

# 线性代数 38 讲

## 目录

<b>1</b>	<b>引子</b>	<b>4</b>
1.1	三角形中线交点的向量证明 . . . . .	4
1.2	鸡兔同笼问题 . . . . .	6
1.3	复数和四元数 . . . . .	8
<b>2</b>	<b>向量空间 <math>\mathbb{R}^n</math>, 线性函数, 线性映射与矩阵</b>	<b>9</b>
<b>3</b>	<b>矩阵基本运算</b>	<b>11</b>
3.1	矩阵的加法和数乘 . . . . .	11
3.2	矩阵的乘法 . . . . .	11
3.2.1	矩阵乘法示例 . . . . .	12
3.2.2	矩阵乘法公式推导 . . . . .	12
3.2.3	斐波那契数列的矩阵解法 . . . . .	12
3.2.4	马的遍历, 分块矩阵的乘法 . . . . .	14
3.3	逆矩阵, Hill 密码 . . . . .	16
<b>4</b>	<b>线性方程组与线性子空间</b>	<b>19</b>
4.1	理解线性方程组的三个视角 . . . . .	19
4.2	$Ax = b$ 的解的情况 (“0, 1, $\infty$ ”) . . . . .	20
<b>5</b>	<b>线性表示, 线性子空间, 线性映射的值域空间和零空间</b>	<b>21</b>
<b>6</b>	<b>线性相关, 线性无关, 空间的基</b>	<b>22</b>
<b>7</b>	<b>基的存在性, 空间维数, 置换引理</b>	<b>23</b>
<b>8</b>	<b>置换引理的证明</b>	<b>25</b>
<b>9</b>	<b>基与坐标, 线性映射和矩阵</b>	<b>27</b>
<b>10</b>	<b>维数与秩, 线性同构与逆矩阵</b>	<b>29</b>

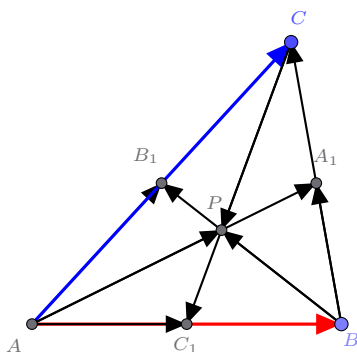
11 矩阵的列变换与行变换	30
12 行变换, Gauss 消元, 矩阵的 LU 分解	34
13 逆矩阵的性质与计算	36
14 行变换的几何意义	38
15 对偶空间和对偶映射, 矩阵的转置	39
16 线性映射和它的对偶映射的关系	41
17 线性方程组应用举例	44
18 线性方程组在经济学中的应用: 投入产出模型	45
19 投入产出理论中的一个关于非负矩阵的结论	47
20 内积空间	48
21 内积空间中的度量	50
22 内积空间中的基, Gram-Schmidt 正交化过程	51
23 正交变换	53
24 线性子空间及其运算	57
25 最小二乘法	61
26 Riesz 表示定理, 内积空间上的线性函数与对偶映射	64
27 双线性型与二次型	66
28 二次型的标准型, 对称矩阵的谱分解	68
29 二次型、二次曲线和二次曲面的分类	71
30 奇异值分解	76
31 多重线性函数与行列式	78
32 行列式的性质与计算	81
33 行列式的应用	83

34 线性变换的特征值与特征向量	86
35 矩阵的特征多项式, Cayley-Hamilton 定理	90
36 幂零矩阵和循环子空间, 根子空间	92
37 矩阵的 Jordan 标准形	96
38 域和向量空间	98

# 1 引子

“线性代数”这个词首次出现是在 1930 年 van der Waerden 所著的书里. 它是研究一些被称为向量的对象是如何通过线性关系（特别是线性映射）彼此之间建立联系的.

## 1.1 三角形中线交点的向量证明



设  $A_1, B_1, C_1$  分别是  $\triangle ABC$  边  $BC, CA, AB$  的中点,  $P$  是中线  $BB_1$  和  $CC_1$  的交点. 我们证明另一条中线  $AA_1$  也经过点  $P$ .

记

$$\overrightarrow{AB} = \mathbf{b}, \quad \overrightarrow{AC} = \mathbf{c}.$$

这里箭头表示有向线段, 从起点到终点. 所谓“向量”就是既有方向（即“向”）, 又有大小（即“量”）的对象. 这里有向线段的长度是它们的大小, 起点和终点表明了有向线段的方向. 这时, 中点向量为:

$$\overrightarrow{AB_1} = \frac{1}{2}\mathbf{b}, \quad \overrightarrow{AC_1} = \frac{1}{2}\mathbf{c}.$$

这里我们可以对向量定义数乘运算.

我们还可以定义向量的加法运算,

$$\overrightarrow{AB_1} = \overrightarrow{AB} + \overrightarrow{BB_1},$$

此时我们把有向线段当作位移, 向量加法就是两次位移的合成.

由此得到

$$\begin{aligned} \overrightarrow{BB_1} &= \overrightarrow{AB_1} - \overrightarrow{AB} = \frac{1}{2}\mathbf{c} - \mathbf{b}, \\ \overrightarrow{CC_1} &= \overrightarrow{AC_1} - \overrightarrow{AC} = \frac{1}{2}\mathbf{b} - \mathbf{c}. \end{aligned}$$

这里减法是加上相反向量，其中

$$-\overrightarrow{AB} = (-1)\overrightarrow{AB} = \overrightarrow{BA}.$$

这样数乘运算允许用任何数（不仅是正数）乘.

因为  $P$  在直线  $BB_1$  和  $CC_1$  上，所以可设  $\overrightarrow{BP} = \lambda \overrightarrow{BB_1}$ ,  $\overrightarrow{CP} = \mu \overrightarrow{CC_1}$ . 表达  $\overrightarrow{AP}$ :

$$\begin{aligned}\overrightarrow{AP} &= \overrightarrow{AB} + \overrightarrow{BP} = \mathbf{b} + \lambda \left( \frac{1}{2} \mathbf{c} - \mathbf{b} \right) = (1 - \lambda) \mathbf{b} + \frac{\lambda}{2} \mathbf{c}, \\ \overrightarrow{AP} &= \overrightarrow{AC} + \overrightarrow{CP} = \mathbf{c} + \mu \left( \frac{1}{2} \mathbf{b} - \mathbf{c} \right) = \frac{\mu}{2} \mathbf{b} + (1 - \mu) \mathbf{c}.\end{aligned}$$

于是我们得到

$$\left( 1 - \lambda - \frac{\mu}{2} \right) \mathbf{b} = \left( 1 - \mu - \frac{\lambda}{2} \right) \mathbf{c},$$

但  $\mathbf{b}, \mathbf{c}$  不在同一条直线上，所以

$$1 - \lambda - \frac{\mu}{2} = 1 - \mu - \frac{\lambda}{2} = 0,$$

整理得到方程组

$$\begin{cases} \mu + 2\lambda = 2, \\ \lambda + 2\mu = 2, \end{cases}$$

解得

$$\lambda = \frac{2}{3}, \quad \mu = \frac{2}{3}.$$

于是

$$\overrightarrow{AP} = \frac{1}{3}(\mathbf{b} + \mathbf{c}).$$

另一方面

$$\overrightarrow{AA_1} = \overrightarrow{AB} + \frac{1}{2} \overrightarrow{BC} = \overrightarrow{AB} + \frac{1}{2}(\overrightarrow{AC} - \overrightarrow{AB}) = \frac{1}{2}(\mathbf{b} + \mathbf{c}),$$

所以

$$\overrightarrow{AP} = \frac{2}{3} \overrightarrow{AA_1},$$

故  $P$  在中线  $AA_1$  上.

**注意：**请找出以上解答中的漏洞.

如果我们用  $P: \begin{bmatrix} x \\ y \end{bmatrix}$  来表示  $\overrightarrow{AP} = x\mathbf{b} + y\mathbf{c}$ , 那么

$$A: \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad B: \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad C: \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

于是

$$A_1 : \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad B_1 : \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} \quad C_1 : \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}.$$

取  $P : \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$ , 则

$$\begin{aligned} \overrightarrow{BP} &= \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}, \\ \overrightarrow{BB_1} &= \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ \frac{1}{2} \end{bmatrix}, \\ \overrightarrow{BP} &= \frac{1}{3} \overrightarrow{BB_1}, \end{aligned}$$

同理可证,  $\overrightarrow{AP} = \frac{1}{3} \overrightarrow{AA_1}$ ,  $\overrightarrow{CP} = \frac{1}{3} \overrightarrow{CC_1}$ , 因此  $P$  位于三条中线上.

**问题 1.1.** 设  $A_1, B_1, C_1$  分别是三角形  $ABC$  三边  $BC, CA, AB$  上的点. 证明  $AA_1, BB_1, CC_1$  交于同一点当且仅当线段长度的比值满足

$$\frac{BA_1}{A_1C} \cdot \frac{CB_1}{B_1A} \cdot \frac{AC_1}{C_1B} = 1.$$

## 1.2 鸡兔同笼问题

笼中有鸡兔若干, 头 10 个, 腿 24 只, 问鸡兔各几只? <sup>1</sup>

解: 设鸡  $x$  只, 兔  $y$  只:

$$\begin{cases} x + y = 10 & (\text{头}) \\ 2x + 4y = 24 & (\text{腿}) \end{cases}$$

每个方程都是未知数  $x, y$  的线性组合, 这样的方程组称为线性方程组.

我们提取两种动物头和腿的特征, 分别用

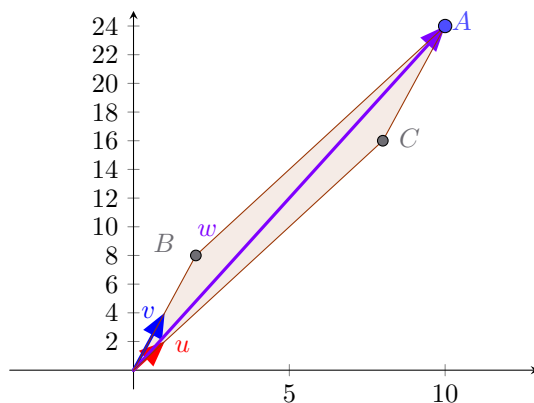
$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

表示鸡和兔子. 于是原问题可以表示为如下向量形式:

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \cdot x + 1 \cdot y \\ 2 \cdot x + 4 \cdot y \end{bmatrix} = \begin{bmatrix} 10 \\ 24 \end{bmatrix},$$

我们可以用自然的方式理解其中分量的运算以及向量相等的关系. 这样问题转化为, 把向量  $\begin{bmatrix} 10 \\ 24 \end{bmatrix}$  表示为另外两个向量  $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$  和  $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$  的线性组合, 求组合系数.

<sup>1</sup> “鸡兔同笼”是一个经典的中国古代数学问题, 它出自于中国古代数学著作《孙子算经》。原文为: “今有雉兔同笼, 上有三十五头, 下有九十四足, 问雉兔各几何?”



我们还可以用函数的观点，

$$H : (x, y) \mapsto H(x, y) = x + y, \quad L : (x, y) \mapsto L(x, y) = 2x + 4y,$$

分别是计算头的总数和腿的总数，它们都是关于鸡和兔子数量的线性函数.

$$H(x, y) = 10, \quad L(x, y) = 24$$

分别是坐标平面中的两条直线，方程组的解就是这两条直线的交点.

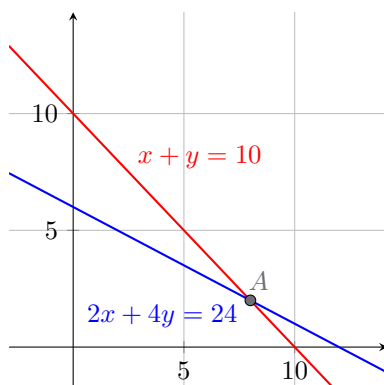
我们用行向量  $(1, 1)$  和  $(2, 4)$  分别表示这两个函数中  $x, y$  的系数，并规定

$$(1, 1) \begin{bmatrix} x \\ y \end{bmatrix} = 1 \cdot x + 1 \cdot y, \quad (2, 4) \begin{bmatrix} x \\ y \end{bmatrix} = 2 \cdot x + 4 \cdot y.$$

所以，列向量是向量，行向量是作用于列向量的函数.

我们也可以把这两个函数摺在一起，写成矩阵形式：

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \cdot x + 1 \cdot y \\ 2 \cdot x + 4 \cdot y \end{bmatrix} = \begin{bmatrix} 10 \\ 24 \end{bmatrix}$$



### 1.3 复数和四元数

在探索求解三次四次多项式方程的过程中，人们发现了  $\mathbf{i} = \sqrt{-1}$ ，并把它和实数一起进行运算得到  $a + b\mathbf{i}$  ( $a, b \in \mathbb{R}$ )，一般地，

$$\begin{aligned}(a_1 + b_1\mathbf{i}) + (a_2 + b_2\mathbf{i}) &= (a_1 + a_2) + (b_1 + b_2)\mathbf{i}, \\(a_1 + b_1\mathbf{i}) \cdot (a_2 + b_2\mathbf{i}) &= (a_1a_2 - b_1b_2) + (a_1b_2 + a_2b_1)\mathbf{i}.\end{aligned}$$

Hamilton<sup>2</sup> 发现可以把  $x + y\mathbf{i}$  看成实数对  $(x, y)$ ，从而作为坐标平面中的点，并规定相应的运算.

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}$$

这个加法符合“平行四边形法则”，满足相应的运算规律.

而乘法  $(a + b\mathbf{i})(x + y\mathbf{i})$  可以看成映射

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} ax - by \\ ay + bx \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

当  $a, b$  不全为零时，可设  $a = \sqrt{a^2 + b^2} \cos \theta, b = \sqrt{a^2 + b^2} \sin \theta$ ，则

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} \sqrt{a^2 + b^2} & 0 \\ 0 & \sqrt{a^2 + b^2} \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

第一个作用是旋转，第二个作用是伸缩，这个分解叫“极分解”.

共轭  $\overline{x + y\mathbf{i}} = x - y\mathbf{i}$  对应于映射

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

它是平面上的一个反射.

Hamilton 尝试把他对复数的发现推广到三维，但他历经十多年寻找“三元数”的尝试始终不成功，直到 1843 年10月16日他步行前往都柏林皇家学会途径皇家运河上的布鲁姆桥时，灵感如闪电般击中他. 他意识到，要得到一个自治的、模守恒的、可除的代数系统，必须牺牲一个代数的基本定律——交换律，并且需要四个分量，而不是三个，这就是四元数

$$a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}, \quad a, b, c, d \in \mathbb{R},$$

其中  $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$ .

为防止这个划时代的灵感消失，他立刻用随身携带的小刀将这一公式刻在了布鲁姆桥的石头上. 如今，桥上有一块石碑纪念这一伟大的科学发现时刻，这一刻不仅是数学史上一个重要的里程碑，也永远改变了代数学的发展方向.

<sup>2</sup>William Rowan Hamilton (1805-1865)，毫无疑问是爱尔兰最伟大的数学家



## 2 向量空间 $\mathbb{R}^n$ ，线性函数，线性映射与矩阵

Hamilton 在布鲁姆桥上闪现的灵感，使得人们不再拘泥于我们熟悉的平面和三维空间，Hamilton 称它是“想象力的一次巨大跨越”<sup>3</sup>.

定义  $n$  维实向量空间：

$$\mathbb{R}^n = \left\{ \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mid x_1, x_2, \dots, x_n \in \mathbb{R} \right\}$$

规定线性运算（加法和数乘）：

$$\lambda \mathbf{x} + \mu \mathbf{y} = \begin{bmatrix} \lambda x_1 + \mu y_1 \\ \vdots \\ \lambda x_n + \mu y_n \end{bmatrix}.$$

不难验证这些运算满足我们熟悉的那些运算规律（交换律、结合律、分配律）.

任意向量都可表示为：

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i,$$

其中

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

称  $\mathbf{e}_1, \dots, \mathbf{e}_n$  为  $\mathbb{R}^n$  的**标准基向量**，称  $x_1, x_2, \dots, x_n$  为向量  $\mathbf{x}$  的**坐标**.

线性函数  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  满足：

$$f(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y}),$$

即  $f$  可与向量的线性运算交换顺序.

由线性可得

$$f(\mathbf{x}) = f\left(\sum_{k=1}^n x_k \mathbf{e}_k\right) = \sum_{k=1}^n x_k f(\mathbf{e}_k) = \underbrace{(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))}_{\text{行向量代表函数}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\text{列向量代表自变量}}.$$

行向量与列向量的乘法是线性代数中最基本的运算形式.

<sup>3</sup>见 Israel Kleiner 的书 A History of Abstract Algebra，第5.5节

线性映射  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  满足:

$$A(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda A\mathbf{x} + \mu A\mathbf{y}$$

对应矩阵

$$A = (a_{ij})_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [\boldsymbol{\alpha}_1 \quad \boldsymbol{\alpha}_2 \quad \cdots \quad \boldsymbol{\alpha}_n] = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix},$$

其中  $a_{ij}$  叫做这个矩阵的第  $i$  行第  $j$  列的元素,  $\boldsymbol{\alpha}_j$  是  $A$  的第  $j$  个列向量,  $f_i$  是  $A$  的第  $i$  个行向量.

列向量视角:

$$A\mathbf{x} = \sum_{j=1}^n x_j \boldsymbol{\alpha}_j$$

是  $A$  的列向量的线性组合, 矩阵的第  $j$  列为

$$\boldsymbol{\alpha}_j = A\mathbf{e}_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

$a_{ij}$  是向量  $A\mathbf{e}_j$  的第  $i$  个坐标.

行向量视角: 矩阵的第  $i$  行为

$$f_i = [a_{i1} \quad a_{i2} \quad \cdots \quad a_{in}], \quad f_i(\mathbf{x}) = \sum_{j=1}^n a_{ij} x_j = (f_i(\mathbf{e}_1), f_i(\mathbf{e}_2), \dots, f_i(\mathbf{e}_n)) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

是一个线性函数,  $A\mathbf{x}$  相当于这  $m$  个线性函数同时作用于  $\mathbf{x}$ ,

$$A\mathbf{x} = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}.$$

坐标视角:

$$\mathbf{y} = A\mathbf{x}, \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y_i = \sum_{j=1}^n a_{ij} x_j$$

**例 2.1.**  $\mathbb{R}^n$  上的恒同映射  $\text{id} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\text{id}(\mathbf{x}) = \mathbf{x}$ , 在矩阵视角下对应于 ( $n$  阶) 单位矩阵,

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

**例 2.2.** 对  $1, 2, \dots, n$  的任何排列  $\sigma$  (看成集合  $\{1, 2, \dots, n\}$  到自身的一一对应), 映射

$$(x_1, \dots, x_n) \mapsto (x_{\sigma(1)}, \dots, x_{\sigma(n)})$$

是  $\mathbb{R}^n$  到  $\mathbb{R}^n$  的一个线性映射, 它对应于  $n$  阶矩阵  $P = (p_{ij})$ ,

$$p_{ij} = \begin{cases} 1, & \text{若 } j = \sigma(i), \\ 0, & \text{否则.} \end{cases}$$

这个矩阵称为一个排列阵. 例如,  $n = 3$ ,  $\sigma(1) = 3, \sigma(2) = 1, \sigma(3) = 2$ ,

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

**定义 2.3.**  $n$  行  $n$  列的矩阵称为  $n$  阶方阵.

## 3 矩阵基本运算

### 3.1 矩阵的加法和数乘

对同为  $m$  行  $n$  列的矩阵  $A, B$ , 定义

$$(A + B)\mathbf{x} = A\mathbf{x} + B\mathbf{x}, \quad (A + B)_{ij} = a_{ij} + b_{ij}$$

$$(\lambda A)\mathbf{x} = \lambda(A\mathbf{x}), \quad (\lambda A)_{ij} = \lambda a_{ij}$$

### 3.2 矩阵的乘法

线性映射  $A, B$  的复合映射  $BA : \mathbb{R}^n \xrightarrow{A} \mathbb{R}^m \xrightarrow{B} \mathbb{R}^p$  也是线性映射, 它对应的矩阵可由  $A, B$  的矩阵得到

$$(BA)_{ij} = \sum_{k=1}^m b_{ik} a_{kj}, \quad \text{要求: } B \text{ 的列数} = A \text{ 的行数}$$

$BA$  是一个  $p$  行  $n$  列的矩阵,  $B$  的行数是乘积矩阵  $BA$  的行数,  $A$  的列数是乘积矩阵  $BA$  的列数,  $B$  的列指标和  $A$  的行指标  $k$  是求和的哑元.

单位矩阵是乘法单位元: 对任何矩阵  $A$  ( $m$  行  $n$  列), 都有

$$AI_n = I_m A = A.$$

### 3.2.1 矩阵乘法示例

对矩阵

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 6 \\ 0 & 5 \end{bmatrix}$$

计算  $BA$ :

$$\begin{aligned} BA \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} -1 & 6 \\ 0 & 5 \end{bmatrix} \left( \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} -1 & 6 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 \cdot x + 2 \cdot y \\ 3 \cdot x + 4 \cdot y \end{bmatrix} \\ &= \begin{bmatrix} (-1) \cdot (x + 2y) + 6 \cdot (3x + 4y) \\ 0 \cdot (x + 2y) + 5 \cdot (3x + 4y) \end{bmatrix} = \begin{bmatrix} 17x + 22y \\ 15x + 20y \end{bmatrix} \\ &= \begin{bmatrix} 17 & 22 \\ 15 & 20 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \end{aligned}$$

因此

$$BA = \begin{bmatrix} 17 & 22 \\ 15 & 20 \end{bmatrix}.$$

类似可得

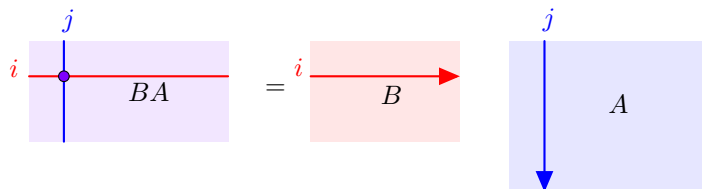
$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 & 6 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} (1)(-1) + (2)(0) & (1)(6) + (2)(5) \\ (3)(-1) + (4)(0) & (3)(6) + (4)(5) \end{bmatrix} = \begin{bmatrix} -1 & 16 \\ -3 & 38 \end{bmatrix}$$

注意  $BA \neq AB$ .

### 3.2.2 矩阵乘法公式推导

$$(BA\mathbf{x})_i = \sum_k b_{ik}(A\mathbf{x})_k = \sum_k b_{ik} \sum_j a_{kj}x_j = \sum_j \left( \sum_k b_{ik}a_{kj} \right) x_j$$

故  $(BA)_{ij} = \sum_k b_{ik}a_{kj}$ .



### 3.2.3 斐波那契数列的矩阵解法

数列定义:  $a_1 = a_2 = 1, a_n = a_{n-1} + a_{n-2} (n \geq 3)$ , 求通项公式.

解： 定义向量序列：

$$\mathbf{b}_n = \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix}, \quad \mathbf{b}_{n+1} = \begin{bmatrix} a_{n+1} \\ a_{n+2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A\mathbf{b}_n$$

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

由初始值  $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  计算得到

$$\mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{b}_4 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \mathbf{b}_5 = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

观察这些点的位置，猜测它们越来越集中于某条特定斜率的直线附近，因此计算斜率值  $\lambda_n = a_{n+1}/a_n$ ：

$$\lambda_1 = 1, \quad \lambda_2 = 2, \quad \lambda_3 = \frac{3}{2} = 1.5, \quad \lambda_4 = \frac{5}{3} \approx 1.667, \quad \lambda_5 = \frac{8}{5} = 1.6$$

它们满足递推关系：

$$\lambda_{n+1} = \frac{a_{n+2}}{a_{n+1}} = \frac{a_{n+1} + a_n}{a_{n+1}} = 1 + \frac{1}{\lambda_n}.$$

假设极限  $\lambda^* = \lim_{n \rightarrow \infty} \lambda_n$  存在，则  $\lambda^* \geq 1$ ，且满足：

$$\lambda^* = 1 + \frac{1}{\lambda^*}.$$

解相应的二次方程，得到  $\lambda^*$  的值：

$$\lambda^{*2} - \lambda^* - 1 = 0 \implies \lambda^* = \frac{1 + \sqrt{5}}{2}.$$

我们来证明这个极限.

$$\begin{aligned} |\lambda_{n+1} - \lambda^*| &= \left| 1 + \frac{1}{\lambda_n} - 1 - \frac{1}{\lambda^*} \right| = \frac{|\lambda_n - \lambda^*|}{\lambda_n \lambda^*} \\ &\leq \frac{2|\lambda_n - \lambda^*|}{3} \quad (\lambda_n > 1, \lambda^* > \frac{3}{2}) \\ &\leq \left(\frac{2}{3}\right)^{n-1} |\lambda_1 - \lambda^*| \rightarrow 0, \quad n \rightarrow +\infty. \end{aligned}$$

注意到上述极限提供了  $A$  的两个不变方向

$$A \begin{bmatrix} 1 \\ \lambda^* \end{bmatrix} = \lambda^* \begin{bmatrix} 1 \\ \lambda^* \end{bmatrix}, \quad A \begin{bmatrix} -\lambda^* \\ 1 \end{bmatrix} = -\frac{1}{\lambda^*} \begin{bmatrix} -\lambda^* \\ 1 \end{bmatrix}$$

由

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = x \begin{bmatrix} 1 \\ \lambda^* \end{bmatrix} + y \begin{bmatrix} -\lambda^* \\ 1 \end{bmatrix}$$

解得

$$x = \frac{1 + \lambda^*}{1 + \lambda^{*2}}, \quad y = \frac{1 - \lambda^*}{1 + \lambda^{*2}}$$

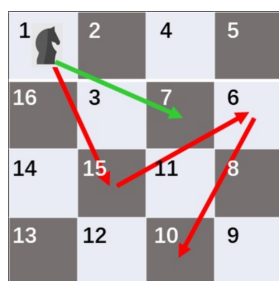
通项公式:

$$\mathbf{b}_n = A^{n-1} \mathbf{b}_1 = \frac{1 + \lambda^*}{1 + \lambda^{*2}} (\lambda^*)^{n-1} \begin{bmatrix} 1 \\ \lambda^* \end{bmatrix} + \frac{1 - \lambda^*}{1 + \lambda^{*2}} \left( -\frac{1}{\lambda^*} \right)^{n-1} \begin{bmatrix} -\lambda^* \\ 1 \end{bmatrix}$$

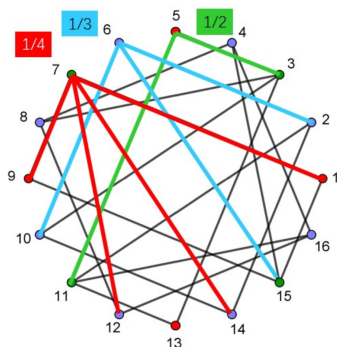
因此得到  $a_n$  的通项公式

$$a_n = \frac{1 + \lambda^*}{1 + \lambda^{*2}} (\lambda^*)^{n-1} + \frac{1}{1 + \lambda^{*2}} \left( -\frac{1}{\lambda^*} \right)^{n-1}.$$

### 3.2.4 马的遍历, 分块矩阵的乘法



国际象棋中的马在一个 4 行 4 列的棋盘区域中移动. 它是否可以从任何一个位置走到另一个位置? 需要多少步? 如果它从一个位置出发, 在所有可能跳到的位置中等可能地随机跳跃, 未来它出现在不同位置的可能性有多大?



我们把 16 个位置进行编号, 然后把它们抽象为点. 如果马从位置  $i$  出发一步可以跳到位置  $j$ , 那么我们就在点  $i$  和点  $j$  之间连一条边. 这样就得到了一个有 16 个顶点的图 (graph).

关联矩阵  $L = (a_{ij})$  反映了这个图的结构, 其中

$$a_{ij} = \begin{cases} 1, & i, j \text{ 间存在边;} \\ 0, & \text{否则.} \end{cases}$$

当前这个图的关联矩阵为

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

$(L^4)_{ij}$  表示从  $i$  出发跳 4 步到  $j$  总共有多少不同的跳法. 计算表明要使矩阵  $L + L^2 + L^3 + L^4 + \cdots + L^k$  的所有元素都非零,  $k$  的最小值为 5. 这说明 5 步之内马可以从任何一个位置出发跳到另一个位置, 少于 5 步就做不到. 比如从位置 1 到位置 5, 4 步以内是走不到的, 但跳 5 步是可以到达的 (例如 1-15-6-2-11-5).

...

☆

计算A+A^2+A^3+A^4+A^5

这是十进制矩阵, 请用python计算

A=A^2+A^3+A^4+A^5

要计算矩阵A的1到5次方之和(A+A^2+A^3+A^4+A^5), 我们可以依次计算各次幂矩阵, 然后进行元素级相加. 以下是实现代码:

计算矩阵A到A^5的和

运行时间: 10.04

代码说明:

- 首先将原始矩阵转换为 NumPy 数组以便进行矩阵运算
- 依次计算矩阵的 2 到 5 次方 (A^2 到 A^5), 每一步都基于前一次的计算结果
- 将 A、A^2、A^3、A^4 和 A^5 进行元素级相加, 得到总和矩阵
- 最后返回打印结果

矩阵加法遵循元素对应相加的规则, 结果矩阵中每个位置 (i,j) 的值等于五个矩阵对应位置元素的总和. 运行代码后, 你将得到一个 16\*16 的矩阵, 包含所有指定次幂矩阵的累加结果.

这是十进制矩阵, 请用python计算

A=A^2+A^3+A^4

要计算矩阵A的1到4次方之和(A+A^2+A^3+A^4), 我们可以依次计算各次幂矩阵后进行元素级相加. 以下是实现代码:

计算矩阵A到A^4的和

运行时间: 10.05

代码说明:

- 首先将原始矩阵转换为 NumPy 数组以便进行高效的矩阵运算
- 依次计算矩阵的 2 到 4 次方 (A^2 到 A^4), 每次计算都基于前一次的结果
- 对 A、A^2、A^3 和 A^4 进行元素级相加, 得到总和矩阵

计算A+A^2+A^3+A^4+A^5

matrix\_sum\_1-5\_check\_zero.py

```

32 # 计算总和: A + A^2 + A^3 + A^4 + A^5
33 sum_matrix = A + A2 + A3 + A4 + A5
34
35 # 打印结果矩阵
36 print("矩阵 A + A^2 + A^3 + A^4 + A^5 的和为: ")
37 for row in sum_matrix:
38     print(row)
39
40 # 检查是否有零元素
41 has_zero = np.any(sum_matrix == 0)
42
43 if has_zero:
44     print("\n结果矩阵中存在零元素")
45     # 找出所有零元素的位置
46     zero_positions = np.where(sum_matrix == 0)
47     print("零元素的位置 (行, 列): ")
48     for i, j in zip(zero_positions[0], zero_positions[1]):
49         print(f"({i}, {j})")
50 else:
51     print("\n结果矩阵中不存在零元素")
52

```

控制台

矩阵 A + A^2 + A^3 + A^4 + A^5 的和为:

[14 26 4 11 8 11 51 26 14 26 4 11 8 11 51 26]

[26 21 48 32 11 53 13 8 26 18 55 32 11 53 13 11]

[14 48 28 13 51 13 34 48 4 55 22 13 51 13 34 48]

[11 32 13 21 26 11 48 53 11 32 13 18 26 8 55 53]

[8 11 51 26 14 26 4 11 8 11 51 26 14 26 4 11]

[11 53 13 11 26 21 48 32 11 53 13 8 26 18 55 32]

[51 13 34 48 4 48 28 13 51 13 34 55 4 55 22 13]

[26 8 55 53 11 32 13 21 26 11 48 53 11 32 13 18]

[14 26 4 11 8 11 51 26 14 26 4 11 8 11 51 26]

[26 18 55 32 11 53 13 11 26 21 48 32 11 53 13 8]

[4 55 22 13 51 13 34 48 4 48 28 13 51 13 34 55]

[11 32 13 18 26 8 55 53 11 32 13 21 26 11 48 53]

[8 11 51 26 14 26 4 11 8 11 51 26 14 26 4 11]

[11 53 13 8 26 18 55 32 11 53 13 11 26 21 48 32]

[51 13 34 55 4 55 22 13 51 13 34 48 4 48 28 13]

[26 11 48 53 11 32 13 18 26 8 55 53 11 32 13 21]

15

**问题 3.1.** 如果把矩阵  $L$  的每一行除以该行元素之和, 则所得矩阵  $P$  的每个元素都是非负数, 且每一行元素的和都是 1. 请给出这个矩阵  $P$  的一个解释, 并计算  $P^2, P^3, \dots, P^n, \dots$ , 观察计算结果, 并给出相应的解释.

仔细观察这个 16 阶的方阵  $L$ , 不难发现它可以写成如下分块形式, 其中每一块都是 4 阶方阵.

$$L = \begin{bmatrix} 0 & A & B & A^T \\ A^T & 0 & A & B \\ B & A^T & 0 & A \\ A & B & A^T & 0 \end{bmatrix}.$$

矩阵乘法的规则同样适用于分块矩阵, 例如

$$\begin{aligned} L^2 &= \begin{bmatrix} AA^T + B^2 + A^T A & BA^T + A^T B & A^2 + (A^T)^2 & AB + BA \\ AB + BA & A^T A + AA^T + B^2 & A^T B + BA^T & (A^T)^2 + A^2 \\ (A^T)^2 + A^2 & BA + AB & B^2 + A^T A + AA^T & BA^T + A^T B \\ BA^T + A^T B & A^2 + (A^T)^2 & AB + BA & AA^T + B^2 + A^T A \end{bmatrix} \\ &= \begin{bmatrix} C_1 & C_4^T & C_3 & C_4 \\ C_4 & C_1 & C_4^T & C_3 \\ C_3 & C_4 & C_1 & C_4^T \\ C_4^T & C_3 & C_4 & C_1 \end{bmatrix}, \quad \begin{cases} C_1 = AA^T + A^T A + B^2, \\ C_3 = A^2 + (A^T)^2 = A^2 + (A^2)^T, \\ C_4 = AB + BA \end{cases} \end{aligned}$$

因此只要计算 6 个 4 阶矩阵  $AA^T, A^T A, B^2, AB, BA, A^2$  即可, 这不仅减少计算量, 而且可以同时并行计算, 从而提高计算效率.

**问题 3.2.** 证明  $L^n$  有如下形式

$$L^n = \begin{bmatrix} P_n & R_n^T & Q_n & R_n \\ R_n & P_n & R_n^T & Q_n \\ Q_n & R_n & P_n & R_n^T \\ R_n^T & Q_n & R_n & P_n \end{bmatrix},$$

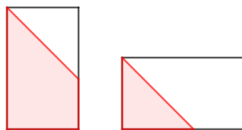
且  $P_n, Q_n, R_n$  是 4 阶方阵,  $P_n, Q_n$  是对称的. 写出它们关于  $n$  的递推关系.

### 3.3 逆矩阵, Hill 密码

**定义 3.3.** 称矩阵  $A$  可逆, 如果存在矩阵  $B_1, B_2$  使得  $AB_1$  和  $B_2A$  都是单位矩阵. 此时, 我们称  $B_1$  是  $A$  的右逆矩阵, 称  $B_2$  是  $A$  的左逆矩阵.

