线性代数

(2024 年 12 月更新)

李思

清华大学

编写这份讲义的起因实属偶然，偶然中承载了一份默契与等待。

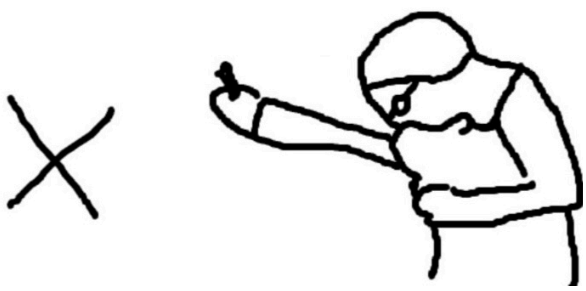
这份讲义是完全开源的，也欢迎读者来信交流。讲义的后续更新将在个人主页上发布

https://sili-math.github.io/

如果你觉得这份讲义有些帮助，并且有兴趣请我喝咖啡的话，我个人以及如下二维码都是非常欢迎的。

“知足常乐，厚积薄发”

特别致谢 2024 年秋季学期清华大学 6A-016 教室全体学生。



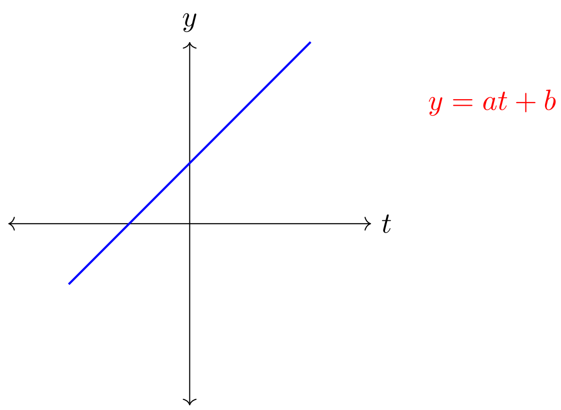
目录

第一章　线性空间

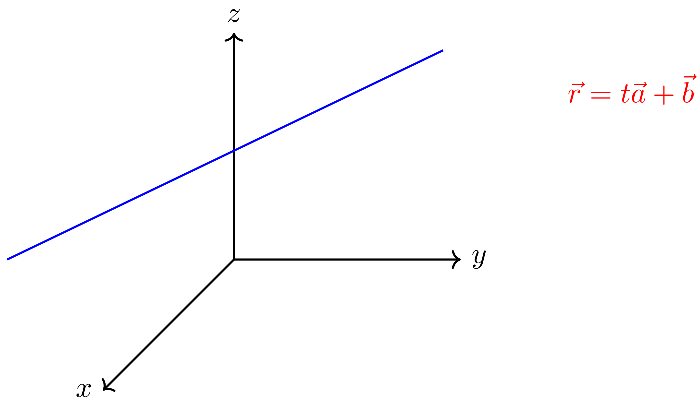
1.1　线性空间与子空间

1.1.1　线性空间

什么是“线性”？我们先了解如何描述一条直线。平面上的一条直线可以通过如下函数刻画：



这里把 t 看作时间，a, b 是常数。当 t 跑遍 $\mathbb{R}$ 时，这个函数关系在 (t, y) 平面上的轨迹是一条直线。类似的，$\mathbb{R}^3$ 中的直线可以用函数关系表示：



这里 $\vec{r} = (x, y, z)$ 是空间向量，$\vec{a}, \vec{b}$ 是两个常数向量，$t \in \mathbb{R}$ 是参数。当 t 跑遍 $\mathbb{R}$ 时，这个参数关系在 3 维空间中的轨迹是一条（经过 $\vec{b}$，沿着 $\vec{a}$ 方向的）直线。

由上述讨论知道，描述直线关系 $\vec{r} = t\vec{a} + \vec{b}$ 我们需要如下两个操作：

1. 加法：可以把两个向量相加，即：$\vec{a} + \vec{b}$。

2. 数乘：可以把一个数乘以一个向量，即：$t\vec{a}$。

简单来说，线性结构就是关于“加法”和“数乘”的结构。一个集合如果有“加法”和“数乘”这两个运算，我们称其为线性空间。为了讨论完备，我们以下给出线性空间的完整数学定义。定义看似很复杂，实际上列出的性质都是比较自然的，很容易在具体例子中检验。

定义 1.1.1. 一个数域 k（$k=\mathbb{R}$ 或者 $\mathbb{C}$）上的线性空间（或称向量空间）指的是一个集合 V 以及二元运算

$$\text{加法}: V\times V \xrightarrow{+} V \qquad \text{数乘}: k\times V\to V$$

使得对任何 $\mathbf{u},\mathbf{v},\mathbf{w}\in V, a,b\in k$，下列性质成立：

1. 加法交换律：$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
2. 加法结合律：$(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$
3. 零元素：存在唯一元素 $\mathbf{0}\in V$ 使得 $\mathbf{0}+\mathbf{u}=\mathbf{u}$
4. 负元素：存在唯一的 $-\mathbf{u}\in V$ 使得 $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$
5. 单位元：对 k 中的单位元 1，满足 $1\mathbf{u}=\mathbf{u}$
6. 数乘结合律：$a(b\mathbf{u})=(ab)\mathbf{u}$
7. 分配律（向量加法）；$a(\mathbf{u}+\mathbf{v})=a\mathbf{u}+a\mathbf{v}$
8. 分配律（数量加法）：$(a+b)\mathbf{u}=a\mathbf{u}+b\mathbf{u}$

线性空间 V 中的元素称为一个向量。

如不特殊说明，我们考虑实线性空间（即 $k=\mathbb{R}$）。

例子 1.1.2. 3 维空间 $\mathbb{R}^3$ 是一个实线性空间

- 加法：对于向量 $\mathbf{u}=(u_1,u_2,u_3),\mathbf{v}=(v_1,v_2,v_3)$

$$\mathbf{u}+\mathbf{v}=(u_1+v_1,u_2+v_2,u_3+v_3)$$

- 数乘：对 $\lambda\in\mathbb{R}$ 和向量 $\mathbf{u}=(u_1,u_2,u_3)$

$$\lambda\mathbf{u}=(\lambda u_1,\lambda u_2,\lambda u_3)$$

例子 1.1.3. 线性空间 $\mathbb{R}^n$ 中的一个向量表达为

$$\mathbf{u}=(u_1,u_2,\cdots,u_n),\quad u_i\in\mathbb{R}$$

- 加法：

$$\mathbf{u}+\mathbf{v}=(u_1+v_1,u_2+v_2,\cdots,u_n+v_n)$$

- 数乘：对 $\lambda\in\mathbb{R}$

$$\lambda\mathbf{u}=(\lambda u_1,\lambda u_2,\cdots,\lambda u_n)$$

例子 1.1.4. 所有的一元多项式组成的集合 $\mathbb{R}[x]$ 构成一个线性空间。给定两个多项式 $f(x), g(x)$，加法即为多项式相加 $f(x)+g(x)$，数乘即为多项式的数乘 $\lambda f(x)$。

类似的，所有的一元光滑函数构成的空间（记为 $C^\infty(\mathbb{R})$）是一个线性空间。

例子 1.1.5. 一个 n 次多项式指的是如下的多项式

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \quad 其中\ a_n \neq 0$$

- 所有 n 次多项式（$n \geq 1$）构成的集合在多项式加法和数乘下，不构成一个线性空间。例如两个 n 次多项式相减可能不再是 n 次多项式。特别地，这个集合不包含零多项式。

- 所有次数 $\leq n$ 的多项式构成的集合在多项式加法和数乘下构成一个线性空间。

例子 1.1.6. 量子力学和经典力学的一个本质区别是量子物理态构成一个复线性空间（*Hilbert* 空间）。例如一个量子自旋系统，一个自旋向上的态 $|\uparrow\rangle$ 和一个自旋向下的态 $|\downarrow\rangle$ 可以叠加得到一个新的物理态

$$\frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{1}{\sqrt{2}}|\downarrow\rangle$$

又例如干涉现象的本质也是做线性叠加。

例子 1.1.7. 设 $D = a_n(x)\frac{d^n}{dx^n} + \cdots + a_1(x)\frac{d}{dx} + a_0(x)$ 是一个微分算子。则所有满足如下微分方程的函数 $f(x)$

$$\{f(x) | Df(x) = 0\}$$

构成一个线性空间。这是因为如果 $f(x), g(x)$ 满足方程，则

$$D(f(x)+g(x)) = Df(x) + Dg(x) = 0 \implies f(x)+g(x)\ 满足方程$$
$$D(cf(x)) = cDf(x) = 0 \implies cf(x)\ 满足方程$$

1.1.2 直和

给定两个线性空间 V, W，我们可以定义一个新的线性空间 $V \oplus W$ 称为 V 和 W 的直和。作为集合

$$V \oplus W = \{(v, w) | v \in V, w \in W\}$$

加法和数乘定义为

$$(v_1, w_1) + (v_2, w_2) := (v_1 + v_2, w_1 + w_2)$$

$$\lambda(v, w) := (\lambda v, \lambda w)$$

类似可以定义 m 个线性空间 $V_1, \cdots, V_m$ 的直和

$$V_1 \oplus V_2 \oplus \cdots \oplus V_m$$

直和中的元素可以表示为 $(u_1, \cdots, u_m), u_i \in V_i$。

例子 1.1.8. 线性空间 $\mathbb{R}^n$ 可以写成 n 个 $\mathbb{R}$ 的直和

$$\mathbb{R}^n = \underbrace{\mathbb{R} \oplus \mathbb{R} \oplus \cdots \oplus \mathbb{R}}_{n}$$

如果 $n = m + k$，我们也可以写成

$$\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^k$$

1.1.3 线性子空间

定义 1.1.9. 设 V 是一个数域 k 上的线性空间。V 的一个子集 W 称为 V 的一个线性子空间，如果 W 中的元素在加法和数乘运算下封闭，即

- 对任意 $w_1, w_2 \in W$，则 $w_1 + w_2 \in W$。
- 对任意 $w \in W$ 和 $\lambda \in k$，则 $\lambda w \in W$。

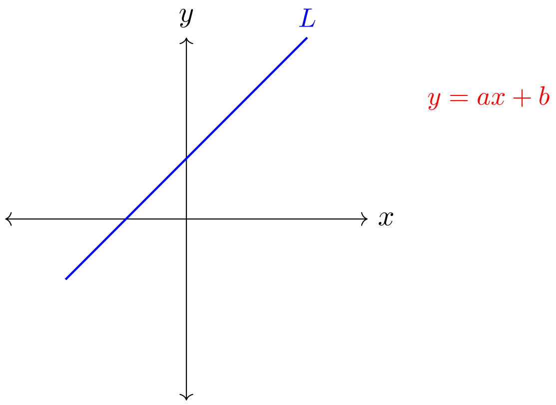
命题 1.1.10. V 的线性子空间是一个线性空间。

证明: 设 W 是 V 的线性子空间。验证 $\mathbf{0} \in W$:

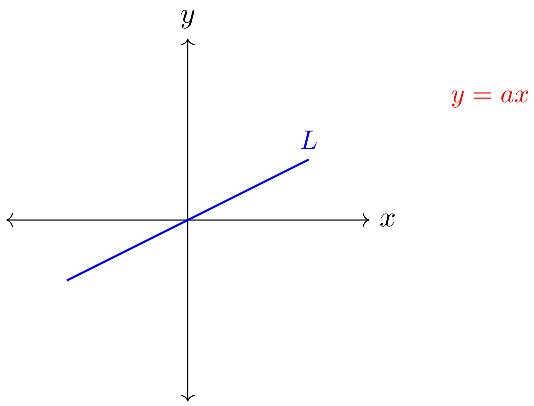
$$\mathbf{0} = w + (-w) \in W \quad \text{这里 } w \in W$$

容易检验 W 在 V 的加法和数乘运算下构成线性空间。 $\square$

例子 1.1.11. 考虑 $\mathbb{R}^2$ 的子集 $L = \{(x, y) | y = ax + b\} \subset \mathbb{R}^2$,



则 L 是 $\mathbb{R}^2$ 的线性子空间当且仅当 L 过原点。



实际上，原点 $\mathbf{0} = (0, 0)$ 是 $\mathbb{R}^2$ 的零元素，如果 L 是线性子空间则一定包含原点。反之，如果 L 过原点，则直线方程为

$$L = \{(x, y) | y = ax\}$$

若 $(x_1, y_1), (x_2, y_2) \in L$，则 $(x_1 + x_2, y_1 + y_2) \in L$

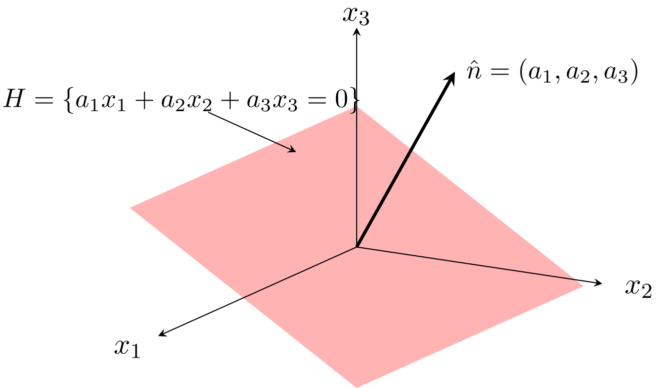
$$y_1 + y_2 = ax_1 + ax_2 = a(x_1 + x_2)$$

并且 $(\lambda x_1, \lambda y_1) \in L$，即

$$\lambda y_1 = \lambda(ax_1) = a(\lambda x_1)$$

因此 L 在加法和数乘下封闭。 $\square$

例子 1.1.12. 空间 $\mathbb{R}^3$ 中的平面 $H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 | a_1x_1 + a_2x_2 + a_3x_3 = b\}$ 是 $\mathbb{R}^3$ 的线性子空间当且仅当 H 过原点，即 $b = 0$。



例子 1.1.13. $V = \mathbb{R}[x]$，$W = \{$次数 $\leq n$ 的一元多项式$\}$。则 $W \subset V$ 是一个线性子空间。

例子 1.1.14. $V = \mathbb{R}[x]$，$W_{x_0} = \{f(x) \in V | f(x_0) = 0\}, x_0 \in \mathbb{R}$。则 $W_{x_0} \subset V$ 是线性子空间。对任意 $f(x), g(x) \in W_{x_0}$

$$(f+g)(x_0) = f(x_0) + g(x_0) = 0$$

因此 $f + g \in W_{x_0}$。类似可证对 $\lambda \in \mathbb{R}$，$\lambda f \in W_{x_0}$。

例子 1.1.15. $V = \mathbb{R}[x]$，$W = \{f(x) \in V | f(0) = 1\}$ 不是 V 的线性子空间。例如零函数不在 W 中。

例子 1.1.16. 考虑两个线性空间的直和 $V \oplus W$。则

$$V \oplus \mathbf{0} = \{(v, \mathbf{0}) | v \in V, \quad \mathbf{0} \in W \text{ 是零元}\}$$
$$\mathbf{0} \oplus W = \{(\mathbf{0}, w) | w \in W, \quad \mathbf{0} \in V \text{ 是零元}\}$$

是 $V \oplus W$ 的线性子空间。这是因为零元满足

$$\mathbf{0} + \mathbf{0} = \mathbf{0}, \quad \lambda\mathbf{0} = \mathbf{0}.$$

例子 1.1.17. $V = C^\infty(\mathbb{R})$，W 为 V 中满足如下微分方程的解

$$f'(x) = 2f(x)$$

W 是 V 的线性子空间。不难解出

$$W = \{ce^{2x}\}_{c \in \mathbb{R}}$$

例子 1.1.18. $V=C^{\infty}(\mathbb{R})$，W 为 V 中满足如下微分方程的解

$$f''(x)+f(x)=0$$

W 是 V 的线性子空间。不难解出

$$W=\{a\sin(x)+b\cos(x)\}_{a,b\in\mathbb{R}}$$

例子 1.1.19. 在量子力学中，满足定态薛定谔方程

$$\left(-\frac{\hbar^2}{2m}\nabla^2+V(x)\right)\psi(x)=E\psi(x)$$

的波函数 $\psi(x)$ 构成态空间的线性子空间。E 称为“能量本征值”，对应子空间包含能量 E 的量子态。

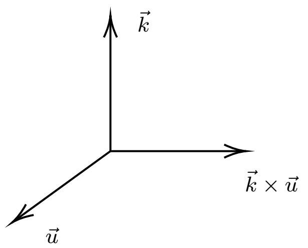
例子 1.1.20. 真空中麦克斯韦方程

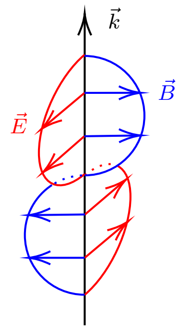
$$\begin{cases}\nabla\cdot\vec{E}=0\\ \nabla\cdot\vec{B}=0\end{cases}\quad\begin{cases}\nabla\times\vec{E}=-\frac{\partial\vec{B}}{\partial t}\\ \nabla\times\vec{B}=\frac{1}{c^2}\frac{\partial\vec{E}}{\partial t}\end{cases}$$

的解是一个线性空间。固定波矢 $\vec{k}$ 的如下解

$$\{\vec{E}|\vec{E}=\vec{u}\cos(\vec{k}\cdot\vec{r}-\omega t)\}_{\vec{u}\text{与}\vec{k}\text{垂直}}$$

是一个由线偏振光构成的线性子空间。





命题 1.1.21. V 的任意个线性子空间的交是线性子空间。

证明：设 $\{W_\alpha\}_{\alpha\in I}$ 是 V 的线性子空间。设

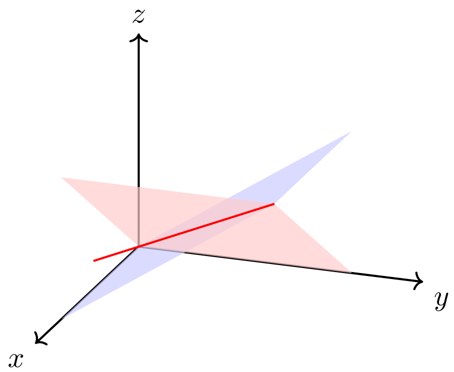
$$v_1,v_2\in\bigcap_{\alpha\in I}W_\alpha$$

则对每个 $\alpha\in I$，我们有 $v_1+v_2\in W_\alpha$，因此

$$v_1+v_2\in\bigcap_{\alpha\in I}W_\alpha$$

类似可证 $\lambda v_1\in\bigcap\limits_{\alpha\in I}W_\alpha$。 $\square$

例子 1.1.22. $\mathbb{R}^3$ 中不平行的两个平面的交是一条直线



例子 1.1.23. $V = C^\infty(\mathbb{R})$，U 为 V 中满足如下方程的解

$$f''(x) + f(x) = 0, \quad f(0) = 0.$$

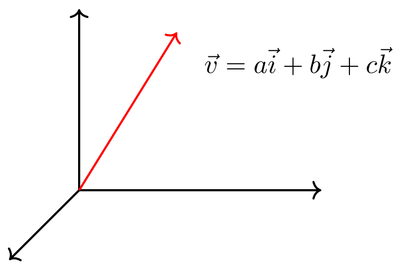
U 是 V 的两个线性子空间 $\{f \in V | f''(x) + f(x) = 0\}$ 和 $\{f \in V | f(0) = 0\}$ 的交，因此是线性子空间。具体而言

$$U = \{a \sin(x)\}_{a \in \mathbb{R}}.$$

1.2 线性组合与线性相关性

1.2.1 线性组合

线性空间中的元素通常可以通过一些比较基本的元素表达出来。例如可以把 $\mathbb{R}^3$ 中的矢量写成



这里 $\vec{v} = (a, b, c)$。$\vec{i}$, $\vec{j}$, $\vec{k}$ 是单位向量

$$\vec{i} = (1, 0, 0) \quad \vec{j} = (0, 1, 0) \quad \vec{k} = (0, 0, 1)$$

定义 1.2.1. 设 $\mathbf{v}_1, \cdots, \mathbf{v}_m \in V$，$\lambda_1, \cdots, \lambda_m \in k$。则向量

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m$$

称为向量 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 的线性组合，或者称 $\mathbf{u}$ 可以由 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 线性表达。所有可以由 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 线性表达的向量构成的集合记为

$$\mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$$

命题 1.2.2. $\mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ 是 V 的线性子空间。

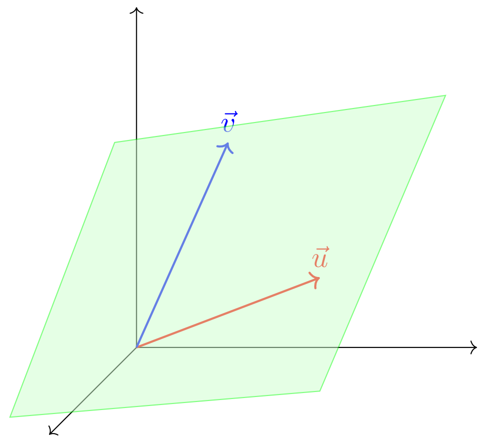
证明：设 $\mathbf{u}, \mathbf{w} \in \mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$，即

$$\mathbf{u} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_m\mathbf{v}_m \qquad \mathbf{v} = \beta_1\mathbf{v}_1 + \cdots + \beta_m\mathbf{v}_m$$

$$\Longrightarrow \mathbf{u} + \mathbf{v} = (\alpha_1 + \beta_1)\mathbf{v}_1 + \cdots + (\alpha_m + \beta_m)\mathbf{v}_m$$

属于 $\mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$。类似可证数乘封闭。 □

例子 1.2.3. $\mathbb{R}^3$ 中 $\mathrm{Span}\{\vec{u}, \vec{v}\}$ 是由 $\vec{u}, \vec{v}$ 两个向量张成的平面（这里假设 $\vec{u}, \vec{v}$ 不共线）



例子 1.2.4. 微分方程

$$f''(x) + f(x) = 0$$

所有的解构成的线性空间可以写成

$$\mathrm{Span}\{\cos(x), \sin(x)\}$$

线性表达的方式有可能并不唯一。例如考虑

$$\vec{v}_1 = (1, -1, 0), \quad \vec{v}_2 = (0, 1, -1), \quad \vec{v}_3 = (1, 0, -1)$$

则向量 $\vec{u} = (2, 0, -2)$ 可以线性表达为

$$\vec{u} = \vec{v}_1 + \vec{v}_2 + \vec{v}_3$$

也可以线性表达为

$$\vec{u} = 2\vec{v}_1 + 2\vec{v}_2$$

那么什么时候线性表达的方式是唯一的？为了回答这个问题，需要引入线性相关的概念。

1.2.2 线性相关性

定义 1.2.5. V 中向量 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 称为是线性相关的，如果存在不全为 0 的数 $\lambda_1, \cdots, \lambda_m \in k$ 使得

$$\lambda_1\mathbf{v}_1 + \cdots + \lambda_m\mathbf{v}_m = \mathbf{0}$$

否则，称 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 是线性无关的。

如果 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 是线性相关的，即存在不全为 0 的数 $\lambda_1, \cdots, \lambda_m \in k$ 使得

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m = \mathbf{0}$$

不妨设其中 $\lambda_i \neq 0$，则

$$\mathbf{v_i} = -\frac{1}{\lambda_i}(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_{i-1}\mathbf{v}_{i-1} + \lambda_{i+1}\mathbf{v}_{i+1} + \cdots \lambda_m \mathbf{v}_m)$$

即 $\mathbf{v}_i$ 可以由其他的向量线性表达。

反之亦然，如果 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 中有某个向量可以被其他向量线性表达，不妨设

$$\mathbf{v}_i = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots \alpha_m \mathbf{v}_m$$

则

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_{i-1}\mathbf{v}_{i-1} + (-1)\mathbf{v}_i + \alpha_{i+1}\mathbf{v}_{i+1} + \cdots \alpha_m \mathbf{v}_m = \mathbf{0}$$

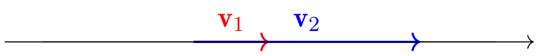
因此 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 线性相关。由此我们证明了如下结论

命题 1.2.6. *向量 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 线性相关当且仅当存在其中某个向量可以被其他向量线性表达。*

因此我们可以把线性相关的向量理解为这些向量中有冗余的线性信息。

例子 1.2.7. *两个非零向量 $\mathbf{v}_1, \mathbf{v_2}$ 是线性相关的当且仅当它们在同一直线上，即存在 $\lambda \neq 0$ 使得*

$$\mathbf{v}_2 = \lambda \mathbf{v}_1$$



例子 1.2.8. *考虑 $\mathbb{R}^3$ 中的向量*

$$\vec{v}_1 = (1, -1, 0), \quad \vec{v}_2 = (0, 1, -1), \quad \vec{v}_3 = (-1, 0, 1)$$

可以看出

$$\vec{v}_1 + \vec{v}_2 + \vec{v}_3 = \mathbf{0}$$

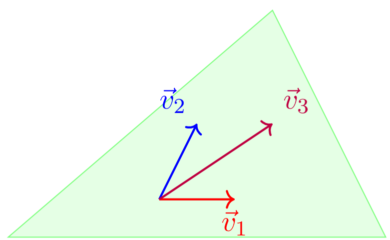
因此它们是线性相关的。比如 $\vec{v}_3$ 可以线性表达为

$$\vec{v}_3 = -\vec{v}_1 - \vec{v}_2$$

例子 1.2.9. *$\mathbb{R}^3$ 中的 3 个向量 $\vec{v}_1, \vec{v}_2, \vec{v}_3$ 是线性相关的当且仅当它们在同一个平面上。实际上，不妨设*

$$\vec{v}_3 = a\vec{v}_1 + b\vec{v}_2.$$

则 $\vec{v}_3$ 包含在 $\vec{v}_1$ 和 $\vec{v}_2$ 张成的平面。



1.2.3 线性无关性

下面我们来考察线性无关性。$\mathbf{v}_1, \cdots, \mathbf{v}_m$ 是线性无关的等价表述为：如果 $\lambda_1, \cdots, \lambda_m \in k$ 满足

$$\lambda_1\mathbf{v}_1 + \cdots + \lambda_m\mathbf{v}_m = \mathbf{0}$$

则一定有

$$\lambda_1 = \cdots = \lambda_m = 0$$

换言之，零元只能由 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 平凡地线性表达。

例子 1.2.10. 考虑 $\mathbb{R}^3$ 中相互垂直的单位向量

$$\vec{i} = (1,0,0) \quad \vec{j} = (0,1,0) \quad \vec{k} = (0,0,1)$$

假如有 $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ 使得

$$\lambda_1\vec{i} + \lambda_2\vec{j} + \lambda_3\vec{k} = \mathbf{0}$$

即 $(\lambda_1, \lambda_2, \lambda_3) = \mathbf{0}$ 是零向量，则每个分量都是 0

$$\lambda_1 = \lambda_2 = \lambda_3 = 0$$

因此 $\vec{i}, \vec{j}, \vec{k}$ 是线性无关的。它们不在同一平面上。

命题 1.2.11. 假设 $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ 线性无关，则 S 的任意子集的向量组都是线性无关的。

证明：不妨考虑子集 $\{\mathbf{v}_1, \cdots, \mathbf{v}_s\}$ $(s \le m)$。假设

$$\lambda_1\mathbf{v}_1 + \cdots + \lambda_s\mathbf{v}_s = \mathbf{0}$$

把它写成

$$\lambda_1\mathbf{v}_1 + \cdots + \lambda_s\mathbf{v}_s + 0\mathbf{v}_{s+1} + \cdots + 0\mathbf{v}_m = \mathbf{0}$$

由 S 的线性无关性可以推知 $\lambda_1 = \cdots = \lambda_s = 0$。因此 $\{\mathbf{v}_1, \cdots, \mathbf{v}_s\}$ 是线性无关的。 □

例子 1.2.12. 我们知道 $\mathbb{R}^3$ 中向量

$$\vec{i} = (1,0,0) \quad \vec{j} = (0,1,0) \quad \vec{k} = (0,0,1)$$

是线性无关的。因此其中任意两个，比如 $\vec{i}$ 和 $\vec{j}$，是线性无关的。

命题 1.2.13. 假设 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 线性无关，$\mathbf{u}$ 是它们的线性组合。则存在唯一的 $\lambda_1, \cdots, \lambda_m \in k$ 使得

$$\mathbf{u} = \lambda_1\mathbf{v}_1 + \cdots + \lambda_m\mathbf{v}_m$$

证明：假设 $\mathbf{u}$ 有两个线性表达

$$\mathbf{u} = \lambda_1\mathbf{v}_1 + \cdots + \lambda_m\mathbf{v}_m \quad \mathbf{u} = \mu_1\mathbf{v}_1 + \cdots + \mu_m\mathbf{v}_m$$

两式相减我们得到

$$(\lambda_1 - \mu_1)\mathbf{v}_1 + \cdots + (\lambda_m - \mu_m)\mathbf{v}_m = \mathbf{0}$$

由线性无关性可知，$\lambda_1 = \mu_1, \cdots, \lambda_m = \mu_m$. □

因此如果 $\mathbf{v}_1, \cdots, \mathbf{v}_m$ 是线性无关的向量，则 $\text{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ 中任一向量 $\mathbf{u}$ 可以唯一表达为

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m$$

因此 $\mathbf{u}$ 可以一一对应于一组数 $(\lambda_1, \cdots, \lambda_m)$。这组数可以想象成标识 $\mathbf{u}$ 的“坐标”。我们之后将会系统讨论这个坐标表达的方法。

例子 1.2.14. 平面 $\mathbb{R}^2$ 中的向量 $\vec{v}_1 = (1, 0)$ 和 $\vec{v}_2 = (1, 1)$ 不共线，因此是线性无关的。它们的线性组合张成整个平面

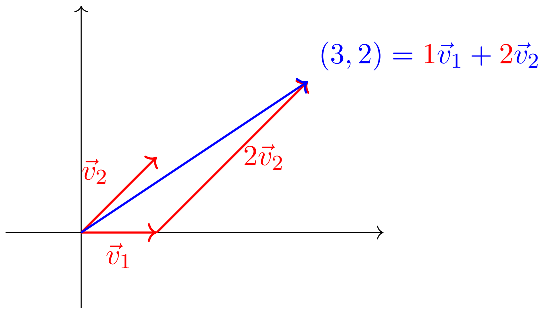
$$\mathbb{R}^2 = \text{Span}\{\vec{v}_1, \vec{v}_2\}$$

对于向量 $\vec{u} = (a, b)$，我们把它线性表达为

$$(a, b) = \lambda_1 \vec{v}_1 + \lambda_2 \vec{v}_2 = (\lambda_1 + \lambda_2, \lambda_2)$$

对比分量我们可以解出 $\lambda_1 = a - b, \lambda_2 = b$。例如

$$(3, 2) = 1\vec{v}_1 + 2\vec{v}_2$$



向量 $(3, 2)$ 如果用 $\vec{v}_1, \vec{v}_2$ 来线性表达的话，它被标记的“坐标”是 1 和 2。这个和原始的坐标 3 和 2 是不一样的。我们之后会详细讨论它们的关系。

定义 1.2.15. 设 $S, T \subset V$ 是线性空间 V 的两个子集。我们称 S 可以由 T 线性表达，如果 S 中的每个向量都可以由 T 线性表达，即

$$S \subset \text{Span} \ \ T$$

S 和 T 称为线性等价，如果 S 和 T 可以相互线性表达，即 S 可以被 T 线性表达，T 可以被 S 线性表达。

命题 1.2.16. S 可以被 T 线性表达当且仅当 $\text{Span}\, S \subset \text{Span}\, T$

证明：假设 $\text{Span}\, S \subset \text{Span}\, T$。由 $S \subset \text{Span}\, S$ 知 $S \subset \text{Span}\, T$，即 S 可以被 T 线性表达。

反之，假设 S 可以被 T 线性表达。对 $\text{Span}\, S$ 中任一向量 $\mathbf{u}$，它可以线性表达为

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m \quad \text{其中} \quad \mathbf{v}_i \in S \subset \text{Span}\, T$$

$\text{Span}\, T$ 是线性空间，因此 $\mathbf{u} \in \text{Span}\, T$。 □

命题 1.2.17. S 和 T 线性等价当且仅当 $\operatorname{Span} S = \operatorname{Span} T$。

证明：由前述命题可知

$$S \text{ 可以被 } T \text{ 线性表达} \Leftrightarrow \operatorname{Span} S \subset \operatorname{Span} T$$
$$T \text{ 可以被 } S \text{ 线性表达} \Leftrightarrow \operatorname{Span} T \subset \operatorname{Span} S$$

因此 S 和 T 线性等价当且仅当 $\operatorname{Span} S = \operatorname{Span} T$。 □

例子 1.2.18. 考虑 $\mathbb{R}^2$ 中的向量。如果两个向量 $\vec{v}_1, \vec{v}_2$ 不共线，另外两个 $\vec{u}_1, \vec{u}_2$ 也不共线，则

$$\operatorname{Span}\{\vec{v}_1, \vec{v}_2\} = \mathbb{R}^2 = \operatorname{Span}\{\vec{u}_1, \vec{u}_2\}$$

都张成整个平面，故 $\{\vec{v}_1, \vec{v}_2\}$ 与 $\{\vec{u}_1, \vec{u}_2\}$ 线性等价。

例子 1.2.19. 设 $S = \{\vec{v}_1, \vec{v}_2\}$ 是 $\mathbb{R}^3$ 不共线的两个向量，$T = \{\vec{u}_1, \vec{u}_2\}$ 是 $\mathbb{R}^3$ 另外两个不共线的向量。则 S 与 T 线性等价当且仅当 S 和 T 张成 $\mathbb{R}^3$ 中同一个平面。

1.3 基与维数

1.3.1 线性空间的基

线性空间中的元素可以通过一些更基本的向量来线性表达。这个想法给出了“基”的概念。

定义 1.3.1. 线性空间 V 的一组向量 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 称为 V 的一组基，如果它们线性无关，并且 V 中的每个向量都可以通过它们来线性表达，即

$$V = \operatorname{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}.$$

我们只考虑有限个向量构成一组基的情况。有时候还需要无穷个向量来构成一组基（例如量子力学中的态空间），这个是泛函分析里讨论的内容。我们这里不考虑无穷的情况。

考虑实线性空间 V。如果 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 构成 V 的一组基，那么 V 中任一向量 $\mathbf{u}$ 可以唯一地写成

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n, \quad \lambda_i \in \mathbb{R}$$

反之，给定实数 $\lambda_1, \cdots, \lambda_n$，上述公式给出了 V 中的一个向量。由此我们证明了如下结论。

命题 1.3.2. 设 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是 V 的一组基。则映射

$$\varphi : \mathbb{R}^n \to V$$
$$(\lambda_1, \cdots, \lambda_n) \mapsto \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n$$

给出了 $\mathbb{R}^n$ 和 V 之间的一一对应。

因此，如果我们找到 V 的一组基，那么就可以用一串数字 $\lambda_1, \cdots, \lambda_n$ 来唯一的表示 V 中的向量。

例子 1.3.3. $\mathbb{R}^3$ 中相互垂直的单位向量

$$\vec{i}=(1,0,0) \quad \vec{j}=(0,1,0) \quad \vec{k}=(0,0,1)$$

构成 $\mathbb{R}^3$ 的一组基。任一向量 (a,b,c) 可以表达为

$$(a,b,c)=a\vec{i}+b\vec{j}+c\vec{k}$$

因此向量 (a,b,c) 可以通过 a,b,c 这 3 个数字来标记。

例子 1.3.4. 类似的，$\mathbb{R}^n$ 中的一组向量

$$\begin{aligned}\vec{e}_1&=(1,0,\cdots,0)\\ \vec{e}_2&=(0,1,\cdots,0)\\ &\cdots\\ \vec{e}_n&=(0,0,\cdots,1)\end{aligned}$$

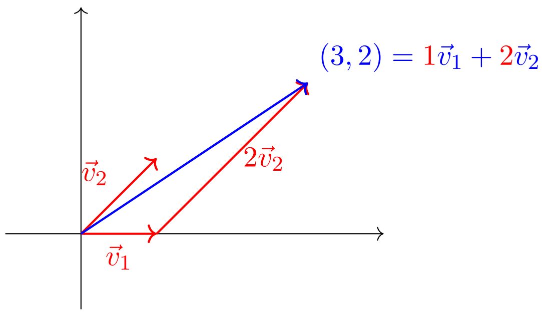
构成了 $\mathbb{R}^n$ 的一组基，称为 $\mathbb{R}^n$ 的标准基。

例子 1.3.5. $V=\mathbb{R}^2$ 中的向量 $\vec{v}_1=(1,0)$ 和 $\vec{v}_2=(1,1)$ 构成一组基。对于向量 $\vec{u}=(a,b)$，我们把它线性表达为

$$(a,b)=\lambda_1\vec{v}_1+\lambda_2\vec{v}_2=(\lambda_1+\lambda_2,\lambda_2)$$

这组基给出了一个一一对应

$$\begin{aligned}\varphi:\mathbb{R}^2 &\to V\\ (\lambda_1,\lambda_2) &\mapsto \lambda_1\vec{v}_1+\lambda_2\vec{v}_2=(\lambda_1+\lambda_2,\lambda_2)\end{aligned}$$



这里 φ 将基坐标 $(1,2)$ 对应到向量 $1\vec{v}_1+2\vec{v}_2$

$$\varphi:(1,2)\to 1\vec{v}_1+2\vec{v}_2=(3,2)$$

1.3.2 维数

自然有如下两个问题。第一，如何构造一组基？第二，注意到基的选取并不唯一，那么不同的基之间是什么关系？

假设 $\{\mathbf{v}_1,\mathbf{v}_2,\cdots,\mathbf{v}_n\}$ 是 V 的一组基，$\{\mathbf{u}_1,\cdots,\mathbf{u}_m\}$ 是 V 的另一组基。我们下面证明 $n=m$，即不同基中的向量个数一定是一样的。

引理 1.3.6. 假设非零向量 $\mathbf{u}$ 可以由一组向量 $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 线性表达，则存在 i 使得

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} \text{ 与 } \{\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \mathbf{u}, \mathbf{v}_{i+1}, \cdots, \mathbf{v}_n\} \text{ 线性等价。}$$

证明：$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n$。不妨设 $\lambda_i \neq 0$，则

$$\mathbf{v}_i = -\frac{1}{\lambda_i}(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_{i-1}\mathbf{v}_{i-1} - \mathbf{u} + \cdots + \lambda_n \mathbf{v}_n)$$

因此

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} \subset \mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \mathbf{u}, \mathbf{v}_{i+1}, \cdots, \mathbf{v}_n\}$$

另一方面显然有

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \mathbf{u}, \mathbf{v}_{i+1}, \cdots, \mathbf{v}_n\} \subset \mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$$

□

命题 1.3.7. 假设 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 线性无关，并可以由一组向量 $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 线性表达。则可以用 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 替换掉 $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 中的某 m 个向量并保持其与 $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 线性等价。特别的，我们有 $m \leq n$。

证明：由假设 $\mathbf{u}_1 \in \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$。由引理1.3.6知存在某个 $\mathbf{v}_i$，不妨设是 $\mathbf{v}_1$，使得

$$\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\} = \mathrm{Span}\{\mathbf{u}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$$

下面考虑 $\mathbf{u}_2$。由假设 $\mathbf{u}_2 \in \mathrm{Span}\{\mathbf{u}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$，于是

$$\mathbf{u}_2 = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{v}_2 + \cdots \lambda_n \mathbf{v}_n$$

其中 $\lambda_2, \cdots, \lambda_n$ 不能全为 0，否则 $\mathbf{u}_2 = \lambda_1 \mathbf{u}_1$ 与 $\mathbf{u}_1, \mathbf{u}_2$ 线性无关的假设矛盾。不妨设 $\lambda_2 \neq 0$，则由引理1.3.6知

$$\mathrm{Span}\{\mathbf{u}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\} = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_3 \cdots, \mathbf{v}_n\}.$$

重复这个过程，我们证明可以用 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 替换掉 $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 中的 m 个向量并保持线性等价性。特别的必然有 $m \leq n$。 □

命题 1.3.8. 假设 $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 是 V 的一组基，$\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 是 V 的另一组基，则 $n = m$。

证明：由假设，$\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ 线性无关，并且可以被 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 线性表达。由命题1.3.7知

$$n \leq m.$$

同理知 $m \leq n$。因此 $n = m$。 □

定义 1.3.9. 设 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 是线性空间 V 的一组基，则 n 称为 V 的维数，记为

$$\dim V = n.$$

V 称为一个 n 维线性空间。

由命题1.3.8知，这样定义的维数其实与基的选取没有关系，是 V 本身的性质。当然也存在无穷维的线性空间，我们这里只讨论有限维的情况。

例子 1.3.10. $\mathbb{R}^n$ 中的一组基为

$$\begin{aligned}\vec{e}_1 &= (1, 0, \cdots, 0)\\ \vec{e}_2 &= (0, 1, \cdots, 0)\\ &\vdots\\ \vec{e}_n &= (0, 0, \cdots, 1)\end{aligned}$$

因此 $\mathbb{R}^n$ 的维数是 n。

例子 1.3.11. 微分方程

$$f''(x) + f(x) = 0$$

解构成的空间 V 的一组基是 $\{\cos(x), \sin(x)\}$，因此 $\dim V = 2$。

命题 1.3.12. 设 V 是 n 维线性空间，$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 是 n 个线性无关的向量。则它们构成 V 的一组基。

证明：由定义，V 存在一组基 $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$。由于 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 线性无关，且可以由 $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ 线性表达，由命题1.3.7知可以用 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 替换 $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ 中的 n 个向量（即全部替换），使得

$$\mathrm{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} = \mathrm{Span}\{\mathbf{u}_1, \cdots, \mathbf{u}_n\} = V$$

因此 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 是 V 的一组基。 □

因此在 n 维空间里找一组基，等价于找 n 个线性无关的向量。

例子 1.3.13. $\mathbb{R}^3$ 中任意 3 个不共面的向量 $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ 线性无关，因此构成 $\mathbb{R}^3$ 的一组基。

命题 1.3.14. 设 V 是一个 n 维线性空间，U 是线性子空间，则

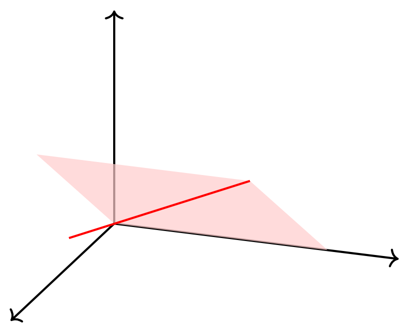
$$\dim U = m \leq \dim V = n$$

并且 U 的任意一组基 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 可以扩展为 V 的一组基 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m, \mathbf{v}_{m+1}, \cdots, \mathbf{v}_n\}$。

证明：选 U 的基 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$，$V$ 的基 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$。由于 $U \subset V$，则 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 线性无关且可以由 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 线性表达。由命题1.3.7知 $m \leq n$，并且可以用 $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ 替换 $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ 中 m 个向量构成 V 的基，从而扩展性得证。 □

例子 1.3.15. 设 H 是 $\mathbb{R}^3$ 中过原点的一个平面，L 是 H 中过原点的一条直线。则 $L \subset H \subset \mathbb{R}^3$

$$\dim L = 1 < \dim H = 2 < \dim \mathbb{R}^3 = 3.$$



1.3.3 向量组的秩

定义 1.3.16. 设 S 是 V 的一组向量。如果 S 中的向量 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 线性无关，并且对任意 $\mathbf{u} \in S$，向量 $\{\mathbf{u}, \mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 均线性相关。我们称 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 是 S 的一个极大线性无关组。

命题 1.3.17. S 中的向量 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 构成 S 的一个极大线性无关组当且仅当它们是 $\operatorname{Span} S$ 的一组基。因此

$$r = \dim \operatorname{Span} S$$

证明：假设 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 是 $\operatorname{Span}\{S\}$ 的一组基。则任意 $\mathbf{u} \in S$ 可以写成 $\mathbf{u} = \lambda_1 \mathbf{u}_1 + \cdots \lambda_r \mathbf{u}_r$，于是 $\{\mathbf{u}, \mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 线性相关。由 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 是线性无关的，因此它们是极大线性无关组。

反之，假设 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 是极大线性无关组。则对任意 $\mathbf{u} \in S$，$\{\mathbf{u}, \mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 是线性相关的，即存在

$$\lambda_0 \mathbf{u} + \lambda_1 \mathbf{u}_1 + \cdots + \lambda_r \mathbf{u}_r = \mathbf{0}$$

这里 λ_i 不全为 0。如果 $\lambda_0 = 0$，由 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 线性无关知 $\lambda_1 = \cdots = \lambda_r = 0$，矛盾。故 $\lambda_0 \neq 0$，因此 $\mathbf{u}$ 可以由 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 线性表达。由 $\mathbf{u}$ 的任意性知 S 中元素均可以由 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 线性表达，因此

$$\operatorname{Span} S = \operatorname{Span}\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}.$$

这说明 $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ 构成 $\operatorname{Span} S$ 的一组基。 □

由此我们知道，不同的极大线性无关组包含的向量个数是一样的，都是 S 张成的线性空间的维数。S 的一组极大线性无关组刨除了 S 中的冗余向量，并保留了 S 中包含的所有线性信息。

定义 1.3.18. 一组向量 $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ 的极大线性无关组的向量个数 $r = \dim \operatorname{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ 称为向量组 $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ 的秩，记为

$$r = \operatorname{rank}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}.$$

一组向量的秩描述了它们张成线性空间的维数。

例子 1.3.19. 如果一组向量 $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ 是线性无关的，那么

$$\operatorname{rank}\{\mathbf{v}_1, \cdots, \mathbf{v}_m\} = m.$$

例子 1.3.20. 考虑 $\mathbb{R}^3$ 中的一组向量

$$\vec{v}_1 = (1, -1, 0), \quad \vec{v}_2 = (0, 1, -1), \quad \vec{v}_3 = (-1, 0, 1)$$

它们满足

$$\vec{v}_1 + \vec{v}_2 + \vec{v}_3 = \mathbf{0}$$

容易看出 $\{\vec{v}_1, \vec{v}_2\}$ 构成极大线性无关组。比如 $\vec{v}_3$ 可以线性表达为 $\vec{v}_3 = -\vec{v}_1 - \vec{v}_2$。同理 $\{\vec{v}_1, \vec{v}_3\}$ 或者 $\{\vec{v}_2, \vec{v}_3\}$ 均是极大线性无关组。这组向量的秩是 2。

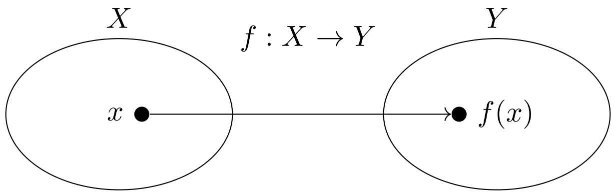
1.4 线性映射与矩阵

1.4.1 集合的映射

两个集合 X, Y 之间的一个映射

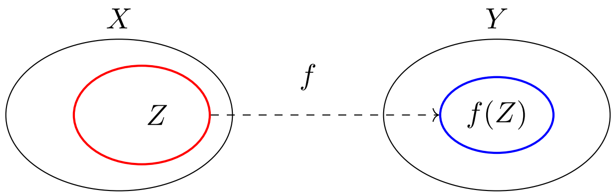
$$f : X \to Y$$

指的是一个对应法则，将 X 中的一个元素 x 唯一对应到 Y 中的一个元素 y，记为 $y = f(x)$



设 $f : X \to Y$ 是一个映射。X 的一个子集 Z 在 f 下的像，指的是所有由 Z 中的元素在映射 f 下构成的 Y 中元素的集合，记为 $f(Z) \subset Y$

$$f(Z) := \{y \in Y |\ \text{存在}\ \ x \in Z\ \ \text{使得}\ y = f(x)\}$$



我们记 $\mathrm{im}(f) := f(X)$，称为映射 f 的像。

定义 1.4.1. 映射 $f : X \to Y$ 称为是一个满射，如果

$$\mathrm{im}(f) = Y$$

即 Y 中每个元素都通过 f 对应于 X 中某个元素。

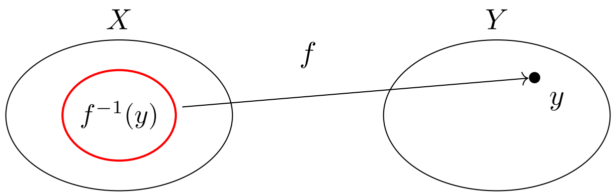
映射 $f : X \to Y$ 称为一个单射，如果对任意两个元素 $x_1, x_2 \in X$

$$x_1 \neq x_2 \quad \Longrightarrow \quad f(x_1) \neq f(x_2)$$

即不同的元素的像也一定是不同的。

$y \in Y$ 在映射 $f : X \to Y$ 下的原像定义为

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



如果 $y \notin \mathrm{im}(f)$，则 $f^{-1}(y) = \varnothing$ 是空集。

不难看出，$f : X \to Y$ 是单射当且仅当对任意 $y \in \mathrm{im}(f)$，$f^{-1}(y)$ 只含有一个元素。

定义 1.4.2. 映射 $f : X \to Y$ 称为是一一映射，如果 f 即是单射也是满射。此时 f 给出了集合 X 和 Y 之间的一个一一对应关系。

1.4.2 线性映射

下面我们考虑两个线性空间之间的映射。由于线性空间上具有线性结构，我们关注那些保持线性结构的映射。这样的映射称为线性映射。

定义 1.4.3. 两个线性空间 V, W 之间的映射 $f: V \to W$ 称为一个线性映射，如果 f 满足

- 保加法：对任意的 $\mathbf{v}_1, \mathbf{v}_2 \in V$

$$f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2),$$

- 保数乘：对任意的 $\mathbf{v} \in V, \lambda \in k$

$$f(\lambda \mathbf{v}) = \lambda f(\mathbf{v}).$$

特别的，f 一定把零向量映到零向量。

例子 1.4.4. 考虑如下映射 $f: \mathbb{R}^2 \to \mathbb{R}$，记作 $(x, y) \mapsto f(x, y)$

- $f(x, y) = x$. f 是一个线性映射
- $f(x, y) = 2x + 3y$. f 是一个线性映射
- $f(x, y) = 2x + 3y - 1$. f 不是一个线性映射
- $f(x, y) = x^2 + y$. f 不是一个线性映射

例子 1.4.5. 设 $V = \mathbb{R}[x]$ 是一元多项式空间。给定 $a \in \mathbb{R}$，我们定义赋值映射

$$\delta_a : \mathbb{R}[x] \to \mathbb{R}, \quad \delta_a(f) := f(a)$$

则 δ_a 是一个线性映射：

- $\delta_a(f + g) = f(a) + g(a) = \delta_a(f) + \delta_a(g)$
- $\delta_a(\lambda f) = \lambda f(a) = \lambda \delta_a(f)$

例子 1.4.6. 类似的，给定 $a_1, \cdots, a_n \in \mathbb{R}$，赋值映射

$$\delta : \mathbb{R}[x] \to \mathbb{R}^n$$
$$\delta(f) := (f(a_1), \cdots, f(a_n))$$

是一个线性映射。而映射

$$\varphi : \mathbb{R}[x] \to \mathbb{R}, \quad \varphi(f) = f(a_1) \cdots f(a_n)$$

不是一个线性映射（$n > 1$）。

命题 1.4.7. 设 $f: V \to W$ 是一个线性映射，$\mathbf{v}_i \in V, \lambda_i \in k$。则

$$f(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m) = \lambda_1 f(\mathbf{v}_1) + \cdots + \lambda_m f(\mathbf{v}_m).$$

即 f 保持线性组合的结构。

证明：

$$f(\lambda_1\mathbf{v}_1+\cdots+\lambda_m\mathbf{v}_m)\overset{\text{保加法}}{=}f(\lambda_1\mathbf{v}_1)+\cdots+f(\lambda_m\mathbf{v}_m)\overset{\text{保数乘}}{=}\lambda_1 f(\mathbf{v}_1)+\cdots+\lambda_m f(\mathbf{v}_m)$$

□

命题 1.4.8. 设 V,W,U 是线性空间，$f:V\to W$ 和 $g:W\to U$ 是线性映射。则复合 $g\circ f:V\to U$ 也是线性映射。

证明：对任意 $\mathbf{v}_1,\mathbf{v}_2\in V$，由 f,g 的线性性

$$g\Big(f(\mathbf{v}_1+\mathbf{v}_2)\Big)=g\Big(f(\mathbf{v}_1)+f(\mathbf{v}_2)\Big)=g\Big(f(\mathbf{v}_1)\Big)+g\Big(f(\mathbf{v}_2)\Big)$$

因此 $g\circ f$ 保加法。类似可证 $g\circ f$ 保数乘。 □

例子 1.4.9.

$$f:\mathbb{R}^2\to\mathbb{R}^2,\quad f(x,y)=(-y,x)$$

是一个线性映射，表示平面上逆时针旋转 $\frac{\pi}{2}$。f 和它自己的复合

$$f\circ f:\mathbb{R}^2\to\mathbb{R}^2,\quad (x,y)\to(-x,-y)$$

也是线性映射，表示平面上逆时针旋转 π。

命题 1.4.10. 设 $\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$ 是 V 的一组基。则一个线性映射 $f:V\to W$ 完全由 f 在 $\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$ 上的值决定。

证明：假设 $\mathbf{w}_1=f(\mathbf{v}_1),\cdots,\mathbf{w}_n=f(\mathbf{v}_n)$。则对 V 中任意向量 $\mathbf{v}$，它可以唯一表达为线性组合

$$\mathbf{v}=\lambda_1\mathbf{v}_1+\cdots+\lambda_n\mathbf{v}_n.$$

则由 f 的线性性，我们得到

$$f(\mathbf{v})=\lambda_1\mathbf{w}_1+\cdots+\lambda_n\mathbf{w}_n$$

被 $\mathbf{w}_1,\cdots,\mathbf{w}_n$ 完全确定了。 □

例子 1.4.11. $\mathbb{R}^n$ 中的一组基为

$$\begin{aligned}\vec{e}_1&=(1,0,\cdots,0)\\ \vec{e}_2&=(0,1,\cdots,0)\\ &\vdots\\ \vec{e}_n&=(0,0,\cdots,1)\end{aligned}$$

设 $f:\mathbb{R}^n\to\mathbb{R}$ 是线性映射并且 $f(\vec{e}_i)=a_i$，则

$$f(\vec{x})=f(x_1\vec{e}_1+\cdots+x_n\vec{e}_n)=a_1x_1+\cdots+a_nx_n$$

1.4.3 矩阵

下面我们讨论线性映射

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

我们说明这样一个线性映射可以通过矩阵来描述。为了方便，我们将 $\mathbb{R}^n$ 中的元素写成列向量

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

$\mathbb{R}^n$ 的一组标准基记为

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \cdots \quad \vec{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

线性映射 $f : \mathbb{R}^n \to \mathbb{R}^m$ 由一组基的值确定。设

$$f(\vec{e}_1) = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, f(\vec{e}_2) = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} \cdots f(\vec{e}_n) = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

则这些向量 $f(\vec{e}_1), \cdots, f(\vec{e}_n) \in \mathbb{R}^m$ 决定了线性映射 f 在所有 $\mathbb{R}^n$ 上的取值。对于 $\mathbb{R}^n$ 中任意一个向量 $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$，我们有

$$f(\vec{x}) = f(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) \overset{\text{线性性}}{=} x_1 f(\vec{e}_1) + \cdots + x_n f(\vec{e}_n)$$

代入 $f(\vec{e}_i)$ 的值，我们得到

$$f(\vec{x}) = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

我们把 $\mathbb{R}^m$ 中的向量 $f(\vec{e}_1), \cdots, f(\vec{e}_n)$ 排列在一起

$$\begin{bmatrix} f(\vec{e}_1) & f(\vec{e}_2) \cdots & f(\vec{e}_n) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

右边这个表达方式称为一个 $m \times n$ 的矩阵。m 表示这个矩阵的行数，n 表示这个矩阵的列数。

命题 1.4.12. 我们有一一对应

$$\{\text{线性映射 } \mathbb{R}^n \to \mathbb{R}^m\} \iff \{m \times n \text{ 的矩阵}\}$$
$$f : \mathbb{R}^n \to \mathbb{R}^m \iff \begin{bmatrix} f(\vec{e}_1) & f(\vec{e}_2) \cdots & f(\vec{e}_n) \end{bmatrix}$$

由上述讨论，给定线性映射 $f : \mathbb{R}^n \to \mathbb{R}^m$，我们将它对应于矩阵（$A$ 称为 f 的矩阵表示）

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

其中 A 的第 k 列对应于 $f(\vec{e}_k)$，$k = 1, ..., n$.

反之，给定矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

A 的列向量决定了一个线性映射 $f : \mathbb{R}^n \to \mathbb{R}^m$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \overset{f}{\mapsto} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

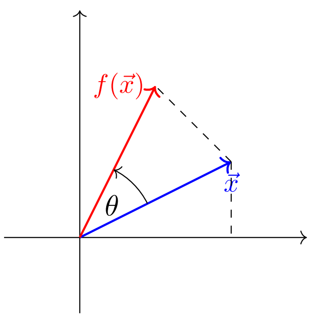
例子 1.4.13. 线性映射 $f : \mathbb{R}^3 \to \mathbb{R}^2$, $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 对应于矩阵

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

这是一个 2 行 3 列的矩阵，即 2×3 矩阵。

例子 1.4.14. $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ 给出一个线性映射 $f : \mathbb{R}^2 \to \mathbb{R}^2$

$$f : \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} \cos\theta\ x_1 - \sin\theta\ x_2 \\ \sin\theta\ x_1 + \cos\theta\ x_2 \end{bmatrix}$$



1.5 习题

1. 判断下列集合是否构成 $\mathbb{R}^3$ 上的线性子空间，并说明理由：

    (a) $V_1 = \{(x, y, z) \in \mathbb{R}^3 |\ x + y + z = 0\}$

    (b) $V_2 = \{(x, y, z) \in \mathbb{R}^3 |\ x + y + z = 1\}$

    (c) $V_3 = \{(x, y, z) \in \mathbb{R}^3 |\ x^2 + y^2 - z^2 = 0\}$

    (d) $V_4 = \{(x, y, z) \in \mathbb{R}^3 |\ x = 2y, z = 3y\}$

    (e) $V_5 = \{(x, y, z) \in \mathbb{R}^3 |\ x \geq 0, y \geq 0\}$

2. 判断下列函数集合是否构成线性空间并说明理由，其中加法和数乘定义为函数的加法和数乘：

    (a) $V = \{f : \mathbb{R} \to \mathbb{R} |\ f(0) = 0\}$

    (b) $V = \{f : \mathbb{R} \to \mathbb{R} |\ f(x) = ax + b,\ a, b \in \mathbb{R}\}$

3. 设向量 $\mathbf{v}_1 = (1, 2, 3)$, $\mathbf{v}_2 = (0, 1, 2)$, $\mathbf{v}_3 = (-1, 0, 1)$。

    (a) 判断向量组 $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ 是否线性相关。若线性相关，找出线性相关的系数。

    (b) 判断 $\mathbf{u}_1 = (1, 5, 9)$ 是否可以表示为 $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ 的线性组合。若可以，找出线性组合的系数。

    (c) 判断 $\mathbf{u}_2 = (4, 7, 11)$ 是否可以表示为 $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ 的线性组合。若可以，找出线性组合的系数。

4. 设 V 是 $\mathbb{R}$ 上所有次数不超过 2 的多项式构成的线性空间。

    (a) 证明向量组 $\{1, x, x^2\}$ 线性无关。

    (b) 向量组 $\{1 + x, 1 - x, x^2\}$ 是否线性相关？说明理由

    (c) 向量组 $\{1, x, x^2, 1 + 2x + 3x^2\}$ 是否线性相关？说明理由

    (d) 设 $p_1(x) = 1 + x$, $p_2(x) = 1 - x$, $p_3(x) = x^2$。多项式 $q(x) = 2x^2 + 3x + 4$ 是否可以表示为 $p_1(x), p_2(x), p_3(x)$ 的线性组合？如果是，请给出具体的线性组合表示。

    (e) 多项式 $r(x) = x^2 - 1$ 是否可以表示为 $p_1(x), p_2(x), p_3(x)$ 的线性组合？如果是，请给出具体的线性组合表示。

5. 证明：若向量组 $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ 线性相关，则向量组 $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n, \mathbf{v}_{n+1}\}$ 也线性相关。

6. 设 V 是 $\mathbb{R}$ 上的线性空间，$\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$。已知 $\mathbf{v}_1, \mathbf{v}_2$ 线性无关，且 $\mathbf{v}_3$ 不能表示为 $\mathbf{v}_1, \mathbf{v}_2$ 的线性组合。证明：$\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 线性无关。

7. 设 V 是 $\mathbb{R}$ 上的线性空间，$\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \in V$。证明：若 $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ 线性无关，且 $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n, \mathbf{u}$ 线性相关，则 $\mathbf{u}$ 可以表示为 $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ 的线性组合。

8. 判断下列向量组是否构成 $\mathbb{R}^4$ 的一组基:

$$\mathbf{v}_1 = (1,0,0,0), \quad \mathbf{v}_2 = (1,1,0,0), \quad \mathbf{v}_3 = (1,1,1,0), \quad \mathbf{v}_4 = (1,1,1,1).$$

9. 设 W 是 $\mathbb{R}^3$ 中所有满足 $x+2y-z=0$ 的向量 (x,y,z) 构成的子空间. 求 W 的一组基和维数.

10. 设 W 是 $\mathbb{R}^4$ 中所有满足 $x_1+x_2-x_3+x_4=0$ 和 $2x_1-x_2+x_3-2x_4=0$ 的向量 (x_1,x_2,x_3,x_4) 构成的子空间. 求 W 的一组基和维数.

11. 设 $M_{m\times n}$ 是所有 $m\times n$ 矩阵构成的向量空间. 求 $M_{m\times n}$ 的一组基和维数.

12. 求下列向量组的秩和一个极大线性无关组:

    (a) $\mathbf{v}_1 = (1,2,3), \quad \mathbf{v}_2 = (2,1,0), \quad \mathbf{v}_3 = (3,3,3).$

    (b) $\mathbf{v}_1 = (1,0,1,0), \quad \mathbf{v}_2 = (0,1,0,1), \quad \mathbf{v}_3 = (1,1,1,1), \quad \mathbf{v}_4 = (2,1,2,1).$

13. 设向量组 $\alpha_1,\alpha_2,\alpha_3$ 线性无关，向量 β 可以由 $\alpha_1,\alpha_2,\alpha_3$ 线性表示，但不能由 α_1,α_2 线性表示. 判断向量组 $\{\alpha_1,\alpha_2,\beta\}$ 是否构成 $\{\alpha_1,\alpha_2,\alpha_3,\beta\}$ 的极大线性无关组，并证明你的结论.

14. 设向量组 $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_s$ 可由向量组 $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_t$ 线性表示. 证明:

$$\operatorname{rank}\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_s\} \le \operatorname{rank}\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_t\}$$

15. 设向量组 $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_s$ 和向量组 $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_t$ 等价. 证明:

$$\operatorname{rank}\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_s\} = \operatorname{rank}\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_t\}$$

16. 求如下线性映射 T 在标准基下的矩阵表示：

    (a) $T:\mathbb{R}^2\to\mathbb{R}^2$ 定义为 $T(x,y)=(x+2y,3x-y)$.

