# 《线性代数I》 讲义

薛博卿

上海科技大学

2023秋

MATH1112.01/MATH1455.01

# 目录

# 第0章　背景与例子

艺术源于生活，数学也是如此. 人类在生产生活中积累了大量实际问题，逐渐地抽象出数学概念、方法与思想，进而形成越来越完善的数学理论。而理论的实践与应用也给人们带来巨大的帮助，推动了社会的发展.

在线性代数课程中，有三个基本的研究对象. 第一是线性方程组，它们来源于具体的问题. 第二是线性空间，这是与线性方程组紧密相伴的数学概念，是一种精美的数学结构，事实上，我们就生活在一个三维的空间里. 第三则是线性变换，它建立起线性空间之间的相互联系，也使我们生活的世界动了起来，丰富多彩。


(a) 刚体旋转　　(b) 人脸识别/图像旋转

我们以平面上的旋转为例. 刚体旋转是物体运动的一种基本方式. 而在人脸识别过程中，为了识别侧脸或斜脸，往往要对拍摄到的画面作一些修正，也会涉及旋转. 我们考虑以下简化后的问题：

**例 0.1.** *在直角坐标平面$xOy$中，一个处于$(x,y)$处的质点绕原点逆时针转动了$\theta_0$ $(0\le\theta_0<2\pi)$弧度后，新位置的坐标$(u,v)$ 是多少？反之，若已知新位置的坐标$(u,v)$，那么原位置的坐标$(x,y)$又是多少？*

*分析.* 在平面直角坐标系中画出$(x,y)$与$(u,v)$，当我们这样做时，就已经使用了线性空间的概念. 记

$$\mathbb{R}^2=\{(x,y):\ x,y\in\mathbb{R}\},$$

每一个数对$(x,y)$与$xOy$平面上的一点一一对应.

由于在旋转的过程中质点与原点的距离是保持不变的，使用极坐标表示（如图）更加合适. 假设原位置的极坐标为$(r,\theta)$，则新位置的极坐标为$(r,\theta+\theta_0)$. 它们满足关系式

$$\begin{cases} x=r\cos\theta, \\ y=r\sin\theta \end{cases},\qquad \begin{cases} u=r\cos(\theta+\theta_0), \\ v=r\sin(\theta+\theta_0). \end{cases} \tag{1}$$

利用三角函数公式

$$\cos(\theta+\theta_0)=\cos\theta\cos\theta_0-\sin\theta\sin\theta_0,\quad \sin(\theta+\theta_0)=\sin\theta\cos\theta_0+\cos\theta\sin\theta_0,$$

我们将(1)式中的辅助变量$r$和$\theta$代换，得到$(x,y)$与$(u,v)$的关系如下：

$$\begin{cases} (\cos\theta_0)\,x\ -\ (\sin\theta_0)\,y\ =\ u, \\ (\sin\theta_0)\,x\ +\ (\cos\theta_0)\,y\ =\ v. \end{cases} \tag{2}$$

当$\theta_0$为给定常数时，$\cos\theta_0$和$\sin\theta_0$均为常数. 对第二个问题而言，$u, v$也是已知的，所以(2)式是关于$x, y$的二元一次方程组，通过解该方程组，便可得到原坐标$(x, y)$.

从几何上看，式(2)中的每个方程都表示$xOy$面上的一条直线，而方程组的解对应于同时位于两条直线上的点. 而平面上的两条直线共有三种位置关系：相交、平行或重合. 当$\theta_0 = 0$或$\pi$时，它们包含一条竖线和一条横线. 而当$\theta_0 \neq 0, \pi$时，注意到

$$(\cos\theta_0)(\cos\theta_0) - (-\sin\theta_0)(\sin\theta_0) = 1 \neq 0, \tag{3}$$

不会出现

$$\frac{\cos\theta_0}{\sin\theta_0} = \frac{-\sin\theta_0}{\cos\theta_0}$$

的情形，即直线方程的一次项系数不成比例，两条直线相交于一点. 这一点的坐标便是方程组(2)的唯一解.

对于第一个问题而言，旋转可以看成$xOy$平面到自身的一个映射. 用$R_{\theta_0}$表示绕原点逆时针转动$\theta_0$弧度，根据(2)式，映射$R_{\theta_0} : \mathbb{R}^2 \to \mathbb{R}^2$满足

$$R_{\theta_0}(x, y) = (u, v) = \big((\cos\theta_0)x - (\sin\theta_0)y,\ (\sin\theta_0)x + (\cos\theta_0)y\big).$$

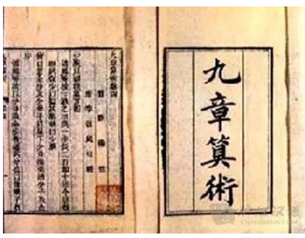
这里的$\mathbb{R}^2$是线性空间，而$R_{\theta_0}$则是线性空间之间的一个线性变换.

回到第二个问题，从新坐标到原坐标是否为一个“相反”的旋转过程？事实上，顺时针旋转$\theta_0$弧度等同于逆时针转动了$-\theta_0$弧度，因此

$$\begin{aligned}(x, y) = R_{-\theta_0}(u, v) = &\ \big((\cos(-\theta_0))u - (\sin(-\theta_0))v,\ (\sin(-\theta_0))u + (\cos(-\theta_0))v\big)\\ = &\ \big((\cos\theta_0)u + (\sin\theta_0)v,\ -(\sin\theta_0)u + (\cos\theta_0)v\big).\end{aligned}$$

到此我们已经回答完了例0.1中的问题. 如果更进一步思考一下，这个例子里的(3)式是否起了很关键的作用？除了旋转之外的运动是否也存在某种与之“相反”的运动呢？ □

通过以上例子可以看出，线性变换通过一些线性方程，建立了线性空间之间的关联. 如何刻画一个线性变换的结构或是变换效果，是线性代数理论中的主要课题之一.





早在西汉年间，由张苍、耿寿昌定稿的《九章算术》中就有这样一个问题：“今有上禾三秉，中禾二秉，下禾一秉，实三十九斗；上禾二秉，中禾三秉，下禾一秉，实三十四斗；上禾一秉，中禾二秉，下禾三秉，实二十六斗；问上、中、下禾实一秉各几何？”

我们用$x, y, z$分别表示上、中、下禾的秉数，就得到线性方程组

$$\left\{\begin{array}{rrrl} 3x & +2y & +z & = 39 \\ 2x & +3y & +z & = 34 \\ x & +2y & +3z & = 26 \end{array}\right. \tag{4}$$

在现代，这样的一个线性方程组并不难解，然而科技发展的突飞猛进也意味着我们会遇到规模非常庞大的方程组，如何找到一般性的方法，并形成理论，是线性代数这门课程的任务之一. 一个自然而然的尝试是，找出方程组中的核心数据，刨除不重要的信息，这是一个非常重要的数学思想. 事实上，古人就是这么做的：

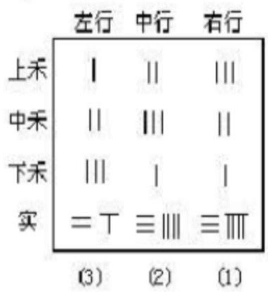



图 1: 《九章算术》中的解题过程

未知变量用$x, y, z$表示，还是用$x_1, x_2, x_3$表示，并无本质区别. 但只要看到上图的数字和它们所在位置，我们就可以重新写出方程组(4). 当方程规模越来越大时，上图的形式显然比(4)的写法简洁明了许多.

将古人的书写习惯更换为现代的书写习惯，我们可以将图1写为如下形式：

$$\left[\begin{array}{ccc:c} 3 & 2 & 1 & 39 \\ 2 & 3 & 1 & 34 \\ 1 & 2 & 3 & 26 \end{array}\right].$$

我们将以上的表达式称为矩阵. 矩阵也是线性代数的一个基本研究对象，没有将它列在三者之中的原因是：矩阵就是线性变换，而当我们在有限维度进行讨论时，线性变换也可以写成矩阵.

# 第1章　线性方程组与矩阵

## 1.1　线性方程组与矩阵

在本讲义中，我们用$\mathbb{F}$表示实数域$\mathbb{R}$或复数域$\mathbb{C}$，并将其中的数称为标量. 在讨论每个问题之前，$\mathbb{F}=\mathbb{R}$或$\mathbb{F}=\mathbb{C}$就已经选定，在讨论的中途是不可以更换的.

包含$n$个未知变量的$m$个方程组成的线性方程组可以写成以下形式：

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1, \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2, \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m, \end{cases} \tag{1.1}$$

其中$a_{ij}\in\mathbb{F}$是给定的常数. 我们也可以将(1.1)写成

$$\sum_{j=1}^{n} a_{ij}x_j=b_i, \quad (1\le i\le n).$$

如果将$s_1,s_2,\ldots,s_n\in\mathbb{F}$分别代入未知元$x_1,x_2,\ldots,x_n$所在的位置，使得(1.1)成立的话，我们就称

$$x_1=s_1,x_2=s_2,\ldots,x_n=s_n$$

是方程组(1.1)的一个解. 为了方便，有时也将这样的解记成$n$元有序数组 $(s_1,s_2,\ldots,s_n)$.

例0.1中的由两个方程组成的二元一次方程组，对应了两条直线相交的情形，这时方程组有唯一解. 当两条直线平行时，方程组无解. 而当两条直线重合时，方程组有无穷多组解. 在本讲义中，我们将使用如下术语：若一个线性方程组有解，我们称其为相容的，否则称其为不相容的.

接下来，我们提取方程组(1.1)中的核心信息:

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \qquad \left[\begin{array}{cccc:c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right]$$

前者被称为方程组(1.1)的系数矩阵，而后者被称为方程组(1.1)的增广矩阵. 后者中的虚线是可以省去的，画上它仅仅是为了一目了然.

更一般地，将一些对象按照长方阵列排列，所得到的便被称为矩阵. 矩阵中的对象则称为元素. 我们用$M_{m\times n}(\mathbb{F})$表示形如

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

的$m\times n$阶$\mathbb{F}$-矩阵全体所构成的集合，这里的$m\times n$表示矩阵具有$m$行、$n$列，而矩阵中的元素$a_{ij}$均来自于$\mathbb{F}$. 当$\mathbb{F}=\mathbb{R}$（或$\mathbb{C}$）时，可直接称矩阵为实矩阵（或复矩阵）. 当行列数相同，即$m=n$时，我们又将矩阵称为$n$阶方阵，我们用$M_n(\mathbb{F})$ 表示所有的$n$阶$\mathbb{F}$-方阵全体所组成的集合. 我们亦将$n\times 1$阶矩阵称为$n$阶列矩阵或$n$维列向量，将$1\times n$ 阶矩阵称为$n$阶行矩阵或$n$维行向量.

**例 1.1.** 以下均为实矩阵：从左到右依次可称为$3\times 2$阶矩阵、4阶行矩阵、3阶方阵、2维列向量、1阶方阵.

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 4 \end{bmatrix},\quad \begin{bmatrix} 2 & 1 & 0 & -3 \end{bmatrix},\quad \begin{bmatrix} e & \pi & -\sqrt{2} \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix},\quad \begin{bmatrix} 1 \\ 3 \end{bmatrix},\quad \begin{bmatrix} 4 \end{bmatrix}$$

特别地，一阶$\mathbb{F}$-方阵$[a]$与$\mathbb{F}$ 中的数$a$在数学上并没有本质区别，在大多数时候我们认为它们是相同的.

除了线性方程组的系数矩阵与增广矩阵之外，在很多其它场合，矩阵都起到了记录数据的作用.

**例 1.2.** 以下三个分数表

| 作业 | 代数 | 分析 |
|---|---|---|
| 学生1 | 100 | 100 |
| 学生2 | 95 | 90 |
| 学生3 | 70 | 65 |

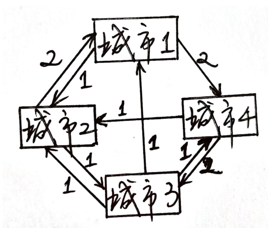
| 期中 | 代数 | 分析 |
|---|---|---|
| 学生1 | 91 | 94 |
| 学生2 | 68 | 85 |
| 学生3 | 47 | 32 |

| 期末 | 代数 | 分析 |
|---|---|---|
| 学生1 | 72 | 68 |
| 学生2 | 64 | 70 |
| 学生3 | 26 | 63 |

可以用三个矩阵分别表示：

$$A=\begin{bmatrix} 100 & 100 \\ 95 & 90 \\ 70 & 65 \end{bmatrix},\quad B=\begin{bmatrix} 91 & 94 \\ 68 & 85 \\ 47 & 32 \end{bmatrix},\quad C=\begin{bmatrix} 72 & 68 \\ 64 & 70 \\ 26 & 63 \end{bmatrix}.$$

**例 1.3.** 假设四个城市之间每天直达航班的数量由以下加权有向图（见左图）表示，箭头指出了航班起飞和降落的城市，而箭头上的数字记录了当天的不同航班数. 这些数据亦可以列成表格，以矩阵的形式记




(a) 加权图

| ?→ | 城市1 | 城市2 | 城市3 | 城市4 |
|---|---|---|---|---|
| 城市1 | 0 | 1 | 0 | 2 |
| 城市2 | 2 | 0 | 1 | 0 |
| 城市3 | 1 | 1 | 0 | 1 |
| 城市4 | 0 | 1 | 2 | 0 |

(b) 邻接矩阵

录. 右侧的矩阵称为左图的邻接矩阵，第$i$行第$j$列的元素表示当天第$i$个城市发往第$j$个城市的直达航班数量.

## 1.2 高斯消元法

让我们回到解线性方程组来，要想解一个较为复杂的方程组，我们采取的数学思想为：在不改变方程组的解的前提下，想办法简化方程组.

我们用以下线性方程组为例，作一些尝试.

$$\begin{cases} 2x+4y+5z &= 4, & (1.2)\\ 99x+99y+198z &= 198, & (1.3)\\ 3x-y+7z &= 0. & (1.4)\end{cases}$$

观察到式(1.3)中的系数相对较大，将整个式子除以99，方程组的解是不会改变的. 于是得到

$$\begin{cases} 2x+4y+5z &= 4, & (1.5)\\ x+y+2z &= 2, & (1.6)\\ 3x-y+7z &= 0. & (1.7)\end{cases}$$

如果可以消去方程中的一个未知数，得到含有未知数更少的方程，对解方程是有利的. 因此我们将式(1.5)、(1.7)分别减去式(1.6)的2倍、3倍，可得

$$\begin{cases} 2y+z &= 0, & (1.8)\\ x+y+2z &= 2, & (1.9)\\ -4y+z &= \text{-}6. & (1.10)\end{cases}$$

接下来，我们交换一下式(1.8)与式(1.9)的位置，即

$$\begin{cases} x+y+2z &= 2, & (1.11)\\ 2y+z &= 0, & (1.12)\\ -4y+z &= \text{-}6. & (1.13)\end{cases}$$

直到这时，整个方程组的解没有被改变，而式(1.12)、(1.13)组成了一个二元一次方程组，这样就达到了降阶简化的目的. 我们对式(1.12)、(1.13)继续上述的简化过程，将式(1.13)加上式(1.12)的2倍，得

$$\begin{cases} x+y+2z &= 2, & (1.14)\\ 2y+z &= 0, & (1.15)\\ 3z &= \text{-}6. & (1.16)\end{cases}$$

这时，可从方程(1.16)解得

$$z=\frac{-6}{3}=-2. \tag{1.17}$$

代入方程(1.15)后，解得

$$y=\frac{-z}{2}=1. \tag{1.18}$$

再一起代入方程(1.14)，得

$$x=2-y-2z=5. \tag{1.19}$$

故原方程组的解为$(5,1,-2)$.

式(1.2)-(1.16)中所展示的化简方法被称为高斯消元法，而式(1.17)-(1.19)则被称为回代过程. 一般而言，常用的线性方程组的初等变换有以下三类：

I 在某个方程上乘以非零常数;

II 交换两个方程的位置;

III 将某个方程的常数倍加到另外一个方程上.

在第一类变换中，在某个方程上乘以0相当于直接去掉了这个方程，乘“非零”常数才能保证原方程组的解不变. 在上述例子中，如果我们不交换式(1.8)、(1.9)的位置，其实对解方程并没有影响. 但是当方程的个数上升到成百上千时，是有必要用第II类初等变换整理一下次序的. 我们习惯把变量个数多的方程置于较上方的位置，把降阶后所得的方程置于较下方的位置. 从数学理论的角度而言，这样的方式更具有结构性，而从应用的角度而言，也更有利于在计算机上编写程序.

事实上，回代过程也可由初等行变换来实现. 我们将式(1.15)、(1.16)分别除以2、3，得

$$\begin{cases} x+y+2z &= 2, & (1.20)\\ \quad\ y+\frac{1}{2}z &= 0, & (1.21)\\ \qquad\quad z &= \text{-}2. & (1.22)\end{cases}$$

进一步，再令式(1.20)、(1.21)分别减去(1.21)式的2倍、1/2倍，得

$$\begin{cases} x+y &= 6, & (1.23)\\ \quad\ y &= 1, & (1.24)\\ \qquad\quad z &= \text{-}2. & (1.25)\end{cases}$$

最后，在式两端同时减去式，得

$$\begin{cases} x &= 5, & (1.26)\\ \quad\ y &= 1, & (1.27)\\ \qquad\quad z &= \text{-}2. & (1.28)\end{cases}$$

这时，我们直接得到了方程组的解$(5,1,-2)$. 式(1.2)-(1.16),以及式(1.20)-(1.28)中所展示的方法合称高斯-约当消元法.

由于将方程组写成增广矩阵会更加简洁，我们对应地定义矩阵的三类初等列变换：

I 矩阵的某一行乘以非零常数;

II 交换矩阵的两行;

III 将矩阵某一行的常数倍加到另外一行上.

具体形式如下：

$$\begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \xrightarrow{c\cdot(\text{行}i)} \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ ca_{i1} & ca_{i2} & \cdots & ca_{in} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

$$\begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \xrightarrow{\text{交换行}i,j} \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

$$\begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \cdots & \cdots & \cdots & \cdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \xrightarrow{\text{行}i+c\cdot(\text{行}j)} \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ a_{i1}+ca_{j1} & a_{i2}+ca_{j2} & \cdots & a_{in}+ca_{jn} \\ \cdots & \cdots & \cdots & \cdots \\ a_{j1} & a_{j2} & \cdots & ca_{jn} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

另外，将“行”替换为“列”所对应的变换称为矩阵的初等列变换.

现在我们将前一个例子的求解过程用该线性方程组的增广矩阵的初等行变换展示出来：

$$
\left[\begin{array}{ccc:c} 2 & 4 & 5 & 4 \\ 99 & 99 & 198 & 198 \\ 3 & -1 & 7 & 0 \end{array}\right]
\xrightarrow{(1/99)\cdot(\text{行}2)}
\left[\begin{array}{ccc:c} 2 & 4 & 5 & 4 \\ 1 & 1 & 2 & 2 \\ 3 & -1 & 7 & 0 \end{array}\right]
\xrightarrow[\text{行}3-3\cdot(\text{行}2)]{\text{行}1-2\cdot(\text{行}2)}
\left[\begin{array}{ccc:c} 0 & 2 & 1 & 0 \\ 1 & 1 & 2 & 2 \\ 0 & -4 & 1 & -6 \end{array}\right]
$$

$$
\xrightarrow{\text{交换行}1,2}
\left[\begin{array}{ccc:c} 1 & 1 & 2 & 2 \\ 0 & 2 & 1 & 0 \\ 0 & -4 & 1 & -6 \end{array}\right]
\xrightarrow{\text{行}3+2\cdot(\text{行}2)}
\left[\begin{array}{ccc:c} 1 & 1 & 2 & 2 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 3 & -6 \end{array}\right]
$$

$$
\xrightarrow[(1/3)\cdot(\text{行}3)]{(1/2)\cdot(\text{行}2)}
\left[\begin{array}{ccc:c} 1 & 1 & 2 & 2 \\ 0 & 1 & 1/2 & 0 \\ 0 & 0 & 1 & -2 \end{array}\right]
\xrightarrow[\text{行}2-(1/2)\cdot(\text{行}3)]{\text{行}1-2\cdot(\text{行}3)}
\left[\begin{array}{ccc:c} 1 & 1 & 0 & 6 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array}\right]
\xrightarrow{\text{行}1-\cdot(\text{行}2)}
\left[\begin{array}{ccc:c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array}\right].
$$

使用高斯消元法对方程组进行变换的最后，后一个方程的变量个数（即系数非零的变量个数）总是严格小于前一个方程的变量个数. 对应地，我们定义矩阵的阶梯型如下.

**定义 1.1.** *我们称一个矩阵$R$为（行）阶梯阵，是指其满足以下两个条件:*

*(i) 零行均位于非零行的下方;*

*(ii) 下方的非零首元总在上方的非零首元右侧.*

*若矩阵$A$可通过初等行变换化为阶梯阵$R$，则我们称$A$具有阶梯型$R$.*

这里非零首元指的是一个非零行中从左到右第一个出现的非零元素. 我们称其所在的位置为首元位置，而它所对应的变量称为主变元或主变量.

**例 1.4.** 以下为阶梯阵的示例：

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

而高斯-约当消元法将下方出现的主变元在上方方程中进一步消去，最后化为形如

$$
\left\{\begin{array}{llllr} x_{k_1} & & & + \sum(\ ) = & b_1' \\ & x_{k_2} & & + \sum(\ ) = & b_2' \\ & \cdots & \cdots & \cdots \quad \cdots & \\ & & x_{k_r} & + \sum(\ ) = & b_r' \\ & & & 0 = & b_{r+1}' \\ & & & \cdots\cdots & \\ & & & 0 = & b_m' \end{array}\right. \tag{1.29}
$$

的形式，从它较容易直接写出方程组的所有解. 对应地，我们定义矩阵的简约阶梯型如下：

**定义 1.2.** *我们称一个（行）阶梯阵$R$为（行）简约阶梯阵，是指其满足以下两个条件:*

*(iii) 非零首元均为1;*

*(iv) 非零首元所在的列中，其他元素均为0.*

*若矩阵$A$可通过初等行变换化为简约阶梯阵$R$，则我们称$A$具有简约阶梯型$R$.*

**例 1.5.** 以下为简约阶梯阵的示例：

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

**定理 1.6.** *（1）任何矩阵都具有阶梯型，也都具有简约阶梯型.*

*（2）给定矩阵的任何一个阶梯型都具有相同的首元位置.*

*（3）给定矩阵的简约阶梯型是唯一的.*

上述定理可以用归纳法证明，留给有兴趣的读者自行完成.

**例 1.7.** *试用不同方法解线性方程组*

$$\begin{cases} x_1+3x_2-2x_3+2x_5=0\\ 2x_1+6x_2-5x_3-2x_4+4x_5-3x_6=-1\\ 5x_3+10x_4+15x_6=5\\ 2x_1+6x_2+8x_4+4x_5+18x_6=6 \end{cases}$$

*(1)高斯消元法与回代过程;*

*(2)高斯-约当消元法.*

解.（略） □

在一个线性方程组中，如果存在不是主变元的变元，我们将其称为自由变元或自由变量. 从以上例子可以看出，自由变元可以在$\mathbb{F}$中随意取值，而主变元的取值由自由变元的值所确定. 当一个线性方程组有无穷多解时，由参数形式给出所有解的公式称为该线性方程组的一个通解.

## 1.3 矩阵的秩

假设一个线性方程组的增广矩阵$\hat{A}$具有阶梯型$R$，而$R$的底部有一些零行，这会意味着什么？这些零行均对应方程$0=0$. 这样的方程并不起任何作用，也就是说，原方程组中具有一些冗余的方程. 比如，将一个线性方程复制一百遍，放在一起组成方程组，那么有效的仅仅只有原先的那个方程.

方程组中有效方程的个数是一个重要的量，它等于增广矩阵的阶梯型中非零行的数量.

**定义 1.3.** *若矩阵$A$的一个阶梯型$R$具有$r$个非零行，则我们称矩阵$A$的秩为$r$，记为$\mathrm{rank}(A)=r$.*

不难得到，若$A\in M_{m\times n}(\mathbb{F})$，则 $\mathrm{rank}(A)\le\min\{m,n\}$. 接下来，我们通过简约阶梯阵，将方程解的情况用秩的语言翻译出来.

**定理 1.8.** *若一个线性方程组的增广矩阵具有阶梯型*

$$\left[\begin{array}{cccccc:c} c_{1j} & & & & \cdots & c_{1n} & d_1\\ & c_{2k} & & & \cdots & c_{2n} & d_2\\ & & c_{3l} & & \cdots & c_{3n} & d_3\\ & & & \ddots & & \vdots & \vdots\\ & & & & c_{rs}\ \cdots & c_{rn} & d_r\\ & & & & & 0 & d_{r+1}\\ & & & & & 0 & 0\\ & & & & & \vdots & \vdots\\ & & & & & 0 & 0 \end{array}\right],$$

其中$c_{1j}, c_{2k}, c_{3l}, \ldots, c_{rs} \neq 0$. 则

*(1) 原方程组是不相容的，当且仅当$d_{r+1} \neq 0$；*

*(2) 原方程组是相容的，当且仅当$d_{r+1} = 0$；*

*(3) 原方程组有唯一解，当且仅当$d_{r+1} = 0$且$r = n$；*

*(4) 原方程组是无穷多解，当且仅当$d_{r+1} = 0$且$r < n$.*

**推论 1.9.** *设一个线性方程组的系数矩阵为$A$，而增广矩阵为$\hat{A}$. 则*

*(1) 原方程组是不相容的，当且仅当$\mathrm{rank}(A) < \mathrm{rank}(\hat{A})$；*

*(2) 原方程组是相容的，当且仅当$\mathrm{rank}(A) = \mathrm{rank}(\hat{A})$；*

*(3) 原方程组有唯一解，当且仅当$\mathrm{rank}(A) = \mathrm{rank}(\hat{A}) = n$；*

*(4) 原方程组是无穷多解，当且仅当$\mathrm{rank}(A) = \mathrm{rank}(\hat{A}) < n$.*

现在，我们考虑一种特殊情形——常数项均为0的方程组.

**例 1.10.** 设实数$a, b$为参数，方程组

$$\begin{cases} -2x & +y & +z & = & -2 \\ x & -2y & +z & = & a \\ x & +y & +(b-2)z & = & a^2+b \end{cases}$$

何时不相容？何时有唯一解？何时有无穷多个解？

解.（略.） □

**定义 1.4.** 一个线性方程组被称为齐次的，是指其中的常数项均为零，即

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \quad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases} \tag{1.30}$$

齐次线性方程组(1.30)一定有解，这可以从两方面看：一、$(0, 0, \ldots, 0)$一定是(1.30)的解，我们称其为平凡解；二、常数项中的0，在初等行变换的过程中始终保持为0，因此增广矩阵$\hat{A}$的阶梯型与系数矩阵$A$的阶梯型具有相同的非零行数，满足$\mathrm{rank}(A) = \mathrm{rank}(\hat{A})$, 根据推论1.9可知一定有解.

## 1.4 矩阵的运算

在这一节中，我们抛开方程组与增广矩阵，探究“矩阵”本身的奥秘.

为了方便起见，我们约定一些符号. 当某个$m \times n$阶矩阵的元素$a_{i,j}$ $(1 \leq i \leq m, 1 \leq j \leq n)$已经被明确指出时，我们记该矩阵为

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

或$[a_{ij}]_{m \times n}$. 在不引起混淆的情况下，我们也将它简记为$[a_{ij}]$. 反之，当一个矩阵已经用$A$明确表示，但并未指出其元素时，我们将其第$i$行第$j$列的元素记为$(A)_{ij}$. 我们有时用粗体字母表示行向量和列向量，如

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

对于方阵$A \in M_n(\mathbb{F})$，其行、列下标相同的位置称为（主）对角线，元素$(A)_{11}, (A)_{22}, \ldots, (A)_n n$称为对角元.

现在，我们来探寻矩阵之间的联系. 首先，以下矩阵是否提供了同样的信息？

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

很明显，前两个矩阵中的数不同，第三个矩阵告诉我们的信息量比第二个更多，而第三第四个矩阵尽管都有两个0，但是所处位置不同，代表的意义一般也不一样.

**定义 1.5.** *两个矩阵相等是指它们的阶相同，且对应位置的元素均相同. 更具体地，$[a_{ij}]_{m\times n} = [b_{kl}]_{p\times q}$是指*

$$m = p, \quad n = q, \quad a_{ij} = b_{ij}\,(1 \le i \le m,\, 1 \le j \le n).$$

以下的前三张表格中记录了三位同学的代数和分析的平时成绩、期中成绩与期末成绩.

| 作业 | 代数 | 分析 |
|---|---|---|
| 学生 *1* | 100 | 100 |
| 学生 *2* | 95 | 90 |
| 学生 *3* | 70 | 65 |

| Midterm | 代数 | 分析 |
|---|---|---|
| 学生 *1* | 91 | 94 |
| 学生 *2* | 68 | 85 |
| 学生 *3* | 47 | 32 |

| Final | 代数 | 分析 |
|---|---|---|
| 学生 *1* | 72 | 68 |
| 学生 *2* | 64 | 70 |
| 学生 *3* | 26 | 63 |

| Total | 代数 | 分析 |
|---|---|---|
| 学生 *1* | ? | ? |
| 学生 *2* | ? | ? |
| 学生 *3* | ? | ? |

假设总评成绩的计算方式为

$$\text{总评成绩} \;=\; \text{作业成绩*30\%} \;+\; \text{期中成绩*30\%} \;+\; \text{期末成绩*40\%}.$$

那么第四个表格中的数据是多少？答案显而易见，将对应位置的几个分数乘以比例后相加即可. 事实上，我们可以让整个表格，也就是矩阵，作这样的运算.

**定义 1.6.** *设$A = [a_{ij}]$,$B = [b_{ij}]$满足$A, B \in M_{m\times n}(\mathbb{F})$. 则$A$与$B$的和$A + B$定义为满足以下条件的矩阵：*

*(i)* $A + B \in M_{m\times n}(\mathbb{F})$.

*(ii)* $(A + B)_{ij} = (A)_{ij} + (B)_{ij} = a_{ij} + b_{ij}$, $(1 \le i \le m,\, 1 \le j \le n)$..

*设$c \in \mathbb{F}$，而$A = [a_{ij}] \in M_{m\times n}(\mathbb{F})$. 则$c$与$A$的数乘$cA$定义为满足以下条件的矩阵：*

*(i)* $cA \in M_{m\times n}(\mathbb{F})$.

*(ii)* $(cA)_{ij} = c \cdot (A)_{ij} = ca_{ij}$, $(1 \le i \le m,\, 1 \le j \le n)$..

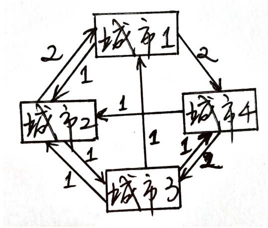
需要注意的是，不同行数或列数的两个矩阵是不能作加法的.

**例 1.11.** *设$2A - 3B = C$，其中*

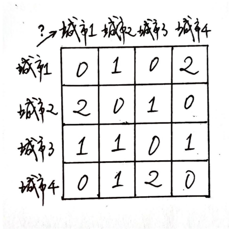
$$A = \begin{bmatrix} 1 & x \\ 3 & y \end{bmatrix}, \quad B = \begin{bmatrix} z & 0 \\ w & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 2 \\ 4 & -2 \end{bmatrix}.$$

*求$x, y, z, w$的值.*

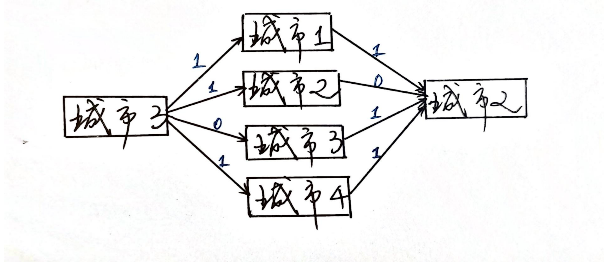
解.（略） □




(c) 加权图




(d) 邻接矩阵



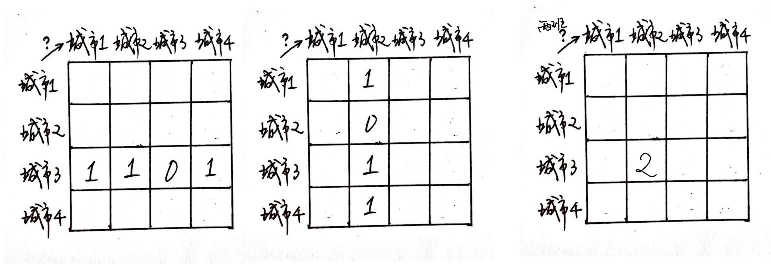

让我们回顾之前的图与邻接矩阵的问题，四个城市之间每天的直达航班数由下列加权图或邻接矩阵矩阵给出.

那么，通过接连两天各坐一次直达航班的方式，从第$i$个城市抵达第$j$个城市的方法有多少种？以从城市3到城市2为例，可以选择从城市1或城市4转机. 为了使得讨论更具一般性，我们以图表示途经所有城市转机的方法数：

不难看出，总方法数为

$$1 \times 1 + 1 \times 0 + 0 \times 1 + 1 \times 1 = 2.$$

如果将上式与邻接矩阵的信息对比，可以发现，正是出发城市对应行中的元素与抵达城市对应列中的元素对应相乘之后求和的结果，即



类似地，从城市4到城市1的总方法数为

$$0 \times 0 + 1 \times 2 + 2 \times 1 + 0 \times 0 = 4.$$

计算所有从城市$i$出发通过接连两天各坐一次直达航班的方式抵达城市$j$的总方法数，列表写成矩阵如下

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

**定义 1.7.** 设$A = [a_{ij}]$, $B = [b_{ij}]$满足$A \in M_{m \times r}(\mathbb{F})$, $B \in M_{r \times n}(\mathbb{F})$. 则$A$与$B$的乘积$AB$定义为满足以下条件的矩阵：

*(i)* $AB \in M_{m\times n}(\mathbb{F})$;

*(ii)* $(AB)_{ij} = \sum_{k=1}^{r} a_{ik}b_{kj}$, $(1 \le i \le m,\ 1 \le j \le n)$.

特别需要强调的是，只有在左侧矩阵的列数等于右侧矩阵的行数时，矩阵才可以相乘. 而乘积所得矩阵的行数与左侧矩阵相同，列数与右侧矩阵相同.

在前面的例子中，

$$\begin{bmatrix} 0 & 1 & 0 & 2 \\ 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 & 2 \\ 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 5 & 0 \\ 1 & 3 & 0 & 5 \\ 2 & 2 & 3 & 2 \\ 4 & 2 & 1 & 2 \end{bmatrix}.$$

**例 1.12.** 试计算$AB$与$BA$，其中

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

解. 略 □

在定义矩阵乘法后，线性方程组

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

就可以写成$A\mathbf{x} = \mathbf{b}$这样更加简明的方式了，这里

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

矩阵还有另一种运算，可以将行转为列，将列转为行.

**定义 1.8.** 设$A \in M_{m\times n}(\mathbb{F})$. 则$A$的转置$A^T$定义为满足以下条件的矩阵:

*(i)* $A^T \in M_{m\times n}(\mathbb{F})$.

*(ii)* $(A^T)_{kl} = (A)_{lk}$, $(1 \le k \le n,\ 1 \le l \le m)$.

即

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

**例 1.13.** 求$A^T$，其中$A = \begin{bmatrix} 2 & 4 \\ 3 & -2 \\ 0 & 1 \end{bmatrix}$.

解. 略 □

**例 1.14.** 令$A=\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$. 试计算$AA^T$与$A^TA$.

解. 略 □

另外，在以后的学习中，大家会看到方阵的对角元之和起着重要的作用.

**定义 1.9.** 设$A\in M_n(\mathbb{F})$. 则$A$的迹定义为

$$tr(A)=\sum_{i=1}^{n}(A)_{ii}.$$

**例 1.15.** 求$tr(A)$，其中$A=\begin{bmatrix} 2 & 4 & 5 \\ 3 & -2 & 0 \\ 0 & 1 & -7 \end{bmatrix}$.

解. 略 □

**例 1.16.** 设$A=[a_{ij}]_{m\times n}$，$B=[b_{ij}]_{n\times m}$. 计算$tr(AB)$与$tr(BA)$.

解. 略 □

## 1.5 分块矩阵

我们已经知道了矩阵乘法的定义，在以下例子中，请观察各种特殊类型的矩阵出现在左、右侧的乘法效果，并归纳总结矩阵乘法的一些运算规律.

**例 1.17.** 计算下列矩阵乘法.

(1) $\begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}$

(2) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$

(3) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$

(4) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$

(5) $\begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$

(6) $\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}$

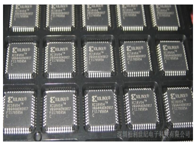
(7) $\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(8) $\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix}$

解. 略 □

正如集成电路上布满了许多子元件，有时为了方便起见，我们用一些水平、竖直的线将矩阵划分为更小的部分，每个部分都称为该矩阵的子矩阵. 例如

$$\left[\begin{array}{c:c:c:c} A_{11} & A_{12} & \dots & A_{1s} \\ \hdashline A_{21} & A_{22} & \dots & A_{2s} \\ \hdashline \dots & \dots & \dots & \dots \\ \hdashline A_{r1} & A_{r2} & \dots & A_{rs} \end{array}\right]$$



**例 1.18.** 试找出对应的子矩阵.

$$A = \left[\begin{array}{c:ccc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = \left[\begin{array}{c:c} A_{11} & A_{12} \\ \hdashline A_{21} & A_{22} \end{array}\right],$$

$$A = \left[\begin{array}{cccc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hdashline a_{21} & a_{22} & a_{23} & a_{24} \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = \left[\begin{array}{c} \mathbf{r}_1 \\ \hdashline \mathbf{r}_2 \\ \hdashline \mathbf{r}_3 \end{array}\right],$$

$$A = \left[\begin{array}{c:c:c:c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = \left[\begin{array}{c:c:c:c} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4 \end{array}\right].$$

利用分块矩阵的概念和记号，我们可以把线性方程组$A\mathbf{x} = \mathbf{b}$的增广矩阵记为$\left[\begin{array}{c:c} A & \mathbf{b} \end{array}\right]$.

如果两个分块矩阵的子矩阵之间的行、列数是匹配的，也可以作相应的分块矩阵乘法. 以两个$2 \times 2$的分块矩阵相乘为例.

**例 1.19.** 设$m = m_1 + m_2$, $r = r_1 + r_2$, $n = n_1 + n_2$. 设

$$A = \left[\begin{array}{c:c} C_{m_1 \times r_1} & D_{m_1 \times r_2} \\ \hdashline E_{m_2 \times r_1} & F_{m_2 \times r_2} \end{array}\right]_{m \times r}, \quad B = \left[\begin{array}{c:c} G_{r_1 \times n_1} & H_{r_1 \times n_2} \\ \hdashline J_{r_2 \times n_1} & K_{r_2 \times n_2} \end{array}\right]_{r \times n}.$$

试证明

$$AB = \left[\begin{array}{c:c} CG + DJ_{m_1 \times n_1} & CH + DK_{m_1 \times n_2} \\ \hdashline EG + FJ_{m_2 \times n_1} & EH + FK_{m_2 \times n_2} \end{array}\right]_{m \times n}.$$

说明.（略.） □

需要强调的是，子矩阵之间相乘的顺序，必须与原矩阵相乘的顺序保持一致.

**例 1.20.** 设 试以不同方式计算$AB$：*(1)*直接计算; *(2)*以分块矩阵的方式计算.

$$A = \left[\begin{array}{ccc|cc} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ \hline 0 & -4 & -2 & 7 & -1 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \left[\begin{array}{cc} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ \hline -1 & 3 \\ 5 & 2 \end{array}\right] = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

解.（略.） □

我们再来观察一下以行、列的方式将矩阵分块. 这些分块作乘法的方式在线性代数中是很常用的.

**例 1.21.** 设$A \in M_{m \times r}$ *and* $B \in M_{r \times n}$. 假设

$$A = \left[\begin{array}{c} \mathbf{a}_1 \\ \hdashline \mathbf{a}_2 \\ \hdashline \vdots \\ \hdashline \mathbf{a}_m \end{array}\right], \quad B = \left[\begin{array}{c:c:c:c} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{array}\right].$$

则有

$$AB = \begin{bmatrix} \mathbf{a}_1\mathbf{b}_1 & \mathbf{a}_1\mathbf{b}_2 & \dots & \mathbf{a}_1\mathbf{b}_n \\ \mathbf{a}_2\mathbf{b}_1 & \mathbf{a}_2\mathbf{b}_2 & \dots & \mathbf{a}_2\mathbf{b}_n \\ \dots & \dots & \dots & \dots \\ \mathbf{a}_m\mathbf{b}_1 & \mathbf{a}_m\mathbf{b}_2 & \dots & \mathbf{a}_m\mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 B \\ \hdashline \mathbf{a}_2 B \\ \hdashline \vdots \\ \hdashline \mathbf{a}_m B \end{bmatrix} = \left[\begin{array}{c:c:c:c} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_n \end{array}\right].$$

另一种理解矩阵乘法的方法，涉及到行或列的组合，我们需要用到下述定义.

**定义 1.10.** 设$A_1, A_2, \dots, A_r \in M_{m\times n}$，而$c_1, c_2, \dots, c_r \in \mathbb{F}$. 我们称

$$c_1A_1 + c_2A_2 + \dots + c_rA_r$$

为矩阵$A_1, A_2, \dots, A_r$的一个线性组合，而$c_1, c_2, \dots, c_r$称为该线性组合的系数.

**定理 1.22.** 设$A \in M_{m\times n}$，而$\mathbf{x} \in M_{n\times 1}$. 假设$A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$，则

$$A\mathbf{x} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \dots + x_n\mathbf{c}_n.$$

即$A\mathbf{x}$可表示为$A$的列向量的线性组合，且系数为$\mathbf{x}$的各个分量.

上述定理的证明可利用分块矩阵的乘法得到. 特别地，线性方程组$A\mathbf{x} = \mathbf{b}$是相容的当且仅当存在$x_1, x_2, \dots, x_n \in \mathbb{F}$，使得

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \dots + x_n\mathbf{c}_n = \mathbf{b},$$

即$\mathbf{b}$是$A$的列向量的线性组合.

## 1.6 矩阵运算的代数性质

若$A$是$3\times 2$阶的矩阵，而$B$是$2\times 5$阶的矩阵，那么$AB$是$3\times 5$阶的矩阵，然而乘积$BA$确不存在，因此，矩阵乘法是有顺序的. 进一步，对于同阶方阵$A$与$B$，乘积$AB$和$BA$均存在且同阶，那么它们相等吗？

**例 1.23.** 设$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. 计算$AB$与$BA$.

解.（略.） □

从以上例子可以看出,即使对同阶的方阵而言,乘法不满足交换律. 当两个矩阵$A, B$满足$AB = BA$时，我们称$A$与$B$可交换. 我们再来看看乘法的结合律.

**例 1.24.** 设$A = [a_{ij}] \in M_{m\times r}$, $B = [b_{ij}] \in M_{r\times s}$, $C = [c_{ij}] \in M_{s\times n}$. 计算$(AB)C$与$A(BC)$.

解. 先观察矩阵的阶，我们有$AB \in M_{m\times s}$，从而$(AB)C \in M_{m\times n}$. 类似地，有$BC \in M_{r\times n}$，从而$A(BC) \in M_{m\times n}$. 下面计算各位置元素的表达式，注意到

$$(AB)_{il} = \sum_{k=1}^{r} a_{ik}b_{kl}, \quad (1 \le i \le m,\, 1 \le l \le r),$$

进而

$$\big((AB)C\big)_{ij} = \sum_{l=1}^{s}(AB)_{il}c_{lj} = \sum_{l=1}^{s}\left(\sum_{k=1}^{r} a_{ik}b_{kl}\right)c_{lj} = \sum_{k=1}^{r}\sum_{l=1}^{s} a_{ik}b_{kl}c_{lj}, \quad (1 \le i \le m,\, 1 \le j \le n).$$

另一方面，由于

$$(BC)_{kj} = \sum_{l=1}^{s} b_{kl}c_{lj}, \quad (1 \le k \le r, 1 \le j \le n),$$

我们可以得到

$$\big(A(BC)\big)_{ij} = \sum_{k=1}^{r} a_{ik}(BC)_{kj} = \sum_{k=1}^{r} a_{ik}\left(\sum_{l=1}^{s} b_{kl}c_{lj}\right) = \sum_{k=1}^{r}\sum_{l=1}^{s} a_{ik}b_{kl}c_{lj}, \quad (1 \le i \le m, 1 \le j \le n).$$

故$\big((AB)C\big)_{ij} = \big(A(BC)\big)_{ij}$对$1 \le i \le m$, $1 \le j \le n$均成立，即$(AB)C = A(BC)$.

□

以上例子证明了当矩阵的阶数匹配时，矩阵乘法是具有结合律的. 因此我们可以将$(AB)C$写为$ABC$，不会引起混淆. 现在假设对一个线性方程组进行变量替换：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \end{cases}, \qquad \begin{cases} x_1 = b_{11}y_1 + b_{12}y_2 \\ x_2 = b_{21}y_1 + b_{22}y_2 \\ x_3 = b_{31}y_1 + b_{32}y_2 \end{cases}$$

设它们的矩阵形式为$A\mathbf{x} = \mathbf{b}$与$\mathbf{x} = B\mathbf{y}$. 这时由结合律可以得到

$$\mathbf{b} = A\mathbf{x} = A(B\mathbf{y}) = (AB)\mathbf{y}.$$

它是以$AB$为系数矩阵，以$\mathbf{y}$为变量的线性方程组.

下面的定理列出了更多矩阵运算的关系式，它们的证明留给读者完成.

**定理 1.25.** 设$a, b \in \mathbb{F}$. 假设矩阵$A$, $B$, $C$的阶能够匹配下述各运算. 则有

- $A + B = B + A$
- $A + (B + C) = (A + B) + C$
- $A(BC) = (AB)C$
- $(B + C)A = BA + CA$
- $A(B + C) = AB + AC$
- $a(B + C) = aB + aC$
- $(a + b)C = aC + bC$
- $a(bC) = (ab)C$
- $a(BC) = (aB)C = B(aC)$

以上各式分别为(1)矩阵加法交换律、(2)矩阵加法结合律、（3）矩阵乘法结合律、（4、5）矩阵乘法对矩阵加法的左、右分配律、（6）数乘对矩阵加法的分配律、（7）数乘对数的加法的分配律、（8）数乘与数的乘法的结合律、（9）数乘与矩阵乘法的结合律.

在这些运算律中，有两类矩阵对加法和乘法起到了关键作用. 若一个矩阵中的所有元素均为0，我们称其为零矩阵，记为$\mathbf{0}_{m\times n}$, $\mathbf{0}$, 或直接简记为0. 以下均为零矩阵的例子.

**例 1.26.** 以下矩阵均为零矩阵.

**定理 1.27.** 设$A$为矩阵，而$c \in \mathbb{F}$. 假设以下矩阵的阶匹配下述各运算，则有

$$\begin{bmatrix}0&0\\0&0\end{bmatrix},\quad \begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix},\quad \begin{bmatrix}0&0&0&0\\0&0&0&0\end{bmatrix},\quad \begin{bmatrix}0\\0\\0\\0\end{bmatrix},\quad [0]$$

- $A+\mathbf{0}=\mathbf{0}+A=A$
- $A+(-A)=(-A)+A=0$
- $\mathbf{0}A=\mathbf{0}$
- 若$cA=\mathbf{0}$，则$c=0$或$A=\mathbf{0}$.

若一个$n$阶方阵的主对角元素均为1，而其它元素为0，则称其为单位阵或恒等矩阵，用$I_n$表示，或直接简记为$I$. 以下均为单位阵的例子.

**例 1.28.** 以下矩阵均为单位阵.

$$[1],\quad \begin{bmatrix}1&0\\0&1\end{bmatrix},\quad \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\quad \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$$

**定理 1.29.** 设$A\in M_{m\times n}$. 则

$$I_mA=A,\quad AI_n=A..$$

**定理 1.30.** 假设$R$是一个$n$阶方阵的简约阶梯阵. 要么$R$有零行，要么$R=I_n$.

证明.（略.） □

实数乘法还有消去律：设$a\neq 0$，若$ab=ac$，则$b=c$. 矩阵乘法是否有类似的性质呢？注意到

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

两个非零方程相乘可以变为零矩阵. 另外，

$$\begin{bmatrix}0&1\\0&2\end{bmatrix}\begin{bmatrix}1&1\\3&4\end{bmatrix}=\begin{bmatrix}3&4\\6&8\end{bmatrix}=\begin{bmatrix}0&1\\0&2\end{bmatrix}\begin{bmatrix}2&5\\3&4\end{bmatrix}.$$

在等式两端，我们并不能同时消去左侧的矩阵$\begin{bmatrix}0&1\\0&2\end{bmatrix}$. 事实上，非零实数$a$是具有倒数$a^{-1}$的，而消去率的本质是我们在等式$ab=ac$两端同时乘上了$a^{-1}$. 而一个实数$b$是非零实数$a$的倒数，需要满足的条件是$ab=1$. 对矩阵乘法而言，由于乘法没有交换律，我们需要同时考虑左、右两侧的情形.

**定义 1.11.** 设$A\in M_n$. 若存在$B\in M_n$，使得

$$AB=BA=I_n,$$

则称矩阵$A$是可逆的或非奇异的，称$B$为$A$的逆矩阵并记为$A^{-1}$. 若不存在满足上式子的矩阵$B$，则称$A$是不可逆的或奇异的.

**例 1.31.** 设

$$A=\begin{bmatrix}3&-2\\-7&5\end{bmatrix},\quad B=\begin{bmatrix}5&2\\7&3\end{bmatrix}.$$

证明$A$和$B$互为逆矩阵.

解.（略.） □

怎样的矩阵可能不可逆呢？我们先来看两个例子.

**例 1.32.** 设

$$(1)\ \ A_1 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 0 & 0 \end{bmatrix}, \quad (2)\ \ A_2 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 2 & 4 & 6 \end{bmatrix}.$$

证明$A_1$与$A_2$均不可逆.

以上两个矩阵的特征为：$A_1$的第三行是零行，而$A_2$的第三行是第一行的2倍. 因此，我们不妨将它们按照行进行分块.

证明. 先将$A_1$和$A_2$均简记为$A$. 反证法，假设$A$可逆，则存在$B \in M_3$，使得$AB = I_3$. 记$A$的第$i$行为$\mathbf{r}_i$ $(i = 1, 2, 3)$. 由分块矩阵的乘法可得

$$\begin{bmatrix} \mathbf{r}_1 B \\ \hdashline \mathbf{r}_2 B \\ \hdashline \mathbf{r}_3 B \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \hdashline \mathbf{r}_2 \\ \hdashline \mathbf{r}_3 \end{bmatrix} B = AB = I_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

(1) 对$A = A_1$，考虑第三行，

$$\mathbf{0} = \mathbf{r}_3 B = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix},$$

得到矛盾.

(2) 对$A = A_2$，考虑第一行与第三行，

$$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \mathbf{r}_3 B = (2\mathbf{r}_1)B = 2(\mathbf{r}_1 B) = 2\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix},$$

推出矛盾.

综上，矩阵$A_1$与$A_2$均不可逆.

□

通过以上证明可以看出，若方阵有零行/列，或两行/列成比例，则一定是不可逆的. 有了可逆以及不可逆的例子后，我们将直面一个数学严密性的问题：对一个矩阵$A$而言，会不会要好几个不同矩阵$B$满足$AB = BA = I$？如果有多个的话，关于逆的定义就不合理了. 以下定理证明了逆的唯一性.

**定理 1.33.** *若$B$与$C$均为$A$的逆矩阵，则$B = C$.*

证明. 由逆矩阵的定义，可知

$$AB = BA = I, \quad AC = CA = I.$$

这时，

$$B = BI = B(AC) = (BA)C = IC = C.$$

□

现在，我们可以叙述关于矩阵的消去率了.

**定理 1.34.** *设$B_1, B_2 \in M_{n\times k}$，$C_1, C_2 \in M_{m\times n}$. 假设矩阵$A \in M_n$是可逆的.*

*(1) 若$AB_1 = AB_2$, 则$B_1 = B_2$.*

*(2) 若$C_1 A = C_2 A$, 则$C_1 = C_2$.*

证明. (1) 由于$A$可逆，在等式两端同时左乘$A^{-1}$，得

$$B_1 = IB_1 = (A^{-1}A)B_1 = A^{-1}(AB_1) = A^{-1}(AB_2) = (A^{-1}A)B_2 = IB_2 = B_2.$$

(2) 由于$A$可逆，可在等式两端同时右乘$A^{-1}$. 因此

$$C_1A = C_2A \implies (C_1A)A^{-1} = (C_2A)A^{-1} \implies C_1I = C_2I \implies C_1 = C_2.$$

□

下面，我们来研究一下2阶方程的可逆性.

**定理 1.35.** *矩阵*$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$*可逆当且仅当*$ad - bc \neq 0$. *可逆时*,

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

证明. 当$ad - bc = 0$时，易知$A$的两行成比例，因此不可逆. 而$ad - bc \neq 0$时，验证得

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\left(\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\left(\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\right)\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

命题得证. □

当$ad - bc \neq 0$时，它出现在分母上才有意义. 在线性代数中，这个量是十分中要的，我们把它定义为$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$的2阶行列式，用$\mathsf{det}(A)$或$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$表示，即

$$\mathsf{det}(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

现在，我们尝试用逆矩阵解线性方程组.

**例 1.36.** *假设*$ad - bc \neq 0$. *试解以下线性方程组*.

$$\begin{cases} ax + by &= u \\ cx + dy &= v. \end{cases}$$

证明. 该方程组的矩阵形式为

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}.$$

当$ad - bc \neq 0$时，系数矩阵是可逆的. 故

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}^{-1}\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{ud-vb}{ad-bc} \\ \frac{-cu+av}{ad-bc} \end{bmatrix}.$$

事实上，方程组有唯一解，且可以写成以下形式：

$$x = \frac{\begin{vmatrix} u & b \\ v & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}, \qquad y = \frac{\begin{vmatrix} a & u \\ c & v \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}.$$

□

接下来，我们探究逆运算与其它矩阵运算之间的关系.

**定理 1.37.** 设$A, B \in M_n$均可逆，则$AB$也可逆，且

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

证明. 验证有

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I,$$
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$

得证. □

**例 1.38.** 试求$(AB)^{-1}$，其中

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}.$$

解.（略.） □

一个方阵可以与自己多次相乘，称为方幂. 不同方幂又可以相加等，构成多项式.

**定义 1.12.** 设$A \in M_n$，归纳定义$A$的非负方幂如下：

$$A^0 = I, \quad A^k = AA^{k-1},\ (k = 1, 2, \ldots).$$

当$A$可逆时，定义$A$的负方幂为

$$A^{-k} = (A^{-1})^k,\ (k = 1, 2, \ldots).$$

**例 1.39.** 设$A, B \in M_n$，试展开$(A + B)^2$与$(A + B)^3$.

解. 计算得

$$(A + B)^2 = (A + B)(A + B) = A(A + B) + B(A + B) = A^2 + AB + BA + B^2.$$

进一步有

$$\begin{aligned} (A + B)^3 \quad &= (A + B)(A + B)^2 = A(A^2 + AB + BA + B^2) + B(A^2 + AB + BA + B^2) \\ &= A^3 + A^2B + ABA + AB^2 + BA^2 + BAB + B^2A + B^3. \end{aligned}$$

□

由于矩阵乘法没有交换律，在大多数情况下，中间得交叉项是不可以合并的. 以下关于矩阵方幂的性质不难验证，我们把证明留给读者.

**定理 1.40.** 设$A$可逆，且$k \in \mathbb{Z}^+$，则

- ◇ $A^{-1}$也可逆，且$(A^{-1})^{-1} = A$.
- ◇ $A^n$也可逆，且$(A^n)^{-1} = A^{-n} = (A^{-1})^n$.
- ◇ 对任意非零$c \in \mathbb{F}$，矩阵$cA$也可逆，且$(cA)^{-1} = c^{-1}A^{-1}$.

**定义 1.13.** 设$p(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0$为多项式. 对$A \in M_n$, 定义矩阵多项式$p(A)$为

$$p(A) = a_m A^m + a_{m-1}A^{m-1} + \cdots + a_0 I_n.$$

这里$p(A)$与$A$为同阶方阵. 特别需要注意的是，多项式中的$a_0$可以看成$a_0x^0$，因此代入矩阵$A$时，变为$a_0I$.

**例 1.41.** 求$p(A)$，其中$p(x)=(x+1)(x-3)$，而

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

解.（略.） □

思考题：是否对任何一个矩阵$A\in M_n$，均能找到一个非零多项式$p(x)$，使得$p(A)=0$呢？满足条件的多项式次数是否可以尽可能低呢？

注意到

$$A^rA^s=A^{r+s}=A^{s+r}=A^sA^r,\quad (s,r\in\mathbb{Z}).$$

我们可以进一步得到以下结论.

**定理 1.42.** 关于同一个方阵$A$的两个矩阵多项式可交换，即

$$p_1(A)p_2(A)=p_2(A)p_1(A).$$

接下来，我们再列举关于转置及迹的一些性质.

**定理 1.43.** 设$A,B\in M_{m\times n}$, $C\in M_{n\times k}$，而$c\in\mathbb{F}$. 则

*(1)* $(A^T)^T=A$.

*(2)* $(A+B)^T=A^T+B^T$.

*(3)* $(cA)^T=cA^T$.

*(4)* $(AC)^T=C^TA^T$.

证明. 这里仅证明(4). 首先可验证$(AC)^T$与$C^TA^T$均为$k\times m$阶矩阵. 设$A=[a_{ij}]$及，$C=[c_{ij}]$. 对$1\le i\le k$及$1\le j\le m$，有

$$\left((CA)^T\right)_{ij}=(CA)_{ji}=\sum_{l=1}^k c_{jl}a_{li}.$$

从而

$$(C^TA^T)_{ij}=\sum_{l=1}^k (C^T)_{il}(A^T)_{lj}=\sum_{l=1}^k c_{li}a_{jl}=\left((CA)^T\right)_{ij}.$$

得证. □

**定理 1.44.** 设$A$为可逆方阵，则$A^T$亦可逆，且

$$(A^T)^{-1}=(A^{-1})^T.$$

**定理 1.45.** 设$A,B\in M_n$，而$c\in\mathbb{F}$. 则

*(1)* $\mathrm{tr}(A+B)=\mathrm{tr}A+\mathrm{tr}B$.

*(2)* $\mathrm{tr}(cA)=c\cdot\mathrm{tr}A$.

*(3)* $\mathrm{tr}(AB)=\mathrm{tr}(BA)$.

以上两个定理的证明留给读者自行完成.

## 1.7 初等矩阵

在例1.17中，我们已经注意到，某些类型的矩阵在左、右乘其它矩阵后所得的效果，相当于给其它矩阵作了初等行、列变换.

**定义 1.14.** *在$M_n(\mathbb{F})$中，我们定义三类矩阵$F_i(c)$，$F_{ij}$和$F_{i,j}(c)$如下，分别称为第一类、第二类、第三类初等矩阵.*

*(1) 设$c\in\mathbb{F}\setminus\{0\}$，而$1\le i\le n$. 定义$F_i(c)\in M_n$为*

$$\big(F_i(c)\big)_{st}=\begin{cases}c, & s=t=i,\\ 1, & s=t\neq i,\\ 0, & s\neq t,\end{cases},\quad (1\le s,t\le n).$$

*(2) 设$1\le i\neq j\neq n$. 定义$F_{i,j}$为*

$$\big(F_{ij}\big)_{st}=\begin{cases}1, & (s,t)=(i,j)\ \text{或}(s,t)=(j,i),\\ 1, & s=t\not\in\{i,j\},\\ 0, & \text{其余情形},\end{cases},\quad (1\le s,t\le n).$$

*(3) 设$c\in\mathbb{F}$，而$1\le i\neq j\neq n$. 定义$F_{i,j}(c)\in M_n$为*

$$\big(F_i(c)\big)_{st}=\begin{cases}1, & s=t,\\ c, & (s,t)=(i,j),\\ 0, & \text{其余情形},\end{cases},\quad (1\le s,t\le n).$$

**例 1.46.** *单位阵$I$也是初等矩阵. 另外，以下矩阵依次为$F_2(4)\in M_2(\mathbb{R})$，$F_{1,3}\in M_4(\mathbb{F})$以及$F_{1,3}(-5)\in M_3(\mathbb{C})$:*

$$\begin{bmatrix}1&0\\0&4\end{bmatrix},\quad \begin{bmatrix}0&0&1&0\\0&1&0&0\\1&0&0&0\\0&0&0&1\end{bmatrix},\quad \begin{bmatrix}1&0&-5\mathrm{i}\\0&1&0\\0&0&1\end{bmatrix}$$

.

三类初等矩阵从左侧乘矩阵$A$，与$A$经过对应的三类初等行变换所得的矩阵是相同的. 而三类初等矩阵从右侧乘矩阵$A$，与$A$经过对应的三类初等列变换所得的矩阵是相同的. 另外，将某一行乘非零常数$c$后再乘$c^{-1}$就回到初始状态；交换两行后再交换一次又回到初始状态；将一行加上另一行的$c$倍后再加后者的$-c$倍，也能回到初始状态. 如此容易得到初等矩阵的逆.

**定理 1.47.** *初等矩阵均可逆，且*

$$F_i(c)^{-1}=F_i(c^{-1})\ (c\neq0),\quad F_{i,j}^{-1}=F_{i,j},\quad F_{i,j}(c)^{-1}=F_{i,j}(-c).$$

**例 1.48.** 设

$$A=\begin{bmatrix}1&0&2&5\\-2&3&-1&7\\2&4&-9&-4\end{bmatrix},\quad B=\begin{bmatrix}0&3&0\\-2&1&1\\5&2&0\\7&0&-2\end{bmatrix},$$

$$E_1=\begin{bmatrix}1&0&0\\0&3&0\\0&0&1\end{bmatrix},\quad E_2=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix},\quad E_3=\begin{bmatrix}1&0&0\\0&1&0\\2&0&1\end{bmatrix}.$$

求以下矩阵表达式

$$E_1A,\ E_2A,\ E_3A,\ BE_1,\ BE_2,\ BE_3,\ E_1^{-1},\ E_2^{-1},\ E_3^{-1}.$$

解.（略.） □

接下来，我们从其它角度来看一看什么是“矩阵可逆”，并尝试寻找判断具体矩阵的可逆性、以及计算具体矩阵的逆的算法.

**定理 1.49.** 设$A\in M_n(\mathbb{F})$. 以下命题等价.

*(i)* 矩阵$A$可逆.