**问题 3.4.** 证明: 可逆矩阵的左逆矩阵和右逆矩阵是相等的, 从而可逆矩阵有唯一的左逆矩阵, 同时它也是唯一的右逆矩阵.

**注 3.5.** 但我们仍不知道可逆矩阵是否为方阵, 这将在后面解决 (问题 10.4).





**问题 3.6.** 下三角矩阵是形如上图那样的矩阵，对角线右侧的元素都为零 ( $a_{ij} = 0, \forall j > i$ )。证明：

1. 下三角矩阵的乘积还是下三角矩阵；
2. 对下三角矩阵  $A$ ， $A$  可逆当且仅当  $A$  是方阵且对角线上的元素都非零；
3. 若下三角矩阵  $A$  可逆，则  $A$  有唯一的左逆矩阵和右逆矩阵，且左右逆矩阵相等。

类似可以定义上三角矩阵，并对上三角矩阵讨论相应的结论。

**问题 3.7.** 证明排列阵（定义见 2.2）都可逆，且有唯一的左逆矩阵和右逆矩阵，且左右逆矩阵相等。

**问题 3.8.** 构造一个矩阵，使它只有左逆矩阵，而没有右逆矩阵。

**例 3.9 (Hill 密码).** 希尔密码由 Lester S. Hill 于 1929 年发明，是一种基于线性代数的测谎替换密码。希尔使用矩阵和矩阵乘法来混合明文。我们约定英文字母和常见标点按以下表格对应于数字，空格对应于数字 0。

表 1: 字母与数字对应表

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	2	3	4	5	6	7	8	9	10	11	12	13	14
O	P	Q	R	S	T	U	V	W	X	Y	Z	,	.
15	16	17	18	19	20	21	22	23	24	25	26	27	28

哈姆雷特的那句著名台词 “To be, or not to be, that is the question” 写成

$$\begin{bmatrix} T & , & O & - & H & S & - & T \\ O & O & T & B & A & - & Q & I \\ - & R & - & E & T & T & U & O \\ B & - & T & , & - & H & E & N \\ E & N & O & T & I & E & S & . \end{bmatrix}$$

其编码数字矩阵为

$$P = \begin{bmatrix} 20 & 27 & 15 & 0 & 8 & 19 & 0 & 20 \\ 15 & 15 & 20 & 2 & 1 & 0 & 17 & 9 \\ 0 & 18 & 0 & 5 & 20 & 20 & 21 & 15 \\ 2 & 0 & 20 & 27 & 0 & 8 & 5 & 14 \\ 5 & 14 & 15 & 20 & 9 & 5 & 19 & 28 \end{bmatrix}$$

加密和解密的关键是用以下矩阵  $K$  和它的（左）逆矩阵  $K^{-1}$

$$K = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 \\ 3 & 5 & 4 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 4 & 6 & 5 & 3 & 2 \end{bmatrix}, \quad K^{-1} = \begin{bmatrix} 2 & 1 & -2 & 0 & 1 \\ -2 & -1 & 1 & -1 & 0 \\ 1 & 0 & 1 & 1 & -1 \\ 1 & 2 & -1 & 1 & 0 \\ -2 & -2 & 0 & -1 & 1 \end{bmatrix}.$$

加密时，用  $K$  乘以矩阵  $P$  得到

$$KP = \begin{bmatrix} 35 & 60 & 35 & 7 & 29 & 39 & 38 & 44 \\ 52 & 75 & 75 & 36 & 30 & 47 & 60 & 67 \\ 144 & 242 & 200 & 104 & 118 & 158 & 198 & 221 \\ 7 & 32 & 35 & 52 & 29 & 33 & 45 & 57 \\ 186 & 316 & 270 & 158 & 156 & 210 & 260 & 307 \end{bmatrix}$$

这个矩阵元素除以 29 所得余数构成加密信息矩阵

$$Q = KP \pmod{29} = \begin{bmatrix} 6 & 2 & 6 & 7 & 0 & 10 & 9 & 15 \\ 23 & 17 & 17 & 7 & 1 & 18 & 2 & 9 \\ 28 & 10 & 26 & 17 & 2 & 13 & 24 & 18 \\ 7 & 3 & 6 & 23 & 0 & 4 & 16 & 28 \\ 12 & 26 & 9 & 13 & 11 & 7 & 28 & 17 \end{bmatrix}$$

它对应的密码矩阵为

$$\tilde{Q} = \begin{bmatrix} F & B & F & G & - & J & I & O \\ W & Q & Q & G & A & R & B & I \\ . & J & Z & Q & B & M & X & R \\ G & C & F & W & - & D & P & . \\ L & Z & I & M & K & G & . & Q \end{bmatrix}$$

解密时，把密码矩阵  $\tilde{Q}$  对应为数字矩阵  $Q$ ，计算  $K^{-1}Q$ ，并把所得矩阵的元素除以 29 计算余数，就可以还原为原文数字矩阵，再对应回字符就得到原文。

**问题 3.10.** 验证上例中的  $K^{-1}$  既是  $K$  的左逆矩阵，也是  $K$  的右逆矩阵。

**问题 3.11.** 请分析上例中对 29 取余数的做法，解释为什么能恢复原文。

## 4 线性方程组与线性子空间

### 4.1 理解线性方程组的三个视角

线性方程组

$$\begin{cases} \alpha_{11}x_1 + \cdots + \alpha_{1n}x_n = b_1, \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_n. \end{cases} \quad A\mathbf{x} = \mathbf{b}$$

例如

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

- 映射观点：给定映射的像，求原像

$$\begin{bmatrix} 1 & -2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

- 行观点：多条直线求交点

- 列观点：向量分解

$$x \begin{bmatrix} 1 \\ 3 \end{bmatrix} + y \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

再如，

$$\begin{cases} x + 2y + 3z = 6 \\ 2x + 5y + 2z = 4 \\ 6x - 3y + z = 2 \end{cases}$$

- 映射观点：给定映射的像，求原像

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

- 行观点：多个平面求交点

- 列观点：向量分解

$$x \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \\ -3 \end{bmatrix} + z \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}$$

又如,

$$\begin{cases} x + 2y + 3z = 6 \\ 2x + 5y + 2z = 4 \end{cases}$$

- 映射观点: 给定映射的像, 求原像

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

- 行观点: 多个平面求交集
- 列观点: 向量分解

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \end{bmatrix} + z \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

总结:

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

- 映射观点: 给定映射的像, 求原像
- 行观点: 多个平面求交点
- 列观点: 把  $\mathbf{b}$  表示为  $A$  的列向量的线性组合

## 4.2 $A\mathbf{x} = \mathbf{b}$ 的解的情况 (“0, 1, $\infty$ ”)

### 基本问题

$A\mathbf{x} = \mathbf{b}$  是否有解? 何时有解, 何时无解?

当  $A\mathbf{x} = \mathbf{b}$  有解时, 有多少解? 是否有唯一解?

当  $A\mathbf{x} = \mathbf{b}$  有多个解时, 解之间有什么关系? 能否用最少个数的解表示所有的解?

### 基本观察

1) 若  $\mathbf{x}_1, \mathbf{x}_2$  都是  $A\mathbf{x} = \mathbf{b}$  (非齐次方程) 的解, 则  $\mathbf{x}_1 - \mathbf{x}_2$  是  $A\mathbf{x} = \mathbf{0}$  (齐次方程) 的解.

2) 若  $\mathbf{x}_1$  是  $A\mathbf{x} = \mathbf{b}$  的解,  $\mathbf{x}_0$  是  $A\mathbf{x} = \mathbf{0}$  的解, 则  $\mathbf{x}_1 + \mathbf{x}_0$  是  $A\mathbf{x} = \mathbf{b}$  的解.

3) 设  $\mathbf{x}_1$  是  $A\mathbf{x} = \mathbf{b}$  的一个解 (特解). 则  $\mathbf{x}$  是  $A\mathbf{x} = \mathbf{b}$  的解当且仅当  $\mathbf{x} - \mathbf{x}_1$  是  $A\mathbf{x} = \mathbf{0}$  的解.

$\text{“}A\mathbf{x} = \mathbf{b} \text{ 的通解”}$	$=$	$\text{“}A\mathbf{x} = \mathbf{b} \text{ 的特解”}$	$+$	$\text{“}A\mathbf{x} = \mathbf{0} \text{ 的通解”}$
非齐次方程		非齐次方程		齐次方程

4) 若  $A\mathbf{x} = \mathbf{b}$  有解, 则  $A\mathbf{x} = \mathbf{b}$  有唯一解当且仅当  $A\mathbf{x} = \mathbf{0}$  只有零解 (唯一解).

若  $A\mathbf{x} = \mathbf{b}$  有解,  $A\mathbf{x} = \mathbf{0}$  有非零解, 则  $A\mathbf{x} = \mathbf{b}$  有无穷多解.

$$A\mathbf{x}_1 = \mathbf{b}, \quad A\mathbf{x}_0 = \mathbf{b}(\mathbf{x}_0 \neq \mathbf{0}) \implies A(\mathbf{x}_1 + t\mathbf{x}_0) = \mathbf{b}, \quad \forall t \in \mathbb{R}.$$

$t \neq 0 \implies t\mathbf{x}_0 \neq \mathbf{0}$  (否则  $\mathbf{x}_0 = \frac{1}{t}t\mathbf{x}_0 = \mathbf{0}$ ).

总结

求解线性方程组  $A\mathbf{x} = \mathbf{b}$  归结为以下三个问题:

1. 如何判断  $A\mathbf{x} = \mathbf{b}$  有解?
2. 当  $A\mathbf{x} = \mathbf{b}$  有解时, 如何求出一个解?
3. 如何求  $A\mathbf{x} = \mathbf{0}$  的所有解?

## 5 线性表示, 线性子空间, 线性映射的值域空间和零空间

定义 5.1. 称  $\mathbf{b}$  可由  $\alpha_1, \dots, \alpha_k$  线性表示, 若存在数  $x_1, \dots, x_k$  使

$$x_1\alpha_1 + \dots + x_k\alpha_k = \mathbf{b},$$

即

$$\begin{bmatrix} \alpha_1 & \dots & \alpha_k \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} = \mathbf{b}.$$

记

$$\begin{aligned} \text{span}\{\alpha_1, \dots, \alpha_k\} &= \{\mathbf{b} \mid \mathbf{b} \text{ 可由 } \alpha_1, \dots, \alpha_k \text{ 线性表示}\} \\ &= \{x_1\alpha_1 + \dots + x_k\alpha_k \mid x_1, \dots, x_k \in \mathbb{R}\} \end{aligned}$$

并称之为由  $\alpha_1, \dots, \alpha_k$  张成的集合.

定义 5.2. 若  $\alpha_1, \dots, \alpha_n$  是  $m$  行  $n$  列矩阵  $A$  的所有列向量, 则

$$\text{span}(A) = \text{span}\{\alpha_1, \dots, \alpha_n\} = \text{Range}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$$

称为  $A$  的值域空间或列空间.

记  $\text{Ker } A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$ , 称为  $A$  的零空间.

注 5.3.  $A\mathbf{x} = \mathbf{b}$  有解当且仅当  $\mathbf{b} \in \text{Range}(A)$ .

当  $A\mathbf{x} = \mathbf{b}$  有解时,  $A\mathbf{x} = \mathbf{b}$  有唯一解当且仅当  $\text{Ker}(A) = \{\mathbf{0}\}$ .

**定义 5.4.** 称  $V \subseteq \mathbb{R}^n$  是一个线性子空间, 若  $V \neq \emptyset$  且  $V$  对线性组合封闭, 即  $\forall \alpha, \beta \in V, \forall \lambda, \mu \in \mathbb{R}, \lambda\alpha + \mu\beta \in V$ .

**注 5.5.** 当  $V \neq \emptyset$  时, 取  $\alpha_0 \in V$ , 则  $\mathbf{0} = 0 \cdot \alpha_0 \in V$ . 因此上述定义中条件  $V \neq \emptyset$  可替换为  $\mathbf{0} \in V$ .

**定理 5.6.** (1) 设  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  是线性映射. 则  $\text{Ker}(A) \subseteq \mathbb{R}^n, \text{Range}(A) \subseteq \mathbb{R}^m$  都是线性子空间.

(2) 如果  $\alpha_1, \dots, \alpha_p$  和  $\beta_1, \dots, \beta_q$  满足每个  $\alpha_i \in \text{span}\{\beta_1, \dots, \beta_q\}$ , 则

$$\text{span}\{\alpha_1, \dots, \alpha_p\} \subseteq \text{span}\{\beta_1, \dots, \beta_q\},$$

存在  $q$  行  $p$  列的矩阵  $(x_{ij})$  使得

$$(\alpha_1, \dots, \alpha_p) = (\beta_1, \dots, \beta_q) \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \cdots & \vdots \\ x_{q1} & \cdots & x_{qp} \end{bmatrix}.$$

证明留作练习.

## 6 线性相关, 线性无关, 空间的基

设  $V \subseteq \mathbb{R}^n$  是线性子空间.

在  $V$  中寻找尽可能少的向量, 使得  $V$  中任何向量可由它们线性表示.

**定义 6.1.** 称  $\alpha_1, \dots, \alpha_k \in V$  线性相关, 若存在不全为零的数  $x_1, \dots, x_k$ , 使得  $x_1\alpha_1 + \dots + x_k\alpha_k = \mathbf{0}$ . 即存在  $\alpha_i$  可由其他  $\alpha_j$  们线性表示:

$$x_i \neq 0 \implies \alpha_i = -\frac{x_1}{x_i}\alpha_1 - \dots - \frac{x_{i-1}}{x_i}\alpha_{i-1} - \frac{x_{i+1}}{x_i}\alpha_{i+1} - \dots - \frac{x_k}{x_i}\alpha_k.$$

称  $\alpha_1, \dots, \alpha_k \in V$  线性无关, 若  $\alpha_1, \dots, \alpha_k$  不是线性相关的.

**注 6.2.** (1)

$\alpha_1, \dots, \alpha_k$  线性无关  $\iff \forall \alpha_i$  都无法由其他那些  $\alpha_j$  们线性表示

$$\iff x_1\alpha_1 + \dots + x_k\alpha_k = \mathbf{0} \text{ 只有零解}$$

(2) 如果  $\alpha_1, \dots, \alpha_k$  线性无关, 但  $\alpha_1, \dots, \alpha_k, \beta$  线性相关, 则  $\beta$  可由  $\alpha_1, \dots, \alpha_k$  线性表示.

(3) 当我们说一组向量线性相关时, 允许一个向量在这组向量中重复出现. 但当一组向量线性无关时, 这组向量中的任何一个向量只能在这组向量中出现一次.

**定义 6.3.** 称  $\alpha_1, \dots, \alpha_n$  是  $V$  的一组基, 若它们线性无关, 但  $V$  中任何向量都可由  $\alpha_1, \dots, \alpha_n$  线性表示.

注 6.4.

$$\begin{aligned}\alpha_1, \dots, \alpha_n \text{ 是 } V \text{ 的基} &\iff \forall \mathbf{b} \in V, \quad x_1\alpha_1 + \dots + x_n\alpha_n = \mathbf{b} \text{ 有唯一解} \\ &\iff \forall \mathbf{b} \in V, \quad \alpha_1, \dots, \alpha_n, \mathbf{b} \text{ 线性相关}\end{aligned}$$

例 6.5. (1)  $V$  中任何非零向量  $\alpha$  都线性无关.

假设存在非零向量  $\alpha$  线性相关. 则存在非零数  $x$  使得  $x\alpha = \mathbf{0}$ . 于是  $\alpha = \frac{1}{x} \cdot x\alpha = \mathbf{0}$ . 这与  $\alpha$  非零矛盾. 所以任何非零向量都是线性无关的.

(2)  $e_1, \dots, e_n$  是  $\mathbb{R}^n$  的一组基.

(3)  $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}$  是  $\mathbb{R}^2$  的一组基.

由

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

解得唯一解  $x = \frac{4b_1 - b_2}{2}, y = \frac{b_2 - 2b_1}{2}$ .

## 7 基的存在性, 空间维数, 置换引理

定理 7.1. 对  $\mathbb{R}^m$  的任何非零线性子空间  $V$ ,

1. 都有一组基;
2. 任何一组基都含有相同个数的向量;  $V$  的维数  $\dim V =$  一组基中的向量个数  $n$
3.  $V$  中任何一组多于  $n$  的向量都是线性相关的;
4. 若  $\dim V = n, e_1, \dots, e_n \in V$  线性无关, 则它们是  $V$  的一组基

注 7.2. 上述定理的结论 1, 即子空间的基的存在性, 是个重要的问题, 但绝大多数教材似乎未给与足够重视, 或者处理的不令人满意.

引理 7.3 (置换引理, Steinitz 1913). 设  $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$  是  $V$  中两组线性无关的向量, 满足每个  $\alpha_i$  都可由  $\beta_1, \dots, \beta_q$  线性表示. 则  $p \leq q$ , 且存在  $\beta_1, \dots, \beta_q$  的一个排列  $\beta_{\sigma(1)}, \dots, \beta_{\sigma(q)}$  使得  $\forall 1 \leq j \leq p$ ,

$$\alpha_1, \dots, \alpha_j, \beta_{\sigma(j+1)}, \dots, \beta_{\sigma(q)} \quad (*)$$

线性无关, 且任何  $\beta_i$  可由 (\*) 中的向量线性表示.

定理 7.1 的证明. (1) 取  $\mathbb{R}^m$  的标准基  $e_1, \dots, e_m$ . 由引理知  $V$  中任何一组线性无关向量的个数必然不超过  $m$ .

因为  $V \neq \{\mathbf{0}\}$ , 所以存在非零向量  $\alpha_1 \in V$ .

于是由引理知存在  $e_1, \dots, e_m$  中的  $m-1$  个向量  $\beta_2, \dots, \beta_m$  使得

$$\alpha_1, \beta_2, \dots, \beta_m$$

是  $\mathbb{R}^m$  的一组基.

如果  $\alpha_1$  不是  $V$  的一组基, 则  $V$  中存在向量  $\alpha_2$  不能用  $\alpha_1$  线性表示, 则  $\alpha_1, \alpha_2$  线性无关, 且存在  $\beta_2, \dots, \beta_m$  中的  $m-2$  个向量  $\gamma_3, \dots, \gamma_m$ , 使得

$$\alpha_1, \alpha_2, \gamma_3, \dots, \gamma_m$$

是  $\mathbb{R}^m$  的一组基.

如果  $\alpha_1, \alpha_2$  不是  $V$  的一组基, 则  $V$  中存在向量  $\alpha_3$  不能用  $\alpha_1, \alpha_2$  线性表示. 此时  $\alpha_1, \alpha_2, \alpha_3$  线性无关.

如此这样, 存在  $n \leq m$  以及  $V$  中  $n$  个线性无关的向量  $\alpha_1, \alpha_2, \dots, \alpha_n$ , 使得  $V$  中任何向量都与  $\alpha_1, \alpha_2, \dots, \alpha_n$  线性相关, 从而可由  $\alpha_1, \alpha_2, \dots, \alpha_n$  线性表示, 从而  $\alpha_1, \alpha_2, \dots, \alpha_n$  是  $V$  的一组基.

(2) 由基的存在性以及引理结论可知  $V$  的任何两组基都含有相同个数的向量.

(3) 设  $\alpha_1, \alpha_2, \dots, \alpha_n$  是  $V$  的一组基,  $\delta_1, \dots, \delta_p$  是  $V$  中一组线性无关向量, 则根据引理知  $p \leq n$ .

(4) 设  $\epsilon_1, \dots, \epsilon_n \in V$  线性无关,  $\alpha_1, \alpha_2, \dots, \alpha_n$  是  $V$  的一组基. 则根据引理, 当把  $\alpha_1, \alpha_2, \dots, \alpha_n$  全部置换为  $\epsilon_1, \dots, \epsilon_n$  时,  $\alpha_1, \alpha_2, \dots, \alpha_n$  可由后者线性表示, 从而  $V$  中任何向量可由  $\epsilon_1, \dots, \epsilon_n$  线性表示, 因此  $\epsilon_1, \dots, \epsilon_n$  是  $V$  的一组基.  $\square$

**推论 7.4.** 设  $V \subseteq W$  是  $\mathbb{R}^m$  的两个线性子空间. 则

1.  $\dim V \leq \dim W$ ;
2. 若  $\dim V < \dim W$ , 则  $V$  的任何一组基可以扩为  $W$  的一组基;
3. 若  $\dim V = \dim W$ , 则  $V = W$ .

**证明.** 设  $\alpha_1, \alpha_2, \dots, \alpha_p$  是  $V$  的一组基,  $\beta_1, \dots, \beta_q$  是  $W$  的一组基.

(1) 因为  $V \subseteq W$ , 所以  $\alpha_1, \alpha_2, \dots, \alpha_p$  可由  $\beta_1, \dots, \beta_q$  线性表示, 所以根据引理结论,  $p \leq q$ , 从而  $\dim V \leq \dim W$ .

(2) 由引理知, 用  $\alpha_1, \alpha_2, \dots, \alpha_p$  替换  $\beta_1, \dots, \beta_q$  中的  $p$  个向量可得  $W$  中  $q$  个线性无关的向量  $\alpha_1, \alpha_2, \dots, \alpha_p, \tilde{\beta}_{q-p}, \dots, \tilde{\beta}_q$ . 由本节定理知, 后面这组向量是  $W$  的一组基.

(3) 若  $\dim V = \dim W$ , 则任取  $V$  的任何一组基用  $\alpha_1, \alpha_2, \dots, \alpha_n$ , 由引理知它们是  $W$  的一组基. 所以

$$V = W = \text{span}\{\alpha_1, \alpha_2, \dots, \alpha_n\}.$$

$\square$



定义 7.5. 规定  $\dim\{\mathbf{0}\} = 0$ .

注 7.6. 上述推论中的结论对  $V = \{\mathbf{0}\}$  或  $W = \{\mathbf{0}\}$  都成立.

问题 7.7. 对非零向量  $\mathbf{v} \in V$  以及线性映射  $A: V \rightarrow V$ , 记

$$\text{Cyc}_A(\mathbf{v}) = \text{span}\{\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots\},$$

称为由  $\mathbf{v}$  生成的  $A$ -循环子空间. 证明:

1. 存在正整数  $k$  使得  $\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots, A^{k-1}\mathbf{v}$  线性无关, 且  $\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots, A^k\mathbf{v}$  线性相关.
2. 对 (1) 中的  $k$ ,  $\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots, A^{k-1}\mathbf{v}$  是  $\text{Cyc}_A(\mathbf{v})$  的一组基.

## 8 置换引理的证明

因为  $\alpha_1$  可由  $\beta_1, \dots, \beta_q$  线性表示, 所以存在数  $b_1, \dots, b_q$  使得

$$\alpha_1 = b_1\beta_1 + \dots + b_q\beta_q.$$

因为  $\alpha_1, \dots, \alpha_p$  线性无关, 所以  $\alpha_1 \neq 0$ , 从而  $b_1, \dots, b_q$  不全为零. 因此存在  $1, 2, \dots, q$  的排列  $\tau$  使得

$$\alpha_1 = b_{\tau(1)}\beta_{\tau(1)} + \dots + b_{\tau(q)}\beta_{\tau(q)}, \quad b_{\tau(1)} \neq 0.$$

因此

$$(\alpha_1, \beta_{\tau(2)}, \dots, \beta_{\tau(q)}) = (\beta_{\tau(1)}, \beta_{\tau(2)}, \dots, \beta_{\tau(q)}) \begin{bmatrix} b_{\tau(1)} & & & \\ b_{\tau(2)} & 1 & & \\ \vdots & & \ddots & \\ b_{\tau(q)} & & & 1 \end{bmatrix} \quad (1)$$

于是

$$(\beta_{\tau(1)}, \beta_{\tau(2)}, \dots, \beta_{\tau(q)}) = (\alpha_1, \beta_{\tau(2)}, \dots, \beta_{\tau(q)}) \begin{bmatrix} \frac{1}{b_{\tau(1)}} & & & \\ -\frac{b_{\tau(2)}}{b_{\tau(1)}} & 1 & & \\ \vdots & & \ddots & \\ -\frac{b_{\tau(q)}}{b_{\tau(1)}} & & & 1 \end{bmatrix}$$

故  $\beta_1, \dots, \beta_q$  可由

$$\alpha_1, \beta_{\tau(2)}, \dots, \beta_{\tau(q)} \quad (2)$$

线性表示. 下证 (2) 中向量线性无关.

假设它们线性相关. 则存在不全为零的数  $y_1, \dots, y_q$  使得

$$y_1 \alpha_1 + y_2 \beta_{\tau(2)} + \dots + y_q \beta_{\tau(q)} = \mathbf{0}.$$

因为  $\beta_{\tau(2)}, \dots, \beta_{\tau(q)}$  线性无关, 所以  $y_1 \neq 0$ . 这表明  $\alpha_1$  可由  $\beta_{\tau(2)}, \dots, \beta_{\tau(q)}$  线性表示.

因此由 (4) 知  $\beta_{\tau(1)}$  可由  $\beta_{\tau(2)}, \dots, \beta_{\tau(q)}$  线性表示. 但这与向量  $\beta_1, \dots, \beta_q$  线性无关矛盾. 因此 (2) 中的向量线性无关.

至此,  $j = 1$  时的结论得证.

下面用数学归纳法完成后续证明.

假设引理结论对某个  $j < \min\{p, q\}$  成立. 此时,

$$\alpha_1, \dots, \alpha_j, \beta_{\varphi(j+1)}, \dots, \beta_{\varphi(q)} \quad (3)$$

线性无关, 且  $\beta_1, \dots, \beta_q$  可由 (3) 中向量线性表示, 其中  $\varphi$  是  $1, 2, \dots, q$  的一个排列.

此时,  $\alpha_{j+1}$  可由  $\beta_1, \dots, \beta_q$  线性表示, 从而可由 (3) 中向量线性表示, 从而存在数  $z_{j+1}, \dots, z_q$  使得

$$\alpha_{j+1} = z_1 \alpha_1 + \dots + z_j \alpha_j + z_{j+1} \beta_{\varphi(j+1)} + \dots + z_q \beta_{\varphi(q)}.$$

因为这里的  $\alpha_i$  们线性无关, 所以  $z_{j+1}, \dots, z_q$  不全为零.

于是存在  $1, 2, \dots, q$  的排列  $\psi$  使得  $\psi(1) = \varphi(1), \dots, \psi(j) = \varphi(j)$ , 且  $z_{\varphi^{-1}(\psi(j+1))} \neq 0$ , 因此  $\beta_{\psi(j+1)}, \dots, \beta_{\psi(q)}$  是  $\beta_{\varphi(j+1)}, \dots, \beta_{\varphi(q)}$  的一个排列. 记  $\tilde{z}_i = z_{\varphi^{-1}(\psi(i))}$ , 则

$$(\alpha_1, \dots, \alpha_j, \alpha_{j+1}, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}) =$$

$$(\alpha_1, \dots, \alpha_j, \beta_{\psi(j+1)}, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}) \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \tilde{z}_1 & & \\ & & \vdots & & \\ & & & \tilde{z}_{j+1} & \\ & & & \vdots & \ddots \\ & & & \tilde{z}_q & & 1 \end{bmatrix} \quad (4)$$

因此

$$(\alpha_1, \dots, \alpha_j, \beta_{\psi(j+1)}, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}) =$$

$$(\alpha_1, \dots, \alpha_j, \alpha_{j+1}, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}) \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & -\frac{\tilde{z}_1}{\tilde{z}_{j+1}} & & \\ & & \vdots & & \\ & & \frac{1}{\tilde{z}_{j+1}} & & \\ & & \vdots & \ddots & \\ & & -\frac{\tilde{z}_q}{\tilde{z}_{j+1}} & & 1 \end{bmatrix}$$

故  $\alpha_1, \dots, \alpha_j, \beta_{\psi(j+1)}, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}$  可由

$$\alpha_1, \dots, \alpha_j, \alpha_{j+1}, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)} \quad (5)$$

线性表示. 因此向量组 (3) 可由向量组 (5) 线性表示.

下面证明向量组 (5) 线性无关.

假设向量组 (5) 线性相关. 因为  $\alpha_1, \dots, \alpha_j, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}$  线性无关, 所以  $\alpha_{j+1}$  可由  $\alpha_1, \dots, \alpha_j, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}$  线性表示. 所以向量组 (3) 可由向量组  $\alpha_1, \dots, \alpha_j, \beta_{\psi(j+2)}, \dots, \beta_{\psi(q)}$  线性表示, 特别是向量组 (3) 中的向量  $\beta_{\psi(j+1)}$  可由该向量组其余向量线性表示, 但这与向量组 (3) 线性无关矛盾. 所以向量组 (5) 线性无关.

因此引理结论对  $j+1$  也成立.

这样由数学归纳法知引理对  $j = \min\{p, q\}$  成立.

假设  $q < p$ . 则上述结论表明  $\alpha_1, \dots, \alpha_q$  可以线性表示  $\beta_1, \dots, \beta_q$ , 从而可以线性表示  $\alpha_p$ , 这与  $\alpha_1, \dots, \alpha_q, \dots, \alpha_p$  线性无关矛盾.

所以  $p \leq q$ .

这样就完成了置换引理的证明.

## 9 基与坐标, 线性映射和矩阵

设  $V$  是  $n$  维线性空间,  $\alpha_1, \dots, \alpha_n$  是  $V$  的一组基.

对任意  $v \in V$  存在唯一一组数  $x_1, \dots, x_n$  使得

$$v = x_1 \alpha_1 + \dots + x_n \alpha_n = (\alpha_1, \dots, \alpha_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

称  $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$  为  $v$  在基  $\alpha_1, \dots, \alpha_n$  下的坐标形式, 称  $x_1, \dots, x_n$  是  $v$  在基  $\alpha_1, \dots, \alpha_n$  下的坐标.

设  $V, W$  分别是  $n$  维和  $m$  维线性空间,  $A: V \rightarrow W$  是线性映射.

设  $\alpha_1, \dots, \alpha_n$  是  $V$  的一组基,  $\beta_1, \dots, \beta_m$  是  $W$  的一组基. 则

$$\begin{aligned} A\mathbf{v} &= A(x_1\alpha_1 + \dots + x_n\alpha_n) = x_1A\alpha_1 + \dots + x_nA\alpha_n \\ &= (A\alpha_1, \dots, A\alpha_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= (\beta_1, \dots, \beta_m) \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

其中矩阵  $(a_{ij})_{m \times n}$  的第  $j$  列是  $A\alpha_j$  在基  $\beta_1, \dots, \beta_m$  下的坐标.

因此我们看到矩阵是线性映射在给定基下的坐标形式. 同一个线性映射在不同的基下会有不同的矩阵表达.

为了更好地研究线性映射  $A$  的性质, 我们往往需要给空间找到与  $A$  相适合的“好的”基.

**例 9.1.** 为了计算矩阵  $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$  的幂  $A^n$ , 我们取  $\mathbb{R}^2$  中两个向量

$$\alpha_1 = \begin{bmatrix} 1 \\ \lambda^* \end{bmatrix}, \quad \alpha_2 = \begin{bmatrix} -\lambda^* \\ 1 \end{bmatrix}, \quad \left( \lambda^* = \frac{1 + \sqrt{5}}{2} \right).$$

对任意  $\beta = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \in \mathbb{R}^2$ , 由方程组

$$x\alpha_1 + y\alpha_2 = \beta = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

解得

$$x = \frac{b_1 + \lambda^* b_2}{1 + \lambda^{*2}}, \quad y = \frac{b_2 - \lambda^* b_1}{1 + \lambda^{*2}}$$

是唯一解, 所以  $\alpha_1, \alpha_2$  是  $\mathbb{R}^2$  的一组基.

$$(A\alpha_1, A\alpha_2) = (\lambda^*\alpha_1, -\frac{1}{\lambda^*}\alpha_2) = (\alpha_1, \alpha_2) \begin{bmatrix} \lambda^* & 0 \\ 0 & \frac{-1}{\lambda^*} \end{bmatrix}$$

因此

$$(A^n\alpha_1, A^n\alpha_2) = (\alpha_1, \alpha_2) \begin{bmatrix} \lambda^{*n} & 0 \\ 0 & (\frac{-1}{\lambda^*})^n \end{bmatrix}, \quad \forall n \geq 1.$$

对研究  $A^n$  而言,  $\alpha_1, \alpha_2$  是比  $\mathbb{R}^2$  的标准基更好的基, 在这个基下,  $A^n$  的表示矩阵是对角矩阵, 这一形式更便于计算矩阵乘法 (映射的迭代).

## 10 维数与秩，线性同构与逆矩阵

**定义 10.1.** 对向量空间  $V$  中的一组向量  $\alpha_1, \dots, \alpha_k$ ，称它们张成的线性子空间  $\text{span}\{\alpha_1, \dots, \alpha_k\}$  的维数为这组向量的**秩**，记为  $\text{rank}(\alpha_1, \dots, \alpha_k)$ . 矩阵的列空间的维数称为矩阵的**列秩**，记为  $\text{rank}(A)$ . 矩阵的行向量张成的线性空间的维数称为矩阵的**行秩**.

**定理 10.2 (维数定理).** 设  $V, W$  是线性空间， $n = \dim V$ . 则对任何线性映射  $A: V \rightarrow W$  都有

$$\dim \text{Ker}(A) + \dim \text{Range}(A) = n.$$

**证明.** 取  $\text{Ker}(A)$  的一组基  $\alpha_1, \dots, \alpha_k$ ，并把它扩充为  $V$  的一组基

$$\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n.$$

则对任意  $v \in V$ ,

$$v = x_1\alpha_1 + \dots + x_k\alpha_k + x_{k+1}\alpha_{k+1} + \dots + x_n\alpha_n,$$

从而

$$Av = x_{k+1}A\alpha_{k+1} + \dots + x_nA\alpha_n,$$

因此

$$\text{Range}(A) = \text{span}\{A\alpha_{k+1}, \dots, A\alpha_n\}.$$

若存在数  $y_{k+1}, \dots, y_n$  使得

$$y_{k+1}A\alpha_{k+1} + \dots + y_nA\alpha_n = 0,$$

则

$$y_{k+1}\alpha_{k+1} + \dots + y_n\alpha_n \in \text{Ker}(A),$$

从而存在数  $y_1, \dots, y_k$  使得

$$y_{k+1}\alpha_{k+1} + \dots + y_n\alpha_n = y_1\alpha_1 + \dots + y_k\alpha_k.$$

由于向量组  $\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_n$  线性无关，所以所有  $y_i = 0$ ，从而向量组  $A\alpha_{k+1}, \dots, A\alpha_n$  线性无关. 因此  $A\alpha_{k+1}, \dots, A\alpha_n$  是  $\text{Range}(A)$  的基. 从而

$$\dim \text{Ker}(A) + \dim \text{Range}(A) = n.$$

□

**推论 10.3.** 设  $V, W$  是有限维线性空间， $A: V \rightarrow W$  是线性映射. 则

1. 若  $A$  是满射，则  $\dim W \leq \dim V$ ;

2. 若  $A$  是单射, 则  $\dim V \leq \dim W$ ;
3. 若  $A$  是双射 (一一对应), 则  $\dim V = \dim W$ ;
4. 当  $\dim V = \dim W$  时,

$A$  是双射  $\iff A$  是单射  $\iff A$  是满射.

**证明.** (1) 若  $A$  是单射, 则  $\text{Ker}(A) = \{\mathbf{0}\}$ , 于是  $\dim \text{Ker}(A) = 0$ , 从而

$$\dim V = \dim \text{Range}(A) \leq \dim W.$$

(2) 若  $A$  是满射, 则  $W = \text{Range}(A)$ , 从而

$$\dim W = \dim \text{Range}(A) \leq \dim \text{Ker}(A) + \dim \text{Range}(A) = \dim V.$$

(3) 是 (1) 和 (2) 的推论.

