

# MIT-18.06: Linear Algebra | 线性代数

[课程视频](#) 和 [配套教材](#) (提取码: whs9) 在这里!

如果我大二的时候学的是这套线代教程, 我的人生轨迹就会发生变化: 可能绩点会有大的改观, 可能拿到保研资格, 可能不会因为考研不考数学而跟医工院深度绑定, 可能离开天津前往上海, 可能硕士毕业就参加工作, 可能不会在一个不喜欢的行业继续读博苟延残喘, 之后的人生理应处处发生改变吧。

这份文档并非是面面俱到的笔记, 而是我在学习相关课程中发现的重要结论与新的收获, 因此部分章节目录存在跳跃。

若无特殊说明, 本文中大写字母表示二维矩阵, 小写字母表示向量。

## 1 Introduction to Vectors

### 1.2 Length and Dot Products

**点乘**是一个需要与矩阵乘法相区分的运算操作, 在泛函里一般称为**内积**。在向量这个低级维度, 可以简单理解为对应元素相乘并相加, 即对于  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{m \times 1}$ :

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^m x_i y_i \iff \mathbf{y}^T \mathbf{x} \quad (1-2-1)$$

虽然说点积与矩阵乘法不一样, 但矩阵乘法里处处都是点积。结果矩阵中的每一个元素其实都来源于一个行向量与列向量的点积。

**向量长度**, 通常又称为向量的2范数, 表示向量空间中的零元至该点的空间距离, 其中  $|\cdot|$  表示取模:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^m |x_i|^2} \quad (1-2-2)$$

#### 1.2.1 向量单位化

在此基础上, 我们可以定义单位向量  $\mathbf{u}$  以及将一个任意非零向量  $\mathbf{v}$  单位化的过程:

$$\|\mathbf{u}\|_2 = 1, \mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \quad (1-2-3)$$

#### 1.2.2 正交向量

对于点乘结果为 0 的特殊向量组合，将其称为正交向量组，即二者相互垂直：

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0 \iff \mathbf{x} \perp \mathbf{y} \quad (1-2-4)$$

### 1.2.3 向量夹角

点乘并非只能判断是否正交，还能进一步量化向量间的夹角。对于单位向量  $\mathbf{u}$ 、 $\mathbf{v}$ ：

$$\mathbf{u} \cdot \mathbf{v} = \cos\theta \quad (1-2-5)$$

### 1.2.4 余弦公式 & 柯西不等式

由 (1-2-5) 进一步可得面向两个非单位向量（如  $\mathbf{x}$ 、 $\mathbf{y}$ ）夹角的余弦公式：

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad (1-2-6)$$

这一公式也常用于高中数学立体几何部分的平面关系判断。说来惭愧，落笔之日距我高考已达7年有余，根本想不起来当初郑永盛是怎么教我们理解这个公式的了（可能单纯就是硬记吧），回首往日甚是怀念。

最后，根据三角函数数值大小的规律可得 *Cauchy-Schwarz-Buniakowsky* 不等式，严格来说，是柯西不等式的其中一种形式：

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\| \|\mathbf{y}\| \quad (1-2-7)$$

## 2 Solving Linear Equations

### 2.1 Vectors and Linear Equations

#### 2.1.1 矩阵的“列视图”

对于多元一次线性方程组：

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

该方程组可以很容易地改写为矩阵  $\mathbf{A} \in \mathbb{R}^{3 \times 3}$  与向量  $\mathbf{x} \in \mathbb{R}^{3 \times 1}$  的乘法：

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \implies \mathbf{Ax} = \mathbf{b} \quad (2-1-1)$$

我们通常习惯以“**元素的加权和**”这一视角来描述这种“数群”至“数字”的运算过程，即存在这样的思维习惯： $b_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3$ 。此时我们接受系数矩阵  $A$  的方式是逐行进行的，是一种略显复杂的动态过程。

**列视图**是一种全新视角，其本质为矩阵列的线性组合，我们可以把式 (2-1-1) 改写为：

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} + x_3 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{2-1-2}$$

在面对形如“矩阵右乘向量”之类的运算时，使用**列视图**能够更好地简化思路，在实际应用中帮助我们明确信息流传递的真实意义。

### 2.1.2 矩阵的“行视图”

对于行向量  $a$  & 矩阵  $X$  的乘法：

$$aX = b \iff \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} \tag{2-1-3}$$

对于“矩阵左乘向量”的运算形式，可以使用**行视图**对其进行观察与分析：

$$a_1 \begin{bmatrix} x_{11} & x_{12} & x_{13} \end{bmatrix} + a_2 \begin{bmatrix} x_{21} & x_{22} & x_{23} \end{bmatrix} + a_3 \begin{bmatrix} x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} \tag{2-1-4}$$

掌握矩阵的**行视图**与**列视图**有助于我们进一步理解矩阵初等变换以及逆矩阵的本质。

### 2.1.3 矩阵乘法

矩阵之间的乘法可视为**左矩阵**与多个**列向量**的乘法，抑或为多个**行向量**与**右矩阵**的乘法。例如  $AB = C$ ，除了最基础、最抽象的**运算型** (*regular way*) 表示：

The regular way: a row of  $A$  times a column of  $B$  to get a number of  $C$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} \sum_i a_{1i}b_{i1} & \sum_i a_{1i}b_{i2} & \sum_i a_{1i}b_{i3} \\ \sum_i a_{2i}b_{i1} & \sum_i a_{2i}b_{i2} & \sum_i a_{2i}b_{i3} \\ \sum_i a_{3i}b_{i1} & \sum_i a_{3i}b_{i2} & \sum_i a_{3i}b_{i3} \end{bmatrix} \tag{2-1-5}$$

还可以表示为**列型** (*column way*) 和**行型** (*row way*)：

The column way: matrix  $A$  times a column of  $B$  to get a column of  $C$   
The row way: a row of  $A$  times matrix  $B$  to get a row of  $C$

$$\mathbf{AB} = \mathbf{A} \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \oplus \mathbf{A} \begin{bmatrix} b_{21} \\ b_{22} \end{bmatrix} \oplus \mathbf{A} \begin{bmatrix} b_{31} \\ b_{32} \end{bmatrix} \quad (2-1-6)$$

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix} \mathbf{B} \oplus \begin{bmatrix} a_{21} & a_{22} \end{bmatrix} \mathbf{B} \oplus \begin{bmatrix} a_{31} & a_{32} \end{bmatrix} \mathbf{B} \quad (2-1-7)$$

此时运算符“ $\oplus$ ”视情况分别表示**列向量在横向**或**行向量在纵向**的拼接。除了上述三种运算角度，其实还有**行视图**、**列视图**混用的第四种形式：

The 4th way: a column of  $\mathbf{A}$  times a row of  $\mathbf{B}$  to get a partial matrix of  $\mathbf{C}$

$$\mathbf{AB} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \end{bmatrix} + \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \begin{bmatrix} b_{21} & b_{22} & b_{23} \end{bmatrix} = \sum_i \mathbf{A}(:, i) \mathbf{B}(i, :) \quad (2-1-8)$$

这种混合视图隐藏的信息非常重要：

- (1)  $\mathbf{AB}$  结果中的**每一列**，都是左矩阵  $\mathbf{A}$  中**各列的线性组合**
- (2)  $\mathbf{AB}$  结果中的**每一行**，都是右矩阵  $\mathbf{B}$  中**各行的线性组合**

上述结论通常与行（列）空间结合，帮助我们判断矩阵是否可逆（满秩）。

## 2.3 Elimination Using Matrices

对矩阵  $\mathbf{X}$  进行单次初等行变换，相当于  $\mathbf{X}$  左乘一个变换矩阵  $\mathbf{E}$ ，例如矩阵**第一行不变、第二行减去第一行的两倍**：

$$\underbrace{\begin{bmatrix} x_{11} & x_{12} \\ x_{21} - 2x_{11} & x_{22} - 2x_{12} \end{bmatrix}}_{\mathbf{Y}} \iff \underbrace{\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}}_{\mathbf{E}} \underbrace{\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}}_{\mathbf{X}}$$

上述过程太抽象？想想矩阵的**行视图**、矩阵乘法的**行型**：

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}}_{1st \text{ row of } \mathbf{Y}} \oplus \underbrace{\begin{bmatrix} -2 & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}}_{2nd \text{ row of } \mathbf{Y}}$$

$$\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = 1 \times \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} + 0 \times \begin{bmatrix} x_{21} & x_{22} \end{bmatrix}$$

$$\begin{bmatrix} -2 & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = -2 \times \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} + 1 \times \begin{bmatrix} x_{21} & x_{22} \end{bmatrix}$$

对应地，单次初等列变换等价于  $\mathbf{X}$  右乘一个操作矩阵  $\mathbf{C}$ ，例如矩阵**第一列除以 2、第二列乘 3 以后加上第一列变换后的 5 倍**：

$$\begin{bmatrix} \frac{x_{11}}{2} & 3x_{12} + \frac{5x_{11}}{2} \\ \frac{x_{21}}{2} & 3x_{22} + \frac{5x_{21}}{2} \end{bmatrix} \iff \underbrace{\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \frac{1}{2} & \frac{5}{2} \\ 0 & 3 \end{bmatrix}}_{\mathbf{C}}$$

朝闻道，夕死可矣。

## 2.5 Inverse Matrices

### 2.5.1 判断矩阵是否可逆

书接上回，例如病态矩阵  $\mathbf{A}$ ：

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 8 \end{bmatrix}$$

要判断  $\mathbf{A}$  是否可逆，即寻找一个矩阵  $\mathbf{A}^{-1}$  使得  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ 。根据上一节末尾的结论， $\mathbf{A}\mathbf{A}^{-1}$  的每一列都是  $\mathbf{A}$  中各列的线性组合。但是  $\mathbf{A}$  的各列是**共线**的，而  $[1, 0]^T$  所在直线并不在方向向量为  $[1, 2]^T$  的线簇中，因此在本例中我们永远无法组合出单位向量，换句话说  $\mathbf{A}$  不可逆。

更一般地，如果我们能找到一个非零列向量  $\mathbf{x} \neq \mathbf{0}$  使得方程  $\mathbf{A}\mathbf{x} = \mathbf{0}$  成立，则  $\mathbf{A}$  不可逆。这一结论可以从两方面理解：

(1)  $\mathbf{A}$  中存在某一列（行）可以由其它列（行）的线性组合表示，即该列（行）对于填充矩阵必要信息没有起到任何作用，因此这个矩阵就是“病态”的，是不完整的，也就不可逆；

(2) 若  $\mathbf{A}$  可逆，则有  $\mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{0}$ ，从而  $\mathbf{x} = \mathbf{0}$ ，这与非零列向量的前提是矛盾的，所以  $\mathbf{A}$  不可逆。

## 2.5.2 求解方阵的逆矩阵

简单起见，我们先具体到一个非奇异的二阶方阵  $B$ ：

$$B = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}$$

关于如何求解逆矩阵  $B^{-1}$ ，有两种主要思路，**第一种**是 *Gauss's way*，即依次求解二元一次线性方程组：

$$B\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, B\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \implies B^{-1} = [\mathbf{x}_1 \quad \mathbf{x}_2]$$

需要注意的是，*Gauss* 通过增广矩阵 (*augmented matrix*) 与初等行变换来实现方程组求解，以  $B\mathbf{x}_1$  为例。如果我们得到了  $\mathbf{x}_1$ ，以此类推可解出  $\mathbf{x}_2$ ：

$$\hat{B} = \left[ \begin{array}{cc|c} 1 & 3 & 1 \\ 2 & 7 & 0 \end{array} \right] \rightarrow \left[ \begin{array}{cc|c} 1 & 3 & 1 \\ 0 & 1 & -3 \end{array} \right] \rightarrow \left[ \begin{array}{cc|c} 1 & 0 & 7 \\ 0 & 1 & -2 \end{array} \right]$$

**第二种**是 *Gauss-Jordan's way*，即一次性求解多个多元一次线性方程组，具体实施方法是在 *Gauss* 的基础上进一步拓展增广矩阵  $\hat{B}$ ，并通过彻底的初等行变换将增广矩阵的左半部分化简为单位阵，则增广部分即为逆矩阵：

$$\hat{B} = \left[ \begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 2 & 7 & 0 & 1 \end{array} \right] \rightarrow \left[ \begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{array} \right] \rightarrow \left[ \begin{array}{cc|cc} 1 & 0 & 7 & -3 \\ 0 & 1 & -2 & 1 \end{array} \right]$$

这一点很好理解，首先对于  $B$  以及  $\hat{B}$  的增广部分 ( $I$ ) 而言，二者进行了同步初等行变换，即左乘了同一个变换矩阵 ( $E$ )。  $B$  经过变换后成为了  $I$ ，这意味着  $E = B^T$ ，那自然也有  $RI = E$ ，所以增广部分就会转换成逆矩阵。即：

$$B^{-1} = \begin{bmatrix} 7 & -3 \\ -2 & 1 \end{bmatrix}$$

## 2.6 Elimination = Factorization

### 2.6.1 LU分解的概念

以三阶方阵  $A \in \mathbb{R}^{3 \times 3}$  为例，假设线性方程组  $A\mathbf{x} = \mathbf{b}$  为一个具有唯一解的三元一次方程组，且方程顺序排列合适（无需进行行交换），则可以通过至多**三次**初等行变换，将系数矩阵  $A$  转变成一个上三角矩阵  $U$  (*Upper triangle matrix*)，其对角线元素称为“主元” (*pivot*)：

$$E_{32}E_{31}E_{21}A = U \implies \hat{E}A = U \quad (2-6-1)$$

显然， $\hat{E}$  主对角线上的元素均为 1。稍作变换即可得到  $A$  的 LU 分解：

$$A = \hat{E}^{-1}U = LU \quad (2-6-2)$$

其中  $L$  是一个下三角矩阵 (*Lower triangle matrix*)。值得注意的是， $U$  不一定是对角阵，但是 LU 分解的结果已经非常接近对角化了。我们只需对  $U$  再进行至多三次初等列变换，即可将其彻底转化为一个对角阵  $D$

(*Diagonal matrix*)，同时列变换操作矩阵的逆将是一个上三角矩阵。因此 LU 分解可以进一步变成 LDU 分解  $A = LDU$ 。

说回到 LU 分解，我们自然而然地会产生一个问题：为什么我们倾向于展示  $A = LU$ ，而不是  $\hat{E}A = U$ ？这里首先给出一个结论：

For  $A = LU$ , if there are no row exchanges, those numbers that we multiplied rows by and subtracted during an elimination step go directly into  $L$ .