*(ii)* 齐次线性方程组$A\mathbf{x}=\mathbf{0}$仅有平凡解.

*(iii)* 矩阵$A$的简约阶梯型为$I_n$.

*(iv)* 矩阵$A$可写成初等矩阵的成绩.

证明. “(i)$\Longrightarrow$(ii)”：首先齐次方程组一定有零解. 设$\mathbf{x}_0$是方程组$A\mathbf{x}=\mathbf{0}$的任何一个解，即$A\mathbf{x}_0=\mathbf{0}$. 由$A$可逆，知

$$\mathbf{x}_0=I_n\mathbf{x}_0=(A^{-1}A)\mathbf{x}_0=A^{-1}(A\mathbf{x}_0)=A^{-1}\mathbf{0}=\mathbf{0}.$$

故该方程组仅有平凡解.

“(ii)$\Longrightarrow$(iii)”：方程组$A\mathbf{x}=\mathbf{0}$仅有平凡解，因此不存在自由变元，所有变元均为主元，从而$\mathsf{rank}(A)=n$(见推论1.9). 此时每一列均为主元列，仅有一个非零元素1，故$A$的简约阶梯型为$I_n$.

“(iii)$\Longrightarrow$(iv)”：设矩阵$A$通过$k$ $(k\geq 0)$次初等行变换化为简约阶梯型$I_n$，可记$E_k\dots E_2E_1A=I_n$，其中$E_1,E_2,\dots,E_k$为初等矩阵. 从而

$$A=(E_k\dots E_2E_1)^{-1}=E_1^{-1}E_2^{-1}\dots E_k^{-1}.$$

由于$E_1^{-1}$, $E_2^{-1}$,...,$E_k^{-1}$也是初等矩阵，因此$A$是初等矩阵的乘积.

“(iv)$\Longrightarrow$(i)”：设$A=F_1F_2\dots F_l$，其中$l\geq 1$，而$F_1$, $F_2$,...,$F_l$均为初等矩阵. 由于$F_1,F_2,\dots,F_n$均可逆，因此$A$可逆. □

从上述等价性定理中的(c)可知，判定一个方阵是否可逆，只需观察它的（简约）阶梯型.

**例 1.50.** 证明以下矩阵不可逆：

$$\begin{bmatrix}1&6&4\\2&4&-1\\-1&2&5\end{bmatrix}.$$

证明.（略.） □

当$A$可逆时，我们从上述定理的证明过程可知$E_k\dots E_2E_1A=I$，这里每个$E_i$ $(i=1,2,\dots,k)$表示一次初等行变换. 这时$A^{-1}=E_k\dots E_2E_1$，而它的具体计算可通过$E_k\dots E_2E_1=(E_k\dots E_2E_1)I$来实现，即在单位阵上作对应的行变换所得. 更一般地，当$A$可逆时，我们有

$$A^{-1}\left[\,A\,\vdots\,I\,\vdots\,B\,\right]=\left[\,A^{-1}A\,\vdots\,A^{-1}I\,\vdots\,A^{-1}B\,\right]=\left[\,I\,\vdots\,A^{-1}\,\vdots\,A^{-1}B\,\right].$$

将$A^{-1}$写成初等矩阵的乘积，则有

$$\left[\,A\,\vdots\,I\,\vdots\,B\,\right]\xrightarrow{\text{行变换}}\left[\,I\,\vdots\,A^{-1}\,\vdots\,A^{-1}B\,\right].$$

即当矩阵$A$被初等行变换化为$I$时，矩阵$I$和$B$就分别变成了$A^{-1}$和$A^{-1}B$.

**例 1.51.** 计算$A^{-1}$与$A^{-1}B$，其中

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & -2 \\ 0 & 3 & 2 \end{bmatrix}.$$

解.（略.） □

如果我们要计算的是矩阵$CA^{-1}$，可以尝试以下方法：一、如上述方法计算$A^{-1}$，之后计算$CA^{-1}$；二、由于$(CA^{-1})^T = (A^{-1})^T C^T = (A^T)^{-1} C^T$，如上述方法计算$(A^T)^{-1}C^T$，并将所得矩阵转置；三、尝试以下初等列变换：

$$\left[\begin{array}{c} A \\ \hdashline C \end{array}\right] \xrightarrow{\text{列变换}} \left[\begin{array}{c} I \\ \hdashline CA^{-1} \end{array}\right],$$

当$A$被初等列变换化为$I$时，$C$就变成了$CA^{-1}$. 最后一种方法的数学原理是什么？请读者们思考一下.

接下来，我们使用等价性定理的(b)来分析一下齐次方程组解的情形.

**例 1.52.** 下列齐次线性方程组仅有平凡解？有无穷多解？

$$\text{(a)} \begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ 2x_1 + 5x_2 + 3x_3 = 0 \\ x_1 \qquad\quad + 8x_3 = 0 \end{cases} \qquad \text{(b)} \begin{cases} x_1 + 6x_2 + 4x_3 = 0 \\ 2x_1 + 4x_2 - x_3 = 0 \\ -x_1 + 2x_2 + 5x_3 = 0 \end{cases}$$

解.（略.） □

将初等行、列变换写成左、有乘初等矩阵，我们可以证明下述定理.

**定理 1.53.** 设矩阵$A \in M_{m\times n}$. 则存在整数$0 \le r \le \min\{m, n\}$，可逆矩阵$P \in M_m$以及可逆矩阵$Q \in M_n$，使得

$$PAQ = \left[\begin{array}{c:c} I_r & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right]$$

说明. 首先，利用行变换将$A$化为简约阶梯阵$R$，得$PA = R$. 这里$P$为初等行变换所对应的初等矩阵的乘积，是可逆的. 接着通过第二类列变换将$R$中各主元列换至左侧，得

$$RQ_1 = \left[\begin{array}{c:c} I_r & C \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right]$$

的形式. 此时，通过第三类初等列变换，可用$I_r$中的第$i$ $(1 \le i \le r)$行里的1，将$C$中第$i$行的元素都消成0，即

$$\left[\begin{array}{c:c} I_r & C \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right] Q_2 = \left[\begin{array}{c:c} I_r & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right].$$

综上，得

$$PAQ_1Q_2 = \left[\begin{array}{c:c} I_r & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right].$$

这里$Q_1, Q_2$均为初等矩阵的乘积，是可逆的. 记$Q = Q_1Q_2$，则$Q$也是可逆的. □

上述定理中的$r$即矩阵$A$的秩. 事实上，上述定理可以作为矩阵$A$的秩的另一种定义方式. 利用上述定理，可以证明非方阵在满秩时也具有消去率.

**定理 1.54.** 设$B \in M_r$，而$C \in M_{r\times n}$. 假设$\mathit{rank}(C) = r$，则下列命题成立.

*(1)* 若$BC = 0$，则$B = 0$.

*(2)* 若$BC = C$，则$B = I_r$

上述定理的证明留给读者作为思考题.

## 1.8 线性方程组与可逆矩阵的更多性质

在这一节中，我们将进一步研究系数矩阵与线性方程组的解之间的关系.

**定理 1.55.** *每个线性方程组都会出现且只出现以下三种情形中的一种：*

*(1) 方程组不相容; (2) 方程组有唯一解；(3) 方程组有无穷多解.*

证明. 将方程组记为$A\mathbf{x}=\mathbf{b}$. 若该方程组无解，则恰出现情形(1)，否则方程组有解. 这时，若有唯一解，则恰出现情形(2)，否则方程组至少有两个不同解$\mathbf{x}_1, \mathbf{x}_2$，即

$$A\mathbf{x}_1=\mathbf{b}, \quad A\mathbf{x}_2=\mathbf{b}.$$

记$\mathbf{x}_0=\mathbf{x}_1-\mathbf{x}_2$. 由于$\mathbf{x}_1$与$\mathbf{x}_2$不同，知$\mathbf{x}_0\neq\mathbf{0}$. 注意到

$$A\mathbf{x}_0=A(\mathbf{x}_1-\mathbf{x}_2)=A\mathbf{x}_1-A\mathbf{x}_2=\mathbf{b}-\mathbf{b}=\mathbf{0}.$$

此时对任何$t\in\mathbb{F}$，均有

$$A(\mathbf{x}_1+t\mathbf{x}_0)=A\mathbf{x}_1+A\mathbf{x}_0=\mathbf{b}+\mathbf{0}=\mathbf{b}.$$

因此方程组有无穷多个解$\mathbf{x}_1+t\mathbf{x}_0$ $(t\in\mathbb{F})$. □

**定理 1.56.** *设线性方程组$A\mathbf{x}=\mathbf{b}$的系数矩阵$A$为可逆方阵时，该方程组有唯一解.*

证明. 当$A$可逆时，容易验证$A(A^{-1}\mathbf{b})=\mathbf{b}$，因此方程组有解. 反之，设$\mathbf{x}$为方程组的任何一个解，则

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

因此解是唯一的. □

**例 1.57.** *利用前面计算过的逆矩阵，求解方程组*

$$\begin{cases} x_1+2x_2+3x_3=5 \\ 2x_1+5x_2+3x_3=3 \\ x_1 \qquad\quad +8x_3=17 \end{cases}$$

$$\begin{cases} x_1+2x_2+3x_3=4 \\ 2x_1+5x_2+3x_3=5 \\ x_1 \qquad\quad +8x_3=9 \end{cases}, \quad \begin{cases} x_1+2x_2+3x_3=1 \\ 2x_1+5x_2+3x_3=6 \\ x_1 \qquad\quad +8x_3=-6 \end{cases}$$

解.（略.） □

当我们遇到多个具有公共的可逆系数矩阵的方程组

$$A\mathbf{x}=\mathbf{b}_1, \quad A\mathbf{x}=\mathbf{b}_2, \quad \ldots \quad A\mathbf{x}=\mathbf{b}_k$$

时，还可以利用逆矩阵与分块矩阵，更加快捷地进行计算. 假设$\mathbf{y}_i$是$A\mathbf{x}=\mathbf{b}_i$的唯一解$(1\leq i\leq k)$. 它们一定满足

$$A\left[\mathbf{y}_1 \,\vdots\, \mathbf{y}_2 \,\vdots\, \cdots \,\vdots\, \mathbf{y}_k\right]=\left[A\mathbf{y}_1 \,\vdots\, A\mathbf{y}_2 \,\vdots\, \cdots \,\vdots\, A\mathbf{y}_k\right]=\left[\mathbf{b}_1 \,\vdots\, \mathbf{b}_2 \,\vdots\, \cdots \,\vdots\, \mathbf{b}_k\right],$$

当$A$可逆时，我们可进一步得到

$$\left[\mathbf{y}_1 \,\vdots\, \mathbf{y}_2 \,\vdots\, \cdots \,\vdots\, \mathbf{y}_k\right]=A^{-1}\left[\mathbf{y}_1 \,\vdots\, \mathbf{y}_2 \,\vdots\, \cdots \,\vdots\, \mathbf{y}_k\right].$$

**例 1.58.** 试同时求解方程组

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 4 \\ 2x_1 + 5x_2 + 3x_3 = 5 \\ x_1 \qquad\quad + 8x_3 = 9 \end{cases}, \quad \begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ 2x_1 + 5x_2 + 3x_3 = 6 \\ x_1 \qquad\quad + 8x_3 = -6 \end{cases}$$

解.（略.） □

方阵$A \in M_n(\mathbb{F})$可逆的定义是：存在$B \in M_n(\mathbb{F})$，使得$AB = BA = I$. 这个式子中包含两部分：$AB = I$与$BA = I$. 下一个定理说明两式中的任意一个都能保证矩阵的可逆性.

**定理 1.59.** 设$A \in M_n(\mathbb{F})$.

*(1) 若存在$B \in M_n(\mathbb{F})$，使得$BA = I$，则$A$可逆，且$A^{-1} = B$.*

*(2) 若存在$B \in M_n(\mathbb{F})$，使得$AB = I$，则$A$可逆，且$A^{-1} = B$.*

证明. (1) 根据等价性定理，要证明$A$可逆，只需证明齐次线性方程组$A\mathbf{x} = \mathbf{0}$仅有平凡解即可. 事实上，由$BA = I$可知

$$\mathbf{x} = I\mathbf{x} = (BA)\mathbf{x} = B(A\mathbf{x}) = B\mathbf{0} = \mathbf{0}.$$

因此$A$可逆得证，由逆的唯一性知$A^{-1} = B$.

(2) 注意到$AB = I$意味着$B^T A^T = (AB)^T = I^T = I$. 根据（1）知矩阵$A^T$可逆且$(A^T)^{-1} = B^T$. 故$A$可逆，且满足$(A^{-1})^T = (A^T)^{-1} = B^T$，即$A^{-1} = B$. □

**定理 1.60.** 设$A, B \in M_n(\mathbb{F})$. 若$AB$可逆，则$A$、$B$均可逆.

证明. 由于$AB$可逆，知存在方阵$C$，使得$C(AB) = (AB)C = I$. 此时有

$$(CA)B = I, \quad A(BC) = I.$$

由上个定理可知，矩阵$A$、$B$均可逆. □

现在，我们进一步扩展等价性定理，使之囊括非齐次线性方程组.

**定理 1.61.** 设$A \in M_n(\mathbb{F})$. 以下命题等价.

*(i) 矩阵$A$可逆.*

*(v) 对任意$\mathbf{b} \in M_{n\times 1}(\mathbb{F})$，线性方程组$A\mathbf{x} = \mathbf{b}$是相容的.*

*(vi) 对任意$\mathbf{b} \in M_{n\times 1}(\mathbb{F})$，线性方程组$A\mathbf{x} = \mathbf{b}$有唯一解.*

证明. 在前文中，我们已经证明了“(i)$\Longrightarrow$(vi)”. 而“(vi)$\Longrightarrow$(v)”是显然的. 以下我们证明“(v)$\Longrightarrow$(i)”. 根据之前的定理，只要找到矩阵$C \in M_n(\mathbb{F})$，使得$AC = I$，就可以推出$A$可逆.

对$1 \le i \le n$，取$\mathbf{e}_i \in M_{n\times 1}(\mathbb{F})$为第$i$个元素为1，其余元素为0的列向量. 由(5)知方程组$A\mathbf{x} = \mathbf{e}_i$有解，记为$\mathbf{c}_i$，满足$A\mathbf{c}_i = \mathbf{b}_i$ $(1 \le i \le n)$. 此时取$C = [\,\mathbf{c}_1 \,\vdots\, \mathbf{c}_2 \,\vdots \ldots \vdots\, \mathbf{c}_n\,]$，便有

$$AC = A\,[\,\mathbf{c}_1 \,\vdots\, \mathbf{c}_2 \,\vdots \ldots \vdots\, \mathbf{c}_n\,] = [\,A\mathbf{c}_1 \,\vdots\, A\mathbf{c}_2 \,\vdots \ldots \vdots\, A\mathbf{c}_n\,] = [\,\mathbf{e}_1 \,\vdots\, \mathbf{e}_2 \,\vdots \ldots \vdots\, \mathbf{e}_n\,] = I.$$

因此矩阵$A$可逆. □

## 1.9 对角、三角及对称矩阵

在本节中，我们将介绍一些特定类型的矩阵.

**定义 1.15.** 设$A = [a_{ij}] \in M_n(\mathbb{F})$. 若对所有$1 \le i \ne j \le n$，均有$a_{ij} = 0$，则称$A$为对角矩阵.

对角矩阵的非零元素只可能出现在对角线上. 当$n$阶对角矩阵$A$的对角元依次为$a_1, a_2, \dots, a_n$时，我们也可记为$A = \mathsf{diag}(a_1, a_2, \dots, a_n)$. 上述对角阵$A$在左侧乘一个$n \times m$阶矩阵$B$的效果是将$B$的第$i$行乘以$a_i$ $(1 \le i \le n)$，而$A$在右侧乘上一个$m \times n$阶矩阵$C$的效果是将$C$的第$i$列乘以$a_i$ $(1 \le i \le n)$.

**例 1.62.** 以下均为对角矩阵：

$$\begin{bmatrix} 2 & 0 \\ 0 & -7 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -\sqrt{2} \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

**例 1.63.** 试计算以下矩阵乘法：

$$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

解.（略.） □

对角矩阵的运算可以在对角线的各个位置上独立进行. 下述定理留给读者们自行验证.

**定理 1.64.** 设$c, a_1, a_2, \dots, a_n, b_1, \dots, b_n \in \mathbb{F}$. 设$p(x)$为系数来自于$\mathbb{F}$的多项式.

*(i)* $c\mathsf{diag}(a_1, a_2, \dots, a_n) = \mathsf{diag}(ca_1, ca_2, \dots, ca_n)$.

*(ii)* $\mathsf{diag}(a_1, a_2, \dots, a_n) + \mathsf{diag}(b_1, b_2, \dots, b_n) = \mathsf{diag}(a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$.

*(iii)* $\mathsf{diag}(a_1, a_2, \dots, a_n)\mathsf{diag}(b_1, b_2, \dots, b_n) = \mathsf{diag}(a_1 b_1, a_2 b_2, \dots, a_n b_n)$.

*(iv)* $\mathsf{diag}(a_1, a_2, \dots, a_n)$可逆当且仅当$a_1, a_2 \dots, a_n \neq 0$. 当其可逆时，

$$\mathsf{diag}(a_1, a_2, \dots, a_n)^{-1} = \mathsf{diag}(a_1^{-1}, a_2^{-1}, \dots, a_n^{-1}).$$

*(v)* $p\big(\mathsf{diag}(a_1, a_2, \dots, a_n)\big) = \mathsf{diag}\big(p(a_1), p(a_2), \dots, p(a_n)\big)$.

**例 1.65.** 求$A^5$, $A^{-2}$以及$A^k$ $(k \in \mathbb{Z})$，其中

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

解.（略.） □

**定义 1.16.** 设$A = [a_{ij}] \in M_n(\mathbb{F})$. 若当$i > j$时，均有$a_{ij} = 0$，则称$A$为上三角矩阵. 若当$i < j$时，均有$a_{ij} = 0$，则称$A$为下三角矩阵.

上（下）三角矩阵的非零元素仅可能出现在矩阵主对角线及其右上（左下）方.

**例 1.66.** 以下矩阵分别为上、下三角矩阵：

$$\begin{bmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{bmatrix}, \quad \begin{bmatrix} * & 0 & 0 & 0 \\ * & * & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \end{bmatrix}.$$

**定理 1.67.** *(i)* 上（下）三角矩阵的转置为下（上）三角矩阵.

*(ii)* 上（下）三角矩阵作数乘后仍为为上（下）三角矩阵.

*(iii)* 同阶上（下）三角矩阵相加后仍为为上（下）三角矩阵.

*(iv)* 同阶上（下）三角矩阵相乘后仍为为上（下）三角矩阵.

*(v)* 上（下）三角矩阵可逆当且仅当主对角元素均非零. 当可逆时，其逆也是上（下）三角矩阵.

证明. 此处仅证明(iv)中关于上三角矩阵的命题，其余的部分留给读者自行完成.

设$A=[a_{ij}]$与$B=[b_{ij}]$均为$n$阶上三角矩阵，即

$$a_{ij}=b_{ij}=0,\quad (1\leq j<i\leq n).$$

这时,对$1\leq i,j\leq n$，有

$$(AB)_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj}.$$

若$(AB)_{ij}\neq 0$，则存在$1\leq k\leq n$使得$a_{ik}b_{kj}\neq 0$，从而$a_{ik},b_{kj}\neq 0$，继而$i\leq k\leq j$. 也就是说，当$1\leq j<i\leq n$时，必有$(AB)_{ij}=0$. 故$AB$是上三角矩阵. □

**例 1.68.** 设$A\in M_n(\mathbb{R})$满足

$$(A)_{ij}=\begin{cases}1, & ruoj=i+1\ (1\leq i\leq n-1)\\ 0, & \text{其余情形}.\end{cases}$$

求$A^k$ $(k=0,1,2,\ldots)$.

解.（待补.） □

在计算机中连续解同系数矩阵的方程组时，为了降低计算复杂度和误差，往往会将方阵分成三角矩阵的乘积.

**定义 1.17.** 设$A\in M_n(\mathbb{F})$，若存在下三角矩阵$L$及上三角矩阵$U$，使得$A=LU$，则称上式为矩阵$A$的一个$LU$-分解.

并不是每一个矩阵都存在$LU$-分解. 比如，若可逆矩阵$A$满足$(A)_{11}=0$，则$A$没有$LU$-分解. 倘若不然，设一个分解为$A=LU$，由于$(A)_{11}$的值等于$L$的第一行乘$U$的第一列，得$0=(A)_{11}=(L)_{11}(U)_{11}$. 从而$(L)_{11}$和$(U)_{11}$中至少有一个元素为0. 于是$L$与$U$中至少有一个矩阵是不可逆的，这与$A$可逆矛盾.

注意到阶梯阵是一个上三角矩阵，如果将方阵$A$化成阶梯阵的过程只用到了$F_i(c)$与$F_{i,j}(c)$ $(i>j)$，那么$A$存在$LU$-分解，且可借助行变换求出.

**例 1.69.** 求$A$的一个$LU$-分解，其中

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

解.（略.） □

当$A=LU$，其中$L$、$U$分别为下、上三角矩阵时，线性方程组$A\mathbf{x}=\mathbf{b}$变成了$LU\mathbf{x}=\mathbf{b}$. 记$U\mathbf{x}=\mathbf{y}$，则$L\mathbf{y}=\mathbf{b}$. 第一步，我们解方程组$L\mathbf{y}=\mathbf{b}$，由于$L$是下三角，求得$\mathbf{y}$的过程是从前向后代入的过程. 第二步，解方程组$U\mathbf{x}=\mathbf{y}$，由于$U$是上三角矩阵，求得$\mathbf{x}$的过程是从后向前的代入过程. 最终，我们求得原方程组$A\mathbf{x}=\mathbf{b}$的解.

**例 1.70.** 利用$A$的$LU$-分解求解线性方程组$A\mathbf{x}=\mathbf{b}$，其中

$$A=\begin{bmatrix}1&1&1\\1&2&2\\1&2&3\end{bmatrix},\quad \mathbf{b}=\begin{bmatrix}2\\0\\3\end{bmatrix}.$$

解.（略.） □

当我们需要反复（非同时）解公共系数矩阵的方程组时，如果每一次都在增广矩阵上作高斯或高斯-约当消元法，那么重复计算的部分是很多的. 如果利用$LU$分解的从前向后、从后先前的代入过程，求解的算法复杂度较低，且由计算机四舍五入照成的误差也会相对较小.

接下来，我们介绍对称矩阵、斜对称矩阵与厄密特矩阵.

**定义 1.18.** 设$A\in M_n(\mathbb{R})$，若$A^T=A$，则称$A$为对称矩阵. 若$A^T=-A$，则称$A$为斜对称（或反对称）矩阵.

**定义 1.19.** 设$A\in M_n(\mathbb{C})$，定义$A^*\in M_n(\mathbb{C})$为$A$的共轭转置，即

$$(A^*)_{ij}=\overline{A_{ji}},\quad (1\le i,j\le n).$$

我们将$A^*$称为$A$的伴随*(adjoint)*. 若$A^*=A$，则称$A$为厄密特（或自伴）矩阵.

从定义不难验证，斜对称矩阵的对角元满足

$$(A)_ii=(A^T)_{ii}=(-A)_{ii}=-(A)_ii,\quad (1\le i\le n),$$

从而必为0. 而厄密特矩阵的对角元

$$\overline{(A)_ii}=(A^*)_{ii}=(A)_{ii},\quad (1\le i\le n),$$

因此必为实数.

**例 1.71.** 以下矩阵均为（实）对称矩阵：

$$[3],\qquad \begin{bmatrix}2&1\\1&0\end{bmatrix},\qquad \begin{bmatrix}1&0&3\\0&2&-1\\3&-1&-\sqrt{2}\end{bmatrix}.$$

以下矩阵均为（实）斜对称矩阵：

$$[0],\qquad \begin{bmatrix}0&1\\-1&0\end{bmatrix},\qquad \begin{bmatrix}0&0&3\\0&0&-1\\-3&1&0\end{bmatrix}.$$

以下矩阵均为（复）厄密特矩阵

$$[\pi],\qquad \begin{bmatrix}1&\mathrm{i}\\-\mathrm{i}&0\end{bmatrix},\qquad \begin{bmatrix}0&1+\mathrm{i}&3\\1-\mathrm{i}&2&-2-5\mathrm{i}\\3&-2+5\mathrm{i}&-1\end{bmatrix}.$$

这些矩阵的基本性质如下.

**定理 1.72.** *设$A, B, C \in M_n(\mathbb{R})$均为对称矩阵. 设$D \in M_{m\times n}(\mathbb{R})$，而$c \in \mathbb{R}$. 设$C$是可逆的.*

*(1) $A^T$对称.*

*(2) $A+B$对称.*

*(3) $cA$对称.*

*(4) $C^{-1}$对称.*

*(5) $DD^T$与$D^TD$均对称.*

*(6) $AB$对称当且仅当$AB = BA$.*

*证明.* 这里仅证明(5)的$D^TD$与(6)，其余部分留给读者自行完成.

(5) 注意到

$$(D^TD)^T = D^T(D^T)^T = D^TD,$$

因此$D^TD$是对称矩阵.

(6) 由于$A, B$对称，即$A^T = A$，$B^T = B$. 于是有

$$(AB)^T = B^TA^T = BA.$$

这时不难看出，$(AB)^T = AB$当且仅当$BA = AB$. □

**定理 1.73.** *设$A, B, C \in M_n(\mathbb{C})$均为厄密特矩阵. 设$D \in M_{m\times n}(\mathbb{C})$，而$c \in \mathbb{R}$. 设$C$是可逆的.*

*(1) $A^*$是厄密特矩阵.*

*(2) $A+B$是厄密特矩阵.*

*(3) $cA$是厄密特矩阵.*

*(4) $C^{-1}$是厄密特矩阵.*

*(5) $DD*$与$D^*D$是厄密特矩阵.*

*(6) $AB$是厄密特矩阵当且仅当$AB = BA$.*

上述定理的证明，以及关于斜对称矩阵的类似命题，留给读者们思考.

**定理 1.74.** *(1) 设$A \in M_n(\mathbb{R})$，则$A$可唯一地表示为$A = B + C$，其中$B \in M_n(\mathbb{R})$为对称矩阵，而$C \in M_n(\mathbb{R})$为斜对称矩阵.*

*(2) 设$A \in M_n(\mathbb{C})$，则$A$可唯一地表示为$A = B + \mathrm{i}C$，其中$B, C \in M_n(\mathbb{C})$均为厄密特矩阵.*

*证明.* 此处仅证明(2)，命题(1)的证明是相似的.

倘若我们并不知道这样的$B$和$C$如何选取，可以先假设存在，则由$B^* = B$和$C^* = C$得

$$A = B + \mathrm{i}C, \quad A^* = B^* + \bar{i}C^* = B - \mathrm{i}C.$$

两式相加、减便得

$$B = \frac{A + A^*}{2}, \quad C = \frac{A - A^*}{2\mathrm{i}}.$$

以上推导过程，在假设存在性得条件下，已经证明了表示的唯一性.

为了说明存在性，取$B = \frac{A+A^*}{2}$, $C = \frac{A-A^*}{2\mathrm{i}}$. 容易验证$B^* = B$, $C^* = C$以及$A = B + \mathrm{i}C$. 证毕. □

未来大家或许能更好地感受到，如上$A = B + \mathrm{i}C$的形式可以看成复矩阵的一种直角坐标表示.

接下来，我们进一步讲讲分块三角矩阵与分块初等矩阵.

**定义 1.20.** *我们称如下的分块矩阵为分块上三角矩阵：*

$$\left[\begin{array}{c:c:c:c} A_{11} & A_{12} & \dots & A_{1s} \\ \hdashline 0 & A_{22} & \dots & A_{2s} \\ \hdashline \cdots & \cdots & \cdots & \cdots \\ \hdashline 0 & 0 & \dots & A_{ss} \end{array}\right]_{n\times n},$$

其中$A_{11}$，$A_{22}$，...，$A_{ss}$均为方阵. 进一步，若当$1 \le i \ne j \le s$时，子矩阵$A_{ij}$均为零矩阵，我们称其为分块对角矩阵.

**例 1.75.** 设 $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, 其中 $A_{11}, A_{22} \in M_n$. 假设$A_{11}$, $A_{22}$均可逆. 求$A^{-1}$的表达式.

解. 如果我们能解出如下的分块矩阵，满足

$$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} C & D \\ E & F \end{bmatrix} = \begin{bmatrix} A_{11}C + A_{12}E & A_{11}D + A_{12}F \\ A_{22}E & A_{22}F \end{bmatrix},$$

那么矩阵$A$就是可逆的. 即

$$\begin{cases} A_{11}C + A_{12}E = I, \\ A_{11}D + A_{12}F = 0, \\ A_{22}E = 0, \\ A_{22}F = I. \end{cases}$$

由于$A_{22}$可逆，我们可以从后两式解得$E = 0$, $F = A_{22}^{-1}$. 再结合$A_{11}$可逆，从前两式可得

$$C = A_{11}^{-1}(I - A_{12}E) = A_{11}^{-1}$$
$$D = A_{11}^{-1}(0 - A_{12}F) = -A_{11}^{-1}A_{12}A_{22}^{-1}.$$

综上有

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

这也是一个分块上三角矩阵. □

对于分块矩阵来说，也有分块初等矩阵的概念，而左、右乘分块矩阵相当与作对应的分块初等行、列变换. 比如，以下矩阵分别是三类分块初等矩阵：

$$\begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix}, \quad \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}, \quad \begin{bmatrix} I & B \\ 0 & I \end{bmatrix},$$

其中$P$为可逆方阵，$I$可以是不同阶的单位矩阵，而$B$为合适阶的矩阵. 它们的乘法效果如下：

$$\begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} C & D \\ E & F \end{bmatrix} = \begin{bmatrix} PC & PD \\ E & F \end{bmatrix} \qquad \begin{bmatrix} C & D \\ E & F \end{bmatrix}\begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} CP & D \\ EP & F \end{bmatrix}$$
$$\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}\begin{bmatrix} C & D \\ E & F \end{bmatrix} = \begin{bmatrix} E & F \\ C & D \end{bmatrix} \qquad \begin{bmatrix} C & D \\ E & F \end{bmatrix}\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} = \begin{bmatrix} D & C \\ F & E \end{bmatrix}$$
$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}\begin{bmatrix} C & D \\ E & F \end{bmatrix} = \begin{bmatrix} C + BE & D + BF \\ E & F \end{bmatrix} \qquad \begin{bmatrix} C & D \\ E & F \end{bmatrix}\begin{bmatrix} I & B \\ 0 & I \end{bmatrix} = \begin{bmatrix} C & CB + D \\ E & EB + F \end{bmatrix}.$$

**例 1.76.** 设 $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, 其中 $A_{11}, A_{22} \in M_n$. 假设$A_{11}$, $A_{22}$均可逆. 求$A^{-1}$的表达式.

解. 注意到$A_{11}$, $A_{22}$均可逆，我们使用分块初等行变换来求矩阵的逆.

$$\left[\begin{array}{cc:cc} A_{11} & A_{12} & I & 0 \\ 0 & A_{22} & 0 & I \end{array}\right] \xrightarrow{A_{ii}^{-1}(\text{行}i)} \left[\begin{array}{cc:cc} I & A_{11}^{-1}A_{12} & A_{11}^{-1} & 0 \\ 0 & I & 0 & A_{22}^{-1} \end{array}\right] \xrightarrow{(\text{行}1)-(A_{11}^{-1}A_{12})(\text{行}2)} \left[\begin{array}{cc:cc} I & 0 & A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & I & 0 & A_{22}^{-1} \end{array}\right].$$

故

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

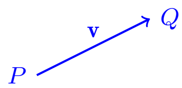
□

# 第2章　欧氏空间与行列式

## 2.1　向量

向量是一种既有长度（大小），又有方向的量. 我们用三种方式表示$n$维欧式空间中的向量.

第一种方法是几何表示：用一个带长度的箭头表示向量. 若箭头的起点为$P$，终点为$Q$，则记为$\overrightarrow{PQ}$. 我们也常在箭头旁写生一个粗体字母$\mathbf{v}$（或$\vec{v}$)来记录它.




第二种方法是数组表示，在$n$维欧式空间中建立直角坐标系后，每个向量都对应一个有序数组$\mathbf{v} = (v_1, v_2, \ldots, v_n)$，我们称其为坐标向量. 而整个空间可以用$\mathbb{R}^n$表示，即

$$\mathbb{R}^n = \{\mathbf{v} = (v_1, v_2, \ldots, v_n) : v_1, \ldots, v_n \in \mathbb{R}\}.$$

第三种方法是列矩阵表示，即

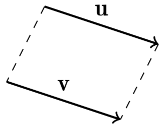
$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

在线性代数中，我们将更多地使用列矩阵来表示向量，并逐步建立涉及矩阵的理论.

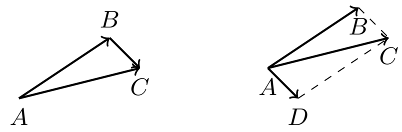
向量$\mathbf{v}$的长度记为$\|\mathbf{v}\|$，称为$\mathbf{v}$的模、模长或范数. 我们将长度为0的向量称为零向量，记为$\mathbf{0}$，它的方向是可以任意指定的. 在几何表示中，起点与终点重合时便得到零向量，即$\overrightarrow{PP} = \mathbf{0}$，而在数组表示或列矩阵表示中，分别有$\mathbf{0} = (0, 0, \ldots, 0)$以及

$$\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

在几何表示中，具有相同方向与长度的向量是相等的，如下图平行四边形对边上的向量$\mathbf{u}$与$\mathbf{v}$，满足$\mathbf{u} = \mathbf{v}$. 在数组表示中，$(u_1, u_2, \ldots, u_n) = (v_1, v_2, \ldots, v_n)$当且仅当$u_1 = v_1, \ldots, u_n = v_n$. 在列矩阵表示中，两个向量相等即列矩阵相等.




向量表示的加法可以使用下图中的三角形法则$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$，亦或平行四边形法则$\overrightarrow{AB} + \overrightarrow{AD} = \overrightarrow{AC}$.




(a) 三角形法则　　(b) 平行四边形法则

数组表示的向量相加为：

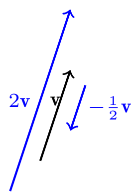
$$(u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n) = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

而列矩阵表示的向量加法即矩阵的加法.

每个向量都有负向量：$\overrightarrow{PQ} + \overrightarrow{QP} = \overrightarrow{PP} = \mathbf{0}$，因此可记$\overrightarrow{QP} = -\overrightarrow{PQ}$. 负向量的长度与原向量相同，而方向相反.


向量的减法可以定义为$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$，或$\overrightarrow{OA} - \overrightarrow{OB} = \overrightarrow{BA}$.

设$k \in \mathbb{R}$，而$\mathbf{v} \in \mathbb{R}^n$. 几何表示的数乘$c\mathbf{v}$如下图所示，它是$\mathbb{R}^n$中的向量，长度满足$\|k\mathbf{v}\| = |k| \cdot \|\mathbf{v}\|$. 当$k > 0$时，方向与$\mathbf{v}$相同；当$k < 0$时，方向与$\mathbf{v}$相反；而$k = 0$时，等于零向量. 特别地，当$\mathbf{u} = c\mathbf{v}$或$\mathbf{v} = c\mathbf{u}$ $(c \in \mathbb{R})$时，我们可以称向量$\mathbf{u}$,$\mathbf{v}$是共线的.




数组表示的数乘如下所示：

$$k(v_1, v_2, \ldots, v_n) = (kv_1, kv_2, \ldots, kv_n).$$

而列矩阵表示的数乘就是矩阵的数乘.

向量的运算具有下列性质.

**定理 2.1.** 设$\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$，而$\lambda, \mu \in \mathbb{F}$. 则

- $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
- $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
- $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
- $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

- $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$.
- $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$.
- $\mu(\lambda\mathbf{v}) = (\mu\lambda)\mathbf{v}$.
- $1\mathbf{v} = \mathbf{v}$.

- $0\mathbf{v} = \mathbf{0}$.
- $\lambda\mathbf{0} = \mathbf{0}$.
- $(-1)\mathbf{v} = -\mathbf{v}$.

设$\mathbf{w}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in \mathbb{R}^n$，而$k_1, k_2, \ldots, k_r \in \mathbb{R}$. 若

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r,$$

则$\mathbf{w}$称为向量$\mathbf{w}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$的一个线性组合，而$k_1, k_2, \ldots, k_r$称为该线性组合的系数.

列矩阵表示的向量的线性组合与前一章中的定义是一致的，且可以写成如下分块矩阵的形式

$$\left[\begin{array}{c} \\ \mathbf{w} \\ \\ \end{array}\right] = \left[\begin{array}{c:c:c:c} & & & \\ \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_r \\ & & & \end{array}\right] \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_r \end{bmatrix}.$$

比如，线性方程组

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

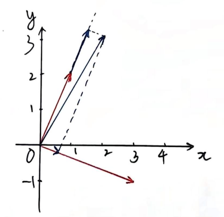
的矩阵形式为

$$\left[\begin{array}{c:c} 1 & 3 \\ 2 & -1 \end{array}\right] \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

将系数矩阵按列作分块，可以得到

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

在$xOy$平面中，我们将上述向量画出. 把向量$(2,3)$按照$(1,2)$和$(3,-1)$的方向作分解，线性方程组的解$x$，




$y$便对应这样的线性组合的系数.

**例 2.2.** 设$\mathbf{u} = (2,1,0,1)$, $\mathbf{v} = (-2,3,1,0)$, $\mathbf{w} = (-8,8,3,-1)$. 试问$\mathbf{w}$是否$\mathbf{u}$，$\mathbf{v}$的线性组合？

解. 考虑线性方程组

$$\left[\begin{array}{c:c} 2 & -2 \\ 1 & 3 \\ 0 & 1 \\ 1 & 0 \end{array}\right] \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} -8 \\ 8 \\ 3 \\ -1 \end{bmatrix}.$$

对增广矩阵作初等行变换，化为简约阶梯阵，可得

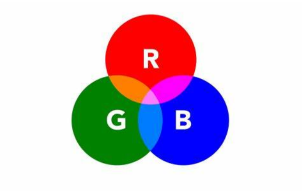
$$\begin{bmatrix} 2 & -2 & -8 \\ 1 & 3 & 8 \\ 0 & 1 & 3 \\ 1 & 0 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

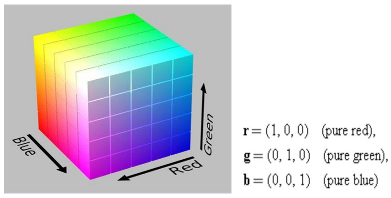
方程组具有唯一解$k_1 = -1$, $k_2 = 3$. 因此$\mathbf{w}$是$\mathbf{u}$与$\mathbf{v}$的线性组合，且

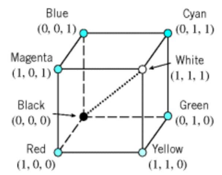
$$\mathbf{w} = -\mathbf{u} + 3\mathbf{v}.$$

□

在生活中，线性组合也是很基础的概念，比如电脑中使用的RGB颜色，就是把每一种颜色都表示成了光学三原色红、绿、蓝的线性组合.








在$\mathbb{R}^n$中，几何表示的向量长度为箭头中线段的长度，而数组表示的向量$\mathbf{v} = (v_1, v_2, \dots, v_n)$的模长为

$$\|\mathbf{v}\| = (v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2}.$$

**定理 2.3.** 设$\mathbf{v}$为$\mathbb{R}^n$中的向量，而$k \in \mathbb{R}$. 则有

- ◇ $\|\mathbf{v}\| \geq 0$
- ◇ $\|\mathbf{v}\| = 0$当且仅当$\mathbf{v} = 0$
- ◇ $\|k\mathbf{v}\| = |k| \cdot \|\mathbf{v}\|$.

我们将模长为1的向量称为单位向量，它给出了一个确定的方向. 对任何一个非零向量$\mathbf{v}$，有

$$\mathbf{v} = \|\mathbf{v}\| \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

其中$\|\mathbf{v}\|$给出了长度的信息，而

$$\left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| = \left| \frac{1}{\|\mathbf{v}\|} \right| \|\mathbf{v}\| = 1,$$

即$\frac{\mathbf{v}}{\|\mathbf{v}\|}$给出了方向的信息，我们称它为$\mathbf{v}$的单位化. 例如，与$\mathbf{v} = (2, -1, -2)$同向的单位向量为

$$\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{(2, -1, 2)}{\sqrt{2^2 + (-1)^2 + (-2)^2}} = \left( \frac{2}{3}, -\frac{1}{3}, -\frac{2}{3} \right).$$

在$\mathbb{R}^n$, 我们称

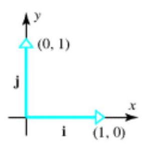
$$\mathbf{e}_1 = (1, 0, 0, \dots, 0), \mathbf{e}_2 = (0, 1, 0, \dots, 0), \dots, \mathbf{e}_n = (0, 0, 0, \dots, 1).$$

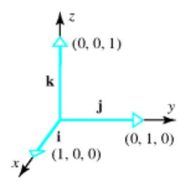
为标准单位向量，它们给出了直角坐标系各个坐标轴的正向. 特别地，二位空间$\mathbb{R}^2$的标准单位向量常记为

$$\mathbf{i} = (1, 0), \quad \mathbf{j} = (0, 1).$$

而三维空间$\mathbb{R}^3$的标准单位向量常记为

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






当我们把一个向量$\mathbf{v} = (v_1, v_2, \dots, v_n)$写成标准单位向量的线性组合时，系数就是对应的分量：

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \ldots + v_n\mathbf{e}_n.$$

在$\mathbb{R}^n$中，两个向量$\mathbf{u}$和$\mathbf{v}$的（欧式）距离定义为$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|$. 设$\mathbf{u} = (u_1, u_2, \ldots, u_n)$，$\mathbf{v} = (v_1, v_2, \ldots, v_n)$，则

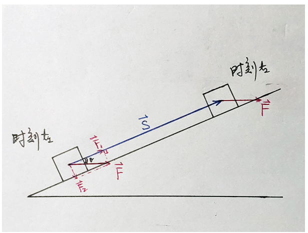
$$d(\mathbf{u},\mathbf{v}) = \sqrt{(u_1-v_1)^2 + (u_2-v_2)^2 + \cdots + (u_n-v_n)^2}.$$

## 2.2 内积与正交性

向量点乘运算的背景与物理中力做的功有关. 若物体在如图水平力$\mathbf{F}$的作用下，在斜坡上的位移为$\mathbf{s}$，在这个过程中力$\mathbf{F}$做的功是多少？我们将力$\mathbf{F}$作平行四边形分解，得到平行于斜坡的分力$\mathbf{F}_1$与垂直于斜坡的分力$\mathbf{F}_2$. 对物体运动真正起作用的部分是分力$\mathbf{F}_1$，因此功为

$$W = \|\mathbf{s}\|\|\mathbf{F}_1\| = \|\mathbf{s}\|\|\mathbf{F}\|\cos\theta,$$

其中$\theta$为向量$\mathbf{F}$与的夹角.




在本章中，对$\mathbb{R}^n$ $(n=2,3)$中的非零向量$\mathbf{u}$和$\mathbf{v}$，我们用$\theta(\mathbf{u},\mathbf{v})$表示它们之间的夹角.

设$\mathbf{u}$和$\mathbf{v}$为$\mathbb{R}^n$ $(n=2,3)$中的非零向量，我们定义

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta(\mathbf{u},\mathbf{v}), \tag{2.1}$$

称为$\mathbf{u}$与$\mathbf{v}$的点乘或内积. 而当$\mathbf{u}=\mathbf{0}$或$\mathbf{v}=\mathbf{0}$时，定义$\mathbf{u}\cdot\mathbf{v}=0$. 当$\mathbf{u}=\mathbf{0}$或$\mathbf{v}=\mathbf{0}$时，该零向量的方向是可以任取的，夹角的写法并不唯一，但它们的内积为零，也符合(2.1)式子. 对于坐标表示的向量，我们可使用以下定义：

**定义 2.1.** 设$\mathbf{u} = (u_1, u_2, \ldots, u_n)$，$\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$. 我们定义

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + \ldots + u_nv_n,$$

称为$\mathbf{u}$与$\mathbf{v}$的内积.

在$\mathbb{R}^2$或$\mathbb{R}^3$中，几何表示与坐标表示的两种定义之间，可以用余弦定理互推，我们把推导过程留给读者.

**定理 2.4.** 设$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. 则

- $\mathbf{v}\cdot\mathbf{v} = \|\mathbf{v}\|^2$.
- $|\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\|\mathbf{v}\|$.

后者被称为柯西不等式，它的坐标形式为

$$\left(\sum_{k=1}^n u_kv_k\right)^2 = \left(\sum_{k=1}^n u_k^2\right)\left(\sum_{k=1}^n v_k^2\right).$$

由于$-1 \le \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \le 1$，在$\mathbb{R}^n$ $(n \ge 4)$中，我们也可以定义两个非零向量$\mathbf{u} = (u_1, u_2, \ldots, u_n)$，$\mathbf{v} = (v_1, v_2, \ldots, v_n)$的夹角：

$$\theta(\mathbf{u},\mathbf{v}) = \arccos\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

进一步，我们可以得到模长以及距离的三角不等式.

**推论 2.5.** 设$\mathbf{u}, \mathbf{v}$是$\mathbb{R}^n$中的向量. 则有

◇ $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\|+\|\mathbf{v}\|$.

◇ $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w})+d(\mathbf{w}, \mathbf{v})$.

向量的内积具有对称性、齐次性，并满足分配律.

**定理 2.6.** 设$k \in \mathbb{F}$，而$\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$. 则有

○ $\mathbf{v} \cdot \mathbf{u}=\mathbf{u} \cdot \mathbf{v}$.

○ $(k\mathbf{u}) \cdot \mathbf{v}=k(\mathbf{u} \cdot \mathbf{v})$.

○ $\mathbf{u} \cdot(\mathbf{v}+\mathbf{w})=\mathbf{u} \cdot \mathbf{v}+\mathbf{u} \cdot \mathbf{w}$.

**例 2.7.** 设$a, b, c, d \in \mathbb{F}$，而$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. 试展开$(a\mathbf{u}+b\mathbf{v}) \cdot(c\mathbf{u}+d\mathbf{v})$.

解. 使用两次分配律，可得

$$
\begin{aligned}
(a\mathbf{u}+b\mathbf{v}) \cdot(c\mathbf{u}+d\mathbf{v}) & =(a\mathbf{u}) \cdot(c\mathbf{u}+d\mathbf{v})+(b\mathbf{v}) \cdot(c\mathbf{u}+d\mathbf{v}) \\
& =(a\mathbf{u}) \cdot(c\mathbf{u})+(a\mathbf{u}) \cdot(d\mathbf{v})+(b\mathbf{v}) \cdot(c\mathbf{u})+(b\mathbf{v}) \cdot(d\mathbf{v}).
\end{aligned}
$$

利用齐次性，上式等于

$$
a(\mathbf{u} \cdot \mathbf{u})+b(\mathbf{u} \cdot \mathbf{v})+c(\mathbf{v} \cdot \mathbf{u})+d(\mathbf{v} \cdot \mathbf{v}).
$$

再利用对称性以及内积与模长的关系，得

$$
\text{原式}=a\|\mathbf{u}\|^2+(b+c)(\mathbf{u} \cdot \mathbf{v})++d\|\mathbf{v}\|^2.
$$

□

特别地，我们有

$$
\|\mathbf{u} \pm \mathbf{v}\|^2=\|\mathbf{u}\|^2 \pm 2(\mathbf{u} \cdot \mathbf{v})+\|\mathbf{v}\|^2. \tag{2.2}
$$

将上述加、减的式子两边同时相加或相减，便得到平行四边形公式和极化恒等式.

**定理 2.8** (平行四边形公式)**.** 设$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. 则

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

**定理 2.9** (极化恒等式)**.** 设$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. 则

$$
\mathbf{u} \cdot \mathbf{v}=\frac{1}{4}\left(\|\mathbf{u}+\mathbf{v}\|^2-\|\mathbf{u}-\mathbf{v}\|^2\right).
$$

平行四边形公式的含义为：平行四边形两条斜边长的平方和等于四条边长的平方和. 而极化恒等式的意义在于，若想证明与内积有关的性质，只需证明与模长有关的对应性质就可以了. 前者与两个任取的向量有关，而后者只涉及一个任取的向量，因此更容易被证明.

在本节的最后，我们再来看看列坐标形式的内积. 设$\mathbf{u}, \mathbf{v} \in M_{n \times 1}$为列向量，则它们的内积为

$$
\mathbf{u} \cdot \mathbf{v}=\mathbf{v}^T \mathbf{u}.
$$

**例 2.10.** 设$A \in M_{m \times n}$, $u \in M_{n \times 1}$, 而$v \in M_{m \times 1}$. 试验证

$$
(A\mathbf{u}) \cdot \mathbf{v}=\mathbf{u} \cdot(A^T \mathbf{v}).
$$

证明. 需要注意的是，等式左端的内积是$\mathbb{R}^m$上的内积，而等式右端的内积是$\mathbb{R}^n$上的内积，它们并不是同一个内积. 计算得

$$
(A\mathbf{u}) \cdot \mathbf{v}=\mathbf{v}^T(A\mathbf{u})=(\mathbf{v}^T A)\mathbf{u}=(A^T \mathbf{v})^T \mathbf{u}=\mathbf{u} \cdot(A^T \mathbf{v}).
$$

□

在二维或三位空间中，我们常说夹角为$\pi/2$的两个非零向量$\mathbf{u},\mathbf{v}$是“垂直的”，这时

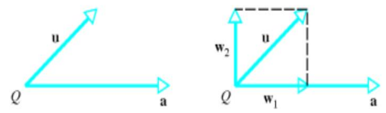
$$\mathbf{u}\cdot\mathbf{v}=\|\mathbf{u}\|\|\mathbf{v}\|\cos\frac{\pi}{2}=0.$$

**定义 2.2.** 在$\mathbb{R}^n$中，我们称两个向量$\mathbf{u}$与$\mathbf{v}$为正交的或垂直的，是指它们满足$\mathbf{u}\cdot\mathbf{v}=0$. 我们称非零向量所构成的集合$S$为正交集，是指$S$中的向量两两正交. 进一步，若$S$中的向量均为单位向量，我们称$S$为标准正交集.

例如，在$\mathbb{R}^4$中，向量$\mathbf{u}=(-2,3,1,4)$与$\mathbf{v}=(1,2,0,-1)$正交，这是因为

$$\mathbf{u}\cdot\mathbf{v}=(-2)\cdot 1+3\cdot 2+1\cdot 0+4\cdot(-1)=0.$$

而在$\mathbb{R}^3$中，$\{\mathbf{i},\mathbf{j},\mathbf{k}\}$构成一个标准正交集合. 正交集这个定义的重要意义在于，我们可以用它的向量构造出直角坐标轴来.




在高中物理中，我们经常对力作垂直分解. 利用正交性的概念，现在我们对这种分解的存在性和唯一性给出证明.

**定理 2.11.** （正交分解定理）设$\mathbf{w}$，$\mathbf{a}\in\mathbb{R}^n$，其中$\mathbf{a}\neq\mathbf{0}$. 则$\mathbf{w}$可被唯一地表示为$\mathbf{w}=\mathbf{w}_1+\mathbf{w}_2$，其中$\mathbf{w}_1$与$\mathbf{a}$共线，而$\mathbf{w}_2$与$\mathbf{a}$正交.

*分析*. 为了证明分解的存在性，我们需要找到$w_1$与$w_2$的表达式. 不妨先假设这样的分解$\mathbf{w}=\mathbf{w}_1+\mathbf{w}_2$是存在的，那么

$$\mathbf{w}_1=k\mathbf{a},\quad \mathbf{w}_2\cdot\mathbf{a}=0. \tag{2.3}$$

其中$k\in\mathbb{F}$. 我们只需找到合适的$k$即可. 注意到$\mathbf{u},\mathbf{a}$都是定理条件中已经给出的量，我们把未知的$\mathbf{w}_2$用$\mathbf{u}-\mathbf{w}_1$代替，得到

$$0=\mathbf{w}_2\cdot\mathbf{w}_a=(\mathbf{u}-\mathbf{w}_1)\cdot\mathbf{a}=\mathbf{u}\cdot\mathbf{a}-(k\mathbf{a})\cdot\mathbf{a}=\mathbf{u}\cdot\mathbf{a}-k\|\mathbf{a}\|.$$

注意到$\mathbf{a}\neq\mathbf{0}$，因此$\|\mathbf{a}\|\neq 0$. 于是

$$k=\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}.$$

□

*证明*. 先证明存在性，取

$$\mathbf{w}_1=\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a},\quad \mathbf{w}_2=\mathbf{w}-\mathbf{w}_1.$$

易知$\mathbf{w}_1$与$\mathbf{a}$共线. 验证得

$$\mathbf{w}_2\cdot\mathbf{a}=(\mathbf{w}-\mathbf{w}_1)\cdot\mathbf{a}=\mathbf{w}\cdot\mathbf{a}-(\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a})\cdot\mathbf{a}.$$

注意到

$$(\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a})\cdot\mathbf{a}=\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}(\mathbf{a}\cdot\mathbf{a})=\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\|\mathbf{a}\|^2=\mathbf{u}\cdot\mathbf{a}.$$

故$\mathbf{w}_2\cdot\mathbf{a}=0$. 存在性得证.

为了证明唯一性，我们考虑$\mathbf{w}$的任何一个分解$\mathbf{w}=\mathbf{w}_1+\mathbf{w}_2$，其中的$\mathbf{w}_1$与$\mathbf{w}_2$满足(2.3)式，由(2.3)式之后的推导可知，系数$k$的取法是唯一的，只能等于$\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}$. 唯一性得证. □

用上述定理的符号，我们记$\mathrm{proj}_{\mathbf{a}}\mathbf{u}=\mathbf{w}_1$，并称它为$\mathbf{u}$平行于$\mathbf{a}$的分量, 而称$\mathbf{w}_2$为$\mathbf{u}$垂直于$\mathbf{a}$的分量. 正交投影的公式为

$$\mathrm{proj}_{\mathbf{a}}\mathbf{u}=\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a},$$

而投影长度为

$$\|\mathrm{proj}_{\mathbf{a}}\mathbf{u}\|=\left|\frac{\mathbf{u}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\right|\cdot\|\mathbf{a}\|=\frac{|\mathbf{u}\cdot\mathbf{a}|}{\|\mathbf{a}\|}.$$

**例 2.12.** *设$\mathbf{u}=(2,-1,3)$，而$\mathbf{a}=(4,-1,2)$. 求$\mathbf{u}$平行于$\mathbf{a}$与垂直于$a$的分量.*

解.（略.） □

**定理 2.13** (勾股定理)**.** *设$\mathbf{u}$,$\mathbf{v}$为$\mathbb{R}^n$中正交的向量，则*

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

*证明.* 在(2.2)式中待入$\mathbf{u}\cdot\mathbf{v}=0$，则定理得证. □

在三维空间中，直线与平面是“对偶”的，平面都具有法向量与垂线，而直线都有法平面. 假设$\mathbf{n}$是一个非零向量，且与平面$\Pi$中的任何一个向量均正交，则称$\mathbf{n}$为平面$\Pi$的一个法向量.

**定理 2.14.** *设平面$\Pi$过$M_0(x_0,y_0,z_0)$并具有法向量$\mathbf{n}=(A,B,C)$. 则$\Pi$满足方程*

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0.$$

*上式称为平面$\Pi$的点法式方程.*

*证明.* 设点M坐标为$(x,y,z)$. 则$M\in\Pi$当且仅当$\overrightarrow{M_0M}$与$\mathbf{n}$正交. 注意到$\overrightarrow{M_0M}=(x-x_0,y-y_0,z-z_0)$，而$\mathbf{n}=(A,B,C)$. 因此$\overrightarrow{M_0M}\cdot\mathbf{n}=0$当且仅当

$$A(x-x_0)+B(y-y_0)+C(z-z_0)=0$$

□

上述点法式方程可写为$Ax+By+Cz+(-Ax_0-By_0-Cz_0)=0$，其中$-Ax_0-By_0-Cz_0=D$为一个数，这时我们可以得到平面的一般式方程

$$Ax+By+Cz+D=0.$$

从以上一般式方程可以立刻读出平面的一个法向量$\mathbf{n}=(A,B,C)$,它是非零向量,因此一次项系数$A,B,C$不能同时为0. 从线性方程组的角度来看，这时系数矩阵的秩为1，因此具有1个主元，2个自由变元. 当自由变元任意变动时，便形成了一个平面.

在一般式方程中，若$D=0$，则意味着平面过原点. 若$A=0$，便意味着平面平行于与$x$轴. 若$A=B=0$，则意味着法向量与$z$轴平行，从而平面平行于$xOy$平面.

**定理 2.15.** *设直线$l$过点$M_0(x_0,y_0,z_0)$，且平行于非零向量$\mathbf{s}=(m,n,p)$. 则直线$l$满足方程*

$$\begin{cases} x=x_0+tm,\\ y=y_0+tn, & ,\quad (t\in\mathbb{R}).\\ z=z_0+tp, \end{cases}$$

*上式称为直线$l$的参数式方程. 而向量$\mathbf{s}$称为直线$l$的方向向量.*

证明. 设点M坐标为$(x, y, z)$. 则$M \in l$当且仅当$\overrightarrow{M_0M}$与$\mathbf{s}$平行. 注意到$\overrightarrow{M_0M} = (x - x_0, y - y_0, z - z_0)$，而$\mathbf{s} = (m, n, p)$，则存在$t \in \mathbb{R}$使得

$$(x - x_0, y - y_0, z - z_0) = t(m, n, p).$$

参数式方程得证. □

如果不想引入新变量$t$，我们可以使用直线的点向式方程：

$$\frac{x - x_0}{m} = \frac{y - y_0}{n} = \frac{z - z_0}{p}.$$

从点向式方程的分子上，可以读取直线上的一点$(x_0, y_0, z_0)$，而从分母上，可以读取直线的一个方向向量$(m, n, p)$. 注意到方向向量的某些分量是有可能为0的，因此我们在点向式中作额外的约定：当分母为0是，表示分子也同时为0. 这样点向式就包含了所有可能的情况了.

另外，当两个平面不平行、不重合时，会交出一条直线. 因此直线方程可以用两个平面的方程联立：

$$\begin{cases} A_1x + B_1y + C_1z + D_1 = 0, \\ A_2x + B_2y + C_2z + D_2 = 0, \end{cases}$$

其中系数$(A_1, B_1, C_1)$与$(A_2, B_2, C_2)$不成比例. 上式称为直线的一般式方程. 从线性方程组的角度看，一次项系数不成比例等价于系数矩阵

$$\mathsf{rank}\left(\begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix}\right) = 2.$$

这时增广矩阵的秩也一定为2，从而方程组有解，且具有2个主元和1个自由变元，当自由变元任意变动时，便形成了一条直线.

**例 2.16.** *设直线$l$的一般式方程为*

$$\begin{cases} A_1x + B_1y + C_1z = 0, \\ A_2x + B_2y + C_2z = 0. \end{cases}$$

*其中*

$$\mathit{rank}\left(\begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix}\right) = 2.$$

*求其点向式方程.*

解. 这是一个齐次线性方程组，根据之前的分析，恰有1个自由变元. 我们不妨先假设$z$是自由变元，将含有$z$的项移到等式右边.

$$\begin{cases} A_1x + B_1y = -C_1z, \\ A_2x + B_2y = -C_2z. \end{cases}$$

由例1.36，当

$$A_1B_2 - A_2B_1 \neq 0 \tag{2.4}$$

时，方程组的解为

$$x = \frac{(-C_1Z)B_2 - (-C_2z)B_1}{A_1B_2 - A_2B_1} = \frac{B_1C_2 - B_2C_1}{A_1B_2 - A_2B_1}z,$$
$$y = \frac{A_1(-C_2Z) - A_2(-C_1z)}{A_1B_2 - A_2B_1} = \frac{C_1A_2 - C_2A_1}{A_1B_2 - A_2B_1}z.$$

这里之所以需要条件(2.4)式，是因为它出现在了分母上. 为了使结果更具有对称性以及一般性，我们不妨令$z=(A_1B_2-A_2B_1)t$，其中$t\in\mathbb{R}$为参数. 这时可得到直线的参数式方程

$$\begin{cases} x=(B_1C_2-B_2C_1)t, \\ y=(C_1A_2-C_2A_1)t \\ z=(A_1B_2-A_2B_1)t. \end{cases}$$

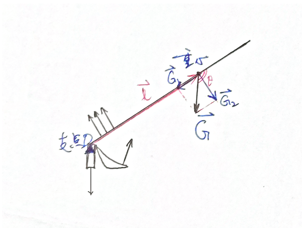
可以验证以上$(x,y,z)$的表达式是原方程组的解，且不再依赖于条件(2.4). 进一步，我们可以将解都写成二阶行列式的格式：

$$x=\begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix} t, \quad y=\begin{vmatrix} C_1 & A_1 \\ C_2 & A_2 \end{vmatrix} t, \quad z=\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} t.$$

□

## 2.3 叉积与混合积

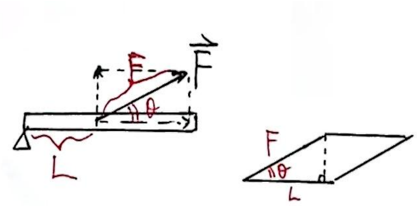
向量叉积的概念与物理里的转动有关.




在高中物理的平面转动中，将力沿着平行与垂直与力臂的方向作分解，转动趋势来自于垂直与力臂的分力. 因而力矩定义为

$$M=FL\sin\theta,$$

其中$F$是力的大小，$L$是力臂的长度，而$\theta$是它们两者的夹角. 物体保持平衡的条件是力矩平衡. 值得注意



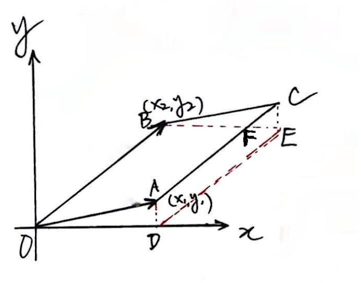

的是，这个量的大小与以$F$和$L$为边的平行四边形的面积相同.

