工科线性代数讲义

刘余
Yu Liu
yliu2018@mail.tsinghua.edu.cn

声明

这个讲义禁止传播. 另外, 这个讲义有很多从纯粹数学的角度来看很不严格的地方; 因此, 请提高自己的鉴别能力. 然而, 这个讲义的初衷仅仅是课堂讲授之外的一点点补充. 我试图用一些直观来代替严密的数学逻辑, 使得某些内容变得可读性强一些. 我的初衷就是展现一种我个人认为理解线性代数的方式, 对我本人而言, 虽然有逻辑上的缺陷, 但是在理解上是顺畅的自然的.

总之, 学习请以教材为准; 对于阅读这个讲义而带来的任何谬误, 本人不承担任何责任.

目录

第零章　引言—-几个实例

例题 0.1. 我们如下定义斐波那契数列：$a_1 = a_2 = 1, a_{n+2} = a_{n+1} + a_n$. 求 $\{a_n\}$ 的通项公式。

例题 0.2 (多输入多输出问题). 多输入多输出是一种利用发射端的多个天线各自独立发送信号，同时在接收端用多个天线接收并复原信息的技术。我们假定发射端有三个天线, 其发射的信号为 x_1, x_2, x_3, 同时接收端也有三个天线，其接收的信号为 y_1, y_2, y_3。已知发送端信号和接收端信号有如下关系：

$$y_i = h_{i1}x_1 + h_{i2}x_2 + h_{i3}x_3.$$

由于任何接收天线的信号中都含有 3 个发送端的信号，处理比较麻烦。但是，如果 $i \neq j$ 时 $h_{ij} = 0$，那么接收信号便成为 $y_i = h_{ii}x_i$. 这看起来像是“单向通信”。但是通常办不到。那怎么处理呢？通过奇异值分解，我们知道

$$H = USV$$

其中 U, V 是酉矩阵，S 是对角矩阵。那么在发送端用 V^* 编码，在接收端用 U^* 解码，那么传输过程看起来就像是用 S 作了“单向通信”。

例题 0.3 (图像压缩存储). 假设一张图片有一亿像素，那么计算机存储这一亿像素的方式是：对每个像素点编号，每个像素点赋值以表示其颜色，那么我们就得到了一个 10^8 个数组。这是最朴素的存储方式，好处是保留了图片的原始信息，坏处是占用空间大。如果我们换一种看法，假设这一亿个像素 $10^4 \times 10^4$ 的格点，那么我们在每个格点处赋值，得到了一个大小为 $10^4 \times 10^4$ 的矩阵，至此，问题并没有任何改变。但是，我们需要注意到，所有的像素点不是孤立存在的，它和它附近的格点处的颜色变化会有规律。因此，如果我们把这种规律能抽象出来，用矩阵语言叙述，那么我们只需要存储这种规律以及若干个像素点的信息，即可压缩图片存储。这样的好处是占用空间小，坏处是图片信息有所损失。

例题 0.4 (页面排序). 假设互联网上有无数个页面。这些页面的重要程度显然不是一致的，那么我们怎么对这些页面按重要性作排序？

假设幼儿园老师给三位小朋友小红，小明和小华各发了 6 个糖果。为了培养他们的分享精神，要求他们将自己手中的糖果分给其余的两位小朋友。小红喜欢小明，因此她毫无犹豫地把自己的糖果全都给了小明；而小明觉得小红和小华都很好，但是更偏向小华，因此把 1/3 给了小红，2/3 给了小华。小华觉得小红和小明都是自己的好朋友，要一视同仁，因此把自己的糖果平均分给小红和小明。结束后，小红有 5 个糖果，小明 9 个，小华 4 个。

如果我们让三位小朋友按照他们刚才的原则，一轮一轮分，那么最终他们各自有几个糖果？设第 n 轮后，小红，小明，小华各有 x_n, y_n, z_n 个糖果，那么

$$\begin{cases} x_{n+1} &= 0\cdot x_n + \frac{1}{3}\cdot y_n + \frac{1}{2}\cdot z_n \\ y_{n+1} &= 1\cdot x_n + 0\cdot y_n + \frac{1}{2}\cdot z_n \\ z_{n+1} &= 0\cdot x_n + \frac{2}{3}\cdot y_n + 0\cdot z_n \end{cases}$$

那么 x_n, y_n, z_n 就反映了小红，小明，小华的受欢迎程度

现在，我们将这些页面编号：$\{1,2,3,\cdots,i\cdots\}$。设 x_n^i 是网民浏览了 n 次网页之后浏览到了第 i 个页面的概率。第 i 个网页有 N_i 个超链接指向其他网页，定义

$$P_{ij} = \begin{cases} \frac{1}{N_i} & \text{从页面 i 有指向页面 j 的超链接} \\ 0 & \text{否则} \end{cases}$$

因此 $x_{n+1}^i = \sum_j P_{ji} x_n^j$. 那么 $\{x_n^i\}_{n\geq 1}$ 的极限就表示了第 i 个页面的重要程度。

第零章　预备知识

0.1　集合论基础

注释 0.1. 这一节关于集合论的叙述, 基本节选自清华大学出版社出版, 张贤科编著的" 高等代数" 附录, 少许增删和修改.

如果想要直接而精确地定义集合, 这不是一件容易的事情; 我们暂且就简单地这样理解: 所谓一个**集合**, 就是被明确地指定的一些对象的全体; 其中的每个对象被称为集合中的**元素**或**元**. 当 a 是集合 S 中的元素时, 我们称 a 属于 S, 或者 S 包含 a, 记作 $a \in S$ 或 $S \ni a$. 不含任何元素的集合称为**空集**, 记为 $\emptyset$.

我们表示一个集合通常有两种方式: 罗列和描述. 前者例如 $\{1, 2, 3, 4, \cdots\}$, 后者例如 $\{x : x \text{ 是正整数}\}$. 最低标准就是: " 明确地指定" 集合中的所有元素.

定义 0.2. 集合 B 称为是集合 A 的子集, 如果集合 B 的元素均属于 A, 记作 $B \subseteq A$ 或者 $A \supseteq B$. 若 B 不是 A 的子集, 我们记作 $B \not\subseteq A$. 若 $B \subseteq A$ 且 $A \not\subseteq B$, 我们称 B 是 A 的真子集.

例题 0.3. 例如, 自然数集 $\mathbb{N}$ 是整数集 $\mathbb{Z}$ 的真子集, 整数集 $\mathbb{Z}$ 是有理数集 $\mathbb{Q}$ 的真子集; 有理数集 $\mathbb{Q}$ 是实数集 $\mathbb{R}$ 的真子集; 实数集 $\mathbb{R}$ 是复数集 $\mathbb{C}$ 的真子集.

注释 0.4. 显然: 若 $C \subseteq B$ 且 $B \subseteq A$, 则 $C \subseteq A$. 若 $A = B$ 当且仅当 $B \subseteq A$ 且 $A \subseteq B$.

定义 0.5. 设 A, B 是两个集合,

- 由属于 A 或 B 的元素全体构成的集合称为 A, B 的并集, 记作 $A \cup B$;
- 由同时属于 A 和 B 的元素全体构成的集合称为 A, B 的交集, 记作 $A \cap B$;
- 由属于 A 但是不属于 B 的元素全体构成的集合称为 A 和 B 的差集, 记作 $A - B$.

注释 0.6. 我们有 $A \cup B = B \cup A, A \cap B = B \cap A$; 但是 $A - B \neq B - A$, 也就是说, A 和 B 的差集一般不等于 B 和 A 的差集.

例题 0.7. 设 $A=(0,2], B=(1,3]$, 那么

$$A\cup B=(1,3];\quad A\cap B=(1,2];\quad A-B=(0,1].$$

固定一个集合 S, 考虑 S 的元素及子集. 对 S 的子集 A, 我们记 $\overline{A}=S-A$. 那么

- $(A\cup B)\cup C=A\cup(B\cup C), (A\cap B)\cap C=A\cap(B\cap C)$.
- $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$; $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$.
- $A\cup(A\cap B)=A$, $A\cap(A\cup B)=A$;
- $\overline{A\cup B}=\overline{A}\cap\overline{B}$; $\overline{A\cap B}=\overline{A}\cup\overline{B}$.

定义 0.8. 集合 A,B 的**直积**或**笛卡尔积**定义为

$$A\times B=\{(a,b):a\in A,b\in B\}.$$

注释 0.9. 这里 (a,b) 是 a 和 b 的有序对, 即 $(a,b)=(c,d)$ 当且仅当 $a=b,c=d$. 类似的, 我们可以定义 (或者归纳地定义) 有点个集合的直积.

从集合 A 到集合 B 的一个**映射**是指一个规则, 这个规则对 A 的每一个元素都**唯一**地指定了 B 中的一个元素. 映射有时也被称为**函数** (特别是当 B 是 $\mathbb{C}$ 的子集时), 有时也被称为**变换** (特别是当 $B=A$ 时). 从 A 到 B 的映射 f 通常记为 $f:A\to B$ 或者 $A\xrightarrow{f}B$. 对 $a\in A$, 若 f 指定了 $b\in B$ 与之对应时, 写为 $f(a)=b$ 或 $f:a\mapsto b$. b 称为 a 的**像**. A 中所有元素的像构成的全体称为 f 的**像集**, 记作 $\mathrm{Im}(f)$, 也就是说

$$\mathrm{Im}(f)=f(A):=\{f(a):a\in A\}.$$

它是 B 的子集, 但未必等于 B. 若 $\mathrm{Im}(f)=B$, 我们称 f 是**满射**. 等价地说

命题 0.10. 映射 $f:A\to B$ 是满射, 当且仅当对任意的 $b\in B$, 存在 $a\in A$ 使得 $f(a)=b$.

若 $f(a)=b$, 我们称 a 是 b 的一个原像, b 的原像的全体记作 $f^{-1}(b)$, 即 $f^{-1}(b)=\{a\in A:f(a)=b\}$. 所以上面的命题可以这么说 "映射 $f:A\to B$ 是满射, 当且仅当 B 中任何元素都有原像".

A 称为 f 的定义域. 从 A 到 B 的两个映射 f,g 相等 (记作 $f=g$) 是指对任何 $a\in A$ 都有 $f(a)=g(a)$.

注释 0.11. 在上述映射的定义中要注意:

1. 必须对 A 中的每个元素均指定像 (而并不要求 B 中每个元素都是像); 例如: 我们用 $\mathbb{R}_+$ 来表示正实数, $f:\mathbb{R}_+\to\mathbb{R}, x\mapsto\sqrt{x}$; 这是一个映射, 但是非正实数不是像.

2. A 中的每个元素都只能指定唯一的像, 即 f 是 "单值" 的 (但是对 $b \in B$, 其原像并不要求是唯一的). 例如 $\mathbb{R}_+ \to \mathbb{R}$, 对 $x \in \mathbb{R}_+$ 我们指定 x 的平方根; 这不是一个映射, 因为平方根有两个: $\pm\sqrt{x}$.

对于映射 $f : A \to B$, 若当 $a_1 \neq a_2$ 时总有 $f(a_1) \neq f(a_2)$ (对任意 $a_1, a_2 \in A$), 则称 f 是**单射**. 这等价于: 若 $f(a_1) = f(a_2)$ 则必有 $a_1 = a_2$; 也等价于每个 $b \in \mathrm{Im}(f)$ 只有唯一的原像; 也等价于对任意 $b \in B$, $\#f^{-1}(b) \leq 1$.

我们称 f 是**双射**, 若 f 既是单射又是满射. 若 $A \subseteq B$, 令 $f : A \to B, f(a) := a$, 则我们确定了一个映射 $f : A \to B$, 称为**包含映射**或者**嵌入**. 若 $A = B$ 且 $f(a) = a$ 对任意 $a \in A$ 成立, 则称 f 为**恒等映射**, 记为 $\mathbf{1}_A$ 或者 id_A.

设 $f : A \to B, f' : A' \to B$ 是两个映射. 若 $A \subseteq A'$, 且对任意 $a \in A$ 都有 $f(a) = f'(a)$, 则称 f 是 f' 在 A 上的**限制**, 记作 $f = f'\Big|_A$; 同时称 f' 是 f 在 A' 上的**延拓**.

设 $f : A \to B, g : B \to C$ 是两个映射. 对任意 $a \in A$, 我们指定 $h(a) = g(f(a)) \in C$. 这样就定义了一个映射 $h : A \to C$, 记作 $h = g \circ f$ 或者 $h = gf$, 称为是 f, g 的**复合映射**. 映射复合满足结合律, 即: $f_3(f_2 f_1) = (f_3 f_2) f_1$, 这里设映射为 $f_1 : A \to B, f_2 : B \to C, f_3 : C \to D$.

设 $f : A \to B$ 是一个映射.

定义 0.12. *我们称映射 $g : B \to A$ 是 f 的*

- **左逆**, *如果 $g \circ f = \mathrm{id}_A$;*
- **右逆**, *如果 $f \circ g = \mathrm{id}_B$.*

当然, 并非所有映射都有左逆, 也并非所有映射都有右逆. 事实上:

命题 0.13. *设 $f : A \to B$ 是一个映射. 则 f*

1. *有左逆当且仅当 f 是单射;*
2. *有右逆当且仅当 f 是满射;*

直接的推论就是

推论 0.14. *$f : A \to B$ 是双射, 当且仅当 f 既有左逆又有右逆.*

命题 0.13 的证明. (1) 设 f 有左逆 g. 若 $f(a_1) = f(a_2)$, 则 $a_1 = g(f(a_1)) = g(f(a_2)) = a_2$; 所以 f 是单射. 反之, 若 f 是单射; 对任意 $b \in \mathrm{Im}(f)$, 存在唯一的 $a \in A$ 使得 $f(a) = b$, 我们令 $g(b) = a$; 我们任取 $a_0 \in A$, 对任意 $b \in B - \mathrm{Im}(f)$, 我们令 $g(b) = a_0$; 即

$$g(b) := \begin{cases} a & \text{若} b = f(a) \\ a_0 & \text{若} b \notin \mathrm{Im}(f) \end{cases}$$

直接验证可知: 对任意 $a \in A$, $g(f(a)) = a$; 即 g 是 f 的左逆.

(2) 若 f 有右逆 g, 对任意 $b \in B$, 有 $b = f(g(b))$, 即 $b \in \mathrm{Im}(f)$, 所以 f 是满射. 反之, 若 f 是满射, 由于对任意 $b \in B$, $f^{-1}(b) \neq \varnothing$, 我们任意取定**一个**元素 $a \in f^{-1}(b)$, 那么我们就定义了一个映射 $g : B \to A$. 直接验证可知 $fg = \mathrm{id}_B$. $\square$

0.1.1 映射复合的结合律

设 $f_i : X_i \to X_{i+1} (i = 1, 2, \cdots, n)$ 是 n 个映射, 我们要证明

命题 0.15. 复合映射 $f_n \circ f_{n-1} \circ \cdots \circ f_1$ 和复合的次序无关.

例如, 我们假设 $n = 4$, 那么我们就有很多种复合的方式, 例如 $(f_4 \circ f_3) \circ (f_2 \circ f_1)$, $f_4 \circ ((f_3 \circ f_2) \circ f_1)$; 若依然不理解, 我们还可以借助加法来理解: 我们计算 $1 + 2 + 3 + 4$, 就可以这么加 $(1 + 2) + (3 + 4) = 3 + 7 = 10$; 也可以这么加 $1 + ((2 + 3) + 4) = 1 + (5 + 4) = 1 + 9 = 10$, 这是两种不同的加的次序, 但是得到的值相同, 因此我们可以说: n 个实数相加和它们加的次序无关. 所以我们可以把所谓的 "次序" 理解为加括号, 使得加括号的那部分映射先复合.

我们要证明的是, 对任意的 $x \in X_1$, 任意的复合的次序, 我们都有 $f_n \circ f_{n-1} \circ \cdots \circ f_1(x) = f_n(f_{n-1}(\cdots(f_1(x))))$. $n = 3$ 时已经证明过了, 我们用数学归纳法来证明. 我们用第二类数学归纳法: 假设对任意的 $k \leq n$, k 个映射的复合都和复合的次序无关, 现在来证明 $n + 1$ 个映射的复合也和次序无关. 我们分两种情况讨论. 第一种情况, f_1 没有在任何括号里面, 因此整个复合映射就是先把前 n 个映射以一定的次序复合完了之后, 假设得到了映射 $g : X_2 \to X_{n+2}$, 再和 f_1 复合, 所以在 x 上我们最终的取值是 $h(f_1(x))$; 这时我们令 $y = f_1(x)$, 根据归纳假设, $h(y) = f_{n+1}(f_n(\cdots(f_2(y)))) = f_{n+1}(f_n(\cdots(f_2(f_1(x)))))$. 第二种情况, f_1 在某个括号里面, 我们假设含有 f_1 的括号的最长的起点为 f_i; 我们把这个括号内的部分记作 g', 那么 g' 就是 $f_i, f_{i-1}, \cdots, f_1$ 的某个次序的复合. 显然, $2 \leq i \leq n$; 若 $i = n + 1$, 我们得到了一个从头 (f_{n+1}) 到尾 (f_1) 的括号, 这个括号显然是无效的; 若 $i = 1$, 这个括号也无效, 意思是说 f_1 不在任何括号里. 这个时候整个复合映射等于 $f_{n+1}, f_n, \cdots, f_{i+1}, g'$ (共有 $n + 2 - i \leq n$ 个映射) 的某个次序的复合; 根据归纳假设, 整个复合映射在 x 上的取值为 $f_{n+1}(f_n(\cdot(g'(x))))$; 再对 g' 使用归纳假设, 我们有 $g'(x) = f_i(f_{i-1} \cdots (f_1(x)))$, 代回去, 就完成了证明.

定义 0.16. 集合 A 上的一个二元运算是指一个映射 $f : A \times A \to A$.

这是说, 对任意 $(a, b) \in A \times A$, 我们有唯一的 $f(a, b) \in A$. 例如

$$f : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}; \quad (a, b) \mapsto a + b$$

就是一个二元运算.

0.2 复数基础

数的扩充都把随着解方程, 从实数到复数的扩充, 就伴随着解二次方程: 若二次方程的判别式小于零, 那么该二次方程没有实数解. 所以我们将实数域扩充为复数域, 这样所有的二次方程就都有复数解了.

定义 0.17. 我们定义集合 $\mathbb{C} := \{z = x + y\boldsymbol{i} : x, y \in \mathbb{R}\}$, 并定义 $\mathbb{C}$ 上的两个二元运算加法和乘法如下:

$$\begin{aligned}(x_1 + y_1\boldsymbol{i}) + (x_2 + y_2\boldsymbol{i}) &:= (x_1 + x_2) + (y_1 + y_2)\boldsymbol{i}, \\ (x_1 + y_1\boldsymbol{i})(x_2 + y_2\boldsymbol{i}) &:= (x_1x_2 - y_1y_2) + (x_1y_2 + x_2y_1)\boldsymbol{i}.\end{aligned}$$

我们将 $\mathbb{C}$ 及这两个二元运算称为复数域; 它的每一个元素称为一个复数.

注释 0.18.

1. $\mathbb{C}$ 上的加法满足交换律和结合律, 乘法满足交换律和结合律; 加法和乘法满足分配率.
2. 令 $x_1 = x_2 = 0, y_1 = y_2 = 1$ 知 $\boldsymbol{i}^2 = -1$. 所以 $\boldsymbol{i}$ 是 -1 的一个" 平方根".
3. 令 $x_1 = x_2 = x, y_1 = -y_2 = y$, 则 $(x + y\boldsymbol{i})(x - y\boldsymbol{i}) = x^2 + y^2 \in \mathbb{R}$. 这说明, 若 x, y 不全为零, 则 $(x + y\boldsymbol{i})(\frac{x}{x^2+y^2} - \frac{y}{x^2+y^2}\boldsymbol{i}) = 1$. 等价地说, 如果 $0 \neq z \in \mathbb{C}$, 则存在 $z' \in \mathbb{C}$ 使得 $zz' = 1$.

定义 0.19. 对任意 $z = x + y\boldsymbol{i} \in \mathbb{C}$, 我们称 $\overline{z} = x - y\boldsymbol{i}$ 是 z 的**复共轭**; 称 $|z| := \sqrt{x^2 + y^2} = \sqrt{z\overline{z}}$ 为 z 的**模长**.

于是 $z = 0$ 当且仅当 $|z| = 0$; $z \in \mathbb{R}$ 当且仅当 $z = \overline{z}$. 对任意的 $x, y \in \mathbb{R}$, 不全为零, 则存在唯一的 $\theta \in [0, 2\pi)$ 使得

$$\frac{x}{x^2 + y^2} = \cos\theta, \quad \frac{y}{x^2 + y^2} = \sin\theta.$$

因此任意复数 $z = x + y\boldsymbol{i}$ 可以唯一的表示为 $z = r(\cos\theta + \boldsymbol{i}\sin\theta) = re^{\boldsymbol{i}\theta}$, 其中 $r \geq 0e^{\boldsymbol{i}\theta} = \cos\theta + \boldsymbol{i}\sin\theta$. 这称为 z 的**极坐标**表示, θ 称为是 z 的**幅角**. 这样表示有一个好处:

命题 0.20.

$$(r_1e^{\boldsymbol{i}\theta_1})(r_1e^{\boldsymbol{i}\theta_2}) = (r_1r_2)e^{\boldsymbol{i}(\theta_1+\theta_2)};$$

证明.

$$\begin{aligned}(r_1e^{\boldsymbol{i}\theta_1})(r_1e^{\boldsymbol{i}\theta_2}) &= r_1(\cos\theta_1 + \boldsymbol{i}\sin\theta_i)r_2(\cos\theta_2 + \boldsymbol{i}\sin\theta_2) \\ &= r_1r_2((\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + \boldsymbol{i}(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2)) \\ &= r_1r_2(\cos(\theta_1 + \theta_2) + \boldsymbol{i}\sin(\theta_1 + \theta_2)) \\ &= r_1r_2e^{\boldsymbol{i}(\theta_1+\theta_2)}\end{aligned}$$

□

也就是说, 计算复数乘法只需要 "模长相乘, 幅角相加". 一个简单的推论就是计算一个复数的高次幂:

$$(re^{i\theta})^n = r^n e^{in\theta}.$$

显而易见的, 任何实系数的二次方程 $ax^2 + bx + c = 0 (a \neq 0)$ 都有两个复数根

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

不仅如此, 我们有如下的 "代数基本定理":

定理 0.21. 任意的复系数多项式方程 $x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n = 0$ $(a_i \in \mathbb{C})$ 都恰好有 n 个复数根 (计重数).

特别的, 当 a_i 都是实数的话, 所有复数根都是 "成对" 出现的, 也就是说, 若 z 是该方程的一个根, 则 $\bar{z}$ 也是该方程的一个根, 这是因为 $\overline{a_i} = a_i$, 对 $z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n = 0$ 取复共轭就有 $\bar{z}^n + a_1\bar{z}^{n-1} + \cdots + a_{n-1}\bar{z} + a_n = 0$. 那么一个推论是: 奇数次的实系数多项式方程一定有实数解.

命题 0.22. 设 $z_1, z_2, \cdots, z_n$ 是复系数多项式方程 $x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n = 0$ 的 n 个根, 那么

$$z_1 + z_2 + \cdots + z_n = -a_1, \quad z_1 z_2 \cdots z_n = (-1)^n a_n.$$

这是因为 $x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n = (x - z_1)(x - z_2)\cdots(x - z_n)$.

0.3 数学归纳法

数学归纳法在处理关于自然数的命题上, 是个强有力的工具. 它的基本形式是这样的:

命题 0.23 (归纳法原理). 设 $\{P(n) : n \geq k\}$ 是一系列的命题. 若 $P(k)$ 成立, 且下面两条中有一条成立:

1. 假设 $P(n)$ 成立时, $P(n+1)$ 也成立.
2. 假设 $P(m)$ 对所有 $m \leq n$ 都成立时, $P(n+1)$ 也成立.

则 $P(n)$ 对所有的 $n \geq k$ 都成立.

我们分别举两个例子.

例题 0.24. 对任意正整数 $n \geq 1$, $1^3+2^3+\cdots+n^3=(1+2+\cdots n)^2$.

证明. 显然 $n=1$ 时, 等式成立. 假设 $1^3+2^3+\cdots+n^3=(1+2+\cdots n)^2$, 只需要证明 $1^3+2^3+\cdots+n^3+(n+1)^3=(1+2+\cdots n+(n+1))^2$. 根据归纳假设:

$$\begin{aligned}
1^3+2^3+\cdots+n^3+(n+1)^3 &= (1+2+\cdots n)^2+(n+1)^3 \\
&= \frac{n^2(n+1)^2}{4}+(n+1)^3 \\
&= (n+1)^2\left(\frac{n^2}{4}+(n+1)\right) \\
&= (n+1)^2\frac{(n+2)^2}{4} \\
&= \left(\frac{(n+1)((n+1)+1)}{2}\right)^2
\end{aligned}$$

□

例题 0.25. 设 $\mathbb{R}[x]$ 是系数在 $\mathbb{R}$ 中的所有多项式, 也就是说,

$$\mathbb{R}[x]=\left\{f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0 : a_i\in\mathbb{R}, n\in\mathbb{N}\right\}.$$

$\mathbb{R}[x]$ 中有我们熟知的加法和乘法. 对于 $f(x)=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0\in\mathbb{R}[x]$, 若 $a_n\neq 0$, 我们称 f 是次数为 n 的多项式, 记作 $\deg(f(x))=n$. 设 $p(x)\in\mathbb{R}[x]$ 且 $\deg(p(x))=r$; 证明: 对任意 $f(x)\in\mathbb{R}[x]$, 都存在多项式 $g(x),q(x)\in\mathbb{R}$, 使得 $f(x)=p(x)g(x)+q(x)$ 且 $\deg(q(x))<r$.

我们不妨假设 $p(x)=x^r+\cdots$, 即它的首项系数等于 1.

证明. 若 $\deg(f(x))<r$, 那么显然 $f(x)=p(x)\cdot 0+f(x)$, 我们取 $g(x)=0, q(x)=f(x)$ 即可; 奠基成立. 下面假设对任意次数小于等于 n 的多项式, 原命题成立; 我们需要证明对次数等于 $n+1$ 的多项式, 命题也成立. 设 $f(x)=a_{n+1}x^{n+1}+\cdots+a_1x+a_0(a_{n+1}\neq 0)$, 令 $\widetilde{f}(x)=f(x)-a_{n+1}x^{n+1-r}p(x)$; 那么 $\widetilde{f}(x)=f(x)=a'_nx^n+a'_{n-1}x^{n-1}+\cdots+a'_1x+a'_0$, 所以 $\deg(\widetilde{f}(x))\leq n$; 根据归纳假设, 存在多项式 $g(x),q(x)\in\mathbb{R}$, 使得 $\widetilde{f}(x)=p(x)g(x)+q(x)$ 且 $\deg(q(x))<r$, 那么

$$f(x)=p(x)\left(g(x)+a_{n+1}x^{n+1-r}\right)+q(x).$$

□

例题 0.26. 正整数 d 被称为正整数 n 的因子, 如果 n/d 也是整数. 正整数 n 被称为素数, 如果 n 没有除了 1 和 n 以外的因子. 证明: 所有正整数 $n\geq 2$ 都可以表示为素数的方幂的乘积.

例如, 一个素数天然已经表示为素数的方幂的成绩了, 再例如 $4 = 2^2, 12 = 2^2 \cdot 3, 100 = 2^2 \cdot 5^2$.

证明. $n = 2$ 时, 命题显然成立. 假设对任意 $k \geq n$, k 都可以表示为素数的方幂的乘积, 只需要证明 $n+1$ 也可以表示为素数的方幂的乘积. 我们分两种情况讨论; 若 $n+1$ 是素数, 那么命题天然成立; 若 $n+1$ 不是素数, 那么 $n+1 = pq$, 其中 p, q 都是大于等于 2 的整数; 于是 $p \leq n, q \leq n$; 根据归纳假设, p, q 都可以写成素数的方幂的乘积, 所以 $n+1 = pq$ 也是素数的方幂的乘积. □

第一章　线性代数的基本概念

考虑 2 元实数组 $\begin{bmatrix} x \\ y \end{bmatrix}, x \in \mathbb{R}, y \in \mathbb{R}$, 例如

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

或者 3 元实数组 $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ 甚至一般的 n 元实数组 $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. 我们把这些数组称为**向量**; 数组中的元素个数称为**维数**. 我们有以下操作:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}; \quad a \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax \\ ay \end{bmatrix}.$$

对于一般的 n 元数组 $\boldsymbol{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, 以及实数 a, b, 我们有

$$a\boldsymbol{v} + b\boldsymbol{w} = \begin{bmatrix} ax_1 + by_1 \\ \vdots \\ ax_n + by_n \end{bmatrix}.$$

这称为 $\boldsymbol{v}, \boldsymbol{w}$ 的一个线性组合.

问题 1.1. 给定 $\boldsymbol{v} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, 集合 $V := \{a\boldsymbol{v} + b\boldsymbol{w} : a, b \in \mathbb{R}\}$ 能表示哪些向量?

若 $\boldsymbol{v}, \boldsymbol{w}$ 其中之一为 0 向量, 例如 $\boldsymbol{v} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, 那么集合 $V = \left\{ \begin{bmatrix} bx_2 \\ by_2 \end{bmatrix} : b \in \mathbb{R} \right\}$.

若 $\boldsymbol{v} \neq 0, \boldsymbol{w} \neq 0$, 当存在实数 c 使得 $\boldsymbol{v} = c\boldsymbol{w}$ 时 (条件成立当且仅当 $x_1y_2 - y_1x_2 = 0$), $V = \left\{ \begin{bmatrix} bx_2 \\ by_2 \end{bmatrix} : b \in \mathbb{R} \right\}$. 若不存在实数 c 使得 $\boldsymbol{v} = c\boldsymbol{w}$ (条件成立当且仅当 $x_1y_2 - y_1x_2 \neq 0$), V 是整个 $\mathbb{R}^2$.

问题 1.2. 请思考一下 $\mathbb{R}^3$ 中 3 个向量的情形. 一般的, 考虑 $\mathbb{R}^n$ 中 m 个向量的情形.

我们总结一下线性代数的一般概念. 设 V 是一个非空集合, 在其中我们指定一个元素 $\mathbf{0}$, 并且有如下的二元运算:

$$\begin{array}{ll} V \times V \xrightarrow{+} V & \qquad (\boldsymbol{v}, \boldsymbol{w}) \mapsto \boldsymbol{v} + \boldsymbol{w} \\ \mathbb{R} \times V \longrightarrow V & \qquad (a, \boldsymbol{v}) \mapsto a\boldsymbol{v} \end{array}$$

我们称 V 是线性空间, 如果 $\mathbf{0}$ 和上述二元运算满足以下条件:

1. $\mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$;

2. $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$, 对任意的 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$;

3. $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$;

4. $a(\boldsymbol{v} + \boldsymbol{w}) = a\boldsymbol{v} + a\boldsymbol{w}, (a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$;

5. $a(b\boldsymbol{v}) = (ab)\boldsymbol{v}$;

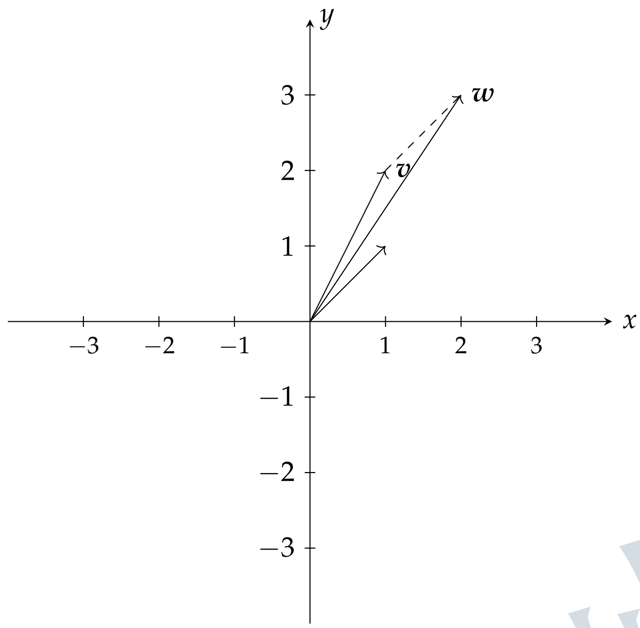
6. $1\boldsymbol{v} = \boldsymbol{v}, 0\boldsymbol{v} = \mathbf{0}$.

V 中的元素称为向量; 形如 $a\boldsymbol{v} + b\boldsymbol{w}$ 的元素称为 $\boldsymbol{v}, \boldsymbol{w}$ 的一个线性组合. 线性代数研究的主要对象有两个:

- 如何表示一个线性空间 V 自身;

- 给两个线性空间 V, W, 研究它们之间的关系.

我们还是回到具体的例子 $\mathbb{R}^2$. 设 $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.



问题 1.3. *是否存在* $x, y \in \mathbb{R}$ *使得* $x_1 \boldsymbol{v} + x_2 \boldsymbol{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$?

将 $\boldsymbol{v}, \boldsymbol{w}$ 的具体的值代入, 上述问题转化为求解二元一次方程组:

$$\begin{cases} x + 2y = 1 \\ 2x + 3y = 1 \end{cases}$$

每个方程定义了平面上的一条直线. 因此方程组有解, 当且仅当这两条直线相交, 并且交点的坐标即方程组的解. 我们把上述方程组写成所谓的" 矩阵形式":

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

根据上面的表述, 等式左边也可以表示向量 $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 和 $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ 的线性组合: $x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

更一般的, 给定线性空间 V 上的 $n + m$ 个向量 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_m$, 我们要问:

问题 1.4. *是否存在* $n \times m$ *个实数* $x_{11}, \cdots, x_{1n}, \cdots, x_{m1}, \cdots, x_{mn} \in \mathbb{R}$, *使得*

$$x_{i1} \boldsymbol{v}_1 + \cdots x_{in} \boldsymbol{v}_n = \boldsymbol{w}_i \quad 1 \leq i \leq m?$$

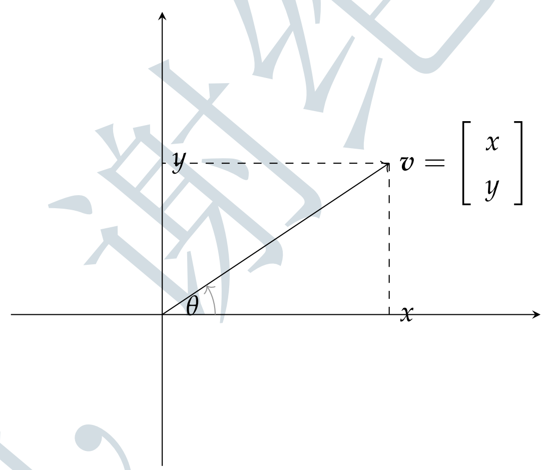
我们可以形式地用矩阵表示这个方程组:

$$\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}.$$

暂且, 这样的矩阵形式只是一个形式记号. 将来, 我们会赋予这个记号本质的含义.

1.1　点积

对于 $\mathbb{R}^2$ 中的向量而言, 虽然它们的坐标表示 $\begin{bmatrix} x \\ y \end{bmatrix}$ 已经表达出了所有的信息. 然而, 我们仍然需要显式地提取出我们所关心的内容. 几何上来看, $\mathbb{R}^2$ 中的一个向量被它的 “方向” 和 “长度” 唯一决定. 因此, 我们所需要做的就是从一个向量的坐标表示出发, 提取出 “方向” 和 “长度” 这两个关键属性.



由勾股定理, 我们知道 v 的长度为

$$\|v\| = (x^2 + y^2)^{1/2}.$$

对于 $\mathbb{R}^3$ 中的向量 $v = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, 同样地由勾股定理, 我们知道它的长度为

$$\|v\| = (x^2 + y^2 + z^2)^{1/2}.$$

因此, 对于 $\mathbb{R}^n$ 中的向量 $\boldsymbol{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, 我们自然地想要定义, 并且也确实如此定义它的长度为

$$\|\boldsymbol{v}\| = (x_1^2 + \cdots + x_n^2)^{1/2}.$$

命题 1.5. 长度满足三角不等式, 即对任意 $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, 都有

$$\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|.$$

证明. 设 $\boldsymbol{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, 那么 $\boldsymbol{v} + \boldsymbol{w} = (x_1 + y_1, x_n + y_n)$, 因此原不等式等价于

$$\left((x_1 + y_1)^2 + \cdots + (x_n + y_n)^2\right)^{1/2} \leq (x_1^2 + \cdots + x_n^2)^{1/2} + (y_1^2 + \cdots + y_n^2)^{1/2}.$$

两边平方, 则原不等式等价于

$$x_1y_1 + \cdots + x_ny_n \leq (x_1^2 + \cdots + x_n^2)^{1/2}(y_1^2 + \cdots + y_n^2)^{1/2}.$$

这个不等式就是著名的 Cauchy-Schwartz 不等式. □

定义 1.6. 长度为 1 的向量被称为**单位向量**.

注释 1.7. 对于任意非零向量 u, $u_1 := \frac{1}{\|u\|}u$ 是一个单位向量.

例题 1.8. 设 $\boldsymbol{u} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ 为单位向量, 那么 $x^2 + y^2 = 1$, 所以

$$\begin{cases} x & = \cos(\theta) \\ y & = \sin(\theta) \end{cases}$$

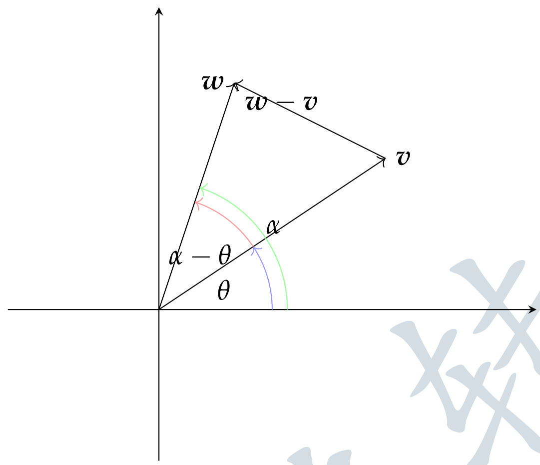
此处, θ 是将 x-正轴逆时针旋转至和由 $\boldsymbol{u}$ 所确定的射线重合所经过的最小角度.

当然, 这个角度取值范围为 $[0, 2\pi)$. 逆时针旋转 θ 角度和顺时针旋转 $2\pi - \theta$ 角度一致. 因此, 我们通常将大于 π 的角度 θ 转化为 $2\pi - \theta$, 因此我们得到的" 角度" 取值范围为 $[0, \pi]$. 而这个所谓的" 角度" 实际上是由 x-正轴和 $\boldsymbol{u}$ 所形成的的夹角.

由于余弦函数在 $[0, \pi]$ 上是单调的, $\theta \in [0, \pi]$ 被它的余弦值 $\cos(\theta)$ 唯一决定.

注释 1.9. 对任意非零向量 u, 若其单位化向量为 $u_1 = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$, 那么 $u = \begin{bmatrix} \|u\| \cos(\theta) \\ \|u\| \sin(\theta) \end{bmatrix}$. 反之亦然.

我们也可以看到, 当我们要谈论 “角度” 的时候我们不能只考虑一个向量, 必须要两个向量才能有一个夹角.



我们来看一个特殊例子. 假设我们有两个单位向量 $\boldsymbol{v}=\begin{bmatrix}\cos(\theta)\\ \sin(\theta)\end{bmatrix}$, $\boldsymbol{w}=\begin{bmatrix}\cos(\alpha)\\ \sin(\alpha)\end{bmatrix}$, 那么从 $\boldsymbol{v}$ 到 $\boldsymbol{w}$ 的角度为 $\alpha-\theta$, 其余弦值为

$$\cos(\alpha-\theta)=\cos(\alpha)\cos(\theta)+\sin(\alpha)\sin(\theta).$$

对任意非零向量 $\boldsymbol{v}=\begin{bmatrix}x_1\\ y_1\end{bmatrix}$, $\boldsymbol{w}=\begin{bmatrix}x_2\\ y_2\end{bmatrix}$, 令它们的单位化向量为 $\boldsymbol{v}_1=\begin{bmatrix}\cos(\theta)\\ \sin(\theta)\end{bmatrix}$, $\boldsymbol{w}_2=\begin{bmatrix}\cos(\alpha)\\ \sin(\alpha)\end{bmatrix}$, 那么 $\boldsymbol{v}$ 和 $\boldsymbol{w}$ 的夹角 δ 等于 $\boldsymbol{v}_1$ 和 $\boldsymbol{w}_1$ 的夹角, 因此 $\delta=\alpha-\theta$, 从而

$$\begin{aligned}\cos(\delta) &= \cos(\alpha)\cos(\theta)+\sin(\alpha)\sin(\theta)\\ &= \frac{1}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}\Big((\|\boldsymbol{w}\|\cos(\alpha))(\|\boldsymbol{v}\|\cos(\theta))+(\|\boldsymbol{w}\|\sin(\alpha))(\|\boldsymbol{v}\|\sin(\theta))\Big)\\ &= \frac{1}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}\Big(x_1x_2+y_1y_2\Big)\end{aligned}$$

于是, 我们定义 $\boldsymbol{v}$ 和 $\boldsymbol{w}$ 的点积为

$$\boldsymbol{v}\cdot\boldsymbol{w}=x_1x_2+y_1y_2.$$

那么它们的夹角 δ 满足: $\cos(\delta)=\dfrac{\boldsymbol{v}\cdot\boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}$. 特别的,

1. $\boldsymbol{v}$ 和 $\boldsymbol{w}$ 方向相同, 当且仅当 $\cos(\delta)=1$;

2. $\boldsymbol{v}$ 和 $\boldsymbol{w}$ 方向相反, 当且仅当 $\cos(\delta)=-1$;

3. $\boldsymbol{v}$ 和 $\boldsymbol{w}$ 垂直, 当且仅当 $\cos(\delta) = 0$;

注释 1.10. 上述点积的公式我们是从两个非零向量出发讨论的. 若其中之一为零向量, 我们补充定义它们的点积为 0. 对任意向量 $\boldsymbol{v}$, 我们有

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

对于 $\mathbb{R}^n$ 中的两个向量 $\boldsymbol{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, 我们也想想定义其点积为

$$\boldsymbol{v} \cdot \boldsymbol{w} := x_1 y_1 + x_2 y_2 + \cdots y_1 y_n.$$

那么我们需要思考一下这个定义是否合理. 无论维数 n 等于多少, $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{w} - \boldsymbol{v}$ 形成一个三角形或者共线, 三条边长分别为 $\|\boldsymbol{v}\|, \|\boldsymbol{w}\|, \|\boldsymbol{w} - \boldsymbol{v}\|$. 由余弦定理,

$$\|\boldsymbol{w} - \boldsymbol{v}\|^2 = \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 - 2\|\boldsymbol{v}\|\|\boldsymbol{w}\| \cos(\delta).$$

而 $\|\boldsymbol{w} - \boldsymbol{v}\|^2 = (y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2, \|\boldsymbol{v}\|^2 = x_1^2 + x_2^2 + \cdots + x_n^2, \|\boldsymbol{w}\|^2 = y_1^2 + y_2^2 + \cdots + x_n^2$, 即得

$$\|\boldsymbol{v}\|\|\boldsymbol{w}\| \cos(\delta) = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

也就是说

$$\cos(\delta) = \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{v}\|\|\boldsymbol{w}\|}.$$

注释 1.11. Cauchy-Schwartz 不等式保证了 $\boldsymbol{v} \cdot \boldsymbol{w} \leq \|\boldsymbol{v}\|\|\boldsymbol{w}\|$, 或者说, $|\cos(\delta)| \leq 1$. 因此 δ 是唯一决定的.

1.2 矩阵

回顾一下，令 $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. 我们可以考虑他们的线性组合，即如下形式的向量 $x_1 \boldsymbol{v} + x_2 \boldsymbol{w}, x_1, x_2 \in \mathbb{R}$. 那么给定一个向量 $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, 它是否是 $\boldsymbol{v}, \boldsymbol{w}$ 的线性组合呢? 即关于 x_1, x_2 的方程

$$x_1 \boldsymbol{v} + x_2 \boldsymbol{w} = \boldsymbol{b}$$

是否有解. 我们有形式记号

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{1}$$

从列的方式看, 等式左边是 $\boldsymbol{v},\boldsymbol{w}$ 的线性组合. 我们也可以用行的方式看. 由于方程 $x_1\boldsymbol{v}+x_2\boldsymbol{w}=\boldsymbol{b}$ 等价于线性方程组:

$$\begin{cases} x_1+2x_2=b_1 \\ 2x_1+3x_2=b_2 \end{cases}$$

而 $x_1+2x_2=\begin{bmatrix}1\\2\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\end{bmatrix}, 2x_1+3x_2=\begin{bmatrix}2\\3\end{bmatrix}\cdot\begin{bmatrix}x_1\\x_2\end{bmatrix}$. 因此, 左边的矩阵乘法可以看成是把一些向量的内积重新排成矩阵.

解这个方程组, 我们得到

$$x_1=-3b_1+2b_2, x_2=2b_1-b_2;$$

因此 $b_1\begin{bmatrix}-3\\2\end{bmatrix}+b_2\begin{bmatrix}2\\-1\end{bmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}$. 我们采用矩阵来做形式记号, 得

$$\begin{bmatrix}-3&2\\2&-1\end{bmatrix}\begin{bmatrix}b_1\\b_2\end{bmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}. \tag{2}$$

看起来, $\begin{bmatrix}-3&2\\2&-1\end{bmatrix}$ 好像是 $\begin{bmatrix}1&2\\2&3\end{bmatrix}$ 的 -1 次方, 而 (2) 式是在 (1) 式等号两边同时乘以 $\begin{bmatrix}1&2\\2&3\end{bmatrix}$ 的 -1 次方得到.

若考虑 $\mathbb{R}^n$ 中的 n 个向量 $\boldsymbol{v}_1=\begin{bmatrix}a_{11}\\\vdots\\a_{n1}\end{bmatrix},\cdots,\boldsymbol{v}_n=\begin{bmatrix}a_{1n}\\\vdots\\a_{nn}\end{bmatrix}$, 那么给定另一个向量 $\boldsymbol{b}=\begin{bmatrix}b_1\\\vdots\\b_n\end{bmatrix}$, 它是否是 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 的线性组合呢? 也就是说, 是否存在 n 个实数 $x_1,\cdots,x_n$, 使得

$$x_1\boldsymbol{v}_1+\cdots+x_n\boldsymbol{v}_n=\boldsymbol{b}.$$

用矩阵语言来描述, 就是

$$\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&\ddots&&a_{2n}\\\vdots&&\ddots&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}.$$

我们也希望矩阵 $A := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ 有 -1 次方: A^{-1}, 使得我们在上式两边同时乘以 A^{-1}, 就能够把这个方程解出来.

定义 1.12. 若对于任意的 $\boldsymbol{b}$, 上述方程有唯一解 $\boldsymbol{x}$, 那么我们称 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 线性无关. 否则称他们线性相关.

如果 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 线性无关, 我们称矩阵 A 是可逆矩阵. 否则称为不可逆矩阵. 当 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 线性无关时, 特别的如果 $\boldsymbol{b} = 0$, $A\boldsymbol{x} = 0$ 有唯一解.

当然, 我们也可以考虑 $\mathbb{R}^n$ 中取 m 个向量的线性组合能表示哪些向量的问题.

例题 1.13. 考虑 $\mathbb{R}^3$ 中的向量 $\boldsymbol{v} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, 问: $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ 是否是 $\boldsymbol{v}, \boldsymbol{w}$ 的线性组合? 用矩阵的形式写就是

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

是否有解.

这个时候, 我们似乎就无法谈论” -1 次方” 这个概念. 而且, 这个方程组也的确没有解. 这提示我们需要系统地研究线性方程组.

1.3 线性变换

在之前的课程当中, 给定若干个向量, 我们考虑了它们的所有线性组合构成的集合. 我们用另一种看法来重新看待这个过程.

例题 1.14. 设 $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^2$, 对任意的 $a, b \in \mathbb{R}$, 有它们的线性组合 $a\boldsymbol{v} + b\boldsymbol{w} \in \mathbb{R}^2$. 我们把 a, b 组合成 $\mathbb{R}^2$ 中的向量 $\begin{bmatrix} a \\ b \end{bmatrix}$, 那么我们就得到一个映射

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \quad \begin{bmatrix} a \\ b \end{bmatrix} \mapsto a\boldsymbol{v} + b\boldsymbol{w}.$$

定义 1.15. 我们称映射 $f:\mathbb{R}^n\to\mathbb{R}^m$ 是一个线性映射 (或者线性变换), 如果对任意 $a,b\in\mathbb{R}$, $\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^n$, 我们都有

$$f(a\boldsymbol{v}+b\boldsymbol{w})=af(\boldsymbol{v})+bf(\boldsymbol{w}).$$

例题 1.16. 设 $m=n+1$, 那么

$$f:\mathbb{R}^n\to\mathbb{R}^{n+1};\quad \begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\mapsto\begin{bmatrix}x_1\\ \vdots\\ x_n\\ 0\end{bmatrix}$$

是一个线性变换. 而

$$f:\mathbb{R}^n\to\mathbb{R}^{n+1};\quad \begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\mapsto\begin{bmatrix}x_1\\ \vdots\\ x_n\\ 1\end{bmatrix}$$

不是一个线性变换.

例题 1.17. 设 $m=n-1$, 那么

$$f:\mathbb{R}^n\to\mathbb{R}^{n-1};\quad \begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\mapsto\begin{bmatrix}x_1\\ \vdots\\ x_{n-1}\end{bmatrix}$$

是一个线性变换.

例题 1.18. 设 $\boldsymbol{v},\boldsymbol{w}$ 是 $\mathbb{R}^m$ 中两个向量, 那么

$$f:\mathbb{R}^2\to\mathbb{R}^m,\quad \begin{bmatrix}x_1\\ x_2\end{bmatrix}\mapsto x_1\boldsymbol{v}+x_2\boldsymbol{w}$$

是一个线性变换. 一般的, 取 $\mathbb{R}^m$ 中 n 个向量 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$, 那么

$$f:\mathbb{R}^n\to\mathbb{R}^m,\quad \begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}\mapsto x_1\boldsymbol{v}_1+\cdots+x_n\boldsymbol{v}_n$$

是一个线性变换.

从上面的例子中, 我们可以得到表达一个线性映射的方式. 事实上, 设 $f:\mathbb{R}^n\to\mathbb{R}^m$ 是一个线性映射, 那么

$$f\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=x_1f\begin{bmatrix}1\\0\\0\end{bmatrix}+x_2f\begin{bmatrix}0\\1\\ \vdots\\0\end{bmatrix}+\cdots+x_nf\begin{bmatrix}0\\0\\ \vdots\\1\end{bmatrix}.$$

因此, 线性映射 f 被它在 $\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\cdots,\begin{bmatrix}0\\0\\1\end{bmatrix}$ 上的取值完全决定. 如果我们令

$$\boldsymbol{v}_1=f\begin{bmatrix}1\\0\\0\end{bmatrix},\boldsymbol{v}_2=f\begin{bmatrix}0\\1\\ \vdots\\0\end{bmatrix},\cdots,\boldsymbol{v}_n=f\begin{bmatrix}0\\0\\ \vdots\\1\end{bmatrix},$$

那么

$$f\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=x_1\boldsymbol{v}_1+x_2\boldsymbol{v}_2+\cdots+x_n\boldsymbol{v}_n. \tag{3}$$

总结上面的讨论, 我们就有如下结论:

命题 1.19. 映射 $f:\mathbb{R}^n\to\mathbb{R}^m$ 是一个线性映射, 当且仅当存在向量 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\in\mathbb{R}^m$, 使得

$$f\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=x_1\boldsymbol{v}_1+x_2\boldsymbol{v}_2+\cdots+x_n\boldsymbol{v}_n.$$

注意到 $\boldsymbol{v}_i\in\mathbb{R}^m$, 我们记

$$\boldsymbol{v}_i=\begin{bmatrix}a_{1i}\\ \vdots\\ a_{mi}\end{bmatrix},1\le i\le n.$$

那么

$$x_1\boldsymbol{v}_1+x_2\boldsymbol{v}_2+\cdots+x_n\boldsymbol{v}_n=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\ a_{21}&\ddots&&a_{2n}\\ \vdots&&\ddots&\vdots\\ a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\ \vdots\\x_n\end{bmatrix} \tag{4}$$

比较 (3) 式和 (4) 式, 形式上我们有

$$f \quad '' = '' \quad \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

在之前的讨论中, 我们只是在描述一个矩阵, 那么, 我们总要给一个严格的定义:

定义 1.20. $\mathbb{R}$ 上的一个 $n \times m$ 矩阵是指一个由 nm 个数组成的有 n 行和 m 列的矩形图表.

注释 1.21. 我们也有行向量和列向量的区分, 行向量形如 $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]$, 我们用 $\boldsymbol{x}^T$ 来表示列向量 $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. Vice versa.

第二章　求解线性方程组

在这一章中, 我们将系统地学习如何判定一个线性方程组有没有解; 在有解时, 如何确定它有多少解; 在有唯一解时, 如何求出这个唯一解; 如果解不唯一, 如何确定由所有的解构成的集合的结构.

2.1　高斯消元法

例题 2.1. 求解线性方程组

$$\begin{cases} x+3y+5z &= 4 \\ x+2y-3z &= 5 \\ 2x+5y+2z &= 8 \end{cases}$$

例题 2.2. 求解线性方程组

$$\begin{cases} x+3y+5z &= 4 \\ x+2y-3z &= 5 \\ 2x+5y+2z &= 9 \end{cases}$$

例题 2.3. 求解线性方程组

$$\begin{cases} x+2y+3z &= 6 \\ 2x+5y+2z &= 4 \\ 6x-3y+1z &= 2 \end{cases}$$

所谓的元, 就是指独立的未知数; 一个线性方程组, 未知数有 N 个, 我们就称这个方程组为 N 元方程组. 那么显然的, 同类型的方程, 未知数越少, 方程越简单. 因此, 我们解线性方程组的一个最基本的想法, 就是消元. 其逻辑链条是这样的:

- 一元线性方程组我们都会解; 如果不会, 请返回小学三年级重新学习相关知识.
- 一个 $N \geq 2$ 元线性方程组, 若我们始终可以将其转化成 $N-1$ 元方程组, 那么逐步地, 最终总能转化成一元方程组, 因此我们可以将这个 N 元线性方程组解出.

所以, 解线性方程组的核心问题就在于第二条, 我们如何减少未知数的个数, 简称" 消元". 然而, 我们并不是始终可以消元, 例如只有一个线性方程的情形, 例如

$$x_1 + 2x_2 + \cdots + nx_n = 0.$$

只有一个方程, 因此束手无策, 无法消元. 但是事实上, 这样的方程我们很容易地就可以确定它的所有的解: 由于

$$x_1 = -(2x_2 + \cdots + nx_n),$$

因此, 对 $x_2, x_3, \cdots, x_n$ 任意取值, 然后就可以唯一地决定 x_1 的取值, 使得它们是原方程的解. 所以这个方程的解空间有 $n-1$ 个" 自由变量", 因此是"$n-1$ 维" 的.

