

# 线性代数

This work is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) 

作者: gtj, 转载请注明出处

本书尝试着为线性代数的初学者们提供一种较为优雅而又不过于抽象的引入方式. 虽然经过了几次检查, 但由于笔者本人水平限制, 仍然难免有疏漏, 如果您对内容有任何建议, 欢迎发送邮件到 [2847562832@qq.com](mailto:2847562832@qq.com)

## 本书包含

第0章 集合、函数和记号

第1章 线性空间和线性映射

第2章 从  $\mathbb{R}^m$  到  $\mathbb{R}^n$  —— 矩阵怎么来的

第3章 从  $\mathbb{R}^n$  到  $V$  —— 基和表示理论

第4章 从  $(\mathbb{R}^n, \cdot)$  到  $H$  —— 内积和正交化

第5章 从  $V$  到  $V$  —— 特征值和行列式

## 第0章 集合、函数和记号

注: 本文中的句号都是点号, 这有利于防止与下标 0 混淆.

### 集合

大家高中都学过, 集合可以包含很多东西. 如果  $x$  是集合  $A$  的元素, 记  $x \in A$ , 否则记  $x \notin A$ . 如果  $x \in A$  可以推出  $x \in B$  则称  $A \subseteq B$ , 如果  $A \subseteq B$  且  $B \subseteq A$  则  $A, B$  有相同元素, 即  $A = B$ .

下文中实数和复数的记号分别为  $\mathbb{R}$  和  $\mathbb{C}$ , 其中  $\mathbb{C}$  中的元素形如  $a + bi$ , 对于集合  $A$ , 记号  $A^n$  表示包含所有元素来自  $A$  的  $n$  元组的集合,  $A \times B$  则包含了所有  $A, B$  中元素构成的有序对, 有序对前一个元素来自  $A$ , 后一个来自  $B$ .

例如, 令  $A = \{1, 2\}, B = \{1, 2, 3\}$  则有  $A \times B = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}$   
 $A^3 = \{(1, 1, 1), (1, 1, 2), (1, 2, 1), (1, 2, 2), (2, 1, 1), (2, 1, 2), (2, 2, 1), (2, 2, 2)\}$ .

在具体用词上, 有时我们用“空间”来代替“集合”, 同时用“点”代替“元素”.

### 函数

从集合  $A$  到集合  $B$  的函数  $f$  记作  $f: A \rightarrow B$ , 对每个  $x \in A$ , 它分配一个  $y \in B$ , 我们称之为  $y = f(x)$ . 如果  $x_1 \neq x_2$  时总有  $f(x_1) \neq f(x_2)$  则称  $f$  是单射 (或称  $f$  单), 反过来说, 单射要求每个  $f(x_1) = f(x_2)$  时可以推出  $x_1 = x_2$ . 如果一个对每个  $y \in B$  都有  $y = f(x)$  对某个  $x \in A$  成立, 则称  $f$  是满射 (或  $f$  满), 既单又满的映射被称为双射.

$A$  被称为定义域,  $B$  被称作陪域 (并非值域). 对于  $A$  的子集  $X, Y = \{f(x) | x \in X\}$  称作  $X$  在  $f$  下的像, 对于  $B$  的子集  $Y, X = \{x \in A | f(x) \in Y\}$  称为  $Y$  在  $f$  下的原象. 注意: 虽然我们在使用反函数的记号, 但是其实并没有依赖于反函数的定义, 当  $f$  是双射时,  $f^{-1}$  才能给出一个  $B$  到  $A$  的反函数. 特别地,  $f(A)$  称作  $f$  的值域.

有时我们使用“映射”来代替“函数”。

## 数学对象

通常我们关心的不是集合其中的元素或集合本身,而是其上的数学结构.例如,当讨论实数集时,我们实际上在讨论附加了额外结构的实数,这里指的是 $+$ , $\times$ , $\leq$ ,其中 $+$ , $\times$ 是实数上的二元运算, $\leq$ 为序关系(减法和除法可以由加和乘自然导出).虽然正式记号是 $(\mathbb{R}, +, \times, \leq)$ ,但是为了方便我们通常只写 $\mathbb{R}$ ,但是一定要有 $\mathbb{R}$ 带有额外结构的认识.

既然“有趣”的对象通常有额外的结构,我们考虑的映射通常保持对应的结构.例如,如果 $f(x_1 + x_2) = f(x_1) + f(x_2)$ 则称 $f: A \rightarrow B$ 保持了加法,这里 $x_1 + x_2$ 和 $f(x_1) + f(x_2)$ 的加法各自是 $A$ 和 $B$ 中的加法(并不一定完全相同).

如果 $A, B$ 之间有保持“一定”结构的双射则称 $f$ 给出 $A$ 和 $B$ 的一个同构(“一定”的具体含义取决于语境).

## 练习

0.1 设 $A = \{1, 2\}, B = \{1, 2, 3\}$ ,分别写出 $A \times B$ 和 $B \times A$ ,并说明这个乘积不具有交换性.

0.2 定义 $A + B = \{x + y | (x, y) \in A \times B\}$ .如果 $A = \{-1, 1, 2\}$ ,写出 $A + A$ .

0.3 考虑映射 $f: \mathbb{R}^2 \rightarrow \mathbb{C}$ ,其中 $f(x, y) = x + yi$ ,证明 $f$ 为双射.

0.4 对函数 $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ ,其中 $f(x) = x + a \ln x$ , $a$ 为实常数, $f$ 何时为(1)单射(2)满射(3)双射.

0.5 写出 $\mathbb{C}$ 的数学结构并指出其与 $\mathbb{R}$ 的异同.

0.6  $\mathbb{Q}$ 为有理数集,你能找到所有保持加法的 $f: \mathbb{Q} \rightarrow \mathbb{Q}$ 吗?证明你的结论.

# 第1章 线性空间和线性映射

## 线性空间

从很多领域都可以引出 $K$ -线性空间(或 $K$ -向量空间)的概念,其正式定义为一个域 $K$ 上的三元组 $(V, +, \cdot)$ ,满足八条性质.通常, $K$ 中的元素称为标量, $V$ 的元素称为向量, $+: V \times V \rightarrow V$ 称作向量加法, $\cdot: K \times V \rightarrow V$ 称作标量乘法.

八条性质分别为:

- (1) 对于所有 $\alpha, \beta \in V$ ,有 $\alpha + \beta = \beta + \alpha$ (向量加法的交换律)
- (2) 对于所有 $\alpha, \beta, \gamma \in V$ ,有 $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ (向量加法的结合律)
- (3) 存在 $0 \in V$ ,使得对于所有 $\alpha \in V$ , $0 + \alpha = \alpha$ (向量加法的零元素)
- (4) 对于每个 $\alpha \in V$ ,存在 $\beta \in V$ 使得 $\alpha + \beta = 0$ ,记作 $-\alpha$ (向量加法的逆元素)
- (5) 对于所有 $\alpha \in V$ ,有 $1 \cdot \alpha = \alpha$ (标量乘法的单位元素)
- (6) 对于所有 $k, l \in K$ 和 $\alpha \in V$ ,有 $(kl) \cdot \alpha = k \cdot (l \cdot \alpha)$ (标量乘法的结合律)
- (7) 对于所有 $k \in K$ 和 $\alpha, \beta \in V$ ,有 $k \cdot (\alpha + \beta) = k \cdot \alpha + k \cdot \beta$ (标量乘法的左分配律)
- (8) 对于所有 $k, l \in K$ 和 $\alpha \in V$ ,有 $(k + l) \cdot \alpha = k \cdot \alpha + l \cdot \alpha$ (标量乘法的右分配律)

对于初学者来说,一次性看到八条公理可能看起来很可怕,但实际上确实有必要.将 $K$ 替换为 $\mathbb{R}$ 或 $\mathbb{C}$ ,我们就得到了实/复线性空间的定义.

**练习:** 如果你想要抱怨,请拿起一本线性代数的书,用手把它举起来10分钟,保持手臂垂直于身体.并且得出举着这本书比读下去更加痛苦的结论.

最平凡的情况是  $\{0\}$ , 其中仅包含一个零向量. 为了方便起见, 我们将其写为  $0$  而不是  $\{0\}$  (本文中  $0$  会用于表示: 数域中的零元素、线性空间中的零向量、平凡线性空间以及例子中会提到的零映射, 这样的符号混用在数学中很普遍, 读者需要根据上下文区分其具体含义). 另一个简单的例子是  $K$  本身, 因为域本身具备满足结合律、交换律以及由分配律连接的加法和乘法运算 (换言之, 矢量和标量的概念并非互斥).

如果  $V_1 \subseteq V$  且  $V_1$  和  $V$  都是  $K$ -线性空间, 那么  $V_1$  称作  $V$  的 ( $K$ -线性)子空间. 根据定义可以推出, 如果  $V_1$  对  $V$  中定义的加法和乘标量乘法封闭,  $V_1$  是  $V$  的线性子空间. 从  $V$  继承的向量加法和数乘保证了  $V_1$  满足其他公理. 如果  $V_1$  和  $V_2$  是  $V$  的两个子空间, 可以证明  $V_1 \cap V_2$  也是一个子空间(留作练习). 然而, 并集  $V_1 \cup V_2$  并不总是一个线性子空间, 因为它可能对加法不封闭. 我们用  $V_1 + V_2$  表示集合  $\{\alpha + \beta \mid \alpha \in V_1, \beta \in V_2\}$ , 并称其为由  $V_1$  和  $V_2$  张成的空间. 这是包含  $V_1$  和  $V_2$  并对向量加法和标量乘法封闭的最小空间(也留作练习).

如果  $V_1, V_2$  是  $K$  线性空间, 则用  $V_1 \oplus V_2$  表示  $\{(\alpha, \beta) \mid \alpha \in V_1, \beta \in V_2\}$  并称之为  $V_1$  与  $V_2$  的直和, 其上的运算定义为  $(\alpha_1, \beta_1) + (\alpha_2, \beta_2) = (\alpha_1 + \alpha_2, \beta_1 + \beta_2)$ ,  $k(\alpha, \beta) = (k\alpha, k\beta)$ . 通过简单观察不难发现直和和前一章所述集合的直积几乎相同, 唯一区别是“直和”这一用语专用于线性空间而不是一般的集合. 虽然  $K$  是其自身上的线性空间可能并不能引起大家足够的重视, 但是  $K^n$ , 其中  $n$  为整数是一族重要的线性空间 (回忆上一章集合  $n$  次方的定义). 它们带有上文所述的向量加法和标量乘法, 是我们下一章研究的主要对象.

## 线性映射

若  $V_1, V_2$  是  $K$ -线性空间, 则一个  $K$ -线性映射  $\varphi: V_1 \rightarrow V_2$  是满足以下条件的映射:

- (1) 对于所有  $\alpha, \beta \in V$ , 有  $\varphi(\alpha + \beta) = \varphi(\alpha) + \varphi(\beta)$
- (2) 对于所有  $k \in K, \alpha \in V$ , 有  $\varphi(k\alpha) = k\varphi(\alpha)$

因此, 我们可以说线性映射保持向量加法和标量乘法.

线性映射的像, 记为  $\text{Im } \varphi$ , 表示  $\varphi(V_1)$ , 或称之为  $\varphi$  的值域. 线性映射的核, 记为  $\text{Ker } \varphi$ , 表示  $\varphi^{-1}(0)$ . 核是零点的集合. 读者可以轻松地验证  $\text{Im } \varphi$  是  $V_2$  的线性子空间, 而  $\text{Ker } \varphi$  是  $V_1$  的线性子空间.

如果  $V_1 = V_2 = V$ , 则称  $\varphi$  为  $V$  上的线性变换.

如果  $\varphi$  是双射, 则称  $\varphi$  给出了  $V_1$  和  $V_2$  之间的线性同构.

## 例子

下面的例子主要是  $\mathbb{R}$  上的例子, 因为  $\mathbb{R}$  是我们最熟悉的数域, 有利于初学者理解.

一些仅违背了一条公理的例子, 这使它们不能成为线性空间, 以下的加法均表示  $\mathbb{R}$  或  $\mathbb{C}$  上原有的加法:

- 人为向实数域中添加一个新的元素, 标记为  $0'$ , 实数域上的加法和乘法不变, 并定义  $0' + 0 = 0$ , 对  $x \neq 0, x + 0' = x$ , 对所有  $k \in \mathbb{R}$ , 定义  $k \cdot 0' = 0'$ , 则  $(\mathbb{R} \cup \{0'\}, +, \cdot)$  不是线性空间, 因为没有满足公理(4)
- 对于所有  $k \in \mathbb{R}, x \in V = \mathbb{R}$ , 定义  $k \cdot x$  恒为  $0$ , 则  $(V, +, \cdot)$  不是  $\mathbb{R}$ -线性空间, 因为没有满足公理(5).
- 对于所有  $k \in \mathbb{C}, z \in V = \mathbb{C}$ , 记  $\text{Re } k$  表示  $k$  的实部, 即  $a + bi$  实部为  $a$ , 定义  $k \cdot z = (\text{Re } k)z$ . 则  $(V, +, \cdot)$  不是  $\mathbb{C}$ -线性空间, 因为没有满足公理(6).
- 对于所有  $k \in \mathbb{C}, z \in V = \mathbb{C}$ , 如果  $z$  的虚部大于等于  $0$ , 定义  $k \cdot z = kz$ ; 如果  $z$  的虚部小于  $0$ , 定义  $k \cdot z = \bar{k}z$ . 则  $(V, +, \cdot)$  不是  $\mathbb{C}$ -线性空间, 因为没有满足公理(7).
- 对于所有  $k \in \mathbb{R}, x \in V = \mathbb{R}$ , 定义  $k \cdot x = x$ . 则  $(V, +, \cdot)$  不是  $\mathbb{R}$ -线性空间, 因为没有满足公理(8).

一些空间中的图形, 有的是线性空间, 有的不是:

- $\mathbb{R}^n$ , 比如  $\mathbb{R}^2$ (平面)和  $\mathbb{R}^3$ (三维空间), 都是  $\mathbb{R}$ -线性空间. (注意, 到此为止其实还没有给出“维度”的正式定义.)
- 直线  $x_2 = x_1$  是  $\mathbb{R}^2$  的一个线性子空间. 实际上, 它是线性映射  $\varphi: \mathbb{R} \rightarrow \mathbb{R}^2$  的像, 其中  $\varphi(t) = (t, t)$ .
- 平面  $\{(x_1, x_2, x_3) \mid x_1 + x_2 + x_3 = 0, x_1, x_2, x_3 \in \mathbb{R}\}$  是  $\mathbb{R}^3$  的一个线性子空间. 它是线性映射  $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}$  的核, 其中  $\varphi(x_1, x_2, x_3) = x_1 + x_2 + x_3$ .
- 考虑两个平面  $V_1 = \{(x_1, x_2, 0) \mid x_1, x_2 \in \mathbb{R}\}$  和  $V_2 = \{(x_1, 0, x_3) \mid x_1, x_3 \in \mathbb{R}\}$ . 它们都是  $\mathbb{R}^3$  的线性子空间. 很明显,  $V_1 \cap V_2 = \{(x_1, 0, 0) \mid x_1 \in \mathbb{R}\}$  是  $\mathbb{R}^3$  的一个子空间, 但是  $V_1 \cup V_2 = \{(x_1, x_2, x_3) \mid x_1 \in \mathbb{R}, x_2 x_3 = 0\}$  不是  $\mathbb{R}^3$  的子空间. 例如,  $(0, 1, 0) + (0, 0, 1) = (0, 1, 1) \notin V_1 \cup V_2$ . 此外,  $\mathbb{R}^3$  中的每个点都可以表示为  $(x_1, x_2, x_3) = (x_1, x_2, 0) + (0, 0, x_3)$ , 其中  $(x_1, x_2, 0) \in V_1$  且  $(0, 0, x_3) \in V_2$ . 因此, 由  $V_1$  和  $V_2$  张成的空间是整个  $\mathbb{R}^3$  (你能找到更多将向量分解为  $V_1$  和  $V_2$  中向量的方法吗? ).
- 平面上的直线  $\{(x_1, x_2) \mid x_2 = x_1 + 1, x_1, x_2 \in \mathbb{R}\}$  不是  $\mathbb{R}$ -线性空间, 尽管它在几何上表示一条直线.
- 平面上的格点, 例如  $\{(n_1, n_2) \mid n_1, n_2 \in \mathbb{Z}\}$ , 不是  $\mathbb{R}$ -线性空间, 尽管它呈现出周期性的模式, 具有对称性, 并沿着直线排列.

此外是一些略微抽象的例子, 不像前面那样可以用直观感觉想象:

- $\mathbb{R}$  是一个  $\mathbb{Q}$ -线性空间.
- 所有次数小于 2 的实系数多项式构成一个  $\mathbb{R}$ -线性空间.
- 所有从区间  $[0, 1]$  到实数集  $\mathbb{R}$  的连续函数构成一个  $\mathbb{R}$ -线性空间. 这个空间记作  $C[0, 1]$
- 所有形如  $f(x) = a \sin(x + \phi)$  的函数, 其中  $a, \phi \in \mathbb{R}$ , 构成一个  $\mathbb{R}$ -线性空间.
- 所有从  $\mathbb{R}$ -线性空间  $V$  到实数集  $\mathbb{R}$  的线性映射  $\varphi$  构成一个  $\mathbb{R}$ -线性空间. 这里, 加法运算定义为  $(\varphi_1 + \varphi_2)(\alpha) = \varphi_1(\alpha) + \varphi_2(\alpha)$ , 标量乘法定义为  $(k \cdot \varphi)(\alpha) = k\varphi(\alpha)$ .

最后考虑一些线性映射的例子. 在读之前, 请先自行验证线性映射一定将零向量映射到零向量.

- $V$  上的标量乘法可以诱导一族以  $K$  中元素为索引的线性映射, 其中  $\varphi_k(\alpha) = k \cdot \alpha$ , 特别地, 由于  $\varphi_1(\alpha) = 1 \cdot \alpha = \alpha$ , 记  $\varphi_1$  为  $\text{Id}$ , 称之为  $V$  上的恒同变换
- 同时, 标量乘法还可以诱导一族以  $V$  中元素索引的  $K$  到  $V$  的线性映射. 即  $\varphi_\alpha(k) = k \cdot \alpha$ . 例如, 设  $V = \mathbb{R}^2, \alpha = (1, 1) \in V$ , 则  $\varphi_\alpha(x) = (x, x)$ .
- 对于  $K$  上的线性空间  $V_1, V_2$ , 总是存在一个零映射, 即  $0(\alpha) \equiv 0$ .
- 平面绕其原点的旋转是一个线性映射, 对于逆时针旋转  $\theta$  角的情况, 画一幅图并思考  $(x_1, x_2)$  旋转到了哪里.
- 然而平移不是线性映射, 举一个最简单的例子, 定义在  $\mathbb{R}$  上的  $f(x) = x + 1$ , 它将实数轴向正方向移动了一个单位长度, 为什么它不是线性变换?
- 对于所有连续实值函数  $f: [0, 1] \rightarrow \mathbb{R}$ , 定义  $\varphi(f) = f(0) + f(1)$ , 则  $\varphi$  将  $C[0, 1]$  线性映射到  $\mathbb{R}$ . 类似这样的映射通常被我们称作线性泛函, 是泛函分析研究的重要对象 (虽然并不是我们现在讨论的线性代数的重点) .

## 练习

1.1 证明对于线性空间  $V$  的子空间  $V_1$  和  $V_2$ , 它们的交集也是一个子空间, 并且  $V_1 + V_2$  是包含  $V_1$  和  $V_2$  的最小子空间.(这里的"最小"是指对于任意包含  $V_1$  和  $V_2$  的子空间  $V_3$ ,  $V_1 + V_2 \subseteq V_3$ ).

1.2 对线性空间中的两个向量  $\alpha, \beta$ , 如果  $k\alpha + l\beta = 0$  不能推出  $k = l = 0$ , 则称向量  $\alpha, \beta$  共线. 我们已经知道  $\mathbb{R}^2$  上有一个标准的加法和标量乘法结构, 现在  $\mathbb{R}^2$  上定义新的运算  $\tilde{+}$ , 如果  $\alpha, \beta$  共线, 定义  $\alpha \tilde{+} \beta = \alpha + \beta$ , 若  $\alpha, \beta$  不共线, 定义  $\alpha \tilde{+} \beta = 0$ , 并记  $\cdot$  仍表示  $\mathbb{R}^2$  上标准的标量乘法.  $(\mathbb{R}^2, \tilde{+}, \cdot)$  是否构成  $\mathbb{R}$  上的线性空间?

1.3 对于那些满足线性空间条件的示例, 请验证公理成立. 对于不满足条件的示例, 请通过给出一个或多个反例来证明它为什么不满足公理. 此外指出并证明它满足了哪些性质.

1.4 对于将  $\mathbb{R}^2$  中的向量  $(x_1, x_2)$  映射为  $(x_2, 0)$  的线性变换  $\varphi$ , 找出它的核和像.

1.5 你能找到所有从  $\mathbb{R}$  到  $\mathbb{R}^n$  的线性映射吗?

1.6 你能找到所有从  $\mathbb{R}^n$  到  $\mathbb{R}$  的线性映射吗?

1.7 证明线性映射  $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$  的核不可能为  $0$ , 并从另一个角度解释这个结果.

1.8 证明线性映射  $\varphi: \mathbb{R} \rightarrow \mathbb{R}^2$  像不可能为  $\mathbb{R}^2$ , 并从另一个角度解释这个结果.

1.9 证明对于线性空间  $V$  的子空间  $V_1$  和  $V_2$ , 如果  $\alpha_1 \in V_1, \alpha_2 \in V_2, \alpha_1 \notin V_2$  且  $\alpha_2 \notin V_1$ , 那么  $\alpha_1 + \alpha_2$  不属于  $V_1 \cup V_2$ .

1.10 记  $V_1 = \{(x_1, x_2, 0) \mid x_1, x_2 \in \mathbb{R}\}$ ,  $V_2 = \{(x_1, 0, x_3) \mid x_1, x_3 \in \mathbb{R}\}$ .  $V_1 + V_2$  和  $V_1 \oplus V_2$  相同吗?  $\mathbb{R}^3$  中的加法将  $V_1 \oplus V_2$  线性映射到  $V_1 + V_2$ , 找到它的核, 并说明这个核与把  $V$  中元素拆分成一对分别来自  $V_1, V_2$  中的向量的不同方法有什么关系.

## 第2章 从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ —— 矩阵怎么来的

### 矩阵 —— 行的视角

看之前请确保你做完了之前一章的习题(什么? 您还没有完成练习题? 回去做练习, 请, 做完再来.) 在本章中, 虽然在  $\mathbb{R}$  还是随便的什么域  $K$  上讨论并不影响我们的结果, 但是我们还是把主要关注点将放在  $\mathbb{R}^n$  上. 然而, 值得注意的是, 将  $\mathbb{R}$  替换为任意域  $K$  对我们的结果并没有太大影响. 为方便起见, 我们将在本文中讨论  $\mathbb{R}$  上的情况. 在本章中, 默认情况下, 当我们提到向量空间时, 我们指的是实向量空间, 其中向量加法和标量乘法是标准的向量加法和标量乘法, 其中标量乘法中的点符号, 比如说  $k \cdot x$  里面的点, 将被省略.

首先考虑最简单的例子: 当  $m = n = 1$  时. 假设有线性函数  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ , 对于任意的  $x \in \mathbb{R}$ , 有  $\varphi(x)$  表示为  $\varphi(x \cdot 1) = x\varphi(1) = kx$ , 其中  $k = \varphi(1)$  是一个常数. 用  $[k]$  表示这个映射, 则  $\varphi(x) = kx$ . 换句话说,  $[k]$  是  $\varphi$  函数的一个别名. 和  $\varphi$  唯一的区别就是记号不同,  $[k]$  直接表明了它是干嘛的.

现在考虑  $\mathbb{R}^n$  上的线性函数, 即  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$ . 回忆上一章的练习, 你会发现这个函数只能是  $\varphi(x_1, x_2, \dots, x_n) = k_1x_1 + k_2x_2 + \dots + k_nx_n$ . 我们可以把这个映射记作  $[k_1 \ k_2 \ \dots \ k_n]$ . 对于  $n = 1$  时, 这个符号与上一段的相兼容. 再次强调, 到目前为止, 这个符号只是映射的一个别名, 我们并没有赋予它任何额外的含义. 在某种程度上, 如果你学过哪门包含 lambda 表达式的编程语言, 你会发现它差不多就是个 lambda 表达式

例如,  $[2 \ 3]$  可以用 Python 表示为:

```
lambda x, y: 2*x + 3*y
```

如果你在 Python 中运行以下代码：

```
print(
    (lambda x, y: 2*x + 3*y)(2, 1)
)

# 正式代码别这么写
# 只是为了解释上文含义才会写出这个
```

结果将与下面的代码相同：

```
def f(x, y):
    return 2*x + 3*y
print(f(2, 1))
```

利用我们刚刚定义的符号，可以写出  $[2 \ 3](2, 1) = 2 \cdot 2 + 3 \cdot 1 = 7$ 。有些人一看，哦，这不向量内积吗，虽然确实有关系，但是我不打算在这章讲。

同样地，来考虑  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m$  的情况。这个映射可以写成  $y = \varphi(x) = (\varphi_1(x), \varphi_2(x), \dots, \varphi_m(x))$ ，我们断言它每个分量  $\varphi_i$  ( $1 \leq i \leq m$ ) 都是一个实值线性函数。这意味着  $\varphi_i$  必须具有形式  $\varphi_i(x_1, x_2, \dots, x_n) = k_{i1}x_1 + k_{i2}x_2 + \dots + k_{in}x_n$ 。把  $\varphi_i$  按行排列，我们得到了一个数表用于表示一般的线性映射：

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

所以这就是矩阵。每一行表示原始空间如何线性映射到新空间的相应分量上。因此，元素  $k_{ij}$  表示  $\mathbb{R}^n$  的第  $j$  个分量如何影响  $\mathbb{R}^m$  的第  $i$  个分量。例如，如果我们想计算：

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

左边的矩阵表示一个从  $\mathbb{R}^2$  到  $\mathbb{R}^3$  的线性映射。（我们没有，也不打算给它一个具体的名称，如  $\varphi$ ）右侧的向量  $(1, 1)$  是  $\mathbb{R}^2$  中的一个元素。计算它需要一步步来。首先，第一个分量，根据我们的定义： $2 \cdot 1 + 1 \cdot 2 = 4$ 。第二个： $1 \cdot 1 + 2 \cdot 2 = 5$ 。第三个： $0 \cdot 1 + 3 \cdot 2 = 6$ 。所以结果是  $(4, 5, 6)$ 。

这个过程可以理解为将  $\mathbb{R}^2$  中的向量从矩阵的上方输入，将相应的项相乘，并将它们求和并从矩阵左侧输出。

$$\begin{array}{r|cc} & 1 & 2 \\ & \downarrow & \downarrow \\ 4 & \leftarrow & 2 \cdot 1 & 1 \cdot 2 \\ 5 & \leftarrow & 1 \cdot 1 & 2 \cdot 2 \\ 6 & \leftarrow & 0 \cdot 1 & 3 \cdot 2 \end{array}$$

## 矩阵 —— 列的视角

在上一节中,我们将注意力集中在将  $\mathbb{R}^m$  分割为  $m$  个分量,并分别考虑  $x \in \mathbb{R}^n$  到它们的函数,这是行的视角.现在,换成把  $\mathbb{R}^n$  分割为其各个分量,这次就是列的视角了.