**例 2.17.** 设$\mathbb{R}^2$中的向量$\overrightarrow{OA}=(x_1,y_1)$和$\overrightarrow{OB}=(x_2,y_2)$如图中所示. 证明它们张成的平行四边形$OACB$的面积为

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = x_1y_2-x_2y_1.$$

证明. 如图作辅助线，用$S$表示对应多边形的面积，利用全等三角形的面积相同，我们可以推断

$$\begin{aligned} S_{OACB} &= S_{OAFB}+S_{BFC}+S_{FEC}-S_{FEC}=S_{OAFB}+S_{BEC}-S_{FEC} \\ &= S_{OAFB}+S_{ODA}-S_{FEC}=S_{OAFB}+S_{ODA}+S_{ADEF}-S_{ADEF}-S_{FEC}=S_{ODEB}-S_{ADEC}. \end{aligned}$$




注意到$A(x_1,y_1)$, $B(x_2,y_2)$以及$C(x_1+x_2,y_1+y_2)$，我们有

$$S_{ODEB}=x_1y_2,\quad S_{ADEC}=y_1(x_1+x_2-x_1)=x_2y_1.$$

定理得证. $\square$

注意到等式右端的$x_1y_2-x_2y_1$是可能取正号、零或负号的，而面积总为非负值. 在图中，向量$\overrightarrow{OA}$按逆时针旋转不超过$\pi$弧度，可以转到向量$\overrightarrow{OB}$的方向，因此得到正号. 若交换两者顺序，变为顺时针旋转的关系，则会得到负号. 我们把这样带着正负号的面积值称为由$\overrightarrow{OA}$与$\overrightarrow{OB}$张成的平行四边形的有向面积. 而2阶行列式的几何意义便是这样的有向面积.

接下来，我们考虑三维空间中的转动. 本节的向量都是三维欧氏空间$\mathbb{R}^3$中的向量.

地球仪有一根固定的轴，而球体可以绕着这条轴旋转. 在之后的章节中，我们会证明三维空间中的旋转一定有固定轴，因此可以分解为固定轴以及与轴垂直的平面中的二维旋转. 二维旋转有两个可能的方向，我们如何来标记想要的方向呢？将右手大拇指伸直而另四指并拢并自然弯曲，当四指弯曲的方向与该二维旋转的方向一致时，大拇指的方向便记为旋转轴的正向，这种方式称为右手螺旋法则，是数学与物理中通用的一个约定. 更近一步，若向量$\mathbf{w}$与$\mathbf{u}$, $\mathbf{v}$均正交，且当右手四指从$\mathbf{u}$转向$\mathbf{v}$时，大拇指的方向与$\mathbf{w}$的夹角小于$\pi$，则称$\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$成右手系.

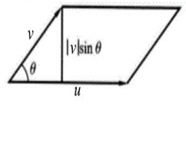
**定义 2.3.** 设$\mathbf{u}$,$\mathbf{v}$是$\mathbb{R}^3$中的两个非零向量. 它们的叉乘或叉积，记为$\mathbf{u}\times\mathbf{v}$，定义为满足以下条件的向量：

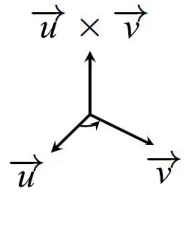
*(i)* $\|\mathbf{u}\times\mathbf{v}\|=\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta(\mathbf{u},\mathbf{v})$;

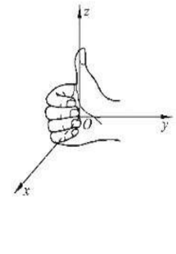
*(ii)* $\mathbf{u}\times\mathbf{v}$与$\mathbf{u}$均正交$\mathbf{v}$，且$\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}\times\mathbf{v}$三者成右手系.

当$\mathbf{u}=0$或$\mathbf{v}=0$时，定义$\mathbf{u}\times\mathbf{v}=\mathbf{0}$.

定义中的(i)告诉我们，向量$\mathbf{u}\times\mathbf{v}$的模长等于由$\mathbf{u}$和$\mathbf{v}$张成的平行四边形的面积. 如果将“面积”更换为“有向面积”，它的方向就给出了$\mathbf{u}\times\mathbf{v}$的方向. 另外，若$\mathbf{u}\times\mathbf{v}=\mathbf{0}$，意味着要么$\mathbf{u}$或$\mathbf{v}$为零向量，要么$\theta=0$或$\theta=\pi$. 即$\mathbf{u}\times\mathbf{v}=\mathbf{0}$当且仅当向量$\mathbf{u}$, $\mathbf{v}$是共线的.




**例 2.18.** 三维欧式空间$\mathbb{R}^3$中的标准单位向量$\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$满足

$$\mathbf{i}\times\mathbf{j}=\mathbf{k},\quad \mathbf{j}\times\mathbf{k}=\mathbf{i},\quad \mathbf{k}\times\mathbf{i}=\mathbf{j},$$

$$\mathbf{j}\times\mathbf{i}=-\mathbf{k},\quad \mathbf{k}\times\mathbf{j}=-\mathbf{i},\quad \mathbf{i}\times\mathbf{k}=-\mathbf{j}.$$

注意到

$$\mathbf{i}\times(\mathbf{j}\times\mathbf{j})=\mathbf{i}\times\mathbf{0}=\mathbf{0},$$

而

$$(\mathbf{i}\times\mathbf{j})\times\mathbf{j}=\mathbf{k}\times\mathbf{j}=-\mathbf{i}.$$

因此向量的叉积不满足结合律.

**定理 2.19.** 设$\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^3$，而$k\in\mathbb{F}$. 则

- ◇ $\mathbf{u}\times\mathbf{v}=-(\mathbf{v}\times\mathbf{u})$.
- ◇ $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v})=0$.
- ◇ $\mathbf{v}\cdot(\mathbf{u}\times\mathbf{v})=0$.
- ◇ $k(\mathbf{u}\times\mathbf{v})=(k\mathbf{u})\times\mathbf{v}=\mathbf{u}\times(k\mathbf{v})$.
- ◇ $\mathbf{u}\times\mathbf{0}=\mathbf{0}\times\mathbf{u}=\mathbf{0}$.
- ◇ $\mathbf{u}\times\mathbf{u}=\mathbf{0}$.
- ◇ $\|\mathbf{u}\times\mathbf{v}\|^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2$.

*证明.* 我们只证明最后一式，其余证明的留给读者完成. 根据定义，有

$$\|\mathbf{u}\times\mathbf{v}\|^2=\big(\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta(\mathbf{u},\mathbf{v})\big)^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2(1-\cos^2\theta(\mathbf{u},\mathbf{v}))=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2.$$

□

下面我们来探索叉积在坐标表示下的公式.

**定理 2.20.** 设$\mathbf{u}=(u_1,u_2,u_3)$, $\mathbf{v}=(v_1,v_2,v_3)$. 则

$$\mathbf{u}\times\mathbf{v}=\left(\begin{vmatrix}u_2 & u_3\\ v_2 & v_3\end{vmatrix},\begin{vmatrix}u_3 & u_1\\ v_3 & v_1\end{vmatrix},\begin{vmatrix}u_1 & u_2\\ v_1 & v_2\end{vmatrix}\right).$$

*思路.* 设$\mathbf{u}\times\mathbf{v}=(x,y,z)$，因为它与$\mathbf{u}$, $\mathbf{v}$均正交，所以满足线性方程组

$$\begin{cases}u_1x+u_2y+u_3z=0,\\ v_1x+v_2y+v_3z=0.\end{cases}$$

由例2.16，知

$$(x,y,z)=t\left(\begin{vmatrix}u_2 & u_3\\ u_2 & v_3\end{vmatrix},\begin{vmatrix}u_3 & u_1\\ v_3 & v_1\end{vmatrix},\begin{vmatrix}u_1 & u_2\\ v_1 & v_2\end{vmatrix}\right),$$

其中$t\in\mathbb{R}$. 将具体表达式带入公式

$$\|\mathbf{u}\times\mathbf{v}\|^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2,$$

对比系数可得$|t|=1$. 为了确定正负号，我们先考虑$\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}\times\mathbf{v}$分别为$\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$的特殊情形，带入具体值可得$t=1$. 而向量$\mathbf{u}\times\mathbf{v}$随着向量$\mathbf{u}$与$\mathbf{v}$的连续变化而连续变化, 因此$t$不可能从1变为$-1$，故恒有$t=1$.

□

**例 2.21.** 求与$\mathbf{u}=(3,-2,4)$和$\mathbf{v}=(1,1,-2)$均垂直的单位向量.

*解.* (略.) □

注意到

$$\|\mathbf{u}\times\mathbf{v}\| = \|\mathbf{u}\|\|\mathbf{v}\|\sin\theta(\mathbf{u},\mathbf{v})$$

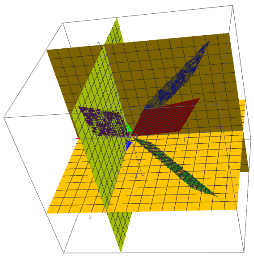
给出的仍然是由$\mathbf{u}$, $\mathbf{v}$张成的平行四边形的面积. 我们亦可以把$\mathbf{u}\times\mathbf{v}$理解为这个平行四边形的“有向面积”，其方向就是$\mathbf{u}\times\mathbf{v}$所指的方向. 这时我们可以把这个平行四边形投影到三个坐标平面中. 设$\mathbf{u}=(u_1,u_2,u_3)$，$\mathbf{v}=(v_1,v_2,v_3)$时，我们用$\mathbf{u}_{yOz}$表示$\mathbf{u}$在$yOz$平面上正交投影所得的向量，其余类似. 则有

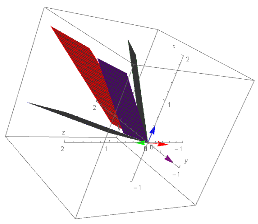
$$\mathbf{v}_{yOz}=(0,v_2,v_3),\quad \mathbf{v}_{xOz}=(v_1,0,v_3),\quad \mathbf{v}_{xOy}=(v_1,v_2,0),$$

$$\mathbf{v}_{yOz}=(0,v_2,v_3),\quad \mathbf{v}_{xOz}=(v_1,0,v_3),\quad \mathbf{v}_{xOy}=(v_1,v_2,0).$$

因此可以得到关于有向面积的勾股定理：

$$\begin{aligned}\|\mathbf{u}\times\mathbf{v}\|^2 &= \begin{vmatrix} u_2 & u_3\\ v_2 & v_3\end{vmatrix}^2+\begin{vmatrix} u_3 & u_1\\ v_3 & v_1\end{vmatrix}^2+\begin{vmatrix} u_1 & u_2\\ v_1 & v_2\end{vmatrix}^2\\ &=\|\mathbf{u}_{yOz}\times\mathbf{v}_{yOz}\|^2+\|\mathbf{u}_{xOz}\times\mathbf{v}_{xOz}\|^2+\|\mathbf{u}_{xOz}\times\mathbf{v}_{xOz}\|^2.\end{aligned}$$





向量叉积运算的一些性质在下述定理中列出，利用定理2.20可进行验证，证明留给读者完成.

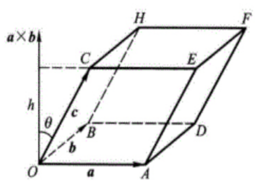
**定理 2.22.** 设$\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^3$.

- ◇ $\mathbf{u}\times(\mathbf{v}+\mathbf{w})=(\mathbf{u}\times\mathbf{v})+(\mathbf{u}\times\mathbf{w})$.
- ◇ $(\mathbf{u}+\mathbf{v})\times\mathbf{w}=(\mathbf{u}\times\mathbf{w})+(\mathbf{v}\times\mathbf{w})$.
- ◇ $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})=(\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{u}\cdot\mathbf{v})\mathbf{w}$.
- ◇ $(\mathbf{u}\times\mathbf{v})\times\mathbf{w}=(\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{u}$.
- ◇ $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})+\mathbf{v}\times(\mathbf{w}\times\mathbf{u})+\mathbf{w}\times(\mathbf{u}\times\mathbf{v})=\mathbf{0}$.

最后一式称为雅可比恒等式，是数学中的一类典型结构的内蕴性质，在物理中也会遇到.

由于两个向量作叉积得到的是一个向量，可以跟另一个向量作内积. 现在我们再来定义混合积.

**定义 2.4.** 在$\mathbb{R}^3$中，三个向量$\mathbf{a},\mathbf{b},\mathbf{c}$的混合即定义为$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$.




混合积$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$是一个数，按几何表示，

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}=\|\mathbf{a}\times\mathbf{b}\|\|\mathbf{c}\|\cos\theta(\mathbf{a}\times\mathbf{b},\mathbf{c}).$$

我们先来说明，其绝对值大小等于由$\mathbf{a},\mathbf{b},\mathbf{c}$所张成的平行六面体的体积. 如图所示将$\mathbf{a}$,$\mathbf{b}$所在平面看成底面，则$\|\mathbf{a}\times\mathbf{b}\|$表示其底面上平行四边形的面积，而平行六面体的高为向量$\mathbf{c}$在向量$\mathbf{a}\times\mathbf{b}$上的正交投影长度，即

$$h=\|\mathrm{proj}_{\mathbf{a}\times\mathbf{b}}\mathbf{c}\|=\frac{|(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}|}{\|\mathbf{a}\times\mathbf{b}\|}.$$

因此由$\mathbf{a},\mathbf{b},\mathbf{c}$所张成的平行六面体的体积为$|(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}|$. 我们再看去掉绝对值后的正负号，它由$\cos\theta(\mathbf{a}\times\mathbf{b},\mathbf{c})$决定，经过分析可以发现，当$\mathbf{a},\mathbf{b},\mathbf{c}$成右手系、共面、左手系时，取值分别为正、零、负. 按照惯例，我们总把右手系作为正的定向，因此混合积$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$表示的是由$\mathbf{a},\mathbf{b},\mathbf{c}$所张成的平行六面体的“有向体积”.

注意到内积具有对称性，而对三个向量的顺序进行轮换仍然是“右手系”，因此以下六项都是同一个混合积：

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}=(\mathbf{b}\times\mathbf{c})\cdot\mathbf{a}=(\mathbf{c}\times\mathbf{a})\cdot\mathbf{b},$$

$$\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})=\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}).$$

设矩阵

$$A=\begin{bmatrix} u_1 & u_2 & u_3\\ v_1 & v_2 & v_3\\ w_1 & w_2 & w_3\end{bmatrix}.$$

我们定义其行列式为向量$\mathbf{u}=(u_1,u_2,u_3)$, $\mathbf{v}=(v_1,v_2,v_3)$, $\mathbf{w}=(w_1,w_2,w_3)$所张成的平行六面体的有向体积，记为$\mathsf{det}(A)$或

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

**定理 2.23.** 设$\mathbf{u}=(u_1,u_2,u_3)$, $\mathbf{v}=(v_1,v_2,v_3)$, $\mathbf{w}=(w_1,w_2,w_3)$. 则

$$\begin{aligned}\begin{vmatrix} u_1 & u_2 & u_3\\ v_1 & v_2 & v_3\\ w_1 & w_2 & w_3\end{vmatrix}&=\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})\\ &=u_1\begin{vmatrix} v_2 & v_3\\ w_2 & w_3\end{vmatrix}-u_2\begin{vmatrix} v_1 & v_3\\ w_1 & w_3\end{vmatrix}+u_3\begin{vmatrix} v_1 & v_2\\ w_1 & w_2\end{vmatrix}\\ &=u_1v_2w_3+u_2v_3w_1+u_3v_1w_2-u_3v_2w_1-u_2v_1w_3-u_1v_3w_2.\end{aligned}$$

证明. 利用定理2.20知

$$\mathbf{v}\times\mathbf{w}=\left(\begin{vmatrix} v_2 & v_3\\ w_2 & w_3\end{vmatrix},\begin{vmatrix} v_3 & v_1\\ w_3 & w_1\end{vmatrix},\begin{vmatrix} v_1 & v_2\\ w_1 & w_2\end{vmatrix}\right).$$

而

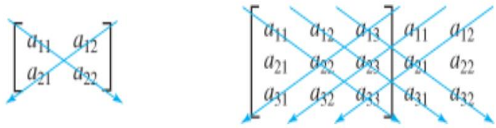
$$\begin{vmatrix} v_3 & v_1\\ w_3 & w_1\end{vmatrix}=v_3w_1-v_1w_3=-(v_1w_3-v_3w_1)=-\begin{vmatrix} v_1 & v_3\\ w_1 & w_3\end{vmatrix},$$

故

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

将这些二阶行列式均展开后便得到了最后一个式子. □

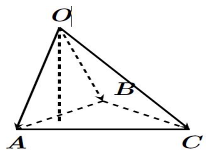
对二阶或三阶行列式，可使用对角线与反对角线的计算技巧. 即与主对角线平行的斜线上的元素乘积取正号，而反向的斜线上的元素乘积取负号.



**例 2.24.** 计算行列式
$$\begin{vmatrix} 1 & 0 & 3 \\ 2 & 5 & 1 \\ 0 & -1 & 2 \end{vmatrix}.$$

解.（略.） □

**例 2.25.** 设$\overrightarrow{OA} = \mathbf{a}$, $\overrightarrow{OB} = \mathbf{b}$, $\overrightarrow{OC} = \mathbf{c}$. 求点$O$到点$A, B, C$所在平面的距离.




解.（略.） □

**例 2.26.** 设$\mathbf{u} = (\mathbf{i} - 2\mathbf{k}) \times (\mathbf{i} + 3\mathbf{j} - 4\mathbf{k})$，而$\mathbf{v} = 3\mathbf{i} - 12\mathbf{j} + 4\mathbf{k}$. 计算$proj_{\mathbf{v}}(\mathbf{u})$的模长.

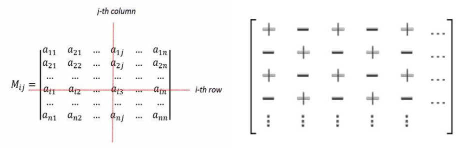
解.（略.） □

## 2.4 行列式

二阶行列式表示平行四边形的有向体积，三阶行列式表示平行六变体的有向体积，那么一阶行列式应该表示什么呢？应该表示的是线段的有向长度. 在一维欧氏空间$\mathbb{R}$中，$\mathbf{a} = (a)$本身就表示一个既有大小有有方向的量，因此定义
$$\det([a]) = a.$$
再回顾二阶、三阶行列式的定义：
$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$
$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = u_1 \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - u_2 \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + u_3 \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix}.$$
发现当$n = 2, 3$时，$n$阶行列式可以展开成一些包含$n-1$阶行列式的正负项的和. 一个$n$阶矩阵$A = [a_{ij}]$的行列式，我们记为$\det(A)$或$|a_{ij}|_{n\times n}$. 为了递归定义高阶行列式，我们先介绍下列定义.

**定义 2.5.** 设$n \geq 2$且$n-1$阶行列式已定义. 在$n$阶行列式$|a_{ij}|_{n\times n}$中划去元素$a_{ij}$所在的第$i$行与第$j$列的元，剩下的元素不改变原来的顺序所构成的$n-1$阶行列式称为$a_{ij}$的余子式，记为$M_{ij}$. 而$(-1)^{i+j}M_{ij}$称为$a_{ij}$的代数余子式，记为$C_{ij}$.

注意正负号，$-1$上的指数为行标加列标的值. 因此正负号如图中所示.




**例 2.27.** 求矩阵

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

的元素$a_{11}$，$a_{22}$，$a_{31}$与$a_{32}$的代数余子式.

解.（略.） □

**定义 2.6.** 设$n \geq 2$且$n-1$阶行列式一定义. 设$A = [a_{ij}] \in M_n(\mathbb{F})$. 设$i_0$为给定行标，而$j_0$为给定列标，则$A$的行列式定义为

$$\begin{aligned} \mathit{det}(A) &= \textstyle\sum_{j=1}^{n} (-1)^{i_0+j} a_{i_0 j} M_{i_0 j} = \sum_{j=1}^{n} a_{i_0 j} C_{i_0 j} \\ &= \textstyle\sum_{i=1}^{n} (-1)^{i+j_0} a_{i j_0} M_{i j_0} = \sum_{i=1}^{n} a_{i j_0} C_{i j_0}. \end{aligned}$$

等式右端分别被称为$A$按第$i_0$行或按$j_0$列的代数余子式展开.

值得一题的是，无论按哪一行亦或哪一列展开，最终的计算结果都是一样的. 这个事实可以利用数学归纳法证明，留给读者作为练习.

**例 2.28.** 使按第2行以及第3列展开，分别计算矩阵

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

的行列式.

解.（略.） □

**例 2.29.** 求下列行列式，并对比不同按行、列展开的计算复杂程度.

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

解.（略.） □

**例 2.30.** 求以下$2n$阶行列式

$$D_n = \begin{vmatrix} a_n & & & & & & & b_n \\ & a_{n-1} & & & & & b_{n-1} & \\ & & \ddots & & & \iddots & & \\ & & & a_1 & b_1 & & & \\ & & & c_1 & d_1 & & & \\ & & \iddots & & & \ddots & & \\ & c_{n-1} & & & & & d_{n-1} & \\ c_n & & & & & & & d_n \end{vmatrix}.$$

解.（略.） □

**例 2.31.** *求以下三角矩阵的行列式.*

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

解.（略.） □

从上述例子不难看出三角阵的行列式的计算公式.

**定理 2.32.** *三角矩阵的行列式等于所有主对角线元素的乘积.*

接下来，我们逐步推导行列式的一些性质，并提炼其它的计算方法.

**定理 2.33.** *若方阵$A$有一行或一列均为0，则$\mathit{det}(A) = 0$.*

证明. 这里只证“行”的情形. 不妨设$A = [a_{ij}]_{n\times n}$的第$i_0$行均为零元素. 按第$i_0$行展开，得

$$\mathsf{det}(A) = \sum_{j=1}^{n} 0 \cdot C_{i_0 j} = 0.$$

□

**定理 2.34.** *若方阵$A$有两行相同或两列相同，则$\mathit{det}(A) = 0$.*

证明. 不失一般性，这里只证第一行与第二行相同的情形. 我们对矩阵的阶进行数学归纳. 当$n = 2$时,

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ba = 0.$$

假设当$n = k\ (k \geq 1)$时定理成立. 现在考虑$n = k + 1$时得情形. 记$A = [a_{ij}]$，其第1行与第2行相同. 将该行列式按第3行展开，则有

$$\mathsf{det}(A) = \sum_{j=1}^{k+1} (-1)^{3+j} a_{3j} M_{3j}.$$

注意到每一个 $M_{3j}\ (1 \leq j \leq k+1)$ 都是$k$阶行列式，且第1行与第2行元素都是相同的，由归纳假设知$M_{3j} = 0$. 故 $\mathsf{det}(A) = 0$. 由数学归纳法，定理得证. □

**定理 2.35.** *对任意方阵$A$，总有*

$$\mathit{det}(A) = \mathit{det}(A^T).$$

证明. 我们对矩阵的阶进行数学归纳. 易知当$n = 1$时结论成立. 假设当$n = k\ (k \geq 1)$时定理成立. 现在考虑$n = k + 1$时得情形. 记$A = [a_{ij}]$，对应代数余子式为$C_{ij}$. 再记$A^T = [\tilde{a_{ij}}]$，对应代数余子式为$\tilde{C}_{ij}$. 根据转置的定义，不难得到

$$a_{ij} = \tilde{a}_{ji}, \quad C_{ij} = \tilde{C}_{ji}.$$

将$\mathsf{det}(A)$按第一行展开，则有

$$\mathsf{det}(A) = \sum_{j=1}^{k+1} a_{1j} C_{ij} = \sum_{j=1}^{k+1} \tilde{a}_{j1} \tilde{C}_{j1} = \mathsf{det}(A^T),$$

其中最后一个等式用到了$\mathsf{det}(A^T)$按第一列的展开. 由数学归纳法，定理得证. □

**定理 2.36.** 设$A \in M_n$.

*(i)* 设$\tilde{A}$是由$A$的第$i$行乘以$c$得到的矩阵（这里$c \in \mathbb{F}$），则$\mathit{det}(\tilde{A}) = c \cdot \mathit{det}(A)$.

*(ii)* 设$\tilde{A}$是由$A$交换第$i,j$行后得到的矩阵（这里$i \neq j$），则$\mathit{det}(\tilde{A}) = -\mathit{det}(A)$.

*(ii)* 设$\tilde{A}$是将$A$中的第$i$行加上第$j$行的$c$倍后得到的矩阵（这里$i \neq j$，而$c \in \mathbb{F}$），则$\mathit{det}(\tilde{A}) = \mathit{det}(A)$.

对列作相应的变换，也得到同样的结论.

*证明.* 设$A = [a_{ij}]$, 对应的余子式与代数余子式分别记为$M_{ij}$和$C_{ij}$. 设$\tilde{A} = [\tilde{a}_{ij}]$, 对应的余子式与代数余子式分别记为$\tilde{M}_{ij}$与$\tilde{C}_{ij}$.

(i) 注意到$\tilde{a}_{ij} = ca_{ij}$. 而$C_{ij} = \tilde{C}_{ij}$ $(1 \leq i,j \leq n)$. 将行列式按第$i$行展开，得

$$\det(\tilde{A}) = \sum_{i=1}^{n} \tilde{a}_{ij}\tilde{C}_{ij} = \sum_{i=1}^{n} ca_{ij}C_{ij} = c\det(A).$$

(ii) 使用数学归纳法，当$n = 2$时，

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = -(cb - da) = \begin{vmatrix} c & d \\ a & b \end{vmatrix}.$$

假设当$n = k$ $(k \geq 2)$时定理成立. 当$n = k+1$时，取异于$i,j$的某个行标$k$. 将行列式按第$k$行展开，得

$$\det(\tilde{A}) = \sum_{j=1}^{k+1} (-1)^{i+j} \tilde{a}_{kj}\tilde{M}_{kj}.$$

注意到$\tilde{a}_{kj} = a_{kj}$. 而$k$阶行列式$\tilde{M}_{kj}$是由$M_{ij}$交换两行后所得到的，由归纳法假设有$\tilde{M}_{kj} = -M_{kj}$. 故

$$\det(\tilde{A}) = \sum_{j=1}^{k+1} (-1)^{i+j} a_{kj}(-M_{kj}) = -\det(A).$$

(iii) 注意到$\tilde{a}_{ij} = ca_{ij}$. 而$C_{ij} = \tilde{C}_{ij}$ $(1 \leq i,j \leq n)$. 将行列式按第$i$行展开，得

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = -(cb - da) = \begin{vmatrix} c & d \\ a & b \end{vmatrix}.$$

假设当$n = k$ $(k \geq 2)$时定理成立. 当$n = k+1$时，取异于$i,j$的某个行标$k$. 将行列式按第$k$行展开，得

$$\det(\tilde{A}) = \sum_{j=1}^{k+1} (-1)^{i+j} \tilde{a}_{kj}\tilde{M}_{kj}.$$

注意到$\tilde{a}_{kj} = a_{kj}$. 而$k$阶行列式$\tilde{M}_{kj}$是由$M_{ij}$交换两行后所得到的，由归纳法假设有$\tilde{M}_{kj} = -M_{kj}$. 故

$$\det(\tilde{A}) = \sum_{j=1}^{k+1} (-1)^{i+j} a_{kj}(-M_{kj}) = -\det(A).$$

□

**推论 2.37.** 对于$n$阶初等矩阵，有

$$\mathit{det}(F_i(c)) = c \ (c \neq 0), \quad \mathit{det}(F_{i,j}) = -1, \quad \mathit{det}(F_{i,j}(c)) = 1.$$

并且对任意矩阵$A \in M_n$，

$$\mathit{det}(F_i(c)A) = \mathit{det}(F_i(c))\mathit{det}(A) \ (c \neq 0),$$
$$\mathit{det}(F_{i,j}A) = \mathit{det}(F_{i,j})\mathit{det}(A),$$
$$\mathit{det}(F_{i,j}(c)A) = \mathit{det}(F_{i,j}(c))\mathit{det}(A).$$

$$\begin{vmatrix}1&0&0&0\\0&3&0&0\\0&0&1&0\\0&0&0&1\end{vmatrix}=3,\quad \begin{vmatrix}0&0&0&1\\0&1&0&0\\0&0&1&0\\1&0&0&0\end{vmatrix}=-1,\quad \begin{vmatrix}1&0&0&7\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{vmatrix}=1$$

**例 2.38.** 以下为初等矩阵的行列式：

利用上述推论，我们可以通过初等行、列变换来求方阵的行列式. 特别地，简约阶梯阵是一个上三角矩阵，其行列式为主对角元素之积.

**例 2.39.** 求行列式

$$\begin{vmatrix}0&1&5\\3&-6&9\\2&6&1\end{vmatrix},\quad \begin{vmatrix}1&0&0&3\\2&7&0&6\\0&6&3&0\\7&3&1&-2\end{vmatrix}.$$

解.（略.） □

**例 2.40.** 设$x_1,x_2,\dots,x_n\in\mathbb{F}$. 计算下列范德蒙行列式

$$\Delta_n=\left|x_j^{i-1}\right|_{n\times n}=\begin{vmatrix}1&1&\dots&1&1\\x_1&x_2&\dots&x_{n-1}&x_n\\x_1^2&x_2^2&\dots&x_{n-1}^2&x_n^2\\\vdots&\vdots&\ddots&\vdots&\vdots\\x_1^{n-1}&x_2^{n-1}&\cdots&x_{n-1}^{n-1}&x_n^{n-1}\end{vmatrix}.$$

解. 为了得到关于$\Delta_n$的递推公式，我们需要在消去最后一列中有关$x_n$的项的同时，使得前$n-1$列中出现形如$x_j^{i-2}$的项. 考虑依次将第$i$行减去第$i-1$行的$x_n$倍($i=n,n-1,\dots,3,2$)，得

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

将第$j$列$(1\le j\le n-1)$的因子$x_j-x_n$均提出，得

$$\Delta_n=\prod_{j=1}^{n}(x_j-x_n)\cdot\begin{vmatrix}1&1&\dots&1&1\\1&1&\dots&1&0\\x_1&x_2&\dots&x_{n-1}&0\\\vdots&\vdots&\ddots&\vdots&\vdots\\x_1^{n-2}&x_2^{n-2}&\cdots&x_{n-1}^{n-2}&0\end{vmatrix}.$$

注意到上式右端的行列式的左下角部分已经出现了$n-1$阶范德蒙行列式. 按最后一列展开，得到$(-1)^{1+n}\cdot 1\cdot\Delta_n$. 由于$(-1)^{1+n}=(-1)^{n-1}$，我们将这$n-1$个负号分别分配到连乘积的每一项中，于是有

$$\Delta_n=\prod_{j=1}^{n}(x_n-x_j)\Delta_{n-1}.$$

利用上述公式递推，结合$\Delta_1=1$. 可得

$$\Delta_n=\prod_{j=1}^{n}(x_n-x_j)\cdot\prod_{j=1}^{n-1}(x_{n-1}-x_j)\cdot\ldots\prod_{j=1}^{1}(x_2-x_j)\cdot\Delta_1=\prod_{1\le j<i\le n}(x_i-x_j).$$

特别地，当$x_1,x_2,\dots,x_n$均不同时，范德蒙行列式非零；而$x_1,x_2,\dots,x_n$中出现相同的数时，范德蒙行列式为零. □

思考题：试将下列行列式写成$\cos\alpha$, $\cos\beta$, $\cos\gamma$的多项式.

$$\begin{vmatrix} 1 & \cos\alpha & \cos 2\alpha \\ 1 & \cos\beta & \cos 2\beta \\ 1 & \cos\gamma & \cos 2\gamma. \end{vmatrix}.$$

思考题：试求下列$n$阶行列式.

$$A_n = \begin{vmatrix} a_1 & 1 & 1 & \dots & 1 \\ 1 & a_2 & & & \\ 1 & & a_3 & & \\ \vdots & & & \ddots & \\ 1 & & & & a_n \end{vmatrix}, \quad B_n = \begin{vmatrix} a_1 & b & b & \dots & b \\ c & a_2 & b & \dots & b \\ c & c & a_3 & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \dots & a_n \end{vmatrix}.$$

$$C_n = \begin{vmatrix} a & b & & & \\ c & a & b & & \\ & c & \ddots & \ddots & \\ & & \ddots & \ddots & b \\ & & & c & a \end{vmatrix}, \quad D_n = \begin{vmatrix} 1+a_1 & a_2 & \dots & a_n \\ a_1 & 1+a_2 & \dots & a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & 1+a_n \end{vmatrix}$$

## 2.5 行列式的性质与克莱默法则

方阵可以作加法、数乘或乘法，那么运算后的矩阵的行列式与原矩阵的行列式有没有什么联系呢？对于数乘的情形，设$A\in M_n(\mathbb{F})$，而$c\in\mathbb{F}$. 这时$cA$的每一行都是$A$对应行的$c$倍. 提出每一行的$c$，得

$$\mathsf{det}(cA) = c^n\mathsf{det}(A).$$

下面，我们重点考察矩阵乘法与行列式的关系.

**定理 2.41.** 设$A,B\in M_n$，则

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

证明. 在推论2.37中，我们已经得到了$A$为初等矩阵的情形. 进一步，我们能证明$A$为初等矩阵乘积的情况.

情形1. 设$A$可逆. 则$A$可写为初等矩阵的乘积，记为$A=F_kF_{k-1}\dots F_1$. 反复利用推论2.37，可得

$$\begin{aligned} \mathsf{det}(AB) &= \mathsf{det}\big(F_k(F_{k-1}\dots F_1B)\big) = \mathsf{det}(F_k)\mathsf{det}(F_{k-1}\dots F_1B) \\ &= \dots = \mathsf{det}(F_k)\mathsf{det}(F_{k-1})\dots\big(\mathsf{det}(F_2)\mathsf{det}(F_1)\big)\mathsf{det}(B) \\ &= \mathsf{det}(F_k)\dots\mathsf{det}(F_3)\mathsf{det}(F_2F_1)\mathsf{det}(B) = \dots \\ &= \mathsf{det}(F_k\dots F_2F_1)\mathsf{det}(B) = \mathsf{det}(A)\mathsf{det}(B). \end{aligned}$$

情形2. 设$C$为任意不可逆方阵，则$C$的简约阶梯阵$R$中含有零行. 由定理2.32，知$\mathsf{det}(R)=0$. 将$C$化简约阶梯阵的过程写成左乘初等矩阵的形式，记为$E_r\dots E_2E_1C=R$. 则$C=(E_1^{-1}E_2^{-1}\dots E_r^{-1})R$. 注意到$E_1^{-1}E_2^{-1}\dots E_r^{-1}$是可逆矩阵，由情形1的结论得

$$\mathsf{det}(C) = \mathsf{det}(E_1^{-1}E_2^{-1}\dots E_r^{-1})\mathsf{det}(R) = 0.$$

当矩阵$A$不可逆时，由定理1.60知，矩阵$AB$也是不可逆的，因此

$$\mathsf{det}(AB) = 0 = 0 \cdot \mathsf{det}(B) = \mathsf{det}(A)\mathsf{det}(B).$$

定理证毕. □

**推论 2.42.** 方阵$A$可逆当且仅当$\mathit{det}(A) \neq 0$. 当方阵可逆时，我们有

$$\mathit{det}(A^{-1}) = \frac{1}{\mathit{det}(A)}.$$

关于方阵$A$是否可逆，现在我们有了新的判定方法：$\mathsf{det}(A) \neq 0$. 在第一章，我们介绍了求逆的算法. 接下来，我们希望找一个可逆矩阵的逆的精确表达式. 下面以$3 \times 3$矩阵为例.

**例 2.43.** 设$A = [a_{ij}] \in M_3$，试证

$$\begin{aligned} a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} &= \mathit{det}(A); \\ a_{11}C_{31} + a_{12}C_{22} + a_{13}C_{23} &= 0. \\ a_{11}C_{31} + a_{12}C_{32} + a_{13}C_{23} &= 0. \end{aligned}$$

证明. 第一个式子即$\mathsf{det}(A)$按第一行展开所得的表达式. 观察第二个式子的左端，它是将$\mathsf{det}(A)$按照第二行展开后，将表达式中的$a_{2j}$对应替换成了$a_{1j}$ $(1 \leq i \leq 3)$ 所得，因此有

$$a_{11}C_{31} + a_{12}C_{22} + a_{13}C_{23} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0.$$

第三个式子的证明也是类似的. □

在上述例子中，我们得到

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = \mathsf{det}(A) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

为了回到矩阵$A$的形式，我们对等式两端的矩阵同时转置，有

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \mathsf{det}(A) \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}.$$

对$A$的第2行、第3行作类似的分析，可以进一步得到

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \hdashline a_{21} & a_{22} & a_{23} \\ \hdashline a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \mathsf{det}(A) \begin{bmatrix} 1 & 0 & 0 \\ \hdashline 0 & 1 & 0 \\ \hdashline 0 & 0 & 1 \end{bmatrix}.$$

当$A$可逆时，$\mathsf{det}(A) \neq 0$，上式便给出了$A$的逆矩阵的一种表达式.

**定义 2.7.** 设$A \in M_n$，记$C_{ij}$为元素$a_{ij}$的代数余子式。我们称

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

为$A$的伴随矩阵，记为$\mathit{adj}(A)$. 即

$$\mathit{adj}(A) = [C_{ij}]^T.$$

**例 2.44.** 求$adj(A)$，其中

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 1 & 2 \\ 2 & 3 & 4 \end{bmatrix}.$$

解.（略.） □

将例2.43及之后的推导推广到$n$阶方阵的情形，再把行、列的角色互换，就得到以下定理.

**定理 2.45.** 设$A \in M_n$. 则有

$$A\,adj(A) = adj(A)\,A = det(A)\,I_n. \tag{2.5}$$

特别地，当$A$可逆时，我们有$\det(A) \neq 0$，从而

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A).$$

**例 2.46.** 设$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$是可逆矩阵，试用伴随矩阵求$A^{-1}$.

解. 因为

$$C_{11} = d, \quad C_{12} = -c, \quad C_{21} = -b, \quad C_{22} = a,$$

所以

$$\mathrm{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

因此当$\det(A) = ad - bc \neq 0$时，

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A) = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

□

**例 2.47.** 试用伴随矩阵求$A^{-1}$，其中

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 1 & 2 \\ 2 & 3 & 4 \end{bmatrix}.$$

解.（略.） □

思考题: 设$A = \begin{bmatrix} 2 & 1 & -1 \\ -1 & 2 & 1 \\ 1 & -1 & 2 \end{bmatrix}$. 若$3\mathrm{adj}(A)B = 7A^{-1} + 14B$，求$B$.

如果知道了一个矩阵$A$的伴随矩阵$\mathrm{adj}(A)$，是否能反过来求出$A$呢？当$A$可逆时，矩阵$\mathrm{adj}(A)$也可逆，且

$$A = \det(A)(\mathrm{adj}(A))^{-1}.$$

因此我们需要在知道$\mathrm{adj}(A)$的情况下，求得$\det(A)$. 事实上，在式(2.5)两端同时求矩阵的行列式，有

$$\det(A)\det(\mathrm{adj}(A)) = \det\big(\det(A)I_n\big) = (\det(A))^n\det(I_n) = (\det(A))^n.$$

因此当$\det(A) \neq 0$时，有

$$\det(\mathrm{adj}(A)) = (\det(A))^n.$$

事实上，当$\det(A) = 0$时，上式同样成立，我们将这部分的证明留到后续的章节.

**例 2.48.** 设$A \in M_4$，满足

$$\mathit{adj}(A) = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 4 & 0 & 3 & 2 \\ -2 & 0 & -1 & 2 \end{bmatrix}.$$

求矩阵$A$.

解.（略.） □

回顾当$ad - bc \neq 0$时，线性方程组

$$\begin{cases} ax + by & = u \\ cx + dy & = v. \end{cases}$$

的解可写为

$$x = \frac{\begin{vmatrix} u & b \\ v & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}, \qquad y = \frac{\begin{vmatrix} a & u \\ c & v \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}.$$

对于更高阶的情形，我们也有类似的结论.

**定理 2.49** (克莱默法则)**.** 设$A \in M_n$为可逆矩阵，而$\mathbf{b} = [b_1 \ldots, b_n]^T$. 则线性方程组$A\boldsymbol{x} = \boldsymbol{b}$的唯一解$\mathbf{x} = [x_1 \ldots x_n]^T$的表达式为

$$x_i = \frac{\mathit{det}(A_i)}{\mathit{det}(A)},$$

其中$A_i$是将$A$中的第$i$列更换为$\mathbf{b}$后所得的矩阵 $(1 \leq i \leq n)$.

证明. 由等价性定理可知方程组有唯一解

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\mathsf{det}(A)}\mathsf{adj}(A)\mathbf{b}.$$

因此

$$\mathsf{det}(A)\, x_i = \big(\mathsf{adj}(A)\mathbf{b}\big)_{i1} = \sum_{j=1}^{n} C_{ji} b_i = \mathsf{det}(A_i), \quad (1 \leq i \leq n).$$

定理得证. □

**例 2.50.** 设$(x_1, x_2, x_3)$为下列方程组的解：

$$\begin{aligned} x_1 \qquad\quad + 2x_3 &= 6 \\ -3x_1 + 4x_2 + 6x_3 &= 30 \\ -x_1 - 2x_2 + 3x_3 &= 8 \end{aligned}$$

*span*

求$x_3$.

解.（略） □

下面重新列出等价性定理.

**定理 2.51.** 设$A \in M_n$，则以下命题等价.

*(i)* 矩阵$A$可逆.

*(ii)* 齐次线性方程组$A\mathbf{x} = \mathbf{0}$仅有平凡解.

*(iii)* 矩阵$A$的简约阶梯型为$I_n$.

*(iv)* 矩阵$A$可写成初等矩阵的成绩.

*(v)* 对任意$\mathbf{b} \in M_{n\times 1}(\mathbb{F})$，线性方程组$A\mathbf{x} = \mathbf{b}$是相容的.

*(vi)* 对任意$\mathbf{b} \in M_{n\times 1}(\mathbb{F})$，线性方程组$A\mathbf{x} = \mathbf{b}$有唯一解.

*(vii)* 行列式$\mathit{det}(A) \neq 0$.

下面我们来看一个等价性定理的应用.

**例 2.52.** 设$(a_1, b_1), \ldots, (a_n, b_n)$为$xOy$平面中的$n$个点，其中$a_1, a_2, \ldots, a_n$两两不同. 则存在唯一的次数不超过$n-1$的多项式$p(x)$，使得

$$p(a_i) = b_i, \quad (1 \leq i \leq n).$$

证明. 不妨设

$$p(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1},$$

其中的系数$c_0, c_1, \ldots, c_{n-1}$看成未知数. 代入条件$p(a_i) = b_i$ $(1 \leq i \leq n)$，得

$$\begin{aligned} c_0 + c_1 a_1 + \ldots + c_{n-1} a_1^{n-1} &= b_1, \\ c_0 + c_1 a_2 + \ldots + c_{n-1} a_2^{n-1} &= b_2, \\ \ldots\ldots\ldots\ldots\ldots \\ c_0 + c_1 a_n + \ldots + c_{n-1} a_n^{n-1} &= b_n. \end{aligned}$$

span

即$A\mathbf{x} = \mathbf{b}$，其中

$$A = \begin{bmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ 1 & a_2 & \cdots & a_2^{n-1} \\ \vdots & \vdots \ddots & \vdots & \\ 1 & a_n & \cdots & a_n^{n-1} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

系数矩阵为范德蒙行列式，因$a_1, a_2, \ldots, a_n$两两不同，因此

$$\mathsf{det}(A) = \prod_{1 \leq j < i \leq n} (a_i - a_j) \neq 0.$$

由等价性定理，知方程组有唯一解，从而得到唯一的次数不超过$n-1$的多项式$p(x)$. □

当$b_1, b_2, \ldots, b_n$均为零时,上述其次方程组的解为平凡解,此时的多项式为零多项式. 而当$b_1, b_2, \ldots, b_n$不全为零时，该多项式是非零的. 利用克莱默法则，事实上可以写出多项式$p(x)$的系数$c_0, c_1, \ldots, c_{n-1}$的具体表达式. 另外，我们也可以用拉格朗日插值法来凑出所要的多项式. 先考虑简单情形：后$n-1$个纵坐标为零，即$b_2 = b_3 = \ldots = b_n = 0$. 这时$a_2, a_3, \ldots, a_n$为多项式$p(x)$的$n-1$个不同零点，所以形如

$$p(x) = c(x - a_2)(x - a_3) \ldots (x - a_n),$$

其中$c \in \mathbb{F}$. 这时代入条件$p(a_1) = b_1$，有

$$b_1 = p(a_1) = c(a_1 - a_2)(a_1 - a_3) \ldots (a_1 - a_n) \quad \Rightarrow \quad c = \frac{b_1}{(a_1 - a_2)(a_1 - a_3) \ldots (a_1 - a_n)}.$$

于是

$$p(x) = b_1 \frac{(x - a_2)(x - a_3) \ldots (x - a_n)}{(a_1 - a_2)(a_1 - a_3) \ldots (a_1 - a_n)}.$$

对于一般的情形，我们可以拆成$n$个简单情形进行讨论. 最后，可以验证多项式

$$p(x) = \sum_{1 \leq i \leq n} b_1 \prod_{\substack{1 \leq j \neq n \\ j \neq i}} \frac{x - a_j}{a_i - a_j}$$

满足$p(a_i) = b_i$ $(1 \leq i \leq n)$.

最后，我们来探索行列式与矩阵加法之间的关系.

**定理 2.53** (行列式的多重线性性). 将行列式的第$k$列拆成两列之和，则有

$$\begin{vmatrix} a_{11} & \ldots & \hat{a}_{1k}+\tilde{a}_{1k} & \cdots & a_{1n} \\ a_{21} & \ldots & \hat{a}_{1k}+\tilde{a}_{1k} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & \hat{a}_{nk}+\tilde{a}_{nk} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \ldots & \hat{a}_{1k} & \cdots & a_{1n} \\ a_{21} & \ldots & \hat{a}_{1k} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & \hat{a}_{nk} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \ldots & \tilde{a}_{1k} & \cdots & a_{1n} \\ a_{21} & \ldots & \tilde{a}_{1k} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & \tilde{a}_{nk} & \cdots & a_{nn} \end{vmatrix}.$$

将“列”更换为“行”，类似的结论也成立.

证明. 将三个行列式都按第$k$列展开，不难得到所求结论. 在这里，我们利用克莱默法则给出另一个证明，并揭示行列式的多重线性性与方程组之间的关联.

情形1：若将上述三个行列式中的第$k$列换为任意列$[a_{1k}\, a_{2k}\, \ldots\, a_{nk}]^T$，行列式均为0，则等式左右均为0，定理得证.

情形2：若存在某列$[a_{1k}\, a_{2k}\, \ldots\, a_{nk}]^T$，使得矩阵$A = [a_{ij}]$满足$\mathsf{det}(A) \neq 0$，我们考虑以下三个线性方程组：

$$Ax = \hat{\mathbf{b}} + \tilde{\mathbf{b}}, \quad Ax = \hat{\mathbf{b}}, \quad Ax = \tilde{\mathbf{b}},$$

其中$\hat{\mathbf{b}} = [\hat{b_1}\, \hat{b_2}\, \ldots\, \hat{b_n}]^T$, $\tilde{\mathbf{b}} = [\tilde{b_1}\, \tilde{b_2}\, \ldots\, \tilde{b_n}]^T$. 设

$$\hat{\mathbf{x}} = [\hat{x_1}\, \hat{x_2}\, \ldots\, \hat{x_n}]^T, \quad \tilde{\mathbf{x}} = [\tilde{x_1}\, \tilde{x_2}\, \ldots\, \tilde{x_n}]^T$$

分别为第二、第三个方程组的唯一解，则

$$A(\hat{\mathbf{x}} + \tilde{\mathbf{x}}) = A\hat{\mathbf{x}} + A\tilde{\mathbf{x}} = \hat{\mathbf{b}} + \tilde{\mathbf{b}}.$$

即$\hat{\mathbf{x}} + \tilde{\mathbf{x}}$是第一个方程组的唯一解. 根据克莱默法则，解的第$k$个分量满足公式

$$\hat{x_k} + \tilde{x_k} = \frac{1}{\mathsf{det}(A)} \begin{vmatrix} a_{11} & \ldots & \hat{a}_{1k}+\tilde{a}_{1k} & \cdots & a_{1n} \\ a_{21} & \ldots & \hat{a}_{1k}+\tilde{a}_{1k} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & \hat{a}_{nk}+\tilde{a}_{nk} & \cdots & a_{nn} \end{vmatrix},$$

$$\hat{x_k} = \frac{1}{\mathsf{det}(A)} \begin{vmatrix} a_{11} & \ldots & \hat{a}_{1k} & \cdots & a_{1n} \\ a_{21} & \ldots & \hat{a}_{1k} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & \hat{a}_{nk} & \cdots & a_{nn} \end{vmatrix}, \quad \tilde{x_k} = \frac{1}{\mathsf{det}(A)} \begin{vmatrix} a_{11} & \ldots & \tilde{a}_{1k} & \cdots & a_{1n} \\ a_{21} & \ldots & \tilde{a}_{1k} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & \tilde{a}_{nk} & \cdots & a_{nn} \end{vmatrix}.$$

定理得证. □

在这一节的最后，我们用以下例题做一个全面的复习.

**例 2.54.** 设$n \geq 2$. 计算$n$阶行列式$\mathit{det}(A_n)$，其中

$$A_n = \begin{bmatrix} a & b & b & \ldots & b \\ c & a & b & \ldots & b \\ c & c & a & \ldots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \ldots & a \end{bmatrix}.$$

解. 记$D_n=\det(A_n)$. 我们先考虑$b=c$的情形，这时行列式的特点是每行的元素之和都是$s=a+(n-1)b$，因此可以用初等行变换将第2行到第$n$行的信息都加到第1行上来：

$$D_n=\begin{vmatrix} s & s & s & \dots & s \\ b & a & b & \dots & b \\ b & b & a & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & a \end{vmatrix}=s\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ b & a & b & \dots & b \\ b & b & a & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & a \end{vmatrix}.$$

这时，我们可以用第1列消去第2列到第$n$列中的1，进而利用下三角矩阵进行行列式计算:

$$D_n=s\begin{vmatrix} 1 & & & & \\ b & a-b & & & \\ b & b & a-b & & \\ \vdots & \vdots & \vdots & \ddots & \\ b & b & b & \dots & a-b \end{vmatrix}=s(a-b)^{n-1}.$$

（注：也可以用第1行的$b$倍消去左下角的$b$，从而得到上三角矩阵的行列式.）综上，当$b=c$时，

$$D_n=\big(a+(n-1)b\big)(a-b)^{n-1}.$$

再考虑$b\neq c$的情形，由之前的经验，我们将矩阵$A_n$的第一行拆成两行之和：

$$\begin{bmatrix} a & b & b & \dots & b \end{bmatrix}=\begin{bmatrix} a-b & 0 & 0 & \dots & 0 \end{bmatrix}+\begin{bmatrix} b & b & b & \dots & b \end{bmatrix}.$$

这时$D_n=D_n'+D_n''$，其中

$$D_n'=\begin{vmatrix} a-b & 0 & 0 & \dots & 0 \\ c & a & b & \dots & b \\ c & c & a & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \dots & a \end{vmatrix},\quad D_n''=\begin{vmatrix} b & b & b & \dots & b \\ c & a & b & \dots & b \\ c & c & a & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \dots & a \end{vmatrix}.$$

将$D_n'$按第1行展开，易得$D_n'=(a-b)D_{n-1}$. 提出$D_n''$第一行中的$b$后，将第2行到第$n$行都减去第1行的$c$倍，得

$$D_n''=b\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ c & a & b & \dots & b \\ c & c & a & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \dots & a \end{vmatrix}=\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ & a-c & b-c & \dots & b-c \\ & & a-c & \dots & b-c \\ & & & \ddots & \vdots \\ & & & & a-c \end{vmatrix}=b(a-c)^{n-1}.$$

于是

$$D_n=(a-b)D_{n-1}+b(a-c)^{n-1}. \tag{2.6}$$

这时我们已经得到了关于$D_n$的递推公式. 另外，观察到$b$与$c$的地位是对称的，

$$D_n=\det(A_n)=\det(A_n^T)=\begin{vmatrix} a & c & c & \dots & c \\ b & a & c & \dots & c \\ b & b & a & \dots & c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & a \end{vmatrix}$$

因此
$$D_n=(a-c)D_{n-1}+c(a-b)^{n-1}. \tag{2.7}$$
联立式(2.6)、(2.7)，得
$$\begin{bmatrix} 1 & b-a \\ 1 & c-a \end{bmatrix}\begin{bmatrix} D_n \\ D_{n-1} \end{bmatrix}=\begin{bmatrix} b(a-c)^{n-1} \\ c(a-b)^{n-1} \end{bmatrix}.$$
这时系数矩阵的行列式为$c-b\neq 0$，应用克莱默法则知
$$D_n=\frac{\begin{vmatrix} b(a-c)^{n-1} & b-a \\ c(a-b)^{n-1} & c-a \end{vmatrix}}{\begin{vmatrix} 1 & b-a \\ 1 & c-a \end{vmatrix}}=\frac{c(a-b)^n-b(a-c)^n}{c-b}.$$

□

## 2.6 矩阵变换

我们将一个向量空间到另一个向量空间的映射称为变换. 线性代数理论的核心问题之一是描述线性变换. 在本书的开头，我们已经举了平面上绕原点逆时针旋转$\theta$弧度的例子. 用$R_\theta$表示这个映射，那么有
$$R_\theta:\mathbb{R}^2\to\mathbb{R}^2,\,(x,y)\mapsto(u,v),$$
其中
$$\begin{cases} u\ =\ (\cos\theta)x+(\sin\theta)y, \\ v\ =\ (-\sin\theta)x+(\cos\theta)y. \end{cases}$$
将其写成矩阵运算的形式
$$\begin{bmatrix} u \\ v \end{bmatrix}=\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$
也就是说，如果我们将旋转前后的点的坐标用列矩阵表示，映射$R_\theta$的效果相当于在在坐标列矩阵的左侧乘上了方阵
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

**定义 2.8.** 设$A\in M_{m\times n}(\mathbb{R})$. 我们称
$$T_A:\mathbb{R}^n\to\mathbb{R}^m,\,\mathbf{x}\mapsto A\mathbf{x}.$$
为矩阵变换，称$A$为$T_A$的标准矩阵. 这里，欧氏空间$\mathbb{R}^n$与$\mathbb{R}^m$中的向量均用列矩阵表示.

当$T$为一个矩阵变换时，我们常用$[T]$表示它的标准矩阵，即
$$T(\mathbf{x})=[T]\mathbf{x}.$$

**例 2.55.** 设$T:\mathbb{R}^3\to\mathbb{R}^2$如下定义：
$$T(x_1,x_2,x_3)=(2x_1-3x_2+x_3,\,4x_1+x_2-2x_3).$$
求其标准矩阵$[T]$，并用该矩阵计算$T(1,-3,0)$.

解.（略.） □

设$T:\mathbb{R}^n\to\mathbb{R}^m$为矩阵变换，则$[T]\in M_{m\times n}$. 将其写成列矩阵的分块形式$[T]=[\mathbf{c}_1\,\mathbf{c}_2\,\ldots\,\mathbf{c}_n]$. 记$\mathbf{e}_j$为第$i$行为1、其余为0的列矩阵，由分块矩阵的运算性质得

$$[T]\mathbf{e}_j=\mathbf{c}_j,\quad (1\le j\le n).$$

即

$$[T]=\left[\begin{array}{c:c:c:c} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n)\end{array}\right].$$

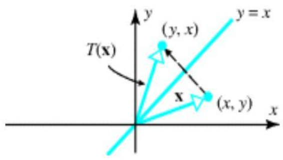
因此对一个由几何直观给出的矩阵映射，可以通过它们在标准向量上的像来计算标准矩阵.

在以下例子中，我们试求相关矩阵映射的标准矩阵，探索单位正方形的形状变化，并观察其面积变化与标准矩阵的行列式之间的关系.

**例 2.56.** *(1)* $\mathbb{R}^2$*中关于*$y$*轴的镜像对称：*$T(x,y)=(-x,y)$.

*(2)* $\mathbb{R}^2$*中关于直线*$y=x$*的镜像对称：*$T(x,y)=(y,x)$.

*(3)* $\mathbb{R}^3$*中关于*$xOy$*平面的镜像对称:* $T(x,y,z)=(x,y,-z)$.



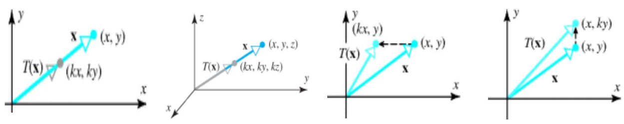

解.（略.） □

**例 2.57.** *(1)* $\mathbb{R}^2$*中以*$k$ $(0<k<1)$*为系数的收缩:* $T(x,y)=(kx,ky)$.

*(2)* $\mathbb{R}^3$*中以*$k$ $(k\ge 1)$*为系数的膨胀:* $T(x,y,z)=(kx,ky,kz)$.

*(3)* $\mathbb{R}^2$*中以*$k$ $(0<k<1)$*为系数关于*$x$*方向的压缩:* $T(x,y)=(kx,y)$.

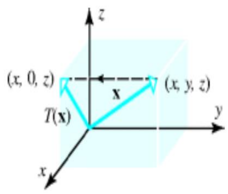
*(4)*$\mathbb{R}^2$*中以*$k$ $(k\ge 1)$*为系数关于*$y$*方向的拉伸:* $T(x,y)=(x,ky)$.




解.（略.） □

**例 2.58.** *(1)* $\mathbb{R}^2$*中向*$y$*轴的正交投影：*$T(x,y)=(0,y)$.

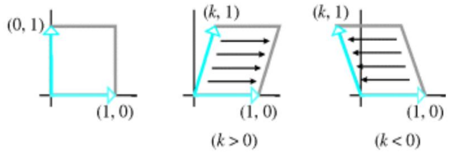
*(2)* $\mathbb{R}^3$*中向*$xOz$*平面的正交投影：*$T(x,y,z)=(x,0,z)$.

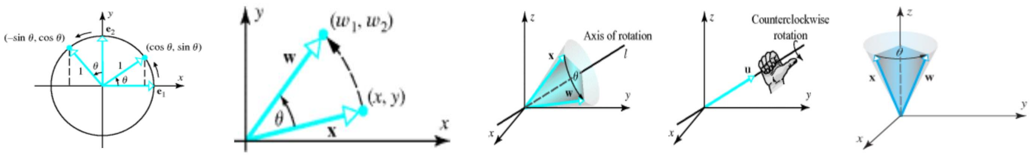



解.（略.） □

**例 2.59.** $\mathbb{R}^2$*中以*$k$*为系数关于*$x$*方向的错切：*$T(x,y)=(x+ky,y)$.

解.（略.） □




**例 2.60.** *(1)* $\mathbb{R}^2$*中绕着原点进行的*$\theta$*弧度的旋转*$R_\theta$.

*(2)* $\mathbb{R}^3$*中以*$z$*轴正方向为旋转轴方向（即右手螺旋法则的大拇指方向）、以*$\theta$*弧度进行的旋转*$R$.

解.（略.） □

**定理 2.61.** 设$A \in M_{m\times n}(\mathbb{R})$. 则矩阵变换$T_A : \mathbb{R}^n \to \mathbb{R}^m$满足

$$T_A(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda T_A(\mathbf{u}) + \mu T_A(\mathbf{v}), \quad (\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n, \forall\lambda, \mu \in \mathbb{R}).$$

证明. 将向量都写成列矩阵. 根据矩阵乘法，我们有

$$T_A(\lambda\mathbf{u} + \mu\mathbf{v}) = A(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda A\mathbf{u} + \mu A\mathbf{v} = \lambda T_A(\mathbf{u}) + \mu T_A(\mathbf{v}).$$

□

**定理 2.62.** 设$T_A : \mathbb{R}^2 \to \mathbb{R}^2$为由$A \in M_2$给出的矩阵变换. 假设$T_A$将一个有向面积为$R$的平行四边形区域映为另一个有向面积为$R'$的平行四边形区域，则有$R' = \mathit{det}(A)R$.

证明. 设平行四边形区域由向量$\mathbf{u}, \mathbf{v}$张成，且从$\mathbf{u}$的方向逆时针转动到$\mathbf{v}$的方向所经过的角度小于$\pi$. 则

$$R = \begin{vmatrix}\mathbf{u} & \mathbf{v}\end{vmatrix}.$$

这时，我们有

$$T_A(\mathbf{u}) = A\mathbf{u}, \quad T_A(\mathbf{v}) = A\mathbf{v}.$$

从而

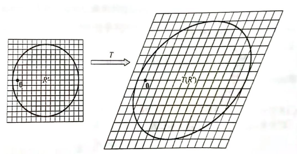
$$R' = \begin{vmatrix}A\mathbf{u} & A\mathbf{v}\end{vmatrix}.$$

由分块矩阵的乘法性质，知

$$A\begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix} = \begin{bmatrix}A\mathbf{u} & A\mathbf{v}\end{bmatrix}.$$

再由行列式保持乘法的性质，得$R' = \det(A)R$. □

当面积变为0时，意味着平行四边形退化成了一条线段或一个点. 对于一般的区域，利用积分的方法也可以得到相似的结论.



**定义 2.9.** 设$U, V, W$为集合，而$f, g$是如下的映射：

$$U \xrightarrow{f} V \xrightarrow{g} W.$$

则$g$与$f$的复合定义为$g \circ f : U \to W$，

$$(g \circ f)(\mathbf{u}) = g(f(\mathbf{u})), \quad (\forall \mathbf{u} \in U).$$

设$T_A : \mathbb{R}^n \to \mathbb{R}^k$, 而$T_B : \mathbb{R}^k \to \mathbb{R}^m$分别为左乘矩阵$A, B$的矩阵变换. 则

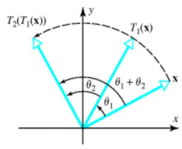
$$T_B \circ T_A = T_{BA}.$$

更进一步，映射的复合具有结合律，当矩阵变换$T_1, T_2, \ldots, T_r$可以依次符合时，

$$[T_r \circ \ldots \circ T_2 \circ T_1] = [T_r] \ldots [T_2][T_1].$$

.