(4) 很明显,  $A$  是双射  $\implies A$  是单射.

当  $\dim V = \dim W$  时, 若  $A$  是单射, 则  $\text{Ker}(A) = \{\mathbf{0}\}$ , 于是  $\dim \text{Ker}(A) = 0$ , 从而

$$\dim \text{Range}(A) = \dim V = \dim W.$$

又  $\text{Range}(A) \subseteq W$ , 所以由第10节定理推论知  $\text{Range}(A) = W$ , 从而  $A$  是双射.

类似可证  $A$  是双射  $\iff A$  是满射.  $\square$

**问题 10.4.** (1) 设矩阵  $A$  有左逆矩阵 (见定义 3.3). 证明  $A$  的列满秩, 即列数等于列秩; 且  $A$  的列数不超过行数.

(2) 设矩阵  $A$  有右逆矩阵 (见定义 3.3). 证明  $A$  的行满秩, 即行数等于行秩; 且  $A$  的行数不超过列数.

(3) 设矩阵  $A$  是可逆矩阵 (见定义 3.3). 证明  $A$  是方阵, 它的左逆矩阵与右逆矩阵相同, 而且是唯一的, 记为  $A^{-1}$ .

**定义 10.5.** 如果线性映射  $A: V \rightarrow W$  是一一对应, 则它的逆映射  $A^{-1}: W \rightarrow V$  也是线性映射. 此时称  $A$  为**线性同构**.

**问题 10.6.** 证明线性同构在给定基下的矩阵是可逆矩阵.

## 11 矩阵的列变换与行变换

回到线性方程组的讨论

讨论  $A\mathbf{x} = \mathbf{b}$  的解的情况, 其中

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

解.  $Ax = b$  是否有解, 决定于  $b$  是否在  $\text{Range}(A)$  (即  $A$  的列空间) 中.

为此, 我们给  $\text{Range}(A)$  找一组形式简单的基:

$$A = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$$

则

$$\text{Range}(A) = \text{span}\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$$

把矩阵  $A$  的第一列  $\alpha_1$  的适当倍数加到其他各列, 可使后面各列第一行元素变成零.

$$\begin{aligned} & (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) \\ &= (\alpha_1, \alpha_2 - 2\alpha_1, \alpha_3 - 2\alpha_1, \alpha_4 + 5\alpha_1, \alpha_5 - 6\alpha_1) \\ &= (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) \begin{bmatrix} 1 & -2 & -2 & 5 & -6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ & \quad \quad \quad C_1 \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & -4 & 5 \\ 4 & 0 & -3 & 12 & -15 \\ 3 & 0 & -5 & 20 & -25 \end{bmatrix} \end{aligned}$$

矩阵  $C_1$  记录了这些对列向量的操作, 它是可逆矩阵. 因此

$$\text{Range}(A) = \text{span}\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$$

把矩阵  $B$  的第三列  $\beta_3$  的适当倍数加到其他各列, 可使其他各列第二行元素变成零.

$$\begin{aligned} & (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) \\ &= (\beta_1 + \beta_3, \beta_2, \beta_3, \beta_4 + 4\beta_3, \beta_5 - 5\beta_3) \\ &= (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 4 & -5 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ & \quad \quad \quad C_2 \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -3 & 0 & 0 \\ -2 & 0 & -5 & 0 & 0 \end{bmatrix} \end{aligned}$$

矩阵  $C_2$  记录了这些对列向量的操作，它是可逆矩阵。另外， $\gamma_2 = \gamma_4 = \gamma_5 = \mathbf{0}$ ，所以

$$\text{Range}(A) = \text{span}\{\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5\} = \text{span}\{\gamma_1, \gamma_3\}.$$

易见  $\gamma_1, \gamma_3$  线性无关，从而是  $\text{Range}(A)$  的基， $\text{rank}(A) = 2$ 。

$$\text{Range}(A) = \{t\gamma_1 + s\gamma_3 | t, s \in \mathbb{R}\} = \left\{ \begin{bmatrix} t \\ s \\ t - 3s \\ -2t - 5s \end{bmatrix} \middle| t, s \in \mathbb{R} \right\}.$$

因此  $A\mathbf{x} = \mathbf{b}$  有解当且仅当  $\begin{cases} b_3 = b_1 - 3b_2, \\ b_4 = -2b_1 - 5b_2. \end{cases}$

注意到  $\gamma_3 = \beta_3 = \alpha_3 - 2\alpha_1$ ， $\gamma_1 = \beta_1 + \beta_3 = \alpha_1 + (\alpha_3 - 2\alpha_1) = -\alpha_1 + \alpha_3$ ，所以当  $A\mathbf{x} = \mathbf{b}$  有解时，

$$\begin{aligned} \mathbf{b} &= b_1\gamma_1 + b_2\gamma_3 \\ &= b_1(-\alpha_1 + \alpha_3) + b_2(\alpha_3 - 2\alpha_1) \\ &= (-b_1 - 2b_2)\alpha_1 + (b_1 + b_2)\alpha_3, \end{aligned}$$

所以  $\mathbf{x} = \begin{bmatrix} -b_1 - 2b_2 \\ 0 \\ b_1 + b_2 \\ 0 \\ 0 \end{bmatrix}$  是  $A\mathbf{x} = \mathbf{b}$  的一个特解。

总结：列变换的目的是寻找  $\text{Range}(A)$  的基，构造形式简单的基，方便判断  $\mathbf{b} \in \text{Range}(A)$  以及求  $\mathbf{b}$  在这个基下的坐标。通常的列变换包括

- 某列乘一个非零数；
- 某列加其他列的线性组合；
- 交换列的位置

为了求  $A\mathbf{x} = \mathbf{b}$  的通解，我们需要了解齐次方程  $A\mathbf{x} = \mathbf{0}$  的通解。为此，我们考察零空间  $\text{Ker}(A)$ 。我们采用行变换

- 某行乘一个非零数；
- 某行加其他行的线性组合；
- 交换行的位置



行变换会改变方程，但不会改变方程组的解.

$$A = \begin{bmatrix} 1 & 2 & 2 & -5 & 6 \\ -1 & -2 & -1 & 1 & -1 \\ 4 & 8 & 5 & -8 & 9 \\ 3 & 6 & 1 & 5 & -7 \end{bmatrix} \xrightarrow{\substack{(2)+(1) \\ (3)-4(1) \\ (4)-3(1)}} \begin{bmatrix} 1 & 2 & 2 & -5 & 6 \\ 0 & 0 & 1 & -4 & 5 \\ 0 & 0 & -3 & 12 & -15 \\ 0 & 0 & -5 & 20 & -25 \end{bmatrix}$$

$$\xrightarrow{\substack{(1)-2(2) \\ (3)+3(2) \\ (5)+5(1)}} \begin{bmatrix} 1 & 2 & 0 & 3 & -4 \\ 0 & 0 & 1 & -4 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

易见这是前两行线性无关， $A\mathbf{x} = \mathbf{0}$  与以下化简后的方程组同解：

$$\begin{cases} x_1 + 2x_2 + 3x_4 - 4x_5 = 0, \\ x_3 - 4x_4 + 5x_5 = 0. \end{cases}$$

这个化简得到的方程组的系数矩阵为阶梯型矩阵，每行首个非零系数对应的自变量可以由其他自变量表示：

$$\begin{cases} x_1 = -2x_2 - 3x_4 + 4x_5 \\ x_3 = 4x_4 - 5x_5 \end{cases}$$

齐次方程通解为

$$\mathbf{x} = \begin{bmatrix} -2u - 3v + 4w \\ u \\ 4v - 5w \\ v \\ w \end{bmatrix} = u \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + v \begin{bmatrix} -3 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} 4 \\ 0 \\ -5 \\ 0 \\ 1 \end{bmatrix}$$

这三个系数向量线性无关，是  $\text{Ker}(A)$  的基， $\dim \text{Ker}(A) = 3$ .

$A\mathbf{x} = \mathbf{0}$  有无穷多解.

当且仅当  $\mathbf{b}$  满足  $\begin{cases} b_3 = b_1 - 3b_2, \\ b_4 = -2b_1 - 5b_2. \end{cases}$  时， $A\mathbf{x} = \mathbf{b}$  有解.， $A\mathbf{x} = \mathbf{b}$  有解

时，有无穷多解，通解为

$$\mathbf{x} = b_1 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + u \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + v \begin{bmatrix} -3 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix} + w \begin{bmatrix} 4 \\ 0 \\ -5 \\ 0 \\ 1 \end{bmatrix}$$

其中  $u, v, w \in \mathbb{R}$  是任意常数.

□

## 12 行变换, Gauss 消元, 矩阵的 LU 分解

我们把线性方程组的系数矩阵和常数项并列写成扩展矩阵  $[A|b]_{m \times (n+1)}$ , 我们对扩展矩阵只做行变换, 求解上一节中的线性方程组  $Ax = b$ .

$$\begin{aligned}
 & \left[ \begin{array}{ccccc|c} \boxed{1} & 2 & 2 & -3 & 6 & b_1 \\ -1 & -2 & -1 & 1 & -1 & b_2 \\ 4 & 8 & 5 & -8 & 9 & b_3 \\ 3 & 6 & 1 & 5 & -7 & b_4 \end{array} \right] \\
 & \xrightarrow{\substack{(2)+(1) \\ (3)-4(1) \\ (4)-3(1)}}} \left[ \begin{array}{ccccc|c} 1 & 2 & 2 & -5 & 6 & b_1 \\ 0 & 0 & \boxed{1} & -4 & 5 & b_1 + b_2 \\ 0 & 0 & -3 & 12 & -15 & b_3 - 4b_1 \\ 0 & 0 & -5 & 20 & -25 & b_4 - 3b_1 \end{array} \right] \\
 & \xrightarrow{\substack{(1)-2(2) \\ (3)+3(2) \\ (4)+5(2)}}} \left[ \begin{array}{ccccc|c} \boxed{1} & 2 & 0 & 3 & -4 & -b_1 - 2b_2 \\ 0 & 0 & \boxed{1} & -4 & 5 & b_1 + b_2 \\ 0 & 0 & 0 & 0 & 0 & -b_1 + 3b_2 + b_3 \\ 0 & 0 & 0 & 0 & 0 & 2b_1 + 5b_2 + b_4 \end{array} \right]
 \end{aligned}$$

由此可见,  $Ax = b$  有解当且仅当系数全为零的行的最后一列必须为零, 即

$$\begin{cases} -b_1 + 3b_2 + b_3 = 0, \\ 2b_1 + 5b_2 + b_4 = 0, \end{cases} \quad \text{即} \quad \begin{cases} b_3 = b_1 - 3b_2, \\ b_4 = -2b_1 - 5b_2. \end{cases} \quad \text{此时方程组的解满足}$$

$$\begin{cases} x_3 = b_1 + b_2 + 4x_4 - 5x_5 \\ x_1 = -b_1 - 2b_2 - 2x_2 - 3x_4 + 4x_5. \end{cases}$$

于是得到方程组的通解

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = b_1 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 4 \\ 0 \\ -5 \\ 0 \\ 1 \end{bmatrix}.$$

从这个通解可知  $\dim \text{Ker}(A) = 3$  (以  $x_3, x_4, x_5$  为系数的部分),  $\dim \text{Range}(A) = 2$  (以  $b_1, b_2$  为系数的部分在  $A$  下的像).

我们可以用矩阵乘法来实现上述消元过程.

$$\begin{aligned}
 & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ -3 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boxed{1} & 2 & 2 & -3 & 6 & | & b_1 \\ -1 & -2 & -1 & 1 & -1 & | & b_2 \\ 4 & 8 & 5 & -8 & 9 & | & b_3 \\ 3 & 6 & 1 & 5 & -7 & | & b_4 \end{bmatrix} \\
 = & \begin{bmatrix} 1 & 2 & 2 & -5 & 6 & | & b_1 \\ 0 & 0 & \boxed{1} & -4 & 5 & | & b_1 + b_2 \\ 0 & 0 & -3 & 12 & -15 & | & b_3 - 4b_1 \\ 0 & 0 & -5 & 20 & -25 & | & b_4 - 3b_1 \end{bmatrix}, \\
 & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & -5 & 6 & | & b_1 \\ 0 & 0 & \boxed{1} & -4 & 5 & | & b_1 + b_2 \\ 0 & 0 & -3 & 12 & -15 & | & b_3 - 4b_1 \\ 0 & 0 & -5 & 20 & -25 & | & b_4 - 3b_1 \end{bmatrix} \\
 = & \begin{bmatrix} \boxed{1} & 2 & 2 & -5 & 6 & | & b_1 \\ 0 & 0 & \boxed{1} & -4 & 5 & | & b_1 + b_2 \\ 0 & 0 & 0 & 0 & 0 & | & -b_1 + 3b_2 + b_3 \\ 0 & 0 & 0 & 0 & 0 & | & 2b_1 + 5b_2 + b_4 \end{bmatrix}
 \end{aligned}$$

最左边的矩阵是可逆矩阵, 它代表对右侧矩阵做行变换. 如果每次消元都是用某一行消去后续行相应未知数系数, 那么这些行变换矩阵都是下三角矩阵. 因此在第二步消元时, 我们没有像之前那样用第二行化简第一行.

消元的结果使得方程组系数矩阵成为**上阶梯型矩阵**. 这时, 后面的方程具有更少的未知数. 我们从最后一个方程开始求解, 并把得到的解代入上面的方程, 这样也减少了未知数个数. 这个消元过程就是 **Gauss 消元**.

Gauss 消元过程体现为

$$L_k \cdots L_2 L_1 A = U,$$

其中  $L_1, L_2, \dots, L_k$  是下三角形可逆矩阵,  $U$  是上阶梯型矩阵. 这样原方程组  $A\mathbf{x} = \mathbf{b}$  就变成更容易求解的同解方程组  $U\mathbf{x} = L_k \cdots L_2 L_1 \mathbf{b}$ .

对于矩阵  $A$ , 我们得到

$$A = L_1^{-1} L_2^{-2} \cdots L_k^{-1} U = LU,$$

其中  $L = L_1^{-1} L_2^{-2} \cdots L_k^{-1}$  是可逆的下三角矩阵. 这就是矩阵的 LU 分解.

对一些矩阵  $A$ , 在消元过程中由于选主元的需要, 必须调整一些列的次序, 这涉及可逆的排列矩阵  $P$ , 于是任何矩阵  $A$  可以分解为  $PLU$  的形式.

**问题 12.1.** 证明排列矩阵的乘积还是排列矩阵, 排列矩阵是可逆矩阵.

**问题 12.2.** 求 Pascal 矩阵  $\left( \frac{(i+j-2)!}{(i-1)!(j-1)!} \right)_{n \times n}$  的 LU 分解, 并证明 Pascal 矩阵可逆.

## 13 逆矩阵的性质与计算

**定理 13.1.** (1) 若  $A$  是可逆矩阵, 则  $A$  是方阵 (行数 = 列数).

(2) 若  $A$  是可逆矩阵, 则  $A$  有唯一的左逆矩阵和唯一的右逆矩阵, 且左逆矩阵与右逆矩阵相等, 记为  $A^{-1}$ .

(3) 设  $A$  是  $n$  阶方阵. 则  $A$  是可逆矩阵当且仅当存在矩阵  $B$  使得  $AB = I_n$  或  $BA = I_n$ .

(4) 设  $A$  是  $n$  阶方阵. 则  $A$  是可逆矩阵当且仅当  $A$  的各列线性无关, 即  $\text{rank}(A) = n$ .

(5) 若  $A, B$  都是  $n$  阶可逆方阵, 则  $AB$  可逆, 且  $(AB)^{-1} = B^{-1}A^{-1}$ .

(6) 若  $A$  是  $n$  阶可逆方阵, 则  $A$  把  $\mathbb{R}^n$  的任何一组基映为  $\mathbb{R}^n$  的基.

**证明.** (1) 设  $A$  是  $m$  行  $n$  列矩阵.

因为  $A$  是可逆矩阵, 存在矩阵  $B_1, B_2$  使得  $AB_1$  和  $B_2A$  都是单位矩阵.

由  $AB_1$  是单位矩阵知,  $A$  (看成  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射) 是满射, 从而  $\dim \text{Range}(A) = m$ .

由  $B_2A$  是单位矩阵知,  $A$  (看成  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射) 是单射, 从而  $\dim \text{Ker}(A) = 0$ .

所以由维数定理 10.2,  $0 + m = n$ , 因此  $m = n$ , 矩阵  $A$  为方阵.

(2) 设  $B_1, B_2$  分别是  $A$  的左逆矩阵和右逆矩阵. 则  $B_1A = I_n = AB_2$ , 从而

$$B_1 = B_1(AB_2) = (B_1A)B_2 = B_2.$$

因此左逆矩阵和右逆矩阵是相同的矩阵.

任取  $A$  的一个右逆矩阵, 则  $A$  的所有左逆矩阵都等于这个右逆矩阵, 从而  $A$  有唯一的左逆矩阵. 于是这个唯一的左逆矩阵就是  $A$  的唯一的右逆矩阵.

(3) 若  $AB = I_n$ , 则  $A$  是满射, 从而由维数定理知  $A$  是单射, 从而  $A$  是双射, 有逆映射, 矩阵  $A$  可逆.

若  $BA = I_n$ , 则  $A$  是单射, 从而由维数定理知  $A$  是满射, 从而  $A$  是双射, 有逆映射, 矩阵  $A$  可逆.

(4)  $A$  可逆当且仅当  $A$  是单射, 当且仅当  $A$  的列向量线性无关, 当且仅当  $\text{rank}(A) = n$ .

(5)

$$B^{-1}A^{-1}(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n.$$

由 (3) 知,  $AB$  可逆, 且  $B^{-1}A^{-1}$  是  $AB$  的左逆矩阵. 再由 (2) 知  $B^{-1}A^{-1}$  也是  $AB$  的右逆矩阵, 于是  $(AB)^{-1} = B^{-1}A^{-1}$ .

(6) 把  $\mathbb{R}^n$  的基以列向量形式组成矩阵  $B$ . 由 (4) 知  $B$  是可逆的, 从而由 (5) 知  $AB$  是可逆矩阵, 因此由 (4) 知  $AB$  的列构成  $\mathbb{R}^n$  的一组基.  $\square$

由上述定理可知, 求矩阵  $A$  的逆矩阵只需要求解线性方程组  $AX = I_n$ , 而这可以通过对扩展矩阵  $[A|I_n]$  进行行变换得到. 当我们通过一系列行变换 (相当于矩阵  $B$ ) 使系数矩阵  $A$  变成单位阵时,

$$B[A|I_n] = [BA|B] = [I_n|B],$$

右侧的矩阵  $B$  自然就是  $A$  的逆矩阵.

**例 13.2.** 求  $A = \begin{bmatrix} 1 & -2 & -3 \\ 2 & -3 & -10 \\ 3 & -2 & -24 \end{bmatrix}$  的逆矩阵

**解.** 行变换

$$\begin{aligned} & \underbrace{\left[ \begin{array}{ccc|ccc} 1 & -2 & -3 & 1 & 0 & 0 \\ 2 & -3 & -10 & 0 & 1 & 0 \\ 3 & -2 & -24 & 0 & 0 & 1 \end{array} \right]}_A \rightarrow \left[ \begin{array}{ccc|ccc} 1 & -2 & -3 & 1 & 0 & 0 \\ 0 & 1 & -4 & -2 & 1 & 0 \\ 0 & 4 & -15 & -3 & 0 & 1 \end{array} \right] \\ & \rightarrow \left[ \begin{array}{ccc|ccc} 1 & 0 & -11 & -3 & 2 & 0 \\ 0 & 1 & -4 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -4 & 1 \end{array} \right] \rightarrow \left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & 52 & -42 & 11 \\ 0 & 1 & 0 & 18 & -15 & 4 \\ 0 & 0 & 1 & 5 & -4 & 1 \end{array} \right] \\ & \hspace{15em} \underbrace{\hspace{10em}}_{A^{-1}} \end{aligned}$$

另解: 通过 LU 分解

$$\begin{aligned} & \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & -3 \\ 2 & -3 & -10 \\ 3 & -2 & -24 \end{bmatrix} = \begin{bmatrix} 1 & -2 & -3 \\ 0 & 1 & -4 \\ 0 & 4 & -15 \end{bmatrix}, \\ & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & -3 \\ 0 & 1 & -4 \\ 0 & 4 & -15 \end{bmatrix} = \begin{bmatrix} 1 & -2 & -3 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix} = U, \end{aligned}$$

所以

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix}_{L_2} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}_{L_1} A = U, \quad U^{-1} = \begin{bmatrix} 1 & 2 & 11 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix},$$

因此

$$\underbrace{U^{-1}L_2L_1}_{A^{-1}}A = U^{-1}U = I_3,$$

从而

$$\begin{aligned} A^{-1} &= U^{-1}L_2L_1 = \begin{bmatrix} 1 & 2 & 11 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -42 & 11 \\ 0 & -15 & 4 \\ 0 & -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 52 & -42 & 11 \\ 18 & -15 & 4 \\ 5 & -4 & 1 \end{bmatrix} \end{aligned}$$

□

## 14 行变换的几何意义

鸡兔同笼问题

$$\begin{cases} x + y = 10 & (1) \\ 2x + 4y = 24 & (2) \end{cases}$$

在  $(x, y)$  坐标平面中，两个方程分别对应两条直线，求解方程组在几何上是求解两条直线的交点坐标.

每条直线的系数构成一个行向量，上述两个方程对应的行向量分别是  $(1, 1)$  和  $(2, 4)$ .

把方程 (1) 的  $t$  倍加到方程 (2) 上，原来的方程 (2) 就变成

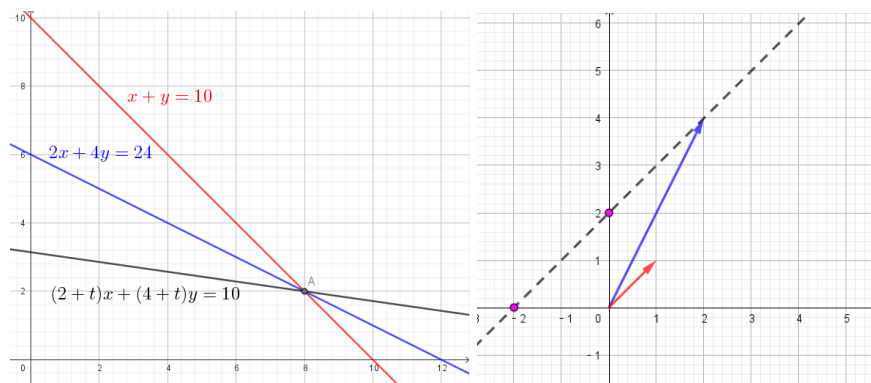
$$(2x + 4y) + t(x + y) = 24 + t \cdot 10$$

即

$$(2 + t)x + (4 + t)y = 24 + 10t \quad (2')$$

相应的行向量变为  $(2 + t, 4 + t)$ .

由于上述过程是可逆的，新的方程 (2') 减去方程 (1) 的  $t$  倍就还原为原来的方程 (2)，所以由 (1) 和 (2') 组成的方程组与原方程组具有相同的解.



这样我们可以选取特殊的  $t$  值, 使得新方程 (2') 中某个变量的系数为零, 也就是这个方程对应的直线在  $(x, y)$  坐标平面中平行于某条坐标轴, 从而实现消元的目的. 上图中的左图为  $(x, y)$  坐标平面中的直线, 右图为行向量空间.

多元线性方程组反映了多个变量之间的耦合联系, 行变换的目的是通过消元解除这种耦合关系. 但行向量本质上是函数, 不是向量, 所以行变换是在函数空间中进行的操作.

另外, 对于线性方程组  $A\mathbf{x} = \mathbf{b}$ , 行变换是在等式两边通过左乘适当矩阵实现方程组形式改变

$$BA\mathbf{x} = B\mathbf{b},$$

这个操作只改变行, 不改变列, 从而不影响要求解的变量  $\mathbf{x}$ .

如果我们对系数矩阵做列变换  $AC$ , 那么原方程就必须改写为

$$ACC^{-1}\mathbf{x} = \mathbf{b}$$

此时记  $\mathbf{y} = C^{-1}\mathbf{x}$ , 原方程组变成

$$\begin{cases} AC\mathbf{y} = \mathbf{b}, \\ \mathbf{x} = C\mathbf{y}. \end{cases}$$

从计算角度上, 行变换要比列变换简单些. 但从几何上理解行变换需要对偶空间的概念.

## 15 对偶空间和对偶映射, 矩阵的转置

设  $V$  是向量空间.  $V$  的成员是向量.

$V$  的对偶空间为

$$V^* = \{f : V \rightarrow \mathbb{R} | f \text{ 是线性函数}\}.$$

$V^*$  的成员是函数.

$V^*$  上可以如下定义加法和数乘：对线性函数  $f, g: V \rightarrow \mathbb{R}$  以及数  $\lambda, \mu \in \mathbb{R}$ ，定义

$$(\lambda f + \mu g)(v) = \lambda f(v) + \mu g(v), \quad \forall v \in V$$

于是  $\lambda f + \mu g: V \rightarrow \mathbb{R}$  也是线性函数， $\lambda f + \mu g \in V^*$ 。

对  $V$  的一组基  $\alpha_1, \dots, \alpha_n$ ，以及任意  $v$ ，存在唯一一组数  $x_1, \dots, x_n$  使得

$$v = x_1 \alpha_1 + \dots + x_n \alpha_n,$$

于是得到坐标函数  $\alpha_k^*: V \rightarrow \mathbb{R}$ ， $\alpha_k^*(v) = x_k$ ， $k = 1, 2, \dots, n$ 。

易见  $\alpha_k^* \in V^*$ ，且

$$\alpha_k^*(\alpha_i) = \begin{cases} 1, & i = k; \\ 0, & i \neq k. \end{cases} \quad (6)$$

对任意  $f \in V^*$ ，

$$f(v) = f(x_1 \alpha_1 + \dots + x_n \alpha_n) = x_1 f(\alpha_1) + \dots + x_n f(\alpha_n) = \sum_{k=1}^n f(\alpha_k) \alpha_k^*(v).$$

因此

$$f = \sum_{k=1}^n f(\alpha_k) \alpha_k^*.$$

若

$$\sum_{k=1}^n \lambda_k \alpha_k^* = 0 \in V^*,$$

则

$$\lambda_i = \sum_{k=1}^n \lambda_k \alpha_k^*(\alpha_i) = 0, \quad \forall i.$$

所以  $\alpha_1^*, \dots, \alpha_n^*$  是  $V^*$  的一组基，它称为对应于  $V$  的基  $\alpha_1, \dots, \alpha_n$  的对偶基。

设  $A: V \rightarrow W$  是线性映射。对任意  $f \in W^*$ ，即  $f: W \rightarrow \mathbb{R}$  是  $W$  上的线性函数，易知  $f \circ A: V \rightarrow \mathbb{R}$  是  $V$  上的线性函数， $f \circ A \in V^*$ ，于是这定义了映射

$$A^*: W^* \rightarrow V^*, \quad f \mapsto f \circ A = A^*(f),$$

称  $A^*$  为  $A$  的对偶映射。

设  $\alpha_1, \dots, \alpha_n$  是  $V$  的基， $\alpha_1^*, \dots, \alpha_n^*$  是  $V^*$  的对偶基； $\beta_1, \dots, \beta_m$  是  $W$  的基， $\beta_1^*, \dots, \beta_m^*$  是  $W^*$  的对偶基。

$A$  的矩阵  $(a_{ij})$ ：

$$(A\alpha_1, \dots, A\alpha_n) = (\beta_1, \dots, \beta_m)(a_{ij})_{m \times n}$$

即  $A\alpha_j = \sum_i a_{ij} \beta_i$ 。于是

$$(A^* \beta_k^*)(\alpha_j) = \beta_k^*(A\alpha_j) = \sum_i a_{ij} \beta_k^*(\beta_i) = a_{kj}$$



$A^*$  的矩阵  $(b_{pq})_{n \times m}$ :

$$(A^*\beta_1, \dots, A^*\beta_m) = (\alpha_1^*, \dots, \alpha_n^*)(b_{ij})_{n \times m}$$

即  $A^*\beta_k^* = \sum_i b_{ik} \alpha_i^*$ . 于是

$$(A^*\beta_k^*)(\mathbf{a}_j) = \sum_i b_{ik} \alpha_i^*(\mathbf{a}_j) = b_{jk}$$

所以

$$a_{kj} = b_{jk}, \quad \forall k, j.$$

即矩阵  $(b_{pq})_{n \times m}$  是矩阵  $(a_{ij})_{m \times n}$  的转置, 即把矩阵  $(b_{ji})_{n \times m}$  的行向量依次写成矩阵  $(a_{ij})_{m \times n}$  的列向量, 记  $(b_{ji}) = (a_{ij})^T$ .

## 16 线性映射和它的对偶映射的关系

设  $A: V \rightarrow W$  是线性映射. 取  $\text{Ker}(A)$  的基  $\alpha_1, \dots, \alpha_k$ , 并把它扩充为  $V$  的基  $\alpha_1, \dots, \alpha_n$ .

**结论1:**  $A\alpha_{k+1}, \dots, A\alpha_n$  是  $\text{Range}(A)$  的一组基.

这是因为

$$\begin{aligned} A(x_1\alpha_1 + \dots + x_n\alpha_n) &= A(x_1\alpha_1 + \dots + x_k\alpha_k) + x_{k+1}A\alpha_{k+1} + \dots + x_nA\alpha_n \\ &= x_{k+1}A\alpha_{k+1} + \dots + x_nA\alpha_n \end{aligned}$$

所以

$$\text{Range}(A) = \text{span}\{A\alpha_{k+1}, \dots, A\alpha_n\}.$$

若  $x_{k+1}A\alpha_{k+1} + \dots + x_nA\alpha_n = \mathbf{0}$ , 则  $x_{k+1}\alpha_{k+1} + \dots + x_n\alpha_n \in \text{Ker}(A)$ , 从而存在数  $x_1, \dots, x_k$  使得

$$x_{k+1}\alpha_{k+1} + \dots + x_n\alpha_n = x_1\alpha_1 + \dots + x_k\alpha_k,$$

所以  $x_1 = \dots = x_n = 0$ , 所以  $A\alpha_{k+1}, \dots, A\alpha_n$  线性无关. 所以它们是  $\text{Range}(A)$  的一组基.  $\square$

把  $A\alpha_{k+1}, \dots, A\alpha_n$  扩充为  $W$  的一组基

$$\beta_1 = A\alpha_{k+1}, \dots, \beta_{n-k} = A\alpha_n, \beta_{n-k+1}, \dots, \beta_m,$$

取  $W^*$  的对偶基  $\beta_1^*, \dots, \beta_{n-k}^*, \beta_{n-k+1}^*, \dots, \beta_m^*$ .

**结论2:**  $\alpha_{k+1}^*, \dots, \alpha_n^*$  是  $\text{Range}(A^*)$  的一组基,  $\beta_{n-k+1}^*, \dots, \beta_m^*$  是  $\text{Ker}(A^*)$  的一组基.

对任何  $f \in W^*$ , 设

$$f = \lambda_1 \beta_1^* + \cdots + \lambda_{n-k} \beta_{n-k}^* + \lambda_{n-k+1} \beta_{n-k+1}^* + \cdots + \lambda_m \beta_m^*.$$

则  $\lambda_i = f(\beta_i)$ ,  $\forall i = 1, 2, \dots, m$ .

对任意  $\mathbf{v} \in V$ ,  $\mathbf{v} = x_1 \alpha_1 + \cdots + x_n \alpha_n$ ,  $x_i = \alpha_i^*(\mathbf{v})$  ( $i = 1, 2, \dots, n$ ),

$$\begin{aligned} A^* f(\mathbf{v}) &= f(A\mathbf{v}) = f(x_1 A\alpha_1 + \cdots + x_k A\alpha_k + x_{k+1} A\alpha_{k+1} + \cdots + x_n A\alpha_n) \\ &= \sum_{i=k+1}^n x_i f(A\alpha_i) \\ &= \sum_{i=k+1}^n \alpha_i^*(\mathbf{v}) \sum_{j=1}^m \lambda_j \beta_j^*(A\alpha_i) = \sum_{i=k+1}^n \alpha_i^*(\mathbf{v}) \lambda_{i-k}, \end{aligned}$$

所以

$$A^* f = \sum_{i=k+1}^n \lambda_{i-k} \alpha_i^* = \sum_{i=k+1}^n f(\beta_i) \alpha_i^*$$

所以  $\alpha_{k+1}^*, \dots, \alpha_n^*$  是  $\text{Range}(A^*)$  的一组基.

另外,  $A^* f = 0$  当且仅当  $\lambda_1 = \cdots = \lambda_{n-k} = 0$ , 当且仅当

$$f \in \text{span}\{\beta_{n-k+1}^*, \dots, \beta_m^*\}.$$

所以  $\beta_{n-k+1}^*, \dots, \beta_m^*$  是  $\text{Ker}(A^*)$  的一组基. □

因此我们得到关于  $A$  和  $A^*$  的以下重要信息:

$V^* : \alpha_1^*, \dots, \alpha_k^*, \underbrace{\alpha_{k+1}^*, \dots, \alpha_n^*}_{\text{Range}(A^*)}$	$\xleftarrow{A^*}$	$W^* : \beta_1^*, \dots, \beta_{n-k}^*, \underbrace{\beta_{n-k+1}^*, \dots, \beta_m^*}_{\text{Ker}(A^*)}$
对偶基 $\uparrow$		$\uparrow$ 对偶基
$V : \underbrace{\alpha_1, \dots, \alpha_k}_{\text{Ker}(A)}, \alpha_{k+1}, \dots, \alpha_n$	$\xrightarrow{A}$	$W : \underbrace{\beta_1, \dots, \beta_{n-k}}_{\text{Range}(A)}, \beta_{n-k+1}, \dots, \beta_m$

对  $V$  的子空间  $V_1$ , 以及  $V^*$  的子空间  $V_2^*$ , 记

$${}^\perp V_1 = \{f \in V^* | f(\mathbf{v}) = 0, \forall \mathbf{v} \in V_1\},$$

$$V_2^{*\perp} = \{\mathbf{v} \in V | f(\mathbf{v}) = 0, \forall f \in V_2^*\},$$

**定理 16.1.** (1)

$${}^\perp \text{Ker}(A) = \text{Range}(A^*), \quad {}^\perp \text{Range}(A) = \text{Ker}(A^*).$$

$$\text{Ker}(A) = \text{Range}(A^*)^\perp, \quad \text{Range}(A) = \text{Ker}(A^*)^\perp.$$

因此  $\mathbf{b} \in \text{Range}(A)$  当且仅当对每个满足  $A^* f = 0$  的  $f$ , 都有  $f(\mathbf{b}) = 0$ .

(2)

$$\dim \text{Range}(A^*) = \dim \text{Range}(A).$$

因此对任何矩阵, 它的行秩等于它的列秩.

(3)

$$\dim \operatorname{Range}(A^*) + \dim \operatorname{Ker}(A) = n.$$

因此  $A^*$  是满射当且仅当  $A$  是单射.

(4)

$$\dim \operatorname{Range}(A) + \dim \operatorname{Ker}(A^*) = m.$$

因此  $A^*$  是单射当且仅当  $A$  是满射.

