteq V$ 是 F^n 的两个子空间, 则 $\dim(U) \le \dim(V)$;

(4) 设 $U \subseteq V$ 是 F^n 的两个子空间, 且 $\dim(U) = \dim(V)$, 则 $U = V$.

证明. (1) 设 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是 V 的一组基, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{r+1} \in V$. 则向量组 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{r+1}$ 可以由 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 线性表示, 从而由定理 5.3.19(3) 知

$$\operatorname{rank}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{r+1}) \le \operatorname{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r) = r,$$

其中的等号是因为 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 线性无关. 这说明向量组 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{r+1}$ 的秩严格小于它的长度. 再由定理 5.3.19(2) 知 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{r+1}$ 线性相关.

(2) 设 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是 V 的一组基, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r \in V$ 线性无关. 则向量组 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 可以由 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 线性表示, 即存在 r 阶矩阵 $\boldsymbol{T}$ 使得

$$\begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_r \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r \end{pmatrix} \boldsymbol{T}.$$

由于 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 线性无关, 利用矩阵的秩等于其列秩, 我们有

$$\operatorname{rank}(\boldsymbol{T}) \ge \operatorname{rank}\left(\begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_r \end{pmatrix} \boldsymbol{T}\right) = \operatorname{rank}\left(\begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_r \end{pmatrix}\right) = r.$$

由于 $\boldsymbol{T}$ 是 r 阶方阵, 我们也有 $\operatorname{rank}(\boldsymbol{T}) \le r$. 这说明 $\operatorname{rank}(\boldsymbol{T}) = r$, 从而 $\boldsymbol{T}$ 可逆, 从而 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 与 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 等价, 即 $V = \langle \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r \rangle = \langle \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_1 \rangle$. 由于 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 线性无关, 这说明 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r$ 是 V 的一组基.

(3) 设 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是 U 的一组基, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s$ 是 V 的一组基. 此时, $U \subseteq V$ 等价于说 $\langle \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r \rangle \subseteq \langle \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s \rangle$, 而这等价于说 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 可以由 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s$ 线性表示 (定理 5.2.7), 从而由定理 5.3.19(3) 可知,

$$\dim(U) = \operatorname{rank}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r) \le \operatorname{rank}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s) = \dim(V).$$

(4) 利用上面的记号, 则 $r = s$. 上面的 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是 r 维子空间 V 中线性无关的 r 个向量, 由 (2) 知 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 也是 V 的一组基, 从而

$$U = \langle \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r \rangle = V.$$

□

5.5　线性方程组解集的结构

对于线性方程组

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{5.4}$$

我们将讨论解的存在性、解的唯一性，以及解的 “形状”.

线性方程组解的存在性和唯一性

引理 5.5.1. *对于向量组 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in F^m$, 以下几条等价:*

(1) *$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ 线性无关;*

(2) *对任意的 $\boldsymbol{b} \in \langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \rangle$, $\boldsymbol{b}$ 可以由 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ 唯一地线性表示出来;*

(3) *存在 $\boldsymbol{b} \in \langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \rangle$, $\boldsymbol{b}$ 可以由 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ 唯一地线性表示出来.*

证明. (1) $\Rightarrow$ (2) $\Rightarrow$ (3) 是显然的. 对于 (3) $\Rightarrow$ (1), 依条件, 假设 $\boldsymbol{b} = \sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \in \langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \rangle$. 用反证法, 反设 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ 线性相关, 则存在不全为零的 $k_1, \ldots, k_n \in F$ 使得 $\sum_{i=1}^{n} k_i \boldsymbol{a}_i = \boldsymbol{0}$. 则此时 $\boldsymbol{b} = \sum_{i=1}^{n} (\lambda_i + k_i) \boldsymbol{a}_i$ 给出了 $\boldsymbol{b}$ 另外一个不同的关于 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ 的线性表示. 这与 (3) 的假设相矛盾. □

定理 5.5.2. *设 $\boldsymbol{A} \in F^{m \times n}$, 而 $\boldsymbol{x} \in F^n$ 和 $\boldsymbol{b} \in F^m$ 为列向量. 记 $\overline{\boldsymbol{A}} = (\boldsymbol{A}\ \boldsymbol{b})$ 为相应的增广矩阵. 则*

(1) *方程组 (5.4) 有解 $\Leftrightarrow \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\overline{\boldsymbol{A}})$;*

(2) *方程组 (5.4) 有解且唯一 $\Leftrightarrow \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\overline{\boldsymbol{A}}) = n$.*

证明. 设 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in F^m$ 为 $\boldsymbol{A}$ 的列向量. 我们有

(1) 我们有

$$\begin{aligned}
\text{方程组 } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \text{ 有解} &\Leftrightarrow \text{ 存在 } \boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}} \in F^n \text{ 使得 } x_1 \boldsymbol{a}_1 + \cdots + x_n \boldsymbol{a}_n = \boldsymbol{b} \\
&\Leftrightarrow \ \boldsymbol{b} \text{ 可以由 } \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \text{ 线性表示} \\
&\xLeftrightarrow{\text{定理 5.3.19(5)}} \operatorname{rank}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = \operatorname{rank}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n, \boldsymbol{b}) \\
&\xLeftrightarrow{\text{列秩等于秩}} \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\overline{\boldsymbol{A}}).
\end{aligned}$$

(2) 我们有

$$\begin{aligned}
\text{方程组 } \boldsymbol{Ax}=\boldsymbol{b} \text{ 有唯一解} &\Leftrightarrow \ \boldsymbol{b} \text{ 可由 } \boldsymbol{a}_1,\dots,\boldsymbol{a}_n \text{ 线性表示且表示唯一} \\
&\Leftrightarrow \ \boldsymbol{b} \in \langle \boldsymbol{a}_1,\dots,\boldsymbol{a}_n \rangle \text{ 且 } \boldsymbol{a}_1,\dots,\boldsymbol{a}_n \text{ 线性无关} \\
&\Leftrightarrow \operatorname{rank}(\overline{\boldsymbol{A}}) = \operatorname{rank}(\boldsymbol{A}) \text{ 且矩阵 } \boldsymbol{A} \text{ 列满秩} \\
&\Leftrightarrow \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\overline{\boldsymbol{A}}) = n.
\end{aligned}$$

□

推论 5.5.3. 关于 n 个未知元的齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{0}$ 总有零解. 它有非零解的充要条件是 $\operatorname{rank}(\boldsymbol{A})<n$. 特别地, 若 $\boldsymbol{A}$ 是 n 阶方阵, 则齐次线性方程组有非零解的充要条件是 $\det(\boldsymbol{A})=0$. 教材推论 5.5.1

齐次线性方程组解集的结构

定理 5.5.4. 关于 n 个未知元的齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{0}$ 的解的全体

$$V=\{\,\boldsymbol{x}\in F^n \mid \boldsymbol{Ax}=\boldsymbol{0}\,\}$$

是 F^n 的子空间, 满足 $\dim(V)=n-\operatorname{rank}(\boldsymbol{A})$, 即

$$\text{解集空间的自由度} = \text{全空间的维数} - \text{约束条件的维数}.$$

证明. 我们已知 V 包含零向量, 从而非空. 若 $\boldsymbol{x},\boldsymbol{y}\in V$, $\boldsymbol{Ax}=\boldsymbol{0}=\boldsymbol{Ay}$. 此时, 对于任意 $\lambda\in F$, 有 $\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y})=\boldsymbol{Ax}+\boldsymbol{Ay}=\boldsymbol{0}$, 以及 $\boldsymbol{A}(\lambda\boldsymbol{x})=\lambda\boldsymbol{Ax}=\lambda\boldsymbol{0}=\boldsymbol{0}$. 故 V 是 F^n 的子空间.

令 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$ 为 $\boldsymbol{A}$ 的列向量. 不妨设 $\operatorname{rank}(\boldsymbol{A})=r$, 且 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_r$ 是 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$ 的一个极大无关组. 此时 $\boldsymbol{a}_{r+1},\dots,\boldsymbol{a}_n$ 可以由 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_r$ 线性表示, 从而存在矩阵 $\boldsymbol{B}'\in F^{r\times(n-r)}$ 使得

$$\begin{pmatrix}\boldsymbol{a}_{r+1} & \cdots & \boldsymbol{a}_n\end{pmatrix} = \begin{pmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_r\end{pmatrix}\boldsymbol{B}'.$$

令 $\boldsymbol{B}=\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{B}'\end{pmatrix}$, 则

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n\end{pmatrix} = \begin{pmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_r\end{pmatrix}\boldsymbol{B}.$$

此时,

$$\begin{aligned}
\boldsymbol{Ax}=\boldsymbol{0} &\Leftrightarrow \begin{pmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_r\end{pmatrix}\boldsymbol{Bx}=\boldsymbol{0} \\
&\Leftrightarrow \text{向量 } \boldsymbol{Bx} \text{ 是齐次线性方程组 } \begin{pmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_r\end{pmatrix}\boldsymbol{y}=\boldsymbol{0} \text{ 的解, 其中 } \boldsymbol{y}\in F^r \text{ 为列向量}
\end{aligned}$$

$\left(\text{由于向量组 } \boldsymbol{a}_1,\dots,\boldsymbol{a}_r \text{ 线性无关, 齐次线性方程组 } \begin{pmatrix}\boldsymbol{a}_1 & \cdots & \boldsymbol{a}_r\end{pmatrix}\boldsymbol{y}=\boldsymbol{0} \text{ 仅有零解}\right)$

$\Leftrightarrow \boldsymbol{Bx}=\boldsymbol{0}.$

这说明 $\boldsymbol{Ax}=\boldsymbol{0}$ 与 $\boldsymbol{Bx}=\boldsymbol{0}$ 具有相同的解空间.

接下来, 不妨设分块矩阵 $\begin{pmatrix}-\boldsymbol{B}'\\ \boldsymbol{I}_{n-r}\end{pmatrix}$ 的列向量依次为 $\boldsymbol{b}_{r+1},\boldsymbol{b}_{r+2},\dots,\boldsymbol{b}_n$. 我们将验证它们构成了 $\boldsymbol{Bx}=\boldsymbol{0}$ 的解空间的一组基, 从而该解空间的维数为 $n-r=n-\operatorname{rank}(\boldsymbol{A})$.

- 先说明 $\boldsymbol{b}_{r+1},\dots,\boldsymbol{b}_n$ 是解. 显然 $(\boldsymbol{I}_r\ \boldsymbol{B}')\begin{pmatrix}-\boldsymbol{B}'\\ \boldsymbol{I}_{n-r}\end{pmatrix}=\boldsymbol{O}$. 这说明列向量 $\boldsymbol{b}_{r+1},\dots,\boldsymbol{b}_n$ 都是方程组 $\boldsymbol{Bx}=\boldsymbol{0}$ 的解.
- 再说明向量组 $\boldsymbol{b}_{r+1},\dots,\boldsymbol{b}_n$ 线性无关. 这是由于子阵 $\boldsymbol{I}_{n-r}$ 的出现, 从而不难看出矩阵 $\begin{pmatrix}-\boldsymbol{B}'\\ \boldsymbol{I}_{n-r}\end{pmatrix}$ 是列满秩的.
- 最后来说明 $\boldsymbol{Bx}=\boldsymbol{0}$ 的任何一个解 $\boldsymbol{\beta}=(l_1,l_2,\dots,l_n)^{\mathrm{T}}$ 都可以表示成 $\boldsymbol{b}_{r+1},\dots,\boldsymbol{b}_n$ 的线性组合. 设 $\boldsymbol{\beta}_1=(l_1,\dots,l_r)^{\mathrm{T}}$, $\boldsymbol{\beta}_2=(l_{r+1},\dots,l_n)^{\mathrm{T}}$. 那么, 由

$$\boldsymbol{0}=\boldsymbol{B\beta}=(\boldsymbol{I}_r\ \boldsymbol{B}')\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$$

可知 $\boldsymbol{\beta}_1=-\boldsymbol{B}'\boldsymbol{\beta}_2$, 从而 $\boldsymbol{\beta}=\begin{pmatrix}-\boldsymbol{B}'\boldsymbol{\beta}_2\\ \boldsymbol{\beta}_2\end{pmatrix}=\begin{pmatrix}-\boldsymbol{B}'\\ \boldsymbol{I}_{n-r}\end{pmatrix}\boldsymbol{\beta}_2$. 这说明 $\boldsymbol{\beta}$ 是矩阵 $\begin{pmatrix}-\boldsymbol{B}'\\ \boldsymbol{I}_{n-r}\end{pmatrix}$ 的列向量 $\boldsymbol{b}_{r+1},\dots,\boldsymbol{b}_n$ 的线性组合. 从而完成了证明. □

注 5.5.5. 上面的证明可用解方程组的初等语言来表述如下. $\boldsymbol{Bx}=\boldsymbol{0}$ 其实是原方程组化为约化标准形后得到的同解方程组, 形如 上课不讲

$$\begin{cases} x_1+\mu_{1,r+1}x_{r+1}+\cdots+\mu_{1,n}x_n=0,\\ x_2+\mu_{2,r+1}x_{r+1}+\cdots+\mu_{2,n}x_n=0,\\ \qquad\vdots\\ x_r+\mu_{r,r+1}x_{r+1}+\cdots+\mu_{r,n}x_n=0.\end{cases}\tag{5.5}$$

在令 $(x_{r+1},\dots,x_n)^{\mathrm{T}}$ 分别为

$$(1,0,\dots,0)^{\mathrm{T}},(0,1,0,\dots,0)^{\mathrm{T}},\dots,(0,\dots,0,1)^{\mathrm{T}}\in F^{n-r}\tag{5.6}$$

后, 方程组 (5.5) 分别有唯一解 $\boldsymbol{b}_{r+1}, \dots, \boldsymbol{b}_n$, 其中

$$\boldsymbol{b}_j = (-\mu_{1,j}, \dots, -\mu_{r,j}, 0, \dots, 0, 1, 0, \dots, 0)^{\mathrm{T}},$$

数值 1 在 $\boldsymbol{b}_j$ 的第 j 个位置上. 由于 (5.6) 中的向量组线性无关, 由教材定理 *5.2.5* 可知, 加长的向量组 $\boldsymbol{b}_{r+1}, \dots, \boldsymbol{b}_n$ 也线性无关, 从而是 $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$ 的一组线性无关的解.

接下来说明 $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$ 的任何一个解 $\boldsymbol{\beta} = (l_1, l_2, \dots, l_n)^{\mathrm{T}}$ 都可以表示成 $\boldsymbol{b}_{r+1}, \dots, \boldsymbol{b}_n$ 的线性组合. 为此, 考虑向量

$$\boldsymbol{\gamma} = \sum_{j=r+1}^{n} l_j \boldsymbol{b}_j - \boldsymbol{\beta} \xlongequal[\text{不妨设为}]{\text{由于后 } n-r \text{ 个坐标为 } 0} (d_1, \dots, d_r, 0, \dots, 0).$$

将其带入方程组 (5.5) 后不难看出, $d_1 = \dots = d_r = 0$, 即 $\boldsymbol{\gamma} = \boldsymbol{0}$. 这说明 $\boldsymbol{\beta} = \sum_{j=r+1}^{n} l_j \boldsymbol{b}_j$ 是 $\boldsymbol{b}_{r+1}, \dots, \boldsymbol{b}_n$ 的线性组合.

注 5.5.6. 受上面讨论的启发, 我们看到, 对于任意的矩阵 $\boldsymbol{A}$, 其化为约化标准形后的主元所在的列, 其实是 $\boldsymbol{A}$ 的列向量组中下标最小的极大无关组. 而约化标准形非主元的列其实给出的是这些列用主元所在列线性表示的系数的信息. 从而, $\boldsymbol{A}$ 的约化标准形是唯一的. 上课不讲

定义 5.5.7. 上面的子空间 V 称为齐次线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 的**解空间**, 也被称为系数矩阵 $\boldsymbol{A}$ 的**零空间** (并被记作 $N(\boldsymbol{A})$ 或 $\mathrm{Null}(\boldsymbol{A})$). V 的任意一组基称为该方程组的**基础解系**. 显然, 一般情形下, 基础解系并不唯一.

例 5.5.8. 求齐次线性方程组的通解:

$$\begin{cases} x_1 + 3x_2 - 5x_3 - x_4 + 2x_5 = 0 \\ 2x_1 + 6x_2 - 8x_3 + 5x_4 + 3x_5 = 0 \\ x_1 + 3x_2 - 3x_3 + 6x_4 + x_5 = 0 \end{cases}$$

证明. 作初等行变换, 我们有

$$\text{系数矩阵 } \boldsymbol{A} = \begin{pmatrix} 1 & 3 & -5 & -1 & 2 \\ 2 & 6 & -8 & 5 & 3 \\ 1 & 3 & -3 & 6 & 1 \end{pmatrix} \xrightarrow{\text{(计算留作热身题)}} \begin{pmatrix} 1 & 3 & 0 & 33/2 & -1/2 \\ 0 & 0 & 1 & 7/2 & -1/2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

故 $\mathrm{rank}(\boldsymbol{A}) = 2$, 从而基础解系的长度为 $5-2 = 3$. x_1, x_3 是主元, 从而可以选取 x_2, x_4, x_5 为自由元. 此时, 方程组可等价地化为

$$\begin{cases} x_1 = -3x_2 - \frac{33}{2}x_4 + \frac{1}{2}x_5, \\ x_3 = -\frac{7}{2}x_4 + \frac{1}{2}x_5. \end{cases}$$

若分别令 $(x_2,x_4,x_5)^{\mathrm{T}}$ 为 $(1,0,0)^{\mathrm{T}},(0,1,0)^{\mathrm{T}},(0,0,1)^{\mathrm{T}}$, 我们可以相应得到 $(x_1,x_3)^{\mathrm{T}}$ 分别为 $(-3,0)^{\mathrm{T}},(-33/2,-7/2)^{\mathrm{T}},(1/2,1/2)^{\mathrm{T}}$. 因此, 基础解系为

$$\boldsymbol{\eta}_1=(-3,1,0,0,0)^{\mathrm{T}},\quad \boldsymbol{\eta}_2=(-33/2,0,-7/2,1,0)^{\mathrm{T}},\quad \boldsymbol{\eta}_3=(1/2,0,1/2,0,1)^{\mathrm{T}}.$$

而方程组的通解为 $\boldsymbol{x}=c_1\boldsymbol{\eta}_1+c_2\boldsymbol{\eta}_2+c_3\boldsymbol{\eta}_3$, 其中 c_1,c_2,c_3 是任意常数. □

例 5.5.9. *若矩阵 $\boldsymbol{A}_{m\times n}$ 和 $\boldsymbol{B}_{n\times s}$ 满足 $\boldsymbol{AB}=\boldsymbol{O}$, 则 $\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})\le n$.*

证明. **(思路一)** 利用例 4.5.17 中的不等式, 我们有

$$\operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})-n\le\operatorname{rank}(\boldsymbol{AB})=\operatorname{rank}(\boldsymbol{O})=0.$$

(思路二) 对于矩阵 $\boldsymbol{B}$ 按列分块 $\boldsymbol{B}=\begin{pmatrix}\boldsymbol{b}_1&\cdots&\boldsymbol{b}_s\end{pmatrix}$. 则 $\boldsymbol{O}=\boldsymbol{AB}=\boldsymbol{A}\begin{pmatrix}\boldsymbol{b}_1&\cdots&\boldsymbol{b}_s\end{pmatrix}=\begin{pmatrix}\boldsymbol{Ab}_1&\cdots&\boldsymbol{Ab}_s\end{pmatrix}$. 从而对于任意的 i, 有 $\boldsymbol{Ab}_i=\boldsymbol{0}$. 这说明列向量 $\boldsymbol{b}_1,\dots,\boldsymbol{b}_s$ 是齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{0}$ 的解. 从而 $\operatorname{rank}(\boldsymbol{b}_1,\dots,\boldsymbol{b}_s)\le$ 解空间的维数, 即, $\operatorname{rank}(\boldsymbol{B})\le n-\operatorname{rank}(\boldsymbol{A})$. □

例 5.5.10. *设 $\boldsymbol{A}$ 是 $m\times n$ 的实矩阵. 证明: $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A})$.*

证明. 在数域 $F=\mathbb{R}$ 上, 我们分别考虑齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{0}$ 与 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{0}$.

我们证明它们的解空间相同:

(1) 若 $\boldsymbol{Ax}=\boldsymbol{0}$, 显然有 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{0}$;

(2) 若 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{0}$, 则 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{0}$. 而 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=(\boldsymbol{Ax})^{\mathrm{T}}(\boldsymbol{Ax})=\|\boldsymbol{Ax}\|^2$, 是向量模长的平方.* 从而 $\|\boldsymbol{Ax}\|=0$, 即 $\boldsymbol{Ax}=\boldsymbol{0}$.

此时, 由于解空间相同, 它们的空间维数相等, 即 $n-\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=n-\operatorname{rank}(\boldsymbol{A})$, 从而 $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A})$. □

注 5.5.11. *在上面的例子中, 若 $\boldsymbol{A}$ 是 $m\times n$ 维复矩阵, 用类似地方法, 我们可以证明: $\operatorname{rank}(\overline{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A})$. 其中用到了对矩阵取复共轭运算. 另一方面, 若 $\boldsymbol{A}$ 是 $m\times n$ 维复矩阵, 我们一般不再有 $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A})$. 比如, 我们可以取 $\boldsymbol{A}=(1,\mathrm{i})^{\mathrm{T}}$.*

例 5.5.12. *设 $\boldsymbol{A}$ 是 $m\times n$ 的实矩阵, $\boldsymbol{b}$ 是 m 维实的列向量. 证明: 关于 $\boldsymbol{x}$ 的线性方程组 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 必有解.*

---

*对于 $\mathbb{R}^n$ 中的向量 $\boldsymbol{x}=(x_1,\dots,x_n)^{\mathrm{T}}$, 我们一般定义其向量模长为 $\|\boldsymbol{x}\|:=(x_1^2+x_2^2+\cdots+x_n^2)^{1/2}$. 显然, $\boldsymbol{x}=\boldsymbol{0}$ 当且仅当 $\|\boldsymbol{x}\|=0$.

证明. 只需证明矩阵的秩 $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b})=\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$. 为此, 我们只需注意到

$$\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b})\underbrace{\geq}_{\text{子阵的秩}}\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}),$$

以及

$$\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A},\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b})=\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A},\boldsymbol{b}))\underbrace{\leq}_{\text{矩阵乘积的秩}}\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}})=\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}).$$ □

例 5.5.13. 若 n 阶方阵 $\boldsymbol{A}^2=\boldsymbol{I}$ 且 $\boldsymbol{A}\neq\boldsymbol{I}$, 证明 $\boldsymbol{A}+\boldsymbol{I}$ 不可逆.

证明. $\boldsymbol{A}^2=\boldsymbol{I}$ 等价于 $(\boldsymbol{A}+\boldsymbol{I})(\boldsymbol{A}-\boldsymbol{I})=\boldsymbol{O}$. 这说明 $\boldsymbol{A}-\boldsymbol{I}$ 的所有列向量是齐次方程组 $(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}$ 的解. 由于 $\boldsymbol{A}-\boldsymbol{I}\neq\boldsymbol{O}$, $\boldsymbol{A}-\boldsymbol{I}$ 存在非零列向量, 从而该齐次方程组的解空间非零. 这说明 $\operatorname{rank}(\boldsymbol{A}+\boldsymbol{I})<n$, 即, $\boldsymbol{A}+\boldsymbol{I}$ 不可逆. □

习题 5.5.14. 求解下列含参数 λ 的线性方程组

$$\begin{cases}\lambda x_1+x_2+x_3=1\\(\lambda^2+1)x_1+2\lambda x_2+(\lambda+1)x_3=\lambda+1\\x_1+x_2+\lambda x_3=1\\2x_1+(\lambda+1)x_2+(\lambda+1)x_3=2\end{cases}$$

习题 5.5.15. 设 $\boldsymbol{A}=(a_{ij})$ 为 n 阶实方阵, 其主对角线上元素为正数, 而其余元素全为负数, 并满足每行元素之和均为 0. 证明: $\operatorname{rank}(\boldsymbol{A})=n-1$. 不布置成作业

非齐次线性方程组解集的结构 我们有如下的事实: 若 γ_1,γ_2 是非齐次的线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的任意两个解, 而 γ_0 是对应的齐次方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的解, 则

(1) $\gamma_1-\gamma_2$ 是齐次方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的解,

(2) $\gamma_1+\gamma_0$ 是方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的解.

其验证很简单:

$$\begin{aligned}\boldsymbol{A}(\gamma_1-\gamma_2)&=\boldsymbol{A}\gamma_1-\boldsymbol{A}\gamma_2=\boldsymbol{b}-\boldsymbol{b}=\boldsymbol{0},\\\boldsymbol{A}(\gamma_1+\gamma_0)&=\boldsymbol{A}\gamma_1+\boldsymbol{A}\gamma_0=\boldsymbol{b}+\boldsymbol{0}=\boldsymbol{b}.\end{aligned}$$

由此可以看出, 若非齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的解集为 W, $\boldsymbol{\gamma}$ 是方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的一个特解, 而对应的齐次线性方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 的解空间为 V, 则

定理 5.5.16. 教材定理 5.5.3

$$W = \boldsymbol{\gamma} + V := \{\, \boldsymbol{\gamma} + \boldsymbol{\alpha} \mid \boldsymbol{\alpha} \in V \,\}.$$

在几何上来看, W 是一个线性空间的过点 $\boldsymbol{\gamma}$ 的平移, 这被称为一个**仿射空间**. 一般而言, 我们不会将其简称为一个空间.

特别地, 若 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{n-r}$ 是 V 的一组基 (即齐次线性方程组的一组基础解系), 则

$$W = \left\{ \boldsymbol{\gamma} + \sum_{i=1}^{n-r} t_i \boldsymbol{\alpha}_i \;\middle|\; t_i \in F \right\}.$$

注 5.5.17. 齐次方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 有时也被称为线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的**导出组**.

例 5.5.18. 已知 $\boldsymbol{\alpha}_1=(0,1,0)^{\mathrm{T}}$ 和 $\boldsymbol{\alpha}_2=(-3,2,2)^{\mathrm{T}}$ 是线性方程组

$$\begin{cases} x_1-\ \ x_2+2x_3=-1\\ 3x_1+\ \ x_2+4x_3=1\\ ax_1+bx_2+cx_3=d\end{cases}$$

的两个解. 求该方程组的通解.

证明. 该线性方程组的系数矩阵为

$$\boldsymbol{A}=\begin{pmatrix}1&-1&2\\3&1&4\\a&b&c\end{pmatrix}.$$

由于该方程组至少有两个不同的解, 其对应的线性方程组有非零解, 从而 $\operatorname{rank}(\boldsymbol{A})<3$. 另一方面, $\boldsymbol{A}$ 有非零的 2 阶子式 $\left|\begin{smallmatrix}1&-1\\3&1\end{smallmatrix}\right|=4\neq 0$, $\operatorname{rank}(\boldsymbol{A})\geq 2$. 这说明 $\operatorname{rank}(\boldsymbol{A})=2$, 从而方程组所对应的齐次方程组的解空间为 $3-2=1$ 维的. 不难看出 $\boldsymbol{\alpha}_3=\boldsymbol{\alpha}_2-\boldsymbol{\alpha}_1=(-3,1,2)^{\mathrm{T}}$ 是该解空间的一个基础解系. 此时可知, 原线性方程组有通解

$$\boldsymbol{x}=\boldsymbol{\alpha}_1+k\boldsymbol{\alpha}_3=(0,1,0)^{\mathrm{T}}+k(-3,1,2)^{\mathrm{T}},$$

其中 k 是任意常数. □

例 5.5.19. 已知有 3 维列向量组 $\boldsymbol{\alpha}_1=(1,2,-1)^{\mathrm{T}}$, $\boldsymbol{\alpha}_2=(-1,-1,2)^{\mathrm{T}}$, $\boldsymbol{\alpha}_3=(2,a,1)^{\mathrm{T}}$ 和 $\boldsymbol{\beta}=(1,2,b)^{\mathrm{T}}$.

(1) a,b 为何值时, $\boldsymbol{\beta}$ 不能由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 线性表示?

(2) a,b 为何值时, $\boldsymbol{\beta}$ 可以由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 唯一地线性表示?

(3) a,b 为何值时, $\boldsymbol{\beta}$ 可以由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 线性表示, 但是表示并不唯一?

证明. 考虑关于 x 的非齐次线性方程组

$$x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+x_3\boldsymbol{\alpha}_3=\boldsymbol{\beta}.$$

该方程组的增广矩阵

$$\overline{\boldsymbol{A}}=(\boldsymbol{A},\boldsymbol{\beta})=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\beta})=\left(\begin{array}{ccc:c}1&-1&2&1\\2&-1&a&2\\-1&2&1&b\end{array}\right)$$

$$\xrightarrow[r_1\to r_3]{-2r_1\to r_2}\begin{pmatrix}1&-1&2&\vdots&1\\0&1&a-4&\vdots&0\\0&1&3&\vdots&b+1\end{pmatrix}\xrightarrow{-r_2\to r_3}\begin{pmatrix}1&-1&2&\vdots&1\\0&1&a-4&\vdots&0\\0&0&7-a&\vdots&b+1\end{pmatrix}$$

由此看出, $\operatorname{rank}(\overline{\boldsymbol{A}})\geq\operatorname{rank}(\boldsymbol{A})\geq 2$, 且 $\operatorname{rank}(\boldsymbol{A})=2$ 的充要条件为 $a=7$, 而 $\operatorname{rank}(\overline{\boldsymbol{A}})=2$ 的充要条件是 $a=7$ 且 $b=-1$.

(1) 显然, $\boldsymbol{\beta}$ 不能由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 线性表示, 当且仅当方程组无解, 当且仅当 $\operatorname{rank}(\boldsymbol{A})=2<\operatorname{rank}(\overline{\boldsymbol{A}})=3$, 即 $a=7$ 且 $b\neq-1$.

(2) $\boldsymbol{\beta}$ 可以由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 唯一地线性表示, 当且仅当方程组有唯一解, 当且仅当 $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\overline{\boldsymbol{A}})=3$, 即 $a\neq 7$.

(3) $\boldsymbol{\beta}$ 可以由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 线性表示, 但是表示方式并不唯一, 该情况出现的充要条件是方程组有解但不唯一, 而这当且仅当 $\operatorname{rank}(\boldsymbol{A})=\operatorname{rank}(\overline{\boldsymbol{A}})=2$, 即 $a=7$ 且 $b=-1$. □

例 5.5.20. 讨论 a,b 取何值时方程组无解？何时方程组有解？在有解时, 求其通解.

$$\begin{cases}x_1+ax_2+\ \ x_3=2\\x_1+\ \ x_2+2x_3=3\\x_1+\ \ x_2+bx_3=4\end{cases}$$

解. 对该线性方程组的增广矩阵作初等行变换, 我们得到

$$\overline{\boldsymbol{A}}=(\boldsymbol{A},\boldsymbol{b})=\begin{pmatrix}1&a&1&\vdots&2\\1&1&2&\vdots&3\\1&1&b&\vdots&4\end{pmatrix}\longrightarrow\begin{pmatrix}1&a&1&\vdots&2\\&1-a&1&\vdots&1\\&&b-2&\vdots&1\end{pmatrix}.$$

由此可以看出, $\operatorname{rank}(\overline{\boldsymbol{A}})\geq\operatorname{rank}(\boldsymbol{A})\geq 2$, 并且 $\operatorname{rank}(\overline{\boldsymbol{A}})=2$ 的充要条件是 $a=1$ 且 $b=3$, 而 $\operatorname{rank}(\boldsymbol{A})=2$ 的充要条件是 $a=1$ 或 $b=2$.

(1) 当 $b=2$ 或者 $a=1$ 且 $b\neq 3$ 时, $\operatorname{rank}(\overline{\boldsymbol{A}})=3>\operatorname{rank}(\boldsymbol{A})=2$, 从而方程组无解.

(2) 当 $a=1$ 且 $b=3$ 时, $\operatorname{rank}(\overline{\boldsymbol{A}})=\operatorname{rank}(\boldsymbol{A})=2<3$, 故方程组有无穷组解. 此时, 方程组可以等价地化为

$$\begin{cases}x_1+x_2+x_3=2\\\qquad\qquad x_3=1\end{cases}$$

于是方程组的通解为

$$x_1 = 1 - t, \quad x_2 = t, \quad x_3 = 1 \qquad (t \in \mathbb{R}).$$

(此处有特解 $\boldsymbol{\gamma}_0 = (1, 0, 1)^{\mathrm{T}}$, 而相应的齐次方程组的基础解系为 $\boldsymbol{\alpha} = (1, -1, 0)^{\mathrm{T}}$. 从而, 通解可以表示成 $\boldsymbol{x} = \boldsymbol{\gamma}_0 + t\boldsymbol{\alpha}$, 其中 t 为任意实数)

(3) 当 $a \neq 1$ 且 $b \neq 2$ 时, $\operatorname{rank}(\overline{\boldsymbol{A}}) = \operatorname{rank}(\boldsymbol{A}) = 3$, 从而方程组有唯一解. 可以解出,

$$x_1 = \frac{3ab - 2b - 8a + 5}{(a-1)\,(b-2)}, \quad x_2 = -\frac{b-3}{(a-1)\,(b-2)}, \quad x_3 = \frac{1}{b-2}. \qquad \square$$

5.6 一般线性空间

在这一节, F 仍然是数域. 我们将 F 上的 n 维数组空间 F^n 推广, 考虑定义了**加法**和**数乘**的非空集合.

一般线性空间的定义

例 5.6.1. 给定非负整数 n 和数域 F, 考虑集合

$$F_n[x] := \{\, a_0 + a_1 x + \cdots + a_n x^n \mid a_0, a_1, \ldots, a_n \in F \,\}.$$

这是 F 上次数不超过 n 的以 F 为系数的多项式的全体. 这个非空集合上定义了**多项式的加法**:

$$\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i = \sum_{i=0}^{n} (a_i + b_i) x^i. \qquad (\text{系数的加法})$$

也定义了**多项式的数乘**:

$$\lambda \sum_{i=0}^{n} a_i x^i = \sum_{i=0}^{n} (\lambda a_i) x^i. \qquad (\text{系数的乘法})$$

该集合 (相对于加法的) 的**零元**:

$$0 = \sum_{i=0}^{n} a_i x^i \Leftrightarrow a_0 = a_1 = \cdots = a_n = 0. \qquad (\text{零多项式})$$

$F_n[x]$ 对 "加法" 和 "数乘" 都**运算封闭**, 即对任意的 $f, g \in F_n[x]$ 和任意的 $\lambda \in F$ 总有 $f + g,\ \lambda f \in F_n[x]$.

例 5.6.2. $F^{m \times n}$ 是数域 F 上所有的 $m \times n$ 矩阵构成的集合, 有自然的加法和数乘运算, 即矩阵的加法与数乘.

定义 5.6.3. 什么叫作域 F 上的**线性空间**或者**向量空间**? 学生自学定义. 要求结合刚才的例子和之前的 n 维数组空间 F^n, 体会这些运算规律. 教材定义 5.6.1

学生自学线性空间的基本性质. 教材 P143.

例 5.6.4 (线性空间的例子). (1) 数组空间 F^n 任意的线性子空间依照数组的加法与乘法构成 (依上面抽象定义的) 一个线性空间.

(2) 例 *5.6.1* 中的数域 F 上次数不超过 n 的以 F 为系数的多项式的全体 $F_n[x]$ 依照多项式的加法与数乘构成一个线性空间. 更一般地, 以 F 为系数的多项式的全体 $F[x]=\bigcup_{n\geq 0}F_n[x]$ 依照多项式的加法与数乘构成一个线性空间, 并且我们有如下的包含关系:

$$F=F_0[x]\subsetneq F_1[x]\subsetneq F_2[x]\subsetneq\cdots\subsetneq F[x]. \tag{5.7}$$

(3) 例 *5.6.2* 中数域 F 上所有的 $m\times n$ 矩阵构成的集合 $F^{m\times n}$ 依照矩阵的加法和数乘构成一个线性空间.

(4) 闭区间 $[a,b]$ 上 n 阶连续可导函数 (n 阶导数是连续函数) 的全体 $C^n[a,b]$, 对于函数的加法及数与函数的乘法, 构成实数域 $\mathbb{R}$ 上的线性空间, 并有如下包含关系:

$$C[a,b]=C^0[a,b]\supsetneq C^1[a,b]\supsetneq C^2[a,b]\supsetneq\cdots\supsetneq C^n[a,b]\supsetneq\cdots\supsetneq C^\infty[a,b], \tag{5.8}$$

其中 $C^\infty[a,b]:=\bigcap_{n\geq 0}C^{(n)}[a,b]$ 是 $[a,b]$ 上无穷阶可导函数的全体.

(5) 复数的全体 $\mathbb{C}$ 上有自然定义的加法. 另外, 对于复数 $z=a+bi\in\mathbb{C}$, 其中 $a,b\in\mathbb{R}$, 我们熟知 $\lambda z=\lambda a+\lambda bi\in\mathbb{C}$, 其中 $\lambda\in\mathbb{R}$. 不难看出, $\mathbb{C}$ 在这些运算下构成实数域 $\mathbb{R}$ 上的线性空间.

例 5.6.5 (不是线性空间的例子). (1) $\mathbb{R}^+$ 对于通常实数的加法和乘法, 不构成 $\mathbb{R}$ 上的线性空间. 它对于数乘不封闭: 可以考虑乘以 -1.

(2) 设 V 是 F^n 中的与某个非零向量不平行的所有向量的全体, 在向量的通常加法和数乘下, 不构成 F 上的线性空间. 例如选取 $F=\mathbb{R}$, V 是 $\mathbb{R}^2$ 中不与向量 $(1,0)$ 平行的所有向量的全体. 则 $(1,1)$ 与 $(0,-1)$ 都是 V 中的向量, 可是它们的向量和为 $(1,0)$, 不在 V 中.

(3) 次数恰好等于 n 的全体实系数多项式, 在多项式的加法与数乘多项式的运算下一般不是线性空间. 它对于加法不封闭, 例如 $n\geq 2$ 时, $f(x)=x^n+x$, $g(x)=1-x^n$, 则 $f(x)+g(x)=x+1$ 的次数不再恰好是 n 了.

(4) 数域 F 上的全体二阶可逆矩阵的集合, 对于矩阵的加法和数乘矩阵运算不构成线性空间. 它对于数乘不封闭, 任何矩阵乘以 0 后得到零矩阵, 而零矩阵不是可逆矩阵. 另外, 它对矩阵的加法也不是封闭的, 例如可以考虑矩阵 $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ 与 $\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right)$ 的矩阵加法.

学生自学教材 P144 的定义 5.6.2 和定义 5.6.3 中关于子空间与生成的子空间的内容, 需要注意 5.6.3 中的集合 S 可以为无穷集合.

例 5.6.6. (1) 在等式 (5.7) 中, 对于每个严格包含关系, 后者都是前者的线性子空间.

(2) 在等式 (5.8) 中, 对于每个严格包含关系, 后者都是前者的线性子空间.

(3) n 阶对称方阵 (*symmetric matrix*) 的全体

$$\mathrm{SM}_n(F) \coloneqq \left\{ \boldsymbol{A} \in F^{n\times n} \mid \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A} \right\}$$

和 n 阶反对称方阵 (*anti-symmetric matrix*) 的全体

$$\mathrm{AM}_n(F) \coloneqq \left\{ \boldsymbol{A} \in F^{n\times n} \mid \boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{A} \right\}$$

都是 $F^{n\times n}$ 的子空间.

(4) 设 V 是 F 上的线性空间, $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m \in V$, 而 $\lambda_1, \dots, \lambda_n \in F$, 则 $\sum_{i=1}^n \lambda_i \boldsymbol{a}_i \in V$ 称为 $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ 的一个**线性组合**或**线性表示**. $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ **生成的子空间**

$$\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_m \rangle \coloneqq \{ \lambda_1\boldsymbol{a}_1 + \lambda_2\boldsymbol{a}_2 + \cdots + \lambda_m\boldsymbol{a}_m \mid \lambda_i \in F \}$$

是 V 的子空间, 它是由 $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ 的线性组合的全体所构成的.

一般线性空间的理论

定义 5.6.7. 设 V 是数域 F 上的线性空间.

(a) 给定 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in V$, 若存在不全为零的系数 $\lambda_i \in F$, 使得线性组合 $\sum_{i=1}^n \lambda_i \boldsymbol{x}_i = \boldsymbol{0} \in V$, 则称向量组 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ **线性相关**; 否则, 则称该向量组**线性无关**.

(b) 设 $S \subseteq V$ 为一组非空向量集合, 其中 S 的元素个数可以为无穷多个.

 (i) 对于给定的向量 $\boldsymbol{\alpha} \in V$, 若存在有限多个向量 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in S$, 使得 $\boldsymbol{\alpha}$ 可以由 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ 线性表示, 则称 **$\boldsymbol{\alpha}$ 可以由 S 线性表示**.

 (ii) 若 S 的任意非空有限子集都是线性无关的, 则称 S 是**线性无关**的.

当 S 是一个有限集合时，上面的定义与前面的有限向量组时的定义一致.

例 5.6.8. 考虑例 *5.6.1* 中定义的线性空间 $F_2[x]$. 我们来判断向量 (多项式)

$$p_1(x) = x^2 - 2x + 3, \quad p_2(x) = 2x^2 + x + 8, \quad p_3(x) = x^2 + 8x + 7$$

是否线性无关. 为此，依定义，我们令

$$\lambda_1 p_1(x) + \lambda_2 p_2(x) + \lambda_3 p_3(x) = 0,$$

即

$$(\lambda_1 + 2\lambda_2 + \lambda_3)x^2 + (-2\lambda_1 + \lambda_2 + 8\lambda_3)x + (3\lambda_1 + 8\lambda_2 + 7\lambda_3) = 0.$$

从而问题转化为解齐次线性方程组

$$\begin{cases} \lambda_1 + 2\lambda_2 + \lambda_3 = 0 \\ -2\lambda_1 + \lambda_2 + 8\lambda_3 = 0 \\ 3\lambda_1 + 8\lambda_2 + 7\lambda_3 = 0. \end{cases}$$

该方程组的系数矩阵的行列式为

$$\begin{vmatrix} 1 & 2 & 1 \\ -2 & 1 & 8 \\ 3 & 8 & 7 \end{vmatrix} = 0,$$

说明原线性方程组有非零解存在. 故，p_1, p_2, p_3 是线性相关的.

学生自学教材定理 5.6.2, 定理 5.6.3, 定义 5.6.7, 定理 5.6.4, 推论 5.6.1, 定理 5.6.4, 推论 5.6.2, 定义 5.6.8, 定理 5.6.6, 定义 5.6.9, 定理 5.6.7, 定理 5.6.8. 即, P146-148 的所有定义、定理、推论. 例子学生课后自己看. 提醒学生, P148 中指出: "除非特别申明, 本书只考虑有限维线性空间. 注意: 有限维的线性空间一定可以由一组向量生成."

注 5.6.9. 若 V 是有限维线性空间，$\dim(V) = n$，$\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 是 V 的一组基，$\boldsymbol{\beta} \in V$ 在 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下有坐标 $(\lambda_1, \ldots, \lambda_n)$，即 $\boldsymbol{\beta} = \sum_{i=1}^n \lambda_i \boldsymbol{\alpha}_i$. 我们可以形式地将其写作

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

故在固定基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 的条件下，我们可以把 $\boldsymbol{\beta}$ 一一对应于 $\begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix}^{\mathrm{T}}$，即 n 维数组空间 F^n 中的一个列向量，从而把有限维线性空间的讨论转换为有限维数组空间中的相应讨论.

注 5.6.9. 若 V 是有限维线性空间, $\dim(V)=n$, $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 是 V 的一组基, $\boldsymbol{\beta}\in V$ 在 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 下有坐标 $(\lambda_1,\ldots,\lambda_n)$, 即 $\boldsymbol{\beta}=\sum_{i=1}^n \lambda_i\boldsymbol{\alpha}_i$. 我们可以形式地将其写作

$$\boldsymbol{\beta}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}.$$

故在固定基 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 的条件下, 我们可以把 $\boldsymbol{\beta}$ 一一对应于 $\begin{pmatrix}\lambda_1 & \cdots & \lambda_n\end{pmatrix}^{\mathrm{T}}$, 即 n 维数组空间 F^n 中的一个列向量, 从而把有限维线性空间的讨论转换为有限维数组空间中的相应讨论.

有了这个工具后, 我们来看如何把有限维线性空间的讨论转换为有限维数组空间中的相应讨论. 下面举两个小例子来说明这一点.

作为第一个例子, 假定我们有 V 中的一组向量 $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m$, 并设它们在这组基下的坐标向量分别是 $\boldsymbol{b}_1,\ldots,\boldsymbol{b}_m$, 而 $\boldsymbol{B}=(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_m)$ 是以这些向量为列向量的 $n\times m$ 矩阵, 则我们可以将其统一写成

$$\begin{pmatrix}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_m\end{pmatrix}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}\boldsymbol{B}.$$

若具体而言有 $\boldsymbol{B}=(b_{ij})$, 那么

$$\text{对于 } 1\le j\le m, \text{ 我们有 } \boldsymbol{\beta}_j=\sum_i \boldsymbol{\alpha}_i b_{ij}. \tag{5.9}$$

这说明我们可以将所含的抽象向量视作列向量来理解其运算规则.

假定此时另有向量组 $\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_p$, 且它可以由向量组 $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m$ 线性表示, 于是存在 $\boldsymbol{C}\in F^{n\times p}$, 使得

$$\begin{pmatrix}\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\gamma}_p\end{pmatrix}=\begin{pmatrix}\boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_m\end{pmatrix}\boldsymbol{C},$$

从而,

$$\begin{pmatrix}\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\gamma}_p\end{pmatrix}=\left(\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}\boldsymbol{B}\right)\boldsymbol{C}.$$

但是我们可以直接验证, 向量组 $\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_p$ 可以由向量组 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 以式 (5.9) 所展示的方式, 通过矩阵 $\boldsymbol{BC}$ 来线性表示. 这说明

$$\left(\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}\boldsymbol{B}\right)\boldsymbol{C}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}(\boldsymbol{BC}).$$

这提示我们, 关于矩阵的常见运算法则在这儿仍然适用.

作为第二个例子, 接着上面的讨论, 我们可以验证, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ 线性无关的充要条件是 $\boldsymbol{B}$ 是列满秩的. 这是因为 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ 是线性无关的, 当且仅当存在不全为零的 $\lambda_1, \ldots, \lambda_m$ 使得 $\sum_i \lambda_i \boldsymbol{\beta}_i = \mathbf{0}$. 但是左边的求和为

$$(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n) \boldsymbol{B} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}.$$

这说明列向量 $\boldsymbol{B}(\lambda_1, \ldots, \lambda_m)^{\mathrm{T}}$ 是 $\mathbf{0}$ 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的坐标列向量, 从而为 $\mathbf{0}$. 于是, 上面 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m$ 的线性无关性等价于 $\boldsymbol{B}(\lambda_1, \ldots, \lambda_m)^{\mathrm{T}} = \mathbf{0}$ 仅存在零解, 而这当然等价于 $\boldsymbol{B}$ 是列满秩的.

例 5.6.10. 继续考虑例 *5.6.1* 中定义的线性空间 $F_2[x]$. 显然向量 (多项式) $1, x, x^2$ 生成了这个线性空间. 它们也是线性无关的: 若

$$\lambda_1 \cdot 1 + \lambda_2 \cdot x + \lambda_3 \cdot x^2 = 0,$$

则必有 $\lambda_1 = \lambda_2 = \lambda_3 = 0$. 这说明 $F_2[x]$ 是 F 上的 3 维线性空间.

这三个向量与向量 $1, 1+x, 1+x^2$ 是等价的. 后者明显可以由前者线性表示. 反过来, 我们有 $x = (1+x) - 1$, $x^2 = (1+x^2) - 1$, 从而前者可以由后者线性表示. 用矩阵来表示, 我们有

$$\begin{pmatrix} 1 & 1+x & 1+x^2 \end{pmatrix} = \begin{pmatrix} 1 & x & x^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

以及

$$\begin{pmatrix} 1 & x & x^2 \end{pmatrix} = \begin{pmatrix} 1 & 1+x & 1+x^2 \end{pmatrix} \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

注意: 这两个“过渡矩阵”互为逆矩阵:

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \boldsymbol{I}_3.$$

例 5.6.11. (1) 更一般地, 线性空间 $F_n[x]$ 是 $n+1$ 维的, 而 $1, x, x^2, \ldots, x^n$ 是它的一组基, 称为 $F_n[x]$ 的**标准基**.

(2) 对于 $F[x]=\bigcup_{n\geq 0}F_n[x]$ (多项式的全体构成的线性空间), 不难验证 $1,x,x^2,\dots$ 是它的一组基, 从而 $F[x]$ 是无穷维线性空间. 类似地可以证明, 对于任意的非负整数 $n\geq 0$, $C^{(n)}[a,b]$ 都是无穷维线性空间; 参见例 *5.6.16* 中的讨论.

例 5.6.12. 矩阵空间 $F^{m\times n}$ 是 mn 维的, 它有一组基 $\{\,\boldsymbol{E}_{ij}\mid 1\leq i\leq m, 1\leq j\leq n\,\}$, 其中 $\boldsymbol{E}_{ij}\in F^{m\times n}$ 仅在 (i,j) 处为 1, 其它位置皆为 0. 这组基称为 $F^{m\times n}$ 的**标准基**.

例 5.6.13. 考虑数域 F 上全体二阶对称矩阵的集合 $\mathrm{SM}_2(F)$. 任何一个"向量" $\boldsymbol{A}\in\mathrm{SM}_2(F)$ 形如

$$\boldsymbol{A}=\begin{pmatrix}a&b\\b&c\end{pmatrix}=a\boldsymbol{\alpha}_1+b\boldsymbol{\alpha}_2+c\boldsymbol{\alpha}_3,$$

其中

$$\boldsymbol{\alpha}_1\coloneqq\boldsymbol{E}_{11}=\begin{pmatrix}1&0\\0&0\end{pmatrix},\quad \boldsymbol{\alpha}_2\coloneqq\boldsymbol{E}_{12}+\boldsymbol{E}_{21}=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \boldsymbol{\alpha}_3\coloneqq\boldsymbol{E}_{22}=\begin{pmatrix}0&0\\0&1\end{pmatrix}.$$

而显然 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\in\mathrm{SM}_2(F)$, 所以这三个方阵生成了 $\mathrm{SM}_2(F)$. 另外, 我们也可以很容易地看出来这三个方阵是线性无关的:

$$a\boldsymbol{\alpha}_1+b\boldsymbol{\alpha}_2+c\boldsymbol{\alpha}_3=\boldsymbol{O}\quad\Leftrightarrow\quad\begin{pmatrix}a&b\\b&c\end{pmatrix}=\boldsymbol{O}\quad\Leftrightarrow\quad a=b=c=0.$$

从而这三个方阵构成了 $\mathrm{SM}_2(F)$ 的一组基. 特别地, $\dim(\mathrm{SM}_2(F))=3$.

与之相关地, 不难看出二阶反对称矩阵的全体 $\mathrm{AM}_2(F)$ 是一维的, 由矩阵 $\boldsymbol{E}_{12}-\boldsymbol{E}_{21}=\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$ 生成.

更一般地, 对称 n 阶方阵的全体 $\mathrm{SM}_n(F)$ 构成了 $F^{n\times n}$ 的 $n(n+1)/2$ 维子空间, 而反对称 n 阶方阵的全体 $\mathrm{AM}_n(F)$ 构成了 $F^{n\times n}$ 的 $n(n-1)/2$ 维子空间. 要求学生写出各自的一组基.

注意到任何一个方阵都可以表示成一个同阶对称矩阵和一个同阶反对称矩阵的线性组合, 而一个方阵若同时为对称矩阵和反对称矩阵, 则必为零矩阵. 用教材 §5.7.3 的语言来说, $F^{n\times n}$ 是子空间 $\mathrm{SM}_n(F)$ 与 $\mathrm{AM}_n(F)$ 的**直和**, 写作 $F^{n\times n}=\mathrm{SM}_n(F)\bigoplus\mathrm{AM}_n(F)$.

