

梁鑫 田垠 杨一龙 编著

# 线性代数入门

Diagram illustrating the relationship between matrices  $A$ ,  $U$ ,  $\Sigma$ , and  $V^T$ . Matrix  $A$  is tall and narrow,  $U$  is tall and narrow,  $\Sigma$  is a diagonal matrix, and  $V^T$  is wide and short. The diagram shows  $A = U\Sigma V^T$ .

清华大学内部讲义  
2020 年秋

## 内容简介

线性代数是当代大学生的必修科目,也是当前科学技术领域的数学基础和通用语言.随着数据科学的发展,线性代数的地位越来越重要.本书前六章是线性代数的基础,主要讨论  $\mathbb{R}^n$  的结构,内容包括向量、矩阵、子空间、内积、行列式、特征值等;后两章是对基本内容的升华,主要讨论抽象的线性空间、线性映射、内积等内容.

本书将定义、定理、命题、推论、例、注等统一编号为  $a.b.c$ ,表示第  $a$  章第  $b$  节第  $c$  项定义、定理或其他内容;将习题、阅读材料等集中放在每节之后,并统一编号为  $a.b.c$ ,表示第  $a$  章第  $b$  节第  $c$  项习题或阅读材料.

本书最后列出名词索引、人名表、符号表,以供读者查阅.正文中出现的人名除中国人外一般以拉丁转写形式出现,并在人名表中提供了原文和常见翻译.

本书可作为线性代数课程的初学教材或参考书.

本书的部分习题来自以下参考文献,供读者参阅:

- [1] G. Strang, 线性代数 (第 5 版), 清华大学出版社, 2019 年.
- [2] 丘维声, 简明线性代数, 北京大学出版社, 2002 年.
- [3] 蓝以中, 高等代数简明教程 (上、下), 北京大学出版社, 2002 年.

## 著作权声明

本作品的内容受中华人民共和国著作权法保护,任何未经授权的使用都可能构成侵犯著作权.

本作品的内容的著作权,未经特殊说明的,均属编撰者所有.

本作品的当前版本由编撰者或其他使用者提供给使用者使用.

除去法律规定的不经编撰者授权即可享有的权利以外,编撰者授予当前版本的使用者任意复制当前版本和分发当前版本给清华大学师生的权利.

# 目录

目录	iii
<b>第一章 线性映射和矩阵</b>	<b>1</b>
1.1 映射	1
1.2 基本概念	4
1.3 线性映射的表示矩阵	15
1.4 线性方程组	25
1.5 线性映射的运算	37
1.5.1 线性运算	37
1.5.2 复合	41
1.6 可逆矩阵	54
1.6.1 初等矩阵	54
1.6.2 可逆矩阵	55
1.6.3 等价关系	61
1.7 分块矩阵	66
1.8 LU 分解	77
<b>第二章 子空间和维数</b>	<b>85</b>
2.1 基本概念	85
2.2 基和维数	95
2.2.1 向量的线性关系	95
2.2.2 基和维数	99
2.3 矩阵的秩	103
2.4 线性方程组的解集	112

<b>第三章 内积和正交性</b>	<b>121</b>
3.1 基本概念	121
3.1.1 内积	121
3.1.2 标准正交基	124
3.2 正交矩阵和 QR 分解	131
3.2.1 正交矩阵	131
3.2.2 QR 分解	133
3.3 正交投影	140
3.3.1 正交补	142
3.3.2 正交投影	144
3.3.3 最小二乘问题	146
<b>第四章 行列式</b>	<b>155</b>
4.1 引子	155
4.2 行列式函数	158
4.3 行列式的展开式	170
<b>第五章 特征值和特征向量</b>	<b>179</b>
5.1 引子	179
5.2 基本概念	183
5.3 对角化和谱分解	192
5.4 相似	200
<b>第六章 实对称矩阵</b>	<b>211</b>
6.1 实对称矩阵的谱分解	211
6.2 正定矩阵	218
6.3 奇异值分解	224
6.3.1 基本性质	224
6.3.2 矩阵范数和低秩逼近	227
6.3.3 主成分分析	229
<b>第七章 线性空间和线性映射</b>	<b>237</b>
7.1 线性空间	237
7.2 基和维数	243
7.3 线性映射	251
7.4 向量的坐标表示	257

7.5 线性映射的矩阵表示 . . . . .	263
<b>第八章 内积空间</b>	<b>277</b>
8.1 欧氏空间 . . . . .	277
8.2 欧氏空间上的线性映射 . . . . .	286
8.3 酉空间 . . . . .	295
<b>名词索引</b>	<b>303</b>
<b>人名表</b>	<b>311</b>
<b>符号表</b>	<b>313</b>

Liang, Tian & Yang

# 第一章 线性映射和矩阵

线性代数这门学科，以研究**线性系统**为目的，是一个从理论到应用都相当完善的数学分支。我们先从系统说起。系统是指由若干个相互作用和相互依赖的事物组合而成的具有特定功能的一个整体。比较简单的一种是集合之间的映射：

$$\text{输入} \xrightarrow{\text{系统}} \text{输出}.$$

读者在中学学习的函数是一种特殊的映射，即从实数集到实数集的映射。在本书中，我们用  $\mathbb{R}$  来表示实数集，而  $\mathbb{Q}$  和  $\mathbb{C}$  分别表示有理数集和复数集。

## 1.1 映射

本节主要回顾一下关于映射的基本知识，包括映射的像、原像，单射、满射、双射，映射的复合，变换，逆映射等。

设有集合  $X$  和  $Y$ ，如果  $X$  中的任意元素  $x$ ，都以某种法则对应于  $Y$  中唯一一个元素，则称这个对应的法则  $f$  是集合  $X$  到集合  $Y$  的一个**映射**， $X$  中的元素  $x$  所对应的  $Y$  中的元素常记作  $f(x)$ 。

两个映射  $f, g$  相等，当且仅当对任意  $x \in X$ ， $f(x) = g(x)$ 。

映射用如下记号表示：

$$\begin{aligned} f: X &\rightarrow Y, \\ x &\mapsto f(x). \end{aligned}$$

这里，集合  $X$  称为映射  $f$  的**定义域**，集合  $Y$  称为映射  $f$  的**陪域**。设  $x \in X, y = f(x) \in Y$ ，我们称  $y$  为  $x$  在映射  $f$  下的**像**， $x$  为  $y$  在  $f$  下的**原像**。注意， $Y$  中的元素不一定都有原像；有原像时，原像也可能不唯一。定义域  $X$  中所有元素在  $f$  下的像构成  $Y$  的一个子集  $\{f(x) \mid x \in X\}$ ，称为映射  $f$  的**像集**或**值域**，记作  $f(X)$ 。

根据映射的像集和原像的性质，我们考虑如下三种特殊情形：

1. 如果对  $X$  中的任意两个不同元素  $x_1, x_2$ ，都有  $f(x_1) \neq f(x_2)$ ，则称  $f$  是一个**单射**；即，任何元素的原像至多只有一个。

2. 如果对  $Y$  中的任意元素  $y$ , 都存在  $X$  中的某个  $x$ , 使得  $y = f(x)$ , 则称  $f$  是一个**满射**; 即,  $f$  的像集  $f(X) = Y$ .
3. 映射  $f$  既是单射又是满射, 则称  $f$  是一个**双射**. 双射  $f$  给出了集合  $X$  和  $Y$  之间的一个一一对应.

映射之间最重要的运算是两个映射的复合. 设  $X, Y, Z$  是三个集合,  $f: X \rightarrow Y$ ,  $g: Y \rightarrow Z$  是两个映射, 则可以定义一个映射

$$\begin{aligned} h: X &\rightarrow Z, \\ x &\mapsto h(x) = g(f(x)), \end{aligned}$$

称其为  $g$  与  $f$  的**复合**, 记作  $h = g \circ f$ . 可以如下表示:

$$\begin{aligned} X &\xrightarrow{f} Y \xrightarrow{g} Z, \\ x &\mapsto f(x) \mapsto g(f(x)). \end{aligned}$$

注意, 只有当  $f$  的陪域与  $g$  的定义域相同时, 才能定义二者的复合. 设  $X, Y, Z, W$  是四个集合,  $f: X \rightarrow Y, g: Y \rightarrow Z, h: Z \rightarrow W$  是三个映射, 则两种不同顺序的复合是同一个  $X$  到  $W$  的映射:  $h \circ (g \circ f) = (h \circ g) \circ f$ , 即映射的复合运算满足结合律. 因此, 这个映射可以写成  $h \circ g \circ f$  而不会引起混淆.

当映射  $f$  的定义域和陪域都是  $X$  时, 称  $f$  是集合  $X$  上的一个**变换**. 容易看到, 一个变换永远可以跟自身复合, 记  $f^2 = f \circ f, f^n = f^{n-1} \circ f, n \geq 1$ . 任意集合  $X$  上都有一个特殊的变换, 它把  $X$  中的任意元素  $x$  映到  $x$  本身, 称为  $X$  上的**恒同变换**, 记作  $\text{id}_X$ . 即  $\text{id}_X(x) = x$ , 对任意  $x \in X$  都成立. 特别地, 恒同变换与映射的复合不改变映射: 对任意映射  $f: X \rightarrow Y$ , 都有  $f \circ \text{id}_X = f = \text{id}_Y \circ f$ .

**定义 1.1.1** 设  $f: X \rightarrow Y$  是一个映射, 若存在另一个映射  $g: Y \rightarrow X$ , 满足  $f \circ g = \text{id}_Y, g \circ f = \text{id}_X$ , 则称映射  $f$  **可逆**, 称  $g$  是  $f$  的一个**逆 (映射)**.

容易看出, 当  $f$  可逆时, 它的逆映射是唯一的, 称为  $f$  的逆, 记作  $f^{-1}$ .

利用映射复合和恒等变换, 可以给出单射、满射和双射的等价描述.

**命题 1.1.2** 对映射  $f: X \rightarrow Y$ , 有

1.  $f$  是单射当且仅当存在另一个映射  $g: Y \rightarrow X$ , 使得  $g \circ f = \text{id}_X$ ;
2.  $f$  是满射当且仅当存在另一个映射  $g: Y \rightarrow X$ , 使得  $f \circ g = \text{id}_Y$ ;
3.  $f$  是双射当且仅当  $f$  可逆, 即存在另一个映射  $g: Y \rightarrow X$ , 使得  $f \circ g = \text{id}_Y, g \circ f = \text{id}_X$ .

命题的证明留作习题.



## 习题

**练习 1.1.1** 判断下列映射是否是单射、满射、双射，并写出双射的逆映射.

1.  $f: \mathbb{R} \rightarrow \mathbb{R},$   
 $x \mapsto x + 1.$
2.  $f: \mathbb{R} \rightarrow \mathbb{R},$   
 $x \mapsto 2x.$
3.  $f: \mathbb{R} \rightarrow \mathbb{R},$   
 $x \mapsto 3.$
4.  $f: \mathbb{R} \rightarrow \mathbb{R},$   
 $x \mapsto x^2.$
5.  $f: (-\infty, 0] \rightarrow \mathbb{R},$   
 $x \mapsto x^2.$
6.  $f: \mathbb{R} \rightarrow (0, \infty),$   
 $x \mapsto e^x.$
7.  $f: \left[-\frac{3\pi}{2}, -\frac{\pi}{2}\right] \rightarrow [-1, 1],$   
 $x \mapsto \sin x.$

**练习 1.1.2** 对复合映射  $f = g \circ h$ ，证明或举出反例.

1.  $g, h$  都是满射，则  $f$  一定是满射.
2.  $g, h$  都是单射，则  $f$  一定是单射.
3.  $h$  不是满射，则  $f$  可能是满射.
4.  $g$  不是满射，则  $f$  可能是满射.
5.  $g$  不是单射，则  $f$  可能是单射.
6.  $h$  不是单射，则  $f$  可能是单射.
7.  $g, h$  都不是双射，则  $f$  可能是双射.

思考其中真命题的逆否命题的含义.

**练习 1.1.3** 1. 在不改变对应法则和定义域的前提下， $\mathbb{R}$  上的变换  $f(x) = x^2 + 2x + 3$ ，把陪域换成哪个集合，得到的映射是满射？

2. 证明映射  $f: [0, 1] \rightarrow \mathbb{R}, f(x) = x$  是单射.

**练习 1.1.4** 下列  $\mathbb{R}$  上的变换，哪些满足交换律  $f \circ g = g \circ f$ ？

1.  $f(x) = x + 1, g(x) = 2x.$
2.  $f(x) = x^2, g(x) = x^3.$
3.  $f(x) = 2^x, g(x) = 3^x.$
4.  $f(x) = 2x + 1, g(x) = 3x + 2.$
5.  $f(x) = 2x + 1, g(x) = 3x + 1.$
6.  $f(x) = \sin x, g(x) = \cos x.$

**练习 1.1.5** 对  $\mathbb{R}$  上的变换  $f(x) = 2x + 1$  和  $g(x) = ax + b$ ，求实数  $a, b$ ，使得  $f \circ g = g \circ f$ .

**练习 1.1.6** 在化简函数  $\arccos \circ \sin: \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \rightarrow [0, \pi]$  的下列步骤中，分别利用了映射的哪些性质？

1. 如果  $y = (\arccos \circ \sin)(x)$ ，则  $\cos y = (\cos \circ (\arccos \circ \sin))(x)$ .
2.  $\cos \circ (\arccos \circ \sin) = (\cos \circ \arccos) \circ \sin.$
3.  $(\cos \circ \arccos) \circ \sin = \text{id}_{[0, 1]} \circ \sin.$
4.  $\text{id}_{[0, 1]} \circ \sin = \sin.$

综上得  $\sin x = \cos y$ ，由此推断化简结果.

**练习 1.1.7** 用数学归纳法证明，任取有限个映射  $f_1, \dots, f_n$ ，如果对  $1 \leq i \leq n-1$ ， $f_{i+1}$  的定义域都包含  $f_i$  的陪域，则复合映射  $f_n \circ \dots \circ f_1$  不依赖于映射两两复合次序的选择.

**练习 1.1.8** 证明，任意映射  $f$  都存在分解  $f = g \circ h$ ，其中  $h$  是满射， $g$  是单射.

**练习 1.1.9** 证明命题 1.1.2.

## 1.2 基本概念

我们从一个简单的系统开始.

**例 1.2.1 (桥墩载荷)** 有一座桥, 桥长为  $d$ , 假设桥两端各有一桥墩, 不考虑桥面和桥墩的重力、厚度. 现在在距左侧  $S_1$  桥墩  $l_1$  处放置一重物, 重力为  $F_1$ ; 在距左侧  $S_1$  桥墩  $l_2$  处放置一重物, 重力为  $F_2$ . 问两个桥墩各承担了多少载荷?

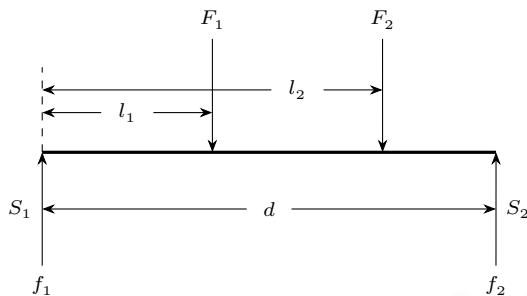


图 1.2.1: 桥墩载荷

利用杠杆原理, 考虑以  $S_1$  为支点、桥面为杆臂的杠杆, 显然  $F_1 l_1 + F_2 l_2 = f_2 d$ , 因此  $f_2 = \frac{l_1}{d} F_1 + \frac{l_2}{d} F_2$ . 考虑以  $S_2$  为支点的杠杆, 有  $F_1(d - l_1) + F_2(d - l_2) = f_1 d$ , 因此  $f_1 = \frac{d - l_1}{d} F_1 + \frac{d - l_2}{d} F_2$ . 也可以利用力平衡,  $f_1 = F_1 + F_2 - f_2$  求出  $f_1$ .

固定重物的位置, 把重力  $F_1, F_2$  视为系统的输入, 两个桥墩的载荷  $f_1, f_2$  视为系统的输出. 系统中的输入和输出满足**叠加原理**, 即 1. 输入放大或缩小某一倍数时, 输出也放大或缩小同一倍数, 2. 两组输入所产生的输出是二者分别产生的独立输出之和. 满足叠加原理的系统称为**线性系统**.

更一般地, 在任意多位置上放置任意重物, 我们都可以直接写出两个桥墩的载荷的表达式. ☺

上面描述的线性系统里, 输入和输出只有有限多个, 二者都可以表示成有序数组, 称之为**向量**.

**定义 1.2.2 (向量)** 一个  $m$  元有序数组  $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$  称为一个  $m$  维**向量**, 实数  $a_1, \dots, a_m$

称为向量  $\mathbf{a}$  的**分量**或**坐标**.

分量都是实数的  $m$  维向量的全体构成的集合记为  $\mathbb{R}^m$ .

**注 1.2.3** 分量都是复数的  $m$  维向量的全体构成的集合记为  $\mathbb{C}^m$ . 分量也可以取其他范围内的数, 本书主要讨论分量为实数和复数的两种情形.

两个向量**相等**是指二者的每个分量都相等. 向量常常用黑体小写字母表示<sup>1</sup>, 如  $\mathbf{a}$ . 由于分量纵向排列, 向量又称为列向量. 根据问题需要, 向量也可以把分量横向排列, 称为行向量. 为了与列向量区分, 用符号  $\mathbf{a}^T$  表示, 即如果  $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$ , 则  $\mathbf{a}^T = [a_1 \ \cdots \ a_m]$ .

现在, 例 1.2.1 中的线性系统可以表示成如下映射:

$$\begin{array}{ccc} \mathbb{R}^2 & \rightarrow & \mathbb{R}^2 \\ \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} & \mapsto & \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \frac{d-l_1}{d}F_1 + \frac{d-l_2}{d}F_2 \\ \frac{l_1}{d}F_1 + \frac{l_2}{d}F_2 \end{bmatrix}. \end{array}$$

一般地,  $n$  个输入  $m$  个输出的线性系统可以表示成如下映射:

$$\begin{array}{ccc} f: & \mathbb{R}^n & \rightarrow \mathbb{R}^m \\ \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} & \mapsto \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} & (1.2.1) \end{array}$$

我们将在下节说明表示的正确性.

线性系统中的输入和输出, 也就是向量, 需要满足叠加原理: 可加性要求任意两个向量之间可以相加, 而齐次性要求任意一个向量可以放缩某一倍数.

**定义 1.2.4 (线性运算)** 为  $\mathbb{R}^m$  中的向量定义两种运算<sup>2</sup>

1. 两个  $m$  维向量的**向量加法**:  $\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} := \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_m + b_m \end{bmatrix};$
2. 一个  $m$  维向量与一个数的**数乘**:  $k \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} := \begin{bmatrix} ka_1 \\ \vdots \\ ka_m \end{bmatrix};$

向量的加法和数乘统称向量的**线性运算**.

带有线性运算的集合  $\mathbb{R}^m$ , 称为**向量空间**  $\mathbb{R}^m$  或**线性空间**  $\mathbb{R}^m$ .

<sup>1</sup>手写时可以直接写字母, 也可以在字母上加箭头, 如  $\vec{a}$ .

<sup>2</sup>表达式中,  $A := B$  表示令  $A = B$  或者把  $B$  记作  $A$ ; 类似地,  $A =: B$  表示  $B := A$ .

注意, “线性空间  $\mathbb{R}^m$ ” 中的向量能做加法和数乘, 而 “集合  $\mathbb{R}^m$ ” 则不能, 二者并不是同一个数学对象. 另外, 线性空间  $\mathbb{R}^n$  和线性空间  $\mathbb{R}^m$  中的向量在  $m \neq n$  时无法做加法; 线性空间  $\mathbb{R}^m$  上的数乘运算中的数, 需要是实数.

线性运算本身需要满足一定的运算法则.

**命题 1.2.5** 向量加法和数乘满足如下八条运算法则:

1. 加法结合律:  $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ ;
2. 加法交换律:  $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ ;
3. 零向量: 存在  $m$  维**零向量**  $\mathbf{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ , 满足  $\mathbf{a} + \mathbf{0} = \mathbf{a}$ ;
4. 负向量: 对任意  $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$ , 记  $-\mathbf{a} = \begin{bmatrix} -a_1 \\ \vdots \\ -a_m \end{bmatrix}$ , 它满足  $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$ , 称它为  $\mathbf{a}$  的**负向量**;
5. 单位数:  $1\mathbf{a} = \mathbf{a}$ ;
6. 数乘结合律:  $(kl)\mathbf{a} = k(l\mathbf{a})$ ;
7. 数乘关于数的分配律:  $(k + l)\mathbf{a} = k\mathbf{a} + l\mathbf{a}$ ;
8. 数乘关于向量的分配律:  $k(\mathbf{a} + \mathbf{b}) = k\mathbf{a} + k\mathbf{b}$ .

命题的验证是直接的, 我们留给读者. 下面讨论运算法则的一些结果和注意事项.

1. 根据负向量性质, 我们可以定义向量的**减法**:  $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$ .
2. 根据加法结合律, 我们不需要区分加法运算的先后次序, 因此我们可以写  $\mathbf{a} + \mathbf{b} + \mathbf{c} = (\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ . 类似地,  $kla = (kl)\mathbf{a} = k(l\mathbf{a})$ .
3. 上面的运算法则使得我们能够 “合并同类项”. 例如,

$$2\mathbf{a} + 4(2\mathbf{a} + (3\mathbf{b} + \mathbf{0})) + 4((\mathbf{a} + (2 + 3)\mathbf{c}) - \mathbf{a}) = 10\mathbf{a} + 12\mathbf{b} + 20\mathbf{c}.$$

4. 实际上有很多和单位数法则类似的规则, 例如,  $0\mathbf{a} = \mathbf{0}$ ,  $(-1)\mathbf{a} = -\mathbf{a}$ .
5. 特别注意, 向量之间**没有**直接的乘除法. 有时我们会写  $\frac{\mathbf{a}}{k}$ , 这时  $k$  必须是实数而非向量. 只要向量出现在了分母上, 就意味着已经犯了十分严重的概念错误.

**例 1.2.6** 读者在中学学习过平面向量. 把平面向量的起点固定在原点, 它的坐标表示就是二维数组; 而平面向量的线性运算的坐标表示就是上面定义的两线性运算. 因此, 配备上线性运算, 所有以原点为起点的平面向量的全体构成的集合就是线性空间  $\mathbb{R}^2$ .  $\odot$

有了线性空间中的线性运算, 线性系统 (1.2.1) 的叠加原理可以具体地描述如下:

$$\begin{aligned} f(\mathbf{x} + \mathbf{x}') &= \begin{bmatrix} a_{11}(x_1 + x'_1) + \cdots + a_{1n}(x_n + x'_n) \\ \vdots \\ a_{m1}(x_1 + x'_1) + \cdots + a_{mn}(x_n + x'_n) \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} + \begin{bmatrix} a_{11}x'_1 + \cdots + a_{1n}x'_n \\ \vdots \\ a_{m1}x'_1 + \cdots + a_{mn}x'_n \end{bmatrix} = f(\mathbf{x}) + f(\mathbf{x}'), \\ f(k\mathbf{x}) &= \begin{bmatrix} a_{11}kx_1 + \cdots + a_{1n}kx_n \\ \vdots \\ a_{m1}kx_1 + \cdots + a_{mn}kx_n \end{bmatrix} = k \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = kf(\mathbf{x}), \end{aligned}$$

其中<sup>3</sup>,  $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ ,  $\mathbf{x}' = \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix} \in \mathbb{R}^n$ ,  $k \in \mathbb{R}$ . 由观察可知, 等式左边的  $\mathbf{x} + \mathbf{x}'$ ,  $k\mathbf{x}$  是  $\mathbb{R}^n$  中的线性运算, 而等式右边的  $f(\mathbf{x}) + f(\mathbf{x}')$ ,  $kf(\mathbf{x})$  是  $\mathbb{R}^m$  中的线性运算.

**定义 1.2.7 (线性映射)** 映射  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 如果满足

1. 任取  $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ , 都有  $f(\mathbf{x} + \mathbf{x}') = f(\mathbf{x}) + f(\mathbf{x}')$ ;
2. 任取  $\mathbf{x} \in \mathbb{R}^n, k \in \mathbb{R}$ , 都有  $f(k\mathbf{x}) = kf(\mathbf{x})$ ,

则称  $f$  为从  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的**线性映射**.

因此, (1.2.1) 中的映射是线性映射.

线性映射把定义域上的加法映射成陪域上的加法, 把定义域上的数乘映射成陪域上同一个数的数乘. 简单来说, 线性映射保持线性运算. 例如, 对任意线性映射  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 都有  $f(k\mathbf{x} + k'\mathbf{x}' + k''\mathbf{x}'') = kf(\mathbf{x}) + k'f(\mathbf{x}') + k''f(\mathbf{x}'')$ . 特别地, 线性映射把零向量映射到零向量: 在定义中取常数  $k = 0$ , 可以得到  $f(\mathbf{0}_n) = \mathbf{0}_m$ , 其中  $\mathbf{0}_n \in \mathbb{R}^n$  是  $\mathbb{R}^n$  中的零向量, 而  $\mathbf{0}_m \in \mathbb{R}^m$  是  $\mathbb{R}^m$  中的零向量. 利用这个性质, 可以对一些映射简单地做出其并非线性的判断.

<sup>3</sup>注意, 这里  $\mathbf{x}'$  表示不同于  $\mathbf{x}$  的另一个向量, 而不是对向量  $\mathbf{x}$  做了某种运算的结果. 类似地, 后文出现的  $\tilde{\mathbf{x}}, \hat{\mathbf{x}}$  等都表示不同的向量.

**例 1.2.8** 线性映射在生活中随处可见.

1. 去超市买水果, 在苹果、香蕉、樱桃之间选择. 潜在的购买空间是由这三种水果组

成的一个线性空间  $\mathbb{R}^3$ , 设苹果  $\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ , 香蕉  $\mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ , 和樱桃  $\mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ . 而结账

这个过程就是一个映射, 它将水果组合, 映射到对应的总价上. 如果苹果、香蕉和樱桃的每千克价格分别是 12.98、9.98 和 129.98 元, 那么“结账”这个映射可以表达为:

$$\begin{aligned} \text{结账: } \mathbb{R}^3 &\rightarrow \mathbb{R}, \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &\mapsto 12.98x_1 + 9.98x_2 + 129.98x_3. \end{aligned}$$

容易看出, 这是一个线性映射. 例如,  $\mathbf{x}, \mathbf{x}'$  分别代表两种水果组合, 对  $\mathbf{x} + 2\mathbf{x}'$  结账时, 总价等于对  $\mathbf{x}$  结账的总价加上对  $\mathbf{x}'$  结账的总价的 2 倍. 即, 结账 $(\mathbf{x} + 2\mathbf{x}')$  = 结账 $(\mathbf{x}) + 2$  结账 $(\mathbf{x}')$ .

2. 假设每千克苹果含有 12 克纤维素和 135 克糖, 每千克香蕉含有 12 克纤维素和 208 克糖, 每千克樱桃含有 3 克纤维素和 99 克糖. 所有纤维素和糖的含量的组合组成

一个二维线性空间  $\mathbb{R}^2$ , 设纤维素  $\mathbf{f} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ , 糖  $\mathbf{s} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ . 把每种水果组合映射到对应的纤维素和糖的含量, 是一个从  $\mathbb{R}^3$  到  $\mathbb{R}^2$  的映射, 称为“营养”:

$$\begin{aligned} \text{营养: } \mathbb{R}^3 &\rightarrow \mathbb{R}^2, \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &\mapsto \begin{bmatrix} 12x_1 + 12x_2 + 3x_3 \\ 135x_1 + 208x_2 + 99x_3 \end{bmatrix}. \end{aligned}$$

容易看出, 这是一个线性映射.

注意, 这两个例子中的线性映射都具有形如 (1.2.1) 的表达式. ☺

**定义 1.2.9 (线性变换)** 从  $\mathbb{R}^n$  到自身的线性映射称为  $\mathbb{R}^n$  上的**线性变换**.

特别地,  $\mathbb{R}^n$  上的**恒同变换**

$$\begin{aligned} \mathbf{I} = \text{id: } \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ \mathbf{x} &\mapsto \mathbf{x} \end{aligned} \tag{1.2.2}$$

是线性变换.

**例 1.2.10** 考虑例 1.2.6 中由平面向量构成的线性空间  $\mathbb{R}^2$ ，下面讨论平面<sup>4</sup>上的几类线性变换.

1. **旋转变换**: 对任意实数  $\theta$ ，将所有向量绕原点逆时针旋转  $\theta$  大小的角. 这定义了一个  $\mathbb{R}^2$  上的变换，记为  $R_\theta$ . 图 1.2.2 证明了  $R_\theta(kx) = kR_\theta(x)$ ,  $R_\theta(x + x') = R_\theta(x) + R_\theta(x')$ .

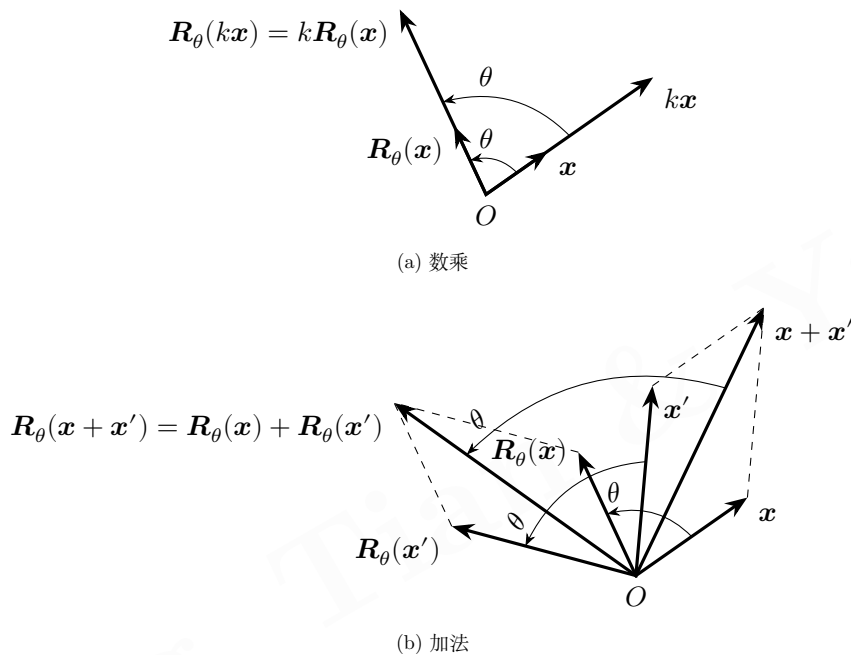


图 1.2.2: 旋转变换

因此，旋转变换  $R_\theta$  是线性变换. 那么，它是否有形如 (1.2.1) 的表达式呢？由于  $R_\theta$  是线性的，我们有

$$R_\theta \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = R_\theta \left( x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = x_1 R_\theta \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) + x_2 R_\theta \left( \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

容易知道，

$$R_\theta \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad R_\theta \left( \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix},$$

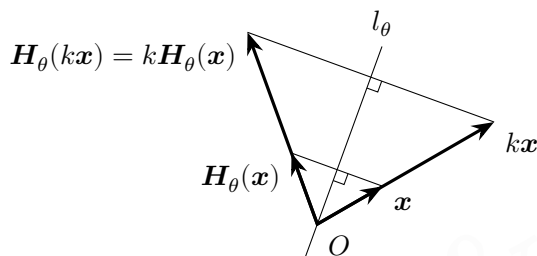
因此

$$R_\theta \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = x_1 \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} + x_2 \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} = \begin{bmatrix} x_1 \cos \theta - x_2 \sin \theta \\ x_1 \sin \theta + x_2 \cos \theta \end{bmatrix}.$$

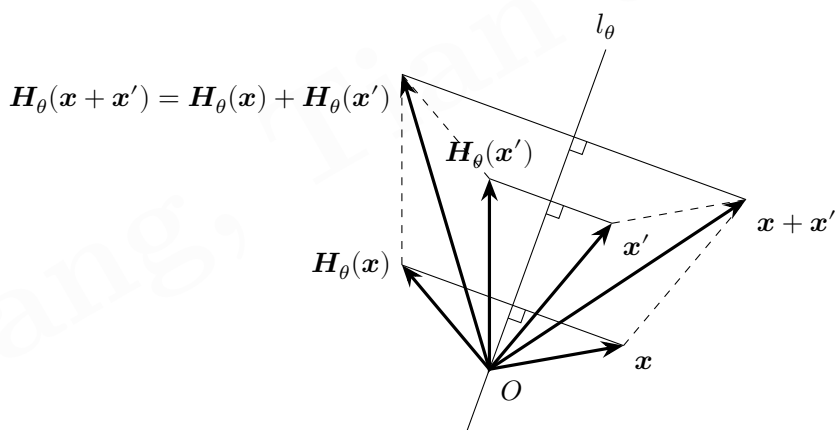
<sup>4</sup>我们以后经常会直接把  $\mathbb{R}^2$  称为平面，但要注意多数情形中  $\mathbb{R}^2$  中的元素是向量，而不是点.

注意, 线性映射  $\mathbf{R}_\theta$  仅仅由它在  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$  和  $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$  这两个向量上的取值决定. 这是线性映射最特殊的地方. 我们称  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$  和  $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$  这两个向量为  $\mathbb{R}^2$  的**标准坐标向量**.

2. **反射变换**: 给定直线  $l_\theta: x_2 \cos \theta - x_1 \sin \theta = 0$ , 其中  $\theta$  是直线与  $x_1$  坐标轴的夹角. 任意向量关于  $l_\theta$  的反射, 定义了一个  $\mathbb{R}^2$  上的变换, 记为  $\mathbf{H}_\theta$ . 图 1.2.3 证明了  $\mathbf{H}_\theta(kx) = k\mathbf{H}_\theta(x)$ ,  $\mathbf{H}_\theta(x + x') = \mathbf{H}_\theta(x) + \mathbf{H}_\theta(x')$ .



(a) 数乘



(b) 加法

图 1.2.3: 反射变换

因此, 反射变换  $\mathbf{H}_\theta$  是线性变换. 类似地, 由于

$$\mathbf{H}_\theta \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} \cos 2\theta \\ \sin 2\theta \end{bmatrix}, \quad \mathbf{H}_\theta \left( \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} \cos(2\theta - \frac{\pi}{2}) \\ \sin(2\theta - \frac{\pi}{2}) \end{bmatrix},$$



因此

$$\mathbf{H}_\theta \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = x_1 \begin{bmatrix} \cos 2\theta \\ \sin 2\theta \end{bmatrix} + x_2 \begin{bmatrix} \sin 2\theta \\ -\cos 2\theta \end{bmatrix} = \begin{bmatrix} x_1 \cos 2\theta + x_2 \sin 2\theta \\ x_1 \sin 2\theta - x_2 \cos 2\theta \end{bmatrix}.$$

特别地, 如果直线就是  $x_1$  坐标轴, 即  $\theta = 0$ , 则  $\mathbf{H}_\theta \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$ .

3. **对换变换**: 对换  $\mathbb{R}^2$  中向量的两个分量  $x_1, x_2$  也构成一个线性变换:

$$\mathbf{P}: \quad \mathbb{R}^2 \quad \rightarrow \quad \mathbb{R}^2 \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mapsto \quad \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}.$$

可以看到, 其实  $\mathbf{P}$  也是关于直线  $x_2 - x_1 = 0$  的反射.

4. **伸缩变换**: 设  $k \in \mathbb{R}$ , 定义一个  $\mathbb{R}^2$  上的变换  $\mathbf{C}_k$ , 它把向量在  $x_1$  方向拉伸  $k$  倍,  $x_2$  方向保持不变, 其表达式为:

$$\mathbf{C}_k: \quad \mathbb{R}^2 \quad \rightarrow \quad \mathbb{R}^2 \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mapsto \quad \begin{bmatrix} kx_1 \\ x_2 \end{bmatrix}.$$

容易验证  $\mathbf{C}_k$  是线性变换. 它把单位正方形  $\{0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$  映射成矩形  $\{0 \leq x_1 \leq k, 0 \leq x_2 \leq 1\}$ .

5. **投影变换**: 对伸缩变换  $\mathbf{C}_k$ , 取  $k = 0$ , 得到的变换是对  $x_2$  轴的投影  $\mathbf{C}_0$ .

6. **错切变换**: 设  $k \in \mathbb{R}$ , 定义  $\mathbb{R}^2$  上的一个变换  $\mathbf{S}_k$ , 它把  $x_1$  方向的  $k$  倍加到  $x_2$  方向上, 并保持  $x_1$  方向不变, 其表达式为:

$$\mathbf{S}_k: \quad \mathbb{R}^2 \quad \rightarrow \quad \mathbb{R}^2 \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mapsto \quad \begin{bmatrix} x_1 \\ kx_1 + x_2 \end{bmatrix}.$$

容易验证  $\mathbf{S}_k$  是线性变换. 它把单位正方形  $\{0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}$  映射成平行四边形  $\{0 \leq x_1 \leq 1, kx_1 \leq x_2 \leq kx_1 + 1\}$ , 而面积保持不变. ☺

**例 1.2.11** 容易验证, 以下这几个映射都不是线性映射.

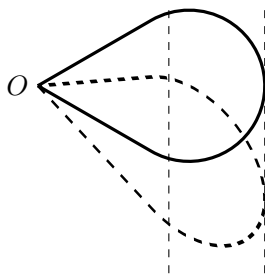


图 1.2.4: 错切变换

1. 平面上的平移变换: 设  $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$  是平面向量,  $T_{\mathbf{a}}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  是关于向量  $\mathbf{a}$  的平移, 其表达式为:

$$T_{\mathbf{a}}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} x_1 + a_1 \\ x_2 + a_2 \end{bmatrix}.$$

当  $\mathbf{a} \neq \mathbf{0}$  时,  $T_{\mathbf{a}}(\mathbf{0}) = \mathbf{0} + \mathbf{a} = \mathbf{a}$ , 即  $T_{\mathbf{a}}$  不保持零向量, 因此**不是**线性映射.

类似地, 如果直线  $l$  不经过原点, 那么关于  $l$  的反射变换就不是线性映射.

2. 取长度: 定义映射  $l: \mathbb{R}^2 \rightarrow \mathbb{R}$ , 它将平面上的任意向量  $\mathbf{x}$  映射到这个向量的长度. 其表达式为:

$$l: \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \sqrt{x_1^2 + x_2^2}.$$

3. 绕圆旋转: 定义映射  $c: \mathbb{R} \rightarrow \mathbb{R}^2$ , 它将输入的角度映射到单位圆上对应的位置. 如果将输入看作时间, 那么这个映射就可以看作是一个点匀速绕定点旋转.

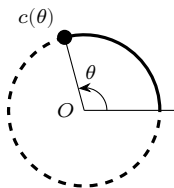


图 1.2.5: 匀速圆周运动

这个映射的表达式为:

$$\begin{aligned} c: \mathbb{R} &\rightarrow \mathbb{R}^2 \\ \theta &\mapsto \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}. \end{aligned}$$

4. 齐次非线性: 定义映射  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ . 其表达式为:

$$\begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &\mapsto (x_1^3 + x_2^3)^{\frac{1}{3}}. \end{aligned}$$

注意, 这个映射虽然满足  $f(k\mathbf{a}) = kf(\mathbf{a})$ , 但是  $f(\mathbf{a} + \mathbf{b}) \neq f(\mathbf{a}) + f(\mathbf{b})$ .  $\odot$

映射  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 单独看  $\mathbb{R}^m$  中任意一个分量的表达式, 都是关于自变量  $x_1, \dots, x_n$  的多元函数. 可以观察到, 例 1.2.10 中的线性映射, 每一个分量都是自变量的线性函数, 即常数项为零的一次函数, 即都有形如 (1.2.1) 的表达式; 而例 1.2.11 中的反例, 映射的分量并不是线性函数, 亦即没有形如 (1.2.1) 的表达式. 那么, 线性映射是不是总有形如 (1.2.1) 的表达式呢?

## 习题

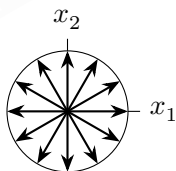


图 1.2.6: 练习 1.2.1

**练习 1.2.1** 如图 1.2.6 所示, 钟表表盘上对应整点有 12 个向量.

1. 计算 12 个向量之和.
2. 不计 2 点方向向量, 计算其他 11 个向量之和.
3. 假设这 12 个向量的起点从表盘中心移到 6 点, 则 12 点对应向量变为  $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$ . 计算此时 12 个向量之和.

**练习 1.2.2** 如果平面上的向量  $\begin{bmatrix} a \\ b \end{bmatrix}$  与  $\begin{bmatrix} c \\ d \end{bmatrix}$  共线, 那么  $\begin{bmatrix} a \\ c \end{bmatrix}$  与  $\begin{bmatrix} b \\ d \end{bmatrix}$  是否共线?

**练习 1.2.3** 证明命题 1.2.5.

**练习 1.2.4** 1. 如果只用加法交换律, 不用加法结合律, 那么  $(a+b)+c$  有多少种与之相等的表达式?

2. 如果只用加法结合律, 不用加法交换律, 那么  $((a+b)+c)+d$  有多少种与之相等的表达式?

**练习 1.2.5** 判断下列映射是否是线性映射.

1.  $f: \mathbb{R} \rightarrow \mathbb{R},$

$$x \mapsto x+1.$$

2.  $f: \mathbb{R} \rightarrow \mathbb{R},$

$$x \mapsto 2x.$$

3.  $f: \mathbb{R} \rightarrow \mathbb{R},$

$$x \mapsto 0.$$

4.  $f: \mathbb{R} \rightarrow \mathbb{R},$

$$x \mapsto 1.$$

5.  $f: \mathbb{R} \rightarrow \mathbb{R},$

$$x \mapsto x^2.$$

6.  $f: \mathbb{R} \rightarrow \mathbb{R},$

$$x \mapsto 2^{\ln x}.$$

7.  $f: \mathbb{R}^2 \rightarrow \mathbb{R}^3,$

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x+y \\ y-x \\ 2x \end{bmatrix}.$$

8.  $f: \mathbb{R}^2 \rightarrow \mathbb{R}^3,$

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x+1 \\ y-x \\ 2x \end{bmatrix}.$$

**练习 1.2.6** 设  $f: \mathbb{R} \rightarrow \mathbb{R}$  是线性映射, 证明存在实数  $k$ , 使得  $f(x) = kx$ .

**练习 1.2.7** 设映射  $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2, f\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} g(x, y, z) \\ h(x, y, z) \end{bmatrix}$ , 其中  $g, h: \mathbb{R}^3 \rightarrow \mathbb{R}$ . 证明  $f$  是线性映射当且仅当  $g, h$  都是线性映射.

**练习 1.2.8** 设  $x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, b_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, b_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, b_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ , 是否存在线性映射  $f: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ , 满足  $f(x_i) = b_i, i = 1, 2, 3$ ?

**练习 1.2.9** 判断下列映射是否是线性映射.

1. 给定  $a \in \mathbb{R}^m, f: \mathbb{R} \rightarrow \mathbb{R}^m, f(k) = ka$ .

2. 给定实数  $k, f: \mathbb{R}^m \rightarrow \mathbb{R}^m, f(a) = ka$ .

3.  $f: \mathbb{R}^{m+1} \rightarrow \mathbb{R}^m, f\left(\begin{bmatrix} a_1 \\ \vdots \\ a_m \\ a_{m+1} \end{bmatrix}\right) = a_{m+1} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}.$

**练习 1.2.10** 给定三维向量  $a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ , 定义:

1. 二者点积为  $a \cdot b := a_1 b_1 + a_2 b_2 + a_3 b_3$ .

2. 二者叉积为  $a \times b := \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}.$

那么给定  $\mathbf{a} \in \mathbb{R}^m$ , 映射  $f: \mathbb{R}^3 \rightarrow \mathbb{R}, \mathbf{b} \mapsto \mathbf{a} \cdot \mathbf{b}$  和  $g: \mathbb{R}^3 \rightarrow \mathbb{R}^3, \mathbf{b} \mapsto \mathbf{a} \times \mathbf{b}$  是否是线性映射?

**练习 1.2.11** 给定平面上任意面积为 1 的三角形, 经过下列变换之后, 其面积是否确定? 如果是, 面积多少? (不需严谨证明, 猜测答案即可.)

1. 旋转变换.
2. 反射变换.
3. 对  $x_2$  投影的投影变换.
4.  $x_1$  方向拉伸 3 倍,  $x_2$  方向不变的伸缩变换.
5. 把  $x_1$  方向的 3 倍加到  $x_2$  方向上, 保持  $x_2$  方向不变的错切变换.

**练习 1.2.12** 设  $\mathbb{R}^2$  上的变换  $f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -y+1 \\ x+2 \end{bmatrix}$ .

1. 证明  $f$  不是线性变换.
2. 构造分解  $f = g \circ h$ , 其中  $g$  是  $\mathbb{R}^2$  上的平移变换,  $h$  是  $\mathbb{R}^2$  上的线性变换. (这种平移与线性变换的复合称为仿射变换. 注意, 平移变换并不是线性变换.)

**练习 1.2.13** 设连续函数  $f: \mathbb{R} \rightarrow \mathbb{R}$  满足  $f(a+b) = f(a) + f(b)$ ,  $f$  是否是线性映射? (需要微积分知识)

## 1.3 线性映射的表示矩阵

我们称线性空间中的一组向量  $\mathbf{a}_1, \dots, \mathbf{a}_n$  为一个向量组. 注意,  $\mathbf{a}_j$  用黑体, 表示它是一组向量中的第  $j$  个向量, 而不是某个向量的第  $j$  个分量. 首先考虑定义域  $\mathbb{R}^n$  中一组特殊的向量

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \quad (1.3.1)$$

这组向量称为  $\mathbb{R}^n$  的**标准坐标向量组**, 其中  $\mathbf{e}_i$  称为第  $i$  个**标准坐标向量**. 这组向量的特

殊之处在于,  $\mathbb{R}^n$  中的任意向量  $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$  都可以轻易地由它们做线性运算得到:

$$\mathbf{x} = x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n.$$

由于线性映射保持线性运算, 因此有

$$f(\mathbf{x}) = x_1 f(\mathbf{e}_1) + \dots + x_n f(\mathbf{e}_n). \quad (1.3.2)$$

**定义 1.3.1 (线性组合与线性表示)** 给定  $\mathbb{R}^m$  中向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  和一组数  $k_1, \dots, k_n \in \mathbb{R}$ , 称向量  $k_1 \mathbf{a}_1 + \dots + k_n \mathbf{a}_n$  是向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  的一个**线性组合**.

设  $\mathbf{b}$  是  $\mathbb{R}^m$  中的向量, 如果存在一组数  $k_1, \dots, k_n \in \mathbb{R}$ , 使得  $\mathbf{b} = k_1 \mathbf{a}_1 + \dots + k_n \mathbf{a}_n$ , 则称  $\mathbf{b}$  可以被向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  **线性表示**.

任意向量  $\mathbf{x} \in \mathbb{R}^n$  都可以被  $\mathbb{R}^n$  的标准坐标向量组  $\mathbf{e}_1, \dots, \mathbf{e}_n$  线性表示, 而  $f(\mathbf{x})$  也以同样方式被它们在  $f$  下的像  $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$  线性表示. 可以说明, 线性映射  $f$  由  $n$  个特殊的像  $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$  所决定.

**命题 1.3.2** 设  $f, g: \mathbb{R}^n \rightarrow \mathbb{R}^m$  是两个线性映射, 如果  $f(\mathbf{e}_i) = g(\mathbf{e}_i), i = 1, \dots, n$ , 则  $f = g$ .

证. 如果  $f(\mathbf{e}_i) = g(\mathbf{e}_i), i = 1, \dots, n$ , 则对任意  $\mathbf{x} \in \mathbb{R}^n$ , 都有

$$f(\mathbf{x}) = x_1 f(\mathbf{e}_1) + \dots + x_n f(\mathbf{e}_n) = x_1 g(\mathbf{e}_1) + \dots + x_n g(\mathbf{e}_n) = g(\mathbf{x}).$$

这意味着两个映射相等. □

这个命题解决了线性映射的唯一性问题, 那么存在性问题呢? 即, 任取  $\mathbb{R}^m$  中的  $n$  个向量  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , 能否找到一个线性映射  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 满足  $f(\mathbf{e}_i) = \mathbf{a}_i$ ? 答案是肯定的. 根据 (1.3.2), 首先定义映射

$$\begin{aligned} f: \quad \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} &\mapsto x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n. \end{aligned}$$

下面根据定义验证它是线性映射:

$$\begin{aligned} f(\mathbf{x} + \mathbf{x}') &= (x_1 + x'_1) \mathbf{a}_1 + \dots + (x_n + x'_n) \mathbf{a}_n \\ &= (x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n) + (x'_1 \mathbf{a}_1 + \dots + x'_n \mathbf{a}_n) \\ &= f(\mathbf{x}) + f(\mathbf{x}'), \end{aligned}$$

$$f(k\mathbf{x}) = kx_1 \mathbf{a}_1 + \dots + kx_n \mathbf{a}_n = k(x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n) = kf(\mathbf{x}),$$

其中,  $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{x}' = \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix} \in \mathbb{R}^n, k \in \mathbb{R}$ . 因此我们有如下结果, 其中的唯一性由命题 1.3.2 保证.

**命题 1.3.3** 任取  $\mathbb{R}^m$  中的  $n$  个向量  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , 都存在唯一的线性映射  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 满足  $f(\mathbf{e}_i) = \mathbf{a}_i$ .

这个命题从理论上完整地描述了线性映射. 一方面, 尽管定义域和陪域都包含无穷多个元素, 但保持线性运算这一性质使得线性映射能够仅仅由它在标准坐标向量组

$e_1, \dots, e_n$  上的值唯一决定; 另一方面, 这  $n$  个值在陪域中又是完全自由的, 可以任意选取.

下面从计算的角度, 具体地描述线性映射  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ . 设  $f(e_i) = a_i$ , 其中

$$a_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, a_2 = \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix}, \dots, a_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

因为  $f(x) = x_1 f(e_1) + \dots + x_n f(e_n)$ , 所以线性映射  $f$  恰好具有 (1.2.1) 中的形式. 把向量组  $a_1, \dots, a_n$  并列排在一起, 得到一个矩形的数表:

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

数学上称之为**矩阵**. 矩阵常常用大写字母表示, 如  $A$ . 上面的矩阵中, 每个元素带有两个下角标, 第一个代表元素所在的行, 第二个代表元素所在的列, 如  $a_{ij}$  就表示该元素位于第  $i$  行第  $j$  列的交叉点处, 称为矩阵  $A$  的  $(i, j)$  元<sup>5</sup>. 竖排的  $n$  个  $\mathbb{R}^m$  中的向量

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \dots, a_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

称为矩阵的**列向量**. 横排的  $m$  个  $\mathbb{R}^n$  中的向量

$$\tilde{a}_1^T = [a_{11} \ a_{12} \ \cdots \ a_{1n}], \dots, \tilde{a}_m^T = [a_{m1} \ a_{m2} \ \cdots \ a_{mn}]$$

称为矩阵的**行向量**. 有  $m$  行  $n$  列的矩阵简称为  $m \times n$  矩阵. 矩阵既可以写成列向量组

的横向排列  $A = [a_1 \ \cdots \ a_n]$ , 又可以写成行向量组的纵向排列  $A = \begin{bmatrix} \tilde{a}_1^T \\ \vdots \\ \tilde{a}_m^T \end{bmatrix}$ . 而  $m \times n$

矩阵  $A$  的  $(i, j)$  元如果是  $a_{ij}$ , 那么可以记作  $A = [a_{ij}]_{m \times n}$ . 两个矩阵**相等**是指二者的每个元素都相等. 根据定义,  $n$  维行向量也是  $1 \times n$  矩阵,  $m$  维列向量也是  $m \times 1$  矩阵.

根据命题 1.3.2, 这个矩阵  $A$  完全决定了一个线性映射  $f$ .

**定义 1.3.4 (线性映射的表示矩阵)** 设线性映射  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $e_i$  为  $\mathbb{R}^n$  的标准坐标向量, 若  $a_i = f(e_i)$ , 则称矩阵  $A = [a_1 \ \cdots \ a_n]$  为线性映射  $f$  在标准坐标向量下的表示矩阵.

<sup>5</sup>具体计算或书写中, 为避免单下标和双下标的歧义, 可以用逗号把双下标隔开, 如  $a_{i,j}$ .

以后我们用  $\mathbf{A}$  表示这个线性映射<sup>6</sup>. 因此,  $\mathbf{A}(\mathbf{x}) = x_1 \mathbf{a}_1 + \cdots + x_n \mathbf{a}_n$ , 其中  $\mathbf{a}_1, \cdots, \mathbf{a}_n$  是矩阵  $A$  的列向量组. 下面我们定义一种新的运算来表示像  $\mathbf{A}(\mathbf{x})$ .

**定义 1.3.5 (矩阵与向量的乘积)** 定义  $m \times n$  矩阵  $A$  和  $n$  维列向量  $\mathbf{x}$  的乘积:

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

把  $A$  写成列向量组的横向排列  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ , 则  $\mathbf{A}\mathbf{x} = x_1 \mathbf{a}_1 + \cdots + x_n \mathbf{a}_n$ , 即矩阵与向量的乘积是矩阵的列向量组的一个线性组合, 系数由  $\mathbf{x}$  的分量给出.

**例 1.3.6** 当行向量  $\mathbf{a}^T = [a_1 \ \cdots \ a_n]$  与列向量  $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$  进行矩阵乘法时, 结果为

$\mathbf{a}^T \mathbf{x} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$ . 这恰好是我们中学学过的二维向量的内积的推广, 我们称其为向量  $\mathbf{a}$  和向量  $\mathbf{x}$  的**内积**. 可见, 矩阵与列向量的乘积, 就是将这个矩阵的每一行对应向量分别与这个列向量做内积而得到的向量, 即, 把  $A$  写成行向量组的纵向排

列  $A = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \vdots \\ \tilde{\mathbf{a}}_m^T \end{bmatrix}$ , 则

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \vdots \\ \tilde{\mathbf{a}}_m^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{x} \\ \vdots \\ \tilde{\mathbf{a}}_m^T \mathbf{x} \end{bmatrix}, \quad (1.3.3)$$

其中  $\tilde{\mathbf{a}}_i^T, 1 \leq i \leq m$ , 是  $1 \times n$  矩阵, 而矩阵与列向量的乘积  $\tilde{\mathbf{a}}_i^T \mathbf{x}$  是  $1 \times 1$  矩阵, 也就是一个实数. ☺

于是, 线性映射  $\mathbf{A}$  就可以写成矩阵与向量的乘积的形式:  $\mathbf{A}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ , 而线性映射保持线性运算的性质就对应于乘积的运算法则.

**命题 1.3.7** 矩阵和向量的乘积对向量的线性运算满足分配律:

$$A(k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2) = k_1 (A\mathbf{x}_1) + k_2 (A\mathbf{x}_2).$$

**例 1.3.8** 对例 1.2.8 中的线性映射, 我们计算其表示矩阵.

<sup>6</sup> 黑体大写字母表示的映射在本书中常暗示其矩阵用相应的大写字母表示.



1. 结账是一个从  $\mathbb{R}^3$  到  $\mathbb{R}$  的线性映射, 其中定义域  $\mathbb{R}^3$  的标准坐标向量分别是一千克苹果、一千克香蕉和一千克樱桃, 它们对应的单价分别是 12.98、9.98 和 129.98 元. 因此结账作为线性映射, 它的表示矩阵是

$$\begin{aligned} A &= [\text{结账(一千克苹果)} \quad \text{结账(一千克香蕉)} \quad \text{结账(一千克樱桃)}] \\ &= [12.98 \quad 9.98 \quad 129.98]; \end{aligned}$$

其表达式可以由矩阵和向量乘法给出:

$$A(\mathbf{x}) = A\mathbf{x} = [12.98 \quad 9.98 \quad 129.98] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 12.98x_1 + 9.98x_2 + 129.98x_3.$$

2. 营养是一个从  $\mathbb{R}^3$  到  $\mathbb{R}^2$  的线性映射. 它的表示矩阵为

$$\begin{aligned} A &= [\text{营养(一千克苹果)} \quad \text{营养(一千克香蕉)} \quad \text{营养(一千克樱桃)}] \\ &= \begin{bmatrix} 12 & 12 & 3 \\ 135 & 208 & 99 \end{bmatrix}. \end{aligned}$$

其表达式为:

$$A(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 12 & 12 & 3 \\ 135 & 208 & 99 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 12x_1 + 12x_2 + 3x_3 \\ 135x_1 + 208x_2 + 99x_3 \end{bmatrix}.$$

☺

特别地,  $\mathbb{R}^n$  上的线性变换的表示矩阵是  $n \times n$  矩阵, 又称为  $n$  阶**方阵**.

**例 1.3.9** 考虑例 1.2.10 中平面向量构成的线性空间  $\mathbb{R}^2$  上的线性变换的表示矩阵.

1. 旋转变换  $R_\theta$  的表示矩阵为

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

其表达式为:

$$R_\theta(\mathbf{x}) = R_\theta \mathbf{x} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \cos \theta - x_2 \sin \theta \\ x_1 \sin \theta + x_2 \cos \theta \end{bmatrix}.$$

2. 反射变换  $\mathbf{H}_\theta$  的表示矩阵为

$$\mathbf{H}_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}.$$

其表达式为:

$$\mathbf{H}_\theta(\mathbf{x}) = \mathbf{H}_\theta \mathbf{x} = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \cos 2\theta + x_2 \sin 2\theta \\ x_1 \sin 2\theta - x_2 \cos 2\theta \end{bmatrix}.$$

3. 对换变换  $\mathbf{P}$  的表示矩阵为

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

其表达式为:

$$\mathbf{P}(\mathbf{x}) = \mathbf{P}\mathbf{x} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}.$$

4. 伸缩变换  $\mathbf{C}_k$  的表示矩阵为

$$\mathbf{C}_k = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}.$$

其表达式为:

$$\mathbf{C}_k(\mathbf{x}) = \mathbf{C}_k \mathbf{x} = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} kx_1 \\ x_2 \end{bmatrix}.$$

5. 错切变换  $\mathbf{S}_k$  的表示矩阵为

$$\mathbf{S}_k = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}.$$

其表达式为:

$$\mathbf{S}_k(\mathbf{x}) = \mathbf{S}_k \mathbf{x} = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ kx_1 + x_2 \end{bmatrix}.$$

☺

容易得到, (1.2.2) 中  $\mathbb{R}^n$  上的恒同变换的表示矩阵为

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix},$$

即对角元素（对角线上的元素，或者所在行数和所在列数相等的元素）都是 1，非对角元素都是 0. 这个矩阵称为  $n$  阶**恒同矩阵**或  $n$  阶**单位矩阵**，记为  $I_n$ . 它还可以用标准坐

标向量组表示：  $I_n = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_n] = \begin{bmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix}$ . 对  $\mathbb{R}^n$  中任意向量  $\mathbf{x}$ ，容易验证  $I_n \mathbf{x} = \mathbf{x}$ .

这意味着，它对应的线性映射  $\mathbf{I}$  能够很容易地由像求出原像（二者相等）.

一般地，非对角元素全为零的方阵称为**对角矩阵**，而其中为零的元素往往省略不写：

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{bmatrix} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} =: \text{diag}(d_i).$$

如果对角元素依次为  $d_1, \dots, d_n$ ，那么这个对角矩阵常用  $\text{diag}(d_i)$  表示.

如果对角矩阵的对角元素都不为零，则向量  $\mathbf{b} = [b_i]$  在对应线性映射  $\mathbf{D}$  下的原像就是向量  $\mathbf{x} = [\frac{b_i}{d_i}]$ . 事实上，由

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \\ d_n x_n \end{bmatrix},$$

就能直接得到  $\mathbf{x}$  的所有元素.

形如  $U = \begin{bmatrix} u_{11} & \cdots & u_{1n} \\ & \ddots & \vdots \\ & & u_{nn} \end{bmatrix}$  的方阵称为  $n$  阶**上三角矩阵**，形如  $L = \begin{bmatrix} l_{11} & & \\ \vdots & \ddots & \\ l_{n1} & \cdots & l_{nn} \end{bmatrix}$

的方阵称为  $n$  阶**下三角矩阵**. 如果上（下）三角矩阵的对角元素都是 0，则称为**严格上（下）三角矩阵**. 如果上（下）三角矩阵的对角元素都是 1，则称为**单位上（下）三角矩阵**.

如果上三角矩阵的对角元素都不为零，那么给定向量  $\mathbf{b}$ ，它在对应线性映射  $\mathbf{U}$  下的原像  $\mathbf{x}$  也容易求得. 事实上，观察

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1n}x_n \\ u_{22}x_2 + \cdots + u_{2n}x_n \\ \vdots \\ u_{nn}x_n \end{bmatrix},$$

由第  $n$  个分量相等  $b_n = u_{nn}x_n$ ，立刻得到  $x_n = \frac{b_n}{u_{nn}}$ . 把  $x_n$  代入第  $n-1$  个分量相等

关系  $b_{n-1} = u_{n-1,n-1}x_{n-1} + u_{n-1,n}x_n$ , 立得  $x_{n-1} = \frac{b_{n-1} - u_{n-1,n}x_n}{u_{n-1,n-1}}$ . 类似进行下去, 可得  $\mathbf{x}$  的所有元素. 这种自下而上逐个代入从而求出每一个分量的方法, 称为**回代法**.

类似地, 如果下三角矩阵的对角元素都不为零, 那么给定向量  $\mathbf{b}$ , 它在对应线性映射  $\mathbf{L}$  下的原像  $\mathbf{x}$  也容易求得. 事实上, 观察

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} l_{11}x_1 \\ l_{21}x_1 + l_{22}x_2 \\ \vdots \\ l_{n1}x_1 + l_{n2}x_2 + \cdots + l_{nn}x_n \end{bmatrix},$$

即可从第 1 个分量相等关系开始计算, 从  $x_1$  到  $x_n$  逐个得出  $\mathbf{x}$  的所有元素. 这种自上而下逐个代入从而求出每一个分量的方法, 称为**前代法**.

利用回代法或前代法可以求出向量在表示矩阵是对角元素都不为零的三角矩阵的线性映射下的原像. 那么对任意线性映射, 有何方法来求解向量的原像呢? 这将在下节详细阐述.

## 习题

**练习 1.3.1** 将下列向量  $\mathbf{b}$  写成矩阵和向量乘积的形式.

$$1. \mathbf{b} = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 5 \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

$$2. \mathbf{b} = 5 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} + 4 \begin{bmatrix} 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}.$$

$$3. \mathbf{b} = \begin{bmatrix} 2b+a+c \\ c-b \\ a+b+c \\ a+b \end{bmatrix}, \text{ 其中 } a, b, c \text{ 为常数.}$$

$$4. \mathbf{b} = f \left( \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} \right), \text{ 其中 } f: \mathbb{R}^5 \rightarrow \mathbb{R}^5 \text{ 是线性变}$$

换, 满足  $f(\mathbf{e}_k) = k\mathbf{e}_{6-k}, k = 1, \dots, 5$ .

5. 假设如果某天下雨, 则第二天下雨概率为 0.8; 如果当天不下雨, 则第二天下雨概率为 0.3. 已知当天有一半的概率会下雨, 令  $\mathbf{b} \in \mathbb{R}^2$ , 其两个分量分别是明天下雨和不下雨的概率.

**练习 1.3.2** 判断下列矩阵和向量的乘积是否良定义. 在可以计算时, 先将其写成列向量的线性组合, 再进行计算.

$$1. \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

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

$$3. \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

$$4. \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}.$$

$$5. \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix}.$$

$$6. \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

$$7. \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}.$$

$$8. \begin{bmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \\ 7 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

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

$$10. \begin{bmatrix} 1 & 3 & 7 \\ 1 & 2 & 4 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}.$$

**练习 1.3.3** 设  $A = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$ ,  $\mathbf{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ,  $\mathbf{v}_0 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ . 对任意自然数  $i$ , 令  $\mathbf{u}_{i+1} = A\mathbf{u}_i$ ,  $\mathbf{v}_{i+1} = A\mathbf{v}_i$ .

1. 对  $i = 1, 2, 3, 4$ , 计算  $\mathbf{u}_i, \mathbf{v}_i$ .
2. 猜测  $\lim_{i \rightarrow \infty} \mathbf{u}_i, \lim_{i \rightarrow \infty} \mathbf{v}_i$ .
3. 任取初始向量  $\mathbf{w}_0$ , 猜测极限  $\lim_{i \rightarrow \infty} \mathbf{w}_i$ , 不同初始向量得到的极限在同一条直线上吗?

**练习 1.3.4** 设线性变换  $f$  的矩阵为  $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ .

$$1. \text{ 令 } \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \text{ 将 } y_1, y_2, y_3 \text{ 分别用 } x_1, x_2, x_3 \text{ 表达出来.}$$

2. 将  $x_1, x_2, x_3$  分别用  $y_1, y_2, y_3$  表达出来.

$$3. \text{ 找到矩阵 } B, \text{ 使得 } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = B \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

4. 设  $g$  是由矩阵  $B$  决定的线性变换, 证明  $f, g$  互为逆变换.

**练习 1.3.5** 计算下列线性映射  $f$  的表示矩阵.

1.  $f$  为  $xy$  平面向  $y$  轴的投影变换.

$$2. f: \mathbb{R}^3 \rightarrow \mathbb{R}, f(\mathbf{x}) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \mathbf{x}, \text{ 其中点积的定义见练习 1.2.10.}$$

$$3. f: \mathbb{R}^3 \rightarrow \mathbb{R}^3, f(\mathbf{x}) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \mathbf{x}, \text{ 其中叉积的定义见练习 1.2.10.}$$

$$4. f: \mathbb{R}^4 \rightarrow \mathbb{R}^4, f \left( \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 + x_2 \end{bmatrix}.$$

$$5. f: \mathbb{R}^4 \rightarrow \mathbb{R}^4, f \left( \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right) = \begin{bmatrix} x_4 \\ x_2 \\ x_3 \\ x_1 \end{bmatrix}.$$

$$6. f: \mathbb{R}^m \rightarrow \mathbb{R}^m, f \left( \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \right) = \begin{bmatrix} x_m \\ \vdots \\ x_1 \end{bmatrix}.$$

**练习 1.3.6** 考虑桥墩载荷问题, 其中  $l_1, l_2, d$  作为常数.

1. 映射  $f$  输入  $F_1, F_2$  得到输出  $f_1, f_2$ , 写出  $f$  的矩阵.
2. 以桥梁的左端为支点, 顺时针方向的力矩为  $df_2 - l_1 F_1 - l_2 F_2$ . 以桥梁的右端为支点或者以桥梁的中点作为支点, 都能类似地得到顺时针方向的力矩. 假设映射  $f$  的输入为  $F_1, F_2, f_1, f_2$ , 而输出为桥梁的左端、中点和右端的顺指针方向的力矩, 写出  $f$  的矩阵.

**练习 1.3.7** 设  $xy$  平面  $\mathbb{R}^2$  上的变换  $f$  是下列三个变换的复合: 先绕原点逆时针旋转  $\frac{\pi}{6}$ ; 然后进行一个保持  $y$  坐标不变, 同时将  $y$  坐标的两倍加到  $x$  坐标上的错切; 最后再沿着直线  $x + y = 0$  反射.

1. 证明  $f$  是线性变换.
2. 计算  $f(e_1)$  和  $f(e_2)$ .
3. 写出  $f$  的矩阵.
4. 计算  $f \left( \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right)$ .

**练习 1.3.8** 1. 幻方矩阵是指元素分别是  $1, \dots, 9$  的 3 阶矩阵  $M$ , 且每行、每列以及两条对角线上

的三个元素之和都相同. 求  $M \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$  的所有可能值.

2. 数独矩阵是指 9 阶矩阵  $M$ , 从上到下、从左到右依次分成九个  $3 \times 3$  的子矩阵, 且每行、每列

以及九个子矩阵中的元素都是  $1, \dots, 9$ . 求  $M \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$  的所有可能值.

**练习 1.3.9** 证明, 如果  $n$  阶方阵  $A$  对任意  $n$  维列向量  $x$ , 都有  $Ax = 0$ , 则  $A = O$ .

**练习 1.3.10** 证明, 如果  $n$  阶方阵  $A$  对任意  $n$  维列向量  $x$ , 都存在依赖于  $x$  的常数  $c(x)$ , 满足  $Ax = c(x)x$ , 则存在常数  $c$ , 使得  $A = cI_n$ .

## 1.4 线性方程组

对线性映射  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 我们考虑它是否是单射或满射.

**命题 1.4.1** 1. 线性映射  $A$  是单射, 当且仅当方程  $Ax = 0_m$  只有唯一解  $0_n$ ;

2. 线性映射  $A$  是满射, 当且仅当对任意  $b \in \mathbb{R}^m$ , 方程  $A(x) = Ax = b$  有解.

证. 第 2 条根据定义显然. 第 1 条: 根据定义,  $A$  是单射, 当且仅当对任意  $x_1 \neq x_2$  都有  $Ax_1 = A(x_1) \neq A(x_2) = Ax_2$ , 它的逆否命题是  $Ax_1 = Ax_2$  给出  $x_1 = x_2$ , 这当且仅当  $A(x_1 - x_2) = 0$  给出  $x_1 - x_2 = 0$ , 由此即得.  $\square$

这两个问题, 可以归结于求解如下线性方程组:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. \end{cases} \quad (1.4.1)$$

**例 1.4.2 (鸡兔同笼)** 一个笼子中有鸡和兔共 8 只, 二者的腿共有 22 只, 那么鸡和兔各有多少只?

利用二元一次方程组来求解. 设鸡有  $x_1$  只, 兔有  $x_2$  只, 那么:

$$\begin{cases} x_1 + x_2 = 8, \\ 2x_1 + 4x_2 = 22. \end{cases} \quad (1.4.2)$$

通过消元法, 把第一式的  $-2$  倍加到第二式上, 就有

$$\begin{cases} x_1 + x_2 = 8, \\ 2x_2 = 6. \end{cases} \quad (1.4.3)$$

再给第二式乘以  $\frac{1}{2}$ , 得

$$\begin{cases} x_1 + x_2 = 8, \\ x_2 = 3. \end{cases} \quad (1.4.4)$$

然后把第二式代入第一式 (也可以解释为把第二式的  $-1$  倍加到第一式上), 即得结果

$$\begin{cases} x_1 = 5, \\ x_2 = 3. \end{cases} \quad (1.4.5)$$

通过观察可知, 在解方程组  $Ax = b$  的过程中, 未知数没有参与任何计算. 因此, 我们完全可以不写出未知数  $x$ , 直接在**系数矩阵**  $A$  和常数项  $b$  上计算, 也就是在如下矩阵

$$[A \quad b] \equiv [A \mid b] = \left[ \begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right]$$

上计算, 其中矩阵  $[A \quad b]$  称为方程组  $Ax = b$  的**增广矩阵**.

我们在增广矩阵上计算例 1.4.2:

$$\left[ \begin{array}{cc|c} 1 & 1 & 8 \\ 2 & 4 & 22 \end{array} \right] \xrightarrow[\text{加到第 2 行}]{\text{第 1 行的 } -2 \text{ 倍}} \left[ \begin{array}{cc|c} 1 & 1 & 8 \\ 0 & 2 & 6 \end{array} \right] \xrightarrow[\text{乘以 } \frac{1}{2}]{\text{第 2 行}} \left[ \begin{array}{cc|c} 1 & 1 & 8 \\ 0 & 1 & 3 \end{array} \right] \xrightarrow[\text{加到第 1 行}]{\text{第 2 行的 } -1 \text{ 倍}} \left[ \begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & 3 \end{array} \right].$$

上述解法中对方程组有如下两种变换: (1) 把某一方程的  $k$  倍加到另一方程上, 用以消去某个变元; (2) 把某一方程乘以非零常数  $k$ , 使得某个变元的系数为 1. 对增广矩阵有相应的两种行变换. 反复利用它们, 当系数矩阵变为恒同矩阵  $I_2$  时, 得到方程有唯一解.

无论未知数有多少个, 如果条件适合, 那么我们都可以用同样的办法逐个消去等式中的未知数, 直到得到方程组的解. 我们再来看另一个例子.

### 例 1.4.3 求解

$$\begin{cases} 2x_2 - x_3 = 1, \\ x_1 - x_2 + x_3 = 0, \\ 2x_1 - 2x_2 + 2x_3 = 1. \end{cases} \quad (1.4.6)$$

在增广矩阵上计算:

$$\left[ \begin{array}{ccc|c} 0 & 2 & -1 & 1 \\ 1 & -1 & 1 & 0 \\ 2 & -2 & 2 & 1 \end{array} \right] \xrightarrow{\text{对换第 1, 2 行}} \left[ \begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 2 & -1 & 1 \\ 2 & -2 & 2 & 1 \end{array} \right] \xrightarrow[\text{加到第 3 行}]{\text{第 1 行的 } -2 \text{ 倍}} \left[ \begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 2 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right].$$

注意到, 第三个方程  $0x_1 + 0x_2 + 0x_3 = 1$  无解. 因此, 原方程组无解. ☹

在这个例子中, 由于第一个方程中  $x_1$  的系数为零, 也就是系数矩阵左上角的元素  $a_{11} = 0$ , 我们需要首先对换第 1 行和第 2 行, 才能再用新得到的第 1 行做行变换消去变元  $x_1$ .

**定义 1.4.4 (初等变换)** 对方程组施加的如下三类变换的每一类都称为方程组的**初等变换**:



1. **对换变换**: 互换两个方程的位置;
2. **倍乘变换**: 把某个方程两边同乘一个非零常数  $k$ ;
3. **倍加变换**: 把某个方程的  $k$  倍加到另一个方程上.

容易看到, 初等变换可逆:

1. 对换变换的逆变换是它本身;
2. 参数为  $k$  的倍乘变换的逆变换也是倍乘变换, 参数为  $\frac{1}{k}$ ;
3. 参数为  $k$  的倍加变换的逆变换也是倍加变换, 参数为  $-k$ .

由此得到如下重要的结论.

**定理 1.4.5** 线性方程组经某个初等变换后得到的新方程组与原方程组同解.

证. 设方程组  $Ax = b$  有一组解  $x_1 = k_1, \dots, x_n = k_n$ , 则有恒等式  $Ak = b$ . 容易验证, 如果对换两个方程的位置, 那么再代入  $x_1 = k_1, \dots, x_n = k_n$ , 依旧得到恒等式, 因此  $x_1 = k_1, \dots, x_n = k_n$  仍然是新方程组的一组解. 类似地, 方程组施加了倍加变换或者倍乘变换后, 代入  $x_1 = k_1, \dots, x_n = k_n$  后也得到恒等式, 即它依旧是新方程组的一组解. 于是, 原方程组的任意一组解, 都是新方程组的解.

反过来, 由于初等变换可逆, 新方程组也可以通过某个初等变换得到原方程组, 因此新方程组的任意一组解都是原方程组的解. 于是两方程组同解.  $\square$

如前解法中, 在方程组上计算和在增广矩阵上计算完全相同. 因此, 对应于方程组的初等变换, 我们可以定义矩阵的初等变换.

**定义 1.4.6 (初等行 (列) 变换)** 对矩阵施加的如下三类变换的每一类都称为矩阵的**初等行 (列) 变换**:

1. **对换变换**: 互换两行 (列) 的位置;
2. **倍乘变换**: 某一行 (列) 乘以非零常数  $k$ ;
3. **倍加变换**: 把某个行 (列) 的  $k$  倍加到另一个行 (列) 上.

**例 1.4.7** 求解

$$\begin{cases} 2x_3 + x_4 = 1, \\ x_1 - x_2 + 4x_3 = 0. \end{cases} \quad (1.4.7)$$

在增广矩阵上计算:

$$\begin{aligned} \left[ \begin{array}{cccc|c} 0 & 0 & 2 & 1 & 1 \\ 1 & -1 & 4 & 0 & 0 \end{array} \right] & \xrightarrow[\text{第 1,2 行}]{\text{对换变换}} \left[ \begin{array}{cccc|c} 1 & -1 & 4 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 \end{array} \right] & \xrightarrow[\text{第 2 行}]{\text{倍乘变换}} \left[ \begin{array}{cccc|c} 1 & -1 & 4 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{array} \right] \\ & \xrightarrow[\text{第 1,2 行}]{\text{倍加变换}} \left[ \begin{array}{cccc|c} 1 & -1 & 0 & -2 & -2 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} \end{array} \right]. \end{aligned}$$

原方程组变换为:

$$\begin{cases} x_1 - x_2 - 2x_4 = -2, \\ x_3 + \frac{1}{2}x_4 = \frac{1}{2}. \end{cases} \quad (1.4.8)$$

把  $x_2, x_4$  移到右边, 得

$$x_1 = -2 + x_2 + 2x_4, \quad x_3 = \frac{1}{2} - \frac{1}{2}x_4. \quad (1.4.9)$$

任取  $x_2, x_4$  的一组值, 就得到方程组的一组解. 注意, 在变换后的系数矩阵中,  $x_1, x_3$  所在的两列恰好构成恒同矩阵  $I_2$ . ☺

如例 1.4.7 中的  $x_2, x_4$  这类可以任意取值的未知数, 称为**自由变量**, 而取值依赖于自由变量的未知数, 称为**主变量**. 而如 (1.4.9) 这样的表达式, 称为**通解**或**一般解**. 通解就是用含自由变量的表达式表示主变量. 注意, 在变换后的系数矩阵中, 主变量所在的列恰好构成恒同矩阵, 这使得我们可以通过移项直接写出通解.

在消元的过程中, 首先选择的是主变量, 其余的变量即为自由变量. 而主变量的选取并不是唯一的. 如例 1.4.7, 可以选择  $x_2, x_4$  为主变量进行消元, 其对应的增广矩阵计算为:

$$\left[ \begin{array}{cccc|c} 0 & 0 & 2 & 1 & 1 \\ 1 & -1 & 4 & 0 & 0 \end{array} \right] \xrightarrow[\text{第 1,2 行}]{\text{对换变换}} \left[ \begin{array}{cccc|c} 1 & -1 & 4 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 \end{array} \right] \xrightarrow[\text{第 1 行}]{\text{倍乘变换}} \left[ \begin{array}{cccc|c} -1 & 1 & -4 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 \end{array} \right].$$

通解为  $x_2 = x_1 + 4x_3, \quad x_4 = 1 - 2x_3$ .

在实际计算中, 我们通常选择角标小的变量 (对应矩阵中靠左的列) 作为主变量进行消元. 观察上面的例子, 对矩阵的计算总体上分两步.

首先, 从角标最小的变量开始, 必要时通过换行将该变量的系数非零的方程换到上面, 然后利用此方程向下消元. 变换后的矩阵满足如下特点:

1. 元素全为 0 的行 (称为**零行**), 只可能存在下方;
2. 元素不全为 0 的行 (称为**非零行**), 从左数第一个不为 0 的元素 (称为**主元**) 的列指标随着行指标的增加而严格增加.

这样的矩阵称为**阶梯形矩阵**. 阶梯形矩阵具有如下形式:

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

其中“\*”表示可能不为 0 的数,“•”表示一定不为 0 的数,即“•”处元素是主元. 一个阶梯形矩阵的非零行数称为它的阶梯数. 当求解方程组时,把增广矩阵化成阶梯形矩阵后,就可以判断方程组解的情形.

其次,先通过倍乘变换将主元化成 1,然后从角标最大的主变量(对应矩阵中靠右的列)开始,依次向上消元. 变换后的矩阵仍是阶梯形矩阵,并满足如下特点:

1. 每个非零行的主元都是 1;
2. 每个主元所在的列的其他元素都是 0.