**推论 16.2.** 对任意  $\mathbf{b}$ , 方程组  $A\mathbf{x} = \mathbf{b}$  有解  $\iff$  方程组  $A^T\mathbf{y} = \mathbf{0}$  只有零解.

例 16.3. 对  $A = \begin{bmatrix} 1 & 2 & 2 & -5 & 6 \\ -1 & -2 & -1 & 1 & -1 \\ 4 & 8 & 5 & -8 & 9 \\ 3 & 6 & 1 & 5 & -7 \end{bmatrix}$ , 得到  $A^T = \begin{bmatrix} 1 & -1 & 4 & 3 \\ 2 & -2 & 8 & 6 \\ 2 & -1 & 5 & 1 \\ -5 & 1 & -8 & 5 \\ 6 & -1 & 9 & -7 \end{bmatrix}$

求四个子空间  $\operatorname{Ker}(A^T)$ ,  $\operatorname{Range}(A^T)$ ,  $\operatorname{Range}(A)$  和  $\operatorname{Ker}(A)$  的基.

解. 对  $A^T$  用行变换得到

$$\begin{bmatrix} 1 & -1 & 4 & 3 \\ 2 & -2 & 8 & 6 \\ 2 & -1 & 5 & 1 \\ -5 & 1 & -8 & 5 \\ 6 & -1 & 9 & -7 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -3 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (*)$$

因此  $A^T\mathbf{y} = \mathbf{0}$  当且仅当  $\begin{cases} y_1 = -y_3 + 2y_4 \\ y_2 = 3y_3 + 5y_4 \end{cases}$  即

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} -t + 2s \\ 3t + 5s \\ t \\ s \end{bmatrix} = t \begin{bmatrix} -1 \\ 3 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 5 \\ 0 \\ 1 \end{bmatrix} \quad (#)$$

把这两个向量写成行向量并组成一个矩阵

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

这个矩阵的零空间就是  $\operatorname{Range}(A)$ ,

$$\begin{cases} x_3 = x_1 - 3x_2, \\ x_4 = -2x_1 - 5x_2 \end{cases}, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = t \begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix} + s \begin{bmatrix} 0 \\ 1 \\ -3 \\ -5 \end{bmatrix},$$

因此  $\text{Range}(A)$  的基是  $\begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -3 \\ -5 \end{bmatrix}$ , 它们是(\*)中矩阵的两个非零行向量的转置.

□

问题 16.4. 求  $\text{Range}(A^T)$  和  $\text{Ker}(A)$  的基, 并证明它们共同组成  $\mathbb{R}^5$  的基.

## 17 线性方程组应用举例

例 17.1 (化学中的应用). 丙烷  $C_3H_8$  燃烧产生二氧化碳  $CO_2$  和水  $H_2O$ , 写出相应的化学反应方程式.

解. 供需平衡

$$\begin{array}{l} C: \\ H: \\ O: \end{array} \begin{bmatrix} 3 & 0 \\ 8 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}$$

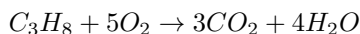
$C_3H_8 \quad O_2 \qquad \qquad CO_2 \quad H_2O$

改写为齐次方程组  $Ax = 0$ , 对系数矩阵做行变换得到

$$\begin{bmatrix} 3 & 0 & -1 & 0 \\ 8 & 0 & 0 & -2 \\ 0 & 2 & -2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 2 & -2 & -1 \\ 0 & 0 & -4 & 3 \end{bmatrix}$$

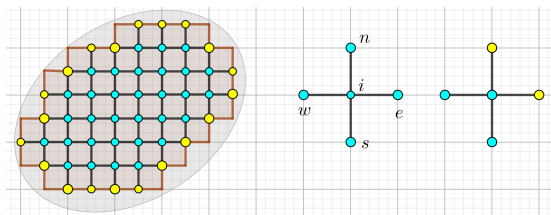
$$\text{解得} \begin{cases} x_3 = \frac{3}{4}x_4 \\ x_2 = x_3 + \frac{1}{2}x_4 = \frac{5}{4}x_4 \\ x_1 = -x_3 + x_4 = \frac{1}{4}x_4 \end{cases}, \text{ 并得到整数特解} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 3 \\ 4 \end{bmatrix}. \text{ 因此反应方程}$$

为



□

例 17.2. 一块铁板已知边缘的温度, 求其内部的温度分布. 根据物理定律, 内部温度满足 Laplace 方程  $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ . 由微积分知识可知每点处的温度等于它周边点的温度的平均值. 我们把铁板划分为网格, 求节点处的温度, 其中最外围节点的温度近似为最近的边界处的温度.



根据以上物理和数学的结论，我们假设内部节点处的温度是相邻节点温度的平均值. 这样就建立了节点温度的线性方程组

$$A\mathbf{u} = \mathbf{f}$$

其中， $\mathbf{f}$  的分量为零或者由外围节点处的已知温度确定， $\mathbf{u}$  是内部节点处的未知温度，矩阵  $A$  由网格结构确定.

例如，如果节点  $i$  周边相邻的四个节点  $e, s, w, n$  也都是内部节点，则相应的方程为

$$u_i - \frac{1}{4}u_e - \frac{1}{4}u_s - \frac{1}{4}u_w - \frac{1}{4}u_n = 0.$$

如果节点  $i$  周边相邻的四个节点中  $s, w$  是内部节点， $e, n$  是外围节点，则相应的方程为

$$u_i - \frac{1}{4}u_s - \frac{1}{4}u_w = \frac{1}{4}f_e + \frac{1}{4}f_n.$$

因此可见  $A = I - M$ ，其中  $M$  的元素都是非负实数，且  $M$  的每行行和不大于 1.

线性方程组  $(I - M)\mathbf{u} = \mathbf{f}$  可以改写为以下不动点形式

$$\mathbf{u} = M\mathbf{u} + \mathbf{f},$$

从而可以使用迭代

$$\mathbf{u}^{(n+1)} = M\mathbf{u}^{(n)} + \mathbf{f}.$$

如果这个迭代收敛，则它的极限是原线性方程组的解.

## 18 线性方程组在经济学中的应用：投入产出模型

国民经济部门分为很多类，我们仅以农业、制造业、服务业为例. 这三个行业的产品除了需要满足最终消费外，它们在正常生产过程中也要消费一部分产品，具体由以下消费矩阵刻画

消费矩阵 $C$		每生产 1 单位产品消耗		
		制造业	农业	服务业
中间消费	制造业	0.5	0.4	0.2
	农业	0.2	0.3	0.1
	服务业	0.1	0.1	0.3

如果最终需求：制造业 50 单位，农业 30 单位，服务业 20 单位，求生产量.

解. 生产量应为最终需求和中间消费的总和.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0.5 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 50 \\ 30 \\ 20 \end{bmatrix}$$

整理成线性方程组的标准形式

$$(I - C) \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \boxed{0.5} & -0.4 & -0.2 \\ -0.2 & \boxed{0.7} & -0.1 \\ -0.1 & -0.1 & \boxed{0.7} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 50 \\ 30 \\ 20 \end{bmatrix}$$

根据以下定义知, 这个系数矩阵  $I - C$  是列对角占优的, 从而根据下面的定理结论知该矩阵可逆, 从而这个方程组有唯一解, 可以用消元法求出方程组的解. 求解过程和最终的解的值留给读者自己完成.  $\square$

**定义 18.1.** 称矩阵  $A = (a_{ij})_{n \times n}$  是行对角占优的, 如果

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad \forall i = 1, 2, \dots, n.$$

类似定义列对角占优的矩阵. 行对角占优矩阵和列对角占优矩阵统称为对角占优的矩阵.

**定理 18.2.** 对角占优矩阵都是可逆的.

**证明.** 设  $A$  是行对角占优矩阵. 因为  $A$  是方阵, 所以只需证明  $A\mathbf{x} = \mathbf{0}$  只有零解.

假设存在非零向量  $\mathbf{x}$  使得  $A\mathbf{x} = \mathbf{0}$ . 于是存在  $i_0$  使得  $|x_{i_0}| = \max_i |x_i| > 0$ . 则

$$a_{i_0 1}x_1 + \cdots + a_{i_0 i_0}x_{i_0} + \cdots + a_{i_0 n}x_n = 0,$$

由此可得如下不等式, 但这不等式与  $A$  行对角占优矛盾.

$$\begin{aligned} |a_{i_0 i_0}x_{i_0}| &= \left| \sum_{j \neq i_0} a_{i_0 j}x_j \right| \\ &\leq \sum_{j \neq i_0} |a_{i_0 j}||x_j| \leq \sum_{j \neq i_0} |a_{i_0 j}||x_{i_0}| \end{aligned}$$

类似处理列对角占优矩阵.  $\square$

对角占优矩阵可以看成是一个可逆的对角矩阵的小扰动

$$\begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix} + \begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & 0 \end{bmatrix}$$

以后我们会知道任何可逆矩阵的任何足够小的扰动也是可逆矩阵.

## 19 投入产出理论中的一个关于非负矩阵的结论

也许上面这个例子的特殊性使得系数矩阵是对角占优的. 但投入产出模型有一套理论, 美籍俄罗斯经济学家 Wassily W. Leontief 因此获得了 1973 年的诺贝尔经济学奖.

**定理 19.1.** 设  $C$  是非负方阵, 记为  $C \geq 0$  (即它的每个元素都是非负数). 若存在  $\mathbf{x} \geq 0$  (即每个  $x_i \geq 0$ ) 使得  $\mathbf{x} > C\mathbf{x}$  (即每个坐标,  $x_i > (C\mathbf{x})_i$ ), 则  $I - C$  可逆, 且  $(I - C)^{-1}$  为非负矩阵.

**注 19.2.** (1) 定理中的条件  $\mathbf{x} > C\mathbf{x}$  的含义是生产量充分满足各种行业的中间需求.

(2) 定理结论是对任何最终需求, 都存在唯一一种生产量恰好满足所有最终消费需求和中间消费需求.

(3) 定理结论实际上是充分必要条件. 事实上, 对任意  $\mathbf{b} > 0$ , 可取  $\mathbf{x} = (I - C)^{-1}\mathbf{b}$ , 则  $\mathbf{x} > 0$  且  $\mathbf{x} = C\mathbf{x} + \mathbf{b} > C\mathbf{x}$ .

**证明.** 设  $\mathbf{x} \geq 0$  满足  $\mathbf{x} > C\mathbf{x}$ . 因  $C$  非负, 故  $C\mathbf{x} \geq 0$ . 从而  $\mathbf{x} > 0$ .

令  $f(t) = \min_{1 \leq i \leq n} \{tx_i - (C\mathbf{x})_i\}$ . 则  $f$  连续,  $f(1) > 0$ .

故存在  $0 < \lambda < 1$  使得  $f(\lambda) > 0$ . 因此  $C\mathbf{x} \leq \lambda\mathbf{x}$ .

所以  $C(\lambda\mathbf{x} - C\mathbf{x}) > 0$ , 由此得  $C^2\mathbf{x} \leq \lambda C\mathbf{x} \leq \lambda\mathbf{x}$ .

用数学归纳法可证明  $\forall n \in \mathbb{N}^*$ ,  $C^n\mathbf{x} \leq \lambda^n\mathbf{x}$

记  $C^n$  最大元素为  $M_n$ , 在第  $\hat{i}$  行第  $\hat{j}$  列.

则

$$M_n x_{\hat{j}} \leq (C^n \mathbf{x})_{\hat{i}} \leq \lambda^n x_{\hat{i}},$$

因此

$$0 \leq (C^n)_{ij} \leq M_n \leq \lambda^n \frac{x_{\hat{i}}}{x_{\hat{j}}} \leq \lambda^n a, \quad a = \frac{\max_i x_i}{\min_j x_j}.$$

于是  $\lim_{n \rightarrow \infty} C^n = 0$ ,

$$(I - C)(I + C + C^2 + \cdots + C^n) = I - C^{n+1} \rightarrow I, \quad n \rightarrow +\infty.$$

设  $K_n = I + C + C^2 + \cdots + C^n$ . 则  $K_{n+p} - K_n = C^{n+1} + \cdots + C^{n+p} \geq 0$ ,

$$\begin{aligned} 0 \leq K_{n+p} - K_n \text{ 的任何元素} &\leq M_{n+1} + M_{n+2} + \cdots + M_{n+p} \\ &< a\lambda^{n+1} + a\lambda^{n+2} + \cdots + a\lambda^{n+p} \\ &\leq \frac{a\lambda^{n+1}}{1 - \lambda} \rightarrow 0, \quad n \rightarrow +\infty, \end{aligned}$$

所以  $(K_n)_{ij}$  都是 Cauchy 数列, 因而收敛, 故存在极限  $\lim_{n \rightarrow +\infty} K_n = K$ . 从而  $(I - C)K = I$ , 故  $I - C$  可逆, 且  $(I - C)^{-1} = K \geq 0$ .  $\square$

**推论 19.3.** 若  $C$  的每个行和  $< 1$  或每个列和  $< 1$ , 则  $I - C$  可逆, 且逆矩阵  $(I - C)^{-1}$  非负.

**证明.** 当  $C$  的行和  $< 1$  时,  $\mathbf{x} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} > C\mathbf{x}$ . 从而由定理结论知  $(I - C)^{-1}$  存在且非负.

当  $C$  的列和  $< 1$  时,  $C^T$  的行和  $< 1$ , 从而  $(I - C^T)^{-1}$  存在且非负, 因此  $(I - C)^{-1} = ((I - C^T)^{-1})^T$ .  $\square$

$(I - C)\mathbf{x} = \mathbf{b}$  可以转化为关于迭代  $\mathbf{x}_{n+1} = C\mathbf{x}_n + \mathbf{b}$  的不动点问题.

这迭代将收敛于不动点 (压缩不动点定理).

## 20 内积空间

**定义 20.1.** 设  $V$  是向量空间, 称二元函数  $\langle \mathbf{x}, \mathbf{y} \rangle$ ,  $V \times V \rightarrow \mathbb{R}$  是  $V$  上的内积, 若它满足

1. 对称:  $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ ,  $\forall \mathbf{u}, \mathbf{v} \in V$ ;

2. 双线性:  $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ ,  $\forall \lambda, \mu \in \mathbb{R}$ ,

$$\langle \lambda \mathbf{u} + \mu \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{u}, \mathbf{w} \rangle + \mu \langle \mathbf{v}, \mathbf{w} \rangle,$$

$$\langle \mathbf{u}, \lambda \mathbf{v} + \mu \mathbf{w} \rangle = \lambda \langle \mathbf{u}, \mathbf{v} \rangle + \mu \langle \mathbf{u}, \mathbf{w} \rangle,$$

即  $\langle \mathbf{u}, \mathbf{v} \rangle$  当其中一个向量固定时, 关于另一个向量是线性的;

3. 正定性:  $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ ,  $\forall \mathbf{u}, \mathbf{v} \in V$ ;  $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0}$

当  $\langle \mathbf{u}, \mathbf{v} \rangle$  是内积时, 记  $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ .

**例 20.2.** (1)  $\mathbb{R}^n$  中标准内积. 对  $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ ,  $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ ,

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$$

(2)  $\mathbb{R}^n$  中非标准内积.

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n \lambda_i x_i y_i$$

其中  $\lambda_1, \dots, \lambda_n > 0$ .



(3)  $\mathbb{R}^n$  中非标准内积.

$$\langle \mathbf{x}, \mathbf{y} \rangle = (x_1, \dots, x_n) A^T A \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = A\mathbf{x} \cdot A\mathbf{y}.$$

$A$  是可逆矩阵.

(4) 由所有  $m$  行  $n$  列的实数矩阵组成一个向量空间  $\mathcal{M}_{m,n}$ . 任何  $A = (a_{ij}) \in \mathcal{M}_{m,n}$  可以看成是一个  $mn$  维的向量, 对  $A, B \in \mathcal{M}_{m,n}$ , 定义

$$\langle A, B \rangle = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij},$$

则  $\langle A, B \rangle$  是  $\mathcal{M}_{m,n}$  上的一个内积. 易见

$$\langle A, B \rangle = \text{tr}(A^T B),$$

这里  $\text{tr}(C)$  表示方阵  $C$  的所有对角元的和, 称为  $C$  的迹 (trace).

**定义 20.3.** 设  $V$  是内积空间, 称  $\mathbf{u}, \mathbf{v}$  是正交的, 如果  $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ .

**定理 20.4.** 设  $V$  是内积空间.

1. (勾股定理) 若  $\mathbf{u}, \mathbf{v}$  是正交的, 则

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

2. (广义勾股定理) 对任意  $\mathbf{u}, \mathbf{v}$ , 都有

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2).$$

**证明.**

$$\begin{aligned} \|\mathbf{u} \pm \mathbf{v}\|^2 &= \langle \mathbf{u} \pm \mathbf{v}, \mathbf{u} \pm \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \pm 2\langle \mathbf{u}, \mathbf{v} \rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \pm 2\langle \mathbf{u}, \mathbf{v} \rangle. \end{aligned}$$

两式相加得到广义勾股定理.

若  $\mathbf{u}, \mathbf{v}$  是正交的, 则得到

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

□

**定理 20.5 (Cauchy-Schwarz-Bunjakovsky 不等式).** 对内积空间  $V$  中任何向量  $\mathbf{u}, \mathbf{v}$ , 都有

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2.$$

**证明.** 不等式两边关于  $u, v$  都有齐次性, 因此不妨设  $\|u\| = \|v\| = 1$ .

令  $w = v - \langle u, v \rangle u$ . 则

$$\langle w, u \rangle = \langle v - \langle u, v \rangle u, u \rangle = \langle u, v \rangle - \langle u, v \rangle \|u\|^2 = 0,$$

所以由勾股定理知

$$1 = \|v\|^2 = \|\langle u, v \rangle u\|^2 + \|w\|^2 = \langle u, v \rangle^2 \|u\|^2 + \|w\|^2 \geq \langle u, v \rangle^2.$$

所以

$$\langle u, v \rangle^2 \leq \|u\|^2 \|v\|^2.$$

等号成立当且仅当  $w = 0$ , 即  $u, v$  共线. □

## 21 内积空间中的度量

长度与距离

**定理 21.1.** 在内积空间  $V$  中, 通过  $\|v\| = \sqrt{\langle v, v \rangle}$  定义了向量  $v$  的长度 (也称为范数), 它满足

1. 正定性: 对任意  $v \in V$ ,  $\|v\| \geq 0$ ;  $\|v\| = 0$  当且仅当  $v = 0$ ;
2. 正齐次性: 对任意  $v \in V$  及任意  $\lambda \in \mathbb{R}$ ,  $\|\lambda v\| = |\lambda| \|v\|$ ;
3. 三角形不等式: 对任意  $u, v \in V$ ,  $\|u + v\| \leq \|u\| + \|v\|$ .

利用向量的长度, 我们可以引入距离,  $d(u, v) = \|u - v\|$ , 它满足

1. 正定性: 对任意  $u, v \in V$ ,  $d(u, v) \geq 0$ ;  $d(u, v) = 0$  当且仅当  $u = v$ ;
2. 三角形不等式: 对任意  $u, v, w \in V$ ,  $d(u, w) \leq d(u, v) + d(v, w)$ ;
3. 平移不变性: 对任意  $u, v, w \in V$ ,  $d(u, v) = d(u + w, v + w)$ .

**证明.** 这里最关键的是范数的三角形不等式, 其他结论都容易被证明.

$$\|u + v\|^2 = \langle u + v, u + v \rangle = \|u\|^2 + \|v\|^2 + 2\langle u, v \rangle \leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\|,$$

最后这个不等式是依据 Cauchy-Schwarz-Bunjakovsky 不等式. □

角度

对非零向量  $u, v \in V$ , 定义

$$\theta = \arccos \frac{\langle u, v \rangle}{\|u\| \|v\|}$$

为  $u$  和  $v$  的夹角 (的大小).

$\theta = 0$  当且仅当  $\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \|\mathbf{v}\|$  即  $\mathbf{u}, \mathbf{v}$  是同方向的两个向量;  $\theta = \pi$  当且仅当  $\langle \mathbf{u}, \mathbf{v} \rangle = -\|\mathbf{u}\| \|\mathbf{v}\|$  即  $\mathbf{u}, \mathbf{v}$  是反方向的两个向量; 同向向量和反向向量都是共线向量, 它们线性相关. 当  $0 < \theta < \pi$  时, 两个向量线性无关. 当  $0 < \theta < \frac{\pi}{2}$  时,  $\mathbf{u}, \mathbf{v}$  有同向的倾向,  $\theta$  越接近 0, 倾向越强; 当  $\frac{\pi}{2} < \theta < \pi$  时,  $\mathbf{u}, \mathbf{v}$  有反向的倾向,  $\theta$  越接近  $\pi$ , 倾向越强; 当  $\theta = \frac{\pi}{2}$  时, 两个向量正交, 是共线倾向最弱的情况, 也是线性无关程度最强的情况.

面积

对两个非零向量  $\mathbf{u}, \mathbf{v} \in V$ , 它们形成的平行四边形面积为

$$\|\mathbf{u}\| \|\mathbf{v}\| \sin \theta = \|\mathbf{u}\| \|\mathbf{v}\| \sqrt{1 - \cos^2 \theta} = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle^2}$$

## 22 内积空间中的基, Gram-Schmidt 正交化过程

**定义 22.1.** 内积空间  $V$  的一组**正交基**, 是一组由彼此正交的向量组成的基.

内积空间  $V$  的一组**单位正交基**, 是一组由彼此正交、长度为 1 的向量组成的基.

**定理 22.2.** 在内积空间  $V$  的任何一组单位正交基  $\mathbf{w}_1, \dots, \mathbf{w}_n$  下,

$$\mathbf{v} = \sum_{k=1}^n \langle \mathbf{v}, \mathbf{w}_k \rangle \mathbf{w}_k,$$

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{k=1}^n \langle \mathbf{u}, \mathbf{w}_k \rangle \langle \mathbf{v}, \mathbf{w}_k \rangle.$$

**证明.** 等式

$$\mathbf{v} = \sum_{k=1}^n x_k \mathbf{w}_k.$$

两边和  $\mathbf{w}_i$  做内积, 得到  $\langle \mathbf{v}, \mathbf{w}_i \rangle = x_i$ .

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= \left\langle \sum_{i=1}^n y_i \mathbf{w}_i, \sum_{j=1}^n x_j \mathbf{w}_j \right\rangle \\ &= \sum_{i,j} y_i x_j \langle \mathbf{w}_i, \mathbf{w}_j \rangle = \sum_{i=1}^n x_i y_i = \sum_{k=1}^n \langle \mathbf{u}, \mathbf{w}_k \rangle \langle \mathbf{v}, \mathbf{w}_k \rangle. \end{aligned}$$

□

**定理 22.3 (Gram-Schmidt 正交化).** 对内积空间  $V$  的任何一组线性无关的向量  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , 都存在一组单位正交的向量  $\mathbf{w}_1, \dots, \mathbf{w}_n$  使得对任意  $1 \leq p \leq n$ ,

$$\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\} = \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_p\}.$$

**证明.** 令  $w_1 = \frac{v_1}{\|v_1\|}$ . 于是  $n = 1$  时结论成立.

假设  $n$  时结论成立. 设  $v_1, \dots, v_{n+1}$  线性无关. 则根据归纳假设存在一组单位正交的向量  $w_1, \dots, w_n$  使得对任意  $1 \leq p \leq n$ ,

$$\text{span}\{v_1, \dots, v_p\} = \text{span}\{w_1, \dots, w_p\}.$$

令

$$u_{n+1} = v_{n+1} - \sum_{k=1}^n \langle v_{n+1}, w_k \rangle w_k.$$

因为  $v_{n+1}$  与  $w_1, \dots, w_n$  线性无关, 所以  $u_{n+1} \neq 0$ . 令

$$w_{n+1} = \frac{u_{n+1}}{\|u_{n+1}\|}.$$

又因为  $\langle u_{n+1}, w_i \rangle = 0$ ,  $i = 1, 2, \dots, n$ , 所以则  $w_1, \dots, w_{n+1}$  是单位正交向量,

$$\text{span}\{v_1, \dots, v_{n+1}\} = \text{span}\{w_1, \dots, w_{n+1}\}.$$

所以由数学归纳法知定理结论对任意  $n$  成立.  $\square$

**推论 22.4.** 设  $V$  是有限维实内积空间. 则

- (1)  $V$  有一组单位正交基;
- (2)  $V$  中任何一组单位正交向量可以扩充为  $V$  的一组单位正交基.  $\square$

**定义 22.5.** 称  $n$  阶矩阵  $A$  是一个**正交矩阵**, 如果它的列向量是  $\mathbb{R}^n$  的一组单位正交基.

**定理 22.6.** 正交矩阵  $A$  满足  $A^T = A^{-1}$ .

**证明.**  $A^T A$  是由  $A$  的列向量配对内积构成的矩阵, 所以  $A^T A = I_n$ . 因此由定理 13.1 知  $A$  可逆, 且  $A^{-1} = A^T$ .  $\square$

**推论 22.7 (矩阵的 QR 分解).** 对任何列满秩的矩阵  $A$  ( $m$  行  $n$  列), 都存在  $m$  阶正交矩阵  $Q$  和  $m$  行  $n$  列矩阵  $R$  使得  $A = QR$ , 其中  $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ ,  $R_1$  是  $n$  阶可逆上三角矩阵.

**证明.** 把  $A$  的列向量用 Gram-Schmidt 正交化方法变成单位正交向量, 并扩充为  $\mathbb{R}^m$  的一组单位正交基, 以这组基为列向量得到矩阵  $Q$ . 从而  $Q^T Q = I_m$ , 并且  $A = QR$ , 其中  $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ ,  $R_1$  是  $n$  阶可逆上三角矩阵.  $\square$

**用 QR 分解求解线性方程组**

$$Ax = b$$

(1) 如果  $A$  列满秩, 则对  $A$  进行 QR 分解,  $A = QR$ ,  $Q$  是正交矩阵,  $R$  是列满秩的上三角矩阵.

$QRx = b$ ,  $Rx = Q^T b$ ,  $R$  是上三角矩阵, 已经完成消元.

如果  $A$  可逆, 则  $R$  可逆, 总有唯一解.

如果  $A$  不是方阵, 则  $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ ,  $R_1$  是可逆的上三角矩阵, 方程组是否有解, 由  $Q^T b$  后半段是否为零确定. 如果有解, 则有唯一解.

(2) 如果  $A$  行满秩, 必要时交换列, 不妨设  $A = (A_1, A_2)$ ,  $A_1$  可逆,  $A_1 = QR$ , 则

$$QRx_1 + A_2x_2 = b,$$

$$Rx_1 = Q^T b - Q^T A_2 x_2,$$

有无穷多解.

**问题 22.8.** 如果内积空间中任给一组基  $v_1, \dots, v_n$ , 如何用内积表示  $v$  在这组基下的坐标?

**问题 22.9.** 设  $A_1, B_1, C_1$  分别是三角形  $ABC$  三边  $BC, CA, AB$  所在直线上的点, 满足  $AA_1 \perp BC, BB_1 \perp CA, CC_1 \perp AB$ . 证明  $AA_1, BB_1, CC_1$  交于同一点.

## 23 正交变换

**定义 23.1.** 设  $V, W$  都是内积空间. 称线性映射  $A: V \rightarrow W$  保内积, 如果对任意  $u, v \in V$ , 都有

$$\langle Au, Av \rangle = \langle u, v \rangle.$$

**定理 23.2.** 设  $A: V \rightarrow W$  是保内积的线性映射. 则

1.  $A$  保长度和角度;
2.  $A$  是单射, 从而  $\dim V \leq \dim W$ ;
3. 若  $\dim V = \dim W$ , 则  $A$  可逆, 且  $A$  映单位正交基为单位正交基. 当  $V = W$  时, 称  $A$  为  $V$  上的一个正交变换. 正交变换在在一组标准正交基下的矩阵为正交矩阵.

**证明.** (1) 是显然的.

(2)  $Av = 0$ , 则  $\|v\| = \|Av\| = 0$ , 从而  $v = 0$ , 因此  $A$  是单射,  $\dim W \geq \dim \text{Range}(A) = \dim V - \dim \text{Ker}(A) = \dim V$ .

(3) 是 (1) (2) 和推论 10.3 的结论.

设  $A\alpha_j = \sum_i a_{ij}\alpha_i$ . 则

$$\langle A\alpha_p, A\alpha_q \rangle = \sum_{ij} a_{ip}a_{jq}\langle \alpha_i, \alpha_j \rangle = \sum_i a_{ip}a_{iq} = (A^T A)_{p,q} = \begin{cases} 1, & p = q; \\ 0, & p \neq q. \end{cases}$$

所以  $A$  的矩阵是正交矩阵. □

**问题 23.3.** 证明保长度的线性映射必然保内积.

$\mathbb{R}^2$  上的正交变换

旋转:

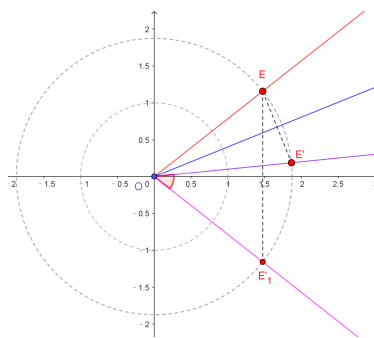
$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

(关于  $x$  轴的) 反射:

$$F_0 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}$$

关于单位向量  $(\cos \theta, \sin \theta)$  所在直线的反射:

$$F = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}.$$



一般地,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

是正交矩阵当且仅当

$$a^2 + b^2 = c^2 + d^2 = 1, \quad a^2 + c^2 = b^2 + d^2 = 1, \quad ab + cd = 0 = ac + bd.$$

所以

$$|a| = |d|, |b| = |c|,$$

情形1:  $a = d \neq 0$ , 此时  $c = -b$ ,  $A = \begin{bmatrix} a & -c \\ c & a \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$  是旋转.

情形2:  $a = -d \neq 0$ , 此时  $c = b$ ,

$$A = \begin{bmatrix} a & c \\ c & -a \end{bmatrix} = \begin{bmatrix} a & -c \\ c & a \end{bmatrix} \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}$$

是旋转与反射的复合.

情形3:  $a = d = 0$ ,

(3.1)  $b = c = 1$ ,  $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$  是旋转与反射的复合.

(3.2)  $b = c = -1$ ,  $A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$  是旋转与反射的复合.

(3.3)  $b = 1, c = -1$ ,  $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$  是旋转.

(3.4)  $b = -1, c = 1$ ,  $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$  是旋转.

平面上的正交变换要么是一次旋转, 要么是一次反射, 要么是一次旋转与一次反射的复合.

**问题 23.4.** 证明平面上任何正交变换要么是一次反射, 要么是两次反射的复合.

$\mathbb{R}^n$  上的正交变换

**问题 23.5.** 给定  $\mathbb{R}^n$  上两个不共线的单位向量  $v_1, v_2$ , 构造一个旋转  $A$  使得  $Av_1 = v_2$ , 这个旋转在与  $v_1, v_2$  都正交的方向上保持不变.

**问题 23.6.** 给定  $\mathbb{R}^n$  上两个不同的单位向量  $v_1, v_2$ , 构造一个反射  $A$  使得  $Av_1 = v_2$ .

**问题 23.7.** 给定  $\mathbb{R}^n$  上两组单位正交基  $v_1, \dots, v_n$  和  $w_1, \dots, w_n$ , 证明存在有限多次旋转以及最多一次反射可以把  $v_1, \dots, v_n$  变成  $w_1, \dots, w_n$ .

**问题 23.8.** 给定  $\mathbb{R}^n$  上两组单位正交基  $v_1, \dots, v_n$  和  $w_1, \dots, w_n$ , 证明存在有限多次反射可以把  $v_1, \dots, v_n$  变成  $w_1, \dots, w_n$ .

**Householder 变换**

给定单位向量  $u \in \mathbb{R}^n$ ,

$$H_u = I_n - 2uu^T$$

是一个正交变换, 它是关于以  $u$  为法向量的平面的反射.

**问题 23.9.** 证明  $H_u$  是对称矩阵, 且  $H_u^{-1} = H_u$ .

**例 23.10.** 分别用 Gram-Schmidt 正交化过程和 Householder 变换对矩阵

$$A = \begin{bmatrix} 2 & 7 & 21 \\ 8 & -8 & 3 \\ 16 & -7 & 6 \end{bmatrix}$$

做 QR 分解.

解：用 **Gram-Schmidt** 正交化. 记矩阵  $A$  的三个列向量依次为  $\beta_1, \beta_2, \beta_3$ .

$$w_1 = \frac{\beta_1}{\|\beta_1\|} = \begin{bmatrix} \frac{1}{9} \\ \frac{4}{9} \\ \frac{8}{9} \end{bmatrix}, \quad \langle \beta_1, w_1 \rangle = 18, \quad \langle \beta_2, w_1 \rangle = -9, \quad \langle \beta_3, w_1 \rangle = 9.$$

$$\gamma_2 = \beta_2 - (-9)w_1 = \begin{bmatrix} 8 \\ -4 \\ 1 \end{bmatrix}$$

$$w_2 = \frac{\gamma_2}{\|\gamma_2\|} = \begin{bmatrix} \frac{8}{9} \\ -\frac{4}{9} \\ \frac{1}{9} \end{bmatrix}, \quad \langle \beta_1, w_2 \rangle = 0, \quad \langle \beta_2, w_2 \rangle = 9, \quad \langle \beta_3, w_2 \rangle = 18.$$

$$\gamma_3 = \beta_3 - 9w_1 - 18w_2 = \begin{bmatrix} 4 \\ 7 \\ -4 \end{bmatrix}$$

$$w_3 = \frac{\gamma_3}{\|\gamma_3\|} = \begin{bmatrix} \frac{4}{9} \\ \frac{7}{9} \\ -\frac{4}{9} \end{bmatrix}, \quad \langle \beta_1, w_3 \rangle = \langle \beta_2, w_3 \rangle = 0, \quad \langle \beta_3, w_3 \rangle = 9.$$

因此

$$A = QR = (w_1, w_2, w_3) \begin{bmatrix} 18 & -9 & 9 \\ & 9 & 18 \\ & & 9 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{8}{9} & \frac{4}{9} \\ \frac{4}{9} & -\frac{4}{9} & \frac{7}{9} \\ \frac{8}{9} & \frac{1}{9} & -\frac{4}{9} \end{bmatrix} \begin{bmatrix} 18 & -9 & 9 \\ & 9 & 18 \\ & & 9 \end{bmatrix}.$$

□

解：用 **Householder** 变换.

$$\|\beta_1\| = 18, \quad R \text{ 的第一列} \begin{bmatrix} 18 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 18 \\ 0 \\ 0 \end{bmatrix} - \beta_1 = \begin{bmatrix} 16 \\ -8 \\ -16 \end{bmatrix}, \quad \text{单位化 } u_1 = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{2}{3} \end{bmatrix}$$

$$H_{u_1} = I_3 - 2u_1u_1^T = \begin{bmatrix} \frac{1}{9} & \frac{4}{9} & \frac{8}{9} \\ \frac{4}{9} & \frac{7}{9} & -\frac{4}{9} \\ \frac{8}{9} & -\frac{4}{9} & \frac{1}{9} \end{bmatrix}$$

$$H_{u_1}A = \begin{bmatrix} 18 & -9 & 9 \\ 0 & 0 & 9 \\ 0 & 9 & 18 \end{bmatrix}$$



交换后两行

$$H_2 = \begin{bmatrix} 1 & & \\ & 0 & 1 \\ & 1 & 0 \end{bmatrix}$$

得到

$$H_2 H_{u_1} A = \begin{bmatrix} 18 & -9 & 9 \\ 0 & 9 & 18 \\ 0 & 0 & 9 \end{bmatrix}$$

所以

$$A = H_{u_1} H_2 \begin{bmatrix} 18 & -9 & 9 \\ 0 & 9 & 18 \\ 0 & 0 & 9 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{8}{9} & \frac{4}{9} \\ \frac{4}{9} & -\frac{4}{9} & \frac{7}{9} \\ \frac{8}{9} & \frac{1}{9} & -\frac{4}{9} \end{bmatrix} \begin{bmatrix} 18 & -9 & 9 \\ & 9 & 18 \\ & & 9 \end{bmatrix}.$$

□

## 24 线性子空间及其运算

子空间的交与和

**定理 24.1.** 设  $W_1, \dots, W_k$  是向量空间  $V$  的线性子空间. 则

1.  $W_1 \cap W_2 \cap \dots \cap W_k$  是  $V$  的线性子空间;
2.  $W_1 + W_2 + \dots + W_k$  是  $V$  的线性子空间, 其中

$$W_1 + \dots + W_k = \{v \mid \text{存在 } v_i \in W_i, i = 1, 2, \dots, k \text{ 使得 } v = v_1 + \dots + v_k\};$$

3. (**Grassmann**)  $\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$ .

