# 2023线性代数1讲义

## 目录

# 1 第一章：线性方程组与矩阵

## 1.1. 线性方程组入门, Introduction to systems of linear equations

**Definition 1.1.** 令$n$为任意自然数，$a_1, \ldots, a_n$，$b$为常数。我们称$n$元**一次**方程

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

为一个(含有$n$个未知数的)**线性**方程。

也就是说，在本课程中，**线性**与**一次**是同一个意思。比如方程$2x_1^2 + \sin x_2 - 5x_3 = 0$不是一个线性方程，因为$x_1$的次数为2，并且$\sin x_2$也不能写成$x_2$的一次方形式。

**Definition 1.2.** 令$m$与$n$为任意自然数，$a_{11}, \ldots, a_{1n}$，$a_{21}, \ldots, a_{2n}$，$\ldots a_{m1}, \ldots, a_{mn}$，$b_1, \ldots, b_m$为常数，我们称以下$m$个$n$元一次方程组成的方程组

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

为一个**线性方程组**。特别地，若$b_1 = \ldots = b_m = 0$，则称该线性方程组为**齐次线性方程组**。显然，$a_{ij}$指代第$i$个方程里第$j$个未知数$x_j$的系数。这里$i$的取值范围为$1, 2, \ldots, m$, $j$的取值范围为$1, 2, \ldots, n$。

给定任何一个(由$m$个$n$元一次方程组成的)线性方程组

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

我们将由未知数系数$a_{ij}, i = 1, \ldots, m, j = 1, \ldots, n$组成的矩形点阵

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

为该线性方程组的**系数矩阵**(coefficient matrix)。显然$a_{ij}$位于该矩阵第$i$行与第$j$列相交的位置上。如果我们把方程组最右侧的常数$b_1, \ldots, b_m$作为**列**添加到系数矩阵的最右边，那么所得到的矩阵

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{bmatrix}$$

被称为该线性方程组的**增广矩阵**(augmented matrix)。
反之，若给定一个由$m$行和$n$列组成的矩阵$A$

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

(这里$a_{ij}$指代该矩阵第$i$行与第$j$列相交的位置上出现的常数)以及一个由$m$个坐标组成的向量$\boldsymbol{b} = (b_1, \ldots, b_m)$, 那么$A$与$b$可以确定一个(由$m$个$n$元一次方程组成的)线性方程组

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

显然，该线性方程组对应的系数矩阵为$A$，对应的增广矩阵为$(A, \boldsymbol{b})$。注意我们经常用记号$(A, \boldsymbol{b})$指代以下矩阵

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{bmatrix}$$

**Remark 1.3.** *1.* 若一个矩阵由$m$行和$n$列所构成，我们称其为一个$m \times n$-矩阵。因此一个(由$m$个$n$元一次方程组成的)线性方程组的系数矩阵为一个$m \times n$-矩阵，它的增广矩阵则为一个$m \times (n+1)$-矩阵。

*2.* 一个由$m$个坐标组成的向量(简称为$m$-向量)$\boldsymbol{b} = (b_1, \ldots, b_m)$可以被视为一个行向量，即一个$1 \times m$-矩阵；也可以被视为一个列向量，即一个$m \times 1$-矩阵, 且在此时记为

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

比如在记号$(A, \boldsymbol{b})$里我们就把$\boldsymbol{b}$看作一个列向量。究竟是把一个向量视为行向量还是列向量取决于具体的应用环境，我们在后面的课程学习中会逐渐了解。

*3.* 诚然每一个(由$m$个$n$元一次方程组成的)线性方程组都可以被它的增广矩阵唯一表示，但是矩阵作为线性代数里的一个独立概念并**不是必须**要与线性方程组联系起来(尽管目前我们使用矩阵的目的是为了求解线性方程组)。更多关于矩阵本身的知识与应用将在后续课程中逐渐展示。

例子：

1. 线性方程组

$$\begin{aligned} x - y &= 1 \\ 2x + y &= 6 \end{aligned}$$

的系数矩阵为

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

增广矩阵则为

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

2. 线性方程组

$$\begin{aligned} x + y &= 4 \\ 3x + 3y &= 6 \end{aligned}$$

的系数矩阵为

$$\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix},$$

增广矩阵则为

$$\begin{bmatrix} 1 & 1 & 4 \\ 3 & 3 & 6 \end{bmatrix}.$$

3. 矩阵$A =$

$$\begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \\ 3 & -3 & 6 \end{bmatrix}$$

与向量$\boldsymbol{b} = (5, 10, 15)$所对应的线性方程组为

$$\begin{aligned} x - y + 2z &= 5 \\ 2x - 2y + 4z &= 10 \\ 3x - 3y + 6z &= 15. \end{aligned}$$

即，该方程组的系数矩阵为$A$, 增广矩阵为$(A, \boldsymbol{b})$。

现在我们来解上面的第一个方程组

$$\begin{aligned} x - y &= 1 \\ 2x + y &= 6. \end{aligned}$$

注意它的增广矩阵为$B_1 =$

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

求解的思路当然是**消元**。具体步骤如下：

1. 首先我们将**第一个方程乘以**$-2$**, 然后加到第二个方程上去**来消掉第二个方程里的未知数$x$。我们所得到的新方程组为

$$\begin{aligned} x - y &= 1 \\ 3y &= 4. \end{aligned}$$

注意这个方程组的增广矩阵变成了$B_2 =$

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

显然，$B_2$是通过**将矩阵**$B_1$**的第一行乘以**$-2$**后加到第二行，同时保留第一行不变**所得到的。

2. 对于方程组

$$\begin{aligned} x - y &= 1 \\ 3y &= 4 \end{aligned}$$

我们将第二个方程等号两边都乘以1/3, 得到另一个方程组

$$\begin{aligned} x - y &= 1 \\ y &= \frac{4}{3} \end{aligned}$$

对应的增广矩阵为$B_3 =$

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

显然，$B_3$是通过**将矩阵$B_2$的第二行乘以$1/3$，同时保留第一行不变**所得到的。

3. 现在我们把方程组

$$\begin{aligned} x - y &= 1 \\ y &= \frac{4}{3} \end{aligned}$$

的**第二个方程加到第一个方程上面**，得到新方程组

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

显然这个最终得到的方程组给出了原方程组的解，其对应的增广矩阵为$B_4 =$

$$\begin{bmatrix} 1 & 0 & \frac{7}{3} \\ 0 & 1 & \frac{4}{3} \end{bmatrix}$$

并且系数矩阵为

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

注意$B_4$是通过**将矩阵$B_3$的第二行加到第一行，同时保留第二行不变**所得到的。

由以上观察我们可以得到以下结论：

- 对以上(由2个二元一次方程组成的)线性方程组进行**消元求解**的过程实际上是对该方程组所对应的**增广矩阵**的行进行一系列操作使其最终变为

$$\begin{bmatrix} 1 & 0 & s_1 \\ 0 & 1 & s_2 \end{bmatrix}$$

这种形式的过程，且此时$x = s_1, y = s_2$即为该方程组**唯一的一组解**。

- 在此过程中矩阵**列与列之间不发生任何关系**。

- 在此过程中矩阵的行与行之间所进行的操作包括:
    1. 将某一行乘以某个**非零**的常数，其他行保持不变;
    2. 将某一行乘以某个常数后加到另外一行上去，且除了被加的行之外其他的行保持不变。

基于此，我们引入一下重要概念:

**Definition 1.4** (初等行变换，elementary row operations)**.** *假设$C$为任意$m \times p$-矩阵。以下三种操作被称为(对矩阵$C$的)初等行变换:*

1. *将$C$的第$i$行 ($i = 1, \dots, m$) 乘以某个**非零**的常数$c$，其他行保持不变 (可记为: i-th row $\times c$ );*
2. *将$C$的第$i$行 ($i = 1, \dots, m$)乘以某个常数$c$后加到第$k$行上去 ($k = 1, \dots, m$且$k \neq i$)，且除了第$k$行之外其他的行保持不变 (可记为: i-th row $\times c +$ k-th row );*
3. *将$C$的第$i$行 ($i = 1, \dots, m$)与第$k$行 ($k = 1, \dots, m$且$k \neq i$) 互换 (可记为: i-th row $\leftrightarrow$ k-th row )，其他行保持不变。*

考虑线性方程组

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

及其对应的增广矩阵为$B = (A, \boldsymbol{b})$ ($A$为其系数矩阵)

$$
\begin{bmatrix}
a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\
a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{m1} & a_{m2} & \ldots & a_{mn} & b_m
\end{bmatrix}.
$$

那么由之前的讨论我们知道

1. “将第$i$个方程的等号两边乘以$c \neq 0$” 对应 “将$B$的第$i$行乘以某个**非零**的常数$c$，其他行保持不变”，即第一类初等行变换;
2. “将第$i$个方程乘以$c$再加到第$k$个方程上去”对应 “将$B$的第$i$行乘以某个常数$c$后加到第$k$行上去 ($k \neq i$)，且除了第$k$行之外其他的行保持不变”，即第二类初等行变换;
3. “将第$i$个方程与第$k$个方程互换”对应 “将$A$的第$i$行与第$k$行 ($k \neq i$) 互换，其他行保持不变”，即第三类初等行变换。

4. 若$m = n$, 即方程个数与未知数个数相等，且通过对增广矩阵$B = (A, \boldsymbol{b})$进行一系列初等行变换操作后得到以下形状的(增广)矩阵$\tilde{B} =$

$$\begin{bmatrix} 1 & 0 & \dots & 0 & s_1 \\ 0 & 1 & \dots & 0 & s_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & s_n \end{bmatrix},$$

那么该方程组具有**唯一的一组解**$x_1 = s_1, x_2 = s_2, \dots, x_n = s_n$。这里我们把$\tilde{B}$记为$\tilde{B} = (I_n, \boldsymbol{s})$, $I_n$**被称为**$n \times n$**–单位矩阵，它的对角线上所有元素均为**$1$**，其他不在对角线上的元素均为**$0$; 列向量$\boldsymbol{s} = (s_1, \dots, s_n)$是方程的解。

因此，从现在开始，我们可以**把对线性方程组** ($m = n$) **的消元求解过程**转化为如何通过一系列初等行变换将其对应的增广矩阵$B = (A, \boldsymbol{b})$转化为$\tilde{B} = (I_n, \boldsymbol{s})$的过程，此时$\boldsymbol{s}$即为方程组的唯一解。特别地，**我们现在操作的对象由方程组变为矩阵**。

在结束本节之前，我们需要注意到另一个非常重要的问题: **并非所有线性方程组对应的增广矩阵都可以通过初等行变换转为**$\tilde{B} = (I_n, \boldsymbol{s})$**的形式!** 比如当方程个数$m$与未知数个数$n$不相等时，原增广矩阵$B$无法变成$\tilde{B} = (I_n, \boldsymbol{s})$的形式 (为什么? ); 同时，即使$m = n$，一些线性方程组的增广矩阵依然不能通过初等行变换转为$\tilde{B} = (I_n, \boldsymbol{s})$的形式，比如方程组

$$\begin{aligned} x - y + 2z &= 5 \\ 2x - 2y + 4z &= 10 \\ 3x - 3y + 6z &= 15 \end{aligned}$$

对应的增广矩阵$B =$

$$\begin{bmatrix} 1 & -1 & 2 & 5 \\ 2 & -2 & 4 & 10 \\ 3 & -3 & 6 & 15 \end{bmatrix}$$

通过消元求解所对应的初等行变换后最终所得到的形式是

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

并不满足$(I_3, \boldsymbol{s})$这样的形式。注意此时可以看出该方程组具有**无穷多个解**$x = 5 + r - 2s, y = r, z = s$，这里$r$和$s$可以为任何数。

作为总结，我们在这一节中的收获为:

- 我们可以**把对线性方程组** ($m = n$) **的消元求解过程**转化为如何通过一系列初等行变换将其对应的增广矩阵$B = (A, \boldsymbol{b})$转化为$\tilde{B} = (I_n, \boldsymbol{s})$的过程。此时向量$\boldsymbol{s}$给出方程的**唯一解**。
- 并非所有方程组的增广矩阵$B$都可以通过一系列初等行变换转化为$\tilde{B} = (I_n, \boldsymbol{s})$的形式。
- **一个线性方程组的增广矩阵$B$通过一系列初等行变换所转化成的某种"最终形态"决定了该方程组是否有解，以及解是否唯一**。

在1.2节我们将学习如何将一个线性方程组的增广矩阵$B$通过一系列初等行变换转化成的一种特殊的"最终形态"，以及如何通过这个"最终形态"确定该方程组是否有解，解是否唯一，以及求出具体解 (在有解的情况下)。

### 1.2. 高斯消元法，Gaussian Elimination

在上一节我们提到，可以通过一系列初等行变换，将一个线性方程组的增广矩阵转化为一种特殊的"最终形态"，这种"最终形态"能够帮助我们确定该方程组是否有解，解是否唯一，以及求出具体解 (在有解的情况下)。在本节我们将会学习到这种"最终形态"就是所谓的阶梯型 (row echelon form) 或者简化阶梯型 (reduced row echelon form)。首先我们给出定义：

**Definition 1.5** (阶梯型与简化阶梯型). *若一个矩阵满足以下三个条件:*

1. *该矩阵中任何不为*0*的行里所包含的第一个非零的项都是*1*。我们称这样的项为主*1 *(leading* 1*);*
2. *该矩阵中任何为*0*的行一定在不为*0*的行的下方;*
3. *对于该矩阵中任意两个**相邻且不为**0的行，**上面**的行里所包含的主*1*一定位于**下面**的行里所包含的主*1*的**左侧**;*

*那么该矩阵被称为**阶梯型**。若一个**阶梯型**额外满足以下条件*

- *该矩阵中任何不为*0*的行里所包含的主*1*是其所在的**列里唯一一个不为**0*的项,*

*那么该矩阵被称为**简化阶梯型**。*

**例子：**

1. 矩阵
$$\begin{bmatrix}1 & -1 & 2\\0 & 2 & 0\\0 & 0 & 0\end{bmatrix}$$
不是阶梯型，因为第二行中第一个非零的项是2，不等于1。
2. 矩阵
$$\begin{bmatrix}1 & -1 & 2\\0 & 0 & 0\\0 & 0 & 1\end{bmatrix}$$
不是阶梯型，因为不为0的第三行在为0的第二行下面。
3. 矩阵
$$\begin{bmatrix}1 & -1 & 2\\0 & 0 & 1\\0 & 1 & 0\end{bmatrix}$$
不是阶梯型，因为第二行的主1在第三行主1的右边。
4. 矩阵
$$\begin{bmatrix}1 & 4 & -3 & 7\\0 & 0 & 6 & 2\\0 & 0 & 1 & 5\end{bmatrix}$$
是阶梯型，但不是简化阶梯型，因为第二行的主1不是它所在列中唯一的非零项，同样，第三行的主1不是它所在列中唯一的非零项。
5. 矩阵
$$\begin{bmatrix}1 & 0 & 0 & 4\\0 & 1 & 0 & 7\\0 & 0 & 1 & -1\end{bmatrix}$$
是简化阶梯型。

下面我们介绍如果将任意给定的一个矩阵通过一系列初等行变化将其转化为一个阶梯型或者简化阶梯型。大家可以参照教材第14页到17页的两个例子的具体计算过程。我们这里使用教材第14页上的矩阵作为示例：考虑矩阵$B =$
$$\begin{bmatrix}0 & 0 & -2 & 0 & 7 & 12\\2 & 4 & -10 & 6 & 12 & 28\\2 & 4 & -5 & 6 & -5 & -1\end{bmatrix}.$$

1. 首先我们知道原矩阵第一行的第一个非零项必然在第一列之后 (实际上通过观察第二列的情况可知在第二列之后)，而原矩阵第二行与第三行的第一个非零项都在第一列里，这意味着目前第一行所包含的主1必然在目前第二行

和第三行所包含的主1右边，与阶梯型要求的条件3不符。因此我们需要将第一行与第二行或者与第三行进行交换 (来保证阶梯型的条件3成立)。这里**第一行与第二行交换**，所得到的矩阵为$B_1 =$

$$\begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}.$$

2. 现在我们将**第一行乘以**$1/2$使得第一行第一个非零项为1，从而得到第一行的主1: $B_2 =$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}.$$

3. 现在我们注意到$B_2$里第三行的第一个非零项2仍在第一列里，这意味着$B_2$的第三行的主1与第一行的主1都在第一列里，不满足阶梯型的条件3。因此我们必须把第三行第一列上的项消掉，即，**将第一行乘以**$-2$**加到第三行上**，得到$B_3 =$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix}.$$

4. 现在我们注意到$B_3$第二行的第一个非零项为$-2$位于第三列上。通过对**第二行乘以**$-1/2$我们得到第二行的主1: $B_4 =$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix}.$$

5. 注意到$B_4$里第三行的第一个非零项为5，它与第二行的主1处于同一列上，不符合阶梯型条件3，因此我们仿照第三步，通过将**第二行乘以**$-5$**加到第三行**，将该项消掉，得到$B_5 =$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}.$$

6. 现在只需对**第三行乘以**2去得到第三行的主1: $B_6 =$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

显然，根据定义，矩阵$B_6$是一个阶梯型。如果我们想将其继续转化为简化阶梯型，则需要将第三列 (即第二行的主1所在的列) 和第五列 (即第三行的主1所在的列) 通过初等行变换分别转化为

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

与

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

最终得到的简化阶梯型为$R =$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & 7 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

在实际操作中，大家可以根据自己的个人喜好，经验，以及矩阵的具体样子来决定如何进行初等行变换将其转化为阶梯型或简化阶梯型。比如可以遵循下面的几个核心步骤：

- 首先浏览矩阵的所有行，找出所有不为0的行，并将所有为0的行全部移动到矩阵的下方，确保所有不为0的行都在所有为0的行的上方。
- 在每个不为0的行里都找出其包含的第一个不为0的项所在的列。假设第$i$行所包含的第一个不为0的项所在的列是其中**最靠左**的，那么将第$i$行与第一行进行交换。
- 在交换过后得到的矩阵里，第一行所包含的第一个不为0的项假设为常数$a$，将第一行乘以$1/a$，得到第一行里的主1。
- 假设现在所得到的矩阵里第一行的主1所在的列为第$j$列，那么通过初等行变换将该列里所有其他的项 (即首行主1下方的所有项)均消为0。注意这一步的意图是为了确保首行所包含的主1处于最左边。
- 现在可以暂时忘记第一行和第一列，我们观察剩下的$m-1$行和$n-1$列 (假设原矩阵为$m \times n$-矩阵)，然后对其**重复以上步骤**。直到将其化为阶梯型为止。
- 将该阶梯型里所有行包含的主1上方和下方的非零项通过初等行变换全部消为0，即可得到简化阶梯型。

现在我们将原矩阵$B =$

$$\begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

视为方程组

$$
\begin{aligned}
-2x_3 + 7x_5 &= 12 \\
2x_1 + 4x_2 - 10x_3 + 6x_4 + 12x_5 &= 28 \\
2x_1 + 4x_2 - 5x_3 + 6x_4 - 5x_5 &= -1
\end{aligned}
$$

的增广矩阵，那么该方程组的解与其阶梯型$B_6 =$

$$
\begin{bmatrix}
1 & 2 & -5 & 3 & 6 & 14 \\
0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\
0 & 0 & 0 & 0 & 1 & 2
\end{bmatrix}
$$

所对应的方程组

$$
\begin{aligned}
x_1 + 2x_2 - 5x_3 + 3x_4 + 6x_5 &= 14 \\
x_3 - \frac{7}{2}x_5 &= -6 \\
x_5 &= 2
\end{aligned}
$$

具有同样的解；类似的，与其简化阶梯型$R =$

$$
\begin{bmatrix}
1 & 2 & 0 & 3 & 0 & 7 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 2
\end{bmatrix}.
$$

所对应的方程组

$$
\begin{aligned}
x_1 + 2x_2 + 3x_4 &= 7 \\
x_3 &= 1 \\
x_5 &= 2
\end{aligned}
$$

具有同样的解。显然不论是方程组

$$
\begin{aligned}
x_1 + 2x_2 - 5x_3 + 3x_4 + 6x_5 &= 14 \\
x_3 - \frac{7}{2}x_5 &= -6 \\
x_5 &= 2
\end{aligned}
$$

还是方程组

$$
\begin{aligned}
x_1 + 2x_2 + 3x_4 &= 7 \\
x_3 &= 1 \\
x_5 &= 2
\end{aligned}
$$

都能容易的给出原方程组的解：

$$x_1 = 7 - 3r - 2s, x_2 = s, x_3 = 1, x_4 = r, x_5 = 2$$

这里$r, s$可以为任意常数。

我们将把一个线性方程组对应的**增广矩阵**通过一系列**初等行变换**转化为阶梯型的过程称之为高斯消元法 (Gauss elimination)，把一个线性方程组对应的**增广矩阵**通过一系列**初等行变换**转化为简化阶梯型的过程称之为高斯-若尔当消元法 (Gauss–Jordan elimination)。

**Definition 1.6.** *假设$B$是一个线性方程组的增广矩阵，其对应的阶梯型$\tilde{B}$或者简化阶梯型$R$里不为0的行里所包含的主1所对应的未知数被称为该方程组的主元 (leading variable)，主元之外的所有未知数被称为该方程组的自由元 (free variable)。*

比如在上面的例子里，由于简化阶梯型为$R =$

$$\begin{bmatrix} 1 & 2 & 0 & 3 & 0 & 7 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix},$$

我们知道$x_1, x_3, x_5$为方程组的主元， $x_2, x_4$为方程组的自由元。

由以上的计算容易看出，若方程组有解，那么该方程组的自由元所对应的解可以取任意值 (即自由取值，这是自由元名字里“自由”一词的由来)；而当我们给自由元赋予特殊的值之后，主元所对应的解则被唯一确定。比如上面的例子中，我们已求出方程组

$$\begin{aligned} -2x_3 + 7x_5 &= 12 \\ 2x_1 + 4x_2 - 10x_3 + 6x_4 + 12x_5 &= 28 \\ 2x_1 + 4x_2 - 5x_3 + 6x_4 - 5x_5 &= -1 \end{aligned}$$

的解为

$$x_1 = 7 - 3r - 2s, x_2 = s, x_3 = 1, x_4 = r, x_5 = 2$$

这里$r$, $s$可以为任意常数。若我们选择$r = 1$, $s = 1$, 那么在对应的解里面，主元$x_1$必然只能取值为$x_1 = 7 - 3 - 2 = 2$，此外$x_3 = 1$和$x_5 = 2$显然已经被唯一确定。

我们将

$$x_1 = 7 - 3r - 2s, x_2 = s, x_3 = 1, x_4 = r, x_5 = 2; \quad r, s \text{ 为任意常数}$$

这样的表示称为方程组的**通解 (general solution)**，将

$$x_1 = 2, x_2 = 1, x_3 = 1, x_4 = 1, x_5 = 2$$

(选择$r = s = 1$后得到的解)这样的写法称为方程组的**特解**。显然**特解是通过对通解里的自由元赋予具体数值后得到**。

通过本节内的计算与观察，我们可以看出来一个线性方程组的增广矩阵经过初等行变换所得到的阶梯型或简化阶梯型的样子能够决定这个方程组是否有解，以及有解时是否有唯一解。总结如下: 设某个(由$m$个$n$元一次方程组成的)方程组的增广矩阵$B = (A, \boldsymbol{b})$ ($A$为系数矩阵)，其对应的简化阶梯型为$R = (U, \boldsymbol{s})$ (这里$\boldsymbol{s}$为矩阵$R$最右边的列)。

- 假如存在某个$i \in \{1, \ldots, m\}$使得矩阵$U$的第$i$行为0，而列向量$\boldsymbol{s}$的第$i$个坐标$s_i \neq 0$，那么该方程组无解。反之该方程组至少有一个解。
- 如果矩阵$U = I_n$为$n \times n$-单位矩阵，那么该方程组有且只有一组解，这个解就是向量$\boldsymbol{s}$。
- 若方程有解，且$U \neq I_n$，那么该方程组有无穷多组解。

我们将在后续的学习中给出以上结论的严格证明。这里我们注意到以上结论已经回答了第1.1节最后提出的猜测。此外，我们很容易看出

- 假设某个(含有$n$个未知数的)线性方程组的增广矩阵为$B$，$\tilde{B}$与$R$分别为由$B$转化的阶梯型和简化阶梯型。那么该方程组主元的个数$k = \tilde{B}$或者$R$里主1的个数 = $\tilde{B}$或者$R$里非零行的个数；该方程组自由元的个数$= n - k$。

在本节的最后我们讨论一类特殊的线性方程组：齐次线性方程组

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

显然，任何齐次线性方程组都至少有一个解：

$$x_1 = 0, x_2 = 0, \ldots, x_n = 0.$$

这样的解被称为**平凡解 (trivial solution)**。如果齐次线性方程组有区别于平凡解的其他解，那么这些解被称为非平凡解 (nontrivial solutions)。注意：**若一个齐次线性方程组有一个非平凡解，那么它必然有无穷多个非平凡解** (请思考为什么)。

## 1.3. 矩阵与矩阵运算，Matrices and Matrix Operations

在本节中大家不需要把矩阵与线性方程组联系起来。我们单独研究矩阵这个对象。

首先我们引入一下记号和定义：

- 所有$m \times n$-矩阵的集合记为$M_{m\times n}$。因此我们经常用$A \in M_{m\times n}$指代$A$是一个$m \times n$-矩阵。
- 若$m = n$, 即矩阵的行数等于列数，那么我们也称$n\times n$-矩阵为$n$阶**方阵 (square matrix)**。
- 对于一个$n \times n$-方阵$A =$
$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$
我们称向量$(a_{11}, a_{22}, \dots, a_{nn})$为$A$的**对角线**。注意对角线是只有**方阵**才有的概念。
- 对于$m \times n$-矩阵$A =$
$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$
有时候我们也用$A_{ij}$指代$a_{ij}$, 用$[a_{ij}]$表示$A$。
- $1 \times 1$-矩阵就是一个常数$a$。
- 对于一个向量$\boldsymbol{b} = (b_1, b_2, \dots, b_n)$，我们可以将其视为一个行向量，即$1\times n$-矩阵，此时将其记为
$$\boldsymbol{b} = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix},$$
也可以将其视为一个列向量，即$n \times 1$-矩阵，此时记为
$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

- 对于$m \times n$-矩阵$A =$
$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$
我们将其第$i$行记为$\boldsymbol{r}_i = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix}, i = 1, \dots, m$，那么$A$可以被表示为 $A =$
$$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix}.$$
我们称其为矩阵$A$的行向量表示。
- 类似的，对于$m \times n$-矩阵$A =$
$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$
我们将其第$j$列记为$\boldsymbol{c}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$，那么$A$可以被表示为 $A =$
$$\begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \dots & \boldsymbol{c}_n \end{bmatrix}.$$
我们称其为矩阵$A$的列向量表示。

本章的主要内容是定义矩阵的运算法则：

1. 矩阵$A \in M_{m_1 \times n_1}$与矩阵$B \in M_{m_2 \times n_2}$**相等**，即$A = B$，当且仅当$m_1 = m_2$，$n_1 = n_2$，并且$A_{ij} = B_{ij}$对一切$i = 1, \dots, m_1 = m_2$和$j = 1, \dots, n_1 = n_2$成立。
2. 对于矩阵$A \in M_{m \times n}$, 常数$c$, 矩阵$A$与$c$的**标量积(scalar product)**，记为$cA$，定义为 $(cA)_{ij} = cA_{ij}$。显然$cA \in M_{m \times n}$。
3. 若矩阵$A \in M_{m \times n}$与矩阵$B \in M_{m \times n}$，那么$A$与$B$的**和**(sum)定义为 $(A+B)_{ij} = A_{ij} + B_{ij}$。类似的，$A$与$B$的**差**(difference)定义为 $(A - B)_{ij} = A_{ij} - B_{ij}$。显然$A + B \in M_{m \times n}$，$A - B \in M_{m \times n}$。注意，若$A \in M_{m_1 \times n_1}$，$B \in M_{m_2 \times n_2}$**不满足**$m_1 = m_2$**,** $n_1 = n_2$，那么$A + B$与$A - B$**无法被定义**。

4. 对于$A_1 \in M_{m\times n}$，$A_2 \in M_{m\times n}$，...，$A_r \in M_{m\times n}$，$c_1, c_2 \ldots, c_r$为常数，定义$c_1A_1 + c_2A_2 + \ldots + c_rA_r$为$m \times n$-矩阵，满足

$$(c_1A_1 + c_2A_2 + \ldots + c_rA_r)_{ij} = c_1(A_1)_{ij} + c_2(A_2)_{ij} + \ldots + c_r(A_r)_{ij}.$$

矩阵$c_1A_1+c_2A_2+\ldots+c_rA_r$被称为$A_1, A_2, \ldots, A_r$关于系数$c_1, c_2, \ldots, c_r$的**线性组合(linear combination of** $A_1, A_2, \ldots, A_r$ **with coefficients** $c_1, c_2, \ldots, c_r$**)**。

5. 现在我们来定义矩阵的**乘法**运算法则。首先我们定义**行向量与列向量之间的乘积**:令$\boldsymbol{r} = \begin{bmatrix} r_1 & r_2 & \ldots & r_n \end{bmatrix}$为一个具有$n$个坐标的行向量，即$1 \times n$-矩阵；令$\boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$为一个具有$n$个坐标的列向量，即$n \times 1$-矩阵，那么它们的**乘积**$\boldsymbol{rc}$为一个常数，即$1 \times 1$-矩阵 $\boldsymbol{rc} = r_1c_1 + r_2c_2 + \ldots r_nc_n$。
利用这个规则，我们可以对更多的矩阵定义乘法运算。具体来说，若$A$为$m\times r$-矩阵，$B$为$r \times n$-矩阵，那么$A$与$B$的**乘积**$AB$为一个$m \times n$-矩阵，它满足：将$A$用行向量表示：$A =$
$$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix},$$
$\boldsymbol{r}_i = \begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{ir} \end{bmatrix}, i = 1, \ldots, m$; 将$B$用列向量表示：$B =$
$$\begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \ldots & \boldsymbol{c}_n \end{bmatrix},$$
$\boldsymbol{c}_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{rj} \end{bmatrix}, j = 1, \ldots, n$，那么乘积$AB$在第$i$行第$j$列上的项定义为
$$(AB)_{ij} = \boldsymbol{r}_i\boldsymbol{c}_j = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{ir}b_{rj},$$
即，$(AB)_{ij}$为矩阵$A$第$i$行与矩阵$B$第$j$列的**乘积**，对任何$i = 1, \ldots, m$，$j = 1, \ldots, n$。可将$AB$写为矩阵的形式：$AB =$
$$\begin{bmatrix} \boldsymbol{r}_1\boldsymbol{c}_1 & \boldsymbol{r}_1\boldsymbol{c}_2 & \ldots & \boldsymbol{r}_1\boldsymbol{c}_n \\ \boldsymbol{r}_2\boldsymbol{c}_1 & \boldsymbol{r}_2\boldsymbol{c}_2 & \ldots & \boldsymbol{r}_2\boldsymbol{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{r}_m\boldsymbol{c}_1 & \boldsymbol{r}_m\boldsymbol{c}_2 & \ldots & \boldsymbol{r}_m\boldsymbol{c}_n. \end{bmatrix}$$

- **例子**：假设$A=$
$$\begin{bmatrix}1&2&4\\2&6&0\end{bmatrix},$$
$B=$
$$\begin{bmatrix}4&1&4&3\\0&-1&3&1\\2&7&5&2\end{bmatrix}$$
那么由于$A$为$2\times 3$-矩阵，$B$为$3\times 4$-矩阵，根据矩阵乘法的定义，$AB$为一个$2\times 4$-矩阵，且有

$$(AB)_{11}=\underbrace{\begin{bmatrix}1&2&4\end{bmatrix}}_{A\text{的第1行}}\underbrace{\begin{bmatrix}4\\0\\2\end{bmatrix}}_{B\text{的第1列}}=1\times 4+2\times 0+4\times 2=12;$$

$$(AB)_{12}=\underbrace{\begin{bmatrix}1&2&4\end{bmatrix}}_{A\text{的第1行}}\underbrace{\begin{bmatrix}1\\-1\\7\end{bmatrix}}_{B\text{的第2列}}=1\times 1+2\times(-1)+4\times 7=27;$$

$$(AB)_{13}=\underbrace{\begin{bmatrix}1&2&4\end{bmatrix}}_{A\text{的第1行}}\underbrace{\begin{bmatrix}4\\3\\5\end{bmatrix}}_{B\text{的第3列}}=1\times 4+2\times 3+4\times 5=30;$$

$$(AB)_{13}=\underbrace{\begin{bmatrix}1&2&4\end{bmatrix}}_{A\text{的第1行}}\underbrace{\begin{bmatrix}3\\1\\2\end{bmatrix}}_{B\text{的第4列}}=1\times 3+2\times 1+4\times 2=13;$$

$$(AB)_{21}=\underbrace{\begin{bmatrix}2&6&0\end{bmatrix}}_{A\text{的第2行}}\underbrace{\begin{bmatrix}4\\0\\2\end{bmatrix}}_{B\text{的第1列}}=2\times 4+6\times 0+0\times 2=8;$$

$$(AB)_{22}=\underbrace{\begin{bmatrix}2&6&0\end{bmatrix}}_{A\text{的第2行}}\underbrace{\begin{bmatrix}1\\-1\\7\end{bmatrix}}_{B\text{的第2列}}=2\times 1+6\times(-1)+0\times 2=-4;$$

$$(AB)_{23} = \underbrace{\begin{bmatrix} 2 & 6 & 0 \end{bmatrix}}_{A\text{的第2行}} \underbrace{\begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}}_{B\text{的第3列}} = 2 \times 4 + 6 \times 3 + 0 \times 5 = 26;$$

$$(AB)_{24} = \underbrace{\begin{bmatrix} 2 & 6 & 0 \end{bmatrix}}_{A\text{的第2行}} \underbrace{\begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}}_{B\text{的第4列}} = 2 \times 3 + 6 \times 1 + 0 \times 2 = 12.$$

因此$AB =$

$$\begin{bmatrix} 12 & 27 & 30 & 13 \\ 8 & -4 & 26 & 12 \end{bmatrix}.$$

- 假设$A$为一个$m \times r$-矩阵，$\boldsymbol{c}$为一个$r \times 1$-矩阵，即一个含有$r$个坐标的列向量，那么根据矩阵乘法定义，$A\boldsymbol{c}$为一个$m \times 1$-矩阵，即一个含有$m$个坐标的列向量。因此，若$B$为一个$r \times n$-矩阵，将$B$用列向量表示为

  $$\begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \ldots & \boldsymbol{c}_n \end{bmatrix},$$

  每个$\boldsymbol{c}_j$为$r \times 1$-矩阵， $j = 1, \ldots, n$，那么容易验证

  $$AB = A \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \ldots & \boldsymbol{c}_n \end{bmatrix} = \begin{bmatrix} A\boldsymbol{c}_1 & A\boldsymbol{c}_2 & \ldots & A\boldsymbol{c}_n \end{bmatrix},$$

  即$A$**与**$B$**的乘积**$AB$**的第**$j$**列为矩阵**$A$**与**$B$**的第**$j$**列**$\boldsymbol{c}_j$ **(作为**$r \times 1$**-矩阵)的乘积**$A\boldsymbol{c}_j$，也即$AB$的**列向量表示为：**

  $$\begin{bmatrix} A\boldsymbol{c}_1 & A\boldsymbol{c}_2 & \ldots & A\boldsymbol{c}_n \end{bmatrix}.$$

- 假设若$B$为一个$r \times n$-矩阵，$\boldsymbol{r}$为一个$1 \times r$-矩阵，即一个含有$r$个坐标的行向量，那么根据矩阵乘法定义， $\boldsymbol{r}B$为一个$1 \times n$-矩阵，即一个含有$n$个坐标的行向量。因此，若$A$为一个$m \times r$-矩阵，将$A$用行向量表示为

  $$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix},$$

  每个$\boldsymbol{r}_i$为$1 \times r$-矩阵，$i = 1, \ldots, m$，那么容易验证

  $$AB = \begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix} B = \begin{bmatrix} \boldsymbol{r}_1 B \\ \boldsymbol{r}_2 B \\ \vdots \\ \boldsymbol{r}_m B, \end{bmatrix}$$

即$A$与$B$的乘积$AB$的第$i$行为矩阵$A$的第$i$行$\boldsymbol{r}_i$ (作为$1\times r$-矩阵)与矩阵$B$的乘积$\boldsymbol{r}_iB$，也即$AB$的行向量表示为

$$\begin{bmatrix} \boldsymbol{r}_1B \\ \boldsymbol{r}_2B \\ \vdots \\ \boldsymbol{r}_mB \end{bmatrix}.$$

- 对任何$m \times n$-矩阵$A$与含有$n$个坐标的向量$\boldsymbol{s} = (s_1, \ldots, s_n)$，将$\boldsymbol{s}$视为列向量，即$n \times 1$-矩阵，那么由矩阵乘法定义可知$A\boldsymbol{s}$是一个$m \times 1$-矩阵，即一个包含$m$个坐标的列向量：

$$A\boldsymbol{s} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} a_{11}s_1 + a_{12}s_2 + \ldots + a_{1n}s_n \\ a_{21}s_1 + a_{22}s_2 + \ldots + a_{2n}s_n \\ \vdots \\ a_{m1}s_1 + a_{m2}s_2 + \ldots + a_{mn}s_n \end{bmatrix}.$$

因此，$\boldsymbol{s} = (s_1, \ldots, s_n)$是线性方程组

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

的一个解**当且仅当**$A\boldsymbol{s} = \boldsymbol{b}$，这里$A$为方程组的系数矩阵，$\boldsymbol{s}$视为列向量，$\boldsymbol{b} =$

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

也视为列向量。

- 将$m \times n$-矩阵$A$用列向量表示：

$$A = \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \ldots & \boldsymbol{c}_n \end{bmatrix},$$

那么显然有

$$A\boldsymbol{s} = \begin{bmatrix} a_{11}s_1 + a_{12}s_2 + \ldots + a_{1n}s_n \\ a_{21}s_1 + a_{22}s_2 + \ldots + a_{2n}s_n \\ \vdots \\ a_{m1}s_1 + a_{m2}s_2 + \ldots + a_{mn}s_n \end{bmatrix} = s_1\boldsymbol{c}_1 + s_2\boldsymbol{c}_2 + \ldots + s_n\boldsymbol{c}_n.$$

即$A\boldsymbol{s}$是$A$的列向量$\boldsymbol{c}_1,\ldots,\boldsymbol{c}_n$(视为$m\times 1$-矩阵)关于系数$s_1,\ldots,s_n$的**线性组合**。

6. 若$A$为一个$m\times n$-矩阵，那么它的**转置**(transpose)，记为$A^\top$，是一个$n\times m$-矩阵，满足
$$(A^\top)_{ij}=(A)_{ji}$$
对所有$i=1,\ldots,n$，$j=1,\ldots,m$。注意，若将$A$用**列向量**表示：
$$A=\begin{bmatrix}\boldsymbol{c}_1 & \boldsymbol{c}_2 & \ldots & \boldsymbol{c}_n\end{bmatrix},$$
那么$A^\top$的**行向量**表示为：
$$A^\top=\begin{bmatrix}\boldsymbol{c}_1^\top\\ \boldsymbol{c}_2^\top\\ \vdots\\ \boldsymbol{c}_n^\top\end{bmatrix},$$
这里$\boldsymbol{c}_i=\begin{bmatrix}a_{1i}\\ a_{2i}\\ \vdots\\ a_{mi}\end{bmatrix}$，$\boldsymbol{c}_i^\top=\begin{bmatrix}a_{1i} & a_{2i} & \ldots & a_{mi}\end{bmatrix}$。例如矩阵
$$B=\begin{bmatrix}2 & 3\\ 1 & 4\\ 5 & 6\end{bmatrix}$$
的转置为
$$B^\top=\begin{bmatrix}2 & 1 & 5\\ 3 & 4 & 6\end{bmatrix}.$$

7. 若$A$为$n\times n$-**方阵**，那么它对角线上所有项的和被称为$A$的**迹**(trace)，记为
$$\mathrm{tr}(A)=a_{11}+a_{22}+\ldots+a_{nn}.$$
例如，矩阵
$$A=\begin{bmatrix}-1 & 2 & 7 & 0\\ 3 & 5 & -8 & 4\\ 1 & 2 & 7 & -3\\ 4 & -2 & 1 & 0\end{bmatrix}$$
的迹为$\mathrm{tr}(A)=-1+5+7+0=11$。

8. 矩阵可以被划分为若干个子矩阵(submatrix)，比如
$$A=\left(\begin{array}{ccc|c}a_{11} & a_{12} & a_{13} & a_{14}\\ a_{21} & a_{22} & a_{23} & a_{24}\\ \hline a_{31} & a_{32} & a_{33} & a_{34}\end{array}\right)=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},$$

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, A_{12} = \begin{bmatrix} a_{14} \\ a_{24} \end{bmatrix}, A_{21} = \begin{bmatrix} a_{31} & a_{32} & a_{33} \end{bmatrix}, A_{22} = \begin{bmatrix} a_{34} \end{bmatrix}$$

就是$A$的子矩阵。像$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$这样的表示也被叫做**分块矩阵**(partitioned matrix)。显然矩阵的行向量表示与列向量表示都是将$A$表示为分块矩阵。

以上所定义的关于矩阵的加，减，乘，转置的运算所遵循的规则请阅读英文教材第1.4节，尤其是**Theorem 1.4.1**，**Theorem 1.4.2**， **Theorem 1.4.8**。这里我们特别注意到矩阵的乘法运算法则与我们高中所熟悉的实数的乘法运算法则在有些地方是有重要区别的：

- 我们知道，若$a$与$b$为两个常数，那么必然有$ab = ba$成立，即实数的乘法是满足交换律的(the commutative law)。然而，**乘法的交换律**对于一般的矩阵$A$与$B$**并不一定成立**，这是因为

  1. 有可能$AB$的乘积可以定义，但$BA$的乘积无法被定义。比如$A$为$2\times 3$-矩阵，$B$为$3 \times 4$-矩阵，此时$AB$可以被定义且为一个$2 \times 4$-矩阵，但$BA$无法被定义。
  2. 有可能$AB$与$BA$都可以被定义，但它们的尺寸不同。比如$A$为$2 \times 3$-矩阵，$B$为$3 \times 2$-矩阵.此时$AB$为$2 \times 2$-矩阵，$BA$为$3 \times 3$-矩阵。
  3. 有可能$AB$与$BA$都可以被定义并且它们的尺寸相同，但仍然有$AB \neq BA$。比如
  $$A = \begin{bmatrix} -1 & 0 \\ 2 & 3 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix},$$
  那么$AB = \begin{bmatrix} -1 & -2 \\ 11 & 4 \end{bmatrix}$， $BA = \begin{bmatrix} 3 & 6 \\ -3 & 0 \end{bmatrix}$，显然$AB \neq BA$。

- 若一个矩阵所有项均为0，那么我们称其为一个零矩阵(zero matrix)。这里我们把零矩阵也记作$0$（实际上，若零矩阵为$m \times n$-矩阵，那么我们最好将其记作$0_{m\times n}$。但在不影响理解的情况下我们可以将其简单写作0）。在处理加法运算时，我们可以认为零矩阵扮演实数中的0的角色，比如对于任何矩阵$A$，$A + 0 = 0 + A = A$。然而在乘法运算中，我们需要小心以下牵扯到零矩阵的情况：

  1. 在实数乘法运算中($a, b, c$均为常数)，若$ab = ac$且$a \neq 0$，那么必有$b = c$，即实数乘法满足消去律(the cancellation law)。该消去律无法照搬到矩阵乘法情况：对于一般的矩阵$A, B, C$，即使$AB = AC$且$A \neq 0$为一个非零矩阵，那么也不一定会有$B = C$。比如
  $$A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}, C = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix},$$

此时满足$A\neq 0$，$B\neq C$，$AB=AC=\begin{bmatrix}3&4\\6&8\end{bmatrix}$，但$B\neq C$。

2. 对于常数$a,b$，我们知道若$ab=0$，那么有$a=0$或$b=0$。该法则对矩阵乘法不成立：

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

显然$A\neq 0$，$B\neq 0$，且$AB=0$。

因此**在矩阵乘法当中，零矩阵不能视为实数中的**0。

由矩阵乘法的定义可知，对于$A,B\in M_{n\times n}$，我们有$AB\in M_{n\times n}$，即两个尺寸相同的方阵之间的乘积还是同尺寸的方阵。而$n$阶单位矩阵$I_n=\begin{bmatrix}1&0&\dots&0\\0&1&\dots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\dots&1\end{bmatrix}$ (除了对角线上的所有项为1，其余不在对角线上的项均为0) 则在$n$阶方阵乘法运算中扮演实数乘法中1的角色：

$$AI_n=I_nA=A$$

对任何$n$阶方阵$A$成立。

以下定理非常重要，我们会在后续的学习中持续使用该定理。

**Theorem 1.7.** *令$R$为一个$n$阶简化阶梯型。那么只可能有如下两种情况：*

1. *$R$包含至少一个为0的行；*
2. *$R=I_n$为$n$阶单位矩阵。*

*证明.* 我们只需证明若一个$n$阶简化阶梯型$R$所有行均非0，那么它一定是一个单位矩阵。在这种情况下，由简化阶梯型的定义可知，$R$矩阵里的每一行都包含一个主1。现在我们只需要证明第$i$行的主1位于第$i$列上对任何$i=1,\dots,n$成立。(若该性质成立，那么由简化阶梯型定义的第四个条件可知$R$必为单位矩阵)。

现在我们用**反证法**与**数学归纳法**证明该性质。假设$R$的第一行里的主1不在第一列上，那么第一行的主1所在的第$j_1$列必然满足$j\geq 2$。那么，由阶梯型定义的第三个条件可知，第二行主1的位置必须在第一行主1的右边，即，第二行主1所在的第$j_2$列必然满足$j_2>j_1\geq 2$，特别地，我们必然有$j_2\geq 3$。以此类推，我们可推出第$n$行的主1所在的第$j_n$列必然满足$j_n\geq n+1$。然而，$R$只包含$n$列，因此第$n$行的主1所在的第$j_n$列也必须满足$j_n\leq n$。这意味着我们现在得到了$n+1\leq j\leq n$，矛盾！因此第一行的主1必须位于第一列上。

现在假设我们对某个$i<n$证明了第$i$行的主1位于第$i$列上。再次利用阶梯型定义

的第三个条件可知，第$i+1$行主1的位置必须在第$i$行主1的右边。那么由归纳假设，可知第$i+1$行主1所在的第$j_{i+1}$列必须满足$j_{i+1} \geq i+1$。再次利用反证法，假设$j_{i+1} \neq i+1$，即$j_{i+1} \geq i+2$，依照证明第一行主1在第一列上的方法，可推得此时第$i+2$行主1所在的第$j_{i+2}$列必须满足$j_{i+2} \geq i+3$，...，第$n$行的主1所在的第$j_n$列必须满足$j_n \geq n+1$，这意味着我们现在又一次得到了$n+1 \leq j \leq n$，矛盾！因此第$i$+1行的主1必须位于第$i$+1列上。因此由归纳法可知该性质对所有$i = 1, \ldots, n$都成立。 □

最后我们注意，对于任何$m \times n$-矩阵$A$，以下等式成立：

$$AI_n = A, \quad I_m A = A$$

这里$I_n$为$n$阶单位矩阵，，$I_m$为$m$阶单位矩阵。

### 1.4. 矩阵的逆， Inverses of matrices

由上一节的讨论我们知道，$n$阶单位矩阵在$n$阶方阵的乘法运算中扮演了与实数乘法中1一样的角色，即

$$AI_n = I_n A = A$$

对任何$n$阶方阵$A$成立。那么我们很自然地会问，**一个$n$阶方阵$A$的"倒数"是什么？** 对于一个实数$a$我们知道，只要$a \neq 0$那么$a$的倒数$\frac{1}{a}$一定存在，即满足$a\frac{1}{a} = \frac{1}{a}a = 1$。基于上一节的相关讨论我们已经知道在矩阵乘法中零矩阵并不能直接视为实数乘法中的0，因此在这里我们也会怀疑并非任何非零矩阵都拥有"倒数"。事实也确实如此。首先我们先给出矩阵"倒数"的定义。

**Definition 1.8.** *设$A$为一个$n$阶**方阵**。我们说$A$是**可逆的**(invertible)或者非奇异的(nonsingular)当且仅当存在一个$n$阶方阵$B$使得*

$$AB = BA = I_n.$$

*若存在这样的$n$阶方阵$B$，则称$B$为$A$的**逆矩阵**或简称为$A$的**逆**(inverse)，并记为$B = A^{-1}$。*
*若这样的$B$**不存在**，那么则称$A$为**奇异的**(singular)。*

显然从以上定义可以看出若$B$满足

$$AB = BA = I_n,$$

那么$B = A^{-1}$就是$A$关于矩阵乘法的"倒数"。

**Remark 1.9.**
- 可逆性这个概念只对**方阵**有效。也可以说任何$m\times n$-矩阵在$m\neq n$时都是奇异的（不可逆的）。
- 一个非零的$n$阶方阵不一定可逆，这与实数的情况不同。比如2阶方阵
$$A=\begin{bmatrix}0&0\\1&2\end{bmatrix}$$
不是零矩阵，但它不可逆，因为对任何2阶矩阵$B$，我们有
$$AB=\begin{bmatrix}0&0\\1&2\end{bmatrix}\begin{bmatrix}b_{11}&b_{12}\\b_{21}&b_{22}\end{bmatrix}=\begin{bmatrix}0&0\\b_{11}+2b_{21}&b_{12}+2b_{22}\end{bmatrix}.$$
因此对任何2阶方阵$B$，$AB$都不可能等于$I_2$。
- 显然若$A$的逆矩阵为$B=A^{-1}$，那么由定义可知矩阵$B$可逆且它的逆为$A$，即$B^{-1}=(A^{-1})^{-1}=A$。
- 在*1.5*节我们会学习如何判断一个方阵是否可逆以及如何求出它的逆矩阵*(*在可逆的情况下*)*。

首先我们容易证明矩阵逆的唯一性：

**Theorem 1.10** (英文教材Theorem 1.4.4)**.** 假设$B$和$C$都是方阵$A$的逆，即$BA=AB=I_n$，$CA=AC=I_n$，那么必然有$B=C$。

*证明.* 由于$B$是$A$的逆，我们有$AB=BA=I_n$。对该等式两边同时**右乘**矩阵$C$可得
$$(BA)C=I_nC=C.$$
另一方面，由矩阵乘法的结合律以及$AC=I_n$，我们还有
$$(BA)C=B(AC)=BI_n=B.$$
因此$B=C$。 □

根据逆矩阵的定义，若$B$为$A$的逆，那么需要满足$AB=BA=I_n$。事实上若要验证$B$为$A$的逆，我们只需要证明$AB=I_n$或者$BA=I_n$成立即可。

**Theorem 1.11.** 对一个$n$阶方阵$A$，若存在一个矩阵$B$使得$BA=I_n$成立，那么矩阵$A$可逆且$A^{-1}=B$，即若$BA=I_n$成立那么必有$AB=I_n$。类似的，若存在一个矩阵$B$使得$AB=I_n$成立，那么矩阵$A$可逆且$A^{-1}=B$，即若$AB=I_n$成立那么必有$BA=I_n$。

我们目前还没有足够的工具对以上定理给出一个简单的证明。所以现在大家只需要记住这个结论，具体证明在后续学习中会给出。以下定理则很容易被证明。

**Theorem 1.12** (英文教材Theorem 1.4.6)**.** *若$A$与$B$均为可逆$n$阶方阵，那么$AB$也可逆，且有$(AB)^{-1} = B^{-1}A^{-1}$。*

证明. 利用矩阵乘法的结合律，易得

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.$$

因此由Theorem 1.11可得$AB$可逆，且$(AB)^{-1} = B^{-1}A^{-1}$。事实上我们也可以直接利用矩阵乘法的结合律验证$(B^{-1}A^{-1})(AB) = I_n$。 □

下面我们介绍矩阵的幂。

**Definition 1.13.** *设$A$为$n$阶方阵。我们定义矩阵$A$的幂(power of matrix)为*

$$A^0 = I_n, A^k = \underbrace{AA\ldots A}_{k\text{个}}, \quad k = 1, 2, \ldots$$

*若$A$可逆，那么它的负幂定义为*

$$A^{-k} = (A^{-1})^k = \underbrace{A^{-1}A^{-1}\ldots A^{-1}}_{k\text{个}}, \quad k = 1, 2, \ldots$$

容易验证以下等式成立：对任何自然数$r, s$，有

$$A^{r+s} = A^rA^s, \quad (A^r)^s = A^{rs},$$

若$A$可逆(因此$A$的负幂也存在)，那么以上等式对任何整数$r, s$成立。特别地我们有以下定理。

**Theorem 1.14** (英文教材Theorem 1.4.7)**.** *设$A$为可逆$n$阶方阵，$k$为任意自然数。那么*

1. *$A^{-1}$可逆且$(A^{-1})^{-1} = A$，*
2. *$A^k$可逆且$(A^k)^{-1} = A^{-k} = (A^{-1})^k$，*
3. *对任何非0的常数$c$，$cA$可逆且$(cA)^{-1} = c^{-1}A^{-1}$。*

该定理的证明较为简单，请参考英文教材。

**Remark 1.15.** *对于实数$a, b$我们有二项式定理:*

$$(a+b)^k = a^kb^0 + \binom{k}{1}a^{k-1}b + \binom{k}{2}a^{k-2}b^2 + \ldots + \binom{k}{k-1}ab^{k-1} + a^0b^k,$$

*但该公式对矩阵的情况不适用。比如存在矩阵$A, B$使得*

$$(A+B)^2 \neq A^2 + 2AB + B^2.$$

这是因为$(A+B)^2 = A^2 + AB + BA + B^2$，因此若$AB \neq BA$，那么就有

$$A^2 + AB + BA + B^2 \neq A^2 + 2AB + B^2.$$

换句话说，**矩阵乘法的非交换性导致了二项式定理对矩阵的失效。**

利用矩阵的幂我们可以定义矩阵多项式(matrix polynomial)。

**Definition 1.16.** 设$A$为一个$n$阶矩阵，令$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m$为一个多项式，$a_0, a_1, \ldots, a_m$为常数。我们定义矩阵多项式$p(A)$为

$$p(A) = a_0 I_n + a_1 A + a_2 A^2 + \ldots + a_m A^m.$$

显然根据矩阵的运算法则可知$p(A)$为一个$n$阶矩阵。

**例子**：设$A = \begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix}$，$p(x) = x^2 - 2x - 3$，那么

$$\begin{aligned} p(A) &= A^2 - 2A - 3I_2 \\ &= \begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix}^2 - 2\begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix} - 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 4 \\ 0 & 9 \end{bmatrix} - \begin{bmatrix} -2 & 4 \\ 0 & 6 \end{bmatrix} - \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

以下定理容易被证明：

**Theorem 1.17** (英文教材Theorem 1.4.9)**.** 设$A$为$n$阶方阵，若$A$可逆，那么它的转置$A^\top$也可逆，且有

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

*证明.* 由于$(A^{-1}A)^\top = A^\top (A^{-1})^\top$ (这里利用运算法则$(AB)^\top = B^\top A^\top$，见英文教材Theorem 1.4.8(e)或者第二次作业Problem A)，以及显然的事实$I_n^\top = I_n$，我们得到

$$A^\top (A^{-1})^\top = (A^{-1}A)^\top = I_n^\top = I_n,$$

类似可证

$$(A^{-1})^\top A^\top = (AA^{-1})^\top = I_n^\top = I_n.$$

这意味着矩阵$B = (A^{-1})^\top$满足$AB = BA = I_n$。因此由定义Definition 1.8可知$A^\top$可逆且有

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

$\square$

在本节的最后，我们介绍一种判断$2\times 2$-方阵是否可逆且求出其逆矩阵的方法。

**Theorem 1.18** (英文教材Theorem 1.4.5). *$2\times 2$-方阵$A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$可逆当且仅当$ad-bc\neq 0$。若该条件满足，那么$A^{-1}=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.*

该定理的严格证明将在第二章给出。目前大家只需要记住该定理的内容以及能够使用即可。

**例子：** 考虑$2\times 2$-方阵阵$A=\begin{bmatrix} 6 & 1 \\ 5 & 2 \end{bmatrix}$。计算可得

$$ad-bc=6\times 2-5\times 1=7\neq 0$$

因此由Theorem1.18可知该矩阵可逆，且有

$$A^{-1}=B=\frac{1}{7}\begin{bmatrix} 2 & -1 \\ -5 & 6 \end{bmatrix}=\begin{bmatrix} \frac{2}{7} & \frac{-1}{7} \\ \frac{-5}{7} & \frac{6}{7} \end{bmatrix}.$$

经简单计算可快速验证$AB=BA=I_2$。

在1.5节我们将学习如何判断一般的$n$阶方阵是否可逆，并在可逆的情况下求出它的逆。

### 1.5. 利用初等矩阵求矩阵的逆, Elementary matrices and a method for finding $A^{-1}$

在本节我们将学习到如何判断一般的$n$阶方阵是否可逆，并在可逆的情况下求出它的逆。要做到这一点我们需要首先回顾初等行变换这个概念。根据Definition 1.4，我们有如下三种初等行变换：(设$A$为$m\times n$-矩阵)

1. 将$A$的第$i$行 ($i=1,\dots,m$) 乘以某个**非零**的常数$c$，其他行保持不变 (可记为: $i$-th row $\times c$ );
2. 将$A$的第$i$行 ($i=1,\dots,m$)乘以某个常数$c$后加到第$k$行上去 ($k=1,\dots,m$且$k\neq i$)，且除了第$k$行之外其他的行保持不变 (可记为: $i$-th row $\times c+k$-th row );
3. 将$A$的第$i$行 ($i=1,\dots,m$)与第$k$行 ($k=1,\dots,m$且$k\neq i$) 互换 (可记为: $i$-th row $\leftrightarrow k$-th row )，其他行保持不变。

以下这个与初等行变换相关的概念非常重要(期中考试重要考点之一)：

**Definition 1.19** (英文教材52页Definition 2). *若一个$m$阶方阵$E$是通过**对**$m$**阶单位矩阵**$I_m$进行一次**初等行变换**得到的，那么$E$被称为一个初等矩阵(elementary matrix)。*

例子：

- $B=\begin{bmatrix}1&0\\0&-3\end{bmatrix}$是一个初等矩阵，因为它显然是通过对2阶单位矩阵的第二行乘以$-3$得到的。
- $B=\begin{bmatrix}1&0&0&0\\0&0&0&1\\0&0&1&0\\0&1&0&0\end{bmatrix}$是一个初等矩阵，因为它显然是通过对4阶单位矩阵交换第2行与第4行得到的。
- $B=\begin{bmatrix}1&0&3\\0&1&0\\0&0&1\end{bmatrix}$是一个初等矩阵，因为它显然是通过对3阶单位矩阵的第三行乘以3再加到第一行上得到的。
- $B=\begin{bmatrix}1&0&0\\0&0&0\\0&0&1\end{bmatrix}$不是一个初等矩阵，因为无论对3阶单位矩阵进行哪种初等行变换都不会得到这种形状的矩阵。
- $B=\begin{bmatrix}1&0&3\\0&0&1\\0&1&0\end{bmatrix}$不是一个初等矩阵，因为它是通过对3阶单位矩阵进行了两次初等行变化得到的：先对3阶单位矩阵的第三行乘以3再加到第一行上，再交换第三行与第二行。

以下定理告诉我们，任何矩阵的初等行变换(注意，这里指对矩阵做一次初等行变换)都可以通过对该矩阵左乘一个初等矩阵得到。

**Theorem 1.20** (英文教材Theorem 1.5.1). *设$A$为$m\times n$-矩阵。若$E$是一个$m$阶初等矩阵，即$E$通过对$m$阶单位矩阵进行一次初等行变换得到，那么$EA$等于对矩阵$A$做与$E$相同的初等行变换所得到的矩阵。具体来讲，*

- *若$E$是通过将$I_m$的第$i$行 $(i=1,\dots,m)$ 乘以某个非零的常数$c$得到，那么$EA$等于对$A$的第$i$行乘以这个非零的常数$c$所得到的的矩阵。*
- *若$E$是通过将$I_m$的第$i$行 $(i=1,\dots,m)$乘以某个常数$c$后加到第$k$行上去 $(k=1,\dots,m$且$k\neq i)$得到的，那么$EA$等于对$A$第$i$行乘以常数$c$后加到第$k$行上去所得到的矩阵。*
- *若$E$是通过将$I_m$的第$i$行 $(i=1,\dots,m)$与第$k$行 $(k=1,\dots,m$且$k\neq i)$ 互换得到的，那么$EA$等于对$A$第$i$行与第$k$行互换所得到的矩阵。*

*证明.* **注意：该证明不要求一定掌握(但建议理解证明过程)，但务必记住该定理的**

**内容(期中考试考点之一)。**

我们首先把$m$阶单位矩阵$I_m$用行向量来表示：

$$I_m = \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_m \end{bmatrix}.$$

显然，由单位矩阵的定义可知，对任何$i = 1, \ldots, m$，$\boldsymbol{e}_i = \begin{bmatrix} 0 & \ldots & 0 & \underbrace{1}_{\text{第i列}} & 0 & \ldots & 0 \end{bmatrix}$，即$\boldsymbol{e}_i$除了第$i$列上的项为1之外，其他项均为0。

对于$m \times n$-矩阵$A$，将其表示为列向量形式$A = \begin{bmatrix} \boldsymbol{c}_1 & \ldots & \boldsymbol{c}_n \end{bmatrix}$，由矩阵乘法的列向量表示(见讲义第19页)可知，

$$\boldsymbol{e}_i A = \begin{bmatrix} \boldsymbol{e}_i \boldsymbol{c}_1 & \boldsymbol{e}_i \boldsymbol{c}_2 & \ldots \boldsymbol{e}_i \boldsymbol{c}_n \end{bmatrix}.$$

由$\boldsymbol{e}_i$的特殊形式，容易计算出

$$\begin{aligned} \boldsymbol{e}_i \boldsymbol{c}_1 &= 0 \times a_{11} + \ldots + 1 \times a_{i1} + \ldots 0 \times a_{m1} = a_{i1} \\ \boldsymbol{e}_i \boldsymbol{c}_2 &= 0 \times a_{12} + \ldots + 1 \times a_{i2} + \ldots 0 \times a_{m2} = a_{i2} \\ &\vdots \\ \boldsymbol{e}_i \boldsymbol{c}_n &= 0 \times a_{1n} + \ldots + 1 \times a_{in} + \ldots 0 \times a_{mn} = a_{in}. \end{aligned}$$

因此，我们得到

$$\boldsymbol{e}_i A = \begin{bmatrix} \boldsymbol{e}_i \boldsymbol{c}_1 & \boldsymbol{e}_i \boldsymbol{c}_2 & \ldots \boldsymbol{e}_i \boldsymbol{c}_n \end{bmatrix} = \begin{bmatrix} a_{i1} & a_{i2} & \ldots a_{in} \end{bmatrix} = \boldsymbol{r}_i = \text{A的第i行}. \tag{1.1}$$

现在我们依次对三种不同的初等行变换进行验证。

- 若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$ 乘以某个非零的常数$c$得到，那么显然$E$的行向量表示为

$$E = \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ c\boldsymbol{e}_i \\ \vdots \\ \boldsymbol{e}_m \end{bmatrix},$$

  由讲义第19页-20页关于矩阵乘积的行向量表示，我们知道$EA$的行向量表示为

$$EA = \begin{bmatrix} \boldsymbol{e}_1 A \\ \vdots \\ c\boldsymbol{e}_i A \\ \vdots \\ \boldsymbol{e}_m A \end{bmatrix},$$

由等式(1.1)可知，

$$EA = \begin{bmatrix} \boldsymbol{e}_1 A \\ \vdots \\ c\boldsymbol{e}_i A \\ \vdots \\ \boldsymbol{e}_m A \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \\ c\boldsymbol{r}_i \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix},$$

这里$\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m$为$A$的行向量。因此我们得到$EA$等于对$A$的第$i$行乘以这个非零的常数$c$所得到的的矩阵。

- 若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$乘以某个常数$c$后加到第$k$行上去 $(k = 1, \ldots, m$且$k \neq i)$得到的，那么显然$E$的行向量表示为

$$E = \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ c\boldsymbol{e}_i + \boldsymbol{e}_k(\text{第k行}) \\ \vdots \\ \boldsymbol{e}_m \end{bmatrix},$$

像第一种情况一样，$EA$的行向量表示为

$$EA = \begin{bmatrix} \boldsymbol{e}_1 A \\ \vdots \\ (c\boldsymbol{e}_i + \boldsymbol{e}_k)A \\ \vdots \\ \boldsymbol{e}_m A \end{bmatrix},$$

由等式(1.1)可知，

$$EA = \begin{bmatrix} \boldsymbol{e}_1 A \\ \vdots \\ (c\boldsymbol{e}_i + \boldsymbol{e}_k)A \\ \vdots \\ \boldsymbol{e}_m A \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_1 \\ \vdots \\ c\boldsymbol{r}_i + \boldsymbol{r}_k \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix},$$

因此$EA$等于对$A$第$i$行乘以常数$c$后加到第$k$行上去所得到的矩阵。

- 与前两种情况类似，若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$与第$k$行 $(k =$

$1,\ldots,m$且$k\neq i$) 互换得到的，那么显然$E$的行向量表示为

$$E=\begin{bmatrix}\boldsymbol{e}_1\\ \vdots\\ \boldsymbol{e}_k(\text{第i行})\\ \vdots\\ \boldsymbol{e}_i(\text{第k行})\\ \boldsymbol{e}_m\end{bmatrix},$$

那么同样的推导可以证明$EA$等于对$A$第$i$行与第$k$行互换所得到的矩阵。具体细节留作练习。

□

**例子**：矩阵$A=\begin{bmatrix}1&0&2&3\\2&-1&3&6\\1&4&4&0\end{bmatrix}$，初等矩阵$E=\begin{bmatrix}1&0&0\\0&1&0\\3&0&1\end{bmatrix}$是通过对$I_3$第1行乘以常数3后加到第3行上得到的，那么通过计算可得

$$EA=\begin{bmatrix}1&0&2&3\\2&-1&3&6\\4&4&10&9\end{bmatrix},$$

显然等于对矩阵$A$的第1行乘以常数3后加到第3行后得到的矩阵。

由初等行变换的定义可知，**每个初等矩阵$E$都可以通过一次初等行变换变成单位矩阵**。具体来讲，我们有：

1. 若$E$是通过将$I_m$的第$i$行 ($i=1,\ldots,m$) 乘以某个非零的常数$c$得到，那么将$E$的第$i$行乘以$1/c$可以得到单位矩阵$I_m$。
2. 若$E$是通过将$I_m$的第$i$行 ($i=1,\ldots,m$)乘以某个常数$c$后加到第$k$行上去 ($k=1,\ldots,m$且$k\neq i$)得到的，那么将$E$的第$i$行乘以$-c$加到第$k$行可以得到单位矩阵$I_m$。
3. 若$E$是通过将$I_m$的第$i$行 ($i=1,\ldots,m$)与第$k$行 ($k=1,\ldots,m$且$k\neq i$) 互换得到的，那么将$E$的第$i$行与第$k$行互换可以得到单位矩阵$I_m$。

以上观察暗示我们**任何初等矩阵都是可逆的，且它的逆也是一个初等矩阵**。

**Theorem 1.21.** *任何初等矩阵都是可逆的，且它的逆也是一个初等矩阵。具体来讲，*

- *若$E$是通过将$I_m$的第$i$行 ($i=1,\ldots,m$) 乘以某个非零的常数$c$得到，那么$E^{-1}$等于对$I_m$的第$i$行乘以$1/c$所得到的的矩阵。*

- 若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$乘以某个常数$c$后加到第$k$行上去 $(k = 1, \ldots, m$且$k \neq i)$得到的，那么$E^{-1}$等于对$I_m$第$i$行乘以常数$-c$后加到第$k$行上去所得到的矩阵。
- 若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$与第$k$行 $(k = 1, \ldots, m$且$k \neq i)$ 互换得到的，那么$E^{-1}$等于对$I_m$第$i$行与第$k$行互换所得到的矩阵。

证明.
- 若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$ 乘以某个非零的常数$c$得到，令$E_0$为通过对$I_m$的第$i$行乘以$1/c$所得到的的矩阵。由初等矩阵的定义可得$E_0$为初等矩阵。再利用Theorem 1.20我们知道$E_0E$等于对$E$的第$i$行乘以$1/c$所得到的的矩阵。基于我们之前的讨论可知，对$E$的第$i$行乘以$1/c$所得到的的矩阵为单位矩阵，即$E_0E = I_m$。由Theorem 1.11可断定$E$可逆且$E_0 = E^{-1}$。
- 若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$乘以某个常数$c$后加到第$k$行上去 $(k = 1, \ldots, m$且$k \neq i)$得到的，令$E_0$为对$I_m$第$i$行乘以常数$-c$后加到第$k$行上去所得到的矩阵。同样的推理可得到$E_0E = I_m$，即这样的$E$可逆且$E_0 = E^{-1}$。
- 若$E$是通过将$I_m$的第$i$行 $(i = 1, \ldots, m)$与第$k$行 $(k = 1, \ldots, m$且$k \neq i)$ 互换得到的，令$E_0$为对$I_m$第$i$行与第$k$行互换所得到的矩阵。同样的推理可得到$E_0E = I_m$，即这样的$E$可逆且$E_0 = E^{-1}$。

□

以下定理给出了判断一个方阵$A$是否可逆的标准，并间接给出了计算$A^{-1}$的一个算法。因此该定理也是期中考试考点之一。

**Theorem 1.22** (英文教材Theorem 1.5.3, Equivalent Theorem)**.** *设$A$为$n$阶方阵。那么以下说法等价(equivalent):*

1. *$A$可逆。*
2. *齐次线性方程组$A\boldsymbol{x} = \boldsymbol{0}$只有一个平凡解(trivial solution)。注意这里$\boldsymbol{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$为$n \times 1$的零列向量，$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$为$n \times 1$的未知数列向量，那么由矩阵乘法定义可知$A\boldsymbol{x} = \boldsymbol{0}$是在描述齐次线性方程组*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= 0. \end{aligned}$$

3. $A$的简化阶梯型为单位矩阵$I_n$。

4. $A$可以被表示为一系列初等矩阵的乘积。

证明. 我们证明1. $\Rightarrow$ 2. $\Rightarrow$ 3. $\Rightarrow$ 4. $\Rightarrow$ 1.。

1. $\Rightarrow$ 2.: 假设$A$可逆。令$\boldsymbol{s}$为齐次线性方程组$A\boldsymbol{x}=\mathbf{0}$的一个解。由矩阵乘法定义(见讲义第20页的讨论)可知$\boldsymbol{s}$满足

$$A\boldsymbol{s}=\mathbf{0}.$$

由于$A$可逆，$A^{-1}$存在，因此对以上等式两边同时左乘$A^{-1}$可得

$$A^{-1}(A\boldsymbol{s})=A^{-1}\mathbf{0}=\mathbf{0},$$

而由矩阵乘法的结合律可知$A^{-1}(A\boldsymbol{s})=(A^{-1}A)\boldsymbol{s}=I_n\boldsymbol{s}=\boldsymbol{s}$。因此我们实际上已经得到$\boldsymbol{s}=\mathbf{0}$。这意味着齐次线性方程组$A\boldsymbol{x}=\mathbf{0}$的任何一个解都等于零向量$\mathbf{0}$，即，该方程组仅有一个平凡解。

2. $\Rightarrow$ 3.: 现在假设齐次线性方程组$A\boldsymbol{x}=\mathbf{0}$仅有一个平凡解。令$R$标记$A$对应的简化阶梯型。那么$R$必然不可能包含一个0行。这是因为$A$和$R$为$n\times n$-矩阵，因此若$R$包含有一个0行意味着$R$最多只能有$n-1$个非零行，也就是说此时$R$最多只能有$n-1$个主1，也即方程组$A\boldsymbol{x}=\mathbf{0}$最多只能有$n-1$个主元，也就是说，方程组$A\boldsymbol{x}=\mathbf{0}$至少有1个自由元(自由元的个数等于未知数个数-主元个数，这里我们有$n$个未知数)。显然自由元的存在意味着方程组$A\boldsymbol{x}=\mathbf{0}$不可能只有一个平凡解而是有无穷多个解，与假设2.矛盾！因此那么$R$必然不可能包含一个0行。

现在，利用Theorem 1.7可知，不包含0行的$n$阶简化阶梯型$R$必然为单位矩阵$I_n$。

3. $\Rightarrow$ 4.: 假设$A$对应的简化阶梯型$R$为单位矩阵$I_n$。我们知道$R=I_n$是通过对原系数矩阵$A$进行一系列初等行变换得到的。因此假设我们一共进行了$k$次初等行变换。**而由Theorem 1.20我们又知道每次初等行变换可以通过对矩阵左乘一个初等矩阵实现。** 这意味着存在$k$个初等矩阵$E_1,\dots E_k$使得

$$E_kE_{k-1}\dots E_1A=I_n.$$

(注意这里我们用$E_1$标记第一次初等行变换对应的初等矩阵，$E_2$代表第二次初等行变换对应的初等矩阵，因此$E_2$需要左乘于$E_1A$，这解释了为何顺序是$E_kE_{k-1}\dots E_1A$。) 显然，以上等式配合Theorem 1.11告诉我们$A$可逆，且$A^{-1}=E_kE_{k-1}\dots E_1$。另外，以上等式同时告诉我们

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

而由Theorem 1.21 我们知道初等矩阵的逆也为初等矩阵，因此以上所有$E_1^{-1},E_2^{-1},\dots,E_k^{-1}$均为初等矩阵，即$A=E_1^{-1}E_2^{-1}\dots E_k^{-1}$ 可以表示为若干个初等矩阵的乘积。

4. $\Rightarrow$ 1.: 假设$A=P_1P_2\dots P_k$可以表示为$k$个初等矩阵$P_1,P_2,\dots,P_k$的乘积，那么首

先因为Theorem 1.21可知所有$P_1, P_2, \ldots, P_k$均可逆，再次利用Theorem 1.12可知作为若干个可逆矩阵的乘积，$A$也可逆，且有

$$A^{-1} = (P_1 P_2 \ldots P_k)^{-1} = P_k^{-1} \ldots P_2^{-1} P_1^{-1}.$$

□

**例子**(2021年线性代数期中考试题)：Find an invertible matrix $P$ such that $PA = B$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad B = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} - a_{21} & a_{32} - a_{22} & a_{33} - a_{23} \end{bmatrix}.$$