为了研究矩阵的列,首先观察一列的元素有什么共同点.对于  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,记  $y = \varphi(x)$ ,并记  $k_{ij}$  为  $\varphi$  的矩阵表示中第  $i$  行第  $j$  列的分量,根据上一节的定义:

$$\begin{aligned}y_1 &= k_{11}x_1 + k_{12}x_2 + \cdots + k_{1n}x_n \\y_2 &= k_{21}x_1 + k_{22}x_2 + \cdots + k_{2n}x_n \\&\cdots \\y_m &= k_{m1}x_1 + k_{m2}x_2 + \cdots + k_{mn}x_n\end{aligned}$$

改写一下这个式子的写法,按照  $x$  的分量整理,我们可以把它重写成:

$$\begin{aligned}(y_1, y_2, \cdots, y_m) &= x_1(k_{11}, k_{21}, \cdots, k_{m1}) \\&\quad + x_2(k_{12}, k_{22}, \cdots, k_{m2}) \\&\quad + \cdots \\&\quad + x_n(k_{1n}, k_{2n}, \cdots, k_{mn})\end{aligned}$$

我们发现这里每个向量都对应到矩阵的一列,矩阵中第  $j$  列的数值都与向量  $x$  的第  $j$  个分量有关.令  $(1, 0, \cdots, 0)$  表示为  $e_1$ ,  $(0, 1, \cdots, 0)$  表示为  $e_2$ , ...,  $(0, 0, \cdots, 1)$  表示为  $e_n$ .显然  $\mathbb{R}^n$  中的任意向量  $x$  可以唯一地表示为  $\sum_{j=1}^n x_j e_j$ .

回顾前一章的练习,我们发现从  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射  $\varphi$  完全由其在  $e_j$  处的取值决定,这是因为

$\varphi\left(\sum_{j=1}^n x_j e_j\right) = \sum_{j=1}^n x_j \varphi(e_j)$ . 将与上述公式进行比较,我们得到了矩阵的列解释: 每一列对应于  $\mathbb{R}^m$  中的一个向量,而每个  $\mathbb{R}^n$  中的向量通过对列进行线性组合而映射到  $\mathbb{R}^m$ .

所以我们可以用另外一种方式考虑矩阵作用,仍然考虑:

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

这次我们换个方法计算:

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 0 & 3 \end{bmatrix} (1, 2) = 1 \cdot (2, 1, 0) + 2 \cdot (1, 2, 3) = (4, 5, 6)$$

虽然计算的方法变了,但是结果没变,因为两种方法本来就是等价的.

## 集合在矩阵下的像 —— 一些例子

既然矩阵表示线性映射,那这个映射的图像长什么样? 来看几个例子.

$[k]: \mathbb{R} \rightarrow \mathbb{R}, y = [k](x) = kx$  就是个正比例函数

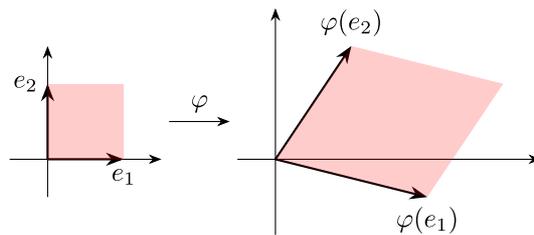
$I_n: \mathbb{R}^n \rightarrow \mathbb{R}^n$ , 用  $I_n(x) = x$  表示  $\mathbb{R}^n$  上的恒同映射,在上一章中我们用  $\text{Id}$  表示它,在映射下  $e_j$  的像就是  $e_j$ ,所以除了对角线上全是 1,其它元素都是 0.这个矩阵也被称为  $n$  阶单位矩阵,它一个图形原封不动地映射到自身.

$[k_1 \ k_2]: \mathbb{R}^2 \rightarrow \mathbb{R}, y = [k_1 \ k_2](x_1, x_2) = k_1x_1 + k_2x_2$  在空间中的图像是一个平面. (如果你对平面方程不熟悉, 打开<https://www.geogebra.org/3d>, 试着输入  $z = 3x + 2y$  来画一个平面)

对  $\varphi = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}: \mathbb{R} \rightarrow \mathbb{R}^2$ , 选择一个  $k_1, k_2$ , 如果你选择困难, 就试试  $k_1 = 1, k_2 = 2$ , 对于若干个  $x$ , 描出对应的  $\varphi(x)$ , 如果你选的  $k$  不全为 0, 这些点会在一条直线上运动, 这条直线就是矩阵的像.

对于  $\varphi = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ , 画出它的图像并不容易, 所以不妨考虑单位正方形  $[0, 1] \times [0, 1]$  在映射下的像, 从而了解它如何作用于平面. 由于  $\varphi(x_1, x_2)$  是  $\varphi(e_1)$  和  $\varphi(e_2)$  的线性组合, 如果二者不共线, 结果应该是一个平行四边形.

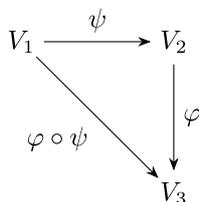
例如, 考虑  $\varphi = \begin{bmatrix} 2 & 1 \\ -0.5 & 1.5 \end{bmatrix}$ , 可以画出单位正方形在映射下的像如下:



通过观察样例, 你大概可以想象它如何作用于平面.

## 映射的复合和矩阵乘法

众所周知, 如果有一个  $\psi: V_1 \rightarrow V_2, \varphi: V_2 \rightarrow V_3$ , 那么可以写出它的复合映射  $\varphi \circ \psi$ , 其将  $x \in V_1$  映射到  $(\varphi \circ \psi)(x) \in V_3$ , 其中  $(\varphi \circ \psi)(x) = \varphi(\psi(x))$ , 它们连接的方式像下图这样, 下图一般被称作叫交换图:

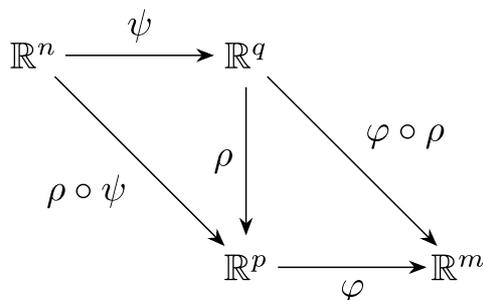


稍微停下来想想为什么  $\varphi$  和  $\psi$  都线性的时候它们的复合也是线性的, 我们可以这样证明 (纯粹的用定义和倒腾记号):

$$\begin{aligned} (\varphi \circ \psi)(k \cdot \alpha) &= \varphi(\psi(k \cdot \alpha)) = \varphi(k \cdot \psi(\alpha)) = k \cdot \varphi(\psi(\alpha)) = k \cdot (\varphi \circ \psi)(\alpha) \\ (\varphi \circ \psi)(\alpha + \beta) &= \varphi(\psi(\alpha + \beta)) = \varphi(\psi(\alpha) + \psi(\beta)) = \\ &= \varphi(\psi(\alpha)) + \varphi(\psi(\beta)) = (\varphi \circ \psi)(\alpha) + (\varphi \circ \psi)(\beta) \end{aligned}$$

取  $V_1$  为  $\mathbb{R}^n, V_2$  为  $\mathbb{R}^p, V_3$  为  $\mathbb{R}^m$ , 则一个  $p \times n$  的矩阵给出一个  $V_1 \rightarrow V_2$  的映射  $\psi$ , 一个  $m \times p$  的矩阵给出一个  $V_2 \rightarrow V_3$  的映射  $\varphi$ , 所以它们的复合  $\varphi \circ \psi$  可以表示为一个  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  的映射, 这可以表示为一个  $m \times n$  的矩阵. 注意到先作用的矩阵写在后面, 这可能与日常书写方向有一定差别. 所以映射的复合实际上给出了矩阵乘法. 由于两个映射复合的要求, 左边这个矩阵的列数和右边这个矩阵的行数必须匹配.

一个重要的结果是矩阵乘法满足结合律, 也就是说如果  $\varphi$  是  $m \times p$  的矩阵,  $\rho$  是  $p \times q$  的矩阵,  $\psi$  是  $q \times n$  的矩阵, 则  $(\varphi \circ \rho) \circ \psi = \varphi \circ (\rho \circ \psi)$  成立, 因为两边的式子都定义为  $\varphi(\rho(\psi(x)))$ , 从下图明显可以看出上式成立:



到目前为止, 我们还只知道映射复合的性质, 不知道映射复合或称之为矩阵乘法的公式, 所以接下来我们要探索这个问题. 我个人比较喜欢从列的角度看矩阵, 所以下面的内容是从列的角度来的, 如果你习惯行, 你大可以从行的角度导出相同的结果.

假设  $\mathbb{R}^n \xrightarrow{\psi} \mathbb{R}^p \xrightarrow{\varphi} \mathbb{R}^m$ , 为了计算  $\varphi \circ \psi$  的每一列, 即  $e_j$  在  $\varphi \circ \psi$  的像, 我们只需要找到  $\psi(e_j)$  在  $\varphi$  下的像. 假设  $\varphi$  矩阵表示中第  $i$  行第  $t$  列元素为  $a_{it}$ ,  $\psi$  矩阵表示中第  $t$  行第  $j$  列的元素为  $b_{tj}$ , 则  $e_j$  在  $\psi$  下的像为  $\psi(e_j) = (b_{1j}, b_{2j}, \dots, b_{pj})$ , 即  $\psi$  矩阵表示的第  $j$  列对应的向量. 于是根据矩阵作为线性映射的定义, 我们有:

$$\begin{aligned} \varphi(\psi(e_j)) &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mp} \end{bmatrix} (b_{1j}, b_{2j}, \dots, b_{pj}) \\ &= (a_{11}b_{1j} + a_{12}b_{2j} + \cdots + a_{1p}b_{pj}, \\ &\quad a_{21}b_{1j} + a_{22}b_{2j} + \cdots + a_{2p}b_{pj}, \\ &\quad \cdots, \\ &\quad a_{m1}b_{1j} + a_{m2}b_{2j} + \cdots + a_{mp}b_{pj}) \end{aligned}$$

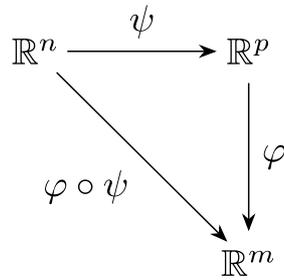
所以复合映射的矩阵第  $i$  行第  $j$  列应该是

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{t=1}^p a_{it}b_{tj}$$

对于所有的  $i, j$  列出来, 得到乘积矩阵长这样

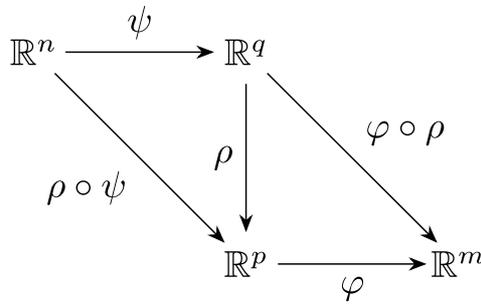
$$\varphi \circ \psi = \begin{bmatrix} \sum_{t=1}^p a_{1t}b_{t1} & \sum_{t=1}^p a_{1t}b_{t2} & \cdots & \sum_{t=1}^p a_{1t}b_{tn} \\ \sum_{t=1}^p a_{2t}b_{t1} & \sum_{t=1}^p a_{2t}b_{t2} & \cdots & \sum_{t=1}^p a_{2t}b_{tn} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^p a_{mt}b_{t1} & \sum_{t=1}^p a_{mt}b_{t2} & \cdots & \sum_{t=1}^p a_{mt}b_{tn} \end{bmatrix}$$

回忆这个图:



我们发现  $\mathbb{R}^n$  中第  $j$  个分量影响  $\mathbb{R}^m$  中第  $i$  个分量的结果（即乘积矩阵第  $i$  行第  $j$  列的值）可以视作把所有通过  $\mathbb{R}^p$  的可能路径加了起来. 如果你仍然对什么叫把所有可能路径加起来感到疑惑, 请思考如下问题: 一个醉鬼在大街上漫无目的地游荡, 每一步只能向前或向后走, 且走下一步时有 20% 的可能转向, 如果他某一步在向前走, 那么他两步后向前走的概率为多少?

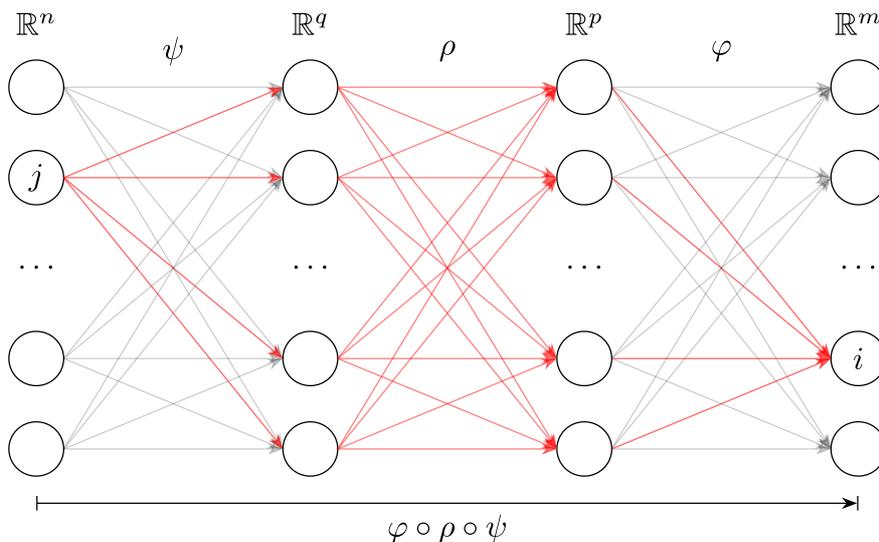
其实非常简单: 为了让方向和两步之前保持相同, 如果他第一步没转向, 这时他第二次也不应该转向, 则可能性为  $80\% \times 80\% = 64\%$ , 如果他第一步转向了, 他第二步应该转回来, 可能性为  $20\% \times 20\% = 4\%$ , 所以总的概率是  $64\% + 4\% = 68\%$ , 这就是我说的把所有中间可能经过的状态加起来. 类似地, 如果有  $\varphi, \rho, \psi$  矩阵表示中的元素分别为  $a_{it_1}, b_{t_1t_2}, c_{t_2j}$ , 且关系如下:



则乘积矩阵中第  $i$  行第  $j$  列的元素应该为

$$\sum_{t_1=1}^p \sum_{t_2=1}^q a_{it_1} b_{t_1t_2} c_{t_2j}$$

这同样可以通过“对所有可能的路径求和”来解释



如果从计算的角度上看, 我们也可以用蛮力证明矩阵乘法的结合律. (但是我认为 not elegant) 我们假设读者读到这里已经清楚矩阵乘法是如何从映射复合中导出的, 既然结合律成立, 接下来我们将会略去矩阵乘法间的复合符号, 因为结合顺序不改变结果.

## 又是符号混用 —— 感觉还行

如果我们把应该  $\mathbb{R}$  到  $\mathbb{R}^n$  的映射和一个  $n \times 1$  矩阵做对比, 我们发现它们干的事情是相同的:

一个  $n \times 1$  矩阵把  $t \in \mathbb{R}$ , 线性映射到  $x \in \mathbb{R}^n$ . 而一个向量通过数乘也能给出  $\mathbb{R}$  到  $\mathbb{R}^n$  的线性映射, 而且二者结果完全相同.

$$\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} (t) = (k_1 t, k_2 t, \dots, k_n t)$$
$$t \cdot (k_1, k_2, \dots, k_n) = (tk_1, tk_2, \dots, tk_n)$$

$\mathbb{R}$  上 (或者说任何数域) 上的乘法可交换, 所以上面两个是相同的, 从作用在  $\mathbb{R}$  上看,  $\mathbb{R}^n$  中向量和  $n \times 1$  矩阵作用完全相同.

接下来考虑一个  $m \times n$  的矩阵对它们有什么影响, 首先, 根据上一节矩阵乘法的公式:

$$\begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} k_{11}x_1 + k_{12}x_2 + \cdots + k_{1n}x_n \\ k_{21}x_1 + k_{22}x_2 + \cdots + k_{2n}x_n \\ \vdots \\ k_{m1}x_1 + k_{m2}x_2 + \cdots + k_{mn}x_n \end{bmatrix}$$

同时根据矩阵的定义:

$$\begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{bmatrix} (x_1, x_2, \dots, x_n)$$
$$= (k_{11}x_1 + k_{12}x_2 + \cdots + k_{1n}x_n, \\ k_{21}x_1 + k_{22}x_2 + \cdots + k_{2n}x_n, \\ \cdots, \\ k_{m1}x_1 + k_{m2}x_2 + \cdots + k_{mn}x_n)$$

通过比较二者, 我们除了符号没有发现任何区别, 所以我们说  $\mathbb{R}^n$  中向量和  $m \times 1$  的矩阵有着典范对应关系. (什么是典范一会再说) 对于每一个  $x \in \mathbb{R}^n$ , 都可以找到一个  $\varphi: \mathbb{R} \rightarrow \mathbb{R}^m$  满足  $tx = \varphi(t)$ , 且反之亦然. 所以我们大胆丢掉原本  $\mathbb{R}^n$  中向量的符号, 把它们全部换成矩阵记号并且称一个单列的矩阵为一个列向量. 所以现在  $\mathbb{R}^n$  中的元素也是矩阵,  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  的映射也是矩阵. 有一句话说到: “人是一切社会关系的总和”, 这也差不多, 一定程度上我们不关心一个数学对象原本是什么, 我们更关心一个它如何和其它对象产生关系.

为了解释数学中“典范”一词的含义, 先来考虑行向量和列向量的关系, 有些人一想就觉得二者肯定是一一对应的, 只需要改个方向写就行. 虽然它们确实可以这么对应起来, 但是通常不认为它是典范的, 这是因为行向量和列向量做的事完全不一样, 行向量把  $\mathbb{R}$  映射到  $\mathbb{R}^n$ , 但是列向量刚好相反, 所以它们本质上并不相同 (有人又要说了, 列向量不也能通过内积把  $\mathbb{R}^n$  映射到  $\mathbb{R}$  吗, 对于一个  $v$ , 内积  $\langle v, - \rangle$  把  $\mathbb{R}^n$  映射到  $\mathbb{R}$ , 其中  $x$  被映射到  $\langle v, x \rangle$  但是问题在于这个同构依赖于内积, 而一个向量空间并不总是定义了内积, 所以二者还是有本质区别的)

总的来说, 典范的意思是一个对应关系或者表示方法在一定的语境下足够自然或者标准. 典范的含义是在一定程度上是最优且唯一的表示方法, 所以, 如果我们用  $\mathbb{R}^{m \times n}$  表示从  $m \times n$  的矩阵全体, 则  $\mathbb{R}^{n \times 1}$  与  $\mathbb{R}^n$  之间有典范的同构, 但是和  $\mathbb{R}^{1 \times n}$  之间则没有.

## 线性方程 —— 高斯消元

一个线性方程(组)是一组如下形式的等式:

$$\begin{aligned}k_{11}x_1 + k_{12}x_2 + \cdots + k_{1n}x_n &= y_1 \\k_{21}x_1 + k_{22}x_2 + \cdots + k_{2n}x_n &= y_2 \\&\dots \\k_{m1}x_1 + k_{m2}x_2 + \cdots + k_{mn}x_n &= y_m\end{aligned}$$

最经典且直接的解法是高斯消元, 一层层地消元直到化简到最简单的方程.

首先假设  $k_{11} \neq 0$ , 通过从  $y_i$  ( $i \geq 2$ ) 中减去  $\frac{k_{i1}}{k_{11}}y_1$  可以消去  $k_{i1}$  所在的项, 在这次操作后, 方程会变成:

$$\begin{aligned}k_{11}x_1 + k_{12}x_2 + \cdots + k_{1n}x_n &= y_1 \\+(k_{22} - \frac{k_{21}}{k_{11}}k_{12})x_2 + \cdots + (k_{2n} - \frac{k_{21}}{k_{11}}k_{1n})x_n &= y_2 \\&\dots \\+(k_{m2} - \frac{k_{m1}}{k_{11}}k_{12})x_2 + \cdots + (k_{mn} - \frac{k_{m1}}{k_{11}}k_{1n})x_n &= y_m\end{aligned}$$

通过从  $y_2 \sim y_m$  中解出  $x_2 \sim x_n$ , 我们可以很容易解出  $x_1$ , 所以解一个  $n$  元方程的问题变成了解  $n - 1$  元方程的问题, 一层层消下去只需要把  $x_1$  解出即可.

如果  $k_{11} = 0$ , 但是它的下方有非 0 的数字, 只需要把这一行和非 0 开头的行交换即可继续消元.

如果第一列的所有元素都是 0, 那么这个等式就和  $x_1$  完全没关系, 所以实际上只剩下了  $n - 1$  个未知数.

通过这一套操作下来, 我们可以一步步把问题化简到剩余一个等式, 例如如果我们想解:

$$\begin{aligned}x_1 + 2x_2 + 3x_3 &= 1 \\4x_1 + 5x_2 + 6x_3 &= 2 \\7x_1 + 8x_2 + 9x_3 &= 3\end{aligned}$$

首先拿第二列减去 4 倍的第一列, 拿第三列减去 7 倍的第一列:

$$\begin{aligned}x_1 + 2x_2 + 3x_3 &= 1 \\-3x_2 - 6x_3 &= -2 \\-6x_2 - 12x_3 &= -4\end{aligned}$$

接下来拿第三列减去两倍的第二列:

$$\begin{aligned}x_1 + 2x_2 + 3x_3 &= 1 \\-3x_2 - 6x_3 &= -2 \\0 &= 0\end{aligned}$$

我们发现  $x_3$  是一个自由变量, 令  $x_3 = t$ , 则

$$-2 = -3x_2 - 6t \Rightarrow x_2 = -2t + \frac{2}{3}$$

$$1 = x_1 + 2(-2t + \frac{2}{3}) + 3t \Rightarrow x_1 = t - \frac{1}{3}$$

所以原方程的解为

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} t - \frac{1}{3} \\ -2t + \frac{2}{3} \\ t \end{bmatrix}$$

如果经过消元得到了类似  $0 = 1$  的式子, 这就导出了矛盾, 这时方程无解.

## 把线性方程用矩阵语言重写一遍吧

显然我们可以用矩阵语言重写线性方程:

$$\begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

使用  $R_{i_2} - kR_{i_1}$  代表将第  $R_{i_2}$  行减去  $k$  倍的第  $R_{i_1}$  行, 用  $R_{i_1 i_2}$  表示交换两行. 此外, 由于  $y$  会和系数一起变化, 上面我们做过的操作可以写成 (其实还可以回代把上面的一些消掉):

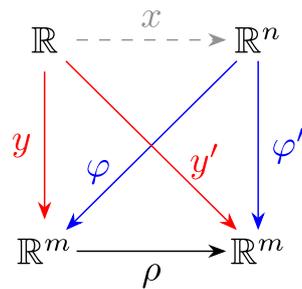
$$\left[ \begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 2 \\ 7 & 8 & 9 & 3 \end{array} \right] \xrightarrow[\begin{array}{l} R_2-4R_1 \\ R_3-7R_1 \end{array}]{R_2-4R_1} \left[ \begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ -3 & -6 & -2 & -2 \\ -6 & -12 & -4 & -4 \end{array} \right] \xrightarrow{R_3-R_2} \left[ \begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ -3 & -6 & -2 & -2 \\ & 0 & 0 & 0 \end{array} \right]$$

矩阵的空白处表示 0, 所以它们被省略掉了. 虽然我们看起来像是给矩阵加了一列然后把它们一起操作, 但是实际上这和  $3 \times 4$  的矩阵没有任何关系, 因为实际上我们并不需要一个  $\mathbb{R}^4$  到  $\mathbb{R}^3$  的映射. 不过它们的操作方式相同并非巧合, 如果用线性方程来解释, 可以说右边的  $y$  必须和左边的系数同时改变, 但是其实有另一个方法可以用线性映射解释这件事.

以第一步为例, 我们通过  $R_2 - 4R_1$  和  $R_3 - 7R_1$ , 将  $y$  变为  $y'$ , 其中

$$y' = \begin{bmatrix} y_1 \\ y_2 - 4y_1 \\ y_3 - 7y_1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -4 & 1 & \\ -7 & & 1 \end{bmatrix} y$$

所以映射  $\rho: y \mapsto y'$  是一个线性映射. 更重要的是  $\rho$  可逆. (可以通过反向操作把  $y'$  变回  $y$ ) 所以从  $\varphi(x) = y$  解出  $x$  和从  $\varphi'(x) = y'$  等价. 由于一个矩阵左乘另一个矩阵等价于左乘它的每一列, 所以  $y$  应当和系数矩阵一起变动,  $\varphi'$  比  $\varphi$  更好处理, 所以我们通过一个行变换将  $\varphi$  变为了  $\varphi'$ , 同时将  $y$  变为了  $y'$ .



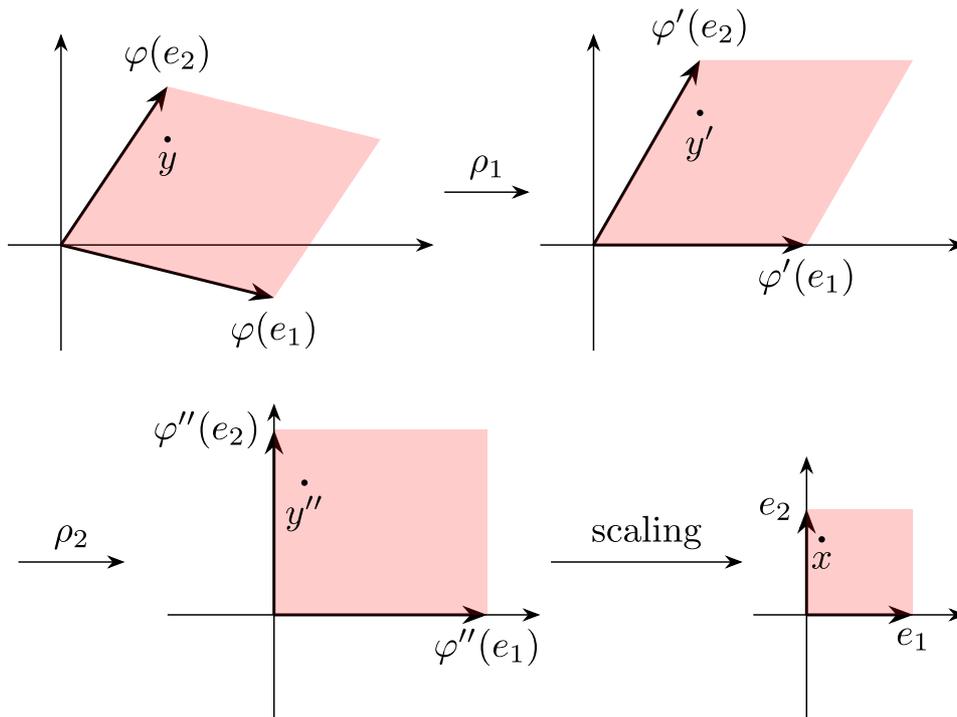
虽然前文已经用几种方式展示了解线性方程的过程, 但我还是想画个图来展现这个过程. 如果有如下的线性方程

$$\begin{bmatrix} 2 & 1 \\ -0.5 & 1.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

它可以通过同步操作  $\varphi$  和  $y$  来解出 ( $R_1 - \frac{4}{7}R_2$  被称作回代, 它把每行第一个非零元素上方的项消除):

$$\left[ \begin{array}{cc|c} 2 & 1 & 1 \\ -0.5 & 1.5 & 1 \end{array} \right] \xrightarrow{R_2+\frac{1}{4}R_1} \left[ \begin{array}{cc|c} 2 & 1 & 1 \\ 1.75 & 1 & 1 \end{array} \right] \xrightarrow{R_1-\frac{4}{7}R_2} \left[ \begin{array}{cc|c} 2 & & 2/7 \\ & 1.75 & 1.25 \end{array} \right] \xrightarrow[\begin{array}{l} R_1 \times \frac{1}{2} \\ R_2 \times \frac{4}{7} \end{array}]{R_1 \times \frac{1}{2}} \left[ \begin{array}{cc|c} 1 & & 1/7 \\ & 1 & 5/7 \end{array} \right]$$

与此同时,  $\varphi$  是这样变化的



消元之后, 称每行第一个非零元素为一个主元. (或阶梯头) 根据消元的规则可知主元下方的数字都是 0, 此时矩阵变为了一个阶梯形 (这通常被称为行阶梯形式), 下面标红的数字是消元后的主元. 如果最后一列 (也就是  $y$  所在的这一列) 出现了主元, 则左侧为 0 而右侧非零, 导出了矛盾从而导致无解. 如果主元全部在左侧 (也就是原本的矩阵所属的范围), 那么这个方程有解. 此时我们可以自由选择非主元列对应的  $x$  分量的值, 主元对应的  $x$  分量可以从其中一一解出, 所以非主元列对应的变量又称为自由变量.

$$\left[ \begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ & -3 & -6 & -2 \\ & & 0 & 0 \end{array} \right]$$

注意到解一个线性方程是从  $y = \varphi(x)$  中解出  $x$  的过程, 如果  $y \in \text{Im } \varphi$  则有解, 如果  $y \notin \text{Im } \varphi$  则无解.