**证明.** (1) 和 (2) 可以根据定义直接计算验证.

- (3) 取  $W_1 \cap W_2$  的一组基  $\gamma_1, \dots, \gamma_p$ , 把它扩充为  $W_1$  的基

$$\gamma_1, \dots, \gamma_p, \alpha_{p+1}, \dots, \alpha_n,$$

以及  $W_2$  的基

$$\gamma_1, \dots, \gamma_p, \beta_{p+1}, \dots, \beta_m.$$

易见

$$W_1 + W_2 \subseteq \text{span}\{\alpha_{p+1}, \dots, \alpha_n, \gamma_1, \dots, \gamma_p, \beta_{p+1}, \dots, \beta_m\}.$$

我们证明这些向量线性无关.

设

$$x_{p+1}\alpha_{p+1} + \dots + x_n\alpha_n + x_1\gamma_1 + \dots + x_p\gamma_p + y_{p+1}\beta_{p+1} + \dots + y_m\beta_m = \mathbf{0}.$$

假设  $y_{p+1}, \dots, y_m$  不全为零, 则  $\beta = y_{p+1}\beta_{p+1} + \dots + y_m\beta_m \neq \mathbf{0}$ , 且

$$\beta = -x_{p+1}\alpha_{p+1} - \dots - x_n\alpha_n - x_1\gamma_1 - \dots - x_p\gamma_p \in W_1 \cap W_2,$$

从而存在数  $z_1, \dots, z_p$  使得

$$\beta = z_1\gamma_1 + \dots + z_p\gamma_p.$$

因此

$$x_{p+1}\alpha_{p+1} + \dots + x_n\alpha_n + (x_1 + z_1)\gamma_1 + \dots + (x_p + z_p)\gamma_p = \mathbf{0}.$$

但这些向量是线性无关的, 所以  $x_{p+1} = \dots = x_n = 0$ . 因此

$$x_1\gamma_1 + \dots + x_p\gamma_p + y_{p+1}\beta_{p+1} + \dots + y_m\beta_m = \mathbf{0}.$$

然而这些向量也是线性无关的, 所以  $y_{p+1} = \dots = y_m = 0$ . 这与假设它们不全为零矛盾.

因此  $y_{p+1} = \dots = y_m = 0$ . 同理可证  $x_{p+1} = \dots = x_n = 0$ . 此时

$$x_1\gamma_1 + \dots + x_p\gamma_p = \mathbf{0}.$$

再由线性无关性得到  $x_1 = \dots = x_p = 0$ . 因此

$$\alpha_{p+1}, \dots, \alpha_n, \gamma_1, \dots, \gamma_p, \beta_{p+1}, \dots, \beta_m$$

线性无关, 是  $W_1 + W_2$  的基. □

**问题 24.2.** 请找出以上证明的漏洞. 你能修补这个漏洞吗?

**问题 24.3.** 对多个子空间的和, 维数公式是怎样的? 你能证明你的结论吗?

**子空间的直和, 补空间, 正交补空间**

**定义 24.4.** 设  $W_1, \dots, W_k$  是向量空间  $V$  的线性子空间. 称  $W_1 + W_2 + \dots + W_k$  是直和, 如果对任意  $\mathbf{v} \in W_1 + \dots + W_k$ , 都存在唯一一组  $\mathbf{v}_1 \in W_1, \dots, \mathbf{v}_k \in W_k$  使得  $\mathbf{v} = \mathbf{v}_1 + \dots + \mathbf{v}_k$ . 记  $W_1 + W_2 + \dots + W_k = W_1 \oplus W_2 \oplus \dots \oplus W_k$ .

如果  $V = W_1 \oplus W_2$ , 则称  $W_1, W_2$  为对方在  $V$  中的补空间.

对于直和  $W_1 \oplus W_2 \oplus \dots \oplus W_k$ ,

$$P_i : W_1 \oplus W_2 \oplus \dots \oplus W_k \rightarrow W_i, P_i(\mathbf{v}) = \mathbf{v}_i$$

是关于这个直和的一个投影映射.

**问题 24.5.** 若线性映射  $P : V \rightarrow V$  满足  $P^2 = P$ , 证明  $P$  是关于  $V$  的一个直和分解的投影.

**定理 24.6.** 以下每个陈述都是  $W_1 + W_2 + \cdots + W_k$  是直和的充分必要条件

1. 若  $\mathbf{v}_1 \in W_1, \dots, \mathbf{v}_k \in W_k$  使得  $\mathbf{0} = \mathbf{v}_1 + \cdots + \mathbf{v}_k$ , 则每个  $\mathbf{v}_i = \mathbf{0}$ ;
2. 对  $\{1, 2, \dots, k\}$  的任意分割  $I, J$  (即  $I \cup J = \{1, 2, \dots, k\}, I \cap J = \emptyset$ ),  

$$\left( \sum_{i \in I} W_i \right) \cap \left( \sum_{j \in J} W_j \right) = \{\mathbf{0}\};$$
3.  $W_1, W_2, \dots, W_k$  各自取一组基, 它们组成  $W_1 + W_2 + \cdots + W_k$  的一组基.

**证明.** 很明显, 直和的定义  $\implies$  (1).

(1)  $\implies$  (2): 设  $\mathbf{v} \in \left( \sum_{i \in I} W_i \right) \cap \left( \sum_{j \in J} W_j \right)$ . 则对每个  $p$ , 存在  $\mathbf{v}_p \in W_p$ , 使得

$$\mathbf{v} = \sum_{i \in I} \mathbf{v}_i = \sum_{j \in J} \mathbf{v}_j.$$

对每个  $p$ , 取  $\delta_p = \begin{cases} 1, & p \in I, \\ -1, & p \in J. \end{cases}$  则

$$\delta_1 \mathbf{v}_1 + \delta_2 \mathbf{v}_2 + \cdots + \delta_k \mathbf{v}_k = \mathbf{v} - \mathbf{v} = \mathbf{0}.$$

由 (1) 知每个  $\delta_p \mathbf{v}_p = \mathbf{0}$ , 从而  $\mathbf{v}_p = \mathbf{0}$ , 因此  $\mathbf{v} = \mathbf{0}$ .

(2)  $\implies$  (3): 设  $\mathbf{v}_1^{(p)}, \dots, \mathbf{v}_{n_p}^{(p)}$  是  $W_p$  的一组基, 则  $W_1 + W_2 + \cdots + W_k$  中的向量可由

$$\mathbf{v}_1^{(1)}, \dots, \mathbf{v}_{n_1}^{(1)}; \dots; \mathbf{v}_1^{(k)}, \dots, \mathbf{v}_{n_k}^{(k)} \quad (*)$$

线性表示. 为证明它们是  $W_1 + W_2 + \cdots + W_k$  的一组基, 只需证明它们线性无关.

假设(\*)中向量线性相关. 则存在不全为零的数  $x_1^{(1)}, \dots, x_{n_1}^{(1)}, \dots, x_1^{(k)}, \dots, x_{n_k}^{(k)}$  使得

$$x_1^{(1)} \mathbf{v}_1^{(1)} + \cdots + x_{n_1}^{(1)} \mathbf{v}_{n_1}^{(1)} + \cdots + x_1^{(k)} \mathbf{v}_1^{(k)} + \cdots + x_{n_k}^{(k)} \mathbf{v}_{n_k}^{(k)} = \mathbf{0}.$$

不妨设  $x_1^{(1)} \neq 0$ . 则

$$-x_1^{(1)} \mathbf{v}_1^{(1)} = x_2^{(1)} \mathbf{v}_2^{(1)} + \cdots + x_{n_1}^{(1)} \mathbf{v}_{n_1}^{(1)} + \cdots + x_1^{(k)} \mathbf{v}_1^{(k)} + \cdots + x_{n_k}^{(k)} \mathbf{v}_{n_k}^{(k)} \in W_1 \cap \left( \sum_{p>1} W_p \right),$$

但这个向量非零, 这与(2)矛盾. 因此(\*)中向量线性无关, 是  $W_1 + W_2 + \cdots + W_k$  的一组基.

(3)  $\implies$  定义: 给每个  $W_p$  取定一组基, 根据(3), 它们构成  $W_1 + W_2 + \cdots + W_k$  的一组基. 任何  $\mathbf{v} \in W_1 + W_2 + \cdots + W_k$  可以唯一表示为这组基的线性组合, 其中用  $W_p$  的基表示的线性组合记为  $\mathbf{v}_p$ , 它是由  $\mathbf{v}$  唯一确定的. 所以定义成立,  $W_1 + W_2 + \cdots + W_k$  是直和.  $\square$

**推论 24.7.** 设  $W_1, \dots, W_k$  是内积空间  $V$  的线性子空间, 满足对任意  $1 \leq i < j \leq k$ , 以及任意  $\mathbf{v}_i \in W_i$  和任意  $\mathbf{v}_j \in W_j$ , 都有  $\mathbf{v}_i \perp \mathbf{v}_j$ . 则  $W_1 + W_2 + \dots + W_k$  是直和.

特别当  $k = 2$  且  $V = W_1 + W_2$  时, 称  $W_1, W_2$  互为对方在  $V$  中的正交补空间.

**证明.** 设  $\mathbf{0} = \mathbf{v}_1 + \dots + \mathbf{v}_k$ , 其中每个  $\mathbf{v}_i \in W_i$ . 则

$$\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, -\mathbf{v}_2 - \dots - \mathbf{v}_k \rangle = 0,$$

所以  $\mathbf{v}_1 = \mathbf{0}$ . 同理每个  $\mathbf{v}_i = \mathbf{0}$ . 由上述定理,  $W_1 + W_2 + \dots + W_k$  是直和.  $\square$

**定理 24.8.** 设  $V$  是向量空间,  $\dim V = n$ . 则

1.  $V$  的任何线性子空间  $W$  都有补空间;
2. 若  $V$  还是内积空间, 则  $V$  的任何线性子空间  $W$  都有唯一的正交补空间  $W^\perp$ .

**证明.** (1) 把  $W$  的一组基  $\mathbf{v}_1, \dots, \mathbf{v}_p$  补齐成  $V$  的一组基  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , 则

$$W_1 = \text{span}\{\mathbf{v}_{p+1}, \dots, \mathbf{v}_n\}$$

是  $W$  的一个补空间.

(2) 把  $W$  的一组基  $\mathbf{v}_1, \dots, \mathbf{v}_p$  补齐成  $V$  的一组基  $\mathbf{v}_1, \dots, \mathbf{v}_n$ , 并用 Gram-Schmidt 正交化成为标准正交基  $\mathbf{w}_1, \dots, \mathbf{w}_n$ . 则

$$W = \text{span}\{\mathbf{w}_1, \dots, \mathbf{w}_p\}, \quad W^\perp = \text{span}\{\mathbf{w}_{p+1}, \dots, \mathbf{w}_n\}.$$

$\square$

**问题 24.9.** 请找出以上证明的漏洞. 你能修补这个漏洞吗?

**空间分解与映射 (矩阵) 分块**

设  $V = V_1 \oplus V_2 \oplus V_3$ ,  $W = W_1 \oplus W_2$  是直和分解. 则线性映射  $A: V \rightarrow W$  可以表示为

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

其中  $A_{ij} = P_i A|_{V_i}: V_j \rightarrow W_i$  是线性映射,  $P_i: W_1 \oplus W_2 \rightarrow W_i$  是投影映射.

**问题 24.10.** 设  $A: V \rightarrow V$  是线性映射,  $W$  是  $V$  的一个线性子空间, 且是  $A$ -不变的 (即对任意  $\mathbf{w} \in W$ ,  $A\mathbf{w} \in W$ ). 证明存在  $V$  的一组基, 使得  $A$  的矩阵为上三角块矩阵  $\begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}$ .

**问题 24.11.** 设  $A: V \rightarrow V$  是线性映射,  $W$  是  $V$  的一个线性子空间, 且是  $A$ -不变的. 是否存在  $A$ -不变的子空间是  $W$  的补空间? 如果存在, 那么  $W$  是否有唯一的  $A$ -不变补空间.

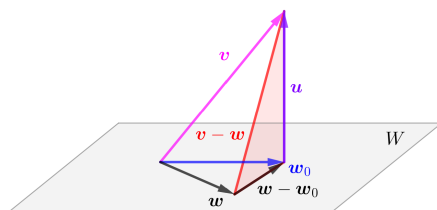
## 25 最小二乘法

### 基本模型

**定理 25.1.** 设  $W$  是内积空间  $V$  的一个线性子空间,  $\{0\} \subset W \subset V$ . 则对任意  $v \in V$ , 存在唯一的  $w_0$  使得

$$\|v - w_0\| = \min_{w \in W} \|v - w\|.$$

这个  $w_0$  是  $v$  在  $W$  中的正交投影, 即  $v - w_0 \in W^\perp$ .



**证明.** 取  $W$  的正交补空间  $W^\perp$ , 则  $v \in V$  可唯一分解为

$$v = w_0 + u, \quad w_0 \in W, u \in W^\perp.$$

于是对任意  $w \in W$ ,

$$v - w = (w_0 - w) + u \in W \oplus W^\perp,$$

由勾股定理,

$$\|v - w\|^2 = \|w_0 - w\|^2 + \|u\|^2 \geq \|w_0 - w\|^2,$$

且不等式中等号成立当且仅当  $w = w_0$ . □

评估最小二乘解的合理性, 不能利用  $\|v - w_0\|$  的大小, 因为在实际问题中这依赖于物理量纲的选择.

统计中用  $v$  和  $W$  的夹角来判断最小二乘法的合理性, 这可以用

$$R^2 = \frac{\|w_0\|^2}{\|v\|^2}, \quad (7)$$

当  $R^2$  接近于 1 时,  $v$  和  $W$  的夹角很小, 说明  $\|v - w_0\|$  很可能是测量中的随机误差造成的.

### 线性方程组的最小二乘解

在科学研究或工程实践中, 常常要通过实验观测确定一些物理量之间的关系. 一般而言, 为了减小随机误差对判断的影响, 观测次数往往要比模型参数

多很多. 当方程个数远多于未知数个数时, 线性方程组  $A\mathbf{x} = \mathbf{b}$  中通常含有相互矛盾的等式, 从而无解. 此时, 我们寻求最小二乘意义下的近似解:

$$\|A\mathbf{x}_0 - \mathbf{b}\| = \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\| = \min_{\mathbf{y} \in \text{Range}(A)} \|\mathbf{y} - \mathbf{b}\|.$$

根据定理 25.1, 我们知道  $A\mathbf{x}_0$  是  $\mathbf{b}$  在  $\text{Range}(A)$  中的正交投影. 因此

$$(A\mathbf{x})^T(\mathbf{b} - A\mathbf{x}_0) = 0, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

上式等价于

$$\mathbf{x}^T(A^T\mathbf{b} - A^TA\mathbf{x}_0) = 0, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

这等价于

$$A^TA\mathbf{x}_0 = A^T\mathbf{b}. \quad (8)$$

所以我们将方程 (8) 称为正交方程.

当  $A$  列满秩时,  $A^TA$  可逆: 一方面, 它是方阵; 另一方面, 若

$$A^TA\mathbf{x} = \mathbf{0},$$

则

$$\|A\mathbf{x}\|^2 = \mathbf{x}^TA^TA\mathbf{x} = 0,$$

从而  $A\mathbf{x} = \mathbf{0}$ , 又因为  $A$  列满秩, 所以  $\mathbf{x} = \mathbf{0}$ , 所以  $A^TA$  是单射, 从而可逆.

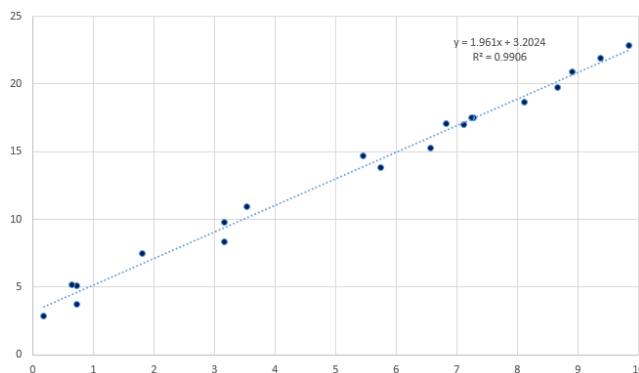
因此正交方程有唯一解

$$\mathbf{x}_0 = (A^*A)^{-1}A^*\mathbf{b}.$$

**例 25.2.** 实验数据记录如下

$x_i$	6.82	5.74	0.73	3.54	5.46	0.65	7.12	0.73	0.18	7.28
$y_i$	17.06	13.81	5.09	10.91	14.7	5.17	17.00	3.69	2.88	17.53
$x_i$	6.57	3.16	1.81	7.25	8.67	8.91	8.11	9.38	3.17	9.85
$y_i$	15.24	9.8	7.49	17.48	19.72	20.91	18.63	21.91	8.35	22.84

实验数据的散点图如下



可以看出  $x, y$  近似成直线关系  $y = ax + b$ . 我们希望找到系数  $a, b$  的值使得

$$\sum_{k=1}^{20} |y_k - ax_k - b|^2$$

最小.

$y = ax + b$  关于待求的系数  $a, b$  是线性组合, 这等价于

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}.$$

因为  $x_1, \dots, x_N$  不全相等, 所以上式左端矩阵列满秩, 从而正交方程

$$\begin{bmatrix} x_1 & \cdots & x_N \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_N \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

有唯一解. 这样求得  $a, b$  的最小二乘解. Excel 对规模不大的数据, 可以给出线性回归方程和决定系数<sup>4</sup>  $R^2$  (见 (7)), 我们看到在这个例子中  $R^2 = 0.9906$ , 这说明线性模型是合理的选择.  $\square$

#### QR 分解在最小二乘法中的应用

列满秩的矩阵  $A$ , 通过添加列把  $A$  扩充为一个可逆矩阵  $(A, A_2)$ , 然后对这个可逆矩阵做 QR 分解,  $(A, A_2) = QR = Q(R_1, R_2)$ , 其中  $Q$  是正交矩阵,  $Q = (Q_1, Q_2)$ ,  $Q_1$  与  $A$  具有相同的列数,  $R_1 = \begin{bmatrix} \tilde{R} \\ 0 \end{bmatrix}$ ,  $\tilde{R}$  是可逆的上三角矩阵, 与  $A$  具有相同的列数.

$$\|Ax - b\|^2 = \|QR_1x - b\|^2 = \|R_1x - Q^Tb\|^2 = \|\tilde{R}x - Q_1^Tb\|^2 + \|Q_2^Tb\|^2 \geq \|Q_2^Tb\|^2,$$

不等式中等号成立当且仅当  $x = \tilde{R}^{-1}Q_1^Tb$ .

#### 一些可以转化为线性最小二乘的非线性模型

$$y = Ax^\beta, \ln y = \ln A + \beta \ln x$$

$$y = Aa^x, \ln y = \ln A + x \ln a$$

$$y = A \ln x + B,$$

$$y = a_mx^m + \cdots + a_1x + a_0,$$

$$\begin{bmatrix} x_1^m & \cdots & x_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_N^m & \cdots & x_N & 1 \end{bmatrix} \begin{bmatrix} a_m \\ \vdots \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

<sup>4</sup> $R^2$  是衡量一个统计模型对观测数据的拟合优度的指标,  $R^2 = 1$  说明完美拟合,  $R^2 = 0$  说明拟合极差

## 26 Riesz 表示定理, 内积空间上的线性函数与对偶映射

内积空间的对偶空间以及内积空间上线性映射的对偶映射要比一般情况简单, 主要是因为如下结论.

**定理 26.1 (Riesz 表示定理).** 设  $V$  是内积空间. 则对任何线性函数  $f \in V^*$  ( $f: V \rightarrow \mathbb{R}$ ), 存在唯一的向量  $\mathbf{u} \in V$  使得对任意  $\mathbf{v} \in V$ ,

$$f(\mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle.$$

**证明.** 若  $f \equiv 0$ , 则  $\mathbf{u} = \mathbf{0}$  满足  $f(\mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle$ .

若  $f \not\equiv 0$ , 则存在  $\mathbf{v}_1 \in V$  使得  $f(\mathbf{v}_1) \neq 0$ . 因此对任意  $t \in \mathbb{R}$ ,

$$f\left(\frac{t}{f(\mathbf{v}_1)}\mathbf{v}_1\right) = t.$$

因此  $\text{Range}(f) = \mathbb{R}$ , 从而由维数定理 10.2 知

$$\dim \text{Ker}(f) = n - 1.$$

故  $(\text{Ker}(f))^\perp$  是  $V$  的一维子空间. 取非零向量  $\mathbf{u}_1 \in (\text{Ker}(f))^\perp$ . 则  $f(\mathbf{u}_1) \neq 0$ .

因此  $\mathbf{u}_1 \neq \mathbf{0}$ . 取  $\mathbf{u} = \frac{f(\mathbf{u}_1)\mathbf{u}_1}{\|\mathbf{u}_1\|^2}$ .

对任何  $\mathbf{v} \in V$ , 存在  $\mathbf{w} \in \text{Ker}(f)$  和数  $\lambda$  使得

$$\mathbf{v} = \mathbf{w} + \lambda\mathbf{u}_1.$$

则

$$\langle \mathbf{v}, \mathbf{u}_1 \rangle = \lambda\|\mathbf{u}_1\|^2,$$

得到

$$\lambda = \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2}.$$

于是

$$f(\mathbf{v}) = f(\mathbf{w} + \lambda\mathbf{u}_1) = f(\mathbf{w}) + \lambda f(\mathbf{u}_1) = \lambda f(\mathbf{u}_1) = \frac{\langle \mathbf{v}, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2} f(\mathbf{u}_1) = \langle \mathbf{v}, \mathbf{u} \rangle.$$

这样证明了向量  $\mathbf{u}$  的存在性.

下面证明唯一性. 若  $\mathbf{u}, \tilde{\mathbf{u}}$  都满足

$$f(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, \tilde{\mathbf{u}} \rangle, \quad \forall \mathbf{v} \in V.$$

则

$$0 = f(\mathbf{u} - \tilde{\mathbf{u}}) - f(\mathbf{u} - \tilde{\mathbf{u}}) = \langle \mathbf{u} - \tilde{\mathbf{u}}, \mathbf{u} \rangle - \langle \mathbf{u} - \tilde{\mathbf{u}}, \tilde{\mathbf{u}} \rangle = \|\mathbf{u} - \tilde{\mathbf{u}}\|^2,$$

因此  $\mathbf{u} = \tilde{\mathbf{u}}$ . □



Riesz 表示定理告诉我们，内积空间上所有线性函数都可以通过内积运算得到.

定义  $\Phi: V^* \rightarrow V$ ,  $\Phi(f) = \mathbf{u}$  (向量  $\mathbf{u}$  如 Riesz 表示定理结论所说). 易见  $\Phi$  是单射, 又  $\dim V^* = \dim V$ , 所以根据推论 10.3,  $\Phi: V^* \rightarrow V$  是可逆的. 因此在映射  $\Phi$  下,  $V^*$  可视同为  $V$ .

#### 内积空间上的对偶积

若  $\alpha_1, \dots, \alpha_n$  是内积空间  $V$  的一组单位正交基, 则

$$\langle \alpha_i, \cdot \rangle, \quad i = 1, 2, \dots, n$$

是  $\alpha_1, \dots, \alpha_n$  的对偶基. 我们把线性函数  $\langle \alpha_i, \cdot \rangle$  视同为向量  $\alpha_i$ , 则单位正交基  $\alpha_1, \dots, \alpha_n$  可以看成是自身的对偶基.

**问题 26.2.** 设  $\beta_1, \dots, \beta_n$  是内积空间  $V$  的一组基, 但未必是单位或正交的. 求它们的对偶基.

#### 内积空间上的对偶映射

设  $A: V \rightarrow W$  是内积空间  $V$  到内积空间  $W$  的线性映射.

任给  $\mathbf{w} \in W$ ,  $\langle \mathbf{w}, \cdot \rangle_W$  是  $W$  上的线性函数,  $\langle \mathbf{w}, A\mathbf{v} \rangle_W$  看成关于  $\mathbf{v} \in V$  的线性函数. 按 Riesz 表示定理, 存在唯一的向量  $\mathbf{u} \in V$  使得

$$\langle \mathbf{w}, A\mathbf{v} \rangle_W = \langle \mathbf{u}, \mathbf{v} \rangle_V, \quad \forall \mathbf{v} \in V.$$

易见  $\mathbf{w} \mapsto \mathbf{u}$  是线性的, 按对偶映射的定义,  $A^*\mathbf{w} = \mathbf{u}$ , 从而

$$\langle \mathbf{w}, A\mathbf{v} \rangle_W = \langle A^*\mathbf{w}, \mathbf{v} \rangle_V, \quad \forall \mathbf{v} \in V.$$

这是内积情形下, 对偶映射 (adjoint map) 的定义.

给定  $V, W$  的基  $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m$ ,  $A$  的表示矩阵为  $(a_{ij})$ , 即

$$(A\alpha_1, \dots, A\alpha_n) = (\beta_1, \dots, \beta_m)(a_{ij})_{m \times n}.$$

$A^*: W \rightarrow V$  的表示矩阵为  $(b_{ij})$ , 即

$$(A^*\beta_1, \dots, A^*\beta_m) = (\alpha_1, \dots, \alpha_n)(b_{ij})_{n \times m}.$$

$$\langle \beta_i, A\alpha_j \rangle_W = \langle \beta_i, \sum_k a_{kj}\beta_k \rangle_W = \sum_k a_{kj}\langle \beta_i, \beta_k \rangle_W,$$

另一方面,

$$\langle A^*\beta_i, \alpha_j \rangle_V = \langle \sum_l b_{li}\alpha_l, \alpha_j \rangle_V = \sum_l b_{li}\langle \alpha_l, \alpha_j \rangle_V,$$

所以

$$(\langle \beta_i, \beta_k \rangle_W)(a_{kj}) = (b_{li})^T(\langle \alpha_l, \alpha_j \rangle_V).$$

所以

$$(a_{kj}) = (\langle \beta_i, \beta_k \rangle_W)^{-1} (b_{li})^T (\langle \alpha_l, \alpha_j \rangle_V).$$

当  $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m$  都是单位正交基时,

$$(a_{kj}) = (b_{jk})^T,$$

即  $A$  和  $A^*$  的表示矩阵互为对方的转置, 一个矩阵的每个行向量依次是另一个矩阵相应的列向量.

当  $W = V$  时, 称  $A$  是对称的 (self-adjoint), 如果  $A^* = A$ , 即

$$\langle w, Av \rangle = \langle Aw, v \rangle, \quad \forall v, w \in V.$$

称  $A$  是反对称的, 如果  $A^* = -A$ , 即

$$\langle w, Av \rangle = -\langle Aw, v \rangle, \quad \forall v, w \in V.$$

$A: V \rightarrow V$  可唯一分解为对称和反对称映射的和

$$A = \frac{A + A^*}{2} + \frac{A - A^*}{2}.$$

## 27 双线性型与二次型

**定义 27.1.** 双线性型 (bilinear form)  $B(v, w)$  是关于一对向量  $v, w$  的映射, 这对向量可以分别来自不同的向量空间  $V, W$ , 也可以来自同一向量空间, 满足:

1. 对任意  $w$ ,  $B(v, w)$  关于  $v$  是线性的;
2. 对任意  $v$ ,  $B(v, w)$  关于  $w$  是线性的.

在本课程中, 我们只关注  $W = V$  的情形, 这时我们称  $B(v, w)$  是  $V$  上的一个双线性型. 我们称  $B(v, w)$  是对称的, 如果对任意  $v, w$ , 都有  $B(v, w) = B(w, v)$ . 称  $B(v, w)$  是反对称的, 如果对任意  $v, w$ , 都有  $B(v, w) = -B(w, v)$ .

任何双线性型可以分解为一对对称双线性型和反对称双线性型的和:

$$\begin{aligned} B_s(v, w) &= \frac{B(v, w) + B(w, v)}{2}, \\ B_a(v, w) &= \frac{B(v, w) - B(w, v)}{2}, \\ B(v, w) &= B_s(v, w) + B_a(v, w). \end{aligned}$$

双线性型的坐标形式

给定  $V$  的一组基:  $\alpha_1, \dots, \alpha_n$ . 对  $v = \sum_{i=1}^n x_i \alpha_i$ ,  $w = \sum_{j=1}^n y_j \alpha_j$ ,

$$B(v, w) = \sum_{i,j} x_i y_j B(\alpha_i, \alpha_j) = \mathbf{x}^T (B(\alpha_i, \alpha_j))_{n \times n} \mathbf{y}.$$

我们称  $(B(\alpha_i, \alpha_j))_{n \times n}$  为双线性型  $B$  在基  $\alpha_1, \dots, \alpha_n$  下的表示矩阵.

如果另外再给  $V$  的一组基:  $\beta_1, \dots, \beta_n$ . 记

$$(\beta_1, \dots, \beta_n) = (\alpha_1, \dots, \alpha_n)P,$$

则

$$v = \sum_{i=1}^n \tilde{x}_i \beta_i = (\beta_1, \dots, \beta_n) \tilde{\mathbf{x}} = (\alpha_1, \dots, \alpha_n) P \tilde{\mathbf{x}}, \quad \mathbf{x} = P \tilde{\mathbf{x}}.$$

此时,

$$\begin{aligned} B(v, w) &= \tilde{\mathbf{x}}^T (B(\beta_i, \beta_j))_{n \times n} \tilde{\mathbf{y}} = \mathbf{x}^T (B(\alpha_i, \alpha_j))_{n \times n} \mathbf{y} \\ &= \tilde{\mathbf{x}}^T P^T (B(\alpha_i, \alpha_j))_{n \times n} P \tilde{\mathbf{y}}. \end{aligned}$$

所以

$$(B(\beta_i, \beta_j))_{n \times n} = P^T (B(\alpha_i, \alpha_j))_{n \times n} P.$$

对  $n$  阶方阵  $A$  和可逆矩阵  $P$ , 称  $P^T A P$  是矩阵  $A$  的一个合同变换, 也称两个矩阵  $A, P^T A P$  合同等价 (或是相合的).

同一个双线性型在不同坐标系下的表示矩阵是合同等价的.

## 二次型

设  $B(v, w)$  是向量空间  $V$  上的一个双线性型. 称  $Q(v) = B(v, v)$  是  $V$  上的一个二次型 (quadratic form).

易见, 二次型只与双线性型  $B$  的对称部分有关

$$Q(v) = B(v, v) = B_s(v, v) + B_a(v, v) = B_s(v, v).$$

二次型  $Q$  在基  $\alpha_1, \dots, \alpha_n$  下的表示矩阵为  $(B_s(\alpha_i, \alpha_j))_{n \times n}$ ,

$$Q(v) = \mathbf{x}^T (B_s(\alpha_i, \alpha_j))_{n \times n} \mathbf{x}.$$

**例 27.2.** 把二次齐次多项式  $x^2 - 2y^2, xy$  分别写成二次型.

解.

$$x^2 - 2y^2 = (x, y) \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad xy = (x, y) \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

□

## 28 二次型的标准型, 对称矩阵的谱分解

$\mathbb{R}^n$  上的二次型

$$\mathbf{x}^T A \mathbf{x}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad A = (a_{ij})_{n \times n}, \quad a_{ji} = a_{ij}.$$

$f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$  是关于  $x_1, \dots, x_n$  的连续函数, 它在有界闭集

$$K = \{\mathbf{x} | x_1^2 + \dots + x_n^2 = 1\}$$

上存在最大值  $f(\mathbf{x}_0)$ .

任取与  $\mathbf{x}_0$  正交的单位向量  $\mathbf{v}$ , 考虑

$$g(\theta) = f(\mathbf{x}_0 \cos \theta + \mathbf{v} \sin \theta) = (\mathbf{x}_0 \cos \theta + \mathbf{v} \sin \theta)^T A (\mathbf{x}_0 \cos \theta + \mathbf{v} \sin \theta).$$

则  $g$  关于  $\theta$  可微,  $g(0)$  是最大值, 从而  $g'(0) = 0$ . 注意到

$$\begin{aligned} g'(\theta) &= (-\mathbf{x}_0 \sin \theta + \mathbf{v} \cos \theta)^T A (\mathbf{x}_0 \cos \theta + \mathbf{v} \sin \theta) \\ &\quad + (\mathbf{x}_0 \cos \theta + \mathbf{v} \sin \theta)^T A (-\mathbf{x}_0 \sin \theta + \mathbf{v} \cos \theta), \end{aligned}$$

所以

$$0 = g'(0) = 2\mathbf{v}^T A \mathbf{x}_0 = 2\mathbf{x}_0^T A \mathbf{v}, \quad (*)$$

这表明  $A\mathbf{x}_0$  与 (所有与  $\mathbf{x}_0$  正交的单位向量)  $\mathbf{v}$  正交, 因此存在数  $\lambda$  使得  $A\mathbf{x}_0 = \lambda\mathbf{x}_0$ . 同时, (\*) 也表明  $A\mathbf{v}$  仍与  $\mathbf{x}_0$  正交, 所以

$$(\text{span}\{\mathbf{x}_0\})^\perp = \{\mathbf{w} \in \mathbb{R}^n | \mathbf{w} \perp \mathbf{x}_0\}$$

是  $A$ -不变子空间. 取  $\mathbb{R}^n$  的一组单位正交基  $\mathbf{x}_0, \beta_2, \dots, \beta_n$ , 则  $\beta_2, \dots, \beta_n$  是  $(\text{span}\{\mathbf{x}_0\})^\perp$  的一组单位正交基.  $A$  在这组基下的表示矩阵为

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{bmatrix},$$

它仍是对称矩阵. 再对  $A_1$  做同样讨论, 结合数学归纳法, 可以证明

**定理 28.1 (谱分解).** 对任何实对称矩阵  $A$ , 存在  $\mathbb{R}^n$  的一组单位正交基  $\alpha_1, \dots, \alpha_n$  以及一组实数

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$$

满足  $A\alpha_k = \lambda_k \alpha_k$ ,  $k = 1, 2, \dots, n$  (称这样的  $\alpha_k$  为  $A$  的特征向量, 称相应的  $\lambda_k$  为  $A$  的特征值). 此时,

$$A = \sum_{k=1}^n \lambda_k \alpha_k \alpha_k^T.$$

**证明.** 满足要求的单位正交基  $\alpha_1, \dots, \alpha_n$  的存在性已在上面完成.

对任意  $v$ ,

$$\begin{aligned} Av &= A(\langle \alpha_1, v \rangle \alpha_1 + \dots + \langle \alpha_n, v \rangle \alpha_n) = \sum_{k=1}^n \langle \alpha_k, v \rangle A \alpha_k \\ &= \sum_{k=1}^n \lambda_k \langle \alpha_k, v \rangle \alpha_k = \sum_{k=1}^n \lambda_k \alpha_k \alpha_k^T v. \end{aligned}$$

所以

$$A = \sum_{k=1}^n \lambda_k \alpha_k \alpha_k^T.$$

□

**推论 28.2.** 对任何实对称矩阵  $A$ , 存在正交矩阵  $P$  使得  $P^T A P$  是对角矩阵.

任何实二次型  $\mathbf{x}^T A \mathbf{x}$  都可经正交坐标变换  $\mathbf{x} = P \mathbf{y}$  变为  $\sum_{k=1}^n \lambda_k y_k^2$  的形式.