答案：显然，$B$是通过先对$A$的第二行乘以$-1$加到第三行上，再通过对$A$的第一行与第二行互换得到的。由于第一次初等行变换对应的初等矩阵为$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$，第二次初等行变化对应的初等矩阵为$E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$，我们得到 (**注意$E_1$与$E_2$相乘的顺序！**)

$$P = \color{red}{E_2 E_1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

由于$P$是两个初等矩阵的乘积且初等矩阵皆可逆，可知$P$也可逆，符合题意要求。

Theorem 1.22 告诉我们可以通过观察$A$的简化阶梯型$R$的形状来判断$A$是否可逆，**即若$R$包含有一个0行，那么$A$必然不可逆；若$R$是一个单位矩阵，那么$A$可逆**。另外，Theorem 1.22 3. $\Rightarrow$ 4. 的证明过程实际上告诉我们，若我们通过一系列初等行变换将$A$转化为简化阶梯型$R = I_n$，即$E_k E_{k-1} \ldots E_1 A = I_n$，那么利用关系$A^{-1} = E_k E_{k-1} \ldots E_1 = E_k E_{k-1} \ldots E_1 I_n$可知**同样的这一系列初等行变换作用在单位矩阵$I_n$上即可得到$A^{-1}$**。因此我们得到了求矩阵$A$的逆的一个算法 (inversion algorithm)。我们结合下面这个简单例子 (英文教材56页Example 4) 详细阐述这个求逆算法的具体步骤：

考虑矩阵$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$。我们现在想要求出它的逆矩阵$A^{-1}$。由以上讨论可知，

我们的核心思路是通过一系列初等行变换将$A$变为它的简化阶梯型$R = I_3$，那么通过将同样的这一系列初等行变换作用在单位矩阵$I_3$上即可得到$A^{-1}$。具体操作如下：

**第一步**：将矩阵$A$与单位矩阵$I_3$写在一起，$A$在左边，$I_3$在右边，中间可以用一条竖线隔开以避免将两者混淆。

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{array}\right)$$

**第二步**：在这一步里我们将左边的矩阵$A$使用初等行变换化为其简化阶梯型，即单位矩阵，同时在右边记录每一次初等行变换对$I_3$作用后所得到的矩阵。

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{array}\right)$$

左边：A的第一行乘以-2加到第二行，第一行乘以-1加到第三行；右边：$I_3$的第一行乘以-2加到第二行，第一行乘以-1加到第三行。

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{array}\right)$$

左边：A的第二行乘以2加到第三行；右边：$I_3$的第二行乘以2加到第三行。

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right)$$

左边：A的第三行乘以-1；右边：$I_3$的第三行乘以-1。

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & -14 & 6 & 3 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right)$$

左边：A的第三行乘以3加到第二行，第三行乘以-3加到第一行；右边：$I_3$的第三行乘以3加到第二行，第三行乘以-3加到第一行。

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -40 & 16 & 9 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right)$$

左边：A的第二行乘以-2加到第一行；右边：$I_3$的第二行乘以-2加到第一行。
**第三步**：现在，左边的矩阵$A$被转化为了它的简化阶梯型$R = I_3$，那么现在出现在右边的矩阵即是$A^{-1}$:

$$A^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}.$$

**例子**(2022年线性代数期中考试填空题)：The inverse of the matrix $A = \begin{bmatrix} 1 & -2 \\ 3 & 5 \end{bmatrix}$ is $\underline{?}$
该题可以通过以上算法求出答案。当然，由于题目中的矩阵为2阶方阵，也可以直接利用Theorem 1.18 进行求解。
在第二章我们将学习到其他判定一个矩阵是否可逆以及求逆的方法。

### 1.6. 线性方程组与可逆矩阵, More on linear system and invertible matrices

在本节我们利用已经学到的知识对线性方程组解的情况进行严格的研究与证明。

**Theorem 1.23** (英文教材Theorem 1.6.1). *一个线性方程组的解只有三种情况：无解，有且仅有一个解，以及有无穷多个解。*

*证明.* 显然我们只需要证明若一个线性方程组有两个不同的解，那么实际上它就有无穷多个解。首先还是注意到一个线性方程组

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ &\vdots \\ &a_{m1}x_1 + a_{n2}x_2 + \ldots + a_{mn}x_n = b_m. \end{aligned}$$

可以被表示为$A\boldsymbol{x} = \boldsymbol{b}$，$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$为$n \times 1$的未知数列向量，$\boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$为$m \times 1$的列向量， $A \in M_{m\times n}$为系数矩阵。
实际上在之前的课堂上我已经简单提到过这个定理的证明思路。这里我给出详细的证明细节。假设$\boldsymbol{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$与$\boldsymbol{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$为方程组$A\boldsymbol{x} = \boldsymbol{b}$的两个不同的解，即$\boldsymbol{s} \neq \boldsymbol{t}$，但

$$A\boldsymbol{s} = \boldsymbol{b} = A\boldsymbol{t}.$$

现在定义列向量$\boldsymbol{u}=\boldsymbol{s}-\boldsymbol{t}$，由于$\boldsymbol{s}\neq\boldsymbol{t}$，可知$\boldsymbol{u}\neq\boldsymbol{0}$不是零列向量。那么这意味着对任何不同的两个常数$c\neq d$，我们必然有$c\boldsymbol{u}\neq d\boldsymbol{u}$。另一方面，因为$A\boldsymbol{s}=\boldsymbol{b}=A\boldsymbol{t}$，我们还有$A\boldsymbol{u}=A\boldsymbol{s}-A\boldsymbol{t}=\boldsymbol{b}-\boldsymbol{b}=\boldsymbol{0}$，即，$\boldsymbol{u}=\boldsymbol{s}-\boldsymbol{t}$作为同一个线性方程组的两个解的差是该方程组系数矩阵$A$诱导出的齐次线性方程组$A\boldsymbol{x}=\boldsymbol{0}$的一个解。
现在，我们对每一个常数$c$定义一个新的列向量$\boldsymbol{s}+c\boldsymbol{u}$。那么由以上讨论可知

$$A(\boldsymbol{s}+c\boldsymbol{u})=A\boldsymbol{s}+cA\boldsymbol{u}=\boldsymbol{b}+\boldsymbol{0}=\boldsymbol{b},$$

这意味着所有这些$\boldsymbol{s}+c\boldsymbol{u}$都是方程组$A\boldsymbol{x}=\boldsymbol{b}$的解。又因为我们已经知道对于任何不同的两个常数$c\neq d$，我们必然有$c\boldsymbol{u}\neq d\boldsymbol{u}$，即$\boldsymbol{s}+c\boldsymbol{u}\neq\boldsymbol{s}+d\boldsymbol{u}$。这意味着该方程组实际上有无穷多个解(因为实数$c$有无穷多个)。证毕。 □

若一个方程组的方程个数与未知数个数相等，且系数矩阵$A$可逆，那么该方程组的唯一解可以直接通过对方程组最右边的常数列向量$\boldsymbol{b}$左乘$A^{-1}$得到。

**Theorem 1.24** (英文教材Theorem 1.6.2)**.** *若$A$为可逆$n$阶方阵，那么对任意$n\times 1$-矩阵(即列向量)$\boldsymbol{b}$，以$(A,\boldsymbol{b})$作为增广矩阵的线性方程组$A\boldsymbol{x}=\boldsymbol{b}$的唯一解为*

$$\boldsymbol{s}=A^{-1}\boldsymbol{b}.$$

证明. 首先容易验证对于列向量$\boldsymbol{s}=A^{-1}\boldsymbol{b}$我们有$A\boldsymbol{s}=A(A^{-1}\boldsymbol{b})=(AA^{-1})\boldsymbol{b}=\boldsymbol{b}$，这意味着$\boldsymbol{s}=A^{-1}\boldsymbol{b}$是该方程组的一个解。
现在我们证明$\boldsymbol{s}=A^{-1}\boldsymbol{b}$是该方程组唯一的一个解。那么假设$\boldsymbol{t}$是方程组的另一个解，即$A\boldsymbol{t}=\boldsymbol{b}$。对该等式两边同左乘$A^{-1}$可得

$$A^{-1}A\boldsymbol{t}=A^{-1}\boldsymbol{b}$$

从而得到$\boldsymbol{t}=A^{-1}\boldsymbol{b}=\boldsymbol{s}$。因此该方程组只有$\boldsymbol{s}$一个解。 □

利用以上定理我们可以将Theorem 1.22 扩展到以下形式。

**Theorem 1.25** (英文教材Theorem 1.6.4)**.** *设$A$为$n$阶方阵。那么以下说法等价(equivalent):*

1. *$A$可逆。*
2. *齐次线性方程组$A\boldsymbol{x}=\boldsymbol{0}$只有一个平凡解(trivial solution)。*
3. *$A$的简化阶梯型为单位矩阵$I_n$。*
4. *$A$可以被表示为一系列初等矩阵的乘积。*
5. *对任何$n\times 1$-矩阵(即列向量)$\boldsymbol{b}$，线性方程组$A\boldsymbol{x}=\boldsymbol{b}$有解。*
6. *对任何$n\times 1$-矩阵(即列向量)$\boldsymbol{b}$，线性方程组$A\boldsymbol{x}=\boldsymbol{b}$有且只有一个解。*

证明. 1. 到 4. 的等价关系我们已经在Theorem 1.22中证明。我们这里证明1. $\Rightarrow$ 6. $\Rightarrow$ 5. $\Rightarrow$ 1.。

1. $\Rightarrow$ 6.: 见Theorem 1.24的证明。

6. $\Rightarrow$ 5.: 显然。

5. $\Rightarrow$ 1.: 由假设，对任何$n \times 1$-矩阵(即列向量)$\boldsymbol{b}$，线性方程组$A\boldsymbol{x} = \boldsymbol{b}$都有解。特别地，考虑$i = 1, \ldots, n$，$\boldsymbol{e}_i = \begin{bmatrix} 0 & \ldots & 0 & \underbrace{1}_{\text{第i列}} & 0 & \ldots & 0 \end{bmatrix}$(即$\boldsymbol{e}_i$除了第$i$列上的项为1之外，其他项均为0)，令$\boldsymbol{b}$依次取$\boldsymbol{b} = \boldsymbol{e}_i^\top$，$i = 1, \ldots, n$，由假设我们可以找到$n \times 1$-列向量$\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$使得$A\boldsymbol{s}_i = \boldsymbol{e}_i^\top$，即

$$A\boldsymbol{s}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, A\boldsymbol{s}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, A\boldsymbol{s}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

令矩阵$C = \begin{bmatrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 & \ldots & \boldsymbol{s}_n \end{bmatrix}$，即$C$的列向量为$\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$。那么由矩阵乘法的列向量表示，见讲义第19页，我们得到

$$AC = \begin{bmatrix} A\boldsymbol{s}_1 & A\boldsymbol{s}_2 & \ldots & A\boldsymbol{s}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1^\top & \boldsymbol{e}_2^\top & \ldots & \boldsymbol{e}_n^\top \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = I_n.$$

这意味着$A$可逆，且$A^{-1} = C$。 □

在Theorem 1.12 里我们知道若$A, B$为$n$阶可逆方阵，那么$AB$也可逆，且有$(AB)^{-1} = B^{-1}A^{-1}$。下面这个定理证明反之亦成立。

**Theorem 1.26** (英文教材Theorem 1.6.5)**.** *设$A, B$为$n$阶方阵，且$AB$可逆。那么$A$与$B$都可逆。*

证明. 首先证明$B$可逆。假设$\boldsymbol{s}$是齐次线性方程组$B\boldsymbol{x} = \boldsymbol{0}$的一个解，即$B\boldsymbol{s} = \boldsymbol{0}$，那么我们有$AB\boldsymbol{s} = A(B\boldsymbol{s}) = \boldsymbol{0}$，即$\boldsymbol{s}$也是齐次线性方程组$(AB)\boldsymbol{x} = \boldsymbol{0}$的一个解。由假设，$AB$可逆，那么由Theorem 1.22 1. $\Rightarrow$ 2.，可知$(AB)\boldsymbol{x} = \boldsymbol{0}$仅有一个平凡解，因此必然有$\boldsymbol{s} = \boldsymbol{0}$。这意味着齐次线性方程组$B\boldsymbol{x} = \boldsymbol{0}$只有平凡解。因此利用Theorem 1.22 2. $\Rightarrow$ 1.，我们可以推出$B$可逆。

由于$B$可逆，我们有$A = AI_n = A(BB^{-1}) = (AB)B^{-1}$。那么$A$作为可逆矩阵$AB$与$B^{-1}$的乘积也必然可逆(利用Theorem 1.12)。证毕。 □

## 1.7. 对角矩阵，三角矩阵与对称矩阵, Diagonal, triangular and symmetric matrices

这一节主要介绍三种特殊的矩阵，即对角矩阵，三角矩阵与对称矩阵。由于该节内容较简单，讲义里将只简单提及一些相对重要的知识点，建议大家在读完讲义后再仔细阅读英文教材的第1.7节。

**注意本节出现的所有矩阵都是方阵。**

首先介绍对角矩阵。这个概念我们已经很熟悉，即若一个**方阵**所有非对角线上的项均为0，那么该矩阵就是一个对角矩阵(diagonal matrix)。比如单位矩阵就是一个典型的对角矩阵。显然，对一个对角矩阵$D$而言，我们只需要知道它对角线上的项。通常我们将对角矩阵写为

$$D = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix},$$

即$D_{11} = d_1, D_{22} = d_2, \dots, D_{nn} = d_n$。关于对角矩阵我们需要掌握以下两点：

- 对于任何自然数$k$，我们有

  $$D^k = \begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix}.$$

  这个特点让我们可以轻易计算对角矩阵的幂。
- **$D$可逆当且仅当它对角线上的所有项均不为零，或者说当且仅当**$d_1 d_2 \dots d_n \neq 0$。

第二个需要介绍的特殊矩阵是三角矩阵(triangular matrix)。三角矩阵又分为上三角矩阵(upper triangular matrix)与下三角矩阵(lower triangular matrix)。具体来讲

- 若一个方阵对角线**右上方**的所有项均为零，那么该方阵是一个**下三角矩阵**。即，方阵$A$是一个下三角矩阵当且仅当$A_{ij} = 0$对于所有$i < j$成立。
- 若一个方阵对角线**左下方**的所有项均为零，那么该方阵是一个**上三角矩阵**。即，方阵$A$是一个上三角矩阵当且仅当$A_{ij} = 0$对于所有$i > j$成立。

关于三角矩阵的特点请仔细阅读英文教材Theorem 1.7.1。这里我们简单提及该定理(b)部分的证明，即证明：若$A, B$为$n$阶下三角矩阵，那么$AB$也为下三角矩阵。(上三角的情况也同样成立。)

假设$A, B$为$n$阶下三角矩阵，即$A_{ij} = 0, B_{ij} = 0$对所有$i < j$成立。令$C = AB$。要

证明$C$是下三角矩阵，我们只需证明$C_{ij}=0$对所有$i<j$成立。由矩阵乘法定义可知，

$$C_{ij}=A_{i1}B_{1j}+A_{i2}B_{2j}+\ldots+A_{in}B_{nj}.$$

那么，对$i<j$的情况，我们可以将以上求和拆成两部分：

$$C_{ij}=A_{i1}B_{1j}+A_{i2}B_{2j}+\ldots+A_{i(j-1)}B_{(j-1)j}+A_{ij}B_{jj}+\ldots+A_{in}B_{nj}.$$

注意在红色部分中出现的$B_{1j},\ldots,B_{(j-1)j}$都为0 (由于$1<j,\ldots,j-1<j$，所以$B_{1j},\ldots,B_{(j-1)j}$都位于$B$对角线的右上方，故而等于零，因为$B$为下三角矩阵)，因此红色部分等于$A_{i1}\times 0+A_{i2}\times 0+\ldots+A_{i(j-1)}\times 0=0$；注意蓝色部分出现的$A_{ij},\ldots,A_{in}$都为0 (由于$i<j,i<j+1,\ldots,i<n$，$A_{ij},\ldots,A_{in}$都位于矩阵$A$右上方，故而等于零，因为$A$为下三角矩阵)，因此蓝色部分等于$0\times B_{jj}+\ldots+0\times B_{nj}=0$。这意味着$C_{ij}=0$。证毕。

最后我们介绍对称矩阵(symmetric matrix)。一个方阵$A$是对称的当且仅当$A^\top=A$，即它的转置矩阵等于它本身。显然这意味着$A_{ij}=A_{ji}$对所有$i,j$成立。

关于对称矩阵的性质请仔细阅读英文教材Theorem 1.7.2, 1.7.3, 1.7.4, 1.7.5。这里值得一提的是若矩阵$B\in M_{m\times n}$，那么$BB^\top$为$m$阶对称矩阵，$B^\top B$为$n$阶对称矩阵：

$$(BB^\top)^\top=(B^\top)^\top B^\top=BB^\top,$$

$$(B^\top B)^\top=B^\top(B^\top)^\top=B^\top B.$$

# 2 第二章：行列式

第一章的 Theorem 1.18 告诉我们一个2阶方阵$A=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$可逆当且仅当$ad-bc\neq 0$，并且在该条件成立时$A^{-1}=\frac{1}{ad-bc}\begin{bmatrix}d & -b\\ -c & a\end{bmatrix}$。实际上，这个定理可以推广到任意$n$阶的方阵，而这个能够决定一个方阵是否可逆的指标 ($n=2$时即是表达式$ad-bc$) 就是第二章的主角：行列式 (determinant)。

笼统来讲，行列式是一个把方阵$A$映射到实数的特殊函数，记作$\det(A)$，使得$A$可逆当且仅当它的行列式$\det(A)\neq 0$。实际上，行列式是一个非常复杂而深刻的数学概念，在第二章里我们仅学习它的定义和一些最基本的性质。本章最重要的任务是让大家学会**如何计算一些特殊矩阵的行列式**，并利用行列式求出矩阵的逆。

## 2.1. 行列式的代数余子式展开, Determinants by cofactor expansion

注意在本节所有矩阵均为**方阵**。

本节的主要目的是**给出行列式的定义**。

**Definition 2.1** (行列式). • 对于$1 \times 1$-矩阵$A = [a]$(即$A$是一个常数$a$)，$A$的行列式$\det(A)$就是$a$：$\det(A) = a$。

- 现在假设已经对所有$n-1$阶方阵($n \geq 2$)都完成了行列式的定义。对任何$n-1$阶方阵$B$，我们用$\det(B)$ ($\det(B)$是一个实数)来指代$B$的行列式。
- 设$A$为$n$阶方阵，我们按照之前的习惯将其表示为

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

对于$i, j = 1, \dots, n$，我们从$n$阶方阵$A$中**删掉第$i$行与第$j$列**，这样得到一个$n-1$-阶的矩阵$\tilde{A}^{ij}$，即

$$\tilde{A}^{ij} = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{n,n} \end{bmatrix}.$$

我们用$M_{ij} = \det(\tilde{A}^{ij})$来指代$n-1$-阶矩阵$\tilde{A}^{ij}$的行列式，称其为$a_{ij}$的余子式(*minor of entry $a_{ij}$*)或者$A$关于第$i$行第$j$列的余子式。相应的，我们称$C_{ij} = (-1)^{i+j}M_{ij}$为$a_{ij}$的代数余子式(*cofactor of entry $a_{ij}$*)或者$A$关于第$i$行第$j$列的代数余子式。

对于$i = 1, \dots, n$，我们称

$$a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in}$$

为$A$**沿第$i$行的代数余子式展开(*cofactor expansion along the i-th row*)**；对于$j = 1, \dots, n$，我们称

$$a_{1j}C_{1j} + a_{2j}C_{2j} + \dots + a_{nj}C_{nj}$$

为$A$**沿第$j$列的代数余子式展开(*cofactor expansion along the j-th column*)**。我们可以证明对任何$i = 1, \dots, n$，任何$j = 1, \dots, n$，**$A$沿第$i$行的代数余子式展开**与**$A$沿第$j$列的代数余子式展开**都相等。令$\det(A)$指代任意**$A$沿第$i$行的代数余子式展开**或**$A$沿第$j$列的代数余子式展开**所取的那个值，我们称这个数$\det(A)$为$n$阶方阵$A$的**行列式**。也就是说，$A$的行列式$\det(A)$满足

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in} \tag{2.1}$$

对任何$i = 1, \ldots, n$；以及

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj} \tag{2.2}$$

对任何$j = 1, \ldots, n$。

- 现在我们完成了对$n$阶方阵行列式的定义。利用$n$阶方阵的行列式我们可以复刻以上步骤去定义$n+1$阶方阵的行列式。通过这种方式我们可以对任意阶方阵定义其行列式。

**Remark 2.2.** 从以上定义可以看出，行列式的定义是迭代式的，即先定义1阶方阵的行列式，再利用1阶方阵的行列式去定义2阶方阵的行列式，再利用2阶方阵的行列式去定义3阶方阵的行列式...再利用$n$阶方阵的行列式去定义$n+1$阶方阵的行列式...这个过程可以类比于我们先对线段(1维物体)定义长度，再利用线段的长度去定义平面上图形的面积(2维物体)，再利用平面图形的面积去定义3维物体的体积。作为例子我们来写出1，2与3阶方阵的行列式表达式。

1. 对于$1 \times 1$-矩阵$A = [a]$(即$A$是一个常数$a$)，$A$的行列式$\det(A)$就是$a$：$\det(A) = a$。

2. 对于$2 \times 2$-矩阵$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$，我们首先计算$a_{11}$的代数余子式：删掉第*1*行与第*1*列后，我们得到的是一个1阶方阵$\tilde{A}^{11} = a_{22}$：

$$\tilde{A}^{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = [a_{22}],$$

   那么由余子式的定义可知$M_{11} = \det(\tilde{A}^{11}) = a_{22}$，由代数余子式的定义可得$C_{11} = (-1)^{1+1}M_{11} = a_{22}$；类似地，我们删掉第*1*行与第*2*列后得到1阶方阵$\tilde{A}^{12} = a_{21}$*:*

$$\tilde{A}^{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = [a_{21}],$$

   因此$a_{12}$的余子式为$M_{12} = \det(\tilde{A}^{12}) = a_{21}$，$a_{12}$的代数余子式为$C_{12} = (-1)^{1+2}M_{12} = -a_{21}$。现在我们可以计算出$A$沿第*1*行的代数余子式展开为

$$a_{11}C_{11} + a_{12}C_{12} = a_{11}a_{22} - a_{12}a_{21},$$

   根据行列式的定义，我们得到2阶方阵的行列式表达式为

$$\det(A) = \det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}C_{11} + a_{12}C_{12} = a_{11}a_{22} - a_{12}a_{21}.$$

3. 现在我们考虑3阶方阵$A = \begin{bmatrix} a_{11} & a_{12} & a_{12} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$。我们利用已经得到的2阶方阵行列式表达式来计算$\det(A)$。我们还是选取$A$的第一行来做代数余子式展开。

首先容易计算出$a_{11}$，$a_{12}$与$a_{13}$各自的代数余子式：

$$C_{11}=(-1)^{1+1}M_{11}=\det(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix})=\det(\begin{bmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{bmatrix})=a_{22}a_{33}-a_{23}a_{32},$$

$$C_{12}=(-1)^{1+2}M_{12}=-\det(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix})=-\det(\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix})=-(a_{21}a_{33}-a_{23}a_{31}),$$

$$C_{13}=(-1)^{1+3}M_{13}=\det(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix})=\det(\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix})=a_{21}a_{32}-a_{22}a_{31},$$

因此$A$沿第一行的代数余子式展开为

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

即以上表达式即是3阶方阵$A$的行列式$\det(A)$。

- 大家只需要记住2阶方阵的行列式表达式即可，上面这个3阶方阵行列式的表达式不需要记住！如果在作业或者考试中遇到3阶方阵行列式的计算，**只需按照定义现场计算矩阵沿某行或者某列的代数余子式展开即可**。实际上作业或者考试中出现的$n$阶方阵的行列式计算都是可以通过技巧化简为一个很容易计算的形式。这些技巧我们将在后续的课堂与作业中陆续学习。
- 英文教材第*110*页介绍了一个计算3阶方阵行列式的小技巧，感兴趣的同学可以了解一下。但这个技巧实际上对简化3阶方阵行列式的计算帮助很有限，也不会作为考点出现。大家直接忽略它也不会有任何影响。

**Remark 2.3.** 我们通常也把方阵$A$的行列式$\det(A)$记作$|A|$，比如

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

就是指代矩阵$\begin{bmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{bmatrix}$的行列式。注意这个符号在矩阵计算的环境下不是绝对值符号。

**Remark 2.4.** 在矩阵行列式的定义中我们使用到一个非常重要的事实：对任何$n$阶方阵$A$，**它沿任何一行或者任何一列的代数余子式展开都相等(即等于它的行列式)**：

$$|A|= a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in} = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$$

对任何$i, j = 1, \ldots, n$。这个结论的严格证明比较复杂因此不在本课程的学习范围内。大家只需牢记这个事实即可。

**例子**：我们分别对矩阵$\begin{bmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{bmatrix}$沿第一行与第一列进行代数余子式展开来计算它的行列式。

$$C_{11} = (-1)^{1+1}\begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix} = \begin{vmatrix} 5 & 6 \\ -8 & 9 \end{vmatrix} = 5 \times 9 - (-8) \times 6 = 93,$$

$$C_{12} = (-1)^{1+2}\begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix} = -\begin{vmatrix} -4 & 6 \\ 7 & 9 \end{vmatrix} = -(-4 \times 9 - 7 \times 6) = 78,$$

$$C_{13} = (-1)^{1+3}\begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix} = -\begin{vmatrix} -4 & 5 \\ 7 & -8 \end{vmatrix} = -4 \times (-8) - 7 \times 5 = -3,$$

因此沿第一行的代数余子式展开为：

$$1 \times C_{11} + 2 \times C_{12} + 3 \times C_{13} = 1 \times 93 + 2 \times 78 - 3 \times 3 = 240.$$

$$C_{21} = (-1)^{2+1}\begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix} = -\begin{vmatrix} 2 & 3 \\ -8 & 9 \end{vmatrix} = -(2 \times 9 - (-8) \times 3) = -42,$$

$$C_{31} = (-1)^{3+1}\begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} = 2 \times 6 - 5 \times 3 = -3,$$

因此沿第一列的代数余子式展开为：

$$1 \times C_{11} + (-4) \times C_{21} + 7 \times C_{31} = 1 \times 93 + (-4) \times (-42) + 7 \times (-3) = 240.$$

这验证了矩阵沿任何一行或任何一列的代数余子式展开都相等。另外我们也得到以上矩阵的行列式为240。

**例子**：(2022年线性代数期中考试题)

For the matrix $A = \begin{bmatrix} -2 & 1 & 0 \\ 0 & 3 & 2 \\ 5 & 1 & -4 \end{bmatrix}$, the cofactor $C_{23}$ =?

该题要求计算矩阵关于第2行第3列的cofactor，即代数余子式，因此依照代数余子式的定义，我们得到

$$C_{23} = (-1)^{2+3} \begin{vmatrix} -2 & 1 & 0 \\ 0 & 3 & 2 \\ 5 & 1 & -4 \end{vmatrix} = - \begin{vmatrix} -2 & 1 \\ 5 & 1 \end{vmatrix} = -(-2 \times 1 - 5 \times 1) = 7.$$

**注意：这里求的是代数余子式(cofactor)$C_{23}$而非余子式(minor)$M_{23}$，因此不要忘记在求出$M_{23}$后再乘上**$(-1)^{2+3}$。

由于计算矩阵行列式时我们可以任选沿一行或一列进行代数余子式展开，有些时候选取合适的行或者列进行代数余子式展开将会极大的减少我们的计算量。这就是计算矩阵行列式的第一个重要的技巧：尽量去找含0最多的那一行或者列去进行代数余子式展开。

**例子**：我们要计算矩阵$A = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 3 & 1 & 2 & 2 \\ 1 & 0 & -2 & 1 \\ 2 & 0 & 0 & 1 \end{bmatrix}$的行列式。由于第二列包含了3个为零的项，较其他列或行都多，我们选择沿第二列进行代数余子式展开。显然，由于$a_{12} = a_{32} = a_{42} = 0$，沿第二列的代数余子式展开为

$$a_{12} \times C_{12} + a_{22} \times C_{22} + a_{32} \times C_{32} + a_{42} \times C_{42} = 0 \times C_{12} + a_{22} \times C_{22} + 0 \times C_{32} + 0 \times C_{42} = a_{22} C_{22}.$$

由于$a_{22} = 1$，我们实际上得到$|A| = C_{22}$。因此

$$|A| = (-1)^{2+2} \begin{vmatrix} 1 & 0 & 0 & -1 \\ 3 & 1 & 2 & 2 \\ 1 & 0 & -2 & 1 \\ 2 & 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \\ 2 & 0 & 1 \end{vmatrix}.$$

对于计算$\begin{vmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \\ 2 & 0 & 1 \end{vmatrix}$我们注意到该矩阵第二列包含最多的0，因此沿第二列进行

代数余子式展开，可得

$$|A|=\begin{vmatrix}1&0&-1\\1&-2&1\\2&0&1\end{vmatrix}=-2\times(-1)^{2+2}\begin{vmatrix}1&0&-1\\1&-2&1\\2&0&1\end{vmatrix}=-2\times\begin{vmatrix}1&-1\\2&1\end{vmatrix}=-2\times(1\times1-2\times(-1))=-6.$$

**例子：**对上三角矩阵$A=\begin{bmatrix}a_{11}&a_{12}&\dots&a_{1n}\\0&a_{22}&\dots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\0&0&\dots&a_{nn}\end{bmatrix}$求行列式我们注意到第一列上首项$a_{11}$下方的所有项全为0，即$a_{21},\dots,a_{n1}$均为0，因此对其沿第一列进行代数余子式展开，可得

$$|A|=a_{11}C_{11}+0\times C_{21}+\dots+0\times C_{n1}=a_{11}C_{11}.$$

显然对$A$删掉首行首列后我们得到的还是一个上三角矩阵：

$$\begin{bmatrix}a_{22}&a_{23}&\dots&a_{2n}\\0&a_{33}&\dots&a_{3n}\\\vdots&\vdots&\ddots&\vdots\\0&0&\dots&a_{nn}\end{bmatrix},$$

继续对该矩阵沿第一列展开，同样的理由告诉我们$|A|=a_{11}a_{22}\begin{vmatrix}a_{33}&a_{34}&\dots&a_{3n}\\0&a_{44}&\dots&a_{4n}\\\vdots&\vdots&\ddots&\vdots\\0&0&\dots&a_{nn}\end{vmatrix}$。继续该步骤(或使用数学归纳法)我们最后得到

$$|A|=a_{11}a_{22}\dots a_{nn},$$

即上三角矩阵$A$的行列式为其对角线上所有项的乘积。

同理，对下三角矩阵沿第一行展开，我们也可以直接得到下三角矩阵$A$的行列式为其对角线上所有项的乘积。

显然，若一个矩阵含有0行或者0列，那么它的行列式一定为0。因为沿着0行或者0列进行代数余子式展开所得到的值必为零。(见英文教材Theorem 2.2.1)

**例子：** 假设$n$阶矩阵$A$具有以下分块矩阵(partitioned matrix)形状：

$$A=\begin{bmatrix}B&\mathbf{0}\\C&D\end{bmatrix},$$

其中$B$为$r$阶方阵，$\mathbf{0}=\mathbf{0}_{r\times k}$为$r\times k$-零矩阵，$C$为$k\times r$-矩阵，$D$为$k$阶方阵，$r+k=n$。换而言之，我们可以将$A$写为

$$\begin{bmatrix} B_{11} & \dots & B_{1r} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ B_{r1} & \dots & B_{rr} & 0 & \dots & 0 \\ C_{11} & \dots & C_{1r} & D_{11} & \dots & D_{1k} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_{k1} & \dots & C_{kr} & D_{k1} & \dots & D_{kk} \end{bmatrix}.$$

我们可以证明$\det(A)=\det(B)\det(D)$。同样的，对于$A=\begin{bmatrix} B & C \\ \mathbf{0} & D \end{bmatrix}$这样的分块矩阵我们也有$\det(A)=\det(B)\det(D)$。具体证明留作习题。

### 2.2. 通过行变换简化行列式的计算，Evaluating determinants by row reduction

在上一节的末尾我们看到三角矩阵的行列式计算非常简单，即，对任何三角矩阵$A$，它的行列式$\det(A)=a_{11}a_{22}\dots a_{nn}$等于它对角线上所有项的乘积。另一方面我们知道任何$n$阶方阵$A$都可以通过**初等行变换化为简化阶梯型 (reduced row echelon form)**$R$，并且由 Theorem 1.7 可知$R$要么包含至少一个0行，要么等于单位矩阵；这意味着$\det(R)=0$ (当$R$包含至少一个0行时，沿该0行的代数余子式展开为0，我们在上一节已经讨论过这个事实) 或者$\det(R)=1$ (当$R$为单位矩阵时，此时因为单位矩阵作为对角矩阵显然为一个三角矩阵，且其对角线上的项全为1，故有$\det(R)=1$)。因此，如果我们能知道**每一次初等行变换对矩阵行列式的影响**，那么通过初等行变换把原矩阵化为简化阶梯型，同时计算这些初等行变换对行列式累积造成的影响，最终结合$\det(R)=0$或1，我们可以计算出原矩阵的行列式。在本节我们将会学习到以上所提到的**“通过行变换简化行列式的计算”**。

**Theorem 2.5** (英文教材Theorem 2.2.3)**.** *令$A$为$n$阶方阵。*

1. *若$B$为通过对$A$的某一行乘以非零常数$c$得到的矩阵，那么$\det(B)=c\det(A)$。*
2. *若$B$为通过交换$A$的任意两行得到的矩阵，那么$\det(B)=-\det(A)$。*
3. *若$B$为通过对$A$的某一行乘以常数$c$加到另外一行上得到的矩阵，那么$\det(B)=\det(A)$。*

*证明.* 注意：该证明本身不在考试内容中，时间较紧张的同学可以忽略该证明。但请务必牢记定理内容！

- 令$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$，那么对$A$的第$i$行乘以非零常数$c$后得到

$$B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{i1} & ca_{i2} & \dots & ca_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}.$$

  显然，若我们沿$B$的第$i$行进行代数余子式展开，那么可得

$$\det(B) = ca_{i1}\tilde{C}_{i1} + ca_{i2}\tilde{C}_{i2} + \ldots + ca_{i1}\tilde{C}_{in},$$

  这里我们用$\tilde{C}_{ij}$指代矩阵$B$关于第i行第j列的代数余子式。由于这个行变换仅影响了第i行，即$B$除第$i$行外的所有其他行都与原矩阵$A$相同，那么**删掉$B$的第$i$行**第$j$列后得到的子矩阵$\tilde{B}^{ij}$与删掉$A$的第$i$行第$j$列后得到的子矩阵$\tilde{A}^{ij}$相同，因此$\tilde{C}_{ij} = (-1)^{i+j}\det(\tilde{B}^{ij}) = (-1)^{i+j}\det(\tilde{A}^{ij}) = C_{ij}$，即$B$**关于第i行第j列的代数余子式与$A$关于第i行第j列的代数余子式相等**。因此，由以上等式，我们实际上得到

$$\det(B) = ca_{i1}\tilde{C}_{i1} + ca_{i2}\tilde{C}_{i2} + \ldots + ca_{i1}\tilde{C}_{in} = c(a_{i1}C_{i1} + \ldots + a_{in}C_{in}) = c\det(A).$$

- 我们首先证明，若$B$是通过对$A$交换**相邻的两行**得到的，那么$\det(B) = -\det(A)$。

  让我们先从最简单的$n = 2$的情况开始。令$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$，交换它的第一行与第二行得到 $B = \begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix}$。注意，若我们把$B$写为$B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$，那么$b_{21} = a_{11}$，$b_{22} = a_{12}$。现在我们对$B$沿第2行进行代数余子式展开，得到

$$\det(B) = b_{21}\tilde{C}_{21} + b_{22}\tilde{C}_{22} = a_{11}\tilde{C}_{21} + a_{12}\tilde{C}_{22},$$

  这里$\tilde{C}_{ij}$还是指代矩阵$B$关于第i行第j列的代数余子式。注意，将$B$删掉第2行第1列后得到的是子矩阵是$a_{22}$，将$B$删掉第2行第2列后得到的是子矩阵是$a_{11}$；它们分别与将$A$删掉第1行第1列后得到的是子矩阵以及将$A$删掉第1行第2列后得到的是子矩阵相等！这是因为$B$的第一行是$A$的第二行，$B$的第二行是$A$的

第一行。根据代数余子式的定义，

$$\begin{aligned}\tilde{C}_{21} &= (-1)^{2+1} \underbrace{a_{22}}_{\text{将B删掉第2行第1列后得到的子矩阵的行列式}} \\ &= (-1)(-1)^{1+1} \underbrace{a_{22}}_{\text{将A删掉第1行第1列后得到的子矩阵的行列式}} \\ &= (-1)C_{11};\end{aligned}$$

同理可得$\tilde{C}_{22} = -C_{12}$。因此，

$$\begin{aligned}\det(B) &= b_{21}\tilde{C}_{21} + b_{22}\tilde{C}_{22} = a_{11}\tilde{C}_{21} + a_{12}\tilde{C}_{22} \\ &= -a_{11}C_{11} - a_{12}C_{12} \\ &= -\det(A).\end{aligned}$$

以上推导可以推广到任意自然数$n$的情况。具体来讲，对$n$阶方阵$A$交换相邻的第$i$行与第$i+1$行得到矩阵$B$，那么必然有$B_{i+1,j} = A_{ij}$，$B_{ij} = A_{i+1,j}$对所有$j = 1, \ldots, n$成立 (即$B$的第$i+1$行是$A$的第$i$行，$B$的第$i$行是$A$的第$i+1$行)，且$B$除了第$i$与第$i+1$行外的其他行与$A$对应的行都相等。现在我们对矩阵$B$沿第$i+1$行进行代数余子式展开 (令$\tilde{B}^{i+1,j}$代指将$B$删掉第$i+1$行第$j$列后得到的子矩阵，令$\tilde{C}_{i+1,j}$代指$B$关于第$i+1$行第j列的代数余子式)，可以得到

$$\begin{aligned}\det(B) &= B_{i+1,1}\underbrace{(-1)^{i+1+1}\det(\tilde{B}^{i+1,1})}_{=\tilde{C}_{i+1,1}} + \ldots + B_{i+1,n}\underbrace{(-1)^{i+1+n}\det(\tilde{B}^{i+1,n})}_{=\tilde{C}_{i+1,n}} \\ &= A_{i1}\underbrace{(-1)^{i+1+1}\det(\tilde{B}^{i+1,1})}_{=\tilde{C}_{i+1,1}} + \ldots + A_{in}\underbrace{(-1)^{i+1+n}\det(\tilde{B}^{i+1,n})}_{=\tilde{C}_{i+1,n}}.\end{aligned}$$

现在请大家自行验证，由于我们只是交换了$A$的**相邻**的$i$与$i+1$两行，所以

$$\underbrace{\tilde{B}^{i+1,j}}_{\text{删掉B第i+1行第j列得到的子矩阵}} = \underbrace{\tilde{A}^{ij}}_{\text{删掉A第i行第j列得到的子矩阵}}$$

对所有$j=1,\dots,n$成立，可参见下图。

$$
\tilde{B}^{i+1,j}=\begin{bmatrix}
b_{11} & \dots & b_{1j} & \dots & b_{1n}\\
\vdots & \vdots & \vdots & \vdots & \vdots\\
b_{i+1,1} & \dots & b_{i+1,j} & \dots & b_{i+1,n}\\
\vdots & \vdots & \vdots & \vdots & \vdots\\
b_{n1} & \dots & b_{nj} & \dots & b_{nn}
\end{bmatrix}
$$

$$
=\begin{bmatrix}
a_{11} & \dots & a_{1,j-1} & a_{1j} & a_{1,j+1} & \dots & a_{1n}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} & \dots & a_{i-1,n}\\
a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} & \dots & a_{i+1,n}\\
a_{i,1} & \dots & a_{i,j-1} & a_{ij} & a_{i,j+1} & \dots & a_{i,n}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
a_{n1} & \dots & a_{n,j-1} & a_{nj} & a_{n,j+1} & \dots & a_{nn}
\end{bmatrix}
\begin{array}{l}
\\ \\ \\ \text{(B的第i行)}\\ \text{(B的第i+1行)}\\ \\ \\
\end{array},
$$

$$
\tilde{A}^{i,j}=\begin{bmatrix}
a_{11} & \dots & a_{1j} & \dots & a_{1n}\\
\vdots & \vdots & \vdots & \vdots & \vdots\\
a_{i,1} & \dots & a_{i,j} & \dots & a_{i,n}\\
\vdots & \vdots & \vdots & \vdots & \vdots\\
a_{n1} & \dots & a_{nj} & \dots & a_{nn}
\end{bmatrix}
$$

$$
=\begin{bmatrix}
a_{11} & \dots & a_{1,j-1} & a_{1j} & a_{1,j+1} & \dots & a_{1n}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} & \dots & a_{i-1,n}\\
a_{i,1} & \dots & a_{i,j-1} & a_{ij} & a_{i,j+1} & \dots & a_{i,n}\\
a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} & \dots & a_{i+1,n}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
a_{n1} & \dots & a_{n,j-1} & a_{nj} & a_{n,j+1} & \dots & a_{nn}
\end{bmatrix}
\begin{array}{l}
\\ \\ \\ \text{(A的第i行)}\\ \text{(A的第i+1行)}\\ \\ \\
\end{array}.
$$

因此，我们有$\tilde{C}_{i+1,j}=(-1)^{i+j+1}\det(\tilde{B}^{i+1,j})=(-1)(-1)^{i+j}\det(\tilde{A}^{i,j})=-C_{ij}$

($C_{ij}$为$A$关于第i行第j列的代数余子式)对任何$j = 1, \ldots, n$成立，故而

$$\begin{aligned}
\det(B) &= B_{i+1,1} \underbrace{(-1)^{i+1+1} \det(\tilde{B}^{i+1,1})}_{=\tilde{C}_{i+1,1}} + \ldots + B_{i+1,n} \underbrace{(-1)^{i+1+n} \det(\tilde{B}^{i+1,n})}_{=\tilde{C}_{i+1,n}} \\
&= A_{i1} \underbrace{(-1)^{i+1+1} \det(\tilde{B}^{i+1,1})}_{=\tilde{C}_{i+1,1}} + \ldots + A_{in} \underbrace{(-1)^{i+1+n} \det(\tilde{B}^{i+1,n})}_{=\tilde{C}_{i+1,n}} \\
&= -A_{i1}C_{i1} - \ldots - A_{in}C_{in} \\
&= -\underbrace{(A_{i1}C_{i1} + \ldots + A_{in}C_{in})}_{\text{A沿第i行的代数余子式展开}} \\
&= -\det(A).
\end{aligned}$$

现在我们证明若$B$是通过对$A$交换任意两行得到的矩阵，那么$\det(B) = -\det(A)$。不妨假设互换的是第$i$行与第$k$行且$k > i+1$。我们首先将第$i$行与它下方相邻的第$i+1$行交换，得到矩阵$A_1$，再将$A_1$的第$i+1$行与第$i+2$行互换，得到矩阵$A_2$，继续$k-i$次后我们得到一个矩阵$A_{k-i}$，显然它的第$i$行是原来$A$的第$i+1$行，它的第$i+1$行是原来$A$的第$i+2$行，...，它的第$k-1$行是原来$A$的第$k$行，它的第$k$行是原来$A$的第$i$行。由于在这个过程中我们共**交换了相邻两行**$k-i$**次**，根据我们之前的分析可知$\det(A_{k-i}) = (-1)^{k-i}\det(A)$。现在将$A_{k-i}$的第$k-1$行(是原矩阵$A$的第$k$行)与第$k-2$行(是原矩阵$A$的第$k-1$行)交换得到矩阵$A_{k-i+1}$，将$A_{k-i+1}$的第$k-2$行(是原矩阵$A$的第$k$行)与第$k-3$行(是原矩阵$A$的第$k-2$行)进行交换得到$A_{k-i+2}$，这样继续$k-i-1$次后我们最终得到矩阵$B = A_{k-i+(k-i-1)}$，它的第$i$行是$A$的第$k$行，第$k$行是$A$的第$i$行，其余行与$A$对应的行一致。从$A_{k-i}$到$B = A_{k-i+(k-i-1)}$**交换了相邻两行**$k-i-1$**次**，因此由之前的分析可得$\det(B) = (-1)^{k-i-1}\det(A_{k-i}) = (-1)^{k-i-1}(-1)^{k-i}\det(A) = (-1)^{2(k-i)-1}\det(A)$。由于$2(k-1)+1$是个奇数，我们得到$\det(B) = -\det(A)$。

- 首先注意，由于已经证明了交换$A$的两行后得到的矩阵$B$满足$\det(B) = -\det(A)$，我们可知：若矩阵$A$包含有相同的两行，比如第$i$行与第$k$行，那么交换这两行后得到的矩阵$B$依然与$A$相同，那么这意味着$\det(B) = \det(A)$与$\det(B) = -\det(A)$必须同时成立，即$\det(A) = 0$。现在假设$B$是通过将$A$的第$i$行乘以常数$c$加到第$k$行上得到的，即$B$第k行为

$$\begin{bmatrix} b_{k1} & b_{k2} & \ldots & b_{kn} \end{bmatrix} = \begin{bmatrix} a_{k1} + ca_{i1} & a_{k2} + ca_{i2} & \ldots & a_{kn} + ca_{in} \end{bmatrix},$$

以及$B$除去第$i$行外的其余行与$A$对应的行都相等。那么后者意味着将$B$删去第$k$行第$j$列后的所有子矩阵都与将$B$删去第$k$行第$j$列后的所有子矩阵相等($j = 1, \ldots, n$)，因此必有$\tilde{C}_{kj} = C_{kj}$，这里$\tilde{C}_{kj}$为$B$关于第k行第j列的代数余子式，$C_{kj}$为$A$关于第k行第j列的代数余子式。将$B$沿着第k行进行代数余子式展开，

我们得到

$$\begin{aligned}\det(B) = b_{k1}\tilde{C}_{k1} + \ldots + b_{kn}\tilde{C}_{kn} &= (a_{k1} + ca_{i1})C_{k1} + \ldots (a_{kn} + ca_{in})C_{kn} \\ &= a_{k1}C_{k1} + \ldots + a_{kn}C_{kn} + c(a_{i1}C_{k1} + \ldots + a_{in}C_{kn}).\end{aligned}$$

现在我们观察$a_{i1}C_{k1} + \ldots + a_{in}C_{kn}$。经过简单思考容易看出，这个表达式实际上是矩阵$P$沿第$k$行的代数余子式，而这个矩阵$P$的第$i$行与第$k$行都与$A$的第$i$行相等，$P$的其余行与$A$的对应行相等。

$$P = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in} \text{ (第i行)} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in} \text{ (第k行)} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}$$

由于$P$含有相同的两行，我们知道$\det(P) = a_{i1}C_{k1} + \ldots + a_{in}C_{kn} = 0$。因此

$$\begin{aligned}\det(B) = b_{k1}\tilde{C}_{k1} + \ldots + b_{kn}\tilde{C}_{kn} &= (a_{k1} + ca_{i1})C_{k1} + \ldots (a_{kn} + ca_{in})C_{kn} \\ &= a_{k1}C_{k1} + \ldots + a_{kn}C_{kn} + c(a_{i1}C_{k1} + \ldots + a_{in}C_{kn}) \\ &= a_{k1}C_{k1} + \ldots + a_{kn}C_{kn} + c\det(P) \\ &= a_{k1}C_{k1} + \ldots + a_{kn}C_{kn} + 0 \\ &= a_{k1}C_{k1} + \ldots + a_{kn}C_{kn}.\end{aligned}$$

这意味着$\det(B) = a_{k1}C_{k1} + \ldots + a_{kn}C_{kn} = \det(A)$，因为$a_{k1}C_{k1} + \ldots + a_{kn}C_{kn}$是$A$沿第k行的代数余子式展开。

□

**Remark 2.6.** *尽管 Theorem 2.5 的证明不需要掌握，在证明过程中得到的两个结论却是我们需要记住的：*

1. *若矩阵$A$包含有相同的两行，比如第$i$行与第$k$行，那么交换这两行后得到的矩阵$B$依然与$A$相同，那么这意味着$\det(B) = \det(A)$与$\det(B) = -\det(A)$ (该等式由 Theorem 2.5 的第2条说法得到) 必须同时成立，即$\det(A) = 0$。*
2. *对于任何$n$阶方阵$A$，$i \neq k$，我们有*

$$a_{i1}C_{k1} + \ldots + a_{in}C_{kn} = 0.$$

*这是因为$a_{i1}C_{k1} + \ldots + a_{in}C_{kn}$实际上是矩阵$P$沿第$k$行展开的代数余子式，而这个矩阵$P$的第$i$行与第$k$行都与$A$的第$i$行相等，$P$的其余行与$A$的对应行相*

等要证实$a_{i1}C_{k1}+\ldots+a_{in}C_{kn}$实际上是矩阵$P$沿第$k$行展开的代数余子式，只需注意，由于$P$与$A$唯一的区别在于$P$的第$k$行等于$A$的第$i$行，因此删掉$P$的第$k$行第$j$列后得到的子矩阵与删掉$A$的第$k$行第$j$列后得到的子矩阵是完全一样的！这意味着$P$关于第$k$行第$j$列的代数余子式$\tilde{C}_{kj}$等于$A$关于第$k$行第$j$列的代数余子式$C_{kj}$对任何$j=1,\ldots,n$成立;因此$P$沿第$k$行的代数余子式展开等于

$$P_{k1}\tilde{C}_{k1}+\ldots+P_{kn}\tilde{C}_{kn}=a_{i1}C_{k1}+\ldots+a_{in}C_{kn}.$$

$$P=\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in}\,(\text{第}i\text{行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in}\,(\text{第}k\text{行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}$$

由于$P$含有相同的两行，我们知道$\det(P)=a_{i1}C_{k1}+\ldots+a_{in}C_{kn}=0$。

以上两条性质非常重要，我们将在下一节使用到它们。此外这两条性质也是潜在的考点，相关的考试题我们将在下一节详细分析。

由以上Remark的第一点我们可以立刻得到以下定理，见英文教材Theorem 2.2.5。

**Theorem 2.7.** 若$A$为$n$阶方阵，且它的第$k$行等于第$i$行乘以某个常数$c$，那么有$\det(A)=0$。

证明. 由假设可知

$$A=\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in}\,(\text{第i行}) \\ \vdots & \vdots & \vdots & \vdots \\ ca_{i1} & ca_{i2} & \ldots & ca_{in}\,(\text{第k行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}$$

是将矩阵

$$P = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in}\ (\text{第i行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in}\ (\text{第k行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

的第k行乘以常数$c$得到的，因此由Theorem 2.5的第1点可知

$$\det(A) = c \det(P).$$

又因为$P$含有相同的两行，我们知道$\det(P) = 0$，因此$\det(A) = c \times 0 = 0$，证毕。

$\square$

由于$\det(I_n) = 1$ (注意单位矩阵$I_n$为对角矩阵，因此由上一节的讨论可知它的行列式等于其对角线上所有项的乘积，即等于1)，下面的定理是 Theorem 2.5 的直接推论，见英文教材 Theorem 2.2.4。

**Theorem 2.8.** *设$E$为$n$阶初等矩阵。*

1. *若$E$是由对单位矩阵$I_n$的某一行乘以非零常数$c$得到的，那么$\det(E) = c$。*
2. *若$E$是由交换单位矩阵$I_n$的某两行得到的，那么$\det(E) = -1$。*
3. *若$E$是由将单位矩阵$I_n$的某一行乘以常数$c$加到另外一行上得到的，那么$\det(E) = 1$。*

注意，由于矩阵沿任何一行或者任何一列的代数余子式都等于它的行列式，而矩阵的转置是将行变成列，列变成行，所以以下定理显然成立。

**Theorem 2.9.** *对任何$n$阶方阵$A$，都有$\det(A) = \det(A^\top)$。*

*证明.* 由于$(A^\top)_{ij} = A_{ji}$对任何$i, j = 1, \dots, n$，对$A^\top$沿第$i$行进行代数余子式展开显然等于对$A$沿第$i$列进行代数余子式展开；类似的，对$A^\top$沿第$i$列进行代数余子式展开显然等于对$A$沿第$i$行进行代数余子式展开。而对$A^\top$沿第$i$行或列进行代数余子式展开等于$\det(A^\top)$，对$A$沿第$i$行或列进行代数余子式展开等于$\det(A)$。故有$\det(A^\top) = \det(A)$。 $\square$

由于矩阵沿任何一行或者任何一列的代数余子式展开都等于它的行列式，本节所有关于行的结论都可以照搬到关于列的情况。比如我们有：

- 若矩阵$A$的第$k$列等于第$i$列乘以某个常数$c$，那么$\det(A)=0$；
- $a_{1j}C_{1k}+a_{2j}C_{2k}+\ldots+a_{nj}C_{nk}=0$对任何$j\neq k$成立；
- 若$B$为通过对$A$的某一列乘以非零常数$c$得到的矩阵，那么$\det(B)=c\det(A)$。

- 若$B$为通过交换$A$的任意两列得到的矩阵，那么$\det(B)=-\det(A)$。
- 若$B$为通过对$A$的某一列乘以常数$c$加到另外一列上得到的矩阵，那么$\det(B)=\det(A)$。

现在我们可以通过对矩阵进行**行变换或者列变换**来化简行列式的计算。

**例子**：矩阵$A=\begin{bmatrix}0&1&5\\3&-6&9\\2&6&1\end{bmatrix}$。通过行变换将其化为阶梯型，同时追踪这些行变换对行列式的影响，我们可以快速计算出它的行列式：

$$\begin{aligned}\det(A)=\begin{vmatrix}0&1&5\\3&-6&9\\2&6&1\end{vmatrix}&=-\begin{vmatrix}3&-6&9\\0&1&5\\2&6&1\end{vmatrix}\quad(\text{第1行与第2行交换})\\&=-3\begin{vmatrix}1&-2&3\\0&1&5\\2&6&1\end{vmatrix}\quad(\begin{bmatrix}3&-6&9\end{bmatrix}=3\begin{bmatrix}1&-2&3\end{bmatrix})\\&=-3\begin{vmatrix}1&-2&3\\0&1&5\\0&10&-5\end{vmatrix}\quad(\text{第1行乘以-2加到第3行})\\&=-3\begin{vmatrix}1&-2&3\\0&1&5\\0&0&-55\end{vmatrix}\quad(\text{第2行乘以-10加到第3行})\\&=(-3)(-55)\begin{vmatrix}1&-2&3\\0&1&5\\0&0&1\end{vmatrix}(\begin{bmatrix}0&0&-55\end{bmatrix}=-55\begin{bmatrix}0&0&1\end{bmatrix})\\&=(-3)(-55)(1)=165.\end{aligned}$$

**例子**：矩阵$A=\begin{bmatrix}1&0&0&3\\2&7&0&6\\0&6&3&0\\7&3&1&-5\end{bmatrix}$。我们通过列变换将其化为一个下三角矩阵，

同时追踪这些行变换对行列式的影响，从而快速计算出它的行列式：

$$\begin{aligned}
\det(A) &= \begin{vmatrix} 1 & 0 & 0 & 3 \\ 2 & 7 & 0 & 6 \\ 0 & 6 & 3 & 0 \\ 7 & 3 & 1 & -5 \end{vmatrix} \\
&= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 7 & 0 & 0 \\ 0 & 6 & 3 & 0 \\ 7 & 3 & 1 & -26 \end{vmatrix} \quad \text{(第1列乘以-3加到第4列上)} \\
&= 1 \times 7 \times 3 \times (-26) = -546.
\end{aligned}$$

**例子：** 矩阵$A = \begin{bmatrix} 3 & 5 & -2 & 6 \\ 1 & 2 & -1 & 1 \\ 2 & 4 & 1 & 5 \\ 3 & 7 & 5 & 3 \end{bmatrix}$。我们通过行(列)变换在某一行或某一列制造出尽可能多的0来化简行列式的计算。

$$\begin{aligned}
\det(A) &= \begin{vmatrix} 3 & 5 & -2 & 6 \\ 1 & 2 & -1 & 1 \\ 2 & 4 & 1 & 5 \\ 3 & 7 & 5 & 3 \end{vmatrix} \\
&= \begin{vmatrix} 0 & -1 & 1 & 3 \\ 1 & 2 & -1 & 1 \\ 0 & 0 & 3 & 3 \\ 0 & 1 & 8 & 0 \end{vmatrix} \quad \text{通过第2行乘以某些常数加到第1，3，4行} \\
&= -\begin{vmatrix} -1 & 1 & 3 \\ 0 & 3 & 3 \\ 1 & 8 & 0 \end{vmatrix} \quad \text{沿第1列进行代数余子式展开} \\
&= -\begin{vmatrix} -1 & 1 & 3 \\ 0 & 3 & 3 \\ 0 & 9 & 3 \end{vmatrix} \quad \text{第1行加到第3行} \\
&= -(-1)\begin{vmatrix} 3 & 3 \\ 9 & 3 \end{vmatrix} \quad \text{沿第1列进行代数余子式展开} \\
&= 3 \times 3 - 9 \times 3 = -18.
\end{aligned}$$

## 2.3. 行列式的性质, Properties of determinants

本节中提到的行列式的几个性质都非常重要且会**作为考点出现在考试中**。另一方面它们的证明并不作要求 (但建议大家在有时间的情况下阅读并理解证明思路)，时间紧张的同学可以先略过证明，**但务必牢记这几条性质的内容!**

**Theorem 2.10.** 设$A$为$n$阶方阵，$c$为常数，那么$\det(cA) = c^n \det(A)$。

证明. 我们使用数学归纳法。当$n = 1$时$A = [a_{11}]$，此时显然有$\det(cA) = \det([ca_{11}]) = ca_{11}$。现在假设对某个自然数$n$以及任何$n$阶方阵$A$，任何常数$c$都有$\det(cA) = c^n \det(A)$。我们考虑$n+1$阶方阵$A$，此时$cA$为

$$\begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1,n+1} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ ca_{n+1,1} & ca_{n+1,2} & \dots & ca_{n+1,n+1} \end{bmatrix}.$$

令$\widetilde{cA}^{1j}$指代矩阵$cA$删去第一行第$j$列后得到的$n$阶子矩阵，令$\tilde{A}^{1j}$指代矩阵$A$删去第一行第$j$列后得到的$n$阶子矩阵。经过简单观察可知$\widetilde{cA}^{1j} = c\tilde{A}^{1j}$。因为它们为$n$阶方阵，由归纳假设可得$\det(\widetilde{cA}^{1j}) = \det(c\tilde{A}^{1j}) = c^n \det(\tilde{A}^{1j})$对任何$j = 1, \dots, n+1$成立。因此对$cA$沿第一行做代数余子式展开可得

$$\begin{aligned} det(cA) &= ca_{11} \underbrace{(-1)^{1+1}\det(\widetilde{cA}^{11})}_{\text{cA关于第一行第一列的代数余子式}} + \dots + ca_{1,n+1} \underbrace{(-1)^{n+1+1}\det(\widetilde{cA}^{1,n+1})}_{\text{cA关于第一行第n+1列的代数余子式}} \\ &= ca_{11}c^n \underbrace{(-1)^{1+1}\det(\tilde{A}^{11})}_{\text{A关于第一行第一列的代数余子式}} + \dots + ca_{1,n+1}c^n \underbrace{(-1)^{n+1+1}\det(\tilde{A}^{1,n+1})}_{\text{A关于第一行第n+1列的代数余子式}} \\ &= c^{n+1}(a_{11} \underbrace{C_{11}}_{\text{A关于第一行第一列的代数余子式}} + \dots + a_{1,n+1} \underbrace{C_{1,n+1}}_{\text{A关于第一行第n+1列的代数余子式}}) \\ &= c^{n+1}\det(A). \end{aligned}$$

因此我们证明了$n+1$阶方阵的情况。由数学归纳法可得该等式对任何自然数$n$均成立。 □

**Theorem 2.11** (英文教材Theorem 2.3.1)**.** 设$C$为$n$阶方阵，并且具有以下形状：

$$C = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1}+b_{i1} & a_{i2}+b_{i2} & \dots & a_{in}+b_{in}\ (\text{第}i\text{行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}.$$

那么，令

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in}\,(\text{第}i\text{行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

令

$$B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{i1} & b_{i2} & \dots & b_{in}\,(\text{第}i\text{行}) \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

(即$C$矩阵的第$i$行等于$A$的第$i$行与$B$的第$i$行之和，$C$的其余行与$A$的其余行相同，也与$B$的其余行相同)，我们有

$$\det(C) = \det(A) + \det(B).$$

类似的，若$C$可表示为

$$\begin{bmatrix} a_{11} & \dots & a_{1j} + b_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nj} + b_{nj} & \dots & a_{nn} \end{bmatrix},$$

那么令

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix},$$

令

$$B = \begin{bmatrix} a_{11} & \dots & b_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & b_{nj} & \dots & a_{nn} \end{bmatrix}$$

(即$C$矩阵的第$j$列等于$A$的第$j$列与$B$的第$j$列之和，$C$的其余列与$A$的其余列相同，也与$B$的其余列相同)，我们有

$$\det(C) = \det(A) + \det(B).$$

证明. 若

$$C=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1}+b_{i1} & a_{i2}+b_{i2} & \dots & a_{in}+b_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

那么显然$C$关于第$i$行第$j$列的代数余子式(记作$C_{ij}$)与$A$关于第$i$行第$j$列的代数余子式相等，也与$B$关于第$i$行第$j$列的代数余子式相等(对任何$j=1,\dots,n$)，因此对$C$沿第$i$行做代数余子式展开得到

$$\begin{aligned} \det(C) &= (a_{i1}+b_{i1})C_{i1}+\dots+(a_{in}+b_{in})C_{in} \\ &= a_{i1}\underbrace{C_{i1}}_{\text{A关于第i行第1列的代数余子式}}+\dots+a_{in}\underbrace{C_{in}}_{\text{A关于第i行第n列的代数余子式}} \\ &\quad+b_{i1}\underbrace{C_{i1}}_{\text{B关于第i行第1列的代数余子式}}+\dots+b_{in}\underbrace{C_{in}}_{\text{B关于第i行第n列的代数余子式}} \\ &= \det(A)+\det(B). \end{aligned}$$

若

$$\begin{bmatrix} C=a_{11} & \dots & a_{1j}+b_{1j} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nj}+b_{nj} & \dots & a_{nn} \end{bmatrix},$$

那么对$C$沿第$j$列做代数余子式展开可得$\det(C)=\det(A)+\det(B)$。 □

注意：一般来说$\det(A+B)\neq\det(A)+\det(B)$。比如令$A=\begin{bmatrix}1&2\\2&5\end{bmatrix}$，$B=\begin{bmatrix}3&1\\1&3\end{bmatrix}$，那么$A+B=\begin{bmatrix}4&3\\3&8\end{bmatrix}$，但$\det(A+B)=23$，$\det(A)+\det(B)=1+b=9$，两者不相等。

**Lemma 2.12** (英文教材Lemma 2.3.2). *令$B$为$n$阶方阵，$E$为$n$阶初等矩阵，那么*$\det(EB)=\det(E)\det(B)$。

证明. 由于初等行变换有三种类型，我们进行如下分类讨论。

1. 假设$E$是由$I_n$的某一行乘以非零常数$c$得到。此时由Theorem 2.8可知$\det(E)=c$。另一方面我们知道$EB$等于将$B$的某一行乘以$c$，因此由Theorem 2.5又有$\det(EB)=c\det(B)$。显然这说明$\det(EB)=\det(E)\det(B)$。

2. 假设$E$是由$I_n$交换某两行得到。那么由Theorem 2.8可知$\det(E) = -1$。另一方面我们知道$EB$等于将$B$的某两行互换，因此由Theorem 2.5又有$\det(EB) = -\det(B)$。显然这说明$\det(EB) = \det(E)\det(B)$。
3. 假设$E$是由$I_n$的某一行乘以常数$c$加到另一行所得到。那么由Theorem 2.8可知$\det(E) = 1$。另一方面我们知道$EB$等于将$B$的某一行乘以常数$c$加到另一行，因此由Theorem 2.5又有$\det(EB) = \det(B)$。显然这说明$\det(EB) = \det(E)\det(B)$。

□

**Theorem 2.13** (英文教材Theorem 2.3.3)**.** *一个$n$阶方阵$A$可逆(invertible)当且仅当$\det(A) \neq 0$。*

证明. 令$R$为$A$的简化阶梯型。由1.5节的内容可知，此时有若干个初等矩阵$E_1, \ldots, E_r$使得$R = E_r \ldots E_2 E_1 A$。那么利用Lemma 2.12我们有

$$\det(R) = \det(E_r) \ldots \det(E_2) \det(E_1) \det(A).$$

若$R$为单位矩阵$I_n$；此时我们有$\det(R) = \det(I_n) = 1 \neq 0$；若$\det(R) \neq 0$，那么必然意味着$R$中不可能包含0行，因此由Theorem 1.7可知此时$R$必是单位矩阵。因此我们实际上得到$A$的简化阶梯型$R$为单位矩阵$I_n$当且仅当$\det(R) \neq 0$，即$\det(R) = \det(E_r) \ldots \det(E_2) \det(E_1) \det(A) \neq 0$。另一方面，由Theorem 2.8我们知道任何初等矩阵$E$的行列式均不为0(若$E$是由$I_n$的某一行乘以非零常数$c$得到，那么$\det(E) = c \neq 0$；若$E$是由$I_n$交换某两行得到，那么$\det(E) = -1 \neq 0$；若$E$是由$I_n$的某一行乘以常数$c$加到另一行所得到，那么$\det(E) = 1 \neq 0$)。因此，利用Theorem 1.22，我们有如下等价关系：

$$\begin{aligned}
A \text{ 可逆 } &\Leftrightarrow R \text{ 是单位矩阵} \quad (\text{Theorem 1.22}) \\
&\Leftrightarrow \det(R) \neq 0 \\
&\Leftrightarrow \text{存在若干初等矩阵} E_1, \ldots, E_r, \det(R) = \underbrace{\det(E_r) \ldots \det(E_2) \det(E_1)}_{\neq 0} \det(A) \neq 0 \\
&\Leftrightarrow \det(A) \neq 0.
\end{aligned}$$

□

**Theorem 2.14** (英文教材Theorem 2.3.4)**.** *对任何$n$阶方阵$A$与$B$，有*

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

证明. 若$A$不可逆，那么$AB$也不可逆(见Theorem 1.26)。此时由Theorem 2.13可知$\det(A) = 0$，$\det(AB) = 0$，因此显然有$\det(AB) = \det(A)\det(B)$。

现在假设$A$可逆。那么由Theorem 1.22可知此时$A$可以被表示为若干个初等矩阵的乘积：$A = E_1 \dots E_r$。此时我们有$\det(AB) = \det(E_1 \dots E_r B)$。那么利用Lemma 2.12可得$\det(AB) = \det(E_1)\dots\det(E_r)\det(B)$。再次利用Lemma 2.12可得$\det(A) = \det(E_1 \dots E_r I_n) = \det(E_1)\dots\det(E_r)\det(I_n) = \det(E_1)\dots\det(E_r)$，因此$\det(AB) = \det(E_1)\dots\det(E_r)\det(B) = \det(A)\det(B)$。 □

**例子：** (2021年线性代数期中考试题)

Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 6 \end{bmatrix}$. Compute the following trace and determinant:

$\mathrm{tr}(A^\top) = ?$, $\quad\quad \det(A^{2021}) = ?$

**答案：** $\mathrm{tr}(A^\top) = \mathrm{tr}(A) = 1 + 2 + 6 = 9$。

$\det(A) = -1$，因此$\det(A^{2021}) = \det(A)^{2021} = (-1)^{2021} = -1$。

**Theorem 2.15** (英文教材Theorem 2.3.5). *若$n$阶方阵$A$可逆，那么*

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

*证明.* 利用Theorem 2.14可得$\det(AA^{-1}) = \det(A)\det(A^{-1})$，因此 $1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$，所以 $\det(A^{-1}) = \frac{1}{\det(A)}$。 □

**Definition 2.16** (英文教材第122页，Definition 1). *对于$n$阶方阵$A$，令$C_{ij}$为$A$关于第$i$行第$j$列的代数余子式。我们称矩阵*

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

*为$A$的代数余子式矩阵(matrix of cofactors from $A$)。我们称$C$的转置$C^\top$为$A$的伴随矩阵(adjoint matrix of $A$)，并记为$adj(A)$，即*

$$adj(A) = \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}.$$

**Theorem 2.17** (英文教材Theorem 2.3.6). *若$A$为可逆$n$阶方阵，那么*

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

实际上，**对任何$n$阶方阵$A$(不论是否可逆)**，我们有

$$A\mathit{adj}(A) = \det(A)I_n.$$

证明. 我们首先证明对任何$n$阶方阵$A$，有

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

由伴随矩阵$\mathrm{adj}(A)$的定义可知，

$$A\mathrm{adj}(A) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & \dots & C_{j1} & \dots & C_{n1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{1n} & \dots & C_{jn} & \dots & C_{nn} \end{bmatrix}.$$

若$i = j = 1, \dots, n$，那么$A\mathrm{adj}(A)$在第$i$行第$j = i$列上的项(即$A\mathrm{adj}(A)$对角线上的项)为

$$\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix} \begin{bmatrix} C_{i1} \\ \vdots \\ C_{in} \end{bmatrix} = a_{i1}C_{i1} + \dots + a_{in}C_{in} = \det(A);$$

若$i \neq j$，那么$A\mathrm{adj}(A)$在第$i$行第$j$列上的项(即$A\mathrm{adj}(A)$不在对角线上的项)为

$$\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix} \begin{bmatrix} C_{j1} \\ \vdots \\ C_{jn} \end{bmatrix} = a_{i1}C_{j1} + \dots + a_{in}C_{jn},$$

而由Remark 2.6我们知道，当$i \neq j$时，$a_{i1}C_{j1} + \dots + a_{in}C_{jn} = 0$。因此，以上观察告诉我们，$A\mathrm{adj}(A)$对角线上所有的项都为$\det(A)$，不在对角线上的项都为0，因此$A\mathrm{adj}(A)$为对角矩阵，其形状为：

$$A\mathrm{adj}(A) = \begin{bmatrix} \det(A) & 0 & \dots & 0 \\ 0 & \det(A) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \det(A) \end{bmatrix} = \det(A)I_n.$$

现在假设$A$可逆，那么由Theorem 2.13可知此时$\det(A) \neq 0$，因此由$A\mathrm{adj}(A) = \det(A)I_n$得到(等式两边除以$\det(A)$)

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

$\square$

例子：(2022年线性代数期中考试题)

1. Let $A$ be an invertible $4 \times 4$ matrix and suppose that the adjoint matrix of $A$ is
$$\text{adj}(A) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}.$$
Compute $\det(A)$ and $A^{-1}$.
2. Let $A$ be a $4 \times 4$ matrix and suppose that the first row of $A$ is $\begin{bmatrix} 1 & 2 & 0 & 4 \end{bmatrix}$, the third column of adj$(A)$ is $\begin{bmatrix} 6 \\ x \\ 19 \\ 2 \end{bmatrix}$. Compute $x$.