在考虑一般的 n-元线性方程组之前, 我们先看一个例子:

$$\left\{\begin{array}{rrrrcrcc} x_1 & -2x_2 & -x_3 & +3x_4 & = & 0 & \cdots\cdots & ① \\ -2x_1 & 4x_2 & +5x_3 & -5x_4 & = & 3 & \cdots\cdots & ② \\ 3x_1 & -6x_2 & -6x_3 & +8x_4 & = & 2 & \cdots\cdots & ③ \end{array}\right.$$

我们还是贯彻消元的思路. 先" 消" x_1: ② $+2\cdot$ ①, ③ $-3\cdot$ ①:

$$\left\{\begin{array}{rrrrcrcc} x_1 & -2x_2 & -x_3 & +3x_4 & = & 0 & \cdots\cdots & ④ \\ & & 3x_3 & +x_4 & = & 3 & \cdots\cdots & ⑤ \\ & & -3x_3 & -x_4 & = & 2 & \cdots\cdots & ⑥ \end{array}\right.$$

现在要消 x_2 了. 显然, 只有 ④ 含有 x_1, 那么 ④ 就不能动了. 而 ⑤ 和 ⑥ 都没有 x_2, 这意味着什么呢? 这意味着消 x_2 这个步骤已经天然地完成了! 于是我们接着消 x_3: ⑤ + ⑥:

$$\left\{\begin{array}{rrrrcr} x_1 & -2x_2 & -x_3 & +3x_4 & = & 0 \\ & & 3x_3 & +x_4 & = & 3 \\ & & & 0 & = & 5 \end{array}\right.$$

最后一个方程显然是不成立的, 这就意味着, 原方程组没有解.

体会一下上面那个例子的具体过程, 现在我们可以尝试一下一般的 n 元方程组了:

$$\left\{\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. \end{array}\right. \tag{1}$$

若 $a_{11}=a_{21}=\cdots=a_{m1}=0$, 那么这个方程组和 x_1 无关, 我们只需要解方程组

$$\begin{cases} a_{12}x_2+\cdots+a_{1n}x_n=b_1\\ a_{22}x_2+\cdots+a_{2n}x_n=b_2\\ \qquad\qquad\vdots\\ a_{m2}x_2+\cdots+a_{mn}x_n=b_m \end{cases}$$

即可. 因此我们成功地 "消去" 了未知数 x_1. 那么我们考虑 x_2 的系数: $a_{12},a_{22},\cdots,a_{m2}$, 若它们也全为零, 那我们也成功地" 消去" 了未知数 x_2; 再考虑 x_3 的系数......., 直至我们得到了某个 x_{j_1}, 使得它的系数 $a_{1j_1},a_{2j_1},\cdots,a_{mj_1}$ 不全为零. 此时, 我们原方程组 (1) 实际上是

$$\begin{cases} a_{1j_1}x_{j_1}+\cdots+a_{1n}x_n=b_1\\ a_{2j_1}x_{j_1}+\cdots+a_{2n}x_n=b_2\\ \qquad\qquad\vdots\\ a_{mj_1}x_{j_1}+\cdots+a_{mn}x_n=b_m \end{cases} \tag{1'}$$

注释 2.4. 极端情形, 若这样的 j_1 找不出来, 也就是说所有未知数的所有系数都是 0, 那么原方程组就是

$$\begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\\vdots\\b_m\end{bmatrix}.$$

那么关于这个方程组的解的讨论是平凡的.

我们考虑个特殊情况: 当 $j_1=1$ 时, 等价地说, $a_{11},a_{21},\cdots,a_{m1}$ 不全为零, 不妨设 $a_{i1}\neq 0$, 我们交换第 1 个方程和第 i 个方程, 得到

$$\begin{cases} a_{i1}x_1 & +a_{i2}x_2 & +\cdots+a_{in}x_n & =b_i\\ a_{21}x_1 & +a_{22}x_2 & +\cdots+a_{2n}x_n & =b_2\\ & \vdots & & \\ a_{i-1,1}x_1 & +a_{i-1,2}x_2 & +\cdots+a_{i-1,n}x_n & =b_{i-1}\\ a_{11}x_1 & +a_{12}x_2 & +\cdots+a_{1n}x_n & =b_1\\ a_{i+1,1}x_1 & +a_{i+1,2}x_2 & +\cdots+a_{i+1,n}x_n & =b_{i+1}\\ & \vdots & & \\ a_{m1}x_1 & +a_{m2}x_2 & +\cdots+a_{mn}x_n & =b_m. \end{cases} \tag{2}$$

注释 2.5. 若 $a_{11}\neq 0$, 这一步我们实际上保持原状即可

将这个新方程组的第 $\ell = 2, 3, \cdots, n$ 个方程减去第一个方程的 $\dfrac{a_{\ell 1}}{a_{i1}}$ 倍 ($\ell = i$ 时, 减去 $\dfrac{a_{11}}{a_{i1}}$ 倍), 得到一个新的线性方程组, 这个方程组除了第 1 个方程, 其他方程均不包含 x_1, 因此我们可以记作:

$$\begin{cases} a'_{11}x_1 + a'_{12}x_2 + \cdots + a'_{1n}x_n &= b'_1 \\ \qquad\quad\; a'_{22}x_2 + \cdots + a'_{2n}x_n &= b'_2 \\ \qquad\qquad\qquad\qquad\quad \vdots \\ \qquad\quad\; a'_{m2}x_2 + \cdots + a'_{mn}x_n &= b'_m. \end{cases} \tag{3}$$

其中 $a'_{11} = a_{i1}, a'_{12} = a_{i2}, \cdots, a'_{1n} = a_{in}, b'_1 = b_i$. 所以我们只需要解线性方程组

$$\begin{cases} a'_{22}x_2 + \cdots + a'_{2n}x_n &= b'_2 \\ \qquad\qquad\qquad \vdots \\ a'_{m2}x_2 + \cdots + a'_{mn}x_n &= b'_m. \end{cases} \tag{3'}$$

它的任何一个解, 利用方程组 (3) 的第一个方程, 可以唯一的决定 x_1 的值, 合在一起, 构成原方程组 (1) 的一个解. 这个时候我们可以做进一步的消元了, 也就是说, 要处理方程组 (3′); 如同消去 x_1 一样, 我们需要做如下三步判断:

1. x_2 的所有系数是否均等于 0. 若是则消元成功, 我们考虑 x_3 的系数;

2. 若否, 取一个 x_2 的系数非零的方程, 将它和原第二行交换;

3. 将新方程组的第 $3, 4, \cdots, m$ 个方程减去第 2 个方程的合适的倍数, 即可消去第 $3, 4, \cdots, m$ 个方程中的变量 x_2.

对于一般的 j_1 (j_1 未必等于 1, 所以实际上我们是在解方程组 (1′)), 因为 $a_{1j_1}, a_{2j_1}, \cdots, a_{mj_1}$ 不全为零, 不妨设 $a_{ij_1} \neq 0$, 我们交换第 1 个方程和第 i 个方程, 得到

$$\begin{cases} a_{ij_1}x_{j_1} & + a_{i,j_1+1}x_{j_1+1} & + \cdots + a_{in}x_n & = b_i \\ a_{2j_1}x_{j_1} & + a_{2,j_1+1}x_{j_1+1} & + \cdots + a_{2n}x_n & = b_2 \\ & \vdots & & \\ a_{i-1,j_1}x_{j_1} & + a_{i-1,j_1+1}x_{j_1+1} & + \cdots + a_{i-1,n}x_n & = b_{i-1} \\ a_{1,j_1}x_{j_1} & + a_{1,j_1+1}x_{j_1+1} & + \cdots + a_{1n}x_n & = b_1 \\ a_{i+1,j_1}x_{j_1} & + a_{i+1,j_1+1}x_{j_1+1} & + \cdots + a_{i+1,n}x_n & = b_{i+1} \\ & \vdots & & \\ a_{m,j_1}x_{j_1} & + a_{m,j_1+1}x_{j_1+1} & + \cdots + a_{mn}x_n & = b_m. \end{cases}$$

将这个新方程组的第 $\ell=2,3,\cdots,n$ 个方程减去第一个方程的 $\dfrac{a_{\ell,j_1}}{a_{i,j_1}}$ 倍 ($\ell=i$ 时, 减去 $\dfrac{a_{1j_1}}{a_{ij_1}}$ 倍), 得到一个新的线性方程组, 这个方程组除了第 1 个方程, 其他方程均不包含 x_{j_1}, 因此我们可以记作:

$$\begin{cases} a'_{1,j_1}x_{j_1}+a'_{1,j_1+1}x_{j_1+1}+\cdots+a'_{1n}x_n &= b'_1 \\ \qquad\qquad a'_{2,j_1+1}x_{j_1+1}+\cdots+a'_{2n}x_n &= b'_2 \\ \qquad\qquad\qquad\qquad\qquad\vdots & \\ \qquad\qquad a'_{m,j_1+1}x_{j_1+1}+\cdots+a'_{mn}x_n &= b'_m. \end{cases}$$

所以我们只需要再解线性方程组

$$\begin{cases} a'_{2,j_1+1}x_{j_1+1}+\cdots+a'_{2n}x_n &= b'_2 \\ \qquad\qquad\qquad\vdots & \\ a'_{m,j_1+1}x_{j_1+1}+\cdots+a'_{mn}x_n &= b'_m. \end{cases}$$

接下来的讨论和 $j_1=1$ 是完全是一致类比, 因此我们不再赘述.

最终, 我们得到的方程组是如下形式:

$$\begin{cases} 0+\cdots+t_{1j_1}x_{j_1}+\cdots+t_{1j_2}x_{j_2}+\cdots+t_{1j_r}x_{j_r}+\cdots+t_{1n}x_n &= c_1 \\ \qquad\qquad\qquad\qquad t_{2j_2}x_{j_2}+\cdots+t_{2j_r}x_{j_r}+\cdots+t_{2n}x_n &= c_2 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\vdots & \\ \qquad\qquad\qquad\qquad\qquad\qquad t_{rj_r}x_{j_r}+\cdots+t_{rn}x_n &= c_r \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 0 &= c_{r+1} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\vdots & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 0 &= c_m \end{cases} \tag{4}$$

其中 $t_{1j_1}\cdots t_{rj_r}\neq 0$, 即 $t_{1j_1},t_{2j_2},\cdots t_{rj_r}$ 均非零.

注释 2.6. 这样的记号的意思是: 如果我们得到的 $r=m$, 则后面的 $0=c_{r+1},\cdots,0=c_m$ 等行是不存在的. 如果 $r<m$, 则那些行存在.

注释 2.7. 这里 j_1 是最小的正整数 k, 使得 x_k 的系数 $a_{1k},a_{2k},\cdots,a_{mk}$ 不全为零. 例如, 若 $a_{i1\neq 0}$, 那么 $j_1=1$; 若 x_1 的系数 $a_{11},a_{21},\cdots,a_{m1}$ 全为零, 而 x_2 的某个系数 $x_{i2}\neq 0$, 那么 $j_1=2$. j_2 是对 x_{j_1} 作了消元之后得到的方程组 "给出的 j_1". 例如, 当 $j_1=1$ 时, 我们对 x_1 作消元, 得到了方程组 $(3')$. 那么 j_2 就是最小的正整数 $k\geq 2$, 使得 x_k 的系数 $a'_{2k},a'_{3k}\cdots,a'_{mk}$ 不全为零. $j_3,\cdots$ 的取法以此类推.

此时, 我们很容易就可以做出如下判断:

- $r \le \min\{m, n\}$.
- 若 $c_{r+1}, \cdots, c_m$ 不全为零, 那么原方程组无解.
- 若 $c_{r+1}, \cdots, c_m$ 全为零, 那么 "有效方程的个数 r" 不大于 n, 且原方程有解.
- 任意确定 $x_{j_r+1}, \cdots, x_n$ 的值, 由第 r 个方程, 可以唯一确定 x_r 的值 (注意, 有 $n - j_r$ 个 "自由变量"); 将他们代回第 $r-1$ 个方程, 再任意确定 $x_{j_{r-1}+1}, \cdots, x_{j_r-1}$ 的值 (注意, 有 $j_r - 1 - j_{r-1}$ 个 "自由变量"), 由第 $r-1$ 个方程, 可以唯一确定 $x_{j_{r-1}}$ 的值. 直至任意确定 $x_1, \cdots, x_{j_1-1}$ 的值 (注意, 有 $j_1 - 1$ 个自由变量). 这样, 我们就可以给出原方程组的所有解.

注释 2.8. *根据上面的讨论, 自由变量的个数等于*

$$(n - j_r) + (j_r - 1 - j_{r-1}) + \cdots + (j_2 - j_1 - 1) + (j_1 - 1) = n - r.$$

因此, 我们应该猜测, 解的空间是 "$n - r$ 维" 的, 尽管我们并没有定义 "维数" 这个概念.

这个过程, 就叫高斯消元法, 它完全地回答了我们章首提出的所有问题.

推论 2.9. *考虑齐次方程组 $A\boldsymbol{x} = 0$, 其中 $A \in M_{m\times n}(\mathbb{R}), m < n$. (也就是说方程个数比未知数个数少), 那么方程组一定有无穷多个解.*

这是因为, 因为 $r \le m < n$, 所以有 $n - r(> 0)$ 个 "自由变元".

例题 2.10.

$$\begin{cases} 2x + y + 3z + t &= 6 \\ x + 2y + z &= 5 \\ y + 2z + t &= 4 \\ z + 2t &= 3 \end{cases}$$

2.2 高斯消元法的矩阵表达

在引言中, 我们已经知道, 一个线性方程组有矩阵表达, (1) 式用矩阵表达为:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

注意, 一个 n-维列向量可以看成一个 n 行 1 列的矩阵, 那么我们首先可以思考, 左边这个形式记号, 有没有一个真正的数学的含义, 即它是否如实地反映了一个 $m\times n$-矩阵和一个 $n\times 1$ 矩阵的某种乘法呢?

在解线性方程的过程中, 我们使用了两种变换: 一是交换两个方程的位置, 二是将一个方程的若干倍加到另一个方程上. 用矩阵表达, 第一个变换是:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \mapsto \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{11} & a_{12} & \cdots & a_{1n} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

第二个变换, 例如将第一行的 e 倍加到第 i 行:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \mapsto \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{i1}+ea_{11} & a_{i2}+ea_{12} & \cdots & a_{in}+ea_{1n} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

和上述一致, 我们也希望用矩阵来表达上述两个过程. 那该怎么处理呢? 我们考虑线性方程右边的变换. 第一个变换, 交换两行:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix} \mapsto \begin{bmatrix} b_i \\ b_2 \\ \vdots \\ b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots \\ 0 & 1 & \cdots & 0 & \cdots \\ & \vdots & & & \\ 1 & 0 & \cdots & 0 & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix}$$

其中, 右边的矩阵是个 $m\times m$ 矩阵 (很重要!), 其对角线除了第 1 和第 i 个位置为 0 外都是 1, 第 1 行第 i 列和第 i 行第 1 列为 1, 其他位置全为 0. 因此我们希望矩阵交换两行可由下面这

个式子表达:

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{11} & a_{12} & \cdots & a_{1n} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots \\ 0 & 1 & \cdots & 0 & \cdots \\ & \vdots & & & \\ 1 & 0 & \cdots & 0 & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

将第一行的 e 倍加到第 i 行的变换为

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix} \mapsto \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i + eb_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots \\ 0 & 1 & \cdots & 0 & \cdots \\ & \vdots & & & \\ e & 0 & 1 & 0 & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix}.$$

右边的矩阵是个 $m \times m$-矩阵 (很重要!). 因此我们希望

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{i1} + ea_{11} & a_{i2} + ea_{12} & \cdots & a_{in} + ea_{1n} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots \\ 0 & 1 & \cdots & 0 & \cdots \\ & \vdots & & & \\ e(i\text{行}1\text{列}) & \cdots & 1 & 0 & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

当然, 我们可以把上面两个变换给稍微一般化:

1. 交换第 i 行和第 j 列的变换由 $m \times m$ 矩阵

$$P(ij) = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots \\ 0 & 1 & \cdots & 0 & \cdots \\ & \vdots & & & \\ 0 & \cdots & & 1(i\text{行}j\text{列}) & \\ & \cdots & & \cdots & \\ 0 & 1(j\text{行}i\text{列}) & \cdots & 0 & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

来表达.

2. 将第 i 行的 e 倍加到第 j 行的变换由 $m \times m$ 矩阵

$$E(ij,e) = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots \\ 0 & 1 & \cdots & 0 & \cdots \\ & \vdots & & & \\ \cdots & e(j\text{行}i\text{列}) & \cdots & 1 & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

来表达.

事实上, 我们还有第三个变换:

3 将第 i 行乘以常数 $e(\neq 0)$ 倍, 这个变换由 $m \times m$ 矩阵

$$C(i,e) = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots \\ 0 & 1 & \cdots & 0 & \cdots \\ & \vdots & & & \\ 0 & \cdots & e(i\text{行}i\text{列}) & 0 & 0 \\ & \vdots & & & \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

来表达.

定义 2.11. 我们把上述三个变换统称为**初等行变换**.

注释 2.12. 我们为什么会想到要“这样希望”呢？一种看法是，既然我们可以把 n-维列向量看成是一个 $n\times 1$-矩阵，那么我们似乎不妨也把一个 $n\times m$-矩阵看成是 m 个 $n\times 1$ 矩阵. 如果这 m 个 $n\times 1$ 矩阵的变化规律是一致的，那么这个变化规律就应该由统一的一个矩阵来表示.

如果能够想清楚上面这个注释，那么我们很自然地能够知道，至少对 $P(ij)$ 和 $E(ij,e)$ 来说，我们的矩阵乘法应该如下定义：设矩阵 X 为某个 $P(ij)$ 或者某个 $E(ij,e)$，若

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix}=[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n]$$

那么

$$XA:=[X\boldsymbol{a}_1,\cdots,X\boldsymbol{a}_n]$$

我们将 X 写成行向量的形式，即 $X=\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_m \end{bmatrix}$，对 $1\le i\le n$，则有

$$X\boldsymbol{a}_i=\begin{bmatrix} \boldsymbol{x}_1^T\cdot\boldsymbol{a}_i \\ \boldsymbol{x}_2^T\cdot\boldsymbol{a}_i \\ \vdots \\ \boldsymbol{x}_n^T\cdot\boldsymbol{a}_i \end{bmatrix},$$

因此

$$XA=[X\boldsymbol{a}_1,\cdots,X\boldsymbol{a}_n]=\begin{bmatrix} \boldsymbol{x}_1^T\cdot\boldsymbol{a}_1 & \boldsymbol{x}_1^T\cdot\boldsymbol{a}_2 & \cdots & \boldsymbol{x}_1^T\cdot\boldsymbol{a}_n \\ \boldsymbol{x}_2^T\cdot\boldsymbol{a}_1 & \boldsymbol{x}_2^T\cdot\boldsymbol{a}_2 & \cdots & \boldsymbol{x}_2^T\cdot\boldsymbol{a}_n \\ \vdots & \vdots & & \vdots \\ \boldsymbol{x}_m^T\cdot\boldsymbol{a}_1 & \boldsymbol{x}_m^T\cdot\boldsymbol{a}_2 & \cdots & \boldsymbol{x}_m^T\cdot\boldsymbol{a}_n \end{bmatrix}. \tag{5}$$

总结一下，到目前为止，我们遇到了如下的矩阵乘法

- $m\times n$ 矩阵乘以 $n\times 1$ 矩阵，
- $m\times m$ 矩阵乘以 $m\times n$ 矩阵，
- $m\times m$ 矩阵乘以 $m\times 1$ 矩阵.

稍加注意就发现有一个共性, 就是前一个矩阵的列数等于后一个矩阵的行数. 根据这些证据, 如果我们考虑定义一般的两个矩阵乘法, 那么第一个矩阵的列数应该等于第二个矩阵的行数. 设 A 是一个 $m \times n$ 矩阵, B 是一个 $n \times k$ 矩阵, 并且我们将 A, B 分别写成:

$$A = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix}, B = [\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_k],$$

模仿 (5) 式, 我们定义

$$AB := \begin{bmatrix} \boldsymbol{a}_1^T \cdot \boldsymbol{b}_1 & \boldsymbol{a}_1^T \cdot \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1^T \cdot \boldsymbol{b}_k \\ \boldsymbol{a}_2^T \cdot \boldsymbol{b}_1 & \boldsymbol{a}_2^T \cdot \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2^T \cdot \boldsymbol{b}_k \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_m^T \cdot \boldsymbol{b}_n & \boldsymbol{a}_m^T \cdot \boldsymbol{b}_n & \cdots & \boldsymbol{a}_m^T \cdot \boldsymbol{b}_k \end{bmatrix}.$$

注意, 在这个矩阵中, 由于 $\boldsymbol{a}_i^T, \boldsymbol{b}_j$ 都是 n 维向量, 因此它们的内积是良定的, 故而 $\boldsymbol{a}_i^T \cdot \boldsymbol{b}_j \in \mathbb{R}$, 所以我们最终得到了一个 $m \times k$ 矩阵. 我们还可以写得更直接一些.

记号 2.13. 对一个矩阵 A, 我们用 A_{ij} 来表示它第 i 行第 j 列的元素. 即, 若

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

那么 $A_{ij} = a_{ij}$.

在这个记号之下, 若

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{ma} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ & \vdots & & \\ b_{n1} & b_{n2} & \cdots & b_{nk} \end{bmatrix}$$

则

$$\boldsymbol{a}_i = [a_{i1}, a_{i2}, \cdots, a_{in}], \boldsymbol{b}_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}, 1 \leq i \leq m, 1 \leq j \leq k,$$

且

$$(AB)_{ij} = \boldsymbol{a}_i^T \cdot \boldsymbol{b}_j = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j}.$$

注释 2.14. 作为一个特殊情形, 一个 $1 \times n$ 矩阵 $\boldsymbol{a}$ 和一个 $n \times 1$ 矩阵 $\boldsymbol{b}$ 相乘, 得到一个 1×1 矩阵, 而 1×1 矩阵天然地和实数没有区别. 而且, 在这样的等同之下, 我们有 $\boldsymbol{ab} = \boldsymbol{a}^T \cdot \boldsymbol{b}$, 这个等式左边是矩阵乘法, 等式右边是两个列向量的点积. 今后我们将一直坚持这样的等同.

练习 2.15. 设 $A = (a_{ij})_{1\le i\le m, 1\le j\le n}$ 是一个 $m \times n$ 矩阵, $P(ij), E(ij,e)$ 为上文所述的 $m \times m$ 矩阵, $1 \le i,j \le m$. 请计算 $P(ij)A$ 和 $E(ij,e)A$.

因为

$$P(ij)_{k\ell} = \begin{cases} 1 & k = \ell \neq i,j \text{或者} \{k,\ell\} = \{i,j\} \\ 0 & \text{否则} \end{cases}$$

因此, 当 $k \neq i,j$ 时, 只有 $t = k$ 时 $P(ij)_{kt} \neq 0$

$$(P(ij)A)_{k\ell} = \sum_{t=1}^{m} P(ij)_{kt} A_{t\ell} = P_{kk} A_{k\ell} = A_{k\ell};$$

当 $k = i$ 时, 只有 $t = j$ 时 $P(ij)_{kt} \neq 0$

$$(P(ij)A)_{k\ell} = (P(ij)A)_{i\ell} = \sum_{t=1}^{m} P(ij)_{it} A_{t\ell} = P(ij)_{ij} A_{j\ell} = A_{j\ell};$$

当 $k = j$ 时, 只有 $t = i$ 时 $P(ij)_{kt} \neq 0$

$$(P(ij)A)_{k\ell} = (P(ij)A)_{j\ell} = \sum_{t=1}^{m} P(ij)_{jt} A_{t\ell} = P(ij)_{ji} A_{i\ell} = A_{i\ell};$$

综上, 矩阵 $P(ij)A$ 由将矩阵 A 的第 i,j 行交换得到.

因为

$$E(ij,e)_{k\ell} = \begin{cases} 1 & k = \ell \\ e & (k,\ell) = (j,i) \\ 0 & \text{否则} \end{cases}$$

因此, 当 $k \neq i$ 时, 只有 $t = k$ 时 $E(ij,e)_{kt} \neq 0$

$$(E(ij,e)A)_{k\ell} = \sum_{t=1}^{m} E(ij,e)_{kt} A_{t\ell} = P_{kk} A_{k\ell} = A_{k\ell};$$

当 $k = j$ 时, 只有 $t = i,j$ 时 $P(ij)_{kt} \neq 0$

$$(E(ij,e)A)_{k\ell} = (E(ij,e)A)_{j\ell} = \sum_{t=1}^{m} E(ij,e)_{it} A_{t\ell} = E(ij,e)_{ji} A_{i\ell} + E(ij,e)_{jj} A_{j\ell} = eA_{i\ell} + A_{j\ell};$$

综上, 矩阵 $E(ij,e)A$ 由将矩阵 A 的第 i 行的 e 倍加到 A 的第 j 行得到.

由上一节 (4) 式知, 对于线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$, 系数矩阵 A 经过初等行变换可以转化为形如

$$\begin{bmatrix} 0 & \cdots & t_{1j_1} & \cdots & t_{1j_2} & \cdots & t_{1j_r} & \cdots & t_{1n} \\ 0 & \cdots & \cdots & 0 & t_{2j_2} & \cdots & t_{1j_r} & \cdots & t_{2n} \\ & & & & \ddots & & \vdots & & \vdots \\ & & & & & \ddots & t_{rj_r} & \cdots & t_{rn} \\ & & & & & & & & 0 \\ & & & & & & & & 0 \end{bmatrix}$$

的矩阵. 这样的矩阵我们称为阶梯形矩阵. 它的特征是: 每一行的首个非零元的位置都在前一行的非零元右边.

定义 2.16. 我们把矩阵经过一系列初等行变换后转化成的阶梯形矩阵的非零行的行数称为该矩阵的秩.

如果把系数矩阵和方程组等号右端的列向量放在一起, 得到的新矩阵

$$A' := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{ma} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

称为增广矩阵.

定理 2.17. 一个线性方程组有解, 当且仅当它的系数矩阵的秩等于增广矩阵的秩. 并且, 在有解的时候, 若系数矩阵的秩小于未知数的个数, 那么方程组有无穷多个解.

证明. 假设系数矩阵 A 经过一系列的初等行变换变成了阶梯形矩阵, 那么列向量 $\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ 在同样的一系列列变换之下变成了 $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$. 因此这个定理只是上一节注释 2.5 之前的判断的重新叙述. □

例题 2.18. 设 A 是一个 $n \times n$ 矩阵, 它对角线非零, 对角线下方全为零. 这样的矩阵我们成为上三角矩阵. 上三角矩阵都是阶梯形矩阵. 那么, 对任意列向量 $\boldsymbol{b}$, 方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 都存在唯一解. 事实上, 设 $A = (a_{ij})$, 满足

$$a_{ii} \neq 0, \quad \forall i; \quad a_{ij} = 0, \quad \forall i > j.$$

那么方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 就是

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = b_1 \\ \qquad\quad a_{22}x_2 + \cdots + a_{2n}x_n & = b_2 \\ \qquad\qquad\qquad\qquad\quad \vdots & \\ \qquad\qquad\qquad\qquad a_{nn}x_n & = b_n \end{cases}$$

问题 2.19. 我们把一个矩阵通过初等行变换转化成了阶梯形矩阵. 但是需要注意到, 我们有很多种行变换的方式, 都能使的该矩阵最终能转化成阶梯形矩阵. 例如, $j_1 = 1$ 时, 我们要在 x_1 的系数中, 选取一个非零系数. 很明显, 这样的非零系数可能有很多选择. 所以最终的阶梯形矩阵不是一个固定的矩阵. 那么所有我们能获得的阶梯形矩阵的非零行的个数是固定的吗? 如果不固定, 这说明我们定义 2.13 是没有意义的. 请同学们思考这个问题.

第三章 矩阵的运算

3.1 矩阵的加法和乘法

我们把实数 $\mathbb{R}$ 上所有的 $m \times n$ 矩阵组成的集合记作 $M_{m\times n}(\mathbb{R})$; 特别的, 当 $m = n$ 时, 我们简记 $M_{n\times n}(\mathbb{R})$ 为 $M_n(\mathbb{R})$. 我们有一个显然的集合上的一一对应:

$$M_{m\times n}(\mathbb{R}) \longrightarrow \mathbb{R}^{mn}, \quad (a_{ij}) \mapsto \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \\ a_{21} \\ \vdots \\ a_{mn} \end{bmatrix}$$

由这个对应, 我们把 $M_{m\times n}(\mathbb{R})$ 也看成是一个线性空间, 它的加法和乘法通过 $\mathbb{R}^{mn}$ 上的加法和乘法转化过来. 用 $M_{m\times n}(\mathbb{R})$ 自身的语言来写, 那就是: 对任意 $A = (a_{ij}), B = (b_{ij}) \in M_{m\times n}(\mathbb{R}), a, b \in \mathbb{R}$,

$$aA + bB = (aa_{ij} + bb_{ij})$$

在上一章中, 我们定义了矩阵乘法, 因此, 我们有映射

$$M_{m\times n}(\mathbb{R}) \times M_{n\times k}(\mathbb{R}) \longrightarrow M_{m\times k}(\mathbb{R}), \quad (A, B) \mapsto AB$$

命题 3.1. 设 $A, B \in M_{m\times n}(\mathbb{R}), C, D \in M_{n\times k}$, $a, b \in \mathbb{R}$, 那么

$$\begin{aligned} (aA + bB)C &= (aA)C + (bB)C = a(AC) + b(BC) =: aAC + bBC; \\ A(aC + bD) &= A(aC) + A(bD) = a(AC) + b(AD) =: aAC + bAD. \end{aligned}$$

证明很简单, 直接计算即可.

命题 3.2. 设 $A \in M_{m\times n}(\mathbb{R}), B \in M_{n\times k}(\mathbb{R}), C \in M_{k\times r}(\mathbb{R})$, 那么

$$(AB)C = A(BC) \in M_{m\times r}(\mathbb{R}).$$

证明. 我们只需证明, 对任意的 $1 \le i \le m, 1 \le j \le r$, 都有 $((AB)C)_{ij} = (A(BC))_{ij}$. 而

$$\begin{aligned} ((AB)C)_{ij} &= \sum_{\ell=1}^{k}(AB)_{i\ell}C_{\ell j} = \sum_{\ell=1}^{k}\left(\sum_{t=1}^{n}A_{it}B_{t\ell}\right)C_{\ell j} \\ &= \sum_{t=1}^{n}A_{it}\left(\sum_{\ell=1}^{k}B_{t\ell}C_{\ell j}\right) = \sum_{t=1}^{n}A_{it}(BC)_{tj} \\ &= (A(BC))_{ij} \end{aligned}.$$

证毕 □

注释 3.3. 一般的, 若 $A, B \in M_{n\times n}(\mathbb{R})$, 那么 AB, BA 都属于 $M_{n\times n}(\mathbb{R})$; 但是, 一般的 $AB \neq BA$. 例如 $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, 而 $AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, BA = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

现在我们考虑 $M_n(\mathbb{R})$. 记 $I_n \in M_n(\mathbb{R})$ 是对角线全为 1 其他位置全为 0 的矩阵. 那么显然, 对任意 $A \in M_n(\mathbb{R})$, 我们有

$$AI_n = I_nA = A.$$

注释 3.4. 对任意的 $a \in \mathbb{R}$, 记 $I(a) = \begin{bmatrix} a & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & a \end{bmatrix} \in M_n(\mathbb{R})$ 是对角线全为 a 其他位置全为 0 的矩阵, 那么对任意的 $A \in M_n(\mathbb{R})$, 我们都有

$$AI(a) = I(a)A = aA.$$

定义 3.5. 矩阵 $A \in M_n(\mathbb{R})$. 我们称 A 有右逆, 如果存在 $B \in M_n(\mathbb{R})$ 使得 $AB = I_n$, B 称为 A 的一个右逆. 我们称 A 有左逆, 如果存在 $C \in M_n(\mathbb{R})$ 使得 $CA = I_n$, C 称为 A 的一个左逆. 我们称 A 可逆, 如果它既有左逆又有右逆.

例题 3.6. 初等行变换对应的矩阵都是可逆矩阵, 而且 $P(ij)$ 的左逆和右逆都是 $P(ij)$, $E(ij, e)$ 的左逆和右逆都是 $E(ij, -e)$, $C(i, e)$ (若 $e \neq 0$) 的左逆和右逆都是 $C(i, e^{-1})$.

注意, 我们定义: 一个矩阵 $A \in M_n(\mathbb{R})$ 可逆, 是说它既有左逆 C 又有右逆 B.

命题 3.7. 若矩阵 $A \in M_n(\mathbb{R})$ 可逆, 那么它的左逆和右逆都唯一并且左逆右逆相等.

证明. 设 B, B' 是 A 的两个右逆, C 是 A 的一个左逆, 那么

$$B = I_n B = (CA)B = C(AB) = CI_n = C;$$

同理 $B' = C$, 因此 $B = B' = C$. 同理, 若 C' 是 A 的左逆, 那么 $C' = B$. □

所以我们可以用统一的记号 A^{-1} 来表示 A 的左逆和右逆.

推论 3.8. 考虑有 n 个未知数, n 个方程的方程组

$$A\boldsymbol{x} = \boldsymbol{b};$$

若 A 是可逆矩阵, 那么它有唯一解 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

证明. 首先, A^{-1} 是 A 的右逆, 所以

$$A(A^{-1}\boldsymbol{b}) = (AA^{-1})\boldsymbol{b} = \boldsymbol{b},$$

也就是说 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ 是方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的一组解.

因为 A 有左逆, 且左逆唯一, 从而若方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 那么它的解 $\boldsymbol{x}$ 满足

$$\boldsymbol{x} = (A^{-1}A)\boldsymbol{x} = A^{-1}(A\boldsymbol{x}) = A^{-1}\boldsymbol{b};$$

也就是说, 解是唯一的. □

命题 3.9. 若 A, B 都可逆, 那么 AB 也可逆, 并且 $(AB)^{-1} = B^{-1}A^{-1}$.

证明. 这是因为

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = I_n.$$

以及

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

□

推论 3.10. 任意的初等行变换矩阵的乘积都是可逆矩阵.

那矩阵 A 会不会只有左逆没有右逆, 或者只有右逆没有左逆呢? 答案是否定的.

命题 3.11. 若 $A \in M_n(\mathbb{R})$ 有左逆, 那么它有右逆, 因此 A 可逆.

证明. 如果 A 有左逆 C: $CA = I_n$, 那么任意的方程组

$$A\boldsymbol{x} = \boldsymbol{b}$$

的解都存在, 而且等于 $\boldsymbol{x} = C\boldsymbol{b}$; 这一个论断见下面的注释 3.13和引理 3.14. 特别的, 当

$$\boldsymbol{b} = \boldsymbol{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix};$$

第 i 行的分量为 1, 其他分量均为 0; $A\boldsymbol{x} = \boldsymbol{e}_i$ 也有解, 我们把这个解记做 $\boldsymbol{x}_i$. 令 $B = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]$, 根据矩阵乘法的定义就有

$$AB = [A\boldsymbol{x}_1, A\boldsymbol{x}_2, \cdots, A\boldsymbol{x}_n] = [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n] = I_n,$$

从而 A 有右逆. 因此 A 既有左逆又有右逆, 也就是说 A 可逆. □

推论 3.12. *若 A 有右逆, 那么它有左逆, 因此 A 可逆.*

证明. 现在假设 A 有右逆 B: $AB = I_n$. 那么 A 就是 B 的左逆, 等价的说, B 有左逆 A; 对 B 应用命题 3.11, 知 B 有右逆, 所以 B 可逆. 设 A' 是 B 的右逆: $BA' = I_n$, 那么 $A = A'$, 即 $BA = I_n$, 所以 A 有左逆, 并且左逆也是 B. □

注释 3.13. *请注意, 若 A 有左逆, 那么方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 至多只有一个解. 至多只有一个解是什么意思呢? 意思就是: 要么没有解, 要么有唯一解; 再说得难懂一点, 那就是如果有解, 那么解一定唯一. 这个论断的证明如下:*

假设 $\Phi_{\boldsymbol{b}} = \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}\} \neq \emptyset$, 即方程的解存在; 任取 A 的左逆 C, 那么对任意的 $\boldsymbol{x} \in \Phi_{\boldsymbol{b}}$, 我们有

$$\boldsymbol{x} = I_n\boldsymbol{x} = CA\boldsymbol{x} = C\boldsymbol{b}.$$

也就是说, $\Phi_{\boldsymbol{b}} = \{C\boldsymbol{b}\}$. 但是, 我们并没有先验地知道解存在: 也就是说, 给一个 $\boldsymbol{b}$, $\Phi_{\boldsymbol{b}}$ 是否是空集, 我们并没有事先知道.

另外我们还需要明白: 取 A 的另一个左逆 C', 我们并得不到 $\Phi_{\boldsymbol{b}}$ 中的“另一个”元素 $C'\boldsymbol{b}$. 这里逻辑上的难点就在于, 即使 $C \neq C'$, 我们也不能断定 $C\boldsymbol{b} \neq C'\boldsymbol{b}$ (千万不要想当然地认为这是成立的). 而是类似上面同样的推导知道 $\Phi_{\boldsymbol{b}} = \{C'\boldsymbol{b}\}$; 从而 $C\boldsymbol{b} = C'\boldsymbol{b}$.

我们也可以形式地描述这句话. 我们可以把 $\mathbb{R}^n$ 分成两个不想交的集合 S_0, S_1:

$$S_0 = \{\boldsymbol{b} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}\text{没有解}, \textit{ i.e. } \Phi_{\boldsymbol{b}} = \varnothing\}$$
$$S_1 = \{\boldsymbol{b} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{b}\text{有解}, \textit{ i.e. } \Phi_{\boldsymbol{b}} \neq \varnothing\}$$

若 A 有左逆, 那么对任意的 $\boldsymbol{b} \in S_1$, $\Phi_{\boldsymbol{b}}$ 是只含有一个元素的非空集合.

引理 3.14. 假设 $A \in M_n(\mathbb{R})$ 有左逆, 那么对任意的 n 维列向量 $\boldsymbol{b}$, 方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的解都存在.

这个引理的意思是说, 若 A 有左逆, 那么注释 3.13中的集合 $S_0 = \varnothing$, 或者等价的 $S_1 = \mathbb{R}^n$.

证明. 设 X 是一些初等行变换矩阵的乘积, 使得 XA 是阶梯形矩阵, 那么方程组 $XA\boldsymbol{x} = X\boldsymbol{b}$ 和方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的解空间相同. 我们断言:

- XA 是对角线非零的上三角矩阵.

若不然, 这个阶梯形矩阵 XA 的最后一行全为 0. 取 $\boldsymbol{b} = \boldsymbol{0}$, 考虑方程组 $A\boldsymbol{x} = \boldsymbol{0}$; 那么 $XA\boldsymbol{x} = \boldsymbol{0}$ 有无穷多个解. 这与注释 3.13矛盾.

再次第注意到: 因为 X 是一些初等行变换矩阵, 因此方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 有解等价于 $XA\boldsymbol{x} = X\boldsymbol{b}$ 有解. 由上述断言, 我们知道 $XA\boldsymbol{x} = X\boldsymbol{b}$ 对任意 $\boldsymbol{b}$ 都有解. 所以 $A\boldsymbol{x} = \boldsymbol{b}$ 有解. □

推论 3.15. 矩阵 $A \in M_n(\mathbb{R})$; 那么下面三条等价:

1. A 可逆;

2. A 有左逆;

3. A 有右逆.

因此在今后, 我们将不再区分这三个术语.

另一方面, 由引理 3.13 的证明, 我们知道一个可逆矩阵, 通过一些初等行变换能转化成的阶梯形矩阵是对角线非零的上三角矩阵. 那么显然, 再经过一些初等行变换, 就能转化成单位矩阵. 例如:

$$\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix} \mapsto \begin{bmatrix} 1 & \frac{3}{2} \\ 0 & 1 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};$$

其中第一个映射是将第一行和第二行的对角线元素都乘以合适的倍数变成 1, 第二个映射是将第二行的合适的倍数加到第一行. 这说明, 任何一个可逆矩阵, 都可以通过一系列的初等行变换, 变成单位阵. 这给了我们求解一个矩阵的逆矩阵的一种方式:

命题 3.16 (Gauss-Jordan 消去). *设 $A \in M_n(\mathbb{R})$ 是可逆矩阵. 考虑 $n \times 2n$ 矩阵 $X = [A, I_n]$. 假设我们对矩阵 X 作初等行变换, 使得 A 变成了 I_n, 那么 I_n 被作了同样的初等行变换后得到的矩阵就是 A 的逆矩阵.*

证明. 我们只需要将这个命题重新叙述一下. 假设 P 是对 X 作的初等行变换使得 A 变成了 I_n, 也就是说 $PX = \begin{bmatrix} I_n & B \end{bmatrix}$. 我们需要证明的是 $BA = I_n$. 而这是因为

$$PX = P\begin{bmatrix} A & I_n \end{bmatrix} = \begin{bmatrix} PA & PI_n \end{bmatrix} = \begin{bmatrix} I_n & B \end{bmatrix},$$

从而 $PA = I_n, PI_n = B$. 证毕. □

例题 3.17. *求下列矩阵的逆矩阵*

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 2 & -1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 2 \\ 1 & 1 & 1 & -1 \\ 1 & 0 & -2 & -6 \end{bmatrix}$$

另外, 命题 3.11 的证明, 我们先证明了若 A 有左逆, 则 A 有右逆; 然后用已经证明的这一条来证明, 若 A 有右逆, 则 A 有左逆. 事实上, 我们也可以反过来, 直接先证明:

命题 3.18. *若 A 有右逆, 则 A 有左逆.*

证明. 设 B 是 A 的右逆, 那么对任意的 $\boldsymbol{b}$ 和方程组 $A\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{x} = B\boldsymbol{b}$ 都是它的一个解. 设 X 是一些初等行变换的乘积, 使得 XA 是行简化阶梯型矩阵. 我们断言

- XA 是对角线非零的上三角矩阵.

我们用反证法证明这个断言. 若不然, XA 的最后一行全为零. 由于 X 是一些初等行变换矩阵的乘积, 因此我们可以取到 $\boldsymbol{b}_0$, 使得

$$X\boldsymbol{b}_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

(我们可以用两种方式去考虑这件事情, 因为 X 是初等行变换的乘积, 其中的每个步骤, 都可以返回; 所以可以从 $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$ 出发逐步返回得到我们想要的 $\boldsymbol{b}_0$; 或者我们已知 X 是逆矩阵, 那么

取 $\boldsymbol{b}_0 = X^{-1}\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$ 即可.) 这样我们就看出来 $XA\boldsymbol{x} = X\boldsymbol{b}_0$ 没有解, 也就是说 $A\boldsymbol{x} = \boldsymbol{b}_0$ 没有解. 矛盾. 因此我们的断言成立.

从这个断言出发, 因为 XA 是对角线非零的上三角矩阵, 我们对 XA 作一系列的初等行变换 Y, 得到对角线非零的对角矩阵 Z, 也就是说 $YXA = Z$. 显然 Z 是可逆矩阵, 因此 $Z^{-1}YXA = I_n$. 所以 $Z^{-1}YX$ 是 A 的左逆. 证毕. □

同样的, 从这个命题出发, 我们可以证明 "若 A 有左逆, 则 A 有右逆." 但是无论从哪种角度出发, 通过高斯消元法把 A 转化为行简化阶梯型矩阵是关键步骤.

3.2 矩阵的 LU 分解

定义 3.19. 矩阵 $X \in M_n(\mathbb{R})$ 称为是上三角矩阵, 若对任意的 $i > j$, $X_{ij} = 0$; 称为是下三角矩阵, 若对任意的 $i < j$, $X_{ij} = 0$; 称为是对角矩阵, 若对任意的 $i \neq j$, $X_{ij} = 0$.

因此上三角矩阵和下三角矩阵分别形如

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \text{和} \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{12} & a_{22} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

定义 3.20. 矩阵 $X \in M_n(\mathbb{R})$ 称为是置换矩阵, 如果 X 的每一行每一列都存在唯一的一个元素非零, 且该非零元等于 1.

注释 3.21. $M_n(\mathbb{R})$ 中的置换矩阵有 $n!$ 个.

为什么这样的矩阵被称为置换矩阵呢?

命题 3.22. 任何置换矩阵 X 都是一些 $P(ij)$ 的乘积.

证明. 首先我们看第一列. 第一列存在唯一的非零元 1, 设它在第 i_1 行, 那么我们交换第 1 行和第 i_1 行, 得到一个新矩阵. 这个新矩阵也是置换矩阵, 但是它第 1 行第 1 列是 1, 也就是说

$$(P(1, i_1)X)_{11} = 1.$$

那么我们再看第二列, 直至第 n 列, 我们有

$$P(n,i_n)P(n-1,i_{n-1})\cdots P(1i_1)X = I_n.$$

而 $P(ij)^2 = I_n$, 所以 $X = P(1i_1)P(2i_2)\cdots P(ni_n)$. 证毕. □

1. 阶梯形矩阵是上三角矩阵;
2. 初等行变换矩阵 $E(ij,e)$: 当 $i<j$ 时是下三角矩阵; 当 $i>j$ 时是上三角矩阵. (请注意 i,j 的大小关系).
3. 初等行变换矩阵 $C(i,e)$ 是对角矩阵.

注释 3.23. 显而易见的, 上 (下) 三角矩阵的和, 乘积都是上 (下) 三角矩阵. 请同学们自行验证这一点.

记号 3.24. 上三角矩阵我们一般用 U 来表示, (英语 upper triangular); 下三角矩阵我们一般用 L 来表示, (英语 lower triangular); 往往下三角矩阵我们指对角线为 1 的下三角矩阵.

命题 3.25. 一个上 (下) 三角矩阵是可逆矩阵, 当且仅当它的对角线元素都非零.

证明. 必要性. 若一个上三角矩阵 U 的某个对角线元素为零, 那么根据 Gauss 消元法, 它作一些行变换之后得到的阶梯形矩阵的最后一行为 0. 因此任意方程组 $U\boldsymbol{x}=\boldsymbol{b}$ 要么无解, 要么有无穷多解; 进而矩阵 U 不可逆.

充分性. 若一个上三角矩阵 U 的所有对角线元素均不为零. 那么根据 Gauss-Jordan 消去, 我们就知道 U 是可逆矩阵.

对于下三角矩阵, 我们可以利用后面要讲的矩阵转置和上三角的结论来论证: 只需要注意到上 (下) 三角矩阵的转置为下 (上) 三角矩阵. 证明在此略过. □

推论 3.26. 一个可逆的上 (下) 三角矩阵它的逆也是上三角矩阵. 置换矩阵的逆矩阵也是置换矩阵.

证明. 一种证明是写出逆矩阵, 逐个分析逆矩阵中元素是零或者非零.

另一种证明是用 Gauss-Jordan 消去; 例如, 对上三角矩阵, 我们只需要注意到用 Gauus-Jordan 消去时所作的行变换对应的矩阵都是上三角矩阵即可. 在此不再赘述. □

现在我们考虑可逆矩阵 A. 根据高斯消元法, 我们知道存在一些初等行变换矩阵 $E's, P's$ 使得它们的乘积再乘以 A 得到上三角矩阵 U. 由于初等行变换矩阵和 A 都可逆, 因此 U 也可逆. 初等行变换矩阵的逆也是初等行变换矩阵, 因此, 我们有

$$A = XU,$$

其中 X 是一些 $E's, P's$ 的乘积.

有一种情况比较理想: 我们在做消元的时候, 不需要交换两行的操作, 那么所有的消元步骤只有将第 i 行的若干倍加到第 j 行, 其中 $i<j$. 也就是说, 那些 E 都是 $E(ij,e)(i<j)$, 从而是对角线为 1 的下三角矩阵. 那么这些 E 的乘积也是对角线为 1 的下三角矩阵. 从而 A 可以写成

$$A = LU$$

的形式. 此时, 我们有两个自然的问题:

1. 是否所有的可逆矩阵 A 都可以写成 LU 的形式? 其中 L 是下三角矩阵, U 是上三角矩阵?

2. 是否所有的对角线为 1 的下三角矩阵都可以写成某些 $E(ij,c)(i<j)$ 的乘积?

第一个问题的答案是否定的. 例如考虑 $A=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. 因为 $A^2=I_2$, 所以 A 是可逆矩阵. 假设 $A=LU$, 其中

$$L=\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}, U=\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}, ad \neq 0;$$

那么

$$LU=\begin{bmatrix} a & b \\ ac & bc+d \end{bmatrix};$$

因为 $a\neq 0$, 所以 $LU\neq A$. 那么这是什么意思呢? 这其实是说明了在做高斯消元法把一个矩阵转化为阶梯形矩阵的过程中, 交换两行的操作是必要的. 但是, 先验地, 我们只知道这些个交换两行的操作是夹杂在 $E's$ 中的.

问题 3.27. *我们是否可以把下面三种变换完全分开:*

1. *把第 i 行的若干倍加到第 j 行 $(i<j)$, 其矩阵是 $E(ij,c)(i<j)$ 下三角矩阵;*

2. *把第 i 行的若干倍加到第 j 行 $(i>j)$, 其矩阵是 $E(ij,c)(i>j)$ 上三角矩阵;*

3. *交换某些行, 其矩阵是 P, 置换矩阵;*

是否存在 P, L, U 使得 $A=PLU, LUP, LPU$, e.t.c ?

在实际应用当中, 我们通常只关注其中的两个问题: 对可逆矩阵 A, 是否存在转置矩阵 P, 上三角矩阵 U, 下三角矩阵 L, 使得

$$A = PLU, resp. LUP?$$

这个问题的答案都是肯定的. 我们看这个例子:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 5 \\ 1 & 3 & 4 & 5 \\ 1 & 3 & 5 & 7 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$\longrightarrow \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

它的严格证明, 在我们学习了分块矩阵的乘法之后再给出.

第二个问题的答案是肯定的, 也是在我们学习了分块矩阵的乘法之后给出.

注释 3.28. 假设 A 是可逆矩阵, 可以写成 LU 的形式, 其中 L 是对角线为 1 的下三角矩阵, U 是对角线全非零的上三角矩阵. 有的时候, 为了形式上的对称性, 我们可以进一步的把 U 写成 DU' 的形式, 其中 D 是一个对角线全非零的对角矩阵, U' 是对角线全为 1 的上三角矩阵. 从而 $A = LDU'$. 这样处理, 至少有个形式地好处, 那就是对称. 比如说, 我们对 A 作转置, 学习了矩阵转置我们会知道, $A^T = U'^T D^T L^T$ 本身已经满足形式上的要求, 即: U'^T 是对角线为 1 的下三角矩阵, D^T 是对角线非零的对角矩阵, L^T 是对角线为 1 的上三角矩阵.

注释 3.29. 那么对称性又有什么好处呢? 在实际的处理当中, 一旦有对称性, 就可以减少运算量. 我们可以这么感觉一下: 如果我们能找到 n 个独立的对称性, 那么就可以把运算量降为原来的 2^{-n}. 这是很可观的. 例如我们将要讨论的对称矩阵, 因为它有一个对称性, 那么实际上做一些运算的时候, 运算量是一般矩阵的一半.

3.3 分块矩阵的乘法

设 $A \in M_{m\times n}(\mathbb{R}), B \in M_{n\times k}(\mathbb{R})$. 我们可以把矩阵 A 切割成一些小矩阵, 例如, 将 A 的行切割成 r 组, 每组的行数分别是 $m_1, m_2, \cdots, m_r$; 将 A 的列切割成 s 组, 每组的列数分别是 $n_1, n_2, \cdots, n_s$. 于是

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rs} \end{bmatrix}$$

其中 A_{ij} 为 $m_i \times n_j$ 矩阵.

对 $1 \le i \le r, 1 \le j \le s$, 我们记

$$m'_i = m_1 + \cdots m_{i-1}, \quad n'_j = n_1 + \cdots + n_j,.$$

暂时地, 为了避免符号的混乱, 我们把矩阵 X 的第 u 行第 v 列的分量记作 $X_{(uv)}$. 注意, 这个符号与以前不同, 以前的符号下标没有括号. 那么根据分块矩阵的定义, A_{ij} 的第 u 行是 A 的第 $m_1 + m_2 + \cdots + m_{i-1} + u = m'_{i-1} + u$ 行, 它的第 v 列是 A 的第 $n_1 + n_2 + \cdots + n_{j-1} + v = n'_{j-1} + v$ 列, 所以我们有

$$(A_{ij})_{(uv)} = A_{m'_{i-1}+u, n'_{j-1}+v}.$$

例题 3.30. 我们把矩阵看成是一些由列向量组成的, 这就是一种分块: 将行分成 1 块, 列分成 n 块; 矩阵本身就可以看成是一个分块矩阵: 行分成 m 块, 列分成 n 块.

例题 3.31. 我们可以将 3×3 矩阵

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

如下分块:

$$\left[\begin{array}{c:cc} 1 & 2 & 3 \\ \hdashline 0 & 1 & 1 \\ 0 & 2 & 4 \end{array}\right] \text{ 或者 } \left[\begin{array}{c:cc} 1 & 2 & 3 \\ 0 & 1 & 1 \\ \hdashline 0 & 2 & 4 \end{array}\right]$$

明显地, 第一种分块的方式优于第二种, 因为它把所有的 0 都集中在一起了, 这在计算中会给我们提供很多便利.

类似的, 我们对 B 作分割, 但是, 要求 B **的行的分割方式和** A **的列的分割方式相同**.

注释 3.32. 我们为什么作这个要求呢? 对于朴素的矩阵而言, 为了研究矩阵乘法, 我们要求后一个矩阵的行数等于后一个矩阵的列数; 当我们研究矩阵的分块的时候, 我们把分块后的" 小矩阵" 看成是组成矩阵的单元, 那么第一步自然地会要求后一个矩阵行数的分块的数目等于前一个矩阵列数的分块的数目; 第二步地深入研究" 小矩阵" 之间的关系, 就会完整地提出这个要求了: 即前一个矩阵分块后的小矩阵的列数等于后一个矩阵分块后的相应位置的小矩阵的行数.

注意, 对 B 的列的分割没有要求. 于是

$$B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ B_{21} & B_{22} & \cdots & B_{2t} \\ \vdots & \vdots & & \vdots \\ B_{s1} & B_{s2} & \cdots & B_{rt} \end{bmatrix}.$$

命题 3.33. 若 $AB = C$, 那么 C 有如下的分块形式:

$$C = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1t} \\ C_{21} & C_{22} & \cdots & C_{2t} \\ \vdots & \vdots & & \vdots \\ C_{r1} & C_{r2} & \cdots & C_{rt} \end{bmatrix},$$

其中 C_{ij} 是 $m_i \times k_j$ 矩阵, 且

$$C_{ij} = \sum_{\ell=1}^{s} A_{i\ell} B_{\ell j}.$$

证明. 设命题中右边的分块矩阵为 C', 我们只需验证 C 和 C' 的任何一个分量都相等. 我们设 $A = (a_{ij})_{\substack{1\le i\le m \\ 1\le j\le n}}, B = (b_{ij})_{\substack{1\le i\le n \\ 1\le j\le k}}, C = (a_{ij})_{\substack{1\le i\le m \\ 1\le j\le k}}, C' = (c'_{ij})_{\substack{1\le i\le m \\ 1\le j\le k}}$. 首先由矩阵乘法的定义, 我们有

$$c_{uv} = \sum_{w=1}^{n} a_{uw} b_{wv};$$

又因为存在 $1 \le i \le r, 1 \le j \le t$, 使得 $m_1 + \cdots + m_{i-1} = m'_{i-1} < u \le m_1 + \cdots + m_i = m'_i$, $k_1 + \cdots + k_{j-1} = k'_{j-1} < v \le k_1 + \cdots + k_j = k'_j$, 那么 $c_{uv} \in C_{ij}$. 记 $u' = u - m_{i-1}, v' = v - k'_{j-1}$.

$$\begin{aligned} c_{uv} &= a_{u1}b_{1v} + \cdots + a_{un_1}b_{n_1v} + a_{u,n_1+1}b_{n_1+1,v} + \cdots + a_{un_2}b_{n_2v} + \cdots + a_{u,n_{r-1}+1}b_{n_{r-1}+1,v} + \cdots + a_{un_r}b_{n_rv} \\ &= \sum_{\ell=1}^{s}\left(\sum_{n'_{\ell-1}<w\le n'_\ell} a_{uw}b_{wv}\right) \\ &= \sum_{\ell=1}^{s}\sum_{1\le w\le n_\ell} (A_{i\ell})_{u'w}(B_{\ell j})_{wv'} \\ &= \sum_{\ell=1}^{s}(A_{i\ell}B_{\ell j})_{u'v'} \\ &= c'_{uv} \end{aligned}$$

其中 $u' = u - (m_1 + \cdots + m_{i-1}), v' = v - (k_1 + \cdots k_{j-1})$.

