



计算语言学中的数学方法

2023 级用

作者：王予沛 & ChatGPT & Claude

组织：北京师范大学中文信息处理研究所



封面图片由Midjourney创作

目录

第一章 矩阵与变换	1
1.1 从方程组到矩阵	1
1.1.1 二元一次方程组	1
1.1.2 系数矩阵和增广矩阵	2
1.1.3 高斯-若尔当消元法	3
1.1.4 矩阵的行列式	4
1.2 通用符号	5
1.2.1 向量	5
1.2.2 矩阵	6
1.2.3 单位矩阵	6
1.2.4 对角阵	6
1.2.5 三角矩阵	6
1.3 矩阵与线性变换	7
1.3.1 几种特殊几何变换	7
1.3.1.1 恒等变换	7
1.3.1.2 旋转变换	7
1.3.1.3 反射变换	8
1.3.1.4 伸缩变换	10
1.3.1.5 投影变换	10
1.3.1.6 切变变换	11
1.3.2 线性变换与线性空间	12
1.4 矩阵的运算	16
1.4.1 向量与向量间的乘法运算	16
1.4.2 矩阵与向量间的乘法运算	17
1.4.3 矩阵与矩阵间的乘法运算	19
1.4.4 初等变换与矩阵的乘法	20
1.5 线性方程解的结构	22
1.5.1 线性方程组	22
1.5.2 线性相关与线性无关	22
1.5.3 矩阵的秩	24
1.5.4 线性子空间	25
1.5.5 线性代数基本定理	26
1.6 基础矩阵操作	26
1.6.1 矩阵的转置	26
1.6.2 矩阵的逆	27
第二章 函数与优化	30
2.1 函数的概念和性质	30
2.1.1 函数的定义	30
2.1.2 具有某些特性的函数	30
2.1.2.1 有界函数	30
2.1.2.2 单调函数	31

2.1.2.3 奇函数和偶函数	31
2.1.2.4 周期函数	32
2.1.3 初等函数	32
2.1.4 复合函数	32
2.1.5 反函数	32
附录 A 基本数学工具	34
A.1 求和算子与描述统计量	34

第一章 矩阵与变换

内容提要


- 从方程组到矩阵
- 从线性变换到矩阵

- 矩阵的运算
- 矩阵的性质

参考资料:

- Machine Learning cheatsheets for Stanford's CS 229 - VIP Refresher: Linear Algebra and Calculus;
- Introduction to Linear Algebra, Fifth Edition (2016);
- 高等代数 (第四版), 高等教育出版;
- 高中数学人教 A 版选修 4-2: 矩阵与变换.

1.1 从方程组到矩阵

 **笔记** 本节旨在帮助学生了解矩阵产生的原因, 它源于人们在解方程组时对简洁表达的需求。

1.1.1 二元一次方程组

让我们首先回顾一下二元一次方程组的基本概念。一个二元一次方程组由两个含有两个变量的一次方程组成。一般形式如下:

$$\begin{cases} ax_1 + bx_2 = c \\ dx_1 + ex_2 = f \end{cases} \quad (1.1)$$

其中, x_1 和 x_2 是我们要找的未知数, 而 a 、 b 、 c 、 d 、 e 和 f 是已知的常数。

要解这样的方程组, 我们通常会使用以下两种方法之一:

- 代数方法:** 这种方法包括“代入法”和“消元法”。
 - 在代入法中, 我们从一个方程中解出一个变量的表达式, 然后将其代入另一个方程;
 - 在消元法中, 我们通过加减两个方程来消除一个变量, 从而找到另一个变量的值。
- 图形方法:** 在这种方法中, 我们将每个方程视为二维空间中的一条直线。方程组的解就是这两条直线的交点。

下面, 我们使用**消元法**来解以下二元一次方程组:

$$\begin{cases} 2x_1 + 3x_2 = 5 \\ 4x_1 + 5x_2 = 7 \end{cases} \quad (1.2)$$

我们的目标是找到满足这两个方程的 x_1 和 x_2 的值。我们可以通过以下步骤来实现这一目标:

- 首先, 我们可以使用第一个方程来消去第二个方程中的 x_1 。为此, 我们可以将第一个方程的两边都乘以 2, 得到新的方程:

$$4x_1 + 6x_2 = 10$$

- 接着, 我们将新得到的方程从第二个方程中减去, 从而消去 x_1 :

$$4x_1 + 5x_2 - (4x_1 + 6x_2) = 7 - 10$$

这将给我们一个只包含 x_2 的方程:

$$-x_2 = -3$$

从中我们可以找到 x_2 的值：

$$x_2 = 3$$

3. 事实上, $x_2 = 3$ 可以看成方程, 我们可以使用它来消去第一个方程中的 x_2 :

$$(2x_1 + 3x_2) + (-3) \cdot x_2 = 5 + (-3) \cdot 3$$

这将给我们:

$$2x_1 = 5 - 9$$

从中我们可以找到 x_1 的值:

$$2x_1 = -4$$

$$x_1 = -2$$

所以, 该方程组的解是:

$$x_1 = -2, \quad x_2 = 3$$

在下一节中, 我们将介绍如何使用高斯消元法和矩阵来更简洁地表达求解过程。

使用 Python 可以很轻松地解出方程组 1.2。

```
import numpy as np

# 定义系数矩阵
A = np.array([[2, 3], [4, 5]])

# 定义常数向量
b = np.array([5, 7])

# 使用线性代数模块中的solve函数求解方程组
x = np.linalg.solve(A, b)

# 打印解向量
print("解向量 x:", x)

# 解向量 x: [-2.  3.]
```

 **笔记** 用方程 u 减去方程 v 指: 用方程 u 的左边减去方程 v 的左边, 用方程 u 的右边减去方程 v 的右边。

1.1.2 系数矩阵和增广矩阵

考虑二元一次方程组 (1.2)。我们将方程组中方程的所有系数按照它们在整理好的方程中的位置取出来写成如式 1.3 的一张数表。有了这张数表我们就可以更加简洁地表达方程组中未知数的系数, 因此我们将其称为方程组的 **系数矩阵**, 我们将其记为大写字母 A 。由于矩阵 A 的尺寸是 2×2 , 因此我们称其为 2×2 矩阵, 或者 2 阶 (方形) 矩阵。

$$A = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \quad (1.3)$$

如果我们在系数矩阵右边增加一列 c , 用于记录方程的右端项 (right hand side, RHS), 则方程组 (1.2) 可以写作 $Ax = c$ 。我们用一根竖线隔开表示系数的数表和表示右端项的数列, 得到的整个新数表称为方程组 1.2 的 **增**

广矩阵，如式1.4所示：

$$[A | c] = \left(\begin{array}{cc|c} 2 & 3 & 5 \\ 4 & 5 & 7 \end{array} \right) \quad (1.4)$$

下面，我们对增广矩阵实施第1.1.1节中所使用的消元法，从而达到解方程的目的：

1. 首先，我们用第一行的系数来将第二行的系数化为0，具体操作如下：
 - (a). 用第二行的第一个系数4除以第一行的第一个系数2，得到商数2；
 - (b). 将第一行整体乘以此商数（2），得到（4, 6, 10）；
 - (c). 再用第二行减去第一行，则第二行的第一个元素变为0。这时增广矩阵变为：

$$\left(\begin{array}{cc|c} 2 & 3 & 5 \\ 0 & -1 & -3 \end{array} \right) \quad (1.5)$$

- (d). 在第二行两边乘以-1
2. 现在用第二行中 x_2 的系数来将第一行中 x_2 的系数化为0，具体操作如下：
 - (a). 用第一行的第二个系数3除以第二行的第二个系数1，得到商数3；
 - (b). 将第二行整体乘以此商数（3），得到（0, 3, 9）；
 - (c). 再用第一行减去第二行，则第一行的第二个元素变为0。这时增广矩阵变为：

$$\left(\begin{array}{cc|c} 2 & 0 & -4 \\ 0 & 1 & 3 \end{array} \right)$$

- (d). 在第一行两边乘以 $\frac{1}{2}$ ，得到增广矩阵：

$$\left(\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 3 \end{array} \right) \quad (1.6)$$

该式还原为方程组为

$$\begin{cases} 1 \cdot x_1 + 0 \cdot x_2 = -2 \\ 0 \cdot x_1 + 1 \cdot x_2 = 3 \end{cases}$$

最终得到 $x_1 = -2$ ， $x_2 = 3$ ，这就是原方程组的解。

矩阵提供了一种组织方程组系数和常数项的方式，通过矩阵运算可以方便地求解方程组。这启发了矩阵理论的产生和发展。

1.1.3 高斯-若尔当消元法

回顾我们在第1.1.2节中的操作，我们在增广矩阵上进行消元法时，仅仅使用了两种操作：

- 把某行乘以一非零常数；
- 把第 i 行加上第 j 行的 k 倍；

可见这些操作都是对矩阵的“行”进行的，因此我们称它们为行变换。

注意到，在进行消元法的过程中，我们得到的增广矩阵1.5的非零元素看起来像是一个“倒过来的楼梯”，其第一行左起第一个非零元下方都是零，第二行左起第一个非零元在第一行左起第一个非零元的右侧，我们称这样的矩阵是行阶梯形矩阵。

定义 1.1 (行阶梯形矩阵)

如果一个矩阵满足下面的条件，我们称其为行阶梯形矩阵（Row Echelon Form）：

1. 所有仅包含零的行都位于矩阵的最底部；
2. 在每一个非零行中，最左边的非零元素（称主元）总是出现在上一行主元的右侧；
3. * 包含非零元素的行中，首个非零元素必须为1.

其中第三个条件是可选的，不同文献中有不同的定义方式，不必纠结。

事实上, 得到增广矩阵1.5后, 我们已经不再需要继续消元了。从第二行我们可以很容易的得到 $-x_2 = -3$, 从而解出 $x_2 = 3$ 。从第一行我们得到方程 $2x_1 + 3x_2 = 5$, 这时候我们使用回代法, 将 $x_2 = 3$ 代到方程 $2x_1 + 3x_2 = 5$ 中得到 $2x_1 + 3 \cdot 3 = 5$, 从而解出 $x_1 = -2$ 。

一般地, 我们称“先通过行变换将矩阵化为行阶梯矩阵, 然后直接通过回代法找到方程组的解”这样的解方程的方法为高斯消元法。可以证明, 任何矩阵都可以通过高斯消元法转换为简化行阶梯形式

此外, 当消元法进行到尾声时, 我们得到的增广矩阵1.6非常有特点, 其每一行的第一个非零元素全是1, 且这些行的第一个元素1所在列的其余元素全为零, 我们称这样的矩阵是简化行阶梯矩阵。显然, 根据简化行阶梯矩阵我们很容易得到方程组的解。

定义 1.2 (简化行阶梯形)

如果一个矩阵满足下面的条件, 我们称其为简化行阶梯形 (Reduced Row Echelon Form, RREF):


1. 所有零行都位于矩阵的底部;
2. 在每一个非零行中, 第一个非零元素 (称主元) 总是出现在上一行主元的右侧;
3. 每一个非零行的主元都是1;
4. 一个主元所在的列的其他元素都是0。

简化行阶梯形和行阶梯形矩阵的主要区别在于最后一条。



一般地, 我们称“先通过行变换将矩阵化为行阶梯矩阵, 然后进一步做行变换, 将其化为简化行阶梯矩阵进而直接读出方程组的解”这样的解方程的方法为高斯-若尔当消元法。可以证明, 任何矩阵都可以通过高斯-若尔当消元法转换为简化行阶梯形式。高斯-若尔当消元法实际上是高斯消元法的一个扩展。需要指出的是, 虽然我们这里借助高斯消元法行变换来引入行变换, 但是行变换并不是只能在增广矩阵上进行, 因为行变换是一种矩阵变形的方法, 并不限于某种特殊的矩阵。

1.1.4 矩阵的行列式

 **笔记** 行列式来自于数学家们在求解方程组时发现的一个可以用来简化表达, 并判断方程组是否有解的记号。

考虑二元一次方程组 (1.1), 我们使用高斯-若尔当消元法求出该方程的解。

$$\left[\begin{array}{cc|c} a & b & c \\ d & e & f \end{array} \right] \rightarrow \left[\begin{array}{cc|c} 1 & \frac{b}{a} & \frac{c}{a} \\ 0 & e - \frac{bd}{a} & f - \frac{cd}{a} \end{array} \right] \rightarrow \left[\begin{array}{cc|c} 1 & 0 & \frac{ce-bf}{ae-bd} \\ 0 & 1 & \frac{af-cd}{ae-bd} \end{array} \right]$$