**证明.**  $P$  的列向量由谱分解定理中的单位正交基构成. □

**推论 28.3 (二次型的标准型).** 任何实二次型  $\mathbf{x}^T A \mathbf{x}$  都可经可逆坐标变换  $\mathbf{x} = T \mathbf{z}$  变为  $\sum_{k=1}^{n_+} z_k^2 - \sum_{k=n_++1}^{n_++n_-} z_k^2$  的形式.

**证明.** 取

$$z_k = \begin{cases} \sqrt{|\lambda_k|} y_k, & \text{若 } \lambda_k \neq 0, \\ y_k, & \text{若 } \lambda_k = 0. \end{cases}$$

$n_+, n_-$  分别是  $\lambda_1, \dots, \lambda_n$  中正数和负数的个数, 称为  $A$  的正惯性指数和负惯性指数. □

**例 28.4.** 把  $x_1^2 + x_2^2 + x_3^2 + 4x_1x_2 + 4x_2x_3 + 4x_3x_1$  化为标准形.

**解.**

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

我们按谱定理定理证明中的方法求  $A$  的特征值和特征向量.

当  $x_1^2 + x_2^2 + x_3^2 = 1$  时,

$$\mathbf{x}^T A \mathbf{x} = 1 + 4x_1x_2 + 4x_2x_3 + 4x_3x_1 \leq 1 + 2(x_1^2 + x_2^2) + 2(x_2^2 + x_3^2) + 2(x_3^2 + x_1^2) = 5,$$

等号成立当且仅当  $x_1 = x_2 = x_3 = \frac{\pm 1}{\sqrt{3}}$ . 所以

$$\alpha_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \quad A \alpha_1 = 5 \alpha_1.$$

$\mathbf{x} \perp \boldsymbol{\alpha}_1$ , 当且仅当  $x_1 + x_2 + x_3 = 0$ , 此时  $1 = x_1^2 + x_2^2 + x_3^2 = x_1^2 + x_2^2 + (-x_1 - x_2)^2 = 2(x_1^2 + x_1x_2 + x_2^2)$ .

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = 1 + 4(x_1 + x_2)(-x_1 - x_2) + 4x_1x_2 = -2(x_1^2 + x_1x_2 + x_2^2) = -1.$$

任取与  $\boldsymbol{\alpha}_1$  正交的两个单位正交向量

$$\boldsymbol{\alpha}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{bmatrix}, \boldsymbol{\alpha}_3 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad A\boldsymbol{\alpha}_k = -\boldsymbol{\alpha}_k, k = 2, 3.$$

取  $P = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$ ,  $\mathbf{y} = P^T \mathbf{x}$ , 则

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = 5y_1^2 - y_2^2 - y_3^2.$$

进一步再做坐标变换

$$z_1 = \sqrt{5}y_1, \quad z_2 = y_2, \quad z_3 = y_3,$$

得到标准形

$$z_1^2 - z_2^2 - z_3^2.$$

□

另解.

$$\begin{cases} \max & x_1^2 + x_2^2 + x_3^2 + 4x_1x_2 + 4x_2x_3 + 4x_3x_1, \\ \text{s.t.} & x_1^2 + x_2^2 + x_3^2 = 1. \end{cases}$$

多元函数求条件极值, 可用 Lagrange 乘子法,

$$F(x_1, x_2, x_3, \lambda) = 1 + 4x_1x_2 + 4x_2x_3 + 4x_3x_1 - \lambda(x_1^2 + x_2^2 + x_3^2 - 1).$$

则

$$\begin{cases} \frac{\partial F}{\partial x_1} = 4x_2 + 4x_3 - 2\lambda x_1 = 0, \\ \frac{\partial F}{\partial x_2} = 4x_3 + 4x_1 - 2\lambda x_2 = 0, \\ \frac{\partial F}{\partial x_3} = 4x_1 + 4x_2 - 2\lambda x_3 = 0, \\ \frac{\partial F}{\partial \lambda} = 1 - x_1^2 - x_2^2 - x_3^2 = 0. \end{cases}$$

这相当于求前面三个方程 (看成关于  $x_1, x_2, x_3$  的线性方程组) 的非零解.

$$\begin{bmatrix} -\lambda & 2 & 2 \\ 2 & -\lambda & 2 \\ 2 & 2 & -\lambda \end{bmatrix} \xrightarrow{\text{行变换}} \begin{bmatrix} 4-\lambda & 4-\lambda & 4-\lambda \\ 2 & -\lambda & 2 \\ 2 & 2 & -\lambda \end{bmatrix}$$

$\lambda = 4$  时, 上述矩阵进一步行变换为

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

从而得到非零解  $x_2 = x_3, x_1 = 2x_2 - x_3 = x_3$ , 再由第四个方程求得  $x_1 = x_2 = x_3 = \frac{\pm 1}{\sqrt{3}}$ .

当  $\lambda \neq 4$  时, 系数矩阵行变换得到

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & -\lambda - 2 & 0 \\ 0 & 0 & -\lambda - 2 \end{bmatrix},$$

线性方程组有非零解当且仅当  $\lambda = -2$ , 此时的非零解满足  $x_1 + x_2 + x_3 = 0$ , 任取两个满足该条件的单位正交向量, 比如

$$\begin{cases} x_1 = \frac{1}{\sqrt{2}}, \\ x_2 = 0, \\ x_3 = -\frac{1}{\sqrt{2}}, \end{cases} \quad \begin{cases} x_1 = \frac{1}{\sqrt{6}}, \\ x_2 = \frac{-2}{\sqrt{6}}, \\ x_3 = \frac{1}{\sqrt{6}}. \end{cases}$$

这样就得了满足条件的三个单位正交向量, 并由它们组成正交矩阵  $P$ , 进而把二次型化为标准形.  $\square$

## 29 二次型、二次曲线和二次曲面的分类

**定义 29.1.** 称二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  (或对称矩阵  $A$ ) 是**正定的**, 如果对任意  $\mathbf{v}$ ,  $\mathbf{v}^T A \mathbf{v} \geq 0$ , 且  $\mathbf{v}^T A \mathbf{v} = 0$  成立当且仅当  $\mathbf{v} = \mathbf{0}$ .

称二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  (或对称矩阵  $A$ ) 是**半正定的**, 如果对任意  $\mathbf{v}$ ,  $\mathbf{v}^T A \mathbf{v} \geq 0$ .

称二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  (或对称矩阵  $A$ ) 是**负定的** (或**半负定的**), 如果  $-A$  是正定的 (相应地, 半正定的).

称二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  (或对称矩阵  $A$ ) 是**不定的**, 如果它既不是半正定的, 也不是半负定的.

称二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  (或对称矩阵  $A$ ) 是**鞍型的**, 如果它是不定的, 且  $A$  是可逆的.

**定理 29.2.** 1. 二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  是正定的 / 半正定的 / 负定的 / 半负定的, 当且仅当  $A$  的特征值都是正数 / 非负数 / 负数 / 非正数.

2. 二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  是不定的, 当且仅当  $A$  既有正特征值, 又有负特征值.

3. 二次型  $Q(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  是鞍型的, 当且仅当  $A$  既有正特征值, 又有负特征值, 但没有零特征值.

**证明.** 由谱分解定理

$$A = \sum_{k=1}^n \lambda_k \boldsymbol{\alpha}_k \boldsymbol{\alpha}_k^T,$$

所以

$$\mathbf{v}^T A \mathbf{v} = \sum_{k=1}^n \lambda_k \mathbf{v}^T \boldsymbol{\alpha}_k \boldsymbol{\alpha}_k^T \mathbf{v} = \sum_{k=1}^n \lambda_k (\boldsymbol{\alpha}_k^T \mathbf{v})^2.$$

因此 (1) 的充分性得证.

取  $\mathbf{v} = \boldsymbol{\alpha}_i$ , 得到

$$\lambda_i = \boldsymbol{\alpha}_i^T A \boldsymbol{\alpha}_i.$$

所以 (1) 的必要性得证.

由 (1) 得到 (2).

另外易知,  $A$  可逆, 当且仅当  $A\mathbf{v} = \mathbf{0}$  只有零解, 后者当且仅当  $A$  没有零特征值. 所以由 (2) 得到 (3).  $\square$

**定义 29.3.** 形如

$$\mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + c = 0$$

的方程确定的几何对象称为二次曲线 ( $n = 2$  时) 或二次曲面 ( $n > 2$  时).

**平面二次曲线的分类**

先做正交变换使曲线方程变成

$$\lambda_1 x^2 + \lambda_2 y^2 + b_1 x + b_2 y + c = 0.$$

(1) 如果  $A$  可逆, 则  $\lambda_1, \lambda_2$  都非零, 此时通过平移消去一次项

$$\lambda_1 \left( x + \frac{b_1}{2\lambda_1} \right)^2 + \lambda_2 \left( y + \frac{b_2}{2\lambda_2} \right)^2 = \frac{b_1^2}{4\lambda_1} + \frac{b_2^2}{4\lambda_2} - c.$$

(1.1.1) 如果  $\lambda_1, \lambda_2$  同号 ( $A$  正定或负定) 但与上式等号右端常数异号, 则方程的解集是空集.

(1.1.2) 如果  $\lambda_1, \lambda_2$  同号 ( $A$  正定或负定), 且上式等号右端常数为零, 则方程的解集是单点.

(1.1.3) 如果  $\lambda_1, \lambda_2$  同号 ( $A$  正定或负定) 且与上式等号右端常数同号, 则方程的解集是椭圆, 特别当  $\lambda_1 = \lambda_2$  时, 它是圆周.

(1.2.1) 如果  $\lambda_1, \lambda_2$  异号 ( $A$  不定), 且上式等号右端常数为零, 则方程的解集是一对相交直线.

(1.2.2) 如果  $\lambda_1, \lambda_2$  异号 ( $A$  不定), 且上式等号右端常数非零, 则方程的解集是一对双曲线.



(2) 如果  $A$  不可逆, 则不妨设  $\lambda_1 > 0, \lambda_2 = 0$ ,

$$\lambda_1 \left( x + \frac{b_1}{2\lambda_1} \right)^2 + b_2 y = \frac{b_1^2}{4\lambda_1} - c.$$

(2.1) 如果  $b_2 \neq 0$ , 则曲线是抛物线.

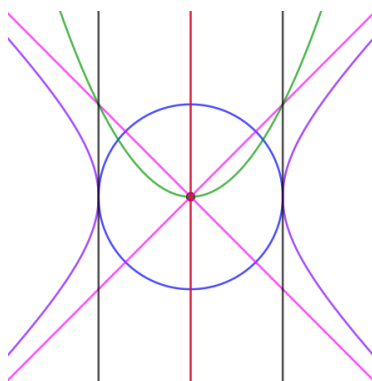
(2.2) 如果  $b_2 = 0$ , 方程右端常数为负数, 则方程解集为空集.

(2.3) 如果  $b_2 = 0$ , 方程右端常数为零, 则方程解集为一条直线 (相当于一对平行直线的特殊情形).

(2.4) 如果  $b_2 = 0$ , 方程右端常数为正数, 则方程解集为一对平行直线.

在正交变换、平移、尺度变换下, 平面二次曲线的标准方程为

方程	曲线类型
$x^2 + y^2 = 1$	圆
$x^2 + y^2 = 0$	单点
$x^2 + y^2 = -1$	$\emptyset$
$x^2 - y^2 = 1$	双曲线
$x^2 - y^2 = 0$	一对相交直线
$x^2 - y = 0$	双曲线
$x^2 = 1$	一对平行直线
$x^2 = 0$	一对重合直线
$x^2 = -1$	$\emptyset$



### $\mathbb{R}^3$ 中二次曲面的分类

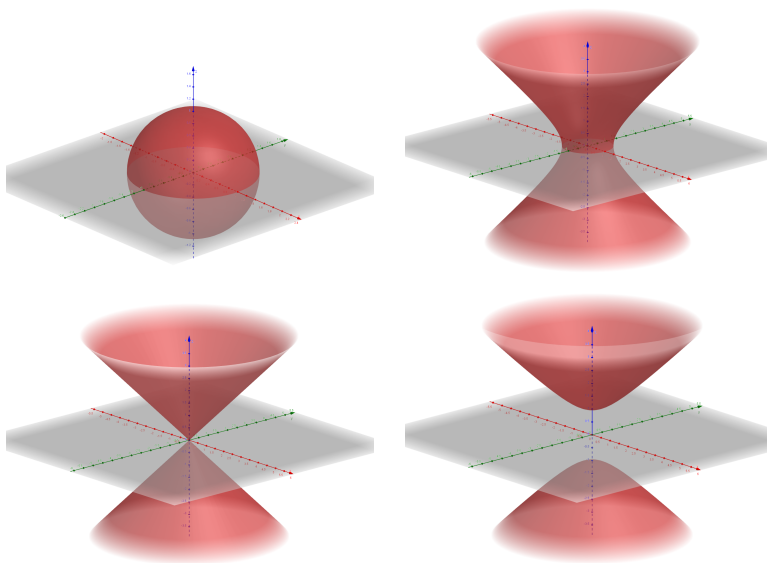
先做正交变换使曲线方程变成

$$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + b_1 x + b_2 y + b_3 z + c = 0.$$

(1) 当  $A$  可逆时,  $\lambda_1, \lambda_2, \lambda_3$  都非零. 此时通过平移可以消去一次项.

$$\lambda_1 \left( x + \frac{b_1}{2\lambda_1} \right)^2 + \lambda_2 \left( y + \frac{b_2}{2\lambda_2} \right)^2 + \lambda_3 \left( z + \frac{b_3}{2\lambda_3} \right)^2 = \frac{b_1^2}{4\lambda_1} + \frac{b_2^2}{4\lambda_2} + \frac{b_3^2}{4\lambda_3} - c$$

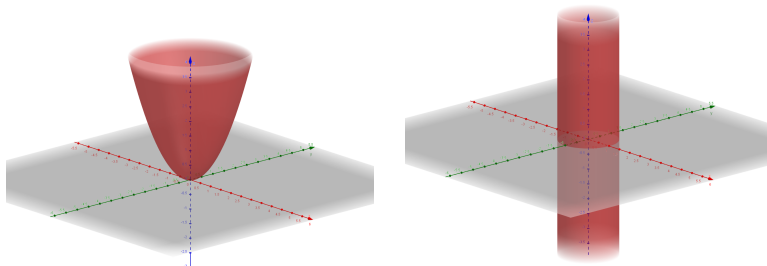
- (1.1.1) 如果  $\lambda_1, \lambda_2, \lambda_3$  以及右端常数都同号, 则曲面为椭球面.  
 (1.1.2) 如果  $\lambda_1, \lambda_2, \lambda_3$  同号, 且右端常数为零, 则方程解集为单点.  
 (1.1.3) 如果  $\lambda_1, \lambda_2, \lambda_3$  同号, 但与右端常数异号, 则方程解集为空集.  
 (1.2) 不妨设  $\lambda_1, \lambda_2$  为正数,  $\lambda_3 < 0$ .  
 (1.2.1) 若方程等式右端常数为正数, 则曲面为单叶双曲面.  
 (1.2.2) 若方程等式右端常数为零, 则曲面为一个锥面.  
 (1.2.3) 若方程等式右端常数为负数, 则曲面为一个双叶双曲面.



- (2)  $\lambda_1, \lambda_2 > 0, \lambda_3 = 0$ .

$$\lambda_1 \left( x + \frac{b_1}{2\lambda_1} \right)^2 + \lambda_2 \left( y + \frac{b_2}{2\lambda_2} \right)^2 + b_3 z = \frac{b_1^2}{4\lambda_1} + \frac{b_2^2}{4\lambda_2} - c$$

- (2.1) 若  $b_3 \neq 0$ , 则曲面是椭圆抛物面.  
 (2.2.1) 若  $b_3 = 0$ , 且方程右端常数为正数, 则曲面为圆柱面.  
 (2.2.2) 若  $b_3 = 0$ , 且方程右端常数为零, 则方程的解集为一条直线 (圆柱面的退化情形).  
 (2.2.3) 若  $b_3 = 0$ , 且方程右端常数为负数, 则方程的解集为空集.



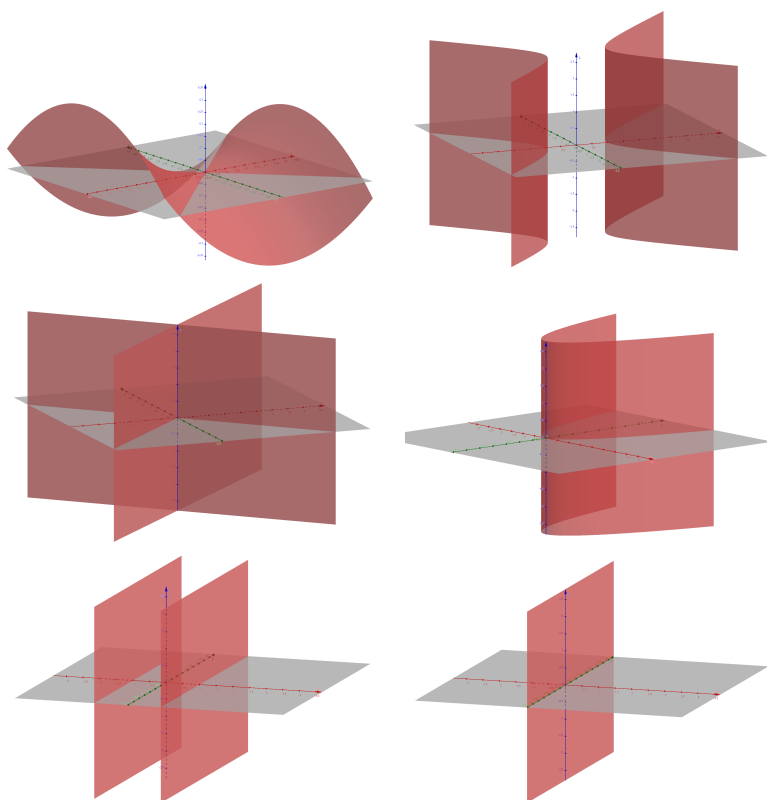
(3)  $\lambda_1 > 0 > \lambda_2 > 0, \lambda_3 = 0$ .

$$\lambda_1 \left( x + \frac{b_1}{2\lambda_1} \right)^2 + \lambda_2 \left( y + \frac{b_2}{2\lambda_2} \right)^2 + b_3 z = \frac{b_1^2}{4\lambda_1} + \frac{b_2^2}{4\lambda_2} - c$$

(3.1) 若  $b_3 \neq 0$ , 则曲面为双曲抛物面 (也称为“马鞍面”).

(3.2.1) 若  $b_3 = 0$ , 且方程右端常数非零, 则曲面为双曲柱面.

(3.2.2) 若  $b_3 = 0$ , 且方程右端常数为零, 则曲面为一对相交平面.



(4)  $\lambda_1 > 0 = \lambda_2 = \lambda_3$ .

$$\lambda_1 \left( x + \frac{b_1}{2\lambda_1} \right)^2 + b_2 y + b_3 z = \frac{b_1^2}{4\lambda_1} - c$$

进一步在  $y, z$  平面里旋转, 方程变成

$$\lambda_1 \left( x + \frac{b_1}{2\lambda_1} \right)^2 + \sqrt{b_2^2 + b_3^2} y = \frac{b_1^2}{4\lambda_1} - c$$

(4.1) 若  $b_2, b_3$  不全为零, 则曲面为抛物柱面.

(4.2.1) 若  $b_2 = b_3 = 0$ , 若方程右端常数为正数, 则曲面是一对平行平面.

(4.2.2) 若  $b_2 = b_3 = 0$ , 若方程右端常数为零, 则曲面是一对重合平面.

(4.2.3) 若  $b_2 = b_3 = 0$ , 若方程右端常数为负数, 则方程解集为空集.

在正交变换、平移、尺度变换下,  $\mathbb{R}^3$  中的二次曲面的标准方程为

方程	曲面类型
$x^2 + y^2 + z^2 = 1$	球
$x^2 + y^2 + z^2 = 0$	单点
$x^2 + y^2 + z^2 = -1$	$\emptyset$
$x^2 + y^2 - z^2 = 1$	单叶双曲面
$x^2 + y^2 - z^2 = 0$	锥面
$x^2 + y^2 - z^2 = -1$	双叶双曲面
$x^2 + y^2 = z$	椭圆抛物面
$x^2 + y^2 = 1$	椭圆柱面
$x^2 + y^2 = 0$	一条直线
$x^2 + y^2 = -1$	$\emptyset$
$x^2 - y^2 = z$	双曲抛物面
$x^2 - y^2 = 1$	双曲柱面
$x^2 - y^2 = 0$	一对相交平面
$x^2 = y$	抛物柱面
$x^2 = 1$	一对平行平面
$x^2 = 0$	一对重合平面
$x^2 = -1$	$\emptyset$

## 30 奇异值分解

### 奇异值与左右奇异向量

任给矩阵  $A$  ( $m$  行  $n$  列, 秩为  $r$ ), 则矩阵  $A^T A$  是半正定的对称矩阵, 它有一组特征向量  $\mathbf{v}_1, \dots, \mathbf{v}_n$  构成  $\mathbb{R}^n$  的一组标准正交基, 相应的特征值

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0 = \lambda_{r+1} = \dots = \lambda_n.$$

因此由谱分解定理知

$$A^T A = \sum_{k=1}^n \lambda_k \mathbf{v}_k \mathbf{v}_k^T,$$

于是  $A^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i$ ,  $\|A \mathbf{v}_i\| = \sqrt{\lambda_i}$  ( $\forall i$ ), 且  $A \mathbf{v}_1, \dots, A \mathbf{v}_n$  彼此正交:

$$(A \mathbf{v}_i)^T A \mathbf{v}_j = \mathbf{v}_i^T A^T A \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j = \begin{cases} \lambda_j > 0, & i = j \leq r; \\ 0, & i \neq j \text{ 或 } i = j > r. \end{cases}$$

取

$$\sigma_i = \|A \mathbf{v}_i\|, \quad \mathbf{u}_i = \frac{A \mathbf{v}_i}{\|A \mathbf{v}_i\|} = \frac{1}{\sigma_i} A \mathbf{v}_i, \quad i = 1, 2, \dots, r.$$

所以  $\mathbf{u}_1, \dots, \mathbf{u}_r$  是单位正交向量, 我们把它补齐为  $\mathbb{R}^m$  的一组标准正交基

$$\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_m.$$

因此  $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$  是  $\text{Ker}(A)$  的一组标准正交基,  $\mathbf{u}_1, \dots, \mathbf{u}_r$  是  $\text{Range}(A)$  的一组标准正交基, 且对  $i = 1, 2, \dots, r$ ,

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i,$$

$$\mathbf{u}_i^T A = \frac{1}{\sigma_i} \mathbf{v}_i^T A^T A = \frac{1}{\sigma_i} \mathbf{v}_i^T \sum_{k=1}^n \lambda_k \mathbf{v}_k \mathbf{v}_k^T = \sum_{k=1}^n \frac{\lambda_k}{\sigma_i} (\mathbf{v}_i^T \mathbf{v}_k) \mathbf{v}_k^T = \frac{\lambda_i}{\sigma_i} \mathbf{v}_i^T = \sigma_i \mathbf{v}_i^T.$$

称  $\mathbf{v}_1, \dots, \mathbf{v}_r$  为  $A$  的右奇异向量,  $\mathbf{u}_1, \dots, \mathbf{u}_r$  为  $A$  的左奇异向量,  $\sigma_1, \dots, \sigma_r$  为  $A$  的奇异值.

奇异值分解

$$A\mathbf{v} = A \left( \sum_{i=1}^n \langle \mathbf{v}_i, \mathbf{v} \rangle \right) \mathbf{v}_i = \sum_{i=1}^n \langle \mathbf{v}_i, \mathbf{v} \rangle A\mathbf{v}_i = \sum_{i=1}^r \langle \mathbf{v}_i, \mathbf{v} \rangle A\mathbf{v}_i = \sum_{i=1}^r \sigma_i \langle \mathbf{v}_i, \mathbf{v} \rangle \mathbf{u}_i,$$

所以

$$A\mathbf{v} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \mathbf{v},$$

因此

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T = U \Sigma V^T, \quad (9)$$

其中  $U$  是以  $\mathbf{u}_1, \dots, \mathbf{u}_r$  为列向量的矩阵,  $V$  是以  $\mathbf{v}_1, \dots, \mathbf{v}_r$  为列向量的矩阵,  $\Sigma$  是以  $\sigma_1, \dots, \sigma_r$  为对角元的对角矩阵. 等式 (9) 称为矩阵  $A$  的奇异值分解.

由 (9) 得到

$$A^T = \sum_{i=1}^r \sigma_i \mathbf{v}_i \mathbf{u}_i^T.$$

这是矩阵  $A^T$  的奇异值分解.

矩阵的奇异值分解与四个重要的子空间

由上述奇异值分解可知:

$\mathbb{R}^n$	$\xrightarrow{A}$	$\mathbb{R}^m$
$\text{Ker}(A) = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$	$\xrightarrow{A}$	$0$
$\oplus$		
$\text{Range}(A^T) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$	$\xrightleftharpoons[A^T \text{ 线性同构}]{A \text{ 线性同构}}$	$\text{Range}(A) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$
		$\oplus$
$0$	$\xleftarrow{A^T}$	$\text{Ker}(A^T) = \text{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$
$\mathbb{R}^n$	$\xleftarrow{A^T}$	$\mathbb{R}^m$

这里需要指出,  $A$  限制在  $\text{Range}(A^T)$  上, 是到  $\text{Range}(A)$  的可逆线性映射, 其逆为

$$\sum_{i=1}^r \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^T,$$

这与  $A^T$  是不同的.

### 奇异值分解的应用

压缩存储: 原来存储一个  $m$  行  $n$  列的矩阵  $A$ , 需要占用  $mn$  个存储单元. 当  $A$  的秩  $r$  相对较小时, 利用奇异值分解, 只需  $(m+n+1)r$  个存储单元.

矩阵的低秩逼近: 两个同行列的矩阵  $A$  和  $B$ , 可以用  $\sqrt{\sum_{i,j} |a_{ij} - b_{ij}|^2}$  度量  $A, B$  之间的距离, 此时  $\sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$  是  $A$  的秩不超过  $k$  的最佳逼近. 当矩阵的奇异值差异较大时, 我们可以只计算较大奇异值部分, 这样可以事先信息压缩.

统计中的主成分分析也是利用奇异值分解发现数据之间的联系.

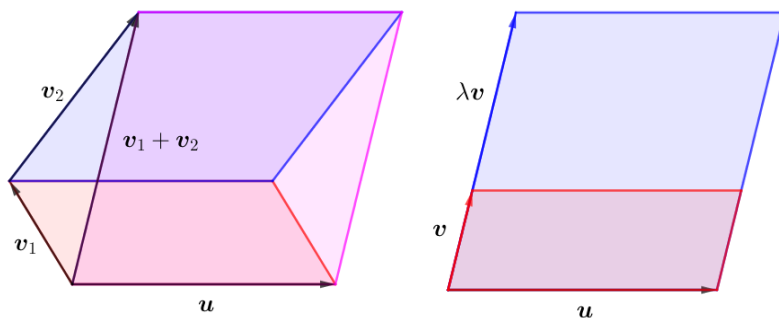
## 31 多重线性函数与行列式

记  $f(\mathbf{u}, \mathbf{v})$  表示以  $\mathbf{u}, \mathbf{v}$  为一组邻边的平行四边形的面积.

则  $f$  满足

$$f(\mathbf{u}, \mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{u}, \mathbf{v}_1) + f(\mathbf{u}, \mathbf{v}_2), \quad (10)$$

$$f(\mathbf{u}, \lambda \mathbf{v}) = \lambda f(\mathbf{u}, \mathbf{v}). \quad (11)$$



即当  $\mathbf{u}$  固定时,  $f(\mathbf{u}, \mathbf{v})$  关于  $\mathbf{v}$  具有线性. 同理, 当  $\mathbf{v}$  固定时,  $f(\mathbf{u}, \mathbf{v})$  关于  $\mathbf{u}$  也具有线性. 即  $f(\mathbf{u}, \mathbf{v})$  是双线性函数.

原本 (11) 只对正数  $\lambda$  成立, 但由于明显的连续性的需要, 它对  $\lambda = 0$  也成立. 又因为 (10), 所以

$$0 = f(\mathbf{u}, 0\mathbf{v}) = f(\mathbf{u}, \lambda\mathbf{v}) + f(\mathbf{u}, -\lambda\mathbf{v}),$$

因此

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

从而 (11) 只对所有实数  $\lambda$  都成立.

因此更确切地说,  $f(\mathbf{u}, \mathbf{v})$  是以  $\mathbf{u}, \mathbf{v}$  为一组邻边的平行四边形的有向面积. 稍后我们详细说明面积的方向.

由面积性质知

$$f(\mathbf{u}, \mathbf{u}) = 0, \quad (12)$$

由双线性, 我们得到

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

因此易见 (12) 与以下反对称性等价

$$f(\mathbf{v}, \mathbf{u}) = -f(\mathbf{u}, \mathbf{v}). \quad (13)$$

**定理 31.1.** 设  $f(\mathbf{u}, \mathbf{v})$  是一个反对称、双线性函数, 则对任何一对向量  $\alpha_1, \alpha_2$ , 都有

$$f(x_1\alpha_1 + y_1\alpha_2, x_2\alpha_1 + y_2\alpha_2) = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} f(\alpha_1, \alpha_2),$$

其中  $\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = x_1y_2 - x_2y_1$  是矩阵  $\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}$  的行列式.

**证明.**

$$\begin{aligned} & f(x_1\alpha_1 + y_1\alpha_2, x_2\alpha_1 + y_2\alpha_2) \\ &= x_1x_2f(\alpha_1, \alpha_1) + y_1y_2f(\alpha_2, \alpha_2) + x_1y_2f(\alpha_1, \alpha_2) + x_2y_1f(\alpha_2, \alpha_1) \\ &= x_1y_2f(\alpha_1, \alpha_2) - x_2y_1f(\alpha_1, \alpha_2) \\ &= \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} f(\alpha_1, \alpha_2), \end{aligned}$$

□

回到面积问题, 我们规定对  $\mathbb{R}^2$  的标准正交基  $\mathbf{e}_1, \mathbf{e}_2$ ,  $f(\mathbf{e}_1, \mathbf{e}_2) = 1$  作为面积单位. 此时由  $\mathbf{u} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ ,  $\mathbf{v} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$  形成的平行四边形的有向面积为  $\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}$ , 所谓有向面积的方向就是这个行列式的正、负.

类似地, 我们可以定义  $\mathbb{R}^n$  中以向量  $\mathbf{v}_1, \dots, \mathbf{v}_n$  为邻边的平行多面体的  $n$  维体积  $f(\mathbf{v}_1, \dots, \mathbf{v}_n)$  是一个  $n$  重线性、反对称的函数.

**定理 31.2.** 设  $f(\mathbf{v}_1, \dots, \mathbf{v}_n)$  是一个反对称、 $n$  重线性函数, 则对任何一对向量  $\alpha_1, \dots, \alpha_n$ , 都有

$$\begin{aligned} & f\left(\sum_{i_1} x_{i_1 1} \alpha_{i_1}, \sum_{i_2} x_{i_2 2} \alpha_{i_2}, \dots, \sum_{i_n} x_{i_n n} \alpha_{i_n}\right) \\ &= \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{vmatrix} f(\alpha_1, \alpha_2, \dots, \alpha_n), \end{aligned}$$

其中

$$\begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{vmatrix} = \sum_{\sigma \in S_n} (-1)^{\# \sigma} x_{\sigma(1)1} x_{\sigma(2)2} \cdots x_{\sigma(n)n}$$

是矩阵  $\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix}$  的行列式<sup>5</sup>. 这里  $S_n$  是  $1, 2, \dots, n$  的全部排列组成的集合,  $\# \sigma = \begin{cases} 0, & \text{若 } \sigma \text{ 可表示为偶数个对换的乘积;} \\ 1, & \text{若 } \sigma \text{ 可表示为奇数个对换的乘积.} \end{cases}$

证明.

$$\begin{aligned} & f\left(\sum_{i_1} x_{i_1 1} \alpha_{i_1}, \sum_{i_2} x_{i_2 2} \alpha_{i_2}, \cdots, \sum_{i_n} x_{i_n n} \alpha_{i_n}\right) \\ &= \sum_{i_1, i_2, \dots, i_n} x_{i_1 1} x_{i_2 2} \cdots x_{i_n n} f(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}) \quad (\text{多重线性}) \\ &= \sum_{i_1, i_2, \dots, i_n \text{ 全不相同}} x_{i_1 1} x_{i_2 2} \cdots x_{i_n n} f(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_n}) \quad (\text{反对称}) \\ &= \sum_{\sigma \in S_n} x_{\sigma(1)1} x_{\sigma(2)2} \cdots x_{\sigma(n)n} (-1)^{\# \sigma} f(\alpha_1, \alpha_2, \dots, \alpha_n) \quad (\text{反对称}) \\ &= \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{vmatrix} f(\alpha_1, \alpha_2, \dots, \alpha_n), \end{aligned}$$

□

**例 31.3.** 对任何三角矩阵  $A$ ,  $\det A$  等于  $A$  的全部对角元的乘积, 特别地,  $\det I_n = 1$ .

**例 31.4.** 对下三角分块矩阵  $A = \begin{bmatrix} A_1 & 0 \\ A_2 & A_3 \end{bmatrix}$ , 其中  $A_1, A_3$  是方阵, 有  $\det A = \det A_1 \det A_3$ .

**推论 31.5.** 对任何关于  $n$  阶方阵  $A$  的列向量多重线性、反对称函数  $f(A)$ , 都有  $f(A) = f(I_n) \det(A)$ .

<sup>5</sup>行列式最早是由 Leibniz 在 1693 年引入的, 而且是用这个组合形式. “行列式”这个词是 Gauss 于 1801 年引入的.



**证明.** 对任何关于  $n$  阶方阵  $A$  的列向量多重线性、反对称函数  $f(A)$ , 存在常数  $C$  使得  $f(A) = C \det(A)$ .

取  $A = I_n$  得到  $f(I_n) = C \det I_n = C$ . 所以  $f(A) = f(I_n) \det(A)$ .  $\square$

**问题 31.6.** 证明 (1)  $1, 2, \dots, n$  的任何排列可以写成若干个轮换

$$(i_1, i_2, \dots, i_{k-1}, i_k) \rightarrow (i_2, i_3, \dots, i_k, i_1)$$

的乘积;

(2) 轮换  $(1, 2, \dots, k-1, k) \rightarrow (2, 3, \dots, k, 1)$  可以写成若干个对换  $(i, j) \rightarrow (j, i)$

的乘积;

(3) 一个排列无论怎样表示为对换的乘积, 其中对换个数的奇偶性是不变的.