$\square$

推论 3.34. 若 $A \in M_m(\mathbb{R}), B = M_n(\mathbb{R})$ 都是可逆矩阵, 那么对于任意的 $C \in M_{m\times n}$, $(m+n) \times (m+n)$-矩阵

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$$

也是可逆矩阵.

直接地验证, $\begin{bmatrix} A^{-1} & -A^{-1}CB^{-1} \\ 0 & B^{-1} \end{bmatrix}$ 是 $\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$ 的逆矩阵.

现在, 我们来证明上一节中未证明的命题:

命题 3.35. 设 A 是可逆矩阵, 存在置换矩阵 P, 对角线为 1 的下三角矩阵 L 和下三角矩阵 U, 使得 $A = PLU$;

证明. 我们对矩阵行数 n 用归纳法. $n = 1, 2$ 时, 命题是显然的. 现在假设原命题对 $n-1$ 时成立. 首先注意到, 由于 A 可逆, 因此第一列存在非零元, 取一个置换矩阵 P_1, 把这个非零元挪到第一行第一列, 根据消元法, 存在 $E_1 = \begin{bmatrix} 1 & 0 \\ \boldsymbol{v}_1 & I_{n-1} \end{bmatrix}$, 使得

$$E_1 P_1 A = \begin{bmatrix} a'_{11} & \boldsymbol{v} \\ 0 & A' \end{bmatrix};$$

其中 A' 是可逆的 $(n-1) \times (n-1)$-方阵. 由归纳假设, $A' = P'E'U'$, 所以

$$\begin{bmatrix} a'_{11} & \boldsymbol{v} \\ 0 & A' \end{bmatrix} = \begin{bmatrix} a'_{11} & \boldsymbol{v} \\ 0 & P'E'U' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & E' \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{v} \\ 0 & U' \end{bmatrix};$$

所以

$$\begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix} E_1 P_1 A = \begin{bmatrix} 1 & 0 \\ 0 & E' \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{v} \\ 0 & U' \end{bmatrix}.$$

如果 $\begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix} E_1 = E_1 \begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix}$, 那我们就大功告成了. 但是很遗憾, 它们的乘法并不交换. (怎么办??) 我们尝试一下:

$$\begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix} E_1 = \begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \boldsymbol{v}_1 & I_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ P'\boldsymbol{v}_1 & P' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ P'\boldsymbol{v}_1 & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix},$$

所以

$$\begin{bmatrix} 1 & 0 \\ P'\boldsymbol{v}_1 & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix} P_1 A = \begin{bmatrix} 1 & 0 \\ 0 & E' \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{v} \\ 0 & U' \end{bmatrix},$$

从而, 我们令

$$P = \left(\begin{bmatrix} 1 & 0 \\ 0 & P' \end{bmatrix} P_1 \right)^{-1}, L = \begin{bmatrix} 1 & 0 \\ P'\boldsymbol{v}_1 & I_{n-1} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & E' \end{bmatrix}, U = \begin{bmatrix} 1 & \boldsymbol{v} \\ 0 & U' \end{bmatrix}$$

即可. 证毕. □

命题 3.36. 设 $L \in M_{n\times n}(\mathbb{R})$ 是对角线为 1 的下三角矩阵, 存在 $E(i_1 j_1, e_1), \cdots, E(i_\ell j_\ell, e_\ell)$, 使得 $L = E(i_1 j_1, e_1) \cdots E(i_\ell j_\ell, e_\ell)$.

证明类似于上一个命题的证明, 也是对矩阵的行数做归纳法. 留给同学们做习题.

最后, 我们要问分块矩阵的运算有什么实际的具体用途呢? 我个人认为, 实际运算中可能没啥用, 但是理论的运算, 能简化很多, 特别是当某个矩阵有很多 0 的时候, 我们可以依据 0 的分布, 对矩阵做合适的分块, 以简化运算. 比如说上面推论 3.25. 再比如, 我们可以用分块矩阵作" 集体消元"; 比如说上面两个命题.

3.4 矩阵的转置

在之前的学习中, 我们已经碰到了行向量和列向量, 以及我们如何把行向量和列向量进行互换. 互换的方式就是" 站起来" 和" 躺下去". 那么这种互换的方式推广到一般的矩阵, 就是所谓的矩阵:

定义 3.37. 设 $A \in M_{m\times n} = (a_{ij})_{\substack{1\le i\le m\\1\le j\le n}}$. 它的置换矩阵, 我们记作 A^T, 是一个 $n\times m$-矩阵: $A^T = (a'_{ij})_{\substack{1\le i\le n\\1\le j\le m}}$, 满足: 对任意的 $1\le i\le n, 1\le j\le m$, 都有

$$a'_{ij} = a_{ji}.$$

例题 3.38. $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, 那么 $A^T = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$; 若 $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, 那么 $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

例题 3.39. 对于初等行变换矩阵, $P(ij)^T = P(ji) = P(ij), E(ij,e) = E(ji,e), E(i,e) = E(i,e)$.

命题 3.40. 1. 设 $A \in M_{m\times n}(\mathbb{R}), B \in M_{n\times k}(\mathbb{R})$. 那么 $(AB)^T = B^T A^T \in M_{k\times m}(\mathbb{R})$.

2. 设 $A \in M_m(\mathbb{R})$ 是一个可逆矩阵, 那么 $(A^T)^{-1} = (A^{-1})^T$

证明. (2) 是 (1) 的推论. 这是因为, 如果我们假设 (1) 成立, 就有

$$(A^T)((A^{-1})^T) = (A^{-1}A)^T = I_n^T = I_n.$$

(1) 的证明如下. 对任意 $1\le i\le k, 1\le j\le m$,

$$(B^TA^T)_{ij} = \sum_{\ell=1}^n B^T_{i\ell}A^T_{\ell j} = \sum_{\ell=1}^n B_{\ell i}A_{j\ell} = \sum_{\ell=1}^n A_{j\ell}B_{\ell i} = (AB)_{ji} = (AB)^T_{ij}.$$

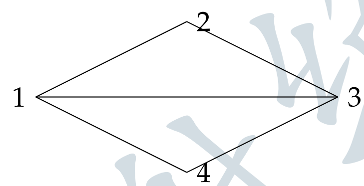
□

定义 3.41. 矩阵 A 称为对称矩阵, 如果 $A = A^T$.

$A = A^T$ 成立的必要条件是矩阵的行数和列数相等. 因此, 当我们谈论对称矩阵的时候, 天然地就是在考虑方阵. 下面我们举一些对称矩阵的例子.

1. 对角矩阵 (即除了对角线元素以外全为 0 的矩阵)
2. 若 A 是一个矩阵 (未必是方阵), 那么 AA^T 和 A^TA 都是对称矩阵.
3. 两个对称矩阵的和是对称矩阵; 两个对称矩阵的积未必是对称矩阵. 例如 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ 和 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 都是对称矩阵, 但是它们的积为 $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ 不是对称矩阵.
4. 设 $A, B \in M_n(\mathbb{R})$, 并且 B 是对称矩阵, 那么 ABA^T 是对称矩阵.

例题 3.42. 考虑无向图



我们可以用矩阵来表示这个无向图. 一个图的所有信息就是它的点和边. 点的表示就很容易, 那么我们想要表示边, 如何表示呢? 一个基本的方法就是" 指示", 类似于化学中的酸碱指示剂. 比如说, 如果点 i 和点 j 之间有一条边, 那么我们就赋值 1; 否则赋值 0. 因此我们得到了一个矩阵

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

S 是个对称矩阵, 因为我们考虑的是无向图, 点 i, j 之间有条边, 那么点 j, i 之间也有条边.

问题: $S^2, S^3, \cdots$ 分别表达了什么意思?

问题 3.43. 上一个例子中, 如果我们把无向图换成有向图, 那么问题该怎么叙述?

问题 3.44. 对于图论而言, 一个基本的问题是最短路径问题, 即给出起始点, 求连接起始点之间的最短的路径的长度. 还是考虑上一个例子, 点 1,4 之间有 3 条路径, 路径 $1-2-4, 1-3-4$ 的长度都是 2, 而路径 $1-4$ 的长度为 1; 因此点 1,4 之间的最短路径为 1; 点 2,3 之间的最短

路径为 2. 那么给定一个无向图图的两个点, 求它们之间的最短路径的一个显而易见的办法是穷举. 但是当点和边的数目都很多的时候, 穷举就变得不可能了. 那么同学们有没有什么思路来理解和解决最短路径问题?

3.4.1 对称矩阵的 LDU 分解

设 A 是一个可逆的对称矩阵, 并且 A 存在 LDU 分解, 即 $A = LDU$, $L(U)$ 是对角线为 1 的下 (上) 三角矩阵, D 是对角线非零的对角矩阵, 那么

命题 3.45. *存在 L 对角线为 1 的上三角矩阵, D 对角线全非零的对角矩阵, 使得 $A = LDL^T$;*

证明. 我们对 A 的行数做归纳. $n = 1$ 时, 显然. $n = 2$ 时, 由于 A 存在 LDU 分解, 因此 $a_1 1 \neq 0$, 又因为 A 是对称矩阵, 所以

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix},$$

$E(12, -\frac{a_{12}}{a_{11}})A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} - \frac{a_{12}^2}{a_{11}} \end{bmatrix}$,

$$E(12, -\frac{a_{12}}{a_{11}})AE(12, -\frac{a_{12}}{a_{11}})^T = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} - \frac{a_{12}^2}{a_{11}} \end{bmatrix} \begin{bmatrix} 1 & -\frac{a_{12}}{a_{11}} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} - \frac{a_{12}^2}{a_{11}} \end{bmatrix}$$

我们取 $L = E(12, -\frac{a_{12}}{a_{11}})^{-1}, D = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} - \frac{a_{12}^2}{a_{11}} \end{bmatrix}$.

对于一般的 n, 由于 A 存在 LDU 分解, 因此 $a_{11} \neq 0$, 又因为 A 是对称矩阵, 我们可以将 A 分块:

$$A = \begin{bmatrix} a_{11} & \boldsymbol{v}^T \\ \boldsymbol{v} & A' \end{bmatrix}$$

其中 $\boldsymbol{v}$ 是一个 $n-1$ 维的列向量, $A' \in M_{n-1}(\mathbb{R})$ 是一个对称矩阵. 取分块矩阵 $E = \begin{bmatrix} 1 & 0 \\ -\boldsymbol{v} & I_{n-1} \end{bmatrix}$, 我们知道

$$EAE^T = \begin{bmatrix} a_{11} & 0 \\ 0 & A_1 \end{bmatrix}$$

其中 $A_1 \in M_{n-1}(\mathbb{R})$ 是一个对称矩阵. 通过可逆矩阵的乘积是可逆矩阵, 我们知道 $\begin{bmatrix} a_{11} & 0 \\ 0 & A_1 \end{bmatrix}$ 是可逆矩阵, 从而 A_1 是可逆矩阵. 另一方面, $A = LDU$, 代入, 利用分块矩阵乘法, 即知 A_1 也

有 LDU 分解. 所以我们用归纳假设: $A_1 = L_1D_1L_1^T$, 所以 $A = E^{-1}\begin{bmatrix} a_{11} & 0 \\ 0 & L_1D_1L_2^T \end{bmatrix}(E^T)^{-1}$, 从而

$$A = \left(E^{-1}\begin{bmatrix} 1 & 0 \\ 0 & L_1 \end{bmatrix}\right)\begin{bmatrix} a_{11} & 0 \\ 0 & D_1 \end{bmatrix}\left(E^{-1}\begin{bmatrix} 1 & 0 \\ 0 & L_1 \end{bmatrix}\right)^T.$$

我们取 $L = \left(E^{-1}\begin{bmatrix} 1 & 0 \\ 0 & L_1 \end{bmatrix}\right), D = \begin{bmatrix} a_{11} & 0 \\ 0 & D_1 \end{bmatrix}$. 证毕. □

注释 3.46. 这个命题就显示了"对称矩阵"这个对称性如何减少了一半的运算量: 因为对于一般的矩阵, 我们在做 LDU 分解的时候, 既要算 L, 也要算 U. 但是对于对称矩阵而言, 我们只需要计算 L 即可.

3.4.2 初等列变换

矩阵转置通俗地来说, 就是"行变成列, 列变成行", 又或者说, 通过这种方式, 实现了行和列的对称. 既然如此, 在之前我们学习过矩阵的初等行变换, 那就应该有矩阵的"初等列变换":

- 将第 i, j 列互换位置;
- 将第 i 列的若干倍加到第 j 列;
- 将第 i 列乘以一个非零的倍数.

同样的, 初等列变换可以由矩阵来表示, 我们怎么看待这件事呢? 假设我们要将矩阵 A 的第 i, j 列互换, 那么我们可以把这一步分解成三步:

$$A \xrightarrow{\text{转置}} A^T \xrightarrow{\text{交换第}i,j\text{行}} P(ij)A^T \xrightarrow{\text{转置}} AP(ij)^T = AP(ij)\ .$$

注释 3.47. 注意此处初等行变换矩阵 $P(ij)$ 的行数和列数. 我们知道, 对一个 $m \times n$ 矩阵做初等行变换用的矩阵是 $m \times m$ 矩阵. 因此, 假设 $A \in M_{m\times n}(\mathbb{R})$, 那么对 A 做初等行变换, 用的矩阵是 m-阶方阵, 但是对 A^T 做初等行变换, 用的是 n-阶方阵. 所以有的时候, 为了避免混淆, 我们会在初等行变换矩阵加上阶数的下标, 以示区别, 例如 $P_n(ij)$.

因此三种初等列变换如下表示: 设 $A \in M_{m\times n}(\mathbb{R})$,

- 将第 i, j 列互换位置: $A \mapsto AP_n(ij)$;
- 将第 i 列的 e 倍加到第 j 列: $A \mapsto AE_n(ij, e)^T = AE_n(ji, e)$;
- 将第 i 列乘以 $e(\neq 0)$ 倍: $A \mapsto AE_n(i, e)^T = AE_n(i, e)$.

其他的关于列向量的讨论我们不再赘述, 其原则就是: 既然我们已经详细地讨论过行, 那么关于列的讨论就通过转置转化为关于行的讨论.

当然, 我们也可以直接讨论列, 但是讨论列的方法和讨论行的方法大同小异, 没有必要再做一遍.

注释 3.48. 从今以后, 我们把 “初等行变换矩阵” 改称 “初等矩阵”; 因为那种形式的矩阵到底是用来作 “初等行变换” 还是 “初等列变换”, 只取决于它是乘在原矩阵左边还是右边. 简而言之, “左乘是行变换, 右乘是列变换”. 唯一需要注意的是, 根据矩阵乘法的要求, 我们对于两个矩阵相乘左边的矩阵列数必须等于右边矩阵的行数.

第四章　线性空间

4.1　定义和例子

在第二章中, 我们介绍了用高斯消元法解线性方程, 用阶梯形矩阵和和增广矩阵的秩 (定理 2.17) 来判断方程组是否有解, 以及有解的时候, 什么时候有无穷多个解; 请同学们重新复习一下这个定理.

假设我们考虑线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$, 其中

$$A \in M_{m\times n}(\mathbb{R}), \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

假设它有解, 我们用高斯消元法判断过它的 “自由变量” 的个数, 在这一章中, 我们通过仔细地研究 “自由变量”, 来给出这个方程组的所有解的结构.

因为我们假设了这个方程组有解, 所以我们先固定这个方程组的一个解 $\boldsymbol{x}_0 \in \mathbb{R}^n$. 一方面, 若 $\boldsymbol{x}'$ 是这个方程组的一个解, 那么 $A\boldsymbol{x}' - A\boldsymbol{x}_0 = \boldsymbol{b} - \boldsymbol{b} = 0$, 所以 $\boldsymbol{x}' - \boldsymbol{x}_0$ 是方程组 $A\boldsymbol{x} = 0$ 的一个解. 另一方面, 若 $\boldsymbol{x}''$ 是方程组 $A\boldsymbol{x} = 0$ 的一个解, 那么 $A(\boldsymbol{x}'' + \boldsymbol{x}_0) = A\boldsymbol{x}'' + A\boldsymbol{x}_0 = 0 + \boldsymbol{b} = \boldsymbol{b}$, 也就是说 $\boldsymbol{x}'' + \boldsymbol{x}_0$ 是方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的一组解. 综上所述, 我们证明了下面这个命题:

命题 4.1. *设 $\boldsymbol{x}_0$ 满足 $A\boldsymbol{x}_0 = \boldsymbol{b}$, 那么我们有一一对应*

$$\left\{\boldsymbol{x} \in \mathbb{R}^n \middle| A\boldsymbol{x} = \boldsymbol{b}\right\} \longrightarrow \left\{\boldsymbol{x} \in \mathbb{R}^n \middle| A\boldsymbol{x} = 0\right\}, \quad \boldsymbol{x} \mapsto \boldsymbol{x} - \boldsymbol{x}_0.$$

所以, 研究 $A\boldsymbol{x} = 0$ 这样的方程, 就有着特殊的意义. 记 $N(A) = \left\{\boldsymbol{x} \in \mathbb{R}^n \middle| A\boldsymbol{x} = 0\right\}$. 我们注意到,

1. 若 $\boldsymbol{v} \in N(A)$, $a \in \mathbb{R}$, 那么 $a\boldsymbol{v} \in N(A)$. 这是因为 $A(a\boldsymbol{v}) = aA\boldsymbol{v} = a0 = 0$;

2. 若 $\boldsymbol{v}, \boldsymbol{w} \in N(A)$, 那么 $\boldsymbol{v} + \boldsymbol{w} \in N(A)$. 这是因为 $A\boldsymbol{v} + A\boldsymbol{w} = 0 + 0 = 0$.

这是什么意思呢? 这两条就说明, $N(A)$ 作为 $\mathbb{R}^n$ 的子集, $N(A)$ 中元素的线性组合也属于 $N(A)$.

我们回顾一下线性空间的定义. 设 V 是一个非空子集;

定义 4.2. 我们称 V 是一个 $\mathbb{R}$ 上的线性空间, 如果在 V 上有

- 加法: $V \times V \to V, \quad (\boldsymbol{v}, \boldsymbol{w}) \mapsto \boldsymbol{v} + \boldsymbol{w}$;
- 数乘: $\mathbb{R} \times V \to V, \quad (a, \boldsymbol{v}) \mapsto a\boldsymbol{v}$;

并且满足以下条件:

1. 存在加法单位元 $\mathbf{0}$, 即对任意 $\boldsymbol{v} \in V, \mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$;
2. 任意 $\boldsymbol{v} \in V$ 都有加法逆元, 即存在 $\boldsymbol{w} \in V$, 使得 $\boldsymbol{v} + \boldsymbol{w} = \mathbf{0}$; 我们把 $\boldsymbol{v}$ 的加法逆元记做 $-\boldsymbol{v}$;
3. 加法满足交换律: 对任意 $\boldsymbol{v}, \boldsymbol{w} \in V$, $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$;
4. 加法满足结合律: 对任意 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$;
5. 数乘零元和单位元: 对任意的 $\boldsymbol{v} \in V$, $0\boldsymbol{v} = \mathbf{0}; 1\boldsymbol{v} = \boldsymbol{v}$;
6. 数乘和加法的分配率 1: 对任意的 $a \in \mathbb{R}, \boldsymbol{v}, \boldsymbol{w} \in V$, $a(\boldsymbol{v} + \boldsymbol{w}) = a\boldsymbol{v} + a\boldsymbol{w}$;
7. 数乘和加法的分配率 2: 对任意的 $a, b \in \mathbb{R}, \boldsymbol{v} \in V$, $(a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$;
8. 数乘结合律: 对任意 $a, b \in \mathbb{R}, \boldsymbol{v} \in V$, $(ab)\boldsymbol{v} = a(b\boldsymbol{v})$

命题 4.3. 设 $A \in M_{m\times n}(\mathbb{R})$, 那么 $N(A) = \left\{\boldsymbol{x} \in \mathbb{R}^n | A\boldsymbol{x} = \mathbf{0}\right\}$ 是一个线性空间.

证明. 首先我们需要定义 $N(A)$ 上的加法和乘法. 因为 $N(A) \subseteq \mathbb{R}^n$, 因此一个自然的定义就是用 V 上的加法和数乘来定义 $N(A)$ 上的加法和数乘. 为了这个加法和数乘是良定义的, 我们需要验证

$$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in N(A), a \in \mathbb{R}, \text{我们都有 } \boldsymbol{x}_1 + \boldsymbol{x}_2 \in N(A), a\boldsymbol{x}_1 \in N(A).$$

我们已知 $N(A)$ 中元素的线性组合也在 $N(A)$ 中, 因此这个验证是显然的. 然后我们再分别验证线性空间定义的 $1-8$ 条要求; 这就说明了 $N(A)$ 是线性空间. 具体验证略, 请同学们自行练习. □

注释 4.4. 我们在上面的定义中, 把 $\mathbb{R}$ 都换成 $\mathbb{C}$, 其他的条件和叙述都不改变, 这样我们就定义了 $\mathbb{C}$ 上的线性空间.

注释 4.5. • 我们通常不区分向量 $\mathbf{0}$ 和实数 0; 混淆这两者并不会引起矛盾, 这是因为实数 0 只出现在数乘中; 结合上下文, 不会有混淆; 也因为第 5 条, 向量 $\mathbf{0}$ 和实数 0 联系很紧密.

- 任意向量的加法逆元都唯一: 设 $v \in V$, w_1, w_2 都是它的加法逆元, 那么

$$w_1 \overset{1}{=} w_1 + \mathbf{0} = w_1 + (v + w_2) \overset{4}{=} (w_1 + v) + w_2 = \mathbf{0} + w_2 \overset{1}{=} w_2.$$

- 对任意 $v \in V$, $-1v = -v$; 这是因为 $v + (-1)v \overset{7}{=} 0v \overset{5}{=} \mathbf{0}$.

- 第 3 条加法交换律事实上是其它几条的推论, 也就是说, 作为一个定义, 它并不需要出现. 证明如下: 首先由分配率 1, $2(v + w) = 2v + 2w$; 其次, 由分配率 2, $2(v + w) = (1 + 1)(v + w) = (v + w) + (v + w)$; 所以 $2v + 2w = (v + w) + (v + w)$, 从而

$$\begin{aligned} v + w &= (-1 + 2)v + (2 - 1)w \\ &\overset{7}{=} (-v) + 2v + 2w + (-w) \\ &= (-v) + (v + w) + (v + w) + (-w) \\ &\overset{4}{=} w + v \end{aligned}$$

  那么, 这一条为什么要写在定义里呢? 这是因为它终究是一条运算规律, 有它自己独立的数学意义. (很难理解对吧, 说了比没说还难理解; 那就不要理解. 但是要理解: 既然它不独立我们也要摆在这里, 说明它很重要, 我们一定要记住它!)

例题 4.6. 1. $\mathscr{H} := \{(x, y) \in \mathbb{R}^2 : y \geq 0\}$ 不是线性空间, 因为它上面的数乘映射不是良定义的.

2. 记 $C(\mathbb{R})$ 是由 $\mathbb{R} \to \mathbb{R}$ 的所有连续函数构成的集合. 它是一个线性空间. $C^+(\mathbb{R}) := \{f \in C(\mathbb{R}) : f\text{是单调增函数}\}$ 不是线性空间, 这也是因为数乘映射不是良定义的: 当 $a < 0, f \in C^+(\mathbb{R})$ 时, af 是个单调递减函数. 记 $C^1(\mathbb{R})$ 是由 $\mathbb{R} \to \mathbb{R}$ 的所有有连续导函数的函数构成的集合 ($f \in C^1(\mathbb{R})$ 当且仅当它处处可导, 且 $f' \in C(\mathbb{R})$), 它也是一个线性空间.

3. 设 S 是一个有限集合, 那么 $V := \{f : S \to \mathbb{R}\}$ 为所有的 S 到 $\mathbb{R}$ 的线性映射构成的集合. 它是一个线性空间. 若 S 的元素个数为 n, 那么 $V \simeq \mathbb{R}^n$.

4. $M_{m\times n}(\mathbb{R})$ 是一个线性空间, 特别的, $M_n(\mathbb{R})$ 是一个线性空间. $M_n(\mathbb{R})$ 中由所有的上 (下) 三角矩阵构成的集合是线性空间, 由所有的对角矩阵构成的集合也是线性空间.

5. $M_n(\mathbb{R})$ 中由可逆矩阵构成的集合不是线性空间.

定义 4.7. V 是一个线性空间, 我们称 W 是 V 的子空间, 如果 W 是 V 的子集, 并且 W 按照 V 中定义的加法和数乘构成了一个线性空间.

注释 4.8. 关于子空间这个概念, W 是 V 的子集是很自然的. 但是后一个条件, 我们要求 W 按照 V 中定义的加法构成一个线性空间, 这是很重要的. 这里的也可以区分出 $\mathbb{R}$ 和 $\mathbb{C}$ 上的线性空间的区别. 在 $V=\mathbb{C}=\{x+iy: x,y\in\mathbb{R}\}$, 加法和数乘定义为

$$(x_1+iy_1)+(x_2+iy_2)=(x_1+x_2)+i(y_1+y_2);\quad a(x_1+iy_1)=ax_1+i(ay_1);$$

这是一个 $\mathbb{R}$ 线性空间, 而且事实上, 是唯一的能使 V 作为 $\mathbb{R}$-线性空间的加法和数乘的定义. 但是, 作为 $\mathbb{C}$ 线性空间, 我们有两种方式:

$$\begin{aligned}&(x_1+iy_1)+(x_2+iy_2)=(x_1+x_2)+i(y_1+y_2);\quad a\circ(x_1+iy_1)=a(x_1+iy_1);\quad x_1,x_2,y_1,y_2\in\mathbb{R},a\in\mathbb{C}\\&(x_1+iy_1)+(x_2+iy_2)=(x_1+x_2)+i(y_1+y_2);\quad a\otimes(x_1+iy_1)=\bar{a}(x_1+iy_1);\quad x_1,x_2,y_1,y_2\in\mathbb{R},a\in\mathbb{C}\end{aligned}$$

其中 $\bar{a}$ 是 a 的复共轭. 可以验证在这两种定义之下, V 都是 $\mathbb{C}$ 线性空间. 如果我们把按照第二种加法和数乘定义的线性空间记作 W, 那么在集合上, $W\subseteq V$, 但是 W 不是 V 的子空间.

注释 4.9. 这个例子也告诉我们, 同一个集合上, 可以定义 "不同" 的线性空间的结构. 那么我们可不可以分类所有的线性空间呢? 请同学们思考一下, 什么叫分类.

往往我们说 V 的一个子集 W 是它的子空间, 或者直接说 W 是 V 的子空间, 默认是指 W 按照 V 的加法和数乘构成了一个线性空间. 当然, 我们整个课程如果没有特殊说明, 都遵循 "往往" 这个约定.

命题 4.10. 若 W_1,W_2 都是 V 的子空间, 那么 $W_1\cap W_2$ 也是 V 的子空间.

这个直接验证定义, 证明省略.

注释 4.11. 若 W_1,W_2 都是 V 的子空间, 一般来说 $W_1\cup W_2$ 不是 V 的子空间, 除非它们之间存在包含关系. 反例很容易找, 例如平面上过原点的两条不同的直线的并集, 对于加法是不封闭的.

设 V 是线性空间, S 是 V 的子集, 定义 $\mathrm{Span}(S)\subseteq V$ 是由 S 的所有有限线性组合构成的集合, 也就是说, $\mathrm{Span}(S)$ 是由所有的形如 $a_1s_1+\cdots+a_ns_n(a_i\in\mathbb{R},s_i\in S,1\le i\le n,n\in\mathbb{N})$ 的元素组成的集合. 或者, 我们可以这么记:

$$\begin{aligned}\mathrm{Span}(S)\quad&:=\left\{a_1s_1+\cdots+a_ns_n\ \middle|a_i\in\mathbb{R},s_i\in S,1\le i\le n,\forall n\in\mathbb{N}\right\}\\&=\left\{\sum_{s\in S}a_s\cdot s\middle|\text{只有有限多个系数 } a_s \text{ 非零}\right\}\end{aligned}.$$

若 S 只含有有限多个元素, 不妨设 $S=\{v_1,v_2,\cdots,v_k\}$, 那么

$$\mathrm{Span}(S)=\{a_1v_1+a_2v_2+\cdots+a_kv_k\Big|a_i\in\mathbb{R},1\le i\le k.\}$$

例题 4.12. $V=\mathbb{R}^2$, 设 $S=\left\{\begin{bmatrix}0\\2^n\end{bmatrix}:n\in\mathbb{N}\right\}$. 那么根据定义, 对 $a_1,\cdots,a_n\in\mathbb{R}$, $\sum_{i=1}^{n}a_i\cdot\begin{bmatrix}0\\2^n\end{bmatrix}$ 属于 $\mathrm{Span}(S)$. 但是, 如果令 $a_i=4^{-i}$, 那么 $\sum_{i=1}^{\infty}a_i\cdot\begin{bmatrix}0\\2^n\end{bmatrix}$, 这是因为这个线性组合有无穷多个系数非零, 所以不是“有限”线性组合. 我们也可以验证:

$$\mathrm{Span}(S)=\left\{\begin{bmatrix}0\\a\end{bmatrix}:a\in\mathbb{R}\right\}$$

命题 4.13. $\mathrm{Span}(S)$ 是 V 的子空间.

直接验证定义.

注释 4.14. 如果 $S'\subseteq S$, 那么 $\mathrm{Span}(S')\subseteq\mathrm{Span}(S)$. 如果 $S'\subseteq\mathrm{Span}(S)$, 那么 $\mathrm{Span}(S')\subseteq\mathrm{Span}(S)$. 前者是显然的: S' 中元素的线性组合自然是 S 中元素的线性组合. 后者是因为: “线性组合的线性组合还是线性组合”, 就是说, 因为 S' 中的元素是 S 中元素的线性组合, 那么 S' 中元素的线性组合就是对“S 中元素的线性组合”再做线性组合, 因此还是 S 中元素的线性组合.

推论 4.15. 若 $W\subseteq V$ 并且 W 中所有元素的任意线性组合都属于 W, 那么 W 是 V 的子空间.

证明. 取 $S=W$, 由条件我们知 $W=\mathrm{Span}(S)$, 再由上面的命题知 W 是 V 的子空间. □

定义 4.16. 设 V 是线性空间, S 是 V 的子集. 如果 $V=\mathrm{Span}(S)$, 我们称 S 是 V 的一组生成元, 称 V 是由 S 生成的 (或者张成的) 线性空间.

设 V 是线性空间, 若存在 V 的有限子集 S 使得 $V=\mathrm{Span}(S)$, 我们称 V 是有限维线性空间.

注释 4.17. 请注意, 我们虽然定义了什么是有限维线性空间, 但是我们并没有定义线性空间的“维数”这个概念. “维数”这个概念是线性空间理论的一个核心概念. 这一章我们都要围绕它展开.

首先, 我们知道, 对任意的线性空间 V, 它都有一组“平凡”生成元: $S=V$; 这组生成元是极大的, 也是最大的: 任意一组其他的生成元 S', 都有 $S'\subseteq S$. 那么, V 有没有极小生成元组呢? 这里所谓的极小生成元组 S, 是说 S 的任意真子集都不是生成元组.

注释 4.18. 我们考虑一个集合的一些子集构成的集合: 设 S 是一个集合, $\mathscr{P}=\{S':S'\subseteq S\}$, $H\subseteq\mathscr{P}$. 对 $X_1,X_2\in H$, 我们称 X_1 比 X_2 小 (或者 X_2 比 X_1 大), 如果 $X_1\subset X_2$. 我们称 X 是 H 中的极小 (大) 元, 如果不存在 X' 比 X 小 (大). 我们称 X 是 H 中的最小 (大) 元, 如果任意 $X'\in H$ 比 X 大 (小). 极小 (大) 元和最小 (大) 元未必存在; 极小元往往也不是最小元.

我们现在考虑两个线性空间 V, W;

定义 4.19. 设 $f: V \to W$ 是一个线性映射, 如果它满足:

$$f(av_1 + bv_2) = af(v_1) + bf(v_2); \forall a, b \in \mathbb{R}, v_1, v_2 \in V.$$

这个要求我们俗称是 “保持线性结构”, 或者 “和线性结构相容”. 有点诘诎聱牙, 说人话, 就是 “把线性组合映为线性组合”.

例题 4.20. 我们考虑映射

$$\frac{\mathrm{d}}{\mathrm{d}x} : C^1(\mathbb{R}) \to C(\mathbb{R}), \quad f \mapsto \frac{\mathrm{d}f}{\mathrm{d}x}.$$

显然, 由求导的基本法则, 这是一个线性映射.

例题 4.21. 设 $A \in M_{m\times n}(\mathbb{R})$, 定义

$$\varphi_A : \mathbb{R}^n \to \mathbb{R}^m, \quad v \mapsto Av.$$

这也是一个线性映射.

事实上, 任意一个线性映射 $\varphi : \mathbb{R}^n \to \mathbb{R}^m$, 都存在唯一的的矩阵 $A \in M_{m\times n}(\mathbb{R})$, 使得 $\varphi = \varphi_A$. 证明如下: 我们取 e_i 是 $\mathbb{R}^n$ 的第 i 个单位向量: 它的第 i 个分量为 1, 其余分量为 0. 令 $\alpha_i = \varphi(e_i) \in \mathbb{R}^m, A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{bmatrix} \in M_{m\times n}(\mathbb{R})$, 那么就有 $\varphi = \varphi_A$. 这是因为: 对任意的 $v = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1e_1 + x_2e_2 + \cdots + x_ne_n$, 我们有

$$\begin{aligned} \varphi(v) &= \varphi(x_1e_1 + x_2e_2 + \cdots + x_ne_n) \\ &= x_1\varphi(e_1) + x_2\varphi(e_2) + \cdots + x_n\varphi(e_n) \\ &= x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n \\ &= A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = Av = \varphi_A(v) \end{aligned}.$$

例题 4.22. 设 $A \in M_{m\times n}(\mathbb{R})$, 定义

$$\begin{aligned} L_A &: M_{n\times k}(\mathbb{R}) \to M_{m\times k}(\mathbb{R}) \quad X \mapsto AX; \\ R_A &: M_{k\times m}(\mathbb{R}) \to M_{k\times n}(\mathbb{R}) \quad X \mapsto XA. \end{aligned}$$

它们都是线性映射. 这是因为矩阵乘法满足分配率.

例题 4.23. 设 $v_0 \in \mathbb{R}^n$, 那么映射

$$\ell : \mathbb{R}^n \to \mathbb{R}, \quad v \mapsto v \cdot v_0$$

是一个线性映射. 如果 $v_0 = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, 那么 ℓ 由 $1 \times n$ 矩阵 $A = [a_1, a_2, \cdots, a_n]$ 给出; 也就是说,

$$\ell = L_A : M_{n\times 1}(\mathbb{R}) \to M_1(\mathbb{R}).$$

对于一个映射而言, 我们要研究它的定义域, 值域, 以及每个取值的原像是什么. 对一个线性映射 $f : V \to W$ 而言, 它的定义域没什么好说的, 就是 V. 它的值域, 也就是像集, 我们定义为

$$\mathrm{Im}(f) := \Big\{ w \in W : \exists v \in V, s.t. w = f(v) \Big\} \subseteq W.$$

对任意 $w \in W$, 我们定义它的原像

$$f^{-1}(w) := \Big\{ v \in V : f(v) = w \Big\} \subseteq V;$$

特别的, 我们把 $f^{-1}(\mathbf{0})$ 称为是 f 的核 (或者零空间), 记为:

$$\ker(f) := \Big\{ v \in V : f(v) = \mathbf{0} \Big\}.$$

命题 4.24. $\mathrm{Im}(f)$ 是 W 的线性子空间; $\ker(f)$ 是 V 的线性子空间.

证明. 对 $\mathrm{Im}(f)$, 我们需要证明对任意的 $w_1, w_2 \in \mathrm{Im}(f), a, b \in \mathbb{R}$, $aw_1 + bw_2 \in \mathrm{Im}(f)$. 因为 $w_1, w_2 \in W$, 那么存在 $v_1, v_2 \in V$ 使得 $f(v_1) = w_1, f(v_2) = w_2$. 所以

$$aw_1 + bw_2 = af(v_1) + bf(v_2) = f(av_1 + bv_2) \in \mathrm{Im}(f).$$

对 $\ker(f)$, 我们需要证明对任意的 $v_1, v_2 \in \ker(f), a, b \in \mathbb{R}$, $av_1 + bv_2 \in \ker(f)$, i.e. $f(av_1 + bv_2) = 0$. 只需要注意到, 因为 $v_1, v_2 \in \ker(f)$, 我们有 $f(v_1) = f(v_2) = \mathbf{0}$, 所以

$$f(av_1 + bv_2) = af(v_1) + bf(v_2) = \mathbf{0}.$$

□

例题 4.25. 在例题 4.20中, $\ker\left(\dfrac{\mathrm{d}}{\mathrm{d}x}\right)$ 是那些函数 f 使得 $\dfrac{\mathrm{d}f}{\mathrm{d}x} = 0$; 所以 f 是常值函数, 故 $\ker\left(\dfrac{\mathrm{d}}{\mathrm{d}x}\right) = \mathbb{R}$. $\mathrm{Im}\left(\dfrac{\mathrm{d}}{\mathrm{d}x}\right) = C(\mathbb{R})$, 这是因为 $\int_0^x f(t)\mathrm{d}t$ 是 $f(x)$ 的一个原函数: $f \in C(\mathbb{R})$ 保证了积分是良定义的.

例题 4.26. 在例题 4.21中, 由于线性映射 φ_A 是由矩阵 $A \in M_{m\times n}(\mathbb{R})$ 给出的, 这个时候, 我们把 $\ker(\varphi_A)$ 也记作 $N(A)$. 那么由定义

$$N(A) = \left\{ \boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \boldsymbol{0} \right\};$$

所以 $N(A)$ 就是方程组 $A\boldsymbol{x} = \boldsymbol{0}$ 的所有解. φ_A 的像集 $\mathrm{Im}(f)$ 是什么呢?

4.2　矩阵的行空间和列空间

暂且, 我们不区分行向量和列向量, 例如, 我们用 $\mathbb{R}^n$ 同时表示由所有的 n 维行向量组成的线性空间, 和由所有的 n 维列向量组成的线性空间.

设 $A \in M_{m\times n}(\mathbb{R})$ 是一个矩阵, 我们可以把它看成是 m 个行向量, 也可以看成是 n 个列向量. 即:

$$A = \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_n \end{bmatrix}$$

我们定义:

- $R(A) = \mathrm{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_m) \subseteq \mathbb{R}^n$ 为 A 的行向量空间
- $C(A) = \mathrm{Span}(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n) \subseteq \mathbb{R}^m$ 为 A 的列向量空间.

那么显然的:

$$\text{方程组 } A\boldsymbol{x} = \boldsymbol{b} \text{ 有解当且仅当 } \boldsymbol{b} \in C(A).$$

注释 4.27. 依然是显然地, 在不区分行向量和列向量的情况下, 我们有下面的等式,

$$R(A) = C(A^T), \quad C(A) = R(A^T).$$

命题 4.28. E 是初等行变换. 那么 $R(EA) = R(A)$.

证明. 若 $E = P(ij)$ 行, 那么

$$R(EA) = \mathrm{Span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_j, \cdots, \boldsymbol{v}_i, \cdots, \boldsymbol{v}_m) = \mathrm{Span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_i, \cdots, \boldsymbol{v}_j, \cdots, \boldsymbol{v}_m) = R(A).$$

若 $E = E(ij, e)$, 那么

$$\begin{aligned} R(EA) &= \mathrm{Span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_i, \cdots, \boldsymbol{v}_j + e\boldsymbol{v}_i, \cdots, \boldsymbol{v}_m), \\ R(A) &= \mathrm{Span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_i, \cdots, \boldsymbol{v}_j, \cdots, \boldsymbol{v}_m) \end{aligned}$$

由 Span 的定义, 因为 EA 的行向量都是 A 的行向量的线性组合, 所以显然 $R(EA) \subseteq R(A)$. 另一方面 $\boldsymbol{v}_j = (\boldsymbol{v}_j + e\boldsymbol{v}_i) - e\boldsymbol{v}_i$, 也就是说 A 的行向量也是 EA 的行向量的线性组合, 所以 $R(A) \subseteq R(EA)$.

若 $E = E(i,e)(e \neq 0)$, 那么

$$R(EA) = \operatorname{Span}(\boldsymbol{v}_1, \cdots, e\boldsymbol{v}_i, \cdots, \cdots, \boldsymbol{v}_m) = \operatorname{Span}(\boldsymbol{v}_1, \cdots, e\boldsymbol{v}_i, \cdots, \cdots, \boldsymbol{v}_m) = R(A),$$

□

注释 4.29. 这个命题说明了初等行变换不改变行空间. 但是显然, 一般地, 它要改变列空间. 利用转置, 我们可以知道, 若 E' 是初等列变换, 那么 $C(AE') = C(A)$.

推论 4.30. 设 X 是一个可逆的 m-阶方阵, $A \in M_{m\times n}(\mathbb{R})$. 那么 $R(XA) = R(A)$.

证明. 由 Gauss-Jordan 消去, 我们知道 X 可以写成初等行变换的乘积, 所以 $R(XA) = R(A)$. □

由 Gauss 消去法, 我们可以把 A 转化成阶梯形矩阵, 我们记作 $\operatorname{ref}(A)$. 由于 $R(\operatorname{ref}(A)) = R(A)$, 所以 $R(A)$ 由 $\operatorname{ref}(A)$ 中的非零行生成. 如果我们记

$$d := \min\{\#S : S\text{是 } R(A) \text{ 的生成元组}\},$$

那么就有 d 小于等于 $\operatorname{ref}(A)$ 中分非零行的数目, 也就是 A 的秩 $\operatorname{rank}(A) =: r$.

定理 4.31.

$$d = r.$$

证明. 由已知 $d \leq r$, 要证明这个命题, 我们只需要证明 $d < r$ 是不可能发生的. 证明的基本思路使用反证法, 就是先假设 $d < r$, 然后推导出矛盾来.

根据 d 的定义, 存在集合 $S \subseteq R(A) \subseteq \mathbb{R}^n$, 使得 S 中元素个数为 d 且 $R(A) = \operatorname{Span}(S)$. 设 $S = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_d\} \subseteq \mathbb{R}^n$. 再设

$$\operatorname{ref}(A) = \begin{bmatrix} 0 & \cdots & t_{1j_1} & \cdots & t_{1j_2} & \cdots & t_{1j_k} & \cdots & t_{1n} \\ 0 & \cdots & \cdots & 0 & t_{2j_2} & \cdots & t_{1j_k} & \cdots & t_{2n} \\ & & & & \ddots & & \vdots & & \vdots \\ & & & & & \ddots & t_{rj_r} & \cdots & t_{rn} \\ & & & & & & & & 0 \\ & & & & & & & & 0 \end{bmatrix},$$

其中 $1 \le j_1 < j_2 < \cdots < j_r \le \min\{m,n\}, t_{1j_1}t_{2j_2}\cdots t_{rj_r} \ne 0$, 对 $1 \le i \le r$. 我们再记 $\boldsymbol{v}_i$ 为 $\mathrm{ref}(A)$ 的第 i 个非零行, 由于 $\mathrm{Span}(S) = R(A) = R(\mathrm{ref}(A))$, 所以对任意的 $1 \le i \le r$, $\boldsymbol{v}_i$ 都是 $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_d\}$ 的线性组合, 所以我们假设

$$\boldsymbol{v}_i = b_{i1}\boldsymbol{w}_1 + b_{i2}\boldsymbol{w}_2 + \cdots + b_{id}\boldsymbol{w}_d,$$

令 $B \in M_{r\times d}(\mathbb{R})$, 使得 $B_{ij} = b_{ij}$, 那么我们有 $r \times n$ 的矩阵的等式:

$$\begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_r \end{bmatrix} = B \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_d \end{bmatrix}.$$

考虑 $B^T \in M_{d\times r}(\mathbb{R})$, 根据我们的反证假设, 它的行数 d 小于 r, 所以方程组

$$B^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

有无穷多个解. 所以我们能取到一个非零解 $[x_1, x_2, \cdots, x_r]^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix}$, 那么就有 $[x_1, x_2, \cdots, x_r]B = [0, 0, \cdots, 0]$, 从而

$$[x_1, x_2, \cdots, x_r] \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_r \end{bmatrix} = [x_1, x_2, \cdots, x_r]B \begin{bmatrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_d \end{bmatrix} = 0.$$

也即 $x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_r\boldsymbol{v}_r = 0$. 考虑这个等式左边的向量的各个分量, 我们得到方程组:

$$\begin{aligned} &t_{1j_1}x_1 &&= 0 \\ &t_{1j_2}x_1 + t_{2j_2}x_2 &&= 0 \\ &\vdots \\ &t_{1j_r}x_1 + t_{2j_r}x_2 + \cdots + t_{rj_r}x_r &&= 0 \end{aligned}$$

解得 $x_1 = x_2 = \cdots = x_r = 0$. 这与我们选取的 $[x_1, x_2, \cdots, x_r]$ 是非零向量矛盾.

总而言之, 我们的反证假设不成立. 所以原命题成立. □

回忆一下, 我们定义过矩阵 A 的秩为对 A 做初等行变换, 将 A 转化成的行简化阶梯型矩阵 $\mathrm{ref}(A)$ 的非零行的个数. 但是 $\mathrm{ref}(A)$ 的形式并不唯一 (取决于我们做 Gauss 消元法的步骤), 因此我们并不知道这个非零行的个数是不是唯一确定的, 也就是说, 我们并不知道当时定义的矩阵 A 的秩是不是良定义的. 但是上面的命题告诉我们, 矩阵的秩等于它的行空间的生成元组的元素个数的最小值, 这个数与对矩阵做的初等行变换无关.

推论 4.32. $\mathrm{rank}(A)$ 跟 $\mathrm{ref}(A)$ 没有关系, 是个良定义的数, 这是因为 d 和 $\mathrm{ref}(A)$ 没有关系, 被 A 唯一决定.

注释 4.33. 注意, d 是 $R(A)$ 的所有生成元组的元素个数的最小值. 那么假设 S 是 $\mathrm{Span}(A)$ 的一个生成元组, 并且 $\#S = d$, 那么 S 本身一定是一个极小生成元组, 但它不是最小生成元组. 事实上, $R(A)$ 的生成元组构成的集合并没有最小元: 我们反证, 设 $S = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_d\}$ 是个最小生成元组, 那么由最小性, 任意其他的生成元组都要包含 S; 但是令 $S' = \{-\boldsymbol{v}_1, \cdots, \boldsymbol{v}_d\}$ 也是一个生成元组, 但是 $S \not\subseteq S'$. 矛盾.

那么问题来了, 是否存在 $R(A)$ 的极小生成元组, 使得它的元素个数不等于 d 呢?

注释 4.34. 我们考虑了 $R(A)$ 的生成元组的元素个数的最小值, 那么 $C(A)$ 的生成元组的元素个数的最小值是多少呢? 是否等于 r 呢?

4.3 线性相关和线性无关

定理 4.35. $R(A)$ 的任何一个极小生成元组都恰好包含 r 个元素.

设 S 是 $R(A)$ 的一个极小生成元组. 首先我们说明 S 只包含有限多个元素. 设 $\boldsymbol{v}_i$ 是 $\mathrm{ref}(A)$ 的第 i 行, 那么存在 $\alpha_{i1}, \cdots, \alpha_{ij_i} \in S$, 使得 $\boldsymbol{v}_i$ 是它们的线性组合, 因此

$$S' := \{\alpha_{11}, \cdots, \alpha_{1j_1}, \cdots, \alpha_{r1}, \cdots, \alpha_{rj_r}\} \subseteq S$$

能生成所有的 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r$, 所以 $\mathrm{Span}(S') = R(A)$. 由极小性, $S' = S$, 所以

$$\#S = \#S' \leq j_1 + j_2 + \cdots + j_r.$$

所以 $\#S$ 是个有限数, 我们设为 N. 我们要证明的定理就是说 $N = r$. 我们记 $S = \{\alpha_1, \cdots, \alpha_N\}$

什么叫极小生成元组呢? 一方面它是生成元组, 一方面是极小的生成元组. 极小的意思就是, 任意去掉一个向量 α_i 都不是一个生成元组. 既然不是生成元组, 就存在一些向量, 不是 $\{\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_N\}$ 的线性组合. 那么, 我们能不能找到一个向量不是它们的线性组合呢?

注释 4.36. 这里的话有点拗口, 我们用一些术语把它们包装一下. 如果向量 β 是向量 $\beta_1, \cdots, \beta_k$ 的线性组合, 我们称向量 β 能被向量 $\beta_1, \cdots, \beta_k$ 线性表出; 如果向量 β 不是向量 $\beta_1, \cdots, \beta_k$ 的线性组合, 我们称向量 β 不能被向量 $\beta_1, \cdots, \beta_k$ 线性表出.

一个自然的直觉就是 α_i 自己就不是剩下那 $N-1$ 个向量的线性组合. 事实上, 若不然, $\{\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_N\}$ 也是 $R(A)$ 的生成元集. 和 S 的极小性矛盾. 这个命题的另一个等价描述是:

命题 4.37. 若 $t_1\alpha_1 + t_2\alpha_2 + \cdots + t_N\alpha_N = 0$, 那么 $t_1 = t_2 = \cdots = t_N = 0$.

证明. 若不然, 存在不全为零的数 $t_1, \cdots, t_N$ 使得 $t_1\alpha_1 + t_2\alpha_2 + \cdots + t_N\alpha_N = 0$, 设 $t_i \neq 0$, 于是

$$\alpha_i = -\frac{t_1}{t_i}\alpha_1 - \cdots - \frac{t_{i-1}}{t_i}\alpha_{i-1} - \frac{t_{i+1}}{t_i}\alpha_{i+1} - \cdots - \frac{t_N}{t_i}\alpha_N.$$

也就是说, α_i 可由余下的 $N-1$ 个向量线性表出. 矛盾. □

这个现象提示我们总结下面的概念:

定义 4.38. 设 V 是一个线性空间, $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$ 是 V 中的向量. 我们称它们线性相关, 如果存在不全为零的是实数 $a_1, \cdots, a_k$ 使得

$$a_1\boldsymbol{v}_1 + \cdots + a_k\boldsymbol{v}_k = 0.$$

若不然, 称它们线性无关. 或者等价地, 如果实数 $a_1, \cdots, a_k$ 使得 $a_1\boldsymbol{v}_1 + \cdots + a_k\boldsymbol{v}_k = 0$, 那么 $a_1 = \cdots = a_k = 0$.

推论 4.39. $R(A)$ 的极小生成元组是线性无关的.

推论 4.40. $\mathrm{ref}(A)$ 的 r 个非零行向量是线性无关的.

证明. 我们可以设

$$\mathrm{ref}(A) = \begin{bmatrix} 0 & \cdots & t_{1j_1} & \cdots & t_{1j_2} & \cdots & t_{1j_r} & \cdots & t_{1n} \\ 0 & \cdots & \cdots & 0 & t_{2j_2} & \cdots & t_{1j_r} & \cdots & t_{2n} \\ & & & & \ddots & & \vdots & & \vdots \\ & & & & & \ddots & t_{rj_r} & \cdots & t_{rn} \\ & & & & & & & & 0 \\ & & & & & & & & 0 \end{bmatrix}.$$

$\boldsymbol{v}_i$ 是 $\mathrm{ref}(A)$ 的第 i 行, 那么我们要证明的就是 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r$ 线性无关.

设 $x_1\boldsymbol{v}_1+\cdots+x_r\boldsymbol{v}_r=0$, 如果我们写成列向量的形式, 就有 $x_1\boldsymbol{v}_1^T+\cdots+x_r\boldsymbol{v}_r^T=0$ 那么 $x_1,\cdots,x_r$ 满足方程组: $\mathrm{ref}(A)^T\begin{bmatrix}x_1\\x_2\\\vdots\\x_r\end{bmatrix}=0$. 忽略 $\mathrm{ref}(A)$ 的前 j_1-1 个全为零的列, 我们有

$$\begin{bmatrix}t_{1j_1} & 0 & \cdots & 0\\ \vdots & 0 & \cdots & 0\\ t_{1j_2} & t_{2j_2} & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ t_{1j_r} & t_{2j_r} & & t_{rj_r}\\ \vdots & \vdots & \vdots & \vdots\\ t_1n & t_{2n} & & t_{rn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_r\end{bmatrix}=0.$$

因为 $t_{1j_1},t_{2j_2},\cdots,t_{rj_r}$ 都不等于 0, 显然地, $x_1=x_2=\cdots=x_r=0$ 是它的唯一解. □

引理 4.41. 对线性空间 V 的任意一个生成元组 $T=\{\beta_1,\cdots,\beta_k\}$, 如果 $\beta_1,\cdots,\beta_k$ 是线性无关的, 那么它是极小生成元组.

证明. 如果 T 不是极小生成元组, 那么存在 β_j, 使得 $T-\{\beta_j\}$ 也是个生成元组, 因此 β_j 可由 $\beta_1,\cdots,\beta_{j-1},\beta_{j+1},\cdots,\beta_k$ 线性表出, 这与 $\beta_1,\cdots,\beta_k$ 的线性无关性矛盾. □

总而言之, 我们把推论 (4.39) 和引理 (4.41) 结合起来, 就有

命题 4.42. 一个生成元组 T, 它是极小生成元组, 当且仅当它包含的向量组是线性无关的.

至此, 我们重新分析一下我们想要证明的定理 (4.35). 这个定理的等价描述, 就是 $R(A)$ 的所有极小生成元组的元素个数都是 r, 即它们都相等. 由推论 (4.40), $\mathrm{ref}(A)$ 的非零行向量是一个极小生成元组, 有 r 个元素, 因此, 定理 (4.35) 是下面这个关于一般线性空间的定理的推论.

定理 4.43. 设 V 是一个线性空间, S 是 V 的一个极小生成元组, 且 $\#S=n$. 那么任意的极小生成元组 T 都满足 $\#T=n$. 或者说, V 的所有极小生成元组的元素个数都相等.

证明. 利用这一节开头的讨论, 我们可以证明 T 也是个有限集, 所以我们设 $\#T=m$. 我们用反证法, 假设 $n\neq m$. 不妨设 $n>m$. 设 $S=\{\alpha_1,\cdots,\alpha_n\},T=\{\beta_1,\cdots,\beta_m\}$. 由于 T 是生成元集, 所以 $\alpha_i\in\mathrm{Span}(T)$, 所以存在 $a_{ij}\in\mathbb{R}$, 使得 $\alpha_i=a_{i1}\beta_1+\cdots+a_{im}\beta_m$, 我们用矩阵的形式写

出来那就是

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = A \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}, \quad A = (a_{ij}) \in M_{n\times m}(\mathbb{R}).$$

考虑矩阵 $A^T \in M_{m\times n}(\mathbb{R})$ 和方程组 $A\boldsymbol{x} = 0$. 由于 A 的行数小于未知数的个数, 那么该方程组有无穷多个解; 我们取一个非零解 $\boldsymbol{x}_0 = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$, 那么 $[t_1, t_2, \cdots, t_n]A = \boldsymbol{x}_0^T A = (A^T \boldsymbol{x}_0)^T = 0$; 所以

$$[t_1, t_2, \cdots, t_n] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \boldsymbol{x}_0^T A \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} = 0,$$

也就是说, 我们找到了不全为零的实数 $t_1, t_2, \cdots, t_n$ 使得 $t_1\alpha_1 + \cdots + t_n\alpha_n = 0$. 这个和注释 (??) 矛盾. 证毕. □

由这个定理, 我们可以作如下的定义:

定义 4.44. 线性空间 V 的维数定义为它的极小生成元组的元素个数, 记作 $\dim V$.