如果我们定义 $D = ae - bd, D_1 = ce - bf, D_2 = af - cd$, 我们可以将原方程的结果表示为:

$$x_1 = \frac{D_1}{D}, \quad x_2 = \frac{D_2}{D} \quad (1.7)$$

这个结果称为二元一次方程组的克莱姆法则 (Cramer's Rule)。其中 D 称为矩阵 $\begin{bmatrix} a & b \\ d & e \end{bmatrix}$ 的行列式, 记作

$$D = \det \left(\begin{bmatrix} a & b \\ d & e \end{bmatrix} \right) = \begin{vmatrix} a & b \\ d & e \end{vmatrix} = ae - bd \quad (1.8)$$

同理有,

$$D_1 = \det \left(\begin{bmatrix} c & b \\ f & e \end{bmatrix} \right) = \begin{vmatrix} c & b \\ f & e \end{vmatrix} = ce - bf, \quad D_2 = \det \left(\begin{bmatrix} a & c \\ d & f \end{bmatrix} \right) = \begin{vmatrix} a & c \\ d & f \end{vmatrix} = af - cd \quad (1.9)$$

考虑一个特殊的方程组

$$\begin{cases} x_1 + x_2 = 5 \\ x_1 + x_2 = 7 \end{cases} \quad (1.10)$$

显然, 该方程组无解, 其中 $D = 0$ 。可见, 凡是系数矩阵的行列式 $D = 0$ 的二元一次方程组都是无解的。

一般地, 考虑任意 n 元一次方程组 (矩阵表示):

$$Ax = c \quad (1.11)$$

其中的 A 是一个 $n \times n$ 的方阵, 而向量 $x = (x_1, x_2, \dots, x_n)^T$ 和 $c = (c_1, c_2, \dots, c_n)^T$ 均为长度为 n 的列向量。

克莱姆法则说明: 如果 $\det A \neq 0$, 那么方程 (1.11) 有解 $x = (x_1, x_2, \dots, x_n)^T$, 其中

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

当中 x_i 是列向量 x 的第 i 行, A_i 是用列向量 c 取代 A 的第 i 列后得到的矩阵。

注意到, 我们仅定义了二阶方阵的行列式, 那么对于任意一个 $n \times n$ 的方阵 A , 如何计算 $\det(A)$ 呢? 具体的计算方式依赖于行列式的三种定义:

- 用排列和逆序数定义 (国内大多数教材上都用这种定义);
- 利用代数余子式和按第一行展开进行归纳定义;
- 利用其几何意义 (有向面积在一般的欧几里得空间中的推广) 进行定义。

本讲义中我们希望大家按照几何意义来理解行列式, 我们可以将其视作在一定的线性变换 (比如旋转、缩放) 下, 形状的面积 (在二维空间) 或体积 (在三维空间) 的缩放因子。想象你有一张纸, 上面画了一个单位正方形。这个正方形的边长都是 1, 所以它的面积也是 1。现在, 如果我们用一个矩阵来表示一个变换, 比如把这个正方形拉长或者压扁, 或者同时进行拉长和压扁的组合, 那么变换后的图形可能就不再是正方形, 而是一个矩形或任意四边形。

行列式的值, 就是这个变换后的图形的面积与原来单位正方形面积的比值。如果行列式的值大于 1, 那么面积被放大了; 如果行列式的值小于 1 但大于 0, 那么面积被缩小了; 如果行列式的值是负数, 这意味着除了面积或体积的缩放, 还发生了一个“翻转”操作, 比如在二维空间中, 图形被沿原点镜像反转了。

在三维空间中, 这个解释同样适用, 但这次是关于体积的缩放而不仅仅是面积。如果你有一个单位立方体, 通过一个矩阵表示的线性变换后, 立方体可能会变成任意的平行六面体。这时候, 行列式的绝对值就告诉我们变换后的平行六面体的体积是如何相对于单位立方体的体积进行缩放的。三阶行列式计算比较好记, 可以顺手补充一下。

简而言之, 行列式给我们提供了一个数值, 这个数值说明了在一个线性变换下, 形状的尺寸如何改变: 面积或体积是如何扩大或缩小的, 以及是否发生了方向的翻转。高阶行列式的具体计算比较复杂, 我们可以使用 NumPy 代劳, 我们只需要理解其意义即可。

1.2 通用符号

 **笔记** 本节最重要却最容易忽略的一点: 数学文本中, 如无特殊说明, 则一般认为向量都是列向量。

1.2.1 向量

我们记 $x \in \mathbb{R}^n$ 为一个 n 维的向量, 其中 $x_i \in \mathbb{R}$ 是第 i 维的元素:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

1.2.2 矩阵

我们记 $A \in \mathbb{R}^{m \times n}$ 为一个 m 行 n 列的矩阵, 其中 $a_{i,j} \in \mathbb{R}$ 是第 i 行 j 列的元素:

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

注意: 如上定义的向量 \mathbf{x} 可以被看做是一个 $n \times 1$ 的矩阵, 常被称为一个列向量。

1.2.3 单位矩阵

单位矩阵 $I \in \mathbb{R}^{n \times n}$ 是一个方形矩阵 (或简称为方阵), 其对角线的元素上均为 1, 其余位置均为 0。

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

1.2.4 对角阵

对角阵 $D \in \mathbb{R}^{n \times n}$ 是一个方阵其对角线上元素均是非零值, 其余位置均为 0。

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

注: 我们记 D 为 $\text{diag}(d_1, \dots, d_n)$ 。

1.2.5 三角矩阵

三角矩阵 (triangular matrix) 因其非零系数的排列呈三角形状而得名。

一个如下形状的方形矩阵:


$$\mathbf{L} = \begin{bmatrix} l_{1,1} & & \cdots & & 0 \\ l_{2,1} & l_{2,2} & & (0) & \\ l_{3,1} & l_{3,2} & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \\ l_{n,1} & l_{n,2} & \cdots & l_{n,n-1} & l_{n,n} \end{bmatrix}$$

被称为下三角矩阵; 同样的, 一个如下形状的方形矩阵:

$$\mathbf{U} = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \cdots & u_{1,n} \\ & u_{2,2} & u_{2,3} & \cdots & u_{2,n} \\ \vdots & & \ddots & \ddots & \vdots \\ & (0) & & \ddots & u_{n-1,n} \\ 0 & & \cdots & & u_{n,n} \end{bmatrix}$$

被称为上三角矩阵。值得注意的是, 三角矩阵一定是方阵。

1.3 矩阵与线性变换

 **笔记** 矩阵与线性变换通过一次方程组建立联系。事实上, 在直角坐标系 Oxy 中, 所有满足“线性”性质的几何变换都具有形式

$$\begin{cases} x' = ax + by, \\ y' = cx + dy, \end{cases}$$

其中 a, b, c, d 均为常数. 该变换可以由二阶矩阵 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 完全确定.

本节将以矩阵为工具, 研究一些几何变换, 并以平面图形的变换为背景, 讨论二阶矩阵的乘法及性质, 用变换的观点理解二元一次方程组的意义, 初步展示矩阵应用的广泛性。

1.3.1 几种特殊几何变换

1.3.1.1 恒等变换

相传, 在古希腊德尔斐的阿波罗神庙中, 刻有“认识你自己”这句箴言, 它鼓励人们深入探索自我, 以实现更深层次的理解和个人成长。这句古老的箴言也在数学中找到了回音。我们在定义任何数学对象之间的关系时, 一定要先思考这种关系在对象和它自身之间是否是存在的。比如, 实数相等和不相等的关系, 就需要建立在每个实数都和它自身相等的基础上。

我们在研究平面上的几何变换时也有类似的概念, 这就是本节要介绍的**恒等变换**, 它是一种将一个对象变换为其自身的变换。这意味着, 在这种变换下, 对象保持不变。

在直角坐标系 Oxy 内, 恒等变换将每个点都映射为它自身 (即相当于没有发生变化)。设点 $P(x, y)$ 经过恒等变换后变成点 $P'(x', y')$, 则 x', y' 与 x, y 有如下关系:

$$\begin{cases} x' = x, \\ y' = y. \end{cases} \quad (1.12)$$

我们将式 (1.12) 称为恒等变换的表达式。对式 1.12 进行简单改写, 如式 1.13 所示:

$$\begin{cases} x' = x + 0 \cdot y, \\ y' = 0 \cdot x + y. \end{cases} \quad (1.13)$$

故其系数矩阵为:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (1.14)$$

可见, 恒等变换的表达式系数矩阵正好是一个单位矩阵, 因此我们将单位矩阵称为恒等变换的**变换矩阵**。单位矩阵也唯一确定了一种变换, 即恒等变换。因此, 在讨论数学问题时, 你常常能听到人们指着单位矩阵说: “这是一个恒等变换”。

1.3.1.2 旋转变换

在第 1.3.1.1 节中, 我们已经得到了平面上的点到它自身的变换, 现在我们可以开始研究平面上的点到其他与它不同的点的变换了。

如图 1.1 所示, 在直角坐标系 Oxy 内, 每个点都绕原点 O 按逆时针方向旋转 180° 。设点 $P(x, y)$ 经过旋转后变成点 $P'(x', y')$, 则 x', y' 与 x, y 有如下关系:

$$\begin{cases} x' = -x, \\ y' = -y. \end{cases} \quad (1.15)$$

我们将式 (1.15) 称为旋转角为 180° 的**旋转变换**的表达式, 它建立了平面内的每个点 P 到其对应点 P' 的对应关系, 我们称 P' 是 P 在这个旋转变换作用下的像.

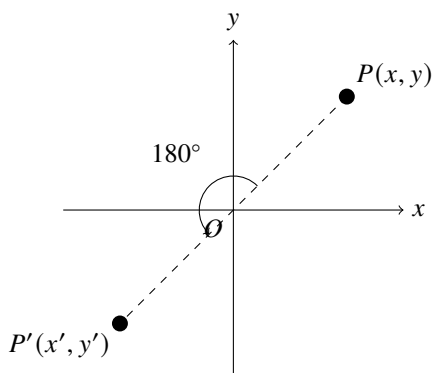


图 1.1: 旋转变换示意图

我们可以对式 1.15 进行简单改写, 如式 1.16 所示:

$$\begin{cases} x' = -1 \cdot x + 0 \cdot y, \\ y' = 0 \cdot x + -1 \cdot y. \end{cases} \quad (1.16)$$

故其系数矩阵为:

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1.17)$$

显然, 在二维平面内旋转 180° 这一变换可以由式 1.17 唯一确定. 反过来, 该矩阵也可以唯一确定这个变换.

练习 1.1

在直角坐标系 Oxy 内, 将每个点绕原点 O 按逆时针方向旋转 30° 的变换称为旋转角是 30° 的旋转变换.

- 求点 $A(1, 0)$ 在这个旋转变换作用下的像 A' ;
- 写出这个旋转变换的表达式.