(回忆像的定义) 为了检验  $y$  是否在像中, 只需要检验最后一起消元之后  $y$  所在的列有没有主元. 当有解时, 如果  $y = \varphi(x_1) = \varphi(x_2)$ , 则有  $\varphi(x_1) - \varphi(x_2) = \varphi(x_1 - x_2) = 0$ , 这意味着  $x_1 - x_2 \in \text{Ker } \varphi$ , 所以方程的解集具有结构

$$\{x \mid y = f(x)\} = \begin{cases} \emptyset, & y \notin \text{Im } \varphi \\ x^* + \text{Ker } \varphi, & y \in \text{Im } \varphi \end{cases}$$

这里  $x^*$  代表原方程某个特定的解,  $x^* + \text{Ker } \varphi$  是  $\{x^* + v \mid v \in \text{Ker } \varphi\}$  的简写

假设一个线性方程有解, 则我们可以通过令所有自由变量为 0 来解出一个特解, 通过把注意力放到  $\text{Ker } \varphi$ , 或称作  $\varphi(x) = 0$  上, 我们可以先把  $y$  放到一边, 先专注地处理  $\varphi$ , 假设  $\varphi$  化为行阶梯形式后矩阵有  $r$  个主元, 则有  $n - r$  个自由变量. 如果  $n = r$ , 则没有自由变量,  $\text{Ker } \varphi = 0$ ,  $x^*$  是唯一解, 如果  $n > r$ , 每次令一个自由变量为 1, 其它为 0, 可以在  $\text{Ker } \varphi$  中找到一系列向量  $u_1 \sim u_{n-r}$ . 根据核是一个线性空间, 这些向量的线性组合  $t_1 u_1 + \dots + t_{n-r} u_{n-r}$  也在核中, 再结合主元对应的变量可以从自由变量的值解出, 我们断言  $\varphi(x) = 0$  解在第  $j$  个自由变量取值为  $t_j$  时只能为  $t_1 u_1 + \dots + t_{n-r} u_{n-r}$ . 故  $\text{Ker } \varphi$  中的元素都可以被唯一表示为  $t_1 u_1 + \dots + t_{n-r} u_{n-r}$  的形式, 再加上特解可得  $\varphi(x) = y$  的解都具有形式  $x = x^* + t_1 u_1 + \dots + t_{n-r} u_{n-r}$ , 这被称为线性方程的通解.

$$\varphi = \begin{bmatrix} k_{11} & k_{12} & k_{13} & \cdots & k_{1n} \\ k_{21} & k_{22} & k_{23} & \cdots & k_{2n} \\ k_{31} & k_{32} & k_{33} & \cdots & k_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & k_{m3} & \cdots & k_{mn} \end{bmatrix} \xrightarrow{\rho} \varphi' = \begin{bmatrix} * & \cdots & * & \cdots & * & \cdots \\ & & * & \cdots & * & \cdots \\ & & & \ddots & \vdots & \ddots \\ & & & & * & \cdots \\ & & & & & 0 \end{bmatrix}$$

(\* 代表一个任意的数, 0 意味着下面的数都是 0. 我们没有给 \* 对应的数指定一个具体的符号, 因为我们不关心它们具体的取值, 只关心这个矩阵长什么样, 标红的表示主元, 是非 0 的数)

注意到  $r \leq m$  成立, 所以对于  $m < n$ , 有  $n - r > 0$  成立, 此时一定存在自由变量. 既然核不为 0,  $\varphi(x) = 0$  有多解, 映射不为单射. 反过来, 如果从  $\mathbb{R}^n$  到  $\mathbb{R}^m$  存在单射, 则  $m \geq n$ . 直观的感觉是从高维空间到低维空间总是要丢失信息 (虽然我们并没有定义维数)

同时, 注意到  $r \leq n$  成立, 所以对于  $n < m$ , 有  $m - r > 0$  成立. 由于前  $r$  行有主元, 剩下的行为 0, 此时可以找到某个  $y'$  使得它在第  $r$  行后面还有非零元素, 这导致它不在  $\varphi'$  的像空间中, 所以可以找到  $y = \rho^{-1}(y')$  使  $y$  不在  $\varphi$  的像空间中. 换言之, 对于  $n < m$ ,  $\text{Im } \varphi$  不为  $\mathbb{R}^m$ , 从而  $\varphi$  不为满射. 这个结果也可以反过来叙述: 如果有  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的满射, 则  $n \geq m$ . 这个结果也可以直观地看成低维空间不能完全表达高维空间.

## 逆矩阵

首先回忆什么是逆. 如果  $f$  给出了  $A, B$  之间的双射, 则它将  $x \in A$  和  $y \in B$  一一对应起来, 所以  $f(x) = y, f^{-1}(y) = x, f^{-1} \circ f = \text{Id}$  且  $f \circ f^{-1} = \text{Id}$ . 第一个是  $A$  上的恒同映射, 第二个是  $B$  上的恒同映射.

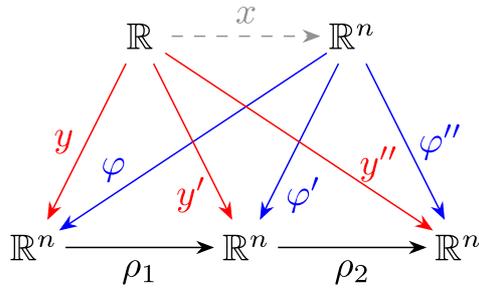
一个线性映射  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m$  可逆显然和它是双射等价, 这要求它既单又满. 所以化为行阶梯形式后  $r = n$  且  $r = m$ , 这推出  $r = m = n$ . 所以  $\varphi$  的行数和列数相同. 此处  $r = n$  排除了自由变量以让它单,  $r = m$  排除了不在像空间中的向量以让它满. 每行都是主元行, 每列都是主元列, 正向消元后得到的行阶梯形式具有形状:

$$\begin{bmatrix} * & * & \cdots & * \\ & * & \cdots & * \\ & & \ddots & \vdots \\ & & & * \end{bmatrix}$$

通过回代, 对角线上方的元素也被消除了, 只剩下一个对角矩阵. 最后将每一行都除以主元把主元变为 1 从而得到了  $n$  阶单位矩阵. 此外如果我们把  $y$  和  $\varphi$  一起操作, 这几步操作把  $\varphi(x) = y$  变为了  $I_n(x) = y'''$ , 所以  $x = y''' = \varphi^{-1}(y)$  可以直接从最终结果得出, 全过程是这样的:

$$\begin{bmatrix} * & * & \cdots & * & y_1 \\ * & * & \cdots & * & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \cdots & * & y_n \end{bmatrix} \xrightarrow{\rho_1} \begin{bmatrix} * & * & \cdots & * & y'_1 \\ & * & \cdots & * & y'_2 \\ & & \ddots & \vdots & \vdots \\ & & & * & y'_n \end{bmatrix} \xrightarrow{\rho_2} \begin{bmatrix} * & & & & y''_1 \\ & * & & & y''_2 \\ & & \ddots & & \vdots \\ & & & * & y''_n \end{bmatrix} \xrightarrow{\rho_3} \begin{bmatrix} 1 & & & & y'''_1 \\ & 1 & & & y'''_2 \\ & & \ddots & & \vdots \\ & & & 1 & y'''_n \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y'''_1 \\ y'''_2 \\ \vdots \\ y'''_n \end{bmatrix}$$

从映射的角度看来, 可以画如下的图:



$\rho_1$  是正向消除, 将  $y$  变为  $y'$ , 并将  $\varphi$  变为  $\varphi'$ , 则  $\varphi'$  是上三角的矩阵. (意思是第  $(i, i)$  个元素以上的元素都是 0)  $\rho_2$  把  $y'$  变为  $y''$ , 将  $\varphi'$  变为  $\varphi''$ , 其中  $\varphi''$  是对角矩阵, 其逆可以直接计算出来, 所以为了计算  $\varphi^{-1}$ , 可以通过  $\varphi^{-1} = (\varphi'')^{-1}\rho_2\rho_1$  得到. 虽然上面的例子中我们讨论的是通过  $\mathbb{R} \rightarrow \mathbb{R}^n$  的映射  $y$  解出映射  $x$ , 但是实际上把  $\mathbb{R}$  换为  $\mathbb{R}^p$  不影响结果, 我们仍然可以通过  $x = (\varphi'')^{-1}\rho_2\rho_1y$  解出  $x$ . 在操作中, 只需要改变拼到右边的矩阵形状就可以, 例如, 对于

$$\varphi = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}, y = \begin{bmatrix} 1 & 1 & 4 & 5 \\ 1 & 4 & 1 & 9 \\ 1 & 9 & 8 & 1 \\ 0 & 8 & 9 & 3 \end{bmatrix}$$

可以通过如下步骤解  $\varphi(x) = y$

$$\begin{aligned} & \left[ \begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 4 & 5 \\ 1 & 2 & 2 & 2 & 1 & 4 & 1 & 9 \\ 1 & 2 & 3 & 3 & 1 & 9 & 8 & 1 \\ 1 & 2 & 3 & 4 & 0 & 8 & 9 & 3 \end{array} \right] \xrightarrow{\substack{R_2-R_1 \\ R_3-R_1 \\ R_4-R_1}} \left[ \begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 4 & 5 \\ 1 & 1 & 1 & 1 & 0 & 3 & -3 & 4 \\ 1 & 1 & 2 & 2 & 0 & 8 & 4 & -4 \\ 1 & 1 & 2 & 3 & -1 & 7 & 5 & -2 \end{array} \right] \xrightarrow{\substack{R_3-R_2 \\ R_4-R_2}} \\ & \left[ \begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 4 & 5 \\ 1 & 1 & 1 & 1 & 0 & 3 & -3 & 4 \\ 1 & 1 & 2 & 2 & 0 & 5 & 7 & -8 \\ 1 & 1 & 2 & 3 & -1 & 4 & 8 & -6 \end{array} \right] \xrightarrow{R_4-R_3} \left[ \begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 4 & 5 \\ 1 & 1 & 1 & 1 & 0 & 3 & -3 & 4 \\ 1 & 1 & 2 & 2 & 0 & 5 & 7 & -8 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 2 \end{array} \right] \xrightarrow{R_1-R_2} \\ & \left[ \begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 4 & 5 \\ 1 & 1 & 1 & 1 & 0 & 3 & -3 & 4 \\ 1 & 1 & 2 & 2 & 0 & 5 & 7 & -8 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 2 \end{array} \right] \xrightarrow{R_2-R_3} \left[ \begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 4 & 5 \\ 1 & 1 & 1 & 1 & 0 & 3 & -3 & 4 \\ 1 & 1 & 2 & 2 & 0 & 5 & 7 & -8 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 2 \end{array} \right] \xrightarrow{R_3-R_4} \\ & \left[ \begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 1 & 4 & 5 \\ 1 & 1 & 1 & 1 & 0 & 3 & -3 & 4 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 2 \\ 1 & 1 & 2 & 2 & 0 & 5 & 7 & -8 \end{array} \right] \Rightarrow x = \begin{bmatrix} 1 & -2 & 7 & 1 \\ 0 & -2 & -10 & 12 \\ 1 & 6 & 6 & -10 \\ -1 & -1 & 1 & 2 \end{bmatrix} \end{aligned}$$

(本例中, 所有的主元都为 1, 所以  $\varphi''$  就是  $I_n$ , 最后的缩放可以省去)

为了找到一个矩阵的逆矩阵, 只需要让  $y = I_n$ , 即可解出  $\varphi^{-1} = \varphi^{-1}I_n$

## LU分解

在前一节, 我们对可逆矩阵找到了一个方法来得到  $\varphi'' = \rho_2\rho_1\varphi$ . 但是实际上前面的向下消元和回代并不依赖于  $\varphi$  可逆, 这个限制只是为了在前一节中让  $\varphi''$  是一个仅对角线不为 0, 其它全为 0 的方阵. 对于通常的  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 它可以被消元到行阶梯形式:

$$\varphi' = \begin{bmatrix} * & \cdots & * & \cdots & * & \cdots \\ & & * & \cdots & * & \cdots \\ & & & \ddots & \vdots & \ddots \\ & & & & * & \cdots \\ & & & & & 0 \end{bmatrix}$$

如果没有经过行交换且有  $r$  个主元, 通常而言进行了  $r$  次消元 (每个主元一次), 所以  $\rho_1$  可以被写为  $\rho_1 = \rho_{1r} \cdots \rho_{12} \rho_{11}$ , 这里  $\rho_{1i}$  表示对第  $i$  个主元下方进行的消元. 举  $i = 1$  为例子, 此时对于每个  $i' > 1$ , 第  $i'$  行都减去了  $c_{i'1}$  倍的第一行 (也就是  $y$  的第 1 个分量影响了变换后  $y$  的第  $i'$  个分量), 所以第  $(i', 1)$  个  $\rho_1$  的分量应该是  $-c_{i'1}$ , 对角线元素应为 1, 而其它元素都为 0. 同样地, 对  $1 \leq i \leq r$ , 只有第  $i$  个对角元素下方的元素会和单位矩阵有差别, 则

$$\rho_{11} = \begin{bmatrix} 1 & & & & \\ -c_{21} & 1 & & & \\ \vdots & & \ddots & & \\ -c_{m1} & & & & 1 \end{bmatrix}$$

为了算  $\rho_1$ , 我们需要先算  $\rho_{12} \rho_{11}$ , 然后算  $\rho_{13} \rho_{12} \rho_{11}$ , ....., 最后算出  $\rho_1 = \rho_{1r} \cdots \rho_{12} \rho_{11}$ , 但是, 问题来了

$$\rho_{12} \rho_{11} = \begin{bmatrix} 1 & & & & \\ -c_{21} & 1 & & & \\ -c_{31} + c_{32}c_{21} & -c_{32} & 1 & & \\ \vdots & & & \ddots & \\ -c_{m1} + c_{m2}c_{21} & -c_{m2} & & & \end{bmatrix}$$

而且当我们越往下算, 底下多项式的项越来越多, 呈现指数级增长, 到最后第  $m$  行第 1 项会变成

$$-c_{m1} + \sum_{1 < t < m} c_{mt} c_{t2} - \sum_{1 < t_1 < t_2 < m} c_{mt_1} c_{t_1 t_2} c_{t_2 m} + \sum_{1 < t_1 < t_2 < t_3 < m} c_{mt_1} c_{t_1 t_2} c_{t_2 t_3} c_{t_3 m} - \cdots$$

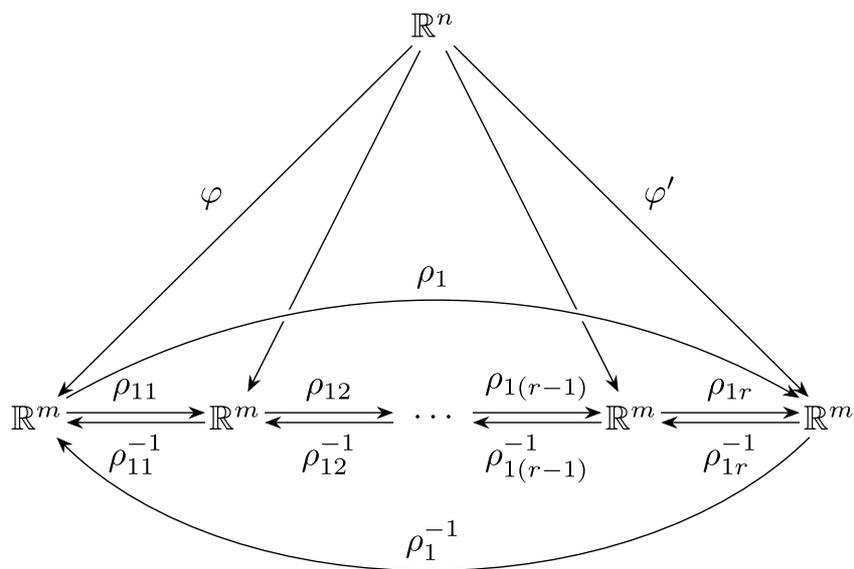
再这么算下去对研究问题也不会有什么用 (这甚至会让我们感觉为什么为什么想不开要合并它们, 分开来写多好)

但是事实上计算  $\rho_1^{-1}$  比计算  $\rho_1$  容易, 这点我等会说明. 为了计算  $\rho_1^{-1}$ , 首先注意到

$$\begin{aligned} & \rho_{11}^{-1} \rho_{12}^{-1} \cdots \rho_{1(r-1)}^{-1} \rho_{1r}^{-1} \rho_{1r} \rho_{1(r-1)} \cdots \rho_{12}^{-1} \rho_{11}^{-1} \\ &= \rho_{11}^{-1} \rho_{12}^{-1} \cdots \rho_{1(r-1)}^{-1} I_m \rho_{1(r-1)} \cdots \rho_{12}^{-1} \rho_{11}^{-1} \\ &= \rho_{11}^{-1} \rho_{12}^{-1} \cdots \rho_{1(r-1)}^{-1} \rho_{1(r-1)} \cdots \rho_{12}^{-1} \rho_{11}^{-1} \\ &= \rho_{11}^{-1} \rho_{12}^{-1} \cdots \rho_{12}^{-1} \rho_{11}^{-1} \\ &= \cdots \\ &= \rho_{11}^{-1} \rho_{12}^{-1} \rho_{12} \rho_{11} \\ &= \rho_{11}^{-1} \rho_{11} \\ &= I_m \end{aligned}$$

所以  $\rho_1$  的逆为  $\rho_{11}^{-1} \rho_{12}^{-1} \cdots \rho_{1r}^{-1}$ ,  $\varphi = \rho_1^{-1} \varphi'$

对于一系列可逆矩阵的乘积, 要小心计算逆矩阵时是反过来乘的, 也就是说对可逆的  $\psi_1$  和  $\psi_2$ ,  $(\psi_1 \psi_2)^{-1} = \psi_2^{-1} \psi_1^{-1}$  是对的, 但是  $\psi_1^{-1} \psi_2^{-1}$  在大部分情况下是错的, 通过跟着下图的箭头走, 你会发现这个结果是合理的.



接下来, 来计算  $\rho_{1i}^{-1}$ , 例如对  $i = 1$ , 我们有 (读者应该自己验证) :

$$\rho_{11} = \begin{bmatrix} 1 & & & & \\ -c_{21} & 1 & & & \\ \vdots & & \ddots & & \\ -c_{m1} & & & & 1 \end{bmatrix}, \rho_{11}^{-1} = \begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & & \ddots & & \\ c_{m1} & & & & 1 \end{bmatrix}$$

类似地,  $\rho_{1i}^{-1}$  和  $\rho_{1i}$  仅在第  $i$  行第  $j$  列的元素下方的元素有差异.

神奇的是, 如果把  $\rho_{11}^{-1} \rho_{12}^{-1} \cdots \rho_{1r}^{-1}$  乘起来, 我们得到了一个非常简单的结果 (读者也应该自己验证) :

$$\rho_1^{-1} = \rho_{11}^{-1} \rho_{12}^{-1} \cdots \rho_{1r}^{-1} = \begin{bmatrix} 1 & & & & \\ c_{21} & 1 & & & \\ \vdots & \vdots & \ddots & & \\ c_{m1} & c_{m2} & \cdots & & 1 \end{bmatrix}$$

这里对角线下方的所有元素都是  $c_{i'i}$ , 而对角线上方的元素都是 0, 且对角线上都为 1. 假设没有行交换, 左边是严格的下三角矩阵并且忠实地表示了对矩阵进行行变换的过程, 而右边的  $\varphi'$  是严格上三角的, 用  $L = \rho_1^{-1}$  表示一个下(lower)三角矩阵, 用  $U = \varphi'$  表示一个上(upper)三角矩阵, 则  $\varphi = LU$  称为矩阵  $\varphi$  的 LU 分解. 为了更清楚地看懂什么是 LU 分解, 看看下面这个例子 (它在前面出现过) 这次我们把  $\varphi_1^{-1}, \varphi_1^{-1} \varphi_2^{-2}, \dots$  也写在矩阵左边, 每次把操作分离出来的变换矩阵乘到左边, 这样我们的操作变成了等式.

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{array}{l} R_2 - 1R_1 \\ R_3 - 1R_1 \\ R_4 - 1R_1 \end{array} \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & & 1 & \\ 1 & & & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & 1 & 2 & 2 \\ & 1 & 2 & 3 \end{bmatrix} \begin{array}{l} R_3 - 1R_2 \\ R_4 - 1R_2 \end{array} \\ \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 1 & 1 & \\ 1 & 1 & & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix} \begin{array}{l} R_4 - 1R_3 \end{array} \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & & 1 & 1 \\ & & & 1 \end{bmatrix}$$

原本的矩阵变为了一个下三角矩阵和一个上三角矩阵的乘积. 这里有 4 个主元但是我们只进行了 3 次行变换, 这并不奇怪, 因为最后一个主元已经在最后一行了, 下方没有元素, 所以不需要用它消元即  $\rho_{14} = I_4$ , 这一步被省略了.

这时有人又要问了, 那万一真要交换行呢? LU分解还能行吗?

答案是否定的. 之前我们通过假设没有行交换来保证消除对应的矩阵是下三角矩阵, 它们的复合也是下三角的. (为什么?) 但是交换矩阵并不是下三角的. 假设有一个矩阵  $\sigma_{23}$ , 它作用在  $\mathbb{R}^3$  上, 交换第二和第三个分量, 所以

$$\sigma_{23} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_3 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

显然它不是下三角矩阵, 所以当我们要用行交换时将会发生:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix} \xrightarrow[\underline{R_3 - R_1}]{\underline{R_2 - R_1}} \rho_{11}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix} \xrightarrow{\underline{R_{23}}} \rho_{11}^{-1} \sigma_{23} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}$$

右边确实是上三角的, 但是左边出问题

$$\rho_{11}^{-1} = \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & & 1 \end{bmatrix}, \sigma_{23} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, \rho_{11}^{-1} \sigma_{23} = \begin{bmatrix} 1 & & \\ 1 & & 1 \\ 1 & 1 & \end{bmatrix}$$

显然  $\rho_{11}^{-1} \sigma_{23}$  不是下三角的, 这表明LU分解失效了

所以此时的解决办法是PLU分解, 它将一个矩阵分解为一个置换矩阵、一个下三角矩阵和一个上三角矩阵.  $\mathbb{R}^m$  上的一个置换矩阵是一个方阵, 其每行每列只有一个元素是 1, 其它元素都是 0. 容易验证这样一个矩阵作用在  $\mathbb{R}$  上只改变各分量的排列顺序, 当作用在  $\rho_{11}$  的右端时它作用在输入端, 所以造成的影响是交换第二和第三列. 为了让它保持下三角的形状, 同时交换  $L$  矩阵的第 2, 3 行, 也就是  $\rho_{11}^{-1} \sigma_{23} = \sigma_{23} \sigma_{23} \rho_{11}^{-1} \sigma_{23}$ .

(注意  $\sigma_{23} \sigma_{23} = I_m$ ) 在这之后  $\sigma_{23} \rho_{11}^{-1} \sigma_{23}$  变回了下三角, 而  $\sigma_{23}$  跑到了左边, 仍然用一个例子来演示:

$$\begin{aligned} & \begin{bmatrix} 1 & & & \\ 1 & 1 & 1 & \\ 2 & 1 & 1 & 1 \\ & 2 & 3 & 3 \end{bmatrix} \xrightarrow[\underline{R_3 - 2R_1}]{\underline{R_2 - R_1}} \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 2 & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & 1 & 1 \\ & 2 & 3 & 3 \end{bmatrix} \xrightarrow{\underline{R_{23}}} \begin{bmatrix} 1 & & & \\ 1 & & 1 & \\ 2 & 1 & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & 1 & 1 \\ & 2 & 3 & 3 \end{bmatrix} = \\ & \begin{bmatrix} 1 & & & \\ & 1 & & \\ 1 & & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 1 & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & 1 & 1 \\ & 2 & 3 & 3 \end{bmatrix} \xrightarrow{\underline{R_4 - 2R_2}} \begin{bmatrix} 1 & & & \\ & 1 & & \\ 1 & & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 1 & & 1 & \\ & 2 & 1 & \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & 1 & 1 \\ & 2 & 3 & 3 \end{bmatrix} \\ & \xrightarrow{\underline{R_{34}}} \begin{bmatrix} 1 & & & \\ & 1 & & \\ 1 & & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 1 & & 1 & \\ & 2 & 1 & \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ 1 & & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 1 & & 1 & \\ & 2 & 1 & \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & 1 & 1 \\ & 1 & 1 & 1 \\ & & & 1 \end{bmatrix} \end{aligned}$$

第一步中, 我们正常地用第一行对下方消元( $\rho_{11}$ )

第二步中, 行交换打破了下三角的形状( $\sigma_{23}$ )

第三步中, 对下三角矩阵的行交换让它恢复了形状

第四步中, 我们正常地用第二行对第四行消元( $\rho_{12}$ )

第五步中, 行交换再一次打破了下三角的形状( $\sigma_{34}$ )

第六步中, 通过对下三角矩阵的另一个行交换, 我们得到了PLU分解结果

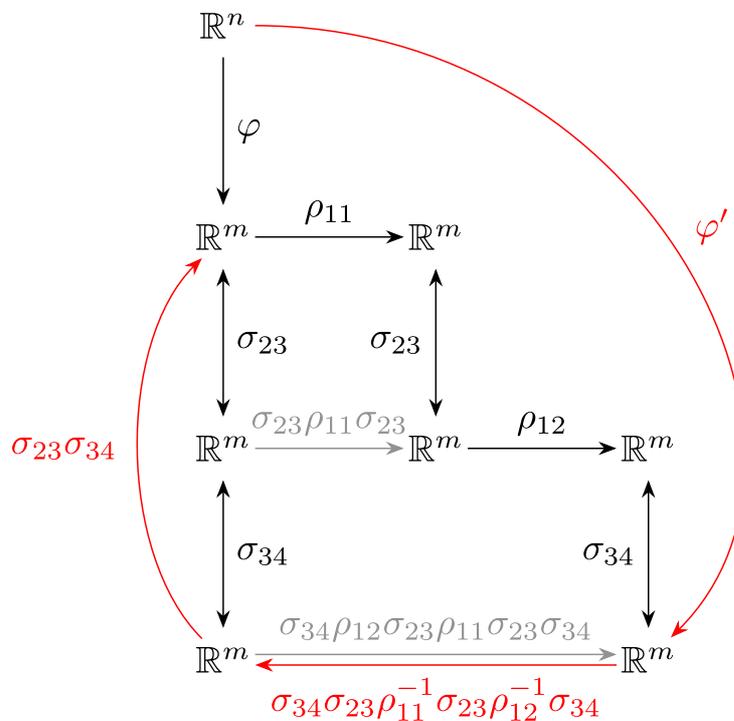
令

$$\rho_{11}^{-1} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 2 & & 1 & \\ & & & 1 \end{bmatrix}, \rho_{12}^{-1} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & 2 & 1 \end{bmatrix}$$

则这个过程可以写成

$$\begin{aligned} \varphi &= \rho_{11}^{-1} \varphi_1 = (\rho_{11}^{-1} \sigma_{23}) \varphi_2 = \sigma_{23} (\sigma_{23} \rho_{11}^{-1} \sigma_{23}) \varphi_2 = \sigma_{23} (\sigma_{23} \rho_{11}^{-1} \sigma_{23} \rho_{12}^{-1}) \varphi_3 \\ &= \sigma_{23} (\sigma_{23} \rho_{11}^{-1} \sigma_{23} \rho_{12}^{-1} \sigma_{34}) \varphi_4 = (\sigma_{23} \sigma_{34}) (\sigma_{24} \sigma_{23} \rho_{11}^{-1} \sigma_{23} \rho_{12}^{-1} \sigma_{34}) \varphi' \end{aligned}$$

又或者有的人希望看图:



通过同时交换行和列,  $\rho_{11}$ ,  $\sigma_{23} \rho_{11} \sigma_{23}$ ,  $\rho_{12} \sigma_{23} \rho_{11} \sigma_{23}$ ,  $\sigma_{34} \rho_{12} \sigma_{23} \rho_{11} \sigma_{23} \sigma_{34}$  和它们的逆都是下三角的,  $\sigma_{ii}$  被扔到了左边,  $\varphi$  变成了  $\varphi'$ , 所以最终有  $P = \sigma_{23} \sigma_{34}$ ,  $L = \sigma_{24} \sigma_{23} \rho_{11}^{-1} \sigma_{23} \rho_{12}^{-1} \sigma_{34}$  和  $U = \varphi'$ , 所以LU分解本身有一定缺陷但是PLU分解就没这个问题. 实际上, 所有的数值计算程序都会有更多关于行调整的规则以保证算法的稳定性, 如果你感兴趣可以自己在网上搜. 别的地方和这里记号可能不同, 但是干的事情是一样的.