这样的阶梯形矩阵称为**行简化阶梯形矩阵**. 行简化阶梯形矩阵具有如下形式:

$$\begin{bmatrix} 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * & \cdots & 0 & * & \cdots & * \\ & & & 1 & * & \cdots & * & 0 & * & \cdots & * & \cdots & 0 & * & \cdots & * \\ & & & & & & 1 & * & \cdots & * & \cdots & 0 & * & \cdots & * \\ & & & & & & & & & \ddots & \vdots & \vdots & & \vdots \\ & & & & & & & & & & 0 & * & \cdots & * \\ & & & & & & & & & & 1 & * & \cdots & * \end{bmatrix}.$$

注意,在行简化阶梯形中,主元所在的行和列的交叉点上的元素组成的矩阵是恒同矩阵. 当求解方程组时,把增广矩阵化成行简化阶梯形矩阵后,就可以写出它的通解.

下面,我们来解决理论问题,即说明使用初等变换就可以把矩阵化为阶梯形和行简化阶梯形.

**定理 1.4.8** 任意矩阵都可以用对换行变换和倍加行变换化为阶梯形;任意矩阵都可以用初等行变换化为行简化阶梯形.

证. 先证明第一部分. 对  $m \times n$  矩阵,对  $m$  用数学归纳法.  $m = 1$  时,矩阵只有一行,自然是阶梯形. 注意保持不变也是一种倍加变换,故  $m = 1$  时,方程组可以通过倍加变

换化为阶梯形. 现假设对任意  $n, (m-1) \times n$  矩阵都可以通过对换变换和倍加变换化为阶梯形, 考察  $m \times n$  矩阵, 分如下情形讨论.

1. 如果  $a_{11} \neq 0$ , 那么把第 1 行的  $-\frac{a_{i1}}{a_{11}}$  倍加到第  $i$  个方程上 (倍加变换), 就可以把其他行中的第一个元素化成 0. 那么第 2 到  $m$  行和第 2 到  $n$  列交叉点上的元素组成的矩阵, 根据归纳假设, 可以化为阶梯形. 从而原矩阵也可以化为阶梯形.
2. 否则  $a_{11} = 0$ . 如果  $a_{21}, \dots, a_{m1}$  中有某个不为 0, 记为  $a_{i1}$ . 则把第 1,  $i$  两方程互换位置 (对换变换), 问题归于第一种情形.
3. 否则  $a_{11}, \dots, a_{m1}$  全为 0, 即矩阵第一列元素全为 0. 类似前面情形考察矩阵的其他列. 如果存在  $j \leq n$  使得矩阵的第 2,  $\dots, j-1$  列全为 0 而第  $j$  列的元素不全为 0, 类似第一二种情形, 可以将原方程组化为阶梯形.
4. 否则矩阵的所有元素全为 0, 自然是阶梯形.

第二部分留给读者. □

阶梯形矩阵可以定性地判断方程组  $Ax = b$  是否有解. 具体来说, 用初等行变换把增广矩阵  $[A \ b]$  化成阶梯形, 系数矩阵  $A$  也同时化成了阶梯形. 此时就可以确定解的情形.

**定理 1.4.9 (判定定理)** 1. 如果  $[A \ b]$  对应的阶梯数比  $A$  对应的阶梯数多 1 (此时方程出现了矛盾), 则方程组无解;

2. 如果  $[A \ b]$  和  $A$  对应的阶梯数相等, 则方程组有解. 其中,

- (a) 如果阶梯形的阶梯数和未知数个数相等, 则方程组有唯一解;
- (b) 如果阶梯形的阶梯数小于未知数个数, 则方程组有无穷多组解.

例 1.4.3 中,  $[A \ b]$  的阶梯数是 3, 而  $A$  的阶梯数是 2, 此时方程组无解. 例 1.4.2 中,  $[A \ b]$  和  $A$  的阶梯数都是 2, 未知数个数也是 2, 此时方程组有唯一解. 例 1.4.7 中,  $[A \ b]$  和  $A$  的阶梯数都是 2, 而未知数个数是 4, 此时方程组有无穷多组解.

方程组有解的情形下, 将增广矩阵化成阶梯形进行消元的办法, 称为 **Gauss 消元法**. 化为阶梯形后, 可以自下而上逐个代入从而求出主变量的表达式, 这种方法仍称为 **回代法**; 也可以进一步将增广矩阵从阶梯形化成行简化阶梯形, 通过移项直接写出通解公式, 这种方法称为 **Gauss-Jordan 消元法**. 由于计算在矩阵上进行, Gauss 消元法和 Gauss-Jordan 消元法又称为矩阵消元法.

**例 1.4.10** 求解

$$\begin{cases} x_1 + x_2 - 3x_3 = -1, \\ 2x_1 + x_2 - 2x_3 = 1, \\ x_1 + x_2 + x_3 = 3, \\ x_1 + 2x_2 - 3x_3 = 3. \end{cases}$$

利用第一个方程对  $x_1$  消元, 在增广矩阵上计算:

$$\begin{aligned} \left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 2 & 1 & -2 & 1 \\ 1 & 1 & 1 & 3 \\ 1 & 2 & -3 & 3 \end{array} \right] & \xrightarrow[\text{第 1,2 行}]{\text{倍加变换}} \left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 0 & -1 & 4 & 3 \\ 1 & 1 & 1 & 3 \\ 1 & 2 & -3 & 3 \end{array} \right] & \xrightarrow[\text{第 1,3 行}]{\text{倍加变换}} \left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 0 & -1 & 4 & 3 \\ 0 & 0 & 4 & 4 \\ 1 & 2 & -3 & 3 \end{array} \right] \\ & \xrightarrow[\text{第 1,4 行}]{\text{倍加变换}} \left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 0 & -1 & 4 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 1 & 0 & 4 \end{array} \right], \end{aligned}$$

可以看到, 对矩阵上面的行计算时下面的行不变, 因此只要按照从上到下的次序对行逐个计算, 一次计算若干行也不会有干扰. 因此如果从头开始计算, 可以写成

$$\left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 2 & 1 & -2 & 1 \\ 1 & 1 & 1 & 3 \\ 1 & 2 & -3 & 3 \end{array} \right] \rightarrow \left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 0 & -1 & 4 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 1 & 0 & 4 \end{array} \right] \rightarrow \left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 0 & -1 & 4 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 4 & 7 \end{array} \right] \rightarrow \left[ \begin{array}{ccc|c} 1 & 1 & -3 & -1 \\ 0 & 1 & -4 & -3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 3 \end{array} \right].$$

增广矩阵的阶梯数大于系数矩阵的阶梯数, 故原方程组无解. ☹

**例 1.4.11** 求解

$$\begin{cases} x_1 + x_2 - 3x_4 - x_5 = 0, \\ x_1 - x_2 + 2x_3 - x_4 = 0, \\ 4x_1 - 2x_2 + 6x_3 + 3x_4 - 4x_5 = 0, \\ 2x_1 + 4x_2 - 2x_3 + 4x_4 - 7x_5 = 0, \end{cases}$$

这个方程组的常数项全是 0, 因此不论何种初等行变换, 对应的计算结果全为 0. 所以我们可以只对系数矩阵  $A$  做初等行变换:

$$\left[ \begin{array}{ccccc} 1 & 1 & 0 & -3 & -1 \\ 1 & -1 & 2 & -1 & 0 \\ 4 & -2 & 6 & 3 & -4 \\ 2 & 4 & -2 & 4 & -7 \end{array} \right] \rightarrow \left[ \begin{array}{ccccc} 1 & 1 & 0 & -3 & -1 \\ 0 & -2 & 2 & 2 & 1 \\ 0 & -6 & 6 & 15 & 0 \\ 0 & 2 & -2 & 10 & -5 \end{array} \right] \rightarrow \left[ \begin{array}{ccccc} 1 & 1 & 0 & -3 & -1 \\ 0 & -2 & 2 & 2 & 1 \\ 0 & 0 & 0 & 9 & -3 \\ 0 & 0 & 0 & 12 & -4 \end{array} \right]$$

$$\rightarrow \begin{bmatrix} 1 & 1 & 0 & -3 & -1 \\ 0 & -2 & 2 & 2 & 1 \\ 0 & 0 & 0 & 9 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

化为阶梯形后, 我们知道方程组有无穷多解. 继续化为行简化阶梯形:

$$\cdots \rightarrow \begin{bmatrix} 1 & 1 & 0 & -3 & -1 \\ 0 & 1 & -1 & -1 & -\frac{1}{2} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 & -2 & -\frac{1}{2} \\ 0 & 1 & -1 & -1 & -\frac{1}{2} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 & 0 & -\frac{7}{6} \\ 0 & 1 & -1 & 0 & -\frac{5}{6} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

于是通解为

$$x_1 = -x_3 + \frac{7}{6}x_5, \quad x_2 = x_3 + \frac{5}{6}x_5, \quad x_4 = \frac{1}{3}x_5,$$

其中  $x_1, x_2, x_4$  是主变量,  $x_3, x_5$  是自由变量. ☺

**例 1.4.11** 中方程组的常数项  $\mathbf{b} = \mathbf{0}$ . 方程组  $A\mathbf{x} = \mathbf{0}$ , 称为**齐次**线性方程组. 齐次方程组显然有一组解  $x_1 = 0, \dots, x_n = 0$ , 称为**零解**或**平凡解**. 除此之外其他的解 (如果存在) 称为**非零解**或**非平凡解**. 对应地, 不齐次的线性方程组, 称为**非齐次**线性方程组.

如果  $A\mathbf{x} = \mathbf{0}$  成立, 则  $A(k\mathbf{x}) = k(A\mathbf{x}) = k\mathbf{0} = \mathbf{0}$ . 因此, 齐次方程组如果有一个非零解, 则有无穷多组解. 阶梯形可以用来判断齐次方程组是否有非零解.

**命题 1.4.12** 一个齐次线性方程组只有零解当且仅当其系数矩阵的 (行简化) 阶梯形的阶梯数等于未知数个数. 等价地, 一个齐次线性方程组有无穷多组解当且仅当它的 (行简化) 阶梯形的阶梯数小于未知数个数.

证. 由于常数项全为零, 只需化简系数矩阵. 由定理 1.4.9 立得结论. □

由此容易得到如下推论.

**命题 1.4.13** 一个齐次线性方程组中的方程个数小于未知数个数, 则一定存在非零解.

证. 齐次方程组对应的阶梯形矩阵的阶梯数小于等于方程的个数, 因此小于未知数的个数, 所以一定存在非零解. □

这个对于齐次线性方程组解的定性判断, 不用通过计算化为阶梯形再得到, 因此常会带来一些便利.

## 习题

**练习 1.4.1** 把下列矩阵化为行简化阶梯形，并回答问题.

1.  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 化简后第一列是否为主列?

2.  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 化简后第二列是否为主列?

3.  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 化简后第三列是否为主列?

4.  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 化简后第四列是否为主列?

5.  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 0 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 化简后第五列是否为主列?

**练习 1.4.2** 下列方程组有解时，找到所有解；方程组无解时，对方程组做初等行变换，得到矛盾表达式  $0 = 1$ .

1.  $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$

2.  $\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$

3.  $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

4.  $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$

5.  $\begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$

6.  $\begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

7.  $\begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$

8.  $\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 5 \end{bmatrix}.$

9.  $\begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 5 \end{bmatrix}.$

**练习 1.4.3** 将下列问题首先化成  $Ax = b$  的形式, 然后求所有解.

1. 对  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 如何将其第三列写成前两列的线性组合?

2. 对  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 如何将其第四列写成前两列的线性组合?

3. 对  $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$ , 如何将其第五列写成前两列的线性组合?

**练习 1.4.4** 求证: 齐次线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = 0, \\ a_{21}x_1 + a_{22}x_2 = 0, \end{cases}$$

有非零解当且仅当  $a_{11}a_{22} - a_{12}a_{21} = 0$ .

**练习 1.4.5** 求满足下列条件的常数  $b, c$ .

1. 方程组  $\begin{bmatrix} 1 & 2 \\ 3 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$  无解.

2. 方程组  $\begin{bmatrix} 3 & 2 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 10 \\ c \end{bmatrix}$  无解.

3. 方程组  $\begin{bmatrix} 2 & 5 & 1 \\ 4 & b & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$  无解.

4. 方程组  $\begin{bmatrix} b & 3 \\ 3 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \end{bmatrix}$  有无穷组解.

5. 方程组  $\begin{bmatrix} 2 & b \\ 4 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 16 \\ c \end{bmatrix}$  有无穷组解.

6. 方程组  $\begin{bmatrix} 1 & b & 0 \\ 1 & -2 & -1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$  有非零解.

7. 方程组  $\begin{bmatrix} b & 2 & 3 \\ b & b & 4 \\ b & b & b \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$  有非零解 (求三个不同的  $b$ ).

**练习 1.4.6** 设齐次线性方程组

$$\begin{cases} 2x_1 + x_2 + x_3 = 0, \\ ax_1 - x_3 = 0, \\ -x_1 + 3x_3 = 0, \end{cases}$$

有非零解, 求  $a$  的值, 并求出所有的解.

**练习 1.4.7** 在下列方程组中, 讨论在  $p$  取不同值时方程组是否有解, 并在有解时, 求出所有的解.

$$\begin{cases} px_1 + x_2 + x_3 = 1, \\ x_1 + px_2 + x_3 = p, \\ x_1 + x_2 + px_3 = p^2. \end{cases}$$



**练习 1.4.8** 设  $A = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}$ ,  $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ . 求证:

1.  $A\mathbf{x} = \mathbf{b}$  有解当且仅当  $b_1 + b_2 + b_3 = 0$ .

2. 齐次线性方程组  $A\mathbf{x} = \mathbf{0}$  的解集是  $\{k\mathbf{x}_1 \mid k \in \mathbb{R}\}$ ,  $\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ .

3. 当  $A\mathbf{x} = \mathbf{b}$  有解时, 设  $\mathbf{x}_0$  是一个解, 则解集是  $\{\mathbf{x}_0 + k\mathbf{x}_1 \mid k \in \mathbb{R}\}$ .

**练习 1.4.9** 求三阶方阵  $A$ , 使得线性方程组  $A\mathbf{x} = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}$  的解集是  $\left\{ \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + c \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + d \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mid c, d \in \mathbb{R} \right\}$ .

**练习 1.4.10** 1. 构造三阶方阵, 其元素各不相同, 且行简化阶梯形有且只有一个主元.

2. 构造 100 阶方阵  $A$ , 所有元素非零, 且行简化阶梯形恰有 99 个主元. 试描述  $A\mathbf{x} = \mathbf{0}$  的解集.

**练习 1.4.11** 给定线性方程组  $A\mathbf{x} = \mathbf{0}$ , 其中  $A$  是 100 阶方阵. 假设 Gauss 消元法计算到最后一行得到  $0 = 0$ .

1. 消元法是在计算  $A$  的行的线性组合. 因此  $A$  的 100 行的某个线性组合是?

2. 计算出  $0 = 0$  说明方程组有无穷多解. 因此  $A$  的 100 列的某个线性组合是?

3. 试说明 Gauss 消元法计算出的零行的个数和自由变量的个数相等.

**练习 1.4.12** 仅用从上往下的倍加变换 (即把上面的行的若干倍加到下面的行上), 将下列矩阵化为阶梯形, 并分析其主元的规律.

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

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

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

这些矩阵都是**三对角矩阵**, 即除对角元素及与其相邻的元素外其余元素都是零的方阵.

**练习 1.4.13** 求解

$$\begin{cases} x_1 & & + x_n = 0, \\ x_1 + x_2 & & = 0, \\ & x_2 + x_3 & = 0, \\ & & \ddots & \vdots \\ & & & x_{n-1} + x_n = 0. \end{cases}$$

**练习 1.4.14** 证明用倍加变换与倍乘变换可以实现对换变换.

**练习 1.4.15** 证明定理 1.4.8 第二部分.

**练习 1.4.16** 练习 1.3.8 中的数独矩阵经历哪些行变换或列变换后还是数独矩阵?

**练习 1.4.17 (初等列变换在方程组上的含义)** 考虑方程组  $\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$ . 进行下列换元, 写出  $x', y'$  满足的方程组. 对比原方程组, 其系数矩阵做了哪种初等变换?

1.  $x' = y, y' = x.$

2.  $x' = 2x, y' = y.$

3.  $x' = x, y' = x + y.$

4.  $x' = x + 1, y' = y.$

**练习 1.4.18 (方程、法向量与超平面)** 给定原点  $O$ , 则空间中的点  $P$  与向量  $\overrightarrow{OP}$  一一对应. 空间中所有点构成的集合称为点空间, 上述一一对应是点空间到线性空间  $\mathbb{R}^3$  的双射. 注意, 点空间与线性空间并不相同. 特别地, 空间中的点  $P$  也可用向量  $\overrightarrow{OP} \in \mathbb{R}^3$  表示.

两个  $\mathbb{R}^3$  中的向量垂直当且仅当其内积为零 (练习 1.2.10). 与空间中某个平面垂直的非零向量称为该平面的法向量. 考虑空间中以非零向量  $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$  为法向量的平面.

1. 给定  $p, q \in \mathbb{R}^3$ , 分别对应该平面上的两个点, 证明,  $\begin{bmatrix} a & b & c \end{bmatrix} p = \begin{bmatrix} a & b & c \end{bmatrix} q$ ;

2. 设该平面经过对应于  $p \in \mathbb{R}^3$  的点, 令  $d = \begin{bmatrix} a & b & c \end{bmatrix} p$ , 证明, 平面是方程  $ax + by + cz = d$  的解集.

3. 设  $a_1x + b_1y + c_1z = d_1$  和  $a_2x + b_2y + c_2z = d_2$  的解集平面平行, 试分析两个方程系数的关系.

4. 设  $\begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}, \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix}$  为两个不共线的非零向量, 则  $a_1x + b_1y + c_1z = d_1$  与  $a_2x + b_2y + c_2z = d_2$  的解集平面不平行, 因此必相交于一条直线  $l$ . 对方程做倍加变换, 证明,  $a_1x + b_1y + c_1z = d_1$  与  $(a_1 + a_2)x + (b_1 + b_2)y + (c_1 + c_2)z = d_1 + d_2$  的解集平面的交集还是直线  $l$ . (从几何上看, 倍加变换将一个平面沿相交直线旋转, 因此不改变解集的交集.)

一个  $m$  元线性方程的解集是对应于  $\mathbb{R}^m$  的点空间中的一个超平面, 因此求解线性方程组等价于求若干超平面的交集.

**练习 1.4.19** 对下列问题, 将其化成  $Ax = b$ , 并找到所有解.

1. 考虑例 1.2.1 中的桥墩载荷问题. 假设重物的重力分别为  $F_1 = 2, F_2 = 3$ , 放置的位置  $l_1, l_2$  为未知变量, 桥梁长度为 5. 如果两个桥墩的载荷分别为  $f_1 = 3, f_2 = 2$ , 求所有可能的  $\begin{bmatrix} l_1 \\ l_2 \end{bmatrix}$ .

2. 笼中有若干鸡兔, 每只鸡有一个头两条腿两只翅膀, 每只兔有一个头四条腿没有翅膀. 笼中现有 4 个头、12 条腿、8 只翅膀, 求鸡兔数目.

3. 设  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  满足  $A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$ , 且  $A$  第一列的元素之和为 2, 求所有可能的  $A$ .

4. 求平面上直线  $y = 2x + 3, y = -x + 5$  的交点.

5. 空间中有三个平面, 分别经过点  $\begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$ , 具有法向量  $\begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix}$ , 求这三个平面的交点 (练习 1.4.18).

6. 空间中有一条经过原点的直线, 并且与向量  $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ ,  $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$  垂直, 求所有直线上的点.

7. 空间中有一个平面经过点  $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ ,  $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ ,  $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ , 求所有与该平面垂直的向量.

**练习 1.4.20** 如果线性方程组  $Ax = b$  和  $Cx = b$ , 对任意  $b$  都有相同的解集, 那么  $A = C$  成立么?

**练习 1.4.21** 设线性映射  $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ . 当  $n > m$  时,  $F$  是否可能是单射? 当  $n < m$  时,  $F$  是否可能是满射?

## 1.5 线性映射的运算

### 1.5.1 线性运算

**定义 1.5.1 (映射的线性运算)** 设  $A, B: \mathbb{R}^n \rightarrow \mathbb{R}^m$  是两个线性映射,  $k \in \mathbb{R}$ , 定义两个新的映射:

$$\begin{aligned} A + B: \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ x &\mapsto A(x) + B(x), \\ kA: \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ x &\mapsto kA(x). \end{aligned} \tag{1.5.1}$$

称  $A + B$  为  $A$  与  $B$  的和,  $kA$  为数  $k$  与  $A$  的数乘.

注意, 定义中涉及的运算是陪域  $\mathbb{R}^m$  中的线性运算. 特别地, 由于向量之间不能做乘除法, 两个映射之间也没有乘除运算.

**命题 1.5.2** 映射  $A + B$  和  $kA$  都是线性映射.

证. 根据定义, 这里只验证  $A + B$  的情形:

$$\begin{aligned} (A + B)(x + x') &= A(x + x') + B(x + x') = A(x) + A(x') + B(x) + B(x') \\ &= A(x) + B(x) + A(x') + B(x') = (A + B)(x) + (A + B)(x'), \\ (A + B)(kx) &= A(kx) + B(kx) = kf(x) + k(x) = k(A + B)(x). \end{aligned}$$

其中,  $x, x' \in \mathbb{R}^n, k \in \mathbb{R}$ . □

在  $kA$  中取  $k = 0$ , 得到一个特殊的线性映射, **零映射**, 也用符号  $O: \mathbb{R}^n \rightarrow \mathbb{R}^m$  表示, 满足  $O(x) = 0$ , 对任意  $x \in \mathbb{R}^n$  都成立. 与零向量在线性空间  $\mathbb{R}^n$  中的性质类似,

零映射在映射的加法之下满足:  $O + A = A = A + O$ . 我们也可以定义线性映射的**减法**:  $A - B = A + (-B)$ , 其中  $-B = (-1)B$ .

所有  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射构成一个集合, 这个集合在映射加法和数乘两个运算下封闭. 这两个线性运算也满足类似于命题 1.2.5 中的八条运算法则.

由于线性映射与矩阵的对应, 对矩阵也可以定义类似的线性运算.

**定义 1.5.3 (矩阵的线性运算)** 设  $A = [a_{ij}]_{m \times n}$ ,  $B = [b_{ij}]_{m \times n}$  是两个  $m \times n$  矩阵,  $k \in \mathbb{R}$ , 定义

$$A + B = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix};$$

$$kA = k \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} ka_{11} & \cdots & ka_{1n} \\ \vdots & & \vdots \\ ka_{m1} & \cdots & ka_{mn} \end{bmatrix}.$$

**命题 1.5.4** 矩阵加法和数乘满足如下八条运算法则:

1. 加法结合律:  $(A + B) + C = A + (B + C)$ ;
2. 加法交换律:  $A + B = B + A$ ;
3. 零矩阵:  $A + O_{m \times n} = A$ , 其中  $O_{m \times n}$  是所有元素全为 0 的矩阵, 称为  $m \times n$  **零矩阵**, 简记为  $O$ ;
4. 负矩阵: 对任意  $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ , 记  $-A = \begin{bmatrix} -a_{11} & \cdots & -a_{1n} \\ \vdots & & \vdots \\ -a_{m1} & \cdots & -a_{mn} \end{bmatrix}$ , 它满足  $A + (-A) = O$ , 称它为  $A$  的**负矩阵**;
5. 单位数:  $1A = A$ ;
6. 数乘结合律:  $(kl)A = k(lA)$ ;
7. 数乘对数的分配律:  $(k + l)A = kA + lA$ ;
8. 数乘对矩阵的分配律:  $k(A + B) = kA + kB$ .

根据负矩阵性质, 我们可以定义矩阵的**减法**:  $A - B = A + (-B)$ . 根据加法结合律, 我们不需要区分加法的计算次序, 因此我们可以写  $A + B + C = (A + B) + C = A + (B + C)$ .

矩阵和向量的乘积对向量的线性运算满足分配律 (命题 1.3.7), 这里有如下类似的结果.

**命题 1.5.5** 矩阵和向量的乘积对矩阵的线性运算满足分配律:

$$(k_1 A_1 + k_2 A_2) \mathbf{x} = k_1 (A_1 \mathbf{x}) + k_2 (A_2 \mathbf{x}).$$

**例 1.5.6 (差分矩阵)**  $\mathbb{R}^n$  中的向量  $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$  可以看成是一个数列  $x_1, \dots, x_n$ . 在数列的

相关计算中, 经常需要计算相邻两项的差. 由此我们定义  $\mathbb{R}^n$  上的一个线性变换, 称为**向前差分**:

$$\begin{array}{ccc} \mathbb{R}^n & \rightarrow & \mathbb{R}^n \\ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} & \mapsto & \begin{bmatrix} x_1 \\ x_2 - x_1 \\ \vdots \\ x_n - x_{n-1} \end{bmatrix} \end{array}.$$

它的表示矩阵是

$$D_{nn} = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}, \quad (1.5.2)$$

其中对角元素全是 1, 其下方相邻的对角线上的元素全是  $-1$ , 其余元素全是 0, 这个矩阵称为**向前差分矩阵**. 它因对应变换取每一项与前一项的差而得名. 根据矩阵的线性运算, 差分矩阵可以分解为:

$$D_{nn} = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & & 1 \end{bmatrix} - \begin{bmatrix} & 1 & & & \\ & & \ddots & & \\ & & & & 1 \end{bmatrix} =: I_n - J_n.$$

$$\text{而 } D_{nn} \mathbf{x} = I_n \mathbf{x} - J_n \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} 0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ \vdots \\ x_n - x_{n-1} \end{bmatrix}.$$

我们给出这个线性变换在一些数列上的作用. 从数列  $a_n = n^2$  开始, 做一次向前差

分得到等差数列, 再做一次得到常数数列 (不考虑第一项):

$$\begin{bmatrix} 1 \\ 4 \\ 9 \\ 16 \\ 25 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix};$$

从 Fibonacci 数列  $a_1 = 1, a_2 = 1, a_n = a_{n-1} + a_{n-2}$  开始, 做一次向前差分后, 不考虑前面若干项, 得到的还是 Fibonacci 数列:

$$\begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 5 \\ 8 \\ 13 \\ 21 \\ 34 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 2 \\ 3 \\ 5 \\ 8 \\ 13 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}.$$

类似地, 也可以取每一项与后一项的差, 得到的是**向后差分**和**向后差分矩阵**:

$$D_{nn}^T = \begin{bmatrix} 1 & -1 & & \\ & 1 & \ddots & \\ & & \ddots & -1 \\ & & & 1 \end{bmatrix}. \quad (1.5.3)$$

这里, 上角标  $\cdot^T$  表示  $D_{nn}^T$  是由  $D_{nn}$  对调行和列的关系得到的. ☺

**定义 1.5.7 (转置)** 定义  $m \times n$  矩阵  $A$  的**转置**:

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

即,  $A^T$  是由对调  $A$  的行和列得到的  $n \times m$  矩阵.

特别地, 对应分量相同的列向量与行向量互为转置. 之前我们用  $\mathbf{a}$  表示列向量, 而用  $\mathbf{a}^T$  表示行向量, 就是为了与此保持一致.

从定义来看, 显然有  $(A^T)^T = A$ . 另外,

$$\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}^T = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}. \quad (1.5.4)$$

**命题 1.5.8** 矩阵的线性运算与转置满足如下运算法则:

$$(A + B)^T = A^T + B^T, (kA)^T = kA^T.$$

**定义 1.5.9 (对称矩阵)** 如果矩阵  $A$  满足  $A = A^T$ , 就称之为**对称矩阵**. 如果矩阵  $A$  满足  $A = -A^T$ , 就称之为**反对称矩阵**或**斜对称矩阵**.

显然, 对称矩阵和反对称矩阵一定是方阵.

## 1.5.2 复合

**例 1.5.10 (弹簧振子)** 读者在中学物理中学习过弹簧振子. 所谓弹簧振子, 就是不考虑摩擦阻力、不考虑弹簧质量、不考虑振子的大小和形状的理想化模型. 假设三根弹簧和三个振子如图 1.5.1 交替依次挂在天花板下方, 弹簧的原始长度为  $l_1, l_2, l_3$ , 劲度系数分别为  $k_1, k_2, k_3$ , 振子的质量分别为  $m_1, m_2, m_3$ . 系统处于静止状态, 问三个振子在哪里?

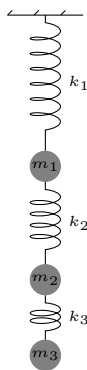


图 1.5.1: 弹簧振子

这一问题并不难. 我们先求三根弹簧的拉伸长度, 设每根弹簧的拉伸长度是  $u_1, u_2, u_3$ . 对振子  $m_3$  受力分析, 利用 Hooke 定律, 易得  $k_3 u_3 = m_3 g$ , 于是  $u_3 = g \frac{m_3}{k_3}$ . 对振子  $m_2, m_3$  整体受力分析, 易得  $k_2 u_2 = m_2 g + m_3 g$ , 于是  $u_2 = g \frac{m_2 + m_3}{k_2}$ . 对振子

$m_1, m_2, m_3$  整体受力分析, 易得  $k_1 u_1 = m_1 g + m_2 g + m_3 g$ , 于是  $u_1 = g \frac{m_1 + m_2 + m_3}{k_1}$ . 这样我们就知道  $m_1$  到天花板的距离是  $u_1 + l_1 = g \frac{m_1 + m_2 + m_3}{k_1} + l_1$ ,  $m_2$  到天花板的距离是  $u_1 + l_1 + u_2 + l_2 = g \frac{m_1 + m_2 + m_3}{k_1} + l_1 + g \frac{m_2 + m_3}{k_2} + l_2$ ,  $m_3$  到天花板的距离是  $u_1 + l_1 + u_2 + l_2 + u_3 + l_3 = g \frac{m_1 + m_2 + m_3}{k_1} + l_1 + g \frac{m_2 + m_3}{k_2} + l_2 + g \frac{m_3}{k_3} + l_3$ .

可以看到, 如果我们增加弹簧和振子的数目, 解将越来越复杂, 虽然很有规律, 但是并不容易用简单的公式表示.

我们用线性映射和矩阵的框架来分析. 首先我们可以看到,  $m_1$  离天花板至少有  $l_1$  那么远,  $m_2$  离天花板至少有  $l_1 + l_2$  那么远,  $m_3$  离天花板至少有  $l_1 + l_2 + l_3$  那么远. 因此我们没有必要去计算振子到天花板的距离. 我们假设  $m_1$  到天花板的距离是  $l_1 + x_1$ ,  $m_2$  的是  $l_1 + l_2 + x_2$ ,  $m_3$  的是  $l_1 + l_2 + l_3 + x_3$ .

把三个振子的位移  $x_1, x_2, x_3$  看作系统的输入, 它们所受的拉力  $f_1, f_2, f_3$  看作系统的输出. 这个系统可以描述成以下三个系统的叠加:

输入 = 振子的位移  $\rightarrow$  弹簧的拉伸长度  $\rightarrow$  弹簧的拉力  $\rightarrow$  振子所受的拉力 = 输出.

我们把输入写成向量  $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ . 那么弹簧的拉伸长度

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -1 & 1 & \\ & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = D_{33} \mathbf{x}.$$

从  $\mathbf{x}$  到  $\mathbf{u}$  构成了一个线性映射, 其表示矩阵是向前差分矩阵  $D_{33}$ .

下一步, 三根弹簧提供的拉力就是

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} k_1 u_1 \\ k_2 u_2 \\ k_3 u_3 \end{bmatrix} = \begin{bmatrix} k_1 & & \\ & k_2 & \\ & & k_3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = K_3 \mathbf{u}.$$

从  $\mathbf{u}$  到  $\mathbf{y}$  构成了一个线性映射, 其表示矩阵为对角矩阵  $K_3$ .

最后, 振子所受拉力为

$$\mathbf{f} = \begin{bmatrix} y_1 - y_2 \\ y_2 - y_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & \\ & 1 & -1 \\ & & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = D_{33}^T \mathbf{y}.$$

从  $\mathbf{y}$  到  $\mathbf{f}$  构成了一个线性映射, 其表示矩阵为向后差分矩阵  $D_{33}^T$ .



记振子的质量向量为  $\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}$ , 则振子所受重力为  $\mathbf{g} = g\mathbf{m} = \begin{bmatrix} m_1 g \\ m_2 g \\ m_3 g \end{bmatrix}$ . 利用力

的平衡有  $\mathbf{f} = \mathbf{g}$ . 注意  $D_{33}, K_3, D_{33}^T$  都是对角元素都不为零的三角矩阵, 因此, 利用前代法和回代法, 我们可以逐步倒推, 由  $\mathbf{f}$  反解出  $\mathbf{x}$ .  $\odot$

另一方面, 在图 1.5.1 中, 如果我们考虑三个系统的叠加, 把三个线性映射的复合映射记为  $\mathbf{A}: \mathbb{R}^3 \rightarrow \mathbb{R}^3, \mathbf{A}(\mathbf{x}) = \mathbf{f}$ . 因此, 只需解方程  $\mathbf{A}(\mathbf{x}) = \mathbf{g}$ . 这里有一个自然的问题: 这个复合映射  $\mathbf{A}$  是线性映射吗? 如果是, 其表示矩阵和三个线性映射的表示矩阵有什么关系?

**命题 1.5.11** 设  $\mathbf{B}: \mathbb{R}^n \rightarrow \mathbb{R}^m, \mathbf{A}: \mathbb{R}^m \rightarrow \mathbb{R}^l$  是两个线性映射, 则二者的复合映射  $\mathbf{A} \circ \mathbf{B}: \mathbb{R}^n \rightarrow \mathbb{R}^l$  也是一个线性映射.

证. 根据定义, 直接验证复合映射保持线性运算, 即对任意  $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n, k \in \mathbb{R}$ ,

$$\begin{aligned} (\mathbf{A} \circ \mathbf{B})(\mathbf{x} + \mathbf{x}') &= \mathbf{A}(\mathbf{B}(\mathbf{x} + \mathbf{x}')) = \mathbf{A}(\mathbf{B}(\mathbf{x}) + \mathbf{B}(\mathbf{x}')) = (\mathbf{A} \circ \mathbf{B})(\mathbf{x}) + (\mathbf{A} \circ \mathbf{B})(\mathbf{x}'), \\ (\mathbf{A} \circ \mathbf{B})(k\mathbf{x}) &= \mathbf{A}(\mathbf{B}(k\mathbf{x})) = \mathbf{A}(k\mathbf{B}(\mathbf{x})) = k\mathbf{A}(\mathbf{B}(\mathbf{x})) = k(\mathbf{A} \circ \mathbf{B})(\mathbf{x}). \end{aligned} \quad \square$$

设  $\mathbf{A}$  和  $\mathbf{B}$  的表示矩阵分别为  $l \times m$  矩阵  $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_m]$  和  $m \times n$  矩阵  $\mathbf{B} = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$ , 那么根据定义, 线性映射  $\mathbf{A} \circ \mathbf{B}$  的表示矩阵  $\mathbf{C}$  可以计算出来:

$$\begin{aligned} \mathbf{C} &= [\mathbf{A}(\mathbf{B}(\mathbf{e}_1)) \ \mathbf{A}(\mathbf{B}(\mathbf{e}_2)) \ \cdots \ \mathbf{A}(\mathbf{B}(\mathbf{e}_n))] \\ &= [\mathbf{A}(\mathbf{b}_1) \ \mathbf{A}(\mathbf{b}_2) \ \cdots \ \mathbf{A}(\mathbf{b}_n)] \\ &= [\mathbf{A}\mathbf{b}_1 \ \mathbf{A}\mathbf{b}_2 \ \cdots \ \mathbf{A}\mathbf{b}_n]. \end{aligned}$$

**定义 1.5.12 (矩阵乘法)** 给定  $l \times m$  矩阵  $\mathbf{A}$  和  $m \times n$  矩阵  $\mathbf{B} = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$ , 定义二者乘积  $\mathbf{AB}$  是如下  $l \times n$  矩阵:

$$\mathbf{AB} = \mathbf{A} [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n] := [\mathbf{A}\mathbf{b}_1 \ \cdots \ \mathbf{A}\mathbf{b}_n]. \quad (1.5.5)$$

矩阵乘法的定义是矩阵与列向量乘法的一个自然推广, 因为矩阵与作为  $n \times 1$  矩阵的  $n$  维列向量乘法已经如前定义. 特别需要注意的是,

1. 两个矩阵  $\mathbf{A}, \mathbf{B}$  能够相乘, 亦即乘积矩阵  $\mathbf{AB}$  能够存在, 必须要求  $\mathbf{A}$  的列数与  $\mathbf{B}$  的行数相同; 此时,  $\mathbf{AB}$  的行数等于  $\mathbf{A}$  的行数,  $\mathbf{AB}$  的列数等于  $\mathbf{B}$  的列数. 这与映射的复合一致, 而这对应着: 如果  $\mathbf{A}$  和  $\mathbf{B}$  可以复合成  $\mathbf{A} \circ \mathbf{B}$ , 那么  $\mathbf{A}$  的定义域必须等于  $\mathbf{B}$  的陪域.

2. 乘积矩阵  $AB$  的第  $k$  列是  $A\mathbf{b}_k$ , 这说明, 只有  $B$  的第  $k$  列  $\mathbf{b}_k$  会影响  $AB$  的第  $k$  列, 而  $B$  的其他列不会影响  $AB$  的第  $k$  列.

类似地, 我们还可以得到  $AB$  的行向量的性质. 根据 (1.3.3) 中矩阵与列向量的乘积, 设

$$A = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \vdots \\ \tilde{\mathbf{a}}_l^T \end{bmatrix}, B = [\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_n], \text{ 我们得到:}$$

$$AB = [A\mathbf{b}_1 \quad \cdots \quad A\mathbf{b}_n] = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{b}_1 & \cdots & \tilde{\mathbf{a}}_1^T \mathbf{b}_n \\ \vdots & & \vdots \\ \tilde{\mathbf{a}}_l^T \mathbf{b}_1 & \cdots & \tilde{\mathbf{a}}_l^T \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T B \\ \vdots \\ \tilde{\mathbf{a}}_l^T B \end{bmatrix}. \quad (1.5.6)$$

这说明,  $AB$  的第  $i$  行是  $\tilde{\mathbf{a}}_i^T B$ , 仅由矩阵  $A$  的第  $i$  行  $\tilde{\mathbf{a}}_i^T$  和矩阵  $B$  决定. 具体到  $C = AB$  的每个元素, 设  $A = [a_{ij}]_{l \times m}$ ,  $B = [b_{jk}]_{m \times n}$ ,  $C = [c_{ik}]_{l \times n}$ , 我们有

$$c_{ik} = \tilde{\mathbf{a}}_i^T \mathbf{b}_k = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{im}b_{mk} = \sum_{j=1}^m a_{ij}b_{jk}. \quad (1.5.7)$$

这里我们引入了求和号  $\sum$ , 其中求和号后面是一个带变动下角标  $j$  的项, 求和号上下提示了变动下角标及其变化范围, 即求和式表示变动下角标取遍变化范围的对应的所有项的和. 注意, 这里一共涉及三个下角标  $i, j, k$ , 求和仅仅针对变动下角标  $j$ , 而下角标  $i, k$  都固定不变. 因此, 矩阵乘法可直接有

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{l1} & \cdots & a_{lm} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^m a_{1j}b_{j1} & \cdots & \sum_{j=1}^m a_{1j}b_{jn} \\ \vdots & & \vdots \\ \sum_{j=1}^m a_{lj}b_{j1} & \cdots & \sum_{j=1}^m a_{lj}b_{jn} \end{bmatrix}.$$

注意, (1.5.5)–(1.5.7) 都可以作为矩阵乘法的定义, 因为三者互相等价.

矩阵乘积的公式在计算中必不可少, 而它和映射复合之间的对应关系, 对我们的理解十分重要. 根据定义 1.5.12, 可以直接得出下面的结果.

**命题 1.5.13** 设  $\mathbf{A}: \mathbb{R}^m \rightarrow \mathbb{R}^l$ ,  $\mathbf{B}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  是两个线性映射, 记二者的表示矩阵分别为  $A$  和  $B$ , 则复合映射  $\mathbf{A} \circ \mathbf{B}: \mathbb{R}^n \rightarrow \mathbb{R}^l$  的表示矩阵为  $C = AB$ .

一个直接的推论是, 对任意列向量  $\mathbf{x} \in \mathbb{R}^n$ , 都有

$$(AB)\mathbf{x} = (\mathbf{A} \circ \mathbf{B})(\mathbf{x}) = \mathbf{A}(\mathbf{B}(\mathbf{x})) = \mathbf{A}(B\mathbf{x}) = A(B\mathbf{x}), \quad (1.5.8)$$

即矩阵与列向量乘积和矩阵乘积有结合律. 更一般地, 我们有如下重要的运算法则.

**命题 1.5.14** 矩阵乘法满足结合律:  $A(BC) = (AB)C$ .

证一. 令  $C = [c_1 \ \dots \ c_n]$ , 则

$$\begin{aligned} A(BC) &= A(B[c_1 \ \dots \ c_n]) \\ &= A[Bc_1 \ \dots \ Bc_n] \\ &= [A(Bc_1) \ \dots \ A(Bc_n)] \quad (\text{由 (1.5.8)}) \\ &= [(AB)c_1 \ \dots \ (AB)c_n] \\ &= (AB)[c_1 \ \dots \ c_n] = (AB)C. \end{aligned} \quad \square$$

证二. 设  $A, B, C$  是三个线性映射, 其表示矩阵分别为  $A, B, C$ , 则复合映射  $A \circ B$  和  $B \circ C$  的表示矩阵分别为  $AB$  和  $BC$ . 因此,  $(A \circ B) \circ C$  和  $A \circ (B \circ C)$  的表示矩阵分别为  $(AB)C$  和  $A(BC)$ . 映射的复合具有结合律  $(A \circ B) \circ C = A \circ (B \circ C)$ , 因此表示矩阵相等, 即  $(AB)C = A(BC)$ .  $\square$

根据命题 1.5.14 的乘法结合律, 我们不需要区分乘法的计算次序, 因此我们可以写  $ABC = (AB)C = A(BC)$ .

矩阵乘法与数的乘法最重要的区别在于, 乘法交换律和消去律一般不成立:

1.  $AB, BA$  不一定同时存在; 即便都存在, 仍然需要  $A, B$  是同阶方阵, 才能保证  $AB, BA$  是同阶方阵.
2. 即使  $A, B, AB, BA$  都是同阶的方阵, 也存在  $A, B$  使得  $AB \neq BA$ , 例如

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

3. 两个非零矩阵的乘积**可能**是零矩阵. 例如,

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

因此  $AB = 0$  **不能**推出  $A = 0$  或  $B = 0$ . 更一般地, 消去律也**不一定**成立:  $AB = AC, A \neq 0$  不能直接推出  $B = C$ .

由于交换律一般不成立, 矩阵乘法的顺序十分重要. 乘积  $AB$ , 我们常读作“ $A$  (左) 乘  $B$ ”或者“ $B$  右乘  $A$ ”.

如果  $AB = BA$ , 我们称  $A, B$  **可交换**. 此时  $A, B$  一定是同阶方阵.

注意, 矩阵乘法的特殊之处, 和映射复合的性质类似. 例如, 两个一般的映射不一定可以复合; 即便是同一集合上的两个变换, 两种不同顺序的复合得到的变换也不一定相同. 类似地, 如果两个映射  $A, B$  满足  $B \circ A = A \circ B$ , 则称  $A, B$  **可交换**.

**例 1.5.15** 考虑例 1.3.9 中平面向量构成的线性空间  $\mathbb{R}^2$  上的线性变换的复合.

1. 旋转变换的复合还是旋转变换:  $\mathbf{R}_{\theta_1} \circ \mathbf{R}_{\theta_2} = \mathbf{R}_{\theta_1+\theta_2} = \mathbf{R}_{\theta_2} \circ \mathbf{R}_{\theta_1}$ . 对应的矩阵乘法是:

$$\begin{bmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{bmatrix} \begin{bmatrix} \cos \theta_2 & -\sin \theta_2 \\ \sin \theta_2 & \cos \theta_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta_1 + \theta_2) & -\sin(\theta_1 + \theta_2) \\ \sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{bmatrix}.$$

特别地,  $\mathbf{R}_{\theta} \circ \mathbf{R}_{-\theta} = \mathbf{R}_{-\theta} \circ \mathbf{R}_{\theta} = \mathbf{R}_0 = \mathbf{I}$ .

2. 反射变换与自己的复合是恒同变换:  $\mathbf{H}_{\theta} \circ \mathbf{H}_{\theta} = \mathbf{I}$ . 对应的矩阵乘法是:

$$\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2.$$

3. 对换变换与旋转变换的复合: 先计算  $\mathbf{P} \circ \mathbf{R}_{\theta}$  和  $\mathbf{R}_{\theta} \circ \mathbf{P}$ . 我们直接计算对应的矩阵乘法,

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} \sin \theta & \cos \theta \\ \cos \theta & -\sin \theta \end{bmatrix},$$

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -\sin \theta & \cos \theta \\ \cos \theta & \sin \theta \end{bmatrix}.$$

显然二者不相等. 事实上, 二者的关系是正弦的系数互为相反数而余弦的系数相等, 而这恰好是  $\theta$  换成  $-\theta$  的效果. 因此, 我们有  $\mathbf{P} \circ \mathbf{R}_{\theta} = \mathbf{R}_{-\theta} \circ \mathbf{P}$ . 从几何直观上看, 旋转再翻折, 等价于翻折再反向旋转. 类似地, 我们可以验证, 反射变换与旋转变换之间满足  $\mathbf{H}_{\theta_2} \circ \mathbf{R}_{\theta_1} = \mathbf{R}_{-\theta_1} \circ \mathbf{H}_{\theta_2}$ .

4. 伸缩变换的复合还是伸缩变换:  $\mathbf{C}_{k_1} \circ \mathbf{C}_{k_2} = \mathbf{C}_{k_1 k_2} = \mathbf{C}_{k_2} \circ \mathbf{C}_{k_1}$ , 对应的矩阵乘法是:

$$\begin{bmatrix} k_1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k_2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} k_1 k_2 & 0 \\ 0 & 1 \end{bmatrix}.$$

特别地, 投影变换与自己的复合保持不变:  $\mathbf{C}_0 \circ \mathbf{C}_0 = \mathbf{C}_0$ .

5. 错切变换的复合还是错切变换:  $\mathbf{S}_{k_1} \circ \mathbf{S}_{k_2} = \mathbf{S}_{k_1+k_2} = \mathbf{S}_{k_2} \circ \mathbf{S}_{k_1}$ . 对应的矩阵乘法是:

$$\begin{bmatrix} 1 & 0 \\ k_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ k_2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ k_1 + k_2 & 1 \end{bmatrix}.$$

6. 我们还可以考虑不同方向的伸缩变换、错切变换的复合, 留给读者思考.

需要指出的是,一般情形下,  $A$  与  $B$  不可交换; 而这里经常出现交换的原因是矩阵具有特殊性.  $\odot$

我们还关心映射的线性运算与复合之间的关系, 以及矩阵的线性运算和乘法之间的关系.

**命题 1.5.16** 设  $B, B': \mathbb{R}^n \rightarrow \mathbb{R}^m, A, A': \mathbb{R}^m \rightarrow \mathbb{R}^l$  是线性映射,  $k \in \mathbb{R}$ , 则有如下运算法则:

1. 对加法的分配律:  $(A + A') \circ B = A \circ B + A' \circ B, A \circ (B + B') = A \circ B + A \circ B'$ ;
2. 对数乘的交换律:  $k(A \circ B) = (kA) \circ B = A \circ (kB)$ .

**命题 1.5.17** 矩阵的乘法还满足如下运算法则:

1. 对加法的分配律:  $(A + B)C = AC + BC, A(B + C) = AB + AC$ ;
2. 对数乘的交换律:  $k(AB) = (kA)B = A(kB)$ .

上面两个命题的验证是直接的, 留给读者. 特别地,  $kI_n$  称为数量矩阵, 因为  $kI_n$  在矩阵乘法下的作用和数乘一样:  $(kI_n)A = k(I_n A) = kA$ .

矩阵的乘法和转置之间, 有如下关系.

**命题 1.5.18** 设  $A, B$  分别是  $l \times m$  矩阵和  $m \times n$  矩阵,  $\mathbf{a}, \mathbf{x}$  是  $m$  维列向量, 则

1.  $\mathbf{a}^T \mathbf{x} = \mathbf{x}^T \mathbf{a}$ ,
2.  $(A\mathbf{x})^T = \mathbf{x}^T A^T$ ;
3.  $(AB)^T = B^T A^T$ .

证. 1. 设  $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$ , 计算可得

$$\mathbf{a}^T \mathbf{x} = a_1 x_1 + \cdots + a_m x_m = x_1 a_1 + \cdots + x_m a_m = \mathbf{x}^T \mathbf{a}.$$

2. 设  $A = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \vdots \\ \tilde{\mathbf{a}}_l^T \end{bmatrix}$ , 则  $A\mathbf{x} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{x} \\ \vdots \\ \tilde{\mathbf{a}}_l^T \mathbf{x} \end{bmatrix}$ , 其中,  $\tilde{\mathbf{a}}_i, \mathbf{x}$  都是  $m$  维列向量, 满足  $\tilde{\mathbf{a}}_i^T \mathbf{x} = \mathbf{x}^T \tilde{\mathbf{a}}_i$ .

根据 (1.5.4), 我们有

$$(A\mathbf{x})^T = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{x} \\ \vdots \\ \tilde{\mathbf{a}}_l^T \mathbf{x} \end{bmatrix}^T = \begin{bmatrix} \mathbf{x}^T \tilde{\mathbf{a}}_1 & \cdots & \mathbf{x}^T \tilde{\mathbf{a}}_l \end{bmatrix} = \mathbf{x}^T \begin{bmatrix} \tilde{\mathbf{a}}_1 & \cdots & \tilde{\mathbf{a}}_l \end{bmatrix} = \mathbf{x}^T A^T.$$

3. 设  $B = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$ , 则

$$(AB)^T = [A\mathbf{b}_1 \ \cdots \ A\mathbf{b}_n]^T = \begin{bmatrix} (A\mathbf{b}_1)^T \\ \vdots \\ (A\mathbf{b}_n)^T \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^T A^T \\ \vdots \\ \mathbf{b}_n^T A^T \end{bmatrix} = B^T A^T. \quad \square$$

命题 1.5.14 和 1.5.18 的证明, 都把矩阵写成行向量组或者列向量组的形式来进行计算, 请读者仔细体会.

最后, 我们回到本小节开头关于弹簧振子的例 1.5.10. 系统对应的映射可以分解为三个映射的复合, 其矩阵分别为  $D_{33}^T, K_3, D_{33}$ , 因此复合映射的矩阵为

$$\begin{aligned} D_{33}^T K_3 D_{33} &= \begin{bmatrix} 1 & -1 & \\ & 1 & -1 \\ & & 1 \end{bmatrix} \begin{bmatrix} k_1 & & \\ & k_2 & \\ & & k_3 \end{bmatrix} \begin{bmatrix} 1 & & \\ -1 & 1 & \\ & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} k_1 + k_2 & -k_2 & \\ -k_2 & k_2 + k_3 & -k_3 \\ & -k_3 & k_3 \end{bmatrix}. \end{aligned}$$

可以看到, 从映射和矩阵的观点考虑问题, 不仅可以很容易地把所列方程推广到任意数目的弹簧和振子上, 而且能够更清楚地表示物理量之间的关系.  $K_3$  反映了弹簧的性质,  $m$  反映了振子的性质. 因此, 当我们对向量和矩阵熟练之后, 我们很容易就可以直接列出方程  $gm = D_{33}^T K_3 D_{33} \mathbf{x}$ , 而不再需要通过分量的计算, 分量的数目也就变得不那么重要了.

## 习题

练习 1.5.1 设  $A, B, C$  分别是  $3 \times 5, 5 \times 3, 3 \times 1$  矩阵, 则  $BA, AB, BC^T, A(B+C)$  中哪些定义良好?

练习 1.5.2 设  $A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \\ 0 & 2 & -2 \end{bmatrix}$ , 求  $AB, BA, AB - BA$ .

练习 1.5.3 计算.

$$\begin{aligned} 1. & \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix}. & 2. & \begin{bmatrix} -7 & 3 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}. & 3. & \begin{bmatrix} 17 & -6 \\ 35 & -12 \end{bmatrix}^5. \\ 4. & \begin{bmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \\ 3 & 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 2 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 & -1 \\ 1 & 1 & -1 \\ -2 & -5 & 4 \end{bmatrix}. & 5. & \begin{bmatrix} 8 & 6 & -6 \\ 4 & 6 & -4 \\ 8 & 8 & -6 \end{bmatrix}^6. \end{aligned}$$

练习 1.5.4 考虑下列  $\mathbb{R}^2$  上的线性变换.

1. 设  $\mathbf{R}_\theta(\mathbf{x}) = R_\theta \mathbf{x}$ , 其中  $R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ . 求证:

(a)  $\mathbf{R}_\theta$  是绕原点逆时针旋转角度  $\theta$  的变换.

(b) 分析当  $\theta$  取何值时, 存在常数  $\lambda$  和非零向量  $\mathbf{x}$ , 满足  $R_\theta \mathbf{x} = \lambda \mathbf{x}$ .

(c) 计算  $R_\theta^n, n > 0$ .

2. 设  $\mathbf{H}_\theta(\mathbf{x}) = H_\theta \mathbf{x}$ , 其中  $H_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$ , 而  $\mathbf{v} = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \mathbf{w} = \begin{bmatrix} \sin \theta \\ -\cos \theta \end{bmatrix}$  是  $\mathbb{R}^2$  中的向量. 求证:

(a)  $\mathbf{v}^T \mathbf{w} = 0, \mathbf{v}^T \mathbf{v} = \mathbf{w}^T \mathbf{w} = 1; H_\theta \mathbf{v} = \mathbf{v}, H_\theta \mathbf{w} = -\mathbf{w}$ . 试分析变换  $\mathbf{H}_\theta$  的几何意义.

(b)  $H_\theta^2 = I_2, H_\theta = I_2 - 2\mathbf{w}\mathbf{w}^T$ .

(c)  $R_{-\theta} H_\phi R_\theta = H_{\phi-\theta}, H_\phi R_\theta H_\phi = R_{-\theta}$ , 并分析其几何意义.

3. 设  $\mathbf{S}(\mathbf{x}) = S\mathbf{x}$ , 其中  $S = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ . 设  $\lambda$  是常数, 求证:  $S\mathbf{x} = \lambda \mathbf{x}$  有非零解当且仅当  $\lambda = 1$ , 并求出所有的非零解; 计算  $S^k, k > 0$ .

练习 1.5.5 计算.

$$1. \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}^2.$$

$$2. \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}^k.$$

练习 1.5.6 设  $E_{ij}$  是  $(i, j)$  元为 1, 其余元素为 0 的矩阵, 而  $n$  阶方阵  $J_n = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$ .

1. 计算  $A = I_2 + J_2, B = I_2 + J_2^T, C = J_2 - J_2^T$ .

2. 先计算  $AB$  再计算  $(AB)C$ , 然后先计算  $BC$  再计算  $A(BC)$ .

3. 先计算  $AB, AC$  再计算  $AB + AC$ , 然后先计算  $B + C$  再计算  $A(B + C)$ .

4. 计算  $AB + BC, B(A + C), (A + C)B$ , 三者是否相等?

5. 计算  $(A + B)^2, A^2 + 2AB + B^2, A^2 + AB + BA + B^2$ , 三者是否相等?

6. 计算  $(AB)^2$  与  $A^2 B^2$ , 二者是否相等?

7. 计算  $E_{13} - \mathbf{e}_1 \mathbf{e}_3^T, \mathbf{e}_1 \mathbf{e}_3^T + \cdots + \mathbf{e}_4 \mathbf{e}_4^T, E_{13} E_{32}, E_{32} E_{13}$ , 其中  $\mathbf{e}_i \in \mathbb{R}^4, E_{ij}$  是 4 阶方阵.

$$8. \text{ 计算 } \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

9. 计算  $J_4 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ ,  $J_4^T \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ ,  $J_4^2 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ ,  $J_4^4$ .

10. 计算  $J_n^k$ , 其中  $k$  是正整数.

11. 设  $P = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}$ , 计算  $PJ_4 + J_4P$ ,  $(P + I_4)(2P + 3I_4) - (2P + 3I_4)(P + I_4)$ .

12. 设  $R = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ ,  $S = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $T = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix}$ , 计算  $SRT$ ,  $SR^T T$ .

**练习 1.5.7** 设 2 阶矩阵  $B$  的  $(1, 2)$  元增加 1, 对下列矩阵讨论该矩阵的哪行哪列一定不变, 并举例说明所有其他行列确实可以变化.

1.  $A + B$ .

2.  $AB$ .

3.  $BA$ .

4.  $B^2$ .

**练习 1.5.8** 判断对错.

1. 如果  $B$  的第一列等于第三列, 则  $AB$  的第一列等于第三列.

2. 如果  $B$  的第一列等于第三列, 则  $BA$  的第一列等于第三列.

3. 如果  $A$  的第一行等于第三行, 则  $ABC$  的第一行等于第三行.

**练习 1.5.9** 求所有满足条件的矩阵  $B$ .

1. 对任意 3 阶方阵  $A$ ,  $BA = 4A$ .

2. 对任意 3 阶方阵  $A$ ,  $BA = 4B$ .

3. 对任意 3 阶方阵  $A$ ,  $BA$  的每一行都是  $A$  的第一行.

4. 对任意 3 阶方阵  $A$ ,  $AB$  的每一行的每一个元素都是  $A$  的对应行的平均值.

5.  $B^2 \neq O$ ,  $B^3 = O$ .

6.  $B \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} B$ .

**练习 1.5.10** 给定矩阵  $A = \begin{bmatrix} p & 0 \\ q & r \end{bmatrix}$ ,  $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ ,  $C = \begin{bmatrix} 0 & z \\ 0 & 0 \end{bmatrix}$ .

1.  $p, q, r$  取何值时, 有  $AB = BA$ ?

2.  $z$  取何值时, 有  $BC = CB$ ?

3.  $p, q, r, z$  取何值时, 有  $ABC = CAB$ ?

**练习 1.5.11** 证明,

1. 设  $n$  维向量  $x$  的每个分量都是 1, 则  $n$  阶方阵  $A$  的各行元素之和为 1 当且仅当  $Ax = x$ .



2. 若  $n$  阶方阵  $A, B$  的各行元素之和均为 1, 则  $AB$  的各行元素之和也均为 1.
3. 若  $n$  阶方阵  $A, B$  的各列元素之和均为 1, 则  $AB$  的各列元素之和也均为 1.

**练习 1.5.12** 证明命题 1.5.17.

**练习 1.5.13** 证明上三角矩阵对加法、数乘、乘法封闭, 即: 设  $U_1, U_2$  是  $n$  阶上三角矩阵,  $k$  是实数, 则  $U_1 + U_2, kU_1, U_1U_2$  都是上三角矩阵. 此外,  $U_1U_2$  的对角元素是  $U_1, U_2$  对应的对角元素的乘积.

**练习 1.5.14** 设  $A$  是  $n$  阶上三角矩阵. 求证:  $A^n = O$  当且仅当  $A$  是严格上三角矩阵.

**练习 1.5.15** 证明, 如果  $n$  阶方阵  $A$  满足  $A^n = O$ , 则  $(I_n - A)(I_n + A + A^2 + \cdots + A^{n-1}) = I_n$ .

**练习 1.5.16** 对  $n$  阶方阵  $A$ , 考虑集合  $\text{Com}(A) = \{n \text{ 阶方阵 } B \mid AB = BA\}$ .

1. 求证:  $\text{Com}(A)$  是所有  $n$  阶方阵的集合当且仅当  $A$  是数量矩阵  $kI_n$ .
2. 设  $A = \text{diag}(d_i)$ ,  $d_i$  互不相同, 求  $\text{Com}(A)$ .
3. 求证: 任取  $B, C \in \text{Com}(A)$ , 都有  $I_n, kB + lC, BC \in \text{Com}(A)$ .

$$4. \text{ 设 } A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \text{ 求证: } J_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \text{Com}(A), \text{ 而且}$$

$$\text{Com}(A) = \{k_1 I_3 + k_2 J_3 + k_3 J_3^2 \mid k_1, k_2, k_3 \in \mathbb{R}\}.$$

**练习 1.5.17** 设  $A = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$ , 求  $A^k$ , 其中  $k$  是正整数.

**练习 1.5.18** 设  $A$  是  $n$  阶方阵, 求证:  $A + A^T, AA^T, A^T A$  都是对称矩阵, 而  $A - A^T$  是反对称矩阵.

**练习 1.5.19** 求证:

1. 任意方阵  $A$  都唯一地表为  $A = B + C$ , 其中  $B$  是对称矩阵,  $C$  是反对称矩阵.
2.  $n$  阶方阵  $A$  是反对称矩阵当且仅当对任意  $n$  维列向量  $x$ , 都有  $x^T A x = 0$ .
3. 设  $A, B$  是  $n$  阶对称矩阵, 则  $A = B$  当且仅当对任意  $n$  维列向量  $x$ , 都有  $x^T A x = x^T B x$ .

**练习 1.5.20** 设  $A, B$  是同阶对称矩阵. 求证:  $AB$  是对称矩阵当且仅当  $AB = BA$ .

**练习 1.5.21** 设  $A$  是实对称矩阵, 如果  $A^2 = O$ , 求证:  $A = O$ .

**练习 1.5.22 (矩阵的迹)** 方阵  $A$  的对角元素的和称为它的迹, 记作  $\text{trace}(A)$ . 验证下列性质.

1. 对任意同阶方阵  $A, B$ ,  $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$ .
2. 对任意方阵  $A$  与实数  $k$ ,  $\text{trace}(kA) = k \text{trace}(A)$ .
3. 对  $m$  阶单位矩阵  $I_m$ ,  $\text{trace}(I_m) = m$ .
4. 对任意方阵  $A$ ,  $\text{trace}(A) = \text{trace}(A^T)$ .

5. 设  $A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$ ,  $B = \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$ , 则  $\text{trace}(A^T B) = \sum_{i=1}^4 a_i b_i$ . 如果  $A$  是  $m$  阶方阵呢?
6. 设  $\mathbf{v}, \mathbf{w}$  是  $m$  维向量, 则  $\text{trace}(\mathbf{v}^T \mathbf{w}) = \text{trace}(\mathbf{w} \mathbf{v}^T)$ .
7. 设  $A, B$  分别是  $m \times n, n \times m$  矩阵, 则  $\text{trace}(AB) = \text{trace}(BA)$ .
8. 设  $A, B$  是任意  $m$  阶方阵, 则  $AB - BA \neq I_m$ .

**练习 1.5.23 (差分矩阵与求导)** 在微积分中,  $f(x)$  的导数  $f'(x) \approx \frac{f(x+\delta) - f(x)}{\delta}$ . 对数列  $a_n$  的相邻两项做减法, 可以看作是一种离散导数  $\frac{a_{n+1} - a_n}{1}$ . 两者具有某种共性. 令  $D$  为 100 阶向前差分矩阵, 而

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_{100} \end{bmatrix}. \text{ 求证:}$$

1. 若  $a_k$  是关于  $k$  的 3 次多项式, 则除第 1 个分量外,  $D\mathbf{a}$  的第  $k$  个分量是关于  $k$  的 2 次多项式.
2. 若  $a_k = e^k$ , 则除第 1 个分量外,  $\frac{e}{e-1} D\mathbf{a}$  的第  $k$  个分量是  $e^k$ .

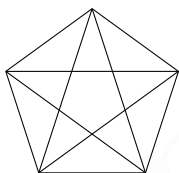


图 1.5.2: 练习 1.5.24

**练习 1.5.24** 如图 1.5.2 的电路包含 5 个顶点, 电势分别为  $v_i, 1 \leq i \leq 5$ ,  $i, j$  两点之间电阻为  $r_{ij} \neq 0$ , 从顶点  $i$  到  $j$  的电流为  $c_{ij}$ , 又记  $c_{ii} = 0$ . 令电势矩阵为  $V = \text{diag}(v_1, \dots, v_5)$ , 电流矩阵为  $C = [c_{ij}]$ , 电导矩阵为  $R = [\frac{1}{r_{ij}}]$ . 求证  $C = VR - RV$ .

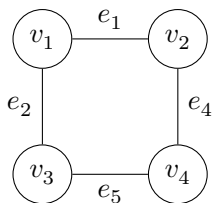


图 1.5.3: 练习 1.5.25

**练习 1.5.25** 图 1.5.3 中的图含有四个顶点  $v_1, v_2, v_3, v_4$ , 顶点之间有边连接. 对称矩阵  $A$  称为该图的邻接矩阵: 如果  $v_i, v_j$  之间有边, 则  $A$  的  $(i, j)$  元是 1; 如果  $v_i, v_j$  之间没有边, 则  $A$  的  $(i, j)$  元是 0.

1. 从  $p_3$  出发, 三次通过连线, 最后回到  $p_3$  的方法有几种?
2. 求矩阵  $A^3$  的  $(3, 3)$  元.
3. 分析  $A^n$  的  $(i, j)$  元的意义.

**练习 1.5.26** 设  $\mathbf{v}^T, \mathbf{w}^T$  是  $k$  维行向量,  $A$  是  $m \times n$  矩阵.

1. 如果矩阵乘积  $\mathbf{v}^T A$  良定义, 需要满足什么条件?
2. 对任意常数  $a, b \in \mathbb{R}$ , 证明  $(a\mathbf{v}^T + b\mathbf{w}^T)A = a\mathbf{v}^T A + b\mathbf{w}^T A$ .
3. 把  $2\mathbf{a}^T + 3\mathbf{b}^T + 4\mathbf{c}^T$  写成一个行向量和矩阵的乘积. 提示: 类比于矩阵乘列向量是矩阵的列的线性组合, 行向量乘矩阵是矩阵的行的线性组合.
4. 求矩阵  $A$ , 使得  $\begin{bmatrix} x_1 & x_2 \end{bmatrix} A \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 2x_1y_1 + 3x_1y_2 + 4x_2y_1 + 5x_2y_2$ .
5. 求矩阵  $A$ , 使得  $\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + 4xy + 5y^2$ . 这样的  $A$  是否唯一?
6. 求 3 阶对称矩阵  $A$ , 使得对任意非零 3 维向量  $\mathbf{v}$ , 都有  $\mathbf{v}^T A \mathbf{v} > 0$ .

**练习 1.5.27** 由标准坐标向量  $\mathbf{e}_i$  定义  $E_{ij} = \mathbf{e}_i \mathbf{e}_j^T$ , 它是  $(i, j)$  元为 1, 其他元素为 0 的矩阵.

1. 证明, 当  $j \neq k$  时,  $E_{ij}E_{kl} = O$ .
2. 对任意矩阵  $A$ , 求向量  $\mathbf{v}$  使得  $A\mathbf{v}$  是  $A$  的第  $i$  列.
3. 对任意矩阵  $A$ , 求向量  $\mathbf{v}$  使得  $\mathbf{v}^T A$  是  $A$  的第  $i$  行.
4. 对任意矩阵  $A$ , 证明,  $\mathbf{e}_i^T A \mathbf{e}_j$  为  $A$  的  $(i, j)$  元.
5. 对  $\mathbf{e}_k \in \mathbb{R}^m$  证明,  $\sum_{k=1}^m \mathbf{e}_k \mathbf{e}_k^T = I_m$ .
6. (阅读) 计算矩阵乘积  $AB$  的  $(i, j)$  元的另一种方法: 设  $A$  有  $m$  列,  $B$  有  $m$  行, 则

$$\mathbf{e}_i^T A B \mathbf{e}_j = \mathbf{e}_i^T A I_m B \mathbf{e}_j = \mathbf{e}_i^T A \left( \sum_{k=1}^m \mathbf{e}_k \mathbf{e}_k^T \right) B \mathbf{e}_j = \sum_{k=1}^m (\mathbf{e}_i^T A \mathbf{e}_k) (\mathbf{e}_k^T B \mathbf{e}_j).$$

**练习 1.5.28** 注意  $m$  维向量是  $m \times 1$  矩阵.

1. 对  $\mathbf{v} \in \mathbb{R}^m, k \in \mathbb{R}$ ,  $k\mathbf{v}, \mathbf{v}k$  是否可以看作矩阵乘法?
2. 对  $\mathbf{v} \in \mathbb{R}^m, \mathbf{w} \in \mathbb{R}^n$ ,  $\mathbf{v}\mathbf{w}^T$  是否良定义? 如果是, 乘积有几行几列?

3. 求  $\begin{bmatrix} 1 \\ 2 \\ \vdots \\ 100 \end{bmatrix} \begin{bmatrix} 10 & \cdots & 1 \end{bmatrix}$  的  $(12, 7)$  元.

4. 求  $\mathbf{v}, \mathbf{w}$ , 使得  $\mathbf{v}\mathbf{w}^T = [(-1)^{i+j}]$ .

5. 求  $\mathbf{v}, \mathbf{w}$ , 使得  $\mathbf{v}\mathbf{w}^T = \begin{bmatrix} i \\ j \end{bmatrix}$ .

6. 令  $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}, B = \begin{bmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_n^T \end{bmatrix}$ . 证明  $AB = \sum_{i=1}^n \mathbf{a}_i \mathbf{b}_i^T$ .

## 1.6 可逆矩阵

### 1.6.1 初等矩阵

回顾 Gauss 消元法中矩阵的初等行变换. 给定矩阵  $A$ , 对  $A$  的第  $i, j$  行做对换变换是互换  $A$  的第  $i, j$  行的位置. 首先考虑  $A$  是列向量的情形, 即  $A = \mathbf{x} \in \mathbb{R}^n$ , 则对换变换定义了如下  $\mathbb{R}^n$  上的变换:

$$\begin{aligned} P_{ij}: \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ \begin{bmatrix} \vdots \\ x_i \\ \vdots \\ x_j \\ \vdots \end{bmatrix} &\mapsto \begin{bmatrix} \vdots \\ x_j \\ \vdots \\ x_i \\ \vdots \end{bmatrix}. \end{aligned}$$

容易验证,  $P_{ij}$  是线性变换, 其表示矩阵为  $P_{ij} = [P_{ij}(\mathbf{e}_1) \cdots P_{ij}(\mathbf{e}_n)]$ .  $P_{ij}$  可由恒同矩阵  $[\mathbf{e}_1 \cdots \mathbf{e}_n]$  的每一列的第  $i, j$  行位置互换得到, 即  $P_{ij}$  可由互换  $I_n$  的第  $i, j$  行位置得到. 根据表示矩阵的定义, 我们有  $P_{ij}(\mathbf{x}) = P_{ij}\mathbf{x}$ , 这意味着列向量的对换行变换可以由左乘矩阵实现. 一般地, 对  $n \times p$  矩阵  $A = [\mathbf{a}_1 \cdots \mathbf{a}_p]$ ,

$$P_{ij}A = P_{ij}[\mathbf{a}_1 \cdots \mathbf{a}_p] = [P_{ij}\mathbf{a}_1 \cdots P_{ij}\mathbf{a}_p],$$

是互换  $A$  的第  $i, j$  行的位置得到的矩阵. 换言之, 任意矩阵的对换行变换也可以由左乘矩阵实现.

容易验证, 以上讨论对其他两种初等行变换也适用. 由此给出如下定义.

**定义 1.6.1 (初等矩阵)** 对恒同矩阵  $I_n$  做一次初等变换, 得到的矩阵统称为**初等矩阵**:

1. 对换行变换: 把  $I_n$  的第  $i, j$  行位置互换, 得到**对换矩阵**

$$P_{ij} = \begin{bmatrix} \ddots & & & & & & \\ & 1 & & & & & \\ & & 0 & & 1 & & \\ & & & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & \\ & 1 & & & & 0 & \\ & & & & & & 1 \\ & & & & & & & \ddots \end{bmatrix} \begin{matrix} \\ \\ i \\ \\ \\ j \\ \\ j \\ \\ \end{matrix};$$

2. 倍乘行变换: 第  $i$  行乘非零常数  $k$ , 得到**倍乘矩阵**

$$E_{ii;k} = \begin{bmatrix} \ddots & & & \\ & 1 & & \\ & & k & \\ & & & 1 \\ & & & & \ddots \end{bmatrix} \begin{matrix} i \\ \\ \\ i \\ \end{matrix};$$

3. 倍加行变换: 把第  $i$  行的  $k$  倍加到第  $j$  行上, 得到**倍加矩阵**

$$E_{ji;k} = \begin{bmatrix} \ddots & & & \\ & 1 & & \\ & & \ddots & \\ & k & & 1 \\ & & & & \ddots \end{bmatrix} \begin{matrix} i \\ \\ j \\ (j > i) \end{matrix}, \quad E_{ji;k} = \begin{bmatrix} \ddots & & & \\ & 1 & & k \\ & & \ddots & \\ & & & 1 \\ & & & & \ddots \end{bmatrix} \begin{matrix} j \\ \\ i \\ (j < i) \end{matrix}.$$

我们还可以考虑初等列变换. 根据前面的分析, 我们容易得到上述初等矩阵左 (右) 乘与初等行 (列) 变换之间的对应关系, 即“左行右列”.

**命题 1.6.2** 1. 若矩阵  $I_m$  经过某一初等行变换得到矩阵  $T$ , 则任意  $m \times n$  矩阵  $A$  经过相同初等行变换得到矩阵  $TA$ .

2. 若矩阵  $I_n$  经过某一初等列变换得到矩阵  $T$ , 则任意  $m \times n$  矩阵  $A$  经过相同初等行变换得到矩阵  $AT$ .

因此, 我们利用 Gauss 消元法或 Gauss-Jordan 消元法求解线性方程组  $A\mathbf{x} = \mathbf{b}$ , 就是找到一系列初等矩阵  $E_1, \dots, E_t$ , 使得矩阵  $E_t \cdots E_1 [A \ \mathbf{b}]$  是阶梯形矩阵或行简化阶梯形矩阵. 而对任意初等矩阵  $E$ , 定理 1.4.5 说明了  $A\mathbf{x} = \mathbf{b}$  与  $E A \mathbf{x} = E \mathbf{b}$  同解.

我们知道, 这些方法能够求解方程组, 根本原因在于初等变换可逆. 那么初等变换诱导的线性变换, 作为线性映射可逆吗? 这又对应了初等矩阵的什么性质?

### 1.6.2 可逆矩阵

我们知道,  $\mathbb{R}^n$  上的任何两个线性变换都可以复合, 得到的还是  $\mathbb{R}^n$  上的线性变换. 其中, 恒同变换  $I$  跟  $\mathbb{R}^n$  上的任何线性变换  $A$  都可交换:  $A \circ I = A = I \circ A$ . 设  $A, B: \mathbb{R}^n \rightarrow \mathbb{R}^n$  是两个线性变换, 如果  $A \circ B = B \circ A = I$ , 则称  $A$  和  $B$  是可逆变换,

称  $B$  是  $A$  的逆变换, 记为  $B = A^{-1}$ ; 类似地,  $A$  是  $B$  的逆变换,  $A = B^{-1}$ . 一个变换可逆当且仅当它既是单射又是满射. 细节参见第 1.1 节.

回顾例 1.5.15,  $R_\theta, H_\theta, P, C_k (k \neq 0), S_k$  都是可逆变换, 其逆变换分别是  $R_{-\theta}, H_\theta, P, C_{k^{-1}}, S_{-k}$ . 而投影变换  $C_0$  不是可逆变换, 因为它既不是单射, 也不是满射.

考虑可逆变换的表示矩阵, 自然地诱导出如下概念.

**定义 1.6.3 (逆矩阵)** 设  $A$  是  $n$  阶方阵, 如果存在  $n$  阶方阵  $B$ , 使得  $AB = BA = I_n$ , 则称  $A$  是**可逆矩阵**或**非奇异矩阵**, 并称  $B$  是  $A$  的**逆 (矩阵)**.

不可逆的矩阵称为**奇异矩阵**.

矩阵的逆如果存在, 则一定唯一: 设  $B, \tilde{B}$  都是  $A$  的逆, 则有  $AB = I_n, \tilde{B}A = I_n$ , 于是  $B = I_n B = \tilde{B}AB = \tilde{B}I_n = \tilde{B}$ . 因此, 我们记  $A$  的 (唯一的) 逆矩阵为  $A^{-1}$ . 另外, 这一点也可以从矩阵和映射的关系得到, 因为映射的逆如果存在则唯一.

根据定义, 恒同矩阵  $I_n$  可逆, 且  $I_n^{-1} = I_n$ . 而零矩阵一定不可逆: 对任意 (可以相乘的)  $A, B, AO = OB = O$ . 例 1.3.9 中的大多数矩阵都可逆:

$$\begin{aligned} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}^{-1} &= \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \\ \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}^{-1} &= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}, \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \\ \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}^{-1} &= \begin{bmatrix} k^{-1} & 0 \\ 0 & 1 \end{bmatrix} \quad (k \neq 0), \\ \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}^{-1} &= \begin{bmatrix} 1 & 0 \\ -k & 1 \end{bmatrix}. \end{aligned}$$

**命题 1.6.4** 初等矩阵可逆, 且  $P_{ij}^{-1} = P_{ij}, E_{ii;k}^{-1} = E_{ii;k^{-1}}, E_{ji;k}^{-1} = E_{ji;-k}$ .

**命题 1.6.5** 对角矩阵  $D$  可逆当且仅当对角元素都不为零, 且  $D$  可逆时有

$$D^{-1} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}^{-1} = \begin{bmatrix} d_1^{-1} & & \\ & \ddots & \\ & & d_n^{-1} \end{bmatrix}.$$

证. 如果对角元素都不为零, 容易验证  $\text{diag}(d_i) \text{diag}(d_i^{-1}) = \text{diag}(d_i^{-1}) \text{diag}(d_i) = I_n$ .

如果对角元素至少有一个为零, 设  $d_i = 0$ , 则  $D$  的第  $i$  行全是零. 任取  $n$  阶方阵  $B$ ,  $DB$  的第  $i$  行也一定全是零, 因此  $DB \neq I_n$ ,  $D$  不可逆.  $\square$

**命题 1.6.6** 设  $A, B$  是可逆矩阵, 则

1.  $A^{-1}$  可逆, 且  $(A^{-1})^{-1} = A$ ;
2.  $A^T$  可逆, 且  $(A^{-1})^T = (A^T)^{-1}$ , 记为  $A^{-T}$ ;
3.  $AB$  可逆, 且  $(AB)^{-1} = B^{-1}A^{-1}$ .

证. 这里只证明第 3 条.

$$\begin{aligned}(B^{-1}A^{-1})(AB) &= B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I, \\(AB)(B^{-1}A^{-1}) &= ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I.\end{aligned}$$

因此  $B^{-1}A^{-1}$  是  $AB$  的逆. □

下面我们就可以初步讨论矩阵可逆的条件了.

**定理 1.6.7** 设  $A$  是  $n$  阶方阵, 以下叙述等价:

1.  $A$  可逆;
2. 任取  $n$  维向量  $\mathbf{b}$ , 方程组  $A\mathbf{x} = \mathbf{b}$  的解唯一, 且  $\mathbf{x} = A^{-1}\mathbf{b}$ ;
3. 齐次方程组  $A\mathbf{x} = \mathbf{0}$  只有零解;
4.  $A$  对应的阶梯形矩阵有  $n$  个主元;
5.  $A$  对应的行简化阶梯形矩阵一定是  $I_n$ ;
6.  $A$  是有限个初等矩阵的乘积.