答案：

1. 因为
$$A\text{adj}(A) = \det(A)I_n$$
我们有
$$\det(A\text{adj}(A)) = \det(\det(A)I_n).$$
这意味着：$\det(A)\det(\text{adj}(A)) = \det(A)^4\det(I_n) = \det(A)^4$。由于$A$可逆，我们有$\det(A) \neq 0$，因此我们得到$\det(\text{adj}(A)) = \det(A)^3$。由于$\det(\text{adj}(A)) = 1 \times 1 \times 1 \times 8 = 8$，我们得到$\det(A)^3 = 8$，因此$\det(A) = 2$。再利用$A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$，最终得到
$$A^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$
2. 由题设可知，$A$的第一行为
$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 4 \end{bmatrix},$$
adj$(A)$的第三列为
$$\begin{bmatrix} C_{31} \\ C_{32} \\ C_{33} \\ C_{34} \end{bmatrix} = \begin{bmatrix} 6 \\ x \\ 19 \\ 2 \end{bmatrix}.$$

那么由等式$a_{11}C_{31}+a_{12}C_{32}+a_{13}C_{33}+a_{14}C_{34}=0$并代入$a_{11}=1, a_{12}=2, a_{13}=0, a_{14}=4, C_{31}=6, C_{32}=x.C_{33}=19, C_{34}=2$，得到

$$1\times 6+2x+0\times 19+4\times 2=0$$

得到$2x=-14$，即$x=-7$。

**由以上解题过程我们可以总结出：若在考试/作业中遇到跟伴随矩阵(adjoint matrix)或者代数余子式(cofactor)有关的计算题或证明题，应该首先想到利用以下等式**

- $A^{-1}=\frac{1}{\det(A)}\mathrm{adj}(A)$,
- $A\mathrm{adj}(A)=\det(A)I_n$,
- $$a_{i1}C_{j1}+\ldots+a_{in}C_{jn}=\begin{cases}\det(A), & i=j\\ 0, & i\neq j.\end{cases}$$

## 2.4. 某些特殊行列式的计算

**例子1：范德蒙行列式, Vandermonde determinant**

对任何自然数$n$，定义Vandermonde matrix为

$$V_n=\begin{bmatrix}1 & x_0 & x_0^2 & \ldots & x_0^n\\ 1 & x_1 & x_1^2 & \ldots & x_1^n\\ \vdots & \vdots & \vdots & \ldots & \vdots\\ 1 & x_n & x_n^2 & \ldots & x_n^n\end{bmatrix},$$

注意$V_n$为$n+1$阶方阵。

现在我们求解$V_n$的行列式通项公式。我们首先依次将$V_n$的第1列乘以$-x_0$加到第2列，第2列乘以$-x_0$加到第3列，...，第$n$列乘以$-x_0$加到第$n+1$列，从而得到

$$\begin{vmatrix}1 & x_0 & x_0^2 & \ldots & x_0^n\\ 1 & x_1 & x_1^2 & \ldots & x_1^n\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_n & x_n^2 & \ldots & x_n^n\end{vmatrix}=\begin{vmatrix}1 & 0 & 0 & 0 & \ldots & 0\\ 1 & x_1-x_0 & x_1(x_1-x_0) & x_1^2(x_1-x_0) & \ldots & x_1^{n-1}(x_1-x_0)\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_n-x_0 & x_n(x_n-x_0) & x_n^2(x_n-x_0) & \ldots & x_n^{n-1}(x_n-x_0)\end{vmatrix}.$$

对上面等式右边的矩阵沿第一行做代数余子式展开，得到

$$
\begin{vmatrix}
1 & 0 & 0 & 0 & \dots & 0 \\
1 & x_1 - x_0 & x_1(x_1 - x_0) & x_1^2(x_1 - x_0) & \dots & x_1^{n-1}(x_1 - x_0) \\
\vdots & \vdots & \vdots & \dots & \vdots & \\
1 & x_n - x_0 & x_n(x_n - x_0) & x_n^2(x_n - x_0) & \dots & x_n^{n-1}(x_1 n - x_0)
\end{vmatrix} = 1 \times
$$

$$
\begin{vmatrix}
x_1 - x_0 & x_1(x_1 - x_0) & x_1^2(x_1 - x_0) & \dots & x_1^{n-1}(x_1 - x_0) \\
\vdots & \vdots & \dots & \vdots & \\
x_n - x_0 & x_n(x_n - x_0) & x_n^2(x_n - x_0) & \dots & x_n^{n-1}(x_1 n - x_0)
\end{vmatrix}
$$

$$
= (x_1 - x_0)(x_2 - x_0)\dots(x_n - x_0)
\begin{vmatrix}
1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\
\vdots & \vdots & \vdots & \dots & \vdots \\
1 & x_n & x_n^2 & \dots & x_n^{n-1}
\end{vmatrix}
$$

$$
= (x_1 - x_0)(x_2 - x_0)\dots(x_n - x_0)|V_{n-1}|.
$$

(因为

$$
\begin{bmatrix} x_i - x_0 & x_i(x_i - x_0) & x_i^2(x_i - x_0) & \dots & x_i^{n-1}(x_i - x_0) \end{bmatrix} = (x_i - x_0) \begin{bmatrix} 1 & x_i & x_i^2 & \dots & x_i^{n-1} \end{bmatrix}
$$

对任何$i = 1, \dots, n$成立，因此对矩阵

$$
\begin{vmatrix}
x_1 - x_0 & x_1(x_1 - x_0) & x_1^2(x_1 - x_0) & \dots & x_1^{n-1}(x_1 - x_0) \\
\vdots & \vdots & \dots & \vdots & \\
x_n - x_0 & x_n(x_n - x_0) & x_n^2(x_n - x_0) & \dots & x_n^{n-1}(x_1 n - x_0)
\end{vmatrix}
$$

的每一行依次使用**Theorem 2.5的第1.条内容** 可得以上等式。)
现在我们得到$\det(V_n) = (x_1 - x_0)(x_2 - x_0)\dots(x_n - x_0)\det(V_{n-1})$。这里

$$
V_{n-1} = \begin{bmatrix}
1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\
\vdots & \vdots & \vdots & \dots & \vdots \\
1 & x_n & x_n^2 & \dots & x_n^{n-1}
\end{bmatrix}.
$$

显然我们可以继续对$V_{n-1}$使用上面的计算步骤，得到 $\det(V_{n-1}) = (x_2 - x_1)(x_3 - x_1)\dots(x_n - x_1)\det(V_{n-2})$，这里

$$
V_{n-2} = \begin{bmatrix}
1 & x_2 & x_2^2 & \dots & x_2^{n-2} \\
\vdots & \vdots & \vdots & \dots & \vdots \\
1 & x_n & x_n^2 & \dots & x_n^{n-2}
\end{bmatrix}.
$$

这样重复下去，经过$n$次迭代最终得到通项公式：

$$
\det(V_n) = \prod_{j=1}^{n}(x_j - x_0)\prod_{j=2}^{n}(x_j - x_1)\dots(x_n - x_{n-1}) = \prod_{0 \le i < j \le n}(x_j - x_i).
$$

(这里记号$\prod_{j=1}^{n}$为对下标$j$从1到$n$的连乘符号，即$\prod_{j=1}^{n}(x_j - x_0) = (x_1 - x_0)(x_2 - x_0)\dots(x_n - x_0)$)

**例子2：2021年线性代数期中考试题**

求矩阵

$$D_n = \begin{bmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 2 & 1 & 2 & \dots & n-2 & n-1 \\ 3 & 2 & 1 & \dots & n-3 & n-2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ n-1 & n-2 & n-3 & \dots & 1 & 2 \\ n & n-1 & n-2 & \dots & 2 & 1 \end{bmatrix}$$

的行列式。

**答案：** 首先将$D_n$的第$n-1$行乘$-1$加到第$n$行，...，第2行乘$-1$加到第3行，第1行乘$-1$加到第2行， **(对比我们计算例1范德蒙行列式时的第一步是依次用每一列乘以常数加到与它相邻的另一列上。显然这里我们使用了相似的技巧)**，得到：

$$\begin{vmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 2 & 1 & 2 & \dots & n-2 & n-1 \\ 3 & 2 & 1 & \dots & n-3 & n-2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ n-1 & n-2 & n-3 & \dots & 1 & 2 \\ n & n-1 & n-2 & \dots & 2 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 1 & -1 & -1 & \dots & -1 & -1 \\ 1 & 1 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & -1 & -1 \\ 1 & 1 & 1 & \dots & 1 & -1 \end{vmatrix}.$$

现在我们将上方等式右边的矩阵的最后一列加到第一列上，得到：

$$\begin{vmatrix} 1 & 2 & 3 & \dots & n-1 & n \\ 1 & -1 & -1 & \dots & -1 & -1 \\ 1 & 1 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & -1 & -1 \\ 1 & 1 & 1 & \dots & 1 & -1 \end{vmatrix} = \begin{vmatrix} n+1 & 2 & 3 & \dots & n-1 & n \\ 0 & -1 & -1 & \dots & -1 & -1 \\ 0 & 1 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 1 & \dots & -1 & -1 \\ 0 & 1 & 1 & \dots & 1 & -1 \end{vmatrix},$$

对上方等式右边的矩阵沿第一列做代数余子式展开，得到

$$\begin{vmatrix} n+1 & 2 & 3 & \dots & n-1 & n \\ 0 & -1 & -1 & \dots & -1 & -1 \\ 0 & 1 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 1 & \dots & -1 & -1 \\ 0 & 1 & 1 & \dots & 1 & -1 \end{vmatrix} = (n+1) \times \begin{vmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & -1 & -1 \\ 1 & 1 & \dots & 1 & -1 \end{vmatrix}.$$

令$F_{n-1}$指代上方等式右边的$n-1$阶方阵，即

$$F_{n-1} = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & -1 & -1 \\ 1 & 1 & \dots & 1 & -1 \end{bmatrix},$$

我们目前得到$\det(D_n) = (n+1)\det(F_{n-1})$。接下来我们计算$\det(F_{n-1})$。首先将它的最后一列加到第一列上，得到

$$\begin{vmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & -1 & -1 \\ 1 & 1 & \dots & 1 & -1 \end{vmatrix} = \begin{vmatrix} -2 & -1 & \dots & -1 & -1 \\ 0 & -1 & \dots & -1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \dots & -1 & -1 \\ 0 & 1 & \dots & 1 & -1 \end{vmatrix},$$

对其沿第一列做代数余子式展开，得到

$$\det(F_{n-1}) = (-2) \times \begin{vmatrix} -1 & \dots & -1 & -1 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \dots & -1 & -1 \\ 1 & \dots & 1 & -1 \end{vmatrix}.$$

将上方等式右边的$n-2$阶方阵记为$F_{n-2}$，显然它与$F_{n-1}$的结构一致，只是阶数减一，因此重复以上步骤对$F_{n-2}$并继续迭代，最后得到

$$\det(F_{n-1}) = (-2)\det(F_{n-2}) = \dots = (-2)^{n-2}\det(F_1).$$

由于$F_1 = -1$，以上等式最终等于：$\det(F_{n-1}) = (-2)^{n-2} \times (-1) = (-1)^{n-1}2^{n-2}$。将该表达式代入之前已经得到的等式$\det(D_n) = (n+1)\det(F_{n-1})$里，我们最后得到

$$\det(D_n) = (n+1)\det(F_{n-1}) = (-1)^{n-1}2^{n-2}(n+1).$$

**例子3：2020年线性代数期中考试题**

$$
D_n = \begin{bmatrix}
2 & 1 & 0 & 0 & \dots & 0\\
1 & 2 & 1 & 0 & \dots & 0\\
0 & 1 & 2 & 1 & \dots & 0\\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots\\
0 & 0 & \dots & 1 & 2 & 1\\
0 & 0 & \dots & 0 & 1 & 2
\end{bmatrix}.
$$

求$D_n$的行列式。

**答案：** 首先注意到将$D_n$的第一行第一列删掉后得到的$n-1$阶方阵为$D_{n-1}$:

$$
\begin{bmatrix}
2 & 1 & 0 & 0 & \dots & 0\\
1 & 2 & 1 & 0 & \dots & 0\\
0 & 1 & 2 & 1 & \dots & 0\\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots\\
0 & 0 & \dots & 1 & 2 & 1\\
0 & 0 & \dots & 0 & 1 & 2
\end{bmatrix},
$$

这意味着$D_n$关于第一行第一列的代数余子式为$C_{11} = \det(D_{n-1})$。
将$D_n$的第二行第一列删掉后得到的$n-1$阶方阵为

$$
\begin{bmatrix}
2 & 1 & 0 & 0 & \dots & 0\\
1 & 2 & 1 & 0 & \dots & 0\\
0 & 1 & 2 & 1 & \dots & 0\\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots\\
0 & 0 & \dots & 1 & 2 & 1\\
0 & 0 & \dots & 0 & 1 & 2
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & \dots & 0\\
1 & 2 & 1 & \dots & 0\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
0 & \dots & 1 & 2 & 1\\
0 & \dots & 0 & 1 & 2
\end{bmatrix},
$$

将上方等式右边的矩阵的第一行第一列删掉后得到的$n-2$阶方阵显然为$D_{n-2}$:

$$
\begin{bmatrix}
1 & 0 & 0 & \dots & 0\\
1 & 2 & 1 & \dots & 0\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
0 & \dots & 1 & 2 & 1\\
0 & \dots & 0 & 1 & 2
\end{bmatrix}
= D_{n-2}.
$$

因此，$D_n$关于第二行第一列的代数余子式$C_{21}$满足

$$C_{21}=(-1)^{2+1}\begin{vmatrix}1&0&0&\dots&0\\1&2&1&\dots&0\\\vdots&\ddots&\ddots&\ddots&\vdots\\0&\dots&1&2&1\\0&\dots&0&1&2\end{vmatrix}=-\det(D_{n-2})\quad(\text{沿第一行做代数余子式展开}).$$

综上，对$D_n$的第一列做代数余子式展开，我们得到

$$\det(D_n)=2\times C_{11}+1\times C_{21}=2\det(D_{n-1})-\det(D_{n-2}),$$

即$\det(D_n)-\det(D_{n-1})=\det(D_{n-1})-\det(D_{n-2})$。由以上关系，以及$\det(D_1)=2$，$\det(D_2)=3$，容易求得$\det(D_n)-\det(D_{n-1})=\det(D_{n-1})-\det(D_{n-2})=\dots=\det(D_2)-\det(D_1)=3-2=1$。由此可得数列$\det(D_n)$的通项公式为

$$\begin{aligned}\det(D_n)&=\det(D_n)-\det(D_{n-1})+\det(D_{n-1})-\det(D_{n-2})+\dots+\det(D_2)-\det(D_1)+\det(D_1)\\&=n-1+\det(D_1)\\&=n-1+2=n+1.\end{aligned}$$

**行列式计算中的一些常见技巧(不完全)**：假设我们要求$n$阶矩阵$D_n$的行列式，那么可以试图

- 通过**一系列行变换或者列变换，构造出含有尽可能多0的一行或者一列，然后沿着这一行或列进行代数余子式展开**。这样通常会得到**与$D_n$结构相同但阶数减少的方阵$D_k$**，$k\le n-1$。由此我们可以找到$D_n$的行列式与$D_k$，$k\le n-1$的行列式之间的关系，从而确定$D_n$行列式的通项公式。比如在例子1中我们得到$\det(D_n)=\prod_{j=1}^n(x_j-x_0)\det(D_{n-1})$，在例子3中我们得到$\det(D_n)-\det(D_{n-1})=\det(D_{n-1})-\det(D_{n-2})$。
- 首先计算几个低阶矩阵的行列式，比如$D_1$，$D_2$，$D_3$，猜出$D_n$的通项公式，然后使用数学归纳法。实际上，要使用数学归纳法，我们依然需要找到$D_n$的行列式与$D_k$，$k\le n-1$的行列式之间的关系，比如第三次作业的Problem E。
- 更多具体技巧需要大家通过做题自己总结，或者参考一些课外教材，比如同济大学出版社的线性代数习题集等。

# 3 第三章：欧式空间

### 3.1. $n$维欧式空间， $n$-Euclidean space

本节的内容非常简单。在高中我们知道平面中的向量(vector)$\boldsymbol{v} = (v_1, v_2)$包含有两个坐标$v_1, v_2$分别指示它在$x$轴与$y$轴上的位置；立体空间中的向量$\boldsymbol{v} = (v_1, v_2, v_3)$具有三个坐标$v_1, v_2, v_3$分别代表它在$x$轴，$y$轴与$z$轴上的位置。那么当然我们也可以考虑具有$n$个坐标的向量，这里$n$可以是任意自然数。

**Definition 3.1** (英文教材136页Definition 1)**.** *对任何自然数$n$，一个$n$维向量$\boldsymbol{v}$是一个排列好的$n$元数组(ordered $n$-tuple) $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$，其中每个$v_i$都是实数，$i = 1, \ldots, n$。我们把所有这些$n$维向量的集合叫做$n$**维欧式空间**，并且记作$\mathbb{R}^n$。*

所有对于二维，三维向量成立的运算法则一样可以推广到$n$维向量的情况。比如

- 所有坐标都为0的$n$维零向量$\mathbf{0} = (0, \ldots, 0)$满足$\boldsymbol{v} + \mathbf{0} = \mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$对任何$\boldsymbol{v} \in \mathbb{R}^n$成立。
- 两个$n$维向量$\boldsymbol{v}$，$\boldsymbol{w}$相等当且仅当它们的每个坐标都相等，即$v_1 = w_1$，...，$v_n = w_n$。
- $\boldsymbol{v} + \boldsymbol{w} = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n)$。
- 对任何常数$c$， $c\boldsymbol{v} = (cv_1, cv_2, \ldots, cv_n)$。

其余运算法则请大家仔细阅读英文教材137页至139页的内容。

**这里请大家特别注意，作为向量，我们不需要区分它是行向量还是列向量。在后面的具体运算中，我们有时会把向量当做行向量，有时当做列向量处理，这取决于具体情形。**

下面这个概念非常重要：

**Definition 3.2** (英文教材139页Definition 4)**.** *给定$n$维向量$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$，我们称一个$n$维向量$\boldsymbol{w} \in \mathbb{R}^n$是关于$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$的线性组合(linear combination)，当且仅当存在实数$k_1, k_2, \ldots, k_r$使得*

$$\boldsymbol{w} = k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 + k_r\boldsymbol{v}_r.$$

*此时我们称$k_1, k_2, \ldots, k_r$为$\boldsymbol{w}$关于$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$的系数(coefficient)。*

**例子**：在第一章的学习中我们引入了**标准单位向量**(standard unit vector)的概念，即在$n$维欧式空间$\mathbb{R}^n$中存在$n$个标准单位向量$\boldsymbol{e}_i = (0, \ldots, 0, \underbrace{1}_{\text{第i个位置}}, 0, \ldots, 0)$，$i = 1, \ldots, n$。显然，任何$n$维向量$\boldsymbol{v} \in \mathbb{R}^n$都可以表示为关于$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$的线性组合：

$$\boldsymbol{v} = (v_1, \ldots, v_n) = v_1\boldsymbol{e}_1 + \ldots + v_n\boldsymbol{e}_n,$$

且此时$v_1, \ldots, v_n$是$\boldsymbol{v}$关于$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$的系数。

**例子**: $\boldsymbol{w} = (1, 2)$是关于$\boldsymbol{e}_1 = (1, 0)$，$\boldsymbol{e}_2 = (0, 1)$的线性组合：

$$\boldsymbol{w} = 1\boldsymbol{e}_1 + 2\boldsymbol{e}_2,$$

此时$\boldsymbol{w}$关于$\boldsymbol{e}_1 = (1, 0)$，$\boldsymbol{e}_2 = (0, 1)$的系数分别为1和2。
同时$\boldsymbol{w} = (1, 2)$是关于$\boldsymbol{v}_1 = (1, -1)$，$\boldsymbol{v}_2 = (-1, -1)$的线性组合：

$$\boldsymbol{w} = -\frac{1}{2}\boldsymbol{v}_1 - \frac{3}{2}\boldsymbol{v}_2,$$

此时$\boldsymbol{w}$关于$\boldsymbol{v}_1 = (1, 0)$，$\boldsymbol{v}_2 = (0, 1)$的系数分别为$-1/2$和$-3/2$。
$\boldsymbol{w} = (1, 2)$是关于$\boldsymbol{v}_1 = (1, 0)$，$\boldsymbol{v}_2 = (0, 1)$，$\boldsymbol{v}_3 = (1, 1)$的线性组合，比如

$$\boldsymbol{w} = 1\boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 0\boldsymbol{v}_3$$

或者

$$\boldsymbol{w} = -\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 2\boldsymbol{v}_3.$$

**这意味着一个向量($\boldsymbol{w}$)关于同一组其他向量($\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$)可能有多个不同的线性组合表示。**

### 3.2. 欧式空间中的范数/内积/距离, Norm, Dot product, Distance in Euclidean space

**Definition 3.3** (英文教材142页Definition 1). *对于$n$维向量$\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$，它的**欧式范数(Euclidean norm)**或者**欧式几何长度(length in Euclidean space)**记为$\|\boldsymbol{v}\|$，并被表示为*

$$\|\boldsymbol{v}\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

显然，当$n = 2$或$n = 3$时，对于$\boldsymbol{v} = (v_1, v_2)$或$\boldsymbol{v} = (v_1, v_2, v_3)$，它的范数$\|\boldsymbol{v}\|$给出的是坐标为$(v_1, v_2)$或坐标为$(v_1, v_2, v_3)$的点距离原点(即零向量)**0**的距离。在$n$维欧式空间里向量$\boldsymbol{v}$的范数$\|\boldsymbol{v}\|$同样反映了坐标为$(v_1, \ldots, v_n)$的点距离原点的距离。

以下定理非常容易证明，请阅读英文教材143页Theorem 3.2.1.

**Theorem 3.4.** 对任何$\boldsymbol{v} \in \mathbb{R}^n$，我们有

1. $\|\boldsymbol{v}\| \geq 0$;
2. $\|\boldsymbol{v}\| = 0$当且仅当 *(if and only if)*$\boldsymbol{v} = 0$;
3. $\|c\boldsymbol{v}\| = |c|\|\boldsymbol{v}\|$ 对任何常数$c$。

我们称任何范数为1的向量为单位向量(unit vector)。显然，对任何**非零向量**$\boldsymbol{v} \in \mathbb{R}^n$，通过除以它自己的范数，我们可以得到一个一个与$\boldsymbol{v}$方向相同但范数为1的单位向量：

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}.$$

容易验证$\|\boldsymbol{u}\| = \|\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\| = \frac{1}{\|\boldsymbol{v}\|}\|\boldsymbol{v}\| = 1$。以上过程被称为$\boldsymbol{v}$的**归一化(normalizing $\boldsymbol{v}$)**。

利用范数我们可以定义两个向量之间的距离(distance)。

**Definition 3.5** (英文教材145页Definition 2)**.** 对于$\boldsymbol{v} \in \mathbb{R}^n$，$\boldsymbol{w} \in \mathbb{R}^n$，它们之间的欧式几何距离，简称距离，记为$d(\boldsymbol{v}, \boldsymbol{w})$，定义为

$$d(\boldsymbol{v}, \boldsymbol{w}) = \|\boldsymbol{v} - \boldsymbol{w}\| = \sqrt{(v_1 - w_1)^2 + \ldots + (v_n - w_n)^2}.$$

另一个非常重要的概念是欧式内积，又称为点乘。

**Definition 3.6** (英文教材147页Definition 4)**.** 向量$\boldsymbol{v} \in \mathbb{R}^n$与$\boldsymbol{w} \in \mathbb{R}^n$之间的**欧式内积*(Euclidean inner product)***，又称为**点乘积*(dot product)***，记为$\boldsymbol{v} \cdot \boldsymbol{w}$，定义为

$$\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n.$$

由欧式内积与欧式范数的定义，容易看出对任何$\boldsymbol{v} \in \mathbb{R}^n$，有

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}.$$

除此之外，利用定义我们很容易验证欧式内积的以下性质，大家也可以查询英文教材Theorem 3.2.2与Theorem 3.2.3.

**Theorem 3.7.** 对任何$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$，我们有

1. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$ (对称性，*symmetry property*);
2. $(c\boldsymbol{u} + k\boldsymbol{v}) \cdot \boldsymbol{w} = c(\boldsymbol{u} \cdot \boldsymbol{w}) + k(\boldsymbol{v} \cdot \boldsymbol{w})$ 对任何常数$c, k$ (第一个变量上的线性);
3. $\boldsymbol{u} \cdot (c\boldsymbol{v} + k\boldsymbol{w}) = c(\boldsymbol{u} \cdot \boldsymbol{v}) + k(\boldsymbol{u} \cdot \boldsymbol{w})$ 对任何常数$c, k$ (第二个变量上的线性);
4. $\boldsymbol{v} \cdot \boldsymbol{v} \geq 0$，$\boldsymbol{v} \cdot \boldsymbol{v} = 0$当且仅当$\boldsymbol{v} = \mathbf{0}$ (正定性，*positivity properyt*)。

**Remark 3.8** (重要!). 从内积的定义与矩阵乘法的定义可知，对于向量$\boldsymbol{v} \in \mathbb{R}^n$，$\boldsymbol{w} \in \mathbb{R}^n$，它们之间的欧式内积等于**将$\boldsymbol{v}$视为$1 \times n$-矩阵，即行向量，以矩阵乘法的形式乘上列向量$\boldsymbol{w}$，即将$\boldsymbol{w}$视为$n \times 1$-矩阵**:

$$\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + \ldots + v_n w_n = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}.$$

因此，假如$\boldsymbol{v}$为一个$n \times 1$-矩阵，$\boldsymbol{w}$也为为一个$n \times 1$-矩阵，那么这两个列向量之间的内积等于$\boldsymbol{v}$的转置(变为行向量)以矩阵乘法的形式乘以列向量$\boldsymbol{w}$:

$$\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + \ldots + v_n w_n = \boldsymbol{v}^\top \boldsymbol{w}.$$

由于内积的对称性，我们还有

$$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v} = \boldsymbol{w}^\top \boldsymbol{v}.$$

特别地，假设$A$为$n$阶方阵，$\boldsymbol{v}$，$\boldsymbol{w}$均为$n \times 1$-矩阵，即列向量，那么$A\boldsymbol{v}$为$n \times 1$-矩阵(列向量)，此时有

$$A\boldsymbol{v} \cdot \boldsymbol{w} = \underbrace{(A\boldsymbol{v})^\top \boldsymbol{w}}_{\text{矩阵乘法}} = \underbrace{\boldsymbol{v}^\top (A^\top \boldsymbol{w})}_{\text{矩阵乘法}} = \boldsymbol{v} \cdot A^\top \boldsymbol{w}.$$

类似可得

$$\boldsymbol{v} \cdot A\boldsymbol{w} = \underbrace{\boldsymbol{v}^\top (A\boldsymbol{w})}_{\text{矩阵乘法}} = \underbrace{(\boldsymbol{v}^\top A)\boldsymbol{w}}_{\text{矩阵乘法}} = \underbrace{(A^\top \boldsymbol{v})^\top \boldsymbol{w}}_{\text{矩阵乘法}} = A^\top \boldsymbol{v} \cdot \boldsymbol{w}.$$

更多相关内容请仔细阅读英文教材*151*页-*153*页。

以下是欧式内积满足的一个非常重要的性质，它的证明将在后续学习中正式给出，目前我们只需牢记定理内容。

**Theorem 3.9** (英文教材Theorem 3.2.4). 对于任何$\boldsymbol{u} \in \mathbb{R}^n$，$\boldsymbol{v} \in \mathbb{R}^n$，有以下***Cauchy–Schwarz*不等式*(Cauchy-Schwarz inequality)***:

$$|\boldsymbol{u} \cdot \boldsymbol{v}| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|.$$

利用以上Cauchy–Schwarz不等式，我们可以证明以下重要性质：

**Theorem 3.10** (英文教材Theorem 3.2.5). 对任何$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$，我们有以下**三角不等式*(triangle inequality)***:

1. $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$;

2. $d(\boldsymbol{u},\boldsymbol{v}) \le d(\boldsymbol{u},\boldsymbol{w}) + d(\boldsymbol{w},\boldsymbol{v})$.

证明.

$$\begin{aligned}
\|\boldsymbol{u}+\boldsymbol{v}\|^2 &= (\boldsymbol{u}+\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v}) \\
&= (\boldsymbol{u}\cdot\boldsymbol{u}) + 2(\boldsymbol{u}\cdot\boldsymbol{v}) + (\boldsymbol{v}\cdot\boldsymbol{v}) \\
&= \|\boldsymbol{u}\|^2 + 2(\boldsymbol{u}\cdot\boldsymbol{v}) + \|\boldsymbol{v}\|^2 \\
&\le \|\boldsymbol{u}\|^2 + 2\|\boldsymbol{u}\|\|\boldsymbol{v}\| + \|\boldsymbol{v}\|^2 \qquad \text{由Cauchy–Schwarz不等式} \\
&= (\|\boldsymbol{u}\| + \|\boldsymbol{v}\|)^2.
\end{aligned}$$

两边开根号即可得到第一个三角不等式。此外，利用第一个不等式，可以推出第二个：

$$\begin{aligned}
d(\boldsymbol{u},\boldsymbol{v}) = \|\boldsymbol{u}-\boldsymbol{v}\| &= \|\boldsymbol{u}-\boldsymbol{w}+(\boldsymbol{w}-\boldsymbol{v})\| \\
&\le \|\boldsymbol{u}-\boldsymbol{w}\| + \|\boldsymbol{w}-\boldsymbol{v}\| \\
&= d(\boldsymbol{u},\boldsymbol{w}) + d(\boldsymbol{w},\boldsymbol{v}).
\end{aligned}$$

□

**Theorem 3.11** (英文教材Theorem 3.2.6, 3.2.7)**.** 对任何$\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$，有

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 + \|\boldsymbol{u}-\boldsymbol{v}\|^2 = 2(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2),$$
$$\boldsymbol{u}\cdot\boldsymbol{v} = \frac{1}{4}\|\boldsymbol{u}+\boldsymbol{v}\|^2 - \frac{1}{4}\|\boldsymbol{u}-\boldsymbol{v}\|^2.$$

证明. 利用内积的运算法则，很容易得到

$$\begin{aligned}
\|\boldsymbol{u}+\boldsymbol{v}\|^2 + \|\boldsymbol{u}-\boldsymbol{v}\|^2 &= (\boldsymbol{u}+\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v}) + (\boldsymbol{u}-\boldsymbol{v})\cdot(\boldsymbol{u}-\boldsymbol{v}) \\
&= 2(\boldsymbol{u}\cdot\boldsymbol{v}) + 2(\boldsymbol{v}\cdot\boldsymbol{v}) \\
&= 2(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2).
\end{aligned}$$

第二个等式类似可证，请参考英文教材Theorem 3.2.7的证明。 □

对任何两个**非零**的向量$\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$，它们之间的**夹角(angle)**定义为

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

或者说$\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$。显然，由Cauchy–Schwarz不等式可得

$$\left|\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right| \le 1,$$

因此以上定义有意义。比如，向量$\boldsymbol{u}=(1,1)$与$\boldsymbol{v}=(1,-1)$之间的夹角$\theta = \arccos(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}) = \arccos(\frac{1\times1+1\times(-1)}{\sqrt{2}\sqrt{2}}) = \arccos(0) = \frac{\pi}{2}$，即两者互相垂直，符合直观图像。

### 3.3. 正交性, Orthogonality

**Definition 3.12** (英文教材155页Definition 1). 两个$n$维向量$\boldsymbol{u}$，$\boldsymbol{v}$正交*(orthogonal)*或者垂直*(perpendicular)*当且仅当$\boldsymbol{u}\cdot\boldsymbol{v}=0$(注意此时它们之间的夹角$\theta=\arccos(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|})=\arccos(0)=\frac{\pi}{2}$)。

显然$\mathbb{R}^n$里的$n$个标准单位向量$\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$互相正交。
以下定理是**考点**。

**Theorem 3.13** (正交投影定理, Projection Theorem). 设$\boldsymbol{u}$与$\boldsymbol{v}$为$n$维向量，且$\boldsymbol{v}\neq\boldsymbol{0}$。那么$\boldsymbol{u}$可以被唯一表示为$\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2$，$\boldsymbol{w}_1=c\boldsymbol{v}$ ($c$是一个常数)，$\boldsymbol{w}_2$与$\boldsymbol{v}$正交($\boldsymbol{w}_2\cdot\boldsymbol{v}=0$)。

证明. 我们先假设存在这样的分解$\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2$，$\boldsymbol{w}_1=c\boldsymbol{v}$，$\boldsymbol{w}_2$与$\boldsymbol{v}$正交。那么必然有

$$\begin{aligned}\boldsymbol{u}\cdot\boldsymbol{v}&=(\boldsymbol{w}_1+\boldsymbol{w}_2)\cdot\boldsymbol{v}\\&=\boldsymbol{w}_1\cdot\boldsymbol{v}+\boldsymbol{w}_2\cdot\boldsymbol{v}\\&=c(\boldsymbol{v}\cdot\boldsymbol{v})+0\quad(\boldsymbol{w}_1=c\boldsymbol{v},\boldsymbol{w}_2\cdot\boldsymbol{v}=0)\\&=c\|\boldsymbol{v}\|^2.\end{aligned}$$

由于已假设$\boldsymbol{v}\neq\boldsymbol{0}$，它的范数$\|\boldsymbol{v}\|$不为0，因此以上等式告诉我们

$$c=\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}.$$

这意味着$\boldsymbol{w}_1=\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}$，且因此有

$$\boldsymbol{w}_2=\boldsymbol{u}-\boldsymbol{w}_1=\boldsymbol{u}-\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}.$$

实际上到此为止我们已经证明了这样分解的存在性。要证明唯一性，假设存在两个分解$\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2$，$\boldsymbol{w}_1=c\boldsymbol{v}$，$\boldsymbol{w}_2$与$\boldsymbol{v}$正交；$\boldsymbol{u}=\boldsymbol{w}_1'+\boldsymbol{w}_2'$，$\boldsymbol{w}_1'=c'\boldsymbol{v}$，$\boldsymbol{w}_2'$与$\boldsymbol{v}$正交。那么必然有

$$\begin{aligned}\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2=\boldsymbol{w}_1'+\boldsymbol{w}_2'=c\boldsymbol{v}+\boldsymbol{w}_2&=c'\boldsymbol{v}+\boldsymbol{w}_2'\\&\Rightarrow(c-c')\boldsymbol{v}=\boldsymbol{w}_2'-\boldsymbol{w}_2\\&\Rightarrow((c-c')\boldsymbol{v})\cdot\boldsymbol{v}=(\boldsymbol{w}_2'-\boldsymbol{w}_2)\cdot\boldsymbol{v}\\&\Rightarrow(c-c')\|\boldsymbol{v}\|^2=\underbrace{\boldsymbol{w}_2'\cdot\boldsymbol{v}}_{=0}-\underbrace{\boldsymbol{w}_2\cdot\boldsymbol{v}}_{=0}\\&\Rightarrow(c-c')\|\boldsymbol{v}\|^2=0\\&\Rightarrow c-c'=0\quad(\|\boldsymbol{v}\|\neq0)\\&\Rightarrow c=c'.\end{aligned}$$

因此我们得到$\boldsymbol{w}_1 = c\boldsymbol{v} = c'\boldsymbol{v} = \boldsymbol{w}_2$，从而有$\boldsymbol{w}_2 = \boldsymbol{u} - \boldsymbol{w}_1 = \boldsymbol{u} - \boldsymbol{w}_2' = \boldsymbol{w}_2'$。这证明了分解的唯一性。 □

- 在以上定理的分解中，我们称$\boldsymbol{w}_1$为"$\boldsymbol{u}$在$\boldsymbol{v}$上的正交投影"(英文：orthogonal projection of $\boldsymbol{u}$ on $\boldsymbol{v}$，或者 the vector component of $\boldsymbol{u}$ along $\boldsymbol{v}$)，并且记为$\mathrm{proj}_{\boldsymbol{v}}(\boldsymbol{u})$。显然以上计算告诉我们
$$\mathrm{proj}_{\boldsymbol{v}}(\boldsymbol{u}) = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}.$$
- 我们称$\boldsymbol{w}_2$为"the vector component of $\boldsymbol{u}$ orthogonal to $\boldsymbol{v}$"，由以上计算可得
$$\boldsymbol{w}_2 = \boldsymbol{v} - \mathrm{proj}_{\boldsymbol{v}}(\boldsymbol{u}) = \boldsymbol{u} - \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}.$$

**例子：** (2022年线性代数期中考试题)
Let $\boldsymbol{u} = (3,-2,1)$, $\boldsymbol{v} = (-1,k,-5)$. If $\|\mathrm{proj}_{\boldsymbol{u}}(\boldsymbol{v})\| = \sqrt{14}$, then $k = ?$
**答案：** 套入以上正交投影公式，首先得到
$$\mathrm{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|^2}\boldsymbol{u},$$
因此$\|\mathrm{proj}_{\boldsymbol{u}}(\boldsymbol{v})\| = \|\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|^2}\boldsymbol{u}\| = |\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|^2}|\|\boldsymbol{u}\| = \frac{|\boldsymbol{u}\cdot\boldsymbol{v}|}{\|\boldsymbol{u}\|}$。现在已知$\|\mathrm{proj}_{\boldsymbol{u}}(\boldsymbol{v})\| = \sqrt{14}$，那么我们实际上得到
$$\frac{|\boldsymbol{u}\cdot\boldsymbol{v}|}{\|\boldsymbol{u}\|} = \sqrt{14}.$$
容易计算得到$\|\boldsymbol{u}\| = \sqrt{14}$，且$\boldsymbol{u}\cdot\boldsymbol{v} = -2k-8$，因此以上等式告诉我们
$$\frac{|\boldsymbol{u}\cdot\boldsymbol{v}|}{\|\boldsymbol{u}\|} = \frac{|-2k-8|}{\sqrt{14}} = \sqrt{14}$$
从而可以算出
$$|2k+8| = 14$$
从而得到$k=3$或者$k=-11$。
**注意：** 此题为填空题，因此需要注意计算的准确性。在具体计算时首先要明确这里要求的是$\boldsymbol{v}$在$\boldsymbol{u}$上的正交投影，此外还要注意这里的$k$有两种取值可能，只计算出一个答案的话会扣分。

以下定理可以被视为$\mathbb{R}^n$空间里的勾股定理。

**Theorem 3.14** (英文教材Theorem 3.3.3). *对任何$\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$，如果$\boldsymbol{u}$与$\boldsymbol{v}$正交，那么*
$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2.$$

证明. 由于$\boldsymbol{u}$与$\boldsymbol{v}$正交，即$\boldsymbol{u}\cdot\boldsymbol{v}=0$，我们容易得到

$$\begin{aligned}\|\boldsymbol{u}+\boldsymbol{v}\|^2&=(\boldsymbol{u}+\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v})\\&=(\boldsymbol{u}\cdot\boldsymbol{u})+2\underbrace{(\boldsymbol{u}\cdot\boldsymbol{v})}_{=0}+(\boldsymbol{v}\cdot\boldsymbol{v})\\&=\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2.\end{aligned}$$

□

假设$S=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_r\}$为$n$维**非零向量**$\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_r$的集合(set)，那么我们称$S$为一个**正交集合(orthogonal set)**，如果它里面包含的所有向量全部都互相正交：$\boldsymbol{v}_i\cdot\boldsymbol{v}_j=0$对任何$i\neq j$，$i,j=1,\ldots,r$都成立。显然，运用以上勾股定理，我们立刻得到，对于任何正交集合$S=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_r\}$，都有

$$\|\boldsymbol{v}_1+\boldsymbol{v}_2+\ldots+\boldsymbol{v}_r\|^2=\|\boldsymbol{v}_1\|^2+\|\boldsymbol{v}_2\|^2+\ldots+\|\boldsymbol{v}_r\|^2.$$

具体证明留给大家验证。

**例子：** (2022年线性代数期中考试题)

Let $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^n$ be three vectors such that $S=\{\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\}$ is an orthogonal set. If $\|\boldsymbol{u}\|=1$, $\|\boldsymbol{v}\|=2$, $\|\boldsymbol{w}\|=3$, then $\|\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}\|=?$

**答案：** 利用勾股定理，

$$\|\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}\|^2=\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2+\|\boldsymbol{w}\|^2=1^2+2^2+3^2=14,$$

因此答案为$\|\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}\|=\sqrt{14}$。

### 3.4. 线性方程组的几何, The geometry of linear systems

考虑齐次线性方程组：

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

将其系数矩阵(coefficient matrix)记为$A$，即

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

显然$A$的行向量表示为

$$\begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_m \end{bmatrix}$$

$\boldsymbol{r}_i = \begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{bmatrix}$， $i = 1, \ldots, m$。
显然，一个$n$维向量$\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$是这个其次线性方程组的一个解 (即，代入$x_1 = s_1$， $x_2 = s_2$ ，...， $x_n = s_n$后，以上方程组成立) 当且仅当

$$\boldsymbol{r}_i \cdot \boldsymbol{s} = 0$$

对所有$i = 1, \ldots, m$成立，即$\boldsymbol{s}$同时正交于系数矩阵$A$的所有行向量$\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m$。
以下定理的证明已经在Theorem 1.23里提及，这里我们再次复习一下这个证明的核心思路。**注意以下定理在整个线性代数课程中都是非常重要的。**

**Theorem 3.15** (英文教材Theorem 3.4.4)**.** 设$A$为$m \times n$-矩阵。$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$为$m \times 1$-列向量。假设线性方程组$A\boldsymbol{x} = \boldsymbol{b}$，即

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

有解。假设$\boldsymbol{s} = (s_1, s_2, \ldots, s_n) \in \mathbb{R}^n$ 是该方程组的**一个特解** (即$A\boldsymbol{s} = \boldsymbol{b}$对于列向量$\boldsymbol{s}$与列向量$\boldsymbol{b}$成立)。那么，该方程组的**任何解**$\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$可以被表示为

$$\boldsymbol{v} = \boldsymbol{s} + \boldsymbol{w}$$

这里$\boldsymbol{w}$**是齐次线性方程组**$A\boldsymbol{x} = \boldsymbol{0}$**的一个解**。反之，若$\boldsymbol{w}$是齐次线性方程组$A\boldsymbol{x} = \boldsymbol{0}$的一个解，那么$\boldsymbol{s} + \boldsymbol{w}$是方程组$A\boldsymbol{x} = \boldsymbol{b}$的一个解。

*证明.* 假设$\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$为$A\boldsymbol{x} = \boldsymbol{b}$的一个解，即$A\boldsymbol{v} = \boldsymbol{b}$成立(显然这里我们将$\boldsymbol{v}$视为列向量)。由于$\boldsymbol{s}$是$A\boldsymbol{x} = \boldsymbol{b}$的一个特解，我们当然也有$A\boldsymbol{s} = \boldsymbol{b}$。令$\boldsymbol{w} = \boldsymbol{v} - \boldsymbol{s}$，那么必然有

$$A\boldsymbol{w} = A(\boldsymbol{v} - \boldsymbol{s}) = A\boldsymbol{v} - A\boldsymbol{s} = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0},$$

即$\boldsymbol{w}=\boldsymbol{v}-\boldsymbol{s}$是齐次线性方程组$A\boldsymbol{x}=\boldsymbol{0}$的一个解。

假设$\boldsymbol{w}$为齐次线性方程组$A\boldsymbol{x}=\boldsymbol{0}$的一个解，即$A\boldsymbol{w}=\boldsymbol{0}$。那么由于$A\boldsymbol{s}=\boldsymbol{b}$，我们当然有

$$A(\boldsymbol{s}+\boldsymbol{w})=A\boldsymbol{s}+A\boldsymbol{w}=\boldsymbol{b}+\boldsymbol{0}=\boldsymbol{b}.$$

因此$\boldsymbol{s}+\boldsymbol{w}$是方程组$A\boldsymbol{x}=\boldsymbol{b}$的一个解。 □

### 3.5. 叉乘, Cross product

除了点乘(内积)之外，对于$\mathbb{R}^3$中的向量我们还有叉乘(cross product)这个概念。但注意，点乘可以对任何$\mathbb{R}^n$里的向量定义，但**叉乘只能对三维空间$\mathbb{R}^3$里的向量定义**。

**Definition 3.16** (英文教材172页Definition 1)**.** 对于$\boldsymbol{u}=(u_1,u_2,u_3)$，$\boldsymbol{v}=(v_1,v_2,v_3)$，它们之间的叉乘$\boldsymbol{u}\times\boldsymbol{v}$定义为

$$\boldsymbol{u}\times\boldsymbol{v}=(u_2v_3-u_3v_2,u_3v_1-u_1v_3,u_1v_2-u_2v_1),$$

也可以表示为

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

注意：一个实用的记住叉乘定义的方法如下：

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

通常在考试中会以填空题的形式考查叉乘的计算。

**例子：** (2022年线性代数期中考试题)

Let $\boldsymbol{u}=(3,-2,1)$, $\boldsymbol{w}=(1,0,-1)$. $\boldsymbol{u}\times\boldsymbol{w}=?$

**答案：**

$$\boldsymbol{u}\times\boldsymbol{w}=\left(\begin{vmatrix}3&-2&1\\1&0&-1\end{vmatrix},-\begin{vmatrix}3&-2&1\\1&0&-1\end{vmatrix},\begin{vmatrix}3&-2&1\\1&0&-1\end{vmatrix}\right)=(2,4,2).$$

以下观察告诉我们当叉乘遇上点乘时，可以得到行列式：

**Definition 3.17** (英文教材177页 Definition 2)**.** 假设$\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^3$，我们称

$$\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})$$

为$\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$的标量三重积***(scalar triple product)***。

利用叉乘，点乘，以及行列式的定义，我们立刻可以验证

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

用语言描述的话，我们得到，$\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$的标量三重积等于以$\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$依次作为第一行，第二行，第三行的矩阵的行列式。

关于点乘与叉乘之间的其他关系请大家仔细阅读**英文教材173页Theorem 3.5.1**的内容与证明。这里我们简单提及一下其中比较重要的部分：

**Theorem 3.18** (英文教材Theorem 3.5.1)**.** 对任何$\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^3$，有

1. $\boldsymbol{u}\cdot(\boldsymbol{u}\times\boldsymbol{v})=0$，即$\boldsymbol{u}$正交于$\boldsymbol{u}\times\boldsymbol{v}$ *($\boldsymbol{u}$ is orthogonal to $\boldsymbol{u}\times\boldsymbol{v}$)* 对任何$\boldsymbol{v}\in\mathbb{R}^3$成立；
2. $\boldsymbol{v}\cdot(\boldsymbol{u}\times\boldsymbol{v})=0$，即$\boldsymbol{v}$正交于$\boldsymbol{u}\times\boldsymbol{v}$ *($\boldsymbol{v}$ is orthogonal to $\boldsymbol{u}\times\boldsymbol{v}$)* 对任何$\boldsymbol{u}\in\mathbb{R}^3$成立；
3. *(Lagrange's identity)* $\|\boldsymbol{u}\times\boldsymbol{v}\|^2=\|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2-(\boldsymbol{u}\cdot\boldsymbol{v})^2$.

证明. 1. 由于$\boldsymbol{u}\cdot(\boldsymbol{u}\times\boldsymbol{v}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$ (见以上关于标量三重积的讨论) 且我们知道若一个方阵有两行相同，那么该方阵的行列式为零，我们当然得到$\boldsymbol{u}\cdot(\boldsymbol{u}\times\boldsymbol{v})=0$.

2. 同上面的证明类似，我们有$\boldsymbol{v}\cdot(\boldsymbol{u}\times\boldsymbol{v}) = \begin{vmatrix} v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = 0$.

3. 直接计算可得，具体细节请自己验证。

□

在本节最后我们简单讨论一下行列式的几何意义。以下定理不作为考试要求，仅供感兴趣的同学了解。

**Theorem 3.19** (英文教材Theorem 3.5.4)**.** *1.* 设$P_1\in\mathbb{R}^2$为平面上一个点，坐标为$(u_1,u_2)$，$P_2\in\mathbb{R}^2$为平面上的另一个点，坐标为$(v_1,v_2)$，令$O$为平面$R^2$的原点，即$O=$

$(0,0)$。那么$\left|\det\left(\begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}\right)\right|$等于由向量$\boldsymbol{u} = \overrightarrow{OP_1} = (u_1, u_2)$与向量$\boldsymbol{v} = \overrightarrow{OP_2} = (v_1, v_2)$作为相邻边所围成的**平行四边形(*parallelogram*)**的面积。

2. 设$P_1 \in \mathbb{R}^3$为$\mathbb{R}^3$中的一个点，坐标为$(u_1, u_2, u_3)$，$P_2 \in \mathbb{R}^3$为另一个点，坐标为$(v_1, v_2, v_3)$，$P_3 \in \mathbb{R}^3$为第三个点，坐标为$(w_1, w_2, w_3)$，令$O$为$R^3$的原点，即$O = (0,0,0)$。那么$\left|\det\left(\begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}\right)\right|$等于由向量$\boldsymbol{u} = \overrightarrow{OP_1} = (u_1, u_2, u_3)$，向量$\boldsymbol{v} = \overrightarrow{OP_2} = (v_1, v_2, v_3)$与向量$\boldsymbol{w} = \overrightarrow{OP_3} = (w_1, w_2, w_3)$作为相邻边所围成的**平行六面体(*parallelepiped*)**的体积。

证明. 1. 首先我们通过对二维向量增加第三个坐标且令该坐标为0的方式将$\mathbb{R}^2$嵌入$\mathbb{R}^3$，即我们将$\boldsymbol{u} = (u_1, u_2)$等同于$\boldsymbol{u} = (u_1, u_2, 0)$，将$\boldsymbol{v} = (v_1, v_2)$等同于$\boldsymbol{v} = (v_1, v_2, 0)$。那么由Lagrange identity，即以上定理Theorem 3.19-(3)，我们有

$$\|\boldsymbol{u} \times \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2 - (\boldsymbol{u} \cdot \boldsymbol{v})^2.$$

这时代入关系式$\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos\theta$ ($\theta$为$\boldsymbol{u}$与$\boldsymbol{v}$之间的夹角) 我们得到

$$\begin{aligned}
\|\boldsymbol{u} \times \boldsymbol{v}\|^2 &= \|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2 - (\boldsymbol{u} \cdot \boldsymbol{v})^2 \\
&= \|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2 - \|\boldsymbol{u}\|\|\boldsymbol{v}\| \cos^2\theta \\
&= \|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2 (1 - \cos^2\theta) \\
&= \|\boldsymbol{u}\|^2\|\boldsymbol{v}\|^2 \sin^2\theta.
\end{aligned}$$

这里我们要求两个向量的夹角取值在0到$\pi$之间，因此$\sin\theta \geq 0$，因此以上等式可开根号并得到

$$\|\boldsymbol{u} \times \boldsymbol{v}\| = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \sin\theta.$$

显然，$\|\boldsymbol{v}\| \sin\theta$给出的是$P_2$这个点往向量$\boldsymbol{u} = \overrightarrow{OP_1}$做垂线所得到的这条垂线的长度，即，以$\boldsymbol{u}$作为平行四边形底边时，这个平行四边形的高。显然$\|\boldsymbol{u}\|$给出的是底边$\boldsymbol{u}$的长度，因此$\|\boldsymbol{v}\| \sin\theta$与$\|\boldsymbol{u}\|$相乘得到这个平行四边形的面积，即

$$\|\boldsymbol{u} \times \boldsymbol{v}\| = \|\boldsymbol{u}\|\|\boldsymbol{v}\| \sin\theta = \text{由向量 } \boldsymbol{u} \text{ 与向量 } \boldsymbol{v} \text{ 作为相邻边所围成的平行四边形的面积}.$$

另一方面，经过简单计算可得

$$\boldsymbol{u} \times \boldsymbol{v} = (u_1, u_2, 0) \times (v_1, v_2, 0) = \det\left(\begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}\right)(0,0,1).$$

因此又有$\|\boldsymbol{u}\times\boldsymbol{v}\|=\left\|\det\left(\begin{bmatrix}u_1 & u_2\\ v_1 & v_2\end{bmatrix}\right)(0,0,1)\right\|=\left|\det\left(\begin{bmatrix}u_1 & u_2\\ v_1 & v_2\end{bmatrix}\right)\right|$。因此平行四边形的面积等于$\left|\det\left(\begin{bmatrix}u_1 & u_2\\ v_1 & v_2\end{bmatrix}\right)\right|$。

2. 我们将以$\boldsymbol{v}$与$\boldsymbol{w}$作为相临边所围成的平行四边形作为平行六面体的底。那么，由以上结果，可知$\|\boldsymbol{v}\times\boldsymbol{w}\|$给出了这个平行六面体的底面积。另一方面，由于$\boldsymbol{v}\cdot(\boldsymbol{v}\times\boldsymbol{w})=0$，$\boldsymbol{w}\cdot(\boldsymbol{v}\times\boldsymbol{w})=0$，见Theorem 3.19-(1),(2)，我们知道$\boldsymbol{v}\times\boldsymbol{w}$垂直于由$\boldsymbol{v}$与$\boldsymbol{w}$组成的底面。那么显然第三条边$\boldsymbol{u}$在$\boldsymbol{v}\times\boldsymbol{w}$上的正交投影的长度即是该平行六面体的高。简单计算可得平行六面体的高为
$$\begin{aligned}\|\mathrm{proj}_{\boldsymbol{v}\times\boldsymbol{w}}\boldsymbol{u}\| &= \left\|\frac{\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})}{\|\boldsymbol{v}\times\boldsymbol{w}\|^2}\boldsymbol{v}\times\boldsymbol{w}\right\|\\ &= \frac{|\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})|}{\|\boldsymbol{v}\times\boldsymbol{w}\|^2}\|\boldsymbol{v}\times\boldsymbol{w}\|\\ &= \frac{|\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})|}{\|\boldsymbol{v}\times\boldsymbol{w}\|}.\end{aligned}$$
那么最终得到平行六面体的体积等于底面积乘以高：
$$\|\mathrm{proj}_{\boldsymbol{v}\times\boldsymbol{w}}\boldsymbol{u}\|\|\boldsymbol{v}\times\boldsymbol{w}\|=|\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})|.$$
此时利用标量三重积的性质，有$|\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})|=\left|\det\left(\begin{bmatrix}u_1 & u_2 & u_3\\ v_1 & v_2 & v_3\\ w_1 & w_2 & w_3\end{bmatrix}\right)\right|$。

□

# 4 第四章：向量空间

## 4.1. 向量空间的定义与重要例子, Real Vector Spaces: Definition and Examples

向量空间是线性代数这门学科研究的核心对象之一。它的定义如下：

**Definition 4.1** (英文教材184页Definition 1). 假设一个**集合(*set*)**$V$上装配有两种**运算(*operations*)**，分别是**加法(*addition*)**与**标量积(*scalar multiplication*)**。我们将加法符号记为“+”，对于$V$里的元素$\boldsymbol{u},\boldsymbol{v}$，将它们由加法运算所连接起来的和(*sum*)记为$\boldsymbol{u}+\boldsymbol{v}$；对于**实数**$c\in\mathbb{R}$，$\boldsymbol{v}\in V$，$c$与$\boldsymbol{v}$的标量积记为$c\boldsymbol{v}$。我们称$V$为一个**向量空间(*vector space*)**，若加法和标量积这两个运算满足以下条件：

1. 对于$V$里的任何元素$\boldsymbol{u}\in V,\boldsymbol{v}\in V$，它们由加法运算所连接起来的和也是$V$里的一个元素，即$\boldsymbol{u}+\boldsymbol{v}\in V$。这个性质在数学上叫做$V$**关于加法的封闭性(*closure under addition*)**。

2. 对于$V$里的任何元素$\boldsymbol{v} \in V$，任何实数$c \in \mathbb{R}$，$c$与$\boldsymbol{v}$的标量积也是$V$里的一个元素，即$c\boldsymbol{v} \in V$。这个性质在数学上叫做**$V$关于标量积的封闭性*(closure under scalar multiplication)***。
3. 对于$V$里的任何元素$\boldsymbol{u} \in V, \boldsymbol{v} \in V$，有$\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{v}+\boldsymbol{u}$，即**加法运算满足交换律**。
4. 对于$V$里的任何元素$\boldsymbol{u} \in V, \boldsymbol{v} \in V, \boldsymbol{w} \in V$，有$\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w}) = (\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}$，即**加法运算满足结合律**。
5. $V$里包含一个元素，记作$\mathbf{0}$，它满足$\mathbf{0}+\boldsymbol{v} = \boldsymbol{v}+\mathbf{0} = \boldsymbol{v}$对任何$\boldsymbol{v} \in V$成立，即**关于加法运算存在一个零向量**。
6. 对$V$里的任何元素$\boldsymbol{v} \in V$，都存在另一个$V$里的元素，记作$-\boldsymbol{v}$，使得$\boldsymbol{v}+(-\boldsymbol{v}) = (-\boldsymbol{v})+\boldsymbol{v} = \mathbf{0}$，**即$V$里的任何元素都有一个关于加法运算"相反数"**。
7. 对任何实数$c \in \mathbb{R}$，任何$V$里的任何元素$\boldsymbol{u} \in V, \boldsymbol{v} \in V$，都有$c(\boldsymbol{u}+\boldsymbol{v}) = c\boldsymbol{u}+c\boldsymbol{v}$，即**标量积与$V$上定义的加法之间运算满足分配律**。
8. 对任何实数$c, k \in \mathbb{R}$，任何$V$里的任何元素$\boldsymbol{v} \in V$，都有$(c+k)\boldsymbol{v} = c\boldsymbol{v}+k\boldsymbol{v}$，即**实数上的加法与标量积之间的运算满足分配律**。这里注意，表达式$c+k$中出现的加法符号"$+$"指经典意义下实数上的加法，表达式$c\boldsymbol{v}+k\boldsymbol{v}$中出现的加法符号"$+$"指集合$V$上所定义的加法。
9. 对任何实数$c, k \in \mathbb{R}$，任何$V$里的任何元素$\boldsymbol{v} \in V$，都有$c(k\boldsymbol{v}) = (ck)\boldsymbol{v}$，即**标量积满足结合律**。
10. 对$V$里的任何元素$\boldsymbol{v} \in V$，$1\boldsymbol{v} = \boldsymbol{v}$，这里1就是实数中的数字1。

对于一个向量空间$V$，它里面的元素$\boldsymbol{v} \in V$被称为**向量*(vector)***。

**Remark 4.2.**
- 要想验证一个集合$V$是否是一个向量空间，首先我们需要明确$V$上的加法和标量积是如何定义；接下来验证$V$关于加法的封闭性与$V$关于标量积的封闭性，即验证对任何$\boldsymbol{u}, \boldsymbol{v} \in V$与任何实数$c \in \mathbb{R}$，都有$\boldsymbol{u}+\boldsymbol{v} \in V$以及$c\boldsymbol{v} \in V$；最后验证$V$上的加法与标量积满足以上定义*Definition 4.1*中所提到的性质3.到10.。
- 注意，$V$上的加法"$+$"需要与我们所熟悉的实数上的加法所区分，尽管我们都用同一个符号"$+$"来指代这两种加法运算。同样，$V$上的标量积也需要与我们所熟悉的实数上的乘法所区分。比如集合$V = \{a \in \mathbb{R} : a > 0\}$为所有**正实数的集合，我们通过实数上的乘法来定义$V$上的加法"$+_V$"**：对于$a, b \in V$，

$$a +_V b = ab.$$

这里我们特地用符号"$+_V$"来强调这是$V$上的加法，以区分实数上的加法符号$+$。此外，我们用以下方式来定义$V$上的标量积：对于任何实数$k \in \mathbb{R}$，$a \in V$，$ka = a^k = \exp(k \log a)$。容易验证$V$关于加法的封闭性与$V$关于标量

积的封闭性(正实数与正实数相乘还是正实数，正实数的任何次方也都还是正实数)；但是注意，由于

$$1 +_V a = 1a = a; \quad a +_V 1 = a1 = a,$$

此时$V = \{a \in \mathbb{R} : a > 0\}$中的零向量是实数1！ 另外，由于$a +_V \frac{1}{a} = a\frac{1}{a} = 1 = \frac{1}{a}a = \frac{1}{a} +_V a$，此时对于$a \in V$，它关于$V$上的加法$+V$的相反数是$1/a$。请大家自行验证$V$装配上以上定义的加法$+_V$与标量积是一个向量空间，而$V$装配上实数上的加法$+$与实数上的乘法作为标量积则不是一个向量空间。

因此，严格来讲，我们最好用"$+_V$"以及"$c \cdot_V \boldsymbol{v}$"来指代$V$上的加法与标量积，以区分实数上的加法与乘法。然而在英文教材以及其他文献中，为了让记号尽可能简单，我们依然使用"$+$"来指代向量空间$V$上的加法，用"$c\boldsymbol{v}$"来指代$V$上的标量积。因此，当你在以上例子中看到$a + b = ab$时不要感到困惑，这不是写错了，而是代表等式左边的$a + b$是向量空间$V$上的加法运算。换句话说，向量空间$V$上的加法符号"$+$"只是个纯粹的抽象运算符号，我们没必要一定要把它与实数上的加法产生联系；向量空间$V$上标量积也只是个纯粹的抽象运算符号，我们没必要一定要把它与实数上的乘法产生联系。

- 由于我们在标量积$c\boldsymbol{v}$的定义中要求了$c \in \mathbb{R}$是个实数，因此$V$更确切的名字是**实向量空间*(real vector space)***。与此对应的，若我们可以取$c \in \mathbb{C}$为复数来定义标量积，那么我们称$V$为一个**复向量空间*(complex vector space)***。在本章我们主要学习实向量空间，简称为向量空间。

**例子：欧式空间**

对任何自然数$n$，$n$维欧式空间$\mathbb{R}^n$是一个向量空间。$\mathbb{R}^n$上的加法运算为我们已经很熟悉的每个坐标相加:

$$\boldsymbol{u} = (u_1, \dots, u_n) \in \mathbb{R}^n, \boldsymbol{v} = (v_1, \dots, v_n) \in \mathbb{R}^n, \boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, \dots, u_n + v_n);$$

$\mathbb{R}^n$上的标量积为我们已经很熟悉的每个坐标乘以实数:

$$c \in \mathbb{R}^n, \boldsymbol{v} = (v_1, \dots, v_n), c\boldsymbol{v} = (cv_1, \dots, cv_n).$$

容易验证这样的加法和标量积满足Definition 4.1里的性质3.–10.

**例子：$m \times n$-矩阵空间**

我们用$M_{m\times n}$指代所有$m \times n$-矩阵的集合。在$V = M_{m\times n}$上我们定义加法运算为我

们已经熟悉的矩阵加法：

$$A \in M_{m\times n}, B \in M_{m\times n}, A+B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$
$$= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{bmatrix};$$

在$V = M_{m\times n}$上我们定义标量积为我们已经熟悉的矩阵上的标量积运算：

$$c \in \mathbb{R}, A \in M_{m\times n}, \quad cA = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{bmatrix}.$$

容易验证$V = M_{m\times n}$装配上以上加法与标量积后是一个向量空间，显然$V = M_{m\times n}$中的零向量即是$m\times n$-零矩阵$\mathbf{0}_{m\times n}$。因此**任何$m\times n$-矩阵$A$都是$V = M_{m\times n}$中的向量**。

**例子：实数上的函数空间**

我们用$F(-\infty,\infty)$指代所有定义域是实数$\mathbb{R}$，值域也为实数$\mathbb{R}$的函数$f:\mathbb{R}\to\mathbb{R}$所组成的集合。在$V = F(-\infty,\infty)$上我们定义加法运算为

$$f \in F(-\infty,\infty), g \in F(-\infty,\infty), \quad (f+g)(x) = f(x)+g(x),$$

即两个函数$f$与$g$之和是一个新函数$f+g$，它在每个$x\in\mathbb{R}$上的取值等于$f$在$x$上的取值$f(x)$加上$g$在$x$上的取值$g(x)$；在 $V = F(-\infty,\infty)$上我们定义标量积为

$$c \in \mathbb{R}, f \in F(-\infty,\infty), \quad (cf)(x) = cf(x),$$

即实数$c$与函数$f$的标量积是一个新函数$cf$，它在每个$x\in\mathbb{R}$上的取值等于$f$在$x$上的取值$f(x)$乘以$c$。容易验证$V = F(-\infty,\infty)$装配上以上加法与标量积是一个向量空间，$V$上的零向量为常值函数$\mathbf{0}(x) = 0$对任何$x\in\mathbb{R}$，对任何$f\in V = F(-\infty,\infty)$，它关于加法的"相反数"为$(-f)(x) = -f(x)$。更多细节请阅读**英文教材187页 Example 6**。因此**任何函数$f:\mathbb{R}\to\mathbb{R}$都是$V = F(-\infty,\infty)$中的向量**。

**总结**：向量空间是一个数学上的抽象空间，它上面的加法运算与标量积亦是抽象运算符。实数或$n$维欧式空间仅仅是一个特殊的向量空间。从现在起，向量不再仅仅局限于样子为$(v_1,\ldots,v_n)$这样的客体(它只是$n$维欧式空间这一特殊向量空间里的向量)，从以上例子我们可以看到矩阵，函数都可以视为向量空间里的向量。同理，从现在起，向量空间里的加法和标量积也不局限于我们所熟悉的实数上的加法与乘法。

以下定理的证明非常简单，请自行阅读英文教材189页Theorem 4.1.1.

**Theorem 4.3.** *令$V$为一个向量空间，$\mathbf{0}\in V$为$V$里的零向量，$\boldsymbol{u}\in V$，$c\in\mathbb{R}$。那么*

1. $0\boldsymbol{u}=\mathbf{0}$。
2. $k\mathbf{0}=\mathbf{0}$。
3. $(-1)\boldsymbol{u}=-\boldsymbol{u}$。
4. $c\boldsymbol{u}=\mathbf{0}\Rightarrow c=0$ *或者* $\boldsymbol{u}=\mathbf{0}$。

### 4.2. 子空间, Subspaces

**Definition 4.4.** *令$V$为一个向量空间。令$W$为$V$中的一个非空子集(non-empty subset)，记为$W\subset V$。我们称$W$是$V$的一个***子空间***，如果$W$在装配上$V$上的加法与标量积后是一个向量空间。更具体来讲，$W$是$V$的一个***子空间***，如果它满足以下条件：*

1. *对任何$\boldsymbol{u}\in W,\boldsymbol{w}\in W$，$\boldsymbol{u}+\boldsymbol{w}\in W$，这里“$+$”为$V$上的加法。即$W$自己满足关于加法的封闭性。*
2. *对任何$\boldsymbol{w}\in W$，$c\in\mathbb{R}$，$c\boldsymbol{w}\in W$，这里$c\boldsymbol{w}$为$V$上的标量积。即$W$自己满足关于标量积的封闭性。*

不难证明在以上子空间的定义中提到的两个定义方式是等价(equivalent)的，具体细节请阅读英文教材192页Theorem 4.2.1.