## 练习

2.1 设线性映射  $\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $\varphi(1, -1) = 0$  且  $\varphi(1, 1) = 2$ , 找出  $\varphi$  并用矩阵表示

2.2 一个映射  $\varphi: \mathbb{R}^3 \rightarrow \mathbb{R}^2$  将  $(x_1, x_2, x_3)$  投影到  $(x_1, x_2)$ , 用矩阵表示它.

2.3 如果在  $\mathbb{R}^2$  上定义内积  $\langle x, y \rangle = x_1 y_1 + x_2 y_2$ , 则  $\langle x, - \rangle$  给出了从  $\mathbb{R}^2$  到  $\mathbb{R}$  的线性映射, 将  $y$  映射到  $\langle x, y \rangle$ . 用矩阵表示这个映射.

2.4  $\mathbb{R}^2$  上的映射  $\varphi$  将每个向量绕着原点旋转  $45^\circ$ , 用矩阵表示它并计算  $\varphi(x)$  的结果, 其中  $x = (3, 1)$ . 通过两种方式计算(行和列的视角)并比较结果.

2.5 对于  $\mathbb{R}^2$  上的线性变换, 设  $\varphi = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$ ,  $\varphi(e_1)$  和  $\varphi(e_2)$  何时共线? 当二者不共线时, 单位正方形的像是一个平行四边形, 并叙述面积表达式和共线条件有何关系.

2.6 设  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 则  $\varphi^2 = \varphi\varphi$  何时有定义? 如果  $\varphi = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ , 计算  $\varphi^2, \varphi^3$  并猜测  $\varphi^n$  的结果, 请使用归纳法证明这一结果.

2.7 设  $\varphi = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ , 计算  $\varphi^2, \varphi^3, \varphi^4, \varphi^5$ , 它们表现出什么规律?

2.8 通常情况下, 对函数  $f$  和  $g$ ,  $f \circ g = g \circ f$  几乎不能成立, 例如, 令  $f(x) = x^2, g(x) = e^x$ , 则  $(f \circ g)(x) = e^{2x}, (g \circ f)(x) = e^{x^2}, f \circ g \neq g \circ f$ . 对于线性映射也是一样, 初学者很容易误以为  $\varphi\psi = \psi\varphi$  成立. 如果  $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m, \psi: \mathbb{R}^m \rightarrow \mathbb{R}^n$  是两个映射, 什么时候它们能同时满足如下三个条件:

(1)  $\varphi\psi$  有定义; (2)  $\psi\varphi$  有定义 (3)  $\varphi\psi$  和  $\psi\varphi$  之间的比较有意义 (这要求两个矩阵大小相同)

2.9 对于多项式  $p(x) = a_0 + \sum_{i=1}^n a_i x^i$  和  $V$  上的线性变换  $\varphi$ , 定义  $p(\varphi) = a_0 \text{Id} + \sum_{i=1}^n a_i \varphi^i$  (与多项式相比, 1 被替换为了  $\text{Id}$ , 这是因为恒同映射的作用类似 1) 证明  $p(\varphi)$  和  $\varphi$  相乘可以交换顺序(即  $\varphi \circ p(\varphi) = p(\varphi) \circ \varphi$ )

2.10 设  $u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, v = [4 \ 5 \ 6]$ , 计算  $uv, (uv)^2$  和  $(uv)^n$ .

2.11 三个人分别有  $x_{01}, x_{02}$  和  $x_{03}$  元, 每一次每个人都把自己的钱分为两半, 给另外两个人一人一半. 如果  $x_n = (x_{n1}, x_{n2}, x_{n3})$  为他们经过  $n$  次之后每个人的钱数, 写出从  $x_n$  到  $x_{n+1}$  的矩阵  $\varphi$  并且解释  $\varphi^n$  的含义

2.12 通过两种方式解下面的线性方程: 使用和不使用矩阵, 并且比较结果.

$$\begin{aligned} 2x - y &= 1 \\ 2y - z &= 2 \\ 2z - x &= 3 \end{aligned}$$

2.13 如果  $\varphi = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ , 求  $\varphi^{-1}$  并解方程  $\varphi(x) = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$ .

2.14 如果  $\varphi = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix}$ , 找到若干  $y$  使  $\varphi(x) = y$  无解.

2.15 当  $a$  为何值时下面的矩阵不可逆? 可逆和不可逆时分别解  $\varphi(x) = 0$ , 主元个数  $r$  分别为多少?

$$\varphi = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 1 & 1 & a \end{bmatrix}$$

2.16 假设  $\varphi, \psi: \mathbb{R}^n \rightarrow \mathbb{R}^n$  的矩阵表示都是下三角矩阵, 证明  $\varphi\psi$  也是下三角矩阵.

2.17 设  $\sigma_{ii'}: \mathbb{R}^m \rightarrow \mathbb{R}^m$  交换  $\mathbb{R}^m$  中的第  $i, i'$  个分量, 通过矩阵运算的公式证明:  $\sigma_{ii'}^2 = I_m$

2.18 证明  $\psi_1$  和  $\psi_2$  都是置换矩阵, 通过计算证明  $\psi_1\psi_2$  也是置换矩阵 (提示: 分别计算对每个  $e_j$  的结果)

2.19 设  $\rho_{11} = \begin{bmatrix} 1 & & & \\ -c_{21} & 1 & & \\ \vdots & & \ddots & \\ -c_{m1} & & & 1 \end{bmatrix}$ , 证明  $\rho_{11}^{-1} = 2I_m - \rho_{11}$ .

2.20 证明公式:

$$\begin{bmatrix} 1 & & & \\ c_{21} & 1 & & \\ \vdots & & \ddots & \\ c_{(m-1)1} & & & 1 \\ c_{m1} & & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & \vdots & \ddots & \\ & c_{(m-1)2} & & 1 \\ & c_{m2} & & & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & & c_{m(m-1)} & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & & & & \\ c_{21} & 1 & & & & & \\ \vdots & \vdots & \ddots & & & & \\ c_{(m-1)1} & c_{(m-1)2} & \cdots & & 1 & & \\ c_{m1} & c_{m2} & \cdots & c_{m(m-1)} & & & 1 \end{bmatrix}$$

2.21 对于交换矩阵  $\sigma_{ii'}$ , 说明为什么它乘在另一个矩阵右侧时表现为列交换, 乘在另一个矩阵左侧时表现为行交换. 如果有另一个下三角矩阵  $L$ ,  $\sigma_{ii'}L\sigma_{ii'}$  并不总是下三角的, 以  $\mathbb{R}^4$  上的  $\sigma_{23}$  和

$L = \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 3 & 4 & 1 & \\ 5 & 6 & 7 & 1 \end{bmatrix}$  为例, 计算  $\sigma_{23}L\sigma_{23}$  的结果, 说明为什么它不再是下三角矩阵并解释为什么在引入

PLU分解时这样的情况没有发生

2.22 对如下矩阵做PLU分解并求出它的核.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 9 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

## 第3章 从 $\mathbb{R}^n$ 到 $V$ —— 基和表示理论

### 维数这东西真的那么显然吗?

接下来我们将会正式地引入维度, 但在此之前, 我们要先证明一些相关结果. 本章中我们仍然在  $\mathbb{R}^n$  上工作, 如无特别解释, 以下的映射和空间都是线性的.

思考第一个问题: 维数是什么? 一个点的维数是 0, 直线的维数是 1, 平面维数是 2, 而我们所在的空间维数是 3, 维数描述了空间中有多少个"不同的方向" (这里左右视为一个方向, 上下视为一个方向). 所以我们自然期待  $\mathbb{R}^n$  的维数是  $n$ , 但是对一个一般的空间, 怎么定义维数呢? 先看几个例子:

一条直线  $\{(t, t) | t \in \mathbb{R}\}$  是  $\mathbb{R}^2$  的子空间, 我们不能因为每个向量有两个分量就说它是二维的, 所以显然我们不应该谈论一个向量的维数, 而是一个空间整体的维数. 我们希望这条线的维数是 1, 因为它只是平面上的

一条线. 所有  $a \sin x + b \cos x$  型的函数形成了一个线性空间, 这时我们不能直接把它看成一个空间中的平面, 因为它甚至不是任何一个  $\mathbb{R}^n$  的子空间. 但我们仍然认为它是二维的 (如果你没有这个直觉就暂时这么接受下来吧), 这是因为有两个参数决定这个函数.

所有  $\mathbb{N}_+ \rightarrow \mathbb{R}$  的映射 (即实数列) 形成一个线性空间, 看起来这里有无穷多个“参数”:  $a_1, a_2, \dots, a_n$  都是互相独立的, 而且它看起来“包含”了  $\mathbb{R}^n$  (通过把  $\mathbb{R}^n$  中的元素看作第  $n$  项之后全部为 0 的有限数列).

所有满足  $a_{n+1} = a_n + a_{n+1}$  的数列  $a: \mathbb{N}_+ \rightarrow \mathbb{R}$  构成一个线性空间, 这时  $a_{n+2}$  完全由  $a_n$  和  $a_{n+1}$  决定, 所以整个数列完全决定于它的前两项, 实际上可以自由变化的量只有两个, 这时我们定义它的维数是 2.

当我们说  $n$  个参数决定 (或称之为编码) 了空间中的一个向量时, 我们可以用更加数学化的语言说这个空间可以用  $\mathbb{R}^n$  中的向量编码, 也就是说每个向量在  $\mathbb{R}^n$  中有唯一的表示. 更准确地说,  $V$  与  $\mathbb{R}^n$  之间存在一个双射. 把维数转到与之存在双射的  $\mathbb{R}^n$  完美地反映了检查需要多少个参数的想法, 同时表明这个空间和  $\mathbb{R}^n$  很像, 在某种意义上我们可以像处理  $\mathbb{R}^n$  一样处理它.

## 准备工作和基础性质

我有必要强调下标可能用来表示不同的映射, 也有可能表示一个映射的不同分量, 这取决于上下文. 虽然我一直尝试着避免这种情况, 但是我为了少从字母表中引入字母, 我放弃了在整个文章中严格用不同的符号表示不同的概念. 在上一章中, 我一直用  $j$  作为  $n$  对应的求和指标, 但是我受够了, 这一章开始我会用  $i$ , 对于我来讲, 这写起来简单多了, 但是我希望不会给读者造成太多困惑.

上一章的想法很好, 但是为了确保维数是“良定义”的, 我们需要解决一些问题.

问题一:  $V$  有没有可能既和  $\mathbb{R}^m$  又和  $\mathbb{R}^n$  同构, 这让它的维数既是  $m$  又是  $n$ , 从而失败去让它良好定义呢.

事实上这不可能, 假设  $V$  和  $\mathbb{R}^n$  同构, 那  $v \in V$  就和  $x \in \mathbb{R}^n$  一一对应. 同时它又和  $y \in \mathbb{R}^m$  一一对应, 所以  $x \in \mathbb{R}^n$  就和  $y \in \mathbb{R}^m$  一一对应, 即给出  $\mathbb{R}^m$  和  $\mathbb{R}^n$  的双射, 但是这对  $m \neq n$  并不可能 (此结果在上一章中已经证明) 所以维度至少不至于是糟糕定义的 (指一个空间不会有两个维度).

问题二: 一定存在与  $V$  同构的  $\mathbb{R}^n$  吗? 有没有可能  $V$  不与任何一个  $\mathbb{R}^n$  同构呢?

事实上这是有可能的, 例如实数列  $a: \mathbb{N}_+ \rightarrow \mathbb{R}$  组成的空间就是, 因为它不和任何一个  $\mathbb{R}^n$  同构 ( $n$  有限), 这时我们称它是无穷维的.

所以当实向量空间与  $\mathbb{R}^n$  同构时, 称它的维数为  $n$ , 记作  $\dim_{\mathbb{R}} V = n$  (没有歧义时下标  $\mathbb{R}$  被省去, 记作  $\dim V = n$ ), 特别地, 称  $\dim 0 = 0$  (这里的 0 是只包含零向量的平凡空间), 否则称其维数为无穷, 即  $\dim V = \infty$ , 接下来回忆我们上一章的结果:

(1) 如果  $\mathbb{R}^n$  到  $\mathbb{R}^m$  有单射, 则  $m \geq n$

(2) 如果  $\mathbb{R}^n$  到  $\mathbb{R}^m$  有满射, 则  $m \leq n$

这一次我们可以更加正式地用维数来叙述这一定理对有限维空间的情况:

对  $\dim V_1 = n, \dim V_2 = m$ , 有:

(3) 若  $m < n$ , 则  $V_1$  到  $V_2$  映射不单

(4) 若  $m > n$ , 则  $V_1$  到  $V_2$  映射不满

(3)(4) 是结论 (1)(2) 通过向量空间与  $\mathbb{R}^n$  和  $\mathbb{R}^m$  的同构自然导出的.

为了加入无穷维的情况, 我们还有必要证明两个结果来表明无穷维空间确实比有限维空间更“大”

(5) 如果  $\dim V_1 = \infty, \dim V_2 = m \in \mathbb{N}$ , 则  $V_1 \rightarrow V_2$  不单

(6) 如果  $\dim V_1 = n \in \mathbb{N}, \dim V_2 = \infty$ , 则  $V_1 \rightarrow V_2$  不满

这个结论将在下一节证明

## 向量组, 基底和坐标

我们知道如果  $\alpha_1$  给出了  $\mathbb{R}^n \rightarrow V_1$  的一个同构,  $\alpha_2$  给出了  $\mathbb{R}^m \rightarrow V_2$  的同构, 那么任何一个  $\varphi: V_1 \rightarrow V_2$  都可以用一个  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  的映射, 即  $m \times n$  矩阵表示, 这个映射是由  $A = \alpha_2^{-1}\varphi\alpha_1$  诱导的, 称为  $\varphi$  在基  $\alpha_1, \alpha_2$  下的表示 (基的定义在后面)

$$\begin{array}{ccc} V_1 & \xrightarrow{\varphi} & V_2 \\ \alpha_1 \uparrow & & \uparrow \alpha_2 \\ \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \end{array}$$

对于一个同构  $\alpha: \mathbb{R}^n \rightarrow V$ , 显然它的值完全由它在  $e_i$  处的值决定, 设  $\alpha(e_i) = \alpha_i$ , 则

$$\alpha(x) = \alpha\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i \alpha(e_i) = \sum_{i=1}^n x_i \alpha_i, \text{ 从列的角度来看, 这个映射可以被写成:}$$

$$\alpha = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]$$

注意到在此我们拓展了矩阵的记号. 狭义地讲, 矩阵每行每列都是数, 但是这里每一列都是一个向量, 而不一定是由数字组成的"列向量".

例如, 在函数空间  $\{a \sin x + b \cos x \mid a, b \in \mathbb{R}\}$  中, 我们可以合理地写出:

$$[\sin x \quad \cos x] \begin{bmatrix} a \\ b \end{bmatrix} = a \sin x + b \cos x$$

这里的  $\sin x$  和  $\cos x$  是空间中的向量, 上述的运算仍然遵守矩阵乘法的法则.

既然我们把(广义)矩阵的概念拓展到了研究一组向量, 我们称一个(有序)向量组是一个可以用  $1 \sim n$  编号的向量列表. 有些教材认为向量组是无序的, 但是为了方便, 我们不妨让它是有顺序的, 这对有限维空间够用了, 于是这个向量组就可以正式地写为:

$$\alpha = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]$$

注意到当  $V$  就是  $\mathbb{R}^m$  时,  $\alpha_i$  刚好就是列向量, 所以这成为了我们熟悉的矩阵, 所以我们把(广义)的矩阵概念拓展到了任意空间  $V$ . 显然一个向量组给出了一个  $\mathbb{R}^n$  到  $V$  的映射, 把每个  $x \in \mathbb{R}^n$  映射到  $\sum_{i=1}^n x_i \alpha_i$ , 我

们称之为  $\alpha$  的一个线性组合, 这个运算过程与矩阵是已知的, 按照我们的习惯, 我们要研究它的核和像以了解这个映射, "由  $\alpha$  张成的空间"指  $\alpha$  的所有线性组合, 即  $\alpha$  作为映射的像.

首先看到核.

- 如果核非 0 (即  $\alpha$  不单), 则根据定义存在一个非零的  $x$  满足  $\sum_{i=1}^n x_i \alpha_i = 0$ , 即某个系数非 0 的线性

组合结果为 0. 此时称  $\alpha$  中的向量线性相关, 不妨设其中  $x_1 \neq 0$ , 则  $x_1 \alpha_1 + \sum_{i=2}^n x_i \alpha_i = 0$ , 从而

$$\alpha_1 = -\frac{1}{x_1} \sum_{i=2}^n x_i \alpha_i = \sum_{i=2}^n \left(-\frac{x_i}{x_1}\right) \alpha_i, \text{ 即某个向量可以表示为其它向量的线性组合}$$

- 相关意味着有向量可以被其它向量的组合表示, 只要我们一个个拿走线性相关相关的向量, 最终它总会变成一个不相关的向量组  $\alpha'$ , 所有  $\alpha$  中的向量都可以由  $\alpha'$  的线性组合表示, 从而  $\text{Im } \alpha' = \text{Im } \alpha$  (不改变像) 且  $\text{Ker } \alpha' = 0$  (没有多余的向量), 如果  $\dim V = m$  而  $\alpha$  是一个  $V$  中的线性无关向量组, 其中  $\alpha$  中共  $n$  个向量, 则可以得到  $n \leq m$ , 这是因为  $\alpha$  给出了一个  $\mathbb{R}^n \rightarrow V$  的单射 (线性无关意味着单射). 换言之, 我们不可能在  $m$  维空间中找出多于  $m$  个线性无关的向量
- 对于一个  $m \times n$  的矩阵  $A$ , 如果高斯消元将它消元到行阶梯型, 且有  $r$  个阶梯头 (对应非自由变量), 则有  $n - r$  个自由变量, 解集可以用自由变量构成的  $\mathbb{R}^{n-r}$  编码, 故  $\dim \text{Ker } A = n - r$

下面讨论像.

- 如果  $\alpha : \mathbb{R}^n \rightarrow V$  是单射, 则每个  $v \in \text{Im } \alpha$  都可以被唯一地表示为  $\alpha(x)$ , 其中  $x \in \mathbb{R}^n$ . 所以  $\alpha$  给出了  $\mathbb{R}^n$  和  $V$  之间的一一对应, 其逆映射满足  $\alpha^{-1}(v) = x$ , 这表明  $\text{Im } \alpha$  和  $\mathbb{R}^n$  同构, 即  $\dim \text{Im } \alpha = n$ .
- 如果  $\alpha : \mathbb{R}^n \rightarrow V$  是满射, 我们可以通过去除线性相关的向量将它约减到  $\alpha' : \mathbb{R}^r \rightarrow V$ , 其中  $\alpha'$  包含  $r$  个线性无关的向量. 这时  $\alpha'$  与  $\alpha$  值域相同, 所以满, 由线性无关, 所以单. 所以  $V$  是有限维的且  $\dim V = r \leq n$ . 特别地, 如果  $\alpha : \mathbb{R}^n \rightarrow V$  是双射,  $\dim \alpha = n$ , 称  $\alpha$  是  $V$  的一个基, 如前文所说, 每个  $v \in V$  都可以唯一地用  $x = \alpha^{-1}(v) \in \mathbb{R}^n$  表示, 使得  $v$  是  $x$  在  $\alpha$  下的像, 称  $x$  是  $v$  在基  $\alpha$  下的坐标.
- 如果  $\alpha : \mathbb{R}^n \rightarrow V$  不满, 我们总可以找到  $v \in V$  不能被  $\alpha$  线性表示. 对于无限维向量空间, 因为任何有限向量组对应的映射都不满(why), 我们总可以往其中添加新的线性无关的向量, 从而在其中找到任意多个线性无关的向量, 从而给定任意  $\mathbb{R}^n$  到  $V$  的单射, 这表明无限维的向量空间确实比有限维的向量空间更"大" (听起来像是正确的废话)
- 对于一个  $m \times n$  的矩阵, 如果高斯消元法留下了  $r$  个阶梯头行, 则上一章提到的行变换  $\rho_1$  给出了  $\text{Im } \varphi$  和剩下前  $r$  行的列向量  $\{(x_1, x_2, \dots, x_r, 0, \dots, 0) \in \mathbb{R}^m \mid x_i \in \mathbb{R}\}$  的一一映射, 所以它经过行变换后是  $r$  维向量空间, 又因为行变换给出了一个一一映射, 所以  $\text{Im } \varphi$  是  $r$  维的.

## 线性代数基本定理

经过这么久, 我们已经研究过了线性方程并定义了什么是向量组, 什么是基底和坐标, 通过这些把有限维向量空间的问题转化到了  $\mathbb{R}^n$  上容易研究的问题. 所以, 线性代数中最重要、最基础也是最精华的定理终于要来了:

线性代数基本定理: 如果  $V_1$  和  $V_2$  是向量空间,  $\varphi : V_1 \rightarrow V_2$  线性且  $V_1$  维度有限, 记  $\varphi$  像的维数为其"秩", 即  $\text{rank } \varphi = \dim \text{Im } \varphi$ , 则有  $\dim \text{Ker } \varphi = \dim V_1 - \text{rank } \varphi$ .

证明: 由  $\text{Im } \varphi$  维度有限(why), 它同构于某个  $\mathbb{R}^r$ , 根据定义,  $r$  作为像的维数就是  $\varphi$  的秩,  $\varphi$  是到其像的满射. 所以可以分别取  $V_1$  和  $\text{Im } \varphi$  的一个基底, 则经过消元刚好有  $r$  个主元, 否则就不是满射了. 前文所述, 记  $n$  为  $V_1$  的维数, 则  $\dim \text{Ker } A = n - r$ , 证毕.

## 练习

3.1 用矩阵语言说明为什么当  $m \neq n$  时  $\mathbb{R}^m$  和  $\mathbb{R}^n$  不同构 (这个结果上一章证过, 这里作为复习)

3.2 证明  $C[0, 1]$  是无穷维空间 (回忆一下,  $C[0, 1]$  是  $[0, 1]$  上的所有连续实函数组成的集合)

3.3 你能叙述  $\mathbb{C}$ -维数的定义吗? 如果  $V = \mathbb{R}^2$ , 把它看成  $\mathbb{R}$ -线性空间, 它的维数是多少? 把它看成  $\mathbb{C}$ -线性空间呢? 更进一步地,  $\dim_{\mathbb{Q}} V$  有限吗? 为什么? ( $\mathbb{Q}$  是有理数域)

3.4 证明线性空间  $V$  和  $0$  之间唯一的映射是  $0$  映射, 并说明  $0 : V \rightarrow 0$  满,  $0 : 0 \rightarrow V$  单. 如果  $V \rightarrow 0$  是满射, 则  $V$  是什么?

3.5 证明一个包含  $0$  的向量组按照定义是线性相关的.

3.6 如果  $n + 1$  个向量  $\alpha_1, \alpha_2, \dots, \alpha_{n+1}$  中线性相关但是其中任意  $n$  个线性无关, 证明任意一个向量都可以用其它向量的线性组合来表示. 更加具体地, 对如下的 4 个向量:

$$\alpha_1 = \begin{bmatrix} 3 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \alpha_2 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ -1 \end{bmatrix}, \alpha_3 = \begin{bmatrix} -1 \\ -1 \\ 3 \\ -1 \end{bmatrix}, \alpha_4 = \begin{bmatrix} -1 \\ -1 \\ -1 \\ 3 \end{bmatrix}$$

证明其中的任意 3 个线性无关, 但是这 4 个向量线性相关, 把每个向量用其它向量的线性组合表示.

3.7 我们知道通过移除线性相关的向量, 可以把一个包含  $n$  个向量的向量组  $\alpha$  变为一个包含  $r$  个向量的向量组  $\alpha'$ . 我们可以优先移除前面的向量, 也可以优先移除后面的向量, 也可以按照任何你喜欢的顺序, 这可能导致留下来的向量不同, 假设一种方式留下了  $r_1$  个向量, 另一种方式留下了  $r_2$  个向量, 证明  $r_1 = r_2$ . 记  $r = r_1 = r_2$ , 证明可以被  $\alpha$  线性表示的  $r + 1$  个向量  $\beta_1, \beta_2, \dots, \beta_r, \beta_{r+1}$  线性相关.

3.8 如果  $\varphi = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ . 通过高斯消元为  $\text{Ker } \varphi$  找到一个基.

3.9 如果  $\alpha = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$ , 在不用消元法的前提下证明  $\alpha$  是单射.