证. 采用轮转证法.

$1 \Rightarrow 2$ : 考虑由  $A$  确定的线性映射  $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^n, \mathbf{x} \mapsto A\mathbf{x}$ . 方阵  $A$  可逆当且仅当  $\mathbf{A}$  是可逆映射. 根据命题 1.1.2, 这等价于  $\mathbf{A}$  既是单射又是满射. 映射  $\mathbf{A}$  是满射等价于  $A\mathbf{x} = \mathbf{b}$  总有解, 而  $\mathbf{A}$  是单射等价于解唯一.

$2 \Rightarrow 3$ : 显然.

$3 \Rightarrow 4$ : 求解此方程组, 先用初等行变换将  $A$  化为阶梯形, 方程组有唯一解等价于阶梯形的阶梯数与未知数个数  $n$  相等, 故主元数等于阶梯数  $n$ .

$4 \Rightarrow 5$ : 显然.

$5 \Rightarrow 6$ : 由假设, 我们可以用一系列初等行变换把  $A$  化成  $I_n$ , 即存在初等矩阵  $E_1, \dots, E_t$  使得  $E_t \cdots E_1 A = I_n$ . 注意到初等矩阵都可逆, 因此  $A = E_1^{-1} \cdots E_t^{-1}$ .

$6 \Rightarrow 1$ : 初等矩阵都可逆, 而可逆矩阵的乘积也可逆. □

**注 1.6.8** 可以证明,  $n$  阶可逆矩阵可以表示成不多于  $n^2$  个初等矩阵的乘积.

设  $A$  可逆, 定理 1.6.7 的证明提示了如何计算逆矩阵: 求解  $AX = I_n$ , 利用 Gauss-Jordan 消元法, 我们可以把  $A$  化成  $I_n$ , 记所需的初等行变换的乘积为  $E$ , 则有  $EA = I$ , 于是  $X = EAX = E$ . 由此得到计算逆矩阵的一种方法, 把  $A$  和  $I_n$  并列排成一个  $n \times 2n$  矩阵  $[A \ I_n]$ , 通过消元法对其做初等行变换, 同样的初等行变换作用在  $A$  和  $I_n$  上, 当  $A$  化成  $I_n$  时,  $I_n$  就化成了  $A^{-1}$ :

$$[A \ I_n] \xrightarrow{\text{Gauss-Jordan}} [I_n \ A^{-1}]. \quad (1.6.1)$$

同样的方法还可以用来求解矩阵方程  $AX = B$ , 其中  $A$  是  $n$  阶可逆矩阵. 对矩阵  $[A \ B]$  做初等行变换, 当  $A$  化成  $I_n$  时,  $B$  就化成了  $A^{-1}B$ .

**例 1.6.9** 回顾例 1.4.2 中的鸡兔同笼问题, 一个笼子中有鸡和兔共 8 只, 二者的腿共有 22 只, 设鸡有  $x_1$  只, 兔有  $x_2$  只, 有方程组

$$\begin{cases} x_1 + x_2 = 8, \\ 2x_1 + 4x_2 = 22. \end{cases}$$

亦即

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 22 \end{bmatrix}.$$

$x_1, x_2$  的求解可以直接用消元法计算出来. 当然, 根据定理 1.6.7, 我们也可以通过计算系数矩阵的逆来求解.

记系数矩阵为  $A$ , 对  $[A \ I_2]$  消元:

$$\begin{aligned} \left[ \begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array} \right] & \xrightarrow{\text{倍加变换}} \left[ \begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 2 & -2 & 1 \end{array} \right] & \xrightarrow{\text{倍乘变换}} \left[ \begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -1 & \frac{1}{2} \end{array} \right] \\ & \xrightarrow{\text{倍加变换}} \left[ \begin{array}{cc|cc} 1 & 0 & 2 & -\frac{1}{2} \\ 0 & 1 & -1 & \frac{1}{2} \end{array} \right]. \end{aligned}$$

因此  $A$  可逆, 且  $A^{-1} = \begin{bmatrix} 2 & -\frac{1}{2} \\ -1 & \frac{1}{2} \end{bmatrix}$ .

注意  $A$  的列向量组, 第一列表示一只鸡的头腿数目, 第二列表示一只兔的头腿数目. 那么  $A^{-1}$  的列向量组有什么实际含义? 由  $AA^{-1} = I_2$  可得,

$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$



也就是说, 两只鸡和负一只兔子的组合刚好有一个头且没有腿; 而负半只鸡和半只兔子的组合刚好没有头但有一条腿. 而一个有 8 头 22 腿的动物组合, 就应该是

$$8 \begin{bmatrix} 2 \\ -1 \end{bmatrix} + 22 \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 2 & -\frac{1}{2} \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 8 \\ 22 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix},$$

这是因为动物头和腿的数量与动物的数量之间满足线性叠加原理. ⑤

定理 1.6.7 可以用来证明某些具有特殊形式的矩阵的可逆性.

先来看前面提及的三角矩阵, 容易验证如下结论.

**命题 1.6.10** 上三角矩阵可逆当且仅当其对角元素都不为零. 此时, 其逆矩阵也是上三角矩阵, 逆矩阵的对角元素是该矩阵的对应对角元素的倒数.

下三角矩阵也有类似性质.

**定义 1.6.11 (置换矩阵)** 单位矩阵经一系列对换行变换得到的矩阵称为置换矩阵.

简单验证可以得到置换矩阵的一些性质.

**命题 1.6.12** 1. 单位矩阵经一系列对换列变换得到的矩阵也是置换矩阵;

2. 置换矩阵可以通过按某种次序排列单位矩阵的行来得到;

3. 置换矩阵可以通过按某种次序排列单位矩阵的列来得到;

4. 不同的  $n$  阶置换矩阵共有  $n!$  个<sup>7</sup>;

5. 置换矩阵的乘积也是置换矩阵;

6. 置换矩阵的逆是其转置, 也是置换矩阵.

**定义 1.6.13 (对角占优)** 如果矩阵  $A = [a_{ij}]_{n \times n}$  对于  $i = 1, \dots, n$ , 都有  $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ , 则称其为 (行) 对角占优矩阵.

**命题 1.6.14** 对角占优矩阵一定可逆.

证. 根据定理 1.6.7, 只需证明  $Ax = 0$  只有零解. 设  $x$  的分量中, 绝对值最大的元素是  $x_i$ . 考虑方程组的第  $i$  个方程:

$$a_{i1}x_1 + \dots + a_{i,i-1}x_{i-1} + a_{ii}x_i + a_{i,i+1}x_{i+1} + \dots + a_{in}x_n = 0.$$

若  $x_i \neq 0$ , 根据对角占优性质, 有

$$|a_{ii}||x_i| = \left| \sum_{j \neq i} a_{ij}x_j \right| \leq \sum_{j \neq i} |a_{ij}||x_j| \leq \sum_{j \neq i} |a_{ij}||x_i| < |a_{ii}||x_i|.$$

矛盾. □

<sup>7</sup>符号  $n!$  表示  $n$  的阶乘, 即  $n! = 1 \cdot 2 \cdot \dots \cdot n$ .

**例 1.6.15** 给定铁丝网如图 1.6.1，铁丝粗细均匀，网格等距，格点温度如图. 假设格点的温度都已经稳定，亦即每个格点的温度等于与其相连的格点的温度的平均值，那么各格点上的温度是多少？

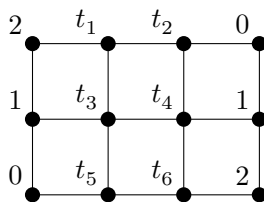


图 1.6.1: 热传导

可以列出方程组：

$$\begin{cases} 3t_1 = t_2 + t_3 + 2, \\ 3t_2 = t_1 + t_4 + 0, \\ 4t_3 = t_1 + t_4 + t_5 + 1, \\ 4t_4 = t_2 + t_3 + t_6 + 1, \\ 3t_5 = t_3 + t_6 + 0, \\ 3t_6 = t_4 + t_5 + 2. \end{cases} \quad (1.6.2)$$

亦即

$$\begin{bmatrix} 3 & -1 & -1 & 0 & 0 & 0 \\ -1 & 3 & 0 & -1 & 0 & 0 \\ -1 & 0 & 4 & -1 & -1 & 0 \\ 0 & -1 & -1 & 4 & -1 & 0 \\ 0 & 0 & -1 & 0 & 3 & -1 \\ 0 & 0 & 0 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}.$$

不难看出，系数矩阵是对角占优矩阵. 根据命题 1.6.14，该矩阵可逆，因此方程组有唯一解. ☺

最后，利用矩阵乘法和矩阵的逆，我们可以定义矩阵的幂.

**定义 1.6.16 (矩阵的幂)** 设  $A$  是  $n$  阶方阵，我们定义其正整数幂  $A^k = \underbrace{AA \cdots A}_{k \text{ 个}}$ ，其中  $k$  是正整数.

如果  $A$  是可逆矩阵，我们定义其负整数幂  $A^{-k} = (A^{-1})^k$  以及零次幂  $A^0 = I_n$ .

注意， $A^{-k} = (A^k)^{-1}$ .

还可以定义矩阵的多项式：给定  $n$  阶方阵  $A$  和多项式  $f(x) = a_0 + a_1x + \cdots + a_px^p$ ，则  $f(A) = a_0I_n + a_1A + \cdots + a_pA^p$ .

### 1.6.3 等价关系

我们已经知道,一系列初等行变换可以把矩阵化成行简化阶梯形矩阵.为描述矩阵在初等行变换前后的关系,我们给出如下定义.

**定义 1.6.17 (左相抵)** 如果矩阵  $A$  可以经过一系列初等行变换化成矩阵  $B$ ,则称  $A$  和  $B$  左相抵.

**命题 1.6.18** 给定两个  $m \times n$  矩阵  $A, B$ .那么二者左相抵,当且仅当存在  $m$  阶可逆矩阵  $P$ ,使得  $PA = B$ .

证.  $A$  与  $B$  左相抵  $\Leftrightarrow A$  经一系列初等行变换得到  $B \Leftrightarrow$  存在  $m$  阶初等矩阵  $P_1, \dots, P_s$ , 使得  $P_s \cdots P_1 A = B \xrightarrow{\text{定理 1.6.7}} \text{存在 } m \text{ 阶可逆矩阵 } P, \text{ 使得 } PA = B. \quad \square$

可以看到在所有和矩阵  $A$  左相抵的矩阵中,形式最简单的应该就是其行简化阶梯形,因此我们也可以称此矩阵为  $A$  的**左相抵标准形**.

容易看到如果两个矩阵左相抵,那么以二者为系数矩阵的齐次线性方程组,或者以二者为增广矩阵的线性方程组,其解集相同.

不难看出,左相抵关系满足如下三条基本性质:

1. 反身性: 每个矩阵和自身左相抵;
2. 对称性: 如果  $A$  和  $B$  左相抵,那么  $B$  和  $A$  左相抵;
3. 传递性: 如果  $A$  和  $B$  左相抵,  $B$  和  $C$  左相抵,那么  $A$  和  $C$  左相抵.

事实上,很多数学对象之间都存在类似的关系.为此,我们抽象出一系列概念.

**定义 1.6.19 (等价关系)** 如果非空集合  $S$  的元素之间定义了一种二元关系“ $\sim$ ”,满足:

1. 反身性: 对任意  $a \in S$ ,  $a \sim a$ ;
2. 对称性: 如果  $a \sim b$ , 那么  $b \sim a$ ;
3. 传递性: 如果  $a \sim b, b \sim c$ , 那么  $a \sim c$ ,

则称此关系为  $S$  上的一个**等价关系**.

$S$  中所有与其中某一元素  $a$  等价的元素组成的集合称为  $a$  所在的**等价类**.由元素  $a$  变成与其等价的元素  $b$  的过程称为对应于这一等价关系的**等价变换**.同一等价类中的元素共享的某个属性称为这一等价关系中(或这一等价变换下)的**不变量**.同一等价类中的形式最简单、性质最好的元素往往称为这一等价关系中(或这一等价变换下)的**标准形**.

显然,  $S$  等于所有等价类的无交并. 等价类的数目可以有限, 也可以无限.

**例 1.6.20** 平面上所有三角形组成的集合上, 相似是一个等价关系, 三个角的角度是等价关系中的不变量.

平面上所有三角形组成的集合上, 全等也是一个等价关系, 三个角的角度和三条边的长度是等价关系中的不变量.

所有  $m \times n$  矩阵组成的集合上, 左相抵也是一个等价关系, 等价变换是初等行变换, 作为系数矩阵得到的齐次线性方程组的解集是等价关系中的不变量, 作为增广矩阵得到的线性方程组的解集也是等价关系中的不变量, 行简化阶梯形是等价关系中的标准形.

所有关于未知数  $x$  的一元方程组成的集合上, 同解也是一个等价关系, 等价变换是同解变换<sup>8</sup>, 解集是等价关系中的不变量, 而解出的  $x = x_1, x_2, \dots$  是等价关系中的标准形. ☺

为了研究某个集合的某个问题, 在此集合上引入等价关系, 然后找到其标准形, 往往能够使问题的研究得到简化: 这是数学中的一个重要方法. 这一方法将在本书中反复体现. 我们将为矩阵定义更多的变换, 并通过找到其标准形来研究矩阵的性质.

## 习题

**练习 1.6.1** 计算下列矩阵乘法.

$$1. \begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} & 1 \\ & & 1 \\ 1 & & \end{bmatrix}.$$

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

$$3. \begin{bmatrix} & 1 & & \\ & & 1 & \\ & & & 1 \\ 1 & & & \end{bmatrix}^k, \quad k \text{ 是正整数}.$$

$$4. \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

$$5. \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ -1 & & 1 & & \\ -1 & & & 1 & \\ -1 & & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 2 & 1 & & \\ 1 & 3 & 3 & 1 & \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}.$$

**练习 1.6.2** 求下列矩阵的逆矩阵.

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

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

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

<sup>8</sup>大家中学都学过, 只有保持解不变的变换才能用来解方程或方程组, 这种变换称为同解变换.

$$4. \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}.$$

练习 1.6.3 设  $A = \begin{bmatrix} 1 & 2 & -1 \\ 3 & a & -2 \\ 5 & -2 & 1 \end{bmatrix}$  不可逆, 求  $a$ .

练习 1.6.4 设  $A = \begin{bmatrix} 1 & & \\ 2 & 3 & \\ 4 & 5 & 6 \end{bmatrix}$ , 解方程  $\boldsymbol{x} = \boldsymbol{e}_i, i = 1, 2, 3$ , 并求  $A^{-1}$ .

练习 1.6.5 对  $[A \quad I_n]$  做初等行变换, 求下列矩阵的逆矩阵.

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

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

$$3. \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ 1 & & 0 & \\ & & 1 & 0 \end{bmatrix}.$$

$$4. \begin{bmatrix} & & & 1 \\ & & 1 & \\ & \ddots & & \\ 1 & & & \end{bmatrix}.$$

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

$$6. \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}.$$

$$7. \begin{bmatrix} 1 & -a & & \\ & 1 & -b & \\ & & 1 & -c \\ & & & 1 \end{bmatrix}.$$

$$8. \begin{bmatrix} 1 & a & & \\ & 1 & \ddots & \\ & & \ddots & a \\ & & & 1 \end{bmatrix}_{n \times n}.$$

练习 1.6.6 说明  $\begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{bmatrix}$  可逆, 并计算其逆.

练习 1.6.7 求  $A = \begin{bmatrix} & & a_1 & \\ & & & \ddots \\ & & & & a_{n-1} \\ a_n & & & & \end{bmatrix}$  的逆矩阵, 其中  $a_i \neq 0$ .

练习 1.6.8 求下列矩阵方程的解:  $\begin{bmatrix} 2 & 2 & 3 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix} X = \begin{bmatrix} 4 & -1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}.$

**练习 1.6.9** 证明二阶方阵  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$  可逆当且仅当  $d = a_{11}a_{22} - a_{12}a_{21} \neq 0$ . 此时,  $A^{-1} = \frac{1}{d} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$ .

**练习 1.6.10** 证明, 有一列元素 (或一行元素) 全为零的方阵不可逆.

**练习 1.6.11** 证明, 对角元素全非零的上三角矩阵  $U$  可逆, 其逆矩阵  $U^{-1}$  也是上三角矩阵, 且  $U^{-1}$  的对角元素是  $U$  的对角元素的倒数.

**练习 1.6.12** 是否只有方阵才有可能可逆? 设  $A$  是  $m \times n$  矩阵.

1. 设  $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ , 能否找到  $B, C$  使得  $AB = I_2, CA = I_3$ ?
2. 设  $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$ , 能否找到  $B$  使得  $AB = I_2, CA = I_3$ ?
3. 设  $CA = I_n$ , 证明  $A\mathbf{x} = \mathbf{0}$  只有零解. 此时  $m, n$  之间有何种关系?
4. 设  $AB = I_m$ , 证明  $A^T\mathbf{x} = \mathbf{0}$  只有零解. 此时  $m, n$  之间有何种关系?
5. 设  $AB = I_m, CA = I_n$ , 证明  $m = n$  且  $B = C$ ; 由此可知, 可逆矩阵一定是方阵.
6. 如果  $m \neq n$ , 那么  $\mathbb{R}^m, \mathbb{R}^n$  之间是否存在线性双射?

**练习 1.6.13** 如果  $n$  阶方阵  $A, B$  满足  $AB = I_n$ , 判断  $A, B$  是否可逆.

**练习 1.6.14** 设

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

1. 求上述矩阵对应的初等行、列变换.
2. 从行变换的角度看, 是否一定有  $AB = BA$ ? 如果否,  $AB - BA$  从行变换的角度意味着什么?
3. 从行变换的角度看, 是否一定有  $AC = CA$ ? 如果否,  $AC - CA$  从行变换的角度意味着什么?
4. 从行变换的角度看, 是否一定有  $BC = CB$ ? 如果否,  $BC - CB$  从行变换的角度意味着什么?
5. 从行变换的角度看,  $D$  是否和  $A, B, C$  可交换? 计算  $DAD^{-1}, DBD^{-1}, DCD^{-1}$ .
6. 从行变换的角度看,  $P$  是否和  $A, B, C$  可交换? 计算  $PAP^{-1}, PBP^{-1}, PCP^{-1}$ .
7. 对三阶方阵  $X$ ,  $(AX)B$  和  $A(XB)$  对  $X$  分别做了何种行、列变换?
8. 对任意矩阵  $X$ , 先做初等行变换, 再做初等列变换, 其结果是否等于先做该初等列变换, 再做该初等行变换? 这对应着矩阵乘法的什么性质?

**练习 1.6.15** 设矩阵  $A$  和  $B$  左相抵. 求证:

1. 如果  $A$  的第一列全是零, 则  $B$  的第一列全是零.

2. 如果  $A$  的所有列都相同, 则  $B$  的所有列都相同.
3. 如果  $A$  的第一列是第二列与第三列的和, 则  $B$  的第一列也是第二列与第三列的和.
4. 如果  $A$  的第一列和第二列不成比例, 则  $B$  的第一列和第二列也不成比例.

**练习 1.6.16** 设有  $M_1, M_2, M_3$  三个城市, 城市之间有人口迁移. 定义矩阵  $A$ , 其  $(i, j)$  元  $a_{ij}$  为人在一年中从  $A_j$  迁移到  $A_i$  的概率. 注意,  $A$  的每个元素都在  $0, 1$  之间, 且  $A$  的每一列的元素之和都是 1.

1. 设今年三个城市的人口分别是  $x_1, x_2, x_3$ , 证明明年它们的预期人口分别是  $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$  的三个分量.
2. 人口迁移满足什么条件时,  $A$  为列对角占优矩阵 (行对角占优矩阵的转置)? 人口迁移满足什么条件时,  $A$  为行对角占优矩阵? 考虑现实生活中的情形, 讨论这些假设是否合理.
3. 如果把矩阵对角占优定义中的大于号换成大于等于号, 则称该矩阵为弱对角占优矩阵. 证明  $A = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$  弱对角占优, 且可逆, 并求其逆矩阵.
4. 对任意  $n$  阶方阵, 设它有  $n-1$  行都是对角占优, 仅有 1 行弱对角占优, 该矩阵可逆吗?

**练习 1.6.17** 设  $A$  是  $n$  阶方阵.

1. 对任意两个多项式  $p(x), q(x)$ , 是否一定有  $p(A)q(A) = q(A)p(A)$ ?
2. 证明, 所有形如  $\begin{bmatrix} a & b & c \\ & a & b \\ & & a \end{bmatrix}$  的矩阵全都彼此交换.
3. 设  $f(x) = p(x) + q(x), g(x) = p(x)q(x)$ , 证明,  $f(A) = p(A) + q(A), g(A) = p(A)q(A)$ .
4. 求  $A$  使得  $A + I_n, A - I_n$  均不为零, 但是  $A^2 - I_n = O$ .
5. 设  $A^2 - I_n = O$ , 证明, 任何  $n$  维向量  $\mathbf{v}$  都存在分解式  $\mathbf{v} = \mathbf{x} + \mathbf{y}$ , 满足  $(A - I_n)\mathbf{x} = (A + I_n)\mathbf{y} = \mathbf{0}$ .
6. 设  $A^3 = O$ , 证明  $A + I_n$  与  $A - I_n$  都可逆, 并求  $p(x), q(x)$  使得  $p(A), q(A)$  分别为其逆.

**练习 1.6.18** 证明, 如果  $n$  阶方阵  $A$  满足  $A^2 = A$ , 则  $I_n - 2A$  可逆.

**练习 1.6.19** 如果一个  $n$  阶方阵从  $(1, n)$  元到  $(n, 1)$  元的对角线下的所有元素均为零, 则称为西北矩阵. 类似地, 可以定义东南矩阵. 如果  $B$  是西北矩阵, 那么  $B^T, B^2, B^{-1}$  是什么矩阵? 西北矩阵和东南矩阵的乘积是什么矩阵?

**练习 1.6.20** 1. 求可逆矩阵  $A, B$ , 使得  $A + B$  不可逆.

2. 求不可逆矩阵  $A, B$ , 使得  $A + B$  可逆.

3. 求 3 阶不可逆矩阵  $A$ , 使得对任意  $k > 0, A + kI_3$  都对角占优.

**练习 1.6.21** 求所有三阶矩阵  $A$ , 满足  $A^2 = I_3$ , 且  $A$  的每个元素只能是 0 或 1.

**练习 1.6.22** 设  $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$  是对称矩阵, 通过“对称化简”求其逆矩阵:

1. 将  $A$  的第一行的二倍从第二行中减去, 第一行的三倍从第三行中减去. 这对应哪个初等矩阵  $E_1$ ?  $E_1^T$  对应的列变换是什么? 计算  $E_1 A$  和  $A_1 = E_1 A E_1^T$ .
2. 将  $A_1$  的第二行与第三行调换. 这对应哪个初等矩阵  $E_2$ ?  $E_2^T$  对应的列变换是什么? 计算  $E_2 A_1$  和  $A_2 = E_2 A_1 E_2^T$ .
3. 将  $A_2$  的第二行的  $\frac{1}{3}$  倍从第三行中减去. 这对应哪个初等矩阵  $E_3$ ?  $E_3^T$  对应的列变换是什么? 计算  $E_3 A_2$  和  $A_3 = E_3 A_2 E_3^T$ .
4. 综上,  $A_3 = E_3 E_2 E_1 A E_1^T E_2^T E_3^T$ , 由此求  $A$  的逆.

**练习 1.6.23** 求证: 对称矩阵的逆矩阵也是对称矩阵; 反对称矩阵的逆矩阵也是反对称矩阵.

**练习 1.6.24** 给定  $n$  阶实反对称矩阵  $A$ , 求证:  $I_n - A$  可逆.

## 1.7 分块矩阵

在矩阵乘法中, 我们经常把矩阵写成列向量组或者行向量组并排的形式. 例如, 设

$A = \begin{bmatrix} \tilde{a}_1^T \\ \vdots \\ \tilde{a}_l^T \end{bmatrix}$ ,  $B = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$ , 其中  $\tilde{a}_i^T$  是  $m$  维行向量,  $\mathbf{b}_1$  是  $m$  维列向量, 则

$$AB = \begin{bmatrix} \tilde{a}_1^T \\ \vdots \\ \tilde{a}_l^T \end{bmatrix} [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n] = \begin{bmatrix} \tilde{a}_1^T \mathbf{b}_1 & \cdots & \tilde{a}_1^T \mathbf{b}_n \\ \vdots & & \vdots \\ \tilde{a}_l^T \mathbf{b}_1 & \cdots & \tilde{a}_l^T \mathbf{b}_n \end{bmatrix}.$$

另一方面, 我们有

$$BA = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n] \begin{bmatrix} \tilde{a}_1^T \\ \vdots \\ \tilde{a}_l^T \end{bmatrix} = \mathbf{b}_1 \tilde{a}_1^T + \cdots + \mathbf{b}_n \tilde{a}_l^T.$$

可以看到, 这和把向量看成元素计算矩阵乘法的规则类似, 最重要的区别是矩阵乘法不能颠倒顺序, 而数的乘法满足交换律, 因此与顺序无关.

事实上, 我们可以把矩阵分成任意大小的块, 如

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rs} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ B_{21} & B_{22} & \cdots & B_{2t} \\ \vdots & \vdots & & \vdots \\ B_{s1} & B_{s2} & \cdots & B_{st} \end{bmatrix},$$



其中,  $A_{ij}$  为  $l_i \times m_j$  矩阵,  $B_{jk}$  为  $m_j \times n_k$  矩阵, 则分块矩阵的乘法和转置分别为:

$$AB = \begin{bmatrix} \sum_{p=1}^s A_{1p}B_{p1} & \sum_{p=1}^s A_{1p}B_{p2} & \cdots & \sum_{p=1}^s A_{1p}B_{pt} \\ \sum_{p=1}^s A_{2p}B_{p1} & \sum_{p=1}^s A_{2p}B_{p2} & \cdots & \sum_{p=1}^s A_{2p}B_{pt} \\ \vdots & \vdots & & \vdots \\ \sum_{p=1}^s A_{rp}B_{p1} & \sum_{p=1}^s A_{rp}B_{p2} & \cdots & \sum_{p=1}^s A_{rp}B_{pt} \end{bmatrix}, \quad A^T = \begin{bmatrix} A_{11}^T & A_{21}^T & \cdots & A_{r1}^T \\ A_{12}^T & A_{22}^T & \cdots & A_{r2}^T \\ \vdots & \vdots & & \vdots \\ A_{1s}^T & A_{2s}^T & \cdots & A_{rs}^T \end{bmatrix}.$$

运算法则和针对元素的法则完全一样, 稍加计算即可证明其正确性, 这里要注意的是:

1. 分块矩阵的运算并不是新的运算, 而仅仅是矩阵运算的另一种书写形式;
2. 由于矩阵乘法没有交换律, 分块矩阵的乘法运算中的每一个矩阵乘法, 一定要左边矩阵的块去左乘右边矩阵的块;
3. 由于矩阵的转置和本身一般不相等, 所以分块矩阵的转置中的块也要转置.

分块矩阵的线性运算是显然的.

**例 1.7.1** 具有如下形式的矩阵

$$A = \begin{bmatrix} A_1 & O & \cdots & O \\ O & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & A_s \end{bmatrix} = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{bmatrix} \quad (1.7.1)$$

称为**分块对角矩阵**, 其中对角线上是可能非零的矩阵  $A_1, \dots, A_s$ , 称为**对角块**. 考虑线性

映射  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , 把向量  $x \in \mathbb{R}^n$  也作对应的分块  $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_s \end{bmatrix}$ , 得到

$$A(x) = Ax = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_s \end{bmatrix} = \begin{bmatrix} A_1 x_1 \\ A_2 x_2 \\ \vdots \\ A_s x_s \end{bmatrix}$$

可以看到, 在分块对角矩阵对应的线性映射之下,  $x$  的各个分块  $x_1, \dots, x_s$  互相独立. 由此得到的线性系统  $A$  等价于  $s$  个独立的子线性系统  $A_1, \dots, A_s$ .

当 (1.7.1) 中的  $A_1, \dots, A_s$  都是可逆矩阵时, 分块对角矩阵  $A$  可逆, 其逆矩阵是

$$A^{-1} = \begin{bmatrix} A_1^{-1} & & \\ & A_2^{-1} & \\ & & \ddots \\ & & & A_s^{-1} \end{bmatrix}.$$

☺

由于分成多个块的情形可以通过反复进行  $2 \times 2$  的分块得到, 下面主要讨论  $2 \times 2$  分块矩阵. 设  $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ , 其中  $A_{ij}$  是  $m_i \times n_j$  矩阵, 则  $\mathbf{A}: \mathbb{R}^{n_1+n_2} \rightarrow \mathbb{R}^{m_1+m_2}$  为:

$$\mathbf{A}(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11}\mathbf{x}_1 + A_{12}\mathbf{x}_2 \\ A_{21}\mathbf{x}_1 + A_{22}\mathbf{x}_2 \end{bmatrix}.$$

如果记  $\mathbf{y} = \mathbf{A}(\mathbf{x})$  并分块, 那么

$$\mathbf{y}_1 = A_{11}\mathbf{x}_1 + A_{12}\mathbf{x}_2,$$

$$\mathbf{y}_2 = A_{21}\mathbf{x}_1 + A_{22}\mathbf{x}_2.$$

可以看到,  $A_{ij}$  表示了  $\mathbf{x}_j$  和  $\mathbf{y}_i$  的关系. 由此可知,  $\mathbf{A}$  等价于四个线性映射  $\mathbf{A}_{ij}: \mathbb{R}^{n_j} \rightarrow \mathbb{R}^{m_i}$ , 可以表示为

$$\begin{array}{ccc} \mathbb{R}^{n_1} & \xrightarrow{\quad} & \mathbb{R}^{m_1} \\ & \searrow & \nearrow \\ & \mathbb{R}^{n_2} & \xrightarrow{\quad} \mathbb{R}^{m_2} \end{array}.$$

特别地, 当  $A_{21} = O$  时,  $\mathbf{A}(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix} \begin{bmatrix} * \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} * \\ A_{22}\mathbf{x}_2 \end{bmatrix}$ , 这里  $*$  表示我们不关心其具体值.

**例 1.7.2** 对例 1.2.8 中的线性映射, 我们考虑其表示矩阵的分块.

1. 结账是一个从  $\mathbb{R}^3$  到  $\mathbb{R}$  的线性映射, 其中定义域  $\mathbb{R}^3$  的标准坐标向量分别是一千克苹果、一千克香蕉和一千克樱桃, 它们对应的单价分别是 12.98、9.98 和 129.98 元. 苹果和香蕉组成廉价组  $\mathbf{x}_1$ , 樱桃组成高价组  $\mathbf{x}_2$ . 那么廉价组和高价组的结账的表示矩阵分别是

$$A_1 = [\text{结账(廉价组)}] = \begin{bmatrix} 12.98 & 9.98 \end{bmatrix},$$

$$A_2 = [\text{结账(高价组)}] = \begin{bmatrix} 129.98 \end{bmatrix}.$$

结账的总表示矩阵是

$$A = \begin{bmatrix} \text{结账(廉价组)} & \text{结账(高价组)} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \end{bmatrix} = \begin{bmatrix} 12.98 & 9.98 & 129.98 \end{bmatrix}.$$

2. 营养是一个从  $\mathbb{R}^3$  到  $\mathbb{R}^2$  的线性映射. 纤维素组成补充组  $\mathbf{y}_1$ , 糖组成供能组  $\mathbf{y}_2$ . 那么廉价组和高价组的补充组营养与供能组营养的表示矩阵分别是

$$A_{11} = \begin{bmatrix} \text{补充(廉价组)} \end{bmatrix} = \begin{bmatrix} 12 & 12 \end{bmatrix},$$

$$A_{21} = \begin{bmatrix} \text{供能(廉价组)} \end{bmatrix} = \begin{bmatrix} 135 & 208 \end{bmatrix},$$

$$A_{12} = \begin{bmatrix} \text{补充(高价组)} \end{bmatrix} = \begin{bmatrix} 3 \end{bmatrix},$$

$$A_{22} = \begin{bmatrix} \text{供能(高价组)} \end{bmatrix} = \begin{bmatrix} 99 \end{bmatrix}.$$

营养的总表示矩阵是

$$A = \begin{bmatrix} \text{补充(廉价组)} & \text{补充(高价组)} \\ \text{供能(廉价组)} & \text{供能(高价组)} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} 12 & 12 & 3 \\ 135 & 208 & 99 \end{bmatrix}.$$

☺

与定义 1.6.1 中的初等矩阵类似, 我们可以定义分块初等矩阵. 为方便起见, 仍仅考虑  $2 \times 2$  分块矩阵:

1. 分块对换矩阵  $\begin{bmatrix} & I_{n_1} \\ I_{n_2} & \end{bmatrix}$ :

$$\begin{bmatrix} & I_{n_1} \\ I_{n_2} & \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} A_2 \\ A_1 \end{bmatrix}, \quad \begin{bmatrix} B_1 & B_2 \end{bmatrix} \begin{bmatrix} & I_{n_1} \\ I_{n_2} & \end{bmatrix} = \begin{bmatrix} B_2 & B_1 \end{bmatrix};$$

2. 分块倍乘矩阵  $\begin{bmatrix} I_{n_1} & \\ & P \end{bmatrix}$ , 其中  $P$  是可逆矩阵:

$$\begin{bmatrix} I_{n_1} & \\ & P \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ PA_2 \end{bmatrix}, \quad \begin{bmatrix} B_1 & B_2 \end{bmatrix} \begin{bmatrix} I_{n_1} & \\ & P \end{bmatrix} = \begin{bmatrix} B_1 & B_2P \end{bmatrix};$$

3. 分块倍加矩阵  $\begin{bmatrix} I_{n_1} & M \\ & I_{n_2} \end{bmatrix}$ , 其中  $M$  是  $n_1 \times n_2$  矩阵:

$$\begin{bmatrix} I_{n_1} & M \\ & I_{n_2} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} A_1 + MA_2 \\ A_2 \end{bmatrix}, \quad \begin{bmatrix} B_1 & B_2 \end{bmatrix} \begin{bmatrix} I_{n_1} & M \\ & I_{n_2} \end{bmatrix} = \begin{bmatrix} B_1 & B_1M + B_2 \end{bmatrix}.$$

容易验证, 三类分块初等矩阵都可逆, 它们的逆矩阵分别为:

$$\begin{bmatrix} & I_{n_2} \\ I_{n_1} & \end{bmatrix}, \quad \begin{bmatrix} I_{n_1} & \\ & P^{-1} \end{bmatrix}, \quad \begin{bmatrix} I_{n_1} & -M \\ & I_{n_2} \end{bmatrix}.$$

现在我们来考虑对分块矩阵  $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$  做 Gauss 消元法, 只是现在数变成了矩阵. 如我们前面所说, 分块矩阵的运算法则和数的基本相同, 但是尤其要注意乘法的运算次序. 设  $A_{11}$  可逆. 直接计算, 得

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}. \quad (1.7.2)$$

类似于命题 1.6.10, 我们知道矩阵  $A_{22} - A_{21}A_{11}^{-1}A_{12}$  可逆, 等价于等式右边矩阵可逆, 又等价于矩阵  $A$  可逆.

**命题 1.7.3** 若矩阵  $A_{11}$  和  $A_{22} - A_{21}A_{11}^{-1}A_{12}$  可逆, 则矩阵  $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$  可逆, 且

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}. \quad (1.7.3)$$

证. 再对 (1.7.2) 右边消去右上角矩阵, 即

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}. \quad (1.7.4)$$

综合 (1.7.2) 和 (1.7.4), 我们有

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \begin{bmatrix} I & A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix}. \quad (1.7.5)$$

两边取逆, 即得结论.  $\square$

矩阵  $A_{22} - A_{21}A_{11}^{-1}A_{12}$  称为  $A_{11}$  关于  $A$  的 **Schur 补**.

类似地, 我们有如下推论.

**推论 1.7.4** 若矩阵  $A_{22}$  和  $A_{11} - A_{12}A_{22}^{-1}A_{21}$  可逆, 则矩阵  $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$  可逆, 且

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -A_{11}^{-1}A_{12} & I \end{bmatrix} \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -A_{21}A_{22}^{-1} \\ 0 & I \end{bmatrix}. \quad (1.7.6)$$

证. 利用命题 1.7.3 和分块对换矩阵. □

矩阵  $A_{11} - A_{12}A_{22}^{-1}A_{21}$  称为  $A_{22}$  关于  $A$  的 Schur 补.

比较 (1.7.3) 和 (1.7.6) 的左上角块, 我们可以得到

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1},$$

这给出了一定条件下一个可逆矩阵加上某个矩阵后的逆的公式. 我们下面给出一个常用的简化版本.

**命题 1.7.5 (Sherman-Morrison 公式)** 设  $A$  是  $n$  阶方阵, 而  $u, v$  是  $n$  维向量. 若  $A$  可逆, 则  $A + uv^T$  可逆, 当且仅当  $1 + v^T A^{-1}u \neq 0$ , 且

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u}.$$

特别地,  $I + uv^T$  可逆, 当且仅当  $1 + v^T u \neq 0$ , 且

$$(I + uv^T)^{-1} = I - \frac{uv^T}{1 + v^T u}.$$

事实上, 命题 1.7.5 中的一般情形, 也可以由特殊情形来证明, 留给读者思考.

下面再看一个稍微复杂的弹簧振子, 来体会矩阵分块和 Sherman-Morrison 公式的用处.

**例 1.7.6 (弹簧振子)** 假设四根弹簧和三个振子如图 1.7.1 交替依次挂在天花板和地板之间, 弹簧的原始长度为  $l_1, l_2, l_3, l_4$ , 劲度系数分别为  $k_1, k_2, k_3, k_4$ , 振子的质量分别为  $m_1, m_2, m_3$ , 天花板和地板的间距是  $d$ , 而  $d > l_1 + l_2 + l_3 + l_4$ . 系统处于静止状态, 问三个振子距离天花板有多远?

和例 1.5.10 相比, 这一问题如果不列方程就很难直接得到解了. 我们仍然假设三个振子到天花板的距离分别是  $l_1 + x_1, l_1 + l_2 + x_2, l_1 + l_2 + l_3 + x_3$ , 欲求  $x_1, x_2, x_3$ . 记

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ . 令  $\delta = d - l_1 - l_2 - l_3 - l_4$ , 那么四根弹簧的拉伸长度是

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \\ \delta - x_3 \end{bmatrix} = \delta \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} =: \delta \mathbf{e}_4 + D_{43} \mathbf{x}.$$

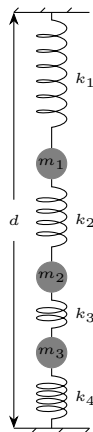


图 1.7.1: 弹簧振子

四根弹簧提供的拉力就是  $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = K_4 \mathbf{u}$ , 其中  $K_4 = \begin{bmatrix} k_1 & & & \\ & k_2 & & \\ & & k_3 & \\ & & & k_4 \end{bmatrix}$ . 振子所受

拉力为  $\mathbf{f} = \begin{bmatrix} y_1 - y_2 \\ y_2 - y_3 \\ y_3 - y_4 \end{bmatrix} = \begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = D_{43}^T \mathbf{y}$ . 因此,

$$\mathbf{f} = D_{43}^T \mathbf{y} = D_{43}^T K_4 \mathbf{u} = D_{43}^T K_4 (D_{43} \mathbf{x} + \delta \mathbf{e}_4) = D_{43}^T K_4 D_{43} \mathbf{x} + \delta D_{43}^T K_4 \mathbf{e}_4.$$

注意, 和例 1.5.10 不同, 因为存在常数项  $\delta D_{43}^T K_4 \mathbf{e}_4$ , 所以从  $\mathbf{x}$  到  $\mathbf{f}$  不构成线性映射.

记振子的质量向量为  $\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}$ , 则振子所受重力为  $\mathbf{g} = g\mathbf{m} = \begin{bmatrix} m_1 g \\ m_2 g \\ m_3 g \end{bmatrix}$ . 利用力

的平衡, 我们有  $\mathbf{g} = \mathbf{f}$ . 因此, 我们得到方程  $(D_{43}^T K_4 D_{43}) \mathbf{x} = g\mathbf{m} - \delta D_{43}^T K_4 \mathbf{e}_4$ , 其中

$$\begin{aligned} D_{43}^T K_4 D_{43} &= \begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & 1 & -1 \end{bmatrix} \begin{bmatrix} k_1 & & & \\ & k_2 & & \\ & & k_3 & \\ & & & k_4 \end{bmatrix} \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} k_1 + k_2 & -k_2 & & \\ -k_2 & k_2 + k_3 & -k_3 & \\ & -k_3 & k_3 + k_4 & \end{bmatrix}. \end{aligned}$$

可以看到, 这个解法可以很容易推广到任意多弹簧和振子上.

另一方面, 由于例 1.5.10 中, 我们已经考虑过三根弹簧和三个振子组成的系统, 因此我们也可以把上面三个弹簧和三个振子作为一个整体来考虑. 我们记这个子系统的弹簧拉伸长度、拉力、振子拉力为  $\hat{u}, \hat{y}, \hat{f}$ . 那么四根弹簧的拉伸长度就是

$$\mathbf{u} = \begin{bmatrix} \hat{u} \\ \delta - x_3 \end{bmatrix} = \begin{bmatrix} D_{33}\mathbf{x} \\ \delta - \mathbf{e}_3^T \mathbf{x} \end{bmatrix} = \delta \mathbf{e}_4 + \begin{bmatrix} D_{33} \\ -\mathbf{e}_3^T \end{bmatrix} \mathbf{x}.$$

四根弹簧提供的拉力就是

$$\mathbf{y} = \begin{bmatrix} \hat{y} \\ k_4 u_4 \end{bmatrix} = \begin{bmatrix} K_3 \hat{u} \\ k_4 u_4 \end{bmatrix} = \begin{bmatrix} K_3 & \\ & k_4 \end{bmatrix} \begin{bmatrix} \hat{u} \\ u_4 \end{bmatrix} = \begin{bmatrix} K_3 & \\ & k_4 \end{bmatrix} \mathbf{u}.$$

振子所受拉力为

$$\mathbf{f} = \hat{\mathbf{f}} - \begin{bmatrix} 0 \\ 0 \\ y_4 \end{bmatrix} = D_{33}^T \hat{\mathbf{y}} - \mathbf{e}_3 y_4 = \begin{bmatrix} D_{33}^T & -\mathbf{e}_3 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{y}} \\ y_4 \end{bmatrix} = \begin{bmatrix} D_{33}^T & -\mathbf{e}_3 \end{bmatrix} \mathbf{y}.$$

因此,

$$\mathbf{f} = \begin{bmatrix} D_{33}^T & -\mathbf{e}_3 \end{bmatrix} \begin{bmatrix} K_3 & \\ & k_4 \end{bmatrix} \begin{bmatrix} D_{33} \\ -\mathbf{e}_3^T \end{bmatrix} \mathbf{x} + \delta \begin{bmatrix} D_{33}^T & -\mathbf{e}_3 \end{bmatrix} \begin{bmatrix} K_3 & \\ & k_4 \end{bmatrix} \mathbf{e}_4.$$

这方程和前述方程一致. 特别地, 系数矩阵

$$D_{43}^T K_4 D_{43} = \begin{bmatrix} D_{33}^T & -\mathbf{e}_3 \end{bmatrix} \begin{bmatrix} K_3 & \\ & k_4 \end{bmatrix} \begin{bmatrix} D_{33} \\ -\mathbf{e}_3^T \end{bmatrix} = D_{33}^T K_3 D_{33} + k_4 \mathbf{e}_3 \mathbf{e}_3^T.$$

我们已经知道  $D_{33}^T K_3 D_{33}$  可逆. 利用命题 1.7.5, 首先  $1 + k_4 \mathbf{e}_3^T (D_{33}^T K_3 D_{33})^{-1} \mathbf{e}_3 = 1 + k_4 (\mathbf{e}_3^T D_{33}^{-T} K_3^{-1} D_{33}^{-1} \mathbf{e}_3) > 0$ , 这是因为  $D_{33}^{-T} \mathbf{e}_3 = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$  的每个元素都是正数. 因此  $D_{43}^T K_4 D_{43}$  可逆.

于是这个方程有唯一解  $\mathbf{x} = (D_{43}^T K_4 D_{43})^{-1} (g\mathbf{m} - \delta D_{43}^T K_4 \mathbf{e}_4)$ . ☺

观察例 1.5.10 和 1.7.6, 我们可以发现共同的模式, 总结在图 1.7.2 中.

## 习题

**练习 1.7.1** 设  $A$  的行简化阶梯形为  $R$ , 行变换对应的可逆矩阵是  $P$ .

1. 求  $\begin{bmatrix} A & 2A \end{bmatrix}$  的行简化阶梯形, 以及行变换对应的可逆矩阵.
2. 求  $\begin{bmatrix} A \\ 2A \end{bmatrix}$  的行简化阶梯形, 以及行变换对应的可逆矩阵.

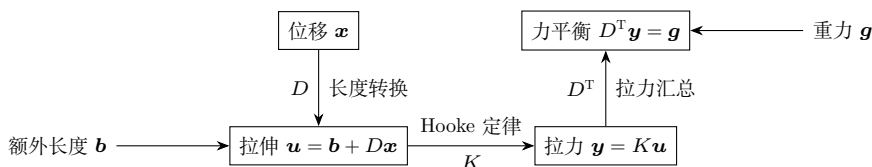


图 1.7.2: 弹簧振子的基本框架

**练习 1.7.2** 求下列矩阵的逆矩阵.

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

2. 
$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

**练习 1.7.3** 计算下列分块矩阵:

1. 
$$\begin{bmatrix} I_n & O \\ A & I_m \end{bmatrix}^{-1}.$$

2. 
$$\begin{bmatrix} O & I_m \\ I_n & A \end{bmatrix}^{-1}.$$

**练习 1.7.4** 设分块矩阵  $X = \begin{bmatrix} O & A \\ B & O \end{bmatrix}$ , 其中  $A, B$  可逆. 试证  $X$  可逆, 并求其逆.

**练习 1.7.5** 设分块矩阵  $U = \begin{bmatrix} A & C \\ O & B \end{bmatrix}$ , 其中  $A, B$  可逆. 试证  $U$  可逆, 并求其逆.

**练习 1.7.6** 设  $A_1, A_2$  分别为  $m, n$  阶方阵, 且存在可逆方阵  $T_1, T_2$ , 使得  $T_1^{-1}A_1T_1$  和  $T_2^{-1}A_2T_2$  都是对角矩阵. 求证: 存在  $m+n$  阶可逆矩阵  $T$ , 使得  $T^{-1} \begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix} T$  是对角矩阵.

**练习 1.7.7** 1. 任取  $m \times n$  矩阵  $X$ , 分块矩阵  $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ , 计算  $\begin{bmatrix} I_m & X \\ O & I_n \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ .

2. 由此判断  $\begin{bmatrix} 1 & & & c_1 \\ & 1 & & c_2 \\ & & \ddots & \vdots \\ & & & 1 & c_{n-1} \\ b_1 & b_2 & \cdots & b_{n-1} & a \end{bmatrix}$  何时可逆, 并在可逆时求其逆.

**练习 1.7.8** 利用 Sherman-Morrison 公式判断下列矩阵何时可逆, 并在可逆时求其逆.

$$\begin{bmatrix} a_1 + 1 & 1 & 1 & \cdots & 1 \\ 1 & a_2 + 1 & 1 & \cdots & 1 \\ 1 & 1 & a_3 + 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 1 & 1 & \cdots & 1 & a_n + 1 \end{bmatrix}.$$



**练习 1.7.9** 1. 求  $A$  使得  $\begin{bmatrix} O & A \\ A & O \end{bmatrix}$  为反对称矩阵.

2. 求 5 阶置换矩阵  $P$ , 满足  $P^5 \neq I_5, P^6 = I_5$ .

3. 求对称矩阵  $A$ , 使得不存在  $B$ , 满足  $A = BB^T$ .

**练习 1.7.10** 给定  $m \times n, n \times m$  矩阵  $A, B$ , 求证:  $I_m + AB$  可逆当且仅当  $I_n + BA$  可逆.

**练习 1.7.11** 设实分块方阵  $X = \begin{bmatrix} A & C \\ O & B \end{bmatrix}$ , 其中  $A$  是方阵. 如果  $X$  与  $X^T$  可交换, 求证:  $C = O$ .

**练习 1.7.12** 设分块对角矩阵  $J = \begin{bmatrix} J_1 & & \\ & J_2 & \\ & & J_3 \end{bmatrix}$ , 其中  $J_1 = \begin{bmatrix} 2 & 1 \\ & 2 \end{bmatrix}, J_2 = \begin{bmatrix} 2 & 1 & \\ & 2 & 1 \\ & & 2 \end{bmatrix}, J_3 = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}$ . 试找出一个五次多项式  $f(x)$ , 使得  $f(J) = O$ .

**练习 1.7.13** 构造  $2n$  阶实方阵  $A$ , 满足  $A^2 = -I_{2n}$ .

**练习 1.7.14** 设  $A, B$  为两个左相抵的行简化阶梯形矩阵.

1. 证明,  $Ax = 0$  与  $Bx = 0$  同解.

2. 证明, 如果  $A$  的最后一列不是主列, 则存在  $x$  使得  $A \begin{bmatrix} x \\ 1 \end{bmatrix} = 0$ .

3. 证明, 如果  $A$  最后一列是主列, 则对任意  $x$ ,  $A \begin{bmatrix} x \\ 1 \end{bmatrix} \neq 0$ .

4. 设  $A, B$  除了最后一列之外, 其他列均相等, 且  $A$  的最后一列不是主列, 证明  $A = B$ .

5. 设  $A, B$  除了最后一列之外, 其他列均相等, 且  $A$  的最后一列是主列, 证明  $A = B$ .

6. 用数学归纳法证明, 左相抵的行简化阶梯形必然相等. 由此证明, 一个矩阵  $A$  的行简化阶梯形唯一.

**练习 1.7.15 (置换的不动点)** 设  $P$  是置换矩阵. 如果  $P$  对应的行变换保持第  $i$  行不变, 则称  $i$  为  $P$  的不动点.

1. 证明  $i$  是  $P$  的不动点当且仅当  $P$  的第  $i$  个对角元为 1.

2. 证明  $P$  的不动点个数等于  $\text{trace}(P)$  (练习 1.5.22).

3. 对任意置换矩阵  $P_1, P_2$ , 证明  $P_1 P_2$  和  $P_2 P_1$  的不动点个数相等.

**练习 1.7.16** 考虑  $\mathbb{R}^n$  上的变换  $f(x) = Ax + b$ , 具有该形式的变换称为**仿射变换**. 仿射变换通常不是线性变换.

1. 证明  $\begin{bmatrix} A & b \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} f(x) \\ 1 \end{bmatrix}$ . 记  $A_f := \begin{bmatrix} A & b \\ 0^T & 1 \end{bmatrix}$ .

2. 证明  $A_f A_g = A_{f \circ g}$ .

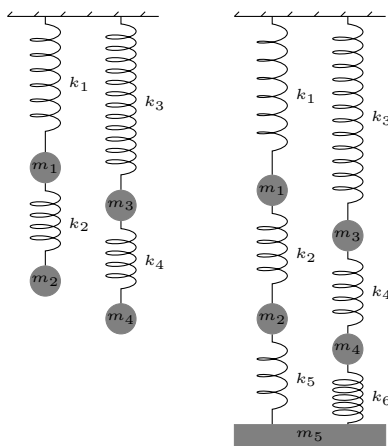


图 1.7.3: 练习 1.7.17

3. 证明当仿射变换  $f$  可逆时, 矩阵  $A_f$  可逆, 且  $A_{f^{-1}} = (A_f)^{-1}$ .

**练习 1.7.17** 图 1.7.3 中有四个弹簧振子. 弹簧的初始长度为  $l_1, l_2, l_3, l_4$ , 振子在稳态的最终长度为  $l_1 + x_1, l_2 + x_2, l_3 + x_3, l_4 + x_4$ .

1. 找到矩阵  $A$  使得  $A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = g \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix}$ . 此时  $A$  的分块结构有什么特点? 你能否从系统中看出原因?

2. 图 1.7.3 中振子下面, 由劲度系数分别为  $k_5, k_6$  的弹簧挂住了一个质量为  $m_5$  的振子. 找到矩阵

$$A \text{ 使得 } A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = g \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{bmatrix}. \text{ 此时 } A \text{ 的分块结构是否还有之前的特点? 为什么?}$$

**练习 1.7.18** 考虑一个元素皆为复数的矩阵. 将每个元素分成实部和虚部, 不难发现对任意复数矩阵, 都存在实数矩阵  $A, B$ , 使得该复数矩阵为  $A + iB$ , 其中  $i$  是虚数单位. 同理, 任意复数向量, 都存在实数向量  $v, w$ , 使得该复数向量为  $v + iw$ . 下面用  $\Re(A), \Re(v), \Im(A), \Im(v)$  来表达复矩阵  $A$  和复向量  $v$  的实部和虚部.

- 用实数矩阵和向量  $A, B, v, w$  来表达  $(A + iB)(v + iw)$  的实部和虚部;
- 对任意实数矩阵  $A, B$ , 求矩阵  $X$ , 使得  $X \begin{bmatrix} \Re(v + iw) \\ \Im(v + iw) \end{bmatrix} = \begin{bmatrix} \Re((A + iB)(v + iw)) \\ \Im((A + iB)(v + iw)) \end{bmatrix}$  对任意实数向量  $v, w$  都成立.
- 考虑映射  $f(a + ib) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ . 验证  $f((a + ib)(c + id)) = f(a + ib)f(c + id)$ .

## 练习 1.7.19 (错位分块对角矩阵) 定义

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \triangle \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & a_{12} & 0 \\ 0 & b_{11} & 0 & b_{12} \\ a_{21} & 0 & a_{22} & 0 \\ 0 & b_{21} & 0 & b_{22} \end{bmatrix}.$$

1. 证明  $(A_1 \triangle B_1)(A_2 \triangle B_2) = (A_1 A_2) \triangle (B_1 B_2)$ .
2. 证明  $A \triangle B$  可逆当且仅当  $A, B$  都可逆; 此时有  $(A \triangle B)^{-1} = A^{-1} \triangle B^{-1}$ .
3. 求  $X$ , 使得对任意 2 阶方阵  $A, B$ , 都有  $X(A \triangle B)X^{-1} = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$ .

## 1.8 LU 分解

在第 1.5 节中可以看到, 以三角矩阵为系数矩阵的方程组不需要消元, 只需要逐个代入就可以求出通解, 即使用回代法或前代法计算. 例 1.5.10 中的方程组是  $D_{33}^T K_3 D_{33} \mathbf{x} = \mathbf{gm}$ . 注意到  $D_{33}^T$  是单位上三角矩阵,  $K_3$  是可逆对角矩阵,  $D_{33}$  是单位下三角矩阵, 因此我们可以逐次求解: 先用回代法求解  $D_{33}^T \mathbf{y} = \mathbf{f}$ ; 然后简单求解  $K_3 \mathbf{u} = \mathbf{y}$ ; 最后用前代法求解  $D_{33} \mathbf{x} = \mathbf{u}$ . 这比先把矩阵  $D_{33}^T K_3 D_{33}$  计算出来再通过消元法计算方程组的解要快得多. (为什么?)

注意, 阶梯形方阵一定是上三角矩阵. 而我们在将矩阵化成阶梯形矩阵时只用到了对换矩阵  $P_{ij}$  和倍加矩阵  $E_{ji;k} (j > i)$ . 在对矩阵  $A$  化简的过程中, 如果没有使用对换矩阵, 那么我们就相当于对矩阵  $A$  的下三角元素按照  $(2, 1), (3, 1), \dots, (n, 1), (3, 2), \dots, (n, 2), \dots, (n, n-1)$  的次序逐一消元. 这类倍加矩阵是下三角矩阵. 记这些初等矩阵的乘积为  $L^{-1}$  (我们知道它可逆), 那么就有  $L^{-1}A = U$ , 其中  $L$  是下三角矩阵,  $U$  是上三角矩阵. 于是, 我们有如下定理.

**定理 1.8.1 (LU 分解)** 如果  $n$  阶方阵  $A$  只使用倍加矩阵  $E_{ji;k} (j > i)$  做行变换就可以化成阶梯形, 那么存在  $n$  阶单位下三角矩阵  $L$  和  $n$  阶上三角矩阵  $U$ , 使得  $A = LU$ .

分解  $A = LU$  称为矩阵  $A$  的 **LU 分解**.

如果  $A$  有 LU 分解  $A = LU$ , 那么求解线性方程组  $A\mathbf{x} = \mathbf{b}$ , 就化成了求解  $L\mathbf{y} = \mathbf{b}$  和  $U\mathbf{x} = \mathbf{y}$  这两个方程组, 而二者的系数矩阵都是三角矩阵, 易于求解.

下面主要考虑可逆矩阵.

**定义 1.8.2 (顺序主子阵)** 方阵  $A$  的左上角  $k \times k$  块, 称为  $A$  的第  $k$  个**顺序主子阵**.

显然,  $n$  阶方阵共有  $n$  个顺序主子阵.

**定理 1.8.3 (可逆矩阵的 LU 分解)** 对  $n$  阶可逆矩阵  $A$ ,  $A$  有 LU 分解  $A = LU$ , 其中  $L$  是  $n$  阶单位下三角矩阵,  $U$  是  $n$  阶上三角矩阵, 当且仅当  $A$  的所有顺序主子阵可逆. 此时,  $A$  的 LU 分解唯一.

证. 充分性: 已知  $A = LU$ .  $A$  可逆,  $L$  为单位下三角阵也可逆, 因此  $U = L^{-1}A$  也可逆, 于是  $L, U$  的所有对角元都不为零. 考虑分块矩阵  $L = \begin{bmatrix} L_{11} & O \\ L_{21} & L_{22} \end{bmatrix}, U = \begin{bmatrix} U_{11} & U_{12} \\ O & U_{22} \end{bmatrix}$ , 其中  $L_{11}, U_{11}$  是  $k \times k$  矩阵. 显然  $L_{11}, U_{11}$  的对角元不为零, 二者都可逆. 注意  $A = LU = \begin{bmatrix} L_{11}U_{11} & L_{11}U_{12} \\ L_{21}U_{11} & L_{21}U_{12} + L_{22}U_{22} \end{bmatrix}$ , 即有  $A$  的第  $k$  个顺序主子阵是  $L_{11}U_{11}$ , 可逆.

必要性: 采用数学归纳法. 当  $n = 1$  时, 显然. 假设命题对  $n - 1$  阶方阵成立, 考虑  $n$  阶方阵  $A$ , 条件已知其所有顺序主子阵可逆.

考虑分块矩阵  $A = \begin{bmatrix} A_1 & \mathbf{x} \\ \mathbf{y}^T & a \end{bmatrix}$ , 其中  $A_1$  是  $(n - 1)$  阶方阵, 且显然满足归纳假设, 因此有 LU 分解  $A_1 = L_1U_1$ , 而  $L_1, U_1$  可逆. 对分块矩阵消元, 有

$$\begin{aligned} A &= \begin{bmatrix} I_{n-1} & \mathbf{0} \\ \mathbf{y}^T A_1^{-1} & 1 \end{bmatrix} \begin{bmatrix} A_1 & \mathbf{x} \\ 0 & a - \mathbf{y}^T A_1^{-1} \mathbf{x} \end{bmatrix} = \begin{bmatrix} I_{n-1} & \mathbf{0} \\ \mathbf{y}^T A_1^{-1} & 1 \end{bmatrix} \begin{bmatrix} L_1 & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} U_1 & L_1^{-1} \mathbf{x} \\ \mathbf{0}^T & a - \mathbf{y}^T A_1^{-1} \mathbf{x} \end{bmatrix} \\ &= \begin{bmatrix} L_1 & \mathbf{0} \\ \mathbf{y}^T U_1^{-1} & 1 \end{bmatrix} \begin{bmatrix} U_1 & L_1^{-1} \mathbf{x} \\ \mathbf{0}^T & a - \mathbf{y}^T U_1^{-1} L_1^{-1} \mathbf{x} \end{bmatrix}, \end{aligned}$$

此即  $A$  的 LU 分解.

唯一性: 设  $A = L_1U_1 = L_2U_2$ . 由  $A$  可逆,  $L_1, L_2$  可逆知,  $U_1, U_2$  可逆. 因此  $L_2^{-1}L_1 = U_2U_1^{-1}$ , 而前者是单位上三角矩阵, 后者是下三角矩阵, 因此二者必为单位矩阵, 即  $L_2^{-1}L_1 = U_2U_1^{-1} = I$ , 因此  $L_1 = L_2, U_1 = U_2$ .  $\square$

**例 1.8.4** 从  $a$  个小球中选  $b$  个, 有多少种选法? 我们中学学过, 这是组合数. 如果  $a \geq b \geq 0$ , 我们有公式  $C_a^b = \frac{a!}{b!(a-b)!}$ . 如果  $a < b$ , 我们定义  $C_a^b = 0$ .

我们考虑对称 **Pascal 矩阵**  $A = [C_{i+j-2}^{i-1}]_{n \times n}$  进行 LU 分解. 为简单起见, 假设

$$n = 4, \text{ 则 } A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}. \text{ 不难看出, 该矩阵的每个元素都是它上边元素和左边}$$

元素的和.

计算可知,  $A$  的 LU 分解为

$$A = LU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

注意,  $U = [C_{j-1}^{i-1}]_{n \times n}$  恰好是上三角 Pascal 矩阵, 每个元素都是它左边元素和左上元素的和; 而  $L = [C_{i-1}^{j-1}]_{n \times n}$  恰好是下三角 Pascal 矩阵, 每个元素都是它上边元素和左上元素的和. 另外, 还可以看到  $L = U^T$ .

如果直接计算矩阵乘法  $LU$ , 可以得到恒等式  $C_{a+b}^t = \sum_{k=0}^t C_a^k C_b^{t-k}$ . 这个恒等式的意义是, 先把  $a+b$  个小球分成各有  $a$  个和  $b$  个的两组, 然后从  $a+b$  个小球中选  $t$  个的选法, 就相当于从  $a$  个小球的组里面选  $k$  个, 再从  $b$  个小球的组里面选  $t-k$  个, 在  $k$  取遍从 0 到  $t$  的自然数时的所有选法.  $\odot$

LU 分解, 就是 Gauss 消元法的简单表达.

可逆矩阵  $A$  存在 LU 分解, 等价于对  $A$  用 Gauss 消元法把  $A$  化成阶梯形, 累积的初等矩阵为  $L^{-1}$ , 而阶梯形矩阵为  $L^{-1}A$ . 之后可以利用 Gauss-Jordan 消元法继续计算: 把主元化成 1, 这可以通过一系列倍乘行变换得到, 记累积的初等矩阵为  $D^{-1}$ , 则此时的矩阵为  $D^{-1}L^{-1}A$ ; 再用一系列倍加行变换把矩阵化成单位矩阵  $I$ , 记累积的初等矩阵为  $U^{-1}$ , 那么  $I = U^{-1}D^{-1}L^{-1}A$ . 这给出了如下结论.

**命题 1.8.5 (LDU 分解)** 如果  $n$  阶可逆方阵  $A$  存在 LU 分解, 那么存在  $n$  阶单位下三角矩阵  $L$ , 对角元素均不为 0 的  $n$  阶对角矩阵  $D$ , 和  $n$  阶单位上三角矩阵  $U$ , 使得  $A = LDU$ , 且该分解唯一.

证. 只需证明唯一性. 假设  $A = L_1 D_1 U_1 = L_2 D_2 U_2$ . 六个矩阵都可逆, 因此  $L_2^{-1} L_1 = D_2 U_2 U_1^{-1} D_1^{-1}$ , 而前者是单位下三角矩阵, 后者是上三角矩阵, 因此二者必为单位矩阵. 立得  $L_1 = L_2, D_2^{-1} D_1 = U_2 U_1^{-1}$ . 类似地,  $D_1 = D_2, U_1 = U_2$ .  $\square$

分解  $A = LDU$  称为矩阵  $A$  的 **LDU 分解**. 注意, 事实上, 我们只需要计算出  $L$  和  $D$ , 就可以直接得到  $U$ , 而不是真需要通过一系列倍加行变换来得到  $U$ .

**推论 1.8.6** 可逆对称矩阵  $A$ , 如果有 LDU 分解  $A = LDU$ , 则  $L = U^T$ .

证. 由  $A$  对称, 有  $LDU = A = A^T = U^T D^T L^T$ . 注意到  $A = LDU = U^T D L^T$  都是  $A$  的 LDU 分解, 根据 LDU 分解的唯一性,  $L = U^T$ .  $\square$

分解  $A = LDL^T$  称为对称矩阵  $A$  的 **LDL<sup>T</sup> 分解**.

注意不是任意对称矩阵都有  $LDL^T$  分解.

我们已经看到, LU 分解可以用来把求解线性方程组  $Ax = b$  的过程简单地形式化. 但是并非任意矩阵, 也并非任意可逆矩阵都可以进行 LU 分解. 另一方面, 我们又知道, 任意可逆矩阵  $A$  通过初等行变换一定可以求出线性方程组  $Ax = b$  的解. 其中的差别就在于对换行变换的使用. 在解方程时, 我们可以知道使用对换行变换只与其出现的先后次序有关系, 而与矩阵元素的值无关. 因此我们可以把解方程的过程中使用的这一系列对换行变换率先施加在矩阵上, 这样得到的矩阵就只需要倍加行变换就可以化成阶梯形了. 这给出了如下结论.

**定理 1.8.7 (PLU 分解)** 给定  $n$  阶可逆矩阵  $A$ , 则存在  $n$  阶置换矩阵  $P$ ,  $n$  阶单位下三角矩阵  $L$ , 和对角元素不为 0 的上三角矩阵  $U$ , 使得  $A = PLU$ . 特别地, 下三角矩阵  $L$  可以选为所有元素的绝对值都不大于 1 的.

证. 采用数学归纳法. 当  $n = 1$  时, 显然. 假设命题对  $n - 1$  阶可逆矩阵成立.

由  $A$  可逆,  $A$  的第一列元素不能全为零, 因此存在置换矩阵  $P_1$ , 使得  $P_1A$  的  $(1, 1)$  元素非零. 于是,

$$P_1A = \begin{bmatrix} a & \mathbf{x}^T \\ \mathbf{y} & A_2 \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \frac{1}{a}\mathbf{y} & I_{n-1} \end{bmatrix} \begin{bmatrix} a & \mathbf{x}^T \\ 0 & A_2 - \frac{1}{a}\mathbf{y}\mathbf{x}^T \end{bmatrix}.$$

由  $A$  可逆且  $a \neq 0$ , 有  $A_2 - \frac{1}{a}\mathbf{y}\mathbf{x}^T$  可逆. 利用归纳假设,  $A_2 - \frac{1}{a}\mathbf{y}\mathbf{x}^T = P_2L_2U_2$ . 因此,

$$\begin{aligned} A &= P_1^{-1} \begin{bmatrix} 1 & \mathbf{0}^T \\ \frac{1}{a}\mathbf{y} & I_{n-1} \end{bmatrix} \begin{bmatrix} a & \mathbf{x}^T \\ 0 & P_2L_2U_2 \end{bmatrix} \\ &= P_1^{-1} \begin{bmatrix} 1 & \mathbf{0}^T \\ \frac{1}{a}\mathbf{y} & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & P_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & L_2 \end{bmatrix} \begin{bmatrix} a & \mathbf{x}^T \\ 0 & U_2 \end{bmatrix} \\ &= P_1^{-1} \begin{bmatrix} 1 & \mathbf{0}^T \\ \frac{1}{a}\mathbf{y} & P_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & L_2 \end{bmatrix} \begin{bmatrix} a & \mathbf{x}^T \\ 0 & U_2 \end{bmatrix} \\ &= P_1^{-1} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & P_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \frac{1}{a}P_2^{-1}\mathbf{y} & I_{n-1} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & L_2 \end{bmatrix} \begin{bmatrix} a & \mathbf{x}^T \\ 0 & U_2 \end{bmatrix} \\ &= P_1^{-1} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & P_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \frac{1}{a}P_2^{-1}\mathbf{y} & L_2 \end{bmatrix} \begin{bmatrix} a & \mathbf{x}^T \\ 0 & U_2 \end{bmatrix}, \end{aligned}$$

其中  $P_1^{-1} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & P_2 \end{bmatrix}$  是置换矩阵,  $\begin{bmatrix} 1 & \mathbf{0}^T \\ \frac{1}{a}P_2^{-1}\mathbf{y} & L_2 \end{bmatrix}$  是单位下三角矩阵,  $\begin{bmatrix} a & \mathbf{x}^T \\ 0 & U_2 \end{bmatrix}$  是上三角矩阵.

特别地, 如果每次选择  $P_1$  时, 使得  $P_1A$  的  $(1, 1)$  元素在第一列中最大, 就可保证  $L$  的所有元素的绝对值都不大于 1, 请读者自行验证.  $\square$

分解  $A = PLU$  称为矩阵  $A$  的 **PLU 分解**.

我们把矩阵消元法写成了矩阵的 LU 分解或 PLU 分解, 从而求解了方程组. 具体来说, 给定条件合适的矩阵  $A$ , 线性方程组  $A\mathbf{x} = \mathbf{b}$  的求解过程为:

1. 首先求出  $A$  的 LU 分解  $A = LU$ ;
2. 然后利用前代法求解下三角矩阵对应的方程组  $L\mathbf{y} = \mathbf{b}$ ;
3. 最后利用回代法求解上三角矩阵对应的方程组  $U\mathbf{x} = \mathbf{y}$ .

对  $n$  阶方阵, 上述算法所需要的四则运算的数目可以计算出来:

1. LU 分解需要做  $\frac{1}{3}(2n^3 - n)$  次四则运算;
2. 前代法和回代法各需要  $n^2$  次四则运算.

因此, 求解线性方程组  $A\mathbf{x} = \mathbf{b}$  的计算量主要用来计算 LU 分解. 注意 LU 分解计算量近似与  $n$  的三次方成正比, 因此, 矩阵的阶数每次翻倍, 对应 LU 分解的计算量将变为原来的 8 倍.

可以看到, LU 分解把矩阵化成两个简单矩阵的乘积. 求解  $A\mathbf{x} = \mathbf{b}$  在得到  $A = LU$  后, 就变为了求解一个以上三角矩阵为系数矩阵的方程组和一个以下三角矩阵为系数矩阵的方程组, 而这两个方程组的求解要简单得多. (当然得到 LU 分解并不简单.) 因此, 把矩阵化为一系列矩阵的乘积, 可能会给解决问题带来很大便利, 也利于分析矩阵的性质, 例如通过 LU 分解, 我们得到了求解系数矩阵是方阵的线性方程组所需计算量与矩阵阶数的关系.

考虑到矩阵和线性变换的关系, 这一过程, 就是把一个复杂的线性映射分解成一系列简单的线性映射的复合的过程; 又考虑到线性变换和线性系统的关系, 这一过程, 也是把一个复杂的线性系统分解成一系列简单的线性系统的“串联”的过程.

一般来说, 把矩阵化成一系列简单矩阵的运算结果的过程, 称为矩阵分解.

我们也可以把矩阵化成若干矩阵的和. 例如, 对任意方阵  $A$ , 都有  $A = \frac{A+A^T}{2} + \frac{A-A^T}{2}$ , 而  $\frac{A+A^T}{2}$  是对称矩阵, 而  $\frac{A-A^T}{2}$  是反对称矩阵, 这对我们把握  $A$  的某些性质有一定帮助.

**例 1.8.8** 回顾例 1.2.10 中的反射变换. 关于直线  $x_2 \cos \theta - x_1 \sin \theta = 0$  的反射变换  $H_\theta$ , 其表示矩阵为  $\begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$ . 为写出表示矩阵, 我们曾考虑  $\mathbf{e}_1, \mathbf{e}_2$  的像, 这并不容易. 然而注意到反射变换的几何含义,  $H_\theta(\mathbf{x})$  和  $\mathbf{x}$  的关系是二者关于该直线对称, 因此  $\mathbf{x} - H_\theta(\mathbf{x}) = 2\mathbf{y}$ , 其中  $\mathbf{y}$  是  $\mathbf{x}$  在直线的法方向 (就是与直线垂直的方向) 的投影. 根据中学学习的平面向量的知识, 我们知道  $\mathbf{y} = (\mathbf{v}^T \mathbf{x})\mathbf{v}$ , 其中  $\mathbf{v}$  是直线的单位法向量. 我

们知道  $\boldsymbol{v} = \begin{bmatrix} \sin \theta \\ -\cos \theta \end{bmatrix}$ , 因此  $H_\theta(\boldsymbol{x}) = \boldsymbol{x} - 2\boldsymbol{v}\boldsymbol{v}^T\boldsymbol{x} = (I_2 - 2\boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}$ , 故反射变换的表示矩阵是  $I_2 - 2\boldsymbol{v}\boldsymbol{v}^T$ . 可以验证, 这和前面含有三角函数的表示矩阵相等. 这个矩阵的优点在于我们不再需要三角函数的参与, 因为法向量和单位法向量容易写出, 具体说来, 直线  $a_1x_1 + a_2x_2 = 0$  的法向量是  $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ .  $\odot$

## 习题

**练习 1.8.1** 求下列矩阵的 LU 分解, 其中  $L$  为单位下三角矩阵.

$$\begin{array}{lll}
 1. \begin{bmatrix} 0 & 3 & -2 \\ -3 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix} & 2. \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}; & 3. \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \\
 4. \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 5 \end{bmatrix} & 5. \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 6 \end{bmatrix} & 6. \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix} \\
 7. \begin{bmatrix} 1 & 5 & 5 & 5 \\ 1 & 2 & 6 & 6 \\ 1 & 2 & 3 & 7 \\ 1 & 2 & 3 & 4 \end{bmatrix} & 8. \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} & 9. \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}
 \end{array}$$

注意观察, 哪些 0 保留在了  $L$  中或者  $U$  中.

**练习 1.8.2** 利用行变换解下列方程

$$\begin{array}{ll}
 1. \begin{bmatrix} 1 & 5 & 5 & 5 \\ 1 & 2 & 6 & 6 \\ 1 & 2 & 3 & 7 \\ 1 & 2 & 3 & 4 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 34 \\ 28 \\ 23 \\ 20 \end{bmatrix} & 2. \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} 34 \\ 28 \\ 23 \\ 20 \end{bmatrix} \\
 3. \begin{bmatrix} 1 & 5 & 5 & 5 \\ 0 & -3 & 1 & 1 \\ 0 & 0 & -3 & 1 \\ 0 & 0 & 0 & -3 \end{bmatrix} \boldsymbol{x} = \boldsymbol{y}, \text{ 这里的 } \boldsymbol{y} \text{ 是上一问的解.} &
 \end{array}$$

**练习 1.8.3** 设  $n$  阶方阵  $A = \begin{bmatrix} a_{11} & \boldsymbol{w}^T \\ \boldsymbol{v} & B \end{bmatrix}$  存在 LU 分解  $A = LU$ , 其中  $\boldsymbol{v} \neq \mathbf{0}$ .

1.  $A$  的左上角元素  $a_{11}$  是否非零?
2. 从上往下对  $A$  做行变换, 将  $\boldsymbol{v}$  变为  $\mathbf{0}$  时, 需要做多少次加法、乘法?
3. 再设  $A$  为对称矩阵. 此时  $\boldsymbol{v}$  和  $\boldsymbol{w}$  有什么关系?  $B$  有什么性质?



4. 利用行变换将  $\boldsymbol{v}$  变为  $\mathbf{0}$  时, 再利用对应的列变换将  $\boldsymbol{w}^T$  变为  $\mathbf{0}$ , 求证此时  $B$  变为对称矩阵. 此时需要做几次加法、乘法?

由此看到, 对称矩阵  $LU$  分解的计算量可以节省一半.

**练习 1.8.4** 设  $n$  阶方阵

$$T = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

利用初等变换证明, 存在分解式  $T = LU$ , 其中  $L$  是下三角矩阵,  $U$  是上三角矩阵. 据此求出  $T^{-1}$ .

Liang, Tian & Yang

## 第二章 子空间和维数

### 2.1 基本概念

给定  $m \times n$  矩阵  $A$ , 矩阵与列向量的乘法给出一个线性映射

$$\begin{aligned} A: \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ \mathbf{x} &\mapsto A\mathbf{x}. \end{aligned}$$

考虑方程组  $A(\mathbf{x}) = A\mathbf{x} = \mathbf{b}$  解是否存在或是否唯一, 就是在考虑  $A$  是否是单射或者满射.

判断  $A$  是否是满射, 需要考虑映射  $A$  的像集, 记为  $\mathcal{R}(A)$ , 即:

$$\mathcal{R}(A) = \{\mathbf{b} \in \mathbb{R}^m \mid \exists \mathbf{x} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{b}\} = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

注意, 对任意矩阵  $A$ ,  $\mathbf{0}_m \in \mathcal{R}(A)$  总成立.

判断  $A$  是否是单射, 需要考虑像集  $\mathcal{R}(A)$  中每个元素的原像. 先看  $\mathbf{0}_m \in \mathcal{R}(A) \subseteq \mathbb{R}^m$  的原像, 记为  $\mathcal{N}(A)$ :

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} \subseteq \mathbb{R}^n.$$

注意, 对任意矩阵  $A$ ,  $\mathbf{0}_n \in \mathcal{N}(A)$  总成立.

**命题 2.1.1** 对  $m \times n$  矩阵  $A$ , 有:

1. 线性映射  $A$  是满射当且仅当  $\mathcal{R}(A) = \mathbb{R}^m$ ;
2. 线性映射  $A$  是单射当且仅当  $\mathcal{N}(A) = \{\mathbf{0}\}$ .

证. 命题 1.4.1 . □

在判断  $A$  是否为单射时, 只需考虑  $\mathbf{0}$  的原像是否唯一, 这是线性映射所具有的很特殊的性质.

**例 2.1.2** 设  $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ ,  $A: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , 则

$$\begin{aligned} \mathcal{R}(A) &= \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^3\} = \left\{ x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \mid x_1, x_2, x_3 \in \mathbb{R} \right\} \\ &= \left\{ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R} \right\}, \end{aligned}$$

是陪域  $\mathbb{R}^3$  中的整个  $x_1x_2$  平面; 而

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^3 \mid A\mathbf{x} = \mathbf{0}\} = \left\{ k \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\},$$

是定义域  $\mathbb{R}^3$  中过原点的一条直线. 映射  $A$  既不是单射, 也不是满射. ⊙

在例 2.1.2 中,  $x_1x_2$  平面和过原点的直线有一个共同点: 它们作为  $\mathbb{R}^3$  的子集, 对  $\mathbb{R}^3$  上的线性运算封闭. 这引导我们作如下定义.

**定义 2.1.3 (子空间)** 设  $\mathcal{M}$  是线性空间  $\mathbb{R}^m$  的非空子集, 如果对任意  $\mathbf{a}, \mathbf{b} \in \mathcal{M}, k \in \mathbb{R}$ , 都满足如下两个条件:

1.  $\mathbf{a} + \mathbf{b} \in \mathcal{M}$ ;

2.  $k\mathbf{a} \in \mathcal{M}$ ;

则称  $\mathcal{M}$  为  $\mathbb{R}^m$  的一个子空间.

注意, 定义 2.1.3 中的两个条件可以合并成一个条件: 对任意  $\mathbf{a}, \mathbf{b} \in \mathcal{M}, k, l \in \mathbb{R}$ , 都有  $k\mathbf{a} + l\mathbf{b} \in \mathcal{M}$ .