解

- 点 $A(1, 0)$ 在这个旋转变换作用下的像为 $A'(\frac{\sqrt{3}}{2}, \frac{1}{2})$;
- $$\begin{cases} x' = \frac{\sqrt{3}}{2}x - \frac{1}{2}y, \\ y' = \frac{1}{2}x + \frac{\sqrt{3}}{2}y. \end{cases}$$

事实上, 在直角坐标系 Oxy 内, 将任意一点 $P(x, y)$ 绕原点 O 按逆时针方向旋转 θ 角得到点 $P'(x', y')$ 的旋转变换的坐标变换公式是

$$\begin{cases} x' = x \cos \theta - y \sin \theta, \\ y' = x \sin \theta + y \cos \theta. \end{cases} \quad (1.18)$$

1.3.1.3 反射变换

如图 1.2 所示, 在直角坐标系 Oxy 内, 关于 x 轴的反射变换把任意一点 $P(x, y)$ 变成它关于 x 轴的对称点 $P'(x', y')$. 此时 x', y' 与 x, y 有如下关系:

$$\begin{cases} x' = -x, \\ y' = y. \end{cases} \quad (1.19)$$

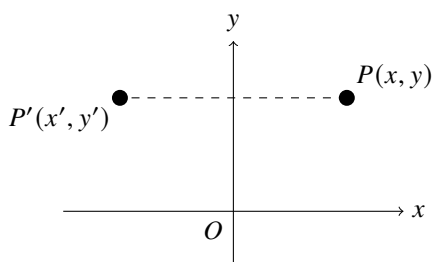


图 1.2: 反射变换示意图

我们将式 (1.19) 称为关于 x 轴的**反射变换**的表达式。我们可以对式 1.19 进行简单改写，如式 1.20 所示：

$$\begin{cases} x' = 1 \cdot x + 0 \cdot y, \\ y' = 0 \cdot x + (-1) \cdot y. \end{cases} \quad (1.20)$$

故其系数矩阵为：

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1.21)$$

练习 1.2

- 关于 y 轴的反射变换把直角坐标系 Oxy 内的任意一点 $P(x, y)$ 变成它关于 y 轴的对称点 $P'(x', y')$. 相应的坐标变换公式和变换矩阵分别是什么?
- 关于 直线 $y = x$ 的反射变换把直角坐标系 Oxy 内的任意一点 $P(x, y)$ 变成它关于直线 $y = x$ 的对称点 $P'(x', y')$. 相应的坐标变换公式和变换矩阵分别是什么?

解

- 关于 y 轴的反射变换，其相应的坐标变换公式是

$$\begin{cases} x' = -x, \\ y' = y. \end{cases}$$

对应的变换矩阵是：

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

- 关于 直线 $y = x$ 的反射变换，其相应的坐标变换公式是

$$\begin{cases} x' = y, \\ y' = x. \end{cases}$$

对应的变换矩阵是：

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

1.3.1.4 伸缩变换

如图1.3所示, 在直角坐标系 Oxy 内, 将任意一点 $P(x, y)$ 的纵坐标变为原来的两倍, 横坐标不变, 得到点 $P'(x', y')$ 。此时 x', y' 与 x, y 有如下关系:

$$\begin{cases} x' = x \\ y' = 2y \end{cases} \quad (1.22)$$

像这样的, 将直角坐标系 Oxy 内每个点的横坐标变为原来的 k_1 倍, 纵坐标变为原来的 k_2 倍, 其中 k_1, k_2 均为非零常数, 我们称这样的几何变换为**伸缩变换** (stretching), 式 (1.22) 称为该伸缩变换的表达式。我们可以对式 (1.22) 进行改写, 得到式:

$$\begin{cases} x' = x + 0 \cdot y \\ y' = 0 \cdot x + 2y \end{cases} \quad (1.23)$$

故其系数矩阵为:

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad (1.24)$$

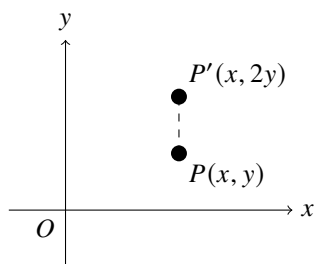


图 1.3: 伸缩变换示意图

练习 1.3

将直角坐标系 Oxy 内每个点的横坐标变为原来的 k_1 倍, 纵坐标变为原来的 k_2 倍 (k_1, k_2 均为非零常数) 的伸缩变换, 其坐标变换公式是

$$\begin{cases} x' = k_1 x, \\ y' = k_2 y. \end{cases} \quad (1.25)$$

对应的变换矩阵为

$$\begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix}. \quad (1.26)$$

注意到, 当 $k_1 = k_2$ 时, 伸缩变换退化为**平移变换**。

1.3.1.5 投影变换

设 l 是平面内一条给定的直线. 对平面内的任意一点 P 作直线 l 的垂线, 垂足为点 P' , 则称点 P' 为点 P 在直线 l 上的**投影** (projection). 将平面上每一点 P 变成它在直线 l 上的投影 P' , 这个变换称为关于直线 l 的**投影变换**.

如图 1.4, 在直角坐标系 Oxy 内, 过任意一点 P 作 x 轴的垂线, 垂足为点 P' , 我们称点 P' 为点 P 在 x 轴上的投影. 如果一个变换把直角坐标系内的每一点变成它在 x 轴上的投影, 那么称这个变换为关于 x 轴的投影变换.

设在关于 x 轴的投影变换的作用下, 点 $P(x, y)$ 变成点 $P'(x', y')$ 。此时 x', y' 与 x, y 有如下关系:

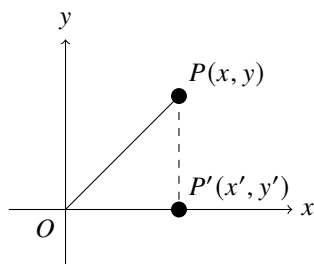


图 1.4: 投影变换示意图

$$\begin{cases} x' = x, \\ y' = 0. \end{cases}$$

即

$$\begin{cases} x' = 1 \cdot x + 0 \cdot y, \\ y' = 0 \cdot x + 0 \cdot y. \end{cases}$$

对应的二阶矩阵为 $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

1.3.1.6 切变变换

切变变换 (shearing) 是一种特殊的几何变换, 用于改变物体的形状。它在一个方向上按照一定的比例“拉伸”或“压缩”物体, 同时在与这个方向垂直的另一方向上保持物体不变。在二维平面上, 切变变换主要有两种:

- **水平切变**: 在水平方向上拉伸或压缩物体, 与物体在垂直方向上的位置成比例。垂直方向上的形状不会改变。如图 1.5¹ 所示。
- **垂直切变**: 在垂直方向上拉伸或压缩物体, 与物体在水平方向上的位置成比例。水平方向上的形状不会改变。

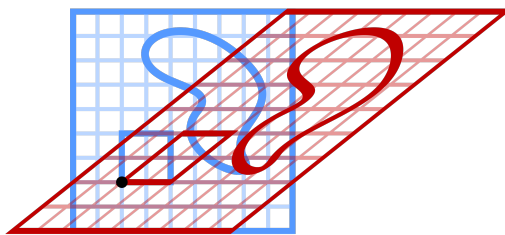


图 1.5: 平面的水平剪切, 将蓝色转变为红色, 黑点为原点

如图 1.6, 在直角坐标系 Oxy 内, 将每一点 $P(x, y)$ 沿着与 x 轴平行的方向平移 ky 个单位变成点 P' , 其中 k 是非零常数, 这就是水平切变 (或平行于 x 轴的切变变换)。设 $P'(x', y')$, 则此时 x', y' 与 x, y 有如下关系:

$$\begin{cases} x' = x + ky, \\ y' = y. \end{cases}$$

从而, 对应的二阶矩阵为 $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$.

练习 1.4

¹ 图源: https://en.wikipedia.org/wiki/Shear_mapping

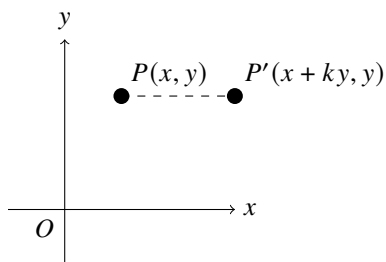


图 1.6: 切变变换示意图

平行于 y 轴的切变变换是指将直角坐标系内的每一点 $P(x, y)$ 沿着与 y 轴平行的方向平移 kx 个单位 (其中 k 是非零常数) 的线性变换. 其坐标变换公式为

$$\begin{cases} x' = x, \\ y' = kx + y. \end{cases}$$


对应的矩阵为 $\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$.

1.3.2 线性变换与线性空间

从高中开始我们就熟知平面直角坐标系, 这是由两条相互平行、且各自向正、负两个方向无限延伸的横轴、纵轴构成的坐标体系. 横轴和纵轴均可以在实数域 \mathbb{R} 上任意取值, 故我们可以将平面记作 $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. 通过第 1.3.1 节的例子, 我们了解到, 二维空间上的点之间可以通过几何变换到达彼此, 同时, 所得结果也一定还是二维空间中的点. 如果我们从原点出发, 引一条到特定点的有向箭头, 就可以将点表示为向量. 因此, 之前所列举的几何变换做的事情都是将平面向量变换为其他平面向量. 高中时, 我们学习过平面向量的加法和数量乘法, 我们先来复习一下.

定义 1.3 (平面向量的加法)

设向量 $\mathbf{u} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$, 向量 \mathbf{u} 与 \mathbf{v} 的加法满足平行四边形法则, 数学表示为 $\mathbf{u} + \mathbf{v} = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}$.

 **笔记** 向量加法就是高中学的平行四边形法则。

定义 1.4 (平面向量的数量乘法)

设向量 $\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$, 规定实数 λ 与向量 \mathbf{u} 的乘积 $\lambda\mathbf{u} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$;

平面向量的数量乘法之所以被称为“数量乘法”是因为它涉及到一个向量与一个数 (或者我们可以称之为标量) 相乘的操作. 这个数可以是任何实数. 当我们将一个向量与一个数相乘时, 结果是一个新的向量, 它的方向与原始向量相同 (或相反), 但长度 (或称为模) 被缩小或拉大了. 这个操作影响到向量的大小, 所以我们称之为“数量乘法”。

练习 1.5

- 【高中数学人教 A 版选修 4-2】设向量 $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. 如图 1.7, 先利用平行四边形法则求 $\mathbf{u} + \mathbf{v}$, 再对向量 $(\mathbf{u} + \mathbf{v})$ 进行关于 x 轴的反射变换; 或者, 先对向量 \mathbf{u}, \mathbf{v} 做关于 x 轴的反射变换, 再利用平行四边形法则求反射变换后的两个向量的和. 请回答下列问题:

- 这两个过程的结果相同吗?
- 如果把 \mathbf{u} 关于 x 轴的反射变换记作 $\text{Reflect}_x(\mathbf{u})$, 把 \mathbf{v} 关于 x 轴的反射变换记作 $\text{Reflect}_x(\mathbf{v})$, 把 $\mathbf{u} + \mathbf{v}$

关于 x 轴的反射变换记作 $\text{Reflect}_x(u + v)$, 那么, 这些变换的坐标变换矩阵分别是什么? 这些变换的矩阵有什么关系?

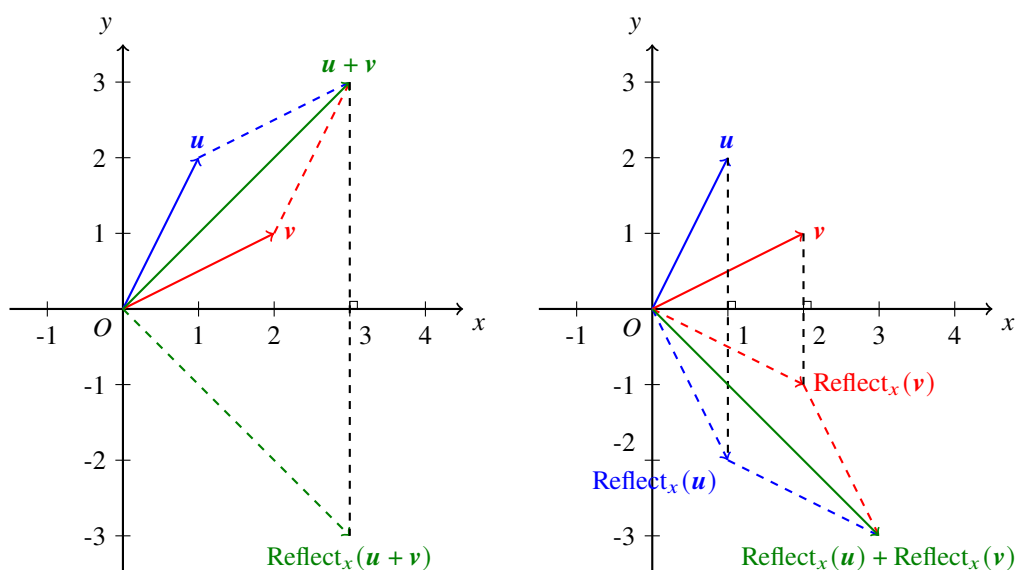


图 1.7: 向量加法与反射变换

2. 设向量 $u = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, 如图 1.8, 先用 2 与该向量进行数量乘法, 再使向量在 x -轴方向的坐标变为原来的 3 倍, 同时使 y -轴方向的坐标变为原来的 2 倍; 或者, 先使向量在 x -轴方向的坐标变为原来的 3 倍, 同时使 y -轴方向的坐标变为原来的 2 倍, 再用 2 与该向量进行数量乘法。请回答下列问题:
- 这两个过程的结果相同吗?
 - 如果把 u 的伸缩变换记作 $\text{Stretch}(u)$, 把 $2u$ 的伸缩变换记作 $\text{Stretch}(2u)$, 那么, 这些变换的坐标变换矩阵分别是什么? 这些变换的矩阵有什么关系?