    (b) $T:\mathbb{R}^3\to\mathbb{R}^2$ 定义为 $T(x,y,z)=(x+y,y-z)$.

    (c) $T:\mathbb{R}^2\to\mathbb{R}^3$ 定义为 $T(x,y)=(x,x+y,x-y)$.

    (d) $T:\mathbb{R}^2\to\mathbb{R}^2$ 定义为关于 x 轴的反射.

    (e) $T:\mathbb{R}^2\to\mathbb{R}^2$ 定义为关于直线 $y=x$ 的反射.

    (f) $T:\mathbb{R}^2\to\mathbb{R}^2$ 定义为先绕原点逆时针旋转 $\frac{\pi}{4}$ 角，再关于 x 轴反射.

第二章　矩阵与线性方程组

2.1　矩阵的基本运算

2.1.1　矩阵的加法与数乘

一个 $m \times n$ 矩阵记为

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

第 i 行第 j 列的元素 a_{ij} 称为矩阵 A 的 (i,j) 分量。我们通常也把矩阵 A 用其元素记作

$$A = (a_{ij})$$

定义 2.1.1. 两个 $m \times n$ 矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

的加法 $A + B$ 定义为如下的 $m \times n$ 矩阵

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

如果用分量来表达的话

$$A = (a_{ij}) \qquad B = (b_{ij})$$

则 $A + B$ 为对应的分量相加

$$A + B = (a_{ij} + b_{ij})$$

由命题1.4.12我们知道 $m \times n$ 矩阵一一对应于 $\mathbb{R}^n \to \mathbb{R}^m$ 的线性映射。考虑两个线性映射 $f, g : \mathbb{R}^n \to \mathbb{R}^m$，它们的和 $f + g$ 定义了一个 $\mathbb{R}^n \to \mathbb{R}^m$ 的线性映射

$$(f + g)(\vec{x}) = f(\vec{x}) + g(\vec{x})$$

如果 f 对应于矩阵 A，g 对应于矩阵 B，容易看出

$$f+g \quad \Longleftrightarrow \quad A+B$$

定义 2.1.2. 给定一个数 $\lambda \in k$ 和一个 $m \times n$ 矩阵

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

我们定义它们的数乘 λA 为 $m \times n$ 矩阵

$$\lambda A=\begin{bmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{bmatrix}$$

我们用 $M_{m\times n}(k)$ 来表示所有数域 k 中的 $m \times n$ 矩阵构成的集合（$k=\mathbb{R}$ 或 $\mathbb{C}$）。如果不特别标记数域，$M_{m\times n}$ 指的是实系数的 $m \times n$ 矩阵的集合

$$M_{m\times n}=M_{m\times n}(\mathbb{R})$$

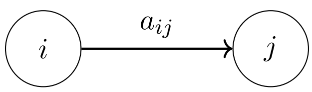
命题 2.1.3. $M_{m\times n}$ 在矩阵的加法和数乘下构成一个线性空间。

通过矩阵的分量，我们可以把矩阵 $A=(a_{ij})$ 一一对应于 mn 个实数 a_{ij}。因此

$$\dim M_{m\times n}=mn$$

2.1.2 矩阵的乘法

矩阵之间可以引进乘法运算。为了说明如何定义矩阵相乘，我们把一个矩阵 $A=(a_{ij})$ 的 (i,j) 分量表示成一个箭头的样子

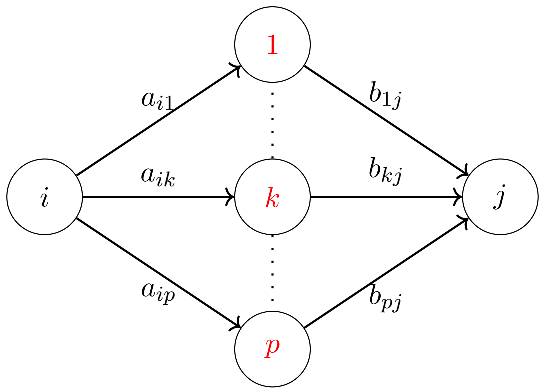


我们把 a_{ij} 想象成从节点 i 到另一个节点 j 的权值。

设 A 是一个 $m \times p$ 矩阵，B 是一个 $p \times n$ 矩阵。

$$A=(a_{ik}), \quad i=1,\cdots,m \quad k=1,\cdots,p.$$
$$B=(b_{kj}), \quad k=1,\cdots,p \quad j=1,\cdots,n.$$

这里 k 的指标都是从 $1,\cdots,p$，我们可以把中间这个节点连接起来



我们考虑从 i 出发，经过中间节点 k，最后到 j 的过程，并对所有的中间节点 k 的位置求和得到一个从 i 到 j 的权值 $c_{ij}=\sum\limits_{k=1}^{p}a_{ik}b_{kj}$。这组新的数字 c_{ij} 定义了一个 $m\times n$ 的矩阵，记作

$$C=(c_{ij})$$

这个新矩阵 C 就是矩阵 A 和 B 的乘积。

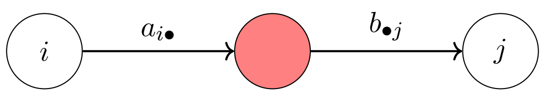
定义 2.1.4. 设 A 是一个 $m\times p$ 矩阵，B 是一个 $p\times n$ 矩阵。它们的乘积 $C=AB$ 定义为 $m\times n$ 矩阵，其中 C 的 (i,j) 分量为

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

这里 a_{ik} 是 A 的 (i,k) 分量，b_{kj} 是 B 的 (k,j) 分量。

我们强调一下，并不是任意两个矩阵都可以相乘。这里两个矩阵 A,B 可以相乘的前提是 A 的列数和 B 的行数一样的，即“首尾相接”。

$$M_{m\times p}\times M_{p\times n}\xrightarrow{\text{相乘}}M_{m\times n}$$



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

这个公式也表明 (AB) 的 (i,j) 分量 c_{ij}，是 A 的第 i 行向量 $\begin{bmatrix}a_{i1} & \cdots & a_{ip}\end{bmatrix}$ 和 B 的第 j 列向量 $\begin{bmatrix}b_{1j}\\ \vdots \\ b_{pj}\end{bmatrix}$ 对应位置的数字乘起来求和

$$a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{ip}b_{pj}.$$

$$
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1p} \\
\cdots & \cdots & \cdots & \cdots \\
a_{i1} & a_{i2} & \cdots & a_{ip} \\
\cdots & \cdots & \cdots & \cdots \\
a_{m1} & a_{m2} & \cdots & a_{mp}
\end{bmatrix}
\begin{bmatrix}
b_{11} & \cdots & b_{1j} & \cdots & b_{1n} \\
b_{21} & \cdots & b_{2j} & \cdots & b_{2n} \\
\vdots & \cdots & \vdots & \cdots & \vdots \\
b_{p1} & \cdots & b_{pj} & \cdots & b_{pn}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\sum_k a_{1k}b_{k1} & \sum_k a_{1k}b_{k2} & \cdots & \sum_k a_{1k}b_{kn} \\
\cdots & \cdots & \cdots & \cdots \\
\sum_k a_{ik}b_{k1} & \cdots & \sum_k a_{ik}b_{kj} & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
\sum_k a_{mk}b_{k1} & \sum_k a_{mk}b_{k2} & \cdots & \sum_k a_{mk}b_{kn}
\end{bmatrix}.
$$

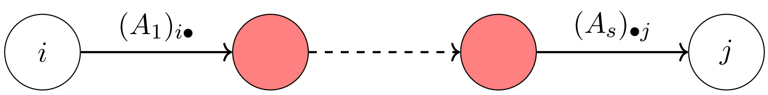
例子 2.1.5.

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

我们同样可以把多个矩阵相乘。设 $A_1, A_2, \cdots, A_s$ 依次是 $m \times p_1, p_1 \times p_2, \cdots, p_{s-1} \times n$ 矩阵。则它们的乘积 $A_1A_2\cdots A_s$ 是一个 $m \times n$ 矩阵，其 (i,j) 分量为

$$
(A_1A_2\cdots A_s)_{ij} = \sum_{k_1=1}^{p_1} \cdots \sum_{k_{s-1}=1}^{p_{s-1}} (A_1)_{ik_1}(A_2)_{k_1k_2}\cdots(A_s)_{k_{s-1}j}
$$

这里 $(A_r)_{kl}$ 是矩阵 A_r 的 (k,l) 分量。我们可以画图表示为



矩阵的乘法满足结合律

$$
(AB)C = A(BC) = ABC,
$$

这里 A 是 $m \times p$ 矩阵，B 是 $p \times q$ 矩阵，C 是 $q \times n$ 矩阵。结合律等价于如下的求和等式

$$
\sum_{l=1}^{q}\sum_{k=1}^{p}(A)_{ik}(B)_{kl}(C)_{lj} = \sum_{k=1}^{p}\sum_{l=1}^{q}(A)_{ik}(B)_{kl}(C)_{lj}
$$

2.1.3 线性映射的复合

我们知道线性映射一一对应于矩阵。下面我们说明矩阵的乘法对应于线性映射的复合，这给出了矩阵乘法定义方式的一个自然的解释。设

$$
f : \mathbb{R}^p \to \mathbb{R}^m, \quad g : \mathbb{R}^n \to \mathbb{R}^p
$$

是两个线性映射。它们的复合得到一个映射

$$
f \circ g : \mathbb{R}^n \to \mathbb{R}^m
$$

$f\circ g$ 依然是一个线性映射：它保加法

$$f(g(\vec{u}+\vec{v}))\overset{g\text{ 线性}}{=}f(g(\vec{u})+g(\vec{v}))\overset{f\text{ 线性}}{=}f(g(\vec{u}))+f(g(\vec{v}))$$

类似可证 $f\circ g$ 保数乘。

命题 2.1.6. 如果线性映射 $f:\mathbb{R}^p\to\mathbb{R}^m$ 对应于 $m\times p$ 矩阵 A，$g:\mathbb{R}^n\to\mathbb{R}^p$ 对应于 $p\times n$ 矩阵 B。则它们的复合 $f\circ g:\mathbb{R}^n\to\mathbb{R}^m$ 对应于 $m\times n$ 矩阵 AB。即矩阵的乘法对应于线性映射的复合。

证明：记 $\mathbb{R}^n$ 的标准基

$$\vec{e}_1=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\quad \vec{e}_2=\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}\quad\cdots\quad \vec{e}_n=\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$$

和 $\mathbb{R}^p$ 的标准基

$$\vec{u}_1=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\quad \vec{u}_2=\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}\quad\cdots\quad \vec{u}_p=\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$$

$g:\mathbb{R}^n\to\mathbb{R}^p$ 对应的 $p\times n$ 矩阵 B 为

$$B=\begin{bmatrix}g(\vec{e}_1) & g(\vec{e}_2) & \cdots & g(\vec{e}_n)\end{bmatrix},$$

用矩阵元可以写成

$$g(\vec{e}_j)=\begin{bmatrix}b_{1j}\\b_{2j}\\\vdots\\b_{pj}\end{bmatrix}=b_{1j}\vec{u}_1+\cdots+b_{pj}\vec{u}_p$$

$f:\mathbb{R}^p\to\mathbb{R}^m$ 对应的 $m\times p$ 矩阵 A 为

$$A=\begin{bmatrix}f(\vec{u}_1) & f(\vec{u}_2)\cdots & f(\vec{u}_p)\end{bmatrix},$$

用矩阵元可以写成

$$f(\vec{u}_k)=\begin{bmatrix}a_{1k}\\a_{2k}\\\vdots\\a_{mk}\end{bmatrix}$$

我们考虑复合 $f\circ g:\mathbb{R}^n\to\mathbb{R}^m$，其对应于矩阵

$$\begin{bmatrix}f(g(\vec{e}_1)) & f(g(\vec{e}_2)) & \cdots & f(g(\vec{e}_n))\end{bmatrix}$$

它的第 j 列为

$$\begin{aligned}
f(g(\vec{e}_j)) =& f(b_{1j}\vec{u}_1 + \cdots + b_{pj}\vec{u}_p) = b_{1j}f(\vec{u}_1) + \cdots + b_{pj}f(\vec{u}_p) \\
=& b_{1j}\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + b_{pj}\begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{mp} \end{bmatrix} = \begin{bmatrix} \sum\limits_{k=1}^{p} a_{1k}b_{kj} \\ \sum\limits_{k=1}^{p} a_{2k}b_{kj} \\ \vdots \\ \sum\limits_{k=1}^{p} a_{mk}b_{kj} \end{bmatrix}.
\end{aligned}$$

因此 $f \circ g$ 对应的矩阵

$$\begin{bmatrix} f(g(\vec{e}_1)) & f(g(\vec{e}_2)) & \cdots & f(g(\vec{e}_n)) \end{bmatrix} = \begin{bmatrix} \sum\limits_{k} a_{1k}b_{k1} & \sum\limits_{k} a_{1k}b_{k2} & \cdots & \sum\limits_{k} a_{1k}b_{kn} \\ \cdots & \cdots & \cdots & \cdots \\ \sum\limits_{k} a_{ik}b_{k1} & \cdots & \sum\limits_{k} a_{ik}b_{kj} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \sum\limits_{k} a_{mk}b_{k1} & \sum\limits_{k} a_{mk}b_{k2} & \cdots & \sum\limits_{k} a_{mk}b_{kn} \end{bmatrix}$$

即为矩阵 A 和 B 的乘积 AB。 $\square$

我们可以对比一下矩阵乘积的表示法

$$\begin{tikzcd} i \arrow[r, "a_{i\bullet}"] & {} \arrow[r, "b_{\bullet j}"] & j \end{tikzcd}$$

和映射复合的表示法

$$\begin{tikzcd} \mathbb{R}^m & \mathbb{R}^p \arrow[l, "f"'] & \mathbb{R}^n \arrow[l, "g"'] \end{tikzcd}$$

矩阵乘法的结合律对应于映射复合的结合律。

我们可以用矩阵把线性映射写成显式的形式。设 $f : \mathbb{R}^n \to \mathbb{R}^m$ 是线性映射，对应于 $m \times n$ 矩阵 A

$$A = \begin{bmatrix} f(\vec{e}_1) & f(\vec{e}_2) \cdots & f(\vec{e}_n) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

对于 $\mathbb{R}^n$ 中任一向量 $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$

$$
\begin{aligned}
f(\vec{x}) &= x_1 f(\vec{e}_1) + \cdots + x_n f(\vec{e}_n) \\
&= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}
\end{aligned}
$$

因此写成向量形式，线性映射 $f : \mathbb{R}^n \to \mathbb{R}^m$

$$
f : \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}
$$

恰好是把 n 维列向量（即 $n \times 1$ 矩阵）乘以 $m \times n$ 矩阵 A，得到 m 维列向量（即 $m \times 1$ 矩阵）。

总结一下，线性映射和矩阵的关系可以用矩阵乘法表达为

$$
f : \mathbb{R}^n \to \mathbb{R}^m
$$

$$
f(\vec{x}) = A\vec{x}, \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}
$$

2.1.4 分块矩阵

分块矩阵是指将矩阵按照某种方式划分成若干个小矩阵（块）。分块矩阵是将大矩阵分解为较小矩阵的一种有效方法，有助于简化计算和表示结构化数据。

一个 $m \times n$ 矩阵可以分块表示为

$$
A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}
$$

其中 A_{ij} 是块矩阵的元素，通常是较小的矩阵。例如，A_{ij} 可以是 $p_i \times q_j$ 矩阵，其中 p_i 和 q_j 是适当选择的维度。

假设我们有两个矩阵 A 和 B，它们分别分块表达为：

$$
A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}
$$

其中 A_{ij} 和 B_{ij} 的维度需要满足乘法的条件。乘积 $C = AB$ 也会是一个分块矩阵：

$$
C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}
$$

其中每个块的计算可以通过类似的矩阵乘法公式得到：

$$C_{ij} = \sum_k A_{ik} B_{kj}$$

具体而言，我们有

$$C = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$$

注记. 注意到 A_{ij}, B_{ij} 都是一些矩阵块而不再是数字，公式里每个矩阵块的乘积表达顺序是重要的。这里我们需要按照乘积顺序把 A 的矩阵块写在前面，B 的矩阵块写在后面。例如 C_{11} 的矩阵元是 $A_{11}B_{11} + A_{12}B_{21}$，而不能写成 $B_{11}A_{11} + B_{21}A_{12}$。类似地矩阵乘法公式对于多个分块也是成立的，具体细节留给读者证明。

例子 2.1.7. 考虑把如下矩阵进行分块：

$$A = \begin{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} & \begin{bmatrix} 5 \\ 6 \end{bmatrix} \\ \begin{bmatrix} 7 & 8 \end{bmatrix} & 9 \end{bmatrix} \qquad B = \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ \begin{bmatrix} 4 & 5 \end{bmatrix} & 6 \end{bmatrix}$$

按照如上分块计算矩阵乘积 $C = AB$

$$\begin{aligned} C_{11} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 5 \\ 6 \end{bmatrix} \begin{bmatrix} 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 20 & 25 \\ 24 & 30 \end{bmatrix} = \begin{bmatrix} 21 & 27 \\ 27 & 34 \end{bmatrix} \\ C_{12} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 6 \end{bmatrix} 6 = \begin{bmatrix} 8 \\ 18 \end{bmatrix} + \begin{bmatrix} 30 \\ 36 \end{bmatrix} = \begin{bmatrix} 38 \\ 54 \end{bmatrix} \\ C_{21} &= \begin{bmatrix} 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 9 \begin{bmatrix} 4 & 5 \end{bmatrix} = \begin{bmatrix} 7 & 8 \end{bmatrix} + \begin{bmatrix} 36 & 45 \end{bmatrix} = \begin{bmatrix} 43 & 53 \end{bmatrix} \\ C_{22} &= \begin{bmatrix} 7 & 8 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} + 9 * 6 = 38 + 54 = 92 \end{aligned}$$

因此我们得到矩阵 AB 的分块表达

$$AB = \begin{bmatrix} \begin{bmatrix} 21 & 27 \\ 27 & 34 \end{bmatrix} & \begin{bmatrix} 38 \\ 54 \end{bmatrix} \\ \begin{bmatrix} 43 & 53 \end{bmatrix} & 92 \end{bmatrix}$$

例子 2.1.8. 设 A 是 $m \times p$ 矩阵，B 是 $p \times n$ 矩阵。它们的乘积 AB 是 $m \times n$ 矩阵。我们可以通过分块来得到乘积 AB 的行和列的表达。

首先我们把 B 分块成列向量的样子

$$B = \begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix}$$

这里 $\vec{\beta}_j$ 是 $\mathbb{R}^p$ 中的列向量。则作为分块矩阵，我们得到如下公式

$$AB = A \begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix} = \begin{bmatrix} A\vec{\beta}_1 & A\vec{\beta}_2 & \cdots & A\vec{\beta}_n \end{bmatrix}$$

即 AB 的第 j 列的列向量为 $A\vec{\beta}_j$。

类似地，我们把 A 分块成行向量的样子

$$A = \begin{bmatrix} \vec{\alpha}_1 \\ \vec{\alpha}_1 \\ \vdots \\ \vec{\alpha}_m \end{bmatrix}$$

这里 $\vec{\alpha}_i$ 是 $\mathbb{R}^p$ 中的行向量。则作为分块矩阵，我们得到如下公式

$$AB = \begin{bmatrix} \vec{\alpha}_1 \\ \vec{\alpha}_2 \\ \vdots \\ \vec{\alpha}_m \end{bmatrix} B = \begin{bmatrix} \vec{\alpha}_1 B \\ \vec{\alpha}_2 B \\ \vdots \\ \vec{\alpha}_m B \end{bmatrix}$$

即 AB 的第 i 行的行向量为 $\vec{\alpha}_i B$。

上面两个公式也可以直接通过矩阵乘法的定义来验证。

2.2 线性方程组

考虑 n 元的线性方程组（这里 $x_1, \cdots, x_n$ 是未知变量）

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

这是最基本的一类代数方程组，有广泛的应用。我们在本节讨论线性方程组的求解方法。

2.2.1 消元法与初等变换

我们首先看一个具体例子。考虑以下线性方程组：

$$x + 2y + z = 6 \tag{1}$$
$$2x + 5y + 3z = 15 \tag{2}$$
$$x + 4y + 6z = 21 \tag{3}$$

将 (1) 的 -2 倍加到 (2)，(1) 的 -1 倍加到 (3)，得到等价的方程组

$$x + 2y + z = 6 \tag{1}$$
$$y + z = 3 \tag{2'}$$
$$2y + 5z = 15 \tag{3'}$$

将 (2') 的 -2 倍加到 (3')，得到

$$\begin{aligned} x+2y+z&=6 && (1)\\ y+z&=3 && (2')\\ 3z&=9 && (3'')\end{aligned}$$

由 (3") 得到 $z=3$。将 $z=3$ 代入 (2')，得到 $y+3=3$，故 $y=0$。将 $y=0$ 和 $z=3$ 代入 (1)，得到 $x+0+3=6$，故 $x=3$。方程组的解为

$$x=3,\quad y=0,\quad z=3.$$

总结如上过程，消元法给出线性方程组的三种初等变换:

1. 交换两行: 将方程组中的两个方程交换位置。
2. 将方程组中的一个方程乘以一个非零常数。
3. 将一个方程乘以一个数加到另一个方程。

这些初等变换不改变方程组的解，可用来简化方程，是求解线性方程组的基本方法。具体而言

$$\begin{aligned} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n&=b_1 && (r_1)\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n&=b_2 && (r_2)\\ &\vdots\\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n&=b_m && (r_m)\end{aligned}$$

三种初等变换可以表示为:

1. 交换第 i 行和第 j 行: $(r_i)\leftrightarrow(r_j)$
2. 第 i 行乘以非零常数 λ: $(r_i)\to\lambda(r_i)$
3. 第 i 行的 λ 倍加到第 j 行: $(r_j)\to(r_j)+\lambda(r_i)$

例子 2.2.1.

$$\begin{aligned} x+2y&=4\\ 2x+3y&=7\end{aligned}$$

我们可以进行以下初等变换：将第一行的 -2 倍加到第二行。得到等价的新的方程组：

$$\begin{aligned} x+2y&=4\\ -y&=-1\end{aligned}$$

由此解出 $y=1,x=2$。

例子 2.2.2. （谷神星的发现和轨道预测）

1801 年，意大利天文学家皮亚齐发现了小行星谷神星（*Ceres*）。然而，在观测了 *40* 天后，谷神星运行到太阳背后，从人们的视野中消失了，要到年底才可能再次见到谷神星。由于观测数据有限，天文学家们无法确定谷神星的轨道，从而无法预测它再次出现的位置。

高斯当时年仅 *24* 岁，利用开普勒定律和皮亚齐的观测数据，将谷神星的轨道计算转化和近似为一个线性方程组，使用消元法求解该方程组，预测出了谷神星的轨道参数。高斯的预测非常精确，谷神星在一年后被重新发现，位置与高斯的预测非常接近。这一事件使得高斯声名鹊起，也大大推动了线性方程组和线性代数的理论和方法。

2.2.2 线性方程组的矩阵形式

线性方程组可以写成矩阵的形式

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

这里

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

我们也可以把整个线性方程组的信息可以写成一个 $m \times (n+1)$ 矩阵

$$\overline{A} = \begin{bmatrix} A & \vec{b} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

A 称为系数矩阵，$\overline{A}$ 称为增广矩阵。

线性方程组的初等变换，等价于对 $\overline{A}$ 作行变换。例如把第一个方程乘以 λ 加到第二个方程

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix} \Longrightarrow \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21}+\lambda a_{11} & a_{22}+\lambda a_{12} & \cdots & a_{2n}+\lambda a_{1n} & b_2+\lambda b_1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

对应于线性方程组的初等变换，我们定义矩阵的三种初等行变换为

1. 交换两行: 将矩阵的两行交换位置。

2. 将矩阵中的一行乘以一个非零常数。

3. 将矩阵的一行乘以一个数加到另一行。

从而利用消去法解线性方程组的过程，可以等价地用对增广矩阵作初等行变换来表示。

命题 2.2.3. 通过初等行变换可以把增广矩阵变为如下阶梯形

$$\begin{bmatrix} a'_{1p_1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & a'_{1n} & b'_1 \\ & & a'_{2p_2} & \cdots & \cdots & \cdots & \cdots & a'_{2n} & b'_2 \\ & & & & \cdots & \cdots & \cdots & \vdots & \vdots \\ & & & & & a'_{rp_r} & \cdots & a'_{rn} & b'_r \\ & & & & & & & & b'_{r+1} \\ & & & & & & & & \vdots \\ & & & & & & & & b'_m \end{bmatrix}$$

其中空白处的矩阵元都是 0。

证明方法和用消去法解方程的过程是一致的。

例子 2.2.4. 考虑如下的线性方程组，我们写成矩阵的形式

$$\begin{cases} x+2y+z=6 \\ 2x+4y+5z=15 \\ x+4y+6z=21 \end{cases} \iff \begin{bmatrix} 1 & 2 & 1 & 6 \\ 2 & 4 & 5 & 15 \\ 1 & 4 & 6 & 21 \end{bmatrix}$$

我们可以通过如下的初等行变换变为阶梯形

$$\begin{bmatrix} 1 & 2 & 1 & 6 \\ 2 & 4 & 5 & 15 \\ 1 & 4 & 6 & 21 \end{bmatrix} \xRightarrow[\text{加到第 2 行}]{\text{第 1 行 }\times(-2)} \begin{bmatrix} 1 & 2 & 1 & 6 \\ 0 & 0 & 3 & 3 \\ 1 & 4 & 6 & 21 \end{bmatrix} \xRightarrow[\text{加到第 3 行}]{\text{第 1 行 }\times(-1)} \begin{bmatrix} 1 & 2 & 1 & 6 \\ 0 & 0 & 3 & 3 \\ 0 & 2 & 5 & 15 \end{bmatrix} \xRightarrow[\text{第 2,3 行}]{\text{交换}} \begin{bmatrix} 1 & 2 & 1 & 6 \\ 0 & 2 & 5 & 15 \\ 0 & 0 & 3 & 3 \end{bmatrix}$$

假如我们已经将一个线性方程组的增广矩阵化成了如下的阶梯形矩阵

$$\begin{bmatrix} a'_{1p_1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & a'_{1n} & b'_1 \\ & & a'_{2p_2} & \cdots & \cdots & \cdots & \cdots & a'_{2n} & b'_2 \\ & & & & \cdots & \cdots & \cdots & \vdots & \vdots \\ & & & & & a'_{rp_r} & \cdots & a'_{rn} & b'_r \\ & & & & & & & & b'_{r+1} \\ & & & & & & & & \vdots \\ & & & & & & & & b'_m \end{bmatrix}$$

那么阶梯形矩阵最后几行对应的方程为

$$\begin{cases} 0 = b'_{r+1} \\ \cdots \\ 0 = b'_m \end{cases}$$

如果其中 $b'_{r+1}, \cdots, b'_m$ 有非零元，则原方程无解。

例子 2.2.5. 考虑如下的线性方程组，将其写成增广矩阵的形式

$$\begin{cases} x+y+z=3 \\ 2x+3y+5z=4 \\ x+2y+4z=2 \end{cases} \iff \begin{bmatrix} 1 & 1 & 1 & 3 \\ 2 & 3 & 5 & 4 \\ 1 & 2 & 4 & 2 \end{bmatrix}$$

通过初等行变换

$$\begin{bmatrix} 1 & 1 & 1 & 3 \\ 2 & 3 & 5 & 4 \\ 1 & 2 & 4 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 3 & -2 \\ 1 & 2 & 4 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 3 & -2 \\ 0 & 1 & 3 & -1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

因此原线性方程组无解。

命题 2.2.6. 线性方程组有解的充要条件是增广矩阵通过初等行变换化成阶梯形矩阵后

$$b'_{r+1} = \cdots = b'_n = 0.$$

证明：我们上面已经说明了必要性。下面我们证明充分性。假设 $b'_{r+1} = \cdots = b'_n = 0$，则原线性方程组等价于如下的阶梯形方程

$$\begin{bmatrix} & a'_{1p_1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & a'_{1n} & b'_1 \\ & & & a'_{2p_2} & \cdots & \cdots & \cdots & \cdots & a'_{2n} & b'_2 \\ & & & & & \cdots & \cdots & \cdots & \vdots & \vdots \\ & & & & & & a'_{rp_r} & \cdots & a'_{rn} & b'_r \\ & & & & & & & & & \end{bmatrix}$$

这里 $a'_{1p_1}, \cdots, a'_{rp_r}$ 均不为 0。从最后一个方程

$$a'_{rp_r} x_{p_r} + \cdots = b'_r$$

可以解出 x_{p_r}，然后代入前面方程解出 $x_{p_{r-1}}$。依次类推可以找到方程的解 (解可能不唯一)。 □

例子 2.2.7. 考虑如下的线性方程组，将其写成矩阵的形式

$$\begin{cases} x+y+z=3 \\ 2x+3y+5z=4 \\ x+2y+4z=1 \end{cases} \iff \begin{bmatrix} 1 & 1 & 1 & 3 \\ 2 & 3 & 5 & 4 \\ 1 & 2 & 4 & 1 \end{bmatrix}$$

做初等行变换

$$\begin{bmatrix} 1 & 1 & 1 & 3 \\ 2 & 3 & 5 & 4 \\ 1 & 2 & 4 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 3 & -2 \\ 1 & 2 & 4 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 3 & -2 \\ 0 & 1 & 3 & -2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

最后这个增广矩阵对应于方程

$$\begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \iff \begin{cases} x+y+z=3 \\ \quad y+3z=-2 \end{cases}$$

由此可以解出 $\begin{cases} x = 2c + 5 \\ y = -2 - 3c \\ z = c \end{cases}$ 这里 c 是任意常数。

总结上述讨论可知：

1. 线性方程组可以通过初等变换化简。
2. 线性方程组可能没有解。
3. 线性方程组如果有解，解可能不唯一。

2.2.3 齐次线性方程组

如下方程称为齐次线性方程组

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

即对应于 $b_1 = \cdots = b_m = 0$ 的情况。矩阵表达为

$$A\vec{x} = 0.$$

之前 b_i 不全为 0 的情况称为非齐次线性方程组。

注意到齐次线性方程组总是有解的：

$$x_1 = \cdots x_n = 0$$

是一个解，称为零解或者平凡解。一个值得关心的问题是，齐次线性方程组是否有非平凡解?

这个问题和线性方程组解的唯一性密切相关。设 $\vec{x}_0, \vec{x}_1$ 是非齐次线性方程 $A\vec{x} = \vec{b}$ 的解，即

$$A\vec{x}_0 = \vec{b} \qquad A\vec{x}_1 = \vec{b}$$

两式相减我们得到 $A(\vec{x}_1 - \vec{x}_0) = 0$，即 $\vec{x}_1 - \vec{x}_0$ 是齐次线性方程组的解。

反之，如果 $\vec{u}$ 是齐次线性方程组的解 $A\vec{u} = 0$，$\vec{x}_0$ 是非齐次线性方程组的解 $A\vec{x}_0 = \vec{b}$。则

$$A(\vec{x}_0 + \vec{u}) = A\vec{x}_0 + A\vec{u} = \vec{b} + 0 = \vec{b}$$

即 $\vec{x}_0 + \vec{u}$ 也是非齐次线性方程组的解。

因此非齐次线性方程组 $A\vec{x} = \vec{b}$ 的一般解是

$$\vec{x} = \vec{x}_0 + \vec{u}$$

这里 $\vec{x}_0$ 是非齐次方程组的一个特解（$A\vec{x}_0 = \vec{b}$），$\vec{u}$ 是齐次方程组的一般解（$A\vec{u} = 0$）。即

$$\text{非齐次通解} = \text{非齐次特解} + \text{齐次通解}$$

当然如果特解不存在，则非齐次方程组没有解。

推论 2.2.8. 如果齐次线性方程组只有零解，则非齐次线性方程组的解是唯一的（如果解存在）。

2.3 线性方程组的解空间

我们这一节将系统讨论线性方程组的解空间的结构。

2.3.1 齐次线性方程组的解空间

考虑齐次线性方程组

$$A\vec{x}=0$$

这里 A 是 $m\times n$ 矩阵，$\vec{x}\in\mathbb{R}^n$，即

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \vec{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

该齐次线性方程组的所有解构成 $\mathbb{R}^n$ 的子集，记为

$$K:=\{\vec{x}\in\mathbb{R}^n|A\vec{x}=0\}.$$

命题 2.3.1. K 是 $\mathbb{R}^n$ 的线性子空间。

证明：假设 $\vec{x}_1,\vec{x}_2\in K,\lambda\in\mathbb{R}$，则

$$A(\vec{x}_1+\vec{x}_2)=A\vec{x}_1+A\vec{x}_2=0 \quad A(\lambda\vec{x}_1)=\lambda A\vec{x}_1=0$$

即 $\vec{x}_1+\vec{x}_2\in K,\lambda\vec{x}_1\in K$。故 K 保持加法和数乘。 □

设 $\vec{u}_1,\cdots,\vec{u}_s$ 是 K 的一组基，$s=\dim K$。则 K 中的元素 $\vec{x}$ 可以唯一地写成线性组合

$$\vec{x}=c_1\vec{u}_1+\cdots+c_s\vec{u}_s, \quad c_i\in\mathbb{R}$$

这构成了齐次线性方程组 $A\vec{x}=0$ 的通解，这里 c_i 是任意常数。

例子 2.3.2. 考虑以下齐次线性方程组：

$$\begin{aligned} x_1+2x_2-x_3+x_4-x_5&=0 \\ 2x_1+4x_2-2x_3+3x_4-x_5&=0 \\ -x_1-2x_2+x_3-2x_4+2x_5&=0 \end{aligned}$$

对应的增广矩阵的形式为

$$\begin{bmatrix} 1 & 2 & -1 & 1 & -1 & 0 \\ 2 & 4 & -2 & 3 & -1 & 0 \\ -1 & -2 & 1 & -2 & 2 & 0 \end{bmatrix}$$

我们用初等变换来解方程

$$\begin{bmatrix} 1 & 2 & -1 & 1 & -1 & 0 \\ 2 & 4 & -2 & 3 & -1 & 0 \\ -1 & -2 & 1 & -2 & 2 & 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 2 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \end{bmatrix} \Longrightarrow \begin{bmatrix} 1 & 2 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \end{bmatrix}$$

得到等价的方程组

$$\begin{aligned} x_1 + 2x_2 - x_3 + x_4 - x_5 &= 0 \\ x_4 + x_5 &= 0 \\ 2x_5 &= 0 \end{aligned}$$

由此可以解得

$$x_5 = 0, \quad x_4 = 0, \quad x_1 = x_3 - 2x_2$$

其中 x_2 和 x_3 可以取任意数。设 $x_2 = c_1, x_3 = c_2$，则通解可以写成

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} c_2 - 2c_1 \\ c_1 \\ c_2 \\ 0 \\ 0 \end{bmatrix} = c_1 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad c_1, c_2 \in \mathbb{R}$$

从而该齐次线性方程组的解空间是 2 维的，一组基为 $\vec{u}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$。

2.3.2 矩阵的秩与解空间维数

矩阵的行秩

我们把一个 $m \times n$ 矩阵 A 写成

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \\ \vec{v}_m \end{bmatrix}$$

这里 $\vec{v}_i$ 是 $\mathbb{R}^n$ 中的行向量

$$\vec{v}_i = (a_{i1}\ a_{i2}\ \cdots a_{in}) \in \mathbb{R}^n$$

定义 2.3.3. 矩阵 A 的行秩定义为 A 的行向量张成的线性空间的维数，即

$$A\text{的行秩} = \dim \operatorname{Span}\{\vec{v}_1, \cdots, \vec{v}_m\}$$

命题 2.3.4. 对矩阵作初等行变换得到的行向量与原行向量是等价的。特别的，初等行变换不改变矩阵的行秩。

为了说明这个命题，我们回顾矩阵的三种初等行变换

1. 交换两行: 将矩阵的两行交换位置。
2. 将矩阵中的一行乘以一个非零常数。
3. 将矩阵的一行乘以一个数加到另一行。

例如交换两行，比如第 1 行和第 2 行，容易看出线性等价

$$\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m\} \Longleftrightarrow \{\vec{v}_2, \vec{v}_1, \cdots, \vec{v}_m\}$$

例如将第 2 行乘以一个非零常数 λ，我们有线性等价

$$\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m\} \Longleftrightarrow \{\vec{v}_1, \lambda\vec{v}_2, \cdots, \vec{v}_m\}$$

例如将矩阵的第 1 行乘以一个数加到第 2 行，容易证明如下的线性等价

$$\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m\} \Longleftrightarrow \{\vec{v}_1, \vec{v}_2 + \lambda\vec{v}_1, \cdots, \vec{v}_m\}$$

因此矩阵的行秩在初等行变换下是不变的。

由命题2.2.3，我们总是可以通过初等行变化将 A 变为阶梯形

$$A' = \begin{bmatrix} & a'_{1p_1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & a'_{1n} \\ & & & a'_{2p_2} & \cdots & \cdots & \cdots & \cdots & a'_{2n} \\ & & & & & \cdots & \cdots & \cdots & \vdots \\ & & & & & & a'_{rp_r} & \cdots & a'_{rn} \\ & & & & & & & & \end{bmatrix} = \begin{bmatrix} \vec{v}'_1 \\ \vec{v}'_2 \\ \vdots \\ \vec{v}'_r \\ 0 \end{bmatrix}$$

这里 $a'_{1p_1}, \cdots, a'_{rp_r}$ 均不为 0。观察到行向量 $\vec{v}'_1, \cdots, \vec{v}'_r$ 是线性无关的。实际上，如果有

$$\lambda_1\vec{v}'_1 + \lambda_2\vec{v}'_2 + \cdots + \lambda_r\vec{v}'_r = 0$$

则比较第 p_1 列系数，我们得到 $\lambda_1 a'_{1p_1} = 0 \Rightarrow \lambda_1 = 0$。然后可以依此得到 $\lambda_2 = \cdots = \lambda_r = 0$。

因此 A' 的行秩是 r。由此我们证明了

$$A \text{ 的行秩 } = A' \text{ 的行秩 } = r$$

这给出了用初等行变换计算行秩的方法。

另一方面，矩阵 A' 对应的齐次线性方程组为

$$\begin{aligned} a'_{1p_1}x_{p_1} + \cdots + \cdots + \cdots + a'_{1p_n}x_n &= 0 \\ &\vdots \\ a'_{1p_r}x_{p_r} + \cdots + a'_{1p_n}x_n &= 0 \end{aligned}$$

这里 $p_1 < p_2 < \cdots < p_r$。因此 $x_{p_1}, \cdots, x_{p_r}$ 可以通过其他变量表达出来，而其他变量可以取任意值。总共有 $n-r$ 个自由变量，因此解空间的维数是 $n-r$。

总结如上，对于 $m \times n$ 矩阵 A，通过初等行变换变成阶梯形矩阵 A'，可以看出

$$A \text{ 的行秩 } = r\text{，且 } \{\vec{x} | A\vec{x} = 0\} \text{ 解空间的维数 } = n - r\text{。}$$

命题 2.3.5. 矩阵 A 对应的齐次线性方程组解空间的维数为

$$A \text{ 的列数（即未知变量数）} - A \text{ 的行秩。}$$

矩阵的列秩

我们把一个 $m \times n$ 矩阵 A 写成

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix}$$

这里 $\vec{u}_i$ 是 $\mathbb{R}^m$ 中的列向量 $\vec{u}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix} \in \mathbb{R}^m$。

定义 2.3.6. 矩阵 A 的列秩定义为 A 的列向量张成的线性空间的维数，即

$$A\text{的列秩} = \dim \operatorname{Span}\{\vec{u}_1, \cdots, \vec{u}_n\}$$

命题 2.3.7. 对矩阵作初等行变换亦不改变矩阵的列秩。

需要注意的是，初等行变换不改变行向量张成的空间，但是会改变列向量张成的空间（但是维数不变）。这个命题成立是因为初等行变换不会改变列向量的线性相关性。具体而言，考虑

$$A = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix} \stackrel{\text{初等行变换}}{\Longrightarrow} A' = \begin{bmatrix} \vec{u}'_1 & \vec{u}'_2 & \cdots & \vec{u}'_n \end{bmatrix}$$

由于线性方程组在行变换下是等价的，我们知道对任意的向量子集 $\{\vec{u}_{i_1}, \vec{u}_{i_2}, \cdots, \vec{u}_{i_s}\}$，矩阵 $\begin{bmatrix} \vec{u}_{i_1} & \vec{u}_{i_2} & \cdots & \vec{u}_{i_s} \end{bmatrix}$ 对应的齐次线性方程组

$$\begin{bmatrix} \vec{u}_{i_1} & \vec{u}_{i_2} & \cdots & \vec{u}_{i_s} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_s \end{bmatrix} = 0$$

和矩阵 $\begin{bmatrix} \vec{u}'_{i_1} & \vec{u}'_{i_2} & \cdots & \vec{u}'_{i_s} \end{bmatrix}$ 对应的齐次线性方程组

$$\begin{bmatrix} \vec{u}'_{i_1} & \vec{u}'_{i_2} & \cdots & \vec{u}'_{i_s} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_s \end{bmatrix} = 0$$

是等价的方程，即

$$\lambda_1\vec{u}_{i_1}+\cdots\lambda_s\vec{u}_{i_s}=0 \quad \text{当且仅当} \quad \lambda_1\vec{u}'_{i_1}+\cdots\lambda_s\vec{u}'_{i_s}=0$$

由此可以证明:如果列向量组 $\{\vec{u}_{k_1},\vec{u}_{k_2},\cdots,\vec{u}_{k_r}\}$ 是 A 的列向量的极大线性无关组,则 $\{\vec{u}'_{k_1},\vec{u}'_{k_2},\cdots,\vec{u}'_{k_r}\}$ 是 A' 的列向量的极大线性无关组。因此 A 和 A' 的列秩相等。证明细节留作练习。

我们通过初等行变化将 A 变为阶梯型

$$A'=\begin{bmatrix} a'_{1p_1} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & a'_{1n} \\ & & a'_{2p_2} & \cdots & \cdots & \cdots & \cdots & a'_{2n} \\ & & & & \cdots & \cdots & \cdots & \vdots \\ & & & & & a'_{rp_r} & \cdots & a'_{rn} \\ & & & & & & & \end{bmatrix}$$

这里 $a'_{1p_1},\cdots,a'_{rp_r}$ 均不为 0。很容易看出列向量

$$\vec{u}'_{p_1}=\begin{bmatrix} a'_{1p_1} \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \vec{u}'_{p_2}=\begin{bmatrix} a'_{1p_2} \\ a'_{2p_2} \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \cdots \quad \vec{u}'_{p_r}=\begin{bmatrix} a'_{1p_r} \\ a'_{2p_r} \\ \vdots \\ a'_{rp_r} \\ \vdots \\ 0 \end{bmatrix}$$

构成 A' 列向量的极大线性无关组，因此 A' 的列秩是 r。

因此通过初等行变换将 A 变成阶梯型，我们证明了如下结论

$$\text{行秩}=r=\text{列秩}$$

命题 2.3.8. 矩阵的行秩等于矩阵的列秩。

因此我们把这个行秩/列秩称为矩阵的秩，记为

$$\operatorname{rank}A.$$

我们把上述关于齐次线性方程组的解总结如下：

定理 2.3.9 (齐次线性方程组解的结构定理)**.** 设 A 是 $m\times n$ 矩阵，$\operatorname{rank}A=r$。

- 如果 $r=n$，则 $A\vec{x}=0$ 只有零解
- 如果 $r<n$，则 $A\vec{x}=0$ 具有非零解，且它的通解具有 $n-r$ 个独立参数

2.3.3 非齐次线性方程组的解空间

下面我们考虑非齐次线性方程组

$$A\vec{x}=\vec{b}$$

的解空间。这里

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \vec{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \vec{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

由于

$$\text{非齐次通解} = \text{非齐次特解} + \text{齐次通解}$$

由命题2.3.5，我们已经知道了齐次方程组的通解结构（解空间的维数即自由参数的个数为 $n-\operatorname{rank} A$），因此我们只需要讨论能否找到非齐次线性方程组的一个特解，即解的存在性问题。

我们把矩阵 A 写成列向量的样子

$$A=\begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix}$$

则线性方程组可以写成

$$x_1\vec{u}_1+\cdots+x_n\vec{u}_n=\vec{b}$$

如果线性方程组有解，即存在 $x_1,\cdots,x_n$ 使得

$$x_1\vec{u}_1+\cdots+x_n\vec{u}_n=\vec{b}$$

即 $\vec{b}$ 可以写成 $\{\vec{u}_1,\cdots,\vec{u}_n\}$ 的线性组合。则

$$\{\vec{u}_1,\cdots,\vec{u}_n\} \overset{\text{等价}}{\Longleftrightarrow} \{\vec{u}_1,\cdots,\vec{u}_n,\vec{b}\}$$

反之，如果

$$\{\vec{u}_1,\cdots,\vec{u}_n\} \overset{\text{等价}}{\Longleftrightarrow} \{\vec{u}_1,\cdots,\vec{u}_n,\vec{b}\}$$

则 $\vec{b}$ 可以通过 $\{\vec{u}_1,\cdots,\vec{u}_n\}$ 线性表达，即方程有解。

因此线性方程组有解当且仅当

$$\{\vec{u}_1,\cdots,\vec{u}_n\} \overset{\text{等价}}{\Longleftrightarrow} \{\vec{u}_1,\cdots,\vec{u}_n,\vec{b}\}$$

即这两组向量张成的线性空间的维数是一样的。由此我们证明了如下结论

命题 2.3.10. *非齐次线性方程组 $A\vec{x}=\vec{b}$ 有解当且仅当 A 的列秩和增广矩阵 $\overline{A}=[A\ \vec{b}]$ 的列秩相等，即*

$$\operatorname{rank} A=\operatorname{rank} \overline{A}$$

我们把上述关于非齐次线性方程组的解总结如下：

定理 2.3.11 (非齐次线性方程组解的结构定理)**.** 设 A 是 $m\times n$ 矩阵，$\operatorname{rank} A=r$。考虑非齐次线性方程组 $A\vec{x}=\vec{b}$，设 $\overline{A}$ 是增广矩阵。

- $A\vec{x}=\vec{b}$ 有解当且仅当 $\operatorname{rank}\overline{A}=r$
- 假设 $\operatorname{rank}\overline{A}=r$
  - 如果 $r=n$，则 $A\vec{x}=\vec{b}$ 有唯一解
  - 如果 $r<n$，则 $A\vec{x}=\vec{b}$ 的解不唯一，且它的通解具有 $n-r$ 个独立参数

例子 2.3.12. 考虑如下线性方程组

$$\begin{cases} x+y+z=3 \\ 2x+3y+5z=4 \\ x+2y+4z=1 \end{cases}$$

这组方程在之前的例子中计算过，其通解为 $\begin{cases} x=2c+5 \\ y=-2-3c \\ z=c \end{cases}$ 这里 c 是任意常数。把其写作

$$\text{通解为}\quad \begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix}+c\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$$

这里 $\begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix}$ 是线性方程组的一个特解，$c\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$ 是对应齐次方程组的通解。

2.3.4 秩-零化度定理

我们讨论线性方阵组解的结构定理的几何解释。

定义 2.3.13. 设

$$f:\mathbb{R}^n\to\mathbb{R}^m$$

是一个线性映射。映射 f 的像是 $\mathbb{R}^m$ 中的子集记为 $\operatorname{im}(f)$。定义 f 的核为 $\mathbb{R}^n$ 中如下子集

$$\ker(f):=\{\vec{x}\in\mathbb{R}^n|\ f(\vec{x})=0\}$$

设线性映射 $f:\mathbb{R}^n\to\mathbb{R}^m$ 对应的矩阵表达为

$$f(\vec{x})=A\vec{x}\qquad \vec{x}\in\mathbb{R}^n$$

这里 A 是 $m\times n$ 矩阵，记为

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

f 的核通过矩阵 A 等价表达为

$$\ker(f) := \{\vec{x} \in \mathbb{R}^n |\ A\vec{x} = 0\}$$

因此核的概念即刻画了齐次线性方程组的解

$$\ker(f) = \{\text{齐次线性方程组 } A\vec{x} = 0 \text{ 的解}\}$$

由命题2.3.1知，$\ker(f) \subset \mathbb{R}^n$ 是线性子空间。

f 的像也可以通过矩阵 A 来刻画。把 A 写成列向量的样子

$$A = \begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} \qquad \text{这里} \quad \vec{\alpha}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \in \mathbb{R}^m$$

则 $\vec{y} \in \operatorname{im}(f)$ 当且仅当存在 $\vec{x} \in \mathbb{R}^n$ 使得 $\vec{y} = A\vec{x}$。写成 A 的列向量的样子

$$\vec{y} = x_1\vec{\alpha}_1 + x_2\vec{\alpha}_2 + \cdots + x_n\vec{\alpha}_n$$

即等价于 $\vec{y}$ 可以通过 $\{\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_n\}$ 线性表达。因此

$$\operatorname{im}(f) = \operatorname{Span}\{\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_n\}$$

特别的，$\operatorname{im}(f) \subset \mathbb{R}^m$ 也是线性子空间。

定理 2.3.14 (秩-零化度定理)**.** 设 $f : \mathbb{R}^n \to \mathbb{R}^m$ 是一个线性映射。则

$$\dim \operatorname{im}(f) + \dim \ker(f) = n$$

证明：由矩阵的秩的定义，我们知道

$$\dim \operatorname{im}(f) = \dim \operatorname{Span}\{\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_n\} = A \text{ 的列秩} = \operatorname{rank} A$$

由定理2.3.9，

$$\dim \ker(f) = n - \operatorname{rank} A = n - \dim \operatorname{im}(f)$$

□

秩-零化度定理的几何是非常直观的。我们可以把映射 f 看作把 $\mathbb{R}^n$ “压缩” 到它的像 $\operatorname{im}(f)$，其中被 “压缩” 的部分是它的核 $\ker(f)$。则 $\mathbb{R}^n$ 的维数是像的维数加上被压缩的维数，即秩-零化度定理。

秩-零化度定理也可以表述在抽象的线性空间里。

定理 2.3.15 (秩-零化度定理)**.** 设 V, W 是有限维线性空间，$f : V \to W$ 是一个线性映射。则

$$\dim \operatorname{im}(f) + \dim \ker(f) = \dim V$$

这里像 $\operatorname{im}(f) \subset W$ 与核 $\ker(f) \subset V$ 的定义与上述类似。通过选取一组基，可以把定理2.3.15转化为定理2.3.14的情形，具体细节留作练习。

2.4 习题

1. 计算如下两个矩阵 A, B 的乘积 AB：

$$(1)\quad \begin{bmatrix} 2 & 1 & 3 \\ 0 & 4 & 1 \\ 5 & 2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 0 \end{bmatrix} \qquad (2)\quad \begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \\ 5 & 2 & 3 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 3 & 4 \end{bmatrix} \qquad (3)\quad \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}\begin{bmatrix} 5 & 6 & 7 & 8 \end{bmatrix}$$

2. 设 $n \times n$ 矩阵 A, B 满足 $AB = BA$，证明 $(A+B)^k = \sum\limits_{i=0}^{k} \frac{k!}{i!(k-i)!} A^i B^{k-i}$。举例说明如果 $AB \neq BA$，那么这个公式不成立。

3. 考虑如下矩阵 A，计算 A^n，这里 n 是正整数

$$(1)\quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad (2)\quad \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} \qquad (3)\quad \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$(4)\quad \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (5)\quad \begin{bmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \end{bmatrix}$$

4. 求如下齐次线性方程组的通解

$$(1)\quad \begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 0 \\ 3x_1 + 2x_2 + x_3 + x_4 - 3x_5 = 0 \\ x_2 + 2x_3 + 2x_4 + 6x_5 = 0 \\ 5x_1 + 4x_2 + 3x_3 + 3x_4 - x_5 = 0 \end{cases} \qquad (2)\quad \begin{cases} x_1 + x_2 - 3x_4 - x_5 = 0 \\ x_1 - x_2 - 2x_3 - x_4 = 0 \\ 4x_1 - 2x_2 + 6x_3 + 3x_4 - 4x_5 = 0 \\ 2x_1 + 4x_2 - 2x_3 + 4x_4 - 7x_5 = 0 \end{cases}$$

5. 如果 $m < n$，证明如下齐次线性方程组一定有非零解

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

6. 补充完整命题2.3.7的证明，即矩阵的列秩在初等行变换下不变。

7. 求如下非齐次线性方程组的通解

(1) $$\begin{cases} 3x-2y+z=7 \\ 2x+y-3z=9 \\ x-y+2z=4 \end{cases}$$

(2) $$\begin{cases} x_1+2x_2-x_3+3x_4=-1 \\ 2x_1-x_2+4x_3-x_4=13 \\ -3x_1+4x_2+x_3+2x_4=6 \\ x_1-x_2+2x_3+x_4=4 \end{cases}$$

(3) $$\begin{cases} 3x_1-x_2+2x_3+x_4=9 \\ x_1+4x_2-2x_3+3x_4=5 \\ -x_1+2x_2+x_3-x_4=2 \\ 5x_1+x_2-x_3+5x_4=10 \end{cases}$$

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

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

8. 当 λ 取何值时，如下线性方程组有解? 有唯一解? 并求其所有解

$$\begin{cases} \lambda x_1+x_2+x_4=1 \\ x_1+\lambda x_2+x_3=1 \\ x_2+\lambda x_3+x_4=1 \\ x_1+x_3+\lambda x_4=1 \end{cases}$$

9. 设 A 是 $m\times n$ 矩阵，$\mathrm{rank}A=n-1$。证明齐次线性方程组 $A\vec{x}=0$ 的任意两个解成比例，即相差一个数值因子。

10. 设 A 是 $m\times n$ 矩阵，且 $\mathrm{rank}A=1$。证明 A 可以写成一个 $m\times 1$ 矩阵和一个 $1\times n$ 矩阵的乘积:

$$A=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}\begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}$$

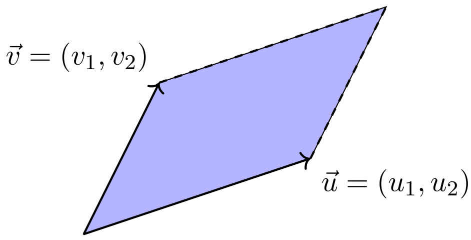
11. 证明定理2.3.15。

第三章　行列式

3.1　面积元与体积元

3.1.1　面积

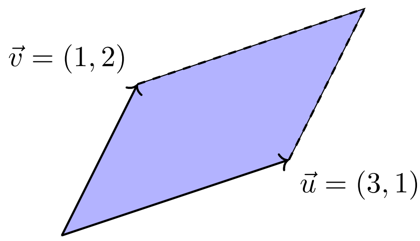
考虑平面上由向量 $\vec{u}=(u_1,u_2)$ 和 $\vec{v}=(v_1,v_2)$ 张成的平行四边形。



它的面积用这两个向量可以表示为

$$\text{面积}=|u_1v_2-u_2v_1|$$

例子 3.1.1. 平行四边形



的面积为

$$\text{面积}=|3\times 2-1\times 1|=5$$

我们引入下面的记号表达

$$\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} := u_1v_2-u_2v_1$$

称为 2×2 矩阵 $A=\begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}$ 的行列式。平行四边形面积为行列式 $\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$ 的绝对值。

观察到如果交换两行

$$\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}=u_1v_2-u_2v_1 \qquad \begin{vmatrix} v_1 & v_2 \\ u_1 & u_2 \end{vmatrix}=v_1u_2-v_2u_1$$

即

$$\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} = -\begin{vmatrix} v_1 & v_2 \\ u_1 & u_2 \end{vmatrix}$$

我们可以把 $\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}$ 理解为带符号的面积。这个符号和定向有关，即两个向量的相互顺序。

如果我们把其中一个向量做伸缩，那么面积也相应的伸缩。因此我们有如下性质

$$\begin{vmatrix} \lambda u_1 & \lambda u_2 \\ v_1 & v_2 \end{vmatrix} = \lambda \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 \\ \lambda v_1 & \lambda v_2 \end{vmatrix}$$

我们可以用一个代数方法来表述 2×2 矩阵的行列式。对于两个向量 $\vec{u}, \vec{v}$，我们引入符号

$$\vec{u} \wedge \vec{v}$$

称为它们的**外积**。外积的基本运算规则是

- 反对称性：$\vec{u} \wedge \vec{v} = -\vec{v} \wedge \vec{u}$, 特别有 $\vec{u} \wedge \vec{u} = 0$
- 线性性：

$$\begin{aligned}(\vec{u} + \vec{v}) \wedge \vec{w} &= \vec{u} \wedge \vec{w} + \vec{v} \wedge \vec{w} \\ \vec{u} \wedge (\vec{v} + \vec{w}) &= \vec{u} \wedge \vec{v} + \vec{u} \wedge \vec{w} \\ (\lambda \vec{u}) \wedge \vec{v} &= \lambda(\vec{u} \wedge \vec{v}) = \vec{u} \wedge (\lambda \vec{v})\end{aligned}$$

我们不详细讨论外积的构造，我们通过一些例子来熟悉这个运算。记 $\mathbb{R}^2$ 的标准基为

$$\vec{e}_1 = (1,0) \quad \vec{e}_2 = (0,1)$$

对向量 $\vec{u} = (u_1, u_2) = u_1\vec{e}_1 + u_2\vec{e}_2, \vec{v} = (v_1, v_2) = v_1\vec{e}_1 + v_2\vec{e}_2$, 其外积为

$$\begin{aligned}\vec{u} \wedge \vec{v} &= (u_1\vec{e}_1 + u_2\vec{e}_2) \wedge \vec{v} = u_1\vec{e}_1 \wedge \vec{v} + u_2\vec{e}_2 \wedge \vec{v} \\ &= u_1\vec{e}_1 \wedge (v_1\vec{e}_1 + v_2\vec{e}_2) + u_2\vec{e}_2 \wedge (v_1\vec{e}_1 + v_2\vec{e}_2) \\ &= u_1v_2\vec{e}_1 \wedge \vec{e}_2 + u_2v_1\vec{e}_2 \wedge \vec{e}_1 \\ &= (u_1v_2 - u_2v_1)\vec{e}_1 \wedge \vec{e}_2\end{aligned}$$

我们观察到

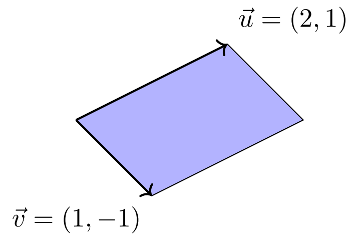
$$\vec{u} \wedge \vec{v} = \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \vec{e}_1 \wedge \vec{e}_2$$

因此外积不仅给出了一个便于记忆的行列式的计算方法，而且给出了它的几何解释：

- $\vec{u} \wedge \vec{v}$ 是 $\vec{u}, \vec{v}$ 张成的平行四边形的面积元
- $\vec{e}_1 \wedge \vec{e}_2$ 是单位向量 $\vec{e}_1, \vec{e}_2$ 张成的单位面积元

它们之间的比例就是平行四边形面积（带符号）。

例子 3.1.2. 计算 $\vec{u}=(2,1),\vec{v}=(1,-1)$ 张成平行四边形面积

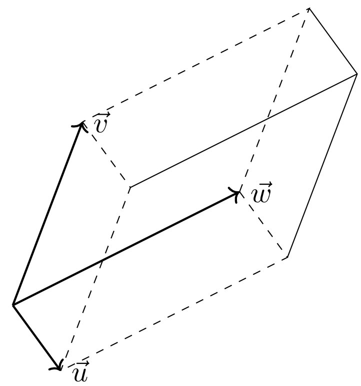


$$\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} = 2\cdot(-1) - 1\cdot 1 = -3$$

这里的负号是由于 $\vec{u},\vec{v}$ 的定向顺序（顺时针）和 $\vec{e}_1,\vec{e}_2$ 的定向顺序（逆时针）相反。平行四边形的面积为 3。

3.1.2 体积

考虑空间中由向量 $\vec{u}=(u_1,u_2,u_3)$，$\vec{v}=(v_1,v_2,v_3)$ 和 $\vec{w}=(w_1,w_2,w_3)$ 张成的平行六面体。



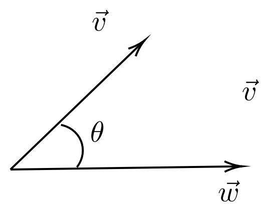
我们同样可以通过这 3 个向量的坐标来计算它的体积。

回顾与空间 $\mathbb{R}^3$ 中的向量相关的基本构造:

- 长度：$\vec{v}=(v_1,v_2,v_3)$ 的长度 $|\vec{v}| := \sqrt{v_1^2+v_2^2+v_3^2}$
- 点乘（内积）：$\vec{v}=(v_1,v_2,v_3)$，$\vec{w}=(w_1,w_2,w_3)$

$$\vec{v}\cdot\vec{w} := v_1w_1+v_2w_2+v_3w_3$$

点乘（内积）的几何含义为



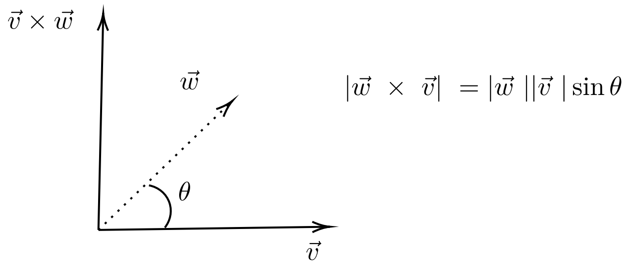
$$\vec{v}\cdot\vec{w} \quad = |\vec{v}\,||\vec{w}\,|\cos\theta$$

例如 $\vec{v}\cdot\vec{v}=|\vec{v}|^2$。

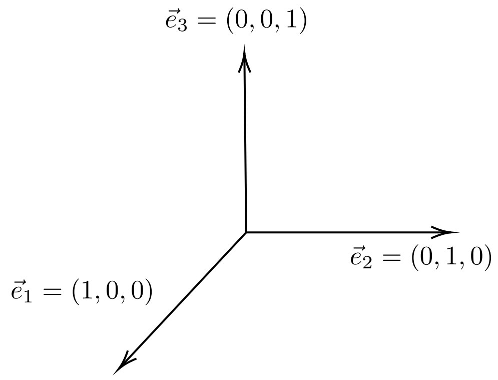
- 叉乘: $\vec{v} = (v_1, v_2, v_3)$, $\vec{w} = (w_1, w_2, w_3)$ 的叉乘为

$$\vec{v} \times \vec{w} := (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1)$$

  $\vec{v} \times \vec{w}$ 是 $\mathbb{R}^3$ 中的向量，其方向与 $\vec{v}$ 和 $\vec{w}$ 垂直，指向按照右手拇指法则，长度为 $\vec{v}$ 和 $\vec{w}$ 围成的平行四边形面积。



例子 3.1.3. 对于 $\mathbb{R}^3$ 中的标准基



$$\begin{aligned} \vec{e}_1 \times \vec{e}_2 = \vec{e}_3 = -\vec{e}_2 \times \vec{e}_1 \\ \vec{e}_2 \times \vec{e}_3 = \vec{e}_1 = -\vec{e}_3 \times \vec{e}_2 \\ \vec{e}_3 \times \vec{e}_1 = \vec{e}_2 = -\vec{e}_1 \times \vec{e}_3 \end{aligned}$$

由平行六面体的体积公式：体积 = 底面积 × 高。我们把 $\vec{v}, \vec{w}$ 张成的面作为底面，得到

$$\text{体积} = \vec{u} \cdot (\vec{v} \times \vec{w}) \text{ 的绝对值}$$

同样，我们可以把 $\vec{u} \cdot (\vec{v} \times \vec{w})$ 理解为带符号的体积。它的正负号和 3 个向量 $\vec{u}, \vec{v}, \vec{w}$ 的定向有关。

定义 3.1.4. 定义 3×3 矩阵的行列式为

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} := \vec{u} \cdot (\vec{v} \times \vec{w})$$

由 $\vec{v} \times \vec{w} = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1)$

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = u_1(v_2w_3 - v_3w_2) - u_2(v_1w_3 - v_3w_1) + u_3(v_1w_2 - v_2w_1)$$

叉乘满足反对称性:

$$\vec{v} \times \vec{w} = -\vec{w} \times \vec{v} \quad \text{特别地} \quad \vec{v} \times \vec{v} = 0$$

由此对于 3×3 的行列式，如果对换其中任何两行则行列式的值差一个负号。例如

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = -\begin{vmatrix} u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = -\begin{vmatrix} v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

特别地，如果有两行向量相同，则行列式为 0

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = 0$$

几何上，当 $\vec{v} = \vec{u}$ 时，平行六面体塌缩成一个二维面，其体积为 0。

如果我们把其中一个向量做伸缩，那么体积也相应的伸缩。即我们有如下性质

$$\begin{vmatrix} \lambda u_1 & \lambda u_2 & \lambda u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 & u_3 \\ \lambda v_1 & \lambda v_2 & \lambda v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \lambda \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

我们同样可以用外积来表示体积元和行列式。对于 $\mathbb{R}^3$ 中的 3 个向量，考虑它们的外积

$$\vec{u} \wedge \vec{v} \wedge \vec{w}$$

外积的代数运算法则和 $\mathbb{R}^2$ 的情况一样，满足：

- 反对称性：交换任意两个向量差一个负号

$$\vec{u} \wedge \vec{v} \wedge \vec{w} = -\vec{v} \wedge \vec{u} \wedge \vec{w} = -\vec{u} \wedge \vec{w} \wedge \vec{v} = -\vec{w} \wedge \vec{v} \wedge \vec{u}$$

- 线性性：关于各分量都具有线性性，例如

$$\vec{u} \wedge (\vec{v}_1 + \vec{v}_2) \wedge \vec{w} = \vec{u} \wedge \vec{v}_1 \wedge \vec{w} + \vec{u} \wedge \vec{v}_2 \wedge \vec{w}$$

记 $\mathbb{R}^3$ 的标准基为

$$\vec{e}_1 = (1,0,0) \quad \vec{e}_2 = (0,1,0) \quad \vec{e}_3 = (0,0,1)$$

对于向量

$$\vec{u} = (u_1, u_2, u_3) = u_1\vec{e}_1 + u_2\vec{e}_2 + u_3\vec{e}_3$$
$$\vec{v} = (v_1, v_2, v_3) = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3$$
$$\vec{w} = (w_1, w_2, w_3) = w_1\vec{e}_1 + w_2\vec{e}_2 + w_3\vec{e}_3$$

我们可以按照外积的运算法则来计算 $\vec{u} \wedge \vec{v} \wedge \vec{w}$

$$\vec{u} \wedge \vec{v} \wedge \vec{w} = u_1\vec{e}_1 \wedge \vec{v} \wedge \vec{w} + u_2\vec{e}_2 \wedge \vec{v} \wedge \vec{w} + u_3\vec{e}_3 \wedge \vec{v} \wedge \vec{w}$$

其中

$$\begin{aligned}\vec{e}_1 \wedge \vec{v} \wedge \vec{w} =& \vec{e}_1 \wedge (v_2\vec{e}_2 + v_3\vec{e}_3) \wedge (w_2\vec{e}_2 + w_3\vec{e}_3)\\ =&(v_2w_3 - v_3w_2)\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3\end{aligned}$$

因此

$$\begin{aligned}&\vec{u} \wedge \vec{v} \wedge \vec{w}\\ =&u_1\vec{e}_1 \wedge \vec{v} \wedge \vec{w} + u_2\vec{e}_2 \wedge \vec{v} \wedge \vec{w} + u_3\vec{e}_3 \wedge \vec{v} \wedge \vec{w}\\ =&u_1(v_2w_3 - v_3w_2)\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3 + u_2(v_1w_3 - v_3w_1)\vec{e}_2 \wedge \vec{e}_1 \wedge \vec{e}_3 + u_3(v_1w_2 - v_2w_1)\vec{e}_3 \wedge \vec{e}_1 \wedge \vec{e}_2\\ =&\left(u_1(v_2w_3 - v_3w_2) - u_2(v_1w_3 - v_3w_1) + u_3(v_1w_2 - v_2w_1)\right)\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3\end{aligned}$$

对比行列式的公式，我们发现

$$\vec{u} \wedge \vec{v} \wedge \vec{w} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3$$

这给出了 3×3 矩阵行列式的几何解释：

- $\vec{u} \wedge \vec{v} \wedge \vec{w}$ 是 $\vec{u}, \vec{v}, \vec{w}$ 张成平行六面体的体积元
- $\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3$ 是单位向量 $\vec{e}_i$ 张成的单位体积元

它们之间的比例就是平行六面体体积（带符号），即 3×3 矩阵的行列式

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

利用外积的反对称性：

$$\vec{u} \wedge \vec{v} \wedge \vec{w} = -\vec{u} \wedge \vec{w} \wedge \vec{v} = -\vec{v} \wedge \vec{u} \wedge \vec{w}$$

我们很容易得出前述行列式的反对称性

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = -\begin{vmatrix} u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = -\begin{vmatrix} v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

利用外积的线性性，例如

$$\vec{u} \wedge (\lambda\vec{v} + \mu\vec{q}) \wedge \vec{w} = \lambda\vec{u} \wedge \vec{v} \wedge \vec{w} + \mu\vec{u} \wedge \vec{q} \wedge \vec{w}$$

我们得出行列式关于任一行的线性性

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ \lambda v_1 + \mu q_1 & \lambda v_2 + \mu q_2 & \lambda v_3 + \mu q_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \lambda\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} + \mu\begin{vmatrix} u_1 & u_2 & u_3 \\ q_1 & q_2 & q_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

最后，由行列式的具体公式

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = u_1(v_2w_3 - v_3w_2) - u_2(v_1w_3 - v_3w_1) + u_3(v_1w_2 - v_2w_1)$$

我们得到 3 阶行列式和 2 阶行列式之间的关系

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = u_1 \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - u_2 \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + u_3 \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}$$

我们将看到一般的行列式都有类似的性质。

例子 3.1.5.