**例 2.63.** 试求旋转变换的复合$R_{\theta_1} \circ R_{\theta_2}$与$R_{\theta_2} \circ R_{\theta_1}$.




解. 我们有

$$\begin{aligned}
[R_{\theta_1} \circ R_{\theta_2}] = [R_{\theta_1}][R_{\theta_2}] &= \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_1 - \sin\theta_1\cos\theta_2 \\ \sin\theta_1\cos\theta_2 + \cos\theta_2\sin\theta_1 & -\sin\theta_1\sin\theta_2 + \cos\theta_1\cos\theta_2 \end{bmatrix} \\
&= \begin{bmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{bmatrix} = [R_{\theta_1+\theta_2}].
\end{aligned}$$

上式最右端中，我们有$\theta_1 + \theta_2 = \theta_2 + \theta_1$. 因此也有$[R_{\theta_2} \circ R_{\theta_1}] = [R_{\theta_1+\theta_2}]$. 事实上，从旋转的效果来看，无论是先转$\theta_2$再转$\theta_1$，还是先转$\theta_1$再转$\theta_2$，最终的效果都是旋转$\theta_1 + \theta_2$. □

**例 2.64.** 在$\mathbb{R}^2$中，设$T_1$为关于直线$y = x$的镜像对称，而$T_2$为向$y$轴的正交投影. 试求$T_1 \circ T_2$与$T_2 \circ T_1$的标准矩阵.

解. 我们用两种不同的方法来分别计算. 首先考虑$T_1 \circ T_2$. 由于

$$T_2(1,0) = (0,0), \quad T_1(0,0) = (0,0),$$
$$T_2(0,1) = (0,1), \quad T_1(0,1) = (1,0).$$

我们有

$$[T] = \left[\begin{array}{c:c} 0 & 1 \\ 0 & 0 \end{array}\right].$$

另一方面，由于

$$[T_2] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad [T_1] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

因而

$$[T_2 \circ T_1] = [T_2][T_1] = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

□

在上个例子中，我们发现$T_1 \circ T_2 \neq T_2 \circ T_1$，矩阵变换的复合不具备交换律.

另外，设$T, S: \mathbb{R}^n \to \mathbb{R}^m$均为矩阵变换，而$c \in \mathbb{F}$. 则可定义$T+S, cT: \mathbb{R}^n \to \mathbb{R}^m$如下：对任意$\mathbf{x} \in \mathbb{R}^n$，

$$(T+S)(\mathbf{x}) = T(\mathbf{x}) + S(\mathbf{x}), \quad (cT)(\mathbf{x}) = c\,T(\mathbf{x}).$$

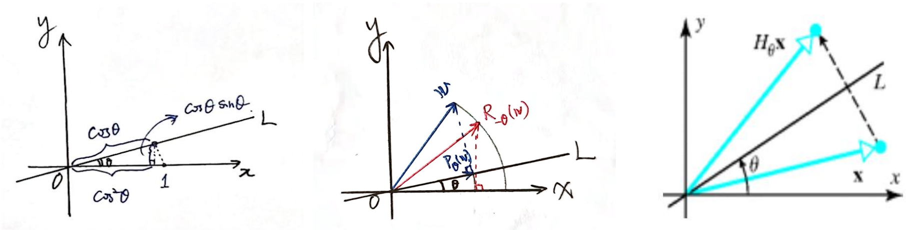
不难验证，我们有

$$[T+S] = [T] + [S], \quad [cT] = c[T].$$

**例 2.65.** *如图，设$L$是过原点的直线，与$x$轴正向的夹角是$\theta$弧度. 试求以下$\mathbb{R}^2$上算子的标准矩阵.*

*(1) $P_\theta$: 向直线$L$的正交投影.*

*(2) $H_\theta$: 关于直线$L$的镜像对称.*




解. (1) 我们尝试不同方法来处理这个问题. 第一种方法，根据定理2.11，取直线的一个方向向量$\mathbf{e} = (\cos\theta, \sin\theta)$,我们有

$$P_\theta(\mathbf{v}) = \mathrm{proj}_{\mathbf{e}}\mathbf{v} = \frac{\mathbf{v}\cdot\mathbf{e}}{\|\mathbf{e}\|^2}\mathbf{e}.$$

记$\mathbf{v} = (x, y)$，则

$$P_\theta(x,y) = \frac{x\cos\theta + y\sin\theta}{1}(\cos\theta, \sin\theta) = (x\cos^2\theta + y\cos\theta\sin\theta, x\cos\theta\sin\theta + y\sin^2\theta).$$

于是

$$[P_\theta] = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}. \tag{2.8}$$

第二种方法，通过计算$P_\theta(1,0)$和$P_\theta(0,1)$来获得$[P_\theta]$. 如图，经过分析可知

$$P_\theta(1,0) = (\cos^2\theta, \cos\theta\sin\theta), \quad P_\theta(0,1) = (\sin\theta\cos\theta, \sin^2\theta).$$

从而得(2.8). 这时，正交投影$P_\theta$的一般表达式可以通过矩阵乘法

$$\begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos^2\theta + y\cos\theta\sin\theta \\ x\cos\theta\sin\theta + y\sin^2 \end{bmatrix}.$$

来得到.

第三种方法，在之前得例子中，我们已经计算了向$x$轴正交投影（即$P_0$）的矩阵. 如果我们将$\mathbf{v}$与直线$L$都顺时针转动$\theta$角，那么向$L$作正交投影就变成了向$x$轴方向作正交投影，这时再逆时针转$\theta$角，就可将投影所得的垂足转回直线$L$上. 也就是说，我们有$P_\theta = R_\theta P_0 R_{-\theta}$. 因此

$$[P_\theta] = [R_\theta][P_0][R_{-\theta}] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}.$$

(2) 对$\mathbb{R}^2$中的一个向量$\mathbf{v}$的顶点，其与关于直线$L$的镜像对称点的连线的中点位置，即向直线$L$的正交投影处. 因此

$$H_\theta(\mathbf{v}) - \mathbf{v} = 2(P_\theta(\mathbf{v}) - \mathbf{v}), \quad (\forall \mathbf{v} \in \mathbb{R}^2).$$

从而

$$H_\theta = 2P_\theta - I,$$

这里的$I$是$\mathbb{R}^2$上的恒等映射. 故

$$[H_\theta] = 2[P_\theta] - [I] = 2\begin{bmatrix}\cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta\end{bmatrix} - \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix} = \begin{bmatrix}\cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta\end{bmatrix}.$$

□

**定义 2.10.** *设$f: V \to W$. 如果$f$将$V$中的不同元素映为不同元素，则称$f$为单射. 若$W$中的每个元素都在$f$的像中，则称$f$为满射. 若$f$即单又满，则称$f$为双射.*

对于$A \in M_n(\mathbb{R})$，"矩阵变换$T_A : \mathbb{R}^n \to \mathbb{R}^n$为单射"，当且仅当"若$A\mathbf{x} = A\mathbf{y}$ $(\mathbf{x}, \mathbf{y} \in \mathbb{R}^n)$，则$\mathbf{x} = \mathbf{y}$". 在后一命题中记$\mathbf{z} = \mathbf{x} - \mathbf{y}$，则有"若$A\mathbf{z} = \mathbf{0}$ $(\mathbf{z} \in \mathbb{R}^n)$，则$\mathbf{z} = \mathbf{0}$"，即齐次线性方程组$A\mathbf{z} = \mathbf{0}$仅有平凡解. 由等价性定理知，这等价于$A$可逆. "矩阵变换$T_A : \mathbb{R}^n \to \mathbb{R}^n$为满射"，当且仅当"对任意$\mathbf{b} \in \mathbb{R}^n$，存在$\mathbf{x} \in \mathbb{R}^n$，使得$A\mathbf{x} = \mathbf{b}$". 同样由等价性命题，知其等价于$A$可逆. 于是$T_A$单射、满射、双射与$A$可逆四者都是相互等价的.

**定义 2.11.** *若矩阵变换$T_A : \mathbb{R}^n \to \mathbb{R}^n$为双射，则称$T_{A^{-1}}$为$T_A$的逆变换（或逆算子），记为$T_A^{-1} = T_{A^{-1}}$.*

**例 2.66.** *求$\mathbb{R}^2$绕原点逆时针转动$\theta$角的算子$R_\theta$的逆.*

解.（略.） □

**例 2.67.** *求$\mathbb{R}^2$上的算子$T$的逆，其中*

$$T(x, y) = (2x + y, 3x + 4y).$$

解.（略.） □

注意到2阶初等矩阵具有如下形式：

$$\begin{bmatrix}k & 0 \\ 0 & 1\end{bmatrix},\ \begin{bmatrix}1 & 0 \\ 0 & k\end{bmatrix},\ \begin{bmatrix}0 & 1 \\ 1 & 0\end{bmatrix},\ \begin{bmatrix}1 & 0 \\ k & 1\end{bmatrix},\ \begin{bmatrix}1 & k \\ 0 & 1\end{bmatrix}.$$

对与前两者，我们将$k$的取值分为$0 < k < 1$, $k \geq 1$, $k < 0$三种情况. 对$k < 0$时，有

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

**命题 2.1.** *设$E \in M_2(\mathbb{R})$是初等矩阵，则矩阵变换$T_E$是以下之一：*

*(a) 沿着某坐标轴错切.*

*(b) 关于直线$y = x$作镜像对称.*

*(c) 沿着某坐标轴压缩.*

*(d) 沿着某坐标轴拉伸.*

*(e) 关于某坐标轴作镜像对称.*

*(f) 先关于某坐标轴作镜像对称，再沿着某坐标轴压缩或拉伸.*

**推论 2.68.** *若$T_A$是$\mathbb{R}^2$上的可逆算子，则它的几何效果是有限次错切、压缩、拉伸与镜像对称的复合.*

# 第3章　线性空间

## 3.1　线性空间的定义

在本课程中，我们记$\mathbb{F}=\mathbb{R}$或$\mathbb{F}=\mathbb{C}$. 选用实数还是复数由我们讨论的问题所决定，并且在讨论开始时就已经取定了，在讨论的过程中不再更换.

**定义 3.1.** 设$V$是一个非空集合，其上定义了加法与数乘，即

*(i)* 对任意$\mathbf{u},\mathbf{v}\in V$，均有$\mathbf{u}+\mathbf{v}\in V$;

*(ii)* 对任意$\mathbf{u}\in V$及$k\in\mathbb{F}$，均有$k\mathbf{u}\in V$.

若以下八条公理成立，则称$V$为数域$\mathbb{F}$上的线性空间或向量空间，而$V$中的元素称为向量.

(公理1)　$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$　$(\forall\mathbf{u},\mathbf{v}\in V)$.

(公理2)　$\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$　$(\forall\mathbf{u},\mathbf{v},\mathbf{w}\in V)$.

(公理3)　$\exists\mathbf{0}\in V$, *s.t.* $\mathbf{0}+\mathbf{u}=\mathbf{u}+\mathbf{0}=\mathbf{u}$　$(\forall\mathbf{u}\in V)$.

(公理4)　$\forall\mathbf{u}\in V$, $\exists\mathbf{v}\in V$, *s.t.* $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}=\mathbf{0}$.

(公理5)　$1\mathbf{u}=\mathbf{u}$　$(\forall\mathbf{u}\in V)$.

(公理6)　$(kh)\mathbf{u}=k(h\mathbf{u})$　$(\forall\mathbf{u}\in V\ \text{and}\ k,h\in\mathbb{F})$.

(公理7)　$k(\mathbf{u}+\mathbf{v})=k\mathbf{u}+k\mathbf{v}$　$(\forall\mathbf{u},\mathbf{v}\in V\ \text{and}\ k\in\mathbb{F})$.

(公理8)　$(k+h)\mathbf{u}=k\mathbf{u}+h\mathbf{u}$　$(\forall\mathbf{u}\in V\ \text{and}\ k,h\in\mathbb{F})$.

对于$\mathbb{F}=\mathbb{R}$（或$\mathbb{F}=\mathbb{C}$）时，我们亦称$V$为实（复）线性空间. 在此定义中，(i)、(ii)表示$V$关于加法与数乘是封闭的. (公理3)中的$\mathbf{0}$称为零向量，而(公理4)中的向量$\mathbf{v}$称为$\mathbf{u}$的负向量，往往被记为$-\mathbf{u}$. 事实上，公理1的加法交换律可以通过其它几条公理推出，感兴趣的读者可以自己尝试一下.

**例 3.1.** 以下均为实线性空间的例子.

*(1)* 在$V=\{\mathbf{0}\}$上定义$\mathbf{0}+\mathbf{0}=\mathbf{0}$，以及$k\mathbf{0}=\mathbf{0}$ $(k\in\mathbb{R})$，则$V$构成线性空间.

*(2)* 欧氏空间$\mathbb{R}^n$在向量加法与数乘下构成线性空间.

*(3)* 实数列全体$\mathbb{R}^{\mathbb{N}}=\{(u_1,u_2,\ldots,u_n,\ldots):\ u_i\in\mathbb{R}\,(i=1,2,\ldots)\}$在数列加法和数乘下构成线性空间，其加法和数乘是对每个分量分别进行的. 其中的零向量为零数列.

*(4)* 区间$I$上的实值函数全体$\mathbb{R}^I=\{f:\ I\to\mathbb{R}\}=\{(f_x)_{x\in I}\}$在函数加法和数乘下构成线性空间，其加法和数乘是对$I$中的每一点分别进行的. 其中的零向量为零函数.

*(5)* 设$n$为给定正整数. 次数小于等于$n-1$的实系数多项式全体$P_n$在多项式加法和数乘下构成线性空间. 其中的零向量为零多项式. 注意到多项式的次数满足

$$\deg(f(x)+g(x))\le\max\{\deg f(x),\deg g(x)\},\quad \deg(cf(x))=\deg(f(x)),\ (c\neq 0),$$

而零多项式的次数定义为$-\infty$. 因此$P_n$关于加法与数乘均封闭.

*(6)* 所有$m\times n$实矩阵全体$M_{m\times n}(\mathbb{R})$在矩阵加法与数乘下构成线性空间. 其零向量为矩阵0.

**例 3.2.** 设$\mathbb{C}^n=\{(z_1,z_2,\ldots,z_n):\ z_i\in\mathbb{C}\,(i=1,2,\ldots,n)\}$. 则

*(i)* $\mathbb{C}^n$是一个复线性空间.

*(ii)* $\mathbb{C}^n$是一个实线性空间.

它们作为集合是相同的，向量加法也是相同的，但是数乘使用的标量不同，前者可用复数作数乘，而后者只能用实数作数乘. 我们会在后文中再次展示它们的区别.

以上是一些常用线性空间的例子. 事实上，线性空间的定义是很宽泛的，下述例子是为了说明我们可以构造非常奇特的线性空间. 当然，在具体应用中，我们还是更多地用到前面地例子.

**例 3.3.** 设$\mathbb{F}=\mathbb{R}$，而$V=\mathbb{R}^+$为全体正数构成的集合. 定义加法与数乘为

$$u \oplus v = uv, \qquad k \odot u = u^k.$$

则$V$在$\oplus$与$\odot$下构成一个实线性空间.

*证明.* 在此，我们仅验证其中的一些公理. 首先正数的乘积和正数的实数方幂均是正数，因此$V$关于加法$\oplus$与数乘$\odot$是封闭的. 注意到

$$1 \odot u = 1u = u, \quad u \odot 1 = u1 = u, \quad (u \in V).$$

因此，起到零向量作用的是1. 对任意$u \in V$，我们有

$$u \oplus (1/u) = u(1/u) = 1, \quad (1/u) \oplus u = (1/u)u = 1,$$

即向量$u$的负向量为其倒数$1/u$. 下面我们来验证两项分配律：对$u, v \in V$及$k, h \in \mathbb{R}$，我们有

$$k \odot (u \odot v) = (u \odot v)^k = (uv)^k = u^k v^k = (u^k) \oplus (v^k) = (k \odot u) \oplus (k \odot v),$$

以及

$$(k+h) \odot u = u^{k+h} = u^k u^h = (u^k) \oplus (u^h) = (k \odot u) \oplus (h \odot u).$$

注意到最后一式中的$k+h$是对$\mathbb{R}$中的$k$、$h$进行的，因此使用的是常规的加法. □

**定理 3.4.** 设$V$为$\mathbb{F}$上的线性空间. 设$\mathbf{u} \in V$，而$k \in \mathbb{F}$. 则

*(i)* $0\mathbf{u} = \mathbf{0}$;

*(ii)* $(-k)\mathbf{u} = -(k\mathbf{u})$;

*(iii)* $k\mathbf{0} = \mathbf{0}$;

*(iv)* 若$k\mathbf{u} = \mathbf{0}$，则要么$k=0$，要么$\mathbf{u} = \mathbf{0}$.

*证明.* (i) 由公理8，得

$$0\mathbf{u} = (0+0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u}. \tag{3.1}$$

由公理4，存在$0\mathbf{u}$的负向量$-0\mathbf{u}$,满足$0\mathbf{u} + (-0\mathbf{u}) = \mathbf{0}$. 在式(3.1)两端同时加上$-0\mathbf{u}$，得

$$\mathbf{0} = 0\mathbf{u} + (-0\mathbf{u}) = (0\mathbf{u} + 0\mathbf{u}) + (-0\mathbf{u}) = 0\mathbf{u} + (0\mathbf{u} + (-0\mathbf{u})) = \mathbf{u} + \mathbf{0} = \mathbf{u}.$$

在上式中，从左向右数第三个等号使用了公理2，而最后一个等号使用了公理3.

(ii) 注意到

$$\mathbf{0} = 0\mathbf{u} = (1 + (-1))\mathbf{u} = 1\mathbf{u} + (-1)\mathbf{u} = \mathbf{u} + (-1)\mathbf{u}.$$

上式从左往右数第一、三、四个等号分别运用了(i)、公理8、公理5. 再根据公理4，得$(-1)\mathbf{u} = -\mathbf{u}$.

(iii) 由公理4、公理8知

$$k\mathbf{0} = k\big(\mathbf{u} + (-\mathbf{u})\big) = k\mathbf{u} + k(-\mathbf{u}).$$

而

$$k(-\mathbf{u}) = k\big((-1)\mathbf{u}\big) = \big(k(-1)\big)\mathbf{u} = \big((-1)k\big)\mathbf{u} = (-1)(k\mathbf{u}) = -(k\mathbf{u}).$$

其中从左向右数第一、五个等号使用了(ii)，而第二、四个等号应用了公理6. 这时，由公理4推出

$$k\mathbf{0} = k\mathbf{u} + \big(-(k\mathbf{u})\big) = \mathbf{0}.$$

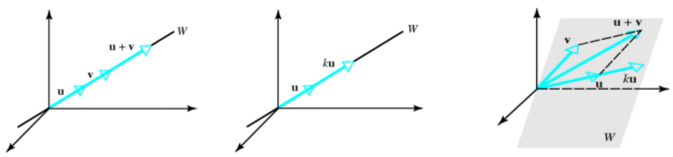
(iv) 当$k \neq 0$时，有

$$\mathbf{u} = 1\mathbf{u} = \big(\frac{1}{k}k\big)\mathbf{u} = \frac{1}{k}(k\mathbf{u}) = \frac{1}{k}\mathbf{0} = \mathbf{0},$$

其中从左向右数第一、三、四、五个等号分别使用了公理5、公理6、定理条件和(iii). □

## 3.2 子空间

在第二章中，我们研究过三维欧氏空间里的直线、平面等. 对于一般的线性空间来说，我们也感兴趣包含于其中的“更小”的线性空间.




**定义 3.2.** *设$V$是$\mathbb{F}$上的线性空间，而$W$是$V$的一个非空子集. 若在$V$中的加法和数乘下，集合$W$也构成一个线性空间，我们就称$W$为$V$的一个子空间.*

如上所述的$W$要成为线性空间，需要满足两条封闭性与八条公理. 那么，哪一些有可能不成立呢？注意到公理1-8中的代数运算，既然在$V$中已经成立，在$W$中也就自然成立了. 因此需要验证的是

(1) $W$对加法封闭；

(2) $W$对数乘封闭；

(3) 公理3中，$\mathbf{0} \in W$;

(4) 公理4中，对任意$\mathbf{u} \in W$，均有$-\mathbf{u} \in W$.

**定理 3.5.** *设$V$是$\mathbb{F}$上的线性空间，而$W$是$V$中的非空子集. 则$W$是$V$的子空间当且仅当以下两个命题成立：*

*(1) 对任意的$\mathbf{u}, \mathbf{v} \in W$，均有$\mathbf{u} + \mathbf{v} \in W$.*

*(2) 对任意的$k \in \mathbb{F}$和任意的$\mathbf{u} \in W$，均有$k\mathbf{u} \in W$.*

值得一题的是，以上(1)、(2)两个命题的一个等价写法是：对任意$\mathbf{u}, \mathbf{v} \in W$和任意$k, \ell \in \mathbb{F}$, 均有$k\mathbf{u} + \ell\mathbf{v} \in W$，即$W$对线性运算封闭. 更进一步，还等价于以下命题：对任意$r \geq 2$，任意$\mathbf{u}_1, \ldots, \mathbf{u}_r \in W$和任意$k_1, k_2, \ldots, k_r \in \mathbb{F}$, 均有$k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \ldots + r\mathbf{u}_r \in W$.

不妨设非空集合$W$中有向量$\mathbf{w}$. 当(1)、(2)成立时，我们有

$$\mathbf{0} = \mathbf{w} + (-1)\mathbf{w} \in W.$$

而对任意的$\mathbf{u} \in W$，亦有$-\mathbf{u} = (-1)\mathbf{u} \in W$. 因此(3)、(4)可由(1)、(2)推出. 结合之前的讨论，可知(1)、(2)可以保证$W$构成线性空间.

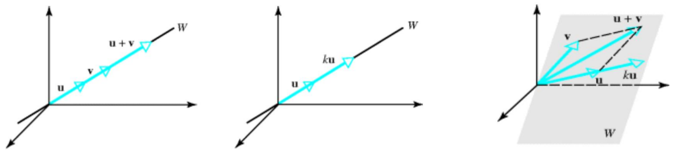
由于非空子集是否构成子空间只与封闭性有关，因此由集合包含关系的传递性，可以推出子空间的传递性.

**推论 3.6.** *设$W$是$\mathbb{F}$上的线性空间，设$V$是$W$的子空间，而$U$是$V$的一个子空间. 那么$U$也是$W$的一个子空间.*

下面我们来看一些关于子空间的例子.

**例 3.7.** 设$V$是$\mathbb{F}$上的线性空间，则$\{\mathbf{0}\}$是$V$的一个子空间，称为零子空间. 而$V$也是$V$的一个子空间，称为全子空间. 任何一个线性空间都具有零子空间与全子空间，因此我们称它们为平凡子空间.

**例 3.8.** 在欧氏空间$\mathbb{R}^3$中，过原点的直线与平面均为$\mathbb{R}^3$的子空间. 需要注意的是，直线$3x+4y+z=1$不




是子空间，因为它不过原点，不包含$\mathbf{0}$.

**例 3.9.** 记$F(-\infty,+\infty)$为区间$(-\infty,+\infty)$上的所有实值函数全体，它在函数加法和数乘下构成实线性空间. 以下集合均为$F(-\infty,+\infty)$的子空间.

*(1)* 区间$(-\infty,+\infty)$上的所有连续函数全体所构成的集合$C(-\infty,+\infty)$;

*(2)* 区间$(-\infty,+\infty)$上的所有具有连续导数的函数全体所构成的集合$C^1(-\infty,\infty)$;

*(3)* 区间$(-\infty,+\infty)$上的所有任意阶可导的函数全体所构成的集合$C^\infty(-\infty,+\infty)$;

*(4)* 多项式全体所构成的集合$P_\infty$;

*(5)* 次数小于$n$的多项式全体所构成的集合$P_n$.

事实上，前一个集合均为后一个的子空间.

**例 3.10.** 以下集合均为$M_n(\mathbb{R})$的子空间.

*(6)* 所有$n$阶对称矩阵全体所构成的集合$U$.

*(7)* 所有$n$阶上三角矩阵全体所构成的集合$V$.

*(8)* 所有$n$阶对角矩阵全体所构成的集合$W$.

事实上，集合$W$也是$U$和$V$的子空间.

当我们有集合$V$的子集$U,W$时，可以通过交、并来得到新的子集. 对线性空间而言，我们考察它们的交与和. 定义

$$U+W=\{\mathbf{u}+\mathbf{w}:\mathbf{u}\in U,\ \mathbf{w}\in W\}.$$

**定理 3.11.** 设$V$是$\mathbb{F}$上的线性空间，而$U$和$W$均为$V$的子空间. 则$U\cap W$与$U+W$也是$V$的子空间.

证明. 我们先验证$U\cap W$对于线性运算均封闭: 对任意$v_i\in U\cap W$及$k_i\in\mathbb{F}$ $(i=1,2)$，我们有$v_i\in U$, $v_i\in W$. 由于$U$和$W$均为$V$的子空间，因此

$$k_1v_1+k_2v_2\in U,\quad k_1v_1+k_2v_2\in W.$$

从而$k_1v_1+k_2v_2\in U\cap W$，封闭性得证.

我们再验证$U+W$对于线性运算均封闭: 对任意$v_i\in U\cap W$及$k_i\in\mathbb{F}$ $(i=1,2)$，不妨设$v_i=u_i+w_i$ $(i=1,2)$，其中$u_i\in U$，$w_i\in W$. 由于$U$和$W$均为$V$的子空间，因此

$$k_1u_1+k_2u_2\in U,\quad k_1w_1+k_2w_2\in W.$$

于是有

$$k_1v_1+k_2v_2=k_1(u_1+w_1)+k_2(u_2+w_2)=(k_1u_1+k_2u_2)+(k_1w_1+k_2w_2)\in U+W.$$

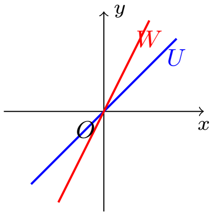
封闭性得证. $\square$

需要注意的是，集合$U\cap W$常常不是子空间.

**例 3.12.** 在$V=\mathbb{R}^2$中，设$U$为直线$y=x$，而$W$为直线$y=2x$. 不难看出

$$U\cap W=\{\mathbf{0}\},\quad U+W=\mathbb{R}^2.$$

而$U\cup W$仅仅是把两条直线上的点放到一个集合里而已.




接下来，我们将线性组合的概念推广到一般线性空间中.

**定义 3.3.** 设$V$为$\mathbb{F}$上的线性空间. 设$\mathbf{v},\mathbf{v}_1,\mathbf{v}_2,\cdots,\mathbf{v}_r\in V$，而$k_1,k_2,\ldots,k_r\in\mathbb{F}$. 若

$$\mathbf{v}=k_1\mathbf{v}_1+k_2\mathbf{v}_2+\cdots k_r\mathbf{v}_r,$$

我们称向量$\mathbf{v}$是向量$\mathbf{v}_1,\mathbf{v}_2,\cdots,\mathbf{v}_r$的一个线性组合，而$k_1,k_2,\ldots,k_r$称为这个线性组合的系数.

**例 3.13.** 在$\mathbb{C}^3$中，考虑

$$\mathbf{u}=(i,2,-1),\quad \mathbf{v}=(0,-i,2).$$

试问以下向量是不是$\mathbf{u}$和$\mathbf{v}$的线性组合?

*(1)* $\mathbf{w}=(-1,i,2-i)$; *(2)* $\mathbf{w}'=(2,2+i,-3)$.

*解.* (1) 列出关于系数$k,l$的线性方程组$k\mathbf{u}+l\mathbf{v}=\mathbf{w}$. 要判断$\mathbf{w}$是不是$\mathbf{u}$和$\mathbf{v}$的线性组合，就是要看这个方程组是否有解. 我们用矩阵形式写出：

$$\left[\begin{array}{c:c} \mathrm{i} & 0\\ 2 & -\mathrm{i}\\ -1 & 2\end{array}\right]\begin{bmatrix}k\\ l\end{bmatrix}=\begin{bmatrix}-1\\ \mathrm{i}\\ 2-\mathrm{i}\end{bmatrix}.$$

对增广矩阵作初等行变换，得

$$\left[\begin{array}{cc:c} \mathrm{i} & 0 & -1\\ 2 & -\mathrm{i} & \mathrm{i}\\ -1 & 2 & 2-\mathrm{i}\end{array}\right]\longrightarrow\left[\begin{array}{cc:c} \mathrm{i} & 0 & -1\\ 0 & -\mathrm{i} & -\mathrm{i}\\ 0 & 2 & 2\end{array}\right]\longrightarrow\left[\begin{array}{cc:c} \mathrm{i} & 0 & -1\\ 0 & -\mathrm{i} & -\mathrm{i}\\ 0 & 0 & 0\end{array}\right].$$

因为在阶梯型中，系数矩阵与增广矩阵的秩相同，因此方程组有解. 故$\mathbf{w}$是$\mathbf{u}$与$\mathbf{v}$的线性组合.

(2) 这个例子留给读者，$\mathbf{w}'$不是$\mathbf{u}$与$\mathbf{v}$的线性组合. □

借助线性组合的概念，我们可以用一些向量去生成子空间. 首先需要下述定理.

**定理 3.14.** 设$V$是$\mathbb{F}$上的线性空间，而$S=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_r\}$为$V$的非空子集. 记$W$为$\mathbf{w}_1,\ldots,\mathbf{w}_r$的所有线性组合全体所构成的集合，即

$$W=\{k_1\mathbf{w}_1+k_2\mathbf{w}_2+\ldots+k_r\mathbf{w}_r:\ k_1,k_2,\ldots,k_r\in\mathbb{F}\}.$$

则

*(1)* $W$构成$V$的一个子空间;

*(2)* 若$U$也是$V$的子空间，且$\mathbf{w}_1,\ldots,\mathbf{w}_r\in U$，则$W\subseteq U$.

上述定理中的(2)指出，这样的$W$是包含$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$的“最小”的子空间. 这个概念是重要且有用的，因为当我们讨论一个问题时，可能仅仅涉及到其中几个向量的线性关系，而不需要用到整个$V$的信息，这时我们限制在$W$上进行讨论就足够了.

证明. (1) 对任意$\mathbf{v}, \mathbf{v}' \in W$及任意$\alpha, \alpha' \in \mathbb{F}$，不妨设

$$\mathbf{v} = k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \ldots + k_r\mathbf{w}_r, \quad \mathbf{v}' = k_1'\mathbf{w}_1 + k_2'\mathbf{w}_2 + \ldots + k_r'\mathbf{w}_r.$$

不难得到

$$\begin{aligned}\alpha\mathbf{v} + \alpha'\mathbf{v}' &= \alpha(k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \ldots + k_r\mathbf{w}_r) + \alpha'(k_1'\mathbf{w}_1 + k_2'\mathbf{w}_2 + \ldots + k_r'\mathbf{w}_r) \\ &= (\alpha k_1 + \alpha' k_1')\mathbf{w}_1 + (\alpha k_2 + \alpha' k_2')\mathbf{w}_2 + \ldots + (\alpha k_r + \alpha' k_r')\mathbf{w}_r \in W.\end{aligned}$$

(2) 由于$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r \in U$，而子空间$U$对线性运算封闭，因此

$$k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \ldots + k_r\mathbf{w}_r \in U$$

对任意$k_1, k_2, \ldots, k_r \in \mathbb{F}$成立. 故$W \subseteq U$. □

**定义 3.4.** 我们将上述定理中的$W$称为集合$S$或向量$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$的*(线性)包络或扩张*，并记为*span*$(S)$或*span*$\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r\}$，即

$$\mathit{span}\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r\} = \{k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \ldots + k_r\mathbf{w}_r : k_1, k_2, \ldots, k_r \in \mathbb{F}\}.$$

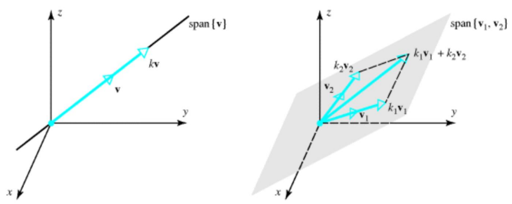
我们也称集合$S$或向量$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$张成了$W$.

**例 3.15.** 标准向量

$$\mathbf{e}_1 = \{1, 0, \ldots, 0\}, \ \mathbf{e}_2 = \{0, 1, \ldots, 0\}, \ \ldots, \ \mathbf{e}_n = \{0, \ldots, 0, 1\}$$

张成了空间$\mathbb{R}^n$.

**例 3.16.** 在图中，试观察*span*$(\{\mathbf{v}\})$与*span*$\{\mathbf{v}_1, \mathbf{v}_2\}$.




**例 3.17.** 证明矩阵

$$A_1 = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

张成了所有2阶上三角矩阵所构成的空间$V$.

证明. 对任意的$\begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \in V$，设

$$k_1\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} + k_2\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} + k_3\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$$

即

$$\begin{cases} k_1 + k_2 + 2k_3 = a, \\ k_1 + k_3 = b, \\ 2k_1 + 2k_2 + 3k_3 = d. \end{cases}$$

注意到系数矩阵的行列式满足

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

根据等价性定理，知方程组有解. 因此任何一个2阶上三角矩阵均为$A_1, A_2, A_3$的线性组合，对应系数即方程组的解$(k_1, k_2, k_3)$. 故span$\{A_1, A_2, A_3\} \supset V$. 另一个方向的包含关系是自然成立的,因此 span$\{A_1, A_2, A_3\} = V$. □

从线性包络的定义，不难得到以下结论.

**定理 3.18.** 设$S_1 = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$，$S_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s\}$均为线性空间$V$的非空子集.则*span*$(S_1) \subseteq$ *span*$(S_2)$当且仅当$S_1$中的每个向量都是$S_2$中向量的线性组合.

**例 3.19.** 试验证

$$\mathit{span}\Big(\{(1,0),(1,2),(-1,1)\}\Big) \subseteq \mathit{span}\Big(\{(1,1),\ (1,-1)\}\Big).$$

证明. 由以下线性组合，知包含关系成立.

$$(1,0) = \frac{1}{2}(1,1) + \frac{1}{2}(1,-1), \quad (1,2) = \frac{3}{2}(1,1) - \frac{1}{2}(1,-1), \quad (-1,1) = -(1,-1).$$

□

**定理 3.20.** 设$U = \mathit{span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r\}$和$W = \mathit{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s\}$均为线性空间$V$的子空间. 则

$$U + W = \mathit{span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s\}.$$

证明. 方便起见，记$V_0 = \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s\}$. 对任意$v \in U + W$，不妨设$v = u + w$，其中$u \in U$, $w \in W$. 由$U$, $W$的构造，设

$$u = k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \ldots + k_r\mathbf{u}_r, \quad w = l_1\mathbf{w}_1 + l_2\mathbf{w}_2 + \ldots + l_s\mathbf{w}_s.$$

则有

$$v = u + v = k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \ldots + k_r\mathbf{u}_r + l_1\mathbf{w}_1 + l_2\mathbf{w}_2 + \ldots + l_s\mathbf{w}_s \in V_0.$$

反之，对任意$v \in V_0$，设

$$v = k_1'\mathbf{u}_1 + k_2'\mathbf{u}_2 + \ldots + k_r'\mathbf{u}_r + l_1'\mathbf{w}_1 + l_2'\mathbf{w}_2 + \ldots + l_s'\mathbf{w}_s.$$

取

$$u = k_1'\mathbf{u}_1 + k_2'\mathbf{u}_2 + \ldots + k_r'\mathbf{u}_r, \quad w = l_1'\mathbf{w}_1 + l_2'\mathbf{w}_2 + \ldots + l_s'\mathbf{w}_s.$$

则有$u \in U$, $w \in W$，进而$v = u + w \in U + V$. 定理得证. □

思考题:若给出$\mathbb{R}^n$的子空间$U = \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r\}$和$W = \mathrm{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s\}$,如何求空间$U \cap W$?

**定理 3.21.** 设$A \in M_{m\times n}(\mathbb{F})$. 则齐次线性方程组$A\mathbf{x} = \mathbf{0}$的解的全体构成$\mathbb{F}^n$的一个子空间.

证明. 对任意的方程$A\mathbf{x}=\mathbf{0}$的解$\mathbf{x}_1,\mathbf{x}_2$以及任意的$k_1,k_2\in\mathbb{F}$，有$A\mathbf{x}_1=A\mathbf{x}_2=\mathbf{0}$，从而

$$A(k_1\mathbf{x}_1+k_2\mathbf{x}_2)=k_1A\mathbf{x}_1+k_2A\mathbf{x}_2=k_1\mathbf{0}+k_2\mathbf{0}=\mathbf{0}.$$

即$k_1\mathbf{x}_1+k_2\mathbf{x}_2$也是方程$A\mathbf{x}=\mathbf{0}$的解. 解的全体构成$\mathbb{F}^n$的一个子空间. □

**定义 3.5.** 设$A\in M_{m\times n}$，我们称方程$A\mathbf{x}=\mathbf{0}$的解的全体所构成的集合为$A$的零空间，记为$\mathit{Null}(A)$. 若$A$可表为

$$A=[a_{ij}]_{m\times n}=\begin{bmatrix}\mathbf{r}_1\\ \hdashline \mathbf{r}_2\\ \hdashline \vdots\\ \hdashline \mathbf{r}_m\end{bmatrix}=\left[\begin{array}{c:c:c:c}\mathbf{c}_1&\mathbf{c}_2&\ldots&\mathbf{c}_n\end{array}\right].$$

我们称$\mathbb{F}^n$的子空间$\mathit{Row}(A):=\mathit{span}\{\mathbf{r}_1,\ldots,\mathbf{r}_m\}$为矩阵$A$的行空间,称$\mathbb{F}^m$的子空间$\mathit{Col}(A):=\mathit{span}\{\mathbf{c}_1,\ldots,\mathbf{c}_n\}$为矩阵$A$的列空间.

现在，定理1.22之后的注记，可以重新写为：若$A\in M_{m\times n}(\mathbb{F})$，则线性方程组$A\mathbf{x}=\mathbf{b}$相容当且仅当$\mathbf{b}\in\mathsf{Col}(A)$. 另外，设$T_A:\mathbb{F}^n\to\mathbb{F}^m$是以$A$为标准矩阵的矩阵变换，则该变换的像集为$A$的列空间，即

$$\{T_A(\mathbf{x}):\,x\in\mathbb{F}^n\}\;=\;\mathsf{Col}(A).$$

事实上，对任何$\mathbf{x}\in\mathbb{F}^n$，有$T_A(\mathbf{x})=A\mathbf{x}\in\mathsf{Col}(A)$. 另一方面，取$\mathbb{F}^n$中的标准向量$\mathbf{e}_i$ $(1\le i\le n)$，则$T_A(\mathbf{e}_i)=A\mathbf{e}_i=\mathbf{c}_i$. 由于矩阵变换$T_A$保持线性性，因此$\mathsf{Col}(A)$中的每一个$\mathbf{c}_1,\mathbf{c}_2,\ldots,\mathbf{c}_n$的线性组合都是$\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n$的对应线性组合在$T_A$下的像.

## 3.3 线性相关与线性无关

当$V=\mathsf{span}\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$时，说明$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$这些向量足够丰富，可以张成整个空间$V$. 然而，用欧氏空间中100个平行的非零向量作成的线性包络，仍然只是一条直线，因为这些向量包含了冗余的信息. 为了描述“无冗余”，我们介绍以下“线性无关”的概念.

**定义 3.6.** 设$S=\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$为$\mathbb{F}$-线性空间$V$的非空子集. 若方程组

$$k_1\mathbf{v}_1+k_2\mathbf{v}_2+\cdots k_r\mathbf{v}_r=\mathbf{0}$$

在$\mathbb{F}$中仅有平凡解$k_i=0$ $(1\le i\le r)$，则称集合$S$（或向量组$S$、或其中的向量）是线性无关的. 反之，若上述方程有非平凡解，则称集合$S$（或向量组$S$、或其中的向量）是线性相关的.

**例 3.22.** 判断下列$\mathbb{R}^3$中的向量

$$\mathbf{v}_1=(1,-2,3),\quad \mathbf{v}_2=(5,6,-1),\quad \mathbf{v}_3=(3,2,1)$$

线性无关还是线性相关？

解.（略.） □

注：由于混合积$(\mathbf{v}_1\times\mathbf{v}_2)\cdot\mathbf{v}_3=0$，这三个向量是共面的.

**例 3.23.** 在$P_\infty$中，以下哪些向量是线性无关的？

*(a)* $S=\{0\}$.

*(b)* $S=\{1+x\}$.

*(c)* $S=\{x,\,x^2\}$.

*(d)* $S=\{2-x,\,4-2x\}$.

*(e)* $S=\{0,\,x,\,2-x^2,\,4+x^3,\,5+x^4\}$.

解.（略.） □

我们将以上例子归纳总结为下述命题.

**命题 3.1.** *(1)* 包含零向量的集合$S=\{\mathbf{0},\mathbf{v}_2,\dots,\mathbf{v}_r\}$是线性相关的.

*(2)* 单元集$S=\{\mathbf{v}\}$线性无关当且仅当$\mathbf{v}\neq\mathbf{0}$.

*(3)* 二元集合$S=\{\mathbf{u},\mathbf{v}\}$线性相关，当且仅当其中一个向量是另一个的常数倍.

**例 3.24.** 证明下述命题.

*(1)* 设$\mathbb{F}=\mathbb{C}$，$V=\mathbb{C}$. 其中的向量$1,i$是线性相关的.

*(2)* 设$\mathbb{F}=\mathbb{R}$，$V=\mathbb{C}$, 其中的向量$1,i$是线性无关的.

证明. (1) 因为存在$c=\mathrm{i}\in\mathbb{F}=\mathbb{C}$，使得$\mathrm{i}=c\cdot 1$. 故$1,\mathrm{i}$线性相关. (2) 设$k\cdot 1+l\cdot\mathrm{i}=0$，其中$k,l\in\mathbb{F}=\mathbb{R}$，则等式左端的实部与虚部均等于0，从而$k=l=0$. 因此$1,\mathrm{i}$线性无关. □

接下来，我们将线性相关的等价描述推广到更一般的情形.

**定理 3.25.** 设$S$是线性空间$V$中的子集，且包含至少2个元素. 则$S$线性相关当且仅当$S$中的某个向量可以写成其余向量的线性组合.

证明. 记$S=\{\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_r\}$ $(r\geq 2)$.

设$S$是线性相关的,即存在不全为零的数$k_1,k_2,\dots,k_r$，使得

$$k_1\mathbf{v}_1+k_2\mathbf{v}_2+\ldots+k_r\mathbf{v}_r=\mathbf{0}.$$

不妨设$k_t\neq 0$，其中$1\leq t\leq r$. 则有

$$\mathbf{v}_t=-\frac{1}{k_t}\left(k_1\mathbf{v}_1+\ldots+k_{t-1}\mathbf{v}_{t-1}+k_{t-1}\mathbf{v}_{t-1}\right).$$

反之，设$\mathbf{v}_s$是其它向量的线性组合,即存在数$l_1,\dots,l_{s-1},l_{s+1},\dots,l_s$，使得

$$\mathbf{v}_s=l_1\mathbf{v}_1+\ldots+l_{s-1}\mathbf{v}_{s-1}+l_{s+1}\mathbf{v}_{s+1}+\ldots+l_r\mathbf{v}_r.$$

于是有

$$l_1\mathbf{v}_1+\ldots+l_{s-1}\mathbf{v}_{s-1}+(-1)\mathbf{v}_s+l_{s+1}\mathbf{v}_{s+1}+\ldots+l_r\mathbf{v}_r=\mathbf{0}.$$

故$S$线性相关. □

**例 3.26.** 试将以下向量中的一个表示为其它向量的线性组合.

$$\mathbf{v}_1=(1,-2,3),\quad \mathbf{v}_2=(5,6,-1),\quad \mathbf{v}_3=(3,2,1).$$

解.（略.） □

**例 3.27.** 在$P_3$中，取

$$p_0(x)=4,\ p_1(x)=1+x,\ p_2(x)=5+3x-2x^2,\ p_3(x)=1+3x-x^2.$$

*(1)* 向量$p_1,p_2,p_3$线性相关还是线性无关？

*(2)* 向量$p_1,p_2,p_3$线性相关还是线性无关？

解.（略.） □

**例 3.28.** 设向量$\mathbf{u}, \mathbf{v}, \mathbf{w}$线性无关. 证明向量

$$\mathbf{x} = \mathbf{u}, \quad \mathbf{y} = \mathbf{u} + \mathbf{v}, \quad \mathbf{z} = \mathbf{u} + \mathbf{v} + \mathbf{w} \tag{3.2}$$

也线性无关.

*证明.* 令$k_1\mathbf{x} + k_2\mathbf{y} + k_3\mathbf{z} = \mathbf{0}$. 将(3.2)式代入后，整理得

$$(k_1 + k_2 + k_3)\mathbf{u} + (k_1 + k_2)\mathbf{v} + k_1\mathbf{w} = \mathbf{0}.$$

由于$\mathbf{u}, \mathbf{v}, \mathbf{w}$线性无关，故有

$$\begin{cases} k_1 + k_2 + k_3 & = 0, \\ k_1 + k_2 & = 0, \\ k_1 & = 0. \end{cases}$$

上述线性方程仅有平凡解$k_1 = k_2 = k_3 = 0$. 故$\mathbf{x}, \mathbf{y}, \mathbf{z}$线性无关. □

**例 3.29.** 设$r \in \mathbb{N}$. 设$V = F(-\infty, +\infty)$为$\mathbb{R}$上所有实值函数全体，证明$V$中得向量

$$\sin x, \, \sin 2x, \, \ldots, \, \sin rx$$

线性无关.

*证明.* 考虑方程

$$k_1 \sin x + k_2 \sin 2x + \ldots + k_r \sin rx = 0,$$

右侧的0表示零函数. 注意到二阶导数$(\sin lx)'' = -l^2 \sin lx$ $(l = 1, 2, \ldots, r)$. 在上式两端依次求$j$ $(j = 0, 2, 4, \ldots, 2(r-1))$阶导，得

$$\begin{cases} k_1 \sin x + k_2 \sin 2x + \ldots + k_{r-1} \sin(r-1)x + k_r \sin rx = 0, \\ k_1 \sin x + 2^2 k_2 \sin 2x + \ldots + (r-1)^2 k_{r-1} \sin(r-1)x + r^2 k_r \sin rx = 0, \\ k_1 \sin x + 2^4 k_2 \sin 2x + \ldots + (r-1)^4 k_{r-1} \sin(r-1)x + r^4 k_r \sin rx = 0, \\ \ldots\ldots\ldots, \\ k_1 \sin x + 2^{2(r-1)} k_2 \sin 2x + \ldots + (r-1)^{2(r-1)} k_{r-1} \sin(r-1)x + r^{2(r-1)} k_r \sin rx = 0. \end{cases}$$

对每个$x$，我们将上式写为矩阵形式$A\mathbf{y} = \mathbf{0}$，其中

$$A = \begin{bmatrix} 1 & 1 & \ldots & 1 & 1 \\ 1 & 2^2 & \ldots & (r-1)^2 & r^2 \\ 1 & 2^4 & \ldots & (r-1)^4 & r^4 \\ & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2^{2(r-1)} & \ldots & (r-1)^{2(r-1)} & r^{2(r-1)} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} k_1 \sin x \\ k_2 \sin 2x \\ k_3 \sin 3x \\ \vdots \\ k_r \sin rx \end{bmatrix}.$$

注意到$A$的行列式是范德蒙矩阵，满足

$$\mathsf{det}(A) = \prod_{1 \le j < i \le r} (i^2 - j^2) \neq 0.$$

因此$A$可逆，而$\mathbf{y} = A^{-1}\mathbf{0} = \mathbf{0}$. 这时，我们取合适的$x$，使得$\sin x, \sin 2x, \ldots, \sin rx$均非零，便得到了$k_1 = k_2 = \ldots = k_r = 0$. 故这组函数是线性无关的. □

## 3.4 基与坐标

如果一个线性空间$V$能够写成$S$的线性包络，那么只要我们研究清楚$S$中的向量，则$V$中其它向量的信息由它们作线性组合就可以得到. 如何选取这样的$S$呢？空间$V$可以是自己的线性包络，但是把$S$取为$V$，就偏离了“化繁为简”的目的. 我们希望能够用尽可能少的向量组成$S$，而线性无关的一组向量是没有冗余信息的.

**定义 3.7.** 设$S$为线性空间$V$中的有限子集，如$S$满足

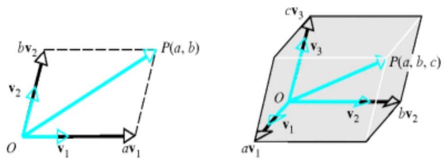
*(1)* $S$张成$V$；

*(2)* $S$线性无关，

则称$S$（或其中的向量）为$V$的一组基.

定义中的(1)指出$S$中的向量足够多，而(2)指出$S$中没有冗余的向量.

**例 3.30.** 向量$\{e_1, e_2, \cdots, e_n\}$构成欧氏空间$\mathbb{R}^n$的一组基，我们称其为$\mathbb{R}^n$的标准基. 下图中的向量也构成$\mathbb{R}^2$或$\mathbb{R}^3$中的一组基. 例如，向量$(1,1,2)$、$(1,0,2)$与$(2,1,3)$构成$\mathbb{R}^3$的一组基.




**例 3.31.** 设$P_n$是由所有次数不超过$n-1$次的实多项式全体所构成的实线性空间. 证明以下两组向量均构成$P_n$的基:

*(A)* $p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = x^2$, $\ldots$, $p_{n-1}(x) = x^{n-1}$, 我们称其为$P_n$的标准基.

*(B)* $q_0(x) = 1$, $q_1(x) = x + c$, $q_2(x) = (x+c)^2$, $\ldots$, $q_{n-1}(x) = (x+c)^{n-1}$，其中$c$为给定非零常数.

证明. 这里只证明(B)的情形. 首先证明$q_0(x), q_1(x), \ldots, q_{n-1}(x)$线性无关. 令

$$0 = \sum_{i=0}^{n-1} k_i q_i(x) = \sum_{i=0}^{n-1} k_i (x+c)^i,$$

左侧为零多项式. 对比$n-1$次项的系数，得$k_{n-1} = 0$. 这时上式变成了$0 = \sum\limits_{i=0}^{n-2} k_i (x+c)^i$. 再对比$n-2$次项的系数，得$k_{n-2} = 0$. 依此类推，可证明$k_{n-1} = k_{n-2} = \ldots = k_1 = k_0 = 0$，即$q_0(x), q_1(x), \ldots, q_{n-1}(x)$线性无关.

现在,我们再证明$q_0(x), q_1(x), \ldots, q_{n-1}(x)$张成$P_n$. 事实上,对任意多项式$f(x) \in P_n(x)$. 利用$f(x)$在$x = -c$处得泰勒展式可知

$$f(x) = \sum_{i=0}^{i-1} \frac{f^{(i)}(-c)}{i!} (x+c)^i = \sum_{i=0}^{i-1} \frac{f^{(i)}(-c)}{i!} q_i(x).$$

(可使用泰勒展式的n次拉格朗日余项,并通过对比$n$次项系数知余项为零.)因此$q_0(x), q_1(x), \ldots, q_{n-1}(x)$张成$P_n$. □

**例 3.32.** 在$\mathbb{F}$-线性空间$M_{m\times n}(\mathbb{F})$中取集合$\{E_{ij} : 1 \le i \le m, 1 \le j \le n\}$，其中$E_{ij}$表示第$i$行、第$j$列的元素为1，其余元素为0的矩阵. 它们构成一组基，称为$M_{m\times n}(\mathbb{F})$的标准基.

接下来，我们探索一下如何寻找与矩阵有关的几个空间的基，基本的方法是初等行变换.

**定理 3.33.** 以下命题成立:

*(1)* 初等行变换不改变矩阵的零空间.

*(2)* 初等行变换不改变矩阵的行空间.

*(3)* 初等行变换不改变矩阵各列之间的线性关系.

上述(3)的具体描述是这样的:设矩阵$A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$经初等行变换变为$A' = \begin{bmatrix} \mathbf{c}_1' & \mathbf{c}_2' & \dots & \mathbf{c}_n' \end{bmatrix}$.倘若矩阵$A$满足

$$k_1\mathbf{c}_1 + k_2\mathbf{c}_2 + \ldots + k_n\mathbf{c}_n = \mathbf{0},$$

其中$k_1, k_2, \ldots, k_n \in \mathbb{F}$，则矩阵$A'$亦满足

$$k_1\mathbf{c}_1' + k_2\mathbf{c}_2' + \ldots + k_n\mathbf{c}_n' = \mathbf{0}.$$

即线性组合的系数在初等行变换下不变. 特别地,向量组$\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$线性相关（或无关）当且仅当$\mathbf{c}_1', \mathbf{c}_2', \ldots, \mathbf{c}_n'$线性相关（或无关）.

*证明.*（略.） □

利用以上定理，我们可以通过初等行变换将矩阵化为（简约）阶梯阵，再通过阶梯阵分析对应空间的信息.

**例 3.34.** 分别求以下矩阵的行空间和列空间的一组基:

$$A = \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix}.$$

*解.* 经过初等行变换，上述矩阵具有阶梯阵

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

由于初等行变换不改变矩阵的行空间，因此

$$\mathsf{Row}(A) = \mathsf{Row}(R) = \mathsf{span}\left\{ \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 2 & 0 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right\}.$$

观察阶梯阵$R$，不难发现其主元所在第一、三、六列是线性无关的，而第二、四、五列可写成第一、三、六列的线性组合. 因此

$$\mathsf{Col}(R) = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

由于初等行变换不改变矩阵各列之间的线性关系，因此$\mathsf{Col}(A)$可由$A$的第一、三、六列张成，即

$$\mathsf{Col}(A) = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ -5 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -3 \\ 15 \\ 18 \end{bmatrix} \right\}.$$

□

现在我们把以上例题的经验总结为下述定理，其具体证明留给读者完成.

**定理 3.35.** *设矩阵$A$具有阶梯型$R$. 则下述命题成立：*

*(1) Null$(A)$ = Null$(R)$;*

*(2) Row$(A)$ = Row$(R)$，且$R$中的所有非零行构成该空间的一组基.*

*(3) $R$中所有非零首元所在列构成$Col(R)$的一组基，而$A$中所有与它们对应的列构成$Col(A)$的一组基.*

**例 3.36.** *在$\mathbb{R}^4$中，取*

$$\mathbf{v}_1 = (1,2,0,2),\ \mathbf{v}_2 = (3,6,0,6),\ \mathbf{v}_3 = (-2,-5,5,0),$$

$$\mathbf{v}_4 = (0,-2,10,8),\ \mathbf{v}_5 = (2,4,0,4),\ \mathbf{v}_6 = (0,-3,15,18)$$

*(1) 试寻找$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_6\}$的一个子集$S_0$，使其构成$Span\{\mathbf{v}_1, \ldots, \mathbf{v}_6\}$的一组基.（我们也称$S_0$为$S$的极大线性无关部分组.)*

*(2) 将每个向量都表示为$S_0$中向量的线性组合.*

解. 此例题的解法不唯一，常规的方法是将向量以列矩阵的形式组成矩阵

$$A = [\,\mathbf{v}_1 \,\vdots\, \mathbf{v}_2 \,\vdots \ldots \vdots\, \mathbf{v}_6\,] = \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix}.$$

(1) 由上一例题可知$S_0$可取为$\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_6\}$.

(2) 为了进一步得到线性组合的具体系数，我们可以将阶梯型$R$进一步化为简约阶梯型

$$R' = \begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} := [\,\mathbf{c}_1 \,\vdots\, \mathbf{c}_2 \,\vdots \ldots \vdots\, \mathbf{c}_6\,].$$

在$R$中可得到如下观察：

$$\mathbf{c}_1 = \mathbf{c}_1, \quad \mathbf{c}_2 = 3\mathbf{c}_1, \quad \mathbf{c}_3 = \mathbf{c}_3, \quad \mathbf{c}_4 = 4\mathbf{c}_1 + 2\mathbf{c}_3, \quad \mathbf{c}_5 = 2\mathbf{c}_1, \quad \mathbf{c}_6 = \mathbf{c}_6.$$

由于初等行变换不改变矩阵各列之间的线性关系，因此

$$\mathbf{v}_1 = \mathbf{v}_1, \quad \mathbf{v}_2 = 3\mathbf{v}_1, \quad \mathbf{v}_3 = \mathbf{v}_3, \quad \mathbf{v}_4 = 4\mathbf{v}_1 + 2\mathbf{v}_3, \quad \mathbf{v}_5 = 2\mathbf{v}_1, \quad \mathbf{v}_6 = \mathbf{v}_6.$$

. □

基在线性空间的理论中起着基石的作用，以下定理说明，基可以起到“坐标轴”的作用.

**定理 3.37.** *设$S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$是$\mathbb{F}$-线性空间$V$的一组基. 则$V$中的每个向量可以唯一地表示为*

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n,$$

*其中$c_1, c_2, \ldots, c_n \in \mathbb{F}$.*

证明. 由于$S$张成$V$，因此每个向量$\mathbf{v} \in V$都能写成$S$中向量的线性组合，这样的表示是存在的. 现在假设有两种表示方法

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = c_1'\mathbf{u}_1 + c_2'\mathbf{u}_2 + \cdots + c_n'\mathbf{u}_n.$$

则有

$$\begin{aligned}\mathbf{0}=\mathbf{v}-\mathbf{v}&=(c_1\mathbf{u}_1+c_2\mathbf{u}_2+\cdots+c_n\mathbf{u}_n)-c_1'\mathbf{u}_1+c_2'\mathbf{u}_2+\cdots+c_n'\mathbf{u}_n\\&=(c_1-c_1')\mathbf{u}_1+(c_2-c_2')\mathbf{u}_2+\ldots+(c_n-c_n')\mathbf{n}.\end{aligned}$$

注意到$\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n$是$V$的一组基，所以是线性无关的,于是上式右端的向量前的系数均为0. 所以$c_i=c_i'$ $(i=1,2,\ldots,n)$，即$\mathbf{v}$的表法是唯一的. □

有了上述定理的保证，我们就沿着基中的向量作坐标轴，坐标轴的正向与向量方向一致.

**定义 3.8.** 设$S=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n\}$是$\mathbb{F}$-线性空间$V$的一组基. 设$\mathbf{v}\in V$具有唯一的表示

$$\mathbf{v}=c_1\mathbf{u}_1+c_2\mathbf{u}_2+\cdots+c_n\mathbf{u}_n,$$

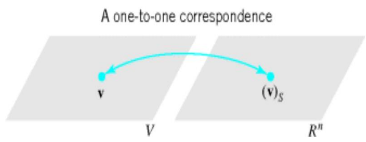
其中$c_1,c_2,\ldots,c_n\in\mathbb{F}$. 则称

$$(\mathbf{v})_S=(c_1,c_2,\ldots,c_n)$$

为$\mathbf{v}$在基$S$下的坐标向量，称

$$[\mathbf{v}]_S=\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}$$

为$\mathbf{v}$在基$S$下的坐标矩阵.




我们也将$(\cdot)_S$或$[\cdot]_S$称为在基$S$下的取坐标变换,不难验证它们是保持向量之间的线性关系的:对任意$\mathbf{v},\mathbf{w}\in V$，以及任意$k,l\in\mathbb{F}$，均有

$$[k\mathbf{v}+l\mathbf{w}]_S=k[\mathbf{v}]_S+l[\mathbf{w}]_S.$$

特别地，空间$V$中一组向量$\{\mathbf{v}_1,\ldots,\mathbf{v}_r\}$线性无关（或相关）当且仅当$\mathbb{R}^n$中的向量$\{[\mathbf{v}_1]_S,\ldots,[\mathbf{v}_r]_S\}$线性无关（或相关）. 利用取坐标变换，可以将抽象的线性空间的结构完全转换到具体的欧氏空间上，以便更好地讨论问题.

**例 3.38.** 求$\mathbb{R}^3$中向量$\mathbf{v}=(x,y,z)$在标准基$\{\mathbf{i},\mathbf{j},\mathbf{k}\}$下的坐标.

解.（略.） □

**例 3.39.** 取$\mathbb{R}^3$的一组基$S=\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$, 其中

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

*(1)* 求向量$\mathbf{v}=(5,-1,9)$在基$S$下的坐标.

*(2)* 求$\mathbf{w}\in\mathbb{R}^3$的的向量$\mathbf{w}$，使得它在基$S$下的坐标为$(\mathbf{w})_S=(-1,3,2)$.

解.（略.） □

**例 3.40.** *(1)* 求$P_n$中向量

$$p(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n$$

在标准基下的坐标.

*(2)* 求$M_2(\mathbb{R})$中矩阵$\begin{bmatrix}a&b\\c&d\end{bmatrix}$在标准基下的坐标.

解.（略.） □

**例 3.41.** 设向量$\mathbf{u},\mathbf{v},\mathbf{w}$线性无关. 证明向量

$$\mathbf{x}=\mathbf{u},\quad \mathbf{y}=\mathbf{u}+\mathbf{v},\quad \mathbf{z}=\mathbf{u}+\mathbf{v}+\mathbf{w} \tag{3.3}$$

也线性无关.

证明. 记$S=\{\mathbf{u},\mathbf{v},\mathbf{w}\}$，并令$V=\mathsf{span}\{\mathbf{u},\mathbf{v},\mathbf{w}\}$. 因$S$是线性无关的，因此构成$V$的一组基. 由(3.3)式得$\mathbf{x},\mathbf{y},\mathbf{z}$在基$S$的坐标如下：

$$(\mathbf{x})_S=(1,0,0),\quad (\mathbf{y})_S=(1,1,0),\quad (\mathbf{z})_S=(1,1,1).$$

注意到

$$\big((\mathbf{x})_S\times(\mathbf{y})_S\big)\cdot(\mathbf{z})_S=\begin{vmatrix}1&1&1\\0&1&1\\0&0&1\end{vmatrix}=1\neq 0,$$

知$(\mathbf{x})_S$, $(\mathbf{y})_S$, $(\mathbf{z})_S$线性无关，从而$\mathbf{x},\mathbf{y},\mathbf{z}$线性无关. □

**例 3.42.** 在$P_\infty$中，取$B=\{p_1(x),p_2(x),\ldots,p_6(x)\}$，其中

$$p_1(x)=1+2x+2x^3,\ p_2(x)=3+6x+6x^3,$$

$$p_3(x)=-2-5x+5x^2,\ p_4(x)=-2x+10x^2+8x^3,$$

$$p_5(x)=2+4x+4x^3,\ p_6(x)=-3x+15x^2+18x^3$$

*(1) 求一个$B$的子集$B_0$，使得$B_0$是span$\{p_1,\ldots,p_6\}$的一组基.*

*(2) 将$B$中的每个多项式表成$B_0$中向量的线性组合.*

解. 取标准基$S=\{1,x,x^2,x^3\}$，计算可得$[p_1(x)]_S=\mathbf{v}_i$ $(1\le i\le 6)$，其中$\mathbf{v}_i$为例3.36中的向量. 这时，我们把多项式空间的问题完全转化为欧氏空间的问题，利用之前的结果便可以得到此题的结果. □

## 3.5 维数与秩

在这一节中，我们将定义线性空间的维数. 若一个线性空间不能由一个有限集张成，则称该空间是无限维的，否则称该控件是有限维的.

思考题：证明多项式全体构成的线性空间$P_\infty$是无穷维的.

**定义 3.9.** 若非空有限集$S$是线性空间$V$的一组基，则定义$V$的维数为

$$\dim(V)=|S|.$$

另外，定义零空间的维数为$\dim(\{\mathbf{0}\})=\mathbf{0}$.