在先前对于 “极小生成元组” 的讨论中, 我们的侧重点在 “极小”, 以 “生成性” 为辅. 但是我们也可以尝试着以 “生成性” 为侧重点, 以 “极小性” 为辅. 事实上, 如果一个集合能线性组合出整个空间, 那么它的元素个数就不能太少, 例如, 我们无法指望一个 $\mathbb{R}^2$ 中的一个元素就能生成整个 $\mathbb{R}^2$; 另一方面, 我们用 “极小性” 取控制元素个数的上界. 因此 “极小生成元组” 是一个矛盾的两方面的结合体. 那么极小性体现在哪一个方面呢?

我们回顾一下推论 (4.41), 它说的是, 一个生成元组, 如果是极小生成元组, 那么它是线性无关的. 所以 “线性无关” 体现的是 “极小性”. 我们也必须有这样的感觉: 元素个数越多, 就越不容易线性无关. 那么我们要问和推论 (4.41) 类似的问题:

一个线性无关组什么时候是极小生成元组?

那个问题的答案也是明显的: 我们已经了解到, 极小生成元组同时体现了多和少两个方面, 而线性无关体现的少的方面, 那么关于上面那个问题的回答, 就是需要加一个 “多” 的条件.

定理 4.45. 设 V 是一个线性线性子空间, S 是它的一个线性无关组. 如果 S 是极大线性无关组, 那么 S 也是极小生成元组.

我们来解释一下, 什么叫 "一个线性无关组是极大的", 或者说怎么理解 "一个极大线性无关组"; 关键就是 "极大", 意思是说, 任意再添加任何一个元素, 构成的新的向量组, 都是线性相关的. 我们做一些推导. 假设 $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 是个极大线性无关组, $\boldsymbol{v} \in V$, 则 $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n, \boldsymbol{v}\}$ 是线性无关的, 所以存在不全为零的实数 $a_1, \cdots, a_n, a$, 使得

$$a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n + a\boldsymbol{v} = 0.$$

我们断言 $a \neq 0.$; 若不然 $a = 0$, 那么 $a_1, \cdots, a_n$ 不全为零, 且 $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n = 0$; 与 $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ 的线性无关性矛盾. 所以断言成立, 于是

$$\boldsymbol{v} = -\frac{a_1}{a}\boldsymbol{v}_1 - \frac{a_2}{a}\boldsymbol{v}_2 - \cdots - \frac{a_n}{a}\boldsymbol{v}_n.$$

这就说明, 对任意的 $\boldsymbol{v} \in V$ 都可以由 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 线性表出; 或者说等价的:

命题 4.46. 若 S 是线性空间 V 的一个极大线性无关组, 那么 $V = \mathrm{Span}(S)$, 即 S 是 V 的一个生成元组.

进一步的, 由推论 (4.41), 我们就证明了定理 (4.45).

推论 4.47. 线性空间 V 的维数等于它的极大线性无关组的元素个数.

例题 4.48. 考虑 $\mathbb{R}^n$, $e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$, $S = \{e_1, e_2, \cdots, e_n\}$. 显然地 $e_1, e_2, \cdots, e_n$ 是线性无关的, 而且 $\mathrm{Span}(S) = \mathbb{R}^n$. 所以 S 是一个极小生成元组. 所以 $\mathbb{R}^n$ 的极大线性无关组的元素个数是 n, 等价地说:

$\mathbb{R}^n$ 中任意 $n+1$ 个向量都是线性相关的.

我们再来说一些抽象废话. 假设 $n+1$ 个向量是 $v_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}$, $1 \le j \le n+1$, $A = (a_{ij})$, 所以

$$\begin{aligned} v_1, v_2, \cdots, v_{n+1} \text{线性相关} &\Leftrightarrow \text{存在不全为零的数 } a_1, a_2, \cdots, a_{n+1}, \text{使得 } a_1 v_1 + a_2 v_2 + \cdots + a_{n+1} v_{n+1} = 0 \\ &\Leftrightarrow \text{存在不全为零的数 } a_1, a_2, \cdots, a_{n+1}, \text{使得 } A \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n+1} \end{bmatrix} = 0 \\ &\Leftrightarrow \text{线性方程组 } A\boldsymbol{x} = 0 \text{ 有非零解.} \end{aligned}$$

而最后一个论断成立, 是因为高斯消元法: 这个齐次方程组有 $n+1$ 个未知数, n 个方程.

讨论完了矩阵 A 的行向量空间, 我们来讨论矩阵的列向量空间.

定理 4.49.

$$\dim R(A) = \dim C(A).$$

和推论 (4.30) 类似, 我们知道, 对 $A \in M_{m\times n}(\mathbb{R})$, 以及可逆矩阵 $X \in M_n(\mathbb{R})$, $\dim C(AX) = \dim C(A)$. 所以, 如果 A 是既约的阶梯形矩阵, 那么上面这个定理显然是对的, (想一想, 为什么?) 所以, 对任意矩阵 A, 都有

$$\dim R(A) = \dim R(\mathrm{ref}(A)) = \dim C(\mathrm{ref}(A)),$$

因此, 要证明这个定理, 我们只需要证明 $\dim C(\mathrm{ref}(A)) = \dim C(A)$. 回顾一下我们证明 $\dim R(\mathrm{ref}(A)) = \dim R(A)$ 的过程, 就可以发现我们只需要证明下面这个命题:

命题 4.50. *初等行变换不改变列空间的维数.*

事实上, 我们证明一个更强的命题:

命题 4.51. *设 $X \in M_m(\mathbb{R})$ 是个可逆矩阵. $A \in M_{m\times n}(\mathbb{R})$, 那么 A 的第 $i_1, \cdots, i_s$ 列线性无关, 当且仅当 XA 的第 $i_1, \cdots, i_s$ 列线性无关.*

证明. 证明: 设 $A = [v_1, \cdots, v_n]$, 那么 $XA = [Xv_1, \cdots, Xv_n]$. 那么

$$\begin{aligned} Xv_{i_1}, \cdots, Xv_{j_s} \text{ 线性相关} &\Leftrightarrow \text{存在不全为零的数 } t_1, \cdots, t_s, \text{使得 } t_1 X v_{i_1} + \cdots + t_s X v_{i_s} = 0 \\ &\Leftrightarrow \text{存在不全为零的数 } t_1, \cdots, t_s, \text{使得 } X(t_1 v_{i_1} + \cdots + t_s v_{i_s}) = 0. \\ &\Leftrightarrow \text{存在不全为零的数 } t_1, \cdots, t_s, \text{使得 } (t_1 v_{i_1} + \cdots + t_s v_{i_s}) = 0. \end{aligned}$$

这最后一个等价是因为 X 是可逆矩阵. $\square$

这就完成了定理 (4.49) 的证明.

作为这个定理的一个显然的推论, 我们有

推论 4.52.

$$\operatorname{rank}(A) \leq \min\{m, n\}$$

4.4 线性方程组的解的结构

我们回顾一下, 我们要解线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$, 其中 $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ 是未知数, $A \in M_{m\times n}(\mathbb{R})$ 是系数矩阵; $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ 是已知的. 通过高斯消元法, 设 X 是可逆矩阵, 使得 XA 是阶梯型矩阵 $\operatorname{ref}(A)$, 因此原方程组和 $\operatorname{ref}(A)\boldsymbol{x} = X\boldsymbol{b}$ 同解. 在之前的章节中, 我们知道 $r(A)$ 是 $\operatorname{ref}(A)$ 的非零行的个数. 并且我们有判定解是否存在及解的个数的结论 (详情请见**第二章: 高斯消元法**):

定理 4.53.

1. $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 当且仅当 $X\boldsymbol{b}$ 从第 $r+1$ 个分量开始都是 0;

2. 若 $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 那么当 $r = n$ 时, 方程组有唯一解;

3. 若 $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 那么当 $r < n$ 时, 方程组有无穷多个解.

这其中 (2) 能成立的必要条件是 $n \leq m$, 这是因为 $n = r \leq \min\{m, n\}$, 当 $\boldsymbol{b} = 0$ 时, 方程组始终都有解: $\boldsymbol{x} = 0$, 因此,

- $r = n$ 时, 方程组有唯一解 $\boldsymbol{x} = 0$;

- $r < n$ 时, 方程组有无穷多个解.

当 $\boldsymbol{b}=0$ 时, 若 $m<n$ 我们有 $r(A)\leq\min\{m,n\}=m<n$, 因此" 方程个数比未知数少的齐次方程组" 一定有无穷多个解. (为什么呢? 因为 m 是方程个数, n 是未知数个数.)

若方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 无解, 那么我们的讨论就终止了. 若方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 有解, 我们任取它的一个解 $\boldsymbol{x}_p$ (即满足 $A\boldsymbol{x}_p=\boldsymbol{b}$), 通常称为特解, 那么我们就可以建立一一对应

$$\left\{\boldsymbol{x}\in\mathbb{R}^n : A\boldsymbol{x}=0\right\}\longleftrightarrow\left\{\boldsymbol{x}\in\mathbb{R}^n : A\boldsymbol{x}=\boldsymbol{b}\right\}\qquad \boldsymbol{x}\mapsto\boldsymbol{x}+\boldsymbol{x}_p.$$

所以, 我们要解方程组 $A\boldsymbol{x}=\boldsymbol{b}$, 只需要先求 (哪怕你是猜出来的, 做梦梦到的都可以) 方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 的一个特解, 然后解齐次线性方程组 $A\boldsymbol{x}=0$.

一方面, 我们知道 $A\boldsymbol{x}=\boldsymbol{b}$ 有解, 当且仅当 $\boldsymbol{b}\in C(A)$; 或者等价地说,

$$\left\{\boldsymbol{b}\in\mathbb{R}^m : \exists\boldsymbol{x}\in\mathbb{R}^n, s.t. A\boldsymbol{x}=\boldsymbol{b}\right\}=C(A).$$

这样的 $\boldsymbol{b}$ 构成了 $\mathbb{R}^m$ 的一个子空间, 其维数为 $\dim C(A)=r(A)$, 这个结论描述了有 "多少" $\boldsymbol{b}$ 使方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 有解.

那么, 下面我们就要解齐次线性方程组 $A\boldsymbol{x}=0$ 了. 注意到

$$N(A):=\left\{\boldsymbol{x}\in\mathbb{R}^n : A\boldsymbol{x}=0\right\}$$

是 $\mathbb{R}^n$ 的线性子空间, 我们参考注释 (2.8) 直接给出我们的结论:

定理 4.54.

$$\dim N(A)=n-r(A).$$

我们来证明下面这个一般的定理, 把定理 4.54 看作是这个定理的推论:

定理 4.55. *设 V 是有限维线性空间, W 是线性空间, $f:V\to W$ 是线性映射, 那么*

$$\dim V=\dim\ker(f)+\dim\operatorname{Im}(f).$$

证明的思路很简单, 就是取 $\ker(f)$ 的一个极大线性无关组和 $\operatorname{Im}(f)$ 的一个极大线性无关组, 把它们拼成 V 的一个极大线性无关组.

证明. 首先我们需要说明的是 $\ker(f)$ 和 $\operatorname{Im}(f)$ 都是有限维的, 前者有限维是显然的. 对于后者, 由于 V 是有限维的, 若有限集合 S 是 V 的一个生成元组, 那么 $f(S):=\{f(s):s\in S\}$ 是 $\operatorname{Im}(f)$ 的一个生成元组, 从而 $\operatorname{Im}(f)$ 是有限维的.

我们设 $\boldsymbol{v}_1 m\cdots,\boldsymbol{v}_r$ 是 $\ker(f)$ 的一个极大线性无关组, $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k$ 是 $\operatorname{Im}(f)$ 的一个极大线性无关组. 由于 $\boldsymbol{w}_i\in\operatorname{Im}(f)$, 我们取 $\boldsymbol{v}_{r+i}\in V$ 使得 $f(\boldsymbol{v}_{r+i})=\boldsymbol{w}_i$ $(1\leq i\leq k)$. 我们断言: $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_r,\boldsymbol{v}_{r+1},\cdots,\boldsymbol{v}_{r+k}$ 是 V 的一个极大线性无关组, 那么我们就有 $\dim(V)=r+k$. 事实上, 我们只需要证明以下两点:

1. $\boldsymbol{v}_1, \cdot, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_{r+k}$ 线性无关.

2. $V = \mathrm{Span}\{\boldsymbol{v}_1, \cdot, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_{r+k}\}$.

我们首先证明 1. 假设 $a_1\boldsymbol{v}_1 + \cdot + a_r\boldsymbol{v}_r + a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_{r+k}\boldsymbol{v}_{r+k} = 0$, 则

$$0 = f(a_1\boldsymbol{v}_1 + \cdot + a_r\boldsymbol{v}_r + a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_{r+k}\boldsymbol{v}_{r+k}) = a_{r+1}\boldsymbol{w}_1 + \cdots + a_{r+k}\boldsymbol{w}_k.$$

因为 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k$ 线性无关, 因此 $a_{r+1} = \cdots = a_{r+k} = 0$. 代回去可得 $a_1\boldsymbol{v}_1 + \cdots + a_r\boldsymbol{v}_r = 0$. 再由 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r$ 的线性无关性可知 $a_1 = \cdots = a_r = 0$. 这就证明了 1.

我们再来证明 2. 对任意 $\boldsymbol{v} \in V$, 因为 $f(v) \in \mathrm{Im}(f)$, 因此存在 $a_{r+1}, \cdots, a_{r+k}$ 使得 $f(v) = a_{r+1}\boldsymbol{w}_1 + \cdots + a_{r+k}\boldsymbol{w}_k = a_{r+1}f(\boldsymbol{v}_{r+1}) + \cdots + a_{r+k}f(\boldsymbol{v}_{r+k})$, 因此

$$f(\boldsymbol{v} - (a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_{r+k}\boldsymbol{v}_{r+k})) = 0;$$

即: $\boldsymbol{v} - (a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_{r+k}\boldsymbol{v}_{r+k}) \in \ker(f)$, 所以存在 $a_1, \cdots, a_r$ 使得

$$\boldsymbol{v} - (a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_{r+k}\boldsymbol{v}_{r+k}) = a_1\boldsymbol{v}_1 + \cdots + a_r\boldsymbol{v}_r,$$

这就是说: $\boldsymbol{v} = a_1\boldsymbol{v}_1 + \cdots + a_r\boldsymbol{v}_r + a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_{r+k}\boldsymbol{v}_{r+k}$. 证毕. □

这个定理被称作维数公式.

定理 4.54的证明. 考虑映射

$$f : \mathbb{R}^n \to \mathbb{R}^m; \quad \boldsymbol{x} \mapsto A\boldsymbol{x},$$

那么根据定义我们就有: $N(A) = \ker(f), C(A) = \mathrm{Im}(f)$. 利用维数公式就证完了. □

定义 4.56. 我们把 $N(A)$ 的一组基称作齐次线性方程组 $A\boldsymbol{x} = 0$ 的一个基础解系.

我们记 A' 为方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的增广矩阵, 即 $A' = [A, \boldsymbol{b}]$, 那么

定理 4.57. 1. $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 当且仅当 $\mathrm{rank}(A) = \mathrm{rank}(A')$. (这是定理 (4.53) 的重新叙述而已; 请同学们参照矩阵秩的定义, 验证这一点.)

2. 设 $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 且 $\boldsymbol{x}_p$ 是它的一个特解, 那么 $A\boldsymbol{x} = \boldsymbol{b}$ 的所有解为 $\{\boldsymbol{x}_p + \boldsymbol{x} : \boldsymbol{x} \in N(A)\}$. 其中 $\dim N(A) = n - r(A)$.

4.4.1 特殊秩的矩阵

4.4.2 满秩方阵

设 $A \in M_n(\mathbb{R})$ 是一个满秩的方阵: 所谓满秩方阵, 是说 $\mathrm{rank}(A) = n$; 或者等价地说 A 的行向量线性无关; 或者等价地说 A 的列向量线性无关. 那么根据第一章定义 (定义 1.2), 这样的矩阵是可逆矩阵. 但是在第三章中, 我们给了可逆矩阵的第二个定义 (定义 3.5, 推论 3.15): 存在矩阵 $B \in M_n(\mathbb{R})$ 使得 $AB = BA = I_n$.

定理 4.58. *上述两个定义等价.*

证明. 一方面, 我们证明, 如果 A 的列向量线性无关, 则存在矩阵 B, 使得 $AB = I_n$. 对 $i = 1, 2, \cdots, n$, 考虑 $\boldsymbol{e}_i$ 为第 i 个单位向量: 第 i 个分量为 1, 其他分量为 0. 由于 A 的列向量线性无关, 那么它的列向量构成 $\mathbb{R}^n$ 的一组基, 所以 $\boldsymbol{e}_i$ 可由 A 的列向量线性表出, 从而存在 $\beta_i \in \mathbb{R}^n$, 使得 $A\beta_i = \boldsymbol{e}_i$. 令 $B = [\beta_1, \beta_2, \cdots, \beta_n]$, 那么

$$AB = [A\beta_1, A-\beta_2, \cdots, A\beta_n] = [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n] = I_n.$$

另一方面, 我们证明, 若存在 $AB = BA = I_n$, 那么 A 的列向量线性无关. 事实上, 设 $A = [\alpha_1, \alpha_2, \cdots, \alpha_n]$, α_i 是 A 的列向量. 假设 $x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = 0$, 用矩阵乘法来写, 就是

$$A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = 0.$$

左乘 B, 由于 $BA = I_n$, 就得到 $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = 0$. 这就说明的 A 的列向量线性无关. □

总而言之, 我们得到了下面的结论:

命题 4.59. *设 $A \in M_n(\mathbb{R})$, 以下各条等价:*

1. *A 的行向量线性无关;*
2. *A 的列向量线性无关;*
3. *A 有左逆;*

4. A 有右逆;

5. 对任意 $\boldsymbol{b} \in \mathbb{R}^n$, 线性方程组至多有一个解;

6. 对任意 $\boldsymbol{b} \in \mathbb{R}^n$, 线性方程组都有一个解;

4.4.3 行满秩矩阵

设 $A \in M_{m\times n}$. 注意, A 只是个矩阵, 未必是方阵.

定义 4.60. 我们称 A 是行满秩矩阵, 如果 A 的行向量线性无关.

首先注意到, 因为 $rank(A) \le \min\{m,n\}$, 所以当 A 行满秩的时候 ($\operatorname{rank}(A) = m$), 有 $m \le n$. 此时矩阵看起来比较" 宽".

命题 4.61. 若 A 行满秩, 那么对任意 $\boldsymbol{b} \in \mathbb{R}^m$, 线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 都有解.

证明. 由于矩阵的行秩等于列秩, 所以 $\dim C(A) = m$. 而 $C(A) \subseteq \mathbb{R}^m$, 所以 $C(A) = \mathbb{R}^m$. 我们又知道, $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 当且仅当 $\boldsymbol{b} \in C(A)$. 这就证明了我们的论断. □

推论 4.62. 若 A 行满秩, 则存在 $B \in M_{n\times m}(\mathbb{R})$ 使得 $AB = I_m$.

简而言之, "若 A 行满秩, 则 A 有右逆".

证明. 对 $i = 1, 2, \cdots, m$, 考虑 $\boldsymbol{e}_i$ 为第 i 个单位向量: 第 i 个分量为 1, 其他分量为 0. 由于 A 的列向量线性无关, 那么它的列向量构成 $\mathbb{R}^n$ 的一组基, 所以 $\boldsymbol{e}_i$ 可由 A 的列向量线性表出, 从而存在 $\beta_i \in \mathbb{R}^n$, 使得 $A\beta_i = \boldsymbol{e}_i$. 令 $B = [\beta_1, \beta_2, \cdots, \beta_m] \in M_{n\times m}(\mathbb{R})$, 那么

$$AB = [A\beta_1, A - \beta_2, \cdots, A\beta_m] = [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m] = I_m.$$

□

另一方面, 若 A 有右逆 B, i.e. $AB = I_m$, 则对任意 $\boldsymbol{b} \in \mathbb{R}^m$, $B\boldsymbol{b}$ 是方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的解, 所以方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 所以 $\dim C(A) = m$, 所以 $\operatorname{rank}(A) = m$, 所以 A 行满秩. 所以我们有

命题 4.63. 设 $A \in M_{m\times n}(\mathbb{R})$, 以下各条等价:

1. A 的行向量线性无关;

2. A 有右逆;

3. 对任意 $\boldsymbol{b} \in \mathbb{R}^m$, 线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 都有一个解;

4.4.4 列满秩矩阵

设 $A \in M_{m\times n}$. 注意, A 只是个矩阵, 未必是方阵.

定义 4.64. 我们称 A 是列满秩矩阵, 如果 A 的列向量线性无关.

首先注意到, 因为 $rank(A) \le \min\{m,n\}$, 所以当 A 列满秩的时候 $(\mathrm{rank}(A) = n)$, 有 $n \le m$.

命题 4.65. 若 A 列满秩, 那么对任意 $\boldsymbol{b} \in \mathbb{R}^n$, 线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 至多有一个解. (等价地说, 解如果存在, 则唯一)

证明. 若 $\boldsymbol{b} \in \mathbb{R}^m$ 使得方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 无解, 那么没什么可证的.

若 $\boldsymbol{b} \in \mathbb{R}^m$ 使得方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 有解, 我们需要证明解是唯一的. 设 $\boldsymbol{x}_1, \boldsymbol{x}_2$ 都是该方程组的解, 那么 $A(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 0$. 而 $A(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 0$ 是 A 的列向量的线性组合, 根据 A 列满秩, 我们知道 $\boldsymbol{x}_1 - \boldsymbol{x}_2 = 0$, 即 $\boldsymbol{x}_1 = \boldsymbol{x}_2$, 也即解唯一. □

反之也对:

命题 4.66. 若对任意 $\boldsymbol{b} \in \mathbb{R}^n$, 线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 至多有一个解, 则 A 列满秩.

证明. 根据条件, 特别地 $A\boldsymbol{x} = 0$ 至多有一个解. 而 $\boldsymbol{x} = 0$ 是它的解, 所以它有唯一解 $\boldsymbol{x} = 0$. 这表明 A 的各列线性无关, 所以 A 列满秩. □

利用矩阵转置, 我们知道: A 列满秩等价于 A^T 行满秩, 等价于 A^T 有右逆 $B : A^TB = I_n$, 等价于 $B^TA = I^n$, 等价于 A 有左逆. 于是我们得到:

命题 4.67. 设 $A \in M_{m\times n}(\mathbb{R})$, 以下各条等价:

1. A 的列向量线性无关;
2. A 有左逆逆;
3. 对任意 $\boldsymbol{b} \in \mathbb{R}^m$, 线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 至多有一个解;

4.4.5 秩为 1 的矩阵

设 $A \in M_{m\times n}(\mathbb{R})$, 且它的秩等于 1. 那么取定它的一个非零列向量, 记为 α, 那么 A 的第 i 个列向量一定是 α 的 a_i 倍, 所以

$$A = [a_1\alpha, a_2\alpha, \cdots, a_n\alpha].$$

所以 $A=\alpha[a_1,a_2,\cdots,a_n]$. 对 $\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\in\mathbb{R}^n$, 我们有

$$A\boldsymbol{x}=\alpha[a_1,a_2,\cdots,a_n]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\alpha(a_1x_1+a_2x_2+\cdots+a_nx_n).$$

因为 α 是非零向量, 所以方程组 $A\boldsymbol{x}=0$ 等价于

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

这和我们之前的结论一致: $A\boldsymbol{x}=0$ 的有效方程个数等于 $\mathrm{rank}(A)=1$.

4.4.6 行简化阶梯型

之前关于“行简化阶梯型矩阵”的叙述实际上是关于“阶梯形矩阵”, 而不是行简化阶梯型. 在这里我们做一些更正. 由于行简化阶梯型相比阶梯形矩阵, 从数学理论上来讲, 并没有什么简化, 只是在实际计算的时候, 更有利于我们求解. 因此我们用一个例子来区分“行简化阶梯型”和“阶梯形矩阵”, 并用它来演示“行简化阶梯型”的作用.

例题 4.68. 解线性方程

$$\begin{cases}\begin{array}{rrrrrl} & -3x_2 & -9x_3 & +4x_4 & +9x_5 & =0\\ -x_1 & -2x_2 & +x_3 & +3x_4 & +x_5 & =0\\ -2x_1 & -3x_2 & & +3x_4 & -x_5 & =0\\ x_1 & +4x_2 & +5x_3 & -9x_4 & -7x_5 & =0\end{array}\end{cases}$$

系数矩阵写出来, 就是

$$\begin{bmatrix}0&-3&-9&4&9\\-1&-2&1&3&1\\-2&-3&0&3&-1\\1&4&5&-9&-7\end{bmatrix}$$

我们对它作初等行变换

$$\begin{bmatrix} 0 & -3 & -9 & 4 & 9 \\ -1 & -2 & 1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & 1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & 9 & 4 & 9 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 6 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -9 & 4 & 9 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 6 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 6 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

这是行阶梯形矩阵.

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 6 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 1 & 3 & -3 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -7 & 3 & 5 \\ 0 & 1 & 3 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -7 & 0 & 5 \\ 0 & 1 & 3 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

这是既约行阶梯形矩阵矩阵. 它比行阶梯形矩阵" 既约" 在哪些方面呢?

- 对行阶梯形矩阵矩阵的每一个非零行除以该行的第一个非零元 (就是所谓的主元), 把第一个非零元变成 1.
- 用每一行的第一个非零元, 对它所在的列做消元, 使得它所在的列只有唯一的非零元.

实际上, 我们就是对行阶梯形矩阵做以上两个操作, 变成了既约行阶梯形矩阵.

注释 4.69. 设矩阵 $A \in M_{m\times n}(\mathbb{R})$, 我们用 $\mathrm{ref}(A)$ 来记它的行阶梯形矩阵, 用 $\mathrm{rref}(A)$ 来记它的既约行阶梯形矩阵. 我们已经知道 $\mathrm{ref}(A)$ 并不唯一, 那么 $\mathrm{rref}(A)$ 是否唯一呢?

命题 4.70. 对任意矩阵 $A \in M_{m\times n}(\mathbb{R})$, $\mathrm{rref}(A)$ 是一个由 A 唯一决定的矩阵.

它的证明不是很明显, 但是利用我们已有的知识可以解决. 然而, 这个结论到目前为止, 对我们的知识结构并没有什么帮助, 所以我们就不在这里证明了.

4.5 概念和主要内容总结

我们来回顾一下我们这一部分的主要概念为:

1. 线性空间的概念, 线性空间由一个子集 S 张成的子空间 $\mathrm{Span}(S)$ 的概念;
2. 线性空间向量的线性相关性和线性无关性;
3. 有限维线性空间中, 极小生成元组和极大线性无关组等价;
4. 线性空间的维数 = 极小生成元组的元素个数 = 极大线性无关组的元素个数.
5. 矩阵的行向量张成的空间 (简称 "行空间") 和列向量张成的空间 (简称 "列空间")

这一部分的主要内容:

1. 首先我们回顾了矩阵 A 的秩 $\mathrm{rank}(A)$ 的概念, 即 $\mathrm{ref}(A)$ 的非零行的个数; 这个被称为 "行秩";
2. 我们证明了矩阵 A 的行秩和 $\mathrm{ref}(A)$ 的形式无关, 是个良定义的数 (这个不重要, 如果你不明白什么意思或者看不懂证明, 那背下来这一句话就行了)
3. 矩阵 A 的行秩 $\mathrm{rank}(A)$ 等于它的行空间的维数 $\dim R(A)$;
4. 矩阵 A 的行秩 $\mathrm{rank}(A)$ 也等于它的列空间的维数 $\dim C(A)$; 相应地, 也等于列秩 (我们可以做初等列变换, 考虑" 既约列阶梯型矩阵" 的非零列的个数).

所以, 我们往往可以混淆 "行秩, 列秩, 行空间维数, 列空间维数" 等等这些个本质上表达了同一个量的概念, 请同学们克服心理困难, 记住它们; 并且在必要的场合区分它们 (例如考题). 我们把这个量统记作 $r(A)$ 或者 $\mathrm{rank}(A)$.

我们总结一下:

定义 4.71. 设 V 是一个有限维线性空间, 我们把它的一个极小生成元组 (或者等价地, 极大线性无关组) 称为是 V 的一组基 (或者基底).

也就是说, "基 = 基底 = 极小生成元组 = 极大线性无关组."

第五章　正交

5.1　基本概念和例子

在上一章, 我们已经解决了解线性方程组的问题. 但是, 从不同的角度来看待解方程, 是我们不断挖掘数学内容的一个重要方法. 我们这一章就从向量点积的角度来看这个问题. 怎么看呢? 这来源于我们理解矩阵乘法的一个方式. 我们把矩阵 $A \in M_n(\mathbb{R})$ 写成列向量的形式:

$$A = \begin{bmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \alpha_m^T \end{bmatrix}, \quad \alpha_i \in \mathbb{R}^n.$$

那么矩阵乘法 $A\boldsymbol{x}$ 就可以写成

$$A\boldsymbol{x} = \begin{bmatrix} \alpha_1 \cdot \boldsymbol{x} \\ \alpha_2 \cdot \boldsymbol{x} \\ \vdots \\ \alpha_m \cdot \boldsymbol{x} \end{bmatrix}.$$

因此解齐次线性方程组 $A\boldsymbol{x} = 0$ 等价于求那些和 $\alpha_1, \alpha_2, \cdots, \alpha_m$ 的点积都为零的 $\boldsymbol{x} \in \mathbb{R}^n$, 等价地说, 求那些 $\boldsymbol{x} \in \mathbb{R}^n$ 使得 $\boldsymbol{x}$ 和 $\alpha_1, \alpha_2, \cdots, \alpha_m$ 都正交.

我们现在改记号:

设 $\alpha, \beta \in \mathbb{R}^n$, α 和 β 的点积我们不再记作 $\alpha \cdot \beta$, 而是改记作 $\langle \alpha, \beta \rangle$.

同时, 我们也把 "点积" 称作" **内积**".

对这个记号, 我们作一点解释.

命题 5.1. 设 $A \in M_{m\times n}(\mathbb{R}), \boldsymbol{v} \in \mathbb{R}^n, \boldsymbol{w} \in \mathbb{R}^m$, 那么

$$\langle A\boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{v}, A^T\boldsymbol{w} \rangle.$$

证明. 这是因为

$$\langle Av, w\rangle = (Av)\cdot w = (Av)^T w = v^T(A^T w) = v\cdot(A^T w) = \langle v, A^T w\rangle.$$

□

在这样的记号之下, 重复上面的叙述, 我们有

$$N(A) = \left\{x \in \mathbb{R}^n : \langle x, \alpha_i\rangle = 0, \forall 1 \le i \le m\right\}.$$

定义 5.2. 1. $v, w \in \mathbb{R}^n$ 称为是正交的, 如果 $\langle v, w\rangle = 0$. 这时我们记作 $v \perp w$.

2. 设 V, W 是 $\mathbb{R}^n$ 中的线性子空间, 我们称 V, W 是正交的, 如果对任意的 $v \in V, w \in W$, 都有 $\langle v, w\rangle = 0$. 这时我们记作 $V \perp W$.

例题 5.3.

1. 在 $\mathbb{R}^2$ 中, 两个坐标轴的单位向量 e_1, e_2 是互相正交的. 一般地, 在 $\mathbb{R}^n$ 中, n 个坐标轴的单位向量 $e_1, e_2, \cdots, e_n$ 都是两两正交的.
2. 在 $\mathbb{R}^3$ 中, 设 V 是 xOy 平面, W 是 yOz 平面, 那么 V, W 不是正交的: 因为 y 轴的单位向量 e_2 同时属于 V, W, 我们取 $v = e_2 \in V, w = e_2 \in W$, 那么就有 $\langle v, w\rangle \neq 0$. 设 W' 是 z 轴, 那么 V, W' 是正交的.

定义 5.4. 设 $V \subseteq \mathbb{R}^n$ 是个集合, 定义

$$V^\perp := \left\{x \in \mathbb{R}^n : \langle x, v\rangle = 0, \forall v \in V\right\}.$$

引理 5.5. 设 $S \subseteq \mathbb{R}^n$ 是个集合, $V = \mathrm{Span}(S)$, 那么

$$S^\perp = V^\perp$$

且 $V^\perp$ 是个线性空间.

证明. 我们只需要证明 $S^\perp \subseteq V^\perp$ 和 $S^\perp \supseteq V^\perp$.

“$\subseteq$”: $x \in S^\perp$. 要证明 $x \in V^\perp$ 等价于证明对任意 $v \in V$, 我们都有 $\langle x, v\rangle = 0$. 而 $V = \mathrm{Span}(S)$, 所以 $v = a_1 s_1 + a_2 s_2 + \cdots + a_n s_n$, $s_i \in S, a_i \in \mathbb{R}$. 从而

$$\langle x, v\rangle = \left\langle x, \sum_{i=1}^n a_i s_i\right\rangle = \sum_{i=1}^n a_i \langle x, s_i\rangle = 0.$$

“$\supseteq$”: $x \in V^\perp$, 因为 $S \subseteq \mathrm{Span}(S)$, 故任意的 $s \in S \subseteq V$, 就有 $\langle x, s\rangle = 0$. □

注释 5.6. 我们定义 $V^{\perp}$ 时, 只需要 V 是个集合, 并不要求它是线性子空间, 例如 V 可以只含有一个元素的集合. 但是通过上面的引理, 我们只需要考虑 V 是个线性空间的情形.

受例题 (5.3 2) 启发, 我们有:

命题 5.7. 若 $V \subseteq \mathbb{R}^n$ 是个线性子空间, 那么 $V \cap V^{\perp} = \{0\}$.

证明. 设 $\boldsymbol{v} \in V \cap V^{\perp}$, 那么一方面 $\boldsymbol{v} \in V$, 另一方面 $\boldsymbol{v} \in V^{\perp}$, 所以 $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0$, 所以 $\boldsymbol{v} = 0$. □

例题 5.8.

1. 在 $\mathbb{R}^2$ 中, 取 V 为 x 轴, 那么 $V^{\perp}$ 等于 y 轴.
2. 在 $\mathbb{R}^3$ 中, 取 V 为 x 轴, 那么 $V^{\perp}$ 等于 yOz 平面. 取 V 为 yOz 平面, 那么 $V^{\perp}$ 等于 x 轴.

在这样的术语之下, 因为 $R(A) = C(A^T)$, 我们有

命题 5.9. $C(A^T)^{\perp} = N(A)$.

在上面这个命题中, 用 A^T 来替换 A, 有

$$C(A)^{\perp} = C((A^T)^T)^{\perp} = N(A^T).$$

这一条显得很突兀, 几乎完全是 “纯粹逻辑” 的产物. 那么我们要试图理解一下这件事. 要理解它, 我们就要先理解 $N(A^T)$ 是什么呢?

$$N(A^T) = \left\{ \boldsymbol{y} \in \mathbb{R}^m : A^T \boldsymbol{y} = 0 \right\} = \left\{ \boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y}^T A = 0 \right\}.$$

所以 $N(A^T)$ 可以等同于方程组 $\boldsymbol{y}A = 0$ 的解, 也就是所谓的 “A 的 ‘左零空间’”. 设 $\boldsymbol{v} \in C(A), \boldsymbol{y} \in N(A^T)$, 那么存在 $\boldsymbol{x} \in \mathbb{R}^n$, 使得 $\boldsymbol{v} = A\boldsymbol{x}$. 所以

$$\langle \boldsymbol{v}, \boldsymbol{y} \rangle = \langle A\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, A^T \boldsymbol{y} \rangle = 0.$$

至此: 我们所说的 “四个基本子空间” 就全出现了:

- 行向量空间 $C(A^T) \subseteq \mathbb{R}^n$;
- 零空间 $N(A) \subseteq \mathbb{R}^n$;
- 列向量空间 $C(A) \subseteq \mathbb{R}^m$;
- 左零空间 $N(A^T) \subseteq \mathbb{R}^m$.

它们的关系是:

- $\dim C(A) = \dim C(A^T) = r$;

- $\dim C(A) + \dim N(A) = n$;
- $\dim C(A^T) + \dim N(A^T) = m$;
- $C(A^T)^\perp = N(A)$;
- $C(A)^\perp = N(A^T)$.

此处应有图, 见教材 198 页. 但是我不会画.

然后我找梁鑫老师帮我画了, 如下:

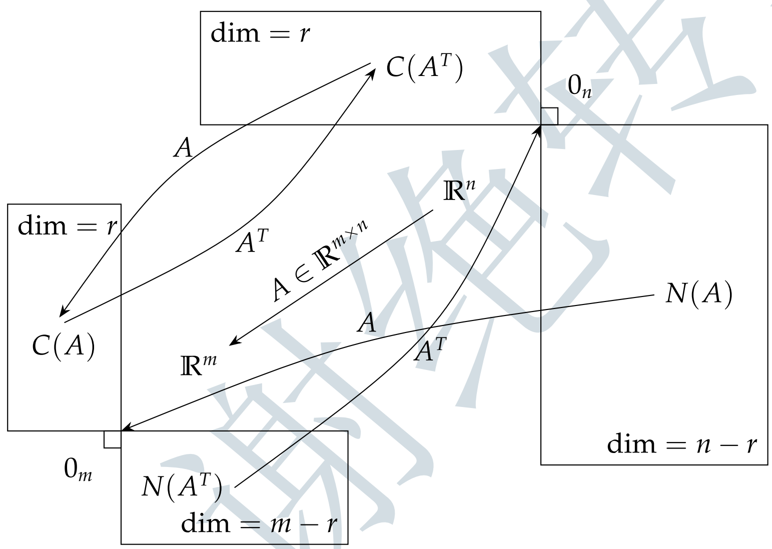


图 5.1: 矩阵导出的四个子空间关系图

关于图 (5.1) 的几点注释：

1. 四个矩形表示四个子空间，维数如图，而图上的相对位置表示：
   (a) $C(A) \perp N(A^T)$, $C(A) \cap N(A^T) = \{0_m\}$, $C(A) + N(A^T) = \mathbb{R}^m$;
   (b) $C(A^T) \perp N(A)$, $C(A^T) \cap N(A) = \{0_n\}$, $C(A^T) + N(A) = \mathbb{R}^n$。
2. $A \in M_{m\times n}(\mathbb{R})$ 定义了一个从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的一个线性映射：
   (a) A 把 $N(A)$ 映射到 $\{0_m\}$，把 $C(A^T)$ 映射到 $C(A)$；
   (b) A^T 把 $N(A^T)$ 映射到 $\{0_n\}$，把 $C(A)$ 映射到 $C(A^T)$。
3. 进一步地，

(a) $C(A^TA) = C(A^T)$, $N(A^TA) = N(A)$;

(b) $C(AA^T) = C(A)$, $N(AA^T) = N(A^T)$;

(c) $A\colon C(A^T) \to C(A)$ 和 $A^T\colon C(A) \to C(A^T)$ 都是双射。

例题 5.10. 考虑 $\mathbb{R}^3$, 直观的我们可以看到若 $V \subseteq \mathbb{R}^3$ 是 1 维的, 那么 $V^\perp$ 是 2 维的; 若 V 是 2 维的, 那么 $V^\perp$ 是 1 维的. 也就是我们始终有 $\dim V + \dim V^\perp = 3$.

一般的, 假设 $V \subseteq \mathbb{R}^n$ 是个线性子空间, 则

命题 5.11. $\dim V + \dim V^\perp = n$.

证明. 设 $r = \dim V$, 取 V 的一组基: $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r \in \mathbb{R}^n$, 它们线性无关. 令矩阵 $A \in M_{r\times n}(\mathbb{R})$ 使得它的行向量分别为 $\boldsymbol{v}_1^T, \boldsymbol{v}_2^T, \cdots, \boldsymbol{v}_r^T$, 即

$$A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix}.$$

那么 $\operatorname{rank}(A) = r$. 另一方面, 由 $V^\perp$ 的定义和引理 (5.5) 知

$$V^\perp = N(A).$$

所以

$$\dim V^\perp = \dim N(A) = n - \operatorname{rank}(A) = n - r.$$

□

这意味着什么呢? 这意味着我们用 V 和 $V^\perp$ 拼成整个空间 $\mathbb{R}^n$!

定理 5.12. 设 $V \subseteq \mathbb{R}^n$ 是线性子空间, 那么对于任意的 $\alpha \in \mathbb{R}^n$, 我们都可以把 α 唯一地写成 $\alpha = \boldsymbol{v} + \boldsymbol{w}$, 其中 $\boldsymbol{v} \in V$, $\boldsymbol{w} \in V^\perp$.

证明. 我们先解释一下唯一性. 所谓的唯一性, 也就是说, 把 α 拆成 V 和 $V^\perp$ 中元素的和的拆法只有一种, 等价地说, 如果 $\alpha = \boldsymbol{v}_1 + \boldsymbol{w}_1 = \boldsymbol{v}_2 + \boldsymbol{w}_2, \boldsymbol{v}_i \in V, \boldsymbol{w}_i \in V^\perp$, 那么 $\boldsymbol{v}_1 = \boldsymbol{v}_2, \boldsymbol{w}_1 = \boldsymbol{w}_2$. 唯一性是容易的, 因为: 我们假设 $\alpha = \boldsymbol{v}_1 + \boldsymbol{w}_1 = \boldsymbol{v}_2 + \boldsymbol{w}_2, \boldsymbol{v}_i \in V, \boldsymbol{w}_i \in V^\perp$, 就有 $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \boldsymbol{w}_2 - \boldsymbol{w}_1$; 这个式子左边属于 V, 而右边属于 $V^\perp$, 所以 $\boldsymbol{v}_1 - \boldsymbol{v}_2 = \boldsymbol{w}_2 - \boldsymbol{w}_1 \in V \cap V^\perp = \{0\}$, 这就得到了 $\boldsymbol{v}_1 = \boldsymbol{v}_2, \boldsymbol{w}_1 = \boldsymbol{w}_2$.

现在我们证明存在性. 设 $r = \dim V$, 根据命题 (5.11), 我们取 V 的一组基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$, $V^\perp$ 的一组基 $\boldsymbol{v}_{r+1}, \boldsymbol{v}_{r+2}, \cdots, \boldsymbol{v}_n$ (共 $n - r$ 个向量), 如果我们能证明这两组基合在一起是线性无

关的, 那么我们就找到了 V 的一个由 n 个元素构成的线性无关组 (此处请回顾维数的定义 (定义 4.71)), 我们就知道它们形成了 V 的一组基, 从而任何 $\alpha \in \mathbb{R}^n$ 都能写成 $\alpha = (a_1\boldsymbol{v}_1 + \cdots + a_r\boldsymbol{v}_r) + (a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_n\boldsymbol{v}_n)$. 我们取 $\boldsymbol{v} = a_1\boldsymbol{v}_1 + \cdots + a_r\boldsymbol{v}_r, \boldsymbol{w} = a_{r+1}\boldsymbol{v}_{r+1} + \cdots + a_n\boldsymbol{v}_n$ 即可. 所以现在只剩下要证明 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 是线性无关的. 设

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_r\boldsymbol{v}_r + x_{r+1}\boldsymbol{v}_{r+1} + \cdots + x_n\boldsymbol{v}_n = 0.$$

那么 $x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_r\boldsymbol{v}_r = -(x_{r+1}\boldsymbol{v}_{r+1} + \cdots + x_n\boldsymbol{v}_n)$, 这个式子左边属于 V, 右边属于 $V^\perp$, 从而它们都属于 $V \cap V^\perp = \{0\}$, 即

$$x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_r\boldsymbol{v}_r = -(x_{r+1}\boldsymbol{v}_{r+1} + \cdots + x_n\boldsymbol{v}_n) = 0.$$

由于 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r$ 是 V 的一组基, 它们是线性无关的, 所以 $x_1 = x_2 = \cdots = x_r = 0$; 同理 $x_{r+1} = x_{r+2} = \cdots = x_n = 0$. 证毕. □

定义 5.13. 设 $V \subseteq \mathbb{R}^n$, 我们把 $V^\perp$ 称为 V 的正交补.

上面的定理解释了 “补” 是什么意思.

推论 5.14.

$$(V^\perp)^\perp = V.$$

证明. $V \subseteq (V^\perp)^\perp$: 要证任意的 $\boldsymbol{v} \in V$, $\boldsymbol{v} \in (V^\perp)^\perp = \left\{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{x}, \boldsymbol{w} \rangle = 0, \forall \boldsymbol{w} \in V^\perp\right\}$. 根据 $V^\perp$ 的定义, 这是显然的.

$(V^\perp)^\perp \subseteq V$: 对任意 $\alpha \in (V^\perp)^\perp \subseteq \mathbb{R}^n$, 设 $\alpha = \boldsymbol{v} + \boldsymbol{w}$, 其中 $\boldsymbol{v} \in V, \boldsymbol{w} \in V^\perp$. 一方面, $\alpha \in (V^\perp)^\perp$, 所以 $\langle \alpha, \boldsymbol{w} \rangle = 0$; 另一方面 $\boldsymbol{w} \in V^\perp$, 所以 $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$. 那么

$$0 = \langle \alpha, \boldsymbol{w} \rangle = \langle \boldsymbol{v} + \boldsymbol{w}, \boldsymbol{w} \rangle = \langle \boldsymbol{v}, \boldsymbol{w} \rangle + \langle \boldsymbol{w}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{w} \rangle.$$

因此 $\boldsymbol{w} = 0$, 也就是说 $\alpha = \boldsymbol{v} \in V$. □

证明. 同上, 我们易知 $V \subseteq (V^\perp)^\perp$, 而且它们的维数相等 (都等于 $n - \dim V^\perp$), 所以这两个线性空间相等. □

5.2 投影和正交分解

首先, 我们从简单的情形出发, 考虑 $\mathbb{R}^2$ 且 V 是 1 维的: $V = \mathbb{R} \cdot \boldsymbol{v}$. 那么投影是什么意思呢?

此处应有一张图, 但是我不会画.

那么 $V^{\perp}$ 也是一维的, 所以我们不妨设 $V^{\perp}=\mathbb{R}\cdot\boldsymbol{w}$. 因为 $\boldsymbol{v},\boldsymbol{w}$ 线性无关 (为什么? 请自己证明一下), 因此 $\{\boldsymbol{v},\boldsymbol{w}\}$ 构成了 $\mathbb{R}^2$ 的一组基. 所以任意的 $\alpha\in\mathbb{R}^2$ 都可以写成 $\alpha=a\boldsymbol{v}+b\boldsymbol{w}$. 我们来确定一下系数:

$$\langle\alpha,\boldsymbol{v}\rangle=\langle a\boldsymbol{v}+b\boldsymbol{w},\boldsymbol{v}\rangle=a\langle\boldsymbol{v},\boldsymbol{v}\rangle+b\langle\boldsymbol{v},\boldsymbol{w}\rangle,$$

所以 $a=\dfrac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}$, 同理 $b=\dfrac{\langle\alpha,\boldsymbol{w}\rangle}{\langle\boldsymbol{w},\boldsymbol{w}\rangle}$. 所以

$$\alpha=\frac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}+\frac{\langle\alpha,\boldsymbol{w}\rangle}{\langle\boldsymbol{w},\boldsymbol{w}\rangle}\cdot\boldsymbol{w}.$$

通过初等的平面几何 (想象一下上面那张我画不出来的图), 我们知道, $\dfrac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}$ 是 V 上使得 $\boldsymbol{v}$ 和它的差有最小的长度的向量, 或者说, 对任意 $\boldsymbol{x}\in V$, 我们都有

$$\|\alpha-\boldsymbol{x}\|\geq\left\|\alpha-\frac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}\right\|.$$

等号成立当且仅当 $\boldsymbol{x}=\dfrac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}$.

现在考虑一般的 n, 但是 $V=\mathbb{R}\cdot\boldsymbol{v}$ 依然是一维的. 假设 $\alpha\in\mathbb{R}^n$ 被写成了 $\alpha=a\boldsymbol{v}+b\boldsymbol{w}$, 那么同样的方法, 我们有 $a=\dfrac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}$, 所以

$$\alpha=\frac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}+\left(\alpha-\frac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}\right).$$

命题 5.15. 对任意 $\boldsymbol{x}\in V$, 我们都有

$$\|\alpha-\boldsymbol{x}\|\geq\left\|\alpha-\frac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}\right\|.$$

等号成立当且仅当 $\boldsymbol{x}=\dfrac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}$.

证明. 设 $\boldsymbol{x}=t\boldsymbol{v}$, 那么

$$\|\alpha-\boldsymbol{x}\|^2=\langle\alpha-t\boldsymbol{v},\alpha-t\boldsymbol{v}\rangle=\langle\boldsymbol{v},\boldsymbol{v}\rangle t^2-2t\langle\alpha,\boldsymbol{v}\rangle+\langle\alpha,\alpha\rangle.$$

这个关于 t 的二次函数当且仅当 $t=\dfrac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}$ 取到最小值, 此时最小值为

$$\left\|\alpha-\frac{\langle\alpha,\boldsymbol{v}\rangle}{\langle\boldsymbol{v},\boldsymbol{v}\rangle}\cdot\boldsymbol{v}\right\|^2$$

□

现在我们考虑一般的 n 和一般的 V.

命题 5.16. 设 $\alpha \in \mathbb{R}^n$ 且 $\alpha = \boldsymbol{v} + \boldsymbol{x}$, 其中 $\boldsymbol{v} \in V$. 则 $\|\boldsymbol{x}\|$ 取最小值时, 当且仅当 $\boldsymbol{x} \in V^{\perp}$.

证明. 由定理 (5.12), 我们 $\alpha = \boldsymbol{w} + \boldsymbol{y}$, 其中 $\boldsymbol{w} \in V, \boldsymbol{y} \in V^{\perp}$, 于是 $\boldsymbol{x} = (\boldsymbol{w} - \boldsymbol{v}) + \boldsymbol{y}$. 那么

$$\begin{aligned} \|\boldsymbol{x}\|^2 = \|(\boldsymbol{w}-\boldsymbol{v})+\boldsymbol{y}\|^2 &= \langle (\boldsymbol{w}-\boldsymbol{v})+\boldsymbol{y}, (\boldsymbol{w}-\boldsymbol{v})+\boldsymbol{y} \rangle \\ &= \langle \boldsymbol{w}-\boldsymbol{v}, \boldsymbol{w}-\boldsymbol{v} \rangle - 2\langle \boldsymbol{w}-\boldsymbol{v}, \boldsymbol{y} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle \\ &= \langle \boldsymbol{w}-\boldsymbol{v}, \boldsymbol{w}-\boldsymbol{v} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle \geq \langle \boldsymbol{y}, \boldsymbol{y} \rangle \end{aligned}$$

最后一个等号是因为 $\boldsymbol{w} - \boldsymbol{v} \in V, \boldsymbol{y} \in V^{\perp}$, 最后一个不等号是平凡的: $\langle \boldsymbol{w}-\boldsymbol{v}, \boldsymbol{w}-\boldsymbol{v} \rangle \geq 0$. □

这样, 正交投影的几何意义就很明显了:

$$\boldsymbol{v} \text{ 是 } \alpha \text{ 到 } V \text{ 的正交投影, 当且仅当 } \|\alpha - \boldsymbol{v}\| = \min\left\{ \|\alpha - \boldsymbol{x}\| : \boldsymbol{x} \in V \right\}.$$

或者说 “减去投影之后的向量长度取到最小值.”

现在, 我们用映射的观点来看待正交投影. 依然假设 $V \subseteq \mathbb{R}^n$ 是线性子空间.

根据定理 (5.12), 任意的 $\alpha \in \mathbb{R}^n$ 可以唯一地写成 $\alpha = \boldsymbol{v} + \boldsymbol{w}$, 其中 $\boldsymbol{v} \in V, \boldsymbol{w} \in V^{\perp}$, 因此我们可以定义映射

$$p : \mathbb{R}^n \longrightarrow \mathbb{R}^n \qquad \alpha \longmapsto \boldsymbol{v}$$

定义 5.17. 我们把上面这个分解中的 $\boldsymbol{v}$ 称作 α 在子空间 V 上的正交投影, p 称为是关于子空间 V 的投影映射.

命题 5.18. p 是个线性映射, 且 $p^2 = p$, 其中 $p^2 = p \circ p$ 是复合映射.

怎么理解 $p^2 = p$ 呢? 这很简单, 因为 $p(\alpha) \in V$, 而对任意的 $\boldsymbol{v} \in V$, $p(\boldsymbol{v}) = \boldsymbol{v}$. 用映射的语言来写, 就是

$$p\Big|_V = \mathrm{id}\Big|_V.$$

设 $\alpha = \boldsymbol{v}_1 + \boldsymbol{w}_1, \beta = \boldsymbol{v}_2 + \boldsymbol{w}_2, a, b \in \mathbb{R}$, 其中 $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V, \boldsymbol{w}_1, \boldsymbol{w}_2 \in V^{\perp}$ 那么 $a\alpha + b\beta = (a\boldsymbol{v}_1 + b\boldsymbol{v}_2) + (a\boldsymbol{w}_1 + b\boldsymbol{w}_2)$, 由正交分解的唯一性, 我们就有 $a\boldsymbol{v}_1 + b\boldsymbol{v}_2$ 是 $a\alpha + b\beta$ 在子空间 V 上的正交投影, 所以

$$p(a\alpha + b\beta) = a\boldsymbol{v}_1 + b\boldsymbol{v}_2 = ap(\alpha) + bp(\beta).$$

这就证明了 p 是个线性映射.

根据例题 (4.21), 存在矩阵 $P \in M_{n \times n}(\mathbb{R})$, 使得对任意 $\alpha \in \mathbb{R}^n$, 我们有 $p(\alpha) = P\alpha$. 我们的目标就是把 P 求出来. 怎么求呢? 其实还是沿袭之前的思路. 我们假设 $r = \dim V$, 那么我们取 V 的一组基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$, 用待定系数法:

$$\alpha = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_r\boldsymbol{v}_r + \boldsymbol{w},$$

其中 $\boldsymbol{w} \in V^{\perp}$, 则 $p(\alpha) = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_r\boldsymbol{v}_r$. 我们需要求解其中的系数 $x_1, x_2, \cdots, x_r$.