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \\ 2 & 1 & -2 \end{vmatrix} =& 1 \cdot \begin{vmatrix} -1 & 4 \\ 1 & -2 \end{vmatrix} - 2 \cdot \begin{vmatrix} 0 & 4 \\ 2 & -2 \end{vmatrix} + 3 \cdot \begin{vmatrix} 0 & -1 \\ 2 & 1 \end{vmatrix} \\ =& 1 \cdot (2-4) - 2 \cdot (0-8) + 3 \cdot (0-(-2)) \\ =& -2 + 16 + 6 = 20 \end{aligned}$$

例子 3.1.6 (3 维空间叉乘的行列式表示). $\mathbb{R}^3$ 中两个向量 $\vec{v} = (v_1, v_2, v_3)$, $\vec{w} = (w_1, w_2, w_3)$ 的叉乘

$$\vec{v} \times \vec{w} = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1)$$

可以表达为行列式

$$\vec{v} \times \vec{w} = \begin{vmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

这里行列式的第一行为 $\mathbb{R}^3$ 的标准基。虽然第一行的元素不是数字，我们还是形式地用如上行列式的公式来计算。可以看出

$$\begin{aligned} \begin{vmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} =& \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} \vec{e}_1 - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} \vec{e}_2 + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} \vec{e}_3 \\ =& (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) = \vec{v} \times \vec{w} \end{aligned}$$

3.2 n 阶行列式

3.2.1 行列式的几何定义

考虑一个 $n \times n$ 矩阵（也称为 n 阶方阵）

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

我们定义一个数称为它的行列式 $\det A$，也记为

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

我们之前讨论了 $n = 2, 3$ 的情况及其几何含义

- 当 $n = 2$ 时

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

- 当 $n = 3$ 时

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

我们下面讨论一般 n 的情况。记 $\mathbb{R}^n$ 的标准基为

$$\vec{e}_1 = (1, 0, \cdots, 0) \quad \vec{e}_2 = (0, 1, \cdots, 0) \quad \cdots \quad \vec{e}_n = (0, \cdots, 0, 1)$$

我们记 A 的行向量为

$$\begin{aligned} \vec{\alpha}_1 &= (a_{11}, a_{12}, \cdots, a_{1n}) = a_{11}\vec{e}_1 + a_{12}\vec{e}_2 + \cdots + a_{1n}\vec{e}_n \\ \vec{\alpha}_2 &= (a_{21}, a_{22}, \cdots, a_{2n}) = a_{21}\vec{e}_1 + a_{22}\vec{e}_2 + \cdots + a_{2n}\vec{e}_n \\ &\vdots \\ \vec{\alpha}_n &= (a_{n1}, a_{n2}, \cdots, a_{nn}) = a_{n1}\vec{e}_1 + a_{n2}\vec{e}_2 + \cdots + a_{nn}\vec{e}_n \end{aligned}$$

我们用外积的代数运算（反对称和线性）计算

$$\begin{aligned} & \vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n \\ =& (a_{11}\vec{e}_1 + a_{12}\vec{e}_2 + \cdots + a_{1n}\vec{e}_n) \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n \\ =& a_{11}\vec{e}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n + \cdots + a_{1n}\vec{e}_n \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n \\ =& \cdots \\ =& D \ \vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n \end{aligned}$$

最后总是可以写成 $\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$ 的一个倍数，这个系数 D 就定义为 A 的行列式

$$\det A := D$$

由这个构造可以看出，当 $n = 2, 3$ 的时候，行列式与面积元和体积元的定义是一致的。我们可以把 $\vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n$ 理解为 $\mathbb{R}^n$ 的 n 个向量 $\vec{\alpha}_1, \cdots, \vec{\alpha}_n$ 张成的 n 维平行体的体积元，$\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$ 是单位体积元。因此行列式的几何含义是代表 $\mathbb{R}^n$ 中平行体的体积（带符号）。

3.2.2 行列式的组合定义

我们可以通过外积的代数运算给出行列式的一个具体的组合表达式。由外积的反对称性，特别有 $\vec{e}_i \wedge \vec{e}_i = 0$，我们可以看出把

$$\vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n$$

展开成 $\vec{e}_i$ 的过程中，如要得到非零元素，只能在 $\vec{\alpha}_1$ 中取一个 $\vec{e}_{\sigma(1)}$，在 $\vec{\alpha}_2$ 中取一个 $\vec{e}_{\sigma(2)}$，$\cdots$，在 $\vec{\alpha}_n$ 中取一个 $\vec{e}_{\sigma(n)}$，并且 $\sigma(1), \sigma(2), \cdots, \sigma(n)$ 各不相同。这样一个取法 σ 对应于集合 $\{1, 2, \cdots, n\}$ 到自身的一个一一映射。

定义 3.2.1. 集合 $\{1, 2, \cdots, n\}$ 到自身的一个一一映射称为一个 n 元置换。所有 n 元置换的集合记为 S_n（其包含 $n!$ 个元素）。

设 σ 是一个 n 元置换。我们可以通过不断作两两对换把

$$\vec{e}_{\sigma(1)} \wedge \vec{e}_{\sigma(2)} \wedge \cdots \wedge \vec{e}_{\sigma(n)}$$

调整为标准形式 $\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$。在外积运算规则中每对换一次变号一次，因此

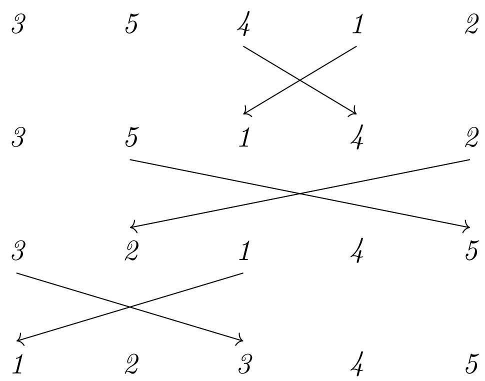
$$\vec{e}_{\sigma(1)} \wedge \vec{e}_{\sigma(2)} \wedge \cdots \wedge \vec{e}_{\sigma(n)} = (\operatorname{sign} \sigma) \vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$$

这里 $\operatorname{sign} \sigma = \pm 1$ 称为置换 σ 的符号。如果需要经过偶数次对换，那么 $\operatorname{sign} \sigma = 1$，这样的 σ 称为偶置换；如果需要经过奇数次对换，那么 $\operatorname{sign} \sigma = -1$，这样的 σ 称为奇置换。可以证明，虽然通过对换来调整为标准形式的方式不唯一，但是需要的对换次数的奇偶性是不变的。

例子 3.2.2. *例如考虑置换 $\sigma : \{1, 2, 3, 4, 5\} \to \{3, 5, 4, 1, 2\}$。我们可以通过 3 次对换把 $\vec{e}_3 \wedge \vec{e}_5 \wedge \vec{e}_4 \wedge \vec{e}_1 \wedge \vec{e}_2$ 调整为 $\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3 \wedge \vec{e}_4 \wedge \vec{e}_5$：*

$$\vec{e}_3 \wedge \vec{e}_5 \wedge \vec{e}_4 \wedge \vec{e}_1 \wedge \vec{e}_2 = -\vec{e}_3 \wedge \vec{e}_5 \wedge \vec{e}_1 \wedge \vec{e}_4 \wedge \vec{e}_2 = \vec{e}_3 \wedge \vec{e}_2 \wedge \vec{e}_1 \wedge \vec{e}_4 \wedge \vec{e}_5 = -\vec{e}_1 \wedge \vec{e}_2 \wedge \vec{e}_3 \wedge \vec{e}_4 \wedge \vec{e}_5$$

因此 σ 是奇置换，$\operatorname{sign} \sigma = -1$。这个外积计算过程也可以用下图来表示



定理 3.2.3. *设 A 是 n 阶方阵，矩阵元为 a_{ij}，则*

$$\det A = \sum_{\sigma \in S_n} \operatorname{sign} \sigma \prod_{i=1}^{n} a_{i\sigma(i)}$$

这里 S_n 是所有 n 元置换的集合。

证明: A 的行向量为

$$\begin{aligned}
\vec{\alpha}_1 &= a_{11}\vec{e}_1 + a_{12}\vec{e}_2 + \cdots + a_{1n}\vec{e}_n \\
\vec{\alpha}_2 &= a_{21}\vec{e}_1 + a_{22}\vec{e}_2 + \cdots + a_{2n}\vec{e}_n \\
&\vdots \\
\vec{\alpha}_n &= a_{n1}\vec{e}_1 + a_{n2}\vec{e}_2 + \cdots + a_{nn}\vec{e}_n
\end{aligned}$$

我们有

$$\begin{aligned}
&\vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n \\
=&\sum_{\sigma\in S_n} a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}\vec{e}_{\sigma(1)} \wedge \vec{e}_{\sigma(2)} \wedge \cdots \wedge \vec{e}_{\sigma(n)} \\
=&\sum_{\sigma\in S_n} \operatorname{sign}\sigma \prod_{i=1}^{n} a_{i\sigma(i)}\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n
\end{aligned}$$

比较行列式的外积定义，我们得到 $\det A = \sum_{\sigma\in S_n} \operatorname{sign}\sigma \prod_{i=1}^{n} a_{i\sigma(i)}$。 □

注记. 在很多教材中，通常从定理3.2.3中的组合公式出发来定义行列式。

例子 3.2.4. 考虑如下行列式

$$\begin{vmatrix}
a_{1,1} & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\
a_{2,1} & a_{2,2} & \cdots & a_{2,n-1} & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
a_{n-1,1} & a_{n-1,2} & 0 & \cdots & 0 \\
a_{n,1} & 0 & 0 & \cdots & 0
\end{vmatrix}$$

我们按照组合公式来计算。先看第 n 行，只能取 $a_{n,1}$。取完后看第 $n-1$ 行，这时只能取 $a_{n-1,2}$。依次从下往上类推，我们发现只有 $a_{1,n}a_{2,n-1}\cdots a_{n,1}$ 这一项有贡献，其对应于置换

$$\sigma : \{1, 2, \cdots, n\} \to \{n, n-1, \cdots, 1\} \qquad \operatorname{sign}\sigma = (-1)^{n(n-1)/2}$$

因此

$$\begin{vmatrix}
a_{1,1} & a_{1,2} & \cdots & a_{1,n-1} & a_{1,n} \\
a_{2,1} & a_{2,2} & \cdots & a_{2,n-1} & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
a_{n-1,1} & a_{n-1,2} & 0 & \cdots & 0 \\
a_{n,1} & 0 & 0 & \cdots & 0
\end{vmatrix} = (-1)^{n(n-1)/2}\, a_{1,n}\, a_{2,n-1}\cdots a_{n,1}$$

3.2.3 行列式的基本性质

命题 3.2.5. 把矩阵的某一行乘以一个系数，则它的行列式的值也乘以同一个系数

$$\begin{vmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
\vdots & \vdots & \vdots & \vdots \\
\lambda a_{i1} & \lambda a_{i2} & \cdots & \lambda a_{in} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{vmatrix} = \lambda \begin{vmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
\vdots & \vdots & \vdots & \vdots \\
a_{i1} & a_{i2} & \cdots & a_{in} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{vmatrix}$$

证明: 由行列式的组合定义显然。从几何角度，这个性质对应于外积的线性性

$$\vec{\alpha}_1 \wedge \cdots \wedge (\lambda \vec{\alpha}_i) \wedge \cdots \wedge \vec{\alpha}_n = \lambda \vec{\alpha}_1 \wedge \cdots \wedge \vec{\alpha}_i \wedge \cdots \wedge \vec{\alpha}_n$$

□

命题 3.2.6. 交换矩阵的任意两行，行列式差一个负号

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{vmatrix} = - \begin{vmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & a_{nn} \end{vmatrix}$$

证明: 这个性质对应于外积的反对称性。 □

命题 3.2.7. 如果矩阵有两行一样，则行列式为 0。

证明: 假设第一行和第二行一样，交换这两行

$$\begin{vmatrix} a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix} = - \begin{vmatrix} a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

由此可得

$$\begin{vmatrix} a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix} = 0$$

□

命题 3.2.8. 如果某一行是另一行的倍数，则行列式为 0。

证明:

$$\begin{vmatrix} \vdots & \vdots & \vdots & \vdots \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \vdots \\ \lambda a_1 & \lambda a_2 & \cdots & \lambda a_n \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix} = \lambda \begin{vmatrix} \vdots & \vdots & \vdots & \vdots \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \vdots \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix} = 0$$

□

命题 3.2.9. 行列式对于任何一行具有如下加法性质

$$\begin{vmatrix} \vdots & \vdots & \vdots \\ a_1 + b_1 & \cdots & a_n + b_n \\ \vdots & \vdots & \vdots \end{vmatrix} = \begin{vmatrix} \vdots & \vdots & \vdots \\ a_1 & \cdots & a_n \\ \vdots & \vdots & \vdots \end{vmatrix} + \begin{vmatrix} \vdots & \vdots & \vdots \\ b_1 & \cdots & b_n \\ \vdots & \vdots & \vdots \end{vmatrix}$$

证明: 这个对应于外积的线性性

$$\cdots\wedge(\vec{\alpha}+\vec{\beta})\wedge\cdots=\cdots\wedge\vec{\alpha}\wedge\cdots+\cdots\wedge\vec{\beta}\wedge\cdots$$

□

命题 3.2.10. 把某一行乘以一个数加到另一行，行列式值不变。

证明:

$$\begin{vmatrix} \vdots & \vdots & \vdots & \vdots \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \vdots & \vdots \\ b_1+\lambda a_1 & b_2+\lambda a_2 & \cdots & b_n+\lambda a_n \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

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

□

命题 3.2.11. 对角方阵的行列式是对角元的乘积

$$\begin{vmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix}=a_{11}a_{22}\cdots a_{nn}$$

证明: 由行列式的组合定义显然。从几何角度，

$$a_{11}\vec{e}_1\wedge a_{22}\vec{e}_2\wedge\cdots\wedge a_{nn}\vec{e}_n=\ a_{11}a_{22}\cdots a_{nn}\ \vec{e}_1\wedge\cdots\wedge\vec{e}_n$$

□

命题 3.2.12. 上三角方阵的行列式是对角元的乘积

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix}=a_{11}a_{22}\cdots a_{nn}$$

证明：由行列式的组合公式容易看出，只有 $a_{11}a_{22}\cdots a_{nn}$ 有贡献，对应的恒等置换是偶置换。 $\square$

命题 3.2.13.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$

证明：由行列式的组合公式，如果第一行不取 a_{11}，则后面一定有某个第 k 行 $(k > 1)$ 需要取 $a_{k1} = 0$，因此贡献为 0。故第一行只能取 a_{11}。由此可以证明命题。 $\square$

3.2.4 初等变换计算行列式

上述关于行列式的性质给出了计算行列式的一般方法。回顾矩阵的三种初等行变换：

1. 交换两行：将矩阵的两行交换位置
2. 将矩阵中的一行乘以一个非零常数
3. 将矩阵的一行乘以一个数加到另一行

对于 n 阶方阵，对应于上述三种初等行变换

1. 行列式变号
2. 行列式乘以同一个非零常数
3. 行列式保持不变

我们可以用初等行变换把矩阵变成梯形。对于 n 阶方阵，即变成上三角矩阵。而上三角矩阵的行列式为对角元相乘。因此我们通过初等变换把 n 阶方阵约化为上三角矩阵的过程，同时给出如何计算 n 阶行列式。

例子 3.2.14. *计算以下 3 阶行列式*

$$\begin{vmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \\ 2 & 1 & -2 \end{vmatrix}$$

将第一行乘以 -2 加到第三行：

$$\begin{vmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \\ 2 & 1 & -2 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \\ 0 & -3 & -8 \end{vmatrix}$$

将第二行乘以 -3 加到第三行:

$$\begin{vmatrix}1 & 2 & 3\\0 & -1 & 4\\0 & -3 & -8\end{vmatrix}=\begin{vmatrix}1 & 2 & 3\\0 & -1 & 4\\0 & 0 & -20\end{vmatrix}=1\cdot(-1)\cdot(-20)=20$$

即原行列式为 20

例子 3.2.15. 计算行列式

$$\begin{vmatrix}2 & 1 & -1 & 1\\1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\1 & 0 & 0 & -1\end{vmatrix}$$

将第一行和第二行交换:

$$\begin{vmatrix}2 & 1 & -1 & 1\\1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\1 & 0 & 0 & -1\end{vmatrix}=-\begin{vmatrix}1 & -1 & 2 & 0\\2 & 1 & -1 & 1\\0 & 1 & -1 & 1\\1 & 0 & 0 & -1\end{vmatrix}$$

将第一行乘以 -2 加到第二行，将第一行乘以 -1 加到第四行:

$$-\begin{vmatrix}1 & -1 & 2 & 0\\2 & 1 & -1 & 1\\0 & 1 & -1 & 1\\1 & 0 & 0 & -1\end{vmatrix}=-\begin{vmatrix}1 & -1 & 2 & 0\\0 & 3 & -5 & 1\\0 & 1 & -1 & 1\\0 & 1 & -2 & -1\end{vmatrix}$$

将第二行和第三行交换:

$$-\begin{vmatrix}1 & -1 & 2 & 0\\0 & 3 & -5 & 1\\0 & 1 & -1 & 1\\0 & 1 & -2 & -1\end{vmatrix}=\begin{vmatrix}1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\0 & 3 & -5 & 1\\0 & 1 & -2 & -1\end{vmatrix}$$

将第二行乘以 -3 加到第三行，将第二行乘以 -1 加到第四行:

$$\begin{vmatrix}1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\0 & 3 & -5 & 1\\0 & 1 & -2 & -1\end{vmatrix}=\begin{vmatrix}1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\0 & 0 & -2 & -2\\0 & 0 & -1 & -2\end{vmatrix}=-2\begin{vmatrix}1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\0 & 0 & 1 & 1\\0 & 0 & -1 & -2\end{vmatrix}$$

这里最后一个等式，我们把第三行提出系数 -2。将第三行加到第四行

$$-2\begin{vmatrix}1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\0 & 0 & 1 & 1\\0 & 0 & -1 & -2\end{vmatrix}=-2\begin{vmatrix}1 & -1 & 2 & 0\\0 & 1 & -1 & 1\\0 & 0 & 1 & 1\\0 & 0 & 0 & -1\end{vmatrix}=-2\cdot(-1)=2$$

由此我们得出原方阵的行列式

$$\begin{vmatrix} 2 & 1 & -1 & 1 \\ 1 & -1 & 2 & 0 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{vmatrix} = 2$$

例子 3.2.16. 我们给出上一个例子的另一个计算方法

$$\begin{aligned} \begin{vmatrix} 2 & 1 & -1 & 1 \\ 1 & -1 & 2 & 0 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{vmatrix} &= - \begin{vmatrix} 1 & -1 & 2 & 0 \\ 2 & 1 & -1 & 1 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{vmatrix} = - \begin{vmatrix} 1 & -1 & 2 & 0 \\ 0 & 3 & -5 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 1 & -2 & -1 \end{vmatrix} \\ &= -1 \cdot \begin{vmatrix} 3 & -5 & 1 \\ 1 & -1 & 1 \\ 1 & -2 & -1 \end{vmatrix} = -3 \begin{vmatrix} -1 & 1 \\ -2 & -1 \end{vmatrix} + (-5) \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} - 1 \cdot \begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix} \\ &= -3 \cdot 3 + (-5) \cdot (-2) - 1 \cdot (-1) = 2 \end{aligned}$$

3.3 Laplace 展开定理

我们在讨论面积元和体积元的时候，观察到 3 阶行列式可以化为 2 阶行列式的展开：

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = u_1 \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - u_2 \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + u_3 \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}$$

实际上 n 阶行列式也可以化为低阶行列式的展开，这个称为 Laplace 展开定理。

3.3.1 余子式与展开定理

给定 n 阶方程

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

定义 3.3.1. A 的第 i 行第 j 列余子式 A_{ij} 定义为把 A 的第 i 行和第 j 列去掉后得到的 $n-1$ 阶方阵的行列式，即

$$A_{ij} = \begin{vmatrix} a_{11} & \cdots & \cancel{a_{1j}} & \cdots & a_{1n} \\ \cdots & \cdots & \vdots & \cdots & \cdots \\ \cancel{a_{i1}} & \cdots & \cancel{a_{ij}} & \cdots & \cancel{a_{in}} \\ \cdots & \cdots & \vdots & \cdots & \cdots \\ a_{n1} & \cdots & \cancel{a_{nj}} & \cdots & a_{nn} \end{vmatrix}$$

例子 3.3.2. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$，则

$$A_{12} = \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = -6 \qquad A_{23} = \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = -6$$

例子 3.3.3. 体积元的展开公式

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

可以写成

$$\det A = a_{11}A_{11} - a_{12}A_{12} + a_{13}A_{13}$$

命题 3.3.4. 设 $A = (a_{ij})$ 是 n 阶方阵，则按照第一行展开有如下公式成立

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} A_{1j}$$

证明: 记 A 的行向量为

$$\begin{aligned} \vec{\alpha}_1 &= (a_{11}, a_{12}, \cdots, a_{1n}) = a_{11}\vec{e}_1 + a_{12}\vec{e}_2 + \cdots + a_{1n}\vec{e}_n \\ \vec{\alpha}_2 &= (a_{21}, a_{22}, \cdots, a_{2n}) = a_{21}\vec{e}_1 + a_{22}\vec{e}_2 + \cdots + a_{2n}\vec{e}_n \\ &\vdots \\ \vec{\alpha}_n &= (a_{n1}, a_{n2}, \cdots, a_{nn}) = a_{n1}\vec{e}_1 + a_{n2}\vec{e}_2 + \cdots + a_{nn}\vec{e}_n \end{aligned}$$

由行列式的外积定义，我们有

$$\vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n = (\det A)\ \vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$$

另一方面，代入 $\vec{\alpha}_1$ 的表达式

$$\vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n = (a_{11}\vec{e}_1 + a_{12}\vec{e}_2 + \cdots + a_{1n}\vec{e}_n) \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n$$

考虑其中的第一项 $a_{11}\vec{e}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n$, 记

$$\begin{aligned} \vec{\alpha}_2 &= a_{21}\vec{e}_1 + a_{22}\vec{e}_2 + \cdots + a_{2n}\vec{e}_n = a_{21}\vec{e}_1 + \vec{\beta}_2 \\ &\vdots \\ \vec{\alpha}_n &= a_{n1}\vec{e}_1 + a_{n2}\vec{e}_2 + \cdots + a_{nn}\vec{e}_n = a_{n1}\vec{e}_1 + \vec{\beta}_n \end{aligned}$$

则行向量 $\vec{\beta}_2, \cdots, \vec{\beta}_n$ 组成把 A 的第 1 行第 1 列去掉后得到的 $n-1$ 阶方阵。由外积代数我们有

$$\vec{\beta}_2 \wedge \vec{\beta}_3 \wedge \cdots \wedge \vec{\beta}_n = A_{11}\ \vec{e}_2 \wedge \vec{e}_3 \wedge \cdots \wedge \vec{e}_n$$

另一方面，由外积代数关系 $\vec{e}_1 \wedge \vec{e}_1 = 0$,

$$\vec{e}_1 \wedge \vec{\alpha}_2 = \vec{e}_1 \wedge (a_{21}\vec{e}_1 + \vec{\beta}_2) = \vec{e}_1 \wedge \vec{\beta}_2$$

类似的我们得到

$$\begin{aligned} & \vec{e}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n \\ = & \vec{e}_1 \wedge (a_{21}\vec{e}_1 + \vec{\beta}_2) \wedge \cdots \wedge (a_{n1}\vec{e}_1 + \vec{\beta}_n) \\ = & \vec{e}_1 \wedge \vec{\beta}_2 \wedge \cdots \wedge \vec{\beta}_n \end{aligned}$$

因此 $\det A$ 展开的第一项可以写成

$$a_{11}\vec{e}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n = a_{11}\vec{e}_1 \wedge \vec{\beta}_2 \wedge \cdots \wedge \vec{\beta}_n = a_{11}A_{11}\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$$

类似的，$\det A$ 展开的第 j 项可以计算为

$$a_{1j}\vec{e}_j \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n = a_{1j}A_{1j}\ \vec{e}_j \wedge \vec{e}_1 \wedge \cdots \cancel{\vec{e}_j} \cdots \wedge \vec{e}_n$$

我们可以用外积的反对称把 $\vec{e}_j$ 从最前面移动到第 j 个位置，共需要交换 $(j-1)$ 次。所以

$$\begin{aligned} & a_{1j}A_{1j}\ \vec{e}_j \wedge \vec{e}_1 \wedge \cdots \cancel{\vec{e}_j} \cdots \wedge \vec{e}_n \\ = & (-1)^{1+j} a_{1j}A_{1j}\ \vec{e}_1 \wedge \cdots \vec{e}_j \cdots \wedge \vec{e}_n \end{aligned}$$

把所有项加起来，我们得到

$$\begin{aligned} & \vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n \\ = & (a_{11}\vec{e}_1 + a_{12}\vec{e}_2 + \cdots + a_{1n}\vec{e}_n) \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n \\ = & \left(\sum_{j=1}^{n} (-1)^{1+j} a_{1j}A_{1j} \right) \vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n \end{aligned}$$

比较行列式的外积定义，我们证明了命题所述

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j}A_{1j}$$

□

命题 3.3.5. 设 $A = (a_{ij})$ 是 n 阶方阵，则按照第一列展开有如下公式成立

$$\det A = \sum_{i=1}^{n} (-1)^{i+1} a_{i1}A_{i1}$$

证明: 记 A 的行向量为

$$\begin{aligned} \vec{\alpha}_1 &= a_{11}\vec{e}_1 + a_{12}\vec{e}_2 + \cdots + a_{1n}\vec{e}_n = a_{11}\vec{e}_1 + \vec{\beta}_1 \\ \vec{\alpha}_2 &= a_{21}\vec{e}_1 + a_{22}\vec{e}_2 + \cdots + a_{2n}\vec{e}_n = a_{21}\vec{e}_1 + \vec{\beta}_2 \\ &\vdots \\ \vec{\alpha}_n &= a_{n1}\vec{e}_1 + a_{n2}\vec{e}_2 + \cdots + a_{nn}\vec{e}_n = a_{n1}\vec{e}_1 + \vec{\beta}_n \end{aligned}$$

利用外积的性质，我们可以展开得到

$$\begin{aligned}\vec{\alpha}_1 \wedge \vec{\alpha}_2 \wedge \cdots \wedge \vec{\alpha}_n =&(a_{11}\vec{e}_1 + \vec{\beta}_1) \wedge (a_{21}\vec{e}_1 + \vec{\beta}_2) \wedge \cdots \wedge (a_{n1}\vec{e}_1 + \vec{\beta}_n) \\ =&\sum_{i=1}^{n} (-1)^{i+1} a_{i1}\vec{e}_1 \wedge \vec{\beta}_1 \wedge \cdots \cancel{\vec{\beta}_i} \cdots \wedge \vec{\beta}_n \\ =&\sum_{i=1}^{n} (-1)^{i+1} a_{i1}A_{i1}\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n\end{aligned}$$

这里最后一步用了和前述命题证明中类似的计算。 □

定理 3.3.6 (Laplace 展开定理)**.** 设 $A = (a_{ij})$ 为 n 阶方阵，A_{ij} 是第 i 行 j 列的余子式。

- 行展开公式：对任意第 i 行

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij}A_{ij}$$

- 列展开公式：对任意第 j 列

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij}A_{ij}$$

证明: 我们在前述命题中证明了第 1 行的展开公式和第 1 列的展开公式。一般的情况证明方法类似，证明细节留作练习。 □

例子 3.3.7. 按照第 1 行展开计算

$$\begin{aligned}\begin{vmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \\ 2 & 1 & -2 \end{vmatrix} =&1 \cdot \begin{vmatrix} -1 & 4 \\ 1 & -2 \end{vmatrix} - 2 \cdot \begin{vmatrix} 0 & 4 \\ 2 & -2 \end{vmatrix} + 3 \cdot \begin{vmatrix} 0 & -1 \\ 2 & 1 \end{vmatrix} \\ =&1 \cdot (2-4) - 2 \cdot (0-8) + 3 \cdot (0-(-2)) = 20\end{aligned}$$

按照第 2 列展开计算

$$\begin{aligned}\begin{vmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \\ 2 & 1 & -2 \end{vmatrix} =&-2 \cdot \begin{vmatrix} 0 & 4 \\ 2 & -2 \end{vmatrix} + (-1) \cdot \begin{vmatrix} 1 & 3 \\ 2 & -2 \end{vmatrix} - 1 \cdot \begin{vmatrix} 1 & 3 \\ 0 & 4 \end{vmatrix} \\ =&(-2) \cdot (0-8) - 1 \cdot (-2-6) - 1 \cdot (4-0) = 20\end{aligned}$$

例子 3.3.8. 我们知道对于上三角方阵

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}$$

这个结论可以很容易地通过展开定理理解。我们把行列式按照第 1 列展开得到

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix}$$

重复这个过程就得到行列式为 $a_{11}a_{22}\cdots a_{nn}$。

3.3.2 转置与行列对称性

定义 3.3.9. 给定 $m \times n$ 矩阵 A，将 A 的行排成列列排成行，得到的 $n \times m$ 矩阵称为 A 的转置，记为 A^T

$$A = \begin{bmatrix} {\color{blue}a_{11}} & {\color{red}a_{12}} & \cdots & a_{1n} \\ {\color{blue}a_{21}} & {\color{red}a_{22}} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ {\color{blue}a_{m1}} & {\color{red}a_{m2}} & \cdots & a_{mn} \end{bmatrix} \qquad A^T = \begin{bmatrix} {\color{blue}a_{11}} & {\color{blue}a_{21}} & \cdots & {\color{blue}a_{m1}} \\ {\color{red}a_{12}} & {\color{red}a_{22}} & \cdots & {\color{red}a_{m2}} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

可以看出，如果 A 是 n 阶方阵，则 A^T 也是 n 阶方阵。

命题 3.3.10. 设 A 是 $m \times k$ 矩阵，B 是 $k \times n$ 矩阵，则

$$(AB)^T = B^T A^T$$

类似地，对于多个矩阵的乘积，$(A_1 A_2 \cdots A_l)^T = A_l^T \cdots A_2^T A_1^T$。

证明：留作练习。 □

定理 3.3.11. n 阶行列式具有转置不变性，即 $\det A = \det A^T$。

证明：当 $n = 1$ 时显然成立，我们对 n 做归纳。对 $\det A$ 作行展开，对 $\det A^T$ 作列展开，得到

$$\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} A_{1j} \qquad \det A^T = \sum_{i=1}^{n} (-1)^{1+i} a_{1i} A_{i1}^T$$

由归纳假设知 $A_{1i} = A_{i1}^T$，因此 $\det A = \det A^T$。 □

转置不变性说明在行列式中行和列的地位是平等的。因此行列式关于行变换的性质对于列变换也成立。例如我们有

- 关于任意一列具有线性性

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

$$\begin{vmatrix} a_{11} & \cdots & \lambda a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & \lambda a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & \lambda a_{nj} & \cdots & a_{nn} \end{vmatrix} = \lambda \begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix}$$

- 交换两列，行列式差一个负号

$$\begin{vmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix} = - \begin{vmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j} & \cdots & a_{2i} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{ni} & \cdots & a_{nn} \end{vmatrix}$$

- 把一列乘以一个数加到另一列，行列式不变

$$\begin{vmatrix} a_{11} & \cdots & a_{1i}+\lambda a_{1j} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i}+\lambda a_{2j} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{ni}+\lambda a_{nj} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix} = - \begin{vmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{ni} & \cdots & a_{nj} & \cdots & a_{nn} \end{vmatrix}$$

3.3.3 一些例子

例子 3.3.12 (Vandermonde 行列式). 如下 n 阶行列式

$$\Delta_n(x_1, \cdots, x_n) := \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}$$

称为 *Vandermonde* 行列式，在数学和物理中有很多应用。

我们用展开定理来计算 $\Delta_n(x_1, \cdots, x_n)$。把第 $(n-1)$ 行乘以 $-x_1$ 加到第 n 行，然后把第 $(n-2)$ 行乘以 $-x_1$ 加到第 $n-1$ 行，$\cdots$，最后把第 1 行乘以 $-x_1$ 加到第 2 行，我们得到

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2-x_1 & \cdots & x_n-x_1 \\ 0 & (x_2-x_1)x_2 & \cdots & (x_n-x_1)x_n \\ \vdots & \vdots & \vdots & \vdots \\ 0 & (x_2-x_1)x_2^{n-2} & \cdots & (x_n-x_1)x_n^{n-2} \end{vmatrix}$$

按照第 1 列展开，我们得到

$$
\begin{aligned}
&=\begin{vmatrix} x_2-x_1 & \cdots & x_n-x_1 \\ (x_2-x_1)x_2 & \cdots & (x_n-x_1)x_n \\ \vdots & \vdots & \vdots \\ (x_2-x_1)x_2^{n-2} & \cdots & (x_n-x_1)x_n^{n-2} \end{vmatrix} \\
&=(x_2-x_1)(x_3-x_1)\cdots(x_n-x_1)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_2 & x_3 & \cdots & x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_2^{n-2} & x_3^{n-2} & \cdots & x_n^{n-2} \end{vmatrix}
\end{aligned}
$$

这里第二个等式是因为提取了每列的共同系数。因此我们得到递推关系

$$\Delta_n(x_1,\cdots,x_n)=\Delta_{n-1}(x_2,\cdots,x_n)\prod_{i=2}^{n}(x_i-x_1).$$

由此得到

$$
\begin{aligned}
&\Delta_n(x_1,\cdots,x_n)=\Delta_{n-1}(x_2,\cdots,x_n)\prod_{i=2}^{n}(x_i-x_1) \\
=&\Delta_{n-2}(x_3,\cdots,x_n)\left(\prod_{i=3}^{n}(x_i-x_2)\right)\left(\prod_{i=2}^{n}(x_i-x_1)\right) \\
=&\cdots=\prod_{i>j}(x_i-x_j)
\end{aligned}
$$

例子 3.3.13. 求如下 n 阶三对角矩阵的行列式

$$\Delta_n=\begin{vmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{vmatrix}_n$$

这里右下角标 $|_n$ 表示是 n 阶方阵。这个矩阵在 *Poisson* 方程边值问题的数值解法中扮演了十分重要的角色。把行列式通过第一行展开，我们得到

$$
\begin{aligned}
&\begin{vmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{vmatrix}_n \\
=&-2\begin{vmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -2 \end{vmatrix}_{n-1}-\begin{vmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & -2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -2 \end{vmatrix}_{n-1}
\end{aligned}
$$

再把上述第二个矩阵按照第一列展开

$$
= -2\begin{vmatrix} -2 & 1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -2 \end{vmatrix}_{n-1} - \begin{vmatrix} -2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & -2 \end{vmatrix}_{n-2}
$$

因此我们得到递推关系

$$
\Delta_n = -2\Delta_{n-1} - \Delta_{n-2}
$$

初始条件为 $\Delta_1 = -2, \Delta_2 = 3$。

我们把递推关系写成 $(\Delta_n + \Delta_{n-1}) = -(\Delta_{n-1} + \Delta_{n-2})$，因此

$$
\begin{aligned}
(\Delta_n + \Delta_{n-1}) &= -(\Delta_{n-1} + \Delta_{n-2}) \\
= \cdots &= (-1)^{n-2}(\Delta_2 + \Delta_1) = (-1)^n
\end{aligned}
$$

代入得

$$
\begin{aligned}
\Delta_n =& (-1)^n - \Delta_{n-1} \\
=& (-1)^n - (-1)^{n-1} + \Delta_{n-2} = \cdots \\
=& (-1)^n - (-1)^{n-1} + \cdots + (-1)^2 + (-1)^{n-1}\Delta_1 \\
=& (n+1)(-1)^n
\end{aligned}
$$

3.4 乘积的行列式

给定两个 n 阶方程

$$
A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}
$$

它们的乘积 AB 也是 n 阶方阵。我们这一节证明如下矩阵乘积的行列式公式：

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

3.4.1 行列式与线性映射

一个 n 阶方阵 A 对应于线性映射 $f : \mathbb{R}^n \to \mathbb{R}^n$

$$
f : \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}
$$

即把 $\mathbb{R}^n$ 中的元素 $\vec{x}$ 写成列向量，我们有

$$f(\vec{x}) = A\vec{x}$$

特别地，f 把 $\mathbb{R}^n$ 的标准基向量映到 A 的列向量

$$\begin{bmatrix} f(\vec{e}_1) & f(\vec{e}_2)\cdots & f(\vec{e}_n) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

例如

$$f(\vec{e}_1) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{n1} \end{bmatrix}$$

回顾我们用矩阵 A 的行向量做外积来定义行列式。由于行列式对于行和列是对称的，我们也可以等价地用列向量做外积来得到行列式。因此矩阵 A 的行列式也可以通过如下表达式给出

$$f(\vec{e}_1) \wedge f(\vec{e}_2) \wedge \cdots \wedge f(\vec{e}_n) = (\det A)\vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$$

这给出了行列式关于线性映射的解释：$\det A$ 是 f 把 $\mathbb{R}^n$ 中的体积元伸缩的倍数（带符号）。

3.4.2 乘积的行列式

设 A, B 是 n 阶行列式，它们分别对应于线性映射

$$f\ ,\ g\ : \mathbb{R}^n \to \mathbb{R}^n$$

我们知道乘积 AB 对应于 f, g 复合的线性映射

$$f \circ g : \mathbb{R}^n \to \mathbb{R}^n$$

通过矩阵 B 的分量 (b_{ij})，我们有

$$\begin{aligned} & g(\vec{e}_1) \wedge g(\vec{e}_2) \wedge \cdots \wedge g(\vec{e}_n) \\ = & \left(\sum_i b_{i1}\vec{e}_i\right) \wedge \left(\sum_i b_{i2}\vec{e}_i\right) \wedge \cdots \wedge \left(\sum_i b_{in}\vec{e}_i\right) \end{aligned}$$

复合 f 并利用 f 的线性性，我们得到

$$\begin{aligned} & f(g(\vec{e}_1)) \wedge f(g(\vec{e}_2)) \wedge \cdots \wedge f(g(\vec{e}_n)) \\ = & \left(\sum_i b_{i1}f(\vec{e}_i)\right) \wedge \left(\sum_i b_{i2}f(\vec{e}_i)\right) \wedge \cdots \wedge \left(\sum_i b_{in}f(\vec{e}_i)\right) \end{aligned}$$

根据行列式的定义，我们可以通过外积运算得到

$$\left(\sum_i b_{i1}\vec{e}_i\right)\wedge\left(\sum_i b_{i2}\vec{e}_i\right)\wedge\cdots\wedge\left(\sum_i b_{in}\vec{e}_i\right)=(\det B)\vec{e}_1\wedge\vec{e}_2\wedge\cdots\wedge\vec{e}_n$$

把 $\vec{e}_i$ 换做 $f(\vec{e}_i)$，通过同样的外积运算可以得到

$$\begin{aligned}&\left(\sum_i b_{i1}f(\vec{e}_i)\right)\wedge\left(\sum_i b_{i2}f(\vec{e}_i)\right)\wedge\cdots\wedge\left(\sum_i b_{in}f(\vec{e}_i)\right)\\&=(\det B)f(\vec{e}_1)\wedge f(\vec{e}_2)\wedge\cdots\wedge f(\vec{e}_n)\end{aligned}$$

因此

$$\begin{aligned}&f(g(\vec{e}_1))\wedge f(g(\vec{e}_2))\wedge\cdots\wedge f(g(\vec{e}_n))\\=&\left(\sum_i b_{i1}f(\vec{e}_i)\right)\wedge\left(\sum_i b_{i2}f(\vec{e}_i)\right)\wedge\cdots\wedge\left(\sum_i b_{in}f(\vec{e}_i)\right)\\=&(\det B)f(\vec{e}_1)\wedge f(\vec{e}_2)\wedge\cdots\wedge f(\vec{e}_n)\end{aligned}$$

由矩阵 A 的行列式的外积定义，

$$f(\vec{e}_1)\wedge f(\vec{e}_2)\wedge\cdots\wedge f(\vec{e}_n)=(\det A)\vec{e}_1\wedge\vec{e}_2\wedge\cdots\wedge\vec{e}_n$$

代入上式得到

$$f(g(\vec{e}_1))\wedge f(g(\vec{e}_2))\wedge\cdots\wedge f(g(\vec{e}_n))=(\det A)(\det B)\vec{e}_1\wedge\vec{e}_2\wedge\cdots\wedge\vec{e}_n$$

另一方面，由于复合 $f\circ g$ 对应于矩阵 AB，由行列式的外积定义我们有

$$f(g(\vec{e}_1))\wedge f(g(\vec{e}_2))\wedge\cdots\wedge f(g(\vec{e}_n))=\det(AB)\vec{e}_1\wedge\vec{e}_2\wedge\cdots\wedge\vec{e}_n$$

对比两方面的计算，我们得到 $\det(AB)=(\det A)(\det B)$。

定理 3.4.1. 设 A,B 是 n 阶方阵，则

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

这个定理的几何解释为：假设 A,B 对应于线性映射 $f,g:\mathbb{R}^n\to\mathbb{R}^n$，故 g 将 n 维体积元伸缩 $\det B$ 倍，f 将 n 维体积元伸缩 $\det A$ 倍。则 f 复合 g 将 n 维体积元伸缩 $(\det A)(\det B)$ 倍。

例子 3.4.2. 矩阵 $A=\begin{bmatrix}1&2\\3&4\end{bmatrix}$，$B=\begin{bmatrix}5&6\\7&8\end{bmatrix}$。直接计算知

$$\det A=-2\quad\det B=-2\quad\Longrightarrow\det A\det B=4$$

另一方面 $AB=\begin{bmatrix}19&22\\43&50\end{bmatrix}$，其行列式为

$$\begin{vmatrix}19&22\\43&50\end{vmatrix}=19\times50-43\times22=950-946=4$$

例子 3.4.3. 方阵 A 称为是幂零矩阵，如果存在一个正整数 N 使得 $A^N=0$，即足够多个 A 乘在一起是零。对于幂零矩阵一定有

$$\det A = 0$$

实际上由 $A^N=0$ 两边取行列式，我们得到

$$\det(A^N) = (\det A)^N = 0 \quad\Longrightarrow\quad \det A = 0$$

例子 3.4.4. 设 $A=(a_{ij})$ 是 n 阶方阵，$\lambda\in\mathbb{R}$。方阵 $\lambda A=(\lambda a_{ij})$ 可以通过矩阵乘积来表示

$$\lambda A = \Lambda A, \qquad \Lambda = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda \end{bmatrix}$$

因此

$$\det(\lambda A) = \det\Lambda \det A = \lambda^n \det A$$

例子 3.4.5. n 阶方阵 A 称为是反对称方阵，如果 $A^T=-A$。假设 A 是奇数阶反对称方阵，则

$$\det A = \det A^T = \det(-A) = (-1)^n \det A = -\det A$$

于是 $\det A=0$。例如

$$\begin{vmatrix} 0 & -2 & 3 \\ 2 & 0 & -1 \\ -3 & 1 & 0 \end{vmatrix} = 0$$

例子 3.4.6. 求如下方阵（称为 n 阶轮回方阵）的行列式

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix}$$

设 $\omega=e^{\frac{2\pi\sqrt{-1}}{n}}$ 是 n 次本原单位根。记多项式

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$$

考虑列向量 $\vec{\beta}_i = \begin{bmatrix} 1 \\ \omega^i \\ \omega^{2i} \\ \vdots \\ \omega^{(n-1)i} \end{bmatrix}$。由 $(\omega^i)^n=1$，我们有

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix} \begin{bmatrix} 1 \\ \omega^i \\ \omega^{2i} \\ \vdots \\ \omega^{(n-1)i} \end{bmatrix} = \begin{bmatrix} f(\omega^i) \\ f(\omega^i)\omega^i \\ f(\omega^i)\omega^{2i} \\ \vdots \\ f(\omega^i)\omega^{(n-1)i} \end{bmatrix} = f(\omega^i) \begin{bmatrix} 1 \\ \omega^i \\ \omega^{2i} \\ \vdots \\ \omega^{(n-1)i} \end{bmatrix}$$

即 $A\vec{\beta}_i = f(\omega^i)\vec{\beta}_i$。比如我们考虑第二行:

$$\begin{aligned}&a_{n-1} + a_0\omega^i + a_1\omega^{2i} + \cdots + a_{n-2}\omega^{(n-1)i}\\=&a_{n-1}\omega^{ni} + a_0\omega^i + a_1\omega^{2i} + \cdots + a_{n-2}\omega^{(n-1)i}\\=&(a_0 + a_1\omega^i + \cdots + a_{n-2}\omega^{(n-2)i} + a_{n-1}\omega^{(n-1)i})\omega^i\\=&f(\omega^i)\omega^i\end{aligned}$$

记矩阵

$$B = \begin{bmatrix}\vec{\beta}_0 & \vec{\beta}_1 & \cdots & \vec{\beta}_{n-1}\end{bmatrix}$$

则

$$AB = \begin{bmatrix}f(\omega^0)\vec{\beta}_0 & f(\omega^1)\vec{\beta}_1 & \cdots & f(\omega^{n-1})\vec{\beta}_{n-1}\end{bmatrix}$$

两边取行列式，我们得到

$$\det A \det B = \det B \prod_{i=0}^{n-1} f(\omega^i)$$

观察到矩阵

$$B = \begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ \omega^0 & \omega & \omega^2 & \cdots & \omega^{n-1}\\ \omega^0 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ \omega^0 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)}\end{bmatrix}$$

由 *Vandermonde* 行列式公式

$$\det B = \prod_{0\le i<j\le n-1} (\omega^j - \omega^i) \neq 0$$

因此公式

$$\det A \det B = \det B \prod_{i=0}^{n-1} f(\omega^i)$$

两边可以消去 $\det B$ 得到轮回矩阵的行列式

$$\det A = \prod_{i=0}^{n-1} f(\omega^i)$$

例子 3.4.7. 集合 $\{1, 2, \cdots, n\}$ 到自身的一个一一映射称为一个 n 元置换。给定 n 元置换 σ，我们可以把它对应到一个 n 阶方阵 A_σ，其矩阵元 a_{ij} 定义为

$$a_{ij} = \begin{cases}1 & \text{如果 } \sigma(i) = j\\ 0 & \text{其它}\end{cases}$$

即 A_σ 的第 i 行只有其第 $\sigma(i)$ 列非 0 且其值是 1。例如 5 元置换

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ & & \Downarrow & & \\ 3 & 5 & 4 & 1 & 2 \end{array}$$

对应的矩阵为

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

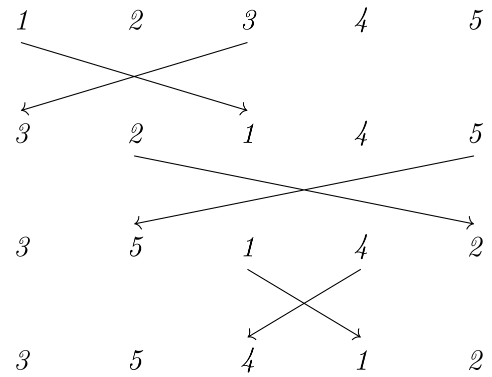
我们可以不断通过交换两行的变换把 A_σ 变为单位矩阵，而每次交换行列式变号。因此

$$\det A_\sigma = \pm 1$$

置换 σ 称为偶置换，如果 $\det A_\sigma = 1$；σ 称为奇置换，如果 $\det A_\sigma = -1$。可以看出，σ 的奇偶性与还原到单位矩阵需要经过的对换次数的奇偶性是一样的。因此这里用行列式定义的置换奇偶性和第3.2.2节定义的奇偶性是等价的，即

$$\operatorname{sign}\sigma = \det A_\sigma$$

例如置换 $\sigma : \{1,2,3,4,5\} \to \{3,5,4,1,2\}$



可以通过 3 个对换复合得到，因此它是奇置换。我们也可以通过其对应的行列式看出

$$\begin{vmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{vmatrix} = -1 \quad \Longrightarrow \quad \operatorname{sign}\sigma = 1$$

两个置换 σ, ρ 的复合 $\sigma \circ \rho$ 依然是一个 n 元置换。容易验证

$$A_{\sigma \circ \rho} = A_\sigma A_\rho$$

$$i \xrightarrow{A_\rho} \rho(i) \xrightarrow{A_\sigma} \sigma(\rho(i))$$

于是 $\det A_{\sigma\circ\rho} = \det A_\sigma \det A_\rho$。因此置换复合满足

$$\begin{gathered}
\text{奇置换} \times \text{奇置换} \to \text{偶置换} \\
\text{奇置换} \times \text{偶置换} \to \text{奇置换} \\
\text{偶置换} \times \text{偶置换} \to \text{偶置换}
\end{gathered}$$

3.5 矩阵的逆

定义 3.5.1. 设 A 是 n 阶方阵，如果存在一个 n 阶方阵 B 使得

$$AB = BA = I_n$$

其中 I_n 是 n 阶单位阵，我们称 A 是可逆的，且 B 是它的逆矩阵，记作 A^{-1}。

注记. 只有方阵（即 $n \times n$ 矩阵）才可能谈逆矩阵。对于一般的矩阵，有一个推广的概念称为“广义逆”（见*5.5.3*节），它没有方阵的逆这样好的结构，但是在解线性方程组方面有一些和逆相近的性质。

3.5.1 可逆矩阵的判别法

假设 n 阶方阵 A 可逆，则对

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

两边取行列式，我们得到 $\det A \det A^{-1} = 1$，因此

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

特别的有 $\det A \neq 0$。因此 A 可逆的一个必要条件是 $\det A \neq 0$。我们下面说明这也是充分条件。

命题 3.5.2. 设 A 是 n 阶方阵，则如下两个条件等价

$$\det A \neq 0 \quad \Longleftrightarrow \quad \operatorname{rank} A = n$$

证明: 我们可以通过初等行变换将 A 变成阶梯形矩阵 A'

$$A' = \begin{vmatrix} a'_{11} & \cdots & \cdots & a'_{1n} \\ 0 & a'_{22} & \cdots & a'_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a'_{nn} \end{vmatrix}$$

由行列式在初等行变换下的变换性质可知

$$\det A \neq 0 \Longleftrightarrow \det A' \neq 0$$

由于 $\det A' = a'_{11}a'_{22}\cdots a'_{nn}$，我们可以看出

$$\det A' \neq 0 \Longleftrightarrow \operatorname{rank} A' = n$$

因此

$$\det A \neq 0 \Longleftrightarrow \det A' \neq 0 \Longleftrightarrow \operatorname{rank} A' = n \Longleftrightarrow \operatorname{rank} A = n$$

□

命题 3.5.3. *设 A 是 n 阶方阵且 $\det A \neq 0$。则对任意列向量 $\vec{b} \in \mathbb{R}^n$，线性方程组*

$$A\vec{x} = \vec{b}$$

有唯一解。

证明: 由 $\det A \neq 0$，我们有 $n - \operatorname{rank} A = 0$，由定理2.3.11得到解的唯一性。我们只需要证明解的存在性。考虑增广矩阵 $\overline{A} = [A\ \vec{b}]$，由

$$n = \operatorname{rank} A \leq \operatorname{rank} \overline{A} = \overline{A} \text{ 的行秩} \leq n$$

知 $\operatorname{rank} \overline{A} = \operatorname{rank} A$。由定理2.3.11知解存在。 □

定理 3.5.4. *A 是 n 阶方阵，则 A 可逆当且仅当 $\det A \neq 0$。*

证明: 我们已经证明了必要性，下面证明充分性。

假设 $\det A \neq 0$。考虑 $\mathbb{R}^n$ 的标准基 $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$。由命题3.5.3，我们可以解出唯一的向量 $\vec{\beta}_i$ 满足

$$A\vec{\beta}_i = \vec{e}_i \qquad i = 1, \cdots, n$$

我们可以把这 n 个方程写成一个矩阵的形式

$$A\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix} = \begin{bmatrix}\vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n\end{bmatrix}$$

记 n 阶方阵 $B = \begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix}$。则上述即为

$$AB = I_n$$

我们下面证明 $BA = I_n$ 也成立。

由于 $AB = I_n$，两边取行列式我们知道 $\det B \neq 0$。把上述对 A 的论述用到 B，我们可以得到一个 n 阶方阵 C 使得 $BC = I_n$。因此

$$A = AI_n = A(BC) = (AB)C = I_nC = C$$

故 $BA = BC = I_n$ 成立。因此 B 是 A 的逆。 □

由上述定理证明我们不难看出，如果

$$AB = I_n$$

则必然有 $BA = I_n$。即“左逆”和“右逆”对方阵是一样的，故统记为矩阵的逆 A^{-1}。

3.5.2 逆矩阵的计算

定理3.5.4的证明过程同时给出了逆矩阵的计算方法：解线性方程组

$$A\vec{\beta}_i = \vec{e}_i$$

则 A 的逆矩阵为

$$A^{-1} = \begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix}$$

由于 $\operatorname{rank} A = n$，我们可以通过初等行变换把 A 变成

$$\begin{vmatrix} 1 & * & \cdots & * \\ 0 & 1 & \cdots & * \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{vmatrix}$$

进而从下往上作行变换消元变为单位阵

$$\begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{vmatrix}$$

把增广矩阵加入如上操作，我们得到

$$\begin{bmatrix} A & \vec{e}_i \end{bmatrix} \overset{\text{行变换}}{\Longrightarrow} \begin{vmatrix} 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \cdots & 0 & b_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & b_n \end{vmatrix}$$

因此线性方程组的解为 $x_1 = b_1, \cdots, x_n = b_n$，即

$$\vec{\beta}_i = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

我们可以用这个方法同时解所有 n 个线性方程组 $A\vec{\beta}_i = \vec{e}_i$。考虑扩大的增广矩阵

$$\begin{bmatrix} A & \vec{e}_1 & \cdots & \vec{e}_n \end{bmatrix} = \begin{bmatrix} A & I_n \end{bmatrix}$$

我们对其作初等行变换将其变为

$$\begin{bmatrix} A & I_n \end{bmatrix} \overset{\text{行变换}}{\Longrightarrow} \begin{bmatrix} I_n & B \end{bmatrix} = \begin{bmatrix} I_n & \vec{\beta}_1 & \cdots & \vec{\beta}_n \end{bmatrix}$$

则 B 的第 i 列即为线性方程组 $A\vec{\beta}_i = \vec{e}_i$ 的解。

定理 3.5.5. *设 A 是 n 阶可逆矩阵。我们对 $n \times 2n$ 矩阵 $\begin{bmatrix} A & I_n \end{bmatrix}$ 作初等行变换将其变为*

$$\begin{bmatrix} A & I_n \end{bmatrix} \overset{\text{行变换}}{\Longrightarrow} \begin{bmatrix} I_n & B \end{bmatrix}$$

则 n 阶矩阵 B 即为 A 的逆。

这个定理给出了逆矩阵的具体计算方法。

例子 3.5.6. 计算矩阵 $\begin{bmatrix}1&2&1\\2&4&5\\1&4&6\end{bmatrix}$ 的逆矩阵。我们对扩展的增广矩阵作初等行变换

$$\begin{bmatrix}1&2&1&1&0&0\\2&4&5&0&1&0\\1&4&6&0&0&1\end{bmatrix}\Longrightarrow\begin{bmatrix}1&2&1&1&0&0\\0&0&3&-2&1&0\\0&2&5&-1&0&1\end{bmatrix}$$
$$\Longrightarrow\begin{bmatrix}1&2&1&1&0&0\\0&2&5&-1&0&1\\0&0&3&-2&1&0\end{bmatrix}\Longrightarrow\begin{bmatrix}1&2&1&1&0&0\\0&1&5/2&-1/2&0&1/2\\0&0&1&-2/3&1/3&0\end{bmatrix}$$
$$\Longrightarrow\begin{bmatrix}1&2&0&5/3&-1/3&0\\0&1&0&7/6&-5/6&1/2\\0&0&1&-2/3&1/3&0\end{bmatrix}\Longrightarrow\begin{bmatrix}1&0&0&-2/3&4/3&-1\\0&1&0&7/6&-5/6&1/2\\0&0&1&-2/3&1/3&0\end{bmatrix}$$