注意到一个线性空间的基不是唯一的. 若有两组基的向量个数不同，上述定义就有歧义. 我们需要以下定理来保证定义的合理性.

**定理 3.43.** 设$S$与$S'$是有限维$\mathbb{F}$-线性空间$V$的两组基，则$|S|=|S'|$.

证明. 不妨设$S=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_s\}$，$S'=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_t\}$，其中$s,t\in\mathbb{N}$.

由于$S'$线性无关，因此$M_{s\times 1}(\mathbb{F})$中的列矩阵$[\mathbf{v}_1]_S,[\mathbf{v}_2]_S,\ldots,[\mathbf{v}_t]_S$也线性无关. 故方程

$$k_1[\mathbf{v}_1]_S+k_2[\mathbf{v}_2]_S+\ldots+k_t[\mathbf{v}_t]_S=\mathbf{0}$$

仅有平凡解$k_1=k_2=\ldots=k_t=0$. 注意到该方程组可以写为$A\mathbf{k}=\mathbf{0}$，其中

$$A=\left[\ [\mathbf{v}_1]_S \,\vdots\, [\mathbf{v}_2]_S \,\vdots\, \ldots \,\vdots\, [\mathbf{v}_t]_S\ \right]\in M_{s\times t},\quad \mathbf{k}=\begin{bmatrix}k_1 & k_2 & \ldots & k_t\end{bmatrix}^T.$$

该方程组有$s$个方程和$t$个变元，且无自由变元. 由推论1.9知$t=\mathsf{rank}(A)\le s$.

另一方面，交换$S$与$S'$的角色作类似推导，亦可得到$s\le t$. 故$s=t$. □

由于$n$维空间的基都恰包含有$n$个向量，那么包含$n$个向量的集合是否能构成空间的基呢？

**定理 3.44.** *设$V$为$n$维$\mathbb{F}$-线性空间，而$S$为$V$中的$n$元子集. 则以下命题等价：*

*(1) $S$是$V$的一组基；*

*(2) $S$是线性无关的；*

*(3) $S$张成$V$.*

证明. 从(1)推导(2)、(3)"是显然的. 下面我们证明(2)与(3)等价，从而根据基的定义都可以推出(1).

设$B$为$V$的一组基，而$S=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_n\}$. 则$S$中向量的线性关系与$[\mathbf{w}_1]_B,[\mathbf{w}_2]_B,\ldots,[\mathbf{w}_n]_B$的线性关系是一致的. 注意到$|B|=\dim(V)=n$，所以这些坐标矩阵都是$n\times 1$阶的. 这时，$S$线性无关等价于线性方程组

$$\left[\ [\mathbf{w}_1]_B \,\vdots\, [\mathbf{w}_2]_B \,\vdots\, \ldots \,\vdots\, [\mathbf{w}_n]_B\ \right]\mathbf{k}=\mathbf{0}$$

仅有平凡解$\mathbf{k}=\mathbf{0}$. 而$S$张成$V$等价于对任意$\mathbf{b}\in\mathbb{F}^n$，方程组

$$\left[\ [\mathbf{w}_1]_B \,\vdots\, [\mathbf{w}_2]_B \,\vdots\, \ldots \,\vdots\, [\mathbf{w}_n]_B\ \right]\mathbf{k}=\mathbf{0}.$$

由等价性定理知，两者均等价于该$n\times n$的系数矩阵是可逆的. 定理得证. □

下面我们来看一些例子.

**例 3.45.** *若$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_r\}$是线性空间$V$中线性无关的集合，则$\dim(\mathit{span}(S))=r$.*

**例 3.46.** *(i) $\mathbb{C}$是1维复线性空间. (ii) $\mathbb{C}$是2维实线性空间.*

**例 3.47.** *以下实线性空间的维数为*

$$\dim(\mathbb{R}^n)=n,\quad \dim(P_n)=n,\quad \dim(M_{m\times n})=mn.$$

**例 3.48.** *设$U$、$V$分别为$n$阶对称、斜对称矩阵全体所构成的实线性空间. 求它们的维数.*

解. （略.） □

**例 3.49.** *求下述齐次线性方程组的解空间的一组基以及维数*

$$\begin{cases} x_1+3x_2-2x_3+2x_5=0\\ 2x_1+6x_2-5x_3-2x_4+4x_5-3x_6=0\\ 5x_3+10x_4+15x_6=0\\ 2x_1+6x_2+8x_4+4x_5+18x_6=0\end{cases}$$

解. （略.） □

基的定义有两条，如果一组向量有所冗余，可以删去一些；如果一组向量个数太少，可以添加一些，最后得到空间的一组基. 以下定理的证明细节留给读者完成.

**定理 3.50.** 设$S$为$V$为线性空间中的非空有限集合.

*(1) 设$S$线性无关. 若$\mathbf{v} \in V$，但$\mathbf{v} \notin \mathit{span}(S)$, 则$S \cup \{\mathbf{v}\}$仍线性无关.*

*(2) 设$S$中的向量$\mathbf{v}$可以表成$S$中其它向量的线性组合，则$\mathit{span}(S) = \mathit{span}(S \setminus \{v\})$.*

**例 3.51.** 在$P_\infty$中，取

$$p_1(x) = 1 - x^2, \quad p_2(x) = 2 - x^2, \quad p_3(x) = x^3.$$

试用上述定理证明$p_1(x), p_2(x), p_3(x)$线性无关.

解. 先考虑$S_1 = \{p_1(x)\}$，由于$p_1(x)$不是零多项式，因此$S_1$线性无关.

再考虑$p_2(x)$，不难看出$p_1(x)$与$p_2(x)$均不是对方的常数被，因此$p_2(x) \notin \mathsf{span}(S_1)$，由上述定理知$S_2 = \{p_1(x), p_2(x)\}$仍线性无关.

最后考虑$p_3(x)$，其次数为3. 因$p_1(x)$与$p_2(x)$的次数均为2，因此$\mathsf{span}(S_2)$中的多项式次数均不超过2，故$p_3(x) \notin \mathsf{span}(S_2)$. 由上述定理知$S_3 = \{p_1(x), p_2(x), p_3(x)\}$是线性无关的. □

以上定理和例子充分展示了，我们可以通过逐步添加的方式，得到更大的线性空间的基.

**推论 3.52.** 设$W$是线性空间$V$的子空间，满足$\dim(W) = r$以及$\dim(V) = n$. 设$\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$是$W$的一组基，则存在向量$\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n \in V \setminus W$，使得

$$\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r, \mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}.$$

构成$V$的一组基.

上述推论被称为基扩充定理. 当我们要探究线性空间$V$与它的子空间$U$之间的关系时，如果先在$V$中找一组基，其中的向量可能都不在$U$中. 但是，如果我们先从$U$中找一组基，再通过及扩充定理得到$V$的基，这时$V$的这组基包含了子空间$U$的信息.

**定理 3.53.** 设$W$是线性空间$V$的子空间，则有$\dim(W) \le \dim(V)$，且等号成立当且仅当$W = V$.

证明. 不妨设$\dim(V) = n$, $\dim(W) = r$. 取$W$的一组基$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$，根据及扩充定理，可将其扩充为$V$的一组基

$$\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r, \mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}.$$

因此$r \le n$. 且$r = n$当且仅当

$$W = \mathsf{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\} = V.$$

□

对于集合$V$的两个子集$S$与$T$来说，集合的交、并运算满足容斥原理：

$$|S \cup T| = |S| + |T| - |S \cap T|.$$

对于子空间的交与和，它们的维数也满足相似的结论.

**定理 3.54** (维数公式)**.** 设$W$是向量空间，而$U, V$是$W$的有限维子空间. 则

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V).$$

证明. 不妨设$\dim(U\cap V)=r$, $\dim(U)=k$, $\dim(V)=l$. 由于$U\cap V$为$U$、$V$的子空间，因此$r\le k$, $r\le l$. 现取$U\cap V$的一组基$\mathbf{w}_1,\dots,\mathbf{w}_r$，根据基扩充定理，存在$U$中的向量$\mathbf{u}_{r+1},\dots,\mathbf{u}_k$，使得$\mathbf{w}_1,\dots,\mathbf{w}_r,\mathbf{u}_{r+1},\dots,\mathbf{u}_k$构成$U$的一组基. 类似地，存在$V$中的向量$\mathbf{v}_{r+1},\dots,\mathbf{v}_l$，使得$\mathbf{w}_1,\dots,\mathbf{w}_r,\mathbf{v}_{r+1},\dots,\mathbf{v}_l$构成$V$的一组基.

这时向量组

$$\{\mathbf{w}_1,\dots,\mathbf{w}_r,\mathbf{u}_{r+1},\dots,\mathbf{u}_k,\mathbf{v}_{r+1},\dots,\mathbf{v}_l\}$$

是线性无关的.（留作思考题）

注意到

$$\begin{aligned}U+V&=\mathsf{span}\{\mathbf{w}_1,\dots,\mathbf{w}_r,\mathbf{u}_{r+1},\dots,\mathbf{u}_k\}+\mathsf{span}\{\mathbf{w}_1,\dots,\mathbf{w}_r,\mathbf{v}_{r+1},\dots,\mathbf{v}_l\}\\&=\mathsf{span}\{\mathbf{w}_1,\dots,\mathbf{w}_r,\mathbf{u}_{r+1},\dots,\mathbf{u}_k,\mathbf{v}_{r+1},\dots,\mathbf{v}_l\}.\end{aligned}$$

因此得到

$$\dim(U+V)=r+(k-r)+(l-r)=k+l-r=\dim U+\dim V-\dim(U\cap V).$$

□

**推论 3.55.** *设$W$是向量空间，而$U,V$是$W$的有限维子空间. 则*

$$\dim(U+V)\le\dim(U)+\dim(V).$$

现在，我们来探究一下与矩阵密切相关的几个空间的维数. 注意到矩阵$A$通过初等行变换化为阶梯型$R$后，它们的行空间可由$R$的所有非零行张成，因此$\dim\mathsf{Row}(A)=\dim\mathsf{Row}(R)=\mathsf{rank}(A)$. 另外，$\mathsf{Col}(R)$可由阶梯起始对应的所有主元列张成，而$Col(A)$可由$A$中的对应列张成，因此亦有$\dim\mathsf{Col}(A)=\dim\mathsf{Col}(R)=\mathsf{rank}(A)$. 我们将这个结论写下述定理,

**定理 3.56.** *对任意矩阵$A$, 有*

$$\mathit{rank}(A)=\dim(Row(A))=\dim(Col(A)).$$

我们也将矩阵$A$的行列、空间的维数称为$A$的行秩、列秩，它们与之前定义的矩阵的秩是相等的. 从行、列空间维数也很容易看出$0\le\mathsf{rank}(A)\le\min\{m,n\}$. 由于$A$的行完全对应$A^T$的列，因此

$$\mathsf{rank}(A^T)=\mathsf{rank}(A).$$

另外，初等、行列变换不改变矩阵的秩，即在一个矩阵左、右乘上初等矩阵，秩不会改变. 更进一步，可逆方阵都可以写成初等矩阵的乘积，因此我们得到以下定理.

**定理 3.57.** *设$A\in M_{m\times n}(\mathbb{F})$. 设$P\in M_m(\mathbb{F})$与$Q\in M_n(\mathbb{F})$均为可逆矩阵，则*

$$\mathit{rank}(PA)=\mathit{rank}(A)=\mathit{rank}(AQ).$$

**推论 3.58.** *分块初等行、列变换不改变矩阵的秩.*

**命题 3.2.** *设矩阵$A$、$B$、$C$的阶数复合下述各运算，则有*

*(1)* $\max\{\mathit{rank}(A),\mathit{rank}(B)\}\le\mathit{rank}\left(\begin{bmatrix}A & B\end{bmatrix}\right)\le\mathit{rank}(A)+\mathit{rank}(B)$;

*(2)** $\mathit{rank}\left(\begin{bmatrix}A & C\\ 0 & B\end{bmatrix}\right)\ge\mathit{rank}\left(\begin{bmatrix}A & 0\\ 0 & B\end{bmatrix}\right)=\mathit{rank}(A)+\mathit{rank}(B)$.

*（注：打星号的章节或定理不在大纲里必须掌握的范围.）*

证明. 这里我们仅证明(1). 而(2)留给读者们作为思考题.

记$C = \begin{bmatrix} A & B \end{bmatrix}$. 易知$\mathsf{Col}(C) \supseteq \mathsf{Col}(A), \mathsf{Col}(B)$，因此

$$\max\{\mathsf{rank}(A), \mathsf{rank}(B)\} \le \mathsf{rank}\left(\begin{bmatrix} A & B \end{bmatrix}\right).$$

另外，观察到$\mathsf{Col}(C) = \mathsf{Col}(A) + \mathsf{Col}(B)$，由维数公式知

$$\begin{aligned}\mathsf{rank}\left(\begin{bmatrix} A & B \end{bmatrix}\right) &= \dim \mathsf{Col}(C) = \dim(\mathsf{Col}(A) + \mathsf{Col}(B)) \\ &= \dim \mathsf{Col}(A) + \dim \mathsf{Col}(B) - \dim(\mathsf{Col}(A) \cap \mathsf{Col}(B)) \le \mathsf{rank}(A) + \mathsf{rank}(B).\end{aligned}$$

□

现在，我们给零空间的维数一个新名称.

**定义 3.10.** *我们将矩阵$A$的零空间维数称为$A$的零化度，记为nullity$(A)$，即nullity$(A) = \dim(\mathit{Null}(A))$.*

**定理 3.59.** *设$A \in M_{m\times n}(\mathbb{R})$，则有*

$$\mathit{rank}(A) + \mathit{nullity}(A) = n.$$

证明. 考虑齐次线性方程组$A\mathbf{x} = \mathbf{0}$，其中$\mathbf{x} \in \mathbb{F}^n$. 注意到方程组共$n$个变元，分为主变元和自由变元. 而

$$\text{主变元个数} = \text{非零阶梯的行数} = \mathsf{rank}(A), \quad \text{自由变元个数} = \text{方程解空间维数} = \mathsf{nullity}(A).$$

定理得证. □

**推论 3.60.** *设$A$为$n$阶方阵，则$A$可逆当且仅当rank$(A) = n$，当且仅当nullity$(A) = 0$.*

下面再介绍一些与秩有关的不等式.

**定理 3.61.** *设矩阵$A, B \in M_{m,n}$，$\lambda, \mu \in \mathbb{F}$. 则*

$$\mathit{rank}(\lambda A + \mu B) \le \mathit{rank}(A) + \mathit{rank}(B).$$

证明. (1) 将矩阵的列设为$A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n\end{bmatrix}$, $B = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n\end{bmatrix}$. 则

$$\begin{aligned}\mathsf{Col}(\lambda A + \mu B) &= \mathsf{span}\{\lambda\mathbf{a}_1 + \mu\mathbf{b}_1, \lambda\mathbf{a}_2 + \mu\mathbf{b}_2, \dots, \lambda\mathbf{a}_n + \mu\mathbf{b}_n\} \\ &\subseteq \mathsf{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n, \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\} \\ &= \mathsf{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\} + \mathsf{span}\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\} = \mathsf{Col}(A) + \mathsf{Col}(B).\end{aligned}$$

结合维数公式，我们得到

$$\begin{aligned}\mathsf{rank}(\lambda A + \mu B) &= \dim \mathsf{Col}(\lambda A + \mu B) \le \dim\big(\mathsf{Col}(A) + \mathsf{Col}(B)\big) \\ &\le \dim \mathsf{Col}(A) + \dim \mathsf{Col}(B) = \mathsf{rank}(A) + \mathsf{rank}(B).\end{aligned}$$

□

**定理 3.62.** **设矩阵$A \in M_{m\times k}$,$B \in M_{k\times n}$. 则*

*(1) Null$(AB) \supseteq$ Null$(B)$;*

*(2) Row$(AB) \subseteq$ Row$(B)$;*

*(3) Col$(AB) \subseteq$ Col$(A)$.*

*进而，我们有*

$$\mathit{rank}(AB) \le \min\{\mathit{rank}(A), \mathit{rank}(B)\}.$$

证明. (1) 对任意$\mathbf{x} \in \mathsf{Null}(B)$，我们有$B\mathbf{x} = \mathbf{0}$. 从而

$$(AB)\mathbf{x} = A(B\mathbf{x}) = A\mathbf{0} = \mathbf{0}.$$

即$\mathbf{x} \in \mathsf{Null}(AB)$.

(2)、(3) 回顾分块矩阵的乘法：记

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \hdashline \mathbf{a}_2 \\ \hdashline \vdots \\ \hdashline \mathbf{a}_m \end{bmatrix}, \quad B = \left[\begin{array}{c:c:c:c} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{array}\right]$$

则有

$$AB = \begin{bmatrix} \mathbf{a}_1 B \\ \hdashline \mathbf{a}_2 B \\ \hdashline \vdots \\ \hdashline \mathbf{a}_m B \end{bmatrix} = \left[\begin{array}{c:c:c:c} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_n \end{array}\right].$$

另外，我们有

$$\mathbf{a}_i B \in \mathsf{Row}(B),\ (1 \le i \le m),$$

$$A\mathbf{b}_j \in \mathsf{Col}(A),\ (1 \le j \le n).$$

因此可得

$$\mathsf{Row}(AB) = \mathsf{span}\{\mathbf{a}_1 B, \mathbf{a}_2 B, \dots, \mathbf{a}_m B\} \subseteq \mathsf{Row}(B),$$

$$\mathsf{Col}(AB) = \mathsf{span}\{A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_n\} \subseteq \mathsf{Col}(A).$$

这时，由(2)、(3)便可证得

$$\mathsf{rank}(AB) \le \min\{\mathsf{rank}(A), \mathsf{rank}(B)\}.$$

另外，由(1)知$\mathsf{nullity}(AB) \ge \mathsf{nullity}(B)$. 注意到矩阵$AB$与$B$的列数均为$n$. 利用秩与零化度的关系知

$$\mathsf{rank}(AB) + \mathsf{nullity}(AB) = n = \mathsf{rank}(B) + \mathsf{nullity}(B).$$

也可由此得到$\mathsf{rank}(AB) \le \mathsf{rank}(B)$. □

**定理 3.63.** *设矩阵$A \in M_{m\times k}$,$B \in M_{k\times n}$. 则

$$\textit{rank}(AB) \ge \textit{rank}(A) + \textit{rank}(B) - k.$$

证明. 注意到$k = \mathsf{rank}(I_k)$，我们将$AB$、$I_k$作成分块矩阵的对角子矩阵，将$A$、$B$也作成分块矩阵的对角子矩阵. 由于分块初等变换不改变矩阵得秩，我们有

$$\begin{bmatrix} AB & 0 \\ 0 & I_k \end{bmatrix} \xrightarrow{(\text{行}1)-A(\text{行}2)} \begin{bmatrix} AB & -A \\ 0 & I_k \end{bmatrix} \xrightarrow{(\text{列}1)+(\text{列}2)B} \begin{bmatrix} 0 & -A \\ B & I_k \end{bmatrix} \longrightarrow \begin{bmatrix} B & I_k \\ 0 & A \end{bmatrix}.$$

最后一步用到了交换两行、接着在第二行左乘$-I_n$. 此时，上式左端矩阵的秩等于$\mathsf{rank}(AB) + k$，而等式右端矩阵的秩大于等于$\mathsf{rank}(A) + \mathsf{rank}(B)$. 定理得证. □

思考题：试证明下述关于秩的不等式（假设相关矩阵的阶合适相应的运算）

(1) $\mathsf{rank}(AB - CD) \le \mathsf{rank}(A - C) + \mathsf{rank}(B - D)$.

(2) $\mathsf{rank}(ABC) \ge \mathsf{rank}(AB) + \mathsf{rank}(BC) - \mathsf{rank}(B)$.

在本节的最后，我们再来进一步阐述一下矩阵$A$与$A^T$的几个基本空间之间的关系.

**定义 3.11.** 设$W$是欧氏空间$\mathbb{R}^n$的子空间. 我们定义$W$的正交补为

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0\,(\forall \mathbf{w} \in W)\}.$$

**定理 3.64.** 设$W$是$\mathbb{R}^n$的子空间.

*(1)* $W^\perp$是$\mathbb{R}^n$的子空间.

*(2)* $W \cap W^\perp = \{\mathbf{0}\}$.

*(3)* $W + W^\perp = \mathbb{R}^n$.

*(4)* $(W^\perp)^\perp = W$.

证明. (1) 我们需要验证$W^T$对线性运算封闭. 设$\mathbf{v}_1, \mathbf{v}_2 \in W^\perp$，而$k_1, k_2 \in \mathbb{R}$. 根据正交补的定义，知

$$\mathbf{v}_i \cdot \mathbf{w} = 0, \quad (\forall \mathbf{w} \in W), \quad (i = 1, 2).$$

于是有

$$(k_1\mathbf{v}_1 + k_2\mathbf{v}_2) \cdot \mathbf{w} = k_1\mathbf{v}_1 \cdot \mathbf{w} + k_2\mathbf{v}_2 \cdot \mathbf{w} = 0, \quad (\forall \mathbf{w} \in W).$$

从而$W^T$构成$\mathbb{R}^n$的子空间.

(2) 由于$W$与$W^T$均为$\mathbb{R}^n$的子空间，故$W \cap W^\perp \supseteq \{\mathbf{0}\}$. 反之，设$\mathbf{u} \in W \cap W^\perp$，则有

$$\|\mathbf{w}\|^2 = \mathbf{w} \cdot \mathbf{w} = 0.$$

在上式的$\mathbf{w} \cdot \mathbf{w}$这一项里，我们把前一个$\mathbf{w}$看成$W^\perp$中的向量，而后一个看成$W$中的向量，所以按照正交补的定义得内积为零. 这时，我们得$\|\mathbf{w}\| = 0$，进而$\mathbf{w} = \mathbf{0}$.

(3) 不妨设$\dim(W) = r$，其中$r \leq n$. 设$W$的一组基为$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$. 这时，

$$W^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{w} = 0,\ (\forall \mathbf{w} \in W)\} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{w}_i = 0,\ (1 \leq i \leq r)\}.$$

将上式右端的$\mathbf{w}_i$与$\mathbf{x}$均写为$n \times 1$矩阵，则$W^\perp$是下述其次线性方程组的解空间：

$$\begin{bmatrix} \mathbf{w}_1^T \\ \hdashline \mathbf{w}_2^T \\ \hdashline \vdots \\ \hdashline \mathbf{w}_r^T \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

该系数矩阵（记为$A$）是$r \times n$的，由$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$线性无关知$\mathsf{rank}(A) = r$. 利用秩与零化度的关系，得

$$\dim(W^\perp) = \mathsf{nullity}(A) = n - \mathsf{rank}(A) = r.$$

再结合维数公式与(2)，便有

$$\dim(W + W^\perp) = \dim(W) + \dim(W^\perp) - \dim(W \cap W^\perp) = r + (n - r) - 0 = n.$$

又$W + W^\perp \subseteq \mathbb{R}^n$. 由定理3.53，得$W + W^\perp \subseteq \mathbb{R}^n$.

(4) 由正交补的定义，不难推出$W \subseteq (W^\perp)^\perp$. 又由(3)知

$$\dim(W) + \dim(W^\perp) = n, \quad \dim(W^\perp) + \dim\left((W^\perp)^\perp\right) = n.$$

因此$\dim(W) = \dim\left((W^\perp)^\perp\right)$. 再次利用定理3.53得$(W^\perp)^\perp = W$. □

在上述定理的证明过程中，我们已经将向量的正交性转化成了线性方程组的解与系数矩阵中的行向量之间的关系. 用类似的方法可以证明下述定理.

**定理 3.65.** *设$A \in M_{m\times n}(\mathbb{R})$. 将其行、列分别看成$\mathbb{R}^n$、$\mathbb{R}^m$中的向量.*

*(1) $\mathsf{Null}(A)^\perp = Row(A)$.*

*(2) $\mathsf{Null}(A^T)^\perp = Col(A)$.*

**定理 3.66.** *设$A \in M_n$，则以下命题等价.*

*(i) 矩阵$A$可逆.*

*(ii) 齐次线性方程组$A\mathbf{x} = \mathbf{0}$仅有平凡解.*

*(iii) 矩阵$A$的简约阶梯型为$I_n$.*

*(iv) 矩阵$A$可写成初等矩阵的成绩.*

*(v) 对任意$\mathbf{b} \in M_{n\times 1}(\mathbb{F})$，线性方程组$A\mathbf{x} = \mathbf{b}$是相容的.*

*(vi) 对任意$\mathbf{b} \in M_{n\times 1}(\mathbb{F})$，线性方程组$A\mathbf{x} = \mathbf{b}$有唯一解.*

*(vii) 行列式$\mathsf{det}(A) \neq 0$.*

*(viii) 矩阵$A$的列向量线性无关.*

*(ix) 矩阵$A$的行向量线性无关.*

*(x) 矩阵$A$的列向量张成$\mathbb{R}^n$.*

*(xi) 矩阵$A$的行向量张成$\mathbb{R}^n$.*

*(xii) 矩阵$A$的列向量构成$\mathbb{R}^n$的一组基.*

*(xiii) 矩阵$A$的行向量线构成$\mathbb{R}^n$的一组基.*

*(xiv) $\mathsf{rank}(A) = n$.*

*(xv) $\mathsf{nullity}(A) = 0$.*

*(xvi) $\mathsf{Null}(A)^\perp = \mathbb{R}^n$.*

*(xvii) $Row(A)^\perp = \{\mathbf{0}\}$.*

## 3.6 基的变换

当一个观测者观察空间中的一样事物时，可以建立坐标系（即找一组基）进行记录. 然而，如果有两个观测者，他们使用的坐标系（即两组基）往往是不一样的. 当他们观察同一件事物时，得到的两组坐标之间应当有明确的关系，这是我们在这一节中将探究的内容.

回顾一下，设$V$是$n$维空间，而$S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$是$V$的一组基. 则$V$中的任何一个向量$\mathbf{v}$均可唯一地表为

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n.$$

我们将$\mathbf{v}$在基$S$下的坐标向量或坐标矩阵记为

$$(\mathbf{v})_S = (c_1, c_2, \ldots, c_n), \qquad [\mathbf{v}]_S = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

现在我们用二维的情况来进行一下分析. 设$B = \{\mathbf{v}_1, \mathbf{v}_2\}$与$B' = \{\mathbf{v}_1', \mathbf{v}_2'\}$是二维空间$V$的两组基. 不妨假设

$$\mathbf{v}_1' = a\mathbf{v}_1 + b\mathbf{v}_2,$$
$$\mathbf{v}_2' = c\mathbf{v}_1 + d\mathbf{v}_2.$$

设$\mathbf{u}$是$V$中的任意向量，下面我们来寻找$[\mathbf{u}]_B$与$[\mathbf{u}]_{B'}$之间的关系. 不妨设

$$\mathbf{u} = x\mathbf{v}_1 + y\mathbf{v}_2 = x'\mathbf{v}_1' + y'\mathbf{v}_2'.$$

将$\mathbf{v}_1', \mathbf{v}_2'$关于$\mathbf{v}_1, \mathbf{v}_2$的线性组合表达式代入，得

$$\mathbf{u} = x'(a\mathbf{v}_1 + b\mathbf{v}_2) + y'(c\mathbf{v}_1 + d\mathbf{v}_2) = (ax' + cy')\mathbf{v}_1 + (bx' + dy')\mathbf{v}_2.$$

由于$\mathbf{u}$在基$B'$下的坐标是唯一得，因此

$$\begin{cases} x = ax' + cy', \\ y = bx' + dy'. \end{cases}$$

这就是两组坐标之间的转换关系. 它的矩阵形式为

$$\begin{bmatrix} x \\ y \end{bmatrix} = \left[\begin{array}{c:c} a & c \\ b & d \end{array}\right] \begin{bmatrix} x' \\ y' \end{bmatrix}$$

或即

$$[\mathbf{u}]_B = \begin{bmatrix} [\mathbf{v}_1']_B & [\mathbf{v}_2']_B \end{bmatrix} [\mathbf{u}]_{B'}.$$

将上述讨论中的2维更换为$n$维，过程都是类似的. 我们直接给出下述一般情况下的定义与定理.

**定义 3.12.** 设$B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$与$B' = \{\mathbf{v}_1', \ldots, \mathbf{v}_n'\}$是向量空间$V$的两组基. 我们定义由基$B'$到$B$的过渡矩阵（或转移矩阵）为

$$P_{B \leftarrow B'} = \left[\begin{array}{c:c:c:c} [\mathbf{v}_1']_B & [\mathbf{v}_2']_B & \ldots & [\mathbf{v}_n']_B \end{array}\right]$$

**命题 3.3.** 设$B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$与$B' = \{\mathbf{v}_1', \ldots, \mathbf{v}_n'\}$是向量空间$V$的两组基. 则对$V$中的任意向量$\mathbf{x}$，均有

$$[\mathbf{x}]_B = P_{B \leftarrow B'} [\mathbf{x}]_{B'}.$$

**例 3.67.** 考虑$\mathbb{R}^2$中的两组基$B' = \{\mathbf{u}_1', \mathbf{u}_2'\}$与$B'' = \{\mathbf{u}_1'', \mathbf{u}_2''\}$，其中

$$\mathbf{u}_1' = (1, 0), \mathbf{u}_2' = (1, 1), \quad \mathbf{u}_1'' = (1, 1), \mathbf{u}_2'' = (2, 1).$$

*(1) 求由基$B''$到基$B'$的过渡矩阵$P_{B' \leftarrow B''}$.*
*(2) 求由基$B'$到基$B''$的过渡矩阵$P_{B'' \leftarrow B'}$.*
*(3) 设$(\mathbf{x})_{B'} = (3, 2)$，求$(\mathbf{x})_{B''}$.*

解. 在这里，我们展示(2)、(3)的计算过程.

(2) 设$k_1\mathbf{u}_1'' + k_2\mathbf{u}_2'' = \mathbf{u}_1'$，即

$$\left[\begin{array}{c:c} 1 & 2 \\ 1 & 1 \end{array}\right] \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

得唯一解$[\mathbf{u}_1']_{B''} = [-1\ 1]^T$. 类似地，解

$$\left[\begin{array}{c:c} 1 & 2 \\ 1 & 1 \end{array}\right] \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

得唯一解$[\mathbf{u}_2']_{B''} = [-1\ 0]^T$. 故

$$P_{B'' \leftarrow B'} = \left[\begin{array}{c:c} [\mathbf{u}_1']_{B''} & [\mathbf{u}_2']_{B''} \end{array}\right] = \left[\begin{array}{c:c} -1 & 1 \\ 1 & 0 \end{array}\right].$$

(3) 注意到$[\mathbf{x}]_{B'} = [3\ 2]^T$，我们有

$$[\mathbf{x}]_{B''} = P_{B'' \leftarrow B'} [\mathbf{x}]_{B'} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

即$(\mathbf{x})_{B''} = (-1, 3)$. □

**例 3.68.** 考虑$P_3$中的两组基$B=\{p_0,p_1,p_2\}$与$B'=\{q_0,q_1,q_2\}$，其中

$$p_0(x)=1,\, p_1(x)=x,\, p_2(x)=x^2,$$
$$q_0(x)=1,\, q_1(x)=x+c,\, q_2(x)=(x+c)^2.$$

求过度矩阵$P_{B'\leftarrow B}$.

*解.* 计算可得

$$\begin{aligned}1&=1+0(x+c)+0(x+c)^2,\\ x&=(-c)1+1(x+c)+0(x+c)^2,\\ x^2&=(c^2)1+(-2c)(x+c)+1(x+c)^2.\end{aligned}$$

故

$$P_{B'\leftarrow B}=\left[\begin{array}{c:c:c}1&-c&c^2\\0&1&-2c\\0&0&1\end{array}\right].$$

□

线性空间中更多组基之间的关系如下.

**定理 3.69.** 设$B,B',B''$为向量空间$V$中的三组基. 则

$$P_{B\leftarrow B''}\;=\;P_{B\leftarrow B'}\,P_{B'\leftarrow B''}$$

*证明.* 对任意$\mathbf{v}\in V$，有

$$P_{B\leftarrow B'}\,P_{B'\leftarrow B''}[\mathbf{x}]_{B''}=P_{B\leftarrow B'}[\mathbf{x}]_{B'}=[\mathbf{x}]_B=P_{B\leftarrow B''}[\mathbf{x}]_{B''}.$$

当$\mathbf{v}$取遍$V$中所有向量时,列矩阵$[\mathbf{x}]_{B''}$取遍对应维数的欧氏空间中的所有向量,因此$P_{B\leftarrow B''}\;=\;P_{B\leftarrow B'}\,P_{B'\leftarrow B''}$.

□

**推论 3.70.** 设$B,B'$是线性空间$V$中的两组向量. 则过度矩阵$P_{B'\to B}$可逆，且

$$P^{-1}_{B\leftarrow B'}=P_{B'\leftarrow B}.$$

*证明.* 不难验证对同一组基$B$，有$P_{B\leftarrow B}=I$. 由前一定理可得该推论中的结论. □

*证明.* 不难验证对同一组基$B$，有$P_{B\leftarrow B}=I$. 由前一定理 □

由$P_{B\leftarrow B''}\;=\;P_{B\leftarrow B'}\,P_{B'\leftarrow B''}$可得$P_{B'\leftarrow B''}=P^{-1}_{B\leftarrow B'}P_{B\leftarrow B''}$. 若$B$是空间的标准基，而$B'$与$B''$中的向量均由在标准基下的坐标给出，则$P_{B'\leftarrow B''}$可由初等行变换

$$\left[\begin{array}{c:c}P_{B\leftarrow B'}&P_{B\leftarrow B''}\end{array}\right]\xrightarrow{\text{row}}\left[\begin{array}{c:c}I&P^{-1}_{B\leftarrow B'}P_{B\leftarrow B''}\end{array}\right]$$

进行求解.

**例 3.71.** 考虑$\mathbb{R}^2$中的两组基$B'=\{\mathbf{u}_1',\mathbf{u}_2'\}$与$B''=\{\mathbf{u}_1'',\mathbf{u}_2''\}$，其中

$$\mathbf{u}_1'=(1,0),\mathbf{u}_2'=(1,1),\quad \mathbf{u}_1''=(1,1),\mathbf{u}_2''=(2,1).$$

*(1)* 求由基$B''$到基$B'$的过渡矩阵$P_{B'\leftarrow B''}$.
*(2)* 求由基$B'$到基$B''$的过渡矩阵$P_{B''\leftarrow B'}$.
*(3)* 设$(\mathbf{x})_{B'}=(3,2)$，求$(\mathbf{x})_{B''}$.

解. 在这里，我们展示(1)的计算过程. 取标准基$B=\{\mathbf{e}_1,\mathbf{e}_2\}$，则

$$P_{B\leftarrow B'}=\left[\,[\mathbf{u}_1']_B \,\vdots\, [\mathbf{u}_2']_B\,\right]=\left[\begin{array}{c:c}1&1\\0&1\end{array}\right].$$

$$P_{B\leftarrow B''}=\left[\,[\mathbf{u}_1'']_B \,\vdots\, [\mathbf{u}_2'']_B\,\right]=\left[\begin{array}{c:c}1&2\\1&1\end{array}\right].$$

我们有$P_{B''\leftarrow B'}=P_{B\leftarrow B''}^{-1}P_{B\leftarrow B'}$，利用行变换

$$[\,P_{B\leftarrow B''} \,\vdots\, P_{B\leftarrow B'}\,]=\left[\begin{array}{cc:cc}1&2&1&1\\1&1&0&1\end{array}\right]\longrightarrow\left[\begin{array}{cc:cc}1&0&0&1\\-1&1&1&0\end{array}\right],$$

得

$$P_{B''\leftarrow B'}=\begin{bmatrix}-1&1\\1&0\end{bmatrix}.$$

□

# 第4章 线性变换

## 4.1 线性变换的定义

在第二章的末尾，我们已经介绍了矩阵变换：设$A \in M_{m\times n}(\mathbb{R})$, 则映射$T_A : \mathbb{R}^n \to \mathbb{R}^m$为

$$T_A(\mathbf{x}) = A\mathbf{x}, \quad (\forall \mathbf{x} \in \mathbb{R}^n).$$

它满足性质

$$T_A(\mathbf{u}+\mathbf{v}) = T_A(\mathbf{u}) + T_A(\mathbf{v}), \quad T_A(k\mathbf{u}) = kT_A(\mathbf{u}),$$

其中$\mathbf{u}, \mathbf{u} \in \mathbb{R}^n$，而$k$为实数.

现在，我们在抽象定义的线性空间之间定义线性变换.

**定义 4.1.** 设$V, W$均为数域$\mathbb{F}$上的线性空间，若映射$T : V \to W$满足

*(1)* 对任意$\mathbf{u}, \mathbf{v} \in V$，有 $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;

*(2)* 对任意$\mathbf{u} \in V$以及$k \in \mathbb{F}$，有 $T(k\mathbf{u}) = kT(\mathbf{u})$,

则我们称$T$为从$V$到$W$的线性变换. 当$V = W$时，我们也称$T$为$V$上的线性算子/算符.

值得一提的是，上述定义中的两个条件还有另一种等价的写法：对任意$\mathbf{u}, \mathbf{v} \in V$，以及任意$k, \ell \in \mathbb{F}$，均有

$$T(k\mathbf{u} + \ell\mathbf{v}) = kT(\mathbf{u}) + \ell T(\mathbf{v}).$$

更近一步，也等价于下述命题：对任意$r \geq 2$，任意$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r \in V$，以及任意$k_1, k_2, \ldots, k_r \in \mathbb{F}$，均有

$$T(k_1\mathbf{u}_1 + \ldots k_n\mathbf{u}_n) = k_1T(\mathbf{u}_1) + \ldots + k_nT(\mathbf{u}_n)$$

即映射$T$保持了线性空间$V$和$W$中的线性结构：“先在定义域空间中作线性组合，再作用映射$T$”，与“先作用映射$T$，再在值域空间中作线性组合”，它们的效果是相同的.

当$T : V \to W$时为线性映射时，我们取$\mathbf{u} \in V$，则有

$$T(\mathbf{0}) = T\big((1+(-1))\mathbf{u}\big) = T(\mathbf{u}) + (-1)T(\mathbf{u}) = \mathbf{0}.$$

因此，线性映射一定把定义域空间中的零向量，映为值域空间中的零向量.

**例 4.1.** 设$A \in M_{m\times n}(\mathbb{C})$，则矩阵变换$T_A : \mathbb{C}^n \to \mathbb{C}^m, \mathbf{x} \mapsto A\mathbf{x}$是线性变换.

**例 4.2.** 设$V$与$W$为$\mathbb{F}$上的线性空间，映射$0 : V \to W$为

$$0(\mathbf{u}) = \mathbf{0}, \quad (\forall \mathbf{u} \in \mathbf{V}).$$

它是一个线性变换，被称为零变换.

**例 4.3.** 设$V$为实线性空间，设$k \in \mathbb{R}$. 定义映射$T : V \to V$为

$$T(\mathbf{u}) = k\mathbf{u}, \quad (\forall \mathbf{u} \in V),$$

则它是一个线性算子.

当$k=0$时，它是零算子；

当$k=1$时，它被称为恒等算子，常被记为$I$.

当$0<k<1$时，它被称为$V$上以$k$为系数的收缩.

当$k>1$时，它被称为$V$上以$k$为系数的膨胀.

**例 4.4.** 对于次数不超过$n-1$次的实系数多项式全体组成的实线性空间$P_n$，以下两个映射均为线性变换：

◇ 映射$M: P_n \to P_{n+1}$由$M(p(x)) = xp(x)$给出；

◇ 映射$D: P_{n+1} \to P_n$由$D(p(x)) = p'(x)$给出.

**例 4.5.** 设$C(a,b)$为$(a,b)$上的实值连续函数全体，而$C^\infty(a,b)$为$(a,b)$任意次可导的函数全体. 以下两个映射均为线性算子.

◇ 映射$D: C^\infty(a,b) \to C^\infty(a,b)$由$D(f(x)) = f'(x)$给出；

◇ 映射$S: C(a,b) \to C(a,b)$定义为

$$S(f(x)) = \int_a^x f(t)dt, \quad (a<x<b).$$

**例 4.6.** 设$V$为所有实柯西列全体所组成的实线性空间. 定义$\mathit{Lim}: V \to \mathbb{R}$为

$$\mathit{Lim}\big(\{a_n\}\big) = \lim_{n\to\infty} a_n, \quad \big(\forall \{a_n\} \in V\big).$$

则$\mathit{Lim}$是线性算子.

**例 4.7.** 定义映射$T_1: \mathbb{R}^3 \to \mathbb{R}^3$与$T_2: \mathbb{R}^3 \to \mathbb{R}^2$为

$$T_1(x,y,z) = (x,y,0), \quad T_2(x,y,z) = (x,y).$$

它们都是线性变换，且效果均提取出了$\mathbb{R}^3$中处于$xOy$平面上的信息.

**例 4.8.** 上述例子是有限个坐标的情形，现在我们来看看可数个坐标的情形. 设

$$\mathbb{R}^{\mathbb{N}} = \big\{(x_1, x_2, \ldots, x_n, x_{n+1}, \ldots) : x_i \in \mathbb{R}\,(i=1,2,\ldots)\big\}$$

是实数列全体. 给定指标$i_1, i_2, \ldots, i_r \in \mathbb{N}$, 定义$T: \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^r$向给定分量的投影为

$$T(x_1, x_2, \ldots, x_n, \ldots) = (x_{i_1}, x_{i_2}, \ldots, x_{i_r}).$$

可以验证它是一个线性变换.

例如，取$r=3$，以及$i_1=1, i_2=2, i_3=3$，则有

$$T(1,2,3,4,\ldots,n,\ldots) = (1,2,3),$$

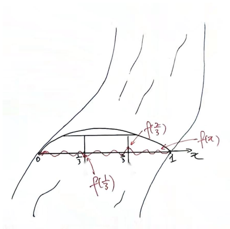
$$T(1,-1,1,-1,\ldots,(-1)^n,\ldots) = (1,-1,1).$$

**例 4.9.** 接下来，我们再看看不可数个坐标的情形. 设$V$为实线性空间$F(a,b)$的一个子空间. 设$r\in\mathbb{N}$, 而$x_1, x_2, \ldots, x_r \in (a,b)$. 定义$V$在点$x_1, x_2, \ldots, x_r$处的取值变换为$T: V \to \mathbb{R}^r$：

$$T(f) = (f(x_1), f(x_2), \ldots, f(x_r)) \quad (\forall f \in V).$$

对任意$f, g \in V$，以及任意$k, l \in \mathbb{R}$，有

$$\begin{aligned} T(kf+lg) \quad &= \big((kf+lg)(x_1), (kf+lg)(x_2), \ldots, (kf+lg)(x_r)\big) \\ &= \big(kf(x_1)+lg(x_1), kf(x_2)+lg(x_2), \ldots, kf(x_r)+lg(x_r),\big) \\ &= k\big(f(x_1), f(x_2), \ldots, f(x_r)\big) + l\big(g(x_1), g(x_2), \ldots, g(x_r)\big) = kT(f)+lT(g). \end{aligned}$$



上式中，从左到右的等号依次用到了$T$的定义、定义域$V$中的运算、值域$\mathbb{R}^r$中的运算与$T$的定义. 这时，我们证明了$T$是一个线性变换.

例如，一座桥横跨一条小河，在两岸处的位置我们分别记为0与1，它在河中有两个桥墩，位置对应$x_1 = 1/3$与$x_2 = 2/3$. 水面的高度是$(0,1)$上的函数. 为了获取桥墩处的水位高度，我们取$r=2$以及点$x_1, x_2$构造取值变换. 于是，

$$T\big(0.2\cos(99\pi x)\big) = \big(0.2\cos(33\pi), 0.2\cos(66\pi)\big) = (-0.2, 0.2).$$

**例 4.10.** 设$\mathbf{v}_0 \in \mathbb{R}^n$为给定向量，定义$T:\mathbb{R}^n \to \mathbb{R}$为

$$T(\mathbf{u}) = \mathbf{u}\cdot\mathbf{v}_0, \quad (\forall \mathbf{u} \in \mathbb{R}^n).$$

则$T$为线性变换.

**例 4.11.** 以下两个映射均为线性变换.

◇ 映射$T: M_{m\times n} \to M_{n\times m}$, $A \mapsto A^T$.

◇ 映射$T: M_n(\mathbb{F}) \to \mathbb{F}$, $A \mapsto \mathit{tr}(A)$.

**例 4.12.** 给定$B, C \in M_n(\mathbb{F})$，则以下$M_n(\mathbb{F})$上的算子均为线性的.

◇ $T_{B,C}(A) = BAC \ (A \in M_n(\mathbb{F}))$.

◇ $ad_B(A) = BA - AB \ (A \in M_n(\mathbb{F}))$.

Remark: We can write .

**例 4.13.** 下列映射不是线性变换：

◇ 映射$T: M_n(\mathbb{R}) \to \mathbb{R}$, $A \mapsto \mathit{det}(A)$. 这里$n \geq 2$.

◇ 映射$T: \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{x} \mapsto \mathbf{x} + \mathbf{x}_0$. 这里$\mathbf{x}_0 \neq \mathbf{0}$.

对于前一个映射，注意到

$$T(2I) = 2^n \neq 2 = 2T(I),$$

因此该映射不是线性的.

对于后一个映射，因$T(\mathbf{0}) = \mathbf{x}_0 \neq \mathbf{0}$，可判断其不是线性变换.

注意到映射之间可以自然定义加法与数乘：设$T, T_1, T_2$均为$\mathbb{F}$上的线性空间$V$到$W$的线性变换，而$c \in \mathbb{F}$，则线性变换的加法和数乘定义为

$$(T_1 + T_2)(\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v}), \quad (cT)(\mathbf{v}) = c\,T(\mathbf{v}), \quad (\mathbf{v} \in V).$$

可以证明上述定义的$T_1 + T_2$和$cT$仍为线性变换（证明留给读者）. 这时，我们记$L(V,W)$为从$V$到$W$的所有线性变换全体所组成的集合，则在上述加法与数乘下，$L(V,W)$也是$\mathbb{F}$上的线性空间.

比如，例(4.12)中的映射满足$ad_B = T_{B,I} - T_{I,B}$. 而$L(\mathbb{R}^n, \mathbb{R}^m) = M_{m\times n}(\mathbb{R})$.

## 4.2 线性变换的矩阵

这一节的主要问题是：如何有效地描述一个线性映射？对于线性空间来说，只要给了一组基，那么空间所有的向量可以自然地由基表示. 因此如果知道了线性映射在基向量上是如何映的，就能推断出其在所有向量上的映射结果.

设$V$, $W$为数域$\mathbb{F}$上的线性空间，而$T: V \to W$是一个线性变换. 设$B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$是定义域空间$V$的一组基，而$\widetilde{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$是值域空间$W$的一组基. 这时，对任意向量$\mathbf{x} \in V$，存在$x_1, x_2, \ldots, x_n \in \mathbb{F}$使得

$$\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_n\mathbf{v}_n = \sum_{j=1}^{n} x_j\mathbf{v}_j.$$

由于$T$保持线性结构，因此

$$T(\mathbf{x}) = x_1T(\mathbf{v}_1) + x_2T(\mathbf{v}_2) + \ldots + x_nT(\mathbf{v}_n) = \sum_{j=1}^{n} x_jT(\mathbf{v}_j).$$

另外，由于$T(\mathbf{v}_j) \in W$，它们可以由$\widetilde{B}$中的向量线性表出，记为

$$T(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + a_{2j}\mathbf{w}_2 + \ldots + a_{mj}\mathbf{w}_m = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i, \quad (1 \le j \le n),$$

其中$a_{ij} \in \mathbb{F}$ $(1 \le i \le m, 1 \le j \le n)$. 最终，我们得到

$$T(\mathbf{x}) = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij}x_j \right) \mathbf{w}_i.$$

记$T(\mathbf{x}) = \sum_{i=1}^{m} y_i\mathbf{w}_i$，则$y_i = \sum_{j=1}^{n} a_{ij}x_j$ $(1 \le i \le m)$.

现在，我们将上述式子都更换为对应的矩阵形式，则有

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

另外，注意到

$$[\mathbf{x}]_B = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad [T(\mathbf{v}_j)]_{\tilde{B}} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}, \qquad [T(x)]_{\tilde{B}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

我们得到关系式

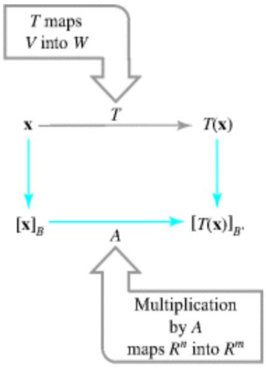
$$[T(x)]_{\tilde{B}} = \left[ [T(\mathbf{v}_1)]_{\tilde{B}} \,\vdots\, [T(\mathbf{v}_2)]_{\tilde{B}} \,\vdots\, \ldots \,\vdots\, [T(\mathbf{v}_n)]_{\tilde{B}} \right] [\mathbf{x}]_B.$$

**定义 4.2.** 设$V$, $W$均为有限维线性空间，而$T: V \to W$为线性变换. 设$B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$与$\widetilde{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$分别为$V$与$W$的基. 我们定义$T$在基$B$与$\widetilde{B}$下的矩阵为

$$[T]_{\widetilde{B},B} = \left[ \; [T(\mathbf{v}_1)]_{\widetilde{B}} \;\vdots\; [T(\mathbf{v}_2)]_{\widetilde{B}} \;\vdots\; \ldots \;\vdots\; [T(\mathbf{v}_n)]_{\widetilde{B}} \; \right].$$

在这一定义下，我们将抽象空间$V$到$W$上的抽象线性映射$T$，完全转化为欧氏空间$\mathbb{F}^n$到$\mathbb{F}^m$上的矩阵变换：

$$[T(\mathbf{x})]_{\widetilde{B}} = [T]_{\widetilde{B},B}\, [\mathbf{x}]_B. \tag{4.1}$$




我们对记号作一些解释，在$T$外侧出现的$[T]$表示是变换$T$对应的矩阵，下标$[]_{\widetilde{B},B}$对应矩阵的阶$[]_{m\times n}$. 特别地，行标对应值域空间的基向量，而列标对应定义域空间的基向量，这与矩阵变换是一致的. 在式(4.1)中的矩阵乘法运算里，我们将$[T]_{\widetilde{B},B}$的列标与$[\mathbf{x}]_B$的行标进行对应求和，最后只留下行标$\widetilde{B}$的信息，因而得到$[T(\mathbf{x})]_{\widetilde{B}}$.

另外，当$W=V$时，我们往往只在空间中取一组基$B$，因此可将$T$在基$B$的矩阵记为

$$[T]_B=\left[\begin{array}{c:c:c:c} [T(\mathbf{v}_1)]_B & [T(\mathbf{v}_2)]_B & \dots & [T(\mathbf{v}_n)]_B \end{array}\right].$$

在不引起混淆的情况下，我们可以把线性变换$T$在给定基下的矩阵简记为$[T]$.

**例 4.14.** *设$A\in M_{m\times n}(\mathbb{F})$. 记$\mathbb{F}^n$与$\mathbb{F}^m$的标准基分别为$S$与$\widetilde{S}$. 不难验证矩阵变换$T_A:\mathbb{F}^n\to\mathbb{F}^m$, $\mathbf{x}\to A\mathbf{x}$在基$S$与$\widetilde{S}$下的矩阵就是$A$本身，即$[T_A]_{\widetilde{S},S}=A$.*

**例 4.15.** *设$V,W$均为有限维线性空间空间.*

*(1) 零变换$0:V\to W$在任何基下的矩阵均为零矩阵.*

*(2) 恒等算子$I:V\to V$在任何基下的矩阵均为恒等矩阵.*

**例 4.16.** *(1) 设$T:P_1\to P_2$的定义为*

$$T(p(x))=xp(3x-5).$$

*求$T$在基$B=\{1,x\}$与$\widetilde{B}=\{1,x,x^2\}$下的矩阵$[T]_{\widetilde{B},B}$.*

*(2) 计算$T(1+2x)$ (a) 直接计算; (b) 利用矩阵$[T]_{\widetilde{B},B}$进行计算.*

解. (1) 记$p_1(x)=1$, $p_2(x)=x$，我们有

$$\begin{aligned}T(p_1(x))&=xp_1(3x-5)=x\,1=x,\\ T(p_1(x))&=xp_2(3x-5)=x(3x-5)=-5x+3x^2.\end{aligned}$$

故

$$[T]_{\widetilde{B},B}=\left[\begin{array}{c:c}0&0\\1&-5\\0&3\end{array}\right].$$

(2) 记$h(x)=1+2x$. 直接计算，得

$$T(h(x))=xh(3x-5)=x\big(1+2(3x-5)\big)=-9x+6x^2.$$

注意到$[h(x)]_B=\begin{bmatrix}1&2\end{bmatrix}^T$. 利用矩阵$[T]_{\widetilde{B},B}$进行计算得

$$[T(h(x))]_B=[T]_{\widetilde{B},B}[h(x)]_B=\begin{bmatrix}0&0\\1&-5\\0&3\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}0\\-9\\6\end{bmatrix}.$$

□

**例 4.17.** 设$V$是$F(-\infty,+\infty)$的子空间，由函数族

$$B=\{e^{2x},xe^{2x},x^2e^{2x},x^3e^{2x}\}.$$

张成. 微分算子$D=\frac{d}{dx}$是$V$上的线性算子. 求算子$D$在基$B$下的矩阵.

解. 计算可得$(e^{2x})'=2e^{2x}$，而对$2\le i\le 4$有

$$(x^{i-1}e^{2x})'=(i-1)x^{i-2}e^{2x}+2x^{i-1}e^{2x}.$$

故

$$[D]_B=\begin{bmatrix}2&1&0&0\\0&2&2&0\\0&0&2&3\\0&0&0&2\end{bmatrix}.$$

□

**例 4.18.** 设$B=\begin{bmatrix}a&b\\c&d\end{bmatrix}$为给定矩阵. 线性变换$ad:M_2(\mathbb{F})\to M_2(\mathbb{F})$的定义由

$$ad(A)=BA-AB,\quad (A\in M_2(\mathbb{F}))$$

给出. 求$ad$在$M_2(\mathbb{F})$的标准基下的矩阵.

解. 标准基为$S=\{E_{11},E_{12},E_{21},E_{22}\}$. 计算得

$$\begin{aligned}
ad(E_{11})&=\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}-\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\begin{bmatrix}0&-b\\c&0\end{bmatrix}=0E_{11}-bE_{12}+cE_{21}+0E_{22},\\
ad(E_{12})&=\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}0&1\\0&0\end{bmatrix}-\begin{bmatrix}0&1\\0&0\end{bmatrix}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\begin{bmatrix}-c&a-d\\0&c\end{bmatrix}=-cE_{11}+(a-d)E_{12}+0E_{21}+cE_{22},\\
ad(E_{11})&=\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}0&0\\1&0\end{bmatrix}-\begin{bmatrix}0&0\\1&0\end{bmatrix}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\begin{bmatrix}b&0\\d-a&-b\end{bmatrix}=dE_{11}+0E_{12}+(b-a)E_{21}-bE_{22},\\
ad(E_{11})&=\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}0&0\\0&1\end{bmatrix}-\begin{bmatrix}0&0\\0&1\end{bmatrix}\begin{bmatrix}a&b\\c&d\end{bmatrix}=\begin{bmatrix}0&b\\-c&0\end{bmatrix}=0E_{11}+bE_{12}-cE_{21}+0E_{22}.
\end{aligned}$$

故

$$[ad]_S=\begin{bmatrix}0&-c&d&0\\-b&a-d&0&b\\c&0&b-a&-c\\0&c&-b&0\end{bmatrix}.$$

□

## 4.3 线性变换的性质

每个矩阵都有它的零空间、行空间和列空间. 既然有限维空间之间的线性变换都能写成矩阵，那么什么是与这些空间有关的结构呢？

**定义 4.3.** 线性变换$T:V\to W$的核与像分别定义为

$$\mathit{Ker}(T)=\{\mathbf{v}\in V:\,T(\mathbf{v})=\mathbf{0}\},$$

以及

$$\mathit{Ran}(T)=\{T(\mathbf{v}):\,\mathbf{v}\in V\}.$$

**例 4.19.** *设$A \in M_{m\times n}(\mathbb{F})$. 则对矩阵变换$T_A : \mathbb{F}^n \to \mathbb{F}^m$，有*

$$\mathit{Ker}(T_A) = \{\mathbf{x} \in \mathbb{F}^n :\ A\mathbf{x} = \mathbf{0}\} = \mathit{Null}(A),$$

$$\mathit{Ran}(T_A) = \{A\mathbf{x} :\ \mathbf{x} \in \mathbb{F}^n\} = \mathit{Col}(A).$$

**定理 4.20.** *设$T : V \to W$为线性变换，则*

*(1) 核$\mathit{Ker}(T)$是$V$的子空间;*

*(2) 像$\mathit{Ran}(T)$是$W$的子空间.*

*证明.* (1) 设$\mathbf{v}_1, \mathbf{v}_2$为$\mathsf{Ker}(T)$中的任意向量，而$k_1, k_2$为任意标量. 则

$$T(\mathbf{v}_1) = \mathbf{0}, \quad T(\mathbf{v}_2) = \mathbf{0}.$$

利用$T$保持线性结构，可得

$$T(k_1\mathbf{v}_1 + k_2\mathbf{v}_2) = k_1T(\mathbf{v}_1) + k_2T(\mathbf{v}_2) = k_1\mathbf{0} + k_2\mathbf{0} = \mathbf{0}.$$

从而$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 \in \mathsf{Ker}(T)$. 故$\mathsf{Ker}(T)$是$V$的子空间.

(2) 设$\mathbf{w}_1, \mathbf{w}_2$为$\mathsf{Ran}(T)$中任意向量，而$l_1, l_2$为任意标量. 则存在$\mathbf{u}_1, \mathbf{u}_2 \in V$，使得

$$T(\mathbf{u}_1) = \mathbf{v}_1, \quad T(\mathbf{u}_2) = \mathbf{v}_2.$$

由于$V$为线性空间，因此$l_1\mathbf{u}_1 + l_2\mathbf{u}_2 \in V$，且

$$T(l_1\mathbf{u}_1 + l_2\mathbf{u}_2) = l_1T(\mathbf{u}_1) + l_2T(\mathbf{u}_2) = l_1\mathbf{v}_1 + l_2\mathbf{v}_2$$

即$l_1\mathbf{v}_1 + l_2\mathbf{v}_2 \in \mathsf{Ran}(T)$. 故$\mathsf{Ran}(T)$是$W$的子空间. □

**例 4.21.** *求下列线性变换的核与像，并确定它们的维数.*

*(i) 线性空间$V$上的零算子$0$.*

*(ii) 线性空间$V$上的恒等算子$I$.*

*(iii) 线性变换$T : M_n(\mathbb{R}) \to \mathbb{R}$, $A \mapsto \mathit{tr}(A)$. 这里$n \geq 2$.*

*(iv) 线性变换$M : P_n \to P_{n+1}$，由$M(p(x)) = xp(x)$给出.*

*(v) 线性变换$D : P_{n+1} \to P_n$，由$D(p(x)) = p'(x)$给出.*

*解.* (i) 我们有$\mathsf{Ker}(T) = V$, $\mathsf{Ran}(T) = \{0\}$. 此时$\dim \mathsf{Ker}(T) = \dim V$, $\dim \mathsf{Ran}(T) = 0$.

(ii) 我们有$\mathsf{Ker}(T) = \{0\}$, $\mathsf{Ran}(T) = V$. 此时$\dim \mathsf{Ker}(T) = 0$, $\dim \mathsf{Ran}(T) = \dim V$.

(iii) 根据定义有

$$\mathsf{Ker}(T) = \{A = [a_{ij}] \in M_n(\mathbb{R}) :\ a_{11} + a_{22} + \ldots + a_{nn} = 0\}.$$

该空间是以$a_{ij}$ $(1 \leq i, j \leq n)$为变元，以$a_{11} + a_{22} + \ldots + a_{nn} = 0$为方程组成的线性方程组的解空间. 由于变元个数为$n^2$，而系数矩阵的秩为1，因此解空间的维数为$n^2 - 1$，即$\dim \mathsf{Ker}(T) = n^2 - 1$. 事实上，可取一组基为

$$\{E_{ij} : 1 \leq i \neq j \leq n\} \cup \{E_{ii} - E_{nn} : 1 \leq i \leq n-1\}.$$

另一方面，对任意$r \in \mathbb{R}$，取$A_r = \mathsf{diag}(r, 0, \ldots, 0) \in M_n(\mathbb{R})$，则有

$$T(A_r) = \mathsf{tr}(A_r) = r.$$

故$r \in \mathsf{Ran}(T)$. 所以$\mathsf{Ran}(T) = \mathbb{R}$, 而$\dim \mathsf{Ran}(T) = 1$.

(iv) 这里$P_n$指次数不超过$n-1$次的多项式全体,$P_{n+1}$指次数不超过$n$次的多项式全体. 不难得到$\mathsf{Ker}(T) = \{\mathbf{0}\}$，$\mathsf{Ran}(T) = \mathsf{span}\{x, x^2, \ldots, x^n\}$. 从而$\dim \mathsf{Ker}(T) = 0$, $\dim \mathsf{Ran}(T) = n$.

(v) 不难得到$\mathsf{Ker}(T) = \mathsf{span}\{1\}$，$\mathsf{Ran}(T) = P_n$. 从而$\dim \mathsf{Ker}(T) = 1$, $\dim \mathsf{Ran}(T) = n$. □

从上述例子可以看出，线性变换的核空间的维数与像空间的维数相加，都等于定义域空间的维数. 这是因为，将线性变换在给定基下写成矩阵，则核空间与像空间分别对应为列空间与零空间，而它们的维数分别为秩与零化度.

**定义 4.4.** 设$T: V \to W$为线性变换，其中$V, W$均为有限维线性空间. 我们将$T$的秩与零化度分别定义为

$$\textit{rank}(T) = \dim \textit{Ran}(T), \quad \textit{nullity}(T) = \dim \textit{Ker}(T).$$

**定理 4.22.** 设$V, W$为有限维线性空间，而$T: V \to W$为线性变换，则

$$\textit{rank}(T) + \textit{nullity}(T) = \dim(V).$$

*证明.* 在$V$与$W$中分别取基$B$与$\widetilde{B}$,则$T$在基$B$与$\widetilde{B}$下的矩阵为$[T]_{\widetilde{B},B}$,简记为$[T]$. 此时$[T]$的列数等于$\dim V$. 根据矩阵的秩与零化度的关系，我们有

$$\mathsf{rank}(T) + \mathsf{nullity}(T) = \mathsf{rank}([T]) + \mathsf{nullity}([T]) = \dim V.$$

□

现在我们来回顾一下单射、满射、双射的定义. 设映射$f: V \to W$，若$f$将$V$中的不同元素映为不同元素，则称$f$为单射；若$W$中的每个元素都在$f$的像中，则称$f$为满射；若$f$即单又满，则称$f$为双射.