特别地,  $\mathbb{R}^m$  有两个平凡子空间, 即  $\{\mathbf{0}\}$  和  $\mathbb{R}^m$  自身. 直观来看, 二者分别是“最小”和“最大”的子空间. 二者之外的子空间, 称为非平凡子空间. 由于  $\mathbf{0} = 0\mathbf{a}$ ,  $\mathbb{R}^m$  的任意子空间都包含零向量. 注意, 空集不是子空间.

**例 2.1.4** 下面给出几个不是子空间的例子.

1. 平面  $\mathbb{R}^2$  中的第一象限, 其中的向量对加法封闭, 对数乘不封闭;
2. 平面  $\mathbb{R}^2$  中  $x_1$  轴和  $x_2$  轴的并集, 对数乘封闭, 但对加法不封闭.

3. 在  $\mathbb{R}^3$  中, 一个不经过原点的平面, 例如  $x_1 + x_2 + x_3 = 1$ , 对加法和数乘都不封闭;
4. 在  $\mathbb{R}^3$  中, 锥面  $x_1^2 + x_2^2 = x_3^2$ , 对数乘封闭, 但对加法不封闭.  $\odot$

**命题 2.1.5** 给定  $m \times n$  矩阵  $A$ , 则

1.  $\mathcal{R}(A)$  是  $\mathbb{R}^m$  的子空间, 称为矩阵  $A$  的**列 (向量) 空间**;
2.  $\mathcal{N}(A)$  是  $\mathbb{R}^n$  的子空间, 称为矩阵  $A$  的**零空间**.

证. 第 1 条: 对任意  $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{R}(A), k_1, k_2 \in \mathbb{R}$ , 都存在  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ , 使得  $A(\mathbf{x}_1) = \mathbf{b}_1, A(\mathbf{x}_2) = \mathbf{b}_2$ . 因此,  $A(k_1\mathbf{x}_1 + k_2\mathbf{x}_2) = k_1\mathbf{b}_1 + k_2\mathbf{b}_2$ , 即  $k_1\mathbf{b}_1 + k_2\mathbf{b}_2 \in \mathcal{R}(A)$ .

第 2 条: 对任意  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{N}(A), k_1, k_2 \in \mathbb{R}$ , 则  $A(k_1\mathbf{x}_1 + k_2\mathbf{x}_2) = k_1A(\mathbf{x}_1) + k_2A(\mathbf{x}_2) = k_1\mathbf{0} + k_2\mathbf{0} = \mathbf{0}$ , 即  $k_1\mathbf{x}_1 + k_2\mathbf{x}_2 \in \mathcal{N}(A)$ .  $\square$

由矩阵导出的子空间  $\mathcal{R}(A)$  和  $\mathcal{N}(A)$  是本章的主要研究对象.

先来看子空间  $\mathcal{R}(A)$ , 它称为矩阵  $A$  的列空间, 其原因在于, 对  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ ,  $\mathcal{R}(A)$  中的任意元素  $A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$  都是  $A$  的列向量组的线性组合 (见定义 1.3.1).

**定义 2.1.6 (线性生成)** 设  $S: \mathbf{a}_1, \cdots, \mathbf{a}_n$  是  $\mathbb{R}^m$  中的向量组, 其线性组合的全体构成  $\mathbb{R}^m$  的一个子集, 记作

$$\text{span}(S) = \text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n) := \{k_1\mathbf{a}_1 + \cdots + k_n\mathbf{a}_n \mid k_1, \cdots, k_n \in \mathbb{R}\},$$

称为由向量组  $\mathbf{a}_1, \cdots, \mathbf{a}_n$  (**线性**) **生成**的子集, 而  $\mathbf{a}_1, \cdots, \mathbf{a}_n$  称为该集合的一组**生成向量**.

**命题 2.1.7** 1. 设  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ , 则  $\mathcal{R}(A) = \text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$ .

2. 向量  $\mathbf{b} \in \mathcal{R}(A)$  当且仅当它可以被  $\mathbf{a}_1, \cdots, \mathbf{a}_n$  线性表示, 这当且仅当方程组  $A\mathbf{x} = \mathbf{b}$  有解.

**命题 2.1.8** 1. 子集  $\text{span}(S)$  是  $\mathbb{R}^m$  的子空间.

2. 如果  $S$  中的向量都在  $\mathbb{R}^m$  的某个子空间中, 则  $\text{span}(S)$  中的向量也都在该子空间中.

证明留给读者.

再来看子空间  $\mathcal{N}(A)$ , 它称为齐次方程组  $A\mathbf{x} = \mathbf{0}$  的**解空间**. 注意, 对非齐次方程组  $A\mathbf{x} = \mathbf{b}$ , 其中  $\mathbf{b} \neq \mathbf{0}$ , 其解集**不是**  $\mathbb{R}^n$  的子空间, 因为它不包含零向量.

对  $m \times n$  矩阵  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ , 子空间  $\mathcal{N}(A)$  中的任意解  $\mathbf{x}$  都给出了零向量的一个线性表示:  $\mathbf{0} = x_1 \mathbf{a}_1 + \cdots x_n \mathbf{a}_n$ . 注意, 零向量总有一个平凡的线性表示:  $\mathbf{0} = 0\mathbf{a}_1 + \cdots 0\mathbf{a}_n$ . 我们考虑  $\mathcal{N}(A) = \{\mathbf{0}\}$  的情形, 有如下关于向量组的定义.

**定义 2.1.9 (线性相关与线性无关)** 给定  $\mathbb{R}^m$  中的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , 如果存在不全为零的一组数  $k_1, \dots, k_n \in \mathbb{R}$ , 使得  $k_1 \mathbf{a}_1 + \cdots + k_n \mathbf{a}_n = \mathbf{0}$ , 则称向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  **线性相关**.

否则,  $k_1 \mathbf{a}_1 + \cdots + k_n \mathbf{a}_n = \mathbf{0}$  一定给出  $k_1 = \cdots = k_n = 0$ , 此时称向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  **线性无关**.

换言之,  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性无关等价于零向量有唯一的线性表示, 即平凡的表示, 也等价于齐次线性方程组  $A\mathbf{x} = \mathbf{0}$  只有零解, 其中  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ . 而  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性相关等价于零向量的线性表示不唯一, 也等价于齐次线性方程组  $A\mathbf{x} = \mathbf{0}$  有非零解.

**例 2.1.10** 1. 平面  $\mathbb{R}^2$  中的两个向量  $\mathbf{a}, \mathbf{b}$  线性相关当且仅当  $\mathbf{a}, \mathbf{b}$  共线; 空间  $\mathbb{R}^3$  中的三个向量  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  线性相关当且仅当  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  共面;

2. 向量组  $\mathbf{a}_1, k\mathbf{a}_1, \mathbf{a}_2$  线性相关:  $-k\mathbf{a}_1 + 1(k\mathbf{a}_1) + 0\mathbf{a}_2 = \mathbf{0}$ ;

3. 如果一个向量组包含零向量, 则它一定线性相关:  $0\mathbf{a}_1 + \cdots + k\mathbf{0} + \cdots + 0\mathbf{a}_s = \mathbf{0}$ , 其中  $k \neq 0$ ;

4. 如果  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性相关, 则任意包含它的向量组<sup>1</sup>  $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{a}_{n+1}, \dots, \mathbf{a}_p$  一定线性相关; 反之, 如果  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性无关, 则它的任意部分组  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  一定线性无关. ☺

下面的例子说明了线性生成和线性无关 (相关) 这两个概念的联系.

**例 2.1.11** 考虑平面  $\mathbb{R}^2$  上的两个向量  $\mathbf{a}, \mathbf{b}$ , 以及二阶方阵  $A = [\mathbf{a} \ \mathbf{b}]$ , 则  $\mathcal{R}(A) = \text{span}(\mathbf{a}, \mathbf{b}) = \{k\mathbf{a} + l\mathbf{b} \mid k, l \in \mathbb{R}\}$  有如下三种可能:

1. 若  $\mathbf{a}, \mathbf{b}$  线性无关, 即不共线, 则  $\text{span}(\mathbf{a}, \mathbf{b})$  是整个平面;
2. 若  $\mathbf{a}, \mathbf{b}$  线性相关, 即共线, 且至少有一个不是零向量, 则  $\text{span}(\mathbf{a}, \mathbf{b})$  是一条过原点的直线;
3. 若  $\mathbf{a} = \mathbf{b} = \mathbf{0}$ , 则  $\text{span}(\mathbf{a}, \mathbf{b})$  只包含原点一个点.

<sup>1</sup>可以称为扩充组.

向量  $\mathbf{a}, \mathbf{b}$  线性无关, 则  $\mathcal{N}(A) = \{\mathbf{0}\}$ . 由定理 1.6.7 可知  $A$  可逆, 且  $\mathcal{R}(A) = \mathbb{R}^2$ .

当  $\mathbf{a}, \mathbf{b}$  线性相关, 且至少有一个不是零向量时, 假设  $A = \begin{bmatrix} a_1 & ka_1 \\ a_2 & ka_2 \end{bmatrix}$ ,  $a_1, a_2$  不全为零, 则

$$\begin{aligned}\mathcal{R}(A) &= \left\{ c \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \mid c \in \mathbb{R} \right\} = \text{span} \left( \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right), \\ \mathcal{N}(A) &= \left\{ c \begin{bmatrix} k \\ -1 \end{bmatrix} \mid c \in \mathbb{R} \right\} = \text{span} \left( \begin{bmatrix} k \\ -1 \end{bmatrix} \right).\end{aligned}$$

几何上看, 两个子空间都是过原点的直线. ⊙

同时考虑线性生成和线性无关的条件, 我们得到如下关于子空间的重要定义.

**定义 2.1.12 (子空间的基)** 给定  $\mathbb{R}^m$  的子空间  $\mathcal{M}$ , 若  $\mathcal{M}$  中存在有限个向量  $\mathbf{a}_1, \dots, \mathbf{a}_n$  满足:

1.  $\mathcal{M}$  中的任意向量都可以被该向量组线性表示, 即  $\mathcal{M} = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ ;
2. 该向量组线性无关;

则称向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  是子空间  $\mathcal{M}$  的一组基.

首先考虑  $\mathcal{M} = \mathbb{R}^m$  的情形. 例 2.1.11 说明, 平面  $\mathbb{R}^2$  中任意两个线性无关的向量都能够线性生成  $\mathbb{R}^2$ , 因此这两个向量是  $\mathbb{R}^2$  的一组基. 而  $\mathbb{R}^2$  中任意三个向量一定线性相关. 一般地, 我们有如下的结论.

**命题 2.1.13** 线性空间  $\mathbb{R}^m$  中, 如果  $n > m$ , 则任意  $n$  个向量都线性相关.

证. 考虑  $\mathbb{R}^m$  中的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$ . 令  $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$ , 则齐次方程组  $A\mathbf{x} = \mathbf{0}$  有  $m$  个方程,  $n$  个未知数, 而未知数个数大于方程个数. 由命题 1.4.13 可知, 该方程组一定有非零解, 即  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性相关. □

因此,  $\mathbb{R}^m$  的任意一组基不可能多于  $m$  个向量. 当恰好有  $m$  个向量时, 我们有如下结果.

**命题 2.1.14** 设  $m$  阶方阵  $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_m]$ , 则  $\mathbf{a}_1, \dots, \mathbf{a}_m$  是  $\mathbb{R}^m$  的一组基当且仅当  $A$  可逆.

证. 必要性: 设  $\mathbf{a}_1, \dots, \mathbf{a}_m$  是  $\mathbb{R}^m$  的一组基, 则向量组线性无关, 即  $A\mathbf{x} = \mathbf{0}$  只有零解. 由定理 1.6.7 可知  $A$  可逆.

充分性: 设  $A$  可逆, 根据定理 1.6.7,  $A\mathbf{x} = \mathbf{0}$  只有零解, 而且  $A\mathbf{x} = \mathbf{b}$  总有解. 前者说明  $\mathbf{a}_1, \dots, \mathbf{a}_m$  线性无关, 后者说明  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_m) = \mathbb{R}^m$ , 因此  $\mathbf{a}_1, \dots, \mathbf{a}_m$  是  $\mathbb{R}^m$  的一组基. □

特别地, 标准坐标向量组  $e_1, \dots, e_m$  是  $\mathbb{R}^m$  的一组基, 称为  $\mathbb{R}^m$  的**标准基**.

我们再来看零向量表示法唯一和一般向量表示法唯一之间的关系.

**命题 2.1.15** 如果向量  $b$  可以被向量组  $a_1, \dots, a_n$  线性表示, 则其表示法唯一当且仅当向量组  $a_1, \dots, a_n$  线性无关.

证. 如果线性表示不唯一, 设  $b = k_1 a_1 + \dots + k_n a_n = l_1 a_1 + \dots + l_n a_n$ , 则  $(k_1 - l_1) a_1 + \dots + (k_n - l_n) a_n = 0$ . 由于两种表示法不相同, 因此  $k_i - l_i$  不全为零, 于是  $a_1, \dots, a_n$  线性相关.

如果  $a_1, \dots, a_n$  线性相关, 即存在不全为零的数  $k_1, \dots, k_n$ , 使得  $k_1 a_1 + \dots + k_n a_n = 0$ . 设  $b = l_1 a_1 + \dots + l_n a_n$  是一种表示法, 则  $b = b + 0 = (l_1 + k_1) a_1 + \dots + (l_n + k_n) a_n$  是另一种不同的表示法.  $\square$

由此可以给出基的一个等价描述.

**命题 2.1.16** 给定  $\mathbb{R}^m$  的子空间  $\mathcal{M}$ , 则  $\mathcal{M}$  中的向量组  $a_1, \dots, a_n$ , 是  $\mathcal{M}$  的一组基, 当且仅当该向量组满足:

1.  $\mathcal{M}$  中的任意向量都可以被  $a_1, \dots, a_n$  线性表示, 即  $\mathcal{M} = \text{span}(a_1, \dots, a_n)$ ;
2. 而且表示法唯一.

**例 2.1.17 (电路网络)** 图 2.1.1 中的有向图  $G$  含有 4 个顶点  $v_j, 1 \leq j \leq 4$  和 5 条带方向的边  $e_i, 1 \leq i \leq 5$ . 它可以描述一个电路网络: 顶点  $v_j$  处的电势为  $x_j$ , 边  $e_i$  上有电阻, 两端的电势差为  $u_i$ .

矩阵  $M$  称为有向图  $G$  的**关联矩阵**  $M_G$ : 如果  $v_j$  是  $e_i$  的头, 则  $M_G$  的  $(i, j)$  元素是 1; 如果  $v_j$  是  $e_i$  的尾, 则  $M_G$  的  $(i, j)$  元素是 -1; 如果  $v_j$  不是  $e_i$  的端点, 则  $M_G$  的  $(i, j)$  元素是 0.

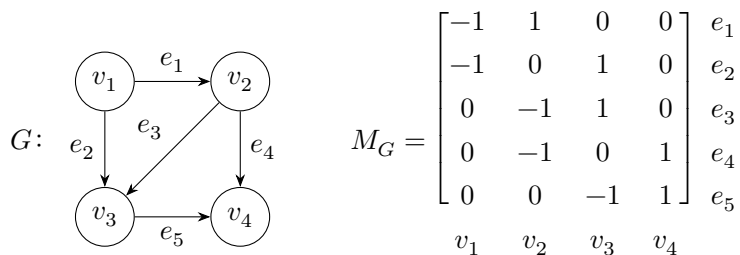


图 2.1.1: 图  $G$  和它的关联矩阵  $M_G$



$$\text{记电势向量为 } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \text{ 电势差向量为 } \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}, \text{ 则 } \mathbf{u} = M_G \mathbf{x} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_1 \\ x_3 - x_2 \\ x_4 - x_2 \\ x_4 - x_3 \end{bmatrix}.$$

矩阵  $M_G$  诱导了线性映射  $M_G: \mathbb{R}^4 \rightarrow \mathbb{R}^5$ , 把电势向量映射到电势差向量. 下面考察  $\mathcal{N}(M_G), \mathcal{R}(M_G)$ . 先看  $\mathcal{N}(M_G)$ . 其中向量, 就是使得任意边上电势差均为零的电势. 不难看出  $M_G \mathbf{x} = \mathbf{0}$  当且仅当  $\mathbf{x} = k\mathbf{1}$ , 其中  $\mathbf{1} = [1 \ 1 \ 1 \ 1]^T$ , 即电势差为零当且仅当各个点的电势都相等. 进一步地, 对任意电势  $\mathbf{x}$ , 都有  $M_G \mathbf{x} = M_G(\mathbf{x} + k\mathbf{1})$ . 这意味着, 所有顶点的电势增加相同数值, 不改变任意边上的电势差.

再看  $\mathcal{R}(M_G)$ . 其中向量, 就是能够实现<sup>2</sup>的电势差. 如果  $\mathbf{u} \in \mathcal{R}(M_G)$ , 那么必然存在电势  $\mathbf{x}$  使得  $\mathbf{u} = M_G \mathbf{x}$ . 容易验证,

$$u_1 - u_2 + u_3 = 0, u_3 - u_4 + u_5 = 0. \quad (2.1.1)$$

注意,  $e_1, e_2, e_3$  或  $e_3, e_4, e_5$  都构成图  $G$  中的一个不带方向的圈. 而  $\mathbf{u}$  的元素之间满足的这种关系, 实际代表着图  $G$  的圈中的边的电势差之和为 0, 这在物理中称为 **Kirchhoff 电压定律**.

$$\text{另一方面, 假设 } \mathbf{u} \text{ 满足 (2.1.1), 那么令 } v_1 \text{ 处电势为 } 0, \text{ 则有 } \mathbf{x} = \begin{bmatrix} 0 \\ u_1 \\ u_1 + u_3 \\ u_1 + u_3 + u_5 \end{bmatrix},$$

于是  $M_G \mathbf{x} = \mathbf{u}$ . 因此,  $\mathbf{u} \in \mathcal{R}(M_G)$  当且仅当它满足 Kirchhoff 电压定律. 换言之, 任意满足 Kirchhoff 电压定律的电势差都可以实现. ( $e_1, e_2, e_5, e_4$  也构成一个圈, 对应的 Kirchhoff 电压定律如何解释?)

☺

例 2.1.17 中, 易知  $\mathcal{N}(M_G) = \text{span}(\mathbf{1})$ , 而  $\mathbf{1}$  是  $\mathcal{N}(M_G)$  的一组基;  $\mathcal{R}(M_G)$  是由  $M_G$  的列向量生成的子空间. 那么后者有基吗? 如果有, 如何计算呢? 对一般的子空间, 我们可以自然地提出两个问题:

1.  $\mathbb{R}^m$  的子空间是否总有一组基?
2.  $\mathbb{R}^m$  的子空间如果有不同的基, 那么这些基有什么共同点?

我们将在下节回答这些问题.

<sup>2</sup>所谓实现, 就是能够实际出现.

## 习题

**练习 2.1.1** 判断下列子集是否是  $\mathbb{R}^n$  的子空间. 其中的子空间是否可以写成线性生成的子空间; 如果可以, 写出一组生成向量.

1.  $\left\{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \mid \sum_{i=1}^n a_i = 0 \right\}.$
2.  $\left\{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \mid \sum_{i=1}^n a_i = 1 \right\}.$
3.  $\left\{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \mid \sum_{i=1}^n a_i^2 = 0 \right\}.$
4.  $\left\{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \mid a_1 = 0 \right\}.$
5.  $\left\{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \mid a_1 \geq 0 \right\}.$
6.  $\left\{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \mid a_1 = \cdots = a_n \right\}.$
7.  $\left\{ \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^T \mid a_1^2 = \cdots = a_n^2 \right\}.$
8.  $\left\{ \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}^T \mid a_1 a_2 a_3 = 0 \right\}.$
9.  $\left\{ \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}^T \mid a_1 = a_2 \right\}.$
10.  $\left\{ \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}^T \mid a_1 \leq a_2 \leq a_3 \right\}.$
11.  $\left\{ \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}^T \mid \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \text{ 是可逆矩阵} \right\}.$
12.  $\left\{ \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}^T \mid \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \text{ 是不可逆矩阵} \right\}.$
13.  $\left\{ \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}^T \mid \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \text{ 是对称矩阵} \right\}.$
14.  $\left\{ \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix}^T \mid \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \text{ 是反对称矩阵} \right\}.$

**练习 2.1.2** 判断满足下列性质的  $\mathbb{R}^3$  子集  $\mathcal{M}$  是否存在. 若存在, 进一步判断哪些是子空间.

1.  $\mathcal{M}$  含有  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ ; 对任意  $\mathbf{v}, \mathbf{w} \in \mathcal{M}$ , 都有  $\mathbf{v} + \mathbf{w} \in \mathcal{M}$ ; 存在  $\mathbf{v} \in \mathcal{M}$ , 使得  $\frac{1}{2}\mathbf{v} \notin \mathcal{M}$ .
2.  $\mathcal{M}$  含有所有形如  $\begin{bmatrix} \cos \theta \\ \sin \theta \\ 1 \end{bmatrix}$  的向量; 对任意  $\mathbf{v} \in \mathcal{M}$ , 都有  $k\mathbf{v} \in \mathcal{M}$ ; 存在  $\mathbf{v}, \mathbf{w} \in \mathcal{M}$ , 使得  $\mathbf{v} + \mathbf{w} \notin \mathcal{M}$ .
3.  $\mathcal{M}$  含有  $\mathbf{e}_1, \mathbf{e}_2$  但不含有  $\mathbf{e}_3$ ; 对任意  $\mathbf{v}, \mathbf{w} \in \mathcal{M}, k \geq 0$ , 都有  $\mathbf{v} + \mathbf{w}, k\mathbf{v} \in \mathcal{M}$ .

**练习 2.1.3** 证明命题 2.1.8.

**练习 2.1.4** 证明,  $\mathcal{R}(O_{m \times n}) = \{\mathbf{0}\}, \mathcal{N}(O_{m \times n}) = \mathbb{R}^n$ ; 如果  $n$  阶方阵  $A$  可逆, 则  $\mathcal{R}(A) = \mathbb{R}^n, \mathcal{N}(A) = \{\mathbf{0}\}$ .

**练习 2.1.5** 证明, 对例 2.1.2 中的  $A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3]$ , 这三个向量中的任意两个都可以作为  $\mathcal{R}(A)$  的一组生成向量, 但是, 其中任意单个向量都不能生成  $\mathcal{R}(A)$ .

**练习 2.1.6** 判断下列  $A, B$  是否具有相同的列空间、零空间, 证明或举出反例.

1.  $A$  为任意矩阵,  $B$  分别为  $2A, \begin{bmatrix} A & A \end{bmatrix}, \begin{bmatrix} A & AC \end{bmatrix}, \begin{bmatrix} A \\ A \end{bmatrix}, \begin{bmatrix} A \\ CA \end{bmatrix}, \begin{bmatrix} A & A \\ O & A \end{bmatrix}, PA, AQ$ , 其中  $P, Q$  可逆.
2.  $A$  为  $n$  阶方阵,  $B$  分别为  $A + I_n, A^2, A^T$ .

**练习 2.1.7** 设矩阵  $A, B$  具有相同的行数和列数, 对下列判断, 证明或举出反例.

1. 如果  $A, B$  有相同的零空间, 那么对任意向量  $\mathbf{b}$ ,  $A\mathbf{x} = \mathbf{b}$  与  $B\mathbf{x} = \mathbf{b}$  一定同解.
2. 如果  $A, B$  有相同的列空间, 那么对任意向量  $\mathbf{b}$ ,  $A\mathbf{x} = \mathbf{b}$  与  $B\mathbf{x} = \mathbf{b}$  一定同解.
3. 如果  $A, B$  有相同的零空间和列空间, 那么对任意向量  $\mathbf{b}$ ,  $A\mathbf{x} = \mathbf{b}$  与  $B\mathbf{x} = \mathbf{b}$  一定同解.

**练习 2.1.8** 设  $A\mathbf{x} = \mathbf{b}$  有解, 证明,

$$1. \mathcal{R}(A) = \mathcal{R}\left(\begin{bmatrix} A & \mathbf{b} \end{bmatrix}\right). \quad 2. \mathcal{N}(A^T) = \mathcal{N}\left(\begin{bmatrix} A^T \\ \mathbf{b}^T \end{bmatrix}\right).$$

**练习 2.1.9** 把  $\mathbb{R}^4$  中的向量  $\mathbf{b}$  表示成  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$  的线性组合.

$$1. \mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix};$$

$$2. \mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

**练习 2.1.10** 判断下列向量组是否线性相关.

$$1. \mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} 0 \\ -7 \\ 3 \\ 1 \end{bmatrix};$$

$$2. \mathbf{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \mathbf{a}_5 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}.$$

**练习 2.1.11** 如果向量组  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  中的任何两个都线性无关, 试问该向量组是否一定线性无关?

**练习 2.1.12** 设  $\mathbb{R}^n$  中向量  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ , 满足  $k_1\mathbf{a}_1 + k_2\mathbf{a}_2 + k_3\mathbf{a}_3 = \mathbf{0}$ , 其中  $k_1k_2 \neq 0$ . 求证:  $\text{span}(\mathbf{a}_1, \mathbf{a}_3) = \text{span}(\mathbf{a}_2, \mathbf{a}_3)$ .

**练习 2.1.13** 证明向量组  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  与向量组  $\mathbf{a}_1 + \mathbf{a}_2, \mathbf{a}_2 + \mathbf{a}_3, \mathbf{a}_3 + \mathbf{a}_1$  线性等价.

**练习 2.1.14** 设向量组  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$  线性无关, 证明向量组  $\mathbf{a}_1, \mathbf{a}_1 + \mathbf{a}_2, \dots, \mathbf{a}_1 + \dots + \mathbf{a}_s$  线性无关.

**练习 2.1.15** 设  $\mathbb{R}^n$  中向量组  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$  线性无关,  $A$  是  $n$  阶可逆矩阵, 求证:  $A\mathbf{a}_1, A\mathbf{a}_2, \dots, A\mathbf{a}_s$  线性无关.

**练习 2.1.16** 证明, 一个线性无关向量组的任意部分组也线性无关; 如果向量组有一个部分组线性相关, 则该向量组也线性相关.

**练习 2.1.17** 证明, 一个向量组线性相关当且仅当其中有一个向量可以被其他向量线性表示.

**练习 2.1.18** 给定  $\mathbb{R}^m$  中向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , 从每个向量中去掉第  $i_1, \dots, i_s$  个分量, 得到  $\mathbb{R}^{m-s}$  中向量组  $\mathbf{a}'_1, \dots, \mathbf{a}'_n$ . 证明,

1. 如果  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性相关, 则  $\mathbf{a}'_1, \dots, \mathbf{a}'_n$  线性相关.
2. 如果  $\mathbf{a}'_1, \dots, \mathbf{a}'_n$  线性无关, 则  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性无关.

**练习 2.1.19 (子空间的和)** 设  $\mathcal{M}, \mathcal{N}$  是  $\mathbb{R}^m$  的两个子空间, 定义集合:

$$\mathcal{M} + \mathcal{N} := \{\mathbf{m} + \mathbf{n} \mid \mathbf{m} \in \mathcal{M}, \mathbf{n} \in \mathcal{N}\}.$$

证明,

1. 集合  $\mathcal{M} + \mathcal{N}$  是  $\mathbb{R}^m$  的子空间, 称为子空间  $\mathcal{M}$  与  $\mathcal{N}$  的**和**.
2. 交集  $\mathcal{M} \cap \mathcal{N}$  是  $\mathbb{R}^m$  的子空间, 称为子空间  $\mathcal{M}$  与  $\mathcal{N}$  的**交**.
3. 集合的交与并满足  $(S_1 \cup S_2) \cap S_3 = (S_1 \cap S_3) \cup (S_2 \cap S_3)$ . 证明或举出反例: 子空间的交与和满足  $(\mathcal{M} + \mathcal{N}) \cap \mathcal{W} = (\mathcal{M} \cap \mathcal{W}) + (\mathcal{N} \cap \mathcal{W})$ .
4. 集合的交与并满足  $(S_1 \cap S_2) \cup S_3 = (S_1 \cup S_3) \cap (S_2 \cup S_3)$ . 证明或举出反例: 子空间的交与和满足  $(\mathcal{M} \cap \mathcal{N}) + \mathcal{W} = (\mathcal{M} + \mathcal{W}) \cap (\mathcal{N} + \mathcal{W})$ .

**练习 2.1.20** 设  $\mathbb{R}^m$  中子空间  $\mathcal{M} = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_s), \mathcal{N} = \text{span}(\mathbf{b}_1, \dots, \mathbf{b}_t)$ , 证明,

$$\mathcal{M} + \mathcal{N} = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_s, \mathbf{b}_1, \dots, \mathbf{b}_t).$$

**练习 2.1.21** 1. 给定  $m \times n, m \times s$  矩阵  $A, B$ , 证明,  $\mathcal{R}(A) + \mathcal{R}(B) = \mathcal{R}(C)$ , 其中  $C = \begin{bmatrix} A & B \end{bmatrix}$ .

2. 给定  $m \times n, l \times n$  矩阵  $A, B$ , 证明,  $\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}(D)$ , 其中  $D = \begin{bmatrix} A \\ B \end{bmatrix}$ .

**练习 2.1.22 (Kirchhoff 电压定律)** 对图 2.1.2 中的电路网络, 令  $M_G$  为对应的关联矩阵. 根据 Kirchhoff 电压定律, 找出一个方程组, 其解集恰为  $\mathcal{R}(M_G)$ . 要求用尽量少的方程 (但不需要写出证明).

**练习 2.1.23 (电路网络拓展)** 设电路网络的关联矩阵是  $M_G$ , 定义图的 Laplace 矩阵  $L_G = M_G^T M_G$ .

1. 矩阵  $L_G$  的元素有什么意义?
2. 求子空间  $\mathcal{N}(L_G)$  的一组生成向量.
3. 求子空间  $\mathcal{R}(L_G)$  的一组生成向量, 要求其中每个向量至少有两个分量为零.
4. 任取向量  $\mathbf{v} \in \mathcal{N}(L_G), \mathbf{w} \in \mathcal{R}(L_G)$ ,  $\mathbf{v}^T \mathbf{w}$  可能取哪些值?
5. 求线性方程组, 其解集恰好是  $\mathcal{R}(L_G)$ .

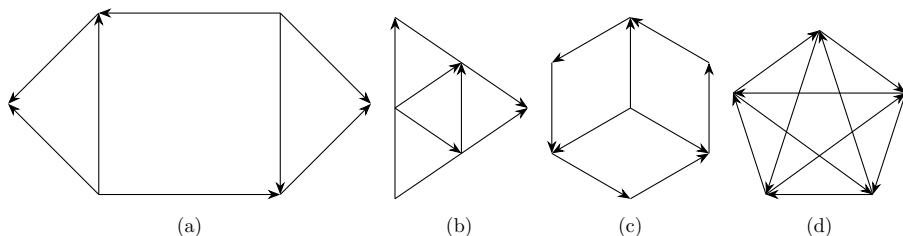


图 2.1.2: Kirchhoff 定律的练习

## 2.2 基和维数

### 2.2.1 向量的线性关系

子空间是否存在基？我们先来讨论矩阵的列空间  $\mathcal{R}(A) = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ . 一个自然的想法是，试图从生成向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  中得到一组基.

设  $S, T$  是  $\mathbb{R}^m$  中的两个向量组，如果  $S$  中的每一个向量都可以被  $T$  线性表示，则称向量组  $S$  可以被  $T$  线性表示.

**定义 2.2.1 (极大线性无关部分组)** 给定  $\mathbb{R}^n$  中的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$ ，设  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  是它的一个部分组，且满足：

1.  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  线性无关；
2.  $\mathbf{a}_1, \dots, \mathbf{a}_n$  可以被  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  线性表示；

则称  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  是  $\mathbf{a}_1, \dots, \mathbf{a}_n$  的一个极大线性无关部分组.

容易看出，部分组  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  是  $\mathbf{a}_1, \dots, \mathbf{a}_n$  的极大线性无关部分组，当且仅当该部分组是  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$  的一组基.

**例 2.2.2** 求向量组  $S: \mathbf{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$  的极大线性无关部分组.

依次考察  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ :  $\mathbf{a}_1 \neq \mathbf{0}$ ，因此  $\mathbf{a}_1$  线性无关；添加  $\mathbf{a}_2$ ，它不能被  $\mathbf{a}_1$  线性表示，而且向量组  $\mathbf{a}_1, \mathbf{a}_2$  线性无关；再考虑  $\mathbf{a}_3$ ， $\mathbf{a}_3 = -\mathbf{a}_1 + \mathbf{a}_2$ ，能被  $\mathbf{a}_1, \mathbf{a}_2$  线性表示，而且  $\mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_3 = \mathbf{0}$ ，即  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  线性相关. 向量组  $\mathbf{a}_1, \mathbf{a}_2$  线性无关，且可以线性表示  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ ，因此是  $S$  的一个极大线性无关部分组.

这种通过依次判断向量组是否线性无关得到极大线性无关部分组的方法称为“筛选法”. 对  $\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_2$  运用筛选法，可得  $\mathbf{a}_1, \mathbf{a}_3$  是极大线性无关部分组. 事实上， $S$  中的任意两个向量构成的部分组都是一个极大线性无关部分组. ☺

从例 2.2.2 可以看出, 理解线性相关(无关)与线性表示的关系对于寻找极大线性无关部分组十分重要.

**命题 2.2.3** 如果向量组  $a_1, \dots, a_n$  线性无关, 那么对任意向量  $b$ , 有

1.  $b$  可以被向量组  $a_1, \dots, a_n$  线性表示, 当且仅当向量组  $a_1, \dots, a_n, b$  线性相关;
2.  $a_1, \dots, a_n, b$  线性无关, 当且仅当  $b$  不能被向量组  $a_1, \dots, a_n$  线性表示.

证. 只证第 1 条, 因为第 2 条是第 1 条的逆否命题.

必要性: 设  $b = k_1 a_1 + \dots + k_n a_n$ , 则  $k_1 a_1 + \dots + k_n a_n + (-1)b = 0$ , 系数不全为零. 根据定义,  $a_1, \dots, a_n, b$  线性相关.

充分性: 由线性相关的定义, 存在不全为零的系数  $k_1, \dots, k_n, k$ , 使得  $k_1 a_1 + \dots + k_n a_n + kb = 0$ . 先证明  $k \neq 0$ . 事实上, 若  $k = 0$ , 则  $k_1 a_1 + \dots + k_n a_n = 0$ . 由  $a_1, \dots, a_n$  线性无关可知,  $k_1 = \dots = k_n = 0$ . 因此  $k_1 = \dots = k_n = k = 0$ , 矛盾. 因此  $k \neq 0$ . 于是  $b = (-\frac{k_1}{k})a_1 + \dots + (-\frac{k_n}{k})a_n$ , 即  $b$  可以被向量组  $a_1, \dots, a_n$  线性表示.  $\square$

下面我们就可以利用“筛选法”得到极大线性无关部分组了.

**命题 2.2.4** 任意向量组  $a_1, \dots, a_n$  都存在极大线性无关部分组.

证. 利用筛选法构造出一个极大线性无关部分组. 首先观察  $a_1$ , 如果  $a_1 = 0$ , 则去掉; 否则保留. 不妨设  $a_1 \neq 0$ . 然后观察  $a_2$ , 如果它可以被前一步得到的向量组  $a_1$  线性表示, 根据命题 2.2.3, 当前向量组  $a_1, a_2$  线性相关, 则去掉  $a_2$ ; 否则保留, 得到的向量组  $a_1, a_2$  线性无关. 类似地逐个考察每个向量, 如果某个向量可以被前一步得到的向量组线性表示, 则去掉; 否则保留. 考察完所有  $n$  个向量后, 得到的部分组就线性无关, 且能线性表示原向量组中的所有向量.  $\square$

命题 2.2.3 是筛选法的关键. 它表明, 在新的向量不能被已经得到的线性无关的部分组线性表示时, 添加它得到的新的部分组仍旧线性无关. 对向量组中的向量按顺序逐个筛选而得到的部分组, 既能线性表示所有的向量, 又线性无关, 因此是一个极大线性无关部分组.

例 2.2.2 还说明了向量组可以有不止一个极大线性无关部分组. 不同的极大线性无关部分组之间有何关系? 首先, 任意极大线性无关部分组都可以线性表示原向量组. 由此, 我们先来具体讨论向量组之间线性表示的关系.

**命题 2.2.5** 设  $S: a_1, \dots, a_n, T: b_1, \dots, b_p$  是  $\mathbb{R}^m$  中的两个向量组, 以下叙述等价:

1.  $S$  可以被  $T$  线性表示;
2. 存在  $p \times n$  矩阵  $U$  满足  $[a_1 \ \dots \ a_n] = [b_1 \ \dots \ b_p] U$ ;

3. 线性生成的子空间满足  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n) \subseteq \text{span}(\mathbf{b}_1, \dots, \mathbf{b}_p)$ .

证. 采用轮转证法.

$1 \Rightarrow 2$ : 对任意  $1 \leq j \leq n$ , 设  $\mathbf{a}_j = u_{1j}\mathbf{b}_1 + \dots + u_{tj}\mathbf{b}_p$ . 令  $U = [u_{ij}]_{p \times n}$ , 由矩阵乘法的定义可得  $[\mathbf{a}_1 \ \dots \ \mathbf{a}_n] = [\mathbf{b}_1 \ \dots \ \mathbf{b}_p] U$ .

$2 \Rightarrow 3$ : 对任意  $\mathbf{c} \in \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ , 都存在  $\mathbf{x} \in \mathbb{R}^n$  使得  $\mathbf{c} = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n] \mathbf{x}$ . 由矩阵乘法的结合律可得

$$\mathbf{c} = ([\mathbf{b}_1 \ \dots \ \mathbf{b}_p] U) \mathbf{x} = [\mathbf{b}_1 \ \dots \ \mathbf{b}_p] (U \mathbf{x}) \in \text{span}(\mathbf{b}_1, \dots, \mathbf{b}_p).$$

$3 \Rightarrow 1$ : 对任意  $1 \leq j \leq n$ , 都有  $\mathbf{a}_j \in \text{span}(\mathbf{b}_1, \dots, \mathbf{b}_p)$ , 即  $\mathbf{a}_j$  可由  $\mathbf{b}_1, \dots, \mathbf{b}_p$  线性表示.  $\square$

由此易得如下结论.

**命题 2.2.6** 给定  $\mathbb{R}^m$  中的三个向量组  $S: \mathbf{a}_1, \dots, \mathbf{a}_n, T: \mathbf{b}_1, \dots, \mathbf{b}_p, R: \mathbf{c}_1, \dots, \mathbf{c}_q$ . 若向量组  $S$  可以被  $T$  线性表示,  $T$  可以被  $R$  线性表示, 则  $S$  可以被  $R$  线性表示. 这称为向量组之间线性表示的传递性.

证. 根据命题 2.2.5, 存在矩阵  $U, V$ , 满足

$$[\mathbf{a}_1 \ \dots \ \mathbf{a}_n] = [\mathbf{b}_1 \ \dots \ \mathbf{b}_p] U, [\mathbf{b}_1 \ \dots \ \mathbf{b}_p] = [\mathbf{c}_1 \ \dots \ \mathbf{c}_r] V.$$

因此,  $[\mathbf{a}_1 \ \dots \ \mathbf{a}_n] = [\mathbf{c}_1 \ \dots \ \mathbf{c}_r] VU$ . 再根据命题 2.2.5 即得.  $\square$

如果两个向量组可以互相线性表示, 则称二者**线性等价**. 线性等价, 指的是向量组在线性组合或线性生成的意义下等价, 尽管两个向量组中的向量可以不全相同.

特别地, 如果一个向量组和某个向量线性等价, 则称该向量组中的向量**共线**.

**例 2.2.7** 以下的向量组都和向量组  $\mathbf{a}_1, \mathbf{a}_2$  线性等价:

1. 向量组  $\mathbf{a}_1, \mathbf{a}_1, \mathbf{a}_2$ ;
2. 向量组  $\mathbf{a}_2, \mathbf{a}_1$ ;
3. 向量组  $c\mathbf{a}_1, \mathbf{a}_2$ , 这里  $c \neq 0$ ;
4. 向量组  $\mathbf{a}_1 + k\mathbf{a}_2, \mathbf{a}_2$ : 这两个向量都可以被  $\mathbf{a}_1, \mathbf{a}_2$  表示, 而反过来,  $\mathbf{a}_1 = (\mathbf{a}_1 + k\mathbf{a}_2) - k\mathbf{a}_2$ , 因此  $\mathbf{a}_1, \mathbf{a}_2$  也可以被这两个向量表示.

注意到, 后三种情形分别对应右乘三种初等矩阵:

$$\begin{aligned} \begin{bmatrix} \mathbf{a}_2 & \mathbf{a}_1 \end{bmatrix} &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \\ \begin{bmatrix} c\mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \\ \begin{bmatrix} \mathbf{a}_1 + k\mathbf{a}_2 & \mathbf{a}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}. \end{aligned}$$

☺

**命题 2.2.8** 设  $S, T$  是  $\mathbb{R}^m$  中的两个向量组, 则  $S$  与  $T$  线性等价当且仅当  $\text{span}(S) = \text{span}(T)$ .

证. 命题 2.2.5 第 3 条. □

可见, 给定向量组的任意两个极大线性无关部分组一定线性等价.

**命题 2.2.9** 向量组的线性等价是等价关系.

证. 只需验证如下三条性质: 反身性, 即任意向量组都和它自己线性等价; 对称性, 即如果向量组  $S$  和  $T$  线性等价, 那么  $T$  和  $S$  线性等价; 传递性, 即如果向量组  $S$  和  $T$  线性等价,  $T$  和  $R$  线性等价, 那么  $S$  和  $R$  线性等价.

尽皆显然. □

向量组的线性等价既然是等价关系, 我们就可以考虑这一等价关系中的不变量. 命题 2.2.8 告诉我们, 向量组生成的子空间是一个不变量.

我们先来讨论向量组中的向量个数与线性表示之间的关系.

**命题 2.2.10** 设向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  可以被  $\mathbf{b}_1, \dots, \mathbf{b}_p$  线性表示.

1. 如果  $n > p$ , 则  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性相关;

2. 如果  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性无关, 则  $n \leq p$ .

证. 只需证明第 1 条, 因为第 2 条是第 1 条的逆否命题.

根据命题 2.2.5, 存在  $p \times n$  矩阵  $C$ , 满足  $[\mathbf{a}_1 \ \dots \ \mathbf{a}_n] = [\mathbf{b}_1 \ \dots \ \mathbf{b}_p] C$ . 考虑齐次方程组  $C\mathbf{x} = \mathbf{0}$ . 因为  $n > p$ , 即未知数个数大于方程个数, 命题 1.4.13 推得方程组存在非零解  $\mathbf{k}$ . 因此,  $[\mathbf{a}_1 \ \dots \ \mathbf{a}_n] \mathbf{k} = [\mathbf{b}_1 \ \dots \ \mathbf{b}_p] C\mathbf{k} = \mathbf{0}$ , 即  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性相关. □



粗略地说, 线性相关的向量组在线性生成的意义下有“多余”的向量. 直观地看命题 2.2.10, 如果一个向量个数多的向量组可以被向量个数少的向量组线性表示, 则多的向量组中一定有多余的向量, 亦即该向量组线性相关; 换言之, 如果一个向量组线性无关, 亦即没有多余的向量时, 那么只有向量个数不少于它的向量组才可能表示出该向量组. 反之, 如果向量组有多余的向量, 那么通过筛选出它的一个极大线性无关部分组, 就可以用向量个数更少的向量组生成它.

考虑到向量组线性等价的定义, 我们有如下重要结论.

**推论 2.2.11** 设向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  和  $\mathbf{b}_1, \dots, \mathbf{b}_p$  线性等价, 如果两个向量组都线性无关, 则  $n = p$ .

由此立得, 一个向量组的任意两个极大线性无关部分组中向量个数都相同.

**定义 2.2.12 (秩)** 一个向量组  $S$  的任意一个极大线性无关部分组中向量的个数称为这个向量组的秩, 记为  $\text{rank}(S)$ . 一个只包含零向量的向量组的秩定义为零.

这说明向量组的秩也是向量组之间线性等价这一等价关系中的不变量.

### 2.2.2 基和维数

命题 2.2.4 说明, 形如  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$  的子空间总存在一组基. 类似的结论很容易推广到一般的子空间上.

**定理 2.2.13 (基存在定理)** 给定  $\mathbb{R}^m$  的子空间  $\mathcal{M}$ , 如果  $\mathcal{M} \neq \{\mathbf{0}\}$ , 则  $\mathcal{M}$  存在一组基, 且基中向量个数不大于  $m$ .

证. 任取  $\mathcal{M}$  中非零向量  $\mathbf{a}_1$ , 如果  $\mathcal{M} = \text{span}(\mathbf{a}_1)$ , 则  $\mathbf{a}_1$  即为  $\mathcal{M}$  的一组基. 否则,  $\mathcal{M} \neq \text{span}(\mathbf{a}_1)$ , 则  $\mathcal{M}$  中必存在向量  $\mathbf{a}_2 \notin \text{span}(\mathbf{a}_1)$ . 因为  $\mathbf{a}_2$  不能被  $\mathbf{a}_1$  线性表示, 且  $\mathbf{a}_1$  线性无关, 根据命题 2.2.3,  $\mathbf{a}_1, \mathbf{a}_2$  线性无关. 如果  $\mathcal{M} = \text{span}(\mathbf{a}_1, \mathbf{a}_2)$ , 则  $\mathbf{a}_1, \mathbf{a}_2$  即为  $\mathcal{M}$  的一组基. 否则,  $\mathcal{M} \neq \text{span}(\mathbf{a}_1, \mathbf{a}_2)$ , 则  $\mathcal{M}$  中必存在向量  $\mathbf{a}_3 \notin \text{span}(\mathbf{a}_1, \mathbf{a}_2)$ . 以此类推. 这个过程一定会在  $m$  步以内终止, 这是因为根据命题 2.1.13, 若  $n > m$ , 任意  $n$  个向量都线性相关. 因此这个过程一定能找到  $r$  个向量  $\mathbf{a}_1, \dots, \mathbf{a}_r$ , 且  $r \leq m$ , 使得  $\mathcal{M} = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_r)$ , 且  $\mathbf{a}_1, \dots, \mathbf{a}_r$  线性无关.  $\square$

基存在定理说明,  $\mathbb{R}^m$  的任意子空间  $\mathcal{M}$  一定是某个矩阵的列空间. 事实上, 如果  $\mathcal{M} = \{\mathbf{0}\}$ , 则它是零矩阵的列空间; 否则,  $\mathcal{M}$  存在一组基  $\mathbf{a}_1, \dots, \mathbf{a}_r$ , 令矩阵  $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_r]$ , 则  $\mathcal{M} = \mathcal{R}(A) = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_r)$ .

与基存在定理类似, 我们也可以考虑有包含关系的两个子空间的基的关系.

**定理 2.2.14 (基扩充定理)** 设  $\mathcal{M}, \mathcal{N}$  是  $\mathbb{R}^m$  的子空间, 且  $\mathcal{M} \neq \{\mathbf{0}\}$ . 如果  $\mathcal{M} \subseteq \mathcal{N}$ , 则  $\mathcal{M}$  的任意一组基都能扩充成  $\mathcal{N}$  的一组基. 特别地, 当  $\mathcal{N} = \mathbb{R}^m$  时, 子空间  $\mathcal{M}$  的任意一组基都能扩充成  $\mathbb{R}^m$  的一组基.

证. 设  $\mathbf{a}_1, \dots, \mathbf{a}_r$  是  $\mathcal{M}$  的一组基, 如果  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_r) = \mathcal{N}$ , 则  $\mathbf{a}_1, \dots, \mathbf{a}_r$  也是  $\mathcal{N}$  的一组基. 否则,  $\mathcal{N}$  中必存在向量  $\mathbf{a}_{r+1} \notin \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_r)$ , 且  $\mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{a}_{r+1}$  线性无关; 如果  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_{r+1}) = \mathcal{N}$ , 证毕. 否则, 以此类推. 因为  $\mathcal{N}$  一定存在一组基, 且基中向量个数不大于  $m$ , 所以有限步之后, 这个过程一定能够得到  $\mathcal{N}$  的一组基.  $\square$

基扩充定理对讨论有包含关系的子空间非常有用. 注意, 考虑问题时要“从小到大”, 如果反过来, 先取定  $\mathcal{N}$  的一组基, 将很难保证这组基与  $\mathcal{M}$  的关系.

与向量组秩的定义类似, 根据推论 2.2.11, 一个子空间的任意两组基包含的向量个数都相同.

**定义 2.2.15 (维数)** 一个子空间  $\mathcal{M}$  的任意一组基中向量的个数称为这个子空间的维数, 记为  $\dim \mathcal{M}$ . 平凡子空间  $\{\mathbf{0}\}$  的维数定义为零.

维数是  $r$  的子空间, 常称为  $r$  维子空间.

在例 2.2.2 中, 向量组  $S$  包含三个向量, 而  $\text{span}(S) = \mathcal{N}(A)$ , 其中  $A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ . 向量组中的任意两个向量都构成  $\mathcal{N}(A)$  的一组基, 因此  $\dim \mathcal{N}(A) = 2$ . 其中三个向量, 各自生成了三个子空间:  $\text{span}(\mathbf{a}_1), \text{span}(\mathbf{a}_2), \text{span}(\mathbf{a}_3)$ , 分别对应三条过原点的直线, 维数是 1. 其中任意两个子空间的交集都是原点一个点, 即平凡子空间  $\{\mathbf{0}\}$ , 维数是 0. 可以看到, 这里定义的子空间的维数与几何直观相符.

**定理 2.2.16** 1. 线性空间  $\mathbb{R}^m$  的维数是  $m$ ;

2. 设  $\mathcal{M}$  是  $\mathbb{R}^m$  的子空间, 则  $\dim \mathcal{M} \leq m$ ;

3. 设  $\mathcal{M}, \mathcal{N}$  是  $\mathbb{R}^m$  的两个子空间, 且  $\mathcal{M} \subseteq \mathcal{N}$ , 则  $\dim \mathcal{M} \leq \dim \mathcal{N}$ .

证. 第 1 条由  $\mathbb{R}^m$  有一组标准基  $\mathbf{e}_1, \dots, \mathbf{e}_m$  立得. 第 2 条由基存在定理 2.2.13 可得; 第 3 条由基扩张定理 2.2.14 可得.  $\square$

注意, 定理 2.2.16 的第 3 条的逆命题并不成立: 两个子空间的维数满足  $\dim \mathcal{M} \leq \dim \mathcal{N}$ , 并不一定有  $\mathcal{M} \subseteq \mathcal{N}$ .

在维数已知的情况下, 可以更简单地判定一个向量组是否为基.

**命题 2.2.17** 设  $\mathcal{M}$  是  $\mathbb{R}^m$  的  $r$  维子空间, 给定  $\mathcal{M}$  中含有  $r$  个向量的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_r$ .

1. 如果  $\mathbf{a}_1, \dots, \mathbf{a}_r$  线性无关, 则  $\mathbf{a}_1, \dots, \mathbf{a}_r$  是  $\mathcal{M}$  的一组基;

2. 如果  $\mathcal{M} = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_r)$ , 则  $\mathbf{a}_1, \dots, \mathbf{a}_r$  是  $\mathcal{M}$  的一组基.

证明留给读者.

事实上,  $\dim \mathcal{M} = r$ ;  $\mathbf{a}_1, \dots, \mathbf{a}_r$  线性无关;  $\mathbf{a}_1, \dots, \mathbf{a}_r$  线性生成  $\mathcal{M}$ : 这三个条件中的任意两个都可以推出另外一个, 因此都可以作为基的判定条件.

**命题 2.2.18** 设  $\mathcal{M}, \mathcal{N}$  是  $\mathbb{R}^m$  的两个子空间, 且  $\mathcal{M} \subseteq \mathcal{N}$ . 如果  $\dim \mathcal{M} = \dim \mathcal{N}$ , 则  $\mathcal{M} = \mathcal{N}$ .

证. 取  $\mathcal{M}$  的一组基  $\mathbf{a}_1, \dots, \mathbf{a}_r$ , 这也是  $\mathcal{N}$  中一个线性无关的向量组. 因为  $\dim \mathcal{N} = \dim \mathcal{M} = r$ , 由命题 2.2.17 第 1 条可知,  $\mathbf{a}_1, \dots, \mathbf{a}_r$  也是  $\mathcal{N}$  的一组基, 所以  $\mathcal{N} = \mathcal{M}$ .  $\square$

作为应用, 我们有如下判断方阵是否可逆的方法.

**命题 2.2.19** 给定  $n$  阶方阵  $A$ ,  $\mathbf{A}: \mathbb{R}^m \rightarrow \mathbb{R}^m, \mathbf{x} \mapsto A\mathbf{x}$  是其诱导的线性变换, 以下叙述等价:

1.  $A$  可逆, 即存在  $n$  阶方阵  $B$ , 满足  $AB = BA = I_m$ ;
2. 存在  $n$  阶方阵  $B$ , 满足  $BA = I_m$ ;
3. 存在  $n$  阶方阵  $B$ , 满足  $AB = I_m$ ;
4.  $A$  是双射;
5.  $A$  是单射;
6.  $A$  是满射.

证.  $1 \Leftrightarrow 4, 2 \Leftrightarrow 5, 3 \Leftrightarrow 6$ : 命题 1.1.2.

$1 \Rightarrow 2, 1 \Rightarrow 3$ : 显然.

记  $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_m]$ .

$2 \Rightarrow 1$ : 由  $BA = I_m$  得,  $A\mathbf{x} = \mathbf{0}$  一定给出  $\mathbf{x} = BA\mathbf{x} = \mathbf{0}$ , 即  $\mathbf{a}_1, \dots, \mathbf{a}_m$  线性无关. 根据命题 2.2.17 第 1 条,  $\mathbf{a}_1, \dots, \mathbf{a}_m$  是  $\mathbb{R}^m$  的一组基. 再由命题 2.1.14,  $A$  可逆.

$3 \Rightarrow 1$ : 由  $AB = I_m$  得,  $\mathbf{e}_i = A(B\mathbf{e}_i)$ , 即  $\mathbf{e}_1, \dots, \mathbf{e}_m$  可以被  $\mathbf{a}_1, \dots, \mathbf{a}_m$  线性表示. 因此  $\mathbb{R}^m = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_m) \subseteq \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_m) \subseteq \mathbb{R}^m$ , 立得  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_m) = \mathbb{R}^m$ . 根据命题 2.2.17 第 2 条,  $\mathbf{a}_1, \dots, \mathbf{a}_m$  是  $\mathbb{R}^m$  的一组基. 再由命题 2.1.14,  $A$  可逆.  $\square$

## 习题

练习 2.2.1 求下列向量组的极大无关部分组和秩.

$$1. \mathbf{a}_1 = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 4 \\ 6 \\ 2 \\ 2 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} 2 \\ 4 \\ 3 \\ 2 \end{bmatrix};$$

$$2. \mathbf{a}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}.$$

练习 2.2.2 在下列向量组中, 除去哪个向量, 得到的向量组与原向量组线性等价?

$$1. \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

$$2. \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

$$3. \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}.$$

$$4. \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

练习 2.2.3 用筛选法去掉方程组中所有多余的方程:

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

练习 2.2.4 给定线性无关的向量组  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ , 求  $\mathbf{a}_1 - \mathbf{a}_2, \mathbf{a}_2 - \mathbf{a}_3, \mathbf{a}_3 - \mathbf{a}_1$  所有的极大线性无关部分组.

练习 2.2.5 证明, 一个向量组的任意线性无关的部分组都可以扩充成它的一个极大线性无关部分组.

练习 2.2.6 证明, 如果向量组  $S$  可以被向量组  $T$  线性表示, 则  $\text{rank}(S) \leq \text{rank}(T)$ .

练习 2.2.7 证明, 如果向量组和一个部分组的秩相同, 则两个向量组线性等价.

练习 2.2.8 举例说明秩相等的向量组未必线性等价.

练习 2.2.9 已知向量组的秩是  $r$ , 设  $S$  是一个包含  $r$  个向量的部分组. 证明:

1. 如果  $S$  线性无关, 则  $S$  是原向量组的一个极大线性无关部分组.
2. 如果  $S$  与原向量组线性等价, 则  $S$  是原向量组的一个极大线性无关部分组.

练习 2.2.10 证明命题 2.2.17.

**练习 2.2.11** 求下列子空间的基和维数.

1. 空间  $\mathbb{R}^3$  中平面  $x - y = 0$  与平面  $x + y - 2z = 0$  的交集.
2. 空间  $\mathbb{R}^3$  中与向量  $\begin{bmatrix} 1 & -1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & -2 \end{bmatrix}$  都垂直的向量组成的子空间.

3. 齐次线性方程组  $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}$  的解集.

**练习 2.2.12** 一个三阶方阵, 如果每行、每列以及两个对角线上的元素之和都相等, 则称为幻方矩阵.

判断  $\left\{ \mathbf{a} = \begin{bmatrix} a_1 & \cdots & a_9 \end{bmatrix}^T \in \mathbb{R}^9 \mid \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \text{ 是幻方矩阵} \right\}$  是否是  $\mathbb{R}^9$  的子空间. 如果是, 求它的一组基.

**练习 2.2.13** 任取非零常数  $k_1, \dots, k_n$  满足  $\frac{1}{k_1} + \cdots + \frac{1}{k_n} + 1 \neq 0$ , 求如下向量组的秩:

$$\mathbf{a}_1 = \begin{bmatrix} 1+k_1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ 1+k_2 \\ \vdots \\ 1 \end{bmatrix}, \dots, \mathbf{a}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1+k_n \end{bmatrix}.$$

**练习 2.2.14** 给定  $r$  阶方阵  $P$  和子空间  $\mathcal{M}$  的一组基  $\mathbf{a}_1, \dots, \mathbf{a}_r$ . 令  $\begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_r \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r \end{bmatrix} P$ , 证明,  $\mathbf{b}_1, \dots, \mathbf{b}_r$  是  $\mathcal{M}$  的一组基当且仅当矩阵  $P$  可逆.

**练习 2.2.15 (Steinitz 替换定理)** 设  $S: \mathbf{a}_1, \dots, \mathbf{a}_r$  线性无关, 可被  $T: \mathbf{b}_1, \dots, \mathbf{b}_t$  线性表示, 求证:

1.  $r \leq t$ .
2. 可以选择  $T$  中的  $r$  个向量换成  $S$ , 得到的新的向量组与  $T$  线性等价.

**练习 2.2.16 (平行的平面)** 考虑方程  $x - 3y - z = 12$  和  $x - 3y - z = 0$  的解集. 这两个解集的交集是什么? 如果  $\mathbf{x}_1, \mathbf{x}_2$  分别是第一个和第二个方程的解, 那么  $\mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_1 + 2\mathbf{x}_2, 2\mathbf{x}_1 + \mathbf{x}_2$  分别是哪个方程的解?

提示: 从几何上看, 这两个解集是  $\mathbb{R}^3$  中两个平行的平面, 其中一个经过原点, 因此是二维子空间; 另一个不经过原点, 因此不是子空间.

**练习 2.2.17** 设  $\mathcal{M}, \mathcal{N}$  是  $\mathbb{R}^m$  的两个子空间, 求证维数公式:

$$\dim(\mathcal{M} + \mathcal{N}) = \dim \mathcal{M} + \dim \mathcal{N} - \dim \mathcal{M} \cap \mathcal{N}.$$

提示: 先取  $\mathcal{M} \cap \mathcal{N}$  的一组基, 根据基扩充定理, 分别扩充成  $\mathcal{M}$  和  $\mathcal{N}$  的一组基, 证明这两组基的并集是  $\mathcal{M} + \mathcal{N}$  的一组基.

## 2.3 矩阵的秩

我们先给出矩阵秩的定义.

**定义 2.3.1 (秩)** 矩阵  $A$  的列空间的维数  $\dim \mathcal{R}(A)$  称为矩阵  $A$  的秩, 记为  $\text{rank}(A)$ .

换言之, 矩阵  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$  的秩就是它的列向量组的秩. 利用命题 2.2.4 中的筛选法来找出  $\mathbf{a}_1, \dots, \mathbf{a}_n$  的极大线性无关部分组, 是我们已知的方法. 但计算过程中需要多次判断向量组的线性关系, 读者可以发现有些计算是不必要的. 有没有更简单的方法呢?

**例 2.3.2** 求下列  $\mathbb{R}^4$  中的向量组的极大线性无关部分组:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 2 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ -1 \\ -2 \\ 4 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 0 \\ 2 \\ 6 \\ -2 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} -3 \\ -1 \\ 3 \\ 4 \end{bmatrix}, \mathbf{a}_5 = \begin{bmatrix} -1 \\ 0 \\ -4 \\ -7 \end{bmatrix}.$$

判断向量组是否线性相关, 等价于判断对应的齐次线性方程组

$$\begin{cases} x_1 + x_2 - 3x_4 - x_5 = 0, \\ x_1 - x_2 + 2x_3 - x_4 = 0, \\ 4x_1 - 2x_2 + 6x_3 + 3x_4 - 4x_5 = 0, \\ 2x_1 + 4x_2 - 2x_3 + 4x_4 - 7x_5 = 0, \end{cases}$$

是否有非零解, 而解又对应了线性表示的表示法. 这个向量组显然线性相关 (为什么?), 不过我们还是先按部就班地计算. 这是例 1.4.11 中的方程组, 对系数矩阵  $A$  做初等行变换得到阶梯形矩阵  $B$ :

$$A = \begin{bmatrix} 1 & 1 & 0 & -3 & -1 \\ 1 & -1 & 2 & -1 & 0 \\ 4 & -2 & 6 & 3 & -4 \\ 2 & 4 & -2 & 4 & -7 \end{bmatrix} \rightarrow B = \begin{bmatrix} 1 & 1 & 0 & -3 & -1 \\ 0 & -2 & 2 & 2 & 1 \\ 0 & 0 & 0 & 9 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

可以看到, 如果我们判断原向量组的部分组, 如  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ , 是否线性相关, 则需要对  $A$  的前三列做初等行变换得到阶梯形:

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 2 \\ 4 & -2 & 6 \\ 2 & 4 & -2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

不难发现, 判断部分组的计算事实上是判断原向量组的计算的一部分. 因此只需对整个向量组做一次计算就足够了. 注意阶梯形对应的方程组与原方程组同解. 换用向量组的

语言来说, 矩阵  $A$  的某些列向量线性相关当且仅当矩阵  $B$  对应的列向量线性相关, 且对应的表示法相同. 因此, 关于矩阵  $A$  的列向量组的极大线性无关部分组的问题就转化为关于矩阵  $B$  的对应问题.

对矩阵  $B$ , 容易根据筛选法看出, 主元所在的列, 即第 1, 2, 4 列, 给出  $B$  的列向量组的一个极大线性无关部分组. 对应地,  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4$  是原向量组的一个极大线性无关部分组.

容易看出,  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5$  和  $\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5$  等都是极大线性无关部分组, 其中包含的向量个数都是 3, 因此向量组的秩为 3, 而矩阵  $A$  的秩也是 3.

最后需要说明的是, 得到极大线性无关部分组只需判断方程组是否有非零解, 因此化成阶梯形就足够了, 并不需要继续化成行简化阶梯形. 但将其他向量用极大线性无关部分组线性表示, 则需要继续计算得到解. 化成行简化阶梯形:

$$B = \begin{bmatrix} 1 & 1 & 0 & -3 & -1 \\ 0 & -2 & 2 & 2 & 1 \\ 0 & 0 & 0 & 9 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow C = \begin{bmatrix} 1 & 0 & 1 & 0 & -\frac{7}{6} \\ 0 & 1 & -1 & 0 & -\frac{5}{6} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

由此即有,  $C = [\mathbf{c}_1 \ \cdots \ \mathbf{c}_5]$  的列满足  $\mathbf{c}_3 = \mathbf{c}_1 - \mathbf{c}_2$ ,  $\mathbf{c}_5 = -\frac{7}{6}\mathbf{c}_1 - \frac{5}{6}\mathbf{c}_2 - \frac{1}{3}\mathbf{c}_4$ . 因此, 对应地,  $\mathbf{a}_3 = \mathbf{a}_1 - \mathbf{a}_2$ ,  $\mathbf{a}_5 = -\frac{7}{6}\mathbf{a}_1 - \frac{5}{6}\mathbf{a}_2 - \frac{1}{3}\mathbf{a}_4$ .  $\odot$

可以看到, Gauss 消去法的计算过程, 对应着列向量组的筛选法, 不过避免了一些重复计算.

在线性方程组的系数矩阵中, 化成阶梯形后主元所在的列, 对应着主变量的系数, 称为主列; 其他列, 对应着自由变量的系数, 称为自由列.

例 2.3.2 的计算过程可以总结出如下结论.

**命题 2.3.3** 设矩阵  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$  与矩阵  $B = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$  左相抵 (即  $A$  经过一系列初等行变换后化成  $B$ , 见定义 1.6.17), 则

1. 部分组  $\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_r}$  线性无关当且仅当对应的部分组  $\mathbf{b}_{i_1}, \cdots, \mathbf{b}_{i_r}$  线性无关;
2.  $A$  的列  $\mathbf{a}_j$  可以被部分组  $\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_r}$  线性表示, 即  $\mathbf{a}_j = k_1\mathbf{a}_{i_1} + \cdots + k_r\mathbf{a}_{i_r}$  当且仅当  $B$  对应的列  $\mathbf{b}_j$  可以被对应的部分组  $\mathbf{b}_{i_1}, \cdots, \mathbf{b}_{i_r}$  线性表示, 且表示法相同, 即  $\mathbf{b}_j = k_1\mathbf{b}_{i_1} + \cdots + k_r\mathbf{b}_{i_r}$ .
3. 部分组  $\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_r}$  是  $A$  的列向量组的极大线性无关部分组当且仅当  $\mathbf{b}_{i_1}, \cdots, \mathbf{b}_{i_r}$  是  $B$  的列向量组的极大线性无关部分组.

上述结论还可以从矩阵乘法的角度解释. 根据命题 1.6.18, 二矩阵左相抵, 等价于存在可逆矩阵  $P$  使得  $B = PA$ , 即  $[\mathbf{b}_{i_1} \ \cdots \ \mathbf{b}_{i_r}] = P[\mathbf{a}_{i_1} \ \cdots \ \mathbf{a}_{i_r}]$ . 而  $\mathbf{a}_{i_1}, \cdots, \mathbf{a}_{i_r}$

线性无关当且仅当  $[a_{i_1} \cdots a_{i_r}]x = Ax = 0$  只有零解. 由于  $P$  可逆, 这等价于  $Bx = PAx = 0$  只有零解. 其他部分类似.

**推论 2.3.4** 矩阵的行简化阶梯形唯一.

证. 记矩阵为  $A$ , 它的一个行简化阶梯形记为  $R$ . 先考虑主列. 根据阶梯形的定义, 阶梯形的某列是主列当且仅当它不是其左边的列的线性组合. 根据命题 2.3.3,  $R$  的列和  $A$  的对应列之间的线性关系相同. 因此  $R$  的某列是主列当且仅当  $A$  的对应列是主列. 而  $A$  的列之间的线性关系是确定的, 因此  $R$  的主列的个数和位置都被唯一确定. 而行简化阶梯形的第  $k$  个主列必须是标准基向量  $e_k$ , 因此  $R$  的主列唯一确定.

再考虑自由列. 可以验证  $R$  的某自由列  $\begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix}$  被主列线性表示, 其表示法中第  $k$

个主列  $e_k$  的系数是  $r_k$ . 根据命题 2.3.3,  $A$  的对应列必被对应主列线性表示, 且表示法相同. 而  $A$  的列之间的线性关系是确定的, 因此  $R$  的自由列的位置和元素都被唯一确定.  $\square$

矩阵  $A$  化成的阶梯形中, 主列给出了  $A$  的列向量组的一个极大线性无关部分组, 这给出如下结论.

**命题 2.3.5**  $\text{rank}(A)$  等于  $A$  化成的阶梯形矩阵的阶梯数.

设  $A$  是  $m \times n$  矩阵,  $\mathbb{R}^n$  的子空间  $\mathcal{R}(A^T)$  由  $A$  的行向量的转置线性生成, 因此称为矩阵  $A$  的**行(向量)空间**.

矩阵  $A$  的行向量对应于齐次方程组  $Ax = 0$  中的方程, 而  $\text{rank}(A^T)$  描述了“有效”方程的个数.

**例 2.3.6** 考虑例 2.3.2 中的线性方程组  $Ax = 0$ . 考虑  $A$  的行空间  $\mathcal{R}(A^T)$ . 对  $A$  做初等行变换, 得到的矩阵的每一行都是  $A$  的行向量的线性组合. 为了记录变换的历史, 我们把行向量记在矩阵左侧, 行向量的记号标在矩阵右侧, 然后将其化成阶梯形:

$$\begin{aligned} & \left[ \begin{array}{ccccc} 1 & 1 & 0 & -3 & -1 \\ 1 & -1 & 2 & -1 & 0 \\ 4 & -2 & 6 & 3 & -4 \\ 2 & 4 & -2 & 4 & -7 \end{array} \right] \begin{array}{l} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \tilde{a}_3^T \\ \tilde{a}_4^T \end{array} \\ \rightarrow & \left[ \begin{array}{ccccc} 1 & 1 & 0 & -3 & -1 \\ 0 & -2 & 2 & 2 & 1 \\ 0 & -6 & 6 & 15 & 0 \\ 0 & 2 & -2 & 10 & -5 \end{array} \right] \begin{array}{l} \tilde{a}_1^T \\ \tilde{a}_2^T - \tilde{a}_1^T \\ \tilde{a}_3^T - 4\tilde{a}_1^T \\ \tilde{a}_4^T - 2\tilde{a}_1^T \end{array} \end{aligned}$$



$$\begin{aligned} &\rightarrow \left[ \begin{array}{ccccc|c} 1 & 1 & 0 & -3 & -1 & \tilde{\mathbf{a}}_1^T \\ 0 & -2 & 2 & 2 & 1 & \tilde{\mathbf{a}}_2^T - \tilde{\mathbf{a}}_1^T \\ 0 & 0 & 0 & 9 & -3 & \tilde{\mathbf{a}}_3^T - 3\tilde{\mathbf{a}}_2^T - \tilde{\mathbf{a}}_1^T \\ 0 & 0 & 0 & 12 & -4 & \tilde{\mathbf{a}}_4^T + \tilde{\mathbf{a}}_2^T - 3\tilde{\mathbf{a}}_1^T \end{array} \right] \\ &\rightarrow \left[ \begin{array}{ccccc|c} 1 & 1 & 0 & -3 & -1 & \tilde{\mathbf{a}}_1^T \\ 0 & -2 & 2 & 2 & 1 & \tilde{\mathbf{a}}_2^T - \tilde{\mathbf{a}}_1^T \\ 0 & 0 & 0 & 9 & -3 & \tilde{\mathbf{a}}_3^T - 3\tilde{\mathbf{a}}_2^T - \tilde{\mathbf{a}}_1^T \\ 0 & 0 & 0 & 0 & 0 & \tilde{\mathbf{a}}_4^T - \frac{4}{3}\tilde{\mathbf{a}}_3^T + 5\tilde{\mathbf{a}}_2^T - \frac{5}{3}\tilde{\mathbf{a}}_1^T \end{array} \right]. \end{aligned}$$

最后一行是  $\tilde{\mathbf{a}}_4^T - \frac{4}{3}\tilde{\mathbf{a}}_3^T + 5\tilde{\mathbf{a}}_2^T - \frac{5}{3}\tilde{\mathbf{a}}_1^T = \mathbf{0}^T$ , 由此可以推出  $\tilde{\mathbf{a}}_4$  可以被  $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_3$  线性表示. 这意味着, 第四个方程是前三个方程的线性组合, 因此是多余的.

另一方面, 阶梯形的每一行都可以被  $A$  的行向量线性表示, 再注意到初等行变换是可逆的, 因此  $A$  的每一行也可以被阶梯形的行向量线性表示. 这说明二者线性等价, 亦即阶梯形的行空间和  $A$  的行空间是同一个子空间. 可见  $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2 - \tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_3 - 3\tilde{\mathbf{a}}_2 - \tilde{\mathbf{a}}_1$  线性无关, 且线性生成行空间. 因此  $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2 - \tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_3 - 3\tilde{\mathbf{a}}_2 - \tilde{\mathbf{a}}_1$  是  $\mathcal{R}(A^T)$  的一组基. 易见  $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_3$  也是  $\mathcal{R}(A^T)$  的一组基.  $\odot$

例 2.3.6 说明, 对矩阵  $A$  做初等行变换不改变行空间  $\mathcal{R}(A^T)$ , 因此  $\mathcal{R}(A^T) = \mathcal{R}(B^T)$ , 其中  $B$  是由  $A$  化成的阶梯形矩阵. 而  $B$  的所有非零行的转置构成  $\mathcal{R}(B^T)$  的一组基, 因此其维数等于阶梯数. 由此得到如下结论.

**命题 2.3.7** 矩阵  $A$  的行空间的维数  $\text{rank}(A^T)$  等于  $A$  化成的阶梯形矩阵的阶梯数.

**命题 2.3.8** 设  $A$  是  $m \times n$  矩阵, 则  $\text{rank}(A^T) = \text{rank}(A)$ .

证. 命题 2.3.5 和 2.3.7.  $\square$

注意,  $\mathcal{R}(A)$  是  $\mathbb{R}^m$  的子空间, 而  $\mathcal{R}(A^T)$  是  $\mathbb{R}^n$  的子空间. 命题 2.3.8 仅仅说明这两个子空间的维数相同.

设  $m \times n$  矩阵  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ , 则  $\mathcal{R}(A) = \text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n) \subseteq \mathbb{R}^m$ . 显然有  $\text{rank}(A) \leq n$ . 当  $\text{rank}(A) = n$  时, 即  $\mathbf{a}_1, \cdots, \mathbf{a}_n$  线性无关时, 称矩阵  $A$  **列满秩**.

由定理 2.2.16 可知,  $\text{rank}(A) \leq m$ . 当  $\text{rank}(A) = m$  时, 称矩阵  $A$  **行满秩**.

特别地, 如果  $\text{rank}(A) = m = n$ , 则称矩阵  $A$  **满秩**.

容易验证如下结论.

**命题 2.3.9** 设  $A$  是  $m \times n$  矩阵,  $\mathbf{A}: \mathbb{R}^n \rightarrow \mathbb{R}^m, \mathbf{x} \mapsto A\mathbf{x}$  是其诱导的线性映射, 则

1.  $A$  是满射当且仅当  $A$  行满秩;

2.  $A$  是单射当且仅当  $A$  列满秩.

3.  $A$  是双射当且仅当  $A$  满秩.

**命题 2.3.10** 1. 矩阵  $A$  可逆当且仅当  $A$  满秩.

2. 矩阵  $A$  是零矩阵当且仅当  $\text{rank}(A) = 0$ .

证明留给读者.

最后, 我们讨论矩阵乘积与列空间的关系.

**命题 2.3.11** 设  $A, B$  分别为  $l \times m$  矩阵和  $m \times n$  矩阵, 则  $\mathcal{R}(AB) \subseteq \mathcal{R}(A)$ . 特别地, 如果  $B$  可逆, 则  $\mathcal{R}(AB) = \mathcal{R}(A)$ .

证. 从列向量的角度看: 设  $B = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$ , 则  $AB = [A\mathbf{b}_1 \ \cdots \ A\mathbf{b}_n]$ . 它的每个列向量  $A\mathbf{b}_j \in \mathcal{R}(A)$ , 因此  $\mathcal{R}(AB) = \text{span}(A\mathbf{b}_1, \dots, A\mathbf{b}_n) \subseteq \mathcal{R}(A)$ .

从映射的角度看:  $A: \mathbb{R}^m \rightarrow \mathbb{R}^l, B: \mathbb{R}^n \rightarrow \mathbb{R}^m, A \circ B: \mathbb{R}^n \rightarrow \mathbb{R}^l$ . 显然,  $A \circ B$  的像集包含在  $A$  的像集里.

当  $B$  可逆时,  $\mathcal{R}(AB) \subseteq \mathcal{R}(A), \mathcal{R}(A) = \mathcal{R}(ABB^{-1}) \subseteq \mathcal{R}(AB)$ . □

**命题 2.3.12** 设  $A, B$  是  $l \times m, m \times n$  矩阵, 则

$$\text{rank}(AB) \leq \text{rank}(A), \quad \text{rank}(AB) \leq \text{rank}(B),$$

即矩阵乘法不增加秩.

证. 由命题 2.3.11 可知,  $\mathcal{R}(AB) \subseteq \mathcal{R}(A)$ , 所以  $\text{rank}(AB) \leq \text{rank}(A)$ . 通过转置, 利用命题 2.3.8, 即有  $\text{rank}(AB) = \text{rank}((AB)^T) = \text{rank}(B^T A^T) \leq \text{rank}(B^T) = \text{rank}(B)$ . □

**命题 2.3.13** 设  $A$  是  $m \times n$  矩阵,  $P, Q$  分别是  $m$  阶和  $n$  阶可逆矩阵, 则  $\text{rank}(PAQ) = \text{rank}(A)$ . 即, 矩阵的秩在初等行变换和初等列变换下不变.

证. 由命题 2.3.3 可知,  $\text{rank}(PA) = \text{rank}(A)$ ; 由命题 2.3.11 可知,  $\mathcal{R}(AQ) = \mathcal{R}(A)$ , 立得  $\text{rank}(AQ) = \text{rank}(A)$ . □

注意, 一般  $\mathcal{R}(PA) \neq \mathcal{R}(A)$ . 例如, 当  $A$  只有一列, 即  $A = \mathbf{a}$  时, 显然  $P\mathbf{a} = k\mathbf{a}$  一般不成立.

根据定理 1.6.7,  $P$  可逆当且仅当它是有限个初等矩阵的乘积. 因此,  $PAQ$  就是对矩阵  $A$  做一系列初等行变换和列变换.

**定义 2.3.14 (相抵)** 如果矩阵  $A$  可以经过一系列初等行变换和初等列变换化成矩阵  $B$ , 则称  $A$  和  $B$  相抵.

类似于通过初等行变换得到行简化阶梯形. 根据命题 2.3.13, 我们有如下结论, 证明留给读者.

**命题 2.3.15** 给定两个  $m \times n$  矩阵  $A, B$ . 那么二者相抵, 当且仅当存在  $m$  阶可逆矩阵  $P$  和  $n$  阶可逆矩阵  $Q$ , 使得  $PAQ = B$ .

**命题 2.3.16** 设  $A$  是  $m \times n$  矩阵, 则存在  $m$  阶可逆矩阵  $P$  和  $n$  阶可逆矩阵  $Q$ , 使得  $PAQ = D_r = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$ , 其中  $r = \text{rank}(A)$ .

命题中的  $D_r$  称为矩阵  $A$  的**相抵标准形**.

**推论 2.3.17** 设  $A, B$  是  $m \times n$  矩阵, 则  $A$  和  $B$  相抵, 当且仅当  $\text{rank}(A) = \text{rank}(B)$ .

矩阵的相抵标准形就是相抵这一等价关系的标准形, 而秩是这一等价关系的不变量. 秩可以用来刻画相抵关系, 而  $m \times n$  矩阵在相抵关系下的等价类共有  $\min\{m, n\} + 1$  个, 分别对应于秩是  $0, 1, \dots, \min\{m, n\}$  的矩阵组成的集合.

## 习题

**练习 2.3.1** 给定矩阵  $A = \mathbf{u}_1 \mathbf{v}_1^T + \mathbf{u}_2 \mathbf{v}_2^T$ .

1. 写出  $A$  的行空间、列空间.
2. 如果  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2$  都不为零, 求  $\text{rank}(A)$ .
3. 讨论这四个向量与  $\text{rank}(A)$  之间的关系.