等式两边分别用 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$ 作内积, 并且注意到 $\langle \boldsymbol{v}_1, \boldsymbol{w}\rangle = \langle \boldsymbol{v}_2, \boldsymbol{w}\rangle = \cdots = \langle \boldsymbol{v}_r, \boldsymbol{w}\rangle = 0$, 我们得到方程组

$$\begin{cases} \langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle x_1 + \langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle x_2 + \cdots + \langle \boldsymbol{v}_1, \boldsymbol{v}_r\rangle x_r &= \langle \boldsymbol{v}_1, \alpha\rangle \\ \langle \boldsymbol{v}_2, \boldsymbol{v}_1\rangle x_1 + \langle \boldsymbol{v}_2, \boldsymbol{v}_2\rangle x_2 + \cdots + \langle \boldsymbol{v}_2, \boldsymbol{v}_r\rangle x_r &= \langle \boldsymbol{v}_2, \alpha\rangle \\ \vdots \\ \langle \boldsymbol{v}_r, \boldsymbol{v}_1\rangle x_1 + \langle \boldsymbol{v}_r, \boldsymbol{v}_2\rangle x_2 + \cdots + \langle \boldsymbol{v}_r, \boldsymbol{v}_r\rangle x_r &= \langle \boldsymbol{v}_r, \alpha\rangle \end{cases}$$

注意到 $\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle = \boldsymbol{v}_i^T\boldsymbol{v}_j$, 那么用矩阵的方式方程组可以改写为

$$\begin{bmatrix} \boldsymbol{v}_1^T\boldsymbol{v}_2 & \boldsymbol{v}_1^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_1^T\boldsymbol{v}_r \\ \boldsymbol{v}_2^T\boldsymbol{v}_1 & \boldsymbol{v}_2^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_2^T\boldsymbol{v}_r \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v}_r^T\boldsymbol{v}_1 & \boldsymbol{v}_r^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r^T\boldsymbol{v}_r \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^T\alpha \\ \boldsymbol{v}_2^T\alpha \\ \vdots \\ \boldsymbol{v}_r^T\alpha \end{bmatrix}$$

根据分块矩阵的乘法, 我们有

$$\begin{bmatrix} \boldsymbol{v}_1^T\boldsymbol{v}_2 & \boldsymbol{v}_1^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_1^T\boldsymbol{v}_r \\ \boldsymbol{v}_2^T\boldsymbol{v}_1 & \boldsymbol{v}_2^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_2^T\boldsymbol{v}_r \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v}_r^T\boldsymbol{v}_1 & \boldsymbol{v}_r^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r^T\boldsymbol{v}_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_r^T \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{v}_1^T\alpha \\ \boldsymbol{v}_2^T\alpha \\ \vdots \\ \boldsymbol{v}_r^T\alpha \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_r^T \end{bmatrix} \alpha.$$

我们做一个换元: 令 $A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix} \in M_{n\times r}(\mathbb{R})$, 它是列满秩矩阵, $A^T = \begin{bmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_r^T \end{bmatrix}$, 所以方程组可以改写为

$$(A^TA)\boldsymbol{x} = A^T\alpha.$$

注意到 $A^TA \in M_{r\times r}(\mathbb{R})$, 如果它是个可逆矩阵那就太完美了, 我们直接可以解出

$$\boldsymbol{x} = (A^TA)^{-1}A^T\alpha,$$

从而

$$p(\alpha) = x_1\boldsymbol{v}_1 + \cdots + x_r\boldsymbol{v}_r = A\boldsymbol{x} = A(A^TA)^{-1}A^T\alpha,$$

这就是说

定理 5.19. 设 $V \subseteq \mathbb{R}^n$ 是由 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$ 为基底的线性子空间, $A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix} \in M_{n \times r}(\mathbb{R})$, $p : \mathbb{R}^n \to \mathbb{R}^n$ 为关于子空间 V 的投影映射. 那么对任意的 $\alpha \in \mathbb{R}^n$, 我们有

$$p(\alpha) = P\alpha,$$

其中 $P = A(A^T A)^{-1} A^T \in M_{n \times n}$.

注释 5.20. 千万不要试图把 $(A^T A)^{-1}$ 拆成 $A^{-1}(A^T)^{-1}$, 因为 A 是个 $n \times r$ 矩阵, 不是方阵. 另一方面, 如果你真的坚持己见非要拆, 拆完了发现 $P = I_n$, 这就完蛋了, 因为 p 不是恒等映射.

注释 5.21.

$$\begin{aligned} P^2 &= (A(A^T A)^{-1} A^T)(A(A^T A)^{-1} A^T) \\ &= A(A^T A)^{-1} (A^T A)(A^T A)^{-1} A^T \\ &= A(A^T A)^{-1} A^T \\ &= P \end{aligned}.$$

这个和 $p^2 = p$ 看起来是一致的. 事实上, 这是更本质的一个事实的在这个情形下的体现: 当我们把线性映射用矩阵来表示的时候, 矩阵乘法就对应了线性映射的复合. 这一点在我们将来有机会具体研究线性映射的时候再谈.

注释 5.22. P 也是一个对称矩阵, 也就是说 $P^T = P$. 请作为练习自己证明.

那么这样完美的事情会不会发生呢? 当然会发生, 因为生活是美好的!

命题 5.23. 设 $A \in M_{n \times r}(\mathbb{R})$, 那么

$$A^T A \in M_{r \times r}(\mathbb{R}) \text{ 可逆, 当且仅当 } A \text{ 是列满秩矩阵.}$$

证明. 我们用等价条件转化一下这个命题, 那么这个命题等价于下面这个论断:

$$\operatorname{rank}(A^T A) = r \Leftrightarrow \operatorname{rank}(A) = r.$$

这个证明很简单, 只需要注意到 $N(A) = N(A^T A)$. 推导如下:

$$\begin{aligned} \operatorname{rank}(A^T A) = r \quad &\Leftrightarrow N(A^T A) = 0 \\ &\Leftrightarrow N(A) = 0 \\ &\Leftrightarrow \operatorname{rank}(A) = r. \end{aligned}$$

□

我们来验证一下 $\dim V = 1$ 的特殊情形, 看看和前面说的是不是一致. 假设 $V = \mathbb{R} \cdot \boldsymbol{v}$, 那么矩阵 A 就等于 $\boldsymbol{v}$, 那么 $P = \boldsymbol{v}(\boldsymbol{v}^T \boldsymbol{v})^{-1} \boldsymbol{v}^T$, 所以

$$p(\alpha) = P\alpha = \boldsymbol{v}(\boldsymbol{v}^T \boldsymbol{v})^{-1} \boldsymbol{v}^T \alpha = \boldsymbol{v} \langle \boldsymbol{v}, \boldsymbol{v} \rangle^{-1} \langle \boldsymbol{v}, \alpha \rangle = \frac{\langle \alpha, \boldsymbol{v} \rangle}{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \cdot \boldsymbol{v}.$$

跟之前的结论一致.

我们再来看一个例子:

例题 5.24. 考虑 $\mathbb{R}^3$, V 由 $\begin{bmatrix}1\\2\\-1\end{bmatrix}$ 和 $\begin{bmatrix}1\\1\\1\end{bmatrix}$ 张成, 所以 $\dim V=2$, 那么 $\dim V^{\perp}=1$, 简单计算可知 $V^{\perp}=\mathbb{R}\cdot\begin{bmatrix}-3\\2\\1\end{bmatrix}$.

那么对任意的 $\alpha=\begin{bmatrix}x\\y\\z\end{bmatrix}$, 它在 $V^{\perp}$ 上的正交投影就是 $\dfrac{-3x+2y+z}{9+4+1}\begin{bmatrix}-3\\2\\1\end{bmatrix}$, 所以它在 V 上的投影是

$$\begin{bmatrix}x\\y\\z\end{bmatrix}-\frac{-3x+2y+z}{9+4+1}\begin{bmatrix}-3\\2\\1\end{bmatrix}=\begin{bmatrix}\frac{5x+6y+3z}{14}\\\frac{6x+10y-2z}{14}\\\frac{3x-2y+13z}{14}\end{bmatrix}=\frac{1}{14}\begin{bmatrix}5&6&3\\6&10&-2\\3&-2&13\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$$

所以 $P=\dfrac{1}{14}\begin{bmatrix}5&6&3\\6&10&-2\\3&-2&13\end{bmatrix}$.

我们现在套公式, 看看和上面的计算结果是不是一样.

$$A=\begin{bmatrix}1&1\\2&1\\-1&1\end{bmatrix},A^TA=\begin{bmatrix}1&2&-1\\1&1&1\end{bmatrix}\begin{bmatrix}1&1\\2&1\\-1&1\end{bmatrix}=\begin{bmatrix}6&2\\2&3\end{bmatrix}$$

所以

$$(A^TA)^{-1}=\frac{1}{6\cdot 3-2\cdot 2}\begin{bmatrix}3&-2\\-2&6\end{bmatrix}$$

根据定理 (5.19) 中的公式, 我们有

$$\begin{aligned} P = A(A^TA)^{-1}A^T &= \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ -1 & 1 \end{bmatrix} \frac{1}{6\cdot 3 - 2\cdot 2} \begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \frac{1}{14} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & -5 \\ 4 & 2 & 8 \end{bmatrix} \\ &= \frac{1}{14} \begin{bmatrix} 5 & 6 & 3 \\ 6 & 10 & -2 \\ 3 & -2 & 13 \end{bmatrix} \end{aligned}$$

计算结果和上面完全一致.

注释 5.25. 我们还需要说明的一点是 $P = A(A^TA)^{-1}A$ 看起来和 A 的选取有关系, 或者说和 V 的一组基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$ 的选取有关系. 事实上, 根据正交投影的几何意义, 它应该只取决于 V, 而不取决于基的选取. 那么我们需要说明, 任取另一组基 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_r$, $B = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_r \end{bmatrix} \in M_{n\times r}$, 我们也有 $P = B(B^TB)^{-1}B^T$. 证明如下: 由于 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$ 是 V 的一组基, 那么 $\boldsymbol{w}_i$ 都可以由他们线性表出, 因此存在矩阵 $C \in M_{r\times r}(\mathbb{R})$, 使得 $B = AC$. 因为 $r = \operatorname{rank}(B) = \operatorname{rank}(AC) \le \operatorname{rank}(C) \le r$, 所以 $\operatorname{rank}(C) = r$, 因此 C 是 r 阶可逆矩阵. 于是

$$\begin{aligned} B(B^TB)^{-1}B^T &= AC((AC)^TAC)^{-1}(AC)^T \\ &= AC(C^T(A^TA)C)^{-1}C^TA^T \\ &= ACC^{-1}(A^TA)^{-1}(C^T)^{-1}C^TA^T \\ &= A(A^TA)^{-1}A^T = P \end{aligned}.$$

5.2.1 投影矩阵

设 $V \subseteq \mathbb{R}^n$ 是一个线性空间, 那么关于 V 的投影映射 p 可以用一个矩阵 $P \in M_{n\times n}(\mathbb{R})$ 来表示. 为了后面的描述不至于引起混淆, 我们加点下标: 用 p_V 来记关于 V 的投影映射, 用 P_V 来记表示 p_V 的矩阵. 我们知道 $P_V^2 = P_V$ 且 $P_V^T = P_V$.

现在反过来, 假设 $P \in M_{n\times n}(\mathbb{R})$ 满足 $P^2 = P$ 且 $P^T = P$, 那么, 它能不能表示关于某个子空间 V 的投影映射呢? 更明确地说, 映射 $\mathbb{R}^n \to \mathbb{R}^n, \boldsymbol{x} \mapsto P\boldsymbol{x}$ 是不是关于某个子空间 V 的投影映射呢? 答案是肯定的.

命题 5.26. 设 $P \in M_{n\times n}(\mathbb{R})$, 那么

$$\text{存在子空间 } V \subseteq \mathbb{R}^n \text{ 使得 } P = P_V \quad \text{当且仅当 } P^2 = P.$$

我们思考一下, 如果 $P = P_V$, 那么 V 必须是什么呢? 首先, 如果 $P = P_V$, 根据投影映射的定义, 我们就知道 $C(P) \subseteq V$. 另一方面, P_V 限制在 V 上是恒等映射, 所以 $V \subseteq C(P)$; 所以 $V = C(P)$. 所以, 这个命题如果成立, 必要条件就是 $V = C(P)$. 剩下来的就简单了, 我们直接证明这一点.

证明. "$\Rightarrow$": 这个已经在注释 (5.21) 中说明了.

"$\Leftarrow$": 此时的条件是 $P^2 = P$ 和 $P^T = P$. 设 $V = C(P)$, 现在只剩下来证明 $P = P_V$. 要证明这一点, 就需要明白我们该怎么转化这个问题. 根据 P_V 的定义, 对任意 $\boldsymbol{x} \in \mathbb{R}^n$, 假设 $\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{w}$ (这样的写法是唯一的), 其中 $\boldsymbol{v} \in V, \boldsymbol{w} \in V^{\perp}$, 那么 $P_V\boldsymbol{x} = \boldsymbol{v}$. 所以我们就只需要证明 $P\boldsymbol{x}$ 也等于 $\boldsymbol{v}$.

$$\boldsymbol{x} = P\boldsymbol{x} + (\boldsymbol{x} - P\boldsymbol{x}),$$

这个分解式中, $P\boldsymbol{x} \in V$, 如果我们能证明 $\boldsymbol{x} - P\boldsymbol{x} \in V^{\perp}$ 就成功了. (先想一想, 再看后面的证明). $V^{\perp} = C(P)^{\perp} = N(P^T) = N(P)$, 所以只需要证明 $\boldsymbol{x} - P\boldsymbol{x} \in N(P)$, 这是很容易的: $P(\boldsymbol{x} - P\boldsymbol{x}) = P\boldsymbol{x} - P^2\boldsymbol{x} = P\boldsymbol{x} - P\boldsymbol{x} = 0$. □

5.3 最小二乘法

首先什么是最小 "二乘法"? 实际上, 这个 "二乘" 的意思, 就是平方, 它是个日语词汇, 我们很多科技词汇的翻译都是转译自日语; 后来有了自己的更好的翻译之后, 在特定的场合的习惯性说法就没有改变, 而是沿袭了习惯.

最小二乘法的来源大体上有殊途同归的两个: 一个是从数学理论上来, 一个从实际应用中来.

数学理论的来源依然是解线性方程组: $A\boldsymbol{x} = \boldsymbol{b}$. 我们如果单纯的把 "求解" 做个解方程的目标的话, 如果我们用高斯消元法判断出来无解, 那么我们已经可以结束我们的讨论了. 但是, 我们还可以有一些更高级的目标, 就是虽然我们找不到一个向量 $\boldsymbol{x}$ 使得 $A\boldsymbol{x} = \boldsymbol{b}$, 但是我对所有的向量 $\boldsymbol{x}$ 都把 $A\boldsymbol{x}$ 计算出来, 看看它和 $\boldsymbol{b}$ 差了多少, 我们把误差最小的那个向量拿出来. 这个所谓的误差就是 $\|A\boldsymbol{x} - \boldsymbol{b}\|$. 另一方面我们也可以看到, 误差等于零等价于方程组有解.

实际应用中的来源大体上是我们做了一些实验, 得到一些实验数据, 把它画在坐标系内: 例如

此处应有一个图, 我依然不会画. 可以想象一下, 就是坐标系内有一些点.

那么我们相用一个方程的图像去拟合, 比如说如果那些点的分布看起来差不多在一条直线上, 我们就用一条直线去拟合: 什么意思呢? 意思是画一条直线, 使得我们得到的数据点当横坐标与数据点的横坐标一致时, 纵坐标的差值的绝对值的和最小. (为什么不用距离求和最小呢? 显

然实际中我们考虑的应该是变量一致时, 理论值和实际值之间的差别.) 用数学语言来描述一下, 就是我们已知一些点 $A_i = (x_i, y_i)$, 要求一条直线 $\ell : y = kx + b$, 使得 $\sum_{i=1}^n |kx_i + b_i - y_i|$ 最小. 但是呢, 这个绝对值处理起来很不方便, 所以我们就变通一下, 改为求直线 $\ell : y = kx + b$ 使得 $\sum_{i=1}^n (kx_i + b - y_i)^2$ 最小. 我们做一些数学推导. 令

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} k \\ b \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

如果数据点不在一条直线上, 那么 $A\boldsymbol{x} = \boldsymbol{b}$ 就没有解; 此时, 我们就是要找一个 $\boldsymbol{x}$, 使得 $\|A\boldsymbol{x} - \boldsymbol{b}\|$ 最小.

所以, 即便问题来源于实际, 我们稍微抽象地归纳一下, 还是回到了原始的数学问题: 设 $A \in M_{m\times n}(\mathbb{R})$ 且方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 没有解, 求 $\boldsymbol{x} \in \mathbb{R}^n$, 使得 $\|A\boldsymbol{x} - \boldsymbol{b}\|$ 使得它取到最小值.

我们继续分析这个问题. 当 $\boldsymbol{x}$ 取遍所有 $\mathbb{R}^n$ 中的元素. 那么 $A\boldsymbol{x}$ 取遍了 $V = C(A) \subseteq \mathbb{R}^m$ 中的元素. 方程组无解等价于 $\boldsymbol{b} \notin V$. 因此原问题转化为在 V 上找一个向量 $\boldsymbol{v}$, 使得 $\|\boldsymbol{v} - \boldsymbol{b}\|$ 取到最小值. 那么答案就显而易见了: $\boldsymbol{v}$ 应取 $\boldsymbol{b}$ 在 V 上的投影! 此时 $\boldsymbol{v} - \boldsymbol{b} \in V^\perp$, 由于 $V = C(A)$, 我们有 $V^\perp = N(A^T)$, 所以 $A^T(\boldsymbol{v} - \boldsymbol{b}) = 0$. 将 $\boldsymbol{v} = A\boldsymbol{x}$ 代入, 我们就知道我们要找的 $\boldsymbol{x}$ 是方程组

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}$$

的解.

反之, 假设 $\boldsymbol{x}$ 满足 $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$, 那么 $A^T(\boldsymbol{b} - A\boldsymbol{x}) = 0$, 即 $\boldsymbol{b} - A\boldsymbol{x} \in N(A^T) = C(A)^\perp$. 由因为 $A\boldsymbol{x} \in C(A)$, 再注意到: $\boldsymbol{b} = (\boldsymbol{b} - A\boldsymbol{x}) + A\boldsymbol{x}$, 那么这个式子就是 $\boldsymbol{b}$ 关于 $C(A)$ 中的正交分解. 从而 $\|\boldsymbol{b} - A\boldsymbol{x}\|$ 取到最小值.

总而言之:

定理 5.27. 若 $A\boldsymbol{x} = \boldsymbol{b}$ 没有解, 那么 $\|A\boldsymbol{x} - \boldsymbol{b}\|$ 取到最小值当且仅当 $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$.

注释 5.28. 当 A^TA 不可逆 (或者等价地 A 不列满秩) 的时候, $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ 的解不唯一. 但是, 如果 $\boldsymbol{x}_1, \boldsymbol{x}_2$ 是上面这个方程的两个解, 那么 $A\boldsymbol{x}_1 = A\boldsymbol{x}_2$. 我们可以用正交投影的唯一性来说明这一点. 我们也可以理论上证明这件事: 因为 $A^TA(\boldsymbol{x}_1 - \boldsymbol{x}_2) = A^T\boldsymbol{b} - A^T\boldsymbol{b} = 0$, 所以 $\boldsymbol{x}_1 - \boldsymbol{x}_2 \in N(A^TA)$. 由于 $N(A^TA) = N(A)$, 那么 $A(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 0$, 即 $A\boldsymbol{x}_1 = A\boldsymbol{x}_2$.

5.4 正交基–Gram Schmidt 正交化

我们回顾一下我们的问题: $V \subseteq \mathbb{R}^n$ 是一个线性子空间, 我们要研究关于它的投影映射

$$p : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

它满足下述性质:

1. $\forall \boldsymbol{v} \in \mathbb{R}^n, p(\boldsymbol{v}) \in V$.
2. $\forall \boldsymbol{v} \in V, p(\boldsymbol{v}) = \boldsymbol{v}$; $\forall \boldsymbol{v} \in \mathbb{R}^n,\ \boldsymbol{v} - p(\boldsymbol{v}) \in V^{\perp}$.
3. $\forall \boldsymbol{v} \in \mathbb{R}^n, p(p(\boldsymbol{v})) = p(\boldsymbol{v})$.

我们可以用一个矩阵来计算: 任取 V 的一组基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$, 令矩阵 $A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix}$, 那么对任意 $\boldsymbol{v} \in \mathbb{R}^n$, $p(\boldsymbol{v}) = A(A^TA)^{-1}A^T\boldsymbol{v}$.

在上面这个矩阵表达式和我们选取的基无关, 所以我们在做计算的时候, 可以选取合适的基, 以尽量简化我们的运算. 什么才叫"合适的基", 怎样才叫简化运算呢? 我们还是观察这个矩阵表达式, 最复杂的部分在于求 $(A^TA)^{-1}$. 所以"合适"的标准就是这个矩阵求逆矩阵要尽量简单, 所以 A^TA 就要尽量简单. 最简单的是单位阵 I_r, 稍微复杂一点点的是对角矩阵, 再复杂一点点的就是上三角矩阵或者下三角矩阵.

我们首先否定上三角矩阵或者下三角矩阵: 这是因为 A^TA 是对称矩阵, 所以如果 A^TA 是上 (下) 三角矩阵, 那么它的对角线下 (上) 方等于 0 ; 由对称性, 它的对角线下 (上) 方也等于零, 所以只能是对角矩阵. 因此我们直接考虑能否取到基, 使得 A^TA 是单位阵或者对角矩阵.

分析完了问题, 我们现在必须要做一些必要的计算了. 依然是设 $A = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix}$, 那么 $A^T = \begin{bmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_r^T \end{bmatrix}$, 所以

$$A^TA = \begin{bmatrix} \boldsymbol{v}_1^T\boldsymbol{v}_1 & \boldsymbol{v}_1^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_1^T\boldsymbol{v}_r \\ \boldsymbol{v}_2^T\boldsymbol{v}_1 & \boldsymbol{v}_2^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_2^T\boldsymbol{v}_r \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v}_r^T\boldsymbol{v}_1 & \boldsymbol{v}_r^T\boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r^T\boldsymbol{v}_r \end{bmatrix} = \begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle & \cdots & \langle \boldsymbol{v}_1, \boldsymbol{v}_r \rangle \\ \langle \boldsymbol{v}_2, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{v}_2 \rangle & \cdots & \langle \boldsymbol{v}_2, \boldsymbol{v}_r \rangle \\ \vdots & \vdots & & \vdots \\ \langle \boldsymbol{v}_r, \boldsymbol{v}_1 \rangle & \langle \boldsymbol{v}_r, \boldsymbol{v}_2 \rangle & \cdots & \langle \boldsymbol{v}_r, \boldsymbol{v}_r \rangle \end{bmatrix}$$

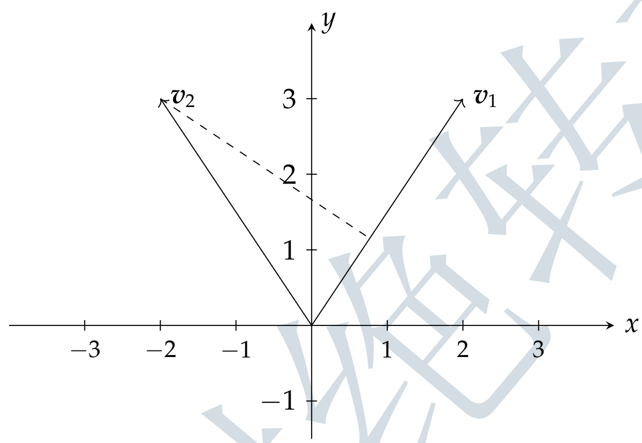
那么 A^TA 是对角矩阵当且仅当 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$ 两两互相正交.

定义 5.29. 设 $V \subseteq \mathbb{R}^n$ 是一个线性子空间, 它的一组基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$ 称为是正交基, 如果 $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0, \forall i \neq j$. 我们称一组正交基为标准正交基, 如果 $\langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle = 1, \forall i$.

注释 5.30. 从一组正交基很容易我们就可以获得一组标准正交基. 设 $v_i' = \dfrac{v_i}{\|v_i\|}$ 即可.

但是一般来说, 我们并不是直接就能取到一组基使得其中的向量两两互相正交, 那么问题来了: 我们是否能从任意一组基出发, 构造出来一组基使得他们两两互相正交. 我们看一个二维的例子.

易知 $v_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, v_2 = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ 是 $\mathbb{R}^2$ 的一组基, 但是它们不正交: $\langle v_1, v_2 \rangle = 5$.



设 p 是关于 $\mathbb{R} \cdot v_1$ 的正交投影, 令 $v_2' = v_2 - p(v_2)$, 那么 $\langle v_1, v_2' \rangle = 0$, 且 v_1, v_2' 是一组正交基.

现在考虑一般的情形. 假设 $V \subseteq \mathbb{R}^n$ 是一个 r-维的线性子空间, 且 $v_1, v_2, \cdots, v_r$ 是它的一组基. 令 $v_1' = v_1, V_1 = \mathrm{Span}(v_1)$, 利用上面二维的例子, 我们也很容易就能知道令 $v_2' = v_2 - p_1(v_1)$ (这里 p_1 是关于子空间 V_1 的投影), 那么 $v_2' \perp v_1'$, 且 $\mathrm{Span}(v_1', v_2') = \mathrm{Span}(v_1, v_2) =: V_2$. 这个时候, 我们考虑关于子空间 V_2 的投影 p_2, 我们可以用矩阵 $A_2 = \begin{bmatrix} v_1' & v_2' \end{bmatrix}$ 来计算 p_2:

$$P_2 = \begin{bmatrix} v_1' & v_2' \end{bmatrix} \begin{bmatrix} \langle v_1', v_1' \rangle & 0 \\ 0 & \langle v_2', v_2' \rangle \end{bmatrix}^{-1} \begin{bmatrix} v_1'^T \\ v_2'^T \end{bmatrix} = \frac{v_1' v_1'^T}{\langle v_1', v_1' \rangle} + \frac{v_2' v_2'^T}{\langle v_2', v_2' \rangle}.$$

所以

$$p_2(v_3) = P_2 v_3 = \frac{v_1' v_1'^T v_3}{\langle v_1', v_1' \rangle} + \frac{v_2' v_2'^T v_3}{\langle v_2', v_2' \rangle} = \frac{\langle v_1', v_3 \rangle}{\langle v_1', v_1' \rangle} \cdot v_1' + \frac{\langle v_2', v_3 \rangle}{\langle v_2', v_2' \rangle} \cdot v_2'$$

我们令

$$v_3' = v_3 - p_2(v_3) = v_3 - \frac{\langle v_1', v_3 \rangle}{\langle v_1', v_1' \rangle} \cdot v_1' - \frac{\langle v_2', v_3 \rangle}{\langle v_2', v_2' \rangle} \cdot v_2',$$

这个时候, 我们就有 v_1', v_2', v_3' 两两互相正交且 $\mathrm{Span}(v_1, v_2, v_3) = \mathrm{Span}(v_1', v_2', v_3') =: V_3$.

利用同样的讨论, 假设我们已经定义出来了 $v_1', v_2', \cdots, v_{k-1}'$, 我们可以归纳地定义

$$v_k' := v_k - p_{k-1}(v_k) = v_k - \sum_{i=1}^{k-2} \frac{\langle v_i', v_k \rangle}{\langle v_i', v_i' \rangle} \cdot v_i',$$

满足

- $\boldsymbol{v}_1', \boldsymbol{v}_2', \cdots, \boldsymbol{v}_k'$ 两两正交;
- $\mathrm{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k) = \mathrm{Span}(\boldsymbol{v}_1', \boldsymbol{v}_2', \cdots, \boldsymbol{v}_k') =: V_k$.

这个过程可以一直重复, 直至我们构造完所有的 $\boldsymbol{v}_1', \boldsymbol{v}_2', \cdots, \boldsymbol{v}_r'$. 这一组基满足下面的性质:

1. $\boldsymbol{v}_1', \boldsymbol{v}_2', \cdots, \boldsymbol{v}_r'$ 两两正交;
2. 对任意的 $1 \leq k \leq r$, $\boldsymbol{v}_k' = \boldsymbol{v}_k - \sum_{i=1}^{k-2} \dfrac{\langle \boldsymbol{v}_i', \boldsymbol{v}_k \rangle}{\langle \boldsymbol{v}_i', \boldsymbol{v}_i' \rangle} \cdot \boldsymbol{v}_i'$.
3. 对任意的 $1 \leq k \leq r$, $\mathrm{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k) = \mathrm{Span}(\boldsymbol{v}_1', \boldsymbol{v}_2', \cdots, \boldsymbol{v}_k')$.

这个过程被称为 Gram-Schmidt **正交化**: 即从一组线性无关向量出发, 得到一组相互正交的向量, 满足上述要求. 接着我们把这些 $\boldsymbol{v}_i'$ 标准化: 令 $\boldsymbol{u}_i = \dfrac{\boldsymbol{v}_i'}{\|\boldsymbol{v}_i'\|}$, 我们就得到:

$$\text{对任意 } 1 \leq k \leq r,\ \boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k \text{ 是 } \mathrm{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k) \text{ 的标准正交基.}$$

现在假设 $A \in M_{n\times n}(\mathbb{R})$ 是可逆矩阵, 设 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 是它的列向量. A 可逆推知它的列向量线性无关. 通过 Gram-Schmidt 正交化, 我们得到 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$. 因为对任意 $1 \leq k \leq n$, $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k$ 是 $\mathrm{Span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k)$ 的标准正交基, 所以 $\boldsymbol{v}_k$ 是 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k$ 的线性组合, 即:

$$\text{存在 } c_{1k}, c_{2k}, \cdots, c_{kk}, \text{ 使得 } \boldsymbol{v}_k = c_{1k}\boldsymbol{u}_1 + c_{2k}\boldsymbol{u}_2 + \cdots + c_{kk}\boldsymbol{u}_k,$$

因此

$$\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} & \cdots & c_{1n} \\ 0 & c_{22} & \cdots & c_{2k} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & c_{kk} & \cdots & c_{kn} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & c_{nn} \end{bmatrix}$$

令 $Q = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}, R = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} & \cdots & c_{1n} \\ 0 & c_{22} & \cdots & c_{2k} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & c_{kk} & \cdots & c_{kn} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & c_{nn} \end{bmatrix}$ 上面这个式子就是

$$A = QR.$$

这其中:

- $Q^TQ = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \cdots \\ \boldsymbol{u}_n^T \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix} = \left(\boldsymbol{u}_i^T\boldsymbol{u}_j\right)_{1\le i,n\le n} = I_n$, 所以 Q 是正交矩阵.
- R 是上三角矩阵.

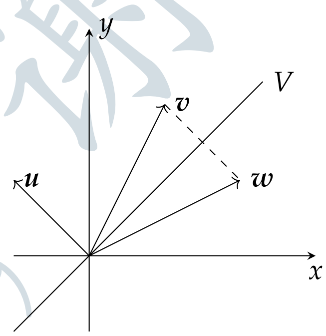
总的来说, 我们就得到了:

定理 5.31 (QR 分解). 任何可逆矩阵 $A \in M_{n\times n}(\mathbb{R})$ 都可以写成一个正交矩阵乘以一个上三角矩阵.

例题 5.32. $\mathbb{R}^2$ 上的旋转的矩阵是正交矩阵, 旋转角度 θ 的矩阵为 $Q(\theta) := \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$. 因此任何一个 2×2 矩阵都可以写成 $Q(\theta)R$ 的形式. 如果不用 Gram-Schmidt 正交化, 你能证明这一点吗?

例题 5.33. 置换矩阵是正交矩阵.

例题 5.34. 设 $V \subseteq \mathbb{R}^n$ 是 $n-1$ 维线性子空间, 关于 V 的反射是正交矩阵. 现在我们把这句话严格化, 什么叫: 关于 V 的反射. 我们看一个 2 维的示意图:



其中 $\boldsymbol{u}$ 是 V 的一个法向量; $\boldsymbol{w}$ 是 $\boldsymbol{v}$ 关于 V 的反射. 它满足这样的条件:

- $\boldsymbol{w}$ 在 V 上的投影和 $\boldsymbol{v}$ 在 V 上的投影一致;
- $\boldsymbol{w}$ 和法向 $\mathbb{R}\cdot\boldsymbol{u}$ 上的投影和 $\boldsymbol{v}$ 在法向的投影的 -1 倍.

$\boldsymbol{v}$ 在法向的投影为 $\dfrac{\langle\boldsymbol{u},\boldsymbol{v}\rangle}{\langle\boldsymbol{u},\boldsymbol{u}\rangle}\boldsymbol{u}$. 所以 $\boldsymbol{w} = \boldsymbol{v} - 2\dfrac{\langle\boldsymbol{u},\boldsymbol{v}\rangle}{\langle\boldsymbol{u},\boldsymbol{u}\rangle}\boldsymbol{u}$. 进一步地, 我们可以要求 $\boldsymbol{u}$ 是**单位法向量**, 所以

$$\boldsymbol{w} = \boldsymbol{v} - 2\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{v}.$$

我们令 $M_u = I_n - 2\boldsymbol{u}\boldsymbol{u}^T$, 那么 $\boldsymbol{w} = M_u\boldsymbol{v}$, 反射所对应的矩阵就是 M_u. 显然 M_u 是对称矩阵, 但是更重要的是:

命题 5.35. M_u 是正交矩阵.

证明.

$$M_u^T M_u = I_n - I_n(2\boldsymbol{u}\boldsymbol{u}^T) - 2\boldsymbol{u}\boldsymbol{u}^T + 4\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{u}\boldsymbol{u}^T = I_n.$$

最后一个等号是因为 $\boldsymbol{u}$ 是单位向量, 从而 $\boldsymbol{u}^T\boldsymbol{u} = 1$. □

我们有一个很深刻的线性代数定理:

定理 5.36 (Cardan-Dieudonne 定理). 任何 n 阶的正交矩阵都可以写成不超过 n 个反射的乘积.

第六章　行列式

6.1　引入

行列式, 英文直译应该称作 "判别式", 所谓 "判别式", 决定啥玩意儿呢? 它决定了未知数个数和方程个数一样多时, 方程组有没有唯一解. 注意, 没有唯一解包含两种情况, 一是没有解, 而是有解但是解不唯一. 一言以蔽之, "方程组有唯一解, 当且仅当 '判别式' 非零".

要说明这点, 我们来看一下高斯消元法, 就可以得到下面的论断:

$$A \in M_n(\mathbb{R}), \boldsymbol{b} \in \mathbb{R}^n, \text{ 那么 } A\boldsymbol{x} = \boldsymbol{b} \text{ 有唯一解当且仅当 } A \text{ 可逆.}$$

因此, "判别式" 首先决定的是方阵 A 是不是可逆. 如何决定 A 是否可逆, 那我们还是看高斯消元法. 根据高斯消元法, 对任意方阵 A, 存在一些交换两行的矩阵 $P_i(1 \le i \le n)$ 和把某一行的若干倍加到另一行的矩阵 E_j 使得他们以某种顺序左乘到 A 上去之后, 得到一个上三角矩阵 U, 设 U 的对角线元素为 $u_1, u_2, \cdots, u_n$, 那么我们就有

$$A \text{ 是可逆矩阵, 当且仅当 } u_1, u_2, \cdots, u_n \text{ 都非零, 当且仅当 } u_1 u_2 \cdots u_n \neq 0.$$

因此, 我们第一个直觉就是把 "判别式" 定义为 U 的对角线元素的乘积: $u_1 u_2 \cdots u_n$. 但是很遗憾, 由于高斯消元法的步骤不唯一, 因此 U 不唯一, 那么按照这样的方式, 我们并得不到一个明确的数.

例题 6.1. 对于对角矩阵 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, 一方面我们什么都不干, 它本身就是 U, 所以我们得到它的 "判别式" 为 1. 另一方面, 我们还可以对它做下面的初等行变换:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix},$$

所以我们得到它的 "判别式" 为 -1.

但是对于低阶矩阵做高斯消元法, 重复大量 (虽然在这里我只做了一个, 但是你们可以多做几个) 的实例, 我们就会发现, 我们期望中的得到的 "判别式"—-$u_1u_2\cdots u_n$—-几乎是良定义的: 它们的绝对值是相等的, 正负号只取决于我们做了多少次 "交换两行", 而且是有规律的: 多交换一次两行, "判别式" 就换一次正负号.

这个时候, 我们就可以总结一下规律了:

- E (某一行的倍数加到另一行) 不改变判别式;
- 对角线为 1 的上三角矩阵的判别式等于对角线元素的乘积;
- 若 P 为交换两行, 那么 PA 的判别式等于 -1 乘以 A 的判别式.

基于这里的第一条, 对于对角线非零的上三角矩阵, 我们可以把它用一些 E 变成对角矩阵; 因此第二条可以改成 "对角矩阵的判别式等于对角线元素的乘积". 我们用稍微归纳一下, 就是

- $\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j + c\alpha_i \\ \vdots \end{bmatrix}$ 的判别式等于 $\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \end{bmatrix}$ 的判别式;
- $\operatorname{diag}(u_1, u_2, \cdots, u_n)$ 的判别式等于 $u_1u_2\cdots u_n$;
- $\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j + c\alpha_i \\ \vdots \end{bmatrix}$ 的判别式等于 $\begin{bmatrix} \vdots \\ \alpha_j \\ \vdots \\ \alpha_i \\ \vdots \end{bmatrix}$ 的判别式的 -1 倍;

这样的描述有两个缺点: 一是太白话, 还不够数学化; 二是第二条和其他两条有点格格不入: 显得不那么 "线性代数": 原因是第一三条是基于行变换的, 而第二条并非. 如何代替第二条呢? 我们看一下除了第一二条的行变换, 还有一种行变换, 那就是某一行乘以若干倍. 例如, 将对角阵的 $\operatorname{diag}(u_1, u_2, \cdots, u_n)$ 的第 i 行乘以 c 倍还是对角阵, 而后者的判别式显然是前者的 c 倍.

基于这些探究, 我们可以数学化地总结一下我们的问题了:

问题 6.2. *是否存在函数*

$$f : M_n(\mathbb{R}) \longrightarrow \mathbb{R}$$

满足下面的条件: 对任意 $\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in M_{n\times n}(\mathbb{R})$,

1. $f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \end{bmatrix}\right) = f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j + c\alpha_i \\ \vdots \end{bmatrix}\right)$;

2. $f\left(\begin{bmatrix} \vdots \\ c\alpha_i \\ \vdots \end{bmatrix}\right) = cf\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$, 且 $f(I_n) = 1$.

3. $f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \end{bmatrix}\right) = -f\left(\begin{bmatrix} \vdots \\ \alpha_j \\ \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$.

我们还有一个几何上的来源. 设 $A \in M_n(\mathbb{R})$; 那么考虑 A 的列向量在 $\mathbb{R}^n$ 中形成的平行多维体 $X(A)$. 即:

$$X(A) = \left\{ t_1\alpha_1 + t_2\alpha_2 + \cdots + t_n\alpha_n \middle| 0 \leq t_i \leq 1, i = 1, 2, \cdots, n \right\}.$$

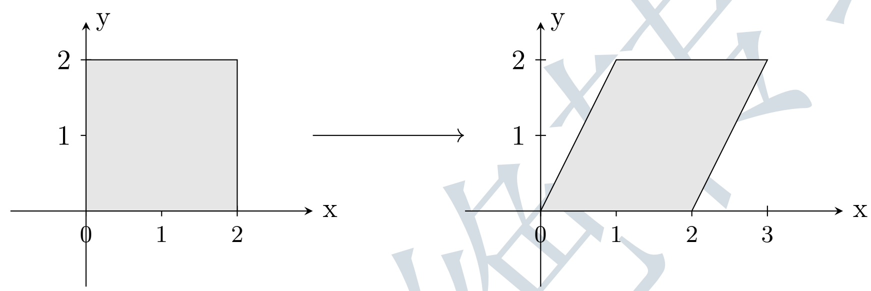
例如, I_2 在 $\mathbb{R}^2$ 中形成的平行二维体, 即平行四边形, 为第一象限的单位正方形; I_3 在 $\mathbb{R}^3$ 中形成的平行三维体为第一象限的单位立方体. 而 $\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ 在 $\mathbb{R}^2$ 中形成的 “平行四边形” 为从 $(0,0)$ 点到 $(2,2)$ 点的线段. 我们可以断言: A 是个满秩方阵, 当且仅当它在 $\mathbb{R}^n$ 中形成的平行多维体有非零的体积.

那么这个体积要满足什么性质呢?

- “第一象限” 的单位立方体的体积应该是 1, 这是什么意思呢? 这相当于说我们给定了一个 “单位” 或者物理学上的一个测量的 “基准”.

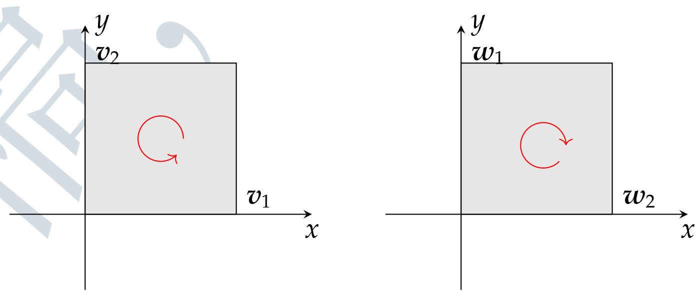
- 一个平行多维体的其中一个方向的边长变成原来的 c 倍, 其他方向的边长不变, 那么体积应该变成原来的 $|c|$ 倍;
- 一个平行多维体的 $n-1$ 个方向的边保持不动, 但是第 n 个方向的边沿着前 $n-1$ 个方向中的某个方向平移, 则体积不变.

例题 6.3. *第三条有点抽象, 我们理解一下: 前 $n-1$ 条边不动, 意思是平行多维体的“底”保持不动; 第 n 个方向的边沿着前 $n-1$ 个方向中的某个方向平移, 此时该平行多维体的“高”也保持不动. 我们画一个二维的示意图:*



这是 x 轴方向不变, y 轴方向沿 x 轴方向平移了 1. 左边这个正方形表面积和右边这个平行四边形面积有相同的“底”和“高”, 所以面积相等.

这里的平行多维体的体积要求和“判别式”的要求几乎一致, 除了体积是非负的, 而“判别式”未必非负. 我们再来观察一下体积. 我们看一下矩阵 $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2]$ 和 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = [\boldsymbol{w}_1 \quad \boldsymbol{w}_2]$ 给出的平行四边形是同一个, 如图



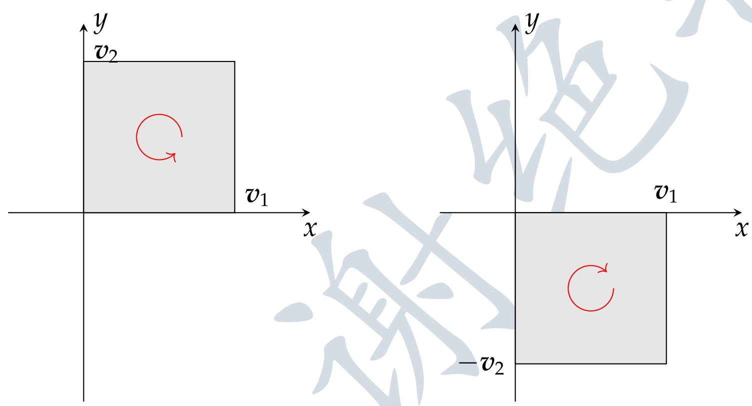
从 $\boldsymbol{v}_1$ 到 $\boldsymbol{v}_2$ 的方向是逆时针, 从 $\boldsymbol{w}_1$ 到 $\boldsymbol{w}_2$ 的方向是顺时针. 当我们以单位阵为基准, 那么 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 的两个列向量的顺序和单位阵的两个列向量的顺序不一致, 所以一个合理的做法是

把 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 给出的平行多面体的体积定义为单位阵 I_2 给出的平行多维体的体积的 -1 倍. 对于三维的体积, 当我们以单位阵 I_3 的体积为基准的时候, 我们会发现, 一个三阶方阵如果它的三个列向量按顺序满足右手定则, 那么它的体积是非负的; 如果不满足右手定则, 则是非正的. 当我们把体积的 “正负” 也考虑进去之后, 我们就发现, 它实际上就是要求问题 (6.2) 中的 (4). 我们可以把这样的体积称作 “**有向体积**”.

如果我们考虑有向体积, 那么体积要求中的第二条就应该改为

- 一个平行多维体的其中一个方向的边长变成原来的 c 倍, 其他方向的边长不变, 那么体积应该变成原来的 c 倍;

为什么要这样呢? 我们看一个例子: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 和 $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, 后者把 v_2 变成了 $-v_2$, 如图:



逆时针变成顺时针了, 因此面积应该是负的. 三维的情形类似, 将一个满足右手系的三个列向量其中两个不变, 另一个变成原来的负数倍, 那么新的三个向量就是左手系, 因此体积应该是负的.

问题 6.4. *有没有一个良定义的 “有向体积” 呢?*

这两个问题 (6.2和 6.4) 最终都指向了 “行列式”!

6.2 定义和性质

开门见山, 我们直接回答问题 (6.2)

定理 6.5. *存在唯一的函数*

$$f : M_n(\mathbb{R}) \longrightarrow \mathbb{R}$$

满足下面的条件: 对任意 $\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in M_{n\times n}(\mathbb{R})$,

1. $f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \end{bmatrix}\right) = f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j + c\alpha_i \\ \vdots \end{bmatrix}\right)$;

2. $f\left(\begin{bmatrix} \vdots \\ c\alpha_i \\ \vdots \end{bmatrix}\right) = cf\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$, 且 $f(I_n) = 1$.

3. $f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \end{bmatrix}\right) = -f\left(\begin{bmatrix} \vdots \\ \alpha_j \\ \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$.

也就是说, 问题 (6.2) 的回答是肯定的.

定义 6.6. 我们把定理 (6.5) 中的函数 f 称作行列式函数, 记作 det. 对 $A \in M_{n\times n}(\mathbb{R})$, 我们把 $\det(A)$ 称作 A 的行列式, 也记作 $|A|$.

证明之前, 我们做一些准备工作:

引理 6.7. 设函数 $f : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ 满足定理 (6.5) 中的条件 3; 那么如果 A 有两行相同, 那么 $f(A)$ 等于 0.

证明. 假设 $\alpha_i = \alpha_j$, 则

$$f(A) = f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \end{bmatrix}\right) = -f\left(\begin{bmatrix} \vdots \\ \alpha_j \\ \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right) = f(A).$$

所以 $f(A)=0$. □

引理 6.8. 假设函数 $f: M_{n\times n}(\mathbb{R})\to\mathbb{R}$ 满足定理 (6.5) 中的条件 2. 如果 A 有一行等于零, 那么 $f(A)=0$;

证明. 设 $\alpha_i=0$, 则对任意 c, 都有 $c\alpha_i=0$. 根据条件 2, 有

$$f(A)=f\left(\begin{bmatrix}\vdots\\ c\alpha_i\\ \vdots\end{bmatrix}\right)=cf\left(\begin{bmatrix}\vdots\\ \alpha_i\\ \vdots\end{bmatrix}\right)=cf(A).$$

所以 $f(A)=0$ □

引理 6.9. 设函数 $f: M_{n\times n}(\mathbb{R})\to\mathbb{R}$ 满足定理 (6.5) 中的条件 1 和 2, 那么对任意的不可逆矩阵 $f(A)=0$.

证明. A 不可逆, 则存在不全为零的数 $k_1,k_2,\cdots,k_n$ 使得 $k_1\alpha_1+\cdots+k_n\alpha_n=0$. 不妨设 $k_1\neq 0$. 则

$$k_1 f(A)\overset{2}{=}f\left(\begin{bmatrix}k_1\alpha_1\\ \vdots\end{bmatrix}\right)\overset{1}{=}f\left(\begin{bmatrix}k_1\alpha_1+\cdots+k_n\alpha_n\\ \vdots\end{bmatrix}\right)=f\left(\begin{bmatrix}0\\ \vdots\end{bmatrix}\right)=0.$$

□

引理 6.10. 设函数 $f: M_{n\times n}(\mathbb{R})\to\mathbb{R}$ 满足定理 (6.5) 中的条件 1 和 2, 那么

$$f\left(\begin{bmatrix}\vdots\\ \alpha_i+\alpha_i'\\ \vdots\end{bmatrix}\right)=f\left(\begin{bmatrix}\vdots\\ \alpha_i\\ \vdots\end{bmatrix}\right)+f\left(\begin{bmatrix}\vdots\\ \alpha_i'\\ \vdots\end{bmatrix}\right).$$

证明.

- 若 $\dim\mathrm{Span}(\alpha_1,\cdots,\alpha_{i-1},\alpha_{i+1},\cdots,\alpha_n)<n-1$, 即 $\alpha_1,\cdots,\alpha_{i-1},\alpha_{i+1},\cdots,\alpha_n$ 线性相关, 那么等式两边全都等于 0, 所以论断成立.

- 若 $\dim\mathrm{Span}(\alpha_1,\cdots,\alpha_{i-1},\alpha_{i+1},\cdots,\alpha_n)=n-1$, 即 $\alpha_1,\cdots,\alpha_{i-1},\alpha_i,\cdots,\alpha_{n-1}$ 线性无关. 我们考虑两种子情形:

– $\alpha_i \in \mathrm{Span}(\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_n)$, 则 $f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right) = 0$, 且由 1,

$$f\left(\begin{bmatrix} \vdots \\ \alpha_i + \alpha_i' \\ \vdots \end{bmatrix}\right) = f\left(\begin{bmatrix} \vdots \\ \alpha_i' \\ \vdots \end{bmatrix}\right)$$

, 故论断成立.

– $\alpha_i \notin \mathrm{Span}(\alpha_1, \cdots, \alpha_{i-1}, \alpha_{i+1}, \cdots, \alpha_n)$, 则 $\alpha_1, \cdots, \alpha_n$ 线性无关, 从而 α_i' 可由它们线性表出. 假设 $\alpha_i' = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n$, 那么由 (1), $f\left(\begin{bmatrix} \vdots \\ \alpha_i + \alpha_i' \\ \vdots \end{bmatrix}\right) = (c_i + 1)f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$, $f\left(\begin{bmatrix} \vdots \\ \alpha_i' \\ \vdots \end{bmatrix}\right) = c_i f\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$. 故论断成立.

于是我们有

•
$$f\left(\begin{bmatrix} \vdots \\ a\alpha_i + b\beta_i \\ \vdots \end{bmatrix}\right) = af\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right) + bf\left(\begin{bmatrix} \vdots \\ \beta_i \\ \vdots \end{bmatrix}\right).$$

□

直接的推论就是

推论 6.11. 设函数 $f: M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ 满足定理 (6.5) 中的条件 1 和 2, 那么

$$f\left(\begin{bmatrix} \vdots \\ a\alpha_i + b\beta_i \\ \vdots \end{bmatrix}\right) = af\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right) + bf\left(\begin{bmatrix} \vdots \\ \beta_i \\ \vdots \end{bmatrix}\right).$$

做了这么多准备工作, 我们看看满足定理 (6.5) 的函数应该长什么样子. 不愿意看也可以跳过不看, 只是一些体力活儿, 暴力展开而已, 毫无技术含量.

设 $\varepsilon_i = \begin{bmatrix} 0 & \cdots & 1 & \cdots 0 \end{bmatrix}$ 是 n-维的行向量, 它的第 i 个分量是 i, 其余分量都是 0. 对 $A = (a_{ij}) = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in M_{n\times n}(\mathbb{R})$, 它的第 i 行等于 $a_{i1}\varepsilon_1 + a_{i2}\varepsilon_2 + \cdots + a_{in}\varepsilon_n$. 由条件 (1):

$$f(A) = \sum_{j_1=1}^{n} a_{1j_1} f\left(\begin{bmatrix} \varepsilon_{j_1} \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}\right)$$

对 α_2 展开, 利用 (2) 和 (1):

$$f(A) = \sum_{j_1=1}^{n} a_{1j_1} \sum_{j_2=1}^{n} f\left(\begin{bmatrix} \varepsilon_{j_1} \\ a_{2j_2}\varepsilon_{j_2} \\ \vdots \\ \alpha_n \end{bmatrix}\right) = \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} a_{1j_1}a_{2j_2} f\left(\begin{bmatrix} \varepsilon_{j_1} \\ \varepsilon_{j_2} \\ \vdots \\ \alpha_n \end{bmatrix}\right);$$

逐步展开, 就得到:

$$f(A) = \sum_{j_1,j_2,\cdots,j_n=1}^{n} a_{1j_1}a_{2j_2}\cdots a_{nj_n} f\left(P(j_1,j_2,\cdots,j_n)\right), \text{其中} P(j_1,j_2,\cdots,j_n) = \begin{bmatrix} \varepsilon_{j_1} \\ \varepsilon_{j_2} \\ \vdots \\ \varepsilon_{j_n} \end{bmatrix}.$$

在这个求和里面, 只有当 $j_1, j_2, \cdots, , j_n$ 互不相等时才有贡献. 因为如果有两项相等, 那么最后一项 $f\left(P(j_1,j_2,\cdots,j_n)\right) = 0$. 所以

$$f(A) = \sum_{\{j_1,j_2,\cdots,j_n\}=\{1,2,\cdots,n\}} a_{1j_1}a_{2j_2}\cdots a_{nj_n} f\left(P(j_1,j_2,\cdots,j_n)\right).$$

每一个矩阵 $P(j_1,j_2,\cdots,j_n)$ 都是置换矩阵, 因此经过若干个交换行的操作能转化成单位矩阵 I_n, 因此由条件 (3), 我们有 $f\left(P(j_1,j_2,\cdots,j_n)\right) = \pm f(I_n)$.

注意条件: $f(I_n) = 1$, 满足定理 (6.5) 条件的函数必须等于

$$f(A) := \sum_{\{j_1,j_2,\cdots,j_n\}=\{1,2,\cdots,n\}} a_{1j_1}a_{2j_2}\cdots a_{nj_n} f(P(j_1,j_2,\cdots,j_n)). \tag{6.1}$$

注释 6.12. 我们知道 $f(P(j_1,j_2,\cdots,j_n)) \in \{\pm f(I_n)\}$ 是个确定的数, 至于它们具体的等于 $f(I_n)$ 还是 $-f(I_n)$, 取决于我们需要经过奇数次交换 (此时等于 $-f(I_n)$) 还是偶数次交换 (此时等于

$f(I_n))$. 注意, 虽然把 $P(j_1,j_2,\cdots,j_n)$ 写成交换两行的矩阵的乘积有很多种方式, 但是交换的次数的奇偶性是确定的, 有兴趣的同学可以阅读下面引理 6.13 和推论 6.14, 但是我建议你放弃这个兴趣, 直接展示你对我的信任即可. 它和排列 $j_1j_2\cdots j_n$ 的"逆序数"奇偶性相同, 记作 $r(j_1j_2\cdots j_n)$, 证明可见任何一本数学系教材. 所以

$$f(A):=f(I_n)\sum_{\{j_1,j_2,\cdots,j_n\}=\{1,2,\cdots,n\}}(-1)^{r(j_1j_2\cdots j_n)}a_{1j_1}a_{2j_2}\cdots a_{nj_n}.$$

再由我们的正规化条件: $f(I_n)=1$, 这就得到了行列式的无比繁琐的看不懂的表达式了.

引理 6.13. 如果我们把单位矩阵 I_n 写成一些交换两行的矩阵的乘积, 那么这些矩阵一共有偶数个, 即: 若 $I_n=P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k}$, 那么 k 一定是偶数.

证明. 首先, $k\neq 1$, 这是因为 $k=1$ 是说 $I_n=P_{i_1j_1}$, 显然是不行的. 若 $k=2$, 那么命题就不用证了. 以下我们用归纳法证明: $k>3$, 且假设对任意 $r<k$, 我们将 I_n 写成 r 个对换的乘积时, r 是偶数; 我们要来证明若 $I_n=P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k}$, 那么 k 一定是偶数. 首先, 我们需要注意对任意 a,b, $P_{ab}=P_{ba}$, 且对任意 c,d

$$(*)\qquad P_{cd}P_{ab}=\begin{cases}P_{ab}P_{cd}=I_n & 若\{c,d\}=\{a,b\}\ 或者\{c,d\}\cap\{a,b\}=\emptyset\\ P_{ab}P_{bd} & 若 c=a\\ P_{ac}P_{bc} & 若 b=d\end{cases}$$

这就看出来, 我们交换的时候, 指标 a 是不增加的, 并且有可能减少. 那么对于表达式 $I_n=P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k}$, 我们考虑 i_1 这个指标, 它在 $i_2,j_2,\cdots,i_k,j_k$ 中也必须出现, 否则第 i_1 被挪了位置却不被换回来. 所以我们不妨设 $i_\ell=i_1$, 那么我们对 $P_{i_{\ell-1}j_{\ell-1}}P_{i_\ell j_\ell}$ 使用上面的交换关系 $(*)$, 这就把 i_1 出现的位置往前移了, 并且在这过程中, 对换的个数 k 没有变, 指标 i_1 的个数不变或者变少. 我们重复这个步骤, 直到把这个指标交换到 $P_{i_1j_1}$ 之后, 那么我们得到 I_n 的一个新的表达式 $I_n=P_{i_1j_1}P_{i_1j_2'}\cdots P_{i_k'j_k'}$, 且这个表达式中 i_1 的个数不变或者变少. 这时, 有两种情况:

- $j_1=j_2'$, 则 $I_n=P_{i_3'j_3'}\cdots P_{i_k'j_k'}$, 这是 $k-2$ 个对换的乘积, 那么根据归纳假设 $k-2$ 是偶数, 所以 k 是偶数.