例 5.6.14. 对于 $\mathbb{R}$ 上次数不超过 n 的多项式的全体 $\mathbb{R}_n[x]$, 我们暂时令 P_e 是 $\mathbb{R}_n[x]$ 中偶函数 (*even function*) 的全体, P_o 是奇函数 (*odd function*) 的全体. 容易验证 P_e 与 P_o 都是 $\mathbb{R}_n[x]$ 的子空间. 任意多项式 $f\in\mathbb{R}_n[x]$, 都可以写成 $f=f_e+f_o$, 其中

$$f_e=\frac{1}{2}\left(f(x)+f(-x)\right)\in P_e,$$

而
$$f_o = \frac{1}{2}\left(f(x) - f(-x)\right) \in P_o.$$
显然同时为奇函数与偶函数的多项式为零多项式. 用 §5.7.3 的语言来说, $\mathbb{R}_n[x]$ 是子空间 P_e 与 P_o 的直和, 写作 $\mathbb{R}_n[x] = P_e \bigoplus P_o$. 容易验证,
$$\{\, x^m \mid\ 0 \le m \le n,\ \text{且 } m \text{ 为偶数} \,\}$$
构成了 P_e 的一组基, 从而 $\dim P_e = 1 + [\frac{n}{2}]$. 其中的方括号表示高斯取整函数. 对称地,
$$\left\{\, x^m \mid\ 0 \le m \le n,\ \text{且 } m \text{ 为奇数} \,\right\}$$
构成了 P_o 的一组基, 从而 $\dim P_o = [\frac{n+1}{2}]$.

例 5.6.15. 考虑实系数的多项式
$$\begin{aligned} g_1 = 1 + x + x^2 + x^3, \quad g_2 = x + x^2 + x^3, \quad g_3 = x^2 + x^3, \quad g_4 = x^3, \\ h_1 = 1 - x - x^3, \quad h_2 = 1 + x, \quad h_3 = 1 + x - x^2, \quad h_4 = 1 - x + x^2. \end{aligned}$$

(1) 证明: g_1, g_2, g_3, g_4 与 h_1, h_2, h_3, h_4 都是线性空间 $\mathbb{R}_3[x]$ 的基.

(2) 求从基 g_1, g_2, g_3, g_4 到基 h_1, h_2, h_3, h_4 的过渡矩阵.

(3) 若 $\mathbb{R}_3[x]$ 中的元素 $p(x)$ 在基 g_1, g_2, g_3, g_4 下的坐标为 $\boldsymbol{X} = (1, -2, 0, 1)^{\mathrm{T}}$, 求 $p(x)$ 在基 h_1, h_2, h_3, h_4 下的坐标 $\boldsymbol{Y}$.

解. (1) 我们已经熟知 $1, x, x^2, x^3$ 是 $\mathbb{R}_3[x]$ 的一组基. 对于给定的向量组, 我们有
$$\begin{pmatrix} g_1 & g_2 & g_3 & g_4 \end{pmatrix} = \begin{pmatrix} 1 & x & x^2 & x^3 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}}_{\boldsymbol{T}_1},$$
以及
$$\begin{pmatrix} h_1 & h_2 & h_3 & h_4 \end{pmatrix} = \begin{pmatrix} 1 & x & x^2 & x^3 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 0 & 0 & -1 & 1 \\ -1 & 0 & 0 & 0 \end{pmatrix}}_{\boldsymbol{T}_2}.$$

由于 $|\boldsymbol{T}_1| = 1$, $|\boldsymbol{T}_2| = -2$, 故 "过渡矩阵" $\boldsymbol{T}_1$ 与 $\boldsymbol{T}_2$ 都是可逆矩阵. 这说明我们有向量组等价:

$$\{ g_1, g_2, g_3, g_4 \} \sim \{ 1, x, x^2, x^3 \} \sim \{ h_1, h_2, h_3, h_4 \}.$$

从而, g_1, g_2, g_3, g_4 与 h_1, h_2, h_3, h_4 都是线性空间 $\mathbb{R}_3[x]$ 的基.

(2) 由上,

$$\left(h_1, h_2, h_3, h_4\right) = \left(g_1, g_2, g_3, g_4\right) \boldsymbol{T}_1^{-1} \boldsymbol{T}_2,$$

从而从基 g_1, g_2, g_3, g_4 到基 h_1, h_2, h_3, h_4 的过渡矩阵为 $\boldsymbol{T} = \boldsymbol{T}_1^{-1} \boldsymbol{T}_2$. 作初等行变换, 我们有

$$\begin{pmatrix} \boldsymbol{T}_1 & \boldsymbol{T}_2 \end{pmatrix} = \left(\begin{array}{cccc:cccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & 0 & 0 & 0 \end{array}\right) \xrightarrow{\text{(计算留作热身题)}} \left(\begin{array}{cccc:cccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & -2 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 1 & -1 & -2 & 2 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 & -1 \end{array}\right).$$

从而过渡矩阵为

$$\boldsymbol{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -2 & 0 & 0 & -2 \\ 1 & -1 & -2 & 2 \\ -1 & 0 & 1 & -1 \end{pmatrix}.$$

(3) 由坐标变换公式, $p(x)$ 在在基 h_1, h_2, h_3, h_4 下的坐标为 $\boldsymbol{Y} = \boldsymbol{T}^{-1} \boldsymbol{X}$. 作初等行变换, 我们有

$$\begin{pmatrix} \boldsymbol{T} & \boldsymbol{X} \end{pmatrix} = \left(\begin{array}{cccc:c} 1 & 1 & 1 & 1 & 1 \\ -2 & 0 & 0 & -2 & -2 \\ 1 & -1 & -2 & 2 & 0 \\ -1 & 0 & 1 & -1 & 1 \end{array}\right) \xrightarrow{\text{(计算留作热身题)}} \left(\begin{array}{cccc:c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right).$$

从而坐标 $\boldsymbol{Y} = (0, -2, 2, 1)^{\mathrm{T}}$. □

例 5.6.16. 这儿, 我们将用行列式的方法来讨论 $(n-1)$ 阶连续可导函数空间 $C^{n-1}[a,b]$ 中的 n 个向量是否线性相关. 假定有函数 $f_1(x), \ldots, f_n(x) \in C^{n-1}[a,b]$.

若这 n 个线性相关, 则存在不全为零的 $\lambda_i \in \mathbb{R}$ 使得 $\sum_{i=1}^{n} \lambda_i f_i(x) = 0$. 对该函数等

式逐次求导, 我们得到了方程组

$$\begin{cases} \lambda_1 f_1(x) + \lambda_2 f_2(x) + \cdots + \lambda_n f_n(x) = 0, \\ \lambda_1 f_1'(x) + \lambda_2 f_2'(x) + \cdots + \lambda_n f_n'(x) = 0, \\ \qquad \vdots \\ \lambda_1 f_1^{(n-1)}(x) + \lambda_2 f_2^{(n-1)}(x) + \cdots + \lambda_n f_n^{(n-1)}(x) = 0. \end{cases} \tag{5.10}$$

从而对于任意的 $x_0 \in [a,b]$, 关于 λ 的齐次线性方程组

$$\begin{pmatrix} f_1(x_0) & f_2(x_0) \cdots\cdots\cdots & f_n(x_0) \\ f_1'(x_0) & f_2'(x_0) \cdots\cdots\cdots & f_n'(x_0) \\ \vdots & \vdots & \vdots \\ f_1^{(n-1)}(x_0) & f_2^{(n-1)}(x_0) \cdots\cdots & f_n^{(n-1)}(x_0) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \mathbf{0} \tag{5.11}$$

都有非零解 $(\lambda_1, \lambda_2, \ldots, \lambda_n)^{\mathrm{T}}$, 而这等价于说线性方程组 (5.11) 的系数矩阵的行列式是零. 又由于该事实对任意的 $x_0 \in [a,b]$ 都成立, 这说明, 若函数 $f_1, f_2, \ldots, f_n$ 线性相关, 则线性方程组 (5.11) 的系数矩阵的行列式是零函数.

为此, 我们称该行列式为函数 $f_1, f_2, \ldots, f_n$ 的**朗斯基行列式** (*Wronskian determinant*), 并将其记作 $W[f_1, f_2, \ldots, f_n]$, 即有

$$W[f_1, f_2, \ldots, f_n](x) \coloneqq \begin{vmatrix} f_1(x) & f_2(x) \cdots\cdots\cdots & f_n(x) \\ f_1'(x) & f_2'(x) \cdots\cdots\cdots & f_n'(x) \\ \vdots & \vdots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) \cdots\cdots & f_n^{(n-1)}(x) \end{vmatrix}.$$

由上面的推导可知:

定理. 设 $f_1, f_2, \ldots, f_n \in C^{n-1}[a,b]$. 若存在 $[a,b]$ 中的一个点 x_0 使得 $W[f_1, f_2, \ldots, f_n](x_0) \neq 0$, 则 $f_1, f_2, \ldots, f_n$ 线性无关. (注意: 这不是充要条件)

注意到多项式函数是无穷阶可微的, 作为上面定理的应用, 我们看到

$$W[1, x, x^2, x^3] = \begin{vmatrix} 1 & x & x^2 & x^3 \\ 0 & 1 & 2x & 3x^2 \\ 0 & 0 & 2 & 6x \\ 0 & 0 & 0 & 6 \end{vmatrix} = 12$$

恒不为零, 从而 $1, x, x^2, x^3$ 是 $\mathbb{R}_3[x] \subsetneq C^3[a,b]$ 中线性无关的向量 (这儿是 $F_3[x]$ 在 $F = \mathbb{R}$ 时的情形). 当然, 这一事实是显而易见的.

受此启发, 我们考察一个特例, 使得上面的定理成为充要条件. 设 $f_1, f_2, \ldots, f_n \in \mathbb{R}_{n-1}[x] \subsetneq C^{n-1}[a,b]$, 那么存在矩阵 $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ 使得

$$\begin{pmatrix} f_1 & f_2 & \cdots & f_n \end{pmatrix} = \begin{pmatrix} 1 & x & \cdots & x^{n-1} \end{pmatrix} \boldsymbol{T}.$$

从而对于 $i = 0, 1, 2, \ldots, n-1$ 有

$$\begin{pmatrix} f_1^{(i)} & f_2^{(i)} & \cdots & f_n^{(i)} \end{pmatrix} = \begin{pmatrix} 1^{(i)} & x^{(i)} & \cdots & (x^{n-1})^{(i)} \end{pmatrix} \boldsymbol{T}.$$

这说明

$$\begin{pmatrix} f_1 & f_2 & \cdots & f_n \\ f_1^{(1)} & f_2^{(1)} & \cdots & f_n^{(1)} \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)} & f_2^{(n-1)} & \cdots & f_n^{(n-1)} \end{pmatrix} = \begin{pmatrix} 1 & x & \cdots & x^{n-1} \\ 1^{(1)} & x^{(1)} & \cdots & (x^{n-1})^{(1)} \\ \vdots & \vdots & & \vdots \\ 1^{(n-1)} & x^{(n-11)} & \cdots & (x^{n-1})^{(n-1)} \end{pmatrix} \boldsymbol{T}.$$

由此不难看出

$$W[f_1, f_2, \ldots, f_n] = W[1, x, \ldots, x^{n-1}] \det(\boldsymbol{T}).$$

由于 $W[1, x, \ldots, x^{n-1}] = \prod_{i=1}^{n} i!$ 是 $\mathbb{R}$ 中的非零数, 同时 $\det(\boldsymbol{T})$ 也是实数, 故 $W[f_1, f_2, \ldots, f_n]$ 也是实数, 并且

$$f_1, f_2, \ldots, f_n \text{ 线性相关} \Leftrightarrow \boldsymbol{T} \text{ 不可逆} \Leftrightarrow \det(\boldsymbol{T}) = 0 \Leftrightarrow W[f_1, f_2, \ldots, f_n] = 0.$$

与上面的定理作比较后, 我们发现, 当把考察的函数空间变小后, 我们得到了更强的结果.

例 5.6.17. 给定矩阵 $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 4 & -2 & 1 \end{pmatrix}$. 令 V 是与 $\boldsymbol{A}$ 乘法可交换的三阶实方阵的全体. 证明: V 在矩阵的加法与数乘下构成实数域上的线性空间, 并求 V 的一组基和维数. 教材作业 #48, 课堂不讲, 留给学生课后看教案

证明. 不难验证 V 是一个线性空间. 为了找到它的一组基, 设 V 中的一般矩阵为 $\boldsymbol{B} =$

$\begin{pmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{pmatrix}$. 由于 $\boldsymbol{AB} = \boldsymbol{BA}$, 我们得到了关于 $x_1, \ldots, x_9$ 的齐次线性方程组:

$$\begin{cases} -x_2 - 4x_3 + x_7 = 0 \\ 2x_3 + x_8 = 0 \\ -x_1 - x_3 + x_9 = 0 \\ x_1 - x_5 - 4x_6 = 0 \\ x_2 + 2x_6 = 0 \\ x_3 - x_4 - x_6 = 0 \\ 4x_1 - 2x_4 + x_7 - x_8 - 4x_9 = 0 \\ 4x_2 - 2x_5 + x_8 + 2x_9 = 0 \\ 4x_3 - 2x_6 - x_7 = 0 \end{cases}$$

将系数矩阵化为约化标准形, 我们得到

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 & -\frac{9}{2} & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

由此可以看出 $\operatorname{rank}(\boldsymbol{A}) = 6$, 从而 $\dim(V) = 9 - 6 = 3$. 进一步地, 我们可以轻松写出 V 的一组基, 其相应的矩阵的形式为

$$\begin{pmatrix} 0 & 2 & 0 \\ 1 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 4 & -1 \\ 1 & 9 & -2 \\ 0 & 2 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

利用特征值与相似对角化的技巧, 我们也有其它更加揭示本质的解法. 可以参考例 6.4.11. □

5.7 子空间的运算 (※)

本节打星号, 这儿仅作简单介绍. 希望学生课后仔细阅读, 有疑问的欢迎讨论.

设 V 是数域 F 上的有限维向量空间, W_1, W_2 是 V 的子空间.

(1) 用定义不难验证子空间的**交集** $W_1 \cap W_2$ 是 V 的子空间.

(2) 一般情形下, 子空间的并集 $W_1 \cup W_2$ 不在是 V 的子空间. 例如: 可以考虑 $V = \mathbb{R}^2$ 为二维平面, $W_1 = \{ (a, 0) \mid a \in \mathbb{R} \}$ 与 $W_2 = \{ (0, b) \mid b \in \mathbb{R} \}$ 分别为 x 轴与 y 轴, 则对于 $(1,0), (0,1) \in W_1 \cup W_2$, 我们有 $(1,0) + (0,1) \notin W_1 \cup W_2$.

作为练习, 大家可以验证: $W_1 \cup W_2$ 是 V 的子空间的充要条件是 W_1 与 W_2 有包含关系, 即 $W_1 \subseteq W_2$ 或 $W_2 \subseteq W_1$.

而另外一个更困难的练习是: 假定 $W_0, W_1, \ldots, W_m$ 是实数域 $\mathbb{R}$ 上的线性空间 V 的子空间, 若 $W_0 \subseteq \bigcup_{i=1}^{n} W_i$, 则必定存在 $1 \le i \le n$, 使得 $W_0 \subseteq W_i$.

(3) 用定义不难验证, 子空间的**和**

$$W_1 + W_2 \coloneqq \{ w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2 \}$$

是 V 的子空间, 并满足 $W_1, W_2 \subseteq W_1 + W_2 \subseteq V$. 进一步地, 大家可以验证: $W_1 + W_2$ 是 V 中同时包含 W_1 与 W_2 的最小的子空间, 即, 若 V' 是 V 的子空间, 且同时包含 W_1 和 W_2, 则 $W_1 + W_2 \subseteq V'$.

之前我们提到过生成的子空间的概念. 对于 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in V$, 这些向量生成的子空间为

$$\langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \rangle = \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \,\middle|\, \lambda_i \in F \right\}.$$

不难验证, $\langle \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \rangle = \langle \boldsymbol{a}_1 \rangle + \cdots + \langle \boldsymbol{a}_n \rangle$. 当然, 大家首先要把两个子空间的和的概念合理地推广到多个子空间的和概念上去.

关于 $W_1, W_2, W_1 \cap W_2, W_1 + W_2$ 这些子空间, 我们有如下的关系.

定理 5.7.1. *设 W_1 和 W_2 是有限维向量空间 V 的子空间, 则*

$$\dim(W_1) + \dim(W_2) = \dim(W_1 \cap W_2) + \dim(W_1 + W_2). \tag{5.12}$$

在这儿我们不给出全部的证明, 仅仅指出其证明要点. 我们可以设 $\dim(W_1) = a$, $\dim(W_2) = b$, $\dim(W_1 \cap W_2) = c$. 由于 $W_1 \cap W_2 \subseteq W_1$, 我们有 $c \le a$. 对称地, 我们也有

$c \leq b$. 假设 $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_c$ 是 $W_1 \cap W_2$ 的一组基. 它们包含于 W_1 中, 并线性无关, 从而可以扩充为 W_1 的一组基 $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_c, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{a-c}$. 类似地, $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_c$ 也可以扩充为 W_2 的一组基 $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_c, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{b-c}$. 最后的任务就是证明: $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_c, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{a-c}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{b-c}$ 是 $W_1 + W_2$ 的一组基.

例 5.7.2. *对于 $V = \mathbb{R}^3$, 我们可以考虑 $W_1 = \{ (a, b, 0) \mid a, b \in \mathbb{R} \}$ 为 Oxy 平面, $W_2 = \{ (a, 0, c) \mid a, c \in \mathbb{R} \}$ 为 Oxz 平面, 则 $W_1 + W_2 = V$ (这是因为任何以过向量 (a, b, c) 都可以表示成 $(a, b, 0) + (0, 0, c)$ 的形式), 而 $W_1 \cap W_2 = \{ (a, 0, 0) \mid a \in \mathbb{R} \}$ 为 x 轴. 不难看出, 公式 (5.12) 中的等式为 $2 + 2 = 3 + 1$.*

由于子空间总是包含零向量, 子空间的交集 $W_1 \cap W_2$ 总是包含零空间的. 对于两个子空间的和, 这两个子空间的交集恰好为零空间的情形值得我们特别关注.

当 $W_1 \cap W_2 = 0$ 时, 我们称 $W_1 + W_2$ 为 W_1 和 W_2 的**直和**, 记作 $W_1 \oplus W_2$. 例如, 二维平面 $\mathbb{R}^2$ 可以视为 x 轴与 y 轴这两个线性子空间的直和. 事实上, $\mathbb{R}^2$ 是过原点的任意两条不重合的直线所代表的线性子空间的直和. 对于子空间的直和, 我们有如下的刻画.

命题 5.7.3. *设 W_1 和 W_2 是向量空间 V 的子空间. 则以下几条等价:*

(1) *对任意的向量 $u \in W_1 + W_2$, 存在唯一的向量 $w_1 \in W_1$ 和 $w_2 \in W_2$ 使得 $u = w_1 + w_2$;*

(2) *$W_1 \cap W_2 = 0$.*

并且当 V 为有限维线性空间时, 以上几条还等价于

(3) *$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2)$;*

(4) *设 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是 W_1 的一组基, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s$ 是 W_2 的一组基, 则 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_s$ 是 $W_1 + W_2$ 的一组基.*

上面的 w_1 和 w_2 将被称为直和 $W_1 \oplus W_2$ 中的向量 u 分别在 W_1 和 W_2 中的**投影**.

例 5.7.4. *在上一节中, 我们已经提到,*

(1) *数域 F 上的 n 阶方阵构成的线性空间是 $\mathrm{AM}_n(F)$ 与 $\mathrm{SM}_n(F)$ 的直和;*

(2) *实数域 $\mathbb{R}$ 上次数不超过 n 的多项式的全体 $\mathbb{R}_n[x]$ 是 P_e 与 P_o 的直和.*

第六章　线性变换

6.1　线性变换的定义与性质

例 6.1.1. 设 F 为数域, $\boldsymbol{A} \in F^{m\times n}$, $\boldsymbol{B} \in F^{n\times p}$, 则 $\boldsymbol{AB} \in F^{m\times p}$. 固定矩阵 $\boldsymbol{A}$, 考虑两个线性空间之间的映射

$$\mathscr{A}: F^{n\times p} \to F^{m\times p}, \qquad \boldsymbol{B} \mapsto \boldsymbol{AB}.$$

则由矩阵乘法的性质可以得到映射 $\mathscr{A}$ 满足以下两个条件:

(1) 对于任意的 $\boldsymbol{B}_1, \boldsymbol{B}_2 \in F^{n\times p}$, 有 $\mathscr{A}(\boldsymbol{B}_1+\boldsymbol{B}_2) = \mathscr{A}(\boldsymbol{B}_1)+\mathscr{A}(\boldsymbol{B}_2)$, 即, $\boldsymbol{A}(\boldsymbol{B}_1+\boldsymbol{B}_2) = \boldsymbol{AB}_1 + \boldsymbol{AB}_2$.

(2) 对于任意的 $\boldsymbol{B} \in F^{n\times p}$, 对于任意的 $\lambda \in F$, 有 $\mathscr{A}(\lambda\boldsymbol{B}) = \lambda\mathscr{A}(\boldsymbol{B})$, 即, $\boldsymbol{A}(\lambda\boldsymbol{B}) = \lambda(\boldsymbol{AB})$.

将此观点推广, 我们有

定义 6.1.2. 设 V, V' 为数域 F 上的两个线性空间. 若映射 $\mathscr{A}: V \to V'$ 满足条件

(保加法) 对于任意的 $\boldsymbol{x}, \boldsymbol{y} \in V$, 有 $\mathscr{A}(\boldsymbol{x}+\boldsymbol{y}) = \mathscr{A}(\boldsymbol{x}) + \mathscr{A}(\boldsymbol{y})$;

(保数乘) 对于任意的 $\boldsymbol{x} \in V$ 和 $\lambda \in F$, 有 $\mathscr{A}(\lambda\boldsymbol{x}) = \lambda\mathscr{A}(\boldsymbol{x})$,

则称 $\mathscr{A}$ 是 V 到 V' 的一个**线性映射**. 若进一步有 $V = V'$, 则称 $\mathscr{A}$ 是 V 上的一个**线性变换** (*linear transformation*) 或**线性算子** (*linear operator*).

我们将主要关注线性变换的情形.

例 6.1.3 (线性变换的例子)**.** (1) 考察数组空间 F^n, 将里面的元素看成列向量. 若 $\boldsymbol{A} \in F^{n\times n}$, 则

$$\mathscr{A}: F^n \to F^n, \qquad \boldsymbol{x} \mapsto \boldsymbol{Ax},$$

是 F^n 上的一个线性变换. 事实上, 我们将证明, F^n 上的每个线性变换都具有这样的形式.

(2) 特别地, 我们看一下平面 $\mathbb{R}^2$ 上的一些典型的线性变换.

(i) $\mathscr{A}_1 : \mathbb{R}^2 \to \mathbb{R}^2$, $\boldsymbol{x} \mapsto 3\boldsymbol{x}$, 是平面上对向量伸长 3 倍的线性变换, 对应的矩阵是 $\left(\begin{smallmatrix} 3 & 0 \\ 0 & 3 \end{smallmatrix}\right)$.

(ii) $\mathscr{A}_2 : \mathbb{R}^2 \to \mathbb{R}^2$, $(a,b)^{\mathrm{T}} \mapsto (a,0)^{\mathrm{T}}$, 是平面上投影到 x 轴的线性变换, 对应的矩阵是 $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$.

(iii) $\mathscr{A}_3 : \mathbb{R}^2 \to \mathbb{R}^2$, $(a,b)^{\mathrm{T}} \mapsto (a,-b)^{\mathrm{T}}$, 是平面上关于 x 轴作对称的线性变换, 对应的矩阵是 $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

(iv) $\mathscr{A}_4 : \mathbb{R}^2 \to \mathbb{R}^2$, $(a,b)^{\mathrm{T}} \mapsto (-b,a)^{\mathrm{T}}$, 是平面上关于原点 $\boldsymbol{O}$ 作逆时针旋转 90° 的线性变换, 对应的矩阵是 $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

(3) 集合

$$C^\infty[a,b] := \{\text{ 闭区间 } [a,b] \text{ 上的无穷可微的函数 }\}$$

是 $\mathbb{R}$ 上的无穷维线性空间. 映射

$$\begin{aligned} \mathscr{A}_1(f) &:= f', && \text{(求导)} \\ \mathscr{A}_2(f) &:= g(x) = \int_a^x f(t)\,\mathrm{d}t, && \text{(求变上限积分)} \end{aligned}$$

是 $C^\infty[a,b]$ 上的两个典型的线性变换.

(4) 对任何线性空间 V,

$$\begin{aligned} &\mathscr{E} : V \to V, \quad \boldsymbol{x} \mapsto \boldsymbol{x}, && \text{(恒等变换)} \\ &\mathscr{O} : V \to V, \quad \boldsymbol{x} \mapsto \boldsymbol{0}, && \text{(零变换)} \end{aligned}$$

是 V 上两个平凡的线性变换.

定理 6.1.4 (线性映射的性质)**.** *用前面定义 6.1.2 中的记号, 则*

(1) $\mathscr{A}(\boldsymbol{0}) = \boldsymbol{0}$. (代入 $\lambda = 0$)

(2) $\mathscr{A}(-\boldsymbol{x}) = -\mathscr{A}(\boldsymbol{x})$. (代入 $\lambda = -1$)

(3) $\mathscr{A}(\sum_{i=1}^n \lambda_i \boldsymbol{x}_i) = \sum_{i=1}^n \lambda_i \mathscr{A}(\boldsymbol{x}_i)$, 其中 $\lambda_i \in F$, $\boldsymbol{x}_i \in V$.

(4) 若 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 是 V 中线性相关的向量, 则 $\mathscr{A}(\boldsymbol{x}_1), \ldots, \mathscr{A}(\boldsymbol{x}_n)$ 亦线性相关. 作为逆否命题, 若 $\mathscr{A}(\boldsymbol{x}_1), \ldots, \mathscr{A}(\boldsymbol{x}_n)$ 线性无关, 则 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 亦线性无关. 显然, 由 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 线性无关, 我们无法推出 $\mathscr{A}(\boldsymbol{x}_1), \ldots, \mathscr{A}(\boldsymbol{x}_n)$ 亦线性无关.

定义 6.1.5. 设 $\mathscr{A}: V_1 \to V_2$ 是向量空间之间的线性映射, 则 $\mathscr{A}$ 的**核** (*kernel*) 定义为

$$\ker(\mathscr{A}) \coloneqq \{\, \boldsymbol{x} \in V_1 \mid \mathscr{A}(\boldsymbol{x}) = \boldsymbol{0} \in V_2 \,\},$$

而 $\mathscr{A}$ 的**像** (*image*) 或**值域** (*range*) 定义为

$$\operatorname{im}(\mathscr{A}) = \mathscr{A}(V_1) \coloneqq \{\, \mathscr{A}(\boldsymbol{x}) \in V_2 \mid \boldsymbol{x} \in V_1 \,\}.$$

不难验证如下的事实 (留作习题):

定理 6.1.6. 设 $\mathscr{A}: V_1 \to V_2$ 是向量空间之间的线性映射, 则 $\ker(\mathscr{A})$ 是 V_1 的子空间, 而 $\operatorname{im}(\mathscr{A})$ 是 V_2 的子空间.

例 6.1.7. 对于由给定方阵 $\boldsymbol{A} \in F^{m\times n}$ 定义的线性映射 $\mathscr{A}: F^n \to F^m$, $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$, 其像空间 $\operatorname{im}(\mathscr{A}) = \mathscr{A}(F^n)$ 是 $\boldsymbol{A}$ 的列空间, 其线性空间维数为 $\boldsymbol{A}$ 的列秩, 即 $\operatorname{rank}(\boldsymbol{A})$. 而 $\mathscr{A}$ 的核空间 $\ker(\mathscr{A})$ 是齐次线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ 的解空间. 特别地, $n = \dim(\ker(\mathscr{A})) + \dim(\operatorname{im}(\mathscr{A}))$.

注 6.1.8. 更一般地, 设 $\mathscr{A}: V_1 \to V_2$ 是向量空间之间的线性映射, 则

$$\dim(V_1) = \dim(\ker(\mathscr{A})) + \dim(\operatorname{im}(\mathscr{A})).$$

其证明可大致勾勒如下: 先找到 $\ker(\mathscr{A})$ 的一组基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$, 接下来将其扩充为 V_1 的一组基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_t$, 并且验证 $\mathscr{A}(\boldsymbol{\beta}_1), \ldots, \mathscr{A}(\boldsymbol{\beta}_t)$ 构成 $\operatorname{im}(\mathscr{A})$ 的一组基.

6.2 线性变换的矩阵

设 V 是数域 F 上的有限维线性空间, $\dim(V) = n$, 而 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 是 V 的一组基. 设 $\mathscr{A}$ 是 V 上的一个线性变换.

线性变换在一组基下的矩阵 对于 $1 \le i \le n$, 向量 $\mathscr{A}(\boldsymbol{\alpha}_i) \in V$ 可以由 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 唯一地线性表示出来:

$$\mathscr{A}(\boldsymbol{\alpha}_i) = \sum_{j=1}^{n} a_{ji}\boldsymbol{\alpha}_j, \qquad \text{其中 } a_{ji} \in F. \tag{6.1}$$

用矩阵的写法, 我们有

$$\begin{pmatrix} \mathscr{A}(\boldsymbol{\alpha}_1) & \cdots & \mathscr{A}(\boldsymbol{\alpha}_n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} \underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}}_{\boldsymbol{A}}.$$

之前我们已经提到, 我们以将抽象向量视为列向量的方式, 即式 (6.1) 中所表达的方式, 来理解上面的矩阵乘法. 其中的矩阵 $\boldsymbol{A}$ 称为**线性变换** $\mathscr{A}$ **在基** $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ **下的矩阵**. 更进一步地, 我们将 $\begin{pmatrix}\mathscr{A}(\boldsymbol{\alpha}_1) & \cdots & \mathscr{A}(\boldsymbol{\alpha}_n)\end{pmatrix}$ 形式地记作 $\mathscr{A}(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n)$, 从而有

$$\mathscr{A}(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n)=\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}\boldsymbol{A}.$$

例 6.2.1. 在之前的讨论中, 我们已经知道 $1,x,x^2,\dots,x^n$ 是 $\mathbb{R}_n[x]$ 的一组基, 并且求导运算 $\mathscr{D}$ 是 $\mathbb{R}_n[x]$ 的一个线性变换. 下面, 我们来求 $\mathscr{D}$ 在这组基下的矩阵.

证明. 对于 $i=0,1,\dots,n$, 我们有 $\mathscr{D}(x^i)=(x^i)'=ix^{i-1}$, 即,

$$\begin{aligned}\mathscr{D}(1,x,x^2,\dots,x^n)&=(0,1,2x,3x^2,\dots,nx^{n-1})\\&=\begin{pmatrix}1 & x & x^2 & \cdots & x^n\end{pmatrix}\underbrace{\begin{pmatrix}0 & 1 & 0 & \cdots & 0\\ \vdots & \ddots & 2 & \ddots & \vdots\\ & & \ddots & \ddots & 0\\ & & & \ddots & n\\ 0 & \cdots & \cdots & \cdots & 0\end{pmatrix}_{(n+1)\times(n+1)}}_{\text{所求的矩阵}}.\end{aligned}$$

□

注 6.2.2. (1) 若 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 是线性空间 V 的一组基, $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 是 V 中任意的一组向量, 则存在唯一一个 V 上的线性变换 $\mathscr{A}$, 使得对任意的 $1\le i\le n$ 有 $\mathscr{A}(\boldsymbol{\alpha}_i)=\boldsymbol{\beta}_i$. 其证明比较简单. 教材习题 #6 只要求证明了它的存在性.

(2) 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 是线性空间 V 的一组基, 对于给定的 V 上的线性变换 $\mathscr{A}$, 它在这组基下的矩阵 $\boldsymbol{A}$ 的列向量是由 $\mathscr{A}(\boldsymbol{\alpha}_1),\mathscr{A}(\boldsymbol{\alpha}_2),\dots,\mathscr{A}(\boldsymbol{\alpha}_n)$ 在这组基下的坐标唯一确定; 反之, 若给定一个 n 阶方阵 $\boldsymbol{A}$ 作为线性变换 $\mathscr{A}$ 在这组基下的矩阵, 也就等价于给出了这组基在线性变换 $\mathscr{A}$ 下的像 $\mathscr{A}(\boldsymbol{\alpha}_1),\mathscr{A}(\boldsymbol{\alpha}_2),\dots,\mathscr{A}(\boldsymbol{\alpha}_n)$, 从而也就确定了线性变换 $\mathscr{A}$. 这说明在基给定的条件下, 线性变换 $\mathscr{A}$ 与矩阵 $\boldsymbol{A}$ 有一个一一对应的关系.

(3) 若已知线性空间 V 上的线性变换 $\mathscr{A}$, 以及 $\mathscr{A}$ 在 V 的某组基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ (未知) 下的矩阵为 $\boldsymbol{A}$ (已知), 我们一般很难把这组基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 再还原出来. 例如, V 上的恒等变换在任何基下的矩阵都是单位阵 $\boldsymbol{I}$, 零变换在任何基下的矩阵都是零矩阵 $\boldsymbol{O}$.

定理 6.2.3. 设 $\mathscr{A}$ 是线性空间 V 上的线性变换, 在基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}$. 设 $\boldsymbol{x}\in V$, $\boldsymbol{y}=\mathscr{A}(\boldsymbol{x})$, 而 $\boldsymbol{x},\boldsymbol{y}$ 在基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 下的坐标分别为 $\boldsymbol{X},\boldsymbol{Y}$, 则 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$. 教材定理 6.2.1

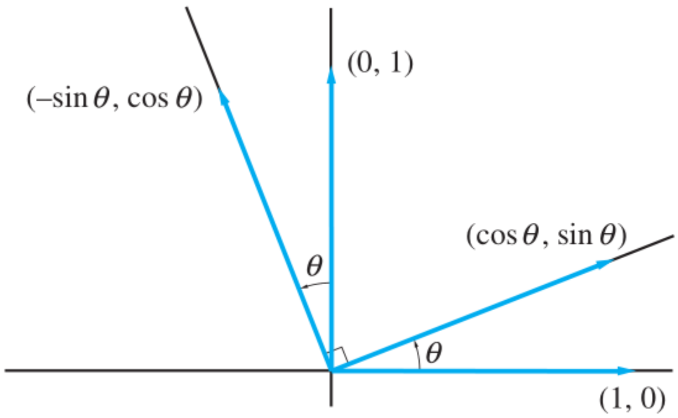
证明. 不妨设 $\boldsymbol{X}=(x_1,\dots,x_n)^{\mathrm{T}}$, 则 $\boldsymbol{x}=\sum_i x_i\boldsymbol{\alpha}_i$. 此时,

$$\begin{aligned}\boldsymbol{y}=&\mathscr{A}(\boldsymbol{x})=\mathscr{A}\left(\sum_i x_i\boldsymbol{\alpha}_i\right)\xlongequal{\mathscr{A}\text{ 的线性性质}}\sum_i x_i\mathscr{A}(\boldsymbol{\alpha}_i)=\begin{pmatrix}\mathscr{A}(\boldsymbol{\alpha}_1) & \cdots & \mathscr{A}(\boldsymbol{\alpha}_n)\end{pmatrix}\boldsymbol{X}\\ =&\left(\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}\boldsymbol{A}\right)\boldsymbol{X}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}(\boldsymbol{A}\boldsymbol{X}).\end{aligned}$$

但是, $\boldsymbol{y}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}\boldsymbol{Y}$. 由坐标表示的唯一性, 我们可以推出 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$. □

例 6.2.4. 考虑平面 $\mathbb{R}^2$ 上的逆时针旋转角度 θ 的线性变换 $\mathscr{A}$. 用几何的方法容易看出, 在 $\mathscr{A}$ 的作用下,

$$\boldsymbol{e}_1=(1,0)^{\mathrm{T}}\mapsto(\cos(\theta),\sin(\theta))^{\mathrm{T}},\qquad \boldsymbol{e}_2=(0,1)^{\mathrm{T}}\mapsto(-\sin(\theta),\cos(\theta))^{\mathrm{T}}.$$



这说明 $\mathscr{A}$ 在 $\mathbb{R}^2$ 的标准基 $\boldsymbol{e}_1,\boldsymbol{e}_2$ 下的矩阵为

$$\boldsymbol{A}=\begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}.$$

对任意的向量 $\boldsymbol{x}=(a,b)^{\mathrm{T}}\in\mathbb{R}^2$, 它在自然基 $\boldsymbol{e}_1,\boldsymbol{e}_2$ 下的坐标为 $\boldsymbol{X}=(a,b)^{\mathrm{T}}$. 由上面的定理 *6.2.3* 可知, $\mathscr{A}(\boldsymbol{x})$ 在基 $\boldsymbol{e}_1,\boldsymbol{e}_2$ 下的坐标为 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$, 从而得到

$$\mathscr{A}(\boldsymbol{x})=\begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}.$$

当然, 用几何的方法, 我们也不难验证这一旋转公式.

注 6.2.5. 比较两组基的过渡矩阵与线性变换在基下的矩阵

(1) 给定两组基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 和 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$. 则基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 到 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 的过渡矩阵 $\boldsymbol{T}$ 反映的是 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 用 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 线性表示的系数 (坐标).

(2) 给定基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 和线性变换 $\mathscr{A}$, 则 $\mathscr{A}$ 在基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 下的矩阵 $\boldsymbol{A}$ 反映的是 $\mathscr{A}(\boldsymbol{\alpha}_1),\dots,\mathscr{A}(\boldsymbol{\alpha}_n)$ 用 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 线性表示的系数 (坐标).

注意: (1) 中过渡矩阵的计算中, 我们没有用到 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 是线性无关的这条性质. 类似地, (2) 中, 我们一般也没有 $\mathscr{A}(\boldsymbol{\alpha}_1),\dots,\mathscr{A}(\boldsymbol{\alpha}_n)$ 线性无关. 关键仅仅为 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 是基. 所以上面 (1) 与 (2) 中的计算实质是一致的, 虽然反映的信息不尽相同.

例 6.2.6. 设 $V=F^n$, $\boldsymbol{A}\in F^{n\times n}$, 而 $\mathscr{A}:V\to V$ 为线性变换. 证明以下几条:

(1) 若 $\mathscr{A}$ 由法则 $\mathscr{A}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}$ 给出, 则 $\mathscr{A}$ 在自然基 $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$ 下的矩阵恰为 $\boldsymbol{A}$;

(2) 若 $\mathscr{A}$ 在自然基 $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$ 下的矩阵为 $\boldsymbol{A}$, 则 $\mathscr{A}$ 由法则 $\mathscr{A}(\boldsymbol{x})=\boldsymbol{A}\boldsymbol{x}$ 给出.

证明. (1) 我们记 $\boldsymbol{A}=(a_{ij})_{n\times n}$, 则

$$\mathscr{A}(\boldsymbol{e}_i)=\boldsymbol{A}\boldsymbol{e}_i=\begin{pmatrix}a_{1i}\\ \vdots\\ a_{ni}\end{pmatrix}=\sum_{j=1}^n a_{ji}\boldsymbol{e}_j=\begin{pmatrix}\boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n\end{pmatrix}\begin{pmatrix}a_{1i}\\ \vdots\\ a_{ni}\end{pmatrix}.$$

从而由定义有

$$\mathscr{A}(\boldsymbol{e}_1,\dots,\boldsymbol{e}_n)=\begin{pmatrix}\mathscr{A}(\boldsymbol{e}_1) & \cdots & \mathscr{A}(\boldsymbol{e}_n)\end{pmatrix}=\begin{pmatrix}\boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n\end{pmatrix}\begin{pmatrix}a_{11} & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{nn}\end{pmatrix}.$$

故 $\mathscr{A}$ 在 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 下的矩阵恰为 $\boldsymbol{A}$.

(2) **(思路一)** 直接证明. 设 $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$ 依次为 $\boldsymbol{A}$ 的 n 个列向量. 由定义, 对于每个 i, 有

$$\mathscr{A}(\boldsymbol{e}_i)=(\boldsymbol{e}_1\ \cdots\ \boldsymbol{e}_n)\boldsymbol{a}_i=\boldsymbol{a}_i.$$

因此, 对任意的 $\boldsymbol{x}=(x_1,\dots,x_n)^{\mathrm{T}}\in F^n$, 我们有

$$\begin{aligned}\mathscr{A}(\boldsymbol{x})&=\mathscr{A}\left(\sum_{i=1}^n x_i\boldsymbol{e}_i\right)=\sum_{i=1}^n x_i\mathscr{A}(\boldsymbol{e}_i)\\ &=\sum_{i=1}^n x_i\boldsymbol{a}_i=(\boldsymbol{a}_1\ \cdots\ \boldsymbol{a}_n)\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\\ &=\boldsymbol{A}\boldsymbol{x}.\end{aligned}$$

(思路二) 由注 6.2.2 知在 F^n 上存在唯一的线性变换 $\mathscr{A}$ 使得其在自然基下的矩阵为给定的 $\boldsymbol{A}$. 另一方面, 由法则 $\mathscr{A}':\boldsymbol{x}\mapsto\boldsymbol{A}\boldsymbol{x}$ 给出的 F^n 上的映射也是符合条件的线性变换. 由唯一性可知 $\mathscr{A}=\mathscr{A}'$. □

线性变换在不同基下的矩阵 接下来讨论的重点是: 例 6.2.6 中 $\mathscr{A}$ 在其它基下的矩阵是否仍然为 $\boldsymbol{A}$. 若否, 则新的矩阵与原来的 $\boldsymbol{A}$ 有什么关系?

定理 6.2.7. 设线性变换 $\mathscr{A}: V \to V$ 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 与 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ 下的矩阵分别为 $\boldsymbol{A}$ 与 $\boldsymbol{B}$. 设基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 到 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ 的过渡矩阵为 $\boldsymbol{T}$, 即, 教材定理 6.2.2

$$\begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} \boldsymbol{T},$$

则 $\boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$.

证明. 我们已知

$$\begin{aligned} \mathscr{A}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n) &= \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} \boldsymbol{A}, \\ \mathscr{A}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n) &= \begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_n \end{pmatrix} \boldsymbol{B}. \end{aligned}$$

若 $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n$ 依次为 $\boldsymbol{T}$ 的列向量, 则它们依次为 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的坐标. 由定理 6.2.3 可知, $\boldsymbol{A}\boldsymbol{t}_1, \ldots, \boldsymbol{A}\boldsymbol{t}_n$ 依次为 $\mathscr{A}(\boldsymbol{\beta}_1), \ldots, \mathscr{A}(\boldsymbol{\beta}_n)$ 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的坐标, 即

$$\mathscr{A}(\boldsymbol{\beta}_i) = \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} (\boldsymbol{A}\boldsymbol{t}_i), \qquad 1 \le i \le n.$$

此时,

$$\begin{aligned} \mathscr{A}(\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_n) &= \begin{pmatrix} \mathscr{A}(\boldsymbol{\beta}_1) & \cdots & \mathscr{A}(\boldsymbol{\beta}_n) \end{pmatrix} \\ &= \left(\begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} (\boldsymbol{A}\boldsymbol{t}_1), \ldots, \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} (\boldsymbol{A}\boldsymbol{t}_n) \right) \\ &= \begin{pmatrix} \boldsymbol{\alpha}_1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{A}\boldsymbol{t}_1 & \cdots & \boldsymbol{A}\boldsymbol{t}_n \end{pmatrix} \\ &= \left(\begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_n \end{pmatrix} \boldsymbol{T}^{-1} \right) \left(\boldsymbol{A} \begin{pmatrix} \boldsymbol{t}_1 & \cdots & \boldsymbol{t}_n \end{pmatrix} \right) \\ &= \begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_n \end{pmatrix} (\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}). \end{aligned}$$

故 $\boldsymbol{B} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$. □

例 6.2.8. 设线性变换 $\mathscr{A}$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 下的矩阵为 $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, 求 $\mathscr{A}$ 在基 $\boldsymbol{\alpha}_3, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1$ 下的矩阵 $\boldsymbol{B}$.

解. 不难看出, 基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3$ 到基 $\boldsymbol{\alpha}_3, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_1$ 的过渡矩阵为 $\boldsymbol{T} = \begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix}$, 并且 $\boldsymbol{T}^{-1} =$

$\boldsymbol{T}$. 从而,

$$\begin{aligned}\boldsymbol{B} =&\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} = \begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}\\ \xlongequal{r_1\leftrightarrow r_3}&\begin{pmatrix}7&8&9\\4&5&6\\1&2&3\end{pmatrix}\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}\xlongequal{c_1\leftrightarrow c_3}\begin{pmatrix}9&8&7\\6&5&4\\3&2&1\end{pmatrix}.\end{aligned}$$

□

例 6.2.9. 设 $\mathscr{A}: F^3\to F^3$ 为线性变换, 满足 教材例题 6.2.2

$$\begin{aligned}\boldsymbol{\alpha}_1 =&(2,3,5)^{\mathrm{T}} &\mapsto&\quad \boldsymbol{\beta}_1=(1,2,0)^{\mathrm{T}},\\ \boldsymbol{\alpha}_2 =&(0,1,2)^{\mathrm{T}} &\mapsto&\quad \boldsymbol{\beta}_2=(2,4,-1)^{\mathrm{T}},\\ \boldsymbol{\alpha}_3 =&(1,0,0)^{\mathrm{T}} &\mapsto&\quad \boldsymbol{\beta}_3=(3,0,5)^{\mathrm{T}}.\end{aligned}$$

求

(1) $\mathscr{A}$ 在 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 下的矩阵 $\boldsymbol{A}$;

(2) $\mathscr{A}$ 在自然基下的矩阵 $\boldsymbol{B}$.

证明. 注意到 $\det(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)=1\neq 0$, 故 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 是 F^3 的一组基.

(1) 由定义,

$$\begin{pmatrix}\boldsymbol{\alpha}_1&\boldsymbol{\alpha}_2&\boldsymbol{\alpha}_3\end{pmatrix}\boldsymbol{A}=\mathscr{A}(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)=\begin{pmatrix}\boldsymbol{\beta}_1&\boldsymbol{\beta}_2&\boldsymbol{\beta}_3\end{pmatrix},$$

从而

$$\begin{aligned}\boldsymbol{A}&=\begin{pmatrix}\boldsymbol{\alpha}_1&\boldsymbol{\alpha}_2&\boldsymbol{\alpha}_3\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{\beta}_1&\boldsymbol{\beta}_2&\boldsymbol{\beta}_3\end{pmatrix}\\&=\begin{pmatrix}2&0&1\\3&1&0\\5&2&0\end{pmatrix}^{-1}\begin{pmatrix}1&2&3\\2&4&0\\0&-1&5\end{pmatrix}\xlongequal{\text{留作热身题}}\begin{pmatrix}4&9&-5\\-10&-23&15\\-7&-16&13\end{pmatrix}.\end{aligned}$$

具体计算如下:

$$\left(\begin{array}{ccc:ccc}2&0&1&1&2&3\\3&1&0&2&4&0\\5&2&0&0&-1&5\end{array}\right)\xrightarrow{-2r_1\to r_2}\left(\begin{array}{ccc:ccc}2&0&1&1&2&3\\1&1&-1&1&2&-3\\5&2&0&0&-1&5\end{array}\right)\xrightarrow{r_1\leftrightarrow r_2}$$

$$
\left(\begin{array}{ccc|ccc}1&1&-1&1&2&-3\\2&0&1&1&2&3\\5&2&0&0&-1&5\end{array}\right)\xrightarrow[-5r_1\to r_3]{-2r_1\to r_2}\left(\begin{array}{ccc|ccc}1&1&-1&1&2&-3\\0&-2&3&-1&-2&9\\0&-3&5&-5&-11&20\end{array}\right)\xrightarrow{-r_3\to r_2}
$$

$$
\left(\begin{array}{ccc|ccc}1&1&-1&1&2&-3\\0&1&-2&4&9&-11\\0&-3&5&-5&-11&20\end{array}\right)\xrightarrow[3r_2\to r_3]{-r_2\to r_1}\left(\begin{array}{ccc|ccc}1&0&1&-3&-7&8\\0&1&-2&4&9&-11\\0&0&-1&7&16&-13\end{array}\right)\xrightarrow[-2r_3\to r_2]{r_3\to r_1}
$$

$$
\left(\begin{array}{ccc|ccc}1&0&0&4&9&-5\\0&1&0&-10&-23&15\\0&0&-1&7&16&-13\end{array}\right)\xrightarrow{-r_3}\left(\begin{array}{ccc|ccc}1&0&0&4&9&-5\\0&1&0&-10&-23&15\\0&0&1&-7&-16&13\end{array}\right).
$$

(2) $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3$ 在自然基下的坐标就为向量本身, 故由定理 6.2.3 可知, 对于 $i=1,2,3$, 有 $\boldsymbol{\beta}_i=\boldsymbol{B}\boldsymbol{\alpha}_i$, 从而

$$
\begin{pmatrix}\boldsymbol{\beta}_1&\boldsymbol{\beta}_2&\boldsymbol{\beta}_3\end{pmatrix}=\boldsymbol{B}\begin{pmatrix}\boldsymbol{\alpha}_1&\boldsymbol{\alpha}_2&\boldsymbol{\alpha}_3\end{pmatrix},
$$

故

$$
\begin{aligned}
\boldsymbol{B}&=\begin{pmatrix}\boldsymbol{\beta}_1&\boldsymbol{\beta}_2&\boldsymbol{\beta}_3\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}_1&\boldsymbol{\alpha}_2&\boldsymbol{\alpha}_3\end{pmatrix}^{-1}\\
&=\begin{pmatrix}1&2&3\\2&4&0\\0&-1&5\end{pmatrix}\begin{pmatrix}2&0&1\\3&1&0\\5&2&0\end{pmatrix}^{-1}\xlongequal{\text{留作热身题}}\begin{pmatrix}3&-20&11\\0&-16&10\\5&-15&7\end{pmatrix}.
\end{aligned}
$$

具体计算如下:

$$
\left(\begin{array}{ccc}1&2&3\\2&4&0\\0&-1&5\\\hline 2&0&1\\3&1&0\\5&2&0\end{array}\right)\xrightarrow{c_1\leftrightarrow c_3}\left(\begin{array}{ccc}3&2&1\\0&4&2\\5&-1&0\\\hline 1&0&2\\0&1&3\\0&2&5\end{array}\right)\xrightarrow{-2c_1\to c_3}\left(\begin{array}{ccc}3&2&-5\\0&4&2\\5&-1&-10\\\hline 1&0&0\\0&1&3\\0&2&5\end{array}\right)\xrightarrow{-3c_2\to c_3}
$$

$$
\left(\begin{array}{ccc}3&2&-11\\0&4&-10\\5&-1&-7\\\hline 1&0&0\\0&1&0\\0&2&-1\end{array}\right)\xrightarrow{-c_3}\left(\begin{array}{ccc}3&2&11\\0&4&10\\5&-1&7\\\hline 1&0&0\\0&1&0\\0&2&1\end{array}\right)\xrightarrow{-2c_3\to c_2}\left(\begin{array}{ccc}3&-20&11\\0&-16&10\\5&-15&7\\\hline 1&0&0\\0&1&0\\0&0&1\end{array}\right).
$$

□

习题 6.2.10. 设 $\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3$ 构成向量空间 V 的一组基, 而 V 上的线性变换 $\mathscr{A}$ 在这组基下的矩阵为 $\begin{pmatrix} 0 & -6 & 1 \\ 0 & 5 & -1 \\ 1 & -2 & 7 \end{pmatrix}$. 计算 $\mathscr{A}(3\boldsymbol{\beta}_1-4\boldsymbol{\beta}_2+5\boldsymbol{\beta}_3)$.