**练习 2.3.2** 求下列矩阵列空间的一组基.

$$1. \begin{bmatrix} 1 & -1 & 2 & 1 & 0 \\ 2 & -2 & 4 & -2 & 0 \\ 3 & 0 & 6 & -1 & 1 \\ 0 & 3 & 0 & 0 & 1 \end{bmatrix} \quad 2. \begin{bmatrix} 1 & 0 & 0 & 1 & 4 \\ 0 & 1 & 0 & 2 & 5 \\ 0 & 0 & 1 & 3 & 6 \\ 1 & 2 & 3 & 14 & 32 \\ 4 & 5 & 6 & 32 & 77 \end{bmatrix} \quad 3. \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

**练习 2.3.3** 求满足下列条件的  $3 \times 4$  矩阵  $A$  的行简化阶梯形和秩.

1.  $A\mathbf{x} = \mathbf{0}$  解集的一组基为  $\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 6 \\ 1 \end{bmatrix}$ .
2.  $A$  的  $(i, j)$  元为  $a_{ij} = 4$ .
3.  $A$  的  $(i, j)$  元为  $a_{ij} = i + j + 1$ .
4.  $A$  的  $(i, j)$  元为  $a_{ij} = (-1)^{i+j}$ .

**练习 2.3.4** 设矩阵  $A$  的行简化阶梯形为  $R$ , 求下列  $B$  的行简化阶梯形.

1.  $B = \begin{bmatrix} A & A \end{bmatrix}$ .

2.  $B = \begin{bmatrix} A & AC \end{bmatrix}$ .

3.  $B = \begin{bmatrix} A \\ A \end{bmatrix}$ .

4.  $B = \begin{bmatrix} A \\ CA \end{bmatrix}$ .

5.  $B = \begin{bmatrix} A & A \\ 0 & A \end{bmatrix}$ .

6.  $B = PA$ , 其中  $P$  是可逆矩阵.

7.  $A$  有 LU 分解  $A = LU$ , 这里  $L$  为单位下三角阵, 令  $B = U$ .**练习 2.3.5** 求矩阵  $A$  中 “\*” 处的元素, 满足相应的条件.

1.  $A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & * & * \\ 4 & * & * \end{bmatrix}$ , 且  $A$  的秩为 1.

2.  $A = \begin{bmatrix} * & 9 & * \\ 1 & * & * \\ 2 & 6 & -3 \end{bmatrix}$ , 且  $A$  的秩为 1.

3.  $A = \begin{bmatrix} a & b \\ c & * \end{bmatrix}$ ,  $a \neq 0$ , 且  $A$  的秩为 1.

4.  $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 7 \\ 4 & 5 & * \end{bmatrix}$ ,  $a \neq 0$ , 且  $A$  的秩为 1.

**练习 2.3.6** 构造满足下列条件的矩阵  $R$ , 要求其中为 1 的元素尽量多.1.  $R$  为  $4 \times 7$  的阶梯形, 其中主列为第二、四、五列.2.  $R$  为  $4 \times 7$  的阶梯形, 其中主列为第一、三、六、七列.3.  $R$  为  $4 \times 7$  的阶梯形, 其中主列为第四、六列.4.  $R$  为  $4 \times 8$  的行简化阶梯形, 其中自由列为第二、四、五、六列.5.  $R$  为  $4 \times 8$  的行简化阶梯形, 其中自由列为第一、三、六、七、八列.**练习 2.3.7** 说明满足下列条件的矩阵  $A$  是否存在. 如果存在, 举例说明, 要求其中为 0 的元素尽量多.1.  $A$  为  $4 \times 7$  的阶梯形, 其中主列为第二、四、五列,  $A^T$  的主列为第一、三列.2.  $A$  为  $4 \times 7$  的阶梯形, 其中主列为第二、四、五列,  $A^T$  的主列为第二、三、四列.**练习 2.3.8** 证明,  $\text{rank}(kA) = \text{rank}(A)$  ( $k \neq 0$ ),  $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$ .**练习 2.3.9** 对分块矩阵  $C = \begin{bmatrix} A & B \end{bmatrix}$ , 证明,

$$\max\{\text{rank}(A), \text{rank}(B)\} \leq \text{rank}(C) \leq \text{rank}(A) + \text{rank}(B).$$

**练习 2.3.10** 1. 对分块对角矩阵  $C = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$ , 证明,  $\text{rank}(C) = \text{rank}(A) + \text{rank}(B)$ .2. 对分块上三角矩阵  $C = \begin{bmatrix} A & X \\ O & B \end{bmatrix}$ , 证明,  $\text{rank}(C) \leq \text{rank}(A) + \text{rank}(B)$ . 由此证明, 当  $A, B$  可逆时,  $C$  也可逆.**练习 2.3.11** 设  $A, B, C$  分别为  $m \times n, n \times k, k \times s$  矩阵, 证明,

$$\text{rank}(AB) + \text{rank}(BC) \leq \text{rank}(ABC) + \text{rank}(B).$$

**练习 2.3.12** 设  $m \times n$  矩阵  $A$  的秩为 1, 证明, 存在非零向量  $\mathbf{a} \in \mathbb{R}^m, \mathbf{b} \in \mathbb{R}^n$ , 使得  $A = \mathbf{a}\mathbf{b}^T$ .

**练习 2.3.13** 试分析矩阵  $A$  满足什么条件时,  $AB = AC$  可以推出  $B = C$ .

**练习 2.3.14** 设  $m \times n$  矩阵  $A$  列满秩, 求证: 存在行满秩的  $n \times m$  矩阵  $B$ , 使得  $BA = I_n$ .

**练习 2.3.15** 证明命题 2.3.10.

**练习 2.3.16** 设  $A, B$  是  $n$  阶方阵, 利用不等式  $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$ , 证明,

1. 如果  $AB = I_n$ , 则  $A, B$  都可逆, 且  $BA = I_n$ .
2. 如果  $AB$  可逆, 则  $A, B$  都可逆.

**练习 2.3.17** 证明命题 2.3.15.

**练习 2.3.18** 证明, 当  $A$  行满秩时, 仅用初等列变换就可以把它化为相抵标准形; 当  $A$  列满秩时, 仅用初等行变换就可以把它化为相抵标准形.

**练习 2.3.19** 对二阶方阵  $A$ , 如果存在  $n > 2$ , 使得  $A^n = O$ , 求证:  $A^2 = O$ .

**练习 2.3.20** 多项式  $f(x)$  满足  $f(0) = 0$ , 求证: 对任意方阵  $A$ , 都有  $\text{rank}(f(A)) \leq \text{rank}(A)$ .

**练习 2.3.21** 证明反对称矩阵的秩是偶数. 由此证明, 奇数阶反对称矩阵一定不可逆.

**练习 2.3.22** 设  $A$  是  $n$  阶可逆实反对称矩阵,  $\mathbf{b}$  是  $n$  维实列向量, 求证:

1.  $\text{rank}(A + \mathbf{b}\mathbf{b}^T) = n$ ;
2.  $\text{rank}\left(\begin{bmatrix} A & \mathbf{b} \\ \mathbf{b}^T & 0 \end{bmatrix}\right) = n$ .

**练习 2.3.23 (满秩分解)** 1. 求向量  $\mathbf{u}, \mathbf{v}$ , 使得  $\mathbf{u}\mathbf{v}^T = \begin{bmatrix} 3 & 6 & 6 \\ 1 & 2 & 2 \\ 4 & 8 & 8 \end{bmatrix}$ .

2. 设  $A$  是秩为  $r > 0$  的  $m \times n$  的矩阵, 令  $C$  为  $A$  的主列按顺序组成的矩阵, 则  $C$  有几行几列? 令  $R$  为  $A$  的行简化阶梯形的非零行按顺序组成的矩阵, 则  $R$  有几行几列? 求证  $A = CR$ .
3. 任意秩为  $r > 0$  的  $m \times n$  矩阵  $A$  可以分解成列满秩矩阵和行满秩矩阵的乘积, 即分别存在  $m \times r, r \times n$  矩阵  $C, R$ , 且  $\text{rank}(C) = \text{rank}(R) = r$ , 使得  $A = CR$ .
4. 证明, 任意线性映射  $f$  都存在分解  $f = g \circ h$ , 其中  $h$  是线性满射,  $g$  是线性单射. 提示: 对一般的映射, 也有类似的结论, 见练习 1.1.8.

**练习 2.3.24 (秩一分解)** 证明, 任意秩为  $r > 0$  的矩阵  $A$  可以分解成  $r$  个秩为 1 的矩阵的和.

## 2.4 线性方程组的解集

先考虑齐次线性方程组  $A\mathbf{x} = \mathbf{0}$ . 下面计算解空间  $\mathcal{N}(A)$  的基. 继续考虑例 2.3.2, 对  $A$  做初等行变换得到行简化阶梯形:

$$\begin{bmatrix} 1 & 0 & 1 & 0 & -\frac{7}{6} \\ 0 & 1 & -1 & 0 & -\frac{5}{6} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

具体计算见例 1.4.11. 由此得到,  $x_1, x_2, x_4$  是主变量,  $x_3, x_5$  是自由变量. 自由变量的任意一组取值都会唯一地确定一个解. 分别取  $x_3 = 1, x_5 = 0$ , 和  $x_3 = 0, x_5 = 1$  得到

两个解  $\mathbf{k}_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{k}_2 = \begin{bmatrix} \frac{7}{6} \\ \frac{5}{6} \\ 0 \\ \frac{1}{3} \\ 1 \end{bmatrix}$ . 通解公式即为  $\mathbf{k} = x_3\mathbf{k}_1 + x_5\mathbf{k}_2$ , 这与例 1.4.11 中的通

解公式一致. 显然  $\mathbf{k}_1, \mathbf{k}_2$  线性无关, 且能够表示方程组的任意一组解, 因此,  $\mathbf{k}_1, \mathbf{k}_2$  是  $\mathcal{N}(A)$  的一组基.

一般地, 设  $A$  是  $m \times n$  矩阵, 秩为  $r$ , 它的行简化阶梯形包含  $r$  个主变量,  $n - r$  个自由变量. 对齐次方程组  $A\mathbf{x} = \mathbf{0}$ , 设  $\mathbf{k}_i$  是第  $i$  个自由变量取 1, 其余自由变量都取 0 时得到的解, 由此得到  $n - r$  个解  $\mathbf{k}_1, \dots, \mathbf{k}_{n-r}$ .

**定理 2.4.1** 对  $m \times n$  矩阵  $A$ , 上述  $n - r$  个解  $\mathbf{k}_1, \dots, \mathbf{k}_{n-r}$  是零空间  $\mathcal{N}(A)$  的一组基, 其中  $r = \text{rank}(A)$ . 进一步地,  $\dim \mathcal{N}(A) = n - \text{rank}(A)$ .

证. 先证  $\mathbf{k}_1, \dots, \mathbf{k}_{n-r}$  可以线性生成  $\mathcal{N}(A)$ . 设  $\mathbf{k} \in \mathcal{N}(A)$  是一个解, 它在  $n - r$  个自由变量上的取值依次为  $c_1, \dots, c_{n-r}$ . 下证  $\mathbf{k} = c_1\mathbf{k}_1 + \dots + c_{n-r}\mathbf{k}_{n-r}$ . 考虑差  $\mathbf{k}' = c_1\mathbf{k}_1 + \dots + c_{n-r}\mathbf{k}_{n-r} - \mathbf{k}$ . 由于  $\mathcal{N}(A)$  对线性运算封闭, 所以  $\mathbf{k}'$  也是一个解, 并且在所有自由变量上的取值都是零. 代入方程组  $A\mathbf{x} = \mathbf{0}$  可得它必为零解:  $\mathbf{k}' = \mathbf{0}$ .

再证  $\mathbf{k}_1, \dots, \mathbf{k}_{n-r}$  线性无关. 设  $c_1\mathbf{k}_1 + \dots + c_{n-r}\mathbf{k}_{n-r} = \mathbf{0}$ , 而系数  $c_1, \dots, c_{n-r}$  恰好是解  $c_1\mathbf{k}_1 + \dots + c_{n-r}\mathbf{k}_{n-r}$  在  $n - r$  个自由变量上的取值, 因此全为零.  $\square$

注意, 事实上,  $\mathcal{N}(A)$  的这组基, 可以从行简化阶梯形中直接得出, 而不需要真地去解  $r$  个变量的方程组.

齐次线性方程组的解, 可以由解空间  $\mathcal{N}(A)$  完全刻画, 因此, 以后齐次线性方程组的解, 就可以直接用  $\mathcal{N}(A)$  来表示, 而不是一组通解.

再考虑非齐次线性方程组. 我们称  $Ax = 0$  是  $Ax = b$  的导出方程组. 先选定  $Ax = b$  的一个解  $k_0$ , 称为特解. 通过平移向量  $k_0$  可以建立两个解集  $\{x \mid Ax = b\}$  和  $\mathcal{N}(A)$  之间的一个一一对应. 事实上, 对  $Ax = b$  的任意解  $k$ , 都有  $A(k - k_0) = b - b = 0$ , 即  $k - k_0$  是  $Ax = 0$  的一个解. 这定义了两个解集之间的一个映射:

$$\begin{aligned} f: \{x \mid Ax = b\} &\rightarrow \mathcal{N}(A), \\ k &\mapsto k - k_0. \end{aligned} \quad (2.4.1)$$

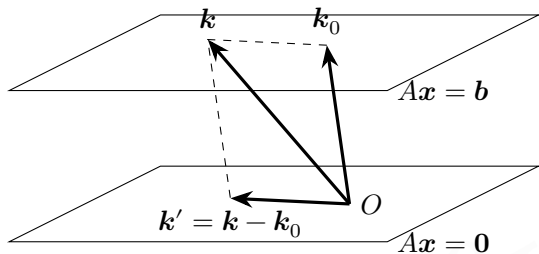


图 2.4.1: 非齐次线性方程组

它可逆, 其逆映射是

$$\begin{aligned} f^{-1}: \mathcal{N}(A) &\rightarrow \{x \mid Ax = b\}, \\ k' &\mapsto k' + k_0. \end{aligned}$$

因此,  $f$  是双射. 例如, 设  $A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ , 则  $Ax = 1$  的解集是  $\mathbb{R}^3$  中的平面  $x_1 + x_2 + x_3 = 1$ , 而  $\mathcal{N}(A)$  是过原点的平面  $x_1 + x_2 + x_3 = 0$ , 两个平面平行. 这就说明, 双射  $f$  通过平移向量  $k_0$  建立了两个解集  $\{x \mid Ax = b\}$  和  $\mathcal{N}(A)$  之间的一一对应, 立刻可得如下结论.

**定理 2.4.2** 设线性方程组  $Ax = b$  的一个特解是  $k_0$ , 其导出方程组的解空间  $\mathcal{N}(A)$  的一组基是  $k_1, \dots, k_{n-r}$ , 其中  $r = \text{rank}(A)$ , 则  $Ax = b$  的解集就是

$$\{k_0 + c_1 k_1 + \dots + c_{n-r} k_{n-r} \mid c_1, \dots, c_{n-r} \in \mathbb{R}\}.$$

注意只要  $b \neq 0$ , 这个解集就不是一个子空间.

下面我们利用  $\mathcal{N}(A)$  的结构来讨论线性方程组的解集. 判定定理 1.4.9, 可以用秩来重新叙述.

**定理 2.4.3 (判别定理)** 对  $n$  个变量的线性方程组  $Ax = b$ , 它的解有如下情形:

1. 它有解, 当且仅当其系数矩阵与增广矩阵秩相等, 即  $\text{rank}(A) = \text{rank}([A \ b])$ ;

2. 它有唯一解, 当且仅当  $\text{rank}(A) = \text{rank}([A \ \mathbf{b}]) = n$ ;
3. 它有无穷多组解, 当且仅当  $\text{rank}(A) = \text{rank}([A \ \mathbf{b}]) < n$ .

假设  $A\mathbf{x} = \mathbf{b}$  有解, 当把增广矩阵  $[A \ \mathbf{b}]$  化成行简化阶梯形时, 系数矩阵  $A$  同时化为了行简化阶梯形. 此时把由  $A$  得到的所有自由变量都取 0, 即可得到一个特解. 下面通过一个例子来说明这一求解过程.

**例 2.4.4** 求解:

$$\begin{cases} x_1 - x_2 - x_4 - x_5 = 1, \\ 2x_1 + x_3 - x_5 = 2, \\ 3x_1 - x_2 - x_3 - x_4 - x_5 = 0. \end{cases}$$

对增广矩阵做初等行变换:

$$\begin{aligned} & \left[ \begin{array}{ccccc|c} 1 & -1 & 0 & 1 & -1 & 1 \\ 2 & 0 & 1 & 0 & -1 & 2 \\ 3 & -1 & -1 & -1 & -1 & 0 \end{array} \right] \rightarrow \left[ \begin{array}{ccccc|c} 1 & -1 & 0 & 1 & -1 & 1 \\ 0 & 2 & 1 & -2 & 1 & 0 \\ 0 & 0 & 2 & 2 & -1 & 3 \end{array} \right] \\ & \rightarrow \left[ \begin{array}{ccccc|c} 1 & -1 & 0 & 1 & -1 & 1 \\ 0 & 1 & \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 1 & -\frac{1}{2} & \frac{3}{2} \end{array} \right] \rightarrow \left[ \begin{array}{ccccc|c} 1 & 0 & 0 & -\frac{1}{2} & -\frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 & -\frac{3}{2} & \frac{3}{4} & -\frac{3}{4} \\ 0 & 0 & 1 & 1 & -\frac{1}{2} & \frac{3}{2} \end{array} \right]. \end{aligned}$$

主变量是  $x_1, x_2, x_3$ , 自由变量为  $x_4, x_5$ . 把  $x_4 = x_5 = 0$  代入, 得到一个特解是  $\mathbf{k}_0 =$

$$\begin{bmatrix} \frac{1}{4} \\ -\frac{3}{4} \\ \frac{3}{2} \\ 0 \\ 0 \end{bmatrix}, \text{ 而导出方程组的一组基为 } \mathbf{k}_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{3}{2} \\ -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{k}_2 = \begin{bmatrix} \frac{1}{4} \\ -\frac{3}{4} \\ \frac{1}{2} \\ 0 \\ 1 \end{bmatrix}. \text{ 因此原方程组的全部解为}$$

$\mathbf{k}_0 + c_1\mathbf{k}_1 + c_2\mathbf{k}_2$ , 其中  $c_1, c_2 \in \mathbb{R}$ .

☺

由定理 2.4.1 得到如下重要的维数公式.

**定理 2.4.5** 设  $A$  是  $m \times n$  矩阵, 则

$$\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = n, \quad \dim \mathcal{R}(A^T) + \dim \mathcal{N}(A^T) = m.$$

其中  $\mathbb{R}^m$  的子空间  $\mathcal{N}(A^T)$ , 称为矩阵  $A$  的左零空间, 得名于其中向量  $\mathbf{x}$  满足  $\mathbf{x}^T A = \mathbf{0}^T$ .

从方程组  $A\mathbf{x} = \mathbf{b}$  的角度来观察定理 2.4.5. 一方面,  $\dim \mathcal{N}(A)$  是自由变量的个数,  $\text{rank}(A) = \dim \mathcal{R}(A)$  是主变量个数, 而  $A$  的列数  $n$  则是变量的总个数. 显然  $\dim \mathcal{N}(A) + \dim \mathcal{R}(A) = n$ . 另一方面,  $\text{rank}(A^T) = \dim \mathcal{R}(A^T)$  是有效方程的个数,  $m$  是总方程个数. 而  $\dim \mathcal{R}(A^T) + \dim \mathcal{N}(A^T) = m$ , 因此  $\dim \mathcal{N}(A^T)$  是多余方程的个数.



**例 2.4.6 (电路网络)** 对例 2.1.17 中的电路网络进行进一步的分析.

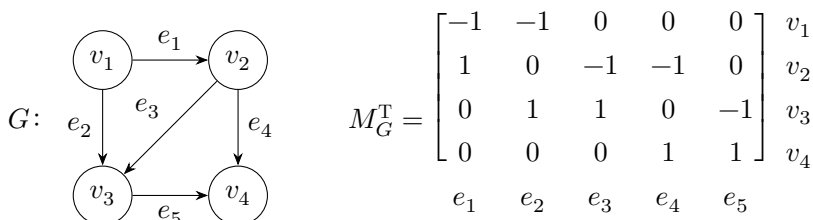


图 2.4.2: 图  $G$  和关联矩阵的转置  $M_G^T$

矩阵  $M_G$  诱导的电势与电势差之间的映射  $\mathbf{M}_G: \mathbb{R}^4 \rightarrow \mathbb{R}^5$ . 由 **Ohm 定律**, 电流与电阻的乘积等于电势差. 假设各边电阻都是 1, 则边上的电流等于两端的电势差. 设  $\mathbf{y} \in \mathbb{R}^5$  表示边上的电流向量, 则转置矩阵  $M_G^T$  诱导了一个映射  $\mathbf{M}_{G^T}: \mathbb{R}^5 \rightarrow \mathbb{R}^4$ , 它把边上的电流映射到顶点流入流出的电流的差.

设  $\mathbf{y} \in \mathcal{N}(M_G^T)$ , 则  $M_G^T \mathbf{y} = 0$  表示在电流  $\mathbf{y}$  中, 每个顶点流入和流出的电流都相等. 零空间  $\mathcal{N}(M_G^T)$  有一组基  $[1 \ -1 \ 1 \ 0 \ 0]^T, [0 \ 0 \ 1 \ -1 \ 1]^T$ , 分别对应图  $G$  上两个圈  $e_1, e_2, e_3$  和  $e_3, e_4, e_5$ . 两个圈组合起来可以得到一个更大的圈  $e_1, e_2, e_4, e_5$ , 对应于向量  $[1 \ -1 \ 0 \ 1 \ -1]^T \in \mathcal{N}(M_G^T)$ . 因此,  $\mathcal{N}(M_G^T)$  的维数等于图  $G$  中独立的圈的个数, 也等于图  $G$  中“面”的个数.

考虑一般的图  $G$ , 设有  $n$  个顶点和  $m$  条边, 则关联矩阵  $M_G$  是  $m \times n$  矩阵. 假设任意两个顶点都能通过边连接 (边可以不止一条), 即图  $G$  是**连通**的. 由维数公式可知,

$$\dim \mathcal{N}(M_G) + \dim \mathcal{R}(M_G) = n, \quad \dim \mathcal{N}(M_G^T) + \dim \mathcal{R}(M_G^T) = m.$$

消去  $\dim \mathcal{R}(M_G) = \dim \mathcal{R}(M_G^T)$  可得

$$n - m + \dim \mathcal{N}(M_G^T) = \dim \mathcal{N}(M_G). \quad (2.4.2)$$

由例 2.1.17 可知  $\dim \mathcal{N}(M_G) = 1$ , 由此可得  $\dim \mathcal{N}(M_G^T) = m - n + 1$ , 这恰好是图  $G$  中“面”的个数. 因此 (2.4.2) 可以写成:

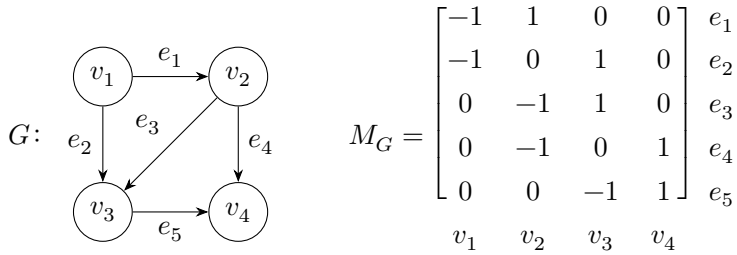
$$\text{“点”的个数} - \text{“边”的个数} + \text{“面”的个数} = 1.$$

这恰是平面上图的 **Euler 公式**.

☺

最后, 来看一个复杂的例子, 以便体会线性方程组的常见应用场景.

**例 2.4.7 (电路网络)** 给定简单连通有向图  $G$ , 为边添加电阻和电源, 构成了一个电路网络. 假设顶点  $v_i$  处的电势为  $x_i$ , 外部流入的电流为  $g_i$ ; 弧  $e_j$  上的电阻为  $r_j$ , 电源提

图 2.4.3: 图  $G$  和关联矩阵  $M_G$ 

供的电压为  $b_j$ ; 其上通过的电流为  $y_j$ . 记电势向量为  $\mathbf{x}$ , 电流向量  $\mathbf{y}$ , 外来电流  $\mathbf{g}$ , 电源  $\mathbf{b}$ , 而电阻矩阵  $R = \text{diag}(r_j)$ .

先来观察  $\mathbf{u} = M_G \mathbf{x} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_1 \\ x_3 - x_2 \\ x_4 - x_2 \\ x_4 - x_3 \end{bmatrix}$ , 它表示的就是边两端的电势差.  $\mathbf{b} - \mathbf{u}$  就是电阻

引起的电势降. Ohm 定律告诉我们,  $r_j y_j = b_j - u_j$ , 亦即而  $R\mathbf{y} = \mathbf{b} - \mathbf{u} = \mathbf{b} - M_G \mathbf{x}$ .

另一方面, 观察  $\mathbf{f} = M_G^T \mathbf{y} = \begin{bmatrix} -y_1 - y_2 \\ y_1 - y_3 - y_4 \\ y_2 + y_3 - y_5 \\ y_4 + y_5 \end{bmatrix}$ , 它表示的是系统内部流入顶点的总电流.

电路稳定时, 每个顶点处流入流出的总电流应该为 0, 这称为 **Kirchhoff 电流定律**. 根据电流平衡,  $\mathbf{g} = \mathbf{f} = M_G^T \mathbf{y}$ . 于是我们列出了线性方程组

$$\begin{cases} R\mathbf{y} + M_G \mathbf{x} = \mathbf{b}, \\ M_G^T \mathbf{y} = \mathbf{g}, \end{cases} \quad \text{或} \quad \begin{bmatrix} R & M_G \\ M_G^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{g} \end{bmatrix}.$$

系数矩阵是对称矩阵, 未知数是电流和电势. 因为对角矩阵  $R$  可逆, 上式可以化简成

$$M_G^T R^{-1} M_G \mathbf{x} = M_G^T R^{-1} \mathbf{b} - \mathbf{g}, \quad (2.4.3)$$

其中未知数是电势. 注意,  $n$  阶方阵  $M_G^T R^{-1} M_G$  不可逆, 它的秩  $\text{rank}(M_G^T R^{-1} M_G) = \text{rank}(M_G) = n - 1$  (为什么?). 零空间  $\mathcal{N}(M_G^T R^{-1} M_G)$  的维数是 1, 它的一组基是  $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T$ . 设  $\mathbf{x}_0$  是方程 (2.4.3) 的特解, 由定理 2.4.2 可知, 方程的解集是  $\{\mathbf{x}_0 + k\mathbf{1} \mid k \in \mathbb{R}\}$ . 这意味着, 各个顶点的电势除相差一个相同常数之外, 解唯一.  $\odot$

类似于图 1.7.2, 我们把例 2.4.7 的模式总结在图 2.4.4 中, 可以看到, 两图的模式也是类似的.

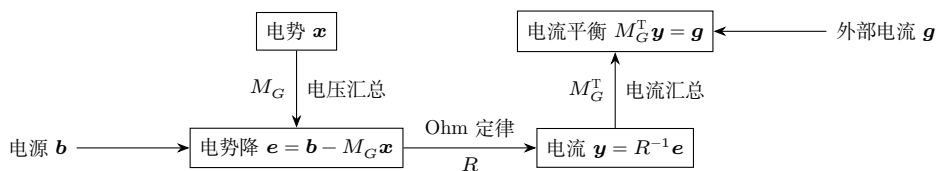


图 2.4.4: 电路网络的基本框架

## 习题

练习 2.4.1 求下列矩阵零空间的一组基:

$$\begin{aligned}
 1. & \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 1 & -3 \\ 0 & 1 & 2 & 2 & 6 \\ 5 & 4 & 3 & 3 & -1 \end{bmatrix}, \quad 2. \begin{bmatrix} 1 & -2 & 1 & 1 & -1 \\ 2 & 1 & -1 & -1 & -1 \\ 1 & 7 & -5 & 5 & 5 \\ 3 & -1 & -2 & 1 & -1 \end{bmatrix}, \quad 3. \begin{bmatrix} 2 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -2 \\ 3 & 3 & -3 & -3 & 4 \\ 4 & 5 & -5 & -5 & 7 \end{bmatrix}.
 \end{aligned}$$

练习 2.4.2 求下列方程组的全部解:

$$\begin{cases} 2x_1 - 2x_2 + x_3 - x_4 + x_5 = 1, \\ x_1 + 2x_2 - x_3 + x_4 - 2x_5 = 1, \\ 4x_1 - 10x_2 + 5x_3 - 5x_4 + 7x_5 = 1. \end{cases}$$

练习 2.4.3 求下列矩阵零空间的一组基.

$$1. \begin{bmatrix} I_n & I_n \end{bmatrix}, \quad 2. \begin{bmatrix} I_n & I_n \\ O & O \end{bmatrix}, \quad 3. \begin{bmatrix} I_n & I_n & I_n \end{bmatrix}.$$

练习 2.4.4 给定  $A = \begin{bmatrix} 1 & & \\ 6 & 1 & \\ 9 & 8 & 1 \end{bmatrix}$  和  $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & \\ & & 1 & 2 \end{bmatrix}$ , 求  $A$  的零空间、列空间、行空间、左零空间的一组基.

练习 2.4.5 线性方程组  $Ax = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$  的全部解是  $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ , 求  $A$ .

练习 2.4.6 求常数  $a, b, c$ , 使得方程  $\begin{bmatrix} 1 & -3 & -1 \end{bmatrix} x = 12$  的所有解都具有如下形式:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} b \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} c \\ 0 \\ 1 \end{bmatrix} x_3.$$

练习 2.4.7 设  $3 \times 4$  矩阵  $A$  的零空间的一组基是  $\begin{bmatrix} 2 \\ 3 \\ 1 \\ 0 \end{bmatrix}$ .

1. 求  $\text{rank}(A)$ .
2. 写出  $A$  的行简化阶梯形.
3. 方程组  $Ax = b$  对哪些  $b$  有解?

**练习 2.4.8** 设  $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ , 对任意  $v \in \mathbb{R}^4$ , 令  $S_v := \{v + x_0 \mid x_0 \in \mathcal{N}(A)\}$ , 即由  $\mathcal{N}(A)$

沿着  $v$  平移后得到的子集. 令  $v_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix}$ ,  $v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix}$ ,  $v_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 6 \end{bmatrix}$ , 求  $b_i$  使得  $S_{v_i}$  是  $Ax = b_i$  的解集,

并分析  $v_1, v_2, v_3$  之间的关系.

**练习 2.4.9** 设  $A, B, C, D$  为二阶方阵, 如果分块矩阵  $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$  满足每一行、每一列以及四个二阶子方阵中的四个元素都是 1, 2, 3, 4, 则称  $M$  为四阶数独矩阵. 写出一个四阶数独矩阵, 并分别求其列空间、行空间、零空间和左零空间的一组基.

**练习 2.4.10** 把国际象棋的棋盘以及棋子的初始位置分别抽象成如下矩阵  $B$  和  $C$ :

$$B = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} r & n & b & q & k & b & n & r \\ p & p & p & p & p & p & p & p \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p & p & p & p & p & p & p & p \\ r & n & b & q & k & b & n & r \end{bmatrix}.$$

分别求其列空间、行空间、零空间和左零空间的一组基.

**练习 2.4.11** 在平面直角坐标系下给定点  $A(a_1, a_2), B(b_1, b_2), C(c_1, c_2)$ , 证明,  $A, B, C$  三点不共线

当且仅当矩阵  $\begin{bmatrix} a_1 & a_2 & 1 \\ b_1 & b_2 & 1 \\ c_1 & c_2 & 1 \end{bmatrix}$  可逆.

**练习 2.4.12** 设  $x_0, x_1, \dots, x_t$  是线性方程组  $Ax = b$  的解, 其中  $b \neq 0$ , 证明,  $c_0x_0 + c_1x_1 + \dots + c_tx_t$  也是解当且仅当  $c_0 + c_1 + \dots + c_t = 1$ .

**练习 2.4.13** 设  $x_0$  是线性方程组  $Ax = b$  的一个解, 其中  $b \neq 0$ , 而  $k_1, \dots, k_t$  是  $\mathcal{N}(A)$  的一组基. 令  $x_i = x_0 + k_i, 1 \leq i \leq t$ , 证明, 线性方程组的任意解都可唯一地表示成  $c_0x_0 + c_1x_1 + \dots + c_tx_t$ , 其中  $c_0 + c_1 + \dots + c_t = 1$ .

**练习 2.4.14** 对任意  $\mathbb{R}^n$  中线性无关的向量组  $x_0, x_1, \dots, x_t$ , 证明, 存在满足如下条件的非齐次线性方程组:

1.  $x_0, x_1, \dots, x_t$  都是此方程组的解;
2. 该方程组的任意解都能被  $x_0, x_1, \dots, x_t$  线性表示.

**练习 2.4.15** 给定线性方程组  $Ax = b$ , 和分块矩阵  $B = \begin{bmatrix} A & b \\ b^T & 0 \end{bmatrix}$ . 证明, 如果  $\text{rank}(A) = \text{rank}(B)$ , 则方程组有解.

**练习 2.4.16 (Fredholm 二择一定理)** 线性方程组  $Ax = b$  有解当且仅当  $\begin{bmatrix} A^T \\ b^T \end{bmatrix} y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$  无解.

**练习 2.4.17** 如果 10 阶方阵  $A$  满足  $A^2 = O$ , 证明  $\text{rank}(A) \leq 5$ . 是否存在  $A^2 = O, \text{rank}(A) = 5$  的 10 阶方阵  $A$ ?

**练习 2.4.18** 设  $A, B$  分别为  $m \times n, n \times k$  矩阵, 证明,  $\text{rank}(AB) \geq \text{rank}(A) + \text{rank}(B) - n$ .

**练习 2.4.19** 对  $n$  阶方阵  $A$ , 求证:

1.  $A^2 = A$  当且仅当  $\text{rank}(A) + \text{rank}(I_n - A) = n$ .
2.  $A^2 = I_n$  当且仅当  $\text{rank}(I_n + A) + \text{rank}(I_n - A) = n$ .

**练习 2.4.20** 证明或否定: 如果对任意  $b$ , 方程组  $A_1x = b$  和  $A_2x = b$  总有相同的解集, 则  $A_1 = A_2$ .

**练习 2.4.21** 证明,  $\mathbb{R}^n$  的任意子空间一定是某个矩阵的零空间.

**练习 2.4.22** 设  $A, B$  分别是  $l \times n, m \times n$  矩阵, 证明,  $\mathcal{N}(A) \subseteq \mathcal{N}(B)$  当且仅当存在  $m \times l$  矩阵  $C$ , 使得  $B = CA$ .

**练习 2.4.23** 设  $A, B$  都是  $m \times n$  矩阵, 证明,  $\mathcal{N}(A) = \mathcal{N}(B)$  当且仅当存在  $m$  阶可逆矩阵  $T$ , 使得  $B = TA$ .

**练习 2.4.24** 设  $A$  是  $n$  阶方阵, 求证: 存在  $k \leq n$ , 满足  $\text{rank}(A^k) = \text{rank}(A^{k+1}) = \dots$ . 由此证明, 如果存在  $p$  使得  $A^p = O$ , 则  $A^n = O$ .

**练习 2.4.25 (Kirchhoff 电流定律)** 对练习 2.1.22 中的电路网络, 令  $M_G$  为对应的关联矩阵. 根据 Kirchhoff 电流定律, 求  $M_G$  左零空间的一组基.

**练习 2.4.26** 在例 2.4.7 中, 有结论  $\text{rank}(M_G^T R^{-1} M_G) = \text{rank } M_G$ , 其中  $R^{-1}$  为对角元素都大于零的对角矩阵. 根据下列思路证明该结论.

1. 若  $y^T R^{-1} y = 0$ , 则  $y = 0$ .
2.  $M_G x = 0$ , 当且仅当  $M_G^T R^{-1} M_G x = 0$ .
3.  $\text{rank}(M_G^T R^{-1} M_G) = \text{rank } M_G$ .

**阅读 2.4.27 (桥墩载荷)** 现在假设桥有三个桥墩, 重力为  $F_1$  的重物和桥墩位置关系如图 2.4.5 所示. 以  $S_1, S_2, S_3$  三处为杠杆, 列出方程

$$\begin{bmatrix} 0 & -d_1 & -d_1 - d_2 \\ d_1 & 0 & -d_2 \\ d_1 + d_2 & d_2 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} -l_1 F_1 \\ (d_1 - l_1) F_1 \\ (d_1 + d_2 - l_1) F_1 \end{bmatrix}$$

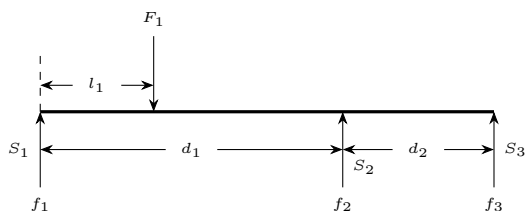


图 2.4.5: 桥墩载荷

可以发现系数矩阵秩是 2，因此方程组有无穷多解。与实际矛盾！添加力平衡条件如何？添加其他支点的杠杆平衡条件如何？全部无用！事实上，不管有多少桥墩，系数矩阵的秩都是 2。问题出在哪里？关键在于我们忽略了桥的形变，以及认为结构横向联系无限强。在实际工程计算中，这种简单实用的杠杆法要假设载荷只由相邻的两个桥墩承担。另一种可行做法是考虑桥的形变以及结构的横向联系对载荷的影响，所需物理知识已大大超出了本书的范围。

## 第三章 内积和正交性

第 1 至 2 章主要介绍了线性映射和线性空间的基本概念，其中重要的是：

1. 方程组  $A\mathbf{x} = \mathbf{b}$  的求解，帮助判断线性映射是否单射或满射；
2. 子空间的基的计算，进一步明确线性映射单射或满射的性质.

本章将处理与之相关的两个问题：

1. 在方程组  $A\mathbf{x} = \mathbf{b}$  无解时，如何找到最佳的逼近解，即找到  $\hat{\mathbf{x}}$ ，使得  $A\hat{\mathbf{x}} - \mathbf{b}$  尽量小？
2. 如果把基类比于坐标系的坐标向量，有没有类似于直角坐标系所具有的两两正交的坐标向量的基？

前者需要刻画向量的大小，后者需要刻画向量的夹角.

### 3.1 基本概念

#### 3.1.1 内积

先看一个平面的例子.

**例 3.1.1** 平面向量有内积、夹角和长度等概念. 其坐标表示需要直角坐标系，即要求  $\mathbb{R}^2$  的标准基  $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ ,  $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$  是互相垂直且每个向量的长度都是 1. 设  $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ ,  $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$  是两个平面向量，则二者的内积为实数  $a_1b_1 + a_2b_2 = \mathbf{a}^T\mathbf{b}$ . 利用内积可以定义向量的长度为  $\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2} = \sqrt{\mathbf{a}^T\mathbf{a}}$ . 而两个向量  $\mathbf{a}, \mathbf{b}$  垂直当且仅当二者内积为零： $a_1b_1 + a_2b_2 = 0$ .

下面考虑方程组  $Ax = b$ , 其中  $A = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ,  $b = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ ,  $x \in \mathbb{R}$ . 方程组显然无解. 退而求其次, 我们希望找到最佳逼近  $\hat{x} \in \mathbb{R}$ , 即满足

$$\|b - A\hat{x}\| = \min_{x \in \mathbb{R}} \|b - Ax\|.$$

注意到, 集合  $\{Ax \mid x \in \mathbb{R}\}$  是一条过原点的直线,  $\|b - Ax\|$  的最小值为向量  $b$  (的终点) 到该直线的距离. 设  $a = A$  为  $A$  的唯一列向量. 为了求此距离, 需要把向量  $b$  向该直线做垂直投影, 即考虑分解  $b = \hat{x}a + r$ , 使得  $a$  与  $r$  垂直, 即内积为零. 于是  $0 = a^T r = a^T (b - \hat{x}a)$ , 因此

$$\hat{x} = \frac{a^T b}{a^T a}, \quad r = b - \frac{a^T b}{a^T a} a.$$

代入可得  $\hat{x} = \frac{8}{5}$ ,  $r = \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \frac{8}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \frac{6}{5} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ , 而距离为  $r$  的长度  $\|r\| = \frac{6}{\sqrt{5}}$ . ☺

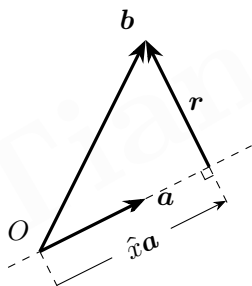


图 3.1.1: 平面向量的逼近

下面考察  $\mathbb{R}^n$  的情形.

**定义 3.1.2 (内积)** 定义  $\mathbb{R}^n$  上的两个列向量  $a, b$  的内积为实数  $a^T b$ , 即如果  $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ ,  $b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ , 则  $a, b$  的内积为  $a_1 b_1 + \cdots + a_n b_n$ .

**命题 3.1.3** 向量内积满足如下性质:

1. 对称性:  $a^T b = b^T a$ ;
2. 双线性性:  $a^T (k_1 b_1 + k_2 b_2) = k_1 a^T b_1 + k_2 a^T b_2$ ,  $(k_1 a_1 + k_2 a_2)^T b = k_1 a_1^T b + k_2 a_2^T b$ ;
3. 正定性:  $a^T a \geq 0$ , 且  $a^T a = 0$  当且仅当  $a = 0$ .



证明留给读者. 内积的正定性使我们能定义  $n$  维向量  $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$  的长度:

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}^T \mathbf{a}} = \sqrt{a_1^2 + \cdots + a_n^2}.$$

一个向量的长度为零当且仅当它是零向量. 长度满足  $\|k\mathbf{a}\| = |k|\|\mathbf{a}\|$ . 长度为 1 的向量称为单位向量. 单位向量  $\frac{\mathbf{a}}{\|\mathbf{a}\|}$  称为非零向量  $\mathbf{a}$  的单位化 (向量). 实数  $\|\mathbf{a} - \mathbf{b}\|$  称为向量  $\mathbf{a}, \mathbf{b}$  间的距离.

**定理 3.1.4 (Cauchy-Schwarz 不等式)**  $|\mathbf{a}^T \mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|$ , 等号成立当且仅当  $\mathbf{a}, \mathbf{b}$  共线.

证. 设  $\mathbf{a} = [a_1 \ \cdots \ a_n]^T, \mathbf{b} = [b_1 \ \cdots \ b_n]^T$ , 则

$$\|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - |\mathbf{a}^T \mathbf{b}|^2 = \sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2 - \left( \sum_{i=1}^n a_i b_i \right)^2 = \sum_{1 \leq i < j \leq n} (a_i b_j - a_j b_i)^2 \geq 0.$$

由此知不等式成立. 等号成立当且仅当对任意  $i < j$ , 都有  $a_i b_j - a_j b_i = 0$ . 这意味着, 若  $b_i \neq 0, b_j \neq 0$ , 则  $a_i$  与  $b_i$  之比等于  $a_j$  与  $b_j$  之比, 即  $\mathbf{a}, \mathbf{b}$  共线; 而这对  $b_i = 0$  或  $b_j = 0$  也成立.  $\square$

由此容易得到如下推论, 证明留给读者.

**推论 3.1.5 (三角不等式)**  $\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$ , 等号成立当且仅当  $\mathbf{a}, \mathbf{b}$  共线.

在  $\mathbb{R}^n$  中的三角形的三条边能写成向量  $\mathbf{a}, \mathbf{b}, \mathbf{a} + \mathbf{b}$ , 而三角不等式说明两边 (长度) 之和大于第三边 (长度).

Cauchy-Schwarz 不等式使我们能定义两个非零向量的夹角为  $\arccos \frac{\mathbf{a}^T \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$ . 两个非零向量的内积为 0 当且仅当夹角为  $\frac{\pi}{2}$ . 如果  $\mathbf{a}^T \mathbf{b} = 0$ , 称二者正交或垂直, 记为  $\mathbf{a} \perp \mathbf{b}$ . 注意, 零向量与任意向量都正交.

**定理 3.1.6 (勾股定理)** 向量  $\mathbf{a}, \mathbf{b}$  正交, 则  $\|\mathbf{a} \pm \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$ .

证. 根据定义,  $\|\mathbf{a} \pm \mathbf{b}\|^2 = (\mathbf{a} \pm \mathbf{b})^T (\mathbf{a} \pm \mathbf{b}) = \mathbf{a}^T \mathbf{a} \pm 2\mathbf{a}^T \mathbf{b} + \mathbf{b}^T \mathbf{b} = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$ .  $\square$

设  $\mathbf{a}, \mathbf{b}$  是  $\mathbb{R}^n$  中的两个向量,  $\mathbf{a} \neq \mathbf{0}$ , 考虑分解式

$$\mathbf{b} = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} \mathbf{a} + \mathbf{r}. \quad (3.1.1)$$

由例 3.1.1 中的计算可知  $\mathbf{a} \perp \mathbf{r}$ , 其中向量  $\frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} \mathbf{a}$  称为向量  $\mathbf{b}$  向直线  $\text{span}(\mathbf{a})$  的正交投影.

**命题 3.1.7** 设  $\mathbf{a}, \mathbf{b}$  是  $\mathbb{R}^n$  中的两个向量,  $\mathbf{a} \neq \mathbf{0}$ , 则

$$\left\| \mathbf{b} - \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} \mathbf{a} \right\| = \min_{x \in \mathbb{R}} \|\mathbf{b} - x\mathbf{a}\|.$$

证一. 设  $\hat{x} = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}$ , 则向量  $(\hat{x} - x)\mathbf{a}, \mathbf{b} - \hat{x}\mathbf{a}, \mathbf{b} - x\mathbf{a}$  组成一个直角三角形, 前两个向量是直角边. 根据勾股定理,  $\|\mathbf{b} - x\mathbf{a}\| \geq \|\mathbf{b} - \hat{x}\mathbf{a}\|$ .  $\square$

证二. 利用一元函数的性质. 考虑

$$\begin{aligned} \|\mathbf{b} - x\mathbf{a}\|^2 &= (\mathbf{b} - x\mathbf{a})^T (\mathbf{b} - x\mathbf{a}) = (\mathbf{a}^T \mathbf{a})x^2 - 2(\mathbf{a}^T \mathbf{b})x + \mathbf{b}^T \mathbf{b} \\ &= (\mathbf{a}^T \mathbf{a})\left(x - \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}\right)^2 + \mathbf{b}^T \mathbf{b} - \left(\frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}\right)^2, \end{aligned}$$

其中  $x = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}}$  时达到最小值.  $\square$

可以看到, 长度、角度、正交等概念, 都能从内积导出, 而这将在第 8 章得到进一步阐述.

### 3.1.2 标准正交基

有了内积和角度的概念, 我们就可以考察子空间的基中向量的夹角. 标准基中的向量, 即  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , 两两正交, 而一般的基并非如此.

**定义 3.1.8 (正交向量组)** 设  $\mathbf{a}_1, \dots, \mathbf{a}_r$  是  $\mathbb{R}^n$  中的向量组, 如果这些向量都非零且两两正交, 则称该向量组为**正交向量组**. 特别地, 如果正交向量组中的向量都是单位向量, 则称其为**正交单位向量组**.

**命题 3.1.9** 正交向量组线性无关.

证. 设向量组为  $\mathbf{a}_1, \dots, \mathbf{a}_r$ . 考虑方程  $k_1 \mathbf{a}_1 + \dots + k_r \mathbf{a}_r = \mathbf{0}$ . 两边和  $\mathbf{a}_i$  做内积, 有  $0 = k_1 \mathbf{a}_i^T \mathbf{a}_1 + \dots + k_i \mathbf{a}_i^T \mathbf{a}_i + \dots + k_r \mathbf{a}_i^T \mathbf{a}_r = k_i \|\mathbf{a}_i\|^2$ . 向量  $\mathbf{a}_i$  非零, 因此  $k_i = 0$ , 对任意  $i$  成立. 于是  $\mathbf{a}_1, \dots, \mathbf{a}_r$  线性无关.  $\square$

**定义 3.1.10 (标准正交基)** 设  $\mathcal{M}$  是  $\mathbb{R}^n$  的子空间, 如果它的一组基是正交向量组, 则称之为  $\mathcal{M}$  的一组**正交基**; 如果它的一组基是正交单位向量组, 则称之为  $\mathcal{M}$  的一组**标准正交基**.

设  $\mathcal{M}$  的维数是  $r$ ,  $S$  是  $\mathcal{M}$  中包含  $r$  个向量的正交向量组, 根据命题 3.1.9,  $S$  线性无关, 因而是  $\mathcal{M}$  的一组正交基.

**例 3.1.11** 标准基  $e_1, \dots, e_n$  是  $\mathbb{R}^n$  的一组标准正交基.

而  $\frac{1}{\sqrt{2}}e_1 + \frac{1}{\sqrt{2}}e_2, \frac{1}{\sqrt{2}}e_1 - \frac{1}{\sqrt{2}}e_2, e_3, \dots, e_n$  也是  $\mathbb{R}^n$  的一组标准正交基.

设  $q_1, \dots, q_n$  是  $\mathbb{R}^n$  的一组标准正交基, 则对任意向量  $a$ , 都有分解式  $a = (q_1^T a)q_1 + \dots + (q_n^T a)q_n$ .  $\odot$

**命题 3.1.12** 设  $\mathcal{M}$  是  $\mathbb{R}^n$  的子空间, 则  $\mathcal{M}$  存在一组正交基, 从而存在一组标准正交基.

证. 对  $\mathcal{M}$  的维数应用数学归纳法. 维数是 1 的子空间中的任意非零向量都构成一组正交基. 假设任意  $r$  维的子空间都存在一组正交基, 下面证明  $r+1$  维的子空间  $\mathcal{M}$  也存在一组正交基. 任取  $\mathcal{M}$  中一组基  $a_1, \dots, a_r, a_{r+1}$ , 其包含的子空间  $\mathcal{N} = \text{span}(a_1, \dots, a_r)$  是  $r$  维的. 根据归纳假设,  $\mathcal{N}$  存在一组正交基  $q_1, \dots, q_r$ . 我们希望找到一个非零向量  $q_{r+1}$ , 它与  $q_1, \dots, q_r$  都正交. 这样  $q_1, \dots, q_r, q_{r+1}$  就是一个  $r+1$  个向量组成的正交向量组, 从而是  $\mathcal{M}$  的一组正交基.

显然  $q_1, \dots, q_r, a_{r+1}$  是  $\mathcal{M}$  的一组基. 设  $q_{n+1} = k_1 q_1 + \dots + k_r q_r + a_{r+1}$ . 假设  $q_{r+1}$  与  $q_1, \dots, q_n$  都正交. 对  $i = 1, \dots, n$ , 两边和  $q_i$  做内积可得  $0 = k_i \|q_i\|^2 + q_i^T a_{r+1}$ , 因此  $k_i = -\frac{q_i^T a_{r+1}}{\|q_i\|^2}$ . 于是  $q_{n+1} = a_{r+1} - \frac{q_1^T a_{r+1}}{\|q_1\|^2} q_1 - \dots - \frac{q_r^T a_{r+1}}{\|q_r\|^2} q_r \neq 0$ . 容易验证它与  $q_1, \dots, q_n$  都正交.

综上即得, 任意子空间都存在一组正交基. 至于标准正交基, 只需将正交基的每个向量单位化即得.  $\square$

基扩张定理也可以推广到标准正交基的情形, 证明留给读者.

**命题 3.1.13** 设  $\mathcal{M}, \mathcal{N}$  是  $\mathbb{R}^n$  的两个子空间, 如果  $\mathcal{M} \subseteq \mathcal{N}$ , 则  $\mathcal{M}$  的任意一组标准正交基都可以扩充成  $\mathcal{N}$  的一组标准正交基.

从命题 3.1.12 的证明中, 可以看到, 从  $\mathcal{M}$  的任意一组基  $a_1, \dots, a_r$  出发, 通过递归能够得到一组正交基. 这种由一组基得到一组正交基的方法称为 **Gram-Schmidt 正交化**. 具体操作如下:

$$\begin{aligned}\tilde{q}_1 &= a_1, \\ \tilde{q}_2 &= a_2 - \frac{\tilde{q}_1^T a_2}{\tilde{q}_1^T \tilde{q}_1} \tilde{q}_1, \\ \tilde{q}_3 &= a_3 - \frac{\tilde{q}_1^T a_3}{\tilde{q}_1^T \tilde{q}_1} \tilde{q}_1 - \frac{\tilde{q}_2^T a_3}{\tilde{q}_2^T \tilde{q}_2} \tilde{q}_2, \\ &\vdots \\ \tilde{q}_r &= a_r - \frac{\tilde{q}_1^T a_r}{\tilde{q}_1^T \tilde{q}_1} \tilde{q}_1 - \dots - \frac{\tilde{q}_{r-1}^T a_r}{\tilde{q}_{r-1}^T \tilde{q}_{r-1}} \tilde{q}_{r-1}.\end{aligned}$$

为了得到标准正交基, 只要再把正交基中的每个向量都单位化即可:  $\mathbf{q}_i = \frac{\tilde{\mathbf{q}}_i}{\|\tilde{\mathbf{q}}_i\|}$ .

这里, 每个等式的右端项是  $\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_{i-1}, \mathbf{a}_i$  这个线性无关向量组的线性组合, 其中前  $i-1$  个向量是正交化过程中得到的新向量, 而最后一个是原有向量. 事实上,  $\tilde{\mathbf{q}}_i$  就是  $\mathbf{a}_i$  中与  $\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_{i-1}$  相垂直的部分.

在 Gram-Schmidt 正交化过程中, 每一步涉及三个向量组

$$\tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_i; \quad \tilde{\mathbf{q}}_1, \dots, \tilde{\mathbf{q}}_{i-1}, \mathbf{a}_i; \quad \mathbf{a}_1, \dots, \mathbf{a}_i$$

线性等价, 都是同一子空间的基. 这是 Gram-Schmidt 正交化的重要性质.

**例 3.1.14** 给定  $\mathbb{R}^3$  中的一组基:

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

利用 Gram-Schmidt 正交化方法计算  $\mathbb{R}^3$  的一组标准正交基. 过程如下:

$$\tilde{\mathbf{q}}_1 = \mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\tilde{\mathbf{q}}_2 = \mathbf{a}_2 - \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_2}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} \tilde{\mathbf{q}}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix},$$

$$\tilde{\mathbf{q}}_3 = \mathbf{a}_3 - \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_3}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} \tilde{\mathbf{q}}_1 - \frac{\tilde{\mathbf{q}}_2^T \mathbf{a}_3}{\tilde{\mathbf{q}}_2^T \tilde{\mathbf{q}}_2} \tilde{\mathbf{q}}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{\frac{1}{3}}{\frac{1}{3}} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 0 \end{bmatrix}.$$

把向量单位化得到标准正交基:  $\mathbf{q}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \mathbf{q}_2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \end{bmatrix}, \mathbf{q}_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}.$

注意,  $\text{span}(\mathbf{a}_1, \mathbf{a}_2) = \text{span}(\mathbf{q}_1, \mathbf{q}_2)$ , 而  $\mathbf{q}_3$  与这个平面正交, 因此  $\mathbf{q}_3$  是该平面的单位法向量.

如果对基  $\mathbf{a}_3, \mathbf{a}_2, \mathbf{a}_1$  应用 Gram-Schmidt 正交化, 得到的标准正交基就是标准基  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ . ☺

注意, 需要说明的是, 实际问题经常在建立模型阶段就会出现正交基, 于是就不再需要利用 Gram-Schmidt 正交化得到正交基. 因此, 在处理实际问题时, 如果能找到更好的模型, 就可以避免无谓计算.

## 习题

练习 3.1.1 证明命题 3.1.3.

练习 3.1.2 在  $\mathbb{R}^4$  中求向量  $\mathbf{a}, \mathbf{b}$  的夹角.

$$1. \mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 1 \end{bmatrix}. \quad 2. \mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ 5 \\ 1 \end{bmatrix}. \quad 3. \mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 0 \end{bmatrix}.$$

练习 3.1.3 求证:

1. 在  $\mathbb{R}^n$  中的非零向量  $\mathbf{a}, \mathbf{b}$  夹角为 0, 当且仅当存在  $k > 0$ , 使得  $\mathbf{a} = k\mathbf{b}$ .
2. 在  $\mathbb{R}^n$  中的两向量  $\mathbf{a}, \mathbf{b}$  正交, 当且仅当对任意实数  $t$ , 有  $\|\mathbf{a} + t\mathbf{b}\| \geq \|\mathbf{a}\|$ .
3. 在  $\mathbb{R}^n$  中的非零向量  $\mathbf{a}, \mathbf{b}$  正交, 当且仅当  $\|\mathbf{a} + \mathbf{b}\| = \|\mathbf{a} - \mathbf{b}\|$ .

练习 3.1.4 证明推论 3.1.5.

练习 3.1.5 (Cauchy-Schwarz 不等式的其他证明) 1. 先证明  $\mathbf{a}, \mathbf{b}$  都是单位向量的情形:  $|\mathbf{a}^T \mathbf{b}| \leq 1$ , 且等号成立当且仅当  $\mathbf{a} = \pm \mathbf{b}$ . 再由单位向量的情形推广到一般的情形. 提示: 利用均值不等式  $\sqrt{xy} \leq \frac{x+y}{2}$ .

2. 根据内积的正定性, 对任意实数  $t$ , 都有  $(\mathbf{a} + t\mathbf{b})^T (\mathbf{a} + t\mathbf{b}) = \mathbf{a}^T \mathbf{a} + 2t\mathbf{a}^T \mathbf{b} + t^2 \mathbf{b}^T \mathbf{b} \geq 0$ . 利用判别式证明结论.

练习 3.1.6 给定  $\mathbf{a} \in \mathbb{R}^3$ . 计算  $\mathbf{a}$  与坐标向量  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  的夹角的余弦, 并计算这三个余弦值的平方和.

练习 3.1.7 设  $\|\mathbf{a}\| = 3, \|\mathbf{b}\| = 4$ , 确定  $\|\mathbf{a} - \mathbf{b}\|$  的取值范围.

练习 3.1.8 设  $\mathbf{a} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \mathbf{b} = \begin{bmatrix} z \\ x \\ y \end{bmatrix}$ , 且  $x + y + z = 0$ . 确定  $\mathbf{a}, \mathbf{b}$  夹角的取值范围.

练习 3.1.9 1. 找到  $\mathbb{R}^4$  中的四个两两正交的向量, 且每个向量的每个分量只能是  $\pm 1$ .

2.  $\mathbb{R}^n$  中最多有多少个两两正交的向量?

练习 3.1.10 1. 找到  $\mathbb{R}^2$  中的三个向量, 使它们之间两两内积为负.

2. 找到  $\mathbb{R}^3$  中的四个向量, 使它们之间两两内积为负.

3.  $\mathbb{R}^n$  中最多有多少个向量, 使它们之间两两内积为负?

练习 3.1.11 在  $\mathbb{R}^4$  中求一单位向量与下列向量正交:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 3 \end{bmatrix}.$$

**练习 3.1.12** 设  $\mathbf{a}_1, \dots, \mathbf{a}_n$  是  $\mathbb{R}^n$  的一组标准正交基, 证明,

1. 如果  $\mathbf{b}^T \mathbf{a}_i = 0$  ( $1 \leq i \leq n$ ), 则  $\mathbf{b} = \mathbf{0}$ .
2. 如果  $\mathbf{b}_1^T \mathbf{a}_i = \mathbf{b}_2^T \mathbf{a}_i$  ( $1 \leq i \leq n$ ), 则  $\mathbf{b}_1 = \mathbf{b}_2$ .

**练习 3.1.13** 设  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  是  $\mathbb{R}^3$  的一组标准正交基, 证明下列向量组也是一组标准正交基:

$$\mathbf{b}_1 = \frac{1}{3}(2\mathbf{a}_1 + 2\mathbf{a}_2 - \mathbf{a}_3), \quad \mathbf{b}_2 = \frac{1}{3}(2\mathbf{a}_1 - \mathbf{a}_2 + 2\mathbf{a}_3), \quad \mathbf{b}_3 = \frac{1}{3}(\mathbf{a}_1 - 2\mathbf{a}_2 - 2\mathbf{a}_3).$$

**练习 3.1.14** 设  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5$  是  $\mathbb{R}^5$  的一组标准正交基, 令

$$\mathbf{b}_1 = \mathbf{a}_1 + \mathbf{a}_5, \quad \mathbf{b}_2 = \mathbf{a}_1 - \mathbf{a}_2 + \mathbf{a}_4, \quad \mathbf{b}_3 = 2\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3.$$

求  $\text{span}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$  的一组标准正交基.

**练习 3.1.15** 求齐次线性方程组

$$\begin{cases} 2x_1 + x_2 - x_3 + x_4 - 3x_5 = 0, \\ x_1 + x_2 - x_3 \quad \quad + x_5 = 0, \end{cases}$$

解空间的一组标准正交基.

**练习 3.1.16** 利用 Gram-Schmidt 正交化方法求由下列向量线性生成的子空间的标准正交基:

$$1. \mathbf{a}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \\ 2 \end{bmatrix}. \quad 2. \mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 1 \end{bmatrix}.$$

**练习 3.1.17** 证明命题 3.1.13.

**练习 3.1.18 (勾股定理的高维推广)** 1. 向量  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  围成一个三角形, 证明其面积的平方为  $\frac{1}{2}(\|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - (\mathbf{a}^T \mathbf{b})^2)$ .

2. 两两垂直的向量  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$  围成一个四面体, 证明其斜面上三角形面积的平方等于其余三个直角三角形面积的平方和.

**练习 3.1.19** 取定非零向量  $\mathbf{a} \in \mathbb{R}^n$ , 考虑  $\mathbb{R}^n$  上的一个变换, 它将每个向量  $\mathbf{b}$  映射到其向直线  $\text{span}(\mathbf{a})$  正交投影后平行于  $\mathbf{a}$  的部分.

1. 证明这是一个线性变换, 其表示矩阵为  $A = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}}$ .
2. 证明  $A^2 = A, A^T = A$ .

**练习 3.1.20 (内积决定转置)** 求证: 设  $m \times n$  矩阵  $A$  和  $n \times m$  矩阵  $B$ , 如果对任意  $\mathbf{v} \in \mathbb{R}^m, \mathbf{w} \in \mathbb{R}^n$ , 都有  $(B\mathbf{v})^T \mathbf{w} = \mathbf{v}^T (A\mathbf{w})$ , 则  $B = A^T$ .

**练习 3.1.21 (Riesz 表示定理)** 设  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  是线性映射, 证明, 存在向量  $\mathbf{b}$ , 使得对任意  $\mathbf{a} \in \mathbb{R}^n$ , 都有  $f(\mathbf{a}) = \mathbf{b}^T \mathbf{a}$ .

**练习 3.1.22** 1. (平行四边形法则) 证明  $\mathbb{R}^n$  中任意平行四边形的两条对角线长度的平方和, 等于其四条边长的平方和.

2. (极化公式) 证明  $\mathbf{v}^T \mathbf{w} = \frac{1}{4}(\|\mathbf{v} + \mathbf{w}\|^2 - \|\mathbf{v} - \mathbf{w}\|^2)$ . 提示: 这意味着, 长度决定角度.
3. 设  $\|\mathbf{a}\|_4 = (\sum_{i=1}^n a_i^4)^{\frac{1}{4}}$ , 定义关于  $\mathbf{v}, \mathbf{w}$  的二元函数  $\frac{1}{4}(\|\mathbf{v} + \mathbf{w}\|_4^2 - \|\mathbf{v} - \mathbf{w}\|_4^2)$ . 这个二元函数是否满足内积定义中的对称性、双线性、正定性?
4. 定义  $\|\mathbf{a}\|_\infty$  为所有分量中绝对值的最大值, 定义关于  $\mathbf{v}, \mathbf{w}$  的二元函数  $\frac{1}{4}(\|\mathbf{v} + \mathbf{w}\|_\infty^2 - \|\mathbf{v} - \mathbf{w}\|_\infty^2)$ . 这个二元函数是否满足内积定义中的对称性、双线性、正定性? 此时, 三角不等式和 Cauchy-Schwarz 不等式是否仍然成立?

**阅读 3.1.23** 计算机在存储图像时, 往往是在照片上打上均匀的正方形网格, 并只存储网格点的信息, 这些网格点称为像素. 为简单起见, 这里只考虑灰度图像. 灰度图像用 0 表示黑色, 1 表示白色, 二者之间的数字表示不同深浅的灰色. 那么所有像素的灰度值按顺序排列就可以得到一个向量, 换言之, 我们就可以用向量来表示具有相同像素数的图像. 图像对应的向量可以生成一个子空间, 它对应的标准基, 就是在某个像素是黑色, 其他像素是白色的图像.

现在我们考虑对图像进行压缩, 即适当减少图像占用的存储. 比较简单的处理办法是减少存储一些像素, 为了保证图像不失真太多, 我们可以选择存储间隔的像素, 例如原本  $256 \times 256$  像素的图像, 只存储间隔的像素就得到了  $128 \times 128$  像素的图像, 所占存储变为原来的四分之一. 这样压缩难以处理图像细节和存储压缩比之间的平衡.

现在考虑另一种办法. 先考虑一个只有四个像素的灰度图像, 四个像素对应的灰度为  $a_1, a_2, a_3, a_4$ ,

组成向量  $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$ . 我们把像素的灰度值看成在其平均值上下的波动, 则图像可以分解为两个图像,

一个是均匀图像  $\mathbf{a}_1 = b_1 \mathbf{q}_1$ , 其中  $b_1 = \frac{a_1 + a_2 + a_3 + a_4}{4}$ ,  $\mathbf{q}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ , 一个是代表波动的图像  $\mathbf{a} - \mathbf{a}_1 =$

$\frac{1}{4} \begin{bmatrix} 3a_1 - a_2 - a_3 - a_4 \\ -a_1 + 3a_2 - a_3 - a_4 \\ -a_1 - a_2 + 3a_3 - a_4 \\ -a_1 - a_2 - a_3 + 3a_4 \end{bmatrix}$ . 代表波动的图像  $\mathbf{a} - \mathbf{a}_1$  又可以类似分解, 一个代表上下两部分分别的平

均值  $\mathbf{a}_2 = \frac{a_1 + a_2 - a_3 - a_4}{4} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \frac{-a_1 - a_2 + a_3 + a_4}{4} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = b_2 \mathbf{q}_2$ , 其中  $b_2 = \frac{a_1 + a_2 - a_3 - a_4}{4}$ ,  $\mathbf{q}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$ , 一

个代表上下两部分分别的波动  $\mathbf{a} - \mathbf{a}_1 - \mathbf{a}_2 = \frac{1}{2} \begin{bmatrix} a_1 - a_2 \\ -a_1 + a_2 \\ a_3 - a_4 \\ -a_3 + a_4 \end{bmatrix}$ . 再对  $\mathbf{a} - \mathbf{a}_1 - \mathbf{a}_2$  类似分解, 一个代

表上下两部分左右两侧分别的平均值（每四分之一图片的平均值），记为  $\mathbf{a}_3 = \frac{a_1 - a_2 + a_3 - a_4}{4} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} +$

$\frac{-a_1 + a_2 - a_3 + a_4}{4} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = b_3 \mathbf{q}_3$ , 其中  $b_3 = \frac{a_1 - a_2 + a_3 - a_4}{4}$ ,  $\mathbf{q}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$ , 另一个代表每四分之一图片的波动

$\mathbf{a}_4 = \mathbf{a} - \mathbf{a}_1 - \mathbf{a}_2 - \mathbf{a}_3 = b_4 \mathbf{q}_4$ , 其中  $b_4 = \frac{a_1 - a_2 - a_3 + a_4}{4}$ ,  $\mathbf{q}_4 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$ . 这样图像  $\mathbf{a}$  就变成了四个图像

的叠加  $\mathbf{a} = b_1 \mathbf{q}_1 + b_2 \mathbf{q}_2 + b_3 \mathbf{q}_3 + b_4 \mathbf{q}_4$ . 可以看到,  $b_1$  是图像  $\mathbf{a}$  整体的平均值,  $b_2$  代表上下二分之一图像的平均值,  $b_3$  代表左右二分之一图像的平均值,  $b_4$  代表了四分之一图像的差异细节. 而整个过程从数学上看, 就是我们用  $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4$  这组基来线性表示  $\mathbf{a}$ . 容易验证,  $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4$  是  $\mathbb{R}^4$  的一组正交基, 而  $\frac{1}{2}\mathbf{q}_1, \frac{1}{2}\mathbf{q}_2, \frac{1}{2}\mathbf{q}_3, \frac{1}{2}\mathbf{q}_4$  是  $\mathbb{R}^4$  的一组标准正交基.

现在考虑有大量像素的图像, 不妨设有  $256 \times 256$  像素. 把图像分割成  $2 \times 2$  像素的小块进行上述计算, 并把对应的  $b_1, b_2, b_3, b_4$  的四组  $128 \times 128$  个值集中在一起, 则这四组值分别代表四个错位的四分之一图像的平均值, 上下错位图像的差异细节, 左右错位图像的差异细节, 四个错位四分之一图像的差异细节. 考虑对图像压缩时, 我们可以舍去后面三个代表差异细节的向量, 得到一个  $128 \times 128$  像素的图像, 所占存储变为原来的四分之一. 如果还想继续压缩, 则对新图如法炮制, 可得四个十六分之一图像的平均值, 如图 3.1.2 所示. 不断进行下去, 最后可得全图像的平均值. 当然这时压缩就没有意义了. 另外, 还可以把不同层次的细节合并在一起, 得到一个在细节与压缩比之间取得一定平衡的图像.

这组基称为二维 Haar 小波基, 在图像处理中很常用.

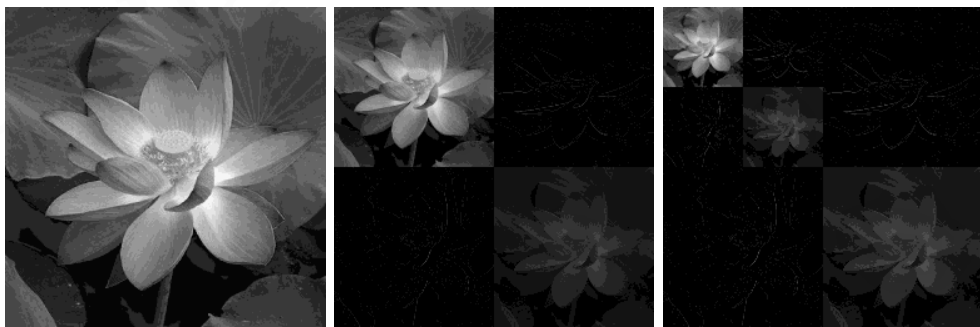


图 3.1.2: 荷花



## 3.2 正交矩阵和 QR 分解

### 3.2.1 正交矩阵

我们知道, 如果  $\mathbf{a}_1, \dots, \mathbf{a}_n$  是  $\mathbb{R}^n$  的一组基, 则  $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$  是可逆矩阵. 新结构带来新性质, 而标准正交基并列排成的矩阵会有何种新性质?

设  $\mathbf{q}_1, \dots, \mathbf{q}_n$  是  $\mathbb{R}^n$  的一组标准正交基, 记  $n$  阶方阵  $Q = [\mathbf{q}_1 \ \dots \ \mathbf{q}_n]$ , 则  $Q^T Q =$

$$\begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} [\mathbf{q}_1 \ \dots \ \mathbf{q}_n] = \begin{bmatrix} \mathbf{q}_1^T \mathbf{q}_1 & \dots & \mathbf{q}_1^T \mathbf{q}_n \\ \vdots & & \vdots \\ \mathbf{q}_n^T \mathbf{q}_1 & \dots & \mathbf{q}_n^T \mathbf{q}_n \end{bmatrix} = I_n.$$

**定义 3.2.1 (正交矩阵)** 一个  $n$  阶方阵  $Q$  如果满足  $Q^T Q = I_n$ , 则称  $Q$  是  $n$  阶正交矩阵.

单位矩阵  $I_n$  是正交矩阵.

下面介绍正交矩阵的性质.

**命题 3.2.2** 对  $n$  阶方阵  $Q$ , 以下叙述等价:

1.  $Q$  是正交矩阵, 即  $Q^T Q = I_n$ ;
2.  $Q$  可逆, 且  $Q^{-1} = Q^T$ ;
3.  $Q Q^T = I_n$ ;
4.  $Q^T$  是正交矩阵;
5.  $Q$  可逆, 且  $Q^{-1}$  是正交矩阵;
6.  $Q$  的列向量组成  $\mathbb{R}^n$  的一组标准正交基;
7.  $Q$  的行向量的转置组成  $\mathbb{R}^n$  的一组标准正交基.

证. 第 1–5 条的等价性是显然的. 第 6 条用矩阵乘法写出即为第 1 条, 第 7 条用矩阵乘法写出来即为第 3 条.  $\square$

例 3.1.14 中的标准正交基并列排成一个 3 阶正交矩阵:

$$Q = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \end{bmatrix}. \quad (3.2.1)$$

因此, 它的行向量的转置也组成  $\mathbb{R}^3$  的一组标准正交基.

**命题 3.2.3** 两个  $n$  阶正交矩阵的乘积还是  $n$  阶正交矩阵.

**命题 3.2.4** 对  $n$  阶方阵  $Q$ , 以下叙述等价:

1.  $Q$  是正交矩阵, 即  $Q^T Q = I_n$ ;
2.  $Q$  为保距变换, 即, 对任意  $\mathbf{x} \in \mathbb{R}^n$ ,  $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ ;
3.  $Q$  为保内积变换, 即, 对任意  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $Q\mathbf{x}$  与  $Q\mathbf{y}$  的内积等于  $\mathbf{x}$  与  $\mathbf{y}$  的内积.

证.  $1 \Rightarrow 2$ : 对任意  $\mathbf{x}$ ,  $\|Q\mathbf{x}\|^2 = (Q\mathbf{x})^T Q\mathbf{x} = \mathbf{x}^T Q^T Q\mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2$ .

$2 \Rightarrow 3$ : 容易验证对任意  $\mathbf{x}, \mathbf{y}$ ,  $\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 = 2\mathbf{x}^T \mathbf{y}$ . 又由于  $Q$  保距, 我们有  $2(Q\mathbf{x})^T (Q\mathbf{y}) = \|Q(\mathbf{x} + \mathbf{y})\|^2 - \|Q\mathbf{x}\|^2 - \|Q\mathbf{y}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 = 2\mathbf{x}^T \mathbf{y}$ . (注意, 这说明保距变换一定保角.)

$3 \Rightarrow 1$ : 对任意矩阵  $A$ ,  $\mathbf{e}_i^T A \mathbf{e}_j$  是  $A$  的  $(i, j)$  元. 由  $Q$  保内积可知,  $\mathbf{e}_i^T (Q^T Q) \mathbf{e}_j = (Q\mathbf{e}_i)^T (Q\mathbf{e}_j) = \mathbf{e}_i^T \mathbf{e}_j = \mathbf{e}_i^T I_n \mathbf{e}_j$ . 因此  $Q^T Q = I_n$ .  $\square$

**例 3.2.5** 给定 2 阶正交矩阵  $Q = [\mathbf{q}_1 \quad \mathbf{q}_2]$ , 则  $\mathbf{q}_1, \mathbf{q}_2$  是  $\mathbb{R}^2$  的一组标准正交基. 首先,  $\mathbf{q}_1, \mathbf{q}_2$  是单位向量, 不妨设  $\mathbf{q}_1 = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$ ,  $\mathbf{q}_2 = \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix}$ . 其次,  $\mathbf{q}_2 \perp \mathbf{q}_1$ , 因此  $0 = \mathbf{q}_1^T \mathbf{q}_2 = \cos \theta \cos \varphi + \sin \theta \sin \varphi = \cos(\theta - \varphi)$ , 于是  $\theta - \varphi = (k + \frac{1}{2})\pi$ . 因此, 任意二阶正交矩阵必然具有如下两种形式之一:

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}. \quad (3.2.2)$$

回顾例 1.3.9, 前者是  $\mathbb{R}^2$  上旋转变换的表示矩阵, 后者是  $\mathbb{R}^2$  上反射变换的表示矩阵. 这说明  $\mathbb{R}^2$  上的保距变换只有两种, 即旋转和反射.

容易验证, 两个反射变换的复合是一个旋转变换, 转角等于两反射轴夹角的二倍; 反之, 任意旋转变换都可以写成两个反射变换的复合. 因此, 任意二阶正交矩阵都能写成至多两个反射矩阵的乘积.  $\odot$

二维的反射和旋转都可以推广到高维.

对旋转, 这里只介绍一个简单情形. 在  $\mathbf{e}_i - \mathbf{e}_j$  平面上转角  $\theta$  的旋转变换的矩阵为

$$G_{\theta; i, j} = \begin{bmatrix} \ddots & & & & & \\ & 1 & & & & \\ & & \cos \theta & & & -\sin \theta \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \\ & \sin \theta & & & & \cos \theta \\ & & & & & & 1 \end{bmatrix}, \text{ 这类旋转常称为 Givens 变换.}$$

下面考虑反射. 例 1.8.8 中反射变换的表示矩阵是  $I_2 - 2\mathbf{v}\mathbf{v}^T$ , 其中  $\mathbf{v}$  是反射轴的单位法向量. 任意  $\mathbb{R}^n$  中的单位向量  $\mathbf{v}$ , 都唯一决定以其为法向量的超平面  $\mathcal{N}(\mathbf{v}^T)$ . 令

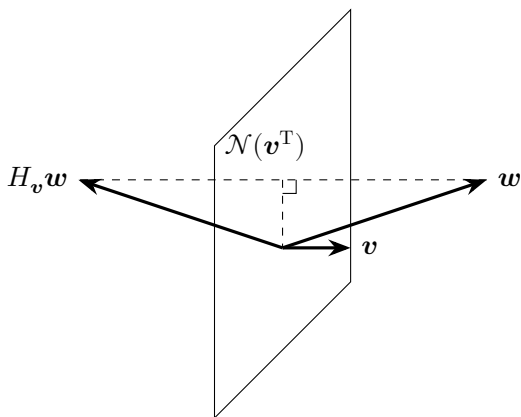


图 3.2.1: 高维空间中的反射

$H_v := I - 2vv^T$ , 则  $H_v$  是  $\mathbb{R}^n$  上的反射变换, 反射面是以  $v$  为法向量的超平面  $\mathcal{N}(v^T)$ , 如图 3.2.1 所示. 事实上, 对任意  $w$ ,

$$\begin{aligned} w &= (I_n - vv^T)w + vv^T w, \\ (I_n - 2vv^T)w &= (I_n - vv^T)w - vv^T w. \end{aligned}$$

可以看到,  $vv^T w = (v^T w)v$  是  $w$  向  $\text{span}(v)$  的投影, 而  $(I_n - vv^T)w \perp v$ , 即  $w$  可以分解为与  $v$  共线的向量和与  $v$  正交的向量之和, 而它在  $I_n - 2vv^T$  下的像是把  $w$  中的与  $v$  共线的成分反向所得到的向量, 可见此变换描述了反射. 这类形如  $H_v$  的变换称为 **Householder 变换**.

简单计算就有如下结论, 证明留给读者.

**命题 3.2.6** 给定  $\mathbb{R}^n$  中向量  $x, y$ , 满足  $\|x\| = \|y\|$ , 则存在反射  $H_v$ , 其中  $v = \frac{y - x}{\|y - x\|}$ , 使得  $H_v(x) = y$ .