接下来我们以一个 4 阶方阵为例，先假设出一些行变换矩阵：

$$\hat{E} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -4 & 1 \end{bmatrix}}_{E_{to3}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix}}_{E_{to2}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ -3 & 0 & 0 & 1 \end{bmatrix}}_{E_{to1}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -3 & 2 & 1 & 0 \\ 3 & -5 & -4 & 1 \end{bmatrix}$$

我们其实不需要严格执行两次矩阵乘法才能获得  $\hat{E}$ 。以第一列为例：

(1)  $\hat{E}(0,0) = 1$  是毫无疑问的（我习惯使用 *pythonic* 风格记录数据下标），之前也有提到对角线上元素都应该是 1；

(2)  $\hat{E}(1,0) = E_{to1}(1,0) = -2$ ；

(3)  $\hat{E}(2,0) = E_{to1}(2,0) + E_{to2}(2,1)\hat{E}(1,0) = -3$ ；

(4)  $\hat{E}(3,0) = E_{to1}(3,0) + E_{to2}(3,1)\hat{E}(1,0) + E_{to3}(3,2)\hat{E}(2,0) = 3$ 。

同理可依次获得第二列、第三列下三角部分的元素值。

不论如何，在  $E_{to3}$ 、 $E_{to2}$  以及  $E_{to1}$  的基础上想获取  $\hat{E}$ ，多多少少需要一些运算。当数字并非示例中如此友好时，这个过程会变得非常痛苦。而如果我们尝试写出  $L$ ，则无需任何运算，甚至比你敲完这段代码的时间还要快得多：

```
import numpy as np
from numpy import linalg as nLA
```

```

E3 = np.array([[1,0,0,0],[0,1,0,0],[0,0,1,0],[0,0,-4,1]])
E2 = np.array([[1,0,0,0],[0,1,0,0],[0,2,1,0],[0,3,0,1]])
E1 = np.array([[1,0,0,0],[-2,1,0,0],[1,0,1,0],[-3,0,0,1]])
hatE = E3 @ E2 @ E1
L = nLA.inv(hatE)

```

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

回顾上面那段英文结论，结合这个示例，我们不难发现：

- (1)  $L(1, 0) = -E_{to1}(1, 0)$ ;
- (2)  $L(2, 0) = -E_{to1}(2, 0)$ ,  $L(2, 1) = -E_{to2}(2, 1)$ ;
- (3)  $L(3, 0) = -E_{to1}(3, 0)$ ,  $L(3, 1) = -E_{to2}(3, 1)$ ,  $L(3, 2) = -E_{to3}(3, 2)$ 。

综上所述，在已知所需的行变换操作与  $U$  时， $A = LU$  的效率远高于  $\hat{E}A = U$ 。这是因为我们可以在进行初等行变换的过程中，同步计算变换结果  $U$ 、轻松获取  $L$  的元素。

## 2.6.2 方阵消元的时间复杂度

我们把单个元素的一次乘法与加法设为基本操作单元（除法等同于乘法，不乘等同于乘 1），在无需行交换的情况下，将  $A \in \mathbb{R}^{n \times n}$  通过初等行变换进行 LU 分解所需的操作次数为：

$$\begin{aligned}
 O(n) &= O(E_{to1}) + O(E_{to2}) + \cdots + O(E_{to(n-1)}) \\
 &= n(n-1) + (n-1)(n-2) + \cdots + 2 \times 1 = \sum_{i=2}^n i^2 - \sum_{i=2}^n i = \sum_{i=1}^n i^2 - \sum_{i=1}^n i \\
 &= \frac{i(i+1)(2i+1)}{6} - \frac{(i+1)i}{2} = \frac{n^3 - n}{3} \approx \frac{1}{3}n^3
 \end{aligned} \tag{2-6-3}$$

以防万一，这里给出一些自然数  $n$  项  $p$  次方的求和公式。需要注意的是，我们可以用微积分的思路来估计级数的和，但是这个估计是有偏差的：



$$\left\{ \begin{aligned} \sum_{i=1}^n i &= C_{n+1}^2 = \frac{n(n+1)}{2} \\ \sum_{i=1}^n i^2 &= C_{n+1}^2 + 2C_{n+1}^3 = \frac{n(n+1)(2n+1)}{6} \\ \sum_{i=1}^n i^3 &= C_{n+1}^2 + 6C_{n+1}^3 + 6C_{n+1}^4 = \frac{n^2(n+1)^2}{4} \\ \sum_{i=1}^n i^4 &= C_{n+1}^2 + 14C_{n+1}^3 + 36C_{n+1}^4 + 24C_{n+1}^5 = \frac{n(n+1)(6n^3+9n^2+n-1)}{30} \\ \sum_{i=1}^n i^5 &= C_{n+1}^2 + 30C_{n+1}^3 + 150C_{n+1}^4 + 240C_{n+1}^5 + 120C_{n+1}^6 = \frac{n^2(n+1)(2n^3+4n^2+n-1)}{12} \end{aligned} \right. \quad (2-6-4)$$

数学真是有意思，以后没准还会开个微积分的专栏（听说是 MIT-18.01？回头看看能不能搜到视频）。言归正传，由于矩阵消元通常发生在线性方程组求解过程（ $A\mathbf{x}^T = \mathbf{b}^T$ ）中，即我们还有一个系数列向量  $\mathbf{b}^T$  需要同步进行行变换以及变量代回处理。在这个阶段所需的运算次数（包含在  $\mathbf{b}^T$  以及  $\mathbf{U}$  上的运算，后者的运算仅限于计算乘法所需的系数）约为：

$$(n-1) + 2 \times (1 + 2 + 3 + \cdots + n) = n^2 + 2n - 1 \approx n^2 \quad (2-6-4)$$

## 2.7 Transposes and Permutations: $A = LU$

之前提到，对矩阵进行初等行变换，等价于左乘一个变换矩阵。当时我们考虑的情况都是不包含行交换的，其实行交换也等价于一次矩阵乘法，即置换（Permutation）矩阵。置换矩阵的本质是**行重新排列的单位矩阵**。对于  $n$  阶方阵，其对应的置换矩阵共有  $n!$  种。以三阶方阵为例，三阶置换矩阵可以表示为 6 个矩阵组成的群：

$$\begin{aligned} P_{123} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, P_{132} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, P_{312} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ P_{321} &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, P_{231} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, P_{213} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

之所以要划定矩阵群的概念，是因为**群内所有矩阵的乘积、逆矩阵依然在群内**，这是一个非常有趣的特性。除此之外，**置换矩阵的逆还等于该矩阵的转置**，即：

$$P^T = P^{-1}, PP^T = P^T P = I \quad (2-7-1)$$

当  $A$  并非我们先前假设的那么“完美”时，通过置换矩阵预先处理  $A$ ，将主元位置调整合适之后，可以继续继续进行 LU 分解：

$$PA = LU \implies A = P^T LU \quad (2-7-2)$$

## 3 Vector Spaces and Subspaces

### 3.1 Spaces of Vectors

#### 3.1.1 线性空间与子空间

在我国高等教育体系中（非数学系），关于这一部分的数学知识通常应用泛函分析这门课程里才会得到真正重视，此时大伙多半已经是研究生了。而年级越高，讲课的、听课的往往都越混，尤以前者为甚。再度惋惜自己没能在初见线代的时候学习 MIT-18.06，后边用得多了发现线代根本没有考试时那么可怕。

对于任意一个线性空间  $\mathbf{M}$ ，属于该空间的元素  $m_1, m_2$  等均需满足对数乘与加法运算的封闭性，即有  $a \times m_1 + b \times m_2 \in \mathbf{M}$  ( $\forall a, b \in \mathbb{R}$ )，且空间内必须存在一个零元。具体到加法、数乘的方式，可以脱离常规定义，只要满足封闭性即可。同理，零元的定义也与数字 0 有所区别。具体定义在此就不详述了，感兴趣的可以参考泛函的相关知识。

上述结论同样适用于线性空间的线性子空间。这里我们需要研究一下常见空间的线性子空间，以  $\mathbb{R}^2$  为例。以下三类都是  $\mathbb{R}^2$  的线性子空间：

- (1)  $\mathbb{R}^2$  自己。换句话说任意线性空间都是自身的子空间；
- (2) 过原点的直线。这里需要区分的是， $\mathbb{R}$  虽然也是一种直线空间，但是它并不是  $\mathbb{R}^2$  的子空间。显然我们只需一个维度的标量即可定义  $\mathbb{R}$  中的元素，而描述任意  $\mathbb{R}^2$  中的元素（即使是原点）都需要两个维度的标量。
- (3)  $\mathbf{0}$ 。即  $\mathbb{R}^2$  的零元。任意线性空间的零元可单独构成该空间的线性子空间。

对于  $\mathbb{R}^3$  呢？我们很容易想到两个极端：(1)  $\mathbb{R}^3$  本身；(2)  $\mathbf{0}$ 。结合  $\mathbb{R}^2$  的经验我们知道还有 (3) 任意穿过原点的平面；以及最后 (4) 穿过原点的直线。

#### 3.1.2 如何构成子空间

当谈论到子空间时，我们的基础语境是已经存在了一个线性空间作为“母空间”。这里给出一种面向一般情况的构建方法，以  $\mathbb{R}^3$  为例。假设  $P, l$  分别是  $\mathbb{R}^3$  中经过原点的某平面、某直线，显然后两者都是前者的子空间。现在我们考虑集合运算“并” ( $\cup$ ) 与“交” ( $\cap$ )，即  $P \cup l$  与  $P \cap l$ ：

(1)  $P \cup l$  不是子空间。因为在  $P$  上取某向量  $p$  与  $l$  中的某向量相加，其结果很可能既不在平面  $P$  上，也不在直线  $l$  上，因此该空间对加法运算不封闭，所以不是  $\mathbb{R}^3$  的子空间。

(2)  $P \cap l$  是子空间。事实上  $P \cap l = \mathbf{0}$ 。

更一般地，对于同属某一线性空间的两个子空间  $S$  与  $T$ ， $S \cap T$  仍然是子空间，但  $S \cup T$  通常不满足子空间条件。简单证明一下， $\forall v, w \in S \cap T$ ，均有：

$$(av + bw) \in S, (av + bw) \in T \implies (av + bw) \in S \cap T, \forall a, b \in \mathbb{R} \quad (5-2-1)$$

除了上述抽象方法，我们当然可以从一个具体的矩阵入手来构造一系列子空间。这些内容将在下一节详细展开。

## 3.2 The Nullspace of $A$ : Solving $Ax = 0$

### 3.2.1 列空间与零空间

以矩阵  $X \in \mathbb{R}^{3 \times 2}$  为例。 $X$  的各列都属于  $\mathbb{R}^3$ ，因此列向量  $X(:, i)$  的全体线性组合可以构成一个  $\mathbb{R}^3$  的线性子空间，称为  $X$  的**列空间**，记为  $C(X)$ 。从欧氏几何的角度来看， $C(X)$  是一个过原点的平面（若列向量不共线）， $X$  的两个列向量位于平面上。列空间有什么作用呢？我们通过一个实例来说明。

给定系数矩阵  $A$ （这个例子将会持续存在一段时间）以及线性方程组  $Ax = b$ ：

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \in \mathbb{R}^{4 \times 3}$$

这种方程数量超过未知数数量的情况称为“**超定**”（*over-determined*）系统。超定系统在大多数情况下是没有精确解的。从感性认识的角度考虑，如果我们通过正定系统获得一个精确解后，再额外添加一个或多个方程，则增添方程必须满足**某些特定条件**才能使得等式依旧成立。相应地，并非所有的系数向量都能使得方程组有解。接下来我们将从  $b$  入手，深入研究  $Ax = b$  有解的条件。

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

显然  $x = b = 0$  对于任意数值、任意维度的系数矩阵  $A$  都成立， $0$  通常称为“**零解**”或“**平凡解**”。除此之外，结合矩阵的列视图，我们还能很快地举出一些其它例子：

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

以上这个拆解形式很容易让人联想到矩阵的列空间  $C(A)$ ，因此有以下结论：当且仅当  $b \in C(A)$  时，方程组  $Ax = b$  确定有解。

说完了列空间，我们再来谈谈零空间。还是以  $A$  为例，满足  $Ax = 0$  的全体  $x$  构成的空间  $N(A)$  称为  $A$  的**零空间**。注意区分，对于  $Ax = b$ ， $C(A)$  关心的是  $b$ ，而  $N(A)$  关心的是  $x$ ；对于  $A \in \mathbb{R}^{m \times n}$ ， $C(A) \in \mathbb{R}^m$  而  $N(A) \in \mathbb{R}^n$ 。

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

显然， $\mathbf{0}$  再一次满足  $N(\mathbf{A})$  的要求。事实上  $[1, 1, -1]^T$  的所有线性组合都满足要求。可见此时  $N(\mathbf{A})$  表现为  $\mathbb{R}^3$  中的一条直线。

需要指出，当  $\mathbf{b} \neq \mathbf{0}$  时，同样可能存在很多  $\mathbf{x}$  使得方程等式平衡。但是这些  $\mathbf{x}$  无法构成线性向量空间。很简单，因为其中一定没有零元  $\mathbf{0}$ 。所以当我们站在线性方程组的角度上回顾子空间的概念时应当明确：

(1) 根据一个给出的矩阵  $\mathbf{X} \in \mathbb{R}^{m \times n}$ ，可以从中构建两种线性空间：列空间  $C(\mathbf{X}) \in \mathbb{R}^m$  与零空间  $N(\mathbf{X}) \in \mathbb{R}^n$ ，而不论哪种空间都必须满足“零元存在”；

(2) 列空间的构建方法是对已有列向量进行线性组合，零空间的构建方法是在矩阵基础上寻找满足条件的解集。二者都是非常重要的子空间构建方法。

### 3.2.2 零空间求解算法

以矩阵  $\mathbf{A}$  为例：

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix}$$

我们来研究求解  $\mathbf{Ax} = \mathbf{0}$  的过程。由于右侧向量已经是  $\mathbf{0}$  了，是否对  $\mathbf{A}$  增广没有区别，因此我们可以直接对  $\mathbf{A}$  进行 LU 分解：

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

在这里先停一下。可以看出：(1)  $\mathbf{U}$  的主元有 2 个，称该矩阵的秩 (*rank*) 为 2，即存在 **2 个有效方程**；(2) **未知数有 4 个**，而根据数学常识，两个方程只能求解 2 个未知数，因此  $\mathbf{Ax} = \mathbf{0}$  不受约束的自由变量还剩  $4-2=2$  个。

关于自由变量，我们可以从更具体的角度理解。比如本例中  $\mathbf{U}(1, :)$  的对应方程为  $2x_3 + 4x_4 = 0$ ，当  $x_3$  固定时， $x_4$  也随之固定，因此两者只有一个为“自由变量”；紧接着看  $\mathbf{U}(0, :)$ ，对应方程为  $x_1 + 2x_2 + 2x_3 + 2x_4 = 0$ ，由于  $x_3, x_4$  均已固定，因此又进入了  $x_1, x_2$  二选一的情况，也是只剩一个自由变量，因此该方程组共有 2 个自由变量。**自由变量的数量即为特解的数量**。这一点从反面很好理解：当方程组正定且有解时，系数矩阵满秩，此时没有自由变量，也就没有所谓的特解。

由于  $\mathbf{U}$  的主元分别对应于  $x_1$  与  $x_3$ ，我们不妨将剩下的  $x_2$  与  $x_4$  定为自由变量（这样做能与自由变量的概念相对应，避免出错），并设为 1、0 或 0、1。对应的特解  $\mathbf{x}_1, \mathbf{x}_2$  以及方程组通解  $\mathbf{x}$  为：

$$\mathbf{x}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix} \implies \mathbf{x} = a\mathbf{x}_1 + b\mathbf{x}_2, a, b \in \mathbb{R}$$

### 3.3 The Rank and the Row Reduced Form

#### 3.3.1 行简化形式

行简化矩阵  $\mathbf{R}$  在上三角矩阵  $\mathbf{U}$  的基础上进一步化简，使得 (1) 所有主元均为 1；(2) 主元上下位置的元素均为 0。依旧以 3.2.2 的矩阵  $\mathbf{A}$  为例：

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{U} = \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

首先需要指出， $\mathbf{Ax} = \mathbf{0}$ 、 $\mathbf{Ux} = \mathbf{0}$  以及  $\mathbf{Rx} = \mathbf{0}$  这三个方程的解是相同的，因此初等行变换并不会导致线性方程组的解产生变化。换句话说， $\mathbf{A}$ 、 $\mathbf{U}$  以及  $\mathbf{R}$  的零空间是相同的。其次，通过上述三者获取解的难度显然是逐次降低的。最后，为了说明行简化矩阵求解的优越性，请注意观察  $\mathbf{R}$  以及  $\mathbf{x}_1$ 、 $\mathbf{x}_2$ ，是否发现其中的元素高度相似？自由变量位置分别为 0 或 1，而主元位置是自由变量系数的相反数？

这个“奇技淫巧”的原理如下：一个理想状态下的  $\mathbf{R}_i^{m \times n}$  可以表示为分块矩阵，其中  $\mathbf{I}$  表示主元矩阵（单位阵）， $\mathbf{F}$  表示自由变量矩阵：

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{I}_{r \times r} & \mathbf{F}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{bmatrix} \quad (3-3-1)$$