**定义 31.7.** 设  $e_1, \dots, e_n$  是  $\mathbb{R}^n$  的标准基. 称  $\mathbb{R}^n$  的基  $v_1, \dots, v_n$  是正向的, 如果  $e_1, \dots, e_n$  在  $v_1, \dots, v_n$  下的表示矩阵的行列式为正数.

## 32 行列式的性质与计算

**定理 32.1.** 对任何  $n$  阶方阵  $A$ , 都有  $\det(A^T) = \det A$ , 因此行列式用列刻画的所有形式也适用于行.

**证明.** 按行列式定义

$$\begin{aligned} \det(A^T) &= \sum_{\sigma \in S_n} (-1)^{\# \sigma} (A^T)_{\sigma(1)1} (A^T)_{\sigma(2)2} \cdots (A^T)_{\sigma(n)n} \\ &= \sum_{\sigma \in S_n} (-1)^{\# \sigma} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \\ &= \sum_{\tau \in S_n} (-1)^{\# \tau} a_{\tau(1)1} a_{\tau(2)2} \cdots a_{\tau(n)n} \quad (\tau = \sigma^{-1}, \quad \# \tau = \# \sigma) \\ &= \det A \end{aligned}$$

$\square$

**定理 32.2 (Cauchy 1815).** 对任何  $n$  阶方阵  $A, B$ , 都有  $\det(AB) = \det A \cdot \det B$ .

**证明.** 记  $A$  和  $B$  的列向量分别为  $\alpha_1, \alpha_2, \dots, \alpha_n, \beta_1, \beta_2, \dots, \beta_n$ . 令

$$f(B) = f(\beta_1, \beta_2, \dots, \beta_n) = \det(A\beta_1, A\beta_2, \dots, A\beta_n).$$

则  $f$  是关于  $B$  的列向量  $\beta_1, \beta_2, \dots, \beta_n$  的  $n$  重线性、反对称的函数,

$$f(I_n) = \det(Ae_1, Ae_2, \dots, Ae_n) = \det A,$$

所以由推论 31.5,  $\det(AB) = f(B) = f(I_n) \det B = \det A \cdot \det B$ .  $\square$

**推论 32.3.** 若  $A$  可逆, 则  $\det A \neq 0$ , 且  $\det(A^{-1}) = (\det A)^{-1}$ .

**证明.**

$$1 = \det I_n = \det(AA^{-1}) = \det A \cdot \det(A^{-1}),$$

所以  $\det A \neq 0$ , 且  $\det(A^{-1}) = (\det A)^{-1}$ . □

**推论 32.4.** 若  $A$  是正交矩阵, 则  $\det A = 1$  或  $-1$ .

**证明.**

$$1 = \det A \cdot \det(A^{-1}) = \det A \cdot \det(A^T) = (\det A)^2,$$

所以  $\det A \in \{1, -1\}$ . □

**推论 32.5.** 若  $A$  是实对称矩阵, 则  $\det A$  的值为  $A$  的全部特征值 (计重数) 的乘积.

**证明.** 根据谱分解定理, 存在正交矩阵  $P$  和对角矩阵  $\Lambda$  使得

$$A = P\Lambda P^T,$$

其中  $\Lambda$  的对角元是  $A$  的全部特征值. 于是

$$\det A = \det P \cdot \det \Lambda \cdot \det(P^T) = \det \Lambda \cdot (\det P)^2 = \det \Lambda.$$

□

**定理 32.6 (Laplace 1772).** 对任何  $n$  阶方阵  $A$ , 以及任何  $1 \leq j \leq n$ ,

$$\det A = a_{1j}A_{1j} + \cdots + a_{nj}A_{nj} = a_{i1}A_{i1} + \cdots + a_{in}A_{in},$$

其中  $A_{ij} = (-1)^{i+j}M_{ij}$  是矩阵  $A$  的代数余子式,  $M_{ij}$  是矩阵  $A$  的余子式, 即删除矩阵  $A$  第  $i$  行和第  $j$  列后剩余的  $n-1$  阶的行列式.

**证明.** 行列式按第  $j$  列线性展开

$$\det A = a_{1j} \begin{vmatrix} a_{11} & \cdots & \boxed{1} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & \boxed{0} & \cdots & a_{nn} \end{vmatrix} + \cdots + a_{nj} \begin{vmatrix} a_{11} & \cdots & \boxed{0} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \cdots & \boxed{1} & \cdots & a_{nn} \end{vmatrix}$$

把行列式  $\det(\alpha_1, \dots, \overset{(j\text{列})}{e_i}, \dots, \alpha_n)$  的第  $i$  行通过与其上面各行逐次交换位置, 再把所得行列式的第  $j$  列与它左边的列逐次交换位置, 最终得到行列式

$$\begin{vmatrix} 1 & 0 \\ 0 & M_{ij} \end{vmatrix} = M_{ij}.$$

因此

$$\det(\alpha_1, \dots, \overset{(j\text{列})}{e_i}, \dots, \alpha_n) = (-1)^{(i-1)+(j-1)} M_{ij} = (-1)^{i+j} M_{ij} = A_{ij}.$$

所以

$$\det A = a_{1j}A_{1j} + \dots + a_{nj}A_{nj}.$$

同理可证

$$\det A = a_{i1}A_{i1} + \dots + a_{in}A_{in}.$$

□

**推论 32.7.** 设  $A_{ij}$  是矩阵  $A$  的代数余子式. 则对任意  $b_1, \dots, b_n \in \mathbb{R}$ ,

$$b_1A_{1j} + \dots + b_nA_{nj} = \det(\alpha_1, \dots, \overset{(j\text{列})}{b}, \dots, \alpha_n),$$

其中  $\alpha_1, \dots, \alpha_n$  是  $A$  的列向量,  $b$  是以  $b_1, \dots, b_n$  为坐标的列向量.

**证明.** 事实上,  $A_{ij}$  只由矩阵  $A$  的除第  $i$  行和第  $j$  列以外的元素决定. 上式等号右端的行列式按第  $j$  列展开即得到左端表达式. □

**推论 32.8.** 设  $A_{ij}$  是矩阵  $A$  的代数余子式. 则

$$a_{1i}A_{1j} + \dots + a_{ni}A_{nj} = \begin{cases} \det A, & i = j; \\ 0, & i \neq j. \end{cases}$$

从而

$$AA^\# = A^\#A = \det A I_n,$$

其中  $A^\# = (A_{ij})^T$  称为  $A$  的伴随矩阵 (adjugate matrix).

**推论 32.9.**  $A$  可逆当且仅当  $\det A \neq 0$ . 且  $A^{-1} = \frac{1}{\det A} A^\#$ . □

### 33 行列式的应用

**线性方程组的解, Cramer 法则**

设  $A$  是  $n$  阶可逆方阵. 于是对任意  $b$ , 线性方程组  $Ax = b$  有唯一解

$$x = A^{-1}b = \frac{A^\#b}{\det A},$$

所以

$$x_i = \frac{(A^\#b)_i}{\det A} = \frac{1}{\det A} \sum_{j=1}^n A_{ji}b_j = \frac{\det(\alpha_1, \dots, \overset{(i\text{列})}{b}, \dots, \alpha_n)}{\det(\alpha_1, \dots, \alpha_i, \dots, \alpha_n)}.$$

最后这个结论叫 **Cramer 法则** (Cramer 1750).

$\mathbb{R}^n$  中  $k$  维平行多面体的体积

$k = 1$  时,  $\mathbf{v}$  的长度为  $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ .

$k = 2$  时,  $\mathbf{u}, \mathbf{v}$  所形成的平行四边形的面积为

$$\sqrt{\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle^2} = \sqrt{\begin{vmatrix} \langle \mathbf{u}, \mathbf{u} \rangle & \langle \mathbf{u}, \mathbf{v} \rangle \\ \langle \mathbf{u}, \mathbf{v} \rangle & \langle \mathbf{v}, \mathbf{v} \rangle \end{vmatrix}}.$$

由此我们猜测  $\mathbb{R}^n$  中由  $k$  个线性无关向量  $\mathbf{v}_1, \dots, \mathbf{v}_k$  所形成的  $k$  维平行多面体的体积为

$$\sqrt{\det(\langle \mathbf{v}_i, \mathbf{v}_j \rangle)_{k \times k}}.$$

记  $A$  为以  $\mathbf{v}_1, \dots, \mathbf{v}_k$  为列向量的矩阵, 则由 Gram-Schmidt 正交化过程可得  $A = QR$ , 其中  $Q$  由  $\mathbb{R}^n$  中  $k$  个彼此正交的单位向量  $\alpha_1, \dots, \alpha_k$  组成,  $R$  是  $k$  阶上三角矩阵. 于是

$$\mathbf{v}_1 = r_{11}\alpha_1,$$

$$\mathbf{v}_2 = r_{22}\alpha_2 + r_{12}\alpha_1,$$

$$\vdots$$

$$\mathbf{v}_k = r_{kk}\alpha_k + r_{1k}\alpha_1 + \dots + r_{(k-1)k}\alpha_{k-1},$$

于是  $\mathbf{v}_1, \dots, \mathbf{v}_k$  所形成的  $k$  维平行多面体的体积为

$$|r_{11}||r_{22}| \cdots |r_{kk}| = |\det R|.$$

$$\begin{aligned} \sqrt{\det(\langle \mathbf{v}_i, \mathbf{v}_j \rangle)_{k \times k}} &= \sqrt{\det(A^T A)} \\ &= \sqrt{\det(R^T Q^T Q R)} = \sqrt{\det(R^T R)} \\ &= \sqrt{\det R^T \cdot \det R} = |\det R| \end{aligned}$$

是  $\mathbf{v}_1, \dots, \mathbf{v}_k$  所形成的  $k$  维平行多面体的体积.

$\mathbb{R}^n$  中超平面的法向量, 向量的叉积

$\mathbb{R}^n$  中  $n-1$  维的线性子空间称为超平面 (hyperplane).  $W$  的正交补空间  $W^\perp$  是一维线性空间 (即直线),  $W^\perp$  中任何非零向量都叫做  $W$  的法向量.

设  $W$  是  $\mathbb{R}^n$  中的一个超平面, 由线性无关向量  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  张成. 我们希望用  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  求出  $W$  的一个法向量.

考虑线性函数

$$L(\mathbf{v}) = \det(\mathbf{v}, \mathbf{w}_1, \dots, \mathbf{w}_{n-1}).$$

上述行列式的代数余子式  $A_{11}, A_{21}, \dots, A_{n1}$  完全由  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  决定, 与  $\mathbf{v}$  无关.

由推论 32.7 知

$$L(\mathbf{v}) = v_1 A_{11} + v_2 A_{21} + \dots + v_n A_{n1} = \langle \mathbf{v}, \mathbf{A} \rangle,$$

其中  $\mathbf{A}$  是以  $A_{11}, A_{21}, \dots, A_{n1}$  为坐标的向量.

因为  $L(\mathbf{w}_i) = 0$  (因为有两列完全一样), 所以  $\mathbf{A}$  与  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  都正交. 另外, 把  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  扩充为  $\mathbb{R}^n$  的一组基  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}, \mathbf{v}_n$ , 从而  $L(\mathbf{v}_n) = \det(\mathbf{v}_n, \mathbf{w}_1, \dots, \mathbf{w}_{n-1}) \neq 0$ , 所以  $\mathbf{A} \neq \mathbf{0}$ , 从而是  $W$  的法向量.

$$\begin{aligned} \|\mathbf{A}\| &= \left\langle \frac{\mathbf{A}}{\|\mathbf{A}\|}, \mathbf{A} \right\rangle = L\left(\frac{\mathbf{A}}{\|\mathbf{A}\|}\right) = \det\left(\frac{\mathbf{A}}{\|\mathbf{A}\|}, \mathbf{w}_1, \dots, \mathbf{w}_{n-1}\right) \\ &= \sqrt{\begin{vmatrix} \left\langle \frac{\mathbf{A}}{\|\mathbf{A}\|}, \frac{\mathbf{A}}{\|\mathbf{A}\|} \right\rangle & \left\langle \frac{\mathbf{A}}{\|\mathbf{A}\|}, \mathbf{w}_1 \right\rangle & \cdots & \left\langle \frac{\mathbf{A}}{\|\mathbf{A}\|}, \mathbf{w}_{n-1} \right\rangle \\ \left\langle \mathbf{w}_1, \frac{\mathbf{A}}{\|\mathbf{A}\|} \right\rangle & \langle \mathbf{w}_1, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{w}_1, \mathbf{w}_{n-1} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \left\langle \mathbf{w}_{n-1}, \frac{\mathbf{A}}{\|\mathbf{A}\|} \right\rangle & \langle \mathbf{w}_{n-1}, \mathbf{w}_1 \rangle & \cdots & \langle \mathbf{w}_{n-1}, \mathbf{w}_{n-1} \rangle \end{vmatrix}} \\ &= \sqrt{\det(\langle \mathbf{w}_i, \mathbf{w}_j \rangle)_{(n-1) \times (n-1)}} \end{aligned}$$

是  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  所成的  $n-1$  维平行多面体的体积, 并且  $\mathbf{A}, \mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  构成  $\mathbb{R}^n$  的一个正向基 (见定义 31.7).

记

$$\mathbf{w}_1 \times \cdots \times \mathbf{w}_{n-1} = \mathbf{A},$$

称为  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  的叉积.

$$n=3 \text{ 时, } \mathbb{R}^3 \text{ 中两个向量 } \mathbf{w}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \text{ 和 } \mathbf{w}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \text{ 的叉积为}$$

$$\mathbf{w}_1 \times \mathbf{w}_2 = \begin{bmatrix} y_1 & y_2 \\ z_1 & z_2 \\ x_1 & x_2 \\ x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} = \begin{vmatrix} \mathbf{i} & x_1 & x_2 \\ \mathbf{j} & y_1 & y_2 \\ \mathbf{k} & z_1 & z_2 \end{vmatrix}.$$

这里  $\mathbf{i}, \mathbf{j}, \mathbf{k}$  是  $x, y, z$  轴上的单位正交向量, 它们构成右手系的一组单位正交基.

叉积  $\mathbf{w}_1 \times \mathbf{w}_2$  与  $\mathbf{w}_1, \mathbf{w}_2$  都正交,  $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_1 \times \mathbf{w}_2$  构成右手系,  $\mathbf{w}_1 \times \mathbf{w}_2$  的长度是  $\mathbf{w}_1, \mathbf{w}_2$  所成平行四边形的面积.

插值

**例 33.1 (插值多项式).** 给定  $n+1$  对数据  $(x_0, y_0), \dots, (x_n, y_n)$ , 满足  $x_0, x_1, \dots, x_n$  全不相同. 则

$$\begin{vmatrix} x^n & x^{n-1} & \cdots & x & 1 & y \\ x_0^n & x_0^{n-1} & \cdots & x_0 & 1 & y_0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ x_n^n & x_n^{n-1} & \cdots & x_n & 1 & y_n \end{vmatrix} = 0 \quad (*)$$

确定了一个关于  $x$  的（不超过  $n$  次的）多项式函数  $y = P(x)$ ，满足  $P(x_i) = y_i$  ( $\forall i = 0, 1, 2, \dots, n$ )。

如果另有关于  $x$  的不超过  $n$  次的多项式函数  $y = Q(x)$ ，满足  $Q(x_i) = y_i$  ( $\forall i = 0, 1, 2, \dots, n$ )，则  $Q = P$ 。

**证明.** 上述  $n+2$  阶行列式按第 1 行展开，得到  $x^n, x^{n-1}, \dots, x, 1, y$  的线性组合，系数完全由数据  $(x_0, y_0), \dots, (x_n, y_n)$  决定。其中  $y$  的系数为（**Vandermonde** 行列式）

$$\begin{vmatrix} x_0^n & x_0^{n-1} & \cdots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \cdots & x_1 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{n-1}^n & x_{n-1}^{n-1} & \cdots & x_{n-1} & 1 \\ x_n^n & x_n^{n-1} & \cdots & x_n & 1 \end{vmatrix} = \prod_{0 \leq i < j \leq n} (x_i - x_j) \neq 0,$$

所以对满足 (\*) 的  $x, y$ ， $y$  是关于  $x$  的（不超过  $n$  次的）多项式  $P(x)$ 。

当  $x = x_i, y = y_i$  时，(\*) 成立，所以  $P(x_i) = y_i$ 。

如果另有关于  $x$  的不超过  $n$  次的多项式函数  $y = Q(x)$ ，满足  $Q(x_i) = y_i$  ( $\forall i = 0, 1, 2, \dots, n$ )，则  $Q(x) - P(x)$  是关于  $x$  的不超过  $n$  次的多项式，但这个多项式有  $n+1$  个不同的零点  $x_0, x_1, \dots, x_n$ ，因此  $Q(x) = P(x)$ 。□

**插值多项式的另一解法：** 设插值多项式为

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0.$$

则

$$\begin{bmatrix} x_0^n & x_0^{n-1} & \cdots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \cdots & x_1 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{n-1}^n & x_{n-1}^{n-1} & \cdots & x_{n-1} & 1 \\ x_n^n & x_n^{n-1} & \cdots & x_n & 1 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix}.$$

左端的系数矩阵，因行列式非零，所以可逆。因此上述方程组有唯一解。□

事实上，(\*) 按最后一列展开，可以直接得到插值多项式的表达式。

**问题 33.2.** 证明对于同一平面上任给的五个不同点，存在唯一的圆锥曲线（可能是退化的）经过这五点。

**问题 33.3.** 给定  $\mathbb{R}^3$  中四个点，写出过该四点的球面的方程。

## 34 线性变换的特征值与特征向量

**定义 34.1.** 称一个向量空间  $V$  到自身的线性映射  $A: V \rightarrow V$  为  $V$  上的一个线性变换。 $V$  的一个线性子空间  $W$  称为  $A$ -不变的，如果  $AW \subseteq W$ ，即对任意  $w \in W$ ，都有  $Aw \in W$ 。

称非零向量  $\boldsymbol{v}$  是  $A$  的一个特征向量 (eigenvector), 如果  $\text{span}\{\boldsymbol{v}\}$  是  $A$ -不变子空间. 此时存在数  $\lambda$  使得  $A\boldsymbol{v} = \lambda\boldsymbol{v}$ , 称  $\lambda$  为  $A$  的一个特征值 (eigenvalue).

$n$  阶方阵可以看成  $\mathbb{R}^n$  上的线性变换, 因此可以定义矩阵  $A$  的特征值与特征向量.

**例 34.2.** 对角矩阵  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  的对角元就是  $\Lambda$  的全部特征值, 相应的特征向量是  $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ .

**证明.** 对角元是特征值, 相应的  $\boldsymbol{e}_i$  是特征向量, 这是很明显的.

若  $\lambda$  不是  $\Lambda$  的对角元, 则  $\Lambda - \lambda I$  可逆, 从而  $\Lambda\boldsymbol{v} = \lambda\boldsymbol{v}$  只有零解, 所以  $\lambda$  不是  $A$  的特征值.  $\square$

**问题 34.3.** 证明对任何三角矩阵  $A$ ,  $A$  的对角元恰好是  $A$  的全部特征值. (请不要使用特征多项式)

**例 34.4.** 矩阵  $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$  没有实数特征值和实数特征向量. 这是因为

$$\begin{aligned} A \begin{bmatrix} x \\ y \end{bmatrix} &= \lambda \begin{bmatrix} x \\ y \end{bmatrix} \text{ 有非零向量解} \\ \iff \begin{cases} -y = \lambda x, \\ x = \lambda y, \end{cases} &\text{ 有非零解} \\ \implies \begin{cases} 0 = x - \lambda y = x + \lambda^2 x = x(\lambda^2 + 1), \\ 0 = y + \lambda x = y + \lambda^2 y = y(\lambda^2 + 1), \end{cases} \end{aligned}$$

$\lambda$  是实数时,  $\lambda^2 + 1 > 0$ , 从而  $x = y = 0$ , 这与向量  $\begin{bmatrix} x \\ y \end{bmatrix} \neq \mathbf{0}$  矛盾.

然而, 如果把  $A$  当成  $\mathbb{C}^2$  上的线性映射, 并且在  $\mathbb{C}^2$  上允许使用复数的数乘, 那么可取  $\lambda = i$  或  $\lambda = -i$ , 且不难发现: 对  $\lambda = i$ ,  $\begin{bmatrix} 1 \\ -i \end{bmatrix}$  是特征向量; 对  $\lambda = -i$ ,  $\begin{bmatrix} 1 \\ i \end{bmatrix}$  是特征向量.

上面这个例子说明, 如果要完整讨论矩阵的特征值和特征向量, 需要在复向量空间 (详见后续“域和向量空间”一节) 上进行.

**定理 34.5.** 任何  $n$  阶复数矩阵都有 (复) 特征值和 (复) 特征向量.

**证明.** 设  $A$  是一个  $n$  阶复数矩阵.

因为全部  $n$  阶矩阵组成的线性空间是  $n^2$  维的, 所以  $I, A, A^2, \dots, A^{n^2}$  这  $n+1$  个  $n$  阶矩阵必然线性相关, 从而存在不全为零的数  $c_0, c_1, \dots, c_{n^2} \in \mathbb{C}$  使

得

$$c_0 I_n + c_1 A + \cdots + c_{n^2} A^{n^2} = 0.$$

很明显  $c_1, \dots, c_{n^2}$  不全为零. 记

$$k = \max\{j | c_j \neq 0\}.$$

于是  $k \geq 1$ , 且

$$c_0 I_n + c_1 A + \cdots + c_k A^k = 0. \quad (*)$$

由代数学基本定理知: 在  $\mathbb{C}$  上, 多项式  $p(\lambda) = c_0 + c_1 \lambda + \cdots + c_k \lambda^k$  可以因式分解为

$$p(\lambda) = c_k (\lambda - \lambda_k) \cdots (\lambda - \lambda_2)(\lambda - \lambda_1).$$

则 (\*) 表明矩阵

$$p(A) = c_k (A - \lambda_k I) \cdots (A - \lambda_2 I)(A - \lambda_1 I)$$

满足对任意非零向量  $\mathbf{v}$ ,  $p(A)\mathbf{v} = \mathbf{0}$ .

记  $\mathbf{v}_0 = \mathbf{v}$ ,  $\mathbf{v}_j = (A - \lambda_j I) \cdots (A - \lambda_1 I)\mathbf{v}$ . 则由 (\*) 知  $\mathbf{v}_k = \mathbf{0}$ .

因此存在  $j$  使得  $\mathbf{v}_j \neq \mathbf{0}$ , 但  $\mathbf{v}_{j+1} = \mathbf{0}$ . 而

$$\mathbf{0} = \mathbf{v}_{j+1} = (A - \lambda_{j+1} I)\mathbf{v}_j,$$

所以  $\lambda_{j+1}$  是  $A$  的特征值,  $\mathbf{v}_j$  是特征向量. □

**推论 34.6.** 任何满足 (\*) 的多项式  $p$  都叫做矩阵  $A$  的**零化多项式**.  $A$  的特征值是任何零化多项式的根.

**证明.** 设  $p$  是  $A$  的一个零化多项式,  $\lambda_0$  是  $A$  的一个特征值,  $\mathbf{u}$  是相应的一个特征向量.

假设  $p(\lambda_0) \neq 0$ . 则由多项式带余除法知

$$p(\lambda) = q(\lambda)(\lambda - \lambda_0) + r, \quad r \neq 0.$$

于是

$$\mathbf{0} = p(A)\mathbf{u} = q(A)(A - \lambda_0 I)\mathbf{u} + r\mathbf{u} = r\mathbf{u},$$

这与  $r \neq 0$  且  $\mathbf{u} \neq \mathbf{0}$  矛盾. 所以  $p(\lambda_0) = 0$ . □

**问题 34.7.** 写出矩阵  $A = \begin{bmatrix} 20 & 63 & -63 \\ -12 & -40 & 42 \\ -6 & -21 & 23 \end{bmatrix}$  的一个零化多项式.

**定理 34.8.** 1. 如果实数矩阵  $A$  在  $\mathbb{C}^n$  上有实数特征值  $\lambda$ , 则  $\lambda$  也是  $A$  在  $\mathbb{R}^n$  上的特征值,  $A$  的复特征向量的实部和虚部中的非零向量都是  $A$  的实特征向量.



2. 如果实数矩阵  $A$  在  $\mathbb{C}^n$  上有复数特征值  $\alpha + i\beta$  ( $\alpha, \beta \in \mathbb{R}, \beta \neq 0$ ), 则  $A$  的复特征向量  $\mathbf{u} + i\mathbf{v}$  ( $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ) 满足  $\mathbf{u}, \mathbf{v}$  都是非零的实数向量, 且

$$\begin{cases} A\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{v}, \\ A\mathbf{v} = -\beta\mathbf{u} + \alpha\mathbf{v}. \end{cases}$$

证明. 直接计算验证. □

**例 34.9.** 实对称矩阵所有特征向量的分量都是实数, 且  $\mathbb{R}^n$  有一组单位正交基由  $A$  的特征向量组成. 由上述定理知,  $A$  有特征值  $\lambda_1 \in \mathbb{C}$ , 从而有 (单位) 特征向量  $\mathbf{v} \in \mathbb{C}^n$ . 因此  $A\mathbf{v} = \lambda_1\mathbf{v}$ . 于是

$$\mathbf{v}^H A \mathbf{v} = \lambda_1 \mathbf{v}^H \mathbf{v},$$

其中  $\mathbf{v}^H$  是  $\mathbf{v}$  的共轭转置. 上式两边取共轭转置, 并注意到  $A$  是实对称矩阵 ( $A^H = A$ ), 得到

$$\mathbf{v}^H A^H \mathbf{v} = \mathbf{v}^H A \mathbf{v} = \overline{\lambda_1} \mathbf{v}^H \mathbf{v},$$

因此比较上面两个等式, 得到

$$\lambda_1 \mathbf{v}^H \mathbf{v} = \overline{\lambda_1} \mathbf{v}^H \mathbf{v},$$

因为  $\mathbf{v}^H \mathbf{v} = \|\mathbf{v}\|^2 > 0$ , 所以  $\lambda_1 = \overline{\lambda_1}$ , 因此  $\lambda_1$  是实数, 从而  $\mathbf{v}$  是实向量.  $\text{span}_{\mathbb{R}}\{\mathbf{v}\}$  是  $\mathbb{R}^n$  中的  $A$ -不变子空间.

由  $A$  的对称性知

$$\langle \mathbf{v}, A\mathbf{u} \rangle_{\mathbb{R}^n} = \langle A\mathbf{v}, \mathbf{u} \rangle_{\mathbb{R}^n} = \lambda_1 \langle \mathbf{v}, \mathbf{u} \rangle_{\mathbb{R}^n},$$

所以  $\mathbf{v}^\perp$  是  $\mathbb{R}^n$  中的  $A$ -不变子空间, 在  $\mathbf{v}^\perp$  中取一组单位正交基  $\mathbf{v}_2, \dots, \mathbf{v}_n$ , 则  $P = (\mathbf{v}, \mathbf{v}_2, \dots, \mathbf{v}_n)$  是正交矩阵, 且

$$P^T A P = \begin{bmatrix} \lambda_1 & \\ & A_1 \end{bmatrix}$$

$A_1$  是  $n-1$  阶实对称矩阵. 用数学归纳法可知  $A$  的所有特征向量都是实数, 且  $\mathbb{R}^n$  有一组单位正交基由  $A$  的特征向量组成.

**定义 34.10.** 当  $\lambda$  是矩阵  $A$  的特征值时,  $\text{Ker}(A - \lambda I)$  称为  $A$  对应于  $\lambda$  的特征子空间, 称  $\dim \text{Ker}(A - \lambda I)$  为特征值  $\lambda$  的几何重数.

**定理 34.11.** 1. 设  $\lambda_1, \dots, \lambda_k$  是矩阵  $A$  的不同的特征值, 则它们的特征子空间  $\text{Ker}(A - \lambda_1 I), \dots, \text{Ker}(A - \lambda_k I)$  构成直和

$$\text{Ker}(A - \lambda_1 I) \oplus \dots \oplus \text{Ker}(A - \lambda_k I),$$

即：若

$$\mathbf{v}_1 + \cdots + \mathbf{v}_k = \mathbf{0},$$

其中  $\mathbf{v}_1 \in \text{Ker}(A - \lambda_1 I), \dots, \mathbf{v}_k \in \text{Ker}(A - \lambda_k I)$ , 则必有  $\mathbf{v}_1 = \mathbf{0}, \dots, \mathbf{v}_k = \mathbf{0}$ . 换言之, 各取  $\text{Ker}(A - \lambda_1 I), \dots, \text{Ker}(A - \lambda_k I)$  的一组基, 它们组成

$$\text{Ker}(A - \lambda_1 I) + \cdots + \text{Ker}(A - \lambda_k I)$$

的基.

2.  $n$  阶矩阵至多有  $n$  个不同的特征值.

证明. (1)

$$\mathbf{0} = (A - \lambda_2 I) \cdots (A - \lambda_k I)(\mathbf{v}_1 + \cdots + \mathbf{v}_k) = (\lambda_1 - \lambda_2) \cdots (\lambda_1 - \lambda_k) \mathbf{v}_1,$$

所以  $\mathbf{v}_1 = \mathbf{0}$ . 同理其他  $\mathbf{v}_i = \mathbf{0}$ . 余下结论由定理 24.6 得到.

(2) 直和子空间

$$\text{Ker}(A - \lambda_1 I) \oplus \cdots \oplus \text{Ker}(A - \lambda_k I)$$

的维数至少是  $k$ , 至多是  $n$ , 所以  $k \leq n$ . □

**推论 34.12.** 设  $\lambda_1, \dots, \lambda_k$  是矩阵  $A$  的全部不同的特征值. 则存在可逆矩阵  $P$  使得  $P^{-1}AP$  为对角阵, 当且仅当  $A$  的特征值的几何重数之和为  $n$ . □

**推论 34.13.** 如果矩阵  $A$  有  $n$  个不同的特征值, 则存在可逆矩阵  $P$  使得  $P^{-1}AP$  为对角阵. □

## 35 矩阵的特征多项式, Cayley-Hamilton 定理

**定理 35.1.**  $\lambda$  是矩阵  $A$  的特征值当且仅当  $(A - \lambda I)\mathbf{v} = \mathbf{0}$  有非零解 (即  $\text{Ker}(A - \lambda I) \neq \{\mathbf{0}\}$ ), 当且仅当  $\det(\lambda I - A) = 0$ . □

**定义 35.2.** 称  $n$  次多项式

$$\chi_A(\lambda) = \det(\lambda I - A)$$

为  $A$  的特征多项式.

上一节中我们证明了任何矩阵  $A$  都有零化多项式, 但那是一个概念性证明, 不是构造性的. 但下述 Cayley-Hamilton 定理以明确的方式构造了一个零化多项式.

另外, 上一节中所说的零化多项式是一个不超过  $n^2$  次的多项式, 它可能有很多根, 我们只知道这些根中包含了所有特征值, 但也有可能有些根不是矩阵  $A$  的特征值. 而下述定理的第一个结论说, 特征多项式的根恰好由  $A$  的全部特征值组成, 从而不同特征值的个数不超过  $n$ . 因此我们只需要关注矩阵的特征多项式即可获得全部特征值.

**定理 35.3.** 1.  $\lambda$  是矩阵  $A$  的特征值当且仅当  $\lambda$  是特征多项式的根.

2. (Cayley-Hamilton) 矩阵  $A$  的特征多项式  $\chi_A$  是  $A$  的一个零化多项式.

**证明.** (1) 是显然的.

(2) 相应于矩阵

$$P(\lambda) = A - \lambda I = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix},$$

考虑  $n^2$  阶矩阵

$$Q(A) = \begin{bmatrix} a_{11}I_n - A & a_{12}I_n & \cdots & a_{1n}I_n \\ a_{21}I_n & a_{22}I_n - A & \cdots & a_{2n}I_n \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}I_n & a_{n2}I_n & \cdots & a_{nn}I_n - A \end{bmatrix}.$$

取  $v = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$ , 则

$$Q(A)v = \begin{bmatrix} a_{11}e_1 - Ae_1 + a_{12}e_2 + \cdots + a_{1n}e_n \\ a_{21}e_1 + a_{22}e_2 - Ae_2 + \cdots + a_{2n}e_n \\ \vdots \\ a_{n1}e_1 + a_{n2}e_2 + \cdots + a_{nn}e_n - Ae_n \end{bmatrix} = \begin{bmatrix} \alpha_1 - \alpha_1 \\ \alpha_2 - \alpha_2 \\ \vdots \\ \alpha_n - \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

由推论 32.8 知矩阵  $P(\lambda)$  的伴随矩阵 (由代数余子式决定)  $P(\lambda)^\#$  满足

$$P(\lambda)^\# P(\lambda) = \det P(\lambda) \cdot I_n = \chi_A(\lambda) I_n = \begin{bmatrix} \chi_A(\lambda) & 0 & \cdots & 0 \\ 0 & \chi_A(\lambda) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_A(\lambda) \end{bmatrix},$$

把上式中的  $\lambda$  换成矩阵  $A$ , 常数  $c$  换成  $cI_n$ , 得到

$$\tilde{Q}(A)Q(A) = \begin{bmatrix} \chi_A(A)I_n & 0 & \cdots & 0 \\ 0 & \chi_A(A)I_n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_A(A)I_n \end{bmatrix}.$$

两边作用于上述向量  $\mathbf{v}$ , 得到

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \tilde{Q}(A)Q(A)\mathbf{v} = \begin{bmatrix} \chi_A(A) & 0 & \cdots & 0 \\ 0 & \chi_A(A) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_A(A) \end{bmatrix} \mathbf{v} = \begin{bmatrix} \chi_A(A)\mathbf{e}_1 \\ \chi_A(A)\mathbf{e}_2 \\ \vdots \\ \chi_A(A)\mathbf{e}_n \end{bmatrix}.$$

因为矩阵  $\chi_A(A)$  作用于  $\mathbb{R}^n$  ( $\mathbb{C}^n$ ) 的一组基上为零, 所以  $\chi_A(A) = 0$ .  $\square$

**推论 35.4.** 对任何方阵  $A$ ,  $A$  的迹  $\text{tr}(A)$  等于  $A$  的全部特征值 (计重数) 之和,  $A$  的行列式的值等于  $A$  的全部特征值 (计重数) 的乘积.

**证明.** 计算特征多项式中常数项的值以及  $\lambda^{n-1}$  的系数. 结合 Vieta 定理即可得到结论.  $\square$

**问题 35.5.** 写出矩阵  $A = \begin{bmatrix} 20 & 63 & -63 \\ -12 & -40 & 42 \\ -6 & -21 & 23 \end{bmatrix}$  的特征多项式  $\chi_A(\lambda)$ , 并通过直接计算验证  $\chi_A(A) = 0$ .

计算  $A$  的全部特征值 (计重数), 并验证  $\text{tr}(A), \det A$  与特征值的关系.