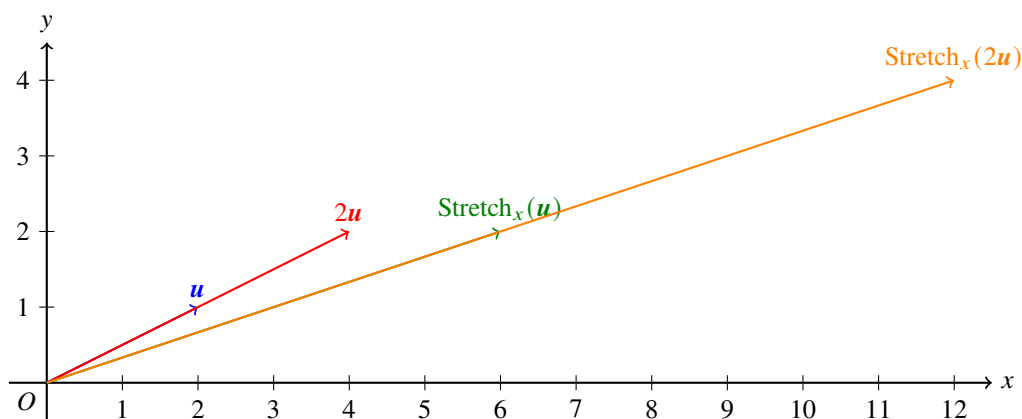



图 1.8: 向量伸缩变换

现在, 我们希望将第 1.3.1 节中的几何变换抽象为更加一般的形式, 这些一般的变换可能没有类似反射变换、旋转变换那样明确的意义, 但能够保持几何图形的一些重要性质, 如直线的形状、直线之间的平行性、相交直线的交点以及直线上不同点之间的相对距离等。

 **笔记** 让学生回顾之前展示每个变换, 并提醒他们每个变换都保持了直线的性质。

命题 1.1 (平面上保持直线性质的变换)

设 T 是一个平面到平面的映射 ($T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$), \mathbf{u}, \mathbf{v} 是平面上的任意两个向量, λ 是一个任意实数。若从原点引出向量 \mathbf{u}, \mathbf{v} , 其末端点分别记为 P_1, P_2 , 显然, 只要选取合适的 \mathbf{u}, \mathbf{v} , P_1, P_2 就可以确定出平面上的任意直线。如果 T 满足下面的条件:

1. 保加法: $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$;
2. 保数乘: $T(\lambda \mathbf{u}) = \lambda T(\mathbf{u})$.

则该映射不改变直线的形状、平行性、与其他直线的交叉点的位置以及直线上点的相对距离。

证明

1. 不改变直线的形状: 直线可以表示为向量的线性组合, 即对于两个向量 \mathbf{u} 和 \mathbf{v} , 直线可以表示为 $\mathbf{u} + t\mathbf{v}$, 其中 t 是标量。根据变换 T 的定义, 我们有:

$$\begin{aligned} T(\mathbf{u} + t\mathbf{v}) &= T(\mathbf{u}) + T(t\mathbf{v}) \\ &= T(\mathbf{u}) + tT(\mathbf{v}) \end{aligned}$$

即直线在经过这样的变换后仍然是一条直线。

2. 保持平行性: 考虑两条平行直线, 它们可以表示为 $\mathbf{u} + t\mathbf{v}$ 和 $\mathbf{u}' + t\mathbf{v}$, 注意这两条直线共享同一个方向向量 \mathbf{v} 。因为变换 T 保持向量加法和标量乘法, 变换后的直线分别是 $T(\mathbf{u}) + tT(\mathbf{v})$ 和 $T(\mathbf{u}') + tT(\mathbf{v})$ 。因为它们仍然共享同一个方向向量 $T(\mathbf{v})$, 所以变换后的直线保持平行。

3. 保持交叉点位置: 假设两条直线 L_1 和 L_2 在点 \mathbf{p} 交叉, 即存在某个 t_1 和 t_2 使得 $\mathbf{u} + t_1\mathbf{v} = \mathbf{u}' + t_2\mathbf{v} = \mathbf{p}$ 。应用变换 T , 我们得到 $T(\mathbf{u} + t_1\mathbf{v}) = T(\mathbf{u}') + t_2T(\mathbf{v}) = T(\mathbf{p})$ 。这表明变换后的直线在 $T(\mathbf{p})$ 点交叉, 保持了交叉点的位置。

4. 保持直线上点的相对距离: 考虑直线 L 上的两点 \mathbf{p} 和 \mathbf{q} , 直线上的任意一点 \mathbf{x} 可以表示为 \mathbf{p} 和 \mathbf{q} 的线性组合:

$$\mathbf{x} = \alpha \mathbf{p} + (1 - \alpha) \mathbf{q}$$

其中, α 是一个标量。我们将线性变换 T 应用于直线 L 上的点 \mathbf{x} :

$$\begin{aligned} T(\mathbf{x}) &= T(\alpha \mathbf{p} + (1 - \alpha) \mathbf{q}) \\ &= \alpha T(\mathbf{p}) + (1 - \alpha) T(\mathbf{q}) \end{aligned}$$

这表明, 经过线性变换 T 后, 点 \mathbf{x} 依然位于 $T(\mathbf{p})$ 和 $T(\mathbf{q})$ 定义的直线上, 并且保持了与原始直线相同的比例 α 。

到这里, 我们证明了平面 (\mathbb{R}^2) 上保加法和保数乘的变换的向量映射 (或称变换) 不会改变直线的形状、直线之间的平行和相交关系, 也不会改变直线上的点之间的相对距离, 我们称这些性质是平面的“线性”性质, 而变换 T 可以在变换前后保持这种“线性”性质, 所以, 我们称变换 T 为平面上的线性变换。

可以验证, 在三维空间 (\mathbb{R}^3) 中, 保加法和保数乘的变换仍然可以保持空间的线性性质, 这是三维空间上的线性变换 (到了 \mathbb{R}^3 中, 线性变换不仅能够直线的性质不变, 还可以保持平面的性质不变)。现在, 我们将研究对象从我们熟悉的二维、三维空间推广到任意维度。我们注意到, 在二维空间中, 变换 T 映射的结果向量一定还在二维空间中; 在三维空间中, 变换 T 映射的结果向量一定还在三维空间中。据此, 我们在心中勾勒出一个任意维空间的“印象”, 我们希望, 在这样的空间中, 变换 T 映射的结果向量仍然属于同一空间, 这样, 我们可以得到 $\mathbb{R}^4, \mathbb{R}^5$ 直到 \mathbb{R}^n 。在 Gilbert Strang 的线性代数教材中就直接使用这种“印象”来定义 \mathbb{R}^n 。

定义 1.5 (\mathbb{R}^n)

\mathbb{R}^n 是包含所有具有 n 个分量的向量的空间。

The space \mathbb{R}^n consists of all column vectors \mathbf{v} with n components.

——Introduction to Linear Algebra by Gilbert Strang

可以证明, 1.1 中的变换 T 在 \mathbb{R}^n 仍然能够保持空间的线性性质, 所以变换 T 也是 \mathbb{R}^n 中的线性变换。在 \mathbb{R}^n 中, 如果要保证线性变换的结果向量仍然在 \mathbb{R}^n 中, 需要满足如下的条件:

- 对向量加法满足:
 - 加法交换律 $\alpha + \beta = \beta + \alpha$;
 - 加法结合律 $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
 - 存在零元, 因为要保证 $\exists x, \text{s.t. } \alpha + x = \alpha$, 这时候 x 就是零元, 记为 0 ;
 - 存在逆元, 因为要保证 $\exists x, \text{s.t. } \alpha + x = 0$, 这时候 x 就是 α 的逆元, 记为 $-\alpha$;
- 对向量数乘满足:
 - 标量间乘法和数乘的结合性: $k(l\alpha) = (kl)\alpha$;
 - 存在数乘单位元: $1\alpha = \alpha$;
- 对向量加法和数乘满足:
 - 数乘对向量加法的分配律: $k(\alpha + \beta) = k\alpha + k\beta$;
 - 数乘对标量间加法的分配律: $(k + l)\alpha = k\alpha + l\alpha$;

定义 1.6 (线性空间)

设 V 是一个非空集合, \mathcal{F} 是一个数域, 在数域 \mathcal{F} 中, 数的加法和乘法能够顺利按照我们小学所学习的规则进行。在集合 V 的元素之间定义了一种代数运算, 叫做加法; 即给出了一个法则, 对于 V 中任意两个元素 α 与 β , 在 V 中都有唯一的一个元素 γ 与它们对应, 称为 α 与 β 的和, 记为 $\gamma = \alpha + \beta$ 。在数域 \mathcal{F} 与集合 V 的元素之间还定义了一种运算, 叫做数量乘法; 这就是说, 对于数域 \mathcal{F} 中任一数 k 与 V 中任一元素 α , 在 V 中都有唯一的一个元素 δ 与它们对应, 称为 k 与 α 的数量乘积, 记为 $\delta = k\alpha$ 。如果加法与数量乘法满足上面 8 条规则, 那么 V 称为数域 \mathcal{F} 上的线性空间 (也称向量空间)。



定义中, 数域这一概念的严格定义一般在抽象代数课程里讨论, 我们这里将其理解为一个保证加法和乘法能够顺利进行的数的集合即可。可以将其想象为一种“通用货币系统”, 在这个系统里, 我们可以自由地进行交易 (加法和乘法运算) 而不用担心交换规则突然改变。就像在日常生活中, 无论我们是在商店购物还是在银行存取款项, 货币的基本价值和使用规则保持不变, 使得经济交易可以顺利进行。



笔记 线性空间的定义可以简记为: 对加法和数量乘法满足封闭性 + 8 条公理。实际上这个定义是有冗余的, 因为满足封闭性就必须要有 8 条公理。

为什么要煞有介事的重新定义加法和数量乘法, 显得多此一举, 直接用平面向量的加法和数量乘法不行吗? 确实不行, 因为定义 1.6 中集合 V 的元素不一定是数字或我们上面讨论的各种空间中的向量, 即我们希望传达一个观点, 一大类数学对象各自形成的集合都是线性空间。下面举几个例子:

1. 多项式空间。 多项式空间是所有具有特定最高次数的多项式的集合。例如, 考虑所有最高次数为 2 的实系数多项式的集合, 记为 $P_2(\mathbb{R})$ 。这个空间包括形如 $ax^2 + bx + c$ 的所有多项式, 其中 a, b, c 是实数系数。

- 向量加法: 两个多项式相加, 就是它们对应系数的和。例如, 如果我们有 $p_1(x) = 2x^2 + 3x + 1$ 和 $p_2(x) = x^2 - 2x + 4$, 它们的和 $p_1(x) + p_2(x) = (2x^2 + x^2) + (3x - 2x) + (1 + 4) = 3x^2 + x + 5$ 也在 $P_2(\mathbb{R})$ 中。
- 标量乘法: 一个多项式乘以一个标量, 就是每个系数乘以这个标量。例如, 如果 $k = 3$ 和 $p_1(x) = 2x^2 + 3x + 1$, 则 $k p_1(x) = 3(2x^2 + 3x + 1) = 6x^2 + 9x + 3$ 。

2. 实函数空间。 考虑所有定义在实数上, 取值也为实数的连续函数的集合, 记为 $C(\mathbb{R})$ 。这是一个非常广泛的向量空间, 包括如直线、抛物线等基本函数。