因此我们可以读出逆矩阵

$$\begin{bmatrix}1&2&1\\2&4&5\\1&4&6\end{bmatrix}^{-1}=\begin{bmatrix}-2/3&4/3&-1\\7/6&-5/6&1/2\\-2/3&1/3&0\end{bmatrix}$$

3.5.3 伴随矩阵

逆矩阵也可以通过余子式直接求出。回顾 A 的第 i 行第 j 列余子式 A_{ij} 定义为把 A 的第 i 行和第 j 列去掉后得到的 $n-1$ 阶矩阵的行列式：

$$A_{ij}=\begin{vmatrix}a_{11}&\cdots&\cancel{a_{1j}}&\cdots&a_{1n}\\\cdots&\cdots&\vdots&\cdots&\cdots\\\cancel{a_{i1}}&\cdots&\cancel{a_{ij}}&\cdots&\cancel{a_{in}}\\\cdots&\cdots&\vdots&\cdots&\cdots\\a_{n1}&\cdots&\cancel{a_{nj}}&\cdots&a_{nn}\end{vmatrix}$$

由行列式关于第 i 行的展开公式，我们有

$$\sum_{j=1}^{n}(-1)^{i+j}a_{ij}A_{ij}=\det A$$

实际上我们有更好的性质

命题 3.5.7. n 阶方阵 A 的余子式 A_{ij} 满足

$$\sum_{j=1}^{n}(-1)^{i+j}a_{kj}A_{ij}=\det A\delta_{ki}$$

证明: 对于 $k=i$，这个即是行列式关于第 i 行的展开公式。下面我们考虑 $k\neq i$。表达式

$$\sum_{j=1}^{n}(-1)^{i+j}a_{kj}A_{ij}$$

可以解释为把矩阵 A 的第 i 行替换成第 k 行，然后关于第 i 行作行列式展开。替换后第 i 行的余子式保持不变。即

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ {\color{red}a_{k1}} & {\color{red}a_{k2}} & \cdots & {\color{red}a_{kn}} \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & \cdots & a_{kn} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \overset{\text{第 } i \text{ 行展开}}{=} \sum_{j=1}^{n}(-1)^{i+j}{\color{red}a_{kj}}A_{ij}$$

另一方面，左边矩阵第 i 行和第 k 行相同，故行列式为 0。因此 $\sum_{j=1}^{n}(-1)^{i+j}{\color{red}a_{kj}}A_{ij}=0$。 □

我们可以把公式

$$\sum_{j=1}^{n}(-1)^{i+j}a_{kj}A_{ij}=\det A\delta_{ki}$$

写成矩阵的样子

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} (-1)^{i+1}A_{i1} \\ (-1)^{i+2}A_{i2} \\ \cdots \\ (-1)^{i+n}A_{in} \end{bmatrix}=\det A\ \vec{e}_i$$

假设 A 可逆，即 $\det A\neq 0$，则上述矩阵表达式说明线性方程组 $A\vec{\beta}_i=\vec{e}_i$ 的解即为

$$\vec{\beta}_i=\frac{1}{\det A}\begin{bmatrix} (-1)^{i+1}A_{i1} \\ (-1)^{i+2}A_{i2} \\ \cdots \\ (-1)^{i+n}A_{in} \end{bmatrix}$$

定义 3.5.8. n 阶矩阵 A 的伴随矩阵 A^* 是一个 n 阶矩阵，其 (i,j) 矩阵元为 $(-1)^{i+j}A_{ji}$，即

$$A^*=\begin{bmatrix} (-1)^{1+1}A_{11} & (-1)^{1+2}A_{21} & \cdots & (-1)^{1+n}A_{n1} \\ (-1)^{2+1}A_{12} & (-1)^{2+2}A_{22} & \cdots & (-1)^{2+n}A_{n2} \\ \cdots & \cdots & \cdots & \\ (-1)^{n+1}A_{1n} & (-1)^{n+2}A_{2n} & \cdots & (-1)^{n+n}A_{nn} \end{bmatrix}$$

我们上述讨论实际上证明了如下命题。

定理 3.5.9. 设 A 是 n 阶可逆方阵，则其逆矩阵为

$$A^{-1}=\frac{1}{\det A}A^*$$

例子 3.5.10. 2 阶方阵 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的伴随矩阵为

$$A^* = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

因此

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

例子 3.5.11. 矩阵 $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 1 & 4 & 6 \end{bmatrix}$ 的伴随矩阵为

$$A^* = \begin{bmatrix} 4 & -8 & 6 \\ -7 & 5 & -3 \\ 4 & -2 & 0 \end{bmatrix}$$

计算知 $\det A = -6$，因此

$$A^{-1} = \begin{bmatrix} -2/3 & 4/3 & -1 \\ 7/6 & -5/6 & 1/2 \\ -2/3 & 1/3 & 0 \end{bmatrix}$$

3.5.4 Cramer 法则

设 A 是 n 阶方阵且 $\det A \neq 0$。由命题3.5.3，对任意列向量 $\vec{b} \in \mathbb{R}^n$，线性方程组

$$A\vec{x} = \vec{b}$$

有唯一解。实际上，由定理3.5.4我们知道 A 可逆，上述方程两边乘以 A^{-1} 可以得到解为

$$\vec{x} = A^{-1}\vec{b}$$

由定理3.5.9，我们可以把上述解具体表示为

$$\vec{x} = \frac{1}{\det A} A^* \vec{b}$$

写成分量有

$$x_i = \frac{1}{\det A} \sum_{j=1}^{n} (-1)^{i+j} A_{ji} b_j$$

考虑矩阵

$$A_i = \begin{bmatrix} a_{11} & \cdots & a_{1,i-1} & b_1 & a_{1,i+1} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2,i-1} & b_2 & a_{2,i+1} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{n,i-1} & b_n & a_{n,i+1} & \cdots & a_{nn} \end{bmatrix}$$

A_i 是把 A 的第 i 列替换成 $\vec{b}$ 构成的方阵。观察到 A_i 的第 k 行第 i 列余子式与 A 的第 k 行第 i 列余子式相同。通过第 i 列展开计算 A_i 的行列式，我们发现

$$\det A_i = \sum_{j=1}^{n} (-1)^{i+j} A_{ji} b_j$$

由此我们证明了如下结论

定理 3.5.12 (Cramer 法则)**.** 设 A 是 n 阶方阵且 $\det A \neq 0$。则线性方程组 $A\vec{x} = \vec{b}$ 的解为

$$\vec{x} = \begin{bmatrix} \det A_1 / \det A \\ \det A_2 / \det A \\ \vdots \\ \det A_n / \det A \end{bmatrix}$$

其中 A_i 是把 A 的第 i 列替换成 $\vec{b}$ 构成的方阵。

3.6 习题

1. 计算如下行列式

$$(1)\ \begin{vmatrix}1&2&3\\0&1&4\\5&6&0\end{vmatrix}\qquad (2)\ \begin{vmatrix}1&2&3&4\\0&1&0&2\\1&0&1&1\\2&3&0&1\end{vmatrix}\qquad (3)\ \begin{vmatrix}1&2&3&4\\2&3&4&1\\3&4&1&2\\4&1&2&3\end{vmatrix}$$

2. 用行列式的组合公式计算如下 n 阶行列式

$$(1)\ \begin{vmatrix}a&b&0&\cdots&\cdots&0\\0&a&b&0&\cdots&0\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&0&\cdots&\cdots&a&b\\b&0&\cdots&\cdots&0&a\end{vmatrix}\qquad (2)\ \begin{vmatrix}x&-1&0&\cdots&\cdots&0\\0&x&-1&0&\cdots&0\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&0&\cdots&\cdots&x&-1\\a_n&a_{n-1}&\cdots&\cdots&a_2&x+a_1\end{vmatrix}$$

3. (a) 说明对 $m\times n$ 矩阵 A 作的三种初等行变换均可以写成对 A 左乘一个 m 阶方阵 P

$$A\to PA$$

写出三种初等行变换对应的矩阵 P 并计算其行列式 $\det P$。

(b) 和初等行变换类似，我们可以定义矩阵的三种初等列变换为

i. 交换两列: 将矩阵的两列交换位置

ii. 将矩阵中的一列乘以一个非零常数

iii. 将矩阵的一列乘以一个数加到另一列

说明对 $m\times n$ 矩阵 A 作的三种初等列变换均可以写成对 A 右乘一个 n 阶方阵 Q

$$A\to AQ$$

写出三种初等列变换对应的矩阵 Q 并计算其行列式 $\det Q$。

4. 设 A 是如下分块上三角形式

$$A=\begin{bmatrix}B&C\\0&D\end{bmatrix}$$

这里 B 是 $k\times k$ 矩阵，C 是 $k\times(n-k)$ 矩阵，D 是 $(n-k)\times(n-k)$ 矩阵。证明 $\det A=\det B\det D$。

5. 设 A 是 $m\times k$ 矩阵，B 是 $k\times n$ 矩阵。证明

$$(AB)^T=B^TA^T$$

类似地，对于多个矩阵的乘积，$(A_1A_2\cdots A_l)^T=A_l^T\cdots A_2^TA_1^T$。

6. 利用 Laplace 展开定理计算

$$(1)\quad \begin{vmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 0 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 3 & 0 & 1 \end{vmatrix} \qquad\qquad (2)\quad \begin{vmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 & 6 \end{vmatrix}$$

7. Fibonacci 数为：$F_1 = 1, F_2 = 2, F_n = F_{n-1} + F_{n-2}$ $(n \geq 3)$。证明 Fibonacci 数可以写成

$$F_n = \begin{vmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -1 & 1 & 1 \\ 0 & 0 & \cdots & \cdots & -1 & 1 \end{vmatrix}$$

8. 设 n 阶方阵 A 的矩阵元 $a_{ij} = a_{ij}(x)$ 都是变量 x 的可微函数。证明

$$\frac{d \det A}{dx} = \sum_{1 \leq i,j \leq n} (-1)^{i+j} \frac{da_{ij}}{dx} A_{ij}$$

其中 A_{ij} 是 A 的第 i 行第 j 列余子式。

9. 设 A 是 $m \times n$ 矩阵，B 是 $n \times m$ 矩阵。考虑 $(m+n) \times (m+n)$ 阶方阵 $\begin{bmatrix} I_m & A \\ B & I_n \end{bmatrix}$，这里 I_m 是 m 阶单位阵，I_n 是 n 阶单位阵。

(a) 证明

$$\begin{bmatrix} I_m & 0 \\ -B & I_n \end{bmatrix} \begin{bmatrix} I_m & A \\ B & I_n \end{bmatrix} = \begin{bmatrix} I_m & A \\ 0 & I_n - BA \end{bmatrix}$$

$$\begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix} \begin{bmatrix} I_m & A \\ B & I_n \end{bmatrix} = \begin{bmatrix} I_m - AB & 0 \\ B & I_n \end{bmatrix}$$

(b) 证明

$$\det(I_m - AB) = \det(I_n - BA)$$

(c) 计算行列式

$$\begin{vmatrix} 1 + a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & 1 + a_2 b_2 & \cdots & a_2 b_n \\ \cdots & \cdots & \cdots & \cdots \\ a_n b_1 & a_n b_2 & \cdots & 1 + a_n b_n \end{vmatrix}$$

10. 求下列矩阵的逆

(1) $\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 1 \\ 1 & 0 & 1 \end{bmatrix}$　　(2) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 4 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 3 & 0 & 1 \end{bmatrix}$

(3) $\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}$　　(4) $\begin{bmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{bmatrix}_{n\times n}$

11. 设 A 是 n 阶幂零方阵，即存在 N 使得 $A^N = 0$。证明 $I_n - A$ 可逆。

12. 设 A 是上三角方阵

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix}$$

且 A 可逆。证明 A^{-1} 也是上三角方阵。

13. 设 A 是 $m \times n$ 矩阵。取出 A 的第 $i_1, i_2, \cdots, i_p$ 行和第 $j_1, j_2, \cdots, j_p$ 列交叉位置上的元素构成的 p 阶方阵的行列式

$$\begin{vmatrix} a_{i_1j_1} & a_{i_1j_2} & \cdots & a_{i_1j_p} \\ a_{i_2j_1} & a_{i_2j_2} & \cdots & a_{i_2j_p} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i_pj_1} & a_{i_pj_2} & \cdots & a_{i_pj_p} \end{vmatrix}$$

称为 A 的一个 p 阶行列子式。记 A 的所有非零行列子式的最高阶数为 $p(A)$。证明

$$p(A) = \operatorname{rank} A$$

14. 设 $m \times n$ 矩阵 A 的秩为 r。证明存在可逆 m 阶方阵 P 和可逆 n 阶方阵 Q 使得

$$A = P \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q$$

$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ 是 $m \times n$ 阶矩阵，左上角 I_r 是 r 阶单位阵。这个结论称为矩阵的相抵标准型。

第四章　特征值理论

4.1　特征值与特征向量

4.1.1　特征值与特征向量

我们知道 n 阶方阵 A 对应于线性映射 $f: k^n \to k^n$（这里数域 $k = \mathbb{R}$ 或者 $\mathbb{C}$）

$$f: \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \mapsto \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

即把列向量 $\vec{x}$ 映射为

$$f(\vec{x}) = A\vec{x}$$

为了更好地刻画这个映射的结构和几何性质，我们可以寻找一些特殊的向量使得这个线性映射把这些特殊向量映射到相对简单的形式。其中最自然的一类向量，就是使得线性映射把它映射到它自己的一个倍数。这类向量称为特征向量，其伸缩的倍数称为特征值。

定义 4.1.1. 设 A 是一个 n 阶方阵。如果存在一个非零向量 $\vec{x}$ 和一个数 λ 使得

$$A\vec{x} = \lambda\vec{x}$$

则 λ 称为 A 的一个特征值，$\vec{x}$ 称为（属于特征值 λ 的）一个特征向量。

这个概念类似地也可以对线性变换来定义。

定义 4.1.2. 设 $f: V \to V$ 是线性空间 V 到自身的一个线性映射。如果存在一个非零向量 $\vec{x} \in V$ 和一个数 λ 使得

$$f(\vec{x}) = \lambda\vec{x}$$

则 λ 称为 f 的一个特征值，$\vec{x}$ 称为（属于特征值 λ 的）一个特征向量。

例子 4.1.3. 考虑矩阵 $A = \begin{bmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 4 & 2 \\ 1 & 4 & 2 & 0 \\ 4 & 2 & 0 & 1 \end{bmatrix}$。观察到

$$\begin{bmatrix} 2 & 0 & 1 & 4 \\ 0 & 1 & 4 & 2 \\ 1 & 4 & 2 & 0 \\ 4 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

因此 7 是 A 的一个特征值，$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ 是一个特征向量。

并不是任何 λ 都能成为给定方阵 A 的特征值。例如考虑 $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$，由于

$$A\vec{x} = 3\vec{x}$$

对任意向量 $\vec{x}$ 成立，因此 A 的特征值只有 $\lambda = 3$。

A 也可能有多个不同的特征值。例如考虑 $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$，容易看出

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

因此 $\lambda = 2$ 和 $\lambda = 3$ 均为 A 的特征值。

下面这个命题是特征值的等价描述。

命题 4.1.4. λ 是 n 阶方阵 A 的特征值当且仅当齐次线性方程组

$$(\lambda I_n - A)\vec{x} = 0$$

具有非零解。

证明: 假设如上齐次线性方程组有非零解 $\vec{x}$，则

$$A\vec{x} = \lambda I_n \vec{x} = \lambda \vec{x}$$

因此 λ 是特征值，$\vec{x}$ 是属于 λ 的特征向量。反之亦然。 □

上述命题给出了特征值的具体判别方法。

命题 4.1.5. λ 是 n 阶方阵 A 的特征值当且仅当

$$\det(\lambda I_n - A) = 0$$

证明: 齐次线性方程组 $(\lambda I_n - A)\vec{x} = 0$ 具有非零解当且仅当 $\det(\lambda I_n - A) = 0$。 □

4.1.2 特征多项式

定义 4.1.6. 设 A 是 n 阶方阵。则 n 次多项式

$$\varphi(\lambda) := \det(\lambda I_n - A)$$

称为 A 的特征多项式。

由命题4.1.5知，A 的特征值即为特征多项式的根。

例子 4.1.7. $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$，特征多项式为

$$\begin{aligned}\varphi(\lambda) = \begin{vmatrix} \lambda-2 & -1 & 0 \\ -1 & \lambda-2 & -1 \\ 0 & -1 & \lambda-2 \end{vmatrix} &= (\lambda-2)\begin{vmatrix} \lambda-2 & -1 \\ -1 & \lambda-2 \end{vmatrix} + \begin{vmatrix} -1 & -1 \\ 0 & \lambda-2 \end{vmatrix} \\ &= (\lambda-2)((\lambda-2)^2-2) = (\lambda-2)(\lambda-2-\sqrt{2})(\lambda-2+\sqrt{2})\end{aligned}$$

因此 A 有 3 个特征值 $2, 2+\sqrt{2}, 2-\sqrt{2}$。

对于 n 阶方阵 $A = (a_{ij})$，其特征多项式

$$\varphi(\lambda) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} = \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n$$

是首项为 λ^n 的 n 次多项式。由代数学基本定理，我们可以把 $\varphi(\lambda)$ 分解为

$$\varphi(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

这里 $\lambda_1, \cdots, \lambda_n$ 是 $\varphi(\lambda)$ 的所有复数根（可以有重根）。

命题 4.1.8. 设 $\lambda_1, \cdots, \lambda_n$ 是 A 的特征多项式 $\varphi(\lambda)$ 的所有复数根（可以有重根），则

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n$$

证明: 由 $\varphi(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$，知 $\varphi(0) = (-1)^n \lambda_1 \lambda_2 \cdots \lambda_n$。另一方面由定义

$$\varphi(0) = \det(-A) = (-1)^n \det A$$

对比两个表达式即得命题。 □

命题 4.1.9. 方阵 A 可逆当且仅当 A 的特征值全不为零。

证明: 由

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n$$

即得命题。 □

例子 4.1.10. 假设 $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$ 是上三角方阵，则

$$\varphi(\lambda) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & \cdots & -a_{1n} \\ 0 & \lambda - a_{22} & -a_{23} & \cdots & -a_{2n} \\ 0 & 0 & \lambda - a_{33} & \cdots & -a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda - a_{nn} \end{vmatrix} = \prod_{i=1}^{n} (\lambda - a_{ii})$$

此时 A 的特征值即为所有对角元的值。

定义 4.1.11. 方阵 $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ 的迹 *(trace)* 定义为

$$\operatorname{Tr} A := a_{11} + a_{22} + \cdots + a_{nn}$$

即为 A 的所有对角元的和。

迹的一个重要性质是如下关于乘法的对称性。

命题 4.1.12. 设 A 是 $m \times n$ 矩阵，B 是 $n \times m$ 矩阵。则

$$\operatorname{Tr} AB = \operatorname{Tr} BA$$

证明：留作练习。 □

命题 4.1.13. 设 n 阶方阵 A 的特征多项式为 $\varphi(\lambda) = \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n$。则

$$\alpha_1 = -\operatorname{Tr} A$$

证明:

$$\varphi(\lambda) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

由行列式组合公式，λ^{n-1} 项只能来源于 $\prod\limits_{i=1}^{n} (\lambda - a_{ii})$ 的贡献，由此易得 $\alpha_1 = -\sum\limits_{i=1}^{n} a_{ii} = -\operatorname{Tr} A$。 □

因此 A 的特征多项式可以表达为

$$\varphi(\lambda) = \lambda^n - (\operatorname{Tr} A)\lambda^{n-1} + \cdots + (-1)^n \det A$$

实际上，特征多项式中其他的系数 α_k 也可以通过 A 的 k 阶余子式来得到，我们这里不作讨论。

命题 4.1.14. 设 $\lambda_1, \cdots, \lambda_n$ 是 A 的特征多项式 $\varphi(\lambda)$ 的所有复数根（可以有重根），则

$$\operatorname{Tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

证明: $\varphi(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$ 的 λ^{n-1} 项系数为 $-\lambda_1 - \lambda_2 - \cdots - \lambda_n$。对比公式 $\varphi(\lambda) = \lambda^n - (\operatorname{Tr} A)\lambda^{n-1} + \cdots$ 即得命题。 □

例子 4.1.15. 设 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 是 2 阶方阵

$$\operatorname{Tr} A = a + d \qquad \det A = ad - bc$$

因此 A 的特征多项式为

$$\varphi(\lambda) = \lambda^2 - (\operatorname{Tr} A)\lambda + \det A = \lambda^2 - (a+d)\lambda + (ad - bc)$$

这个公式也可以通过计算行列式直接验证。

4.1.3 Cayley-Hamilton 定理

定理 4.1.16 (Cayley-Hamilton)**.** 设 n 阶方阵 A 的特征多项式为 $\varphi(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n$。则矩阵 $\varphi(A) := A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1}A + \alpha_n I_n$ 是零矩阵，即满足 $\varphi(A) = 0$。

证明: 令 $B = \lambda I_n - A$，则 $\varphi(\lambda) = \det B$。观察到 B 的每个余子式 B_{ij} 至多是 λ 的 $(n-1)$ 次多项式。因此 B 的伴随矩阵 B^* 可以写成

$$B^* = \lambda^{n-1}C_1 + \lambda^{n-2}C_2 + \cdots + C_n$$

其中 C_i 是不含 λ 的 n 阶方阵。由伴随矩阵的性质

$$BB^* = \det(B)I_n = \varphi(\lambda)I_n$$

代入上述表达式，我们得到等式

$$(\lambda I_n - A)(\lambda^{n-1}C_1 + \lambda^{n-2}C_2 + \cdots + C_n) = \varphi(\lambda)I_n$$

比较两边的 λ 多项式的系数矩阵，我们得到

$$\begin{aligned} C_1 &= I_n \\ C_2 - AC_1 &= \alpha_1 I_n \\ &\vdots \\ C_n - AC_{n-1} &= \alpha_{n-1} I_n \\ -AC_n &= \alpha_n I_n \end{aligned}$$

把这组方程依次左乘以 $A^n, A^{n-1}, \cdots, A, I_n$ 相加，即得到 $\varphi(A) = 0$。 □

定义 4.1.17. 设 A 是 n 阶方阵。一个非零多项式 $f(\lambda)$ 称为 A 的一个化零多项式，如果 $f(A) = 0$ 是零矩阵。

Cayley-Hamilton 定理表明，每个 n 阶方阵 A 都会满足一个多项式的代数关系：它的特征多项式是一个化零多项式。需要指出的是，特征多项式不一定是 A 的最小化零多项式：可能存在次数比 n 小的化零多项式。

例如对于 $A=\begin{bmatrix}2&0&0\\0&2&0\\0&0&3\end{bmatrix}$，其特征多项式为 $\varphi(\lambda)=(\lambda-2)^2(\lambda-3)$。另一方面

$$A-2I=\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}\qquad A-3I=\begin{bmatrix}-1&0&0\\0&-1&0\\0&0&0\end{bmatrix}$$

容易看出 $(A-2I)(A-3I)=0$。因此 $f(\lambda)=(\lambda-2)(\lambda-3)$ 是 A 的一个化零多项式，它的次数比特征多项式 $\varphi(\lambda)$ 的次数低。

Cayley-Hamilton 定理有很多重要的应用。我们这里举一个例子，它给出了一种矩阵逆的代数表达方法。设 A 是 n 阶方阵，其特征多项式为

$$\varphi(\lambda)=\lambda^n+\alpha_1\lambda^{n-1}+\cdots+\alpha_{n-1}\lambda+\alpha_n$$

假设 A 可逆，则 $\alpha_n=(-1)^n\det A\neq 0$。由 Cayley-Hamilton 定理，我们有等式

$$A^n+\alpha_1A^{n-1}+\cdots+\alpha_{n-1}A+\alpha_nI_n=0$$

把 α_nI_n 移到右边，两边同除以 $-\alpha_n$

$$-\frac{1}{\alpha_n}(A^{n-1}+\alpha_1A^{n-2}+\cdots+\alpha_{n-1}I_n)A=I_n$$

由此得

$$A^{-1}=-\frac{1}{\alpha_n}(A^{n-1}+\alpha_1A^{n-2}+\cdots+\alpha_{n-1}I_n)$$

特别地，A^{-1} 可以通过 A 的一个多项式来表达。

例子 4.1.18. 设 $A=\begin{bmatrix}2&-2&4\\2&3&2\\-1&1&-1\end{bmatrix}$。它的特征多项式为

$$\varphi(\lambda)=\det(\lambda I_3-A)=\lambda^3-4\lambda^2+7\lambda-10$$

因此

$$A^{-1}=\frac{1}{10}(A^2-4A+7I_3)=\frac{1}{10}\begin{bmatrix}-5&2&-16\\0&2&4\\5&0&10\end{bmatrix}$$

4.2 相似变换

4.2.1 相似与对角化

定义 4.2.1. 两个 n 阶方阵 A 与 B 称为是相似的，如果存在 n 阶可逆方阵 P 使得

$$A=PBP^{-1}$$

容易验证，相似关系构成一个等价关系，即满足

- 自反性：A 与 A 相似
- 对称性：若 A 与 B 相似，则 B 与 A 相似
- 传递性：若 A 与 B 相似，B 与 C 相似，则 A 与 C 相似

命题 4.2.2. 如果 A 与 B 相似，$f(x)$ 是任意一个多项式。则 $f(A)$ 与 $f(B)$ 相似。

证明: 设 $A = PBP^{-1}$。则对任意正整数 m

$$A^m = (PBP^{-1})(PBP^{-1})\cdots(PBP^{-1}) = PB^mP^{-1}$$

由此易知 $f(A) = P\ f(B)\ P^{-1}$。 □

命题 4.2.3. 如果 A 与 B 相似，则它们具有相同的特征多项式。特别地，我们有

$$\operatorname{Tr} A = \operatorname{Tr} B \qquad \det A = \det B$$

证明: 设 $A = PBP^{-1}$。则

$$\begin{aligned}
\det(\lambda I - A) &= \det(\lambda I - PBP^{-1}) = \det(P(\lambda I - B)P^{-1}) \\
&= \det(P)\det(\lambda I - B)\det(P^{-1}) \\
&= \det(\lambda I - B)
\end{aligned}$$

□

定义 4.2.4. n 阶方阵 A 称为可对角化，如果 A 相似于一个对角阵 $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$。这里

$$\operatorname{diag}(\lambda_1, \cdots, \lambda_n) := \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

如果 A 相似于 $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$，则 A 的特征多项式和这个对角阵相同

$$\varphi(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

因此这个对角阵的元素即为 A 的特征值。

假设 A 可对角化，$A = P\operatorname{diag}(\lambda_1, \cdots, \lambda_n)P^{-1}$，即

$$AP = P\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$$

记 P 的列向量为 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$

$$P = \begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix}$$

由 P 可逆知 $\operatorname{rank} P = n$，因此 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 构成一组基。

等式 $A=P\operatorname{diag}(\lambda_1,\cdots,\lambda_n)P^{-1}$ 可以写成

$$A\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix}=\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \cdots & \cdots & \cdots & 0\\ 0 & \cdots & 0 & \lambda_n\end{bmatrix}$$

即

$$A\vec{\beta}_i=\lambda_i\vec{\beta}_i,\qquad i=1,\cdots,n$$

这说明 $\vec{\beta}_i$ 是属于 λ_i 的特征向量。由上述讨论，我们证明了如下结论

命题 4.2.5. n 阶方阵 A 可对角化当且仅当存在一组基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 使得每个 $\vec{\beta}_i$ 都是 A 的特征向量。

在这个情况下，我们把任意向量 $\vec{x}$ 通过基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 来线性表达：$\vec{x}=\sum\limits_i c_i\vec{\beta}_i$。则矩阵 A 乘在向量 $\vec{x}$ 上很容易计算出

$$A\vec{x}=\sum_i c_i\lambda_i\vec{\beta}_i$$

例子 4.2.6. 并不是每个方阵都可以对角化，例如

$$A=\begin{bmatrix}\lambda_0 & 1 & 0 & \cdots & 0\\ 0 & \lambda_0 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & \lambda_0 & 1\\ 0 & 0 & \cdots & 0 & \lambda_0\end{bmatrix}$$

A 的特征值只有 λ_0，我们考虑它的特征向量。齐次线性方程组

$$(\lambda_0 I_n-A)\vec{x}=0\Longleftrightarrow\begin{bmatrix}0 & -1 & 0 & \cdots & 0\\ 0 & 0 & -1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & 0 & -1\\ 0 & 0 & \cdots & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_{n-1}\\ x_n\end{bmatrix}=0$$

的解为 $x_1=c,x_2=x_3=\cdots=x_n=0$。因此 λ_0 的特征向量只有一个线性无关的元素，它不可能构成 n 维空间 $(n>1)$ 的一组基。这说明方阵 A 不可对角化。

命题 4.2.7. 如果数域 k 上 n 阶方阵 A 具有 n 个互不相同的特征值在 k 中，则 A 可以对角化。

证明：设 A 的特征值为 $\lambda_1,\cdots,\lambda_n$。由假设这些特征值互不相同。设 $\vec{\beta}_i$ 是属于 λ_i 的一个特征向量。我们下面证明 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 线性无关。由此知 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 构成 n 维空间的一组基，因此 A 可对角化。

假设有系数 $c_1,\cdots,c_n$ 使得

$$c_1\vec{\beta}_1+c_2\vec{\beta}_2+\cdots+c_n\vec{\beta}_n=0$$

两边依次乘以 $A, A^2, \cdots, A^{n-1}$，我们得到

$$\begin{cases} c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \cdots + c_n\vec{\beta}_n = 0 \\ \lambda_1 c_1\vec{\beta}_1 + \lambda_2 c_2\vec{\beta}_2 + \cdots + \lambda_n c_n\vec{\beta}_n = 0 \\ \qquad\qquad \cdots\cdots \\ \lambda_1^{n-1} c_1\vec{\beta}_1 + \lambda_2^{n-1} c_2\vec{\beta}_2 + \cdots + \lambda_n^{n-1} c_n\vec{\beta}_n = 0 \end{cases}$$

我们可以把这组方程写成矩阵的形式

$$\begin{bmatrix} c_1\vec{\beta}_1 & c_2\vec{\beta}_2 & \cdots & c_n\vec{\beta}_n \end{bmatrix} \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{bmatrix} = 0$$

记 $P = \begin{bmatrix} c_1\vec{\beta}_1 & c_2\vec{\beta}_2 & \cdots & c_n\vec{\beta}_n \end{bmatrix}$，$Q = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{bmatrix}$，则上式为

$$PQ = 0$$

由 Vandermonde 行列式

$$\det Q = \prod_{i>j} (\lambda_i - \lambda_j) \neq 0$$

故 Q 可逆。因此 $PQ = 0 \Longrightarrow P = 0$，即 P 的每列 $c_i\vec{\beta}_i = 0$。由于 $\vec{\beta}_i$ 都是非零向量，我们得出 $c_i = 0, i = 1, \cdots, n$。这证明了 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 是线性无关的向量。 □

例子 4.2.8. 设 $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 1 & 1 \end{bmatrix}$，计算 A^{100}。

我们首先计算 A 的特征多项式

$$\begin{vmatrix} \lambda - 2 & -1 & 0 \\ 0 & \lambda - 3 & 0 \\ 0 & -1 & \lambda - 1 \end{vmatrix} = (\lambda - 2)(\lambda - 3)(\lambda - 1)$$

A 有 3 个不同的特征值 $2, 3, 1$，因此可对角化。

我们计算特征值 $\lambda_1=2,\lambda_2=3,\lambda_3=1$ 对应的特征向量 $\vec{\beta}_1,\vec{\beta}_2,\vec{\beta}_3$

$$(\lambda_1-A)\vec{\beta}_1=\begin{bmatrix}0&-1&0\\0&-1&0\\0&-1&1\end{bmatrix}\vec{\beta}_1=0 \qquad \text{取} \qquad \vec{\beta}_1=\begin{bmatrix}1\\0\\0\end{bmatrix}$$

$$(\lambda_2-A)\vec{\beta}_2=\begin{bmatrix}1&-1&0\\0&0&0\\0&-1&2\end{bmatrix}\vec{\beta}_2=0 \qquad \text{取} \qquad \vec{\beta}_2=\begin{bmatrix}2\\2\\1\end{bmatrix}$$

$$(\lambda_3-A)\vec{\beta}_3=\begin{bmatrix}-1&-1&0\\0&-2&0\\0&-1&0\end{bmatrix}\vec{\beta}_3=0 \qquad \text{取} \qquad \vec{\beta}_3=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

记矩阵

$$P=\begin{bmatrix}\vec{\beta}_1&\vec{\beta}_2&\vec{\beta}_3\end{bmatrix}=\begin{bmatrix}1&2&0\\0&2&0\\0&1&1\end{bmatrix}$$

容易计算

$$P^{-1}=\begin{bmatrix}1&-1&0\\0&1/2&0\\0&-1/2&1\end{bmatrix}$$

我们得到相似变化 $A=P\begin{bmatrix}2&0&0\\0&3&0\\0&0&1\end{bmatrix}P^{-1}$。因此

$$\begin{aligned}A^{100}=&P\begin{bmatrix}2&0&0\\0&3&0\\0&0&1\end{bmatrix}^{100}P^{-1}=P\begin{bmatrix}2^{100}&0&0\\0&3^{100}&0\\0&0&1\end{bmatrix}P^{-1}\\=&\begin{bmatrix}1&2&0\\0&2&0\\0&1&1\end{bmatrix}\begin{bmatrix}2^{100}&0&0\\0&3^{100}&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&-1&0\\0&1/2&0\\0&-1/2&1\end{bmatrix}\\=&\begin{bmatrix}2^{100}&3^{100}-2^{100}&0\\0&3^{100}&0\\0&(3^{100}-1)/2&1\end{bmatrix}\end{aligned}$$

4.2.2 相似变换的几何含义

设 $f:k^n\to k^n$ 是一个线性映射。我们知道可以把 f 对应于一个 n 阶方阵 A。具体而言，取 k^n 的标准基 $\{\vec{e}_1,\cdots,\vec{e}_n\}$，计算

$$f(\vec{e}_j)=\sum_i a_{ij}\vec{e}_i$$

则 $A=(a_{ij})$。实际上，除了标准基，我们也可以取另外一组基作类似的构造。

定义 4.2.9. 设 $f : k^n \to k^n$ 是一个线性映射。给定 k^n 的一组基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$，设

$$f(\vec{\alpha}_j) = \sum_i a_{ij}\vec{\alpha}_i$$

则方阵 $A = (a_{ij})$ 称为 f 在这组基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 下的表示矩阵。

我们把 f 在基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 下的表示矩阵写成矩阵关系

$$\begin{bmatrix} f(\vec{\alpha}_1) & f(\vec{\alpha}_2) & \cdots & f(\vec{\alpha}_n) \end{bmatrix} = \begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

如果我们取 k^n 中两组不同的基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 和 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$。设 f 在基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 下的表示矩阵为 A，在基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 下的表示矩阵为 B。那么矩阵 A 和 B 是什么关系？

由于 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 是一组基，我们可以把向量 $\vec{\beta}_j$ 在这组基下作线性展开

$$\vec{\beta}_j = \sum_i p_{ij}\vec{\alpha}_i$$

写成矩阵的形式

$$\begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix} = \begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

这个 n 阶方阵 $P = (p_{ij})$ 称为从基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 到基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 的过渡矩阵。

命题 4.2.10. 一组基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 到另一组基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 的过渡矩阵 P 是可逆矩阵，且 P^{-1} 是基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 到基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 的过渡矩阵。

证明：设 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 到基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 的过渡矩阵为 Q。则

$$\begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix} = \begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} P$$
$$\begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} = \begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix} Q$$

把两个等式复合，我们得到

$$\begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} = \begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} PQ$$

即

$$\begin{bmatrix} \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n \end{bmatrix} (I_n - PQ) = 0$$

由 $\{\vec{\alpha}_i\}$ 的线性无关性，我们得到 $PQ = I_n$。 □

定理 4.2.11. 设线性映射 $f: k^n \to k^n$ 在基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 下的表示矩阵为 A, 在基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 下的表示矩阵为 B。设基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 到基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 的过渡矩阵为 P。则

$$A = PBP^{-1}$$

证明：过渡矩阵关系为

$$\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix} = \begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix} P$$

由于 f 是线性映射，两边作用映射 f 给出

$$\begin{bmatrix}f(\vec{\beta}_1) & f(\vec{\beta}_2) & \cdots & f(\vec{\beta}_n)\end{bmatrix} = \begin{bmatrix}f(\vec{\alpha}_1) & f(\vec{\alpha}_2) & \cdots & f(\vec{\alpha}_n)\end{bmatrix} P$$

代入 f 在基下表示矩阵的关系

$$\begin{bmatrix}f(\vec{\alpha}_1) & f(\vec{\alpha}_2) & \cdots & f(\vec{\alpha}_n)\end{bmatrix} = \begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix} A$$
$$\begin{bmatrix}f(\vec{\beta}_1) & f(\vec{\beta}_2) & \cdots & f(\vec{\beta}_n)\end{bmatrix} = \begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix} B$$

我们得到

$$\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix} B = \begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix} AP$$

再次代入过渡矩阵关系，上述等式变为

$$\begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix} PB = \begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix} AP$$

即

$$\begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix} (PB - AP) = 0$$

由 $\{\vec{\alpha}_i\}$ 的线性无关性，我们得到 $PB = AP$。 □

注记. 这个命题给出了相似变换的几何含义：相似变换是同一个线性映射 $f: k^n \to k^n$ 在不同基下表示矩阵之间的变换。

4.3 特征子空间与根子空间

4.3.1 特征子空间

定义 4.3.1. 设 A 是数域 k 上的 n 阶方阵，λ_0 是 A 一个特征值。定义向量空间 k^n 的子集

$$V_{\lambda_0} := \{\vec{x} \in k^n | A\vec{x} = \lambda_0 \vec{x}\}$$

称为特征值 λ_0 的特征子空间。

换言之，V_{λ_0} 是由所有属于 λ_0 的特征向量以及零向量构成的集合。

命题 4.3.2. 特征子空间 V_{λ_0} 是一个线性空间。

证明: 设 $\vec{x},\vec{y}\in V_{\lambda_0}$，$\alpha$ 是一个数。则

$$\begin{aligned} A(\vec{x}+\vec{y}) =& A\vec{x}+A\vec{y}=\lambda_0\vec{x}+\lambda_0\vec{y}=\lambda_0(\vec{x}+\vec{y}) \\ A(\alpha\vec{x}) =& \alpha A\vec{x}=\alpha\lambda_0\vec{x}=\lambda_0(\alpha\vec{x}) \end{aligned}$$

即 $\vec{x}+\vec{y}\in V_{\lambda_0},\alpha\vec{x}\in V_{\lambda_0}$。由此命题得证。 □

例子 4.3.3. 考虑 n 阶方阵 $A=\begin{bmatrix} \lambda_0 & 1 & 0 & \cdots & 0 \\ 0 & \lambda_0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_0 & 1 \\ 0 & 0 & \cdots & 0 & \lambda_0 \end{bmatrix}$。它只有一个特征值 λ_0，其特征子空间 V_{λ_0} 是 1 维的，由向量 $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$ 张成。

定义 4.3.4. 设 $\varphi(\lambda)$ 是 A 的特征多项式，λ_0 是 A 的特征值。

- 记 $\varphi(\lambda)=(\lambda-\lambda_0)^m g(\lambda)$，其中 $g(\lambda_0)\neq 0$。称 m 为特征值 λ_0 的代数重数，即 λ_0 作为特征多项式的根的重数
- $\dim V_{\lambda_0}$ 称为特征值 λ_0 的几何重数

例子 4.3.5. n 阶方阵 $A=\begin{bmatrix} \lambda_0 & 1 & 0 & \cdots & 0 \\ 0 & \lambda_0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_0 & 1 \\ 0 & 0 & \cdots & 0 & \lambda_0 \end{bmatrix}$ 的特征多项式

$$\varphi(\lambda)=(\lambda-\lambda_0)^n$$

因此特征值 λ_0 的代数重数为 n，几何重数为 1。

命题 4.3.6. 特征值的几何重数 $\leq$ 代数重数。

证明: 设 λ_0 是 n 阶方阵 A 的特征值，其几何维数是 r，即 $\dim V_{\lambda_0}=r$。设 $\{\vec{\beta}_1,\cdots,\vec{\beta}_r\}$ 是 V_{λ_0} 的一组基。我们可以把它扩展为整个 n 维空间的一组基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_r,\vec{\beta}_{r+1},\cdots,\vec{\beta}_n\}$。把这组基作为列向量，记 P 为方阵

$$P=\begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix}$$

由于 $\{\vec{\beta}_1,\cdots,\vec{\beta}_r\}$ 是 λ_0 的特征向量，我们有

$$\begin{aligned} A\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix} &= \begin{bmatrix}A\vec{\beta}_1 & A\vec{\beta}_2 & \cdots & A\vec{\beta}_n\end{bmatrix} \\ &= \begin{bmatrix}\lambda_0\vec{\beta}_1 & \cdots & \lambda_0\vec{\beta}_r & A\vec{\beta}_{r+1}\cdots & A\vec{\beta}_n\end{bmatrix} \\ &= \begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix}\begin{bmatrix}\lambda_0 I_r & C \\ 0 & D\end{bmatrix}\end{aligned}$$

这里 C 是 $r\times(n-r)$ 矩阵，D 是 $(n-r)\times(n-r)$ 矩阵，它们表示 $A\vec{\beta}_{r+1},\cdots,A\vec{\beta}_n$ 按照基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 展开的系数。我们得到相似变换关系 $A=P\begin{bmatrix}\lambda_0 I_r & C \\ 0 & D\end{bmatrix}P^{-1}$ 。因此

$$\det(\lambda I_n-A)=\det(\lambda I_n-\begin{bmatrix}\lambda_0 I_r & C \\ 0 & D\end{bmatrix})=\begin{vmatrix}(\lambda-\lambda_0)I_r & -C \\ 0 & \lambda I_{n-r}-D\end{vmatrix}=(\lambda-\lambda_0)^r\det(\lambda I_{n-r}-D)$$

这说明 λ_0 的代数重数 $\geq r$。 □

4.3.2 子空间的和与直和

定义 4.3.7. 设 V 是数域 k 上的线性空间，$V_1,\cdots,V_m$ 是 V 的线性子空间。定义 V 的子集

$$V_1+\cdots+V_m=\{\vec{\alpha}_1+\cdots+\vec{\alpha}_m|\vec{\alpha}_i\in V_i\}\subset V$$

称为子空间 $V_1,\cdots,V_m$ 的和。

命题 4.3.8. V 的线性子空间的和 $V_1+\cdots+V_m$ 是 V 的线性子空间。

证明: 设 $\vec{x},\vec{y}\in V_1+\cdots+V_m$。则存在 $\vec{\alpha}_i,\vec{\beta}_i\in V_i$ 使得

$$\vec{x}=\vec{\alpha}_1+\cdots+\vec{\alpha}_m \qquad \vec{y}=\vec{\beta}_1+\cdots+\vec{\beta}_m$$

则

$$\vec{x}+\vec{y}=(\vec{\alpha}_1+\vec{\beta}_1)+\cdots+(\vec{\alpha}_m+\vec{\beta}_m)$$

由于 V_i 是线性子空间，$(\vec{\alpha}_i+\vec{\beta}_i)\in V_i$，因此 $\vec{x}+\vec{y}\in V_1+\cdots+V_m$。同理可证保乘法。 □

例子 4.3.9. 设 $V=\mathbb{R}^4$，$\{e_i\}$ 是标准基。考虑子空间 $V_1=\mathrm{Span}\{e_1,e_2\}$，$V_2=\mathrm{Span}\{e_2,e_3\}$，$V_3=\mathrm{Span}\{e_3,e_4\}$。则

$$V_1+V_2=\mathrm{Span}\{e_1,e_2,e_3\} \qquad V_1+V_3=V$$

定义 4.3.10. 设 V 是数域 k 上的线性空间，$V_1,\cdots,V_m$ 是 V 的线性子空间。如果 $V_1+\cdots+V_m$ 中的任意向量 $\vec{\alpha}$，存在唯一的向量 $\alpha_i\in V_i$ 使得

$$\vec{\alpha}=\vec{\alpha}_1+\cdots+\vec{\alpha}_m$$

则此时 $V_1+\cdots+V_m$ 称为子空间 $V_1,\cdots,V_m$ 的直和，记为 $V_1\oplus\cdots\oplus V_m$。

命题 4.3.11. V 的子空间 $V_1, \cdots, V_m$ 的和是直和的充分必要条件是如下性质成立:

$$\text{若 } \vec{\alpha}_i \in V_i \quad \vec{\alpha}_1 + \cdots + \vec{\alpha}_m = 0 \quad \Longrightarrow \quad \vec{\alpha}_1 = \cdots = \vec{\alpha}_m = 0$$

即只需要验证零向量表达的唯一性(零向量只能通过各个子空间的零向量相加得到)。

证明: 假设 V 的子空间 $V_1, \cdots, V_m$ 的和是直和,则零向量只能唯一地表达为零向量相加,因此命题所述性质成立。

反之,假设命题所述性质成立。设 $\vec{\alpha}$ 是 $V_1 + \cdots + V_m$ 中的向量,有两种表达方式

$$\vec{\alpha} = \vec{\alpha}_1 + \cdots + \vec{\alpha}_m \qquad \vec{\alpha} = \vec{\beta}_1 + \cdots + \vec{\beta}_m \qquad \text{这里} \quad \vec{\alpha}_i, \vec{\beta}_i \in V_i$$

两式相减得到

$$(\vec{\alpha}_1 - \vec{\beta}_1) + \cdots + (\vec{\alpha}_m - \vec{\beta}_m) = 0$$

由零向量表达的唯一性知 $\vec{\alpha}_1 = \vec{\beta}_1, \cdots, \vec{\alpha}_m = \vec{\beta}_m$。由此证得表达的唯一性。 □

例子 4.3.12. 设 $V = \mathbb{R}^4$, $\{\vec{e}_i\}$ 是标准基。考虑子空间

$$V_1 = \mathrm{Span}\{\vec{e}_1, \vec{e}_2\} \qquad V_2 = \mathrm{Span}\{\vec{e}_2, \vec{e}_3\} \qquad V_3 = \mathrm{Span}\{\vec{e}_3, \vec{e}_4\}$$

- $V_1 + V_2$ 不是直和,例如我们有两种不一样的方法来表达相加

  $$\vec{e}_2 = \vec{e}_2 + 0 \qquad \vec{e}_2 = 0 + \vec{e}_2$$

  其中第一种表达里 $\vec{e}_2$ 看作是 V_1 中的向量,第二种表达里 $\vec{e}_2$ 看作是 V_2 中的向量。

- $V_1 + V_3 = V_1 \oplus V_3$ 是直和,可以由 $\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4$ 的线性无关性证明。

命题 4.3.13. V 的子空间 $V_1, \cdots, V_m$ 的和是直和的充分必要条件是

$$\dim(V_1 + \cdots + V_m) = \dim V_1 + \cdots + \dim V_m$$

此时每个 V_i 选一组基合在一起构成 $V_1 \oplus \cdots \oplus V_m$ 的一组基。

证明: 设 $\{\vec{\beta}_1^{(i)}, \cdots, \vec{\beta}_{t_i}^{(i)}\}$ 是 V_i 的一组基。由定义,

$$V_1 + \cdots + V_m = \mathrm{Span}\{\vec{\beta}_1^{(1)}, \cdots, \vec{\beta}_{t_1}^{(1)}, \cdots, \vec{\beta}_1^{(m)}, \cdots, \vec{\beta}_{t_m}^{(m)}\}$$

假设 $V_1 + \cdots + V_m = V_1 \oplus \cdots \oplus V_m$ 是直和。我们说明 $\{\vec{\beta}_1^{(1)}, \cdots, \vec{\beta}_{t_1}^{(1)}, \cdots, \vec{\beta}_1^{(m)}, \cdots, \vec{\beta}_{t_m}^{(m)}\}$ 是线性无关的,因此构成 $V_1 + \cdots + V_m$ 的一组基。由此即得 $\dim(V_1 + \cdots + V_m) = \dim V_1 + \cdots + \dim V_m$。

若

$$c_1^{(1)} \vec{\beta}_1^{(1)} + \cdots + c_{t_1}^{(1)} \vec{\beta}_{t_1}^{(1)} + \cdots + c_1^{(m)} \vec{\beta}_1^{(m)} + \cdots + c_{t_m}^{(m)} \vec{\beta}_{t_m}^{(m)} = 0$$

记 $\vec{\beta}_i = c_1^{(i)} \vec{\beta}_1^{(i)} + \cdots + c_{t_i}^{(i)} \vec{\beta}_{t_i}^{(i)}$。则 $\vec{\beta}_i \in V_i$ 且

$$0 = \vec{\beta}_1 + \cdots + \vec{\beta}_m$$

另一方面零向量显然可以写成零向量的和。由直和的表达唯一性,对每个 i 我们有 $\vec{\beta}_i = 0$。又由于 $\vec{\beta}_1^{(i)}, \cdots, \vec{\beta}_{t_i}^{(i)}$ 是 V_i 的一组基,$c_1^{(i)} = \cdots = c_{t_i}^{(i)} = 0$。这证明了 $\{\vec{\beta}_1^{(1)}, \cdots, \vec{\beta}_{t_1}^{(1)}, \cdots, \vec{\beta}_1^{(m)}, \cdots, \vec{\beta}_{t_m}^{(m)}\}$ 是线性无关的。

反之，假设 $\dim(V_1+\cdots+V_m)=\dim V_1+\cdots+\dim V_m$ 成立。由

$$V_1+\cdots+V_m=\mathrm{Span}\{\vec{\beta}_1^{(1)},\cdots,\vec{\beta}_{t_1}^{(1)},\cdots,\vec{\beta}_1^{(m)},\cdots,\vec{\beta}_{t_m}^{(m)}\}$$

知 $\{\vec{\beta}_1^{(1)},\cdots,\vec{\beta}_{t_1}^{(1)},\cdots,\vec{\beta}_1^{(m)},\cdots,\vec{\beta}_{t_m}^{(m)}\}$ 是线性无关的。因此 $V_1+\cdots+V_m$ 中的任一向量可以唯一地写成 $\{\vec{\beta}_1^{(1)},\cdots,\vec{\beta}_{t_1}^{(1)},\cdots,\vec{\beta}_1^{(m)},\cdots,\vec{\beta}_{t_m}^{(m)}\}$ 的线性组合。由此易证直和性。 □

命题 4.3.14. 设 $\lambda_1,\cdots,\lambda_s$ 是 A 的 s 个互不相同的特征值。则它们的特征子空间的和是直和。

$$V_{\lambda_1}+\cdots+V_{\lambda_s}=V_{\lambda_1}\oplus\cdots\oplus V_{\lambda_s}$$

证明：设 $\{\vec{\beta}_1^{(j)},\cdots,\vec{\beta}_{t_j}^{(j)}\}$ 是 V_{λ_j} 中的一组基。我们只需证明这些向量 $\{\vec{\beta}_1^{(1)},\cdots,\vec{\beta}_{t_1}^{(1)},\cdots,\vec{\beta}_1^{(s)},\cdots,\vec{\beta}_{t_s}^{(s)}\}$ 合在一起构成的向量集是线性无关的。假设有线性组合满足

$$c_1^{(1)}\vec{\beta}_1^{(1)}+\cdots+c_{t_1}^{(1)}\vec{\beta}_{t_1}^{(1)}+\cdots+c_1^{(s)}\vec{\beta}_1^{(s)}+\cdots+c_{t_s}^{(s)}\vec{\beta}_{t_s}^{(s)}=0$$

记

$$\vec{\beta}^{(j)}=c_1^{(j)}\vec{\beta}_1^{(j)}+\cdots+c_{t_j}^{(j)}\vec{\beta}_{t_j}^{(j)}$$

则 $\vec{\beta}^{(j)}\in V_{\lambda_j}$ 并且

$$\vec{\beta}^{(1)}+\vec{\beta}^{(2)}+\cdots+\vec{\beta}^{(s)}=0$$

两边依次乘以 $A,A^2,\cdots,A^{s-1}$，我们得到

$$\begin{cases}\vec{\beta}^{(1)}+\vec{\beta}^{(2)}+\cdots+\vec{\beta}^{(s)}=0\\ \lambda_1\vec{\beta}^{(1)}+\lambda_2\vec{\beta}^{(2)}+\cdots+\lambda_s\vec{\beta}^{(s)}=0\\ \qquad\qquad\cdots\\ \lambda_1^{s-1}\vec{\beta}^{(1)}+\lambda_2^{s-1}\vec{\beta}^{(2)}+\cdots+\lambda_s^{s-1}\vec{\beta}^{(s)}=0\end{cases}$$

把这组方程写成矩阵的形式

$$\begin{bmatrix}\vec{\beta}^{(1)} & \vec{\beta}^{(2)} & \cdots & \vec{\beta}^{(s)}\end{bmatrix}\begin{bmatrix}1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{s-1}\\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{s-1}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & \lambda_s & \lambda_s^2 & \cdots & \lambda_s^{s-1}\end{bmatrix}=0$$

记 s 阶矩阵 $P=\begin{bmatrix}1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{s-1}\\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{s-1}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & \lambda_s & \lambda_s^2 & \cdots & \lambda_s^{s-1}\end{bmatrix}$，由 Vandermonde 行列式知 P 可逆。上式两边同时从右边乘以 P^{-1}，我们得到

$$\begin{bmatrix}\vec{\beta}^{(1)} & \vec{\beta}^{(2)} & \cdots & \vec{\beta}^{(s)}\end{bmatrix}=0$$

即每个向量 $\vec{\beta}^{(j)}=c_1^{(j)}\vec{\beta}_1^{(j)}+\cdots+c_{t_j}^{(j)}\vec{\beta}_{t_j}^{(j)}=0$。由 $\{\vec{\beta}_1^{(j)},\cdots,\vec{\beta}_{t_j}^{(j)}\}$ 的线性无关性，我们得到 $c_1^{(j)}=\cdots=c_{t_j}^{(j)}=0$ 对每个 $j=1,\cdots,s$ 成立。 □

这个命题说明，不同特征值的特征向量之间是相互线性无关的。特别地，如果 n 阶方阵 A 有 n 个互不相同的特征值，则其对应的特征向量构成 n 维空间一组基，即得到命题4.2.7。

例子 4.3.15. 设 n 阶方阵 A 满足 $A^2 = A$（称为幂等方阵）。则 A 相似于对角阵 $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$，这里 $r = \operatorname{rank} A$。实际上，考虑如下两个线性子空间

$$V_1 = \{\vec{x} \in k^n | A\vec{x} = \vec{x}\} \quad V_0 = \{\vec{x} \in k^n | A\vec{x} = 0\}$$

如果 i 是 A 的特征值（$i = 0, 1$），则 V_i 是特征子空间，否则 $V_i = \{0\}$。

对任意向量 $\vec{x} \in \mathbb{R}^n$，我们把它分解为

$$\vec{x} = A\vec{x} + (I_n - A)\vec{x}$$

由 A 的幂等性

$$A(A\vec{x}) = A\vec{x}, \quad A(I_n - A)\vec{x} = 0 \quad \Longrightarrow \quad A\vec{x} \in V_1, \quad (I_n - A)\vec{x} \in V_0$$

这个分解说明

$$V_1 + V_0 = k^n$$

设 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_r\}$ 是 V_1 的一组基，$\{\vec{\beta}_1, \cdots, \vec{\beta}_s\}$ 是 V_0 的一组基。由命题*4.3.14*知，$\{\vec{\alpha}_1, \cdots, \vec{\alpha}_r, \vec{\beta}_1, \cdots, \vec{\beta}_s\}$ 构成 k^n 的一组基。由命题*4.2.5*，A 相似于对角阵 $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$。

4.3.3 根子空间

这一节我们假设数域 k 上 n 阶方阵 A 的所有特征值都在数域 k 中（若 $k = \mathbb{R}$，即要求特征值均为实数）。A 的特征多项式可以分解为

$$\varphi(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_s)^{m_s}$$

这里 $\lambda_1, \cdots, \lambda_s$ 是互不相同的特征值，m_i 是 λ_i 的代数重数。

设 V_{λ_i} 是 λ_i 的特征子空间。我们知道

$$m_1 + m_2 + \cdots + m_s = n$$

由命题4.3.6

$$\dim V_{\lambda_i} \leq m_i, \quad i = 1, \cdots, s$$

如果 $\dim V_{\lambda_i} = m_i$ 对每个 i 成立，那么每个 V_{λ_i} 选一组基，合起来构成整个 k^n 的一组基。若 $\dim V_{\lambda_i} < m_i$，我们似乎“缺失”了某些向量。这些“缺失”的向量可以通过根子空间找回来。

定义 4.3.16. 设 λ_0 是 n 阶方阵 A 的一个特征值。定义 λ_0 的根子空间为

$$R_{\lambda_0} = \{\vec{x} \mid \text{存在某个 } m > 0 \text{ 使得} (\lambda_0 I_n - A)^m \vec{x} = 0\}$$

即 $\vec{x} \in R_{\lambda_0}$ 当且仅当对 $\vec{x}$ 乘以 $\lambda_0 I_n - A$ 足够多次后会变成 0。

λ_0 的特征子空间可以刻画为

$$V_{\lambda_0} = \{\vec{x} | (\lambda_0 I_n - A)\vec{x} = 0\}$$

即对 $\vec{x}$ 乘以 $\lambda_0 I_n - A$ 一次变成 0。因此 $V_{\lambda_0} \subset R_{\lambda_0}$。

命题 4.3.17. 根子空间 R_{λ_0} 是线性空间。

证明: 设 $\vec{x},\vec{y}\in R_{\lambda_0}$。则存在 $m_1,m_2>0$ 使得

$$(\lambda_0 I_n - A)^{m_1}\vec{x}=0 \qquad (\lambda_0 I_n - A)^{m_2}\vec{y}=0$$

取 $m=\max\{m_1,m_2\}$，则

$$(\lambda_0 I_n - A)^m(\vec{x}+\vec{y})=(\lambda_0 I_n - A)^m\vec{x}+(\lambda_0 I_n - A)^m\vec{y}=0$$

即 $\vec{x}+\vec{y}\in R_{\lambda_0}$。易知 R_{λ_0} 也保数乘。 □

例子 4.3.18. n 阶方阵 $A=\begin{bmatrix}\lambda_0 & 1 & 0 & \cdots & 0\\ 0 & \lambda_0 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & \lambda_0 & 1\\ 0 & 0 & \cdots & 0 & \lambda_0\end{bmatrix}$。易知

$$(\lambda_0 I_n - A)^n=0$$

因此任意向量 $\vec{x}$ 都属于根子空间 R_{λ_0} 中。可以看出根子空间的确找回来了“缺失”的向量。

定义 4.3.19. 设 A 是 n 阶矩阵。线性子空间 $V\subset k^n$ 称为是 A 的不变子空间如果

$$\text{对任意 } \vec{x}\in V,\quad A\vec{x}\in V$$

命题 4.3.20. 根子空间 R_{λ_0} 是 A 的不变子空间。

证明: 设 $\vec{x}\in R_{\lambda_0}$，$(\lambda_0 I_n - A)^m\vec{x}=0$。则

$$(\lambda_0 I_n - A)^m(A\vec{x})=A((\lambda_0 I_n - A)^m\vec{x})=0$$

故 $A\vec{x}\in R_{\lambda_0}$。 □

由于 R_{λ_0} 是 A 的不变子空间，把 A 的作用限制在 R_{λ_0} 上得到一个线性映射

$$A: R_{\lambda_0}\to R_{\lambda_0}$$

命题 4.3.21. 线性映射 $A: R_{\lambda_0}\to R_{\lambda_0}$ 在 R_{λ_0} 的任一组基下的表示矩阵的特征值只有 λ_0。

证明: 考虑这个线性映射在 R_{λ_0} 的一组基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_t\}$ 下的表示矩阵 s 阶方阵 $B=(b_{ij})$，即

$$A\vec{\beta}_j=\sum_i \vec{\beta}_i b_{ij}$$

写成矩阵的形式

$$A\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_t\end{bmatrix}=\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_t\end{bmatrix}B$$

由根子空间定义，存在充分大 $m>0$ 使得 $(\lambda_0 I_n - A)^m\vec{\beta}_i=\vec{0}$，对 $i=1,\cdots,t$ 成立。由

$$(\lambda_0 I_t - A)^m\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_t\end{bmatrix}=\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_t\end{bmatrix}(\lambda_0 I_t - B)^m$$

左边是零，因此
$$\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_t\end{bmatrix}(\lambda_0 I_t - B)^m = \mathbf{0}$$
由 $\{\vec{\beta}_1, \cdots, \vec{\beta}_t\}$ 的线性无关性，得 $(\lambda_0 I_t - B)^m = 0$。由此易知（或参考命题4.4.3）$\lambda_0 I_t - B$ 的特征值只有 0, 即 B 的特征值只有 λ_0。 □

命题 4.3.22. 设 $\lambda_1, \cdots, \lambda_s$ 是互不相同的特征值。则它们的根子空间的和是直和。

$$R_{\lambda_1} + \cdots + R_{\lambda_s} = R_{\lambda_1} \oplus \cdots \oplus R_{\lambda_s}$$

证明：设 $\vec{\beta}_i \in R_{\lambda_i}$ 并且
$$\vec{\beta}_1 + \vec{\beta}_2 + \cdots + \vec{\beta}_s = 0$$
我们只需证明一定有 $\vec{\beta}_1 = \cdots = \vec{\beta}_s = 0$。

假设 $\vec{\beta}_1 \neq 0$。则存在正整数 m 使得 $(\lambda_1 I_n - A)^{m-1}\vec{\beta}_1 \neq 0$，$(\lambda_1 I_n - A)^m \vec{\beta}_1 = 0$。记 $\vec{\gamma}_1 = (\lambda_1 I_n - A)^{m-1}\vec{\beta}_1 \neq 0$，则
$$(\lambda_1 I_n - A)\vec{\gamma}_1 = 0$$
即 $\vec{\gamma}_1$ 是 λ_1 的特征向量。则对任意多项式 $f(x)$

$$f(A)\vec{\gamma}_1 = f(\lambda_1)\vec{\gamma}_1$$

取 N 充分大使得
$$(\lambda_i I_n - A)^N \vec{\beta}_i = 0, \qquad i = 1, \cdots, s$$
令 $f(x) = (\lambda_2 - x)^N (\lambda_3 - x)^N \cdots (\lambda_s - x)^N$。则

$$f(A)\vec{\gamma}_1 = f(\lambda_1)\vec{\gamma}_1, \quad f(A)\vec{\beta}_2 = \cdots = f(A)\vec{\beta}_s = 0$$

其中 $f(\lambda_1) = (\lambda_2 - \lambda_1)^N (\lambda_3 - \lambda_1)^N \cdots (\lambda_s - \lambda_1)^N \neq 0$。在等式

$$\vec{\beta}_1 + \vec{\beta}_2 + \cdots + \vec{\beta}_s = 0$$

两边乘以 $(\lambda_1 I_n - A)^{m-1} f(A)$，我们得到

$$f(\lambda_1)\vec{\gamma}_1 = 0$$

由于 $f(\lambda_1) \neq 0 \Longrightarrow \vec{\gamma}_1 = 0$，与假设矛盾。因此 $\vec{\beta}_1 = 0$。同理可证 $\vec{\beta}_1 = \cdots = \vec{\beta}_s = 0$。 □

命题 4.3.23. 设 $\lambda_1, \cdots, \lambda_s$ 是 n 阶方阵 A 所有互不相同的特征值。则

$$k^n = R_{\lambda_1} \oplus \cdots \oplus R_{\lambda_s}$$

证明：由命题4.3.22，我们只需证明任意向量 $\vec{x} \in k^n$ 可以分解为

$$\vec{x} = \vec{\beta}_1 + \cdots + \vec{\beta}_s, \qquad \vec{\beta}_i \in R_{\lambda_i}$$

n 阶方阵 A 的特征多项式为

$$\varphi(\lambda) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_s)^{m_s}$$

这里 $\lambda_1,\cdots,\lambda_s$ 是互不相同的特征值。考虑如下的多项式

$$\varphi_i(\lambda)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_{i-1})^{m_{i-1}}(\lambda-\lambda_{i+1})^{m_{i+1}}\cdots(\lambda-\lambda_s)^{m_s}$$

即把 $\varphi(\lambda)$ 中去掉 $(\lambda-\lambda_i)^{m_i}$ 因子得到的多项式。多项式 $\varphi_1(\lambda),\cdots,\varphi_s(\lambda)$ 的公因子只有常数函数，因此存在多项式 $h_i(\lambda)$ 使得

$$\varphi_1(\lambda)h_1(\lambda)+\cdots+\varphi_s(\lambda)h_s(\lambda)=1$$

代入矩阵 A，我们得到

$$\varphi_1(A)h_1(A)+\cdots+\varphi_s(A)h_s(A)=I_n$$

因此

$$\vec{x}=(\varphi_1(A)h_1(A)+\cdots+\varphi_s(A)h_s(A))\vec{x}=\vec{\beta}_1+\cdots+\vec{\beta}_s$$

这里 $\vec{\beta}_i=\varphi_i(A)h_i(A)\vec{x}$。由 Cayley-Hamilton 定理，

$$(\lambda_iI_n-A)^{m_i}\varphi_i(A)=\varphi(A)=0$$

因此 $(\lambda_iI_n-A)^{m_i}\vec{\beta}_i=\varphi(A)h_i(A)\vec{x}=0$，即 $\vec{\beta}_i\in R_{\lambda_i}$。 $\square$

设 $\lambda_1,\cdots,\lambda_s$ 是 A 的所有互不相同的特征值。设 $\{\vec{\beta}_1^{(j)},\cdots,\vec{\beta}_{t_j}^{(j)}\}$ 是 R_{λ_j} 的一组基。由命题4.3.23，向量组

$$\{\vec{\beta}_1^{(1)},\cdots,\vec{\beta}_{t_1}^{(1)},\cdots,\vec{\beta}_1^{(s)},\cdots,\vec{\beta}_{t_s}^{(s)}\}$$

构成 k^n 的一组基。由命题4.3.20，R_{λ_j} 是 A 的不变子空间，因此 $A\vec{\beta}_i^{(j)}$ 是 $\{\vec{\beta}_1^{(j)},\cdots,\vec{\beta}_{t_j}^{(j)}\}$ 的线性组合。写成矩阵的形式

$$\begin{aligned}&A\begin{bmatrix}\vec{\beta}_1^{(1)} & \cdots & \vec{\beta}_{t_1}^{(1)} & \cdots & \vec{\beta}_1^{(s)} & \cdots & \vec{\beta}_{t_s}^{(s)}\end{bmatrix}\\ =&\begin{bmatrix}\vec{\beta}_1^{(1)} & \cdots & \vec{\beta}_{t_1}^{(1)} & \cdots & \vec{\beta}_1^{(s)} & \cdots & \vec{\beta}_{t_s}^{(s)}\end{bmatrix}\begin{bmatrix}B_1 & 0 & \cdots & 0\\ 0 & B_2 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & \cdots & 0 & B_s\end{bmatrix}\end{aligned}$$

这里 B_j 是 t_j 阶方阵，其特征多项式为 $(\lambda-\lambda_j)^{t_j}$（命题4.3.21）。因此 A 的特征多项式为

$$\det(\lambda I_n-A)=\det(\lambda-B_1)\cdots\det(\lambda-B_s)=(\lambda-\lambda_1)^{t_1}\cdots(\lambda-\lambda_s)^{t_s}$$

对比

$$\det(\lambda I_n-A)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$$

我们得到

$$t_j=m_j$$

由此我们证明了如下命题。

命题 4.3.24. $\dim R_{\lambda_i}$ 等于 λ_i 的代数重数。

总结如上，我们有 $V_{\lambda_i} \subset R_{\lambda_i}$

$$\dim V_{\lambda_i} = \text{几何重数} \quad \leq \quad \dim R_{\lambda_i} = \text{代数重数}$$

根子空间给出直和分解

$$k^n = R_{\lambda_1} \oplus R_{\lambda_2} \oplus \cdots \oplus R_{\lambda_s}$$

.