矩阵的相似

回忆. 在定理 6.2.7 中, 设线性变换 $\mathscr{A}:V\to V$ 在基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 与 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 下的矩阵分别为 $\boldsymbol{A}$ 与 $\boldsymbol{B}$. 设基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 到 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 的过渡矩阵为 $\boldsymbol{T}$, 即,

$$\begin{pmatrix}\boldsymbol{\beta}_1 & \dots & \boldsymbol{\beta}_n\end{pmatrix}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \dots & \boldsymbol{\alpha}_n\end{pmatrix}\boldsymbol{T},$$

则 $\boldsymbol{B}=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$.

一般地, 考察 $\boldsymbol{B}=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$, 其中 $\boldsymbol{T}$ 可逆, 则矩阵 $\boldsymbol{B}$ 和 $\boldsymbol{A}$ 本质上反映了同一个东西 (线性变换 $\mathscr{A}$).

定义 6.2.11. 设 $\boldsymbol{A},\boldsymbol{B}$ 为数域 F 上的两个 n 阶方阵. 若存在 F 上的同阶可逆方阵 $\boldsymbol{T}$ 使得 $\boldsymbol{B}=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$, 则称 $\boldsymbol{B}$ 是由 $\boldsymbol{A}$ 做**相似变换**得到的, 称 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ **(在 F 上) 相似** (*similar*), 记作 $A\sim B$. 若 $F=\mathbb{R}$, 则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ **实相似**; 若 $F=\mathbb{C}$, 则称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ **复相似**.

命题 6.2.12. 同阶方阵的相似关系为等价关系, 即 教材命题 6.2.1

(反身性) $\boldsymbol{A}$ 与自身相似;

(对称性) 若 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似, 则反过来 $\boldsymbol{B}$ 也与 $\boldsymbol{A}$ 相似;

(传递性) 若 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似, $\boldsymbol{B}$ 与 $\boldsymbol{C}$ 相似, 则 $\boldsymbol{A}$ 也与 $\boldsymbol{C}$ 相似.

证明. 用定义可以直接验证, 略. □

因此, 当 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 是相似的时候, 我们也称它们为**相似等价**的.

推论 6.2.13. 有限维线性空间 V 上的一个线性变换在 V 的不同基下的矩阵是相似的.

命题 6.2.14 (相似矩阵的基本性质)**.** (1) 若 $\boldsymbol{A}\sim\boldsymbol{B}$, 则 $\boldsymbol{A}^m\sim\boldsymbol{B}^m$, 其中 m 为正整数.

(2) 若 $\boldsymbol{A}\sim\boldsymbol{B}$, 设 $f(x)$ 是一个一元多项式, 则矩阵多项式 $f(\boldsymbol{A})\sim f(\boldsymbol{B})$.

(3) 若 $\boldsymbol{A}_i\sim\boldsymbol{B}_i$, $i=1,2,\dots,s$, 则 $\operatorname{diag}(\boldsymbol{A}_1,\boldsymbol{A}_2,\dots,\boldsymbol{A}_s)\sim\operatorname{diag}(\boldsymbol{B}_1,\boldsymbol{B}_2,\dots,\boldsymbol{B}_s)$.

(4) 若 $\boldsymbol{A}\sim\boldsymbol{B}$, 且 $\boldsymbol{A}$ 可逆, 则 $\boldsymbol{B}$ 也可逆, 且 $\boldsymbol{A}^{-1}\sim\boldsymbol{B}^{-1}$.

(5) 相似的矩阵有相同的行列式、迹和秩.

证明. (1) 若 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 则对任意的正整数 m, 有

$$\boldsymbol{A}^m=\underbrace{(\boldsymbol{T}^{-1}\boldsymbol{B}\underbrace{\boldsymbol{T})(\boldsymbol{T}^{-1}}_{\text{相乘抵消}}\boldsymbol{B}\boldsymbol{T})\cdots(\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T})}_{m\text{组}}=\boldsymbol{T}^{-1}\boldsymbol{B}^m\boldsymbol{T},$$

(2) 若 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 而 $f(x)=\sum_{i=0}^m a_ix^i$, 那么

$$\begin{aligned}f(\boldsymbol{A})&=\sum_{i=0}^m a_i\boldsymbol{A}^i=\sum_{i=0}^m a_i(\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T})^m=\sum_{i=0}^m a_i\boldsymbol{T}^{-1}\boldsymbol{B}^m\boldsymbol{T}\\&=\boldsymbol{T}^{-1}\left(\sum_{i=0}^m a_i\boldsymbol{B}^i\right)\boldsymbol{T}=\boldsymbol{T}^{-1}f(\boldsymbol{B})\boldsymbol{T}.\end{aligned}$$

(3) 若 $\boldsymbol{A}_i=\boldsymbol{T}_i^{-1}\boldsymbol{B}_i\boldsymbol{T}_i$, $i=1,2,\dots,s$, 那么

$$\begin{aligned}\boldsymbol{A}&=\operatorname{diag}(\boldsymbol{A}_1,\dots,\boldsymbol{A}_s)=\operatorname{diag}(\boldsymbol{T}_1^{-1}\boldsymbol{B}_1\boldsymbol{T}_1,\dots,\boldsymbol{T}_s^{-1}\boldsymbol{B}_s\boldsymbol{T}_s)\\&=\operatorname{diag}(\boldsymbol{T}_1^{-1},\dots,\boldsymbol{T}_s^{-1})\operatorname{diag}(\boldsymbol{B}_1,\dots,\boldsymbol{B}_s)\operatorname{diag}(\boldsymbol{T}_1,\dots,\boldsymbol{T}_s)\\&=\underbrace{\operatorname{diag}(\boldsymbol{T}_1,\dots,\boldsymbol{T}_s)^{-1}}_{\boldsymbol{T}^{-1}}\operatorname{diag}(\boldsymbol{B}_1,\dots,\boldsymbol{B}_s)\underbrace{\operatorname{diag}(\boldsymbol{T}_1,\dots,\boldsymbol{T}_s)}_{\text{记作 }\boldsymbol{T}}.\end{aligned}$$

(4) 若 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 且 $\boldsymbol{A}$ 可逆, 那么 $\boldsymbol{B}=\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}$ 是可逆矩阵的乘积, 从而也可逆, 并有 $\boldsymbol{A}^{-1}=\boldsymbol{T}^{-1}\boldsymbol{B}^{-1}\boldsymbol{T}$.

(5) 若 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 则

$$\begin{aligned}\det(\boldsymbol{B})&=\det((\boldsymbol{T}^{-1}\boldsymbol{A})\boldsymbol{T})=\det(\boldsymbol{T}(\boldsymbol{T}^{-1}\boldsymbol{A}))=\det(\boldsymbol{A}),\\\operatorname{tr}(\boldsymbol{B})&=\operatorname{tr}((\boldsymbol{T}^{-1}\boldsymbol{A})\boldsymbol{T})=\operatorname{tr}(\boldsymbol{T}(\boldsymbol{T}^{-1}\boldsymbol{A}))=\operatorname{tr}(\boldsymbol{A}).\end{aligned}$$

又由于 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相抵等价, 从而有相同的秩. □

注 6.2.15. (1) 若同阶方阵满足 $\boldsymbol{A}_1\sim\boldsymbol{A}_2$, $\boldsymbol{B}_1\sim\boldsymbol{B}_2$, 我们一般情况下是无法推出 $(\boldsymbol{A}_1+\boldsymbol{B}_1)\sim(\boldsymbol{A}_2+\boldsymbol{B}_2)$ 的, 也无法推出 $(\boldsymbol{A}_1\boldsymbol{B}_1)\sim(\boldsymbol{A}_2\boldsymbol{B}_2)$. 请自行举例说明这一点.

(2) 在第四章中我们见到, 一般矩阵的相抵关系也是等价关系. 相抵的同型矩阵构成一个等价类, 我们称之为相抵等价类. 同型的矩阵相抵当且仅当它们有相同的相抵标准形, 当且仅当它们有相等的秩. 秩被称为一个**相抵不变量**, 即, 相抵的矩阵通过计算都会得到相等的量.

(3) 相似的同阶方阵也构成一个等价类, 称为**相似等价类**. 对于相似等价类, 我们也可以讨论相应的**相似标准形** (请参考教材 *§6.5* 若尔当标准形这一小节的内容) 和**相似不变量** (即相似的矩阵通过计算都会得到相等的量).

(4) 谈相抵矩阵时, 不要求矩阵为方阵. 谈相似矩阵时, 必须要求为方阵.

(5) 在前面的讨论中, 我们已经见到了行列式、迹和秩这三个常见的相似不变量. 由此很容易看出, 虽然相似的方阵必然相抵, 但是反过来一般不对.

(6) 用定义容易验证, 与单位矩阵 $\boldsymbol{I}_n$ 相似的矩阵仅有 $\boldsymbol{I}_n$ 自身, 这说明 $\boldsymbol{I}_n$ 所在的相似等价类仅有 $\boldsymbol{I}_n$ 一个元素. 更一般地, 不难验证, 若一个相似等价类仅有一个元素, 则它必为某个数量阵 (形如 $a\boldsymbol{I}_n$) 的相似等价类 (其证明与教材 $P114$ 作业题 $\#10$ 是类似的, 思路如下. 首先注意到 $\boldsymbol{A} = (a_{ij})_{n\times n}$ 所在的相似等价类仅有 $\boldsymbol{A}$ 的充要条件是 $\boldsymbol{A}$ 与任何可逆同阶方阵 $\boldsymbol{T}$ 乘法可交换. 接下来, 对于 $1 \leq i \neq j \leq n$, 考虑 $\boldsymbol{T} = \boldsymbol{I} + \boldsymbol{E}_{ij}$, 利用乘法可交换这一条件看出 $a_{ii} = a_{jj}$, $a_{ki} = 0$ $(k \neq i)$, $a_{jk} = 0$ $(k \neq j)$.). 除此之外, 一般而言, 相似等价类很大, 包含了很多不同的矩阵.

基于上面的讨论, 当我们谈方阵 $\boldsymbol{A}$ 可能的相似标准形, 一般不能指望其为数量阵, 那么退而求其次, 我们希望它能为对角阵.

定义 6.2.16. 给定一个方阵 $\boldsymbol{A}$, 若存在可逆阵 $\boldsymbol{T}$, 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 是对角阵, 则称 $\boldsymbol{A}$ 为**可相似对角化的**, 或简称为**可对角化的** (*diagonalizable*), 而称 $\boldsymbol{T}$ 将 $\boldsymbol{A}$ **对角化** (*diagonalize*).

注 6.2.17. 并不是每个相似等价类都有对角阵, 即, 不是每个方阵都可以对角化. 例如, 我们之后将要在例 *6.4.1* 中证明 $\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ 不可对角化.

注 6.2.18. 若 $f(x)$ 为一个一元多项式, 且 $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 在前面我们已经看到了 $f(\boldsymbol{A}) = \boldsymbol{T}^{-1}f(\boldsymbol{B})\boldsymbol{T}$. 而若 $\boldsymbol{B} = \mathrm{diag}(\mu_1, \ldots, \mu_n)$ 为对角阵, 则 $f(\boldsymbol{B}) = \mathrm{diag}(f(\mu_1), \ldots, f(\mu_n))$, 其计算非常简单, 从而 $f(\boldsymbol{A})$ 的计算也会相对比较简单. 这给出了我们关心可对角化矩阵的一个理由.

例 6.2.19. 在教材例 *4.2.6* 中, 我们考虑了 *Fibonacci* 数列, $F_1 = F_2 = 1$, $F_n = F_{n-1} + F_{n-2}$. 我们推出了递推公式

$$F_n = (1, 0)\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

这儿我们需要计算矩阵 $\boldsymbol{A}=\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$ 的高阶幂. 显然, 直接计算是比较复杂的. 但是我们可以找到可逆矩阵 $\boldsymbol{T}=\begin{pmatrix}\frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2}\\ 1 & 1\end{pmatrix}$, 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{D}=\begin{pmatrix}\frac{1+\sqrt{5}}{2} & \\ & \frac{1-\sqrt{5}}{2}\end{pmatrix}$. 从而

$$\begin{aligned}
F_n=&(1,0)\left(\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^{-1}\right)^{n-2}\begin{pmatrix}1\\ 1\end{pmatrix}=(1,0)\boldsymbol{T}\boldsymbol{D}^{n-2}\boldsymbol{T}^{-1}\begin{pmatrix}1\\ 1\end{pmatrix}\\
=&(1,0)\begin{pmatrix}\frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2}\\ 1 & 1\end{pmatrix}\begin{pmatrix}\left(\frac{1+\sqrt{5}}{2}\right)^{n-2} & \\ & \left(\frac{1-\sqrt{5}}{2}\right)^{n-2}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{5}} & \frac{\sqrt{5}-1}{2\sqrt{5}}\\ -\frac{1}{\sqrt{5}} & \frac{\sqrt{5}+1}{2\sqrt{5}}\end{pmatrix}\begin{pmatrix}1\\ 1\end{pmatrix}\\
=&\frac{\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}}.
\end{aligned}$$

在之后我们会系统地讨论这样的 $\boldsymbol{T}$ 的找法.

例 6.2.20. 如果方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=\boldsymbol{A}$, 则称 $\boldsymbol{A}$ 为一个**幂等矩阵**. 证明*:* 与幂等矩阵相似的矩阵仍然是幂等矩阵.

证明. 只需验证: 对于任意的可逆矩阵 $\boldsymbol{T}$ 有 $(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T})^2=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 当且仅当 $\boldsymbol{A}^2=\boldsymbol{A}$. 而这是显然的. □

例 6.2.21. 如果方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=\boldsymbol{I}$, 则称 $\boldsymbol{A}$ 为一个**对合矩阵**. 证明*:* 与对合矩阵相似的矩阵仍然是对合矩阵.

证明. 只需验证: 对于任意的可逆矩阵 $\boldsymbol{T}$ 有 $(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T})^2=\boldsymbol{I}$ 当且仅当 $\boldsymbol{A}^2=\boldsymbol{I}$. 而这是显然的. □

例 6.2.22. 如果方阵 $\boldsymbol{A}$ 的某个正整数次幂对于零矩阵, 则称 $\boldsymbol{A}$ 为一个**幂零矩阵**. 使得 $\boldsymbol{A}^s=\boldsymbol{O}$ 成立的最小正整数 s 称为 $\boldsymbol{A}$ 的**幂零指数**. 证明*:* 与幂零矩阵相似的矩阵仍是幂零矩阵, 并且它们的幂零指数相等.

证明. 只需验证: 对于任意的可逆矩阵 $\boldsymbol{T}$ 和任意的正整数 s, 有 $(\boldsymbol{T}^{-1}\boldsymbol{A}^s\boldsymbol{T})^s=\boldsymbol{O}$ 当且仅当 $\boldsymbol{A}^s=\boldsymbol{O}$. 而这是显然的. □

例 6.2.23. 证明*:* 如果 $\boldsymbol{A}$ 可以对角化, 则 $\boldsymbol{A}\sim\boldsymbol{A}^{\mathrm{T}}$. (事实上, 在命题 *6.5.2* 中我们将利用若尔当标准形的理论来证明该结论对于一般的复方阵也成立, 从而不需要加"可对角化"这一条件.)

证明. 若 $\boldsymbol{A}$ 可对角化, 则存在可逆矩阵 $\boldsymbol{T}$ 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{B}=\mathrm{diag}(\mu_1,\ldots,\mu_n)$. 取转置运算, 我们有 $\boldsymbol{T}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{T}^{\mathrm{T}})^{-1}=\boldsymbol{B}^{\mathrm{T}}$. 但是由于 $\boldsymbol{B}$ 为对角阵, 我们有 $\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{B}$. 从而 $\boldsymbol{A}\sim\boldsymbol{B}=\boldsymbol{B}^{\mathrm{T}}\sim\boldsymbol{A}^{\mathrm{T}}$. 这说明 $\boldsymbol{A}\sim\boldsymbol{A}^{\mathrm{T}}$. □

6.3 特征值与特征向量

一个很自然的问题是, 例 6.2.19 中的 $\boldsymbol{T}$ 与 $\boldsymbol{D}$ 是如何找到的? 我们可以换一个角度来看这个问题. 若记 $\boldsymbol{T}=(\boldsymbol{x}_1,\boldsymbol{x}_2)$, $\boldsymbol{D}=\operatorname{diag}(\lambda_1,\lambda_2)$, 则 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{D}$ 当且仅当 $\boldsymbol{A}\boldsymbol{T}=\boldsymbol{T}\boldsymbol{D}$, 当且仅当 $\boldsymbol{A}(\boldsymbol{x}_1,\boldsymbol{x}_2)=(\lambda_1\boldsymbol{x}_1,\lambda_2\boldsymbol{x}_2)$. 从而, 这样的 $\boldsymbol{T}$ 与 $\boldsymbol{D}$ 若存在, 必满足如下的两条.

(1) 由于 $\boldsymbol{T}$ 可逆, 列向量 $\boldsymbol{x}_1,\boldsymbol{x}_2$ 线性无关. 特别地, 它们都是非零向量.

(2) 解矩阵方程组, 我们有 $\boldsymbol{A}\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i$ $(i=1,2)$.

由上面的讨论, 我们引入特征值与特征向量的概念.

特征值与特征向量的定义

定义 6.3.1. 设 $\boldsymbol{A}$ 为数域 F 上的 n 阶方阵. 如果存在 $\lambda\in F$ 和非零列向量 $\boldsymbol{x}\in F^n$ 使得 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$, 则称 λ 为 $\boldsymbol{A}$ 的**一个特征值** (*eigenvalue, characteristic value*), 并称 $\boldsymbol{x}$ 为**属于 λ 的一个特征向量** (*eigenvector, characteristic vector*).

注 6.3.2. (1) 若 $\boldsymbol{x}$ 是矩阵 $\boldsymbol{A}$ 关于 λ 的特征向量, 则 $\boldsymbol{A}\boldsymbol{x}$ 是与 $\boldsymbol{x}$ 平行的向量, λ 是拉伸的系数.

(2) 由前面例子的引入, 我们知道特征值是相似对角阵 (如果存在) 的主对角线上的元素.

(3) 特征值 λ 作为一个标量并不一定非零.

习题 6.3.3. (1) 设 $\boldsymbol{u}=(3,1)^{\mathrm{T}}$ 和 $\boldsymbol{v}=(1,2)^{\mathrm{T}}$ 是 $\mathbb{R}^2$ 中的向量, 并设 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 是一个 2 阶方阵 $\boldsymbol{A}$ 分别相应于特征值 2 和 3 的特征向量. 设 $\mathscr{A}:\mathbb{R}^2\to\mathbb{R}^2$ 是线性变换, 由 $\boldsymbol{x}\mapsto\boldsymbol{A}\boldsymbol{x}$ 给出. 记 $\boldsymbol{w}=\boldsymbol{u}+\boldsymbol{v}$. 画出 $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, $\mathscr{A}(\boldsymbol{u})$, $\mathscr{A}(\boldsymbol{v})$ 和 $\mathscr{A}(\boldsymbol{w})$.

(2) 设 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 是 $\boldsymbol{A}$ 分别对应于特征值 -1 和 3 的特征向量, 重新做上面的小题.

例 6.3.4. 在例 *6.2.19* 中, 由

$$\begin{pmatrix}1&1\\1&0\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt5}{2}&\frac{1-\sqrt5}{2}\\1&1\end{pmatrix}=\begin{pmatrix}\frac{1+\sqrt5}{2}&\frac{1-\sqrt5}{2}\\1&1\end{pmatrix}\begin{pmatrix}\frac{1+\sqrt5}{2}&\\&\frac{1-\sqrt5}{2}\end{pmatrix}$$

可知, $\boldsymbol{x}_1=\left(\frac{1+\sqrt5}{2},1\right)^{\mathrm{T}}$ 是方阵 $\begin{pmatrix}1&1\\1&0\end{pmatrix}$ 相对于特征值 $\lambda_1=\frac{1+\sqrt5}{2}$ 的一个特征向量, $\boldsymbol{x}_2=\left(\frac{1-\sqrt5}{2},1\right)^{\mathrm{T}}$ 是方阵 $\begin{pmatrix}1&1\\1&0\end{pmatrix}$ 相对于特征值 $\lambda_2=\frac{1-\sqrt5}{2}$ 的一个特征向量.

考察集合 $V_{\boldsymbol{A}}(\lambda) := \{\, \boldsymbol{x} \in F^n \mid \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \,\}$. 它是齐次方程组 $(\lambda \boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ 的解空间, 从而构成 F^n 的一个子空间. 由定义 6.3.5 知, λ 是 $\boldsymbol{A}$ 的一个特征值当且仅当 $V_{\boldsymbol{A}}(\lambda)$ 存在非零向量. 若该条件满足, 我们称 $V_{\boldsymbol{A}}(\lambda)$ 为**属于 λ 的特征子空间** (eigenspace, characteristic space). 一个向量是属于 λ 的特征向量当且仅当它是 $V_{\boldsymbol{A}}(\lambda)$ 中的一个非零向量.

类似地, 我们可以讨论线性变换的特征值与特征向量的问题.

定义 6.3.5. *设 F 为数域, V 为 F 上的线性空间, $\mathscr{A}$ 为 V 上的线性变换. 若存在 $\lambda \in F$ 以及非零向量 $\boldsymbol{x} \in V$, 满足 $\mathscr{A}(\boldsymbol{x}) = \lambda \boldsymbol{x}$, 则称 λ 为 $\mathscr{A}$ 的***一个特征值***, 并称 $\boldsymbol{x}$ 为***属于 λ 的一个特征向量***.

类似地, 我们可以考察集合 $V_{\mathscr{A}}(\lambda) := \{\, \boldsymbol{x} \in V \mid \mathscr{A}(\boldsymbol{x}) = \lambda \boldsymbol{x} \,\}$. 由于 $\mathscr{A}$ 为线性变换, $\boldsymbol{0} \in V_\lambda$, 从而该集合非空. 容易用定义验证, $V_{\mathscr{A}}(\lambda)$ 对于向量的加法与数乘运算封闭, 从而构成 V 的一个子空间. 由定义 6.3.5 知, λ 是 $\mathscr{A}$ 的一个特征值当且仅当 $V_{\mathscr{A}}(\lambda)$ 存在非零向量. 若该条件满足, 我们称 $V_{\mathscr{A}}(\lambda)$ 为**属于 λ 的特征子空间** (eigenspace, characteristic space). 一个向量是属于 λ 的特征向量当且仅当它是 $V_{\mathscr{A}}(\lambda)$ 中的一个非零向量.

例 6.3.6. *若 $\mathscr{A}$ 为恒等变换, 则 $\mathscr{A}$ 仅有特征值 1, 特征子空间 $V_{\mathscr{A}}(1) = V$ 为全空间, 即, V 中任意的非零向量都是属于 1 的特征向量.*

习题 6.3.7. *求矩阵 $\boldsymbol{A} = \begin{pmatrix} 4 & 2 & 3 & 3 \\ 0 & 2 & h & 3 \\ 0 & 0 & 4 & 14 \\ 0 & 0 & 0 & 2 \end{pmatrix}$ 中的 h, 使得 $\lambda = 5$ 的特征子空间的维数为 2.*

若 $\mathscr{A}$ 是数组空间 F^n 上的一个线性变换, 在自然基 $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ 下的矩阵为 $\boldsymbol{A}$, 则容易验证, $\boldsymbol{A}$ 与 $\mathscr{A}$ 的特征值和特征向量完全吻合. 关于这一点的更进一步的讨论, 我们有如下的命题.

命题 6.3.8. *若 $\mathscr{A}$ 是数组空间 V 上的一个线性变换, 在基 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}$. 那么, 我们有以下几条.* 教材命题 6.3.1

(1) *$\boldsymbol{A}$ 与 $\mathscr{A}$ 具有相同的特征值;*

(2) *设 λ 是 $\mathscr{A}$ 的一个特征值, 则 $\boldsymbol{x}$ 是 $\mathscr{A}$ 的关于 λ 的特征向量的充要条件是 $\boldsymbol{x}$ 在基 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 下的坐标列向量 $\boldsymbol{X}$ 是 $\boldsymbol{A}$ 关于 λ 的特征向量. 从而,*

$$V_{\mathscr{A}}(\lambda) = \{\, \boldsymbol{x} \in V \mid \boldsymbol{x} \text{ 在基 } \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n \text{ 下的坐标 } \boldsymbol{X} \text{ 属于 } V_{\boldsymbol{A}}(\lambda) \,\}.$$

证明. 设 $\lambda \in F$, $\boldsymbol{x} \in V$ 在 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 下的坐标为 $\boldsymbol{X}$, 即, $\boldsymbol{x} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n)\boldsymbol{X}$. 此时,

$$\begin{aligned}\mathscr{A}(\boldsymbol{x}) =& \mathscr{A}((\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n)\boldsymbol{X}) = \mathscr{A}(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n)\boldsymbol{X} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n)\boldsymbol{A}\boldsymbol{X},\\ \lambda\boldsymbol{x} =& \lambda((\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n)\boldsymbol{X}) = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n)\lambda\boldsymbol{X}.\end{aligned}$$

由坐标表示的唯一性可知 $\boldsymbol{A}\boldsymbol{X} = \lambda\boldsymbol{X}$ 的充要条件是 $\mathscr{A}(\boldsymbol{x}) = \lambda\boldsymbol{x}$. 另一方面, $\boldsymbol{x}$ 是零向量当且仅当 $\boldsymbol{X}$ 是零向量. 从而, 命题的各条容易验证. □

注 6.3.9. *线性变换的特征值与特征向量的求解问题, 一般都是利用命题 6.3.8, 化成在某组基下的矩阵的相应问题的求解. 关于这方面具体的例子, 见教材 P175 的例 6.3.6. 学生课后自习.*

习题 6.3.10. *在以下题目中, 设 $\boldsymbol{A}$ 是线性变换 $\mathscr{A}$ 在自然基下的矩阵. 不写出 $\boldsymbol{A}$, 直接描述 (求出) $\boldsymbol{A}$ 的特征值与特征子空间.*

(1) *$\mathscr{A}$ 是 $\mathbb{R}^2$ 的相对于某条过原点的直线的反射变换 (镜像映射).*

(2) *$\mathscr{A}$ 是 $\mathbb{R}^3$ 的相对于某条过原点的直线的旋转变换.*

特征值与特征向量的计算 设 $\boldsymbol{A} = (a_{ij})$ 为数域 F 上的一个 n 阶方阵, 则

$$\begin{aligned}\lambda \in F \text{ 为 } \boldsymbol{A} \text{ 的特征值} &\Leftrightarrow V_A(\lambda) \text{ 有非零向量} \Leftrightarrow (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \text{ 有非零解}\\ &\Leftrightarrow \lambda\boldsymbol{I} - \boldsymbol{A} \text{ 不可逆} \Leftrightarrow \text{行列式 } |\lambda\boldsymbol{I} - \boldsymbol{A}| = 0.\end{aligned}$$

我们可以观察到

$$\begin{aligned}|\lambda\boldsymbol{I} - \boldsymbol{A}| &= \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -a_{n-1,n} \\ -a_{n1} & \cdots & -a_{n,n-1} & \lambda - a_{nn} \end{vmatrix}\\ &= \lambda^n + \text{中间次数的复杂项} + (-1)^n \det(\boldsymbol{A}).\end{aligned}$$

最后的这个多项式是关于未知元 λ 的 F 上的 n 次多项式, 其中间次数的系数较为复杂, 我们稍后会作进一步解释. 该多项式将被记作 $p_{\boldsymbol{A}}(\lambda)$, 称为 $\boldsymbol{A}$ 的**特征多项式*** (characteristic polynomial). 相应地, 方程 $p_{\boldsymbol{A}}(\lambda) = 0$ 称为矩阵 $\boldsymbol{A}$ 的**特征方程** (characteristic equation). 具体取值 $\lambda_0 \in F$ 为 $\boldsymbol{A}$ 的特征值, 当且仅当 λ_0 是 $p_{\boldsymbol{A}}(\lambda) = 0$ 的根. 由域上的多项式的理论, 我们知 $\boldsymbol{A}$ 的特征值最多有 n 个. 特别地, $\boldsymbol{A}$ 的特征值的个数有限.

*有的教材和数学软件将 $|\boldsymbol{A} - \lambda\boldsymbol{I}|$ 称为特征多项式. 这个变动不是本质的.

注 6.3.11. 类似的讨论可以说明, $|\lambda \boldsymbol{I}+\boldsymbol{A}|$ 也是关于未知元 λ 的 F 上的 n 次多项式, 从而仅有有限多个 $\lambda \in F$ 使得 $\lambda \boldsymbol{I}+\boldsymbol{A}$ 不可逆. 这是 "微小摄动法" 的理论前提.

注 6.3.12. 我们观察到多项式 λ^2+1 在 $\mathbb{R}$ 上无根, 在 $\mathbb{C}$ 上有一对不同的根. 故, 考虑特征值时, 我们更加倾向于在 $\mathbb{C}$ 上考察.

问题: 如何在 $\mathbb{C}$ 上求一个 n 阶方阵 $\boldsymbol{A}$ 的特征值和相应的特征向量?

(1) 首先, 计算矩阵 $\boldsymbol{A}$ 的特征多项式 $p_{\boldsymbol{A}}(\lambda)$. 由**代数学基本定理**可知, $p_{\boldsymbol{A}}(\lambda)$ 恰好有 n 个复根 (可能有重根), 从而可以写成

$$p_{\boldsymbol{A}}(\lambda)=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\cdots(\lambda-\lambda_s)^{n_s},$$

其中 $\lambda_1,\ldots,\lambda_s \in \mathbb{C}$ 是 $\boldsymbol{A}$ 的全部不同的特征值, 相应的重数 $n_i \geq 1$ 被称为 λ_i 的**代数重数**, 满足

$$n_1+\cdots+n_s=n.$$

(2) 对于每个特征值 λ_i, 求解齐次方程组 $(\lambda_i \boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$. 相应的解空间 $V_{\boldsymbol{A}}(\lambda_i)$ 不是零空间, 不妨设 $\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},\ldots,\boldsymbol{x}_{im_i}$ 是一个基础解系 (m_i 被称为 λ_i 的**几何重数**). 则 $\boldsymbol{A}$ 的属于 λ_i 的所有特征向量为 $V_{\boldsymbol{A}}(\lambda_i)$ 中的所有非零向量, 即 $\boldsymbol{x}_{i1},\ldots,\boldsymbol{x}_{im_i}$ 的非零线性组合的全体.

例 6.3.13. 求矩阵 $\boldsymbol{A}=\begin{pmatrix}1&2&2\\2&1&2\\2&2&1\end{pmatrix}$ 的特征值及特征向量.

证明. (1) 求解特征多项式

$$|\lambda \boldsymbol{I}-\boldsymbol{A}|=\begin{vmatrix}\lambda-1&-2&-2\\-2&\lambda-1&-2\\-2&-2&\lambda-1\end{vmatrix}=\lambda^3-3\lambda^2-9\lambda-5=(\lambda-5)(\lambda+1)^2.$$

正确计算简单的三阶方阵的特征多项式, 并将其分解因式, 是非常重要的

故矩阵 $\boldsymbol{A}$ 的特征值为 $\lambda_1=5$ 和 $\lambda_2=-1$ (两重).

(2) 对于 $\lambda_1=5$,

$$\lambda_1 \boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}4&-2&-2\\-2&4&-2\\-2&-2&4\end{pmatrix}\xrightarrow{\text{初等行变换}}\begin{pmatrix}1&0&-1\\0&1&-1\\0&0&0\end{pmatrix}.$$

解相应的方程组, 我们可以以 x_3 为自由元, 解得 $x_1=x_2=x_3$. 这说明解空间有基础解系 $(1,1,1)^{\mathrm{T}}$.

(3) 对于 $\lambda_2=-1$,

$$\lambda_2-\boldsymbol{A}=\begin{pmatrix}-2&-2&-2\\-2&-2&-2\\-2&-2&-2\end{pmatrix}\xrightarrow{\text{初等行变换}}\begin{pmatrix}1&1&1\\0&0&0\\0&0&0\end{pmatrix}.$$

我们以 x_2,x_3 为自由元, 解得 $x_1=-x_2-x_3$. 即

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}-x_2-x_3\\x_2\\x_3\end{pmatrix}=x_2\begin{pmatrix}-1\\1\\0\end{pmatrix}+x_3\begin{pmatrix}-1\\0\\1\end{pmatrix}.$$

从而解空间有基础解系 $(-1,1,0)^{\mathrm{T}},(-1,0,1)^{\mathrm{T}}$.

综上, $\boldsymbol{A}$ 有两个特征值 $\lambda_1=5$ 和 $\lambda_2=-1$. 属于 $\lambda_1=5$ 的特征向量为 $c_1(1,1,1)^{\mathrm{T}}$, 其中 $c_1\neq 0$. 属于 $\lambda_2=-1$ 的特征向量为 $c_2(-1,1,0)^{\mathrm{T}}+c_3(-1,0,1)^{\mathrm{T}}$, 其中 c_2,c_3 不全为 0. □

例 6.3.14. 设 3 阶矩阵 $\boldsymbol{A}$ 的 3 个特征值为 $1,1,2$, 对应地分别有特征向量

$$\boldsymbol{\xi}_1=\begin{pmatrix}1\\2\\1\end{pmatrix},\quad \boldsymbol{\xi}_2=\begin{pmatrix}1\\1\\0\end{pmatrix},\quad \boldsymbol{\xi}_3=\begin{pmatrix}2\\0\\-1\end{pmatrix}.$$

求矩阵 $\boldsymbol{A}$.

解. 由定义, $\boldsymbol{A}\boldsymbol{\xi}_1=\boldsymbol{\xi}_1$, $\boldsymbol{A}\boldsymbol{\xi}_2=\boldsymbol{\xi}_2$, $\boldsymbol{A}\boldsymbol{\xi}_3=2\boldsymbol{\xi}_3$. 将这些由矩阵表示, 我们得到

$$\boldsymbol{A}\underbrace{(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\boldsymbol{\xi}_3)}_{\boldsymbol{T}}=(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\boldsymbol{\xi}_3)\underbrace{\begin{pmatrix}1&&\\&1&\\&&2\end{pmatrix}}_{\boldsymbol{D}},$$

可以验证, 我们有 $\det(\boldsymbol{T})\neq 0$, 这说明 $\boldsymbol{T}$ 可逆, 从而有

$$\begin{aligned}\boldsymbol{A}=&\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^{-1}\\=&\begin{pmatrix}1&1&2\\2&1&0\\1&0&-1\end{pmatrix}\begin{pmatrix}1&&\\&1&\\&&2\end{pmatrix}\begin{pmatrix}1&1&2\\2&1&0\\1&0&-1\end{pmatrix}^{-1}\\=&\begin{pmatrix}1&1&2\\2&1&0\\1&0&-1\end{pmatrix}\begin{pmatrix}1&&\\&1&\\&&2\end{pmatrix}\begin{pmatrix}1&-1&2\\-2&3&-4\\1&-1&1\end{pmatrix}\end{aligned}$$

$$= \begin{pmatrix} 3 & -2 & 2 \\ 0 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix}. \qquad \square$$

例 6.3.15. 设 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1s} \\ \boldsymbol{O} & \ddots & \vdots \\ \vdots & \ddots & \\ \boldsymbol{O} & \cdots\ \boldsymbol{O} & \boldsymbol{A}_{ss} \end{pmatrix}$ 为准上三角的分块矩阵, 其中主对角线上的子块 $\boldsymbol{A}_{11}, \boldsymbol{A}_{22}, \ldots, \boldsymbol{A}_{ss}$ 都是方阵, 那么由定义不难验证, 特征多项式

$$p_{\boldsymbol{A}}(\lambda) = p_{\boldsymbol{A}_{11}}(\lambda) p_{\boldsymbol{A}_{22}}(\lambda) \cdots p_{\boldsymbol{A}_{ss}}(\lambda).$$

特别地, 若 $\boldsymbol{A} = (a_{ij})$ 是上三角的方阵, 则 $p_{\boldsymbol{A}}(\lambda) = (\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn})$. 当然, 若 $\boldsymbol{A}$ 是 (准) 下三角的方阵时, 讨论也是类似的.

注 6.3.16. 由多项式的理论可知，实多项式的虚根都是成对出现的. 这说明实方阵的虚复特征值也是成对出现的.

事实上，若 λ 是某个 n 阶实方阵 $\boldsymbol{A}$ 的复特征值，$\boldsymbol{x} \in \mathbb{C}^n$ 是对应的一个特征向量，则

$$\boldsymbol{A}\overline{\boldsymbol{x}} = \overline{\boldsymbol{A}}\overline{\boldsymbol{x}} = \overline{\boldsymbol{A}\boldsymbol{x}} = \overline{\lambda \boldsymbol{x}} = \overline{\lambda}\overline{\boldsymbol{x}}.$$

由于 $\overline{\boldsymbol{x}} \neq \boldsymbol{0}$，这说明 $\overline{\boldsymbol{x}}$ 是实矩阵 $\boldsymbol{A}$ 关于特征值 $\overline{\lambda}$ 的特征向量. 特别地，这说明实方阵的复特征向量也是成对出现的. (当然，这儿需要简要说明一下 $\boldsymbol{x} \neq \overline{\boldsymbol{x}}$，从而 $\boldsymbol{x}$ 不是实向量. 若否，则我们会得到

$$\overline{\lambda}\boldsymbol{x} = \overline{\lambda}\overline{\boldsymbol{x}} = \overline{\boldsymbol{A}\boldsymbol{x}} \xlongequal[\text{从而 } \boldsymbol{A}\boldsymbol{x} \text{ 也是实的}]{\boldsymbol{A}, \boldsymbol{x} \text{ 都是实的}} \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x},$$

这说明 $(\lambda - \overline{\lambda})\boldsymbol{x} = \boldsymbol{0}$. 但是 $\lambda \neq \overline{\lambda}$，从而可以推出 $\boldsymbol{x} = \boldsymbol{0}$. 而这与 $\boldsymbol{x}$ 为特征值相矛盾.)

另一方面，如果实矩阵 $\boldsymbol{A}$ 的特征值是实数，那么为了讨论的方便，作为解空间的基本解系的特征向量我们一般可以取为实向量.

若 $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}} \in \mathbb{C}^n$，则

$$\|\boldsymbol{x}\| := \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} = \sqrt{\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}}$$

称为 $\boldsymbol{x}$ 在标准内积下的**模长**. 显然，$\boldsymbol{x} = \boldsymbol{0}$ 当且仅当 $\|\boldsymbol{x}\| = 0$.

例 6.3.17. 设 λ 为 n 阶方阵 $\boldsymbol{A}$ 的一个特征值，$\boldsymbol{x}$ 为对应的一个特征向量，则 教材例题 6.3.3

(1) λ^k 是 $\boldsymbol{A}^k$ 的特征值，其中 k 为正整数，并且更一般地，若 $f(t)$ 是一个一元多项式，则 $f(\lambda)$ 是 $f(\boldsymbol{A})$ 的特征值;

(2) λ 为 $\boldsymbol{A}^{\mathrm{T}}$ 的特征值;

(3) 若 $\lambda \neq 0$，则 $\frac{\det(\boldsymbol{A})}{\lambda}$ 为 $\boldsymbol{A}$ 的伴随矩阵 $\boldsymbol{A}^*$ 的特征值;

(4) 若 $\boldsymbol{A}$ 为实方阵且满足 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$ (即 $\boldsymbol{A}$ 为正交矩阵)，则 $|\lambda| = 1$. (此时 $\lambda \in \mathbb{C}$)

证明. (1) 可以推出，

$$\boldsymbol{A}^k\boldsymbol{x} = \boldsymbol{A}^{k-1}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}^{k-1}(\lambda\boldsymbol{x}) = \lambda\boldsymbol{A}^{k-1}\boldsymbol{x} = \cdots = \lambda^k\boldsymbol{x},$$

以及 $f(\boldsymbol{A})\boldsymbol{x} = f(\lambda)\boldsymbol{x}$. 从而 $\boldsymbol{x}$ 是 $\boldsymbol{A}^k$ 关于特征值 λ^k 的一个特征向量，也是 $f(\boldsymbol{A})$ 关于特征值 $f(\lambda)$ 的一个特征值.

(2) 此时 $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=0$, 但是转置运算不改变方阵的行列式, 从而 $\det(\lambda\boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}})=0$. 这说明 λ 也是 $\boldsymbol{A}^{\mathrm{T}}$ 的特征值.

(3) 我们有 $\boldsymbol{A}^*\boldsymbol{A}=\det(\boldsymbol{A})\boldsymbol{I}$, 从而可以推出 $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{x}=\det(\boldsymbol{A})\boldsymbol{x}$. 可是 $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^*(\lambda\boldsymbol{x})=\lambda\boldsymbol{A}^*\boldsymbol{x}$, 这说明 $\boldsymbol{A}^*\boldsymbol{x}=\frac{\det(\boldsymbol{A})}{\lambda}\boldsymbol{x}$, 从而 $\boldsymbol{x}$ 是 $\boldsymbol{A}^*$ 的属于 $\frac{\det(\boldsymbol{A})}{\lambda}$ 的特征向量.

(4) 由于 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$, 取转置运算和复共轭后, 我们有 $\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}=\overline{\lambda}\overline{\boldsymbol{x}}^{\mathrm{T}}$, 再分别右乘等式 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$ 的两边, 我们有 $\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\lambda\overline{\lambda}\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}$, 即 $\|\boldsymbol{x}\|^2=|\lambda|^2\cdot\|\boldsymbol{x}\|^2$. 由于 $0\neq\|\boldsymbol{x}\|\in\mathbb{C}$, 我们推出 $|\lambda|=1$. □

注 6.3.18. (1) 设 n 阶方阵 $\boldsymbol{A}$ 的每行元素之和都等于 s. 容易看出, 对于全为 1 的列向量 $\boldsymbol{x}=(1,1,\dots,1)^{\mathrm{T}}$ 有 $\boldsymbol{A}\boldsymbol{x}=s\boldsymbol{x}$. 从而, $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的关于特征值 s 的特征向量.

(2) 设 n 阶方阵 $\boldsymbol{B}$ 的每列元素之和都等于 t. 由于转置运算不改变矩阵的特征值, 利用上面一条可以看出, t 是 $\boldsymbol{B}$ 的特征值. 不过, 此时的特征向量就不一定容易描述了.

命题 6.3.19. *若 $\boldsymbol{A}=\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$, 则 $p_{\boldsymbol{A}}(\lambda)=p_{\boldsymbol{B}}(\lambda)$, 即它们的特征多项式相等. 特别地, 相似的矩阵的特征值 (重根按重数计算) 相同.*

证明. 直接计算, 有

$$p_{\boldsymbol{A}}(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\det(\lambda\boldsymbol{I}-\boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T})=\det(\boldsymbol{T}^{-1}(\lambda\boldsymbol{I}-\boldsymbol{B})\boldsymbol{T})=\det(\lambda\boldsymbol{I}-\boldsymbol{B})=p_{\boldsymbol{B}}(\lambda).$$

□

注 6.3.20. 设 $\lambda_1,\dots,\lambda_n$ 是 n 阶方阵 $\boldsymbol{A}$ 的所有特征值, $f(t)$ 为一元多项式, 则 $f(\lambda_1),\dots,f(\lambda_n)$ 皆为 $f(\boldsymbol{A})$ 的特征值. 由于可能出现 $i\neq j$ 但是 $f(\lambda_i)=f(\lambda_j)$ 的情形, 目前这并没有表明 $f(\lambda_1),\dots,f(\lambda_n)$ 为 $f(\boldsymbol{A})$ 全部的特征值. 为此, 我们利用下一节要提到的 *Schur* 定理. 不妨设所考虑的矩阵为复方阵, 由该定理可知, 存在可逆方阵 $\boldsymbol{T}$ 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 为上三角矩阵, 其对角线上的元素依次为 $\lambda_1,\dots,\lambda_n$. 可以直接验证, $\boldsymbol{T}^{-1}f(\boldsymbol{A})\boldsymbol{T}$ 仍然是上三角矩阵, 其对角线上的元素依次为 $f(\lambda_1),\dots,f(\lambda_n)$. 这足以说明, $f(\lambda_1),\dots,f(\lambda_n)$ 为 $f(\boldsymbol{A})$ 全部的特征值.

命题 6.3.21. *设 $\boldsymbol{A},\boldsymbol{B}$ 为同阶方阵, 则 $\boldsymbol{A}\boldsymbol{B}$ 和 $\boldsymbol{B}\boldsymbol{A}$ 具有相同的特征多项式.*

证明. 我们只需说明 $\det(\lambda\boldsymbol{I}-\boldsymbol{A}\boldsymbol{B})=\det(\lambda\boldsymbol{I}-\boldsymbol{B}\boldsymbol{A})$. 当 $\lambda=0$ 时, 这等价于证明 $\det(-\boldsymbol{A}\boldsymbol{B})=\det(-\boldsymbol{B}\boldsymbol{A})$, 而这是显然的. 当 $\lambda\neq 0$ 时, 我们将教材第四章习题 #25 用于 $\frac{1}{\lambda}\boldsymbol{A}$ 和 $\boldsymbol{B}$, 从而有 $\det(\boldsymbol{I}-\frac{1}{\lambda}\boldsymbol{A}\boldsymbol{B})=\det(\boldsymbol{I}-\frac{1}{\lambda}\boldsymbol{B}\boldsymbol{A})$. 对其稍加变形即可. □

我们下面开始具体研究复方阵的特征多项式. 设 $\boldsymbol{A}=(a_{ij})\in\mathbb{C}^{n\times n}$, 则

$$p_{\boldsymbol{A}}(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\begin{vmatrix}\lambda-a_{11} & -a_{12} & \cdots & -a_{1n}\\ -a_{21} & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -a_{n-1,n}\\ -a_{n1} & \cdots & -a_{n,n-1} & \lambda-a_{nn}\end{vmatrix}.$$

很明显, $p_{\boldsymbol{A}}(\lambda)$ 是关于 λ 的多项式. 由于 λ 仅出现在主对角线上, 不难看出 $p_{\boldsymbol{A}}(\lambda)$ 是关于 λ 的最高次系数为 1 的 n 次多项式, 从而

$$\begin{aligned}p_{\boldsymbol{A}}(\lambda)&\xlongequal{\text{记作}}\lambda^n+\sigma_1\lambda^{n-1}+\cdots+\sigma_{n-1}\lambda+\sigma_n\\ &=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n).\end{aligned}$$

其中 $\lambda_1,\ldots,\lambda_n$ 为 $p_{\boldsymbol{A}}(\lambda)$ 的根, 即 $\boldsymbol{A}$ 的特征值. 由**韦达定理 (多项式系数与根的关系)** 可知,

$$\sigma_i=(-1)^i\sum_{1\le j_1<\cdots<j_i\le n}\lambda_{j_1}\lambda_{j_2}\cdots\lambda_{j_i},$$

例如, 当 $n=4$ 而 $i=2$, 则

$$\sigma_2=(-1)^2(\lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_1\lambda_4+\lambda_2\lambda_3+\lambda_2\lambda_4+\lambda_3\lambda_4).$$

需要特别关注的是,

$$\sigma_1=-\sum_{i=1}^n\lambda_i,\qquad \sigma_n=(-1)^n\lambda_1\cdots\lambda_n.$$

另一方面, 可以看出

$$\sigma_n=p_{\boldsymbol{A}}(0)=\det(0\boldsymbol{I}-\boldsymbol{A})=\det(-\boldsymbol{A})=(-1)^n\det(\boldsymbol{A}).$$

这说明

$$\det(\boldsymbol{A})=\lambda_1\cdots\lambda_n. \tag{6.2}$$

另外, λ 仅出现在 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的主对角线上. 由行列式 $|\lambda\boldsymbol{I}-\boldsymbol{A}|$ 的完全展开公式中, 通项里若有 $n-1$ 个元素出现在主对角线上, 则剩下的第 n 个元素也出现在主对角线上. 这说明, λ^{n-1} 仅在唯一的展开项 $(\lambda-a_{11})(\lambda-a_{22})\cdots(\lambda-a_{nn})$ 中出现. 不难看出, 其中 λ^{n-1} 的系数为 $\sigma_1=-\sum_{i=1}^n a_{ii}$ (这儿可以直接展开看出, 或者再次用一下韦达定理). 这说明 回忆这一公式

$$\operatorname{tr}(\boldsymbol{A})=\sum_{i=1}^n a_{ii}=\sum_{i=1}^n\lambda_i. \tag{6.3}$$

注 6.3.22. 对于给定 n 阶矩阵 $\boldsymbol{A}$, 形如 $\left|\boldsymbol{A}\binom{i_1\ i_2\ \cdots\ i_k}{i_1\ i_2\ \cdots\ i_k}\right|$ 的 k 阶子式称为行列式 $|\boldsymbol{A}|$ 的 k 阶**主子式**, 其中 $1 \le i_1 < i_2 < \cdots < i_k \le n$. 可以证明, 上面的 $(-1)^k\sigma_k$ 是所有 k 阶主子式的和.

习题 6.3.23. 设 $\boldsymbol{A} = (a_{ij})_{3\times 3}$. 若 $\boldsymbol{A}$ 的特征多项式为 $|\lambda \boldsymbol{I}_3 - \boldsymbol{A}| = \lambda^3 + \sigma_1\lambda^2 + \sigma_2\lambda + \sigma_3$, 验证 σ_2 是 $\boldsymbol{A}$ 的所有 2 阶主子式的和.

由公式 (6.2) 出发, 我们立刻得到

推论 6.3.24. n 阶方阵可逆当且仅当零不是它的特征值.

注 6.3.25. 由于相似的方阵具有相同的特征多项式, 方阵的迹和行列式作为特征多项式的系数都是相似不变量. 当然, 我们在之前的讨论中已经直接看出了这一点.

例 6.3.26. 学生课堂上自学教材 *P174-175* 的例 *6.3.4, 6.3.5.*

例 6.3.27. 设 $\boldsymbol{A}$ 为 3 阶方阵, 有特征值 $1, 2, 3$. 设 $\boldsymbol{B} = \boldsymbol{A}^3 + 2\boldsymbol{A}^2 + \boldsymbol{A} + 2\boldsymbol{I}$. 若以 $f(x) = x^3 + 2x^2 + x + 2$, 则 $\boldsymbol{B} = f(\boldsymbol{A})$, 从而以 $f(1) = 6$, $f(2) = 20$ 和 $f(3) = 50$ 为特征值. 特别地, $|\boldsymbol{B}| = 6 \cdot 20 \cdot 50 = 6000$, $\operatorname{tr}(\boldsymbol{B}) = 6 + 20 + 50 = 76$.

习题 6.3.28. 证明*:* 若 $\boldsymbol{A}^{2021} = \boldsymbol{O}$, 则 $\det(\boldsymbol{I} - 1958\boldsymbol{A}) = 1$.

例 6.3.29. 若 $\boldsymbol{A} = (a_{ij})_{4\times 4}$, 以 1 为二重特征值, 以 -2 为一重特征值, 求 $\boldsymbol{A}$ 的特征多项式.

解. 设 $\lambda_{1,2} = 1$, $\lambda_3 = -2$, 则 $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = \operatorname{tr}(\boldsymbol{A}) = a_{11} + a_{22} + a_{33} + a_{44}$. 故, $\lambda_4 = a_{11} + a_{22} + a_{33} + a_{44}$. 这说明

$$p_{\boldsymbol{A}}(\lambda) = (\lambda - 1)^2(\lambda + 2)(\lambda - (a_{11} + a_{22} + a_{33} + a_{44})). \qquad \square$$

定义 6.3.30. 在数域 F 上的有限维线性空间 V 上的线性变换 $\mathscr{A}$, 在 V 的一组基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下有矩阵 $\boldsymbol{A}$. 在命题 *6.3.8* 中我们已经见到了 $\boldsymbol{A}$ 的特征向量与 $\mathscr{A}$ 的特征向量之间的关系. 由此出发, 我们定义如下*:*

$$\begin{aligned} \mathscr{A}\ \text{的}\textbf{特征多项式}\ p_{\mathscr{A}}(\lambda) &\coloneqq p_{\boldsymbol{A}}(\lambda), \\ \mathscr{A}\ \text{的}\textbf{秩}\ \operatorname{rank}(\mathscr{A}) &\coloneqq \operatorname{rank}(\boldsymbol{A}), \\ \mathscr{A}\ \text{的}\textbf{行列式}\ \det(\mathscr{A}) &\coloneqq \det(\boldsymbol{A}), \\ \mathscr{A}\ \text{的}\textbf{迹}\ \operatorname{tr}(\mathscr{A}) &\coloneqq \operatorname{tr}(\boldsymbol{A}). \end{aligned}$$

由于特征多项式、秩、行列式、迹是方阵的相似不变量, 故上面的定义不依赖于特定的基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ (从而 $\boldsymbol{A}$) 的选取.