- 向量加法: 两个函数的和是一个新的函数, 其值在每一点上都是原来两个函数值的和。例如, 如果 $f(x) = x^2$ 和 $g(x) = 2x$, 那么它们的和 $h(x) = f(x) + g(x) = x^2 + 2x$ 。
- 标量乘法: 一个函数乘以一个标量, 结果是一个新的函数, 其值在每一点上都是原来函数值乘以该标量。例如, 如果 $k = 2$ 和 $f(x) = x^2$, 那么 $k f(x) = 2f(x) = 2x^2$ 。

3. 实矩阵空间。考虑所有 2×2 的实数矩阵的集合, 记为 $M_{2,2}(\mathbb{R})$ 。这个空间包括所有形如 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 的矩阵, 其中 a, b, c, d 是实数。

- 向量加法: 两个矩阵相加, 就是它们对应元素的和。例如, 如果我们有 $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ 和 $B = \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix}$, 它们的

$$\text{和 } A + B = \begin{pmatrix} 1+4 & 2+3 \\ 3+2 & 4+1 \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ 5 & 5 \end{pmatrix}。$$

- 标量乘法: 一个矩阵乘以一个标量, 就是每个元素乘以这个标量。例如, 如果 $k = 2$ 和 $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, 则

$$kA = 2 \cdot \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}。$$

定义 1.7 (线性变换)

线性变换是从一个线性空间到另一个线性空间的映射 $\mathcal{A}: V \rightarrow W$, 对于所有的 $\alpha, \beta \in V$ 和任意标量 k_1, k_2 , 满足以下两个条件:


1. 加法保持性: $\mathcal{A}(\alpha + \beta) = \mathcal{A}(\alpha) + \mathcal{A}(\beta)$
2. 数量乘法保持性: $\mathcal{A}(k\alpha) = k\mathcal{A}(\alpha)$



上述两条可以写成一条: 设 \mathcal{A} 是平面上一个线性变换, α, β 是平面上的任意两个向量, k_1, k_2 是任意两个实数, 则

$$\mathcal{A}(k_1\alpha + k_2\beta) = k_1\mathcal{A}(\alpha) + k_2\mathcal{A}(\beta) \quad (1.27)$$

1.4 矩阵的运算

 **笔记** 矩阵乘法的规则是“**行乘列**”。这一点在计算结果时很重要, 但无助于理解矩阵乘法的本质, 我们应该从线性变换和线性组合的角度理解矩阵乘法。

1.4.1 向量与向量间的乘法运算


在介绍矩阵的运算之前我们首先回顾向量之间的乘法运算。存在两种类型的向量与向量的乘积: 内积和外积。

定义 1.8 (内积)

对 $x, y \in \mathbb{R}^n$, 我们有:

$$x^T y = \sum_{i=1}^n x_i y_i \in \mathbb{R}$$



 **练习 1.6** 假设我们有两个二维向量 $x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 和 $y = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ 。我们可以使用定义 1.8 中的公式来计算它们的内积:

$$x^T y = xy^T = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = (1 \cdot 3) + (2 \cdot 4) = 3 + 8 = 11$$


因此, 向量 x 和 y 的内积为 11。

定义 1.9 (外积)

对 $x \in \mathbb{R}^m, y \in \mathbb{R}^n$, 我们有:

$$xy^T = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{pmatrix} \in \mathbb{R}^{m \times n}$$



 **练习 1.7** 假设我们有一个列向量 $x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 和一个三维向量 $y = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$ 。我们可以使用定义 1.9 中的公式来计算它们的外积:


$$xy^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 4 & 5 \end{bmatrix}$$

计算这个外积:

$$xy^T = \begin{bmatrix} 1 \cdot 3 & 1 \cdot 4 & 1 \cdot 5 \\ 2 \cdot 3 & 2 \cdot 4 & 2 \cdot 5 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \end{bmatrix}$$

因此, 列向量 x 和三维向量 y 的外积是一个 2×3 的矩阵。

1.4.2 矩阵与向量间的乘法运算

 **笔记** 矩阵乘向量就是对向量做某一个线性变换。

在第 1.3.1 节中, 我们通过将线性变换写成坐标变换方程的方法, 指出平面上线性变换和二阶方阵之间存在着——对应的关系, 即我们实现了用矩阵表示线性变换。

此外, 我们之前提到, 矩阵源于人们在解方程组是对简洁表达的需求。那么, 不借助坐标变换方程, 我们可否直接用矩阵来实现坐标变换呢? 答案是肯定的。


让我们回顾一下式 (1.18), 在直角坐标系 Oxy 内, 将任意一点 $P(x, y)$ 绕原点 O 按逆时针方向旋转 θ 角得到点 $P'(x', y')$ 的旋转变换的坐标变换公式

$$\begin{cases} x' = x \cos \theta - y \sin \theta, \\ y' = x \sin \theta + y \cos \theta. \end{cases}$$

我们进行如下规定, 就可以实现用矩阵来做坐标变换

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (\cos \theta) \cdot x + (-\sin \theta) \cdot y \\ (\sin \theta) \cdot x + (\cos \theta) \cdot y \end{pmatrix} \quad (1.28)$$

我们把第二个等号对应的关系称为**矩阵与向量间的乘法运算**。将一个矩阵与一个向量相乘, 就是对该向量进行该矩阵所表示的线性变换。

 **笔记** 学生可以尝试把之前的几何变换的坐标变换公式写成矩阵与向量间的乘法运算。


定义 1.10 (矩阵与向量间的乘法运算)

矩阵 $A \in \mathbb{R}^{m \times n}$ 和向量 $x \in \mathbb{R}^n$ 的乘积是一个大小为 \mathbb{R}^m 的向量, 满足:

$$Ax = \begin{pmatrix} a_{r,1}^T x \\ \vdots \\ a_{r,m}^T x \end{pmatrix} = \sum_{i=1}^n a_{c,i} x_i \in \mathbb{R}^m$$

其中 $a_{r,i}^T$ 是行向量, $a_{c,j}$ 是 A 的列向量, x_i 是 x 的元素。



 **练习 1.8** 假设我们有一个 2×3 的矩阵 A 与一个 3 维向量 x 相乘。

矩阵 A 和向量 x 的值如下:

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

现在, 让我们根据定义, 计算 Ax 的乘积:


$$Ax = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

我们首先计算每一行与向量 x 的内积, 然后将结果组合成一个新的向量:

$$1. \text{ 第一行与 } x \text{ 的内积: } a_{1,1}^T x = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = 1 \cdot 2 + 2 \cdot 3 + 3 \cdot 1 = 2 + 6 + 3 = 11$$

$$2. \text{ 第二行与 } x \text{ 的内积: } a_{2,1}^T x = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = 4 \cdot 2 + 5 \cdot 3 + 6 \cdot 1 = 8 + 15 + 6 = 29$$

因此, Ax 的结果是一个大小为 \mathbb{R}^2 的向量, 其元素分别是 11 和 29。

 **练习 1.9** 假设我们有一个 2×2 的矩阵 A 与一个 2 维向量 x 相乘。

矩阵 A 和向量 x 的值如下:

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

单位矩阵乘以任何向量等于向量本身, 单位矩阵表示恒等变换。

我们注意到, 式 (1.28) 还有另一种写法:

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} + y \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} \quad (1.29)$$


我们称式 (1.29) 为对向量 $\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$ 和向量 $\begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}$ 的**线性组合**, x, y 称为组合系数。在这个例子中, 初学

者可能会可能会混淆混淆线性变换和线性组合, 线性变换指的是利用 $\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ 对 $\begin{pmatrix} x \\ y \end{pmatrix}$ 进行变换, 而

线性组合指的是 x, y 作为组合系数对向量 $\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$ 和向量 $\begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}$ 进行组合。线性组合是理解矩阵和向量的乘法的关键, 这是 Gilbert Strang 的线性代数教材中介绍的第一个概念。下面给出 Strang 教材中的定义。

定义 1.11 (线性组合)

The heart of linear algebra is in two operations—both with vectors. We add vectors to get $v + w$. We multiply them by numbers c and d to get cv and dw . Combining those two operations (adding cv to dw) gives the **linear combination** $cv + dw$.

 **笔记** 特别地，零向量是任何一组向量的线性组合，请思考原因。

有了我们上面的诸多铺垫，我们早已认识到加法和乘法在线性代数中的核心地位，也比较容易理解为什么“线性组合”这一名称中带有“线性”二字，因为这两个操作是线性变换的基本组成部分。线性变换保留向量空间的线性结构，意味着向量加法和标量乘法的结果仍然在原向量空间内。

在上面的例子告诉我们，矩阵和向量的乘法可以视为用向量的每一个分量对矩阵的各个列对应的向量进行线性组合，我们称矩阵的列对应的向量是**矩阵的列向量**，当然，对偶地，我们也称矩阵的行对应的向量为**矩阵的行向量**。

1.4.3 矩阵与矩阵间的乘法运算

 **笔记** 矩阵和矩阵的乘法是两个线性变换的复合。

我们从线性变换的角度引入矩阵和矩阵之间的乘法。我们考虑平面上的点 $P(1, 0)$ ，我们首先将其逆时针旋转 90° ，对应的变换矩阵为 $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ ，得到点 $P'(0, 1)$ ，然后按照 x 轴对称，变换矩阵为 $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ ，得到点 $P''(0, -1)$ 。

注意到，这两个变换的复合结果等效于将点 $P(1, 0)$ 沿着直线 $y = -x$ 进行反射，该变换的变换矩阵是 $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ 。这样的现象在线性空间中普遍存在。因为我们把线性变换的复合在矩阵上的表现抽象出来定义为一种运算，这种运算就是矩阵和矩阵之间的乘法。


$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

定义 1.12 (矩阵和矩阵的乘法)

矩阵 $A \in \mathbb{R}^{m \times n}$ 和 $B \in \mathbb{R}^{n \times p}$ 的乘积是一个大小为 $\mathbb{R}^{m \times p}$ 的矩阵，满足：

$$AB = \begin{pmatrix} a_{r,1}^T b_{c,1} & \cdots & a_{r,1}^T b_{c,p} \\ \vdots & & \vdots \\ a_{r,m}^T b_{c,1} & \cdots & a_{r,m}^T b_{c,p} \end{pmatrix} = \sum_{i=1}^n a_{c,i} b_{r,i}^T \in \mathbb{R}^{n \times p}$$

其中 $a_{r,i}^T, b_{r,i}^T$ 是行向量， $a_{c,j}, b_{c,j}$ 分别是 A 和 B 的列向量。

 **练习 1.10** 我们将使用一个 2×3 的矩阵 A 与一个 3×2 的矩阵 B 相乘。

首先，让我们定义矩阵 A 和矩阵 B ：

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}$$

现在，让我们使用这个新的矩阵 B 与之前定义的矩阵 A 相乘，以演示矩阵乘法的过程：

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}$$

计算每一行与每一列的内积，并生成新的矩阵 AB ：

1. 第一行与第一列的内积： $a_{1,1}^T b_{1,1} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = 1 \cdot 2 + 2 \cdot 4 + 3 \cdot 6 = 2 + 8 + 18 = 28$
2. 第一行与第二列的内积： $a_{1,1}^T b_{2,1} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} = 1 \cdot 3 + 2 \cdot 5 + 3 \cdot 7 = 3 + 10 + 21 = 34$
3. 第二行与第一列的内积： $a_{2,1}^T b_{1,1} = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = 4 \cdot 2 + 5 \cdot 4 + 6 \cdot 6 = 8 + 20 + 36 = 64$
4. 第二行与第二列的内积： $a_{2,1}^T b_{2,1} = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} = 4 \cdot 3 + 5 \cdot 5 + 6 \cdot 7 = 12 + 25 + 42 = 79$

因此， AB 的结果是一个大小为 $\mathbb{R}^{2 \times 2}$ 的矩阵，其元素如下：

$$AB = \begin{bmatrix} 28 & 34 \\ 64 & 79 \end{bmatrix}$$

1.4.4 初等变换与矩阵的乘法

在第1.1.3节中，我们给出了行变换的概念。事实上，在解方程组的过程中，只有三种基本的行变换：

- 行交换 (Row-switching transformations): 交换矩阵中的两行；
- 行乘以非零系数 (Row-multiplying transformations): 将某一行所有的元素乘以一个非零常数；
- 行加减 (Row-addition transformations): 将某一行的若干倍加到另一行上。