4.4 Jordan 标准型

4.4.1 幂零变换与循环子空间

定义 4.4.1. 一个线性映射 $f : V \to V$ 称为幂零变换，如果存在正整数 m 使得 f 的 m 次复合 $f^m = 0$。类似地，一个 n 阶方阵 A 称为幂零矩阵，如果存在正整数 m 使得 $A^m = 0$。使得 $f^m = 0$（或 $A^m = 0$）成立的最小正整数 m 称为 f（或 A）的幂零指数，此时也称 f（或 A）为 m 次幂零变换（或幂零方阵）。

命题 4.4.2. 设 A 是一个 n 阶幂零矩阵，则 $I_n - A$ 可逆。

证明: 设 $A^m = 0$，则

$$(I_n - A)(I_n + A + \cdots + A^{m-1}) = I_n$$

因此 $(I_n - A)^{-1} = I_n + A + \cdots + A^{m-1}$。 □

命题 4.4.3. 一个 n 阶方阵 A 是幂零矩阵当且仅当 A 只有零特征值，即特征多项式为 λ^n。

证明：假设 A 是幂零矩阵。设 $\lambda_0 \neq 0$，则 $\lambda_0^{-1}A$ 也是幂零矩阵。由命题4.4.2$(I_n - \lambda_0^{-1}A)$ 可逆，$\lambda_0 I_n - A = \lambda_0(I_n - \lambda_0^{-1}A)$ 可逆，故 λ_0 不是 A 的特征值。由 λ_0 的任意性，A 的特征值只有 0。

反之假设 A 只有零特征值，即特征多项式为 λ^n。由 Cayley-Hamilton 定理知 $A^n = 0$。

□

定义 4.4.4. 设 $f : V \to V$ 是一个线性变换（或 A 是一个 n 阶矩阵），$\vec{x}$ 是一个非零向量。

$$C_{\vec{x}} = \mathrm{Span}\{\vec{x}, f(\vec{x}), \cdots, f^m(\vec{x}), \cdots\} \quad (\quad \text{或} \quad C_{\vec{x}} = \mathrm{Span}\{\vec{x}, A\vec{x}, \cdots, A^m\vec{x}, \cdots\} \quad)$$

称为在 A 作用下由 $\vec{x}$ 生成的循环子空间。

由 Cayley-Hamilton 定理知，$\{\vec{x}, A\vec{x}, \cdots, A^{n-1}\vec{x}\}$ 已经可以张成整个 $C_{\vec{x}}$。

命题 4.4.5. 设 A 是幂零方阵，r 是满足 $A^r\vec{x} = 0$ 的最小正整数。则 $\dim C_{\vec{x}} = r$，并且

$$\{\vec{x}, A\vec{x}, \cdots, A^{r-1}\vec{x}\}$$

构成这个在 A 作用下由 $\vec{x}$ 生成的循环子空间 $C_{\vec{x}}$ 的一组基。

证明：设 r 是满足 $A^r\vec{x}=0$ 的最小正整数。我们只需证明 $\{\vec{x}, A\vec{x}, \cdots, A^{r-1}\vec{x}\}$ 是线性无关的。

假设

$$c_0\vec{x}+c_1A\vec{x}+\cdots+c_{r-1}A^{r-1}\vec{x}=0$$

并且系数 c_i 不全为零。假设 $c_0=\cdots=c_{k-1}=0, c_k\neq 0$，$k\leq r-1$，不妨设 $c_k=1$。则我们有

$$A^k\vec{x}+c_{k+1}A^{k+1}\vec{x}+\cdots+c_{r-1}A^{r-1}\vec{x}=\Big(I_n+g(A)\Big)A^k\vec{x}=0$$

这里 $g(A)=c_{k+1}A+\cdots+c_{r-1}A^{r-1-k}$。由于 A 是幂零方阵，$g(A)$ 也是幂零方阵，由命题4.4.2知 $I_n+g(A)$ 可逆。两边乘以 $\big(I_n+g(A)\big)^{-1}$，我们得到

$$A^k\vec{x}=0$$

这与 r 的最小性矛盾。 □

定义 4.4.6. 设 f 是幂零变换（或 A 是幂零方阵）。定义向量 $\vec{x}$ 在 f（或 A）作用下的深度 $d_{\vec{x}}$ 为满足 $f^r(\vec{x})=0$（或 $A^r\vec{x}=0$）的最小正整数 r。由命题*4.4.5*知

$$d_{\vec{x}}=\dim C_{\vec{x}}$$

命题 4.4.7. 设 V 是 n 维线性空间，$f:V\to V$ 是一个 m 次幂零变换。则存在向量 $\vec{\alpha}_1,\cdots,\vec{\alpha}_t\in V$ 使得

$$V=C_{\vec{\alpha}_1}\oplus C_{\vec{\alpha}_2}\oplus\cdots\oplus C_{\vec{\alpha}_t}$$

即 V 可以分解为循环子空间的直和。此时 $\max\{d_{\vec{\alpha}_i}\}=\max\{\dim C_{\vec{\alpha}_i}\}=$ 幂零指数m。

证明: 我们对 m 作归纳。当 $m=1$ 时命题显然成立。考虑 V 的线性子空间

$$f(V)=\operatorname{im} f\subset V$$

易知 $f(V)$ 是 V 的不变子空间，且 f 限制在 $f(V)$ 上得到的线性变换

$$f:f(V)\to f(V)$$

是一个 $m-1$ 次幂零变换。由归纳假设，存在 $\vec{\gamma}_1,\cdots,\ \vec{\gamma}_r\in f(V)$ 使得

$$f(V)=C_{\vec{\gamma}_1}\oplus\cdots\oplus C_{\vec{\gamma}_r}$$

由 $\vec{\gamma}_i\in f(V)$，存在向量 $\vec{\alpha}_i$ 使得 $\vec{\gamma}_i=f(\vec{\alpha}_i), i=1,\cdots,r$。我们首先说明

$$C_{\vec{\alpha}_1}+\cdots+C_{\vec{\alpha}_r}=C_{\vec{\alpha}_1}\oplus\cdots\oplus C_{\vec{\alpha}_r}$$

是直和。假设有 $\vec{\beta}_i=c_{i,0}\vec{\alpha}_i+c_{i,1}f(\vec{\alpha}_i)+\cdots\in C_{\vec{\alpha}_i}$ 使得

$$\vec{\beta}_1+\vec{\beta}_2+\cdots+\vec{\beta}_r=0$$

两边作用 f 得到 $f(V)$ 中的关系

$$c_{1,0}\vec{\gamma}_1+\cdots+c_{r,0}\vec{\gamma}_r+\cdots=0$$

由如上 $f(V)$ 的直和分解，知 $c_{1,0}=\cdots=c_{r,0}=0$。因此 $\vec{\beta}_i\in C_{\vec{\gamma}_i}$。再由直和分解知 $\vec{\beta}_i=0$。这证明了 $C_{\vec{\alpha}_1}+\cdots+C_{\vec{\alpha}_r}=C_{\vec{\alpha}_1}\oplus\cdots\oplus C_{\vec{\alpha}_r}$ 是直和。

由于 f 是幂零变换，只有零特征值。考虑其对应的特征子空间

$$V_0=\{\vec{x}\in V|f(\vec{x})=0\}$$

设 $d_i=d_{\vec{\alpha}_i}, i=1,\cdots,r$。则 $\{f^{d_1-1}(\vec{\alpha}_1),\cdots,f^{d_r-1}(\vec{\alpha}_r)\}$ 是 V_0 中线性无关的向量。我们添加向量 $\{\vec{\alpha}_{r+1},\cdots,\vec{\alpha}_t\}$ 使得

$$\{f^{d_1-1}(\vec{\alpha}_1),\cdots,f^{d_r-1}(\vec{\alpha}_r),\vec{\alpha}_{r+1},\cdots,\vec{\alpha}_t\}$$

构成 V_0 的一组基。注意到对于 $i=r+1,\cdots,t$，$C_{\vec{\alpha}_i}=\mathrm{Span}\{\vec{\alpha}_i\}$。我们下面说明

$$C_{\vec{\alpha}_1}+\cdots+C_{\vec{\alpha}_r}+C_{\vec{\alpha}_{r+1}}+\cdots+C_{\vec{\alpha}_t}=C_{\vec{\alpha}_1}\oplus\cdots\oplus C_{\vec{\alpha}_r}\oplus C_{\vec{\alpha}_{r+1}}\oplus\cdots\oplus C_{\vec{\alpha}_t}$$

是直和。假设有 $\vec{\beta}_i\in C_{\vec{\alpha}_i}$ 使得

$$\vec{\beta}_1+\cdots+\vec{\beta}_r+\vec{\beta}_{r+1}+\cdots+\vec{\beta}_t=0$$

两边作用 f 得到 $f(\vec{\beta}_1+\cdots+\vec{\beta}_r)=0$，即 $\vec{\beta}_1+\cdots+\vec{\beta}_r\in V_0$。因此我们有

$$\vec{\beta}_1+\cdots+\vec{\beta}_r\in(C_{\vec{\alpha}_1}\oplus\cdots\oplus C_{\vec{\alpha}_r})\cap V_0=\mathrm{Span}\{f^{d_1-1}(\vec{\alpha}_1),\cdots,f^{d_r-1}(\vec{\alpha}_r)\}$$

因此

$$\vec{\beta}_{r+1}+\cdots+\vec{\beta}_t=-(\vec{\beta}_1+\cdots+\vec{\beta}_r)\in\mathrm{Span}\{f^{d_1-1}(\vec{\alpha}_1),\cdots,f^{d_r-1}(\vec{\alpha}_r)\}$$

因为 $\{f^{d_1-1}(\vec{\alpha}_1),\cdots,f^{d_r-1}(\vec{\alpha}_r),\vec{\alpha}_{r+1},\cdots,\vec{\alpha}_t\}$ 线性无关，我们得到 $\vec{\beta}_{r+1}=\cdots=\vec{\beta}_t=0$。于是

$$\vec{\beta}_1+\cdots+\vec{\beta}_r=0$$

又由于 $C_{\vec{\alpha}_1}\oplus\cdots\oplus C_{\vec{\alpha}_r}$ 是直和，我们进一步得到 $\vec{\beta}_1=\cdots=\vec{\beta}_r=0$。这证明了如上的直和性质。

最后我们说明 $\{C_{\vec{\alpha}_1},\cdots,C_{\vec{\alpha}_t}\}$ 张成了整个 V。实际上，对 V 中任一向量 $\vec{x}$，由 $f(\vec{x})\in f(V)=C_{\vec{\gamma}_1}\oplus\cdots\oplus C_{\vec{\gamma}_r}$ 知存在 $\vec{y}\in C_{\vec{\alpha}_1}\oplus\cdots\oplus C_{\vec{\alpha}_r}$ 使得 $f(\vec{x})=f(\vec{y})$。这说明 $\vec{x}-\vec{y}\in V_0$，即可以被 $\{f^{d_1-1}(\vec{\alpha}_1),\cdots,f^{d_r-1}(\vec{\alpha}_r),\vec{\alpha}_{r+1},\cdots,\vec{\alpha}_t\}$ 线性表达。因此

$$\vec{x}=\vec{y}+(\vec{x}-\vec{y})\in C_{\vec{\alpha}_1}+\cdots+C_{\vec{\alpha}_r}+C_{\vec{\alpha}_{r+1}}+\cdots+C_{\vec{\alpha}_s}$$

结合直和分解性质，我们得到

$$V=C_{\vec{\alpha}_1}\oplus\cdots\oplus C_{\vec{\alpha}_r}\oplus C_{\vec{\alpha}_{r+1}}\oplus\cdots\oplus C_{\vec{\alpha}_s}$$

由归纳知原命题成立。 □

命题 4.4.8. 设 V 是 n 维线性空间，$f:V\to V$ 是一个 m 次幂零变换。设

$$V=C_{\vec{\alpha}_1}\oplus C_{\vec{\alpha}_2}\oplus\cdots\oplus C_{\vec{\alpha}_t}$$

是命题4.4.7给出的关于 f 的循环子空间的直和。则

$$\dim\ker f^m-\dim\ker f^{m-1}=\text{集合}\{i|\dim C_{\vec{\alpha}_i}\geq m\}\text{的元素个数}$$

这里 $\ker f^m=\{\vec{x}\in V|f^m(\vec{x})=0\}$。

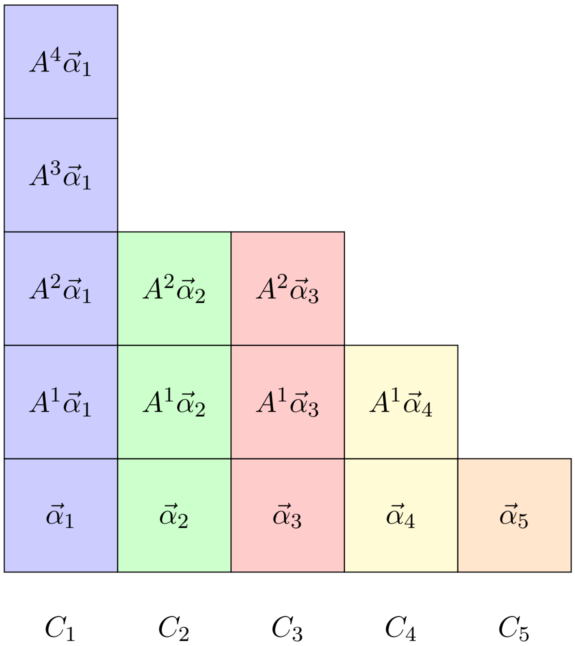


图 4.1: 循环子空间分解

证明：留作练习。 □

例子 4.4.9. 设 A 是一个幂零方阵，对应一个幂零线性映射 f。假设如下的循环子空间分解

则由上图容易看出

$$\begin{aligned}
\dim\ker A &= \dim \operatorname{Span}\{A^4\vec{\alpha}_1, A^2\vec{\alpha}_2, A^2\vec{\alpha}_3, A\vec{\alpha}_4, \vec{\alpha}_5\} = 5 \\
\dim\ker A^2 &= \dim \operatorname{Span}\{A^4\vec{\alpha}_1, A^3\vec{\alpha}_1, A^2\vec{\alpha}_2, A\vec{\alpha}_2, A^2\vec{\alpha}_3, A\vec{\alpha}_3, A\vec{\alpha}_4, \vec{\alpha}_4, \vec{\alpha}_5\} = 5+4 \\
\dim\ker A^3 &= \dim \operatorname{Span}\{A^4\vec{\alpha}_1, A^3\vec{\alpha}_1, A^2\vec{\alpha}_1, A^2\vec{\alpha}_2, A\vec{\alpha}_2, \vec{\alpha}_2, A^2\vec{\alpha}_3, A\vec{\alpha}_3, \vec{\alpha}_3, A\vec{\alpha}_4, \vec{\alpha}_4, \vec{\alpha}_5\} = 5+4+3 \\
\dim\ker A^4 &= \dim \operatorname{Span}\{A^4\vec{\alpha}_1, A^3\vec{\alpha}_1, A^2\vec{\alpha}_1, A\vec{\alpha}_1, A^2\vec{\alpha}_2, A\vec{\alpha}_2, \vec{\alpha}_2, A^2\vec{\alpha}_3, A\vec{\alpha}_3, \vec{\alpha}_3, A\vec{\alpha}_4, \vec{\alpha}_4, \vec{\alpha}_5\} \\
&= 5+4+3+1 \\
\dim\ker A^5 &= \dim \operatorname{Span}\{A^4\vec{\alpha}_1, A^3\vec{\alpha}_1, A^2\vec{\alpha}_1, A\vec{\alpha}_1, \vec{\alpha}_1, A^2\vec{\alpha}_2, A\vec{\alpha}_2, \vec{\alpha}_2, A^2\vec{\alpha}_3, A\vec{\alpha}_3, \vec{\alpha}_3, A\vec{\alpha}_4, \vec{\alpha}_4, \vec{\alpha}_5\} \\
&= 5+4+3+1+1 \\
\dim\ker A^6 &= \dim\ker A^5
\end{aligned}$$

由这些维数信息不难反解出来有 1 个维数 5 的循环子空间，2 个维数 3 的循环子空间，1 个维数 2 的循环子空间，1 个维数 1 的循环子空间。

给定 n 阶幂零方阵 A，由命题4.4.7我们知道 n 维空间可以分解为 A 的一些循环子空间的直和。由命题4.4.8，虽然这些循环子空间的选法并不唯一，但是循环子空间的个数以及每个循

环子空间的维数是由 A 决定的。具体而言，可以通过解线性方程组依次计算维数

$$\begin{cases} \dim\ker A = \dim\{\vec{x}|A\vec{x}=0\} \\ \dim\ker A^2 = \dim\{\vec{x}|A^2\vec{x}=0\} \\ \vdots \\ \dim\ker A^m = \dim\{\vec{x}|A^m\vec{x}=0\} \\ \vdots \end{cases}$$

得到。这里 $\dim\ker A$ 即特征子空间的维数表示有多少个循环子空间 C_i。然后依次有

$$\dim\ker A^m - \dim\ker A^{m-1} = \sharp\{\text{维数至少是 } m \text{ 的 } C_i \text{ 的个数}\}$$

如果发现当到达某个 N 使得

$$\dim\ker A^{N+1} = \dim\ker A^N$$

时就可以停止计算。由所有这些信息 $\{\dim\ker A^m\}_{1\le m\le N}$ 可以推出每个循环子空间的维数。

4.4.2 Jordan 块

定义 4.4.10. 如下 m 阶方阵称为一个 m 阶 *Jordan* 块，记为

$$J_m(\lambda) := \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$$

命题 4.4.11. 设 A 是一个 n 阶幂零方阵。则存在正整数 $m_1, \cdots, m_k$，使得 A 相似于由 *Jordan* 块构成的如下方阵

$$\begin{bmatrix} J_{m_1}(0) & 0 & \cdots & 0 \\ 0 & J_{m_2}(0) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_{m_k}(0) \end{bmatrix}$$

这里 $m_1 + \cdots + m_k = n$，且 A 的幂零指数是 $\max\{m_1, \cdots, m_k\}$。

证明: 由命题4.4.7，存在向量 $\vec{\alpha}_1, \cdots, \vec{\alpha}_k$ 使得

$$k^n = C_{\vec{\alpha}_1} \oplus \cdots \oplus C_{\vec{\alpha}_k}$$

把这个直和分解中的基作为列向量排起来，构成 n 阶可逆方阵 P

$$P = \begin{bmatrix} A^{m_1-1}\vec{\alpha}_1 & \cdots & A\vec{\alpha}_1 & \vec{\alpha}_1 & \cdots\cdots & A^{m_i-1}\vec{\alpha}_i & \cdots & A\vec{\alpha}_i & \vec{\alpha}_i & \cdots\cdots & A^{m_k-1}\vec{\alpha}_k & \cdots & A\vec{\alpha}_k & \vec{\alpha}_k \end{bmatrix}$$

这里 $m_i = d_{\vec{\alpha}_i}$。则易知

$$A = P \begin{bmatrix} J_{m_1}(0) & 0 & \cdots & 0 \\ 0 & J_{m_2}(0) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_{m_k}(0) \end{bmatrix} P^{-1}$$

$\square$

命题 4.4.12. 设 A 是一个 n 阶方阵，且只有一个特征值 λ_0。则存在正整数 $m_1, \cdots, m_k$，使得 A 相似于由 Jordan 块构成的如下方阵

$$\begin{bmatrix} J_{m_1}(\lambda_0) & 0 & \cdots & 0 \\ 0 & J_{m_2}(\lambda_0) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_{m_k}(\lambda_0) \end{bmatrix}$$

这里 $m_1 + \cdots + m_k = n$。

证明：$A - \lambda_0 I_n$ 只有 0 特征值，因此是幂零矩阵。将命题4.4.11用于 $A - \lambda_0 I_n$，即得结论。 $\square$

4.4.3 Jordan 标准型

复方阵的相似标准型

我们首先考虑复数域上的方阵。

定理 4.4.13. 设 A 是 n 阶复方阵，$\lambda_1, \cdots, \lambda_s$ 是 A 的所有互不相同的特征值。则 A 相似于

$$\begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_s \end{bmatrix}$$

这里 J_i 是形如由 Jordan 块构成的方阵，$m_i^{(1)} + \cdots + m_i^{(t_i)} = \lambda_i$ 的代数重数

$$J_i = \begin{bmatrix} J_{m_i^{(1)}}(\lambda_i) & 0 & \cdots & 0 \\ 0 & J_{m_i^{(2)}}(\lambda_i) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_{m_i^{(t_i)}}(\lambda_i) \end{bmatrix}$$

如上形式称为复方阵 A 在相似变换下的 Jordan 标准型。

证明：由命题4.3.23，我们有根子空间的直和分解

$$\mathbb{C}^n = R_{\lambda_1} \oplus \cdots \oplus R_{\lambda_s}$$

设 $\{\vec{\beta}_1^{(i)}, \cdots, \vec{\beta}_{m_i}^{(i)}\}$ 为 R_{λ_i} 的一组基。由命题4.3.20，R_{λ_i} 是 A 的不变子空间，我们有

$$\begin{bmatrix} A\vec{\beta}_1^{(i)} & \cdots & A\vec{\beta}_{m_i}^{(i)} \end{bmatrix} = \begin{bmatrix} \vec{\beta}_1^{(i)} & \cdots & \vec{\beta}_{m_i}^{(i)} \end{bmatrix} B_i$$

这里 B_i 是 m_i 阶方阵，是线性映射 $A : R_{\lambda_i} \to R_{\lambda_i}$ 在 R_{λ_i} 的这组基下的表示矩阵。记 P 为 n 阶可逆方阵

$$P = \begin{bmatrix} \vec{\beta}_1^{(1)} & \cdots & \vec{\beta}_{m_1}^{(1)} & \cdots\cdots & \vec{\beta}_1^{(s)} & \cdots & \vec{\beta}_{m_s}^{(s)} \end{bmatrix}$$

则我们得到相似变换

$$A = P \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & B_s \end{bmatrix} P^{-1}$$

由命题4.3.21，方阵 B_i 只有一个特征值 λ_i，其特征多项式为 $(\lambda-\lambda_i)^{m_i}$。由命题4.4.12，存在 m_i 阶可逆方阵 Q_i 使得

$$B_i = Q_i J_i Q_i^{-1} \quad 其中 \quad J_i = \begin{bmatrix} J_{m_i^{(1)}}(\lambda_i) & 0 & \cdots & 0 \\ 0 & J_{m_i^{(2)}}(\lambda_i) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_{m_i^{(t_i)}}(\lambda_i) \end{bmatrix}$$

因此我们得到相似变换

$$A = P \begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ 0 & Q_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & Q_s \end{bmatrix} \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_s \end{bmatrix} \begin{bmatrix} Q_1^{-1} & 0 & \cdots & 0 \\ 0 & Q_2^{-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & Q_s^{-1} \end{bmatrix} P^{-1}$$

□

定义 4.4.14. *设 A 是 n 阶复方阵，其 Jordan 标准型如定理4.4.13所示。单项式 $(\lambda-\lambda_i)^{m_i^{(j)}}$ 称为属于特征值 λ_i 的一个初等因子。初等因子的全体*

$$\{(\lambda-\lambda_1)^{m_1^{(1)}}, \cdots, (\lambda-\lambda_1)^{m_1^{(t_1)}}, \cdots, (\lambda-\lambda_s)^{m_s^{(1)}}, \cdots, (\lambda-\lambda_s)^{m_s^{(t_s)}}\}$$

称为 A 的初等因子组。

命题 4.4.15. *初等因子组由方阵完全决定。两个方阵相似当且仅当它们具有相同的初等因子组。*

证明：略。 □

例子 4.4.16. *考虑方阵*

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$

A 和 B 具有相同的特征多项式 $(\lambda-2)^4(\lambda-3)^2$。上述矩阵已经是 Jordan 标准型的样子，我们可以读出对应的初等因子组

$$A\ 的初等因子组 = \{(\lambda-2)^3, (\lambda-2), (\lambda-3)^2\}$$
$$B\ 的初等因子组 = \{(\lambda-2)^2, (\lambda-2)^2, (\lambda-3)^2\}$$

A 和 B 的初等因子组不同，因此 A 与 B 不相似。

实方阵的相似标准型

我们简要说明一下实数域上的 Jordan 标准型。

一个 n 阶实方阵 A 总是可以通过一个实可逆矩阵 P 相似变换为如下标准形

$$A = P\begin{bmatrix} \boxed{\quad} & & & \\ & \boxed{\quad} & & \\ & & \ddots & \\ & & & \boxed{\quad} \end{bmatrix} P^{-1}$$

这里 $\boxed{\quad}$ 是如下形式的 Jordan 块

- 对于 A 的实特征值 λ_i，其 Jordan 块的形式为

$$\boxed{\quad} = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}$$

- 对于 A 的复特征值，其一定是成对出现 $\lambda_j = \alpha \pm i\beta$。这对复特征值对应的实 Jordan 块为

$$\boxed{\quad} = \begin{bmatrix} \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & & & 0 \\ & \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & & \\ & & \ddots & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \\ 0 & & & \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} & \end{bmatrix}$$

特别地，如果实方阵 A 作为复方阵可以对角化的话，那么存在一个可逆的实矩阵 P 使得

$$A = P\begin{bmatrix} \lambda_1 & & & & & & 0 \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_k & & & \\ & & & & \begin{bmatrix} \alpha_1 & -\beta_1 \\ \beta_1 & \alpha_1 \end{bmatrix} & & \\ & & & & & \begin{bmatrix} \alpha_2 & -\beta_2 \\ \beta_2 & \alpha_2 \end{bmatrix} & \\ 0 & & & & & & \ddots \end{bmatrix} P^{-1}$$

4.4.4 Jordan 标准型的计算

给定 n 阶复方阵 A，通过它的特征多项式可以得到所有不同的特征值 $\lambda_1,\cdots,\lambda_s$。我们可以对特征值 λ_i 依次解线性方程组

$$\begin{cases}\ker(A-\lambda_i)=\{\vec{x}|((A-\lambda_i))\vec{x}=0\}\\ \ker(A-\lambda_i)^2=\{\vec{x}|(A-\lambda_i)^2\vec{x}=0\}\\ \vdots\\ \ker(A-\lambda_i)^m=\{\vec{x}|(A-\lambda_i)^m\vec{x}=0\}\\ \vdots\end{cases}$$

直到 $\ker(A-\lambda_i)^N=\ker(A-\lambda_i)^{N+1}$ 为止。由命题4.4.8，通过上述空间的维数可以解出 λ_i 的根子空间对应的 Jordan 块的大小，即 λ_i 的初等因子组。由此可以得到 A 的 Jordan 标准型。进一步通过上述方程组可以得到对应的可逆相似变换。我们通过例子来具体说明。

例子 4.4.17. *计算* $A=\begin{bmatrix}3&1&0&1\\-1&5&4&1\\0&0&2&0\\0&0&0&4\end{bmatrix}$ *的 Jordan 标准型。A 的特征多项式为*

$$\varphi(\lambda)=\begin{vmatrix}\lambda-3&-1&0&-1\\1&\lambda-5&-4&-1\\0&0&\lambda-2&0\\0&0&0&\lambda-4\end{vmatrix}=(\lambda-2)(\lambda-4)^3$$

因此 A 有两个不同的特征值 $\lambda_1=2,\lambda_2=4$。

λ_1 的代数重数是 1，因此 λ_1 的根子空间是 1 维的，由特征向量张成。由

$$(A-\lambda_1)\vec{\beta}_1=\begin{bmatrix}1&1&0&1\\-1&3&4&1\\0&0&0&0\\0&0&0&2\end{bmatrix}\vec{\beta}_1=\vec{0}\quad\Longrightarrow\quad\vec{\beta}_1=\begin{bmatrix}1\\-1\\1\\0\end{bmatrix}\text{且是 } A \text{ 的特征向量,}$$

所以 λ_1 有一个1 阶 Jordan 块，贡献一个A 的特征向量；相应的，λ_1 的几何重数是 1。

λ_2 的代数重数是 3，因此 λ_2 的根子空间是 3 维的。解方程

$$(A-\lambda_2)\vec{x}=\begin{bmatrix}-1&1&0&1\\-1&1&4&1\\0&0&-2&0\\0&0&0&0\end{bmatrix}\vec{x}=\vec{0}\quad\Longrightarrow\quad\vec{x}=\begin{bmatrix}c_1\\c_1-c_2\\0\\c_2\end{bmatrix}$$

解方程

$$(A-\lambda_2)^2\vec{x}'=\begin{bmatrix}0&0&4&0\\0&0&-4&0\\0&0&4&0\\0&0&0&0\end{bmatrix}\vec{x}'=\vec{0}\quad\Longrightarrow\quad\vec{x}'=\begin{bmatrix}c_1'\\c_2'\\0\\c_3'\end{bmatrix}$$

$\ker(A-\lambda_2)^2$ 的维数是 3，已经得到 $\lambda_2=4$ 的根子空间。由

$$\dim\ker(A-\lambda_2)=2 \qquad \dim\ker(A-\lambda_2)^2=3$$

所以 λ_2 有一个2 阶 *Jordan* 块，一个1 阶 *Jordan* 块，贡献两个A 的特征向量；相应的，λ_2 的几何重数是 2。

因此 A 的 *Jordan* 标准型可以写为

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

通过上述方程组的解，可以把 λ_2 的根子空间的循环子空间分解找出来。

$$\text{取向量 } \vec{\alpha}_1 \text{ 使得 } \begin{cases} \vec{\alpha}_1 \in \ker(A-\lambda_2)^2, \ \ \vec{\alpha}_1 \notin \ker(A-\lambda_2) \\ \ker(A-\lambda_2)^2 = Span\{\vec{\alpha}_1\} \oplus \ker(A-\lambda_2) \end{cases} \quad \text{例如取 } \vec{\alpha}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\text{然后取向量 } \vec{\alpha}_2 \text{ 使得 } \begin{cases} \vec{\alpha}_2 \in \ker(A-\lambda_2) \\ \ker(A-\lambda_2) = Span\{(A-\lambda_2)\vec{\alpha}_1, \vec{\alpha}_2\} \end{cases} \quad \text{例如取 } \vec{\alpha}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

则 $\Big\{(A-\lambda_2)\vec{\alpha}_1, \vec{\alpha}_1, \vec{\alpha}_2\Big\}$ 构成 λ_2 的根子空间的一组基，其循环子空间分解为

$$\begin{aligned} R_{\lambda_2} &= C_{\vec{\alpha}_1} \ \oplus \ C_{\vec{\alpha}_2} \\ &= Span\{(A-\lambda_2)\vec{\alpha}_1, \vec{\alpha}_1\} \ \oplus \ Span\{\vec{\alpha}_2\} \end{aligned}$$

且 $(A-\lambda_2)\vec{\alpha}_1$，$\vec{\alpha}_2$ 是 λ_2 的两个特征向量。

由构造，按照对应顺序，则有

$$A\begin{bmatrix} \vec{\beta}_1 & (A-\lambda_2)\vec{\alpha}_1 & \vec{\alpha}_1 & \vec{\alpha}_2 \end{bmatrix} = \begin{bmatrix} \vec{\beta}_1 & (A-\lambda_2)\vec{\alpha}_1 & \vec{\alpha}_1 & \vec{\alpha}_2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

记可逆矩阵 P 为

$$P = \begin{bmatrix} \vec{\beta}_1 & (A-\lambda_2)\vec{\alpha}_1 & \vec{\alpha}_1 & \vec{\alpha}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

得到 A 与 A 的 *Jordan* 标准型之间的相似变换

$$A = P\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}P^{-1}$$

例子 4.4.18. 计算 $A=\begin{bmatrix}2&1&-1&8&-3\\0&2&0&7&5\\0&0&2&7&5\\0&0&0&2&0\\0&0&0&0&2\end{bmatrix}$ 的 *Jordan* 标准型。A 为上三角方阵，只有一个特征值 $\lambda_1=2$，代数重数是 *5*。

解方程

$$(A-\lambda_1)\vec{x}=\begin{bmatrix}0&1&-1&8&-3\\0&0&0&7&5\\0&0&0&7&5\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}\vec{x}=\vec{0}\quad\Longrightarrow\quad\vec{x}=\begin{bmatrix}c_1\\c_2\\c_2+61c_3\\5c_3\\-7c_3\end{bmatrix}$$

计算得 $(A-\lambda_1)^2=\mathbf{0}$，解方程 $(A-\lambda_1)^2\vec{x}=\vec{0}$，其解为整个空间，即 $\ker(A-\lambda_1)^2=$ 整个空间。因此

$$\dim\ker(A-\lambda_1)=3\qquad\dim\ker(A-\lambda_1)^2=5$$

由此知 λ_1 有两个2 阶 *Jordan* 块，一个1 阶 *Jordan* 块，贡献三个A 的特征向量；相应的，λ_1 的几何重数是 *3*。

取两个向量使得

$$\begin{cases}\vec{\alpha}_1,\vec{\alpha}_2\in\ker(A-\lambda_1)^2,\ \ \vec{\alpha}_1,\vec{\alpha}_2\notin\ker(A-\lambda_1)\\ \ker(A-\lambda_1)^2=Span\{\vec{\alpha_1},\vec{\alpha_2}\}\oplus\ker(A-\lambda_1)\end{cases}\qquad\text{例如取}\quad\vec{\alpha}_1=\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}\quad\vec{\alpha}_2=\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}$$

然后取一个向量 $\vec{\alpha}_3$ 使得

$$\begin{cases}\vec{\alpha}_3\in\ker(A-\lambda_1)\\ \ker(A-\lambda_1)=\mathrm{Span}\{(A-\lambda_1)\vec{\alpha}_1,(A-\lambda_1)\vec{\alpha}_2,\vec{\alpha}_3\}\end{cases}\qquad\text{例如取}\qquad\vec{\alpha}_3=\begin{bmatrix}0\\0\\61\\5\\-7\end{bmatrix}$$

则 $\{(A-\lambda_1)\vec{\alpha}_1,\ \vec{\alpha}_1,(A-\lambda_1)\vec{\alpha}_2,\ \vec{\alpha}_2,\ \vec{\alpha}_3\}$ 构成 λ_1 的根子空间 *(*即整个空间*)* 的一组基，其循环子空间分解为

$$\begin{aligned}R_{\lambda_1}&=C_{\vec{\alpha_1}}\oplus C_{\vec{\alpha_2}}\oplus C_{\vec{\alpha_3}}\\&=Span\{(A-\lambda_1)\vec{\alpha}_1,\ \vec{\alpha}_1\}\oplus Span\{(A-\lambda_1)\vec{\alpha}_2,\ \vec{\alpha}_2\}\oplus Span\{\vec{\alpha}_3\}\end{aligned}$$

且 $(A-\lambda_1)\vec{\alpha}_1$，$(A-\lambda_1)\vec{\alpha}_2$，$\vec{\alpha}_3$ 是 λ_1 的三个特征向量。

由构造，按照对应顺序，则有

$$A\begin{bmatrix}(A-\lambda_1)\vec{\alpha}_1 & \vec{\alpha}_1 & (A-\lambda_1)\vec{\alpha}_2 & \vec{\alpha}_2 & \vec{\alpha}_3\end{bmatrix}=\begin{bmatrix}(A-\lambda_1)\vec{\alpha}_1 & \vec{\alpha}_1 & (A-\lambda_1)\vec{\alpha}_2 & \vec{\alpha}_2 & \vec{\alpha}_3\end{bmatrix}\begin{bmatrix}2&1&0&0&0\\0&2&0&0&0\\0&0&2&1&0\\0&0&0&2&0\\0&0&0&0&2\end{bmatrix}$$

记可逆矩阵 P 为

$$P=\begin{bmatrix}(A-\lambda_1)\vec{\alpha}_1 & \vec{\alpha}_1 & (A-\lambda_1)\vec{\alpha}_2 & \vec{\alpha}_2 & \vec{\alpha}_3\end{bmatrix}=\begin{bmatrix}1&0&-3&0&0\\0&1&5&0&0\\0&0&5&0&61\\0&0&0&0&5\\0&0&0&1&-7\end{bmatrix}$$

得到 A 与 A 的 *Jordan* 标准型之间的相似变换

$$A=P\begin{bmatrix}2&1&0&0&0\\0&2&0&0&0\\0&0&2&1&0\\0&0&0&2&0\\0&0&0&0&2\end{bmatrix}P^{-1}$$

4.5 一些例子

4.5.1 可逆矩阵开根

命题 4.5.1. 设 A 是 n 阶可逆复方阵。则存在可逆方阵 B 使得 $A=B^2$。

我们首先考虑 $A=J_n(\lambda)$ 是一个 Jordan 块的情况，即

$$A=\lambda I_n+N \qquad N=J_n(0)=\begin{bmatrix}0&1&&0\\&0&\ddots&\\&&\ddots&1\\0&&&0\end{bmatrix}$$

我们把 A 写成 $A=\lambda(I_n+\lambda^{-1}N)$。利用函数 $(1+x)^{1/2}$ 在 $x=0$ 处的 Taylor 展开

$$(1+x)^{1/2}=1+\sum_{k=1}^{\infty}\frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}x^k$$

我们得到级数恒等式

$$1+x=\left(1+\sum_{k=1}^{\infty}\frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}x^k\right)^2$$

代入矩阵 $x=\lambda^{-1}N$ 并利用 $N^n=0$，我们得到

$$I_n+\lambda^{-1}N=\left(I_n+\sum_{k=1}^{n-1}\frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}\lambda^{-k}N^k\right)^2$$

因此我们可以取

$$B=\sqrt{\lambda}\left(I_n+\sum_{k=1}^{n-1}\frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}\lambda^{-k}N^k\right)$$

对于一般的情况，我们可以作相似变换

$$A=P\begin{bmatrix}J_{m_1}(\lambda_1) & & 0\\ & \cdots & \\ 0 & \cdots & J_{m_k}(\lambda_k)\end{bmatrix}P^{-1}$$

这里每个 $J_{m_i}(\lambda_1)$ 都是 Jordan 块，因此可以如上构造 B_i 使得 $J_{m_i}(\lambda_1)=B_i^2$。取

$$B=P\begin{bmatrix}B_1 & & 0\\ & \cdots & \\ 0 & \cdots & B_k\end{bmatrix}P^{-1}$$

则满足 $B^2=A$。

注记. 类似的方法可以证明，对于 n 阶可逆复方阵和正整数 k，存在可逆方阵 B 使得 $A=B^k$。

4.5.2 指数矩阵

命题 4.5.2. 设 A 是 n 阶复方阵。则如下矩阵级数

$$e^A:=\sum_{k=0}^{\infty}\frac{A^k}{k!}=I_n+A+\frac{A^2}{2!}+\cdots+\frac{A^m}{m!}+\cdots$$

收敛到一个可逆矩阵，并且

$$\det(e^A)=e^{\operatorname{Tr}A}$$

我们同样也是先考虑 $A=J_n(\lambda)$ 是一个 Jordan 块的情况，即

$$A=\lambda I_n+N\qquad N=J_n(0)=\begin{bmatrix}0 & 1 & & 0\\ & 0 & \ddots & \\ & & \ddots & 1\\ 0 & & & 0\end{bmatrix}$$

利用 $N^n=0$ 易知 e^A 收敛到（计算留作练习）

$$\begin{aligned}e^A=&e^{\lambda I_n}e^N=e^{\lambda}\left(I_n+N+\frac{N^2}{2!}\cdots+\frac{N^{n-1}}{(n-1)!}\right)\\=&\begin{bmatrix}e^{\lambda} & e^{\lambda} & \cdots & \cdots & \frac{e^{\lambda}}{(n-2)!} & \frac{e^{\lambda}}{(n-1)!}\\ & e^{\lambda} & \cdots & \cdots & \cdots & \frac{e^{\lambda}}{(n-2)!}\\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & e^{\lambda} & e^{\lambda}\\ 0 & & & & & e^{\lambda}\end{bmatrix}\end{aligned}$$

此时有 $\det(e^A)=e^{n\lambda}=e^{\operatorname{Tr}A}$。

对于一般的情况，我们可以作相似变换

$$A=P\begin{bmatrix}J_{m_1}(\lambda_1) & & 0\\ & \cdots & \\ 0 & \cdots & J_{m_k}(\lambda_k)\end{bmatrix}P^{-1}$$

则

$$e^A=P\begin{bmatrix}e^{J_{m_1}(\lambda_1)} & & 0\\ & \cdots & \\ 0 & \cdots & e^{J_{m_k}(\lambda_k)}\end{bmatrix}P^{-1}$$

并且

$$\det(e^A)=e^{m_1\lambda_1+\cdots+m_k\lambda_k}=e^{\operatorname{Tr}A}$$

成立。由于 $e^{\operatorname{Tr}A}\neq 0$，$e^A$ 是可逆矩阵。实际上，e^A 的逆易知是 e^{-A}。

例子 4.5.3. *考虑方阵*

$$A=\begin{bmatrix}3 & -2 & 2\\ 0 & 1 & 2\\ 0 & -1/2 & 3\end{bmatrix}$$

可以验证它相似于如下 Jordan 标准型

$$A=P\begin{bmatrix}3 & 0 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix}P^{-1}\qquad \text{这里}\qquad P=\begin{bmatrix}1 & 2 & 0\\ 0 & 2 & 0\\ 0 & 1 & 1\end{bmatrix}\quad P^{-1}=\begin{bmatrix}1 & -1 & 0\\ 0 & 1/2 & 0\\ 0 & -1/2 & 1\end{bmatrix}$$

A 有两个 Jordan 块。因此

$$e^A=P\begin{bmatrix}e^3 & 0 & 0\\ 0 & e^2 & e^2\\ 0 & 0 & e^2\end{bmatrix}P^{-1}=\begin{bmatrix}1 & 2 & 0\\ 0 & 2 & 0\\ 0 & 1 & 1\end{bmatrix}\begin{bmatrix}e^3 & 0 & 0\\ 0 & e^2 & e^2\\ 0 & 0 & e^2\end{bmatrix}\begin{bmatrix}1 & -1 & 0\\ 0 & 1/2 & 0\\ 0 & -1/2 & 1\end{bmatrix}=\begin{bmatrix}e^3 & -e^3 & 2e^2\\ 0 & 0 & 2e^2\\ 0 & -e^2/2 & 2e^2\end{bmatrix}$$

4.5.3 极小多项式

由 Cayley-Hamilton 定理，我们知道 n 阶方阵 A 的特征多项式 $\varphi(\lambda)=\det(\lambda I_n-A)$ 是 A 的一个化零多项式，即满足

$$\varphi(A)=0$$

实际上满足 $f(A)=0$ 的多项式，即 A 的化零多项式中，特征多项式 $\varphi(\lambda)$ 并不一定是次数最低的。例如对于方阵

$$A=\begin{bmatrix}2 & 1 & 0 & 0\\ 0 & 2 & 0 & 0\\ 0 & 0 & 3 & 0\\ 0 & 0 & 0 & 3\end{bmatrix}\qquad \varphi(\lambda)=(\lambda-2)^2(\lambda-3)^2$$

容易验证 A 满足一个 3 次代数方程

$$(A-2I_4)^2(A-3I_4)=0$$

定义 4.5.4. 方阵 A 的所有非平凡化零多项式中次数最小的首一多项式称为 A 的极小多项式，记为 $d_A(\lambda)$。这里首一多项式指的是多项式最高次项的系数是 1，即 $d_A(\lambda) = \lambda^d + \cdots$.

由 Cayley-Hamilton 定理知 A 有化零多项式，因此 A 的极小多项式是存在的且次数不超过 A 的阶数。另一方面，A 的极小多项式是唯一的。假设 $d_A(\lambda)$ 和 $d'_A(\lambda)$ 均为 A 的极小多项式。则 $d_A(\lambda)$ 和 $d'_A(\lambda)$ 的次数相同，并且 $d_A(\lambda) - d'_A(\lambda)$ 是次数更低的化零多项式。由极小多项式的次数最小性，说明 $d_A(\lambda) - d'_A(\lambda) = 0$ 是平凡的，即 $d_A(\lambda) = d'_A(\lambda)$。

首先我们注意到，如果 λ_0 不是 A 的特征值，则线性方程组

$$(A - \lambda_0 I_n)\vec{x} = 0$$

只有零解，即 $A - \lambda_0 I_n$ 是可逆矩阵。因此如果 $f(\lambda)$ 是 A 的化零多项式且 λ_0 是 f 的一个根，即有 $f(\lambda) = (\lambda - \lambda_0)g(\lambda)$，则

$$f(A) = (A - \lambda_0 I_n)g(A) = 0 \quad \xRightarrow{\text{左乘 } (A-\lambda_0 I_n)^{-1}} \quad g(A) = 0$$

故 $g(\lambda)$ 也是 A 的化零多项式且次数比 f 低。这说明 $d_A(\lambda)$ 的根只包含 A 的特征值。

例子 4.5.5. 设 $A = J_n(\lambda_0)$ 是 n 阶 *Jordan* 块，则 A 的最小多项式为

$$d_A(\lambda) = (\lambda - \lambda_0)^n$$

实际上，A 只有一个特征值 λ_0，其最小多项式形如 $(\lambda - \lambda_0)^m$。由于

$$N = A - \lambda_0 I_n = \begin{bmatrix} 0 & 1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix}$$

满足 $N^{n-1} \neq 0, N^n = 0$，我们得到 $m = n$。此时 A 的极小多项式等于特征多项式。

命题 4.5.6. 设 A 是 n 阶复方阵，$\lambda_1, \cdots, \lambda_s$ 是 A 的所有互不相同的特征值。则 A 的最小多项式为

$$d_A(\lambda) = (\lambda - \lambda_1)^{d_1} \cdots (\lambda - \lambda_s)^{d_s}$$

其中 d_i 是特征值 λ_i 的 *Jordan* 块的最大阶数，即 $(\lambda - \lambda_i)^{d_i}$ 是特征值 λ_i 最高次数的初等因子。

证明：A 相似于 Jordan 标准型

$$A = P \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_s \end{bmatrix} P^{-1}$$

对任何多项式 $f(\lambda)$，我们有

$$f(A) = P \begin{bmatrix} f(J_1) & 0 & \cdots & 0 \\ 0 & f(J_2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & f(J_s) \end{bmatrix} P^{-1}$$

因此 $f(A)=0$ 当且仅当对每个 J_i 均有 $f(J_i)$=0。而每个 J_i 是形如由 Jordan 块构成的方阵

$$J_i = \begin{bmatrix} J_{m_i^{(1)}}(\lambda_i) & 0 & \cdots & 0 \\ 0 & J_{m_i^{(2)}}(\lambda_i) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & J_{m_i^{(t_i)}}(\lambda_i) \end{bmatrix}$$

因此 $f(J_i)=0$ 当且仅当对每个 $f(J_{m_i^{(k)}}(\lambda_i))=0$。由例4.5.5易知

$$d_i = \max\{m_i^{(1)}, m_i^{(2)}, \cdots, m_i^{(t_i)}\}$$

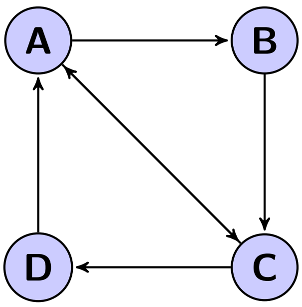
□

例子 4.5.7. 方阵 $A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ 的极小多项式为 $(\lambda-2)^2(\lambda-3)$。

4.5.4 Google PageRank

PageRank 是 Google 搜索引擎用来排名网页相关性和重要性的算法。其核心思想是一个网页的重要性与其链接到的网页的重要性的数量和质量有关。基本数学原理是通过网页数据得到一个大型的矩阵，然后通过求解最大特征值的特征向量得到排名数据。

我们将互联网看作一个巨大的有向图，每个节点代表一个网页，每条有向边代表一个链接，边的方向表示链接的方向。



上图表示有 4 个网页 A、B、C、D，链接关系如下：

- A 链接到 B 和 C
- B 链接到 C
- C 链接到 A 和 D
- D 链接到 A

PageRank 算法将给每个网页 i 匹配一个 PageRank 值 r_i，r_i 的值越大表示其越重要。其基本思想是这个 PageRank 值是通过网络关系来体现的：如果更多更重要的网页关联到某个网页，那么这个网页本身也更重要。具体而言，引入链接关系矩阵 $M=(m_{ij})$：

- 如果网页 j 链接到网页 i，则

$$m_{ij} = \frac{1}{L(j)}$$

  其中 $L(j)$ 是所有从网页 j 链出的边数。

- 如果网页 j 没有链接到网页 i，则 $m_{ij} = 0$

m_{ij} 表现了从网页 j 链接到网页 i 的权重。例如对上图所示 4 个网络，其链接关系矩阵为

$$M = \begin{bmatrix} 0 & 0 & 1/2 & 1 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \end{bmatrix}$$

这里我们用 $i = 1, 2, 3, 4$ 来分别表示网页 A,B,C,D。例如

- $m_{13} = 1/2$，因为网页 C 链接到网页 A，且网页 C 的出链数量为 2（链接到 A 和 D）。
- $m_{21} = 1/2$，因为网页 A 链接到网页 B，且网页 A 的出链数量为 2（链接到 B 和 C）。
- $m_{32} = 1$，因为网页 B 链接到网页 C，且网页 B 的出链数量为 1（链接到 C）。
- $m_{23} = 0$，因为没有链接从网页 C 链接到网页 B。

PageRank 算法用来计算每个网页 PageRank 权重 $r(i)$ 的方法为：一个网页的 PageRank 值等于所有链接到它的网页的 PageRank 值的加权平均，权重为链接概率。具体公式为：

$$r_i = \sum_j m_{ij} r_j$$

我们把所有网页的 PageRank 权重 $\{r_i\}$ 组合为列向量 $\vec{r}$，则上述方程可以写成矩阵形式

$$M\vec{r} = \vec{r}$$

因此寻找网页的 PageRank 值即为求解链接关系矩阵 M 属于特征值 1 的特征向量。

上述 4 个网页的例子中，求解 $M\vec{r} = \vec{r}$ 可以得到归一化的 PageRank 值

$$\vec{r} = \begin{bmatrix} 1/3 \\ 1/6 \\ 1/3 \\ 1/6 \end{bmatrix}$$

因此 A, C 重要性最高，B, D 次之。这个结果和图示是相符合的。

M **是否有特征值** 1 **?**

首先观察到每个网页链出的总权重是 1，即对每个 j 有

$$\sum_i m_{ij} = 1$$

写成矩阵的形式为

$$M^T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

这表明 M 的转置 M^T 有一个属于特征值 1 的特征向量。由于 M^T 和 M 具有相同的特征多项式，这说明 M 一定有特征值 1，因此方程 $M\vec{r} = \vec{r}$ 是有解的。

M **的属于特征值** 1 **的特征子空间** V_1 **是否是** 1 **维的？**

如果特征子空间 V_1 是 1 维的话，那么得到唯一的一个 PageRank 向量 $\vec{r}$（不同特征向量差一个系数，可以通过设置归一化条件 $\sum_i r_i = 1$ 把这个系数固定）。然而实际上很容易找到 $\dim V_1 > 1$ 的例子，这给用 PageRank 算法来准确地衡量网页的重要性造成了困难。

为了解决这个困难，PageRank 算法引入了一个阻尼因子 d(通常设置为 0.85)。阻尼因子表示用户有 d 的概率点击链接，$1-d$ 的概率随机跳转到任意一个网页。改进后的链接关系矩阵为

$$\widehat{M} = dM + (1-d)S$$

这里 S 的每个矩阵元都是 $1/n$（n 是网页的总个数）

$$S = \begin{bmatrix} 1/n & 1/n & \cdots & 1/n \\ 1/n & 1/n & \cdots & 1/n \\ \cdots & \cdots & \cdots & \cdots \\ 1/n & 1/n & \cdots & 1/n \end{bmatrix}$$

$\widehat{M}$ 的每一列加起来仍然是 1，因此仍然有特征值 1。改进的 PageRank 计算公式为

$$\widehat{M}\vec{r} = \vec{r}$$

命题 4.5.8. *$\widehat{M}$ 的特征子空间 V_1 是 1 维的。*

证明: 首先观察到 $\widehat{M}$ 的每个矩阵元 $\hat{m}_{ij} > 0$。假设 $\vec{r} \in V_1$ 是任意一个特征向量，即满足

$$\widehat{M}\vec{r} = \vec{r}$$

我们说明 $\vec{r}$ 的每个分量不可能即有正数又有负数。实际上，假如 r_i 中即有正数也有负数。由于 $\hat{m}_{ij} > 0$，我们有

$$|\sum_j \hat{m}_{ij} r_j| < \sum_j \hat{m}_{ij} |r_j|$$

由特征向量方程 $r_i = \sum_j \hat{m}_{ij} r_j$，得到

$$|r_i| = |\sum_j \hat{m}_{ij} r_j| < \sum_j \hat{m}_{ij} |r_j|$$

两边对 i 求和，并利用 $\sum_i \hat{m}_{ij} = 1$，我们得到

$$\sum_i |r_i| < \sum_{i,j} \hat{m}_{ij} |r_j| = \sum_j |r_j|$$

矛盾。这说明 $\vec{r}$ 的每个分量不可能即有正数又有负数。

现在假设 $\dim V_1 > 1$。设 $\vec{r}$ 和 $\vec{s}$ 是 V_1 中两个线性无关的特征向量。不妨设 $\vec{r}, \vec{s}$ 的分量都是非负的，因此 $\sum_i r_i > 0, \sum_i s_i > 0$。取非零实数 a, b 使得

$$a(\sum_i r_i) + b(\sum_i s_i) = 0$$

考虑线性组合 $a\vec{r} + b\vec{s}$。由于 $a\vec{r} + b\vec{s} \in V_1$，上述讨论知它每个分量不可能即有正数又有负数。由 a, b 的取法，向量 $a\vec{r} + b\vec{s}$ 的所有分量求和是 0，因此必然有 $a\vec{r} + b\vec{s} = 0$。这与 $\vec{r}, \vec{s}$ 线性无关的假设矛盾。因此假设 $\dim V_1 > 1$ 不成立，故 $\dim V_1 = 1$。 □

这个命题说明，在考虑阻尼因子后，用改进的算法的确得到一个确定的 PageRank 向量。可以进一步证明（见习题），$\widehat{M}$ 的根子空间 R_1 也是 1 维的，即 $V_1 = R_1$。

如何计算 PageRank 向量？

实际应用中网页的数量是非常多的，改进的链接关系矩阵 $\widehat{M}$ 是一个巨大的矩阵。如何有效的计算 PageRank 向量 $\vec{r}$？这里关键的一个性质是如下命题。

命题 4.5.9. 设 λ 是改进的链接关系矩阵 $\widehat{M}$ 的一个特征值，则 $|\lambda| \le 1$。

证明: 设复列向量 $\vec{\alpha} \in \mathbb{C}^n$ 是属于特征值 λ 的一个特征向量：$\widehat{M}\vec{\alpha} = \lambda\vec{\alpha}$。考虑如下行向量

$$E_{\vec{\alpha}} = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \quad \text{这里} \quad b_i = \begin{cases} \frac{\bar{\alpha}_i}{|\alpha_i|} & \alpha_i \neq 0 \\ 0 & \alpha_i = 0 \end{cases}$$

记

$$E_{\vec{\alpha}}\widehat{M} = \begin{bmatrix} k_1 & k_2 & \cdots & k_n \end{bmatrix}$$

由于 $\sum_i \hat{m}_{ij} = 1$ 并且每个 $\hat{m}_{ij} > 0$ 知

$$|k_j| = |\sum_i b_i \hat{m}_{ij}| \le \sum_i \hat{m}_{ij} = 1$$

由方程 $\widehat{M}\vec{\alpha} = \lambda\vec{\alpha}$ 得到 $E_{\vec{\alpha}}\widehat{M}\vec{\alpha} = \lambda E_{\vec{\alpha}}\vec{\alpha}$，即

$$\sum_i k_i \alpha_i = \lambda \sum_i |\alpha_i|$$

因此

$$\lambda = \frac{\sum\limits_i k_i \alpha_i}{\sum\limits_i |\alpha_i|} \quad \Longrightarrow \quad |\lambda| \le \frac{\sum\limits_i |k_i||\alpha_i|}{\sum\limits_i |\alpha_i|} \le 1$$

□

可以进一步证明，满足 $|\lambda| = 1$ 的特征值只有 $\lambda = 1$（见习题）。由命题4.5.8和命题4.5.9以及 Jordan 标准型知，改进的链接关系矩阵 $\widehat{M}$ 相似于

$$\widehat{M} = P \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & J_{m_1}(\lambda_1) & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & J_{m_k}(\lambda_k) \end{bmatrix} P^{-1}$$

这里 Jordan 块 $J_{m_i}(\lambda_i)$ 对应的特征值 λ_i 均满足 $|\lambda_i| < 1$。对于这样的 Jordan 块，可以证明（见习题）

$$\lim_{N\to\infty} J_{m_i}(\lambda_i)^N = 0$$

因此

$$\lim_{N\to\infty} \widehat{M}^N = \lim_{N\to\infty} P \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & J_{m_1}(\lambda_1)^N & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & J_{m_k}(\lambda_k)^N \end{bmatrix} P^{-1} = P \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} P^{-1}$$

设 P 的列向量为

$$P = \begin{bmatrix} \vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n \end{bmatrix}$$

由构造知 $\vec{\beta}_1$ 是特征值 1 的特征向量。我们得到

$$\lim_{N\to\infty} \widehat{M}^N = \begin{bmatrix} \vec{\beta}_1 & 0 & \cdots & 0 \end{bmatrix} P^{-1}$$