3.10 如果  $\alpha = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}$ , 证明  $\alpha$  是双射, 所以给出了  $\mathbb{R}^3$  的一个基, 写出向量  $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$  在基  $\alpha$  下的坐标

3.11 如果  $\alpha = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $\alpha$  是单射吗?  $\alpha$  是满射吗? 如果  $\alpha$  不是满射, 向  $\alpha$  中添加更多向量让它变为满射, 如果你希望让它保持单射, 你最多可以添加多少个向量? 从  $(\mathbb{R}^n, \cdot)$  到  $H$  —— 内积和正交化

## 第4章 从 $(\mathbb{R}^n, \cdot)$ 到 $H$ —— 内积和正交化

本章中我将会始终讨论  $\mathbb{R}$  上的向量空间, 而非像之前的章节那样总可以自然地拓展到任何一个域  $K$  上. 在某些教科书中, 在复空间上定义内积是合理的, 它从某些意义上比实数内积更复杂. 但是简洁起见, 我们只讨论实向量空间. 字母  $\alpha$  和  $\beta$  可能被用来指代向量或者基底, 这取决于上下文.

### 什么是内积?

如果  $H$  是一个实向量空间, 那么一个内积以两个向量作为参数并且得到一个实数, 它反映了两个向量之间的相似或者说同方向性, 我们高中学过的点乘就是一个重要例子. 所以内积从形式上是

$\langle -, - \rangle : H \times H \rightarrow \mathbb{R}$ . 在高中, 点乘通常被定义为  $\alpha \cdot \beta = |\alpha||\beta| \cos \angle(\alpha, \beta)$ , 其中  $\angle(\alpha, \beta)$  是两个向量之间的夹角. 如果内积是正的, 这意味着两个向量处在相似的方向 (成锐角), 如果内积是负的, 这意味着两个向量处在相反的方向 (成钝角), 如果内积是 0, 它们两个向量就是正交的 (成直角). 为了用这种方式定义内积, 我们需要两个函数. 第一个是模 (或称之为模长、长度、甚至是绝对值)  $|-| : H \rightarrow \mathbb{R}_{\geq 0}$ . 第二个是向量间夹角  $\angle : H \times H \rightarrow [0, \pi]$ . 这种方法在欧氏空间中走得通, 但是来到一般的实向量空间时就开

始效果堪忧了,因为这个空间不一定自然拥有模和夹角.此外这么定义会让一些性质的证明变得困难,比如说:  $\alpha \cdot (\beta_1 + \beta_2) = \alpha \cdot \beta_1 + \alpha \cdot \beta_2$ .这就是为什么我们需要重新定义一个内积以让我们处理问题更加方便.

重新定义一个概念并不是一件容易的事情,我们需要确保它和之前的概念兼容.之前的路线问题出在:一个一般的空间上并不天然拥有一个度量模长和角度的函数,而我们最终的解决方案是先定义内积,然后反过来定义模和角度.

就像我们之前定义向量空间的方法一样,我们先通过几条规则来定义什么可以称之为内积.内积是一个  $\langle -, - \rangle : H \times H \rightarrow \mathbb{R}$  的函数,对任意的  $\alpha, \beta \in H$  和  $k_1, k_2 \in \mathbb{R}$  满足:

1.  $\langle \alpha, \beta \rangle = \langle \beta, \alpha \rangle$  (对称性)
2.  $\langle k_1\alpha_1 + k_2\alpha_2, \beta \rangle = k_1\langle \alpha_1, \beta \rangle + k_2\langle \alpha_2, \beta \rangle$  (对第一个参数线性)

注:前两条规则可以推导出

$$\langle \alpha, k_1\beta_1 + k_2\beta_2 \rangle = \langle k_1\beta_1 + k_2\beta_2, \alpha \rangle = k_1\langle \beta_1, \alpha \rangle + k_2\langle \beta_2, \alpha \rangle = k_1\langle \alpha, \beta_1 \rangle + k_2\langle \alpha, \beta_2 \rangle,$$

所以事实上它也对第二个变量线性,这个性质被称为双线性

3. 对于  $\alpha \neq 0, \langle \alpha, \alpha \rangle > 0$  (正定性)

注:  $\langle 0, 0 \rangle = 0\langle 0, 0 \rangle$  可以推出  $\langle 0, 0 \rangle = 0$

为了和公式  $\langle \alpha, \alpha \rangle = |\alpha||\alpha| \cos \angle(\alpha, \alpha) = |\alpha|^2 \cos 0 = |\alpha|^2$  兼容,我们需要重新定义模长为  $|\alpha| = \sqrt{\langle \alpha, \alpha \rangle}$ , 并且重新定义角度为  $\angle(\alpha, \beta) = \arccos \frac{\langle \alpha, \beta \rangle}{|\alpha||\beta|}$ . 如果  $\langle \alpha, \beta \rangle = 0$ , 称  $\alpha$  和  $\beta$  正交(垂直), 记作  $\alpha \perp \beta$ . 这个定义和某些读者的直觉可能不太相符,但是众所周知度量角的大小是为了描述一个物体相对另一个物体的定向,这与内积的目的基本相同.此外,在计算中,计算内积通常比计算角度更加容易.几何的基本对象是长度和角度,而二者都可以由内积导出,所以一个向量空间的"几何"完全由其上定义的内积决定.这里最重要的一点是:内积取代了长度和角度这些,成为了最基本的概念.

## 转置和 $\mathbb{R}^n$ 中的标准内积

对于两个向量  $x, y \in H = \mathbb{R}^n$ , 读者可能已经在高中课上学过它们的点积是  $x \cdot y = \sum_{i=1}^n x_i y_i$ . 在我们已

经重定义了内积的前提下,为了确认点积确实是一种内积,我们需要验证它满足内积的三个条件.这留给读者,作为练习.点积也被称作  $\mathbb{R}^n$  上的标准内积,但是我们只在讨论  $\mathbb{R}^n$  时使用"标准内积"这个说法,这是因为  $\mathbb{R}^n$  有一组标准的基底  $e_1, e_2, \dots, e_n$ , 但是大部分的线性空间并没有.

为了把这个过程看的更清楚,不妨引入转置的概念.一个矩阵的转置就是把它的列重新写成行,  $A$  的转置被记作  $A^T$ , 例如:

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

所以一个行向量的转置是一个列向量,而一个列向量的转置是一个行向量.在引入转置的记号后,  $\mathbb{R}^n$  上的标准内积可以被写成  $x \cdot y = x^T y$ , 也就是说:

$$x \cdot y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i$$

注：转置的重要性质是  $(AB)^T = B^T A^T$ ，最后一节中将会用到这个性质。这可以通过把每一项展开证明：设  $a_{ik}, b_{kj}, c_{ij}$  分别为  $A$  的第  $(i, k)$  项,  $B$  的  $(k, j)$  项,  $AB$  的第  $(i, j)$  项, 则  $c'_{ji} = c_{ij} = \sum_k a_{ik} b_{kj} = \sum_k b'_{jk} a'_{ki}$ , 在  $(AB)^T$  中,  $c'_{ji} = c_{ij} = \sum_k a_{ik} b_{kj} = \sum_k b'_{jk} a'_{ki}$ . 对每个元素都有  $c'_{ji} = \sum_k b'_{jk} a'_{ki}$ , 这推出  $(AB)^T = B^T A^T$

## 非标准内积和正定矩阵

既然我们把熟悉的内积叫做"标准"内积, 你可能在想, 有什么"非标准"的内积吗? 事实上, 在内积的新定义下, 有无穷多个内积, 而标准内积只是其中之一. 例如, 考虑  $\langle x, y \rangle = 2x^T y$ , 我们就得到了一个"非标准"的内积. 现在摆在我们面前的有三个问题:

1. 我们真的需要这些"非标准"的内积吗?
2. 我们可以找到所有的内积吗?
3. 这些内积有什么共同点?

以下是我对这些问题的答案

1. 有如下几个原因:

(1) 加权内积: 加权内积会为向量的不同分量加上不同的权重, 通常用在数值分析或有限元分析中用于模拟具有不同性质的物理现象. 当某些分量比较重要, 而另一个不那么重要的时候, 通常使用"加权"的内积, 给每个变量分配一个权重, 例如, 在  $H = \mathbb{R}^2$  上定义的  $\langle x, y \rangle = 2x_1 y_1 + x_2 y_2$  给第一个分量权重 2, 给第二个分量权重 1

(2) 内积中的交叉项: 基向量并不总是垂直的, 这时可能出现交叉项. 例如, 假设我们在讨论空间  $H = \{x \in \mathbb{R}^3 \mid x_1 + x_2 = x_3\}$ , 映射  $\alpha: \mathbb{R}^2 \rightarrow H, (x_1, x_2) \mapsto (x_1, x_2, x_1 + x_2)$  是一个同构, 并且给出了  $H$  的一个基, 其中基向量  $\alpha_1 = \alpha(e_1) = (1, 0, 1), \alpha_2 = \alpha(e_2) = (0, 1, 1)$ , 所以  $(x_1, x_2, x_1 + x_2)$  在基  $\alpha$  下的坐标为  $(x_1, x_2)$ . 坐标为  $(x_1, x_2)$  的向量和坐标为  $(y_1, y_2)$  的向量的内积是  $(x_1, x_2, x_1 + x_2) \cdot (y_1, y_2, y_1 + y_2) = 2x_1 y_1 + 2x_2 y_2 + x_1 y_2 + x_2 y_1$ . 2 表明基向量的长度不是 1 (实际上长度是  $\sqrt{2}$ ), 而  $x_1 y_2$  和  $x_2 y_1$  这样的交叉项表明基向量并非垂直 (实际上角度为  $\pi/3$ )

2. 为了研究  $H \times H$  上双线性函数的性质, 我们只需要知道它如何作用于基向量就可以了. 先说  $\mathbb{R}^n, \mathbb{R}^n$  有一组向量  $e_1, e_2, \dots, e_n$  作为基底, 所以

$$\begin{aligned} \langle x, y \rangle &= \left\langle \sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right\rangle \\ &= \sum_{i=1}^n x_i \left\langle e_i, \sum_{j=1}^n y_j e_j \right\rangle \\ &= \sum_{i=1}^n x_i \sum_{j=1}^n y_j \langle e_i, e_j \rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n \langle e_i, e_j \rangle x_i y_j \end{aligned}$$

记  $g_{ij} = \langle e_i, e_j \rangle$ , 则一个内积可以写成

$$\langle x, y \rangle = \sum_{1 \leq i, j \leq n} g_{ij} x_i y_j$$

在展开式子的过程中, 第二条性质 (双线性) 自动满足了, 所以我们需要验证其它的两条性质. 第一条性质是对称性, 这要求

$$\langle x, y \rangle = \sum_{1 \leq i, j \leq n} g_{ij} x_i y_j = \langle y, x \rangle = \sum_{1 \leq i, j \leq n} g_{ij} y_i x_j = \sum_{1 \leq i, j \leq n} g_{ji} x_i y_j$$

所以对任意的  $1 \leq i, j \leq n$ , 有  $g_{ij} = g_{ji}$  成立

而最后一条性质: 正定性, 这要求非零向量和自身的内积是正的, 也就是说对  $x \neq 0$ :

$$\langle x, x \rangle = \sum_{1 \leq i, j \leq n} g_{ij} x_i x_j > 0$$

而内积也可以用矩阵语言来表示

$$\langle x, y \rangle = x^T G y, \text{ 其中 } G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix}$$

为了看得更清楚,  $G$  可以被视作接受两个向量输入并且给出一个数字作为输出 (如下的公式块只是为了便于理解, 并非正式数学语言)

$$x^T G y \rightarrow \begin{bmatrix} & y_1 & y_2 & \cdots & y_n \\ & \downarrow & \downarrow & \cdots & \downarrow \\ x_1 & \rightarrow & g_{11} & g_{12} & \cdots & g_{1n} \\ x_2 & \rightarrow & g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \rightarrow & \vdots & \vdots & \ddots & \vdots \\ x_n & \rightarrow & g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix} \rightarrow \sum_{\text{全部}} \begin{bmatrix} g_{11} x_1 y_1 & g_{12} x_1 y_2 & \cdots & g_{1n} x_1 y_n \\ g_{21} x_2 y_1 & g_{22} x_2 y_2 & \cdots & g_{2n} x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} x_n y_1 & g_{n2} x_n y_2 & \cdots & g_{nn} x_n y_n \end{bmatrix}$$

这个矩阵必须对称 ( $G^T = G$ , 这是因为  $g_{ij} = g_{ji}$ ), 此外对  $x \neq 0$  有  $x^T G x > 0$ . 满足这一条件的矩阵被称作正定矩阵. 需要注意的是  $f(x) = x^T G x$  是关于  $x$  的齐次二次函数. (不要和线性函数混淆) 更多关于正定的内容将会在第6章中呈现, 在此我们只需要对它在 0 处取得极小值有印象就可以了. 所以  $\mathbb{R}^n$  上的内积和正定矩阵一一对应, 矩阵  $G$  通常被称作度量矩阵. (或度规) 特别地, 令  $G = I_n$ , 即  $n \times n$  单位矩阵, 则它很显然对称, 且  $x^T G x = x^T I_n x = x^T x = x_1^2 + x_2^2 + \cdots + x_n^2 \geq 0$ , 仅在 0 处取到 0 值, 这恰好就是标准内积. 对于一般的空间, 只需把这里的基底换成对应的基底即可.

3. 定义内积的时候, 我们只依赖于三条性质: 对称性、双线性和正定性, 所以在证明相关性质时我们只有, 实际上也只需要它们. (性质直接从三条规则导出, 不依赖于  $\mathbb{R}^n$ ) 在高中时我们已经学过了关于向量, 特别是关于内积和长度的很多结果: 比如柯西不等式告诉我们  $\alpha \cdot \beta \leq |\alpha| |\beta|$ , 此外还有余弦定理, 三角不等式, 又或许你学过一个向量在另一个向量的投影. 下文将会证明这些结果可以直接拓展到任意的内积空间.

#### (1) 柯西不等式

你是否想过, 在没有验证  $|\langle \alpha, \beta \rangle| \leq |\alpha| |\beta|$  的情况下, 前文为什么能够贸然取  $\arccos$  呢? 接下来我将会使用一点点技巧来证明它确实是对的.

考虑如下函数

$$f(t) = \langle \alpha + t\beta, \alpha + t\beta \rangle = \langle \alpha, \alpha \rangle + 2t\langle \alpha, \beta \rangle + t^2\langle \beta, \beta \rangle$$

由于夹角只对非零向量定义, 这里有  $\langle \beta, \beta \rangle \neq 0$ , 所以它是二次函数. 正定推出  $f(t) \geq 0$ , 当且仅当  $\alpha + t\beta = 0$  时取等号. 由于这个函数最多有一个实零点, 其

$$\Delta = b^2 - 4ac = 4\langle \alpha, \beta \rangle^2 - 4\langle \alpha, \alpha \rangle \langle \beta, \beta \rangle \leq 0, \text{ 仅在 } \alpha + t\beta = 0 \text{ 对某个 } t \text{ 成立时取等号,}$$

$$\langle \alpha, \beta \rangle^2 \leq \langle \alpha, \alpha \rangle \langle \beta, \beta \rangle = |\alpha|^2 |\beta|^2 \Rightarrow |\langle \alpha, \beta \rangle| \leq |\alpha| |\beta| \text{ 等号成立当且仅当 } \alpha \text{ 和 } \beta \text{ 共线}$$

#### (2) 余弦定理

$$|\alpha - \beta| = \sqrt{\langle \alpha - \beta, \alpha - \beta \rangle} = \sqrt{\langle \alpha, \alpha \rangle + \langle \beta, \beta \rangle - 2\langle \alpha, \beta \rangle} = \sqrt{|\alpha|^2 + |\beta|^2 - 2|\alpha||\beta|\cos\theta}$$

用我们更为熟悉的语言, 如果在三角形  $OAB$  中,  $\alpha = \vec{OA}, \beta = \vec{OB}$ , 则

$$|\vec{AB}| = |\alpha - \beta| = \sqrt{|\vec{OA}|^2 + |\vec{OB}|^2 - 2|\vec{OA}||\vec{OB}|\cos\theta}$$

特别地, 当  $\alpha \perp \beta$  时,  $\langle \alpha, \beta \rangle$  这一项变为了 0, 得到了勾股定理

(3) 三角不等式

$$|\alpha + \beta| = \sqrt{|\alpha|^2 + |\beta|^2 + 2|\alpha||\beta|\cos\theta} \leq \sqrt{|\alpha|^2 + |\beta|^2 + 2|\alpha||\beta|} = \sqrt{(|\alpha| + |\beta|)^2} = |\alpha| + |\beta|$$

(4) 一个向量在其它向量上的投影

对  $\alpha, \beta \in H$ , 向量  $\beta$  在  $\alpha$  上的投影为  $\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha$ . 这是因为

$$\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \alpha = \frac{|\alpha||\beta|\cos\angle(\alpha, \beta)}{|\alpha|^2} |\alpha| \frac{\alpha}{|\alpha|} = (|\beta|\cos\angle(\alpha, \beta)) \frac{\alpha}{|\alpha|}$$

$|\beta|\cos\angle(\alpha, \beta)$  是投影长度, 而  $\frac{\alpha}{|\alpha|}$  是单位方向向量, 为了方便, 记  $\beta$  在  $\alpha$  上的投影为  $\text{proj}_\alpha(\beta)$ .

## 等距同态和施密特正交化

现在我要讲讲我们的标题了, " $(\mathbb{R}^n, \cdot)$  到  $H$ " 也可以指一种映射. 众所周知一个线性映射保持了加法和标量乘法, 而现在应该内积空间在向量空间的基础上多了一个内积, 这意味着  $H$  可以看成是一个 4 元组

$(H, \text{向量加法}, \text{标量乘法}, \text{内积})$

$(\mathbb{R}^n, \cdot)$  是一个特例, 以其上的点积作为内积:

$(\mathbb{R}^n, \text{向量加法}, \text{标量乘法}, \text{点积})$

如果一个映射保持了"结构", 除了要是线性的, 他还应该保持内积. 这样的映射被称作等距同态, 这是由"等距"和"同态"构成的复合词, 反映了它不仅是线性映射(同态), 还保持度量.

初等线性代数和表示论的一个主要目标就是把(有限维)抽象空间中的向量用它们在某个基底下的坐标来表示, 即  $\mathbb{R}^n$  中的向量表示, 并且通过更为具体的矩阵来研究抽象的映射. 对于一个有限维向量空间  $H$  的一个基底  $\alpha$ ,  $H$  上的度量自然诱导出一个  $\mathbb{R}^n$  上的度量 ( $g_{ij} = \langle e_i, e_j \rangle$  由  $\langle \alpha(e_i), \alpha(e_j) \rangle$  决定). 既然  $\mathbb{R}^n$  上有一个标准内积, 一个好想法是找到一个基底, 使空间的度量诱导出的  $\mathbb{R}^n$  上的标准内积. 换言之, 对  $i = j, g_{ij} = 1$ , 对  $i \neq j, g_{ij} = 0$ . 所以  $\langle \alpha_i, \alpha_i \rangle = 1$ , 这表明  $|\alpha_i| = 1$ , 而对  $i \neq j, \langle \alpha_i, \alpha_j \rangle = 0$ , 这表明  $\alpha_i \perp \alpha_j$ . 这样的一个基  $[\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$  被称为一个标准正交基 (或归一正交基) 被称作"标准"是因为它诱导出的  $\mathbb{R}^n$  上的内积是标准内积.

回想我们先前如何获得有限维向量空间的一个基底: 先找到一组足以"表达"整个空间的向量, 然后逐渐去除线性相关的直到获得一组线性无关的向量. 为了从一个基底得到一个正交基, 我们要做的大体上差不多. 不过这次我们要去除的不是多余的向量, 而是向量中不垂直的多余的"部分". 如下的过程称作施密特正交化:

从  $H$  的一个基底  $\alpha = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$  开始

1. 令  $\beta_1 = \alpha_1$

令  $\beta_2 = \alpha_2 - \text{proj}_{\beta_1}(\alpha_2)$ , 则  $\beta_2 \perp \beta_1$

令  $\beta_3 = \alpha_3 - \text{proj}_{\beta_1}(\alpha_3) - \text{proj}_{\beta_2}(\alpha_3)$ , 则  $\beta_3 \perp \beta_1$  且  $\beta_3 \perp \beta_2$ .

.....

令  $\beta_n = \alpha_n - \text{proj}_{\beta_1}(\alpha_n) - \text{proj}_{\beta_2}(\alpha_n) - \cdots - \text{proj}_{\beta_{n-1}}(\alpha_n)$ , 则  $\beta_n \perp \beta_i$  对  $i < n$  成立

2. 最终把每个向量除以它的模, 以让它们模长为 1, 即归一化

$$\varepsilon_i = \frac{\beta_i}{|\beta_i|}$$

$$\langle \varepsilon_i, \varepsilon_j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \text{ 是 } H \text{ 的一个归一正交基}$$

在 python 代码中, 这个过程写作 (设  $\alpha_i$  均为  $\mathbb{R}^n$  的向量)

```
# numpy 提供了方便的数组运算
import numpy as np
# ndarray 在 numpy 中是 n 维数组类型
def Gram_Schmidt(a: list[np.ndarray]) -> list[np.ndarray]:
    # 获得向量数量
    n: int = len(a)
    # 从 alpha_1 开始
    b: list[np.ndarray] = [a[0]]
    for j: int in range(1, n):
        v: np.ndarray = a[j]
        for i: int in range(i):
            bi_dot_aj: float = b[i].dot(a[j])
            bi_dot_bi: float = b[i].dot(b[i])
            v -= bi_dot_aj / bi_dot_bi * b[i]
        # 设置 b[j] 为 v
        b.append(v)
    # np.linalg.norm 计算 |beta_i|
    return [bi / np.linalg.norm(bi) for bi in b]
```

你是否认为这个操作看起来很熟悉? 在第2章中高斯消元引出了矩阵的LU(或者PLU)分解. 因为我们在高斯消元中做行操作, 我们从左侧分离出了一个下三角矩阵, 它编码了我们操作矩阵的过程. 而在正交化过程中, 我们在对向量操作, 这些向量作为矩阵的列, 应该从右侧分离出一个记录了我们如何操作向量组的矩阵.

同样地, 首先把这个过程改写成矩阵形式 (这些向量作为列, 但本身不必是列向量) 用  $\gamma_1$  表示向量之间的相减,  $\gamma_2$  表示除以模长的过程:

$$\begin{aligned} [\beta_1 \ \alpha_2 \ \cdots \ \alpha_n] &\xrightarrow{\gamma_{12}} [\beta_1 \ \beta_2 \ \cdots \ \alpha_n] \rightarrow \cdots \\ &\xrightarrow{\gamma_{1n}} [\beta_1 \ \beta_2 \ \cdots \ \beta_n] \xrightarrow{\gamma_2} [\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_n] \end{aligned}$$

对于每个  $j \geq 2$ , 在  $\gamma_{1j}$  中,  $\alpha_j$  被减去每个满足  $1 \leq i < j$  的  $\beta_i$  的倍数, 记这个系数为  $c_{ij} = \frac{\langle \beta_i, \alpha_j \rangle}{\langle \beta_i, \beta_i \rangle}$ , 则这些列操作可以为

$$\alpha \xrightarrow{\gamma_{12}: C_2 - c_{12}C_1} \xrightarrow{\gamma_{13}: C_3 - c_{13}C_1 - c_{23}C_2} \cdots$$

现在来关注  $\gamma_{1j}$ :

对  $i < j$  (右侧的矢量只有 1 个非零的数字)

$$\beta_i = [\beta_1 \quad \dots \quad \beta_i \quad \dots \quad \alpha_n] \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

对于  $i > j$  (右侧的矢量只有 1 个非零的数字)

$$\alpha_i = [\beta_1 \quad \dots \quad \alpha_i \quad \dots \quad \alpha_n] \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

对于第  $j$  列, (1 上方的数为  $-c_{1j}, -c_{2j}, \dots, -c_{(j-1)j}$ , 下方的数字全都是 0)

$$\beta_j = [\beta_1 \quad \dots \quad \alpha_j \quad \dots \quad \alpha_n] \begin{bmatrix} -c_{1j} \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

把它们写到一起, 就得到了  $\gamma_{1j}$  的表达式

$$[\beta_1 \quad \dots \quad \beta_j \quad \dots \quad \alpha_n] = [\beta_1 \quad \dots \quad \alpha_j \quad \dots \quad \alpha_n] \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & -c_{1j} & & & & & & \\ & & \vdots & & & & & & \\ & & 1 & & & & & & \\ & & & \ddots & & & & & \\ & & & & \ddots & & & & \\ & & & & & & & & 1 \end{bmatrix}$$

$\gamma_1$  实际上就是把这些列操作合并到一起

$$\beta = \alpha \gamma_{12} \gamma_{13} \dots \gamma_{1n} \Rightarrow \gamma_1 = \gamma_{12} \gamma_{13} \dots \gamma_{1n}$$

但是你会发现合并这些项很困难, 原因和第2章的一样, 解决方案也和第二章的一样: 找  $\gamma$  的逆, 它会更直接告诉我们关于正交化的信息:

容易验证 (读者应当自己验证)

$$\gamma_{1j}^{-1} = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & c_{1j} & & & & & & \\ & & \vdots & & & & & & \\ & & 1 & & & & & & \\ & & & \ddots & & & & & \\ & & & & \ddots & & & & \\ & & & & & & & & 1 \end{bmatrix}$$

其中对角线都是 1, 在第  $j$  列主对角线上方第  $i$  行的元素为  $c_{ij}$ , 其它所有元素都是 0. 把它们乘到一起得到了

$$\gamma_1^{-1} = \gamma_{1n}^{-1} \gamma_{1(n-1)}^{-1} \cdots \gamma_{12}^{-1} = \begin{bmatrix} 1 & c_{12} & \cdots & c_{1n} \\ & 1 & \cdots & c_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

这并没有结束,  $\gamma_2$  把  $\beta$  中的每个向量除以它的模长让它们归一化, 所以

$$\gamma_2 = \begin{bmatrix} |\beta_1|^{-1} & & & \\ & |\beta_2|^{-1} & & \\ & & \ddots & \\ & & & |\beta_n|^{-1} \end{bmatrix} \Rightarrow \gamma_2^{-1} = \begin{bmatrix} |\beta_1| & & & \\ & |\beta_2| & & \\ & & \ddots & \\ & & & |\beta_n| \end{bmatrix}$$

而  $\gamma^{-1}$  的表达式

$$\begin{aligned} \varepsilon &= \beta \gamma_2 = \alpha \gamma_1 \gamma_2 = \alpha \gamma \Rightarrow \gamma = \gamma_1 \gamma_2 \Rightarrow \\ \gamma^{-1} &= \gamma_2^{-1} \gamma_1^{-1} \\ &= \begin{bmatrix} |\beta_1| & & & \\ & |\beta_2| & & \\ & & \ddots & \\ & & & |\beta_n| \end{bmatrix} \begin{bmatrix} 1 & c_{12} & \cdots & c_{1n} \\ & 1 & \cdots & c_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix} \\ &= \begin{bmatrix} |\beta_1| & |\beta_1|c_{12} & \cdots & |\beta_1|c_{1n} \\ & |\beta_2| & \cdots & |\beta_2|c_{2n} \\ & & \ddots & \vdots \\ & & & |\beta_n| \end{bmatrix} \end{aligned}$$

在这么做的过程中  $\alpha = \varepsilon \gamma^{-1}$  被分解为一个归一正交基和一个上三角矩阵, 对于  $H = \mathbb{R}^n$  且使用标准内积的情况, 这把一个矩阵分解为两个  $M = QR$ . 其中  $Q$  中的向量是归一正交的, 每列模长都是 1, 且每两个不同的列都垂直, 这样的矩阵被称为正交阵. (这与  $Q^T Q = I_n$  等价)  $R$  是一个上三角矩阵, 其中的系数表示出我们是如何操作向量组让它正交的.