**Theorem 4.5** (英文教材Theorem 4.2.1.)**.** *令$V$为一个向量空间。令$W$为$V$中的一个非空子集。那么$W$在装配上$V$上的加法与标量积后是一个向量空间当且仅当它满足以下两个条件*

1. *对任何$\boldsymbol{u}\in W,\boldsymbol{w}\in W$，$\boldsymbol{u}+\boldsymbol{w}\in W$，这里“$+$”为$V$上的加法。即$W$自己满足关于加法的封闭性。*
2. *对任何$\boldsymbol{w}\in W$，$c\in\mathbb{R}$，$c\boldsymbol{w}\in W$，这里$c\boldsymbol{w}$为$V$上的标量积。即$W$自己满足关于标量积的封闭性。*

因此，对于任何向量空间$V$里的非空子集$W$，若要验证$W$是$V$的一个**子空间**，我们只需验证$W$满足

1. 对任何$\boldsymbol{u}\in W, \boldsymbol{w}\in W$，$\boldsymbol{u}+\boldsymbol{w}\in W$，这里“+”为$V$上的加法。即$W$自己满足关于加法的封闭性。
2. 对任何$\boldsymbol{w}\in W$，$c\in\mathbb{R}$，$c\boldsymbol{w}\in W$，这里$c\boldsymbol{w}$为$V$上的标量积。即$W$自己满足关于标量积的封闭性。

**例子：零空间，the zero subspaces**
对任何向量空间$V$，由零向量$\mathbf{0}$自己做组成的子集$W=\{\mathbf{0}\}$是一个子空间，被称作零空间。显然这样的$W$满足加法与标量积封闭性的两个条件：

$$\mathbf{0}+\mathbf{0}=\mathbf{0}, \quad c\mathbf{0}=\mathbf{0} \quad (c\in\mathbb{R}).$$

**例子：欧式空间里的一些子空间**

- 令$\boldsymbol{v}=(v_1,v_2)$为$\mathbb{R}^2$里的一个向量且$\boldsymbol{v}\neq\mathbf{0}$。那么由$\boldsymbol{v}$与原点$\mathbf{0}=(0,0)$所确定的直线$W=\{k\boldsymbol{v}=(kv_1,kv_2):k\in\mathbb{R}\}$是一个$\mathbb{R}^2$里的子空间。容易验证$W$满足加法与标量积封闭性的两个条件：

$$k\boldsymbol{v}+c\boldsymbol{v}=(k+c)\boldsymbol{v}, \quad c(k\boldsymbol{v})=(ck)\boldsymbol{v}.$$

- 令$\boldsymbol{u}=(1,1,1)\in\mathbb{R}^3$，$\boldsymbol{v}=(1,1,0)\in\mathbb{R}^3$与$\mathbf{0}=(0,0,0)\in\mathbb{R}^3$。由这三个点所确定的平面$W=\{k\boldsymbol{u}+c\boldsymbol{v}:k\in\mathbb{R},c\in\mathbb{R}\}$是$\mathbb{R}^3$里的一个子空间。容易验证$W$满足加法与标量积封闭性的两个条件：

$$\boldsymbol{w}_1=k_1\boldsymbol{u}+c_1\boldsymbol{v}\in W, \boldsymbol{w}_2=(k_2\boldsymbol{u}+c_2\boldsymbol{v})\in W\Rightarrow \boldsymbol{w}_1+\boldsymbol{w}_2=(k_1+k_2)\boldsymbol{u}+(c_1+c_2)\boldsymbol{v}\in W;$$

$$\boldsymbol{w}=k\boldsymbol{u}+c\boldsymbol{v}\in W, a\in\mathbb{R}\Rightarrow a\boldsymbol{w}=(ak)\boldsymbol{u}+(ac)\boldsymbol{v}\in W.$$

类似的，令$\boldsymbol{e}_1=(1,0,0)\in\mathbb{R}^3$，$\boldsymbol{e}_2=(0,1,0)\in\mathbb{R}^3$。这两个点与$\mathbf{0}=(0,0,0)\in\mathbb{R}^3$所确定的平面，即$\mathbb{R}^3$里的$xy$-平面，$W=\{(x,y,0)=x\boldsymbol{e}_1+y\boldsymbol{e}_2:x,y\in\mathbb{R}\}$是一个$\mathbb{R}^3$里的子空间。
关于这种类型的子空间构造我们会在稍后的篇幅中继续学习。

**例子：矩阵空间里的一些子空间**
$M_{n\times n}$($n$阶方阵空间)至少包含以下几个子空间：

- $W=\{A\in M_{n\times n}:A^\top=A\}$，即由所有对称矩阵所组成的集合。容易验证，两个对称矩阵之和仍是对称矩阵，任何实数与一个对称矩阵相乘还是一个对称矩阵。因此$W$满足加法与标量积封闭性的两个条件。

- $W=\{A\in M_{n\times n}: A$为上三角矩阵$\}$，即由所有上三角矩阵组成的集合。容易验证，两个上三角矩阵之和仍是上三角矩阵，任何实数与一个上三角矩阵相乘还是一个上三角矩阵。因此$W$满足加法与标量积封闭性的两个条件。
- $W=\{A\in M_{n\times n}: A$为下三角矩阵$\}$，即由所有下三角矩阵组成的集合。
- $W=\{A\in M_{n\times n}: A$为对角矩阵$\}$，即由所有对角矩阵组成的集合。

**例子：函数空间里的一些子空间**

$F(-\infty,\infty)=\{f:\mathbb{R}\to\mathbb{R}\}$(定义域为实数，值域为实数的函数的集合)至少包含以下几个子空间：

- $W=C(-\infty,\infty)=\{f:\mathbb{R}\to\mathbb{R}: f$ 为连续函数$\}$。
- $W=C^1(-\infty,\infty)=\{f:\mathbb{R}\to\mathbb{R}: f$ 为一阶连续可微函数$\}$，即$f$可微/可导，且它的导数是连续函数。
- $W=C^m(-\infty,\infty)=\{f:\mathbb{R}\to\mathbb{R}: f$ 为m阶连续可微函数$\}$ ($m$为自然数)，即$f$可以求导$m$次，且它的$m$阶导数$f^{(m)}$是连续函数。
- $W=C^\infty(-\infty,\infty)=\{f:\mathbb{R}\to\mathbb{R}: f$ 为无穷次连续可微函数$\}$，即$f$可以求导无穷多次。
- $W=P_\infty=\{f(x)=a_0+a_1x+\ldots+a_nx^n: n\geq 0, a_0,a_1,\ldots,a_n\in\mathbb{R}\}$为所有多项式(polynomials)所组成的集合。
- 固定一个非负整数$n\geq 0$，$W=P_n=\{f(x)=a_0+a_1x+\ldots+a_nx^n: a_0,a_1,\ldots,a_n\in\mathbb{R}\}$为**所有次数(degree)小于或等于$n$的多项式**所组成的集合(英文：subspace of polynomials of degree $\leq n$)。

注意以上子空间满足以下包含关系(对任何非负整数$n\geq 0$，任何自然数$m\geq 1$)：

$$P_n\subset P_\infty\subset C^\infty(-\infty,\infty)\subset C^m(-\infty,\infty)\subset C^1(-\infty,\infty)\subset C(-\infty,\infty)\subset F(-\infty,\infty).$$

**例子：可逆矩阵组成的集合不是一个子空间**

令$S=\{A\in M_{n\times n}: A$ 可逆$\}$。显然，对于$A\in S$，$0A=\mathbf{0}_{n\times n}$作为零矩阵不可逆，这意味着$S$不满足关于标量积的封闭性，故不是一个子空间。类似的，对于$A\in S$可逆(注意这等价于$\det(A)\neq 0$)，显然有$-A$也可逆(因为$\det(-A)=(-1)^n\det(A)\neq 0$)，即$-A\in S$；但是$A$与$-A$的和等于零矩阵不可逆，即$A+(-A)$不再属于$S$，因此$S$不满足关于矩阵加法的封闭性，故不是一个子空间。

下面我们介绍几种常见的构造子空间的方法。

**Theorem 4.6** (英文教材Theorem 4.2.2)**.** *若$W_1,\ldots,W_r$为$V$里的子空间，那么它们的交集(intersection)$W_1\cap W_2\cap\ldots\cap W_r$也为$V$里的子空间。*

该定理证明非常直接，可参考英文教材里的证明。

**Theorem 4.7** (英文教材Theorem 4.2.3)**.** 令$V$为一个向量空间，$r \geq 1$为给定自然数，$S = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\}$为$V$里的一个子集(即$\boldsymbol{w}_1 \in V, \ldots, \boldsymbol{w}_r \in V$)。令$W$为所有关于$\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$的线性组合*(linear combination)*:

$$W = \{k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r : k_1, k_2, \ldots, k_r \in \mathbb{R}\}.$$

那么$W$是$V$里的一个子空间。另外，对任何$V$里的子空间$U$，若$U$也包含集合$S$，那么必然有$W \subset U$，即$W$是$V$里所有包含集合$S$的子空间里"最小"的那个。

证明. 首先很容易验证

$$W = \{k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r : k_1, k_2, \ldots, k_r \in \mathbb{R}\}.$$

满足加法与标量积的封闭性：

$$\begin{aligned}\boldsymbol{u} = c_1\boldsymbol{w}_1 + \ldots + c_r\boldsymbol{w}_r \in W, \boldsymbol{v} = k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r \in W \\ \Rightarrow \boldsymbol{u} + \boldsymbol{v} = (c_1 + k_1)\boldsymbol{w}_1 + \ldots + (c_r + w_r)\boldsymbol{w}_r \in W;\end{aligned}$$

$$\boldsymbol{v} = k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r \in W, c \in \mathbb{R} \Rightarrow c\boldsymbol{v} = ck_1\boldsymbol{w}_1 + ck_2\boldsymbol{w}_2 + \ldots + ck_r\boldsymbol{w}_r \in W.$$

因此$W = \{k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r : k_1, k_2, \ldots, k_r \in \mathbb{R}\}$是$V$里的一个子空间。
现在假设$U \subset V$是一个子空间且$S \subset U$。那么由于$\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r \in S \subset U$且$U$作为一个子空间满足加法与标量积的封闭性，对任何实数$k_1, \ldots, k_r$我们必然有

$$k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r \in U.$$

因此有$W = \{k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r : k_1, k_2, \ldots, k_r \in \mathbb{R}\} \subset U$。 □

**Definition 4.8** (英文教材196页Definition 3)**.** 令$V$为一个向量空间，$r \geq 1$为给定自然数，$S = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\}$为$V$里的一个子集(即$\boldsymbol{w}_1 \in V, \ldots, \boldsymbol{w}_r \in V$)。令$W$为所有关于$\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$的线性组合*(linear combination)*:

$$W = \{k_1\boldsymbol{w}_1 + k_2\boldsymbol{w}_2 + \ldots + k_r\boldsymbol{w}_r : k_1, k_2, \ldots, k_r \in \mathbb{R}\}.$$

我们称$W$为由$S = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\}$所生成的子空间*(subspace generated by S)*，称$\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$张成了向量空间$W$*($\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$ span W)*。我们将$W$记作

$$W = span\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\}$$

或者

$$W = span(S).$$

**例子：**

- 令$\boldsymbol{e}_i = (0, \ldots, 0, \underbrace{1}_{\text{第i个}}, 0, \ldots, 0)$，$i = 1, \ldots, n$为$\mathbb{R}^n$里的标准单位向量，那么显然有$\mathbb{R}^n = \text{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$。
- 令$S = \{1, x, x^2, \ldots, x^n\}$为次数小于等于$n$的所有单项式(monomials)的集合。那么显然有

$$P_n = \{f(x) = a_0 + a_1 x + \ldots + a_n x^n : a_0, a_1, \ldots, a_n \in \mathbb{R}\} = \text{span}(S).$$

假设我们现在有$n$个$\mathbb{R}^n$里的向量$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$，并将它们的坐标记为

$$\boldsymbol{v}_1 = (a_{11}, a_{21}, \ldots, a_{n1}), \ldots, \boldsymbol{v}_i = (a_{1i}, a_{2i}, \ldots, a_{ni}), \ldots \boldsymbol{v}_n = (a_{1n}, a_{2n}, \ldots, a_{nn}).$$

因此对于给定的系数$k_1, \ldots, k_n \in \mathbb{R}$，线性组合$k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n$的坐标为

$$(k_1 a_{11} + k_2 a_{12} + \ldots + k_n a_{1n}, \ldots, k_1 a_{n1} + k_2 a_{n2} + \ldots + k_n a_{nn}).$$

因此，

$$\begin{aligned}
&\mathbb{R}^n = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \\
&\Leftrightarrow \text{任何}\mathbb{R}^n\text{里的向量}\boldsymbol{b} = (b_1, \ldots, b_n)\text{都可以表示为}\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\text{的线性组合} \\
&\Leftrightarrow \text{对任何}\boldsymbol{b} = (b_1, \ldots, b_n)\text{都存在系数}k_1, \ldots, k_n\text{使得}\boldsymbol{b} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n : \\
&\qquad k_1 a_{11} + k_2 a_{12} + \ldots + k_n a_{1n} = b_1, \ldots, k_1 a_{n1} + k_2 a_{n2} + \ldots + k_n a_{nn} = b_n \\
&\Leftrightarrow A\boldsymbol{x} = \boldsymbol{b}\text{对任何}\boldsymbol{b}\text{都有解}, A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \\
&\Leftrightarrow A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \text{可逆} \quad \text{(见讲义Theorem 1.25)} \\
&\Leftrightarrow \det(A) \neq 0.
\end{aligned}$$

**例子：** 判断$\boldsymbol{v}_1 = (1, 1, 2)$，$\boldsymbol{v}_2 = (1, 0, 1)$，$\boldsymbol{v}_3 = (2, 1, 3)$是否张成整个$\mathbb{R}^3$。
由以上讨论可知，令$A$为依次由$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$作为列向量的三阶方阵，那么$\mathbb{R}^3 = \text{span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$当且仅当$\det(A) \neq 0$。因此我们只需计算$\det(A) = \det\left(\begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 3 \end{bmatrix}\right)$。经计算可得此时$\det(A) = 0$，因此$\boldsymbol{v}_1 = (1, 1, 2)$，$\boldsymbol{v}_2 = (1, 0, 1)$，$\boldsymbol{v}_3 = (2, 1, 3)$无法张成整个$\mathbb{R}^3$。

**Theorem 4.9** (英文教材Theorem 4.2.4)**.** *令$A \in M_{m\times n}$。以$A$作为系数矩阵的齐次线性方程组*

$$A\boldsymbol{x} = \boldsymbol{0},$$

*即*

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

*的所有解所组成的集合$W = \{\boldsymbol{s} \in \mathbb{R}^n : A\boldsymbol{s} = \boldsymbol{0}\}$是$\mathbb{R}^n$里的一个子空间。我们称$W$为齐次线性方程组$A\boldsymbol{x} = \boldsymbol{0}$的解空间(solution space)。*

证明. 若$\boldsymbol{s}_1 \in W, \boldsymbol{s}_2 \in W$为方程组的两个解，即$A\boldsymbol{s}_1 = A\boldsymbol{s}_2 = \boldsymbol{0}$，那么由于$A(\boldsymbol{s}_1 + \boldsymbol{s}_2) = A\boldsymbol{s}_1 + A\boldsymbol{s}_2 = \boldsymbol{0} + \boldsymbol{0} = \boldsymbol{0}$，得到$\boldsymbol{s}_1 + \boldsymbol{s}_2 \in W$。类似的，若$\boldsymbol{s} \in W$，$c \in \mathbb{R}$，那么$A(c\boldsymbol{s}) = cA\boldsymbol{s} = c\boldsymbol{0} = \boldsymbol{0}$，因此$c\boldsymbol{s} \in W$。以上证明了$W$关于向量加法与标量积的封闭性，因此它是一个子空间。 □

### 4.3. 线性无关性, Linear Independence

在本节我们试图回答这么一个问题：给定一组向量$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \in V$，$W = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为由$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \in V$生成的子空间，那么任何一个$W$里的向量$\boldsymbol{w}$都可以表示为关于$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \in V$的线性组合：

$$\boldsymbol{w} = c_1\boldsymbol{v}_1 + \ldots + c_r\boldsymbol{v}_r.$$

那么，**当$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$满足何种条件时，这样的线性组合表示是唯一的？**(这里的唯一性是指如果$\boldsymbol{w}$同时满足$\boldsymbol{w} = c_1\boldsymbol{v}_1 + \ldots + c_r\boldsymbol{v}_r$，$\boldsymbol{w} = k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r$，那么只有可能是$k_1 = c_1, \ldots, k_r = c_r$。)

注意，关于同一组向量的不同的线性组合是可以表示同一个向量的，比如我们在第三章已经看到：$\boldsymbol{w} = (1, 2)$是关于$\boldsymbol{v}_1 = (1, 0)$，$\boldsymbol{v}_2 = (0, 1)$，$\boldsymbol{v}_3 = (1, 1)$的线性组合，它同时可以被表示为

$$\boldsymbol{w} = 1\boldsymbol{v}_1 + 2\boldsymbol{v}_2 + 0\boldsymbol{v}_3$$

与

$$\boldsymbol{w} = -\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 2\boldsymbol{v}_3.$$

以下概念是线性代数里极其重要的概念：

**Definition 4.10** (英文教材203页Definition 1)**.** 令$V$为一个向量空间，令$r \geq 2$。$V$里的一个集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$被称为一个*线性无关集合(linearly independent set)*或者说*向量$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$线性无关(linearly independent)*，**如果$S$里的任何一个向量$\boldsymbol{v}_i(i = 1, \ldots, r)$都不能表示为$S$里除了$\boldsymbol{v}_i$之外的其他向量的线性组合。**

当然，若集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$不是一个线性无关集合，那么我们就称它为线性相关集合(linearly dependent set)或者$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$线性相关。

实际上线性无关性还有以下的等价定义。

**Theorem 4.11** (英文教材Theorem 4.3.1)**.** 令$V$为一个向量空间，令$r \geq 2$。$V$里的一个集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是一个*线性无关集合*当且仅当以下等式

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$$

只有在$k_1 = k_2 = \ldots = k_r = 0$的时候成立。

*证明.* 假设$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是一个线性无关集合。如果存在不全为零的系数$k_1, \ldots, k_r$使得等式$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$成立，比如说$k_1 \neq 0$，那么我们有

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0} \Rightarrow k_1\boldsymbol{v}_1 = -k_2\boldsymbol{v}_2 - \ldots - k_r\boldsymbol{v}_r \Rightarrow \boldsymbol{v}_1 = \frac{-k_2}{k_1}\boldsymbol{v}_2 + \ldots + \frac{-k_r}{k_1}\boldsymbol{v}_r,$$

即$\boldsymbol{v}_1$可以被表示为$\boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$的线性组合，与$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是一个线性无关集合的假设矛盾。因此在$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是一个线性无关集合的假设下，等式

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$$

只有在$k_1 = k_2 = \ldots = k_r = 0$的时候成立。
现在假设等式

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$$

只有在$k_1 = k_2 = \ldots = k_r = 0$的时候成立。如果$S$不是一个线性无关集合，那么不妨令$\boldsymbol{v}_1$可以被$\boldsymbol{v}_2, \ldots, \boldsymbol{v}_r$的线性组合表示，即存在系数$c_2, \ldots, c_r$使得

$$\boldsymbol{v}_1 = c_2\boldsymbol{v}_2 + \ldots + c_r\boldsymbol{v}_r.$$

显然这意味着$\boldsymbol{v}_1 - c_2\boldsymbol{v}_2 - \ldots - c_r\boldsymbol{v}_r = \boldsymbol{0}$，也即等式

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$$

在$k_1 = 1, k_2 = c_2, \ldots, k_r = c_r$的时候成立了，与我们的假设：等式 $k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$ 只有在$k_1 = k_2 = \ldots = k_r = 0$的时候成立，矛盾。因此如果等式 $k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$ 只有在$k_1 = k_2 = \ldots = k_r = 0$的时候成立，那么必然有$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是一个线性无关集合。 $\square$

从现在开始，我们经常使用条件“等式

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \mathbf{0}$$

只有在$k_1 = k_2 = \ldots = k_r = 0$的时候成立”来验证集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是否是一个线性无关集合。

**例子：**$\mathbb{R}^n$里的$n$个标准单位向量$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$线性无关：若等式$k_1\boldsymbol{e}_1 + \ldots + k_n\boldsymbol{e}_n = \mathbf{0}$成立，那么必然意味着$k_1\boldsymbol{e}_1 + \ldots + k_n\boldsymbol{e}_n = (k_1, \ldots, k_n) = \mathbf{0}$，即必须$k_1 = 0, k_2 = 0, \ldots, k_n = 0$。

**例子：** 考虑$V = P_n = \{f(x) = a_0 + a_1x + \ldots + a_nx^n : a_0, a_1, \ldots, a_n \in \mathbb{R}\}$次数不高于$n$的多项式所组成的向量空间。令$\boldsymbol{p}_0(x) = 1, \boldsymbol{p}_1(x) = x, \ldots, \boldsymbol{p}_n(x) = x^n$为次数不高于$n$的单项式。若系数$c_0, c_1, \ldots, c_n$使得等式

$$c_0\boldsymbol{p}_0 + c_1\boldsymbol{p}_1 + \ldots + c_n\boldsymbol{p}_n = \mathbf{0}$$

成立，那么这意味着对**任何**$x \in \mathbb{R}$，都有

$$c_0 + c_1x + c_2x^2 + \ldots + c_nx^n = 0.$$

此时利用**代数基本定理**：任何**非零**的一元$n$次方程最多只有$n$个根，可知以上等式$c_0 + c_1x + c_2x^2 + \ldots + c_nx^n = 0$.对**任何**$x \in \mathbb{R}$都成立的唯一可能是方程$c_0 + c_1x + c_2x^2 + \ldots + c_nx^n$的所有系数$c_0, c_1, \ldots, c_n$都为0。(否则这个方程为**非零**的一元$n$次方程，至多有$n$个根，因此不可能使得任何$x \in \mathbb{R}$都是方程的根。) 由以上观察可得，$\boldsymbol{p}_0(x) = 1, \boldsymbol{p}_1(x) = x, \ldots, \boldsymbol{p}_n(x) = x^n$线性无关。

**考试中可能出现的一类题型：**判断给定的一组向量是否线性无关。

**例子：** 判断$\boldsymbol{v}_1 = (1, -2, 3)$， $\boldsymbol{v}_2 = (5, 6, -1)$，$\boldsymbol{v}_3 = (3, 2, 1)$在$\mathbb{R}^3$里是否线性无关。

**答案：**

$$
\begin{aligned}
&\boldsymbol{v}_1=(1,-2,3),\boldsymbol{v}_2=(5,6,-1),\boldsymbol{v}_3=(3,2,1)\text{线性无关}\\
&\Leftrightarrow k_1\boldsymbol{v}_1+k_2\boldsymbol{v}_2+k_3\boldsymbol{v}_3=\mathbf{0}\text{只有在}k_1=k_2=k_3=0\text{时成立}\\
&\Leftrightarrow \quad k_1+5k_2+3k_3=0\\
&\qquad -2k_1+6k_2+2k_3=0\\
&\qquad 3k_1-k_2+k_3=0\text{仅有平凡解}\\
&\Leftrightarrow A=\begin{bmatrix}1&5&3\\-2&6&2\\3&-1&1\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_1&\boldsymbol{v}_2&\boldsymbol{v}_3\end{bmatrix}\text{可逆(见讲义Theorem 1.25)}\\
&\Leftrightarrow \det(A)\neq 0.
\end{aligned}
$$

经计算可得$\det(A)=0$，因此由以上分析可知$\boldsymbol{v}_1=(1,-2,3)$，$\boldsymbol{v}_2=(5,6,-1)$，$\boldsymbol{v}_3=(3,2,1)$在$\mathbb{R}^3$里线性相关。另外，通过解齐次方程组$A\boldsymbol{x}=\mathbf{0}$，可以验证$\boldsymbol{v}_3=\frac{1}{2}\boldsymbol{v}_1+\frac{1}{2}\boldsymbol{v}_2$。

**例子：(2021年线性代数期中考试题)** Let $c\in\mathbb{R}$. Suppose that

$p_1(x)=1-2x, p_2(x)=3+x-cx^2, p_3(x)=-1+3x^2, p_4(x)=1+2021x+2021^2x^2+2021^3x^3.$

Find a value $c$ such that $p_1,p_2,p_3,p_4$ are **linearly dependent** in the vector space $P_\infty$.

**答案：** 对任何系数$k_1,k_2,k_3,k_4\in\mathbb{R}$，我们有

$$
\begin{aligned}
k_1p_1+k_2p_2+k_3p_3+k_4p_4=&(k_1+3k_2-k_3+k_4)+(-2k_1+k_2+2021k_2)x\\
&+(-ck_2+3k_3+2021^2k_4)x^2+2021^3x^3.
\end{aligned}
$$

因此，$k_1p_1+k_2p_2+k_3p_3+k_4p_4=\mathbf{0}$当且仅当

$(k_1+3k_2-k_3+k_4)+(-2k_1+k_2+2021k_2)x+(-ck_2+3k_3+2021^2k_4)x^2+2021^3k_4x^3=\mathbf{0}.$

现在我们知道$1,x,x^2,x^3$线性无关，因此以上等式成立必然意味着

$$
\begin{aligned}
k_1+3k_2-k_3+k_4&=0\\
-2k_1+k_2+2021k_2&=0\\
-ck_2+3k_3+2021^2k_4&=0\\
2021^3k_4&=0.
\end{aligned}
$$

因此，我们需要找到一个$c\in\mathbb{R}$使得以上齐次方程组(以$k_1,k_2,k_3,k_4$为未知数)有**非平凡解**。将其系数矩阵记为$A=\begin{bmatrix}1&3&-1&1\\-2&1&0&2021\\0&-c&3&2021^2\\0&0&0&2021^3\end{bmatrix}$，那么我们需要找到一

个$c\in\mathbb{R}$使得$\det(A)=0$。计算可得

$$\det(A)=2021^3(21-2c).$$

因此$c=\frac{21}{2}$为所求。

之前我们讨论了$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$在$r\geq 2$的时候的线性无关性，对于$S=\{\boldsymbol{v}_1\}$的情况，我们称$S$为一个线性无关集合，如果$\boldsymbol{v}_1\neq\mathbf{0}$。
显然，如果一个集合$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$里包含了一个零向量$\mathbf{0}$，那么此时$S$必然线性相关：不妨假设$\boldsymbol{v}_1=\mathbf{0}$，那么等式$k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r=\mathbf{0}$在$k_1=1,k_2=0,\ldots,k_r=0$的时候成立。

**Theorem 4.12.** 令$V$为一个线性空间，$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$为$V$里的一个集合。那么$S$是一个线性无关集合当且仅当$W=\mathit{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$里的任何向量都有唯一的线性组合表示，即如果$\boldsymbol{w}=c_1\boldsymbol{v}_1+\ldots+c_r\boldsymbol{v}_r\in W$与$\boldsymbol{w}=k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r$同时成立，那么必然有$k_1=c_1,\ldots,k_r=c_r$。

证明. 假设$S$为线性无关集合，那么如果$\boldsymbol{w}\in W=\mathrm{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$有两种表示

$$\boldsymbol{w}=c_1\boldsymbol{v}_1+\ldots+c_r\boldsymbol{v}_r,\quad \boldsymbol{w}=k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r.$$

那么两者相减可得

$$(c_1-k_1)\boldsymbol{v}_1+\ldots+(c_r-k_r)\boldsymbol{v}_r=\mathbf{0}.$$

由线性无关性可得系数$c_1-k_1=c_2-k_2=\ldots=c_r-k_r=0$，即$c_1=k_1,\ldots,c_r=k_r$。这意味着$W=\mathrm{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$里的任何向量都有唯一的线性组合表示。
反之，假设$W=\mathrm{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$里的任何向量都有唯一的线性组合表示。那么由于$\mathbf{0}=0\boldsymbol{v}_1+0\boldsymbol{v}_2+\ldots+0\boldsymbol{v}_r$是零向量关于$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r$的一个线性组合表示，由唯一性可知不存在不全为零的系数$c_1,\ldots,c_r$使得$\mathbf{0}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\ldots+c_r\boldsymbol{v}_r$成立。因此由定义可得$S$为线性无关集合。 □

**Theorem 4.13.** 令$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$为$\mathbb{R}^n$里的一个集合。如果$r>n$，那么$S$必然是一个线性相关集合。

证明. 将向量$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r$的坐标记为

$$\boldsymbol{v}_1=(a_{11},a_{21},\ldots,a_{n1}),\ldots,\boldsymbol{v}_r=(a_{1r},a_{2r},\ldots,a_{nr}).$$

显然，存在实数$k_1,\ldots,k_r$使得$k_r\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r=\mathbf{0}$当且仅当$k_1,\ldots,k_r$满足

$$\begin{aligned}&a_{11}k_1+a_{12}k_2+\ldots+a_{1r}k_r=0\\&a_{21}k_1+a_{22}k_2+\ldots+a_{2r}k_r=0\\&\vdots\\&a_{n1}k_1+a_{n2}k_2+\ldots+a_{nr}k_r=0\end{aligned}$$

即$\boldsymbol{s}=(k_1,\ldots,k_r)$是齐次方程组$A\boldsymbol{x}=\mathbf{0}$，这里$A=\begin{bmatrix}a_{11}&a_{12}&\ldots&a_{1r}\\a_{21}&a_{22}&\ldots&a_{2r}\\\vdots&\vdots&\ddots&\vdots\\a_{n1}&a_{n2}&\ldots&a_{nr}\end{bmatrix}$。由于我们假设$r>n$，即未知数的个数$r$大于方程个数$n$，那么由第三次作业Problem A的第一问可得该方程组有无穷多组解，特别地，存在一个**非零解**$\boldsymbol{s}=(k_1,\ldots,k_r)$。这意味着存在不全为零的$k_1,\ldots,k_r$使得等式$k_r\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r=\mathbf{0}$成立。也即$S$为线性相关集合。 □

### 4.4. 坐标与基底, Coordinates and basis

**Definition 4.14.** *我们称一个向量空间$V$为一个有限维向量空间(finite-dimensional vector space)，如果存在一个集合$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$，$n\geq 1$为一个自然数，使得*

1. *$S$张成(spans)空间$V$，即任何$\boldsymbol{v}\in V$都可以表示为$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$的线性组合；*
2. *$S$为一个线性无关集合。*

*满足以上两个条件的集合$S$被称为$V$的一组基底(basis)，它里面包含的向量个数$n$称为$V$的维数(dimension)。*

**例子：$n$维欧式空间$\mathbb{R}^n$里的标准基底与基底的不唯一性**

$\mathbb{R}^n$中的$n$个标准单位向量$\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$组成了$\mathbb{R}^n$的一组基底，被称为$\mathbb{R}^n$的**标准基底(the standard basis for $\mathbb{R}^n$)**。因为在4.2节我们已经验证了$\mathrm{span}(\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n)=\mathbb{R}^n$，在4.3节我们证明了$\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$线性无关。因此$\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$满足了基底的两个性质，从而成为了$\mathbb{R}^n$的一组基底。

然而一个向量空间中可能包含不同的基底。比如向量$\boldsymbol{v}_1=(1,2,1),\boldsymbol{v}_2=(2,9,0),\boldsymbol{v}_3=(3,3,4)$组成了$\mathbb{R}^3$的一组基底。让我们来验证它们满足基底的两个性质。首先我们证明$\mathrm{span}\{\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3\}=\mathbb{R}^3$，即任何$\boldsymbol{b}=(b_1,b_2,b_3)\in\mathbb{R}^3$都可以表示为$\boldsymbol{b}=c_1\boldsymbol{v}_1+$

$c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3$。这显然意味着线性方程组

$$\begin{aligned} c_1 + 2c_2 + 3c_3 &= b_1 \\ 2c_1 + 9c_2 + 3c_3 &= b_2 \\ c_1 + 4c_3 &= b_3 \end{aligned}$$

对任何$\boldsymbol{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$都有解。那么利用讲义Theorem 1.25我们知道，这等价于要求系数矩阵

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 3 \\ 1 & 0 & 4 \end{bmatrix}$$

可逆，也等价于$\det(A) \neq 0$。**注意矩阵$A$的列向量依次为$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$。**经简单计算可得$\det(A) = -1$，因此以上分析告诉我们$\mathbb{R}^3 = \text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$。

现在我们证明$\boldsymbol{v}_1 = (1, 2, 1), \boldsymbol{v}_2 = (2, 9, 0), \boldsymbol{v}_3 = (3, 3, 4)$线性无关，即等式

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \boldsymbol{0}$$

成立的唯一可能是令$c_1 = c_2 = c_3 = 0$。代入$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$的具体坐标后可以看出它们的线性无关性实际上等价于要求线性方程组

$$\begin{aligned} c_1 + 2c_2 + 3c_3 &= 0 \\ 2c_1 + 9c_2 + 3c_3 &= 0 \\ c_1 + 4c_3 &= 0 \end{aligned}$$

仅有一组平凡解。再次利用讲义Theorem 1.25我们知道，这等价于要求系数矩阵

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 3 \\ 1 & 0 & 4 \end{bmatrix}$$

可逆，也等价于$\det(A) \neq 0$。**注意矩阵$A$的列向量依次为$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$。**由于我们已经计算了$\det(A) = -1$，因此$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$线性无关。综上，$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$为$\mathbb{R}^3$的一组基底。

实际上，这个例子里告诉我们，对于$\mathbb{R}^n$里的$n$个向量$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$，**它们构成$\mathbb{R}^n$的一组基底，当且仅当以$\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$依次作为列向量的$n$阶方阵$A$可逆**，更具体来说，此时我们有以下等价关系：

$$\begin{aligned} S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset \mathbb{R}^n\text{线性无关} &\Leftrightarrow \text{span}(S) = \mathbb{R}^n \\ &\Leftrightarrow S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset \mathbb{R}^n\text{是一组基底} \\ &\Leftrightarrow A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots \boldsymbol{v}_n \end{bmatrix}\text{可逆}. \end{aligned}$$

**注意：** 上面这个例子告诉我们向量空间里可能具有多个不同的基底。但是我们可以证明，对于**有限维向量空间**来说，每个基底包含的向量个数都是相同的，因此向量空间的维数等于**任何**一个基底里包含的向量个数。具体证明请见第4.5节。

**例子：次数$\leq n$的多项式空间里的标准基底**
还记得我们用$P_n$指代所有次数$\leq n$的多项式组成的集合，即

$$P_n = \{p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n : a_0, a_1, \ldots, a_n \in \mathbb{R}\}.$$

令$S = \{1, x, x^2, \ldots, x^n\}$。显然任何$P_n$里的多项式$p(x)$都可以表示为$S$中这些单项式(monomials)的线性组合；另外在4.3节我们证明了$S$是一个线性无关集合。综上，$S = \{1, x, x^2, \ldots, x^n\}$是$P_n$的一组基底，被称为$P_n$的**标准基底(standard basis)**。由定义可知$P_n$的维数(dimension)为$S$里向量的个数，即等于$n+1$。

**例子：$m \times n$-矩阵空间里的标准基底**
还记得所有$m \times n$-矩阵组成的集合$M_{m\times n}$是一个向量空间。对于每个$i = 1, \ldots, n$，每个$j = 1, \ldots, n$，令$M^{ij}$为满足以下条件的$m \times n$-矩阵：

$$M^{ij} = \begin{bmatrix} 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \underbrace{1}_{\text{第i行第j列}} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & \ldots & 0 \end{bmatrix}$$

即$M^{ij}$除了第$i$行第$j$列上的项为1，其余项均为0。容易验证

$$S = \{M^{11}, M^{12}, \ldots, M^{1n}, M^{21}, \ldots, M^{2n}, \ldots, M^{m1}, \ldots, M^{mn}\}$$

是$M_{m\times n}$的一组基底。我们称其为$M_{m\times n}$的**标准基底(standard basis)**。由定义可知$M_{m\times n}$的维数(dimension)为$S$里向量的个数，即等于$mn$。

**例子：无穷维空间**
我们称一个向量空间$V$为**无穷维向量空间(infinite dimensional vector space)**，如果不存在一个**有限**的集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$使得$V = \text{span}(S)$。比如多项式空间$P_\infty = \{a_0 + a_1x + \ldots + a_nx^n : n \geq 0, a_0, \ldots, a_n \in \mathbb{R}\}$是一个无穷维向量空间。简单来说，要表示$P_\infty$里的一个多项式$p(x) = a_0 + a_1x + \ldots + a_nx^n$，我们需要$1, x, \ldots, x^n$这$n+1$个单项式，但由于这里$n$可以取任意的自然数，要表示$P_\infty$里的任何一个多项式我们

需要无穷多个单项式$1, x, \ldots, x^n, \ldots$(因为自然数有无穷多个)，因此不存在一个有限集合$S$张成整个空间$P_\infty$，所以$P_\infty$为无穷维。

**注意：在本学期的线性代数课程里我们只学习有限维向量空间。因此不需要关注无穷维向量空间。**

假设$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$是向量空间$V$的一组基底。那么任何$\boldsymbol{v} \in V$都可以表示为$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$的线性组合，即存在$c_1, \ldots, c_n$使得

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n.$$

又由于$S$为线性无关集合，由Theorem 4.12可知这些系数$c_1, \ldots, c_n$是唯一的。

**Definition 4.15.** 假设$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$是向量空间$V$的一组基底，对于$\boldsymbol{v} \in V$，存在唯一的一组实数$c_1, \ldots, c_n$使得

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n.$$

我们称$n$维向量$(c_1, \ldots, c_n) \in \mathbb{R}^n$为**$\boldsymbol{v}$关于基底$S$的坐标向量*(coordinate vector of $v$ relative to $S$)***，并记为

$$(\boldsymbol{v})_S = (c_1, \ldots, c_n).$$

如果我们将坐标向量视为列向量，则记为

$$[\boldsymbol{v}]_S = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

**例子：**

- $V = \mathbb{R}^n$，$S = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$为$\mathbb{R}^n$的标准基底。对于任何$\boldsymbol{v} = (v_1, \ldots, v_n)$，由于$\boldsymbol{v} = v_1\boldsymbol{e}_1 + \ldots + v_n\boldsymbol{e}_n$，$\boldsymbol{v}$关于标准基底的坐标向量$(\boldsymbol{v})_S$即是它自己：

$$(\boldsymbol{v})_S = (v_1, \ldots, v_n).$$

- $V = P_n = \{a_0 + a_1x + \ldots + a_nx^n : a_0, a_1, \ldots, a_n \in \mathbb{R}\}$，$S = \{1, x, x^2, \ldots, x^n\}$为$P_n$的标准基底。对于任何$p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \in P_n$，显然有

$$(p(x))_S = (a_0, a_1, \ldots, a_n) \in \mathbb{R}^{n+1}.$$

- $V = M_{m\times n}$，$S = \{M^{11}, M^{12}, \ldots, M^{1n}, M^{21}, \ldots, M^{2n}, \ldots, M^{m1}, \ldots, M^{mn}\}$为$M_{m\times n}$的标准基底。对于任何$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$显然有

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}M^{ij}.$$

因此

$$(A)_S = (a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{m1}, \ldots, a_{mn}).$$

**例子：(2022年线性代数期中考试题)**

Let $S = \{\boldsymbol{v}_1 = (1,0,0), \boldsymbol{v}_2 = (2,2,0), \boldsymbol{v}_3 = (3,3,3)\}$ be a basis of $\mathbb{R}^3$. Then the coordinate vector of $\boldsymbol{v} = (4,-2,6)$ relative to $S$, $(\boldsymbol{v})_S =$?

**答案：** 假设$(\boldsymbol{v})_S = (c_1, c_2, c_3)$，那么它们必须满足

$$\boldsymbol{v} = (4,-2,6) = (c_1 + 2c_2 + 3c_3, 2c_2 + 3c_3, 3c_3),$$

即满足

$$\begin{aligned} c_1 + 2c_2 + 3c_3 &= 4 \\ 2c_2 + 3c_3 &= -2 \\ 3c_3 &= 6. \end{aligned}$$

用高斯消元法求解以上方程组得到答案

$$c_1 = 6, c_2 = -4, c_3 = 2 \Rightarrow (\boldsymbol{v})_S = (6,-4,2).$$

## 4.5. 向量空间的维数, Dimension

在上一节关于基底的定义Definition 4.14里我们将(有限维)向量空间$V$的维数(dimension)，记为dim$(V)$，定义为它里面任何一组基底所包含的**向量个数**。在本节我们证明任何一组基底所包含的**向量个数都相等**(因此维数的定义是有意义的)。

在证明这个性质之前，我们首先计算一些主要向量空间的维数。实际上在上一节我们已经通过观察标准基底的个数得到以下结果：

- 对任何自然数$n \geq 1$，$\mathbb{R}^n$的维数dim$(\mathbb{R}^n) = n$。
- 对任何非负整数$n \geq 0$，$P_n$的维数dim$(P_n) = n+1$。
- 对任何自然数$n \geq 1, m \geq 1$，$M_{m\times n}$的维数dim$(M_{m\times n}) = mn$。
- 对于零空间(the zero vector space)$W = \{\mathbf{0}\}$，我们规定dim$(V) = 0$。

除此之外，我们注意到

**例子：** 令$V$为一个向量空间，$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$为$V$里的一个(非空)线性无关集合。令$W=\text{span}(S)$为由$S$生成的子空间。那么$S$是$W$的一组基底(显然，任何$\boldsymbol{w}\in W$根据span$(S)$的定义都可以表示为关于$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r$的线性组合，且由假设$S$为线性无关集合，因此$S$**作为**$W$**里的子集**满足基底的两个条件)，我们得到

$$\dim(W)=\dim(\text{span}(S))=\dim(\text{span}\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\})=r.$$

下面我们证明一系列与基底和维数有关的定理。

**Theorem 4.16.** *令$V$为一个向量空间。假设$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$为$V$的一组基底。*

1. *若$m>n$，那么$V$里任何包含$m$个向量的集合$M=\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m\}\subset V$必然为线性相关集合。*
2. *若$m<n$，那么$V$里任何包含$m$个向量的集合$M=\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m\}\subset V$必然无法张成整个空间$V$，即$span(M)\neq V$。*

*证明.* 1. 假设$m>n$，$M=\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m\}\subset V$为任一$V$里的子集。由于$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$为$V$里的一组基底，$M$里的向量$\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m$均可以表示为关于$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$的线性组合。我们将这些线性组合记为：

$$\begin{aligned}\boldsymbol{w}_1&=a_{11}\boldsymbol{v}_1+a_{21}\boldsymbol{v}_2+\ldots+a_{n1}\boldsymbol{v}_n,\\ \boldsymbol{w}_2&=a_{12}\boldsymbol{v}_1+a_{22}\boldsymbol{v}_2+\ldots+a_{n2}\boldsymbol{v}_n,\\ &\vdots\\ \boldsymbol{w}_m&=a_{1m}\boldsymbol{v}_1+a_{2m}\boldsymbol{v}_2+\ldots+a_{nm}\boldsymbol{v}_n.\end{aligned}$$

现在我们考虑等式$k_1\boldsymbol{w}_1+\ldots+k_m\boldsymbol{w}_m=\mathbf{0}$。将$\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m$用以上的关于$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$的线性组合代入$k_1\boldsymbol{w}_1+\ldots+k_m\boldsymbol{w}_m=\mathbf{0}$，我们可以将其表示为一个关于$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$的等式：

$$(k_1a_{11}+k_2a_{12}+\ldots+k_ma_{1m})\boldsymbol{v}_1+(k_1a_{21}+k_2a_{22}+\ldots+k_ma_{2m})\boldsymbol{v}_2+\\ \ldots+(k_1a_{n1}+k_2a_{n2}+\ldots+k_ma_{nm})\boldsymbol{v}_n=\mathbf{0}.$$

由于$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$是一组基底，$S$为线性无关集合，因此以上等式只有当所有系数都等于零时才成立，即$(k_1,\ldots,k_m)$需要满足

$$\begin{aligned}&a_{11}k_1+a_{12}k_2+\ldots+a_{1m}k_m=0\\ &a_{21}k_1+a_{22}k_2+\ldots+a_{2m}k_m=0\\ &\vdots\\ &a_{n1}k_1+a_{n2}x_2+\ldots+a_{nm}k_m=0.\end{aligned}$$

总结一下，我们实际上得到以下等价关系：$k_1\boldsymbol{w}_1+\ldots+k_m\boldsymbol{w}_m=\mathbf{0}$当且仅当$(k_1,\ldots,k_m)\in\mathbb{R}^m$是齐次线性方程组

$$\begin{aligned}
&a_{11}x_1+a_{12}x_2+\ldots+a_{1m}x_m=0\\
&a_{21}x_1+a_{22}x_2+\ldots+a_{2m}x_m=0\\
&\vdots\\
&a_{n1}x_1+a_{n2}x_2+\ldots+a_{nm}x_m=0.
\end{aligned}$$

的一组解。

显然，由于$m>n$，即以上线性方程组的方程个数$n$小于未知数个数$m$，由**第三次作业Problem A的第一问**可知此时以上齐次线性方程组有无穷多个解；特别地，存在一个非零解$(k_1,\ldots,k_m)\neq(0,\ldots,0)\in\mathbb{R}^m$。由于这个非零解$(k_1,\ldots,k_m)$使得等式

$$k_1\boldsymbol{w}_1+\ldots+k_m\boldsymbol{w}_m=\mathbf{0}$$

成立，由定义可知$M=\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m\}$为线性相关集合。

2. 假设$m<n$，$M=\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m\}\subset V$为任一$V$里的子集。基于反证法，假设$\text{span}(M)=V$。那么对基底$S$里的任何向量$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$，它们都可以表示为关于$\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m$的线性组合：

$$\begin{aligned}
\boldsymbol{v}_1&=a_{11}\boldsymbol{w}_1+a_{21}\boldsymbol{w}_2+\ldots+a_{m1}\boldsymbol{w}_m,\\
\boldsymbol{v}_2&=a_{12}\boldsymbol{w}_1+a_{22}\boldsymbol{w}_2+\ldots+a_{m2}\boldsymbol{w}_m,\\
&\vdots\\
\boldsymbol{v}_n&=a_{1n}\boldsymbol{w}_1+a_{2n}\boldsymbol{w}_2+\ldots+a_{mn}\boldsymbol{w}_m.
\end{aligned}$$

注意，如果$(k_1,\ldots,k_n)\in\mathbb{R}^n$使得等式

$$k_1\boldsymbol{v}_1+\ldots+k_n\boldsymbol{v}_n=\mathbf{0}.$$

将$\boldsymbol{v}_1=a_{11}\boldsymbol{w}_1+a_{21}\boldsymbol{w}_2+\ldots+a_{m1}\boldsymbol{w}_m,\ldots,\boldsymbol{v}_n=a_{1n}\boldsymbol{w}_1+a_{2n}\boldsymbol{w}_2+\ldots+a_{mn}\boldsymbol{w}_m$代入以上等式，可得

$$\begin{aligned}
&(k_1a_{11}+k_2a_{12}+\ldots+k_na_{1n})\boldsymbol{w}_1+(k_1a_{21}+k_2a_{22}+\ldots+k_na_{2n})\boldsymbol{w}_2+\\
&\qquad\ldots+(k_1a_{m1}+k_2a_{m2}+\ldots+k_na_{mn})\boldsymbol{w}_m=\mathbf{0}.
\end{aligned}$$

反之，假如$(k_1,\ldots,k_n)\in\mathbb{R}^n$使得等式

$$\begin{aligned}
&(k_1a_{11}+k_2a_{12}+\ldots+k_na_{1n})\boldsymbol{w}_1+(k_1a_{21}+k_2a_{22}+\ldots+k_na_{2n})\boldsymbol{w}_2+\\
&\qquad\ldots+(k_1a_{m1}+k_2a_{m2}+\ldots+k_na_{mn})\boldsymbol{w}_m=\mathbf{0}.
\end{aligned}$$

成立，那么通过用$\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m$表示$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$，我们可得等式$k_1\boldsymbol{v}_1+\ldots+k_n\boldsymbol{v}_n=\mathbf{0}$成立。因此，$k_1\boldsymbol{v}_1+\ldots+k_n\boldsymbol{v}_n=\mathbf{0}$当且仅当

$$(k_1a_{11}+k_2a_{12}+\ldots+k_na_{1n})\boldsymbol{w}_1+(k_1a_{21}+k_2a_{22}+\ldots+k_na_{2n})\boldsymbol{w}_2+\\ \ldots+(k_1a_{m1}+k_2a_{m2}+\ldots+k_na_{mn})\boldsymbol{w}_m=\mathbf{0}.$$

显然，假如$(k_1,\ldots,k_n)\in\mathbb{R}^n$为齐次线性方程组

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

的解，那么我们有

$$\begin{aligned}
(k_1a_{11}+k_2a_{12}+\ldots+k_na_{1n})\boldsymbol{w}_1+(k_1a_{21}+k_2a_{22}+\ldots+k_na_{2n})\boldsymbol{w}_2+\\
\ldots+(k_1a_{m1}+k_2a_{m2}+\ldots+k_na_{mn})\boldsymbol{w}_m\\
=0\boldsymbol{w}_1+\ldots+0\boldsymbol{w}_m=\mathbf{0},
\end{aligned}$$

从而利用以上标为蓝色的等价关系得到$k_1\boldsymbol{v}_1+\ldots+k_n\boldsymbol{v}_n=\mathbf{0}$。然而，由于假设$m<n$，方程组

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

的方程个数$m$小于未知数个数$n$，再次利用**第三次作业Problem A的第一问**可知此时以上齐次线性方程组有无穷多个解；特别地，存在一个非零解$(k_1,\ldots,k_m)\neq(0,\ldots,0)\in\mathbb{R}^n$。由以上讨论可知，这样的$(k_1,\ldots,k_m)\neq(0,\ldots,0)\in\mathbb{R}^n$使得$k_1\boldsymbol{v}_1+\ldots+k_n\boldsymbol{v}_n=\mathbf{0}$成立。这意味着$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$为线性相关集合，与$S$是基底的假设矛盾。因此$\text{span}(M)=V$不能成立。

□

利用以上定理，我们证明**有限维向量空间$V$里任何基底的个数都相同**。

**Theorem 4.17.** 令$V$为一个有限维向量空间。如果$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$与$M=\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m\}$都是$V$的基底，那么必然有$n=m$。

证明. 首先由于$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$是$V$的一组基底，由定义可知$\text{span}(S) = V$且$S$为线性无关集合。同样，由于$M = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$也是$V$的一组基底，我们也有$\text{span}(M) = V$且$M$为线性无关集合。
如果$m > n$，那么由Theorem 4.16-(1)可知，含有$m$个向量的集合$M$必然为线性相关集合，与$M$是基底的假设矛盾。
如果$m < n$，那么由Theorem 4.16-(2)可知，含有$m$个向量的集合$M$必然无法满足$\text{span}(M) = V$，再次与$M$是基底的假设矛盾。
综上，我们只能有$m = n$。 □

**Theorem 4.18** (Plus/Minus Theorem). *令$V$为一个向量空间，$S \subset V$为一个非空子集且包含有有限个向量。*

1. *如果$S$为一个线性无关集合，$\boldsymbol{v} \in V$且$\boldsymbol{v} \notin span(S)$，那么$S \cup \{\boldsymbol{v}\}$依然为一个线性无关集合。*
2. *如果$\boldsymbol{v} \in S$且$\boldsymbol{v}$可以被$S$里的其他向量的线性组合表示，即$\boldsymbol{v} \in span(S - \{\boldsymbol{v}\})$，那么$span(S) = span(S - \{\boldsymbol{v}\})$。*

证明. 该证明不作为考试内容，时间紧张的同学可以跳过。但理解该证明过程对解答与基底相关的证明题有一定帮助。

1. 设$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为线性无关集合。令$\boldsymbol{v} \in V$且$\boldsymbol{v} \notin \text{span}(S)$。我们现在证明集合$M = S \cup \{\boldsymbol{v}\} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r, \boldsymbol{v}\}$为线性无关集合，即等式

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r + k_{r+1}\boldsymbol{v} = \boldsymbol{0}$$

只有当$k_1 = \ldots = k_r = k_{r+1} = 0$时成立。
基于反证法，我们假设存在$k_1, \ldots, k_{r+1}$且$k_{r+1} \neq 0$使得等式$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r + k_{r+1}\boldsymbol{v} = \boldsymbol{0}$成立。那么显然有

$$k_{r+1}\boldsymbol{v} = -(k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r) \Rightarrow \boldsymbol{v} = -\frac{k_1}{k_{r+1}}\boldsymbol{v}_1 - \ldots - -\frac{k_r}{k_{r+1}}\boldsymbol{v}_r,$$

即此时$\boldsymbol{v} \in \text{span}(S)$，与假设$\boldsymbol{v} \notin \text{span}(S)$矛盾。因此若存在$k_1, \ldots, k_{r+1}$使得等式$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r + k_{r+1}\boldsymbol{v} = \boldsymbol{0}$成立，那么$k_{r+1}$必须等于零。这也意味着$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r + k_{r+1}\boldsymbol{v} = \boldsymbol{0}$实际上可以写成

$$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}.$$

因为$S = \{\boldsymbol{v}_1, \ldots, \dot{_}r\}$为线性无关集合，只有当$k_1 = \ldots = k_r = 0$该等式才成立。综上，$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$只有当$k_1 = \ldots = k_r = k_{r+1} = 0$时成立。证毕。

2. 设$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$。已知$\boldsymbol{v} \in S$，故不妨假设$\boldsymbol{v} = \boldsymbol{v}_r$。由假设，$\boldsymbol{v} = \boldsymbol{v}_r$可以写成关于$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{r-1}$的线性组合：

$$\boldsymbol{v}_r = c_1\boldsymbol{v}_1 + \ldots + c_{r-1}\boldsymbol{v}_{r-1}.$$

那么对于任何$\boldsymbol{w} \in \text{span}(S)$，我们有

$$\boldsymbol{w} = k_1\boldsymbol{v}_1+\ldots+k_{r-1}\boldsymbol{v}_{r-1}+k_r\boldsymbol{v}_r = k_1\boldsymbol{v}_1+\ldots+k_{r-1}\boldsymbol{v}_{r-1}+k_r(c_1\boldsymbol{v}_1+\ldots+c_{r-1}\boldsymbol{v}_{r-1}).$$

简单整理后得到

$$\boldsymbol{w} = (k_1 + k_rc_1)\boldsymbol{v}_1 + \ldots + (k_{r-1} + k_rc_{r-1})\boldsymbol{v}_{r-1}.$$

这说明任何$\boldsymbol{w} \in \text{span}(S)$都可以写成关于$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{r-1}$的线性组合，即$\boldsymbol{w} \in \text{span}(S - \{\boldsymbol{v}\})$。由此我们得到

$$\text{span}(S) \subset \text{span}(S - \{\boldsymbol{v}\}).$$

另一方面，由于$S - \{\boldsymbol{v}\} \subset S$，显然有$\text{span}(S - \{\boldsymbol{v}\}) \subset \text{span}(S)$。综上，得到$\text{span}(S - \{\boldsymbol{v}\}) = \text{span}(S)$。

□

以下定理非常重要且为重要考点之一。

**Theorem 4.19.** 令$n \geq 1$，$V$为$n$维向量空间，即$dim(V) = n$。令$S = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$为$V$的一组基底。那么以下说法成立：

1. 对于任何$V$里的子集$M = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$，$M$为线性无关集合当且仅当$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$关于$S$的坐标向量$(\boldsymbol{v}_1)_S \in \mathbb{R}^n, \ldots, (\boldsymbol{v}_r)_S \in \mathbb{R}^n$是$\mathbb{R}^n$里的线性无关集合。
2. 对于任何$V$里的含有$n$个向量的子集$M = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$，$M$是$V$的一组基底当且仅当$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$关于$S$的坐标向量$(\boldsymbol{v}_1)_S \in \mathbb{R}^n, \ldots, (\boldsymbol{v}_n)_S \in \mathbb{R}^n$是$\mathbb{R}^n$里的基底，当且仅当以$[\boldsymbol{v}_1]_S, \ldots, [\boldsymbol{v}_n]_S$为列向量的$n$阶方阵$A$可逆。

证明. 该证明本身不作为考试内容，时间紧张的同学可以跳过。但请务必牢记定理内容，且理解该证明过程对解答与基底相关的证明题有一定帮助。

1. 我们将坐标向量$(\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_r)_S$记为

$$(\boldsymbol{v}_1)_S = (a_{11}, a_{21}, \ldots, a_{n1}),$$

$$(\boldsymbol{v}_2)_S = (a_{12}, a_{22}, \ldots, a_{n2}),$$

$$\vdots$$

$$(\boldsymbol{v}_r)_S = (a_{1r}, a_{2r}, \dots, a_{nr}),$$

即

$$\begin{aligned}
\boldsymbol{v}_1 &= a_{11}\boldsymbol{w}_1 + a_{21}\boldsymbol{w}_2 + \dots + a_{n1}\boldsymbol{w}_n,\\
\boldsymbol{v}_2 &= a_{12}\boldsymbol{w}_1 + a_{22}\boldsymbol{w}_2 + \dots + a_{n2}\boldsymbol{w}_n,\\
&\ \ \vdots\\
\boldsymbol{v}_r &= a_{1r}\boldsymbol{w}_1 + a_{2r}\boldsymbol{w}_2 + \dots + a_{nr}\boldsymbol{w}_n.
\end{aligned}$$

我们现在注意到

$$\begin{aligned}
k_1\boldsymbol{v}_1 + \dots + k_r\boldsymbol{v}_r = \mathbf{0} &\Leftrightarrow k_1(a_{11}\boldsymbol{w}_1 + a_{21}\boldsymbol{w}_2 + \dots + a_{n1}\boldsymbol{w}_n) + \dots\\
&\qquad + k_r(a_{1r}\boldsymbol{w}_1 + a_{2r}\boldsymbol{w}_2 + \dots + a_{nr}\boldsymbol{w}_n) = \mathbf{0}\\
&\Leftrightarrow (k_1a_{11} + k_2a_{12} + \dots + k_ra_{1r})\boldsymbol{w}_1 + \dots\\
&\qquad + (k_1a_{n1} + k_2a_{n2} + \dots + k_ra_{nr})\boldsymbol{w}_n = \mathbf{0}.
\end{aligned}$$

由于$S = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$为$V$的一组基底，它是一个线性无关集合，因此由以上分析我们得到：

$$k_1\boldsymbol{v}_1 + \dots + k_r\boldsymbol{v}_r = \mathbf{0}$$

当且仅当

$$\begin{aligned}
&k_1a_{11} + k_2a_{12} + \dots + k_ra_{1r} = 0\\
&k_1a_{21} + k_1a_{22} + \dots + k_ra_{2r} = 0\\
&\vdots\\
&k_1a_{n1} + k_1a_{n2} + \dots + k_ra_{nr} = 0
\end{aligned}$$

显然后者又等价于

$$k_1\underbrace{\begin{bmatrix} a_{11}\\ a_{21}\\ \vdots\\ a_{n1}\end{bmatrix}}_{=[\boldsymbol{v}_1]_S} + k_2\underbrace{\begin{bmatrix} a_{12}\\ a_{22}\\ \vdots\\ a_{n2}\end{bmatrix}}_{=[\boldsymbol{v}_2]_S} + \dots + k_r\underbrace{\begin{bmatrix} a_{1r}\\ a_{2r}\\ \vdots\\ a_{nr}\end{bmatrix}}_{=[\boldsymbol{v}_r]_S} = \begin{bmatrix} 0\\ 0\\ \vdots\\ 0\end{bmatrix}.$$

总结：我们实际上得到，

$$k_1\boldsymbol{v}_1 + \dots + k_r\boldsymbol{v}_r = \mathbf{0} \in V \Leftrightarrow k_1[\boldsymbol{v}_1]_S + \dots + k_r[\boldsymbol{v}_r]_S = \mathbf{0} \in \mathbb{R}^n.$$

如果$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$线性无关，因此$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0} \in V$只有当$k_1 = \ldots = k_r = 0$成立，因此利用以上等价关系，我们得到$k_1[\boldsymbol{v}_1]_S + \ldots + k_r[\boldsymbol{v}_r]_S = \boldsymbol{0} \in \mathbb{R}^n$只有当$k_1 = \ldots = k_r = 0$成立，故$\{[\boldsymbol{v}_1]_S, \ldots, [\boldsymbol{v}_r]_S\}$为$\mathbb{R}^n$里的线性无关集合。反之，如果$\{[\boldsymbol{v}_1]_S, \ldots, [\boldsymbol{v}_r]_S\}$为$\mathbb{R}^n$里的线性无关集合，相同的推理得到$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$线性无关。

2. 我们已经证明了$M = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset V$线性无关当且仅当坐标向量

$$\{(\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_n)_S\}$$

是$\mathbb{R}^n$里的线性无关集合。因此我们现在只需证明$\text{span}(M) = V$当且仅当

$$\text{span}\{(\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_n)_S\} = \mathbb{R}^n.$$

先假设$\text{span}(M) = V$。那么对于任何$\boldsymbol{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$，令$\boldsymbol{u} = b_1\boldsymbol{w}_1 + \ldots + b_n\boldsymbol{w}_n \in V$(注意$(\boldsymbol{u})_S = \boldsymbol{b}$)。由于$\text{span}(M) = V$，存在系数$c_1, \ldots, c_n$使得

$$c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n = \boldsymbol{u}.$$

现在将

$$\begin{aligned}
\boldsymbol{v}_1 &= a_{11}\boldsymbol{w}_1 + a_{21}\boldsymbol{w}_2 + \ldots + a_{n1}\boldsymbol{w}_n, \\
\boldsymbol{v}_2 &= a_{12}\boldsymbol{w}_1 + a_{22}\boldsymbol{w}_2 + \ldots + a_{n2}\boldsymbol{w}_n, \\
&\vdots \\
\boldsymbol{v}_n &= a_{1n}\boldsymbol{w}_1 + a_{2n}\boldsymbol{w}_2 + \ldots + a_{nn}\boldsymbol{w}_n.
\end{aligned}$$

以及$\boldsymbol{u} = b_1\boldsymbol{w}_1 + \ldots + b_n\boldsymbol{w}_n \in V$代入等式$c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n = \boldsymbol{u}$，我们得到一个关于$\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$的等式：

$$\begin{aligned}
(c_1a_{11} + c_2a_{12} + \ldots + c_na_{1n})\boldsymbol{w}_1 + \ldots + (c_1a_{n1} + c_2a_{n2} + \ldots + c_na_{nn})\boldsymbol{w}_n = \\
b_1\boldsymbol{w}_1 + \ldots + b_n\boldsymbol{w}_n \\
\Leftrightarrow (c_1a_{11} + c_2a_{12} + \ldots + c_na_{1n} - b_1)\boldsymbol{w}_1 \\
+ \ldots (c_1a_{n1} + c_2a_{n2} + \ldots + c_na_{nn} - b_n)\boldsymbol{w}_n = \boldsymbol{0}.
\end{aligned}$$

由于$\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$线性无关，以上等式实际上告诉我们

$$\begin{aligned}
c_1a_{11} + c_2a_{12} + \ldots + c_na_{1n} - b_1 = 0, \\
\vdots \\
c_1a_{n1} + c_2a_{n2} + \ldots + c_na_{nn} - b_n = 0
\end{aligned}$$

也即

$$c_1\underbrace{\begin{bmatrix}a_{11}\\a_{21}\\\vdots\\a_{n1}\end{bmatrix}}_{=[\boldsymbol{v}_1]_S}+c_2\underbrace{\begin{bmatrix}a_{12}\\a_{22}\\\vdots\\a_{n2}\end{bmatrix}}_{=[\boldsymbol{v}_2]_S}+\ldots+c_n\underbrace{\begin{bmatrix}a_{1n}\\a_{2n}\\\vdots\\a_{nn}\end{bmatrix}}_{=[\boldsymbol{v}_n]_S}=\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}.$$

这当然意味着$\boldsymbol{b}\in\text{span}\{[\boldsymbol{v}_1]_S,\ldots,[\boldsymbol{v}_n]_S\}$。由于$\boldsymbol{b}\in\mathbb{R}^n$可以取任何向量，我们得到$\text{span}\{[\boldsymbol{v}_1]_S,\ldots,[\boldsymbol{v}_n]_S\}=\mathbb{R}^n$。

现在假设$\text{span}\{[\boldsymbol{v}_1]_S,\ldots,[\boldsymbol{v}_n]_S\}=\mathbb{R}^n$。那么对任何$\boldsymbol{u}\in V$，将其关于$S$的坐标向量记为$(\boldsymbol{u})_S=(b_1,\ldots,b_n)\in\mathbb{R}^n$。由假设，存在$c_1,\ldots,c_n$使得

$$c_1[\boldsymbol{v}_1]_S+\ldots+c_n[\boldsymbol{v}_n]_S=\boldsymbol{b}.$$

那么使用与之前相同的推理，容易得到

$$c_1\boldsymbol{v}_1+\ldots+c_n\boldsymbol{v}_n=\boldsymbol{u},$$

即$\boldsymbol{u}\in\text{span}(S)$。由于$\boldsymbol{u}\in V$为任意向量，我们得到$\text{span}(S)=V$。

现在我们已经证明对于任何$V$里的含有$n$个向量的子集$M=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$，$M$是$V$的一组基底当且仅当$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r$关于$S$的坐标向量$(\boldsymbol{v}_1)_S\in\mathbb{R}^n,\ldots,(\boldsymbol{v}_n)_S\in\mathbb{R}^n$是$\mathbb{R}^n$里的基底。另一方面，在**4.4节Definition 4.14下方的例子**里我们已经证明了$\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\in\mathbb{R}^n$是$\mathbb{R}^n$的一组基底当且仅当以$\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n$为列向量的矩阵可逆。因此我们最终得到: 任何$V$里的含有$n$个向量的子集$M=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$，$M$是$V$的一组基底当且仅当以$[\boldsymbol{v}_1]_S,\ldots,[\boldsymbol{v}_n]_S$为列向量的$n$阶方阵$A$可逆。

□

**例子：(2022年线性代数考试题)**

Let $p_1(x)=1+3x$,$p_2(x)=2+4x$,$p_3(x)=-4x^2$ be three polynomials in $P_2$.

1. Let $M=\{1,x,x^2\}$ be the standard basis of $P_2$. Let $A=\begin{bmatrix}[p_1(x)]_M & [p_2(x)]_M & [p_3(x)]_M\end{bmatrix}$ be the matrix such that its columns are $[p_1(x)]_M$, $[p_2(x)]_M$ and $[p_3(x)]_M$. Compute the adjoint matrix of $A$.
2. Prove that $S=\{p_1(x),p_2(x),p_3(x)\}$ is a basis of $P_2$.

**答案：**

1. 由$p_1(x)=1+3x$,$p_2(x)=2+4x$,$p_3(x)=-4x^2$可知

$$[p_1(x)]_M=\begin{bmatrix}1\\3\\0\end{bmatrix},[p_2(x)]_M=\begin{bmatrix}2\\4\\0\end{bmatrix},[p_3(x)]_M=\begin{bmatrix}0\\0\\-4\end{bmatrix},$$

因此$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & -4 \end{bmatrix}$。经过计算后可得

$$\mathrm{adj}(A) = \begin{bmatrix} -16 & 8 & 0 \\ 12 & -4 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

2. 利用Theorem 4.19我们知道要证$S = \{p_1(x), p_2(x), p_3(x)\}$是$P_2$的一组基底只需证明它们的坐标向量$[p_1(x)]_M$，$[p_2(x)]_M$与$[p_3(x)]_M$所组成的矩阵$A$可逆。简单计算可得$\det(A) = 8 \neq 0$，因此$A$可逆，所以$S = \{p_1(x), p_2(x), p_3(x)\}$是$P_2$的一组基底。

以下定理告诉我们在理论上可以通过在任一个向量的集合里通过增减向量使其变成一组基底。

**Theorem 4.20.** *令$V$为一个$n$维向量空间。令$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$为$V$中的一个子集。*

*1. 若span$(S) = V$且$m > n$且我们可以通过删掉$S$里的某些向量得到$V$的一组基底。*

*2. 若$S$为一个线性无关集合且$m < n$且我们可以通过给$S$增加某些向量得到$V$的一组基底。*

*证明.* 1. 由Theorem 4.16以及$m > n = \dim(V)$，可知$S$必然为线性相关集合。由定义，$S$里存在一个向量，不妨设其为向量$\boldsymbol{v}_m$，可以表示为$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-1}$的线性组合。此时利用Theorem 4.18-(2)，将$\boldsymbol{v}_m$从$S$中删掉后得到的集合$M = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-1}\}$满足$\mathrm{span}(M) = \mathrm{span}(S)$。如果$M$已经是一个线性无关集合，那么显然$M$已经是$V$的一组基底(注意此时$m$必然等于$n+1$)。如果$M$依然是线性相关集合，那么重复以上步骤我们可以删掉$M$里的一个向量，设为$\boldsymbol{v}_{m-1}$，从而得到一个更小的集合$N = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{m-2}\}$，满足$\mathrm{span}(N) = \mathrm{span}(M) = \mathrm{span}(S)$。如果$N$为线性无关，那么$N$是$V$的一组基底，如果$N$仍然线性相关，那么继续以上步骤删掉$N$里的向量，直到我们得到一组基底为止。

2. 由Theorem 4.16以及$m < n = \dim(V)$，可知$\mathrm{span}(S) \neq V$，即存在$\boldsymbol{v} \in V - \mathrm{span}(S)$。那么由Theorem 4.18，$M = S \cup \{\boldsymbol{v}\}$仍是一个线性无关集合。如果此时$\mathrm{span}(M) = V$，那么$M$是$V$的一组基底(注意此时$m = n - 1$)。如果$\mathrm{span}(M) \neq V$，那么重复以上步骤向$M$里增加向量，最终得到$V$的一组基底。

□

**例子：(2022年线性代数期中考试题)**

1. Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ be a linearly independent set in the $n$–dimsnional Euclidean vector space $\mathbb{R}^n$, and $\boldsymbol{v}_3 \in \mathbb{R}^3, \boldsymbol{v}_3 \neq \mathbf{0}$ be another non–zero vector such that $\boldsymbol{v}_3$ is orthogonal to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Prove that $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is also a linearly independent set in $\mathbb{R}^n$.
2. Now let $S = \{\boldsymbol{v}_1 = (1,0,2), \boldsymbol{v}_2 = (3,2,0)\}$ be a linearly independent set in $\mathbb{R}^3$. Find a vector $\boldsymbol{v}_3 \in \mathbb{R}^3$ such that $S' = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is a basis of $\mathbb{R}^3$.

**答案：**

1. 考虑等式
$$k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 + k_3\boldsymbol{v}_3 = \mathbf{0}.$$
假设$k_3 \neq 0$，那么$\boldsymbol{v}_3 = -\frac{k_1}{k_3}\boldsymbol{v}_1 - \frac{k_2}{k_3}\boldsymbol{v}_2$。然而，由于$\boldsymbol{v}_3$同时正交于$\boldsymbol{v}_1, \boldsymbol{v}_2$，可得$\boldsymbol{v}_3 \cdot (-\frac{k_1}{k_3}\boldsymbol{v}_1 - \frac{k_2}{k_3}\boldsymbol{v}_2) = -\frac{k_1}{k_3}\boldsymbol{v}_3 \cdot \boldsymbol{v}_1 - \frac{k_2}{k_3}\boldsymbol{v}_3 \cdot \boldsymbol{v}_2 = 0$，这与$\boldsymbol{v}_3 \cdot (-\frac{k_1}{k_3}\boldsymbol{v}_1 - \frac{k_2}{k_3}\boldsymbol{v}_2) = \boldsymbol{v}_3 \cdot \boldsymbol{v}_3 = \|\boldsymbol{v}_3\|^2 > 0$矛盾。因此$k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 + k_3\boldsymbol{v}_3 = \mathbf{0}$.的必要条件是$k_3 = 0$。此时该等式化简为$k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 = \mathbf{0}$。那么由于$\boldsymbol{v}_1, \boldsymbol{v}_2$线性无关，可知$k_1 = k_2 = 0$，即$k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 + k_3\boldsymbol{v}_3 = \mathbf{0}$.只有当$k_1 = k_2 = k_3 = 0$时成立。
2. 由第一问可知，若$\boldsymbol{v}_3 \neq \mathbf{0}$同时垂直于$\boldsymbol{v}_1, \boldsymbol{v}_2$，那么$S' = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$为线性无关集合。由于$S'$包含了三个向量，它必然是$\mathbb{R}^3$里的一组基底。由于$\boldsymbol{v}_1 \times \boldsymbol{v}_2$同时正交于$\boldsymbol{v}_1, \boldsymbol{v}_2$，经计算可得
$$\boldsymbol{v}_3 = \boldsymbol{v}_1 \times \boldsymbol{v}_2 = (-4, 6, 2)$$
即为所求向量。

**Theorem 4.21.** *令$V$为$n$维向量空间，$W \subset V$为$V$的一个子空间。*

1. *令$m = dim(W)$，则必有$m \leq n$。*
2. *$W = V$当且仅当$m = n$。*

证明. 1. 令$S = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$为$W$的一组基底，$m = \dim(W)$。由于$S$在$W$里是一个线性无关集合，它在$V$里也是一个线性无关集合。如果$m > n$，那么Theorem 4.16-(1)告诉我们$S$必然为线性相关集合，矛盾！因此$m \leq n$必须成立。
2. 若$W = V$，那么必然有$\dim(W) = m = n = \dim(V)$。
若$\dim(W) = m = n = \dim(V)$，则$W$的任何基底都包含$n$个向量。假设$S$为$W$的一组基底，但$\mathrm{span}(S) = W \neq V$，那么存在$\boldsymbol{v} \in V - \mathrm{span}(S)$。此时由Theorem 4.18-(1)可知$S \cup \{\boldsymbol{v}\}$为$V$里的一个线性无关集合但包含有$n$+1个向量，与$\dim(V) =$

$n$矛盾(这里使用Theorem 4.16-(2)，即对于$n = \dim(V)$，任何含有多于$n$个向量的集合必然线性相关)。因此$W = V$必须成立。

$\square$

### 4.6. 基底变换, Change of basis

通过在**4.4节**的学习我们知道，对于$n$维向量空间$V$，$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$为$V$里一组给定的基底，那么任何$\boldsymbol{v} \in V$都可以被**唯一**的表示为

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \ldots + c_n\boldsymbol{v}_n,$$

我们将以上线性组合里的系数$c_1, \ldots, c_n$组成的$n$维向量

$$(\boldsymbol{v})_S = (c_1, \ldots, c_n) \in \mathbb{R}^n \quad \text{或} \quad [\boldsymbol{v}]_S = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$$

称为$\boldsymbol{v}$关于$S$的坐标向量(coordinate vector)。
我们称以上映射

$$\boldsymbol{v} \in V \to (\boldsymbol{v})_S \in \mathbb{R}^n$$

为(从$n$维向量空间$V$到$n$维欧式空间的)**关于基底$S$坐标映射(coordinate map relative to $S$)**。
注意，若$\boldsymbol{v} \in V$关于基底$S$的坐标向量为$(\boldsymbol{v})_S = (c_1, \ldots, c_n)$，即$\boldsymbol{v} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n$；$\boldsymbol{w} \in V$关于基底$S$的坐标向量为$(\boldsymbol{w})_S = (k_1, \ldots, k_n)$，即$\boldsymbol{w} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n$，那么$\boldsymbol{v} + \boldsymbol{w} = (c_1 + k_1)\boldsymbol{v}_1 + \ldots + (c_n + k_n)\boldsymbol{v}_n$是$\boldsymbol{v} + \boldsymbol{w}$关于$S$的**唯一**线性组合表示，因此由坐标向量的定义可得

$$(\boldsymbol{v} + \boldsymbol{w})_S = (c_1 + k_1, \ldots, c_n + k_n) = (c_1, \ldots, c_n) + (k_1, \ldots, k_n) = (\boldsymbol{v})_S + (\boldsymbol{w})_S.$$

类似的，若$\boldsymbol{v} \in V$关于基底$S$的坐标向量为$(\boldsymbol{v})_S = (c_1, \ldots, c_n)$，即$\boldsymbol{v} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n$，那么对于任何实数$k \in \mathbb{R}$，由于$k\boldsymbol{v} = (kc_1)\boldsymbol{v}_1 + \ldots + (kc_n)\boldsymbol{v}_n$，我们有

$$(k\boldsymbol{v})_S = (kc_1, \ldots, kc_n) = k(c_1, \ldots, c_n) = k(\boldsymbol{v})_S.$$

这其实意味着关于基底$S$的坐标映射$\boldsymbol{v} \to (\boldsymbol{v})_S$是一个**线性映射(linear map)**。线性映射这个概念我们将在后续章节中系统学习。这里大家暂时可以忽略这个术语。**但务必牢记坐标映射的以上两个特殊性质**。

另一方面，我们知道$V$里通常存在**多组不同的基底**，且**同一个向量$\boldsymbol{v}$关于不同基底的坐标向量也不同**。比如，令$V = \mathbb{R}^3$，$S = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$为$\mathbb{R}^3$的标准基底，

$S' = \{\boldsymbol{v}_1 = (1,2,1), \boldsymbol{v}_2 = (2,9,0), \boldsymbol{v}_3 = (3,3,4)\}$为$\mathbb{R}^3$的另一组基底。那么对于$\boldsymbol{v} = (5,-1,9)$，我们显然有

$$(\boldsymbol{v})_S = (5,-1,9);$$

而通过关系$(5,-1,9) = c_1(1,2,1)+c_2(2,9,0)+c_3(3,3,4)$我们可以求解出$c_1 = 1, c_2 = -1, c_3 = 2$，即

$$(\boldsymbol{v})_{S'} = (1,-1,2).$$

在本节我们要解决以下问题：

**基底变换问题, The Change-of-Basis Problem:** 假设$V$为$n$维向量空间，$B = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$为$V$的一组基底，$B' = \{\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_n\}$为$V$的另一组基底，那么对于$\boldsymbol{v} \in V$，它关于$B$的坐标向量$(\boldsymbol{v})_B$与关于$B'$的坐标向量$(\boldsymbol{v})_{B'}$满足怎样的关系？

**解答：** 我们将$B' = \{\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_n\}$里的每个向量都用基底$B = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$表示，即

$$\boldsymbol{u}'_1 = a_{11}\boldsymbol{u}_1 + a_{21}\boldsymbol{u}_2 + \ldots + a_{n1}\boldsymbol{u}_n \Leftrightarrow [\boldsymbol{u}'_1]_B = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix};$$

$$\vdots$$

$$\boldsymbol{u}'_i = a_{1i}\boldsymbol{u}_1 + a_{2i}\boldsymbol{u}_2 + \ldots + a_{ni}\boldsymbol{u}_n \Leftrightarrow [\boldsymbol{u}'_i]_B = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix};$$

$$\vdots$$

$$\boldsymbol{u}'_n = a_{1n}\boldsymbol{u}_1 + a_{2n}\boldsymbol{u}_2 + \ldots + a_{nn}\boldsymbol{u}_n \Leftrightarrow [\boldsymbol{u}'_n]_B = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{bmatrix}.$$

假设$[\boldsymbol{v}]_{B'} = \begin{bmatrix} k'_1 \\ k'_2 \\ \vdots \\ k'_n \end{bmatrix}$，即$\boldsymbol{v} = k'_1\boldsymbol{u}'_1 + k'_2\boldsymbol{u}'_2 + \ldots + k'_n\boldsymbol{u}'_n$；将以上$\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_n$关于$B$的线

性组合表示代入该等式，我们得到

$$\begin{aligned}\boldsymbol{v} = k_1'\boldsymbol{u}_1' + k_2'\boldsymbol{u}_2' + \ldots + k_n'\boldsymbol{u}_n' &= k_1'(a_{11}\boldsymbol{u}_1 + a_{21}\boldsymbol{u}_2 + \ldots + a_{n1}\boldsymbol{u}_n) \\ &\quad + \ldots + k_n'(a_{1n}\boldsymbol{u}_1 + a_{2n}\boldsymbol{u}_2 + \ldots + a_{nn}\boldsymbol{u}_n) \\ &= (a_{11}k_1' + a_{12}k_2' + \ldots + a_{1n}k_n')\boldsymbol{u}_1 \\ &\quad + \ldots + (a_{n1}k_1' + a_{n2}k_2' + \ldots + a_{nn}k_n')\boldsymbol{u}_n;\end{aligned}$$

那么如果$[\boldsymbol{v}]_B = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$，即$\boldsymbol{v} = k_1\boldsymbol{u}_1 + k_2\boldsymbol{u}_2 + \ldots + k_n\boldsymbol{u}_n$；由坐标表示的唯一性，必然有

$$\begin{aligned}k_1 &= a_{11}k_1' + a_{12}k_2' + \ldots + a_{1n}k_n', \\ k_2 &= a_{21}k_1' + a_{22}k_2' + \ldots + a_{2n}k_n', \\ &\vdots \\ k_n &= a_{n1}k_1' + a_{n2}k_2' + \ldots + a_{nn}k_n'.\end{aligned}$$

因此，假如我们用

$$P_{B\leftarrow B'} = \begin{bmatrix} [\boldsymbol{u}_1']_B & [\boldsymbol{u}_2']_B & \ldots & [\boldsymbol{u}_n']_B \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}$$

来指代由坐标向量$[\boldsymbol{u}_1']_B, [\boldsymbol{u}_2']_B, \ldots, [\boldsymbol{u}_n']_B$作为列向量的$n$阶方阵，那么以上计算告诉我们

$$[\boldsymbol{v}]_B = P_{B\leftarrow B'}[\boldsymbol{v}]_{B'}.$$

交换$B$与$B'$的角色，令

$$P_{B'\leftarrow B} = \begin{bmatrix} [\boldsymbol{u}_1]_{B'} & [\boldsymbol{u}_2]_{B'} & \ldots & [\boldsymbol{u}_n]_{B'} \end{bmatrix}$$

来指代由坐标向量$[\boldsymbol{u}_1]_{B'}, [\boldsymbol{u}_2]_{B'}, \ldots, [\boldsymbol{u}]_{B'}$作为列向量的$n$阶方阵，那么重复以上计算我们得到

$$[\boldsymbol{v}]_{B'} = P_{B'\leftarrow B}[\boldsymbol{v}]_B.$$

我们称$P_{B\leftarrow B'}$为从$B'$到$B$的转移矩阵(transition matrix from $B'$ to $B$)，称$P_{B'\leftarrow B}$为从$B$到$B'$的转移矩阵(transition matrix from $B$ to $B'$)。

**注意**：英文教材里用$P_{B'\to B}$标记从$B'$到$B$的转移矩阵，用$P_{B\to B'}$标记从$B$到$B'$的转移矩阵。我们在这里使用符号$P_{B\leftarrow B'}$以及$P_{B'\leftarrow B}$是因为这样可以**更自然地表示坐标向量转换的起点与终点**：

$$[\boldsymbol{v}]_B = P_{B\leftarrow B'}[\boldsymbol{v}]_{B'}, \quad [\boldsymbol{v}]_{B'} = P_{B'\leftarrow B}[\boldsymbol{v}]_B.$$

两个转移矩阵$P_{B\leftarrow B'}$与$P_{B'\leftarrow B}$互为逆矩阵：

**Theorem 4.22** (英文教材Theorem 4.6.1)**.** $P_{B\leftarrow B'} = P^{-1}_{B'\leftarrow B}$， $P_{B'\leftarrow B} = P^{-1}_{B\leftarrow B'}$。

证明. 对于任何$\boldsymbol{v}\in V$，我们有

$$[\boldsymbol{v}]_B = P_{B\leftarrow B'}[\boldsymbol{v}]_{B'};$$

$$[\boldsymbol{v}]_{B'} = P_{B'\leftarrow B}[\boldsymbol{v}]_B,$$

因此

$$[\boldsymbol{v}]_B = P_{B\leftarrow B'}[\boldsymbol{v}]_{B'} = P_{B\leftarrow B'}P_{B'\leftarrow B}[\boldsymbol{v}]_B.$$

这意味着，对**任何**$\boldsymbol{b}\in\mathbb{R}^n$ ($\boldsymbol{b}\in\mathbb{R}^n$对应$[\boldsymbol{v}]_B$)，矩阵$A = P_{B\leftarrow B'}P_{B'\leftarrow B}$满足

$$A\boldsymbol{b} = \boldsymbol{b}.$$

容易证明**满足这样的条件的矩阵只可能是单位矩阵**(留作作业题)，即

$$A = P_{B\leftarrow B'}P_{B'\leftarrow B} = I_n.$$

□

如果$V = \mathbb{R}^n$，那么我们可以用以下步骤计算转移矩阵$P_{B'\leftarrow B}$。

1. 对于$B = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}\subset\mathbb{R}^n$，我们不妨用同一个符号$B$来指代由$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$作为列向量的$n$阶方阵：

$$B = \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_n\end{bmatrix}.$$

类似的，若$B' = \{\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_n\}$为$\mathbb{R}^n$的另一组基底，我们用同一个符号$B'$来指代由$\boldsymbol{u}'_1, \ldots, \boldsymbol{u}'_n$作为列向量的$n$阶方阵：

$$B' = \begin{bmatrix}\boldsymbol{u}'_1 & \boldsymbol{u}'_2 & \ldots & \boldsymbol{u}'_n\end{bmatrix}.$$

类似我们在第一章求矩阵的逆的方法，我们将$B'$放在左边，$B$放在右边：

$$[B'|B].$$

2. 同时将**初等行变换**作用在$B'$与$B$上，直到左边的$B'$变成单位矩阵$I_n$，此时将右边$B$经过行变换所得到的的矩阵记为$P_{B'\leftarrow B}$:

$$[B'|B] \Rightarrow [I_n|P_{B'\leftarrow B}].$$

3. 此时右边的矩阵$P_{B'\leftarrow B}$即为所求的转移矩阵。

**对以上算法的数学解释：**按照之前的记号，对于给定的一组$\mathbb{R}^n$里的基底$B=\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$，我们依然用$B$来指代由$\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n$作为列向量的$n$阶方阵：

$$B=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_n\end{bmatrix}.$$

那么如果一个向量$\boldsymbol{v}\in\mathbb{R}^n$关于$B$的坐标向量为$[\boldsymbol{v}]_B=\begin{bmatrix}k_1\\k_2\\\vdots\\k_n\end{bmatrix}$，即$\boldsymbol{v}=k_1\boldsymbol{u}_1+k_2\boldsymbol{u}_2+\ldots+k_n\boldsymbol{u}_n$，那么由矩阵乘法的定义，容易验证实际上

$$\boldsymbol{v}=B[\boldsymbol{v}]_B,$$

以上等式里的$B$显然是指矩阵$B=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_n\end{bmatrix}$。
同理，对于另一组$\mathbb{R}^n$里的基底$B'=\{\boldsymbol{u}'_1,\ldots,\boldsymbol{u}'_n\}$，我们也用同一个符号$B'$来来指代由$\boldsymbol{u}'_1,\ldots,\boldsymbol{u}'_n$作为列向量的$n$阶方阵：

$$B'=\begin{bmatrix}\boldsymbol{u}'_1 & \boldsymbol{u}'_2 & \ldots & \boldsymbol{u}'_n\end{bmatrix}.$$

如果向量$\boldsymbol{v}\in\mathbb{R}^n$关于$B'$的坐标向量为$[\boldsymbol{v}]_{B'}=\begin{bmatrix}k'_1\\k'_2\\\vdots\\k'_n\end{bmatrix}$，即$\boldsymbol{v}=k'_1\boldsymbol{u}'_1+k'_2\boldsymbol{u}'_2+\ldots+k'_n\boldsymbol{u}'_n$，那么由矩阵乘法的定义，我们有

$$\boldsymbol{v}=B'[\boldsymbol{v}]'_B,$$

以上等式里的$B'$显然是指矩阵$B'=\begin{bmatrix}\boldsymbol{u}'_1 & \boldsymbol{u}'_2 & \ldots & \boldsymbol{u}'_n\end{bmatrix}$。
经过以上观察，我们实际上得到等式

$$\boldsymbol{v}=B'[\boldsymbol{v}]'_B=B[\boldsymbol{v}]_B.$$

由于$B$与$B'$都是$\mathbb{R}^n$的基底，由Theorem 4.19可知矩阵$B$与$B'$都可逆，因此对等式

$$B'[\boldsymbol{v}]'_B=B[\boldsymbol{v}]_B$$

两边同乘以$(B')^{-1}$，我们得到

$$[\boldsymbol{v}]'_B = (B')^{-1}B[\boldsymbol{v}]_B.$$

那么由转移矩阵的定义可得$P_{B'\leftarrow B} = (B')^{-1}B$。

现在我们回到以上求转移矩阵$P_{B'\leftarrow B}$的计算步骤。我们知道利用初等行变换将$B'$转化为单位矩阵的过程实际上是在$B'$左边乘上一系列初等矩阵$E_r \dots E_1$，而这个乘积$E_r \dots E_1$等于$(B')^{-1}$；那么同样的行变换作用在矩阵$B$上则得到$E_r \dots E_1 B = (B')^{-1}B$。由于我们已经知道$P_{B'\leftarrow B} = (B')^{-1}B$，我们可以确认之前的计算步骤里所最终得到的右手边的矩阵即是要求的从$B$到$B'$的转移矩阵。

**例子：** 考虑$\mathbb{R}^3$的两组基底$B = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$与$B' = \{\boldsymbol{u}'_1, \boldsymbol{u}'_2, \boldsymbol{u}'_3\}$:

$$\boldsymbol{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{u}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \boldsymbol{u}_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix};$$

$$\boldsymbol{u}'_1 = \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}, \boldsymbol{u}'_2 = \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix}, \boldsymbol{u}'_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}.$$

我们利用之前介绍的计算步骤求解转移矩阵$P_{B'\leftarrow B}$。

1. 此时$B' = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 1 & 0 \\ -5 & -3 & 2 \end{bmatrix}$，$B = \begin{bmatrix} 2 & 2 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$。因此

$$[B'|B] = \left(\begin{array}{ccc|ccc} 3 & 1 & -1 & 2 & 2 & 1 \\ 1 & 1 & 0 & 1 & -1 & 2 \\ -5 & -3 & 2 & 1 & 1 & 1 \end{array}\right).$$

2. 经过初等行变换同时作用在$B$与$B'$上后，我们得到

$$\left(\begin{array}{ccc|ccc} 3 & 1 & -1 & 2 & 2 & 1 \\ 1 & 1 & 0 & 1 & -1 & 2 \\ -5 & -3 & 2 & 1 & 1 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & 2 & \frac{5}{2} \\ 0 & 1 & 0 & -2 & -3 & -\frac{1}{2} \\ 0 & 0 & 1 & 5 & 1 & 6 \end{array}\right).$$

3. $P_{B'\leftarrow B} = \begin{bmatrix} 3 & 2 & \frac{5}{2} \\ -2 & -3 & -\frac{1}{2} \\ 5 & 1 & 6 \end{bmatrix}$.