这说明, 对任意等长的向量  $x, y$ , 都能找到一个反射, 使得  $H_v(x) = y$ .

容易验证,  $G_{\theta; i, j}$  和  $H_v$  都是正交矩阵, 从而是保距变换.

### 3.2.2 QR 分解

类似于用矩阵乘法来描述 Gauss 消元法得到 LU 分解, 下面用矩阵乘法描述 Gram-Schmidt 正交化得到 QR 分解.

设  $a_1, \dots, a_n$  是  $\mathbb{R}^n$  的一组基, Gram-Schmidt 正交化的计算过程分为两步, 第一步正交化, 得到一组正交基:

$$\tilde{q}_1 = a_1,$$

$$\begin{aligned}
\tilde{q}_2 &= a_2 - \frac{\tilde{q}_1^T a_2}{\tilde{q}_1^T \tilde{q}_1} \tilde{q}_1, \\
\tilde{q}_3 &= a_3 - \frac{\tilde{q}_1^T a_3}{\tilde{q}_1^T \tilde{q}_1} \tilde{q}_1 - \frac{\tilde{q}_2^T a_3}{\tilde{q}_2^T \tilde{q}_2} \tilde{q}_2, \\
&\vdots \\
\tilde{q}_n &= a_n - \frac{\tilde{q}_1^T a_n}{\tilde{q}_1^T \tilde{q}_1} \tilde{q}_1 - \cdots - \frac{\tilde{q}_{n-1}^T a_n}{\tilde{q}_{n-1}^T \tilde{q}_{n-1}} \tilde{q}_{n-1}.
\end{aligned}$$

第二步再单位化每个向量, 得到标准正交基:  $q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$ .

设  $A = [a_1 \ \cdots \ a_n]$ ,  $\tilde{Q} = [\tilde{q}_1 \ \cdots \ \tilde{q}_n]$ ,  $Q = [q_1 \ \cdots \ q_n]$ , 第一步正交化, 能得到  $a_i$  被  $\tilde{q}_1, \dots, \tilde{q}_i$  线性表示的表示法. 这可用矩阵乘法表示:

$$A = \tilde{Q}\tilde{R}, \quad \tilde{R} = \begin{bmatrix} 1 & \frac{\tilde{q}_1^T a_2}{\tilde{q}_1^T \tilde{q}_1} & \cdots & \frac{\tilde{q}_1^T a_n}{\tilde{q}_1^T \tilde{q}_1} \\ & 1 & \ddots & \vdots \\ & & \ddots & \frac{\tilde{q}_{n-1}^T a_n}{\tilde{q}_{n-1}^T \tilde{q}_{n-1}} \\ & & & 1 \end{bmatrix},$$

第二步单位化可以写成  $\tilde{Q} = Q \operatorname{diag}(\|\tilde{q}_i\|)$ , 因此

$$A = Q \operatorname{diag}(\|\tilde{q}_i\|) \tilde{R} = QR,$$

其中  $R = \operatorname{diag}(\|\tilde{q}_i\|) \tilde{R}$  是对角元素都是正数的上三角矩阵,  $Q$  是正交矩阵. 注意,  $\tilde{Q}$  一般不是正交矩阵.

**定理 3.2.7 (可逆矩阵的 QR 分解)** 设  $A$  是  $n$  阶可逆矩阵, 则存在唯一的分解  $A = QR$ , 其中  $Q$  是正交矩阵,  $R$  是对角元都是正数的上三角矩阵.

证. 存在性: 由 Gram-Schmidt 正交化可得.

唯一性: 设  $A = Q_1 R_1 = Q_2 R_2$ , 其中  $Q_1, Q_2$  是正交矩阵,  $R_1, R_2$  是具有正对角元的上三角矩阵, 且所有矩阵都可逆. 因此,  $Q_2^{-1} Q_1 = R_2 R_1^{-1}$ , 其中,  $Q_2^{-1} Q_1$  是正交矩阵,  $R_2 R_1^{-1}$  是具有正对角元的上三角矩阵. 既是正交矩阵又是具有正对角元的上三角矩阵的矩阵只有单位矩阵. 于是  $Q_1 = Q_2, R_1 = R_2$ , 即分解唯一.  $\square$

分解  $A = QR$  称为矩阵  $A$  的 **QR 分解**.

观察 Gram-Schmidt 正交化的计算过程, 第一步正交化, 是一系列对  $A$  的从左往右的倍加列变换, 从而得到  $\tilde{Q}$ . 注意, 这种倍加矩阵都是单位上三角矩阵, 所以其乘积也是单位上三角矩阵. 因此  $A = \tilde{Q}\tilde{R}$ , 这里  $\tilde{R}$  是单位上三角矩阵. 第二步单位化, 显然可得  $R$  是具有正对角元的上三角矩阵.

例 3.2.8 回顾例 3.1.14 ,

$$A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

利用 Gram-Schmidt 正交化方法得到:

$$\widetilde{Q} = [\tilde{\mathbf{q}}_1 \quad \tilde{\mathbf{q}}_2 \quad \tilde{\mathbf{q}}_3] = \begin{bmatrix} 1 & \frac{1}{3} & \frac{1}{2} \\ 1 & \frac{1}{3} & -\frac{1}{2} \\ 1 & -\frac{2}{3} & 0 \end{bmatrix}, \quad \widetilde{R} = \begin{bmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ & 1 & \frac{1}{2} \\ & & 1 \end{bmatrix},$$

再把正交向量单位化得到:

$$Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3] = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \end{bmatrix},$$

$$R = \text{diag}(\|\tilde{\mathbf{q}}_i\|)\widetilde{R} = \begin{bmatrix} \sqrt{3} & & \\ & \frac{\sqrt{6}}{3} & \\ & & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ & 1 & \frac{1}{2} \\ & & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ & \frac{\sqrt{6}}{3} & \frac{1}{\sqrt{6}} \\ & & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

这就是  $A$  的 QR 分解  $A = QR$ .

由  $A = \widetilde{Q}\widetilde{R}$ , 有  $A\widetilde{R}^{-1} = \widetilde{Q}$ . 我们可以将  $\widetilde{R}^{-1}$  分解成倍加矩阵的乘积:  $\widetilde{R}^{-1} = \begin{bmatrix} 1 & -\frac{2}{3} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} =: R_1 R_2 R_3$ . 如果以  $A$  的列向量作为平行六面体的三条同一点出发的边, 那么可以看到, 第一步正交化的过程  $A \mapsto AR_1 \mapsto AR_1 R_2 \mapsto AR_1 R_2 R_3 = \widetilde{Q}$ , 就是将这个平行六面体成长方体的过程. 参见图 3.2.2. ☺

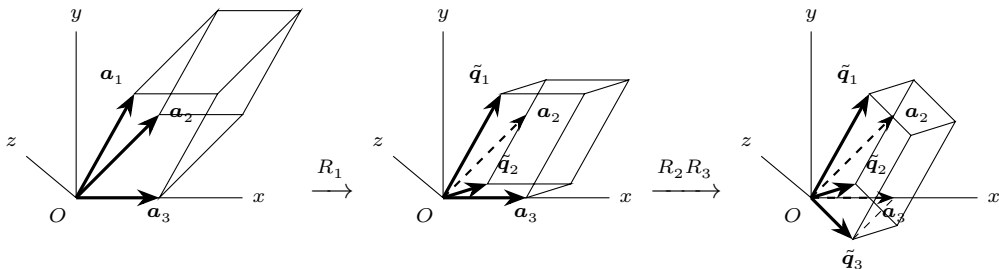


图 3.2.2: 向量的正交化

对  $m \times n$  矩阵, 当  $m \geq n$  时, 也有 QR 分解.

**定义 3.2.9 (列正交矩阵)** 矩阵  $Q$ , 如果满足  $Q^T Q = I_n$ , 则称为列正交矩阵.

显然,  $Q$  是  $m \times n$  列正交矩阵等价于  $Q$  的列向量组构成  $\mathbb{R}^m$  的一个单位向量组, 其中有  $n$  个向量. 此时必有  $m \geq n$ .

**定理 3.2.10 (QR 分解)** 对  $m \times n$  矩阵  $A$ , 其中  $m \geq n$ , 则

1. 存在  $m \times n$  列正交矩阵  $Q_1$  和具有非负对角元的  $n$  阶上三角矩阵  $R_1$ , 使得  $A = Q_1 R_1$ ;
2. 进一步地, 存在  $m$  阶正交矩阵  $Q$  和  $m \times n$  矩阵  $R$ , 使得  $A = QR$ , 其中  $R = \begin{bmatrix} R_1 \\ O \end{bmatrix}$ ,  $Q = [Q_1 \quad Q_2]$ , 即  $Q$  的列向量组由  $Q_1$  的列向量组扩充而成.

证一. 记  $A = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$ . 我们先来证明分解  $A = Q_1 R_1$  的存在性, 其中  $Q_1$  是  $m \times n$  列正交矩阵,  $R_1$  是  $n$  阶具有非负对角元的上三角矩阵.

对  $n$  用数学归纳法. 当  $n = 1$  时, 如果  $\mathbf{a}_1 \neq \mathbf{0}$ , 则令  $Q_1 = \mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$ ,  $R_1 = \|\mathbf{a}_1\|$ , 即有  $A = Q_1 R_1$ , 满足条件. 如果  $\mathbf{a}_1 = \mathbf{0}$ , 则任取一个单位向量做  $Q_1$ , 取  $R_1 = 0$  即可.

假设结论对  $m \times (n-1)$  矩阵成立. 考虑  $m \times n$  矩阵  $A = [A_0 \quad \mathbf{a}_n]$ . 根据归纳假设,  $A_0$  存在分解  $A_0 = Q_0 R_0$ , 其中  $Q_0$  列正交,  $R_0$  是具有非负对角元的  $n-1$  阶上三角矩阵. 如果  $\mathbf{a}_n \in \mathcal{R}(Q_0)$ , 那么存在  $\mathbf{r}_n$ , 使得  $\mathbf{a}_n = Q_0 \mathbf{r}_n$ , 于是任选一个与  $Q_0$  的列向量都正交的单位向量  $\mathbf{q}_n$ , 就有

$$A = [A_0 \quad \mathbf{a}_n] = [Q_0 R_0 \quad Q_0 \mathbf{r}_n] = [Q_0 \quad \mathbf{q}_n] \begin{bmatrix} R_0 & \mathbf{r}_n \\ 0 & 0 \end{bmatrix} = Q_1 R_1,$$

则  $Q_1, R_1$  为满足条件的矩阵. 如果  $\mathbf{a}_n \notin \mathcal{R}(Q_0)$ , 那么用一步 Gram-Schmidt 正交化, 就有  $\tilde{\mathbf{q}}_n = \mathbf{a}_n - Q_0 \mathbf{r}_n$ , 且  $\tilde{\mathbf{q}}_n$  与  $Q_0$  的列向量组都正交. 令  $\mathbf{q}_n = \frac{\tilde{\mathbf{q}}_n}{\|\tilde{\mathbf{q}}_n\|}$ . 于是

$$A = [A_0 \quad \mathbf{a}_n] = [Q_0 R_0 \quad Q_0 \mathbf{r}_n + \|\tilde{\mathbf{q}}_n\| \mathbf{q}_n] = [Q_0 \quad \mathbf{q}_n] \begin{bmatrix} R_0 & \mathbf{r}_n \\ 0 & \|\tilde{\mathbf{q}}_n\| \end{bmatrix} = Q_1 R_1,$$

则  $Q_1, R_1$  为满足条件的矩阵. 综上, 分解  $A = Q_1 R_1$  存在.

最后证明  $A = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$  的存在性. 事实上, 只要将  $\mathbf{q}_1, \dots, \mathbf{q}_n$  扩充成  $\mathbb{R}^m$  的一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_m$ , 那么将这组基向量组成的矩阵记为  $Q$ , 就有  $A = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ .  $\square$

证二. 证一的关键是处理  $A$  的列线性相关的情形. 如果  $A$  列线性无关, 则直接对子空间  $\mathcal{R}(A)$  中  $A$  的列向量 Gram-Schmidt 正交化即得  $Q_1$ . 但当  $A$  的列线性相关时, 我们不得不处理不作为基的列, 处理办法类似于命题 2.3.3 中对行简化阶梯形的讨论. 事实上, 我们也可以利用行简化阶梯形来证明.

用初等行变换把  $A$  化成行简化阶梯形, 数学化就是  $A = BC$ , 其中  $B$  是  $m$  阶可逆矩阵,  $C$  是  $A$  的行简化阶梯形. 由于  $m \geq n$ , 在  $C$  的下方  $m - n$  行都是零行, 因此  $C = \begin{bmatrix} C_1 \\ O \end{bmatrix}$ , 其中  $C_1$  为  $n$  阶上三角矩阵. 而  $B$  可逆, 根据定理 3.2.7, 存在 QR 分解  $B = QR_B$ , 则  $A = QR_B C$ . 我们只需再证明  $R_B C$  的下方  $m - n$  行都是零行. 由于  $R_B$  是上三角矩阵, 可记  $R_B = \begin{bmatrix} R_{11} & R_{12} \\ O & R_{22} \end{bmatrix}$ , 其中  $R_{11}$  是  $n$  阶上三角矩阵,  $R_{22}$  为  $m - n$  阶上三角矩阵. 则  $R_B C = \begin{bmatrix} R_{11} & R_{12} \\ O & R_{22} \end{bmatrix} \begin{bmatrix} C_1 \\ O \end{bmatrix} = \begin{bmatrix} R_{11}C_1 \\ O \end{bmatrix}$ . 记  $Q = [Q_1 \ Q_2]$ , 其中  $Q_1$  是  $m \times n$  矩阵. 令  $R_1 = R_{11}C_1$ , 则  $R_1$  是上三角矩阵, 且  $A = QR_B C = [Q_1 \ Q_2] \begin{bmatrix} R_1 \\ O \end{bmatrix} = Q_1 R_1$ .

注意, 考虑  $A$  的列与  $B$  的列的关系, 事实上  $B$  的前面的列就是  $A$  的所有主列, 后面的列就是利用前面的列扩充成  $\mathbb{R}^m$  的一组基得到的那些向量. 这说明  $Q_1$  的列就是  $A$  的主列 Gram-Schmidt 正交化的结果.  $\square$

证三. 还可以利用 Householder 变换来证明, 我们直接证明第 2 条.

对  $n$  用数学归纳法. 证一已说明  $n = 1$  的情形. 假设结论对任意  $p \times (n - 1)$  矩阵成立, 其中  $p \geq n - 1$ . 考虑  $m \times n$  矩阵  $A = [A_0 \ \mathbf{a}_n]$ .

考虑  $m \times n$  矩阵. 如果  $\mathbf{a}_1 \neq \mathbf{0}$ , 根据命题 3.2.6, 存在 Householder 变换矩阵  $Q_1$ , 使得  $Q_1 \mathbf{a}_1 = \|\mathbf{a}_1\| \mathbf{e}_1$ . 因此  $Q_1 A$  有如下形式  $Q_1 A = \begin{bmatrix} \|\mathbf{a}_1\| & \mathbf{b}^T \\ \mathbf{0} & A_1 \end{bmatrix}$ , 其中  $A_1$  是  $(m - 1) \times (n - 1)$  矩阵. 而  $m - 1 \geq n - 1$ , 因此  $A_1$  满足归纳假设, 即存在正交矩阵

$Q_2$  和上三角矩阵  $R_2$ , 使得  $A_1 = Q_2 \begin{bmatrix} R_2 \\ O \end{bmatrix}$ . 因此  $\begin{bmatrix} 1 & \\ & Q_2^T \end{bmatrix} Q_1 A = \begin{bmatrix} \|\mathbf{a}_1\| & \mathbf{b}^T \\ \mathbf{0} & R_2 \\ \mathbf{0} & O \end{bmatrix}$ . 令

$Q = Q_1^T \begin{bmatrix} 1 & \\ & Q_2 \end{bmatrix}, R_1 = \begin{bmatrix} \|\mathbf{a}_1\| & \mathbf{b}^T \\ \mathbf{0} & R_2 \end{bmatrix}$ . 容易验证,  $A = QR$ , 且  $Q, R$  满足条件.

如果  $\mathbf{a}_1 = \mathbf{0}$  但  $\mathbf{a}_j \neq \mathbf{0}$ , 则通过对换矩阵  $P_{1j}$  将  $\mathbf{a}_j$  换到第一列, 注意到对换矩阵是正交矩阵, 就化为上一种情形. 否则  $A = O$ , 命题显然成立.  $\square$

分解  $A = QR$  称为  $A$  的 **QR 分解**,  $A = Q_1 R_1$  称为  $A$  的 **简化 QR 分解**.

根据以上分析, 简化 QR 分解可以通过 Gram-Schmidt 正交化方法来得到. 这一计算简化 QR 分解的算法常称为经典 Gram-Schmidt 算法, 简称 CGS.

通常, 为了计算上的数值稳定性, 我们常常使用修正 Gram-Schmidt 算法, 简称 MGS. 这一算法和 CGS 相比, 只是改变了计算次序: MGS 中, 我们每算出一个正交基向量, 就在后面的还未计算好的基向量上先减去与当前正交基向量的合适倍数. 换言之, MGS 每个迭代步骤除了计算出一个正交基向量外, 还通过计算使得其他未计算好的基向量都与算好的正交基向量正交.

二者对比, 用矩阵的语言描述, MGS 的每次迭代步骤都计算出  $R$  的一个行; 而 CGS 的每次迭代步骤都计算出  $R$  的一个列. 可以看到 CGS 和 MGS 的计算量相同, 二者都需要约  $2mn^2$  次四则运算和  $n$  次开平方运算.

根据证三, QR 分解和简化 QR 分解, 也可利用 Householder 变换来计算, 这一算法需要约  $2mn^2 - \frac{2}{3}n^3$  次四则运算<sup>1</sup>.

**注 3.2.11** 证三还说明,  $n$  阶正交矩阵可以表示成不多于  $n$  个反射 (Householder 变换) 的乘积.

## 习题

**练习 3.2.1** 求一个四阶正交矩阵, 其中前两个列向量分别为:  $\frac{1}{3} \begin{bmatrix} 1 \\ -2 \\ 0 \\ 2 \end{bmatrix}$ ,  $\frac{1}{\sqrt{6}} \begin{bmatrix} -2 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ .

**练习 3.2.2** 写出元素都是 0 或 1 的所有三阶正交矩阵.

**阅读 3.2.3 (Hadamard 矩阵)** 给定  $n$  阶矩阵  $A$ , 如果  $A$  的元素都是 1 或  $-1$ , 且  $A^T A = nI_n$ , 则称  $A$  是一个  $n$  阶 **Hadamard 矩阵**.

显然 Hadamard 是所有元素绝对值相同的正交矩阵的倍数. 可以证明 Hadamard 矩阵的阶只能是 1, 2 或  $4k, k = 1, 2, \dots$ .

然而是否存在  $4k$  阶 Hadamard 矩阵, 还是一个悬而未决的问题, 称为 **Hadamard 猜想**.

Hadamard 矩阵在信号处理中有应用.

**练习 3.2.4** 1. 列举所有的 1, 2 阶 Hadamard 矩阵.

2. 说明不存在 3 阶 Hadamard 矩阵.

3. 找出一个 4 阶 Hadamard 矩阵.

4. 证明如果  $A$  是 Hadamard 矩阵, 则  $\begin{bmatrix} A & A \\ A & -A \end{bmatrix}$  也是 Hadamard 矩阵. 以此说明存在  $2^n$  阶

Hadamard 矩阵. 提示: 这与阅读 3.1.23 中的 Haar 小波基相关.

<sup>1</sup>开平方可以避免, 算法细节可以参考数值线性代数的相关教材或专著.



**练习 3.2.5** 证明, 上三角矩阵是正交矩阵时, 必是对角矩阵, 且对角元素是  $\pm 1$ .

**练习 3.2.6** 证明, 分块上三角矩阵  $\begin{bmatrix} c & \mathbf{a}^T \\ \mathbf{0} & Q \end{bmatrix}$  是正交矩阵时, 必有  $c = \pm 1, \mathbf{a} = \mathbf{0}$ ,  $Q$  是正交矩阵.

**练习 3.2.7** 对标准基  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , 显然  $\sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T = I_n$ . 对任意标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_n$ , 求证  $\sum_{i=1}^n \mathbf{q}_i \mathbf{q}_i^T = I_n$ .

**练习 3.2.8** 设  $\mathbf{a}_1, \dots, \mathbf{a}_n$  和  $\mathbf{b}_1, \dots, \mathbf{b}_n$  是  $\mathbb{R}^n$  的两组标准正交基, 证明存在正交矩阵  $Q$ , 使得  $Q\mathbf{a}_i = \mathbf{b}_i, 1 \leq i \leq n$ .

**练习 3.2.9** 设  $\mathbf{a}_1, \dots, \mathbf{a}_s$  和  $\mathbf{b}_1, \dots, \mathbf{b}_s$  是  $\mathbb{R}^n$  的两个向量组, 证明, 存在正交矩阵  $Q$ , 使得  $Q\mathbf{a}_i = \mathbf{b}_i, 1 \leq i \leq s$ , 当且仅当  $\mathbf{a}_i^T \mathbf{a}_j = \mathbf{b}_i^T \mathbf{b}_j, 1 \leq i, j \leq s$ .

**练习 3.2.10** 回顾命题 3.2.4, 设  $\mathbf{a}_1, \dots, \mathbf{a}_n$  是  $\mathbb{R}^n$  的一组基, 考虑下列放松的条件.

1.  $A$  在一组基上保距, 即如果对  $1 \leq i \leq n, \|A\mathbf{a}_i\| = \|\mathbf{a}_i\|$ , 那么  $A$  是否一定是正交矩阵?
2.  $A$  在一组基上保内积, 即如果对  $1 \leq i, j \leq n, A\mathbf{a}_i$  与  $A\mathbf{a}_j$  的内积等于  $\mathbf{a}_i$  与  $\mathbf{a}_j$  的内积, 那么  $A$  是否一定是正交矩阵?

**练习 3.2.11** 设  $H_v := I_n - 2\mathbf{v}\mathbf{v}^T$  是  $\mathbb{R}^n$  上反射变换的表示矩阵,  $Q$  是  $n$  阶正交矩阵, 证明  $Q^{-1}H_vQ$  也是某个反射变换的表示矩阵.

**练习 3.2.12** 计算 QR 分解.

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

**练习 3.2.13** 设向量组  $\mathbf{v}_1, \dots, \mathbf{v}_k$  线性无关, 首先令  $\mathbf{q}_1$  为与  $\mathbf{v}_1$  平行的单位向量, 然后令  $\mathbf{q}_2$  为二维子空间  $\text{span}(\mathbf{v}_1, \mathbf{v}_2)$  中垂直于直线  $\text{span}(\mathbf{v}_1)$  的单位向量, 再令  $\mathbf{q}_3$  为三维子空间  $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$  中垂直于平面  $\text{span}(\mathbf{v}_1, \mathbf{v}_2)$  的单位向量, 以此类推. 这样得到的  $\mathbf{q}_1, \dots, \mathbf{q}_k$  与 Gram-Schmidt 正交化得到的结果是否一致? 如果有区别的话, 区别在哪里? 从 QR 分解的角度如何解释?

**练习 3.2.14 (QR 分解的其他理解)** 考虑  $A = \begin{bmatrix} 1 & 0 \\ -2 & 5 \\ 0 & 1 \end{bmatrix}$  的列向量围成的平行四边形.

1. 通过列变换, 把第一列的若干倍加到第二列, 使得平行四边形变成长方形. 这对应着  $A$  右乘哪个矩阵?
2. 通过旋转和反射, 把平行四边形的第一条边变到  $x_1$  轴正半轴上, 第二条边变到  $x_1x_2$  平面中  $x_2 > 0$  的那一半. 这对应着  $A$  左乘哪个正交矩阵?

**练习 3.2.15** 证明命题 3.2.6.

**练习 3.2.16** 证明任意  $n$  阶正交矩阵可以表示成不多于  $n$  个反射的乘积.

**练习 3.2.17 (保角变换)** 设可逆矩阵  $A$  对应的线性变换保持向量之间的角度不变.

1. 对  $A$  进行  $QR$  分解, 证明  $R$  也保持向量之间的角度不变.
2. 证明  $R$  为对角矩阵.
3. 证明  $R = kI_n$ , 这里  $k$  为常数. 由此得到,  $A$  必是某个正交矩阵的倍数.

**练习 3.2.18** 设  $\mathbf{a}_1, \dots, \mathbf{a}_m$  是  $\mathbb{R}^n$  中的  $s$  个向量, 定义矩阵

$$G(\mathbf{a}_1, \dots, \mathbf{a}_m) := \begin{bmatrix} \mathbf{a}_1^T \mathbf{a}_1 & \cdots & \mathbf{a}_1^T \mathbf{a}_m \\ \vdots & & \vdots \\ \mathbf{a}_m^T \mathbf{a}_1 & \cdots & \mathbf{a}_m^T \mathbf{a}_m \end{bmatrix},$$

称为  $\mathbf{a}_1, \dots, \mathbf{a}_s$  的 **Gram 矩阵**. 证明,

1.  $\mathbf{a}_1, \dots, \mathbf{a}_s$  是标准正交向量组当且仅当  $G(\mathbf{a}_1, \dots, \mathbf{a}_m) = I_m$ .
2. Gram 矩阵  $G = G(\mathbf{a}_1, \dots, \mathbf{a}_m)$  是  $m$  阶对称矩阵, 且对任意  $\mathbf{x} \in \mathbb{R}^m$ , 都有  $\mathbf{x}^T G \mathbf{x} \geq 0$ .
3.  $\mathbf{a}_1, \dots, \mathbf{a}_s$  线性无关当且仅当  $G = G(\mathbf{a}_1, \dots, \mathbf{a}_m)$  可逆, 也等价于对任意非零  $\mathbf{x} \in \mathbb{R}^m$ , 都有  $\mathbf{x}^T G \mathbf{x} > 0$ .

### 3.3 正交投影

先看例 3.1.1 的一个高维推广.

**例 3.3.1** 考虑方程组  $A\mathbf{x} = \mathbf{b}$ , 其中  $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 3 \end{bmatrix}$ ,  $\mathbf{b} = \begin{bmatrix} 0 \\ 3 \\ 5 \end{bmatrix}$ ,  $\mathbf{x} \in \mathbb{R}^2$ . 简单计算可知方

程组无解. 那么, 如何找到  $\hat{\mathbf{x}} \in \mathbb{R}^2$ , 满足

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \min_{\mathbf{x} \in \mathbb{R}^2} \|\mathbf{b} - A\mathbf{x}\|?$$

这里  $\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2\} = \mathcal{R}(A)$  是  $\mathbb{R}^3$  中过原点的平面, 而  $\|\mathbf{b} - A\mathbf{x}\|$  的最小值为向量  $\mathbf{b}$  到该平面的距离, 见图 3.3.1. 对向量  $\mathbf{b}$  向该平面做垂直投影, 即考虑分解  $\mathbf{b} = \hat{\mathbf{b}} + \mathbf{r}$ , 满足  $\hat{\mathbf{b}} \in \mathcal{R}(A)$ , 且  $\hat{\mathbf{b}}$  与  $\mathbf{r}$  垂直. 由于  $\mathcal{R}(A)$  是二维子空间, 若取它的一组基, 则可对  $\hat{\mathbf{b}}$  进一步分解. 记  $A = [\mathbf{a}_1 \ \mathbf{a}_2]$ , 则  $\mathbf{a}_1, \mathbf{a}_2$  为  $\mathcal{R}(A)$  的一组基. 考虑分解  $\mathbf{b} = \hat{x}_1 \mathbf{a}_1 + \hat{x}_2 \mathbf{a}_2 + \mathbf{r}$ , 求实数  $\hat{x}_1, \hat{x}_2$ , 满足  $\mathbf{r}$  与  $\mathbf{a}_1, \mathbf{a}_2$  同时正交. 这等价于解方程  $\mathbf{a}_1^T \mathbf{r} = \mathbf{a}_2^T \mathbf{r} = 0$ , 即  $A^T \mathbf{r} = \mathbf{0}$ , 其中  $\mathbf{r} = \mathbf{b} - A\hat{\mathbf{x}}$ . 计算留给读者.

如果对  $\mathbf{a}_1, \mathbf{a}_2$  做 Gram-Schmidt 正交化, 则得到  $\mathcal{R}(A)$  的一组标准正交基  $\mathbf{q}_1, \mathbf{q}_2$ . 再考虑分解  $\mathbf{b} = \hat{y}_1 \mathbf{q}_1 + \hat{y}_2 \mathbf{q}_2 + \mathbf{r}$ , 其中  $\mathbf{q}_1, \mathbf{q}_2, \mathbf{r}$  两两正交. 易得  $\hat{y}_1 = \mathbf{q}_1^T \mathbf{b}$ ,  $\hat{y}_2 = \mathbf{q}_2^T \mathbf{b}$ . 我们得到了与向量向直线的垂直投影 (3.1.1) 类似的分解:

$$\mathbf{b} = (\mathbf{q}_1^T \mathbf{b}) \mathbf{q}_1 + (\mathbf{q}_2^T \mathbf{b}) \mathbf{q}_2 + \mathbf{r}. \quad (3.3.1)$$

其成立的关键是  $\mathbf{q}_1, \mathbf{q}_2$  是  $\mathcal{R}(A)$  的标准正交基.

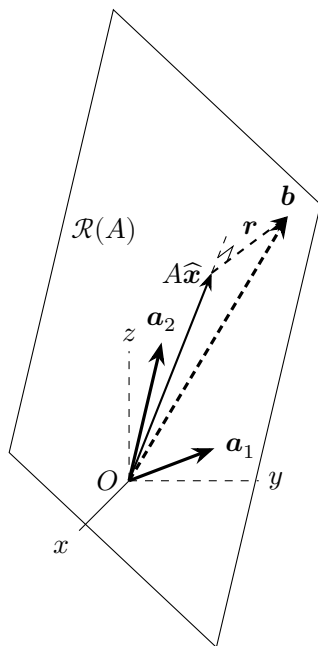


图 3.3.1: 向量的逼近

用矩阵语言表达, 就是  $\mathbf{b} = A\hat{\mathbf{x}} + \mathbf{r} = Q\hat{\mathbf{y}} + \mathbf{r}$ , 其中  $Q = [\mathbf{q}_1 \ \mathbf{q}_2]$  是列正交矩阵, 且满足  $Q^T \mathbf{r} = \mathbf{0}$ . 因此  $\hat{\mathbf{y}} = Q^T \mathbf{b}$ . Gram-Schmidt 正交化对应于 QR 分解  $A = QR$ , 其中  $R$  是二阶可逆上三角矩阵. 由  $Q\hat{\mathbf{y}} = A\hat{\mathbf{x}} = QR\hat{\mathbf{x}}$  可知,  $\hat{\mathbf{x}} = R^{-1}\hat{\mathbf{y}}$ .

下面计算 QR 分解. 首先计算正交向量组  $\tilde{\mathbf{q}}_1, \tilde{\mathbf{q}}_2$ :

$$\tilde{\mathbf{q}}_1 = \mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \tilde{\mathbf{q}}_2 = \mathbf{a}_2 - \frac{\tilde{\mathbf{q}}_1^T \mathbf{a}_2}{\tilde{\mathbf{q}}_1^T \tilde{\mathbf{q}}_1} \tilde{\mathbf{q}}_1 = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} - \frac{6}{6} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}.$$

矩阵  $\tilde{R} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ . 再单位化可得  $\mathbf{q}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ ,  $\mathbf{q}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}$ . 因此  $A = QR$ , 其中

$$Q = \begin{bmatrix} \frac{1}{\sqrt{6}} & 0 \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} \end{bmatrix}, \quad R = \begin{bmatrix} \sqrt{6} & \\ & \sqrt{5} \end{bmatrix} \tilde{R} = \begin{bmatrix} \sqrt{6} & \sqrt{6} \\ 0 & \sqrt{5} \end{bmatrix}.$$

易知

$$R^{-1} = \tilde{R}^{-1} \begin{bmatrix} \frac{1}{\sqrt{6}} & \\ & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \\ & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{bmatrix}.$$

于是

$$\hat{\mathbf{y}} = Q^T \mathbf{b} = \begin{bmatrix} \frac{11}{\sqrt{6}} \\ \frac{7}{\sqrt{5}} \end{bmatrix}, \hat{\mathbf{x}} = R^{-1} \hat{\mathbf{y}} = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{5}} \\ 0 & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{11}{\sqrt{6}} \\ \frac{7}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{13}{30} \\ \frac{7}{5} \end{bmatrix}.$$

因此,  $\mathbf{r} = \mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} -\frac{11}{6} \\ \frac{11}{15} \\ \frac{11}{30} \end{bmatrix}$ , 最小距离  $\|\mathbf{r}\| = \frac{11}{\sqrt{30}}$ . ☺

在例 3.3.1 中, 为了求距离的最小值  $\min_{\mathbf{x} \in \mathbb{R}^2} \|\mathbf{b} - A\mathbf{x}\| = \min_{\mathbf{y} \in \mathcal{R}(A)} \|\mathbf{b} - \mathbf{y}\|$ , 需要对平面  $\mathcal{R}(A)$  做正交投影.

下面我们将把类似概念推广到任意子空间上.

### 3.3.1 正交补

由内积的线性性质易得如下结论.

**命题 3.3.2** 如果  $\mathbf{b}$  与  $\mathbf{a}_1, \dots, \mathbf{a}_s$  都正交, 则  $\mathbf{b}$  与子空间  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_s)$  中的任意向量都正交.

在  $\mathbb{R}^3$  中, 命题 3.3.2 的几何描述, 就是向量垂直于某个平面当且仅当它垂直于平面内两条相交直线.

根据内积的定义, 齐次方程组  $A\mathbf{x} = \mathbf{0}$  的任意解向量  $\mathbf{x}$  与矩阵  $A$  的所有行向量都正交. 因此, 零空间  $\mathcal{N}(A)$  中的任意向量和行空间  $\mathcal{R}(A^T)$  中的任意向量都正交.

**定义 3.3.3 (子空间正交)** 给定  $\mathbb{R}^n$  的子空间  $\mathcal{M}, \mathcal{N}$ , 如果  $\mathcal{M}$  中任意向量和  $\mathcal{N}$  中任意向量都正交, 则称  $\mathcal{M}$  与  $\mathcal{N}$  正交, 记为  $\mathcal{M} \perp \mathcal{N}$ .

特别地, 如果  $\text{span}(\mathbf{a}) \perp \mathcal{M}$ , 则简称向量  $\mathbf{a}$  与子空间  $\mathcal{M}$  正交, 记为  $\mathbf{a} \perp \mathcal{M}$ .

对任意矩阵  $A$ ,  $\mathcal{N}(A) \perp \mathcal{R}(A^T)$ ,  $\mathcal{N}(A^T) \perp \mathcal{R}(A)$ . 进一步地, 所有与  $\mathcal{R}(A^T)$  正交的向量都在  $\mathcal{N}(A)$  中 (为什么?).

**定义 3.3.4 (正交补)** 给定  $\mathbb{R}^n$  的子空间  $\mathcal{M}$ , 则  $\mathbb{R}^n$  的子集  $\mathcal{M}^\perp := \{\mathbf{a} \in \mathbb{R}^n \mid \mathbf{a} \perp \mathcal{M}\}$ , 称为  $\mathcal{M}$  的正交补.

**例 3.3.5** 设  $A = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ , 则  $\mathcal{N}(A) = \text{span}(\mathbf{e}_2, \mathbf{e}_3)$ ,  $\mathcal{R}(A^T) = \text{span}(\mathbf{e}_1)$ , 两个子空间互为正交补. ☺

**命题 3.3.6** 如果  $\mathcal{M}$  是  $\mathbb{R}^n$  的子空间, 则其正交补  $\mathcal{M}^\perp$  也是  $\mathbb{R}^n$  的子空间.

证. 首先,  $\mathcal{M}^\perp$  非空:  $\mathbf{0} \in \mathcal{M}^\perp$ .

其次,  $\mathcal{M}^\perp$  对线性运算封闭: 如果  $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{M}^\perp$ , 则对任意  $\mathbf{b} \in \mathcal{M}$ ,  $\mathbf{a}_1^\mathrm{T} \mathbf{b} = \mathbf{a}_2^\mathrm{T} \mathbf{b} = 0$ , 于是对任意线性组合  $k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2$ , 有  $(k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2)^\mathrm{T} \mathbf{b} = k_1 \mathbf{a}_1^\mathrm{T} \mathbf{b} + k_2 \mathbf{a}_2^\mathrm{T} \mathbf{b} = 0$ , 即  $k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2 \in \mathcal{M}^\perp$ .  $\square$

对两个子空间  $\mathcal{M}, \mathcal{N}$ , 如果  $\mathcal{M} \perp \mathcal{N}$ , 那么  $\mathcal{M} \subseteq \mathcal{N}^\perp, \mathcal{N} \subseteq \mathcal{M}^\perp$ ; 如果  $\mathcal{M} \subseteq \mathcal{N}$ , 那么  $\mathcal{N}^\perp \subseteq \mathcal{M}^\perp$ .

正交补具有如下性质.

**命题 3.3.7** 对  $\mathbb{R}^n$  的子空间  $\mathcal{M}$ , 有

1.  $\mathcal{M} \cap \mathcal{M}^\perp = \{\mathbf{0}\}$ ;
2.  $\dim \mathcal{M}^\perp = n - \dim \mathcal{M}$ ;
3.  $(\mathcal{M}^\perp)^\perp = \mathcal{M}$ ;
4. 对任意  $\mathbf{a} \in \mathbb{R}^n$ , 都存在唯一的分解  $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$ , 使得  $\mathbf{a}_1 \in \mathcal{M}, \mathbf{a}_2 \in \mathcal{M}^\perp$ .

证. 第 1 条: 对任意  $\mathbf{a} \in \mathcal{M} \cap \mathcal{M}^\perp$ , 则  $\mathbf{a} \perp \mathbf{a}$ , 因此  $\mathbf{a}^\mathrm{T} \mathbf{a} = 0$ , 可知  $\mathbf{a} = \mathbf{0}$ .

第 2 条: 取  $\mathcal{M}$  的一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_r$ , 根据正交基扩充定理 (命题 3.1.13), 将其扩充成  $\mathbb{R}^n$  的一组基  $\mathbf{q}_1, \dots, \mathbf{q}_r, \mathbf{q}_{r+1}, \dots, \mathbf{q}_n$ . 容易验证  $\mathbf{q}_{r+1}, \dots, \mathbf{q}_n$  是  $\mathcal{M}^\perp$  的一组标准正交基. 计算维数可得.

第 3 条: 显然有  $\mathcal{M} \subseteq (\mathcal{M}^\perp)^\perp$ . 根据第 2 条,  $\dim(\mathcal{M}^\perp)^\perp = n - \dim \mathcal{M}^\perp = \dim \mathcal{M}$ . 再由命题 2.2.18,  $\mathcal{M} = (\mathcal{M}^\perp)^\perp$ .

第 4 条: 取  $\mathcal{M}$  的一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_r$ , 对任意  $\mathbf{a} \in \mathbb{R}^n$ , 令  $\mathbf{a}_1 = (\mathbf{q}_1^\mathrm{T} \mathbf{a}) \mathbf{q}_1 + \dots + (\mathbf{q}_r^\mathrm{T} \mathbf{a}) \mathbf{q}_r$ . 容易验证  $\mathbf{q}_i^\mathrm{T} \mathbf{a} = \mathbf{q}_i^\mathrm{T} \mathbf{a}_1, 1 \leq i \leq r$ . 因此  $\mathbf{a} - \mathbf{a}_1$  与每个  $\mathbf{q}_i$  都正交, 即  $\mathbf{a}_2 = \mathbf{a} - \mathbf{a}_1 \in \mathcal{M}^\perp$ . 唯一性: 假设有两个分解  $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2 = \mathbf{a}'_1 + \mathbf{a}'_2$ , 则  $\mathbf{a}_1 - \mathbf{a}'_1 = \mathbf{a}'_2 - \mathbf{a}_2 \in \mathcal{M} \cap \mathcal{M}^\perp = \{\mathbf{0}\}$ , 因此  $\mathbf{a}_1 = \mathbf{a}'_1, \mathbf{a}_2 = \mathbf{a}'_2$ , 即分解唯一.  $\square$

应用到  $\mathcal{M} = \mathcal{R}(A), \mathcal{R}(A^\mathrm{T})$  的情形, 我们有如下的结果.

**定理 3.3.8** 给定  $m \times n$  矩阵  $A$ , 则

1.  $\mathcal{R}(A^\mathrm{T})^\perp = \mathcal{N}(A), \mathcal{R}(A)^\perp = \mathcal{N}(A^\mathrm{T})$ ;
2.  $\mathcal{R}(A^\mathrm{T}A) = \mathcal{R}(A^\mathrm{T}), \mathcal{N}(A^\mathrm{T}A) = \mathcal{N}(A)$ ;
3.  $\mathcal{R}(AA^\mathrm{T}) = \mathcal{R}(A), \mathcal{N}(AA^\mathrm{T}) = \mathcal{N}(A^\mathrm{T})$ .

证. 第 1 条: 显然.

第 2 条: 先证  $\mathcal{N}(A^\mathrm{T}A) = \mathcal{N}(A)$ : 显然  $\mathcal{N}(A) \subseteq \mathcal{N}(A^\mathrm{T}A)$ , 下证  $\mathcal{N}(A^\mathrm{T}A) \subseteq \mathcal{N}(A)$ . 对任意  $\mathbf{x} \in \mathcal{N}(A^\mathrm{T}A)$ , 有  $A^\mathrm{T}A\mathbf{x} = \mathbf{0}$ , 两边同时左乘行向量  $\mathbf{x}^\mathrm{T}$  可得  $\mathbf{x}^\mathrm{T}A^\mathrm{T}A\mathbf{x} = 0$ . 设

$\mathbf{y} = A\mathbf{x}$ , 则  $\mathbf{y}^T = \mathbf{x}^T A^T$ ,  $0 = \mathbf{y}^T \mathbf{y} = y_1^2 + \cdots + y_m^2$ . 而  $y_i$  都是实数, 因此  $\mathbf{y} = \mathbf{0}$ , 即  $\mathbf{x} \in \mathcal{N}(A)$ .

根据零空间维数定理 2.4.1,  $\text{rank}(A^T A) = n - \dim \mathcal{N}(A^T A) = n - \dim \mathcal{N}(A) = \text{rank}(A)$ . 而  $\mathcal{R}(A^T A) \subseteq \mathcal{R}(A^T)$ , 由命题 2.2.18 知,  $\mathcal{R}(A^T A) = \mathcal{R}(A^T)$ .

第 3 条: 对  $A^T$  应用第 2 条. □

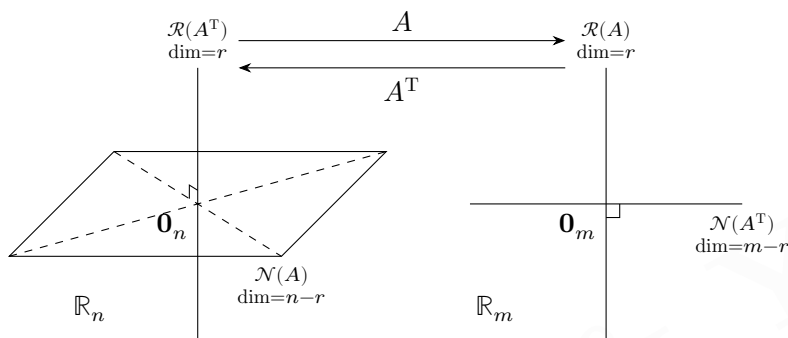


图 3.3.2: 矩阵导出的四个子空间关系图

这四个子空间的关系可以用图 3.3.2 来表示, 其中  $r = \text{rank}(A)$ , 有如下解释:

1.  $\mathcal{R}(A^T)$  和  $\mathcal{N}(A)$  在  $\mathbb{R}^n$  中互为正交补;  $\mathcal{R}(A)$  和  $\mathcal{N}(A^T)$  在  $\mathbb{R}^m$  中互为正交补.
2.  $A$  对应的线性映射  $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$  把  $\mathcal{N}(A)$  映射到  $\{\mathbf{0}_m\}$ , 把  $\mathcal{R}(A^T)$  映射到  $\mathcal{R}(A)$ ;
3.  $A^T$  对应的线性映射  $A^*: \mathbb{R}^m \rightarrow \mathbb{R}^n$  把  $\mathcal{N}(A^T)$  映射到  $\{\mathbf{0}_n\}$ , 把  $\mathcal{R}(A)$  映射到  $\mathcal{R}(A^T)$ .

**例 3.3.9** 回顾例 3.3.1,  $A = \begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 3 \end{bmatrix}$ ,  $\mathbf{r} = \frac{11}{30} \begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}$ , 则  $\mathcal{N}(A^T) =$

$\text{span}(\mathbf{r}) = \mathcal{R}(A)^\perp$ . 而向量  $\mathbf{b}$  的分解是  $\mathbf{b} = A\hat{\mathbf{x}} + \mathbf{r}$ ,  $A\hat{\mathbf{x}} \in \mathcal{R}(A)$ ,  $\mathbf{r} \in \mathcal{N}(A^T)$ .

另一方面,  $\mathcal{R}(A^T) = \mathbb{R}^2$ , 所以  $\mathcal{N}(A) = \mathcal{R}(A^T)^\perp = \{\mathbf{0}\}$ . ☺

### 3.3.2 正交投影

给定  $\mathbb{R}^n$  的子空间  $\mathcal{M}$ , 根据命题 3.3.7 可知, 对任意  $\mathbf{a} \in \mathbb{R}^n$ , 都有唯一的分解  $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$ , 其中  $\mathbf{a}_1 \in \mathcal{M}$ ,  $\mathbf{a}_2 \in \mathcal{M}^\perp$ .

定义  $\mathbb{R}^n$  上的一个变换

$$P_{\mathcal{M}}: \mathbf{a} \mapsto \mathbf{a}_1,$$

它是线性变换: 如果  $\mathbf{a}$  和  $\mathbf{b}$  分别有分解  $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2, \mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$ , 则  $\mathbf{a} + \mathbf{b} = (\mathbf{a}_1 + \mathbf{b}_1) + (\mathbf{a}_2 + \mathbf{b}_2)$  是  $\mathbf{a} + \mathbf{b}$  唯一的分解, 因此  $\mathbf{P}_{\mathcal{M}}(\mathbf{a} + \mathbf{b}) = \mathbf{a}_1 + \mathbf{b}_1 = \mathbf{P}_{\mathcal{M}}(\mathbf{a}) + \mathbf{P}_{\mathcal{M}}(\mathbf{b})$ ; 类似可证  $\mathbf{P}_{\mathcal{M}}(k\mathbf{a}) = k\mathbf{P}_{\mathcal{M}}(\mathbf{a})$ .

**定义 3.3.10** 给定  $\mathbb{R}^n$  的子空间  $\mathcal{M}$ , 线性变换  $\mathbf{P}_{\mathcal{M}}$  称为子空间  $\mathcal{M}$  上的**正交投影 (变换)**, 而  $\mathbf{a}_1 = \mathbf{P}_{\mathcal{M}}(\mathbf{a})$  称为向量  $\mathbf{a}$  在  $\mathcal{M}$  上的**正交投影**.

特别地,  $\mathbf{a} \in \mathcal{M}$  当且仅当  $\mathbf{P}_{\mathcal{M}}(\mathbf{a}) = \mathbf{a}$ , 而  $\mathbf{a} \in \mathcal{M}^\perp$  当且仅当  $\mathbf{P}_{\mathcal{M}}(\mathbf{a}) = \mathbf{0}$ .

线性变换  $\mathbf{P}_{\mathcal{M}^\perp}: \mathbf{a} \mapsto \mathbf{a}_2$  是  $\mathcal{M}^\perp$  上的正交投影 (变换), 而  $\mathbf{a}_2$  是  $\mathbf{a}$  在  $\mathcal{M}^\perp$  上的正交投影. 注意,  $\mathbf{a}_1 \perp \mathbf{a}_2$ , 因此一个向量在一个子空间上的正交投影, 与其在该子空间的正交补上的投影总正交, 这就是这种变换称为正交投影的原因.

显然  $\mathbf{I} = \mathbf{P}_{\mathcal{M}} + \mathbf{P}_{\mathcal{M}^\perp}$ .

**例 3.3.11** 给定  $\mathbb{R}^3$  的子空间  $\mathcal{M} = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$  是  $\mathbf{e}_1$ - $\mathbf{e}_2$  坐标平面. 其正交补  $\mathcal{M}^\perp = \text{span}(\mathbf{e}_3)$ , 是  $\mathbf{e}_3$  坐标轴. 正交投影  $\mathbf{P}_{\mathcal{M}} \left( \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \right) = \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix}, \mathbf{P}_{\mathcal{M}^\perp} \left( \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \\ a_3 \end{bmatrix}$ , 分

别是  $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$  向  $\mathbf{e}_1$ - $\mathbf{e}_2$  平面和  $\mathbf{e}_3$  轴的正交投影. ⊙

利用正交投影可以求得最小距离.

**命题 3.3.12** 给定  $\mathbb{R}^n$  的子空间  $\mathcal{M}$  和向量  $\mathbf{a}$ , 而  $\mathbf{a}_1 = \mathbf{P}_{\mathcal{M}}(\mathbf{a})$  为  $\mathbf{a}$  在  $\mathcal{M}$  上的正交投影, 则  $\|\mathbf{a} - \mathbf{a}_1\| = \min_{\mathbf{x} \in \mathcal{M}} \|\mathbf{a} - \mathbf{x}\|$ .

证明与命题 3.1.7 的证明类似, 留给读者.

如何计算正交投影? 设  $\mathbf{q}_1, \dots, \mathbf{q}_r$  是  $\mathcal{M}$  的一组标准正交基. 对任意  $\mathbf{a} \in \mathbb{R}^n$ , 容易验证  $\mathbf{P}_{\mathcal{M}}(\mathbf{a}) = (\mathbf{q}_1^T \mathbf{a})\mathbf{q}_1 + \dots + (\mathbf{q}_r^T \mathbf{a})\mathbf{q}_r$ . 令  $\mathbf{Q}_r = [\mathbf{q}_1 \ \dots \ \mathbf{q}_r]$ , 它是列正交矩阵, 因为  $\mathbf{Q}_r^T \mathbf{Q}_r = \mathbf{I}_r$ . 于是

$$\mathbf{P}_{\mathcal{M}}(\mathbf{a}) = \mathbf{q}_1 \mathbf{q}_1^T \mathbf{a} + \dots + \mathbf{q}_r \mathbf{q}_r^T \mathbf{a} = \mathbf{Q}_r \mathbf{Q}_r^T \mathbf{a}.$$

因此正交投影  $\mathbf{P}_{\mathcal{M}}$  的表示矩阵就是  $\mathbf{Q}_r \mathbf{Q}_r^T$ , 记为  $\mathbf{P}_{\mathcal{M}}$ . 注意, 表示矩阵  $\mathbf{P}_{\mathcal{M}}$  与  $\mathcal{M}$  的正交基和  $\mathbf{Q}_r$  的选取无关 (为什么?).

下面讨论  $\mathcal{M} = \mathcal{R}(A)$  的情形, 此时  $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ .

**定义 3.3.13** 给定矩阵  $A$ , 其列空间上的正交投影的表示矩阵  $\mathbf{P}_{\mathcal{R}(A)}$ , 称为关于  $A$  的**正交投影矩阵**, 简记为  $\mathbf{P}_A$ .

当  $A$  是可逆方阵时,  $\mathcal{R}(A) = \mathbb{R}^n$ , 此时正交投影就是恒同变换, 因此  $P_A = I_n$ .

如果  $P$  是关于  $A$  的正交投影矩阵, 则  $P = P_{\mathcal{R}(A)}$ , 于是  $I_n - P = P_{\mathcal{R}(A)^\perp} = P_{\mathcal{N}(A^T)}$  也是正交投影矩阵 (关于哪个矩阵?). 正交投影矩阵有如下性质.

**命题 3.3.14** 给定  $n$  阶方阵  $P$ ,  $P$  是正交投影矩阵, 当且仅当  $P^2 = P^T = P$ .

证. “ $\Rightarrow$ ”: 假设  $P$  是关于  $A$  的投影矩阵, 且  $\mathcal{R}(A)$  的一组标准正交基组成矩阵  $Q_r$ , 那么  $P = Q_r Q_r^T$  是对称矩阵. 同时,  $P^2 = (Q_r Q_r^T)(Q_r Q_r^T) = Q_r Q_r^T = P$ .

“ $\Leftarrow$ ”: 下证  $P$  是关于矩阵  $P$  本身的投影矩阵. 对任意向量  $\mathbf{x}$ ,  $P\mathbf{x} \in \mathcal{R}(P)$ . 由于  $P^2 = P = P^T$ , 因此  $P^T(\mathbf{x} - P\mathbf{x}) = 0$ , 即  $\mathbf{x} - P\mathbf{x} \in \mathcal{N}(P^T) = \mathcal{R}(P)^\perp$ . 因此  $\mathbf{x} = P\mathbf{x} + (\mathbf{x} - P\mathbf{x})$  是正交投影对应的唯一分解, 该正交投影的表示矩阵是  $P$ .  $\square$

**例 3.3.15** 继续讨论例 3.3.1, 取  $\mathcal{R}(A)$  的一组标准正交基并列排成的矩阵是  $Q =$

$\begin{bmatrix} \frac{1}{\sqrt{6}} & 0 \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} \end{bmatrix}$ . 正交投影矩阵  $P_A = QQ^T = \frac{1}{6} \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{13}{15} & -\frac{1}{15} \\ \frac{1}{6} & -\frac{1}{15} & \frac{29}{30} \end{bmatrix}$ . 向量  $\mathbf{b}$  的正交投影分解为  $\mathbf{b} = P_A \mathbf{b} + (I_3 - P_A) \mathbf{b}$ , 可以验证  $A\hat{\mathbf{x}} = P_A \mathbf{b}$ ,  $(I_2 - P_A) \mathbf{b} = \mathbf{r}$ , 与例 3.3.1 中的分解  $\mathbf{b} = A\hat{\mathbf{x}} + \mathbf{r}$  一致.  $\odot$

如果不计算  $\mathcal{R}(A)$  的标准正交基, 能否求出正交投影矩阵?

对任意向量  $\mathbf{b}$ , 记其在  $\mathcal{R}(A)$  上的正交投影为  $A\mathbf{x}$ , 则  $\mathbf{b} - A\mathbf{x} \perp \mathcal{R}(A)$ . 因此  $\mathbf{b} - A\mathbf{x} \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ , 即  $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$ , 于是  $A^T A \mathbf{x} = A^T \mathbf{b}$ . 考虑这个线性方程组. 根据定理 3.3.8,  $\mathcal{R}(A^T A) = \mathcal{R}(A^T)$ , 因此  $A^T \mathbf{b} \in \mathcal{R}(A^T A)$ . 这说明  $A^T A \mathbf{x} = A^T \mathbf{b}$  有解.

如果  $A^T A$  可逆, 那么唯一解  $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}$ , 于是  $\mathbf{b}$  在  $\mathcal{R}(A)$  上的正交投影是  $A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}$ . 这意味着, 关于  $A$  的正交投影矩阵是

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

特别地, 如果  $A$  是列正交矩阵, 则  $A^T A = I_r$ ,  $P_A = AA^T$ .

注意,  $A^T A$  可逆, 并不意味着  $A^T$  和  $A$  可逆, 事实上, 两个矩阵都不一定是方阵.

对  $A^T A$  不可逆的情形, 现在还无法将正交投影矩阵写成  $A$  的表达式, 相关内容我们将在第 6 章中讲述. 另一方面, 正交投影矩阵仍能用前述方法计算. 事实上,  $A^T A$  可逆当且仅当  $A$  列满秩, 即  $A$  的列向量组线性无关. 可以先求出  $\mathcal{R}(A)$  的一组基, 并列排成矩阵  $B$ , 则  $B$  列满秩, 且  $\mathcal{R}(B) = \mathcal{R}(A)$ . 因此,  $P_A = P_B = B(B^T B)^{-1} B^T$ .

### 3.3.3 最小二乘问题

现在我们考虑一个实际问题. 假设两个量  $y, t$  之间存在某种规律, 不妨用函数  $y = f(t)$  表示, 但是函数的具体形式并不清楚. 那么, 能否通过测量  $y$  根据  $t$  的变化, 亦即



通过测量  $y$  在  $t$  的若干个取值  $t_1, \dots, t_m$  上的值  $y_1, \dots, y_m$  (这一过程称为**采样**), 来总结出函数  $f$  的具体形式?

**例 3.3.16** 我们想研究人群受全日制教育的年数与其收入是否有关系. 首先通过调查, 我们得到了  $n$  个人的数据  $(t_1, y_1), \dots, (t_n, y_n)$ , 这里  $t_i$  是第  $i$  个人接受全日制教育的年数, 而  $y_i$  是其 35 岁时的年收入.

先拟设  $t_i$  和  $y_i$  满足线性关系, 即函数  $y = f(t) = kt + b$ . 这意味着, 每增加一年全日制教育, 35 岁时的年收入就会多  $k$  (当然这不大现实).

令  $\mathbf{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$ ,  $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ ,  $\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ , 则假设引出方程  $k\mathbf{t} + b\mathbf{1} = \mathbf{y}$ , 其中  $k, b$  是未知

数. 一般来说, 这方程组无解. 但我们可以试图去找一组  $k, b$  使得  $\mathbf{y} - k\mathbf{t} - b\mathbf{1}$  尽可能小.

记  $A = [\mathbf{t} \ \mathbf{1}]$ ,  $\mathbf{x} = \begin{bmatrix} k \\ b \end{bmatrix}$ , 则问题变为求解  $\min_{\mathbf{x} \in \mathbb{R}^2} \|\mathbf{y} - A\mathbf{x}\|$ .

上节已经给出了问题的解法, 即  $A\mathbf{x}$  应该是  $\mathbf{y}$  在  $\mathcal{R}(A)$  上的正交投影. 不妨假设  $t_i$  不全相同, 则  $A$  列满秩, 问题的解  $\mathbf{x} = (A^T A)^{-1} A^T \mathbf{y}$ . 计算可得  $k = \frac{\mathbf{t}^T \mathbf{y} - \frac{1}{n} (\mathbf{1}^T \mathbf{t})(\mathbf{1}^T \mathbf{y})}{\mathbf{t}^T \mathbf{t} - \frac{1}{n} (\mathbf{1}^T \mathbf{t})^2}$ , 其中分子是  $t_i$  的方差, 分母是  $t_i$  与  $y_i$  的协方差.

上述模型称为线性拟合. 类似地, 我们也可以考虑二次拟合, 即拟设  $y = f(t) = at^2 + bt + c$ , 计算过程类似. 事实上, 拟设的函数可以具有任意形式.  $\odot$

一般地, 我们总会假定两个量之间的函数  $f$  具有某种特殊形式, 比如  $f$  是线性函数、反比例函数、多项式函数、指数函数等. 假设  $f(t) = x_1 f_1(t) + \dots + x_n f_n(t)$ . 那么求解  $f$  的具体形式, 就是求解这些系数  $x_1, \dots, x_n$ .

这个模型称为**数据拟合**, 它在各种领域中都有广泛的应用.

通过上面的已知条件, 可以列出方程组:

$$\begin{cases} f_1(t_1)x_1 + f_2(t_1)x_2 + \dots + f_n(t_1)x_n = y_1, \\ f_1(t_2)x_1 + f_2(t_2)x_2 + \dots + f_n(t_2)x_n = y_2, \\ \vdots \\ f_1(t_m)x_1 + f_2(t_m)x_2 + \dots + f_n(t_m)x_n = y_m, \end{cases}$$

记  $\mathbf{x} = [x_j]_{n \times 1}$ ,  $\mathbf{b} = [y_i]_{m \times 1}$ ,  $A = [f_j(t_i)]_{m \times n}$ , 则方程组就写为

$$A\mathbf{x} = \mathbf{b}.$$

理想情形当然是方程组有解. 虽然可能有不止一组解, 但当采样点足够多时, 一般只会有唯一解, 因为规律应该唯一确定.

当然, 现实情形不会如此理想, 比如: 采样过程总有误差; 选择的函数形式不足以精确描述所需要的函数; 等等. 这些都会导致方程组无解.

定义残量  $\mathbf{r} := \mathbf{b} - A\mathbf{x}$ , 我们希望残量在某种意义上越小越好, 残量越小, 说明  $A\mathbf{x}$  接近  $\mathbf{b}$ , 或者说  $A\mathbf{x} = \mathbf{b}$  越接近于有解.

特别地, 我们选择残量的长度最小来求解  $\mathbf{x}$ , 即求解问题

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{r}\| = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|, \quad (3.3.2)$$

这一问题称为**最小二乘**问题. 这一最小二乘问题的解又常常称为线性方程组  $A\mathbf{x} = \mathbf{b}$  的**最小二乘解**.

根据命题 3.3.12, 我们知道  $\|\mathbf{b} - P_A \mathbf{b}\| = \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|$ . 由于  $P_A \mathbf{b} \in \mathcal{R}(A)$ , 因此一定存在  $\mathbf{x}_0 \in \mathbb{R}^n$  使得  $A\mathbf{x}_0 = P_A \mathbf{b}$ . 因此  $\|\mathbf{b} - A\mathbf{x}_0\| = \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|$ , 即  $\mathbf{x}_0$  是最小二乘问题 (3.3.2) 的一个解. 由此即知, 最小二乘问题总有解.

假设 (3.3.2) 有不只一个解. 任选另一个解  $\mathbf{x}_1$ , 那么  $\|\mathbf{b} - A\mathbf{x}_0\| = \|\mathbf{b} - A\mathbf{x}_1\| = \min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|$ . 根据正交投影的唯一性,  $A\mathbf{x}_0 = A\mathbf{x}_1 = P_A \mathbf{b}$ . 因此  $A(\mathbf{x}_1 - \mathbf{x}_0) = \mathbf{0}$ , 即  $\mathbf{x}_1 - \mathbf{x}_0 \in \mathcal{N}(A)$ . 因此最小二乘问题 (3.3.2) 的解集就是  $\{\mathbf{x}_0 + \mathbf{x}_1 : \mathbf{x}_1 \in \mathcal{N}(A)\}$ , 其中  $\mathbf{x}_0$  是一个特解.

结合对正交投影的讨论, 我们有定理 3.3.17.

**定理 3.3.17** 向量  $\mathbf{x}$  是线性方程组  $A\mathbf{x} = \mathbf{b}$  的最小二乘解, 当且仅当  $A^T A\mathbf{x} = A^T \mathbf{b}$ .

证. 如前所述,  $\mathbf{x}$  是  $A\mathbf{x} = \mathbf{b}$  的最小二乘解, 当且仅当  $A\mathbf{x} = P_A \mathbf{b}$ . 根据  $\mathbf{b} - P_A \mathbf{b} \perp \mathcal{R}(A)$  可得  $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$ , 即  $A^T A\mathbf{x} = A^T \mathbf{b}$ . 另一方面, 如果  $A^T A\mathbf{x} = A^T \mathbf{b}$ , 则  $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$ , 于是  $\mathbf{b} - A\mathbf{x} \in \mathcal{N}(A^T) = \mathcal{R}(A)^\perp$ . 注意到  $\mathbf{b} - P_A \mathbf{b} \in \mathcal{R}(A)^\perp$ , 就有  $P_A \mathbf{b} - A\mathbf{x} \in \mathcal{R}(A)^\perp$ . 但  $P_A \mathbf{b} - A\mathbf{x} \in \mathcal{R}(A)$ , 因此  $A\mathbf{x} - P_A \mathbf{b} = \mathbf{0}$ .  $\square$

事实上, 我们已经得到了最小二乘问题的一种解法, 即求解线性方程组  $A^T A\mathbf{x} = A^T \mathbf{b}$ . 这称为**正则化方法**. 正则化方法的基本步骤为:

1. 计算  $C = A^T A, \mathbf{d} = A^T \mathbf{b}$ ;
2. 计算  $C$  的 LDL<sup>T</sup> 分解  $C = LDL^T$ ;
3. 求解下三角方程组  $L\mathbf{z} = \mathbf{d}$ ;
4. 求解对角方程组  $D\mathbf{y} = \mathbf{z}$ ;
5. 求解上三角方程组  $L^T \mathbf{x} = \mathbf{y}$ .

我们还可以使用**正交化方法**, 这可以利用 QR 分解得到. 假设  $A$  的 QR 分解为  $A = QR = [Q_1 \ Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1$ , 令  $Q^T \mathbf{b} = \begin{bmatrix} Q_1^T \mathbf{b} \\ Q_2^T \mathbf{b} \end{bmatrix}$ , 那么

$$\|\mathbf{b} - A\mathbf{x}\|^2 = \|Q^T(\mathbf{b} - QR\mathbf{x})\|^2 = \left\| \begin{bmatrix} Q_1^T \mathbf{b} \\ Q_2^T \mathbf{b} \end{bmatrix} - \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{x} \right\|^2 = \|Q_1^T \mathbf{b} - R_1 \mathbf{x}\|^2 + \|Q_2^T \mathbf{b}\|^2.$$

这样,  $\mathbf{x}$  是最小二乘问题 (3.3.2) 的解当且仅当  $\mathbf{x}$  是线性方程组  $R_1 \mathbf{x} = Q_1^T \mathbf{b}$  的解. 而  $R_1$  是上三角矩阵, 这意味着这个线性方程组易于求解. 正交化方法的基本步骤为:

1. 计算  $A$  的简化 QR 分解  $A = Q_1 R_1$ ;
2. 计算  $\mathbf{c} = Q_1^T \mathbf{b}$ ;
3. 求解上三角方程组  $R_1 \mathbf{x} = \mathbf{c}$ .

最后, 我们看一个实际例子, 来体会最小二乘和数据拟合的关系.

**例 3.3.18** 我们用 CT 扫描张三来寻找他体内的肿瘤. 为了简化问题, 假设张三体内有四个同样大小的正方形区域, 区域边长是单位长度, 每个区域可能是骨、肉、血液、肿瘤之一. 已知数据是 X 射线穿过每个单位长度的骨头, 肉, 血液或者肿瘤时, 衰减程度为 1, 2, 3, 4. 我们用 X 射线穿过张三体内这些区域, 并通过测量射线穿过不同区域的衰减程度来判断每个区域的组成. 设不同区域内的单位长度的衰减程度分别为  $x_1, x_2, x_3, x_4$ . 五个不同方向的 X 射线穿过之后, 衰减程度分别为 3, 7, 7, 4, 6.1. 如图 3.3.3 所示. 得到

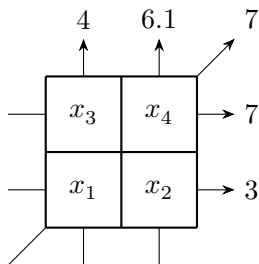


图 3.3.3: CT 扫描衰减

的数据列出方程组

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ \sqrt{2} & 0 & 0 & \sqrt{2} \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 7 \\ 4 \\ 6.1 \end{bmatrix}.$$

计算可得, 该方程无解, 这往往是实际观测出现的误差引起的. 利用最小二乘法, 可以解得  $x_1 \approx 0.950, x_2 \approx 2.075, x_3 \approx 3.025, x_4 \approx 4.000$ . 和已知数据对照, 可以得到四个区域内分别为骨、肉、血液、肿瘤.  $\odot$

## 习题

**练习 3.3.1** 设  $\mathcal{M}$  是下列齐次线性方程组的解空间:

$$\begin{cases} 2x_1 + x_2 + 3x_3 - x_4 = 0, \\ 3x_1 + 2x_2 - 2x_4 = 0, \\ 3x_1 + x_2 + 9x_3 - x_4 = 0, \end{cases}$$

分别求  $\mathcal{M}$  和  $\mathcal{M}^\perp$  的一组标准正交基.

**练习 3.3.2** 设  $\mathbb{R}^4$  中的向量  $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$  生成子空间  $\mathcal{M} = \text{span}(\mathbf{a}_1, \mathbf{a}_2)$ , 求  $\mathcal{M}^\perp$  的一组标准正交基.

**练习 3.3.3** 1. 设  $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 2 \end{bmatrix}, B = \begin{bmatrix} 5 & 4 \\ 6 & 3 \\ 5 & 1 \end{bmatrix}$ , 求两个矩阵列空间的交集中的一个非零向量; 由此判断两个列空间是否正交.

2. 求标准正交基  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ , 使得  $\mathbf{v}_1 \in \mathcal{R}(A) \cap \mathcal{R}(B)$ ,  $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{R}(A)$ , 且  $\mathbf{v}_1, \mathbf{v}_3 \in \mathcal{R}(B)$ .

3. 求一组向量  $\mathbf{x}, \mathbf{y}$ , 使得  $A\mathbf{x} = B\mathbf{y}$ , 并计算  $\begin{bmatrix} A & -B \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ .

4. 求  $\begin{bmatrix} A & -B \end{bmatrix}$  零空间的一组基, 并求所有的  $\mathbf{x}, \mathbf{y}$ , 满足  $A\mathbf{x} = B\mathbf{y}$ .

**练习 3.3.4** 1.  $\mathbb{R}^5$  中的两个三维子空间是否可能正交?

2. 设  $A = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 0 & \\ & & & 0 & 1 \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}$ , 则  $\mathcal{R}(A)$  和  $\mathcal{N}(A)$  是否正交? 是否互为正交补空间?

**练习 3.3.5** 设 6 阶方阵  $A$  满足  $A^3 = 0$ , 它的秩最大为多少? 举例说明. 在  $A, A^2$  的行空间、列空间、零空间和左零空间之中, 哪些相互正交?

**练习 3.3.6** 设  $\mathbb{R}^n$  的子空间  $\mathcal{M} = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_s)$ , 证明,  $\mathcal{M}^\perp = \{\mathbf{b} \in \mathbb{R}^n \mid \mathbf{b}^T \mathbf{a}_i = 0, 1 \leq i \leq s\}$ .

**练习 3.3.7** 对向量组  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ , 定义  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}^\perp$  为与这些向量都正交的向量所构成的子集.

1. 证明  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}^\perp$  是一个子空间.

- 构造矩阵  $A$ , 使得  $\mathcal{N}(A) = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}^\perp$ .
- 证明  $(\{\mathbf{v}_1, \dots, \mathbf{v}_k\}^\perp)^\perp = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ .

**练习 3.3.8** 集合运算有 De Morgan 定律: 对给定集合的两个子集  $X, Y$ ,  $X \cap Y$  的补集等于  $X$  的补集与  $Y$  的补集的并集;  $X \cup Y$  的补集等于  $X$  的补集与  $Y$  的补集的交集. 子空间是否也有类似的法则呢? 设  $\mathbb{R}^m$  的两个子空间  $\mathcal{M}, \mathcal{N}$ , 不妨设存在矩阵  $A, B$ , 使得  $\mathcal{M} = \mathcal{R}(A), \mathcal{N} = \mathcal{R}(B)$ .

- $\mathcal{M} + \mathcal{N}$  是哪个矩阵的列空间? 因此,  $(\mathcal{M} + \mathcal{N})^\perp$  是该矩阵的什么空间?
- $\mathcal{M}^\perp, \mathcal{N}^\perp, \mathcal{M}^\perp \cap \mathcal{N}^\perp$  分别是哪个矩阵的零空间?
- 证明 De Morgan 定律的子空间版本:  $(\mathcal{M} + \mathcal{N})^\perp = \mathcal{M}^\perp \cap \mathcal{N}^\perp, (\mathcal{M} \cap \mathcal{N})^\perp = \mathcal{M}^\perp + \mathcal{N}^\perp$ .

**练习 3.3.9** 设  $\mathbf{a}_1 = \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \mathcal{M} = \text{span}(\mathbf{a}_1, \mathbf{a}_2)$  是  $\mathbb{R}^3$  的子空间, 求  $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$  在  $\mathcal{M}$  上的正交投影.

**练习 3.3.10** 设  $\mathcal{L}$  是  $\mathbb{R}^3$  中的直线:

$$\mathcal{L}: \begin{cases} x_1 + x_2 + x_3 = 0, \\ 2x_1 - x_2 - 2x_3 = 0. \end{cases}$$

求  $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$  在直线  $\mathcal{L}$  上的正交投影.

**练习 3.3.11** 设  $\mathcal{M}$  是  $\mathbb{R}^3$  中由方程  $x_1 - x_2 + x_3$  决定的平面, 求  $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$  在平面  $\mathcal{M}$  上的正交投影, 并求  $\mathbf{b}$  到平面  $\mathcal{M}$  的距离.

**练习 3.3.12** 将下列问题中的  $\mathbf{x}$  分解成  $\mathcal{N}(A)$  与  $\mathcal{R}(A^T)$  中向量的和.

$$1. \mathbf{x} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, A = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \quad 2. \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad 3. \mathbf{x} = \begin{bmatrix} 5 \\ 5 \\ 9 \end{bmatrix}, A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 2 & 3 & 5 \end{bmatrix}.$$

**练习 3.3.13** 证明命题 3.3.12.

**练习 3.3.14** 说明满足下列条件的矩阵是否存在, 如果存在, 举例说明.

- $\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} \in \mathcal{R}(A), \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \mathcal{N}(A)$ .
- $\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} \in \mathcal{R}(A^T), \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \mathcal{N}(A)$ .

3.  $Ax = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$  有解, 且  $A$  的左零空间包含  $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ .
4.  $A$  不是零矩阵, 且  $A$  的每一行的转置垂直于  $A$  的每一列.
5.  $A$  非零, 且  $\mathcal{R}(A) = \mathcal{N}(A)$ .
6.  $A$  的所有列向量的和是零向量, 且所有行向量的和是分量均为 1 的向量.
7.  $A, B$  均为非零的正交投影矩阵, 且  $A + B$  仍是正交投影矩阵.
8.  $A, B$  均为正交投影矩阵, 但  $A + B$  并不是正交投影矩阵.
9.  $A, B, C$  均为非零的三阶正交投影矩阵, 且  $A + B + C = I_3$ .

**练习 3.3.15** 下列说法中, 哪些正确?

1.  $A$  的所有行的转置与  $A^{-1}$  的所有行的转置两两正交.
2.  $A$  的所有行的转置与  $A^{-1}$  的所有列两两正交.
3.  $A$  的所有列与  $A^{-1}$  的所有行的转置两两正交.
4.  $A$  的所有列与  $A^{-1}$  的所有列两两正交.
5. 如果向量  $v$  与  $w$  正交, 则  $v^T x = 0$  与  $w^T x = 0$  的解集互相正交.
6. 如果  $A$  是正交投影矩阵, 则  $A$  的第  $k$  列的长度的平方等于  $A$  的第  $k$  个对角元素.
7. 如果  $A, B$  是正交投影矩阵, 则  $AB = BA$  当且仅当  $AB$  也是正交投影矩阵.
8. 如果  $A, B$  是正交投影矩阵, 则  $A + B$  是正交投影矩阵当且仅当  $\mathcal{R}(A), \mathcal{R}(B)$  互相正交.

**练习 3.3.16** 如果矩阵  $A$  的列向量线性无关, 那么向  $\mathcal{R}(A)$  的正交投影矩阵为  $A(A^T A)^{-1} A^T$ . 试分析以下化简中可能出现的问题:  $A(A^T A)^{-1} A^T = AA^{-1}(A^T)^{-1} A^T = I_n I_n = I_n$ .

1. 证明  $(A^T A)^{-1} = A^{-1}(A^T)^{-1}$  并不一定总成立.
2. 当  $A$  满足什么条件时, 上式一定成立? 试分析此时正交投影矩阵等于  $I_n$  的原因.

**练习 3.3.17** 设  $A = \begin{bmatrix} 3 & 6 & 6 \\ 4 & 8 & 8 \end{bmatrix}$ .

1. 求向  $A$  的列空间的正交投影矩阵  $P_1$ .
2. 求向  $A$  的行空间的正交投影矩阵  $P_2$ .
3. 计算  $P_1 A P_2$ .

**练习 3.3.18** 给定  $A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}$ . 设  $P_1$  是关于  $A$  的第一列的正交投影矩阵,  $P_2$  是关于  $A$  的正交投影矩阵, 计算  $P_2 P_1$ .

**练习 3.3.19** 一个  $n$  阶方阵  $P$  是关于  $A$  的正交投影矩阵, 当且仅当对任意向量  $x \in \mathbb{R}^n$ , 都有  $Px \in \mathcal{R}(A), x - Px \in \mathcal{N}(A^T)$ .

**练习 3.3.20** 当  $A$  分别为对称矩阵、反对称矩阵、正交矩阵或上三角矩阵时, 判断  $\mathcal{R}(A)$  和  $\mathcal{N}(A)$  是否互为正交补. 证明或给出反例.

**练习 3.3.21** 给定  $\mathbb{R}^m$  中的子空间  $\mathcal{M}_1, \mathcal{M}_2$ ,  $\mathbb{R}^n$  中的子空间  $\mathcal{N}_1, \mathcal{N}_2$ .

1. 是否一定存在矩阵  $A$ , 使得  $\mathcal{R}(A) = \mathcal{M}_1, \mathcal{N}(A^T) = \mathcal{M}_2, \mathcal{R}(A^T) = \mathcal{N}_1, \mathcal{N}(A) = \mathcal{N}_2$ ?
2. 如果不一定存在, 那么当四个子空间满足什么条件时, 这样的矩阵才一定存在?

**练习 3.3.22** 给定向量  $\mathbf{b}_1, \dots, \mathbf{b}_m$ , 不难确知其平均值  $\mathbf{x}_m = \frac{1}{m} \sum_{i=1}^m \mathbf{b}_i$ . 再给出向量  $\mathbf{b}_{m+1}$ , 这  $m+1$  个向量的平均值就是  $\mathbf{x}_{m+1} = \frac{1}{m+1} \sum_{i=1}^{m+1} \mathbf{b}_i$ . 则  $\mathbf{x}_{m+1}$  可以用  $\mathbf{b}_{m+1}, \mathbf{x}_m, m$  来表示:  $\mathbf{x}_{m+1} = \mathbf{x}_m + \frac{1}{m+1} (\mathbf{b}_{m+1} - \mathbf{x}_m)$ .

若给出向量组成的矩阵  $B_n = [\mathbf{b}_{m+1} \quad \dots \quad \mathbf{b}_{m+n}]$ , 如何用  $B_n, \mathbf{x}_m, m, n$  来表示  $\mathbf{b}_1, \dots, \mathbf{b}_{m+n}$  这  $m+n$  个向量的平均值  $\mathbf{x}_{m+n}$ ?

**练习 3.3.23** 一个方阵如果仅仅满足  $P^2 = P$ , 则称之为斜投影矩阵, 其对应的线性变换称为斜投影. 给定一个  $n$  阶斜投影矩阵  $P$ .

1. 证明  $I_n - P$  也是  $n$  阶斜投影矩阵.
2. 证明  $\mathcal{R}(P) = \mathcal{N}(I_n - P), \mathcal{R}(I_n - P) = \mathcal{N}(P)$ .
3. 对任意向量  $\mathbf{v} \in \mathbb{R}^n$ , 是否一定存在分解  $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ , 满足  $\mathbf{v}_1 \in \mathcal{R}(P), \mathbf{v}_2 \in \mathcal{R}(I_n - P)$ ? 分解如果存在, 是否唯一?
4. 构造一个二阶斜投影矩阵, 但不是正交投影矩阵.