例如,  $\alpha = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$  是  $\mathbb{R}^3$  的一个基, 但并非归一正交, 让我们用施密特正交化方法把它正交化:

$$1. \beta_1 = \alpha_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\frac{\langle \beta_1, \alpha_2 \rangle}{\langle \beta_1, \beta_1 \rangle} = \frac{1}{2}, \beta_2 = \alpha_2 - \frac{1}{2} \beta_1 = \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \end{bmatrix},$$

$$\frac{\langle \beta_1, \alpha_3 \rangle}{\langle \beta_1, \beta_1 \rangle} = \frac{1}{2}, \frac{\langle \beta_2, \alpha_3 \rangle}{\langle \beta_2, \beta_2 \rangle} = \frac{1}{3}, \beta_3 = \alpha_3 - \frac{1}{2} \beta_1 - \frac{1}{3} \beta_2 = \begin{bmatrix} -2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

$$2. \varepsilon_1 = |\beta_1|^{-1} \beta_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\varepsilon_2 = |\beta_2|^{-1}\beta_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

$$\varepsilon_3 = |\beta_3|^{-1}\beta_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

基底  $[\varepsilon_1 \ \varepsilon_2 \ \varepsilon_3]$  归一正交, 相应的分解出的上三角矩阵  $R$  为

$$R = \begin{bmatrix} \sqrt{2} & & \\ & \sqrt{3}/\sqrt{2} & \\ & & 2/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & -1/2 & -1/2 \\ & 1 & -1/3 \\ & & 1 \end{bmatrix}$$

## 一般空间上内积的例子

1. 区间  $[a, b]$  上的连续函数构成向量空间, 微积分告诉我们这样的函数总是可积的, 可以定义两个函数的内积为

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$

注意到积分可以视作指标  $x$  在连续集合内取值时的求和, 此外我们也可以乘以一个连续函数  $\mu(x)$  作为权重以得到  $[a, b]$  上的加权内积

$$\langle f, g \rangle = \int_a^b f(x)g(x)\mu(x)dx$$

这样定义的内积对泛函分析和信号处理等很有用

2. 在概率论中, 两个随机变量的内积被定义为乘积的期望值,  $\langle X, Y \rangle = E(XY)$ . 即使你不知道这是什么意思, 它为什么成立也没关系, 你只需要知道线性代数对概率论也很重要
3. 为了比较两幅图片或者两个离散信号的相似度, 可以把它们视作向量. 内积可以反映它们有多相似.

## 关于李群的简介 \*

初学者如果不熟悉群可以跳过这一节

一个群是带有一个二元运算的集合  $(G, \cdot)$

1. 运算满足结合律
2. 存在单位元
3. 所有元素都存在逆

设  $Q \in \mathbb{R}^{n \times n}$  是正交阵, 则每一列模长都是 1 而不同列正交, 所以  $Q^T Q = I_n$ , 即  $Q^{-1} = Q^T$ . 在变换后两个向量的内积是

$$(Qu) \cdot (Qv) = (Qu)^T(Qv) = u^T Q^T Q v = u^T I_n v = u^T v = u \cdot v$$

这表明内积在正交阵作用下保持不变, 所以长度和角度也保持不变. 注意到  $e_i$  的模是 1, 而  $e_i$  和  $e_j$  的点积是 0, 所以一个保持内积的矩阵的列模长总是 1 并且互相垂直. 如果  $P, Q$  都是正交矩阵, 那么

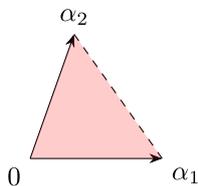
$(PQu) \cdot (PQv) = (Qu) \cdot (Qv) = u \cdot v$ , 所以  $PQ$  也保持内积不变, 是正交阵, 故正交阵组成的集合在矩阵乘法下封闭,  $n \times n$  矩阵全体被称为  $\mathbb{R}^n$  上的正交群, 记作  $O_n$ , 或者  $O_n(\mathbb{R})$ . 因为元素可以连续取值, 所以这又被称作一个李群.

## 练习

4.1 对 4 个向量  $\alpha_1 = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 1 \end{bmatrix}$ ,  $\alpha_2 = \begin{bmatrix} 2 \\ 1 \\ -2 \\ -1 \end{bmatrix}$ ,  $\alpha_3 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -2 \end{bmatrix}$ ,  $\alpha_4 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}$ . 计算两两之间的点积, 哪些向量对是互相垂直的?

4.2 先检验  $\alpha_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ ,  $\alpha_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$  是  $\text{Ker} [1 \ 1 \ 1]$  的一个基, 然后计算这个基下的度量矩阵, 设在这个基下,  $v_1$  坐标为  $x_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ,  $v_2$  坐标为  $x_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ , 计算  $v_1$  和  $v_2$  的模和夹角.

4.3 设一个二维的向量空间在某基底  $\alpha$  下有度量矩阵  $G = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ . 两个基向量的模长是多少?  $\alpha_1$  和  $\alpha_2$  点的距离是多少? 根据边长, 计算三角形  $0\alpha_1\alpha_2$  的面积 (面积可以用长度和角度计算, 所以面积也由内积决定)



4.4 证明  $\alpha - \text{proj}_\beta(\alpha)$  和  $\beta$  垂直, 对多项式定义如下内积:  $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ . 如果  $f(x) = 1$  是常值函数,  $g(x) = x$ , 计算  $g - \text{proj}_f g$

4.5 (这一问需要一些计算能力) 假设对小于 3 次的多项式(即  $a + bx + cx^2$  形式的多项式)定义如下内积:

$$\langle f, g \rangle = \int_0^{+\infty} f(x)g(x)e^{-x} dx$$

显然  $\alpha_1 = 1, \alpha_2 = x, \alpha_3 = x^2$  是一个基, 计算这个基下的度量矩阵并且用施密特正交化方法为这个空间找到一个归一正交基.

4.6 注意到  $\alpha = \varepsilon R$  把一个向量组  $\alpha$  写成一个归一正交向量组  $\varepsilon$  的线性组合, 其中  $R$  作用在右侧作为过渡矩阵. 为这个基变换过程画一个交换图. 如果某向量原来的坐标是  $x$ , 它在归一正交基下的坐标是什么?

## 第5章 从 $V$ 到 $V$ —— 特征值和行列式

许多教科书引入行列式的方式莫名其妙

### 线性变换会在哪里出现

先回顾什么是线性变换: 我们称一个把线性空间映射到自身的线性映射称为线性变换, 即一个  $V$  上的线性变换是一个映射  $\varphi: V \rightarrow V$

线性变换作用在一个空间上, 当我们需要输入一个向量并且输出同一个空间内的向量时线性变换就出现了. 比如, 当一个光子碰到一面镜子并且反射的时候, 镜子接受了它的动量并且给了它一个新的动量作为输出, 所以镜子可以看作给出了动量所处空间的线性变换. 对于一个线性递推数列  $a: \mathbb{R} \rightarrow \mathbb{R}$  (递推公式为  $a_{n+m} = k_0 a_n + k_1 a_{n+1} + \dots + k_{m-1} a_{n+m-1}$ ). 去除其前  $t$  项并且把每一个  $a_{n+t}$  向前移动到  $a_n$  并

不改变它的递推公式. 注意到所有满足递推公式的数列组成一个线性空间, 且去除前  $t$  项的操作是线性的 (去除两个数列的和数列的前几项等价于两个去除了前几项的数列的和) 所以去除前  $t$  项的操作是线性的并且把数列映射到数列. 这是一个满足递推公式的数列空间上的线性变换. 设去除前  $t$  项的操作为  $R_t$ . 则容易验证  $R_{t_1+t_2} = R_{t_1} \circ R_{t_2}$ . 此外, 行变换和列变换作用在矩阵空间上, 我们可以使用它们消元或者完成正交化.

到处都有线性变换, 它们通常会乘在一起或者和自身相乘 (这意味着取一个线性变换的幂)

## 没有哪个基更加特殊 —— 为什么我们需要不变量

在从第三章开始的针对一般的 (特别是有限维) 线性空间的讨论中, 我们通过把抽象线性空间中的向量"翻译"为它们在某个基底得到的  $\mathbb{R}^n$  中的坐标. 这让我们感到更加亲切, 也更容易处理. 我们已经讨论过如何在两个基底之间转换, 也学会了如何在给定的内积下从任何一个基底得到一个正交基. 然而, 这一切都建立在一个前提下: 你不会为了如何选择基底而感到困难. 然而, 如果空间中并没有什么条件可以告诉你你该怎么选取基底导致你产生选择困难, 你解决问题就会困难起来. 这并不完全是一个笑话, 数学家们总是想要做出"好"的选择并且最终加入了选择公理告诉我们存在这样一个选择函数, 它总是可以从一个非空集合中选择一个元素. (这并非本文的话题, 如果你感兴趣你可以自行搜索以了解更多)

通常有很多种方式为一个向量空间选择基底并且研究映射的性质. 但是随着基向量的变化, 许多性质, 例如某个向量的坐标也会随之变化. 而那些不随基底变化的性质就被称作不变量. 找到和基底选取无关的性质通常很重要, 因为这样的性质反映了映射的本质属性. 换句话说, 当 Alice 和 Bob 从不同角度看一个空间时, 他们能得到什么共同结论?

在继续讨论这个话题之前, 我们需要回忆一些我们学过的关于(有限维)线性空间或者其上的线性映射的不变量:

1. 线性相关性: 一个向量组是否线性相关与基底选取无关
2. 维数: 一个空间的维数不会随着基底的选取而变化, 所以基底可能不同, 但是基向量的数量一定相同, 这就是空间的维数
3. 秩: 秩反映了一个线性映射的像的维数. 对于一个  $\mathbb{R}^n$  到其它空间的映射, 这个映射作为一个向量组, 即使向量组本身可能线性相关, 但是它的秩告诉我们一个向量组中有多少个"有用"的向量
4. 零化度: 此前我没提过这个概念, 一个映射  $\varphi$  的零化度就是  $\dim \text{Ker } \varphi$ . 它通过测量核的大小表明有"多少"向量会被映射到 0. 对于一个从  $\mathbb{R}^n$  到某个空间的映射, 即一个向量组, 这告诉我们有多少种"不同"的方式让它的线性组合为 0 (如果  $\alpha$  是一个向量组, 其中  $\alpha_2 - 2\alpha_1 = 0$  且  $\alpha_3 - 2\alpha_2 = 0$ , 这时我们不认为  $\alpha_3 - 4\alpha_1$  是一个"不同的"线性组合因为它可以看作前面两种组合的线性组合), 由于  $\dim \text{Ker } \varphi = n - r$ . 向量组的零化度可以视作其中"无用"的向量的数量. 所以线性代数基本定理又被称为秩-零化度定理. 对于  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  的情况, 秩和零化度分别对应阶梯头和自由变量的数量.

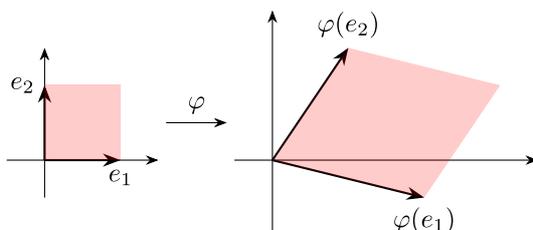
可以证明对于  $\varphi: V_1 \rightarrow V_2$ , 其中  $\dim V_1 = n$ ,  $\dim V_2 = m$ , 秩  $r$ , 与维数  $m, n$  可以生成所有关于这个线性映射的所有不变量. 这可以通过把所有矩阵都归约到一个标准型来实现. 此处"生成"的意思是所有的其它不变量都可以根据它们来计算. 例如, 零化度可以通过  $\dim \text{Ker } \varphi = n - r$  计算. 证明这一点并不难但不是我们现在要讨论的话题. 一定程度上, 这是因为你可以自由地选择  $V_1$  和  $V_2$  的基, 然而, 当我们为线性映射加上其它条件时, 可能会出现新的不变量. 在接下来的内容中, 我们把线性映射限制在一个空间上, 变成  $\varphi: V \rightarrow V$  并且谈论这是如何带来新性质的.

## 体积的变化 —— 一步步构建起行列式

### 第一步 —— 找到线性映射的额外不变量

#### 先自己思考一下

为了找到线性映射有什么额外的性质, 首先考虑一些简单的例子. 比如, 首先考虑  $\mathbb{R}^n$ . 但是你要假设你忘掉了它原本的坐标系统, 认为你的基向量是随便选取的, 它不够自然. 第一个例子是  $\mathbb{R}^1 = \mathbb{R}$ , 此时所有的线性变换都是一个缩放, 只由一个数字  $k$  决定, 即  $\varphi(x) = [k](x)$ . 在这里讨论线性无关都没有意义. 所以接下来让我们直接继续考虑  $\mathbb{R}^2$ . 下图和第二章初见矩阵的图是相同的图. 看看图并想想, 即使我们随便选择一个基, 我们可以知道什么关于它的性质.



#### 几何量

首先, 我们不可能讨论长度、角度或者面积的确切大小, 因为通常来讲这个空间不一定定义了度量, 不同的度量会给出不同的结果. 所以这些结果实际上是度量而非空间本身的性质. 然而我们可以讨论向量关于原来向量的变化. 在一般的空间中, 只有当两个向量线性相关的时候我们才能拿它们作比较. 所以如果我们要这么干我们就需要找到一些向量  $v$  使得  $\varphi(v)$  和  $v$  共线, 换言之,  $\varphi(v) = \lambda v$ . 这样的值  $\lambda$  被称作特征值, 对应的向量被称作特征向量. 我们会在之后详细讨论这些内容.

#### 关于空间中的面积

虽然我们不能讨论向量的长度, 在一个 2 维空间中, 我们总是可以讨论"面积"的变化. 你可能在想, 我们都没有度量, 哪来的面积呢? 好消息是, 可以证明面积变化的比例与度量的选取无关, 这可以通过用正方形近似并且证明所有的正方形缩放方式相同来证明. 如果你对此感到困惑, 请移步 3blue1brown 的视频再回到本文, 视频链接:

<https://www.bilibili.com/video/BV1ys411472E?p=7>

### 第二步 —— 定义二维空间的有向面积

#### 预备知识

什么是定向? 这并不难理解. 简而言之, 定向就是字母 'p' 和 'b' 之间的差异. 或者它是你左手和右手之间的差异. 如果你在一张有弹性的纸上写字母 'p', 无论如何旋转或拉伸纸张, 都无法将其变成字母 'b'. 尽管你可以翻转纸张, 将 'p' 变成 'b', 但这个动作不在原始的纸面二维空间内, 所以可以改变定向. 就手的例子来说, 无论你怎么旋转或移动左手, 它都不会变得和右手一样. 在化学中, 这种现象通常被称为手性. 如果存在一个 4 维空间, 理论上, 4 维空间中的生物可以将你"翻转"回到我们的 3 维空间. 但在我们的空间中, 这是不可能的.

可以为任何  $n$  维空间定义一个"定向". 注意到通过翻转垂直轴, 字母 'p' 变成了具有不同定向的字母 'b', 但接下来通过翻转水平轴, 它又变成了字母 'd', 可以旋转  $180^\circ$  与 'p' 重叠. 因此, 如果我们翻转偶数个轴, 定向保持不变, 而当我们翻转奇数个轴时, 定向会改变. 这看起来和  $(-1)^k$  很像, 当  $k$  是偶数时得到结果 1, 当  $k$  是奇数时得到结果  $-1$ .

然后让我们谈谈轴的交换. 如果我们绕着直线  $y = x$  翻转, 这会交换水平和垂直的坐标轴, 字母 'p' 会变成顺时针旋转了  $90^\circ$  的字母 'b', 它与 'p' 有不同的定向. 另一次翻转或轴的改变会将其定向还原. 因此, 似乎只有两种定向, 就像我们有两只不同的手一样. 如果一个物体的定向改变了偶数次, 那么与原始定向相比, 它是正定向的 (+1). 如果一个物体的定向改变了奇数次, 那么与原始定向相比, 它是负定向的 (-1).

我们经常使用的坐标系是右手坐标系, 因此通常认为左手坐标系为负定向.

## 为什么要有向面积

虽然这一小节我们在讨论  $\mathbb{R}^2$ , 之后我们将讨论的会是一个一般的二维线性空间

假设  $\mathbb{R}^2$  中有一个平行四边形, 其顶点为

$$A(0, 0), B(a, c), C(a + b, c + d), D(b, d)$$

这是向量  $\alpha_1 = (a, c)$  和  $\alpha_2 = (b, d)$  根据平行四边形法则确定的平行四边形, 请自己画个图并且计算它的面积.  $\alpha_2$  在  $\alpha_1$  的左侧或者右侧会引出不同的结果, 前者的面积为  $S = ad - bc$ , 而后者的为  $S = bc - ad$ . 这个公式可以总结为  $S = |ad - bc|$ . 此外我们还需要排除  $\alpha_1$  与  $\alpha_2$  共线的情况, 这让它不可能成为平行四边形.

定义有向面积的主要原因之一是, 它允许负面积作为负定向出现且允许 0 作为共线出现. 其中通过符号来确定定向只是一个副产物, 不过这也许也有其作用. 例如, 如果我们只是测量 'p' 和 'b' 上点间的距离, 我们会发现它们全等, 所以不能区分出它们. 然而通过计算给定两条轴之间的有向面积, 我们可以将二者分开. 有向面积其实就是面积加上了一个符号来表示定向, 它使用  $ad - bc$  代替了  $|ad - bc|$  并且告诉我们更多关于平行四边形的信息. 唯一的缺陷是, 我们不能自由地交换  $\alpha_1$  和  $\alpha_2$ , 因为这会改变定向. 虽然有向面积有一些缺陷, 但瑕不掩瑜, 它仍然是一个好东西. 我们将会像之前定义其它概念一样, 用几条规则来定义有向面积, 解释这些规则的合理性并且得到相关性质.

## 有向面积遵循的规则和基础性质

这几段定义了有向面积并且给出了一些解释

为了正式定义有向面积 (或称作有符号面积), 我们将会从最简单的对象: 平行四边形开始. 这里的平行四边形指的并非我们平常的平行四边形, 而是一个向量组  $\alpha$  中的  $\alpha_1$  和  $\alpha_2$  由平行四边形定则确定的四边形. 有向面积应该满足如下的性质:

有向面积以向量组  $\alpha$  作为输入, 其输出满足:

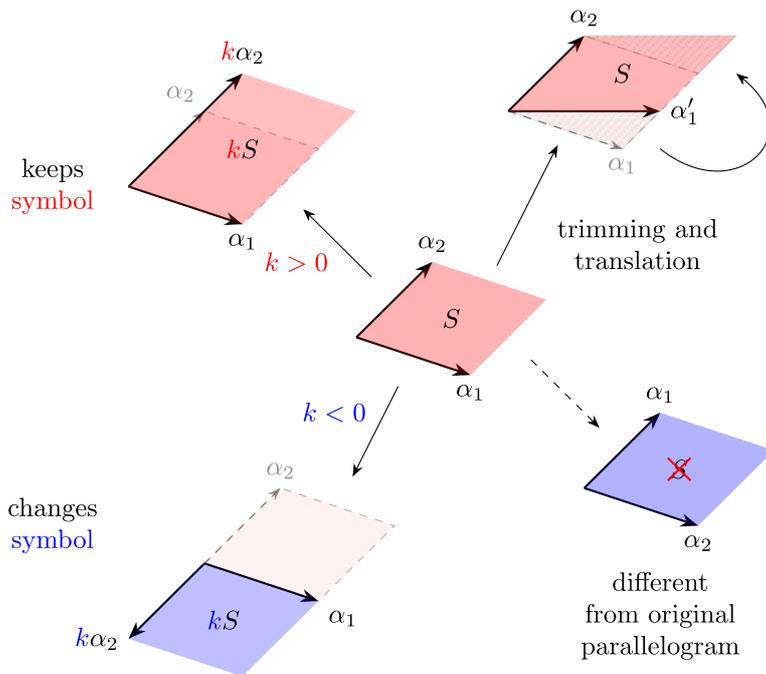
1. 把平行四边形的一条边缩放一个因子  $k$  且不改变其它边会让最终的面积变为原来的  $k$  倍, 即

$$f(k\alpha_1, \alpha_2) = f(\alpha_1, k\alpha_2) = kf(\alpha_1, \alpha_2)$$

注: 当  $k$  为负数时, 平行四边形的定向会发生改变, 所以有向面积会变号, 考虑定向有助于去掉绝对值, 用一种更加线性的方法来考虑问题

2. 如果一个平行四边形可以通过适当的切割和平移得到其它的四边形, 那么这两个四边形的面积应该是相同的, 然而我们不认为交换两个向量得到的平行四边形是相同的四边形.

如下的图展示了上面的两条规则,  $S$  是  $[\alpha_1 \quad \alpha_2]$  的有向面积:



这一段讨论了有向面积在列加减中不变

右上角的图告诉我们对常数  $k$ , 当把  $\alpha_1$  改为  $\alpha'_1 = \alpha_1 + k\alpha_2$  时, 平行四边形改变的方式可以视作把一部分切下来然后平移到另一边得到平行四边形  $[\alpha'_1 \quad \alpha_2]$ . 所以  $f(\alpha_1, \alpha_2) = f(\alpha_1 + k\alpha_2, \alpha_2)$ . 同样地, 可以证明把  $\alpha_1$  的某个倍数加到  $\alpha_2$  上时结果也成立.

接下来的段落承接上一段, 证明两个共线向量的有向面积为 0. 特别地, 这对一个向量和它自身成立.

设  $\alpha_1$  和  $\alpha_2$  共线, 例如  $\alpha_2 = k\alpha_1$ , 则

$$f(\alpha_1, \alpha_2) = f(\alpha_1, k\alpha_1) = f(\alpha_1, k\alpha_1 - k\alpha_1) = f(\alpha_1, 0) = 0f(\alpha_1, \alpha_1) = 0$$

所以对于两个共线向量, 可以推出面积为 0. 这和我们的直觉一致, 它的核不为 0, 把线性空间压缩到了一条线.

以下段落通过选取适当基底证明了有向面积的多线性

在第一个性质的注中, 我提到了这让有向面积变得线性. 所以我们会期望

$f(\alpha_1, \alpha_2 + \beta) = f(\alpha_1, \alpha_2) + f(\alpha_1, \beta)$ . 证明分为两部分, 如果  $\beta$  和  $\alpha_2$  都与  $\alpha_1$  线性相关, 则三项都为 0, 等式成立. 接下来假设其中一个与  $\alpha_1$  线性无关, 不妨设  $\alpha_2$  与  $\alpha_1$  线性无关, 则  $\alpha$  包含两个线性无关的向量, 从而成为这个二维空间的一个基. 可以用基  $\alpha$  的一个线性组合来表示  $\beta$ . 设  $\beta = x_1\alpha_1 + x_2\alpha_2$ . 则左侧的结果为

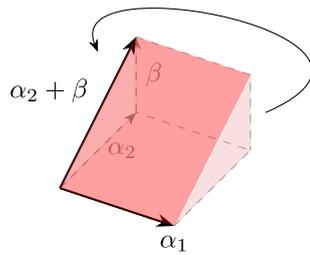
$$f(\alpha_1, \alpha_2 + \beta) = f(\alpha_1, x_1\alpha_1 + (x_2 + 1)\alpha_2) = f(\alpha_1, (x_2 + 1)\alpha_2) = (x_2 + 1)f(\alpha_1, \alpha_2)$$

右侧的结果是

$$f(\alpha_1, \alpha_2) + f(\alpha_1, x_1\alpha_1 + x_2\alpha_2) = f(\alpha_1, \alpha_2) + f(\alpha_1, x_2\alpha_2) = (x_2 + 1)f(\alpha_1, \alpha_2)$$

这一段讨论为什么有向面积在一定程度上比原始的面积有更好的性质

在以上的证明中, 我们不关系  $x_1, x_2$  或者  $x_2 + 1$  的具体符号, 因为有向面积遵循的规则支持我们直接把这些系数提出来. 这对计算来说难道不方便吗? 当  $\alpha_2$  在  $\alpha_1$  左侧时, 结果是正的. 而当  $\alpha_2$  在  $\alpha_1$  右侧时, 结果是负的. 在一定程度上, 有向面积比面积本身更加自然. 下面的图是一个图像证明 (这个图只是一个二维的图, 不要误以为它是三维的. 它只在  $x_2 > 0$  的时候才能有效论证结果成立)



## 有向面积的新符号

本段将  $u \wedge v$  作为  $f(u, v)$  的别名引入. 这就像我们使用矩阵代替  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  时做的一样. 主要目的是为了简化记号. 请不要认为它满足交换律, 否则当你看到反交换的时候会感到非常困惑.

为了方便, 从现在开始我将会用  $u \wedge v$  来表示  $f(u, v)$ . 根据前面的段落, 我们已经证明了  $f$  的多线性, 用  $\wedge$  来写就是  $(k_1 u_1 + k_2 u_2) \wedge v = k_1 u_1 \wedge v + k_2 u_2 \wedge v$  和  $u \wedge (k_1 v_1 + k_2 v_2) = k_1 u \wedge v_1 + k_2 u \wedge v_2$  (双线性). 此外我们发现把一个向量的倍数加到其它向量上并不改变结果, 这与双线性可以共同推出

$$\begin{aligned} 0 &= (u + v) \wedge (u + v) \\ &= u \wedge (u + v) + v \wedge (u + v) \\ &= u \wedge v + v \wedge u \\ &\Rightarrow v \wedge u = -u \wedge v \end{aligned}$$

你完全可以用  $f$  的记号来描述这些, 但对我来讲, 写这个会节省时间并且让它更加清晰. 既然  $u$  和  $v$  是任意的, 这表明交换两个变量位置会导致计算的结果符号改变. 这被称作反交换性, 也是上文一直强调交换两个向量不同于原本向量的原因.

这一段为有向面积找了一个单位并证明有向面积构成一维线性空间

在前面的讨论中, 我们并未讨论平行四边形面积的具体值是什么, 这是因为如果你把面积函数  $f$  乘以某个常数, 它应该满足的规则仍然成立, 所以我们只能相对于某个"单位"测量面积. 只有当你选择了单位, 例如  $\text{m}^2$ , 你才能讨论我们空间中什么东西的面积并且得到它(相对于这个单位)的值. 任意选一个平行四边形, 设它由基  $\alpha$  定义 (这意味着面积非 0, 所以可以作为一个单位), 不妨将其有向面积作为一个单位  $\alpha_1 \wedge \alpha_2$ . 设  $\beta_1 = a\alpha_1 + c\alpha_2, \beta_2 = b\alpha_1 + d\alpha_2$ , 则由  $\beta_1$  和  $\beta_2$  定义的平行四边形面积为

$$\begin{aligned} &\beta_1 \wedge \beta_2 \\ &= (a\alpha_1 + c\alpha_2) \wedge (b\alpha_1 + d\alpha_2) \\ &= a\alpha_1 \wedge (b\alpha_1 + d\alpha_2) + c\alpha_2 \wedge (b\alpha_1 + d\alpha_2) \\ &= a\alpha_1 \wedge b\alpha_1 + a\alpha_1 \wedge d\alpha_2 + c\alpha_2 \wedge b\alpha_1 + c\alpha_2 \wedge d\alpha_2 \\ &= ab(\alpha_1 \wedge \alpha_1) + ad(\alpha_1 \wedge \alpha_2) + bc(\alpha_2 \wedge \alpha_1) + cd(\alpha_2 \wedge \alpha_2) \\ &= ad(\alpha_1 \wedge \alpha_2) + bc(-\alpha_1 \wedge \alpha_2) \\ &= (ad - bc)(\alpha_1 \wedge \alpha_2) \end{aligned}$$

所以任意的有向面积都可以写成  $\alpha_1 \wedge \alpha_2$  的一个倍数, 所以有向面积的空间是一个一维向量空间, 每一个线性变换都可以用一个缩放因子表示 (即具有形式  $\varphi(\text{面积}) = k(\text{面积})$ ). 为了了解线性变换  $\varphi$  作用在有向面积上的结果, 我们只需要观察某个平行四边形  $\alpha$  的面积  $S$  并且找到改变后的面积  $\varphi(S)$ . 把它们相处就得到了系数  $k = \varphi(S)/S$ . 然后可以得到  $\varphi$  作用在有向面积上是一个正比例函数, 比例为  $k$ .