### 4.7. 行空间/列空间/零空间, Row space, Column space and Null space

对于$m \times n$-矩阵$A$，

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

我们按照之前的记号将其第$i$行$(i = 1, \dots, m)$记为

$$\boldsymbol{r}_i = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix}$$

并称其为$A$的第$i$个行向量(the i-th row vector of $A$)。显然$\boldsymbol{r}_i, i = 1, \dots, m$为$\mathbb{R}^n$里的向量。类似的，$A$的第$j$个列向量(the j-th column vector of $A$)

$$\boldsymbol{c}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}, j = 1, \dots, n$$

是$\mathbb{R}^m$里的向量。

**Definition 4.23** (英文教材237页Definition 2)**.** 对于$m \times n$-矩阵$A$，

- *span*$\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_m\} \subset \mathbb{R}^n$被称为$A$的**行空间*(the row space of* $A$*)***，通常记为*Row*$(A)$；
- *span*$\{\boldsymbol{c}_1, \dots, \boldsymbol{c}_n\} \subset \mathbb{R}^m$被称为$A$的**列空间*(the row space of* $A$*)***，通常记为*Col*$(A)$；
- 齐次线性方程组$A\boldsymbol{x} = \boldsymbol{0}$的解空间*(solution space)*，通常记为*Null*$(A)$:

$$\{\boldsymbol{s} \in \mathbb{R}^n : A\boldsymbol{s} = \boldsymbol{0}\}$$

  被称为$A$的**零空间*(the null space of* $A$*)***。

**Theorem 4.24.** 对任何$m \times n$-矩阵$A$，任何$\boldsymbol{b} \in \mathbb{R}^m$，线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$有解*(consistent)*当且仅当$\boldsymbol{b} \in$ *span*$\{\boldsymbol{c}_1, \dots, \boldsymbol{c}_n\}$，即$\boldsymbol{b}$属于$A$的**列空间**。

证明. 见第五次作业Problem A的第二小题的参考答案。 □

**Theorem 4.25.** 假设$A\boldsymbol{x} = \boldsymbol{b}$有解且$\boldsymbol{s}_0 \in \mathbb{R}^n$为一个**特解*(special solution)***，假设$S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$为$A$的零空间*Null*$(A)$的一组基底。那么$A\boldsymbol{x} = \boldsymbol{b}$的**通解*(general solution)***可以表示为

$$\boldsymbol{s} = \boldsymbol{s}_0 + c_1\boldsymbol{v}_1 + \dots + c_k\boldsymbol{v}_k, \quad c_1, \dots, c_k \in \mathbb{R}.$$

反之，任何$\boldsymbol{s} = \boldsymbol{s}_0 + c_1\boldsymbol{v}_1 + \dots + c_k\boldsymbol{v}_k, \quad c_1, \dots, c_k \in \mathbb{R}$都是$A\boldsymbol{x} = \boldsymbol{b}$的解。

证明. 实际上我们已经在第一章与第三章谈及了该定理的证明。这里我们再复习一下证明过程：

假设$\boldsymbol{s} \in \mathbb{R}^n$是方程组$A\boldsymbol{x} = \boldsymbol{b}$的一个解，即$A\boldsymbol{s} = \boldsymbol{b}$，那么由于$\boldsymbol{s}_0$是$A\boldsymbol{x} = \boldsymbol{b}$的特解，也有$A\boldsymbol{s}_0 = \boldsymbol{b}$，因此$A(\boldsymbol{s} - \boldsymbol{s}_0) = A\boldsymbol{s} - A\boldsymbol{s}_0 = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}$。由$A$的零空间的定义可知，$\boldsymbol{s} - \boldsymbol{s}_0 \in \text{Null}(A)$。由假设，$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$为$A$的零空间的一组基底，从而有$\boldsymbol{s} - \boldsymbol{s}_0 \in \text{span}(S) = \text{Null}(A)$，所以存在系数$c_1, \ldots, c_k \in \mathbb{R}$使得

$$\boldsymbol{s} - \boldsymbol{s}_0 = c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k.$$

因此$A\boldsymbol{x} = \boldsymbol{b}$的通解可以表示为

$$\boldsymbol{s} = \boldsymbol{s}_0 + c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k, \quad c_1, \ldots, c_k \in \mathbb{R}.$$

反之，假设$\boldsymbol{s} = \boldsymbol{s}_0 + c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k$，那么由于$A\boldsymbol{v}_i = \boldsymbol{0}$对任何$i = 1, \ldots, k$(因为$\boldsymbol{v}_i \in \text{Null}(A)$)成立，所以$A\boldsymbol{s} = A\boldsymbol{s}_0 + A(c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k) = \boldsymbol{b} + c_1A\boldsymbol{v}_1 + \ldots + c_kA\boldsymbol{v}_k = \boldsymbol{b} + \boldsymbol{0} = \boldsymbol{b}$。因此任何可以被写成$\boldsymbol{s} = \boldsymbol{s}_0 + c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k$的向量$\boldsymbol{s}$都是方程组$A\boldsymbol{x} = \boldsymbol{b}$的解。 □

本节最重要的内容在于学习如何找到$A$的行空间Row$(A)$，$A$的列空间Col$(A)$，$A$的零空间Null$(A)$的基底。

我们先从寻找**$A$的零空间Null**$(A)$的基底开始。这个问题对应着一类很重要的考试题型。

**例子：(1996年考研真题)**

已知线性方程组

$$\begin{aligned}
&x_1 + x_2 - 2x_3 + 3x_4 = 0 \\
&2x_1 + x_2 - 6x_3 + 4x_4 = -1 \\
&3x_1 + 2x_2 + px_3 + 7x_4 = -1 \\
&x_1 - x_2 - 6x_3 - x_4 = t,
\end{aligned}$$

讨论参数$p, t$取哪些值时方程组有解，无解；当有解时，用**基础解系**(即$\boldsymbol{s} = \boldsymbol{s}_0 + c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k$，$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$为Null$(A)$的一组基底，这样的形式)来表示通解。

**解答：** 对增广矩阵

$$B = \begin{bmatrix} 1 & 1 & -2 & 3 & 0 \\ 2 & 1 & -6 & 4 & -1 \\ 3 & 2 & p & 7 & -1 \\ 1 & -1 & -6 & -1 & t \end{bmatrix}$$

用初等行变换化为阶梯型

$$R=\begin{bmatrix}1&0&-4&1&-1\\0&1&2&2&1\\0&0&p+8&0&0\\0&0&0&0&t+2\end{bmatrix}.$$

显然，如果$t+2\neq 0$，那么该方程组无解。如果$t=-2$，那么方程组有解。
假设$t=-2$，那么当$p=-8$时，以上阶梯型为

$$R=\begin{bmatrix}1&0&-4&1&-1\\0&1&2&2&1\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}.$$

由此我们可以确定$x_1,x_2$为方程组的主元，$x_3,x_4$为自由元。令$x_3=r$，令$x_4=s$，$r,s\in\mathbb{R}$，我们可以求得此时方程组的通解为

$$\boldsymbol{s}=\begin{bmatrix}-1-s+4r\\1-2s-2r\\r\\s\end{bmatrix},\quad r\in\mathbb{R},s\in\mathbb{R}.$$

现在我们将$\boldsymbol{s}$中包含的常数，$r$与$s$全部分离出来：

$$\boldsymbol{s}=\begin{bmatrix}-1-s+4r\\1-2s-2r\\r\\s\end{bmatrix}=\begin{bmatrix}-1-s+4r\\1-2s-2r\\0+0s+r\\0+s+0r\end{bmatrix}=\begin{bmatrix}-1\\1\\0\\0\end{bmatrix}+s\begin{bmatrix}-1\\-2\\0\\1\end{bmatrix}+r\begin{bmatrix}4\\-2\\1\\0\end{bmatrix}.$$

容易验证$\boldsymbol{v}_1=\begin{bmatrix}-1\\-2\\0\\1\end{bmatrix},\boldsymbol{v}_2=\begin{bmatrix}4\\-2\\1\\0\end{bmatrix}$是**Null**$(A)$的一组基底(将在后面的学习中证明该事实，这里请大家直接记住该结论即可)，且$\begin{bmatrix}-1\\1\\0\\0\end{bmatrix}$是方程组的一个特解。因此此时方程组的通解的基础解系为

$$\boldsymbol{s}=\begin{bmatrix}-1\\1\\0\\0\end{bmatrix}+s\begin{bmatrix}-1\\-2\\0\\1\end{bmatrix}+r\begin{bmatrix}4\\-2\\1\\0\end{bmatrix}.$$

假设$t=-2$，那么当$p\neq -8$时，通过对$R$的第三行乘以$\frac{1}{p+8}$我们得到新的阶梯型

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

显然此时方程组的主元为$x_1, x_2, x_3$，自由元为$x_4$。令$x_4 = s$可得通解为

$$\boldsymbol{s} = \begin{bmatrix} -1-s \\ 1-2s \\ 0 \\ s \end{bmatrix} = \begin{bmatrix} -1-s \\ 1-2s \\ 0+0s \\ 0+s \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -1 \\ -2 \\ 0 \\ 1 \end{bmatrix}.$$

接下来我们讨论如何找到$A$**的行空间与列空间**的基底。这里我们再一次需要用到初等行变换与阶梯型。

**Theorem 4.26.** 任何初等行变换都不会改变$A$的行空间$Row(A)$。

证明. 如果对$A$的第$i$行乘以一个非零常数$c$，那么所形成的矩阵$\tilde{A}$的行空间为

$$\text{Row}(\tilde{A}) = \text{span}\{\boldsymbol{r}_1, \ldots, c\boldsymbol{r}_i, \ldots, \boldsymbol{r}_m\}.$$

显然$\text{Row}(\tilde{A}) = \text{span}\{\boldsymbol{r}_1, \ldots, c\boldsymbol{r}_i, \ldots, \boldsymbol{r}_m\} = \text{span}\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_i, \ldots, \boldsymbol{r}_m\} = \text{Row}(A)$，因为

$$\begin{aligned} \boldsymbol{v} \in \text{span}\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_i, \ldots, \boldsymbol{r}_m\} &\Leftrightarrow \boldsymbol{v} = k_1\boldsymbol{r}_1 + \ldots + k_i\boldsymbol{r}_i + \ldots + k_m\boldsymbol{r}_m \\ &\Leftrightarrow \boldsymbol{v} = k_1\boldsymbol{r}_1 + \ldots + \frac{k_i}{c}(c\boldsymbol{r}_i) + \ldots + k_m\boldsymbol{r}_m \\ &\Leftrightarrow \boldsymbol{v} \in \text{span}\{\boldsymbol{r}_1, \ldots, c\boldsymbol{r}_i, \ldots, \boldsymbol{r}_m\}. \end{aligned}$$

类似的，如果$A$的第$i$行乘以一个常数$c$后加到第$j$行，那么所形成的矩阵$\tilde{A}$的行空间为

$$\text{Row}(\tilde{A}) = \text{span}\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_i, \ldots, c\boldsymbol{r}_i + \boldsymbol{r}_j, \ldots, \boldsymbol{r}_m\}.$$

显然$\text{Row}(\tilde{A}) = \text{Row}(A)$，因为

$$\begin{aligned} \boldsymbol{v} \in \text{span}\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_i, \ldots, \boldsymbol{r}_j, \ldots \boldsymbol{r}_m\} &\Leftrightarrow \boldsymbol{v} = k_1\boldsymbol{r}_1 + \ldots + k_i\boldsymbol{r}_i + \ldots + k_j\boldsymbol{r}_j + \ldots + k_m\boldsymbol{r}_m \\ &\Leftrightarrow \boldsymbol{v} = k_1\boldsymbol{r}_1 + \ldots + (k_i - ck_j)\boldsymbol{r}_i + \ldots + k_j(c\boldsymbol{r}_i + \boldsymbol{r}_j) + \ldots \\ &\qquad + k_m\boldsymbol{r}_m \\ &\Leftrightarrow \boldsymbol{v} \in \text{span}\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_i, \ldots, c\boldsymbol{r}_i + \boldsymbol{r}_j, \ldots, \boldsymbol{r}_m\}. \end{aligned}$$

最后，如果交换$A$的第$i$行与第$j$行，显然也不会改变$A$的行空间。具体证明留给大家自行验证。 □

**Theorem 4.27.** 初等行变换可能会改变$A$的列空间$Col(A)$。但任何初等行变换都不会改变$A$的列空间$Col(A)$的维数。

证明. **注意：请大家牢记该定理内容，证明过程仅供感兴趣的同学参考。**
比如矩阵$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$，将$A$的第一行乘以$-2$加到第二行可得矩阵$\tilde{A} = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$。显然$\text{Col}(A) = \text{span}\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix}\} = \{k\begin{bmatrix} 1 \\ 2 \end{bmatrix} : k \in \mathbb{R}\}$不等于$\text{Col}(\tilde{A}) = \text{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \end{bmatrix}\} = \{k\begin{bmatrix} 1 \\ 0 \end{bmatrix} : k \in \mathbb{R}\}$。
现在我们来证明任何初等行变换都不会改变$A$的列空间$\text{Col}(A)$的维数。
假设$\dim(\text{Col}(A)) = k$，且不妨假设$\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$是$\text{Col}(A)$的一组基底(利用Theorem 4.18)。假设$E$为一个初等矩阵，那么$E$对应的行变换作用在$A$上所形成的的矩阵为$EA$。由矩阵乘法定义可知$EA$的列向量表示为

$$EA = \begin{bmatrix} E\boldsymbol{c}_1 & \ldots & E\boldsymbol{c}_n \end{bmatrix}.$$

不难证明$\{E\boldsymbol{c}_1, \ldots, E\boldsymbol{c}_k\}$是一个线性无关集合：如果$d_1, \ldots, d_k$使得

$$d_1 E\boldsymbol{c}_1 + \ldots + d_k E\boldsymbol{c}_k = \mathbf{0}$$

那么对等式两边同时左乘$E^{-1}$，可得

$$E^{-1}(d_1 E\boldsymbol{c}_1 + \ldots + d_k E\boldsymbol{c}_k) = \mathbf{0} \Rightarrow d_1\boldsymbol{c}_1 + \ldots + d_k\boldsymbol{c}_k = \mathbf{0}.$$

由于$\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$是$\text{Col}(A)$的一组基底，其线性无关性导致以上等式成立必须要求$d_1 = \ldots = d_k = 0$。这意味着$d_1 E\boldsymbol{c}_1 + \ldots + d_k E\boldsymbol{c}_k = \mathbf{0}$只有可能在$d_1 = \ldots = d_k = 0$时成立，即$\{E\boldsymbol{c}_1, \ldots, E\boldsymbol{c}_k\}$是一个线性无关集合。
由于$\{E\boldsymbol{c}_1, \ldots, E\boldsymbol{c}_k\}$是一个线性无关集合，我们得到

$$\dim(\text{Col}(EA)) \geq k = \dim(\text{Col}(A)).$$

另一方面，对$EA$左乘$E^{-1}$(对应着一个行变换)，以上分析告诉我们

$$k = \dim(\text{Col}(A)) = \dim(\text{Col}(E^{-1}EA)) \geq \dim(\text{Col}(EA)).$$

综上，$k = \dim(\text{Col}(A)) = \dim(\text{Col}(EA))$。 □

以下定理非常重要。同样，**请大家务必牢记定理内容，但可以忽略证明过程。**

**Theorem 4.28.** 对$m \times n$-矩阵$A$，令$R$为$A$的一个阶梯型。那么

1. $R$里所有含有主1(*leading one*)的行向量组成$A$的行空间$Row(A)$的一组基底。
2. $R$所有含有主1(*leading one*)的列向量组成$R$的列空间$Col(R)$的一组基底。

证明. 令$R$为$A$的一个阶梯型。假设包含主1的$R$的行向量为$\boldsymbol{u}_1,\dots,\boldsymbol{u}_k$，假设$\boldsymbol{u}_1$的主1所在的列为第$j_1$列，...，$\boldsymbol{u}_k$的主1所在的列为第$j_k$列，即

$$\boldsymbol{u}_1=\begin{bmatrix}b_{11} & b_{12} & \dots & b_{1n}\end{bmatrix},\quad b_{1j_1}=1, b_{1j}=0\text{若}j<j_1;$$

$$\vdots$$

$$\boldsymbol{u}_k=\begin{bmatrix}b_{k1} & b_{k2} & \dots & b_{kn}\end{bmatrix},\quad b_{kj_k}=1, b_{kj}=0\text{若}j<j_k.$$

这里注意由阶梯型的定义，我们必然有$j_1<j_2<\dots<j_k$。现在考虑等式

$$c_1\boldsymbol{u}_1+\dots+c_k\boldsymbol{u}_k=\mathbf{0}$$

即

$$\begin{bmatrix}c_1b_{11}+\dots c_kb_{k1} & \dots & c_1b_{1n}+\dots c_kb_{kn}\end{bmatrix}=\begin{bmatrix}0 & \dots & 0\end{bmatrix}.$$

显然，以上等式意味着$c_1b_{1j_1}+\dots c_kb_{kj_1}=0$ (即$\begin{bmatrix}c_1b_{11}+\dots c_kb_{k1} & \dots & c_1b_{1n}+\dots c_kb_{kn}\end{bmatrix}$的第$j_1$个坐标需要等于0)。由于$b_{1j_1}=1$，$b_{2j_1}=\dots=b_{kj_1}=0$(已知$b_{ij}=0$对任何$j<j_i$，那么对$i=2,\dots,k$，由于$j_1<j_2<\dots<j_k$，$b_{2j_1}=\dots=b_{kj_1}=0$)，我们实际上有

$$c_1b_{1j_1}+\dots c_kb_{kj_1}=0\Rightarrow c_1\times 1+c_2\times 0+\dots+c_k\times 0=c_1=0.$$

因此等式$c_1\boldsymbol{u}_1+\dots+c_k\boldsymbol{u}_k=\mathbf{0}$成立的必要条件是$c_1=0$，即我们实际上有

$$c_2\boldsymbol{u}_2+\dots+c_k\boldsymbol{u}_k=\mathbf{0}.$$

现在对$c_2\boldsymbol{u}_2+\dots+c_k\boldsymbol{u}_k=\mathbf{0}$使用以上分析，可得$c_2\boldsymbol{u}_2+\dots+c_k\boldsymbol{u}_k=\mathbf{0}$成立的必要条件是$c_2=0$。以此类推，最后我们得到$c_1\boldsymbol{u}_1+\dots+c_k\boldsymbol{u}_k=\mathbf{0}$成立的必要条件是$c_1=c_2=\dots=c_k=0$，即$\boldsymbol{u}_1,\dots,\boldsymbol{u}_k$线性无关。
另一方面，有阶梯性的定义，对$R$的其他行向量$\boldsymbol{u}_{k+1},\dots,\boldsymbol{u}_m$，我们必然有$\boldsymbol{u}_{k+1}=\dots=\boldsymbol{u}_m=\mathbf{0}$(因为它们不包含主1)，所以$\text{Row}(R)=\text{span}\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_m\}=\text{span}\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_k\}$。
综上，我们得到$R$里所有含有主1的行向量组成$R$的行空间Row$(R)$的一组基底。又由Theorem 4.26，我们有$\text{Row}(R)=\text{Row}(A)$，因此$R$里所有含有主1的行向量组成$A$的行空间Row$(A)$的一组基底。
利用同样的证明思路我们可以验证$R$所有含有主1的列向量组成$R$的列空间Col$(R)$的一组基底，但过程更加繁琐一点。具体细节留给感兴趣的同学自行补全。 □

**例子：**

1. 从$S=\{\boldsymbol{v}_1=(1,-2,0,3),\boldsymbol{v}_2=(2,-5,-3,6),\boldsymbol{v}_3=(0,1,3,0),\boldsymbol{v}_4=(2,-1,4,-7),\boldsymbol{v}_5=(5,-8,1,2)\}$里找出一个子集$M$使得$M$是span$(S)$的一组基底。
2. 将$S-M$里的不属于$M$的向量表示为基底$M$的线性组合。

**答案：**

1. 将$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_5$作为列向量组成矩阵

$$A=\begin{bmatrix}1&2&0&2&5\\-2&-5&1&-1&-8\\0&-3&3&4&1\\3&6&0&-7&2\end{bmatrix}.$$

因此现在的任务为寻找$A$的列空间的一组基底。基于Theorem 4.28我们需要先将$A$通过行变换化为阶梯型：

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

由Theorem 4.28我们知道$R$里含有主1的三个列向量$\boldsymbol{w}_1=\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\boldsymbol{w}_2=\begin{bmatrix}0\\1\\0\\0\end{bmatrix},\boldsymbol{w}_4=\begin{bmatrix}0\\0\\1\\0\end{bmatrix}$组成$R$的列空间的基底。从Theorem 4.27的证明中我们知道$A$**里对应的三个列向量**$M=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_4\}$**组成**$A$**的列空间的一组基底。**

2. 由第一问可知$S-M=\{\boldsymbol{v}_3,\boldsymbol{v}_5\}$。令$\boldsymbol{v}_3=a\boldsymbol{v}_1+b\boldsymbol{v}_2+c\boldsymbol{v}_4$，代入具体坐标并解方程组可得

$$a=2,b=-1,c=0.$$

类似可计算得到$\boldsymbol{v}_5=\boldsymbol{v}_1+\boldsymbol{v}_2+\boldsymbol{v}_4$。

## 4.8. 秩/零化度/矩阵的基本空间, Rank, Nullity and Fundamental Matrix Spaces

由上一节的Theorem 4.28以及Theorem 4.27我们知道对于任何$m\times n$-矩阵$A$，令$R$为它的一个阶梯型，那么有

$$\text{dim}(\text{Row}(A))=\text{dim}(\text{Row}(R))=R\text{里主1的个数}$$

$$\dim(\mathrm{Col}(A)) = \dim(\mathrm{Col}(R)) = R\text{里主1的个数},$$

因为阶梯型里每一行/每一列都包含至多一个主1。

**Definition 4.29.** 对于任何$m \times n$-矩阵$A$，令$R$为它的一个阶梯型，我们定义

- $\mathit{rank}(A) = \mathit{dim}(\mathit{Row}(A)) = \mathit{dim}(\mathit{Col}(A)) = R$里主*1*的个数，被称为$A$的秩*(rank)*；

- $\mathit{nullity}(A) = \mathit{dim}(\mathit{Null}(A))$，被称为$A$的零化度*(Nullity)*。

从以上定义我们可以立刻看出，对于任何$m \times n$-矩阵$A$，我们有

$$\mathrm{rank}(A) \leq \min(n, m), \quad \mathrm{nullity}(A) \leq n.$$

**以下定理是线性代数中最具代表性的定理**。它的证明将在下一章给出，现在我们只需牢记定理内容。

**Theorem 4.30** (Rank-Nullity Theorem/Dimension Theorem for Matrices; 英文教材250页Theorem 4.8.2)**.** 对于任何$m \times n$-矩阵$A$，有

$$\mathit{rank}(A) + \mathit{nullity}(A) = n.$$

对于一个给定的$m \times n$-矩阵$A$，令$R$为$A$的一个阶梯型。由秩的定义我们知道$\mathrm{rank}(A) = R$里主1的个数，**即$\mathbf{rank}(A)$等于齐次线性方程组$A\boldsymbol{x} = \mathbf{0}$里主元的个数**。由于$A$为$m \times n$-矩阵，齐次线性方程组$A\boldsymbol{x} = \mathbf{0}$里含有$n$个未知数。因此，由以上Rank-Nullity Theorem (Theorem 4.30)，即等式$\mathrm{rank}(A) + \mathrm{nullity}(A) = n$，我们实际上得到

$$\mathrm{nullity}(A) = n - \mathrm{rank}(A) = \text{方程组}A\boldsymbol{x} = \mathbf{0}\text{里自由元的个数}.$$

**Definition 4.31.** 对$m \times n$-矩阵$A$，以下六个向量空间被称为$A$的基本空间*(the fundamental spaces of $A$)*:

- $A$的行空间$\mathit{Row}(A)$,
- $A$的列空间$\mathit{Col}(A)$,
- $A^\top$的行空间$\mathit{Row}(A^\top)$,
- $A^\top$的列空间$\mathit{Col}(A^\top)$,
- $A$的零空间$\mathit{Null}(A)$,
- $A^\top$的零空间$\mathit{Null}(A^\top)$。

由以上定义可知：

- $A$的行空间Row$(A)$等于$A^\top$的列空间Col$(A^\top)$，且它们都是$\mathbb{R}^n$里的子空间；
- $A$的列空间Col$(A)$等于$A^\top$的行空间Row$(A^\top)$，且它们都是$\mathbb{R}^m$里的子空间；
- Null$(A)$是$\mathbb{R}^n$里的子空间；
- Null$(A^\top)$是$\mathbb{R}^m$里的子空间；
- rank$(A)$ = rank$(A^\top)$，因为

$$\text{rank}(A) = \text{dim}(\text{Row}(A)) = \text{dim}(\text{Col}(A^\top)) = \text{rank}(A^\top).$$

除此之外，我们容易得到以下等价关系

**Theorem 4.32** (英文教材254页 Theorem 4.8.8)**.** 对于$n$阶方阵$A$，以下说法等价

- $A$可逆；
- $A\boldsymbol{x} = \mathbf{0}$只有平凡解；
- $A$的简化阶梯型为单位矩阵；
- $A$是一组初等矩阵的乘积；
- $A\boldsymbol{x} = \boldsymbol{b}$对任何$n \times 1$的列向量$\boldsymbol{b}$都有解；
- $A\boldsymbol{x} = \boldsymbol{b}$对任何$n \times 1$的列向量$\boldsymbol{b}$都有且只有一个解；
- $\det(A) \neq 0$；
- $A$的所有$n$个行向量线性无关；
- $A$的所有$n$个列向量线性无关；
- *span*$(Row(A)) = \mathbb{R}^n$；
- *span*$(Col(A)) = \mathbb{R}^n$；
- $A$的所有$n$个行向量构成$\mathbb{R}^n$的一组基底；
- $A$的所有$n$个列向量构成$\mathbb{R}^n$的一组基底；
- *rank*$(A) = n$；
- *Null*$(A) = 0$。

现在我们利用本节学习的知识研究$\mathbb{R}^2, \mathbb{R}^3$**里由方程确定的直线/平面**。这部分内容属于**英文教材的第3.3节**。
首先我们给出以下几何物体的数学定义：

- 平面(plane)$\mathbb{R}^2$里的每一条直线(line)都可以表示为：

$$\{(\bar{x}, \bar{y}) + k(v_1, v_2) : k \in \mathbb{R}\}, \quad (v_1, v_2) \neq \mathbf{0}.$$

  显然，如果$(\bar{x}, \bar{y}) = (0, 0)$，那么这条直线穿过原点且因此表示了一个一维子空间；容易验证$\tan(\theta) = \frac{v_2}{v_1}$给出这条直线的斜率。
- 设$\mathbb{R}^2$里的一条直线表示为$\{(\bar{x}, \bar{y}) + k(v_1, v_2) : k \in \mathbb{R}\}$。如果一个向量$\boldsymbol{n} \in \mathbb{R}^2, \boldsymbol{n} \neq \mathbf{0}$满足

$$\boldsymbol{n} \cdot (v_1, v_2) = 0$$

那么我们称$\boldsymbol{n}$为这条直线的**法向量(normal)**。

- 立体空间$\mathbb{R}^3$里的每一个平面(plane)都可以表示为：

$$\{(\bar{x},\bar{y},\bar{z})+k_1(v_1,v_2,v_3)+k_2(w_1,w_2,w_3):k_1,k_2\in\mathbb{R})\},\quad (v_1,v_2,v_3)\text{与}(w_1,w_2,w_3)\text{ 线性无关.}$$

显然，如果$(\bar{x},\bar{y},\bar{z})=(0,0,0)$，那么这个平面穿过原点且因此表示了一个二维子空间，且此时这个平面被由$(0,0,0)$，$(v_1,v_2,v_3)$，$(w_1,w_2,w_3)$作为顶点的三角形唯一确定。

- 设$\mathbb{R}^3$里的一个平面表示为

$$\{(\bar{x},\bar{y},\bar{z})+k_1(v_1,v_2,v_3)+k_2(w_1,w_2,w_3):k_1,k_2\in\mathbb{R})\},\quad (v_1,v_2,v_3)\text{与}(w_1,w_2,w_3)\text{ 线性无关.}$$

如果一个向量$\boldsymbol{n}\in\mathbb{R}^3,\boldsymbol{n}\neq\boldsymbol{0}$满足

$$\boldsymbol{n}\cdot(v_1,v_2,v_3)=0,\quad \boldsymbol{n}\cdot(w_1,w_2,w_3)=0$$

那么我们称$\boldsymbol{n}$为这个平面的**法向量(normal)**。

**Theorem 4.33** (英文教材157页Theorem 3.3.1)**.** *1. 若$(a,b)\in\mathbb{R}^2$，$(a,b)\neq(0,0)$，$c\in\mathbb{R}$，那么方程*

$$ax+by+c=0$$

*的所有解组成的集合是$\mathbb{R}^2$里的一条直线，且这条直线的法向量为$\boldsymbol{n}=(a,b)$。我们简称这条直线为“直线$ax+by+c=0$”。*

*2. 若$(a,b,c)\in\mathbb{R}^3$，$(a,b,c)\neq(0,0,0)$，$d\in\mathbb{R}$，那么方程*

$$ax+by+cz+d=0$$

*的所有解组成的集合是$\mathbb{R}^3$里的一个平面，且这个平面的法向量为$\boldsymbol{n}=(a,b,c)$。我们简称这个平面为“平面$ax+by+cz+d=0$”。*

证明. 1. 由于已经假设$(a,b)\neq(0,0)$，不妨令$a\neq0$，此时方程$ax+by+c=0\Leftrightarrow ax+by=-c$必然有解，比如$(\bar{x},\bar{y})=(\frac{-c}{a},0)$是这个方程的一个特解。因此，利用Theorem 4.25我们知道这个方程的通解可以被表示为

$$(x,y)=(\bar{x},\bar{y})+k(v_1,v_2),$$

这里向量$(v_1,v_2)$是齐次线性方程$ax+by=0$的(任意)一个解，同时也是$1\times2$-矩阵$A=\begin{bmatrix}a & b\end{bmatrix}$的零空间的一组基底。注意，这里我们使用了以下论点：

- 齐次线性方程组$ax+by=0$的系数矩阵为$1\times2$-矩阵$A=\begin{bmatrix}a & b\end{bmatrix}$。由假设，$(a,b)\neq(0,0)$，因此$A$的秩rank$(A)=\dim(\text{Row}(A))=1$。那么由Rank-Nullity Theorem，即Theorem 4.30，我们知道nullity$(A)=2-1=1$，即$\dim(\text{Null}(A))=1$。我们取$(v_1,v_2)\neq(0,0)$为Null$(A)$的一组基底，那么有span$\{(v_1,v_2)\}=\{k(v_1,v_2):k\in\mathbb{R}\}=\text{Null}(A)$。

由平面中直线的数学表示我们知道，方程$ax+by+c=0 \Leftrightarrow ax+by=-c$的通解组成的集合

$$\{(x,y)=(\bar{x},\bar{y})+k(v_1,v_2): k\in\mathbb{R}\}$$

是$\mathbb{R}^2$里的一条直线；且由于$(v_1,v_2)$是$ax+by=0$的一个解，我们还有$av_1+bv_2=(a,b)\cdot(v_1,v_2)=0$，因此由法向量的定义可知$\boldsymbol{n}=(a,b)$是这条直线的法向量。

2. 由于已经假设$(a,b,c)\neq(0,0,0)$，不妨令$a\neq 0$，此时方程$ax+by+cz+d=0 \Leftrightarrow ax+by+cz=-d$必然有解，比如$(\bar{x},\bar{y},\bar{z})=(\frac{-c}{a},0,0)$是这个方程的一个特解。因此，利用Theorem 4.25我们知道这个方程的通解可以被表示为

$$(x,y,z)=(\bar{x},\bar{y},\bar{z})+k_1(v_1,v_2,v_3)+k_2(w_1,w_2,w_3),$$

这里向量$(v_1,v_2,v_3)$，$(w_1,w_2,w_3)$是齐次线性方程$ax+by+cz=0$的两个解，同时也是$1\times 3$-矩阵$A=\begin{bmatrix}a & b & c\end{bmatrix}$的零空间的一组基底。与之前的证明类似，这里我们使用以下论点：

- 齐次线性方程组$ax+by+cz=0$的系数矩阵为$1\times 3$-矩阵$A=\begin{bmatrix}a & b & c\end{bmatrix}$。由假设，$(a,b,c)\neq(0,0,0)$，因此$A$的秩$\text{rank}(A)=\dim(\text{Row}(A))=1$。那么由Rank-Nullity Theorem，即Theorem 4.30，我们知道$\text{nullity}(A)=3-1=2$，即$\dim(\text{Null}(A))=2$。我们取$(v_1,v_2,v_3),(w_1,w_2,w_3)$为$\text{Null}(A)$的一组基底，那么有$\text{span}\{(v_1,v_2,v_3),(w_1,w_2,w_3)\}=\{k_1(v_1,v_2,v_3)+k_2(w_1,w_2,w_3): k_1,k_2\in\mathbb{R}\}=\text{Null}(A)$。

由立体空间中平面的数学表示我们知道，方程$ax+by+cz+d=0 \Leftrightarrow ax+by+cz=-d$的通解组成的集合

$$\{(x,y,z)=(\bar{x},\bar{y},\bar{z})+k_1(v_1,v_2,v_3)+k_2(w_1,w_2,w_3): k_1,k_2\in\mathbb{R}\}$$

是$\mathbb{R}^3$里的一个平面(注意$(v_1,v_2,v_3)$与$(w_1,w_2,w_3)$线性无关)；
且由于$(v_1,v_2,v_3),(w_1,w_2,w_3)$是$ax+by+cz=0$的解，我们还有$av_1+bv_2+cv_3=(a,b,c)\cdot(v_1,v_2,v_3)=0$，$aw_1+bw_2+cw_3=(a,b,c)\cdot(w_1,w_2,w_3)=0$。因此由法向量的定义可知$\boldsymbol{n}=(a,b,c)$是这个平面的法向量。

□

**Definition 4.34.** 假设$H\subset\mathbb{R}^3$是立体空间中的一个平面，且它的方程为

$$ax+by+cz+d=0.$$

令$(\bar{x},\bar{y},\bar{z})\in\mathbb{R}^3$为该平面$H$上的一个点，即$(\bar{x},\bar{y},\bar{z})$为方程$ax+by+cz+d=0.$的一个特解，满足$a\bar{x}+b\bar{y}+c\bar{z}+d=0$。那么我们称

$$a(x-\bar{x})+b(y-\bar{y})+c(z-\bar{z})=0$$

为平面$H$的*point–norm form*.

注意，如果$(x,y,z)\in\mathbb{R}^3$满足

$$a(x-\bar{x})+b(y-\bar{y})+c(z-\bar{z})=0$$

那么由于$a\bar{x}+b\bar{y}+c\bar{z}+d=0\Leftrightarrow a\bar{x}+b\bar{y}+c\bar{z}=-d$，我们有

$$a(x-\bar{x})+b(y-\bar{y})+c(z-\bar{z})=0\Rightarrow ax+by+cz-(a\bar{x}+b\bar{y}+c\bar{z})=ax+by+cz+d=0,$$

即$(x,y,z)\in H$。

显然，如果我们要确定$\mathbb{R}^3$里的一个平面$H$的point–norm form，我们需要知道两件事：**一是$H$的法向量$\boldsymbol{n}=(a,b,c)$，二是$H$里面包含的一个点的坐标$(\bar{x},\bar{y},\bar{z})$**。

**以下定理在过去的考试中以填空题的形式出现，因此请大家务必牢记计算公式。**

**Theorem 4.35** (英文教材161页Theorem 3.3.4). *1.* 设$(a,b)\neq(0,0)$，$P_0=(x_0,y_0)\in\mathbb{R}^2$为$\mathbb{R}^2$里的一个点。那么$P_0$与直线$ax+by+c=0$的距离$D$满足

$$D=\frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}.$$

2. 设$(a,b,c)\neq(0,0,0)$，$P_0=(x_0,y_0,z_0)\in\mathbb{R}^3$为$\mathbb{R}^3$里的一个点。那么$P_0$与平面$ax+by+cz+d=0$的距离$D$满足

$$D=\frac{|ax_0+by_0+cz_0+d|}{\sqrt{a^2+b^2+c^2}}.$$

*证明.* 我们证明第二项，第一项的证明留给大家自行验证。

取平面$ax+by+cz+d=0$上的任意一点$Q=(x_1,y_1,z_1)$。那么从$Q$指向$P_0$的向量为

$$\overrightarrow{QP_0}=P_0-Q=(x_0-x_1,y_0-y_1,z_0-z_1).$$

由Theorem 4.33我们知道平面$ax+by+cz+d=0$的法向量为$\boldsymbol{n}=(a,b,c)$，因此画图后易知**从$P_0$到平面$ax+by+cz+d=0$的距离$D$为向量$\overrightarrow{QP_0}$到法向量$\boldsymbol{n}$的正交投影的长度**，即

$$D=\|\mathrm{proj}_{\boldsymbol{n}}(\overrightarrow{QP_0})\|=\frac{\overrightarrow{QP_0}\cdot\boldsymbol{n}}{\|\boldsymbol{n}\|}.$$

代入$\overrightarrow{QP_0}\cdot\boldsymbol{n}=a(x_0-x_1)+b(y_0-y_1)+c(z_0-z_1)$，$\|\boldsymbol{n}\|=\sqrt{a^2+b^2+c^2}$，得到

$$D=\frac{\overrightarrow{QP_0}\cdot\boldsymbol{n}}{\|\boldsymbol{n}\|}=\frac{|a(x_0-x_1)+b(y_0-y_1)+c(z_0-z_1)|}{\sqrt{a^2+b^2+c^2}}.$$

又因为$Q=(x_1,y_1,z_1)$是平面$ax+by+cz+d=0$上的一点，即$-(ax_1+by_1+cz_1)=d$，代入以上$D$的表达式我们得到

$$D=\frac{|ax_0+by_0+cz_0+d|}{\sqrt{a^2+b^2+c^2}}.$$

□

**例子：(2021年线性代数期中考试题)**

In $\mathbb{R}^3$, suppose that $\Pi$ is the plane given by $2x - y - 3z + 2 = 0$. The distance between the point $P_0 = (1, -1, 0)$ and the plane $\Pi$ is ?

**答案：** 注意$\boldsymbol{n} = (a, b, c) = (2, -1, -3)$，$d = 2$，将其与$P_0 = (x_0, y_0, z_0) = (1, -1, 0)$的坐标直接套入公式：

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}$$

可以计算得到答案为:$\frac{5}{\sqrt{14}}$。

**关于矩阵的基本空间与秩(rank)的一些补充：**

**Theorem 4.36.** 令$A \in M_{m\times n}$，令$W = \{A\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^n\}$。那么有$W = \mathit{Col}(A)$。

证明. 将$A$的$n$个列向量依次表示为$\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n \in \mathbb{R}^m$。对于任何$\boldsymbol{v} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$，由矩阵乘法定义可知

$$A\boldsymbol{v} = k_1\boldsymbol{c}_1 + k_2\boldsymbol{c}_2 + \ldots + k_n\boldsymbol{c}_n \in \text{span}\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\} = \text{Col}(A).$$

因此我们有$W = \{A\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^n\} \subset \text{Col}(A)$。反之，对于任何$\boldsymbol{w} \in \text{Col}(A) = \text{span}\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}$，我们知道存在系数$k_1, k_2, \ldots, k_n$使得$\boldsymbol{w} = k_1\boldsymbol{c}_1 + k_2\boldsymbol{c}_2 + \ldots + k_n\boldsymbol{c}_n = A\boldsymbol{v}$，这里$\boldsymbol{v} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} \in \mathbb{R}^n$。这意味着$\text{Col}(A) \subset W$。综上，我们得到

$$W \subset \text{Col}(A) \subset W \Rightarrow W = \text{Col}(A).$$

□

**Theorem 4.37.** 令$A \in M_{m\times n}, B \in M_{n\times k}$，那么必然有

$$\mathit{rank}(AB) \leq \mathit{rank}(A).$$

证明. 由于rank$(AB) = \dim(\text{Col}(AB))$，rank$(A) = \dim(\text{Col}(A))$，要证明rank$(AB) \leq$ rank$(A)$只需证明Col$(AB) \subset$ Col$(A)$。利用以上Theorem 4.36 我们知道

$$\text{Col}(AB) = \{(AB)\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^k\},$$

$$\text{Col}(A) = \{A\boldsymbol{w} : \boldsymbol{w} \in \mathbb{R}^n\}.$$

显然，对于任何$\boldsymbol{v} \in \mathbb{R}^k$，我们有$\boldsymbol{w} = B\boldsymbol{v} \in \mathbb{R}^n$，因此

$$(AB)\boldsymbol{v} = A(B\boldsymbol{v}) = A\boldsymbol{w} \in \text{Col}(A),$$

即$\{(AB)\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^k\} \subset \{A\boldsymbol{w} : \boldsymbol{w} \in \mathbb{R}^n\}$。 □

**Theorem 4.38.** 对于任何$A \in M_{n\times n}$，都有

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

证明. 利用Theorem 4.30，即

$$\text{rank}(A) + \text{nullity}(A) = n, \quad \text{rank}(A^\top A) + \text{nullity}(A^\top A) = n,$$

要证明rank$(A^\top A)$ = rank$(A)$，我们只需证明nullity$(A^\top A)$ = nullity$(A)$。实际上我们可以证明Null$(A^\top A)$ = Null$(A)$：首先，对任何$\boldsymbol{x} \in$ Null$(A)$，即$A\boldsymbol{x} = \boldsymbol{0}$，必然有$(A^\top A)\boldsymbol{x} = A^\top(A\boldsymbol{x}) = A^\top\boldsymbol{0} = \boldsymbol{0}$，即$\boldsymbol{x} \in$ Null$(A^\top A)$，从而有Null$(A) \subset$ Null$(A^\top A)$。反之，令$\boldsymbol{x} \in$ Null$(A^\top A)$，即$A^\top A\boldsymbol{x} = \boldsymbol{0}$。那么我们必然有

$$\boldsymbol{0} = \boldsymbol{x}^\top \underbrace{A^\top A\boldsymbol{x}}_{=\boldsymbol{0}} = (A\boldsymbol{x})^\top(A\boldsymbol{x}) = \|A\boldsymbol{x}\|^2,$$

从而得到$A\boldsymbol{x} = \boldsymbol{0}$，即$\boldsymbol{x} \in$ Null$(A)$。这意味着Null$(A^\top A) \subset$ Null$(A)$。综上，我们得到

$$\text{Null}(A^\top A) \subset \text{Null}(A) \subset \text{Null}(A^\top A) \Rightarrow \text{Null}(A^\top A) = \text{Null}(A).$$

□

**关于维数公式的一个补充：**

**Theorem 4.39.** 令$U$为一个向量空间，$V, W \subset U$为两个$U$里的有限维子空间。那么有

$$\textit{dim}(V + W) = \textit{dim}(V) + \textit{dim}(W) - \textit{dim}(V \cap W).$$

证明. 该证明与第七次作业Bonus题目的参考答案一致。假设$V \cap W$的一组基底为$S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$。利用Theorem 4.20我们通过向$S$里加入向量$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$使得$M = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$为$V$的一组基底；通过向$S$里加入向量$\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_j\}$使得$B =$

$\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_r,\boldsymbol{w}_1,\dots,\boldsymbol{w}_j\}$为$W$的一组基底。注意此时我们实际上假设$\dim(V)=r+k$，$\dim(W)=r+j$，$\dim(V\cap W)=r$。此外，注意$\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}\cap(V\cap W)=\{\boldsymbol{0}\}$，因为增加的这些向量$\boldsymbol{v}_1,\dots,\boldsymbol{v}_k$都与$V\cap W$的基底$S$线性无关: 如果$\boldsymbol{u}\in\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}\cap(V\cap W)$，那么$\boldsymbol{u}=c_1\boldsymbol{v}_1+\dots+c_k\boldsymbol{v}_k$同时$\boldsymbol{u}=k_1\boldsymbol{u}_1+\dots+k_r\boldsymbol{u}_r$，即$\boldsymbol{u}-\boldsymbol{u}=\boldsymbol{0}=c_1\boldsymbol{v}_1+\dots+c_k\boldsymbol{v}_k-k_1\boldsymbol{u}_1-\dots-k_r\boldsymbol{u}_r$，由线性无关性可得唯一可能为$c_1=\dots=c_k=k_1=\dots=k_r=0$，即$\boldsymbol{u}=0\boldsymbol{v}_1+\dots+0\boldsymbol{v}_k=\boldsymbol{0}$。同理可得$\text{span}\{\boldsymbol{w}_1,\dots,\boldsymbol{w}_j\}\cap(V\cap W)=\{\boldsymbol{0}\}$。
另一方面，我们容易验证$M\cup B=\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_r,\boldsymbol{v}_1,\dots,\boldsymbol{v}_k,\boldsymbol{w}_1,\dots,\boldsymbol{w}_j\}$是$V+W$的一组基底：$\text{span}(M\cup B)=V+W$显然成立；考虑等式

$$k_1\boldsymbol{u}_1+\dots+k_r\boldsymbol{u}_r+c_1\boldsymbol{v}_1+\dots+c_k\boldsymbol{v}_k+d_1\boldsymbol{w}_1+\dots+d_j\boldsymbol{w}_j=\boldsymbol{0},$$

若该等式成立，则有

$$c_1\boldsymbol{v}_1+\dots+c_k\boldsymbol{v}_k=-(\boldsymbol{u}_1+\dots+k_r\boldsymbol{u}_r+d_1\boldsymbol{w}_1+\dots+d_j\boldsymbol{w}_j),$$

显然等式左边属于$\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}$，等式右边属于$W$，因此该等式成立意味着$c_1\boldsymbol{v}_1+\dots+c_k\boldsymbol{v}_k=-(\boldsymbol{u}_1+\dots+k_r\boldsymbol{u}_r+d_1\boldsymbol{w}_1+\dots+d_j\boldsymbol{w}_j)\in\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}\cap W$。然而我们之前已经证明$\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}\cap(V\cap W)=(\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}\cap V)\cap W=\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}\cap W=\{\boldsymbol{0}\}$，因此必有$c_1\boldsymbol{v}_1+\dots+c_k\boldsymbol{v}_k=-(\boldsymbol{u}_1+\dots+k_r\boldsymbol{u}_r+d_1\boldsymbol{w}_1+\dots+d_j\boldsymbol{w}_j)\in\text{span}\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_k\}=\boldsymbol{0}$。利用这些向量的线性无关性可得$c_1=\dots=c_k=0,k_1=\dots=k_r=0,d_1=\dots+d_j=0$。因此证$M\cup B=\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_r,\boldsymbol{v}_1,\dots,\boldsymbol{v}_k,\boldsymbol{w}_1,\dots,\boldsymbol{w}_j\}$线性无关。
综上，我们得到

$$\dim(V+W)=r+k+j$$

等于

$$\dim(V)+\dim(W)-\dim(V\cap W)=r+k+r+j-r=r+k+j.$$

□

**例子：(第七次作业Bonus)** Let $U$ be a vector space of dimension 6. Let $V$ and $W$ be two subspaces of $U$ such that $\dim(V)=2$, $\dim(W)=3$. Find all possible dimensions of $V+W$, and explain your argument.

**答案：** 由以上维数公式，有$\dim(V+W)=\dim(V)+\dim(W)-\dim(V\cap W)=2+3-\dim(V\cap W)$。由于$\dim(V\cap W)\subset\dim(V)$可能取值为$0,1,2$，因此$\dim(V+W)$可能取值为$5-0=5,5-1=4,5-2=3$。

**关于正交补的一些补充：**

**Definition 4.40.** 设$W \subset \mathbb{R}^n$为一个子空间。令

$$W^{\perp} = \{\boldsymbol{v} \in \mathbb{R}^n : \boldsymbol{v} \cdot \boldsymbol{w} = 0 \text{ 对所有 } \boldsymbol{w} \in W \text{ 成立}\}.$$

$W^{\perp}$被称为$W$的**正交补*(orthogonal complement)***。

**Theorem 4.41.** 设$W \subset \mathbb{R}^n$为一个子空间，那么

*1.* $W^{\perp} \subset \mathbb{R}^n$是一个子空间。
*2.* 若$\boldsymbol{w} \in W \cap W^{\perp}$，那么$\boldsymbol{w} = \boldsymbol{0}$，即$W \cap W^{\perp} = \{\boldsymbol{0}\}$。。
*3.* $(W^{\perp})^{\perp} = W$。

证明. 第三条性质我们暂时跳过。前两条性质非常容易证明，比如第二条：假设$\boldsymbol{w} \in W \cap W^{\perp}$，那么$\boldsymbol{w} \cdot \boldsymbol{w} = 0$，即$\|\boldsymbol{w}\|^2 = 0$。由欧式范数性质可得$\boldsymbol{w} = \boldsymbol{0}$。 □

**Theorem 4.42.** 设$W \subset \mathbb{R}^n$为一个子空间，$S = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_k\}$为$W$的一组基底。那么如果$\boldsymbol{v} \in \mathbb{R}^n$同时正交于$S$里的所有向量：

$$\boldsymbol{v} \cdot \boldsymbol{w}_i = 0$$

对所有$i = 1, \dots, k$成立，则必然有$\boldsymbol{v} \in W^{\perp}$。

证明. 由于$S = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_k\}$为$W$的一组基底，那么任何$\boldsymbol{w} \in W$都可以表示为

$$\boldsymbol{w} = c_1\boldsymbol{w}_1 + \dots + c_k\boldsymbol{w}_k.$$

因此$\boldsymbol{v} \cdot \boldsymbol{w} = c_1\boldsymbol{v} \cdot \boldsymbol{w}_1 + \dots + c_k\boldsymbol{v} \cdot \boldsymbol{w}_k = 0 + \dots + 0 = 0$，即$\boldsymbol{v} \in W^{\perp}$。 □

**考试常见题型概览：(重要性不分先后，且不完全覆盖期中考试的内容)**

- 一个向量空间内的子集是否为一个子空间，即验证加法与标量积封闭性。
  代表题目：期中模拟卷第一题(a)，第六次作业Problem A，Problem E.
- 找出一个给定向量空间或子空间的一组基底，这里包括寻找一个给定矩阵$A$的基本空间的一组基底。
  代表题目：期中模拟卷第五题，第六次作业Problem C， Problem D，第七次作业Problem B，Problem C，Problem D， 2022年期中考试薛老师班级第7题。
- 方阵$A$可逆的若干个判断标准。
  代表题目：期中模拟卷第一题(b)。
- 矩阵乘法的消去律：$AB = AC$在哪种条件下一定可以推出$B = C$?
  代表题目：期中模拟卷第一题(b)。
- 维数公式Theorem 4.39的运用。
  代表题目：期中模拟卷第一题(c)，第七次作业Problem E， Bonus题目。
- 对于给定基底求坐标向量。
  代表题目：期中模拟卷第二题(a)，期中模拟卷第6题(c)，第六次作业Problem C， Problem D，2022年期中考试薛老师班级第1题(5)。
- 矩阵与伴随矩阵之间的关系，不同代数余子式展开的不同取值。
  代表题目：期中模拟卷第二题(b)，期中模拟卷第四题，第四次作业Problem A，Problem D，讲义第65页的例子(2022年期中考试题)，第五次作业Problem A-(1)。
- 矩阵的秩的计算。
  代表题目：期中模拟卷第二题(c)，第八次作业Problem C， Problem D。
- 具体矩阵的乘法与求逆计算。
  代表题目：期中模拟卷第三题，第四次作业Bonus题目，2022年期中考试薛老师班级第6题。
- 不同基底之间转移矩阵的计算。
  代表题目：期中模拟卷第六题(b),(c)，第七次作业Problem A，讲义第118页例子。
- 正交性的定义，正交投影计算公式，与正交性有关的一些等式与不等式(比如勾股定理，，Cauchy–Schwarz不等式)。
  代表题目：期中模拟卷第七题，第五次作业Problem B, Problem C, Problem D.
- 内积与矩阵乘法的关系。
  代表题目：期中模拟卷第八题，第五次作业Problem E。
- 由方程决定的平面与直线，平面的point-norm form。
  代表题目：第八次作业Problem B。

补充题目：令$H = \{x + y + z + 2 = 0\} \subset \mathbb{R}^3$为立体空间里的一个平面。令$W$为由$(-3, -2, 0)$与$H$的法向量作为基底的子空间。求$W$的正交补$W^\perp$的一组基底，并将$W$表示为point-norm form。

- 特殊$n$阶行列式的通项公式求解。

  代表题目：讲义第2.4节，第四次作业Problem B。

### 4.9. 矩阵变换, Matrix Transformations

本节的主要内容来自**英文教材第1.8节**。

首先，假设$A$与$B$为两个集合，我们称$f : A \to B$为从$A$到$B$的一个**函数(function)**，如果$f$是一种对每个$a \in A$都分配**唯一**一个$B$中对应的元素$f(a) \in B$的规则。换而言之，对每个$a \in A$，我们都有一个对应的$f(a) \in B$且这样的$f(a)$只有一个。
如果$f : A \to B$是一个函数，那么我们称$f(a)$**为$a$关于函数$f$的像(the image of $a$ under $f$)**或者**函数$f$在$a$上的取值(the value of $f$ at $a$)**; 集合$A$被称为函数$f$的**定义域(domain)**，集合$B$被称为函数$f$的**到达域(codomain)**，也称为**陪域，余定义域，目标集**；我们称由定义域$A$里的所有的元素$a$关于函数$f$的像$f(a)$所组成的集合

$$\text{RAN}(f) = \{f(a) : a \in A\}$$

为函数$f$的**值域(range)**。

**Definition 4.43.** 对任何$A \in M_{m\times n}$，我们称以下函数$f = T_A$

$$T_A : \mathbb{R}^n \times \mathbb{R}^m, \boldsymbol{x}(\in \mathbb{R}^n) \mapsto T_A(\boldsymbol{x}) = A\boldsymbol{x}(\in \mathbb{R}^m)$$

*为$A$**的矩阵变换(matrix transformation of $A$)**。*

显然，对于$A \in M_{m\times n}$，函数$T_A$的定义域为$\mathbb{R}^n$，到达域为$\mathbb{R}^m$，且$T_A$的实质是用$m \times n$–矩阵$A$左乘$n$维列向量(即$n \times 1$–矩阵)$\boldsymbol{x} \in \mathbb{R}^n$，从而得到它的像$A\boldsymbol{x} \in \mathbb{R}^m$(为$m \times 1$–矩阵)。
利用矩阵乘法的运算法则，我们容易验证$T_A$满足以下性质。

**Theorem 4.44.** *对任何$A \in M_{m\times n}$，$\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$以及实数$k \in \mathbb{R}$，有*

1. $T_A(\mathbf{0}) = \mathbf{0}$;
2. $T_A(k\boldsymbol{u}) = kT_A(\boldsymbol{u})$;
3. $T_A(\boldsymbol{u} + \boldsymbol{v}) = T_A(\boldsymbol{u}) + T_A(\boldsymbol{v})$.

实际上满足以上性质的从$\mathbb{R}^n$到$\mathbb{R}^m$的函数都可以表示为一个矩阵变换。

**Theorem 4.45.** *一个函数$T : \mathbb{R}^n \to \mathbb{R}^m$是一个矩阵变换，即$T = T_A$对某个$A \in M_{m\times n}$成立，当且仅当对任何$\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$以及实数$k \in \mathbb{R}$，有*

1. $T(k\boldsymbol{u}) = kT(\boldsymbol{u})$ ***(homogeneity property)***;
2. $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$ ***(additive property)***.

证明. 显然我们只需证明满足以上两个性质的任何函数都是一个矩阵变换。假设$T: \mathbb{R}^n \to \mathbb{R}^m$满足以上两条性质，那么对$\mathbb{R}^n$里的标准基底$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$以及系数$k_1, \ldots, k_n \in \mathbb{R}$，必然有

$$T(k_1\boldsymbol{e}_1 + \ldots + k_n\boldsymbol{e}_n) = k_1T(\boldsymbol{e}_1) + \ldots + k_nT(\boldsymbol{e}_n).$$

那么，令$A$为以下$m \times n$-矩阵：(即将$T(\boldsymbol{e}_1), \ldots, T(\boldsymbol{e}_n)$作为矩阵的列向量)

$$A = \begin{bmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & \ldots & T(\boldsymbol{e}_n) \end{bmatrix}$$

对于任何$\mathbb{R}^n$里的向量$\boldsymbol{x} = (x_1, \ldots, x_n) = x_1\boldsymbol{e}_1 + \ldots + x_n\boldsymbol{e}_n$，我们有

$$T(\boldsymbol{x}) \underbrace{=}_{T\text{满足的两个性质}} x_1T(\boldsymbol{e}_1) + \ldots + x_nT(\boldsymbol{e}_n) \underbrace{=}_{\text{矩阵乘法的列向量表示}} A\boldsymbol{x}.$$

因此我们得到$T(\boldsymbol{x}) = A\boldsymbol{x}$对任何$\boldsymbol{x} \in \mathbb{R}^n$成立，即$T = T_A$。 □

以下定义在下一章会再次出现，这里我们简单提及：

**Definition 4.46.** 设$V$与$W$为两个向量空间，函数$T: V \to W$是一个线性映射*(linear mapping)*或者**线性变换*(linear transformation)***，如果它对任何$\boldsymbol{u}, \boldsymbol{v} \in V$以及实数$k \in \mathbb{R}$满足以下两个性质：

1. $T(k\boldsymbol{u}) = kT_A(\boldsymbol{u})$ ***(homogeneity property)***;
2. $T(\boldsymbol{u} + \boldsymbol{v}) = T_A(\boldsymbol{u}) + T_A(\boldsymbol{v})$ ***(additive property)***.

显然，我们实际上已经证明，**任何矩阵变换$T_A$都是一个从$\mathbb{R}^n$到$\mathbb{R}^m$的线性变换；反之任何从$\mathbb{R}^n$到$\mathbb{R}^m$的线性变换$T$都能够表示为一个矩阵变换$T = T_A$。**
这里很自然的引申出一个问题，对于同一个线性变换$T: \mathbb{R}^n \to \mathbb{R}^m$，是否可能存在两个不同的矩阵$A \in M_{m\times n}, B \in M_{m\times n}, A \neq B$使得

$$T = T_A = T_B?$$

以下定理告诉我们这种情况不会发生，即对于每个线性变换$T: \mathbb{R}^n \to \mathbb{R}^m$，都有唯一确定的矩阵$A \in M_{m\times n}$使得$T = T_A$。

**Theorem 4.47.** 假设$A \in M_{m\times n}, B \in M_{m\times n}$满足$T_A = T_B$，那么$A = B$。

证明. 设$A$的列向量表示为

$$A = \begin{bmatrix} \boldsymbol{c}_1 & \ldots & \boldsymbol{c}_n \end{bmatrix}$$

$B$的列向量表示为

$$B = \begin{bmatrix} \boldsymbol{w}_1 & \ldots & \boldsymbol{w}_n \end{bmatrix}$$

那么对于$\mathbb{R}^n$的标准基底$\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$，我们对任何$i=1,\ldots,n$都有$T_A(\boldsymbol{e}_i)=A\boldsymbol{e}_i=\boldsymbol{c}_i$，$T_B(\boldsymbol{e}_i)=B\boldsymbol{e}_i=\boldsymbol{w}_i$。因此$T_A=T_B$意味着

$$T_A(\boldsymbol{e}_i)=T_B(\boldsymbol{e}_i)\Rightarrow \boldsymbol{c}_i=\boldsymbol{w}_i$$

对任何$i=1,\ldots,n$都成立，即$A$与$B$的每一列都相等。因此得到$A=B$。 □

**Definition 4.48.** 设$T:\mathbb{R}^n\to\mathbb{R}^m$为一个线性变换，$A\in M_{m\times n}$为唯一满足$T(\boldsymbol{x})=T_A(\boldsymbol{x})=A\boldsymbol{x}$的矩阵，即$A=\begin{bmatrix}T(\boldsymbol{e}_1) & \ldots & T(\boldsymbol{e}_n)\end{bmatrix}$($\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$为$\mathbb{R}^n$的标准基底)。我们称$A$为$T$的**标准矩阵*(standard matrix)***。我们通常将这个标准矩阵记为$[T]$。

以下是一些常见的欧式空间上的矩阵变换/线性变换：

- 令$A=\mathbf{0}_{m\times n}$为零矩阵，那么显然$T_A$把所有$\mathbb{R}^n$里的向量都映射到$\mathbb{R}^m$里的零向量。我们称这个矩阵变换为**零变换(zero transformation)**。
- 如果线性变换$T:\mathbb{R}^n\to\mathbb{R}^n$的定义域与到达域都是$\mathbb{R}^n$，那么我们也称$T$为**$\mathbb{R}^n$上的线性算子(linear operator on $\mathbb{R}^n$)**。此时$T$的标准矩阵$A$显然为$n$阶方阵。比如$A=I_n$为单位矩阵时，$T_A(\boldsymbol{x})=\boldsymbol{x}$对任何$\boldsymbol{x}\in\mathbb{R}^n$都成立。反之，容易验证，**如果一个线性算子$T:\mathbb{R}^n\to\mathbb{R}^n$满足$T(\boldsymbol{x})=\boldsymbol{x}$对任何$\boldsymbol{x}\in\mathbb{R}^n$都成立，那么它的标准矩阵必然为单位矩阵$I_n$**。这是因为如果$A\boldsymbol{x}=\boldsymbol{x}$对任何$\boldsymbol{x}\in\mathbb{R}^n$成立，则必有$A\boldsymbol{e}_i=\boldsymbol{e}_i$对任何$i=1,\ldots,n$成立，那么这意味着$A$的第$i$列是第$i$个标准单位向量$\boldsymbol{e}_i$。显然满足这个条件的矩阵只能是单位矩阵$I_n$。
- 假设$\boldsymbol{v}\in\mathbb{R}^n$，$\boldsymbol{v}\neq\mathbf{0}$为非零向量。我们定义$T:\mathbb{R}^n\to\mathbb{R}^n$为所谓的**正交投影算子**，即

  $$T(\boldsymbol{w})=\text{proj}_{\boldsymbol{v}}(\boldsymbol{w})$$

  给出向量$\boldsymbol{w}\in\mathbb{R}^n$在$\boldsymbol{v}$上的正交投影。容易验证这样的$T$为$\mathbb{R}^n$上的线性算子：对任何$\boldsymbol{w}_1,\boldsymbol{w}_2\in\mathbb{R}^n$，任何$k\in\mathbb{R}^n$，通过正交投影计算公式可得

  $$\begin{aligned}T(\boldsymbol{w}_1+\boldsymbol{w}_2)=\text{proj}_{\boldsymbol{v}}(\boldsymbol{w}_1+\boldsymbol{w}_2)=\frac{(\boldsymbol{w}_1+\boldsymbol{w}_2)\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}&=\frac{\boldsymbol{w}_1\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}+\frac{\boldsymbol{w}_2\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}\\ &\text{proj}_{\boldsymbol{v}}(\boldsymbol{w}_1)+\text{proj}_{\boldsymbol{v}}(\boldsymbol{w}_2)\\ &=T(\boldsymbol{w}_1)+T(\boldsymbol{w}_2);\end{aligned}$$

  $$\begin{aligned}T(k\boldsymbol{w}_1)=\text{proj}_{\boldsymbol{v}}(k\boldsymbol{w}_1)=\frac{(k\boldsymbol{w}_1)\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}&=\frac{k(\boldsymbol{w}_1\cdot\boldsymbol{v})}{\|\boldsymbol{v}\|^2}\boldsymbol{v}\\ &=k\text{proj}_{\boldsymbol{v}}(\boldsymbol{w}_1)\\ &=kT(\boldsymbol{w}_1).\end{aligned}$$

  特别地，如果选取$\boldsymbol{v}=\boldsymbol{e}_i$为$\mathbb{R}^n$的第$i$个标准单位向量，那么容易验证此时对于任何$\boldsymbol{w}=(w_1,\ldots,w_n)\in\mathbb{R}^n$，

  $$T(\boldsymbol{w})=\text{proj}_{\boldsymbol{e}_i}(\boldsymbol{w})=w_i\boldsymbol{e}_i=(0,\ldots,0,w_i,0,\ldots,0)$$

给出的是$\boldsymbol{w}$关于标准基底的第$i$个坐标$w_i$与$\boldsymbol{e}_i$的标量积。关于正交投影算子的标准矩阵的计算留作作业。

- 考虑$T:\mathbb{R}^2\to\mathbb{R}^2$为将$\mathbb{R}^2$里的所有向量**逆时针(counterclockwise)**旋转$\theta(\in[0,2\pi])$度的操作。显然，通过画图可知，$T$将$x$轴上的标准单位向量$\boldsymbol{e}_1=(1,0)$映射到$T(\boldsymbol{e}_1)=(\cos\theta,\sin\theta)$，将$y$轴上的标准单位向量$\boldsymbol{e}_2=(0,1)$映射到$T(\boldsymbol{e}_2)=(-\sin\theta,\cos\theta)$。那么利用平面上向量的极坐标表示，我们可以证明这个**"旋转算子"(rotation operator)**是一个$\mathbb{R}^2$上的线性算子，且它的标准矩阵为

$$A=\begin{bmatrix}T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2)\end{bmatrix}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}.$$

  更多关于$\mathbb{R}^3$上的旋转算子的分析请参阅英文教材261–263页的内容。

- 对于给定的非负数$k\geq 0$，考虑函数$T:\mathbb{R}^n\to\mathbb{R}^n$，$T(\boldsymbol{x})=k\boldsymbol{x}$。显然$T$的作用是将向量$\boldsymbol{x}$在保持方向不变的情况下改变它的长度。当$k>1$时我们称为向量的**拉伸(dilation)**，当$k\leq 1$时称为向量的**收缩(contraction)**。容易验证这样的$T$为线性变换，且它的标准矩阵为对角矩阵

$$\begin{bmatrix}k & & & \\ & k & & \\ & & \ddots & \\ & & & k\end{bmatrix}.$$

更多$\mathbb{R}^2$与$\mathbb{R}^3$空间上的线性变换与其对应的几何意义请仔细阅读英文教材4.9节。

### 4.10. 矩阵变换的性质, Properties of matrix transformations

假设$T:\mathbb{R}^n\to\mathbb{R}^k$为从$\mathbb{R}^n$到$\mathbb{R}^k$的线性变换，$S:\mathbb{R}^k\to\mathbb{R}^m$为从$\mathbb{R}^k$到$\mathbb{R}^m$的线性变换，那么我们可以定义$S$与$T$的**复合(composition)**并记为$S\circ T$，这是一个从$\mathbb{R}^n$到$\mathbb{R}^m$的函数，满足

$$(S\circ T)(\boldsymbol{x})=S(T(\boldsymbol{x})),\quad \boldsymbol{x}\in\mathbb{R}^n.$$

用语言来描述的话，$S\circ T$的作用是先把$\boldsymbol{x}\in\mathbb{R}^n$通过$T$映射到$T(\boldsymbol{x})\in\mathbb{R}^k$，再把$T(\boldsymbol{x})\in\mathbb{R}^k$通过$S$映射到$S(T(\boldsymbol{x}))\in\mathbb{R}^m$。显然，对于任何$\boldsymbol{x},\boldsymbol{w}\in\mathbb{R}^n$，任何$k\in\mathbb{R}$，都有

$$(S\circ T)(\boldsymbol{x}+\boldsymbol{w})=S(T(\boldsymbol{x})+T(\boldsymbol{w}))=S(T(\boldsymbol{x}))+S(T(\boldsymbol{w}))=(S\circ T)(\boldsymbol{x})+(S\circ T)(\boldsymbol{w});$$
$$(S\circ T)(k\boldsymbol{x})=S(T(k\boldsymbol{x}))=S(kT(\boldsymbol{x}))=k(S(T(\boldsymbol{x})))=k(S\circ T)(\boldsymbol{x})$$