不失一般性地，我令  $\mathbf{F}$  为一个非方阵，它的底部甚至可以是一些纯 0 行，只要保证  $\mathbf{R}_i$  与  $\mathbf{F}$  的行数相等即可。接下来我们考察方程  $\mathbf{R}_i \mathbf{x} = \mathbf{0}$  的解。由于方程是欠定的，因此存在一些自由变量和秩次数个特解。换句话说我们可以构建一个零空间矩阵  $\mathbf{N} \in \mathbb{R}^{n \times (n-r)}$ ，使得  $\mathbf{R}_i \mathbf{N} = \mathbf{0} \in \mathbb{R}^{m \times (n-r)}$ ， $\mathbf{N}$  中的每一列构成线性方程组的一个特解。这一点同时照应了上一节提到的“特解数量为自由变量个数”的结论：

$$\mathbf{N} = \begin{bmatrix} -\mathbf{F}_{r \times (n-r)} \\ \mathbf{I}_{(n-r) \times (n-r)} \end{bmatrix} \quad (3-3-2)$$

我们不妨再换个角度来审视这个过程， $\mathbf{R}_i \mathbf{x} = \mathbf{0}$  可以展开为：

$$\begin{bmatrix} \mathbf{I} & \mathbf{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{pivot} \\ \mathbf{x}_{free} \end{bmatrix} = \mathbf{0} \quad (3-3-3)$$

显然该等式成立的条件是：主元变量为自由变量的相反数，自由变量构成单位阵（即挨个等于 1，剩下为 0）。

上述步骤说明了行简化矩阵与特解之间的对应关系，但前提是理想情况。对于需要列交换的情况还成立吗？显然成立，只不过我们需要从一个更具体的视角入手。以本节初给出的  $\mathbf{R}$  为例，将非主元的非零元素用字母替代，方便我们看清本质：

$$\mathbf{R} = \begin{bmatrix} 1 & r_1 & 0 & r_2 \\ 0 & 0 & 1 & r_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

自由变量分别为第 2、第 4 变量，因此提前写出两个特解：

$$\mathbf{x}_1 = \begin{bmatrix} ? \\ 1 \\ ? \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} ? \\ 0 \\ ? \\ 1 \end{bmatrix}$$

第一自由变量所在列为  $\mathbf{R}(:, 1)$ ，且仅  $\mathbf{R}(0, 1)$  不为 0，则在特解  $\mathbf{x}_1$  中由上至下依次填入  $-r_1$  与 0；第二自由变量所在列为  $\mathbf{R}(:, 3)$ ，在特解  $\mathbf{x}_2$  中由上至下依次填入  $-r_2$ 、 $-r_3$ 。以由于我们已经把数值变量化了，因此通过表达式可以轻松证明以上操作步骤的成立。

### 3.3.2 矩阵的秩

矩阵的形状可以千奇百怪，但是其中包含的有效信息往往并不像表面看上去那么丰富，真正决定有效信息量的参数是秩。对于矩阵  $\mathbf{M} \in \mathbb{R}^{m \times n}$ ， $r(\mathbf{M})$  表示主元的个数，它是一个不大于  $m$  或  $n$  的正整数。 $r = m$ 、 $r = n$  分别对应于“行满秩”与“列满秩”两种情况：

(1) 行满秩：把  $\mathbf{M}$  化简为  $\mathbf{R}$  后，每一行都有一个主元。即**不存在全零行**。仅适用于  $m \leq n$  的情况（正定或欠定，矩阵形如横着的长方形）：

(2) 列满秩： $\mathbf{R}$  的每一列均有一个主元。即**不存在自由变量**。仅适用于  $m \geq n$  的情况（正定或超定，矩阵形如竖着的长方形），事实上  $\mathbf{R}$  的有效部分为单位阵。

(3) **矩阵的秩次数不随矩阵的转置而变化**。即  $r(\mathbf{M}) = r(\mathbf{M}^T)$ 。

目前想要证明结论 (3) 有些困难，我们当然可以用字母表达式的方式在  $\mathbb{R}^{3 \times 3}$ 、 $\mathbb{R}^{3 \times 4}$  等矩阵上进行验证，无非是表达式比较冗长罢了。未来我们一定会用更成熟的理论来解释这个问题。

## 3.4 The Complete Solution to $\mathbf{Ax}=\mathbf{b}$

### 3.4.1 可解性与解的结构

先前我们讨论了线性方程组  $\mathbf{Ax} = \mathbf{b}$  有解的条件，即  $\mathbf{b} \in \mathbf{C}(\mathbf{A})$ ，之后以  $\mathbf{Ax} = \mathbf{0}$  为例讲解了齐次线性方程组的解法。现在来看看非齐次方程组的求解流程。例如：

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}$$

(1) 通过 LU 分解化简增广系数矩阵  $\tilde{\mathbf{A}}$ :

$$\mathbf{E}_{to2}\mathbf{E}_{to1}\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{U}$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \left[ \begin{array}{cccc|c} 1 & 2 & 2 & 2 & 1 \\ 0 & 0 & 2 & 4 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

(2) 将所有自由变量 (本例中为  $x_2$ 、 $x_4$ ) 设为 0, 求出特解  $\mathbf{x}_p$ :

$$\begin{cases} x_1 + 2x_3 = 1 \\ 2x_3 = 3 \end{cases} \implies \mathbf{x}_p = \begin{bmatrix} -2 \\ 0 \\ 1.5 \\ 0 \end{bmatrix}$$

(3) 求出  $\mathbf{A}$  的零空间  $\mathbf{N}(\mathbf{A})$  作为通解  $\mathbf{x}_n$ :

$$\mathbf{Ax} = \mathbf{0} \implies \mathbf{Ux} = \mathbf{0} \implies \mathbf{Rx} = \mathbf{0}, \mathbf{R} = \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{x}_n = k_1 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}, k_1, k_2 \in \mathbb{R}$$

(4) 合并通解与特解, 得到全解  $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$ :

$$\mathbf{x} = \begin{bmatrix} -2 \\ 0 \\ 1.5 \\ 0 \end{bmatrix} + k_1 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}, k_1, k_2 \in \mathbb{R}$$

$$\mathbf{A}(\mathbf{x}_p + \mathbf{x}_n) = \mathbf{b} + \mathbf{0} = \mathbf{b} \quad (3-4-1)$$

在本例中, 全解表现为  $\mathbb{R}^4$  空间中的一个二维平面, 注意它并非子空间。我们可以说  $\mathbf{N}(\mathbf{A})$  是  $\mathbb{R}^4$  的一个子空



间，它在欧式几何上表现为过零元的一个平面。而  $\mathbf{x}$  是一个经过特解点  $\mathbf{x}_p$  的平面，相当于将  $N(\mathbf{A})$  进行平移后的结果。因为这个空间内没有零元，所以它不是子空间。

### 3.4.2 矩阵秩次对于解的影响

我们知道对于一个系数矩阵  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ，其秩次数  $r(\mathbf{A})$  可决定矩阵不满秩、行满秩或列满秩。这三种情况分别对应了线性方程组  $\mathbf{Ax} = \mathbf{b}$  的不同解情况。

首先是列满秩 ( $r(\mathbf{A}) = n$ )：结合 3.3.2 节末尾的结论，此时不存在自由变量， $N(\mathbf{M})$  仅有  $\mathbf{0}$  一个向量，即不存在通解  $\mathbf{x}_n$ 。线性方程组  $\mathbf{Ax} = \mathbf{b}$  只有 1 个特解  $\mathbf{x}_p$  或无解 ( $\mathbf{b} \notin C(\mathbf{A})$ )；

其次是行满秩 ( $r(\mathbf{A}) = m$ )：行满秩时行简化矩阵没有全零行。与列满秩不同，此时方程组的可解性对于  $\mathbf{b}$  没有任何要求（不论超定或正定）。自由变量有  $n - m$  个，即通解  $\mathbf{x}_n$  由  $n - m$  项组成，全解要么只有 1 个（特解， $m = n$ ），要么有无数个（特解 + 通解）；

这里有必要强调一种特殊情况：满秩 ( $r(\mathbf{A}) = m = n$ )。此时方程组的系数矩阵是一个可逆的方阵，其行简化矩阵为单位阵。由于它是列满秩，因此  $N(\mathbf{A})$  依然只有  $\mathbf{0}$ 。而它又是行满秩，故  $\mathbf{Ax} = \mathbf{b}$  必有解，且解为  $\mathbf{A}^{-1}\mathbf{b}$ ；

最后谈谈不满秩的情况 ( $r(\mathbf{A}) < m, r(\mathbf{A}) < n$ )：此时行简化矩阵有全零行（当且仅当  $\mathbf{b} \in C(\mathbf{A})$  时有解），也有自由变量（通解有无数个）。即要么无解，要么有无数个解。

## 3.5 Independence, Basis and Dimension

对于欠定线性方程组的系数矩阵  $\mathbf{A} \in \mathbb{R}^{m \times n}$  ( $m < n$ )，根据上一节的相关内容我们知道， $\mathbf{A}$  至多可达行满秩，而不可能列满秩，必然存在至少  $n - m$  个自由变量，所以  $N(\mathbf{A})$  中的元素不止  $\mathbf{0}$  一个，即  $\mathbf{Ax} = \mathbf{0}$  包含非零解。上述分析过程其实对应于矩阵列向量组的一个特性：相关性。

### 3.5.1 线性相关性

向量组线性相关性的定义可以从多个方面进行阐释，注意以下的各种说法本质上都是一样的。对于前言中的矩阵  $\mathbf{A}$ ，当满足以下条件之一时，其列向量组**线性相关**：

- (1) 其列向量可通过某个非零的线性组合构成零向量；
- (2)  $\mathbf{Ax} = \mathbf{0}$  存在非零解；
- (3) 零空间  $N(\mathbf{A})$  内除了零元  $\mathbf{0}$  还有其它元素。

反之，若上述条件之一不满足，则称向量组**线性无关**。

### 3.5.2 生成空间 & 基 & 维数



类似于列空间，我们可以在广义层面上定义一个由向量组  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  张成（生成）的空间  $\mathcal{S}$ ：

$$\forall k_i \in \mathbb{R}, \sum_{i=1}^n k_i \mathbf{v}_i \in \mathcal{S} \quad (3-5-1)$$

对于一个生成空间，其内部自然由无数个向量构成。这些向量可能是相关的，也可能是无关的。我们关心的问题常常是：能否找到一组个数最少的向量，从而足以构成整个空间。换句话说，我们总是希望能够寻找到空间的一组基来描述整个空间的某些特性。基向量一般需满足以下特点：

- (1) 彼此线性无关。例如对于空间  $\mathbf{R}^n$ ，假设有  $n$  个向量构成一组基，则向量组合成的  $n$  阶方阵是可逆矩阵；
- (2) 可生成完整的目标空间
- (3) 基向量的数目应当与空间维数（即向量元素个数）一致。

空间维数这个概念与矩阵的秩也存在关联：矩阵  $\mathbf{A}$  的秩 ( $r$ ) 等于主元列的个数，等于列空间  $\mathcal{C}(\mathbf{A})$  的维数； $\mathbf{A}$  自由变量的个数 ( $n - r$ ) 等于零空间  $\mathcal{N}(\mathbf{A})$  的维数。

不难发现，秩是矩阵的概念，维数是空间的概念，而两者在某种程度上是相互连通的。事实上，矩阵只是空间的一种表现形式而已。

### 3.5.3 矩阵空间、函数空间的基

独立性、基和维数的概念不仅限于向量，从泛函分析的角度，它们还能拓展至矩阵空间、函数空间等各种空间。例如一个包含全体二阶方阵的矩阵空间  $\mathcal{M}$ ，它的基  $\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\}$  以及生成空间  $\text{span}\{\mathbf{A}\}$  分别为：

$$\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\text{span}\{\mathbf{A}\} = \sum_{i=1}^4 k_i \mathbf{A}_i = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}$$

此时，我们考虑的对象不再是单个维度的向量，而是二阶方阵。其中每一个元素的“0/1”状态都是彼此独立的，因此空间  $\mathcal{M}$  的维数为 4，与矩阵元素个数相同。更一般地， $\mathbb{R}^{n \times n}$  的矩阵空间  $\mathcal{M}$  的维数为  $n^2$ 。

注意到  $\mathbf{A}_1$ 、 $\mathbf{A}_2$  以及  $\mathbf{A}_4$  都是上三角矩阵空间  $\mathcal{U}$  ( $\mathcal{M}$  的子空间) 的基，在  $\mathbb{R}^{2 \times 2}$  矩阵空间中， $\mathcal{U}$  的维数显然为 3。进一步地，在  $\mathbb{R}^{n \times n}$  矩阵空间中， $\mathcal{U}$  的维数与矩阵中可变状态的元素个数相等，即：

$$\begin{bmatrix} m_{0,0} & m_{0,1} & \cdots & m_{0,n-1} \\ 0 & m_{1,1} & \cdots & m_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_{n-1,n-1} \end{bmatrix} \implies \frac{1}{2}(n^2 - n) + n = \frac{1}{2}n^2 + \frac{1}{2}n$$

类似地，我们可以推导出其它子空间的维数，例如  $n$  阶对角阵空间  $\mathbf{D}$  的维数为  $n$ 、 $n$  阶对称阵  $\mathbf{S}$  的维数为  $\frac{1}{2}n^2 + \frac{1}{2}n$  等等。

类似向量空间、矩阵空间，函数空间指一群具有特定结构，或满足特殊功能的函数集合。我们来看三个微分特征方程以及它们的解集：

$$\frac{d^2 y}{dx^2} = 0 \implies y = cx + d$$

$$\frac{d^2 y}{dx^2} = -y \implies y = c \sin x + d \cos x$$

$$\frac{d^2 y}{dx^2} = y \implies y = ce^x + de^{-x}$$

对于第一个微分方程，它的解空间有两个基： $x$  与  $1$ ，它们是二阶导数的“零空间”。以此类推，第二个微分方程解空间的基为  $\sin x$  与  $\cos x$ ，第三个为  $e^x$  与  $e^{-x}$ 。这些空间的基不再是向量或者矩阵，而是函数，因而得名“函数空间”。需要注意的是，并非所有微分方程的解都能构成函数空间，正如并非所有向量（矩阵）能够成对应空间一样。例如非齐次微分方程  $y'' = 2$ ，其特解为  $y = x^2$ ，通解为对应齐次微分方程的解  $y = cx + d$ ，全解为  $y = x^2 + cx + d$ ，显然这些函数簇并不满足线性可加性（二次项的系数固定为 1），因此它们无法构成线性子空间。

最后需要提到的一个向量空间是只含有零向量的空间  $\mathbf{Z}$ 。数学上定义该空间的维数为 0，基是空集  $\emptyset$ ，并不是零向量。**零向量永远无法成为某个线性空间的基**，因为在定义上它是**任意方向**的，因此与任何一个非零元基的方向都重合。

## 3.6 Dimensions of the Four Subspaces

### 3.6.1 四种基本子空间的定义

对于一个矩阵  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ，我们可以从中构建四种基本子空间：

- (1) 列空间： $\mathbf{C}(\mathbf{A})$ ， $\mathbb{R}^m$  的子空间；
- (2) 行空间： $\mathbf{C}(\mathbf{A}^T)$ ， $\mathbb{R}^n$  的子空间；
- (3) 零空间： $\mathbf{N}(\mathbf{A})$ ， $\mathbb{R}^n$  的子空间；
- (4) 左零空间： $\mathbf{N}(\mathbf{A}^T)$ ， $\mathbb{R}^m$  的子空间。

### 3.6.2 子空间的基与维数

(1) 对于列空间  $C(\mathbf{A})$ ，根据先前章节的知识，我们知道维数  $\dim[C(\mathbf{A})] = \text{rank}(\mathbf{A}) = r$ ，基即为各主元所在列。

(2) 对于同一矩阵的行空间  $C(\mathbf{A}^T)$ ，其维数不变： $\dim[C(\mathbf{A}^T)] = \text{rank}(\mathbf{A}^T) = \text{rank}(\mathbf{A})$ ，基为  $\mathbf{A}^T$  的各主元所在列。