请先自行认真阅读教材上 P175 的例 6.3.6, 然后完成下面习题.

习题 6.3.31. 设 $\boldsymbol{A}=\begin{pmatrix}1 & 2\\ 0 & 4\end{pmatrix}$, $V=F^{2\times 2}$ 为线性空间, 定义 V 上的线性变换 $\mathscr{A}:\boldsymbol{M}\mapsto \boldsymbol{AM}$, $\boldsymbol{M}\in V$.

(1) 求 $\mathscr{A}$ 的特征值和特征向量.

(2) 求 $\mathscr{A}$ 在基 $\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$, $\begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}$, $\begin{pmatrix}1 & -1\\ 1 & -1\end{pmatrix}$, $\begin{pmatrix}1 & 1\\ -1 & -1\end{pmatrix}$ 下的矩阵.

定义 6.3.32. 对于线性空间 V 上的一个线性变换 $\mathscr{A}$, 若存在 V 的一组基, 使得 $\mathscr{A}$ 在这组基下的矩阵为对角阵, 那么我们称 $\mathscr{A}$ 为**可对角化的**.

注 6.3.33. 若 $\mathscr{A}$ 是数组空间 F^n 上的一个线性变换, 在自然基 $\boldsymbol{e}_1,\dots,\boldsymbol{e}_n$ 下的矩阵为 $\boldsymbol{A}$, 则由定理 *6.2.7* 容易验证, $\boldsymbol{A}$ 可对角化当且仅当 $\mathscr{A}$ 可对角化.

注 6.3.34. 对于相似的矩阵, 虽然其特征值相同, 对应的特征向量还是会改变的. 事实上, 若 $\boldsymbol{B}=\boldsymbol{T}^{-1}\boldsymbol{AT}$, 则 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$ 当且仅当 $\boldsymbol{TBT}^{-1}\boldsymbol{x}=\lambda\boldsymbol{x}$, 当且仅当 $\boldsymbol{B}(\boldsymbol{T}^{-1}\boldsymbol{x})=\lambda(\boldsymbol{T}^{-1}\boldsymbol{x})$. 这说明 $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的属于 λ 的特征向量, 当且仅当 $\boldsymbol{T}^{-1}\boldsymbol{x}$ 是 $\boldsymbol{B}$ 的属于 λ 的特征向量.

6.4 矩阵的相似对角化

首先我们需要指出, 并不是所有的方阵都可以相似对角化.

例 6.4.1. *考察矩阵* $\boldsymbol{A}=\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}$. *由于该矩阵为上三角矩阵, 在例 6.3.15 中我们已经看到,* $\boldsymbol{A}$ *的特征值为* $2,2,2$, *故若* $\boldsymbol{B}$ *与* $\boldsymbol{A}$ *相似, 则* $\boldsymbol{B}$ *的特征值也同样地为* $2,2,2$. *注意:* $\boldsymbol{B}$ *不能为对角阵, 否则* $\boldsymbol{B}$ *必为* $2\boldsymbol{I}_3$. *但是* $2\boldsymbol{I}_3$ *的相似等价类中仅有一个元素* $2\boldsymbol{I}_3$, *并不包含* $\boldsymbol{A}$. *故* $\boldsymbol{A}$ *与* $2\boldsymbol{I}_3$ *并不相似等价, 即,* $\boldsymbol{A}$ *不可对角化.*

接下来我们研究矩阵可以相似对角化的条件.

矩阵可对角化的充要条件

定理 6.4.2. n *阶方阵* $\boldsymbol{A}$ *相似于对角阵的充要条件是* $\boldsymbol{A}$ *有* n *个线性无关的特征向量.*

证明. (1) 设 $\boldsymbol{A}$ 可以对角化, 即存在可逆矩阵 $\boldsymbol{T}$ 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{B}$ 为对角阵. 设 $\boldsymbol{B}=\operatorname{diag}(\lambda_1,\ldots,\lambda_n)$, 而 $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ 为 $\boldsymbol{T}$ 的各个列向量. 此时 $\boldsymbol{A}\boldsymbol{T}=\boldsymbol{T}\boldsymbol{B}$, 即有

$$\boldsymbol{A}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\begin{pmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_n\end{pmatrix}.$$

这说明 $\boldsymbol{A}\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i$, $i=1,2,\ldots,n$. 由于 $\boldsymbol{T}$ 可逆, 向量组 $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ 线性无关. 特别地, 每个 $\boldsymbol{x}_i$ 都非零, 从而是 $\boldsymbol{A}$ 的属于特征值 λ_i 的特征向量.

(2) 反过来, 设 $\boldsymbol{A}\boldsymbol{x}_i=\lambda_i\boldsymbol{x}_i$, 满足 $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ 线性无关. 此时, 用矩阵表示即有

$$\boldsymbol{A}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\begin{pmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_n\end{pmatrix}.$$

将矩阵 $(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$ 记作 $\boldsymbol{T}$. 由于 $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ 线性无关, 方阵 $\boldsymbol{T}$ 列满秩, 从而可逆. 此时, 对于上式左乘矩阵 $\boldsymbol{T}^{-1}$, 即有 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\operatorname{diag}(\lambda_1,\ldots,\lambda_n)$. 这说明方阵 $\boldsymbol{A}$ 可以对角化. □

注 6.4.3. (1) 前面的证明说明了, 若方阵 $\boldsymbol{A}$ 可对角化, 则将其对角化的矩阵 $\boldsymbol{T}$ 的列向量为 $\boldsymbol{A}$ 的特征向量, 而对角化后的对角阵上的主对角线上的元素为 $\boldsymbol{A}$ 的特征值.

(2) 前面的定理说明, 方阵 $\boldsymbol{A}$ 可对角化的充要条件是 $\boldsymbol{A}$ 有足够多的特征向量来形成 F^n 的基. 我们不妨称这样的基为**特征向量基**.

(3) 用来对角化的矩阵 $\boldsymbol{T}$ 并不是唯一的. 例如, 将 $\boldsymbol{T}$ 的各列重新排列, 或者将它们分别乘以不同的非零标量, 这样得到的新矩阵仍然可以将 $\boldsymbol{A}$ 对角化.

接下来讨论矩阵特征向量之间的线性相关性.

命题 6.4.4. 设 $\boldsymbol{A}$ 是数域 F 上的 n 阶方阵, $\lambda_1, \ldots, \lambda_m$ 为 $\boldsymbol{A}$ 的两两互不相等的特征值, $\{\, \boldsymbol{x}_{i,j} \mid 1 \le j \le k_i \,\}$ 为属于 λ_i 的线性无关的特征向量, 则 $\{\, \boldsymbol{x}_{i,j} \mid 1 \le i \le m, 1 \le j \le k_i \,\}$ 线性无关.

证明. 设 $\mu_{i,j} \in F$ 使得

$$\sum_{i=1}^{m} \underbrace{\sum_{j=1}^{k_i} \mu_{i,j} \boldsymbol{x}_{i,j}}_{\text{记作 } \boldsymbol{x}_i} = \boldsymbol{0},$$

从而有 $\sum_i \boldsymbol{x}_i = \boldsymbol{0}$. 由于 $\boldsymbol{A}\boldsymbol{x}_{i,j} = \lambda_i \boldsymbol{x}_{i,j}$, 容易验证, $\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$. 对于 $k = 1, 2, \ldots, m-1$, 左乘 $\boldsymbol{A}^k$, 我们有

$$\boldsymbol{0} = \boldsymbol{A}^k \sum_{i=1}^{m} \boldsymbol{x}_i = \sum_{i=1}^{m} \boldsymbol{A}^k \boldsymbol{x}_i = \sum_{i=1}^{m} \lambda_i^k \boldsymbol{x}_i.$$

写成矩阵形式后, 我们有

$$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \begin{pmatrix} \lambda_1^k \\ \vdots \\ \lambda_m^k \end{pmatrix} = \boldsymbol{0},$$

从而进一步有

$$(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m) \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{m-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{m-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_m & \lambda_m^2 & \cdots & \lambda_m^{m-1} \end{pmatrix} = \boldsymbol{O}.$$

由于 $\lambda_1, \ldots, \lambda_m$ 互不相等, 其中的 Vandermonde 行列式对应的矩阵可逆, 这说明

$$(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m) = \boldsymbol{O}.$$

从而每个 $\boldsymbol{x}_i$ 都是零向量, 即 $\sum_j \mu_{ij}\boldsymbol{x}_{ij} = \boldsymbol{0}$. 对于这个 i, 由于向量组 $\{\, \boldsymbol{x}_{i,j} \mid 1 \le j \le k_i \,\}$ 线性无关, 故每个 $\mu_{i,j} = 0$. 这说明 $\{\, \boldsymbol{x}_{i,j} \mid 1 \le i \le m, 1 \le j \le k_i \,\}$ 线性无关. □

推论 6.4.5. *设 $\boldsymbol{A}$ 是数域 F 上的 n 阶方阵, $\lambda_1, \ldots, \lambda_m$ 为 $\boldsymbol{A}$ 的两两互不相等的特征值, $\boldsymbol{x}_i$ 为属于 λ_i 的特征向量, 则 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ 线性无关.*

证明. 对于每个 i $(1 \le i \le m)$, 由一个非零向量 $\boldsymbol{x}_i$ 构成的向量组显然是线性无关的, 从而可以使用命题 6.4.4 中的结果. □

例 6.4.6. *在例 6.3.13 中, 对于矩阵 $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$, 我们求出了对于特征值 $\lambda_1 = 5$ 的一个特征向量 $\boldsymbol{x}_1 = (1,1,1)^{\mathrm{T}}$, 和对于特征值 $\lambda_2 = \lambda_3 = -1$ 的两个线性无关的特征向量 $\boldsymbol{x}_2 = (-1,1,0)^{\mathrm{T}}$ 和 $\boldsymbol{x}_3 = (-1,0,1)^{\mathrm{T}}$. 由命题 6.4.4 可知, $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ 线性无关, 从而矩阵 $\boldsymbol{T} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ 是可逆矩阵. 由前面的讨论可知, $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3) = \operatorname{diag}(5, -1, -1)$.*

习题 6.4.7. *设 $\mathscr{A}$ 是 $\mathbb{R}^2$ 上的线性变换, 由 $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$ 给出, 其中 $\boldsymbol{A} = \begin{pmatrix} 7 & 2 \\ -4 & 1 \end{pmatrix}$. 求 $\mathbb{R}^2$ 的一组基, 使得 $\mathscr{A}$ 在这组基下的矩阵是对角阵.*

另外, 如下的推论给出了一个常用的矩阵可对角化的充分条件.

推论 6.4.8. *如果 n 阶方阵 $\boldsymbol{A}$ 的 n 个特征值两两不同, 则 $\boldsymbol{A}$ 可以对角化.*

证明. 对于 $\boldsymbol{A}$ 的每个特征值 λ_i, $i = 1, 2, \ldots, n$, 由定义知至少存在一个属于 λ_i 的特征向量 $\boldsymbol{x}_i$. 推论 6.4.5 保证了 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 是线性无关的. 再利用定理 6.4.2, 由此可以推出 $\boldsymbol{A}$ 可以对角化. □

注 6.4.9. *若 $\boldsymbol{A}$ 为 n 阶对角阵, 其对角线上的元素各不相等. 那么我们可以证明: 同阶方阵 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 乘法可交换的充要条件是 $\boldsymbol{B}$ 为对角阵. 教材习题 P192#22 给出了与之等价的形式, 留作课后作业.*

例 6.4.10. *设 $\boldsymbol{A}$ 为数域 F 上的 n 阶方阵, 有 n 个互不相等的特征值 $\lambda_1, \ldots, \lambda_n$, 对应分别有特征向量 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. 记 $\boldsymbol{T} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, 则 $\boldsymbol{T}$ 是一个可逆方阵, 并且 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \xlongequal{\text{记作}} \boldsymbol{B}$.*

容易验证 $V = \{\, \boldsymbol{C} \in F^{n\times n} \mid \boldsymbol{C}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{C} \,\}$ 是 $F^{n\times n}$ 的一个子空间. 同样地, 易验证 $\boldsymbol{C} \in V$ 当且仅当 $\boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^{-1}\boldsymbol{C} = \boldsymbol{C}\boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^{-1}$, 当且仅当 $\boldsymbol{B}(\boldsymbol{T}^{-1}\boldsymbol{C}\boldsymbol{T}) = (\boldsymbol{T}^{-1}\boldsymbol{C}\boldsymbol{T})\boldsymbol{B}$. 由于

与教材 P157 的作业题 #48 题相关

$\boldsymbol{B}$ 是主对角线上元素互不相等的对角阵，这等价于说 $\boldsymbol{T}^{-1}\boldsymbol{C}\boldsymbol{T}$ 是一个 F 上的对角阵. 由此，我们看出 $\dim(V)=n$，而 V 有一组基 $\{\,\boldsymbol{T}\boldsymbol{E}_{ii}\boldsymbol{T}^{-1}\mid 1\le i\le n\,\}$.

接下来，我们给出 V 的另外一组基. 若记之前提到的对角阵 $\boldsymbol{T}^{-1}\boldsymbol{C}\boldsymbol{T}=\operatorname{diag}(\mu_1,\dots,\mu_n)$，由 *Lagrange* 插值公式 (教材 *P115#32*) 可知，存在 $f(x)\in F_{n-1}[x]$，使得 $f(\lambda_i)=\mu_i$，$1\le i\le n$. 这说明矩阵的多项式 $f(\boldsymbol{B})=\boldsymbol{T}^{-1}\boldsymbol{C}\boldsymbol{T}$，即

$$\boldsymbol{C}=\boldsymbol{T}f(\boldsymbol{B})\boldsymbol{T}^{-1}=f(\boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^{-1})=f(\boldsymbol{A}).$$

这说明 $\boldsymbol{C}$ 可由 $\{\,\boldsymbol{I},\boldsymbol{A},\boldsymbol{A}^2,\dots,\boldsymbol{A}^{n-1}\,\}$ 生成. 由于这组向量显然在 V 中，它们构成了 V 的一组生成元. 又由于已知 $\dim(V)=n$，这足以说明 $\{\,\boldsymbol{I},\boldsymbol{A},\boldsymbol{A}^2,\dots,\boldsymbol{A}^{n-1}\,\}$ 也构成了 V 的一组基.

若更一般地，$\boldsymbol{A}$ 不可对角化，我们也有相应的结果；请参考习题 *6.5.3*.

特征值的代数重数与几何重数 (※) 对于复数域 $\mathbb{C}$ 上的 n 阶矩阵 $\boldsymbol{A}$，设其特征多项式为

$$p_{\boldsymbol{A}}(\lambda)=(\lambda-\lambda_1)^{n_1}\cdots(\lambda-\lambda_s)^{n_s},$$

其中 $\lambda_1,\dots,\lambda_s$ 是 $\boldsymbol{A}$ 的所有两两不同的特征值 (不计重数). $n_i\ge 1$ 称为 λ_i 的**代数重数** (algebraic multiplicity). 对于 λ_i，考察特征子空间

$$V_{\boldsymbol{A}}(\lambda_i)=\{\,\boldsymbol{x}\in F^n\mid \boldsymbol{A}\boldsymbol{x}=\lambda_i\boldsymbol{x}\,\}.$$

则向量空间维数 $m_i=\dim(V_{\boldsymbol{A}}(\lambda_i))\ge 1$ 称为 λ_i 的**几何重数** (geometric multiplicity).

定理 6.4.11. 用上面的记号，则对于每个 i，我们总有 $m_i\le n_i$. 而 $\boldsymbol{A}$ 可以对角化的充要条件是对于每个 i，等号成立：$m_i=n_i$. 教材引理 6.4.3, 定理 6.4.2

定理的证明与本节的例子，留给学生课后阅读.

注 6.4.12. 综合之前的讨论，我们总结一下求相似对角阵的方法. 设 $\boldsymbol{A}$ 是给定的 n 阶方阵.

(1) 求 $\boldsymbol{A}$ 的特征值，得到 $p_{\boldsymbol{A}}(\lambda)=\det(\lambda\boldsymbol{I}_n-\boldsymbol{A})=\prod_{i=1}^{s}(\lambda-\lambda_i)^{n_i}$，其中 $\lambda_1,\lambda_2,\dots,\lambda_s$ 是 $\boldsymbol{A}$ 的所有两两不同的特征值.

(2) 若对每个 i 都有 $\operatorname{rank}(\lambda_i\boldsymbol{I}_n-\boldsymbol{A})=n-n_i$，则 $\boldsymbol{A}$ 可对角化；否则，$\boldsymbol{A}$ 不可对角化.

(3) 在可对角化的前提下，对每个 λ_i，求出方程组 $(\lambda_i\boldsymbol{I}_n-\boldsymbol{A})$ 的一组基础解系 $\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},\dots,\boldsymbol{x}_{in_i}$.

(4) 令 $\boldsymbol{T} = (\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \ldots, \boldsymbol{x}_{1n_1}, \boldsymbol{x}_{21}, \boldsymbol{x}_{22}, \ldots, \boldsymbol{x}_{in_2}, \ldots, \boldsymbol{x}_{s1}, \boldsymbol{x}_{s2}, \ldots, \boldsymbol{x}_{sn_s})$, 则 $\boldsymbol{T}$ 是可逆方阵, 并满足

$$\boldsymbol{T}^{-1} A \boldsymbol{T} = \operatorname{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{n_1\text{ 个}}, \underbrace{\lambda_2, \ldots, \lambda_2}_{n_2\text{ 个}}, \ldots, \underbrace{\lambda_s, \ldots, \lambda_s}_{n_s\text{ 个}}).$$

例 6.4.13. 重新考察矩阵 $\boldsymbol{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$. 则 $p_{\boldsymbol{A}}(\lambda) = (\lambda - 2)^3$, 故 $\boldsymbol{A}$ 的特征值 2 的代数重数为 3. 另一方面, 相应的特征子空间为

$$V_{\boldsymbol{A}}(2) = \left\{ \boldsymbol{x} \in F^3 \mid \boldsymbol{A}\boldsymbol{x} = 2\boldsymbol{x} \right\},$$

即线性方程组 $(2\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ 的解空间. 由于

$$\operatorname{rank}(2\boldsymbol{I} - \boldsymbol{A}) = \operatorname{rank} \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} = 2,$$

故解空间的维数为 $3 - 2 = 1$, 即特征值 2 的几何重数为 1. 由于 $1 < 3$, 故原矩阵 $\boldsymbol{A}$ 不可对角化.

例 6.4.15. 对于下三角方阵 $\boldsymbol{A}=\begin{pmatrix}1&0&0&0\\a&1&0&0\\2&b&2&0\\3&2&c&2\end{pmatrix}$, 问 a,b,c 各为何值时, $\boldsymbol{A}$ 可以相似对角化?

解. 容易看出, 特征多项式 $p_{\boldsymbol{A}}(\lambda)=(\lambda-1)^2(\lambda-2)^2$, 从而特征值为 $\lambda_1=1$, $\lambda_2=2$, 皆为两重. 此时, $\boldsymbol{A}$ 可以相似对角化的充要条件是 $\lambda_1=1$ 与 $\lambda_2=2$ 的几何重数皆为 2.

(1) 当 $\lambda=1$, $\lambda\boldsymbol{I}-\boldsymbol{A}=\begin{pmatrix}0&0&0&0\\-a&0&0&0\\-2&-b&\boxed{\text{-1}}&0\\-3&-2&-c&\boxed{\text{-1}}\end{pmatrix}$. 则 $\lambda_1=1$ 的几何重数为 2, 当且仅当 $2=4-\operatorname{rank}(\boldsymbol{I}-\boldsymbol{A})$, 当且仅当 $\operatorname{rank}(\boldsymbol{A})=2$. 由于上面的矩阵的第三行于第四行已经线性无关, 我们可以由此推出 $a=0$, 此时对于 b 和 c 没有限制.

(2) 类似地, 考虑 $\lambda=2$. 可以推出 $c=0$, 对于 a 和 b 没有限制.

综上, $a=c=0$, 而 b 任意. □

例 6.4.16. 已知 $\boldsymbol{A}_t=\begin{pmatrix}t&t-2&4-2t\\3&-1&0\\1+t&t-2&3-2t\end{pmatrix}$. 若 $\boldsymbol{A}_t$ 可对角化, 描述此时的 t, 并求出 $\boldsymbol{P}$, 使得 $\boldsymbol{P}^{-1}\boldsymbol{A}_t\boldsymbol{P}$ 是对角阵.

解. 特征多项式 $p_{\boldsymbol{A}_t}(\lambda)=(\lambda-2)(\lambda+1)(\lambda-1+t)$, 故特征值为 $2,-1,1-t$.

(1) 若 $t\neq-1,2$, 则 $\boldsymbol{A}_t$ 有三个不同的特征值, 可以对角化. 容易求得 $\boldsymbol{P}=\begin{pmatrix}1&0&t-2\\1&2&-3\\1&1&t-2\end{pmatrix}$, 使得 $\boldsymbol{P}^{-1}\boldsymbol{A}_t\boldsymbol{P}=\operatorname{diag}(2,-1,1-t)$.

(2) 若 $t=2$, 容易求得 $\boldsymbol{P}=\begin{pmatrix}1&0&0\\1&2&-3\\1&1&0\end{pmatrix}$, 使得 $\boldsymbol{P}^{-1}\boldsymbol{A}_t\boldsymbol{P}=\operatorname{diag}(2,-1,-1)$.

(3) 若 $t=-1$, 则特征值 2 的几何重数为 1, 小于代数重数, 从而矩阵不可对角化. □

例 6.4.17. 设方阵 $\boldsymbol{A}$ 为**幂等矩阵**: $\boldsymbol{A}^2=\boldsymbol{A}$. 证明:

(1) $\boldsymbol{A}$ 的特征值只有 0 和 1;

(2) $\boldsymbol{A}$ 相似于其相抵标准形 $\begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}$, 其中 $r = \operatorname{rank}(\boldsymbol{A})$;

(3) $\operatorname{tr}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$. (教材第六章作业题 #27)

证明. (1) 设 $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量: $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. 同时左乘 $\boldsymbol{A}$, 我们有 $\boldsymbol{A}^2\boldsymbol{x} = \boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda\boldsymbol{A}\boldsymbol{x} = \lambda^2\boldsymbol{x}$. 另一方面, 利用幂等性质, $\boldsymbol{A}^2\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. 由于 $\boldsymbol{x}$ 不是零向量, 这说明 $\lambda^2 = \lambda$, 从而 $\lambda = 0$ 或 1.

(2) 对于 (可能的) 特征值 $\lambda_1 = 0$, 其几何重数为

$$m_1 = \dim\{\, \boldsymbol{x} \mid (0\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \,\} = n - \operatorname{rank}(\boldsymbol{A}).$$

类似地, 对于 (可能的) 特征值 $\lambda_2 = 1$, 其几何重数为 $m_2 = n - \operatorname{rank}(\boldsymbol{I} - \boldsymbol{A})$. 以前, 我们已经推导过, 对于幂等矩阵 $\boldsymbol{A}$, 有

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{I} - \boldsymbol{A}) = n.$$

这说明 $m_1 + m_2 = n$. 但是, 我们知道特征值的代数重数不小于几何重数: $n_1 \geq m_1$, $n_2 \geq m_2$. 另一方面, 我们总有 $n_1 + n_2 = n$. 这说明 $n_1 = m_1$, $n_2 = m_2$. 从而对于 $\boldsymbol{A}$ 的所有特征值, 其相应的代数重数总是等于几何重数. 从而 $\boldsymbol{A}$ 可以相似对角化. 由定理 6.4.2 的证明, 我们可以假设这些 1 都在对角形的左上角. (注: 若 λ 不是真正的特征值, 则相应的“代数重数”与“几何重数”皆为零, 不实质影响这儿的讨论.) 从而此时, 有 $\boldsymbol{A} \sim \boldsymbol{D} \coloneqq \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}$. 由于矩阵的秩和迹是相似不变量, 我们有 $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{D}) = r = \operatorname{rank}(\boldsymbol{D}) = \operatorname{rank}(\boldsymbol{A})$.

(3) 从前一问的讨论来看, 这是显然的. 另外, 在例 4.5.20 中, 我们给出了另外一个证明. 除此之外, 也可以看看教材上用若尔当标准形的方法重新考虑这道例题 (P188 例 6.5.6). $\square$

例 6.4.18. 计算行列式 $\begin{vmatrix} x & 1 & 2 & \cdots & n-1 \\ -1 & \ddots & \ddots & \ddots & \vdots \\ -2 & \ddots & \ddots & \ddots & 2 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 1-n & \cdots & -2 & -1 & x \end{vmatrix}$.

解. **(解法一)** 若令 $\boldsymbol{A}=(a_{ij})_{n\times n}$, 其中 $a_{ij}=i-j$, 那么所求就是 $\boldsymbol{A}$ 的特征多项式 $p_{\boldsymbol{A}}(x)$. 由于 $\boldsymbol{A}$ 的各行的元素构成等差数列 (的一部分), 对于 $1\le i\le n-2$, 容易看出 $(0,\ldots,0,\underbrace{1}_{\text{第 } i \text{ 个位置}},-2,1,0,\ldots,0)$ 是 $\boldsymbol{A}$ 的一个关于 0 的特征向量. 由于这些特征向量线性无关, 0 的几何重数至少为 $n-2$, 从而代数重数至少为 $n-2$, 即 $\boldsymbol{A}$ 至少有 $n-2$ 重的 0 特征根, 即 $\boldsymbol{A}$ 的特征根可以设为 $\lambda_1,\lambda_2,\lambda_3=\cdots=\lambda_n=0$. 由此看出 $p_{\boldsymbol{A}}(x)=(x-\lambda_1)(x-\lambda_2)x^{n-2}=x^n+a_{n-2}x^{n-2}$ (这儿用到了 $a_{n-1}=-\operatorname{tr}(\boldsymbol{A})=0$ 这一事实). 于是, 对所求的行列式, 只需观察 x^{n-2} 前面的系数. 很明显, 作行列式的全展开时, 若得到次数为 x^{n-2} 的项, 则这些 x 来源于主对角线上 $n-2$ 个位置, 从而留下了主对角线上的 2 个位置, 比如说, 第 i,j 两位置, 其中 $i<j$. 于是, 相应的项的系数为 $\begin{vmatrix} 0 & i-j \\ j-i & 0 \end{vmatrix}$. 于是, 所求的 a_{n-2} 为这些二阶行列式的和, 即

$$a_{n-2}=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(j-i)^2=\frac{n^2(n^2-1)}{12}.$$

这儿用到了公式 $\sum_{i=1}^n i=\frac{n(n+1)}{2}$, $\sum_{i=1}^n i^2=\frac{n(n+1)(2n+1)}{6}$, 以及 $\sum_{i=1}^n i^3=\frac{n^2(n+1)^2}{4}$.

(解法二)

第二种解法课堂不讲

$$\begin{vmatrix} x & 1 & 2 & \cdots & n-1 \\ -1 & x & 1 & \ddots & \vdots \\ -2 & -1 & x & \ddots & 2 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 1-n & \cdots & -2 & -1 & x \end{vmatrix} \xlongequal[\substack{-c_{n-2}\to c_{n-1} \\ \cdots \\ -c_1\to c_2}]{\text{依次}-c_{n-1}\to c_n} \begin{vmatrix} x & 1-x & 1 & \cdots & 1 \\ -1 & x+1 & \ddots & \ddots & \vdots \\ -2 & 1 & \ddots & \ddots & 1 \\ \vdots & \vdots & \ddots & \ddots & 1-x \\ -(n-1) & 1 & \cdots & 1 & x+1 \end{vmatrix}$$

$$\xlongequal[\substack{r_3\to r_1 \\ \vdots \\ r_n\to r_1}]{\text{依次}r_2\to r_1} \begin{vmatrix} x-\frac{n(n-1)}{2} & n & \cdots & \cdots & n \\ -1 & x+1 & 1-x & 1\cdots & 1 \\ -2 & 1 & \ddots & \ddots & 1 \\ \vdots & \vdots & \ddots & \ddots & 1-x \\ -(n-1) & 1 & \cdots & 1 & x+1 \end{vmatrix}$$

$$
\xlongequal{\frac{1}{n}r_1} n\begin{vmatrix}
a:=\frac{x}{n}-\frac{n-1}{2} & 1 & \cdots & \cdots & \cdots & 1 \\
-1 & x+1 & 1-x & 1 & \cdots & 1 \\
-2 & 1 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & 1 \\
\vdots & \vdots & & \ddots & \ddots & 1-x \\
-(n-1) & 1 & \cdots & \cdots & 1 & x+1
\end{vmatrix}
$$

$$
\xlongequal[\substack{-r_1\to r_3 \\ \vdots \\ -r_1\to r_n}]{\text{依次}-r_1\to r_2} n\begin{vmatrix}
a & 1 & \cdots & \cdots & \cdots & 1 \\
-1-a & x & -x & 0 & \cdots & 0 \\
-2-a & 0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & 0 \\
\vdots & \vdots & & \ddots & \ddots & -x \\
-(n-1)-a & 0 & \cdots & \cdots & 0 & x
\end{vmatrix}
$$

$$
\xlongequal[\substack{r_{n-1}\to r_{n-2} \\ \vdots \\ r_3\to r_2}]{\text{依次}r_n\to r_{n-1}} n\begin{vmatrix}
a & 1 & \cdots & \cdots & 1 \\
b_2:=-(n-1)-(n-2)-\cdots-1-(n-1)a & x & 0 & \cdots & 0 \\
b_3:=-(n-1)-(n-2)-\cdots-2-(n-2)a & 0 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
b_{n-1}:=-(n-1)-(n-2)-2a & \vdots & & \ddots & 0 \\
b_n:=-(n-1)-a & 0 & \cdots & 0 & x
\end{vmatrix}
$$

$$
\xlongequal[\substack{\frac{-b_{n-1}}{x}c_{n-1}\to c_1 \\ \vdots \\ \frac{-b_2}{x}c_2\to c_1}]{(\star)\text{依次}\frac{-b_n}{x}c_n\to c_1} n\begin{vmatrix}
a-\sum_{k=2}^{n}\frac{b_k}{x} & 1 & \cdots & \cdots & 1 \\
0 & x & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & 0 \\
0 & \cdots & \cdots & 0 & x
\end{vmatrix} = nx^{n-1}\left(a-\sum_{k=2}^{n}\frac{b_k}{x}\right).
$$

在上面的 $(\star)$ 这一步, 我们视 x 为未知元, 或者利用连续性, 不妨假定 x 不为 0, 因此可以在这一步除以 x. 化简最后的表达式, 即可. □

习题 6.4.19. 设 $\mathbb{R}_2[x]$ 表示次数不超过 2 的多项式全体. 记线性变换

$$
\varphi:\mathbb{R}_2[x]\to\mathbb{R}_2[x],\quad f(x)\mapsto f(x+1)+f'(x).
$$

判断线性变换 φ 是否可对角化.

相似于上三角矩阵 由于不是所有的复方阵都可以对角化, 我们退而求其次, 考虑在相似变换后其它可能的简单情形: 上三角化.

定理 6.4.20 (Schur 定理)**.** 设 $\boldsymbol{A}$ 是数域 F 上的 n 阶方阵, 在 F 中有 n 个特征值 $\lambda_1, \lambda_2, \ldots, \lambda_n$ (不要求互不相等), 那么一定存在 F 上的可逆矩阵 $\boldsymbol{T}$ 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 为上三角矩阵.

证明. 我们对于 n 用归纳法. $n=1$ 的情形是显然的. 对于一般的 n 阶方阵 $\boldsymbol{A}$, 不妨设 $\boldsymbol{x}_1$ 是 $\boldsymbol{A}$ 的属于 λ_1 的一个特征向量. 非零向量 $\boldsymbol{x}_1$ 可以扩充为 F^n 的一组基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. 将这些列向量按行排列, 得到可逆矩阵 $\boldsymbol{T}_1 = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in F^{n\times n}$. 直接计算, 我们有

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{T}_1 &= \boldsymbol{A}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = (\boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2, \ldots, \boldsymbol{A}\boldsymbol{x}_n) = (\lambda_1\boldsymbol{x}_1, *, \ldots, *) \\
&= (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\begin{pmatrix} \lambda_1 & * \cdots\cdots * \\ 0 & * \cdots\cdots * \\ \vdots & \vdots \qquad \vdots \\ 0 & * \cdots\cdots * \end{pmatrix} = \boldsymbol{T}_1\begin{pmatrix} \lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{A}_1 \end{pmatrix}
\end{aligned}$$

其中 $\boldsymbol{A}_1 \in F^{(n-1)\times(n-1)}$. 这说明 $\boldsymbol{T}_1^{-1}\boldsymbol{A}\boldsymbol{T}_1$ 为准上三角矩阵 $\begin{pmatrix} \lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{A}_1 \end{pmatrix}$. 我们观察到,

$$(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n) = p_{\boldsymbol{A}}(\lambda) = p_{\boldsymbol{T}_1^{-1}\boldsymbol{A}\boldsymbol{T}_1}(\lambda) = (\lambda-\lambda_1)p_{\boldsymbol{A}_1}(\lambda).$$

这说明 $\boldsymbol{A}_1$ 的特征值为 $\lambda_2, \ldots, \lambda_n$. 由归纳假设, 对于 $n-1$ 阶方阵 $\boldsymbol{A}_1$, 存在可逆矩阵 $\boldsymbol{T}_2$ 使得 $\boldsymbol{T}_2^{-1}\boldsymbol{A}\boldsymbol{T}_2$ 为上三角矩阵.

考虑矩阵 $\boldsymbol{T} = \boldsymbol{T}_1\begin{pmatrix} 1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2 \end{pmatrix}$. 直接计算, 我们有

$$\begin{aligned}
\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} &= \begin{pmatrix} 1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1} \end{pmatrix}\boldsymbol{T}_1^{-1}\boldsymbol{A}\boldsymbol{T}_1\begin{pmatrix} 1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2 \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1} \end{pmatrix}\begin{pmatrix} \lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{A}_1 \end{pmatrix}\begin{pmatrix} 1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2 \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1}\boldsymbol{A}_1 \end{pmatrix}\begin{pmatrix} 1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{T}_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & * \\ \boldsymbol{O} & \boldsymbol{T}_2^{-1}\boldsymbol{A}_1\boldsymbol{T}_2 \end{pmatrix}.
\end{aligned}$$

由归纳假设, 这是一个上三角矩阵矩阵. □

注 6.4.21. (1) 定理 *6.4.20* 中的条件在 $F=\mathbb{C}$ 时自动成立, 这说明任何复方阵都复相似于一个上三角矩阵.

(2) 用转置运算, 我们可以看出, 复方阵也可以相似于一个下三角矩阵.

(3) 显然, 定理中的上三角矩阵的主对角线上的元素就是全部这些特征值, 并且我们的证明可以进一步保证, 这些特征值按照指定的顺序在对角线上来排列.

(4) 我们上面定理的证明相对比较粗略. 事实上, 任何复方阵都可以相似于某个若尔当矩阵 (称为该矩阵的若尔当标准型). 这种上三角矩阵可视为原矩阵的相似等价类中的最简形式的代表元. 书上这一块的内容, 感兴趣的学生可以课后自学.

(5) 对于上面定理中用来上三角化的可逆复矩阵 $\boldsymbol{T}$, 我们可以通过正交化的方法, 进一步假定其为一个酉矩阵.

注 6.4.22. 若方阵 $\boldsymbol{A}$ 为实方阵, 并且 n 个复特征值 (带重数) 都是实数, 那么上面的定理说明 $\boldsymbol{A}$ 可以实相似于实的上三角矩阵. 其逆命题也显然成立: 若实矩阵 $\boldsymbol{A}$ 有特征值不是实数, 那么它显然无法实相似于实的上三角矩阵.

事实上, 特征值不全是实数的实方阵 $\boldsymbol{A}$ 只能实相似于某种准上 (或下) 三角阵: 其主对角线上是 2 阶方阵或 $\boldsymbol{A}$ 的特征值, 其中特征值为原实方阵的特征多项式的实根, 2 阶方阵块对应于成对出现的共轭复根. 大致证明思路如下:

(1) 若 $\lambda = a + b\mathrm{i}$ 为实方阵 $\boldsymbol{A}$ 的虚特征值 (故 $b \neq 0$), 而 $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2\mathrm{i} \in \mathbb{C}^n$ 是相应的复特征向量, 其中 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^2$, 即有 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. 可以验证 $\boldsymbol{x}_1, \boldsymbol{x}_2$ 在 $\mathbb{R}$ 上线性无关 (留作练习), 从而可以扩充为 $\mathbb{R}^n$ 的一组基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. 令 $\boldsymbol{P} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in \mathbb{R}^{n\times n}$, 由于 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, 而 $\boldsymbol{A}\overline{\boldsymbol{x}} = \overline{\lambda}\overline{\boldsymbol{x}}$, 故 $\boldsymbol{A}\boldsymbol{x}_1 = a\boldsymbol{x}_1 - b\boldsymbol{x}_2$, 以及 $\boldsymbol{A}\boldsymbol{x}_2 = b\boldsymbol{x}_1 + a\boldsymbol{x}_2$. 这说明 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{A}_1 & * \\ \boldsymbol{O} & * \end{pmatrix}$, 其中 $\boldsymbol{A}_1 = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$.

(2) 利用定理 *6.4.20* 的证明思路不难推出, 存在可逆的实矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{A}_1 & & * \\ & \ddots & \\ & & \boldsymbol{A}_m \end{pmatrix}$ 为准上三角阵, 其中的方阵 $\boldsymbol{A}_j$ 为形如 $\begin{pmatrix} a_j & b_j \\ -b_j & a_j \end{pmatrix}$ 的实方阵 (此时, $\boldsymbol{A}$ 有复特征值 $a \pm b\mathrm{i}$), 或为由实数 ($\boldsymbol{A}$ 的实特征值) 给出的 1 阶矩阵.

习题 6.4.23. 假定 $\lambda = a + b\mathrm{i}$ 为实方阵 $\boldsymbol{A}$ 的虚特征值 (故 $b \neq 0$), 而 $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2\mathrm{i} \in \mathbb{C}^n$ 是相应的一个复特征向量, 其中 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^2$. 证明: $\boldsymbol{x}_1, \boldsymbol{x}_2$ 在 $\mathbb{R}$ 上线性无关,

例 6.4.24. 重新考虑习题 4.5.29 中的情形, 即有 n 阶复方阵 $\boldsymbol{A}$ 满足 $\operatorname{rank}(\boldsymbol{A}) = 1$. 此时, 存在两个非零列向量 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 使得 $\boldsymbol{A} = \boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}$, 从而 $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}) = \operatorname{tr}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\alpha}) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\alpha}$, 并有 $\boldsymbol{A}^2 = \boldsymbol{\alpha}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\alpha})\boldsymbol{\beta}^{\mathrm{T}} = \operatorname{tr}(\boldsymbol{A})\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}} = \operatorname{tr}(\boldsymbol{A})\boldsymbol{A}$.

在 $\operatorname{tr}(\boldsymbol{A}) \neq 0$ 的条件下, 我们来证明 $\boldsymbol{A}$ 可以对角化. 设 $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的属于特征值 λ 的特征向量. 由于 $\boldsymbol{A}^2 = \operatorname{tr}(\boldsymbol{A})\boldsymbol{A}$. 右乘 $\boldsymbol{x}$, 得到 $\lambda^2\boldsymbol{x} = \operatorname{tr}(\boldsymbol{A})\lambda\boldsymbol{x}$. 说明 $\lambda = \operatorname{tr}(\boldsymbol{A})$ 或者为

0. 又由于 $\boldsymbol{A}$ 的特征值之和为 $\operatorname{tr}(\boldsymbol{A})$, 这说明, 特征值 $\operatorname{tr}(\boldsymbol{A})$ 的重数为 1, 特征值 0 的重数为 $n-1$. 由 *Schur* 定理 *6.4.20* 可知, $\boldsymbol{A}$ 可相似于上三角矩阵 $\boldsymbol{B}$, 其中 $\boldsymbol{B}$ 的主对角线上的元素依次为 $\operatorname{tr}(\boldsymbol{A}),0,\ldots,0$. 由于 $\operatorname{tr}(\boldsymbol{A})\neq 0$, 而 $\operatorname{rank}(\boldsymbol{B})=\operatorname{rank}(\boldsymbol{A})=1$, 这说明 $\boldsymbol{B}=\begin{pmatrix}\operatorname{tr}(\boldsymbol{A}) & * \\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$. 此时可以直接验证, $\boldsymbol{B}$ 关于特征值 $\operatorname{tr}(\boldsymbol{A})$ 的代数重数和几何重数皆为 1, 关于特征值 0 的代数重数和几何重数皆为 $n-1$. 故矩阵 $\boldsymbol{B}$ (从而矩阵 $\boldsymbol{A}$) 相似于对角阵 $\begin{pmatrix}\operatorname{tr}(\boldsymbol{A}) & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$. 另外, 利用若尔当标准形的方法, 也可以很容易地给出该相似对角化的证明.

当然, 若 $\operatorname{tr}(\boldsymbol{A})=0$, 则 $\boldsymbol{A}$ 具有 n 重的特征值 0, 其几何重数为 $n-\operatorname{rank}(-\boldsymbol{A})=n-1$, 从而 $\boldsymbol{A}$ 不可对角化.

例 6.4.25 (Caylay–Hamilton)**.** 令 $\boldsymbol{A}$ 是复数域上的一个 n 阶方阵, $p_{\boldsymbol{A}}(\lambda)$ 是 $\boldsymbol{A}$ 的特征多项式. 证明: $p_{\boldsymbol{A}}(\boldsymbol{A})=\boldsymbol{O}$.

在给出定理的证明之前, 我们解释一下上面 “零化多项式” 的意思. 例如在例 6.3.13 中, 对于矩阵 $\boldsymbol{A}=\begin{pmatrix}1&2&2\\2&1&2\\2&2&1\end{pmatrix}$, 我们有 $p_{\boldsymbol{A}}(\lambda)=\lambda^3-3\lambda^2-9\lambda-5$. 那么, 上面的定理指出

$$p_{\boldsymbol{A}}(\boldsymbol{A})=\boldsymbol{A}^3-3\boldsymbol{A}^2-9\boldsymbol{A}-5\boldsymbol{I}_3=\boldsymbol{O}_3.$$

下面, 我们给出 Caylay–Hamilton 定理的证明.

课堂不讲证明

证明. **(思路一)** 设 $p_{\boldsymbol{A}}(\lambda)=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\cdots(\lambda-\lambda_s)^{n_s}$, 其中特征值 $\lambda_1,\ldots,\lambda_s$ 互不相等. 方阵 $\boldsymbol{A}$ 可以相似上三角化, 即存在可逆阵 $\boldsymbol{P}$, 使得 $\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}=\boldsymbol{J}$ 为上三角阵. 此时,

$$\begin{aligned}p_{\boldsymbol{A}}(\boldsymbol{A})=&(\boldsymbol{A}-\lambda_1\boldsymbol{I})^{n_1}\cdots(\boldsymbol{A}-\lambda_s\boldsymbol{I})^{n_s}\\=&(\boldsymbol{P}^{-1}\boldsymbol{J}\boldsymbol{P}-\lambda_1\boldsymbol{I})^{n_1}\cdots(\boldsymbol{P}^{-1}\boldsymbol{J}\boldsymbol{P}-\lambda_s\boldsymbol{I})^{n_s}\\=&\left(\boldsymbol{P}^{-1}(\boldsymbol{J}-\lambda_1\boldsymbol{I})\boldsymbol{P}\right)^{n_1}\cdots\left(\boldsymbol{P}^{-1}(\boldsymbol{J}-\lambda_s\boldsymbol{I})\boldsymbol{P}\right)^{n_s}\\=&\boldsymbol{P}^{-1}(\boldsymbol{J}-\lambda_1\boldsymbol{I})^{n_1}\cdots(\boldsymbol{J}-\lambda_s\boldsymbol{I})^{n_s}\boldsymbol{P}.\end{aligned}$$

通过适当选取可逆阵 $\boldsymbol{P}$, 我们可以假定 $\boldsymbol{J}$ 为如下的上三角阵:

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_1 & * & * & * \\ & \boldsymbol{J}_2 & * & * \\ & & \ddots & \vdots \\ & & & \boldsymbol{J}_s \end{pmatrix},$$

其中

$$\boldsymbol{J}_i = \begin{pmatrix} \lambda_i & * & * & * \\ & \lambda_i & * & * \\ & & \ddots & \vdots \\ & & & \lambda_i \end{pmatrix}, \qquad i = 1, 2, \cdots, s.$$

(在 Schur 定理 6.4.20 的证明中, 我们有挑选特征值的自由, 从而可以假定特征值按上面指定的顺序排列) 此时,

$$(\boldsymbol{J}_i - \lambda_i \boldsymbol{I}_{n_i})^{n_i} = \begin{pmatrix} 0 & * & * & * \\ & 0 & * & * \\ & & \ddots & \vdots \\ & & & 0 \end{pmatrix}^{n_i}.$$

对于 $1 \le k \le n_i$, 若记 $(\boldsymbol{J}_i - \lambda_i \boldsymbol{I}_{n_i})^k = \left(s_{x,y}^{(i,k)}\right)_{1 \le x,y \le n_i}$, 我们用归纳法可以验证, 当 $y < x + k$ 时, $s_{x,y}^{(i,k)} = 0$. 特别地, 当 $k = n_i$, 所有的 $s_{x,y}^{i,n_i} = 0$, 即 $(\boldsymbol{J}_i - \lambda_i \boldsymbol{I}_{n_i})^{n_i} = \boldsymbol{O}_{n_i}$ 为 n_i 阶零方阵. 故

$$(\boldsymbol{J} - \lambda_i \boldsymbol{I})^{n_i} = \begin{pmatrix} (\boldsymbol{J}_1 - \lambda_i \boldsymbol{I}_{n_1})^{n_s} & * & * & * & * & * & * \\ & \ddots & * & * & * & * & * \\ & & (\boldsymbol{J}_{i-1} - \lambda_i \boldsymbol{I}_{n_{i-1}})^{n_i} & * & * & * & * \\ & & & \boldsymbol{O}_{n_i} & * & * & * \\ & & & & (\boldsymbol{J}_{i+1} - \lambda_i \boldsymbol{I}_{n_{i+1}})^{n_i} & * & * \\ & & & & & \ddots & \vdots \\ & & & & & & (\boldsymbol{J}_s - \lambda_i \boldsymbol{I}_{n_s})^{n_i} \end{pmatrix}.$$

通过归纳法, 我们可以直接验证 $(\boldsymbol{J} - \lambda_1 \boldsymbol{I})^{n_1} \cdots (\boldsymbol{J} - \lambda_i \boldsymbol{I})^{n_i}$ 是前 $(n_1 + \cdots + n_i)$ 列为零的上三角阵. 特别地, $(\boldsymbol{J} - \lambda_1 \boldsymbol{I})^{n_1} \cdots (\boldsymbol{J} - \lambda_s \boldsymbol{I})^{n_s} = \boldsymbol{O}$. 从而 $p_{\boldsymbol{A}}(\boldsymbol{A}) = \boldsymbol{O}$.

(思路二) 设方阵 $\lambda \boldsymbol{I}_n - \boldsymbol{A}$ 的伴随矩阵为 $\boldsymbol{B}$. 显然 $\boldsymbol{B}$ 的每个元素都是关于 λ 的 $\mathbb{C}$ 上的次数不超过 $n-1$ 的多项式. 因此, 我们可以设 $\boldsymbol{B} = \sum_{i=0}^{n-1} \lambda^i \boldsymbol{B}_i$, 其中 $\boldsymbol{B}_i \in \mathbb{C}^{n \times n}$.

利用伴随矩阵的性质, 我们有

$$(\lambda \boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{B} = p_{\boldsymbol{A}}(\lambda)\boldsymbol{I}_n.$$

我们不妨设 $p_{\boldsymbol{A}}(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$. 另一方面, 直接乘, 我们有

$$(\lambda \boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{B} = (\lambda \boldsymbol{I}_n - \boldsymbol{A})\sum_{i=0}^{n-1}\lambda^i \boldsymbol{B}_i = \lambda^n \boldsymbol{B}_{n-1} + \sum_{i=1}^{n-1}\lambda^i(\boldsymbol{B}_{i-1} - \boldsymbol{A}\boldsymbol{B}_i) - \boldsymbol{A}\boldsymbol{B}_0.$$

比较 “系数”, 我们有

$$\begin{cases} \boldsymbol{B}_{n-1} = \boldsymbol{I}_n, \\ \boldsymbol{B}_{n-2} - \boldsymbol{A}\boldsymbol{B}_{n-1} = a_{n-1}\boldsymbol{I}_n, \\ \qquad \vdots \\ \boldsymbol{B}_0 - \boldsymbol{A}\boldsymbol{B}_1 = a_1\boldsymbol{I}_n, \\ -\boldsymbol{A}\boldsymbol{B}_0 = a_0\boldsymbol{I}_n. \end{cases}$$

上面的各式依次分别左乘矩阵 $\boldsymbol{A}^n, \boldsymbol{A}^{n-1}, \ldots, \boldsymbol{A}, \boldsymbol{I}$, 并相加, 即得到

$$\begin{aligned} p_{\boldsymbol{A}}(\boldsymbol{A}) =& \boldsymbol{A}^n + a_{n-1}\boldsymbol{A}^{n-1} + \cdots + a_1\boldsymbol{A} + a_0\boldsymbol{I}_n \\ =& \boldsymbol{A}^n\boldsymbol{B}_{n-1} + \boldsymbol{A}^{n-1}(\boldsymbol{B}_{n-2} - \boldsymbol{A}\boldsymbol{B}_{n-1}) + \cdots + \boldsymbol{A}(\boldsymbol{B}_0 - \boldsymbol{A}\boldsymbol{B}_1) - \boldsymbol{A}\boldsymbol{B}_0 \\ =& \boldsymbol{O}. \end{aligned}$$

□

例 6.4.26. 设 $\boldsymbol{A} = \begin{pmatrix} 2 & -2 & 4 \\ 2 & 3 & 2 \\ -1 & 1 & -1 \end{pmatrix}$, 它的特征多项式是 $p_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I}_3 - \boldsymbol{A}) = \lambda^3 - 4\lambda^2 + 7\lambda - 10$. 利用 *Caylay–Hamilton* 定理, 这说明了

$$\boldsymbol{A}^{-1} = \frac{1}{10}\left(\boldsymbol{A}^2 - 4\boldsymbol{A} + 7\boldsymbol{I}_3\right) = \frac{1}{10}\begin{pmatrix} -5 & 2 & -16 \\ 0 & 2 & 4 \\ 5 & 0 & 10 \end{pmatrix}.$$

例 6.4.27. 求 $\boldsymbol{A}^{100}$, 其中

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

证明. 不难验证, 该矩阵不可相似对角化. 因此, 将矩阵对角化后求解的方法暂时不适用. 当然, 若接用若尔当标准形, 我们也可以稍微简化一下计算. 现在我们看一下如何利用 Caylay–Hamilton 定理来计算.