**例 4.23.** 问以下线性变换是否单射？满射？双射？

*(i)* 线性空间$V$上的零算子$0$.

*(ii)* 线性空间$V$上的恒等算子$I$.

*(iii)* 线性变换$T: M_n(\mathbb{R}) \to \mathbb{R}$, $A \mapsto \textit{tr}(A)$. 这里$n \geq 2$.

*(iv)* 线性变换$M: P_n \to P_{n+1}$，由$M(p(x)) = xp(x)$给出.

*(v)* 线性变换$D: P_{n+1} \to P_n$，由$D(p(x)) = p'(x)$给出.

*(vi)* 数列空间$\mathbb{R}^{\mathbb{N}}$上的右平移算子$T_1$和左平移算子$T_2$，它们的定义为

$$T_1(x_1, x_2, \ldots, x_n, \ldots) = (0, x_1, \ldots, x_{n-1}, \ldots),$$

$$T_2(x_1, x_2, \ldots, x_n, \ldots) = (x_2, x_3, \ldots, x_{n+1}, \ldots).$$

*证明.* 前五个例子都关于有限维空间之间的线性变换. 回顾例4.21，我们可以得到下述结论.

(i) 当$V = \{\mathbf{0}\}$时，零算子是单射、满射、双射；当$V \neq \{\mathbf{0}\}$时，零算子非单、非满、非双.

(ii) 在任何情况下，恒等算子都是单射、满射、双射.

(iii) 不难验证，线性变换$T$非单、满、非双.

(iv) 不难验证，线性变换$T$单、非满、非双.

(v) 不难验证，线性变换$T$非单、满、非双.

我们再看看无限维空间之间的线性变换. 对于(vi)，线性算子$T_1$是单、非满、非双的，而线性算子$T_2$是非单、满、非双的. 下面我们给出具体的证明细节.

设$(x_1, x_2, \ldots)$和$(x_1', x_2', \ldots)$是$\mathbb{R}^{\mathbb{N}}$中的两个数列，满足$T_1(x_1, x_2, \ldots) = T_1(x_1', x_2', \ldots)$. 则有

$$(0, x_1, x_2, \ldots) = (0, x_1', x_2', \ldots).$$

于是各个分量相等，即$x_i = x_i'$, $(i = 1, 2, \ldots)$，进而$(x_1, x_2, \ldots) = (x_1', x_2', \ldots)$. 因此$T_1$是单射. 要说明$T_1$不是满射，只需举出反例: 比如，数列$(1, 0, 0, \ldots)$不在$T_1$的像中.

算子$T_2$不是单的，我们有下述反例：

$$T_2(1, 0, 0, \ldots) = (0, 0, 0, \ldots) = T_2(0, 0, 0, \ldots).$$

特别地，我们有$\mathsf{Ker}(T_2) \neq \{(0,0,\ldots)\}$. 另外，对任意值域空间中的数列$(y_1,y_2,\ldots)$，取定义空间中的数列$(0,y_1,y_2,\ldots)$，则有

$$T_2(0,y_1,y_2,\ldots) = (y_1,y_2,\ldots).$$

因此$\mathsf{Ran}(T) = \mathbb{R}^{\mathbb{N}}$，算子$T_2$是满射. □

从以上例子中，我们能初步观察出一些规律：比如，线性算子为单射当且仅当其核空间是零空间，为满射当且仅当其像空间是值域空间. 又比如，从“小空间”映到“大空间”的线性变换是不满的，而从“大空间”映到“小空间”的线性变换是不单的.

**定理 4.24.** *设$V, W$为有限维线性空间，维数分别为$\dim(V) = n$, $\dim(W) = m$. 设$T : V \to W$为线性变换.*

*以下三个命题等价：*
*(1) $T$为单射；*
*(2) $\mathsf{Ker}(T) = \{\mathbf{0}\}$，或即$\mathsf{nullity}(T) = 0$；*
*(3) $\mathsf{rank}(T) = n$.*

*以下三个命题等价:*
*(1) $T$为满射；*
*(2) $\mathsf{Ran}(T) = W$，或即$\mathsf{rank}(T) = m$；*
*(3) $\mathsf{nullity}(T) = n - m$.*

*当$m = n$时，以下三个命题等价：*
*(1) $T$为单射；*
*(2) $T$为满射；*
*(3) $T$为双射.*

*证明.* 在$V$与$W$中取定基后，将线性变换$T$写成给定基下的矩阵$[T]$，这是一个$m \times n$阶的矩阵. 此时，“线性变换$T$为单射”等价于“对任意$\mathbf{b} \in \mathsf{Col}([T])$，线性方程组$[T]\mathbf{x} = \mathbf{b}$有唯一解”，即该方程组无自由变元. 可以证明，上述命题等价于“$\mathsf{rank}([T]) = n$”. 结合定理4.22，可得第一部分的等价性.

对于第二部分，按照定义知“线性变换$T$为满射”等价于$\mathsf{Ran}(T) = W$. 结合定理4.22可得所需结论.

对于第三部分，当$m = n$时，第一部分的(3)与第二部分的(2)是相同的，从而单射等价于满射，进而等价于双射. □

当$T$从“小空间”映到“大空间”时，我们有$n < m$，从而$\mathsf{nullity}(T) \geq 0 > n - m$，因此$T$不可能是满射. 当$T$从“大空间”映到“小空间”时，我们有$n > m$，从而$\mathsf{rank}(T) \leq m < n$，因此$T$不可能是单射.

需要注意的是，上述定理第三部分的等价性仅在$V, W$的维数均有限时成立. 对于无穷维空间的情况，例4.23(vi)给出了反例.

对线性变换而言，我们给双射一个新的名词.

**定义 4.5.** *设$T : V \to W$为线性变换. 若$T$为双射，则称$T$为（线性）同构. 这时，我们也称线性空间$V$与$W$是同构的.*

顾名思义，“同构”的意思就是这两个空间的线性结构是完全相同的. 线性代数代数理论的一个重要目的，就是将抽象的情形全部转化为具体的情形来进行研究.

**定理 4.25.** *任何一个$n$维的$\mathbb{F}$-线性空间均同构于$\mathbb{F}^n$.*

证明. 不妨设$V$为$\mathbb{F}$上的线性空间，其维数维$n$. 任取其一组基$B$，则取坐标映射

$$[\cdot]_B:\ V\to\mathbb{F}^n,\ \mathbf{v}\mapsto[\mathbf{v}]_B$$

是一个线性同构. 因此$V$同构于$\mathbb{F}^n$. □

**例 4.26.** *(1) 多项式空间$P_{n-1}$同构于$\mathbb{R}^n$，一个自然的同构映射的构造为*

$$a_0+a_1x+\ldots+a_{n-1}x^{n-1}\quad\mapsto\quad(a_0,a_1,\ldots,a_{n-1}).$$

*(2) 矩阵空间$M_2(\mathbb{R})$同构于$\mathbb{R}^4$，一个自然的同构映射的构造为*

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}\quad\mapsto\quad(a,b,c,d).$$

在第二章中，我们已经介绍了映射的复合：设$U,V,W$为三个集合，而映射$f,g$满足

$$U\xrightarrow{f}V\xrightarrow{g}W,$$

则$g$与$f$的复合映射$g\circ f:U\to W$的定义为

$$(g\circ f)(\mathbf{u})=g(f(\mathbf{u})),\quad(\forall\mathbf{u}\in U).$$

有时，为方便起见，我们会省去中间的符号“$\circ$”，记$T_2T_1=T_2\circ T_1$，并把它定义为映射$T_2$与$T_1$的乘法. 当一个映射$T$的定义域与值域相同时，就可以反复作用映射$T$，这时，我们记$T^n=T\circ\ldots\circ T$（$n$个$T$的复合映射）.

**定理 4.27.** *若$T_1:U\to V$与$T_2:V\to W$均为线性变换，则$T_2\circ T_1:U\to W$也是线性变换.*

证明. 对任意$\mathbf{u}_1,\mathbf{u}_2\in U$，以及任意标量$k_1,k_2$，利用$T_1,T_2$均为线性变换，可得

$$\begin{aligned}(T_2\circ T_1)(k_1\mathbf{u}_1+k_2\mathbf{u}_2)&=T_2\big(T_1((k_1\mathbf{u}_1+k_2\mathbf{u}_2)\big)=T_2\big(k_1T_1(\mathbf{u}_1)+k_2T_1(\mathbf{u}_2)\big)\\&=k_1T_2(T_1(\mathbf{u}_1))+k_2T_2(T_1(\mathbf{u}_2))=k_1(T_2\circ T_1)(\mathbf{u}_1)+k_2(T_2\circ T_1)(\mathbf{u}_2).\end{aligned}$$

因此$T_2\circ T_1$是线性变换. □

**定理 4.28.** *若$T_1:U\to V$与$T_2:V\to W$均为线性变换. 设$B,B',B''$分别为线性空间$U,V,W$的基. 则*

$$[T_2\circ T_1]_{B'',B}=[T_2]_{B'',B'}[T_1]_{B',B}.$$

证明. 对任意$\mathbf{u}\in U$，我们有

$$[T_2]_{B'',B'}[T_1]_{B',B}[u]_B=[T_2]_{B'',B'}[T_1(u)]_{B'}=[T_2(T_1(u))]_{B''}=[T_2\circ T_1]_{B'',B}[u]_B.$$

当$\mathbf{u}$在$U$中取遍时，向量$[u]_B$取遍整个$\mathbb{R}^{\dim U}$. 因此得$[T_2\circ T_1]_{B'',B}=[T_2]_{B'',B'}[T_1]_{B',B}$. □

**例 4.29.** *我们来看看有关零变换0、恒等算子$I$的线性变换的复合.*

*设$0:V\to W$，而$T_1:U\to V$，$\tilde{T}_1:W\to\widetilde{U}$均为线性变换. 容易验证$0\circ T_1=0:U\to W$，以及$\widetilde{T_1}\circ 0=0:V\to\widetilde{U}$.*

*设$I:V\to V$，而$T_2:U\to V$，$\widetilde{T_2}:V\to\widetilde{U}$均为线性变换. 容易验证$T_2\circ I=T_2$，以及$I\circ T_1=T_1$.*

**例 4.30.** *求线性变换$T_2$与$T_1$的复合，其中*

$$T_1:\ P_2\to P_3,\ p(x)\mapsto xp(x),$$

$$T_2:\ P_3\to P_3,\ p(x)\mapsto p(3x+1).$$

*求$T_2\circ T_1$以及它在标准基下的矩阵.*

解. 取$P_2$的基$B=\{p_0(x),p_2(x)\}$，以及$P_3$的基$\tilde{B}=\{\widetilde{p_0}(x),\widetilde{p_1}(x),\widetilde{p_2}(x)\}$，其中

$$p_0(x)=\widetilde{p_0}(x)=1,\quad p_1(x)=\widetilde{p_1}(x)=x,\quad \widetilde{p_2}(x)=x^2.$$

则有

$$\begin{aligned}T_1(p_0(x))&=xp_0(x)=x,\\T_1(p_1(x))&=xp_1(x)=x^2.\end{aligned}$$

以及

$$\begin{aligned}T_2(\widetilde{p_0}(x))&=\widetilde{p_0}(3x+1)=1,\\T_2(\widetilde{p_1}(x))&=\widetilde{p_1}(3x+1)=1+3x,\\T_2(\widetilde{p_2}(x))&=\widetilde{p_2}(3x+1)=(3x+1)^2=1+6x+9x^2.\end{aligned}$$

因此，我们有

$$[T_1]_{\widetilde{B},B}=\begin{bmatrix}0&0\\1&0\\0&1\end{bmatrix},\qquad [T_2]_{\widetilde{B},\widetilde{B}}=\begin{bmatrix}1&1&1\\0&3&6\\0&0&9\end{bmatrix}.$$

从而

$$[T_2\circ T_1]_{\widetilde{B},B}=[T_2]_{\widetilde{B},\widetilde{B}}[T_1]_{\widetilde{B},B}=\begin{bmatrix}1&1&1\\0&3&6\\0&0&9\end{bmatrix}\begin{bmatrix}0&0\\1&0\\0&1\end{bmatrix}=\begin{bmatrix}1&1\\3&6\\0&9\end{bmatrix}.$$

另一方面，我们也可以直接计算线性变换的复合，得

$$(T_2\circ T_1)(p(x))=T_2\big(T_1(p(x))\big)=T_2(xp(x))=(3x+1)p(3x+1).$$

特别地，

$$\begin{aligned}(T_2\circ T_1)(p_0(x))&=(3x+1)p_0(3x+1)=(3x+1)1=1+3x,\\(T_2\circ T_1)(p_1(x))&=(3x+1)p_1(3x+1)=(3x+1)(3x+1)=1+6x+9x^2.\end{aligned}$$

这与之前计算的$T_2\circ T$在基$B,\widetilde{B}$下的矩阵是吻合的. □

**例 4.31.** 设$U$是$\mathbb{R}$上某类函数构成的线性空间. 证明下述量子版本的不确定性原理：

$$P\circ Q-Q\circ P=I.$$

其中，算子$P$和$Q$的定义为

$$(Pf)=f'(x),\quad (Qf)(x)=xf(x),\quad (\forall f\in U).$$

注：$P$表示动量算符；$Q$表示位置算符.

解. 对任意函数$f\in U$，有

$$(P\circ Q)(f)=P(Q(f))=P(xf(x))=f(x)+xf'(x),$$

$$(Q\circ P)(f)=Q(P(f))=Q(f'(x))=xf'(x).$$

因此

$$(P\circ Q-Q\circ P)(f)=f(x)+xf'(x)-xf'(x)=f(x)=I(f).$$

故$P\circ Q-Q\circ P=I$. □

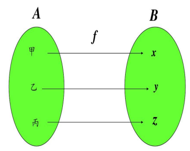
思考题：证明对有限维线性空间上的任何线性算子$P$与$Q$，都不可能满足$P \circ Q - Q \circ P = I$.

在第二章中，我们已经介绍了矩阵变换的逆. 现在，我们正式定义一个映射的可逆性.

**定义 4.6.** 设$U$, $V$为集合，用$I_U$和$I_V$分别表示$U$和$V$上的恒等映射. 对映射$f : U \to V$，若存在映射$g : V \to U$，使得

$$g \circ f = I_U, \quad f \circ g = I_V,$$

则称映射$f$为可逆的，且将$g$称为$f$的逆映射，常常记为$f^{-1}$.




思考题：试证明映射$f : U \to V$为双射，当且仅当$f$可逆.

**定理 4.32.** *(1) 设$T : V \to W$为可逆线性变换，则其逆变换$T^{-1} : W \to V$也是线性的.*

*(2) 设$B$与$B'$分别为空间$V$与$W$的基，则矩阵$[T]_{B',B}$可逆，且*

$$[T^{-1}]_{B,B'} = [T]^{-1}_{B',B}.$$

证明.（略.） □

**定理 4.33.** *设$T_1 : U \to V$与$T_2 : V \to W$均为可逆的线性变换，则$T_2 \circ T_1$也可逆，且*

$$(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}.$$

证明.（略.） □

**例 4.34.** 设线性算子$T : \mathbb{R}^3 \to \mathbb{R}^3$由下式给出：

$$T(x_1, x_2, x_3) = (3x_1 + x_2,\ -2x_1 - 4x_2 + 3x_3,\ 5x_1 + 4x_2 - 2x_3).$$

问$T$是否可逆？若可逆，求出$T^{-1}$.

解. 在标准基$S = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$下，我们有

$$[T] = \begin{bmatrix} 3 & 1 & 0 \\ -2 & -4 & 3 \\ 5 & 4 & -2 \end{bmatrix}.$$

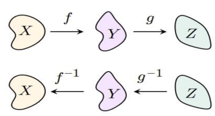
可通过行列式或初等行变换判断$[T]$是可逆的，从而$T$可逆. 计算逆矩阵得

$$[T^{-1}] = [T]^{-1} = \begin{bmatrix} 4 & -2 & -3 \\ -11 & 6 & 9 \\ -12 & 7 & 10 \end{bmatrix}.$$

故

$$T^{-1}(x_1, x_2, x_3) = (4x_1 - 2x_2 - 3x_3,\ -11x_1 + 6x_2 + 9x_3,\ -12x_1 + 7x_2 + 10x_3).$$

□



**例 4.35.** 设线性算子$T: P_3 \to P_3$的定义为

$$T(p(x)) = p(2x+1).$$

问$T$是否可逆？若可逆，求出$T^{-1}$.

解. 取$P_3$的标准基$B = \{p_0(x), p_1(x), p_2(x)\}$，其中$p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = x^2$. 我们有

$$\begin{aligned} T(p_0(x)) &= \widetilde{p_0}(2x+1) = 1, \\ T(p_1(x)) &= \widetilde{p_1}(2x+1) = 1 + 2x, \\ T(p_2(x)) &= \widetilde{p_2}(2x+1) = (1+2x)^2 = 1 + 4x + 4x^2. \end{aligned}$$

因此，我们有

$$[T]_B = \left[\begin{array}{c:c:c} 1 & 1 & 1 \\ 0 & 2 & 4 \\ 0 & 0 & 4 \end{array}\right].$$

计算逆矩阵可得

$$[T^{-1}]_B = [T_B]^{-1} = \begin{bmatrix} 1 & -1/2 & 1/4 \\ 0 & 1/2 & -1/2 \\ 0 & 0 & 1/4 \end{bmatrix}.$$

设$p(x) = a + bx + cx^2$，则

$$[T^{-1}(p(x))]_B = [T]_B[p(x)]_B = \begin{bmatrix} 1 & -1/2 & 1/4 \\ 0 & 1/2 & -1/2 \\ 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a - \frac{1}{2}b + \frac{1}{4}c \\ \frac{1}{2}b - \frac{1}{2}c \\ \frac{1}{4}c \end{bmatrix},$$

即

$$T^{-1}(a + bx + cx^2) = \left(a - \frac{1}{2}b + \frac{1}{4}c\right) + \left(\frac{1}{2}b - \frac{1}{2}c\right)x + \frac{1}{4}cx^2.$$

以上是标准的解题方法. 对本题而言，线性变换$T$的效果是对多项式做了变量替换$y = 2x+1$，因此可以直观地断言$T^{-1}$的效果是作变量替换$x = (y-1)/2$，即$T^{-1}(p(y)) = p\big((y-1)/2\big)$. 读者们可以验证这两种结果是相同的. □

由于有限维空间上的线性变换都可以写成基下的矩阵，我们将第一章中关于矩阵的一些定理重新用线性变换的形式展示出来.

**定理 4.36.** *设$V$和$W$均为有限维线性空间，且维数相同. 设$T: V \to W$为线性变换.*

*(1) 若$T_1$为单射，则$T_1$可逆.*

*(2) 若$T_1$为满射，则$T_1$可逆.*

**定理 4.37.** *设$V$和$W$均为有限维线性空间，且维数相同. 设$T_1: V \to W$为与$T_2: W \to V$均为线性变换.*

*(1) 若$T_2 \circ T_1 = I_V$，则$T_1$可逆，且$T_1^{-1} = T_2$.*

*(2) 若$T_2 \circ T_1 = I_V$，则$T_2$可逆，且$T_2^{-1} = T_1$.*

需要注意的是，当$V$和$W$为无限维线性空间时，上述定理便不在成立了，一个典型的反例是例4.23中的右平移$T_1$和左平移$T_2$满足$T_2 \circ T_1 = I$，但$T_1$与$T_2$均不可逆.

思考题：(1) 证明若$T_2 \circ T_1$为单射，则$T_1$也是单射.

(2) 证明若$T_2 \circ T_1$为满射，则$T_2$也是满射

(3) 举一个例子，使得$T_2 \circ T_1$为单射，但$T_2$不单.

(4) 举一个例子，使得$T_2 \circ T_1$为满射，但$T_1$不满.

## 4.4 相似性

我们可以将线性变换理解为某种运动，如拉伸、错切、旋转等. 而线性空间的一组基则是一位观测者建立的坐标系. 将线性变换表示成这组基下的矩阵，就是由该观测者得到的运动的具体表达式. 当两位观测者建立两个不同的坐标系，而观察同一种运动时，就会得到两种不同的表达式，这两种表达式之间一定存在确切的联系. 这一节的目的就是要明确一个线性变换在两组不同基下的矩阵之间的关系.

在上一章中,我们已经提到了基的变换,它给出了同一向量在不同基下的坐标之间的转换关系. 设$B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$与$B' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$为空间$V$的两组基，则从基$B$'到基$B$的过渡矩阵为

$$P_{B \leftarrow B'} = \left[\begin{array}{c:c:c:c} [\mathbf{v}'_1]_B & [\mathbf{v}'_2]_B & \ldots & [\mathbf{v}'_n]_B \end{array}\right]$$

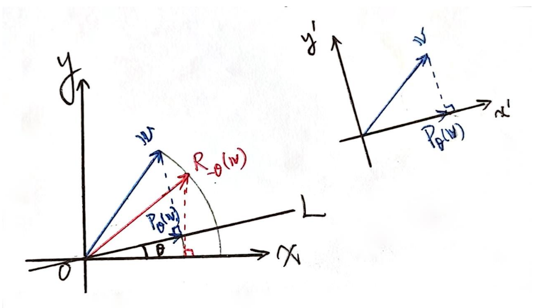
对任何向量$\mathbf{x} \in V$，它满足关系式

$$[\mathbf{x}]_B = P_{B \leftarrow B'} [\mathbf{x}]_{B'}.$$

现在，设$T$是$V$上的一个线性算子，它在基$B$、$B'$下的矩阵分别为$[T]_B$、$[T]_{B'}$. 那么$[T]_B$与$[T]_{B'}$之间有怎样的关系呢？

然我们先回顾第二章中的一个例子.

**例 4.38.** 如图，设$L$是过原点的直线，与$x$轴正向的夹角是$\theta$弧度. 试求$\mathbb{R}^2$中向直线$L$作正交投影的算子$P_\theta$.




在第二章中，我们展示的一种方法是：第一步，将向量$\mathbf{v}$与直线$L$先一起绕原点顺时针转动$\theta$角（这时直线与$x$轴重合）；第二步，向$x$轴作正交投影；第三步，将垂足与直线一起逆时针转动$\theta$角，就得到了向直线$L$作正交投影的垂足. 用线性映射的复合写出，有

$$P_\theta = R_\theta \circ P_0 \circ R_{-\theta} = (R_{-\theta})^{-1} \circ P_0 \circ R_{-\theta}.$$

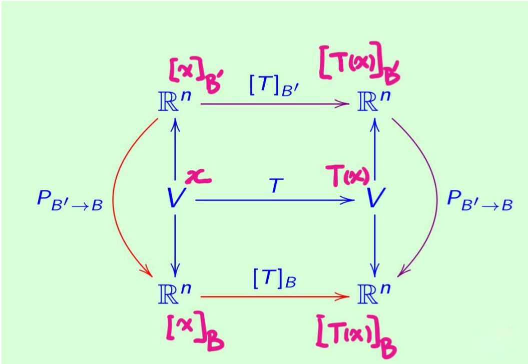
它们的标准矩阵满足

$$[P_\theta] = [R_\theta][P_0][R_{-\theta}].$$

现在我们用另一种方式理解上式:第一步,我们将原坐标系$xOy$整体逆时针转动$\theta$角,得到坐标系$x'Oy'$，这对应矩阵$[R_{-\theta}]$，事实上它是从坐标系$xOy$到坐标系$x'Oy'$的过度矩阵（原坐标下的角度等于新坐标下的角度减去$\theta$)；第二步，在$x'Oy'$坐标系里，向横轴$x'$轴作正交投影，这对应矩阵$[P_0]$，事实上它是所求正交投影在$x'Oy'$下的矩阵；第三步，再将新坐标系$x'Oy'$整体顺时针转动$\theta$角，回到原坐标系$xOy$，它对应矩阵$[R_\theta]$，事实上它是从坐标系$x'Oy'$到坐标系$xOy$的过度矩阵. 最终我们得到了$[P_\theta]$，它是所求正交投影在原坐标系下的矩阵.

**定理 4.39.** 设$V$为有限维线性空间，而$T : V \to V$是线性算子. 设$B$与$B'$是空间$V$的两组基. 则有

$$[T]_{B'} \;=\; P^{-1}_{B \leftarrow B'} \, [T]_B \, P_{B \leftarrow B'}.$$




证明. 对任意向量$\mathbf{x} \in V$，我们有

$$[T]_B \, P_{B \leftarrow B'} \, [\mathbf{x}]_{B'} = [T]_B \, [\mathbf{x}]_B = [T(\mathbf{x})]_B.$$

另一方面，

$$P_{B \leftarrow B'} \, [T]_{B'} \, [\mathbf{x}]_{B'} = P_{B \leftarrow B'} [T(\mathbf{x})]_{B'} = [T(\mathbf{x})]_B.$$

因此两式的左端是相等的. 定理得证. □

**例 4.40.** 设矩阵变换$T : \mathbb{R}^2 \to \mathbb{R}^2$的标准矩阵为

$$[T] = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}.$$

求$T$在另一组基$B' = \{\mathbf{u}_1', \mathbf{u}_2'\}$下的矩阵. 这里

$$\mathbf{u}_1' = (1,1), \quad \mathbf{u}_2' = (1,2).$$

并给出这两个矩阵的具体转换关系.

解. 记$\mathbb{R}^2$的标准基为$S = \{\mathbf{e}_1, \mathbf{e}_2\}$，则

$$[T]_S = [T] = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}, \quad [\mathbf{u}_1']_S = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad [\mathbf{u}_2']_S = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

我们有

$$P_{S \leftarrow B'} = \left[ [\mathbf{u}_1']_S \,\vdots\, [\mathbf{u}_2']_S \right] = \left[\begin{array}{c:c} 1 & 1 \\ 1 & 2 \end{array}\right], \quad P_{S \leftarrow B'}^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

从而

$$[T]_{B'} = P_{S \leftarrow B'}^{-1} [T]_S P_{S \leftarrow B'} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}.$$

□

现在我们将上述定理和例子中的矩阵关系进一步抽象出来，作成定义.

**定义 4.7.** 设$A, B \in M_n(\mathbb{F})$. 我们成$A$与$B$相似，是指存在可逆矩阵$P \in M_n(\mathbb{F})$，使得

$$P^{-1}AP = B.$$

这里定义的"相似"是一种"等价关系"，即满足下列三个性质.

(i) 自反性：方阵$A$与自身是相似的. 事实上，我们有$I^{-1}AI = A$.

(ii) 对称性：若$A$与$B$相似，则$B$与$A$也相似. 事实上，假设存在可逆矩阵$P$使得$P^{-1}AP = B$，则亦有$(P^{-1})^{-1}BP^{-1} = A$.

(iii) 传递性：若$A$与$B$相似，且$B$与$C$相似，则$A$与$C$也相似. 事实上，假设存在可逆矩阵$P$、$Q$，使得$P^{-1}AP = B$、$Q^{-1}BQ = C$，则有$(PQ)^{-1}A(PQ) = C$，这里的$PQ$也是可逆的.

**例 4.41.** 验证矩阵$A$与$D$相似，其中

$$A=\begin{bmatrix}1&1\\-2&4\end{bmatrix},\quad D=\begin{bmatrix}2&0\\0&3\end{bmatrix}.$$

解.（略.）此题起承上启下的作用，读者可以忽略，因为它跟后续的对角化是类似的. □

需注意的是，用到的过度矩阵$P$的取法并不是唯一的. 另外，对于对角矩阵$D=\mathrm{diag}(\mu_1,\mu_2,\ldots,\mu_n)$而言，我们总有$D\mathbf{e}_j=\mu_j\mathbf{e}_j$ $(1\le j\le n)$.

相似的两个方阵可以看成同一个线性算子在不同基下的矩阵，因此它们应当有很多相同的特征信息，这些信息往往被称为“相似不变量”，它们都是矩阵背后的线性算子本身的内蕴信息.

**定理 4.42.** *设矩阵$A,B\in M_n$是相似的. 则*

*(1) $\mathit{tr}(A)=\mathit{tr}(B)$;*

*(2) $\mathit{det}(A)=\mathit{det}(B)$;*

*(3) $\mathit{rank}(A)=\mathit{rank}(B)$;*

*(4) $\mathit{nullity}(A)=\mathit{nullity}(B)$.*

证明. 由于$A$与$B$相似，设$P$为可逆矩阵，满足$P^{-1}AP=B$.

(1) $\mathrm{tr}(B)=\mathrm{tr}\big(P^{-1}(AP)\big)=\mathrm{tr}\big((AP)P^{-1}\big)=\mathrm{tr}(A)$.

(2) $\det(B)=\det(P^{-1}AP)=\big(\det(P)\big)^{-1}\det(A)\det(P)=\det(A)$.

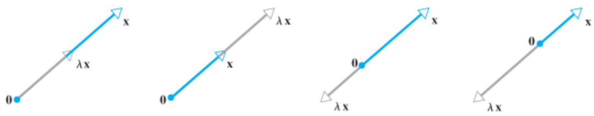
(3) 由于左、右乘可逆矩阵不改变矩阵的秩，因此$\mathrm{rank}(B)=\mathrm{rank}(P^{-1}AP)=\mathrm{rank}(A)$.

(4) $\mathrm{nullity}(A)=n-\mathrm{rank}(A)=n-\mathrm{rank}(B)=\mathrm{nullity}(B)$. □

## 4.5 特征值与特征向量

一个线性变换的结构可能是很复杂的. 在数学思想的层面，我们总是先考虑较简单的情形，再将简单的结构组合起来，去探究更复杂的情况；或是将复杂的情形化简成较简单的情形，经过层层转化，最终达到目标.

如果一个线性变换将某一维子空间映到这个子空间本身，那么它在这一维上的变换效果将是拉伸、收缩、镜像反射、正交投影到零向量等（见下图），这种情况是我们比较容易理解和处理的.




(a) $0\le\lambda\le1$　(b) $\lambda\le1$　(c) $-1\le\lambda\le0$　(d) $\lambda\le-1$

**定义 4.8.** *设$V$是数域$\mathbb{F}$上的向量空间，而$T:V\to V$为线性算子. 若纯量$\lambda\in\mathbb{F}$以及非零向量$\mathbf{x}\in V$满足*

$$T(\mathbf{x})=\lambda\mathbf{x},$$

*则称$\lambda$为线性算子$T$的特征值，称$\mathbf{x}$为$T$关于特征值$\lambda$的特征向量.*

当空间是有限维时，算子$T$可写成基下的矩阵$[T]$，下面我们重新写一下矩阵特征值与特征向量的定义.

**定义 4.9.** *设$A\in M_n(\mathbb{F})$，若纯量$\lambda\in\mathbb{F}$以及非零向量$\mathbf{x}\in\mathbb{F}^n$满足*

$$A\mathbf{x}=\lambda\mathbf{x},$$

*则称$\lambda$为矩阵$A$的特征值，称$\mathbf{x}$为$A$关于特征值$\lambda$的特征向量.*

这里需要强调一下，特征向量都是非零的，只有非零向量才能张成它所在的一维子空间. 在前一节中，我们知道相似的矩阵可以看成同一个线性变换在不同基下的两种不同具体表现，因此它们在某一维子空间上的变换效果也应该有对应关系.

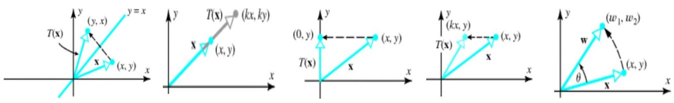
**定理 4.43.** *相似矩阵具有相同的特征值.*

*证明.* 设$A, B$为$M_n(\mathbb{F})$中的相似矩阵，则存在可逆矩阵$P \in M_n(\mathbb{F})$使得$P^{-1}AP = B$. 假设$B\mathbf{x} = \lambda\mathbf{x}$，其中$\lambda \in \mathbb{F}$，而$\mathbf{x}$为$\mathbb{F}^n$中的非零向量. 则有$P^{-1}AP\mathbf{x} = \lambda\mathbf{x}$，进而$AP\mathbf{x} = \lambda P\mathbf{x}$. 记$y = P\mathbf{x}$，则$A\mathbf{y} = \lambda\mathbf{y}$. 因$P$可逆，而$\mathbf{x} \neq \mathbf{0}$，可得$\mathbf{y} \neq \mathbf{0}$. 所以$\lambda$也是$A$的特征向量. □

从上述证明中可以看出，对于$A, B$的公共特征值$\lambda$的特征向量可能不同，事实上，特征向量之间也差一个过渡矩阵的转换，即$\mathbf{y} = P\mathbf{x}$.

**例 4.44.** *通过几何上的变换效果来观察下述线性算子的特征值与特征向量.*

*(1) 算子$T$为$\mathbb{R}^2$上关于直线$y = x$的镜面反射，即$T(x, y) = (y, x)$.*
*(2) 算子$T$为$\mathbb{R}^2$上的膨胀或收缩，即$T(x, y) = (kx, ky)$.*
*(3) 算子$T$为$\mathbb{R}^2$上关于$y$轴的正交投影，即$T(x, y) = (0, y)$.*
*(4) 算子$T$为$\mathbb{R}^2$上关于$x$方向的错切，即$T(x, y) = (x + ky, y)$.*
*(5) 算子$R_\theta$为$\mathbb{R}^2$上关于原点进行的逆时针$\theta$角的旋转.*




*解.* (略.) □

**例 4.45.** *设$D : C^\infty(-\infty, +\infty) \to C^\infty(-\infty, +\infty)$为微分算子，即$D = \frac{d}{dx}$. 对任意$\lambda \in \mathbb{R}$, 函数$e^{\lambda x}$是$D$的特征向量，即*

$$D(e^{\lambda x}) = \lambda e^{\lambda x}.$$

**例 4.46.** *设$\Delta : C^\infty(-\pi, \pi) \to C^\infty(-\pi, \pi)$为二阶微分算子$\Delta = -\frac{d^2}{dx^2}$. 则对任意$k \in \mathbb{Z}$, 函数$\sin kx$与$\cos kx$为$\Delta$关于特征值$k^2$的特征向量，即*

$$T(\sin kx) = k^2 \sin kx, \quad T(\cos kx) = k^2 \cos kx.$$

*在物理中，简谐波是拉普拉斯算子的特征向量.*

接下来，我们来探索一下求一个具体矩阵的特征值和特征向量的方法. 注意到表达式$A\mathbf{x} = \lambda\mathbf{x}$可写为$(\lambda I - A)\mathbf{x} = \mathbf{0}$，这是一个齐次线性方程组.

**定理 4.47.** *设$A \in M_n(\mathbb{F})$. 则$\lambda \in \mathbb{F}$是$A$的特征值，当且仅当*

$$\mathsf{det}(\lambda I_n - A) = 0.$$

*证明.* 纯量$\lambda$为矩阵$A$的特征值，当且仅当存在$\mathbf{x} \neq \mathbf{0}$使得$A\mathbf{x} = \lambda\mathbf{x}$，当且仅当线性方程组$(\lambda I_n - A)\mathbf{x} = \mathbf{0}$有非零解，当且仅当$(\lambda I_n - A)\mathbf{x} = \mathbf{0}$有无穷多解，当且仅当$\mathsf{det}(\lambda I_n - A) = 0$. □

当$A = [a_{ij}] \in M_n(\mathbb{F})$时，若将$\mathsf{det}(\lambda I - A)$具体展开，会得到一个关于$\lambda$的$n$次多项式，其最高项系数为1，而其它项系数为$a_{ij}$的函数，因此系数来自于$\mathbb{F}$.

**定义 4.10.** 设$A \in M_n$. 我们称$det(\lambda I - A) = 0$为$A$的特征方程，称

$$f_A(\lambda) = det(\lambda I_n - A)$$

为$A$的特征多项式.

根据代数学基本定理，任何一个$n$次复多项式有$n$个（复）根（带重数），或即任何一个$n$次复多项式可以写成$n$个一次因子的乘积. 另外，任何一个$n$次实多项式可以写成一次或二次因子的乘积，其中二次因子的判别式小于0，因此该多项式的（实）根数量不超过$n$个. 因此复矩阵必有特征值与特征向量，而实矩阵可能没有特征值与特征向量.

思考题：设$A \in M_n(\mathbb{R})$, 证明$A$与$A^T$具有相同特征值.

思考题：证明相似矩阵具有相同的特征多项式.

**例 4.48.** 令

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

*(1)* 设$A \in M_2(\mathbb{R})$（即作为实矩阵），求$A$的特征值.
*(2)* 设$A \in M_2(\mathbb{C})$（即作为复矩阵），求$A$的特征值.

解. 令

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & 1 \\ -1 & \lambda - 1 \end{vmatrix} = (\lambda - 1)^2 + 1 = \lambda^2 - 2\lambda + 2.$$

(1) 当$\mathbb{F} = \mathbb{R}$时，关于$\lambda$的二次方程判别式为小于0，方程无解，因此不存在特征值.

(2) 当$\mathbb{F} = \mathbb{C}$时，解方程得两个特征值$\lambda_\pm = \frac{2 \pm \sqrt{-4}}{2} = 1 \pm \mathrm{i}$. □

**例 4.49.** 求3阶实方阵

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

的所有特征值.

解.（略.） □

**例 4.50.** 求矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

的所有特征值.

解.（略.） □

**定理 4.51.** 设$A$为三角矩阵，则其特征值恰为所有对角元素（带重数）.

对于$\mathbb{R}^2$上的膨胀或收缩，我们已经看到，同一个特征值可以有多个线性无关的特征向量. 下面，我们把关于同一个特征值的特征向量全部放到一起进行刻画.

**定义 4.11.** 设$A \in M_n(\mathbb{F})$，而$\lambda$为$A$的一个特征值. 称

$$E_\lambda(A) := \{\mathbf{x} \in \mathbb{F}^n : A\mathbf{x} = \lambda \mathbf{x}\} = Null(\lambda I_n - A)$$

为矩阵$A$关于特征值$\lambda$的特征子空间.

由于$E_\lambda(A)$是矩阵$\lambda I_n - A$的零空间，因此它是$\mathbb{F}^n$的一个子空间．这个定义的主要意义在于，将矩阵$A$看作线性变换$T_A$，它在子空间$E_\lambda(A)$上的变换效果为数乘$\lambda$，即相应的膨胀、收缩、镜像反射、正交投影到零向量等．

思考题：设矩阵$A$与$B$相似，而$\lambda$为它们公共的特征值．试证明$\dim(E_\lambda(A)) = \dim(E_\lambda(B))$．

**例 4.52.** *求矩阵$A$的所有特征子空间的各一组基，其中*

$$A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \in M_3(\mathbb{R}).$$

解．首先求特征值，令

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda & 0 & 2 \\ -1 & \lambda - 2 & -1 \\ -1 & 0 & \lambda - 3 \end{vmatrix} = \lambda(\lambda-2)(\lambda-3) + 2(\lambda-2) = (\lambda-1)(\lambda-2)^2.$$

得两个不同特征值1与2．

对于特征值1，令

$$\mathbf{0} = (I - A)\mathbf{v} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & -1 & -1 \\ -1 & 0 & -2 \end{bmatrix},$$

得解空间的一组基$\mathbf{v}_1 = \begin{bmatrix} -2 & 1 & 1 \end{bmatrix}^T$．

对于特征值2，令

$$\mathbf{0} = (2I - A)\mathbf{v} = \begin{bmatrix} 2 & 0 & 2 \\ -1 & 0 & -1 \\ -1 & 0 & -1 \end{bmatrix},$$

得解空间的一组基$\mathbf{v}_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$, $\mathbf{v}_3 = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T$．

最终，我们有

$$E_1(A) = \mathsf{span}\left\{\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}\right\}, \quad E_2(A) = \mathsf{span}\left\{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}.$$

□

**例 4.53.** *通过几何上的变换效果来观察下述线性算子的特征值与特征向量．*

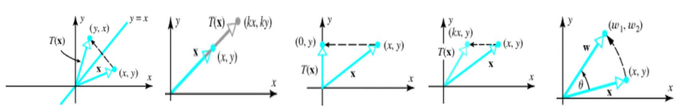
*(1) 算子$T$为$\mathbb{R}^2$上关于直线$y = x$的镜面反射，即$T(x, y) = (y, x)$．*

*(2) 算子$T$为$\mathbb{R}^2$上的膨胀或收缩，即$T(x, y) = (kx, ky)$．*

*(3) 算子$T$为$\mathbb{R}^2$上关于$y$轴的正交投影，即$T(x, y) = (0, y)$．*

*(4) 算子$T$为$\mathbb{R}^2$上关于$x$方向的错切，即$T(x, y) = (x + ky, y)$．*

*(5) 算子$R_\theta$为$\mathbb{R}^2$上关于原点进行的逆时针$\theta$角的旋转．*



解．（略.）

□

接下来，我们来探究一下特征值与矩阵可逆性以及矩阵代数运算之间的一些关系.

**定理 4.54.** 设$A \in M_n(\mathbb{F})$. 则以下命题等价：

*(a)* 矩阵$A$可逆.

*(t)* 纯量0不是$A$的特征值.

证明. 我们已经知道$A$可逆等价于$\mathsf{det}(A) \neq 0$. 注意到

$$\mathsf{det}(0I - A) = \mathsf{det}(-A) = (-1)^n\mathsf{det}(A).$$

因此$A$可逆等价于0不是$A$的特征值. □

从上述证明中，我们可以得到

$$\mathsf{det}(A) = (-1)^n\mathsf{det}(0I - A) = (-1)^n f_A(0).$$

设$f_A(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0$，则$f_A(0) = c_0$. 所以矩阵$A$的行列式等于其特征多项式的常数项乘以$(-1)^n$.

**例 4.55.** 设

$$A = \begin{bmatrix} 1 & 1+\mathrm{i} \\ 2 & \mathrm{i} \end{bmatrix} \in M_2(\mathbb{C}).$$

求$A^{-1}$的所有特征值，已经它们对应的各一个特征向量.

解. 令

$$0 = \mathsf{det}(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -1-\mathrm{i} \\ -2 & \lambda - \mathrm{i} \end{vmatrix} = \lambda^2 - (1+\mathrm{i})\lambda - 2 - \mathrm{i} = (\lambda+1)(\lambda - 2 - \mathrm{i}).$$

得特征值$\lambda_1 = -1$, $\lambda_2 = 2 + \mathrm{i}$. 解方程组$\mathbf{0} = (\lambda_i I - A)\mathbf{v}_i$ $(i = 1, 2)$，分别得特征向量

$$\mathbf{v}_1 = \begin{bmatrix} -\frac{1+\mathrm{i}}{2} & 1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

□

**例 4.56.** 设$A \in M_n(\mathbb{F})$，而$q(x)$为系数来自于$\mathbb{F}$的多项式. 则有

$$\{q(A) \text{ 的特征值}\} \supseteq \{q(\lambda) : \lambda \text{ 为} A \text{ 的特征值}\}.$$

更具体地，若$A\mathbf{x} = \lambda\mathbf{x}$，则有$q(A)\mathbf{x} = q(\lambda)\mathbf{x}$.

证明. 先考虑$A^k$的情形. 对$k \geq 1$进行归纳法，我们有$A^0\mathbf{x} = I\mathbf{x} = 1\mathbf{x}$，以及

$$A^k\mathbf{x} = A^{k-1}(A\mathbf{x}) = A^{k-1}(\lambda\mathbf{x}) = \lambda(A^{k-1}\mathbf{x}) = \lambda(\lambda^{k-1}\mathbf{x}) = \lambda^k\mathbf{x}.$$

再利用矩阵加法与数乘的分配律，不难得到$q(A)\mathbf{x} = q(\lambda)\mathbf{x}$. □

值得一提的是，当$\mathbb{F} = \mathbb{C}$时，上述定理中的 “$\supseteq$”更更换为“=”（其证明是超纲的）.

我们已经熟悉了矩阵$A$与其转置$A^T$之间的关系. 对于复矩阵$A$来说，更加契合的对应结构为共轭转置，我们在这里展示一下关于共轭转置的初步性质.

**定义 4.12.** 设$A \in M_{m\times n}(\mathbb{C})$, 则$A$的伴随（或称共轭转置）$A^*$定义为$\overline{A^T}$，即满足下列条件的矩阵：

*(i)* $A^* \in M_{n\times m}(\mathbb{C})$;

*(ii)* $(A^*)_{ij} = \overline{(A)_{ji}}, \quad (1 \leq i \leq m, 1 \leq j \leq n).$

**命题 4.1.** 设$A, B \in M_{m\times n}(\mathbb{C})$, $C \in M_{n\times k}(\mathbb{C})$，而$c \in \mathbb{C}$. 则

*(i)* $(A+B)^* = A^* + B^*$.

*(ii)* $(cA)^* = \overline{c}A^*$.

*(iii)* $(A^*)^* = A$.

*(iv)* $(AC)^* = C^*A^*$.

**定理 4.57.** 设$A \in M_n(\mathbb{C})$. 若$\lambda$是$A$的一个特征值，则$\overline{\lambda}$是$A^*$的一个特征值.

证明. 由于$\lambda$是$A$的特征值，因此$\mathsf{det}(\lambda I - A) = 0$. 注意到$(\lambda I - A)^* = \overline{\lambda}I - A^*$，于是

$$\mathsf{det}(\overline{\lambda}I - A^*) = \mathsf{det}\big((\lambda I - A)^*\big) = \mathsf{det}\big(\overline{(\lambda I - A)^T}\big) = \overline{\mathsf{det}\big((\lambda I - A)^T\big)} = \overline{\mathsf{det}(\lambda I - A)} = 0.$$

所以$\overline{\lambda}$是$A^*$的特征值. □

**例 4.58.** 设

$$A = \begin{bmatrix} 1 & 1+\mathrm{i} \\ 2 & \mathrm{i} \end{bmatrix} \in M_2(\mathbb{C}).$$

求下述矩阵的所有特征值与对应的各一个特征向量.

*(i)* $A^5$; *(ii)* $A^2 + A - 3I_2$.

*(iii)* 求 $A^*$ 的特征值.

解. 在之前得例子中，我们已经得到矩阵$A$的特征值$\lambda_1 = -1$, $\lambda_2 = 2+\mathrm{i}$，以及相应的特征向量

$$\mathbf{v}_1 = \begin{bmatrix} -\frac{1+\mathrm{i}}{2} & 1 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

(1) 矩阵$A^5$的特征值分别为$\lambda_i' = \lambda_i^5$ $(i = 1, 2)$，分别有特征向量$\mathbf{v}_1, \mathbf{v}_2$.

(2) 矩阵$A^2 + A - 3I_2$的特征值分别为$\lambda_i'' = \lambda_i^2 + \lambda_i - 3$ $(i = 1, 2)$，分别有特征向量$\mathbf{v}_1, \mathbf{v}_2$.

(3) 矩阵$A^*$的特征值分别为$\lambda_i''' = \overline{\lambda_i}$ $(i = 1, 2)$.

这些特征值的具体数值，留给读者们完成计算. □

## 4.6 矩阵的对角化

一个$n$维空间上的线性算子$T$可能很复杂，当$T$有特征值和特征向量时，在特征向量所在的这一维上，算子$T$的变换效果是直观的，于是我们进一步考虑剩下的$n-1$维即可. 最简单的情形，是全空间有一组基，均由$T$的特征向量构成，这时$T$在每一维上的变换效果都是直观的，而在每个向量上的映射效果可由线性组合得到.

以下例子是最简单的一种情形：

**例 4.59.** 对角矩阵

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

的特征值为$\lambda_1, \lambda_2, \ldots, \lambda_n$. 且$\mathbb{R}^n$的标准基$\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$恰为这些特征值对应的特征向量.

我们将上述例子略微作一些推广，形成以下定义.

**定义 4.13.** 方阵$A \in M_n$（或$n$维线性空间$V$上的线性算子$T$）称为可对角化的，是指它有$n$个线性无关的特征向量.

**例 4.60.** 下述线性算子是否可对角化？若是，通过合适的特征向量来分析算子的变换效果.

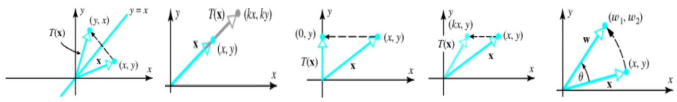
*(1)* 算子$T$为$\mathbb{R}^2$上关于直线$y=x$的镜面反射，即$T(x,y)=(y,x)$.

*(2)* 算子$T$为$\mathbb{R}^2$上的膨胀或收缩，即$T(x,y)=(kx,ky)$.

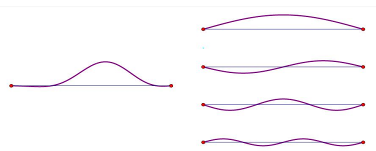
*(3)* 算子$T$为$\mathbb{R}^2$上关于$y$轴的正交投影，即$T(x,y)=(0,y)$.

*(4)* 算子$T$为$\mathbb{R}^2$上关于$x$方向的错切，即$T(x,y)=(x+ky,y)$.

*(5)* 算子$R_\theta$为$\mathbb{R}^2$上关于原点进行的逆时针$\theta$角的旋转.



解.（略.） □



**例 4.61.** 设$V$是$[-\pi,\pi]$上任意阶可导的且满足$f(-\pi)=f(\pi)=0$的函数$f$所构成的线性空间. 函数的傅里叶级数为

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

其中$a_n,b_n$为傅里叶系数. 事实上，函数$\sin nx$与$\cos nx$均为二阶微分算子$\Delta=-\frac{d^2}{dx^2}$的特征向量，它们也构成空间$V$的一组基.（这是无穷维空间，我们不具体介绍无穷维空间上基的定义，这是超纲的内容）

矩阵可对角化还有更多的描述方式.

**定理 4.62.** 设$A\in M_n(\mathbb{F})$. 以下命题等价：

*(i)* 矩阵$A$有$n$个线性无关的特征向量.

*(ii)* 存在可逆矩阵$P\in M_n(\mathbb{F})$与对角矩阵$D\in M_n(\mathbb{F})$，使得$P^{-1}AP=D$..

证明. 我们先证明"(i)$\Longrightarrow$(ii)". 设$A\mathbf{v}_i=\lambda_i\mathbf{v}_i$ $(1\le i\le n)$，其中$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$是$A$的$n$个线性无关的特征向量，分别对应特征值$\lambda_1,\lambda_2,\ldots,\lambda_n$. 这时$B=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$构成$\mathbb{F}^n$的一组基. 设$T_A$为乘$A$的矩阵变换，即$T_A:\mathbb{F}^n\to\mathbb{F}^n$, $\mathbf{x}\mapsto A\mathbf{x}$. 根据线性映射在基下的矩阵的定义，可以得到

$$[T_A]_B=\left[\begin{array}{c:c:c:c}[A\mathbf{v}_1]_B & [A\mathbf{v}_2]_B & \cdots & [A\mathbf{v}_n]_B\end{array}\right]=\left[\begin{array}{c:c:c:c}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{array}\right].$$

另一方面，设$S=\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}$为$\mathbb{F}^n$的标准基. 易知对于，有$[T]_S=A$. 故

$$[T_A]_B=P_{S\leftarrow B}^{-1}[T_A]_SP_{S\leftarrow B}=P_{S\leftarrow B}^{-1}AP_{S\leftarrow B}.$$

这里$P_{S\leftarrow B}$可逆，而$[T_A]_B$为对角矩阵.

我们再来证明"(ii)$\Longrightarrow$(i)". 设$D=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n)$，则有$D\mathbf{e}_i=\lambda_i\mathbf{e}_i$ $(1\le i\le n)$. 这时，取$\mathbf{v}_i=P\mathbf{e}_i$ $(1\le i\le n)$，则有

$$A\mathbf{v}_i=(PP^{-1})A(P\mathbf{e}_i)=P(P^{-1}AP\ \mathbf{e}_i=PD\mathbf{e}_i=\lambda_iP\mathbf{e}_i=\lambda_i\mathbf{v}_i.$$

由$P$可逆及$\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n$线性无关，知$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$仍线性无关. 因此$A$有$n$个线性无关的特征向量. □

从上述证明过程中，我们也得到了将$A$对角化的具体方法：计算$A$的特征值$\lambda_i$与对应的$n$个线性无关的特征向量$\mathbf{v}_i$，令$P$为$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$作为列向量所构成的分块矩阵，令$D=\mathsf{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n)$，则得到$P^{-1}AP=D$.

**例 4.63.** 在$M_3(\mathbb{R})$中，试对角化$A$，并写出对应的过渡矩阵与对角阵. 这里

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

解. 在之前的例题中，我们已经得到$A$的两个不同特征值1和2，以及特征子空间的基：

$$E_1(A)=\mathsf{span}\left\{\begin{bmatrix}-2\\1\\1\end{bmatrix}\right\},\quad E_2(A)=\mathsf{span}\left\{\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}-1\\0\\1\end{bmatrix}\right\}.$$

令

$$P=\left[\begin{array}{c:c:c}-2&0&-1\\1&1&0\\1&0&1\end{array}\right],\qquad D=\begin{bmatrix}1&&\\&2&\\&&2\end{bmatrix},$$

则有$P^{-1}AP=D$. □

接下来，我们再进一步分析一下矩阵可对角化的一些充分或必要条件.

**定理 4.64.** 设$\lambda_1,\lambda_2,\ldots,\lambda_k$为$A$的不同特征值，而$E_{\lambda_i}(A)$ $(1\le i\le k)$具有基

$$\mathbf{v}_{i,1},\mathbf{v}_{i,1},\ldots,\mathbf{v}_{i,m_k}.$$

则向量组

$$\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\quad \mathbf{v}_{2,1},\ldots,\mathbf{v}_{2,m_2},\quad ,\ldots,\quad \mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$$

是线性无关的.

证明. 令

$$c_{1,1}\mathbf{v}_{1,1}+\ldots+c_{1,m_1}\mathbf{v}_{1,m_1}+c_{2,1}\mathbf{v}_{2,1}+\ldots+c_{2,m_2}\mathbf{v}_{2,m_2}+\ldots+c_{k,1}\mathbf{v}_{k,1}+c_{k,m_k}\mathbf{v}_{k,m_k}=\mathbf{0}.$$

我们需要证明上式左端的线性组合的系数只能全为零.

方便起见，我们记$\mathbf{u}_i=c_{i,1}\mathbf{v}_{i,1}+\ldots+c_{i,m_i}\mathbf{v}_{i,m_i}$ $(1\le i\le k)$. 则有

$$\mathbf{u}_1+\mathbf{u}_2+\ldots+\mathbf{u}_k=\mathbf{0}, \tag{4.2}$$

以及

$$A\mathbf{u}_i=c_{i,1}A\mathbf{v}_{i,1}+\ldots+c_{i,m_i}A\mathbf{v}_{i,m_i}=c_{i,1}\lambda_i\mathbf{v}_{i,1}+\ldots+c_{i,m_i}\lambda_i\mathbf{v}_{i,m_i}=\lambda_i\mathbf{u}_i.$$

从而对$1\le j\le k-1$，有

$$\lambda_1^j\mathbf{u}_1+\lambda_2^j\mathbf{u}_2+\ldots+\lambda_k^j\mathbf{u}_k=A^j(\mathbf{u}_1+\mathbf{u}_2+\ldots+\mathbf{u}_k)=A^j\mathbf{0}=\mathbf{0}. \tag{4.3}$$

即

$$\begin{aligned}
&1\cdot\mathbf{u}_1+1\cdot\mathbf{u}_2+\ldots+1\cdot\mathbf{u}_k=\mathbf{0},\\
&\lambda_1\cdot\mathbf{u}_1+\lambda_2\cdot\mathbf{u}_2+\ldots+\lambda_k\cdot\mathbf{u}_k=\mathbf{0},\\
&\qquad\cdots\quad\cdots\\
&\lambda_1^{k-1}\cdot\mathbf{u}_1+\lambda_2^{k-1}\cdot\mathbf{u}_2+\ldots+\lambda_k^{k-1}\cdot\mathbf{u}_k=\mathbf{0}.
\end{aligned}$$

当我们将以上方程组中的$\mathbf{u}_i$和$\mathbf{0}$都替换为它们的第$r$个分量时，就得到了齐次线性方程组. 注意到系数矩阵对应的行列式为范德蒙行列式，当$\lambda_1, \lambda_2, \ldots, \lambda_k$不同时，该行列式值为$\prod_{1\le i<j\le n}(\lambda_j - \lambda_i) \neq 0$. 因此，我们推断$\mathbf{u}_i$ $(1 \le i \le k)$的第$r$个分量均为零. 这一结论对每个分量都是成立的，故有$\mathbf{u}_i = \mathbf{0}$ $(1 \le i \le k)$.

这时，由$c_{i,1}\mathbf{v}_{i,1} + \ldots + c_{i,m_i}\mathbf{v}_{i,m_i} = \mathbf{0}$以及$\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,m_i}$线性无关知所有系数均为零. 证毕.

□

由上述定理可知，在求解矩阵对角化的过程中，仅需找出各不同特征子空间的基，它们的并集仍然是线性无关的.

**推论 4.65.** 设$A \in M_n(\mathbb{F})$，且$A$的所有不同特征值为$\lambda_1, \lambda_2, \ldots, \lambda_k$. 则$A$可对角化当且仅当

$$\dim E_{\lambda_1}(A) + \dim E_{\lambda_2}(A) + \ldots + \dim E_{\lambda_k}(A) = n.$$

**推论 4.66.** 设$A \in M_n(\mathbb{F})$. 若$A$有$n$个不同特征值，则$A$可对角化.

**例 4.67.** 试说明矩阵$B \in M_3(\mathbb{R})$不可对角化. 这里

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

解. 矩阵$B$是下三角矩阵，特征值均来自于对角元，因此有两个不同特征值1和2. 我们有

$$I - A = \begin{bmatrix} 0 & & \\ -1 & 1 & \\ 0 & -1 & 1 \end{bmatrix}, \quad 2I - A = \begin{bmatrix} -1 & & \\ -1 & 0 & \\ 0 & -1 & 0 \end{bmatrix}.$$

可知$\mathsf{rank}(I-A) = 2$, $\mathsf{rank}(2I-A) = 2$，进而$\mathsf{nullity}(I-A) = 3-2 = 1$, $\mathsf{nullity}(2I-A) = 3-2 = 1$. 于是

$$\dim E_1(A) + \dim E_2(A) = \mathsf{nullity}(I-A) + \mathsf{nullity}(2I-A) = 1+1 = 2 < 3.$$

故矩阵$B$不可对角化. □

对角矩阵的数乘、加法和乘法可以直接在对角的位置分别进行，是非常方便的. 对于可对角化的矩阵来说，有关它们的运算也可在对角化之后再进行. 假设矩阵$A$可对角化，则存在可逆矩阵$P$，使得$P^{-1}AP = D$，从而$A = PDP^{-1}$. 此时，我们有

$$A^2 = (PDP^{-1})^2 = PDP^{-1}PDP^{-1} = PD^2P^{-1}.$$

利用归纳法，可以进一步证明对任意$k \in \mathbb{N} \cup \{0\}$，均有$A^k = PD^kP^{-1}$. 从而，对任意多项式$q(x)$，

$$q(A) = Pq(D)P^{-1}.$$

**例 4.68.** 计算$A^5$, 其中

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

解. 在之前的例子中，我们已经得到了$A$的对角化$P^{-1}AP = D$. 因此$A^5 = (PDP^{-1})^5 = P^{-1}D^5P$. 我们把具体的计算留给读者. □

**例 4.69.** 斐波那契额数列有下述递推关系式给出：

$$a_0 = 0, \quad a_1 = 1, \quad a_n = a_{n-1} + a_{n-2}\ (n \ge 2).$$

利用$2 \times 2$矩阵求其通项公式.

解. 递推关系式可写成如下矩阵形式：

$$\begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix}.$$

因此

$$\begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}.$$

记$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$，以及$A^n = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$，则有

$$a_n = ca_1 + da_0 = c \cdot 1 + d \cdot 0 = c.$$

现在我们利用对角化的方法计算$A^n$的表达式. 令

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 - \lambda - 1.$$

因此$A$有两个特征值$\lambda_\pm = \frac{1\pm\sqrt{5}}{2}$. 令

$$\mathbf{0} = (\lambda_\pm I - A)\mathbf{v} = \begin{bmatrix} \frac{-1\pm\sqrt{5}}{2} & -1 \\ -1 & \frac{1\pm\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

解方程组可得关于$\lambda_\pm$的特征向量

$$\mathbf{v}_\pm = \begin{bmatrix} 1 \\ \frac{-1\pm\sqrt{5}}{2} \end{bmatrix}.$$

令

$$P = \left[\mathbf{v}_+ \,\vdots\, \mathbf{v}_-\right] = \begin{bmatrix} 1 & 1 \\ \frac{-1+\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \end{bmatrix}, \quad D = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix},$$

则$P^{-1}AP = D$. 因此$A = PDP^{-1}$，进而

$$\begin{aligned} A^n = PD^nP^{-1} &= \begin{bmatrix} 1 & 1 \\ \frac{-1+\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \frac{-1+\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & \left(\frac{1-\sqrt{5}}{2}\right)^n \\ \left(\frac{1+\sqrt{5}}{2}\right)^{n-1} & \left(\frac{1-\sqrt{5}}{2}\right)^{n-1} \end{bmatrix} \left(\frac{1}{-\sqrt{5}} \begin{bmatrix} \frac{-1-\sqrt{5}}{2} & -1 \\ -\frac{-1+\sqrt{5}}{2} & 1 \end{bmatrix}\right). \end{aligned}$$

于是，矩阵$A^n$的左下角元素为

$$\begin{aligned} a_n = c &= \frac{1}{-\sqrt{5}} \left( \left(\frac{1+\sqrt{5}}{2}\right)^{n-1} \frac{-1-\sqrt{5}}{2} + \left(\frac{1-\sqrt{5}}{2}\right)^{n-1} \left(-\frac{-1+\sqrt{5}}{2}\right) \right) \\ &= \frac{1}{\sqrt{5}} \left( \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \right). \end{aligned}$$

□

# 第5章　内积空间

## 5.1　内积

任何一个数域$\mathbb{F}$上的$n$维线性空间均同构于$\mathbb{F}^n$. 然而，欧氏空间$\mathbb{R}^n$中的向量有模长、距离、角度、正交性等. 在抽象的向量空间上，我们是否有相关的概念呢？在这一节中，我们将引进“内积”的概念，它使得我们能够讨论多项式、函数、矩阵等的模长、距离、角度、正交性等.