我们在单位矩阵上进行三种行变换，看看会发生什么。二阶单位矩阵是一个简单的矩阵，其形式为：

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

这个单位矩阵将作为我们变换的基础。

行交换。如果我们交换这个单位矩阵的两行，我们得到一个矩阵：

$$T_{1,2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$T_{i,j}$ 表示交换矩阵的第 i 行和第 j 行。

行倍乘。如果我们将第一行乘以一个非零常数，比如 2，我们得到：

$$D_1(2) = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$


$D_i(m)$ 代表对矩阵的第 i 行的所有元素乘以标量 m 。按照我们之前的讨论知，它起到伸缩变换的效果。其形式可概括为：除了第 i 个位置（此处为 m ）外，所有对角元素均为 1 的对角矩阵。

行加减。如果我们将第二行的 2 倍加到第一行上，我们得到：

$$L_{12}(2) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

矩阵 $L_{ij}(m)$ 将行 j 乘以标量 m 加到行 i ，其形式可概括为单位矩阵在 (i, j) 位置有一个 m 。

$T_{i,j}, D_i(m)$ 和 $L_{ij}(m)$ 三种矩阵是初等行变换对应的矩阵，我们将其称为**初等矩阵**。现在，让我们通过一个例子来展示如何使用这些初等矩阵。

 **练习 1.11** 考虑一个二阶矩阵 A ：

$$A = \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix}$$

当我们使用 $T_{1,2}$ 左乘矩阵 A （又称**左乘** A ）时，我们有

$$T_{1,2} \cdot A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 5 & 6 \\ 3 & 4 \end{pmatrix}$$

可见，相乘的结果是矩阵 A 的第一行和第二行被交换了。

当我们使用 $D_1(2)$ 从左乘矩阵 A 时，我们有

$$D_1(2) \cdot A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 2 \times 3 & 2 \times 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 5 & 6 \end{pmatrix}$$

这个结果展示了行倍乘的效果，即矩阵 A 的第一行的每一个元素都乘了 2。

当我们使用 $L_{ij}(m)$ 左乘矩阵 A 时，我们有


$$L_{ij}(m) \cdot A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 3 + 2 \times 5 & 4 + 2 \times 6 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 13 & 16 \\ 5 & 6 \end{pmatrix}$$

这个操作将第二行的两倍加到第一行上，改变了矩阵 A 的第一行，而第二行保持不变，展现了切变变换的效果。

至此，我们通过对单位矩阵应用行交换、行倍乘和行相加构造了初等矩阵，并通过具体的例子理解了这些初等矩阵如何通过矩阵乘法初等行变换作用于其他矩阵。

 **笔记** 用初等矩阵左乘矩阵 A 等同于对 A 做对应的初等行变换。

一个很自然的问题是，为什么一定要左乘？为什么不右乘？我们可以试试看。

 **练习 1.12** 考虑一个二阶矩阵 A ：

$$A = \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix}$$

当我们使用 $T_{1,2}$ 右乘矩阵 A 时，我们有

$$A \cdot T_{1,2} = \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 6 & 5 \end{pmatrix}$$

可见，相乘的结果是矩阵 A 的第一列和第二列被交换了。

当我们使用 $D_1(2)$ 右乘矩阵 A 时，我们有

$$A \cdot D_1(2) = \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 \times 3 & 4 \\ 2 \times 5 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 4 \\ 10 & 6 \end{pmatrix}$$

这个结果展示了列倍乘的效果，即矩阵 A 的第一列 的每一个元素都乘了 2。

当我们使用 $L_{ij}(m)$ 右乘矩阵 A 时，我们有

$$L_{ij}(m) \cdot A = \begin{pmatrix} 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 4+3 \times 2 \\ 5 & 6+5 \times 2 \end{pmatrix} = \begin{pmatrix} 3 & 10 \\ 5 & 16 \end{pmatrix}$$

这个操作将第一列 的两倍加到第二列 上，改变了矩阵 A 的第二列。

我们看到，初等变换右乘矩阵 A 的效果是将初等行变换那样的操作作用到列上，这些变换称为初等列变换。

 **笔记** 用初等矩阵右乘矩阵 A 等同于对 A 做对应的初等列变换。

1.5 线性方程解的结构

至此，我们已经探讨了向量之间的线性组合（这有助于我们理解矩阵和向量的乘法）、线性变换，以及线性空间的概念。在这一节中，我们将讨论线性方程组与线性空间的关系。我们将看到，线性方程组的每个解集合形成了原空间中的一个子空间，称为解空间，这个解空间包含了所有满足方程组条件的向解量，即给定线性方程组的所有可能解。

在介绍线性方程的解的结构之前，我们先来认识向量空间中的向量可能存在的关系，这种关系有助于我们理解线性方程解的性质。

1.5.1 线性方程组

线性方程（Linear equation）是指最高次数为一次的多项式方程。在线性代数中，一个线性方程通常表示为一系列变量的和，每个变量都乘以一个常数（该变量的系数），等于一个常数。在一元线性方程的情况下，其一般形式可以表示为：

$$ax + b = 0$$

这里的 x 是变量，而 a 和 b 是系数，其中 $a \neq 0$ 。如果 $a = 0$ ，那么方程退化成一个常数项等于 0 的非线性方程。

在多元线性方程的情况下，其一般形式可以表示为：

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + b = 0$$

这里的 x_1, x_2, \dots, x_n 是变量，而 a_1, a_2, \dots, a_n 和 b 是系数。

线性方程的主要特点是：

- 每项是常数或者常数乘以一个变量；
- 没有变量的乘方、根号或其他非线性操作；
- 变量之间没有乘积（例如没有项是 x_1x_2 或 x_1^2 ）

在几何上，二维空间中的一个线性方程代表一条直线，三维空间中的一个线性方程代表一个平面。更高维度的线性方程代表超平面。

线性方程的一个重要特性是解的叠加原理：如果 x_1 和 x_2 分别是两个线性方程的解，则它们的任意线性组合 $c_1x_1 + c_2x_2$ 也是该线性方程的解，其中 c_1 和 c_2 是任意常数。

线性方程组是由多个线性方程构成的集合，通常我们寻求这些方程共同满足的解。

1.5.2 线性相关与线性无关

 **笔记** 复习：

- 线性组合
- n 维单位向量可以线性表出任意 n 维向量, 就像直角坐标系里的坐标一样, 就用二维和三位空间举例,

定义 1.13 (线性表出)

当向量 α 是向量组 $\beta_1, \beta_2, \dots, \beta_s$ 的一个线性组合时, 我们也说 α 可以经向量组 $\beta_1, \beta_2, \dots, \beta_s$ **线性表出**.

定义 1.14 (向量组等价)

如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_t$ 中每一个向量 $\alpha_i (i = 1, 2, \dots, t)$ 都可以经向量组 $\beta_1, \beta_2, \dots, \beta_s$ 线性表出, 那么向量组 $\alpha_1, \alpha_2, \dots, \alpha_t$ 就称为可以经向量组 $\beta_1, \beta_2, \dots, \beta_s$ 线性表出. 如果两个向量组互相可以线性表出它们就称为**等价**.

练习 1.13 设

$$\alpha_1 = (1, 1, 1), \alpha_2 = (1, 2, 0);$$

$$\beta_1 = (1, 0, 2), \beta_2 = (0, 1, -1).$$

则向量组 α_1, α_2 与向量组 β_1, β_2 是等价的.

定义 1.15 (向量组的线性相关)

如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_s (s \geq 2)$ 中有一个向量可以由其余的向量线性表出, 那么向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 称为**线性相关**的.

练习 1.14 向量组 $\alpha_1 = (2, -1, 3, 1), \alpha_2 = (4, -2, 5, 4), \alpha_3 = (2, -1, 4, -1)$ 是线性相关的, 因为

$$3\alpha_1 - \alpha_2 = \alpha_3.$$

从定义可以看出任意一个包含零向量的向量组一定是线性相关的. 还可看出, 向量组 α_1, α_2 线性相关就表示 $\alpha_1 = k\alpha_2$ 或者 $\alpha_2 = k\alpha_1$ (这两个式子不一定能同时成立, 例如当其中有一个是零向量的时候).

定义 1.16 (线性相关难懂版)

向量组 $\alpha_1, \alpha_2, \dots, \alpha_s (s \geq 1)$ 称为**线性相关**, 如果有数域 \mathcal{F} 中不全为零的数 k_1, k_2, \dots, k_s , 使

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_s\alpha_s = \mathbf{0}$$

为什么要写这个难懂版呢? 是因为要对称地给出线性无关的定义.

定义 1.17 (向量组线性无关)

一向量组 $\alpha_1, \alpha_2, \dots, \alpha_s (s \geq 1)$ 不线性相关, 即没有不全为零的数 k_1, k_2, \dots, k_s 使

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_s\alpha_s = \mathbf{0},$$

就称为**线性无关**; 或者说, 一向量组 $\alpha_1, \alpha_2, \dots, \alpha_s$ 称为**线性无关**, 如果由

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_s\alpha_s = \mathbf{0}$$

可以推出

$$k_1 = k_2 = \dots = k_s = 0$$

一些结论, 如果这些结论一看就懂, 那说明你初步理解了上面的概念:


- 如果一向量组的一部分线性相关, 那么这个向量组就线性相关.
- 设 $\alpha_1, \alpha_2, \dots, \alpha_r$ 与 $\beta_1, \beta_2, \dots, \beta_s$ 是两个向量组. 有下面一组逆否命题:
 - 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 可以经 $\beta_1, \beta_2, \dots, \beta_s$ 线性表出且 $r > s$, 那么向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 必线性相关.
 - 如果向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 可以经向量组 $\beta_1, \beta_2, \dots, \beta_s$ 线性表出, 且 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性无关, 那

么 $r \leq s$.

- 任意 $n+1$ 个 n 维向量必线性相关.
- 两个线性无关的等价的向量组, 必含有相同个数的向量.

定义 1.18 (最大线性无关组)

一向量组的一个部分组称为一个最大线性无关组, 如果这个部分组本身是线性无关的, 并且从这向量组中任意添一个向量 (如果还有的话), 所得的部分向量组都线性相关.

 **练习 1.15** 在向量组 $\alpha_1 = (2, -1, 3, 1)$, $\alpha_2 = (4, -2, 5, 4)$, $\alpha_3 = (2, -1, 4, -1)$ 中, 由 α_1, α_2 组成的部分组就是一个最大线性无关组. 首先, α_1, α_2 线性无关, 因为由


$$\begin{aligned} k_1\alpha_1 + k_2\alpha_2 &= k_1(2, -1, 3, 1) + k_2(4, -2, 5, 4) \\ &= (2k_1 + 4k_2, -k_1 - 2k_2, 3k_1 + 5k_2, k_1 + 4k_2) \\ &= (0, 0, 0, 0), \end{aligned}$$

就有 $k_1 = k_2 = 0$. 同时我们知道, $\alpha_1, \alpha_2, \alpha_3$ 线性相关. 不难看出, α_2, α_3 也是一个最大线性无关组. 一些结论:

- 一个线性无关向量组的最大线性无关组就是这个向量组自身.
- 最大线性无关组的一个基本性质是, 任意一个最大线性无关组都与向量组本身等价.

定义 1.19

向量组的极大线性无关组所含向量的个数称为这个向量组的秩.

 **练习 1.16** 向量组 $\alpha_1 = (2, -1, 3, 1)$, $\alpha_2 = (4, -2, 5, 4)$, $\alpha_3 = (2, -1, 4, -1)$ 的秩就是 2.

因为线性无关的向量组就是它自身的极大线性无关组, 所以一向量组线性无关的充分必要条件为它的秩与它所含向量的个数相同.

1.5.3 矩阵的秩

考虑矩阵 A , 将其化为简化行阶梯形

$$A = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 14 \end{bmatrix} \rightarrow \text{rref}(A) = \begin{bmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

我们发现第一列和第二列无法被其他列线性表出. 这个现象不是巧合, 因为我们在利用行变换将矩阵 A 变为简化行阶梯形的过程中选取了第一列的第一个元素和第二列的第二个元素作为基准元素 (它们被称为**主元**, 见第 1.1.3 节), 并且通过行变换将它们所在的列中的其他元素变为 0. 在这个过程中, 主元的位置是固定的, 并且在每一行的最左边. 主元的存在保证了矩阵的行简化形式是唯一的. 我们将主元所在的列以外的列成为**主元列**(pivot column), 主元列以外的列称为**自由列**(free column). 自由列可以通过主元的线性组合来表示, 因此它们是自由的. 自由列的数量决定了线性方程组的自由度 (注意方程的自由度不是解的个数).