(3) 对于零空间  $N(\mathbf{A})$ ，基为  $\mathbf{A}\mathbf{x} = \mathbf{0}$  的特解，因此基的个数为  $\mathbf{A}$  的自由变量个数  $n - r$ ，即维数为  $n - r$ 。这里需注意， $N(\mathbf{A})$  的维数加上  $C(\mathbf{A})$  的维数等于  $\mathbf{A}$  的列数  $n$ 。

(4) 最后是左零空间  $N(\mathbf{A}^T)$ ，类似  $N(\mathbf{A})$  我们知道， $N(\mathbf{A}^T)$  的维数应当是  $m - r$ ，基为  $\mathbf{A}^T \mathbf{x} = \mathbf{0}$  的特解。

### 3.6.3 求解子空间的基

在现有阶段，初等行变换（LU 分解）始终是我们分析矩阵性质的重要技术手段。

(1) 对于列空间  $C(\mathbf{A})$ ，我们可以将其转置后进行初等行变换化简为行简化矩阵，通过转置的行空间间接获得原矩阵的列空间。

(2) 对于行空间  $C(\mathbf{A}^T)$ ，我们需要在行向量中寻找独立的向量组。对  $\mathbf{A}$  进行初等行变换得到行简化矩阵  $\mathbf{R}$ ，主元所在行即为  $C(\mathbf{A}^T)$  的基。一般来说， $C(\mathbf{R}) \neq C(\mathbf{A})$ ，但是  $\mathbf{A}$  与  $\mathbf{R}$  的行空间是相同的。因为初等行变换的过程本质上就是一种线性组合，因此对  $\mathbf{R}$  进行逆向变换（线性组合）可以完全恢复至  $\mathbf{A}$ 。

(3) 对于零空间  $N(\mathbf{A})$ ，根据已有知识，我们通常对  $\mathbf{A}$  进行初等行变换（或 LU 分解）获得行简化矩阵  $\mathbf{R}$ ，以此来确定主元所在列、自由变量以及特解（基）。

(4) 对于左零空间  $N(\mathbf{A}^T)$ ，根据定义有  $\mathbf{A}^T \mathbf{y} = \mathbf{0}$ ，即  $\mathbf{y}^T \mathbf{A} = \mathbf{0}$ ，这个结果很像我推导空间滤波器的形式。这里我们还是先用列向量形式。以如下所示的矩阵  $\mathbf{A}$  为例：

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

通过 *Guass-Jordan* 消元法，可以同步获取行简化矩阵  $\mathbf{R}$  与左乘操作矩阵  $\mathbf{E}$ （当矩阵为方阵时， $\mathbf{E} = \mathbf{A}^{-1}$ ）：

$$\mathbf{E}_{m \times m} [\mathbf{A}_{m \times n} \quad \mathbf{I}_{m \times m}] = [\mathbf{R}_{m \times n} \quad \mathbf{E}_{m \times m}]$$

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

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

在本例中，事实上我们已经找到了一个  $N(A^T)$  的基，即行向量  $E(-1, :)$ 。需要明确的概念是， $N(A)$  寻找的是使得  $A$  各列组合为  $\mathbf{0}$  的向量，而  $N(A^T)$  寻找的是使得  $A$  各行组合为  $\mathbf{0}$  的向量，即对应于  $E$ 。当然我们不太可能总是从  $E$  中能够直接找到答案，因此将原矩阵转置后求解零空间依然不失为一种合理且靠谱的手段。

## 3.7 Matrix Spaces & Rank One Matrices

### 3.7.1 矩阵空间的基

以一个三阶方阵空间  $M \in \mathbb{R}^{3 \times 3}$  (9 维) 为例，上三角矩阵空间  $U$  (6 维)、对称矩阵空间  $S$  (6 维) 都是  $M$  的子空间。显然  $U$  的基是  $M$  的基的一部分，但  $S$  并非如此。通常对于子空间，我们需要重新寻找它的基。

在此之前，我们先来回顾一下如何在已有空间的基础上构建子空间。以对角阵空间  $D$  (3 维) 为例：

$$D = S \cap U \quad (3-7-1)$$

(子) 空间的交集一般仍是子空间，然而 (子) 空间的并集往往不再是线性空间。从广义的空间角度来看，子空间是原始空间的一个“切片”（例如 3 维空间中过零点的平面），子空间的交集是“切片”的“切片”（例如两个相交平面的过零点交线），因此它能够保留形成空间所需的特性；子空间的并集是两个“切片”的简单加和，例如 3 维空间中的两条相交直线，这个空间还有大量空白的部分需要“填补”才能形成一个完备的线性空间。

除了交集以外，空间加和也是一种简单直观的线性空间构建方法：

$$S + U = \{k_1 s + k_2 u\}, \forall k_1, k_2 \in \mathbb{R} \quad (3-7-2)$$

注意  $S + U$  的范围 (9 维) 已经达到了  $M$ ，即它包含了任意 3 阶方阵。一般地：

$$\dim(S) + \dim(U) = \dim(S \cap U) + \dim(S + U) \quad (3-7-3)$$

### 3.7.2 秩 1 矩阵

以  $A$  为例：

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 8 & 10 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}_u \underbrace{\begin{bmatrix} 1 & 4 & 5 \end{bmatrix}}_{v^T}$$

这是一个典型的秩 1 矩阵，即  $\dim(C(A)) = \text{rank}(A) = \dim(C(A^T)) = 1$ 。任意一个秩 1 矩阵  $A$  都能表示为列向量与行向量的乘积  $uv^T$ 。类似乐高积木，秩 1 矩阵可视为其它矩阵的基本构成单位：一个秩为  $r$  的矩阵可由  $r$  个秩 1 矩阵线性组合而成。也正因此，秩 1 矩阵集合并不是一个子空间，因为它不满足可加性。

接下来是另外一个案例，在  $\mathbb{R}^4$  中，对于一个长度为 4 的列向量  $\mathbf{v}$ ，寻找使得各分量之和为 0 的所有向量构成的集合  $\mathbf{S}$ ：

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}, \mathbf{S} = \{\mathbf{v} \mid \sum_{i=1}^4 v_i = 0\}$$

$\mathbf{S}$  是  $\mathbb{R}^4$  的一个线性子空间。尽管它由四维向量组成，但是它的维度是 3。后者似乎不如前者那么明显，我们可以从零空间的角度来看  $\mathbf{S}$ ：

$$\mathbf{S} = \mathbf{N}(\mathbf{A}), \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \implies \mathbf{A}\mathbf{v} = \mathbf{0}$$

$\mathbf{A}$  是一个秩 1 矩阵，自由变量个数为 3，因此  $\mathbf{N}(\mathbf{A})$ （即  $\mathbf{S}$ ）是 3 维空间，它的基为：

$$\mathbf{s}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{s}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{s}_3 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

同理我们可以得知， $\mathbf{C}(\mathbf{A})$  的维度是 1， $\mathbf{C}(\mathbf{A}^T)$  的维度是 1， $\mathbf{N}(\mathbf{A}^T)$  的维度是 0。结合 3.6.2 的知识不难发现以下规律：

$$\forall \mathbf{A} \in \mathbb{R}^{m \times n}, \begin{cases} \dim[\mathbf{C}(\mathbf{A})] + \dim[\mathbf{N}(\mathbf{A})] = n \\ \dim[\mathbf{C}(\mathbf{A}^T)] + \dim[\mathbf{N}(\mathbf{A}^T)] = m \end{cases} \quad (3-7-4)$$

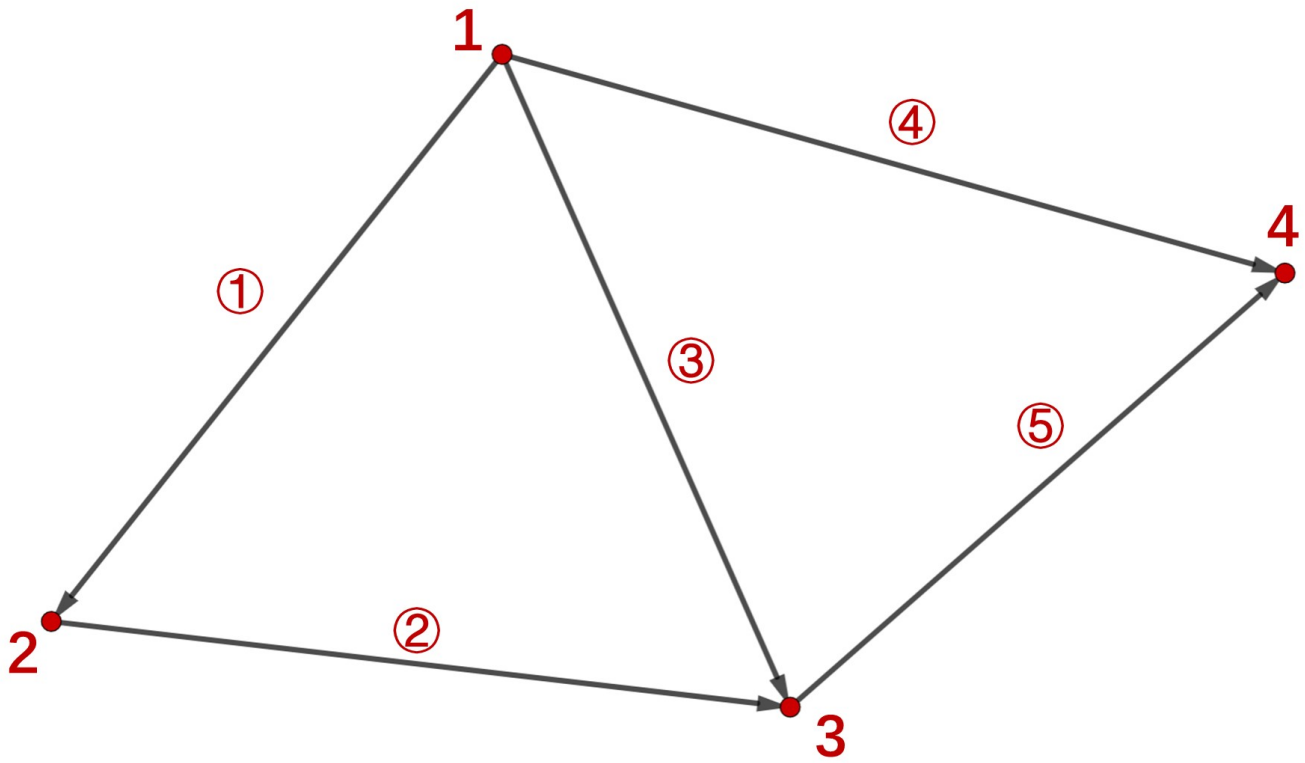
## 3.8 Graphs & Networks

### 3.8.1 图与矩阵

首先需要简单介绍一下“图”的概念：图是“边”与“结点”的集合，“边”连通各个“结点”。我们可以用一个关联矩阵 (*Incidence matrix*)  $\mathbf{A}$  来描述图：

$$\text{Graph} = \{\text{edges}(m), \text{nodes}(n)\} \implies \mathbf{A} \in \mathbb{R}^{m \times n} \quad (3-8-1)$$

以下图所示的有向网络为例，接下来的各种分析都将基于类似的有向图展开。该网络由 4 个结点、5 条边构成：



关联矩阵的每一行表示一条边，每一列代表各个结点在不同边上的流动情况，流起点的值为 -1，流终点的值为 1。若该结点不在对应边上，则数值为 0。据此可以得出图示网络的关联矩阵  $\mathbf{A} \in \mathbb{R}^{5 \times 4}$ ：

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

此图存在两个回路（内部无回路的最小结构），观察对应的关联矩阵：

$$\mathbf{A}_{123} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{A}_{345} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

不难发现，**回路关联矩阵的行向量组（边）是线性相关的**，而且线性组合的方式与信息流向是完全一致的。例如回路  $\mathbf{A}_{123}$ ，信息从结点 1 经过 结点 2 流向结点 3，另一条通路是直接从结点 1 至结点 3。按照向量的视角，应有  $\mathbf{v}_{12} + \mathbf{v}_{23} = \mathbf{v}_{13}$ ，对应地  $\mathbf{A}_{123}(0, :) + \mathbf{A}_{123}(1, :) = \mathbf{A}_{123}(2, :)$ ；同理在回路  $\mathbf{A}_{345}$  中，有  $\mathbf{A}_{345}(0, :) + \mathbf{A}_{345}(1, :) = \mathbf{A}_{345}(2, :)$ 。相应地，如果图中没有回路，这种图称为“树”，**树的各行是线性无关的**。

接下来我们假设该图是一个电路网络，以此来窥探应用数学的一斑。结合 3.6 的知识，我们将通过子空间分析关联矩阵的特征对应于现实场景的物理意义。

### 3.8.2 电路网络分析：零空间

首先是零空间  $\mathbf{N}(\mathbf{A})$ ，对应线性方程组为：

$$\mathbf{Ax} = \mathbf{0} \implies \begin{cases} x_2 - x_1 = 0 \\ x_3 - x_2 = 0 \\ x_3 - x_1 = 0 \\ x_4 - x_1 = 0 \\ x_4 - x_3 = 0 \end{cases}$$

其中  $\{x_1, x_2, x_3, x_4\}$  在电路中表示**结点电势**， $\mathbf{Ax}$  表示各边上的**电势差**。由于该图是一个无源电路，因此电势差应当为  $\mathbf{0}$ 。落实到具体方程的解，显然全零电势 ( $x_1 = x_2 = x_3 = x_4 = 0$ ) 与全等电势 ( $x_1 = x_2 = x_3 = x_4 = 1$ ) 均属于  $N(\mathbf{A})$ ，更进一步地：

$$\mathbf{x} = k \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, k \in \mathbb{R}$$

即有  $\dim[N(\mathbf{A})] = 1$ ，根据 (3-7-4) 可知， $\text{rank}(\mathbf{A}) = n - 1 = 3$ 。事实上， $\forall \mathbf{A} \in \mathbb{R}^{m \times n}$ ，均有  $\dim[N(\mathbf{A})] = 1$ ， $\text{rank}(\mathbf{A}) = n - 1$ 。这一结论很好证明，对于有向图  $\mathbf{A}$  中的一条有向线段  $\vec{l}$  (排序为  $l$ )，其两端端点分别为  $I$ 、 $J$  (排序为  $i$ 、 $j$ )：

(1) 求解零空间的过程中有方程  $\mathbf{A}(l, i) - \mathbf{A}(l, j) = 0$ ，等电势条件  $\mathbf{A}(l, i) = \mathbf{A}(l, j) = 1$  显然满足要求。而图中的任意线段都可以表示为类似形式，因此全等电势必然是  $\mathbf{Ax} = \mathbf{0}$  的特解，即  $N(\mathbf{A})$  的一个基；

(2)  $\mathbf{A}$  的每一行都代表一条边，相应的方程对应一组等（零）电势解。图网络由边连接而成，各结点之间均存在关联，差别仅在于距离远近，即不存在孤立结点。因此也不可能出现部分结点电势为 0，而另一部分结点电势为 1 的情况。

综上所述， $N(\mathbf{A})$  的基有且仅有全等电势一个，因此关联矩阵零空间的维度始终为 1，即：**关联矩阵的秩始终等于图中点数减一**。

上述分析面向的对象都是无源电路。当问题延申至有源（电压源）场景时，电势差  $\mathbf{Ax}$  依然有其用武之地。我们只需构建一个描述了各边电源分布及正负极向信息的电压源矩阵  $\mathbf{e}$ ，并将系数矩阵  $\mathbf{0}$  改为  $\mathbf{e}$  即可。形如  $\mathbf{e} = \mathbf{Ax}$  的问题对应的电路基本定理为**基尔霍夫电压定律** (Kirchoff's Voltage Law, KVL)，即“**任何一个闭合回路中，各元件上的电压降的代数和等于电动势的代数和**”。