可以求出, $\boldsymbol{A}$ 的特征多项式为 $p_{\boldsymbol{A}}(\lambda)=\lambda(\lambda-1)^3$. 对多项式 λ^{100} 关于 $p_{\boldsymbol{A}}(\lambda)$ 作带余除法:

$$\lambda^{100}=p_{\boldsymbol{A}}(\lambda)f(\lambda)+r(\lambda), \tag{6.4}$$

其中 $r(\lambda)=a\lambda^3+b\lambda^2+c\lambda+d$. 由于 0 是 $\boldsymbol{A}$ 的单特征值, 1 是 $\boldsymbol{A}$ 的 3 重特征值, 从而, 0 是 $p_{\boldsymbol{A}}(\lambda)$ 的根, 1 是 $p_{\boldsymbol{A}}(\lambda), p'_{\boldsymbol{A}}(\lambda), p''_{\boldsymbol{A}}(\lambda)$ 的根. 基于此观察, 我们在等式 (6.4) 中代入 $\lambda=0$ 和 $\lambda=1$, 对等式 (6.4) 两边依次求导和求二次导后代入 $\lambda=1$, 我们可以建立如下四个方程:

$$\begin{cases}0^{100}=p_{\boldsymbol{A}}(0)f(0)+r(0)=r(0)=d,\\ 1^{100}=p_{\boldsymbol{A}}(1)f(1)+r(1)=r(1)=a+b+c+d,\\ 100\cdot 1^{99}=p'_{\boldsymbol{A}}(1)f(1)+p_{\boldsymbol{A}}(1)f'(1)+r'(1)=r'(1)=3a+2b+c,\\ 100\cdot 99\cdot 1^{98}=p''_{\boldsymbol{A}}(1)f(1)+2p'_{\boldsymbol{A}}(1)f'(1)+p_{\boldsymbol{A}}(1)f''(1)+r''(1)=r''(1)=6a+2b.\end{cases}$$

可以解得 $a=4851$, $b=-9603$, $c=4753$, $d=0$, 从而 $r(\lambda)=4851\lambda^3-9603\lambda^2+4753\lambda$.

再由 Caylay–Hamilton 定理, 我们知 $p_{\boldsymbol{A}}(\boldsymbol{A})=\boldsymbol{O}$, 从而

$$\begin{aligned}\boldsymbol{A}^{100}=&p_{\boldsymbol{A}}(\boldsymbol{A})f(\boldsymbol{A})+r(\boldsymbol{A})\\ =&\boldsymbol{O}\,f(\boldsymbol{A})+r(\boldsymbol{A})=r(\boldsymbol{A})\\ =&4851\boldsymbol{A}^3-9603\boldsymbol{A}^2+4753\boldsymbol{A}\\ =&4851\begin{pmatrix}0&1&3&6\\0&1&3&6\\0&0&1&3\\0&0&0&1\end{pmatrix}-9603\begin{pmatrix}0&1&2&3\\0&1&2&3\\0&0&1&2\\0&0&0&1\end{pmatrix}+4753\begin{pmatrix}0&1&1&1\\0&1&1&1\\0&0&1&1\\0&0&0&1\end{pmatrix}\\ =&\begin{pmatrix}0&1&100&5050\\0&1&100&5050\\0&0&1&100\\0&0&0&1\end{pmatrix}.\end{aligned}$$

□

注 6.4.28. 当然, 在上题中, 我们也可以直接用归纳法证明

$$\boldsymbol{A}^n = \begin{pmatrix} 0 & 1 & n & \frac{n(n+1)}{2} \\ 0 & 1 & n & \frac{n(n+1)}{2} \\ 0 & 0 & 1 & n \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

6.5 若尔当标准形简介 (※)

在这儿, 我们将非常简要地介绍若尔当 (Jordan) 矩阵. 感兴趣的同学需要认真阅读教材上的相关内容.

设 λ 是一个复数, m 是一个正整数. 我们称如下矩阵

$$\boldsymbol{J}_m(\lambda) := \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}_{m\times m}$$

为一个 m **阶若尔当块**. 该矩阵的特征值为 m 重的 λ, 但是相应的几何重数为 1. 因此, 这样的矩阵可对角化的充要条件是 $m = 1$.

进一步地, 我们称由若尔当块构成的准对角的矩阵

$$\boldsymbol{J} = \mathrm{diag}(\boldsymbol{J}_{m_1}(\lambda_1), \boldsymbol{J}_{m_2}(\lambda_2), \ldots, \boldsymbol{J}_{m_s}(\lambda_s))$$

为一个**若尔当形矩阵**.

关于若尔当形矩阵的理论的核心是如下的结构定理.

定理 6.5.1. 任何一个复方阵 $\boldsymbol{A}$ 都相似于一个若尔当形矩阵 $\boldsymbol{J}$, 并且若不计若尔当块的排列顺序, 则这样的 $\boldsymbol{J}$ 是唯一的. 教材定理 6.5.1

该定理中的 $\boldsymbol{J}$ 被称为 $\boldsymbol{A}$ 的**若尔当标准形** (Jordan canonical form, Jordan's normal form). 所谓的不计排列顺序是指, 我们将 $\mathrm{diag}(\boldsymbol{J}_3(2), \boldsymbol{J}_1(2), \boldsymbol{J}_2(5))$ 与 $\mathrm{diag}(\boldsymbol{J}_1(2), \boldsymbol{J}_2(5), \boldsymbol{J}_3(2))$ 视为同一矩阵的若尔当标准形.

上面定理中的存在性与唯一性的证明都很复杂, 故这儿略去. 我们在实际运用中, 经常仅需要确定 $\boldsymbol{A}$ 的若尔当标准形中所有若尔当块的个数, 为此, 我们首先注意到

$$\mathrm{rank}(\boldsymbol{J}_m(0)^k) = \begin{cases} m-k, & \text{若 } k \le m, \\ 0, & \text{若 } k > m. \end{cases}$$

接下来设 $\boldsymbol{A}$ 的特征值为 $\lambda_1,\dots,\lambda_s$, 对于 $k\geq 0$, 记 $r_k^i := \operatorname{rank}\big((\boldsymbol{A}-\lambda_i\boldsymbol{I})^k\big)$, 并记 $d_k^i := r_{k-1}^i - r_k^i$, 则利用上面关于秩的计算 (以及其它的计算技巧) 可以证明, $\boldsymbol{A}$ 的若尔当标准形中 λ_i 的 k 阶若尔当块的个数为 $d_k^i - d_{k+1}^i$. 关于其具体证明, 可以看一下教材的 P186 中的讨论.

教材上给出了若尔当标准形的两个应用, 这儿我们再介绍一个应用.

命题 6.5.2. 若 $\boldsymbol{A}$ 是 n 阶复方阵, 则 $\boldsymbol{A}$ 相似于其转置矩阵 $\boldsymbol{A}^{\mathrm{T}}$.

证明. 由定理 6.5.1 可知, $\boldsymbol{A}$ 相似于某个若尔当标准形

$$\boldsymbol{J} = \operatorname{diag}(\boldsymbol{J}_{m_1}(\lambda_1), \boldsymbol{J}_{m_2}(\lambda_2), \dots, \boldsymbol{J}_{m_s}(\lambda_s)),$$

从而 $\boldsymbol{A}^{\mathrm{T}}$ 相似于

$$\boldsymbol{J}^{\mathrm{T}} = \operatorname{diag}(\boldsymbol{J}_{m_1}(\lambda_1)^{\mathrm{T}}, \boldsymbol{J}_{m_2}(\lambda_2)^{\mathrm{T}}, \dots, \boldsymbol{J}_{m_s}(\lambda_s)^{\mathrm{T}}).$$

接下来, 由于有命题 6.2.14(3), 我们只需验证, 对于 $i=1,2,\dots,s$, 有矩阵的相似 $\boldsymbol{J}_{m_i}(\lambda_i)^{\mathrm{T}} \sim \boldsymbol{J}_{m_i}(\lambda_i)$. 为此, 我们取 $\boldsymbol{T}_i = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}_{m_i\times m_i}$. 可直接验证, $\boldsymbol{T}_i^{-1}\boldsymbol{J}_{m_i}(\lambda_i)^{\mathrm{T}}\boldsymbol{T}_i = \boldsymbol{J}_{m_i}(\lambda_i)$.

当然, 利用上面提到的计算公式, 直接验证 $\boldsymbol{J}_{m_i}(\lambda_i)^{\mathrm{T}}$ 的若尔当标准形的仅含 $\boldsymbol{J}_{m_i}(\lambda_i)$ 这一个若尔当块, 也是可行的. □

习题 6.5.3. 若复方阵 $\boldsymbol{A}$ 相似于若尔当型矩阵 $\boldsymbol{J} = \operatorname{diag}(\boldsymbol{J}_{m_1}(\lambda_1), \boldsymbol{J}_{m_2}(\lambda_2), \dots, \boldsymbol{J}_{m_s}(\lambda_s))$, 证明与 $\boldsymbol{A}$ 乘法可交换的所有矩阵所构成的线性空间的维数*为

$$\sum_{\substack{i,j=1\\ \lambda_i=\lambda_j}}^{s} \min(m_i, m_j).$$

习题 6.5.4. 设 $\boldsymbol{A}$ 是一个 $n\times n$ 的实矩阵, 它的所有 (复) 特征值都是正的实数. 证明: 对于每个正整数 m, 都存在某个实的矩阵 $\boldsymbol{B}$, 满足 $\boldsymbol{B}^m = \boldsymbol{A}$. (提示: 若 a 是一个正实数, $b=\sqrt[m]{a}$ 是它的 m 次的正根, 首先验证 $\boldsymbol{J}_t(b)^m$ 与 $\boldsymbol{J}_t(a)$ 是相似的)

*若 $d_1(\lambda) \mid d_2(\lambda) \mid \cdots \mid d_s(\lambda)$ 是 $\lambda\boldsymbol{I}-\boldsymbol{A}$ 的所有不是 1 的不变因子, 则该维数也可以表示为 $\sum_{i=1}^{s}(2s-2i+1)\deg(d_i(\lambda))$. 这被称为 Frobenius 定理, 参见 Jacobson 的 Basic Algebra I, §3.11.

第七章 欧几里得空间

在熟知的三维 $\mathbb{R}^3$ 中, 我们可以建立直角坐标系 $Oxyz$. 对于任意的向量 $\boldsymbol{a} = (a_1, a_2, a_3)^{\mathrm{T}}$ 和 $\boldsymbol{b} = (b_1, b_2, b_3)^{\mathrm{T}} \in \mathbb{R}^3$, 我们有**向量的内积** (**点乘**, **标量积**)

$$\boldsymbol{a} \cdot \boldsymbol{b} \coloneqq a_1 b_1 + a_2 b_2 + a_3 b_3.$$

这儿, 我们将把内积记成 $(\boldsymbol{a}, \boldsymbol{b})$. 内积的几何意义是: 向量 $\boldsymbol{a}$ 的模长 $|\boldsymbol{a}|$ 为 $\sqrt{(\boldsymbol{a}, \boldsymbol{a})}$, 而且

$$(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a}| \cdot |\boldsymbol{b}| \cdot \cos(\theta),$$

其中的 θ 是向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 之间的夹角. 这一套处理方法可以推广到高维情形.

7.1 定义与基本性质

定义 7.1.1. *设 V 是实数域 $\mathbb{R}$ 上的线性空间. 如果对于 V 内任意两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 都按照某一法则对应于一个实数, 记作 $(\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}$,* 满足:*

(对称性) $(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{b}, \boldsymbol{a})$,

(对于第一位置的线性性) $\lambda(\boldsymbol{a}, \boldsymbol{b}) = (\lambda \boldsymbol{a}, \boldsymbol{b})$ *以及* $(\boldsymbol{a} + \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{a}, \boldsymbol{c}) + (\boldsymbol{b}, \boldsymbol{c})$,

(正定性) *对于任意的 $\boldsymbol{a} \in V$ 有 $(\boldsymbol{a}, \boldsymbol{a}) \geq 0$, 且不等式中等号成立的充要条件为 $\boldsymbol{a} = \boldsymbol{0}$,*

*则称该二元运算法则为 V 上的***内积** *(inner product), 而 $(\boldsymbol{a}, \boldsymbol{b})$ 称为 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的***内积***. 定义了内积的 $\mathbb{R}$ 上的线性空间称为***内积空间** *(inner product space) 或***欧几里得空间***, 简称为***欧氏空间***. 此时, 把线性空间 V 的维数叫作欧氏空间 V 的***维数***.*

注 7.1.2. *(1) 从内积的对称性和对于第一位置的线性性, 我们可以马上推出对于第二位置的线性性.*

*有些教材里将其记作 $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$, 用来与有序对的记号作区别; 这样的记号也许更为合理.

(2) 对于任意的 $\boldsymbol{a} \in V$, 我们有 $(\boldsymbol{a}, \mathbf{0}) = (\mathbf{0}, \boldsymbol{a}) = 0$.

(3) 同一个线性空间在不同的内积下将被视作不同的内积空间.

例 7.1.3 (内积空间的例子). (1) 对于数组空间 $\mathbb{R}^n$, 将里面的向量视为列向量. 任取其中两个向量 $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$, $\boldsymbol{b} = (b_1, \ldots, b_n)^{\mathrm{T}}$, 我们可以定义

$$(\boldsymbol{a}, \boldsymbol{b}) \coloneqq a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \boldsymbol{a}^{\mathrm{T}} \boldsymbol{b}.$$

可以验证, 这是一个内积, 称为 $\mathbb{R}^n$ 的**标准内积** (*standard inner product*) 或者**欧几里得内积**. 若不作特别说明的话, 我们默认 $\mathbb{R}^n$ 为采用了如此定义的内积空间.

当然, 对于给定的一个元素全为正的向量 $\boldsymbol{w} = (w_1, \ldots, w_n)^{\mathrm{T}}$, 我们也可以定义 $\mathbb{R}^n$ 上的一个新的内积为

$$(\boldsymbol{a}, \boldsymbol{b}) \coloneqq \sum_{i=1}^{n} a_i b_i w_i.$$

其中这些标量 w_i 称为这个新的内积的**权** (*weights*).

(2) 设 V 是 $\mathbb{R}$ 上的 n 维线性空间, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 为 V 的一组基. 对于任意的 $\boldsymbol{x}, \boldsymbol{y} \in V$, 设它们在 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的坐标分别为 $\boldsymbol{X} = (x_1, \ldots, x_n)^{\mathrm{T}}$ 和 $\boldsymbol{Y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, 即 $\boldsymbol{x} = \sum_i x_i \boldsymbol{\alpha}_i$, $\boldsymbol{y} = \sum_i y_i \boldsymbol{\alpha}_i$. 定义

$$(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \sum_i x_i y_i,$$

即 $\mathbb{R}^n$ 中的向量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的标准内积. 不难验证, 这给出了 V 上的一个内积.

(3) 由全体 $m \times n$ 实矩阵构成的线性空间 $\mathbb{R}^{m \times n}$ 中, 定义一个二元函数为

$$(\boldsymbol{A}, \boldsymbol{B}) \coloneqq \operatorname{tr}(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{B}).$$

容易验证, 该二元函数是 $\mathbb{R}^{m \times n}$ 上的一个内积.

(4) $C[a, b]$ 为有界闭区间 $[a, b]$ 上的实连续函数的全体, 是 $\mathbb{R}$ 上的无穷维线性空间. 这儿假定 $a < b$. 对于 $f, g \in C[a, b]$, 定义

$$(f, g) \coloneqq \int_a^b f(x) g(x) \, \mathrm{d}x.$$

这给出了 $C[a, b]$ 的一个内积. 为了验证正定性, 我们注意到

$$(f, f) = \int_a^b f(x)^2 \, \mathrm{d}x \geq 0.$$

若存在 $[a,b]$ 中某个 x_0 满足 $f(x_0)\neq 0$, 那么由函数 f^2 的连续性可知, 存在 $[a,b]$ 中包含 x_0 的小区间 I, 使得在 I 上有 $f^2 \geq \frac{f(x_0)^2}{2}$. 若令 I 的长度为 $p>0$, 那么

$$(f,f)=\int_a^b f(x)^2\,\mathrm{d}x \geq \int_I f(x)^2\,\mathrm{d}x \geq \frac{f(x_0)^2 p}{2} > 0.$$

这说明若 $(f,f)=0$, 则 f 在 $[a,b]$ 上必为零函数.

更一般地, 若 $w(x)\in C[a,b]$ 是一个正的连续函数, 那么

$$(f,g) \coloneqq \int_a^b f(x)g(x)w(x)\,\mathrm{d}x$$

给出了 $C[a,b]$ 的一个新的内积, 而 $w(x)$ 称为该内积的**权函数** (*weight function*).

(5) 对于 $F=\mathbb{R}$, 我们考虑线性空间 $F_n[x]$. 这是次数不超过 n 的实系数多项式构成的实线性空间. 设 $a_0,a_1,\ldots,a_n$ 为不同的实数, 对于 $f,g\in F_n[x]$, 定义

$$(f,g) \coloneqq \sum_{i=0}^n f(a_i)g(a_i).$$

这给出了 $F_n[x]$ 的一个内积. 为了验证正定性, 我们注意到

$$(f,f)=\sum_{i=0}^n f(a_i)^2 \geq 0.$$

若 $(f,f)=0$, 则 $a_0,a_1,\ldots,a_n$ 为 $f(x)$ 的不同的根. 而 $f(x)$ 的次数不超过 n, 这说明 f 必为零多项式.

更一般地, 若 $w_0,w_1,\ldots,w_n$ 为一组正数, 那么

$$(f,g) \coloneqq \sum_{i=0}^n f(a_i)g(a_i)w_i$$

也定义了 $F_n[x]$ 上的一个内积.

定理 7.1.4 (Cauchy–Schwarz)**.** 设 V 是欧氏空间, $(\cdot,\cdot)$ 是 V 的内积, 则对 V 中的任意两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$, 有 $|(\boldsymbol{a},\boldsymbol{b})| \leq \sqrt{(\boldsymbol{a},\boldsymbol{a})(\boldsymbol{b},\boldsymbol{b})}$.

证明. 不妨设 $\boldsymbol{a}\neq \boldsymbol{0}$, 从而 $(\boldsymbol{a},\boldsymbol{a})\neq 0$. 对于任意实数 λ, 有

$$0 \leq (\lambda\boldsymbol{a}+\boldsymbol{b},\lambda\boldsymbol{a}+\boldsymbol{b}) = (\boldsymbol{a},\boldsymbol{a})\lambda^2 + 2(\boldsymbol{a},\boldsymbol{b})\lambda + (\boldsymbol{b},\boldsymbol{b}).$$

右式是关于 λ 的实系数二次多项式, 故其判别式非正:

$$4(\boldsymbol{a},\boldsymbol{b})^2 - 4(\boldsymbol{a},\boldsymbol{a})(\boldsymbol{b},\boldsymbol{b}) \leq 0.$$

化简即可. □

注 7.1.5. (1) 由证明过程易知, *Cauchy–Schwarz* 不等式中等号成立的条件是向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 线性相关.

(2) 将 *Cauchy–Schwarz* 不等式运用于例 *7.1.3* (1), 我们得到, 若 $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$, 则

$$|a_1b_1 + a_2b_2 + \cdots + a_nb_n| \leq \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}\sqrt{b_1^2 + b_2^2 + \cdots + b_n^2}.$$

将 *Cauchy–Schwarz* 不等式运用于例 *7.1.3* (4), 我们得到, 若 $f(x), g(x) \in C[a, b]$, 则

$$\left|\int_a^b f(x)g(x)\,\mathrm{d}x\right| \leq \sqrt{\int_a^b f(x)^2\,\mathrm{d}x}\sqrt{\int_a^b g(x)^2\,\mathrm{d}x}.$$

定义 7.1.6. 设 V 是欧氏空间, $(\cdot, \cdot)$ 是 V 的内积, $\boldsymbol{a}, \boldsymbol{b} \in V$. 仿照三维的情形, 我们定义如下.

(1) $|\boldsymbol{a}| := \sqrt{(\boldsymbol{a}, \boldsymbol{a})} \geq 0$ 称为 $\boldsymbol{a}$ 的**长度** (*length*) 或**模** (*module*). 有时模也记作 $\|\boldsymbol{a}\|$.* 由内积的正定性容易推出, $\|\boldsymbol{a}\| = 0$ 当且仅当 $\boldsymbol{a} = \boldsymbol{0}$. 另外, 对任意的 $\lambda \in \mathbb{R}$, 我们都有 $\|\lambda\boldsymbol{a}\| = |\lambda| \cdot \|\boldsymbol{a}\|$.

(2) 两个非零向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 之间的**夹角**为 $\theta = \arccos\left(\frac{(\boldsymbol{a}, \boldsymbol{b})}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|}\right)$. 注意, 由 *Cauchy–Schwarz* 不等式, 我们有

$$-1 \leq \frac{(\boldsymbol{a}, \boldsymbol{b})}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|} \leq 1,$$

从而定义是可行的.

(3) 进一步地, 若 $(\boldsymbol{a}, \boldsymbol{b}) = 0$ (允许 $\boldsymbol{a}, \boldsymbol{b}$ 为 $\boldsymbol{0}$), 则称 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ **垂直** (*perpendicular*) 或**正交** (*orthogonal*), 记作 $\boldsymbol{a} \perp \boldsymbol{b}$. 容易验证, $\boldsymbol{a} \perp \boldsymbol{b}$ 当且仅当 $\boldsymbol{b} \perp \boldsymbol{a}$, 当且仅当**勾股定理** (也被称为**毕达哥拉斯定理**) 成立:

$$\|\boldsymbol{a} + \boldsymbol{b}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2.$$

另外, 零向量 $\boldsymbol{0}$ 与 V 中任意的向量都正交, 并且是 V 中满足这一条件的唯一的向量, 也是 V 中唯一的与自身正交的向量.

(4) 长度为 1 的向量称为**单位向量**.

(5) 向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 之间的**距离**定义为 $d(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|$. 我们很容易验证:

*为了与标量的绝对值以示区别, 在我们的教案里会采用这一记号.

(对称性) $d(\boldsymbol{\alpha},\boldsymbol{\beta})=d(\boldsymbol{\beta},\boldsymbol{\alpha})$,

(正定性) $d(\boldsymbol{\alpha},\boldsymbol{\beta})\geq 0$, 且 $d(\boldsymbol{\alpha},\boldsymbol{\beta})=0$ 当且仅当 $\boldsymbol{\alpha}=\boldsymbol{\beta}$,

(三角不等式) $d(\boldsymbol{\alpha},\boldsymbol{\beta})\leq d(\boldsymbol{\alpha},\boldsymbol{\gamma})+d(\boldsymbol{\beta},\boldsymbol{\gamma})$.

容易看出,

$$
\begin{aligned}
\text{关于距离的三角不等式} &\xLeftrightarrow[\boldsymbol{b}=\boldsymbol{\gamma}-\boldsymbol{\beta}]{\text{记 } \boldsymbol{a}=\boldsymbol{\alpha}-\boldsymbol{\gamma}} \|\boldsymbol{a}+\boldsymbol{b}\|\leq\|\boldsymbol{a}\|+\|\boldsymbol{b}\| \quad (\text{模的三角不等式})\\
&\Leftrightarrow \|\boldsymbol{a}+\boldsymbol{b}\|^2\leq(\|\boldsymbol{a}\|+\|\boldsymbol{b}\|)^2\\
&\Leftrightarrow (\boldsymbol{a}+\boldsymbol{b},\boldsymbol{a}+\boldsymbol{b})\leq\|\boldsymbol{a}\|^2+2\|\boldsymbol{a}\|\cdot\|\boldsymbol{b}\|+\|\boldsymbol{b}\|^2\\
&\Leftrightarrow (\boldsymbol{a},\boldsymbol{a})+(\boldsymbol{a},\boldsymbol{b})+(\boldsymbol{b},\boldsymbol{a})+(\boldsymbol{b},\boldsymbol{b})\leq(\boldsymbol{a},\boldsymbol{a})+2\|\boldsymbol{a}\|\cdot\|\boldsymbol{b}\|+(\boldsymbol{b},\boldsymbol{b})\\
&\Leftrightarrow (\boldsymbol{a},\boldsymbol{b})\leq\|\boldsymbol{a}\|\cdot\|\boldsymbol{b}\|. \quad (\text{Cauchy–Schwarz 不等式})
\end{aligned}
$$

习题 7.1.7. 证明 $\mathbb{R}^n$ 中的向量 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 满足**平行四边形法则**:

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2+\|\boldsymbol{u}-\boldsymbol{v}\|^2=2\|\boldsymbol{u}\|^2+2\|\boldsymbol{v}\|^2.$$

习题 7.1.8. 设 4 维实的列向量 $\boldsymbol{\alpha}$ 的长度为 5, 求 $|\boldsymbol{I}_4-\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}}|$ 的值.

7.2 内积的表示与标准正交基

回忆. 设 V 是有限维向量空间, $\mathscr{A}$ 为 V 上的线性变换, 在基 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}$, 则 $\mathscr{A}$ 的具体作用由 $\boldsymbol{A}$ 唯一决定. (定理 6.2.3)

若 V 是有限维欧氏空间, 对于其上的内积, 我们也有类似地表述.

设 $(\cdot,\cdot)$ 是 V 上的内积, 任给 $\boldsymbol{x},\boldsymbol{y}\in V$, 假定它们在基 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 下的坐标分别为 $\boldsymbol{X}=(x_1,\ldots,x_n)^{\mathrm{T}}$ 和 $\boldsymbol{Y}=(y_1,\ldots,y_n)^{\mathrm{T}}$, 我们有

$$(\boldsymbol{x},\boldsymbol{y})=\left(\sum_{i=1}^n x_i\boldsymbol{\alpha}_i,\sum_{j=1}^n y_j\boldsymbol{\alpha}_j\right)\xlongequal{\text{双线性性}}\sum_{i=1}^n\sum_{j=1}^n x_iy_j(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j).$$

令 $\boldsymbol{G}=(g_{ij})\in\mathbb{R}^{n\times n}$, 其中 $g_{ij}=(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j)$. 则上式可以表述成

$$(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Y}.$$

注意, 由内积的对称性, 有 $(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j)=(\boldsymbol{\alpha}_j,\boldsymbol{\alpha}_i)$, 即 $g_{ij}=g_{ji}$. 这说明方阵 $\boldsymbol{G}$ 是实对称阵. 在接下来的讨论中, 方阵 $\boldsymbol{G}$ 将被称为内积 $(\cdot,\cdot)$ 在基 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 下的**度量矩阵**.

一般地, 对于 n 阶实对称阵 $\boldsymbol{A}$, 若是对于任意的非零列向量 $\boldsymbol{X}\in\mathbb{R}^n$, 皆有 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}>0$, 则称 $\boldsymbol{A}$ 是一个**正定方阵**. 故由内积的正定性, 知其度量矩阵 $\boldsymbol{G}$ 必为一个正定方阵.

反过来, 给定 $\mathbb{R}$ 上的有限维线性空间 V 和上面的一组基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$, 以及一个 n 阶正定的实对称方阵 $\boldsymbol{G}$, 可以定义 V 上的一个内积如下. 任给 $\boldsymbol{x},\boldsymbol{y}\in V$, 假定它们在基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 下的坐标分别为 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$, 定义 $(\boldsymbol{x},\boldsymbol{y})\coloneqq\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Y}$. 则 V 变成了一个内积空间.

习题 7.2.1. 我们考虑线性空间 $V=\mathbb{R}_4[x]$, 这是次数不超过 4 的实系数多项式构成的实线性空间. 设 a_0,a_1,a_2,a_3,a_4 依次为 $-2,-1,0,1,2$. 对于 $f,g\in\mathbb{R}_4[x]$, 定义

$$(f,g)\coloneqq\sum_{i=0}^{4}f(a_i)g(a_i).$$

这给出了 V 的一个内积. 求它在自然基 $1,x,x^2,x^3,x^4$ 下的度量矩阵.

现在出现了一个很自然的问题: 在考虑线性空间 V 的一组新的基时, 线性变换相应的矩阵要做相似变换. 那么对于内积而言, 相应的度量矩阵该如何改变?

引理 7.2.2. 设 $\boldsymbol{A},\boldsymbol{B}\in F^{n\times n}$. 则 $\boldsymbol{A}=\boldsymbol{B}$ 当且仅当对于任意的 n 维列向量 $\boldsymbol{X},\boldsymbol{Y}$, 有 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Y}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Y}$.

证明. 必要性是显然的. 对于充分性, 只需选取数组空间的标准基中的向量: 取 $\boldsymbol{X}=\boldsymbol{e}_i$, $\boldsymbol{Y}=\boldsymbol{e}_j$, 那么 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Y}$ 是 $\boldsymbol{A}$ 的 (i,j) 元素, 而 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Y}$ 是 $\boldsymbol{B}$ 的 (i,j) 元素. □

设 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 是欧氏空间 V 的另外一组基. 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 到 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 的过渡矩阵为 $\boldsymbol{P}$, 即有

$$(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n)\boldsymbol{P}.$$

设内积在 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 和 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 下的度量矩阵分别为 $\boldsymbol{G},\widetilde{\boldsymbol{G}}$. 对于任意的 $\boldsymbol{x},\boldsymbol{y}\in V$, 假设它们在基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 下的坐标分别为 $\boldsymbol{X},\boldsymbol{Y}$. 从而, 它们在基 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 下的坐标分别为 $\widetilde{\boldsymbol{X}}=\boldsymbol{P}^{-1}\boldsymbol{X}$ 和 $\widetilde{\boldsymbol{Y}}=\boldsymbol{P}^{-1}\boldsymbol{Y}$. 由定义, 有

$$\boxed{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Y}}=(\boldsymbol{\alpha},\boldsymbol{\beta})=\widetilde{\boldsymbol{X}}^{\mathrm{T}}\widetilde{\boldsymbol{G}}\widetilde{\boldsymbol{Y}}=\boxed{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}})^{-1}\widetilde{\boldsymbol{G}}\boldsymbol{P}^{-1}\boldsymbol{Y}}.$$

由上面的引理可知, 这说明 $\boldsymbol{G}=(\boldsymbol{P}^{\mathrm{T}})^{-1}\widetilde{\boldsymbol{G}}\boldsymbol{P}^{-1}$, 即

$$\widetilde{\boldsymbol{G}}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{P}.$$

定义 7.2.3. 给定两个同阶方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$. 若存在可逆矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{A}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}$, 则称 $\boldsymbol{A}$ 是 $\boldsymbol{B}$ 由**相合变换** (congruent transformation) 得到的, 称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 是**相合**的.

注 7.2.4. (1) 容易验证相合关系也是一个等价关系, 满足反身性, 对称性以及传递性.

(2) 上面的讨论说明内积在不同基下的矩阵是相合的.

(3) 用正交矩阵来作相似变换与相合变换, 效果是一样的.

(4) 我们一般是在实方阵的框架下讨论相合. 不难看出, 定义中的 $\boldsymbol{A}$ 是实对称的, 当且仅当 $\boldsymbol{B}$ 是实对称的.

习题 7.2.5. 设 $\mathbb{R}_2[x]$ 中的某个内积在基 $\boldsymbol{a}_1 = x-1$, $\boldsymbol{a}_2 = x+1$, $\boldsymbol{a}_3 = x^2$ 下的度量矩阵为 $\begin{pmatrix} 2 & 2 & 1 \\ 2 & 5 & 0 \\ 1 & 0 & 1 \end{pmatrix}$, 则该内积在基 $\boldsymbol{b}_1 = 2x$, $\boldsymbol{b}_2 = -x+1$, $\boldsymbol{b}_3 = x^2+2$ 下的度量矩阵是什么?

定义 7.2.6. 在欧氏空间 V 中, 由两两正交的非零向量构成的向量组称为一个**正交向量组**, 由正交向量组构成的基称为一组**正交基**, 由单位向量构成的正交基称为一组**标准正交基** (*orthonormal basis*), **单位正交基**或**规范正交基**.

例 7.2.7. (1) $\mathbb{R}^n$ 的自然基在标准内积下是最常用的一组标准正交基.

(2) $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^{\mathrm{T}}$, $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)^{\mathrm{T}}$, $(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}})^{\mathrm{T}}$ 是 $\mathbb{R}^3$ 的一组标准正交基.

(3) 在 $C[-\pi, \pi]$ 上定义内积

$$(f, g) := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) g(x) \, \mathrm{d}x.$$

对于任意给定的正整数 n, 不难验证集合

$$\left\{ \frac{1}{\sqrt{2}}, \cos(x), \cos(2x), \ldots, \cos(nx), \sin(x), \sin(2x), \ldots, \sin(nx) \right\}$$

是一个由单位向量构成的正交向量组. 由于已知 $C[-\pi, \pi]$ 是实数域上的无穷维线性空间, 故这组向量不可能构成 $C[-\pi, \pi]$ 的标准正交基.

注 7.2.8. 欧氏空间中的内积在标准正交基下的度量矩阵为单位阵.

命题 7.2.9. 若 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_r$ 是欧氏空间的一个正交向量组, 那么它一定是线性无关的, 从而构成所生成的子空间的一组基.

证明. 由条件可知, 对于任意的 $i \neq j$, 有 $(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j) = 0$. 若 $\sum_{i=1}^{r} \lambda_i \boldsymbol{\alpha}_i = \mathbf{0}$, 则对于任意的 j, 有

$$0 = \left(\sum_{i=1}^{r} \lambda_i \boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j \right) = \sum_{i=1}^{r} \lambda_i (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j) = \lambda_j (\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j) = \lambda_j \|\boldsymbol{\alpha}_j\|^2.$$

因为 $\boldsymbol{\alpha}_j$ 不是零向量 (正交向量组的定义中的特别要求), $\|\boldsymbol{\alpha}_j\| \neq 0$, 从而 $\lambda_j = 0$ 对于任意的 j 成立. 这说明该向量组是线性无关的. □

推论 7.2.10. 在 n 维欧氏空间中两两正交的非零向量的个数不会超过 n 个.

注 7.2.11. 若 $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ 是 V 的一组标准正交基, 向量 $\boldsymbol{x}$ 在这组基下的坐标为 $(x_1, \ldots, x_n)^{\mathrm{T}}$, 即有 $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{\varepsilon}_i$. 将该式与 $\boldsymbol{\varepsilon}_j$ 作内积, 我们得到

$$(\boldsymbol{x}, \boldsymbol{\varepsilon}_j) = \left(\sum_{i=1}^{n} x_i \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j \right) = \sum_{i=1}^{n} x_i (\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = x_j.$$

这说明向量坐标的分量可以用内积表达:

$$\boldsymbol{x} = \sum_{i=1}^{n} (\boldsymbol{x}, \boldsymbol{\varepsilon}_i) \boldsymbol{\varepsilon}_i.$$

该式称为 $\boldsymbol{x}$ 的 ***Fourier* 展开**, 其中每个系数 $(\boldsymbol{x},\boldsymbol{\varepsilon}_i)$ 都称为 $\boldsymbol{x}$ 的 ***Fourier* 系数**. 设向量 $\boldsymbol{y}\in V$ 在这组标准正交基下的坐标为 $(y_1,\dots,y_n)^{\mathrm{T}}$, 即有 $\boldsymbol{y}=\sum_{j=1}^n y_j\boldsymbol{\varepsilon}_j$. 此时,

$$\boxed{(\boldsymbol{x},\boldsymbol{y})}=\left(\sum_{i=1}^n x_i\boldsymbol{\varepsilon}_i,\sum_{j=1}^n x_y\boldsymbol{\varepsilon}_j\right)=\sum_{i=1}^n\sum_{j=1}^n x_iy_j(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_j)=\sum_{i=1}^n\sum_{j=1}^n x_iy_j\delta_{ij}=\boxed{\textstyle\sum_{i=1}^n x_iy_i}.$$

特别地, 我们有**帕塞瓦尔等式**

$$(\boldsymbol{x},\boldsymbol{x})=\sum_{i=1}^n x_i^2,$$

即向量的模满足**勾股定理**:

$$\|\boldsymbol{x}\|=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}.$$

既然标准正交基有这么好的性质, 它是否一定存在呢?

注 7.2.12. 向量组 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\in V$ 构成 n 维欧氏空间 V 的标准正交基的充要条件是对于任意的 i,j, 有 $(\boldsymbol{x}_i,\boldsymbol{x}_j)=\delta_{ij}$. 回忆一下, 这是 Kronecker 符号, 是这样定义的:

$$\delta_{ij}=\begin{cases}1, & \text{当 } i=j,\\ 0, & \text{当 } i\neq j.\end{cases}$$

特别地, 例 *7.1.3* 的 (2) 表明, 相对于有限维实线性空间 V 的任意一组基 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$, V 上都存在一个相应的内积, 使得这组基在该内积下成为一组标准正交基.

接下来我们说明, 有限维的欧氏空间在给定的内积下, 也必然存在一组标准正交基.

回忆. 有限维线性空间的任意一组线性无关的向量都可以扩充为向量空间的一组基.

类似地, 有限维欧氏空间的任意一个正交向量组都可以扩充为该欧氏空间的一组正交基. 为此, 我们打算证明如下的定理.

定理 7.2.13. 从 n 维欧氏空间 V 的任意一组基出发, 可以构造出一组标准正交基. 特别地, V 的标准正交基一定存在.

为了证明上面的结果, 我们需要先做一点准备.

讨论. (1) 若向量 $\boldsymbol{\alpha}\in V$ 非零, 对其做**归一化** (也称作**标准化**或**单位化**), 可以得到单位向量:

$$\boldsymbol{\alpha}\mapsto\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}.$$

新得到的向量, 从几何意义上来说, 与原向量具有相同的方向.

(2) 在欧氏空间 V 中任意取定一个非零向量 $\boldsymbol{\alpha}$, 对于 V 的中任意向量 $\boldsymbol{\beta}$, 我们希望找到

正交分解 $\boldsymbol{\beta}=\boldsymbol{\beta}_1+\boldsymbol{\beta}_2$, 其中 $\boldsymbol{\beta}_1$ 与 $\boldsymbol{\alpha}$ 平行, 而 $\boldsymbol{\beta}_2$ 与 $\boldsymbol{\alpha}$ 垂直.

为了找到这样的分解, 我们不妨设 $\boldsymbol{\beta}_1=k\boldsymbol{\alpha}$, $k\in\mathbb{R}$. 由于 $\boldsymbol{\beta}_2=\boldsymbol{\beta}-\boldsymbol{\beta}_1$ 与 $\boldsymbol{\alpha}$ 垂直, 我们有

$$0=(\boldsymbol{\beta}_2,\boldsymbol{\alpha})=(\boldsymbol{\beta},\boldsymbol{\alpha})-k(\boldsymbol{\alpha},\boldsymbol{\alpha}).$$

这说明 $k=\frac{(\boldsymbol{\beta},\boldsymbol{\alpha})}{(\boldsymbol{\alpha},\boldsymbol{\alpha})}$, 从而

$$\boldsymbol{\beta}_1=\frac{(\boldsymbol{\beta},\boldsymbol{\alpha})}{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\boldsymbol{\alpha},\qquad \boldsymbol{\beta}_2=\boldsymbol{\beta}-\frac{(\boldsymbol{\beta},\boldsymbol{\alpha})}{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\boldsymbol{\alpha}.$$

不难直接验证, 这样的分解符合要求, 并且满足要求的分解是唯一的. 这样得到的 k 被称为称为 $\boldsymbol{\beta}$ 到向量 $\boldsymbol{\alpha}$ 上的**标量投影** (*scalar projection*), 而 $\boldsymbol{\beta}_1$ 被称为 $\boldsymbol{\beta}$ 到向量 $\boldsymbol{\alpha}$ 上的**向量投影** (*vector projection*), 或简称为**投影**. 而 $\boldsymbol{\beta}_2$ 将被称作 $\boldsymbol{\beta}$ 的与 $\boldsymbol{\alpha}$ 正交的分量.

(3) 若前面讨论中的向量 $\boldsymbol{\alpha},\boldsymbol{\beta}$ 线性无关, 那么 $\boldsymbol{\beta}_2=\boldsymbol{\beta}-k\boldsymbol{\alpha}\neq\mathbf{0}$, 从而我们得到了正交向量组 $\boldsymbol{\alpha},\boldsymbol{\beta}_2$.

(4) 更一般地, 设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m$ 是欧氏空间 V 中的一个正交向量组, 对于任意的 $\boldsymbol{\beta}\in V$, 我们可以考虑分解 $\boldsymbol{\beta}=\boldsymbol{\beta}_1+\boldsymbol{\beta}_2$, 其中

$$\boldsymbol{\beta}_1=\sum_{i=1}^m\frac{(\boldsymbol{\beta},\boldsymbol{\alpha}_i)}{(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i)}\boldsymbol{\alpha}_i,\qquad \boldsymbol{\beta}_2=\boldsymbol{\beta}-\boldsymbol{\beta}_1.$$

显然, $\boldsymbol{\beta}_1$ 属于线性子空间 $\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$, 并且对于任意的 $\boldsymbol{\gamma}:=\sum_{i=1}^m k_i\boldsymbol{\alpha}_i\in\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$, 我们有

$$\begin{aligned}(\boldsymbol{\beta}_2,\boldsymbol{\gamma})&=\left(\boldsymbol{\beta}-\sum_{i=1}^m\frac{(\boldsymbol{\beta},\boldsymbol{\alpha}_i)}{(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i)}\boldsymbol{\alpha}_i,\sum_{j=1}^m k_j\boldsymbol{\alpha}_j\right)\\&=\sum_{j=1}^m k_j(\boldsymbol{\beta},\boldsymbol{\alpha}_j)-\underbrace{\sum_{i=1}^m\sum_{j=1}^m k_j\frac{(\boldsymbol{\beta},\boldsymbol{\alpha}_i)}{(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i)}(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j)}_{\text{当 } j\neq i \text{ 时 } (\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j)=0}\\&=\sum_{j=1}^m k_j(\boldsymbol{\beta},\boldsymbol{\alpha}_j)-\sum_{i=1}^m k_i\frac{(\boldsymbol{\beta},\boldsymbol{\alpha}_i)}{(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i)}(\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_i)\\&=0.\end{aligned}$$

这说明上面构造的 $\boldsymbol{\beta}_2$ 与子空间 $\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$ 中的所有向量都正交. 由此, 我们称 $\boldsymbol{\beta}_1$ 为向量 $\boldsymbol{\beta}$ 到子空间 $\langle\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_m\rangle$ 的**向量投影**.

特别地, 若 $\boldsymbol{\beta} \notin \langle \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m \rangle$, 那么 $\boldsymbol{\beta}_2 \neq \mathbf{0}$, 从而得到更大的正交向量组 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m, \boldsymbol{\beta}_2$.

有了上面的准备, 我们首先来看 **Gram–Schmidt 正交化过程** (orthogonalization).

定理 7.2.14. 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_s$ 是欧氏空间 V 的一个线性无关的向量组, 令

$$\begin{cases} \boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1, \\ \boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - \frac{(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1)} \boldsymbol{\beta}_1, \\ \quad \vdots \\ \boldsymbol{\beta}_s = \boldsymbol{\alpha}_s - \sum_{j=1}^{s-1} \frac{(\boldsymbol{\alpha}_s, \boldsymbol{\beta}_j)}{(\boldsymbol{\beta}_j, \boldsymbol{\beta}_j)} \boldsymbol{\beta}_j, \end{cases}$$

则 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_s$ 是一个正交向量组, 并且与 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_s$ 等价.

证明. 对于 s 作归纳法. $s = 1$ 时显然成立. 假设 $s = k$ 时命题成立, 现在来看 $s = k + 1$ 的情形. 此时,

$$\boldsymbol{\beta}_{k+1} = \boldsymbol{\alpha}_{k+1} - \sum_{j=1}^{k} \frac{(\boldsymbol{\alpha}_{k+1}, \boldsymbol{\beta}_j)}{(\boldsymbol{\beta}_j, \boldsymbol{\beta}_j)} \boldsymbol{\beta}_j.$$

由归纳假设可知, 向量组 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ 可由向量组 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ 线性表示. 从而由上面的等式可知, $\boldsymbol{\beta}_{k+1}$ 可以由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_k, \boldsymbol{\alpha}_{k+1}$ 线性表示. 当然, 上面等式的简单变形也指出 $\boldsymbol{\alpha}_{k+1}$ 可以由 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_k, \boldsymbol{\beta}_{k+1}$ 线性表示. 从而由归纳假设, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{k+1}$ 与 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{k+1}$ 是等价的. 特别地, $\boldsymbol{\beta}_{k+1} \neq \mathbf{0}$.

另外, 在前面 (4) 中的讨论说明, $\boldsymbol{\beta}_{k+1}$ 与 $\boldsymbol{\beta}_i$ $(i = 1, 2, \ldots, k)$ 正交. 由归纳假设, 这说明 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{k+1}$ 构成一个正交向量组. □

若将归一化的方法结合在一起, 我们就可以完成预定证明的证明了.

定理 *7.2.13* 的证明. 设 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 是 n 维欧氏空间 V 的一组基. 由上面的 Gram–Schmidt 正交化过程, 我们得到 V 的一组正交基 $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$. 对于每个 i, 令 $\boldsymbol{\varepsilon}_i = \frac{\boldsymbol{\beta}_i}{\|\boldsymbol{\beta}_i\|}$. 则 $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ 构成了 V 的一组标准正交基. □

注 7.2.15. 若是将 *Gram–Schmidt* 正交化过程与归一化过程结合在一起, 则算法大致如

下

$$
\begin{tikzcd}
\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1 \arrow[r] & \boldsymbol{\varepsilon}_1 = \boldsymbol{\beta}_1/\|\boldsymbol{\beta}_1\| \arrow[dl] \\
\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - (\boldsymbol{\alpha}_2, \boldsymbol{\varepsilon}_1)\boldsymbol{\varepsilon}_1 \arrow[r] & \boldsymbol{\varepsilon}_2 = \boldsymbol{\beta}_2/\|\boldsymbol{\beta}_2\| \arrow[dl] \\
\boldsymbol{\beta}_3 = \boldsymbol{\alpha}_3 - (\boldsymbol{\alpha}_3, \boldsymbol{\varepsilon}_1)\boldsymbol{\varepsilon}_1 - (\boldsymbol{\alpha}_3, \boldsymbol{\varepsilon}_2)\boldsymbol{\varepsilon}_2 \arrow[r] & \boldsymbol{\varepsilon}_3 = \boldsymbol{\beta}_3/\|\boldsymbol{\beta}_3\| \arrow[dl, dotted] \\
\boldsymbol{\beta}_n = \boldsymbol{\alpha}_n - \sum_{j=1}^{n-1}(\boldsymbol{\alpha}_s, \boldsymbol{\varepsilon}_j)\boldsymbol{\varepsilon}_j \arrow[r] & \boldsymbol{\varepsilon}_n = \boldsymbol{\beta}_n/\|\boldsymbol{\beta}_n\|
\end{tikzcd}
$$

例 7.2.16. 用 *Gram–Schmidt* 正交化方法把 $\mathbb{R}^3$ 的一组基

$$
\boldsymbol{\alpha}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\alpha}_2 = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix}, \quad \boldsymbol{\alpha}_3 = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}
$$

化成一组标准正交基.

解.

$$
\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \xrightarrow{\text{由于 } \|\boldsymbol{\beta}_1\| = \sqrt{3}} \boldsymbol{\varepsilon}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},
$$

$$
\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - (\boldsymbol{\alpha}_2, \boldsymbol{\varepsilon}_1)\boldsymbol{\varepsilon}_1 = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix} - \frac{-2}{\sqrt{3}} \left(\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) = \frac{1}{3} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix},
$$

由于 $\|\boldsymbol{\beta}_2\| = \frac{\sqrt{6}}{3}$,

$$
\boldsymbol{\varepsilon}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}.
$$

$$
\begin{aligned}
\boldsymbol{\beta}_3 =& \boldsymbol{\alpha}_3 - (\boldsymbol{\alpha}_3, \boldsymbol{\varepsilon}_1)\boldsymbol{\varepsilon}_1 - (\boldsymbol{\alpha}_3, \boldsymbol{\varepsilon}_2)\boldsymbol{\varepsilon}_2 \\
=& \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} - \frac{4}{\sqrt{3}} \left(\frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) - \frac{2}{\sqrt{6}} \left(\frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \right) = \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}.
\end{aligned}
$$

由于 $\|\boldsymbol{\beta}_3\| = 2\sqrt{2}$,

$$\boldsymbol{\varepsilon}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\0\\1\end{pmatrix}.$$

则 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3$ 是所需的标准正交基. $\square$

类似于上面的正交化的计算方法, 必须熟练掌握.

习题 7.2.17. 用 *Gram–Schmidt* 正交化方法把 $\mathbb{R}^4$ 的一组向量

$$\boldsymbol{\alpha}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \quad \boldsymbol{\alpha}_2 = \begin{pmatrix}0\\1\\1\\1\end{pmatrix}, \quad \boldsymbol{\alpha}_3 = \begin{pmatrix}0\\0\\1\\1\end{pmatrix}$$

化成一组标准正交向量.

习题 7.2.18. 用 *Gram–Schmidt* 正交化方法把 $\mathbb{R}^4$ 的一组向量

$$\boldsymbol{\alpha}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \quad \boldsymbol{\alpha}_2 = \begin{pmatrix}0\\1\\1\\1\end{pmatrix}, \quad \boldsymbol{\alpha}_3 = \begin{pmatrix}0\\0\\1\\1\end{pmatrix}$$

化成一组标准正交向量.

习题 7.2.19. 考虑线性空间 $V = \mathbb{R}_2[x]$, 运算为多项式的加法和数乘. 对于 $f(x) = a_0 + a_1x + a_2x^2$ 和 $g(x) = b_0 + b_1x + b_2x^2$, 定义 $(f, g) = \int_0^1 f(x)g(x)\,\mathrm{d}x$, 则 $(V, (\ ,\ ))$ 为欧氏空间. 用 *Schmidt* 正交化方法将 $1, x, x^2$ 按顺序改造成标准正交基.

接下来, 我们看一下 Gram–Schmidt 正交化的几个理论性的应用.

定义 7.2.20. 对于欧氏空间 V 中的向量 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$, 令 $\boldsymbol{G} = (g_{ij})_{m\times m}$, 其中 $g_{ij} = (\boldsymbol{x}_i, \boldsymbol{x}_j)$. 这样得到的对称矩阵 $\boldsymbol{G}$ 称为向量组 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ 的 ***Gram* 矩阵**.

注 7.2.21. (1) 内积在 V 的基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的度量矩阵是该向量组的 *Gram* 矩阵.

(2) 由定义不难看出, 向量组 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ 是一组由单位向量构成的正交组的充要条件是对于任意的 i, j, 有 $(\boldsymbol{x}_i, \boldsymbol{x}_j) = \delta_{ij}$. 用这儿的语言表述, 这等价于说, 它们的 *Gram* 矩阵为 m 阶的单位阵 $\boldsymbol{I}_m$.

(3) 若 V 是具有标准内积的欧氏空间 $\mathbb{R}^n$, 设 $\boldsymbol{X}$ 是以 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ 为列向量的 $n\times m$ 矩阵, 那么 $\boldsymbol{G}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$.

(4) 在 (3) 中, 若 $m=n$ 时, 这意味着 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ 是数组空间 $\mathbb{R}^n$ 的一组标准正交基的充要条件是以其为列向量组的矩阵 $\boldsymbol{X}$ 满足 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{I}_n$, 即 $\boldsymbol{X}$ 为正交矩阵. 进一步地, 由于 $\boldsymbol{X}$ 是正交矩阵当且仅当 $\boldsymbol{X}^{\mathrm{T}}$ 也是一个正交矩阵, 故这也等价于说 $\boldsymbol{X}$ 的行向量组在数组空间 $\mathbb{R}^n$ 的标准内积下构成一个正交向量组.

(5) 仍然考虑(3) 中的情形. 在例 *5.5.10* 中我们已经证明了 $\operatorname{rank}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})=\operatorname{rank}(\boldsymbol{X})$. 由此可知, $\det(\boldsymbol{G})\neq 0$ 当且仅当 $\boldsymbol{G}$ 是满秩矩阵, 当且仅当 $\boldsymbol{X}$ 是列满秩的, 而这也当且仅当 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ 线性无关. 另外, 这儿的的充要条件在一般的欧氏空间里本质上也成立; 见下面的习题.

(6) 一般而言, 欧氏空间 V 中的向量 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ 的 *Gram* 矩阵 $\boldsymbol{G}$ 是半正定的对称矩阵, 并且 $\boldsymbol{G}$ 是正定的当且仅当向量组 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ 线性无关 (参见教材最后一章最后一节的定义). 另外, 我们有 $\operatorname{rank}(\boldsymbol{G})=\operatorname{rank}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_m)$.