- $j_1\neq j_2'$, 则 $I_n=P_{i_1j_2'}P_{j_1j_2'}\cdots P_{i_k'j_k'}$. 这是 k 个对换的乘积, 但是指标 i_1 出现的次数比 $I_n=P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k}$ 中 i_1 出现的次数要少. 于是我们还可以进行上面的步骤, 直至把指标 i_1 的个数消成 0. 但是在交换关系 $(*)$ 式中, 只有第 1 个式子, 才有可能把指标 i_1 从存在消成不存在, 但是这个式子正好会把 I_n 的写成对换的乘积的个数减少 2 个, 这就回到了上面的步骤了: 利用归纳假设 $k-2$ 是偶数, 进而 k 是偶数.

证毕. □

推论 6.14. 一个置换矩阵写成对换矩阵的乘积, 那么对换的个数奇偶性不变, 即: 若 $P = P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k} = P_{i'_1j'_1}P_{i'_2j'_2}\cdots P_{i'_mj'_m}$, 那么 k, m 同时为奇数或同时为偶数, 或者等价地说: $k+m$ 是偶数.

证明. 若 $P = P_{i_1j_1}P_{i_2j_2}\cdots P_{i_kj_k} = P_{i'_1j'_1}P_{i'_2j'_2}\cdots P_{i'_mj'_m}$, 那么

$$I_n = P_{i_kj_k}\cdots P_{i_1j_1}P = P_{i_kj_k}\cdots P_{i_1j_1}P_{i'_1j'_1}P_{i'_2j'_2}\cdots P_{i'_mj'_m}$$

由上面的引理知: $k+m$ 是偶数. □

注释 6.15. 我们很容易就可以验证, (6.1) 式中的函数满足定理 (6.5) 的 3 个条件, 于是存在性也被验证了. 请作为习题做一遍.

注释 6.16. 由我们的证明过程可知, 满足定理 (6.5) 中除了 $f(I_n) = 1$ 的条件以外的所有条件的函数都形如 $c \cdot \det$, 其中 $c = f(I_n)$.

推论 6.17. 对任意 $A, B, C \in M_n(\mathbb{R})$, 则

$$\det\left(\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}\right) = \det(A)\det(B)$$

证明. 若 A 或 B 不可逆, 则 $\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$ 不可逆, 此时等式两边都等于 0.

若 A, B 都可逆, 那么 C 的行向量是 B 的行向量的线性组合, 我们只需要做 "把某一行的若干倍加到另一行" 的初等行变换, 就可以把 $\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$ 变为 $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$. 所以

$$\det\left(\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}\right) = \det\left(\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right).$$

我们只需证 $\det\left(\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right) = \det(A)\det(B)$. 对任意固定 B, 我们考察函数

$$f_B : M_{n\times n}(\mathbb{R}) \longrightarrow \mathbb{R} \qquad A \longmapsto \det\left(\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right),$$

那么它满足定理定理 (6.5) 中除了 $f_B(I_n) = 1$ 以外所有的条件, 所以我们有函数 $M_{n\times n}(\mathbb{R})$ 上的函数相等: $f_B = f_B(I_n)\det$, 即对任意 $A \in M_{n\times n}(\mathbb{R})$,

$$f_B(A) = f_B(I_n)\det(A).$$

而 $f_B(I_n) = \det\left(\begin{bmatrix} I_n & 0 \\ 0 & B \end{bmatrix}\right)$, 类似的, 我们考察函数

$$M_{n\times n}(\mathbb{R}) \longrightarrow \mathbb{R} \qquad B \longmapsto \det\left(\begin{bmatrix} I_n & 0 \\ 0 & B \end{bmatrix}\right),$$

那么它满足定理定理 (6.5) 中所有的条件, 所以任意 $B \in M_{n\times n}(\mathbb{R})$,

$$\det\left(\begin{bmatrix} I_n & 0 \\ 0 & B \end{bmatrix}\right) = \det(B).$$

这就证明了我们的推论. □

下面我们看一下初等矩阵的行列式的值:

命题 6.18. 假设函数 $f : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ 满足定理 (6.5) 中的条件, 那么:

1. 若 E 是把某一行的若干倍加到另一行, 那么 $f(E) = 1$;
2. 若 $E(c)$ 是把某一行乘以 c 倍, 那么 $f(E) = c$;
3. 若 P 是交换两行, 那么 $f(P) = -1$.

证明. 1. 条件 1 中取单位阵可得.

2. 条件 2 中取单位阵可得.
3. 条件 3 中取单位阵可得.

□

例题 6.19. 1. 对于二阶矩阵 $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, 直接根据定义, 我们有

$$\det(A) = a_{11}a_{22}\det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + a_{12}a_{21}\det\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$

2. 对于三阶矩阵 $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, 根据定义, 它的行列式有 6 项. 我们直接把它写出来:

$$\det(A) = (a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}) - (a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33} + a_{11}a_{22}a_{31}).$$

怎么记忆这个公式呢? 见下图:

$$\begin{tikzcd}
a_{11} \arrow[ddrr] & a_{12} \arrow[ddrr] & a_{13} \arrow[ddrr] \arrow[ddll] & a_{11} \arrow[ddll] & a_{12} \arrow[ddll] \\
a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\
a_{31} & a_{32} & a_{33} & a_{31} & a_{32}
\end{tikzcd}$$

将指向右下方的三个箭头 $\searrow$ 经过的三项乘起来相加, 减去将指向左下方的三个箭头 $\swarrow$ 经过的三项乘起来相加的和.

这两个公式务必背熟! 这两个公式务必背熟! 这两个公式务必背熟!

说了半天废话, 我们来点实际的:

例题 6.20. $a_1, a_2, \cdots, a_n \neq 0$, 计算 n 阶矩阵的行列式:

$$\begin{bmatrix}
x_1 - a_1 & x_2 & x_3 & \cdots & x_n \\
x_1 & x_2 - a_2 & x_3 & \cdots & x_n \\
x_1 & x_2 & x_3 - a_3 & \cdots & x_n \\
\vdots & \vdots & \vdots & & \vdots \\
x_1 & x_2 & x_3 & \cdots & x_n - a_n
\end{bmatrix}$$

解. 不要害怕, 做行列变换永远是计算行列式的首选! 这里我们假装知道作初等列变换和作初

等行变换, 行列式满足相同的变化规律, i.e. 假装我们知道了推论 (6.26).

$$\begin{bmatrix} x_1-a_1 & x_2 & x_3 & \cdots & x_n \\ x_1 & x_2-a_2 & x_3 & \cdots & x_n \\ x_1 & x_2 & x_3-a_3 & \cdots & x_n \\ \vdots & \vdots & \vdots & & \vdots \\ x_1 & x_2 & x_3 & \cdots & x_n-a_n \end{bmatrix} \xrightarrow{Row_{i+1}-Row_1} \begin{bmatrix} x_1-a_1 & x_2 & x_3 & \cdots & x_n \\ a_1 & -a_2 & 0 & \cdots & 0 \\ a_1 & 0 & -a_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_1 & 0 & 0 & \cdots & -a_n \end{bmatrix}$$

$$\xrightarrow{a_i^{-1}C_i} a_1a_1\cdots a_n \begin{bmatrix} \frac{x_1}{a_1}-1 & \frac{x_2}{a_2} & \frac{x_3}{a_3} & \cdots & \frac{x_n}{a_n} \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

$$\xrightarrow{C_1+C_2+\cdots+C_n} a_1a_1\cdots a_n \begin{bmatrix} \sum\limits_{i=1}^n \frac{x_i}{a_i}-1 & \frac{x_2}{a_2} & \frac{x_3}{a_3} & \cdots & \frac{x_n}{a_n} \\ 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

这是一个上三角矩阵, 所以答案是

$$(-1)^{n-1}a_1a_2\cdots a_n\left(\sum_{i=1}^n \frac{x_i}{a_i}-1\right).$$

□

例题 6.21. 计算行列式

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ 1 & \binom{2}{1} & \cdots & \binom{n}{1} \\ 1 & \binom{3}{2} & \cdots & \binom{n+1}{2} \\ \vdots & \vdots & & \vdots \\ 1 & \binom{n}{n-1} & & \binom{2n-2}{n-1} \end{vmatrix}$$

其中 $\binom{n}{m}$ 是 n 个不同元素中取 m 个的组合数.

解. 这个只需要每次都把每一行减去上一行. 作为练习题, 自己做. □

6.2.1 行列式的性质

首先, 在问题 (6.2) 中涉及的所有要求都是行列式的性质. 我们不厌其烦地再抄一遍: 对任意 $A \in M_{n\times n}(\mathbb{R})$:

1. $\det(U) = u_1u_2\cdots u_n$, 如果 U 是对角线为 $u_1, u_2, \cdots, u_n$ 的上三角矩阵.
2. $\det(EA) = \det(A)$, E 是把其中一行的若干倍加到另一行;
3. $\det(PA) = -\det(A)$, P 是交换两行.
4. $\det\left(\begin{bmatrix} \vdots \\ c\alpha_i \\ \vdots \end{bmatrix}\right) = c\det\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$.
5. $\det\left(\begin{bmatrix} \vdots \\ \alpha_i+\beta_i \\ \vdots \end{bmatrix}\right) = \det\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right) + \det\left(\begin{bmatrix} \vdots \\ \beta_i \\ \vdots \end{bmatrix}\right).$;
6. $\det\left(\begin{bmatrix} \vdots \\ \alpha_i \\ \vdots \\ \alpha_j \\ \vdots \end{bmatrix}\right) = -1\cdot\det\left(\begin{bmatrix} \vdots \\ \alpha_j \\ \vdots \\ \alpha_i \\ \vdots \end{bmatrix}\right)$.

命题 6.22. *对任意 $A, B \in M_{n\times n}(\mathbb{R})$, 我们有*

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

$A \in M_{n\times n}(\mathbb{R})$, 则 A 是可逆矩阵当且仅当 $\det(A) \neq 0$.

证明. 如果 A 不可逆, 那么 $\operatorname{rank}(A) < n$, 所以 $\operatorname{rank}(AB) \le \operatorname{rank}(A) < n$, 所以 $\det(A)\det(B) = 0 = \det(AB)$;

如果 A 可逆, 那么根据高斯约当消去, A 可以写成初等矩阵的乘积, 故可设 $A = X_1X_2\cdots X_r$, 其中 X_i 都是初等矩阵. 我们首先对三种初等矩阵证明这个命题:

- 若 $X = E$, 那么这个命题就是上边的第 2 条.

- 若 $X = P$, 那么这个命题就是上边的第 3 条.
- 若 $X = E_i(c)$, 那么这个命题就是上边的第 4 条.

因此, 若 A 是初等矩阵, 命题成立. 我们反复使用这一点: 首先

$$\det(A) = \det(X_1X_2\cdots X_r) = \det(X_1)\det(X_2\cdots X_r) = \cdots = \det(X_1)\det(X_2)\cdots\det(X_r);$$

其次,

$$\begin{aligned}\det(AB) &= \det(X_1X_2\cdots X_rB) = \det(X_1)\det(X_2\cdots X_rB)\\ &= \cdots = \det(X_1)\det(X_2)\cdots\det(X_r)\det(B)\\ &= \det(A)\det(B).\end{aligned}$$

证毕 □

推论 6.23. $A \in M_{n\times n}$, 那么 A 可逆当且仅当 $\det(A) \neq 0$.

证明. 我们已知若 A 不可逆, 则 $\det(A) = 0$; 我们还知道若 A 是初等矩阵, 则 $\det(A) \neq 0$.

若 A 可逆, 那么 A 可以写成初等矩阵的乘积, 从而根据上面的命题, 就有 $\det(A) \neq 0$. □

推论 6.24. 若 $A \in M_{n\times n}$ 可逆, 那么 $\det(A^{-1}) = \det(A)^{-1}$.

证明. 若 A 可逆, 则 A^{-1} 存在且 $\det(A) \neq 0$, 所以

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

□

命题 6.25. 对任意 $A \in M_{n\times n}(\mathbb{R})$, 我们有

$$\det(A) = \det(A^T).$$

证明. 若 A 不可逆, 则 A^T 不可逆, 所以 $\det(A) = 0 = \det(A^T)$.

若 A 可逆, 那么 A 可以写成初等矩阵的乘积, 故可设 $A = X_1X_2\cdots X_r$, 其中 X_i 都是初等矩阵. 直接验证可知初等矩阵显然满足这个命题, 从而

$$\begin{aligned}\det(A^T) &= \det((X_1X_2\cdots X_r)^T)\\ &= \det(X_r^T\cdots X_2^TX_1^T)\\ &= \det(X_r^T)\cdots\det(X_2^T)\det(X_1^T)\\ &= \det(X_r)\cdots\det(X_2)\det(X_1)\\ &= \det(X_1)\det(X_2)\cdots\det(X_r)\\ &= \det(X_1X_2\cdots X_r)\\ &= \det(A).\end{aligned}$$

□

推论 6.26. 对矩阵作初等列变换, 行列式的变化规律和作初等行变换行列式的变换规律一致; 即:

- 将一列的若干倍加到另一列上, 行列式的值不变;
- 将一列乘以 c 倍, 行列式的值也变为原来的 c 倍;
- 交换两列, 行列式的值变为原来的 -1 倍;

证明. 证明很简单, 利用上面的公式, 可以把初等列变换通过转置变成初等行变化. □

例题 6.27. 设 A 是 n 阶正交矩阵, 且 $\det(A) < 0$. 证明: $\det(A + I_n) = 0$.

例题 6.28 (范德蒙行列式).

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

形如等式右边的矩阵称作范德蒙矩阵.

证明. 作初等行变换:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix} \xrightarrow{\text{Row}_i - x_1\text{Row}_{i-1}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \\ \vdots & & & \vdots \\ 0 & x_2^{n-1} - x_2^{n-2}x_1 & \cdots & x_n^{n-1} - x_n^{n-2}x_1 \end{bmatrix}$$

由推论 (6.17) 知

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

而

$$\begin{bmatrix} x_2 - x_1 & \cdots & x_n - x_1 \\ \vdots & & \vdots \\ x_2^{n-1} - x_2^{n-2}x_1 & \cdots & x_n^{n-1} - x_n^{n-2}x_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_2 & x_3 & \cdots & x_n \\ \vdots & \vdots & & \vdots \\ x_2^{n-2} & x_3^{n-2} & \cdots & x_n^{n-2} \end{bmatrix} \begin{bmatrix} x_2 - x_1 & 0 & \cdots & 0 \\ 0 & x_3 - x_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_n - x_1 \end{bmatrix}$$

所以

$$V_n(x_1,x_2,\cdots,x_n)=\left(\prod_{i=2}^{n}(x_i-x_1)\right)\cdot V_{n-1}(x_2,x_3,\cdots,x_n).$$

逐步递归可知

$$V_n(x_1,x_2,\cdots,x_n)=\prod_{1\le i<j\le n}(x_j-x_i).$$

□

另解. 我们把范德蒙矩阵的行列式看成是关于 $x_1,x_2,\cdots,x_n$ 的多项式. 当 $x_i=x_j$ 时, 矩阵不可逆, 因此行列式等于零, 所以 $V(x_1,x_2,\cdots,x_n)$ 有因子 x_i-x_j. 所以

$$\prod_{1\le i<j\le n}(x_i-x_j)\Big|V_n(x_1,x_2,\cdots,x_n).$$

我们计算一下两边的多项式:

- $\prod_{1\le i<j\le n}(x_i-x_j)$ 的次数是 $\dfrac{n(n-1)}{2}$. 这是因为一共只有 $\binom{n}{2}=\dfrac{n(n-1)}{2}$ 个一次因子. 我们对它做字典排序, 其首项是 $x_n^{n-1}x_{n-1}^{n-2}\cdots x_2$, 它的系数是 $(-1)^{n-1}(-1)^{n-2}\cdots(-1)^1=(-1)^{\frac{n(n-1)}{2}}$
- 对 $V_n(x_1,x_2,\cdots,x_n)$ 利用行列式大公式知它的次数也是 $\dfrac{n(n-1)}{2}$, 并且 $x_n^{n-1}x_{n-1}^{n-2}\cdots x_2$ 的系数是 1.

所以

$$V_n(x_1,x_2,\cdots,x_n)=(-1)^{\frac{n(n-1)}{2}}\prod_{1\le i<j\le n}(x_i-x_j)=\prod_{1\le i<j\le n}(x_j-x_i).$$

□

例题 6.29. 给定平面上的 n 个点 $(x_1,y_1),\cdots,(x_n,y_n)$, x_i 互不相等. 那么存在唯一的 n 次多项式函数 $f(x)$ 使得 $f(x_i)=y_i$, $1\le i\le n$.

证明. 我们用待定系数法: 设 $f(x)=a_0+a_1x+\cdots+a_nx^{n-1}$; 则

$$\begin{cases} a_0+a_1x_1+a_2x_1^2+\cdots+a_nx_1^{n-1} & =y_1\\ a_0+a_1x_2+a_2x_2^2+\cdots+a_nx_2^{n-1} & =y_2\\ \vdots & \vdots\\ a_0+a_1x_n+a_2x_n^2+\cdots+a_nx_n^{n-1} & =y_n\end{cases}$$

所以系数 $a_0, a_1, \cdots, a_n$ 必须满足方程组

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

系数矩阵是范德蒙矩阵的转置, 所以可逆, 因此方程组有唯一解. □

6.3 余子式展开, 逆矩阵和 Cramer 法则

设矩阵 $A \in M_{n\times n}(\mathbb{R})$, 对任意的 $1 \le i, j \le n$, 我们记 $M_{ij} \in M_{(n-1)\times(n-1)}$ 是将 A 的第 i 行第 j 列全部删除剩下的部分, 如图:

$$M_{ij} = \begin{pmatrix} a_{11} & \cdots & a_{1j} & & \cdots a_{in} \\ \vdots & & & \vdots & \\ a_{i1} & & a_{ij} & \cdots & a_{in} \\ \vdots & & & \vdots & \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{pmatrix}$$

定义:

$$C_{ij} = (-1)^{i+j} \det(M_{ij}).$$

定义 6.30. 我们称 C_{ij} 为 A 的第 (i, j) 个代数余子式.

定理 6.31 (Laplace 展开). 设矩阵 $A = (a_{ij}) \in M_{n\times n}(\mathbb{R})$, 对任意 $1 \le i \le n$, 我们都有:

$$\det(A) = \begin{cases} a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} & \text{按第 } i \text{ 行展开} \\ a_{1i}C_{1i} + a_{2i}C_{2i} + \cdots + a_{ni}C_{ni} & \text{按第 } i \text{ 列展开} \end{cases}$$

这个证明并不复杂, 只是繁琐而已, 我们只需要验证上面的等号右边的表达式作为 A 的函数也满足定理 (6.5) 中的三条要求, 再由唯一性可以得到等号.

怎么想到 Laplace 展开的呢? 其实也很简单, 在定理 (6.5) 的证明中, 我们已经给出了 det 的表达式, 我们只需要适当地合并同类项, 然后观察即可.

推论 6.32. 若 $j \ne i$, 那么

$$a_{j1}C_{i1} + a_{j2}C_{i2} + \cdots + a_{jn}C_{in} = 0$$
$$a_{1j}C_{1i} + a_{2j}C_{2i} + \cdots + a_{nj}C_{ni} = 0$$

证明. 我们只证明第一个等式. 设 $A' \in M_{n\times n}(\mathbb{R})$, 它的第 i 行为 A 的第 j 行, 其他行和 A 相同, 那么

$$\det(A') = a_{j1}C_{i1} + a_{j2}C_{i2} + \cdots + a_{jn}C_{in}.$$

但是 A' 的第 i 行和第 j 行都是 A 的第 j 行, 所以 $\det(A') = 0$. □

我们定义矩阵 $C = (C_{ij})^T \in M_{n\times n}(\mathbb{R})$, 称为 A 的**伴随矩阵**.

命题 6.33.

$$AC = CA = \det(A)I_n.$$

也就是说 AC, CA 的对角线元素都等于 $\det(A)$, 其余元素都等于 0.

注释 6.34. 注意, 此处极易混淆. 因为 C 的第 (i,j) 分量等于 C_{ji}, 而非 C_{ij}.

证明. 我们记 $C = (c_{ij})$, 则 $c_{ij} = C_{ji}$. 那么

$$(AC)_{(ij)} = \sum_{k=1}^{n} a_{ik}c_{kj} = \sum_{k=1}^{n} a_{ik}C_{jk} = \begin{cases} 0 & i \neq j \\ \det(A) & i = j \end{cases}$$

□

推论 6.35. 若 A 是可逆矩阵, 那么

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

例题 6.36. 求范德蒙矩阵

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}$$

的逆矩阵.

证明. 记

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}$$

我们先计算计算 C_{ji}. 将范德蒙行列式按第 i 列展开, 那么

$$V_n(x_1,x_2,\cdots,x_n)=C_{1i}+x_iC_{2i}+\cdots+x_i^jC_{j+1,i}+\cdots+x_i^{n-1}C_{ni}.$$

我们把等式两边看成是关于 x_i 的多项式, 那么 C_{ji} 就是 x_i^{j-1} 的系数. 于是我们展开范德蒙矩阵的行列式

$$\begin{aligned}
V_n(x_1,x_2,\cdots,x_n) &= \prod_{1\le j<k\le n}(x_k-x_i)\\
&= \left((-1)^{n-i}\prod_{\substack{1\le j<k\le n\\ j,k\neq i}}(x_k-x_j)\right)\prod_{\ell\neq i}(x_i-x_\ell).
\end{aligned}$$

而 $\prod\limits_{\ell\neq i}(x_i-x_\ell)$ 中 x_i^{j-1} 项的系数为

$$(-1)^{n-j}\sum_{\substack{1\le t_1<t_2<\cdots<t_{n-j}\\ t_1,t_2,\cdots,t_{n-j}\neq i}}x_{t_1}x_{t_2}\cdots x_{t_{n-j}}.$$

所以

$$C_{ji}=(-1)^{i+j}\left(\prod_{\substack{1\le j<k\le n\\ j,k\neq i}}(x_k-x_j)\right)\left(\sum_{\substack{1\le t_1<t_2<\cdots<t_{n-j}\le n\\ t_1,t_2,\cdots,t_{n-j}\neq i}}x_{t_1}x_{t_2}\cdots x_{t_{n-j}}\right)$$

所以 A 的逆矩阵的 (i,j) 位置的分量就等于 (注意此处有转置)

$$\frac{C_{ji}}{\det(A)}=(-1)^{n-j}\frac{\sum\limits_{\substack{1\le t_1<t_2<\cdots<t_{n-j}\le n\\ t_1,t_2,\cdots,t_{n-j}\neq i}}x_{t_1}x_{t_2}\cdots x_{t_{n-j}}}{\prod\limits_{\ell\neq i}(x_i-x_\ell)}.$$

□

另解. 我们也可以根据 Lagrange 插值公式来计算范德蒙矩阵的逆矩阵. 首先我们注意下面这个差值公式:

$$f(x)=\sum_{i=0}^{n}\frac{(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}\cdot y_i$$

那么 $f(x_i)=y_i$. 根据唯一性, 它就是插值平面上的点 $(x_1,y_1),\cdots,(x_n,y_n)$ 的唯一的次数为 $n-1$ 的多项式.

对 $0\le i\le n$, 令 $L_i(x)=\dfrac{(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}$, 它满足

$$L_i(x_j)=\begin{cases}0 & i\neq j\\ 1 & i=j\end{cases}$$

那么, $L_i(x)$ 是插值了平面上的点: $(x_0,0),\cdots,(x_{i-1},0),(x_i,1),(x_{i+1},0),\cdots,(x_n,0)$ 的唯一的次数为 n 的多项式.

另一方面, 假设 $A^{-1}=(u_{ij})$, 直接根据矩阵乘法的定义: $A^{-1}A$ 的第 (i,j) 个分量为 $u_{i1}\cdot 1+u_{i2}\cdot x_j^2+\cdots u_{in}x_j^{n-1}$, 所以

$$u_{i1}\cdot 1+u_{i2}\cdot x_j^2+\cdots u_{in}x_j^{n-1}=\begin{cases}0 & j\neq i\\ 0 & j=i\end{cases}$$

那么, 我们令 $\widetilde{L}_i(x)=u_{i1}+u_{i2}x+\cdots u_{in}x^{n-1}$, 那么我们有

$$\widetilde{L}_i(x_j)=\begin{cases}0 & i\neq j\\ 1 & i=j\end{cases}$$

这说明多项式 $L_i(x)$ 和 $\widetilde{L}_i(x)$ 相等! 所以 u_{ij} 等于 $L_i(x)$ 的 x^{j-1} 的系数. □

为什么要多此一举地再给一个解法呢? 这不是为了说明我很牛, 而是这种解法有深刻的内涵: 它和快速傅里叶变换 (FFT) 有紧密的联系, FFT 的主要内容就出自于范德蒙行列式的基本性质. 所以我在这里写出来, 同学们将来学到了相关内容, 再来回顾理解.

推论 6.37 (Cramer 法则). 设 $A\in M_{n\times n}(\mathbb{R})$ 是可逆矩阵, $\boldsymbol{b}=\begin{bmatrix}b_1 & b_2 & \cdots & b_n\end{bmatrix}^T\in\mathbb{R}^n$, 设 B_i 是将 A 的第 i 个列向量替换成 $\boldsymbol{b}$ 得到的矩阵. 则线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 的解为

$$x_1=\frac{\det(B_1)}{\det(A)},\quad x_2=\frac{\det(B_2)}{\det(A)},\quad\cdots,\quad x_n=\frac{\det(B_n)}{\det(A)}.$$

证明. A 可逆, 那么 $A\boldsymbol{x}=\boldsymbol{b}$ 的解为 $\boldsymbol{x}=A^{-1}\boldsymbol{b}=\frac{1}{\det(A)}C\boldsymbol{b}$, 若 $C=(c_{ij})$, 则 $c_{ij}=C_{ji}$, 且

$$\begin{aligned}x_i &= \frac{1}{\det(A)}\left(c_{i1}b_1+c_{i2}b_2+\cdots+c_{in}b_n\right)\\ &= \frac{1}{\det(A)}\left(b_1C_{1i}+b_2C_{2i}+\cdots+b_nC_{ni}\right)\\ &= \frac{\det(B_i)}{\det A}.\end{aligned}$$

□

我们来玩一点儿花活儿, 重新证明以下 Cramer 法则:

证明. 我们把 A 写成列向量的形式: $A=\begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_n\end{bmatrix}$, 那么 $A\boldsymbol{x}=\boldsymbol{b}$ 等价于 $\boldsymbol{b}=$

$x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n$, 于是

$$\begin{aligned}
\det(B_i) &= \det\left(\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \boldsymbol{b} & \cdots & \alpha_n \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n & \cdots & \alpha_n \end{bmatrix}\right) \\
&= \sum_{j=1}^{n} \det\left(\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & x_j\alpha_j & \cdots & \alpha_n \end{bmatrix}\right)
\end{aligned}$$

上面的求和式中, 只有 $j = i$ 的项有非零贡献, 这一项等于

$$\det\left(\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & x_i\alpha_i & \cdots & \alpha_n \end{bmatrix}\right) = x_i \det(A).$$

证毕. □

6.4 行列式的计算

行列式的计算有两大基本方法: 递归和打洞. 所谓的递归, 是对方阵的阶数做归纳; 这要求矩阵内部的元素分布很整齐. 打洞, 就是作行列变换, 人为地制造很多零出来, 然后利用 Laplace 展开. 事实上, Laplace 展开用来具体地计算行列式的值也只有当矩阵内有很多零时才显示出优势.

例题 6.38. 计算矩阵的行列式

$$\begin{vmatrix}
\lambda & 0 & \cdots & & 0 & a_{n-1} \\
-1 & \lambda & \cdots & & 0 & a_{n-2} \\
0 & -1 & \ddots & & 0 & a_{n-3} \\
\vdots & \vdots & \ddots & \ddots & & \vdots \\
0 & & & -1 & \lambda & a_1 \\
0 & 0 & & \cdots & -1 & a_0 - \lambda
\end{vmatrix}$$

打洞: 逐行作行变化.

例题 6.39. $a \neq b$, 计算矩阵的行列式

$$\begin{vmatrix}
a+b & ab & 0 & \cdots & 0 & 0 \\
1 & a+b & ab & \cdots & 0 & 0 \\
0 & 1 & a+b & \cdots & 0 & \\
\vdots & \vdots & \ddots & \ddots & & \vdots \\
0 & 0 & 0 & & 1 & a+b
\end{vmatrix}$$

递归.

第七章　特征值特征向量

7.1　引入

我们回顾一下 Fibonacci 数列.

例题 7.1. 我们如下定义斐波那契数列：$a_1 = a_2 = 1, a_{n+2} = a_{n+1} + a_n$. 求 $\{a_n\}$ 的通项公式。

我们设向量 $\alpha_n = \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix}$, 则

$$\alpha_{n+1} = \begin{bmatrix} a_{n+1} \\ a_{n+2} \end{bmatrix} = \begin{bmatrix} a_{n+1} \\ a_n + a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix}$$

令 $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$, 则 $\alpha_{n+1} = A^n \alpha_1$. 所以我们要求 a_n 的通项公式, 只需要把 A^n 算出来.

我们尝试几次就会发现, 直接计算 A^n 和计算 Fibonacci 数列本身没有任何区别, 所以直接计算是行不通的. 那么我们就要把问题转化一下. 怎么转化呢? 在第二次习题课的时候, 就给过同学们这样的例子: 寻找可逆矩阵 P 使得 $P^{-1}AP$ 是对角矩阵 $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. 如果能做到这一点, 那么

$$(P^{-1}AP)^n = \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix},$$

于是 $A^n = P \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} P^{-1}$. 问题就转化为寻找可逆矩阵 P, 使得 $P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. 我们继续推导. 令 $\boldsymbol{e}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$, 那么

$$P^{-1}AP\boldsymbol{e}_i = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i;$$

上式两边同时左乘 P 得: $AP\boldsymbol{e}_i = \lambda_i P\boldsymbol{e}_i$. 令 $\boldsymbol{v}_i = P\boldsymbol{e}_i$, 那么

$$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i.$$

这启发我们给出下面的定义:

定义 7.2. $A \in M_{n\times n}(\mathbb{R})$. 我们称 $\lambda \in \mathbb{R}$ 是 A 的特征值, 如果存在非零向量 $\boldsymbol{v} \in \mathbb{R}^n$ 使得 $A\boldsymbol{v} = \lambda\boldsymbol{v}$; 这个 $\boldsymbol{v}$ 称为是属于特征值 λ 的特征向量.

在这个定义里面, 一定要注意是存在**非零向量** $\boldsymbol{v}$; 否则, 如果允许零向量, 那么一个后果就是任意实数都是 A 的特征值, 这显然是很荒谬的.

注释 7.3. 一个特征值可以有很多属于这个特征值的特征向量, 例如对于单位阵 I_2 而言, 任何向量都是属于特征值 1 的特征向量. 但是一个特征向量只能属于一个特征值, 若不然: $A\boldsymbol{v} = \lambda_1\boldsymbol{v} = \lambda_2\boldsymbol{v}$, 那么 $(\lambda_1 - \lambda_2)\boldsymbol{v} = 0$; 于是 $\boldsymbol{v} = 0$, 矛盾.

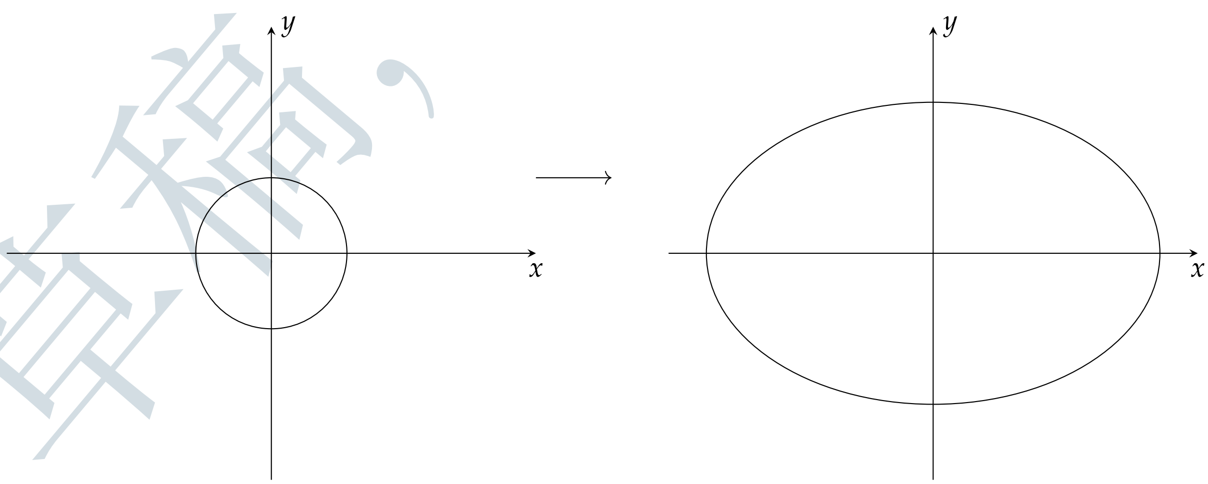
定义 7.4. 对于矩阵 $A, B \in M_{n\times n}(\mathbb{R})$, 若存在可逆矩阵 P 使得 $A = P^{-1}BP$, 那么我们称 A 和 B **共轭** (**或者: 相似**) 或者 A 共轭于 (或者: 相似于) B. 如果 A 共轭于某个对角矩阵, 那么我们称 A **可相似对角化**.

习惯上, 我们用可对角化来简称可相似对角化, 虽然还有其他形式的对角化, 例如对二次型而言, 有合同对角化. 但是如果没有特殊说明, 可对角化只指代可相似对角化.

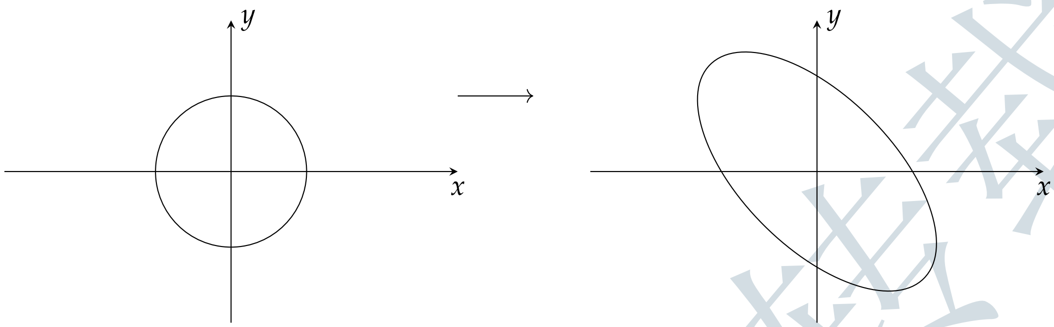
我们再从几何上理解一下特征值特征向量. 还是以二维为例, 设 $A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$, 我们把它看成是线性映射:

$$A : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \qquad \boldsymbol{x} \longmapsto A\boldsymbol{x}\,.$$

我们画个图



这个变换很简单, 就是在 x 轴拉伸 3 倍, y 轴拉伸 2 倍. 但如果 A 不是对角矩阵, 事情就变得复杂, 例如 $A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$,



这个事情就变得很复杂了: 是否还存在两个向量 $\boldsymbol{v}_1, \boldsymbol{v}_2$, 使得该变换就是在这两个方向上的拉伸呢? 等价的说, 是否存在 $\boldsymbol{v}_1, \boldsymbol{v}_2$ 使得 $A\boldsymbol{v}_i$ 是 $\boldsymbol{v}_i$ 的倍数, i.e. $A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$? 就这个问题而言, 我们可以定性地分析一下: $A\boldsymbol{e}_1 = [3, -1]^T$, 在第四象限; $A\boldsymbol{e}_2 = [-1, 3]^T$ 在第二象限, 也就是说, "变量" 的角度变化速度比 "自变量" 大; 因此总有自变量和变量季度重合的时候, 所以一定存在 $\boldsymbol{v}$ 使得 $A\boldsymbol{v}$ 和 $\boldsymbol{v}$ 方向相同. 所以, 对于这个特殊的矩阵, 特征值和特征向量一定存在.

那么理论上我们怎么解决一般的矩阵的特征值和特征向量的问题呢?

7.2 特征多项式, 特征值, 特征向量

我们回顾一下定义:

定义 7.5. $A \in M_{n\times n}(\mathbb{R})$. 我们称 $\lambda \in \mathbb{R}$ 是 A 的特征值, 如果存在非零向量 $\boldsymbol{v} \in \mathbb{R}^n$ 使得 $A\boldsymbol{v} = \lambda\boldsymbol{v}$; 这个 $\boldsymbol{v}$ 称为是属于特征值 λ 的特征向量.

我们的目标是求出 $A \in M_{n\times n}(\mathbb{R})$ 的所有的特征值和特征向量.

- 若 λ 是 A 的特征值, $0 \neq \boldsymbol{v}$ 是属于 λ 的特征向量, 那么 $A\boldsymbol{v} = \lambda\boldsymbol{v}$, 或者等价地 $(A - \lambda I_n)\boldsymbol{v} = 0$, 即: $\boldsymbol{v} \in N(A - \lambda I_n)$, 所以 $A - \lambda I_n$ 不可逆, 从而 $\det(A - \lambda I_n) = 0$.
- 反之, 若 $\lambda \in \mathbb{R}$ 使得 $\det(A - \lambda I_n) = 0$, 那么 $A - \lambda I_n$ 不可逆, 即 $\operatorname{rank}(A - \lambda I_n) \leq n - 1$; 所以 $\dim N(A - \lambda I_n) \geq 1$; 因此存在 $0 \neq \boldsymbol{v} \in N(A - \lambda I_n)$, 于是 $A\boldsymbol{v} = \lambda\boldsymbol{v}$.

综上所述:

命题 7.6. λ 是 A 的特征值, 当且仅当 $\det(A - \lambda I_n) = 0$; 属于 λ 的特征向量全体是 $N(A - \lambda I_n)$ 中的所有非零向量.

这个命题实际上给了我们计算特征值和特征向量的算法:

1. 首先计算行列式: $\det(A-\lambda I_n)$; 关于 λ 这是一个 n 次的多项式, 称作 A 的**特征多项式**, 记作 $f_A(\lambda)$

2. 求出多项式方程 $f_A(\lambda)\det(A-\lambda I_n)=0$ 的所有解: $\lambda_1,\lambda_2,\cdots,\lambda_m$; 它们就是 A 的所有特征值;

3. 对任意 λ_i, 求零空间 $N(A-\lambda_i I_n)$; 我们把 $N(A-\lambda_i I_n)$ 称作 A 的属于 λ_i 的**特征子空间**, 记作 V_{λ_i}. 它的非零元素就是属于特征值 λ_i 的特征向量.

注释 7.7. 在很多教材或者书籍上面, A 的特征多项式是指 $\det(\lambda I_n - A)$, 请注意区分. 事实上, 它们并没有本质区别, 这是因为

$$\det(A-\lambda I_n)=(-1)^n\det(\lambda I_n - A).$$

例题 7.8. 矩阵 $A=\begin{bmatrix}0&1\\-1&0\end{bmatrix}$ 的特征多项式是 λ^2+1, 它没有实数解, 因此没有实特征值. 矩阵 $B=\begin{bmatrix}0&1\\1&0\end{bmatrix}$ 的特征多项式是 λ^2-1, 它有两个实数解: $\lambda_1=1,\lambda_2=-1$; $N(B-\lambda_1)=\mathbb{R}\cdot\begin{bmatrix}1\\1\end{bmatrix}$, 因此属于 λ_1 的所有特征向量是 $c_1\begin{bmatrix}1\\1\end{bmatrix}, c_1\neq 0$; $N(B-\lambda_2)=\mathbb{R}\cdot\begin{bmatrix}1\\-1\end{bmatrix}$, 因此属于 λ_2 的所有特征向量是 $c_2\begin{bmatrix}1\\-1\end{bmatrix}, c_2\neq 0$.

例题 7.9. 我们现在来解决 Fibonacci 数列的例子. $A=\begin{bmatrix}0&1\\1&1\end{bmatrix}$. 先计算行列式:

$$\det(A-\lambda I_2)=\begin{vmatrix}-\lambda&1\\1&1-\lambda\end{vmatrix}=\lambda(1-\lambda)-1=\lambda^2-\lambda-1.$$

它有两个不等的实根 $\lambda_1=\dfrac{1+\sqrt{5}}{2},\lambda_2=\dfrac{1-\sqrt{5}}{2}$. 我们用 Vieta 定理来帮助我们计算: $\lambda_1+\lambda_2=1,\lambda_1\lambda_2=-1$.

$$A-\lambda_1 I_2=\begin{bmatrix}-\lambda_1&1\\1&1-\lambda_1\end{bmatrix}=\begin{bmatrix}-\lambda_1&1\\-\lambda_1\lambda_2&\lambda_2\end{bmatrix},$$

所以 $N(A-\lambda_1 I_2)$ 以 $\boldsymbol{v}_1=\begin{bmatrix}1\\\lambda_1\end{bmatrix}$ 为基; 同理 $N(A-\lambda_2 I_2)$ 以 $\boldsymbol{v}_2=\begin{bmatrix}1\\\lambda_2\end{bmatrix}$ 为基.

注意: $P\boldsymbol{e}_i = \boldsymbol{v}_i$, 所以 $P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}$; 于是 $P^{-1} = \dfrac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix}$; 此时 $P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, 从而

$$P^{-1}A^nP = \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix}$$

即

$$A = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix} \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix} = \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_1^n\lambda_2 - \lambda_1\lambda_2^n & \lambda_2^n - \lambda_1^n \\ \lambda_1^{n+1}\lambda_2 - \lambda_1\lambda_2^{n+1} & \lambda_2^{n+1} - \lambda_1^{n+1} \end{bmatrix}.$$

稍微化简一下, 即得:

$$A = \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2^{n-1} - \lambda_1^{n-1} & \lambda_2^n - \lambda_1^n \\ \lambda_2^n - \lambda_1^n & \lambda_2^{n+1} - \lambda_1^{n+1} \end{bmatrix}$$

因此

$$\begin{bmatrix} a_{n+1} \\ a_{n+2} \end{bmatrix} = A^n \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \dfrac{\lambda_1^n(\lambda_2 - 1) - \lambda_2^n(\lambda_1 - 1)}{\lambda_2 - \lambda_1} \\ \dfrac{\lambda_1^{n+1}(\lambda_2 - 1) - \lambda_2^{n+1}(\lambda_1 - 1)}{\lambda_2 - \lambda_1} \end{bmatrix} = \begin{bmatrix} \dfrac{\lambda_2^{n+1} - \lambda_1^{n+1}}{\lambda_2 - \lambda_1} \\ \dfrac{\lambda_2^{n+2} - \lambda_1^{n+2}}{\lambda_2 - \lambda_1} \end{bmatrix}.$$

这样, 我们就求出了 Fibonacci 数列的通项公式:

$$a_n = \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2}, \quad \lambda_1 = \frac{1+\sqrt{5}}{2}, \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

从 Fibonacci 数列的例子来看, 只要矩阵可以对角化, 那么计算它的高次幂就变得简单了. 那么, 是不是所有矩阵都可以对角化呢? 如果不是, 那哪些矩阵可以对角化? 对于可对角化的矩阵, 我们如何将它对角化?

类似于分析特征值和特征向量, 我们先分析一下矩阵对角化. 假设 $A \in M_{n\times n}(\mathbb{R})$ 可对角化, 那么存在可逆矩阵 P, 使得 $P^{-}1AP = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 这里 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 可以重复. 令 $\boldsymbol{v}_i = P\boldsymbol{e}_i$, 那么

$$A\boldsymbol{v}_i = AP\boldsymbol{e}_i = P(P^{-1}AP)\boldsymbol{e}_i = P \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \boldsymbol{e}_i = P(\lambda_i\boldsymbol{e}_i) = \lambda_i\boldsymbol{v}_i;$$

因此 $\boldsymbol{v}_i$ 就是 A 的属于特征值 λ_i 的特征向量. 另一方面, 由于 $\boldsymbol{v}_i = P\boldsymbol{e}_i$, 注意到 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是线性无关的, 且 P 可逆, 所以 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 也是线性无关的. 所以我们得到了矩阵可对角化的必要条件:

命题 7.10. 若 $A \in M_{n\times n}$ 可对角化, 那么它有 n 个线性无关的特征向量.

例题 7.11. 这个命题的一个推论就是: 并不是所有矩阵都可以对角化. 我们只需要找到矩阵它没有 n 个线性无关的特征向量. 例如 $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, 由于它没有实数特征值, 所以不能对角化. 但是这个例子有点特殊, 虽然它没有实特征值, 但是它有两个不相等的复数特征值: $\pm\sqrt{-1}$. 如果我们类似的考虑 $\mathbb{C}^2$ 和 $M_2(\mathbb{C})$, 那么 $\begin{bmatrix} 1 \\ \sqrt{-1} \end{bmatrix}$ 是属于特征值 $\sqrt{-1}$ 的特征向量, $\begin{bmatrix} 1 \\ -\sqrt{-1} \end{bmatrix}$ 是属于特征值 $-\sqrt{-1}$ 的特征向量. 而且计算发现:

$$\begin{bmatrix} 1 & 1 \\ \sqrt{-1} & -\sqrt{-1} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \sqrt{-1} & \sqrt{-1} \end{bmatrix} = \begin{bmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{bmatrix}.$$

这个现象说明, 虽然有些矩阵在实数域上不可对角化, 但是在复数域上可对角化. 这个现象来源于: 有些系数在实数域上的多项式方程没有实根, 但是它有复数根. 这种障碍可以在复数域中克服, 我们不赘述.

例题 7.12. 设 $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. $\det(A - \lambda I_2) = (\lambda - 1)^2$, 所以 A 只有唯一的特征值 $\lambda = 1$. $A - I_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, 所以 $N(A - I_2) = \mathbb{R} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 所以 A 只有 1 个特征向量. 所以 A 不可对角化. 而且, 即使我们考虑 $\mathbb{C}^2, M_2(\mathbb{C})$, 依然不能将 A 对角化.

如果我们将上一个例子称为 "本质可对角化", 那么这个例子就是 "本质不可对角化". 学习更多的理论, 我们会发现, 所有的 "本质不可对角化" 的矩阵在相似的意义下, 都和这个例子中的矩阵是同一个类型. 如果有兴趣, 可以继续学习有关 Jordan 标准型的内容.

有了必要条件, 我们可以思考它是不是充分条件. 那是不是呢? 当然是, 因为生活很美好!

命题 7.13. 若矩阵 $A \in M_{n\times n}(\mathbb{R})$ 有 n 个线性无关的特征向量, 那么 A 可以对角化.

证明. 设 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 是 A 的 n 个线性无关的特征向量, 它们分别属于特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 即 $A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$. 令 $P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} \in M_{n\times n}(\mathbb{R})$. 由 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 线性无关可知 P 是可逆矩阵. 另一方面, 直接根据矩阵乘法, 我们就有 $P\boldsymbol{e}_i = \boldsymbol{v}_i$. 考虑 $D = P^{-1}AP$; 它的第 i 个列向量等于 $D\boldsymbol{e}_i$, 而

$$D\boldsymbol{e}_i = P^{-1}AP\boldsymbol{e}_i = P^{-1}A\boldsymbol{v}_i = P^{-1}(\lambda_i \boldsymbol{v}_i) = \lambda_i(P^{-1}\boldsymbol{v}_i) = \lambda_i \boldsymbol{e}_i.$$

所以 $D = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. □

总的来说, 我们就得到了下面的定理:

定理 7.14. 矩阵 $A \in M_{n\times n}(\mathbb{R})$ 可对角化当且仅当 A 有 n 个线性无关的特征向量.

事实上, 在方阵 A 可对角化的条件下, 命题 (7.13) 也给出了我们对角化的算法, 也可以参考 Fibonacci 数列的例子. 为了加深印象, 我们还是再给个例子

例题 7.15. 判断矩阵 $A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 5 & -2 \\ -2 & 4 & -1 \end{bmatrix}$ 是否可对角化; 若可对角化, 则将它对角化; 若否, 则给出理由.

我们只需要计算 A 有多少个线性无关的特征向量.

证明.

$$\det(A-\lambda I_2) = \begin{vmatrix} 1-\lambda & 0 & 0 \\ -2 & 5-\lambda & -2 \\ -2 & 4 & -1-\lambda \end{vmatrix} = -(\lambda-1)^2(\lambda-3).$$

因此它有两个特征值: $\lambda_1 = 1, \lambda_2 = 3$.

$$A - \lambda_1 I_3 = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 4 & -2 \\ -2 & 4 & -2 \end{bmatrix}$$

$N(A-\lambda_1 I_3)$ 的一组基为 $\boldsymbol{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

$$A - \lambda_2 I_3 = \begin{bmatrix} -2 & 0 & 0 \\ -2 & 2 & -2 \\ -2 & 4 & -4 \end{bmatrix}$$

$N(A-\lambda_2 I_3)$ 的一组基为 $\boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

所以 $\boldsymbol{v}_1, \boldsymbol{v}_2$ 是 A 的属于 $\lambda_1 = 1$ 的特征向量, $\boldsymbol{v}_3$ 是属于 λ_3 的特征向量. 令

$$P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

则

$$P^{-1}AP = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

□

推论 7.16. *如果 $A \in M_{n\times n}(\mathbb{R})$ 有 n 个不同的实数特征值, 那么 A 可对角化.*

证明. 因为每个特征值都至少有一个特征向量, 所以 A 至少有 n 个线性无关的特征向量. 这是为什么呢? 这是因为属于不同特征值的特征向量是线性无关的, 见下面的引理. □

引理 7.17. *设 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 是 A 的 k 个互不相等的特征值, (未必是所有的不同的特征值), $\boldsymbol{v}_i$ 是属于 λ_i $(1 \le i \le k)$ 的特征向量. 那么 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k$ 线性无关.*

证明. 设 $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k = 0$, 则 $A^i(a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k) = 0$, 所以

$$a_1\lambda_1^i\boldsymbol{v}_1 + a_2\lambda_2^i\boldsymbol{v}_2 + \cdots + a_k\lambda_k^i\boldsymbol{v}_k = 0; \forall i.$$

对 i 取 $0, 1, 2, \cdots, k-1$, 得

$$\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \end{bmatrix} \begin{bmatrix} a_1 & a_1\lambda_1 & \cdots & a_1\lambda_1^{k-1} \\ a_2 & a_2\lambda_2 & \cdots & a_2\lambda_2^{k-1} \\ \vdots & \vdots & & \vdots \\ a_k & a_k\lambda_k & \cdots & a_k\lambda_k^{k-1} \end{bmatrix} = 0. \tag{7.1}$$

注意:

$$\begin{bmatrix} a_1 & a_1\lambda_1 & \cdots & a_1\lambda_1^{k-1} \\ a_2 & a_2\lambda_2 & \cdots & a_2\lambda_2^{k-1} \\ \vdots & \vdots & & \vdots \\ a_k & a_k\lambda_k & \cdots & a_k\lambda_k^{k-1} \end{bmatrix} = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & a_k \end{bmatrix} \begin{bmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{k-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_k & \cdots & \lambda_k^{k-1} \end{bmatrix}$$

上式最后一个矩阵是范德蒙矩阵的转置, 所以它可逆; 我们把 7.1式等号两边右乘该范德蒙矩阵的逆, 就得到:

$$\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \end{bmatrix} \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & a_k \end{bmatrix} = 0;$$

即: $\begin{bmatrix} a_1\boldsymbol{v}_1 & a_2\boldsymbol{v}_2 & \cdots & a_k\boldsymbol{v}_k \end{bmatrix} = 0$. 由于 $\boldsymbol{v}_i$ 都是非零矩阵, 所以 $a_1 = a_2 = \cdots = a_k = 0$. □

那问题来了, 若 A 只有 k $(k<n)$ 个不同的实数特征值咋办呢? 对 $\lambda \in \mathbb{R}$ 是 A 的特征值, 定义它的**几何重数**:

$$d_\lambda := \dim V_\lambda.$$

注释 7.18. 对任意 $\lambda \in \mathbb{R}$, 我们都可以定义 $V_\lambda = N(A-\lambda I_n)$ 和 $d_\lambda = \dim V_\lambda$. 那么 λ 是 A 的特征值, 当且仅当 $V_\lambda \neq 0$, 当且仅当 $d_\lambda \geq 1$.

命题 7.19. A 可对角化当且仅当它有互不相等的实数特征值 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 使得 $d_{\lambda_1} + d_{\lambda_2} + \cdots + d_{\lambda_k} = n$.

证明. "$\Rightarrow$": A 可对角化, 则 A 有 n 个线性无关的特征向量. 而每个特征向量都属于某个特征子空间. 所以 $d_{\lambda_1} + d_{\lambda_2} + \cdots + d_{\lambda_k} = n$.

"$\Leftarrow$": 设 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{d_1}\}$ 是 V_{λ_1} 的一组基, $\{\boldsymbol{v}_{d_1+1}, \boldsymbol{v}_{d_1+2}, \cdots, \boldsymbol{v}_{d_1+d_2}\}$ 是 V_{λ_2} 的一组基, ..., $\{\cdots, \boldsymbol{v}_{d_1+d_2+\cdots+d_k}\}$ 是 V_{λ_k} 的一组基. 若 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{d_1+d_2+\cdots+d_k}$ 线性无关, 则根据条件 $d_1 + d_2 + \cdots + d_k = n$ 知 A 有 n 个线性无关的特征向量, 于是 A 可对角化. 现在我们证明它们线性无关. 设 $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_{d_1+d_2+\cdots+d_k}\boldsymbol{v}_{d_1+d_2+\cdots+d_k} = 0$, 那么我们要证明所有系数都等于 0. 我们给两个证明;

(I). 令 $\beta_1 = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_{d_1}\boldsymbol{v}_{d_1}, \beta_2 = a_{d_1+1}\boldsymbol{v}_{d_1+1} + \cdots + a_{d_1+d_2}\boldsymbol{v}_{d_1+d_2}$, 等等. 那么 $\beta_i \in V_{\lambda_i}$ 且 $\beta_1 + \beta_2 + \cdots + \beta_k = 0$. 首先我们证 $\beta_1, \beta_2, \cdots, \beta_k$ 全都等于 0. 反证: 若不然, 假设其中一共有 ℓ 项非零; $\ell \geq 1$. 我们不妨设 $\beta_1, \beta_2, \cdots, \beta_\ell$ 非零, 那么 $\beta_1 + \beta_2 + \cdots + \beta_\ell = 0$. 这说明 $\beta_1, \cdots, \beta_\ell$ 线性相关, 从而和引理 (7.17) 矛盾. 所以 $\beta_1 = \beta_2 = \cdots = \beta_k = 0$. $\beta_1 = 0$ 就是 $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_{d_1}\boldsymbol{v}_{d_1} = 0$; 由 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{d_1}\}$ 的线性无关性知 $a_1 = a_2 = \cdots = a_{d_1} = 0$. 同理, 其它所有的系数都等于 0. 因此 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{d_1+d_2+\cdots+d_k}$ 线性无关.

(II). β_i 同上, 我们有 $\beta_1 + \beta_2 + \cdots + \beta_k = 0$. 那么 $A^i(\beta_1 + \beta_2 + \cdots + \beta_k) = 0$, 所以

$$\lambda_1^i \beta_1 + \cdots \lambda_k^i \beta_k = 0.$$

对 i 取 $0, 1, 2, \cdots, k-1$, 和引理 (7.17) 同样的可以证明 $\beta_1 = \beta_2 = \cdots = \beta_k = 0$. 剩下来的就和 (I) 是一样的.

□

最后, 我们再研究一下特征多项式. 首先给个预备知识:

定理 7.20 (代数基本定理). 任何一个 n 次多项式 $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$, 它的系数属于 $\mathbb{C}$, 那么它恰好有 n 个复数根.