现在取一个初始向量 $\vec{r}_0$，设

$$P^{-1}\vec{r}_0 = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix}$$

并且 $a_1 \neq 0$。则

$$\lim_{N\to\infty} \widehat{M}^N \vec{r}_0 = a_1\beta_1$$

收敛到特征值 1 的一个特征向量！

这给出了如下计算 PageRank 向量 $\vec{r}$ 的算法：取初始向量

$$\vec{r}_0 = \begin{bmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

即初始时把所有网页的 PageRank 值都设置为 $1/n$。然后开始迭代计算

$$\begin{aligned}
\vec{r}_1 &= \widehat{M}\vec{r}_0 \\
\vec{r}_2 &= \widehat{M}\vec{r}_1 = \widehat{M}^2\vec{r}_0 \\
&\vdots \\
\vec{r}_N &= \widehat{M}^N\vec{r}_0
\end{aligned}$$

注意到

$$\begin{aligned}
\begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}\vec{r}_k &= \begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}\widehat{M}\vec{r}_{k-1} \\
&= \begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}\vec{r}_{k-1} \\
&= \cdots \\
&= \begin{bmatrix}1 & 1 & \cdots & 1\end{bmatrix}\vec{r}_0 = 1
\end{aligned}$$

因此每个迭代的向量 $\vec{r}_k$ 的所有分量的和都是 1，即都是归一化的。当 N 充分大时，多次迭代得到的向量 $\vec{r}_N$ 即给出了 PageRank 向量 $\vec{r}$ 的近似值。

4.6 习题

1. 求下列方阵的特征多项式、特征值以及每个特征值的特征向量

$$a)\begin{bmatrix}2 & -1 & 2\\ 5 & -3 & 3\\ 1 & 0 & -2\end{bmatrix}\qquad b)\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 1\end{bmatrix}\qquad c)\begin{bmatrix}a_0 & a_1 & a_2 & \cdots & a_{n-1}\\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2}\\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_{n-3}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ a_1 & a_2 & a_3 & \cdots & a_0\end{bmatrix}$$

2. 已知 $A=\begin{bmatrix}2 & 0 & 0\\ -1 & 3 & 1\\ -1 & 1 & 3\end{bmatrix}$，求 A^{20}。

3. 设 n 阶可逆方阵 A 的特征多项式为 $\varphi(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_n$。求 A^{-1} 的特征多项式，并说明 A^{-1} 的特征值与 A 的特征值的关系。

4. 设 A 是 $n\times m$ 矩阵, B 是 $m\times n$ 矩阵。证明 $\mathrm{Tr}\,AB=\mathrm{Tr}\ BA$。

5. 证明不存在 n 阶方阵 A,B 使得 $AB-BA=I_n$。注：这说明量子力学中的坐标和动量算子不可能通过有限维矩阵来实现。

6. 设 A,B 是 n 阶方阵。证明 AB 和 BA 的特征多项式相同。

7. 设 A,B 是两个 n 阶对角方阵。证明 A 和 B 相似当且仅当它们的特征多项式相同，即对角元素相同（排列顺序可能不同）。

8. 举例说明两个特征多项式相同的 n 阶方阵可能不相似。

9. 设 A,B 是两个 n 阶复方阵，A 的特征多项式为 $\varphi(\lambda)$。证明 n 阶复方阵 $\varphi(B)$ 可逆当且仅当 A 和 B 没有共同的特征值。

10. 设 A,B 是两个 n 阶复方阵，且没有共同的特征值。设 n 阶复方阵 X 满足 $AX=XB$。证明 $X=0$。

11. 求下列方阵所有的特征值、特征子空间和根子空间，并判断其是否可以对角化

$$a)\ \begin{bmatrix}2 & 0 & 1\\ -1 & 3 & 2\\ 0 & 0 & 2\end{bmatrix}\qquad b)\ \begin{bmatrix}0 & -2 & 3 & 2\\ 1 & 1 & -1 & -1\\ 0 & 0 & 2 & 0\\ 1 & -1 & 0 & 1\end{bmatrix}$$

12. (a) 证明 V 的两个子空间 V_1,V_2 的和是直和当且仅当 $V_1\cap V_2=\{0\}$ 是零子空间。

    (b) 举例说明，三个线性子空间 V_1,V_2,V_3 满足 $V_1\cap V_2=V_1\cap V_3=V_2\cap V_3=\{0\}$，但 $V_1+V_2+V_3$ 不一定是直和。

(c) 证明如果三个线性子空间 V_1, V_2, V_3 满足 $V_1 \cap V_2 = \{0\}, (V_1 + V_2) \cap V_3 = 0$，则 $V_1 + V_2 + V_3 = V_1 \oplus V_2 \oplus V_3$ 是直和。

13. 设 $V_1, V_2 \subset k^n$ 是 n 阶方阵 A 的两个不变子空间，$\dim V_1 = r, \dim V_2 = n - r$，并且 $k^n = V_1 \oplus V_2$。证明 A 相似于如下形式的矩阵

$$\begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$$

这里 B 是 r 阶方阵，C 是 $n - r$ 阶方阵。

14. 证明任意 n 阶复方阵都可以相似于一个上三角方阵。

15. (a) 证明任意 n 阶复方阵 A 都可以分解为 $A = D + N$。这里 D 是 n 阶可对角化方阵，N 是 n 阶幂零方阵，并且 $DN = ND$。

(b) 对于矩阵 $A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$，找出 (a) 中所述 D 和 N。

16. (a) 证明 Jordan 块 $J_n(0) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$ 与它的转置 $J_n(0)^T$ 相似。

(b) 证明 n 阶复方阵 A 与它的转置 A^T 相似。

17. 设 A, B 是两个可对角化的 n 阶复方阵。证明如下两个性质等价：

(a) A, B 可以同时对角化，即存在可逆 P 使得 PAP^{-1}, PBP^{-1} 均为对角阵

(b) A 与 B 可交换，即 $AB = BA$

18. (a) 设 $J_m(\lambda)$ 是一个 Jordan 块，$|\lambda| < 1$。证明 $\lim\limits_{N \to \infty} J_m(\lambda)^N = 0$。

(b) 设方阵 A 的所有特征值的绝对值均小于 1，证明 $\lim\limits_{N \to \infty} A^N = 0$。

19. 设 $\widehat{M}$ 是 Google PageRank 算法中改进的链接关系矩阵。

(a) 证明 -1 不是 $\widehat{M}$ 的特征值。

(b) 讲义中我们证明了 $\widehat{M}$ 关于特征值 $\lambda = 1$ 的几何重数是 1。证明 $\widehat{M}$ 关于特征值 $\lambda = 1$ 的代数重数也是 1，即特征值 1 的根子空间和特征子空间一致。

第五章　内积空间

在线性空间的代数结构基础上，可以进一步引入几何结构。基本的几何概念是“距离”，线性空间上与距离相关的数学结构称为“内积”。根据数域 k 是实数 $\mathbb{R}$ 还是复数 $\mathbb{C}$，主要有两类内积空间

- 带内积的实线性空间称为欧几里得空间
- 带内积的复线性空间称为酉空间

这一章我们将首先讨论欧几里得空间，然后在简述酉空间时相关概念与性质的对应和推广。

5.1　欧几里得空间

5.1.1　实线性空间的内积

定义 5.1.1. *设 V 是一个实线性空间。V 上的一个内积是一个映射*

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

满足以下性质：

1. **正定性：** *对任意 $\vec{x}\in V$, $\langle\vec{x},\vec{x}\rangle\geq 0$, 且 $\langle\vec{x},\vec{x}\rangle=0$ 当且仅当 $\vec{x}=0$*
2. **对称性：** *对任意 $\vec{x},\vec{y}\in V$, $\langle\vec{x},\vec{y}\rangle=\langle\vec{y},\vec{x}\rangle$*
3. **线性性：** *对任意 $\vec{x},\vec{y},\vec{z}\in V$ 和 $a,b\in\mathbb{R}$*

$$\langle a\vec{x}+b\vec{y},\vec{z}\rangle=a\langle\vec{x},\vec{z}\rangle+b\langle\vec{y},\vec{z}\rangle$$

注意到线性性是对内积的第一个分量来陈述的。由对称性，内积对第二个分量也是线性的：

$$\langle\vec{z},a\vec{x}+b\vec{y}\rangle=a\langle\vec{z},\vec{x}\rangle+b\langle\vec{z},\vec{y}\rangle$$

因此我们称内积是双线性的。

定义 5.1.2. *带内积的实线性空间称为实内积空间，或称为欧几里得空间。*

n 维实内积空间通常也称为 n 维欧几里得空间。实内积空间也可能是无穷维，此时如果对应的范数具备完备性，则称为希尔伯特空间。

例子 5.1.3. $V=\mathbb{R}^n$，我们定义两个向量的内积为

$$\langle\vec{x},\vec{y}\rangle=\sum_{i=1}^{n}x_iy_i$$

这里 x_i,y_i 是向量 $\vec{x},\vec{y}$ 的分量。此时

$$\sqrt{\langle\vec{x},\vec{x}\rangle}=\sqrt{\sum_{i=1}^{n}x_i^2}$$

表示向量 $\vec{x}$ 的欧氏长度。容易验证，$(\mathbb{R}^n,\langle\cdot,\cdot\rangle)$ 构成一个实内积空间，称为 n 维标准欧几里得空间。如果我们把 $\mathbb{R}^n$ 中的向量写成列向量的样子，则这个内积可以用矩阵乘法表示为

$$\langle\vec{x},\vec{y}\rangle=\vec{x}^T\cdot\vec{y}=\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

这里 $\vec{x}^T$ 是 $\vec{x}$ 的转置。

例子 5.1.4. $V=\mathbb{R}^n$，设 $a_i>0$ 是正实数。我们定义两个向量的内积为

$$\langle\vec{x},\vec{y}\rangle=\sum_{i=1}^{n}a_ix_iy_i$$

对称性和线性性显然满足。由

$$\langle\vec{x},\vec{x}\rangle=\sum_{i=1}^{n}a_ix_i^2$$

可以很容易看出正定性。令 A 为对角方阵

$$A=\begin{bmatrix}a_1 & 0 & \cdots & 0\\ 0 & a_2 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & \cdots & 0 & a_n\end{bmatrix}$$

则这个内积也可以写成矩阵形式

$$\langle\vec{x},\vec{y}\rangle=\vec{x}^TA\vec{y}=\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}\begin{bmatrix}a_1 & 0 & \cdots & 0\\ 0 & a_2 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & \cdots & 0 & a_n\end{bmatrix}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

例子 5.1.5. 设 V 是由有限闭区间 $[a,b]$ 上实连续函数构成的线性空间。对任意两个函数 $f(x),g(x)\in V$，我们定义它们的内积为

$$\langle f,g\rangle:=\int_a^b f(x)g(x)dx$$

对称性和线性性显然满足。由

$$\langle f,f\rangle:=\int_a^b f(x)^2dx$$

知

$$\langle f,f\rangle \geq 0$$

且 $\langle f,f\rangle = 0$ 当且仅当 $f(x)=0$ 是零函数，即 V 中的零向量。因此 $\langle\cdot,\cdot\rangle$ 定义了函数空间 V 上的一个内积。这个内积空间是无穷维的。

5.1.2 Gram 矩阵

设 V 是 n 维实线性空间，我们如何具体来刻画 V 上的一个内积 $\langle\cdot,\cdot\rangle$? 回顾对于一个线性映射 $f:V\to V$，我们具体描述 f 的方法是取 V 的一组基，根据 f 在基上的作用得到一个矩阵，则这个矩阵完全刻画了 f 本身。我们可以用类似的想法来刻画一个内积。设

$$\{\vec{\alpha}_1,\vec{\alpha}_2,\cdots,\vec{\alpha}_n\}$$

是 V 的一组基。考虑这些基向量之间的内积

$$\text{定义}\quad G_{ij}:=\langle\vec{\alpha}_i,\vec{\alpha}_j\rangle,\qquad i,j=1,\cdots,n$$

由此我们得到一个 n 阶方阵 $G=(G_{ij})$。

定义 5.1.6. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间，$\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 为 V 的一组基。我们称 n 阶方阵

$$G=\begin{bmatrix}\langle\vec{\alpha}_1,\vec{\alpha}_1\rangle & \langle\vec{\alpha}_1,\vec{\alpha}_2\rangle & \cdots & \langle\vec{\alpha}_1,\vec{\alpha}_n\rangle\\ \langle\vec{\alpha}_2,\vec{\alpha}_1\rangle & \langle\vec{\alpha}_2,\vec{\alpha}_2\rangle & \cdots & \langle\vec{\alpha}_2,\vec{\alpha}_n\rangle\\ \cdots & \cdots & \cdots & \cdots\\ \langle\vec{\alpha}_n,\vec{\alpha}_1\rangle & \langle\vec{\alpha}_n,\vec{\alpha}_2\rangle & \cdots & \langle\vec{\alpha}_n,\vec{\alpha}_n\rangle\end{bmatrix}$$

为内积 $\langle\cdot,\cdot\rangle$ 在基 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 下的 *Gram* 矩阵。

我们下面说明内积的 Gram 矩阵完全刻画了内积本身，因此给出了内积的矩阵表达方法。

设 $G=(G_{ij})$ 是内积 $\langle\cdot,\cdot\rangle$ 在 V 的一组基 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 下的 Gram 矩阵。对 V 中任意两个向量 $\vec{x},\vec{y}$，它们通过这组基的线性组合记为

$$\begin{aligned}\vec{x}&=x_1\vec{\alpha}_1+\cdots x_n\vec{\alpha}_n\\ \vec{y}&=y_1\vec{\alpha}_1+\cdots y_n\vec{\alpha}_n\end{aligned}$$

我们计算内积 $\langle\vec{x},\vec{y}\rangle$。利用内积的线性性

$$\begin{aligned}\langle\vec{x},\vec{y}\rangle=&\langle\sum_i x_i\vec{\alpha}_i,\vec{y}\rangle=\sum_i x_i\langle\vec{\alpha}_i,\vec{y}\rangle\\ =&\sum_i x_i\langle\vec{\alpha}_i,\sum_j y_j\vec{\alpha}_j\rangle\\ =&\sum_{i,j}x_iy_j\langle\vec{\alpha}_i,\vec{\alpha}_j\rangle\\ =&\sum_{i,j}x_iy_jG_{ij}\end{aligned}$$

写成矩阵的形式，即为

$$\langle\vec{x},\vec{y}\rangle=\begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}G\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

这说明如果我们知道了基向量之间的内积（即 Gram 矩阵），则上述公式给出了任意两个向量之间的内积。因此内积完全由其 Gram 矩阵确定。

我们下面讨论 Gram 矩阵的性质。首先，由内积的对称性

$$G_{ij}=\langle\vec{\alpha}_i,\vec{\alpha}_j\rangle=\langle\vec{\alpha}_j,\vec{\alpha}_i\rangle=G_{ji}$$

写成矩阵的形式

$$G=G^T$$

即 G 是一个实对称矩阵。

其次，由内积的正定性，对任意向量 $\vec{x}=\sum_i x_i\vec{\alpha}_i$

$$\langle\vec{x},\vec{x}\rangle=\sum_{i,j}G_{ij}x_ix_j\geq 0$$

且等号成立当且仅当 $x_1=\cdots=x_n=0$。

定义 5.1.7. *一个 n 阶实对称方阵 A 称为正定矩阵，如果对任意非零列向量 $\vec{x}\in\mathbb{R}^n$，都有*

$$x^TAx>0 \quad \text{即} \quad \sum_{i,j=1}^n A_{ij}x_ix_j>0$$

因此 Gram 矩阵 G 是一个正定矩阵。反之，给定一个 n 阶正定矩阵 A，对 V 中任意两个向量 $\vec{x}=\sum_i x_i\vec{\alpha}_i,\vec{y}=\sum_i y_i\vec{\alpha}_i$，我们定义

$$\langle\vec{x},\vec{y}\rangle_A:=\sum_{i,j=1}^n A_{ij}x_iy_j$$

容易验证 $\langle\cdot,\cdot\rangle_A$ 定义了 V 上的一个内积。因此在给定 V 的一组基的情况下，Gram 矩阵

$$\text{内积}\ \overset{\text{Gram}}{\Longleftrightarrow}\ \text{正定矩阵}$$

给出了内积和正定矩阵之间的一一对应。

5.1.3 长度与夹角

定义 5.1.8. *设 $(V,\langle\cdot,\cdot\rangle)$ 是一个欧几里得空间。V 中向量 $\vec{x}$ 的长度 (或范数) 定义为*

$$\|\vec{x}\|=\sqrt{\langle x,x\rangle}$$

命题 5.1.9 (Cauchy-Schwarz 不等式)**.** *设 $(V,\langle\cdot,\cdot\rangle)$ 是一个欧几里得空间。则对任意 $\vec{x},\vec{y}\in V$*

$$\langle\vec{x},\vec{y}\rangle^2\leq\langle\vec{x},\vec{x}\rangle\langle\vec{y},\vec{y}\rangle \qquad \text{即} \qquad |\langle\vec{x},\vec{y}\rangle|\leq\|\vec{x}\|\|\vec{y}\|$$

等号成立当且仅当 $\vec{x},\vec{y}$ 线性相关。

证明：不妨设 $\vec{x},\vec{y}$ 均是非零向量。记

$$\begin{cases} a=\langle\vec{x},\vec{x}\rangle>0 \\ b=\langle\vec{x},\vec{y}\rangle \\ c=\langle\vec{y},\vec{y}\rangle>0 \end{cases}$$

需证明 $b^2\le ac$。任取 $\lambda\in\mathbb{R}$，考虑向量 $\beta=\lambda\vec{x}+\vec{y}$ 与自己的内积

$$\begin{aligned}\langle\vec{\beta},\vec{\beta}\rangle=&\langle\lambda\vec{x}+\vec{y},\lambda\vec{x}+\vec{y}\rangle\\ =&\lambda^2\langle\vec{x},\vec{x}\rangle+\lambda\langle\vec{x},\vec{y}\rangle+\lambda\langle\vec{y},\vec{x}\rangle+\langle\vec{y},\vec{y}\rangle\\ =&a\lambda^2+2b\lambda+c\\ =&a(\lambda+b/a)^2+\frac{ac-b^2}{a}\end{aligned}$$

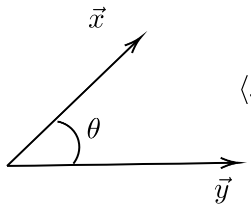
由内积正定性，$\langle\vec{\beta},\vec{\beta}\rangle\ge 0$ 对任意 $\lambda\in\mathbb{R}$ 成立。因此 $b^2\le ac$，即得命题不等式。等号成立当且仅当存在 λ 使得 $\vec{\beta}=0$，即 $\vec{x},\vec{y}$ 线性相关。 □

设 $\vec{x},\vec{y}$ 是 V 中两个非零向量，由 Cauchy-Schwarz 不等式

$$-1\le\frac{\langle\vec{x},\vec{y}\rangle}{\|\vec{x}\|\|\vec{y}\|}\le 1$$

我们定义 $\vec{x},\vec{y}$ 之间的夹角 $\theta\in[0,\pi]$ 为

$$\cos\theta=\frac{\langle\vec{x},\vec{y}\rangle}{\|\vec{x}\|\|\vec{y}\|}$$



$$\langle\vec{x},\vec{y}\rangle=\|\vec{x}\|\|\vec{y}\|\cos\theta$$

定义 5.1.10. 如果向量 $\vec{x}$ 和 $\vec{y}$ 的夹角是 $\theta=\pi/2$，即

$$\langle\vec{x},\vec{y}\rangle=0$$

我们称 $\vec{x}$ 和 $\vec{y}$ 是正交的，记为 $\vec{x}\perp\vec{y}$。

命题 5.1.11 (三角不等式)**.** 对内积空间中任意两个向量 $\vec{x},\vec{y}$

$$\|\vec{x}+\vec{y}\|\le\|\vec{x}\|+\|\vec{y}\|$$

证明：

$$\begin{aligned}\|\vec{x}+\vec{y}\|^2=&\langle\vec{x}+\vec{y},\vec{x}+\vec{y}\rangle\\ =&\langle\vec{x},\vec{x}\rangle+2\langle\vec{x},\vec{y}\rangle+\langle\vec{y},\vec{y}\rangle\\ \le&\|\vec{x}\|^2+2\|\vec{x}\|\|\vec{y}\|+\|\vec{y}\|^2\\ =&(\|\vec{x}\|+\|\vec{y}\|)^2\end{aligned}$$

□

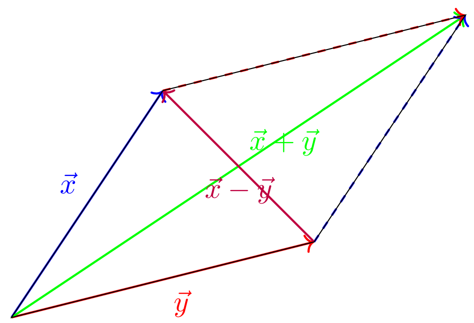
命题 5.1.12 (平行四边形法则)**.** 对内积空间中任意两个向量 $\vec{x},\vec{y}$

$$\|\vec{x}+\vec{y}\|^2+\|\vec{x}-\vec{y}\|^2=2\|\vec{x}\|^2+2\|\vec{y}\|^2$$

证明:

$$\begin{aligned}&\|\vec{x}+\vec{y}\|^2+\|\vec{x}-\vec{y}\|^2\\=&\langle\vec{x}+\vec{y},\vec{x}+\vec{y}\rangle+\langle\vec{x}-\vec{y},\vec{x}-\vec{y}\rangle\\=&\langle\vec{x},\vec{x}\rangle+2\langle\vec{x},\vec{y}\rangle+\langle\vec{y},\vec{y}\rangle+\langle\vec{x},\vec{x}\rangle-2\langle\vec{x},\vec{y}\rangle+\langle\vec{y},\vec{y}\rangle\\=&2\|\vec{x}\|^2+2\|\vec{y}\|^2\end{aligned}$$

□



$$\|\vec{x}+\vec{y}\|^2+\|\vec{x}-\vec{y}\|^2=2\|\vec{x}\|^2+2\|\vec{y}\|^2$$

5.2 正交基与正交化

设 $(V,\langle\cdot,\cdot\rangle)$ 是一个 n 维欧几里得空间。这一节我们说明可以在 V 中找一组方便的基，使得在这组基下 V 上的几何结构与 n 维标准欧几里得空间 $\mathbb{R}^n$ 一样。这样我们可以把对一般的 n 维欧几里得空间的讨论约化到标准欧几里得空间 $\mathbb{R}^n$。

5.2.1 Gram–Schmidt 正交化

命题 5.2.1. 设 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_m\}$ 是欧几里得空间 V 中 m 个两两正交的非零向量，即

$$\langle\vec{\alpha}_i,\vec{\alpha}_j\rangle=0\qquad i\neq j,\quad i,j=1,\cdots,m$$

则 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_m\}$ 线性无关。

证明：设 $\vec{\beta}=c_1\vec{\alpha}_1+\cdots+c_m\vec{\alpha}_m=0$，则对 $i=1,\cdots,m$

$$0=\langle\vec{\alpha}_i,\beta\rangle=c_i\langle\vec{\alpha}_i,\vec{\alpha}_i\rangle\quad\Longrightarrow\quad c_i=0$$

因此 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_m\}$ 线性无关。 □

这个命题说明，在 n 维欧几里得空间中两两正交的非零向量组最多包含 n 个向量。

定义 5.2.2. 欧几里得空间 V 的一组基如果由两两正交的向量构成，则称这组基为正交基。

下面这个命题给出了具体构造正交基的方法。

命题 5.2.3 (Gram–Schmidt 正交化). 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间，$\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 是 V 的一组基。则存在一组正交基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$，使得从基 $\{\vec{\alpha}_i\}$ 到 $\{\vec{\beta}_i\}$ 的过渡矩阵 P 是上三角矩阵

$$\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix}=\begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix}P$$

证明：我们依次构造 $\vec{\beta}_1,\cdots,\vec{\beta}_n$ 使得对任意 k

$$\mathrm{Span}\{\vec{\alpha}_1,\cdots,\vec{\alpha}_k\}=\mathrm{Span}\{\vec{\beta}_1,\cdots,\vec{\beta}_k\}$$

首先对 $k=1$，我们取 $\vec{\beta}_1=\vec{\alpha}_1$，显然满足

$$\mathrm{Span}\{\vec{\alpha}_1\}=\mathrm{Span}\{\vec{\beta}_1\}$$

假设我们已构造了两两正交的 $\vec{\beta}_1,\cdots,\vec{\beta}_{k-1}$ 使得

$$\mathrm{Span}\{\vec{\alpha}_1,\cdots,\vec{\alpha}_{k-1}\}=\mathrm{Span}\{\vec{\beta}_1,\cdots,\vec{\beta}_{k-1}\}$$

我们下面构造 $\vec{\beta}_k=\vec{\alpha}_k+c_1\vec{\beta}_1+\cdots+c_{k-1}\vec{\beta}_{k-1}$ 使得 $\vec{\beta}_k$ 与 $\vec{\beta}_1,\cdots,\vec{\beta}_{k-1}$ 正交，这里系数 c_i 待定。

由归纳假设，向量 $\vec{\beta}_1,\cdots,\vec{\beta}_{k-1}$ 两两正交。对 $i=1,\cdots,k-1$，要求正交性

$$\begin{aligned}\langle\vec{\beta}_k,\vec{\beta}_i\rangle=&\langle\vec{\alpha}_k+c_1\vec{\beta}_1+\cdots+c_{k-1}\vec{\beta}_{k-1},\vec{\beta}_i\rangle\\=&\langle\vec{\alpha}_k,\vec{\beta}_i\rangle+c_i\langle\vec{\beta}_i,\vec{\beta}_i\rangle=0\end{aligned}$$

可以解出系数

$$c_i=-\frac{\langle\vec{\alpha}_k,\vec{\beta}_i\rangle}{\langle\vec{\beta}_i,\vec{\beta}_i\rangle}$$

因此我们可以具体写下 $\vec{\beta}_k$ 为

$$\vec{\beta}_k=\vec{\alpha}_k-\sum_{i=1}^{k-1}\frac{\langle\vec{\alpha}_k,\vec{\beta}_i\rangle}{\langle\vec{\beta}_i,\vec{\beta}_i\rangle}\vec{\beta}_i$$

由构造，$\{\vec{\beta}_1,\cdots,\vec{\beta}_k\}$ 是两两正交的向量，且

$$\mathrm{Span}\{\vec{\alpha}_1,\cdots,\vec{\alpha}_{k-1},\vec{\alpha}_k\}=\mathrm{Span}\{\vec{\beta}_1,\cdots,\vec{\beta}_{k-1},\vec{\alpha}_k\}=\mathrm{Span}\{\vec{\beta}_1,\cdots,\vec{\beta}_{k-1},\vec{\beta}_k\}$$

由此我们依次构造了两两正交的向量 $\vec{\beta}_1,\cdots,\vec{\beta}_n$ 使得对任意的 k，

$$\vec{\beta}_k\in\mathrm{Span}\{\vec{\alpha}_1,\cdots,\vec{\alpha}_k\}$$

写成矩阵的形式，

$$\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_n\end{bmatrix}=\begin{bmatrix}\vec{\alpha}_1 & \cdots & \vec{\alpha}_n\end{bmatrix}\begin{bmatrix}* & * & \cdots & *\\ 0 & * & \cdots & *\\ \cdots & \cdots & \cdots & \cdots\\ 0 & \cdots & 0 & *\end{bmatrix}$$

即过渡矩阵是上三角矩阵。 $\square$

命题证明中由欧几里得空间中的任一组基 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 得到正交基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 的过程

$$\vec{\beta}_k=\vec{\alpha}_k-\sum_{i=1}^{k-1}\frac{\langle\vec{\alpha}_k,\vec{\beta}_i\rangle}{\langle\vec{\beta}_i,\vec{\beta}_i\rangle}\vec{\beta}_i\qquad k=1,\cdots,n$$

称为 Gram–Schmidt 正交化。

定义 5.2.4. n 维欧几里得空间 V 的一组基 $\{\vec{\gamma}_1,\cdots,\vec{\gamma}_n\}$ 称为是标准正交基，如果它们两两正交并且长度为 1

$$\|\vec{\gamma}_i\|=1\qquad i=1,\cdots,n$$

命题 5.2.5. n 维欧几里得空间 V 存在标准正交基。

证明：设 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 是 V 的任一组基。由 Gram–Schmidt 正交化，我们得到一组正交基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$。通过归一化定义

$$\vec{\gamma}_i=\frac{\vec{\beta}_i}{\|\vec{\beta}_i\|}\qquad i=1,\cdots,n$$

则 $\{\vec{\gamma}_1,\cdots,\vec{\gamma}_n\}$ 是一组标准正交基。 $\square$

设 $\{\vec{\gamma}_1,\cdots,\vec{\gamma}_n\}$ 是 V 的标准正交基，则其对应的 Gram 矩阵为

$$G_{ij}=\langle\vec{\gamma}_i,\vec{\gamma}_j\rangle=\delta_{ij}$$

即 $G=I_n$ 是单位矩阵。对于任意两个向量 $\vec{x}=\sum_{i=1}^n x_i\vec{\gamma}_i$ 和 $\vec{y}=\sum_{i=1}^n y_i\vec{\gamma}_i$，我们有

$$\langle\vec{x},\vec{y}\rangle=\sum_{i=1}^n x_iy_i$$

因此在标准正交基展开的坐标下，向量的内积公式和标准欧几里得空间一样。

例子 5.2.6. 考虑标准欧几里得空间 $\mathbb{R}^3$ 的一组基

$$\vec{\alpha}_1=\begin{bmatrix}1\\1\\1\end{bmatrix}\quad\vec{\alpha}_2=\begin{bmatrix}0\\1\\1\end{bmatrix}\quad\vec{\alpha}_3=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

这组基不是正交基。我们对这组基作 *Gram-Schmidt* 正交化。依次有

$$\vec{\beta}_1=\vec{\alpha}_1$$

$$\vec{\beta}_2=\vec{\alpha}_2-\frac{\langle\vec{\alpha}_2,\vec{\beta}_1\rangle}{\langle\vec{\beta}_1,\vec{\beta}_1\rangle}\vec{\beta}_1=\begin{bmatrix}0\\1\\1\end{bmatrix}-\frac{2}{3}\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}-2/3\\1/3\\1/3\end{bmatrix}$$

$$\begin{aligned}\vec{\beta}_3 =& \vec{\alpha}_3 - \frac{\langle \vec{\alpha}_3, \vec{\beta}_1\rangle}{\langle \vec{\beta}_1, \vec{\beta}_1\rangle}\vec{\beta}_1 - \frac{\langle \vec{\alpha}_3, \vec{\beta}_2\rangle}{\langle \vec{\beta}_2, \vec{\beta}_2\rangle}\vec{\beta}_2 \\ =& \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1/3}{2/3}\begin{bmatrix} -2/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -1/2 \\ 1/2 \end{bmatrix}\end{aligned}$$

进一步作归一化，我们得到标准正交基

$$\vec{\gamma}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \quad \vec{\gamma}_2 = \begin{bmatrix} -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} \quad \vec{\gamma}_3 = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

这组标准正交基与 $\mathbb{R}^3$ 的标准基 $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ 是不一样的。这个例子也说明标准正交基并不唯一。

5.2.2 正交投影

定义 5.2.7. 设 $(V, \langle\cdot,\cdot\rangle)$ 是一个 n 维欧几里得空间，$U \subset V$ 是线性子空间。我们定义 U 在 V 中的正交补空间为

$$U^{\perp} := \{\vec{\alpha} \in V | \langle \vec{\alpha}, \vec{\beta}\rangle = 0 \text{ 对 } U \text{ 中任意向量 } \vec{\beta} \text{ 成立}\}$$

即 $\vec{\alpha} \in U^{\perp}$ 如果 $\vec{\alpha}$ 与 U 中的任意向量都正交。

我们选 U 中的一组基 $\vec{\beta}_1, \cdots, \vec{\beta}_m$，则 $\vec{\alpha} \in U^{\perp}$ 当且仅当 $\vec{\alpha}$ 与这组基向量正交

$$\langle \vec{\alpha}, \vec{\beta}_i\rangle = 0 \qquad i = 1, \cdots, m$$

如果上式满足，则对于 U 中的任意向量 $\vec{\beta}$，把它按照基做线性展开：$\vec{\beta} = a_1\vec{\beta}_1 + \cdots + a_m\vec{\beta}_m$。则

$$\langle \vec{\alpha}, \vec{\beta}\rangle = a_1\langle \vec{\alpha}, \vec{\beta}_1\rangle + \cdots + a_m\langle \vec{\alpha}, \vec{\beta}_m\rangle = 0$$

例子 5.2.8. $\mathbb{R}^4$ 是 *4* 维标准欧几里得空间。考虑线性子空间

$$U_1 = \mathrm{Span}\{\vec{e}_1\} \qquad U_2 = \mathrm{Span}\{\vec{e}_1, \vec{e}_2\}$$

则

$$U_1^{\perp} = \mathrm{Span}\{\vec{e}_2, \vec{e}_3, \vec{e}_4\} \qquad U_2^{\perp} = \mathrm{Span}\{\vec{e}_3, \vec{e}_4\}$$

命题 5.2.9. 设 $(V, \langle\cdot,\cdot\rangle)$ 是一个 n 维欧几里得空间，$U \subset V$ 是线性子空间。则我们有直和分解

$$V = U \oplus U^{\perp}$$

证明：将内积 $\langle\cdot,\cdot\rangle$ 限制在 U 上，易知 $(U, \langle\cdot,\cdot\rangle)$ 是一个欧几里得空间。由命题5.2.5，我们取 $\vec{\beta}_1, \cdots, \vec{\beta}_m$ 为 U 的一组标准正交基。对 V 中任意向量 $\vec{x}$，定义

$$\vec{x}^{\parallel} := \langle \vec{x}, \vec{\beta}_1\rangle\vec{\beta}_1 + \cdots + \langle \vec{x}, \vec{\beta}_m\rangle\vec{\beta}_m \qquad \vec{x}^{\perp} = \vec{x} - \vec{x}^{\parallel}$$

则 $\vec{x}^{\parallel} \in U$ 且

$$\langle \vec{x}^{\perp}, \vec{\beta}_i\rangle = \langle \vec{x}, \vec{\beta}_i\rangle - \langle \vec{x}^{\parallel}, \vec{\beta}_i\rangle = 0 \quad i = 1, \cdots, m$$

即 $\vec{x}^{\perp} \in U^{\perp}$。因此我们得到分解

$$\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp} \qquad \vec{x}^{\parallel} \in U, \quad \vec{x}^{\perp} \in U^{\perp}$$

由 $\vec{x}$ 的任意性，说明 $V = U + U^{\perp}$。我们下面证明这个是直和。

假设有 $\beta \in U, \gamma \in U^{\perp}$ 使得 $\vec{\beta} + \vec{\gamma} = 0$。则

$$\langle \vec{\beta}, \vec{\beta} \rangle = -\langle \vec{\beta}, \vec{\gamma} \rangle = 0 \quad \Longrightarrow \quad \vec{\beta} = 0$$

因此 $\vec{\beta} = \vec{\gamma} = 0$。这说明 $V = U \oplus U^{\perp}$。 □

定义 5.2.10. 设 $(V, \langle \cdot, \cdot \rangle)$ 是一个 n 维欧几里得空间，$U \subset V$ 是线性子空间。任意向量 $\vec{x} \in V$ 可以唯一地分解为

$$\vec{x} = \vec{x}^{\parallel} + \vec{x}^{\perp} \quad 其中 \quad \vec{x}^{\parallel} \in U, \quad \vec{x}^{\perp} \in U^{\perp}$$

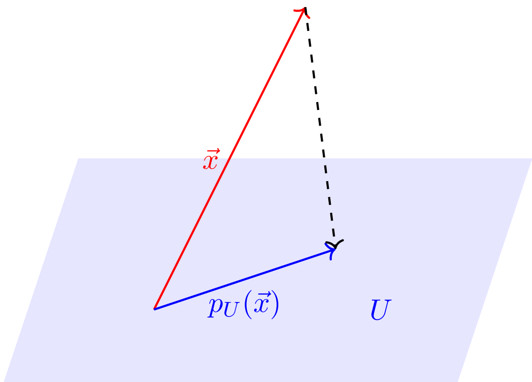
称为 $\vec{x}$ 关于 U 的正交分解。线性映射

$$p_U : V \to U, \qquad \vec{x} \to \vec{x}^{\parallel}$$

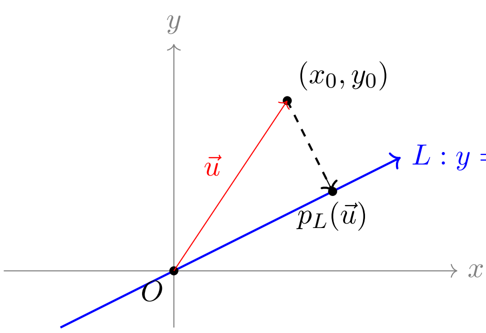
称为 V 向 U 的正交投影。

由上述命题的证明，如果取 $\vec{\beta}_1, \cdots, \vec{\beta}_m$ 为 U 的一组标准正交基，则正交投影 p_U 为

$$p_U(\vec{x}) = \langle \vec{x}, \vec{\beta}_1 \rangle \vec{\beta}_1 + \cdots + \langle \vec{x}, \vec{\beta}_m \rangle \vec{\beta}_m$$



例子 5.2.11. 考虑平面 $\mathbb{R}^2$ 的由直线 $L : \{(x, y) \in \mathbb{R}^2 | y = ax\}$ 构成的一维子空间。考虑向量 $\vec{u} = (x_0, y_0)$ 向 L 的正交投影。



直线 $L:\{(x,y)\in\mathbb{R}^2|y=ax\}$ 的标准正交基是一个 L 中的单位向量，可以选为

$$\vec{\beta}=\left(\frac{1}{\sqrt{1+a^2}},\frac{a}{\sqrt{1+a^2}}\right)$$

因此 $\vec{u}$ 向 L 的正交投影为

$$p_L(\vec{u})=\langle\vec{u},\vec{\beta}\rangle\vec{\beta}=\left(\frac{x_0+ay_0}{1+a^2},\frac{(x_0+ay_0)a}{1+a^2}\right)$$

命题 5.2.12. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间，$U\subset V$ 是线性子空间，$\vec{x}$ 是 V 中向量。则

$$\|\vec{x}-p_U(\vec{x})\|=\min_{\vec{y}\in U}\|\vec{x}-\vec{y}\|$$

即 $p_U(\vec{x})$ 是 U 中与 $\vec{x}$ 的距离最短的向量。

证明：设 $\vec{x}=\vec{x}^{\parallel}+\vec{x}^{\perp}$ 是关于 U 的正交分解，即

$$p_U(\vec{x})=\vec{x}^{\parallel}\in U,\qquad \vec{x}^{\perp}\in U^{\perp}$$

对任意 $\vec{y}\in U$，$\vec{x}^{\parallel}-\vec{y}\in U$ 因此与 $\vec{x}^{\perp}$ 正交。故

$$\begin{aligned}\|\vec{x}-\vec{y}\|^2=&\langle\vec{x}^{\parallel}+\vec{x}^{\perp}-\vec{y},\vec{x}^{\parallel}+\vec{x}^{\perp}-\vec{y}\rangle\\=&\langle\vec{x}^{\parallel}-\vec{y},\vec{x}^{\parallel}-\vec{y}\rangle+2\langle\vec{x}^{\parallel}-\vec{y},\vec{x}^{\perp}\rangle+\langle\vec{x}^{\perp},\vec{x}^{\perp}\rangle\\=&\|\vec{x}^{\parallel}-\vec{y}\|^2+\|\vec{x}^{\perp}\|^2\geq\|\vec{x}^{\perp}\|^2\end{aligned}$$

等号当且仅当 $\vec{y}=\vec{x}^{\parallel}=p_U(\vec{x})$ 成立。 □

等价而言，如果向量 $\vec{x}$ 关于 U 的正交分解为

$$\vec{x}=\vec{x}^{\parallel}+\vec{x}^{\perp}$$

则 $\vec{x}$ 与 U 中向量间的最短距离为分量 $\vec{x}^{\perp}$ 的长度

$$\min_{\vec{y}\in U}\|\vec{x}-\vec{y}\|=\|\vec{x}^{\perp}\|$$

这个值也称为向量 $\vec{x}$ 与子空间 U 的距离。

5.3 正交方阵

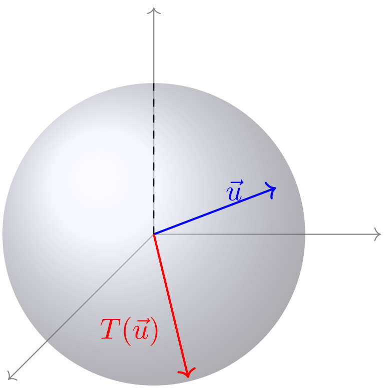
5.3.1 正交变换与正交方阵

定义 5.3.1. 设 V 是一个欧几里得空间。线性变换 $T:V\to V$ 称为正交变换，如果对于任意向量 $\vec{x}\in V$，都有

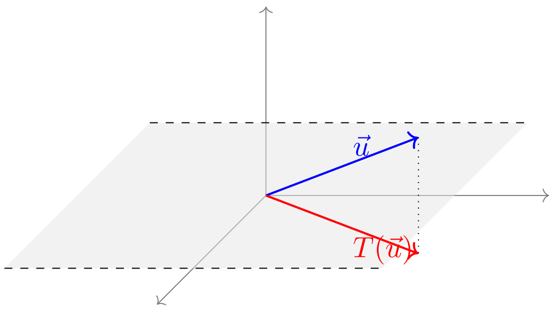
$$\|T(\vec{x})\|=\|\vec{x}\|$$

即正交变换保持向量的范数（长度）不变。

例子 5.3.2. $\mathbb{R}^3$ 中的转动保持长度，是一个正交变换



例子 5.3.3. $\mathbb{R}^3$ 中沿着过原点的平面作反射是一个正交变换



我们在5.3.2节中将更详细地讨论转动和反射的性质。

命题 5.3.4. 欧几里得空间 V 上的线性变换 $T : V \to V$ 是一个正交变换当且仅当 T 保持内积不变，即对 V 中的任意向量 $\vec{x}, \vec{y}$

$$\langle T(\vec{x}), T(\vec{y})\rangle = \langle \vec{x}, \vec{y}\rangle$$

证明：若 T 保持内积，则由 $\langle T(\vec{x}), T(\vec{x})\rangle = \langle \vec{x}, \vec{x}\rangle$ 知 T 保长度，即为正交变换。

反之，假设 T 是正交变换，即保长度。由

$$\langle \vec{x}+\vec{y}, \vec{x}+\vec{y}\rangle = \langle \vec{x}, \vec{x}\rangle + \langle \vec{y}, \vec{y}\rangle + 2\langle \vec{x}, \vec{y}\rangle$$

知 $\langle \vec{x}, \vec{y}\rangle = \frac{1}{2}\|\vec{x}+\vec{y}\|^2 - \frac{1}{2}\|\vec{x}\|^2 - \frac{1}{2}\|\vec{y}\|^2$。因此

$$\begin{aligned}\langle T(\vec{x}), T(\vec{y})\rangle =& \frac{1}{2}\|T(\vec{x})+T(\vec{y})\|^2 - \frac{1}{2}\|T(\vec{x})\|^2 - \frac{1}{2}\|T(\vec{y})\|^2\\ =& \frac{1}{2}\|T(\vec{x}+\vec{y})\|^2 - \frac{1}{2}\|T(\vec{x})\|^2 - \frac{1}{2}\|T(\vec{y})\|^2\\ =& \frac{1}{2}\|\vec{x}+\vec{y}\|^2 - \frac{1}{2}\|\vec{x}\|^2 - \frac{1}{2}\|\vec{y}\|^2 = \langle \vec{x}, \vec{y}\rangle\end{aligned}$$

即 T 保持内积。 □

特别地，两个向量 $\vec{x}, \vec{y}$ 之间的夹角 $\theta \in [0, \pi]$ 为

$$\cos\theta = \frac{\langle \vec{x}, \vec{y}\rangle}{\|\vec{x}\|\|\vec{y}\|}$$

因此正交变换也保持向量间的夹角不变。

下面我们通过选一组基来刻画正交变换。欧几里得空间中比较好的一组基是标准正交基。

命题 5.3.5. 设 $\{\vec{\gamma}_1,\cdots,\vec{\gamma}_n\}$ 是 n 维欧几里得空间 V 的一组标准正交基，$T:V\to V$ 是线性映射。则 T 是正交变换当且仅当

$$\{T(\vec{\gamma}_1),\cdots,T(\vec{\gamma}_n)\}$$

也是 V 的标准正交基。

证明：假设 $T:V\to V$ 是正交变换，则 T 保内积

$$\langle T(\vec{\gamma}_i),T(\vec{\gamma}_j)\rangle=\langle\vec{\gamma}_i,\vec{\gamma}_j\rangle=\delta_{ij}$$

因此 $\{T(\vec{\gamma}_1),\cdots,T(\vec{\gamma}_n)\}$ 是一组标准正交基。

反之，假设 $\{T(\vec{\gamma}_1),\cdots,T(\vec{\gamma}_n)\}$ 是一组标准正交基。对任意向量 $\vec{x}\in V$，按照基展开

$$\vec{x}=x_1\vec{\gamma}_1+\cdots+x_n\vec{\gamma}_n$$

则

$$\begin{aligned}\langle T(\vec{x}),T(\vec{x})\rangle=&\langle T(\sum_i x_i\vec{\gamma}_i),T(\sum_j x_j\vec{\gamma}_j)\rangle\\=&\sum_{i,j}x_ix_j\langle T(\vec{\gamma}_i),T(\vec{\gamma}_j)\rangle\\=&\sum_i x_i^2=\langle\vec{x},\vec{x}\rangle\end{aligned}$$

即 T 保长度，因此是正交变换。 □

这个命题说明正交变换给出了 V 中不同标准正交基之间的联系。设 $T:V\to V$ 是正交变换，$\{\vec{\gamma}_1,\cdots,\vec{\gamma}_n\}$ 是标准正交基。我们考虑基变换的过渡矩阵

$$\begin{bmatrix}T(\vec{\gamma}_1)&\cdots&T(\vec{\gamma}_n)\end{bmatrix}=\begin{bmatrix}\vec{\gamma}_1&\cdots&\vec{\gamma}_n\end{bmatrix}A$$

或写成分量的形式 $T(\vec{\gamma}_j)=\sum\limits_k a_{kj}\vec{\gamma}_k$。等价而言，$A$ 即为 T 在标准基 $\{\vec{\gamma}_1,\cdots,\vec{\gamma}_n\}$ 下的表示矩阵。代入计算内积

$$\begin{aligned}\langle T(\vec{\gamma}_i),T(\vec{\gamma}_j)\rangle=&\langle\sum_k a_{ki}\vec{\gamma}_k,\sum_l a_{lj}\vec{\gamma}_l\rangle\\=&\sum_{k,l}a_{ki}a_{lj}\langle\vec{\gamma}_k,\vec{\gamma}_l\rangle\\=&\sum_k a_{ki}a_{kj}\end{aligned}$$

T 是正交变换等价于

$$\langle T(\vec{\gamma}_i),T(\vec{\gamma}_j)\rangle=\delta_{ij}$$

即

$$\sum_k a_{ki}a_{kj}=\delta_{ij}$$

写成矩阵的形式，上式等价于

$$A^TA=I_n$$

定义 5.3.6. 一个 n 阶实矩阵 A 称为正交方阵，如果它满足

$$A^TA = AA^T = I_n$$

其中 A^T 表示 A 的转置矩阵，I_n 表示 n 阶单位矩阵。所有实正交矩阵组成的集合记为 $\mathrm{O}(n)$。

A 是正交方阵的一个等价写法是

$$A^T = A^{-1}$$

即 A 可逆且 A 的逆是它的转置。由上述讨论，我们证明了如下命题

命题 5.3.7. 设 $T: V \to V$ 是 n 维欧几里得空间 V 的线性变换。则 T 是正交变换当且仅当 T 在 V 的一组标准正交基下的表示矩阵 A 是 n 阶正交方阵。

设 $A \in \mathrm{O}(n)$ 是 n 阶正交方阵。记 A 的列向量为 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$，即

$$A = \begin{bmatrix} \vec{\beta}_1 & \cdots & \vec{\beta}_n \end{bmatrix} \qquad A^T = \begin{bmatrix} \vec{\beta}_1^T \\ \vdots \\ \vec{\beta}_n^T \end{bmatrix}$$

则

$$A^TA = \begin{bmatrix} \vec{\beta}_1^T\vec{\beta}_1 & \vec{\beta}_1^T\vec{\beta}_2 & \cdots & \vec{\beta}_1^T\vec{\beta}_n \\ \vec{\beta}_2^T\vec{\beta}_1 & \vec{\beta}_2^T\vec{\beta}_2 & \cdots & \vec{\beta}_2^T\vec{\beta}_n \\ \cdots & \cdots & \cdots & \cdots \\ \vec{\beta}_n^T\vec{\beta}_1 & \vec{\beta}_n^T\vec{\beta}_2 & \cdots & \vec{\beta}_n^T\vec{\beta}_n \end{bmatrix}$$

因此 $A^TA = I_n$ 等价于

$$\vec{\beta}_i^T\vec{\beta}_j = \langle \vec{\beta}_i, \vec{\beta}_j \rangle = \delta_{ij}$$

即 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$ 是标准欧几里得空间 $\mathbb{R}^n$ 的标准正交基。

类似的，我们把 A 的行向量记为 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$，即

$$A = \begin{bmatrix} \vec{\alpha}_1 \\ \vdots \\ \vec{\alpha}_n \end{bmatrix} \qquad A^T = \begin{bmatrix} \vec{\alpha}_1^T & \cdots & \vec{\alpha}_n^T \end{bmatrix}$$

则 $AA^T = I_n$ 等价于行向量 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 是标准欧几里得空间 $\mathbb{R}^n$ 的标准正交基。

由此我们证明了如下命题

命题 5.3.8. 设 A 是 n 阶实方阵，则如下条件等价：

1. A 是正交方阵

2. A 的列向量构成 $\mathbb{R}^n$ 的标准正交基

3. A 的行向量构成 $\mathbb{R}^n$ 的标准正交基

例子 5.3.9. 由 $\mathbb{R}^3$ 的一组基

$$\vec{\alpha}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \vec{\alpha}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \vec{\alpha}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

做 *Gram-Schmidt* 正交化和归一化（见例5.2.6），得到标准正交基

$$\vec{\gamma}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \quad \vec{\gamma}_2 = \begin{bmatrix} -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} \quad \vec{\gamma}_3 = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

因此我们得到如下 3 阶正交方阵

$$\begin{bmatrix} 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \end{bmatrix}$$

命题 5.3.10. 设 A, B 是 n 阶正交方阵，则 $A^{-1}(= A^T), AB$ 均为 n 阶正交方阵。我们说 $\mathrm{O}(n)$ 构成一个群，称为 n 阶正交群。

证明：$AA^T = A^TA = I_n$ 知 $A^T \in \mathrm{O}(n)$，即 $A^{-1} \in \mathrm{O}(n)$。由

$$(AB)^T(AB) = B^TA^TAB = B^TB = I_n$$

知 $AB \in \mathrm{O}(n)$。 □

设 $A \in \mathrm{O}(n)$ 是 n 阶正交方阵。由

$$A^TA = I_n$$

两边取行列式得

$$\det(A^T)\det(A) = 1 \quad 即 \quad (\det A)^2 = 1$$

因此

$$\det A = \pm 1$$

定义 5.3.11. n 阶方阵 A 称为特殊正交方阵，如果 A 是正交方阵并且 $\det A = 1$。所有 n 阶特殊正交方阵构成的集合记为 $\mathrm{SO}(n)$，即

$$\mathrm{SO}(n) = \{A \in \mathrm{O}(n) | \det A = 1\}$$

$\mathrm{SO}(n)$ 也构成一个群，称为 n 阶特殊正交群。

5.3.2 反射与旋转

例子 5.3.12 (反射变换)**.** 设 $\vec{u}\in\mathbb{R}^n$ 是非零向量。考虑

$$H=\{\vec{y}\in\mathbb{R}^n|\,\langle\vec{y},\vec{u}\rangle=0\}$$

为所有与 $\vec{u}$ 垂直的向量构成的集合。H 可以看作是 $\mathbb{R}^n$ 中的一个 $n-1$ 维超平面，$\vec{u}$ 是 H 的法向量。我们考虑 $\mathbb{R}^n$ 中关于 H 的反射变换 $R_{\vec{u}}$。具体而言

$$R_{\vec{u}}(\vec{x})=\vec{x}-2\frac{\langle\vec{u},\vec{x}\rangle}{\langle\vec{u},\vec{u}\rangle}\vec{u}$$

可以验证，$R_{\vec{u}}$ 是一个线性变换，将 $\vec{u}$ 变为 $-\vec{u}$，并且保持超平面 H 不动。易知这是一个等距变换，对应于一个正交矩阵。

为简化讨论，不妨设 $\vec{u}$ 为单位向量，即 $\langle\vec{u},\vec{u}\rangle=1$。容易验证，此时 $R_{\vec{u}}$ 对应的正交矩阵为

$$I_n-2\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}\begin{bmatrix}u_1&u_2&\cdots&u_n\end{bmatrix}=\begin{bmatrix}1-2u_1u_1&-2u_1u_2&\cdots&-2u_1u_n\\-2u_2u_1&1-2u_2u_2&\cdots&-2u_2u_n\\\vdots&\cdots&\cdots&\vdots\\-2u_nu_1&-2u_nu_2&\cdots&1-2u_nu_n\end{bmatrix}$$

特别的，如果 $\vec{u}=\vec{e_i}$ 为第 i 个坐标方向的单位向量，则

$$R_{\vec{e_i}}=\begin{bmatrix}1&0&\cdots&\cdots&\cdots&0\\0&1&\cdots&\cdots&\cdots&0\\\vdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&\cdots&\cdots&-1&\cdots&0\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&\cdots&\cdots&\cdots&\cdots&1\end{bmatrix}$$

其中第 i 行 i 列的元素为 -1，其它对角元为 1。

反射的行列式是 -1，即

$$\det R_{\vec{u}}=-1$$

例子 5.3.13 (2 维正交变换)**.** 考虑平面 $\mathbb{R}^2$ 中的正交变换。设

$$A=\begin{bmatrix}a&b\\c&d\end{bmatrix}\in\mathrm{O}(2)$$

由

$$A^TA=\begin{bmatrix}a&c\\b&d\end{bmatrix}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\begin{bmatrix}a^2+c^2&ab+cd\\ab+cd&b^2+d^2\end{bmatrix}$$

我们得到方程组

$$a^2+c^2=b^2+d^2=1\,,\quad ab+cd=0$$

第一组方程的解为

$$(a,c)=(\cos\theta,\sin\theta)\,,\quad(b,d)=(\sin\phi,\cos\phi)$$

代入第二组方程，我们得到 $\sin(\theta+\phi)=0$，即 $\phi=-\theta$ 或 $-\theta+\pi$。所以我们得到解

$$A=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$$

或

$$A=\begin{bmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{bmatrix}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}.$$

容易看出，前者属于 SO(2)，表示为逆时针转 θ 的旋转；后者 $\det A=-1$，表示为与 x 轴夹角为 $\theta/2$ 的直线的反射。由矩阵的行列式我们很容易知道两个平面反射的复合为一个旋转。

5.3.3 Cartan–Dieudonné 定理

我们首先考虑 $\mathbb{R}^3$ 中的正交变换。O(3) 中的元素可以分类为如下三种

1. 绕某个轴的转动。如
$$\begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}$$
为绕 z-轴逆时针旋转 θ。

2. 关于某个平面的反射。如
$$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -1\end{bmatrix}$$
为关于 xy 平面的反射。

3. 一个绕轴的转动和一个与旋转轴垂直平面的反射的复合。如
$$\begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & -1\end{bmatrix}$$
为绕 z-轴转动和 xy-平面反射的复合。

这三种分类可以通过如下的几何结论得到。

命题 5.3.14. *O(3) 中任一元素可以写成至多 3 个反射的乘积。*

证明：设 $A\in O(3)$。不妨设 $A\neq I_n$，于是存在非零向量 $\vec{x}$ 使得 $A(\vec{x})\neq\vec{x}$。我们不妨设 $\vec{x}=\vec{e}_3$ 为第 3 个标准基向量。令 $\vec{u}=A\vec{e}_3-\vec{e}_3$。考虑沿与 $\vec{u}$ 垂直平面的反射 $R_{\vec{u}}$，易知

$$R_{\vec{u}}(A\vec{e}_3)=\vec{e}_3$$

考虑乘积 $B=R_{\vec{u}}A$。由于 $B(\vec{e}_3)=\vec{e}_3$，B 形如

$$B=\begin{bmatrix}* & * & 0\\ * & * & 0\\ * & * & 1\end{bmatrix}$$

由正交性 $B^TB = I_3$ 可以得到

$$B = \begin{bmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \in \mathrm{O}(2)$$

由例5.3.13，我们已知 O(2) 中的元素可以写成不超过两个反射的乘积，于是 B 可以写成不超过两个反射的乘积。因此 $A = R_{\vec{u}}^{-1}B$ 可以写成不超过 3 个反射的乘积。 □

定理 5.3.15 (Cartan–Dieudonné 定理)**.** $\mathbb{R}^n$ 上任意正交变换都可以写成至多 n 个反射的复合。

这个定理可以通过对 n 做归纳证明，证明方法和上述 $n = 3$ 的情况类似。

5.4 正交相似与奇异值分解

5.4.1 正交相似

定义 5.4.1. 设 A 和 B 是两个 n 阶实方阵。如果存在一个 n 阶正交矩阵 Q，使得

$$B = QAQ^{-1}$$

则称矩阵 A 和 B 正交相似。由于 Q 是正交方阵，此时也可以写成 $B = QAQ^T$。

正交相似是比相似更强的关系。如果两个矩阵正交相似，则它们一定相似；但反之则不一定成立。

例子 5.4.2. 考虑如下两个矩阵

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$

由于 B 有两个不同的特征值 $1, 2$，B 可以对角化为 A，因此 A 与 B 相似。

然而 A 与 B 不是正交相似的。假如存在正交方阵 Q 使得 $B = QAQ^{-1}$。由 Q 是正交方阵，$Q^{-1} = Q^T$，我们也可以写成 $B = QAQ^T$。因此

$$B^T = (QAQ^T)^T = QA^TQ^T = QAQ^T = B$$

由此得出 B 是对称方阵，矛盾。

命题 5.4.3. 设 A 是 n 阶实方阵，且所有特征值均为实数。则 A 正交相似于一个上三角方阵。

证明：我们对 n 作归纳。设 λ_1 是 A 的一个特征值。由假设 $\lambda_1 \in \mathbb{R}$，设 $\vec{\beta}_1 \in \mathbb{R}^n$ 是属于 λ_1 的一个特征向量。不妨设 $\|\vec{\beta}_1\| = 1$。通过 Gram-Schmidts 正交化，我们可以把 $\vec{\beta}_1$ 延拓为 $\mathbb{R}^n$ 的一组标准正交基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_n\}$。把这组列向量组成 n 阶方阵

$$P = \begin{bmatrix} \vec{\beta}_1 & \cdots & \vec{\beta}_n \end{bmatrix}$$

由于 $\{\vec{\beta}_1, \cdots, \cdots, \vec{\beta}_n\}$ 是标准正交基，P 是正交方阵。由 $\vec{\beta}_1$ 是特征向量，有

$$AP = P \begin{bmatrix} \lambda_1 & * \\ 0 & B \end{bmatrix}$$

这里 B 是 $n-1$ 阶方阵。由于 A 与 $\begin{bmatrix} \lambda_1 & * \\ 0 & B \end{bmatrix}$ 相似，B 的特征值均为 A 的特征值故为实数。由归纳假设，存在 $n-1$ 阶正交方阵 Q 使得

$$B = Q \begin{bmatrix} \lambda_2 & * & \cdots & * \\ 0 & \lambda_3 & \cdots & * \\ 0 & \cdots & \cdots & * \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} Q^{-1}$$

因此

$$A = P \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ 0 & \cdots & \cdots & * \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q^{-1} \end{bmatrix} P^{-1}$$

$P \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ 是正交方阵，由归纳知命题成立。 □

5.4.2 实对称方阵

命题 5.4.4. 设 A 是实对称方阵，即 $A = A^T$。则 A 的特征值都是实数。

证明：设 λ_0 是 A 的特征值。设复列向量 $\vec{\beta} \in \mathbb{C}^n$ 是属于 λ_0 的特征向量，即 $A\vec{\beta} = \lambda_0 \vec{\beta}$。设

$$\vec{\beta} = \begin{bmatrix} z_1 \\ \cdots \\ z_n \end{bmatrix}$$

则特征向量说明

$$\sum_j A_{ij} z_j = \lambda_0 z_i$$

两边乘以 $\bar{z}_i$ 相加，得到

$$\sum_{ij} A_{ij} \bar{z}_i z_j = \lambda_0 \sum_i |z_i|^2$$

由于 A 是对称实方阵，$A_{ij} = A_{ji}$

$$\overline{\sum_{ij} A_{ij} \bar{z}_i z_j} = \sum_{ij} A_{ij} z_i \bar{z}_j = \sum_{ij} A_{ji} \bar{z}_j z_i$$

因此 $\sum\limits_{ij} A_{ij} \bar{z}_i z_j$ 是实数。故 $\lambda_0 = \frac{\sum_{ij} A_{ij} \bar{z}_i z_j}{\sum_i |z_i|^2}$ 是实数。 □

命题 5.4.5. 设 A 是实对称方阵，则 A 正交相似于对角阵。

证明：由命题5.4.4知 A 的特征值均为实数。由命题5.4.3知存在正交方阵 P 使得 $B = PAP^{-1}$ 是上三角方阵。由

$$B^T = (PAP^{-1})^T = (PAP^T)^T = PA^TP^T = PAP^T = B$$

知 B 也是对称方阵，因此 B 是对角方阵。 □

命题 5.4.6. 实对称方阵 A 是正定方阵当且仅当 A 的每个特征值均大于零。

证明：由 A 是实对称方阵，存在正交方阵 P 使得

$$A = P^{-1}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} P$$

这里 $\lambda_1, \cdots, \lambda_n$ 为 A 的所有特征值。对任意向量 $\vec{x} \in \mathbb{R}^n$，令 $\vec{y} = P\vec{x}$。则

$$\vec{x}^T A\vec{x} = \vec{y}^T PAP^{-1}\vec{y} = \vec{y}^T \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \vec{y} = \sum_{i=1}^{n} \lambda_i y_i^2$$

由此易知 $\vec{x}^T A\vec{x} > 0$ 对任意非零向量 $\vec{x}$ 成立当且仅当每个特征值 $\lambda_i > 0$。 □

定义 5.4.7. 一个 n 阶实对称方阵 A 称为半正定矩阵，如果对任意列向量 $\vec{x} \in \mathbb{R}^n$，都有

$$\vec{x}^T A\vec{x} \geq 0 \quad \text{即} \quad \sum_{i,j=1}^{n} A_{ij}x_ix_j \geq 0$$