我们可以再看一个例子. 考虑矩阵 B ,

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \rightarrow \text{rref}(B) = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

这个矩阵只有一个主元列, 另外两列都是自由列. 自由列也都可以被主元列线性表出.

综合以上两个例子, 我们看到, 主元在矩阵的行变换过程中非常重要, 主元的个数就是矩阵 A 对应的线性方程组中有效方程的个数, 这个数字可以认为是矩阵 A “真正的大小” (true size), 这个“真正的大小”很重要, 我们将其称为**矩阵的秩**.

定义 1.20 (矩阵的秩)

矩阵 A 的主元个数成为矩阵 A 的**秩** (rank), 一般简记为 r 。

1.5.4 线性子空间

 **笔记** 本节核心概念:

- 子空间
- “张成一个空间”的含义
- 子空间的基
- 空间的维度
- 矩阵的列空间和行空间

定义 1.21 (子空间—Gilbert Strang 教材)

向量空间中一组包含 $\mathbf{0}$ 的向量对加法和数乘满足封闭性, 则这组向量构成原来向量空间的一个**子空间**。

举个例子, 平面上的直线都是平面的子空间, 三维空间中的平面和直线都是三维空间的子空间。此外, 平面和三维空间分别是它们自己的子空间。

命题 1.2

包含向量 \mathbf{u}, \mathbf{v} 的子空间一定包含二者的所有线性组合。

定义 1.22 (Span—Gilbert Strang)

一组向量**张成**一个空间, 如果它们的线性组合布满的整个空间。

A set of vectors **span** a space if their linear combinations fill the space.

定义 1.23 (基和基向量)

向量空间的**基** (basis) 是能够张成这个空间的一组线性无关的向量。我们称构成基的向量是该空间的**基向量** (base)。

 **笔记** 结合最大线性无关组的概念, 我们知道, 线性空间的基就是该空间内所有向量的最大线性无关组。

在 n 维线性空间 V 中, n 个线性无关的向量 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 称为 V 的一组基。设 α 是 V 中任一向量, 于是 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \alpha$ 线性相关, 因此 α 可以被基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 线性表出:

$$\alpha = a_1\varepsilon_1 + a_2\varepsilon_2 + \dots + a_n\varepsilon_n,$$

其中系数 a_1, a_2, \dots, a_n 是被向量 α 和基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 唯一确定的, 这组数就称为 α 在基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 下的坐标, 记为 (a_1, a_2, \dots, a_n) 。

定义 1.24 (向量空间的维度)

向量空间的**维度** (维数) 是它的每一组基中基向量的个数。

这个定义是良定义的 (well-defined), 当且仅当向量空间的每一组基中基向量的数量都相等。

现在, 我们使用上面的概念来重新审视矩阵。我们之前有如下的理解矩阵的方法:

- 矩阵 A 表示线性方程组 $Ax = c$ 的系数;
- 矩阵 A 表示空间中某一个线性变换;

我们发现, 矩阵 A 的列向量也可以张成一个维度至少为 1 的空间, 我们称这个空间为矩阵 A 的**列空间**, 记为 $C(A)$ 。对偶地, 矩阵 A 的行向量也可以张成一个维度至少为 1 的空, 我们称这个空间为矩阵 A 的**行空间**。

从现在开始我们几乎不再像中学时那样写出完整的线性方程组，而更多地使用向量形式，即 $Ax = c$ ，这看起来像是一个公式，称之为方程组虽然是对的但有些怪，故从现在起我们使用它的一个贴近本质的名称——**线性系统** (linear system)，从而线性方程组的解就是线性系统的解。借助列空间的概念，我们可以获得对线性系统的一种解释。

定理 1.1

线性系统 $Ax = c$ 有解当且仅当 c 在 A 的列空间中。



考察第1.5.3节中的两个例子。我们发现，例子1.5.3中列空间的维度是2，行空间的维度也是2；而例子1.5.3中列空间的维度是1，行空间的维度也是1。

定理 1.2

矩阵的行空间的维度等于矩阵列空间的维度等于矩阵的秩。



你可能会到别处看到矩阵的行秩和列秩的概念，这二者分别对应于矩阵行空间的维度和列空间的维度，二者永远相等。

定义 1.25

一个 $m \times n$ 的矩阵 A ，如果满足以下任一条件，则称 A 是**满秩矩阵**：

- 矩阵 A 的秩等于 $\min(m, n)$ 。也就是说，矩阵的秩达到了最大可能值；
- 矩阵 A 的所有行向量线性无关；
- 矩阵 A 的所有列向量线性无关；
- 如果 $m \geq n$ ，则 A 有 n 个线性无关的列向量；如果 $m \leq n$ ，则 A 有 m 个线性无关的行向量

**1.5.5 线性代数基本定理** **笔记**

- 线性系统 $Ax = 0$ 的解答
- 矩阵的零空间
- 线性系统 $Ax = c$ 的解答等于一个 $Ax = c$ 的特解加上 $Ax = 0$ 的通解
- 秩-零化度定理

1.6 基础矩阵操作**1.6.1 矩阵的转置****定义 1.26**

矩阵 $A \in \mathbb{R}^{m \times n}$ 的转置，记作 A^T ，是其中元素的翻转

$$\forall i, j, \quad A_{i,j}^T = A_{j,i}$$



练习 1.17 假设我们有一个 2×3 的矩阵：

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

现在，让我们计算矩阵 A 的转置 A^T 。转置操作将原矩阵的行变成列，列变成行，如下所示：

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

这里，我们可以清晰地看到原始矩阵 A 的行变成了 A^T 的列，而原始矩阵的列变成了 A^T 的行。

定理 1.3 (转置运算的性质)

对矩阵 A, B ，我们有 $(AB)^T = B^T A^T$ 。（提示：想一下矩阵相乘的尺寸对齐关系）



1.6.2 矩阵的逆

定义 1.27

设 A 为一个 $n \times n$ 的方阵。如果存在另一个同样大小的矩阵 B 使得：

$$AB = BA = I_n$$

其中 I_n 是 $n \times n$ 的单位矩阵。则称矩阵 B 为 A 的逆矩阵，记作 A^{-1} 。



换句话说，矩阵 A 的逆矩阵是一个矩阵 A^{-1} ，使得与 A 相乘，无论是左乘还是右乘，都能得到单位矩阵。逆矩阵“抵消”了原矩阵的效果，很像除法抵消了实数乘法的效果。即根据定义 1.27，我们有

$$AA^{-1} = A^{-1}A = I$$

矩阵的逆可以被看作是对应线性变换的逆变换。如果有一个线性变换可以通过矩阵 A 来表示，那么该线性变换的逆变换（如果存在）可以通过矩阵 A 的逆矩阵 A^{-1} 来表示。这意味着，如果 A 是将向量 x 映射到向量 y 的矩阵，则 A^{-1} 将 y 映射回 x 。

为了更具体地说明这一点，让我们考虑两个具有直观几何意义的线性变换：反射和旋转。

反射变换。假设我们有一个关于 x 轴的反射变换。这个变换可以通过以下矩阵表示：

$$A_{\text{反射}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

这个矩阵将点 (x, y) 反射到 $(x, -y)$ 。因为关于 x 轴的反射是自己的逆变换（即，再次反射将会恢复原状）， $A_{\text{反射}}$ 的逆矩阵仍然是它自己：

$$A_{\text{反射}}^{-1} = A_{\text{反射}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

旋转变换。考虑一个围绕原点逆时针旋转 90° 的变换。这个变换可以通过以下矩阵表示：

$$A_{\text{旋转}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

这个矩阵将点 (x, y) 旋转到 $(-y, x)$ 。其逆变换，也就是顺时针旋转 90° ，可以通过计算 $A_{\text{旋转}}$ 的逆得到。根据矩阵逆的定义，我们找到 $A_{\text{旋转}}^{-1}$ ：

$$A_{\text{旋转}}^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

这个矩阵将点 (x, y) 旋转到 $(y, -x)$ ，也就是顺时针旋转 90° 。

这两个例子清晰地展示了线性变换的逆变换在几何上是如何工作的。反射变换的逆变换就是它本身，因为两次反射将会抵消彼此的效果。旋转变换的逆变换是反方向的旋转，以此来抵消原始旋转的效果。

然而，并不是所有的方阵都有逆矩阵。一个矩阵 A 是可逆的，当且仅当其行列式非零，即 $\det(A) \neq 0$ 。根据第 1.1.4 节中介绍的行列式的几何意义可知，如果行列式为零，则变换将正方形或立方体分别折叠成一条线或一个平面，即行列式为零的矩阵对应的线性变换会导致几何图形维度的坍塌。可以说，行列式为零的矩阵就是线性代数世界中的二向箔，被降维的过程是不可逆的，因为损失了还原原始几何题的必要信息，

不可逆矩阵也叫奇异矩阵或退化矩阵，它们的行列式的值为 0。下面是一些不可逆矩阵的例子：

1. 全零矩阵：

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

2. 含有线性相关行或列的矩阵：

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

3. 非方阵也是不可逆的，例如：

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

4. 任何行或列全为 0 的矩阵，例如：

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

这些例子中的矩阵都不能找到一个逆矩阵来满足矩阵乘法的单位矩阵条件。

定理 1.4 (矩阵的逆的运算性质)

对可逆矩阵 A, B ，我们有 $(AB)^{-1} = B^{-1}A^{-1}$ 。（提示：套定义和原矩阵相乘就可以证明）



那么，如何计算矩阵 A 的逆呢？一个最容易理解的方法是借助初等变换的求逆法。初等变换包括行交换、行乘以非零常数、行加上另一行的非零常数倍。正如第 1.4.4 节所述，这些操作对应于左乘不同的初等矩阵。矩阵 A 的逆可以通过对矩阵 A 和单位矩阵 I 同时进行初等行变换来求得。具体步骤如下：

1. **设置矩阵：**从矩阵 A 开始，旁边写上同阶的单位矩阵 I 。这个组合记为 $[A|I]$ 。
2. **进行初等行变换：**目标是通过初等行变换将 A 转换为单位矩阵 I 。对于每一个在 A 上执行的操作，同样的操作也要在 I 上执行。
3. **获得逆矩阵：**如果 A 可以通过行变换变为单位矩阵 I ，那么在 I 位置上变换得到的矩阵就是 A 的逆矩阵 A^{-1} 。

该方法的动机和合理性来源于线性代数中的一个基本定理：**任何可逆矩阵都可以通过一系列初等行变换变为单位矩阵。**²这种方法背后的直觉是，当我们对矩阵 A 和单位矩阵 I 同时执行相同的初等行变换时，实际上我们是在记录下从 A 变换到 I 所需的每一步操作。这些操作，反过来应用于单位矩阵 I ，可以生成 A 的逆矩阵 A^{-1} ，因为它们描述了如何通过一系列操作将 A 变换为 I 。

练习 1.18 假设我们有一个矩阵 A ：

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

我们要找到它的逆。首先，我们将 A 与单位矩阵 I 并排写下：

$$[A|I] = \left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 3 & 2 & 0 & 1 \end{array} \right]$$

²反证法：给定一个可逆矩阵 A ，我们首先执行一系列初等行变换将其转换为行最简形式。由于 A 是可逆的，其行最简形式必定是单位矩阵 I 。这是因为如果 A 的行最简形式不是 I ，则会存在零行，意味着 A 不是满秩的，与 A 可逆（意味着满秩）相矛盾。

步骤 1: 将第一行乘以 $\frac{1}{2}$, 使 A 的第一行第一个元素变为 1:

$$\left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 3 & 2 & 0 & 1 \end{array} \right]$$

步骤 2: 将第一行的 -3 倍加到第二行上, 使 A 的第二行第一个元素变为 0:

$$\left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{3}{2} & 1 \end{array} \right]$$

步骤 3: 将第二行乘以 2, 使 A 的第二行第二个元素变为 1:

$$\left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & -3 & 2 \end{array} \right]$$