### 3.8.3 电路网络分析：左零空间

接下来是左零空间  $N(\mathbf{A}^T)$ ，对应的线性方程组为：



$$\mathbf{A}^T = \begin{bmatrix} -1 & 0 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\mathbf{A}^T \mathbf{y} = \mathbf{0} \implies \begin{cases} -y_1 - y_3 - y_4 = 0 \\ y_1 - y_2 = 0 \\ y_2 + y_3 - y_5 = 0 \\ y_4 + y_5 = 0 \end{cases}$$

既然  $\mathbf{A}\mathbf{x}$  中的  $\mathbf{x}$  表示**电势**，那么  $\mathbf{A}^T \mathbf{y}$  中的  $\mathbf{y}$  呢？当然是与它不离不弃的**电流**了。由于结点上不会积累电荷，因此  $\mathbf{A}^T \mathbf{y} = \mathbf{0}$  描述的物理现象是：“**电路中任何一个结点上，任意时刻流入结点的电流之和等于流出结点的电流之和**”，这就是**基尔霍夫电流定律**（*Kirchoff's Current Law, KCL*）。

我们来看方程组的全解：

$$\mathbf{y} = k_1 \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad k_1, k_2 \in \mathbb{R}$$

根据物理知识可知，电流可以仅存在于某个回路中，而不经其它回路（例如短路），因此左零空间基向量的物理意义是**回路电流的分布情况**，且仅包含小回路。结合图示可知，小回路有  $\mathbf{A}_{123}$ 、 $\mathbf{A}_{345}$  两个，对应边为  $\{1, 2, 3\}$  和  $\{3, 4, 5\}$ ，根据 *KCL* 定律可以轻松获得如上所示的两个基向量。换句话说，**关联矩阵左零空间的维度等于小回路个数**。当然，根据已有的线代知识，我们也应该知道  $\dim[\mathbf{N}(\mathbf{A}^T)] = m - r = 2$ 。

我们来看看大回路有什么“问题”。本例中的大回路即为  $\mathbf{A}_{1254}$ ，对应的解向量为：

$$\mathbf{y}_{1254} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \mathbf{y}_{123} + \mathbf{y}_{345}$$

由上可见  $\mathbf{y}_{1254}$  无法作为  $\mathbf{N}(\mathbf{A}^T)$  的基。因此在图网络分析中，我们只关注内部没有回路的最小回路，大回路对应的解向量通常可由内含的小回路线性组合而成。

同 3.8.2， $\mathbf{A}^T \mathbf{y}$  在有源（电流源）场景中也有应用。我们只需构建一个描述各边电流分布及流向信息的电流源矩阵  $\mathbf{f}$ ，并将系数矩阵  $\mathbf{0}$  改为  $\mathbf{f}$ ，即  $\mathbf{A}^T \mathbf{y} = \mathbf{f}$ 。

### 3.8.4 欧拉公式与欧姆定律



结合前两节的分析可知：

$$\underbrace{\dim [N(A^T)]}_{\text{loops}} = \underbrace{m}_{\text{edges}} - \underbrace{r}_{\text{nodes}-1} \implies \text{nodes} - \text{edges} + \text{loops} = 1 \quad (3-8-2)$$

这一结论其实是**欧拉定理** (Euler's formula) 在**二维平面**上的表现。空间拓扑学中的欧拉定理内容是：“**在任何一**  
**个规则球面上，区域个数记作  $R$ ，顶点个数记作  $V$ ，边界个数记作  $E$ ，则有  $R + V - E = 2$ ”。**

当三维空间降维至二维平面时， $z$  轴信息消失，大回路失去了独立构成区域的能力，相当于总区域数  $R$  减少了一个，因此有  $R + V - E = 1$ ，即 (3-8-2)。

至此，我们已经得到了描述电势、电流的两个电学基本公式  $\mathbf{e} = \mathbf{A}\mathbf{x}$ 、 $\mathbf{A}^T \mathbf{y} = \mathbf{f}$ 。根据我残存的电学知识（对不住我的各位物理老师），**欧姆定律**是连接它们的桥梁： $\mathbf{y} = \mathbf{C}\mathbf{e}$ ，其中  $\mathbf{C}$  表示各边上的电导（电阻的倒数）。最后，让我们用一个平衡等式来结束短暂的应用数学之旅，它完整地描述了一个稳态电路在各结点、各边上的电学信息：

$$\mathbf{A}^T \mathbf{C} \mathbf{A} \mathbf{x} = \mathbf{f} \quad (3-8-3)$$

最后的最后，还需要再提醒几点：

(1) 形如 (3-8-3) 的方程不仅适用于电路分析，在很多场景中都能见到。至少在我目前研究的多通道电生理信号分析中是抬头不见低头见了。当我们赋予  $\mathbf{A}$ 、 $\mathbf{C}$ 、 $\mathbf{x}$  以及  $\mathbf{f}$  不同的物理意义时，它描述的问题自然也会随之改变；

(2) 形如  $\mathbf{A}^T \mathbf{C} \mathbf{A}$  或  $\mathbf{A}^T \mathbf{A}$  的矩阵往往具有一些独特的魅力（比如对称），持续不断地吸引着研究者们埋头深耕。这些矩阵我们在下一章以及 MIT-18.065 课程中会进一步详细探讨。

## 4 Orthogonality

### 4.1 Orthogonality of the Four Subspaces

#### 4.1.1 正交向量

在大多数情况下，“正交”是“垂直”的另一种说法。在  $\mathbb{R}^n$  空间中，两个列向量  $\mathbf{x}$ 、 $\mathbf{y}$  正交意味着：

$$\mathbf{x}^T \mathbf{y} = 0 \text{ or } \langle \mathbf{y}, \mathbf{x} \rangle = 0 \quad (4-1-1)$$

这一点可以在二维空间中通过**勾股定理**加以验证。对于  $\mathbb{R}^2$  中由向量  $\mathbf{x}$ 、 $\mathbf{y}$  以及  $\mathbf{x} + \mathbf{y}$  围成的**直角三角形**，勾股定理告诉我们：

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2 \quad (4-1-2)$$

对于向量而言,“长度”的概念可以从几何的角度理解:例如  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ , 从空间原点到坐标点  $(x_1, x_2, \dots, x_n)$  的距离为  $\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ , 即:

$$\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x} \quad (4-1-3)$$

在目前的知识体系下,这样表示是没有问题的。当然等以后我们接触了矩阵范数的概念时,会对 (4-1-3) 有一个更为严谨的理解。把上述结果代回 (4-1-2) 可得:

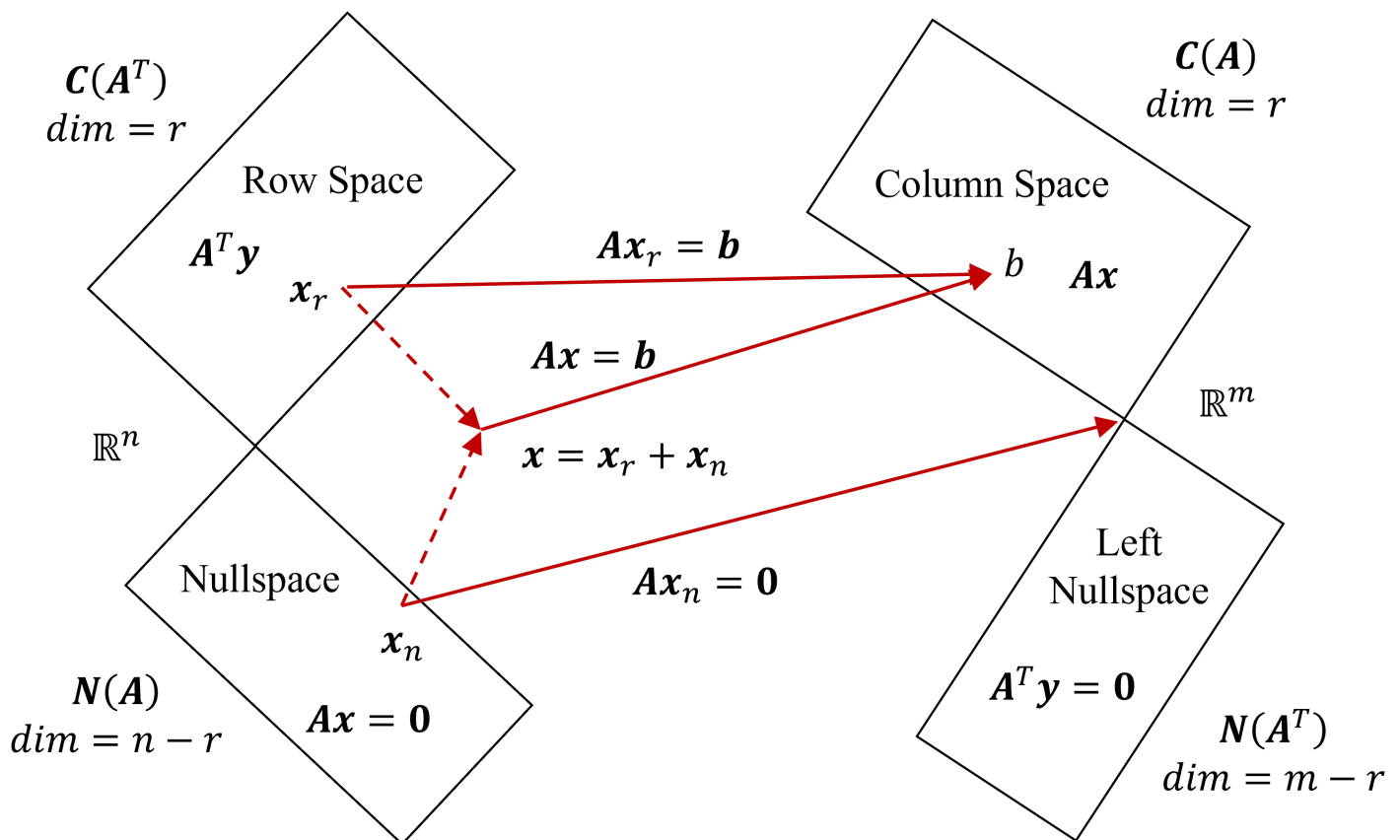
$$\mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y} = (\mathbf{x} + \mathbf{y})^T (\mathbf{x} + \mathbf{y}) \implies \mathbf{y}^T \mathbf{x} + \mathbf{x}^T \mathbf{y} = 0$$

$$\because \mathbf{y}^T \mathbf{x} = (\mathbf{x}^T \mathbf{y})^T \in \mathbb{R}$$

$$\therefore \mathbf{y}^T \mathbf{x} = \mathbf{x}^T \mathbf{y} \implies \mathbf{x}^T \mathbf{y} = 0$$

### 4.1.2 正交子空间

下图展示了四种子空间的基本信息以及相互关系:



这张图还描述了两组正交关系,在说明它们之前,我们有必要先了解一下空间正交的含义,对于两个  $n$  维子空间  $S$ 、 $T$ ,当我们说“子空间  $S$ 、 $T$  二者正交”时,其含义为:

$$S \perp T \implies \forall \mathbf{s} \in S, \mathbf{t} \in T, \mathbf{s}^T \mathbf{t} \text{ or } \mathbf{t}^T \mathbf{s} = 0 \quad (4-1-4)$$

向量空间的“正交”与我们在欧式几何中谈论的“垂直”并不完全一致。譬如三维笛卡尔坐标系中的  $xOy$  平面与

$yOz$  平面。这两个平面在立体几何的概念里是**垂直**的；在线代领域里，它们都是  $\mathbb{R}^3$  的子空间。然而根据 (4-1-4) 的定义，两者并不**正交**。很显然位于  $y$  轴上的向量同时属于两个平面，却相互平行而非垂直。需要注意的是，问题的根源并不是二者有交集，而是它们的交集包含了非零元素。举个例子，我们知道  $x$  轴和  $y$  轴是垂直且正交的，二者的唯一交集就是零元（原点）。更一般地，**如果子空间  $S$  与  $T$  存在非零元的交集，则二者必不可能正交**。

接下来我们说明图中的正交关系。首先是行空间与零空间，二者不仅正交，还将母空间  $\mathbb{R}^n$  划分成了两个**互补**空间，即：

$$C(A^T), N(A) \subseteq \mathbb{R}^n, C(A^T) = N(A)^\perp \quad (4-1-5)$$

我们先从零空间入手，看看它们正交的原因。对于线性方程组  $Ax = 0$ ：

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \begin{cases} \mathbf{a}_1^T \mathbf{x} = 0 \\ \mathbf{a}_2^T \mathbf{x} = 0 \\ \vdots \\ \mathbf{a}_n^T \mathbf{x} = 0 \end{cases}$$

其中  $\mathbf{a}_i^T$  表示  $A$  中的第  $i$  行向量。这个方程组说明零空间中的任意向量  $\mathbf{x}$  都与  $A$  的**每个行向量**正交，进而满足 (4-1-4) 提出的定义，即有  $C(A^T) \perp N(A)$ 。至于正交补，是因为零空间与列空间的维度之和为母空间维度，且二者的交集只包含零元。举例来说，对于  $\mathbb{R}^3$  中的直线簇  $L$ ：

$$L = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$$

$C(L^T)$  表示经过原点与  $(1, 2, 5)$  的直线， $\dim[C(L^T)] = \text{rank}(L^T) = 1$ ， $\dim[N(L)] = 2$ 。更具体地：

$$N(L) = k_1 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix} = k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2, k_1, k_2 \in \mathbb{R}$$

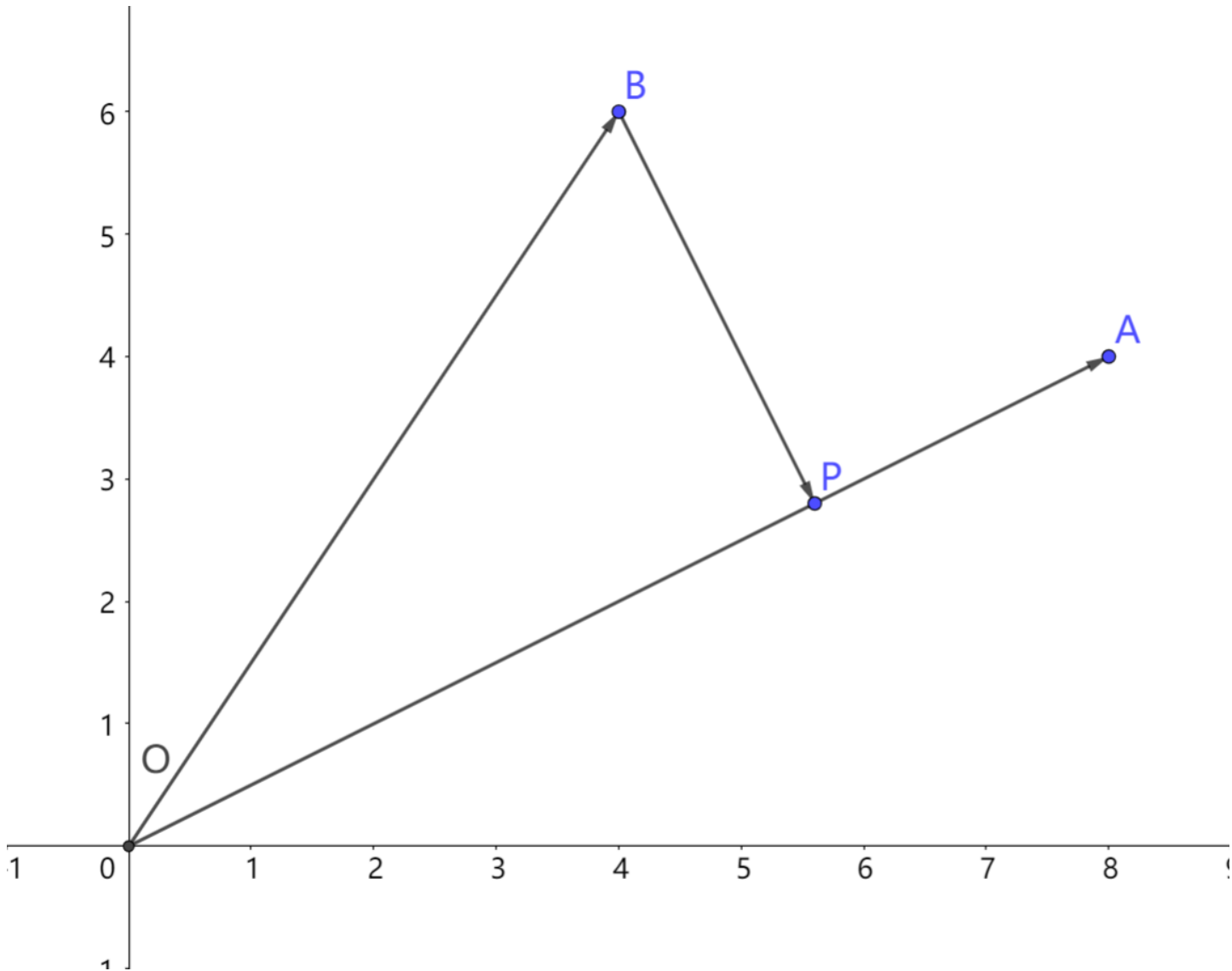
可见  $N(L)$  的几何表现为一个平面。根据立体几何的相关知识， $\mathbf{x}_1$  与  $\mathbf{x}_2$  是平面  $N(L)$  上的非共线向量。而方向向量  $[1, 2, 5]^T$  经计算验证可知与  $\mathbf{x}_1$ 、 $\mathbf{x}_2$  均垂直，因此  $C(L^T) \perp N(L)$ 。更进一步地， $C(L^T)$  与  $N(L)$  相交于原点，二者空间相加  $C(L^T) + N(L)$  的几何含义是：以  $C(L^T)$  为中心轴，平移平面  $N(L)$  所形成的立体空间。根据 (3.7.1) 的结论，相加空间依然是线性子空间。显然这个子空间与  $\mathbb{R}^3$  是等价的，即  $C(L^T)$  是  $N(L)$  的正交补，反之亦然。