也即$S\circ T$**作为两个线性变换的复合也是一个线性变换**。实际上利用数学归纳法可以证明，对于任何向量空间$V_1,V_2,\ldots,V_{k+1}$，任何线性变换$T_1:\circ V_1\to V_2$，$T_2:V_2\to V_3$，$\ldots$，$T_k:V_k\to V_{k+1}$，它们的复合

$$T_k\circ T_{k-1}\circ\ldots\circ T_2\circ T_1:V_1\to V_{k+1}$$

是从$V_1$到$V_{k+1}$的线性变换。
由Theorem 4.45我们知道对于线性变换$T:\mathbb{R}^n\circ\mathbb{R}^k$，我们可以将其表示为$T(\boldsymbol{x})=A\boldsymbol{x}$，$A=[T]\in M_{k\times n}$为$T$的标准矩阵。同理，对于线性变换$S:\mathbb{R}^k\circ\mathbb{R}^m$，我们可以将其表示为$S(\boldsymbol{v})=B\boldsymbol{v}$，$B=[S]\in M_{m\times k}$为$S$的标准矩阵。容易验证线性变换$S\circ T:\mathbb{R}^n\to\mathbb{R}^m$的标准矩阵为

$$[S\circ T]=[S][T]=BA,$$

即等于$S$的标准矩阵$[S]$左乘$T$的标准矩阵$[T]$。显然该等式可以推广到任意多个线性变换的复合，即对于$V_1=\mathbb{R}^{n_1},V_2=\mathbb{R}^{n_2},\ldots,V_{k+1}=\mathbb{R}^{n_{k+1}}$，任何线性变换$T_1:\circ V_1\to V_2$，$T_2:V_2\to V_3$，$\ldots$，$T_k:V_k\to V_{k+1}$，都有

$$[T_k\circ\ldots\circ T_2\circ T_1]=[T_k]\ldots[T_2][T_1]$$

即复合后的线性变换$T_k\circ\ldots\circ T_2\circ T_1$的标准矩阵等于它们各自的标准矩阵按照复合顺序的矩阵乘积。
由于矩阵乘法不服从交换律，因此线性变换的复合通常也不满足可交换的性质，即一般情况下，对于$S:\mathbb{R}^n\to\mathbb{R}^n$与$T:\mathbb{R}^n\to\mathbb{R}^n$，$S\circ T\neq T\circ S$。但也有两个线性变换的复合可交换的特殊情况，比如：
令$T_1:\mathbb{R}^2\to\mathbb{R}^2$为将$\mathbb{R}^2$里的向量逆时针旋转$\theta_1$度的线性变换，令$T_2:\mathbb{R}^2\to\mathbb{R}^2$为将$\mathbb{R}^2$里的向量逆时针旋转$\theta_2$度的线性变换。那么从直观上考虑，$T_2\circ T_1$实现的是先将向量逆时针旋转$\theta_1$度再旋转$\theta_2$度；而$T_1\circ T_2$实现的是先将向量逆时针旋转$\theta_2$度再旋转$\theta_1$度。显然这两种操作的最终结果都是把向量逆时针旋转了$\theta_1+\theta_2$度，从而有$T_2\circ T_1=T_1\circ T_2$。这一观察可以被矩阵乘法

$$[T_2][T_1]=\begin{bmatrix}\cos\theta_2 & -\sin\theta_2\\ \sin\theta_2 & \cos\theta_2\end{bmatrix}\begin{bmatrix}\cos\theta_1 & -\sin\theta_1\\ \sin\theta_1 & \cos\theta_1\end{bmatrix}=\begin{bmatrix}\cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2)\\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2)\end{bmatrix};$$

$$[T_1][T_2]=\begin{bmatrix}\cos\theta_1 & -\sin\theta_1\\ \sin\theta_1 & \cos\theta_1\end{bmatrix}\begin{bmatrix}\cos\theta_2 & -\sin\theta_2\\ \sin\theta_2 & \cos\theta_2\end{bmatrix}=\begin{bmatrix}\cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2)\\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2)\end{bmatrix}$$

所验证。
显然$S\circ T=T\circ S$等价于$[S][T]=[T][S]$，即**两个欧式空间上线性算子的复合可交换等价于它们标准矩阵的乘积可交换**。
更多关于线性变换复合的几何意义的讨论请阅读英文教材272页。

由以上讨论我们显然可以看出，欧式空间上的线性变换所满足的性质实际上被它(们)的标准矩阵所满足的对应性质所刻画。接下来我们继续验证这一事实。

**Definition 4.49.** 对于向量空间$V,W$，线性变换$T:V\to W$，我们称

1. $T$是一个**一一映射*(one-to-one)***或者**单射*(injective)***，如果对于任何$\boldsymbol{v}_1 \neq \boldsymbol{v}_2$，都有$T(\boldsymbol{v}_1) \neq T(\boldsymbol{v}_2)$。
2. $T$是一个**满射*(surjective)***，如果对于任何$\boldsymbol{w} \in W$都存在一个$\boldsymbol{v} \in V$使得$T(\boldsymbol{v}) = \boldsymbol{w}$，也即$T$的值域*RAN*$(T) = W$。此时我们也称**"$T$ *is onto* W"**.
3. $T$是一个**双射*(bijective)***，如果$T$同时是单射和满射。

**Definition 4.50.** 对于向量空间$V, W$，线性变换$T : V \to W$，我们称集合

$$\{\boldsymbol{v} \in V : T(\boldsymbol{v}) = \boldsymbol{0} \in W\}$$

为$T$的**核*(kernel)***，并记为*Ker*$(T)$。

**Theorem 4.51.** 令$V = \mathbb{R}^n$，$W = \mathbb{R}^m$，$T : \mathbb{R}^n \to \mathbb{R}^m$为线性变换，$[T] \in M_{m\times n}$为$T$的标准矩阵。那么我们有

1. *Ker*$(T) =$ *Null*$([T])$;
2. *RAN*$(T) =$ *Col*$([T])$;
3. $T$是单射当且仅当*Ker*$(T) =$ *Null*$([T]) = \{\boldsymbol{0}\}$，当且仅当*nullity*$([T]) = 0$;
4. $T$是满射当且仅当*Col*$([T]) = \mathbb{R}^m$，当且仅当*rank*$([T]) = m$;
5. 如果$m = n$，那么$T : \mathbb{R}^n \to \mathbb{R}^n$为双射当且仅当$[T]$可逆，当且仅当$T$是单射，当且仅当$T$是满射。

证明. 我们将$[T]$记为$A$。那么由于$A$是$T$的标准矩阵，我们有$T(\boldsymbol{x}) = A\boldsymbol{x}$对任何$\boldsymbol{x} \in \mathbb{R}^n$成立。显然由核的定义，

$$\text{Ker}(T) = \{\boldsymbol{x} \in \mathbb{R}^n : T(\boldsymbol{x}) = \boldsymbol{0}\} = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{0}\} = \text{Null}(A).$$

类似可证，对于$A$的列向量表示$A = \begin{bmatrix} \boldsymbol{c}_1 & \dots & \boldsymbol{c}_n \end{bmatrix}$，

$$\begin{aligned} \text{RAN}(T) = \{T(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^n\} &= \{A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\} \\ &= \{x_1\boldsymbol{c}_1 + \ldots + x_n\boldsymbol{c}_n : x_1, \ldots, x_n \in \mathbb{R}\} \\ &= \text{span}\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\} \\ &= \text{Col}(A). \end{aligned}$$

首先假设$T : \mathbb{R}^n \to \mathbb{R}^m$是一个单射，那么对于任何$\boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^n$，必然有$T(\boldsymbol{x}) = A\boldsymbol{x} \neq A\boldsymbol{0} = T(\boldsymbol{0}) = \boldsymbol{0} \in \mathbb{R}^m$，这意味着Ker$(T) =$ Null$(A) = \{\boldsymbol{0}\}$。反之，假设Ker$(T) =$ Null$(A) = \{\boldsymbol{0}\}$，但$T$不是单射，即存在$\boldsymbol{x} \neq \boldsymbol{y}$使得$T(\boldsymbol{x}) = T(\boldsymbol{y})$。由于$T$的线性特点，此时必然有$T(\boldsymbol{x} - \boldsymbol{y}) = T(\boldsymbol{x}) - T(\boldsymbol{y}) = \boldsymbol{0}$，这意味着$\boldsymbol{x} - \boldsymbol{y} \in$ Ker$(T)$。另一方面，因为$\boldsymbol{x} \neq \boldsymbol{y}$，必然有$\boldsymbol{x} - \boldsymbol{y} \neq \boldsymbol{0}$，即Ker$(T)$此时包含了一个非零向量，与假设Ker$(T) =$ Null$(A) = \{\boldsymbol{0}\}$矛盾。因此$T$只能是单射。我们因此证明

了$T$是单射当且仅当$\mathrm{Ker}(T) = \mathrm{Null}([T]) = \{\mathbf{0}\}$，当且仅当$\mathrm{nullity}([T]) = 0$。
由于已经证明了$\mathrm{RAN}(T) = \mathrm{Col}([T])$，我们可以得到$T$为满射当且仅当$\mathrm{RAN}(T) = \mathbb{R}^m$当且当今$\mathrm{Col}(A) = \mathbb{R}^m$当且仅当$\mathrm{rank}(A) = \dim(\mathrm{Col}(A)) = m$。
假设$m = n$，此时$A$为$n$阶方阵。由定义可知$T$是双射意味着它既是单射又是满射。由以上讨论可知，$T$是满射当且仅当$\mathrm{nullity}(A) = 0$同时$\mathrm{rank}(A) = n$。然而，利用rank–nullity theorem 4.30，即$\mathrm{nullity}(A)+\mathrm{rank}(A) = n$，我们可以看出$\mathrm{nullity}(A) = 0$成立当且仅当$\mathrm{rank}(A) = n$，也即$T$是满射只需$\mathrm{nullity}(A) = 0$(即$T$为单射)**或者**$\mathrm{rank}(A) = n$(即$T$为满射)两个条件之一成立即可。此外注意对于$n$阶方阵$A$，$\mathrm{nullity}(A) = 0$和$\mathrm{rank}(A) = n$都等价于$A$可逆。 □

假设$T : \mathbb{R}^n \to \mathbb{R}^n$为一个双射的线性变换，那么对于任何$\boldsymbol{y} \in \mathbb{R}^n$，都存在唯一的一个**原像(preimage)**$\boldsymbol{x} \in \mathbb{R}^n$使得$\boldsymbol{y} = T(\boldsymbol{x})$。因此我们可以定义它的逆映射(inverse)$T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$满足$T^{-1}(\boldsymbol{y}) = \boldsymbol{x}$。容易验证

**Theorem 4.52.** 对于任何双射的线性变换$T : \mathbb{R}^n \to \mathbb{R}^n$，它的逆映射$T^{-1}$也为一个线性变换，且如果$T$的标准矩阵为$[T]$，那么$T^{-1}$的标准矩阵为$[T]^{-1}$。

证明. 对于$\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^n$，假设它们的原像分别为$\boldsymbol{x}_1, \boldsymbol{x}_2$，即$T(\boldsymbol{x}_1) = \boldsymbol{y}_1, T(\boldsymbol{x}_2) = \boldsymbol{y}_2$，那么我们有$T^{-1}(\boldsymbol{y}_1) = \boldsymbol{x}_1$，$T^{-1}(\boldsymbol{y}_2) = \boldsymbol{x}_2$。由于$T$为线性变换，由$T(\boldsymbol{x}_1 + \boldsymbol{x}_2) = T(\boldsymbol{x}_1) + T(\boldsymbol{x}_2) = \boldsymbol{y}_1 + \boldsymbol{y}_2$可得$\boldsymbol{y}_1 + \boldsymbol{y}_2$的原像为$\boldsymbol{x}_1 + \boldsymbol{x}_2$且由于$T$为单射，我们知道$\boldsymbol{x}_1 + \boldsymbol{x}_2$为$\boldsymbol{y}_1 + \boldsymbol{y}_2$的唯一原像，因此$T^{-1}(\boldsymbol{y}_1 + \boldsymbol{y}_2) = x_1 + x_2 = T^{-1}(\boldsymbol{y}_1) + T^{-1}(\boldsymbol{y}_2)$。
类似的，对于任何$\boldsymbol{y} \in \mathbb{R}^n$，$k \in \mathbb{R}$，若$T^{-1}(\boldsymbol{y}) = \boldsymbol{x}$，即$T(\boldsymbol{x}) = \boldsymbol{y}$，那么由$T(k\boldsymbol{x}) = kT(\boldsymbol{x}) = k\boldsymbol{y}$可知$k\boldsymbol{y}$关于$T$的原像为$k\boldsymbol{x}$，即$T^{-1}(k\boldsymbol{y}) = k\boldsymbol{x} = kT^{-1}(\boldsymbol{y})$。综上，可得$T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$也为一个线性变换。
现在我们验证$T^{-1}$的标准矩阵为$[T]^{-1}$。首先注意由于已经证明$T^{-1}$是一个线性变换，由Theorem 4.45可知存在一个$n$阶方阵$[T^{-1}]$使得$T^{-1}(\boldsymbol{y}) = [T^{-1}]\boldsymbol{y}$对任何$\boldsymbol{y} \in \mathbb{R}^n$成立。那么，对于任何给定的$\boldsymbol{y} \in \mathbb{R}^n$，假设它的原像为$T^{-1}(\boldsymbol{y}) = \boldsymbol{x}$(即$T(\boldsymbol{x}) = \boldsymbol{y}$)，由于$T(T^{-1}(\boldsymbol{y})) = T(\boldsymbol{x}) = \boldsymbol{y}$，我们实际上得到

$$T(T^{-1}(\boldsymbol{y})) = [T]([T^{-1}]\boldsymbol{y}) = \boldsymbol{y},$$

即$([T][T^{-1}])\boldsymbol{y} = \boldsymbol{y}$对任何$\boldsymbol{y} \in \mathbb{R}^n$成立。显然由之前的证明我们知道这样的矩阵$[T][T^{-1}]$只能是单位矩阵$I_n$，即$[T][T^{-1}] = I_n$。同理易得对任何$\boldsymbol{x} \in \mathbb{R}^n$都有$T^{-1}(T(\boldsymbol{x})) = \boldsymbol{x} \Rightarrow ([T][T^{-1}])\boldsymbol{x} = \boldsymbol{x}$，即$[T][T^{-1}] = I_n$。综上可得$[T^{-1}] = [T]^{-1}$。 □

注意以上定理也可以表示为，对任何可逆方阵$A \in M_{n\times n}$，$A$所对应的矩阵变换$T_A(\boldsymbol{x}) = A\boldsymbol{x}$满足

$$(T_A)^{-1} = T_{A^{-1}}.$$

# 5 第五章：线性变换

注意，本章对应**英文教材的第八章General Linear Transformations**。

## 5.1. 一些重要的线性变换与线性变换的一些基本性质

首先我们回顾一下线性变换的定义。

**Definition 5.1.** 令$V$与$W$为向量空间。一个从$V$到$W$的函数$T : V \to W$是一个**线性变换*(linear transformation)***，如果它满足以下两个条件

1. $T(\boldsymbol{u}+\boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$对任何$\boldsymbol{u}, \boldsymbol{v} \in V$成立， *(additive property)*
2. $T(k\boldsymbol{u}) = kT(\boldsymbol{u})$对任何$\boldsymbol{u} \in V$，任何$k \in \mathbb{R}$成立。 *(homogeneity property)*

如果$T : V \to V$的定义域与到达域都是$V$，那么也称线性变换$T$为$V$**上的线性算子*(linear operator on $V$)***。

注意，由以上线性变换满足的性质可知，任何线性变换$T : V \to W$都满足

$$T(\mathbf{0}_V) = \mathbf{0}_W,$$

这里我们用$\mathbf{0}_V, \mathbf{0}_W$分别代指$V$与$W$里的零向量。当然也可以简写为$T(\mathbf{0}) = \mathbf{0}$。因此，如果一个函数$f : V \to W$**不满足**$f(\mathbf{0}_V) = \mathbf{0}_W$，那么它一定**不是线性变换**。这一标准在判断一个函数是否是线性变换时通常很有用。
在本节余下的部分，我们介绍一些常见的线性变换。**注意其中一些例子出现在过往考试中**。

**例子1**：矩阵变换(Matrix Transformations)
当$V = \mathbb{R}^n$，$W = \mathbb{R}^m$时，我们在Theorem 4.45中证明了任何从$\mathbb{R}^n$到$\mathbb{R}^m$的线性变换$T$都可以表示为

$$T(\boldsymbol{x}) = T_A(\boldsymbol{x}) = A\boldsymbol{x},$$

这里$A = [T] \in M_{m\times n}$为$T$的标准矩阵(standard matrix)。因此任何欧式空间上的线性变换都是矩阵变换。

**例子2**：恒等算子(The identity operator)
令$V$为任意向量空间，令$I : V \to V$为满足$I(\boldsymbol{v}) = \boldsymbol{v}$的函数。那么显然$I$是一个$V$上的线性算子，它把$V$里的任意向量$\boldsymbol{v}$映射到$\boldsymbol{v}$本身。这样的线性算子被称为恒等算

子(the identity operator)。

**例子3**：矩阵空间上的线性变换(Transformations on Matrix Spaces)

- 令$V = M_{m\times n}$，令$W = M_{n\times m}$，那么

$$T(A) = A^{\top}$$

即，矩阵的转置，是一个线性变换。

- 令$V = M_{n\times n}$，令$W = \mathbb{R}$，那么

$$T(A) = \det(A)$$

即，方阵的行列式，在$n > 1$的情况下不是一个线性变换。因为此时能够找到一些实数$k \in \mathbb{R}$，有$\det(kA) = k^n \det(A) \neq k \det(A)$。

- 令$V = M_{n\times n}$，令$W = \mathbb{R}$，那么

$$T(A) = \text{tr}(A)$$

即，方阵的迹，是一个线性变换。

**例子4**：多项式空间上的两个线性变换

令$V = P_n = \{a_0 + a_1 x + \ldots + a_n x^n : a_0, a_1, \ldots, a_n \in \mathbb{R}\}$为次数不大于$n$的所有多项式(polynomials)的集合，令$W = P_{n+1} = \{a_0 + a_1 x + \ldots + a_{n+1} x^{n+1} : a_0, a_1, \ldots, a_{n+1} \in \mathbb{R}\}$为次数不大于$n+1$的所有多项式(polynomials)的集合。那么函数

$$T : P_n \to P_{n+1}, T(p(x)) = xp(x) = x(a_0 + a_1 x + \ldots + a_n x^n)$$

显然是一个从$P_n$到$P_{n+1}$的线性变换。

令$n > 1$，$V = P_n$，$W = P_{n-1}$，那么函数

$$T : P_n \to P_{n+1}, T(p(x)) = p'(x) = a_1 + 2a_2 x + \ldots + n a_n x^{n-1}$$

即，对多项式的求导运算，显然是一个从$P_n$到$P_{n+1}$的线性变换。

**例子5**：赋值变换(The Evalutation Transformation)

令$V = F(-\infty, \infty)$为所有$\mathbb{R}$上的实数取值的函数的集合。对于**给定**的$n$个实数$x_1 \in \mathbb{R}, x_2 \in \mathbb{R}, \ldots, x_n \in \mathbb{R}$，我们称以下函数

$$T : F(-\infty, \infty) \to \mathbb{R}^n, T(f) = (f(x_1), \ldots, f(x_n))$$

为函数空间上的赋值变换。比如，取$x_1 = -1, x_2 = 2, x_3 = 4$，取$f(x) = x^2 - 1 \in F(-\infty, \infty)$，那么

$$T(x^2 - 1) = ((-1)^2 - 1, (2)^2 - 1, 4^2 - 1) = (0, 3, 15).$$

容易验证这样的$T$是一个线性变换。

**例子6**：求导与积分变换(The Derivation and Integration Transformations)

- 令$V = C^1(-\infty, \infty)$为所有连续可微(continuously differentiable)函数的集合，即如果一个函数$f : \mathbb{R} \to \mathbb{R}$可微且它的导数(derivative)为一个连续函数，则有$f \in V = C^1(-\infty, \infty)$；令$W = C(-\infty, \infty)$为所有连续函数的集合。令$D : V \to W$为求导运算：
$$D(f(x)) = f'(x).$$
比如对于$f(x) = \sin x$，$D(\sin x) = \sin'(x) = \cos x$。容易验证这样的求导运算$D$为线性变换。
- 令$V = C(-\infty, \infty)$为所有连续函数的集合，令$W = C^1(-\infty, \infty)$为所有连续可微(continuously differentiable)函数的集合。令$J : V \to W$为不定积分运算：
$$J(f(t))(x) = \int_0^x f(t)dt.$$
比如对于$f(t) = t^2$，那么$J(f(t)) = \int_0^x f(t)dt = \int_0^x t^2 dt = \frac{t^3}{3}|_0^x = \frac{x^3}{3}$。容易验证这样的积分运算$J$为线性变换。

接下来我们介绍一些关于线性变换的基本性质。首先回顾一下线性变换的核(kernel)与值域(range)的定义。

**Definition 5.2.** 令$V, W$为向量空间，$T : V \to W$为一个线性变换。那么

- 我们称集合$\{\boldsymbol{v} \in V : T(\boldsymbol{v}) = \mathbf{0}_W\}$为$T$的***核(kernel)***或者核空间，并记为*Ker*$(T)$。
- 我们称集合$\{\boldsymbol{w} \in W :$ 存在 $\boldsymbol{v} \in V, T(\boldsymbol{v}) = \boldsymbol{w}\} = \{T(\boldsymbol{v}) \in W : \boldsymbol{v} \in V\}$为$T$的***值域(range)***，并记为*RAN*$(T)$或$R(T)$。

**Theorem 5.3.** 对任何线性变换$T : V \to W$，都有

1. *Ker*$(T) \subset V$是$V$里的一个子空间。
2. *RAN*$(T) \subset W$是$W$里的一个子空间。

*证明.* 1. 我们证明Ker$(T)$满足加法与标量积的封闭性。但这两点都比较明显：对于$\boldsymbol{v}, \boldsymbol{u} \in$ Ker$(T)$，我们有

$$T(\boldsymbol{v}+\boldsymbol{u}) = T(\boldsymbol{v}) + T(\boldsymbol{u}) = \mathbf{0}+\mathbf{0} = \mathbf{0} \Rightarrow \boldsymbol{v}+\boldsymbol{u} \in \text{Ker}(T);$$

$$T(k\boldsymbol{v}) = kT(\boldsymbol{v}) = k\mathbf{0} = \mathbf{0} \Rightarrow k\boldsymbol{v} \in \text{Ker}(T).$$

因此Ker$(T)$为$V$里的一个子空间。

2. 我们证明RAN$(T)$满足加法与标量积的封闭性。如果$\boldsymbol{w}_1, \boldsymbol{w}_2 \in$ RAN$(T)$，那么存在$\boldsymbol{v}_1 \in V, \boldsymbol{v}_2 \in V$使得$T(\boldsymbol{v}_1) = \boldsymbol{w}_1, T(\boldsymbol{v}_2) = \boldsymbol{w}_2$。这意味着$T(\boldsymbol{v}_1+\boldsymbol{v}_2) = T(\boldsymbol{v}_1)+T(\boldsymbol{v}_2) = \boldsymbol{w}_1+\boldsymbol{w}_2$，即$\boldsymbol{w}_1+\boldsymbol{w}_2 \in$ RAN$(T)$。类似可得，对于$\boldsymbol{w} = T(\boldsymbol{v}) \in$ RAN$(T)$，$k \in \mathbb{R}$，那么由于$k\boldsymbol{w} = T(k\boldsymbol{v})$且$k\boldsymbol{v} \in V$，可得$k\boldsymbol{w} \in$ RAN$(T)$。综上，RAN$(T)$为$W$里的一个子空间。

□

**Definition 5.4.** *令$T: V \to W$为一个线性变换。那么我们称$dim(Ker(T)) = nullity(T)$为$T$的零化度，称$dim(RAN(T)) = rank(T)$为$T$的秩。*

显然，如果$V = \mathbb{R}^n$，$W = \mathbb{R}^m$，那么对于矩阵变换$T = T_A : \mathbb{R}^n \to \mathbb{R}^m$，$A = [T] \in M_{m\times n}$为$T$的标准矩阵，我们4.10节Theorem 4.51知道

$$\text{Ker}(T) = \text{Null}(A), \text{RAN}(T) = \text{Col}(A),$$

此时有nullity$(T)$ = dim(Ker$(T)$) = dim(Null$(A)$) = nullity$(A)$，rank$(T)$ = dim(RAN$(T)$) = dim(Col$(A)$) = rank$(A)$。即**此时$T$的零化度/秩与它对应的标准矩阵的零化度/秩相同**。
在本节的最后我们证明秩-零化度定理。

**Theorem 5.5.** *令$V$为一个有限维向量空间，$W$为一个向量空间，$T: V \to W$为一个线性变换。那么我们有*

$$rank(T) + nullity(T) = dim(V).$$

*证明.* 假设$V$为$n$维向量空间，即dim$(V) = n$；假设nullity$(T) = r$，显然我们有$0 \le r \le n$。下面我们对$r$的取值进行分类讨论。

- 首先考虑$1 \le r < n$。此时我们知道nullity$(T)$ = dim(Ker$(T)$) $= r$，即Ker$(T)$的基底包含$r$个向量。假设$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为Ker$(T) \subset V$的一组基底。那么利用Theorem 4.20，通过向$B$里补充$n-r$个向量$\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n$，我们可以得到$V$的一组基底：$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$。如果我们可以证明$S = \{T(\boldsymbol{v}_{r+1}), \ldots, T(\boldsymbol{v}_n)\} \subset$

RAN$(T)$构成$T$的值域RAN$(T) \subset W$的一组基底，那么我们可以得到dim(RAN$(T)) =$ rank$(T) = n - r$，从而证明了等式

$$\text{rank}(T) + \text{nullity}(T) = n - r + r = n = \dim(V).$$

那么我们现在就来证明$S = \{T(\boldsymbol{v}_{r+1}), \ldots, T(\boldsymbol{v}_n)\} \subset$ RAN$(T)$构成$T$的值域RAN$(T) \subset W$的一组基底。首先可以看出span$(S) =$ RAN$(T)$：对于任何$\boldsymbol{w} \in$ RAN$(T)$，
由值域的定义可知存在一个$\boldsymbol{v} \in V$使得$T(\boldsymbol{v}) = \boldsymbol{w}$。因为$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$是$V$的一组基底，$\boldsymbol{v} = c_1\boldsymbol{v}_1 + \ldots + c_r\boldsymbol{v}_r + c_{r+1}\boldsymbol{v}_{r+1} + \ldots + c_n\boldsymbol{v}_n$；又因为$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$是Ker$(T)$里的向量，有$T(\boldsymbol{v}_1) = \ldots = T(\boldsymbol{v}_r) = \mathbf{0}$。因此，我们实际上得到

$$\begin{aligned}
\boldsymbol{w} = T(\boldsymbol{v}) &= T(c_1\boldsymbol{v}_1 + \ldots + c_r\boldsymbol{v}_r + c_{r+1}\boldsymbol{v}_{r+1} + \ldots + c_n\boldsymbol{v}_n) \\
&= \underbrace{c_1T(\boldsymbol{v}_1) + \ldots + c_rT(\boldsymbol{v}_r)}_{=\mathbf{0}} + c_{r+1}T(\boldsymbol{v}_{r+1}) + \ldots + c_nT(\boldsymbol{v}_n) \\
&= c_{r+1}T(\boldsymbol{v}_{r+1}) + \ldots + c_nT(\boldsymbol{v}_n)
\end{aligned}$$

从而有$\boldsymbol{w} \in$ span$(S)$，即RAN$(T) \subset$ span$(S)$。由于span$(S) \subset$ RAN$(T)$显然成立，我们得到span$(S) =$ RAN$(T)$。
接下来我们证明$S = \{T(\boldsymbol{v}_{r+1}), \ldots, T(\boldsymbol{v}_n)\} \subset W$为线性无关集合。考虑等式

$$k_{r+1}T(\boldsymbol{v}_{r+1}) + \ldots + k_nT(\boldsymbol{v}_n) = \mathbf{0}.$$

显然如果存在实数$k_{r+1}, \ldots, k_n$使得以上等式成立，那么必然有

$$k_{r+1}T(\boldsymbol{v}_{r+1}) + \ldots + k_nT(\boldsymbol{v}_n) = T(k_{r+1}\boldsymbol{v}_{r+1} + \ldots + k_n\boldsymbol{v}_n) = \mathbf{0}$$

从而有$k_{r+1}\boldsymbol{v}_{r+1} + \ldots + k_n\boldsymbol{v}_n \in$ Ker$(T)$。又因为已假设$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为Ker$(T)$的基底，我们可以将$k_{r+1}\boldsymbol{v}_{r+1} + \ldots + k_n\boldsymbol{v}_n \in$ Ker$(T)$用$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$的线性组合表示：

$$k_{r+1}\boldsymbol{v}_{r+1} + \ldots + k_n\boldsymbol{v}_n = k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r,$$

这意味着$k_{r+1}\boldsymbol{v}_{r+1} + \ldots + k_n\boldsymbol{v}_n - k_1\boldsymbol{v}_1 - \ldots - k_r\boldsymbol{v}_r = \mathbf{0}$成立。因为已经假设$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$是$V$的一组基底，它们之间的线性无关性质导致$k_1, \ldots, k_r, k_{r+1}, \ldots, k_n$只能取0。因此等式

$$k_{r+1}T(\boldsymbol{v}_{r+1}) + \ldots + k_nT(\boldsymbol{v}_n) = \mathbf{0}.$$

只能在$k_{r+1} = \ldots = k_n = 0$时成立。
综上，我们得到$S = \{T(\boldsymbol{v}_{r+1}), \ldots, T(\boldsymbol{v}_n)\}$是RAN$(T)$的一组基底。

- 如果$r = 0$，那么任取$V$的一组基底$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$，按照以上的方法可以证明$S = \{T(\boldsymbol{v}_1), \ldots, T(\boldsymbol{v}_n)\}$此时构成RAN$(T)$的一组基底。因此rank$(T)$+nullity$(T) = n + 0 = n = \dim(V)$依然成立。

- 如果$r = n$，即nullity$(T) = \dim(\mathrm{Ker}(T)) = n$。此时可知Ker$(T) = V$(因为Ker$(T) \subset V$作为$V$的一个子空间与$V$具有相同的维数)，即$T(\boldsymbol{v}) = \mathbf{0}$对所有$\boldsymbol{v} \in V$都成立。此时RAN$(T) = \{\mathbf{0}\}$，从而有rank$(T) = \dim(\mathrm{RAN}(T)) = 0$。显然此时rank$(T) +$ nullity$(T) = 0 + n = n = \dim(V)$依然成立。

□

## 5.2. 线性变换的复合，逆，与同构, Compositions, Inverse Transformations and Isomorphism

本节的内容对应英文教材8.2-8.3节的内容。
首先我们回顾以下概念：

- 令$U, V, W$为向量空间，$T_1 : U \to V$，$T_2 : V \to W$为线性变换，那么$T_2 \circ T_1$代表$T_2$与$T_1$的复合(composition of $T_2$ with $T_1$)，即对于任何$\boldsymbol{u} \in U$，有

$$(T_2 \circ T_1)(\boldsymbol{u}) = T_2(T_1(\boldsymbol{u})) : \boldsymbol{u} \mapsto T_1(\boldsymbol{u}) \mapsto T_2(T_1(\boldsymbol{u})).$$

  显然，**两个线性变换的复合仍是一个线性变换**。
- 我们称一个线性变换$T : V \to W$为**一一映射(one-to-one)或者单射(injective)**，如果对于任何$\boldsymbol{v}_1 \neq \boldsymbol{v}_2$，都有$T(\boldsymbol{v}_1) \neq T(\boldsymbol{v}_2)$。见Definition 4.49。
- 我们称一个线性变换$T : V \to W$为**满射(surjective,onto)**如果对于任何$\boldsymbol{w} \in W$都存在一个$\boldsymbol{v} \in V$使得$T(\boldsymbol{v}) = \boldsymbol{w}$，也即$T$的值域RAN$(T) = W$。见Definition 4.49。
- 我们称$T : V \to W$是一个**双射(bijective)**，如果$T$同时是单射和满射。见Definition 4.49。

特别地，我们对双射的线性变换还有以下称呼。

**Definition 5.6.** 令$V, W$为向量空间，$T : V \to W$为一个双射*(bijective)*的线性变换。那么$T$也被称为一个***同构(Isomorphism)***。对于两个向量空间$V$与$W$，如果**存在**一个双射的线性变换$T : V \to W$，那么我们称$V$与$W$**是同构的*(V is isomorphic to W)***。

**例子：** 令$V = P_n = \{a_0 + a_1x + \ldots + a_nx^n : a_0, a_1, \ldots, a_n \in \mathbb{R}\}$，令$W = \{a_0 + a_1x + \ldots + a_{n+1}x^{n+1} : a_0, a_1, \ldots, a_{n+1} \in \mathbb{R}\}$。定义一个从$V$到$W$的线性变换

$$T : V \to W, \quad T(p(x)) = xp(x),$$

即对于多项式$p(x) = a_0 + a_1x + \ldots + a_nx^n$，$T(p(x)) = a_0x + a_1x^2 + \ldots + a_nx^{n+1}$。显然，我们有如下结论：

- $T: V \to W$是一个一一映射：如果$p(x) \in V, q(x) \in V$满足$p(x) \neq q(x)$，那么将$p(x)$记为$p(x) = a_0 + a_1x + \ldots + a_nx^n$，将$q(x)$记为$q(x) = b_0 + b_1x + \ldots + b_nx^n$，我们必然可以找到某个$i = 0, \ldots, n$使得$a_i \neq b_i$。因此，对于$W = P_{n+1}$的标准基底$B = \{1, x, \ldots, x^{n+1}\}$，我们有

$$[T(p(x))]_B = (0, a_0, \ldots, a_i, \ldots, a_n) \neq (0, b_0, \ldots, b_i, \ldots, b_n) = [T(q(x))]_B,$$

  那么由坐标向量表示的唯一性可知，$T(p(x)) \neq T(q(x))$。
- $T: V \to W$不是一个满射：显然，由$T$的定义可知，对于常数函数$1 \in W = P_{n+1}$，不存在任何一个多项式$p(x) \in V = P_n$使得$T(p(x)) = 1$。更确切的说，$T$的值域RAN$(T)$等于

$$\{f(x) \in P_{n+1} : [f(x)]_B = (0, c_1, \ldots, c_{n+1}) \in \mathbb{R}^{n+2}\}.$$

在第四章我们已经学习到，对于一个$n$维向量空间$V$，一旦我们固定它的一组基底$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$，那么$V$里的任何向量$\boldsymbol{v} \in V$都可以被它的坐标向量(coordinate vector)$[\boldsymbol{v}]_B \in \mathbb{R}^n$唯一表示，并且许多$V$空间上的问题可以借此转化为$\mathbb{R}^n$空间上的问题，比如Theorem 4.19($\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\} \subset V$线性无关当且仅当它们的坐标向量$\{[\boldsymbol{w}_1]_B, \ldots, [\boldsymbol{w}_k]_B\}$在$\mathbb{R}^n$里线性无关)。坐标向量的这个性质可以被总结为以下定理：它是$n$维向量空间$V$与$n$维欧式空间$\mathbb{R}^n$之间的一个同构。

**Theorem 5.7.** 任何$n$维向量空间$V$都与$n$维欧式空间$\mathbb{R}^n$同构。

证明. 假设$V$为$n$维向量空间，固定它的一组基底$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$，那么我们知道，任何$\boldsymbol{v} \in V$都可以被唯一的表示为$\boldsymbol{v} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n$，此时我们有$[\boldsymbol{v}]_B = (k_1, \ldots, k_n) \in \mathbb{R}^n$。现在定义$T : V \to \mathbb{R}^n$为$V$关于基底$B$的坐标向量映射，即对于$\boldsymbol{v} \in V$，定义

$$T(\boldsymbol{v}) = [\boldsymbol{v}]_B \in \mathbb{R}^n.$$

我们现在证明$T$是一个同构。

- 首先验证$T$是一个线性变换。对于$\boldsymbol{v} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n$，$\boldsymbol{u} = c_1\boldsymbol{v}_1 + \ldots + c_n\boldsymbol{v}_n$，我们有

$$\boldsymbol{u} + \boldsymbol{v} = (k_1 + c_1)\boldsymbol{v}_1 + \ldots + (k_n + c_n)\boldsymbol{v}_n$$

  为$\boldsymbol{u} + \boldsymbol{v}$关于基底$B$的线性组合表示。显然，由坐标向量的定义，可得

$$\begin{aligned} T(\boldsymbol{u} + \boldsymbol{v}) = [\boldsymbol{u} + \boldsymbol{v}]_B = (k_1 + c_1, \ldots, k_n + c_n) &= (k_1, \ldots, k_n) + (c_1, \ldots, c_n) \\ &= [\boldsymbol{v}]_B + [\boldsymbol{u}]_B = T(\boldsymbol{u}) + T(\boldsymbol{v}). \end{aligned}$$

  同理，可证$T(k\boldsymbol{v}) = [k\boldsymbol{v}]_B = k[\boldsymbol{v}]_B = kT(\boldsymbol{v})$。综上，$T : V \to \mathbb{R}^n$是一个线性变换。

- 现在证明$T$是一个一一映射。假如$\boldsymbol{v} \neq \boldsymbol{u}$，但有$T(\boldsymbol{u}) = T(\boldsymbol{v})$，即$[\boldsymbol{u}]_B = [\boldsymbol{v}]_B = (k_1, \ldots, k_n)$，那么后者意味着$\boldsymbol{v} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n = \boldsymbol{u}$，矛盾！因此必然有$\boldsymbol{u} \neq \boldsymbol{v} \Rightarrow T(\boldsymbol{u}) \neq T(\boldsymbol{v})$。
- 最后证明$T$是满射。对于任何$(k_1, \ldots, k_n) \in \mathbb{R}^n$，我们令$\boldsymbol{v} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n \in V$，那么显然有

$$T(\boldsymbol{v}) = [\boldsymbol{v}]_B = (k_1, \ldots, k_n),$$

从而有$\text{RAN}(T) = \mathbb{R}^n$。

综上，可知$T : V \to \mathbb{R}^n$为一个同构，因此也有$V$与$\mathbb{R}^n$同构。 □

以下定理可以让我们快速判断一个线性变换是否是一一映射，它的证明与Theorem 4.51的部分证明完全相同，因此在这里省略推导过程。具体证明可参考英文教材Theorem 8.2.1.

**Theorem 5.8.** *令$T : V \to W$为一个线性变换，那么以下说法等价：*

1. *$T$为一个一一映射。*
2. *$Ker(T) = \{\mathbf{0}\}$。*

利用以上定理结合秩–零化度定理Theorem 5.5我们可以证明(可以参考Theorem 4.51的证明)：

**Theorem 5.9.** *假设$V$是一个有限维向量空间，$T : V \to W$是一个从$V$到$W$的线性变换，并假设$dim(V) = dim(W) = n$。那么以下说法等价：*

1. *$T$为一个一一映射。*
2. *$Ker(T) = \{\mathbf{0}\}$。*
3. *$T$为一个满射。*
4. *$T$为一个同构。*

证明. "$T$为一个一一映射"与"$\text{Ker}(T) = \{\mathbf{0}\}$"等价已经在上一个定理中被证明。
由秩–零化度定理Theorem 5.5可知

$$n = \dim(V) = \text{rank}(T) + \text{nullity}(T).$$

如果$\text{Ker}(T) = \{\mathbf{0}\}$，那么必然有$\text{rank}(T) = \dim(\text{RAN}(T)) = n - 0 = n$，即$\text{RAN}(T)$作为$W$的子空间与$W$拥有同样的维数。显然此时只能有$\text{RAN}(T) = W$，即$T : V \to W$为一个满射。反之，如果$T : V \to W$是一个满射，那么有$\text{RAN}(T) = W$，从而

有$\text{rank}(T) = \dim(\text{RAN}(T)) = \dim(W) = n$，因此利用秩–零化度定理Theorem 5.5可得

$$\text{nullity}(T) = \dim(V) - \text{rank}(T) = n - n = 0.$$

即$\text{Ker}(T) = \{\mathbf{0}\}$。

我们已经证明$T : V \to W$在$\dim(V) = \dim(W) = n$的情况下是一一映射当且仅当它是一个满射。因此可得“$T$为一个一一映射”与“$T$为一个满射”与“$T$为一个同构”等价。 □

**注意：以上定理Theorem 5.9 只在$\dim(V) = \dim(W) = n$时成立！若没有该假设则不能使用这个结论！**

实际上，利用秩–零化度定理Theorem 5.5我们可以得到以下结论，具体证明细节留给大家自己验证，或参考第九次作业Problem D.

**Theorem 5.10.** *令$V, W$为两个向量空间，$T : V \to W$为一个线性变换。*

1. *如果$\dim(V) > \dim(W)$，那么$T$必然不可能是一一映射。*
2. *如果$\dim(V) < \dim(W)$，那么$T$必然不可能是满射。*

如果$T : V \to W$是一个同构，那么模仿Theorem 2.17的证明，我们知道$T$存在一个逆映射$T^{-1} : W \to V$，且这个$T^{-1}$仍为一个同构。如果$T : V \to W$**仅为一一映射但不为满射**，那么由于$\text{RAN}(T) \neq W$，即对于某些$\boldsymbol{w} \in W$不存在$\boldsymbol{v} \in V$使得$T(\boldsymbol{v}) = \boldsymbol{w}$，我们无法将$T$的逆映射定义在整个空间$W$上。然而，如果我们只考虑$T$的值域$\text{RAN}(T)$，那么显然每个$\boldsymbol{w} \in \text{RAN}(T)$都拥有一个唯一的原像$\boldsymbol{v} \in V$使得$T(\boldsymbol{v}) = \boldsymbol{w}$，这个唯一性是由$T$的一一映射性质所决定的。因此，我们能够在$\text{RAN}(T)$上定义一个映射$S : \text{RAN}(T) \to V$使得$S(\boldsymbol{w}) = \boldsymbol{v}$，$\boldsymbol{v} \in V$是唯一能够满足$T(\boldsymbol{v}) = \boldsymbol{w}$的原像。这样的$S$显然是一个线性变换，我们称其为**$T$的逆映射(inverse)**。

**Theorem 5.11.** *对任何线性变换$T : V \to W$，如果它是一一映射，那么总存在定义在$RAN(T)$上的逆映射$T^{-1} : RAN(T) \to V$使得对于任何$\boldsymbol{w} \in RAN(T)$，$T^{-1}(\boldsymbol{w}) = \boldsymbol{v}$，这个$\boldsymbol{v} \in V$满足$T(\boldsymbol{v}) = \boldsymbol{w}$。此外，$T^{-1} : RAN(T) \to V$满足：*

- *对于任何$\boldsymbol{v} \in V$，有$(T^{-1} \circ T)(\boldsymbol{v}) = \boldsymbol{v}$。*
- *对于任何$\boldsymbol{w} \in RAN(T)$，有$(T \circ T^{-1})(\boldsymbol{w}) = \boldsymbol{w}$。*
- *$T^{-1} : RAN(T) \to V$是一一映射也是满射，因此对于一一映射$T$，$RAN(T)$与$V$是同构的。*

**例子：** 令$V = P_n = \{a_0 + a_1x + \ldots + a_nx^n : a_0, a_1, \ldots, a_n \in \mathbb{R}\}$，令$W = \{a_0 + a_1x + \ldots + a_{n+1}x^{n+1} : a_0, a_1, \ldots, a_{n+1} \in \mathbb{R}\}$。定义一个从$V$到$W$的线性变换

$$T : V \to W, \quad T(p(x)) = xp(x),$$

即对于多项式$p(x) = a_0 + a_1x + \ldots + a_nx^n$，$T(p(x)) = a_0x + a_1x^2 + \ldots + a_nx^{n+1}$。我们已经知道这样的线性变换$T$是一个一一映射，且它的值域为

$$\begin{aligned}\mathrm{RAN}(T) &= \{f(x) \in P_{n+1} : f(x) = c_1x + \ldots + c_{n+1}x^{n+1}\} \\ &= \{f(x) \in P_{n+1} : [f(x)]_B = ({\color{red}0}, c_1, \ldots, c_{n+1}) \in \mathbb{R}^{n+2}\}, B = \{1, x, x^2, \ldots, x^{n+1}\}.\end{aligned}$$

容易验证，此时$T^{-1} : \mathrm{RAN}(T) \to V$是以下映射

$$T^{-1}(c_1x + \ldots + c_{n+1}x^{n+1}) = c_1 + c_2x + \ldots + c_{n+1}x^n.$$

比如，对于$V = P_3$，$W = P_4$，$T^{-1}(2x - x^2 + 5x^3 + 3x^4) = 2 - x + 5x^2 + 3x^3$。

以下定理的证明很直接，可参考英文教材Theorem 8.2.4.

**Theorem 5.12.** *如果$T_1 : U \to V$与$T_2 : V \to W$为一一映射的线性变换，那么*

1. *$T_2 \circ T_1 : U \to W$为一一映射。*
2. *令$(T_2 \circ T_1)^{-1} : RAN(T_2 \circ T_1) \to U$为$T_2 \circ T_1$的逆映射，那么有*

$${\color{red}(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}}.$$

### 5.3. 线性变换的矩阵表示, Matrices for general linear transformations

本节的核心内容是将两个(抽象)向量空间的线性变换$T : V \to W$用欧式空间之间的**矩阵变换**$T_A : \mathbb{R}^n \to \mathbb{R}^m, T_A(\boldsymbol{x}) = A\boldsymbol{x}$表示(这里假设$\dim(V) = n, \dim(W) = m$)，并且证明矩阵$A \in M_{m\times n}$完全刻画了线性变换$T$的性质。这里沟通抽象向量空间$V$(或$W$)与欧式空间$\mathbb{R}^n$(或$\mathbb{R}^m$)之间的桥梁是**坐标向量映射**。

更具体来讲，假设$V$为$n$维向量空间，$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$是$V$的一组基底；假设$W$为$m$维向量空间，$B' = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m\}$是$W$的一组基底；对于任意给定的线性变换$T : V \to W$，我们试图找到一个$m \times n$的矩阵$A \in M_{m\times n}$使得下图成立：

$$\begin{tikzcd} \boldsymbol{v} \in V \arrow[r, "T"] \arrow[d, "{\color{blue}f_B}"] & T(\boldsymbol{v}) \in W \arrow[d, "{\color{blue}g_{B'}}"] \\ ([\boldsymbol{v}]_B \in \mathbb{R}^n) \arrow[r, red, "T_A"] & ([T(\boldsymbol{v})]_{B'} \in \mathbb{R}^m) \end{tikzcd}$$

**记号解释：**

- $f_B: V \to \mathbb{R}^n$为$V$上关于$B$的坐标映射，对于$\boldsymbol{v} \in V$，如果$\boldsymbol{v} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n$，那么$f_B(\boldsymbol{v}) = [\boldsymbol{v}]_B = (k_1, \ldots, k_n) \in \mathbb{R}^n$；
- $g_{B'}: W \to \mathbb{R}^m$为$W$上关于$B'$的坐标映射，对于$\boldsymbol{w} \in W$，如果$\boldsymbol{w} = c_1\boldsymbol{w}_1 + \ldots + c_m\boldsymbol{w}_m$，那么$g_{B'}(\boldsymbol{w}) = [\boldsymbol{w}]_{B'} = (c_1, \ldots, c_m) \in \mathbb{R}^m$；
- $T_A: \mathbb{R}^n \to \mathbb{R}^m$为由矩阵$A \in M_{m\times n}$决定的矩阵变换，即$T_A(\boldsymbol{x}) = A\boldsymbol{x}$。

换句话说，我们的目的在于找到这样的矩阵$A \in M_{m\times n}$使得对于任何$\boldsymbol{v} \in V$，都有

$$[T(\boldsymbol{v})]_{B'} = A[\boldsymbol{v}]_B$$

成立，即$T(\boldsymbol{v}) \in W$**关于基底**$B'$**的坐标向量**$[T(\boldsymbol{v})]_{B'} \in \mathbb{R}^m$**可以通过矩阵**$A$**乘以**$\boldsymbol{v} \in V$**关于**$B'$**的坐标向量**$[\boldsymbol{v}]_B \in \mathbb{R}^n$**得到**。特别地，我们应该注意到：
这样的矩阵$A \in M_{m\times n}$依赖于$V$的基底$B$与$W$的基底$B'$的选取，如果我们更换$V$的基底或者$W$的基底，则$A$也随之改变。我们将在下一节详细学习这样的矩阵$A$是如何随着基底选取而改变的。

下面我们证明这样的矩阵$A \in M_{m\times n}$的存在性以及确定它的具体构造。
假如存在$A \in M_{m\times n}$使得

$$[T(\boldsymbol{v})]_{B'} = A[\boldsymbol{v}]_B$$

对任何$\boldsymbol{v} \in V$都成立，那么代入$\boldsymbol{v} = \boldsymbol{v}_1, \ldots, \boldsymbol{v} = \boldsymbol{v}_n$(这里我们设$V$的基底$B$为$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$)，我们得到

$$[T(\boldsymbol{v}_1)]_{B'} = A[\boldsymbol{v}_1]_B, \quad , [T(\boldsymbol{v}_n)]_{B'} = A[\boldsymbol{v}_n]_B.$$

另一方面，对于$V$的基底$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$，我们当然有

$$[\boldsymbol{v}_1]_B = \boldsymbol{e}_1, \ldots, [\boldsymbol{v}_n]_B = \boldsymbol{e}_n,$$

这里$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$为$\mathbb{R}^n$的标准单位向量。因此，我们实际上得到了

$$\begin{aligned}
[T(\boldsymbol{v}_1)]_{B'} &= A[\boldsymbol{v}_1]_B = A\boldsymbol{e}_1 = \text{A 的第1列} \\
&\vdots \\
[T(\boldsymbol{v}_n)]_{B'} &= A[\boldsymbol{v}_n]_B = A\boldsymbol{e}_n = \text{A 的第n列}
\end{aligned}$$

至此，我们可以得到结论：如果这样的矩阵$A$存在，那么它必须满足第一列等于$[T(\boldsymbol{v}_1)]_{B'} \in \mathbb{R}^m$，第二列为$[T(\boldsymbol{v}_2)]_{B'} \in \mathbb{R}^m$，...，第$n$列为$[T(\boldsymbol{v}_n)]_{B'} \in \mathbb{R}^m$，即

$$A = \begin{bmatrix} [T(\boldsymbol{v}_1)]_{B'} & [T(\boldsymbol{v}_2)]_{B'} & \ldots & [T(\boldsymbol{v}_n)]_{B'}. \end{bmatrix}$$

现在我们验证以上矩阵$A$确实实现了$[T(\boldsymbol{v})]_{B'} = A[\boldsymbol{v}]_B$对任何$\boldsymbol{v} \in V$都成立。对于任何$\boldsymbol{v} \in V$，假设$\boldsymbol{v} = k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n$，即$[\boldsymbol{v}]_B = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$，我们先利用$T$的线性性质得到

$$T(\boldsymbol{v}) = T(k_1\boldsymbol{v}_1 + \ldots + k_n\boldsymbol{v}_n) = k_1T(\boldsymbol{v}_1) + \ldots + k_nT(\boldsymbol{v}_n).$$

接下来，由于在Theorem 5.7中我们已经证明坐标向量映射$g_{B'} : W \to \mathbb{R}^m$，$g_{B'}(\boldsymbol{w}) = [\boldsymbol{w}]_{B'}$为一个线性变换，因此以上等式告诉我们

$$\begin{aligned}
[T(\boldsymbol{v})]_{B'} = g_{B'}(T(\boldsymbol{v})) = g_{B'}(k_1T(\boldsymbol{v}_1) + \ldots + k_nT(\boldsymbol{v}_n)) &= k_1g_{B'}(T(\boldsymbol{v}_1)) + \ldots + k_ng_{B'}(T(\boldsymbol{v}_n)) \\
&= k_1[T(\boldsymbol{v}_1)]_{B'} + \ldots + k_1[T(\boldsymbol{v}_n)]_{B'} \\
&= \begin{bmatrix} [T(\boldsymbol{v}_1)]_{B'} & [T(\boldsymbol{v}_2)]_{B'} & \ldots & [T(\boldsymbol{v}_n)]_{B'} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} \\
&= A[\boldsymbol{v}]_B.
\end{aligned}$$

现在我们可以做出结论：$A = \begin{bmatrix} [T(\boldsymbol{v}_1)]_{B'} & [T(\boldsymbol{v}_2)]_{B'} & \ldots & [T(\boldsymbol{v}_n)]_{B'} \end{bmatrix}$为所求矩阵。

**Definition 5.13.** 令$V$为$n$维向量空间，$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$为它的一组基底；令$V$为$n$维向量空间，$B' = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$为它的一组基底。那么对任何线性变换$T : V \to W$，我们称矩阵

$$\begin{bmatrix} [T(\boldsymbol{v}_1)]_{B'} & [T(\boldsymbol{v}_2)]_{B'} & \ldots & [T(\boldsymbol{v}_n)]_{B'} \end{bmatrix}$$

为$T$关于基底$B$与$B'$的矩阵表示***(the matrix for $T$ relative to $B$ and $B'$)***，并将其记为$[T]_{B',B}$。

我们现在将以上分析总结为以下定理。

**Theorem 5.14.** 令$V$为$n$维向量空间，$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$为它的一组基底；令$V$为$n$维向量空间，$B' = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$为它的一组基底。那么对任何线性变换$T : V \to W$，任何$\boldsymbol{v} \in V$，都有

$$[T(\boldsymbol{v})]_{B'} = [T]_{B',B}[\boldsymbol{v}]_B.$$

这里$[T]_{B',B} = \begin{bmatrix} [T(\boldsymbol{v}_1)]_{B'} & [T(\boldsymbol{v}_2)]_{B'} & \ldots & [T(\boldsymbol{v}_n)]_{B'} \end{bmatrix}$。

**例子：** 对于$V=\mathbb{R}^n$，$W=\mathbb{R}^m$，我们知道任何线性变换$T:\mathbb{R}^n\to\mathbb{R}^m$都是一个矩阵变换，其标准矩阵为

$$[T]=\begin{bmatrix}T(\boldsymbol{e}_1) & \dots & T(\boldsymbol{e}_n)\end{bmatrix}.$$

令$B=\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$为$\mathbb{R}^n$上的标准基底，令$B'=\{\boldsymbol{e}_1',\dots,\boldsymbol{e}_m'\}$为$\mathbb{R}^m$上的标准基底，容易验证$T$的标准矩阵事实上就是$T$关于欧式空间上的标准基底$B$与$B'$的矩阵表示，即

$$[T]=[T]_{B',B}.$$

**例子：2022-2023年线性代数期末考试填空题**

Let $V$ be the subspace in $F(-\infty,\infty)$ spanned by $B=\{\sin x,\cos x\}$. Let $T:V\to\mathbb{R}^2$ be defined by $T(f)=(f(0),f(\frac{\pi}{2}))$ for $f\in V$. Take $B'=\{(1,0),(0,1)\}$ as a basis of $\mathbb{R}^2$. Then the matrix for $T$ relative to $B$ and $B'$ is

$$[T]_{B',B}=?$$

**答案：** 由定义，可知

$$[T]_{B',B}=\begin{bmatrix}[T(\sin x)]_{B'} & [T(\cos x)]_{B'}\end{bmatrix},$$

其中

$$[T(\sin x)]_{B'}=(\sin 0,\sin\frac{\pi}{2})=(0,1),$$
$$[T(\cos x)]_{B'}=(\cos 0,\cos\frac{\pi}{2})=(1,0).$$

因此，所求矩阵为

$$[T]_{B',B}=\begin{bmatrix}[T(\sin x)]_{B'} & [T(\cos x)]_{B'}\end{bmatrix}=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.$$

有了以上的知识储备，我们可以轻松得到线性变换复合与线性变换的逆映射的矩阵表示：

**Theorem 5.15.** 设$U$为$n$维向量空间，$B$为$U$的一组基底，$V$为$k$维向量空间，$\tilde{B}$为$V$的一组基底，$W$为$m$维向量空间，$B'$为$W$的一组基底。那么对线性变换$T_1:U\to V$，线性变换$T_2:V\to W$，有

$$[T_2\circ T_1]_{B',B}=[T_2]_{B',\tilde{B}}[T_1]_{\tilde{B},B}.$$

即$T_2\circ T_1$关于基底$B$与$B'$的矩阵等于$T_2$关于基底$B'$与$\tilde{B}$的矩阵左乘以$T_1$关于基底$\tilde{B}$与$B$的矩阵。

证明. 该定理的证明过程包含在以下图表中，只需注意到$A=[T_1]_{\tilde{B},B}$与$C=[T_2]_{B',\tilde{B}}$。具体细节留给大家自行验证。

$$\begin{array}{ccccc}
\boldsymbol{u}\in U & \xrightarrow{\quad T_1 \quad} & T_1(\boldsymbol{u})\in V & \xrightarrow{\quad T_2 \quad} & T_2(T_1(\boldsymbol{u}))\in W \\
{\scriptstyle f_B}\downarrow & & {\scriptstyle h_{\tilde{B}}}\downarrow & & {\scriptstyle g_{B'}}\downarrow \\
([\boldsymbol{u}]_B\in\mathbb{R}^n) & \xrightarrow{T_A} & ([T_1(\boldsymbol{u})]_{\tilde{B}}=A[\boldsymbol{u}]_B\in\mathbb{R}^k) & \xrightarrow{T_C} & ([T_2(T_1(\boldsymbol{u}))]_{B'}=(CA)[\boldsymbol{u}]_B\in\mathbb{R}^m)
\end{array}$$

□

正如矩阵变换$T:\mathbb{R}^n\to\mathbb{R}^m$的一一映射/满射/同构性质可以通过观察它的标准矩阵$[T]$的对应性质来得到，任何向量空间之间的线性变换$T:V\to W$的一一映射/满射/同构性质也反映在它关于(任何)基底的矩阵表示中。具体来讲，我们有以下定理：

**Theorem 5.16.** 设$V$，$W$为有限维向量空间，$T:V\to W$为一个线性变换。

1. 如果$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}\subset V$为$V$里的一个线性无关集合，且$T:V\to W$为**一一映射**，那么$S'=\{T(\boldsymbol{v}_1),\ldots,T(\boldsymbol{v}_r)\}\subset W$也是一个线性无关集合。
2. 如果$B=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}\subset V$为$V$里的一组基底，且$T:V\to W$为**同构**，那么$B'=\{T(\boldsymbol{v}_1),\ldots,T(\boldsymbol{v}_n)\}\subset W$是$W$的一组基底。

证明. 如果$T:V\to W$是一个一一映射，那么由Theorem 5.8我们知道$\mathrm{Ker}(T)=\{\mathbf{0}\}$。现在假设$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}\subset V$为$V$里的一个线性无关集合，我们考虑等式

$$k_1T(\boldsymbol{v}_1)+\ldots+k_rT(\boldsymbol{v}_r)=\mathbf{0}.$$

利用$T$的线性性质，以上等式可以写成

$$T(k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r)=\mathbf{0},$$

□

因此可得$k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r\in\mathrm{Ker}(T)$。那么因为已知 $\mathrm{Ker}(T)=\{\mathbf{0}\}$，我们必然有

$$k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r=\mathbf{0}.$$

利用$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$的线性无关性质，可得以上等式成立的必要条件是$k_1=\ldots=k_r=0$，也即等式$k_1T(\boldsymbol{v}_1)+\ldots+k_rT(\boldsymbol{v}_r)=\mathbf{0}$.只有当$k_1=\ldots=k_r=0$时成立，从而有$S'=\{T(\boldsymbol{v}_1),\ldots,T(\boldsymbol{v}_r)\}\subset W$也是一个线性无关集合。
现在假设$T:V\to W$是一个同构，那么必然有$\dim(V)=\dim(W)$。假设$B=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}\subset V$为$V$里的一组基底，那么根据$T$的满射性质可得$\mathrm{RAN}(T)=\mathrm{span}\{T(\boldsymbol{v}_1),\ldots,T(\boldsymbol{v}_n)\}$ $W$。又由于$T$是一一映射，我们已经证明此时$\{T(\boldsymbol{v}_1),\ldots,T(\boldsymbol{v}_n)\}\subset W$为$W$里的线性无关集合。综上，可得$B'=\{T(\boldsymbol{v}_1),\ldots,T(\boldsymbol{v}_n)\}\subset W$是$W$的一组基底。

**Theorem 5.17.** 令$V$为$n$维向量空间，令$W$为$m$维向量空间，$T:V\to V$为$V$上的线性算子，且令$B$为$V$的一组基底，$B'$为$W$的一组基底。令$[T]_{B',B}\in M_{m\times n}$为$T$关于基底$B$与$B'$的矩阵表示。那么我们有

1. $\mathit{rank}(T)=\mathit{rank}([T]_{B',B})$，$\mathit{nullity}(T)=\mathit{nullity}([T]_{B',B})$。
2. $T$是一个一一映射当且仅当$\mathit{nullity}([T]_{B',B})=0$。
3. $T$是一个满射当且仅当$\mathit{rank}([T]_{B',B})=m=\dim(W)$。
4. 若$n=m$，那么$T$是一个同构当且仅当$[T]_{B',B}\in M_{n\times n}$可逆。此时$T^{-1}:W\to V$关于基底$B'$与$B$的矩阵表示为
$$[T^{-1}]_{B,B'}=([T]_{B',B})^{-1}.$$

证明. 该定理的整个证明思路体现在以下图表中：

$$\begin{array}{ccc}
\boldsymbol{v}\in V & \xrightarrow{\quad T\quad} & T(\boldsymbol{v})\in W \\
\big\downarrow f_B & & \big\uparrow g_{B'}^{-1} \\
([\boldsymbol{v}]_B\in\mathbb{R}^n) & \xrightarrow[T_A]{} & ([T(\boldsymbol{v})]_{B'}=A[\boldsymbol{v}]_B\in\mathbb{R}^m)
\end{array}$$

这里$A=[T]_{B',B}$，$f_B:V\to\mathbb{R}^n$为关于基底$B$的坐标向量映射($f_B(\boldsymbol{v})=[\boldsymbol{v}]_B$)，$g_{B'}:W\to\mathbb{R}^m$为关于基底$B'$的坐标向量映射($g_{B'}(\boldsymbol{w})=[\boldsymbol{w}]_{B'}$)。显然以上图表告诉我们
$$T=g_{B'}^{-1}\circ T_A\circ f_B.$$
由于$f_B:V\to\mathbb{R}^n$与$g_{B'}:W\to\mathbb{R}^m$是同构，它们的逆映射$f_B^{-1}:\mathbb{R}^n\to V$与$g_{B'}^{-1}:\mathbb{R}^m\to W$均存在且也为同构(见Theorem 5.11)，故由以上等式我们又可以得到
$$\begin{aligned}T=g_{B'}^{-1}\circ T_A\circ f_B &\Rightarrow g_{B'}\circ T\circ f_B^{-1}=g_{B'}\circ(g_{B'}^{-1}\circ T_A\circ f_B)\circ f_B^{-1}=(g_{B'}\circ g_{B'}^{-1})\circ T_A\circ(f_B\circ f_B^{-1})\\ &\Rightarrow g_{B'}\circ T\circ f_B^{-1}=T_A,\end{aligned}$$
注意这里我们当然利用了等式$f_B\circ f_B^{-1}=\mathrm{Id}_{\mathbb{R}^n}$为$\mathbb{R}^n$上的恒等算子(identity operator)，即$(f_B\circ f_B^{-1})(\boldsymbol{x})=\boldsymbol{x}$对任何$\boldsymbol{x}\in\mathbb{R}^n$成立，同理$g_{B'}\circ g_{B'}^{-1}=\mathrm{Id}_{\mathbb{R}^m}$为$\mathbb{R}^m$上的恒等算子。现在我们观察等式
$$T=g_{B'}^{-1}\circ T_A\circ f_B.$$
由于$f_B:V\to\mathbb{R}^n$为同构，特别地，$f_B$是一个满射，由第九次作业Problem E的第二问(rank$(S\circ T)=$ rank$(S)$，如果$T$为满射)我们有
$$\mathrm{rank}(T_A\circ f_B)=\mathrm{rank}(T_A);$$
由于$g_{B'}^{-1}:\mathbb{R}^m\to W$也是一个同构，特别地，它是一个一一映射，那么再次利用Problem E的第二问(rank$(S\circ T)=$ rank$(T)$，如果$S$为一一映射)可得
$$\mathrm{rank}(T)=\mathrm{rank}(g_{B'}^{-1}\circ T_A\circ f_B)=\mathrm{rank}(T_A\circ f_B)=\mathrm{rank}(T_A)=\mathrm{rank}(A)=\mathrm{rank}([T]_{B',B}).$$

此时利用rank–nullity theorem 5.5 我们可以立刻得到

$$\text{nullity}(T) = \dim(V) - \text{rank}(T) = n - \text{rank}([T]_{B',B}) = \text{nullity}([T]_{B',B}).$$

利用等式$\text{rank}(T) = \text{rank}([T]_{B',B})$，$\text{nullity}(T) = \text{nullity}([T]_{B',B})$，我们容易推出：

$$\begin{aligned} T \text{ 是一一映射} &\iff \text{Ker}(T) = \{\mathbf{0}\} \\ &\iff \text{nullity}(T) = 0 \\ &\iff \text{nullity}([T]_{B',B}) = 0 \\ &\iff ([T]_{B',B}) = \{\mathbf{0}\} \\ &\iff T_A : \mathbb{R}^n \to \mathbb{R}^m \text{ 是一一映射}, A = [T]_{B',B}, \end{aligned}$$

以及

$$\begin{aligned} T \text{ 是满射} &\iff \text{RAN}(T) = W \\ &\iff \text{rank}(T) = \dim(W) = m \\ &\iff \text{rank}([T]_{B',B}) = m \\ &\iff \text{Col}([T]_{B',B}) = \mathbb{R}^m \\ &\iff T_A : \mathbb{R}^n \to \mathbb{R}^m \text{ 是满射}, A = [T]_{B',B}. \end{aligned}$$

如果$T : V \to W$是一个同构，那么我们知道$T \circ T^{-1} : W \to W$满足$(T \circ T^{-1})(\boldsymbol{w}) = \boldsymbol{w}$对任何$\boldsymbol{w} \in W$成立，即$[T \circ T^{-1}]_{B',B'} = I_n$。那么利用Theorem 5.15可得

$$[T]_{B',B}[T^{-1}]_{B,B'} = I_n$$

从而可以推出

$$[T^{-1}]_{B,B'} = ([T]_{B',B})^{-1}.$$

□

作为以上重要定理的一个推论，我们有以下结果：

**Theorem 5.18.** *令$V$为$n$维向量空间，$T : V \to V$为$V$上的线性算子，且令$B$为$V$上的一组基底。那么以下说法等价：*

1. *$T$是一个一一映射。*
2. *$[T]_{B,B}$可逆。*

*此时$T^{-1}$关于基底$B$与$B$的矩阵表示为*

$$[T^{-1}]_{B,B} = ([T]_{B,B})^{-1}.$$

证明. 我们首先证明，如果$T : V \to V$是一个一一映射，那么$[T]_{B,B}$为可逆矩阵。因为$T$的定义域和到达域都是$V$，由Theorem 5.9可知$T : V \to V$是一一映射等价于$T : V \to V$是满射，即$\mathrm{RAN}(T) = V$。因此对于$T : V \to V$，它是一个一一映射等价于它是满射。$[T]_{B,B}$可逆与$[T^{-1}]_{B,B} = ([T]_{B,B})^{-1}$这两个性质可以从Theorem 5.17中得到。 □

### 5.4. 相似矩阵, Similarity

在本节我们回答以下问题：

设$V$为一个$n$维向量空间，$T : V \to V$为一个$V$上的线性算子。假设$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$为$V$的一组基底，那么我们已经知道$T$关于$B$的矩阵表示$[T]_{B,B}$满足

$$[T]_{B,B} = \begin{bmatrix} [T(\boldsymbol{v}_1)]_B & [T(\boldsymbol{v}_2)]_B & \ldots & [T(\boldsymbol{v}_n)]_B \end{bmatrix}.$$

假设$B' = \{\boldsymbol{v}_1', \ldots, \boldsymbol{v}_n'\}$是$V$的另一组基底，那么类似可得$T$关于$B'$的矩阵表示为

$$[T]_{B',B'} = \begin{bmatrix} [T(\boldsymbol{v}_1')]_{B'} & [T(\boldsymbol{v}_2')]_{B'} & \ldots & [T(\boldsymbol{v}_n')]_{B'} \end{bmatrix}.$$

**那么这两个矩阵$[T]_{B,B}$与$[T]_{B',B'}$满足怎样的关系？**

要回答这个问题，我们需要引入以下重要概念：

**Definition 5.19.** 设$A \in M_{n\times n}, B \in M_{n\times n}$为两个$n$阶方阵。如果存在一个可逆矩阵$P \in M_{n\times n}$使得

$$B = P^{-1}AP$$

(等价于$A = PBP^{-1}$)，那么我们称$A$与$B$相似, *A is similar to B, A and B are similar*。

**注意：** 关于相似性，我们有以下基本性质：

1. 任何$A \in M_{n\times n}$都与它自己相似：只需取$P = I_n$为单位矩阵，则显然有$A = I_n^{-1}AI_n$。
2. 如果$A$与$B$相似，那么$B$与$A$相似：$B = P^{-1}AP \iff A = PBP^{-1} = (P^{-1})^{-1}BP^{-1}$。
3. 如果$A$与$B$相似，$B$与$C$相似，那么$A$与$C$相似：若$A$与$B$相似，那么存在可逆的$P \in M_{n\times n}$使得$B = P^{-1}AP$；若$B$与$C$相似，则存在可逆的$Q \in M_{n\times n}$使得$C = Q^{-1}BQ$，由此可得

$$C = Q^{-1}BQ = Q^{-1}(P^{-1}AP)Q = (Q^{-1}P^{-1})A(PQ) = (PQ)^{-1}A(PQ).$$

4. 如果$A$与$B$相似，即$B=P^{-1}AP$，那么对于任何多项式$f(x)=a_0+a_1x+\ldots+a_mx^m$，多项式矩阵$f(A)$与$f(B)$也相似：

$$\begin{aligned}
f(B) &= a_0I + a_1B + a_2B^2 + \ldots + a_mB^m \\
&= a_0(P^{-1}IP) + a_1(P^{-1}AP) + a_2(P^{-1}AP)^2 + \ldots + a_m(P^{-1}AP)^m \\
&= a_0(P^{-1}IP) + a_1(P^{-1}AP) + a_2(P^{-1}A^2P) + \ldots + a_m(P^{-1}A^mP) \\
&= P^{-1}(a_0I + a_1A + a_2A^2 + \ldots + a_mA^m) \\
&= P^{-1}f(A)P.
\end{aligned}$$

以下定理回答了我们之前提出的问题：**对于$V$上的不同基底$B$与$B'$，同一个线性算子$T:V\to V$关于它们的矩阵表示$[T]_{B,B}$与$[T]_{B',B'}$相似**。

**Theorem 5.20.** *令$V$为$n$维向量空间，$B$与$B'$为$V$的两组基底，$T:V\to V$为$V$上的线性算子。那么$[T]_{B,B}$与$[T]_{B',B'}$相似。更具体地说，令$P_{B\leftarrow B'}$为从基底$B'$到基底$B$的转移矩阵(transition matrix)，那么有*

$$[T]_{B',B'} = (P_{B\leftarrow B'})^{-1}[T]_{B,B}P_{B\leftarrow B'}.$$

*证明.* 令$I:V\to V$为$V$上的恒等算子，即$I(\boldsymbol{v})=\boldsymbol{v}$对任何$\boldsymbol{v}\in V$成立。那么显然有

$$T = I\circ T\circ I.$$

对于右边的恒等算子$I:V\to V$，我们考虑它关于基底$B'$与基底$B$的矩阵表示$[I]_{B,B'}$；对于左边的恒等算子$I:V\to V$，我们考虑它关于基底$B$与基底$B'$的矩阵表示$[I]_{B',B}$。那么利用Theorem 5.15可得

$$[T]_{B',B'} = [I\circ T\circ I]_{B',B'} = [I]_{B',B}[T]_{B,B}[I]_{B,B'}.$$

这个过程如下图所示：

$$\begin{tikzcd}
\boldsymbol{v}\in V \arrow[r, "I"] \arrow[d, "f_{B'}"'] & \boldsymbol{v}\in V \arrow[r, "T"] \arrow[d, "f_B"'] & T(\boldsymbol{v})\in V \arrow[r, "I"] \arrow[d, "f_B"'] & T(\boldsymbol{v})\in V \arrow[d, "f_{B'}"'] \\
([\boldsymbol{v}]_{B'}) \arrow[r, "T_P"] & ([I]_{B,B'}[\boldsymbol{v}]_{B'}=[\boldsymbol{v}]_B) \arrow[r, "T_A"] & (A[\boldsymbol{v}]_B=[T(\boldsymbol{v})]_B) \arrow[r, "T_Q"] & ([I]_{B',B}[T(\boldsymbol{v})]_B=[T(\boldsymbol{v})]_{B'})
\end{tikzcd}$$

这里$P=[I]_{B,B'}$，$A=[T]_{B,B}$，$Q=[I]_{B',B}$。
现在我们只需证明$[I]_{B,B'}=P_{B\leftarrow B'}$以及$[I]_{B',B}=(P_{B\leftarrow B'})^{-1}=P_{B'\leftarrow B}$即可。但显然，根据$[I]_{B,B'}$以及转移矩阵$P_{B\leftarrow B'}$的定义，令$B=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}$，令$B'=\{\boldsymbol{v}'_1,\ldots,\boldsymbol{v}'_n\}$，我们有

$$[I]_{B,B'} = \begin{bmatrix}[I(\boldsymbol{v}'_1)]_B & [I(\boldsymbol{v}'_2)]_B & \ldots & [I(\boldsymbol{v}'_n)]_B\end{bmatrix} = \begin{bmatrix}[\boldsymbol{v}'_1]_B & [\boldsymbol{v}'_2]_B & \ldots & [\boldsymbol{v}'_n]_B\end{bmatrix} = P_{B\leftarrow B'}.$$

同理可证$[I]_{B',B}=P_{B'\leftarrow B}=(P_{B\leftarrow B'})^{-1}$。 □

下面这个定理告诉我们，如果$A$与$B$相似，那么它们的很多性质都相同。

**Theorem 5.21.** 对于任何$A, B \in M_{n\times n}$，如果$A$与$B$相似，那么

1. $\det(A) = \det(B)$;
2. *rank*$(A) =$ *rank*$(B)$;
3. *nullity*$(A) =$ *nullity*$(B)$;
4. *tr*$(A) =$ *tr*$(B)$，即$A$与$B$有相同的迹。

以上指标因此也被称为**相似不变量*(the similarity invariants)***。

证明. 因为$A$与$B$相似，那么存在可逆矩阵$P$使得$B = P^{-1}AP$。显然此时有

$$\det(B) = \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \det(P)^{-1}\det(P)\det(A) = \det(A).$$

又因为$P$可逆，矩阵变换$T_P : \mathbb{R}^n \to \mathbb{R}^n$与$T_{P^{-1}} : \mathbb{R}^n \to \mathbb{R}^n$是同构，因此利用Theorem 5.17的证明技巧可得

$$\text{rank}(B) = \text{rank}(P^{-1}AP) = \text{rank}(T_{P^{-1}} \circ T_A \circ T_P) = \text{rank}(A),$$

利用rank–nullity theorem 5.5可得nullity$(B) =$ nullity$(A)$。
从$B = P^{-1}AP$推出tr$(A) =$ tr$(B)$的过程较为繁琐，不在考试内容中，因此省略证明步骤。**但请大家记住这个结论。** □