注: 一般来讲向量相除不是好的运算, 但是在一维空间中是可以的.

### 第三步 —— 定义二维的行列式

#### 行列式遵循的基本规则

注：Gabriel Cramer 在18世纪引入了行列式，用于解线性方程，英文中"determinant"和"determine"是同源词，它决定了矩阵的一些重要性质。然而中文的"行列式"可谓是一个奇怪的翻译，暂时把它接受下来就好了。

如果有映射  $\varphi: V \rightarrow V$ ，其中  $V$  是一个二维向量空间，它给出了有向面积所属线性空间上的线性映射。记  $\det \varphi = \varphi(S)/S$  并称之为  $\varphi$  的行列式。这个结果对每个非零的有向面积都相同，所以是良定义的。所以我们可以随意选取一个基底  $\alpha$  并且继续用它计算（因为基底不影响结果的选取）

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & V \\ \alpha \uparrow & & \uparrow \alpha \\ \mathbb{R}^2 & \xrightarrow{A} & \mathbb{R}^2 \end{array}$$

设  $\alpha$  是一个基， $\varphi$  具有矩阵表示  $A$ 。也就是说  $\varphi(\alpha) = \alpha A$ 。我们可以通过  $A$  在基底  $\alpha$  下的行列式为  $\det_{\alpha}(A) = \det(\varphi) = \det(\alpha A \alpha^{-1})$ 。（之后我会证明矩阵的行列式与基底无关，所以可以直接写成  $\det A$ ）。更明确地说，它是这样定义的

$$\det_{\alpha} [x \ y] = \frac{\alpha(x) \wedge \alpha(y)}{\alpha_1 \wedge \alpha_2}$$

接下来的段落将会把有向面积的结果翻译为矩阵行列式的结果并且对矩阵行列式得到相同的结果  $ad - bc$

我们已经在一些规则的基础上证明了两条重要的性质，现在我们希望得到行列式的性质，此外还要加上另外一条。

#### 1. 多线性

对于第一个变量，

$$(k_1 u_1 + k_2 u_2) \wedge v = k_1 u_1 \wedge v + k_2 u_2 \wedge v$$

如果我们把  $u$  和  $v$  用基底  $\alpha$  下的坐标来表示，我们有

$$\begin{aligned} & \alpha(k_1 x_1 + k_2 x_2) \wedge \alpha(y) \\ &= (k_1 \alpha(x_1) + k_2 \alpha(x_2)) \wedge \alpha(y) \\ &= k_1 \alpha(x_1) \wedge \alpha(y) + k_2 \alpha(x_2) \wedge \alpha(y) \end{aligned}$$

因为我们定义  $\det_{\alpha} A = \det \varphi$  为面积的比例，这意味着

$$\det_{\alpha} [x \ y] = \frac{\alpha(x) \wedge \alpha(y)}{\alpha_1 \wedge \alpha_2}$$

所以  $\alpha(x) \wedge \alpha(y) = (\det_{\alpha} [x \ y])(\alpha_1 \wedge \alpha_2)$

有向面积对第一个变量线性可以写作

$$\begin{aligned} & (\det_{\alpha} [k_1 x_1 + k_2 x_2 \ y])(\alpha_1 \wedge \alpha_2) = \\ & (k_1 \det_{\alpha} [x_1 \ y])(\alpha_1 \wedge \alpha_2) + (k_2 \det_{\alpha} [x_2 \ y])(\alpha_1 \wedge \alpha_2) \end{aligned}$$

其中  $\alpha_1 \wedge \alpha_2$  可以从等式两边消去（因为它们不共线，所以面积不为 0）

$$\det_{\alpha} [k_1 x_1 + k_2 x_2 \ y] = k_1 \det_{\alpha} [x_1 \ y] + k_2 \det_{\alpha} [x_2 \ y]$$

同理可证

$$\det_{\alpha} [x \quad k_1 y_1 + k_2 y_2] = k_1 \det_{\alpha} [x \quad y_1] + k_2 \det_{\alpha} [x \quad y_2]$$

所以行列式对每个变量线性, 但是注意:

$$\det_{\alpha} [x_1 + x_2 \quad y_1 + y_2] \neq \det_{\alpha} [x_1 \quad y_1] + \det_{\alpha} [x_2 \quad y_2]$$

对每个变量线性意味着你需要一个个地使用线性

$$\begin{aligned} \det_{\alpha} [x_1 + x_2 \quad y_1 + y_2] &= \det_{\alpha} [x_1 \quad y_1 + y_2] + \det_{\alpha} [x_2 \quad y_1 + y_2] \\ &= \det_{\alpha} [x_1 \quad y_1] + \det_{\alpha} [x_1 \quad y_2] + \det_{\alpha} [x_2 \quad y_1] + \det_{\alpha} [x_2 \quad y_2] \end{aligned}$$

如果你希望展开它, 你需要做的就像多项式乘法一样.

## 2. 反交换性.

这意味着  $v \wedge u = -u \wedge v$ . 仍然假设  $u = \alpha(x), v = \alpha(y)$ , 根据

$$\alpha(x) \wedge \alpha(y) = (\det_{\alpha} [x \quad y])(\alpha_1 \wedge \alpha_2)$$

可以得到

$$\begin{aligned} v \wedge u &= (\det_{\alpha} [y \quad x])(\alpha_1 \wedge \alpha_2) \\ &= -u \wedge v \\ &= -(\det_{\alpha} [x \quad y])(\alpha_1 \wedge \alpha_2) \end{aligned}$$

于是  $\det_{\alpha} [y \quad x] = -\det_{\alpha} [x \quad y]$

交换向量会改变定向, 在矩阵中就是说交换矩阵的两列改变行列式的符号.

如果令  $x = y$ ,  $\det_{\alpha} [x \quad x] = -\det_{\alpha} [x \quad x]$  推出一个含有重复的列的矩阵行列式为 0 (此时  $\alpha A$  中出现重复的向量, 从而线性相关)

## 3. 对单位矩阵结果为 1

虽然我们不能讨论一个面积的绝对大小, 我们至少可以说  $\alpha$  的面积等于它自己, 也就是说

$$1 = \frac{\alpha_1 \wedge \alpha_2}{\alpha_1 \wedge \alpha_2}$$

注意到  $I_2$  作为单位矩阵作用在  $\alpha$  上满足  $\alpha I_2 = \alpha$ , 所以

$$\det_{\alpha} I_2 = 1$$

有了这些结果, 我们可以纯粹地用矩阵和行列式的语言重新叙述  $ad - bc$  的结果.

$$\begin{aligned}
& \det_{\alpha} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\
&= \det_{\alpha} \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} + \det_{\alpha} \begin{bmatrix} 0 & b \\ c & d \end{bmatrix} \\
&= \det_{\alpha} \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} + \det_{\alpha} \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} + \det_{\alpha} \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} + \det_{\alpha} \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix} \\
&= \underbrace{ab \det_{\alpha} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}}_{\text{有重复的列}} + \underbrace{ad \det_{\alpha} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{单位阵}} \\
&\quad + \underbrace{bc \det_{\alpha} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\substack{\text{列交换后} \\ \text{的单位阵}}} + \underbrace{cd \det_{\alpha} \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}}_{\text{有重复的列}} \\
&= ad - bc
\end{aligned}$$

在此过程中, 我们只用到了多线性, 反交换和单位阵行列式为 1 的性质. 虽然这些性质是通过选取了某个基底来证明的, 但是结果最终和特定的基底无关. 所以实际上我们并不需要  $\det$  下的下标  $\alpha$ , 而是可以直接写出

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

所以对于一个线性变换,  $\det \varphi$  是一个不变量, 它不随基底而改变. 同样地, 对于一个矩阵,  $\det A$  也是不变量, 且等于它表示的任意线性映射的行列式. 这里"任意它表示的线性映射"是说  $A$  在基  $\alpha$  下可能表示映射  $\varphi$ , 而在基  $\beta$  下可能表示映射  $\psi$ , 这些映射可能不同, 但是它们的行列式一定相同.

#### 第四步 —— 把定义拓展到 $n$ 维

##### 把有向面积的定义推进

你还记得为什么我们选择了二维空间吗? 是的, 这是因为它相对简单, 接下来我们将会推广这个结果并且得到  $n$  维空间中的行列式. 但在此之前, 我们需要找到有向面积的替代品.

在二维空间中, 一个基的定向由第二个向量在第一个向量的左侧还是右侧决定. 假设  $\alpha_2$  在变动, 除非  $\alpha_2$  穿过  $\alpha_1$  所在的直线, 否则基的定向不会改变. 然而在加上一个维度之后我们可以轻易地翻转改变它的方向从而不需要跨过  $\alpha_1$ . 因此在 3 维空间中, 你可能已经想到了, 我们应该使用"有向体积", 这时平行四边形变成了平行六面体. 请回忆, 有向面积构成一个向量空间. 如果我们称一个有向面积是一个 2-向量, 并且称有向体积是一个 3-向量, 那么对于  $n$  维空间, 应当有对应的  $n$ -向量概念. 虽然对于我们来讲想象  $n$ -向量如何变化, 但是我们可以用 2-向量 (有向面积) 和 3-向量 (有向体积) 作为例子类比.

如果你感到困惑, 在理解下文的时候把 3-向量这个词换成"有向体积"并且把  $u \wedge v \wedge w$  换成  $f(u, v, w)$

在一个 3 维空间中, 3-向量应当遵循的规则应当包括:

1. 对每个变量线性, 即

$$(k_1 u_1 + k_2 u_2) \wedge v \wedge w = k_1 u_1 \wedge v \wedge w + k_2 u_2 \wedge v \wedge w$$

对  $v$  和  $w$  也有相同的公式

(可以这么解释: 从  $u_1 + u_2$  到平面  $v \wedge w$  的有向距离等于  $u_1$  与  $u_2$  到平面  $v \wedge w$  的有向距离的和)

2. 反交换性, 如果交换两个向量的位置, 结果改变符号.

$$u \wedge v \wedge w = -u \wedge w \wedge v = -v \wedge u \wedge w = -w \wedge v \wedge u$$

根据以上两条规则, 任何 3-向量都可以在给定的基  $\alpha$  下分解, 从而得到  $\alpha_1 \wedge \alpha_2 \wedge \alpha_3$  的某个倍数, 所以三维空间中的 3-向量组成的空间是一个一维空间.

### 定义 $n$ 维行列式

请回忆: 一个  $2 \times 2$  矩阵的行列式对每一列线性, 满足反交换律且对单位矩阵取值为 1

仍然将  $n$ -向量的比例定义为线性变换  $\varphi$  的行列式. 关于  $n \times n$  矩阵的行列式的结果仍然可以从  $n$ -向量运算的方式得出. 这次我们直接通过三条规则来定义  $n \times n$  矩阵的行列式:

矩阵的行列式是一个函数  $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ , 满足:

1. 多线性: 对每一列线性, 即对任意的  $1 \leq i \leq n$ ,

$$\det [\cdots \quad k_1 v_i + k_2 v'_i \quad \cdots] = k_1 \det [\cdots \quad v_i \quad \cdots] + k_2 \det [\cdots \quad v'_i \quad \cdots]$$

2. 反交换性: 对任意  $i \neq j$ ,

$$\det [\cdots \quad v_i \quad \cdots \quad v_j \quad \cdots] = -\det [\cdots \quad v_j \quad \cdots \quad v_i \quad \cdots]$$

3.  $\det I_n = 1$

接下来回忆我们证明过的更多关于有向面积的性质, 例如, 从一行中加或者减去另一行的若干倍, 结果不变. 也就是说如果  $C_i - kC_j$  把矩阵  $A$  变为矩阵  $B$ , 则  $\det A = \det B$ . 证明如下, 其中反交换性推出带有重复行的矩阵行列式为 0

$$\begin{aligned} & \det [\cdots \quad v_i - kv_j \quad \cdots \quad v_j \quad \cdots] \\ &= \det [\cdots \quad v_i \quad \cdots \quad v_j \quad \cdots] - k [\cdots \quad v_j \quad \cdots \quad v_j \quad \cdots] \\ &= \det [\cdots \quad v_i \quad \cdots \quad v_j \quad \cdots] - k \cdot 0 \\ &= \det [\cdots \quad v_i \quad \cdots \quad v_j \quad \cdots] \end{aligned}$$

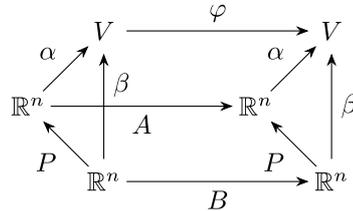
为了计算一个矩阵的行列式, 我们只需要对它做列变换把它变成单位矩阵算出结果, 又或者得到包含 0 或重复列的矩阵得到 0, 以两个三阶矩阵为例:

$$\begin{aligned} \det \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} & \xrightarrow[\frac{C_3-3C_1}{C_2-2C_1}]{} \det \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \\ 7 & -6 & -12 \end{bmatrix} \xrightarrow{C_3-2C_2} \det \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \\ 7 & -6 & 0 \end{bmatrix} = 0 \\ \det \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} & \xrightarrow[\frac{C_3-C_1/2}{C_2-C_1/2}]{} \det \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1.5 & 0.5 \\ 1 & 0.5 & 1.5 \end{bmatrix} \xrightarrow{C_3-C_2/3} \\ \det \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1.5 & 0 \\ 1 & 0.5 & 4/3 \end{bmatrix} &= 2 \cdot 1.5 \cdot (4/3) \det \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.5 & 1/3 & 1 \end{bmatrix} \xrightarrow[\frac{C_2-C_3/3}{C_1-C_3/2}]{} \\ 4 \det \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} & \xrightarrow{\frac{C_1-C_2/2}}{} 4 \det \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = 4 \end{aligned}$$

这个过程是对矩阵的列进行高斯消元. 在此过程中如果有全为 0 的列, 我们马上得到了行列式为 0. (这意味着列并非先行独立, 所以核不为 0). 如果没有出现全为 0 的列, 应该得到一个下三角矩阵. 在此之后我们只需要把对角线上的系数从每一列提出来让对角线全为 1. 接下来总是可以很容易地在不改变对角线 1 的情况下消掉对角线下方的所有元素. 所以行列式会等于这些阶梯头, 即消元后对角线上元素的乘积. (所以事实上我们并不需要关心把它变成下三角矩阵以后的其它过程, 直接把对角线上的元素相乘就可以)

## 行列式的重要性质

- 行列式在基变换下保持不变（从某种意义上这是最重要的性质，因为我们最开始的目标就是找到一个不变量）所以如果  $\alpha$  和  $\beta$  都是  $V$  的基底，而  $\varphi: V \rightarrow V$  在这两个基底下的矩阵表示分别为  $A$  和  $B$ ，则  $\det A = \det B$ 。此外，任意一个可逆矩阵都可以把基  $\alpha$  变为一个基  $\alpha P$ 。在基  $\beta = \alpha P$  下， $\varphi$  的表示应该为  $\beta^{-1}\varphi\beta = (\alpha P)^{-1}\varphi(\alpha P) = P^{-1}(\alpha^{-1}\varphi\alpha)P = P^{-1}AP$ 。这推出对任意一个矩阵  $A$  和任意可逆矩阵  $P$  都有  $\det P^{-1}AP = \det A$ 。



- 多线性、反交换性和  $\det I_n = 1$ 。这些事我们直接从  $n$ -向量（或称之为  $n$  维有向体积）推导出的性质。
- 在列加减下保持不变。
- 如果  $A$  的列的某个线性组合为 0，则  $\det A = 0$ 。换言之，此时  $\dim \text{Ker } A > 0$ ，且  $\text{rank } A < n$ 。
- 4 的逆否命题：如果  $\det A \neq 0$ ，则矩阵的列线性无关且  $\dim \text{Ker } A = 0$ 。这意味着每个等式  $y = A(x)$ ，其中  $y \in \mathbb{R}^n$  总是有唯一解，此时  $A$  是双射，存在逆矩阵。
- 4 的逆命题：如果  $\det A = 0$ ，则  $A$  中存在线性相关的行。如果  $\det A = 0$ ，则它不可能被消元到一个对角线不全为 0 的下三角矩阵，再此过程中，完全是 0 的列会出现，所以  $y = A(x)$  对有些  $y$  无解，而对有些有无穷多解。
- $\det(\varphi \circ \psi) = \det(\varphi) \det(\psi)$ 。这可以通过映射符合的缩放因子等于两次缩放因子的乘积来解释。因为缩放因子与所选的基底无关，若  $\det \psi \neq 0$ ，则  $\psi(\alpha)$  也是一个基底，所以

$$\begin{aligned} & \det(\varphi \circ \psi) \\ &= \frac{(\varphi \circ \psi)(\alpha_1) \wedge (\varphi \circ \psi)(\alpha_2) \wedge \cdots \wedge (\varphi \circ \psi)(\alpha_n)}{\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_n} \\ &= \frac{\varphi(\psi(\alpha_1)) \wedge \varphi(\psi(\alpha_2)) \wedge \cdots \wedge \varphi(\psi(\alpha_n))}{\psi(\alpha_1) \wedge \psi(\alpha_2) \wedge \cdots \wedge \psi(\alpha_n)} \frac{\psi(\alpha_1) \wedge \psi(\alpha_2) \wedge \cdots \wedge \psi(\alpha_n)}{\alpha_1 \wedge \alpha_2 \wedge \cdots \wedge \alpha_n} \\ &= \det(\varphi) \det(\psi) \end{aligned}$$

如果  $\det \psi \neq 0$ ，则  $\text{Ker}(\psi) = \{0\} \Rightarrow \text{Ker}(\varphi \circ \psi) = \{0\}$ ，所以  $\det(\varphi \circ \psi) \neq 0$ ，在这种情况下仍然成立。特别地，取  $V = \mathbb{R}^n$ ，则  $\varphi$  和  $\psi$  都是  $n \times n$  矩阵，且  $\det(AB) = \det(A) \det(B)$  对任意  $A, B \in \mathbb{R}^{n \times n}$  成立

- 排列矩阵的行列式是  $\pm 1$ 。在一个排列矩阵中，每行每列都仅包含一个 1。所以我们可以通过交换它的行来重新得到  $I_n$ 。例如：

$$\begin{aligned} \det \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix} & \stackrel{C_{12}}{=} - \det \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} \stackrel{C_{24}}{=} \\ \det \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} & \stackrel{C_{34}}{=} - \det I_n = -1 \end{aligned}$$

在每次交换过程中, 我们把第  $i$  列和第  $i$  行的 1 所在的列交换, 最终把行列式转变成  $\det I_n = 1$ , 最多相差一个符号.

### 9. $\det A = \det A^T$

请确保你还记得 PLU 分解, 并且确认自己明白  $P^T P = I_n$ . 如果你忘记了这一点又或者你不关心它是怎么证的, 你可以暂时把  $\det A = \det A^T$  当作一个既定的结果接受下来并且直接向下读. 这个性质告诉我们可以像做列变换一样做行变换. 也就是说你在不使用列变换的前提下得到一个矩阵的行列式. 此外, 教材中通常更喜欢做行变换, 好处是它让记号更统一, 操作更为一致.

证明这一点需要一些技巧. 首先先对下三角和上三角矩阵证明这个事实:

对于一个下三角矩阵, 它的特征值就是对角线上所有元素的乘积. 而对上三角矩阵, 也是这样, 我们总是可以消除对角线上方的所有元素.

$$\det \begin{bmatrix} 1 & & & \\ 1 & 4 & & \\ 5 & 1 & 4 & \\ & & & \end{bmatrix} \xrightarrow[\text{(2) } C_1 - C_2/4 - 5C_3/4]{\text{(1) } C_2 - C_3/4} \det \begin{bmatrix} 1 & & & \\ & 4 & & \\ & & 4 & \\ & & & \end{bmatrix} = 1 \cdot 4 \cdot 4 = 16$$

$$\det \begin{bmatrix} 1 & 1 & 5 & \\ & 4 & 1 & \\ & & & 4 \\ & & & \end{bmatrix} \xrightarrow[\text{(2) } C_3 - 5C_1 - C_2/4]{\text{(1) } C_2 - C_1} \det \begin{bmatrix} 1 & & & \\ & 4 & & \\ & & 4 & \\ & & & \end{bmatrix} = 1 \cdot 4 \cdot 4 = 16$$

(如果 0 出现在对角线上, 那么这个三角矩阵的秩就不是  $n$ , 从而特征值为 0)

所以对下三角和上三角矩阵, 可以写出

$$\det L = \det L^T, \det U = \det U^T$$

对于任意  $n \times n$  的矩阵, PLU 分解把它分解为

$$A = PLU \Rightarrow \det A = \det P \det L \det U$$

由于

$$A^T = U^T L^T P^T \Rightarrow \det A^T = \det U^T \det L^T \det P^T$$

且其中  $\det L = \det L^T, \det U = \det U^T$ . 所以我们只需要证明  $\det P = \det P^T$ .

由于  $P^T P = I_n$ , 这意味着  $\det P^T \det P = \det I_n = 1$ . 当  $\det P = 1$  时有  $\det P^T = 1$ . 当  $\det P = -1$  时有  $\det P^T = -1$ . 这表明  $\det P = \det P^T$ , 完成了证明.

一会我们还会见到行列式的更多性质.

### 第五步 —— 行列式的第二定义

请不要被这一节吓到, 它的目的是解释某些教材上吓人的公式. 它基本没有提供关于行列式的直觉, 而是把学生淹没在计算的海洋当中.

有些教科书甚至完全不关心行列式的不变性以及一般线性变换的行列式. 它们把一个  $n \times n$  矩阵的行列式定义为

$$\det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \sum_{j_1 \sim j_n} (-1)^{\tau(j_1, j_2, \dots, j_n)} a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

其中  $\tau(j_1, j_2, \dots, j_n) = \tau(j_1) + \tau(j_2) + \dots + \tau(j_n)$ , 而  $\tau(j_i)$  对  $i_j$  之后但是比  $i_j$  小的数计数. 你可能正在想: 这什么鬼啊? 这个  $\tau$  有意义吗? 为什么我们要把这些项加起来还要用一个巨大的跑遍所有  $j$  的求和? 它为什么加起来等于  $\det A$ . 又或者更确切地说这一大堆东西它真的有意义吗? 为了回答这些问题, 让我们回到

$$\det(\varphi) = \frac{\varphi(\alpha_1) \wedge \varphi(\alpha_2) \wedge \dots \wedge \varphi(\alpha_n)}{\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n}$$

如果  $A$  的第  $(i, j)$  个元素是  $a_{ij}$ , 且  $\varphi$  在基  $\alpha$  下的表示是  $A$ . 则  $\varphi(\alpha) = \alpha A$ , 且

$$\varphi(\alpha_1) = \sum_{i=1}^n a_{i1} \alpha_i, \varphi(\alpha_2) = \sum_{i=1}^n a_{i2} \alpha_i, \dots, \varphi(\alpha_n) = \sum_{i=1}^n a_{in} \alpha_i$$

所以

$$\det(\varphi) = \frac{(\sum_{i=1}^n a_{i1} \alpha_i) \wedge (\sum_{i=1}^n a_{i2} \alpha_i) \wedge \dots \wedge (\sum_{i=1}^n a_{in} \alpha_i)}{\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n}$$

接下来把这个公式打开并且把所有的求和提出到括号之外 (为了方便, 省去了  $1 \sim n$  的范围) 从而得到

$$\det(\varphi) = \sum_{i_1 \sim i_n} \frac{a_{i_1 1} a_{i_2 2} \dots a_{i_n n} \alpha_{i_1} \wedge \alpha_{i_2} \wedge \dots \wedge \alpha_{i_n}}{\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n}$$

如果  $i_p = i_q$ , 则  $\alpha_{i_p}$  和  $\alpha_{i_q}$  同时出现在  $\alpha_{i_1} \wedge \alpha_{i_2} \wedge \dots \wedge \alpha_{i_n}$  中, 所以这些项是 0 (因为它们并非线性无关). 只有当  $i_j$  正好取遍 1 到  $n$  中的每个值一次的时候才能得到非零值. 所以  $\alpha_{i_1} \wedge \alpha_{i_2} \wedge \dots \wedge \alpha_{i_n}$  是  $\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n$  的一个重排列. 其值为  $\pm \alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n$ . 符号由其定向决定. 不妨设  $\alpha_{i_1} \wedge \alpha_{i_2} \wedge \dots \wedge \alpha_{i_n}$  的定向为一个函数  $\varepsilon(i_1, i_2, \dots, i_n) = \frac{\alpha_{i_1} \wedge \alpha_{i_2} \wedge \dots \wedge \alpha_{i_n}}{\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n}$ . 则  $\varepsilon$  只能取 1 和  $-1$ , 我们将会在下文中讨论这个符号. 所以我们可以简化等式, 并且把它写成

$$\det A = \det(\varphi) = \sum_{i_1 \sim i_n} \varepsilon(i_1, i_2, \dots, i_n) a_{i_1 1} a_{i_2 2} \dots a_{i_n n}$$

因为  $\det A = \det A^T$ , 我们可以交换  $i$  和  $j$  并且把它等价地变成

$$\det A = \det A^T = \sum_{j_1 \sim j_n} \varepsilon(j_1, j_2, \dots, j_n) a_{1j_1} a_{2j_2} \dots a_{nj_n}$$

现在终于该解决  $\varepsilon(j_1, j_2, \dots, j_n)$  的符号了. 为了计算它, 首先看一些例子并且找规律. 最简单的非平凡的例子是, 当  $n = 2$  时,

$$\varepsilon(2, 1) = \frac{\alpha_2 \wedge \alpha_1}{\alpha_1 \wedge \alpha_2} = \frac{-\alpha_1 \wedge \alpha_2}{\alpha_1 \wedge \alpha_2} = -1$$

接下来, 可以不用  $\alpha$  来写:

$$\varepsilon(2, 1) = -\varepsilon(1, 2) = -1$$

可以用  $2 \times 2$  行列式来检验这个结果:

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \underbrace{\varepsilon(1, 2) a_{11} a_{22}}_{i_1=1, i_2=2} + \underbrace{\varepsilon(2, 1) a_{12} a_{21}}_{i_1=2, i_2=1} = a_{11} a_{22} - a_{12} a_{21}$$

如果记  $a_{11} = a, a_{12} = b, a_{21} = c, a_{22} = d$ , 它就是熟悉的  $ad - bc$ .

这里  $\varepsilon(1, 2)$  对应到  $\alpha_1 \wedge \alpha_2$ , 我们称其为标准排列. 而交换两个数字对应交换  $\alpha$  中的两个向量, 所以改变定向.