步骤 4: 将第二行的 $-\frac{1}{2}$ 倍加到第一行上, 使 A 的第一行第二个元素变为 0:

$$\left[\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -3 & 2 \end{array} \right]$$

在完成这些步骤之后, A 变为了单位矩阵, 而我们在 I 的位置得到的矩阵就是 A 的逆:

$$A^{-1} = \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix}$$

使用 Python 求矩阵的逆:

```
import numpy as np

# 定义一个矩阵
A = np.array([[1, 2], [3, 4]])

# 使用numpy的linalg.inv()函数来找到它的逆
A_inv = np.linalg.inv(A)

print(f'逆矩阵 A^{-1}: \n{A_inv}')
# [[-2.  1.]
#  [ 1.5 -0.5]]
```

第二章 函数与优化


在数学中，**函数**是一个基本但强大的工具，它简单地将每一个输入值映射到一个输出值。你可以将它想象为一个机器：你放进一个数（输入），它便会给你另一个数（输出）。

接下来我们将探讨**优化**的概念，它涉及到一个目标函数的最大化或最小化过程，帮助我们在可能的解决方案中找到“最优”的一种。在现实生活中，这可以是找到最省钱的购物清单，或是找到最快的到达目的地的路线。

在本章中，我们将初步探讨函数和优化的基本理论，为你打开一扇理解和掌握这两个强大工具的大门，从而在实际应用中找到最佳的解决方案。

2.1 函数的概念和性质

2.1.1 函数的定义

 **笔记** 函数是一种数学关系，它将一个集合的元素映射到另一个集合的元素。

定义 2.1 (函数)

给定两个**实数集** D 和 M ，若有**对应法则** f ，使对 D 内每一个数 x ，都有**唯一**的一个数 $y \in M$ 与它相对应，则称 f 是定义在数集 D 上的**函数**，记作

$$\begin{aligned} f: D &\rightarrow M, \\ x &\mapsto y. \end{aligned} \quad (2.1)$$

数集 D 称为函数 f 的**定义域**， x 所对应的数 y 称为 f 在点 x 的**函数值**，常记为 $f(x)$ 。全体函数值的集合

$$f(D) = \{y \mid y = f(x), x \in D\} (\subset M)$$

称为函数 f 的**值域**。

关于函数的定义，我们做如下几点说明：


1. 什么是函数相同？相同的函数形式可能不同；
2. 使运算式子有意义的自变量值的全体，称为存在域。函数的定义域常取存在域；
3. 函数是一种映射关系，原象和象是映射关系的要素；
4. 在函数定义中，对每一个 $x \in D$ ，只能有唯一的一个 y 值与它对应，这样定义的函数称为**单值函数**。若同一个 x 值可以对应多于一个的 y 值，则称这种函数为**多值函数**。在本讲义范围内，我们只讨论单值函数。

2.1.2 具有某些特性的函数

2.1.2.1 有界函数

定义 2.2 (有界集)

设 S 为 \mathbf{R} 中的一个数集。若存在数 $M(L)$ ，使得对一切 $x \in S$ ，都有 $x \leq M(x \geq L)$ ，则称 S 为**有上界 (下界) 的数集**，数 $M(L)$ 称为 S 的一个**上界 (下界)**。若数集 S 既有上界又有下界，则称 S 为**有界集**。若 S 不是**有界集**，则称 S 为**无界集**。

 **练习 2.1** 证明数集 $\mathbf{N}_+ = \{n \mid n \text{ 为正整数} \}$ 有下界而无上界。

定义 2.3 (有上界函数)

设 f 为定义在 D 上的函数. 若存在数 M , 使得对每一个 $x \in D$ 有

$$f(x) \leq M,$$

则称 f 为 D 上的**有上界函数**, M 称为 f 在 D 上的一个**上界**.

f 在 D 上有上界, 意味着值域 $f(D)$ 是一个有上界的数集. 又若 M 为 f 在 D 上的上界, 则任何大于 M 的数也是 f 在 D 上的上界.

定义 2.4 (有下界函数)

设 f 为定义在 D 上的函数. 若存在数 L , 使得对每一个 $x \in D$ 有

$$f(x) \geq L,$$

则称 f 为 D 上的**有下界函数**, L 称为 f 在 D 上的一个**下界**.

f 在 D 上有下界, 意味着值域 $f(D)$ 是一个有下界的数集. 又若 L 为 f 在 D 上的下界, 则任何小于 L 的数也是 f 在 D 上的下界.


定义 2.5 (有界函数)

设 f 为定义在 D 上的函数. 若存在正数 M , 使得对每一个 $x \in D$ 有


$$|f(x)| \leq M, \quad (2.2)$$

则称 f 为 D 上的**有界函数**.

f 在 D 上有界的充要条件是 f 在 D 上既有上界又有下界. (2.2) 式的几何意义是: 若 f 为 D 上的有界函数, 则 f 的图像完全落在直线 $y = M$ 与 $y = -M$ 之间.

 **练习 2.2** 证明正弦函数 $\sin x$ 和余弦函数 $\cos x$ 为 \mathbf{R} 上的有界函数.

关于函数 f 在数集 D 上无上界、无下界或无界的定义, 可按上述相应定义的否定说法来叙述. 例如, 设 f 为定义在 D 上的函数, 若对任何 M (无论 M 多大), 都存在 $x_0 \in D$, 使得 $f(x_0) > M$, 则称 f 为 D 上的无上界函数.

 **练习 2.3** 证明 $f(x) = \frac{1}{x}$ 为 $(0, 1]$ 上的无上界函数.

2.1.2.2 单调函数**定义 2.6**

设 f 为定义在 D 上的函数. 若对任何 $x_1, x_2 \in D$, 当 $x_1 < x_2$ 时, 总有

1. $f(x_1) \leq f(x_2)$, 则称 f 为 D 上的**增函数**, 特别当成立严格不等式 $f(x_1) < f(x_2)$ 时, 称 f 为 D 上的**严格增函数**;
2. $f(x_1) \geq f(x_2)$, 则称 f 为 D 上的**减函数**, 特别当成立严格不等式 $f(x_1) > f(x_2)$ 时, 称 f 为 D 上的**严格减函数**.

增函数和减函数统称为**单调函数**, 严格增函数和严格减函数统称为**严格单调函数**.

2.1.2.3 奇函数和偶函数

奇函数和偶函数是两种特殊类型的函数, 它们具有特定的对称性质。

定义 2.7 (奇函数)

如果一个定义在关于原点对称的数集 D 上的函数 $f(x)$ 满足 $f(x) = -f(-x)$, 那么它被称为**奇函数**。

显然, 奇函数关于原点对称。

定义 2.8 (偶函数)

如果一个定义在关于原点对称的数集 D 上函数 $f(x)$ 满足 $f(x) = f(-x)$, 那么它被称为**偶函数**。

显然, 偶函数关于 y 轴对称。

2.1.2.4 周期函数**定义 2.9 (周期函数)**

设 f 为定义在数集 D 上的函数. 若存在 $\sigma > 0$, 使得对一切 $x \in D, x \pm \sigma \in D$, 有 $f(x \pm \sigma) = f(x)$, 则称 f 为**周期函数**, σ 称为 f 的一个**周期**.

显然, 若 σ 为 f 的周期, 则 $n\sigma$ (n 为正整数) 也是 f 的周期. 若在周期, 则称此最小周期为 f 的**基本周期**, 或简称**周期**.

练习 2.4

1. $\sin x$ 的周期为 2π , $\tan x$ 的周期为 π .
2. 如图2.1所示, 函数

$$f(x) = x - [x], x \in \mathbf{R}$$

的周期为 1.

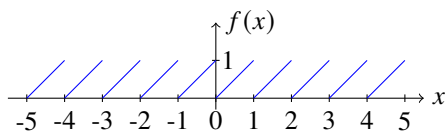


图 2.1: $f(x) = x - [x]$ 的图像

3. 常量函数 $f(x) = c$ 是以任何正数为周期的周期函数, 但不存在基本周期.

2.1.3 初等函数

我们在中学接触过的基本初等函数有如表2.1所示的六类。它们的函数图像如图2.2所示。

幂函数 $y = x^a$ 和指数函数 $y = a^x$ 都涉及乘幂, 而在中学数课程中只给出了有理指数乘幂的定义. 这里我们指出, 无理数也可以作为幂指数。无理指数幂和有理指数幂一起构成了实数指数幂, 并保持有理指数幂的基本性质。

定义 2.10 (初等函数)

由基本初等函数经过有限次四则运算与复合运算所得到的函数, 统称为**初等函数**。

2.1.4 复合函数**2.1.5 反函数**

表 2.1: 基本初等函数分类表

函数类型	函数表达式
常量函数	$y = c$ (c 是常数)
幂函数	$y = x^u$ (u 为实数)
指数函数	$y = a^x$ ($a > 0, a \neq 1$)
对数函数	$y = \log_a x$ ($a > 0, a \neq 1$)
三角函数	$y = \sin x$ (正弦函数)
	$y = \cos x$ (余弦函数)
	$y = \tan x$ (正切函数)
	$y = \cot x$ (余切函数)
反三角函数	$y = \arcsin x$ (反正弦函数)
	$y = \arccos x$ (反余弦函数)
	$y = \arctan x$ (反正切函数)
	$y = \operatorname{arccot} x$ (反余切函数)

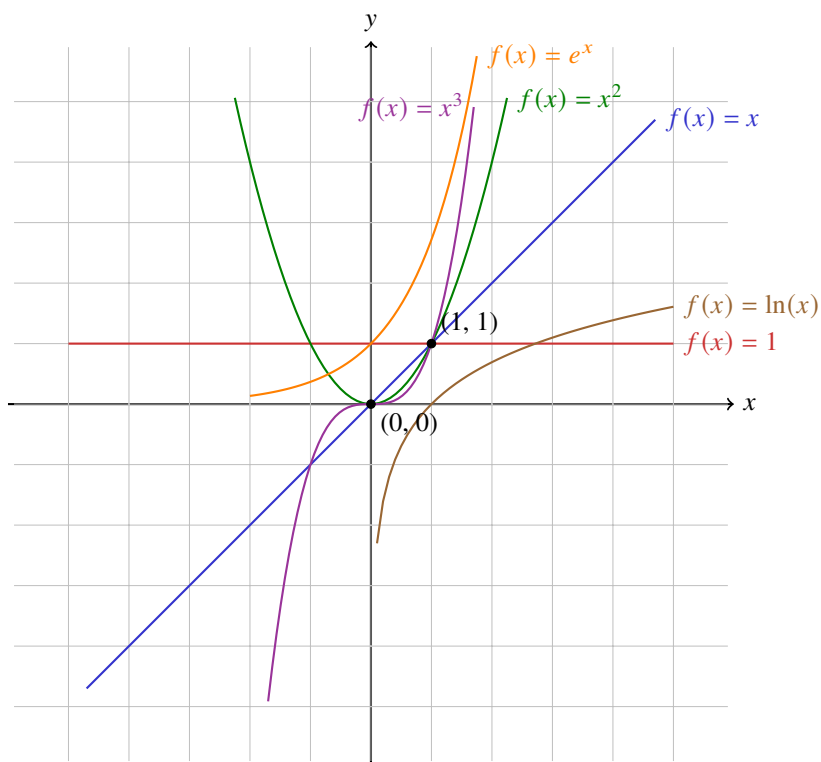


图 2.2: 基本初等函数的图像

附录 A 基本数学工具

本附录包括了计量经济学中用到的一些基本数学，我们扼要论述了求和算子的各种性质，研究了线性和某些非线性方程的性质，并复习了比例和百分数。我们还介绍了一些在应用计量经济学中常见的特殊函数，包括二次函数和自然对数，前 4 节只要求基本的代数技巧，第 5 节则对微分学进行了简要回顾；虽然要理解本书的大部分内容，微积分并非必需，但在一些章末附录和第 3 篇某些高深专题中，我们还是用到了微积分。

A.1 求和算子与描述统计量

求和算子是用以表达多个数求和运算的一个缩略符号，它在统计学和计量经济学分析中扮演着重要作用。如果 $\{x_i : i = 1, 2, \dots, n\}$ 表示 n 个数的一個序列，那么我们就把这 n 个数的和写为：

$$\sum_{i=1}^n x_i \equiv x_1 + x_2 + \cdots + x_n \quad (\text{A.1})$$