习题 7.2.22. 设 $\boldsymbol{G}$ 是欧氏空间 V 中的向量 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ 的 *Gram* 矩阵.

(1) 证明: $\det(\boldsymbol{G})\neq 0$ 当且仅当 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ 线性无关.

(2) 证明: $\operatorname{rank}(\boldsymbol{G})=\operatorname{rank}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_m)$.

定理 7.2.23 (可逆矩阵的 QR 分解)**.** 对任意的 n 阶可逆实矩阵 $\boldsymbol{A}$, 都存在一个 n 阶的正交矩阵 $\boldsymbol{Q}$ 和一个 n 阶的主对角元素为正数的上三角矩阵 $\boldsymbol{R}$, 使得 $\boldsymbol{A}=\boldsymbol{QR}$. 这称为 $\boldsymbol{A}$ 的 $\boldsymbol{QR}$ **分解**, 并且这样的分解是唯一的.

证明. 先证分解的存在性. 对于 $\boldsymbol{A}$ 的列向量 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_n$, 由于 $\boldsymbol{A}$ 可逆, 这组向量线性无关, 构成 $\mathbb{R}^n$ 的一组基. 对这组向量使用 Gram–Schmidt 正交化方法, 我们会得到

$$\begin{cases}\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1,\\ \boldsymbol{\beta}_2=\boldsymbol{\alpha}_2-\frac{(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1,\boldsymbol{\beta}_1)}\boldsymbol{\beta}_1,\\ \qquad\vdots\\ \boldsymbol{\beta}_n=\boldsymbol{\alpha}_n-\sum_{j=1}^{n-1}\frac{(\boldsymbol{\alpha}_n,\boldsymbol{\beta}_j)}{(\boldsymbol{\beta}_j,\boldsymbol{\beta}_j)}\boldsymbol{\beta}_j.\end{cases}$$

上面的表达式换一种写法就是

$$\boldsymbol{\alpha}_s=\sum_{j=1}^{s-1}\frac{(\boldsymbol{\alpha}_s,\boldsymbol{\beta}_j)}{(\boldsymbol{\beta}_j,\boldsymbol{\beta}_j)}\boldsymbol{\beta}_j+\boldsymbol{\beta}_s,\qquad s=1,2,\dots,n.$$

将这些用矩阵表示, 于是有

$$\begin{pmatrix}\boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}=\begin{pmatrix}\boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_n\end{pmatrix}\begin{pmatrix}1 & & * & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{pmatrix}.$$

再做归一化, 对于每个 s, 令 $\boldsymbol{\varepsilon}_s=\frac{\boldsymbol{\beta}_s}{\|\boldsymbol{\beta}_s\|}$, 从而

$$\begin{pmatrix}\boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_n\end{pmatrix}=\begin{pmatrix}\boldsymbol{\varepsilon}_1 & \boldsymbol{\varepsilon}_2 & \cdots & \boldsymbol{\varepsilon}_n\end{pmatrix}\operatorname{diag}(\|\boldsymbol{\beta}_1\|,\|\boldsymbol{\beta}_2\|,\ldots,\|\boldsymbol{\beta}_n\|).$$

此时,

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_n\end{pmatrix}=\underbrace{\begin{pmatrix}\boldsymbol{\varepsilon}_1 & \boldsymbol{\varepsilon}_2 & \cdots & \boldsymbol{\varepsilon}_n\end{pmatrix}}_{\boldsymbol{Q}}\underbrace{\operatorname{diag}(\|\boldsymbol{\beta}_1\|,\|\boldsymbol{\beta}_2\|,\ldots,\|\boldsymbol{\beta}_n\|)\begin{pmatrix}1 & & * & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{pmatrix}}_{\boldsymbol{R}}.$$

可以直接验证, 上面定义的 $\boldsymbol{Q}$ 与 $\boldsymbol{R}$ 满足要求.

接下来证明分解的唯一性. 若有正交矩阵 $\boldsymbol{Q}_1,\boldsymbol{Q}_2$ 和主对角线上为正数的上三角矩阵 $\boldsymbol{R}_1,\boldsymbol{R}_2$ 满足

$$\boldsymbol{Q}_1\boldsymbol{R}_1=\boldsymbol{A}=\boldsymbol{Q}_2\boldsymbol{R}_2,$$

那么

$$\boldsymbol{Q}_1^{-1}\boldsymbol{Q}_2=\boldsymbol{R}_1\boldsymbol{R}_2^{-1}.$$

上式的左边是正交矩阵, 右边是主对角线上为正数的上三角矩阵. 同时满足这两个条件的方阵仅为单位矩阵 (见下面的习题), 从而 $\boldsymbol{Q}_1=\boldsymbol{Q}_2$ 以及 $\boldsymbol{R}_1=\boldsymbol{R}_2$. □

习题 7.2.24. 证明: *若一个正交矩阵同时是主对角线上为正数的上三角矩阵, 那么它必定是一个单位阵.*

稍后我们会看到关于矩阵的 QR 分解的应用.

习题 7.2.25. *若 $\boldsymbol{A}$ 为 n 阶实矩阵, 有 n 个实特征值 (不要求互不相同), 那么存在 $\mathbb{R}^n$ 的一组标准正交基, 使得 $\boldsymbol{A}$ 在这组基下的矩阵为上三角阵.*

7.3 欧几里得空间中的线性变换

在这一节中，我们将线性变换的讨论与欧氏空间的内积联系起来.

正交变换与正交矩阵

定义 7.3.1. 设 $\mathscr{A}$ 是欧氏空间 V 上的线性变换. 若对于任意的 $\boldsymbol{a}, \boldsymbol{b} \in V$, 有 $(\mathscr{A}(\boldsymbol{a}), \mathscr{A}(\boldsymbol{b})) = (\boldsymbol{a}, \boldsymbol{b})$, 则称 $\mathscr{A}$ 为 V 上的**正交变换** (*orthogonal transformation*).

简言之，正交变换是保持内积不变的线性变换.

例 7.3.2. (1) 恒等变换是 V 上的正交变换. 若线性空间 V 不是零空间，则零变换不是正交变换.

(2) 在例 *6.2.4* 中我们已经见到，对于具有标准内积的欧氏空间 $\mathbb{R}^2$, 以原点为中心，逆时针旋转角度 θ 的变换 $\mathscr{A}_\theta$ 是 $\mathbb{R}^2$ 上的线性变换，它在 $\mathbb{R}^2$ 的自然基下的矩阵为

$$\boldsymbol{A}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

任取向量 $\boldsymbol{x} = (x_1, x_2)^{\mathrm{T}}, \boldsymbol{y} = (y_1, y_2)^{\mathrm{T}} \in \mathbb{R}^2$, 我们有

$$\begin{aligned}
(\mathscr{A}_\theta(\boldsymbol{x}), \mathscr{A}_\theta(\boldsymbol{y})) &= (\boldsymbol{A}_\theta\boldsymbol{x}, \boldsymbol{A}_\theta\boldsymbol{y}) \\
&= \left(\begin{pmatrix} \cos(\theta)x_1 - \sin(\theta)x_2 \\ \sin(\theta)x_1 + \cos(\theta)x_2 \end{pmatrix}, \begin{pmatrix} \cos(\theta)y_1 - \sin(\theta)y_2 \\ \sin(\theta)y_1 + \cos(\theta)y_2 \end{pmatrix} \right) \\
&= (\cos(\theta)x_1 - \sin(\theta)x_2)(\cos(\theta)y_1 - \sin(\theta)y_2) \\
&\qquad + (\sin(\theta)x_1 + \cos(\theta)x_2)(\sin(\theta)y_1 + \cos(\theta)y_2) \\
&= x_1y_1 + x_2y_2 = (\boldsymbol{x}, \boldsymbol{y}).
\end{aligned}$$

这说明，旋转变换 $\mathscr{A}_\theta$ 是 $\mathbb{R}^2$ 上相对于标准内积的正交变换.

注 7.3.3. 在欧氏空间中，向量的长度和向量之间的夹角由空间的内积来决定:

- $\|\boldsymbol{a}\| = \sqrt{(\boldsymbol{a}, \boldsymbol{a})}$;
- 非零向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 之间的夹角的余弦为 $\frac{(\boldsymbol{a},\boldsymbol{b})}{\sqrt{(\boldsymbol{a},\boldsymbol{a})}\sqrt{(\boldsymbol{b},\boldsymbol{b})}}$.

故正交变换保持向量的长度和向量之间的夹角不变，即 $\|\mathscr{A}(\boldsymbol{a})\| = \|\boldsymbol{a}\|$, 且 $\mathscr{A}(\boldsymbol{a})$ 与 $\mathscr{A}(\boldsymbol{b})$ 之间的夹角与 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 之间的夹角一致. 特别地，正交变换把正交的向量组变成正交的向量组，把标准正交基变成标准正交基.

事实上, 我们有更多的刻划:

定理 7.3.4. 设 $\mathscr{A}$ 是 n 维欧氏空间 V 上的一个线性变换, 则以下几个条件等价: 教材定理 7.3.1

(1) $\mathscr{A}$ 是正交变换;

(2) $\mathscr{A}$ 保持向量的模不变;

(3) $\mathscr{A}$ 将任意的标准正交基变为标准正交基;

(4) $\mathscr{A}$ 将给定的一组标准正交基变为标准正交基.

证明. 我们已经看到了 (1)⇒(2) 以及 (3)⇒(4). 关于 (2)⇒(3), 我们只需注意到之前提到的事实: $\boldsymbol{a}\perp\boldsymbol{b}$ 当且仅当 $\|\boldsymbol{a}+\boldsymbol{b}\|^2=\|\boldsymbol{a}\|^2+\|\boldsymbol{b}\|^2$. 接下来, 我们证明 (4)⇒(1):

设 $\boldsymbol{\varepsilon}_1,\dots,\boldsymbol{\varepsilon}_n$ 是欧氏空间 V 上给定的一组标准正交基. 对于任意的 $\boldsymbol{a},\boldsymbol{b}\in V$, 设 $(a_1,\dots,a_n)^{\mathrm{T}}$ 与 $(b_1,\dots,b_n)^{\mathrm{T}}$ 是它们在这组基下的坐标. 此时

$$(\mathscr{A}(\boldsymbol{a}),\mathscr{A}(\boldsymbol{b}))=\left(\mathscr{A}\left(\sum_{i=1}^n a_i\boldsymbol{\varepsilon}_i\right),\mathscr{A}\left(\sum_{j=1}^n b_j\boldsymbol{\varepsilon}_j\right)\right)=\sum_{i=1}^n\sum_{j=1}^n a_ib_j(\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j))$$

请回忆注 7.2.11中的讨论

$$\xlongequal[\text{为标准正交基}]{\mathscr{A}(\boldsymbol{\varepsilon}_1),\dots,\mathscr{A}(\boldsymbol{\varepsilon}_n)}\sum_{i=1}^n\sum_{j=1}^n a_ib_j\delta_{ij}=\sum_{i=1}^n a_ib_i\xlongequal[\text{为标准正交基}]{\boldsymbol{\varepsilon}_1,\dots,\boldsymbol{\varepsilon}_n}(\boldsymbol{a},\boldsymbol{b}).$$

故 $\mathscr{A}$ 是一个正交变换. □

例 7.3.5. 设 $\mathscr{A}$ 是平面 $\mathbb{R}^2$ 上的线性变化, 将 $(1,0)^{\mathrm{T}}$ 变成了 $(1,0)^{\mathrm{T}}$ 自身, 将 $(0,1)^{\mathrm{T}}$ 变成了 $(1/\sqrt{2},1/\sqrt{2})^{\mathrm{T}}$. $\mathscr{A}$ 虽然将 $\mathbb{R}^2$ 的一组由单位向量构成的基变成了另外一组由单位向量构成的基, 但是 $\mathscr{A}$ 不是正交变换.

定理 7.3.6. 设 $\mathscr{A}$ 为 n 维欧氏空间 V 上的线性变换, 在 V 的一组标准正交基 $\boldsymbol{\varepsilon}_1,\dots,\boldsymbol{\varepsilon}_n$ 下的矩阵为 $\boldsymbol{A}$. 那么, $\mathscr{A}$ 为正交变换的充要条件是 $\boldsymbol{A}$ 为正交矩阵. 教材定理 7.3.2

证明. 由定义

$$\mathscr{A}(\boldsymbol{\varepsilon}_1,\dots,\boldsymbol{\varepsilon}_n)=(\boldsymbol{\varepsilon}_1,\dots,\boldsymbol{\varepsilon}_n)\boldsymbol{A}.$$

若设 $\boldsymbol{A}=(a_{ij})_{n\times n}\in\mathbb{R}^{n\times n}$, 则对于 $i=1,2,\dots,n$, 我们有

$$\mathscr{A}(\boldsymbol{\varepsilon}_i)=\sum_{k=1}^n a_{ki}\boldsymbol{\varepsilon}_k.$$

注意 a 的下标

由定理 7.3.4, $\mathscr{A}$ 为正交变换当仅当 $\mathscr{A}(\boldsymbol{\varepsilon}_1),\dots,\mathscr{A}(\boldsymbol{\varepsilon}_n)$ 为标准正交基, 当且仅当在 $1\le i,j\le n$ 时我们有 $\delta_{ij}=(\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j))$. 但是,

$$(\mathscr{A}(\boldsymbol{\varepsilon}_i),\mathscr{A}(\boldsymbol{\varepsilon}_j))=\left(\sum_{k=1}^n a_{ki}\boldsymbol{\varepsilon}_k,\sum_{l=1}^n a_{lj}\boldsymbol{\varepsilon}_l\right)=\sum_{k=1}^n\sum_{l=1}^n a_{ki}a_{lj}(\boldsymbol{\varepsilon}_k,\boldsymbol{\varepsilon}_l)$$

$$\xlongequal[\text{为标准正交基}]{\varepsilon_1,\ldots,\varepsilon_n} \sum_{k=1}^{n}\sum_{l=1}^{n} a_{ki}a_{lj}\delta_{kl} = \sum_{k=1}^{n} a_{ki}a_{kj} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})_{i,j},$$

故, 这也当且仅当 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_n$, 即 $\boldsymbol{A}$ 为正交矩阵. □

例 7.3.7. 在例 *7.3.2* (2) 中的旋转变换 $\mathscr{A}_\theta$ 为正交变换, 从而 $\mathscr{A}_\theta$ 在自然基下的矩阵 $\boldsymbol{A}_\theta$ 为正交矩阵. 当然, 我们可以很轻松地利用定义 *(*旋转变换保持内积*)* 来直接验证这一点.

注 7.3.8. 在定理 *7.3.6* 中的基 $\varepsilon_1,\ldots,\varepsilon_n$ 为标准正交基这一条件非常重要, 因为一般而言, 正交变换在非标准正交基的基下的矩阵不为正交矩阵, 即, 正交矩阵在相似变换下一般不再是正交矩阵了. 例如, 考虑正交矩阵 $\boldsymbol{A} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ 和可逆矩阵 $\boldsymbol{T} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$, 其中 $\theta, a \in \mathbb{R}$. 此时,

$$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} = \begin{pmatrix} \cos(\theta) - a\sin(\theta) & -a\ (a\sin(\theta) + \cos(\theta)) - \sin(\theta) + a\cos(\theta) \\ \sin(\theta) & a\sin(\theta) + \cos(\theta) \end{pmatrix}.$$

若其为正交矩阵, 其第一个列向量必为单位向量. 可是并不是所有的 $a, \theta \in \mathbb{R}$ 都使得

$$(\cos(\theta) - a\sin(\theta))^2 + \sin^2(\theta) = 1.$$

故 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 一般而言不再是正交矩阵了.

与定理 7.3.6 相关的是下面的一个命题, 它告诉我们, 在有限维欧氏空间 V 中, 给定了一组标准正交基, 则 V 的所有标准正交基与正交矩阵一一对应.

命题 7.3.9. 若欧氏空间 V 的一组标准正交基 $\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n$ 到另外一组基 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 的过渡矩阵为 $\boldsymbol{P}$, 则 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 仍然是 V 的标准正交基的充要条件是 $\boldsymbol{P}$ 为正交方阵. (注意, 若 V 是数组空间 $\mathbb{R}^n$, $\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n$ 是它的自然基, 那么 $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ 是 $\boldsymbol{P}$ 的列向量)

证明. **(证法一, 硬算)** 设 $\boldsymbol{P} = (p_{ij})_{n\times n}$, 则由定义知

$$\boldsymbol{\alpha}_i = \sum_{k=1}^{n} p_{ki}\boldsymbol{\varepsilon}_k.$$

对于任意的 i, j 有

$$(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j) = \left(\sum_{k=1}^{n} p_{ki}\boldsymbol{\varepsilon}_k, \sum_{l=1}^{n} p_{lj}\boldsymbol{\varepsilon}_l\right) = \sum_{k=1}^{n}\sum_{l=1}^{n} p_{ki}p_{lj}(\boldsymbol{\varepsilon}_k, \boldsymbol{\varepsilon}_l)$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{n} p_{ki} p_{lj} \delta_{kl} = \sum_{k=1}^{n} p_{ki} p_{kj} = (\boldsymbol{P}^{\mathrm{T}} \boldsymbol{P})_{i,j}.$$

该计算表明 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 是 V 的标准正交基的充要条件是 $\boldsymbol{P}$ 为正交方阵.

(证法二, 逻辑推导) 对于 V 的另外一组基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$, 存在唯一的 V 上的线性映射 $\mathscr{A}$, 使得 $\mathscr{A}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\alpha}_i$ $(1 \leq i \leq n)$. 此时, $\mathscr{A}$ 在基 $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ 下的矩阵 (记作 $\boldsymbol{A}$) 就是 $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ 到 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 的过渡矩阵 (记作 $\boldsymbol{P}$). 于是, $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 是一组标准正交基, 当且仅当线性变换 $\mathscr{A}$ 把已知的标准正交基 $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ 变为标准正交基, 当且仅当 $\mathscr{A}$ 为正交变换, 当且仅当线性变换 $\mathscr{A}$ 在已知的标准正交基 $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ 的矩阵 $\boldsymbol{A}$ 为正交矩阵, 即过渡矩阵 $\boldsymbol{P}$ 为正交矩阵. □

注 7.3.10. 我们再小结一下正交矩阵的常见性质.

(1) 若 $\boldsymbol{A}, \boldsymbol{B}$ 为两个同阶的正交矩阵, 则 $\boldsymbol{AB}$ 也是正交矩阵:

$$(\boldsymbol{AB})^{\mathrm{T}}(\boldsymbol{AB}) = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AB} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I}.$$

(2) 若 $\boldsymbol{A}$ 为正交矩阵, 则 $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$ 也是正交矩阵:

$$(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}.$$

(3) 若 $\boldsymbol{A}$ 为正交矩阵, 则 $\det(\boldsymbol{A}) = \pm 1$. 这是由于 $1 = \|\boldsymbol{I}\| = \|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\| = \|\boldsymbol{A}^{\mathrm{T}}\| \cdot \|\boldsymbol{A}\| = \|\boldsymbol{A}\|^2$, 而 $\|\boldsymbol{A}\| \in \mathbb{R}$.

由此, 若 $|\boldsymbol{A}| = 1$, 则称 $\boldsymbol{A}$ 为**第一类正交阵**, 否则, 则称之为**第二类正交阵**. 注意, 线性变换的行列式不依赖于具体基的选取, 故我们有如下的定义.

- 若正交变换 $\mathscr{A}$ 的行列式为 1, 则称 $\mathscr{A}$ 为**第一类变换**. 典型的例子是平面的关于原点的旋转变换.
- 若正交变换 $\mathscr{A}$ 的行列式为 -1, 则称 $\mathscr{A}$ 为**第二类变换**. 典型的例子是平面的关于任何一条指定的过原点的直线的镜面反射.

可以证明, 上面指出的旋转变换与镜面反射是平面 $\mathbb{R}^2$ 上所有的正交变换 (教材作业 *P221#14*).

(4) 若 $\boldsymbol{A}$ 是正交矩阵, 则伴随矩阵 $\boldsymbol{A}^*$ 也是正交矩阵. 这是因为

$$\boldsymbol{A}\boldsymbol{A}^* = \det(\boldsymbol{A})\boldsymbol{I} = \pm\boldsymbol{I},$$

从而 $\boldsymbol{A}^* = \pm\boldsymbol{A}^{-1}$.

(5) 我们再复习一下正交矩阵的常见判定准则. 对于 n 阶实方阵 $\boldsymbol{A}$, 假定它的行向量组为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, 它的列向量组为 $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$. 那么, 以下几条等价:

(a) $\boldsymbol{A}$ 为正交矩阵;

(b) 对任意的 i, j 有 $\boldsymbol{x}_i \boldsymbol{x}_j^{\mathrm{T}} = \delta_{ij}$;

(c) 对任意的 i, j 有 $\boldsymbol{y}_i^{\mathrm{T}} \boldsymbol{y}_j = \delta_{ij}$;

(d) $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 是 $\mathbb{R}^n$ 在标准内积下的一组标准正交基;

(e) $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ 是 $\mathbb{R}^n$ 在标准内积下的一组标准正交基.

例 7.3.11. 置换矩阵 (*permutation matrix*) 是将单位矩阵的各列重新排列得到的矩阵, 即形如 $(\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \ldots, \boldsymbol{e}_{i_n})$ 的矩阵, 其中列向量 $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ 为 $\mathbb{R}^n$ 的自然基, 而 $\sigma = (i_1\ i_2\ \cdots\ i_n) \in \mathcal{S}_n$ 是一个置换 (也称作排列; 请复习教材在定义 *4.3.2* 中给出的关于排列的记号). 显然, 置换矩阵的列向量组仍然构成标准正交基, 从而该矩阵是正交矩阵, 并且, 这样的矩阵是一个第一类正交矩阵的充要条件是其对应的置换 σ 为一个偶置换. 另外, 置换矩阵也可以视为由单位矩阵的各行重新排列得到的矩阵.

例 7.3.12. (1) 设 $\mathscr{A}$ 与 $\mathscr{B}$ 是有限维线性空间 V 上的线性变换, 在基 $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ 下的矩阵分别为 $\boldsymbol{A}$ 与 $\boldsymbol{B}$. 不难验证, 映射的复合 $\mathscr{A} \circ \mathscr{B}$ 仍然是 V 上的线性变换, 并且相应的矩阵为 $\boldsymbol{AB}$.

事实上, 若 $V = F^n$ 是数组空间, $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ 是自然基, $\boldsymbol{x} \in V$ 写成列向量形式, 则 $\mathscr{A} : \boldsymbol{x} \mapsto \boldsymbol{Ax}$, $\mathscr{B} : \boldsymbol{x} \mapsto \boldsymbol{Bx}$. 通过直接验证, 我们看到 $\mathscr{A} \circ \mathscr{B} : \boldsymbol{x} \rightsquigarrow \boldsymbol{Bx} \rightsquigarrow \boldsymbol{ABx}$. 这间接表明了 $\mathscr{A} \circ \mathscr{B}$ 对应的矩阵应当为 $\boldsymbol{AB}$.

进一步地, 若 V 是欧氏空间, $\mathscr{A}$ 与 $\mathscr{B}$ 都是正交变换, 则对于任意的 $\boldsymbol{x}, \boldsymbol{y} \in V$,

$$(\mathscr{A}(\mathscr{B}(\boldsymbol{x})), \mathscr{A}(\mathscr{B}(\boldsymbol{y}))) = (\mathscr{B}(\boldsymbol{x}), \mathscr{B}(\boldsymbol{y})) = (\boldsymbol{x}, \boldsymbol{y}).$$

这说明正交变换的复合仍然是正交变换. 这对应于前面提到的事实: 同阶正交矩阵的乘积仍然是正交矩阵.

(2) 设线性变换 $\mathscr{A}$ 是有限维线性空间 V 上的一一的满射, 即, V 中的任何元素在 $\mathscr{A}$ 下都存在唯一一个原像. 此时, 不难看出, $\mathscr{A}$ 存在**逆映射** $\mathscr{A}^{-1} : V \to V$, 将 V 中的元素映射成为其在 $\mathscr{A}$ 下的原像. $\mathscr{A}^{-1}$ 仍然是线性映射, 从而是 V 上的线性变换, 称为 $\mathscr{A}$ 的**逆变换**. 进一步地, $\mathscr{A}^{-1}$ 仍然是一一的满射, 满足 $(\mathscr{A}^{-1})^{-1} = \mathscr{A}$.

设 V 是有限维的, 而 $\mathscr{A}^{-1}$ 与 $\mathscr{A}$ 在 V 的基 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ 下的矩阵为 $\boldsymbol{B}$ 和 $\boldsymbol{A}$. 由于 $\mathscr{A}^{-1} \circ \mathscr{A} = \mathrm{id}_V$, 这说明 $\boldsymbol{BA} = \boldsymbol{I}_n$, 从而 $\mathscr{A}^{-1}$ 对应的矩阵 $\boldsymbol{B}$ 恰为 $\boldsymbol{A}^{-1}$.

设 $\mathscr{A}$ 是有限维欧氏空间 V 上的正交变换, 不难验证 $\mathscr{A}$ 存在逆变换 (参见下面的习题). 对于任意的 $\boldsymbol{x},\boldsymbol{y}\in V$, 由于 $\mathscr{A}^{-1}\circ\mathscr{A}=\mathrm{id}_V=\mathscr{A}\circ\mathscr{A}^{-1}$,

$$(\mathscr{A}^{-1}(\boldsymbol{x}),\mathscr{A}^{-1}(\boldsymbol{y}))\xlongequal{\mathscr{A}\text{ 为正交变换}}(\mathscr{A}(\mathscr{A}^{-1}(\boldsymbol{x})),\mathscr{A}(\mathscr{A}^{-1}(\boldsymbol{y})))=(\boldsymbol{x},\boldsymbol{y}),$$

这说明 $\mathscr{A}^{-1}$ 仍然是正交变换. 这对应于前面提到的事实: 正交方阵的逆仍然是正交方阵.

习题 7.3.13. 若 $\mathscr{A}$ 是有限维欧氏空间 V 上的正交变换, 证明 V 中的任何元素在 $\mathscr{A}$ 下都存在唯一一个原像. (可以参考注 *6.1.8*)

由于我们现在是在实数域上考虑, 线性变换一般不一定有 (实) 特征值.

命题 7.3.14. 设 $\mathscr{A}$ 是 n 维欧氏空间 V 上的正交变换.

(1) 若 $\lambda\in\mathbb{R}$ 是 $\mathscr{A}$ 的特征值, 则 $\lambda=\pm1$.

(2) 若 V 是 n 维欧氏空间, 其中 n 为奇数, 而 $\mathscr{A}$ 为第一类正交变换, 则 $\mathscr{A}$ 一定存在值为 1 的特征值.

证明. (1) $\mathscr{A}$ 在标准正交基下的矩阵为正交方阵, 而我们之前的讨论 (例 6.3.17(4)) 说明了正交矩阵的特征值 λ 的模长必为 1. 由于线性变换的特征值是它在任何一组基下的矩阵的特征值, 从而 $\lambda=\pm1$.

注意, 若 V 的维数不一定有限, 但是 $\lambda\in\mathbb{R}$ 是 $\mathscr{A}$ 的特征值, 则我们也可以直接证明 $\lambda=\pm1$. 证明如下. 设 $\boldsymbol{x}\in V$ 是属于 λ 的特征向量, 从而 $\mathscr{A}(\boldsymbol{x})=\lambda\boldsymbol{x}$. 此时, 由正交性,

$$(\mathscr{A}(\boldsymbol{x}),\mathscr{A}(\boldsymbol{x}))=(\boldsymbol{x},\boldsymbol{x}).$$

另一方面, 左式为

$$(\lambda\boldsymbol{x},\lambda\boldsymbol{x})=\lambda^2(\boldsymbol{x},\boldsymbol{x}).$$

由于 $\boldsymbol{x}\neq\mathbf{0}$, 故 $(\boldsymbol{x},\boldsymbol{x})\neq0$. 从而上面两式说明 $\lambda^2=1$, 即 $\lambda=\pm1$.

(2) 特征多项式 $p_{\mathscr{A}}(\lambda)$ 是实系数的 n 次多项式. 由于 n 为奇数, $p_{\mathscr{A}}(\lambda)$ 必有实根. 注意到实系数多项式的虚根必然成对出现, 我们可以假定 $\mathscr{A}$ 的所有特征值为

$$\lambda_1,\overline{\lambda_1},\lambda_2,\overline{\lambda_2},\ldots,\lambda_m,\overline{\lambda_m},\lambda_{m+1},\ldots,\lambda_{n-m}.$$

其中 $\lambda_i,\overline{\lambda_i}$ 为成对出现的复根, $1\leq i\leq m$; 而 $\lambda_{m+1},\ldots,\lambda_{n-m}$ 为实根, 并且由上面的讨论知其为 ±1.

注意到每对虚根的乘积 $\lambda_i\overline{\lambda_i} = \|\lambda_i\|^2$ 为正实数, 而所有这些特征值的乘积为 $\mathscr{A}$ 的行列式 1, 这说明这 $n-2m$ (由于 n 是奇数, 这也是个奇数) 个实根 $\lambda_{m+1}, \ldots, \lambda_{n-m}$ 不可能全为 -1, 从而其中必然存在值为 1 的特征值. $\square$

注 7.3.15. 上面命题的一个几何意义: 三维欧氏空间的第一类正交变换, 其必保持一个对称轴不变; 进一步的讨论可以表明, 它是绕该对称轴的旋转变换.

该事实, 学生应该记住, 其具体证明课堂不讲

设 $\mathscr{A}$ 是这样的一个第一类正交变换, 则 $\lambda = 1$ 是它的一个特征值. 设单位向量 ε_1 是 $\lambda = 1$ 的一个特征向量, 则对任意 $k \in \mathbb{R}$, $\mathscr{A}(k\varepsilon_1) = k\mathscr{A}(\varepsilon_1) = k\varepsilon_1$. 这说明 $\mathscr{A}$ 保持直线 $\mathbb{R}\varepsilon_1 = \{\, k\varepsilon_1 \mid k \in \mathbb{R} \,\}$ 上的向量不变. 利用正交化的方法, 我们可以找到 $\varepsilon_2, \varepsilon_3 \in \mathbb{R}^3$ 使得 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 为一组标准正交基.

任取 $\boldsymbol{\alpha} \in \mathbb{R}^3$, 假设它在这组基下的坐标为 $(x, y, z)^{\mathrm{T}}$, 它在 $\varepsilon_2\varepsilon_3$ 平面上的投影 $\boldsymbol{\alpha}' = y\varepsilon_2 + z\varepsilon_3$ 与 ε_1 垂直, 从而 $\mathscr{A}(\boldsymbol{\alpha}')$ 与 $\mathscr{A}(\varepsilon_1) = \varepsilon_1$ 垂直. 这说明 $\mathscr{A}$ 将 $\varepsilon_2\varepsilon_3$ 平面上的点仍然映射到该平面上, 从而 $\mathscr{A}$ 局限在该平面上后成为该平面的线性变换, 记作 $\mathscr{A}|_{\varepsilon_2\varepsilon_3}$. 另外需要指出的是, $\mathbb{R}^3$ 的标准内积局限在 $\varepsilon_2\varepsilon_3$ 平面后成为该平面的一个内积, 而 $\varepsilon_2, \varepsilon_3$ 在该内积下是一组标准正交基. 设 $\mathscr{A}|_{\varepsilon_2\varepsilon_3}$ 在基 $\varepsilon_2, \varepsilon_3$ 下的矩阵为 $\boldsymbol{A}'$, 则 $\mathscr{A}$ 在 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 下的矩阵为

$$\boldsymbol{A} = \begin{pmatrix} 1 & \\ & \boldsymbol{A}' \end{pmatrix}.$$

由于 $\boldsymbol{A}$ 是行列式为 1 的正交矩阵, 这迫使 $\boldsymbol{A}'$ 是行列式为 1 的正交矩阵, 从而 $\mathscr{A}|_{\varepsilon_2\varepsilon_3}$ 是 $\varepsilon_2\varepsilon_3$ 平面上的第一类正交变换, 即旋转变换. 因此, $\mathscr{A}$ 是 $\mathbb{R}^3$ 的绕 ε_1 轴的旋转.

习题 7.3.16. 设欧氏空间 V 中有一个指定的非零向量 $\boldsymbol{\alpha}$, 定义映射

$$\varphi : V \to V, \quad \boldsymbol{x} \mapsto \boldsymbol{x} - \frac{2(\boldsymbol{x}, \boldsymbol{\alpha})}{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}\boldsymbol{\alpha}.$$

证明 φ 是一个第二类的正交变换.

对称变换与对称矩阵

定义 7.3.18. 设 $\mathscr{A}$ 是欧氏空间 V 上的线性变换. 若对于任意的 $\boldsymbol{a},\boldsymbol{b}\in V$ 有 $(\mathscr{A}(\boldsymbol{a}),\boldsymbol{b})=(\boldsymbol{a},\mathscr{A}(\boldsymbol{b}))$, 则称 $\mathscr{A}$ 是 V 上的**对称变换** (*symmetric transformation*).

例 7.3.19. (1) 零变换和恒等变换都是对称变换.

(2) 在平面 $\mathbb{R}^2$ 上的旋转变换 $\mathscr{A}_\theta$ 一般不为对称变换.

下面我们讨论对称变换与对称矩阵的关系.

定理 7.3.20. 设 $\mathscr{A}$ 是 n 维欧氏空间 V 上的线性变换, 则以下几条等价.

(1) $\mathscr{A}$ 是 V 上的对称变换.

(2) $\mathscr{A}$ 在任何一组标准正交基下的矩阵都是实对称方阵.

(3) $\mathscr{A}$ 在给定的一组标准正交基下的矩阵是实对称方阵.

证明. (1) $\Rightarrow$ (2): 任取 V 的一组标准正交基 $\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n$, 假设 $\mathscr{A}$ 在其上的矩阵为 $\boldsymbol{A}=(a_{ij})_{n\times n}$. 依定义, 这说明对于任意的 i 有

$$\mathscr{A}(\boldsymbol{\varepsilon}_i)=\sum_{k=1}^n a_{ki}\boldsymbol{\varepsilon}_k.$$

此时,

$$\begin{aligned}(\mathscr{A}(\boldsymbol{\varepsilon}_i),\boldsymbol{\varepsilon}_j)&=\left(\sum_{k=1}^n a_{ki}\boldsymbol{\varepsilon}_k,\boldsymbol{\varepsilon}_j\right)=\sum_{k=1}^n a_{ki}(\boldsymbol{\varepsilon}_k,\boldsymbol{\varepsilon}_j)=\sum_{k=1}^n a_{ki}\delta_{kj}=a_{ji},\\(\boldsymbol{\varepsilon}_i,\mathscr{A}(\boldsymbol{\varepsilon}_j))&=\left(\boldsymbol{\varepsilon}_i,\sum_{k=1}^n a_{kj}\boldsymbol{\varepsilon}_k\right)=\sum_{k=1}^n a_{kj}(\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_k)=\sum_{k=1}^n a_{kj}\delta_{ik}=a_{ij}.\end{aligned}$$

由于 $\mathscr{A}$ 为对称变换, 这说明对于任意的 i,j 有 $(\mathscr{A}(\boldsymbol{\varepsilon}_i),\boldsymbol{\varepsilon}_j)=(\boldsymbol{\varepsilon}_i,\mathscr{A}(\boldsymbol{\varepsilon}_j))$, 即 $a_{ij}=a_{ji}$, 从而 $\boldsymbol{A}$ 为对称方阵.

(2) $\Rightarrow$ (3): 这是显然的.

(3) $\Rightarrow$ (1): 设 $\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n$ 为 V 上事先给定的一组标准正交基, $\mathscr{A}$ 在其上的矩阵为 $\boldsymbol{A}=(a_{ij})_{n\times n}$. 由假设, $\boldsymbol{A}$ 为对称方阵. 从而由上面的推导可以看到, 对任意的 i,j 有

$$(\mathscr{A}(\boldsymbol{\varepsilon}_i),\boldsymbol{\varepsilon}_j)=(\boldsymbol{\varepsilon}_i,\mathscr{A}(\boldsymbol{\varepsilon}_j)).\tag{7.1}$$

此时, 考虑一般的 $\boldsymbol{a}=\sum_{i=1}^n a_i\boldsymbol{\varepsilon}_i,\boldsymbol{b}=\sum_{i=1}^n b_i\boldsymbol{\varepsilon}_i\in V$. 直接计算, 有

$$(\mathscr{A}(\boldsymbol{a}),\boldsymbol{b})=\left(\mathscr{A}\left(\sum_{i=1}^n a_i\boldsymbol{\varepsilon}_i\right),\sum_{j=1}^n b_j\boldsymbol{\varepsilon}_j\right)=\left(\sum_{i=1}^n a_i(\mathscr{A}\boldsymbol{\varepsilon}_i),\sum_{j=1} b_j\boldsymbol{\varepsilon}_j\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j(\mathscr{A}(\boldsymbol{\varepsilon}_i), \boldsymbol{\varepsilon}_j) \xlongequal{(7.1)} \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j(\boldsymbol{\varepsilon}_i, \mathscr{A}(\boldsymbol{\varepsilon}_j)) \\
&= \cdots = (\boldsymbol{a}, \mathscr{A}(\boldsymbol{b})).
\end{aligned}
$$

故 $\mathscr{A}$ 为对称变换. $\square$

注 7.3.21. 接着上面的定理的证明, 设 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 是 V 的另外一组基, $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ 到 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 的过渡矩阵为 $\boldsymbol{P}$, 则 $\mathscr{A}$ 在 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{P}^{-1}\boldsymbol{AP}$. 在一般情形下, $\boldsymbol{P}^{-1}\boldsymbol{AP}$ 不再是对称方阵. 另一方面, 若 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 仍然是 V 的一组标准正交基, 则由定理 7.3.20 知 $\boldsymbol{P}^{-1}\boldsymbol{AP}$ 必为对称方阵. 事实上, 由命题 *7.3.9* 我们知此时的过渡矩阵 $\boldsymbol{P}$ 为实正交方阵, 故由 $\boldsymbol{A}$ 为对称方阵, 我们可以直接验证 $\boldsymbol{P}^{-1}\boldsymbol{AP} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}$ 为对称方阵.

实对称阵的对角化 一般而言, 实方阵的特征多项式可能有虚根, 从而不能实相似对角化. 但是实对称的矩阵的特征值都是实数.

命题 7.3.22. 设 $\boldsymbol{A}$ 为 n 阶实对称方阵, 则 $\boldsymbol{A}$ 所有的复特征值其实都是实数, 而 $\boldsymbol{A}$ 的属于不同特征值的实特征向量在 $\mathbb{R}^n$ 的标准内积下必然正交.

证明. 设复数 λ 是 $\boldsymbol{A}$ 的特征值, 从而存在非零复向量 $\boldsymbol{x} \in \mathbb{C}^n$, 使得 $\boldsymbol{Ax} = \lambda\boldsymbol{x}$. 对该式两边取复共轭有 $\overline{\boldsymbol{Ax}} = \overline{\lambda\boldsymbol{x}}$, 即 $\overline{\boldsymbol{A}}\overline{\boldsymbol{x}} = \overline{\lambda}\overline{\boldsymbol{x}}$. 由于 $\boldsymbol{A}$ 为实对称矩阵, 即 $\overline{\boldsymbol{A}} = \boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$. 这说明

$$
\lambda\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x} = \overline{\boldsymbol{x}}^{\mathrm{T}}(\lambda\boldsymbol{x}) = \overline{\boldsymbol{x}}^{\mathrm{T}}(\boldsymbol{AX}) = (\overline{\boldsymbol{x}}^{\mathrm{T}}\overline{\boldsymbol{A}}^{\mathrm{T}})\boldsymbol{X} = (\overline{\boldsymbol{Ax}})^{\mathrm{T}}\boldsymbol{x} = (\overline{\lambda\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{x} = \overline{\lambda}\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}.
$$

这说明 $(\lambda - \overline{\lambda})\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x} = 0$. 由于 $\boldsymbol{x} \neq \boldsymbol{0}$, 我们有 $\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|^2 \neq 0$, 从而 $\lambda = \overline{\lambda}$, 即 $\lambda \in \mathbb{R}$.

接下来设 $\lambda_1, \lambda_2 \in \mathbb{R}$ 为 $\boldsymbol{A}$ 不同的特征值, 而 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ 为相应的特征向量. 于是有

$$
\lambda_1\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 = (\lambda_1\boldsymbol{x}_1)^{\mathrm{T}}\boldsymbol{x}_2 = (\boldsymbol{Ax}_1)^{\mathrm{T}}\boldsymbol{x}_2 = \boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_2 = \boldsymbol{x}_1^{\mathrm{T}}(\boldsymbol{Ax}_2) = \boldsymbol{x}_1^{\mathrm{T}}(\lambda_2\boldsymbol{x}_2) = \lambda_2\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2.
$$

由于 $\lambda_1 \neq \lambda_2$, 故有 $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 = 0$. $\square$

对称变换也有相应的性质.

定理 7.3.23. 设 $\mathscr{A}$ 是有限维欧氏空间 V 上的对称变换, 则 $\mathscr{A}$ 的特征值都是实数, 且 $\mathscr{A}$ 在不同特征值下的特征向量相互正交.

证明. 关于第一部分断言, 我们先取 V 的一组标准正交基, 并设 $\mathscr{A}$ 在这组基下的矩阵为 $\boldsymbol{A}$. 于是, $\boldsymbol{A}$ 为实对称阵, 而 $\boldsymbol{A}$ 的特征值就是 $\mathscr{A}$ 的特征值. 于是利用命题 7.3.22 即可.

关于第二部分断言, 我们设 V 中的向量 $\boldsymbol{x}_1$ 是属于 λ_1 的特征值, $\boldsymbol{x}_2$ 是属于 λ_2 的特征值, 并且 $\lambda_1 \neq \lambda_2$. 此时,

注意这一部分证明在 V 不是有限维时也成立

$$\lambda_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\lambda_1 \boldsymbol{x}_1, \boldsymbol{x}_2) = (\mathscr{A}\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}_1, \mathscr{A}\boldsymbol{x}_2) = (\boldsymbol{x}_1, \lambda_2 \boldsymbol{x}_2) = \lambda_2(\boldsymbol{x}_1, \boldsymbol{x}_2).$$

由于 $\lambda_1 \neq \lambda_2$, 该式说明 $(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0$, 从而 $\boldsymbol{x}_1$ 与 $\boldsymbol{x}_2$ 正交. □

下面的定理指明, 实对称阵一定**正交相似**于对角阵. 注意, 用正交阵做相似变换等同于用其做相合变换.

定理 7.3.24. *对于 n 阶实对称方阵 $\boldsymbol{A}$, 总存在同阶正交阵 $\boldsymbol{T}$ 使得 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 为对角阵.*

证明. **(思路一)** 我们对于阶数 n 作归纳法. $n=1$ 的情形是显然的. 假设 $n-1$ 时结论成立. 对于 n 的情形, 设 λ_1 是 $\boldsymbol{A}$ 的一个特征值. 由命题 7.3.22, $\lambda_1 \in \mathbb{R}$. 对于特征子空间

这是教材上的证明, 课堂上不讲, 留作课后阅读

$$V_{\boldsymbol{A}}(\lambda_1) = \{\, \boldsymbol{x} \in \mathbb{R}^n \mid (\lambda_1 \boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \,\},$$

由条件已知 $V_{\boldsymbol{A}}(\lambda_1)$ 不是零空间, 从而可以找到非零向量 $\boldsymbol{x}_1 \in V_{\boldsymbol{A}}(\lambda_1)$. 通过归一化, 我们可以假定 $\|\boldsymbol{x}_1\| = 1$. 将 $\boldsymbol{x}_1$ 可以扩充为欧氏空间 $\mathbb{R}^n$ 的一组标准正交基 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ 构成了矩阵 $\boldsymbol{T}_n$ 的列向量序列, 则 $\boldsymbol{T}_n$ 是一个正交矩阵, 满足 $\boldsymbol{A}\boldsymbol{T}_n = \boldsymbol{T}_n \begin{pmatrix} \lambda_1 & * \\ \boldsymbol{0} & \boldsymbol{A}_{n-1} \end{pmatrix}$, 即

$$\boldsymbol{T}_n^{-1}\boldsymbol{A}\boldsymbol{T}_n = \begin{pmatrix} \lambda_1 & * \\ \boldsymbol{0} & \boldsymbol{A}_{n-1} \end{pmatrix}.$$

由于 $\boldsymbol{T}_n$ 为正交矩阵, 上面的左式为实对称阵, 从而上面的右式实际上为 $\begin{pmatrix} \lambda_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{n-1} \end{pmatrix}$, 并且子方阵 $\boldsymbol{A}_{n-1}$ 为 $(n-1)$ 阶实对称矩阵.

由归纳假设, 存在 $(n-1)$ 阶正交矩阵 $\boldsymbol{T}_{n-1}$ 使得 $\boldsymbol{T}_{n-1}^{-1}\boldsymbol{A}_{n-1}\boldsymbol{T}_{n-1} = \operatorname{diag}(\lambda_2, \ldots, \lambda_n)$ 为对角阵. 此时, 令 $\boldsymbol{T} = \boldsymbol{T}_n \begin{pmatrix} 1 & \\ & \boldsymbol{T}_{n-1} \end{pmatrix}$. 可以直接验证, $\boldsymbol{T}$ 是正交阵, 满足

$$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

(思路二) 由于实对称阵 $\boldsymbol{A}$ 的特征值都是实数, 利用 Schur 定理 (定理 6.4.21) 及其随后的注, 我们知道 $\boldsymbol{A}$ 可以通过实矩阵相似上三角化, 即存在可逆的实矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ 为实的上三角阵. 由定理 7.2.23 可知, 实的可逆矩阵 $\boldsymbol{P}$ 存在 QR 分解:

课堂上讲这个证明

$\boldsymbol{P}=\boldsymbol{Q}\boldsymbol{R}$, 其中 $\boldsymbol{Q}$ 为正交矩阵, $\boldsymbol{R}$ 为上三角矩阵. 故 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{R}^{-1}\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}\boldsymbol{R}$. 我们将选取 $\boldsymbol{T}=\boldsymbol{Q}$, 于是只需证明 $\boldsymbol{D}:=\boldsymbol{R}\boldsymbol{B}\boldsymbol{R}^{-1}=\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 是一个对角阵. 我们注意到: 一方面, $\boldsymbol{D}=\boldsymbol{R}\boldsymbol{B}\boldsymbol{R}^{-1}$ 是三个上三角矩阵的乘积, 故仍然是上三角阵; 另一方面, $\boldsymbol{D}=\boldsymbol{T}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{T}$ 是通过对对称阵 $\boldsymbol{A}$ 作相合变换得到的, 从而也是对称阵. 因此, $\boldsymbol{D}$ 为对角阵. □

注 7.3.25. *下面给出定理 7.3.24 中的正交矩阵 $\boldsymbol{T}$ 计算步骤.*

(1) 求出实对称矩阵 $\boldsymbol{A}$ 的特征多项式

$$p_{\boldsymbol{A}}(\lambda)=\det(\lambda\boldsymbol{I}_n-\boldsymbol{A})=(\lambda-\lambda_1)^{n_1}\cdots(\lambda-\lambda_s)^{n_s},$$

其中当 $i\neq j$ 时, 特征值 $\lambda_i\neq\lambda_j$, 而 n_i 为 λ_i 的代数重数.

(2) 由于 $\boldsymbol{A}$ 可相似对角化, λ_i 的代数重数 n_i 等于它的几何重数 m_i. 这说明特征子空间 $V_{\boldsymbol{A}}(\lambda_i)$ 存在一组基 $\boldsymbol{x}_{i,1},\boldsymbol{x}_{i,2},\ldots,\boldsymbol{x}_{i,n_i}$. 利用 Gram–Schmidt 正交化过程, 不妨假定 $\boldsymbol{x}_{i,1},\boldsymbol{x}_{i,2},\ldots,\boldsymbol{x}_{i,n_i}$ 是一个标准正交向量组.

(3) 观察到 $\boldsymbol{A}$ 的不同特征值之间的特征向量是相互正交, 我们得到了 $\mathbb{R}^n$ 的一组标准正交基

$$\{\,\boldsymbol{x}_{i,j}\mid 1\le i\le s, 1\le j\le n_i\,\}.$$

将这些列向量有序排列得到矩阵

$$\boldsymbol{T}=(\boldsymbol{x}_{1,1},\boldsymbol{x}_{1,2},\ldots,\boldsymbol{x}_{1,n_1},\boldsymbol{x}_{2,1},\boldsymbol{x}_{2,2},\ldots,\boldsymbol{x}_{2,n_2},\ldots,\boldsymbol{x}_{s,1},\boldsymbol{x}_{s,2},\ldots,\boldsymbol{x}_{s,n_s}),$$

则 $\boldsymbol{T}$ 是一个正交矩阵, 满足

$$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\operatorname{diag}(\lambda_1,\lambda_1,\ldots,\lambda_1,\lambda_2,\lambda_2,\ldots,\lambda_2,\ldots,\lambda_s,\lambda_s,\ldots,\lambda_s).$$

例 7.3.26. *考虑例 6.3.13 中的对称阵* $\boldsymbol{A}=\begin{pmatrix}1&2&2\\2&1&2\\2&2&1\end{pmatrix}$. *我们已经计算出, 对于特征值* 这种计算一定要熟练
$\lambda=5$, $\boldsymbol{A}$ *有特征向量* $\boldsymbol{x}_1=(1,1,1)^{\mathrm{T}}$, *对于二重的特征值* $\lambda=-1$, $\boldsymbol{A}$ *有线性无关的特征向量* $\boldsymbol{x}_2=(-1,1,0)^{\mathrm{T}}$ *和* $\boldsymbol{x}_3=(-1,0,1)^{\mathrm{T}}$. *对* $\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3$ *做正交化, 我们得到* $\mathbb{R}^3$ *的一组标准正交基*

$$\varepsilon_1=\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix},\quad \varepsilon_2=\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\end{pmatrix},\quad \varepsilon_3=\frac{1}{\sqrt{6}}\begin{pmatrix}-1\\-1\\2\end{pmatrix}.$$

由此, 若我们取

$$
\boldsymbol{T}=\begin{pmatrix}\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}}\end{pmatrix},
$$

那么 $\boldsymbol{T}$ 为正交矩阵, 满足

$$
\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\operatorname{diag}(5,-1,-1).
$$

习题 7.3.27. 设 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 都是 n 阶的实对称方阵. 证明: $\boldsymbol{AB}=\boldsymbol{BA}$ 的充要条件是存在正交方阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{-1}\boldsymbol{AP}$ 与 $\boldsymbol{P}^{-1}\boldsymbol{BP}$ 同时为对角阵. 课堂不讲, 留作思考题

证明. 由于充分性是显然的, 我们只证明必要性.

