

# 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:

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

#### 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  $\mathbf{A}$  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：

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

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  $\mathbf{a}$  & 矩阵  $\mathbf{X}$  的乘法：

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

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：

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

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  $\mathbf{AB} = \mathbf{C}$ ，除了最基础、最抽象的**运算型** (*regular way*) 表示：

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

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

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

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

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

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

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

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**：

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

$$\begin{cases} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = 1 \times \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} + 0 \times \begin{bmatrix} x_{21} & x_{22} \end{bmatrix} \\ \begin{bmatrix} -2 & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = -2 \times \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} + 1 \times \begin{bmatrix} x_{21} & x_{22} \end{bmatrix} \end{cases}$$

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**：

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

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  $\mathbf{B}$ ：

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

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)

```

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

回顾上面那段英文结论，结合这个示例，我们不难发现：

- (1)  $\mathbf{L}(1, 0) = -\mathbf{E}_{to1}(1, 0)$ ;
- (2)  $\mathbf{L}(2, 0) = -\mathbf{E}_{to1}(2, 0)$ ,  $\mathbf{L}(2, 1) = -\mathbf{E}_{to2}(2, 1)$ ;
- (3)  $\mathbf{L}(3, 0) = -\mathbf{E}_{to1}(3, 0)$ ,  $\mathbf{L}(3, 1) = -\mathbf{E}_{to2}(3, 1)$ ,  $\mathbf{L}(3, 2) = -\mathbf{E}_{to3}(3, 2)$ 。

综上所述，在已知所需的行变换操作与  $\mathbf{U}$  时， $\mathbf{A} = \mathbf{LU}$  的效率远高于  $\hat{\mathbf{E}}\mathbf{A} = \mathbf{U}$ 。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  $\mathbf{U}$ 、轻松获取  $\mathbf{L}$  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  $\mathbf{A} \in \mathbb{R}^{n \times n}$  通过初等行变换进行 LU 分解所需的操作次数为：

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

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（ $\mathbf{Ax}^T = \mathbf{b}^T$ ）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: $\mathbf{A} = \mathbf{LU}$

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：

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

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：

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

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

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

## 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。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  $\mathbf{R}$  以及两个特解，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）；

(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。