**例子：2022-2023年线性代数期末延期考试填空题**

Let $x, y$ be two real numbers. Suppose that $A$ and $B$ are similar, where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & x \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Then $x =?, y =?$

**答案：** 这道题考察的是相似不变量。已知$A$和$B$相似，那么它们必然有相同的行列式：

$$\det(A) = -1 = \det(B) = -y,$$

由此可得$y = 1$。再利用相似矩阵有同样的迹，可得tr$(A) = 1 + x =$ tr$(B) = 1 + y + (-1) = y = 1$，可得$x = 0$为所求。

一般来说，与相似矩阵相关的考试题会结合下一章我们要学习的特征值/特征向量的相关知识一起考察。故在本节暂不提供更多例子。

# 6 第六章：特征值与特征向量

本章对应英文教材的第五章。

矩阵的特征值(eignevalue)与特征向量(eigenvector)在数学与物理中的偏微分方程，统计学中的主成分分析，计算机视觉中的人脸识别等众多领域都有广泛的应用，因此本章节的内容将会在大学本科四年以及硕博期间的研究中频繁的用到。

## 6.1. 特征值与特征向量, eigenvalue and eigenvector

**Definition 6.1.** 令$A \in M_{n\times n}$为$n$阶方阵。我们称一个实数$\lambda \in \mathbb{R}$是矩阵$A$的一个特征值*(eigenvalue)*，如果存在一个非零的向量$\boldsymbol{x} \in \mathbb{R}^n$，$\boldsymbol{x} \neq \boldsymbol{0}$，使得

$$A\boldsymbol{x} = \lambda \boldsymbol{x}.$$

此时，我们称这样的$\boldsymbol{x}$为关于$\lambda$的一个特征向量*(eigenvector corresponding to $\lambda$)*。

令$T_A : \mathbb{R}^n \to \mathbb{R}^n$为$A$对应的矩阵变换，我们有时也称$\lambda$为$T_A$的特征值，$\boldsymbol{x}$为$T_A$的特征向量。

**例子：** $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$是矩阵$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$关于$\lambda = 3$的一个特征向量：

$$A\boldsymbol{x} = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 3\boldsymbol{x}.$$

**Theorem 6.2.** $\lambda \in \mathbb{R}$是$A$的一个特征值当且仅当$\det(\lambda I_n - A) = 0$。

*证明.*

$$\begin{aligned} \lambda \text{ 是A的特征值} &\iff \text{存在 } \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0}, A\boldsymbol{x} = \lambda\boldsymbol{x} \\ &\iff \text{存在 } \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0}, (\lambda I_n - A)\boldsymbol{x} = \boldsymbol{0} \\ &\iff \text{Ker}(\lambda I_n - A) \neq \{\boldsymbol{0}\} \\ &\iff \lambda I_n - A \text{ 不可逆} \\ &\iff \det(\lambda I_n - A) = 0. \end{aligned}$$

□

**Definition 6.3.** 对于给定的矩阵$A \in M_{n\times n}$，函数

$$p : \lambda \in \mathbb{R} \mapsto p(\lambda) = \det(\lambda I_n - A) \in \mathbb{R}$$

被称为$A$的**特征多项式*(characteristic polynomial)***。方程$p(\lambda) = \det(\lambda I_n - A) = 0$被称为$A$的**特征方程*(characteristic equation)***。

**Theorem 6.4.** 对于$n$阶方阵$A \in M_{n\times n}$，它的特征多项式$p(\lambda) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_n$，这里$c_1, \ldots, c_n$为实数。另外，$\lambda \in \mathbb{R}$是$A$的一个特征值当且仅当$p(\lambda) = 0$，即$\lambda$是特征多项式的一个实数根*(root)*。

证明. 对矩阵阶数$n$使用数学归纳法，可证明

$$p(\lambda) = \det(\lambda I_n - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \ldots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \ldots & \lambda - a_{nn} \end{vmatrix} = \lambda^n + c_1\lambda^{n-1} + \ldots + c_n.$$

利用Theorem 6.2可轻松证明定理剩下的部分。 □

**Remark 6.5.** 考虑2阶方阵$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$。简单计算可得它的特征多项式为

$$p(\lambda) = \begin{vmatrix} \lambda - a & -b \\ -c & \lambda - d \end{vmatrix} = (\lambda - a)(\lambda - d) - bc = \lambda^2 - (a+d)\lambda + (ad - bc).$$

因此此时$p(\lambda) = \lambda^n + c_1\lambda^{n-1} + c_2$的系数分别为

$$c_1 = -tr(A), \quad c_2 = \det(A).$$

实际上，我们可以通过计算验证，对于任意$A \in M_{n\times n}$，它特征多项式$p(\lambda) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_n$中$\lambda^{n-1}$前面的系数$c_1$满足

$$c_1 = -tr(A),$$

常数项$c_n$则满足

$$c_n = (-1)^n \det(A).$$

**Remark 6.6.** 注意，我们目前要求$A$的特征值$\lambda$必须是个实数。因此，并非每个矩阵$A \in M_{n\times n}$都有实数特征值。比如矩阵

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

的特征多项式$p(\lambda) = \lambda^2 + 1$，显然$p(\lambda) = 0$没有实数根(当然它有复数根，分别为$i$与$-i$)。

**例子：** 矩阵

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

的特征多项式为

$$p(\lambda) = \begin{vmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -4 & -17 & \lambda - 8 \end{vmatrix} = \lambda^3 - 8\lambda^2 + 17\lambda - 4.$$

求解特征方程$p(\lambda) = \lambda^3 - 8\lambda^2 + 17\lambda - 4 = 0$，我们得到

$$\begin{aligned} 0 = p(\lambda) &= \lambda^3 - 8\lambda^2 + 17\lambda - 4 \\ &= \lambda^3 - 8\lambda^2 + 16\lambda + \lambda - 4 \\ &= \lambda(\lambda^2 - 8\lambda + 16) + \lambda - 4 \\ &= \lambda(\lambda - 4)^2 + \lambda - 4 \\ &= (\lambda - 4)(\lambda(\lambda - 4) + 1) \\ &= (\lambda - 4)(\lambda^2 - 4\lambda + 1) \end{aligned}$$

因此方程的根，即$A$的特征值为

$$\lambda_1 = 4, \lambda_2 = 2 + \sqrt{3}, \lambda_3 = 2 - \sqrt{3}.$$

以下关于特征值与特征向量的一些基本性质容易证明，留作习题。

**Theorem 6.7.** *以下说法都成立：*

1. 对于任何$n$阶方阵$A \in M_{n\times n}$，它最多有$n$个不同的特征值。
2. 对于上三角矩阵$A = \begin{bmatrix} a_{11} & \star & \star & \dots & \star \\ 0 & a_{22} & \star & \dots & \star \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}$它有特征值为对角线上的元素$a_{11}, \dots, a_{nn}$。该结论对下三角矩阵也成立。
3. 假设$\lambda \in \mathbb{R}$为$A$的一个特征值，$\boldsymbol{x} \in \mathbb{R}^n$为$A$关于$\lambda$的一个特征向量，那么对于任何多项式$f(x) = a_0 + a_1 x + \dots + a_m x^m$，$f(\lambda)$是多项式矩阵$f(A) = a_0 I_n + a_1 A + \dots + a_m A^m$的一个特征值，$\boldsymbol{x}$为$f(A)$关于$f(\lambda)$的一个特征向量。
4. $A$可逆当且仅当$\lambda = 0$不是$A$的特征值。
5. 如果$\lambda \in \mathbb{R}$为$A$的一个特征值，$\boldsymbol{x} \in \mathbb{R}^n$为$A$关于$\lambda$的一个特征向量，且$A$可逆，那么对于任何多项式$f(x) = a_0 + a_1 x + \dots + a_m x^m$，$f(\frac{1}{\lambda})$是多项式矩阵$f(A^{-1}) = a_0 I_n + a_1 A^{-1} + \dots + a_m A^{-m}$的一个特征值，$\boldsymbol{x}$为$f(A^{-1})$关于$f(\frac{1}{\lambda})$的一个特征向量。

**例子：2017年南京大学数学系线性代数期中考试题**

设$A$为2阶可逆方阵，$\lambda_1, \lambda_2$为$A$的两个**整数**特征值。已知矩阵$B = A^{-2} - 6A^{-1}$的特

征值为$-5$和7。求$\lambda_1, \lambda_2$。

**答案：** 由以上定理Theorem 6.7可知，如果$\lambda_1, \lambda_2$为$A$的两个特征值，那么$B = A^{-2} - 6A^{-1}$的两个特征值分别为

$$\frac{1}{\lambda_1^2} - \frac{6}{\lambda_1}, \frac{1}{\lambda_2^2} - \frac{6}{\lambda_2}.$$

现在已知$B = A^{-2} - 6A^{-1}$的特征值为$-5$和7，因此解方程

$$\frac{1}{\lambda_1^2} - \frac{6}{\lambda_1} = -5, \frac{1}{\lambda_2^2} - \frac{6}{\lambda_2} = 7$$

可知$\lambda_1$需要满足

$$(5\lambda_1 - 1)(\lambda_1 - 1) = 0,$$

$\lambda_2$需要满足

$$(7\lambda_2 - 1)(\lambda_2 + 1) = 0.$$

由于要求$\lambda_1, \lambda_2$均为整数，所以求得$\lambda_1 = 1, \lambda_2 = -1$。

在本节的最后，我们学习如何求得一个特征值对应的特征向量。

**Theorem 6.8.** 假设$A \in M_{n\times n}$，$\lambda$为$A$的一个特征值。假设$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为*Null*$(\lambda I_n - A)$的一组基底。那么任何一个$A$关于$\lambda$的特征向量$\boldsymbol{x}$都可以表示为

$$\boldsymbol{x} = k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r$$

这里系数$k_1, \ldots, k_r$不全为0。

证明. 显然我们有以下等价关系：

$$\begin{aligned} \boldsymbol{x}\text{ 是A关于 }\lambda\text{ 的特征向量} &\iff A\boldsymbol{x} = \lambda\boldsymbol{x} \\ &\iff (\lambda I_n - A)\boldsymbol{x} = \mathbf{0} \\ &\iff \boldsymbol{x} \in \mathrm{Null}(\lambda I_n - A). \end{aligned}$$

因此，如果$B = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为Null$(\lambda I_n - A)$的一组基底，那么$\boldsymbol{x} = k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r$。因为特征向量不能为零向量，所以$k_1, \ldots, k_r$不全为0。 □

**Definition 6.9.** 假设$A \in M_{n\times n}$，$\lambda$为$A$的一个特征值。我们称*Null*$(\lambda I_n - A)$为$A$关于$\lambda$的**特征空间*(the eigensapce of $A$ corresponding to the eigenvalue $\lambda$)***。

以下结果非常容易验证，但同时非常重要：

**Theorem 6.10.** 假设$A \in M_{n\times n}$，$\lambda_1, \lambda_2$为$A$的两个不同的特征值。将$V_{\lambda_1}$记为$A$关于$\lambda_1$的特征空间，将$V_{\lambda_2}$记为$A$关于$\lambda_2$的特征空间.那么有

$$V_{\lambda_1} \cap V_{\lambda_2} = \{\mathbf{0}\}.$$

特别地，我们有$\dim(V_{\lambda_1} + V_{\lambda_2}) = \dim(V_{\lambda_1}) + \dim(V_{\lambda_2})$。

证明. 令$\boldsymbol{x} \in V_{\lambda_1} \cap V_{\lambda_2}$，那么由特征空间的定义可知

$$A\boldsymbol{x} = \lambda_1 \boldsymbol{x}, \quad A\boldsymbol{x} = \lambda_2 \boldsymbol{x}$$

从而有

$$\mathbf{0} = A\boldsymbol{x} - A\boldsymbol{x} = \lambda_1 \boldsymbol{x} - \lambda_2 \boldsymbol{x} = (\lambda_1 - \lambda_2)\boldsymbol{x}.$$

由于$\lambda_1 - \lambda_2 \neq 0$，以上等式只有当$\boldsymbol{x} = \mathbf{0}$时成立，从而有$V_{\lambda_1} \cap V_{\lambda_2} = \{\mathbf{0}\}$。利用维数公式$\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W)$可得等式$\dim(V_{\lambda_1} + V_{\lambda_2}) = \dim(V_{\lambda_1}) + \dim(V_{\lambda_2})$。 □

### 6.2. 对角化, Diagonalization

在5.4节我们学习了一个很重要的概念：矩阵的相似。回顾Definition 5.19，对于两个矩阵$A, B \in M_{n \times n}$，我们称$A$与$B$相似，如果存在一个可逆矩阵$P \in M_{n \times n}$使得

$$B = P^{-1}AP.$$

如果两个矩阵$A, B$是相似的，那么它们的很多性质是相同的，这些指标被叫做相似不变量(similarity invariants)。在5.4节我们已经知道，矩阵的行列式，秩，零化度，以及迹都是相似不变量。这里我们补充额外的几个**相似不变量**：

- 如果两个矩阵$A, B$是相似的，那么它们的**特征多项式**相同。具体来讲，令$p_A$标记$A$的特征多项式，令$p_B$标记$B$的特征多项式，那么对任何$\lambda \in \mathbb{R}$都有

$$\begin{aligned} p_B(\lambda) = \det(\lambda I_n - B) &= \det(\lambda I_n - P^{-1}AP) \\ &= \det(P^{-1}(\lambda I_n - A)P) \\ &= \det(P^{-1}) \det(\lambda I_n - A) \det(P) \\ &= \det(P^{-1}) \det(P) \det(\lambda I_n - A) \\ &= \det(\lambda I_n - A) = p_A(\lambda), \end{aligned}$$

  这里我们显然利用了事实$\det(P^{-1}) = \frac{1}{\det(P)}$。
- 如果两个矩阵$A, B$是相似的，那么它们的**特征值**相同。这是因为特征值是特征多项式的根。我们刚刚证明了对于相似的两个矩阵$A$和$B$，它们具有相同的特征多项式，因此它们当然也有一样的特征值。
- 如果两个矩阵$A, B$是相似的，令$\lambda \in \mathbb{R}$为$A$的一个特征值，那么我们已经知道$\lambda$也是$B$的一个特征值。令$V_\lambda$标记$A$关于$\lambda$的特征空间，即$V_\lambda = \mathrm{Null}(\lambda I_n - A)$；令$W_\lambda$标记$B$关于$\lambda$的特征空间，即$W_\lambda = \mathrm{Null}(\lambda I_n - B)$，那么我们有

$$\dim(V_\lambda) = \dim(W_\lambda).$$

要证明这个结果，我们先假设$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是$V_\lambda$的一组基底(即$\dim(V_\lambda) = r$)，显然根据特征空间的定义，我们有$A\boldsymbol{v}_j = \lambda\boldsymbol{v}_j$对任何$j = 1, \ldots, r$都成立。现在考虑集合$M = \{P^{-1}\boldsymbol{v}_1, \ldots, P^{-1}\boldsymbol{v}_r\}$。注意，对于任何$j = 1, \ldots, r$我们都有

$$B(P^{-1}\boldsymbol{v}_j) = (P^{-1}AP)P^{-1}\boldsymbol{v}_j = P^{-1}A(PP^{-1}\boldsymbol{v}_j) = P^{-1}(A\boldsymbol{v}_j) = P^{-1}(\lambda\boldsymbol{v}_j) = \lambda(P^{-1}\boldsymbol{v}_j).$$

这当然意味着集合$M = \{P^{-1}\boldsymbol{v}_1, \ldots, P^{-1}\boldsymbol{v}_r\} \subset W_\lambda$属于$B$关于$\lambda$的特征空间$W_\lambda$。由于$P^{-1}$是个可逆矩阵，它所对应的矩阵变换$T_{P^{-1}} : \mathbb{R}^n \to \mathbb{R}^n$是一个同构，因此利用Theorem 5.16 我们可得集合$M = \{P^{-1}\boldsymbol{v}_1, \ldots, P^{-1}\boldsymbol{v}_r\}$仍是一个线性无关结合。那么$M = \{P^{-1}\boldsymbol{v}_1, \ldots, P^{-1}\boldsymbol{v}_r\} \subset W_\lambda$则意味着$W_\lambda$的基底个数不小于$M$里包含的元素个数$r$(见Theorem 4.16)，即$\dim(W_\lambda) \geq r = \dim(V_\lambda)$。
类似的，我们假设$S' = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\}$是$W_\lambda$的一组基底，那么利用关系$A = PBP^{-1}$，重复以上分析，我们可以证明集合 $M' = \{P\boldsymbol{w}_1, \ldots, P\boldsymbol{w}_k\} \subset V_\lambda$是$A$关于$\lambda$的特征空间$V_\lambda$的线性无关子集，从而有$\dim(V_\lambda) = r \geq k = \dim(W_\lambda)$。综上，我们得到$\dim(V_\lambda) = \dim(W_\lambda)$。

我们将上面的讨论结果总结为以下定理。

**Theorem 6.11.** 假设两个矩阵$A, B \in M_{n\times n}$相似。那么它们具有相同的**特征多项式**，相同的**特征值**，并且对任何共同的特征值$\lambda \in \mathbb{R}$，它们对应的**特征空间的维数(*eigenspace dimension*)**也相同。此外，如果$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$是$A$关于$\lambda$的特征空间$V_\lambda$的一组基底，那么$\{P^{-1}\boldsymbol{v}_1, \ldots, P^{-1}\boldsymbol{v}_r\}$是$B$关于$\lambda$的特征空间$W_\lambda$的一组基底。反之，如果$\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\}$是$B$关于$\lambda$的特征空间$W_\lambda$的一组基底，那么$\{P\boldsymbol{w}_1, \ldots, P\boldsymbol{w}_r\}$是$A$关于$\lambda$的特征空间$V_\lambda$的一组基底。特别地，如果$\boldsymbol{x}$是$A$关于$\lambda$的一个特征向量，那么$P^{-1}\boldsymbol{x}$是$B$关于$\lambda$的一个特征向量；如果$\boldsymbol{y}$是$B$关于$\lambda$的一个特征向量，那么$P\boldsymbol{y}$是$A$关于$\lambda$的一个特征向量。

**注意：** 一般来说对于相似的矩阵$A$和$B$，它们对于共同的特征值$\lambda$的特征向量未必相同：它们之间的差异在于要左乘矩阵$P^{-1}$或矩阵$P$，如果$P \neq I_n$，那么它们就可能不一样，比如对于矩阵$A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$与矩阵$B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$，可以证明它们是相似的(见下方讨论)，且特征值都为2和3；另一方面也容易验证向量$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$是$A$关于2的特征向量，但不是$B$关于2的特征向量。

我们特别关心一个给定的矩阵$A$是否可以跟一个**对角矩阵(diagonal matrix)**相似。

**Definition 6.12.** 对于一个$A \in M_{n\times n}$，如果存在一个**对角矩阵**$D \in M_{n\times n}$使得$A$与$D$相似，即

$$D = P^{-1}AP$$

对某个可逆矩阵$P \in M_{n\times n}$成立，那么我们称矩阵$A$**可对角化*(diagonalizable)***。

**例子：** 对于上一个例子中出现的矩阵$A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$，我们可以验证，令$P = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$，我们有$P^{-1}AP = B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$，因此$A$与对角矩阵$B$相似，即这个矩阵$A$可对角化。

如果一个矩阵$A$可以被对角化，那么很多与$A$有关的计算会变得简单很多。比如对于上面的矩阵$A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$，要求$A^{2023}$，我们只需利用相似关系$B = P^{-1}AP = B$得到

$$B^{2023} = (P^{-1}AP)^{2023} = P^{-1}A^{2023}P$$

即$A^{2023} = PB^{2023}P^{-1}$。显然由于$B = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$为对角矩阵，它的高次幂计算非常简单：$B^{2023} = \begin{bmatrix} 2^{2023} & 0 \\ 0 & 3^{2023} \end{bmatrix}$。现在只需计算$PB^{2023}P^{-1}$即可得到$A^{2023}$。

我们接下来最重要的任务是

- **判断一个给定的$A \in M_{n\times n}$都否可以对角化；**
- **如果可以对角化，如何求出可逆矩阵$P$与对角矩阵$D$使得$D = P^{-1}AP$。**

我们首先回答第一个问题。以下定理给出了一个方阵是否能够对角化的充分必要条件。

**Theorem 6.13.** *对于任何$n$阶方阵$A \in M_{n\times n}$，以下说法等价：*

1. *$A$可对角化；*
2. *$A$具有$n$个线性无关的特征向量，即存在一组$\mathbb{R}^n$的基底$B$，它里面的向量全部是$A$的特征向量。*

*证明.* • 假设$A \in M_{n\times n}$可被对角化，即存在可逆矩阵$P \in M_{n\times n}$，对角矩阵$D$，使得

$$D = P^{-1}AP \iff AP = PD.$$

这里我们将$P$记为$P = \begin{bmatrix} \boldsymbol{c}_1 & \dots & \boldsymbol{c}_n \end{bmatrix}$，即$\boldsymbol{c}_1, \dots, \boldsymbol{c}_n$为$P$的列向量；将对角矩阵$D$记为$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$。基本的矩阵乘法运算告诉我们

$$AP = \begin{bmatrix} A\boldsymbol{c}_1 & A\boldsymbol{c}_2 & \dots & A\boldsymbol{c}_n \end{bmatrix}$$

$$PD = \begin{bmatrix} \lambda_1\boldsymbol{c}_1 & \lambda_2\boldsymbol{c}_2 & \dots & \lambda_n\boldsymbol{c}_n \end{bmatrix}.$$

因此等式$AP = PD$意味着

$$A\boldsymbol{c}_1 = \lambda_1\boldsymbol{c}_1, A\boldsymbol{c}_2 = \lambda_2\boldsymbol{c}_2, \dots, A\boldsymbol{c}_n = \lambda_n\boldsymbol{c}_n,$$

即**$\boldsymbol{c}_j$是$A$关于$\lambda_j$的一个特征向量**($j = 1, \dots, n$)。又由于$P$为可逆矩阵，它的列向量$\boldsymbol{c}_1, \dots, \boldsymbol{c}_n$线性无关，因此集合$B = \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_n\}$完全由$A$的特征向量组成，且它是$\mathbb{R}^n$的一组基底。

- 假设$B = \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_n\}$是$\mathbb{R}^n$的一组基底，且它们全都是$A$的特征向量，即存在$\lambda_j$，$j = 1, \dots, n$使得

$$A\boldsymbol{c}_j = \lambda_j\boldsymbol{c}_j.$$

那么只需令$P = \begin{bmatrix} \boldsymbol{c}_1 & \dots & \boldsymbol{c}_n \end{bmatrix}$，令$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$，那么通过之前步骤的计算，容易验证$AP = PD$成立。由于$\text{rank}(P) = n$，可知$P$可逆，因此等式$AP = PD$可以写成$D = P^{-1}AP$，即$A$可对角化。

□

下面我们证明来自$A$的不同特征值的特征向量组成的集合线性无关。

**Theorem 6.14.** *假设$\lambda_1, \dots, \lambda_k$为$A$的特征值且它们各不相等(英文： they are distinct)。那么以下说法成立。*

1. *若$\boldsymbol{v}_j$是$A$关于$\lambda_j$，$j = 1, \dots, k$的一个特征向量，那么它们所组成的集合$S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$线性无关。*
2. *令$V_{\lambda_j} = Null(\lambda_j I_n - A)$为$A$关于$\lambda_j$的特征空间，$j = 1, \dots, k$，且$\{\boldsymbol{x}_1^{\lambda_j}, \dots, \boldsymbol{x}_{d_j}^{\lambda_j}\}$为$V_{\lambda_j}$的一组基底，$j = 1, \dots, k$，$d_j$为$V_{\lambda_j}$的维数，那么它们所组成的集合*

$$B = \{\boldsymbol{x}_1^{\lambda_j}, \dots, \boldsymbol{x}_{d_j}^{\lambda_j} : j = 1, \dots, k\}$$

为线性无关集合。

证明. 该证明不要求一定要掌握，但请务必记住定理内容。

1. 已知$\boldsymbol{v}_j$是$A$关于$\lambda_j$，$j=1,\ldots,k$的一个特征向量。假设$r$为使得$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$线性无关的**最大整数**。由于$\boldsymbol{v}_1$是$A$的一个特征向量，根据定义它不为零向量，因此$r\geq 1$。我们要**证明**$r=k$**成立**。利用反证法，我们假设$r<k$，即$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$线性无关，但是$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r,\boldsymbol{v}_{r+1}\}$线性相关(注意由于假设$r<k$，所以$r+1\leq k$)。后者意味着存在不全为0的系数$c_1,\ldots,c_{r+1}$使得

$$c_1\boldsymbol{v}_1+\ldots+c_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}.$$

等式两边同时左乘矩阵$A$，我们得到

$$A(c_1\boldsymbol{v}_1+\ldots+c_{r+1}\boldsymbol{v}_{r+1})=A\mathbf{0}=\mathbf{0},$$

另一方面，由假设我们知道$A\boldsymbol{v}_1=\lambda_1\boldsymbol{v}_1,\ldots,A\boldsymbol{v}_{r+1}=\lambda_{j+1}\boldsymbol{v}_{r+1}$，因此以上等式实际上可以写为

$$c_1\lambda_1\boldsymbol{v}_1+\ldots+c_{r+1}\lambda_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}.$$

现在将等式$c_1\boldsymbol{v}_1+\ldots+c_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}$两边同时乘以实数$\lambda_{r+1}$，可得

$$c_1\lambda_{r+1}\boldsymbol{v}_1+\ldots+c_{r+1}\lambda_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}.$$

将以上两个蓝色标记的等式相减，我们得到：

$$\begin{aligned}\mathbf{0}&=c_1\lambda_1\boldsymbol{v}_1+\ldots+c_{r+1}\lambda_{r+1}\boldsymbol{v}_{r+1}-(c_1\lambda_{r+1}\boldsymbol{v}_1+\ldots+c_{r+1}\lambda_{r+1}\boldsymbol{v}_{r+1})\\&=c_1(\lambda_{r+1}-\lambda_1)\boldsymbol{v}_1+c_2(\lambda_{r+1}-\lambda_2)\boldsymbol{v}_2+\ldots+c_r(\lambda_{r+1}-\lambda_r)\boldsymbol{v}_r.\end{aligned}$$

由我们对$r$的假设，$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r$线性无关，因此以上等式成立的必要条件是$c_1(\lambda_{r+1}-\lambda_1)=c_2(\lambda_{r+1}-\lambda_2)=\ldots=c_r(\lambda_{r+1}-\lambda_r)=0$。又因为$\lambda_{r+1}-\lambda_1\neq 0,\ldots,\lambda_{r+1}-\lambda_r\neq 0$，以上等式成立的必要条件变成了$c_1=c_2=\ldots=c_r=0$。因此，等式$c_1\boldsymbol{v}_1+\ldots+c_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}$实际上化简为$c_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}$。然而，因为$\boldsymbol{v}_{r+1}\neq\mathbf{0}$(它是一个特征向量，故不能为零向量)，$c_{r+1}=0$必须成立。这意味着$c_1\boldsymbol{v}_1+\ldots+c_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}$只有当$c_1=\ldots=c_{r+1}=0$时才能成立，这与存在不全为0的$c_1,\ldots,c_{r+1}$使得$c_1\boldsymbol{v}_1+\ldots+c_{r+1}\boldsymbol{v}_{r+1}=\mathbf{0}$成立矛盾。因此$r=k$必须成立，即 $S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k\}$线性无关。

2. 已知$\{\boldsymbol{x}_1^{\lambda_j},\ldots,\boldsymbol{x}_{d_j}^{\lambda_j}\}$为$V_{\lambda_j}$的一组基底，$j=1,\ldots,k$，$d_j$为$V_{\lambda_j}$的维数。考虑等式

$$(c_1^1\boldsymbol{x}_1^{\lambda_1}+\ldots+c_{d_1}^1\boldsymbol{x}_{d_1}^{\lambda_1})+\ldots+(c_1^k\boldsymbol{x}_1^{\lambda_k}+\ldots+c_{d_k}^k\boldsymbol{x}_{d_k}^{\lambda_k})=\mathbf{0}.$$

显然，令$\boldsymbol{v}_1 = c_1^1\boldsymbol{x}_1^{\lambda_1} + \ldots + c_{d_1}^1\boldsymbol{x}_{d_1}^{\lambda_1}, \ldots, \boldsymbol{v}_k = c_1^k\boldsymbol{x}_1^{\lambda_k} + \ldots + c_{d_k}^k\boldsymbol{x}_{d_k}^{\lambda_k}$，以上等式可以写为

$$\boldsymbol{v}_1 + \ldots + \boldsymbol{v}_k = \mathbf{0}$$

并且$\boldsymbol{v}_1 \in V_{\lambda_1}, \ldots, \boldsymbol{v}_k \in V_{\lambda_k}$。假设存在若干个指标$j_1, \ldots, j_r$(它们的取值范围为$1, \ldots, k$)使得$\boldsymbol{v}_{j_1} \neq \mathbf{0}, \ldots, \boldsymbol{v}_{j_r} \neq \mathbf{0}$，且对于所有$j \in \{1, \ldots, k\} \setminus \{j_1, \ldots, j_r\}$，有$\boldsymbol{v}_j = \mathbf{0}$。此时等式$\boldsymbol{v}_1 + \ldots + \boldsymbol{v}_k = \mathbf{0}$可以写为

$$\boldsymbol{v}_{j_1} + \ldots + \boldsymbol{v}_{j_r} = \mathbf{0}.$$

然而，对于不为零向量的$\boldsymbol{v}_{j_1}, \ldots, \boldsymbol{v}_{j_r}$，它们显然都是$A$的特征向量，我们刚刚证明了它们所组成的集合是线性无关的，这意味着$\boldsymbol{v}_{j_1} + \ldots + \boldsymbol{v}_{j_r} = \mathbf{0}$ 不能成立，矛盾！因此$\boldsymbol{v}_1 + \ldots + \boldsymbol{v}_k = \mathbf{0}$成立只有当$\boldsymbol{v}_1 = \ldots = \boldsymbol{v}_k = \mathbf{0}$。现在代入$\boldsymbol{v}_j = c_1^j\boldsymbol{x}_1^{\lambda_j} + \ldots + c_{d_j}^j\boldsymbol{x}_{d_j}^{\lambda_j}$，利用$\boldsymbol{x}_1^{\lambda_j}, \ldots, \boldsymbol{x}_{d_j}^{\lambda_j}$的线性无关性质(它们是$V_{\lambda_j}$的基底)，可得$c_1^j = \ldots = c_{d_j}^j = 0$对任何$j = 1, \ldots, k$成立。综上，$B = \{\boldsymbol{x}_1^{\lambda_j}, \ldots, \boldsymbol{x}_{d_j}^{\lambda_j} : j = 1, \ldots, k\}$线性无关。

□

作为直接推论，我们得到矩阵可对角化的一个**充分条件(sufficient condition)**。

**Theorem 6.15.** *如果$A \in M_{n\times n}$拥有$n$个不同的特征值，那么$A$可对角化。*

*证明.* 假设$A$的特征值为$\lambda_1, \ldots, \lambda_n$且它们各不相等，令$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$分别为$\lambda_1, \ldots, \lambda_n$对应的特征向量。由Theorem 6.14可知$\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^n$线性无关，因此它们构成$\mathbb{R}^n$的一组基底。利用Theorem 6.13我们得到$A$可对角化。 □

**注意：** $A$具有$n$个不同的特征值只是$A$可对角化的**充分而非必要条件**。比如矩阵$A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$的特征多项式为$p(\lambda) = (\lambda-2)^2(\lambda-1)$，仅此它作为3阶方阵只拥有两个不同的特征值2与1。但是$A$可以对角化：可以计算出它关于2的特征空间的一组基底为$\boldsymbol{v}_1 = (-1, 0, 1)$与$\boldsymbol{v}_2 = (0, 1, 0)$，关于1的特征空间的一组基底为$\boldsymbol{v}_3 = (-2, 1, 1)$，且容易验证$\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$线性无关。因此$A$可对角化。

以上观察实际上暗示给我们当矩阵$A \in M_{n\times n}$不具有$n$个不同的特征值时，一个判断它是否可对角化的标准。我们首先引入以下定义。

**Definition 6.16.** 令$A \in M_{n\times n}$。假设$a \in \mathbb{R}$是$A$的一个特征值，即对于$A$的特征多项式$p(\lambda)$而言，有$p(a) = 0$。那么我们知道$p(\lambda)$可以被写为

$$p(\lambda) = (\lambda - a)^c q(\lambda),$$

这里$c$为正整数，$q(\lambda)$为另一个多项式且它的最高次幂为$n - c$，并且$q(a) \neq 0$。令

$$d = \dim(V_a) = \dim(\mathit{Null}(aI_n - A)),$$

即$d$为$A$关于$a$的特征空间$V_a$的维数。那么我们称*$c$为$A$关于特征值$a$的代数重数(algebraic multiplicity)*，称*$d$为为$A$关于特征值$a$的几何重数(geometric multiplicity)*。

**注意：** 假设$A$全部的特征值为$\lambda_1, \ldots, \lambda_k$且它们互不相等。令$V_{\lambda_1}, \ldots, V_{\lambda_k}$为它们各自的特征空间，令$d_1, \ldots, d_k$为它们各自的几何重数。由Theorem 6.10我们知道，对于$\lambda_i \neq \lambda_j$，我们必然有$V_{\lambda_i} \cap V_{\lambda_j} = \{\mathbf{0}\}$，因此多次利用维数公式$\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W)$，容易推导出

$$\dim(V_{\lambda_1} + \ldots + V_{\lambda_k}) = \dim(V_{\lambda_1}) + \ldots + \dim(V_{\lambda_k}) = d_1 + \ldots + d_k.$$

显然，由于$V_{\lambda_1} + \ldots + V_{\lambda_k} \subset \mathbb{R}^n$，我们一直都有

$$d_1 + \ldots + d_k \leq n.$$

但是$d_1 + \ldots + d_k = n$未必对任何矩阵$A$都成立。如以下例子所示：

**例子：** 对于$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ -3 & 5 & 2 \end{bmatrix}$，可计算得到它的特征多项式为 $p(\lambda) = (\lambda - 2)^2(\lambda - 1)$。令$\lambda_1 = 2$，$\lambda_2 = 1$为它的两个特征值，由于

$$\lambda_1 I_3 - A = 2I_3 - A = \begin{bmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ 3 & -5 & -1 \end{bmatrix}$$

的秩为2，可得$d_1 = \dim(\mathbf{Null}(\lambda_1 I_3 - A)) = \mathbf{nullity}(\lambda_1 I_3 - A) = 3 - 2 = 1$；由于

$$\lambda_2 I_3 - A = I_3 - A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 3 & -5 & 0 \end{bmatrix}$$

的秩为2，可得$d_2 = \dim(\mathbf{Null}(\lambda_2 I_3 - A)) = \mathbf{nullity}(\lambda_2 I_3 - A) = 3 - 2 = 1$。因此此时有$d_1 + d_2 = 1 + 1 = 2 < 3$。注意，此时$\lambda_1 = 2$的代数重数为$c_1 = 2$，我们也有$1 = d_1 < c_1 = 2$，即它的几何重数小于它的代数重数。

以下定理的证明不要求掌握因此省略，但内容非常重要请务必记住!

**Theorem 6.17.** 对于任何一个$n$阶方阵$A \in M_{n\times n}$而言，我们有

1. 对于$A$的任何一个特征值，它的几何重数都不会大于它的代数重数。
2. $A$可对角化当且仅当$A$的特征多项式$p(\lambda)$的所有根都是实数，并且它所有特征值的几何重数等于代数重数。

综上，我们得到了如何判断一个矩阵$A \in M_{n\times n}$是否可对角化的标准，以及在它可对角化的情形下，将其对角化的计算过程。总结如下：

1. 计算$A$的特征多项式$p(\lambda) = \det(\lambda I_n - A)$，并求解特征方程$p(\lambda) = 0$。如果该方程存在不属于实数的解，那么$A$不能被对角化，停止；否则进入下方第二步。
2. 现在已知$p(\lambda) = 0$的所有根均为实数，假设其为$\lambda_1, \ldots, \lambda_k$，且它们互不相等，那么此时我们知道$p(\lambda)$可被表示为
$$p(\lambda) = (\lambda - \lambda_1)^{c_1}(\lambda - \lambda_2)^{c_2} \ldots (\lambda - \lambda_k)^{c_k}.$$
这里$c_j$对应特征值$\lambda_j$的**代数重数**，$j = 1, \ldots, k$。将其记录下来。
接下来依次计算$\lambda_1, \ldots, \lambda_k$的**几何重数**$d_1, \ldots, d_k$，即依次求出
$$\text{Null}(\lambda_1 I_n - A), \ldots, \text{Null}(\lambda_k I_n - A)$$
的**维数**。如果出现了某个$j$使得$d_j < c_j$，那么$A$不可对角化，停止；否则进入下方第三步。
3. 通过第二步，现在的情况是所有特征值的代数重数与几何重数都相等，即$A$可以对角化，即存在一个对角矩阵$D$与一个可逆矩阵$P$使得$D = P^{-1}AP$。由Theorem 6.13的证明过程与Theorem 6.14可知，如果$A$关于$\lambda_j$的特征空间$V_{\lambda_j}$的一组基底为$\{\boldsymbol{x}_1^{\lambda_j}, \ldots, \boldsymbol{x}_{d_j}^{\lambda_j}\}$，那么令$P$的列向量表示为：
$$P = \begin{bmatrix} \boldsymbol{x}_1^{\lambda_1} & \ldots & \boldsymbol{x}_{d_1}^{\lambda_1} & \boldsymbol{x}_1^{\lambda_2} & \ldots & \boldsymbol{x}_{d_2}^{\lambda_2} & \ldots & \boldsymbol{x}_1^{\lambda_k} & \ldots & \boldsymbol{x}_{d_j}^{\lambda_j} \end{bmatrix}$$
那么我们有
$$D = P^{-1}AP,$$
此时$D = \text{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{d_1\text{个}}, \underbrace{\lambda_2, \ldots, \lambda_2}_{d_2\text{个}}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{d_k\text{个}})$。因此，在这一步里我们要做的是**分别求出$A$关于$\lambda_j$的特征空间$V_{\lambda_j}$的一组基底$\{\boldsymbol{x}_1^{\lambda_j}, \ldots, \boldsymbol{x}_{d_j}^{\lambda_j}\}$，$j = 1, \ldots, k$，将其作为列向量拼成矩阵$P$，并写出对角矩阵$D$**。

我们现在以矩阵$A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$为例，判断它是否可对角化，并且在可以对角化的情况下，求出可逆矩阵$P$和对角矩阵$D$使得$D = P^{-1}AP$。

1. 我们首先计算$A$的特征多项式

$$p(\lambda) = \det(\lambda I_3 - A) = \det\left(\begin{bmatrix} \lambda & 0 & 2 \\ -1 & \lambda - 2 & -1 \\ -1 & 0 & \lambda - 3 \end{bmatrix}\right) = (\lambda - 2)(\lambda^2 - 3\lambda + 2)$$
$$= (\lambda - 2)^2(\lambda - 1).$$

求解特征方程$p(\lambda) = (\lambda - 2)^2(\lambda - 1) = 0$可得特征多项式的根为$\lambda_1 = 2, \lambda_2 - 1$；特别地，特征方程的所有解均为实数。因此我们进入下一个步骤。

2. 从第一步的计算中我们已知$\lambda_1 = 2$的代数重数为$c_1 = 2$，$\lambda_2 = 1$的代数重数为$c_2 = 1$。现在我们计算它们的几何重数。首先对于$\lambda_1 = 2$，有

$$\lambda_1 I_3 - A = 2I_3 - A = \begin{bmatrix} 2 & 0 & 2 \\ -1 & 0 & -1 \\ -1 & 0 & -1 \end{bmatrix}.$$

我们求出它的简化阶梯型为
$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
显然可以得到$\text{rank}(\lambda_1 I_3 - A) = 1$，从而可以推出$\lambda_1 = 2$的特征空间的维数为$\dim(\text{Null}(\lambda_1 I_3 - A)) = \text{nullity}(\lambda_1 I_3 - A) = 3 - \text{rank}(\lambda_1 I_3 - A) = 3 - 1 = 2$，即它的几何重数$d_1 = 2 = c_1$。
对于$\lambda_2 = 1$，由于它的代数重数$c_2 = 1$，且有它的几何重数$1 \leq d_2 \leq c_2 = 1$，因此必然有$c_2 = d_2 = 1$。
因此，$\lambda_1$的代数重数与几何重数相同，$\lambda_2$的代数重数与几何重数也相同，我们可以断定$A$可以对角化。因此转入下一步。

3. 我们现在来求特征空间$V_{\lambda_1} = \text{Null}(\lambda_1 I_3 - A)$和$V_{\lambda_2} = \text{Null}(\lambda_2 I_3 - A)$的一组基底。对于$\lambda_1 = 2$，我们已经在上一步中得到$\lambda_1 I_3 - A$的阶梯型为

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

利用高斯消元法容易求得它解空间的通解为

$$
\boldsymbol{s}=\begin{bmatrix}-t\\ r\\ t\end{bmatrix}=t\begin{bmatrix}-1\\ 0\\ 1\end{bmatrix}+r\begin{bmatrix}0\\ 1\\ 0\end{bmatrix},
$$

因此我们得到$V_{\lambda_1}$的一组基底为$\boldsymbol{x}_1^{\lambda_1}=\begin{bmatrix}-1\\ 0\\ 1\end{bmatrix}$，$\boldsymbol{x}_2^{\lambda_1}=\begin{bmatrix}0\\ 1\\ 0\end{bmatrix}$。

对于$\lambda_2=1$，有

$$
\lambda_2 I_3-A=I_3-A=\begin{bmatrix}1&0&2\\ -1&-1&-1\\ -1&0&-2\end{bmatrix}.
$$

我们求出它的简化阶梯型为

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

因此高斯消元法告诉我们其通解为$\boldsymbol{s}=t\begin{bmatrix}-2\\ 1\\ 1\end{bmatrix}$，即$\boldsymbol{x}_1^{\lambda_2}=\begin{bmatrix}-2\\ 1\\ 1\end{bmatrix}$是特征空间$V_{\lambda_2}$的一组基底。因此，矩阵

$$
P=\begin{bmatrix}\boldsymbol{x}_1^{\lambda_1}&\boldsymbol{x}_2^{\lambda_1}&\boldsymbol{x}_1^{\lambda_2}\end{bmatrix}=\begin{bmatrix}-1&0&-2\\ 0&1&1\\ 1&0&1\end{bmatrix}
$$

以及对角矩阵

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

满足$D=P^{-1}AP$。

# 7 第七章：内积，正交矩阵与二次型

## 7.1. 正交矩阵与正交对角化, Orthogonal Matrices and Orthogonal Diagonalization

本节对应**英文教材的7.1节与7.2节**。

首先我们给出正交矩阵的定义。

**Definition 7.1.** 我们称$A \in M_{n\times n}$为一个正交矩阵*(orthogonal matrix)*，如果它满足

$$A^\top A = AA^\top = I_n,$$

即$A^{-1} = A^\top$。

比如对于任何$\theta \in [0, 2\pi]$，矩阵

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

是一个正交矩阵，因为容易验证$A^\top A = I_2$(注意矩阵$A$实现平面上逆时针旋转$\theta$度的线性变换)。

**Definition 7.2.**

- 我们称$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为$\mathbb{R}^n$里的一个正交集合*(orthogonal set)*，如果对于任何$i \neq j, i, j \in \{1, \ldots, r\}$，都有

$$\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0,$$

即$S$里不同的$\boldsymbol{v}_i$与$\boldsymbol{v}_j$互相正交。

- 我们称$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为$\mathbb{R}^n$里的一个正交规范集合*(orthogonal normal set)*，如果$S$是一个正交集合且对任何$\boldsymbol{v}_i \in S, i = 1, \ldots, r$，都有$\|\boldsymbol{v}_i\| = 1$(这里$\|\boldsymbol{v}_i\|$指欧式范数)。
- 如果$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$是$\mathbb{R}^n$的一个正交集合且同时为$\mathbb{R}^n$的一组基底，那么我们称$S$是$\mathbb{R}^n$的一组正交基底*(orthogonal basis)*。
- 如果$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$是$\mathbb{R}^n$的一个正交规范集合且同时为$\mathbb{R}^n$的一组基底，那么我们称$S$是$\mathbb{R}^n$的一组标准/规范正交基底*(orthonormal basis)*。

实际上，**任何一个不包含零向量正交集合$S$都是线性无关集合**。因此如果一个$\mathbb{R}^n$里的正交集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$包含$n$个非零向量，那么它必然是$\mathbb{R}^n$的一组正交基底。如果一个$\mathbb{R}^n$里的正交规范集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$包含$n$个向量，那么它必然是$\mathbb{R}^n$的一组标准正交基底。

**Theorem 7.3.** 如果$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} \subset \mathbb{R}^n$是一个正交集合且$\boldsymbol{v}_i \neq \boldsymbol{0}$对所有$i = 1, \ldots, r$成立，那么$S$是一个线性无关集合。

*证明.* 考虑等式$k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r = \boldsymbol{0}$，对任何$i = 1, \ldots, r$，利用向量彼此之间的正交性质，即$\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$对任何$j \neq i$成立，可得

$$\begin{aligned} 0 = \boldsymbol{0} \cdot \boldsymbol{v}_i &= (k_1\boldsymbol{v}_1 + \ldots + k_r\boldsymbol{v}_r) \cdot \boldsymbol{v}_i \\ &= k_1(\boldsymbol{v}_1 \cdot \boldsymbol{v}_i) + \ldots + k_r(\boldsymbol{v}_r \cdot \boldsymbol{v}_i) \\ &= k_i(\boldsymbol{v}_i \cdot \boldsymbol{v}_i) = k_i\|\boldsymbol{v}_i\|^2, \end{aligned}$$

因此等式$k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r=\mathbf{0}$如果成立，那么对任何$i=1,\ldots,r$都有$k_i\|\boldsymbol{v}_i\|^2=0$。由于已经假设$\boldsymbol{v}_i\neq\mathbf{0}$，我们有$\|\boldsymbol{v}_i\|^2>0$，因此只能有$k_i=0$对任何$i=1,\ldots,r$成立。

□

**Theorem 7.4.** 对于$A\in M_{n\times n}$，以下说法等价：

- $A$为正交矩阵，即$A^\top A=AA^\top=I_n$；
- $A$的$n$个行向量构成$\mathbb{R}^n$的一个**标准正交基底**。
- $A$的$n$个列向量构成$\mathbb{R}^n$的一个**标准正交基底**。
- $A$是一个正交矩阵当且仅当对任何$\boldsymbol{x}\in\mathbb{R}^n$，有$\|A\boldsymbol{x}\|=\|\boldsymbol{x}\|$，这里$\|\boldsymbol{x}\|$是$\boldsymbol{x}$的欧式范数。
- $A$是一个正交矩阵当且仅当对任何$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$，有$A\boldsymbol{x}\cdot A\boldsymbol{y}=\boldsymbol{x}\cdot\boldsymbol{y}$。

证明. 将$A$的行向量表示记为

$$A=\begin{bmatrix}\boldsymbol{r}_1\\ \boldsymbol{r}_2\\ \vdots\\ \boldsymbol{r}_n\end{bmatrix},$$

每个$\boldsymbol{r}_i$为$1\times n$-行向量。由矩阵乘法定义，显然有

$$AA^\top=\begin{bmatrix}\boldsymbol{r}_1\cdot\boldsymbol{r}_1 & \boldsymbol{r}_1\cdot\boldsymbol{r}_2 & \ldots & \boldsymbol{r}_1\cdot\boldsymbol{r}_n\\ \vdots & \vdots & \vdots & \vdots\\ \boldsymbol{r}_n\cdot\boldsymbol{r}_1 & \boldsymbol{r}_n\cdot\boldsymbol{r}_2 & \ldots & \boldsymbol{r}_n\cdot\boldsymbol{r}_n\end{bmatrix}.$$

显然，$AA^\top=I_n$当且仅当

$$\boldsymbol{r}_i\cdot\boldsymbol{r}_j=\begin{cases}0, \text{如果} i\neq j\\ 1, \text{如果} i=j\end{cases}$$

并且后者实际上意味着$A$的$n$个行向量$S=\{\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n\}$构成一个$\mathbb{R}^n$里的正交规范集合，那么利用之前证明的Theorem 7.3，可知实际上此时$S=\{\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n\}$是$\mathbb{R}^n$的一组标准正交基底。类似可证$A$为正交矩阵当且仅当它的$n$个列向量$\boldsymbol{c}_1,\ldots,\boldsymbol{c}_n$组成$\mathbb{R}^n$的一组标准正交基底。

现在我们证明最后两个等价关系。我们证明的过程为先证明“$A$是一个正交矩阵”能够推出对任何$\boldsymbol{x}\in\mathbb{R}^n$，有$\|A\boldsymbol{x}\|=\|\boldsymbol{x}\|$，再证明$\|A\boldsymbol{x}\|=\|\boldsymbol{x}\|$对一切$\boldsymbol{x}\in\mathbb{R}^n$成立意味着对任何$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$，有$A\boldsymbol{x}\cdot A\boldsymbol{y}=\boldsymbol{x}\cdot\boldsymbol{y}$，最后证明如果$A\boldsymbol{x}\cdot A\boldsymbol{y}=\boldsymbol{x}\cdot\boldsymbol{y}$对任何$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$成立，那么$A$必然是一个正交矩阵。

如果$A$是正交矩阵，那么由于$A^\top A=I_n$，我们有

$$\|A\boldsymbol{x}\|^2=A\boldsymbol{x}\cdot A\boldsymbol{x}=(A\boldsymbol{x})^\top(A\boldsymbol{x})=\boldsymbol{x}^\top(A^\top A)\boldsymbol{x}=\boldsymbol{x}^\top I_n\boldsymbol{x}=\|\boldsymbol{x}\|^2,$$

从而得到$\|A\boldsymbol{x}\|=\|\boldsymbol{x}\|$。

现在假设对任何$\boldsymbol{x}\in\mathbb{R}^n$，有$\|A\boldsymbol{x}\|=\|\boldsymbol{x}\|$，那么利用平行四边形等式，见讲义Theorem 3.11，英文教材Theorem 3.2.7，我们有

$$\begin{aligned}A\boldsymbol{x}\cdot A\boldsymbol{y} &= \frac{1}{4}\|A\boldsymbol{x}+A\boldsymbol{y}\|^2-\frac{1}{4}\|A\boldsymbol{x}-A\boldsymbol{y}\|^2\\ &= \frac{1}{4}\underbrace{\|A(\boldsymbol{x}+\boldsymbol{y})\|^2}_{=\|\boldsymbol{x}-\boldsymbol{y}\|^2}-\frac{1}{4}\underbrace{\|A(\boldsymbol{x}-\boldsymbol{y})\|^2}_{=\|\boldsymbol{x}-\boldsymbol{y}\|^2}\\ &= \frac{1}{4}\|\boldsymbol{x}+\boldsymbol{y}\|^2-\frac{1}{4}\|\boldsymbol{x}-\boldsymbol{y}\|^2\\ &= \boldsymbol{x}\cdot\boldsymbol{y}.\end{aligned}$$

如果$A\boldsymbol{x}\cdot A\boldsymbol{y}=\boldsymbol{x}\cdot\boldsymbol{y}$对任何$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$成立，那么

$$A\boldsymbol{x}\cdot A\boldsymbol{y}=(A\boldsymbol{x})^\top(A\boldsymbol{y})=\boldsymbol{x}^\top(A^\top A)\boldsymbol{y}=\boldsymbol{x}\cdot\boldsymbol{y}=\boldsymbol{x}^\top I_n\boldsymbol{x},$$

也即$\boldsymbol{x}^\top(A^\top A)\boldsymbol{y}=\boldsymbol{x}^\top I_n\boldsymbol{x}$对任何$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$成立。我们用$(A^\top A)_{ij}$指代$A^\top A$第$i$行第$j$列上的项，$(I_n)_{ij}$指代$I_n$第$i$行第$j$列上的项，那么选取$\boldsymbol{x}=\boldsymbol{e}_i$，$\boldsymbol{y}=\boldsymbol{e}_j$为$\mathbb{R}^n$的标准单位向量，容易验证

$$\boldsymbol{e}_i^\top(A^\top A)\boldsymbol{e}_j=(A^\top A)_{ij},\quad \boldsymbol{e}_i^\top(I_n)\boldsymbol{e}_j=(I_n)_{ij}$$

因此等式$\boldsymbol{x}^\top(A^\top A)\boldsymbol{y}=\boldsymbol{x}^\top I_n\boldsymbol{x}$对任何$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$成立实际上给出了$(A^\top A)_{ij}=(I_n)_{ij}$对任何$i,j=1,\ldots,n$都成立，也即$A^\top A=I_n$成立。 □

利用正交矩阵的定义，我们容易证明

**Theorem 7.5.** *对于$A,B\in M_{n\times n}$：*

1. *如果$A$为正交矩阵，那么它可逆且$A^{-1}$也是一个正交矩阵。*
2. *如果$A,B$都是正交矩阵，那么$AB$也是正交矩阵。*
3. *如果$A$是一个正交矩阵，那么$\det(A)=1$或者$\det(A)=-1$。*

证明. 1. $A$是一个正交矩阵意味着$A^{-1}=A^\top$，因此利用$(A^{-1})^\top=(A^\top)^{-1}$我们得到

$$(A^{-1})^\top A^{-1}=(\underbrace{A^\top}_{=A^{-1}})^{-1}A^{-1}=(A^{-1})^{-1}A^{-1}=AA^{-1}=I_n,$$

即$A^{-1}$也是正交矩阵。

2. 如果$A,B$都是正交矩阵，那么$A^\top A=A^\top A=I_n, B^\top B=BB^\top=I_n$，因此

$$(AB)^\top AB=(B^\top A^\top)AB=B^\top(A^\top A)B=B^\top I_nB=B^\top B=I_n,$$

因此$AB$也为正交矩阵。

3. 如果$A$是一个正交矩阵，那么$A^{-1}=A^{\top}$，因此

$$1=\det(I_n)=\det(AA^{-1})=\det(A)\det(A^{-1})=\det(A)\det(A^{\top})=\det(A)^2,$$

因此$\det(A)=1$或$-1$。

□

下面的定理告诉我们在向量的坐标表示中选取**标准正交基底**的好处。

**Theorem 7.6.** 令$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n\}\subset\mathbb{R}^n$为$\mathbb{R}^n$的一组标准正交基底。假设$\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^n$，它们关于$S$的坐标向量为

$$(\boldsymbol{x})_S=(u_1,\ldots,u_n)\quad(\boldsymbol{x}=u_1\boldsymbol{v}_1+\ldots+u_n\boldsymbol{v}_n),$$

$$(\boldsymbol{y})_S=(v_1,\ldots,v_n)\quad(\boldsymbol{y}=v_1\boldsymbol{v}_1+\ldots+v_n\boldsymbol{v}_n).$$

那么我们有

- $\|\boldsymbol{x}\|^2=u_1^2+\ldots+u_n^2$。
- $d(\boldsymbol{x},\boldsymbol{y})^2=\|\boldsymbol{x}-\boldsymbol{y}\|^2=(u_1-v_1)^2+\ldots+(u_n-v_n)^2$。
- $\boldsymbol{x}\cdot\boldsymbol{y}=u_1v_1+\ldots+u_nv_n$。

证明.

- 由于$\|\boldsymbol{x}\|^2=\boldsymbol{x}\cdot\boldsymbol{x}=(u_1\boldsymbol{v}_1+\ldots+u_n\boldsymbol{v}_n)\cdot(u_1\boldsymbol{v}_1+\ldots+u_n\boldsymbol{v}_n)$，利用

$$\boldsymbol{v}_i\cdot\boldsymbol{v}_j=\begin{cases}0,\text{如果}i\neq j\\1,\text{如果}i=j\end{cases}$$

容易验证

$$(u_1\boldsymbol{v}_1+\ldots+u_n\boldsymbol{v}_n)\cdot(u_1\boldsymbol{v}_1+\ldots+u_n\boldsymbol{v}_n)=u_1^2+\ldots+u_n^2.$$

类似可证明$d(\boldsymbol{x},\boldsymbol{y})^2=\|\boldsymbol{x}-\boldsymbol{y}\|^2=(u_1-v_1)^2+\ldots+(u_n-v_n)^2$与$\boldsymbol{x}\cdot\boldsymbol{y}=u_1v_1+\ldots+u_nv_n$。

□

以下定理告诉我们，两个标准正交基底之间的转移矩阵(transition matrix)是正交矩阵。

**Theorem 7.7.** 设$S$与$S'$是$\mathbb{R}^n$的两个**标准正交基底**。那么$P_{S\leftarrow S'}$和$P_{S'\leftarrow S}$都是**正交矩阵**。

证明. 根据Theorem 7.4，要证明$P_{S\leftarrow S'}$是一个正交矩阵，只需证明对任何$\boldsymbol{x} \in \mathbb{R}^n$，都有$\|P_{S\leftarrow S'}\boldsymbol{x}\| = \|\boldsymbol{x}\|$。
注意，对任何$\boldsymbol{u} \in \mathbb{R}^n$，假设$[\boldsymbol{u}]_S = (u_1, \ldots, u_n)$，$[\boldsymbol{u}]_{S'} = (u'_1, \ldots, u'_n)$，那么因为$S$和$S'$都是标准正交基底，由之前证明的定理可知，此时有

$$\|\boldsymbol{u}\|^2 = \|[\boldsymbol{u}]_S\|^2 = u_1^2 + \ldots + u_n^2, \quad \|\boldsymbol{u}\|^2 = \|[\boldsymbol{u}]_{S'}\|^2 = (u'_1)^2 + \ldots + (u'_n)^2,$$

特别地，我们有$\|[\boldsymbol{u}]_S\|^2 = \|[\boldsymbol{u}]_{S'}\|^2 = \|\boldsymbol{u}\|^2$。
另一方面，由于$[\boldsymbol{u}]_S = P_{S\leftarrow S'}[\boldsymbol{u}]_{S'}$，我们还得到

$$\|[\boldsymbol{u}]_S\| = \|P_{S\leftarrow S'}[\boldsymbol{u}]_{S'}\|.$$

综上，我们得到

$$\|[\boldsymbol{u}]_S\| = \|[\boldsymbol{u}]_{S'}\| = \|P_{S\leftarrow S'}[\boldsymbol{u}]_{S'}\|,$$

对任何$\boldsymbol{u} \in \mathbb{R}^n$成立。特别地，对于任何$\boldsymbol{x} \in \mathbb{R}^n$，令$\boldsymbol{u} \in \mathbb{R}^n$满足$[\boldsymbol{u}]_{S'} = \boldsymbol{x}$，以上分析告诉我们

$$\|\boldsymbol{x}\| = \|[\boldsymbol{u}]_{S'}\| = \|P_{S\leftarrow S'}[\boldsymbol{u}]_{S'}\| = \|P_{S\leftarrow S'}\boldsymbol{x}\|,$$

因此得证。 □

现在我们引入一个很重要的概念：正交对角化(Orthogonal Diagonalization)。

**Definition 7.8.** *对于$A, B \in M_{n\times n}$，我们称$A$与$B$**正交相似(orthogonally similar)**，如果存在一个正交矩阵$P \in M_{n\times n}$使得*

$$P^\top AP = B.$$

*如果存在一个**对角矩阵**$D$使得$A$与$D$正交相似，那么我们称矩阵$A$可以**正交对角化(orthogonally diagonalizable)**。*

下面的定理告诉我们一个非常有趣的结果：一个矩阵可以正交对角化当且仅当它是一个对称矩阵(symmetric matrix)。回顾：一个方阵$A \in M_{n\times n}$是对称的，如果它的转置等于它自己：$A^\top = A$。

**Theorem 7.9.** *对于$A \in M_{n\times n}$，以下说法等价：*

1. *$A$可以正交对角化。*
2. *可以找到$A$的$n$个特征向量使其构成$\mathbb{R}^n$的一组标准正交基底。*
3. *$A^\top = A$，即$A$是一个对称矩阵。*

证明. “$A$可以正交对角化” $\Rightarrow$ “可以找到$A$的$n$个特征向量使其构成$\mathbb{R}^n$的一组标准正交基底”：如果$A$可以被正交对角化，那么存在一个正交矩阵$P$与一个对角矩阵$D$使得

$$D = P^\top AP.$$

从Theorem 6.13的证明过程中我们知道$P$的$n$个列向量都是$A$的特征向量，而由Theorem 7.4我们知道$P$作为正交矩阵，它的$n$个列向量为$\mathbb{R}^n$的一组标准正交基底。因此该方向得证。

“可以找到$A$的$n$个特征向量使其构成$\mathbb{R}^n$的一组标准正交基底” $\Rightarrow$ “$A$可以正交对角化”：令$\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$为$A$的特征向量且构成$\mathbb{R}^n$的一组标准正交基底。那么Theorem 6.13的证明可知以$\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$作为列向量的矩阵$P = \begin{bmatrix} \boldsymbol{c}_1 & \ldots_b c & \boldsymbol{c}_n \end{bmatrix}$满足

$$P^{-1}AP = D,$$

这里$D$为对角矩阵，对角线上的元素均为$A$的特征值。那么再次利用Theorem 7.4我们知道这样的$P$必然为正交矩阵，因此$P^{-1} = P^\top$，从而得到

$$P^\top AP = D,$$

即$A$可被正交对角化。因此该方向得证。

“$A$可以正交对角化” $\Rightarrow$ “$A^\top = A$”：由于$A$可以正交对角化，存在正交矩阵$P$和对角矩阵$D$使得

$$P^\top AP = D,$$

从而得到$A = PDP^\top$ (这里注意由于$P$是正交矩阵，我们有$P^\top P = PP^\top = I_n$)。因此可得

$$A^\top = (PDP^\top)^\top = (P^\top)^\top D^\top P^\top = PDP^\top,$$

因为对于对角矩阵$D$总有$D^\top = D$成立。

“$A^\top = A$” $\Rightarrow$ “$A$可以正交对角化”：该证明目前不要求掌握。 □

尽管我们暂时不提供“$A^\top = A$” $\Rightarrow$ “$A$可以正交对角化”的证明，我们还是可以证明以下事实：

**Theorem 7.10.** 令$A \in M_{n\times n}$为一个对称矩阵。那么如果$\lambda, \mu$是$A$的两个不同的特征值，那么对任何$\boldsymbol{u} \in Null(\lambda I_n - A)$，任何$\boldsymbol{v} \in Null(\mu I_n - A)$，必然有$\boldsymbol{u} \cdot \boldsymbol{v} = 0$，即**对称矩阵不同特征空间必然相互正交。**

证明. 令$\boldsymbol{u} \in \mathrm{Null}(\lambda I_n - A)$，令$\boldsymbol{v} \in \mathrm{Null}(\mu I_n - A)$，那么由于$A$是对称矩阵，即$A^\top = A$，我们有

$$A\boldsymbol{u} \cdot \boldsymbol{v} = (A\boldsymbol{u})^\top \boldsymbol{v} = \boldsymbol{u}^\top A^\top \boldsymbol{v} = \boldsymbol{u}^\top (A^\top \boldsymbol{v}) = \boldsymbol{u} \cdot A^\top \boldsymbol{v} \underbrace{=}_{A^\top = A} \boldsymbol{u} \cdot A\boldsymbol{v}.$$

由于$A\boldsymbol{u}=\lambda\boldsymbol{u}, A\boldsymbol{v}=\mu\boldsymbol{v}$，又有

$$A\boldsymbol{u}\cdot\boldsymbol{v}=\lambda\boldsymbol{u}\cdot\boldsymbol{v},\quad \boldsymbol{u}\cdot A\boldsymbol{v}=\mu\boldsymbol{u}\cdot\boldsymbol{v}.$$

综上，我们最终得到

$$A\boldsymbol{u}\cdot\boldsymbol{v}=\boldsymbol{u}\cdot A\boldsymbol{v}\Rightarrow\lambda\boldsymbol{u}\cdot\boldsymbol{v}=\mu\boldsymbol{u}\cdot\boldsymbol{v}\Rightarrow(\lambda-\mu)\boldsymbol{u}\cdot\boldsymbol{v}=0.$$

因为$\lambda\neq\mu$，以上等式成立必然要求$\boldsymbol{u}\cdot\boldsymbol{v}=0$。 □

给定一个对称矩阵$A\in M_{n\times n}$，我们现在考虑如何将其正交对角化。回忆我们在第6.2节中提到的将矩阵对角化的过程：

1. 计算$A$的特征多项式$p(\lambda)=\det(\lambda I_n-A)$，并求解特征方程$p(\lambda)=0$。如果该方程存在不属于实数的解，那么$A$不能被对角化，停止；否则进入下方第二步。
   对于对称矩阵$A$，我们已知它必然可以被对角化，因此要对角化对称矩阵$A$，我们必然会进入第二步。
2. 现在已知$p(\lambda)=0$的所有根均为实数，假设其为$\lambda_1,\ldots,\lambda_k$，且它们互不相等，那么此时我们知道$p(\lambda)$可被表示为

$$p(\lambda)=(\lambda-\lambda_1)^{c_1}(\lambda-\lambda_2)^{c_2}\ldots(\lambda-\lambda_k)^{c_k}.$$

   这里$c_j$对应特征值$\lambda_j$的**代数重数**，$j=1,\ldots,k$。将其记录下来。
   接下来依次计算$\lambda_1,\ldots,\lambda_k$的**几何重数**$d_1,\ldots,d_k$，即依次求出

$$\text{Null}(\lambda_1 I_n-A),\ldots,\text{Null}(\lambda_k I_n-A)$$

   的**维数**。如果出现了某个$j$使得$d_j<c_j$，那么$A$不可对角化，停止；否则进入下方第三步。
   对于对称矩阵$A$，我们已知它必然可以被对角化，因此它的所有特征值的代数重数与几何重数必然相等，即对于对称矩阵$A$，我们必然会进入第三步。
3. 通过第二步，现在的情况是所有特征值的代数重数与几何重数都相等，即$A$可以对角化，即存在一个对角矩阵$D$与一个可逆矩阵$P$使得$D=P^{-1}AP$。由Theorem 6.13的证明过程与Theorem 6.14可知，如果$A$关于$\lambda_j$的特征空间$V_{\lambda_j}$的一组基底为$\{\boldsymbol{x}_1^{\lambda_j},\ldots,\boldsymbol{x}_{d_j}^{\lambda_j}\}$，那么令$P$的列向量表示为：

$$P=\begin{bmatrix}\boldsymbol{x}_1^{\lambda_1} & \ldots & \boldsymbol{x}_{d_1}^{\lambda_1} & \boldsymbol{x}_1^{\lambda_2} & \ldots & \boldsymbol{x}_{d_2}^{\lambda_2} & \ldots & \boldsymbol{x}_1^{\lambda_k} & \ldots & \boldsymbol{x}_{d_j}^{\lambda_j}\end{bmatrix}$$

   那么我们有

$$D=P^{-1}AP,$$

此时$D=\mathrm{diag}(\underbrace{\lambda_1,\ldots,\lambda_1}_{d_1\text{个}},\underbrace{\lambda_2,\ldots,\lambda_2}_{d_2\text{个}},\ldots,\underbrace{\lambda_k,\ldots,\lambda_k}_{d_k\text{个}})$。因此，在这一步里我们要做的是**分别求出$A$关于$\lambda_j$的特征空间$V_{\lambda_j}$的一组基底$\{\boldsymbol{x}_1^{\lambda_j},\ldots,\boldsymbol{x}_{d_j}^{\lambda_j}\}$，$j=1,\ldots,k$，将其作为列向量拼成矩阵$P$，并写出对角矩阵$D$**。

在这一步里，我们找到了$A$的$n$个特征向量，方便起见，这里记为$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$。将它们作为列向量拼成可逆矩阵$P$。然而，在正交对角化中我们要求$P$必须是一个正交矩阵，因此我们必须将$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$转化为一组$\mathbb{R}^n$里的标准正交基底$\boldsymbol{c}_1,\ldots,\boldsymbol{c}_n$，且这些$\boldsymbol{c}_1,\ldots,\boldsymbol{c}_n$必须依然是$A$的特征向量。我们将在下一节讨论如何操作才能实现将$A$的特征向量"标准正交化"。

## 7.2. 内积空间与Gram–Schmidt程序, Inner Product Space and Gram–Schmidt Process

在本节我们主要学习如何将一组普通的基底转化为一组标准正交基底，即所谓的Gram–Schmidt程序。在介绍这个算法之前，我们首先讨论内积空间这一概念。

**Definition 7.11.** 令$V$为一个向量空间。令$\langle\cdot,\cdot\rangle:V\times V\to\mathbb{R}$为一个定义在$V\times V$上的函数，即对任何的$\boldsymbol{u}\in V,\boldsymbol{v}\in V$，$\langle\boldsymbol{u},\boldsymbol{v}\rangle$是一个对应的实数。如果$\langle\cdot,\cdot\rangle$满足以下条件：

1. $\langle\boldsymbol{u},\boldsymbol{v}\rangle=\langle\boldsymbol{v},\boldsymbol{u}\rangle$ (对称性，*symmetry*)；
2. $\langle\boldsymbol{u}+\boldsymbol{w},\boldsymbol{v}\rangle=\langle\boldsymbol{u},\boldsymbol{v}\rangle+\langle\boldsymbol{w},\boldsymbol{v}\rangle$，$\langle\boldsymbol{u},\boldsymbol{v}+\boldsymbol{w}\rangle=\langle\boldsymbol{u},\boldsymbol{v}\rangle+\langle\boldsymbol{u},\boldsymbol{w}\rangle$ (对两个变量都满足可加性，*additivity*)；
3. 对任何实数$k\in\mathbb{R}$，都有$\langle k\boldsymbol{u},\boldsymbol{v}\rangle=k\langle\boldsymbol{u},\boldsymbol{v}\rangle$，$\langle\boldsymbol{u},k\boldsymbol{v}\rangle=k\langle\boldsymbol{u},\boldsymbol{v}\rangle$ (对两个变量都满足齐次性，*homogeneity*)；
4. $\langle\boldsymbol{v},\boldsymbol{v}\rangle\geq 0$对任何$\boldsymbol{v}\in V$，且$\langle\boldsymbol{v},\boldsymbol{v}\rangle=0$当且仅当$\boldsymbol{v}=\mathbf{0}\in V$ (正定性，*positivity*),

那么我们称$\langle\cdot,\cdot\rangle$是$V$上的一个内积，称$V$是一个内积空间(*inner product space*)。
如果$V$是一个内积空间且$\langle\cdot,\cdot\rangle$是$V$上的一个内积，那么对于$\boldsymbol{v}\in V$，我们称$\|\boldsymbol{v}\|=\sqrt{\langle\boldsymbol{v},\boldsymbol{v}\rangle}$为向量$\boldsymbol{v}$的范数(*norm*)，称$d(\boldsymbol{u},\boldsymbol{v})=\|\boldsymbol{u}-\boldsymbol{v}\|=\sqrt{\langle\boldsymbol{u}-\boldsymbol{v},\boldsymbol{u}-\boldsymbol{v}\rangle}$为向量$\boldsymbol{u}$与$\boldsymbol{v}$之间的距离(*distance*)，对于非零向量$\boldsymbol{u},\boldsymbol{v}\in V$，称$\theta=\arccos(\frac{\langle\boldsymbol{u},\boldsymbol{v}\rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|})$为向量$\boldsymbol{u}$与$\boldsymbol{v}$之间的夹角(*angle*)，称满足$\langle\boldsymbol{u},\boldsymbol{v}\rangle=0$的向量$\boldsymbol{u}$与$\boldsymbol{v}$互相正交，称满足$\|\boldsymbol{v}\|=1$的向量$\boldsymbol{v}$为单位向量(*unit vector*)。

显然，**内积空间是对装备有欧式内积的欧式空间在一般向量空间上的推广**。特别地，任何对于欧式空间成立的结论(即所有第三章提到的关于欧式内积的结果)，都可以照搬到一般内积空间的情况。我们总结如下：

**Theorem 7.12.** 假设$V$是一个内积空间且$\langle\cdot,\cdot\rangle$是$V$上的一个内积。那么以下说法全部成立：

1. $\|\boldsymbol{v}\|=0$当且仅当$\boldsymbol{v}=\mathbf{0}$。
2. $\|k\boldsymbol{v}\|=|k|\|\boldsymbol{v}\|$对任何$k\in\mathbb{R}$成立。
3. $d(\boldsymbol{u},\boldsymbol{v})=d(\boldsymbol{v},\boldsymbol{u})$。
4. $d(\boldsymbol{u},\boldsymbol{v})=0 \iff \boldsymbol{u}=\boldsymbol{v}$。
5. $|\langle\boldsymbol{u},\boldsymbol{v}\rangle|\leq\|\boldsymbol{u}\|\|\boldsymbol{v}\|$ (*Cauchy–Schwarz* 不等式)，且等号成立当且仅当$\boldsymbol{u}$与$\boldsymbol{v}$线性相关。
6. $\|\boldsymbol{u}+\boldsymbol{v}\|\leq\|\boldsymbol{u}\|+\|\boldsymbol{v}\|$， $d(\boldsymbol{u},\boldsymbol{v})\leq d(\boldsymbol{u},\boldsymbol{w})+d(\boldsymbol{w},\boldsymbol{v})$ (*Triangle* 不等式)。
7. $\langle\boldsymbol{u},\boldsymbol{v}\rangle=0\Rightarrow\|\boldsymbol{u}+\boldsymbol{v}\|^2=\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2$ (勾股定理)。
8. 令$W\subset V$为$V$的一个子空间，我们称$W^\perp=\{\boldsymbol{u}\in V:\langle\boldsymbol{u},\boldsymbol{w}\rangle=0\}$为$W$的**正交补(orthogonal complement)**。那么有
   - $W^\perp$是一个子空间。
   - $W\cap W^\perp=\{\mathbf{0}\}$。
   - 若$W$为有限维空间，那么$(W^\perp)^\perp=W$。

**例子：**令$V=\mathbb{R}^n$为$n$维欧式空间，$w_1,\ldots,w_n$为**大于0的实数**，对$\boldsymbol{u}=(u_1,\ldots,u_n)\in\mathbb{R}^n,\boldsymbol{v}=(v_1,\ldots,v_n)\in\mathbb{R}^n$，我们定义

$$\langle\boldsymbol{u},\boldsymbol{v}\rangle_{w_1,\ldots,w_n}=w_1(u_1v_1)+\ldots+w_n(u_nv_n).$$

容易验证这样的$\langle\boldsymbol{u},\boldsymbol{v}\rangle$是$\mathbb{R}^n$上的一个内积。此时我们称实数$w_1,\ldots,w_n$为该内积的权重(weights)。显然，如果$w_1=w_2=\ldots=w_n=1$，那么$\langle\boldsymbol{u},\boldsymbol{v}\rangle_{1,\ldots,1}$即是欧式内积。比如在$\mathbb{R}^2$上我们令$w_1=\frac{1}{9}$，令$w_2=\frac{1}{4}$，那么$\langle\boldsymbol{u},\boldsymbol{v}\rangle_{w_1,w_2}=\frac{1}{9}u_1v_1+\frac{1}{4}u_2v_2$。注意，此时$\mathbb{R}^2$关于该内积$\langle\boldsymbol{u},\boldsymbol{v}\rangle_{w_1,w_2}$的单位圆$\{\boldsymbol{u}\in\mathbb{R}^2:\|\boldsymbol{u}\|_{w_1,w_2}=1\}$实际上是$\{\boldsymbol{u}=(x,y)\in\mathbb{R}^2:\frac{1}{9}x^2+\frac{1}{4}y^2=1\}$是$\mathbb{R}^2$关于欧式内积的椭圆。这实际上意味着**空间的内积决定了空间的几何**。

实际上以上带有权重的内积是所谓"由矩阵定义的内积"的一个特例：令$V=\mathbb{R}^n$，令$A\in M_{n\times n}$为一个**可逆矩阵**，定义

$$\langle\boldsymbol{u},\boldsymbol{v}\rangle_A=A\boldsymbol{u}\cdot A\boldsymbol{v}=(A\boldsymbol{u})^\top(A\boldsymbol{v})=\boldsymbol{u}^\top(A^\top A)\boldsymbol{v},$$

容易验证这样的$\langle\boldsymbol{u},\boldsymbol{v}\rangle_A$是$\mathbb{R}^n$上的一个内积。容易看出，如果我们选取$A=\begin{bmatrix}\sqrt{w_1}&&&\\&\sqrt{w_2}&&\\&&\ddots&\\&&&\sqrt{w_n}\end{bmatrix}$，$w_1>0,\ldots,w_n>0$，那么$\langle\boldsymbol{u},\boldsymbol{v}\rangle_A=\langle\boldsymbol{u},\boldsymbol{v}\rangle_{w_1,\ldots,w_n}$，并且$\langle\boldsymbol{u},\boldsymbol{v}\rangle_{I_n}=\boldsymbol{u}\cdot\boldsymbol{v}$为欧式内积。

**例子：** 令$V=M_{n\times n}$为$n$阶方阵组成的向量空间。对于$A,B\in M_{n\times n}$，定义

$$\langle A,B\rangle=\mathrm{tr}(A^\top B).$$

那么容易验证这样的$\langle A,B\rangle$是矩阵空间上的一个内积。

**例子：** 令$V=C([a,b])$为定义在区间$[a,b]$上的连续函数组成的向量空间。对于$f(x),g(x)\in V$，定义

$$\langle f(x),g(x)\rangle=\int_a^b f(x)g(x)dx.$$

那么这样的$\langle f(x),g(x)\rangle$是$V$上的一个内积。

**例子：取自2021-2022年线性代数期末考试题**

Let $V=P_2$. Define $\langle\cdot,\cdot\rangle$ on $V$ by

$$\langle p(x),q(x)\rangle=p(1)q(1)+p'(1)q'(1)+p''(1)q''(1)$$

for $p(x),q(x)\in V$. Prove that $\langle\cdot,\cdot\rangle$ is an inner product on $V$.