由于 $\boldsymbol{B}$ 为实对称阵, 存在正交方阵 $\boldsymbol{P}$ 使得

$$
\boldsymbol{P}^{\mathrm{T}}\boldsymbol{BP}=\operatorname{diag}(\lambda_1\boldsymbol{I}_{n_1},\dots,\lambda_s\boldsymbol{I}_{n_s}),
$$

其中 $\lambda_1,\dots,\lambda_s$ 是 $\boldsymbol{B}$ 的所有不同的特征值, n_i 是 λ_i 的代数重数, 也是其几何重数. 令 $\widetilde{\boldsymbol{A}}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}$. 则 $\boldsymbol{AB}=\boldsymbol{BA}$ 当且仅当

$$
\operatorname{diag}(\lambda_1\boldsymbol{I}_{n_1},\dots,\lambda_s\boldsymbol{I}_{n_s})\widetilde{\boldsymbol{A}}=\widetilde{\boldsymbol{A}}\operatorname{diag}(\lambda_1\boldsymbol{I}_{n_1},\dots,\lambda_s\boldsymbol{I}_{n_s}).
$$

又由于 $\lambda_1,\dots,\lambda_s$ 互不相同, 上式成立当且仅当 $\widetilde{\boldsymbol{A}}$ 为准对角形

$$
\widetilde{\boldsymbol{A}}=\operatorname{diag}(\widetilde{\boldsymbol{A}}_1,\dots,\widetilde{\boldsymbol{A}}_s),
$$

其中 $\widetilde{\boldsymbol{A}}_i$ 为 n_i 阶方阵. 由于 $\boldsymbol{A}$ 与 $\widetilde{\boldsymbol{A}}$ 相合, 故 $\widetilde{\boldsymbol{A}}$ 是实对称阵, 从而主对角线上每个子块 $\widetilde{\boldsymbol{A}}_i$ 也都是实对称阵, 于是存在 n_i 阶正交方阵 $\boldsymbol{P}_i$ 使得 $\boldsymbol{P}_i^{\mathrm{T}}\widetilde{\boldsymbol{A}}_i\boldsymbol{P}_i=\boldsymbol{D}_i$ 为对角阵. 此时,

$$
\begin{aligned}
&\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)\\
=&\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)^{\mathrm{T}}\operatorname{diag}(\widetilde{\boldsymbol{A}}_1,\dots,\widetilde{\boldsymbol{A}}_s)\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)\\
=&\operatorname{diag}(\boldsymbol{D}_1,\dots,\boldsymbol{D}_s),
\end{aligned}
$$

而同时有

$$
\begin{aligned}
&\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{BP}\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)\\
=&\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)^{\mathrm{T}}\operatorname{diag}(\lambda_1\boldsymbol{I}_{n_1},\dots,\lambda_s\boldsymbol{I}_{n_s})\operatorname{diag}(\boldsymbol{P}_1,\dots,\boldsymbol{P}_s)\\
=&\operatorname{diag}(\lambda_1\boldsymbol{I}_{n_1},\dots,\lambda_s\boldsymbol{I}_{n_s}).
\end{aligned}
$$

这说明它们确实可以同时正交相似对角化. □

第八章　实二次型

教材的 §2.2.5 简要地介绍了三维欧氏空间中的二次曲面. 由二次多项式定义, 我们有

$$a_1x^2+a_2xy+a_3xz+a_4y^2+a_5yz+a_6z^2+a_7x+a_8y+a_9z+a_{10}=0.$$

其中,

$$a_1x^2+a_2xy+a_3xz+a_4y^2+a_5yz+a_6z^2$$

为二次项, 不为零,

$$a_7x+a_8y+a_9z$$

为一次项、线性项, 而 a_{10} 为常数项. 通过适当的坐标变换 (正交变换 + 平移变换; 这两种变换都能保持向量之间的距离不变, 其中正交变换是线性变换, 而平移变换一般不是线性变换), 可以把二次曲面 (quadratic surface) 归结成几种标准曲面:

(1) 椭球面, 单叶双曲面, 双叶双曲面, 二次锥面, 椭圆柱面, 双曲柱面; (它们没有线性项)

(2) 椭圆抛物面, 双曲抛物面, 抛物柱面. (它们有线性项)

其中, 柱面型二次曲面是退化的二次曲面. 各位同学需要自行阅读教材的 §2.2.5 中的相关内容, 做到能将各类二次曲面与其标准方程熟练对应.

在这一章里, 我们将系统学习如何将二次多项式化成标准形. 特别地,

(1) 我们在多元 (n 维) 中考虑;

(2) 我们研究什么是 “标准” 的.

8.1　二次型的矩阵表示

定义 8.1.1. 关于 n 个变量 $x_1,\dots,x_n$ 的二次齐次多项式

$$Q(x_1,\dots,x_n)=\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j,$$

其中 $a_{ij} = a_{ji}$, 称为 $x_1, \ldots, x_n$ 的**二次型** (*quadratic form*) 或**二次形式**. 当所有 a_{ij} 都是实数、复数或者整数时, Q 称为**实二次型**、**复二次型**或者**整二次型**. 在本书里, 我们只考虑实二次型.

例 8.1.2. 对于

$$Q(x_1, x_2, x_3) = 3x_1^2 + x_2^2 - x_3^2 + 2x_1x_2 - x_1x_3 + 4x_2x_3,$$

有 $a_{11} = 3$, $a_{22} = 1$, $a_{33} = -1$, $a_{12} = a_{21} = \frac{1}{2} \cdot 2 = 1$, $a_{13} = a_{31} = \frac{1}{2} \cdot (-1) = -\frac{1}{2}$, $a_{23} = a_{32} = \frac{1}{2} \cdot 4 = 2$.

一般地, 实二次型可以表示成为

$$Q(x_1, \ldots, x_n) = \boldsymbol{X}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{X},$$

其中 $\boldsymbol{X} = (x_1, \ldots, x_n)^{\mathrm{T}}$, $\boldsymbol{A} = (a_{ij})_{n \times n}$. 由于 $a_{ij} = a_{ji}$, 矩阵 $\boldsymbol{A}$ 是实对称阵. 在刚刚的例子里, 我们有

$$\boldsymbol{A} = \begin{pmatrix} 3 & 1 & -1/2 \\ 1 & 1 & 2 \\ -1/2 & 2 & -1 \end{pmatrix}.$$

这样的矩阵 $\boldsymbol{A}$ 称为二次型 Q 的**矩阵**.

课堂练习 8.1.3. 写出以

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & -1 \end{pmatrix}$$

为矩阵的二次型.

注 8.1.4. 若将上面的二次型 Q 视为关于 $x_1, \ldots, x_n \in \mathbb{R}^n$ 的函数, 则相应的矩阵 $\boldsymbol{A}$ 中的元素也可以由该函数的特殊取值表示出来. 例如, 不妨设 $n = 3$, 那么, $a_{11} = Q(1, 0, 0)$, $a_{22} = Q(0, 1, 0)$, 而 $a_{12} = a_{21} = \frac{1}{2}\left(Q(1, 1, 0) - Q(1, 0, 0) - Q(0, 1, 0)\right)$.

如果把 $\boldsymbol{X} = (x_1, \ldots, x_n)^{\mathrm{T}}$ 看成 n 维线性空间 V 中的向量 $\boldsymbol{\gamma}$ 在某组基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的坐标, 那么二次型 Q 可以看成关于向量 $\boldsymbol{\gamma}$ 的函数, 从而记作

$$Q(\boldsymbol{\gamma}) \coloneqq \boldsymbol{X}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{X}.$$

此时, $\boldsymbol{A}$ 也称为二次型 $\boldsymbol{Q}$ 在基 $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ 下的**矩阵**. 同一个向量 $\boldsymbol{\gamma}$ 在不同基下的坐标可能不同. 接下来, 我们考察二次型 Q 在不同基下的矩阵的变换公式.

设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 与 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 为 $\mathbb{R}^n$ 的两组基. 设 $\boldsymbol{\gamma}\in\mathbb{R}^n$ 在这两组基下的坐标分别为 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$, 即

$$\boldsymbol{\gamma}=(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n)\boldsymbol{X}=(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n)\boldsymbol{Y}.$$

设 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 到 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 的过渡矩阵为 $\boldsymbol{P}$, 即

$$(\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n)\boldsymbol{P}.$$

则坐标之间有变换公式

$$\boldsymbol{Y}=\boldsymbol{P}^{-1}\boldsymbol{X},\qquad \text{或等价地},\qquad \boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}.$$

设二次型 Q 在 $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n$ 和 $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_n$ 下的矩阵分别为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$. 此时,

$$Q(\boldsymbol{\gamma})=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Y}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}=(\boldsymbol{P}\boldsymbol{Y})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{P}\boldsymbol{Y})=\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P})\boldsymbol{Y}.$$

故 $\boldsymbol{B}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$. 这说明二次型$\boldsymbol{Q}$ 在不同基下的矩阵是相合的.

我们知道相合关系是一种等价关系, 可以保持二次型 Q 的矩阵的秩不变, 因此, 二次型 Q 在任意一组基下的矩阵 $\boldsymbol{A}$ 的秩也被称为二次型 Q 的**秩**.

对于可逆方阵 $\boldsymbol{P}$, 我们称坐标变换

$$\boldsymbol{X} = \boldsymbol{P}\boldsymbol{Y}$$

给出了一个**可逆**的或者**满秩**的**线性变换**. 请注意, 这只是坐标的一个保数乘保加法的变换操作, 并不是我们之前提到的线性空间到自身的线性变换.

例 8.1.5. 考虑二元的二次型

$$Q(\boldsymbol{\gamma}) = 3x_1^2 - 4x_1x_2 - 4x_2^2 = (x_1, x_2)\begin{pmatrix} 3 & -2 \\ -2 & -4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

若采用可逆的线性变换 $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{Y}$, 其中 $\boldsymbol{P} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, 那么

$$\begin{aligned} Q(\boldsymbol{\gamma}) &= (x_1, x_2)\begin{pmatrix} 3 & -2 \\ -2 & -4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= (y_1, y_2)\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{\mathrm{T}}\begin{pmatrix} 3 & -2 \\ -2 & -4 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= (y_1, y_2)\begin{pmatrix} 3 & 1 \\ 1 & -5 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= 3y_1^2 + 2y_1y_2 - 5y_2^2. \end{aligned}$$

当然, 由于 $x_1 = y_1 + y_2$, $x_2 = y_2$, 最后的结果也可由

$$Q = 3(y_1 + y_2)^2 - 4(y_1 + y_2)y_2 - 4y_2^2$$

直接化简得到.

8.2 二次型的标准形

假定二次型 $Q(x_1, \ldots, x_n) = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$ 通过某个坐标的可逆线性变换 $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{Y}$ 化为不含混合项 (交叉项) 的形式:

$$Q(x_1, \ldots, x_n)|_{\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}} \coloneqq \boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P})\boldsymbol{Y} = \mu_1 y_1^2 + \mu_2 y_2^2 + \cdots + \mu_n y_n^2,$$

即

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_n).$$

我们称该形式为二次型 Q 的**标准形** (canonical form).

注 8.2.1. 标准形显然不唯一. 例如

$$Q(x_1, x_2) = 2x_1^2 + 3x_2^2 = y_1^2 + y_2^2,$$

其中 $y_1 = \sqrt{2}x_1$, $y_2 = \sqrt{3}x_2$.

本节的主题就是讨论如何将二次型化为标准形.

主轴化方法 很显然, 将二次型化简成标准形的过程, 就是寻找可逆矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{B} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ 为对角阵的过程. 由于 $\boldsymbol{A}$ 为实对称阵, 我们可以找到正交矩阵 $\boldsymbol{P}$ 将其相似对角化. 由于 $\boldsymbol{P}$ 为正交阵, 对角阵 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$, 且其主对角线上的元素为 $\boldsymbol{A}$ 的所有特征值. 借此, 我们证明了

定理 8.2.2 (主轴定理)**.** 任何实二次型都可以通过坐标的正交变换化为标准形

$$Q(x_1, \ldots, x_n)|_{\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

这儿的 $\lambda_1, \ldots, \lambda_n$ 为 $\boldsymbol{A}$ 的所有特征值.

例 8.2.3. 考虑二次型 $Q(\boldsymbol{\gamma}) = x_1^2 + x_2^2 + x_3^2 + 4x_1x_2 + 4x_1x_3 + 4x_2x_3 = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$, 其中

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}.$$

在例 *7.3.26* 中, 我们已经找到了正交矩阵

$$\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix},$$

满足

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(5, -1, -1).$$

这说明

$$Q(\boldsymbol{\gamma}) = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{Y}^{\mathrm{T}}\,\mathrm{diag}(5, -1, -1)\boldsymbol{Y} = 5y_1^2 - y_2^2 - y_3^2,$$

从而得到 Q 的一个标准形.

例 8.2.4. 设二次型 $Q(x_1,x_2,x_3)=x_1^2+x_2^2+x_3^2+2\alpha x_1x_2+2\beta x_2x_3+2x_1x_3$ 经过正交变换 $\boldsymbol{X}=\boldsymbol{PY}$ 化为 $Q=y_2^2+2y_3^2$, 其中 $\boldsymbol{X}=(x_1,x_2,x_3)^{\mathrm{T}}$, $\boldsymbol{Y}=(y_1,y_2,y_3)^{\mathrm{T}}$, $\boldsymbol{P}$ 是正交阵. 试求 α,β 的值.

证明. 变换前后二次型的矩阵分别为

$$\boldsymbol{A}=\begin{pmatrix}1&\alpha&1\\ \alpha&1&\beta\\ 1&\beta&1\end{pmatrix}\qquad 和 \qquad \boldsymbol{B}=\begin{pmatrix}0&0&0\\0&1&0\\0&0&2\end{pmatrix},$$

且有 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}=\boldsymbol{B}$. 因为 $\boldsymbol{P}$ 是正交阵, 从而有 $\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{B}$, 于是有特征多项式 $p_{\boldsymbol{A}}(\lambda)=p_{\boldsymbol{B}}(\lambda)$, 即

这儿若只是比较两个方阵的迹与行列式, 显然是不够的

$$\lambda^3-3\lambda^2+(2-\alpha^2-\beta^2)\lambda+(\alpha^2-2\alpha\beta+\beta^2)=\lambda^3-3\lambda^2+2\lambda.$$

比较系数, 我们有 $2-\alpha^2-\beta^2=2$, 且 $\alpha^2-2\alpha\beta+\beta^2=0$. 该方程组的解为 $\alpha=\beta=0$. □

例 8.2.5. 设 $\boldsymbol{A}$ 为 3 阶实对称阵, 特征值为 $1,1,2$, 且有特征向量 $\boldsymbol{\alpha}_1=(1,0,1)^{\mathrm{T}}$, $\boldsymbol{\alpha}_2=(0,1,1)^{\mathrm{T}}$. 求正交变换将此实对称阵对应的二次型 $Q=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{AX}$ 化为标准形.

解. 注意到题目中的两个特征向量不正交, 故它们同属于特征值 1. 此时, 设 $(u,v,w)^{\mathrm{T}}$ 是属于 2 的一个特征向量, 由正交性可知, $u+w=0$, $v+w=0$. 于是不妨设 $\boldsymbol{\alpha}_3=(1,1,-1)^{\mathrm{T}}$. 对 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 用 Gram–Schmidt 正交化, 可以得到标准正交基 $\boldsymbol{\varepsilon}_1=(1/\sqrt{2},0,1/\sqrt{2})^{\mathrm{T}}$, $\boldsymbol{\varepsilon}_2=(-1/\sqrt{6},2/\sqrt{6},1/\sqrt{6})^{\mathrm{T}}$, $\boldsymbol{\varepsilon}_3=(1/\sqrt{3},1/\sqrt{3},-1/\sqrt{3})^{\mathrm{T}}$. 于是所用的正交变换为 $\boldsymbol{X}=\boldsymbol{PY}$, 其中 $\boldsymbol{P}=\begin{pmatrix}\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{6}}&\frac{1}{\sqrt{3}}\\ 0&\frac{2}{\sqrt{6}}&\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}&-\frac{1}{\sqrt{3}}\end{pmatrix}$, 而在正交变换后得到的二次型为 $Q(y_1,y_2,y_3)=y_1^2+y_2^2+2y_3^2$. □

这样的 $\boldsymbol{P}$ 并不唯一

例 8.2.6 (Rayleigh 原理)**.** 设 n 元二次型 $Q(x_1,\dots,x_n)=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{AX}$, 其中 $\boldsymbol{X}=(x_1,\dots,x_n)^{\mathrm{T}}$, 而实对称阵 $\boldsymbol{A}$ 的特征值满足 $\lambda_1\le\lambda_2\le\cdots\le\lambda_n$. 证明: 在条件 $x_1^2+x_2^2+\cdots+x_n^2=1$ 的条件下, $Q(x_1,\dots,x_n)$ 的最小值为 λ_1, 最大值为 λ_n.

证明. 对于二次型 Q, 必有正交变换 $\boldsymbol{X}=\boldsymbol{PY}$ 使得

$$Q|_{\boldsymbol{X}=\boldsymbol{PY}}=\lambda_1y_1^2+\lambda_2y_2^2+\cdots+\lambda_ny_n^2.$$

由特征值之间的大小关系, 知

$$\lambda_1(y_1^2+y_2^2+\cdots+y_n^2)\le Q\le\lambda_n(y_1^2+y_2^2+\cdots+y_n^2).$$

由于 $\boldsymbol{P}$ 为正交矩阵, 我们有

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = (\boldsymbol{PY})^{\mathrm{T}}(\boldsymbol{PY}) = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{PY} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}^{-1}\boldsymbol{PY} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y},$$

请牢记坐标的正交变换的这一性质

这说明, $x_1^2+x_2^2+\cdots+x_n^2=1$ 当且仅当 $y_1^2+y_2^2+\cdots+y_n^2=1$, 从而在该条件下, 我们有

$$\lambda_1 \le Q(x_1,\ldots,x_n) \le \lambda_n.$$

显然, 上界 λ_n 在 $\boldsymbol{Y}=(0,0,\ldots,0,1)^{\mathrm{T}}$ 时可以取到, 下界 λ_1 在 $\boldsymbol{Y}=(1,0,0,\ldots,0)^{\mathrm{T}}$ 时可以取到. □

例 8.2.7. 设 $\boldsymbol{A}=(a_{ij})$ 是 n 阶实对称阵, 它的 n 个特征值为 $\lambda_1\le\lambda_2\le\cdots\le\lambda_n$. 证明: $\lambda_1\le a_{ii}\le\lambda_n$, 其中 $i=1,2,\ldots,n$.

证明. 设 $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$ 为 $\mathbb{R}^n$ 的自然基, 则 $a_{ii}=\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{e}_i$. 接下来, 我们只需要利用上题的结果. □

配方法 由于正交阵 $\boldsymbol{P}$ 的计算比较复杂, 在将二次型化为标准形时, 若没有特别要求, 我们有时仅仅通过配方的方法, 消除掉二次型中的交叉项. 这样的好处是计算简便, 不利之处在于由于做了不同方向的不同比例的拉伸, 画出的 (可能是高维的) 几何图形的形状会和原来的形状不一致.

注 8.2.8 (配方中的两种基本情形)**.** 设

$$Q(x_1,x_2,\ldots,x_n)=a_{11}x_1^2+2a_{12}x_1x_2+\cdots+2a_{1n}x_1x_n+(\text{与 } x_1 \text{ 无关的项}).$$

(1) 假设 $a_{11}\ne 0$. 此时,

$$\begin{aligned}Q(x_1,x_2,\ldots,x_n) &= a_{11}\left(x_1+\frac{a_{12}}{a_{11}}x_2+\cdots+\frac{a_{1n}}{a_{11}}x_n\right)^2+(\text{配方后新的项})\\ &\quad +(\text{之前与 } x_1 \text{ 无关的项})\\ &= a_{11}\left(x_1+\frac{a_{12}}{a_{11}}x_2+\cdots+\frac{a_{1n}}{a_{11}}x_n\right)^2+(\text{与 } x_1 \text{ 无关的项}).\end{aligned}$$

在这种情况下, 我们会设

$$y_1=x_1+\frac{a_{12}}{a_{11}}x_2+\cdots+\frac{a_{1n}}{a_{11}}x_n,\quad y_2=x_2,\quad \ldots,\quad y_n=x_n,$$

从而

$$x_1=y_1-\frac{a_{12}}{a_{11}}y_2-\cdots-\frac{a_{1n}}{a_{11}}y_n,\quad x_2=y_2,\quad \ldots,\quad x_n=y_n.$$

相应的坐标变换的矩阵形式为

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & -\frac{a_{12}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

(2) 假设 $a_{12} \neq 0$, 而 $\underset{\sim}{a_{11} = a_{22} = \cdots = a_{nn} = 0}$. 我们利用等式

$$x_1 x_2 = \frac{1}{4}\left[(x_1 + x_2)^2 - (x_1 - x_2)^2\right] = \left(\frac{x_1 + x_2}{2}\right)^2 - \left(\frac{x_1 - x_2}{2}\right)^2,$$

令

$$y_1 = \frac{x_1 + x_2}{2}, \quad y_2 = \frac{x_1 - x_2}{2}, \quad y_3 = x_3, \quad \ldots, \quad y_n = x_n,$$

从而

$$x_1 = y_1 + y_2, \quad x_2 = y_1 - y_2, \quad x_3 = y_3, \quad \ldots, \quad x_n = y_n.$$

这种情形下的坐标的基本变换的矩阵形式为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \left(\begin{array}{cc|ccc} 1 & 1 & & & \\ 1 & -1 & & & \\ \hline & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{array}\right) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}.$$

在此之后, 我们可以化归为第一种情形, 其中 y_1^2 前的系数为原式

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = (a_{11} + 2a_{12} + a_{22})y_1^2 + \cdots$$

中 y_1^2 的系数, 即 $2a_{12}$, 不为零. 注意, 这里仅仅 $a_{11} = 0$, $a_{12} \neq 0$, 是不够的.

例 8.2.9. 考察 $Q(x_1, x_2, x_3) = x_1^2 + 2x_3^2 + x_1x_2$.

解. 直接配方, 有

$$\begin{aligned} Q(x_1, x_2, x_3) &= \left(x_1 + \frac{1}{2}x_2\right)^2 - \frac{1}{4}x_2^2 + 2x_3^2 \\ &= y_1^2 - \frac{1}{4}y_2^2 + 2y_3^2, \qquad \text{(二次型的标准形)} \end{aligned}$$

其中

$$\begin{cases} y_1 = x_1 + \frac{1}{2}x_2, \\ y_2 = x_2, \\ y_3 = x_3, \end{cases} \Leftrightarrow \begin{cases} x_1 = y_1 - \frac{1}{2}y_2, \\ x_2 = y_2, \\ x_3 = y_3, \end{cases} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

二次型对应的矩阵为 $\boldsymbol{A} = \begin{pmatrix} 1 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. 若我们记坐标变换矩阵 $\boldsymbol{P} = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, 则 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(1, -1/4, 2)$, 与我们上面计算得到的二次型的标准形相对应. □

例 8.2.10. *考察* $Q(x_1, x_2, x_3) = 2x_1x_2 + 4x_1x_3$.

解. 直接配方, 有

$$Q(x_1, x_2, x_3) = 2\left(\left(\frac{x_1 + x_2}{2}\right)^2 - \left(\frac{x_1 - x_2}{2}\right)^2\right) + 4x_1x_3.$$

若令

$$\begin{cases} y_1 = \frac{x_1+x_2}{2} \\ y_2 = \frac{x_1-x_2}{2} \\ y_3 = x_3 \end{cases} \Leftrightarrow \begin{cases} x_1 = y_1 + y_2 \\ x_2 = y_1 - y_2 \\ x_3 = y_3 \end{cases} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

则

$$\begin{aligned} Q =& 2(y_1 + y_2)(y_1 - y_2) + 4(y_1 + y_2)y_3 \\ =& 2y_1^2 + 4y_1y_3 - 2y_2^2 + 4y_2y_3 \\ =& 2(y_1 + y_3)^2 - 2y_2^2 + 4y_2y_3 - 2y_3^2 \\ =& 2(y_1 + y_3)^2 - 2(y_2 - y_3)^2 \\ =& 2z_1^2 - 2z_2^2. \qquad (\text{二次型的标准形}) \end{aligned}$$

其中

$$\begin{cases} z_1 = y_1 + y_3, \\ z_2 = y_2 - y_3, \\ z_3 = y_3. \end{cases} \Leftrightarrow \begin{cases} y_1 = z_1 - z_3, \\ y_2 = z_2 + z_3, \\ y_3 = z_3 \end{cases} \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

此时

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}1&1&0\\1&-1&0\\0&0&1\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\begin{pmatrix}1&1&0\\1&-1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&-1\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}z_1\\z_2\\z_3\end{pmatrix}=\begin{pmatrix}1&1&0\\1&-1&-2\\0&0&1\end{pmatrix}\begin{pmatrix}z_1\\z_2\\z_3\end{pmatrix}.$$

二次型对应的矩阵为 $\boldsymbol{A}=\begin{pmatrix}0&1&2\\1&0&0\\2&0&0\end{pmatrix}$. 若令 $\boldsymbol{P}=\begin{pmatrix}1&1&0\\1&-1&-2\\0&0&1\end{pmatrix}$, 则 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\operatorname{diag}(2,-2,0)$, 与我们上面计算得到的二次型的标准形相对应. □

矩阵的初等变换法 在主轴化定理 8.2.2 中我们已经见到, 存在可逆矩阵 $\boldsymbol{P}$ 将二次型的矩阵 $\boldsymbol{A}$ 相合对角化. 由于 $\boldsymbol{P}$ 可以表示成一系列初等矩阵的乘积, 于是我们有

定理 8.2.11. *对每个实对称阵 $\boldsymbol{A}$, 存在同阶的初等矩阵 $\boldsymbol{P}_1,\boldsymbol{P}_2,\ldots,\boldsymbol{P}_s$ 使得*

$$\boldsymbol{P}_S^{\mathrm{T}}\cdots\boldsymbol{P}_2^{\mathrm{T}}\boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}_1\boldsymbol{P}_2\cdots\boldsymbol{P}_s$$

为对角阵.

由此出发, 为了计算二次型的标准形, 我们将对实矩阵采取如下的操作 (其中矩阵 $\boldsymbol{A}$ 为对称阵)

$$\begin{pmatrix}\boldsymbol{A}\\\boldsymbol{B}\end{pmatrix}\longrightarrow\begin{pmatrix}\boldsymbol{AP}\\\boldsymbol{BP}\end{pmatrix}\longrightarrow\begin{pmatrix}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}\\\boldsymbol{BP}\end{pmatrix},$$

即, 我们先对整个矩阵作列变换 (右乘 $\boldsymbol{P}$), 再对分块矩阵中上面的方阵做相应的行变换 (左乘 $\boldsymbol{P}^{\mathrm{T}}$).

注 8.2.12. (1) *若 $\boldsymbol{P}=\boldsymbol{S}_{ij}:c_i\leftrightarrow c_j$, 则 $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{S}_{ij}:r_i\leftrightarrow r_j$.*

(2) *若 $\boldsymbol{P}=\boldsymbol{D}_i(\lambda):\lambda c_i$, 则 $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{D}_i(\lambda):\lambda r_i$.*

(3) *若 $\boldsymbol{P}=\boldsymbol{T}_{ij}(\lambda):\lambda c_i\to c_j$, 则 $\boldsymbol{P}^{\mathrm{T}}=\boldsymbol{T}_{ji}(\lambda):\lambda r_i\to r_j$.*

若最开始 $\boldsymbol{B}=\boldsymbol{I}$, 而经过一系列这样的操作后, $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}$ 为对角阵, 则分块矩阵中下面的矩阵 $\boldsymbol{BP}=\boldsymbol{P}$ 为所求.

例 8.2.13. *用初等变换的方法把二次型 $Q(x_1,x_2,x_3)=x_1^2+4x_1x_3-x_2^2+2x_2x_3+x_3^2$ 化为标准形.*

解. 二次型的矩阵为 $\boldsymbol{A}=\begin{pmatrix}1&0&2\\0&-1&1\\2&1&1\end{pmatrix}$. 于是,

$$\begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{I}\end{pmatrix}=\left(\begin{array}{ccc}1&0&\boxed{2}\\0&-1&1\\2&1&1\\ \hline 1&0&0\\0&1&0\\0&0&1\end{array}\right)\xrightarrow{-2c_1\to c_3}\left(\begin{array}{ccc}1&0&0\\0&-1&1\\2&1&-3\\ \hline 1&0&-2\\0&1&0\\0&0&1\end{array}\right)\xrightarrow{-2r_1\to r_3}\left(\begin{array}{ccc}1&0&0\\0&-1&\boxed{1}\\0&1&-3\\ \hline 1&0&-2\\0&1&0\\0&0&1\end{array}\right)$$

$$\xrightarrow{c_2\to c_3}\left(\begin{array}{ccc}1&0&0\\0&-1&0\\0&1&-2\\ \hline 1&0&-2\\0&1&1\\0&0&1\end{array}\right)\xrightarrow{r_2\to r_3}\left(\begin{array}{ccc}1&0&0\\0&-1&0\\0&0&-2\\ \hline 1&0&-2\\0&1&1\\0&0&1\end{array}\right).$$

最后的分块矩阵中, 上面的方阵为对角阵, 故下面的矩阵 $\boldsymbol{P}=\begin{pmatrix}1&0&-2\\0&1&1\\0&0&1\end{pmatrix}$ 为所求, 满足 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\operatorname{diag}(1,-1,-2)$. 此时的坐标变换为

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}1&0&-2\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix},$$

而二次型为

$$Q(x_1,x_2,x_3)|_{\boldsymbol{X}=\boldsymbol{PY}}=y_1^2-y_2^2-2y_3^2.$$

□

例 8.2.14. 用初等变换的方法把二次型 $Q(x_1,x_2,x_3)=4x_1x_2+2x_1x_3+x_2^2+2x_2x_3+x_3^2$ 化为标准形.

解. 二次型的矩阵为 $\boldsymbol{A}=\begin{pmatrix}0&2&1\\2&1&1\\1&1&1\end{pmatrix}$. 与例 8.2.13 中的情形略有不同, 这儿的 $\boldsymbol{A}$ 的

$(1,1)$ 元素为零, 但是我们注意到其 $(2,2)$ 位置的元素非零. 于是,

$$\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{I} \end{pmatrix} = \left(\begin{array}{ccc} 0 & 2 & 1 \\ \boxed{2} & 1 & \boxed{1} \\ 1 & 1 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \xrightarrow[-c_2 \to c_3]{-2c_2 \to c_1} \left(\begin{array}{ccc} -4 & 2 & -1 \\ 0 & 1 & 0 \\ -1 & 1 & 0 \\ \hline 1 & 0 & 0 \\ -2 & 1 & -1 \\ 0 & 0 & 1 \end{array}\right) \xrightarrow[-r_2 \to r_3]{-2r_2 \to r_1} \left(\begin{array}{ccc} -4 & 0 & \boxed{-1} \\ 0 & 1 & 0 \\ -1 & 0 & 0 \\ \hline 1 & 0 & 0 \\ -2 & 1 & -1 \\ 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{-\frac{1}{4}c_1 \to c_3} \left(\begin{array}{ccc} -4 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & \frac{1}{4} \\ \hline 1 & 0 & -\frac{1}{4} \\ -2 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{array}\right) \xrightarrow{-\frac{1}{4}r_1 \to r_3} \left(\begin{array}{ccc} -4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{4} \\ \hline 1 & 0 & -\frac{1}{4} \\ -2 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{array}\right).$$

最后的分块矩阵中, 上面的方阵为对角阵, 故下面的矩阵 $\boldsymbol{P} = \begin{pmatrix} 1 & 0 & -\frac{1}{4} \\ -2 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}$ 为所求, 满足 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(-4, 1, \frac{1}{4})$. 此时的坐标变换为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -\frac{1}{4} \\ -2 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

而二次型为

$$Q(x_1, x_2, x_3)|_{\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}} = -4y_1^2 + y_2^2 + \tfrac{1}{4}y_3^2. \qquad \square$$

例 8.2.15. *用初等变换的方法把二次型* $Q(x_1, x_2, x_3) = 2x_1x_2 + 4x_1x_3$ *化为标准形.*

解. 二次型的矩阵为 $\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix}$. 与例 8.2.14 中的情形不同, 这儿的 $\boldsymbol{A}$ 的对角线上的元素全为零. 我们的策略是用 $\boldsymbol{A}$ 的非零的 $(1,2)$ 与 $(2,1)$ 元素造出一个对角线上的

非零元. 于是

$$\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{I} \end{pmatrix} = \begin{pmatrix} 0 & \boxed{1} & 2 \\ \boxed{1} & 0 & 0 \\ 2 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{c_2 \to c_1} \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{r_2 \to r_1} \begin{pmatrix} 2 & \boxed{1} & \boxed{2} \\ 1 & 0 & 0 \\ 2 & 0 & 0 \\ \hline 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[-c_1 \to c_3]{-\frac{1}{2}c_1 \to c_2}$$

$$\begin{pmatrix} 2 & 0 & 0 \\ 1 & -\frac{1}{2} & -1 \\ 2 & -1 & -2 \\ \hline 1 & -\frac{1}{2} & -1 \\ 1 & \frac{1}{2} & -1 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow[-r_1 \to r_3]{-\frac{1}{2}r_1 \to r_2} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & \boxed{\text{-1}} \\ 0 & -1 & -2 \\ \hline 1 & -\frac{1}{2} & -1 \\ 1 & \frac{1}{2} & -1 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{-2c_2 \to c_3} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & -1 & 0 \\ \hline 1 & -\frac{1}{2} & 0 \\ 1 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{-2r_2 \to r_3} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \\ \hline 1 & -\frac{1}{2} & 0 \\ 1 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

最后的分块矩阵中, 上面的方阵为对角阵, 故下面的矩阵 $\boldsymbol{P} = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 1 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \end{pmatrix}$ 为所求, 满足 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(2, -1/2, 0)$. 此时的坐标变换为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ 1 & \frac{1}{2} & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

而二次型为

$$Q(x_1, x_2, x_3)|_{\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}} = 2y_1^2 - \frac{1}{2}y_2^2. \qquad \square$$

而二次型为

$$Q(x_1,x_2,x_3)|_{\boldsymbol{X}=\boldsymbol{PY}}=2y_1^2-\tfrac{1}{2}y_2^2. \qquad \square$$

注 8.2.16. (1) 由于初等变换的选取的不同, 最终结果中的标准形可能不同.

(2) 我们也可以把矩阵写成 $(\boldsymbol{A},\boldsymbol{I})$ 的形式, 同时对 $\boldsymbol{A},\boldsymbol{I}$ 做行变换, 然后只对 $\boldsymbol{A}$ 做相应列变换. 变换的最终目标是得到 $(\boldsymbol{D},\boldsymbol{P}^{\mathrm{T}})$,* 其中 $\boldsymbol{D}$ 为对角阵. 做法与我们上面的操作是等价的, 具体例子可以参考书上例 *8.2.4*. 学生只需要熟悉一种方法即可.

注 8.2.17. 本节介绍了三种方法用于将二次型化为标准形. 其中的主轴化的方法 *(*坐标的正交变换*)* 是最重要的, 要求大家熟练掌握. 另一方面, 在下一节中, 我们会谈到二次型的正负惯性指数, 利用配方法容易帮助大家快速找到这些重要的指标.

8.3 相合不变量与分类

之前我们已经证明了任意实二次型都可以化成标准形, 但是标准形不唯一. 例如, 二次型 $Q(x_1,x_2,x_3)=4x_1^2-3x_2^2$ 已是 Q 的标准形. 但是若采用坐标变换

$$\begin{cases} y_1=2x_1,\\ y_2=\sqrt{3}x_2,\\ y_3=x_3, \end{cases}$$

则 $Q(y_1,y_2,y_3)=y_1^2-y_2^2$ 是 Q 的另外一个标准形. 那么, 什么是二次型 "最终极" 的标准形呢? 等价地, 实对称阵可以通过相合变换得到一个什么形式的最简单的对角阵呢?

定理 8.3.1 (惯性定理)**.** 设 $\boldsymbol{A}$ 是一个 n 阶实对称矩阵, 则存在可逆矩阵 $\boldsymbol{P}$, 使得

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\begin{pmatrix}\boldsymbol{I}_r & & \\ & -\boldsymbol{I}_s & \\ & & \boldsymbol{O}\end{pmatrix}. \tag{8.1}$$

并且这里的 r 与 s 是由 $\boldsymbol{A}$ 唯一决定的.

显然, 在等式 (8.1) 中, $r+s=\operatorname{rank}(\boldsymbol{A})$. 我们将把等式右边的矩阵称为 $\boldsymbol{A}$ 的**规范形** (normal form), 而相应的二次型称为对应于 $\boldsymbol{A}$ 的**规范二次型**. 上面的定理说明 $\boldsymbol{A}$ 的规范形是存在且唯一的.

*注意这儿是 $\boldsymbol{P}^{\mathrm{T}}$, 而不再是 $\boldsymbol{P}$

证明. **存在性的证明.** 在前面一节中, 我们已经见到, 对于实对称阵 $\boldsymbol{A}$, 存在实矩阵 $\boldsymbol{P}_0$, 使得 $\boldsymbol{P}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}_0 = \operatorname{diag}(\mu_1, \ldots, \mu_n)$ 为对角阵. 通过交换 $\boldsymbol{P}_0$ 的列向量组的排列顺序, 我们不妨假设这些对角线上的元素满足

$$\mu_1, \ldots, \mu_r > 0, \quad \mu_{r+1}, \ldots, \mu_{r+s} < 0, \quad \mu_{r+s+1} = \cdots = \mu_n = 0.$$

若令

$$\boldsymbol{P} = \boldsymbol{P}_0 \operatorname{diag}\left(\frac{1}{\sqrt{\mu_1}}, \ldots, \frac{1}{\sqrt{\mu_r}}, \frac{1}{\sqrt{-\mu_{r+1}}}, \ldots, \frac{1}{\sqrt{-\mu_{r+s}}}, 1, \ldots, 1\right),$$

则有 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{I}_r & & \\ & -\boldsymbol{I}_s & \\ & & \boldsymbol{O} \end{pmatrix}$.

唯一性的证明. 假设 $\begin{pmatrix} \boldsymbol{I}_r & & \\ & -\boldsymbol{I}_s & \\ & & \boldsymbol{O} \end{pmatrix}$ 与 $\begin{pmatrix} \boldsymbol{I}_p & & \\ & -\boldsymbol{I}_q & \\ & & \boldsymbol{O} \end{pmatrix}$ 都是 $\boldsymbol{A}$ 的规范形, 则 $r + s = p + q = \operatorname{rank}(\boldsymbol{A})$. 考虑二次型 $Q(x_1, \ldots, x_n) \coloneqq \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$. 此时, 存在两个不同的坐标的可逆变换 $\boldsymbol{X} = \boldsymbol{P}_1\boldsymbol{Y}$ 和 $\boldsymbol{X} = \boldsymbol{P}_2\boldsymbol{Z}$ 分别使得

$$\begin{aligned} Q(\boldsymbol{X})|_{\boldsymbol{X}=\boldsymbol{P}_1\boldsymbol{Y}} &= y_1^2 + \cdots + y_r^2 - y_{r+1}^2 - \cdots - y_{r+s}^2, \\ Q(\boldsymbol{X})|_{\boldsymbol{X}=\boldsymbol{P}_2\boldsymbol{Z}} &= z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_{p+q}^2. \end{aligned}$$

若 $r \neq p$, 不妨假设 $r < p$. 此时, $\boldsymbol{Y} = \boldsymbol{P}_1^{-1}\boldsymbol{X}$ 与 $\boldsymbol{Z} = \boldsymbol{P}_2^{-1}\boldsymbol{X}$. 不妨设 $\boldsymbol{P}_1^{-1}$ 的行向量依次为 $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n$, 而 $\boldsymbol{P}_2^{-1}$ 的行向量依次为 $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_n$. 则对任意 i, 我们有 $y_i = \boldsymbol{B}_i\boldsymbol{X}$, $z_i = \boldsymbol{C}_i\boldsymbol{X}$. 接下来, 我们来考察如下特别挑选的齐次线性方程组

$$\begin{cases} y_1 = 0, \\ \quad\vdots \\ y_r = 0, \\ z_{p+1} = 0, \\ \quad\vdots \\ z_n = 0, \end{cases} \qquad \Leftrightarrow \qquad \begin{pmatrix} \boldsymbol{B}_1 \\ \vdots \\ \boldsymbol{B}_r \\ \boldsymbol{C}_{p+1} \\ \vdots \\ \boldsymbol{C}_n \end{pmatrix} \boldsymbol{X} = \boldsymbol{0},$$

其中系数矩阵为 $(r + (n-p)) \times n$ 维的. 由于 $r < p$, 这导致 $r + (n-p) < n$. 这意味

着上面的齐次线性方程组的系数矩阵不是列满秩的, 从而存在非零解 $\boldsymbol{X}_0$. 此时,

$$\boldsymbol{Y}_0=\boldsymbol{P}_1^{-1}\boldsymbol{X}_0=\begin{pmatrix}0\\ \vdots\\ 0\\ y_{r+1}\\ \vdots\\ y_n\end{pmatrix}\neq\boldsymbol{0},\qquad \boldsymbol{Z}_0=\boldsymbol{P}_2^{-1}\boldsymbol{X}_0=\begin{pmatrix}z_1\\ \vdots\\ z_p\\ 0\\ \vdots\\ 0\end{pmatrix}\neq\boldsymbol{0}.$$

则我们同时有

$$\begin{aligned}Q(\boldsymbol{X}_0)=&Q|_{\boldsymbol{X}=\boldsymbol{P}_1\boldsymbol{Y}}(\boldsymbol{Y}_0)=-y_{r+1}^2-\cdots-y_{r+s}^2\leq 0,\quad (r+s\text{ 可能严格小于 }n,\text{ 从而 }Q\text{ 可能为 }0)\\ Q(\boldsymbol{X}_0)=&Q|_{\boldsymbol{X}=\boldsymbol{P}_2\boldsymbol{Z}}(\boldsymbol{Z}_0)=z_1^2+\cdots+z_p^2>0,\quad (p>r\geq 0,\text{ 从而 }Q\text{ 严格大于 }0)\end{aligned}$$

矛盾. 故 $r=p$, $s=q$, 由矩阵 $\boldsymbol{A}$ 唯一确定. □

定理 8.3.1 中的 r 称为 $\boldsymbol{A}$ 的**正惯性指数** (positive index of inertia), s 称为 $\boldsymbol{A}$ 的**负惯性指数** (negative index of inertia)*. 由上面定理的证明可知, 这两个指数分别是 $\boldsymbol{A}$ 的任意一个标准形中正项的项数与负项的项数, 从而由主轴定理 (定理 8.2.2) 可知, 它们也分别是 $\boldsymbol{A}$ 的特征值的正项与负项的个数. 相应地, $r-s$ 称为 $\boldsymbol{A}$ 的**符号差** (signature). 这儿的 r, s 和 $r-s$ 也分别被称为 $\boldsymbol{A}$ 所对应的二次型 Q 的**正惯性指数**、**负惯性指数**和**符号差**.

请想清楚原因

综合前面的结果, 我们不难得到如下的定理.

定理 8.3.2. 任何实对称阵都可以 (实) 相合等价于一个实对角阵, 并且两个相同维数的实对称阵相合等价, 当且仅当它们的正、负惯性指数相等, 当且仅当它们的秩和符号差相等.

请想清楚原因

例 8.3.3. 设 $\boldsymbol{A}$ 为 n 阶实可逆矩阵, 若 $\boldsymbol{A}$ 与 $-\boldsymbol{A}$ 相合, 则 n 为偶数. 进一步地, 若 $\boldsymbol{A}$ 为实对称矩阵, 则 $\boldsymbol{A}$ 的正惯性指数为 $\frac{n}{2}$.

证明. 由于 $\boldsymbol{A}$ 与 $-\boldsymbol{A}$ 相合, 存在可逆矩阵 $\boldsymbol{P}$ 使得 $-\boldsymbol{A}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$, 故 $|-\boldsymbol{A}|=\left|\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}\right|$, 即 $(-1)^n|\boldsymbol{A}|=|\boldsymbol{A}|\cdot|\boldsymbol{P}|^2$. 又 $|\boldsymbol{A}|\neq 0$, 故 $(-1)^n=|\boldsymbol{P}|^2>0$, 从而 n 为偶数.

若进一步地, $\boldsymbol{A}$ 为实对称阵, 关于惯性指数, 我们显然有 $r_{\boldsymbol{A}}=s_{-\boldsymbol{A}}$, $s_{\boldsymbol{A}}=r_{-\boldsymbol{A}}$. 而若 $\boldsymbol{A}$ 与 $-\boldsymbol{A}$ 相合, 则 $r_{\boldsymbol{A}}=r_{-\boldsymbol{A}}$, $s_{\boldsymbol{A}}=s_{-\boldsymbol{A}}$, 从而皆等于 $\frac{n}{2}$. □

*因此, 有些教材里分别选用 p 和 n 来表示实对称阵的正、负惯性指数

例 8.3.4. 设 n 阶实对称阵 $\boldsymbol{A}$, $\boldsymbol{B}$ 和 $\boldsymbol{A}+\boldsymbol{B}$ 的正惯性指数分别为 $r_{\boldsymbol{A}}$, $r_{\boldsymbol{B}}$ 和 $r_{\boldsymbol{A}+\boldsymbol{B}}$. 证明: $r_{\boldsymbol{A}}+r_{\boldsymbol{B}}\geq r_{\boldsymbol{A}+\boldsymbol{B}}$.

证明. 记 $\boldsymbol{A}$, $\boldsymbol{B}$ 和 $\boldsymbol{A}+\boldsymbol{B}$ 的正惯性指数分别为 $r_{\boldsymbol{A}}$, $r_{\boldsymbol{B}}$ 和 $r_{\boldsymbol{A}+\boldsymbol{B}}$, 负惯性指数分别为 $s_{\boldsymbol{A}}$, $s_{\boldsymbol{B}}$ 和 $s_{\boldsymbol{A}+\boldsymbol{B}}$. 于是, 存在可逆矩阵 $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ 和 $\boldsymbol{P}_3$ 使得

$$\begin{aligned}
\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|_{\boldsymbol{X}=\boldsymbol{P}_1\boldsymbol{Y}} =&y_1^2+\cdots+y_{r_{\boldsymbol{A}}}^2-y_{r_{\boldsymbol{A}}+1}^2-\cdots-y_{r_{\boldsymbol{A}}+s_{\boldsymbol{A}}}^2,\\
\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{X}|_{\boldsymbol{X}=\boldsymbol{P}_2\boldsymbol{Z}} =&z_1^2+\cdots+z_{r_{\boldsymbol{B}}}^2-z_{r_{\boldsymbol{B}}+1}^2-\cdots-z_{r_{\boldsymbol{B}}+s_{\boldsymbol{B}}}^2,\\
\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{X}|_{\boldsymbol{X}=\boldsymbol{P}_3\boldsymbol{W}} =&w_1^2+\cdots+w_{r_{\boldsymbol{A}+\boldsymbol{B}}}^2-w_{r_{\boldsymbol{A}+\boldsymbol{B}}+1}^2-\cdots-w_{r_{\boldsymbol{A}+\boldsymbol{B}}+s_{\boldsymbol{A}+\boldsymbol{B}}}^2.
\end{aligned}$$

用反证法, 假设 $r_{\boldsymbol{A}}+r_{\boldsymbol{B}}<r_{\boldsymbol{A}+\boldsymbol{B}}$. 类似于惯性定理的证明中的处理手法, 我们解关于 $\boldsymbol{X}$ 的如下方程组:

$$\begin{cases}
y_i=0, & \text{若 } 1\leq i\leq r_{\boldsymbol{A}},\\
z_j=0, & \text{若 } 1\leq j\leq r_{\boldsymbol{B}},\\
w_k=0, & \text{若 } r_{\boldsymbol{A}+\boldsymbol{B}}+1\leq k\leq n.
\end{cases}$$

由于此时 $r_{\boldsymbol{A}}+r_{\boldsymbol{B}}+(n-r_{\boldsymbol{A}+\boldsymbol{B}})<n$, 该齐次方程组存在非零解 $\boldsymbol{X}_0$. 对于它, 我们同时有

$$\boldsymbol{X}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}_0\leq 0,\quad \boldsymbol{X}_0^{\mathrm{T}}\boldsymbol{B}\boldsymbol{X}_0\leq 0\quad \text{以及}\quad \boldsymbol{X}_0^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{X}_0>0.$$

这显然有矛盾. □

8.4 正定二次型

回忆. 设 $\boldsymbol{A}$ 是实对称矩阵, 则 $\boldsymbol{A}$ 是正定矩阵当且仅当对任意的非零列向量 $\boldsymbol{X}\in\mathbb{R}^n$, 有 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}>0$. 此时, 我们将之简记为 $\boldsymbol{A}>0$.

定理 8.4.1. 设 $\boldsymbol{A}$ 为 n 阶实对称阵, 则以下几条等价.

(1) $\boldsymbol{A}>0$.

(2) $\boldsymbol{A}$ 的正惯性指数 $r=n$, 即 $\boldsymbol{A}$ 的相合标准形中主对角元全大于零.

(3) $\boldsymbol{A}$ 的负惯性指数 $s=0$ 且 $\operatorname{rank}(\boldsymbol{A})=n$.

(4) $\boldsymbol{A}$ 相合于单位方阵, 即存在 n 阶可逆矩阵 $\boldsymbol{B}$ 使得 $\boldsymbol{A}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$.

(5) 存在可逆上三角矩阵 $\boldsymbol{R}$ 使得 $\boldsymbol{A}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$. *

*这称为正定矩阵的楚列斯基分解 (Cholesky decomposition).

(6) $\boldsymbol{A}$ 的所有特征值 (必为实数) 皆大于零.

(7) 存在可逆上三角矩阵 $\boldsymbol{R}$ 使得 $\boldsymbol{A}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{R}$, 其中 $\boldsymbol{R}$ 的对角线上的元素都是 1, 而 $\boldsymbol{D}$ 为对角矩阵, 其对角元素均为正的.

证明. 存在可逆阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\begin{pmatrix}\boldsymbol{I}_r & & \\ & -\boldsymbol{I}_s & \\ & & \boldsymbol{O}\end{pmatrix}$ 为 $\boldsymbol{A}$ 的规范形. 令 $Q(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$.

"(1) $\Leftarrow$ (2)": 若 $r=n$, 则 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{I}_n$. 考虑坐标变换 $\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}$, 则 $Q(\boldsymbol{X})|_{\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}}=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}=|\boldsymbol{Y}|^2$. 此时, 若 $\boldsymbol{X}\neq\boldsymbol{0}$, 则 $\boldsymbol{Y}\neq\boldsymbol{0}$, 从而 $|\boldsymbol{Y}|^2>0$, 故 $Q(\boldsymbol{X})>0$.

"(1) $\Rightarrow$ (2)": 反设 $r\neq n$. 仍然令 $\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}$, 并选取 $\boldsymbol{Y}_0=(0,\ldots,0,1)^{\mathrm{T}}\neq\boldsymbol{0}$. 则 $\boldsymbol{X}_0=\boldsymbol{P}\boldsymbol{Y}_0\neq\boldsymbol{0}$, 且

$$Q(\boldsymbol{X}_0)=Q(\boldsymbol{X})|_{\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}_0}=\boldsymbol{Y}_0^{\mathrm{T}}\begin{pmatrix}\boldsymbol{I}_r & & \\ & -\boldsymbol{I}_s & \\ & & \boldsymbol{O}\end{pmatrix}\boldsymbol{Y}_0.$$

由于对角阵 $\begin{pmatrix}\boldsymbol{I}_r & & \\ & -\boldsymbol{I}_s & \\ & & \boldsymbol{O}\end{pmatrix}$ 的第 (n,n) 元素为 0 或 -1, 从而相应地 $Q(\boldsymbol{X}_0)=0$ 或 $=-1$, 故 $\boldsymbol{A}$ 不是正定矩阵.

"(2) $\Leftrightarrow$ (3)": 这是因为 $\operatorname{rank}(\boldsymbol{A})=r+s$.

"(2) $\Leftrightarrow$ (4)": 利用 $\boldsymbol{A}$ 的规范形.

"(4) $\Leftrightarrow$ (5)": 我们只需要证明 "$\Rightarrow$". 假定 $\boldsymbol{A}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$, 其中 $\boldsymbol{B}$ 为可逆矩阵. 由 QR 分解可知, 存在正交矩阵 $\boldsymbol{Q}$ 与上三角矩阵 $\boldsymbol{R}$ 使得 $\boldsymbol{B}=\boldsymbol{Q}\boldsymbol{R}$. 此时,

$$\boldsymbol{A}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=(\boldsymbol{Q}\boldsymbol{R})^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}.$$

"(2) $\Leftrightarrow$ (6)": 这是因为 r 是 $\boldsymbol{A}$ 的特征值中正项的个数.