以上分析过程同样适用于  $C(A)$  与  $N(A^T)$ ，即  $C(A) = N(A^T)^\perp$ ，证明过程在此不再赘述。

## 4.2 Projections

### 4.2.1 平面投影

在二维笛卡尔坐标系中，假设有如下图所示的点  $A$ 、 $B$ ，现需要在射线  $OA$  寻找一点  $P$  使得  $BP$  长度最短。即寻找向量  $\vec{OB}$  在  $\vec{OA}$  方向上的投影  $\vec{OP}$ ：



在中学数学中，这顶多只能算是课堂开场白难度的问题，我们可以很熟练地用三角函数把它从里到外分析得清清楚楚。但是线性代数追求的就是公式上的极简风格，哪怕能少写几个字母都是快乐的。这种极简主义看似抽象，但在从低维向高维延申拓展时会有奇效。

从向量的视角来看：

$$\vec{OB} + \vec{BP} = \vec{OP} \implies \vec{BP} = \vec{OP} - \vec{OB}$$

简单起见，我们令  $\vec{OA} = \mathbf{a}$ 、 $\vec{OB} = \mathbf{b}$ 、 $\vec{OP} = \mathbf{p}$ 、 $\vec{BP} = \mathbf{e}$ ，则有  $\mathbf{e} \perp \mathbf{a}$ 。根据上一节的定义，有：

$$\mathbf{a}^T \mathbf{e} = 0 \implies \mathbf{a}^T (\mathbf{b} - \mathbf{p}) = 0$$

由于  $\mathbf{p}$  与  $\mathbf{a}$  共线，则二者必然存在倍数关系  $x$  ( $x \in \mathbb{R}$ ) 使得  $\mathbf{p} = \mathbf{a}x$ ，进而上述正交关系可以表示为：

$$\mathbf{a}^T(\mathbf{b} - \mathbf{a}x) = 0 \quad (4-2-1)$$

由于  $\mathbf{a}$ 、 $\mathbf{b}$  都是一维向量（序列），系数  $x$  很容易化简求解：

$$x = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} \quad (4-2-2)$$

同时投影过程可以表示为：

$$\mathbf{p} = \mathbf{a}x \implies \mathbf{p} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}} \mathbf{b}$$

至此，我们得以走向最终解决问题所需的“总抓手”——投影矩阵  $\mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}}$ 。之所以说它是“抓手”，是因为我们在已知投影方向  $\mathbf{a}$  和待投影目标  $\mathbf{b}$  的情况下，最需要的不是投影结果  $\mathbf{p}$ ，而是一个能够帮助我们稳定获取  $\mathbf{p}$  的工具，即投影矩阵  $\mathbf{P}$ 。之后不论遇到怎么样的  $\mathbf{b}$ ，都可以方便快捷地获取投影结果  $\mathbf{p} = \mathbf{P}\mathbf{b}$ 。

投影结果  $\mathbf{p}$  与待投影向量  $\mathbf{b}$  是同步变化的，然而不论投影方向向量  $\mathbf{a}$  在长度上作何变化（非零），投影结果  $\mathbf{p}$  都不会改变，也就是说投影矩阵  $\mathbf{P}$  没有发生变化。因此我们可以作一个合理推测：尽管  $\mathbf{P}$  的计算公式由  $\mathbf{a}$  组成，但其本质上是由  $\mathbf{a}$  的单位方向向量  $\mathbf{i}_a$  决定的。

将上述结果从  $\mathbb{R}^2$  稍稍拓展至  $\mathbb{R}^n$ ，可以得到类似的结果，即  $\mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}} \in \mathbb{R}^{n \times n}$ 。除此之外，我们还能得出以下结论：

- (1) 由于  $\mathbf{a}$  可视为秩 1 矩阵，因此  $\text{rank}(\mathbf{a}\mathbf{a}^T) = 1$ ，即  $\mathbf{P}$  也是秩 1 矩阵， $\mathcal{C}(\mathbf{P})$  是  $\mathbf{a}$  方向的射线簇；
- (2)  $\mathbf{P}$  是 *Hermitte* 矩阵，在实数范围内有  $\mathbf{P} = \mathbf{P}^T$ ，进一步地，**投影矩阵都是对称矩阵**；
- (3) 对投影结果  $\mathbf{p}$ （或  $\mathbf{P}\mathbf{b}$ ）再投影一次，结果不变。因此  $\mathbf{P}$  是幂等矩阵，即  $\mathbf{P}\mathbf{P} = \mathbf{P}$ ：

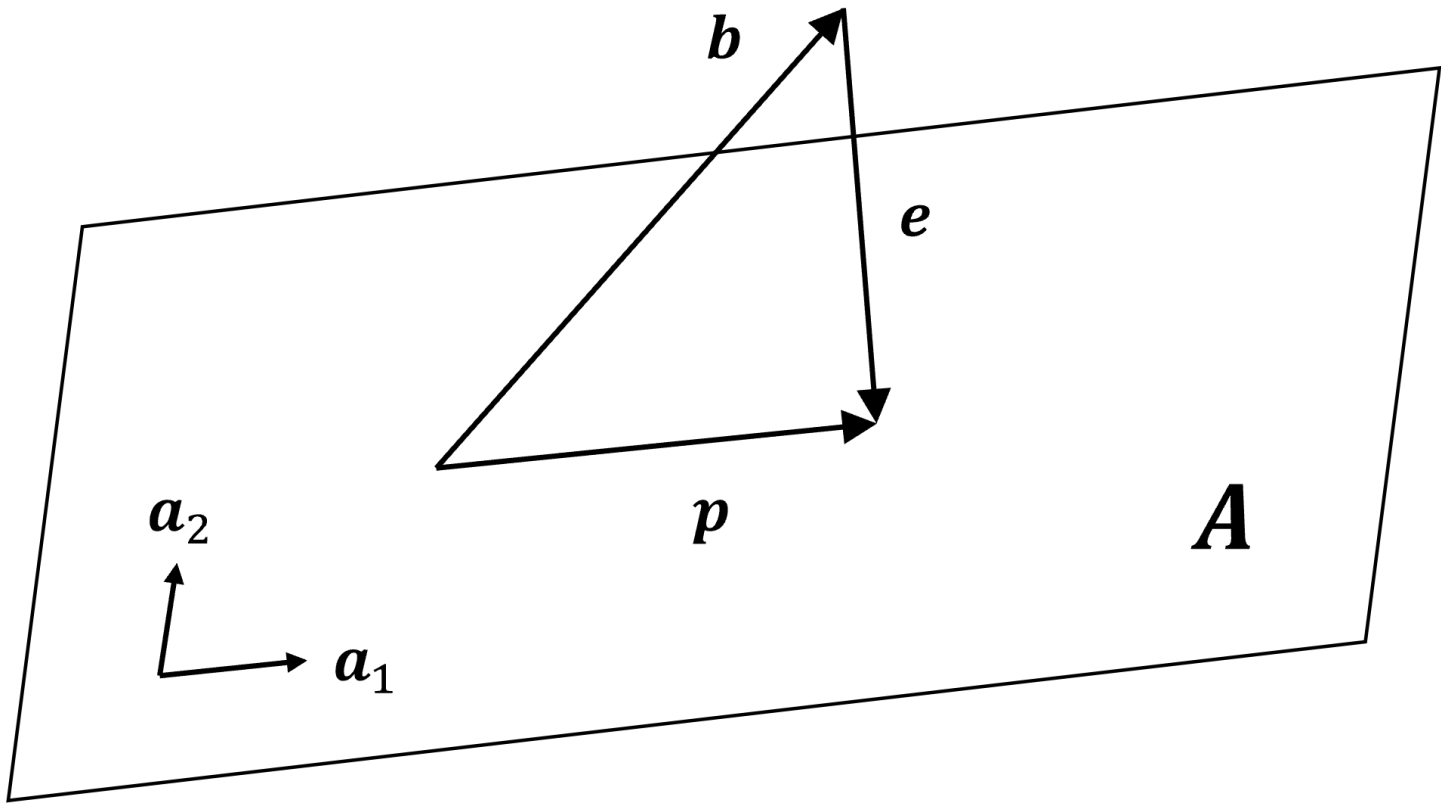
$$\mathbf{P}\mathbf{P} = \frac{\mathbf{a}(\mathbf{a}^T \mathbf{a})\mathbf{a}^T}{(\mathbf{a}^T \mathbf{a})\mathbf{a}^T \mathbf{a}} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}} = \mathbf{P}$$

在了解了投影的过程之后，我们有必要从一个更深层次的角度思考投影的本质。 $\mathbf{P}\mathbf{b}$  与  $\mathbf{b}$  的关系不仅仅是直角边与斜边、 $\cos\theta$ 、光学投射等等表象性质的关系， **$\mathbf{P}\mathbf{b}$  是  $\mathbf{b}$  在  $\mathbf{a}$  方向上的最佳逼近之一**。

说“之一”，是因为在不同误差标准的评判尺度下，最佳结果可能不止一个；说“最佳”，是因为我们当前对于误差的定义是误差向量的 2 范数平方  $\|\mathbf{P}\mathbf{b} - \mathbf{b}\|_2^2$ ，通俗地说即为向量长度的平方。而在  $\mathbf{a}$  上取其它任意投影点  $\mathbf{P}_i$  构成的投影向量  $\mathbf{p}_i$ ，均有  $\|\mathbf{p}_i - \mathbf{b}\|_2^2 > \|\mathbf{P}\mathbf{b} - \mathbf{b}\|_2^2$ 。从平面几何的角度很好理解： $\|\mathbf{P}\mathbf{b} - \mathbf{b}\|_2$  是垂线段  $BP$  的长度， $\|\mathbf{p}_i - \mathbf{b}\|_2$  是直角三角形  $P_iBP$  的斜边，斜边自然是长于直角边的。

## 4.2.2 空间投影

当我们把问题的视角从平面向量升格至多维空间时，遵循  $\mathbb{R}^2$  中的分析方法可以获得形式上相似的结果。我们先从稍微简单的情景入手，例如  $\mathbb{R}^3$  中的向量  $\mathbf{b}$  至平面  $A$  上的投影：



类似二维平面，我们同样可以在三维空间中寻找  $\mathbf{b}$  的末端点至  $A$  的垂线，连接  $\mathbf{b}$  的起点与垂足形成投影向量  $\mathbf{p}$  以及误差向量  $\mathbf{e}$ 。尽管我个人非常喜欢这种几何学的脑内风暴，但是我们还是要了解线代的描述语言。本例与上一节案例的最大差别在于目标向量  $\mathbf{a}$  变成了平面  $A$ ，所以我们需要换一种线代语言来描述一个平面，基向量组（矩阵）就是一个很好的方法。

假设  $A$  的基（列）向量为  $\mathbf{a}_1$ 、 $\mathbf{a}_2$ ，构造矩阵  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2]$ ，则  $\mathbf{A}$  的列向量生成空间（或列空间）即为平面  $A$ 、 $A$  上的任意向量都可以用  $\mathbf{a}_1$ 、 $\mathbf{a}_2$  的线性组合表示：

$$\begin{aligned}\mathbf{p} &= \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 = [\mathbf{a}_1 \quad \mathbf{a}_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{A}\mathbf{x} \\ \mathbf{e} &= \mathbf{b} - \mathbf{p} = \mathbf{b} - \mathbf{A}\mathbf{x}\end{aligned}$$

误差向量  $\mathbf{e}$  垂直于平面  $A$ ，所以  $\mathbf{e}$  正交于平面上任意向量，即正交于平面的两个基向量：

$$\begin{cases} \mathbf{a}_1^T (\mathbf{b} - \mathbf{A}\mathbf{x}) = 0 \\ \mathbf{a}_2^T (\mathbf{b} - \mathbf{A}\mathbf{x}) = 0 \end{cases} \Rightarrow \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \end{bmatrix} (\mathbf{b} - \mathbf{A}\mathbf{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{A}^T (\mathbf{b} - \mathbf{A}\mathbf{x}) = \mathbf{0} \quad (4-2-3)$$

(4-2-3) 与 (4-2-1) 简直如出一辙对吧，这就是极简风格的魅力所在。同时，我们还能发现一个小彩蛋： $\mathbf{b} - \mathbf{A}\mathbf{x} \in \mathbf{N}(\mathbf{A}^T)$ ，而误差向量  $\mathbf{b} - \mathbf{A}\mathbf{x}$  与平面垂直，即正交于平面列空间  $\mathbf{C}(\mathbf{A})$ 。这一点与 4.1.2 图片中的空间正交关系是对应的。类似地，我们可以解出投影矩阵  $\mathbf{P}$ ：

$$A^T (\mathbf{b} - A\mathbf{x}) = \mathbf{0} \implies \mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$$

$$\mathbf{p} = A\mathbf{x} = P\mathbf{b} \implies P = A (A^T A)^{-1} A^T \quad (4-2-4)$$