**证明：** 对称性，可加性与齐次性都容易验证。我们现在证明以上定义的$\langle\cdot,\cdot\rangle$也满足正定性。显然对任何$p(x)\in V$，都有

$$\langle p(x),p(x)\rangle=p(1)^2+(p'(1))^2+(p''(1))^2\geq 0.$$

假设$p(x)\in V$满足

$$\langle p(x),p(x)\rangle=p(1)^2+(p'(1))^2+(p''(1))^2=0$$

那么必然有$p(1)^2=(p'(1))^2=(p''(1))^2=0$。将$p(x)$记为$p(x)=a_0+a_1x+a_2x^2$，那么$p''(x)=2a_2$(对任何$x\in\mathbb{R}$)，因此我们有$(p''(1))^2=(2a_2)^2=4a_2^2=0$，从而得到$a_2=0$，也即$p(x)=a_0+a_1x$。此时有$p'(x)=a_1$(对任何$x\in\mathbb{R}$)，因此由$(p'(1))^2=(a_1)^2=a_1^2=0$可得$a_1=0$，因此$p(x)=a_0$。那么显然由$p(1)^2=a_0^2=0$可得$a_0=0$。综上，$p(x)=0+0x+0x^2$是$V$里的零向量。

与装配有欧式内积的$\mathbb{R}^n$空间上的情况完全类似，对于一般的内积空间$V$(装配内积$\langle\cdot,\cdot\rangle$)，我们也有正交集合，正交规范集合，标准/规范正交基底这些概念：

**Definition 7.13.** 令$V$为一个内积空间，其内积记为$\langle\cdot,\cdot\rangle$。

1. 我们称集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} \subset V$是一个**正交集合(orthogonal set)**，如果对任何$i \neq j$都有$\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0$，即该集合内任意两个不同的向量相互关于内积$\langle \cdot, \cdot \rangle$正交。
2. 若$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} \subset V$是一个**正交集合(orthogonal set)**同时满足$\|\boldsymbol{v}_i\| = 1$对所有$i = 1, \ldots, r$成立，那么称$S$为一个**正交规范集合(orthonormal set)**。
3. 如果$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset V$是一个**正交集合(orthogonal set)**且构成$V$的一组基底，那么称$S$为$V$的一组**正交基底(orthogonal basis)**。
4. 如果$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subset V$是一个**正交规范集合(orthonormal set)**且构成$V$的一组基底，那么称$S$为$V$的一组**标准/规范正交基底(orthonormal basis)**。

同样，模仿上一节针对欧式内积的证明，我们可以轻松证明以下结果对一般的内积依然成立：

**Theorem 7.14.** 令$V$为一个内积空间，其内积记为$\langle \cdot, \cdot \rangle$。那么我们有以下结果：

1. 对于任何正交集合$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} \subset V$，如果$\boldsymbol{v}_i \neq \mathbf{0}$对所有$i = 1, \ldots, r$成立，那么$S$一定是线性无关集合。
2. 对于任何$V$的标准正交基底$S$，任何$\boldsymbol{u} \in V, \boldsymbol{v} \in V$，将其关于$S$的坐标向量记为
$$(\boldsymbol{u})_S = (u_1, \ldots, u_n), \quad (\boldsymbol{v})_S = (v_1, \ldots, v_n),$$
那么总是有
$$\|\boldsymbol{u}\|^2 = \langle \boldsymbol{u}, \boldsymbol{u} \rangle = u_1^2 + \ldots + u_n^2,$$
$$d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\| = (\langle \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \rangle)^{\frac{1}{2}} = \sqrt{(u_1 - v_1)^2 + \ldots + (u_n - v_n)^2},$$
$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = u_1 v_1 + \ldots + u_n v_n.$$
3. 若$V$为有限维向量空间，$S$与$S'$为$V$的两组标准正交基底。那么$P_{S \leftarrow S'}$与$P_{S' \leftarrow S}$都是正交矩阵。

与欧式内积的情况完全类似，我们也可以**对一般的内积空间定义投影(projection)，这里我们只需把欧式内积$\boldsymbol{u} \cdot \boldsymbol{v}$替换为$\langle \boldsymbol{u}, \boldsymbol{v} \rangle$**。

**Theorem 7.15** (Projection Theorem for inner product space)**.** 令$V$为一个内积空间，其内积记为$\langle \cdot, \cdot \rangle$。令$W \subset V$为一个有限维子空间。那么任何$\boldsymbol{u} \in V$都可以唯一表示为
$$\boldsymbol{u} = \boldsymbol{w}_1 + \boldsymbol{w}_2,$$
其中$\boldsymbol{w}_1 \in W, \boldsymbol{w}_2 \in W^{\perp}$。我们将$\boldsymbol{w}_1$记为$\boldsymbol{w}_1 = proj_W \boldsymbol{u}$，将$\boldsymbol{w}_2$记为$\boldsymbol{w}_2 = proj_{W^{\perp}}(\boldsymbol{u})$。

如果$S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$为$W$的一组**正交基底**，那么对任何$\boldsymbol{u} \in V$，都有
$$proj_W \boldsymbol{u} = \frac{\langle \boldsymbol{u}, \boldsymbol{v}_1 \rangle}{\|\boldsymbol{v}_1\|^2} \boldsymbol{v}_1 + \ldots + \frac{\langle \boldsymbol{u}, \boldsymbol{v}_r \rangle}{\|\boldsymbol{v}_r\|^2} \boldsymbol{v}_r.$$

如果$S=\{\boldsymbol{v}_1,...,\boldsymbol{v}_r\}$为$W$的一组**标准正交基底**，那么对任何$\boldsymbol{u}\in V$，都有

$$proj_W\boldsymbol{u}=\langle\boldsymbol{u},\boldsymbol{v}_1\rangle\boldsymbol{v}_1+...+\langle\boldsymbol{u},\boldsymbol{v}_r\rangle\boldsymbol{v}_r.$$

特别地，如果$S=\{\boldsymbol{v}_1,...,\boldsymbol{v}_n\}$是$W=V$的一组标准正交基底，那么任何$\boldsymbol{v}\in V$都可以表示为

$$\boldsymbol{v}=\langle\boldsymbol{v},\boldsymbol{v}_1\rangle\boldsymbol{v}_1+...+\langle\boldsymbol{v},\boldsymbol{v}_n\rangle\boldsymbol{v}_n.$$

证明. 我们先证明以下分解

$$\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2,$$

(其中$\boldsymbol{w}_1\in W,\boldsymbol{w}_2\in W^{\perp}$)的唯一性。假设$\boldsymbol{u}=\boldsymbol{v}_1+\boldsymbol{v}_2$，且$\boldsymbol{v}_1\in W,\boldsymbol{v}_2\in W^{\perp}$。那么我们必然有

$$\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2=\boldsymbol{v}_1+\boldsymbol{v}_2\Rightarrow\boldsymbol{v}_1-\boldsymbol{w}_1=\boldsymbol{w}_2-\boldsymbol{v}_2.$$

由于$\boldsymbol{v}_1-\boldsymbol{w}_1\in W$，$\boldsymbol{w}_2-\boldsymbol{v}_2\in W^{\perp}$(因为$W^{\perp}$为一个子空间)，以上等式告诉我们$\boldsymbol{v}_1-\boldsymbol{w}_1\in W\cap W^{\perp},\boldsymbol{w}_2-\boldsymbol{v}_2\in W\cap W^{\perp}$。由Theorem 7.12我们知道$W\cap W^{\perp}=\{\boldsymbol{0}\}$，因此可得

$$\boldsymbol{v}_1-\boldsymbol{w}_1=\boldsymbol{w}_2-\boldsymbol{v}_2=\boldsymbol{0}.$$

现在我们证明分解

$$\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2,$$

(其中$\boldsymbol{w}_1\in W,\boldsymbol{w}_2\in W^{\perp}$)的存在性。假设$S=\{\boldsymbol{v}_1,...,\boldsymbol{v}_r\}$是$W$的一组**正交基底**。注意，这里我们实际上还要证明，对任何有限维子空间$W\subset V$都存在一组$W$的正交基底。这一部分内容我们将在该定理后面学习。现在我们先假设$W\subset V$总是拥有一组正交基底。我们直接**定义**

$$\boldsymbol{w}_1=\frac{\langle\boldsymbol{u},\boldsymbol{v}_1\rangle}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1+...+\frac{\langle\boldsymbol{u},\boldsymbol{v}_r\rangle}{\|\boldsymbol{v}_r\|^2}\boldsymbol{v}_r,$$

定义$\boldsymbol{w}_2=\boldsymbol{u}-\boldsymbol{w}_1$。显然，对任何$\boldsymbol{v}_i\in S$，都有

$$\begin{aligned}\langle\boldsymbol{u}-\boldsymbol{w}_1,\boldsymbol{v}_i\rangle&=\langle\boldsymbol{u},\boldsymbol{v}_i\rangle-\langle\boldsymbol{w}_1,\boldsymbol{v}_i\rangle\\&\langle\boldsymbol{u},\boldsymbol{v}_i\rangle-\langle\sum_{j=1}^{r}\frac{\langle\boldsymbol{u},\boldsymbol{v}_j\rangle}{\|\boldsymbol{v}_j\|^2}\boldsymbol{v}_j,\boldsymbol{v}_i\rangle\\&\underbrace{=}_{j\neq i\Rightarrow\langle\boldsymbol{v}_j,\boldsymbol{v}_i\rangle=0}\langle\boldsymbol{u},\boldsymbol{v}_i\rangle-\langle\frac{\langle\boldsymbol{u},\boldsymbol{v}_i\rangle}{\|\boldsymbol{v}_i\|^2}\boldsymbol{v}_i,\boldsymbol{v}_i\rangle\\&=\langle\boldsymbol{u},\boldsymbol{v}_i\rangle-\frac{\langle\boldsymbol{u},\boldsymbol{v}_i\rangle}{\|\boldsymbol{v}_i\|^2}\langle\boldsymbol{v}_i,\boldsymbol{v}_i\rangle\\&=\langle\boldsymbol{u},\boldsymbol{v}_i\rangle-\frac{\langle\boldsymbol{u},\boldsymbol{v}_i\rangle}{\|\boldsymbol{v}_i\|^2}\|\boldsymbol{v}_i\|^2\\&=\langle\boldsymbol{u},\boldsymbol{v}_i\rangle-\langle\boldsymbol{u},\boldsymbol{v}_i\rangle\\&=0.\end{aligned}$$

因此我们实际上证明了对于以上定义的$\boldsymbol{w}_1$，对任何$i=1,\ldots,r$，都有$\boldsymbol{u}-\boldsymbol{w}_1$与$\boldsymbol{v}_i$正交，那么因为$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$是$W$的一组基底，我们容易利用这一事实推出$\boldsymbol{v}-\boldsymbol{w}_1$与$W$里的任何向量正交：

$$\langle \boldsymbol{u}-\boldsymbol{w}_1,\boldsymbol{w}\rangle=0$$

对任何$\boldsymbol{w}=k_1\boldsymbol{v}_1+\ldots+k_r\boldsymbol{v}_r$。因此由定义，必然有$\boldsymbol{w}_2=\boldsymbol{u}-\boldsymbol{w}_1\in W^{\perp}$。因此我们实际上证明了分解

$$\boldsymbol{u}=\boldsymbol{w}_1+\boldsymbol{w}_2,$$

(其中$\boldsymbol{w}_1\in W,\boldsymbol{w}_2\in W^{\perp}$)的存在性，并且由该分解的唯一性，可知

$$\boldsymbol{w}_1=\mathrm{proj}_W\boldsymbol{u}=\frac{\langle \boldsymbol{u},\boldsymbol{v}_1\rangle}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1+\ldots+\frac{\langle \boldsymbol{u},\boldsymbol{v}_r\rangle}{\|\boldsymbol{v}_r\|^2}\boldsymbol{v}_r.$$

如果$S=\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$为$W$的一组**标准正交基底**，那么由于$\|\boldsymbol{v}_i\|=1$对任何$i=1,\ldots,r$成立，我们可以化简$\mathrm{proj}_W\boldsymbol{u}$的表达式为

$$\mathrm{proj}_W\boldsymbol{u}=\langle \boldsymbol{u},\boldsymbol{v}_1\rangle\boldsymbol{v}_1+\ldots+\langle \boldsymbol{u},\boldsymbol{v}_r\rangle\boldsymbol{v}_r.$$

□

现在我们介绍一个将有限维子空间$W$的任意一组基底$\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_r\}$转化为一组标准正交基底$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$的算法，即著名的**Gram–Schmidt Process**。这一算法间接证明了任何内积空间$V$的有限维子空间$W$总是拥有一组标准正交基底。
该算法总结如下：

**Gram–Schmidt算法：**

1. 令$\boldsymbol{v}_1'=\boldsymbol{u}_1$。如果$r=1$，令$\boldsymbol{v}_1=\frac{1}{\|\boldsymbol{v}_1'\|}\boldsymbol{v}_1'$。此时该算法结束。
2. 若$r=2$，令$\boldsymbol{v}_2'=\boldsymbol{u}_2-\langle \boldsymbol{u}_2,\boldsymbol{v}_1\rangle\boldsymbol{v}_1$，定义$\boldsymbol{v}_2=\frac{1}{\|\boldsymbol{v}_2'\|}\boldsymbol{v}_2'$。此时(即如果$r=2$)该算法结束。
3. 若$r=3$，令$\boldsymbol{v}_3'=\boldsymbol{u}_3-\langle \boldsymbol{u}_3,\boldsymbol{v}_1\rangle\boldsymbol{v}_1-\langle \boldsymbol{u}_3,\boldsymbol{v}_2\rangle\boldsymbol{v}_2$，定义$\boldsymbol{v}_3=\frac{1}{\|\boldsymbol{v}_3'\|}\boldsymbol{v}_3'$。此时(即如果$r=3$)该算法结束。
4. 对于一般的$r$，我们重复以上步骤。即，如果我们已经利用以上步骤得到$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{r-1}$，那么定义

$$\boldsymbol{v}_r'=\boldsymbol{u}_r-\langle \boldsymbol{u}_r,\boldsymbol{v}_1\rangle\boldsymbol{v}_1-\langle \boldsymbol{u}_r,\boldsymbol{v}_2\rangle\boldsymbol{v}_2-\ldots-\langle \boldsymbol{u}_r,\boldsymbol{v}_{r-1}\rangle\boldsymbol{v}_{r-1},$$

并令$\boldsymbol{v}_r=\frac{1}{\|\boldsymbol{v}_r'\|}\boldsymbol{v}_r'$。那么我们此时可以得到$W$的一组标准正交基底$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$。算法结束。

**例子：** 令$V=P_2$，定义$V$上的一个内积为

$$\langle p(x),q(x)\rangle=\int_{-1}^{1}p(x)q(x)dx.$$

令$B=\{u_1(x)=1,u_2(x)=x,u_3(x)=x^2\}$为$V$的标准基底。我们现在利用Gram–Schmidt算法，通过$B$得到$V$的一组标准正交基底。

1. 令$\boldsymbol{v}_1'(x)=u_1(x)=1$，由于$\|\boldsymbol{v}_1'(x)\|=\int_{-1}^{1}1dx=2$，可得$\boldsymbol{v}_1(x)=\frac{1}{\sqrt{2}}\boldsymbol{v}_1'(x)=\frac{1}{\sqrt{2}}$。
2. 令$\boldsymbol{v}_2'(x)=\boldsymbol{u}_2(x)-\langle\boldsymbol{u}_2(x),\boldsymbol{v}_1(x)\rangle\boldsymbol{v}_1(x)=\boldsymbol{u}_2(x)-\int_{-1}^{1}\frac{1}{\sqrt{2}}xdx\boldsymbol{v}_1(x)$。显然由于$\int_{-1}^{1}\frac{1}{\sqrt{2}}xdx=\frac{1}{2}\int_{-1}^{1}xdx=0$，我们有$\boldsymbol{v}_2'(x)=\boldsymbol{u}_2(x)=x$，因此利用$\|\boldsymbol{v}'(x)\|=\|\boldsymbol{u}_2(x)\|=(\int_{-1}^{1}x^2dx)^{1/2}=\sqrt{\frac{2}{3}}$，得到
$$\boldsymbol{v}_2(x)=\frac{1}{\|\boldsymbol{v}_2'(x)\|}\boldsymbol{v}_2'(x)=\sqrt{\frac{3}{2}}x.$$
3. 令$\boldsymbol{v}_3'(x)=\boldsymbol{u}_3(x)-\langle\boldsymbol{u}_3(x),\boldsymbol{v}_1(x)\rangle\boldsymbol{v}_1(x)-\langle\boldsymbol{u}_3(x),\boldsymbol{v}_2(x)\rangle\boldsymbol{v}_2(x)$。经计算可得
$$\langle\boldsymbol{u}_3(x),\boldsymbol{v}_1(x)\rangle=\int_{-1}^{1}\frac{1}{\sqrt{2}}x^2dx=\frac{1}{\sqrt{2}}\int_{-1}^{1}x^2dx=\frac{1}{\sqrt{2}}\times\frac{2}{3}=\frac{\sqrt{2}}{3},$$
$$\langle\boldsymbol{u}_3(x),\boldsymbol{v}_2(x)\rangle=\int_{-1}^{1}\sqrt{\frac{3}{2}}x^3dx=\sqrt{\frac{3}{2}}\int_{-1}^{1}x^3dx=\sqrt{\frac{3}{2}}\times 0=0,$$
因此$\boldsymbol{v}_3'(x)=x^2-\frac{\sqrt{2}}{3}\frac{1}{\sqrt{2}}=x^2-\frac{1}{3}$。从而有$\|\boldsymbol{v}_3'(x)\|^2=\int_{-1}^{1}(x^2-\frac{1}{3})^2dx=\frac{8}{45}$，故
$$\boldsymbol{v}_3(x)=\frac{1}{\|\boldsymbol{v}_3'(x)\|}\boldsymbol{v}_3'(x)=\frac{x^2-\frac{1}{3}}{\frac{8}{45}}.$$

如果我们只需要求出一组**正交基底**，那么对Gram–Schmidt算法进行一些微调即可：

**Gram–Schmidt算法的另一个版本：** 令$V$为一个内积空间，我们的目标是将有限维子空间$W$的任意一组基底$\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_r\}$转化为一组正交基底$\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r\}$。

1. 令$\boldsymbol{v}_1=\boldsymbol{u}_1$。如果$r=1$，此时该算法结束。
2. 若$r=2$，令$\boldsymbol{v}_2=\boldsymbol{u}_2-\frac{\langle\boldsymbol{u}_2,\boldsymbol{v}_1\rangle}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1$。此时(即如果$r=2$)该算法结束。
3. 若$r=3$，令$\boldsymbol{v}_3=\boldsymbol{u}_3-\frac{\langle\boldsymbol{u}_3,\boldsymbol{v}_1\rangle}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1-\frac{\langle\boldsymbol{u}_3,\boldsymbol{v}_2\rangle}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2$。此时(即如果$r=3$)该算法结束。
4. 对于一般的$r$，我们重复以上步骤。即，如果我们已经利用以上步骤得到$\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{r-1}$，那么定义
$$\boldsymbol{v}_r=\boldsymbol{u}_r-\frac{\langle\boldsymbol{u}_r,\boldsymbol{v}_1\rangle}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1-\frac{\langle\boldsymbol{u}_r,\boldsymbol{v}_2\rangle}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2-\ldots-\frac{\langle\boldsymbol{u}_r,\boldsymbol{v}_{r-1}\rangle}{\|\boldsymbol{v}_{r-1}\|^2}\boldsymbol{v}_{r-1}.$$

那么我们此时可以得到$W$的一组正交基底$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$。算法结束。

5. 如果我们最终还是想要得到一组标准正交基底，那么只需要将以上正交基底**规范化(normalizing)**即可，即：$\{\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}, \ldots, \frac{\boldsymbol{v}_r}{\|\boldsymbol{v}_r\|}\}$是所求标准正交基底。

现在我们回到**对称矩阵的正交对角化算法**。我们先回顾之前已经得到的几个步骤：

1. 计算$A$的特征多项式$p(\lambda) = \det(\lambda I_n - A)$，并求解特征方程$p(\lambda) = 0$。如果该方程存在不属于实数的解，那么$A$不能被对角化，停止；否则进入下方第二步。
   对于对称矩阵$A$，我们已知它必然可以被对角化，因此要对角化对称矩阵$A$，我们必然会进入第二步。
2. 现在已知$p(\lambda) = 0$的所有根均为实数，假设其为$\lambda_1, \ldots, \lambda_k$，且它们互不相等，那么此时我们知道$p(\lambda)$可被表示为
$$p(\lambda) = (\lambda - \lambda_1)^{c_1}(\lambda - \lambda_2)^{c_2} \ldots (\lambda - \lambda_k)^{c_k}.$$
   这里$c_j$对应特征值$\lambda_j$的**代数重数**，$j = 1, \ldots, k$。将其记录下来。
   接下来依次计算$\lambda_1, \ldots, \lambda_k$的**几何重数**$d_1, \ldots, d_k$，即依次求出
$$\mathrm{Null}(\lambda_1 I_n - A), \ldots, \mathrm{Null}(\lambda_k I_n - A)$$
   的**维数**。如果出现了某个$j$使得$d_j < c_j$，那么$A$不可对角化，停止；否则进入下方第三步。
   对于对称矩阵$A$，我们已知它必然可以被对角化，因此它的所有特征值的代数重数与几何重数必然相等，即对于对称矩阵$A$，我们必然会进入第三步。
3. 通过第二步，现在的情况是所有特征值的代数重数与几何重数都相等，即$A$可以对角化。对每个$j = 1, \ldots, k$，通过求解齐次线性方程组$(\lambda_j I_n - A)\boldsymbol{x} = \boldsymbol{0}$找到$A$关于$\lambda_j$的特征空间$V_{\lambda_j}$的一组基底，并记为$\{\boldsymbol{x}_1^{\lambda_j}, \ldots, \boldsymbol{x}_{d_j}^{\lambda_j}\}$。
4. 对每个$j = 1, \ldots, k$，利用Gram–Schmidt算法作用于$\mathbb{R}^n$上的欧式内积，我们将$V_{\lambda_j}$的基底$\{\boldsymbol{x}_1^{\lambda_j}, \ldots, \boldsymbol{x}_{d_j}^{\lambda_j}\}$转化为$V_{\lambda_j}$的一组标准正交基底$\{\boldsymbol{v}_1^{\lambda_j}, \ldots, \boldsymbol{v}_{d_j}^{\lambda_j}\}$。现在令$P$的列向量表示为：
$$P = \begin{bmatrix} \boldsymbol{v}_1^{\lambda_1} & \ldots & \boldsymbol{v}_{d_1}^{\lambda_1} & \boldsymbol{v}_1^{\lambda_2} & \ldots & \boldsymbol{v}_{d_2}^{\lambda_2} & \ldots & \boldsymbol{v}_1^{\lambda_k} & \ldots & \boldsymbol{v}_{d_j}^{\lambda_j} \end{bmatrix}$$
   那么我们有
$$D = P^{-1}AP,$$
   此时$D = \mathrm{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{d_1\text{个}}, \underbrace{\lambda_2, \ldots, \lambda_2}_{d_2\text{个}}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{d_k\text{个}})$。显然，由于$P$的列向量现在组成$\mathbb{R}^n$的一组标准正交基底(请思考为什么)，矩阵$P$是一个正交矩阵，

因此$P^{-1}=P^{\top}$，因此实际上我们得到了

$$D=P^{\top}AP.$$

计算结束。

**例子：** 令

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

我们按照以上步骤，将$A$正交对角化。

1. 计算可得$p_A(\lambda)=\det(\lambda I_3-A)=\det(\begin{bmatrix}\lambda-4&-2&-2\\-2&\lambda-4&-2\\-2&-2&\lambda-4\end{bmatrix})=(\lambda-2)^2(\lambda-8)$。显然$A$的两个特征值为$\lambda_1=2,\lambda_2=8$。
2. 求解齐次线性方程组$(\lambda_1I_3-A)\boldsymbol{x}=\mathbf{0}$，可得$\text{Null}(\lambda_1I_3-A)$的一组基底为

$$\boldsymbol{x}_1^{\lambda_1}=\begin{bmatrix}-1\\1\\0\end{bmatrix},\quad \boldsymbol{x}_2^{\lambda_1}=\begin{bmatrix}-1\\0\\1\end{bmatrix}.$$

求解齐次线性方程组$(\lambda_2I_3-A)\boldsymbol{x}=\mathbf{0}$，可得$\text{Null}(\lambda_2I_3-A)$的一组基底为

$$\boldsymbol{x}_1^{\lambda_2}=\begin{bmatrix}1\\1\\1\end{bmatrix}.$$

3. 现在我们用Gram–Schmidt算法将$\boldsymbol{x}_1^{\lambda_1},\boldsymbol{x}_2^{\lambda_1}$正交规范化：

$$\boldsymbol{v}_1^{\lambda_1}=\frac{1}{\|\boldsymbol{x}_1^{\lambda_1}\|}\boldsymbol{x}_1^{\lambda_1}=\begin{bmatrix}-1/\sqrt{2}\\1/\sqrt{2}\\0\end{bmatrix},$$

$$\boldsymbol{v}_2^{\lambda_1}=\frac{\boldsymbol{x}_2^{\lambda_1}-(\boldsymbol{x}_2^{\lambda_1}\cdot\boldsymbol{v}_1^{\lambda_1})\boldsymbol{v}_1^{\lambda_1}}{\|\boldsymbol{x}_2^{\lambda_1}-\boldsymbol{x}_2^{\lambda_1}\cdot\boldsymbol{v}_1^{\lambda_1}\boldsymbol{v}_1^{\lambda_1}\|}=\begin{bmatrix}-1/\sqrt{6}\\-1/\sqrt{6}\\2/\sqrt{6}\end{bmatrix}.$$

我们现在将$\boldsymbol{x}_1^{\lambda_2}$正交规范化。显然此时只需除以它的范数即可：

$$\boldsymbol{v}_1^{\lambda_2}=\frac{1}{\|\boldsymbol{x}_1^{\lambda_2}\|}\boldsymbol{x}_1^{\lambda_2}=\begin{bmatrix}1/\sqrt{3}\\1/\sqrt{3}\\1/\sqrt{3}\end{bmatrix}.$$

$$因此P = \begin{bmatrix} \boldsymbol{v}_1^{\lambda_1} & \boldsymbol{v}_2^{\lambda_1} & \boldsymbol{v}_1^{\lambda_2} \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}，D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix}。$$

## 7.3. 二次型, Quadratic Forms

本节对应英文教材的7.3节。

**Definition 7.16.** 令$A \in M_{n\times n}$为对称矩阵，即$A^\top = A$。我们称函数

$$Q_A : \mathbb{R}^n \to \mathbb{R}, Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x}, \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

为$A$的二次型*(quadratic form associated with $A$)*。

我们首先注意二次型具有以下基本性质：

- $Q_A : \mathbb{R}^n \to \mathbb{R}$不是线性变换。
- 将$A$记为$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$，那么利用$A$的对称性，即$a_{ij} = a_{ji}$对任何$i, j = 1, \dots, n$成立，可得

$$\begin{aligned} Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x)ix_j \\ &= (a_{11}x_1^2 + \dots + a_{nn}x_n^2) + \sum_{i\neq j} a_{ij}x_ix_j \\ &= (a_{11}x_1^2 + \dots + a_{nn}x_n^2) + 2\sum_{i<j} a_{ij}x_ix_j. \end{aligned}$$

  我们称$2\sum_{i<j} a_{ij}x_ix_j$为二次型$Q_A$的交叉项(cross product term)。
- 当$A$为对角矩阵时，即$A = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$，那么显然有

$$Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2.$$

  即对于对角矩阵$A$，它的二次型不含有交叉项。

**例子：** 找出下面两个二次型对应的对称矩阵$A$。

1. $2x^2+6xy-5y^2$

   假设$A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$且$a_{12}=a_{21}$。那么

   $$Q_A(\boldsymbol{x})=a_{11}x^2+a_{22}y^2+\underbrace{a_{12}xy+a_{21}xy}_{=2a_{12}xy},\boldsymbol{x}=\begin{bmatrix} x \\ y \end{bmatrix}.$$

   因此如果$Q_A(\boldsymbol{x})=2x^2+6xy-5y^2$，那么比较$x^2,y^2,xy$的系数必然有

   $$a_{11}=2,a_{22}=-5,a_{12}=a_{21}=3$$

   即$A=\begin{bmatrix} 2 & 3 \\ 3 & -5 \end{bmatrix}$。

2. $x_1^2+7x_2^2-3x_3^2+4x_1x_2-2x_1x_3+8x_2x_3$。

   假设$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$，且$a_{21}=a_{12},a_{13}=a_{31},a_{23}=a_{32}$。那么可以计算

   得到($\boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$)

   $$Q_A(\boldsymbol{x})=a_{11}x_1^2+a_{22}x_2^2+a_{33}x_3^2+2a_{12}x_1x_2+2a_{13}x_1x_3+2a_{23}x_2x_3.$$

   因此如果$Q_A(\boldsymbol{x})=x_1^2+7x_2^2-3x_3^2+4x_1x_2-2x_1x_3+8x_2x_3$，那么比较$x_1^2,x_2^2,x_3^2,x_1x_2,x_1x_3,x_2x_3$的系数，可得

   $$a_{11}=1,a_{22}=7,a_{33}=-3,a_{12}=a_{21}=2,a_{13}=a_{31}=-1,a_{23}=a_{32}=4.$$

   因此$A=\begin{bmatrix} 1 & 2 & -1 \\ 2 & 7 & 4 \\ -1 & 4 & -3 \end{bmatrix}$。

显然，如果二次型$Q_A$不包含交叉项，那么它的表达式会化简为

$$Q_A(\boldsymbol{x})=\lambda_1x_1^2+\ldots+\lambda_nx_n^2.$$

而对于非对角的对称矩阵$A$，它的二次型$Q_A$并不能写成这样的形式。为了将其中的交叉项消去，我们需要进行**换元(change of variable)**的操作。具体细节如下：

由于$A$是对称矩阵，由Theorem 7.9可知$A$能被正交对角化，即存在正交矩阵$P$与对角矩阵$D$使得

$$D=P^\top AP.$$

对于$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$，我们令$\boldsymbol{y} = P^{-1}\boldsymbol{x} = P^{\top}\boldsymbol{x}$，即$\boldsymbol{x} = P\boldsymbol{y}$，那么对于$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$，我们有(设$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$):

$$Q_A(\boldsymbol{x}) = \boldsymbol{x}^{\top}A\boldsymbol{x} = (P\boldsymbol{y})^{\top}A(P\boldsymbol{y}) = \boldsymbol{y}^{\top}(P^{\top}AP)\boldsymbol{y} = \boldsymbol{y}^{\top}D\boldsymbol{y} = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2.$$

我们将以上观察总结为以下定理。

**Theorem 7.17** (The principal axes theorem)**.** *令$A \in M_{n\times n}$为对称矩阵，那么$A$可被正交对角化为*

$$D = P^{\top}AP, D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

*令$\boldsymbol{y} = P^{\top}\boldsymbol{x}$，我们称其为**orthogonal change of variables**。那么我们有*

$$Q_A(\boldsymbol{x}) = Q_A(P\boldsymbol{y}) = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2.$$

**例子：** 令$Q(\boldsymbol{x}) = x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$。找出一个orthogonal change of variables $\boldsymbol{y} = P^{\top}\boldsymbol{x} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$，并计算$Q(P\boldsymbol{y})$。

**解：** 我们首先找到一个对称矩阵$A$使得$Q(\boldsymbol{x}) = Q_A(\boldsymbol{x}) = \boldsymbol{x}^{\top}A\boldsymbol{x}$。由于$\boldsymbol{x}^{\top}A\boldsymbol{x} = (a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2) + 2a_{12}x_1x_2 + 2a_{23}x_2x_3 + 2a_{13}x_1x_3$，如果

$$\boldsymbol{x}^{\top}A\boldsymbol{x} = (a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2) + 2a_{12}x_1x_2 + 2a_{23}x_2x_3 + 2a_{13}x_1x_3 = x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$$

成立，那么比较$x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3$的系数可得

$$a_{11} = 1, a_{22} = 0, a_{33} = -1, a_{12} = a_{21} = -2, a_{13} = a_{31} = 0, a_{23} = a_{32} = 2,$$

因此对称矩阵$A$的表达式为

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

要找到orthogonal change of variables中出现的正交矩阵$P$，我们需要将$A$正交对角化。那么按照**192页上介绍的正交对角化算法**，我们依次进行以下操作：

1. $A$的特征多项式为

$$p_A(\lambda) = \det(\lambda I_3 - A) = \begin{vmatrix} \lambda - 1 & 2 & 0 \\ 2 & \lambda & -2 \\ 0 & -2 & \lambda + 1 \end{vmatrix} = \lambda^3 - 9\lambda = \lambda(\lambda - 3)(\lambda + 3).$$

   解特征方程$p_A(\lambda) = 0$，可得$A$的特征值为

$$\lambda_1 = 0, \lambda_2 = -3, \lambda_3 = 3.$$

2. 现在求出特征空间$V_{\lambda_1} = \text{Null}(\lambda_1 I_3 - A)$，$V_{\lambda_2} = \text{Null}(\lambda_2 I_3 - A)$与$V_{\lambda_3} = \text{Null}(\lambda_3 I_3 - A)$的一组基底。那么用高斯消元法依次求解方程组$(\lambda_1 I_3 - A)\boldsymbol{x} = \boldsymbol{0}$，$(\lambda_2 I_3 - A)\boldsymbol{x} = \boldsymbol{0}$与$(\lambda_3 I_3 - A)\boldsymbol{x} = \boldsymbol{0}$，可求得各自特征空间的基底为

$$\boldsymbol{x}_1^{\lambda_1} = \begin{bmatrix} 1 \\ \frac{1}{2} \\ 1 \end{bmatrix}, \boldsymbol{x}_1^{\lambda_2} = \begin{bmatrix} -\frac{1}{2} \\ -1 \\ 1 \end{bmatrix}, \boldsymbol{x}_1^{\lambda_3} = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}.$$

3. 现在利用Gram–Schmidt算法将$\boldsymbol{x}_1^{\lambda_1}$正交规范化为$V_{\lambda_1}$的一个标准正交基底，将$\boldsymbol{x}_1^{\lambda_2}$正交规范化为$V_{\lambda_2}$的一个标准正交基底，将$\boldsymbol{x}_1^{\lambda_3}$正交规范化为$V_{\lambda_3}$的一个标准正交基底。显然，此时我们只需规范化(normalizing)$\boldsymbol{x}_1^{\lambda_1}$，$\boldsymbol{x}_1^{\lambda_2}$，$\boldsymbol{x}_1^{\lambda_3}$即可。因此得到

$$\boldsymbol{v}_1^{\lambda_1} = \frac{1}{\|\boldsymbol{x}_1^{\lambda_1}\|}\boldsymbol{x}_1^{\lambda_1} = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}, \boldsymbol{v}_1^{\lambda_2} = \frac{1}{\|\boldsymbol{x}_1^{\lambda_2}\|}\boldsymbol{x}_1^{\lambda_2} = \begin{bmatrix} -1/3 \\ -2/3 \\ 2/3 \end{bmatrix}, \boldsymbol{v}_1^{\lambda_3} = \frac{1}{\|\boldsymbol{x}_1^{\lambda_3}\|}\boldsymbol{x}_1^{\lambda_3} = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix}.$$

   从而有$P = \begin{bmatrix} \boldsymbol{v}_1^{\lambda_1} & \boldsymbol{v}_1^{\lambda_2} & \boldsymbol{v}_1^{\lambda_3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$，也即所求的orthogonal change of variables为

$$\boldsymbol{y} = P^\top \boldsymbol{x} \text{ 或者 } \boldsymbol{x} = P\boldsymbol{y}.$$

   此时我们有$Q(P\boldsymbol{y}) = Q_A(P\boldsymbol{y}) = (P\boldsymbol{y})^\top AP\boldsymbol{y} = \boldsymbol{y}^\top(P^\top AP)\boldsymbol{y} = \boldsymbol{y}^\top \begin{bmatrix} 0 & & \\ & -3 & \\ & & 3 \end{bmatrix} \boldsymbol{y} = -3y_2^2 + 3y_3^2.$

**orthogonal change of variables的数学意义：**

考虑二次型$Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x}$，$A \in M_{n\times n}$为对称矩阵。设$A$的正交对角化为

$$P^\top AP = D,$$

其中$P$为正交矩阵，$D$为对角矩阵。设$P$的列向量表示为

$$P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \end{bmatrix}.$$

由Theorem 6.13的证明我们知道$\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$都是$A$的特征向量，并且$B' = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$是$\mathbb{R}^n$的一组(关于欧式内积的)正交基底。

现在令$B = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$为$\mathbb{R}^n$的标准基底，显然它也是一组(关于欧式内积的)正交基底。容易看出

$$P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} [\boldsymbol{v}_1]_B & [\boldsymbol{v}_2]_B & \dots & [\boldsymbol{v}_n]_B \end{bmatrix} = P_{B \leftarrow B'}$$

是从$B'$到$B$的转移矩阵。因此orthogonal change of variables$\boldsymbol{y} = P^\top \boldsymbol{x} = P^{-1}\boldsymbol{x} = P_{B\leftarrow B'}^{-1}\boldsymbol{x} = P_{B'\leftarrow B}\boldsymbol{x}$实际上是将$\boldsymbol{x} \in \mathbb{R}^n$关于标准基底$B$的坐标向量$[\boldsymbol{x}]_B = \boldsymbol{x}$替换为$\boldsymbol{x} \in \mathbb{R}^n$关于另一组正交基底$B' = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$的坐标向量$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \boldsymbol{y}$，即$\boldsymbol{x} = x_1\boldsymbol{e}_1 + \dots + x_n\boldsymbol{e}_n = y_1\boldsymbol{v}_1 + \dots + y_n\boldsymbol{v}_n$。

**例子：** 在这个例子里我们试图找出方程

$$5x^2 - 4xy + 8y^2 = 36$$

在平面$\mathbb{R}^2$上所描述的图形形状，即集合$\{(x,y) \in \mathbb{R}^2 : 5x^2 - 4xy + 8y^2 = 36\}$的几何形状。

带着这个目的，我们进行如下操作：首先利用**orthogonal change of variables**将二次型$Q(\boldsymbol{x} = (x,y)) = 5x^2 - 4xy + 8y^2$化简为$a(x')^2 + b(y')^2$的形式，此时原方程$5x^2 - 4xy + 8y^2 = 36$可以用$(x', y')$表示为

$$a(x')^2 + b(y')^2 = 36,$$

显然这是一个平面上的椭圆方程。接下来我们分析**坐标变换**$(x,y) \mapsto (x',y')$**的几何意义**，最终确定原集合$\{(x,y) \in \mathbb{R}^2 : 5x^2 - 4xy + 8y^2 = 36\}$的几何形状。

具体计算过程如下：容易确定二次型$Q(\boldsymbol{x} = (x,y)) = 5x^2 - 4xy + 8y^2$对应的对称矩阵为

$$A = \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix},$$

即$Q(\boldsymbol{x} = (x,y)) = 5x^2 - 4xy + 8y^2 = \boldsymbol{x}^\top A\boldsymbol{x}$。对$A$进行正交对角化，可得

$$D = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}, P = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix},$$

因此对于$\boldsymbol{y}=(x',y')=P^{\top}\boldsymbol{x} \iff \boldsymbol{x}=P\boldsymbol{y}$，我们有

$$Q(\boldsymbol{x})=Q_A(P\boldsymbol{y})=4(x')^2+9(y')^2,$$

从而得到

$\boldsymbol{x}=(x,y)$ 满足 $5x^2-4xy+8y^2=36 \iff \boldsymbol{y}=(x',y')=P^{\top}\boldsymbol{x}$ 满足 $4(x')^2+9(y')^2=36.$

容易看出$4(x')^2+9(y')^2=36 \iff \frac{(x')^2}{9}+\frac{(y')^2}{4}=1$是一个**椭圆**。

由于$P=P_{B\leftarrow B'}$，这里$B=\{(1,0),(0,1)\}$，$B'=\{\boldsymbol{v}_1=\begin{bmatrix}\frac{2}{\sqrt{5}}\\ \frac{1}{\sqrt{5}}\end{bmatrix},\boldsymbol{v}_2=\begin{bmatrix}-\frac{1}{\sqrt{5}}\\ \frac{2}{\sqrt{5}}\end{bmatrix}\}$，我们知道$\boldsymbol{y}=(x',y')=P^{\top}\boldsymbol{x}=P^{-1}\boldsymbol{x}=P_{B'\leftarrow B}\boldsymbol{x}$实际上是$\boldsymbol{x}=(x,y)$关于基底$B'$的坐标向量，也即$\boldsymbol{x}=\begin{bmatrix}x\\ y\end{bmatrix}=x'\boldsymbol{v}_1+y'\boldsymbol{v}_2=x'\begin{bmatrix}\frac{2}{\sqrt{5}}\\ \frac{1}{\sqrt{5}}\end{bmatrix}+y'\begin{bmatrix}-\frac{1}{\sqrt{5}}\\ \frac{2}{\sqrt{5}}\end{bmatrix}$。另一方面，我们令$\cos\theta=\frac{2}{\sqrt{5}}$，那么$\sin\theta=\frac{1}{\sqrt{5}}$，那么显然有

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

即$P$表示的是**逆时针旋转$\theta$度的操作**。特别地，因为$\boldsymbol{v}_1=\begin{bmatrix}\frac{2}{\sqrt{5}}\\ \frac{1}{\sqrt{5}}\end{bmatrix}=\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}=P\begin{bmatrix}1\\ 0\end{bmatrix}$，$\boldsymbol{v}_2=\begin{bmatrix}-\frac{1}{\sqrt{5}}\\ \frac{2}{\sqrt{5}}\end{bmatrix}=\begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}=P\begin{bmatrix}0\\ 1\end{bmatrix}$，我们可以确定在$B'=\{\boldsymbol{v}_1,\boldsymbol{v}_2\}$所决定的平面直角对角系中的横轴为$\boldsymbol{v}_1=\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}$是由标准直角坐标系的横轴$\boldsymbol{e}_1=\begin{bmatrix}1\\ 0\end{bmatrix}$逆时针旋转$\theta$度得到，纵轴$\boldsymbol{v}_2=\begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}$是由标准直角坐标系的纵轴$\boldsymbol{e}_2=\begin{bmatrix}0\\ 1\end{bmatrix}$逆时针旋转$\theta$度得到。
综上，利用已经得到的所有结果我们可以确定

$$\{(x,y):5x^2-4xy+8y^2=36\}$$

所描述的图形是将标准平面直角坐标系逆时针旋转$\theta$度得到的另一个平面直角对角系中的椭圆$\frac{(x')^2}{9}+\frac{(y')^2}{4}=1$，这里$\theta=\arccos(\frac{2}{\sqrt{5}})$。

在本节的最后我们介绍几种特殊的二次型：

**Definition 7.18.** 二次型$Q_A(\boldsymbol{x})=\boldsymbol{x}^{\top}A\boldsymbol{x}$被称为

- ***positive definite***，如果对任何$\boldsymbol{x}\neq\boldsymbol{0}$都有$Q_A(\boldsymbol{x})=\boldsymbol{x}^{\top}A\boldsymbol{x}>0$。
- ***negative definite***，如果对任何$\boldsymbol{x}\neq\boldsymbol{0}$都有$Q_A(\boldsymbol{x})=\boldsymbol{x}^{\top}A\boldsymbol{x}<0$。

- ***indefinite***，如果同时存在一个$\boldsymbol{x} \neq \mathbf{0}$使得$Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} > 0$，一个$\boldsymbol{u} \neq \mathbf{0}$使得$Q_A(\boldsymbol{u}) = \boldsymbol{u}^\top A\boldsymbol{u} < 0$。
- ***positive semidefinite***，如果对任何$\boldsymbol{x} \in \mathbb{R}^n$都有$Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} \geq 0$。
- ***negative semidefinite***，如果对任何$\boldsymbol{x} \in \mathbb{R}^n$都有$Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} \leq 0$。

下面的定理告诉我们二次型的以上性质与它对应的**对称矩阵的特征值**有关。

**Theorem 7.19.** 令$A \in M_{n\times n}$为一个对称矩阵。

1. $Q_A$是*positive definite*当且仅当$A$的所有特征值都大于零。
2. $Q_A$是*negative definite*当且仅当$A$的所有特征值都小于零。
3. $Q_A$是*indefinite*当且仅当$A$具有正特征值也有负特征值。
4. $Q_A$是*positive semidefinite*当且仅当$A$的所有特征值都非负。
5. $Q_A$是*negative semidefinite*当且仅当$A$的所有特征值都非正。

证明.

1. 由Theorem 7.17，通过orthogonal change of variables，即令$\boldsymbol{y} = P^\top\boldsymbol{x} \iff \boldsymbol{x} = P\boldsymbol{y}$，对任何$\boldsymbol{x} \in \mathbb{R}^n$，都有
$$\boldsymbol{x}^\top A\boldsymbol{x} = \boldsymbol{y}^\top(P^\top AP)\boldsymbol{y} = \boldsymbol{y}^\top D\boldsymbol{y} = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2.$$
由于$P$作为正交矩阵是可逆的，$\boldsymbol{x} \neq \mathbf{0} \iff \boldsymbol{y} = P^{-1}\boldsymbol{x} \neq \mathbf{0}$，因此我们有以下等价关系
$$\boldsymbol{x}^\top A\boldsymbol{x} > 0 \text{ 对任何非零的 } \boldsymbol{x} \iff \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 > 0 \text{ 对任何非零的 } \boldsymbol{y}.$$
显然，$\lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 > 0$对任何$\boldsymbol{y} \neq \mathbf{0}$成立当且仅当$\lambda_j > 0$对任何$j = 1, \ldots, n$都成立。现在只需注意$\lambda_j$，$j = 1, \ldots, n$都是$A$的特征值。
2. 对于$Q_A$是negative definite的情况类似可证。
3. 对于$Q_A$是indefinite的情况类似可证。只需注意存在一个$\boldsymbol{y} \in \mathbb{R}^n$使得$\lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 > 0$当且仅当至少有一个$\lambda_j > 0$，存在一个$\boldsymbol{y} \in \mathbb{R}^n$使得$\lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 < 0$当且仅当至少有一个$\lambda_j < 0$。
4. 对于$Q_A$是positive semidefinite的情况类似可证。
5. 对于$Q_A$是negative semidefinite的情况类似可证。

□

最后我们介绍一个实用的判断对称矩阵是否positive definite的标准。

**Definition 7.20.** 令$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}$为$n$阶对称矩阵。我们称

- $\begin{bmatrix} a_{11} \end{bmatrix} \in M_{1\times 1}$为$A$的***first principal submatrix***.
- $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_{2\times 2}$为$A$的***second principal submatrix***.
- 对于任何$k = 1, \ldots, n$，我们称$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \ldots & a_{kk} \end{bmatrix} \in M_{k\times k}$为$A$的***k–th principal submatrix***.

**Theorem 7.21.** 对任何对称矩阵$A \in M_{n\times n}$，$A$是*positive definite*当且仅当$A$的$k$*–th* ***principal submatrix***的行列式

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \ldots & a_{kk} \end{vmatrix}$$

大于0对所有$k = 1, \ldots, n$成立。

**例子：** 对于对称矩阵

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

容易计算出

$$\text{first principal submatrix} = \begin{bmatrix} 2 \end{bmatrix} \Rightarrow \det(\begin{bmatrix} 2 \end{bmatrix}) = 2 > 0;$$

$$\text{second principal submatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \Rightarrow \det(\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}) = 3 > 0;$$

$$\text{third principal submatrix} = \begin{bmatrix} 2 & -1 & -3 \\ -1 & 2 & 4 \\ -3 & 4 & 9 \end{bmatrix} \Rightarrow \det(\begin{bmatrix} 2 & -1 & -3 \\ -1 & 2 & 4 \\ -3 & 4 & 9 \end{bmatrix}) = 1 > 0.$$

因此可以判定$A$是positive definite。
注意，**一个对称矩阵$A$是negative definite的当且仅当$-A$是positive definite**，因此我们可以用以上方法通过判断$-A$是否是positive definite来判断$A$是否是negative definite的。但是，请注意$A$是negative definite并不等价于它的任何k–th principal submatrix的行列式都小于0!

## 7.4. QR分解与奇异值分解, QR Decomposition and Singular Value Decomposition

在本节我们学习两种非常重要的矩阵分解方法。首先是QR分解，见英文教材6.3节。

**Theorem 7.22.** 假设$A \in M_{m\times n}$满足$rank(A) = n$。那么$A$可以被表示为

$$A = QR,$$

这里$Q \in M_{m\times n}$且$Q$的$n$个列向量组成$\mathbb{R}^m$里的一个标准正交集合，$R \in M_{n\times n}$为一个可逆的上三角矩阵。

证明. 设$A \in M_{m\times n}$的列向量表示为

$$A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_n \end{bmatrix}$$

其中每个$\boldsymbol{u}_j \in \mathbb{R}^m$为$m\times 1$-列向量。由于rank$(A) = n$，可知这些列向量$\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$线性无关，即它们是$A$的列空间Col$(A)$的一组基底。**那么对$\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$使用Gram–Schmidt算法，可以得到$A$的列空间Col$(A)$的一组标准正交基底**，记为

$$\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}.$$

现在定义矩阵$Q$为

$$Q = \begin{bmatrix} \boldsymbol{q}_1 & \dots & \boldsymbol{q}_n \end{bmatrix}.$$

那么利用标准正交基底的特性，见Theorem 7.15，得到

$$\boldsymbol{u}_j = (\boldsymbol{u}_j \cdot \boldsymbol{q}_1)\boldsymbol{q}_1 + (\boldsymbol{u}_j \cdot \boldsymbol{q}_2)\boldsymbol{q}_2 + \dots + (\boldsymbol{u}_j \cdot \boldsymbol{q}_n)\boldsymbol{q}_n, j = 1, \dots, n.$$

因此，我们定义矩阵$R$为

$$R = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{q}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_1 & \dots & \boldsymbol{u}_n \cdot \boldsymbol{q}_1 \\ \boldsymbol{u}_1 \cdot \boldsymbol{q}_2 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_2 & \dots & \boldsymbol{u}_n \cdot \boldsymbol{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_1 \cdot \boldsymbol{q}_n & \boldsymbol{u}_2 \cdot \boldsymbol{q}_n & \dots & \boldsymbol{u}_n \cdot \boldsymbol{q}_n \end{bmatrix}.$$

利用矩阵乘法的运算法则以及等式$\boldsymbol{u}_j = (\boldsymbol{u}_j \cdot \boldsymbol{q}_1)\boldsymbol{q}_1 + (\boldsymbol{u}_j \cdot \boldsymbol{q}_2)\boldsymbol{q}_2 + \dots + (\boldsymbol{u}_j \cdot \boldsymbol{q}_n)\boldsymbol{q}_n, j = 1, \dots, n.$，容易验证

$$QR = \begin{bmatrix} \boldsymbol{q}_1 & \dots & \boldsymbol{q}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{q}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_1 & \dots & \boldsymbol{u}_n \cdot \boldsymbol{q}_1 \\ \boldsymbol{u}_1 \cdot \boldsymbol{q}_2 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_2 & \dots & \boldsymbol{u}_n \cdot \boldsymbol{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_1 \cdot \boldsymbol{q}_n & \boldsymbol{u}_2 \cdot \boldsymbol{q}_n & \dots & \boldsymbol{u}_n \cdot \boldsymbol{q}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_n \end{bmatrix} = A.$$

显然，由以上定义，$Q$的列向量$\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$是$A$的列空间Col$(A)$的一组标准正交基底，因此$Q$的列向量$\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$组成了一个$\mathbb{R}^m$里的标准正交集合。我们现在证明$R$是上三角矩阵：由Gram-Schmidt算法的内容可知，

$$\boldsymbol{q}_2 \cdot \boldsymbol{u}_1 = 0,$$

$$\boldsymbol{q}_3 \cdot \boldsymbol{u}_1 = 0, \quad \boldsymbol{q}_3 \cdot \boldsymbol{u}_2 = 0,$$

$$\boldsymbol{q}_j \cdot \boldsymbol{u}_1 = 0, \quad \ldots, \quad \boldsymbol{q}_j \cdot \boldsymbol{u}_{j-1} = 0$$

对任何$2 \leq j \leq n$成立。因此可知

$$R = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{q}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_1 & \ldots & \boldsymbol{u}_n \cdot \boldsymbol{q}_1 \\ 0 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_2 & \ldots & \boldsymbol{u}_n \cdot \boldsymbol{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{u}_n \cdot \boldsymbol{q}_n \end{bmatrix}$$

必然为上三角矩阵。

现在我们证明$R$可逆，即证明$\det(R) \neq 0$。再次利用Gram–Schmidt算法的具体内容，我们有

$$\boldsymbol{q}_1 = \frac{1}{\|\boldsymbol{u}_1\|}\boldsymbol{u}_1 \Rightarrow \boldsymbol{u}_1 \cdot \boldsymbol{q}_1 = \|\boldsymbol{u}_1\| > 0;$$

对于$\boldsymbol{u}_2 \cdot \boldsymbol{q}_2$，注意Gram–Schmidt算法实际上保证了

$$\text{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\} = \text{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\},$$

因此如果$\boldsymbol{u}_2 \cdot \boldsymbol{q}_2 = 0$，那么$\boldsymbol{q}_2$同时正交于$\boldsymbol{u}_1$和$\boldsymbol{u}_2$，因此$\boldsymbol{q}_2$正交于span$\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$，因此$\boldsymbol{q}_2 = \boldsymbol{0}$，这显然不可能。因此$\boldsymbol{u}_2 \cdot \boldsymbol{q}_2 \neq 0$类似可证$\boldsymbol{u}_j \cdot \boldsymbol{q}_j \neq 0$对任何$j = 1, \ldots, n$成立。故有$\det(R) = \boldsymbol{u}_1 \cdot \boldsymbol{q}_1 \times \boldsymbol{u}_2 \cdot \boldsymbol{q}_2 \times \ldots \times \boldsymbol{u}_n \cdot \boldsymbol{q}_n \neq 0$。 □

**例子：** 求矩阵

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

的$QR$分解。

**解：** $A$的列向量为

$$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

由Gram–Schmidt算法，可求得$A$的列空间的一组标准正交基底：

$$\boldsymbol{q}_1 = \frac{1}{\|\boldsymbol{u}_1\|}\boldsymbol{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix},$$

$$\boldsymbol{q}_2 = \frac{\boldsymbol{u}_2 - (\boldsymbol{u}_2 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1}{\|\boldsymbol{u}_2 - (\boldsymbol{u}_2 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1\|} = \begin{bmatrix} -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix},$$

$$\boldsymbol{q}_3 = \frac{\boldsymbol{u}_3 - (\boldsymbol{u}_3 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1 - (\boldsymbol{u}_3 \cdot \boldsymbol{q}_2)\boldsymbol{q}_2}{\|\boldsymbol{u}_3 - (\boldsymbol{u}_3 \cdot \boldsymbol{q}_1)\boldsymbol{q}_1 - (\boldsymbol{u}_3 \cdot \boldsymbol{q}_2)\boldsymbol{q}_2\|} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

因此$Q = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix}$。

按照之前定理的证明，可求得上三角矩阵$R$：

$$R = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{q}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_1 & \boldsymbol{u}_3 \cdot \boldsymbol{q}_1 \\ 0 & \boldsymbol{u}_2 \cdot \boldsymbol{q}_2 & \boldsymbol{u}_3 \cdot \boldsymbol{q}_2 \\ 0 & 0 & \boldsymbol{u}_3 \cdot \boldsymbol{q}_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

因此$A$的QR分解为

$$A = QR = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

我们接下来学习矩阵的**奇异值分解(Singular Value Decomposition)**，简称SVD。计算奇异值分解的背景如下：我们知道，对于$A \in M_{n\times n}$且$A$为对称矩阵，那么存在正交矩阵$P \in M_{n\times n}$以及对角矩阵$D \in M_{n\times n}$使得

$$A = PDP^\top.$$

这个分解也被称为$A$的**特征值分解(eigenvalue decomposition)**。

现在我们考虑一个$m \times n$的矩阵$A$。我们想知道，在$m \neq n$的时候，是否还存在类似的分解。奇异值分解则告诉我们，对于$A \in M_{m\times n}$，存在分解

$$A = U\Sigma V^\top,$$

其中$U$为$m \times m$正交矩阵，$V$为$n \times n$正交矩阵，$\Sigma$为$m \times n$的对角矩阵。

我们首先对**不是方阵**的矩阵$A \in M_{m\times n}$，$m \neq n$定义对角线。

**Definition 7.23.** • 如果$m < n$，那么矩阵$A \in M_{m\times n}$的对角线为$a_{11}, a_{22}, \ldots, a_{mm}$。

- 如果$m > n$，那么矩阵$A \in M_{m\times n}$的对角线为$a_{11}, a_{22}, \ldots, a_{nn}$。
- 我们称一个矩阵$\Sigma \in M_{m\times n}$是一个对角矩阵，如果它的对角线之外的项都是0。

比如$3\times 4$–矩阵$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}$的对角线是$1, 2, 3$，且这样的$A$为一个对角矩阵。$4\times 3$–矩阵$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \\ 3 & 2 & 1 \end{bmatrix}$的对角线是$1, 2, 3$，且这样的$A$不是一个对角矩阵。

接下来我们定义**奇异值(singular values)**。对于$A \in M_{m\times n}$，显然有$A^\top A$为$n\times n$–对称矩阵。

**Theorem 7.24.** 对于$A \in M_{m\times n}$，我们有

1. $A^\top A \in M_{n\times n}$可正交对角化。
2. $A^\top A \in M_{n\times n}$的所有特征值都非负*(nonnegative)*，因此$A^\top A$是*positive semidefinite*。

*证明.* 由于$A^\top A \in M_{n\times n}$为对称矩阵，由Theorem 7.9可知$A^\top A$可以正交对角化。假设$\lambda \in \mathbb{R}$是$A^\top A$的一个特征值，$\boldsymbol{x} \neq \mathbf{0}$是对应的一个特征向量，即

$$(A^\top A)\boldsymbol{x} = \lambda \boldsymbol{x}.$$

因此，可以推出

$$\|A\boldsymbol{x}\|^2 = A\boldsymbol{x} \cdot A\boldsymbol{x} = (A\boldsymbol{x})^\top (A\boldsymbol{x}) = \boldsymbol{x}^\top (A^\top A\boldsymbol{x}) = \boldsymbol{x}^\top (\lambda \boldsymbol{x}) = \lambda \|\boldsymbol{x}\|^2.$$

这意味着$\lambda = \frac{\|A\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2} \geq 0$。这说明$A^\top A$的所有特征值非负。因此由Theorem 7.19可知$A^\top A$是positive definite。 □

对于$A \in M_{m\times n}$，我们假设$A^\top A$的正交对角化$D = P^\top (A^\top A)P$中的对角矩阵$D$对角线上的元素为$\lambda_1, \ldots, \lambda_n$。这里我们额外要求$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$。

**Definition 7.25.** 对于$A \in M_{m\times n}$，假设$A^\top A$的正交对角化$D = P^\top (A^\top A)P$中的对角矩阵$D$对角线上的元素为$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$。我们称

$$\sigma_1 = \sqrt{\lambda_1} \geq \sigma_2 = \sqrt{\lambda_2} \geq \ldots \geq \sigma_n = \sqrt{\lambda_n}$$

为$A$的奇异值(***singular values of*** $A$)。

**Theorem 7.26** (Singular Value Decomposition)**.** 任何$A \in M_{m\times n}$都可以表示为

$$A = U\Sigma V^\top,$$

其中$U$为$m\times m$正交矩阵，$V$为$n\times n$正交矩阵，$\Sigma$为$m\times n$的对角矩阵。

该定理的证明不需要掌握，感兴趣的同学可以自行阅读讲义最后部分对该定理的证明。这里我们需要掌握的是具体的奇异值分解算法。

**奇异值分解算法：**

对于$A \in M_{m\times n}$，设$\mathrm{rank}(A) = k \leq \min\{m, n\}$。

1. $V \in M_{n\times n}$为对称矩阵$A^\top A$的正交对角化中出现的正交矩阵，即

$$V^\top(A^\top A)V = D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix},$$

   注意这里我们要求$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$。
   将$V$的列向量表示记为

$$V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n \end{bmatrix}.$$

2. 由于$\mathrm{rank}(A) = k$，我们可以证明(具体过程可参考讲义最后部分对以上定理的证明)这时必然有$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$，$\lambda_{k+1} = \ldots = \lambda_n = 0$。令$\sigma_1 = \sqrt{\lambda_1} \geq \ldots \geq \sigma_k = \sqrt{\lambda_k} > 0$为$A$的前$k$个奇异值，定义$m\times n$–对角矩阵$\Sigma$为

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}.$$

3. 定义

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1}A\boldsymbol{v}_1, \ldots, \boldsymbol{u}_k = \frac{1}{\sigma_k}A\boldsymbol{v}_k$$

注意这里$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$是$V$的前$k$个列向量。我们可以证明$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$是$A$的列空间Col$(A)$的一组标准正交基底(可参考讲义最后对以上定理的证明)。接下来通过基底扩充与Gram–Schmidt算法将$\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$扩展为$\mathbb{R}^m$的一组**标准正交基底**$\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_m\}$。定义

$$U = \begin{bmatrix} \boldsymbol{u}_1 & \ldots & \boldsymbol{u}_k & \boldsymbol{u}_{k+1} & \ldots & \boldsymbol{u}_m \end{bmatrix}.$$

4. 最终我们得到

$$A = U\Sigma V^\top.$$

**例子：** 令矩阵$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$。求$A$的奇异值分解$A = U\Sigma V^\top$。

我们按照之前介绍的算法求出$U$，$V$和$\Sigma$。

1. 容易计算$A^\top A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$。由此可得对称矩阵$A^\top A$的特征多项式为

$$p(\lambda) = \det(\lambda I_2 - A) = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1).$$

求解特征方程$p(\lambda) = 0$，得到$A^\top A$的特征值为

$$\lambda_1 = 3, \lambda_2 = 1$$

(注意我们这里要求$\lambda_1 > \lambda_2$)。因此我们得到$A$的两个奇异值为$\sigma_1 = \sqrt{3}, \sigma_2 = \sqrt{1} = 1$。所以对角矩阵$\Sigma \in M_{3\times 2}$等于

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

2. 接下来我们求正交矩阵$V \in M_{2\times 2}$。我们知道它是由对称矩阵$A^\top A$的特征空间$V_{\lambda_1} = \text{Null}(\lambda_1 I_2 - A^\top A)$的标准正交基底$\boldsymbol{v}_1$与特征空间$V_{\lambda_2} = \text{Null}(\lambda_2 I_2 - A^\top A)$的标准正交基底$\boldsymbol{v}_2$组成(即$V^\top(A^\top A)V = D = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$正交对角化$A^\top A$)。因此求解齐次方程组

$$(\lambda_1 I_2 - A^\top A)\boldsymbol{x} = \mathbf{0}$$

并规范化(normalizing)所得的解空间基底，可得$\boldsymbol{v}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$；求解齐次方程组

$$(\lambda_2 I_2 - A^\top A)\boldsymbol{x} = \mathbf{0}$$

并规范化(normalizing)所得的解空间基底，可得$\boldsymbol{v}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$。由此可得

$$V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

3. 现在我们求出正交矩阵$U \in M_{3\times 3}$。由奇异值分解算法的第三步可知，$U$的前两个列向量依次为

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \frac{\sqrt{3}}{3} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix},$$

$$\boldsymbol{u}_2 = \frac{1}{\sigma_2} A\boldsymbol{v}_2 = 1 \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}.$$

我们现在需要找到第三个列向量$\boldsymbol{u}_3 \in \mathbb{R}^3$使得$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$组成$\mathbb{R}^3$的一组标准正交基底。对此我们假设$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$属于span$\{\boldsymbol{u}_1, \boldsymbol{u}_2\}^\perp$，那么

$$\boldsymbol{x} \cdot \boldsymbol{u}_1 = \frac{\sqrt{6}}{3} x_1 + \frac{\sqrt{6}}{6} x_2 + \frac{\sqrt{6}}{6} x_3 = 0,$$

$$\boldsymbol{x} \cdot \boldsymbol{u}_2 = 0x_1 - \frac{\sqrt{2}}{2} x_2 + \frac{\sqrt{2}}{2} x_3 = 0.$$

因此求解方程组

$$\begin{bmatrix} \frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

求得span$\{\boldsymbol{u}_1, \boldsymbol{u}_2\}^\perp$的一组基底为$\boldsymbol{x} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$。将其规范化，可得$\boldsymbol{u}_3 = \boldsymbol{x}/\|\boldsymbol{x}\| = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$。因此$U$等于

$$U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

*Theorem 7.26*的证明. 我们证明206–207页上介绍的奇异值分解算法的有效性，即按这个算法得到的矩阵$U, V, \Sigma$确实满足$A = U\Sigma V^\top$。我们假设$\mathrm{rank}(A) = k$。

- 首先我们证明如果$\mathrm{rank}(A) = k$，那么$A^\top A$的特征值$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$满足$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$，$\lambda_{k+1} = \ldots = \lambda_n = 0$。
  由Theorem 4.38我们知道$\mathrm{rank}(A^\top A) = \mathrm{rank}(A) = k$。另一方面，考虑对称矩阵$A^\top A$的正交对角化
  $$D = V^\top(A^\top A)V, \quad D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$
  我们当然有$D$与$A^\top A$相似。那么由于rank是相似不变量(见Theorem 5.21)，我们得到
  $$\mathrm{rank}(D) = \mathrm{rank}(A^\top A) = \mathrm{rank}(A) = k.$$
  显然，作为对角矩阵$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$的秩等于$k$必然意味着$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$且$\lambda_{k+1} = \ldots = \lambda_n = 0$。
- 我们证明关于矩阵$U$的一些性质。我们令$V = \begin{bmatrix} \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_n \end{bmatrix} \in M_{n\times n}$为$A^\top A$正交对角化中出现的正交矩阵，即$D = V^\top(A^\top A)V$。
  首先证明$\{A\boldsymbol{v}_1, \ldots, A\boldsymbol{v}_k\}$是$\mathbb{R}^m$里的一个正交集合。注意由$V$的定义可知，$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$是$A^\top A$关于$\lambda_1, \ldots, \lambda_n$的特征向量且$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$本身是$\mathbb{R}^n$里的一个标准正交基底，因此$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$本身是$\mathbb{R}^n$里的一个正交集合。
  对于任何$i, j = 1, \ldots, k$，我们有
  $$A\boldsymbol{v}_i \cdot A\boldsymbol{v}_j = (A\boldsymbol{v}_i)^\top (A\boldsymbol{v}_j) = \boldsymbol{v}_i^\top (A^\top A)\boldsymbol{v}_j = \boldsymbol{v}_i \cdot (A^\top A)\boldsymbol{v}_j = \boldsymbol{v}_i \cdot (\lambda_j)\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_i \cdot \boldsymbol{v}_j.$$
  因此，基于$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$本身是$\mathbb{R}^n$里的一个正交规范集合，可得
  $$A\boldsymbol{v}_i \cdot A\boldsymbol{v}_j = \begin{cases} 0 \text{ 如果 } i \neq j \\ \lambda_j \text{ 如果 } i = j \end{cases}$$
  同理可证，对于$i = k+1, \ldots, n$，因为$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$，我们有
  $$\|A\boldsymbol{v}_i\|^2 = A\boldsymbol{v}_i \cdot A\boldsymbol{v}_i = \underbrace{\lambda_i}_{=0} \|\boldsymbol{v}_i\|^2 = 0.$$
  因此，$A\boldsymbol{v}_i = \mathbf{0}$对$i = k+1, \ldots, n$成立。
  现在，对$i = 1, \ldots, k$，令
  $$\boldsymbol{u}_i = \frac{1}{\sigma_i} A\boldsymbol{v}_i, \quad \sigma_i = \sqrt{\lambda_i} > 0,$$

对$i,j=1,\dots,k$，我们有

$$\boldsymbol{u}_i\cdot\boldsymbol{u}_j=\frac{1}{\sigma_i}A\boldsymbol{v}_i\cdot\frac{1}{\sigma_j}A\boldsymbol{v}_j=\frac{1}{\sigma_i\sigma_j}(A\boldsymbol{v}_i\cdot A\boldsymbol{v}_j)=\begin{cases}0\text{ 如果 }i\neq j\\ \frac{1}{\lambda_i}\times\lambda_i=1\text{ 如果 }i=j\end{cases}.$$

因此$\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_k\}$是$\mathbb{R}^m$里的一个正交规范集合。利用基底扩充与Gram–Schmidt算法，可将其扩充为$\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_k,\boldsymbol{u}_{k+1},\dots,\boldsymbol{u}_m\}$是$\mathbb{R}^m$里的一个标准正交基底。

- 最后我们验证在算法里定义的矩阵$U,V,\Sigma$确实满足$A=U\Sigma V^\top$。
  对于$U=\begin{bmatrix}\boldsymbol{u}_1&\dots&\boldsymbol{u}_m\end{bmatrix}$，由于它的列向量组成了$\mathbb{R}^m$里的一个标准正交基底，$U\in M_{m\times m}$是一个正交矩阵。令

$$\Sigma=\begin{bmatrix}\sigma_1&&&&&&\\&\sigma_2&&&&&\\&&\ddots&&&&\\&&&\sigma_k&&&\\&&&&0&&\\&&&&&\ddots&\\&&&&&&0\end{bmatrix},$$

那么利用矩阵乘法容易求得

$$\begin{aligned}U\Sigma&=\begin{bmatrix}\sigma_1\boldsymbol{u}_1&\sigma_2\boldsymbol{u}_2&\dots&\sigma_k\boldsymbol{u}_k&\mathbf{0}&\dots&\mathbf{0}\end{bmatrix}\\&=\begin{bmatrix}\sigma_1\frac{1}{\sigma_1}A\boldsymbol{v}_1&\sigma_2\frac{1}{\sigma_2}A\boldsymbol{v}_2&\dots&\sigma_k\frac{1}{\sigma_k}A\boldsymbol{v}_k&\mathbf{0}&\dots&\mathbf{0}\end{bmatrix}\\&=\begin{bmatrix}A\boldsymbol{v}_1&A\boldsymbol{v}_2&\dots&A\boldsymbol{v}_k&\mathbf{0}&\dots&\mathbf{0}\end{bmatrix}.\end{aligned}$$

由于我们之前证明了$A\boldsymbol{v}_i=\mathbf{0}$对$i=k+1,\dots,n$成立，所以

$$U\Sigma=\begin{bmatrix}A\boldsymbol{v}_1&A\boldsymbol{v}_2&\dots&A\boldsymbol{v}_k&\mathbf{0}&\dots&\mathbf{0}\end{bmatrix}=\begin{bmatrix}A\boldsymbol{v}_1&A\boldsymbol{v}_2&\dots&A\boldsymbol{v}_k&A\boldsymbol{v}_{k+1}&\dots&A\boldsymbol{v}_n\end{bmatrix}=AV,$$

因此，由于$V$是一个正交矩阵，由$U\Sigma=AV$可得

$$A=U\Sigma V^\top.$$

□