**练习 3.3.24** 设平面上的四个点  $\begin{bmatrix} x_i \\ y_i \end{bmatrix}$  分别是  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \end{bmatrix}, \begin{bmatrix} 4 \\ 20 \end{bmatrix}$ , 利用最小二乘法求下列直线或曲线:

1. 求平行于  $x$  轴的直线  $y = b$  使得  $\sum_{i=1}^4 |y_i - b|^2$  最小.
2. 求经过原点的直线  $y = kx$  使得  $\sum_{i=1}^4 |y_i - kx|^2$  最小.
3. 求直线  $y = kx + b$  使得  $\sum_{i=1}^4 |y_i - (kx_i + b)|^2$  最小.
4. 求抛物线  $y = ax^2 + bx + c$  使得  $\sum_{i=1}^4 |y_i - (ax_i^2 + bx_i + c)|^2$  最小.

Liang, Tian & Yang



## 第四章 行列式

### 4.1 引子

给定平面  $\mathbb{R}^2$  上的两个向量  $\mathbf{a}_1, \mathbf{a}_2$ , 考虑以二者为边的平行四边形的面积  $S(\mathbf{a}_1, \mathbf{a}_2)$ . 简单观察可知, 当一条边被拉伸若干倍时, 面积也会变成原面积的相应倍数, 即对任意  $k \geq 0$ ,  $S(k\mathbf{a}_1, \mathbf{a}_2) = kS(\mathbf{a}_1, \mathbf{a}_2)$ ; 当  $\mathbf{a}_1, \mathbf{a}'_1$  都在  $\mathbf{a}_2$  的同侧时,  $S(\mathbf{a}_1 + \mathbf{a}'_1, \mathbf{a}_2) = S(\mathbf{a}_1, \mathbf{a}_2) + S(\mathbf{a}'_1, \mathbf{a}_2)$ . 如图 4.1.1 所示. 那么, 面积  $S(\mathbf{a}_1, \mathbf{a}_2)$  关于两条边对应的向

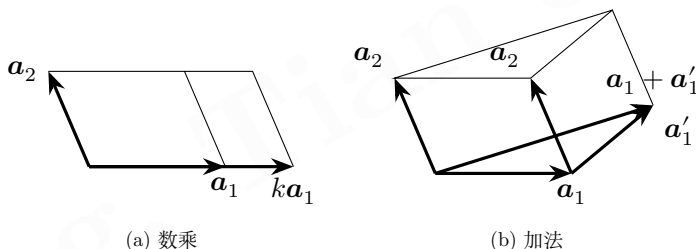


图 4.1.1: 向量与面积

量  $\mathbf{a}_1, \mathbf{a}_2$  满足线性关系吗? 显然并非如此, 例如,  $S(-k(-\mathbf{a}_1), \mathbf{a}_2) = kS(-\mathbf{a}_1, \mathbf{a}_2)$ ; 而  $S(\mathbf{a}_1 + \mathbf{a}'_1 + (-\mathbf{a}_1), \mathbf{a}_2) = S(\mathbf{a}_1 + \mathbf{a}'_1, \mathbf{a}_2) - S(-\mathbf{a}_1, \mathbf{a}_2)$ .

容易验证, 如果面积可以是负数, 线性关系就能满足了. 事实上, 带正负号的面积, 读者在中学学习定积分已经接触过它. 一个函数在某个区间上的定积分, 就是函数的曲线和  $x$  轴之间围成区域的面积的代数和, 其中区域在  $x$  轴上方的, 面积带正号; 区域在  $x$  轴下方的, 面积带负号. 如图 4.1.2 所示. 带符号的面积, 就是所谓有向面积. 对上述向量生成的平行四边形, 具体说来, 平行四边形在  $\mathbf{a}_1$  方向左侧时, 面积带正号; 在  $\mathbf{a}_1$  方向右侧时, 面积带负号.

为符合直观, 有向面积需要满足如下条件:

1. 双线性性:  $\delta(k\mathbf{a}_1 + k'\mathbf{a}'_1, \mathbf{a}_2) = k\delta(\mathbf{a}_1, \mathbf{a}_2) + k'\delta(\mathbf{a}'_1, \mathbf{a}_2)$ ,  
 $\delta(\mathbf{a}_1, k\mathbf{a}_2 + k'\mathbf{a}'_2) = k\delta(\mathbf{a}_1, \mathbf{a}_2) + k'\delta(\mathbf{a}_1, \mathbf{a}'_2)$ ;

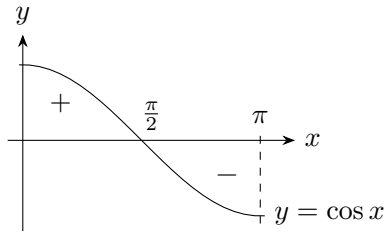


图 4.1.2: 有向面积

2. 反对称性:  $\delta(\mathbf{a}_1, \mathbf{a}_2) = -\delta(\mathbf{a}_2, \mathbf{a}_1)$ ;

3. 共线为零:  $\delta(\mathbf{a}, \mathbf{a}) = 0$ ;

4. 归一化条件:  $\delta(\mathbf{e}_1, \mathbf{e}_2) = 1$ .

事实上, 反对称性和共线为零互相等价.  $2 \Rightarrow 3$ : 令  $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}$ , 立得;  $3 \Leftarrow 2$ : 令  $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$ , 利用双线性性展开即得.

利用上述性质能得到有向面积  $\delta(\mathbf{a}_1, \mathbf{a}_2)$  的表达式. 设  $\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}$ ,  $\mathbf{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}$ , 则

$$\begin{aligned} \delta(\mathbf{a}_1, \mathbf{a}_2) &= \delta(a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2, a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2) \\ &= a_{11}a_{12}\delta(\mathbf{e}_1, \mathbf{e}_1) + a_{11}a_{22}\delta(\mathbf{e}_1, \mathbf{e}_2) + a_{21}a_{12}\delta(\mathbf{e}_2, \mathbf{e}_1) + a_{21}a_{22}\delta(\mathbf{e}_2, \mathbf{e}_2) \\ &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned}$$

这个计算过程还说明了有向面积被所需性质唯一确定.

有向面积  $\delta(\mathbf{a}_1, \mathbf{a}_2)$  还可以看作是矩阵  $A = [\mathbf{a}_1 \ \mathbf{a}_2]$  的函数, 记为  $\delta(A)$ . 则所需性质用矩阵语言描述就是:

1.  $\delta(AE_{ii;k}) = k\delta(A)$ ,  $\delta(AE_{ij;k}) = \delta(A)$ ;

2.  $\delta(AP_{12}) = -\delta(A)$ ;

3. 如果  $A$  不满秩, 则  $\delta(A) = 0$ ;

4.  $\delta(I_2) = 1$ .

注意, 前两条是对  $A$  做初等列变换.

倍加变换不改变有向面积, 因此, 如下计算

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \xrightarrow{AE_{21;k}} B = \begin{bmatrix} a_{11} - \frac{a_{21}}{a_{22}}a_{12} & a_{12} \\ 0 & a_{22} \end{bmatrix} \xrightarrow{BE_{12;k'}} C = \begin{bmatrix} a_{11} - \frac{a_{21}}{a_{22}}a_{12} & 0 \\ 0 & a_{22} \end{bmatrix},$$

满足  $\delta(A) = \delta(B) = \delta(C)$ . 根据双线性条件和归一化条件,  $\delta(C) = (a_{11} - \frac{a_{21}}{a_{22}}a_{12})a_{22}$ . 因此,  $\delta(A) = a_{11}a_{22} - a_{12}a_{21}$ . 计算过程如图 4.1.3 所示, 注意三个平行四边形等底等高的关系.

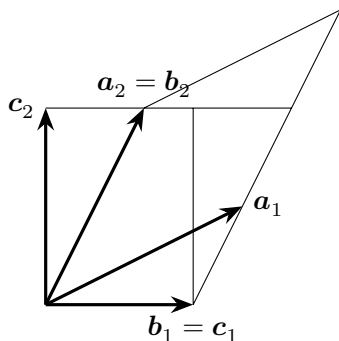


图 4.1.3: 列变换与面积

现在考虑有向面积在线性变换下会如何变化. 换言之, 给定矩阵  $A$ ,  $\delta(\mathbf{a}_1, \mathbf{a}_2)$  与  $\delta(A\mathbf{a}_1, A\mathbf{a}_2)$  之间有何关系?

先考虑初等矩阵对应的线性变换, 即错切变换、伸缩变换和对换变换. 通过设置坐标系, 可令  $\mathbf{a}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ ,  $\mathbf{a}_2 = \begin{bmatrix} x \\ y \end{bmatrix}$ . 下面来看三者对有向面积的影响, 如图 4.1.4 所示. 错切

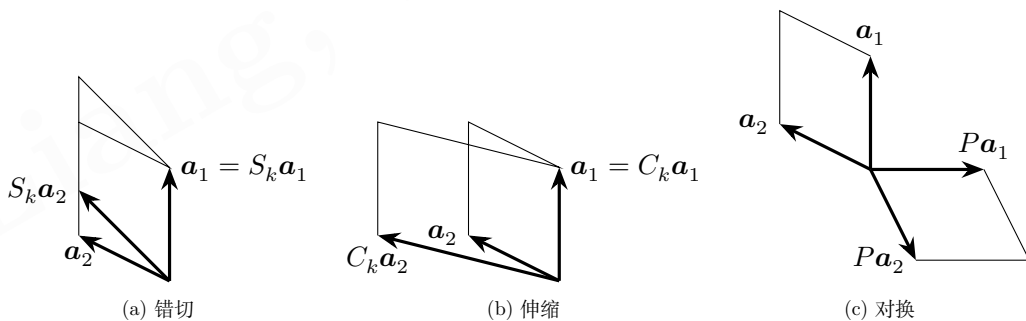


图 4.1.4: 变换与面积

变换  $S_k$  的表示矩阵是  $S_k = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$ , 显然  $\delta(S_k\mathbf{a}_1, S_k\mathbf{a}_2) = \delta(\mathbf{a}_1, \mathbf{a}_2)$ , 因为对应两个平行四边形等底等高. 伸缩变换  $C_k$  的表示矩阵是  $C_k = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$ , 显然  $\delta(C_k\mathbf{a}_1, C_k\mathbf{a}_2) = k\delta(\mathbf{a}_1, \mathbf{a}_2)$ , 因为对应两个平行四边形等底, 高变为  $k$  倍. 反射变换  $P$  的表示矩阵是

$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ , 显然  $\delta(P\mathbf{a}_1, P\mathbf{a}_2) = -\delta(\mathbf{a}_1, \mathbf{a}_2)$ , 因为对应两个平行四边形关于直线  $y = x$  对称, 但面积的符号发生了改变. 另一轴上的错切变换和伸缩变换, 可以通过上述变换的复合得到. 而任意可逆矩阵  $A$  都可以分解成初等矩阵的乘积, 因此任意可逆线性变换都分解为上述变换的复合, 而有向面积也随之变化了一定倍数.

为了说明所变倍数被线性变换唯一确定, 进行如下计算. 对方阵  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$  和向量  $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ ,  $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ , 有

$$\begin{aligned} \delta(A\mathbf{x}, A\mathbf{y}) &= \delta(x_1\mathbf{a}_1 + x_2\mathbf{a}_2, y_1\mathbf{a}_1 + y_2\mathbf{a}_2) \\ &= x_1y_1\delta(\mathbf{a}_1, \mathbf{a}_1) + x_1y_2\delta(\mathbf{a}_1, \mathbf{a}_2) + x_2y_1\delta(\mathbf{a}_2, \mathbf{a}_1) + x_2y_2\delta(\mathbf{a}_2, \mathbf{a}_2) \\ &= (x_1y_2 - x_2y_1)\delta(\mathbf{a}_1, \mathbf{a}_2) \\ &= \delta(\mathbf{x}, \mathbf{y})\delta(\mathbf{a}_1, \mathbf{a}_2) \\ &= \delta(A)\delta(\mathbf{x}, \mathbf{y}). \end{aligned}$$

因此,  $\mathbb{R}^2$  上任意线性变换, 如果其表示矩阵为  $A$ , 则它将任意平行四边形的有向面积变为原来的  $\delta(A)$  倍, 该数称之为线性变换的**变积系数**.

**注 4.1.1** 1. 事实上, 任意图形的有向面积都变为原来的  $\delta(A)$  倍.

2. 有时我们只需要面积的大小而不需要面积的符号, 就只需考虑上述有向面积的绝对值. 这时, 面积是  $\delta$  的绝对值, 对应的变积系数也是  $\delta$  的绝对值.

以上讨论刻画了有向面积  $\delta$  的一些性质:

1. 对任意二阶矩阵  $A$ ,  $\delta(A)$  既是以  $A$  的列向量为边的平行四边形的有向面积, 也是  $A$  表示的线性变换的变积系数.
2.  $\delta$  具有多线性性、反对称性, 且满足归一化条件. 共线为零条件也满足.
3.  $\delta(AB) = \delta(A)\delta(B)$ , 这说明线性变换的复合的变积系数是线性变换的变积系数的乘积.

接下来, 本章会把有向面积和变积系数推广到  $\mathbb{R}^n$ .

## 4.2 行列式函数

我们可以把方阵的函数等价地定义为它的列向量组的函数, 为简单起见, 我们用相同的记号表示这两个定义域不同但值相同的函数, 即  $\delta(\mathbf{a}_1, \dots, \mathbf{a}_n) := \delta \left( \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \right)$ ,

$\mathbf{a}_i \in \mathbb{R}^n$ .

**定义 4.2.1 (行列式)** 定义在全体  $n$  阶方阵上的函数  $\delta$ , 如果满足如下性质:

1. 列多线性性: 对每个列向量都满足线性性, 即对任意  $1 \leq i \leq n$ , 都有  $\delta(\cdots, k\mathbf{a}_i + k'\mathbf{a}'_i, \cdots) = k\delta(\cdots, \mathbf{a}_i, \cdots) + k'\delta(\cdots, \mathbf{a}'_i, \cdots)$ ;
2. 列反对称性: 对任意  $1 \leq i < j \leq n$ , 都有  $\delta(\cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j, \cdots) = -\delta(\cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots)$ ;
3. 单位化条件:  $\delta(I_n) = 1$ ;

则  $\delta$  就称为一个  $n$  阶**行列式函数**.

我们将证明  $n$  阶方阵的行列式函数存在且唯一. 这个唯一的行列式函数在矩阵  $A$  的值称为  $A$  的**行列式**, 记为<sup>1</sup> $\det(A)$  或  $|A|$ .

某个  $n$  阶方阵的行列式可以直接称为一个  $n$  阶行列式.

行列式函数的存在性需要具体的构造才能证明, 见第 4.3 节.

**例 4.2.2** 根据定义, 1 阶行列式函数就是  $\mathbb{R}$  上的一个线性变换, 满足在 1 处的取值是

1. 因此, 1 阶矩阵  $A = [a_{11}]$  的行列式定义为  $\det(A) = a_{11}$ .

对 2 阶矩阵  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ , 容易验证  $f(A) = a_{11}a_{22} - a_{12}a_{21}$  是一个行列式函数. 由行列式函数的唯一性, 即有

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

容易看到,  $\det(A) \neq 0$  当且仅当  $A$  可逆.

☺

**命题 4.2.3** 1. 如果方阵  $A$  有两列相等, 则  $\det(A) = 0$ ;

2. 如果方阵  $A$  不满秩, 即不可逆, 则  $\det(A) = 0$ ;

3. 如果方阵  $A$  有一列为零或有一行为零, 则  $\det(A) = 0$ .

证. 第 1 条: 利用列反对称性, 令  $\mathbf{a}_i = \mathbf{a}_j$  即得.

第 2 条: 若  $A$  不满秩, 则有至少一列可以写成其他列的线性组合, 不妨设  $\mathbf{a}_n = k_1\mathbf{a}_1 + \cdots + k_{n-1}\mathbf{a}_{n-1}$ . 根据列线性性,  $\delta(\mathbf{a}_1, \cdots, \mathbf{a}_{n-1}, \mathbf{a}_n) = k_1\delta(\mathbf{a}_1, \cdots, \mathbf{a}_{n-1}, \mathbf{a}_1) + \cdots + k_{n-1}\delta(\mathbf{a}_1, \cdots, \mathbf{a}_{n-1}, \mathbf{a}_{n-1})$ . 由第 1 条立得  $\det(A) = 0 + \cdots + 0 = 0$ .

第 3 条: 由第 2 条, 显然.

□

<sup>1</sup>在可能与绝对值的记号混淆时, 应注意避免或提前说明. 为解决这一问题, 我们只在  $n$  行  $n$  列记号 ( $n > 1$ ) 组成的矩阵上使用  $|\cdot|$  这种记号.

由此即知, 行列式函数只有在可逆矩阵上的值才可能不是零. 根据定理 1.6.7, 可逆矩阵可以写成有限多个初等矩阵的乘积. 初等矩阵的行列式由定义容易得到.

**命题 4.2.4** 行列式函数在初等矩阵上的取值均不为零, 分别是:

1.  $\det(P_{ij}) = -1$ ;
2.  $\det(E_{ii;k}) = k$ ;
3.  $\det(E_{ji;k}) = 1$ .

证. 三类初等矩阵都可以从单位矩阵  $I_n$  做一次初等列变换得到.

第 1 条:  $P_{ij}$  对调了  $I_n$  的第  $i, j$  列, 由列反对称性可知.

第 2 条:  $E_{ii;k}$  把  $I_n$  的第  $i$  列乘以  $k$ , 由列线性性可知.

第 3 条:  $E_{ji;k}$  把  $I_n$  的第  $j$  列的  $k$  倍加到第  $i$  列, 因此  $\det(E_{ji;k}) = \delta(\cdots, \mathbf{e}_i + k\mathbf{e}_j, \cdots, \mathbf{e}_j, \cdots) = \delta(\cdots, \mathbf{e}_i, \cdots, \mathbf{e}_j, \cdots) + k\delta(\cdots, \mathbf{e}_j, \cdots, \mathbf{e}_j, \cdots) = \det(I_n) = 1$ .  $\square$

矩阵的行列式函数在初等列变换下的变化规律可以类似地推得, 证明留给读者.

**命题 4.2.5** 行列式函数满足

1. 对  $A$  的第  $i, j$  列位置互换得到  $B = AP_{ij}$ , 则  $\det(B) = \det(AP_{ij}) = -\det(A)$ ;
2. 对  $A$  的第  $i$  列乘非零常数  $k$  得到  $B = AE_{ii;k}$ , 则  $\det(B) = \det(AE_{ii;k}) = k\det(A)$ ;
3. 把  $A$  的第  $j$  列的  $k$  倍加到第  $i$  列上得到  $B = AE_{ji;k}$ , 则  $\det(B) = \det(AE_{ji;k}) = \det(A)$ .

换言之, 对换变换改变行列式的符号, 倍乘变换把行列式乘以相同的倍数, 而倍加变换不改变行列式. 几何直观上, 这对应着: 反射变换不改变有向体积大小, 但改变符号; 参数为  $k$  的伸缩变换将有向体积改变  $k$  倍; 错切变换不改变有向体积.

注意, 当  $n > 1$  时, 行列式函数不是线性函数. 对  $n$  阶方阵  $A$ ,  $\det(kA) = k^n \det(A)$ , 这与行列式和体积的关系相一致.

另一方面, 我们有如下重要结论.

**定理 4.2.6** 行列式函数有如下性质:

1. 对初等矩阵  $E$ , 则  $\det(AE) = \det(A)\det(E)$ ;
2. 设可逆矩阵  $A = E_1 \cdots E_m$ , 其中  $E_i$  为初等矩阵, 则  $\det(A) = \det(E_1) \cdots \det(E_m)$ ;
3.  $\det(A) \neq 0$  当且仅当  $A$  可逆;

$$4. \det(AB) = \det(A) \det(B);$$

$$5. \det(A^T) = \det(A).$$

证. 第 1 条: 由命题 4.2.4 和 4.2.5 立得.

第 2 条: 反复运用第 1 条可得.

第 3 条: 如果  $A$  可逆, 则由第 2 条和初等矩阵行列式不为零可知  $\det(A) \neq 0$ ; 如果  $A$  不可逆, 根据命题 4.2.3,  $\det(A) = 0$ .

第 4 条: 如果  $B$  可逆, 令  $B = E_1 \cdots E_m$ , 则  $\det(AB) = \det(AE_1 \cdots E_m) = \det(AE_1 \cdots E_{m-1}) \det(E_m) = \cdots = \det(A) \det(E_1) \cdots \det(E_m) = \det(A) \det(B)$ ; 如果  $B$  不可逆, 则  $B$  不满秩, 由  $\text{rank}(AB) \leq \text{rank}(B)$  可知  $AB$  也不满秩, 因此  $\det(AB) = \det(B) = 0$ ,  $\det(AB) = \det(A) \det(B)$  也成立.

第 5 条: 如果  $A$  可逆, 令  $A = E_1 \cdots E_m$ , 则  $A^T = E_m^T \cdots E_1^T$ , 由初等矩阵和它的转置具有相同的行列式可知  $\det(A^T) = \det(A)$ ; 如果  $A$  不可逆, 则  $A^T$  也不可逆, 因此  $\det(A^T) = \det(A) = 0$ .  $\square$

定理 4.2.6 中, 第 3 条是矩阵可逆的等价描述.

第 4 条说明,  $\det(AB) = \det(A) \det(B) = \det(B) \det(A) = \det(BA)$ , 即尽管矩阵  $AB$  和  $BA$  并不总相等, 但二者行列式相等.

第 5 条说明对调矩阵的行和列不改变行列式的值, 由此立得如下结论.

**命题 4.2.7** 行列式也满足行线性和行反对称性.

容易看出, 我们也可以用行线性和行反对称性来定义行列式.

**命题 4.2.8** 满足行线性, 行反对称性和单位化条件的关于方阵的函数是行列式函数.

**例 4.2.9** 下面计算一些具有特殊形式的矩阵的行列式.

1. 对角矩阵: 其行列式是对角元素的乘积. 这可由列线性性推得.

在二维 (三维) 情形的几何直观上, 对角矩阵对应了长方形 (长方体), 面积或体积就是各边长乘积.

2. 上三角矩阵: 其行列式是对角元素的乘积. 如果有对角元素为零, 则矩阵不可逆, 行列式为零; 否则, 可以通过倍加变换消去非对角元, 得到对角矩阵, 而它的行列式就是对角元素的乘积, 而倍加变换不改变行列式, 因此上三角矩阵的行列式也是对角元素的乘积. 类似结论对下三角矩阵也成立.

在二维 (三维) 情形的几何直观上, 上三角矩阵的对角元素对应了底边长、底边上的高 (以及平行六面体的高), 而乘积就是面积 (体积). 事实上倍加变换, 就是一

系列等底等高的错切, 把原来的平行四边形 (平行六面体) 化成了等面积 (体积) 的长方形 (长方体), 可以参考图 4.1.3.

3. 分块对角矩阵,  $X = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$ , 其中  $A, B$  是方阵:  $\det(X) = \det(A) \det(B)$ .

4. 分块上三角矩阵  $X = \begin{bmatrix} A & C \\ O & B \end{bmatrix}$ , 其中  $A, B$  是方阵:  $\det(X) = \det(A) \det(B)$ .

5. 正交矩阵: 行列式是 1 或  $-1$ . 记正交矩阵为  $Q$ , 由  $Q^T Q = I_n$  可得,  $\det(Q)^2 = \det(Q^T) \det(Q) = \det(I_n) = 1$ , 立得  $\det(Q) = \pm 1$ .

在二维情形的几何直观上, 正交矩阵只能是旋转或反射 (例 3.2.5). 旋转不改变有向面积, 反射只改变有向面积的符号. 因此旋转的表示矩阵  $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$  的

行列式是 1, 反射的表示矩阵  $\begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}$  的行列式是  $-1$ .

高维情形下, Givens 变换和 Householder 变换的行列式, 请读者思考.

6. 可逆矩阵的逆: 其行列式是原矩阵的倒数. 记原矩阵为  $P$ , 则  $\det(P^{-1}) \det(P) = \det(P^{-1}P) = \det(I_n) = 1$ , 因此  $\det(P^{-1}) = \frac{1}{\det(P)}$ . 特别地,  $\det(P^{-1}AP) = \det(A)$ .

☺

根据定理 4.2.6, 可以证明行列式函数的唯一性.

**命题 4.2.10** 定义在  $n$  阶方阵上的行列式函数如果存在, 则唯一.

证. 假设有两个行列式函数. 对任意不满秩矩阵, 两个行列式都为 0; 对单位矩阵, 两个行列式都为 1; 对初等矩阵, 两个行列式的值都相等; 对满秩矩阵, 根据定理 1.6.7, 它可以写成初等矩阵的乘积, 于是它的两个行列式的值也相等, 都为对应的初等矩阵行列式的乘积. 因此, 对任意矩阵, 两个行列式函数的值都相等, 这意味着两个函数相等. 即行列式函数如果存在, 则唯一.  $\square$

命题 4.2.10 容易推得如下结论, 证明留给读者.

**命题 4.2.11** 定义在  $n$  阶方阵上的函数  $\delta$ , 如果满足如下性质:

1.  $\delta(AB) = \delta(A)\delta(B)$ ;
2. 如果  $A$  不可逆, 那么  $\delta(A) = 0$ ;
3. 如果  $A = P_{ij}$  是对换矩阵, 那么  $\delta(A) = -1$ ;



4. 如果  $A = E_{ii;k}$  是参数为  $k \neq 0$  的倍乘矩阵, 那么  $\delta(A) = k$ ;
5. 如果  $A = E_{ji;k}$  是倍加矩阵, 那么  $\delta(A) = 1$ ;

则  $\delta$  就是行列式函数.

事实上命题 4.2.11 暗示了行列式函数的一个等价定义.

注意, 在二维情形的几何直观上, 如果规定变积系数为  $\det(A)$ , 则可将命题 4.2.11 和线性变换的变积系数联系起来:

1. 第 1 条是变积系数应有的性质, 即两个线性变换复合时, 变积系数应该相乘;
2. 第 2 条意味着, 如果存在向量  $\mathbf{x}$  使得  $A\mathbf{x} = 0$ , 那么变积系数为 0, 这符合直观, 因为这个线性变换把某个方向的厚度变为 0;
3. 第 3 条, 注意到  $A = P_{ij} = I - 2\left(\frac{\mathbf{e}_i - \mathbf{e}_j}{\sqrt{2}}\right)\left(\frac{\mathbf{e}_i - \mathbf{e}_j}{\sqrt{2}}\right)^T$ , 这对应了反射变换, 变积系数为 -1;
4. 第 4 条, 这对应了沿  $\mathbf{e}_i$  方向、伸缩系数为  $k$  的伸缩变换, 变积系数为伸缩系数;
5. 第 5 条, 这对应了将  $\mathbf{e}_j$  沿着  $\mathbf{e}_i$  的方向、错切系数为  $k$  的错切变换, 变积系数为 1;
6. 这五条都是变积系数应有的性质, 而命题 4.2.11 说明, 只要一个函数满足这五条性质, 它就一定是  $\det(A)$ , 换言之, 这五条决定了变积系数是  $\det(A)$ , 而行列式函数的唯一性又确定了变积系数的唯一性.

行列式函数, 作为列向量组的函数, 定义 4.2.1 可以类比第 4.1 节中平行四边形的有向面积; 作为矩阵的函数, 命题 4.2.11 可以类比第 4.1 节中线性变换的变积系数. 因此, 我们可以利用行列式来定义高维空间中平行体的有向体积, 以及高维空间中线性变换的变积系数, 这将在多元微积分中得到应用.

对任意方阵  $A$ , 考虑 QR 分解  $A = QR$ . 由于  $Q$  正交,  $\det(Q) = \pm 1$ . 而  $R$  是对角元皆为非负数的上三角矩阵, 故  $\det(R) \geq 0$ . 而  $\det(A) = \det(Q)\det(R)$ , 其中  $\det(R)$  是有向面积 (体积) 的绝对值, 即传统意义上的面积 (体积);  $\det(Q)$  代表有向面积 (体积)  $\det(A)$  的符号, 即所谓“向”<sup>2</sup>.

命题 4.2.11 给出了计算行列式的实用方法, 即通过初等行列变换来化简. 如果  $E$  是初等矩阵, 那么  $EA$  是对  $A$  施加了初等行变换,  $AE$  是对  $A$  施加了初等列变换. 利用  $\det(EA) = \det(AE) = \det(A)\det(E)$ , 可得如下计算规律:

1. 把  $A$  的某行的倍数加到另一行, 或某列的倍数加到另一列, 其行列式不变;

<sup>2</sup>这可以简单推出, 无论空间维数如何, “向”只可能有两种. 事实上, 向与空间  $\mathbb{R}^n$  的连续性有关.

2. 把  $A$  的两行或两列对调, 其行列式变为原来的相反数;
3. 把  $A$  的某行或某列乘以  $k$ , 其行列式变为原来的  $k$  倍.

下面来看几个具体例子.

**例 4.2.12** 计算  $\begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 7 \\ 3 & 7 & 10 \end{vmatrix}$ . 首先, 将第一列的  $-2$  倍和  $-3$  倍分别加到第二列和第三

列上, 得  $\begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 7 \\ 3 & 7 & 10 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 3 & 1 & 1 \end{vmatrix}$ . 然后, 将第一行的  $-2$  倍和  $-3$  倍分别加到第二行和

第三行上, 得  $\begin{vmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 3 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix}$ . 交换第二行和第三行, 就得到上三角矩阵, 行列式

立得. 计算过程如下:

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 7 \\ 3 & 7 & 10 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 3 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = - \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} = -1.$$

计算过程类似于 PLU 分解. 事实上, PLU 分解和初等行列变换等价, 因此也可以直接计算矩阵的 PLU 分解, 再计算  $P, U$  的行列式的乘积即可 (为什么?). ☺

**例 4.2.13** 设多项式  $p(x) = k_3x^3 + k_2x^2 + k_1x + k_0$  经过四个不同的点  $(\lambda_i, \mu_i), i = 1, 2, 3, 4$ . 这样的多项式存在吗? 如果存在, 有几个呢?

读者在中学已经学过两点确定唯一直线, 三点唯一确定抛物线. 可以猜测四点唯一确定三次多项式. 下面来证明猜测. 根据条件列出方程组:

$$\begin{cases} k_3\lambda_1^3 + k_2\lambda_1^2 + k_1\lambda_1 + k_0 = \mu_1, \\ k_3\lambda_2^3 + k_2\lambda_2^2 + k_1\lambda_2 + k_0 = \mu_2, \\ k_3\lambda_3^3 + k_2\lambda_3^2 + k_1\lambda_3 + k_0 = \mu_3, \\ k_3\lambda_4^3 + k_2\lambda_4^2 + k_1\lambda_4 + k_0 = \mu_4, \end{cases} \quad \text{或} \quad \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 \\ 1 & \lambda_2 & \lambda_2^2 & \lambda_2^3 \\ 1 & \lambda_3 & \lambda_3^2 & \lambda_3^3 \\ 1 & \lambda_4 & \lambda_4^2 & \lambda_4^3 \end{bmatrix} \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}.$$

下面来说明系数矩阵  $A$  可逆. 根据定理 4.2.6, 只需说明  $\det(A) \neq 0$ .

计算该行列式, 首先自右而左把每一列的  $-\lambda_1$  倍加到右边相邻列, 则有

$$\begin{vmatrix} 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 \\ 1 & \lambda_2 & \lambda_2^2 & \lambda_2^3 \\ 1 & \lambda_3 & \lambda_3^2 & \lambda_3^3 \\ 1 & \lambda_4 & \lambda_4^2 & \lambda_4^3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & \lambda_2 - \lambda_1 & \lambda_2^2 - \lambda_2\lambda_1 & \lambda_2^3 - \lambda_2^2\lambda_1 \\ 1 & \lambda_3 - \lambda_1 & \lambda_3^2 - \lambda_3\lambda_1 & \lambda_3^3 - \lambda_3^2\lambda_1 \\ 1 & \lambda_4 - \lambda_1 & \lambda_4^2 - \lambda_4\lambda_1 & \lambda_4^3 - \lambda_4^2\lambda_1 \end{vmatrix}$$

$$\begin{aligned}
&= \begin{vmatrix} \lambda_2 - \lambda_1 & \lambda_2^2 - \lambda_2\lambda_1 & \lambda_2^3 - \lambda_2^2\lambda_1 \\ \lambda_3 - \lambda_1 & \lambda_3^2 - \lambda_3\lambda_1 & \lambda_3^3 - \lambda_3^2\lambda_1 \\ \lambda_4 - \lambda_1 & \lambda_4^2 - \lambda_4\lambda_1 & \lambda_4^3 - \lambda_4^2\lambda_1 \end{vmatrix} \\
&= (\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_4 - \lambda_1) \begin{vmatrix} 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \\ 1 & \lambda_4 & \lambda_4^2 \end{vmatrix}
\end{aligned}$$

类似计算即得

$$= (\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_4 - \lambda_1)(\lambda_3 - \lambda_2)(\lambda_4 - \lambda_2)(\lambda_4 - \lambda_3).$$

由于  $\lambda_i$  两两不同, 因此  $\det(A) \neq 0$ , 系数矩阵  $A$  可逆, 方程组有唯一解. 这说明四点唯一确定三次多项式.

给定平面上若干点, 寻找一条经过这些点的曲线, 这个模型称为**插值**, 在各种领域都有广泛应用.

事实上, 参考上面的计算, 利用数学归纳法, 可以证明:

$$\begin{vmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{vmatrix} = \prod_{1 \leq j < i \leq n} (\lambda_i - \lambda_j),$$

其中  $\prod$  是连乘号, 它的上下提示了变动角标的变换范围, 而连乘式表示变动下角标取遍变化范围的对应所有项的积, 即

$$\prod_{1 \leq j < i \leq n} (\lambda_i - \lambda_j) = (\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1) \cdots (\lambda_n - \lambda_1)(\lambda_3 - \lambda_2) \cdots (\lambda_n - \lambda_2) \cdots \cdots (\lambda_n - \lambda_{n-1}).$$

这类矩阵被称为 **Vandermonde 矩阵**. 当  $\lambda_i$  两两不同时, 其行列式不为零, Vandermonde 矩阵可逆. 这说明, 给定平面上  $n+1$  个横坐标两两不同的点, 可以唯一确定  $n$  次多项式. ☺

## 习题

**练习 4.2.1** 计算下列行列式.

$$1. \begin{vmatrix} 1 & 2 & 1 & 4 \\ 0 & -1 & 2 & 1 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 3 \end{vmatrix}.$$

$$2. \begin{vmatrix} 1 & -1 & 2 \\ 3 & 2 & 1 \\ 0 & 1 & 4 \end{vmatrix}.$$

$$3. \begin{vmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{vmatrix}.$$

$$4. \begin{vmatrix} x & y & x+y \\ y & x+y & x \\ x+y & x & y \end{vmatrix}, \quad 5. \begin{vmatrix} 1+x & 1 & 1 & 1 \\ 1 & 1-x & 1 & 1 \\ 1 & 1 & 1+y & 1 \\ 1 & 1 & 1 & 1-y \end{vmatrix}.$$

**练习 4.2.2** 设  $A$  是三阶方阵,  $\det(A) = 5$ , 求下列矩阵  $B$  的行列式.

1.  $B = 2A, -A, A^2, A^{-1}$ .

2.  $B = \begin{bmatrix} \mathbf{a}_1^T - \mathbf{a}_3^T \\ \mathbf{a}_2^T - \mathbf{a}_1^T \\ \mathbf{a}_3^T - \mathbf{a}_2^T \end{bmatrix}, \begin{bmatrix} \mathbf{a}_1^T + \mathbf{a}_3^T \\ \mathbf{a}_2^T + \mathbf{a}_1^T \\ \mathbf{a}_3^T + \mathbf{a}_2^T \end{bmatrix}$ , 其中  $A = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \mathbf{a}_3^T \end{bmatrix}$ .

**练习 4.2.3** 设  $A_n = \begin{bmatrix} -1 & 1 \\ -6 & 4 \end{bmatrix} + nI_2$ .

1. 求  $A_0, A_1, A_2, A_3$  的行列式.

2. 求  $\begin{bmatrix} -1 & 1 \\ -6 & 4 \end{bmatrix} + xI_2$  的行列式, 并将其写成  $(x+a)(x+b)$  的形式.

3. 分别求  $A_0^2, A_0^2 + I_2, A_0^2 + 3A_0 + 2I_2, A_0^3 - 2A_0^2 + 3A_0 - 4I_2$  的行列式, 并分析它们与  $a, b$  的关系.

**练习 4.2.4** 计算  $\det(A)$ .

1.  $A = [i+j]_{n \times n}$ .

2.  $A = [ij]_{n \times n}$ .

**练习 4.2.5** 计算  $\begin{vmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \\ -1 & & & 1 \end{vmatrix}$ .

**练习 4.2.6** 计算  $\begin{vmatrix} 1+x_1y_1 & 1+x_1y_2 & \cdots & 1+x_1y_n \\ 1+x_2y_1 & 1+x_2y_2 & \cdots & 1+x_2y_n \\ \vdots & \vdots & & \vdots \\ 1+x_ny_1 & 1+x_ny_2 & \cdots & 1+x_ny_n \end{vmatrix}$ .

**练习 4.2.7** 计算  $\begin{vmatrix} a_{11} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & \cdots & a_{2,n-1} & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & 0 & \cdots & 0 \end{vmatrix}$ .

**练习 4.2.8** 1. 令  $A_n$  是从右上到左下对角线上的元素全为 1, 其余元素全为 0 的  $n$  阶方阵. 求  $A_2, A_3, A_4, A_5$  的行列式, 分析其规律, 推断出  $A_n$  的行列式.

2. 令  $A_2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$ , 以此类推  $A_n$ . 求  $A_2, A_3, A_4$  的行列式, 分析其规律, 推

断出  $A_n$  的行列式. 提示: 利用  $LU$  分解.

3. 设  $A$  具有 QR 分解  $A = Q \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ , 求  $\det(A)$  的所有可能值.

4. 定义 Hilbert 矩阵  $H_n = \left[ \frac{1}{i+j-1} \right]_{n \times n}$ . 计算  $H_2, H_3$ .

Hilbert 矩阵是一种常见的难于计算的矩阵, 常用来测试算法.

#### 练习 4.2.9 证明或举出反例.

- $AB - BA$  的行列式必然是零.
- $A$  的行列式等于其行简化阶梯形的行列式.
- $A$  为  $n$  阶反对称矩阵, 当  $n$  为奇数时,  $\det(A) = 0$ .
- $A$  为  $n$  阶反对称矩阵, 当  $n$  为偶数时,  $\det(A) = 0$ .
- 如果  $|\det(A)| > 1$ , 那么当  $n$  趋于无穷时,  $A^n$  中必然有元素的绝对值趋于无穷.
- 如果  $|\det(A)| < 1$ , 那么当  $n$  趋于无穷时,  $A^n$  中的所有元素都趋于 0.

练习 4.2.10 1. 对二阶可逆矩阵  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , 计算其逆矩阵的行列式:

$$\det(A^{-1}) = \det\left(\frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\right) = \frac{ad-bc}{ad-bc} = 1.$$

- 对分块矩阵  $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ , 计算其行列式:  $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = AD - BC$ , 得到的是矩阵, 而不是数. 如果  $A$  可逆, 正确的公式是什么?
- 计算正交投影矩阵  $P$  的行列式  $\det(P) = \det(A(A^T A)^{-1} A^T) = \frac{\det(A) \det(A^T)}{\det(A^T A)} = 1$ , 然而正交投影矩阵常常不可逆. 错在哪里?
- 如果  $AB = -BA$ , 那么  $\det(A) \det(B) = -\det(B) \det(A)$ , 由此得到  $2 \det(A) \det(B) = 0$ , 所以  $A, B$  必有一个矩阵不可逆. 这是否正确? 如果不是, 请指出错误并举出反例.
- 对矩阵  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , 同时做两个行变换得到  $\begin{bmatrix} a+sc & b+sd \\ c+ta & d+tb \end{bmatrix}$ , 行列式是否一定保持不变?  $s, t$  满足什么条件时, 行列式一定保持不变?

练习 4.2.11 设  $n$  阶方阵  $A$  的对角元素全为 0, 其他元素全为 1, 令  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ .

- 求向量  $\mathbf{u}$ , 使得  $\mathbf{a}_i$  可以写成  $\mathbf{e}_i$  和  $\mathbf{u}$  的线性组合.
- 利用列多线性性质, 求  $\det(A)$ .

练习 4.2.12 证明命题 4.2.5.

练习 4.2.13 证明命题 4.2.11.

练习 4.2.14 证明奇数阶反对称矩阵不可逆.

**练习 4.2.15** 证明任意可逆矩阵  $A$  都可以只用倍加变换化为  $\text{diag}(1, 1, \dots, \det(A))$ .

**练习 4.2.16** 给定  $A = [a_{ij}]_{n \times n}$ , 而  $B = [a_{ij}c^{i-j}]_{n \times n}$ , 其中  $c \neq 0$ , 证明,  $\det(A) = \det(B)$ .

**练习 4.2.17** 给定  $n-1$  个互不相同的数  $a_1, \dots, a_{n-1}$ , 令

$$P(x) = \begin{vmatrix} 1 & x & \cdots & x^{n-1} \\ 1 & a_1 & \cdots & a_1^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_{n-1} & \cdots & a_{n-1}^{n-1} \end{vmatrix}.$$

证明  $P(x)$  是一个关于  $x$  的  $n-1$  次多项式, 并求  $P(x)$  的  $n-1$  个根.

**练习 4.2.18** 设  $f_i(x)$  是  $i$  次多项式,  $0 \leq i \leq n-1$ , 其首项系数是  $a_i$ . 又设  $b_0, b_1, \dots, b_{n-1}$  是  $n$  个数, 计算下列  $n$  阶行列式:

$$\begin{vmatrix} f_0(b_0) & f_0(b_1) & \cdots & f_0(b_{n-1}) \\ f_1(b_0) & f_1(b_1) & \cdots & f_1(b_{n-1}) \\ \vdots & \vdots & & \vdots \\ f_{n-1}(b_0) & f_{n-1}(b_1) & \cdots & f_{n-1}(b_{n-1}) \end{vmatrix}.$$

**练习 4.2.19** 1. 对分块对角矩阵  $X = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$ , 其中  $A, B$  是方阵, 证明,  $\det(X) = \det(A) \det(B)$ .

2. 对分块上三角矩阵  $X = \begin{bmatrix} A & C \\ O & B \end{bmatrix}$ , 其中  $A, B$  是方阵, 证明,  $\det(X) = \det(A) \det(B)$ .

**练习 4.2.20** 设  $A$  可逆,  $D$  是方阵, 证明,  $\det \left( \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right) = \det(A) \det(D - CA^{-1}B)$ .

**练习 4.2.21** 设  $A, B$  是  $n$  阶方阵, 证明,  $\det \left( \begin{bmatrix} A & B \\ B & A \end{bmatrix} \right) = \det(A+B) \det(A-B)$ .

**练习 4.2.22** 设  $n$  阶方阵  $A, B$  满足  $AB = BA$ , 证明,  $\det \left( \begin{bmatrix} A & B \\ -B & A \end{bmatrix} \right) = \det(A^2 + B^2)$ .

**练习 4.2.23** 设  $A, B$  分别是  $m \times n, n \times m$  矩阵, 证明,  $\det(I_m + AB) = \det(I_n + BA)$ . 由此推出,  $I_m + AB$  可逆当且仅当  $I_n + BA$  可逆.

**练习 4.2.24** 计算  $\begin{vmatrix} 1+a_1^2 & a_1a_2 & \cdots & a_1a_n \\ a_2a_1 & 1+a_2^2 & \cdots & a_2a_n \\ \vdots & \vdots & & \vdots \\ a_na_1 & a_na_2 & \cdots & 1+a_n^2 \end{vmatrix}$ .

**练习 4.2.25** 设  $A$  是三阶矩阵, 已知  $\det(A - I_3) = \det(A - 2I_3) = \det(A - 3I_3) = 0$ .

1. 证明存在非零向量  $v_1, v_2, v_3$ , 满足  $Av_i = iv_i$ .

2. 设  $k_1v_1 + k_2v_2 + k_3v_3 = 0$ , 证明  $k_1v_1 + 2k_2v_2 + 3k_3v_3 = 0, k_1v_1 + 4k_2v_2 + 9k_3v_3 = 0$ .

3. 证明存在可逆 Vandermonde 矩阵  $V$ , 使得

$$\begin{bmatrix} k_1 \mathbf{v}_1 & k_2 \mathbf{v}_2 & k_3 \mathbf{v}_3 \end{bmatrix} V = O.$$

4. 证明  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  构成  $\mathbb{R}^3$  的一组基, 因此矩阵  $B = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]$  可逆.

5. 证明存在对角矩阵  $D$ , 使得  $AB = BD$ , 并计算  $\det(A)$ .

**练习 4.2.26** 对函数  $f(t) = \det(I_n + tA)$  在  $t = 0$  处求导. 设  $A = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n]$ , 则  $I_n + tA$  的第  $i$  列是  $\mathbf{e}_i + t\mathbf{a}_i$ .

1. 当  $n = 1, 2, 3$  时, 用  $A$  的元素表示  $f'(0)$ ; 分析其规律, 求  $f'(0)$  的一般表达式.

2. 利用  $\det(I_n + AB) = \det(I_m + BA)$ , 证明  $\text{trace}(AB) = \text{trace}(BA)$  ( $\text{trace}$  的定义见练习 1.5.22).

**练习 4.2.27** 设  $A$  是  $m \times n$  矩阵, 任取其中的  $k$  行和  $k$  列构成  $k$  阶方阵, 它的行列式称为  $A$  的一个  $k$  阶子式. 定义  $\text{rank}_{\det}(A) = \max\{k \mid A \text{ 有非零的 } k \text{ 阶子式}\}$ . 证明,  $\text{rank}_{\det}(A) = \text{rank}(A)$ .

**练习 4.2.28 (Hadamard 不等式)** 1. 利用 QR 分解证明, 对任意  $n$  阶矩阵  $T = [\mathbf{t}_1 \quad \cdots \quad \mathbf{t}_n]$ ,  
 $|\det(T)| \leq \|\mathbf{t}_1\| \|\mathbf{t}_2\| \cdots \|\mathbf{t}_n\|$ .

2. 说明阅读 3.2.3 中的 Hadamard 矩阵使得等号成立.

**练习 4.2.29** 设  $n$  阶对称矩阵  $A$  有  $LDL^T$  分解  $A = LDL^T$ , 其中  $D = \text{diag}(d_i)$ , 并记  $A$  的第  $i$  个顺序主子阵为  $A_i$ . 证明  $d_i = \frac{\det(A_i)}{\det(A_{i-1})}$ .

矩阵  $A$  的第  $i$  个顺序主子阵的行列式称为其第  $i$  个顺序主子式.

**练习 4.2.30 (行列式在多元微积分中的应用)** 一个多元函数  $f(x_1, \cdots, x_n)$ , 把  $x_i$  之外的变量都看做常数, 对  $x_i$  的导数称为  $f$  对  $x_i$  的偏导数, 记作  $\frac{\partial f}{\partial x_i}$ . 例如,  $f(x, y) = x^2 y$ , 则  $\frac{\partial f}{\partial x} = 2xy$ ,  $\frac{\partial f}{\partial y} = x^2$ .

平面  $\mathbb{R}^2$  有直角坐标系  $(x, y)$ , 和极坐标系  $(r, \theta)$ , 其中  $r \geq 0$  是该点到原点的距离,  $\theta \in [0, 2\pi)$  是从  $x$  轴正方向开始, 逆时针旋转, 到达该点和原点连线所需要的的角度. 两种坐标之间的关系是  $x = r \cos \theta, y = r \sin \theta$ .

分别计算  $J_1 = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix}$  和  $J_2 = \begin{bmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{bmatrix}$  的行列式, 将结果都写成  $r, \theta$  的函数. 这两个矩阵  $J_1, J_2$  有什么关系?

**练习 4.2.31** 设  $f(a, b, c, d) = \ln(ad - bc)$ .

1. 接上题, 求偏导数  $\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}, \frac{\partial f}{\partial c}, \frac{\partial f}{\partial d}$ .

2. 证明  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\partial f}{\partial a} & \frac{\partial f}{\partial b} \\ \frac{\partial f}{\partial c} & \frac{\partial f}{\partial d} \end{bmatrix}^T$ .

3. 三阶时是否有类似的结论?

**阅读 4.2.32** 在 Vandermonde 矩阵中, 如果对于不同的  $i, j$  有  $\lambda_i = \lambda_j$ , 则其  $i, j$  两行相同, 行列式为零. 根据多项式理论,  $\lambda_i - \lambda_j$  是 Vandermonde 行列式的一个因式. 事实上, Vandermonde 行列式就是这些因式的乘积.

### 4.3 行列式的展开式

本节的主要内容是通过数学归纳法从低阶行列式推导高阶行列式,从而得到行列式的表达式,以此证明行列式的存在性.

我们已经知道,一阶行列式  $\det([a_{11}]) = a_{11}$ ,二阶行列式  $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$ ,下面来看三阶行列式.

**例 4.3.1 (三阶行列式)** 根据列线性性和行反对称性,有

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11} \begin{vmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + a_{21} \begin{vmatrix} 0 & a_{12} & a_{13} \\ 1 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 1 & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11} \begin{vmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} 1 & a_{22} & a_{23} \\ 0 & a_{12} & a_{13} \\ 0 & a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} 1 & a_{32} & a_{33} \\ 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \end{vmatrix} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}. \end{aligned}$$

因此我们可以用定义好的二阶行列式来得到三阶行列式.

继续计算即有

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

可以验证,等式右端项满足列线性性、列反对称性、单位化条件,从而是唯一的三阶行列式函数.  $\odot$

例 4.3.1 提示我们,可以用  $n-1$  阶行列式去递归地表示  $n$  阶行列式.为简便起见,我们引入如下定义.

**定义 4.3.2 (代数余子式)** 给定  $n$  阶方阵  $A$ ,令  $A \begin{pmatrix} i \\ j \end{pmatrix}$  表示从  $A$  划去第  $i$  行和第  $j$  列得到的  $n-1$  阶方阵,则  $M_{ij} = \det(A \begin{pmatrix} i \\ j \end{pmatrix})$ ,称为元素  $a_{ij}$  的余子式;而  $C_{ij} = (-1)^{i+j}M_{ij} = (-1)^{i+j}\det(A \begin{pmatrix} i \\ j \end{pmatrix})$ ,称为元素  $a_{ij}$  的代数余子式.

二阶行列式  $\det(A) = a_{11}a_{22} - a_{21}a_{12} = a_{11}C_{11} + a_{21}C_{21}$ .而例 4.3.1 中三阶行列式  $\det(A) = a_{11}C_{11} + a_{21}C_{21} + a_{31}C_{31}$ .

下面来说明可以用  $n-1$  阶行列式去定义  $n$  阶行列式.



**定理 4.3.3** 给定  $n$  阶方阵  $A = [a_{ij}]$ , 则函数  $a_{11}C_{11} + \cdots + a_{n1}C_{n1}$  是行列式函数, 即

$$\det(A) = a_{11}C_{11} + \cdots + a_{n1}C_{n1},$$

这称为行列式按第一列的展开式.

证. 对  $n$  应用数学归纳法. 记  $\delta(A) = a_{11}C_{11} + \cdots + a_{n1}C_{n1}$ , 根据命题 4.2.8, 只需证明  $\delta(A)$  满足行线性性、行反对称性和单位化条件. 当  $n = 1$  时, 显然, 现假设命题对  $n - 1$  成立.

行线性性: 这里只证明第一行的情形, 其他情形同理. 设对任意  $j$ , 我们都有  $a_{1j} =$

$$ka'_{1j} + la''_{1j}. \text{ 令 } A' = \begin{bmatrix} a'_{11} & \cdots & a'_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, A'' = \begin{bmatrix} a''_{11} & \cdots & a''_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}. \text{ 根据归纳假设, } i \neq 1 \text{ 时,}$$

代数余子式满足  $C_{i1} = kC'_{i1} + lC''_{i1}$ . 而  $i = 1$  时, 代数余子式  $C_{11} = C'_{11} = C''_{11}$ . 因此,

$$\begin{aligned} \delta(A) &= a_{11}C_{11} + \cdots + a_{n1}C_{n1} \\ &= (ka'_{11} + la''_{11})C_{11} + a_{21}(kC'_{21} + lC''_{21}) + \cdots + a_{n1}(kC'_{n1} + lC''_{n1}) \\ &= k(a'_{11}C'_{11} + a_{21}C'_{21} + \cdots + a_{n1}C'_{n1}) + l(a''_{11}C''_{11} + a_{21}C''_{21} + \cdots + a_{n1}C''_{n1}) \\ &= k\delta(A') + l\delta(A''). \end{aligned}$$

行反对称性: 行反对称性等价于任意两行相同的行列式为零. 这里只证明第一行和第二行的情形, 其他情形同理. 设  $A$  前两行相同, 根据归纳假设,  $i \neq 1, 2$  时, 代数余子式  $C_{i1} = 0$ . 而  $i = 1, 2$  时,  $A$  的两个余子式相同, 但代数余子式的符号相反, 即  $C_{11} = -C_{21}$ . 因此,

$$\det(A) = a_{11}C_{11} + \cdots + a_{n1}C_{n1} = a_{11}C_{11} + a_{11}(-C_{11}) = 0.$$

单位化条件:  $\det(I_n) = a_{11}C_{11} = 1 \cdot \det(I_{n-1}) = 1$ .

因此, 命题对  $n$  也成立. □

定理 4.3.3 说明了行列式函数的存在性. 从证明中还能看出, 代数余子式中的正负号对应了行列式的反对称性.

上三角矩阵  $U = \begin{bmatrix} u_{11} & * & \cdots & * \\ & u_{22} & \ddots & \vdots \\ & & \ddots & * \\ & & & u_{nn} \end{bmatrix}$  的行列式, 可以用按列展开式简单解释:  $u_{11}$

的余子式仍是上三角矩阵的行列式, 以此类推, 反复运用展开式可得:  $\det(U) = u_{11}C_{11} = u_{11}u_{22} \cdots u_{nn}$ .

**例 4.3.4** 计算 (3.2.1) 中的正交矩阵

$$Q = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \end{bmatrix}$$

的行列式. 按第一列展开, 有

$$\det(Q) = \frac{1}{\sqrt{3}} \begin{vmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ -\frac{2}{\sqrt{6}} & 0 \end{vmatrix} - \frac{1}{\sqrt{3}} \begin{vmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{2}{\sqrt{6}} & 0 \end{vmatrix} + \frac{1}{\sqrt{3}} \begin{vmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{vmatrix} = -1.$$

也可以利用倍加变换,

$$\begin{vmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \end{vmatrix} = \begin{vmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & -\sqrt{2} \\ 0 & -\frac{3}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{vmatrix} = \frac{1}{\sqrt{3}} \begin{vmatrix} 0 & -\sqrt{2} \\ -\frac{3}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{vmatrix} = \frac{1}{\sqrt{3}} \left( -\frac{3}{\sqrt{6}} \sqrt{2} \right) = -1.$$

☺

根据行列式的性质, 可以得到与定理 4.3.3 中不同的展开式.

**命题 4.3.5** 行列式按任意一行或任意一列展开:

1. 按第  $j$  列展开:  $\det(A) = a_{1j}C_{1j} + \cdots + a_{nj}C_{nj}$ ;

2. 按第  $i$  行展开:  $\det(A) = a_{i1}C_{i1} + \cdots + a_{in}C_{in}$ .

证. 第 1 条: 设  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ , 而  $A' = [\mathbf{a}_j \ \mathbf{a}_1 \ \cdots \ \mathbf{a}_{j-1} \ \mathbf{a}_{j+1} \ \cdots \ \mathbf{a}_n]$  是把  $A$  的第  $j$  列移到第 1 列之前得到的矩阵, 这能够通过调相邻两列  $j-1$  次实现, 因此  $\det(A) = (-1)^{j-1} \det(A')$ .

注意到  $A'$  的第 1 列是  $A$  的第  $j$  列, 因此  $a'_{i1} = a_{ij}$ ,  $M'_{i1} = M_{ij}$ . 对  $A'$  的第 1 列展开可得:

$$\det(A') = \sum_{1 \leq i \leq n} (-1)^{i+1} a'_{i1} M'_{i1} = (-1)^{j+1} \sum_{1 \leq i \leq n} (-1)^{i+j} a_{ij} M_{ij}.$$

因此,  $\det(A) = (-1)^{j-1} \det(A') = \sum_{1 \leq i \leq n} a_{ij} C_{ij}$ .

第 2 条: 对  $A$  的第  $i$  行展开就是对  $A^T$  的第  $i$  列展开, 由第 1 条和  $\det(A) = \det(A^T)$  可得.  $\square$

**命题 4.3.6** 令  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ , 再记第  $j$  列元素的代数余子式组成的向量为  $\mathbf{c}_j =$

$$\begin{bmatrix} C_{1j} \\ \vdots \\ C_{nj} \end{bmatrix}. \text{ 则当 } j' \neq j \text{ 时, } \mathbf{a}_{j'}^T \mathbf{c}_j = 0; \text{ 当 } j' = j \text{ 时, 有 } \mathbf{a}_{j'}^T \mathbf{c}_j = \det(A).$$

对某行元素的代数余子式组成的向量, 也有类似结论.

证. 只考虑  $j' \neq j$  的情形. 在  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$  中, 令  $\mathbf{a}_j = \mathbf{a}_{j'}$ , 按第  $j$  列展开得到行列式  $a_{1j'}C_{1j} + \cdots + a_{nj'}C_{nj}$ . 另一方面这个矩阵有两列相同, 因此行列式为 0.

关于行的结论, 利用转置立得.  $\square$

某列元素的代数余子式组成的向量有明确的几何含义. 以三阶情形为例, 记  $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3]$ , 则  $\det(A)$  是以  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  为棱的平行六面体的有向体积. 根据命题 4.3.6,  $\mathbf{c}_1 \perp \mathbf{a}_2, \mathbf{c}_1 \perp \mathbf{a}_3$ , 即  $\mathbf{c}_1$  与  $\mathbf{a}_1, \mathbf{a}_2$  所在平面正交. 换言之, 如果  $\mathbf{a}_2, \mathbf{a}_3$  所在的面看作底面, 那么  $\mathbf{c}_1$  就是底面的法向量, 因此  $\mathbf{a}_1$  向  $\mathbf{c}_1$  方向的投影  $\frac{\mathbf{c}_1^T \mathbf{a}_1}{\mathbf{c}_1^T \mathbf{c}_1} \mathbf{c}_1$  的 (有向) 长度  $\frac{\mathbf{c}_1^T \mathbf{a}_1}{\|\mathbf{c}_1\|}$  就是该底面的 (有向) 高. 体积等于底面积与高的乘积, 结合  $\det(A) = \mathbf{a}_1^T \mathbf{c}_1 = \frac{\mathbf{c}_1^T \mathbf{a}_1}{\|\mathbf{c}_1\|} \|\mathbf{c}_1\|$ , 就有  $\|\mathbf{c}_1\|$  是  $\mathbf{a}_2, \mathbf{a}_3$  所在底面的面积. 综上,  $\mathbf{c}_1$  是底面的一个长度等于底面积的向量.

命题 4.3.6 中元素及其代数余子式的内积可以得到不少推论.

对矩阵  $A = [a_{ij}]$ , 记  $C = [C_{ij}]_{n \times n}$ , 即  $C$  的  $(i, j)$  元素是  $a_{ij}$  的代数余子式, 矩阵  $C^T$  常称为  $A$  的伴随矩阵.

**推论 4.3.7 (逆矩阵公式)** 对可逆矩阵  $A$ ,  $A^{-1} = \frac{1}{\det(A)} C^T$ .

证. 命题 4.3.6 说明  $C^T A = \det(A) I_n$ . 立得.  $\square$

**推论 4.3.8 (Cramer 法则)** 给定方阵  $A$ , 线性方程组  $A\mathbf{x} = \mathbf{b}$  有唯一解, 当且仅当  $\det(A) \neq 0$ , 且有唯一解时, 唯一解为

$$x_1 = \frac{\det(B_1)}{\det(A)}, \cdots, x_n = \frac{\det(B_n)}{\det(A)},$$

其中  $B_j$  是把  $A$  的第  $j$  列换成  $\mathbf{b}$  得到的矩阵.

证. 这里只计算唯一解. 记  $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ . 线性方程组  $A\mathbf{x} = \mathbf{b}$  可以写作

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{b}.$$

求两边与  $\mathbf{c}_i$  的内积, 得  $x_i \mathbf{c}_i^T \mathbf{a}_i = \mathbf{c}_i^T \mathbf{b}$ . 对  $A$  和  $B_i$  的行列式分别按第  $i$  列展开, 有  $\det(A) = \mathbf{c}_i^T \mathbf{a}_i, \det(B_i) = \mathbf{c}_i^T \mathbf{b}$ . 因此  $x_i = \frac{\det(B_i)}{\det(A)}$ .  $\square$

事实上, Cramer 法则也有明确的几何含义. 仍以三阶情形为例. 如果  $x_1$  是解中对应元素, 则  $\mathbf{b} - x_1 \mathbf{a}_1 = x_2 \mathbf{a}_2 + x_3 \mathbf{a}_3 \subseteq \text{span}(\mathbf{a}_2, \mathbf{a}_3)$ , 即  $\mathbf{b} - x_1 \mathbf{a}_1$  落在  $\mathbf{a}_2, \mathbf{a}_3$  生成的平面. 因此  $\mathbf{b}$  在该平面法向的投影与  $x_1 \mathbf{a}_1$  在该平面法向的投影相等. 而  $\mathbf{c}_1$  是该平面的法向量, 因此  $\frac{\mathbf{b}^T \mathbf{c}_1}{\mathbf{c}_1^T \mathbf{c}_1} \mathbf{c}_1 = \frac{x_1 \mathbf{a}_1^T \mathbf{c}_1}{\mathbf{c}_1^T \mathbf{c}_1} \mathbf{c}_1$ . 进一步地,  $x_1$  是以  $\mathbf{b}, \mathbf{a}_2, \mathbf{a}_3$  为棱的平行六面体与以  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  为棱的平行六面体的有向体积之比 (为什么?).

注意, 逆矩阵一般不应通过伴随矩阵计算, 线性方程组的解一般也不应利用 Cramer 法则计算. 因为伴随矩阵或 Cramer 法则中的行列式终归需要利用 PLU 分解来计算, 而利用 PLU 分解已经能直接求出逆矩阵或方程组的解了.

最后考虑行列式的完全展开, 即行列式关于元素的表达式.

在行列式按列展开中, 余子式都是低一阶的行列式, 继续展开, 反复进行, 就能得到行列式用元素表示的形式.

二阶矩阵的完全展开式是

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

三阶行列式的完全展开式是

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

可以看到, 完全展开之后的公式是若干元素乘积的代数和, 其中一半带正号, 一般带负号. 还可以看到, 每个乘积的因数中, 每行都有且仅有一个元素出现, 每列都有且仅有一个元素出现. 这种乘积共有  $n!$  个. 可以猜测任意阶行列式的完全展开式也具有上述规律. 下面就来说明猜测的正确性.

首先确定符号.

**定义 4.3.9 (排列)** 将正整数  $1, \dots, n$  按一定顺序排列起来得到  $\sigma_1, \dots, \sigma_n$ , 称为一个**排列**或**置换**. 这里称为排列  $\sigma$ .

对调排列中两个数的顺序, 称为对该排列施加一次**对换**.

排列  $\sigma$ , 如果可以经过奇数次对换得到  $1, 2, \dots, n$ , 则称为**奇排列**; 如果可以经过偶数次对换得到  $1, 2, \dots, n$ , 则称为**偶排列**.

我们可以在排列和置换矩阵之间建立一一对应. 事实上, 映射  $\sigma \mapsto [e_{\sigma_1} \ \dots \ e_{\sigma_n}]$  是双射 (为什么?). 而奇排列对应的置换矩阵的行列式是  $-1$ , 偶排列对应的置换矩阵的行列式是  $1$ .

任意排列是奇偶排列之一, 因为利用对换逐渐把其中较小的数排到前面, 最后一定能得到  $1, 2, \dots, n$ . 但排列不能既是奇排列又是偶排列, 因为置换矩阵的行列式不能既是  $1$  又是  $-1$ .

**定义 4.3.10 (排列的符号)** 对排列  $\sigma$ , 如果它是奇排列, 则定义其符号为  $\text{sign}(\sigma) = -1$ ; 否则它是偶排列, 定义其符号为  $\text{sign}(\sigma) = 1$ .

然后来验证猜测的正确性.

**命题 4.3.11 (行列式完全展开)** 如下等式成立:

$$1. \det(A) = \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a_{\sigma_1 1} \cdots a_{\sigma_n n}.$$

$$2. \det(A) = \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a_{1\sigma_1} \cdots a_{n\sigma_n}.$$

证. 只证明第 1 条, 第 2 条利用转置立得. 令  $\delta(A) = \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a_{\sigma_1 1} \cdots a_{\sigma_n n}$ .

列线性性: 这里只证明第一列的情形, 其他情形同理. 设对任意  $i$ , 我们都有  $a_{i1} =$

$$ka'_{i1} + la''_{i1}. \text{ 令 } A' = \begin{bmatrix} a'_{11} & \cdots & a'_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, A'' = \begin{bmatrix} a''_{11} & \cdots & a''_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \text{ 则}$$

$$\begin{aligned} \delta(A) &= \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a_{\sigma_1 1} \cdots a_{\sigma_n n} \\ &= \sum_{\text{排列 } \sigma} \text{sign}(\sigma) (ka'_{\sigma_1 1} + la''_{\sigma_1 1}) a_{\sigma_2 2} \cdots a_{\sigma_n n} \\ &= k \left( \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a'_{\sigma_1 1} a_{\sigma_2 2} \cdots a_{\sigma_n n} \right) + l \left( \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a''_{\sigma_1 1} a_{\sigma_2 2} \cdots a_{\sigma_n n} \right) \\ &= k\delta(A') + l\delta(A''). \end{aligned}$$

列反对称性: 这里只证明第一列和第二列的情形, 其他情形同理. 假设  $A'$  为  $A$  交换前两列所得, 因此我们有  $a'_{i1} = a_{i2}, a'_{i2} = a_{i1}$ . 而当  $j \neq 1, 2$  时,  $a'_{ij} = a_{ij}$ . 对一个排序  $\sigma$ , 我们将前两项进行一次对调后, 称得到的新排序为  $\sigma'$ . 显然  $\text{sign}(\sigma') = -\text{sign}(\sigma)$ . 因此,

$$\begin{aligned} \delta(A) &= \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a_{\sigma_1 1} \cdots a_{\sigma_n n} \\ &= \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a_{\sigma_2 1} a_{\sigma_1 2} a_{\sigma_3 3} \cdots a_{\sigma_n n} \\ &= \sum_{\text{排列 } \sigma} \text{sign}(\sigma) a'_{\sigma_1 1} a'_{\sigma_2 2} a'_{\sigma_3 3} \cdots a'_{\sigma_n n} \\ &= - \sum_{\text{排列 } \sigma'} \text{sign}(\sigma') a'_{\sigma'_1 1} a'_{\sigma'_2 2} a'_{\sigma'_3 3} \cdots a'_{\sigma'_n n} \\ &= -\delta(A'). \end{aligned}$$

单位化条件:  $\delta(I_n)$  对应的表达式中, 除了  $a_{11} \cdots a_{nn}$  之外均为零. 因此  $\delta(I_n) = 1$ . 因此,  $\delta(A) = \det(A)$ .  $\square$

完全展开式也可以作为行列式存在性的证明.

注意, 行列式一般不应通过完全展开来计算.

## 习题

练习 4.3.1 计算.

1. 
$$\begin{vmatrix} 1 & 1 & 4 \\ 1 & 2 & 2 \\ 1 & 2 & 5 \end{vmatrix}.$$

2. 
$$\begin{vmatrix} 1 & 1 & 10 \\ 1 & 2 & 2 \\ 1 & 2 & 5 \end{vmatrix}.$$

练习 4.3.2 利用按行(列)展开求下列行列式;按哪一行(列)展开,使得计算最简单?

1. 
$$\begin{vmatrix} 1 & 0 & 0 & 2 \\ 0 & 3 & 4 & 5 \\ 5 & 4 & 0 & 3 \\ 2 & 0 & 0 & 1 \end{vmatrix}.$$

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

练习 4.3.3 计算

$$\begin{vmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ b_1 & b_2 & b_3 & b_4 & b_5 \\ c_1 & c_2 & 0 & 0 & 0 \\ d_1 & d_2 & 0 & 0 & 0 \\ e_1 & e_2 & 0 & 0 & 0 \end{vmatrix}.$$

练习 4.3.4 求

$$\begin{vmatrix} 2x & x & 1 & 2 \\ 1 & x & 1 & -1 \\ 3 & 2 & x & 1 \\ 1 & 1 & 1 & x \end{vmatrix}$$
 中  $x^4, x^3$  的系数.

练习 4.3.5 计算

$$\begin{vmatrix} \lambda & & & a_n \\ -1 & \lambda & & a_{n-1} \\ & \ddots & \ddots & \vdots \\ & & -1 & \lambda & a_2 \\ & & & -1 & \lambda + a_1 \end{vmatrix}.$$

练习 4.3.6 回顾例 1.8.4 中的对称、上三角和下三角 Pascal 矩阵. 在四阶的情形, 这三种矩阵分别为

$$S_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}, U_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, L_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{bmatrix}. \text{ 另外, 存在 LU 分解 } S_n = L_n U_n.$$

1. 求  $\det(L_n), \det(U_n), \det(S_n)$ .
2. 求  $S_n$  右下角元素的代数余子式.
3. 将  $S_n$  右下角的元素减 1 得到矩阵  $A_n$ , 求  $\det(A_n)$ .

练习 4.3.7 给定  $A_n = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}, B_n = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}.$

1. 利用展开式得到  $\det(B_n)$  关于  $n$  的递推关系, 并计算  $\det(B_n)$ .
2. 利用  $\det(A_n)$  与  $\det(B_n)$  的关系计算  $\det(A_n)$ .

**练习 4.3.8 (行列式中的 Fibonacci 数列)** 如果一个矩阵比上(下)三角矩阵仅仅多一排非零对角

元, 则称之为上(下) Hessenberg 矩阵. 例如,  $H_4 = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$  就是上 Hessenberg 矩阵. 上

Hessenberg 矩阵在数值代数中很有用.

1. 令  $H_n$  为  $n$  阶上 Hessenberg 矩阵, 其对角元素都是 2, 其他非零元素都是 1. 证明  $\det(H_{n+2}) = \det(H_{n+1}) + \det(H_n)$ , 即这些行列式组成了 Fibonacci 数列.
2. 令  $S_n$  是对角元素为 3, 与对角元相邻的元素为 1 的  $n$  阶三对角矩阵 (见练习 1.4.12), 例如,  $S_4 = \begin{bmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$ . 它既是上 Hessenberg 矩阵, 也是下 Hessenberg 矩阵. 求  $S_n$  的递归公式, 并分析与 Fibonacci 数列的关系.
3. 设  $n$  阶三对角矩阵的行列式的完全展开式中, 最多有  $t_n$  项非零, 求  $t_n$  的递归公式.

**练习 4.3.9** 求下列推广的 Vandermonde 行列式:

$$\begin{vmatrix} 1 & x_1 & \cdots & x_1^{n-2} & x_1^n \\ 1 & x_2 & \cdots & x_2^{n-2} & x_2^n \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & x_n & \cdots & x_n^{n-2} & x_n^n \end{vmatrix}.$$

**练习 4.3.10** 对  $n$  阶方阵  $A$ , 证明,  $\det(\lambda I_n - A)$  是  $\lambda$  的首项系数为 1 的  $n$  次多项式.

- 练习 4.3.11**
1. 设  $A$  是正交矩阵, 且  $\det(A) < 0$ , 证明  $I_n + A$  不可逆. 由此可得, 存在非零向量  $x$ , 使得  $Ax = -x$ .
  2. 设  $A$  是奇数阶正交矩阵, 且  $\det(A) > 0$ , 证明  $I_n - A$  不可逆. 由此可得, 存在非零向量  $x$ , 使得  $Ax = x$ . 偶数阶的情形, 结论是否成立?

**练习 4.3.12** 设  $n$  阶方阵  $A = [a_{ij}]$  对角占优 (见定义 1.6.13), 证明  $\det(A) > 0$ .

**练习 4.3.13** 证明若  $A$  不可逆, 则其伴随矩阵的秩是 0 或 1.

**练习 4.3.14** 给定所有元素全为整数的可逆矩阵  $A$ , 证明,  $A^{-1}$  的所有元素全为整数, 当且仅当  $|\det(A)| = 1$ .

**练习 4.3.15** 利用 Cramer 法则把未知数  $x_1, x_2$  表示成  $t$  的函数:

$$\begin{cases} e^t x_1 + e^{-2t} x_2 = 3 \sin t, \\ e^t x_1 - 2e^{-2t} x_2 = t \cos t. \end{cases}$$

**阅读 4.3.16 (Cramer 法则的另一证明)** 设  $\mathbf{b} = A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3$ , 则

$$\det [\mathbf{b} \quad \mathbf{a}_2 \quad \mathbf{a}_3] = \det [x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 \quad \mathbf{a}_2 \quad \mathbf{a}_3] = \det [x_1\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3] = x_1 \det(A).$$

**练习 4.3.17** 利用完全展开式求行列式.

1. 设  $A$  为 4 阶矩阵, 所有元素均为 1, 在其行列式的完全展开式中, 多少项为 1? 多少项为 -1? 由此计算  $A$  的行列式.
2. 把  $A$  的  $(i, j)$  元乘以  $\frac{i}{j}$  得到  $B$ , 在  $A$  和  $B$  行列式的完全展开式中, 每一项如何变化? 行列式如何变化?

3. 设  $A = \begin{bmatrix} a & 0 & b & 0 \\ 0 & c & 0 & d \\ e & 0 & f & 0 \\ 0 & g & 0 & h \end{bmatrix}$ , 在行列式的完全展开式中, 有多少项非零? 这个完全展开式是否有因式分解? (跟分块对角矩阵的情形进行类比)

**练习 4.3.18** 由  $\begin{vmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{vmatrix} = 0$ , 证明,  $1, \dots, n$  的所有排列中, 奇、偶排列各占一半.

**练习 4.3.19** 在空间  $\mathbb{R}^3$  中, 证明由向量  $\mathbf{a}_1, \mathbf{a}_2$  围成的平行四边形面积是  $\sqrt{\det(A^T A)}$ , 其中  $A = [\mathbf{a}_1 \quad \mathbf{a}_2]$ .

**练习 4.3.20 (行列式游戏)** 甲和乙两个人构造一个  $n$  阶方阵, 轮流填写矩阵中的元素, 直到填满为止. 如果矩阵的行列式非零, 则甲胜; 否则, 乙胜.

1. 如果乙先开始, 且  $n = 2$ , 谁有必胜策略?
2. 如果乙先开始, 且  $n = 3$ , 谁有必胜策略?
3. 如果甲先开始, 且  $n = 3$ , 谁有必胜策略?
4. 如果甲先开始, 且  $n$  为偶数, 则谁有必胜策略?



## 第五章 特征值和特征向量

第 1 至 4 章主要研究了线性映射的像与原像的关系, 重点解决了与线性方程组  $A\mathbf{x} = \mathbf{b}$  有关的一系列问题. 从本章开始, 我们将把注意力转向更加复杂的问题.

### 5.1 引子

先看一个人口流动模型.

**例 5.1.1** 人口在甲地和乙地之间流动. 假设甲地的人口, 每年有 40% 移居到乙地, 剩下 60% 留在甲地; 而乙地的人口, 每年有 10% 移居到甲地, 剩下 90% 留在乙地. 用二维向量来表示两地的人口, 则一年人口变化可由如下线性变换描述:

$$\begin{aligned} A: \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &\mapsto \begin{bmatrix} 0.6 & 0.1 \\ 0.4 & 0.9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.6x_1 + 0.1x_2 \\ 0.4x_1 + 0.9x_2 \end{bmatrix}, \end{aligned}$$

表示矩阵是  $A = \begin{bmatrix} 0.6 & 0.1 \\ 0.4 & 0.9 \end{bmatrix}$ . 现在希望得到这个系统在充分长时间后的变化规律, 以及是否存在一个稳定的人口分布, 即在人口移动下不变的人口分布. 这两个问题的数学叙述就是:

1. 线性变换  $A^t$  在  $t$  充分大时的规律, 特别地,  $A^t$  当  $t \rightarrow \infty$  的极限;
2. 是否存在  $\mathbf{x}$  使得  $A\mathbf{x} = \mathbf{x}$ .

第二个问题较为简单. 事实上, 将等式右端的  $\mathbf{x}$  移到左端即得齐次线性方程组  $(I_2 - A)\mathbf{x} = \mathbf{0}$ , 系数矩阵  $I_2 - A = \begin{bmatrix} -0.4 & 0.1 \\ 0.4 & -0.1 \end{bmatrix}$ . 显然  $\text{rank}(I_2 - A) = 1$ , 解空间是一维子空间  $\left\{ k\mathbf{x}_1 \mid \mathbf{x}_1 = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} \right\}$ , 其中  $\mathbf{x}_1$  满足  $A\mathbf{x}_1 = \mathbf{x}_1$ , 即当甲乙两地的人口分别占两地总人口的 20% 和 80% 时, 人口分布保持不变.

第一个问题需要深入考虑. 直观上, 向量或矩阵的极限, 就是其分量或元素的极限组成的向量或矩阵, 这一点这里不详细说明. 下面考察线性变换  $A^t$  的变化规律, 只需要考虑对任意  $\mathbf{x}$ ,  $A^t \mathbf{x}$  的变化规律. 由于任意向量  $\mathbf{x}$  都能表示为一组基的线性组合, 因此只需要考虑一组基的变化规律. 而  $\mathbf{x}_1$  在  $A^t$  下的变化规律已得, 即  $A^t \mathbf{x}_1 = \mathbf{x}_1$ . 基不妨由  $\mathbf{x}_1$  扩张得到, 设  $\mathbf{x}_1, \mathbf{x}_2$  是  $\mathbb{R}^2$  的一组基. 于是  $A\mathbf{x}_2$  可由其表示, 即  $A\mathbf{x}_2 = k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2$ . 类似地,

$$\begin{aligned} A^2 \mathbf{x}_2 &= A(A\mathbf{x}_2) = A(k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2) = k_1 A\mathbf{x}_1 + k_2 A\mathbf{x}_2 \\ &= k_1 \mathbf{x}_1 + k_2 (k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2) = (k_1 + k_1 k_2) \mathbf{x}_1 + k_2^2 \mathbf{x}_2. \end{aligned}$$

以此类推, 可以得到越来越复杂的表达式来表示  $A^t \mathbf{x}_2$ . 易见, 如果  $k_1 = 0$ , 表达式就有明显的规律  $A^t \mathbf{x}_2 = k_2^t \mathbf{x}_2$ , 那么对任意向量  $\mathbf{x} = l_1 \mathbf{x}_1 + l_2 \mathbf{x}_2$ , 有  $A^t \mathbf{x} = l_1 \mathbf{x}_1 + l_2 k_2^t \mathbf{x}_2$ . 那么, 存在满足  $A\mathbf{x}_2 = k_2 \mathbf{x}_2$  的基向量  $\mathbf{x}_2$  吗?