**定义 5.1.** 设$V$为数域$\mathbb{F}$上的线性空间. 我们称$\langle\cdot,\cdot\rangle : V\times V\to\mathbb{F}$为空间$V$上的内积，是指它满足以下三条性质：

*(i)* 【（共轭）对称性】对任意$\mathbf{u},\mathbf{v}\in V$，有$\langle\mathbf{u},\mathbf{v}\rangle=\overline{\langle\mathbf{v},\mathbf{u}\rangle}$

*(ii)* 【（共轭）双线性性】对任意$\mathbf{u},\mathbf{v},\mathbf{w}\in V$及任意$k,l\in\mathbb{F}$，有$\langle k\mathbf{u}+\ell\mathbf{v},\mathbf{w}\rangle=k\langle\mathbf{u},\mathbf{w}\rangle+\ell\langle\mathbf{v},\mathbf{w}\rangle$.

*(iii)* 【正定性】对任意$\mathbf{v}\in V$，有$\langle\mathbf{v},\mathbf{v}\rangle\geq 0$，且等号成立当且仅当$\mathbf{v}=0$.

这时，我们称$\big(V,\langle\cdot,\cdot\rangle\big)$为内积空间.

在上述定义中，如果$\mathbb{F}=\mathbb{R}$，那么我们可以忽略共轭的部分，此时我们称$V$为实内积空间. 当$\mathbb{F}=\mathbb{C}$时，我们称$V$为酉空间（或复内积空间）. 根据定义，我们总有$\langle\mathbf{0},\mathbf{v}\rangle=\langle\mathbf{u},\mathbf{0}\rangle=0$. 另外，利用定义中的 (i) 和(ii)，可以推出内积对于第二个位置也具有（共轭）线性性：（证明留给读者）

$$\langle\mathbf{w},k\mathbf{u}+l\mathbf{v}\rangle=\overline{k}\langle\mathbf{w},\mathbf{u}\rangle+\overline{l}\langle\mathbf{w},\mathbf{v}\rangle.$$

特别需要强调的是，当$\mathbb{F}=\mathbb{C}$时，从第二个位置提取出来的纯量需要加上共轭.

**例 5.1.** 设$\mathbb{F}=\mathbb{R}$，$V=\mathbb{R}^n$. 对$\mathbf{u}=(u_1,\ldots,u_n)$与$\mathbf{v}=(v_1,\ldots,v_n)$，定义

$$\langle\mathbf{u},\mathbf{v}\rangle=u_1v_1+\ldots+u_nv_n.$$

则$(V,\langle\cdot,\cdot\rangle)$构成实内积空间，我们称其为欧氏空间，而该内积称为欧式内积.

**例 5.2.** 设$\mathbb{F}=\mathbb{C}$，$V=\mathbb{C}^n$. 对$\mathbf{z}=(z_1,\ldots,z_n)$与$\mathbf{v}=(w_1,\ldots,w_n)$，定义

$$\langle\mathbf{z},\mathbf{w}\rangle=z_1\overline{w_1}+\ldots+z_n\overline{w_n}.$$

则$(V,\langle\cdot,\cdot\rangle)$构成酉空间. 在教材上，将该内积空间称为复欧式空间.

在$\mathbb{R}^n$上，除了欧式内积外，还可以构造其它形式的内积.

**例 5.3.** 设$\mathbf{u}=(u_1,u_2)$, $\mathbf{v}=(v_1,v_2)$. 试证$\mathbb{R}^2$上由$\langle\mathbf{u},\mathbf{v}\rangle=u_1v_1+u_1v_2+u_2v_1+2u_2v_2$给出的二元运算$\langle\cdot,\cdot\rangle$构成实内积.

证明.（略.） □

**例 5.4.** 设矩阵$A\in M_n(\mathbb{R})$可逆. 设$\langle\cdot,\cdot\rangle$是$\mathbb{R}^n$上的一个内积. 定义$\langle\cdot,\cdot\rangle_A$为（将$\mathbf{u},\mathbf{v}\in\mathbb{R}^n$看成列矩阵）

$$\langle\mathbf{u},\mathbf{v}\rangle_A=\langle A\mathbf{u},A\mathbf{v}\rangle,\quad(\mathbf{u},\mathbf{v}\in\mathbb{R}^n).$$

证明$\langle\cdot,\cdot\rangle_A$也是$\mathbb{R}^n$上的内积.

证明. 先证明对称性，对于任意$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$，有

$$\langle \mathbf{u}, \mathbf{v} \rangle_A = \langle A\mathbf{u}, A\mathbf{v} \rangle = \langle A\mathbf{v}, A\mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle_A.$$

上式中间的的等号用到了内积$\langle \cdot, \cdot \rangle$的对称性.

再证共轭双线性性，对于$\mathbf{u}_1, \mathbf{u}_2, \mathbf{v} \in \mathbb{R}^n$以及$k_1, k_2 \in \mathbb{R}$，有

$$\begin{aligned} \langle k_1\mathbf{u}_1 + k_2\mathbf{u}_2, \mathbf{v} \rangle_A &= \langle A(k_1\mathbf{u}_1 + k_2\mathbf{u}_2), A\mathbf{v} \rangle = \langle k_1 A\mathbf{u}_1 + k_2 A\mathbf{u}_2, A\mathbf{v} \rangle \\ &= k_1 \langle A\mathbf{u}_1, A\mathbf{v} \rangle + k_2 \langle A\mathbf{u}_2, A\mathbf{v} \rangle = k_1 \langle \mathbf{u}_1, \mathbf{v} \rangle + k_2 \langle \mathbf{u}_2, \mathbf{v} \rangle. \end{aligned}$$

上式中的一步用到了内积$\langle \cdot, \cdot \rangle$的共轭双线性性.

最后我们证明正定性，对于$\mathbf{v} \in \mathbb{R}^n$，我们有

$$\langle \mathbf{v}, \mathbf{v} \rangle_A = \langle A\mathbf{v}, A\mathbf{v} \rangle \geq 0.$$

这里我们用到了内积$\langle \cdot, \cdot \rangle$的正定性，并且可以进一步得到：等号成立当且仅当$A\mathbf{v} = \mathbf{0}$. 由于$A$可逆，这也等价于$\mathbf{v} = \mathbf{0}$. 证毕. □

**定义 5.2.** 设$\big(V, \langle \cdot, \cdot \rangle\big)$是内积空间，则向量$\mathbf{v} \in V$的模长（或范数）定义为

$$\|\mathbf{v}\| = \sqrt{\langle v, v \rangle}.$$

模长为1的向量被称为单位向量. 另外，向量$\mathbf{u}, \mathbf{v} \in V$的距离定义为

$$d(u, v) = \|u - v\|.$$

上述定义很大程度下依赖于内积的正定性，因为向量的模长都应该是非负的，并且仅有零向量的模长为零才是合理的.

**例 5.5.** 在$\mathbb{R}^2$中，取$\mathbf{u} = (1, 0)$，$\mathbf{v} = (0, 1)$. 在下列内积下，分别计算$\|\mathbf{u}\|, \|\mathbf{v}\|$以及$d(\mathbf{u}, \mathbf{v})$:

*(a)* $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_2v_2$;

*(b)* $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_1v_2 + u_2v_1 + 2u_2v_2$.

解.（略.） □

**定理 5.6.** 设$\mathbf{u}, \mathbf{v}$为内积空间中的向量，设$k$为纯量，则有

- ◇ $\|\mathbf{v}\| \geq 0$，且等号成立当且仅当$\mathbf{v} = 0$.
- ◇ $\|k\mathbf{v}\| = |k| \cdot \|\mathbf{v}\|$.
- ◇ $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$.
- ◇ $d(\mathbf{u}, \mathbf{v}) \geq 0$，且等号成立当且仅当$\mathbf{v} = \mathbf{u}$.

以上定理中的四条分别为范数的正定性、齐次性，以及距离的对称性、正定性，它们的推导留给读者进行. 下面我们列出更多关于模长的性质，它们与欧氏空间中的情形类似，因此留给读者自行证明.

**定理 5.7** (平行四边形公式)**.** 设$\mathbf{u}, \mathbf{v}$为实内积中的向量，则有

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

**定理 5.8** (极化恒等式)**.** 设$\mathbf{u}, \mathbf{v}$为内积中的向量. 当$\mathbb{F} = \mathbb{R}$时，我们有

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4} \sum_{k=0}^{1} \left\|\mathbf{u} + (-1)^k \mathbf{v}\right\|^2 = \frac{1}{4}\left(\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2\right).$$

当$\mathbb{F}=\mathbb{C}$时，则有

$$\begin{aligned}\langle\mathbf{u},\mathbf{v}\rangle&=\frac{1}{4}\sum_{k=0}^{3}\|\mathbf{u}+\mathrm{i}^k\mathbf{v}\|^2\\&=\frac{1}{4}\left(\|\mathbf{u}+\mathbf{v}\|^2+\mathrm{i}\|\mathbf{u}+\mathrm{i}\mathbf{v}\|^2-\|\mathbf{u}-\mathbf{v}\|^2-\mathrm{i}\|\mathbf{u}-\mathrm{i}\mathbf{v}\|^2\right)\end{aligned}$$

接下来，让我们来看看大量非欧氏空间的例子.

**例 5.9.** 证明以下二元运算$\langle\cdot,\cdot\rangle$是实线性空间$V$上的内积，并计算相应向量的内积、模长或距离.

*(1)* 设$\mathbb{F}=\mathbb{R}$, $V=C[-\pi,\pi]$，定义$\langle f,g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)g(t)dt$. 取$f(t)=t$，$g(t)=t^2$，计算$\|f\|$, $\|g\|$和$\langle f,g\rangle$.

*(2)* 设$\mathbb{F}=\mathbb{R}$, $V=M_n(\mathbb{R})$，定义$\langle A,B\rangle=\mathrm{tr}(B^TA)$. 取$A=\begin{bmatrix}1&2\\3&4\end{bmatrix}$, $B=\begin{bmatrix}-1&0\\3&2\end{bmatrix}$，计算$\langle A,B\rangle$, $\|A\|$与$d(A,B)$.

*(3)* 设$\mathbb{F}=\mathbb{R}$, $V=P_n$, 定义$\left\langle\sum_{k=0}^n a_kx^k,\ \sum_{k=0}^n b_kx^k\right\rangle=\sum_{k=0}^n a_kb_k$.

*(4)* 设$\mathbb{F}=\mathbb{R}$, $V=P_n$，给定不同的数$x_0,x_1,\ldots,x_n\in\mathbb{R}$，定义$\langle p,q\rangle=\sum\limits_{j=0}^{n}p(x_j)q(x_j)$. 取$n=2$, $x_0=-2$, $x_1=0$, $x_2=2$，以及$p(x)=x^2$, $q(x)=1+x$，计算$\langle p,q\rangle$和$\|p\|$.

*证明.* 这里，我们仅验证(1)、(2)、(4)的正定性. 其余部分留给读者进行.

(1) 对$f\in V$, 利用积分的保号性，得

$$\langle f,f\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}f^2(t)dt\geq 0.$$

且$\langle f,f\rangle=0$当且仅当$\int_{-\pi}^{\pi}f^2(t)dt=0$. 由于$f^2(t)$为连续非负函数，因此还等价于$f^2(t)=0$，或即$f(t)=0$. 正定性得证.

(2) 设$A\in M_n(\mathbb{R})$，则

$$\langle A,A\rangle=\mathrm{tr}(A^TA)=\sum_{i=1}^{n}\sum_{j=1}^{n}(A)_{ij}^2\geq 0.$$

且$\langle A,A\rangle=0$当且仅当$(A)_{ij}=0$ $1\leq i,j\leq n$，当且仅当$A=0$. 正定性得证.

(4) 设$p(x)\in V$，则

$$\langle p,p\rangle=\sum_{j=0}^{n}p^2(x_j)\geq 0.$$

这时，$\langle p,p\rangle=0$等价于$p(x_j)=0$ $(j=1,2,\ldots,n)$. 由于$x_1,x_2,\ldots,x_n$是不同的数，而次数不超过$n-1$次的多项式之多有$n-1$个根，因此也等价于$p(x)$为零多项式. 正定性得证. □

上述例子有对应的复内积版本：

(1') 设$\mathbb{F}=\mathbb{C}$, $V=C_{\mathbb{C}}[-\pi,\pi]$（连续的复值函数全体），定义$\langle f,g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\overline{g(t)}dt$，则$(V,\langle\cdot,\cdot\rangle)$构成一个酉空间.

(2') 设$\mathbb{F}=\mathbb{C}$, $V=M_n(\mathbb{C})$, 定义$\langle A,B\rangle=\mathrm{tr}(B^*A)$，则$(V,\langle\cdot,\cdot\rangle)$构成一个酉空间.

(3') 设$\mathbb{F}=\mathbb{C}$, $V=P_{n,\mathbb{C}}$ (复系数的多项式),定义$\left\langle\sum_{k=0}^n a_kx^k,\ \sum_{k=0}^n b_kx^k\right\rangle=\sum_{k=0}^n a_k\overline{b_k}$.

**例 5.10.** 取$\mathbb{F}=\mathbb{R}$，令

$$V=\left\{(x_1,x_2,\ldots,x_n,\ldots):\sum_{i=1}^{\infty}x_i^2<\infty\right\}.$$

定义

$$\langle(x_i),\,(y_i)\rangle=\sum_{i=1}^{\infty}x_iy_i.$$

则$\langle\cdot,\cdot\rangle$是$V$上的内积.

在$\mathbb{C}^n$上，标准的复内积有如下矩阵形式.

**定义 5.3.** 设$\mathbf{u}, \mathbf{v} \in M_{n\times 1}(\mathbb{C})$为列矩阵，则$\mathbf{u}$与$\mathbf{v}$的内积由下式给出：

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{v}^*\mathbf{u}.$$

**例 5.11.** 设$A \in M_{m\times n}(\mathbb{C})$, $u \in M_{n\times 1}(\mathbb{C})$，而$v \in M_{m\times 1}(\mathbb{C})$. 则

$$(A\mathbf{u})\cdot\mathbf{v} = \mathbf{u}\cdot(A^*\mathbf{v}).$$

证明. 需要注意的是，上式左端的内积是$n$维空间上的，而右端的内积是$m$维空间上的，并不是同一个内积. 我们有

$$(A\mathbf{u})\cdot\mathbf{v} = \mathbf{v}^*(A\mathbf{u}) = (A^*\mathbf{v})^*\mathbf{u} = \mathbf{u}\cdot(A^*\mathbf{v}).$$

□

**定义 5.4.** 矩阵$A \in M_n(\mathbb{C})$被称为自伴的（或厄密特的），是指$A^* = A$.

**例 5.12.** 证明厄密特矩阵的特征值均为实数.

证明. 设$A \in M_n(\mathbb{C})$为厄密特矩阵，设$\lambda$为$A$的一个特征值，而$\mathbf{x} \in \mathbb{C}^n$为其一个对应的特征向量. 则

$$(A\mathbf{x})\cdot\mathbf{x} = (\lambda\mathbf{x})\cdot\mathbf{x} = \lambda(\mathbf{x}\cdot\mathbf{x}) = \lambda\|\mathbf{x}\|^2.$$

另一方面，由前面的例子、厄米特矩阵$A^* = A$以及内积的共轭双线性性，可得

$$(A\mathbf{x})\cdot\mathbf{x} = \mathbf{x}\cdot(A^*\mathbf{x}) = \mathbf{x}\cdot(A\mathbf{x}) = \mathbf{x}\cdot(\lambda\mathbf{x}) = \overline{\lambda}(\mathbf{x}\cdot\mathbf{x}) = \overline{\lambda}\|\mathbf{x}\|^2.$$

因$\mathbf{x} \neq \mathbf{0}$，有$\|\mathbf{x}\|^2 \neq 0$，从而$\lambda = \overline{\lambda}$. 故$\lambda$为实数，得证.

□

## 5.2 角度与正交性

在这一节中，我们总假设$(V, \langle\cdot,\cdot\rangle)$是一个内积空间.

**定理 5.13.** 对任意$\mathbf{u}, \mathbf{v} \in V$，我们有

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

证明. 当$\mathbf{v} = \mathbf{0}$时，结论成立. 下面我们设$\mathbf{v} \neq \mathbf{0}$. 注意到

$$\begin{aligned} 0 &\leq \big\|\|\mathbf{v}\|^2\mathbf{u} - \langle\mathbf{u},\mathbf{v}\rangle\mathbf{v}\big\|^2 = \big\langle\|\mathbf{v}\|^2\mathbf{u} - \langle\mathbf{u},\mathbf{v}\rangle\mathbf{v}, \|\mathbf{v}\|^2\mathbf{u} - \langle\mathbf{u},\mathbf{v}\rangle\mathbf{v}\big\rangle \\ &= \|\mathbf{v}\|^4\langle\mathbf{u},\mathbf{u}\rangle - \|\mathbf{v}\|^2\overline{\langle\mathbf{u},\mathbf{v}\rangle}\langle\mathbf{u},\mathbf{v}\rangle - \|\mathbf{v}\|^2\langle\mathbf{u},\mathbf{v}\rangle\langle\mathbf{v},\mathbf{u}\rangle + \langle\mathbf{u},\mathbf{v}\rangle\overline{\langle\mathbf{u},\mathbf{v}\rangle}\langle\mathbf{v},\mathbf{v}\rangle \\ &= \|\mathbf{v}\|^4\|\mathbf{u}\|^2 - |\langle\mathbf{u},\mathbf{v}\rangle|^2\,\|\mathbf{v}\|^2. \end{aligned}$$

定理得证. □

**推论 5.14.** 设$\mathbf{u}, \mathbf{v} \in V$，以下三角不等式成立：

- ◇ $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.
- ◇ $d(\mathbf{u},\mathbf{v}) \leq d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v})$.

证明. 注意到

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\mathrm{Re}\langle\mathbf{u},\mathbf{v}\rangle + \|\mathbf{v}\|^2. \tag{5.1}$$

由于$|\langle\mathbf{u},\mathbf{v}\rangle| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$. 因此等式右端不超过$(\|\mathbf{u}\| + \|\mathbf{v}\|)^2$. 第一个式子得证. 对于第二个式子，我们有

$$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| \leq \|\mathbf{u}-\mathbf{w}\| + \|\mathbf{w}-\mathbf{v}\| = d(\mathbf{u},\mathbf{w}) + d(\mathbf{w},\mathbf{v}).$$

□

我们总有

$$-1 \le \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \le 1$$

因此可以定义向量$\mathbf{u}$和$\mathbf{v}$之间的夹角$\theta$为

$$\theta = \arccos\left(\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\ \|\mathbf{v}\|}\right).$$

当然，描述两个矩阵或者两个函数之间的角度也并不具备实际的意义，除了一个特殊的角度：$\pi/2$.

**定义 5.5.** *我们称向量$\mathbf{u}$和$\mathbf{v}$是正交的，是指$\langle \mathbf{u}, \mathbf{v}\rangle = 0$.*

**例 5.15.** *验证下述内积空间中的向量是正交的：*

*(1) 设$\mathbb{F} = \mathbb{R}$, $V = M_2(\mathbb{R})$，内积为$\langle A, B\rangle = tr(B^T A)$. 取*

$$A = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}.$$

*(2) 设$\mathbb{F} = \mathbb{R}$, $V = C[-\pi, \pi]$，内积为$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)g(t)dt$. 取*

$$f_m(t) = \sin mt, \quad g_n(t) = \cos nt, \quad h(t) = 1/\sqrt{2},$$

*其中$m, n = 1, 2, 3, \ldots$.*

*(3) 设$\mathbb{F} = \mathbb{C}$, $V = C_{\mathbb{C}}[-\pi, \pi]$（复值连续函数全体），内积为$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\overline{g(t)}dt$. 取$f_n(t) = e^{int}$ $(n \in \mathbb{Z})$.*

解.（略.） □

**定理 5.16** (勾股定理)**.** *设$\mathbf{u}$与$\mathbf{v}$为内积空间中的两个正交的向量，则*

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

证明. 由(5.1)及$\langle \mathbf{u}, \mathbf{v}\rangle = 0$可得. □

**例 5.17.** *设$\mathbb{F} = \mathbb{R}$, $V = P_2$，内积为$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx$. 试以$p(x) = x$和$q(x) = x^2$验证勾股定理.*

解.（略.） □

**定义 5.6.** *设$W$是内积空间$V$的子空间，定义$W$的正交补为*

$$W^{\perp} = \big\{\mathbf{v} : \langle \mathbf{v}, \mathbf{w}\rangle = 0\,(\forall \mathbf{w} \in W)\big\}.$$

与欧氏空间的情形类似，正交补具有下述性质：

**定理 5.18.** *设$W$是内积空间$V$的子空间.*

*(1) $W^{\perp}$是$V$的子空间.*

*(2) $W^{\perp} \cap W = \{\mathbf{0}\}$.*

*(3) $W + W^{\perp} = V$.*

*(4) 若$V$是有限维的，则有$(W^{\perp})^{\perp} = W$.*

特别地，我们有$\{\mathbf{0}\}^{\perp} = W$，而$W^{\perp} = \{\mathbf{0}\}$. 需要注意的是，上述性质(4)在无限维空间中并不成立. 例如，取实内积空间

$$V = \{(x_n)_{n-1}^{\infty} : \sum_{n=1}^{\infty} x_n^2 < \infty\},$$

其内积由

$$\langle (x_n)_{n-1}^{\infty}, (y_n)_{n=1}^{\infty}\rangle = \sum_{n=1}^{\infty} x_n y_n$$

给出. 取$\mathbf{e}_i$ $(i=1,2,\ldots)$为第$i$个分量为1，其余分量为零的数列. 取子空间$W=\mathsf{span}\{\mathbf{e}_i : i=1,2,\ldots\}$.这个线性包络仅包含了相应向量的有限线性组合，其中的向量至多在有限个分量上非零，所以$W\neq V$. 现在我们来说明$W^{\perp}=\{0\}$. 事实上，任取$(x_n)_{n=1}^{\infty}\in W^{\perp}$，由正交补的定义知

$$0=\langle (x_n), \mathbf{e}_i\rangle = x_i, \quad (i=1,2,\ldots).$$

因此$(x_n)_{n=1}^{\infty}=(0)_{n=1}^{\infty}=\mathbf{0}$. 最后，我们得到$(W^{\perp})^{\perp}=\{\mathbf{0}\}^{\perp}=V$. 对于无穷维的情况，性质(4)一般对“闭”子空间$W$成立，其具体的含义不属于线性代数I的内容.

**例 5.19.** 设$W$为$\mathbb{R}^6$的子空间，由下述4个向量张成：

$$\mathbf{w}_1=(1,3,-2,0,2,0), \quad \mathbf{w}_2=(2,6,-5,-2,4,-3)$$

$$\mathbf{w}_3=(0,0,5,10,0,15), \quad \mathbf{w}_4=(2,6,0,8,4,18).$$

求$W^{\perp}$在欧氏空间$\mathbb{R}^6$中的一组基.

解.（略.） □

**例 5.20.** 设$\mathbb{F}=\mathbb{R}$, $V=C[-1,1]$，内积取为

$$\langle f,g\rangle=\int_{-1}^{1} f(x)g(x)dx, \quad (f,g\in V).$$

设$V_e$、$V_o$分别为$V$中所有偶函数、奇函数所构成的子空间.证明$V_o^{\perp}=V_e$.

证明. 先证“$V_e\subseteq V_o^{\perp}$”. 设$f\in V_e$. 对任意$g\in V_o$，函数$fg$为奇函数，因此

$$\langle f,g\rangle=\int_{-1}^{1} f(x)g(x)dx=0.$$

故$f\in V_o^{\perp}$.

再证“$V_o^{\perp}\subseteq V_e$”. 设$f\in V_o^{\perp}$. 记$f=f_o+f_e$，其中

$$f_e(x)=\frac{f(x)+f(-x)}{2}\in V_e, \quad f_o(x)=\frac{f(x)-f(-x)}{2}\in V_o.$$

由$f\in V_o^{\perp}$，对任意$\langle f,f_o\rangle=0$. 而由证明的前半部分，知$\langle f_e,f_o\rangle=0$. 因而

$$0=\langle f,f_o\rangle=\langle f_e,f_o\rangle+\langle f_o,f_o\rangle=\|f_o\|^2.$$

于是$f_o=0$，进而$f=f_e\in V_e$. 证毕. □

## 5.3 Gram-Schmidt正交化与$QR$-分解

在第三章中，基给出了线性空间的“坐标系”. 到这一章，我们有了正交的概念，自然也应该引入“直角坐标系”.

**定义 5.7.** 设$V$为内积空间，而$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_r\}$ $(r\geq 2)$是$V$中的一个子集.

若$S$中的向量两两正交，则称$S$为正交集. 若正交集$S$中的向量都是单位向量，则称$S$为标准正交集.

若$V$的一组基$S$是正交集（或标准正交集），则我们称$S$为正交基（或标准正交基）.

**例 5.21.** 在欧式空间$\mathbb{R}^3$中，取$S=\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$，其中

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

验证$S$为正交基，并通过将其中的向量单位化得到一组标准正交基.

解.（略.） □

**例 5.22.** 验证下列集合$B$是给定内积空间$V$中的标准正交集.

*(1)* 设$\mathbb{F}=\mathbb{R}$, $V=M_n(\mathbb{R})$, 内积为$\langle A,B\rangle=tr(B^TA)$，取其标准基$S=\{E_{ij}:1\le i,j\le n\}$.

*(2)* 设$\mathbb{F}=\mathbb{R}$, $V=P_n$,内积为$\left\langle\sum_{k=0}^n a_kx^k,\ \sum_{k=0}^n b_kx^k\right\rangle=\sum_{k=0}^n a_kb_k$. 取标准基$B=\{1,x,x^2,\ldots,x^n\}$.

*(3)* 设$\mathbb{F}=\mathbb{R}$, $V=C[-\pi,\pi]$，内积为$\langle f,g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)g(t)dt$. 给定$k,l\in\mathbb{N}$，取集合

$$B=\left\{\frac{1}{\sqrt{2}},\sin x,\sin 2x,\ldots,\sin kx,\cos x,\cos 2x,\ldots,\cos lx\right\}.$$

*(4)* 设$\mathbb{F}=\mathbb{C}$, $V=C_{\mathbb{C}}[-\pi,\pi]$, 内积为$\langle f,g\rangle=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)\overline{g(t)}dt$. 给定$N\in\mathbb{N}$,取集合$B=\left\{e^{\mathrm{i}nx}:|n|\le N\right\}$.

解. 这里只验证(1)的情形. 记

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

注意到

$$\langle E_{ij},E_{kl}\rangle=\mathrm{tr}(E_{kl}^TE_{ij})=\mathrm{tr}(E_{lk}E_{ij}).$$

由于$E_{lk}E_{ij}=\delta_{ik}E_{lj}$（左侧矩阵的1出现在第$k$列，右侧矩阵的1出现在第$i$行，仅在$k=i$时这两个1才能够对应相乘，相乘所得的1的位置在第$l$行第$j$列）. 故

$$\langle E_{ij},E_{kl}\rangle=\delta_{ik}\mathrm{tr}(E_{lj})=\delta_{ik}\delta_{lj}=\begin{cases}1, & 若(i,j)=(k,l),\\ 0 & 若(i,j)\neq(k,l).\end{cases}$$

因此$S$构成$V$的一组标准正交基.

□

下面我们来推导一些有关正交集的性质.

**定理 5.23.** 若$S$是内积空间$(V,\langle\cdot,\cdot\rangle)$中有限个非零向量所组成的正交集，则$S$线性无关.

证明. 不妨设$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$. 令$k_1\mathbf{v}_1+k_2\mathbf{v}_2+\ldots+k_n\mathbf{v}_n=\mathbf{0}$. 则对$1\le j\le n$，均有

$$0=\langle\mathbf{0},\mathbf{v}_j\rangle=\langle k_1\mathbf{v}_1+k_2\mathbf{v}_2+\ldots+k_n\mathbf{v}_n,\mathbf{v}_j\rangle=k_1\langle\mathbf{v}_1,\mathbf{v}_j\rangle+k_2\langle\mathbf{v}_2,\mathbf{v}_j\rangle+\ldots+k_n\langle\mathbf{v}_n,\mathbf{v}_j\rangle.$$

因$S$是正交集，当$i\neq j$时，有$\langle\mathbf{v}_i,\mathbf{v}_j\rangle=0$. 于是

$$0=k_j\langle\mathbf{v}_j,\mathbf{v}_j\rangle=k_j\|\mathbf{v}_j\|^2.$$

由于$\mathbf{v}_j\ (1\le j\le n)$为非零向量，故$\|\mathbf{v}_j\|^2\neq 0$，从而$k_j=0$. 所以$S$是线性无关的. □

**定理 5.24.** 设$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$为内积空间$(V,\langle\cdot,\cdot\rangle$的一组标准基，而$\mathbf{u}\in V$. 则有

$$\mathbf{u}=\frac{\langle\mathbf{u},\mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1+\frac{\langle\mathbf{u},\mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2+\cdots+\frac{\langle\mathbf{u},\mathbf{v}_n\rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n.$$

证明. 由于$S$是$V$的一组基，因此存在$c_1, c_2, \ldots, c_n \in \mathbb{F}$，使得

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n.$$

为了确定系数$c_j$ $(1 \le j \le n)$，我们在上式两端与$\mathbf{v}_j$作内积，利用正交性可得

$$\langle \mathbf{u}, \mathbf{v}_j\rangle = \langle c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n, \mathbf{v}_j\rangle = c_1\langle \mathbf{v}_1, \mathbf{v}_j\rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_j\rangle + \ldots + c_n\langle \mathbf{v}_n, \mathbf{v}_j\rangle = c_j\langle \mathbf{v}_j, \mathbf{v}_j\rangle.$$

定理得证. □

需要注意的是，在线性组合系数分子上的内积里，向量$\mathbf{u}$出现在第一个位置，如果把$\mathbf{u}$换到第二个位置而$\mathbb{F} = \mathbb{C}$，则需要加上共轭. 另外，当$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$为标准正交基时，有

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle\mathbf{v}_2 + \cdots \langle \mathbf{u}, \mathbf{v}_n\rangle\mathbf{v}_n.$$

**例 5.25.** 令$\mathbf{v}_1 = (0, 2, 0)$, $\mathbf{v}_2 = (3, 0, 3)$, $\mathbf{v}_3 = (-4, 0, 4)$，而$S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

*(1)* 验证$S$构成欧氏空间$\mathbb{R}^3$中的一组正交基.

*(2)* 求$[\mathbf{u}]_B$，其中$\mathbf{u} = (1, 2, 4)$.

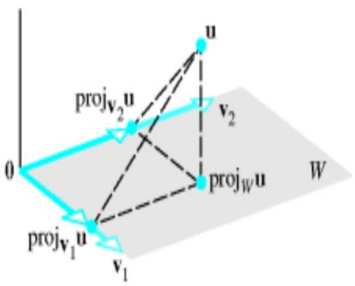
解.（略.） □

在第二章中，我们已经证明了欧氏空间中特殊情形的投影定理. 下述定理是投影定理的一般形式，我们省去其证明（利用直和证明会比较方便）.

**定理 5.26** (投影定理)**.** 设$W$是有限维内积空间$V$的子空间，则$V$中的任意向量$\mathbf{u}$可以唯一地分解为

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2,$$

其中$\mathbf{w}_1 \in W$，而$\mathbf{w}_2 \in W^\perp$.




在上述分解式中，我们将使用记号

$$\mathrm{proj}_W(\mathbf{u}) := \mathbf{w}_1, \quad \mathrm{proj}_{W^\perp}(\mathbf{u}) = \mathbf{w}_2.$$

即有

$$\mathbf{u} = \mathrm{proj}_W(\mathbf{u}) + \mathrm{proj}_{W^\perp}(\mathbf{u}) = \mathbf{w}_2.$$

值得一题的是，将$\mathrm{proj}_W$看成$V$上的一个算子，它是线性的. 由于上述的$\mathbf{u}_1$的分解为

$$\mathbf{u}_1 = \mathrm{proj}_W(\mathbf{u}_1) + \mathrm{proj}_{W^\perp}(\mathbf{u}_2) = \mathbf{u}_1 + \mathbf{0}.$$

该正交投影算子满足

$$(\mathrm{proj}_W \circ \mathrm{proj}_W)(\mathbf{u}) = \mathrm{proj}_W(\mathbf{u}_1) = \mathbf{u}_1 = \mathrm{proj}_W(\mathbf{u}).$$

即

$$\mathrm{proj}_W \circ \mathrm{proj}_W = \mathrm{proj}_W.$$

**定理 5.27.** 设$W$是内积空间$(V,\langle\cdot,\cdot\rangle)$的一个子空间，而$S=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_r\}$是$W$的一组正交基. 则对任意$\mathbf{v}\in V$，均有

$$proj_W(\mathbf{v})=\frac{\langle\mathbf{u},\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1+\frac{\langle\mathbf{u},\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2+\ldots+\frac{\langle\mathbf{u},\mathbf{w}_r\rangle}{\|\mathbf{w}_r\|^2}\mathbf{w}_r.$$

解. 注意到$\mathrm{proj}_W(\mathbf{v})$，根据定理5.24，我们有

$$\mathrm{proj}_W(\mathbf{v})=\frac{\langle\mathrm{proj}_W(\mathbf{v}),\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1+\frac{\langle\mathrm{proj}_W(\mathbf{v}),\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2+\cdots+\frac{\langle\mathrm{proj}_W(\mathbf{v}),\mathbf{w}_r\rangle}{\|\mathbf{w}_r\|^2}\mathbf{w}_r.$$

另外，由于$\mathrm{proj}_{W^\perp}(\mathbf{v})\in W^\perp$，而$\mathbf{w}_j\in W$ $(1\le j\le r)$，所以$\langle\mathrm{proj}_{W^\perp}(\mathbf{v}),\mathbf{w}_j\rangle=0$，进而有

$$\langle\mathbf{v},\mathbf{w}_j\rangle=\langle\mathrm{proj}_W(\mathbf{v})+\mathrm{proj}_{W^\perp}(\mathbf{v}),\mathbf{w}_i\rangle=\langle\mathrm{proj}_W(\mathbf{v}),\mathbf{w}_i\rangle+\langle\mathrm{proj}_{W^\perp}(\mathbf{v}),\mathbf{w}_j\rangle=\langle\mathrm{proj}_W(\mathbf{v}),\mathbf{w}_i\rangle.$$

定理得证. □

当上述定理中的$S=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_r\}$为标准正交基时，则有

$$\mathrm{proj}_W\mathbf{v}=\langle\mathbf{u},\mathbf{w}_1\rangle\mathbf{w}_1+\langle\mathbf{v},\mathbf{w}_2\rangle\mathbf{w}_2+\cdots,+\langle\mathbf{v},\mathbf{w}_r\rangle\mathbf{w}_r.$$

**例 5.28.** 设$\mathbb{F}=\mathbb{R}$, $V=C[-\pi,\pi]$，内积为

$$\langle f,g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)g(t)dt,\quad(\forall f,g\in V).$$

求向量$h(t)=t$在子空间$W=\mathit{span}\{\cos t,\sin t\}$上的正交投影.

解. 记$f(t)=\cos t$, $g(t)=\sin t$，可验证$\{f(t),g(t)\}$是标准正交的. 因此，我们有

$$\mathrm{proj}_W(h(t))=\langle h,f\rangle f(t)+\langle h,g\rangle g(t).$$

利用奇偶性，计算可得

$$\langle h,f\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}t\cos tdt=0,$$

$$\langle h,g\rangle=\frac{1}{\pi}\int_{-\pi}^{\pi}t\sin tdt=-\frac{1}{\pi}t\cos t\Big|_{-\pi}^{\pi}+\frac{1}{\pi}\int_{-\pi}^{\pi}\cos tdt=2.$$

故$\mathrm{proj}_W(h(t))=2\sin t$. □

我们虽然已经有了正交基的概念，但是还是需要再问一问以下两个问题：
(i) 对任何一个有限维内积空间$V$，一定存在正交基吗？
(ii) 若存在，如何寻找一组正交基？
在下述定理的推导过程中，我们同时回答这两个问题.

**定理 5.29.** 每个有限维内积空间$V$均存在标准正交基.

不妨设$\dim(V)=n$，从而可以找到$V$的一组基$S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$. 我们先通过$S$构造一组正交基.
取$\mathbf{u}_1=\mathbf{v}_1$.
取$\mathbf{u}_2=\mathbf{v}_2-\frac{\langle\mathbf{v}_2,\mathbf{u}_1\rangle}{\|\mathbf{u}_1\|^2}\mathbf{u}_1$.
取$\mathbf{u}_3=\mathbf{v}_3-\frac{\langle\mathbf{v}_3,\mathbf{u}_1\rangle}{\|\mathbf{u}_1\|^2}\mathbf{u}_1-\frac{\langle\mathbf{v}_3,\mathbf{u}_2\rangle}{\|\mathbf{u}_2\|^2}\mathbf{u}_2$.
......
取$\mathbf{u}_n=\mathbf{v}_n-\frac{\langle\mathbf{v}_n,\mathbf{u}_1\rangle}{\|\mathbf{u}_1\|^2}\mathbf{u}_1-\ldots-\frac{\langle\mathbf{v}_n,\mathbf{u}_{n-1}\rangle}{\|\mathbf{u}_{n-1}\|^2}\mathbf{u}_{n-1}$.

这时，我们得到$V$的一组正交基$S'=\{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n\}$. 接下来，我们通过单位化，可以得到标准正交基，具体过程为：令$\mathbf{w}_i=\frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$ $(1\le i\le n)$，得到$V$的标准正交基$S''=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_n\}$. 以上方法被称为Gram-Schmidt正交化过程.

之所以$S'$为正交基，是因为我们取第$r$ $(2 \le r \le n)$个向量$\mathbf{u}_r$时，我们的取法为

$$\mathbf{u}_r = \mathrm{proj}_{\mathsf{span}\{\mathbf{u}_1,\dots,\mathbf{u}_{r-1}\}^\perp}(\mathbf{v}_r) = \mathbf{u}_r - \mathrm{proj}_{\mathsf{span}}\{\mathbf{u}_1,\dots,\mathbf{u}_{r-1}\}(\mathbf{v}_r).$$

因此$\mathbf{u}_r$与$\mathbf{u}_1,\dots,\mathbf{u}_{r-1}$均正交. 特别地，我们有

$$\mathbf{u}_r \in \mathsf{span}\{\mathbf{u}_1,\dots,\mathbf{u}_{r-1}\}^\perp = \mathsf{span}\{\mathbf{v}_1,\dots,\mathbf{v}_{r-1}\}^\perp.$$

**例 5.30.** 设$\mathbb{F}=\mathbb{R}$，$V=P_3$，内积为

$$\langle p,q\rangle = \int_{-1}^{1} p(x)q(x)dx, \quad (\forall p,q\in P_2).$$

对标准基$\{1,x,x^2\}$使用*Gram-Schmidt*正交化过程，将其转化为正交基$\{q_1(x),q_2(x),q_3(x)\}$.

*解.* 记$p_1(x)=1$, $p_2(x)=x$, $p_3(x)=x^2$. 首先进行正交化，取

$$\begin{aligned}
q_1(x) &= p_1(x) = 1,\\
q_2(x) &= p_2(x) - \frac{\langle p_2,q_1\rangle}{\|q_1\|^2}q_1(x),\\
q_3(x) &= p_3(x) - \frac{\langle p_3,q_1\rangle}{\|q_1\|^2}q_1(x) - \frac{\langle p_3,q_2\rangle}{\|q_2\|^2}q_2(x).
\end{aligned}$$

将内积、模长都代入计算后，得到

$$h_1(x)=1,\quad h_2(x)=x,\quad h_3(x)=x^2-\frac{1}{3}.$$

□

如果需要，还可以将上述正交基进一步化为标准正交基. 若将$P_3$更换为$P_n$，我们可以用同样的方式得到一组正交基$\{q_1(x),q_2(x),\dots,q_n(x)\}$，这些$q_j(x)$被称为勒让德多项式.

利用Gram-Schmidt正交化过程以及基扩充定理，可以得到下述结论.

**推论 5.31.** 设$W$为有限维内积空间. 则

*(a)* 任何一个非零向量组成的$W$中的正交集可以被扩充为$W$的一组正交基.

*(b)* 任何一个非零向量组成的$W$中的标准正交集可以被扩充为$W$的一组标准正交基.

================== 分割线:创艺(MATH1455)同学的讲义到此为止 ==================

在计算机中，为了更加节省的储存和更加快捷的计算，我们往往要把矩阵作一些分解. 利用Gram-Schmidt正交化过程，我们可对矩阵作$QR$-分解.

**定理 5.32.** 设$A\in M_{m\times n}$满足$\mathsf{rank}(A)=n$. 则$A$可分解为$A=QR$，其中矩阵$Q\in M_{m\times n}$的列是标准正交的，而$R$是可逆上三角矩阵.

定理中$\mathsf{rank}(A)=n$意味着$A$是满秩的，且其列数不超过行数. 作$QR$-分解的具体过程是这样的：记

$$A = \left[\begin{array}{c:c:c:c} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{array}\right].$$

其中的列是线性无关的. 设$\mathbf{q}_1,\dots,\mathbf{q}_n$是对向量组$\mathbf{c}_1,\dots,\mathbf{c}_n$使用Gram-Schmidt正交化过程所得到的. 则有

$$A = \left[\begin{array}{c:c:c:c} \\ \mathbf{q}_1 & \mathbf{q}_2 & \dots & \mathbf{q}_n \\ \\ \end{array}\right]\begin{bmatrix} \langle \mathbf{c}_1,\mathbf{q}_1\rangle & \langle \mathbf{c}_2,\mathbf{q}_1\rangle & \cdots & \langle \mathbf{c}_n,\mathbf{q}_1\rangle \\ & \langle \mathbf{c}_2,\mathbf{q}_2\rangle & \cdots & \langle \mathbf{c}_n,\mathbf{q}_1\rangle \\ & & \ddots & \vdots \\ & & & \langle \mathbf{c}_n,\mathbf{q}_n\rangle \end{bmatrix} := QR.$$

事实上，根据定理5.24，我们有

$$\mathbf{c}_j = \langle \mathbf{c}_j, \mathbf{q}_1\rangle \mathbf{q}_1 + \langle \mathbf{c}_j, \mathbf{q}_2\rangle \mathbf{q}_2 + \ldots + \langle \mathbf{c}_j, \mathbf{q}_n\rangle \mathbf{q}_n, \quad (1 \le j \le n).$$

将上式写成矩阵形式，可得

$$A = \left[\begin{array}{c:c:c:c} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n \end{array}\right] = \left[\begin{array}{c:c:c:c} \mathbf{q}_1 & \mathbf{q}_2 & \ldots & \mathbf{q}_n \end{array}\right] \begin{bmatrix} \langle \mathbf{c}_1, \mathbf{q}_1\rangle & \langle \mathbf{c}_2, \mathbf{q}_1\rangle & \cdots & \langle \mathbf{c}_n, \mathbf{q}_1\rangle \\ \langle \mathbf{c}_1, \mathbf{q}_2\rangle & \langle \mathbf{c}_2, \mathbf{q}_2\rangle & \cdots & \langle \mathbf{c}_n, \mathbf{q}_2\rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{c}_1, \mathbf{q}_n\rangle & \langle \mathbf{c}_2, \mathbf{q}_n\rangle & \cdots & \langle \mathbf{c}_n, \mathbf{q}_n\rangle \end{bmatrix}.$$

在Gram-Schmidt正交化过程中，对$2 \le r \le n$，我们有

$$\mathbf{q}_r \in \mathsf{span}\{\mathbf{c}_1, \ldots, \mathbf{c}_{r-1}\}^{\perp}.$$

因此当$1 \le j < i \le n$时，有$\langle \mathbf{c}_i, \mathbf{q}_j\rangle = 0$. 得到的矩阵是上三角的. 利用矩阵乘法与秩的关系，可以推出$R$是可逆的.

**例 5.33.** *求矩阵$A$的$QR$-分解，其中*$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$.

*解.* 记$A = [\mathbf{c}_1 \vdots \mathbf{c}_2 \vdots \mathbf{c}_3]$. 我们先将$\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$正交化，计算后得到

$$\mathbf{u}_1 = \mathbf{c}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\mathbf{u}_2 = \mathbf{c}_2 - \frac{\mathbf{c}_2 \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 = \begin{bmatrix} -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}.$$

$$\mathbf{u}_3 = \mathbf{c}_3 - \frac{\mathbf{c}_3 \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 - \frac{\mathbf{c}_3 \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2}\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1/2 \\ 1/2 \end{bmatrix}.$$

再将$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$单位化，有

$$\mathbf{q}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \quad \mathbf{q}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} = \begin{bmatrix} -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathbf{q}_3 = \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

这时，令

$$Q = \left[\begin{array}{c:c:c} 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \end{array}\right], \quad R = \begin{bmatrix} \mathbf{c}_1 \cdot \mathbf{q}_1 & \mathbf{c}_2 \cdot \mathbf{q}_1 & \mathbf{c}_3 \cdot \mathbf{q}_1 \\ & \mathbf{c}_2 \cdot \mathbf{q}_2 & \mathbf{c}_3 \cdot \mathbf{q}_2 \\ & & \mathbf{c}_3 \cdot \mathbf{q}_3 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 2/\sqrt{3} & 1/\sqrt{3} \\ & 2/\sqrt{6} & 1/\sqrt{6} \\ & & 1/\sqrt{2} \end{bmatrix}.$$

便有$A = QR$. □

## 5.4 正交矩阵与酉矩阵

有了正交基，我们进一步考虑两组正交基之间的转换. 先考虑特殊的情形，从欧氏空间$\mathbb{R}^n$的一组标准正交基到标准基的过渡矩阵是怎样的？

**定义 5.8.** 设$A \in M_n(\mathbb{R})$. 我们称$A$为正交阵，是指它满足下述四个等价的命题:

*(1)* $A^TA = AA^T = I_n$;

*(2)* $A^{-1} = A^T$;

*(3)* $A$的行向量构成欧式空间$\mathbb{R}^n$的一组标准正交基;

*(4)* $A$的列向量构成欧式空间$\mathbb{R}^n$的一组标准正交基.

等价性的证明. 我们主要证明(4)等价于$A^TA = I_n$,其余部分的等价性留给读者. 记$A$的各列分别为$\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$. 则有

$$A^TA = \begin{bmatrix} \mathbf{c}_1^T \\ \hdashline \mathbf{c}_2^T \\ \hdashline \vdots \\ \hdashline \mathbf{c}_n^T \end{bmatrix} [\, \mathbf{c}_1 \,\vdots\, \mathbf{c}_2 \,\vdots\, \dots \,\vdots\, \mathbf{c}_n \,] = [\mathbf{c}_i^T \mathbf{c}_j]_{n\times n} = [\langle \mathbf{c}_j, \mathbf{c}_i \rangle]_{n\times n}.$$

因此，$\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$构成$\mathbb{R}^n$的标准正交基，当且仅当$\langle \mathbf{c}_j, \mathbf{c}_i \rangle = \delta_{ij}$ $(1 \le i, j \le n)$，当且仅当$A^TA = I_n$. □

从上述等价性定理可知，若$B = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$是$\mathbb{R}^n$的一组标准正交基，则有

$$P_{S\leftarrow B} = [\, \mathbf{c}_1 \,\vdots\, \mathbf{c}_2 \,\vdots\, \dots \,\vdots\, \mathbf{c}_n \,],$$

这是一个正交矩阵.

**例 5.34.** 验证矩阵

$$\begin{bmatrix} 3/7 & 2/7 & 6/7 \\ -6/7 & 3/7 & 2/7 \\ 2/7 & 6/7 & -3/7 \end{bmatrix}$$

是正交阵.

解.（略.） □

**例 5.35.** 记$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ 矩阵$R_\theta$, $R_\theta^T$, $R_\theta^{-1}$, $R_\theta R_{\theta'}$是否正交？

解.（略.） □

下面，我们介绍与实矩阵中正交阵所对应的——复矩阵中酉矩阵. 酉矩阵在复欧氏空间$\mathbb{C}^n$上所起的作用，与正交阵在欧氏空间$\mathbb{R}^n$上起的作用是一样的，它们的定义与性质也都是相似的.

**定义 5.9. Definition.** 设$A \in M_n(\mathbb{C})$. 我们称$A$为酉矩阵，是指它满足下述四个等价的命题:

*(1)* $A^*A = AA^* = I_n$;

*(2)* $A^{-1} = A^*$;

*(3)* $A$的行向量构成复欧式空间$\mathbb{C}^n$的一组标准正交基;

*(4)* $A$的列向量构成复欧式空间$\mathbb{C}^n$的一组标准正交基.

**例 5.36.** 证明1阶方阵$[z] \in M_1(\mathbb{C})$是酉矩阵当且仅当$|z| = 1$.

证明. 矩阵$[z]$是酉矩阵，当且仅当$[z]^*[z] = I_1$，即$\overline{z}z = 1$，这等价于$|z| = 1$. □

下面我们来探究一下正交阵和酉矩阵的性质. 我们给出酉矩阵情形的证明，对正交矩阵的证明是类似的.

**定理 5.37.** 设$A, B \in M_n(\mathbb{R})$均为正交阵，则

- $A^T$是正交矩阵;
- $A^{-1}$是正交矩阵;
- $AB$是正交矩阵;
- $|det(A)| = 1$.

**例 5.38.** 设$A, B \in M_n(\mathbb{R})$均为酉矩阵，则

- $A^*$是酉矩阵;
- $A^{-1}$是酉矩阵;
- $AB$是酉矩阵;
- $|det(A)| = 1$.

证明. 由于$A, B$为酉矩阵，因此$A^*A = B^*B = I_n$，且$A^{-1} = A^*$. 从定义可以得到$A^*$和$A^{-1}$均为酉矩阵. 另外，我们有

$$(AB)^*(AB) = B^*(A^*A)B = B^*B = I_n,$$

所以$AB$也是酉矩阵. 最后，我们有

$$1 = \det(I_n) = \det(A^*A) = \det(A^*)\det(A) = \overline{\det(A^T)}\det(A) = |\det(A)|^2.$$

故$|\det(A)| = 1$. □

由以上两个定理可知，正交矩阵的行列式只能取±1，而酉矩阵的行列式可以写成$e^{i\theta}$，其中$\theta \in [0, 2\pi)$. 另外，设$B_1$, $B_2$是$\mathbb{F}^n$的两组标准正交基，则

$$P_{B_2 \leftarrow B_1} = P_{S \leftarrow B_2}^{-1} P_{S \leftarrow B_1}.$$

因此过渡矩阵$P_{B_2 \leftarrow B_1}$也是正交/酉矩阵.

接下来，我们探究一下正交阵和酉矩阵作为线性变换的映射效果. 我们给出正交阵情形的证明，对酉矩阵的证明是类似的.

**定理 5.39.** 设$A \in M_n(\mathbb{R})$，则以下命题等价：

*(1)* $A$是正交阵.

*(2)* 对任意向量$\mathbf{x} \in \mathbb{R}^n$，有$\|A\mathbf{x}\| = \|\mathbf{x}\|$.

*(3)* 对任意向量$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，有$A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

证明. “(1)⇒(3)”：由于$A$是正交的，因此$A^TA = I_n$. 从而对任意$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，均有

$$(A\mathbf{x}) \cdot (A\mathbf{y}) = (A\mathbf{y})^T(A\mathbf{x}) = \mathbf{y}^T(A^TA)\mathbf{x} = \mathbf{y}^T\mathbf{x} = \mathbf{x} \cdot \mathbf{y}.$$

“(3)⇒(1)”：设$A$的各列分别为$\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$，则有$A\mathbf{e}_i = \mathbf{c}_i$ $(1 \le i \le n)$. 取(3)中的$\mathbf{x}, \mathbf{y}$分别为$\mathbf{e}_j, \mathbf{e}_i$ $(1 \le i, j \le n)$，得

$$\mathbf{c}_j \cdot \mathbf{c}_i = (A\mathbf{e}_j) \cdot (A\mathbf{e}_i) = \mathbf{e}_j \cdot \mathbf{e}_i = \delta_{ij}.$$

因此$A$的列构成$\mathbb{R}^n$的一组标准正交基，即$A$为正交阵.

“(3)⇒(2)”是显然的，下面我们利用极化恒等式证明“(2)⇒(3)”：对任意$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，有

$$\begin{aligned} A\mathbf{x} \cdot A\mathbf{y} &= \frac{1}{4}\big(\|A\mathbf{x} + A\mathbf{y}\| - \|A\mathbf{x} - A\mathbf{y}\|\big) = \frac{1}{4}\big(\|A(\mathbf{x} + \mathbf{y})\| - \|A(\mathbf{x} - \mathbf{y})\|\big) \\ &= \frac{1}{4}\big(\|\mathbf{x} + \mathbf{y}\| - \|\mathbf{x} - \mathbf{y}\|\big) = \mathbf{x} \cdot \mathbf{y}. \end{aligned}$$

在上式中，第一个和最后一个等号用到了极化恒等式，倒数第二个等号用到了(2)的条件. □

从上述定理可以看出，正交矩阵作为线性变换是保持向量模长与向量之间的角度的，结合行列式等于1，它还是保持面积不变的. 特别地，当行列式为1时，它保持定向不变，因此可以理解为$\mathbb{R}^n$中的旋转. 当行列式为−1时，除了旋转还复合了一个镜面反射.

正交矩阵可能没有特征值，比如平面上的旋转$R_\theta$ $(0 < \theta < \pi)$. 当一个正交矩阵存在特征值与特征向量时，它的映射效果一定保持该特征向量的长度不变，因此对应的特征值只可能取±1.

**推论 5.40.** 设$A \in M_n(\mathbb{R})$为正交阵. 若$\lambda$是$A$的一个特征值，则$\lambda = \pm 1$.

我们把证明细节留到后续酉矩阵对应的定理. 在我们的日常生活中，地球仪有一条固定轴是不动的，在与轴垂直的平面上展现了2维的旋转效果. 这条固定轴对应了一个特征值为1的特征向量.

思考题：设$A \in M_n(\mathbb{R})$为正交阵，设$n$为奇数，且$\det(A) = 1$. 证明1必为$A$的特征值.

**定理 5.41.** 设$A \in M_n(\mathbb{C})$，则以下命题等价：

*(1)* $A$是酉矩阵.

*(2)* 对任意向量$\mathbf{x} \in \mathbb{C}^n$，有$\|A\mathbf{x}\| = \|\mathbf{x}\|$.

*(3)* 对任意向量$\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$，有$A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

**推论 5.42.** 设$A \in M_n(\mathbb{C})$为酉矩阵. 则$A$的特征值$\lambda$满足$|\lambda| = 1$.

*证明.* 设$A\mathbf{x} = \lambda\mathbf{x}$，其中$\mathbf{x} \neq 0$. 利用酉矩阵保持向量长度不变的性质，得

$$\|\mathbf{x}\| = \|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\|.$$

由于$\|\mathbf{x}\| \neq 0$，我们得到结论$|\lambda| = 1$. □

在本节的最后，我们看看在内积空间中，向量在标准正交基下的运算性质.

**定理 5.43.** 设$S$为$n$维内积空间$(V, \langle\cdot,\cdot\rangle)$的一组标准正交基. 记$(\mathbf{u})_S = (u_1, u_2, \cdots, u_n)$，而$(\mathbf{v})_S = (v_1, v_2, \cdots, v_n)$. 则有

- ◇ $\langle \mathbf{u}, \mathbf{v}\rangle = u_1\overline{v_1} + \cdots + u_n\overline{v_n}$;
- ◇ $\|\mathbf{u}\| = \sqrt{|u_1|^2 + \cdots |u_n|^2}$;
- ◇ $d(\mathbf{u}, \mathbf{v}) = \sqrt{|u_1 - v_1|^2 + \ldots + |u_n - v_n|^2}$.

注：当$\mathbb{F} = \mathbb{R}$时，可以忽略式子中的“共轭”.

*证明.* 我们仅证明第一个关于内积的式子. 记$S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$. 将线性组合

$$\mathbf{u} = u_1\mathbf{w}_1 + \ldots + u_n\mathbf{w}_n, \quad \mathbf{v} = v_1\mathbf{w}_1 + \ldots + v_n\mathbf{v}_n$$

代入内积，利用共轭双线性性可得

$$\langle \mathbf{u}, \mathbf{v}\rangle = \left\langle \sum\nolimits_{i=1}^n u_i\mathbf{w}_i, \sum\nolimits_{j=1}^n v_j\mathbf{w}_j \right\rangle = \sum_{i=1}^n\sum_{j=1}^n u_i\overline{v_j}\langle \mathbf{w}_i, \mathbf{w}_j\rangle.$$

由于$S$为标准正交基，有$\langle \mathbf{w}_i, \mathbf{w}_j\rangle = \delta_{ij}$. 因此上式右端的非零项一定来自于$i = j$的情形. 故

$$\langle \mathbf{u}, \mathbf{v}\rangle = u_1\overline{v_1} + \cdots u_n\overline{v_n}.$$

□

在这一节中，我们定义了正交/酉矩阵. 事实上，可以在抽象的内积空间上定义正交/酉算子. 大致是这样的（这些是超纲的内容，仅供有兴趣的同学了解）：设$\big(V, \langle\cdot,\cdot\rangle_V\big)$是有限维酉空间，而$T$是$V$上的一个线性算子，则存在唯一的线性算子$T^*$，满足

$$\langle T\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, T^*\mathbf{y}\rangle, \quad (\forall \mathbf{x}, \mathbf{y} \in V).$$

在一组给定的标准正交基下，该伴随算子$T^*$满足$[T^*] = [T]^*$. 这时，一个线性算子$U : V \to V$被称为酉算子，是指它满足$UU^* = U^*U = I$.

## 5.5 正交对角化

在第四章中，我们考虑过结构较简单的算子：可对角化的矩阵. 对于这一类矩阵，我们可以找到一组由特征向量构成的基，因此在对应的坐标轴上，算子的作用效果都是拉伸、收缩、镜面对称等. 现在，我们有了由内积给出的正交性，自然希望把对角化中出现的“坐标轴”更换为“直角坐标轴”.

**例 5.44.** 三维欧氏空间$\mathbb{R}^3$中，向$xOz$-平面作正交投影的算子为

$$T(x,y,z)=(x,0,z),\quad (x,y,z)\in\mathbb{R}^3.$$

它的标准矩阵是

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

向量$\mathbf{i}$与$\mathbf{k}$是关于特征值1的特征向量，而$\mathbf{j}$是关于特征值0的特征向量. 这里，$\{\mathbf{i},\mathbf{j},\mathbf{k}\}$构成$\mathbb{R}^3$的标准正交基.

**定义 5.10.** 设$V$是内积空间，而$T$是$V$上的一个线性算子. 我们称$T$是可正交对角化的，是指存在$T$的一组特征向量构成$V$的标准正交基.

**定义 5.11.** 设$A\in M_n(\mathbb{F})$. 我们称矩阵$A$可正交对角化，是指存在$A$的一组特征向量构成$\mathbb{F}^n$的标准正交基.

在矩阵的对角化中，我们知道$A\in M_n(\mathbb{F})$有$n$个线性无关的特征向量构成$\mathbb{F}$的基，等价于存在可逆矩阵$P$与对角阵$D$，使得$P^{-1}AP=D$. 这里$D$的对角元素为$A$的特征值（带重数），而$P$是对应特征向量按列组成的分块矩阵，也是这从组基到标准基的过渡矩阵.（我们之前只讲了$\mathbb{F}=\mathbb{R}$的情形，对$\mathbb{F}=\mathbb{C}$的情形是完全类似的.）现在，基都变成了正交基，因此过渡矩阵$P$是一个正交矩阵（或酉矩阵），它满足$P^{-1}=P^T$（或$P^{-1}=P^*$).

**定理 5.45.** 设$A\in M_n(\mathbb{R})$，则以下三个命题等价:

*(1) 存在正交矩阵$P$和对角矩阵$D$，使得$P^TAP=D$.*

*(2) 存在$A$的一组特征向量构成$\mathbb{R}^n$的标准正交基.*

*(3) 矩阵$A$是对称的.*

**例 5.46.** 设$A\in M_n(\mathbb{C})$，则以下三个命题等价:

*(1) 存在酉矩阵$P$和对角矩阵$D$，使得$P^*AP=D$.*

*(2) 存在$A$的一组特征向量构成$\mathbb{C}^n$的标准正交基.*

*(3) 矩阵$A$是正规的，即$A^*A=AA^*$.*

在以上两个定理中，(3)与(1)、(2)的等价性是《线性代数II》的内容，在《线性代数I》中，我们承认这样的结论. 假设$A\mathbf{v}_i=\lambda_i\mathbf{v}_i$，其中$\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$是$n$个标准正交的特征向量，则定理中的$P$和$D$分别可取为

$$P=\left[\begin{array}{c:c:c:c} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{array}\right],\qquad D=\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

因此，求一个具体矩阵的正交对角化的关键，在于寻找$\mathbb{F}^n$的一组标准正交基. 在第四章对角化的过程中，我们可以求得所有的特征子空间的各一组基，有定理保证：这些基向量的并集仍然是线性无关的，因此可以构成$\mathbb{F}^n$的基. 在正交对角化过程中，我们可以用Gram-Schmidt正交化过程来得到每个特征子空间的正交基，而不同特征值的特征子空间之间的关系有下列定理给出.

**定理 5.47.** 设$A$是$M_n(\mathbb{R})$中的对称矩阵. 则$A$的关于不同特征值的特征向量是相互正交的.

证明. 设$\lambda, \nu$为$A$的不同特征值，对应的特征向量分别为$\mathbf{x}, \mathbf{y}$. 根据正交矩阵的性质，有

$$\lambda \mathbf{x} \cdot \mathbf{y} = (A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^*\mathbf{y}) = \mathbf{x} \cdot (A\mathbf{y}) = \mu \mathbf{x} \cdot \mathbf{y}.$$

因$\lambda \neq \mu$，我们推出$\mathbf{x} \cdot \mathbf{y} = 0$. □

类似结论对复数情形的厄密特矩阵也是成立的. 下面我们用一个例子来熟悉这些与正交对角化有关的定理.

**例 5.48.** 设$A \in M_3(\mathbb{R})$是对称矩阵，且$rank(A) = 2$. 假设

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

求$A$的所有特征值，及相应的各一个特征向量.

解. 通过矩阵运算式观察，可以得到两个特征值与特征向量：取$\mathbf{v}_1 = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T$, $\mathbf{v}_2 = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}^T$. 则

$$A\mathbf{v}_1 = (-2)\mathbf{v}_1, \qquad A\mathbf{v}_2 = 1\mathbf{v}_2.$$

由于$A$是对称矩阵，可以正交对角化，所以存在正交矩阵$P$和对角矩阵$D$，使得$P^TAP = D$. 注意到$P$可逆，因此$\mathsf{rank}(D) = \mathsf{rank}(P^TAP) = \mathsf{rank}(A) = 2$. 所以除了$-2$与1以外，$D$中的第三个对角元素为0，故第三个特征值为0. 设$\mathbf{v}_3$是$A$关于特征值0的一个特征向量. 则$\mathbf{v}_3$与$\mathbf{v}_1$，$\mathbf{v}_2$均正交. 列出方程组

$$\begin{bmatrix} 1 & 0 & -1 \\ \hdashline 1 & 2 & 1 \end{bmatrix} \mathbf{v}_3 = \mathbf{0}.$$

该方程组的解空间是1维的，我们取一个特解$\mathbf{v}_3 = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}^T$. 综上，矩阵$A$有三个特征值$-2, 1$与0，对应的特征向量分别为$\mathbf{v}_1, \mathbf{v}_2$和$\mathbf{v}_3$.