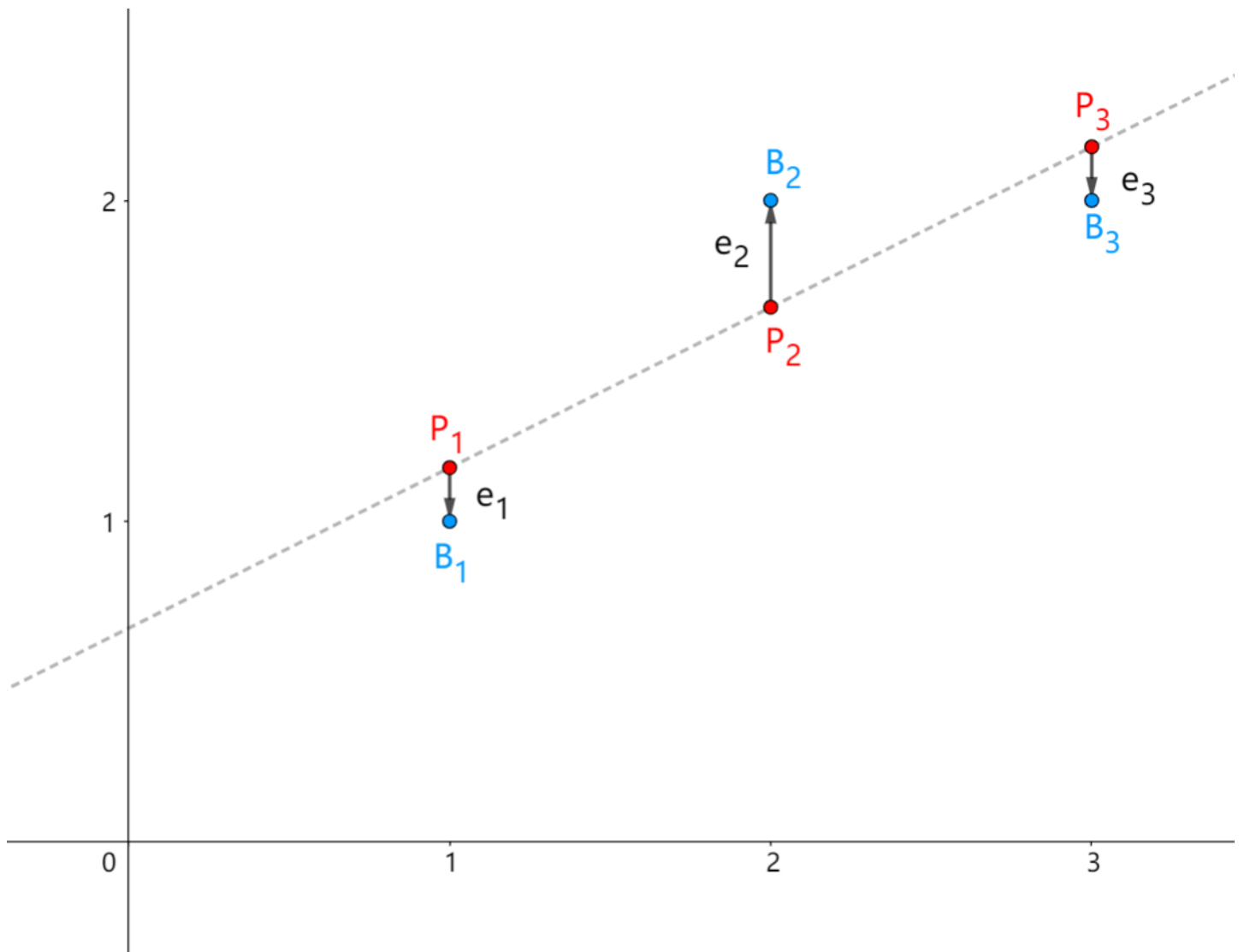
如果我们进一步拓展维度，将  $\mathbf{b}$  升格为矩阵空间  $\mathbb{R}^{n \times n}$  中的矩阵  $B \in \mathbb{R}^{b_1 \times b_2}$  ( $b_1, b_2 \leq n$ )，平面  $A$  升格为子空间  $A \in \mathbb{R}^{a_1 \times a_2}$  ( $a_1, a_2 \leq n$ )，投影矩阵的形式依然为 (4-2-4) 所示，投影结果升格为  $PB$ 。同理可以证明空间投影中的矩阵  $P$  依然满足 *Hermitte* 矩阵、幂等矩阵的性质，且  $PB$  是  $B$  在  $A$  空间上的最佳逼近之一。

最后我们不得不面临的一个问题就是：为什么（或者什么时候）需要投影？这个问题的答案其实与线性方程组的近似解有关。根据列空间的相关知识我们知道， $A\mathbf{x} = \mathbf{b}$  并非总是有解的：对于超定系统，大多数情况下线性方程组无解，可超定系统却是现实物理世界中遍布各地的存在。例如描述某个物理过程的理论模型设计了 5 个参数，而由于测量误差等干扰因素的存在，实际测试往往需要重复成百上千次，此时方程数量远多于未知数。对于随机系统，方程系数矩阵往往还是满秩的。除了删减优化样本（方程个数），最常用、也是最合理的办法就是投影，即“最小二乘逼近法”，相关内容将在下一节展开。

## 4.3 Least Squares Approximations

### 4.3.1 最小二乘回归

继续上一节末尾的问题，投影是解决超定系统方程组的重要手段之一。对于  $A\mathbf{x} = \mathbf{b}$ ，我们知道  $A\mathbf{x}$  总在  $C(A)$  中，而  $\mathbf{b}$  不一定。只要我们对  $\mathbf{b}$  进行微调，使得  $\hat{\mathbf{b}} \approx \mathbf{b}$  且  $\hat{\mathbf{b}} \in C(A)$ ，则方程组  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$  有解。微调的结果就是将  $\mathbf{b}$  替换成其在列空间  $C(A)$  上的投影，即最小二乘法。这里我们以一个简单的直线回归问题作为案例：



假设  $\{B_1, B_2, B_3\}$  为数据采样点集：

$$\begin{cases} B_1 : (1, 1) \\ B_2 : (2, 2) \\ B_3 : (3, 2) \end{cases} \Rightarrow y = kx + b, \begin{cases} k + b = 1 \\ 2k + b = 2 \\ 3k + b = 2 \end{cases}$$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \Rightarrow \mathbf{Ax} = \mathbf{b}$$

显然没有任何一条直线能够恰好同时经过这三个点，即  $\mathbf{Ax} = \mathbf{b}$  无解。这里尤其要注意  $\mathbf{x}$  与  $x$ 、 $\mathbf{b}$  与  $b$  没有关联，单纯只是用惯了这些字母而已。我们需要拟合一条直线，其表达式为  $y = \hat{k}x + \hat{b}$ （虚线所示），使得在对应横坐标下的采样点  $\{P_1, P_2, P_3\}$  与原始数据的**误差平方和**  $e$  最小：

$$e = e_1^2 + e_2^2 + e_3^2 = (\hat{k} + \hat{b} - 1)^2 + (2\hat{k} + \hat{b} - 2)^2 + (3\hat{k} + \hat{b} - 2)^2$$

这里我们规定向上为正，向下为负（即拟合点减去实际点的偏差）。将误差表示为向量，并将一切都用线代语言翻译一下：



$$\begin{bmatrix} \hat{k} \\ \hat{b} \end{bmatrix} = \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{e}\|^2 = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \quad (4-3-1)$$

我们不妨先通过我们熟悉的微积分来考虑总体误差  $e$ ：

$$f(\hat{k}, \hat{b}) = e = 14\hat{k}^2 + 12\hat{k}\hat{b} + 3\hat{b}^2 - 22\hat{k} - 10\hat{b} + 9$$

$$\begin{cases} \frac{\partial f}{\partial \hat{k}} = 28\hat{k} + 12\hat{b} - 22 = 0 \\ \frac{\partial f}{\partial \hat{b}} = 6\hat{b} + 12\hat{k} - 10 = 0 \end{cases} \Rightarrow \begin{cases} 14\hat{k} + 6\hat{b} = 11 \\ 6\hat{k} + 3\hat{b} = 5 \end{cases} \Rightarrow \begin{cases} \hat{k} = \frac{1}{2} \\ \hat{b} = \frac{2}{3} \end{cases}$$

因此拟合直线为  $y = \frac{1}{2}x + \frac{2}{3}$ ，拟合采样点的纵坐标依次为  $\frac{7}{6}$ 、 $\frac{5}{3}$ 、 $\frac{13}{6}$ ，对应误差依次为  $-\frac{1}{6}$ 、 $\frac{2}{6}$ 、 $-\frac{1}{6}$ 。

接下来看看线代如何解决以上问题：

$$\begin{aligned} \mathbf{A}\hat{\mathbf{x}} &= \mathbf{b} \\ \mathbf{A}^T \mathbf{A}\hat{\mathbf{x}} &= \mathbf{A}^T \mathbf{b} \Rightarrow \begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} \hat{k} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 11 \\ 5 \end{bmatrix} \end{aligned}$$

至此我们已经得到了与微积分方法相同的二元一次方程组，我们甚至可以一步得到  $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ 。根据上一节的知识，我们知道  $\mathbf{A}\hat{\mathbf{x}}$  是  $\mathbf{b}$  在空间  $\mathbf{A}$  上的投影。这里如果你深想了，但没有深透的话，会发现一件非常“诡异”的事：

我们知道  $\mathbf{A}\mathbf{x} = \mathbf{b}$  是无解的，事实上  $\mathbf{A}\hat{\mathbf{x}} \neq \mathbf{b}$ 。但我们从  $\mathbf{A}^T \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$  求解得到的  $\hat{\mathbf{x}}$ ，为什么从这一步逆向回到  $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$  就行不通了呢？

(我强烈建议各位读者不要急着浏览接下来的部分。尽管这段推理本身无足轻重，但是费尽心机解开一个 bug 之后，我的多巴胺释放水平显著提升了，本着公平起见的原则，大伙也痛苦一下吧。)

这个问题可以从两种角度解释。

(1) 首先是比较本质、抽象一点的方法：好比  $D\mathbf{A}\mathbf{B} = D\mathbf{C}$ ，得到  $D(\mathbf{A}\mathbf{B} - \mathbf{C}) = \mathbf{0}$  是没问题的，但不能进一步等价于  $\mathbf{A}\mathbf{B} - \mathbf{C} = \mathbf{0}$ ，更不能等价于  $\mathbf{A}\mathbf{B} = \mathbf{C}$ 。原因很简单，各位难道没见过  $\mathbf{A}\mathbf{x} = \mathbf{0}$  有多个解的情况吗？ $\mathbf{A}\mathbf{B} - \mathbf{C} \in N(D)$ ，只要矩阵  $D$  不满秩，零空间就必然包含除零元以外的其它向量元素，换句话说  $\mathbf{A}\mathbf{B} - \mathbf{C}$  有无数个也毫不奇怪。由已知条件有：

$$\mathbf{A}^T (\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) = \mathbf{0} \Rightarrow \mathbf{A}\hat{\mathbf{x}} - \mathbf{b} \in N(\mathbf{A}^T)$$

而  $\dim[N(\mathbf{A}^T)] = 1$ ，即左零空间包含无限多个元素，因此存在不同的  $\mathbf{b}$  使得上式成立是正常情况。自然地，不同的  $\mathbf{b}$  可能对应不同的解  $\hat{\mathbf{x}}$ （或者无解）。

(2) 其次是更表层、直观一点的方法：正因为  $\mathbf{A}\mathbf{x} = \mathbf{b}$  无解，我们才需要把  $\mathbf{b}$  微调成  $\hat{\mathbf{b}}$ ，在这个过程中：

$$\mathbf{A}^T \mathbf{b} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 11 \\ 5 \end{bmatrix}$$

$$\mathbf{A}^T \hat{\mathbf{b}} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{bmatrix} = \begin{bmatrix} 11 \\ 5 \end{bmatrix}$$

什么叫微调？这就叫微调。尽管  $\mathbf{Ax} = \mathbf{b}$  确实无解，但还是可以通过  $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b} = \mathbf{A}^T \hat{\mathbf{b}}$  抢救一下，在等式维持成立的情况下对右侧进行调整，最终  $\mathbf{Ax} = \hat{\mathbf{b}}$  成功化险为夷。

最后总结一下，当超定系统线性方程组  $\mathbf{Ax} = \mathbf{b}$  无解时，可以尝试将  $\mathbf{b}$  投影至  $\mathbf{A}$  的列空间中，从而使得原方程有解。特别地，投影矩阵为  $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ 。

### 4.3.2 需要补充的一些特性

对于  $\mathbf{Ax} = \mathbf{b}$ ，有投影矩阵  $\mathbf{P} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ 。通过对  $\mathbf{b}$  投影 ( $\mathbf{Pb}$ ) 可将其投射至  $\mathbf{A}$  的列空间  $\mathbf{C}(\mathbf{A})$ ，这里存在两种特殊情况：

(1) 若  $\mathbf{b} \in \mathbf{C}(\mathbf{A})$ ，则原方程有解，投影没有实际意义，即有  $\mathbf{Pb} = \mathbf{b}$ ：

$$\mathbf{b} = \mathbf{Ay}, \implies \mathbf{Pb} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Ay} = \mathbf{Ay} = \mathbf{b}$$

根据上一节的分析，我们需要注意  $\mathbf{Pb} = \mathbf{b}$  并不意味着  $\mathbf{P} = \mathbf{I}$ ；

(2) 若  $\mathbf{b} \in \mathbf{N}(\mathbf{A}^T)$  (或  $\mathbf{b} \perp \mathbf{C}(\mathbf{A})$ )，则有  $\mathbf{Pb} = \mathbf{0}$ ，即投影过程会使得  $\mathbf{b}$  完全衰减：

$$\mathbf{A}^T \mathbf{b} = \mathbf{0} \implies \mathbf{Pb} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} = \mathbf{0}$$

通常情况下， $\mathbf{b}$  中既有属于列空间的成分，也有属于左零空间的成分。因此投影过程能够削减一部分无关分量，保留有效成分，这一点在信号处理过程中经常使用。

最后是一点关于误差向量的内容。对于原始向量  $\mathbf{b}$ ，通过投影我们将其分解为  $\mathbf{p} + \mathbf{e}$ 。其中误差（属于左零空间）与投影分量（属于列空间）是正交的，即  $\mathbf{e} \perp \mathbf{p}$ 。换个角度看，其实误差向量也是某种投影，其目标空间是左零空间，即与目标空间正交的子空间。因此误差向量也存在投影矩阵  $\mathbf{I} - \mathbf{P}$ ，且该矩阵同样是 *Hermitte* 矩阵，满足幂等性质。

## 4.4 Orthogonal Basis and Gram-Schmidt

### 4.4.1 正交基与正交矩阵

正交基与正交向量组的概念很接近，其差异仅在于正交基是单位化的。即对于一组正交基  $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ ，它们满足以下数学关系：

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad (4-4-1)$$

由正交基向量组成的矩阵满足  $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ 。当  $\mathbf{Q}$  为方阵时，还有  $\mathbf{Q}^T = \mathbf{Q}^{-1}$ ，这样的  $\mathbf{Q}$  称为正交矩阵。

接下来展示一些典型正交矩阵的示例，如  $n$  阶方阵的置换矩阵、如下所示的  $\mathbf{Q}_1, \mathbf{Q}_2$ ：

$$\mathbf{Q}_1 = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \mathbf{Q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

这两个矩阵的示例可不是随便举的： $\mathbf{Q}_1$  又称旋转矩阵，对于任何二维向量  $\mathbf{x}$ ， $\mathbf{Q}_1 \mathbf{x}$  的几何表现为将其逆时针旋转  $\theta$  角度。事实上， $\mathbf{Q}_1$  的列向量就是由  $\mathbf{i}, \mathbf{j}$  旋转而来。 $\mathbf{Q}_1^T = \mathbf{Q}_1^{-1}$  与之相反，其作用是将向量顺时针旋转  $\theta$  角度。其原理很简单，从数形结合的角度来看， $\mathbf{Q}_1^{-1}$  的作用与  $\mathbf{Q}_1$  应当是相互抵消的。

而关于  $\mathbf{Q}_2$ ，Hadamard 基于分块矩阵的思想对其进行了扩展：

$$\mathbf{Q}_3 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} = \left( \frac{1}{\sqrt{2}} \right)^2 \begin{bmatrix} \mathbf{Q}_2 & \mathbf{Q}_2 \\ \mathbf{Q}_2 & -\mathbf{Q}_2 \end{bmatrix}$$

$$\mathbf{Q}_4 = \frac{1}{2\sqrt{2}} \begin{bmatrix} \mathbf{Q}_2 & \mathbf{Q}_2 & \mathbf{Q}_2 & \mathbf{Q}_2 \\ \mathbf{Q}_2 & -\mathbf{Q}_2 & \mathbf{Q}_2 & -\mathbf{Q}_2 \\ \mathbf{Q}_2 & \mathbf{Q}_2 & -\mathbf{Q}_2 & -\mathbf{Q}_2 \\ \mathbf{Q}_2 & -\mathbf{Q}_2 & -\mathbf{Q}_2 & \mathbf{Q}_2 \end{bmatrix} = \left( \frac{1}{\sqrt{2}} \right)^3 \begin{bmatrix} \mathbf{Q}_3 & \mathbf{Q}_3 \\ \mathbf{Q}_3 & -\mathbf{Q}_3 \end{bmatrix}$$