## 36 幂零矩阵和循环子空间, 根子空间

**例 36.1.** 矩阵  $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$  的特征多项式为  $\chi_A(\lambda) = \lambda^3$ , 其零点为三重根

0. 由  $(A - 0I)\mathbf{v} = A\mathbf{v} = \mathbf{0}$  解得特征向量  $\mathbf{v} = \mathbf{e}_1$ , 特征子空间  $\text{Ker}(A - 0I) = \text{span}\{\mathbf{e}_1\}$ , 几何重数为 1.

$A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ ,  $A^3 = 0$ . 这样的矩阵称为**幂零矩阵** (nilpotent matrix), 即存在正整数  $k$  使得  $A^k = 0$ .

由  $A^2\mathbf{v} = \mathbf{0}$  解得  $\mathbf{v} = t\mathbf{e}_1 + s\mathbf{e}_2$ , 所以

$$\text{Ker}(A^2) = \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}.$$

$$\text{Ker}(A^3) = \mathbb{R}^3 = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}.$$

于是我们看到

$$\{\mathbf{0}\} \subset \text{Ker}A \subset \text{Ker}(A^2) \subset \text{Ker}(A^3) = \text{Ker}(A^4) = \cdots.$$

$$\mathbf{e}_3 \in \text{Ker}(A^3) \setminus \text{Ker}(A^2), \quad A\mathbf{e}_3 = \mathbf{e}_2, \quad A^2\mathbf{e}_3 = A\mathbf{e}_2 = \mathbf{e}_1.$$

所以

$$\mathbb{R}^3 = \text{span}\{\mathbf{e}_3, A\mathbf{e}_3, A^2\mathbf{e}_3\}.$$

**定理 36.2.** (1) 若  $A^{k+1}\mathbf{v} = \mathbf{0}$ , 则循环子空间 (定义见问题 7.7)

$$\text{Cyc}_A(\mathbf{v}) = \text{span}\{\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots, A^k\mathbf{v}\},$$

是  $A$ -不变子空间.

(2) 若  $A^k\mathbf{v} \neq \mathbf{0}$ ,  $A^{k+1}\mathbf{v} = \mathbf{0}$ , 即

$$\mathbf{v} \xrightarrow{A} A\mathbf{v} \xrightarrow{A} A^2\mathbf{v} \cdots \xrightarrow{A} A^k\mathbf{v} \xrightarrow{A} \mathbf{0},$$

则  $\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \dots, A^k\mathbf{v}$  是循环子空间  $\text{Cyc}_A(\mathbf{v})$  的一组基, 在这组基下,  $A$  的表示矩阵为

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

**证明.** 留作练习. □

**定理 36.3.** 设  $\lambda$  是  $A$  的一个特征值. 则

1. 存在正整数  $m \leq n$  使得

$$\begin{aligned} \{\mathbf{0}\} &\subset \text{Ker}(A - \lambda I) \subset \cdots \subset \text{Ker}(A - \lambda I)^{m-1} \subset \text{Ker}(A - \lambda I)^m \\ &= \text{Ker}(A - \lambda I)^{m+1} = \text{Ker}(A - \lambda I)^{m+2} = \cdots, \end{aligned}$$

2.  $\mathbb{R}^n = \text{Ker}(A - \lambda I)^m \oplus \text{Range}(A - \lambda I)^m$  是  $A$  不变的子空间直和分解.

3.  $A - \lambda I$  在  $\text{Range}(A - \lambda I)^m$  上可逆.

**证明.** (1) 如果  $\text{Ker}(A - \lambda I)^k = \text{Ker}(A - \lambda I)^{k+1}$ , 则对任意正整数  $j$ ,

$$\text{Ker}(A - \lambda I)^{k+j} = \text{Ker}(A - \lambda I)^{k+j+1}.$$

这是因为: 若

$$(A - \lambda I)^{k+j+1}\mathbf{v} = \mathbf{0},$$

则

$$(A - \lambda I)^j\mathbf{v} \in \text{Ker}(A - \lambda I)^{k+1} = \text{Ker}(A - \lambda I)^k,$$

所以

$$(A - \lambda I)^{k+j}\mathbf{v} = \mathbf{0},$$

因此  $\text{Ker}(A - \lambda I)^{k+j} = \text{Ker}(A - \lambda I)^{k+j+1}$ .

$\dim \text{Ker}(A - \lambda I)^k$  关于  $k$  单调不减, 有上界  $n$ . 所以存在最小的正整数  $m$  使得  $\dim \text{Ker}(A - \lambda I)^m = \dim \text{Ker}(A - \lambda I)^{m+1}$ . 此时

$$\{\mathbf{0}\} \subset \text{Ker}(A - \lambda I) \subset \cdots \subset \text{Ker}(A - \lambda I)^{m-1} \subset \text{Ker}(A - \lambda I)^m = \text{Ker}(A - \lambda I)^{m+1}.$$

在严格上升部分, 空间维数每次至少增加 1, 而最大子空间维数不超过  $n$ , 所以  $m \leq n$ .

(2) 设  $\mathbf{v} \in \text{Range}(A - \lambda I)^m \cap \text{Ker}(A - \lambda I)^m$ . 则  $(A - \lambda I)^m \mathbf{v} = \mathbf{0}$  且存在  $\mathbf{u} \in \mathbb{R}^n$  使得  $\mathbf{v} = B^m \mathbf{u}$ . 于是  $B^{2m} \mathbf{u} = \mathbf{0}$ . 因为  $\text{Ker}(A - \lambda I)^{2m} = \text{Ker}(A - \lambda I)^m$ , 从而  $(A - \lambda I)^m \mathbf{u} = \mathbf{0}$ , 因此  $\mathbf{v} = (A - \lambda I)^m \mathbf{u} = \mathbf{0}$ . 因此  $\text{Range}(A - \lambda I)^m \cap \text{Ker}(A - \lambda I)^m = \{\mathbf{0}\}$ , 从而  $\text{Range}(A - \lambda I)^m + \text{Ker}(A - \lambda I)^m$  是直和.

由

$$\dim \text{Ker}(A - \lambda I)^m = n - \dim \text{Range}(A - \lambda I)^m,$$

所以

$$\mathbb{R}^n = \text{Ker}(A - \lambda I)^m \oplus \text{Range}(A - \lambda I)^m.$$

因为  $A$  与  $A - \lambda I$  交换, 所以  $\text{Ker}(A - \lambda I)^m$  和  $\text{Range}(A - \lambda I)^m$  都是  $A$ -不变子空间.

(3) 因为  $\text{Ker}(A - \lambda I) \subset \text{Ker}(A - \lambda I)^m$ , 所以  $\text{Ker}(A - \lambda I) \cap \text{Range}(A - \lambda I)^m = \{\mathbf{0}\}$ , 因此  $A - \lambda I$  在  $\text{Range}(A - \lambda I)^m$  上可逆.  $\square$

**定义 36.4.** 称

$$R_\lambda = \{\mathbf{v} | \text{存在 } k \text{ 使得 } (A - \lambda I)^k \mathbf{v} = \mathbf{0}\}$$

为  $\lambda$  所对应的根子空间.

由上述定理知, 存在正整数  $m$  使得  $R_\lambda = \text{Ker}(A - \lambda I)^m$ .

**定理 36.5.** 设  $\lambda_1, \dots, \lambda_k$  是  $A$  的全部的不同的特征值. 则

- (1) 根子空间  $R_{\lambda_1}, \dots, R_{\lambda_k}$  都是  $A$ -不变子空间,
- (2) 当  $\mu \neq \lambda_i$  时,  $A - \mu I$  在  $R_{\lambda_i}$  上是可逆的,
- (3) 根子空间  $R_{\lambda_1}, \dots, R_{\lambda_k}$  构成直和,
- (4)  $\mathbb{R}^n = R_{\lambda_1} \oplus \cdots \oplus R_{\lambda_k}$ .

**证明.** (1)  $A - \lambda I$  与  $A$  可交换, 由此易知根子空间是  $A$ -不变子空间.

(2) 若  $\mathbf{v} \in R_{\lambda_i}$  满足  $(A - \mu I)\mathbf{v} = \mathbf{0}$ , 则

$$(A - \lambda_i I)\mathbf{v} = (A - \mu I)\mathbf{v} + (\mu - \lambda_i)\mathbf{v} = (\mu - \lambda_i)\mathbf{v},$$

从而

$$\mathbf{0} = (A - \lambda_i I)^m \mathbf{v} = (\mu - \lambda_i)^m \mathbf{v},$$

所以  $\mathbf{v} = \mathbf{0}$ . 因此  $A - \mu I$  在  $R_{\lambda_i}$  上是可逆的.

(3) 设

$$\mathbf{v}_1 + \cdots + \mathbf{v}_k = \mathbf{0},$$

其中  $\mathbf{v}_1 \in R_{\lambda_1}, \dots, \mathbf{v}_k \in R_{\lambda_k}$ , 则必有

$$\mathbf{0} = (A - \lambda_2 I)^n \cdots (A - \lambda_k I)^n (\mathbf{v}_1 + \cdots + \mathbf{v}_k) = (A - \lambda_2 I)^n \cdots (A - \lambda_k I)^n \mathbf{v}_1.$$

因为  $(A - \lambda_2 I)^n \cdots (A - \lambda_k I)^n$  在  $R_{\lambda_1}$  上可逆, 所以  $\mathbf{v}_1 = \mathbf{0}$ . 同理可证其他  $\mathbf{v}_i = \mathbf{0}$ . 所以  $R_{\lambda_1}, \dots, R_{\lambda_k}$  构成直和.

(4) 记  $W = R_{\lambda_1} \oplus \cdots \oplus R_{\lambda_k}$ . 则  $W$  是  $A$ -不变子空间.

假设  $W \neq \mathbb{R}^n$ .

任取  $W$  的补空间  $W_1$ , 则在  $W \oplus W_1$  的一组基下,

$$A = \begin{bmatrix} A_W & C \\ 0 & B \end{bmatrix},$$

于是  $A_W = A|W$ . 所以  $A_W$  只有特征值  $\lambda_1, \dots, \lambda_k$ .

若  $\mu$  是  $B$  的特征值且  $\mu \notin \{\lambda_1, \dots, \lambda_k\}$ , 则存在  $\mathbf{u} \in W_1$  是  $B$  的特征向量,  $A_W - \mu I$  可逆, 取

$$\mathbf{w} = -(A_W - \mu I)^{-1} C \mathbf{u}.$$

则  $\begin{bmatrix} \mathbf{w} \\ \mathbf{u} \end{bmatrix}$  是  $A$  对应于  $\mu$  的特征向量. 这与  $\lambda_1, \dots, \lambda_k$  是  $A$  的全部特征值矛盾.

若  $\mu$  是  $B$  的特征值, 则存在  $i$  使得  $\mu = \lambda_i$ . 不妨设  $i = 1$ .

取  $B$  对应于特征值  $\lambda_1$  的特征向量  $\mathbf{u}$ .

记

$$A = \begin{bmatrix} A_1 & 0 & C_1 \\ 0 & A_2 & C_2 \\ 0 & 0 & B \end{bmatrix},$$

其中  $A_1 = A|R_{\lambda_1}$ ,  $A_2 = A|R_{\lambda_2} \oplus \cdots \oplus R_{\lambda_k}$ ,  $C_1 : W_1 \rightarrow R_{\lambda_1}$ ,  $C_2 : W_1 \rightarrow R_{\lambda_2} \oplus \cdots \oplus R_{\lambda_k}$ . 则

$$(A - \lambda_1 I) \begin{bmatrix} \mathbf{0} \\ -(A_2 - \lambda_1 I)^{-1} C_2 \mathbf{u} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} C_1 \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

所以

$$(A - \lambda_1 I)^{m+1} \begin{bmatrix} \mathbf{0} \\ -(A_2 - \lambda_1 I)^{-1} C_2 \mathbf{u} \\ \mathbf{u} \end{bmatrix} = (A - \lambda_1 I)^m \begin{bmatrix} C_1 \mathbf{u} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \mathbf{0},$$

因此

$$\begin{bmatrix} \mathbf{0} \\ -(A_2 - \lambda_1 I)^{-1} C_2 \mathbf{u} \\ \mathbf{u} \end{bmatrix} \in R_{\lambda_1}.$$

因为  $R_{\lambda_1} \oplus \cdots \oplus R_{\lambda_k} \oplus W_1$  是直和, 所以  $\mathbf{u} = \mathbf{0}$ , 但这与  $\mathbf{u}$  是特征向量矛盾.

因此  $W = \mathbb{R}^n$ .  $\square$

于是我们希望在每个根子空间中找到“好的”基, 使  $A$  可以得到简单表达. 由定理 36.2 知找到如下形式的最长向量链可以获得最大个数的线性无关的向量组,

$$\begin{aligned} \xrightarrow{A-\lambda I} \mathbf{v}_0 \xrightarrow{A-\lambda I} \mathbf{v}_1 \xrightarrow{A-\lambda I} \mathbf{v}_2 \xrightarrow{A-\lambda I} \cdots \xrightarrow{A-\lambda I} \mathbf{v}_k \xrightarrow{A-\lambda I} \mathbf{0} \\ \xrightarrow{A-\lambda I} \mathbf{u}_0 \xrightarrow{A-\lambda I} \mathbf{u}_1 \xrightarrow{A-\lambda I} \cdots \xrightarrow{A-\lambda I} \mathbf{u}_l \xrightarrow{A-\lambda I} \mathbf{0}, \end{aligned}$$

如果  $\mathbf{v}_k, \mathbf{u}_l$  线性无关, 则上述向量链上的向量都是线性无关的. 对更多的链, 这结论也成立.

## 37 矩阵的 Jordan 标准形

定义 37.1. 形如

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

的矩阵称为一个 **Jordan 块**.

**定理 37.2 (Jordan 标准型).** 对复数域上的任何  $n$  阶方阵  $A$ , 存在  $n$  阶可逆复矩阵  $P$  使得  $P^{-1}AP$  为对角块矩阵, 其中每个对角块为对角矩阵或 *Jordan* 块矩阵.

**例 37.3.** 求以下矩阵的 Jordan 标准形.

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ -1 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 1 & 0 \\ 0 & 0 & -2 & 2 & 1 \\ 2 & -1 & 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} A_2 & 0 \\ B & A_3 \end{bmatrix}$$

解.

$$\det(\lambda I - A) = \begin{vmatrix} \lambda I_2 - A_2 & 0 \\ -B & \lambda I_3 - A_3 \end{vmatrix} = (\lambda - 2)^5$$



所以 2 是唯一特征值.

$$A - 2I = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & -2 & 0 & 1 \\ 2 & -1 & 3 & 1 & -1 \end{bmatrix}, (A - 2I)^2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 1 \\ 0 & -1 & 1 & -1 & -1 \\ 0 & 2 & -2 & 2 & 2 \end{bmatrix},$$

$$(A - 2I)^3 = 0.$$

$$\mathbb{R}^5 = K_3 \xrightarrow{A-2I} K_2 \xrightarrow{A-2I} K_1 \xrightarrow{A-2I} \{\mathbf{0}\}, \quad K_j = \text{Ker}(A - 2I)^j.$$

由  $\text{rank}(A - 2I) = 3$  和  $\text{rank}(A - 2I)^2 = 1$  知  $\dim \text{Ker}(A - 2I) = 2$ ,  $\dim \text{Ker}(A - 2I)^2 = 4$ . 由  $(A - 2I)^3 = 0$ , 所以根子空间为  $K_3 = \text{Ker}(A - 2I)^3$ .

任取根子空间一组基 (这里取  $\mathbb{R}^5$  的标准基  $e_1, \dots, e_5$ ), 用  $A - 2I$  迭代, 从中发现

$$e_1 \xrightarrow{A-2I} v_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 2 \end{bmatrix} \xrightarrow{A-2I} \mathbf{0},$$

$$e_2 \xrightarrow{A-2I} v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \xrightarrow{A-2I} v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 2 \end{bmatrix} \xrightarrow{A-2I} \mathbf{0},$$

这样得到  $\mathbb{R}^5$  (实际上是根子空间  $K_3$ ) 的一组基为

$$v_1, e_1, v_3, v_2, e_2,$$

在这组基下

$$(A - 2I)(v_1, e_1, v_3, v_2, e_2) = (v_1, \mathbf{0}, v_2, v_3, \mathbf{0})$$

$$= (v_1, e_1, v_3, v_2, e_2) \left[ \begin{array}{cc|ccc} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

$$A(v_1, e_1, v_3, v_2, e_2) = (v_1, e_1, v_3, v_2, e_2) \left[ \begin{array}{cc|ccc} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ \hline 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{array} \right].$$

记

$$P = (v_1, e_1, v_3, v_2, e_2) = \left[ \begin{array}{cc|ccc} -1 & 1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 2 & 0 & 2 & -1 & 0 \end{array} \right]$$

则

$$P^{-1} = \left[ \begin{array}{ccccc} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & -1 \\ 0 & 1 & -1 & 1 & 1 \end{array} \right], \quad P^{-1}AP = \left[ \begin{array}{cc|ccc} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ \hline 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{array} \right].$$

这样就得到了矩阵  $A$  的 Jordan 标准形.

□

$e_1$	$e_2$	$e_3$	$e_4$	$e_5$	
		$\downarrow$			
$v_{11}$	$v_{12}$	$v_{13}$	$v_{14}$	$v_{15}$	$\dim K_3 = 5$
-1	1	0	0	0	$K_3$ 的基:
-1	1	0	0	0	$v_{11}, v_{22}$ 特征向量
1	0	1	1	0	$e_1, v_{12}$ 一阶广义特征向量
0	0	-2	0	1	$e_2$ 二阶广义特征向量
2	-1	3	1	-1	
		$\downarrow$			$\dim K_2 = 4$
$v_{21}$	$v_{22}$	$v_{23}$	$v_{24}$	$v_{25}$	$K_2$ 的基:
0	0	0	0	0	$v_{11}, v_{22}$ 特征向量
	0	0	0	0	$e_1, v_{12}$ 一阶广义特征向量
	1	-1	1	1	
	-1	1	-1	-1	$\dim K_1 = 2$
	2	-2	2	2	$K_1$ 的基:
		$\downarrow$			$v_{11}, v_{22}$ 特征向量
$v_{31}$	$v_{32}$	$v_{33}$	$v_{34}$	$v_{35}$	
0	0	0	0	0	

## 38 域和向量空间

**定义 38.1.** 设集合  $\mathbb{F}$  上有两个运算, 分别称为加法和乘法,  $+: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$ ,  $\cdot: \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{F}$ , 并且  $\mathbb{F}$  包含两个不同的元素  $0, 1$  满足:

(1)  $\mathbb{F}$  关于加法满足交换律、结合律, 0 是加法单位元, 每个  $a \in \mathbb{F}$  都有加法逆元;

(2)  $\mathbb{F}$  关于乘法满足交换律、结合律, 1 是乘法单位元, 每个非零的  $a \in \mathbb{F}$  都有乘法逆元;

(3)  $\mathbb{F}$  的乘法对加法有分配律.

我们称满足上述性质的集合  $\mathbb{F}$  为一个域.

**例 38.2.** (1) 有理数集  $\mathbb{Q}$ , 实数集  $\mathbb{R}$ , 复数集  $\mathbb{C}$  都是域.

(2)  $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} | a, b \in \mathbb{Q}\}$  是域.

(3) 如果  $\mathbb{F}$  是一个域, 则以  $\mathbb{F}$  中元素为系数的全体有理分式组成的集合  $\mathbb{F}\{x\}$  是一个域.

(4) 如果  $p$  是一个素数 (即大于 1 的正整数, 其正整数因子只有 1 和  $p$  两个), 则  $\mathbb{Z}_p = \{0, 1, 2, \dots, p-1\}$  是一个域, 其中的加法和乘法都按  $p$  取余数, 例如在  $\mathbb{Z}_5$  中,

$$3 + 2 = 0, \quad 3 \times 2 = 1.$$

**定义 38.3.** 称集合  $V$  是域  $\mathbb{F}$  上的一个向量空间 (也称为线性空间)<sup>6</sup>, 如果  $V$  上存在两种运算, 分别称为加法和数乘,  $+: V \times V \rightarrow V$ ,  $\cdot: \mathbb{F} \times V \rightarrow V$ , 并且  $V$  包含一个元素  $\mathbf{0}$  满足:

(1)  $V$  关于加法满足交换律、结合律,  $\mathbf{0}$  是加法单位元 (称为零向量), 每个  $v \in V$  都有加法逆元;

(2)  $V$  关于数乘满足以下性质:

$$(2.1) a(bv) = (ab)v, \quad 1 \cdot v = v;$$

$$(2.2) a(u+v) = au + av, \quad (a+b)v = av + bv;$$

**定理 38.4.** 在向量空间  $V$  中,

(1) 零向量是唯一的, 且  $0 \cdot v = \mathbf{0}$ ;

(2) 每个向量有唯一的加法逆元, 且为  $(-1) \cdot v$ ;

(3) 加法消去律成立: 若  $u + w = v + w$ , 则  $u = v$ ;

(4) 乘法消去律成立: 若  $\lambda v = \mathbf{0}$ , 则  $v = \mathbf{0}$  或  $\lambda = 0$ ;

**证明.** (1) 设  $\mathbf{0}_1, \mathbf{0}_2$  是两个零向量. 则  $\mathbf{0}_1 = \mathbf{0}_1 + \mathbf{0}_2 = \mathbf{0}_2$ .

对任意  $v$ , 取  $v$  的加法逆元  $u$ , 则

$$\begin{aligned} 0 \cdot v &= \mathbf{0} + 0 \cdot v = (u + v) + 0 \cdot v \\ &= u + (v + 0 \cdot v) \\ &= u + (1 + 0) \cdot v = u + 1 \cdot v = u + v \\ &= \mathbf{0}. \end{aligned}$$

<sup>6</sup>向量空间的公理化定义是 1888 年由 Peano 引进的, 但当时不太被人所知. 1918 年 Weyl 用公理化方式定义了有限维实向量空间. 1920 年 Banach 引入了实完备赋范空间的公理化定义, 其中前 13 条是现在向量空间的定义

(2) 加法逆元唯一, 且为  $(-1)v$ :

设  $u_1, u_2$  都是  $v$  的加法逆元. 则

$$u_2 = (u_1 + v) + u_2 = u_1 + (v + u_2) = u_1.$$

$$(-1) \cdot v + v = (-1) \cdot v + 1 \cdot v = [(-1) + 1] \cdot v = 0 \cdot v = 0.$$

(3)

$$u + w = v + w \implies u = u + w + (-w) = v + w + (-w) = v.$$

(4) 若  $\lambda \neq 0$  且  $\lambda v = 0$ , 则

$$v = \frac{1}{\lambda} \lambda v = \frac{1}{\lambda} 0 = 0.$$

□

例 38.5. 我们之前在讨论 Hill 密码时使用的矩阵实际上是  $\mathbb{Z}_{29}$  上的线性代数.

例 38.6. 四元数  $x + yi + zj + wk$  ( $x, y, z, w \in \mathbb{R}$ ) 可以看成是一个 4 维的实数向量  $(x, y, z, w)$ . 从而所有四元数组成一个 4 维的实向量空间  $\mathbb{H}$ .

$i, j, k$  是虚数单位, 即

$$i^2 = j^2 = k^2 = -1.$$

设  $ij = a + bi + cj + dk$ . 但考虑到  $1, i, j, k$  彼此垂直,  $ij$  与  $i, j$  都垂直, 所以  $b = c = 0$ , 因此  $ij = a + dk$ .

两边左乘  $i$ , 得到  $-j = ai + dijk$ . 考虑到  $j$  和  $ik$  都与  $i$  垂直, 所以  $a = 0$ , 从而  $-j = dijk$ . 再由  $i, j, k$  长度都为 1, 所以  $d = 1$  (其实,  $d = -1$  也是可以的, 此时  $k$  取相反方向即可). 因此我们得到

$$ik = -j, \quad ij = k.$$

由此得

$$kj = ij^2 = -i, \quad jk = -ik^2 = i, \quad ji = j^2k = -k, \quad ki = -k^2j = j.$$

实数与任何数作乘法时可交换,  $i, j, k$  之间的乘法满足以上要求, 所有数的乘法满足结合律. 于是我们就可以进行四元数乘法了:

$$\begin{aligned} & (a + bi + cj + dk)(x + yi + zj + wk) \\ &= (ax - by - cz - dw) + (bx + ay - dz + cw)i \\ & \quad + (cx + dy + az - bw)j + (dx - cy + bz + aw)k \end{aligned}$$

如果把  $x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k}$  当成 4 维实向量  $(x, y, z, w)^T$ , 则上述乘法可以写成矩阵形式

$$\begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}.$$

于是四元数  $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$  也对应于矩阵

$$\begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}.$$

**定理 38.7.**

$$\varphi(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}$$

是  $\mathbb{H}$  到  $\varphi(\mathbb{H})$  的  $\mathbb{R}$ -线性同构, 其中  $\varphi(H)$  是 4 阶实数矩阵组成的向量空间的一个线性子空间.

我们也可以把四元数  $x + y\mathbf{i} + z\mathbf{j} + w\mathbf{k}$  写成  $(x + y\mathbf{i}) + (z + w\mathbf{i})\mathbf{j}$ , 并看成一对复数  $(u, v) = (x + y\mathbf{i}, z + w\mathbf{i})$ . 于是  $\mathbb{H}$  也可以看成复数域  $\mathbb{C}$  上的二维向量空间. 类似用实数对表示复数那样, 定义加法

$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2),$$

乘法的定义稍显复杂<sup>7</sup>:

$$(u_1, v_1) \cdot (u_2, v_2) = (u_1 u_2 - \bar{v}_1 v_2, \bar{u}_1 v_2 + u_2 v_1), \quad (14)$$

这是因为它需要与共轭

$$\overline{(u, v)} = (\bar{u}, -v)$$

和以下类似复数的模的性质  $u\bar{u} = |u|^2$  协调一致

$$(u, v)(\bar{u}, -v) = (u\bar{u} - \bar{v}(-v), \bar{u}(-v) + v\bar{u}) = (|u|^2 + |v|^2, 0).$$

于是 (14) 可以看成关于  $\mathbb{C}^2$  中向量的线性变换

$$\begin{bmatrix} u_1 & -\bar{v}_1 \\ v_1 & \bar{u}_1 \end{bmatrix} \begin{bmatrix} u_2 \\ v_2 \end{bmatrix}$$

<sup>7</sup> 详见《普林斯顿数学指南 (第一卷)》III.76

此时四元数  $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$  可以视同于矩阵

$$\begin{bmatrix} a + b\mathbf{i} & -c + d\mathbf{i} \\ c + d\mathbf{i} & a - b\mathbf{i} \end{bmatrix}.$$

**定理 38.8.**

$$\psi(a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = \begin{bmatrix} a + b\mathbf{i} & -c + d\mathbf{i} \\ c + d\mathbf{i} & a - b\mathbf{i} \end{bmatrix}$$

是  $\mathbb{H}$  到  $\psi(\mathbb{H})$  的一个  $\mathbb{C}$ -线性同构, 其中  $\psi(\mathbb{H})$  是 2 阶复数矩阵组成的复 4 维向量空间的一个线性空间.

物理学 (特别是量子力学) 中一个重要的矩阵集合是

$$SU(2) = \left\{ \begin{bmatrix} u & -\bar{v} \\ v & \bar{u} \end{bmatrix} \middle| u, v \in \mathbb{C}, |u|^2 + |v|^2 = 1 \right\},$$

它构成一个群 (即成员之间可定义乘法, 有单位元, 每个成员有逆元). 在数学中, 群是关于对称的一个数学概念. 在量子力学中,  $SU(2)$  群用于描述粒子自旋的对称性.

## 附录1: 线性代数的历史

本节是 Israel Kleiner 所著《A history of abstract algebra》一书和 Robyn Arianrhod 所著《Vector: a surprising story of space, time, and mathematical transformation》的读书笔记.

### 线性方程组与行列式

时间	人物	事项
4000多年前	古巴比伦人	$2 \times 2$ 线性方程组
公元前250年	中国人	《九章算术》 $3 \times 3$ 线性方程组， 矩阵（只表示系数，没有运算）， 使用了消元法
1693年	Leibniz	发明行列式，组合方式定义，但 不为时人所知
1750年	Maclaurin	2 阶 3 阶行列式，关于行列式的 最早的出版物
1750年	Cramer Euler	Cramer 法则， $n \times n$ 线性方程组 发现 $n \times n$ 线性方程组未必有唯 一解
1772年	Vandermonde	关于行列式的研究
1772年	Laplace	推广了 Vandermonde 的工作， 行列式的余子式展开
18世纪，关于方程组的讨论都是在行列式框架下，所以只讨论 $n \times n$ 的线性方程组		
1801年	Gauss	首次使用“行列式”的说法
1811年	Gauss	发明最小二乘法，消元法，方程 个数 $\neq$ 未知量个数
1815年	Cauchy	系统研究了行列式， $\det(AB) = \det A \det B$ ，提供了研究 $n$ 维代 数、几何、分析的强大的代数工 具
1860年代	Weierstrass, Kronecker	行列式的公理化定义，他们的工 作直到1903年才为人所知
19世纪后半叶，关于线性方程组的研究是受化简二次型和双线性型推动的. 行列式在一些领域中的应用：判断两个多项式是否具有相同零点， 坐标变换化简代数表达式（例如二次型）、重积分换元、天体力学 19世纪，关于行列式的研究盛极一时， 但在20世纪却销声匿迹——线性代数的主要结论的证明已无需行列式.		

### 矩阵和线性变换

时间	人物	事项
公元前250年	中国人	《九章算术》 $3 \times 3$ 线性方程组，矩阵（只表示系数，没有运算），使用了消元法
17-18 世纪		解析几何，坐标变换 $m = n \leq 3$
17-19 世纪初		射影几何
1801年	Gauss	把矩阵作为线性变换，隐晦地引入. 隐晦地引入了 $2 \times 2$ 、 $3 \times 3$ 矩阵乘法，作为线性变换复合
	Eisenstein, Hermite	同单个字母表示线性变换，定义了加法、乘法（复合）
1850年, 1858年	Cayley	受四元数启发， $m \times n$ 矩阵，矩阵代数形式化，矩阵加法、乘法、数乘，单位矩阵和逆矩阵
1850年	Sylvester	命名“矩阵”
1858年	Cayley	Cayley-Hamilton 定理（ $n = 2, 3$ , Cayley认为没必要对一般情况给出形式证明；Hamilton 证明了 $n = 4$ ，四元数） Cayley 把矩阵当作代数对象的思想极其重要，但他的工作在 1880 年前不被英国以外的人所知
1820-1870年	Cauchy, Jacobi, Jordan, Weierstrass	深入研究了矩阵，创造了谱理论，对称、正交、酉矩阵，特征值和矩阵标准型理论，Jordan 标准型（Weierstrass, Jordan）
1878年	Frobenius	双线性型的标准型，“矩阵历史上的重要地标：他把 Cauchy, Jacobi, Weierstrass 的谱理论与 Eisenstein, Hermite, Cayley 的符号传统结合起来”. 把矩阵用于群表示论，对代数学后续发展起到重要作用
1960年代	杰瑞·萨尔顿	向量、矩阵被用于编程搜索引擎
1990年代末	拉里·佩奇，谢尔盖·布林	向量、矩阵被用于谷歌网页排名算法



### 线性无关、基、维数、向量空间

时间	人物	事项
1572年	拉斐尔·邦贝利（意）	《代数学》开创了复数的数学
17 世纪初	Descartes, Fermat	解析几何
1619年	托马斯·哈利奥特（英）	矢量概念，平行四边形法则
17世纪		向量的概念出现在物理中
18 世纪	Euler	解析几何，线性齐次微分方程的解
18 世纪	Lagrange	线性非齐次微分方程的解
1797年-1831年	Wessel, Gauss	复数的几何表示
19 世纪初	Monge	现代形式的解析几何
1835年	Hamilton	复数作为数对，运算及运算律
1843年	Hamilton	四元数和矢量
1840年代	Cayley, Hamilton, Grassmann	任意维数向量
1844年	Grassmann	$n$ 维向量空间，子空间，张成集 (span)，线性无关，基，维数，线性变换， $\dim(V + W) = \dim V + \dim W - \dim(V \cap W)$ . 但不被数学界接受，直到1862年新版.
1871年	Dedekind	代数数
1880年	Gibbs, Heaviside	3 维向量分析
直到 1880 年，线性代数很多基本结论已建立，但没有一个一般的理论		
1888年	Peano	向量空间的公理化定义，但被人们忽视了——公理化方法尚处婴儿期
1893年	Weber	有限群和有限域
1910年	Steinitz	域的代数扩张
1918年	Weyl	有限维实向量空间的公理化定义
1920年	Banach	完备的赋范实向量空间
1921年	Noether	定义了比向量空间更一般的模
1920年代	Artin	线性 Galois 理论
1930年	van der Waerden	在近世代数书里包含了一章“线性代数”，这个词第一次在数学文献中以现代意义出现
1981年	美国航空航天局	四元数被常规用于航天器导航
21世纪初		四元数被用于机器人技术、计算机生成图像、分子动力学、手机平面旋转、航空器控制

## 附录2：用 Excel 做向量矩阵的数值运算

向量和矩阵的数值计算过程是初等的，但当维数较大时，计算过程还是比较繁琐而无趣的。当矩阵或向量的维数不超过 20 时，可以利用 Excel 代替手工进行计算。

### 1 向量和矩阵的存储

选中一个矩形区域，按行列填写矩阵的元素。一行或一列的矩阵就是向量。

### 2 向量和矩阵的线性组合

$$= \text{倍数1} * \text{矩阵1} + \text{倍数2} * \text{矩阵2}$$

选中输出结果的区域，然后填写上述命令，执行命令时需要同时按 Ctrl-Shift-Enter 键。两个矩阵必须是同样格式的。

### 3 向量内积

**Sumproduct** (Sum “求和”，product “乘积”)

$$= \text{Sumproduct}(\text{向量1}, \text{向量2})$$

这里两个向量必须同时为行向量，或者同时为列向量。

**SumSQ** (Sum “求和”，SQ “平方” square 的字头)

$$= \text{SumSQ}(\text{向量})$$

求一个向量的坐标平方和。

Sumproduct 和 SumSQ 也可以用于矩阵。

### 4 矩阵乘法

**MMult** (首字母 M 表示矩阵，Mult 是英文单词“乘法” Multiplication 的词头)

$$= \text{MMult}(\text{矩阵1}, \text{矩阵2})$$

需要为乘积矩阵预选出相应的表格区域，然后输入上述命令，执行命令时需要同时按 Ctrl-Shift-Enter 键。

### 5 矩阵转置

**Transpose**

$$= \text{Transpose}(\text{矩阵})$$

需要为乘积矩阵预选出相应的表格区域，然后输入上述命令，执行命令时需要同时按 Ctrl-Shift-Enter 键。

### 6 矩阵的逆

**Minverse** (首字母 M 表示矩阵，inverse 是英文单词“逆”)

$$= \text{Minverse}(\text{矩阵})$$

需要为乘积矩阵预选出相应的表格区域，然后输入上述命令，执行命令时需要同时按 Ctrl-Shift-Enter 键。

## 7 矩阵的行列式

**Mdeterm**（首字母 M 表示矩阵，determ 是英文单词“行列式”determination 的词头）

= Mdeterm(矩阵)

选择一个单元格，然后输入上述命令，执行命令时按 Enter 键。

## 8 单位矩阵

**Munit**（首字母 M 表示矩阵，unit 是英文单词“单位”）

= Munit(维数)

选择一个表格区域，然后输入上述命令，执行命令时需要同时按 Ctrl-Shift-Enter 键。