设 $x_1, x_2, \cdots, x_n$ 是 $f(X) = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$ 的 n 个根, 则

$$f(X) = a_n(X-x_1)(X-x_2)\cdots(X-x_n).$$

那么

$$\begin{aligned} x_1 + x_2 + \cdots + x_n &= -\frac{a_{n-1}}{a_n} \\ \sum_{1\le i_1<i_2<\cdots<i_\ell\le n} x_{i_1}x_{i_2}\cdots x_{i_\ell} &= (-1)^\ell \frac{a_{n-\ell}}{a_n} \\ x_1x_2\cdots x_n &= (-1)^n\frac{a_0}{a_n} \end{aligned}$$

回顾一下, A 的特征多项式就是 $f_A(\lambda)=\det(A-\lambda I_n)$. 我们假设 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 是 $f_A(\lambda)$ 互不相等的复数根, 因此根据代数基本定理, 我们可以假设

$$\det(A-\lambda I_n) = (-1)^n(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_k)^{m_k}.$$

那么 $m_1+m_2+\cdots+m_k=n$.

定义 7.21. 我们把 m_i 称为是 λ_i 的**代数重数**.

命题 7.22. 同一个特征值的几何重数小于等于其代数重数: 即对任意 i, $d_i\le m_i$.

证明. 我们只对特征多项式的实数根证明这一点. 这个命题对复数根也成立, 证明类似, 只不过我们需要考虑复数域上的线性空间. 不妨假设 $\lambda_1\in\mathbb{R}$ 且不妨只证明 $d_1\le m_1$. 由于 $d_1=\dim V_{\lambda_1}$, 我们取它的一组基: $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_{d_1}$. 我们再取 $\boldsymbol{v}_{d_1+1},\boldsymbol{v}_{d_1+2},\cdots,\boldsymbol{v}_n\in\mathbb{R}^n$ 使得 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_{d_1},\boldsymbol{v}_{d_1+1},\cdots,\boldsymbol{v}_n$ 构成 $\mathbb{R}^n$ 的一组基. 于是:

$$A\boldsymbol{v}_1=\lambda_1\boldsymbol{v}_1,\quad A\boldsymbol{v}_2=\lambda_1\boldsymbol{v}_2,\quad \cdots,\quad A\boldsymbol{v}_{d_1}=\lambda_1\boldsymbol{v}_{d_1}.$$

而当 $j\ge d_1+1$ 时, $A\boldsymbol{v}_j\in\mathbb{R}^n$ 是 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n$ 的线性组合, 所以我们可以设

$$A\boldsymbol{v}_j = a_{1j}\boldsymbol{v}_1 + a_{2j}\boldsymbol{v}_2 + \cdots + a_{d_1j}\boldsymbol{v}_{d_1} + a_{d_1+1,j}\boldsymbol{v}_{d_1+1} + \cdots + a_{nj}\boldsymbol{v}_n;$$

写成矩阵形式:

$$\begin{aligned} & A\begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_{d_1} & \boldsymbol{v}_{d_1+1} & \cdots & \boldsymbol{v}_n\end{bmatrix} \\ = & \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_{d_1} & \boldsymbol{v}_{d_1+1} & \cdots & \boldsymbol{v}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & \cdots & 0 & a_{1,d_1+1} & \cdots & a_{1n} \\ & \ddots & & & & \\ 0 & \cdots & \lambda_1 & a_{d_1,d_1+1} & \cdots & a_{d_1n} \\ 0 & \cdots & 0 & a_{d_1+1,d_1+1} & \cdots & a_{d_1+1,n} \\ \vdots & & & \vdots & & \vdots \\ 0 & \cdots & 0 & a_{n,d_1+1} & \cdots & a_{nn}\end{bmatrix}. \\ = & \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_{d_1} & \boldsymbol{v}_{d_1+1} & \cdots & \boldsymbol{v}_n\end{bmatrix}\begin{bmatrix}\lambda_1 I_{d_1} & B \\ 0 & C\end{bmatrix} \end{aligned}$$

令 $P = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$, 则 P 是可逆矩阵, 且 $A = P\begin{bmatrix} \lambda_1 I_{d_1} & B \\ 0 & C \end{bmatrix}P^{-1}$. 所以

$$\begin{aligned} \det(A-\lambda I_n) &= \det\left(P\begin{bmatrix} \lambda_1 I_{d_1} & B \\ 0 & C \end{bmatrix}P^{-1} - \lambda I_n\right) \\ &= \det\left(\begin{bmatrix} (\lambda_1-\lambda) I_{d_1} & B \\ 0 & C-\lambda I_{n-d_1} \end{bmatrix}\right) = (\lambda_1-\lambda)^{d_1}\det(C-\lambda I_{n-d_1}) \end{aligned}$$

如果 λ_1 是 $\det(C-\lambda I_{n-d_1}) = 0$ 的根, 那么 $m_1 > d_1$; 如果不是, 那么 $m_1 = d_1$. 总而言之: $d_1 \le m_1$. □

现在我们假设 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 是它的所有的复数根, 代数重数分别为 $m_1, m_2, \cdots, m_k$, 几何重数分别为 $d_1, d_2, \cdots, d_k$. 假设 $\lambda_1, \lambda_2, \cdots, \lambda_\ell$ 是其中的实数根. 那么

$$d_1 + d_2 \cdots + d_\ell \le m_1 + m_2 + \cdots + m_\ell \le m_1 + m_2 \cdots + m_k = n;$$

所以

$$d_1 + d_2 + \cdots + d_\ell = n \text{ 当且仅当 } \ell = k \text{ 且对任意 } 1 \le i \le k \text{ 都有 } d_i = m_i.$$

也就是说:

定理 7.23. 矩阵 $A \in M_{n\times n}(\mathbb{R})$ 可对角化, 当且仅当它的特征多项式的根全都是实数且每个特征值的几何重数都等于代数重数.

注释 7.24. 这个定理的更重要的意义在于其反面: 若 A 的特征多项式有非实数根, 或者某个实数根的几何重数小于其代数重数, 那么 A 不可对角化.

这就是的我们不必要找出一组极大的线性无关的特征向量组然后数个数, 而是只需要一个一个特征值去验证即可. 这就减少了工作量.

我们最后介绍一些矩阵 A 的不变量和它的特征多项式之间的联系.

首先, 根据行列式大公式, 我们知道 $\det(A-\lambda I_n)$ 的首项系数为 $(-1)^n$, 所以我们有:

$$f_A(\lambda) = (-1)^n(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n).$$

首先, 取 $\lambda = 0$, 就有

$$\det(A) = \lambda_1\lambda_2\cdots\lambda_n.$$

即

命题 7.25. 矩阵 $A \in M_{n\times n}(\mathbb{R})$ 的行列式等于它的特征多项式的所有复数根 (计重数) 的乘积.

再由行列式大公式, 我们知 $\det(A-\lambda I_n)$ 的 λ^{n-1} 项的系数为 $(-1)^{n-1}(a_{11}+a_{22}+\cdots a_{nn})$. 因此我们定义

定义 7.26. A 的对角元的和称为 A 的迹, 记作 $\mathrm{tr}(A)$, 即

$$\mathrm{tr}(A) = a_{11}+a_{22}+\cdots+a_{nn}$$

那么利用根与系数的关于, 考虑特征多项式的 λ^{n-1} 项的系数, 就有

$$\mathrm{tr}(A) = \lambda_1+\lambda_2+\cdots+\lambda_n.$$

即

命题 7.27. 矩阵 $A \in M_{n\times n}(\mathbb{R})$ 的迹等于它的特征多项式的所有复数根 (计重数) 的和.

根据 $\det(P^{-1}AP-\lambda I_n) = \det(A-\lambda I_n)$, 其中 P 是任意一个可逆矩阵, 知:

推论 7.28. 若 P 是可逆矩阵, 则 $\mathrm{tr}(P^{-1}AP) = \mathrm{tr}(A)$.

假设 $f_A(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1}+\cdots+c_1\lambda+c_0$, 那么我们令 $f_A(A) = c_0I_n + c_1A\cdots + c_nA^n \in M_{n\times n}(\mathbb{R})$, 则我们有

定理 7.29 (Calay-Hamilton 定理).

$$f_A(A) = 0.$$

证明. 设 $B(\lambda)$ 是 $A-\lambda I_n$ 的伴随矩阵, 那么 B 的么一个分量都是关于 λ 的次数不超过 $n-1$ 的多项式. 因此存在矩阵 $B_0, B_1, \cdots, B_{n-1} \in M_{n\times n}(\mathbb{R})$ 使得 $B(\lambda) = B_0+\lambda B_1+\cdots+\lambda^{n-1}B_{n-1}$. 另一方面, 我们有 $B(\lambda)(A-\lambda I_n) = f_A(\lambda)I_n$, 即:

$$(B_0+\lambda B_1+\cdots+\lambda^{n-1}B_{n-1})(A-\lambda I_n) = (c_0+c_1\lambda+\cdots+c_n\lambda^n)I_n.$$

将上式左右两边展开, 比较关于 λ 的每个单项, 得到

$$\begin{cases} B_0A &= c_0I_n \qquad 0 \\ B_1A - B_0 &= c_1I_n \qquad 1 \\ \vdots & \\ B_{n-1}A - B_{n-2} &= c_{n-1}I_n \qquad \text{n-1} \\ -B_{n-1} &= c_nI_n \qquad \text{n} \end{cases}$$

我们考虑 0 + 1 $A+\cdots+$ n-1 $A^{n-1}+$ n A^n, 这个式子左边等于 0, 右边等于 $c_0I_n+c_1A+\cdots+c_{n-1}A^{n-1}+c_nA^n$. 证毕. □

7.3 实对称矩阵的对角化

开门见山:

定理 7.30. $A \in M_{n\times n}(\mathbb{R})$, 且 $A = A^T$. 那么存在正交矩阵 Q 使得 $Q^{-1}AQ$ 是对角矩阵.

分析一下问题, 如果 A 可对角化, 那么根据定理 (7.14) 知它必须有 n 线性无关的特征向量, 设为 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$, 那么它可以用可逆矩阵 $P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ 来对角化. 所以第一个问题:

$$A \text{ 是否一定有 } n \text{ 个线性无关的特征向量呢?}$$

我们这个定理还要求 A 可被正交矩阵对角化, 这就意味着我们可以找到 A 的 n 个两两互相正交的单位特征向量. 首先, 对属于 A 的同一特征值的特征向量, 我们可以将其作 Gram-Schmidt 正交化, 使得他们两两正交. 但是属于不同特征值的特征向量的线性组合不再是特征向量, 因此我们无法对属于不同特征值的特征向量作 Gram-Schmidt 正交化; 那怎么办? 这意味着属于 A 的不同特征值的特征向量必须天然地已经正交了!

命题 7.31. 设 $A = A^T$, $\lambda_1 \neq \lambda_2$ 是 A 的两个特征值, $A\boldsymbol{v}_1 = \lambda_1\boldsymbol{v}_1, A\boldsymbol{v}_2 = \lambda_2\boldsymbol{v}_2$, 则 $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = 0$

证明. 这是因为 $\langle A\boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \boldsymbol{v}_1^T A^T \boldsymbol{v}_2 = \boldsymbol{v}_1^T A \boldsymbol{v}_2 = \langle \boldsymbol{v}_1, A\boldsymbol{v}_2 \rangle$, 所以 $\lambda_1 \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \lambda_2 \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle$. $\lambda_1 \neq \lambda_2$ 推出 $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = 0$ □

这个命题有直接的推论:

推论 7.32. 如果 A 有 n 个不同的实数特征值, 则存在正交矩阵 Q 使得 $Q^{-1}AQ$ 是对角矩阵.

证明. 设 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 A 的所有不同特征值, $\boldsymbol{v}_i$ 是属于 λ_i $(1 \leq i \leq n)$ 的特征向量; 我们设 $\|\boldsymbol{v}_i\| = 1$, 否则就用 $\dfrac{\boldsymbol{v}_i}{\|\boldsymbol{v}_i\|}$ 来替换 $\boldsymbol{v}_i$. 于是 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 是两两正交的单位向量; 所以 $Q = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ 是正交矩阵, 且 $Q^{-1}AQ = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. □

那么 A 有没有 n 个不同的实数特征值呢? 结论当然是未必, 例如 $A = I_n$. 甚至连 A 是否有实数特征值我们暂且都不知道. 所以我们先解决这个问题:

命题 7.33. 若 $A = A^T$, $\lambda \in \mathbb{C}$ 是 $\det(A - \lambda I_n) = 0$ 的复数根, 则 $\lambda \in \mathbb{R}$.

证明. 如果我们对复数矩阵用高斯消元法, 那么我们就知道存在 $0 \neq \boldsymbol{z} = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}^T \in \mathbb{C}^n$ 使得 $(A - \lambda I_n)\boldsymbol{z} = 0$, 即: $A\boldsymbol{z} = \lambda\boldsymbol{z}$. (分析一下, 我们要证明 $\lambda \in \mathbb{R}$, 也就是证明 $\lambda = \bar{\lambda}$, 所以我们要取共轭. 条件是 $A = A^T$, 所以我们要取转置. 做题的套路就是这样: 条件结论两边凑凑.) 于是

$$A\bar{\boldsymbol{z}} = \bar{\lambda}\bar{\boldsymbol{z}}, \quad \boldsymbol{z}^T A = \lambda \boldsymbol{z}^T;$$

于是

$$z^T A\bar{z} = \begin{cases} z^T(A\bar{z}) & = \lambda z^T\bar{z} \\ (z^T A)\bar{z} & = \bar{\lambda} z^T\bar{z} \end{cases}$$

而 $z^T\bar{z} = z^T\bar{z} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2$, 所以

$$(\lambda - \bar{\lambda})(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2) = 0,$$

从而 $\lambda = \bar{\lambda}$, 即 $\lambda \in \mathbb{R}$. □

这就是说, A 的所有特征值都是实数. 于是我们取一个实数特征值 λ_1 以及属于 λ_1 的一个单位特征向量 v_1. 那么我们还需要找 $n-1$ 个和 v_1 相互正交的特征向量. 这个暂且找不出来, 但是能找出来 $n-1$ 个和 v_1 相互正交的向量. 那就这样吧, 我们先找出来再说: 取 $\mathrm{Span}(v_1)^\perp$ 的一组基 $w_2, w_3, \cdots, w_n$, 然后将他们做 Gram-Schmidt 正交化, 再单位化, 就得到了 $\mathrm{Span}(v_1)^\perp$ 的一组标准正交基: $v_2, v_3, \cdots, v_n$, 这样 $v_1, v_2, \cdots, v_n$ 就是两两正交的单位向量了. 因为 $Av_1 = \lambda_1 v_1$, 因此 $A\mathrm{Span}(v_1) \subseteq \mathrm{Span}(v_1)$, 根据我给同学们出过的选做题, 因为 A 是对称矩阵, 所以 $A\mathrm{Span}(v_1)^\perp \subseteq \mathrm{Span}(v_1)^\perp$. 当然, 我们也可以直接验证: 对任意 $w \in \mathrm{Span}(v_1)^\perp$, 都有

$$\langle v_1, Aw\rangle = v_1^T Aw = v_1 A^T w = (Av_1)^T w = \langle Av_1, w\rangle = \lambda v_1, w = 0.$$

所以 $Aw \in \mathrm{Span}(v_1)^\perp$. 这就意味着: Av_i $(2 \le i \le n)$ 是 $v_2, v_3, \cdots v_n$ 的线性组合, 即

$$\begin{aligned} Av_2 &= b_{22}v_2 + b_{32}v_3 + \cdots + b_{n2}v_n \\ Av_3 &= b_{23}v_2 + b_{33}v_3 + \cdots + b_{n3}v_n \\ &\vdots \\ Av_2 &= b_{2n}v_2 + b_{3n}v_3 + \cdots + b_{nn}v_n \end{aligned}$$

写成矩阵形式, 就是

$$A\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & b_{22} & b_{23} & \cdots & b_{2n} \\ 0 & b_{32} & b_{33} & \cdots & b_{3n} \\ \vdots & \vdots & & & \vdots \\ 0 & b_{n2} & b_{n3} & \cdots & b_{nn} \end{bmatrix}.$$

记

$$Q_1 = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}, B = \begin{bmatrix} b_{22} & b_{23} & \cdots & b_{2n} \\ b_{32} & b_{33} & \cdots & b_{3n} \\ \vdots & & & \vdots \\ b_{n2} & b_{n3} & \cdots & b_{nn} \end{bmatrix}.$$

那么 Q 是 $n\times n$ 的正交矩阵, B 是 $n-1$ 阶方阵. (啥? $n-1$ 阶? 降阶了? 莫不是骗我用归纳法? 且待我看看 B 是不是对称矩阵!) 用分块矩阵来写, 就是

$$Q_1^TAQ_1=\begin{bmatrix}\lambda_1 & 0\\ 0 & B\end{bmatrix}.$$

等式左边是对称矩阵, 因为 A 是对称的; 所以右边也是对称矩阵, 所以 $B=B^T$ 是对称矩阵. 好的, 由归纳假设存在 $n-1$ 阶正交矩阵 Q_2 使得 $Q_2^TBQ_2=\mathrm{diag}(\lambda_2,\cdots,\lambda_n)$, 所以

$$Q_1^TAQ_1=\begin{bmatrix}\lambda_1 & 0\\ 0 & Q_2\mathrm{diag}(\lambda_2,\cdots,\lambda_n)Q_2^T\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & Q_2\end{bmatrix}\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)\begin{bmatrix}1 & 0\\ 0 & Q_2^T\end{bmatrix}$$

$\begin{bmatrix}1 & 0\\ 0 & Q_2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & Q_2^T\end{bmatrix}=I_n$ 且 $\begin{bmatrix}1 & 0\\ 0 & Q_2\end{bmatrix}^T=\begin{bmatrix}1 & 0\\ 0 & Q_2^T\end{bmatrix}$, 所以 $\begin{bmatrix}1 & 0\\ 0 & Q_2\end{bmatrix}$ 是 n 阶正交阵! 令

$$Q_3=\begin{bmatrix}1 & 0\\ 0 & Q_2\end{bmatrix},$$

那么

$$Q_3^TQ_1^TAQ_1Q_3=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n).$$

令 $P=Q_1Q_3$, 那么 $PP^T=Q_1Q_3Q_3^TQ_1^T=I_n$, 所以 P 也是 n 阶正交阵. 那么

$$P^{-1}AP=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n).$$

也就是说 P 将 A 对角化了.

最后我们还剩下奠基了: $n=1$ 时的情形. 这是平凡的. 因此定理 (八) 就证完了.

反过来, 如果矩阵 A 可以正交对角化: 即 $A=Q^{-1}DQ$, 其中 Q 是正交矩阵, $D=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$ 是对角矩阵. 那么 A 是对称矩阵: 这是因为 Q 是正交矩阵, 则 $Q^{-1}=Q^T$. 所以

$$A^T=\left(Q^TDQ\right)^T=Q^TD^T(Q^T)^T=A.$$

所以:

定理 7.34. 矩阵 $A\in M_n(\mathbb{R})$, $A=A^T$ 当且仅当 A 可以正交对角化.

当 A 可正交对角化的时候, $A=Q^T\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)Q$, $Q=\begin{bmatrix}\boldsymbol{q}_1\\ \boldsymbol{q}_2\\ \vdots\\ \boldsymbol{q}_n\end{bmatrix}$ 是正交矩阵, $\boldsymbol{q}_i$ 是

互相正交的单位行向量, 那么

$$\begin{aligned} A &= \begin{bmatrix} \boldsymbol{q}_1^T & \boldsymbol{q}_2^T & \cdots & \boldsymbol{q}_n^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q}_n \end{bmatrix} \\ &= \lambda_1 \boldsymbol{q}_1^T \boldsymbol{q}_1 + \lambda_2 \boldsymbol{q}_2^T \boldsymbol{q}_2 + \cdots \lambda_n \boldsymbol{q}_n^T \boldsymbol{q}_n. \end{aligned}$$

$\lambda_i \boldsymbol{q}_i^T \boldsymbol{q}_i$ 都是秩为 1 的对称矩阵; 因此 A 可以写成至多 n 个秩为 1 的对称矩阵的和.

注释 7.35. 任何一个秩为 1 的矩阵都可以写成一个列向量 α 和一个行向量 β 的乘积: $\alpha\beta$. 而实对称矩阵的组成成分尤其整齐: $\boldsymbol{q}_i^T \boldsymbol{q}_i$.

实际上, 我们有一个一般性的定理:

定理 7.36 (Schur 分解). 对任意的 $A \in M_{n\times n}(\mathbb{R})$, 若它的特征多项式的根都是实根, 那么存在正交矩阵 P 和上三角矩阵 U, 使得

$$A = P^T U P.$$

证明方法也是归纳法, 证明过程和实对称矩阵可对角化一致, 我们不再赘述. 但是, 利用 Schur 分解, 我们可以给实对称矩阵可对角化一个快速的证明: 因为我们已知实对称矩阵的特征值都是实数, 那么就有 Schur 分解 $A = P^T U P$, 于是 $A^T = P^T U^T P$, $A = A^T$ 知 $U = U^T$, 所以 U 只能是对角矩阵.

7.3.1 Lagrange 极值法的应用

如题, 我们的目的是用 Lagrange 求带约束条件的函数极值的方法, 来证明:

定理 7.37. 对任意 $A \in M_n(\mathbb{R})$, 若 $A = A^T$, 则 A 有实特征值, 即存在 $\lambda \in \mathbb{R}$ 和 $0 \neq \boldsymbol{v} \in \mathbb{R}^n$, 使得 $A\boldsymbol{v} = \lambda \boldsymbol{v}$.

有了这个定理, 我们就可以利用归纳法证明实对称矩阵都可以正交对角化.

我们简单叙述一下 Lagrange 的方法. 设 F 是关于 n 个变元 $x_1, x_2, \cdots, x_n$ 的函数, G 是约束条件, 即 $x_1, \cdots, x_n$ 满足 $G(x_1, \cdots, x_n) = 0$; 例如: $F(x_1, x_2) = x_1 + 2x_2$, 约束条件是 $x_1^2 + x_2^2 = 1$, 或者说 $G(x_1, x_2) = x_1^2 + x_2^2 - 1$. 我们怎么求 F 在哪些点处取得极值呢? Lagrange 引入 $n+1$ 个变元的函数

$$H(x_1, \cdots, x_n) = F(x_1, \cdots, x_n) + \lambda G(x_1, \cdots, x_n),$$

若函数 F, G 的性质比较好, 我们考虑方程组:

$$\frac{\partial H}{\partial x_1}=0,\cdots,\frac{\partial H}{\partial x_n}=0,\frac{\partial H}{\partial \lambda}=0;$$

那么 F 只有在上面方程组的解处才可能取得极值; 或者说, F 的符合约束条件 G 的极值点一定满足上面的方程组.

具体的证明, 就不在这里写了, 请参考多元微积分的教材. 我们用上面的例子演示一遍: $F(x_1,x_2)=x_1+2x_2$, 约束条件是 $x_1^2+x_2^2=1$, 于是 $H=x_1+2x_2-\lambda(x_1^2+x_2^2-1)$; 我们就得到了方程组:

$$\begin{cases} \frac{\partial H}{\partial x_1} & =1+2\lambda x_1 & =0 \\ \frac{\partial H}{\partial x_2} & =2+2\lambda x_2 & =0 \\ \frac{\partial H}{\partial \lambda} & =x_1^2+x_2^2-1 & =0 \end{cases}$$

求得两组解

$$\lambda=\frac{\sqrt{5}}{2},x_1=-\frac{1}{\sqrt{5}},x_2=-\frac{2}{\sqrt{5}} \quad \text{和} \quad \lambda=-\frac{\sqrt{5}}{2},x_1=\frac{1}{\sqrt{5}},x_2=\frac{2}{\sqrt{5}}.$$

简单验证, 就知道在第一组解处取得极小值, 在第二组解处取得极大值. 你们可以利用三角函数的替换, 来看看这个结果是不是和初等方法吻合.

了解了基本方法, 我们就可以证明我们的定理了. 考虑函数:

$$F(x_1,\cdots,x_n)=\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}=\sum_{1\le i,j\le n} a_{ij}x_ix_j;$$

约束条件为 $G(x_1,\cdots,x_n)=x_1^2+\cdots+x_n^2-1$, 即变元在单位球 S 上. 由于单位球是有界闭集, 因此 F 在 S 上能取到最大值 (当然我们考虑最小值也可以, 证明雷同). 而最大值一定是极大值, 因此, 若我们假设 F 在 $(x_1,\cdots,x_n)\in S$ 处取得最大值, 那么它们要满足方程组

$$\frac{\partial H}{\partial x_1}=0,\cdots,\frac{\partial H}{\partial x_n}=0,\frac{\partial H}{\partial \lambda}=0,$$

其中 $H=F+\lambda G=\sum_{1\le i,j\le n} a_{ij}x_ix_j+\lambda(x_1^2+\cdots+x_n^2-1)$. 对任意 $1\le k\le n$,

$$\frac{\partial H}{\partial x_k}=\sum_{1\le i,j\le n} a_{ij}\left(x_j\frac{\partial x_i}{\partial x_k}+x_i\frac{\partial x_j}{\partial x_k}\right)+2\lambda x_k=\sum_{1\le i\le n} a_{kj}x_j+\sum_{1\le i\le n} a_{ik}x_i+2\lambda x_k.$$

我们利用 $A=A^T$ (即 $a_{ij}=a_{ji}$) 和 Lagrange 的结果, 并且对求和指标做一下变换, 就有

$$\sum_{1\le i\le n} a_{ki}x_i+\lambda x_k=0;$$

这就推出了:

$$A\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=-\lambda\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$$

这个意思是说, 极大值处的坐标是 A 的特征向量, 它所属于的特征值可以由 Lagrange 的方法求出来.

重复上一节中的归纳法的步骤, 我们可以证明: A 的所有特征值都是实数, 并且可以正交对角化.

7.3.2 正定的实对称矩阵

实对称矩阵都可以对角化, 那么将它对角化以后的对角矩阵就反映了原矩阵的很多信息.

定义 7.38. 我们称一个实对称矩阵是正定 (半正定) 的, 如果它所有的特征值都是正实数 (非负实数).

也就是说, 对称矩阵 $A\in M_n(\mathbb{R})$ 是正定 (半正定) 的, 如果存在正交矩阵 P 和正实数 (非负实数) $\lambda_1,\lambda_2,\cdots,\lambda_n$, 使得 $P^{-1}AP=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$.

注释 7.39. 显然, 正定的实对称矩阵一定是半正定的: 正实数都是非负实数.

一个半正定的实对称矩阵是正定的, 当且仅当它可逆: 这是因为行列式是所有特征值的乘积.

命题 7.40. 若 A 是半正定 (正定) 的实对称矩阵, 那么存在矩阵 (可逆矩阵) B 使得 $A=B^TB$.

证明. 设 P,λ_i 如上, 令

$$B=\begin{bmatrix}\sqrt{\lambda_1}&0&\cdots&0\\ 0&\sqrt{\lambda_2}&\cdots&0\\ \vdots&&\ddots&\vdots\\ 0&0&\cdots&\sqrt{\lambda_n}\end{bmatrix}P.$$

即可. □

这是什么意思呢? 这个意思基本上是说 “半正定的实对称矩阵可以开方”, 可不可以开方是区别非负数和负数的一个特征, 所以这个命题可以看做是对 “半正定” 的” 正” 的一个刻画. 下面的命题更具体地阐释了 “正定” 的含义:

命题 7.41. 实对称矩阵 $A\in M_n(\mathbb{R})$ 是正定 (半正定) 的, 当且仅当对任意非零 $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{x}^TA\boldsymbol{x}>0$ ($\boldsymbol{x}^TA\boldsymbol{x}\geq 0$).

证明. 我们证明正定的情形, 半正定的情形类似.

由于 A 是实对称矩阵, 存在正交矩阵 P 使得 $P^{-1}AP = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

"⇒". 设 A 是正定的: $\lambda_i > 0, \forall i$, $0 \neq \boldsymbol{x} \in \mathbb{R}^n$. 令 $\boldsymbol{y} = P^{-1}\boldsymbol{x}$, 那么 $\boldsymbol{x} = P\boldsymbol{y}$,

$$\boldsymbol{x}^T A \boldsymbol{x} = \boldsymbol{y}^T P^T A P \boldsymbol{y} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 > 0.$$

"⇐" 设对任意的 $\boldsymbol{x} \neq 0, \boldsymbol{x}^T A \boldsymbol{x} > 0$. 那么设 $\boldsymbol{x}_i$ 是属于特征值 λ_i 的特征向量, 那么 $A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$, 两边左乘 $\boldsymbol{x}_i^T$ 得: $\boldsymbol{x}_i^T A \boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i^T \boldsymbol{x}_i$, 于是

$$\lambda_i = \frac{\boldsymbol{x}_i^T A \boldsymbol{x}_i}{\boldsymbol{x}_i^T \boldsymbol{x}_i} > 0.$$

□

到这里, 我们意识到命题 (7.40) 的逆命题也是正确的:

命题 7.42. 对任意矩阵 (可逆矩阵) $B \in M_{n\times n}(\mathbb{R})$, $A = B^T B$ 是半正定 (正定) 的实对称矩阵.

证明. 实对称性是显然的. 我们证明半正定. 对任意非零 $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{x}^T A \boldsymbol{x} = \boldsymbol{x}^T B^T B \boldsymbol{x} = \|B\boldsymbol{x}\|^2 \geq 0.$$

由命题 (7.41) 知 A 半正定.

进一步地, 如果 B 可逆, 那么 $B\boldsymbol{x} \neq 0$, 所以

$$\boldsymbol{x}^T A \boldsymbol{x} = \boldsymbol{x}^T B^T B \boldsymbol{x} = \|B\boldsymbol{x}\|^2 > 0.$$

同样由命题 (7.41) 知 A 正定. □

一个显然的事实是对于正定的实对称矩阵 A, $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n > 0$. 我们还有更强的结论. 我们先引入一个概念:

定义 7.43. 设 $A = (a_{ij}) \in M_{n\times n}(\mathbb{R})$, 对 $1 \leq k \leq n$, 我们称

$$A_k := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}$$

是 A 的第 k 个顺序子矩阵, 它的行列式 $\det(A_k)$ 称为 A 的第 k 个**顺序主子式**.

注意, 我们对任意的方阵定义了顺序子矩阵和顺序主子式的概念, 而不限于正定的实对称矩阵.

例题 7.44. n 阶方阵 $A=(a_{ij})$ 的第 1 个顺序主子式是 a_{11}, 第 n 个顺序主子式是 $\det(A)$.

命题 7.45. 若 A 是正定的实对称矩阵, 则它的所有顺序子矩阵都是实对称正定矩阵.

证明. 对任意 $1\le k\le n$, A 的第 k 个顺序子矩阵 A_k 是一个 k 阶的实对称矩阵. 要证明它是正定的, 只需要证明对任意非零向量 $\boldsymbol{y}\in\mathbb{R}^k$, $\boldsymbol{y}^TA_k\boldsymbol{y}>0$. 在 $\boldsymbol{y}$ 后面接 $n-k$ 个零, 得非零向量 $\boldsymbol{x}=\begin{bmatrix}\boldsymbol{y}\\0\end{bmatrix}\in\mathbb{R}^n$. 由 A 正定知 $\boldsymbol{x}^TA\boldsymbol{x}>0$. 用分块矩阵乘法知 $\boldsymbol{y}^TA_k\boldsymbol{y}=\boldsymbol{x}^TA\boldsymbol{x}$. □

推论 7.46. 正定的实对称矩阵的所有顺序主子式都是正的.

反过来对不对呢? 答案是肯定的:

命题 7.47. 一个对称矩阵, 如果它的顺序主子式都是正的, 那么它是正定的.

证明. 我们对 A 的阶数 n 作归纳法.

$n=1$ 时结论是平凡的. 假设阶数等于 $n-1$ 时结论成立, 我们来证明阶数等于 n 时的情形. 设

$$A=(a_{ij})=\begin{bmatrix}B&\alpha\\\alpha^T&a_{nn}\end{bmatrix},$$

其中 $\alpha\in\mathbb{R}^{n-1},B\in M_{(n-1)\times(n-1)}(\mathbb{R})$ 是 A 的第 $n-1$ 个顺序子矩阵那么 B 是对称矩阵. 且 B 的第 k $(1\le k\le n-1)$ 个顺序主子式等于 A 的第 k 个顺序主子式. 根据归纳假设, B 是正定的, 于是存在可逆矩阵 C, 使得 $B=C^TC$. 我们对 A 作分块消元:

$$\begin{bmatrix}I_{n-1}&0\\-B^{-1}\alpha^T&1\end{bmatrix}A\begin{bmatrix}I_{n-1}&-B^{-1T}\alpha\\0&1\end{bmatrix}=\begin{bmatrix}B&0\\0&a_{nn}-\alpha^TB^{-1}\alpha\end{bmatrix}.$$

两边取行列式知 $a_{nn}-\alpha^TB^{-1}\alpha=\dfrac{\det(A)}{\det(B)}>0$. 令 $b=(a_{nn}-\alpha^TB^{-1}\alpha)^{1/2}$, $E=\begin{bmatrix}I_{n-1}&-B^{-1T}\alpha\\0&1\end{bmatrix}^{-1}$, 那么有

$$A=E^T\begin{bmatrix}C^TC&0\\0&bb\end{bmatrix}E.$$

再令 $P=\begin{bmatrix}C&0\\0&b\end{bmatrix}E$, 则有 $A=P^TP$. 所以 A 是正定的实对称矩阵. □

总结一下:

定理 7.48. $A \in M_{n\times n}(\mathbb{R})$ 是实对称矩阵, 那么以下等价:

- A 是正定的;
- 存在可逆矩阵 P 使得 $A = P^T P$;
- 对任意非零向量 $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x}^T A \boldsymbol{x} > 0$;
- A 的所有顺序主子式都大于零.

当然并非所有的实对称矩阵都正定, 例如 $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. 由于任何实对称矩阵都可对角化, 我们定义:

定义 7.49. 设 A 是一个实对称矩阵, 它的大于零的特征值的个数 (计重数) 称为 A 的正惯性指数, 小于零的特征值的个数 (计重数) 称为 A 的负惯性指数.

等价地说, 我们可以将 A 对角化

$$A = P\mathrm{diag}(\lambda_1, \cdots, \lambda_p, \lambda_{p+1}, \cdots, \lambda_r, 0, \cdots)P^{-1},$$

其中 $r = \mathrm{rank}(A), \lambda_1, \cdots, \lambda_p > 0, \lambda_{p+1}, \cdots, \lambda_r < 0$.

显然的, 若 $A \in M_{n\times n}(\mathbb{R})$ 是实对称矩阵, 对任意 $B \in M_{n\times n}(\mathbb{R})$, $B^T AB$ 也是实对称矩阵. 那么它的惯性指数是什么关系呢?

命题 7.50. 若 $B \in M_{n\times n}(\mathbb{R})$ 是可逆矩阵, 则 $B^T AB$ 的惯性指数和 A 的惯性指数相同.

证明. 设 $A, B^T AB$ 的正惯性指数分别为 p, q, 那么它们的负惯性指数分别为 $r-p, r-q$; 于是我们就是要证 $p = q$. 我们用反证法, 不妨设 $q > p$.

我们将 $A, B^T AB$ 分别正交对角化:

$$A = U\Lambda U^T, B^T AB = V\Lambda' V^T,$$

其中 $\Sigma = \mathrm{diag}(\lambda_1, \cdots, \lambda_p, \lambda_{p+1}, \cdots, \lambda_r, 0, \cdots)$, $\lambda_1, \cdots, \lambda_p > 0, \lambda_{p+1}, \cdots, \lambda_r < 0$;
$\Sigma' = \mathrm{diag}(\lambda_1', \cdots, \lambda_q', \lambda_{q+1}', \cdots, \lambda_r', 0, \cdots)$, $\lambda_1', \cdots, \lambda_q' > 0, \lambda_{q+1}', \cdots, \lambda_r' < 0$. 设

$$U = \begin{bmatrix} \alpha_1 & \cdots & \alpha_p & \alpha_{p+1} & \cdots & \alpha_n \end{bmatrix}, BV = \begin{bmatrix} \beta_1 & \cdots & \beta_q & \beta_{q+1} & \cdots & \beta_n \end{bmatrix}.$$

考虑向量组 $\beta_1, \cdots, \beta_q, \alpha_{p+1}, \cdots, \alpha_n$; 共 $q + (n-p) = n + (q-p) > n$ 个向量, 所以它们线性无关, 即存在不全为零的实数 $x_1, \cdots, x_q, y_{p+1}, \cdots, y_n$, 使得:

$$x_1\beta_1 + \cdots + x_q\beta_q + y_{p+1}\alpha_{p+1} + \cdots + y_n\alpha_n = 0.$$

令

$$
\begin{aligned}
\boldsymbol{x} &= x_1\beta_1 + \cdots + x_q\beta_q & &= -(y_{p+1}\alpha_{p+1} + \cdots + y_n\alpha_n) \\
&= BV\mathrm{diag}(x_1, \cdots, x_q, 0, \cdots, 0) & &= -U\mathrm{diag}(0, \cdots, 0, y_{p+1}, \cdots, y_n).
\end{aligned}
$$

于是 $\boldsymbol{x} \neq 0$ 且:

- $\boldsymbol{x}^T A\boldsymbol{x} = \mathrm{diag}(x_1, \cdots, x_q, 0, \cdots, 0)(BV^T)ABV\mathrm{diag}(x_1, \cdots, x_q, 0, \cdots, 0) = \lambda_1' x_1^2 + \cdots + \lambda_q x_q^2 > 0$;
- $\boldsymbol{x}^T A\boldsymbol{x} = \mathrm{diag}(0, \cdots, 0, y_{p+1}, \cdots, y_n)U^T AU\mathrm{diag}(0, \cdots, 0, y_{p+1}, \cdots, y_n) = \lambda_{p+1} y_{p+1}^2 + \cdots + \lambda_r y_r^2 < 0$.

矛盾. 因此反证假设不成立, 从而 $q = p$. 证毕. □

定义 7.51. 设 $A, B \in M_{n\times n}(\mathbb{R})$, 我们称 A, B 合同, 如果存在可逆矩阵 $C \in M_{n\times n}(\mathbb{R})$ 使得 $B = C^T AC$.

注释 7.52. 在这个定义之下, 上面的命题就是在说 "合同不改变实对称矩阵的惯性指数".

7.3.3 矩阵的 LU 分解

我们以前说过矩阵的 LU 分解; 当时谈论的一直是若矩阵有 LU 分解, 然后如何如何. 我们现在给出矩阵存在 LU 分解的充分必要条件.

定理 7.53. 设 $A \in M_{n\times n}(\mathbb{R})$ 是可逆矩阵. 那么 A 存在 LU 分解当且仅当 A 的所有顺序子矩阵都可逆 (等价地, A 的所有顺序主子式都非零).

证明. "$\Rightarrow$": 设 $A = LU$, 其中 L 是对角线为 1 的下三角矩阵, U 是对角线非零的上三角矩阵. 我们将 L 和 U 分块:

$$
L = \begin{bmatrix} L_k & L_1' \\ L_2' & L_3' \end{bmatrix}, \quad U = \begin{bmatrix} U_k & U_1' \\ U_2' & U_3' \end{bmatrix},
$$

其中 $L_k, U_k \in M_{k\times k}(\mathbb{R})$. L 下三角, U 上三角, 那么 $L_1' = 0, U_2' = 0$. 所以 $A = \begin{bmatrix} L_kU_k & L_kU_1' \\ L_2'U_k & L_3'U_3' \end{bmatrix}$, 那么 $A_k = L_kU_k$, 于是 $\det(A_k) = \det(U_k) \neq 0$.

"$\Leftarrow$": 现在假设 A 的所有顺序主子式都非零, 需要证明 A 有 LU 分解. 我们还是用归纳法. 证明过程和命题 (7.47) 如出一辙. $n = 1$ 时是平凡的. 假设阶数等于 $n-1$ 时结论成立, 我们来证明阶数等于 n 时的情形. 设

$$
A = (a_{ij}) = \begin{bmatrix} B & \alpha \\ \beta^T & a_{nn} \end{bmatrix},
$$

其中 $\alpha \in \mathbb{R}^{n-1}, B \in M_{(n-1)\times(n-1)}(\mathbb{R})$ 是 A 的第 $n-1$ 个顺序子矩阵, 那么 B 的第 k $(1 \le k \le n-1)$ 个顺序主子式等于 A 的第 k 个顺序主子式. 根据归纳假设, B 有 LU 分解, 于是存在对角线为 1 的下三角矩阵 $\widetilde{L}$ 和可逆上三角矩阵 $\widetilde{U}$, 使得 $B = \widetilde{L}\widetilde{U}$. 我们对 A 作分块消元:

$$\begin{bmatrix} I_{n-1} & 0 \\ -B^{-1}\beta^T & 1 \end{bmatrix} A \begin{bmatrix} I_{n-1} & -B^{-1}\alpha \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} B & 0 \\ 0 & a_{nn} - \beta^T B^{-1}\alpha \end{bmatrix}.$$

两边取行列式知 $\det(B)\,(a_{nn} - \beta^T B^{-1}\alpha) = \det(A) \neq 0$. 所以 $b = a_{nn} - \alpha^T B^{-1}\alpha \neq 0$, 令

$$L = \begin{bmatrix} I_{n-1} & 0 \\ B^{-1}\beta^T & 1 \end{bmatrix} \begin{bmatrix} \widetilde{L} & 0 \\ 0 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} \widetilde{U} & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} I_{n-1} & B^{-1}\alpha \\ 0 & 1 \end{bmatrix},$$

则有 $A = LU$. □

注释 7.54. 现在我们来考察一下“主元”. 所谓的主元, 就是在做初等行变换时 (一行的倍数加往另一行和交换两行), 每一列第一次出现的非零元. 假设 A 有 LU 分解, 那么做初等行变换时就没有交换两行. 且 A 的主元就是 U 的对角线元素. 所以一个有 LU 分解的方阵, 它的行列式等于所有主元的乘积.

由于 A 的第 k 个顺序子矩阵可逆, 因此若我们将 A 分块: $A = \begin{bmatrix} A_k & B \\ C & D \end{bmatrix}$ 那么

$$\begin{bmatrix} I_k & 0 \\ -CA_k^{-1} & I_{n-k} \end{bmatrix} A = \begin{bmatrix} A_k & B \\ 0 & D - CA_k^{-1}B \end{bmatrix}.$$

左乘矩阵 $\begin{bmatrix} I_k & 0 \\ -CA_k^{-1} & I_{n-k} \end{bmatrix}$ 是对 A 作了一系列的”将某一行的倍数加到另一行”上的初等行变换; 所以 A 的前 k 个主元都出现在对 A_k 作初等行变换的过程中. 所以 A 的前 k 个主元的乘积就等于 $\det(A_k)$, 所以 A 的第 k 个主元

$$a_k = \frac{\det(A_k)}{\det(A_{k-1})}.$$

7.4 不可对角化矩阵举例

设 $A \in M_{n\times n}(\mathbb{R})$. 若 A 的特征多项式 $\det(A - \lambda I_n)$ 有非实数解, 那么它显然不可对角化. 这是不可对角化的” 平凡情形”. 下面我们研究非平凡的不可对角化 2, 3 阶方阵.

7.4.1 不可对角化的 2 阶方阵.

设 $A \in M_{2\times 2}(\mathbb{R})$, 它的特征多项式的根都是实数, 但是不可对角化. 由于它有两个实根, 因此它的根相等: 因为若两个实根不等, 那么根据推论 (7.16), A 可对角化. 所以 A 的特征多项

式等于 $(\lambda_0-\lambda)^2$. 特征值 λ_0 的代数重数等于 2, 所以它的几何重数等于 1. 取 $0\neq\alpha\in V_{\lambda_0}$, 再取 $0\neq\beta\in\mathbb{R}^2$ 使得 α,β 构成 $\mathbb{R}^2$ 的一组基. 就有 $P=\begin{bmatrix}\alpha & \beta\end{bmatrix}\in M_{2\times 2}(\mathbb{R})$ 是可逆矩阵. 我们假设 $A\beta=a\alpha+b\beta$. 由假设, β 不是特征向量, 所以 $a\neq 0$. 于是我们有

$$AP=\begin{bmatrix}A\alpha & A\beta\end{bmatrix}=\begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}\lambda_0 & a\\ 0 & b\end{bmatrix}=PC$$

其中 $C=\begin{bmatrix}\lambda_0 & a\\ 0 & b\end{bmatrix}$. 因为 $P^{-1}AP=C$, 所以 A,C 有相同的特征多项式, 而 C 的特征多项式等于 $(\lambda_0-\lambda)(b-\lambda)$, 所以 $b=\lambda_0$. 至此, 我们证明了:

命题 7.55. *若 $A\in M_{2\times 2}(\mathbb{R})$ 是有两个实数特征值的不可对角化矩阵, 则存在可逆矩阵 $P\in M_{2\times 2}(\mathbb{R})$ 和 $a\neq 0$ 使得 $P^{-1}AP=\begin{bmatrix}\lambda_0 & a\\ 0 & \lambda_0\end{bmatrix}$.*

7.4.2 不可对角化的 3 阶方阵.

设 $A\in M_{3\times 3}(\mathbb{R})$, 它的特征多项式的根都是实数, 但是不可对角化. 由于它有三个实根, 同样根据推论 (7.16) 可知有重根, 那么就有以下两种情形:

- A 的特征多项式有一个单根, 一个二重根;
- A 的特征多项式有一个三重根.

我们考虑第一种情形. 假设 λ_1 是单根, λ_2 $(\neq\lambda_1)$ 是二重根. 那么同样由不可对角化知 λ_2 的几何重数等于 1. 设 α_1,α_2 分别是属于特征值 λ_1,λ_2 的特征向量, 再取 $0\neq\beta\in\mathbb{R}^2$ 使得 α_1,α_2,β 构成 $\mathbb{R}^3$ 的一组基. 就有 $P=\begin{bmatrix}\alpha_1 & \alpha_2 & \beta\end{bmatrix}\in M_{3\times 3}(\mathbb{R})$ 是可逆矩阵. 我们假设 $A\beta=a_1\alpha_1+a_2\alpha_2+b\beta$. 于是我们有

$$AP=\begin{bmatrix}A\alpha_1 & A\alpha_2 & A\beta\end{bmatrix}=\begin{bmatrix}\alpha_1 & \alpha_2 & \beta\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & a_1\\ 0 & \lambda_2 & a_2\\ 0 & 0 & b\end{bmatrix}=PC$$

其中 $C=\begin{bmatrix}\lambda_1 & 0 & a_1\\ 0 & \lambda_2 & a_2\\ 0 & 0 & b\end{bmatrix}$. 因为 $P^{-1}AP=C$, 所以 A,C 有相同的特征多项式, 而 C 的特征多项式等于 $(\lambda_1-\lambda)(\lambda_2-\lambda)(b-\lambda)$, 所以 $b=\lambda_2$. 因此

$$P^{-1}AP=C=\begin{bmatrix}\lambda_1 & 0 & a_1\\ 0 & \lambda_2 & a_2\\ 0 & 0 & b\end{bmatrix}.$$

引理 7.56. *存在可逆矩阵 Q 使得*

$$QCQ^{-1} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & a_2 \\ 0 & 0 & b \end{bmatrix}.$$

证明. 我们用待定系数法. 观察一下 C 是上三角矩阵, 所以我们希望找最简单的 Q. 我们尝试一下, 设 $Q = \begin{bmatrix} 1 & \boldsymbol{x}^T \\ 0 & I_2 \end{bmatrix}$, 其中 $\boldsymbol{x} \in \mathbb{R}^2$. 那么 $Q^{-1} = \begin{bmatrix} 1 & -\boldsymbol{x}^T \\ 0 & I_2 \end{bmatrix}$. 直接计算

$$QCQ^{-1} = \begin{bmatrix} \lambda_1 & -\lambda_1\boldsymbol{x}^T + \boldsymbol{w}^T + \boldsymbol{x}^T D \\ 0 & D \end{bmatrix}, \text{ 其中}\boldsymbol{w} = \begin{bmatrix} 0 \\ a_1 \end{bmatrix}, D = \begin{bmatrix} \lambda_2 & a_2 \\ 0 & \lambda_2 \end{bmatrix},$$

所以我们只需要证明关于 $\boldsymbol{x}$ 的方程 $-\lambda_1\boldsymbol{x}^T + \boldsymbol{w}^T + \boldsymbol{x}^T D = \begin{bmatrix} 0 & 0 \end{bmatrix}$ 有解, 等价地: $(D^T - \lambda_1 I_2)\boldsymbol{x} = -\boldsymbol{w}$ 有解. 而 $(D^T - \lambda_1 I_2) = \begin{bmatrix} \lambda_2 - \lambda_1 & 0 \\ a_2 & \lambda_2 - \lambda_1 \end{bmatrix}$ 且 $\lambda_1 \neq \lambda_2$, 所以 $D^T - \lambda_1 I_2$ 可逆, 那么所需方程有解. 证毕. □

将引理中的 Q 代入, 那么 $(PQ)^{-1}A(PQ) = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & a_2 \\ 0 & 0 & \lambda_2 \end{bmatrix}$. 至此, 我们证明了:

命题 7.57. *若 $A \in M_{3\times3}(\mathbb{R})$ 不可对角化矩阵, 它的特征多项式有一个单实数根 λ_1 和另一个不等二重实数根 λ_2, 则存在可逆矩阵 $P \in M_{3\times3}(\mathbb{R})$ 和 $a \neq 0$ 使得 $P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & a \\ 0 & 0 & \lambda_2 \end{bmatrix}$.*

现在考虑第三种情形: A 的特征多项式有一个三重实数根 λ_0. 这时候又有两个情形: λ_0 的几何重数等于 1 或者 2.

若 λ_0 的几何重数等于 1, 取 $0 \neq \alpha \in V_{\lambda_0}$, 以及 β_1, β_2, 使得 α, β_1, β_2 是 $\mathbb{R}^3$ 的一组基. 令 $P_1 = \begin{bmatrix} \alpha & \beta_1 & \beta_2 \end{bmatrix}$, 那么

$$P_1^{-1}AP = \begin{bmatrix} \lambda_0 & a_1 & b_1 \\ 0 & a_2 & b_2 \\ 0 & a_3 & b_3 \end{bmatrix} = B$$

令 $C = \begin{bmatrix} a_2 & b_2 \\ a_3 & b_3 \end{bmatrix}$, 那么 $\det(B - \lambda I_3) = (\lambda_0 - \lambda)\det(C - \lambda I_2)$. 所以 C 的特征多项式等于 $(\lambda_0 - \lambda)^2$, 根据我们已知的 2 阶矩阵的情形知存在可逆矩阵 $P_2 \in M_{2\times2}(\mathbb{R})$, 使得 $P_2CP_2 =$

$\begin{bmatrix} \lambda_0 & c \\ 0 & \lambda_0 \end{bmatrix}$ (若 $c = 0$ 则 C 可对角化. 若 $c \neq 0$ 则 C 不可对角化). 令 $P_3 = P_1 \begin{bmatrix} 1 & 0 \\ 0 & P_2 \end{bmatrix}$, 那么

$$P_3^{-1} A P_3 = \begin{bmatrix} \lambda_0 & a & b \\ 0 & \lambda_0 & c \\ 0 & 0 & \lambda_0 \end{bmatrix}.$$

由于 $\dim V_{\lambda_0} = 1$, 则 $\operatorname{rank}(A - \lambda_0 I_3) = 2$, 因此 $ac \neq 0$. 再令 $P_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{b}{a} \\ 0 & 0 & 1 \end{bmatrix}$, 那么

$$P_4 \begin{bmatrix} \lambda_0 & a & b \\ 0 & \lambda_0 & c \\ 0 & 0 & \lambda_0 \end{bmatrix} P_4^{-1} = \begin{bmatrix} \lambda_0 & a & 0 \\ 0 & \lambda_0 & c \\ 0 & 0 & \lambda_0 \end{bmatrix}$$

所以令 $P = P_3 P_4$, 就有

$$P^{-1} A P = \begin{bmatrix} \lambda_0 & a & 0 \\ 0 & \lambda_0 & c \\ 0 & 0 & \lambda_0 \end{bmatrix}.$$

若 λ_0 的几何重数等于 2, 取 V_{λ_0} 的一组基 α_1, α_2, 以及 β, 使得 $\alpha_1, \alpha_2, \beta$ 是 $\mathbb{R}^3$ 的一组基. $P_1 = \begin{bmatrix} \alpha_1 & \alpha_2 & \beta \end{bmatrix}$, 那么

$$P_1^{-1} A P_1 = \begin{bmatrix} \lambda_0 & 0 & a \\ 0 & \lambda_0 & b \\ 0 & 0 & c \end{bmatrix}.$$

老规矩, 考虑特征多项式, 我们有 $c = \lambda_0$. 由于 A 不可对角化, 那么 a, b 不全为零. 若 $a \neq 0$, 令 $P_2 = \begin{bmatrix} 0 & a & 0 \\ 1 & b & 0 \\ 0 & 0 & 1 \end{bmatrix}$, 则

$$P_2 \begin{bmatrix} \lambda_0 & 0 & a \\ 0 & \lambda_0 & b \\ 0 & 0 & \lambda_0 \end{bmatrix} P_2^{-1} = \begin{bmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_0 & a + b^2 \\ 0 & 0 & \lambda_2 \end{bmatrix}.$$

所以令 $P = P_1 P_2$, 就有

$$P^{-1} A P = \begin{bmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_0 & a + b^2 \\ 0 & 0 & \lambda_0 \end{bmatrix}.$$

总之,

命题 7.58. 若 $A \in M_{3\times 3}(\mathbb{R})$ 不可对角化矩阵, 它的特征多项式有一个三重实数根 λ_0. 则存在可逆矩阵 $P \in M_{3\times 3}(\mathbb{R})$ 和 $a, b \neq 0$ 使得

- λ_0 的几何重数等于 1 时, $P^{-1}AP = \begin{bmatrix} \lambda_0 & a & 0 \\ 0 & \lambda_0 & b \\ 0 & 0 & \lambda_0 \end{bmatrix}$. P 的第一列是 A 的特征向量.
- λ_0 的几何重数等于 2 时, $P^{-1}AP = \begin{bmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_0 & b \\ 0 & 0 & \lambda_0 \end{bmatrix}$. P 的第一二列是 A 的特征向量.

第八章　奇异值分解

奇异值分解的基本原理和对角化的想法一致: 能否把一个矩阵写成对角矩阵. 那么这个答案显然是不能: 首先方阵就未必可以对角化; 其次如果不是方阵, 天然地不能对角化. 那么问题来了: 退而求次的话, 应该是什么形式?

我们回顾一下方阵 A 可对角化的意思是存在可逆矩阵 P 和对角矩阵 D 使得 $A = P^{-1}DP$. 现在由于 A 未必是方阵, 那么显然无法要求用一个可逆矩阵去共轭. 那么一个妥协就是是否存在两个可逆矩阵 P_1, P_2 和对角矩阵 D, 使得 $A = P_1 D P_2$? 这也办不到: 设 $A \in M_{m\times n}(\mathbb{R})$, 但是 $m \neq n$, 那么 $P_1 \in M_{m\times m}(\mathbb{R}), P_2 \in M_{n\times n}(\mathbb{R}), D \in M_{m\times n}(\mathbb{R})$, 所以 D 不可能是方阵, 更谈不上对角阵. 当然我们还可以再妥协, D 尽量 "对角", 例如 $D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. 这个时候问题就变得太简单了, 作初等行列变换就可以知道了, 同学们作为练习自行证明下一个命题:

命题 8.1. 设 $A \in M_{m\times n}(\mathbb{R})$, $\operatorname{rank}(A) = r$, 则存在可逆矩阵 $P_1 \in M_{m\times m}(\mathbb{R}), P_2 \in M_{n\times n}(\mathbb{R})$, 使得

$$A = P_1 \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} P_2.$$

那怎么办? 这说明我们妥协得太多, 所以要额外提一些要求, 例如: 不能仅仅要求 P_1, P_2 是可逆矩阵, 能不能要求它们是正交矩阵呢?