"(5) $\Rightarrow$ (7)": 由条件, 存在可逆上三角矩阵 $\boldsymbol{R}_0$ 使得 $\boldsymbol{A}=\boldsymbol{R}_0^{\mathrm{T}}\boldsymbol{R}_0$. 设 $\boldsymbol{R}_0=(r_{ij})_{n\times n}$. 若取 $\boldsymbol{D}=\operatorname{diag}(r_{11}^2,r_{22}^2,\ldots,r_{nn}^2)$, 而 $\boldsymbol{R}=(q_{ij}\coloneqq r_{ij}/r_{ii})_{n\times n}$, 不难直接验证, 这样的构造符合要求.

"(7) $\Rightarrow$ (2)": 这是因为 $\boldsymbol{A}$ 的正惯性指数与 $\boldsymbol{D}$ 的正惯性指数一致, 皆为 n. □

习题 8.4.2. 已知 2021 阶实对称阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=2021\boldsymbol{A}$. 证明: $\boldsymbol{I}+\boldsymbol{A}+\cdots+\boldsymbol{A}^{2021}$ 必然是正定的.

习题 8.4.3. 设 n 阶实矩阵 $\boldsymbol{A}$ 是正定矩阵, $\boldsymbol{B}$ 为 $m\times n$ 实矩阵. 证明: $\operatorname{rank}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{B})=\operatorname{rank}(\boldsymbol{B})$.

定理 8.4.4. 设 $\boldsymbol{A}$ 为 n 阶实对称阵.

(1) 设 $\boldsymbol{P}$ 为 n 阶可逆方阵, $\boldsymbol{B}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$, 则 $\boldsymbol{A}>0$ 当且仅当 $\boldsymbol{B}>0$.

(2) 若 $\boldsymbol{A}>0$, 则 $\det(\boldsymbol{A})>0$ 且 $\operatorname{tr}(\boldsymbol{A})>0$.

(3) 若 $\boldsymbol{A}>0$, 则 $\boldsymbol{A}$ 可逆, 且 $\boldsymbol{A}^{-1}$ 也是实对称的正定方阵.

证明. (1) $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相合等价, 从而有相同的正、负惯性指数.

(2) 一方面, $\det(\boldsymbol{A})$ 是 $\boldsymbol{A}$ 的特征值之积, $\operatorname{tr}(\boldsymbol{A})$ 是 $\boldsymbol{A}$ 的特征值之和. 另一方面, 由定理 8.4.1 已知, $\boldsymbol{A}$ 的所有特征值都是正数.

(3) 由上, $\det(\boldsymbol{A})\neq 0$, 从而 $\boldsymbol{A}$ 可逆, 并且 $(\boldsymbol{A}^{-1})^{\mathrm{T}}=(\boldsymbol{A}^{\mathrm{T}})^{-1}=\boldsymbol{A}^{-1}$ 是实对称阵. $\boldsymbol{A}^{-1}$ 的所有特征值是 $\boldsymbol{A}$ 的相应特征值的逆, 皆为大于零的实数. 故 $\boldsymbol{A}^{-1}$ 也是正定的. □

例 8.4.5. 设 $\boldsymbol{A}$ 是 n 阶可逆实对称矩阵. 证明: $\boldsymbol{A}$ 是正定的充要条件是对于任意的同阶正定矩阵 $\boldsymbol{B}$, 有 $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B})>0$.

证明. 必要性. 设 $\boldsymbol{A}$ 是一个正定矩阵. 任取一个同阶正定矩阵 $\boldsymbol{B}$, 则存在同阶可逆矩阵 $\boldsymbol{C}$ 使得 $\boldsymbol{B}=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$. 此时,

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B})=\operatorname{tr}(\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})=\operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}}).$$

由于 $\boldsymbol{A}$ 是正定的, $\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}}$ 也是正定的, 从而 $\operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}})>0$.

充分性. 由于 $\boldsymbol{A}$ 是实对称矩阵, 存在同阶正交矩阵 $\boldsymbol{T}$ 使得

$$\boldsymbol{A}=\boldsymbol{T}^{-1}\operatorname{diag}(\lambda_1,\lambda_2,\dots,\lambda_n)\boldsymbol{T},$$

其中 $\lambda_1,\lambda_2,\dots,\lambda_n\in\mathbb{R}$ 为 $\boldsymbol{A}$ 的全部特征值. 对于任意固定的 $i=1,2,\dots,n$, 考虑矩阵

$$\boldsymbol{B}_t=\boldsymbol{T}^{-1}\operatorname{diag}(t,\dots,t,1,t,\dots,t)\boldsymbol{T},$$

其中的 1 在第 i 个位置上. 则对于任意的 $t>0$, $\boldsymbol{B}_t$ 是正定矩阵, 从而由条件知

$$0<\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}_t)=\operatorname{tr}(\boldsymbol{T}^{-1}\operatorname{diag}(t\lambda_1,\dots,t\lambda_{i-1},\lambda_i,t\lambda_{i+1},\dots,t\lambda_n)\boldsymbol{T})=\lambda_i+t\sum_{j\neq i}\lambda_j.$$

令 $t\to 0^+$, 我们得到 $\lambda_i\geq 0$. 又由于 $\boldsymbol{A}$ 可逆, 故 $\lambda_i\neq 0$, 从而有 $\lambda_i>0$. 再由 i 的任意性可知 $\boldsymbol{A}$ 的所有特征值都严格大于零, 从而实对称阵 $\boldsymbol{A}$ 是正定的. □

定理 8.4.6. 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是两个 n 阶实对称矩阵, 其中 $\boldsymbol{A}$ 正定. 则存在可逆矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{I}_n$, 而 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}=\boldsymbol{D}$ 为对角阵. 特别地, $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 可以同时相合对角化.

证明. 由于 $\boldsymbol{A}$ 正定, 存在可逆矩阵 $\boldsymbol{P}_1$ 使得 $\boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}_1=\boldsymbol{I}_n$. 此时, $\boldsymbol{B}_1\coloneqq\boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}_1$ 仍然为实对称阵, 故存在正交阵 $\boldsymbol{P}_2$, 使得 $\boldsymbol{B}_1$ 相似 (相合) 于对角阵: $\boldsymbol{D}\coloneqq\boldsymbol{P}_2^{\mathrm{T}}\boldsymbol{B}_1\boldsymbol{P}_2$ 为对角阵. 若令 $\boldsymbol{P}=\boldsymbol{P}_1\boldsymbol{P}_2$, 则 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{I}$ 且 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}=\boldsymbol{D}$. □

例 8.4.7. 设 $\boldsymbol{A},\boldsymbol{B}$ 均为 n 阶正定阵, 若 $\boldsymbol{A}-\boldsymbol{B}$ 正定, 证明: $\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}$ 也是正定阵.

证明. 因为 $\boldsymbol{A}$ 为正定阵, 由定理 8.4.6 可知, 存在可逆方阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{I}_n$, 而 $\boldsymbol{D}\coloneqq\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}=\operatorname{diag}(\mu_1,\dots,\mu_n)$ 为对角阵. 又由于 $\boldsymbol{B}$ 是正定的, 故其中每个 $\mu_i>0$. 由于 $\boldsymbol{A}-\boldsymbol{B}$ 为正定阵, 故 $\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{P}=\boldsymbol{I}-\boldsymbol{D}=\operatorname{diag}(1-\mu_1,1-\mu_2,\dots,1-\mu_n)$ 也是正定阵, 这说明每个 $\mu_i<1$. 此时, $\boldsymbol{P}^{-1}(\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1})(\boldsymbol{P}^{-1})^{\mathrm{T}}=\boldsymbol{D}^{-1}-\boldsymbol{I}=\operatorname{diag}(\mu_1^{-1}-1,\mu_2^{-1}-1,\dots,\mu_n^{-1}-1)$. 由于每个 $\mu_i^{-1}-1>0$, 该实对称阵正定, 从而在相合变换后, $\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}$ 也是正定的. □

例 8.4.8. 设 $\boldsymbol{A},\boldsymbol{B}$ 均为 n 阶实对称阵, 且 $\boldsymbol{A}$ 正定. 证明: 当实数 t 充分大时, $t\boldsymbol{A}+\boldsymbol{B}$ 也是正定阵.

证明. 因为 $\boldsymbol{A}$ 为正定阵, 存在可逆方阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{I}_n$, 而 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}=\boldsymbol{D}=\operatorname{diag}(\mu_1,\dots,\mu_n)$. 此时, $\boldsymbol{P}^{\mathrm{T}}(t\boldsymbol{A}+\boldsymbol{B})\boldsymbol{P}=t\boldsymbol{I}_n+\boldsymbol{D}=\operatorname{diag}(t+\mu_1,t+\mu_2,\dots,t+\mu_n)$. 显然, 当 $t>\max\{-\mu_1,-\mu_2,\dots,-\mu_n\}$ 时, $t\boldsymbol{A}+\boldsymbol{B}$ 是正定的. □

定义 8.4.9. 对于二次型 $Q(x_1,\dots,x_n)=\sum_{i,j=1}^{n}a_{ij}x_ix_j=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$, 若 $\boldsymbol{A}$ 为正定阵, 则称 Q 为**正定的** (*positive definite*).

例 8.4.10. $Q_1(x_1,x_2,\dots,x_n)=x_1^2+x_2^2+\cdots+x_n^2$ 是正定的二次型, 而当 $r<n$ 时, $Q_2(x_1,x_2,\dots,x_n)=x_1^2+x_2^2+\cdots+x_r^2$ 不是正定的二次型, 这是因为 $Q(0,0,\dots,0,1)=0$.

推论 8.4.11. 二次型经过可逆的线性变换, 其正定性不变.

证明. 这是二次型经过可逆的线性变换得到的矩阵与原矩阵是相合的. □

研究二次型的正定性就是研究其对应矩阵的正定性, 与之相关的最实用的条件是以下的定理.

定理 8.4.12. n 阶实对称阵 $\boldsymbol{A}$ 是正定阵的充要条件是 $\boldsymbol{A}$ 的各阶**顺序主子式**均大于零, 即对任意的 $r=1,2,\dots,n$ 有

$$\begin{vmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & & \vdots \\ a_{r1} & \cdots & a_{rr} \end{vmatrix} > 0.$$

(注意: 这儿只需验证 n 个子阵的行列式, 不是所有的主子式.)

证明. 必要性. 对于任意的 $r=1,2,\ldots,n$, 我们考察 r 个变元的二次型

$$Q_r(x_1,\ldots,x_r)\coloneqq Q(x_1,\ldots,x_r,0,\ldots,0)=\sum_{i=1}^{r}\sum_{j=1}^{r}a_{ij}x_ix_j.$$

显然, 若 $x_1,\ldots,x_r$ 不全为零, 则 $(x_1,\ldots,x_r,0,\ldots,0)\neq\mathbf{0}$, 故

$$Q_r(x_1,\ldots,x_r)=Q(x_1,\ldots,x_r,0,\ldots,0)>0.$$

这等价于说 Q_r 是变元 $x_1,\ldots,x_r$ 上的正定实二次型. 它对应的方阵

$$\boldsymbol{A}_r=\begin{pmatrix}a_{11}&\cdots&a_{1r}\\ \vdots&&\vdots\\ a_{r1}&\cdots&a_{rr}\end{pmatrix}$$

是 r 阶实正定方阵, 故由定理 8.4.4 可知 $|\boldsymbol{A}_r|>0$.

充分性. 我们关于阶数 n 作归纳. $n=1$ 时显然成立. 接下来假设 $n-1$ 时结论成立, 考虑 n 时的情形.

$$\boldsymbol{A}=\left(\begin{array}{ccc|c}a_{1,1}&\cdots&a_{1,n-1}&a_{1,n}\\ \vdots&&\vdots&\vdots\\ a_{n-1,1}&\cdots&a_{n-1,n-1}&a_{n-1,n}\\ \hline a_{n,1}&\cdots&a_{n,n-1}&a_{n,n}\end{array}\right)=\begin{pmatrix}\boldsymbol{A}_{n-1}&\boldsymbol{c}\\ \boldsymbol{c}^{\mathrm{T}}&a_{n,n}\end{pmatrix}.$$

其中 $\boldsymbol{A}_{n-1}$ 是 $n-1$ 阶实对称方阵. 由条件, $\boldsymbol{A}_{n-1}$ 的各阶顺序主子式也都大于 0, 故由归纳假设可知, $\boldsymbol{A}_{n-1}>0$. 此时, 由定理 8.4.1 可知, 存在可逆矩阵 $\boldsymbol{P}_{n-1}$ 使得 $\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{A}_{n-1}\boldsymbol{P}_{n-1}=\boldsymbol{I}_{n-1}$. 此时, 对于 $\boldsymbol{A}$ 我们有相合变换

$$\begin{pmatrix}\boldsymbol{P}_{n-1}&\\ &1\end{pmatrix}^{\mathrm{T}}\underbrace{\begin{pmatrix}\boldsymbol{A}_{n-1}&\boldsymbol{c}\\ \boldsymbol{c}^{\mathrm{T}}&a_{n,n}\end{pmatrix}}_{\boldsymbol{A}}\begin{pmatrix}\boldsymbol{P}_{n-1}&\\ &1\end{pmatrix}=\begin{pmatrix}\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{A}_{n-1}\boldsymbol{P}_{n-1}&\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}\\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{P}_{n-1}&a_{n,n}\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_{n-1}&\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}\\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{P}_{n-1}&a_{n,n}\end{pmatrix}.$$

接下来, 我们继续操作相合变换, 利用左上角的 $\boldsymbol{I}_{n-1}$ 将右上角的 $\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}$ 消去, 这等价于说整个 n 阶方阵首先右乘 $\begin{pmatrix}\boldsymbol{I}_{n-1}&-\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}\\ \mathbf{0}&1\end{pmatrix}$. 从而整个相合变换为

$$\begin{pmatrix}\boldsymbol{I}_{n-1}&-\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}\\ \mathbf{0}&1\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\boldsymbol{I}_{n-1}&\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}\\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{P}_{n-1}&a_{n,n}\end{pmatrix}\begin{pmatrix}\boldsymbol{I}_{n-1}&-\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}\\ \mathbf{0}&1\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_{n-1}&\\ &\underbrace{a_{n,n}-\boldsymbol{c}^{\mathrm{T}}\boldsymbol{P}_{n-1}\boldsymbol{P}_{n-1}^{\mathrm{T}}\boldsymbol{c}}_{\text{记作 } a}\end{pmatrix}.$$

接下来, 利用

- 条件: $|\boldsymbol{A}| > 0$;

- 事实: 相合变换不改变实矩阵的行列式的正负号.

因此, 最后得到的方阵的行列式 $\begin{vmatrix} \boldsymbol{I}_{n-1} & \\ & a \end{vmatrix} = a > 0$. 由于 $\begin{pmatrix} \boldsymbol{I}_{n-1} & \\ & a \end{pmatrix}$ 是 $\boldsymbol{A}$ 的一个标准形, 这说明 $\boldsymbol{A}$ 的正惯性指数为 n, 从而 $\boldsymbol{A} > 0$. □

对上面的定理的必要性部分的证明稍作修改, 我们不难看出如下的结果.

推论 8.4.13. *设 $\boldsymbol{A}$ 是一个 n 阶正定的矩阵, $1 \le i_1 < i_2 < \cdots < i_s \le n$, 那么由 $\boldsymbol{A}$ 的第 $i_1, i_2, \cdots, i_r$ 行第 $i_1, i_2, \ldots, i_r$ 列元素构成的 r 阶主子阵 $\boldsymbol{A}\left(\begin{smallmatrix} i_1 & i_2 & \cdots & i_r \\ i_1 & i_2 & \cdots & i_r \end{smallmatrix}\right)$ 是一个 r 阶的正定矩阵. 特别地, 它的行列式是正的.*

例 8.4.14. *若 $\boldsymbol{A} = (a_{ij})_{n\times n}$ 为正定矩阵, 证明: $\det(\boldsymbol{A}) \le a_{11}a_{22}\cdots a_{nn}$.* 教材习题 #17

证明. 对阶数 n 作归纳, 并假定命题在 $n-1$ 时成立. 利用定理 8.4.12 中的记号, 我们可以将 $\boldsymbol{A}$ 写成分块矩阵并作初等列变换, 得到

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{n-1} & \boldsymbol{C} \\ \boldsymbol{C}^{\mathrm{T}} & a_{nn} \end{pmatrix} \xrightarrow{-c_1\boldsymbol{A}_{n-1}^{-1}\boldsymbol{C}\to c_2} \begin{pmatrix} \boldsymbol{A}_{n-1} & \boldsymbol{0} \\ \boldsymbol{C}^{\mathrm{T}} & a_{nn} - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}_{n-1}^{-1}\boldsymbol{C} \end{pmatrix}.$$

这说明

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}_{n-1})(a_{nn} - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}_{n-1}^{-1}\boldsymbol{C}).$$

这其实是例 4.4.6 中的 Schur 行列式公式的特例

由归纳假设可知 $0 < \det(\boldsymbol{A}_{n-1}) \le a_{11}a_{22}\cdots a_{n-1,n-1}$. 另外, $\boldsymbol{A}_{n-1}^{-1}$ 是正定矩阵, 从而 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}_{n-1}^{-1}\boldsymbol{C} \ge 0$, 并因此有 $a_{nn} - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}_{n-1}^{-1}\boldsymbol{C} \le a_{nn}$. 从而命题在 n 时也成立. □

例 8.4.15. *判断二次型 $Q(x_1, x_2, x_3) = 2x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3$ 是否正定.*

解. 二次型的矩阵为 $\boldsymbol{A} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 1 \end{pmatrix}$. 由于各阶顺序主子式依次为 $2 > 0$, $\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3 > 0$, $|\boldsymbol{A}| = -3 < 0$, $\boldsymbol{A}$ 不是正定的, 从而二次型 Q 不是正定的. □

定义 8.4.16 (与正定阵相关的定义). *设 $\boldsymbol{A}$ 为实对称阵, 考虑二次型 $Q(\boldsymbol{X}) = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$.*

(1) *称 Q 以及 $\boldsymbol{A}$ 为***半正定的** *(positive semidefinite)* 当且仅当 Q 的负惯性指数 $s = 0$, 即, 当且仅当对任意 $\boldsymbol{X} \in \mathbb{R}^n$, 有 $Q(\boldsymbol{X}) \ge 0$. 此时, 记 $\boldsymbol{A} \ge 0$.*

*直接翻译的话, 应该称之为正半定的; 不过我们这儿不采取这样的说法.

(2) 称 Q 以及 $\boldsymbol{A}$ 为**负定的** (*negative definite*) 当且仅当 Q 的负惯性指数 $s=n$, 即, 当且仅当对任意 $\boldsymbol{X}\neq\mathbf{0}\in\mathbb{R}^n$, 有 $Q(\boldsymbol{X})<0$. 此时, 记 $\boldsymbol{A}<0$.

(3) 称 Q 以及 $\boldsymbol{A}$ 为**半负定的** (*negative semidefinite*) 当且仅当 Q 的正惯性指数 $r=0$, 即, 当且仅当对任意 $\boldsymbol{X}\in\mathbb{R}^n$, 有 $Q(\boldsymbol{X})\leq 0$. 此时, 记 $\boldsymbol{A}\leq 0$.

(4) 称 Q 以及 $\boldsymbol{A}$ 为**不定的** (*indefinite*) 当且仅当 Q 处于其它情形, 即, 当且仅当 $r>0$ 且 $s>0$, 而这也当且仅当存在 $\boldsymbol{X}_1,\boldsymbol{X}_2\neq\mathbf{0}\in\mathbb{R}^n$, 有 $Q(\boldsymbol{X}_1)<0<Q(\boldsymbol{X}_2)$.

命题 8.4.17. 对实二次型 $Q(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$, 对应于定理 *8.4.4*, 我们有如下的等价刻划.

(1) $\boldsymbol{A}\geq 0$.

这些性质
也请牢记

(2) $\boldsymbol{A}$ 相合于它的相抵标准形 $\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$.

(3) $\boldsymbol{A}$ 的特征值全为非负实数.

(4) 存在 $m\times n$ 维矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{A}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$. (可以由此推出, $m\geq\operatorname{rank}(\boldsymbol{A})$)

(5) 存在 $m\times n$ 维矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{A}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$, 其中 $m=\operatorname{rank}(\boldsymbol{A})$.

(6) 存在 $n\times n$ 维矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{A}=\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$.

(7) 所有的主子式皆非负.

证明. 这儿, 我们只证明 “(7)⇒(3)”, 其余的部分留作练习. 考虑特征多项式

$$p_{\boldsymbol{A}}(t)=|t\boldsymbol{I}-\boldsymbol{A}|=t^n+a_{n-1}t^{n-1}+\cdots+a_1t+a_0.$$

对于 $i=1,2,\ldots,n$, 我们熟知 $(-1)^i a_{n-i}$ 是 $\boldsymbol{A}$ 的所有 i 阶主子式之和, 而由条件可知其非负. 这说明, 此时, 当 $t<0$ 时, $(-1)^n p_{\boldsymbol{A}}(t)$ 为正数, 从而不可能为零. 换言之, 实对称阵 $\boldsymbol{A}$ 的特征值皆非负. □

例 8.4.18. 在命题 *8.4.17* 中, 若只有 $\boldsymbol{A}$ 的顺序主子式非负, 不能推出 $\boldsymbol{A}\geq 0$. 例如, 考虑 $\boldsymbol{A}=\operatorname{diag}(1,0,-1)$.

例 8.4.19. 若 $\boldsymbol{A}$ 为 $m\times n$ 维实矩阵, 则 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 为半正定的实对称阵. 这是因为对任意的 m 维列向量 $\boldsymbol{X}$, 二次型 $Q(\boldsymbol{X})=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}=\left|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\right|^2\geq 0$, 从而 Q 是半正定的二次型.

由此看出, $\det(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})\geq 0$. 当然, 在 $m\geq n$ 时, 利用秩的性质, 我们很容易证明这一点. 在一般情形, 也可以采用 *Binet–Cauchy* 公式. 另外, 若 $\boldsymbol{A}$ 为 $m\times n$ 维复矩阵, 则 $\det(\boldsymbol{A}\overline{\boldsymbol{A}}^{\mathrm{T}})$ 为非负实数. 这儿的证明类似, 但是要用到酉空间的性质.

例 8.4.20. 设 $\boldsymbol{B}$ 是 n 阶实矩阵, $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ 的全部特征值为 $\lambda_1 \leq \cdots \leq \lambda_n$. 证明: 如果 μ 是 $\boldsymbol{B}$ 的实特征值, 那么 $\sqrt{\lambda_1} \leq |\mu| \leq \sqrt{\lambda_n}$.

证明. 容易验证 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ 为半正定实对称阵, 从而 $\lambda_1 \geq 0$. 设 $\boldsymbol{x}$ 是 $\boldsymbol{B}$ 的属于 μ 的单位特征向量. 由例 8.2.6 (Rayleigh 原理) 有

$$\lambda_1 \leq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x} \leq \lambda_n.$$

但是 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x} = (\boldsymbol{B}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{B}\boldsymbol{x}) = (\mu\boldsymbol{x})^{\mathrm{T}}(\mu\boldsymbol{x}) = \mu^2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \mu^2\,|\boldsymbol{x}|^2 = \mu^2$. 化简即可. □

例 8.4.21. 设 $\boldsymbol{A}, \boldsymbol{B}$ 为 n 阶实对称正定阵, 则 $\boldsymbol{AB}$ 为对称正定阵的充要条件是 $\boldsymbol{AB} = \boldsymbol{BA}$.

教材习题 #24(2)

证明. 由于 $(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{BA}$, 这意味着 $\boldsymbol{AB}$ 为对称阵, 当且仅当 $\boldsymbol{AB} = (\boldsymbol{AB})^{\mathrm{T}}$, 当且仅当 $\boldsymbol{AB} = \boldsymbol{BA}$.

我们先证明必要性. 显然, 若 $\boldsymbol{AB}$ 为正定阵, 则其首先为对称阵, 从而 $\boldsymbol{AB} = \boldsymbol{BA}$.

下面, 我们证明充分性. 于是, $\boldsymbol{AB}$ 为实对称阵.

(思路一) 由思考题 7.3.28 的结果 (可乘法交换的两个实对称阵可以同时正交相似对角化) 可知, 存在正交矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP} = \mathrm{diag}(\lambda_1', \ldots, \lambda_n')$, 而 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{BP} = \mathrm{diag}(\lambda_1'', \ldots, \lambda_n'')$. 由于 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 正定, 这些 λ_i' 与 λ_i'' 皆大于 0. 此时, $\boldsymbol{PABP}^{\mathrm{T}} = (\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP})(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{BP}) = \mathrm{diag}(\lambda_1'\lambda_1'', \ldots, \lambda_n'\lambda_n'')$. 故 $\boldsymbol{AB}$ 的所有特征值为 $\lambda_1'\lambda_1'', \ldots, \lambda_n'\lambda_n''$, 皆大于 0, 从而实对称阵 $\boldsymbol{AB}$ 是正定的.

(思路二) 由于 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 正定, 分别存在可逆方阵 $\boldsymbol{P}$ 与 $\boldsymbol{Q}$ 使得 $\boldsymbol{A} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$, $\boldsymbol{B} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$. 则 $\boldsymbol{AB} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{PQ}^{\mathrm{T}}\boldsymbol{Q}$, 且 $\boldsymbol{Q}(\boldsymbol{AB})\boldsymbol{Q}^{-1} = \boldsymbol{QP}^{\mathrm{T}}\boldsymbol{PQ}^{\mathrm{T}} = (\boldsymbol{PQ}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{PQ}^{\mathrm{T}}) > 0$. 这说明, $\boldsymbol{AB}$ 与正定阵相似, 从而所有的特征值大于 0, 从而实对称阵 $\boldsymbol{AB}$ 是正定的.

(思路三) 任取实对称阵 $\boldsymbol{AB}$ 的特征值 λ, 假设 $\boldsymbol{x}$ 是对应的一个特征向量: $\boldsymbol{ABx} = \lambda\boldsymbol{x}$. 从而 $\boldsymbol{Bx} = \lambda\boldsymbol{A}^{-1}\boldsymbol{x}$, 故 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Bx} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x}$. 由于 $\boldsymbol{x} \neq \boldsymbol{0}$, 而 $\boldsymbol{B}, \boldsymbol{A}^{-1}$ 都是正定的, 故

$$\lambda = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Bx}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{x}} > 0.$$

这说明实对称阵 $\boldsymbol{AB}$ 是正定的.

(思路四) 存在正交矩阵 $\boldsymbol{Q}$ 使得 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{AQ} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. 此时, $\boldsymbol{C} := \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{BQ}$ 是正定方阵, 并且实对称阵

$$\boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{AB})\boldsymbol{Q} = (\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{AQ})(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{BQ}) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{C}.$$

可以直接验证, $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{C}$ 的 k 阶顺序主子式恰为 $\lambda_1 \cdots \lambda_k$ 倍的 $\boldsymbol{C}$ 的 k 阶顺序主子式, 从而大于零. 这说明实对称阵 $\boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{AB})\boldsymbol{Q}$ 是正定的, 从而 $\boldsymbol{AB}$ 是正定的. $\square$

例 8.4.22. 证明: n 阶半正定矩阵 $\boldsymbol{A}$ 必满足 $\det(\boldsymbol{A}) \le \left(\frac{1}{n}\operatorname{tr}(\boldsymbol{A})\right)^n$.

教材习题 #27

证明. 由于 $\boldsymbol{A}$ 半正定, 它的所有特征值 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 皆为非负实数. 由于 $\det(\boldsymbol{A}) = \lambda_1\lambda_2\cdots\lambda_n$, 而 $\operatorname{tr}(\boldsymbol{A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$, 欲证的不等式为熟知的算数平均–几何平均不等式. □

例 8.4.23. 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是 n 阶半正定实对称方阵. 证明:

$$\det(\boldsymbol{A})^{1/n}\det(\boldsymbol{B})^{1/n} \le \frac{1}{n}\operatorname{tr}(\boldsymbol{AB}).$$

证明. 分别存在 n 阶矩阵 $\boldsymbol{P}$ 与 $\boldsymbol{Q}$ 使得 $\boldsymbol{A} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}$ 和 $\boldsymbol{B} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$. 则

$$\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}) = \operatorname{tr}(\boldsymbol{P}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{P}^{\mathrm{T}}).$$

另一方面,

$$0 \le \det(\boldsymbol{A})\det(\boldsymbol{B}) = \det(\boldsymbol{P}^{\mathrm{T}})\det(\boldsymbol{P})\det(\boldsymbol{Q}^{\mathrm{T}})\det(\boldsymbol{Q}) = \det(\boldsymbol{P}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{P}^{\mathrm{T}}).$$

由于 $\boldsymbol{C} := \boldsymbol{P}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{P}^{\mathrm{T}} = (\boldsymbol{Q}\boldsymbol{P}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{Q}\boldsymbol{P}^{\mathrm{T}})$ 为半正定对称阵, 故由例 8.4.22 可知

$$\det(\boldsymbol{C}) \le \left(\frac{\operatorname{tr}(\boldsymbol{C})}{n}\right)^n.$$

化简即可. □

习题 8.4.24. 设 $\boldsymbol{A}$ 为实对称矩阵. 证明: 二次型 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$ 不是不定的, 当且仅当由 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} = 0$ 可以推出 $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{0}$.

思考题 8.4.25. 设 $\boldsymbol{A}, \boldsymbol{B}$ 是同阶实对称矩阵, 证明:

$$\operatorname{tr}(\boldsymbol{ABAB}) \le \operatorname{tr}(\boldsymbol{A}^2\boldsymbol{B}^2).$$

证明. **(思路一)** 考虑 $\boldsymbol{C} = \boldsymbol{AB} - \boldsymbol{BA}$, 则 $\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}$ 是半正定的, 从而 $\operatorname{tr}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}) \ge 0$, 而这等价于

$$\operatorname{tr}(\boldsymbol{ABBA}) + \operatorname{tr}(\boldsymbol{BAAB}) \ge \operatorname{tr}(\boldsymbol{ABAB}) + \operatorname{tr}(\boldsymbol{BABA}).$$

不难看出, $\operatorname{tr}(\boldsymbol{ABBA}) = \operatorname{tr}(\boldsymbol{BAAB}) = \operatorname{tr}(\boldsymbol{A}^2\boldsymbol{B}^2)$, 而 $\operatorname{tr}(\boldsymbol{ABAB}) = \operatorname{tr}(\boldsymbol{BABA})$.

(思路二) 对于一般的同阶实方阵 $\boldsymbol{C}, \boldsymbol{D}$, 由 Cauchy-Schwarz 不等式可知,

$$\operatorname{tr}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{D}) \le \sqrt{\operatorname{tr}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})\operatorname{tr}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})}.$$

运用该不等式, 我们有

$$\begin{aligned}\operatorname{tr}(\boldsymbol{ABAB}) &= \operatorname{tr}((\boldsymbol{BA})^{\mathrm{T}}(\boldsymbol{BA})) \le \sqrt{\operatorname{tr}((\boldsymbol{BA})^{\mathrm{T}}(\boldsymbol{BA}))\operatorname{tr}((\boldsymbol{BA})^{\mathrm{T}}(\boldsymbol{BA}))}\\ &= \sqrt{\operatorname{tr}(\boldsymbol{ABBA})\operatorname{tr}(\boldsymbol{ABBA})} = \sqrt{(\operatorname{tr}(\boldsymbol{A}^2\boldsymbol{B}^2))^2} = \operatorname{tr}(\boldsymbol{A}^2\boldsymbol{B}^2).\end{aligned}$$

其中用到了 $\operatorname{tr}((\boldsymbol{BA})^{\mathrm{T}}(\boldsymbol{BA})) = \operatorname{tr}(\boldsymbol{ABBA}) = \operatorname{tr}(\boldsymbol{A}^2\boldsymbol{B}^2)$ 是半正定矩阵 $(\boldsymbol{BA})^{\mathrm{T}}(\boldsymbol{BA})$ 的迹, 从而非负. □

思考题 8.4.26. (1) 证明: n 阶实对称阵 $\boldsymbol{A}$ 是正定的充分必要条件是存在可逆实对称阵 $\boldsymbol{C}$ 使得 $\boldsymbol{A} = \boldsymbol{C}^2$.

(2) 证明: n 阶实对称阵 $\boldsymbol{A}$ 是半正定的充分必要条件是存在实对称阵 $\boldsymbol{C}$ 使得 $\boldsymbol{A} = \boldsymbol{C}^2$.

证明. 我们只证第一个命题. 第二个命题的证明是类似的.

必要性. 设 $\boldsymbol{A}$ 是 n 阶正定矩阵, 则 $\boldsymbol{A}$ 的特征值 $\lambda_1, \ldots, \lambda_n$ 全大于零, 并存在正交矩阵 $\boldsymbol{T}$ 使得

$$\begin{aligned}\boldsymbol{A} =& \boldsymbol{T}^{-1}\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{T}\\ =& \left(\boldsymbol{T}^{-1}\operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})\boldsymbol{T}\right)\left(\boldsymbol{T}^{-1}\operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})\boldsymbol{T}\right)\\ =& \boldsymbol{C}^2,\end{aligned}$$

其中 $\boldsymbol{C} \coloneqq \boldsymbol{T}^{-1}\operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})\boldsymbol{T}$ 是符合要求的矩阵.

充分性是显然的, 这是因为 $\boldsymbol{C}$ 是实对称的, 从而 $\boldsymbol{A} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}$. □

思考题 8.4.27. (1) 设 $\boldsymbol{A}$ 为正定矩阵, 证明: 存在唯一的正定矩阵 $\boldsymbol{C}$ 使得 $\boldsymbol{A} = \boldsymbol{C}^2$.

(2) 设 $\boldsymbol{A}$ 为半正定矩阵, 证明: 存在唯一的半正定矩阵 $\boldsymbol{C}$ 使得 $\boldsymbol{A} = \boldsymbol{C}^2$.

证明. 我们仍然只证明第一个命题. 上面例题的证明给出了存在性.

唯一性. 设 $\boldsymbol{C}_1 > 0$ 也满足 $\boldsymbol{C}_1^2 = \boldsymbol{A}$. 设 $\boldsymbol{A}$ 的全部特征值为 $\lambda_1, \ldots, \lambda_n$. 则 $\boldsymbol{C}$ 与 $\boldsymbol{C}_1$ 的全部特征值为 $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$. 于是存在正交矩阵 $\boldsymbol{T}$ 与 $\boldsymbol{T}_1$ 使得

$$\boldsymbol{C} = \boldsymbol{T}^{-1}\operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})\boldsymbol{T}, \quad \boldsymbol{C}_1 = \boldsymbol{T}_1^{-1}\operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})\boldsymbol{T}_1.$$

由于 $\boldsymbol{C}^2 = \boldsymbol{A} = \boldsymbol{C}_1^2$, 我们有

$$\boldsymbol{T}_1^{-1}\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{T}_1 = \boldsymbol{T}^{-1}\operatorname{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{T},$$

即

$$\boldsymbol{T}\boldsymbol{T}_1^{-1}\operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\boldsymbol{T}\boldsymbol{T}_1^{-1}.$$

若记正交矩阵 $\boldsymbol{T}\boldsymbol{T}_1^{-1}=(t_{ij})_{n\times n}$, 则比较上式的 (i,j) 元, 我们有

$$t_{ij}\lambda_j=\lambda_i t_{ij}.$$

若 $t_{ij}\neq 0$, 则有 $\lambda_i=\lambda_j$, 从而

$$t_{ij}\sqrt{\lambda_j}=\sqrt{\lambda_i}t_{ij}. \tag{8.2}$$

当然, 若 $t_{ij}=0$, 等式 (8.2) 仍然成立. 这说明

$$\boldsymbol{T}\boldsymbol{T}_1^{-1}\operatorname{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_n})=\operatorname{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_n})\boldsymbol{T}\boldsymbol{T}_1^{-1}.$$

从而可以推出

$$\boldsymbol{C}=\boldsymbol{T}^{-1}\operatorname{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_n})\boldsymbol{T}=\boldsymbol{T}_1^{-1}\operatorname{diag}(\sqrt{\lambda_1},\dots,\sqrt{\lambda_n})\boldsymbol{T}_1=\boldsymbol{C}_1.$$ □

思考题 8.4.28. *设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是 n 阶半正定实对称方阵. 证明: $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 可以同时相合于对角阵, 即存在可逆方阵 $\boldsymbol{P}$, 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ 和 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P}$ 都是对角方阵.*

证明. 对任意的 $\boldsymbol{X}\in\mathbb{R}^n$, $\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{X}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{X}\geq 0+0=0$, 故 $\boldsymbol{A}+\boldsymbol{B}$ 仍然是半正定矩阵, 从而存在可逆矩阵 $\boldsymbol{Q}_1$ 使得 $\boldsymbol{Q}_1^{\mathrm{T}}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{Q}_1=\operatorname{diag}(\boldsymbol{I}_r,\boldsymbol{O})$. 写成分块矩阵的形式:

$$\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_1=(a_{ij})_{n\times n}=\begin{pmatrix}\boldsymbol{A}_{11}&\boldsymbol{A}_{12}\\ \boldsymbol{A}_{12}^{\mathrm{T}}&\boldsymbol{A}_{22}\end{pmatrix},\qquad \boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q}_1=(b_{ij})_{n\times n}=\begin{pmatrix}\boldsymbol{B}_{11}&\boldsymbol{B}_{12}\\ \boldsymbol{B}_{12}^{\mathrm{T}}&\boldsymbol{B}_{22}\end{pmatrix}.$$

其中 $\boldsymbol{A}_{11}$ 与 $\boldsymbol{B}_{11}$ 都是 r 阶实对称方阵.

由于 $\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_1,\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q}_1$ 半正定, 对于 $r<i\leq n$, $a_{ii},b_{ii}\geq 0$. 另一方面, $a_{ii}+b_{ii}=0$, 故迫使 $a_{ii}=b_{ii}=0$. 对于 $r<i<j\leq n$, 由半正定性, 主子式

$$\begin{vmatrix}a_{ii}=0&a_{ij}\\ a_{ij}&a_{jj}=0\end{vmatrix}\geq 0,$$

故 $a_{ij}=0$. 类似地, 有 $b_{ij}=0$. 从而 $\boldsymbol{A}_{22}=\boldsymbol{B}_{22}=\boldsymbol{O}$.

对于 $1\leq i\leq r<j\leq n$, 同样由半正定性, 主子式

$$\begin{vmatrix}a_{ii}&a_{ij}\\ a_{ij}&a_{jj}=0\end{vmatrix}\geq 0.$$

这迫使 $a_{ij}=0$. 从而 $\boldsymbol{A}_{12}=\boldsymbol{O}$. 类似地, $\boldsymbol{B}_{12}=\boldsymbol{O}$. 故, $\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_1=\operatorname{diag}(\boldsymbol{A}_{11},\boldsymbol{O})$, $\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{B}\boldsymbol{Q}_1=\operatorname{diag}(\boldsymbol{B}_{11},\boldsymbol{O})$.

对于实对称阵 $\boldsymbol{A}_{11}$, 存在 r 阶正交矩阵 $\boldsymbol{Q}_2$ 使得 $\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{A}_{11}\boldsymbol{Q}_2=\boldsymbol{D}_1$ 为对角阵. 此时,

$$\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{B}_{11}\boldsymbol{Q}_2=\boldsymbol{Q}_2^{\mathrm{T}}(\boldsymbol{A}_{11}+\boldsymbol{B}_{11})\boldsymbol{Q}_2-\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{A}_{11}\boldsymbol{Q}_2=\boldsymbol{Q}_2^{\mathrm{T}}\boldsymbol{I}_r\boldsymbol{Q}_2-\boldsymbol{D}_1=\boldsymbol{I}_r-\boldsymbol{D}_1$$

为 r 阶对角矩阵. 故, 我们取 $\boldsymbol{P}=\boldsymbol{Q}_1\operatorname{diag}(\boldsymbol{Q}_2,\boldsymbol{I}_{n-r})$ 即可. □

8.5　二次曲线与曲面的分类

在这一节里, 我们要通过坐标变换将一般的二次曲线 (曲面) 的方程化简. 保持图像形状不变的坐标变换有两种基本形式: 正交变换和平移变换. 我们一般默认是采用这两种坐标的变换. 注意, 平移变换不是线性变换.

学生提前自学教材 §2.2.5 的内容

二维情形　首先, 考察平面上的二次曲线 (quadratic curve)

$$a_{11}x^2+2a_{12}xy+a_{22}y^2+2b_1x+2b_2y+c=0.$$

第一步, 我们把图形摆正, 即将二次曲线的对称轴调整到与坐标轴平行的状态. 这意味着消去交叉项 xy. 注意到曲线中的二次项为

$$a_{11}x^2+2a_{12}xy+a_{22}y^2=(x,y)\begin{pmatrix}a_{11}&a_{12}\\a_{12}&a_{22}\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

将该二次型的矩阵记作 $\boldsymbol{A}$. 这个 $\boldsymbol{A}$ 为实对称阵, 可以通过正交矩阵, 相似 (相合) 于对角阵, 即存在正交阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\operatorname{diag}(\lambda_1,\lambda_2)$, 其中 λ_1,λ_2 为 $\boldsymbol{A}$ 的特征值.

此时, 若 $\boldsymbol{P}$ 为第一类正交矩阵, 则 $\boldsymbol{P}$ 对应于一个旋转, 可以写成

$$\boldsymbol{P}=\begin{pmatrix}\cos(\theta)&-\sin(\theta)\\\sin(\theta)&\cos(\theta)\end{pmatrix}$$

的形式. 若 $\boldsymbol{P}$ 为第二类正交矩阵, 设 $\boldsymbol{P}=(\boldsymbol{P}_1,\boldsymbol{P}_2)$, 考虑 $\boldsymbol{P}'=(\boldsymbol{P}_1,-\boldsymbol{P}_2)$, 则 $\boldsymbol{P}'$ 为第一类正交矩阵, 且 $(\boldsymbol{P}')^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{P}')=\operatorname{diag}(\lambda_1,\lambda_2)$. 故, 我们可以直接假设 $\boldsymbol{P}$ 是一个如上形式的旋转.

第二步, 我们来确定 $\boldsymbol{P}$ 的具体形式, 这意味着

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}=\begin{pmatrix}\cos(\theta)&\sin(\theta)\\-\sin(\theta)&\cos(\theta)\end{pmatrix}\begin{pmatrix}a_{11}&a_{12}\\a_{12}&a_{22}\end{pmatrix}\begin{pmatrix}\cos(\theta)&-\sin(\theta)\\\sin(\theta)&\cos(\theta)\end{pmatrix}$$

需要成为一个对角阵. 由于这已经是一个实对称阵了, 我们接下来只需要 $(1,2)$ 位置的元素为 0, 即

$$(-a_{11}+a_{22})\cos(\theta)\sin(\theta)+a_{12}(-\sin^2(\theta)+\cos^2(\theta))=0.$$

在 $a_{12}\neq 0$ 的条件下, 我们可以得到

$$\cot(2\theta)=\frac{a_{11}-a_{22}}{2a_{12}}.$$

解出适当的 θ 后, 我们可以写出 $\boldsymbol{P}$ (即通过三角函数的公式, 将 $\cos(\theta)$ 和 $\sin(\theta)$ 都用 $\cot(2\theta)$ 表示出来). 此时考虑坐标变换

$$\begin{pmatrix} x \\ y \end{pmatrix} = \boldsymbol{P} \begin{pmatrix} x' \\ y' \end{pmatrix}.$$

则二次曲线的方程可以化成

$$\lambda_1(x')^2 + \lambda_2(y')^2 + 2b_1'x' + 2b_2y' + c' = 0,$$

其中 λ_1, λ_2 为 $\boldsymbol{A}$ 的特征值 (注意顺序, 它们依赖于 $\boldsymbol{P}$ 的选取), 不全为 0, 而 $c' = c$.

在最后一步里, 我们通过配方来进一步化简曲线的方程. 其中, 若 $\lambda_1 \neq 0$, 我们令 $\tilde{x} = x' + b_1'/\lambda_1$, 否则, 令 $\tilde{x} = x'$. 类似地, 若 $\lambda_2 \neq 0$, 我们令 $\tilde{y} = y' + b_2'/\lambda_2$, 否则, 令 $\tilde{y} = y'$. 此时, 方程可以化成如下的标准形式.

(1) (椭圆型) 此时, $\lambda_1\lambda_2 > 0$, 即 λ_1 与 λ_2 同号:

$$\lambda_1\tilde{x}^2 + \lambda_2\tilde{y}^2 + \lambda_3 = 0.$$

当然, 依赖于 λ_3 是否与 λ_1 同号, 或为零, 该方程组可能无解, 有退化解, 或有正常解.

(2) (双曲线型) 此时, $\lambda_1\lambda_2 < 0$, 即 λ_1 与 λ_2 异号:

$$\lambda_1\tilde{x}^2 + \lambda_2\tilde{y}^2 + \lambda_3 = 0.$$

若 $\lambda_3 = 0$, 该方程为两条相交直线. 若 $\lambda_3 \neq 0$, 该方程为正常的双曲线.

(3) (抛物线型) λ_1 与 λ_2 之中有一个为零. 不妨设 $\lambda_2 = 0$. 则此时的方程为

$$\lambda_1\tilde{x}^2 + 2b_2'\tilde{y} + c' = 0.$$

(a) 若 $b_2' \neq 0$, 我们可以令 $\tilde{y}' = \tilde{y} + \frac{c'}{2b_2'}$, 可以进一步简化方程.

(b) 若 $b_2' = 0$, 则方程退化成 $\lambda_1\tilde{x}^2 + c' = 0$.

习题 8.5.1. 设二阶实对称矩阵 $\boldsymbol{A} = (a_{ij})$ 的特征值为 λ_1 和 λ_2 的, 求其元素 a_{12} 可能取的值之最大者和最小者.

三维情形　考察由如下方程给出的二次曲面 (quadratic surface)

$$a_{11}x^2+a_{22}y^2+a_{33}z^2+2a_{12}xy+2a_{13}xz+2a_{23}yz+2b_1x+2b_2y+2b_3z+c=0.$$

对于其二次项

$$(x,y,z)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

令二次型的矩阵为 $\boldsymbol{A}$. 我们寻找正交阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ 为对角阵. 此时, 令

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\boldsymbol{P}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix},$$

则二次曲面的方程可以化为

$$\lambda_1(x')^2+\lambda_2(y')^2+\lambda_3(z')^2+2b_1'x'+2b_2'y'+2b_3'z'+c'=0.$$

其中 $\lambda_1,\lambda_2,\lambda_3$ 为 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ 主对角线上的元素, 即 $\boldsymbol{A}$ 的所有特征值; 而 $c'=c$. 接下来, 根据参数是否为零, 我们做适当的坐标轴平移, 进一步简化方程. 由于具体情况较多, 我们就不具体讨论了.

教学的要求是学生能根据简化后的方程判断曲面的类型, 参见教材 §2.2.5 中的分类.

注 8.5.2. *此时维数为 3, 正交阵 $\boldsymbol{P}$ 一般不会有二维中的简单情形. 那么该如何求呢? 之前已经提过: 通过特征值来计算特征向量, 然后对特征向量作正交归一化. 当然, 二维的情形也可以采用这一办法来处理.*

例 8.5.3. *试指出二次曲面*

$$x^2+(2+\lambda)y^2+\lambda z^2+2xy-2xz-yz-5=0$$

中参数 λ 取什么值时, 该曲面为椭球面.

解. 设相应的二次型对应的矩阵为

$$\boldsymbol{A}=\begin{pmatrix} 1 & 1 & -1 \\ 1 & 2+\lambda & -1/2 \\ -1 & -1/2 & \lambda \end{pmatrix}.$$

于是, 存在正交矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$, 其中 $\lambda_1, \lambda_2, \lambda_3$ 是 $\boldsymbol{A}$ 的特征值. 此时, 在坐标变换 $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{Y}$ 下, 曲面的方程可化为 $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 - 5 = 0$. 这是一个椭球面的充要条件是 $\lambda_1, \lambda_2, \lambda_3$ 都是正数, 即 $\boldsymbol{A}$ 是正定的.

由于 $\boldsymbol{A}$ 的顺序主子式依次为

$$D_1 \coloneqq 1, \quad D_2 \coloneqq \begin{vmatrix} 1 & 1 \\ 1 & 2+\lambda \end{vmatrix} = 1 + \lambda, \quad D_3 \coloneqq \det(\boldsymbol{A}) = \lambda^2 - \frac{5}{4}.$$

为了使 $D_1, D_2, D_3 > 0$, 我们需要 $\lambda > \frac{\sqrt{5}}{2}$. □

例 8.5.4. 已知二次型 $f(x_1, x_2, x_3) = 5x_1^2 + 5x_2^2 + cx_3^2 - 2x_1x_2 + 6x_1x_3 - 6x_2x_3$ 的秩为 2.

(1) 求参数 c 及此二次型对应矩阵的特征值.

(2) 指出方程 $f(x_1, x_2, x_3) = 1$ 表示何种二次曲面.

解. (1) 二次型对应的矩阵为

$$\boldsymbol{A} = \begin{pmatrix} 5 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & c \end{pmatrix}.$$

由于 $\operatorname{rank}(\boldsymbol{A}) = 2$, 矩阵 $\boldsymbol{A}$ 不是满秩的, 从而 $\det(\boldsymbol{A}) = 24(c-3) = 0$, 从而 $c = 3$. 此时, 特征多项式 $p_{\boldsymbol{A}}(\lambda) = \lambda(\lambda-4)(\lambda-9)$, 从而有特征值 $\lambda_1 = 0$, $\lambda_2 = 4$, $\lambda_3 = 9$.

(2) 此时二次曲面的方程可以化为 $4y_2^2 + 9y_3^2 = 1$, 这对应于一个椭圆柱面. □

例 8.5.5. 用正交变换和平移将二次方程

$$x^2 + 2y^2 + 2z^2 - 4yz - 2x + 2\sqrt{2}y - 6\sqrt{2}z + 5 = 0$$

化为标准形式, 并判断它是什么曲面.

解. 曲面方程的二次部分对应的方阵为 $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 2 \end{pmatrix}$. 我们寻找正交矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ 为对角阵.

计算得特征多项式 $p_{\boldsymbol{A}}(\lambda) = \lambda(\lambda-1)(\lambda-4)$, 故 $\boldsymbol{A}$ 有特征值 $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 4$. 由于特征值互不相等, 对应的特征向量已经相互正交. 经具体计算知

- $\lambda_1 = 0$ 有特征向量 $\boldsymbol{\xi}_1 = (0, 1, 1)^{\mathrm{T}}$;

- $\lambda_2 = 1$ 有特征向量 $\boldsymbol{\xi}_2 = (1, 0, 0)^{\mathrm{T}}$;
- $\lambda_3 = 4$ 有特征向量 $\boldsymbol{\xi}_3 = (0, 1, -1)^{\mathrm{T}}$.

故, 可以选

$$\boldsymbol{P} = \left(\frac{1}{\sqrt{2}}\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \frac{1}{\sqrt{2}}\boldsymbol{\xi}_3\right) = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

这样得到的 $\boldsymbol{P}$ 为正交矩阵. 此时, $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \mathrm{diag}(0, 1, 4)$.

考虑坐标变换

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \boldsymbol{P} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}.$$

则二次曲面的方程化为

$$\underline{(y')^2 + 4(z')^2} - 2y' + 2\sqrt{2}\left(\frac{1}{\sqrt{2}}x' + \frac{1}{\sqrt{2}}z'\right) - 6\sqrt{2}\left(\frac{1}{\sqrt{2}}x' - \frac{1}{\sqrt{2}}z'\right) + 5 = 0,$$

其中的二次项部分可以由特征值的信息直接给出, 不需要额外计算

即

$$(y')^2 + 4(z')^2 - 4x' - 2y' + 8z' + 5 = 0.$$

进一步地, 令 $\tilde{x} = x'$, $\tilde{y} = y' - 1$, $\tilde{z} = z' + 1$, 则有

$$(\tilde{y})^2 + 4(\tilde{z})^2 - 4\tilde{x} = 0.$$

这是一个椭圆抛物面. □

习题 8.5.6. 对于曲面 $xy - yz + zx + 2x - z + 2 = 0$, 请通过正交变换和平移求出其标准方程, 并指出其曲面类型.