诸如此类的方阵都是正交矩阵。Hadamard 矩阵的数学定义是“一个仅由 1 和 -1 组成的正交矩阵序列”，且序列中的相邻矩阵存在递归关系。Hadamard 矩阵常用于纠错码，如 Reed-Muller 码等。

正交矩阵的神奇性质之一表现在投影上，即正交投影。假设我们需要将向量  $\mathbf{x} \in \mathbb{R}^{n \times 1}$  投影至具有正交列向量组的矩阵  $\mathbf{Q}$  的列空间  $\mathcal{C}(\mathbf{Q})$  中，根据已有知识我们知道投影矩阵为  $\mathbf{P} = \mathbf{Q} (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T$ 。而  $\mathbf{Q}$  满足  $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}$ ，即正交投影矩阵的形式可简化为  $\mathbf{Q} \mathbf{Q}^T$ 。

注意这一结论的前提是  $\mathbf{Q}$  仅需满足具有正交列向量组即可，当  $\mathbf{Q}$  为方阵（正交矩阵）时， $\mathbf{Q} \mathbf{Q}^T = \mathbf{I}$ ，即正交投影没有实质性作用。这是因为  $\mathbf{x}$  的维度与  $\mathcal{C}(\mathbf{Q})$  的维度相同，而  $\mathcal{C}(\mathbf{Q})$  本质上等价于  $\mathbb{R}^n$ ，把向量投影至原空间当然不需要额外的任何操作，单位阵足矣。

正交矩阵的第二性质是范数不变性，严格来说是向量的长度不变，即  $\|\mathbf{Q}\mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2$ ：

$$\|\mathbf{Q}\mathbf{x}\|_2^2 = (\mathbf{Q}\mathbf{x})^T (\mathbf{Q}\mathbf{x}) = \mathbf{x} \mathbf{Q} \mathbf{Q}^T \mathbf{x} = \mathbf{x} \mathbf{x}^T = \|\mathbf{x}\|_2^2 \quad (4-4-2)$$

### 4.4.2 Gram-Schmidt 正交化

除了上一节举例的那些具有特殊几何或物理意义的正交矩阵，当然还存在很多相对比较“平凡”的正交矩阵。事实上，我们完全可以对一个非正交方阵进行某种变换获得正交矩阵，*Gram-Schmidt* 正交化就是其中一种非常经典的方法。对于  $\mathbb{R}^n$  中的非正交向量  $\mathbf{a}$  与  $\mathbf{b}$ ，现演示通过 *Gram-Schmidt* 正交化获得正交向量  $\boldsymbol{\alpha}$ 、 $\boldsymbol{\beta}$  的过程：

(1) 固定初始向量  $\mathbf{a}$  的方向，并将其标准化：

$$\boldsymbol{\alpha} = \frac{\mathbf{a}}{\|\mathbf{a}\|} \quad (4-4-3)$$

(2) 将第二向量  $\mathbf{b}$  分解为平行与垂直于  $\boldsymbol{\alpha}$  (或  $\mathbf{a}$ ) 的两个部分，并将后者标准化。平行部分即为  $\mathbf{b}$  在该方向上的投影  $\mathbf{p}_{\mathbf{b},\mathbf{a}}$ ，垂直部分表示误差向量，即  $\mathbf{e}_b = \mathbf{b} - \mathbf{p}_{\mathbf{b},\mathbf{a}}$ ：

$$\mathbf{e}_b = \mathbf{b} - \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}\mathbf{b}, \quad \boldsymbol{\beta} = \frac{\mathbf{e}_b}{\|\mathbf{e}_b\|} \quad (4-4-4)$$

上述过程当然可以继续拓展，例如此时我们再引入第三个向量  $\mathbf{c}$ ：

(3) 将第三向量  $\mathbf{c}$  依次面向  $\mathbf{b}$  与  $\mathbf{a}$  进行如 (2) 所示的步骤，最终可得到同时与  $\mathbf{b}$  与  $\mathbf{a}$  正交的标准向量  $\boldsymbol{\theta}$ ：

$$\mathbf{e}_c = \mathbf{c} - \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}\mathbf{c} - \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}}\mathbf{c}, \quad \boldsymbol{\theta} = \frac{\mathbf{e}_c}{\|\mathbf{e}_c\|} \quad (4-4-5)$$

以此类推，正交化可以面向更多数目的向量组。在 (2) 中提到了一个名词即“正交分解”，这提醒我们 *Gram-Schmidt* 正交化本质上与 LU 分解一样，可以将普通矩阵分解为正交向量组矩阵与另一个矩阵的乘积，事实上这就是正交三角分解，它更装逼的名字叫约化 QR 分解 (*QR decomposition*)。对于矩阵  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ，约化 QR 分解将其列向量转化为标准正交向量 (基) 并进行重构  $\mathbf{A} = \mathbf{Q}\mathbf{R}$ ：

$$\mathbf{A} = \underbrace{[\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n]}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \mathbf{a}_1^T \mathbf{q}_1 & \mathbf{a}_2^T \mathbf{q}_1 & \cdots & \mathbf{a}_n^T \mathbf{q}_1 \\ 0 & \mathbf{a}_2^T \mathbf{q}_2 & \cdots & \mathbf{a}_n^T \mathbf{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{a}_n^T \mathbf{q}_n \end{bmatrix}}_{\mathbf{R}} \quad (4-4-6)$$

其中  $\mathbf{R}$  原本的形态应该是：

$$\mathbf{R} = \begin{bmatrix} \mathbf{a}_1^T \mathbf{q}_1 & \mathbf{a}_2^T \mathbf{q}_1 & \cdots & \mathbf{a}_n^T \mathbf{q}_1 \\ \mathbf{a}_1^T \mathbf{q}_2 & \mathbf{a}_2^T \mathbf{q}_2 & \cdots & \mathbf{a}_n^T \mathbf{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_1^T \mathbf{q}_n & \mathbf{a}_2^T \mathbf{q}_n & \cdots & \mathbf{a}_n^T \mathbf{q}_n \end{bmatrix}$$

由于正交化的步骤中保证了新向量必然正交于原有的全部向量，因此  $\mathbf{a}_1^T \mathbf{q}_2 = \mathbf{a}_1^T \mathbf{q}_3 = \cdots = \mathbf{a}_{n-1}^T \mathbf{q}_n = 0$ 。

接下来以具体矩阵  $\mathbf{A}$  为例：

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}, \mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

$$\mathbf{e}_{\mathbf{a}_2} = \mathbf{a}_2 - \frac{\mathbf{a}_1 \mathbf{a}_1^T}{\mathbf{a}_1^T \mathbf{a}_1} \mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{q}_2 = \frac{\mathbf{a}_2}{\|\mathbf{a}_2\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

## 5. Determinants

当我学习线代时，这一部分的内容在同济版线代课本中是第一节（天津大学自用的课本也一样），试想一下对于一个纯新手而言，开场白如同天书一般难以消化，往后的日子大概率也是遍布阴霾。而在 MIT-18.06 这里，行列式被安排在了第五章。此时我们已经对矩阵的初等变换、LU 分解、可逆性等诸多性质都有了充分的感性与理性理解，关于行列式的各种性质推导也都变得顺理成章。

我相信作为教师，国内高校老师的水平与国外没有显著差别，甚至本土老师更了解国情与学生的个性，本应提供质量更高的课堂。然而受限于违反教育规律的各种教材，最终质量往往不如人意，这一现象值得每一位教育从业者以及有意向投身教育事业的人反思。

### 5.1 The Properties of Determinants

方阵的行列式是一个简单的数字，但是这个数字包含了非常丰富的矩阵信息。其中最基本、最重要的信息是  $\det(\mathbf{A}) = 0$ ，则矩阵奇异，即不可逆；反之  $\det(\mathbf{A}) \neq 0$ ，则矩阵非奇异、可逆。对于二阶方阵而言，行列式的具体计算公式为：

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \quad (5-1-1)$$

在此基础上，我们接下来依次介绍行列式的十条基本性质。

#### (1) 单位阵的行列式为 1：

$$\det(\mathbf{I}) = 1 \quad (5-1-2)$$

**(2) 每进行一次行交换，行列式变号一次。**关于这一条的延申是：任何置换矩阵的行列式不是 1 就是 -1，符号取决于它由单位阵经过了多少次行变换得来。

**(3-A) 矩阵单行乘上常系数，其行列式的值也乘上相同系数：**

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad (5-1-3)$$

**(3-B) 矩阵单行对应的线性组合同样存在于行列式中：**

$$\begin{vmatrix} a + a_0 & b + b_0 \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a_0 & b_0 \\ c & d \end{vmatrix} \quad (5-1-4)$$

前三条性质是基础中的基础，根据它们以及一些学过的矩阵知识（无需（5-1-1））即可以推导出剩下的七条性质！

**(4) 矩阵中存在相同行时，行列式为 0：**

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = 0 \quad (5-1-5)$$

这一条性质可由（2）证明，交换两相同行不会改变矩阵的任何元素，所以行列式应当相等；而进行了一次行交换意味着行列式应当变号。唯一能同时满足以上两个条件的情况即行列式为 0。

**(5) 线性组合矩阵中的行向量不会改变矩阵的行列式：**

$$\begin{vmatrix} a & b \\ c + ka & d + kb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \underbrace{k \begin{vmatrix} a & b \\ a & b \end{vmatrix}}_0 = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \quad (5-1-6)$$

这一条可由（3）（4）联合证明：

$$\begin{vmatrix} a & b \\ c + ka & d + kb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + k \underbrace{\begin{vmatrix} a & b \\ a & b \end{vmatrix}}_0 = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

**(6) 矩阵存在全零行时，行列式为 0：**

$$\begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} = 0 \quad (5-1-7)$$

这一条可以通过（3-B）证明：

$$\begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} = \begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} \implies \begin{vmatrix} a & b \\ 0 & 0 \end{vmatrix} = 0$$

(7) 三角矩阵 (或对角阵) 的行列式为主元乘积:

$$\begin{aligned} \mathbf{A} = \mathbf{LU} \in \mathbb{R}^{n \times n}, \det(\mathbf{U}) &= \prod_{i=1}^n U(i, i) \\ \mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n) \in \mathbb{R}^{n \times n}, \det(\mathbf{D}) &= \prod_{i=1}^n d_i \end{aligned} \quad (5-1-8)$$

首先需要明确, 经过 LU 分解得到的上三角矩阵  $\mathbf{U}$  可以进一步行变换得到对角阵  $\mathbf{D}$ 。接下来对  $\det(\mathbf{D})$  反复应用 (3-A) 提取每一行的系数:

$$\det(\mathbf{D}) = d_1 \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{vmatrix} = \cdots = (d_1 d_2 \cdots d_n) \times \det(\mathbf{I}_n) = \prod_{i=1}^n d_i$$

(8)  $\det(\mathbf{A}) = 0$  则  $\mathbf{A}$  是奇异矩阵, 反之  $\det(\mathbf{A}) \neq 0$  则  $\mathbf{A}$  可逆。由于后半句与前半句严格对立, 因此我们着重证明前半句:

$\Leftarrow$ : 奇异矩阵通过行变换能化简出全零行, 所以行列式为 0;

$\Rightarrow$ : 以二阶方阵行列式为例:

$$\det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = 0$$

假设  $abcd \neq 0$  (若其中存在某元素为 0, 则必存在全零行或全零列), 现证明  $\mathbf{A}$  的两行向量存在线性关系:

$$ad = \left(a \times \frac{b}{a}\right) \times \left(d \times \frac{a}{b}\right) = b \times \frac{ad}{b} = bc$$

因为四个元素均不为零, 则必有  $\frac{ad}{b} = c$ , 即有:

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

即行列式为零的二阶方阵中, 行向量组存在线性关系, 由此可以推断出矩阵奇异。

**(9) 矩阵乘积的行列式等于行列式的乘积：**

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \times \det(\mathbf{B}) \quad (5-1-9)$$

这一条可结合 LU 分解的性质来加以证明。已知在**无需行交换**的情况下，LU 分解的结果是一个下三角矩阵  $\mathbf{L}$ （对角线元素全为 1）以及上三角矩阵  $\mathbf{U}$ ，因此  $\det(\mathbf{L}) = 1$ 。考虑  $\mathbf{A}$  与  $\mathbf{U}$  的关系： $\mathbf{U}$  由  $\mathbf{A}$  在不经过行交换（**行列式不变号**）的情况下，仅通过行向量的线性组合（**行列式不变**）得来，因此必有  $\det(\mathbf{U}) = \det(\mathbf{A})$ ，即：

$$\det(\mathbf{A}) = \det(\mathbf{LU}) = \underbrace{\det(\mathbf{L})}_1 \times \det(\mathbf{U})$$

这是无需行交换的情况。那么当需要行交换时，其本质是在  $\mathbf{L}$  的基础上新增一些置换矩阵，假设它们的乘积是  $\mathbf{P}$  且  $\det(\mathbf{P}) = (-1)^n$ ， $n$  表示行交换次数。考虑  $\mathbf{A}$  与  $\mathbf{U}$  的关系： $\mathbf{U}$  由  $\mathbf{A}$  经过  $n$  次行交换（**行列式变号  $n$  次**）、行向量线性组合（**行列式不变**）得来，因此有  $\det(\mathbf{U}) \times (-1)^n = \det(\mathbf{A})$ ，即：

$$\det(\mathbf{A}) = \det(\mathbf{P}^{-1}\mathbf{LU}) = \underbrace{\frac{1}{\det(\mathbf{P})}}_{(-1)^n} \times \underbrace{\det(\mathbf{L})}_1 \times \det(\mathbf{U})$$

综上所述，不论是否需要行交换，我们始终可以通过 LU 分解证明性质 (9) 的成立。

**(10) 矩阵转置后行列式不变：**

$$\det(\mathbf{A}^T) = \det(\mathbf{A}) \quad (5-1-10)$$

在 (9) 的基础上，最后一条其实很好证明了。假设有  $\mathbf{A} = \mathbf{P}^{-1}\mathbf{LU} \in \mathbb{R}^{n \times n}$ ，则  $\mathbf{A}^T = \mathbf{U}^T \mathbf{L}^T \mathbf{P}$ （置换矩阵的转置即为其逆矩阵），故有：

$$\det(\mathbf{A}) = \det(\mathbf{P}^T) \times \det(\mathbf{L}) \times \det(\mathbf{U})$$

$$\det(\mathbf{A}^T) = \det(\mathbf{U}^T) \times \det(\mathbf{L}^T) \times \det(\mathbf{P})$$

由于  $\mathbf{L}$ 、 $\mathbf{U}$  都是三角阵，因此它们的转置不改变行列式的值。现在唯一的问题在于  $\det(\mathbf{P}^T)$  是否等于  $\det(\mathbf{P})$ 。

首先毫无疑问的是  $|\det(\mathbf{P})| = 1$ ，即只存在符号差别；

其次单位阵以及部分  $\mathbf{P}$  本身即为 *Hermitte* 矩阵 ( $\mathbf{P}^T = \mathbf{P}$ )，这类  $\mathbf{P}$  显然有  $\det(\mathbf{P}^T) = \det(\mathbf{P})$ 。需要强调一下什么时候会产生对称置换矩阵：**变更位置的行向量有且仅有一次交换**。

最后是部分非对称的  $\mathbf{P}$ ，假设它们主对角线上有  $m$  个 1 ( $m < n$ )，说明进行了  $n - m - 1$  次**不回头**的行交换（即不考虑交换以后又换回去的情况），当它转置后，偏离主对角线的元素数量不变，还是相当于进行了  $n - m - 1$  次**不回头**的行交换，所以行列式的值不会改变，同样有  $\det(\mathbf{P}^T) = \det(\mathbf{P})$ 。



综上所述 (5-1-10) 成立。