类似的方法可以证明

命题 5.4.8. 实对称方阵 A 是半正定方阵当且仅当 A 的每个特征值均非负。

命题 5.4.9. 设 A 是 n 阶半正定对称方阵。则存在唯一的半正定对称方阵 B 使得 $A = B^2$。

证明：设 A 的特征值为 $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$。存在正交方阵 P 使得

$$A = P^{-1}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} P$$

取

$$B = P^{-1}\begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{bmatrix} P$$

则 B 是半正定对称方阵且 $B^2 = A$ 成立。下证唯一性。

假设有两个半正定对称方阵 $B, \tilde{B}$ 使得

$$B^2 = \tilde{B}^2 = A$$

设 B 和 $\tilde{B}$ 的特征值分别为 $\mu_1 \geq \cdots \geq \mu_n \geq 0$ 和 $\tilde{\mu}_1 \geq \cdots \geq \tilde{\mu}_n \geq 0$。存在正交方阵 $P, \tilde{P}$ 使得

$$B = P^{-1} \begin{bmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \mu_n \end{bmatrix} P \qquad \tilde{B} = \tilde{P}^{-1} \begin{bmatrix} \tilde{\mu}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\mu}_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \tilde{\mu}_n \end{bmatrix} \tilde{P}$$

因此 $B^2 = A$ 的特征值为 $\mu_1^2 \geq \cdots \geq \mu_n^2 \geq 0$，故 $\mu_i = \sqrt{\lambda_i}$。同理 $\tilde{\mu}_i = \sqrt{\lambda_i}$，即 B 与 $\tilde{B}$ 的特征值相同。由 $B^2 = \tilde{B}^2 = A$ 知

$$P^{-1} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} P = \tilde{P}^{-1} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \tilde{P}$$

令 $Q = \tilde{P}P^{-1}$，其为正交方阵。则

$$Q \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} Q$$

这个等式写成矩阵元 $Q = (Q_{ij})$ 的形式为

$$Q_{ij}(\lambda_i - \lambda_j) = 0 \qquad i, j = 1, \cdots, n$$

无论对 λ_i 是否等于 λ_j，我们均有

$$Q_{ij}(\sqrt{\lambda_i} - \sqrt{\lambda_j}) = 0 \qquad i, j = 1, \cdots, n$$

转化为矩阵形式，我们得到

$$Q \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{bmatrix} Q$$

带回 $Q = \tilde{P}P^{-1}$，即得

$$B = P^{-1} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{bmatrix} P = \tilde{P}^{-1} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{bmatrix} \tilde{P} = \tilde{B}$$

□

由命题5.4.9知对于半正定对称方阵 A，我们有唯一确定的半正定对称方阵 B 使得 $B^2 = A$ 成立。我们把这个半正定对称方阵记作 $\sqrt{A}$，其满足

$$(\sqrt{A})^2 = A$$

$\sqrt{A}$ 称为 A 的平方根。类似的方法可以证明，存在唯一的开 k 次方的半正定对称方阵 $\sqrt[k]{A}$。

5.4.3 奇异值分解

设 A 是 $m \times n$ 实矩阵。考虑 n 阶矩阵 A^TA。对于任意 $\vec{x} \in \mathbb{R}^n$

$$\vec{x}^T(A^TA)\vec{x} = (A\vec{x})^T(A\vec{x}) \geq 0$$

因此 A^TA 是 n 阶半正定对称方阵。同理，AA^T 是 m 阶半正定对称方阵。

由第三章习题 9 知

$$\lambda^m \det(\lambda I_n - A^TA) = \lambda^n \det(\lambda I_m - AA^T)$$

这说明 A^TA 和 AA^T 具有相同的非零特征值。

定义 5.4.10. 设 A 是 $m \times n$ 实矩阵。则 n 阶半正定对称方阵 A^TA（或 AA^T）的非零特征值的平方根称为 A 的奇异值。

设 A^TA 的所有非零特征值为 $\mu_1, \cdots, \mu_r$，则 A 的所有奇异值为 $\sqrt{\mu_1}, \cdots, \sqrt{\mu_r}$。

命题 5.4.11. 设 A 是 n 阶实对称方阵。则 A 的所有奇异值为所有非零特征值的绝对值。

证明：设 A 的所有特征值为 $\lambda_1, \cdots, \lambda_n$。由命题5.4.5知 A 可对角化。由此易知 $A^TA = A^2$ 的特征值为 $\lambda_1^2, \cdots, \lambda_n^2$。由此即得命题。 □

例子 5.4.12. 考虑方阵 $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$。

$$A^TA = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

A^TA 的特征多项式为 $\begin{vmatrix} \lambda - 1 & -1 \\ -1 & \lambda - 2 \end{vmatrix} = \lambda^2 - 3\lambda + 1$，解得 A^TA 的两个特征值为 $\frac{3\pm\sqrt{5}}{2}$。故 A 有两个奇异值

$$\sqrt{\frac{3 \pm \sqrt{5}}{2}}$$

而 A 的两个特征值均为 1。因此当方阵 A 不对称时候，奇异值和特征值的绝对值并不一样。

定理 5.4.13 (奇异值分解). 设 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 是 $m \times n$ 阶实矩阵 A 的所有奇异值。则存在 m 阶正交方阵 U 和 n 阶正交方阵 V 使得 $A = U\Sigma V^T$，其中

$$\Sigma = \begin{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$$

这里 $0_{k\times l}$ 表示 $k \times l$ 的零矩阵。

证明：由奇异值的定义，存在 n 阶正交方阵 V 使得

$$V^TA^TAV = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_r^2, 0, \cdots, 0)$$

把 $AV = \begin{bmatrix} \vec{\alpha}_1 & \cdots & \vec{\alpha}_n \end{bmatrix}$ 写成 $\mathbb{R}^m$ 中列向量，则

$$V^TA^TAV = (AV)^T(AV) = \begin{bmatrix} \vec{\alpha}_1^T \\ \vdots \\ \vec{\alpha}_n^T \end{bmatrix} \begin{bmatrix} \vec{\alpha}_1 & \cdots & \vec{\alpha}_n \end{bmatrix} = \begin{bmatrix} \vec{\alpha}_1^T\vec{\alpha}_1 & \cdots & \vec{\alpha}_1^T\vec{\alpha}_n \\ \cdots & \cdots & \cdots \\ \vec{\alpha}_n^T\vec{\alpha}_1 & \cdots & \vec{\alpha}_n^T\vec{\alpha}_n \end{bmatrix}$$

对 $i > r$，我们有

$$\vec{\alpha}_i^T\vec{\alpha}_i = 0 \quad \Longrightarrow \quad \vec{\alpha}_i = 0$$

对 $1 \leq i, j \leq r$，

$$\vec{\alpha}_i^T\vec{\alpha}_j = \delta_{ij}\sigma_i^2$$

记 $\vec{\beta}_1 = \sigma_1^{-1}\vec{\alpha}_1, \cdots, \vec{\beta}_r = \sigma_r^{-1}\vec{\alpha}_r$。则 $\{\vec{\beta}_i\}$ 两两正交且长度均为 1。我们把它们扩展为 $\mathbb{R}^m$ 中的一组标准正交基 $\{\vec{\beta}_1, \cdots, \vec{\beta}_r, \vec{\beta}_{r+1}, \cdots, \vec{\beta}_m\}$。则

$$\begin{aligned} AV &= \begin{bmatrix} \vec{\alpha}_1 & \cdots & \vec{\alpha}_r & 0 \cdots & 0 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1\vec{\beta}_1 & \cdots & \sigma_r\vec{\beta}_r & 0 \cdots & 0 \end{bmatrix} \\ &= \begin{bmatrix} \vec{\beta}_1 & \cdots & \vec{\beta}_r & \vec{\beta}_{r+1} \cdots & \vec{\beta}_m \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \end{aligned}$$

令 $U = \begin{bmatrix} \vec{\beta}_1 & \cdots & \vec{\beta}_r & \vec{\beta}_{r+1} \cdots & \vec{\beta}_m \end{bmatrix}$，则 U 是 m 阶正交方阵。上述等式即得奇异值分解。 □

5.5 一些例子

5.5.1 QR 分解

命题 5.5.1 (可逆方阵的 QR 分解). 设 A 是 n 阶可逆方阵，则存在唯一的分解

$$A = QR$$

其中 Q 是 n 阶正交方阵，R 是对角元都是正数的上三角矩阵。

证明：记 A 的列向量为 $A=\begin{bmatrix}\vec{\alpha}_1 & \cdots & \vec{\alpha}_n\end{bmatrix}$。由 Gram-Schmidt 正交化和归一化的过程，我们得到一个对角元都是正数的上三角矩阵 R 使得

$$\begin{bmatrix}\vec{\alpha}_1 & \cdots & \vec{\alpha}_n\end{bmatrix}=\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_n\end{bmatrix}R$$

这里 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 是标准正交基。记 $Q=\begin{bmatrix}\vec{\beta}_1 & \cdots & \vec{\beta}_n\end{bmatrix}$，则 Q 是正交方阵且 $A=QR$。

下证唯一性。假设有 $Q_1R_1=Q_2R_2$，这里 Q_1,Q_2 是正交阵，R_1,R_2 是对角元都是正数的上三角矩阵。则 $R_1R_2^{-1}=Q_1^{-1}Q_2$ 是正交阵并且是对角元都是正数的上三角矩阵

$$R_1R_2^{-1}=\begin{bmatrix}a_{11} & * & \cdots & * \\ 0 & a_{22} & \cdots & * \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_{nn}\end{bmatrix} \qquad a_{ii}>0$$

由列向量的标准正交性易知 $R_1R_2^{-1}$ 只能是对角阵，且 $a_{ii}=1$。即 $R_1R_2^{-1}=I_n$。由此得 $R_1=R_2,Q_1=Q_2$。 □

QR 分解有很多应用。比如考虑线性方程组

$$A\vec{x}=\vec{b}$$

这里 A 是可逆矩阵。我们知道它有唯一解 $\vec{x}=A^{-1}\vec{b}$。在实际应用中计算矩阵的逆 A^{-1} 通常很复杂。如果我们有矩阵的 QR 分解 $A=QR$，则上述方程等价于

$$R\vec{x}=Q^T\vec{b}$$

此时 R 是上三角方阵，方程很容易求解。

更一般的，QR 分解可以推广到任意长方形矩阵。

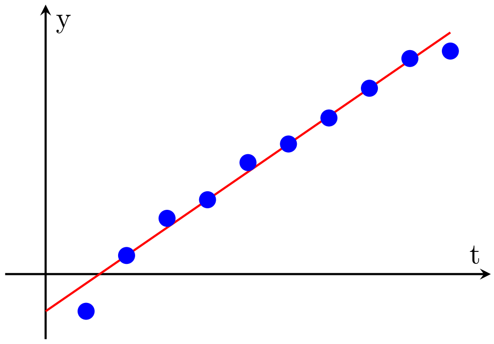
命题 5.5.2 (QR 分解)**.** 设 A 是 $m\times n$ 实矩阵，这里 $m\geq n$。则存在 m 阶正交阵 Q 和 $m\times n$ 矩阵 $R=\begin{bmatrix}R_1\\0\end{bmatrix}$，其中 R_1 是对角元非负的 n 阶上三角矩阵，使得

$$A=QR$$

这也可以通过 Gram-Schmidt 正交化证明，具体细节留作练习。

5.5.2 最小二乘法

最小二乘法（Least Squares Method）是一种优化方法，用于拟合数据点并找到最佳的拟合曲线。其基本目标是通过最小化拟合曲线与数据点之间的误差平方和来找到最佳的参数。



例如我们有一组数据 (t_i, y_i) 如图，希望用一个线性函数 $y = at + b$ 来拟合这组数据。即目标是找到 a, b 使得

$$\begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

我们把待求的拟合参数记作未知变量 $\vec{x} = \begin{bmatrix} a \\ b \end{bmatrix}$，则上述关系变成求解线性方程组 $A\vec{x} = \vec{y}$。

然而实际上数据不可能严格落在一条直线上，因此这个线性方程组是没有解的。退而求其次，我们可以寻找 $\vec{x}$ 使得如下的平方误差最小

$$\min_{\vec{x}} \|A\vec{x} - \vec{y}\|^2 \qquad \text{或} \qquad \min_{\vec{x}} \|A\vec{x} - \vec{y}\|$$

从而拟合出一条相对比较符合原始数据的直线。

一般地，给定 $m \times n$ 实矩阵 A 和向量 $\vec{y} \in \mathbb{R}^m$，求解

$$\min_{\vec{x} \in \mathbb{R}^n} \|A\vec{x} - \vec{y}\|$$

这一问题称为最小二乘问题。这个问题的解 $\vec{x}$ 称为线性方程组 $A\vec{x} = \vec{y}$ 的最小二乘解。

当然如果线性方阵组 $A\vec{x} = \vec{y}$ 本身有解的话，那么它的解自然是最小二乘解。这个问题的核心是处理线性方阵组 $A\vec{x} = \vec{y}$ 没有解的情况。从几何上看，矩阵 A 定义了一个线性映射

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

令 $W \subset \mathbb{R}^m$ 是这个映射的象，即所有形如 $A\vec{x}$ 的向量构成的集合。W 是 $\mathbb{R}^m$ 的一个线性子空间。由命题5.2.12我们知道

$$\min_{\vec{x} \in \mathbb{R}^n} \|A\vec{x} - \vec{y}\| = \|\vec{y} - p_W(\vec{y})\|$$

这里 $p_W(\vec{y})$ 是 $\vec{y}$ 向 W 的正交投影。因此最小二乘问题的解即为线性方程组

$$A\vec{x} = p_W(\vec{y})$$

的解。由 $p_W(\vec{y}) \in W$，这个线性方程组的解存在，其通解的自由参数个数是 $n - \operatorname{rank} A$。

命题 5.5.3. *$\vec{x}$ 是线性方程组 $A\vec{x} = \vec{y}$ 的最小二乘解当且仅当*

$$A^T A\vec{x} = A^T\vec{y}$$

证明：对任意的向量 $\vec{z} \in \mathbb{R}^n, \vec{c} \in \mathbb{R}^m$，我们有

$$\langle A\vec{z}, \vec{c}\rangle = (A\vec{z})^T \vec{c} = \vec{z}^T A^T \vec{c}$$

故 $\vec{c} \in W^\perp$ 当且仅当 $\langle A\vec{z}, \vec{c}\rangle = \vec{z}^T A^T \vec{c} = 0$ 对所有 $\vec{z}$ 成立，即等价于 $A^T \vec{c} = 0$。

$\vec{x}$ 是最小二乘解当且仅当 $A\vec{x} = p_W(\vec{y})$，即

$$A\vec{x} - \vec{y} \in W^\perp$$

由上述讨论这等价于

$$A^T(A\vec{x} - \vec{y}) = 0 \quad \text{即} \quad A^T A\vec{x} = A^T \vec{y}$$

□

注记. 我们也可以通过微积分的方法来证明这个结论。考虑 n 个变量的函数

$$f(\vec{x}) = \|A\vec{x} - \vec{y}\|^2$$

最小二乘解即为函数 f 的最小值。另一方面，

$$f(\vec{x}) = (A\vec{x} - \vec{y})^T(A\vec{x} - \vec{y}) = \vec{x}^T A^T A\vec{x} - 2\vec{x}^T A^T \vec{y} + \vec{y}^T \vec{y}$$

由于 $A^T A$ 半正定，这是一个凸函数，最小值即为 f 的极值。求解极值方程

$$\frac{\partial f}{\partial x_i} = 0 \quad \text{即得} \quad A^T A\vec{x} = A^T \vec{y}$$

由上述命题，我们可以得到如下求解最小二乘问题的算法。不妨设 A 的列向量是线性无关的，此时最小二乘法的解是唯一的。设 A 的 QR 分解为

$$A = QR = Q\begin{bmatrix} R_1 \\ 0 \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}\begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

由假设，R_1 是可逆 n 阶上三角矩阵。代入最小二乘解的方程 $A^T A\vec{x} = A^T \vec{y}$，我们得到

$$R_1^T R_1 \vec{x} = \begin{bmatrix} R_1^T & 0 \end{bmatrix}\begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}\vec{y} = R_1^T Q_1^T \vec{y}$$

即

$$R_1 \vec{x} = Q_1^T \vec{y}$$

由于 R_1 是上三角方阵，这个方程组很容易求解。

5.5.3 广义逆

矩阵的广义逆是一种对矩阵的“逆”这个概念的推广。广义逆在矩阵不可逆或者矩阵不是方阵的情况下，提供了一种类似逆矩阵的替代方法，广泛应用于求解线性方程组。

定义 5.5.4. 设 A 是 $m \times n$ 矩阵。如果 $n \times m$ 矩阵 X 满足

$$AXA = A$$

则称 X 为 A 的一个广义逆。

矩阵的广义逆总是存在的，而且并不唯一。我们用符号 A^- 来表示 A 的一个广义逆。我们考虑实矩阵的情况。对 A 作奇异值分解

$$A = U\Sigma V^T$$

这里 U 是 m 阶正交阵，V 是 n 阶正交阵。则

$$(U\Sigma V^T)X(U\Sigma V^T) = U\Sigma V^T \quad \text{即} \quad \Sigma V^T XU\Sigma = \Sigma$$

代入 $\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$，这里 $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$

$$\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^T XU \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$$

得到 X 的通解为

$$V^T XU = \begin{bmatrix} \Sigma_r^{-1} & * \\ * & * \end{bmatrix} \quad \text{即} \quad X = V \begin{bmatrix} \Sigma_r^{-1} & * \\ * & * \end{bmatrix} U^T$$

因此对于 $m \times n$ 矩阵

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^T$$

它的广义逆的一般形式为 $m \times n$ 矩阵

$$A^- = V \begin{bmatrix} \Sigma_r^{-1} & * \\ * & * \end{bmatrix} U^T$$

这里 $*$ 是任意数。这里面有一个特别的广义逆

$$V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T$$

称为 Moore–Penrose 逆。我们将在5.8.2详细讨论 Moore–Penrose 逆的刻画和性质。

命题 5.5.5. *线性方程组 $A\vec{x} = \vec{b}$ 有解的充分必要条件是*

$$\vec{b} = AA^-\vec{b}$$

这里 A^- 是 A 的任意一个广义逆。此时 $A^-\vec{b}$ 是方程的一个特解。

证明： 假设 $\vec{x}_0$ 是线性方程组 $A\vec{x} = \vec{b}$ 的解，则

$$\vec{b} = A\vec{x}_0 = AA^-A\vec{x}_0 = AA^-\vec{b}$$

反之，假设 $\vec{b} = AA^-\vec{b}$ 成立。令 $\vec{x}_0 = A^-\vec{b}$，则

$$A\vec{x}_0 = AA^-\vec{b} = \vec{b} \quad \text{成立}$$

□

我们知道，对于 n 阶可逆方阵 A，线性方阵组 $A\vec{x} = \vec{b}$ 具有唯一解

$$\vec{x} = A^{-1}\vec{b}$$

对于一般的 $m \times n$ 矩阵 A，广义逆给出了类似的公式：假设方阵组 $A\vec{x} = \vec{b}$ 有解，则 $\vec{x} = A^-\vec{b}$ 给出了一个具体的解。

命题 5.5.6. A 是 $m \times n$ 矩阵，A^- 是 A 的某个取定的广义逆。则齐次线性方程组 $A\vec{x}=0$ 的通解为

$$\vec{x}=(I_n-A^-A)\vec{z}$$

这里 $\vec{z}$ 是任意 n 维列向量。特别地，如果线性方程组 $A\vec{x}=\vec{b}$ 有解，则它的通解为

$$\vec{x}=A^-\vec{b}+(I_n-A^-A)\vec{z}$$

证明：由

$$A(I_n-A^-A)\vec{z}=(A-AA^-A)\vec{z}=0$$

知对任意 $\vec{z}$，$(I_n-A^-A)\vec{z}$ 是齐次方程组的解。

设 $\vec{x}_0$ 是齐次线性方程组 $A\vec{x}=0$ 的任意一个解，即 $A\vec{x}_0=0$。取 $\vec{z}=\vec{x}_0$，则

$$\vec{x}_0=(I_n-A^-A)\vec{x}_0=(I_n-A^-A)\vec{z}$$

即 $\vec{x}_0$ 可以表达为命题要求形式。 □

如果线性方程组 $A\vec{x}=\vec{b}$ 无解，我们可以考虑它的最小二乘问题。设 A 的奇异值分解

$$A=U\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}V^T$$

我们考虑一个特殊的广义逆（即 Moore–Penrose 逆）

$$A^-=V\begin{bmatrix}\Sigma_r^{-1} & 0\\ 0 & 0\end{bmatrix}U^T$$

这个广义逆满足

$$\begin{aligned}A^TAA^- =&V\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}U^TU\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}V^TV\begin{bmatrix}\Sigma_r^{-1} & 0\\ 0 & 0\end{bmatrix}U^T\\ =&V\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}U^T=A^T\end{aligned}$$

即 $A^T(I_m-AA^-)=0$。考虑 $\vec{x}_0=A^-\vec{b}$。对任意向量 $\vec{y}\in\mathbb{R}^n$

$$\langle A\vec{y},\vec{b}-A\vec{x}_0\rangle=\vec{y}^TA^T(I_m-AA^-)\vec{b}=0$$

因此 $A\vec{x}_0$ 即为 $\vec{b}$ 向 A 的像的投影。故 $\vec{x}_0=A^-\vec{b}$ 给出了一个最小二乘解。在5.8.2节中，我们将说明在复线性空间中也有类似的结论。

5.5.4 矩阵的谱范数

定义 5.5.7. 设 A 是 $m \times n$ 实矩阵，其谱范数 $\|A\|$ 定义为：

$$\|A\|=\max_{\vec{x}\neq 0}\frac{\|A\vec{x}\|}{\|\vec{x}\|}=\max_{\|\vec{x}\|=1}\|A\vec{x}\|$$

其中，$\|A\vec{x}\|$ 是 $\mathbb{R}^m$ 中向量 $A\vec{x}$ 的欧几里得范数，$\|\vec{x}\|$ 是 $\mathbb{R}^n$ 中向量 $\vec{x}$ 的欧几里得范数。

命题 5.5.8. 矩阵的谱范数满足如下基本性质：

1. **非负性**：$\|A\| \geq 0$，且 $\|A\| = 0 \iff A = 0$
2. **齐次性**：对任意 $\alpha \in \mathbb{R}$，$\|\alpha A\| = |\alpha|\|A\|$
3. **三角不等式**：$\|A + B\| \leq \|A\| + \|B\|$
4. 乘积性质：$\|AB\| \leq \|A\|\|B\|$
5. 若 U, V 是正交方阵，则 $\|UAV^T\| = \|A\|$

证明：我们证明最后一条性质，其他留作练习。由于正交方阵保持欧几里得范数，我们有

$$\|UAV^T\| = \max_{\vec{x}\neq 0} \frac{\|UAV^T\vec{x}\|}{\|\vec{x}\|} = \max_{\vec{x}\neq 0} \frac{\|AV^T\vec{x}\|}{\|V^T\vec{x}\|} \overset{\vec{y}=V^T\vec{x}}{=} \max_{\vec{y}\neq 0} \frac{\|A\vec{y}\|}{\|\vec{y}\|} = \|A\|$$

□

命题 5.5.9. 矩阵的谱范数 $\|A\|$ 等于 A 的最大奇异值。

证明：设 A 的奇异值分解为 $A = U\Sigma V^T$，则 $\|A\| = \|\Sigma\|$。由

$$\Sigma = \begin{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix} & 0 \\ 0 & 0 \end{bmatrix}$$

得知

$$\|\Sigma\| = \max_{\vec{x}\neq 0} \frac{\|A\vec{x}\|}{\|\vec{x}\|} = \max_{\vec{x}\neq 0} \frac{\sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_r^2 x_r^2}}{\sqrt{x_1^2 + \cdots + x_n^2}} = \max_{1\leq i\leq r} \sigma_i$$

□

这个命题给出了最大奇异值的几何解释，即线性映射

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

把向量长度伸缩最大的倍数。我们下面说明，其他奇异值均有类似的几何解释。

例子 5.5.10. 考虑一个 n 阶实对称方阵 S。由命题5.4.5，存在一组标准正交基 $\vec{\beta}_1, \cdots, \vec{\beta}_n$ 使得每个 $\vec{\beta}_i$ 都是 S 的特征向量。不妨设其对应的特征值 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$。

对任意向量 $\vec{x} \in \mathbb{R}^n$，在这组基下的线性表达为

$$\vec{x} = c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n$$

由此可得

$$\frac{\vec{x}^T S\vec{x}}{\vec{x}^T\vec{x}} = \frac{\lambda_1 c_1^2 + \cdots + \lambda_n c_n^2}{c_1^2 + \cdots + c_n^2}$$

易知
$$\lambda_1 = \max_{\vec{x}\neq 0} \frac{\vec{x}^T S\vec{x}}{\vec{x}^T\vec{x}}$$
类似的，我们有
$$\lambda_i = \max_{\substack{\vec{x}\neq 0\\ \vec{x}\in\mathrm{Span}\{\vec{\beta}_i,\cdots,\vec{\beta}_n\}}} \frac{\vec{x}^T S\vec{x}}{\vec{x}^T\vec{x}} = \max_{\substack{\vec{x}\neq 0\\ \vec{x}\in\mathrm{Span}\{\vec{\beta}_1,\cdots,\vec{\beta}_{i-1}\}^{\perp}}} \frac{\vec{x}^T S\vec{x}}{\vec{x}^T\vec{x}}$$

设 A 是 $m\times n$ 实矩阵。我们把上述例子中的方法应用到半正定实对称方阵 $S=A^TA$。设 A 的奇异值为 $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>0$，即 A^TA 的非零特征值为 $\sigma_1^2\geq\sigma_2^2\geq\cdots\geq\sigma_r^2$。设 $\vec{\beta}_1,\cdots,\vec{\beta}_r$ 是 $\mathbb{R}^n$ 的标准正交基并且构成 A^TA 的特征向量
$$A^TA\vec{\beta}_i = \begin{cases} \sigma_i^2\vec{\beta}_i & 1\leq i\leq r \\ 0 & i>r \end{cases}$$

则我们有
$$\sigma_1 = \sqrt{\max_{\vec{x}\neq 0}\frac{\vec{x}^TA^TA\vec{x}}{\vec{x}^T\vec{x}}} = \max_{\vec{x}\neq 0}\frac{\|A\vec{x}\|}{\|\vec{x}\|}$$
并且 $\vec{\beta}_1$ 达到这个最大值
$$\sigma_1 = \frac{\|A\vec{\beta}_1\|}{\|\vec{\beta}_1\|}$$

然后考虑 $\vec{\beta}_1$ 的正交补空间，得到次大的奇异值
$$\sigma_2 = \sqrt{\max_{\substack{\vec{x}\neq 0\\ \vec{x}\in\mathrm{Span}\{\vec{\beta}_1\}^{\perp}}}\frac{\vec{x}^TA^TA\vec{x}}{\vec{x}^T\vec{x}}} = \max_{\substack{\vec{x}\neq 0\\ \vec{x}\in\mathrm{Span}\{\vec{\beta}_1\}^{\perp}}}\frac{\|A\vec{x}\|}{\|\vec{x}\|}$$
这样我们依次得到第 i 个奇异值的几何解释
$$\sigma_i = \max_{\substack{\vec{x}\neq 0\\ \vec{x}\in\mathrm{Span}\{\vec{\beta}_1,\cdots,\vec{\beta}_{i-1}\}^{\perp}}}\frac{\|A\vec{x}\|}{\|\vec{x}\|}$$

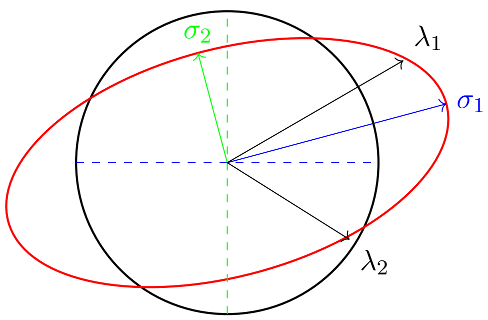


图 5.1: 蓝色和绿色表示奇异值方向，黑色表示特征值方向

5.5.5 矩阵极限

矩阵的谱范数给出了一个描述矩阵之间“距离”的方法。

定义 5.5.11. 我们称一组 $m \times n$ 矩阵 A_k 收敛于 $m \times n$ 矩阵 B，如果

$$\lim_{k\to\infty} \|A_k - B\| = 0$$

此时我们记

$$\lim_{k\to\infty} A_k = B$$

命题 5.5.12. 设 $A = (a_{ij})$ 是 $m \times n$ 实矩阵。则

$$\max_{i,j} |a_{ij}| \le \|A\| \le \sum_{i,j} |a_{ij}|$$

证明：设

$$A = \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix} \qquad \text{这里} \qquad \vec{u}_k = \begin{bmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{bmatrix} \in \mathbb{R}^n$$

对任意 $\vec{x} \in \mathbb{R}^n$ 且 $\|\vec{x}\| = 1$，我们有

$$A\vec{x} = \begin{bmatrix} \vec{u}_1^T \cdot \vec{x} \\ \vdots \\ \vec{u}_m^T \cdot \vec{x} \end{bmatrix} = \begin{bmatrix} \langle \vec{u}_1, \vec{x} \rangle \\ \vdots \\ \langle \vec{u}_m, \vec{x} \rangle \end{bmatrix}$$

由 Cauchy–Schwarz 不等式，

$$\|A\vec{x}\| \le \sum_i |\langle \vec{u}_i, \vec{x} \rangle| \le \sum_i \|\vec{u}_i\|\|\vec{x}\| = \sum_i \|\vec{u}_i\| \le \sum_i \sum_j |a_{ij}|$$

另一方面，考虑 $\mathbb{R}^n$ 中的标准基 $\vec{e}_j$。由

$$A\vec{e}_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

因此对任意指标 i, j

$$\|A\| \ge \|A\vec{e}_j\| \ge |a_{ij}|$$

命题得证。 □

这个命题说明，矩阵极限成立

$$\lim_{k\to\infty} A_k = B$$

当且仅当 A_k 的每个矩阵元都收敛到 B 中对应的矩阵元。

作为一个应用，我们考虑 n 阶实方阵 A。由矩阵范数的乘积性质，我们有

$$\left\| \frac{A^n}{n!} \right\| \le \frac{\|A\|^n}{n!}$$

由于级数

$$\sum_{n=0}^{\infty} \frac{\|A\|^n}{n!} = e^{\|A\|}$$

收敛，上述谱范数的估计可以证明矩阵级数

$$\lim_{N\to\infty} \sum_{k=0}^{N} \frac{A^k}{k!}$$

收敛。我们把这个极限矩阵方便地记为 e^A

$$e^A := \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

并且

$$\|e^A\| = \left\|\sum_{k=0}^{\infty} \frac{A^k}{k!}\right\| \le \sum_{k=0}^{\infty} \left\|\frac{A^k}{k!}\right\| \le \sum_{k=0}^{\infty} \frac{\|A\|^k}{k!} = e^{\|A\|}$$

由如上级数公式容易证明，e^A 是可逆矩阵，并且它的逆是 e^{-A}。即

$$e^A e^{-A} = I_n$$

指数矩阵 e^A 具有和指数函数 e^x 非常类似的性质。

命题 5.5.13. 设 A, B 是两个相互交换的 n 阶方阵，即满足 $AB = BA$。则

$$e^A e^B = e^{A+B}$$

证明：

$$\begin{aligned}
e^{A+B} &= \sum_{n=0}^{\infty} \frac{(A+B)^n}{n!} \\
&\underset{AB=BA}{\overset{\text{利用}}{=}} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} \\
&= \sum_{n=0}^{\infty} \sum_{k+m=n} \frac{A^k B^m}{k!m!} \\
&= \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right)\left(\sum_{m=0}^{\infty} \frac{B^m}{m!}\right) = e^A e^B
\end{aligned}$$

□

注记. 两个方阵如果不交换，$AB \neq BA$，则有如下的 *Baker-Campbell-Hausdorff* 公式

$$e^A \cdot e^B = e^{A+B+\frac{1}{2}[A,B]+\frac{1}{12}[A,[A,B]]-\frac{1}{12}[B,[A,B]]+\cdots}$$

5.5.6 微分方程中的应用

现在我们引入一个变量 t，考虑随 t 变化的矩阵

$$e^{tA}=\sum_{k=0}^{\infty}\frac{t^kA^k}{k!}$$

对矩阵等式

$$\frac{d}{dt}\left(\frac{t^kA^k}{k!}\right)=\frac{t^{k-1}A^k}{(k-1)!}.$$

求和，可以得到

$$\frac{d}{dt}e^{tA}=Ae^{tA}$$

对任意向量 $\vec{y}_0\in\mathbb{R}^n$，考虑如下向量函数

$$\vec{y}(t)=e^{tA}\vec{y}_0$$

由上述讨论知

$$\frac{d\vec{y}}{dt}=Ae^{tA}\vec{y}_0=A\vec{y}$$

并且在 $t=0$ 时有 $\vec{y}(0)=\vec{y}_0$。因此 $\vec{y}(t)=e^{tA}\vec{y}_0$ 给出了如下齐次线性常微分初值问题的解

$$\begin{cases}\frac{d\vec{y}}{dt}=A\vec{y}\\ \vec{y}(0)=\vec{y}_0\end{cases}$$

进一步，我们可以考虑非齐次线性常微分方程组

$$\begin{cases}\frac{d\vec{y}}{dt}=A\vec{y}+\vec{b}(t)\\ \vec{y}(0)=\vec{y}_0\end{cases}$$

这里 $\vec{b}(t)=\begin{bmatrix}b_1(t)\\ \vdots\\ b_n(t)\end{bmatrix}$ 是随时间变化的向量。上述方程等价于

$$\frac{d}{dt}(e^{-tA}\vec{y})=e^{-tA}\vec{b}(t)$$

两边积分，我们得到

$$e^{-tA}\vec{y}-\vec{y}_0=\int_0^t e^{-sA}\vec{b}(s)ds$$

因此

$$\vec{y}=e^{tA}\vec{y}_0+\int_0^t e^{(t-s)A}\vec{b}(s)ds$$

给出了非齐次线性常微分方程组

$$\begin{cases}\frac{d\vec{y}}{dt}=A\vec{y}+\vec{b}(t)\\ \vec{y}(0)=\vec{y}_0\end{cases}$$

的解。

5.6 酉空间

下面我们考虑复数域，也就是复向量空间上的内积空间，这样的空间称为酉空间。酉空间是欧几里得空间在复数域上的自然推广，很多处理方法和欧几里得空间类似。我们在论述相似的结论时也只叙不证，细节留作练习。

定义 5.6.1. 复线性空间 V 上的厄米内积指的是一个映射

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

满足以下条件：

1. **正定性**：对任意 $\mathbf{u} \in V$，$\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$ 且 $\langle \mathbf{u}, \mathbf{u}\rangle = 0 \iff \mathbf{u} = 0$
2. **共轭对称性**：对任意 $\mathbf{u}, \mathbf{v} \in V$，$\langle \mathbf{u}, \mathbf{v}\rangle = \overline{\langle \mathbf{v}, \mathbf{u}\rangle}$
3. **线性性**：对任意 $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ 和复数 $a, b \in \mathbb{C}$，

$$\langle \mathbf{u}, a\mathbf{v} + b\mathbf{w}\rangle = a\langle \mathbf{u}, \mathbf{v}\rangle + b\langle \mathbf{u}, \mathbf{w}\rangle$$

注意到这里的线性性是对 $\langle\cdot,\cdot\rangle$ 的第二个分量描述的。由共轭对称性，厄米内积对第一个分量是共轭线性的，即

$$\langle a\mathbf{v} + b\mathbf{w}, \mathbf{u}\rangle = \bar{a}\langle \mathbf{v}, \mathbf{u}\rangle + \bar{b}\langle \mathbf{w}, \mathbf{u}\rangle$$

注记. 有的参考文献讨论厄米内积时也会定义为对第一个分量线性，第二个分量共轭线性。这两种定义方式是等价的，只是表达方式不同。我们这里采取如上的约定，即对第二个分量线性。

定义 5.6.2. 带厄米内积的复线性空间称为酉空间。

酉空间也可能是无穷维。我们主要讨论有限维的情况。

例子 5.6.3. $V = \mathbb{C}^n$，定义两个复向量的厄米内积为

$$\langle \mathbf{z}, \mathbf{w}\rangle = \sum_{i=1}^{n} \bar{z}_i w_i$$

这里 z_i, w_i 是复向量 $\mathbf{z}, \mathbf{w}$ 的分量。此时

$$\sqrt{\langle \mathbf{z}, \mathbf{z}\rangle} = \sqrt{\sum_{i=1}^{n} |z_i|^2}$$

表示向量 $\mathbf{z}$ 的长度。容易验证，$(\mathbb{C}^n, \langle\cdot,\cdot\rangle)$ 构成一个酉空间，称为 n 维标准酉空间。如果我们把 $\mathbb{C}^n$ 中的向量写成列向量的样子，则这个厄米内积可以用矩阵乘法表示为

$$\langle \mathbf{z}, \mathbf{w}\rangle = \bar{\mathbf{z}}^T \cdot \mathbf{w} = \begin{bmatrix} \bar{z}_1 & \cdots & \bar{z}_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

例子 5.6.4. $V = \mathbb{C}^n$，设 $a_i > 0$ 是正实数。定义两个复向量的厄米内积为

$$\langle \mathbf{z}, \mathbf{w} \rangle = \sum_{i=1}^{n} a_i \bar{z}_i w_i$$

对称性和线性性显然满足。由

$$\langle \mathbf{z}, \mathbf{z} \rangle = \sum_{i=1}^{n} a_i |z_i|^2$$

可以很容易看出正定性。令 A 为对角方阵

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_n \end{bmatrix}$$

则这个厄米内积也可以写成矩阵形式

$$\langle \mathbf{z}, \mathbf{w} \rangle = \bar{\mathbf{z}}^T A \mathbf{w} = \begin{bmatrix} \bar{z}_1 & \cdots & \bar{z}_n \end{bmatrix} \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

例子 5.6.5. 设 V 是由有限闭区间 $[a, b]$ 上取值在复数的连续函数构成的线性空间。对任意两个函数 $f(x), g(x) \in V$，我们定义它们的厄米内积为

$$\langle f, g \rangle := \int_a^b \overline{f(x)} g(x) dx$$

对称性和线性性显然满足。由

$$\langle f, f \rangle := \int_a^b |f(x)|^2 dx$$

知

$$\langle f, f \rangle \geq 0$$

且 $\langle f, f \rangle = 0$ 当且仅当 $f(x) = 0$ 是零函数，即 V 中的零向量。因此 $(V, \langle \cdot, \cdot \rangle)$ 定义了一个无穷维的酉空间。

5.6.1 Gram 矩阵

设 V 是一个酉空间，$\{\vec{\alpha}_1, \vec{\alpha}_2, \cdots, \vec{\alpha}_n\}$ 是 V 的一组基。考虑这些基向量之间的厄米内积

$$\text{定义} \quad G_{ij} := \langle \vec{\alpha}_i, \vec{\alpha}_j \rangle, \qquad i, j = 1, \cdots, n$$

由此我们得到一个 n 阶复方阵 $G = (G_{ij})$。

定义 5.6.6. 设 $(V, \langle\cdot,\cdot\rangle)$ 是 n 维酉空间，$\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 为 V 的一组基。我们称 n 阶方阵

$$G = \begin{bmatrix} \langle\vec{\alpha}_1, \vec{\alpha}_1\rangle & \langle\vec{\alpha}_1, \vec{\alpha}_2\rangle & \cdots & \langle\vec{\alpha}_1, \vec{\alpha}_n\rangle \\ \langle\vec{\alpha}_2, \vec{\alpha}_1\rangle & \langle\vec{\alpha}_2, \vec{\alpha}_2\rangle & \cdots & \langle\vec{\alpha}_2, \vec{\alpha}_n\rangle \\ \cdots & \cdots & \cdots & \cdots \\ \langle\vec{\alpha}_n, \vec{\alpha}_1\rangle & \langle\vec{\alpha}_n, \vec{\alpha}_2\rangle & \cdots & \langle\vec{\alpha}_n, \vec{\alpha}_n\rangle \end{bmatrix}$$

为厄米内积 $\langle\cdot,\cdot\rangle$ 在基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 下的 *Gram* 矩阵。

设 $G = (G_{ij})$ 是厄米内积 $\langle\cdot,\cdot\rangle$ 在 V 的一组基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 下的 Gram 矩阵。对 V 中任意两个向量 $\mathbf{u}, \mathbf{v}$，它们通过这组基的线性展开记为

$$\mathbf{u} = z_1\vec{\alpha}_1 + \cdots z_n\vec{\alpha}_n$$
$$\mathbf{v} = w_1\vec{\alpha}_1 + \cdots w_n\vec{\alpha}_n$$

则

$$\langle\mathbf{u}, \mathbf{v}\rangle = \begin{bmatrix} \bar{z}_1 & \cdots & \bar{z}_n \end{bmatrix} G \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

定义 5.6.7. 对于复矩阵 $A = (a_{ij})$，我们定义它的复共轭 $\bar{A} = (\bar{a}_{ij})$，其中 $\bar{a}_{ij}$ 是复数 a_{ij} 的共轭。复共轭和转置的复合 $\bar{A}^T$ 称为 A 的共轭转置。

命题 5.6.8. 对于复方阵的乘积，我们有

$$\overline{A_1 A_2 \cdots A_m} = \bar{A}_1 \bar{A}_2 \cdots \bar{A}_m$$
$$\overline{A_1 A_2 \cdots A_m}^T = \bar{A}_m^T \cdots \bar{A}_2^T \bar{A}_1^T$$

证明：留作练习。 □

定义 5.6.9. n 阶复方阵 A 称为是厄米方阵，如果

$$A = \bar{A}^T$$

复数域上的厄米方阵类比于实数域上的对称方阵。

定义 5.6.10. 一个 n 阶厄米方阵 $A = (a_{ij})$ 称为正定的（半正定的），如果对任意非零列向量 $\mathbf{z} \in \mathbb{C}^n$，都有

$$\bar{\mathbf{z}}^T A\mathbf{z} > 0 \quad (\geq 0) \quad \text{即} \quad \sum_{i,j=1}^{n} a_{ij}\bar{z}_i\bar{z}_j > 0 \quad (\geq 0)$$

注意到对于厄米方阵 A

$$\overline{\bar{\mathbf{z}}^T A\mathbf{z}} = \overline{(\bar{\mathbf{z}}^T A\mathbf{z})}^T = \bar{\mathbf{z}}^T \bar{A}^T \mathbf{z} = \bar{\mathbf{z}}^T A\mathbf{z}$$

即 $\bar{\mathbf{z}}^T A\mathbf{z}$ 一定是实数。

命题 5.6.11. 设 G 是酉空间一组基 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_n\}$ 下的 *Gram* 矩阵。则 G 是正定厄米方阵。

证明与欧几里得空间类似：厄米内积的共轭对称性和正定性保证了 G 是正定厄米方阵。因此在给定 V 的一组基的情况下，Gram 矩阵

$$\text{厄米内积} \overset{\text{Gram}}{\Longleftrightarrow} \text{厄米正定矩阵}$$

给出厄米内积和正定厄米矩阵之间的一一对应。

5.6.2 范数

定义 5.6.12. 设 $(V, \langle\cdot,\cdot\rangle)$ 是一个酉空间。对于任意向量 $\mathbf{u} \in V$，$\mathbf{u}$ 的长度 *(或范数)* 定义为

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}$$

命题 5.6.13 (Cauchy-Schwarz 不等式)**.** 设 $(V, \langle\cdot,\cdot\rangle)$ 是一个酉空间。则对任意两个向量 $\mathbf{u}, \mathbf{v} \in V$

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

等号成立当且仅当 $\mathbf{u}, \mathbf{w}$ 线性相关。

证明：类似命题5.1.9。 □

命题 5.6.14 (三角不等式)**.** 对酉空间中任意两个向量 $\mathbf{u}, \mathbf{v}$

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

证明；类似命题5.1.11。 □

命题 5.6.15 (平行四边形法则)**.** 对酉空间中任意两个向量 $\mathbf{u}, \mathbf{v}$

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$$

证明：类似命题5.1.12。 □

5.6.3 正交性

定义 5.6.16. 酉空间中满足如下条件的两个向量

$$\langle \mathbf{u}, \mathbf{v}\rangle = 0$$

称为是正交的，记作 $\mathbf{u} \perp \mathbf{v}$。

命题 5.6.17. 设 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_m\}$ 是酉空间 V 中 m 个两两正交的非零向量，即

$$\langle \vec{\alpha}_i, \vec{\alpha}_j\rangle = 0 \qquad i \neq j, \quad i, j = 1, \cdots, m$$

则 $\{\vec{\alpha}_1, \cdots, \vec{\alpha}_m\}$ 线性无关。

证明：类似命题5.2.1。 □

定义 5.6.18. 酉空间 V 的一组基如果由两两正交的向量构成，则称这组基为正交基。如果正交基的每个基向量范数都是 1，则这组基称为酉空间的标准正交基。

给定酉空间 V 的任一组基 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$，我们同样可以用 Gram-Schmidt 正交化依次定义

$$\vec{\beta}_k=\vec{\alpha}_k-\sum_{i=1}^{k-1}\frac{\langle\vec{\alpha}_k,\vec{\beta}_i\rangle}{\langle\vec{\beta}_i,\vec{\beta}_i\rangle}\vec{\beta}_i$$

得到一组正交基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_k\}$。此时从基 $\{\vec{\alpha}_i\}$ 到 $\{\vec{\beta}_i\}$ 的过渡矩阵 P 是上三角矩阵

$$\begin{bmatrix}\vec{\beta}_1 & \vec{\beta}_2 & \cdots & \vec{\beta}_n\end{bmatrix}=\begin{bmatrix}\vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_n\end{bmatrix}P$$

我们可以进一步对 $\vec{\beta}_i$ 作归一化，得到一组标准正交基。由此证明了如下结论。

命题 5.6.19. n 维酉空间存在标准正交基。

定义 5.6.20. 设 $(V,\langle\cdot,\cdot\rangle)$ 是一个 n 维酉空间，$W\subset V$ 是复线性子空间。我们定义 W 在 V 中的正交补空间为

$$W^{\perp}:=\{\mathbf{u}\in V|\langle\mathbf{u},\mathbf{v}\rangle=0\text{ 对 }W\text{ 中任意向量 }\mathbf{v}\text{ 成立}\}$$

即 $\mathbf{u}\in W^{\perp}$ 如果 $\mathbf{u}$ 与 W 中的任意向量都正交。

如果我们选 W 中的一组基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_m\}$，则 $\mathbf{u}\in W^{\perp}$ 当且仅当 $\mathbf{u}$ 与这组基向量正交

$$\langle\mathbf{u},\vec{\beta}_i\rangle=0\qquad i=1,\cdots,m$$

这给出了正交补空间向量的一个判别方法。

命题 5.6.21. 设 $(V,\langle\cdot,\cdot\rangle)$ 是一个 n 维酉空间，$W\subset V$ 是线性子空间。则我们有直和分解

$$V=W\oplus W^{\perp}$$

证明：类似命题5.2.9。 □

因此任意向量 $\mathbf{u}\in V$ 可以唯一地写为

$$\mathbf{u}=\mathbf{u}^{\parallel}+\mathbf{u}^{\perp}\quad\text{ 其中 }\quad\mathbf{u}^{\parallel}\in W,\quad\mathbf{u}^{\perp}\in W^{\perp}$$

称为 $\mathbf{u}$ 关于 W 的正交分解。线性映射

$$p_W:V\to W,\qquad\mathbf{u}\to\mathbf{u}^{\parallel}$$

称为 V 向 W 的正交投影。

命题 5.6.22. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维酉空间，$W\subset V$ 是线性子空间。设 $\mathbf{u}$ 是 V 中任意向量，则

$$\|\mathbf{u}-p_W(\mathbf{u})\|=\min_{\mathbf{w}\in W}\|\mathbf{u}-\mathbf{w}\|$$

即 $p_W(\mathbf{u})$ 是 W 中与 $\mathbf{u}$ 的距离最短的向量。

证明：类似命题5.2.12。 □

等价而言，如果向量 $\mathbf{u}$ 关于 W 的正交分解为

$$\mathbf{u}=\mathbf{u}^{\parallel}+\mathbf{u}^{\perp}$$

则 $\mathbf{u}$ 与 W 中向量间的最短距离为分量 $\mathbf{u}^{\perp}$ 的长度

$$\min_{\mathbf{w}\in W}\|\mathbf{u}-\mathbf{w}\|=\|\mathbf{u}^{\perp}\|$$

这个值也称为向量 $\mathbf{u}$ 与子空间 W 的距离。

5.7 酉矩阵

5.7.1 酉矩阵

定义 5.7.1. 设 V 是一个酉空间。线性变换 $T: V \to V$ 称为正交变换，如果对于任意向量 $\mathbf{u} \in V$，都有

$$\|T(\mathbf{u})\| = \|\mathbf{u}\|$$

即正交变换保持向量的范数（长度）不变。

命题 5.7.2. 酉空间 V 上的线性变换 $T: V \to V$ 是一个正交变换当且仅当 T 保持厄米内积不变，即对 V 中的任意向量 $\mathbf{u}, \mathbf{v}$

$$\langle T(\mathbf{u}), T(\mathbf{v})\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$$

证明：类似命题5.3.4。 □

与欧几里得空间类似，我们同样可以选标准正交基来具体刻画正交变换。

命题 5.7.3. 设 $\{\vec{\gamma}_1, \cdots, \vec{\gamma}_n\}$ 是 n 维酉空间 V 的一组标准正交基。则线性映射 $T: V \to V$ 是正交变换当且仅当

$$\{T(\vec{\gamma}_1), \cdots, T(\vec{\gamma}_n)\}$$

也是 V 的标准正交基。

证明：类似命题5.3.5。 □

定义 5.7.4. 一个 n 阶复方阵 A 称为酉矩阵，如果它满足

$$\bar{A}^T A = A\bar{A}^T = I_n$$

其中 $\bar{A}^T$ 表示 A 的共轭转置，I_n 表示 n 阶单位矩阵。所有 n 阶酉矩阵组成的集合记为 $\mathrm{U}(n)$。

即 A 是酉矩阵当且仅当 A 的共轭转置是它的逆

$$\bar{A}^T = A^{-1}$$

命题 5.7.5. 设 $T: V \to V$ 是 n 维酉空间 V 的线性变换。则 T 是酉变换当且仅当 T 在 V 的一组标准正交基下的表示矩阵 A 是 n 阶酉矩阵。

证明：类似命题5.3.7。 □

命题 5.7.6. 设 A 是 n 阶复方阵，则如下条件等价：

1. A 是酉矩阵
2. A 的列向量构成 $\mathbb{C}^n$ 的标准正交基
3. A 的行向量构成 $\mathbb{C}^n$ 的标准正交基

证明：类似命题5.3.8。 □

命题 5.7.7. 设 A, B 是 n 阶酉矩阵，则 $A^{-1}(=\bar{A}^T), AB$ 均为 n 阶酉矩阵。即 $\mathrm{U}(n)$ 构成一个群，称为 n 阶酉群。

证明：类似命题5.3.10。 □

定义 5.7.8. 行列式为 1 的 n 阶酉矩阵构成的集合记为 $\mathrm{SU}(n)$，称为 n 阶特殊酉群

$$\mathrm{SU}(n) = \{A \in \mathrm{U}(n) | \det A = 1\}$$

$\mathrm{SU}(n)$ 是 $U(n)$ 的子群。

定义 5.7.9. 设 A 和 B 是两个 n 阶复方阵。如果存在一个 n 阶酉矩阵 U，使得

$$B = UAU^{-1} \quad (\text{即} = UA\bar{U}^T)$$

则称矩阵 A 和 B 酉相似。

酉相似是比相似更强的关系。如果两个矩阵酉相似，则它们一定相似；反之不一定成立。

命题 5.7.10. 任意 n 阶复方阵酉相似于一个上三角方阵。

证明：类似命题5.4.3。 □

5.7.2 正规矩阵

可对角矩阵具有非常好的性质。我们这里给出判断 n 阶复方阵是否可以通过酉相似对角化的一个简单方法。

定义 5.7.11. n 阶复方阵 A 称为正规矩阵，如果 A 与它的共轭转置 $\bar{A}^T$ 交换

$$A\bar{A}^T = \bar{A}^T A$$

命题 5.7.12. n 阶复方阵 A 酉相似于对角方阵当且仅当 A 是正规矩阵。

证明：假设 A 酉相似于对角方阵，即存在酉矩阵 U 和对角方阵 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ 使得

$$A = U\Lambda\bar{U}^T$$

则 $\bar{A}^T = U\bar{\Lambda}\bar{U}^T$，因此 $A\bar{A}^T = \bar{A}^T A = U\Lambda\bar{\Lambda}\bar{U}^T$。这里我们用到 $\Lambda\bar{\Lambda} = \bar{\Lambda}\Lambda$。

反之，假设 A 是正规矩阵。由命题5.7.10，存在酉矩阵 U 和上三角方阵

$$\Lambda = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$$

使得 $A = U\Sigma\bar{U}^T$。代入正规条件 $A\bar{A}^T = \bar{A}^T A$，得到 $\Sigma\bar{\Sigma}^T = \bar{\Sigma}^T\Sigma$，即

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} \begin{bmatrix} \bar{a}_{11} & 0 & \cdots & 0 \\ \bar{a}_{12} & \bar{a}_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \bar{a}_{1n} & \cdots & * & \bar{a}_{nn} \end{bmatrix} = \begin{bmatrix} \bar{a}_{11} & 0 & \cdots & 0 \\ \bar{a}_{12} & \bar{a}_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \bar{a}_{1n} & \cdots & * & \bar{a}_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$$

比较两边第 1 行第 1 列矩阵元得

$$|a_{11}|^2 + |a_{12}|^2 + \cdots + |a_{1n}|^2 = |a_{11}|^2$$

由此得 $a_{12} = \cdots = a_{1n} = 0$。然后比较两边第 2 行第 2 列矩阵元，依次类推我们得到

$$\Lambda = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$$

是对角方阵。 □

命题 5.7.13. 设 A 是 n 阶厄米方阵，即 $A = \bar{A}^T$，则 A 可以酉相似于对角方阵。

证明：由 A 是 n 阶厄米方阵

$$A\bar{A}^T = A^2 = \bar{A}^T A$$

故 A 是正规矩阵。由命题5.7.12，A 可以酉相似于对角方阵。 □

命题 5.7.14. 设 A 是 n 阶厄米方阵，则 A 的特征值均为实数。A 是正定（半正定）当且仅当 A 的特征值均为正数（非负数）。

证明：类似命题5.4.4、命题5.4.6和命题5.4.8。 □

命题 5.7.15. 设 A 是 n 阶半正定厄米方阵。则存在唯一的半正定厄米方阵 B 使得 $A = B^2$。记 $B = \sqrt{A}$。

证明：类似命题5.4.9。 □

5.7.3 奇异值分解

设 A 是 $m \times n$ 复矩阵。考虑 n 阶方阵 $\bar{A}^T A$。对于任意 $\mathbf{z} \in \mathbb{C}^n$

$$\bar{\mathbf{z}}^T (\bar{A}^T A)\mathbf{z} = \overline{(A\mathbf{z})}^T (A\mathbf{z}) \geq 0$$

因此 $\bar{A}^T A$ 是 n 阶半正定厄米方阵。同理，$A\bar{A}^T$ 是 m 阶半正定厄米方阵。并且 $\bar{A}^T A$ 和 $A\bar{A}^T$ 具有相同的非零特征值。

定义 5.7.16. 设 A 是 $m \times n$ 复矩阵。半正定厄米方阵 $\bar{A}^T A$（或 $A\bar{A}^T$）的非零特征值的平方根称为 A 的奇异值。

设 $\bar{A}^T A$ 的所有非零特征值为 $\mu_1, \cdots, \mu_r$，则 A 的所有奇异值为 $\sigma_1 = \sqrt{\mu_1}, \cdots, \sigma_r = \sqrt{\mu_r}$。

命题 5.7.17. 设 A 是 n 阶厄米方阵。则 A 的所有奇异值为所有非零特征值的绝对值。

证明：类似命题5.4.11。 □

对于非厄米方阵，奇异值和特征值没有这样的直接联系。

定理 5.7.18 (奇异值分解). 设 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 是 $m \times n$ 阶复矩阵 A 的所有奇异值。则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V 使得 $A = U\Sigma\bar{V}^T$，其中

$$\Sigma = \begin{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}$$

这里 $0_{k\times l}$ 表示 $k \times l$ 的零矩阵。

证明：类似定理5.4.13。 □

5.8 一些例子

5.8.1 极分解

我们知道一个复数 $z \in \mathbb{C}$ 可以写成

$$z = re^{i\theta}$$

这里 $r \geq 0$ 表示 z 的长度，$e^{i\theta}$ 的模长为 1。(r, θ) 称为复平面的极坐标。

n 阶复方阵也可以作类似的分解。

命题 5.8.1. 设 A 是 n 阶复方阵。则存在 n 阶半正定厄米方阵 R（或 $\tilde{R}$）和 n 阶酉矩阵 Θ（或 $\tilde{\Theta}$）使得

$$A = R\Theta \quad (= \tilde{\Theta}\tilde{R})$$

其中 R（或 $\tilde{R}$）由 A 唯一确定。这个称为矩阵的极分解。

证明：设 $A = U\Sigma\bar{V}^T$ 是 A 的奇异值分解。则

$$A = U\Sigma\bar{U}^T U\bar{V}^T$$

取 $R = U\Sigma\bar{U}^T, \Theta = U\bar{V}^T$ 即满足要求。

另一方面

$$A\bar{A}^T = R\bar{R}^T = R^2$$

由命题5.7.15知 $R = \sqrt{A\bar{A}^T}$ 是唯一确定的。 □

5.8.2 Moore–Penrose 逆

在复数域的情形，一个 $m \times n$ 矩阵 A 的广义逆 A^- 也是定义为满足

$$AA^-A = A$$

的 $n \times m$ 矩阵 A^-。广义逆总是存在而且并不唯一。我们可以通过引入额外条件在所有广义逆中找一个唯一的代表元，称为 Moore–Penrose 逆。

命题 5.8.2. 设 A 是 $m\times n$ 复矩阵。则存在唯一的 $n\times m$ 复矩阵 X 满足如下条件

$$\begin{cases} AXA = A & XAX = X \\ \overline{(AX)}^T = AX & \overline{(XA)}^T = XA \end{cases}$$

这个 X 称为 A 的 *Moore–Penrose* 逆，记为 A^+。

证明：设 A 的奇异值分解为

$$A = U\Sigma\bar{V}^T = U\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}\bar{V}^T$$

考虑

$$X = V\begin{bmatrix}\Sigma_r{}^{-1} & 0\\ 0 & 0\end{bmatrix}\bar{U}^T$$

直接计算易知

$$AXA = U\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}\bar{V}^TV\begin{bmatrix}\Sigma_r{}^{-1} & 0\\ 0 & 0\end{bmatrix}\bar{U}^TU\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}\bar{V}^T = A$$
$$XAX = V\begin{bmatrix}\Sigma_r{}^{-1} & 0\\ 0 & 0\end{bmatrix}\bar{U}^TU\begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix}\bar{V}^TV\begin{bmatrix}\Sigma_r{}^{-1} & 0\\ 0 & 0\end{bmatrix}\bar{U}^T = X$$

并且

$$AX = U\begin{bmatrix}I_r & 0\\ 0 & 0\end{bmatrix}\bar{U}^T \quad \text{和} \quad XA = V\begin{bmatrix}I_r & 0\\ 0 & 0\end{bmatrix}\bar{V}^T$$

都是厄米方阵。因此 X 满足所述方程。下证唯一性。

记 $X = VY\bar{U}^T$，则 Y 满足

$$\begin{cases} \Sigma Y\Sigma = \Sigma & Y\Sigma Y = Y \\ \overline{(\Sigma Y)}^T = \Sigma Y & \overline{(Y\Sigma)}^T = Y\Sigma \end{cases}$$

写成分块矩阵的样子

$$\Sigma = \begin{bmatrix}\Sigma_r & 0\\ 0 & 0\end{bmatrix} \qquad Y = \begin{bmatrix}B & C\\ D & E\end{bmatrix}$$

由方程 $\Sigma Y\Sigma = \Sigma$ 和 $Y\Sigma Y = Y$ 得到

$$Y = \begin{bmatrix}\Sigma_r^{-1} & C\\ D & D\Sigma_r C\end{bmatrix}$$

再由

$$\Sigma Y = \begin{bmatrix}I_r & \Sigma_r C\\ 0 & 0\end{bmatrix} \quad \text{和} \quad Y\Sigma = \begin{bmatrix}I_r & 0\\ D\Sigma_r & 0\end{bmatrix}$$

的厄米性，得到 $C = D = 0$。唯一性得证。 □

通过 Moore–Penrose 逆可以直接给出最小二乘问题的一个解。在复向量的情形，最小二乘问题为

$$\min_{\mathbf{z}\in\mathbb{C}^n}\|A\mathbf{z}-\mathbf{w}\|$$

由 Moore–Penrose 逆 A^+ 的性质可以得到

$$\bar{A}^T=\overline{(AA^+A)}^T=\bar{A}^T\overline{(AA^+)}^T=\bar{A}^TAA^+$$

即

$$\bar{A}^T(I_m-AA^+)=0$$

因此对任意 $\mathbf{y}\in\mathbb{C}^n$

$$\bar{\mathbf{y}}^T\bar{A}^T(I_m-AA^+)\mathbf{w}=0$$

即

$$\langle A\mathbf{y},\mathbf{w}-AA^+\mathbf{w}\rangle=0$$

这说明 $AA^+\mathbf{w}$ 是 $\mathbf{w}$ 向 A 的像的投影，即 $z_0=A^+\mathbf{w}$ 给出了最小二乘问题的一个解

$$\|A\mathbf{z}_0-\mathbf{w}\|=\min_{\mathbf{z}\in\mathbb{C}^n}\|A\mathbf{z}-\mathbf{w}\|$$

5.8.3 Pauli 矩阵与自旋

SU(2) **群**

特殊酉群 SU(2) 在物理中扮演着重要的角色，例如在量子力学中被用来描述粒子自旋的对称性。在几何上，SU(2) 给出了 3 维空间转动 SO(3) 的双重覆盖，也称为 3 维自旋群 $\mathrm{SU}(2)=\mathrm{Spin}(3)$。由定义，SU(2) 群由满足 $\det(A)=1$ 的 2 阶酉矩阵 A 组成。

设 $A=\begin{bmatrix}a&b\\c&d\end{bmatrix}$，则

$$A\bar{A}^T=\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}\bar{a}&\bar{c}\\\bar{b}&\bar{d}\end{bmatrix}=\begin{bmatrix}a\bar{a}+b\bar{b}&a\bar{c}+b\bar{d}\\c\bar{a}+d\bar{b}&c\bar{c}+d\bar{d}\end{bmatrix}$$