接下来计算对  $n = 3$  所有排列的结果:

$$\begin{aligned}\varepsilon(1, 2, 3) &= 1 \\ \varepsilon(1, 3, 2) &= -\varepsilon(1, 2, 3) = -1 \\ \varepsilon(2, 1, 3) &= -\varepsilon(1, 2, 3) = -1 \\ \varepsilon(2, 3, 1) &= -\varepsilon(1, 3, 2) = \varepsilon(1, 2, 3) = 1 \\ \varepsilon(3, 1, 2) &= -\varepsilon(1, 3, 2) = \varepsilon(1, 2, 3) = 1 \\ \varepsilon(3, 2, 1) &= -\varepsilon(1, 2, 3) = -1\end{aligned}$$

虽然在第一步中将 1 放在第一位, 然后在下一步中将 2 放在第二位非常方便, 但事实证明如果用这种方法, 在看到一个排列时很难预测会做多少交换 (当维数增加时这种现象尤为显著). 为了理解为什么教材上会出现逆序数, 让我们从冒泡排序开始——这是一种不高效但对理解和解决这个问题很有帮助的排序算法.

假设我们有一个序列, 这个序列是 1 到  $n$  的一个排列. 我们希望重新排列它, 使其变得有序 ( $1, 2, \dots, n$ ). 按照以下步骤进行:

从第一个元素开始, 将其与第二个元素进行比较, 如果它大于第二个元素, 则交换它们并将计数加 1. 继续处理第二个元素, 将其与第三个元素进行比较...直到我们到达最后两个元素, 然后从第一个元素重新开始. 继续这个过程, 直到序列变成有序状态.

以数列 3, 1, 4, 5, 2 为例:

1.  $3 > 1 \Rightarrow 1, 3, 4, 5, 2$ , count = 1.
- $3 < 4 \Rightarrow 1, 3, 4, 5, 2$ , count = 1
- $4 < 5 \Rightarrow 1, 3, 4, 5, 2$ , count = 1
- $5 > 2 \Rightarrow 1, 3, 4, 2, 5$ , count = 2
2.  $1 < 3 \Rightarrow 1, 3, 4, 2, 5$ , count = 2
- $3 < 4 \Rightarrow 1, 3, 4, 2, 5$ , count = 2
- $4 > 2 \Rightarrow 1, 3, 2, 4, 5$ , count = 3
3.  $1 < 3 \Rightarrow 1, 3, 2, 4, 5$ , count = 3
- $3 > 2 \Rightarrow 1, 2, 3, 4, 5$ , count = 4

共交换了四次, 所以  $\varepsilon(3, 1, 4, 2, 5) = (-1)^4 = 1$

对应的python代码是:

```
def bubble_sort(arr: list[int]) -> int:
    "返回交换的次数"
    n: int = len(arr)
    count: int = 0
    for i in range(n):
        # 为了优化算法的标志, 无交换时退出
        swapped: bool = False
        for j in range(0, n-i-1):
            if arr[j] > arr[j+1]:
                arr[j], arr[j+1] = arr[j+1], arr[j]
                count += 1
                swapped = True
```

```

        if not swapped:
            break
    return count

# 示例
index_list = [3, 1, 4, 2, 5]
exchanges = bubble_sort(index_list)
print(exchanges)

```

让我们关注一种数对：逆序对。在一个序列中，如果一个数字放在另一个数字之后，且比前面一个数字小（这两个数字并不一定直接相邻，例如 3, 1, 4, 5, 2 中的 4 和 2），我们称之为逆序对。请不要急着问为什么我们要这样做。让我们接着看这个例子：

在序列 3, 1, 4, 5, 2 中，有 4 个逆序对：(3, 1), (3, 2), (4, 2), (5, 2)。

我们在冒泡排序的过程中进行了 4 次交换：

交换 1 和 3 会去除逆序对 (3, 1)，因为它们变成了 (1, 3)，不再是逆序。

交换 2 和 5 会去除逆序对 (5, 2)，因为它们变成了 (2, 5)，不再是逆序。

类似地，其他 2 次交换分别去除了 (4, 2) 和 (3, 2)。

事实上，我们共进行了 4 次交换，每次去除一个逆序对。这并非巧合。注意到我们每次只交换相邻的数字。这只是交换了它们的相对位置，但它们相对于其他数字的位置并没有改变。因此，一次交换对应一个逆序对，通过计数逆序对，我们就能知道进行了多少次交换。

最终，记  $\tau(j_i)$  为出现在  $j_i$  之后但是比  $j_i$  小的数的计数（严格意义上应该写作  $\tau_{j_1, j_2, \dots, j_n}(j_i)$ ，因为结果不但依赖于  $j_i$  自身，还和所有的  $j_1 \sim j_n$  有关）。所以  $\tau(j_i)$  计数了所有前一个元素是  $j_i$  的逆序对。把它对  $1 \sim n$  求和，总的逆序数应该为  $\tau(j_1, j_2, \dots, j_n) = \tau(j_1) + \tau(j_2) + \dots + \tau(j_n)$ 。由于逆序数计数了通过冒泡排序将数列排序所需的交换数， $\tau(j_1, j_2, \dots, j_n)$  的奇偶性决定了  $\varepsilon(j_1, j_2, \dots, j_n)$ ，其中  $\varepsilon = (-1)^\tau$ 。最终我们可以清楚地得到结果：

$$\det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \sum_{j_1 \sim j_n} (-1)^{\tau(j_1, j_2, \dots, j_n)} a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

这个公式是通过把  $n$ -向量的所有分量展开得到的，因为这个定义并不依赖于线性变换，并且定义行列式的方式更加“明确”，所以包括我们教科书在内的一些人更喜欢用这个定义，因为它们认为这更直接。然而，这个式子的形式意义大于实际意义，此外使用它计算大于  $3 \times 3$  的矩阵行列式非常困难。就我个人来讲，我更喜欢把这个式子看作  $n$  维有向体积的和，这种几何直观对我理解并记住它起到了很大帮助。

### 代数余子式

这一小节只是为了解释我们教材中的代数余子式，然而它对理解线性映射的本质并无太大影响，所以如果你在此过程中感到枯燥，不妨跳过。

我们知道一个矩阵的行列式对每列线性，所以如果只让一列变化，同时固定其它  $n - 1$  列，就得到了一个线性映射，把这一列线性映射到行列式，即一个数，所以它是  $\mathbb{R}^n \rightarrow \mathbb{R}$  的。假设矩阵的第  $2 \sim n$  列都确定了，那么只有  $a_{i1}$  这些项可以改变，映射具有形式

$$\det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = c_1 a_{11} + c_2 a_{21} + \cdots + c_n a_{n1}$$

这里的  $c_i$  是与  $A$  的第一列无关的常数, 但是它们会随着其它列的改变而改变. 当把其它列都定下来之后它们就成为了常数.

接下来让我们为  $c_i$  找到显式的表达式

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} &= \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \cdots & a_{nn} \end{bmatrix} + \det \begin{bmatrix} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \cdots & a_{nn} \end{bmatrix} \\ &+ \cdots + \det \begin{bmatrix} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \\ &= a_{11} \det \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \cdots & a_{nn} \end{bmatrix} + a_{21} \det \begin{bmatrix} & a_{12} & \cdots & a_{1n} \\ 1 & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \cdots & a_{nn} \end{bmatrix} \\ &+ \cdots + \det \begin{bmatrix} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ 1 & a_{n2} & \cdots & a_{nn} \end{bmatrix} \end{aligned}$$

所以

$$c_1 = \det \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \det \begin{bmatrix} 1 & & & \\ & a_{22} & \cdots & a_{2n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

显然, 右侧的行列式等于

$$\det \begin{bmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

(这时因为对前面的行列式高斯消元不改变第一行第一列, 所以行列式等于右下角的更小的块的行列式乘以对角线上的 1, 它们是一样的)

然而为了计算  $c_2$ , 我们有必要把第二行换到第一行并且得到

$$c_2 = -\det \begin{bmatrix} 1 & & & & \\ & a_{12} & a_{13} & \cdots & a_{1n} \\ & a_{32} & a_{33} & \cdots & a_{3n} \\ & \vdots & \vdots & \ddots & \vdots \\ & a_{n2} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = -\det \begin{bmatrix} a_{12} & a_{13} & \cdots & a_{1n} \\ a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

为了计算  $c_3$ , 首先交换第二和第三行, 然后交换第一和第二行. 这样让  $a_3$  浮到了顶端, 而同时没有改变其它行之间的相对位置. 这里进行了两次交换, 所以前面的符号又变为了正的.

不难发现这样的规律

$$c_i = (-1)^{i-1} \det \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{i-1,2} & a_{i-1,3} & \cdots & a_{i-1,n} \\ a_{i+1,2} & a_{i+1,3} & \cdots & a_{i+1,n} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

这里的  $(-1)^{i-1}$  意味着进行了  $i-1$  次的行交换, 矩阵原本的第  $i$  行和第一列被删除了.

当我们想让第  $j$  列变化而固定其它的列时, 我们发现对应的因子和第一列的类似

$$c_{ij} = (-1)^{i+j-2} \det \begin{bmatrix} \ddots & \vdots & \vdots & \ddots \\ \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots \\ \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots \\ \ddots & \vdots & \vdots & \ddots \end{bmatrix}$$

$(-1)^{i+j-2}$  可以被写为  $(-1)^{i+j}$ . 但是此处我写成  $(-1)^{i+j-2}$  的原因是为了把第  $(i, j)$  个元素换到左上角而不改变其它行和列之间的相对位置, 需要经过  $(-1)^{i+j-2}$  次交换. 在现在的矩阵中, 第  $i$  行和第  $j$  列被删除了. 这个系数  $c_{ij}$  被称作  $a_{ij}$  的代数余子式, 对每个  $1 \leq j \leq n$ , 有

$$\det A = \sum_{i=1}^n a_{ij} c_{ij}$$

这是因为固定下除了第  $j$  列之外的其它列, 行列式变成了第  $j$  列的线性函数, 其系数分别为  $c_{ij}$ . 我们称其为行列式的列展开.

类似的讨论可以导出行列式的行展开, 对每个  $1 \leq i \leq n$ :

$$\det A = \sum_{j=1}^n a_{ij} c_{ij}$$

## 线性方程 —— 行列式的另一个角度

### Cramer 法则

先回忆一个矩阵  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  的行为. 当  $m > n$  时,  $A$  不可能是满射, 总是有一些  $y$  让  $y = Ax$  无解. 而对  $n > m$ ,  $A$  不可能是单射, 每个值域中的  $y$  都对应到很多的  $x$ . 所以  $m = n$  是最"有趣"的情况, 此时  $n$  个方程对应  $n$  个未知数.

我们知道对  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , 如果  $\det A = 0$ , 那么消元会在主对角线上产生 0, 此时就有了自由变量, 于是  $A$  不可逆. 但是对  $\det A \neq 0$ ,  $\dim \text{Ker } A = 0$ . 线性代数基本定理告诉我们  $\dim \text{Im } A = n - \dim \text{Ker } A = n$ . 这意味着像是  $\mathbb{R}^n$ . 在本节中, 我们将会为这样的线性方程的解写出显式的表达式.

把一个线性方程写成如下形式, 其中  $\alpha_1 \sim \alpha_n$  为列向量:

$$y = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$$

例如, 为了解  $x_1$ , 我们应该消掉所有其它的像. 神奇的是, 我们可以通过计算行列式完成这一过程.

$$\det(y, \alpha_2, \cdots, \alpha_n) = \det(x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n, \alpha_2, \cdots, \alpha_n)$$

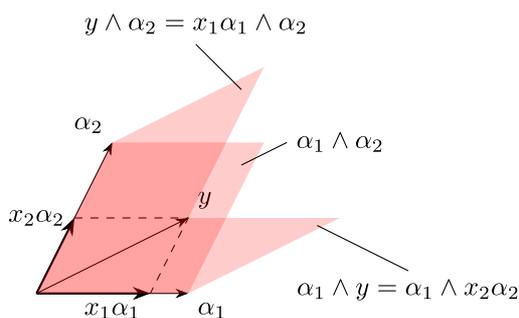
注意到列相减不改变行列式的值, 所以

$$\det(y, \alpha_2, \cdots, \alpha_n) = \det(x_1\alpha_1, \alpha_2, \cdots, \alpha_n) = x_1 \det(\alpha_1, \alpha_2, \cdots, \alpha_n)$$

至少从形式上, 我们可以从  $y_1$  直接解出  $x_1$ . (类似地对于  $x_i$  的情况, 分子上的  $y$  替换的是  $\alpha_i$ )

$$x_1 = \frac{\det(y, \alpha_2, \cdots, \alpha_n)}{\det(\alpha_1, \alpha_2, \cdots, \alpha_n)}$$

分母恰好就是  $\det A$ . 这就是 Cramer 法则, 下图展示了  $n = 2$  时 Cramer 法则为什么合理.



## 线性方程的判别式

本小节讨论的是我个人关于行列式对于线性方程作用的理解.

大部分人对判别式的第一反应可能是  $\Delta = b^2 - 4ac$ . 但是这只是判别式的一部分, 判别式实际上可以对任意多项式  $p(x)$ , 甚至可以对多变量多项式定义. 判别式反映的是  $p(x) = 0$  的根的情况. 对于一个二次函数  $ax^2 + bx + c = 0$ , 它"倾向于"拥有两个复根, 只有当  $\Delta = b^2 - 4ac = 0$  时才会出现例外. 而对一个方阵  $A$ , 如果你随意地在  $\mathbb{R}$  中取它的系数, 它的行列式几乎总是非 0 的. 所以它几乎总是可逆, 此时  $Ax = 0$  有唯一解  $x = 0$ . 但是如果  $\det A$  恰好为 0, 则  $Ax = 0$  有无穷多解. 这意味着此时出现了"重根", 所以判别式可以视作是一种类型的判别式, 它显示出是否有重根.

## 伴随矩阵

本小节只是为了解释课本中的伴随矩阵概念, 它对理解线性映射的本质并无太大影响. 所以如果你感到无聊, 可以直接跳过它.

我们知道对一个  $\det A \neq 0$  的线性方程  $y = Ax$ , 有

$$x_1 = \frac{\det(y, x_2, \cdots, x_n)}{\det A}$$

我们可以用代数余子式把这个式子展开:

$$x_1 = (\det A)^{-1} \sum_{i=1}^n y_i c_{i1} = (\det A)^{-1} [c_{11} \quad c_{21} \quad \cdots \quad c_{n1}]y$$

类似地

$$x_j = (\det A)^{-1} \sum_{i=1}^n y_i c_{ij} = (\det A)^{-1} [c_{1j} \quad c_{2j} \quad \cdots \quad c_{nj}]y$$

实际上在这个过程中我们已经用  $y$  解出了  $x$  :

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (\det A)^{-1} \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ c_{12} & c_{22} & \cdots & c_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix} y$$

于是我们得到了一个把  $A^{-1}$  用余子式表示的显式表达式.

$$A^{-1} = (\det A)^{-1} \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ c_{12} & c_{22} & \cdots & c_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix}$$

这里有  $(A^{-1})_{ji} = (\det A)^{-1} c_{ij}$  (注意此处下标的顺序). 右侧的矩阵被称作伴随矩阵, 记作  $\text{adj } A$  (或  $A^*$ ). 对于  $\det A \neq 0$ ,  $A^{-1} = (\det A)^{-1} \text{adj } A$ , 这可以推出

$$A \text{adj } A = (\text{adj } A) A = (\det A) I_n$$

然而这个矩阵乘以其伴随的性质并不依赖于  $\det A \neq 0$ . 由余子式的定义, 容易验证  $A \text{adj } A = (\det A) I_n = 0$  对  $\det A = 0$  成立, 所以这个式子对所有方阵  $A$  成立.

## 特征值和特征向量

本章中我们将会在  $\mathbb{C}$  上讨论, 因为一个  $n$  次的多项式在  $\mathbb{C}$  中总有  $n$  个根 (计重数)

### 最初的想法

先回忆, 当我们希望比较一个变化前和变化后的向量时我们需要找到一些向量, 满足线性变换作用在它们上时表现的像一个伸缩. 这给了我们另外一种看待线性变换的角度并且可以帮助我们找到一个好的基底来表示它. 线性变换像一个伸缩一样意味着对某个非零向量  $v$ , 有  $\varphi(v) = \lambda v$ . 注意到  $\lambda$  倍的恒同映射总是把  $v$  映射到  $(\lambda \text{Id})(v) = \lambda v$ . 所以只需要把两个式子详见, 就存在某些非零向量满足

$$(\lambda \text{Id} - \varphi)(v) = 0$$

这样一组  $(\lambda, v)$  被称作特征值和一个对应的特征向量. 对有限维空间而言, 这意味着  $\det(\lambda \text{Id} - \varphi) = 0$ . 这是关于  $\lambda$  的多项式方程, 我们可以先解出  $\lambda$ , 然后通过解  $(\lambda \text{Id} - \varphi)(v) = 0$  得到  $v$ .

更具体地说, 首先选一个基并写出  $\varphi$  的矩阵表示  $A$ . 则  $\lambda \text{Id} - \varphi$  的矩阵表示为  $\lambda I_n - A$ . 它们的行列式相同. 然后只需要解出  $(\lambda I_n - A)x = 0$ , 就可以找到所有特征向量的坐标, 例如, 假设

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

为了找到  $\lambda$  和  $x$  使  $Ax = \lambda x$ . 先计算

$$\det(\lambda I_2 - A) = \det\left(\lambda \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} - \begin{bmatrix} 3 & 1 \\ 6 & 4 \end{bmatrix}\right) = \det\left(\begin{bmatrix} \lambda - 3 & 1 \\ 6 & \lambda - 4 \end{bmatrix}\right)$$

$$= (\lambda - 3)(\lambda - 4) - 1 \cdot 6 = \lambda^2 - 7\lambda + 12 - 6 = (\lambda - 1)(\lambda - 6)$$

多项式  $p(\lambda) = (\lambda - 1)(\lambda - 6)$  称作特征多项式,  $p(\lambda) = 0$  被称作特征方程. 解特征方程可得  $\lambda_1 = 1, \lambda_2 = 6$ . 对  $\lambda_1 = 1$ , 有  $(\lambda_1 I_2 - A)x = 0 \Rightarrow x = t \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ . 取  $u_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$  为一个特征向量. 对  $\lambda_2 = 6$ ,  $(\lambda_2 I_2 - A)x = 0 \Rightarrow x = t \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ , 取  $u_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$  为另一个特征向量, 则对所有的  $x = tu_1$ ,  $Ax = \lambda_1 x$ , 而对所有的  $x = tu_2$ ,  $Ax = \lambda_2 x$ . 所以我们找到了所有的特征值和特征向量.

对于一个  $n$  维空间,  $\deg p = n$ . 代数基本定理告诉我们一个  $n$  次多项式总可以被因式分解为  $p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$ , 这里每个  $\lambda_i$  都是一个复数.

### 没有重根的情况

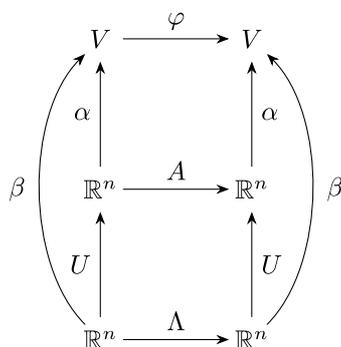
最一般的情况是多项式没有重根, 此时对  $i \neq j$  有  $\lambda_i \neq \lambda_j$ . 这给出我们  $n$  个特征值, 所以可以从  $(\lambda_i I_n - A)x = 0$  中解出  $n$  个特征向量  $u_1 \sim u_n$ . 矩阵可以写作如下形式:

$$A [u_1 \quad u_2 \quad \cdots \quad u_n] = [u_1 \quad u_2 \quad \cdots \quad u_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

可以通过归纳法证明  $u_1 \sim u_n$  总是先行独立的, 所以  $[u_1 \quad u_2 \quad \cdots \quad u_n]$  作为过渡矩阵 (基变换矩阵) 把原先选定的基底  $\alpha$  变成一个新的基底  $\beta$ , 使得基  $\beta$  下的矩阵表示是一个对角矩阵.

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

这一过程称作"对角化". 这对没有重根出现的情况总是适用. 注意到行列式不依赖于基变化, 所以特征多项式实际上是一个不变量, 从中解出的特征值也是不变量, 这会把每个基底都变到一个只与线性变换  $\varphi$  有关的基  $\beta$  (最多相差一个特征向量乘的系数, 因为特征向量的倍数也是特征向量)



### 有重根的情况

在讨论之前, 我们先定义一个叫特征子空间的概念. 对于向量空间  $V$  上线性变换  $\varphi$  的某个特征值  $\lambda$ . 称对应的特征子空间为  $V_\lambda = \{v \in V | \varphi(v) = \lambda v\}$ . 这确实是一个子空间, 因为它可以写成  $V_\lambda = \text{Ker}(\lambda \text{Id} - \varphi)$ . 它包含了所有对应于特征值  $\lambda$  的特征向量以及 0 向量.

不幸的是,当重根出现时,不一定总能找到一组特征向量组把它们作为基.这是因为  $\det(\lambda \text{Id} - \varphi) = 0$  只能确保存在特征向量  $v$  满足  $\varphi(v) = \lambda v$ ,也就是说对应的特征子空间维数至少是 1 维的.但是当重根出现时,不能保证重数一定等于子空间的维数.例如,矩阵  $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$  的特征方程是  $\lambda^2 = 0 \Rightarrow \lambda = 0$ .然而方程  $Ax = 0$  的解只有  $x = \begin{bmatrix} t \\ 0 \end{bmatrix}$ .这意味着子空间是 1 维的,然而特征值  $\lambda = 0$  作为根的重数是 2.这意味着没有足够的向量构成一个基,矩阵因此不可对角化.

也有时即使出现了特征值,但是矩阵仍然可以对角化.例如,所有的非零向量都是恒同映射  $\text{Id}$  特征值为 1 的特征向量.所以  $\lambda = 1$  对应的子空间是整个空间.设空间维数为  $n$ .由于没有其它的特征值,所以对应的特征多项式是  $p(\lambda) = (\lambda - 1)^n$ .任意取一个  $V$  的基底,则有  $n$  个基向量,这个线性映射可以被对角化为一个对角线上全为 1 而其它位置全为 0 的矩阵(即单位矩阵)

称特征方程中特征值  $\lambda$  的重数为代数重数,而特征子空间维数  $\dim V_\lambda$  为几何重数.可以给出估计  $1 \leq \text{几何重数} \leq \text{代数重数}$ .如果没有重根,则所有的代数重数都是 1,由不等式可知必须有每个特征值对应的特征子空间维数都为 1.而当出现重根时,代数重数大于 1.只有当所有的几何重数都等于代数重数时这个线性变换才能对角化.

## 练习

5.1 注意到行变换和列变换都是矩阵空间上的线性变换,先验证下面 4 个矩阵组成  $\mathbb{R}^{2 \times 2}$  的一个基

$$e_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, e_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, e_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, e_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

如果记  $T: \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}^{2 \times 2}$  为将第二行减掉两倍的的第一行的变换,它在基  $e_{11}, e_{12}, e_{21}, e_{22}$  下的矩阵表示是什么?

5.2 先验证所有满足  $a_{n+2} = a_n + a_{n+1}$  的数列构成线性空间,并且证明这两个数列

$s_1: 1, 0, 1, 1, 2, 3, \dots$  和  $s_2: 0, 1, 1, 2, 3, 5, \dots$  构成线性空间的一个基.由于去除前  $t$  项的操作  $R_t$  是这个空间上的线性映射,写出  $R_1, R_2, R_3$  的矩阵表示  $A_1, A_2, A_3$ .并且叙述  $A_1, A_2, A_3$  之间有何联系.

5.3 计算  $\mathbb{R}^2$  中顶点为  $(0, 0), (3, 1), (9, 5), (6, 4)$  的平行四边形通常的(即不包含定向的)面积.

5.4 使用  $(u - v) \wedge (u - v) = 0$  证明  $u \wedge v = -v \wedge u$ .

5.5 如果  $[\beta_1 \ \beta_2 \ \beta_3] = [\alpha_1 \ \alpha_2 \ \alpha_3]A$ ,也就是说

$$\beta_1 = a_{11}\alpha_1 + a_{21}\alpha_2 + a_{31}\alpha_3$$

$$\beta_2 = a_{12}\alpha_1 + a_{22}\alpha_2 + a_{32}\alpha_3$$

$$\beta_3 = a_{13}\alpha_1 + a_{23}\alpha_2 + a_{33}\alpha_3$$

把  $\beta_1 \wedge \beta_2 \wedge \beta_3$  的表达式展开并把它用  $\alpha_1 \wedge \alpha_2 \wedge \alpha_3$  表示.观察其中有多少非零的项,并把它们的符号和所有三个数字排列的定向比较.

5.6 分别通过消元和你在 5.5 中得到的结果计算如下行列式:

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

5.7 由于  $\det A = \det A^T$ .对正交矩阵  $Q$ ,证明  $|\det Q| = 1$ .

5.8 计算  $\varepsilon(3, 1, 5, 2, 4)$  并计算下列矩阵的行列式:

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

5.9 计算如下的行列式, 结果会是关于向量  $x$  的线性函数,  $x$  的每个分量前的系数分别是什么? 如果使用  $\mathbb{R}^3$  上的标准内积, 它可以表示成哪个向量和  $x$  的内积, 结果何时为 0?

$$\det \begin{bmatrix} x_1 & 1 & 4 \\ x_2 & 2 & 5 \\ x_3 & 3 & 6 \end{bmatrix}$$

5.10 分别使用 Cramer 法则和消元解下列方程:

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

5.11 为  $\mathbb{R}^2$  找一个基  $U$  以对角化  $A = \begin{bmatrix} 1 & 1 \\ & 2 \end{bmatrix}$ . 则  $AU = U\Lambda \Rightarrow A = U\Lambda U^{-1}$ . 根据  $A = U\Lambda U^{-1}$  计算  $A^{127}$ , 你可以从中体会到对角化为什么让计算矩阵的幂更加方便.

5.12 证明如果一个线性映射  $\varphi$  可对角化, 则  $\det \varphi = \lambda_1 \lambda_2 \cdots \lambda_n$ . (提示: 行列式无关于基底选取)

5.13 证明一个方阵不可逆当且仅当它有 0 特征值, 为什么这个结果成立.

5.14 如果  $\lambda$  是  $\varphi$  的一个特征值, 证明  $\lambda^2$  是  $\varphi^2 = \varphi \circ \varphi$  的一个特征值, 且  $v$  仍然是  $\varphi^2$  的一个特征向量.

5.15 如果  $p(\lambda)$  是某个可对角化矩阵  $A$  的特征多项式, 根据  $A = U\Lambda U^{-1}$  证明  $p(A) = 0$ . 这里  $p(A)$  的定义是  $p(A) = p_0 I_n + p_1 A + p_2 A^2 + \cdots + p_{n-1} A^{n-1} + p_n A^n$ . 齐总  $p_i$  是  $\lambda^i$  项的系数

5.16 如果  $A \in \mathbb{R}^{m \times m}, B \in \mathbb{R}^{n \times n}, C \in \mathbb{R}^{m \times n}$ . 证明  $\det \begin{bmatrix} A & C \\ & B \end{bmatrix} = \det A \cdot \det B$ . (提示: 把问题分成两半:  $\det A = 0$  与  $\det A \neq 0$ )

5.17  $A = \begin{bmatrix} 1 & 1 & \\ & 1 & 1 \\ & & 1 \end{bmatrix}$  可对角化吗? 计算  $A^2, A^3, A^4$  并猜想  $A^k$  的结果.

5.18 证明除了恒同映射以外, 没有线性映射可以被相似对角化为单位矩阵.