能! 这就是奇异值分解的结果:

定理 8.2 (奇异值分解). 设 $A \in M_{m\times n}(\mathbb{R})$, $\operatorname{rank}(A) = r$, 则存在正交矩阵 $U \in M_{m\times m}(\mathbb{R}), V \in M_{n\times n}(\mathbb{R})$, 以及可逆对角矩阵 $D_r \in M_{n\times n}(\mathbb{R})$ 使得

$$A = U \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} V^T.$$

注释 8.3. 为什么叫 "奇异值" 分解呢? 如果能懂线性映射, 那么就接着看这个注释. 如果不懂线性映射, 那么跳过这个注释. 首先在方阵的情形, 特征值是那些 $\lambda \in \mathbb{R}$ 使得 $A - \lambda I_n$ 是奇异

矩阵. 所以特征值就是"奇异值". 其次, 在非方阵的情形, 考察映射

$$C(A^T) \longrightarrow C(A) \qquad x \longmapsto Ax$$

由于 $\dim C(A^T) = \dim C(A) = \operatorname{rank}(A) = r$, 那么在合适的基之下, 这个矩阵可以由一个 r 阶方阵来表达. D_r 中的对角元就是在这个 r 阶方阵的特征值—-"奇异值".

在证明奇异值分解之前, 我们先看看它的实际用途.

例题 8.4. 计算机存储图片, 比如说一张一万像素的图片, 实际上是一个 100×100 的格, 每个格子用一个数字指示颜色, 所以实际上这张图片就是个 100×100 的矩阵. 我们存储这张图片, 需要的存储空间是 100×100. 那么能不能减少一些存储空间呢? 这就是图像压缩. 其原理就是: 将这个矩阵做奇异值分解, 然后只取其中最大的奇异值, 例如 $D = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{100})$, $\lambda_1 = \max\{\lambda_i\}$. 那我们只存储 $U\operatorname{diag}(\lambda_1, 0, \cdots, 0)V$. 这个矩阵秩为 1, 所以可以写成 $\alpha\beta^T$ 的形式, 所以实际上我们只需要存储 α, β, 这样存储空间只需要 200. 当然, 图像压缩肯定是损失了信息的; 我们也可以取多几个奇异值, 这样得到的压缩图像会与原图像更吻合.

所以问题就变成了我们如何证明定理 (8.2) 以及如何实现奇异值分解. 定理的证明当然给了一个实现方式, 但是往往不是实际使用中最好的方式; 所以这个有赖于我们同学们走上工作岗位不断研究.

我们还是对结果做一番探索. 假设结论正确, 那么 $A = U\begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}V$ 等价于 $AV = U\begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}$. U 可逆意味着 $\operatorname{rank}\left(U\begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}\right) = r$, 而 V 可逆意味着

$$C(A) = C(AV) = C\left(U\begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}\right) = \operatorname{Span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r).$$

所以 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r$ 是 A 的列向量空间的一组标准正交基. 取转置, 同理可得 V 的前 r 行是 A 的行向量空间的一组标准正交基.

另一方面, $A^TA = V\begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}V^T$, 这意味着我们用 V 将 A 做了正交对角化! 如果 $D_r = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r)$, 那么 $\lambda_1^2, \lambda_2^2, \cdots, \lambda_r^2, 0$ 是 A^TA 的所有特征值, $\boldsymbol{v}_i$ $(1 \le i \le r)$ 是属于特征值 λ_i^2 的特征向量, $\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_m$ 是属于特征值 0 的特征向量, 所以 $\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_m$ 是 $N(A^TA) = N(A)$ 的一组标准正交基. 我们也可以再研究一遍 AA^T.

注释 8.5. 也许有同学想问, 为什么会想到要考虑 A^TA 呢? 首先, 我们拿到一个问题, 总是要多方面进行尝试. 其次呢, A^TA 以前在做投影的时候就研究过, 因此有一些成熟的经验可以借鉴.

定理 (8.2) 的证明. 考虑 $B = A^T A \in M_{m\times m}$ 是实对称矩阵, 根据定理 (), 它一定可以正交对角化. 所以存在 $\mathbb{R}^n$ 的一组单位正交基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 是 A^A 的特征向量. 设 $A^T A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$. 我们知道 $A^T A$ 的非零特征值的个数 (计重数) 等于 $\operatorname{rank}(A^T A) = \operatorname{rank}(A) = r$. 所以我们假设 $\lambda_1, \lambda_2, \cdots, \lambda_r \neq 0$. 对 $1 \le i \le r$, 令 $\boldsymbol{u}_i' = A\boldsymbol{v}_i \in C(A) \subseteq \mathbb{R}^m$. 首先 $A^T \boldsymbol{u}_i' = A^T A \boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \neq 0$, 所以 $\boldsymbol{u}_i'$ 均非零. 其次, 注意到当 $1 \le i \neq j \le r$ 时,

$$\left\langle \boldsymbol{u}_i', \boldsymbol{u}_j' \right\rangle = \boldsymbol{u}_i'^T \boldsymbol{u}_j' = \boldsymbol{v}_i^T A^T A \boldsymbol{v}_j = \lambda_i \left\langle \boldsymbol{v}_i, \boldsymbol{v}_j \right\rangle = 0;$$

所以 $\boldsymbol{u}_1', \cdots, \boldsymbol{u}_r'$ 两两互相正交, 因此它们线性无关, 这就构成了 $C(A) \subseteq \mathbb{R}^m$ 的一组正交基. 而 $\|\boldsymbol{u}_i'\|^2 = \langle \boldsymbol{u}_i', \boldsymbol{u}_i' \rangle = \lambda_i \langle \boldsymbol{v}_i \rangle^2 = \lambda_i$, 这一方面说明 $\lambda_i > 0$, 另一方面 $\boldsymbol{u}_i'$ 并不是单位向量, 所以我们将它们单位化: 令 $\boldsymbol{u}_i = \dfrac{\boldsymbol{u}_i'}{\|\boldsymbol{u}_i'\|} = \dfrac{\boldsymbol{u}_i'}{\sqrt{\lambda_i}}$, 它们就是 $C(A)$ 的一组标准正交基. 我们再任取 $C(A)^\perp = N(A^T)$ 的一组标准正交基: $\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m$, 这样, 我们就得到了 $\mathbb{R}^m$ 的一组标准正交基. 以及两个正交矩阵:

$$V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} \in M_{n\times n}(\mathbb{R}), \quad U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m \end{bmatrix} \in M_{m\times m}(\mathbb{R}).$$

$$\begin{aligned} AV &= A\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} \\ &= \begin{bmatrix} A\boldsymbol{v}_1 & A\boldsymbol{v}_2 & \cdots & A\boldsymbol{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{\lambda_1}\boldsymbol{u}_1 & \sqrt{\lambda_2}\boldsymbol{u}_2 & \cdots & \sqrt{\lambda_r}\boldsymbol{u}_r & 0 & \cdots & 0 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_r} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix} \end{aligned}$$

因此取

$$D_r = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_r} \end{bmatrix},$$

就有 $A = U\begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} V^T$. 这就证明了定理 (8.2). □

注释 8.6. 假设 $A \in M_{n\times n}(\mathbb{R})$, 那么奇异值分解的几何意义是: $A : \mathbb{R}^n \to \mathbb{R}^n, \boldsymbol{x} \mapsto A\boldsymbol{x}$ 可以分

解为：一个旋转 + 坐标轴的拉伸 + 一个旋转. 若 A 不可逆, 那么肯定有若干坐标轴被“彻底拍扁”了：缩成 0 了.

注释 8.7. 我们可以比较一下实对称矩阵对角化和任意矩阵的奇异值分解：设 $S, A \in M_{n\times n}(\mathbb{R})$, $S = S^T$.

- 存在对角矩阵 D 和正交矩阵 Q, 使得 $S = QDQ^T$;
- 存在对角矩阵 D 和正交矩阵 U, V, 使得 $S = UDV^T$.

8.0.1 矩阵范数

设 $A \in M_{m\times n}(\mathbb{R})$, 定义

$$\|A\| := \sup\left\{\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \,\middle|\, 0 \neq \boldsymbol{x} \in \mathbb{R}^n\right\}.$$

设 A 的奇异值分解为 $A = UDV^T$, 其中 $D = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r, 0, \cdots)$, $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r$, 那么:

命题 8.8.

$$\|A\| = \sigma_1.$$

证明. 注意到, 对任意 n 阶正交矩阵 Q 以及 $\boldsymbol{x} \in \mathbb{R}^n$, 有 $\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$, 所以

$$\|A\boldsymbol{x}\| = \|DV^T\boldsymbol{x}\|, \|V^T\boldsymbol{x}\| = \|\boldsymbol{x}\|.$$

从而

$$\|A\| = \sup_{0\neq \boldsymbol{x}\in\mathbb{R}^n}\left\{\frac{\|DV^T\boldsymbol{x}\|}{\|V^T\boldsymbol{x}\|}\right\} \overset{V^T\boldsymbol{x}\to\boldsymbol{y}}{=} \sup_{0\neq \boldsymbol{y}\in\mathbb{R}^n}\left\{\frac{\|D\boldsymbol{y}\|}{\|\boldsymbol{y}\|}\right\}.$$

设 $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$, 则

$$\|D\boldsymbol{y}\| = \left(\sigma_1^2 y_1^2 + \lambda_2^2 y_2^2 + \cdots + \sigma_r y_r^2\right)^{1/2} \leq \sigma_1 \left(y_1^2 + y_2^2 + \cdots + y_n^2\right)^{1/2} = \sigma_1\|\boldsymbol{y}\|,$$

所以 $\|A\| \leq \sigma_1$. 另一方面, 取 $\boldsymbol{y} = \boldsymbol{e}_1$, 则 $D\boldsymbol{y} = \sigma_1\boldsymbol{e}_1$, 于是就有

$$\|A\| \geq \frac{\|D\boldsymbol{e}_1\|}{\|\boldsymbol{e}_1\|} = \sigma_1;$$

综合起来就有 $\|A\| = \sigma_1$. □

推论 8.9.

$$\|A\| = \|A^T\|.$$

直接利用奇异值分解和上述命题.

注释 8.10. 结合几何意义, $\|A\|$ 是 A 沿某个方向拉伸的最大的倍数.

例题 8.11. 设 $\boldsymbol{v} \in \mathbb{R}^n$ 是个列向量, 我们也可以把它看成是 $n \times 1$ 矩阵. 一方面, 我们用内积定义过它的长度; 另一方面, 有矩阵范数. 那么这两者是什么关系呢; 很简单, 是相等的. 请直接验证.

另一方面, 设 $\boldsymbol{w}$ 是个 n 维行向量; 那么 $\boldsymbol{w}$ 的矩阵范数等于 $\boldsymbol{w}^T$ 的矩阵范数, 也等于它由内积定义的长度. 我们还可以直接证明; 为了不引起混淆我们暂且记用内积定义的向量长度为 $\|\cdot\|_2$: 对 $\boldsymbol{x} \in \mathbb{R}^n, \|\boldsymbol{x}\|_2 = 1$, 则

$$\|\boldsymbol{w}\boldsymbol{x}\|_2 = |\boldsymbol{w}^T \cdot \boldsymbol{x}| \leq \|\boldsymbol{w}^T\|_2 \|\boldsymbol{x}\|_2 = \|\boldsymbol{w}^T\|_2.$$

所以

$$\|\boldsymbol{w}\| = \sup_{\|\boldsymbol{x}\|_2 = 1} \|\boldsymbol{w}\boldsymbol{x}\|_2 \leq \|\boldsymbol{w}^T\|_2.$$

另一方面, 取 $\boldsymbol{x} = \dfrac{\boldsymbol{w}^T}{\|\boldsymbol{w}^T\|_2}$, 则 $\|\boldsymbol{w}\boldsymbol{x}\|_2 = \|\boldsymbol{w}^T\|_2$, 所以 $\|\boldsymbol{w}\| \geq \|\boldsymbol{w}\boldsymbol{x}\| = \|\boldsymbol{w}^T\|_2$. 所以只能有 $\|\boldsymbol{w}\| = \|\boldsymbol{w}^T\|_2$.

既然向量看成矩阵的矩阵范数和用内积定义的长度一致, 往后我们就不再区分它们.

命题 8.12. 矩阵范数满足:

1. $A \in M_{m\times n}(\mathbb{R}), B \in M_{n\times k}(\mathbb{R})$, 则 $\|AB\| \leq \|A\|\|B\|$;

2. $A \in M_{m\times n}(\mathbb{R}), B \in M_{m\times n}(\mathbb{R})$, 则 $\|A+B\| \leq \|A\| + \|B\|$.

证明. 对任意 $0 \neq \boldsymbol{x} \in \mathbb{R}^k$, 如果 $B\boldsymbol{x} = 0$, 则 $\dfrac{\|AB\boldsymbol{x}\|}{\boldsymbol{x}} = 0 \leq \|A\|\|B\|$. 如果 $B\boldsymbol{x} \neq 0$, 则

$$\frac{\|AB\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \frac{\|AB\boldsymbol{x}\|}{\|B\boldsymbol{x}\|} \cdot \frac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|}.$$

而 $\dfrac{\|AB\boldsymbol{x}\|}{\|B\boldsymbol{x}\|} \leq \|A\|, \dfrac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \|B\|$, 所以 $\dfrac{\|AB\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \|A\|\|B\|$. 总而言之, 对任意 $\boldsymbol{x} \in \mathbb{R}^n$, 都有 $\dfrac{\|AB\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \|A\|\|B\|$. 这就证明了第 1 条.

第 2 条由向量长度的三角不等式可得, 证明如下: 对任意 $0 \neq \boldsymbol{x} \in \mathbb{R}^n$,

$$\frac{\|(A+B)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \frac{\|A\boldsymbol{x} + B\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \frac{\|A\boldsymbol{x}\| + \|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \|A\| + \|B\|.$$

□

我们还是假设 $A \in M_{m\times n}(\mathbb{R})$. 对任意 $1 \le k \le r$, 定义

$$A_k = U\Sigma_k V^T, \Sigma_k = \begin{bmatrix} D_k & 0 \\ 0 & 0 \end{bmatrix}, D_k = \mathrm{diag}(\sigma_1, \cdots, \sigma_k, 0, \cdots, 0).$$

命题 8.13 (Eckart-Young-Mirsky 定理). 对任意的 $B \in M_{m\times n}(\mathbb{R}), \mathrm{rank}(B) = k$, 则

$$\|A - B\| \ge \|A - A_k\| = \sigma_{k+1}.$$

第二个等号是明显的, 前一个不等号的证明关键在于取

$$0 \neq \boldsymbol{x} \in N(B) \cap C(A_{k+1}^T) \subseteq \mathbb{R}^n.$$

证明. 由维数公式:

$$\begin{aligned} \dim(N(B) \cap C(A_{k+1}^T)) &= \left(\dim N(B) + \dim C(A_{k+1}^T)\right) - \dim(N(B) + C(A_{k+1}^T)) \\ &\ge (n-k) + (k+1) - n = 1, \end{aligned}$$

所以我们可以取 $0 \neq \boldsymbol{x} \in N(B) \cap C(A_{k+1}^T) \subseteq \mathbb{R}^n$. 将 $\boldsymbol{x}$ 单位化, 我们假设 $\|\boldsymbol{x}\| = 1$. 由于 $\boldsymbol{x} \in C(A_{k+1}^T)$, 我们设 $\boldsymbol{x} = x_1\boldsymbol{v}_1 + \cdots + x_{k+1}\boldsymbol{v}_{k+1}$, 其中 $V = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix}$; 于是 $\|\boldsymbol{x}\| = \left(x_1^2 + \cdots + x_{k+1}^2\right)^{1/2}$. 那么

$$\|(A-B)\boldsymbol{x}\| = \|A\boldsymbol{x}\| = \|U\Sigma V^T\boldsymbol{x}\| = \left(\sigma_1^2 x_1^2 + \cdots \sigma_{k+1}^2 x_{k+1}^2\right)^{1/2} \ge \sigma_{k+1}\left(x_1^2 + \cdots x_{k+1}^2\right)^{1/2} = \sigma_{k+1}.$$

□

以下选读.

我们以前在习题课内学过另外两个范数:

$$\begin{aligned} \|A\|_1 &= \left(\sum_{i,j} a_{ij}^2\right)^{1/2} \\ \|A\|_2 &= \sup_{i,j} |a_{ij}| \end{aligned}$$

设 $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, 那么

$$\|A\boldsymbol{x}\| = \left(\sum_i \left(\sum_j a_{ij}x_j\right)^2\right)^{1/2} \le \sqrt{n}\|A\|_2\|\boldsymbol{x}\|,$$

所以我们有

$$\|A\| \leq \sqrt{n}\|A\|_2$$

再由算术-几何平均不等式, 我们有

$$\left(\sum_j a_{ij}x_j\right)^2 \leq \sum_j a_{ij}^2 \sum_j x_j^2,$$

所以

$$\|A\boldsymbol{x}\| \leq \left(\sum_i \left(\sum_j a_{ij}^2 \sum_j x_j^2\right)\right)^{1/2} = \sqrt{n}\|\boldsymbol{x}\|\Big(\sum_{i,j} a_{ij}^2\Big)^{1/2} = \|A\|_1\sqrt{n}\|\boldsymbol{x}\|;$$

于是我们有

$$\|A\| \leq \sqrt{n}\|A\|_1.$$

另一方面, 显然地我们有

$$\|A\|_1 = \left(\sum_{i,j} a_{ij}^2\right)^{1/2} \leq \left(\sum_{i,j} \|A\|_2\right)^{1/2} = n\|A\|_2.$$

设 $|a_{k\ell}| = \max_{i,j}|a_{ij}|$, 那么根据定义 $\|A\|_2 = |a_{k\ell}|$. 我们取 $\boldsymbol{x}_1 = \boldsymbol{e}_\ell$, 那么

$$\|A\boldsymbol{e}_\ell\| = \left(\sum_i a_{i\ell}^2\right)^{1/2} \geq |a_{k\ell}| = \|A\|_2,$$

所以

$$\|A\| \geq \frac{\|A\boldsymbol{e}_\ell\|}{\|\boldsymbol{e}_\ell\|} \geq \|A\|_2.$$

所以我们有范数之间的不等式:

- $\frac{1}{n}\|A\|_1 \leq \|A\| \leq \sqrt{n}\|A\|_1$;
- $\|A\|_2 \leq \|A\| \leq \sqrt{n}\|A\|_2$
- $\|A\|_2 \leq \|A\|_1 \leq n\|A\|_2$

用数学术语来说, 这几个范数是互相“等价”的. 所谓的等价, 是我们在考虑所谓的矩阵函数 (就是自变量是矩阵的函数) 的连续性或者其它分析性质时, 我们可以自由地用其它范数取替换.

第九章 线性变换

这一章本质上没有新的内容; 我们可以将它看做对前面的知识的归纳整理, 抽象化和重新理解. 首先请大家仔细阅读, 努力理解第四章第一节的内容.

9.1 线性映射

我们不再回顾线性空间的定义, 但是回顾一下线性映射:

定义 9.1. 设 $f : V \to W$ 是一个线性映射, 如果它满足:

$$f(a\boldsymbol{v}_1 + b\boldsymbol{v}_2) = af(\boldsymbol{v}_1) + bf(\boldsymbol{v}_2); \forall a, b \in \mathbb{R}, \boldsymbol{v}_1, \boldsymbol{v}_2 \in V.$$

设 $f : V \to W$ 是一个映射, 我们定义:

定义 9.2.

- 0_W 的原像 (即 V 中取值为 0_W 的元素) 称为 f 的核, 记作 $\ker(f)$:

$$\ker(f) = f^{-1}(0_W) = \{\boldsymbol{v} \in V : f(\boldsymbol{v}) = 0_W\} \subseteq V.$$

  若 $\ker(f) = 0_V$, 我们称 f 是单射.

- W 的由所有 $f(\boldsymbol{v})(\boldsymbol{v} \in V)$ 组成的集合称为 f 的像, 记作 $\mathrm{Im}(f)$:

$$\mathrm{Im}(f) = f(V) = \{f(\boldsymbol{v}) : \boldsymbol{v} \in V\} = \{\boldsymbol{w} \in W : \exists \boldsymbol{v} \in V, s.t. \boldsymbol{w} = f(\boldsymbol{v})\}.$$

  若 $\mathrm{Im}(f) = W$, 我们称 f 是满射.

- 若线性映射 $f : V \to W$ 既是单射又是满射, 我们称 f 是同构.

我们对线性映射最熟悉的例子就是:

例题 9.3. 设 $A \in M_{m\times n}(\mathbb{R})$, 则

$$\varphi_A : \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad \boldsymbol{x} \longmapsto A\boldsymbol{x}$$

是线性映射. $\ker(\varphi_A) = N(A), \mathrm{Im}(\varphi_A) = C(A)$. 所以 φ_A 是同构当且仅当 $m = n$ 且 A 是可逆矩阵. 请根据定义, 把这个例子理解好!

这个例子无论如何一定一定要理解, 一定一定要理解, 一定一定要理解.

例题 9.4. 设 V 是所有的次数不大于 n 的多项式构成的集合. 请验证 V 是一个线性空间. 考虑

$$\partial : V \longrightarrow V \qquad f \longmapsto \frac{\mathrm{d}}{\mathrm{d}x} f$$

即: $\partial(f) := \dfrac{\mathrm{d}f}{\mathrm{d}x} = f'$. 请根据定义, 验证 ∂ 是一个线性映射, 求 $\ker(\partial)$ 和 $\mathrm{Im}(\partial)$.

那么我们怎么描述一个线性映射呢? 我们还是先回顾第一章中的一个命题:

命题 9.5. 映射 $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ 是一个线性映射, 当且仅当存在向量 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n \in \mathbb{R}^m$, 使得

$$\varphi \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

其中 A 是 $m \times n$ 矩阵 $\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$

我们回顾一下证明, 就会给我们一些提示:

$$\varphi \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\varphi(\boldsymbol{e}_1) + \cdots + x_n\varphi(\boldsymbol{e}_n),$$

我们令 $\boldsymbol{v}_i = \varphi(\boldsymbol{e}_i)$ 即可. 所以我们干了一件什么事情呢? 我们取了 $\mathbb{R}^n$ 的一组基!

现在我们考虑一般的线性映射: $f : V \to W$. 设 $n = \dim V, m = \dim W$. 那么我们可以选取

$$V \text{ 的一组基: } \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n; \quad W \text{ 的一组基 } \boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m.$$

那么 $f(\boldsymbol{v}_i)$ 是 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m$ 的线性组合, 所以存在唯一的一组 $a_{1i}, a_{2i}, \cdots, a_{mi}$ 使得

$$f(\boldsymbol{v}_i) = a_{1i}\boldsymbol{w}_1 + a_{2i}\boldsymbol{w}_2 + \cdots + a_{mi}\boldsymbol{w}_m.$$

组合在一起:

$$\begin{bmatrix} f(\boldsymbol{v}_1) & f(\boldsymbol{v}_2) & \cdots & f(\boldsymbol{v}_n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

我们记

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

那么

$$f\left(\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_m \end{bmatrix} A.$$

对任意 $\boldsymbol{v} = x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, 我们有

$$\begin{aligned} f(\boldsymbol{v}) &= x_1 f(\boldsymbol{v}_1) + x_2 f(\boldsymbol{v}_2) + \cdots + x_n f(\boldsymbol{v}_n) \\ &= \begin{bmatrix} f(\boldsymbol{v}_1) & f(\boldsymbol{v}_2) & \cdots & f(\boldsymbol{v}_n) \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} . \\ &= \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_m \end{bmatrix} A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

这个意思是, 一旦取定了 V 和 W 的基, 线性映射在这组基下的表现就由矩阵给出来了! 具体地说, 就是

命题 9.6. 设 V, W 分别是 n, m 维的线性空间, $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}, \{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m\}$ 分别是它们的一组基. 那么: $f : V \to W$ 是一个线性映射, 当且仅当存在唯一的矩阵 $A \in M_{m\times n}(\mathbb{R})$ 使得对任意 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T \in \mathbb{R}^n$, 我们有

$$f(x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n) = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_n \end{bmatrix} A\boldsymbol{x}.$$

例题 9.7. 令 $A=\begin{bmatrix}1&1&2\\1&2&2\\2&3&4\end{bmatrix}$, $\varphi:\mathbb{R}^3\to\mathbb{R}^3, \varphi(\boldsymbol{x}):=A\boldsymbol{x}$. 那么

$$N(A)=\mathbb{R}\begin{bmatrix}2\\0\\-1\end{bmatrix}, C(A^T)=\mathbb{R}\begin{bmatrix}0\\1\\0\end{bmatrix}+\mathbb{R}\begin{bmatrix}1\\0\\2\end{bmatrix}, C(A)=\mathbb{R}\begin{bmatrix}1\\1\\2\end{bmatrix}+\mathbb{R}\begin{bmatrix}1\\2\\3\end{bmatrix}$$

我们令 $V=C(A^T)$, $\boldsymbol{v}_1=\begin{bmatrix}0\\1\\0\end{bmatrix}, \boldsymbol{v}_2=\begin{bmatrix}1\\0\\2\end{bmatrix}$ 是它的一组基. 令 $W=C(A)$, $\boldsymbol{w}_1=\begin{bmatrix}1\\1\\2\end{bmatrix}, \boldsymbol{w}_2=\begin{bmatrix}1\\2\\3\end{bmatrix}$ 是它的一组基. 那么我们就有线性映射:

$$\varphi: V\longrightarrow W \qquad \varphi(\boldsymbol{v})=A\boldsymbol{v}$$

那么 φ 在基 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 和 $\boldsymbol{w}_1,\boldsymbol{w}_2$ 之下的矩阵是什么呢? 是不是 A?

证明. 显然不是: φ 在基 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 和 $\boldsymbol{w}_1,\boldsymbol{w}_2$ 之下的矩阵必须是个 2×2 矩阵, 但是 A 是 3 阶方阵.

$\varphi(\boldsymbol{v}_1)=\boldsymbol{w}_2, \varphi(\boldsymbol{v}_2)=3\boldsymbol{w}_1$, 所以

$$\begin{bmatrix}\varphi(\boldsymbol{v}_1) & \varphi(\boldsymbol{v}_2)\end{bmatrix}=\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2\end{bmatrix}\begin{bmatrix}0&3\\1&0\end{bmatrix},$$

所以 $\varphi: V\to W$ 在基 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 和 $\boldsymbol{w}_1,\boldsymbol{w}_2$ 之下的矩阵是 $\begin{bmatrix}0&3\\1&0\end{bmatrix}$. □

定义 9.8. 我们把命题 (9.6) 中的矩阵 A 称作线性映射 f 在基 $\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$ 和 $\{\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_m\}$ 之下的矩阵.

言下之意, 如果我们选取不同的基, 得到的矩阵是不一样的. 下面我们要讨论在不同的基之下 f 的矩阵的变换规律. 设 $\{\boldsymbol{v}_1',\boldsymbol{v}_2',\cdots,\boldsymbol{v}_n'\}$ 是 V 的另一组基, 那么存在唯一的可逆矩阵 $P\in M_{n\times n}(\mathbb{R})$, 使得

$$\begin{bmatrix}\boldsymbol{v}_1' & \boldsymbol{v}_2' & \cdots & \boldsymbol{v}_n'\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}P.$$

同样的设 $\{\boldsymbol{w}_1',\boldsymbol{w}_2',\cdots,\boldsymbol{w}_m'\}$ 是 W 的另一组基, 那么存在唯一的可逆矩阵 $Q\in M_{m\times m}(\mathbb{R})$, 使得

$$\begin{bmatrix}\boldsymbol{w}_1' & \boldsymbol{w}_2' & \cdots & \boldsymbol{w}_m'\end{bmatrix}=\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_m\end{bmatrix}Q.$$

命题 9.9. 记号如上. 那么 f 在基 $\begin{bmatrix} \boldsymbol{v}_1' & \boldsymbol{v}_2' & \cdots & \boldsymbol{v}_n' \end{bmatrix}, \begin{bmatrix} \boldsymbol{w}_1' & \boldsymbol{w}_2' & \cdots & \boldsymbol{w}_m' \end{bmatrix}$ 之下的矩阵为 $Q^{-1}AP$, 即对任意 $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T \in \mathbb{R}^n$, 我们有

$$f(x_1\boldsymbol{v}_1' + x_2\boldsymbol{v}_2' + \cdots + x_n\boldsymbol{v}_n') = \begin{bmatrix} \boldsymbol{w}_1' & \boldsymbol{w}_2' & \cdots & \boldsymbol{w}_n' \end{bmatrix} Q^{-1}AP\boldsymbol{x}.$$

这个直接计算, 建议同学们先自己做一遍. 下面给出证明.

证明. 设 $P = (p_{ij})$, 那么 $\boldsymbol{v}_j' = \sum_{i=1}^n p_{ij}\boldsymbol{v}_i$, 所以 $f(\boldsymbol{v}_j') = \sum_{i=1}^n p_{ij} f(\boldsymbol{v}_i)$, 写成矩阵形式就是

$$\begin{bmatrix} f(\boldsymbol{v}_1') & f(\boldsymbol{v}_2') & \cdots & f(\boldsymbol{v}_n') \end{bmatrix} = \begin{bmatrix} f(\boldsymbol{v}_1) & f(\boldsymbol{v}_2) & \cdots & f(\boldsymbol{v}_n) \end{bmatrix} P,$$

所以

$$\begin{aligned} f(x_1\boldsymbol{v}_1' + x_2\boldsymbol{v}_2' + \cdots + x_n\boldsymbol{v}_n') &= \begin{bmatrix} f(\boldsymbol{v}_1') & f(\boldsymbol{v}_2') & \cdots & f(\boldsymbol{v}_n') \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \\ &= \begin{bmatrix} f(\boldsymbol{v}_1) & f(\boldsymbol{v}_2) & \cdots & f(\boldsymbol{v}_n) \end{bmatrix} P\boldsymbol{x} \\ &= \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_m \end{bmatrix} AP\boldsymbol{x} \\ &= \begin{bmatrix} \boldsymbol{w}_1' & \boldsymbol{w}_2' & \cdots & \boldsymbol{w}_m' \end{bmatrix} Q^{-1}AP\boldsymbol{x} \end{aligned}$$

□

推论 9.10. 给定线性映射 $f: V \to V$, 设 f 在 V 的基 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 之下的矩阵为 A:

$$f(x_1\boldsymbol{v}_1 + x_2\boldsymbol{v}_2 + \cdots + x_n\boldsymbol{v}_n) = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} A\boldsymbol{x}.$$

那么如果我们取定 V 的另一组基 $\begin{bmatrix} \boldsymbol{v}_1' & \boldsymbol{v}_2' & \cdots & \boldsymbol{v}_n' \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} P$, 则 f 在这组基之下的矩阵是

$$P^{-1}AP.$$

所以, 如果我们得到 f 在基 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ 之下的矩阵 A, 知道 A 可以对角化, 即存在可逆矩阵 P 使得 $P^{-1}AP$ 是对角矩阵. 令 $\begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} P$; 那么 f 在基 $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n\}$ 之下的矩阵就是对角阵. 这说起来有些拗口, 整体思路如下:

- 选取线性空间的一组基;
- 写出线性映射在这组基下的矩阵 A;
- 通过矩阵 P 相似简化 A, 那么我们就得到了线性映射在另一组基之下的矩阵 $P^{-1}AP$.

线性映射还是那个线性映射, 但是通过不同的基, 体现出来的外在表现是复杂度不一样的. 所以矩阵对角化的含义就是: 选取合适的基, 使得线性映射的外在表现尽量简单.

现在假定我们有三个线性空间 U,V,W, 分别是 k,n,m 维的, $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k\}$, $\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n\}$, $\{\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_m\}$ 分别是它们的一组基. $f:U\to V, g:V\to W$ 是两个线性映射, 在先前取定的基之下的矩阵分别是 A,B. 那么 $g\circ f:U\to W$ 的矩阵是什么呢?

命题 9.11. 在上面的假设之下, $g\circ f$ 的矩阵是 BA.

证明. 设 $A=(a_{ij})_{\substack{1\le i\le n\\ 1\le j\le n}}, B=(b_{st})_{\substack{1\le s\le m\\ 1\le t\le n}}$, 则

$$f(\boldsymbol{u}_i)=a_{1i}\boldsymbol{v}_1+a_{2i}\boldsymbol{v}_2+\cdots+a_{ni}\boldsymbol{v}_n,\quad g(\boldsymbol{v}_j)=b_{ij}\boldsymbol{w}_1+b_{2j}\boldsymbol{w}_2+\cdots+b_{mj}\boldsymbol{w}_m.$$

那么

$$\begin{aligned}(g\circ f)(\boldsymbol{u}_i)=g(f(\boldsymbol{u}_i)) &= g(a_{1i}\boldsymbol{v}_1+a_{2i}\boldsymbol{v}_2+\cdots+a_{ni}\boldsymbol{v}_n)\\ &=\sum_{j=1}^{n}a_{ji}g(\boldsymbol{v}_j)\\ &=\sum_{j=1}^{n}a_{ji}\left(\sum_{\ell=1}^{m}b_{\ell j}\boldsymbol{w}_\ell\right)\\ &=\sum_{\ell=1}^{m}\left(\sum_{j=1}^{n}b_{\ell j}a_{ji}\right)\boldsymbol{w}_\ell\end{aligned}$$

注意到 $\sum\limits_{j=1}^{n}b_{\ell j}a_{ji}$ 正是矩阵 $C:=BA$ 的第 (ℓ,i) 分量, 那么上式就是 $(g\circ f)(\boldsymbol{u}_i)=\sum\limits_{\ell=1}^{m}c_{\ell i}\boldsymbol{w}_\ell$. 所以对任意 $\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_k\end{bmatrix}^T\in\mathbb{R}^k$, 我们有

$$\begin{aligned}(g\circ f)(x_1\boldsymbol{u}_1+x_2\boldsymbol{u}_2+\cdots+x_k\boldsymbol{u}_k) &= \begin{bmatrix}(g\circ f)(\boldsymbol{u}_1) & (g\circ f)(\boldsymbol{u}_2) & \cdots & (g\circ f)(\boldsymbol{u}_k)\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_k\end{bmatrix}\\ &=\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_m\end{bmatrix}\begin{bmatrix}c_{11} & c_{12} & \cdots & c_{1n}\\ c_{21} & c_{22} & \cdots & c_{2n}\\ \vdots & \vdots & & \vdots\\ c_{m1} & c_{m2} & \cdots & c_{mn}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_k\end{bmatrix}\end{aligned}$$

□

这个命题的计算看起来很繁杂, 但是毫无困难: 仅仅是利用定义朝着结论直线计算, 没有任何技巧, 没有任何需要额外思考的部分. 所以唯一需要的就是: 勇气和仔细.

这个命题说明了: **矩阵乘法对应了线性映射的复合**.

注释 9.12. 我们最后提醒一下, 映射复合的方向和矩阵乘法的方向: $U \xrightarrow{f} V \xrightarrow{g} W$ 的复合映射是 $g \circ f$, 不是 $f \circ g$. f 对应于 A, g 对应于 B, 复合映射对应于 BA, 而不是 AB.

最后我们解释一下, 为什么我们之前的所有课程都在讲 $\mathbb{R}^n$ 而不是一般的线性空间.

命题 9.13. 两个线性空间是同构的, 当且仅当它们的维数相等. 特别的, 任何一个 n 维线性空间都和 $\mathbb{R}^n$ 同构.

证明. 假设 V, W 是两个线性空间.

如果 $f : V \to W$ 是一个同构; 令 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 是 V 的一组基, 我们只需要证明 $f(\boldsymbol{v}_1), \cdots, f(\boldsymbol{v}_n)$ 是 W 一组基. 这等价于要证明: 它们线性相关, 且任何 W 中的元素都是它们的线性组合.

设 $x_1 f(\boldsymbol{v}_1) + \cdots + x_n f(\boldsymbol{v}_n) = 0$, 则 $f(x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n) = 0$. f 是单射, 所以 $x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n = 0$; 而 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 线性无关, 所以 $x_1 = \cdots = x_n = 0$. 这就证明了 $f(\boldsymbol{v}_1), \cdots, f(\boldsymbol{v}_n)$ 线性无关.

设 $\boldsymbol{w} \in W$. 因为 f 是满射, 所以存在 $\boldsymbol{v} \in V$ 使得 $\boldsymbol{w} = f(\boldsymbol{v})$. 而 $b\boldsymbol{v} = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n$, 所以

$$\boldsymbol{v} = f(\boldsymbol{v}) = f(x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n) = x_1 f(\boldsymbol{v}_1) + \cdots + x_n f(\boldsymbol{v}_n).$$

反之. 假设 V, W 的维数相等, 我们需要证明他们之间存在一个同构映射. 设 $\dim V = \dim W = n$, 我们取 V 的基 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 的基, 和 W 的基 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n$. 定义

$$f : V \to W; \quad f(x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n) = x_1\boldsymbol{w}_1 + \cdots + x_n\boldsymbol{w}_n.$$

这是一个良定义的映射, 因为 V 中的元素都可以唯一地写成 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 的线性组合. 显然是线性的. 于是我们需要证明 f 是单射满射. 设 $f(\boldsymbol{v}) = 0$, 由于 $\boldsymbol{v} = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n$, 则 $x_1\boldsymbol{w}_1 + \cdots + x_n\boldsymbol{w}_n = f(\boldsymbol{v}) = 0$. 由 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n$ 的线性无关性可知 $x_1 = \cdots = x_n = 0$, 所以 $\boldsymbol{v} = 0$; 即 f 是单射. 另一方面, 对任意的 $\boldsymbol{w} \in W$, 由于 $\boldsymbol{w} = x_1\boldsymbol{w}_1 + \cdots + x_n\boldsymbol{w}_n$, 我们令 $\boldsymbol{v} = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n$, 就有 $\boldsymbol{w} = f(\boldsymbol{v})$. 所以 f 是满射. □

正是因此, 我们可以把任意的一个线性空间, 取一组基之后, 就实现了它和某个 $\mathbb{R}^n$ 的同构; 进而关于这个线性空间的研究就转化成了关于具体的 $\mathbb{R}^n$ 的研究了.

注释 9.14. 请注意, 我们说两个线性空间 V, W 同构, 是说存在同构映射 $f : V \to W$; 而不是说任何的线性映射 $V \to W$ 都是同构映射. 例如, 如果 $A \in M_{m\times n}(\mathbb{R})$, 我们考虑 $\varphi_A : \mathbb{R}^n \to \mathbb{R}^n$. 同一个空间 $\mathbb{R}^n$ 当然是同构的; 但是 φ_A 是同构当且仅当 A 是可逆矩阵.

9.1.1 四个基本子空间, 奇异值分解和广义逆

假设 $A \in M_{m\times n}(\mathbb{R})$, 那么关于 A 我们有四个基本子空间:

- $C(A), N(A^T) \subseteq \mathbb{R}^m$, 且 $C(A)^\perp = N(A^T)$;
- $C(A^T), N(A) \subseteq \mathbb{R}^n$, 且 $C(A^T) = N(A)^\perp$;
- $\dim C(A) = \dim C(A^T) = r = \operatorname{rank}(A)$, $\dim N(A) = n - r$, $\dim N(A^T) = m - r$.

具体的图示请参考第五章. 现在我们理解一下奇异值分解的相关内容.

假设 A 的奇异值分解为 $A = U\Sigma V^T$, 其中: $U \in M_{m\times m}(\mathbb{R}), V \in M_{n\times n}(\mathbb{R})$ 都是正交矩阵, $\Sigma = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} \in M_{m\times n}(\mathbb{R}), D_r = \operatorname{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$, $\sigma_1 > \sigma_2 > \cdots > \sigma_r > 0$. 考虑

$$\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad \boldsymbol{v} \longmapsto A\boldsymbol{v}\ ;$$

我们把 U, V 都写成列向量的形式:

$$U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m \end{bmatrix}; \quad V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$$

这一节以下选读

9.2 交换图示

这一节的主要目的是用图示来理解线性空间取基之后的坐标表示, 取不同的基的过度矩阵, 以及线性映射在不同的基之下的矩阵的变化关系. 我们把上一节的命题 9.13 重新叙述一下:

命题 9.15. 设 V 是一个 n 为线性空间. 对 V 的任意一组基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$, 存在唯一的线性同构 $\varphi : V \to \mathbb{R}^n$ 使得 $\varphi(\boldsymbol{v}_i) = \boldsymbol{e}_i$, $i = 1, 2, \cdots, n$.

反之, 给定线性空间的同构 $\varphi : V \to \mathbb{R}^n$, 则 $\boldsymbol{v}_i := \varphi^{-1}(\boldsymbol{e}_i)$ $(1 \le i \le n)$ 构成了 V 的一组基, 所以 $\dim V = n$.

这就可以看出来 “同构” 的意思: 同构就是有同样的结构, 或者说 “看起来一样”. 怎样 “才能看起来一样呢”? 那就取一组基, 取完了基之后, 就跟标准的 $\mathbb{R}^n$ 一样了; 或者说取基就是实现 “看起来” 这个动作的过程.

例题 9.16. 设 $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ 是一个线性映射, 那么 φ 是同构当且仅当 $\varphi(\boldsymbol{e}_1), \cdots, \varphi(\boldsymbol{e}_n)$ 线性无关, 当且仅当矩阵 $\begin{bmatrix} \varphi(\boldsymbol{e}_1) & \varphi(\boldsymbol{e}_n) & \cdots & \varphi(\boldsymbol{e}_n) \end{bmatrix}$ 是可逆矩阵.

因此, 除了用标准正交基来 “看” $\mathbb{R}^n$, 我们可以用任何一组基来看待它.

那过渡矩阵是什么呢? 给定 n-维线性空间, 给定两组基 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 和 $\boldsymbol{v}_1',\cdots,\boldsymbol{v}_n'$, 所谓的过渡矩阵 P 是唯一的 n 阶可逆矩阵使得

$$\begin{bmatrix}\boldsymbol{v}_1' & \boldsymbol{v}_2' & \cdots & \boldsymbol{v}_n'\end{bmatrix}=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}P.$$

设 $\varphi,\varphi':V\to\mathbb{R}^n$ 分别是这两组基对应的同构, 我们有交换图表

$$\begin{tikzcd} & V \arrow[dl, "\varphi'"'] \arrow[dr, "\varphi"] & \\ \mathbb{R}^n \arrow[rr, dashed, "P"] & & \mathbb{R}^n \end{tikzcd}$$

注释 9.17. 所谓的"交换图表", 意思是按照映射的箭头不同的路径进行复合最终得到的复合映射是相同的. 我们来验证一下上面的图表是交换图表, 也就是说, $P\circ\varphi'=\varphi$. 一方面,

$$P\circ\varphi'\left(\begin{bmatrix}\boldsymbol{v}_1' & \cdots & \boldsymbol{v}_n'\end{bmatrix}\right)=P\begin{bmatrix}\varphi'(\boldsymbol{v}_1') & \cdots & \varphi'(\boldsymbol{v}_n')\end{bmatrix}=P\begin{bmatrix}\boldsymbol{e}_1 & \cdots & \boldsymbol{e}\end{bmatrix}=P;$$

另一方面,

$$\varphi\left(\begin{bmatrix}\boldsymbol{v}_1' & \cdots & \boldsymbol{v}_n'\end{bmatrix}\right)=\varphi\left(\begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n\end{bmatrix}P\right)=\begin{bmatrix}\boldsymbol{e}_1 & \cdots & \boldsymbol{e}\end{bmatrix}P=P.$$

这就得到了我们的结论.

注释 9.18. 下面的所有交换图表都不再验证, 证明方法和上面雷同, 请自己练习.

设 V 是 n 维线性空间, W 是 m 维线性空间, $f:V\to W$ 是线性变换, 那么给定 V 的一组基 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$, W 的一组基 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_m$, 它们分别给出同构: $\varphi:V\to\mathbb{R}^n,\psi:W\to\mathbb{R}^m$, 在这两组基之下, f 也给出了矩阵 A, 那么我们有交换图表:

$$\begin{tikzcd} V \arrow[r] \arrow[d, "\varphi"] & W \arrow[d, "\psi"] \\ \mathbb{R}^n \arrow[r, dashed, "A"] & \mathbb{R}^m \end{tikzcd}.$$

如果还有线性空间 U 和线性映射 $g:W\to U$, 取定 U 的基, 就有同构 $\phi:U\to\mathbb{R}^k$, 那么我们有交换图表

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\varphi"] & W \arrow[r, "g"] \arrow[d, "\psi"] & U \arrow[d, "\phi"] \\ \mathbb{R}^n \arrow[r, dashed, "A"] & \mathbb{R}^m \arrow[r, dashed, "B"] & \mathbb{R}^k \end{tikzcd}.$$

设 V 是 n 维线性空间, W 是 m 维线性空间, $f : V \to W$ 是线性变换, 那么给定 V 的一组基 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$, W 的一组基 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_m$, 它们分别给出同构: $\varphi : V \to \mathbb{R}^n, \psi : W \to \mathbb{R}^m$, 且在这两组基之下 f 对应于矩阵 A. 给定 V 的另一组基 $\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n$ 和 W 的另一组基 $\boldsymbol{w}'_1, \cdots, \boldsymbol{w}'_m$, 它们分别给出同构: $\varphi' : V \to \mathbb{R}^n, \psi' : W \to \mathbb{R}^m$, 且在这两组基之下 f 对应于另一个矩阵 A'. 若 V 的两组基之间的过渡矩阵为 P, W 的两组基之间的过渡矩阵为 Q, 我们已知 $A' = Q^{-1}AP$. 我们有如下的交换图表图表

$$\begin{tikzcd}
& V \arrow[rr, "f"] \arrow[dr, "\varphi'"] \arrow[ddl, "\varphi"'] & & W \arrow[dr, "\psi'"] \arrow[ddl, "\psi"] & \\
& & \mathbb{R}^n \arrow[rr, dashed, "A'"] \arrow[dll, dashed, "P"'] & & \mathbb{R}^m \arrow[dll, dashed, "Q"] \\
\mathbb{R}^n \arrow[rr, dashed, "A"] & & \mathbb{R}^m & &
\end{tikzcd}$$

特别地, 我们考虑 V 上的线性变换. 设 V 是 n 维线性空间, $f : V \to V$ 是线性变换, 那么给定 V 的一组基 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$, 则在这组基之下 f 对应于一个矩阵 A; 给定另一组基 $\boldsymbol{v}'_1, \cdots, \boldsymbol{v}'_n$, 则在这组基之下 f 对应于另一个矩阵 A'. 若这两组基之间的过渡矩阵为 P, 我们已知 $A' = P^{-1}AP$. 我们有如下的交换图表图表

$$\begin{tikzcd}
& V \arrow[rr, "f"] \arrow[dr, "\varphi'"] \arrow[ddl, "\varphi"'] & & V \arrow[dr, "\varphi'"] \arrow[ddl, "\varphi"] & \\
& & \mathbb{R}^n \arrow[rr, dashed, "A'"] \arrow[dll, dashed, "P"'] & & \mathbb{R}^n \arrow[dll, dashed, "P"] \\
\mathbb{R}^n \arrow[rr, dashed, "A"] & & \mathbb{R}^n & &
\end{tikzcd}$$

9.3 举例

9.3.1 投影矩阵

考虑矩阵 $A \in M_{\times n}(\mathbb{R})$, 以及 $p : \mathbb{R}^m \to \mathbb{R}^m$ 为向 A 的列向量空间 $C(A)$ 的正交投影映射. 那么我们之前证明过: 若我们取 $C(A)$ 的任意一组基 $\alpha_1, \cdots, \alpha_r$, 令 $B = \begin{bmatrix} \alpha_1 & \cdots & \alpha_r \end{bmatrix}$, 它是列满秩矩阵, 那么投影映射 p 可以用矩阵 $B(B^TB)^{-1}B^T$ 来表示. 这个矩阵和 $\alpha_1, \cdots, \alpha_r$ 的选取无关. 精确地说, 对任意列向量 $\alpha \in \mathbb{R}^m$, 那么 $p(\alpha) = B(B^TB)^{-1}B^T\alpha$. 事实上, $B(B^TB)^{-1}B^T$ 是 p

在 $\mathbb{R}^m$ 的自然的标准正交基之下的矩阵, 也就是说:

$$p(x_1\boldsymbol{e}_1+\cdots+x_n\boldsymbol{e}_n)=\begin{bmatrix}\boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n\end{bmatrix}B(B^TB)^{-1}B^T\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$$

好的, 现在我们尝试着取一组具体的基 $\alpha_1,\cdots,\alpha_r$ 来实现这一点. 事实上, A 的奇异值分解就可以给我们提供这么一组基. 设 A 的奇异值分解为 $A=P\begin{bmatrix}D_r & 0\\ 0 & 0\end{bmatrix}Q$, 其中 $P\in M_{m\times m}(\mathbb{R}), Q\in M_{n\times n}(\mathbb{R})$ 都是正交矩阵. 那么 $AQ^T=P\begin{bmatrix}D_r & 0\\ 0 & 0\end{bmatrix}$. 令 $P=\begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m\end{bmatrix}$, 则

$$P\begin{bmatrix}D_r & 0\\ 0 & 0\end{bmatrix}=\begin{bmatrix}\sigma_1\boldsymbol{u}_1 & \cdots & \sigma_r\boldsymbol{u}_r & 0 & \cdots & 0\end{bmatrix}.$$

由于 Q 可逆, 那么 $C(A)=C(AQ^T)$; 所以 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_r$ 是 $C(A)$ 的一组基. 因此我们可以取 $B=\begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r\end{bmatrix}$, 于是 $B^TB=I_r$ 是 r 阶单位阵. 那么 p 在 $\mathbb{R}^m$ 的自然的标准正交基之下的矩阵为

$$BB^T=\boldsymbol{u}_1\boldsymbol{u}_1^T+\cdots+\boldsymbol{u}_r\boldsymbol{u}_r^T.$$

那么, 我们也可以尝试着选取别的基, 使得 p 在这一组基之下的矩阵能够简单一点. 例如, 我们可以以 P 的列向量为一组基. 注意:

$$p(\boldsymbol{u}_i)=(\boldsymbol{u}_1\boldsymbol{u}_1^T+\cdots+\boldsymbol{u}_r\boldsymbol{u}_r^T)(\boldsymbol{u}_i)=\sum_{j=1}^{r}\langle\boldsymbol{u}_j,\boldsymbol{u}_i\rangle\,\boldsymbol{u}_j;$$

所以

$$p(\boldsymbol{u}_i)=\begin{cases}\boldsymbol{u}_i & 1\le i\le r\\ 0 & r+1\le i\le n\end{cases}$$

所以 (这个 "所以" 如果看不懂, 请回顾命题 (9.6) 前面的那段论述) p 在 $\mathbb{R}^m$ 基 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$ 之下的矩阵就是 $\begin{bmatrix}I_r & 0\\ 0 & 0\end{bmatrix}$.

事实上, 根据 p 的定义, 我们有

$$p(\alpha)=\begin{cases}\alpha & \alpha\in C(A)\\ 0 & \alpha\in C(A)^{\perp}\end{cases}$$

我们任意选取 $C(A)$ 的一组基 $\beta_1,\cdots,\beta_r$, $C(A)^{\perp}$ 的一组基 $\beta_{r+1},\cdots,\beta_n$, 那么一方面, $\beta,\cdots,\beta_r,\beta_{r+1},\cdots,\beta_m$ 是 $\mathbb{R}^m$ 的一组基; 另一方面我们有

$$p(\beta_i)=\begin{cases}\beta_i & 1\le i\le r\\ 0 & r+1\le i\le n\end{cases}$$

所以 p 在 $\mathbb{R}^m$ 的基 $\beta_1, \cdots, \beta_n$ 之下的矩阵就是 $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. 事实上, 我们对 A 作了奇异值分解 $P\begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix}Q$, 那么 P 的列向量 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r$ 是 $C(A)$ 的一组基, 而 $\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_n$ 是 $C(A)^\perp$ 的一组基. 所以这里的结论会和上面的一致.

9.3.2 幂零矩阵

$A \in M_{n\times n}(\mathbb{R})$ 被称作幂零矩阵, 如果存在正整数 k 使得 $A^k = 0$. 事实上, 我们可以证明, 如果 A 是幂零矩阵, 那么存在正整数 $1 \le k \le n$ 使得 $A^k = 0$. 这是为什么呢?

引理 9.19. 假设非零向量 $\boldsymbol{v} \in \mathbb{R}^n$ 满足 $A^k\boldsymbol{v} = 0$ 但是 $A^{k-1}\boldsymbol{v} \neq 0$; 则 $\boldsymbol{v}, A\boldsymbol{v}, \cdots, A^{k-1}\boldsymbol{v}$ 线性无关.

证明请参考第十二周选做题第 8 题.

那么, 我们假设 k 是最小的正整数使得 $A^k = 0$; 最小就意味着 $k = 1$ 或者 $A^{k-1} \neq 0$. 如果 $k = 1$ 那么论证结束. 现在假设 $A^{k-1} \neq 0$. 那么就存在非零向量 $\boldsymbol{v} \in \mathbb{R}^n$ 使得 $A^{k-1}\boldsymbol{v} \neq 0$. 于是 $\boldsymbol{v}, A\boldsymbol{v}, \cdots, A^{k-1}\boldsymbol{v}$ 线性无关; 一共是 k 个元素. 而 $\mathbb{R}^n$ 中线性无关的向量组中的向量个数至多有 n 个, 所以 $k \le n$.

这里, 我们有两种极端情况: 一是 $k = 1$ 是使得 $A^k = 0$ 的最小正整数, 这是说 $A = 0$; 二是 $k = n$ 是使得 $A^k = 0$ 的最小正整数. 如上所述, 我们任意选取非零向量 $\boldsymbol{v} \in \mathbb{R}^n$ 使得 $A^{n-1}\boldsymbol{v} \neq 0$. 于是 $\boldsymbol{v}, A\boldsymbol{v}, \cdots, A^{n-1}\boldsymbol{v}$ 线性无关, 那他们就构成了 $\mathbb{R}^n$ 的一组基. 我们考虑线性映射 $\varphi_A : \mathbb{R}^n \to \mathbb{R}^n, \varphi_A(\alpha) = A\alpha$; 那么 φ_A 在这组基之下的矩阵等于什么呢? 它就等于

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \vdots & \ddots & 1 & 0 \\ 0 & 0 & \vdots & \ddots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

这样的矩阵我们记作 J_n, 即 J_n 是这样的 n 阶方阵: 它的对角线上方都是 1; 其他所有位置都等于 0. 特别的, $J_1 = \begin{bmatrix} 0 \end{bmatrix}$.

一般的, 我们有这样的结论:

命题 9.20. 对任意的 n 阶幂零矩阵 A, 存在 $\mathbb{R}^n$ 的一组基以及正整数 $n_1, \cdots, n_t$, 使得 φ_A 在这组基之下的矩阵为分块对角矩阵 $\mathrm{diag}(J_{n_1}, J_{n_1}, \cdots, J_{n_t})$.

用幂零矩阵和纯量矩阵 (或者说数量矩阵, 形如 cI_n 这样的矩阵) 为基本元素, 我们可以" 拼出" 所有的矩阵, 这就是所谓的约当 (Jordan) 标准型理论:

命题 9.21. 对任意的 $A \in M_{n\times n}(\mathbb{R})$, 存在可逆矩阵 $P \in M_{n\times n}(\mathbb{C})$, 正整数 $n_1, \cdots, n_t$ 和 $\lambda_1, \cdots, \lambda_t \in \mathbb{C}$, 使得 $P^{-1}AP$ 等于分块对角矩阵

$$\mathrm{diag}(\lambda_1 I_{n_1} + J_{n_1}, \lambda_2 I_{n_2} + J_{n_1}, \cdots, \lambda_t I_{n_t} + J_{n_t}).$$