□

现在我们来进行具体的正交对角化. 假设可正交对角化的矩阵$A$的所有不同特征值为$\lambda_1, \ldots, \lambda_k$，而$\{\mathbf{u}_{i,1}, \ldots, \mathbf{u}_{i,i_{n_i}}\}$是特征子空间$E_{\lambda_i}(A)$ $(1 \leq i \leq k)$的一组基. 首先，我们在每个特征子空间上运用Gram-Schmidt正交化过程，得到标准正交基$\{\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,i_{n_i}}\}$. 这时，它们的并集

$$\{\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,n_1}, \mathbf{v}_{2,1}, \ldots, \mathbf{v}_{2,n_2}, \ldots, \mathbf{v}_{k,1}, \ldots, \mathbf{v}_{k,n_k}\}$$

构成$\mathbb{F}^n$的一组标准正交基. 将它们按列矩阵写成分块矩阵$P$，记

$$D = \mathsf{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_n, \ldots, \lambda_n),$$

其中$\lambda_i$ $(1 \leq i \leq k)$出现的次数为$n_i = \dim E_{\lambda_i}(A)$. 则正交对角化由$P^TAP = D$给出.

**例 5.49.** 试将对称矩阵$A$正交对角化，其中

$$A = \begin{bmatrix} 2 & 3 & 3 \\ 3 & 2 & 3 \\ 3 & 3 & 2 \end{bmatrix} \in M_3(\mathbb{R}).$$

解. 因$A$为对称矩阵，是可正交对角化的. 令

$$0 = \mathsf{det}(\lambda I - A) = \begin{vmatrix} \lambda-2 & -3 & -3 \\ -3 & \lambda-2 & -3 \\ -3 & -3 & \lambda-2 \end{vmatrix} = \lambda^3 - 6\lambda^2 - 15\lambda - 8 = (\lambda+1)^2(\lambda-8).$$

有两个不同特征值$-1$和8.

对于特征值$-1$，令

$$\mathbf{0} = (-I - A)\mathbf{u} = \begin{bmatrix} -3 & -3 & -3 \\ -3 & -3 & -3 \\ -3 & -3 & -3 \end{bmatrix} \mathbf{u}.$$

取解空间的一组基

$$\mathbf{u}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

对于特征值8,令

$$\mathbf{0} = (8I - A)\mathbf{u} = \begin{bmatrix} 6 & -3 & -3 \\ -3 & 6 & -3 \\ -3 & -3 & 6 \end{bmatrix} \mathbf{u}.$$

取解空间的一组基

$$\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

将特征子空间$E_{-1}(A)$的基$\{\mathbf{u}_1, \mathbf{u}_2\}$作Gram-Schmidt正交化，得

$$\mathbf{w}_1 = \mathbf{u}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{w}_2 = \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{w}_1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix}.$$

$$\mathbf{v}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}.$$

将特征子空间$E_8(A)$的基$\{\mathbf{u}_3\}$作Gram-Schmidt正交化，得

$$\mathbf{v}_3 = \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

这时，向量组$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$构成$\mathbb{R}^3$的一组标准正交基，取

$$P = \left[\begin{array}{c:c:c} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{array}\right], \quad D = \begin{bmatrix} -1 & & \\ & -1 & \\ & & 8 \end{bmatrix}.$$

则有$P^TAP = D$. □

**例 5.50.** 试将厄密特矩阵$A$正交对角化，其中

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in M_2(\mathbb{C}).$$

解.（略.） $\square$

现在，我们将对称矩阵的正交对角化用另一种方式写出来. 设$A \in M_n(\mathbb{R})$，而$A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ $(1 \le i \le n)$，其中$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$构成$\mathbb{R}^n$的一组标准正交基. 记

$$P = [\, \mathbf{v}_1 \mid \mathbf{v}_2 \mid \ldots \mid \mathbf{v}_n \,], \qquad D = \mathsf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

则$P^T AP = D$. 进一步，我们有

$$A = PDP^T = [\, \mathbf{v}_1 \mid \mathbf{v}_2 \mid \ldots \mid \mathbf{v}_n \,] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \hdashline \mathbf{v}_2 \\ \hdashline \vdots \\ \hdashline \mathbf{v}_n \end{bmatrix}$$

$$= [\, \mathbf{v}_1 \mid \mathbf{v}_2 \mid \ldots \mid \mathbf{v}_n \,] \begin{bmatrix} \lambda_1 \mathbf{v}_1 \\ \hdashline \lambda_2 \mathbf{v}_2 \\ \hdashline \vdots \\ \hdashline \lambda_n \mathbf{v}_n \end{bmatrix} = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \cdots \lambda_n \mathbf{v}_n \mathbf{v}_n^T.$$

我们把这个结论整理成下述定理.

**定理 5.51** (谱分解定理). 设$A \in M_n(\mathbb{R})$是对称矩阵，则存在$\lambda_1, \ldots, \lambda_n \in \mathbb{R}$，以及标准正交的向量组$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$，使得

$$A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \cdots \lambda_n \mathbf{v}_n \mathbf{v}_n^T.$$

定理中的等式称为矩阵$A$的谱分解（特征值是谱的一种情形），我们来解释一下它的含义：每个$E_r := \mathbf{v}_r \mathbf{v}_r^T$ $(1 \le r \le n)$都是秩为1的$n$阶方阵，且$E_r$作为线性算子的映射效果是在$\mathbb{R}^n$中向子空间$V_r := \mathsf{span}\{\mathbf{v}_r\}$作正交投影（思考题）. 并且，这些子空间$V_1, V_2, \ldots, V_n$是两两正交的. 用代数的语言，有

$$E_r^2 = E_r = E_r^T \ (1 \le r \le n), \qquad E_i E_j = 0 \ (1 \le i \ne j \le n),$$

以及

$$I = \sum_{r=1}^n E_r, \qquad A = \sum_{r=1}^n \lambda_r E_r.$$

这时，若$p(x)$是一个多项式，进一步还有

$$p(A) = \sum_{r=1}^n p(\lambda_r) E_r,$$

即$A$的代数运算完全可以在特征值上分别进行.

另外，上述谱分解中的特征值是带重数的，我们也可以把相同的特征值合并，这时对应的特征向量张成对应的特征子空间. 设$\mu_1, \ldots, \mu_k$是矩阵$A \in M_n(\mathbb{R})$的所有不同特征值，则

$$A = \sum_{i=1}^k \mu_i \widetilde{E_i}.$$

其中$\widetilde{E_i}$是$\mathbb{R}^n$中向特征子空间$E_{\mu_i}(A)$作正交投影的算子. 并且，我们有

$$p(A) = \sum_{i=1}^k p(\mu_i) \widetilde{E_i},$$

对应复数的情形，正规矩阵也可以作谱分解.

**定理 5.52** (谱分解定理). 设$A \in M_n(\mathbb{C})$是正规矩阵，则存在$\lambda_1, \ldots, \lambda_n \in \mathbb{C}$，以及标准正交的向量组$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{C}^n$，使得

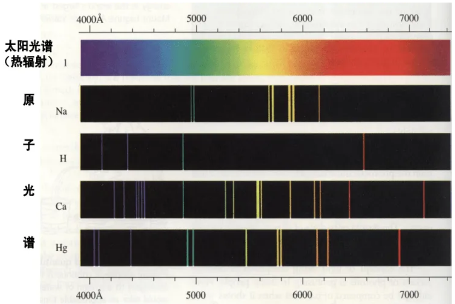
$$A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^* + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^* + \cdots \lambda_n \mathbf{v}_n \mathbf{v}_n^*.$$

（以下灰色的部分供有兴趣的同学们了解，并不在考试范围之内）

在量子力学中，观测量都用酉空间上的厄密特算子（即自伴算子）表示，所有可能的观测值都是该算子的特征值. 仪表上读出的数值都是实数，正对应于厄密特算子的特征值都是实的. 把所有不同的实特征值从小到大排列：$\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$，算子的谱分解

$$A = \sum\nolimits_{i=1}^{k} \mu_i \widetilde{E_i},$$

可对应与物理中的光谱.




当酉空间是无线维时（这时要求是希尔伯特空间），“谱”可以包含“连续”的部分，这时谱分解公式里的求和号“$\sum$”可以变成积分号“$\int$”. 在物理中，位置与动量（速度）是无法同时同时准确测量的，这被称为不确定性原理/测不准原理. 在量子物理中，能被同时测量的观测量是可交换的算子.

**定理 5.53.** *设$A, B \in M_n(\mathbb{C})$交换（即$AB = BA$）. 则它们有公共的特征向量.*

需要注意的是，上述定理中的矩阵$A$、$B$虽然有公共的特征向量，但对应的特征值未必相同. 在量子物理中，特征值所对应的单位特征向量是物理系统的“定态”，当系统处于该定态时，观测到的值便是这个特征值. 而空间中的任何一个单位向量都表示“叠加态”，该单位向量写成基的线性组合的系数的平方则表示系统处于对应定态的概率. 在观测之前，我们无法确定系统的状态，因此用“叠加态”表示. 可进行同时观测的两个观测量有公共的“定态”，在这个定态上观测到的值是两个算子分别的特征值.

可对角化的矩阵作为线性变换的作用效果是非常直观的. 对于不可对角化的矩阵，我们也有相应的结构性定理. 在此，我们展示一些结论，它们的细节不属于《线性代数I》的内容.

**定理 5.54** (Schur 定理)**.** *若$A \in M_n(\mathbb{R})$的特征多项式有$n$个根$\lambda_1, \ldots, \lambda_n \in \mathbb{R}$，则存在正交矩阵$P$，使得$P^T AP$有如下的上三角形式：*

$$P^T AP = \begin{bmatrix} \lambda_1 & * & * & \cdots & * \\ 0 & \lambda_2 & * & \cdots & * \\ 0 & 0 & \lambda_3 & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} := S$$

*特别地，公式$A = PSP^T$被称为$A$的$Schur$分解.*

回顾一下代数学基本定理，实系数多项式在实数域内可分解为一次因子与二次因子的乘积，其中的二次因子判别式均小于零. 在以上定理的条件中，“特征多项式有$n$个根”意味着该多项式在实数域内可分解为一次因子的乘积. 对于特征多项式存在二次因子的情形，我们有下述结论.

**定理 5.55** (Hessenberg 定理)**.** *设$A \in M_n(\mathbb{R})$，则存在正交矩阵$P$，使得*

$$P^TAP = \begin{bmatrix} * & * & \dots & * & * & * \\ * & * & \dots & * & * & * \\ 0 & * & \ddots & * & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & * & * & * \\ 0 & 0 & \dots & 0 & * & * \end{bmatrix} := H.$$

*特别地，公式$A = PHP^T$被称为$A$的Hessenberg分解.*

这几种分解在数值计算上的意义有以下几点：矩阵$S$与$H$的形式均比一般的矩阵$A$本身简单，这可以加快计算速度并减小储存空间；其次，“正交矩阵”可以保证在计算过程中因“四舍五入”产生的误差（如$1/3*3 = 0.333333*3 = 0.999999.$）不会在后续的计算中被放大（设$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e}$，其中$\mathbf{e}$表示误差，这时由与正交变换保持向量长度不变，故$\|P\mathbf{x} - P\hat{\mathbf{x}}\| = \|P\mathbf{e}\| = \|\mathbf{e}\|$，误差不会变大.）

## 5.6 奇异值分解

在应用中，我们拿到的一组数据组成矩阵$A \in M_{m\times n}(\mathbb{R})$，往往$A$的行数不等于列数，构不成方阵. 如何将其转换为一个方阵，且保持基本的信息不被改变呢？一个自然的想法是：$A^TA \in M_n(\mathbb{R})$，$AA^T \in M_m(\mathbb{R})$!

**定理 5.56.** *设$A \in M_{m\times n}(\mathbb{R})$. 则*

*(1) $\mathsf{Null}(A) = \mathsf{Null}(A^TA)$;*

*(2) $\mathsf{Row}(A) = \mathsf{Row}(A^TA)$;*

*(3) $\mathsf{Col}(A^T) = \mathsf{Col}(A^TA)$;*

*(4) $\mathsf{rank}(A) = \mathsf{rank}(A^TA)$.*

*证明.* (1) 对任意$\mathbf{x} \in \mathsf{Null}(A)$，有$A\mathbf{x} = \mathbf{0}$. 从而

$$(A^TA)\mathbf{x} = A^T(A\mathbf{x}) = A^T\mathbf{0} = \mathbf{0}.$$

因此$\mathbf{x} \in A^TA$. 故$\mathsf{Null}(A) \subseteq \mathsf{Null}(A^TA)$.

反之，设$\mathbf{y} \in \mathsf{Null}(A^TA)$，则$(A^TA)\mathbf{y} = \mathbf{0}$. 这时有

$$\mathbf{0} = \mathbf{y}^T(A^TA)\mathbf{y} = (A\mathbf{y})^T(A\mathbf{y}) = \|A\mathbf{y}\|^2.$$

从而$A\mathbf{y} = \mathbf{0}$，即$\mathbf{y} \in \mathsf{Null}(A)$. 故$\mathsf{Null}(A^TA) \subseteq \mathsf{Null}(A)$.

(2) 我们有$\mathsf{Row}(A) = \mathsf{Null}(A)^T = \mathsf{Null}(A^TA)^T = \mathsf{Row}(A^TA)$.

(3) 我们有$\mathsf{Col}(A^TA) = \mathsf{Row}\big((A^TA)^T\big) = \mathsf{Row}(A^TA) = \mathsf{Row}(A) = \mathsf{Col}(A^T)$.

(4) 我们有$\mathsf{rank}(A) = \dim \mathsf{Row}(A) = \dim \mathsf{Row}(A^TA) = \mathsf{rank}(A^TA)$. □

**定理 5.57.** *设$A \in M_{m\times n}(\mathbb{R})$. 则有矩阵$A^TA$可正交对角化，且特征值均为非负实数.*

*证明.* 矩阵$A^TA$是对称的，所以可正交对角化. 设$\lambda$为$A^TA$的一个特征值，而$\mathbf{x}$为对应的一个特征向量. 则有

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T(A^TA\mathbf{x}) = \mathbf{x}^T(\lambda\mathbf{x}) = \lambda\|\mathbf{x}\|^2.$$

因$\mathbf{x} \neq \mathbf{0}$，知$\|\mathbf{x}\| \neq 0$，我们有

$$\lambda = \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2} \geq 0.$$

□

假设$A = \mathsf{diag}(\mu_1, \mu_2, \ldots, \mu_n)$为对角矩阵，则$A^TA = \mathsf{diag}(\mu_1^2, \mu_2^2, \ldots, \mu_n^2)$. 后者的特征值是前者特征值的平方. 当$A$不是方阵时，方阵$A^TA$的特征值仍然包含了矩阵$A$的信息.

**定义 5.12.** 设$A \in M_{m\times n}(\mathbb{R})$，而$\lambda_1, \lambda_2, \ldots, \lambda_n$为$A^TA$的特征值. 我们称

$$\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_n = \sqrt{\lambda_n}$$

为矩阵$A$的奇异值.

在应用中，我们往往将特征值/奇异值按从大到小的顺序排列，即$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$，从而$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$.

**例 5.58.** 设$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$. 求矩阵$A$和$A^T$的特征值.

解. 计算可得

$$A^TA = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad AA^T = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

进一步，有

$$\mathsf{det}(\lambda I - A^TA) = \begin{vmatrix} \lambda - 2 & -1 \\ -1 & \lambda - 2 \end{vmatrix} = (\lambda - 3)(\lambda - 1).$$

因此$A^TA$的特征值为$\lambda_1 = 3$, $\lambda_2 = 1$，而$A$的奇异值为$\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$. 另外，

$$\mathsf{det}(\lambda I - AA^T) = \begin{vmatrix} \lambda - 2 & -1 & -1 \\ -1 & \lambda - 1 & 0 \\ -1 & 0 & \lambda - 1 \end{vmatrix} = (\lambda - 3)(\lambda - 1)\lambda.$$

因此$(A^T)^TA^T = AA^T$的特征值为$\lambda_1 = 3$, $\lambda_2 = 1$, $\lambda_3 = 0$，而$A^T$的奇异值为$\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$, $\sigma_3 = 0$. □

思考题：试证明$A$与$A^T$的奇异值除了一些可能的0之外，均是相同的.

现在，我们先快速地展示一下“奇异值分解”中的矩阵结构，读者可以把它与矩阵的正交对角化作一些对比，而具体细节在之后再慢慢展开. 设$A \in M_{m\times n}(\mathbb{R})$的秩为$r$. 则存在正交矩阵$U \in M_m(\mathbb{R})$与$V \in M_n(\mathbb{R})$，使得

$$A = U\Sigma V,$$

其中

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}_{m\times n}.$$

在这一节中，我们总使用下述记号：矩阵$A \in M_{m\times n}(\mathbb{R})$满足$\mathsf{rank}(A) = r$. 方阵$A^TA$的特征值为

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0,$$

相应的特征向量分别为$\mathbf{v}_1, \ldots, \mathbf{v}_n$，它们构成$\mathbb{R}^n$的一组标准正交基. 这时，矩阵$A$的奇异值为$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0$.

**定理 5.59** (奇异值分解之几何版). *设$A \in M_{m\times n}(\mathbb{R})$满足rank$(A) = r$. 则存在*

*(i) 空间$\mathbb{R}^n$的一组标准正交基$\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$;*

*(ii) 空间$\mathbb{R}^m$的一组标准正交基$\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$;*

*(iii) 非负实数$\sigma_1, \dots, \sigma_r \in \mathbb{R}$, 使得*

$$A\mathbf{v}_i = \begin{cases} \sigma_i \mathbf{u}_i, & (1 \le i \le r), \\ 0, & (r < i \le n), \end{cases} \quad A^T\mathbf{u}_i = \begin{cases} \sigma_i \mathbf{v}_i, & (1 \le i \le r), \\ 0, & (r < i \le m). \end{cases}$$

证明. 在(i)和(iii)中出现的标准正交基和非负实数分别为$A^TA$的特征向量和$A$的奇异值. 我们来构造(ii)中的标准正交基. 由于$\mathsf{rank}(A^TA) = \mathsf{rank}(A) = r$，故前$r$个特征值和奇异值是正的，而后$n-r$个为0. 因此

$$\mathsf{Col}(A^TA) = \mathsf{span}\{(A^TA)\mathbf{v}_1, \dots, (A^TA)\mathbf{v}_n\} = \mathsf{span}\{\lambda_1\mathbf{v}_1, \dots, \lambda_r\mathbf{v}_r, \mathbf{0}, \dots, \mathbf{0}\} = \mathsf{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}.$$

对$1 \le i, j \le r$, 我们有

$$\mathbf{u}_i \cdot \mathbf{u}_j = \left(\frac{1}{\sigma_i} A\mathbf{v}_i\right) \cdot \left(\frac{1}{\sigma_j} A\mathbf{v}_j\right) = \frac{1}{\sigma_i\sigma_j} \mathbf{v}_i \cdot (A^TA\mathbf{v}_j) = \frac{\lambda_j}{\sigma_i\sigma_j} \mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij}.$$

因此$\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$构成标准正交集. 我们将其扩充为$\mathbb{R}^m$中的一组标准正交基$\{\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$.

根据$\mathbf{u}_i$ $(1 \le i \le r)$的取法，我们有$A\mathbf{v}_i = \sigma_i\mathbf{u}_i$，以及

$$A^T\mathbf{u}_i = A^T\left(\frac{1}{\sigma_i} A\mathbf{v}_i\right) = \frac{1}{\sigma_i}(A^TA)\mathbf{v}_i = \frac{\lambda_i}{\sigma_i}\mathbf{u}_i = \sigma_i\mathbf{u}_i.$$

对于$r+1 \le j \le n$，我们有

$$\mathbf{v}_j \in \mathsf{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}^T = \mathsf{Col}(A^TA)^\perp = \mathsf{Col}(A^T)^\perp = \mathsf{Null}(A),$$

故$A\mathbf{v}_j = \mathbf{0}$. 另外，注意到

$$\mathsf{span}\{\mathbf{u}_1, \dots, \mathbf{u}_n\} = \mathsf{span}\{A\mathbf{v}_1, \dots, A\mathbf{v}_n\} = \mathsf{Col}(A).$$

对任意$\mathbf{u}_i$ $(1 \le i \le r)$，有

$$\mathbf{u}_j \in \mathsf{span}\{\mathbf{u}_1, \dots, \mathbf{u}_n\}^\perp = \mathsf{Col}(A)^\perp = \mathsf{Null}(A^T).$$

故$A^T\mathbf{u}_j = \mathbf{0}$. 定理得证. □

奇异值分解的几何版本的含义是这样的：把一个$m \times n$阶实矩阵$A$看成线性变换，存在定义域$\mathbb{R}^n$的一组标准正交基$\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$，变换把它们映为值域$\mathbb{R}^m$中的一组正交集

$$\{A\mathbf{v}_1, \dots, A\mathbf{v}_n\} = \{\sigma_1\mathbf{u}_1, \dots, \sigma_r\mathbf{u}_r, \mathbf{0}, \dots, \mathbf{0}\}.$$

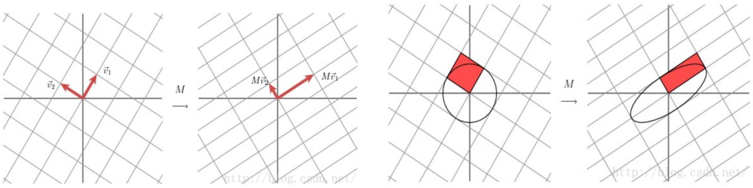
相应地，矩阵$A^T$作为线性变换，将$\mathbb{R}^m$中的标准正交基$\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$映为$\mathbb{R}^n$中的正交集

$$\{A^T\mathbf{u}_1, \dots, A^T\mathbf{u}_m\} = \{\sigma_1\mathbf{v}_1, \dots, \sigma_r\mathbf{v}_r, \mathbf{0}, \dots, \mathbf{0}\}.$$

以$2 \times 2$的可逆矩阵为例，根据奇异值分解，存在一组标准正交基$\{\mathbf{v}_1, \mathbf{v}_2\}$，被映为正交集$\{\mathbf{u}_1, \mathbf{u}_2\}$. 这时以$\mathbf{v}_1, \mathbf{v}_2$为边的单位正方形被映为以$\mathbf{u}_1, \mathbf{u}_2$为边的长方形. 而以原点为圆心的单位圆被映为椭圆，其中$\mathbf{u}_1, \mathbf{u}_2$给出了长短半轴.

**例 5.60.** *平面$\mathbb{R}^2$上的错切变换是可逆且不可对角化的:*

$$M = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \quad M^T = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}, \qquad M^TM = \begin{bmatrix} 1 & k \\ k & 1+k^2 \end{bmatrix}.$$

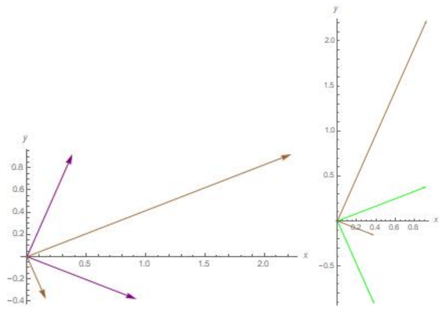


取$k=2$，使用定理中的记号，计算可得$\lambda_{1,2}=3\pm2\sqrt{2}$, $\sigma_{1,2}=\sqrt{2}\pm1$，以及

$$\mathbf{v}_{1,2}=\left(\frac{1}{\sqrt{4\pm2\sqrt{2}}},\frac{1\pm\sqrt{2}}{\sqrt{4+2\sqrt{2}}}\right),$$

$$\mathbf{u}_{1,2}=\left(\frac{\pm1+\sqrt{2}}{\sqrt{4\pm2\sqrt{2}}},\frac{\pm1}{\sqrt{4\pm2\sqrt{2}}}\right).$$

从下图可以看到奇异值分解的几何效果.



(c) $\mathbf{v}_i$, $A\mathbf{v}_i$ (d) $\mathbf{u}_i$, $A^T\mathbf{u}_i$

在奇异值分解的几何版本中，我们得到

$$A\mathbf{v}_i=\begin{cases}\sigma_i\mathbf{u}_i, & (1\le i\le r),\\ 0, & (r<i\le n).\end{cases}$$

写成矩阵形式，得

$$A\left[\begin{array}{c:c:c:c:c:c}\mathbf{v}_1 & \dots & \mathbf{v}_r & \mathbf{v}_{r+1} & \dots & \mathbf{v}_n\end{array}\right]=\left[\begin{array}{c:c:c:c:c:c}\sigma_1\mathbf{u}_1 & \dots & \sigma_r\mathbf{u}_r & \mathbf{0} & \dots & \mathbf{0}\end{array}\right]$$

$$=\left[\begin{array}{c:c:c:c:c:c}\mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m\end{array}\right]\left[\begin{array}{c:c:c:c:c:c}\sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & & & \end{array}\right].$$

我们使用记号

$$U=\left[\begin{array}{c:c} U_r & U_{m-r}\end{array}\right]_{m\times m}=\left[\begin{array}{ccc:ccc} \mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m \end{array}\right],$$

$$V=\left[\begin{array}{c:c} V_r & V_{n-r}\end{array}\right]_{n\times n}=\left[\begin{array}{ccc:ccc} \mathbf{v}_1 & \dots & \mathbf{v}_r & \mathbf{v}_{r+1} & \dots & \mathbf{v}_n \end{array}\right],$$

$$\Sigma=\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}=\left[\begin{array}{ccc:c} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ \hdashline & & & \\ \end{array}\right].$$

则有$AV=U\Sigma$. 注意到$U$和$V$均为正交阵，因此

$$A=U\Sigma V^{-1}=U\Sigma V^T.$$

另外，将等式右侧的表达式按照分块矩阵展开，有

$$U\Sigma V^T=\begin{bmatrix} U_r & U_{m-r}\end{bmatrix}\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix}=\begin{bmatrix} U_r & U_{m-r}\end{bmatrix}\begin{bmatrix} \Sigma_r V_r^T \\ 0 \end{bmatrix}=U_r\Sigma_r V_r^T.$$

所以还可以得到$A=U_r\Sigma_r V_r$.

**定理 5.61** (奇异值分解之完整版)**.** 设$A\in M_{m\times n}(\mathbb{R})$满足*rank*$(A)=r$. 则存在

*(i)* 正交矩阵$V\in M_n(\mathbb{R})$;

*(ii)* 正交矩阵$V\in M_m(\mathbb{R})$;

*(iii)* 非负实数$\sigma_1,\dots,\sigma_r\in\mathbb{R}$,

使得$A=U\Sigma V^T$，其中

$$\Sigma=\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}=\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}.$$

**定理 5.62** (奇异值分解之简化版)**.** 设$A\in M_{m\times n}(\mathbb{R})$满足*rank*$(A)=r$. 则存在

*(i)* 空间$\mathbb{R}^n$的一组标准正交集$\{\mathbf{v}_1,\dots,\mathbf{v}_r\}$;

*(ii)* 空间$\mathbb{R}^m$的一组标准正交集$\{\mathbf{u}_1,\dots,\mathbf{u}_r\}$;

*(iii)* 非负实数$\sigma_1,\dots,\sigma_r\in\mathbb{R}$, 使得

$$A=\sigma_1\mathbf{u}_1\mathbf{v}_1^T+\cdots+\sigma_r\mathbf{u}_r\mathbf{v}_r^T.$$

上式的矩阵形式为

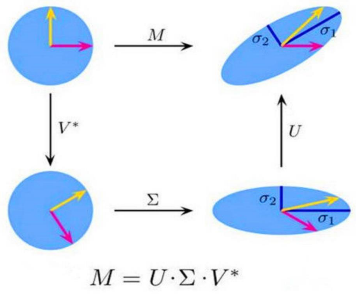
$$A=U_r\Sigma_r V_r^T=\left[\begin{array}{c:c:c:c} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_r \end{array}\right]\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}\begin{bmatrix} \mathbf{v}_1^T \\ \hline \mathbf{v}_2^T \\ \hline \vdots \\ \hline \mathbf{v}_r^T \end{bmatrix}.$$

下面我们来看两个具体的例子.

**例 5.63.** 平面$\mathbb{R}^2$上的错切变换$(k=2)$的奇异值分解为

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}} & \frac{\sqrt{2}-1}{\sqrt{4-2\sqrt{2}}} \\ \frac{1}{\sqrt{4+2\sqrt{2}}} & \frac{-1}{\sqrt{4-2\sqrt{2}}} \end{bmatrix} \begin{bmatrix} \sqrt{2}+1 & 0 \\ 0 & \sqrt{2}-1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{4+2\sqrt{2}}} & \frac{1+\sqrt{2}}{\sqrt{4+2\sqrt{2}}} \\ \frac{1}{\sqrt{4-2\sqrt{2}}} & \frac{1-\sqrt{2}}{\sqrt{4-2\sqrt{2}}} \end{bmatrix}.$$

如图，作为线性变换，其效果相当于依次作用变换$V^T$、$\Sigma$和$U$. 最先作用的正交矩阵$V^T$的效果是旋转




（或复合镜像反射），之后$\Sigma$的作用是在坐标轴上的拉伸、收缩（或复合镜像反射）等，最后作用的正交矩阵$U$的效果是再一次的旋转（或复合镜像反射）.

**例 5.64.** 求矩阵$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$的奇异值分解（完整版）.

解. 根据之前的计算，我们有

$$A^TA = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

其特征值为$\lambda_1 = 3$, $\lambda_2 = 1$. 从而$A$的奇异值为$\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$，且$\mathrm{rank}(A) = 2$.

现在我们来找一组由$A^TA$的特征向量构成的标准正交基. 解方程组

$$\mathbf{0} = (3I - A)\mathbf{x} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\mathbf{x}, \qquad \mathbf{0} = (I - A)\mathbf{x} = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}\mathbf{x},$$

分别特征子空间$E_3(A^TA)$的一组基$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$，与$E_1(A^TA)$的一组基$\mathbf{x}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

在每个特征子空间内作Gram-Schmidt正交化，分别得

$$\mathbf{v}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} = \begin{bmatrix} 1/\sqrt{2} \\ 1\sqrt{2} \end{bmatrix}, \quad \mathbf{v}_2 = \frac{\mathbf{x}_2}{\|\mathbf{x}_2\|} = \begin{bmatrix} -1/\sqrt{2} \\ 1\sqrt{2} \end{bmatrix}.$$

这时，向量组$\{\mathbf{v}_1, \mathbf{v}_2\}$构成$\mathbb{R}^2$的一组标准正交基.

接下来，注意到$\mathrm{rank}(A) = 2$，我们令

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ 1\sqrt{2} \end{bmatrix} = \begin{bmatrix} 2/\sqrt{6} \\ 1\sqrt{6} \\ 1\sqrt{6} \end{bmatrix},$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{1}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -1/\sqrt{2} \\ 1\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1\sqrt{2} \\ -1\sqrt{2} \end{bmatrix}.$$

这时$\{\mathbf{u}_1, \mathbf{u}_2\}$构成$\mathbb{R}^3$中的一组标准正交集，我们将它扩充为标准正交基$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. 注意到$\mathbf{u}_3$与$\mathbf{u}_1, \mathbf{u}_2$均正交，我们构造方程组

$$\mathbf{0} = \begin{bmatrix} \mathbf{u}_1 \\ \hdashline \mathbf{u}_2 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 2/\sqrt{6} & 1\sqrt{6} & 1\sqrt{6} \\ \hdashline 0 & 1\sqrt{2} & -1\sqrt{2} \end{bmatrix} \mathbf{y},$$

取解空间的一组基$\mathbf{y}_3 = \begin{bmatrix} -1 & 1 & 1 \end{bmatrix}^T$. 将其进行Gram-Schmidt正交化，得

$$\mathbf{u}_3 = \frac{\mathbf{y}_3}{\|\mathbf{y}_3\|} = \begin{bmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

这时$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$便是$\mathbb{R}^3$的一组标准正交基.

现在，我们令

$$U = \left[\begin{array}{c:c:c} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1\sqrt{6} & 1\sqrt{2} & 1/\sqrt{3} \\ 1\sqrt{6} & -1\sqrt{2} & 1/\sqrt{3} \end{array}\right], \quad V = \left[\begin{array}{c:c} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{array}\right], \quad \Sigma = \left[\begin{array}{cc} \sqrt{3} & 0 \\ 0 & 1 \\ \hdashline 0 & 0 \end{array}\right].$$

则有$A = U\Sigma V^T$，即

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \left[\begin{array}{c:c:c} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1\sqrt{6} & 1\sqrt{2} & 1/\sqrt{3} \\ 1\sqrt{6} & -1\sqrt{2} & 1/\sqrt{3} \end{array}\right] \left[\begin{array}{cc} \sqrt{3} & 0 \\ 0 & 1 \\ \hdashline 0 & 0 \end{array}\right] \left[\begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \\ \hdashline -1/\sqrt{2} & 1/\sqrt{2} \end{array}\right].$$

□

复矩阵$A \in M_{m\times n}(\mathbb{C})$也有对应的奇异值分解$A = U\Sigma V^*$，其中$U, V$均为酉矩阵. 我们不再展示其中的细节.

（以下灰色部分供有兴趣的同学了解，并不在大纲要求的范围之内.）

矩阵$A \in M_{m\times n}(\mathbb{R})$的许多信息包含在奇异值分解$A = U\Sigma V^T$内. 事实上，我们有

$$\mathsf{Col}(A) = \mathsf{Col}(U_r), \quad \mathsf{Null}(A) = \mathsf{Col}(V_{n-r}), \quad \mathsf{Row}(A) = \mathsf{Col}(A^T) = \mathsf{Col}(V_r).$$

在奇异值分解中，我们有类似谱分解的等式

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T,$$

这里$\sigma_1$ 设$\|\cdot\|$是$M_{m\times n}(\mathbb{R})$上定义的某范数. 设$k \le r$. 若我们想用秩为$k$的矩阵去近似秩为$r$的矩阵$A$，则最佳逼近为

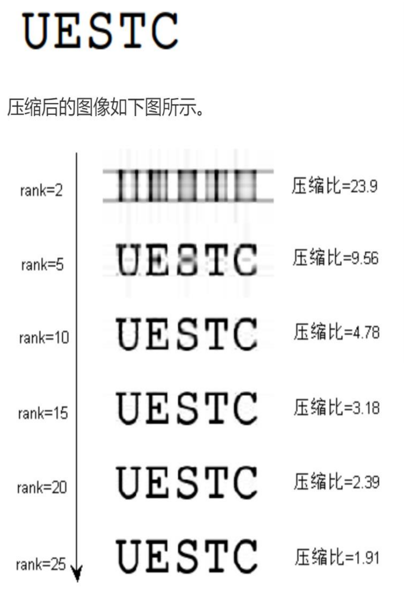
$$\|A - A_k\| = \min_{\mathsf{rank}(B)\le k} \|A - B\|,$$

其中

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T.$$

也就是说，矩阵$A$中最主要的信息都集中在奇异值分解式的左上角部分.

根据应用的需要，我们可以舍弃小奇异值对应的部分，而保留大奇异值对应的部分，所得的矩阵与原矩阵的差异是较小的. 比如，将图像化成矩阵数据，所得的矩阵维数是很大的，而图像上的不同点之间往往有很强的相关性，比如说眼睛位置的许多点都是黑色的，因此矩阵的秩往往较小. 这时，使用奇异值分解可以节省大量的数据储存空间. 比如，左图中的图片是不同灰度的10个横条，对应的$10 \times 10$矩阵秩为1，矩阵里包含了100个数. 而它的奇异值分解只需要$\mathbf{u}_1$、$\mathbf{v}_1$这两个10维向量，以及奇异值$\sigma_1$，共11个数. 另外，将右图中的字母图片写成矩阵并作奇异值分解，当我们取前5个奇异值时，在近似的图片中就已经能够辨认出所有字母了，而当我们取前10个奇异值时，近似的图片已经具有足够的清晰度了. 奇异值分解是压缩图片的一种有效方法，也降噪的常用手段.




$$uv^H = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0.1 & 0.1 & \cdots & 0.1 \\ 0.2 & 0.2 & \cdots & 0.2 \\ \vdots & \vdots & & \vdots \\ 0.9 & 0.9 & \cdots & 0.9 \end{pmatrix}$$

## 5.7 最佳逼近与最小二乘法

在这一节中，我们仅考虑实数的情形，即$\mathbb{F} = \mathbb{R}$.

如果我们得到了一组二维数据

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_m, y_m).$$

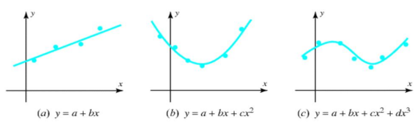
并希望用一个$n-1$次的多项式$p(x)$去拟合这些数据. 可以设

$$p(x) = c_0 + c_1 x + \ldots + c_{n-1}x^{n-1},$$

其中$c_0, c_1, \ldots, c_{n-1}$是待定的系数. 我们希望求这组系数，使得平方误差

$$\sum\nolimits_{k=1}^{m} \left| y_k - (c_0 + c_1 x_k + \ldots + c_{n-1}x_k^{n-1}) \right|^2$$

达到最小.




将这个问题用矩阵的语言写出：令

$$A = \begin{bmatrix} 1 & x_1 & \ldots & x_1^{n-1} \\ 1 & x_2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & \ldots & x_m^{n-1} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

我们要求解$\mathbf{x}$，使得$\|\mathbf{y} - A\mathbf{x}\|$达到最小. 我们将满足上述性质的解称为方程$A\mathbf{x} = \mathbf{b}$的“最小二乘解”.

现在我们假设矩阵$A \in M_{m\times n}(\mathbb{R})$, 向量$\mathbf{x} \in \mathbb{R}^n$和$\mathbf{b} \in \mathbb{R}^m$均为任意取的. 线性方程组$A\mathbf{x} = \mathbf{b}$未必有解. 我们的目标是找一个最接近该方程组的解，即寻找向量$\hat{\mathbf{x}} \in \mathbb{R}^n$，使得

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| = \min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|.$$

注意到

$$\{A\mathbf{x} :\ \mathbf{x} \in \mathbb{R}^n\} = \mathsf{Ran}(T_A) = \mathsf{Col}(A),$$

因此有

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\| = \min_{\mathbf{w}\in\mathsf{Col}(A)} \|\mathbf{w} - \mathbf{b}\|.$$

所以最小二乘解$\hat{\mathbf{x}}$必然是离子空间$\mathsf{Col}(A)$最近的那个向量. 根据几何直观，最近的向量应当是$\mathbf{b}$向子空间$\mathsf{Col}(A)$作正交投影所得的向量.

**定理 5.65.** *设$W$是有限维内积空间$(V, \langle\cdot,\cdot\rangle)$的子空间，而$\mathbf{b} \in V$. 则$proj_W(\mathbf{b})$是$\mathbf{b}$在$W$中的最佳逼近，即对其他向量$\mathbf{w} \in W \setminus \{proj_W(\mathbf{b})\}$，均有*

$$\big\|\mathbf{b} - proj_W(\mathbf{b})\big\| < \|\mathbf{b} - \mathbf{w}\|.$$

*证明.* 因为

$$\mathbf{b} = \mathrm{proj}_W(\mathbf{b}) + \mathrm{proj}_{W^\perp}(\mathbf{b}),$$

我们有

$$\mathbf{b} - \mathbf{w} = \big(\mathrm{proj}_W(\mathbf{b}) - \mathbf{w}\big) + \mathrm{proj}_{W^\perp}(\mathbf{b}).$$

由于$\mathbf{w} \in W$，知 $\mathrm{proj}_W(\mathbf{b}) - \mathbf{w} \in W$，它与$W^\perp$中的向量是正交的，故

$$\big(\mathrm{proj}_W(\mathbf{b}) - \mathbf{w}\big) \cdot \mathrm{proj}_{W^\perp}(\mathbf{b}) = 0.$$

根据勾股定理，我们得到

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathrm{proj}_W(\mathbf{b}) - \mathbf{w}\|^2 + \|\mathrm{proj}_{W^\perp}(\mathbf{b})\|^2 \geq \|\mathrm{proj}_{W^\perp}(\mathbf{b})\|^2 = \|\mathbf{b} - \mathrm{proj}_W(\mathbf{b})\|^2.$$

即$\|\mathbf{b} - \mathbf{w}\| \geq \big\|\mathbf{b} - \mathrm{proj}_W(\mathbf{b})\big\|$. 等号成立当且仅当$\|\mathrm{proj}_W(\mathbf{b}) - \mathbf{w}\| = 0$，当且仅当$\mathbf{w} = \mathrm{proj}_W(\mathbf{b})$. □

**定理 5.66.** *(1) 线性方程组$A\mathbf{x} = \mathbf{b}$的最小二乘解存在，且其任何一个最小二乘解$\hat{\mathbf{x}}$满足$proj_{\mathsf{Col}(A)}\mathbf{b} = A\hat{\mathbf{x}}$，也满足$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. (2) 考虑线性方程组$A\mathbf{x} = \mathbf{b}$，则线性方程组$A^TA\mathbf{x} = A^T\mathbf{b}$必定有解. 且$A^TA\mathbf{x} = A^T\mathbf{b}$的所有解均为$A\mathbf{x} = \mathbf{b}$的最小二乘解.*

*证明.* 首先，我们注意到$A^T\mathbf{u} = \mathbf{0}$等价于$\mathbf{u} \in \mathsf{Null}(A^T) = \mathsf{Col}(A)^\perp$.

(1) 由之前的推导及前一个定理，知最小二乘解$\hat{\mathbf{x}}$存在，且满足$\mathrm{proj}_{\mathsf{Col}(A)}\mathbf{b} = A\hat{\mathbf{x}} \in \mathsf{Col}(A)$. 进而

$$\mathbf{b} - A\hat{\mathbf{x}} = \mathbf{b} - \mathrm{proj}_{\mathsf{Col}(A)}\mathbf{b} = \mathrm{proj}_{\mathsf{Col}(A)^\perp}\mathbf{b} \in \mathsf{Col}(A)^\perp.$$

因而$A^T(A\hat{\mathbf{x}} - \mathbf{b}) = \mathbf{0}$.

(2) 注意到$A^T\mathbf{b} \in \mathsf{Col}(A^T)$，而$\mathsf{Col}(A^TA) = \mathsf{Col}(A^T)$. 故存在$\hat{\mathbf{x}}$，使得$(A^TA)\hat{\mathbf{x}} = A^T\mathbf{b}$. 此时有$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$，于是$A\hat{\mathbf{x}} - \mathbf{b} \in \mathsf{Col}(A)^\perp$. 这时，我们有正交分解式

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}}),$$

其中$A\hat{\mathbf{x}} \in \mathsf{Col}(A)$，而$\mathbf{b} - A\hat{\mathbf{x}} \in Col(A)^\perp$. 故$A\hat{\mathbf{x}} = \mathrm{proj}_{\mathsf{Col}(A)}(\mathbf{b}) = \min_{\mathbf{w}\in\mathsf{Col}(A)} \|\mathbf{w} - \mathbf{b}\|$，即$\hat{\mathbf{x}}$是$A\mathbf{x} = \mathbf{b}$的最小二乘解. □

需要注意的是，矩阵$A$未必满秩，矩阵$A^TA$不一定可逆，但总可以求出$A^TA\mathbf{x} = A^T\mathbf{b}$的解.

**例 5.67.** *求下列线性方程组的所有最小二乘解：*

$$\begin{cases} x_1 + x_2 &= 0 \\ \qquad\ \ x_2 &= 1 \\ x_1 &= 2 \end{cases}$$

解. 将方程组写为$A\mathbf{x}=\mathbf{b}$，有

$$A=\begin{bmatrix}1&1\\0&1\\1&0\end{bmatrix},\quad \mathbf{b}=\begin{bmatrix}0\\1\\2\end{bmatrix}.$$

计算得

$$A^TA=\begin{bmatrix}2&1\\1&2\end{bmatrix},\quad A^T\mathbf{b}=\begin{bmatrix}1\\2\end{bmatrix}.$$

解方程$(A^TA)\mathbf{x}=A^T\mathbf{b}$，得唯一解$\widehat{\mathbf{x}}=\begin{bmatrix}0\\1\end{bmatrix}$，即为所求的最小二乘解. □

**例 5.68.** 设$W=\mathit{span}\{\mathbf{w}_1,\mathbf{w}_2\}$，其中

$$\mathbf{w}_1=(1,0,1),\quad \mathbf{w}_2=(1,1,0).$$

求$\mathbb{R}^3$中向量$\mathbf{b}=(0,1,2)$向子空间$W$作正交投影所得的向量.

解. 令

$$A=[\mathbf{w}_1 \,\vdots\, \mathbf{w}_2]=\left[\begin{array}{c:c}1&1\\0&1\\1&0\end{array}\right],$$

则$\mathsf{Col}(A)=\mathsf{span}\{\mathbf{w}_1,\mathbf{w}_2\}=W$. 已经计算得到线性方程组$A\mathbf{x}=\mathbf{b}$的最小二乘解为

$$\widehat{\mathbf{x}}=\begin{bmatrix}0\\1\end{bmatrix}.$$

我们有

$$\mathrm{proj}_W(\mathbf{b})=\mathrm{proj}_{\mathsf{Col}(A)}(\mathbf{b})=A\widehat{\mathbf{x}}=\begin{bmatrix}1&1\\0&1\\1&0\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}1\\1\\0\end{bmatrix}.$$

□

在本节的最后，回顾多项式拟合的问题，我们可以得到以下结论.

**例 5.69.** 考虑用不超过$n-1$次的多项式

$$y(x)=c_0+c_1x+\ldots+c_{n-1}x^{n-1}$$

拟合$m$个二元数组$(x_1,y_1),(x_2,y_2),\ldots,(x_m,y_m)$. 则多项式系数$c_0,c_1,\ldots,c_{n-1}$由线性方程组

$$M^TM\mathbf{c}=M^T\mathbf{y},$$

给出，其中

$$M=\begin{bmatrix}1&x_1&\ldots&x_1^{n-1}\\1&x_2&\ldots&x_2^{n-1}\\\vdots&\vdots&\vdots&\vdots\\1&x_m&\cdots&x_m^{n-1}\end{bmatrix},\quad \mathbf{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix},\quad \mathbf{c}=\begin{bmatrix}c_0\\c_1\\\vdots\\c_{n-1}\end{bmatrix}.$$

这样的向量$\mathbf{c}$满足$\|\mathbf{y}-M\mathbf{c}\|=\min_{\mathbf{v}\in\mathbb{R}^n}\|\mathbf{y}-M\mathbf{v}\|$.

## 5.8 二次型

空间$\mathbb{R}^n$上所有可能的内积$\langle\cdot,\cdot\rangle_B$形式为

$$\langle\mathbf{x},\mathbf{y}\rangle_P = P\mathbf{x}\cdot P\mathbf{y} = \mathbf{y}^T P^T P\mathbf{x}, \qquad (\mathbf{x},\mathbf{y}\in\mathbb{R}^n).$$

前者为欧式内积的点乘形式，后者为矩阵形式，而$P\in M_n(\mathbb{R})$是可逆矩阵. 我们记$M^TM = C = [c_{ij}]$，$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n\end{bmatrix}^T$，$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n\end{bmatrix}^T$. 则内积的具体表达式为

$$\langle\mathbf{x},\mathbf{y}\rangle_M = \mathbf{y}^T C\mathbf{x} = \sum_{k=1}^n\sum_{l=1}^n y_k c_{kl} x_l = \sum_{k=1}^n\sum_{l=1}^n c_{kl} y_k x_l.$$

而模长的表达式为

$$\|\mathbf{x}\|^2 = \mathbf{x}^T C\mathbf{x} = \sum_{k=1}^n\sum_{l=1}^n c_{kl} x_k x_l.$$

**定义 5.13.** *我们称2次齐次多项式为二次型，当它有$n$个变元$x_1,x_2,\dots,x_n$时，形如*

$$Q(x_1,x_2,\dots,x_n) = \sum_{1\le i\le j\le n} c_{ij}x_ix_j,$$

*其中$c_{ij}\in\mathbb{R}$ $(1\le i\le j\le n)$.* （在这一节中，我们仅考虑系数为实数的情形.）

**例 5.70.** *以下均为二次型：*

- ◇ $Q(x_1,x_2,x_3) = x_1^2 + x_2^2 + x_3^3$.
- ◇ $Q(x,y) = (x/a)^2 + (y/b)^2$.
- ◇ $Q(x,y) = (x/a)^2 - (y/b)^2$.
- ◇ $Q(x,y) = xy$.
- ◇ $Q(x_1,x_2,x_3) = 2x_1^2 + 6x_1x_2 - 5x_2^2 + 3x_2x_3$.

作为多项式，我们总有$x_1x_2 = x_2x_1$，可以写成对称的形式：

$$c_{12}x_1x_2 = \frac{c_{12}}{2}x_1x_2 + \frac{c_{12}}{2}x_2x_1.$$

这时，我们有

$$Q(x_1,x_2,\dots,x_n) = \sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j,$$

这里$a_{ii} = c_{ii}$ $(1\le i\le n)$，而

$$a_{ij} = \frac{1}{2}c_{ij}, (i<j), \quad a_{ij} = \frac{1}{2}c_{ji}, (i>j).$$

将其写成矩阵的形式，则有

$$Q(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 & \dots & x_n\end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n}\\ a_{21} & a_{22} & \dots & a_{2n}\\ \dots & \dots & \dots & \dots\\ a_{n1} & a_{n2} & \dots & a_{nn}\end{bmatrix}\begin{bmatrix} x_1\\ x_2\\ \dots\\ x_n\end{bmatrix} = \mathbf{x}^T A\,\mathbf{x}.$$

这里出现的方阵$A$是对称矩阵，我们将其称为关于二次型$Q(\mathbf{x})$的矩阵.

反之，对任意矩阵$C = [c_{ij}]\in M_n(\mathbb{R})$，记

$$Q_C(\mathbf{x}) = \mathbf{x}^T C\,\mathbf{x} = \sum_{i=1}^n\sum_{j=1}^n c_{ij}x_ix_j,$$

则$Q_C(\mathbf{x})$是一个二次型. 由于每个方阵都可以写成一个对称矩阵与一个斜对称矩阵之和，记$C = A + B$，其中$A$对称、$B$斜对称. 这时我们得到

$$Q_C(\mathbf{x}) = \mathbf{x}^T C\mathbf{x} = \mathbf{x}^T A\mathbf{x} + \mathbf{x}^T B\mathbf{x} := Q_A(\mathbf{x}) + Q_B(\mathbf{x}).$$

对于斜对称矩阵$B$，我们有$B^T = -B$. 又$\mathbf{x}^T B\mathbf{x}$是一个数，或1阶矩阵. 于是

$$Q_B(\mathbf{x}) = \mathbf{x}^T B\mathbf{x} = (\mathbf{x}^T B\mathbf{x})^T = \mathbf{x}^T B^T\mathbf{x} = \mathbf{x}^T(-B)\mathbf{x} = -Q_B(\mathbf{x}).$$

因此我们总有$Q_B(\mathbf{x}) = 0$，以及$Q_C(\mathbf{x}) = Q_A(\mathbf{x})$. 所以在实二次型的理论中，我们总假设二次型与对称矩阵一一对应.

在抽象的内积空间$(V, \langle\cdot,\cdot\rangle)$上,我们也可以定义二次型:设$L$为$V$上的算子,满足$L^T = L$,则定义$Q_L(\mathbf{v}) = \langle L\mathbf{v}, \mathbf{v}\rangle$ $(\mathbf{v} \in V)$.

**例 5.71.** 将下列二次型写成$\mathbf{x}^T A\mathbf{x}$的形式，其中$A$对称.

*(1)* $2x^2 + 6xy - 5y^2$.

*(2)* $x_1^2 + 7x_2^2 - 3x_3^2 + 4x_1x_2 - 2x_1x_3 + 8x_2x_3$.

解.（略.） □

本节开头的内积$\langle\cdot,\cdot\rangle_M$与常规的欧式内积的区别，只在于乘了矩阵$P$.

**定义 5.14.** 当$P \in M_n(\mathbb{R})$为可逆矩阵时，我们称$\mathbf{x} = P\mathbf{y}$为一个变量替换.

当$P \in M_n(\mathbb{R})$为正交矩阵时，我们称$\mathbf{x} = P\mathbf{y}$为一个正交变量替换.

注意到

$$\mathbf{x}^T A\,\mathbf{x} = (P\mathbf{y})^T A\,(P\mathbf{y}) = \mathbf{y}^T(P^T AP)\mathbf{y}.$$

由于$A$对称,我们总能通过正交对角化,找到正交阵$P$与对角阵$D$,使得$P^T AP = D$. 记$D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$，则

$$\mathbf{x}^T A\,\mathbf{x} == \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2.$$

**定理 5.72.** 设$A \in M_n(\mathbb{R})$为对称矩阵. 则存在正交变量替换$\mathbf{x} = P\mathbf{y}$，使得二次型$Q_A(\mathbf{x})$变为对角二次型$Q_A(\mathbf{y})$（即无交叉项的二次型）.

**例 5.73.** 试通过正交变量替换将下列二次型对角化

$$Q = 2x_1^2 + 2x_2^2 + 2x_3^2 + 6x_1x_2 + 6x_1x_3 + 6x_2x_3,$$

并将$Q$用新的变量表出.

解. 我们有$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$，其中

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

在之前的例子里，我们已经得到$A$的正交对角化，即$P^T AP = D$，其中

$$P = \left[\begin{array}{c:c:c} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{array}\right], \quad D = \begin{bmatrix} -1 & & \\ & -1 & \\ & & 8 \end{bmatrix}.$$

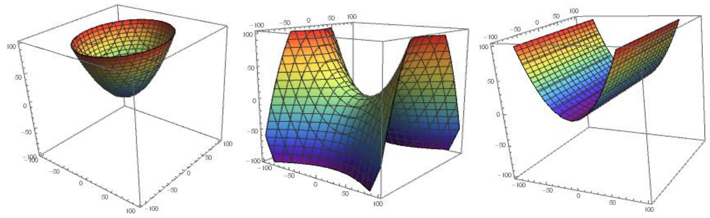
做正交变换替换$\mathbf{x} = P\mathbf{y}$，则有

$$Q(\mathbf{y}) = \mathbf{y}^T D\mathbf{y} = -y_1^2 - y_2^2 + 8y_3^2.$$

□

下面，我们来研究一下两个变量的二次型的图像. 通过正交变量替换，我们假设二次型具有对角形式

$$Q(x,y)=\lambda_1x^2+\lambda_2y^2.$$



(h) $\lambda_1>0,\ \lambda_2>0$  (i) $\lambda_1>0,\ \lambda_2<0$  (j) $\lambda_1>0,\ \lambda_2=0$

情形1. $\lambda_1>0,\ \lambda_2>0$: $Q(x,y)\geq 0$，且$Q(x,y)=0$仅在$(x,y)=(0,0)$时取到.
情形2. $\lambda_1>0,\ \lambda_2<0$: $Q(1,0)>0$，而$Q(0,1)<0$.
情形3. $\lambda_1>0,\ \lambda_2=0$: $Q(x,y)\geq 0$, 而除了原点外，亦有$Q(0,1)=0$.
在其余情形，我们只需更改$\lambda_1$和/或$\lambda_2$的正负号，便可以得到相应的结论.

**定义 5.15.** 矩阵$A\in M_n(\mathbb{R})$，或二次型$\mathbf{x}^TA\,\mathbf{x}$, 称为

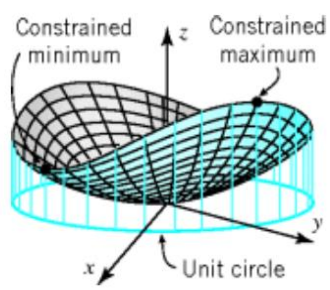
- ◇ 正定的，是指$\mathbf{x}^TA\,\mathbf{x}>0$对任意$\mathbf{x}\neq 0$成立;
- ◇ 负定的，是指$\mathbf{x}^TA\,\mathbf{x}<0$对任意$\mathbf{x}\neq 0$成立;
- ◇ 半正定的，是指$\mathbf{x}^TA\,\mathbf{x}\geq 0$对任意$\mathbf{x}$均成立;
- ◇ 半负定的，是指对任意$\mathbf{x}$均成立;
- ◇ 不定的，是指$\mathbf{x}^TA\,\mathbf{x}\leq 0$可以取到正值和负值.

根据之前的讨论以及二元二次型的例子，我们不难得到以下结论.

**定理 5.74.** 设$A$为对称矩阵，则

*(a)* 矩阵$A$或二次型$\mathbf{x}^TA\,\mathbf{x}$正定，当且仅当$A$的特征值均为正数；
*(b)* 矩阵$A$或二次型$\mathbf{x}^TA\,\mathbf{x}$负定，当且仅当$A$的特征值均为负数；
*(c)* 矩阵$A$或二次型$\mathbf{x}^TA\,\mathbf{x}$半正定，当且仅当$A$的特征值均为非负实数；
*(d)* 矩阵$A$或二次型$\mathbf{x}^TA\,\mathbf{x}$半负定，当且仅当$A$的特征值均为非正实数；
*(e)* 矩阵$A$或二次型$\mathbf{x}^TA\,\mathbf{x}$不定，当且仅当$A$的特征值即有正数、又有负数.

接下来我们一类考虑条件极值问题：二次型$\mathbf{x}^TA\,\mathbf{x}$在单位球面上（即限制条件 $\|\mathbf{x}\|=1$下）的最大、最小值.




**定理 5.75.** 设$A\in M_n(\mathbb{R})$为对称矩阵，其特征值为$\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n$. 则

*(1)* 二次型$\mathbf{x}^TA\mathbf{x}$在单位球面上可取到最大、最小值;
*(2)* 其最大值在关于特征值$\lambda_1$的特征向量上取到;
*(2)* 其最小值在关于特征值$\lambda_n$的特征向量上取到.

证明. 将矩阵$A$正交对角化，设$A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ $(1 \le i \le n)$，其中$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$构成$\mathbb{R}^n$的标准正交基. 取$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}$以及$D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$，则$P^T A P = D$. 作正交变量替换$\mathbf{x} = P\mathbf{y}$下，我们有$\mathbf{v}_i = P\mathbf{e}_i$ $(1 \le i \le n)$，以及

$$\|\mathbf{y}\| = \|P\mathbf{x}\| = \|\mathbf{x}\| = 1.$$

上式用到了正交矩阵作为线性变换是保持向量长度不变的. 另外，我们有

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \sum_{i=1}^n \lambda_i y_i^2.$$

由于$\lambda_n \le \lambda_i \le \lambda_1$ $(1 \le \le i)$，得

$$\mathbf{x}^T A\mathbf{x} \le \sum_{i=1}^n \lambda_1 y_i^2 = \lambda_1 \|\mathbf{y}\|^2 = \lambda_1.$$

类似地，有$\mathbf{x}^T A\mathbf{x} \ge \lambda_n$. 另外，可以验证

$$\mathbf{v}_1^T A\mathbf{v}_1 = (P\mathbf{e}_1)^T A(P\mathbf{e}_1) = \mathbf{e}_1^T P^T A P\mathbf{e}_1 = \mathbf{e}_1^T D\mathbf{e}_1 = \lambda_1.$$

类似有$\mathbf{v}_n^T A\mathbf{v}_n = \lambda_n$. 定理得证. □

**例 5.76.** *求二次型$z = 5x^2 + 5y^2 + 4xy$在限制条件$x^2 + y^2 = 1$上的最大、最小值.*

解.（略.） □

## 5.9 极分解

（此灰色部分共有兴趣的同学们了解，不在大纲要求的范围内.）

对于复数$z \in \mathbb{C}$，我们有极坐标表达$z = re^{i\theta}$，其中$r \ge 0$，而$|e^{i\theta}| = 1$. 将$z$看成1阶复矩阵，则$r$和$e^{i\theta}$分别对应半正定矩阵和酉矩阵. 在这一节中，我们证明高阶的矩阵也有类似的分解形式. 方便起见，我们仅考虑实矩阵的情形，对于复矩阵，类似的结论也是成立的.

**定理 5.77** (极分解定理)**.** *设$A \in M_n(\mathbb{R})$. 则存在正交矩阵$P$以及半正定矩阵$H, H'$，使得*

$$A = HP = PH'.$$

证明. 根据奇异值分解，存在正交阵$U$、$V$以及对角矩阵$\Sigma$，使得$A = U\Sigma V^T$. 这时，我们有

$$A = (U\Sigma U^T)(UV) = (UV)(V^T \Sigma V).$$

取$P = UV$, $H = U\Sigma U^T$, $H' = V^T\Sigma V$. 因$U$、$V$均是正交阵，所以$P$为正交矩阵. 另外，不难验证$H$是对称的，且

$$\mathrm{det}(\lambda I - H) = \mathrm{det}(\lambda I - U\Sigma U^T) = \mathrm{det}\big(U(\lambda I - \Sigma)U^T\big) = \mathrm{det}(\lambda I - \Sigma),$$

即$H$与$\Sigma$的特征值相同，为非负实数. 所以矩阵$H$是半正定的. 类似可得$H'$是半正定矩阵. □

在课程的最后，我们用一张图标来展示复矩阵全体所具有的一些结构.

| | $\mathbb{C}$ | $M_n(\mathbb{C})$ | Eigenvalue |
|---|---|---|---|
| Unitary | $\bar{z} z = z \bar{z} = 1$<br>$\lvert z \rvert = 1$ | $U^* U = U U^* = I_n$ | $\lvert\lambda\rvert = 1$. |
| Hermitian | $\bar{z} = z$<br>$z \in \mathbb{R}$ | $H^* = H$ | $\lambda \in \mathbb{R}$ |
| Semi-positive | $z \geq 0$ | $\forall x \in \mathbb{C}^n$, $x^* Q x \geq 0$<br>$\exists B \in M_n(\mathbb{C})$, $Q = B^* B$ | $\lambda \geq 0$ |
| Orthogonal Projection | $z = 0$ or $1$. | $E^2 = E = E^*$ | $\lambda \in \{0, 1\}$ |
| Rectangular Decomposition | $z = x + iy$<br>$(x, y \in \mathbb{R})$ | $A = H + iK$<br>$(H, K$ Hermitian$)$ | |
| Polar Decomposition | $z = re^{i\theta}$<br>$(r \geq 0, 0 \leq \theta < 2\pi)$ | $A = QU$<br>$Q$ semi-positive, $U$ unitary | |