由酉方阵 $A\bar{A}^T=I_2$ 得

$$|a|^2+|b|^2=|c|^2+|d|^2=1\qquad a\bar{c}=-b\bar{d}$$

由 $\det A=ad-bc=1$，我们得到联立方程组

$$\begin{cases}|a|^2+|b|^2=|c|^2+|d|^2=1\\a\bar{c}=-b\bar{d}\\ad-bc=1\end{cases}$$

由此解出

$$d=\bar{a}\qquad c=-\bar{b}\qquad |a|^2+|b|^2=1$$

因此

$$\mathrm{SU}(2)=\left\{\left.\begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix}\right| a,b\in\mathbb{C} \text{ 满足 } |a|^2+|b|^2=1\right\}$$

如果用实数来表示 $a=x_1+ix_2, b=x_3+ix_4$，则 $SU(2)$ 中的元素一一对应于 3 维单位球面

$$S^3=\{(x_1,x_2,x_3,x_4)\in\mathbb{R}^4|x_1^2+x_2^2+x_3^2+x_4^2=1\}$$

Pauli 矩阵

Pauli 矩阵包括三个矩阵，定义如下：

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

这些矩阵都是厄米方阵并且满足代数关系

$$\begin{cases} \sigma_1^2=\sigma_2^2=\sigma_3^2=I_2 \\ \sigma_1\sigma_2=-\sigma_2\sigma_1=i\sigma_3 \\ \sigma_2\sigma_3=-\sigma_3\sigma_2=i\sigma_1 \\ \sigma_3\sigma_1=-\sigma_1\sigma_3=i\sigma_2 \end{cases}$$

设 V 为所有迹为零的 2 阶厄米方阵构成的集合

$$V=\{2 \text{ 阶复方阵 } A|\ \bar{A}^T=A, \operatorname{Tr} A=0\}$$

容易验证，任一矩阵 $Q\in V$ 可以写成

$$Q=\begin{bmatrix} x_3 & x_1-ix_2 \\ x_1+ix_2 & -x_3 \end{bmatrix}=x_1\sigma_1+x_2\sigma_2+x_3\sigma_3$$

这里 x_1,x_2,x_3 是实数。因此我们可以把 $\mathbb{R}^3$ 和 V 一一对应起来

$$\begin{aligned} \sigma:\mathbb{R}^3 &\to V \\ \vec{x} &\mapsto \sigma(\vec{x}):=x_1\sigma_1+x_2\sigma_2+x_3\sigma_3 \end{aligned}$$

即 V 是 3 维实线性空间，Pauli 矩阵给出一组基。

命题 5.8.3. *$\mathbb{R}^3$ 中的向量长度可以通过 Pauli 矩阵实现为*

$$\|\vec{x}\|^2=-\det\sigma(\vec{x})$$

证明：

$$\det\sigma(\vec{x})=\begin{vmatrix} x_3 & x_1-ix_2 \\ x_1+ix_2 & -x_3 \end{vmatrix}=-(x_1^2+x_2^2+x_3^2)$$

□

命题 5.8.4. 设 $\vec{x},\vec{y}\in\mathbb{R}^3$，则

$$\sigma(\vec{x})\sigma(\vec{y})=\langle\vec{x},\vec{y}\rangle I_2+i\sigma(\vec{x}\times\vec{y})$$

即 *Pauli* 矩阵同时表达了 3 维向量的内积和叉乘。

证明：设 $\epsilon_{ijk}(i,j,k\in\{1,2,3\})$ 是 3 维 Levi-Civita 符号。ϵ_{ijk} 关于指标 i,j,k 是全反对称的：

$$\epsilon_{ijk}=-\epsilon_{jik}=-\epsilon_{kji}=-\epsilon_{ikj}$$

ϵ_{ijk} 只有当 i,j,k 互不相同时才非零，它的值为置换 $\{1,2,3\}\to\{i,j,k\}$ 的符号。例如 $\epsilon_{123}=1$。

Pauli 矩阵相乘可以通过 Levi-Civita 符号表示为

$$\sigma_j\sigma_k=\delta_{jk}I_2+i\sum_l\epsilon_{jkl}\sigma_l$$

例如

$$\sigma_2^2=I_2\qquad\sigma_1\sigma_3=i\epsilon_{132}\sigma_2=-i\sigma_2$$

于是

$$\begin{aligned}\sigma(\vec{x})\sigma(\vec{y})&=\sum_j x_j\sigma_j\sum_k y_k\sigma_k\\&=\sum_{j,k}x_jy_k\delta_{jk}I_2+i\sum_{j,k,l}\epsilon_{jkl}x_jy_k\sigma_l\\&=\langle\vec{x},\vec{y}\rangle I_2+i\sigma(\vec{x}\times\vec{y})\end{aligned}$$

这里我们用到叉乘 $\vec{x}\times\vec{y}$ 的第 l 分量可以表达为 $\sum\limits_{j,k}\epsilon_{jkl}x_jy_k$。 □

设 $A\in\mathrm{SU}(2)$，$Q\in V$，则矩阵 $AQ\bar{A}^T$ 满足

$$\begin{cases}\overline{AQ\bar{A}^T}^T=A\bar{Q}^T\bar{A}^T=AQ\bar{A}^T\\\mathrm{Tr}(AQ\bar{A}^T)=\mathrm{Tr}(Q\bar{A}^TA)=\mathrm{Tr}\,Q=0\end{cases}$$

即 $AQ\bar{A}^T$ 也是 V 中的矩阵。因此给定 $A\in\mathrm{SU}(2)$，我们可以定义映射

$$\begin{aligned}f_A:V&\to V\\Q&\mapsto f_A(Q):=AQ\bar{A}^T\end{aligned}$$

容易验证 f_A 是一个线性映射。记 f_A 在 V 的基 $\{\sigma_1,\sigma_2,\sigma_3\}$ 下的表示矩阵为 $\rho(A)$，即

$$f_A(\sigma_j)=\sum_{i=1}^3\sigma_i\rho(A)_{ij}$$

这里 $\rho(A)_{ij}$ 表示矩阵 $\rho(A)$ 的 (i,j) 分量。则对于 $\sigma(\vec{x})=x_1\sigma_1+x_2\sigma_2+x_3\sigma_3\in V$，我们有

$$f_A(\sigma(\vec{x}))=\sum_j x_jf_A(\sigma_j)=\sum_{i,j}\sigma_i\rho(A)_{ij}x_j=\sum_i\sigma_iy_i$$

这里 $y_i=\sum\limits_j\rho(A)_{ij}x_j$，或写成向量形式 $\vec{y}=\rho(A)\vec{x}$。

因此我们发现如下关系

$$f_A(\sigma(\vec{x}))=\sigma(\rho(A)\vec{x})$$

命题 5.8.5. ρ 定义了一个映射 $\rho : \mathrm{SU}(2) \to \mathrm{SO}(3)$。

证明：设 $A \in \mathrm{SU}(2)$，需证明 $\rho(A) \in \mathrm{SO}(3)$。由

$$f_A(\sigma(\vec{x})) = \sigma(\rho(A)\vec{x})$$

两边取行列式并利用 $\det \sigma(\vec{x}) = -\|\vec{x}\|^2$，我们得到

$$\|\rho(A)\vec{x}\|^2 = -\det \sigma(\rho(A)\vec{x}) = -\det f_A(\sigma(\vec{x})) = -\det(A\sigma(x)\bar{A}^T) = -\det \sigma(x) = \|\vec{x}\|^2$$

因此 $\rho(A)$ 保持 $\mathbb{R}^3$ 的范数，故 $\rho(A) \in \mathrm{O}(3)$。

由 $\rho(A) \in \mathrm{O}(3)$ 知 $\det \rho(A) = \pm 1$。由于 $\mathrm{SU}(2)$ 拓扑上和 3 维球面 S^3 一样，而 S^3 是连通的，所以 A 可以在 $\mathrm{SU}(2)$ 中连续变化到恒等矩阵 I_2。因此

$$\det \rho(A) = \det \rho(I_2) = 1$$

这说明 $\rho(A) \in \mathrm{SO}(3)$。 □

例子 5.8.6. 设 $\vec{u} \in \mathbb{R}^3$ 是一个单位向量，即 $\|\vec{u}\| = 1$。考虑

$$A_{\vec{u}} := \begin{bmatrix} iu_3 & iu_1 + u_2 \\ iu_1 - u_2 & -iu_3 \end{bmatrix} \in \mathrm{SU}(2)$$

我们可以用 *Pauli* 矩阵把它写为

$$A_{\vec{u}} = i(u_1\sigma_1 + u_2\sigma_2 + u_3\sigma_3) = i\sigma(\vec{u})$$

因此

$$f_{A_{\vec{u}}}(\sigma(\vec{x})) = A_{\vec{u}}\sigma(\vec{x})\overline{A_{\vec{u}}}^T = \sigma(\vec{u})\sigma(\vec{x})\sigma(\vec{u})$$

利用 $\|\vec{u}\| = 1$ 和 Pauli 矩阵的乘法性质

$$\begin{aligned} \sigma(\vec{u})\sigma(\vec{x})\sigma(\vec{u}) =& (\langle \vec{u}, \vec{x} \rangle I_2 + i\sigma(\vec{u} \times \vec{x}))\sigma(\vec{u}) \\ =& \sigma(\langle \vec{u}, \vec{x} \rangle \vec{u}) - \sigma((\vec{u} \times \vec{x})) \times \vec{u}) \\ =& \sigma(2\langle \vec{x}, \vec{u} \rangle \vec{u} - \vec{x}) \end{aligned}$$

写成矩阵表达式

$$2\langle \vec{x}, \vec{u} \rangle \vec{u} - \vec{x} = \begin{bmatrix} 2u_1u_1 - 1 & 2u_1u_2 & 2u_1u_3 \\ 2u_2u_1 & 2u_2u_2 - 1 & 2u_2u_3 \\ 2u_3u_1 & 2u_3u_2 & 2u_3u_3 - 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

我们得到

$$\rho(A_{\vec{u}}) = \begin{bmatrix} 2u_1u_1 - 1 & 2u_1u_2 & 2u_1u_3 \\ 2u_2u_1 & 2u_2u_2 - 1 & 2u_2u_3 \\ 2u_3u_1 & 2u_3u_2 & 2u_3u_3 - 1 \end{bmatrix}$$

观察到

$$\begin{cases} \rho(A_{\vec{u}})(\vec{u}) = \vec{u} \\ \rho(A_{\vec{u}})(\vec{x}) = -\vec{x} \quad \text{若 } \vec{x} \perp \vec{u} \end{cases}$$

因此 $\rho(A_{\vec{u}})$ 表示在 $\vec{u}$ 的正交补平面上作对径映射。

命题 5.8.7. *$\rho : \mathrm{SU}(2) \to \mathrm{SO}(3)$ 满足如下性质：*

1. *保单位：$\rho(I_2) = I_3$*
2. *保乘法结构：$\rho(AB) = \rho(A)\rho(B)$*
3. *保逆结构：$\rho(A^{-1}) = \rho(A)^{-1}$*

证明：保单位显然。我们只需证明 ρ 保乘法。则

$$\rho(A)\rho(A^{-1}) = \rho(I_2) = I_3 \quad \Longrightarrow \quad \rho(A^{-1}) = (\rho(A))^{-1}$$

由

$$\begin{aligned}\sigma(\rho(A)\rho(B)\vec{x}) =& f_{\vec{A}}(\sigma(\rho(B)\vec{x})) = f_{\vec{A}}(f_{\vec{B}}(\sigma(\vec{x})))\\ =& A(B\sigma(\vec{x})\bar{B}^T)\bar{A}^t = (AB)\sigma(x)\overline{AB}^T\\ =& f_{AB}(\sigma(\vec{x})) = \sigma(\rho(AB)\vec{x})\end{aligned}$$

即得

$$\rho(A)\rho(B) = \rho(AB)$$

□

我们称 $\rho : \mathrm{SU}(2) \to \mathrm{SO}(3)$ 是一个群同态，即保持群结构（乘法和逆）的映射。进一步分析容易验证 ρ 是一个 $2:1$ 的覆盖映射。例如

$$\rho(\pm I_2) = I_3$$

实际上对任意 $A \in \mathrm{SU}(2)$ 有

$$\rho(A) = \rho(-A)$$

这个 $2:1$ 的覆盖解释了一个非常有趣的现象：即自旋 1/2 粒子绕某一轴旋转 360 度后，状态不完全恢复，而是变为原状态的负号。

例如考虑 SU(2) 矩阵

$$R_\theta = \cos(\theta/2)I_2 + i\sin(\theta/2)\sigma_1 = \begin{bmatrix}\cos(\theta/2) & i\sin(\theta/2)\\ i\sin(\theta/2) & \cos(\theta/2)\end{bmatrix}$$

直接计算易知

$$\rho(R_\theta) = \begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & \sin\theta\\ 0 & -\sin\theta & \cos\theta\end{bmatrix}$$

因此 $\rho(R_\theta)$ 对应于在 3 维空间中以 x_1-方向为轴在 x_2x_3-平面旋转 θ 角度。在旋量空间或者在 SU(2) 矩阵中，$\rho(R_\theta)$ 相当于旋转 $\theta/2$。可以看出

$$R_{2\pi} = -R_0 = -I_2 \qquad \rho(R_{2\pi}) = I_3$$

Pauli 矩阵还可以实现 4 维时空的 Lorentz 变换。设 M 为所有 2 阶厄米方阵构成的集合

$$M = \{2 \text{ 阶复方阵 } A | \bar{A}^T = A\}$$

M 中任一矩阵 T 可以写成

$$T = \begin{bmatrix} x_0 + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_0 - x_3 \end{bmatrix} = x_0 I_2 + x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3$$

这 $x_0, x_1, x_2, x_3 \in \mathbb{R}$。因此 M 是 4 维实线性空间，矩阵 $\{I_2, \sigma_1, \sigma_2, \sigma_3\}$ 构成 M 的一组基。

我们可以把 M 中的矩阵和 4 维时空对应起来，其中 x_0 表示时间，x_1, x_2, x_3 表示空间。用

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

来表示 4 维时空向量。记

$$T(\mathbf{x}) = \begin{bmatrix} x_0 + x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_0 - x_3 \end{bmatrix}$$

则

$$\det T(\mathbf{x}) = x_0^2 - x_1^2 - x_2^2 - x_3^2$$

给出 4 维时空的 Minkowski 长度。因此保持 M 中矩阵行列式不变的线性变换都将给出 Minkowski 空间的 Lorentz 变换。

例如设 A 是一个 2 阶复方阵，满足 $\det A = \pm 1$。对 2 阶厄米方阵 $T \in M$，矩阵 $AT\bar{A}^T$ 也是厄米方阵，并且

$$\det(AT\bar{A}^T) = |\det A|^2 \det T = \det T$$

因此我们得到保行列式的线性映射

$$\begin{aligned} g_A : M &\to M \\ T &\mapsto AT\bar{A}^T \end{aligned}$$

这个线性映射在基 $\{I_2, \sigma_1, \sigma_2, \sigma_3\}$ 下的表示矩阵是一个 Lorentz 变换。这个实际上给出了 4 维时空 Lorentz 群的旋量构造。记

$$SL^*(2, \mathbb{C}) := \{2\text{阶可逆复方阵} | \det A = \pm 1\}$$

则上述构造给出了 4 维 Minkowski 时空的旋量群 $\mathrm{Spin}(3, 1) = SL^*(2, \mathbb{C})$。

5.9 习题

1. 设 V 是由所有 $m\times n$ 实矩阵构成的线性空间。定义 $\langle\cdot,\cdot\rangle$ 如下

$$\langle A,B\rangle=\mathrm{Tr}(A^TB),\qquad A,B\in V$$

证明 $\langle\cdot,\cdot\rangle$ 是 V 上的一个内积。

2. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间。设内积 $\langle\cdot,\cdot\rangle$ 在一组基 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 下的 Gram 矩阵是 G，在另一组基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 下的 Gram 矩阵是 G'。设基 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 到基 $\{\vec{\beta}_1,\cdots,\vec{\beta}_n\}$ 的过渡矩阵是 P。问 G,G',P 之间是什么关系?

3. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间。对于 V 中任意 n 个向量 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$，我们定义其 Gram 矩阵 $G=(G_{ij})$ 为

$$G_{ij}=\langle\vec{\alpha}_i,\vec{\alpha}_j\rangle$$

证明 $\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 线性无关（即构成 V 的一组基）当且仅当 G 是可逆矩阵。

4. 证明 2 阶实对称方阵 A 是正定方阵当且仅当 $\mathrm{Tr}A>0\ \&\ \det A>0$。

5. 证明欧几里得空间中的向量 $\vec{x},\vec{y}$ 正交当且仅当 $\|\vec{x}+\vec{y}\|^2=\|\vec{x}\|^2+\|\vec{y}\|^2$。

6. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间，$\{\vec{\alpha}_1,\cdots,\vec{\alpha}_n\}$ 是 V 的标准正交基。证明对任意 $\vec{x},\vec{y}\in V$

$$\langle\vec{x},\vec{y}\rangle=\sum_{i=1}^n\langle\vec{x},\vec{\alpha}_i\rangle\langle\vec{\alpha}_i,\vec{y}\rangle\qquad\text{(Parseval 等式)}$$

7. 设 $V=\mathrm{Span}\{1,x,x^2,x^3\}$ 是由所有次数小于等于 3 的实系数多项式构成的线性空间。定义

$$\langle f,g\rangle:=\int_0^1 f(x)g(x)dx$$

$(V,\langle\cdot,\cdot\rangle)$ 构成 4 维欧几里得空间。找一组 $(V,\langle\cdot,\cdot\rangle)$ 的标准正交基。

8. 举例说明，方阵 A 的列向量两两相互正交，但 A 的行向量不一定两两相互正交。

9. 在标准欧几里得空间 $\mathbb{R}^3$ 中，设 U 是由方程 $x_1+2x_2+2x_3=0$ 定义的平面。计算点 $\mathbf{p}=(1,2,3)$ 在平面 U 上的正交投影 $\mathbf{p}'$。

10. 设 $(V,\langle\cdot,\cdot\rangle)$ 是由所有 n 阶实方阵构成的欧几里得空间，其内积为 $\langle A,B\rangle=\mathrm{Tr}(A^TB)$。对于如下 V 的子空间 U，计算 U 的正交补 $U^\perp$ 以及正交投影 $p_U:V\to U$。

    (a) $U=\mathrm{Span}\{I_n\}$ 是由单位方阵张成的子空间

    (b) U 是由所有对角方阵构成的线性子空间

    (c) U 是由所有上三角方阵构成的线性子空间

11. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间，U 是 V 的线性子空间。证明 $(U^\perp)^\perp=U$。

12. 设 $(V,\langle\cdot,\cdot\rangle)$ 是 n 维欧几里得空间，U_1,U_2 是 V 的线性子空间。证明 $(U_1+U_2)^\perp=U_1^\perp\cap U_2^\perp$。

13. 设 $\vec{\alpha},\vec{\beta}$ 是 n 维欧几里得空间 V 中的两个向量，满足 $\|\vec{\alpha}\|=\|\vec{\beta}\|$。证明一定存在一个正交变换 $T:V\to V$ 使得 $T(\vec{\alpha})=\vec{\beta}$。

14. 找出所有使得 $A\vec{\alpha}=\vec{\beta}$ 的 3 阶正交矩阵 A。

$$a)\quad \vec{\alpha}=\begin{bmatrix}1\\0\\0\end{bmatrix}\quad \vec{\beta}=\begin{bmatrix}0\\1\\0\end{bmatrix}\qquad\qquad b)\quad \vec{\alpha}=\begin{bmatrix}2\\1\\1\end{bmatrix}\quad \vec{\beta}=\begin{bmatrix}1\\1\\2\end{bmatrix}$$

15. 设 λ_0 是 n 阶正交方阵 A 是一个特征值（可能是复数）。

    (a) 证明 $|\lambda_0|=1$。

    (b) 假设 $\lambda_0\notin\mathbb{R}$，$\vec{z}\in\mathbb{C}^n$ 是 λ_0 的一个复特征向量。记 $\vec{z}=\vec{x}+i\vec{y}$，这里 $\vec{x},\vec{y}\in\mathbb{R}^n$。证明

$$\vec{x}\perp\vec{y}\quad \text{并且}\quad \|\vec{x}\|=\|\vec{y}\|$$

16. 判断如下实对称方阵是否是正定方阵或者半正定方阵。

$$a)\begin{bmatrix}5&3&0\\3&5&4\\0&4&5\end{bmatrix}\qquad b)\begin{bmatrix}1&-1&0\\-1&1&-1\\0&-1&1\end{bmatrix}\qquad c)\begin{bmatrix}2&-1&0&0\\-1&2&-1&0\\0&-1&2&-1\\0&0&-1&2\end{bmatrix}$$

17. 设 A 是 n 阶反对称方阵，即 $A^T=-A$。证明 $e^A\in\mathrm{SO}(n)$ 是 n 阶特殊正交方阵。

18. 设 A,B 是 n 阶正定实对称方阵。证明 AB 的特征值均为正实数。

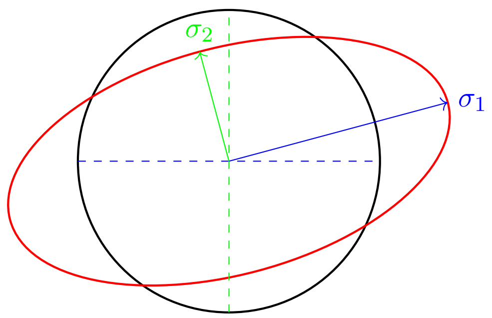
19. 计算如下矩阵的奇异值分解

$$a)\quad =\begin{bmatrix}3&0\\4&5\end{bmatrix}\qquad\qquad b)\quad =\begin{bmatrix}1&-1\\-2&2\\2&-2\end{bmatrix}$$

20. 设 $f:\mathbb{R}^2\to\mathbb{R}^2$ 是一个可逆线性映射，其在标准基下对应于矩阵 $A=\begin{bmatrix}a&b\\c&d\end{bmatrix}$，$\det A\neq 0$

$$f:\begin{bmatrix}x_1\\x_2\end{bmatrix}\mapsto\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$$

设 A 的奇异值为 $\sigma_1\geq\sigma_2>0$。证明单位圆 $\{x_1^2+x_2^2=1\}$ 在 f 的映射下变为一个椭圆，椭圆的长半轴长为 σ_1，短半轴长为 σ_2。



21. 计算如下矩阵的 QR 分解

$$a)\quad \begin{bmatrix} -1 & 3 \\ 1 & 5 \end{bmatrix} \qquad\qquad b)\quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

22. 用最小二乘法找出最佳拟合如下数据的线性方程 $y = at + b$

t_i	-2	-1	0	2
y_i	4	3	1	0

23. 证明矩阵范数的如下两个性质：

(a) 三角不等式：$\|A+B\| \leq \|A\| + \|B\|$

(b) 乘积性质：$\|AB\| \leq \|A\|\|B\|$

24. 设 A 是一个 n 阶正定方阵。证明 $\lim\limits_{t\to+\infty} e^{-tA} = 0$。

25. 假设函数 $x(t), y(t)$ 满足如下微分方程组

$$\begin{cases} x'(t) = -4x(t) + y(t) \\ y'(t) = -2x(t) - y(t) \end{cases}$$

证明 $\lim\limits_{t\to+\infty} x(t) = \lim\limits_{t\to+\infty} y(t) = 0$。

26. 酉空间中两个向量 $\vec{\alpha}, \vec{\beta}$ 正交当且仅当对任意复数 $a, b \in \mathbb{C}$,

$$\|a\vec{\alpha} + b\vec{\beta}\|^2 = \|a\vec{\alpha}\|^2 + \|b\vec{\beta}\|^2$$

27. 设 V 是由所有 $m \times n$ 复矩阵构成的复线性空间。定义 $\langle \cdot, \cdot \rangle$ 如下

$$\langle A, B \rangle = \mathrm{Tr}(\bar{A}^T B), \qquad A, B \in V$$

证明 $\langle \cdot, \cdot \rangle$ 是 V 上的一个厄米内积。

28. 设 A, B 是 n 阶厄米方阵。证明 $\mathrm{Tr}(AB)^2$ 和 $\mathrm{Tr}(A^2B^2)$ 都是实数并且

$$\mathrm{Tr}(AB)^2 \leq \mathrm{Tr}(A^2B^2)$$

等号成立当且仅当 $AB = BA$。

29. 证明任意一个 2 阶酉方阵可以表达为如下形式

$$\begin{bmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} e^{i\theta_3} & 0 \\ 0 & e^{i\theta_4} \end{bmatrix}$$

这里 $\theta_1, \theta_2, \theta_3, \theta_4, \theta$ 均是实数。

30. 设 A 是一个 n 阶复方阵。定义一个 $2n$ 阶实方阵 $\varphi(A)$ 如下:

$$\varphi(A)=\begin{bmatrix} B & C \\ -C & B \end{bmatrix} \qquad \text{这里 } A=B+iC,\, B \text{ 和 } C \text{ 是实方阵}$$

证明 φ 满足如下性质:

(a) $\varphi(A_1A_2)=\varphi(A_1)\varphi(A_2)$, $\varphi(\bar{A}^T)=\varphi(A)^T$

(b) A 是厄米方阵当且仅当 $\varphi(A)$ 是对称方阵

(c) A 是酉方阵当且仅当 $\varphi(A)$ 是正交方阵

31. 设 n 阶复方阵 A 有如下分块结构

$$A=\begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \qquad B \text{ 是 } m \text{ 阶复方阵},\ D \text{ 是 } n-m \text{ 阶复方阵}$$

证明 A 是酉方阵当且仅当 B, D 是酉方阵并且 $C=0$。

32. 设 n 阶复方阵 A 的所有特征值为 $\lambda_1,\cdots,\lambda_n$。证明

$$\sum_{i=1}^{n}|\lambda_i|^2 \le \mathrm{Tr}(\bar{A}^TA)$$

等号成立当且仅当 A 是正规矩阵。

33. 设 A 是 n 阶厄米方阵。证明 A 是正定方阵当且仅当存在 n 阶可逆矩阵 P 使得 $A=\bar{P}^TP$。

34. 计算矩阵 $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ -2 & 1 & 0 \end{bmatrix}$ 的极分解。

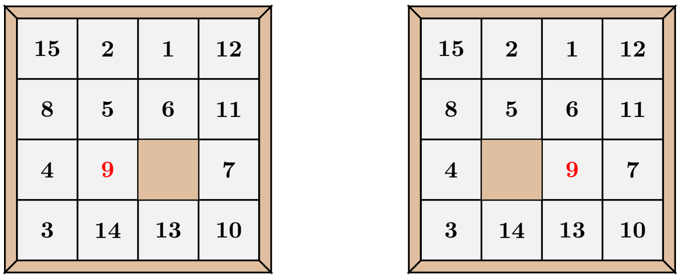
附录 A　几何与对称

本文来源于作者 2018 年在清华大学为首届丘成桐英才班开设的教改课程《几何与对称》的部分内容 [1]，通过几何直观和棋盘游戏实例来解释“群”这个重要数学概念的一些基本思想和应用方法。我们把部分内容作为附录放在这份讲义里，希望能帮助读者更好地了解我们在定义行列式时使用的置换变换，并初步了解群这个概念。

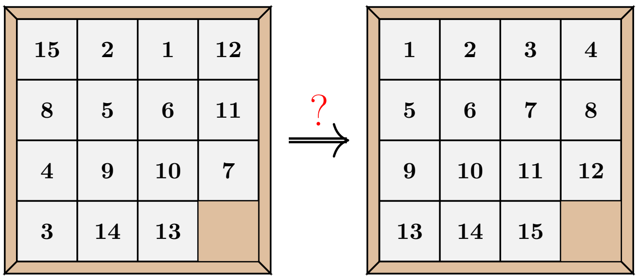
引言

大学数学系里通常有一门基础课称为“抽象代数”，主要内容是群、环、域等这些现代代数学中的基本概念和方法。读书的时候感觉这些概念初学起来的确很“抽象”，就像这门课的名称一样。随着不断学习和积累，逐渐发现这些概念其实非常直观，朋友们也经常笑谈到这门课应该被教为“非抽象代数”。本文通过一个具体的数字游戏来聊聊非抽象代数。

我们探讨的这个经典例子称为数字推盘问题，玩法类似于华容道。考虑一个 4×4 的方形棋盘以及标记 $1-15$ 的数字方块。游戏者每次移动一个数字方块到相邻空白格，例如



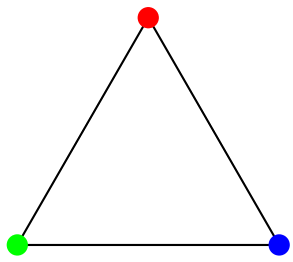
问题: 是否可以由下图左还原到下图右的原始状态?



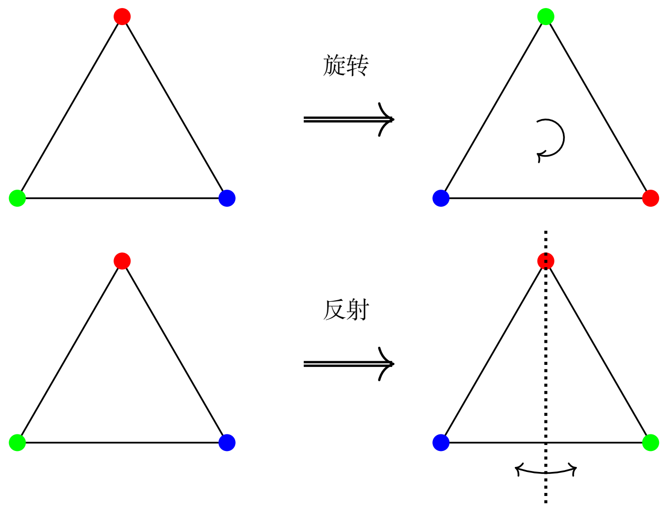
我们将围绕如何解决这个经典问题来解释“群”这个基本数学概念。

几何变换与对称性

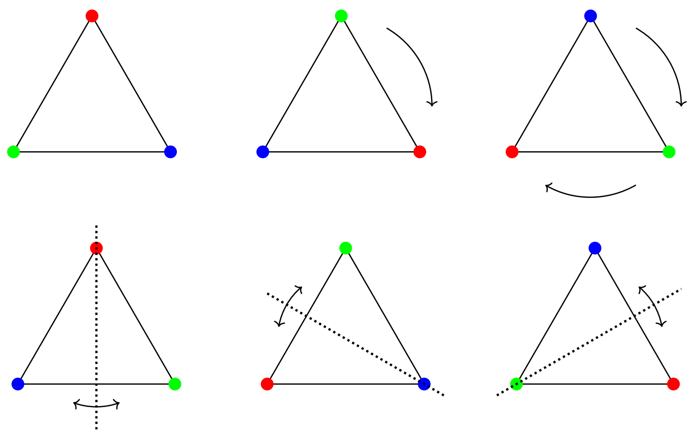
具有几何对称的物体通常都有强烈美感和丰富结构，“群”即刻画如此。考虑等边三角形



这个图形具有很强的几何对称性：我们可以通过如下的几何变换把三角形变换到它自己

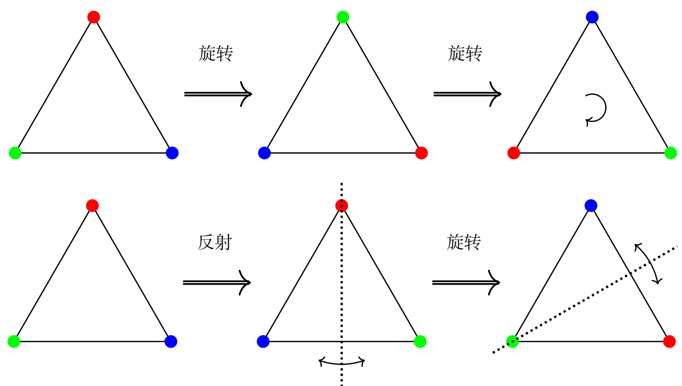


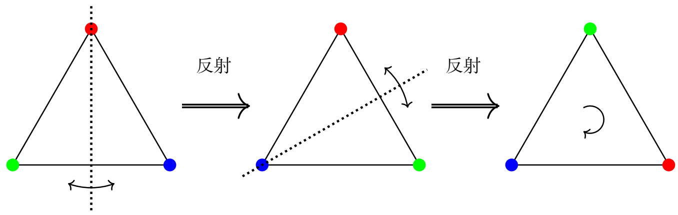
我们总共得到如下 6 个对称变换



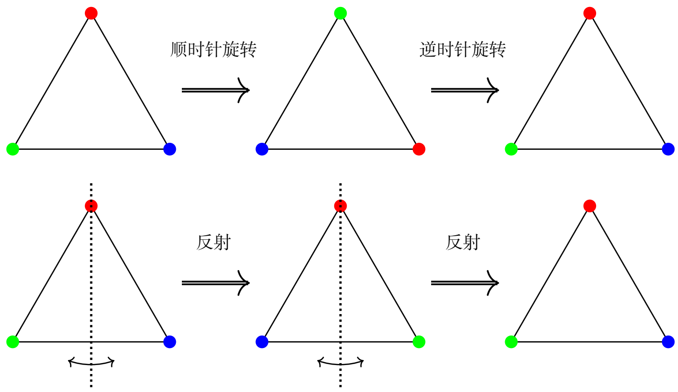
观察到这些几何变换满足如下两个重要性质：

1. 对称变换的复合还是对称变换





2. 对称变换可以由逆变换来还原



群的概念

几何对称变换的这些性质构成的数学对象称为“群”。

抽象定义而言，一个群由一个集合 G 和满足结合律的二元运算

$$\bullet : G \times G \to G$$

（称为群 G 中的乘积）组成，满足

- 单位元：存在单位元1_G，使得：$1_G \bullet g = g \bullet 1_G = g$
- 可逆性：对任一元素 g，存在逆元g^{-1} 使得：$g \bullet g^{-1} = g^{-1} \bullet g = 1_G$

例如：

1. $G = \mathbb{R} - \{0\}$，$\bullet$ 为实数的乘法 $a \bullet b := ab$

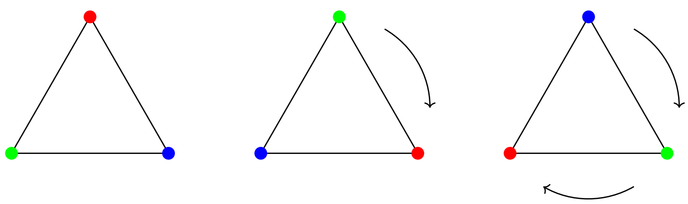
$$单位元 = 1, \quad a \text{ 的逆元是 } \frac{1}{a}$$

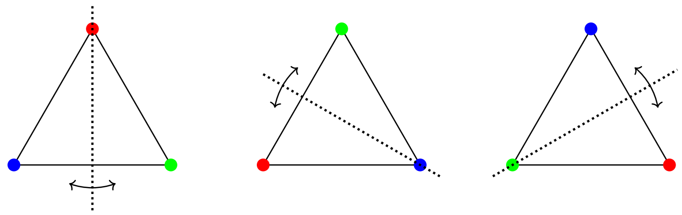
2. $G = \mathbb{Z}$，$\bullet$ 为整数的加法 $m \bullet n := m + n$

$$单位元 = 0, \quad m \text{ 的逆元是 } -m$$

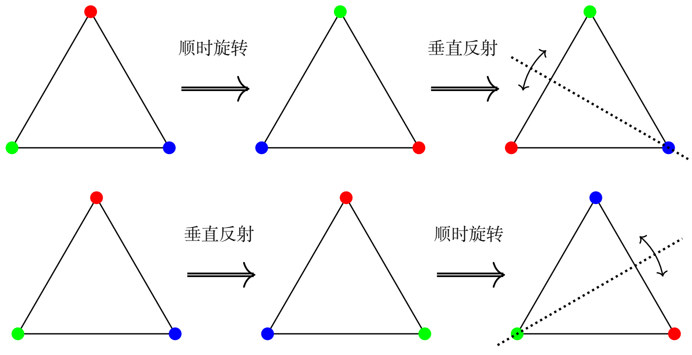
上述例子中 $\bullet$ 是交换的，即满足：$a \bullet b = b \bullet a$。乘积满足交换性质的群称为阿贝尔群。

例如：如下 6 个几何变换组合成一个群（称为二面体群D_3）

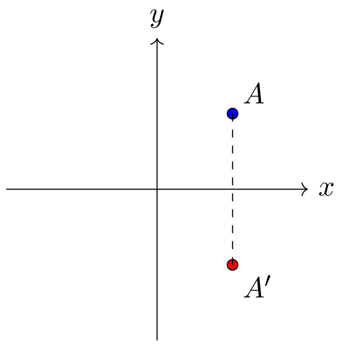


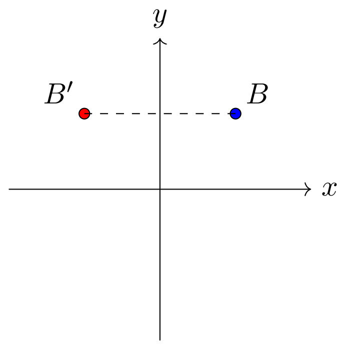


D_3 是非阿贝尔群：旋转和反射不交换！



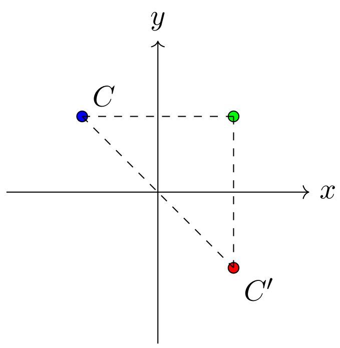
例如：考虑平面上的几何变换





σ_x= 沿 x-轴作镜像反射　　　　σ_y= 沿 y-轴作镜像反射

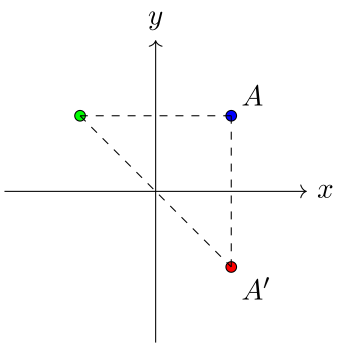
这两个变换满足 $\sigma_x \cdot \sigma_x = \sigma_y \cdot \sigma_y =$ 恒等变换。我们将两个变换 σ_x, σ_y 相复合，记为 σ_{xy}。



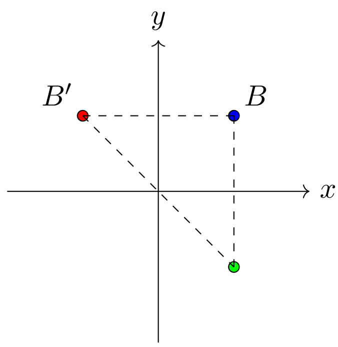
σ_{xy}= 沿原点作对径映射

$\sigma_{xy} = \sigma_x \cdot \sigma_y = \sigma_y \cdot \sigma_x$ 且 $\sigma_{xy} \cdot \sigma_{xy} =$ 恒等变换。

我们考虑变换 σ_{xy} 与 σ_x 以及 σ_y 的复合



$\sigma_{xy} \cdot \sigma_y = \sigma_x$



$\sigma_{xy} \cdot \sigma_x = \sigma_y$

我们得到如下的变换复合关系：

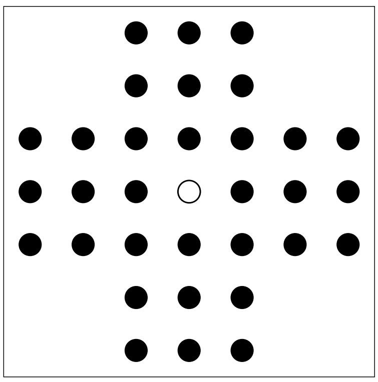
$$\sigma_x \cdot \sigma_x = \sigma_y \cdot \sigma_y = \sigma_{xy} \cdot \sigma_{xy} = \text{恒等变换}$$
$$\sigma_x \cdot \sigma_y = \sigma_y \cdot \sigma_x = \sigma_{xy}$$
$$\sigma_y \cdot \sigma_{xy} = \sigma_{xy} \cdot \sigma_y = \sigma_x$$
$$\sigma_{xy} \cdot \sigma_x = \sigma_x \cdot \sigma_{xy} = \sigma_y$$

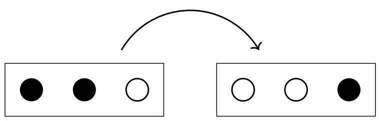
这些变换 $\{1, \sigma_x, \sigma_y, \sigma_{xy}\}$ 构成一个阿贝尔群，称为Klein 四元群。特别的我们有

$$\sigma_x \cdot \sigma_y \cdot \sigma_{xy} = \text{恒等变换（单位元 1）}$$

一个古老的游戏 (Peg Solitaire)

群的结构表现为变换。在实际运用中，一个重要的思想是在一系列丰富的变换中找到不变的对象。作为一个例子，我们考虑一个古老的游戏：

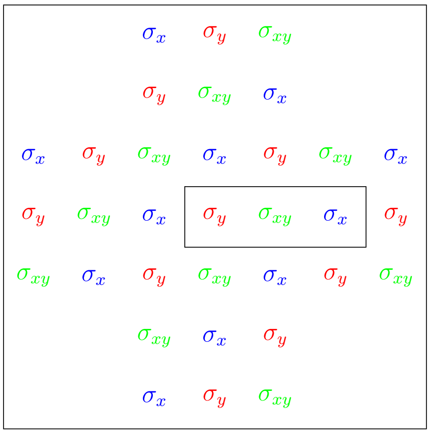




游戏规则: 每次把一个黑子隔着相邻 (水平或者垂直) 的另一个黑子跳到空白处，同时移除被跳过的黑子。如上图。

问题: 是否可以使得棋盘上最后只剩下一个黑子? 如果可以，这个黑子落在哪里?

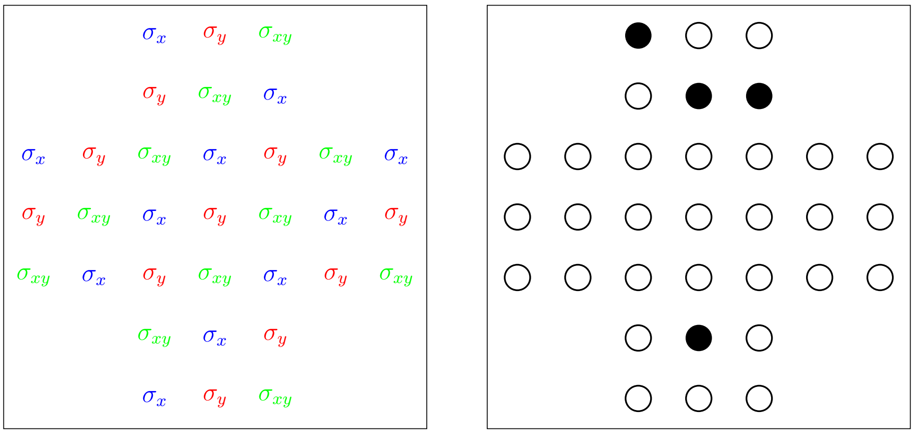
通过实践很容易发现第一问的答案是肯定的，我们用群论方法来分析第二问。把 Klein 四元群元素如下标记到棋盘中



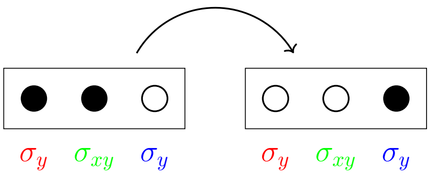
这个标记方法使得直线上任意三个相邻元素的乘积都是恒等变换单位元 1。

设 C 是棋盘上的一个黑子分布。定义 C 的权为

$$[C] = \text{所有的黑子位置上的 Klein 四元群元素相乘}$$



$$\text{上图右的权} = \sigma_x \cdot \sigma_{xy} \cdot \sigma_x \cdot \sigma_x = \sigma_y$$



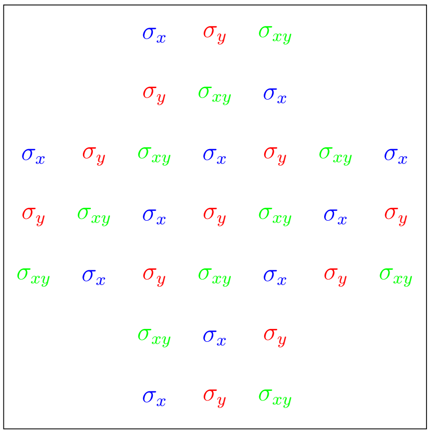
$$\sigma_y \cdot \sigma_{xy} = \sigma_x$$

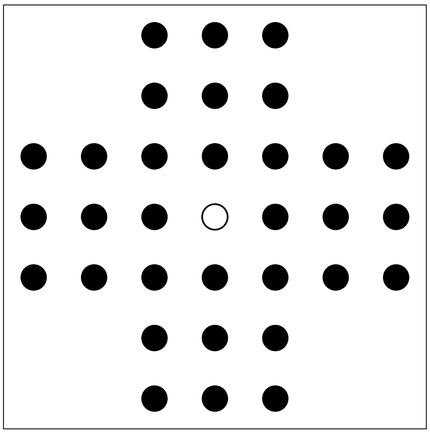
观察到：如果 C 走一步得到 C'，则

$$[C] = [C']$$

具有相同的权。因此“权”为游戏过程中的一个不变量。

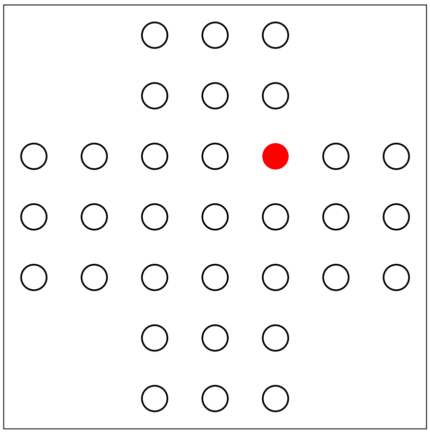
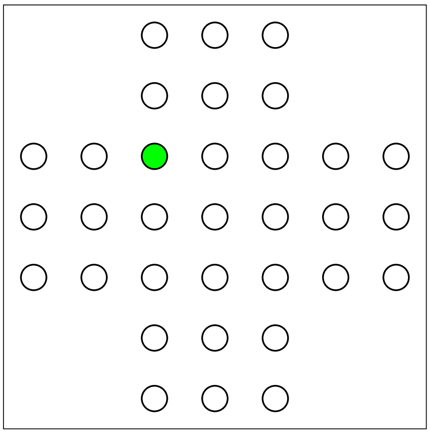
容易计算初始黑子分布的权为 σ_y



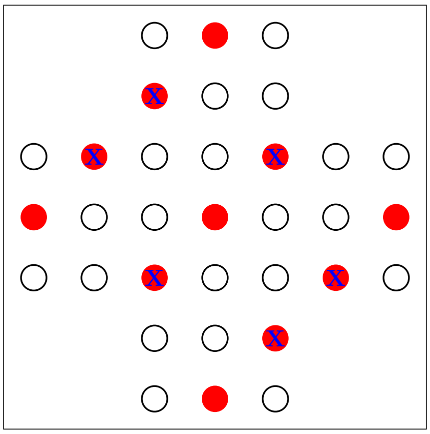


因此最后一个黑子只可能落在标记为 σ_y 的位置。

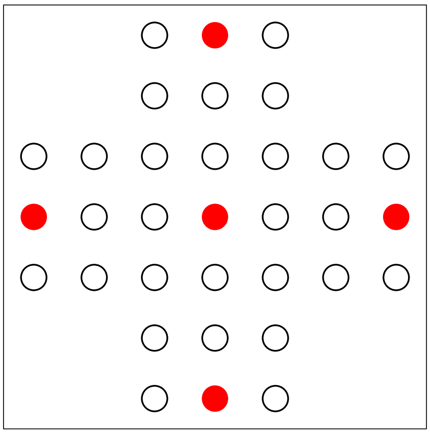
黑子最后不可能落在如下右图绿点的位置 (权为σ_{xy})。由左右对称性，可以排除最后一个黑子落在如下左图的可能性（否则可以采取左右镜像的走法使得最后落在右图绿点）。

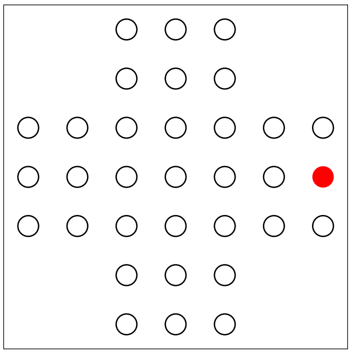
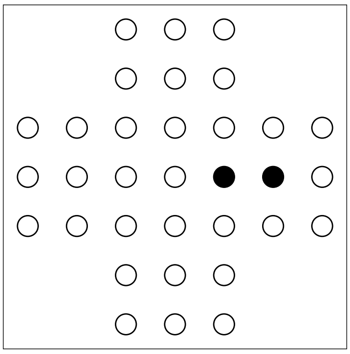
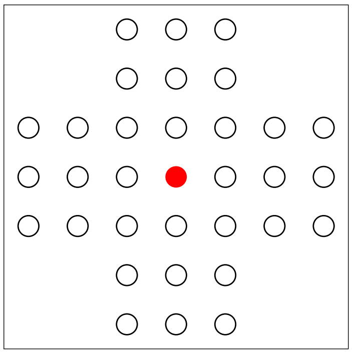
类似对称性分析，我们可以共排除如下 6 种可能



最后一个黑子只能落在如下图中的某个位置



实际上这 5 种可能性都可以达到，而且我们总可以使得最后一个黑子落在正中间。假如最后一个黑子落在如下左图的位置，则上一步如中间图。可以改变走法使得最后一个黑子落在右图的位置。这样我们就完整地解决了这个问题。

置换群

集合 $\{1,2,\cdots,n\}$ 到自己的一个一一映射称为一个n 元置换。例如：如下映射 σ 是一个 5 元置换

$$\sigma : 1,2,3,4,5 \to 3,1,5,2,4$$

两个置换的复合也是置换。所有 n 元置换构成一个群 (称置换群)，记为 S_n。

一个 n 元置换 σ 可以记为

$$\sigma = \begin{bmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{bmatrix}$$

σ 表示将 1 变成 i_1，2 变成 i_2，...，n 变成 i_n。例如

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 2 & 4 \end{bmatrix}$$

表示置换 $\sigma : 1,2,3,4,5 \to 3,1,5,2,4$。乘积 $\alpha\beta$ 表示先作用 β 置换，再作用 α 置换。例如

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

如果置换 σ 把其中 k 个元素映射为

$$\sigma : i_1, i_2, i_3, \cdots, i_k \to i_2, i_3, \cdots, i_k, i_1$$

而把其他元素映射到自己，我们称 σ 是一个长度为 k 的轮换，并把 σ 简化标记为

$$\sigma = (i_1\ i_2\ \cdots\ i_k)$$

例如：

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 4 & 2 \end{bmatrix} = (2\ 3\ 5)$$

任意一个置换总可以表示为一些轮换的乘积，例如

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 6 & 2 & 3 & 5 \end{bmatrix} = (1)(24)(365) = (24)(365)$$

其中 (1) 表示的是恒等置换，可以省略不记。

交错群（偶置换群）

长度为 2 的轮换 $(i\ j)$ 称为一个对换。例如

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 4 & 2 \end{bmatrix} = (2\ 5)$$

每个轮换均可以表示为一些对换的乘积，但是这种表达方式不唯一。例如

$$(1\ 2\ 3) = (1\ 3)(1\ 2) = (1\ 2)(2\ 3)$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}$$

由于每个置换总可以表示为一些轮换的乘积，因此每个置换也总可以表示成一些对换的乘积。虽然表达法不唯一，但是容易发现不同表达法里包含的对换个数的奇偶性是确定不变的，称为该置换的奇偶性。例如：长度 k 的轮换可以写成 $k-1$ 个对换相乘

$$(1\ 2\ 3\ \ldots\ k) = (1\ k)(1\ k-1)\cdots(1\ 3)(1\ 2)$$

因此我们知道奇长度的轮换是偶置换，偶长度的轮换是奇置换。例如

$$\begin{aligned} \text{奇置换：} &\quad (2\ 3), \quad (1\ 3\ 4\ 7) = (1\ 7)(1\ 4)(1\ 3) \\ \text{偶置换：} &\quad (1\ 2\ 3) = (1\ 3)(1\ 2), \quad (4\ 3\ 2\ 1\ 6) \end{aligned}$$

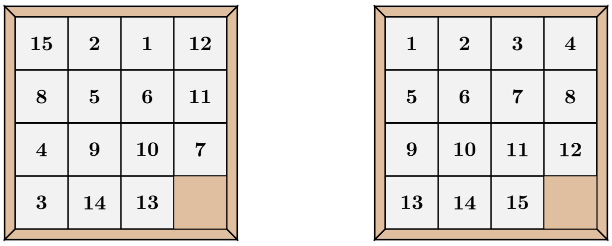
置换的复合满足

$$\begin{aligned} \text{奇置换} \times \text{奇置换} &\to \text{偶置换} \\ \text{奇置换} \times \text{偶置换} &\to \text{奇置换} \\ \text{偶置换} \times \text{偶置换} &\to \text{偶置换} \end{aligned}$$

因此 S_n 中所有偶置换构成一个群（称为交错群），记为 A_n 。

数字推盘问题的解-必要条件

问题: 是否可以由图左还原到图右原始位置?

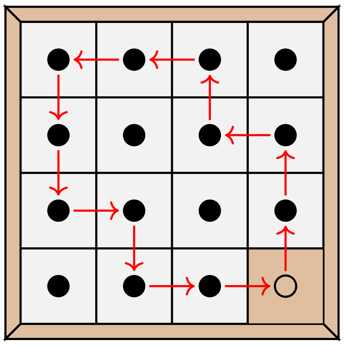


上图左对应于一个置换

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 15 & 2 & 1 & 12 & 8 & 5 & 6 & 11 & 4 & 9 & 10 & 7 & 3 & 14 & 13 & 16 \end{bmatrix}$$

这里我们把空白标记为 16。

如果棋盘的布置可以从原始的位置（上图右）移动得到，它对应的置换 σ 应该满足什么性质? 移动过程可以通过记录空白位置的走动来标志



每个移动为 S_{16} 中的一个对换 (空白标志为 16)。如果要求将空白移回到出发点，则

往上走步数 = 往下走步数，往左走步数 = 往右走步数

总共要移动偶数步! 因此数字推盘问题可解的一个必要条件是

推盘对应的数字置换是偶置换

回到这节开始的问题，推盘对应的置换

$$\begin{aligned} \sigma &= \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 15 & 2 & 1 & 12 & 8 & 5 & 6 & 11 & 4 & 9 & 10 & 7 & 3 & 14 & 13 & 16 \end{bmatrix} \\ &= (1\ 15\ 13\ 3)(4\ 12\ 7\ 6\ 5\ 8\ 11\ 10\ 9) \end{aligned}$$

是奇置换，因此我们知道它对应的数字推盘不可能还原到原始状态。

进一步，我们可以考虑

问题：如果数字推盘对应的数字置换是偶置换，是否一定可以还原?

答案是肯定的。为了证明这个结论，我们需要引入群的生成元的概念。

群的生成元

群 G 的一组元素称为生成元，如果 G 中每个元素都可以通过这组元素反复相乘得到。

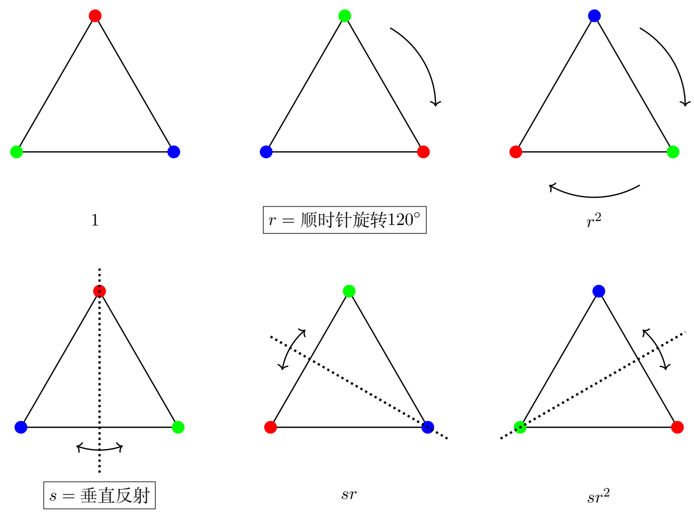
例如：Klein 群　$K = \{1, a, b, c\}$

$$a \bullet b = c \qquad b \bullet c = a \qquad c \bullet a = b$$
$$a^2 = 1 \qquad b^2 = 1 \qquad c^2 = 1$$

- $\{a, b\}$ 是一组生成元。比如 1 可以由 $a \bullet a$ 得到，c 可以由 $a \bullet b$ 得到。
- $\{b, c\}$ 也是一组生成元。因此生成元的选法并不唯一。

例如：二面体群 D_3 的一组生成元是 $\{r, s\}$，它们满足

$$r^3 = 1, \quad s^2 = 1, \quad sr = r^2 s$$



例如：置换群 S_n 可以由如下对换生成

$$(1\ 2), \quad (1\ 3) \quad , \quad \cdots \quad , \quad (1\ n)$$

比如任意的置换 $(i\ j)$ 可以写成

$$(i\ j) = (1\ j)(1\ i)(1\ j)$$

由此可以生成所有的置换。类似的，给定其他一个数比如 3，如下对换

$$(3\ 1), \quad (3\ 2) \quad ,(3\ 4) \quad \cdots \quad , \quad (3\ n)$$

也构成 S_n 的一组生成元。

例如：交错群 A_n (即 S_n 中的偶置换) 的一组生成元是

$$(1\ 2\ 3), \quad (1\ 2\ 4) \quad , \quad \cdots \quad , \quad (1\ 2\ n)$$

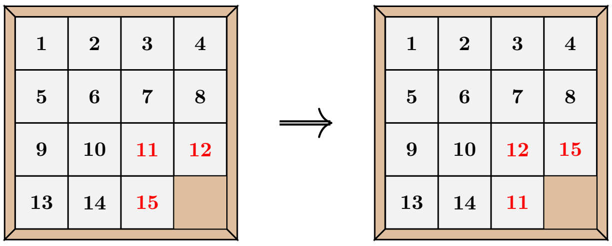
类似的，给定其他两个数比如 2, 4，如下轮换

$$(2\ 4\ 1), \quad (2\ 4\ 3) \quad ,(2\ 4\ 5) \quad \cdots \quad , \quad (2\ 4\ n)$$

也构成 A_n 的一组生成元。读者可以尝试着证明这个结论。

数字推盘问题的解-充分条件

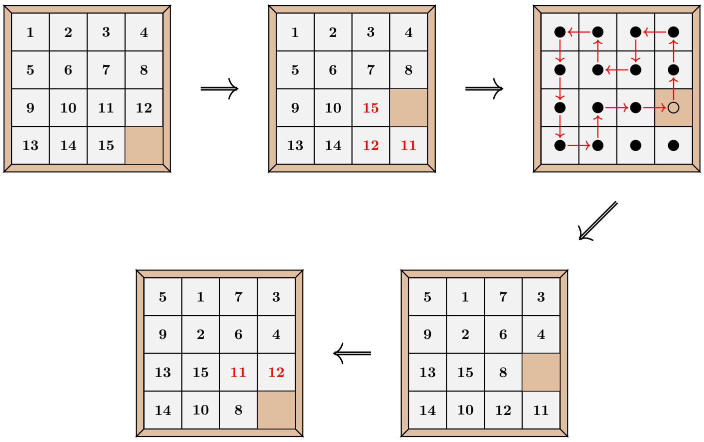
我们下面来分析求解数字推盘问题的充分条件。首先我们考虑两个特殊的移动。第一个移动为：



通过如上移动，我们知道如下置换是可以还原的

$$\tau = (11\ \ 12\ \ 15)$$

第二个移动为：



通过如上移动，我们知道如下置换是可以还原的

$$\begin{aligned}\rho &= \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ 5 & 1 & 7 & 3 & 9 & 2 & 6 & 4 & 13 & 15 & 11 & 12 & 14 & 10 & 8 \end{bmatrix} \\ &= (1\ 5\ 9\ 13\ 14\ 10\ 15\ 8\ 4\ 3\ 7\ 6\ 2)\end{aligned}$$

观察到：如果置换 σ_1, σ_2 可以还原，那么

- $\sigma_1\sigma_2$ 可以还原：通过复合操作实现
- σ_1^{-1} 可以还原：通过逆向操作实现

换言之，所有可以还原的置换构成了一个群。

通过上述两个移动，我们知道如下两个置换是可以还原的

$$\tau = (11\ \ 12\ \ 15), \quad \rho = (1\ 5\ 9\ 13\ 14\ 10\ 15\ 8\ 4\ 3\ 7\ 6\ 2)$$

因此由它们及其逆元复合生成的置换均可还原。直接计算发现

$$\tau, \quad \rho\tau\rho^{-1}, \quad \rho^2\tau\rho^{-2}, \quad \cdots \quad , \rho^{12}\tau\rho^{-12}$$

恰好构成集合（顺序不一致）

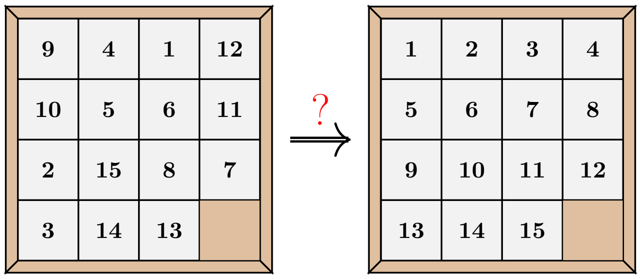
$$(11\ 12\ 1),\cdots(11\ 12\ 10),(11\ 12\ 13),(11\ 12\ 14),(11\ 12\ 15)$$

因此这些置换均可以被还原。由上述讨论，我们知道元素

$$(11\ 12\ 1),\cdots(11\ 12\ 10),(11\ 12\ 13),(11\ 12\ 14),(11\ 12\ 15)$$

生成所有的 15 元的偶置换。因此任意一个对应于偶置换的数字推盘均可以还原！总结上述讨论，我们通过置换群的结构证明了如下结论：

定理. 数字推盘可以被还原到初始状态当且仅当其对应的数字置换是偶置换。



例如对于上面这个问题，它对应的置换

$$\begin{aligned}\sigma=&\begin{bmatrix}1&2&3&4&5&6&7&8&9&10&11&12&13&14&15\\9&4&1&12&10&5&6&11&2&15&8&7&3&14&13\end{bmatrix}\\=&(1\ 9\ 2\ 4\ 12\ 7\ 6\ 5\ 10\ 15\ 13\ 3)(8\ 11)\end{aligned}$$

是偶置换，因此我们知道它一定可以被还原。

参考文献

[1] Li, Si. https://sili-math.github.io/