类似上面对第二个问题的讨论, 可得带参数  $k_2$  的齐次方程组  $(k_2 I_2 - A)\mathbf{x}_2 = \mathbf{0}$ . 齐次方程组有非零解当且仅当  $k_2 I_2 - A$  不可逆, 而根据定理 4.2.6, 又当且仅当行列式  $\det(k_2 I_2 - A) = \begin{vmatrix} k_2 - 0.6 & 0.1 \\ 0.4 & k_2 - 0.9 \end{vmatrix} = 0$ . 由此得到关于  $k_2$  的一元二次方程  $k_2^2 - 1.5k_2 + 0.5 = 0$ , 方程有两个实根 1, 0.5. 根 1 对应  $I_2 - A$  不可逆, 其解空间的基  $\mathbf{x}_1$  已在对第二个问题的讨论中得到,  $\mathbf{x}_1$  满足  $(I_2 - A)\mathbf{x}_1 = \mathbf{0}$ . 根 0.5 对应  $0.5I_2 - A$  不可逆, 类似可得其解空间的基  $\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ , 满足  $A\mathbf{x}_2 = 0.5\mathbf{x}_2$ .

简单验证,  $\mathbf{x}_1, \mathbf{x}_2$  是  $\mathbb{R}^2$  的一组基, 因此  $A^t$  在其上的取值  $A^t \mathbf{x}_1 = \mathbf{x}_1, A^t \mathbf{x}_2 = 0.5^t \mathbf{x}_2$ , 完全决定了线性变换  $A^t$ . 事实上, 记  $X = [\mathbf{x}_1 \quad \mathbf{x}_2]$ , 则

$$A^t X = [A^t \mathbf{x}_1 \quad A^t \mathbf{x}_2] = [\mathbf{x}_1 \quad 0.5^t \mathbf{x}_2] = X \begin{bmatrix} 1 & \\ & 0.5^t \end{bmatrix} =: X \Lambda^t,$$

其中  $\Lambda = \text{diag}(1, 0.5)$ . 而  $X$  可逆, 因此

$$A^t = X \Lambda^t X^{-1} = \begin{bmatrix} 0.2 & 1 \\ 0.8 & -1 \end{bmatrix} \begin{bmatrix} 1 & \\ & 0.5^t \end{bmatrix} \begin{bmatrix} 0.2 & 1 \\ 0.8 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 0.2 + \frac{0.8}{2^t} & 0.2 - \frac{0.2}{2^t} \\ 0.8 - \frac{0.8}{2^t} & 0.8 + \frac{0.2}{2^t} \end{bmatrix}.$$

当  $t \rightarrow \infty$  时, 对应的线性变换  $A^t \rightarrow \begin{bmatrix} 0.2 & 0.2 \\ 0.8 & 0.8 \end{bmatrix}$ . 容易验证, 对任意初始人口分布

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ ,  $A^s \mathbf{x} \rightarrow \begin{bmatrix} 0.2(x_1 + x_2) \\ 0.8(x_1 + x_2) \end{bmatrix} = (x_1 + x_2)\mathbf{x}_1$ , 即在两地总人口保持不变时, 无论初始分布如何, 长时间流动后, 人口分布都趋向于达到稳定状态, 甲地占 80%, 乙地占 20%. ☺

从例 5.1.1 中可以看到, 方程  $A\mathbf{x} = \lambda\mathbf{x}$  对理解  $A$  所对应的线性变换至关重要.

例 5.1.1 启发我们思考方程  $A\mathbf{x} = \lambda\mathbf{x}$  的性质.

首先,  $\lambda$  满足该方程, 当且仅当  $\det(\lambda I - A) = 0$ , 而  $\lambda$  就是这个一元方程的根.

其次, 如果线性空间  $\mathbb{R}^n$  存在一组基  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , 且每个基向量都满足该方程, 即  $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$ , 那么记  $X = [\mathbf{x}_1 \ \dots \ \mathbf{x}_n]$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ , 就有  $A = X\Lambda X^{-1}$ , 因此  $A^t = X\Lambda^t X^{-1}$ , 于是  $A^t$  的长期演化就尽在掌握了.

再次, 方程  $A\mathbf{x} = \lambda\mathbf{x}$  蕴涵着  $A\mathbf{x} \parallel \mathbf{x}$ , 即  $\mathbf{x}$  在  $A$  对应的线性变换下的像与  $\mathbf{x}$  共线.

下面从该方程的角度去观察  $\mathbb{R}^2$  上的两类保距变换.

**例 5.1.2** 反射变换的表示矩阵为  $H_v = I_2 - 2\mathbf{v}\mathbf{v}^T$ , 其中  $\mathbf{v}$  是单位向量. 令  $\mathbf{w}$  是与  $\mathbf{v}$  垂直的单位向量, 则  $H_v\mathbf{w} = \mathbf{w}$ ,  $H_v\mathbf{v} = -\mathbf{v}$ , 即

$$H_v \begin{bmatrix} \mathbf{v} & \mathbf{w} \end{bmatrix} = \begin{bmatrix} -\mathbf{v} & \mathbf{w} \end{bmatrix}.$$

反射变换的几何性质能够简单说明, 如果  $\mathbf{x} \nparallel \mathbf{v}$ ,  $\mathbf{x} \nparallel \mathbf{w}$ , 则  $H_v\mathbf{x} \nparallel \mathbf{x}$ . 换言之, 当  $\lambda = \pm 1$  时, 方程  $H_v\mathbf{x} = \lambda\mathbf{x}$  关于  $\mathbf{x}$  有非零解; 而其他  $\lambda$  不会使方程关于  $\mathbf{x}$  有非零解.

另一方面, 记  $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ , 考虑方程

$$0 = \det(\lambda I_2 - H_v) = \begin{vmatrix} \lambda - 1 + 2v_1^2 & 2v_1v_2 \\ 2v_1v_2 & \lambda - 1 + 2v_2^2 \end{vmatrix} = \lambda^2 - 1,$$

即知满足该方程的  $\lambda$  只能是  $\pm 1$ , 和几何直观一致.

旋转变换的表示矩阵为  $R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ . 如果  $\theta \neq k\pi$ , 则任意非零向量在变换下的像一定不会与原向量共线, 即对任意实数  $\lambda$ ,  $R_\theta\mathbf{x} = \lambda\mathbf{x}$  都没有非零解.

另一方面, 考虑方程

$$0 = \det(\lambda I_2 - R_\theta) = \begin{vmatrix} \lambda - \cos \theta & \sin \theta \\ -\sin \theta & \lambda - \cos \theta \end{vmatrix} = \lambda^2 - 2\cos \theta \lambda + 1.$$

这个关于  $\lambda$  的一元二次方程的判别式  $(-2\cos \theta)^2 - 4 = -4\sin^2 \theta < 0$ , 因此没有实根, 和几何直观一致.

然而这个方程有两复根  $\cos \theta + i\sin \theta, \cos \theta - i\sin \theta$ . 复数的出现有意义吗? ☺

如前所述, 给定  $n$  阶方阵  $A$ , 寻找满足  $A\mathbf{x} = \lambda\mathbf{x}$  的  $\lambda$ , 就是求解一元方程  $\det(\lambda I_n - A) = 0$  的根. 先考虑关于参数  $\lambda$  的表达式  $\det(\lambda I_n - A)$ . 在行列式的完全展开式中, 所有对角元素的乘积是关于  $\lambda$  的  $n$  次多项式, 且首项 (即次数最高的项) 系数为 1, 而其他项是关于  $\lambda$  的次数小于  $n$  的多项式, 因此  $\det(\lambda I_n - A)$  是关于  $\lambda$  的  $n$  次的多项式, 且首项系数为 1.

因此问题归结于求解一个一元多项式的根<sup>1</sup>. 而在求解多项式的根时, 复数的出现是一种必需.

**定理 5.1.3 (代数学基本定理)** 复系数一元  $n$  次多项式在  $\mathbb{C}$  上至少有一个根.

该定理十分重要, 但不能通过纯粹的代数方法证明, 这里仅仅列出结论. 容易由此得到如下推论.

**推论 5.1.4** 复系数一元  $n$  次多项式  $p(x)$ , 在  $\mathbb{C}$  上恰好有  $n$  个根 (可能相同), 即存在因式分解  $p(x) = a_0(x - x_1)^{n_1} \cdots (x - x_s)^{n_s}$ , 其中  $n_1 + \cdots + n_s = n$ .

证. 根据代数学基本定理,  $p(x)$  总存在一个复根  $x_1$ , 即  $p(x_1) = 0$ . 记  $p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ , 则  $p(x_1) = a_0x_1^n + a_1x_1^{n-1} + \cdots + a_{n-1}x_1 + a_n$ . 因此

$$p(x) - p(x_1) = a_0(x^n - x_1^n) + a_1(x^{n-1} - x_1^{n-1}) + \cdots + a_{n-1}(x - x_1).$$

注意到

$$x^k - x_1^k = (x - x_1)(x^{k-1} + x^{k-2}x_1 + \cdots + xx_1^{k-2} + x_1^{k-1}),$$

因此存在  $n-1$  次多项式  $q(x)$ , 使得  $p(x) - p(x_1) = q(x)(x - x_1)$ . 而  $p(x_1) = 0$ , 于是  $p(x) = q(x)(x - x_1)$ . 对  $n-1$  次多项式  $q(x)$  重复上述步骤, 类似进行下去, 即得结论.  $\square$

在上述因式分解中,  $n_i$  称为复根  $x_i$  的**重数**,  $x_i$  称为  $p(x)$  的  $n_i$  重根.

多项式的根与其系数满足 Vieta 定理, 证明留给读者.

**定理 5.1.5 (Vieta 定理)** 复系数一元  $n$  次多项式  $p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$  的  $n$  个根 (计重数)  $x_1, \cdots, x_n$  满足:

$$\begin{aligned} -\frac{a_1}{a_0} &= x_1 + x_2 + \cdots + x_n, \\ &\vdots \\ (-1)^k \frac{a_k}{a_0} &= \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k}, \\ &\vdots \\ (-1)^n \frac{a_n}{a_0} &= x_1 x_2 \cdots x_n. \end{aligned}$$

例 5.1.1 和 5.1.2 中计算了实系数多项式的根, 也观察到实系数多项式不一定有实根. 但实系数多项式的复根共轭成对出现, 如例中的  $\cos \theta \pm i \sin \theta$ . 读者在中学已经学

<sup>1</sup>一般地, 根用来指称方程  $p(x) = 0$  的解, 而使得  $p(x) = 0$  的值称为  $p(x)$  的零点, 不过本书对这两个概念不加区分.

过共轭运算: 对实数  $a, b$ , 复数  $a - bi$  是复数  $a + bi$  的**共轭**, 记为  $a - bi = \overline{a + bi}$ . 共轭运算保持加法和乘法, 即  $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ ,  $\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$ . 容易证明, 对实系数多项式  $p(x)$ , 如果  $a + bi$  是它的根, 则  $a - bi$  也是它的根 (为什么?).

复系数多项式有更好的性质, 我们理应考虑复向量和复矩阵的对应概念. **如果不加特别说明, 本章将出现的矩阵都是复矩阵.** 前述矩阵运算、线性方程组、子空间、基和维数、行列式等概念和相应性质, 都可以毫不困难地从实向量和实矩阵推广到复向量和复矩阵.

复数的共轭运算也可以推广到复向量和复矩阵. 给定复向量  $\mathbf{x} = [x_j] \in \mathbb{C}^n$ , 则由它的每个分量的共轭组成的向量记为  $\overline{\mathbf{x}} = [\overline{x_j}]$ ; 给定复矩阵  $A = [a_{ij}]$ , 则由它的每个元素的共轭组成的矩阵记为  $\overline{A} = [\overline{a_{ij}}]$ .

**注 5.1.6** 实向量和实矩阵的概念中, 内积和正交性不能简单地推广到复向量和复矩阵. 例如, 对复向量  $\mathbf{x}$ ,  $\mathbf{x}^T \mathbf{x} = x_1^2 + \cdots + x_n^2 \geq 0$  不一定成立, 而  $\mathcal{R}(AA^T) = \mathcal{R}(A)$  也不一定成立. 复向量和复矩阵的对应概念, 将在第 8 章中阐述.

接下来, 本章将在  $\mathbb{C}^n$  上讨论方程  $A\mathbf{x} = \lambda\mathbf{x}$  及其相关概念.

## 习题

练习 5.1.1 证明定理 5.1.5.

## 5.2 基本概念

**定义 5.2.1 (特征值)** 给定  $n$  阶方阵  $A$ , 如果对  $\lambda \in \mathbb{C}$ , 存在非零向量  $\mathbf{x} \in \mathbb{C}^n$ , 使得  $A\mathbf{x} = \lambda\mathbf{x}$ , 则称数  $\lambda$  为方阵  $A$  (在  $\mathbb{C}$  上) 的一个**特征值**, 而称非零向量  $\mathbf{x}$  为方阵  $A$  (在  $\mathbb{C}$  上) 的一个属于特征值  $\lambda$  的**特征向量**.

二元组  $(\lambda, \mathbf{x})$  常称为方阵  $A$  的一个**特征对**.

特别地, 对实方阵  $A$ , 如果特征对  $(\lambda, \mathbf{x})$  满足  $\lambda \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$ , 则分别称二者为  $A$  在  $\mathbb{R}$  上的特征值和特征向量, 称该二元组为  $A$  在  $\mathbb{R}$  上的特征对.

注意:

1. 只有方阵才有特征值和特征向量;
2. 零向量**不是**特征向量;
3. 如果  $\mathbf{x}$  是特征向量, 则对任意  $k \neq 0$ ,  $k\mathbf{x}$  都是特征向量.

例 5.1.1 中的矩阵  $A$ , 由  $A\mathbf{x}_1 = \mathbf{x}_1, A\mathbf{x}_0 = 0.5\mathbf{x}_0$  可知, 1 和 0.5 是  $A$  的两个特征值,  $\mathbf{x}_1$  是属于  $\lambda = 1$  的特征向量,  $\mathbf{x}_2$  是属于  $\lambda = 0.5$  的特征向量. 两个特征值不同, 对应的特征向量线性无关.

例 5.2.2 对例 5.1.2 中的反射变换, 由  $H_v\mathbf{w} = \mathbf{w}, H_v\mathbf{v} = -\mathbf{v}$  可知, 1 和  $-1$  是  $H_v$  的两个特征值,  $\mathbf{w}$  是属于 1 的特征向量,  $\mathbf{v}$  是属于  $-1$  的特征向量. 两个特征值不同, 对应的特征向量线性无关.

对旋转变换, 当  $\theta \neq k\pi$  时, 没有实特征值, 因而也没有对应的特征向量.

对伸缩变换, 其表示矩阵为  $C_k = \begin{bmatrix} k & \\ & 1 \end{bmatrix}$ , 则  $C_k\mathbf{e}_1 = k\mathbf{e}_1, C_k\mathbf{e}_2 = \mathbf{e}_2$ , 即对角元素  $k, 1$  是  $C_k$  的特征值, 对应的特征向量分别是  $\mathbf{e}_1, \mathbf{e}_2$ . 特别地, 恒同矩阵  $I_2$  只有一个特征值 1, 任意非零向量都是属于 1 的特征向量, 其中  $\mathbf{e}_1, \mathbf{e}_2$  是两个线性无关的特征向量.

对错切变换, 其表示矩阵为  $S_k = \begin{bmatrix} 1 & \\ k & 1 \end{bmatrix}, k \neq 0$ . 设  $S_k\mathbf{x} = \lambda\mathbf{x}$ , 即  $\begin{bmatrix} 1 & \\ k & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ kx_1 + x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ . 如果  $x_2 \neq 0$ , 则  $\lambda = 1$ , 但  $x_1 + kx_2 \neq x_1$ , 因此无解, 即当  $x_2 \neq 0$  时, 向量在错切变换下的像不可能与原向量共线. 如果  $x_2 = 0$ , 易知  $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$  是  $S_k$  的属于特征值 1 的特征向量. 因此, 1 是  $S_k$  的特征值, 但只有一个线性无关的特征向量  $\mathbf{e}_2$ .  $\odot$

对任意  $n$  阶方阵  $A$ , 如何计算其特征值和特征向量? 根据定义 5.2.1,  $(\lambda_0, \mathbf{x}_0)$  是一个特征对, 当且仅当  $(\lambda_0 I_n - A)\mathbf{x}_0 = \mathbf{0}$ , 即  $\mathbf{x}_0 \in \mathcal{N}(\lambda_0 I_n - A)$ . 注意到, 特征向量  $\mathbf{x}_0 \neq \mathbf{0}$ , 因此  $\lambda_0$  是  $A$  的特征值当且仅当  $(\lambda_0 I_n - A)\mathbf{x}_0 = \mathbf{0}$  有非零解. 这等价于方阵  $\lambda_0 I_n - A$  不可逆, 也等价于其行列式为 0. 由此我们首先得到如下结论.

**命题 5.2.3** 数  $\lambda_0$  是  $A$  的特征值, 当且仅当  $\det(\lambda_0 I_n - A) = 0$ . 特别地, 0 是  $A$  的特征值当且仅当  $A$  不可逆.

多项式

$$p_A(\lambda) := \det(\lambda I_n - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \ddots & \vdots \\ \vdots & & \ddots & -a_{n-1,n} \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

称为矩阵  $A$  的**特征多项式**.

上述讨论总结如下.

**定理 5.2.4** 设  $n$  阶方阵  $A$  的特征多项式为  $p_A(\lambda)$ , 那么

1. 数  $\lambda_0$  是  $A$  的特征值, 当且仅当  $p_A(\lambda_0) = 0$ , 即  $\lambda_0$  是特征多项式  $p_A(\lambda)$  的根.
2. 向量  $\mathbf{x}_0 \in \mathbb{R}^n$  是  $A$  的属于  $\lambda_0$  的特征向量, 当且仅当  $\mathbf{x}_0 \in \mathcal{N}(\lambda_0 I_n - A)$  且  $\mathbf{x}_0 \neq \mathbf{0}$ , 即  $\mathbf{x}_0$  是  $\lambda_0 I_n - A$  的零空间中的非零向量.

解空间  $\mathcal{N}(\lambda_0 I_n - A)$  称为  $A$  的属于  $\lambda_0$  的**特征子空间**. 注意到特征向量不为零,

$$\{A \text{ 的属于 } \lambda_0 \text{ 的特征向量}\} = \mathcal{N}(\lambda_0 I_n - A) \setminus \{\mathbf{0}\}.$$

定理 5.2.4 给出了求解矩阵的特征值、特征子空间和特征向量的方法:

1. 先计算  $A$  的特征多项式  $p_A(\lambda)$ ;
2. 然后计算  $p_A(\lambda)$  的所有根, 即为  $A$  的全部特征值;
3. 如果有特征值, 则对每个特征值  $\lambda_i$ , 计算其特征子空间  $\mathcal{N}(\lambda_i I_n - A)$  的一组基  $\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,r_i}$ , 则属于  $\lambda_i$  的全部特征向量就是

$$\mathcal{N}(\lambda_i I_n - A) \setminus \{\mathbf{0}\} = \{k_1 \mathbf{x}_{i,1} + \dots + k_{r_i} \mathbf{x}_{i,r_i} \mid k_1, \dots, k_{r_i} \text{ 不全为 } 0\}.$$

简单修改上述算法的计算范围, 就可以计算实矩阵的在  $\mathbb{R}$  上的特征值、特征子空间和特征向量.

特征值的计算只与行列式  $\det(\lambda I_n - A)$  有关, 容易得到如下结论.

**命题 5.2.5** 上(下)三角矩阵的全部特征值就是其所有对角元素.

**例 5.2.6** 观察二阶上三角矩阵  $A = \begin{bmatrix} a_{11} & a_{12} \\ & a_{22} \end{bmatrix}$ . 显然,  $A$  的特征多项式是  $(\lambda - a_{11})(\lambda - a_{22})$ .

当  $a_{11} \neq a_{22}$  时,  $A$  的全部特征值是  $a_{11}, a_{22}$ , 二者对应的特征子空间分别为

$$\mathcal{N}(a_{11} I_2 - A) = \text{span}(\mathbf{e}_1), \quad \mathcal{N}(a_{22} I_2 - A) = \text{span}((a_{22} - a_{11})\mathbf{e}_2 + \mathbf{e}_1).$$

特别地,  $A$  有两个线性无关的特征向量  $\mathbf{e}_1, (a_{22} - a_{11})\mathbf{e}_2 + \mathbf{e}_1$ .

当  $a_{11} = a_{22}$  时,  $A$  的全部特征值是  $a_{11}$ . 如果  $a_{12} = 0$ , 则其对应的特征子空间为  $\mathcal{N}(a_{11} I_2 - A) = \mathbb{R}^2$ . 特别地,  $A$  有两个线性无关的特征向量  $\mathbf{e}_1, \mathbf{e}_2$ .

如果  $a_{12} \neq 0$ , 则其对应的特征子空间为  $\mathcal{N}(a_{11} I_2 - A) = \text{span}(\mathbf{e}_1)$ .  $A$  只有一个线性无关的特征向量  $\mathbf{e}_1$ .

可以看到, 对对角矩阵和上三角矩阵, 尽管特征值都是对角元素, 但对角元素是否相等对特征子空间的影响很大. 对角元素从不等到相等, 对对角矩阵, 特征子空间从两个一维子空间变成一个二维子空间; 而对非对角的三角矩阵, 特征子空间从两个一维子空间变成一个一维子空间. ☺

**例 5.2.7** 例 5.1.2 中的旋转变换的表示矩阵有两个复特征值  $\lambda_1 = \cos \theta + i \sin \theta, \lambda_2 = \cos \theta - i \sin \theta$ . 而

$$\lambda_1 I_2 - A = \begin{bmatrix} i \sin \theta & \sin \theta \\ \sin \theta & i \sin \theta \end{bmatrix}, \quad \lambda_2 I_2 - A = \begin{bmatrix} i \sin \theta & \sin \theta \\ \sin \theta & i \sin \theta \end{bmatrix},$$

其中  $\theta \neq k\pi, \sin \theta \neq 0$ , 因此对应的特征子空间分别为

$$\mathcal{N}(\lambda_1 I_2 - A) = \text{span} \left( \begin{bmatrix} 1 \\ -i \end{bmatrix} \right), \quad \mathcal{N}(\lambda_2 I_2 - A) = \text{span} \left( \begin{bmatrix} 1 \\ i \end{bmatrix} \right).$$

当把该线性变换看作是复线性空间  $\mathbb{C}^2$  上的线性变换时, 表示矩阵有两个线性无关的特征向量  $\begin{bmatrix} 1 \\ -i \end{bmatrix}, \begin{bmatrix} 1 \\ i \end{bmatrix}$ . ☺

**例 5.2.8** 设  $A = \begin{bmatrix} 7 & -5 & 5 \\ -2 & 3 & -2 \\ -10 & 10 & -8 \end{bmatrix}$ . 求  $A$  在  $\mathbb{C}$  上和在  $\mathbb{R}$  上的特征值和特征向量.

先计算特征多项式

$$\begin{aligned} p_A(\lambda) &= \det(\lambda I - A) \\ &= \begin{vmatrix} \lambda - 7 & 5 & -5 \\ 2 & \lambda - 3 & 2 \\ 10 & -10 & \lambda + 8 \end{vmatrix} = \begin{vmatrix} \lambda - 2 & 5 & -5 \\ \lambda - 1 & \lambda - 3 & 2 \\ 0 & -10 & \lambda + 8 \end{vmatrix} \\ &= (\lambda - 2) \begin{vmatrix} \lambda - 3 & 2 \\ -10 & \lambda + 8 \end{vmatrix} - (\lambda - 1) \begin{vmatrix} 5 & -5 \\ -10 & \lambda + 8 \end{vmatrix} \\ &= (\lambda - 2) ((\lambda - 3)(\lambda + 8) - 2(-10)) - (\lambda - 1) \begin{vmatrix} 5 & 0 \\ -10 & \lambda - 2 \end{vmatrix} \\ &= (\lambda - 2)(\lambda^2 + 1). \end{aligned}$$

矩阵  $A$  只有三个特征值  $\lambda_1 = 2, \lambda_2 = i, \lambda_3 = -i$ , 其中  $\lambda_1$  是实特征值.

再来计算特征向量. 当  $\lambda_1 = 2$  时,  $\lambda_1 I_3 - A = \begin{bmatrix} -5 & 5 & -5 \\ 2 & -1 & 2 \\ 10 & -10 & 10 \end{bmatrix}$ , 特征子空间

$$\mathcal{N}(\lambda_1 I_3 - A) \text{ 的一组基 } \mathbf{x}_{1,1} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$



当  $\lambda_2 = i$  时,  $\lambda_2 I_3 - A = \begin{bmatrix} -7+i & 5 & -5 \\ 2 & -3+i & 2 \\ 10 & -10 & 8+i \end{bmatrix}$ , 特征子空间  $\mathcal{N}(\lambda_2 I_3 - A)$  的一

$$\text{组基 } \mathbf{x}_{2,1} = \begin{bmatrix} 1-2i \\ -1+i \\ -2+4i \end{bmatrix}.$$

当  $\lambda_3 = -i$  时,  $\lambda_3 I_3 - A = \begin{bmatrix} -7-i & 5 & -5 \\ 2 & -3-i & 2 \\ 10 & -10 & 8-i \end{bmatrix}$ , 特征子空间  $\mathcal{N}(\lambda_3 I_3 - A)$  的

$$\text{一组基 } \mathbf{x}_{3,1} = \begin{bmatrix} 1+2i \\ -1-i \\ -2-4i \end{bmatrix}.$$

注意, 虽然无论在  $\mathbb{C}$  上还是在  $\mathbb{R}$  上考虑, 矩阵  $A$  的属于  $\lambda_1 = 2$  的特征子空间都可由  $\mathbf{x}_{1,1}$  生成, 但二者并不相同. 矩阵  $A$  的在  $\mathbb{R}$  上的属于 2 的所有特征向量组成的集合是  $\{k\mathbf{x}_{1,1} \mid k \in \mathbb{R}, k \neq 0\}$ ; 矩阵  $A$  的在  $\mathbb{C}$  上的属于 2 的所有特征向量组成的集合是  $\{k\mathbf{x}_{1,1} \mid k \in \mathbb{C}, k \neq 0\}$ .  $\odot$

特征值的大概范围能够通过对角元素和非对角元素确定. 换言之, 能得到一个区域, 区域中含有矩阵的特征值.

**定理 5.2.9 (Gershgorin 圆盘定理)** 对矩阵  $A = [a_{ij}]_{n \times n}$ , 定义如下  $n$  个圆盘

$$G_i(A) := \left\{ z \mid |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\},$$

那么矩阵的任意特征值  $\lambda$  一定落在某个圆盘中. 因此, 矩阵的全部特征值一定落在这  $n$  个圆盘中.

证. 对任意特征值  $\lambda_0$ , 考虑矩阵  $A - \lambda_0 I$ . 如果  $\lambda_0$  不在任意一个圆盘中, 那么对  $i = 1, \dots, n$ ,  $|a_{ii} - \lambda_0| > \sum_{j \neq i} |a_{ij}|$ , 即  $A - \lambda_0 I$  是对角占优矩阵. 根据命题 1.6.14, 对角占优矩阵可逆, 于是  $A - \lambda_0 I$  可逆. 这与特征值的定义矛盾.  $\square$

最后, 由于  $\lambda_0$  是矩阵  $A$  的特征值当且仅当  $\lambda_0$  是特征多项式  $p_A(\lambda)$  的根, 通过特征多项式能得到特征值的一些性质.

**定义 5.2.10 (代数重数)** 给定  $n$  阶方阵  $A$  及  $A$  的一个特征值  $\lambda_0 \in \mathbb{C}$ , 如果  $\lambda_0$  是特征多项式  $p_A(\lambda)$  的  $n_0$  重根, 则称  $n_0$  为  $\lambda_0$  作为  $A$  的特征值的**代数重数** (简称**重数**), 称  $\lambda_0$  是  $A$  的一个  $n_0$  重特征值.

一个 1 重特征值, 又称为**单特征值**.

**命题 5.2.11** 给定  $n$  阶实方阵  $A$ , 如果  $\lambda_0$  是它的一个非实数特征值, 则  $\bar{\lambda}_0$  也是它的特征值, 且其代数重数和  $\lambda_0$  的代数重数相等. 进一步地, 如果复向量  $\mathbf{x}_0$  是属于  $\lambda_0$  的特征向量, 则  $\bar{\mathbf{x}}_0$  是属于  $\bar{\lambda}_0$  的特征向量.

证. 关于特征值的结论显然. 对特征向量, 因为  $\bar{A} = A$ , 而共轭保持加法和乘法运算, 因此  $A\mathbf{x}_0 = \lambda_0\mathbf{x}_0$  当且仅当  $A\bar{\mathbf{x}}_0 = \bar{\lambda}_0\bar{\mathbf{x}}_0$ .  $\square$

例 5.2.8 中的两个特征向量就互为共轭:  $\mathbf{x}_{3,1} = \bar{\mathbf{x}}_{2,1}$ .

**命题 5.2.12** 给定  $n$  阶方阵  $A$ , 其特征多项式具有如下形式:

$$p_A(\lambda) = \lambda^n - \text{trace}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A),$$

其中  $\text{trace}(A) = a_{11} + a_{22} + \cdots + a_{nn}$  是方阵  $A$  的对角元素的和, 称为方阵  $A$  的迹.

证. 根据定义,

$$p_A(\lambda) := \det(\lambda I_n - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{m1} & -a_{m2} & \cdots & \lambda - a_{mn} \end{vmatrix}.$$

我们在前面已经说明它是一个首项系数为 1 的  $n$  次多项式.

现在考虑其中  $\lambda^{n-1}$  的系数. 在完全展开式包含的  $n!$  项中, 每一项都是矩阵  $\lambda I_n - A$  中  $n$  个元素的乘积, 且因数中每行只有一个元素, 每列只有一个元素. 而只有对角元素包含  $\lambda$ , 因此对应项的因数至少需要从对角线上的  $n$  个元素中选取  $n-1$  个, 这使得剩下的一个元素也必须是对角元素. 于是,  $p_A(\lambda)$  与对角元素的乘积  $(\lambda - a_{11}) \cdots (\lambda - a_{nn})$  二者的  $\lambda^{n-1}$  的系数相等. 容易计算后者的相应系数等于  $-(a_{11} + \cdots + a_{nn}) = -\text{trace}(A)$ .

最后考虑常数项, 代入  $\lambda = 0$  即得  $p_A(0) = \det(-A) = (-1)^n \det(A)$ .  $\square$

根据推论 5.1.4,  $n$  阶方阵有且只有  $n$  个特征值 (计重数). 利用 Vieta 定理 5.1.5, 有

$$\lambda_1 + \cdots + \lambda_n = \text{trace}(A), \quad \lambda_1 \cdots \lambda_n = \det(A).$$

最后作一简单注记. 前述求解矩阵的特征值和特征向量的纸面方法, 即先求特征多项式的根以得到特征值, 再求解线性方程组以得到特征向量, 在实践中难以实现. 抽象代数知识表明<sup>2</sup>, 利用有限步代数运算 (加减乘除乘方开方) 不能准确地计算出高阶多项

<sup>2</sup>五次以上的一元高次方程没有求根公式, 这是抽象代数在历史上首个重要结论. 对这一命题的研究改变了代数学的面貌, 使得代数学不再只是研究求解代数方程的数学分支.

式的根. 在实际生产生活中, 计算稠密型矩阵的特征值问题时, 往往会利用 Francis 算法<sup>3</sup>计算特征值的近似值.

注意, 对任意首项系数为 1 的多项式

$$p(x) = c_0 + c_1x + \cdots + c_{n-1}x^{n-1} + x^n,$$

令

$$\mathbf{y} = \begin{bmatrix} 1 \\ \lambda \\ \vdots \\ \lambda^{n-1} \end{bmatrix}, \quad C_p = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ -c_0 & \cdots & -c_{n-2} & -c_{n-1} \end{bmatrix},$$

容易验证  $C_p \mathbf{y} = \lambda \mathbf{y}$  当且仅当  $p(\lambda) = 0$ . 这说明  $\lambda$  是多项式  $p$  的根, 当且仅当  $\lambda$  是  $C_p$  的特征值. 事实上,  $C_p$  的特征多项式是  $p(x)$  (为什么?). 矩阵  $C_p$  称为多项式  $p(x)$  的友矩阵. 因此在实践中, 任意多项式的根的近似值, 就可以通过求解其友矩阵的特征值来得到.

## 习题

**练习 5.2.1** 求下列复矩阵的全部特征值和特征向量.

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

2.  $\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}.$

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

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

5.  $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$

**练习 5.2.2** 构造符合要求的矩阵  $A$ .

1.  $A$  的特征多项式为  $\lambda^2 - 9\lambda + 20$ , 构造三个不同的  $A$ .

2.  $A = \begin{bmatrix} 0 & 1 \\ * & * \end{bmatrix}$ , 且  $A$  的特征值为 4, 7.

3.  $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ * & * & * \end{bmatrix}$ , 且  $A$  的特征值为 1, 2, 3.

**练习 5.2.3** 设  $A = \begin{bmatrix} 2 & -2 & 0 \\ -2 & x & -2 \\ -2 & -2 & 0 \end{bmatrix}$ ,  $B = \begin{bmatrix} 2 & & \\ & 2 & \\ & & y \end{bmatrix}$ , 已知  $A, B$  特征多项式相同, 求  $x, y$ .

<sup>3</sup>这一算法是数值线性代数的基础内容, 其设计十分精巧, 被认为是 20 世纪对科学和工程领域产生最大影响力的十大算法之一.

**练习 5.2.4** 设  $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & a \end{bmatrix}$  可逆, 且  $\mathbf{x} = \begin{bmatrix} 1 \\ b \\ 1 \end{bmatrix}$  是  $A$  的特征向量, 求  $a, b$  的值.

**练习 5.2.5** 设  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ .

1. 利用一元二次方程求根公式, 写出  $A$  的两个特征值  $\lambda_1, \lambda_2$  的表达式.
2. 构造一个非对角的  $A$ , 满足  $\lambda_1 = \lambda_2$ .
3. 设  $\lambda$  为  $A$  的特征值, 证明  $A \begin{bmatrix} b \\ \lambda - a \end{bmatrix} = \lambda \begin{bmatrix} b \\ \lambda - a \end{bmatrix}$ ,  $A \begin{bmatrix} \lambda - d \\ c \end{bmatrix} = \lambda \begin{bmatrix} \lambda - d \\ c \end{bmatrix}$ . 提示: 如果这两个向量不是零向量, 那么它们就是特征向量.
4. 若上述两个向量中有且仅有一个是零向量, 求  $A$  的特征值和特征向量.
5. 若上述两个向量都是零向量, 求  $A$  的特征值和特征向量.

**练习 5.2.6** 给定向量  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ , 计算  $A = \mathbf{a}\mathbf{b}^T$  的所有特征对.

**练习 5.2.7** 已知  $A$  的特征值为  $1, 2, 3$ , 求下列矩阵的特征值.

1.  $2A, A + I_3, A^2, \bar{A}, A^T, A^{-1}$  ( $A$  为何可逆?).
2.  $\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}, \begin{bmatrix} A & A \\ 0 & A \end{bmatrix}$ .

**练习 5.2.8** 证明, 如果  $(\lambda, \mathbf{x})$  是  $A$  的特征对, 则  $(f(\lambda), \mathbf{x})$  是  $f(A)$  的特征对, 其中  $f(x)$  是任意多项式.

**练习 5.2.9** 设  $A$  是可逆矩阵, 证明,  $A$  的特征值都不为 0; 若  $\lambda_0$  是  $A$  的一个特征值, 则  $\frac{1}{\lambda_0}$  是  $A^{-1}$  的一个特征值.

**练习 5.2.10** 设  $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3 \in \mathbb{R}^3$  为一组标准正交基, 分别求  $\mathbf{q}_1\mathbf{q}_1^T, \mathbf{q}_1\mathbf{q}_1^T + \mathbf{q}_2\mathbf{q}_2^T, \mathbf{q}_1\mathbf{q}_1^T + \mathbf{q}_2\mathbf{q}_2^T + \mathbf{q}_3\mathbf{q}_3^T$  所有特征值和特征向量.

**练习 5.2.11** 设  $n$  阶方阵  $A$  满足  $A^T\mathbf{v} = \lambda\mathbf{v}$ , 其中  $\mathbf{v} \neq \mathbf{0}$ .

1. 设  $A\mathbf{w} = \mu\mathbf{w}$ , 且  $\lambda \neq \mu$ , 证明  $\mathbf{v}, \mathbf{w}$  正交.
2. 证明, 对称矩阵的属于不同特征值的特征向量正交.

**练习 5.2.12** 证明:

1. 若存在正整数  $k$ , 使得  $A^k = O$ , 则  $A$  的特征值只能是 0.
2. 若  $A^2 = I_n$ , 则  $A$  的特征值只能是 1 或 -1.
3. 若  $A^2 = A$ , 则  $A$  的特征值只能是 1 或 0.

**练习 5.2.13** 对方阵  $A$ , 若多项式  $f(x)$  满足  $f(A) = O$ , 则称  $f(x)$  是  $A$  的化零多项式.

给定  $A$  的特征值  $\lambda$ , 证明, 若  $f(x)$  是  $A$  的化零多项式, 则  $f(\lambda) = 0$ .

**练习 5.2.14** 设方阵  $A, B$  可交换,  $\lambda_0$  是  $A$  的一个特征值,  $V_{\lambda_0}$  是  $A$  的特征值为  $\lambda_0$  的特征子空间. 证明, 对任意  $\mathbf{x} \in V_{\lambda_0}$ , 都有  $B\mathbf{x} \in V_{\lambda_0}$ . 当  $A, B$  不可交换时, 结论是否成立?

**练习 5.2.15** 证明,  $A$  和  $T^{-1}AT$  具有相同的特征多项式.

**练习 5.2.16** 设  $\lambda_1, \lambda_2$  是  $A$  的两个不同特征值,  $\boldsymbol{x}_1, \boldsymbol{x}_2$  是分别属于  $\lambda_1, \lambda_2$  的特征向量. 证明,  $\boldsymbol{x}_1 + \boldsymbol{x}_2$  不是  $A$  的特征向量.

**练习 5.2.17** 证明或举出反例.

1. 如果  $A, B$  具有相同的特征值、代数重数和特征向量, 则  $A = B$ .
2. 如果  $A, B$  有相同的特征值和代数重数, 则  $A - B$  所有特征值之和为零.
3.  $A + B$  的特征值等于  $A$  与  $B$  的特征值之和.
4.  $AB$  的特征值是  $A$  与  $B$  的特征值之积.
5.  $AB$  和  $BA$  具有相同的特征值和代数重数.
6. 如果  $A$  的特征值全为零, 则  $A$  是零矩阵.
7. 将  $A$  的第  $i$  行加到第  $j$  行上, 再将第  $i$  列从第  $j$  列中减去, 得到的矩阵  $B$  和  $A$  有相同的特征值. 若正确, 则对应的特征向量有何联系?
8. 将  $A$  的第  $i$  行加到第  $j$  行上, 再将第  $j$  列从第  $i$  列中减去, 得到的矩阵  $B$  和  $A$  有相同的特征值. 若正确, 则对应的特征向量有何联系?
9. 将  $A$  的第  $i, j$  行交换, 再将第  $i, j$  列交换, 得到的矩阵  $B$  和  $A$  有相同的特征值. 若正确, 则对应的特征向量有何联系?
10. 对角阵的特征向量一定是标准基向量.
11. 正交矩阵的特征值都是绝对值等于 1 的复数.
12. 所有  $n$  阶置换矩阵都有一个共同的特征向量.

**练习 5.2.18** 设  $A, B$  分别是  $m \times n, n \times m$  矩阵, 证明,

$$\lambda^n \det(\lambda I_m - AB) = \lambda^m \det(\lambda I_n - BA).$$

特别地, 当  $m = n$  时,  $\det(\lambda I_n - AB) = \det(\lambda I_n - BA)$ .

**练习 5.2.19** 如果复矩阵  $A, B$  可交换, 证明  $A, B$  至少有一个公共的特征向量.

**练习 5.2.20** 设  $A$  是四阶数独矩阵 (见练习 2.4.9), 证明其绝对值最大的特征值为 10, 且属于该特征值的特征向量的所有分量都相等.

**练习 5.2.21** 设方阵  $A$  的每个元素都是整数, 证明  $\frac{1}{2}$  一定不是  $A$  的特征值.

**练习 5.2.22** 回顾例 5.1.1 中的矩阵  $A = \begin{bmatrix} 0.6 & 0.1 \\ 0.4 & 0.9 \end{bmatrix}$ , 以及稳定状态对应的矩阵  $A^\infty = \begin{bmatrix} 0.2 & 0.2 \\ 0.8 & 0.8 \end{bmatrix}$ .

1. 如果  $A^n$  和  $A^\infty$  中对应的元素相差不超过 0.01, 那么  $n$  至少是多少?
2. 交换  $A$  的两行, 特征值是否不变?

**练习 5.2.23** 给定  $m$  阶方阵  $A_1$ ,  $n$  阶上三角矩阵  $A_2$  和  $m \times n$  矩阵  $B$ . 证明如果  $A_1$  和  $A_2$  没有相同的特征值, 关于  $m \times n$  矩阵  $X$  的矩阵方程  $A_1 X - X A_2 = B$  有唯一解.

矩阵方程  $A_1 X - X A_2 = B$  称为 **Sylvester 方程**, 在控制论中有不少应用.

### 5.3 对角化和谱分解

在第 5.1 节中, 我们已经看到, 如果线性空间存在一组基  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , 且基向量  $\mathbf{x}_i$  是矩阵  $A$  的属于  $\lambda_i$  的特征向量, 那么  $A$  就可以写作  $A = X\Lambda X^{-1}$ , 其中  $X = [\mathbf{x}_1 \ \dots \ \mathbf{x}_n]$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ . 这种分解对考虑矩阵的性质有很大便利. 首先, 这种分解利于计算矩阵的幂和逆 (如果存在). 容易计算, 对方阵  $A$ ,  $A^t = X\Lambda^t X^{-1}$ ,  $A^{-1} = X\Lambda^{-1} X^{-1}$ . 其次, 这种分解利于分析矩阵对应的线性变换的性质. 例 5.1.1 中的二阶矩阵的两个特征值分别是 1 和 0.5, 因此线性变换  $A^t$  的长期趋势由属于特征值 1 的特征向量决定, 即对任意初始向量  $\mathbf{x} = k_1\mathbf{x}_1 + k_2\mathbf{x}_2$ , 当  $t \rightarrow \infty$  时,  $A^t\mathbf{x} = k_1\mathbf{x}_1 + 0.5^t k_2\mathbf{x}_2 \rightarrow k_1\mathbf{x}_1$ . 事实上,  $A^t \rightarrow \mathbf{x}_1 \mathbf{e}_1^T X^{-1}$ .

考虑  $n$  阶方阵  $A$  对应的线性变换在向量  $\mathbf{x}$  上的作用, 由于  $\mathbf{x}_1, \dots, \mathbf{x}_n$  是一组基,  $\mathbf{x}$  可以被其线性表示  $\mathbf{x} = k_1\mathbf{x}_1 + k_2\mathbf{x}_2 + \dots + k_n\mathbf{x}_n$ , 如果  $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$ , 则

$$\begin{aligned} A^t\mathbf{x} &= k_1\lambda_1^t\mathbf{x}_1 + k_2\lambda_2^t\mathbf{x}_2 + \dots + k_n\lambda_n^t\mathbf{x}_n \\ &= \lambda_1^t \left( k_1\mathbf{x}_1 + k_2 \left( \frac{\lambda_2}{\lambda_1} \right)^t \mathbf{x}_2 + \dots + k_n \left( \frac{\lambda_n}{\lambda_1} \right)^t \mathbf{x}_n \right) \approx \lambda_1^t k_1\mathbf{x}_1. \end{aligned}$$

这意味着, 特征值的相对大小直接影响线性变换  $A^t$  的长期趋势.

容易看到, 这些好处来源于  $A$  是对角矩阵, 为此作如下定义.

**定义 5.3.1 (谱分解)** 对方阵  $A$ , 如果存在可逆矩阵  $X$  使得  $X^{-1}AX = \Lambda$  是对角矩阵, 则称  $A$  是 (在  $\mathbb{C}$  上) **可对角化** 的,  $X$  把  $A$  对角化, 或  $X$  对角化  $A$ .

如果方阵  $A, X, \Lambda$  都是实矩阵, 则称  $A$  在  $\mathbb{R}$  上可对角化.

当  $A$  可对角化时, 分解  $A = X\Lambda X^{-1}$  称为  $A$  的**谱分解**<sup>4</sup>.

对角矩阵显然可对角化, 可以取  $X = I_n$ .

我们已经知道, 有  $n$  个线性无关的特征向量的矩阵可对角化. 对角矩阵总有  $n$  个线性无关的特征向量  $\mathbf{e}_1, \dots, \mathbf{e}_n$ . 注意到,  $n$  阶矩阵有  $n$  个特征值 (计重数), 因此非对角的矩阵也可能有  $n$  个线性无关的向量, 如例 5.1.1, 5.2.7 和 5.2.8 中的矩阵. 然而, 并非任意矩阵都有  $n$  个线性无关的特征向量. 而例 5.2.2 中的二阶矩阵, 就没有两个线性无关的特征向量.

那么何种矩阵可对角化, 从而有谱分解? 如下结论说明, 只有有  $n$  个线性无关的特征向量的矩阵才可对角化.

**命题 5.3.2** 对  $n$  阶方阵  $A$ ,  $A$  可对角化, 当且仅当  $A$  有  $n$  个线性无关的特征向量.

<sup>4</sup>之所以称为谱分解, 是因为特征值也称为谱.

证. “ $\Leftarrow$ ”: 前面讨论已得. “ $\Rightarrow$ ”: 如果  $A$  可对角化, 那么存在可逆矩阵  $X = [\mathbf{x}_1 \cdots \mathbf{x}_n]$  使得  $X^{-1}AX = \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$  是对角矩阵. 于是  $A[\mathbf{x}_1 \cdots \mathbf{x}_n] = AX = X\Lambda = [\lambda_1 \mathbf{x}_1 \cdots \lambda_n \mathbf{x}_n]$ . 因此  $A\mathbf{x}_i = \lambda_i \mathbf{x}_i, i = 1, \cdots, n$ . 而  $X$  可逆, 因此  $\mathbf{x}_1, \cdots, \mathbf{x}_n$  线性无关, 这意味着  $A$  有  $n$  个线性无关的特征向量.  $\square$

命题 5.3.2 的证明说明, 如果矩阵  $A$  可对角化, 则  $A$  存在谱分解  $A = X\Lambda X^{-1}$ , 且  $\Lambda$  的对角元素就是  $A$  的特征值,  $X$  的第  $i$  列就是  $A$  的属于  $\lambda_i$  的第  $i$  个对角元素的特征向量.

判断一个矩阵是否可对角化, 等价于判断它是否有  $n$  个线性无关的特征向量. 那么, 矩阵的特征向量何时线性无关?

**命题 5.3.3** 方阵的属于不同特征值的特征向量线性无关.

证. 设方阵  $A$  有特征向量  $\mathbf{x}_1, \cdots, \mathbf{x}_r$ , 分别属于特征值  $\lambda_1, \cdots, \lambda_r$ , 且  $\lambda_1, \cdots, \lambda_r$  两两不同. 采用数学归纳法. 当  $r = 1$  时, 因为特征向量不为零, 因此线性无关. 现假设任意  $r - 1$  个不同特征值的特征向量都线性无关. 观察方程  $k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2 + \cdots + k_r \mathbf{x}_r = \mathbf{0}$ . 等式两边左乘  $A$ , 则有  $\mathbf{0} = A(k_1 \mathbf{x}_1 + k_2 \mathbf{x}_2 + \cdots + k_r \mathbf{x}_r) = k_1 \lambda_1 \mathbf{x}_1 + k_2 \lambda_2 \mathbf{x}_2 + \cdots + k_r \lambda_r \mathbf{x}_r$ . 再减去原方程的  $\lambda_1$  倍, 就有  $k_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + k_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0}$ . 根据归纳假设,  $\mathbf{x}_2, \cdots, \mathbf{x}_r$  线性无关, 于是  $k_i(\lambda_i - \lambda_1) = 0, i = 2, \cdots, r$ . 由于  $\lambda_1$  和  $\lambda_2, \cdots, \lambda_r$  不同, 因此  $k_i = 0, i = 2, \cdots, r$ . 又得  $k_1 \mathbf{x}_1 = \mathbf{0}$ , 由特征向量不为零得  $k_1 = 0$ . 故  $\mathbf{x}_1, \cdots, \mathbf{x}_r$  线性无关.  $\square$

由此立得矩阵可对角化的一个充分条件.

**推论 5.3.4** 有  $n$  个不同特征值的  $n$  阶方阵, 即特征值都是单特征值的方阵, 可对角化.

显然这不是必要条件, 例如单位矩阵只有一个特征值 1, 但它可对角化.

**例 5.3.5** 空间  $\mathbb{R}^2$  上旋转变换的表示矩阵  $R_\theta$  在  $\mathbb{C}$  上可对角化, 因为它有两个不同的复特征值  $\pm i$ ; 但它在  $\mathbb{R}$  上不可对角化, 因为它没有实特征值.

空间  $\mathbb{R}^2$  上错切变换的表示矩阵  $S_k$  只有一个特征值 1, 且没有两个线性无关的特征向量, 因此不可对角化.  $\odot$

当方阵  $A$  有代数重数大于 1 的特征值时, 是否可对角化会变得复杂. 为此作如下定义.

**定义 5.3.6 (几何重数)** 给定  $n$  阶方阵  $A$  及其特征值  $\lambda_0$ , 称特征子空间  $\mathcal{N}(\lambda_0 I_n - A)$  的维数为  $\lambda_0$  作为  $A$  的特征值的**几何重数**.

任意特征值的几何重数都不小于 1, 因此特征值至少对应一个特征向量.

**命题 5.3.7** 方阵的特征值的几何重数不大于其代数重数.

证. 设  $\lambda_0$  是  $n$  阶方阵  $A$  的几何重数是  $r$  的一个特征值, 那么存在  $r$  个线性无关的特征向量  $\mathbf{x}_1, \dots, \mathbf{x}_r$ , 使得  $A\mathbf{x}_i = \lambda_0\mathbf{x}_i, i = 1, \dots, r$ . 则  $\mathbf{x}_1, \dots, \mathbf{x}_r$  可以扩充成  $\mathbb{C}^n$  的一组基  $\mathbf{x}_1, \dots, \mathbf{x}_r, \mathbf{y}_{r+1}, \dots, \mathbf{y}_n$ .

记  $X_{\lambda_0} := [\mathbf{x}_1 \ \dots \ \mathbf{x}_r], Y := [\mathbf{y}_{r+1} \ \dots \ \mathbf{y}_n], X := [X_{\lambda_0} \ Y]$ , 而  $X$  可逆. 由于  $X$  的列是一组基,  $AX$  的列能被其线性表示, 于是存在方阵  $M$ , 使得  $AX = XM$ . 由于  $X_{\lambda_0}$  的列是  $A$  的属于  $\lambda_0$  的特征向量, 所以  $M$  具有如下分块形式:

$$AX = A[X_{\lambda_0} \ Y] = [X_{\lambda_0} \ Y] \begin{bmatrix} \lambda_0 I_r & M_{12} \\ O & M_{22} \end{bmatrix} = XM.$$

而  $A$  和  $M = X^{-1}AX$  具有相同的特征多项式, 因为

$$\det(\lambda I_n - A) = \det(X^{-1}(\lambda I_n - A)X) = \det(\lambda I_n - M).$$

另一方面,

$$\det(\lambda I_n - M) = \det \left( \begin{bmatrix} \lambda I_r - \lambda_0 I_r & -M_{12} \\ O & \lambda I_{n-r} - M_{22} \end{bmatrix} \right) = (\lambda - \lambda_0)^r \det(\lambda I_{n-r} - M_{22}).$$

这说明  $\lambda_0$  是  $A$  的一个代数重数至少  $r$  的特征值.  $\square$

**定义 5.3.8** 几何重数和代数重数相等的特征值, 称为**半单特征值**. 几何重数小于代数重数的特征值, 称为**亏损特征值**.

如果一个特征值的代数重数是 1, 那么由几何重数不小于 1 可知, 它是半单特征值, 即单特征值是半单特征值.

利用几何重数, 就能得到矩阵可对角化的等价条件.

**定理 5.3.9** 1.  $n$  阶方阵  $A$  可对角化, 当且仅当其特征值都半单.

2.  $n$  阶实方阵  $A$  在  $\mathbb{R}$  上可对角化, 当且仅当其特征多项式的根都是实根, 且其特征值都半单.

证. 第 1 条: “ $\Rightarrow$ ”: 如果  $A$  可对角化, 设  $\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,k_1}, \dots, \mathbf{x}_{s,1}, \dots, \mathbf{x}_{s,k_s}$  是  $n$  个线性无关的特征向量, 满足  $A\mathbf{x}_{i,j} = \lambda_i \mathbf{x}_{i,j}$ . 因此,  $\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,k_i}$  是特征子空间  $\mathcal{N}(\lambda_i I_n - A)$  中一个线性无关的向量组, 于是  $k_i$  小于等于  $\lambda_i$  的几何重数  $m_i$ , 且  $\sum_{1 \leq i \leq s} k_i = n$ . 设  $n_i$  是  $\lambda_i$  的代数重数, 则  $\sum_{1 \leq i \leq s} n_i = n$ . 由命题 5.3.7 可知,  $k_i \leq m_i \leq n_i$ . 比较两个和式可知  $k_i = m_i = n_i$ , 即任意特征值半单.



“ $\Leftarrow$ ”: 如果  $A$  的特征值都半单, 设  $\lambda_1, \dots, \lambda_s$  是所有特征值, 则几何重数  $m_i$  等于代数重数  $n_i$ , 因此  $\sum_{1 \leq i \leq s} m_i = n$ . 取每个特征子空间  $\mathcal{N}(\lambda_i I_n - A)$  的一组基  $\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,m_i}$ , 合并起来的向量组  $\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,m_1}, \dots, \mathbf{x}_{s,1}, \dots, \mathbf{x}_{s,m_s}$  仍旧线性无关, 证明与命题 5.3.3 的证明类似. 而合并得到的向量组中有  $\sum_{1 \leq i \leq s} m_i = n$  个向量, 即  $A$  有  $n$  个线性无关的特征向量, 因此可对角化.

第 2 条: 证明与第 1 条类似, 留给读者.  $\square$

**例 5.3.10** 单位矩阵  $I_n$  只有一个特征值 1, 其代数重数和几何重数都是  $n$ , 1 是半单特征值. 该矩阵显然可对角化.

矩阵  $S_k = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$  ( $k \neq 0$ ), 只有一个特征值 1, 其代数重数是 2, 几何重数是 1, 1 是亏损特征值. 该矩阵不可对角化.

矩阵  $J_n(\lambda) := \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}_{n \times n}$  只有一个特征值  $\lambda$ , 其代数重数是  $n$ , 几何重数

是 1, 当  $n > 1$  时  $\lambda$  是亏损特征值. 形如  $J_n(\lambda)$  的矩阵称为关于  $\lambda$  的  $n$  阶 **Jordan 块**. 当  $n > 1$  时, 它不可对角化.  $\odot$

下面通过分块对角矩阵的对角化, 来帮助理解特征值半单概念.

**命题 5.3.11** 设分块对角矩阵  $A = \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{bmatrix}$ , 其中  $A_i, i = 1, \dots, r$  都是方阵, 则  $A$  可对角化当且仅当所有  $A_i$  都可对角化.

证. 不妨设  $r = 2$ , 否则反复应用  $r = 2$  的情形可得.

充分性: 如果  $A_1, A_2$  都可对角化, 则存在可逆矩阵  $X_1, X_2$ , 使得  $X_1^{-1}A_1X_1 = D_1$ ,  $X_2^{-1}A_2X_2 = D_2$  都是对角矩阵. 则

$$\begin{bmatrix} X_1^{-1} & \\ & X_2^{-1} \end{bmatrix} \begin{bmatrix} A_1 & \\ & A_2 \end{bmatrix} \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix} = \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix}$$

也是对角矩阵, 而  $X = \begin{bmatrix} X_1 & \\ & X_2 \end{bmatrix}$  可逆, 且  $X^{-1} = \begin{bmatrix} X_1^{-1} & \\ & X_2^{-1} \end{bmatrix}$ .

必要性: 首先  $p_A(\lambda) = p_{A_1}(\lambda)p_{A_2}(\lambda)$ , 因此  $A_1, A_2$  的特征值必为  $A$  的特征值, 于是只需讨论特征值的重数. 设  $\lambda_0$  是  $A$  的一个特征值,  $n_0, m_0$  分别为它的代数重数和几

何重数, 则  $n_0 = m_0$ . 考虑代数重数, 若  $\lambda_0$  分别是  $A_1, A_2$  的  $n_{01}, n_{02}$  重特征值, 则  $n_{01} + n_{02} = n_0$ . 考虑几何重数, 容易证明

$$\dim \mathcal{N} \left( \begin{bmatrix} \lambda_0 I_{r_1} - A_1 & \\ & \lambda_0 I_{r_2} - A_2 \end{bmatrix} \right) = \dim \mathcal{N}(\lambda_0 I_{r_1} - A_1) + \dim \mathcal{N}(\lambda_0 I_{r_2} - A_2).$$

设  $m_{01} = \dim \mathcal{N}(\lambda_0 I_{r_1} - A_1)$ ,  $m_{02} = \dim \mathcal{N}(\lambda_0 I_{r_2} - A_2)$ , 因此  $m_0 = m_{01} + m_{02}$ . 于是  $m_{01} + m_{02} = n_{01} + n_{02}$ , 由  $m_{01} \leq n_{01}, m_{02} \leq n_{02}$  可知  $m_{01} = n_{01}, m_{02} = n_{02}$ . 因此  $A_1$  的任意特征值作为  $A_1$  的特征值半单, 根据定理 5.3.9,  $A_1$  可对角化. 类似地,  $A_2$  也可对角化.  $\square$

最后我们来看谱分解的应用.

**例 5.3.12 (Fibonacci 数列)** 给定  $F_0 = 0, F_1 = 1$  和递推关系  $F_{n+1} = F_n + F_{n-1}$ . 试给出  $F_n$  的通项公式.

这是一个两步递推, 因此难以直接得到通项. 然而, 可以写出矩阵形式的一步递推:

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix}.$$

将其重写为  $\mathbf{u}_{n+1} = A\mathbf{u}_n$ , 那么  $\mathbf{u}_n = A^{n-1}\mathbf{u}_1$ . 而  $\mathbf{u}_1 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ . 则  $A^{n-1}$  只需要计算  $A$  的谱分解  $A = X\Lambda X^{-1}$ , 就有  $A^{n-1} = X\Lambda^{n-1}X^{-1}$ . 通过计算特征值和特征向量, 我们知道  $A$  可对角化, 它的谱分解是

$$A = \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{bmatrix},$$

其中  $\lambda_1 = \frac{1+\sqrt{5}}{2}, \lambda_2 = \frac{1-\sqrt{5}}{2}$  是两个特征值. 因此

$$F_n = \mathbf{e}_1^T \mathbf{u}_n = \mathbf{e}_1^T \frac{1}{\lambda_1 - \lambda_2} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^{n-1} & \\ & \lambda_2^{n-1} \end{bmatrix} \begin{bmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2}.$$

注意到  $\left| \frac{\lambda_2^n}{\lambda_1 - \lambda_2} \right| \approx \frac{|-0.618|^n}{2.236} \leq \frac{1}{2.236} \approx 0.4472$ , 因此只需计算  $\frac{\lambda_1^n}{\lambda_1 - \lambda_2} = \frac{1}{\sqrt{5}} \left( \frac{1+\sqrt{5}}{2} \right)^n$  最靠近的整数即可.  $\odot$

**例 5.3.13 (随机游走)** 给定一个有向无权图, 如果不断重复从一个顶点出发按某种概率选择以之为尾的弧到达下一顶点的过程, 那么生成的途径就称为**随机游走**.

如果从一个顶点出发去往相邻顶点的概率不随着时间变化, 我们就可以把相应的概率看作对应弧的权重. 矩阵  $P$  称为该图的邻接矩阵: 如果顶点  $v_i, v_j$  之间有弧, 则  $P$  的

$(i, j)$  元素是弧的权重；如果顶点  $v_i, v_j$  之间没有弧，则  $P$  的  $(i, j)$  元素是 0. 容易验证  $P$  就满足  $P\mathbf{1} = \mathbf{1}$ ，其中  $\mathbf{1}$  是所有元素都是 1 的向量. 这种每个元素都非负且任意一行元素的和都是 1 的矩阵称为**随机矩阵**. 随机矩阵有一个特征值是 1,  $\mathbf{1}$  是对应的特征向量，且随机矩阵的特征值的绝对值不大于 1 (为什么?).

任意一个元素都非负且和为 1 的向量都可以代表一个顶点集合上的概率分布，不妨称这样的向量为随机向量. 记  $\mathbf{p}$  是一个随机向量，那么  $P^T\mathbf{p}$  就是经过一步随机游走后的概率分布. 类似地， $(P^T)^k\mathbf{p}$  就是经过  $k$  步随机游走后的概率分布.

假设  $v_1, v_2, v_3, v_4$  代表四个火车站，各火车站间乘车方向概率分布见图 5.3.1. 图

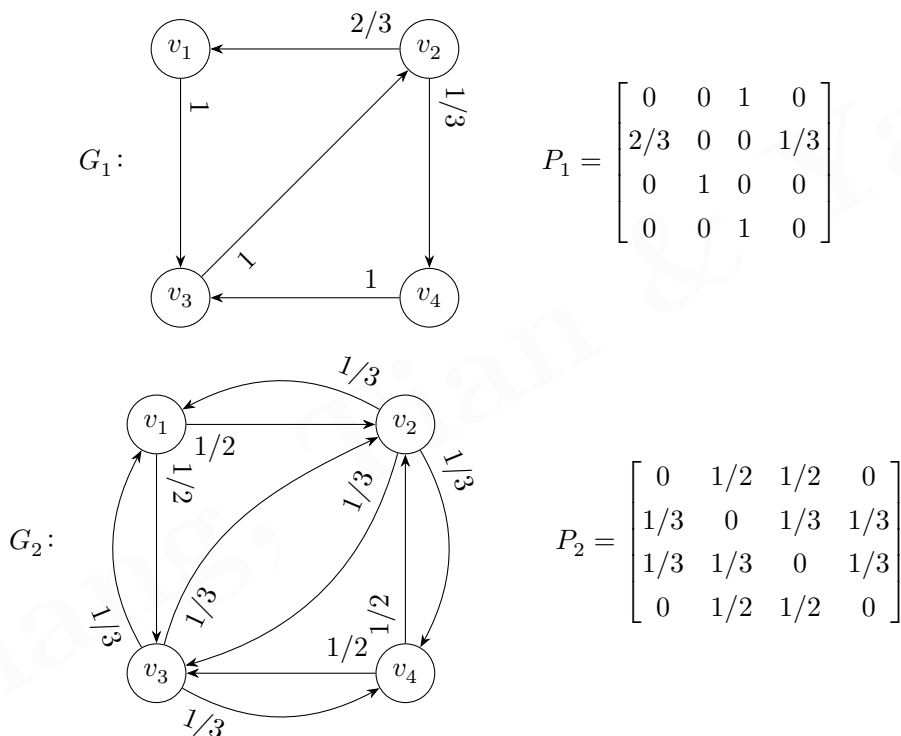


图 5.3.1: 随机游走

$G_1$  对应的随机游走，假设从  $v_1$  出发，即初始概率分布为  $\mathbf{e}_1$ ，那么序列  $(P_1^T)^k\mathbf{e}_1$  就是  $\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2, \frac{1}{3}\mathbf{e}_1 + \frac{2}{3}\mathbf{e}_4, \mathbf{e}_3, \dots$ ，序列不会收敛；假设初始概率分布为  $\mathbf{p} = \frac{1}{9} \begin{bmatrix} 1 & 3 & 3 & 2 \end{bmatrix}^T$ ，则序列是  $\mathbf{p}, \mathbf{p}, \dots$ ，序列收敛.

图  $G_2$  对应的随机游走，从任意初始概率分布出发，都会收敛到  $\frac{1}{10} \begin{bmatrix} 2 & 3 & 3 & 2 \end{bmatrix}^T$ ，该向量是  $P^T$  的属于 1 的特征向量.

两个随机游走对应的随机矩阵有何区别?  $P_1$  的绝对值为 1 的特征值有不只一个；而

$P_2$  的只有一个. 请读者思考这个区别如何导致不同初始分布产生的随机游走的收敛性的不同. ☺

## 习题

**练习 5.3.1** 设三阶方阵  $A$  的特征值及对应特征向量是  $1, 1, 3$  和  $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ , 求  $A$ .

**练习 5.3.2** 判断下列方阵是否可对角化.

$$1. \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}. \quad 2. \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

**练习 5.3.3** 计算  $A^n$ .

$$1. A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 4 & 0 & 1 \end{bmatrix}. \quad 2. A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 4 & -2 \\ -3 & -3 & 5 \end{bmatrix}.$$

**练习 5.3.4** 设  $A = \begin{bmatrix} 3 & 2 & -2 \\ -k & -1 & k \\ 4 & 2 & -3 \end{bmatrix}$ . 当  $k$  取何值时,  $A$  可对角化? 当  $A$  可对角化时, 写出其谱分解.

**练习 5.3.5** 设  $A$  是 4 阶方阵, 其对角元都是 4, 非对角元都是  $-1$ . 令  $H$  为 4 阶 Hadamard 矩阵,

即  $H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$  (见阅读 3.2.3). 计算  $AH$ , 由此得出  $A$  的谱分解, 并求  $\det(A), A^{-1}$ .

**练习 5.3.6** 利用矩阵, 求下列数列的通项公式和极限.

1. (Lucas 数)  $L_1 = 1, L_2 = 3, L_{n+2} = L_{n+1} + L_n$ .

2.  $a_0 = 0, a_1 = 1, a_{n+2} = \frac{1}{2}a_{n+1} + \frac{1}{2}a_n$ .

3.  $a_0 = a_1 = 0, a_2 = 1, a_{n+3} = \frac{1}{2}a_{n+2} + \frac{1}{4}a_{n+1} + \frac{1}{4}a_n$ .

4.  $a_0 = 1, a_{n+1} = a_n + 2 \times 3^n$ . 提示: 考虑向量  $\begin{bmatrix} a_n \\ 3^n \end{bmatrix}$ .

5.  $a_0 = 0, b_0 = 1, a_{n+1} = a_n + b_n, b_{n+1} = 3a_n + 3b_n$ . 提示: 考虑向量  $\begin{bmatrix} a_n \\ b_n \end{bmatrix}$ .

**练习 5.3.7** 利用谱分解说明如下事实.

$$1. \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}^{1024} \text{ 的每个元素都大于 } 10^{700}.$$

$$2. \begin{bmatrix} 3 & 2 \\ -5 & -3 \end{bmatrix}^{1024} = I_2.$$

$$3. \begin{bmatrix} 5 & 7 \\ -3 & -4 \end{bmatrix}^{1024} = - \begin{bmatrix} 5 & 7 \\ -3 & -4 \end{bmatrix}.$$

$$4. \begin{bmatrix} 5 & 6.9 \\ -3 & -4 \end{bmatrix}^{1024} \text{ 的每个元素都小于 } 10^{-70}.$$

**练习 5.3.8** (Pavel Grinfeld) 利用谱分解计算  $\begin{bmatrix} 110 & 55 & -164 \\ 42 & 21 & -62 \\ 88 & 44 & -131 \end{bmatrix}^{2017}$ .

**练习 5.3.9** 设有谱分解  $A = X\Lambda X^{-1}$ , 求下列矩阵的谱分解.

$$1. A^{-1}.$$

$$2. A^T.$$

$$3. \begin{bmatrix} A & O \\ O & 2A \end{bmatrix}.$$

$$4. \begin{bmatrix} O & A \\ A & O \end{bmatrix}.$$

**练习 5.3.10** 利用特征值计算下列  $n$  阶矩阵的行列式.

$$1. \begin{vmatrix} a & b & \cdots & b \\ b & a & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{vmatrix}.$$

$$2. \begin{vmatrix} 1+a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & 1+a_2b_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1}b_n \\ a_nb_1 & \cdots & a_nb_{n-1} & 1+a_nb_n \end{vmatrix}.$$

**练习 5.3.11** 证明:

$$1. n \text{ 阶方阵 } J_n(\lambda_0) = \begin{bmatrix} \lambda_0 & 1 & & \\ & \lambda_0 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_0 \end{bmatrix} \text{ 只有一个特征值 } \lambda_0, \text{ 其代数重数是 } n, \text{ 几何重数是 } 1.$$

$$2. \text{ 分块对角矩阵 } A = \begin{bmatrix} J_{n_1}(\lambda_0) & O \\ O & J_{n_2}(\lambda_0) \end{bmatrix} \text{ 只有一个特征值 } \lambda_0, \text{ 其代数重数是 } n_1 + n_2, \text{ 几何重数是 } 2.$$

**练习 5.3.12** 试分析秩为 1 的方阵何时可对角化.

**练习 5.3.13** 设  $A$  是实二阶方阵, 且  $\det(A) < 0$ , 证明  $A$  在  $\mathbb{R}$  上可对角化.

**练习 5.3.14** 证明  $A = \begin{bmatrix} I_r & O \\ B & -I_{n-r} \end{bmatrix}$  可对角化.

**练习 5.3.15** 证明,

1. 若  $A^2 = A$ , 则  $A$  可对角化.
2. 若  $A^2 = O$ , 且  $A \neq O$ , 则  $A$  不可对角化.
3. 若  $A^2 + A + I_n = O$ , 则  $A$  在  $\mathbb{R}$  上不可对角化.

**练习 5.3.16** 证明或举出反例.

1. 如果  $A$  所有的特征向量是  $k \begin{bmatrix} 1 \\ 4 \end{bmatrix}, k \neq 0$ , 则  $A$  一定有不单的特征值.

2. 如果  $A$  所有的特征向量是  $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ ,  $k \neq 0$ , 则  $A$  一定不可对角化.
3. 如果  $A$  是上三角矩阵但不是对角矩阵, 则  $A$  不可对角化.
4. 如果  $A$  是上三角矩阵但不是对角矩阵, 则存在对角矩阵  $D$ , 使得  $A - D$  不可对角化.
5. 如果  $A$  可以被对角矩阵对角化, 则  $A$  也是对角矩阵.

**练习 5.3.17** 证明定理 5.3.9 第 2 条.

**练习 5.3.18** 证明, 对反射矩阵  $H_v = I_n - 2vv^T$ , 其中  $\|v\| = 1$ , 存在正交矩阵  $Q$ , 使得  $Q^{-1}H_vQ = \text{diag}(-1, 1, \dots, 1)$ .

**练习 5.3.19 (Householder 矩阵的推广)** 设  $\mathcal{M}$  是  $\mathbb{R}^n$  的子空间, 对任意  $v \in \mathbb{R}^n$ , 存在唯一的分解  $v = v_1 + v_2$ , 其中  $v_1 \in \mathcal{M}, v_2 \in \mathcal{M}^\perp$ . 定义  $\mathbb{R}^n$  上的变换  $H$ , 使得  $H(v_1 + v_2) = v_1 - v_2$ .

1. 证明  $H$  是线性变换.
2. 设  $H$  为该线性变换的表示矩阵, 证明  $H^2 = I_n$ .
3. 求  $H$  的特征值、代数重数及特征子空间.  $H$  能否被对角化?
4. 证明, 存在矩阵  $A$ , 使得  $H = I_n - 2AA^T$ .

线性变换  $H$  保持  $\mathcal{M}$  不变, 把  $\mathcal{M}^\perp$  中的向量映射到其负向量, 因此是关于子空间  $\mathcal{M}$  的反射变换.

**练习 5.3.20** 设  $A$  是  $n$  阶实方阵, 满足  $A^2 = I_n$ , 证明  $A$  在  $\mathbb{R}$  上可对角化, 并判断  $A$  是否是关于某个子空间的反射变换.

**练习 5.3.21** 设矩阵  $A$  的特征多项式是  $p_A(x)$ .

1. 设  $A$  为对角矩阵, 证明  $p_A(A) = O$ .
2. 设  $A$  为可对角化的矩阵, 证明  $p_A(A) = O$ .
3. 设  $A = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}$ , 它是否可对角化? 是否满足  $p_A(A) = O$ ?

**练习 5.3.22** 求下列矩阵的特征值, 并判断是否可以对角化.

$$1. \begin{bmatrix} 10 & 1 & & \\ & 10 & 1 & \\ & & 10 & \end{bmatrix}, \quad \begin{bmatrix} 10 & 1 & & \\ & 10.001 & 1 & \\ & & 10.002 & \end{bmatrix}, \quad 2. \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ 0.001 & & & 0 \end{bmatrix}.$$

提示: 由此可见, 矩阵变化不大时, 特征值和可对角化性质可能变换很大.

## 5.4 相似

**定义 5.4.1 (相似)** 对方阵  $A, B$ , 如果存在可逆矩阵  $X$  使得  $X^{-1}AX = B$ , 则称  $A$  和  $B$  相似, 或  $A$  相似于  $B$ .

根据定义 5.3.1, 一个矩阵可对角化当且仅当它相似于对角矩阵.

**例 5.4.2** 数量矩阵只与自己相似:  $X^{-1}(kI_n)X = kI_n$ .

对角矩阵  $\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}$  与  $\begin{bmatrix} \lambda_2 & \\ & \lambda_1 \end{bmatrix}$  相似:  $\begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}$ .

一般地, 两个对角矩阵的对角元素如果只相差一个排列, 则二者相似, 其中的可逆矩阵是该排列对应的置换矩阵.  $\square$

**命题 5.4.3** 方阵的相似关系是等价关系.

证. 反身性:  $A = I_n^{-1}AI_n$ , 因此  $A$  和  $A$  相似.

对称性: 若  $A, B$  相似, 则存在可逆矩阵  $X$ , 使得  $X^{-1}AX = B$ ; 于是  $(X^{-1})^{-1}BX^{-1} = A$ , 而  $X^{-1}$  可逆, 因此  $B$  和  $A$  相似.

传递性: 若  $A, B$  相似, 且  $B, C$  相似, 则存在可逆矩阵  $X, Y$ , 使得  $X^{-1}AX = B$ ,  $Y^{-1}BY = C$ ; 于是  $(XY)^{-1}A(XY) = C$ , 而  $XY$  可逆, 因此  $A$  和  $C$  相似.  $\square$

下面考虑相似关系作为等价关系的从属概念.

对应于相似关系的等价变换称为**相似变换**, 即  $T: A \mapsto X^{-1}AX$ . 相似变换可以用来转化问题. 例如, 解决  $A$  对应的线性变换的原像问题  $Ax = y$ , 等价于解决相似于  $A$  的矩阵  $B = X^{-1}AX$  对应的原像问题  $Bx' = y'$ , 其中  $x' = X^{-1}x, y' = X^{-1}y$ . 特别地, 如果  $A$  可对角化, 即相似于对角矩阵  $\Lambda$  时, 原像问题  $\Lambda x' = y'$  就一目了然. 又如, 要判断矩阵  $A$  是否可对角化, 等价于判断相似于  $A$  的矩阵  $B$  是否可对角化.

下面来讨论相似关系中的不变量.

**命题 5.4.4** 方阵的相似关系有如下不变量:

1. 秩;
2. 特征多项式、特征值、特征值的代数重数、迹、行列式;
3. 特征值的几何重数.

证. 显然  $\text{rank}(X^{-1}AX) = \text{rank}(A)$ , 秩是不变量. 而  $\det(\lambda I_n - X^{-1}AX) = \det(\lambda I_n - A)$ , 特征多项式是不变量. 特征多项式决定了特征值、特征值的代数重数、迹、行列式, 因此它们都是不变量.

最后考虑几何重数. 如果  $u_1, \dots, u_r$  是  $A$  的属于  $\lambda_0$  的特征子空间的一组基, 则  $X^{-1}u_1, \dots, X^{-1}u_r$  就是  $X^{-1}AX$  的属于  $\lambda_0$  的  $r$  个线性无关的特征向量. 事实上, 容易验证  $X^{-1}AX(X^{-1}u_i) = X^{-1}(\lambda_0 u_i) = \lambda_0(X^{-1}u_i), i = 1, \dots, r$ , 即它们是属于  $\lambda_0$  的特征向量. 另一方面, 由于  $X^{-1} \begin{bmatrix} u_1 & \dots & u_r \end{bmatrix} = \begin{bmatrix} X^{-1}u_1 & \dots & X^{-1}u_r \end{bmatrix}$ , 而  $X^{-1}$  可逆, 因此  $\begin{bmatrix} u_1 & \dots & u_r \end{bmatrix}$  和  $\begin{bmatrix} X^{-1}u_1 & \dots & X^{-1}u_r \end{bmatrix}$  这两个矩阵的秩相等, 因此  $X^{-1}u_1, \dots, X^{-1}u_r$

线性无关. 于是  $\lambda_0$  作为  $A$  的特征值的几何重数就不大于  $\lambda_0$  作为  $X^{-1}AX$  的特征值的几何重数. 根据相似关系的对称性, 这两个几何重数相等.  $\square$

**推论 5.4.5** 两个对角矩阵相似当且仅当它们的对角元素除排列次序外相同.

证. 充分性: 对角元素相同的两个对角矩阵可以用置换矩阵得到相似关系.

必要性: 如果两个对角矩阵相似, 则它们具有相同的特征多项式, 而对角元素除了相差一个排列次序外, 被特征多项式唯一确定.  $\square$

如果  $A$  可对角化, 则称对角化得到的对角矩阵为  $A$  的**相似标准形**. 由推论 5.4.5 可知, 在  $A$  可对角化时, 它的相似标准形在对角元素相差一个排列次序的意义下唯一.

一般的方阵, 只要有亏损特征值, 就不可对角化. 那么, 我们能找到哪种具有简单形式的矩阵与之相似? 换言之, 利用相似变换, 我们能把方阵化简成什么形式?

**命题 5.4.6** 对  $n$  阶方阵  $A$ , 存在可逆矩阵  $X$ , 使得  $X^{-1}AX = T$  是上三角矩阵, 且  $T$  的对角元素是  $A$  的  $n$  个特征值 (计重数). 进一步地, 通过选择特定的  $X$ , 能够令  $T$  的对角元素是  $A$  的特征值的任意排列.

证. 采用数学归纳法.  $n = 1$  时, 显然. 假设任意  $n - 1$  阶方阵都有如上分解, 观察  $n$  阶方阵.

任取  $A$  的一个特征对  $(\lambda_1, \mathbf{x}_1)$ , 把  $\mathbf{x}_1$  扩充成  $\mathbb{C}^n$  的一组基  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . 记  $X_1 = [\mathbf{x}_1 \ X_{12}] = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]$ . 则  $AX_1$  的列能被  $X$  的列线性表示, 故存在矩阵  $M_1$  使得  $AX_1 = X_1M_1$ . 由于  $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ , 因此

$$AX_1 = [A\mathbf{x}_1 \ AX_{12}] = [\lambda_1\mathbf{x}_1 \ AX_{12}] = [\mathbf{x}_1 \ X_{12}] \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & A_1 \end{bmatrix} = X_1M_1,$$

即  $M_1$  是分块上三角矩阵. 而  $A_1$  是  $n - 1$  阶方阵, 根据归纳假设, 存在可逆矩阵  $X_2$ , 使得  $A_1 = X_2T_2X_2^{-1}$ , 其中  $T_2$  上三角. 于是

$$A = X_1 \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & X_2T_2X_2^{-1} \end{bmatrix} X_1^{-1} = X_1 \begin{bmatrix} 1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & X_2 \end{bmatrix}^{-1} X_1^{-1}.$$

记  $X = X_1 \begin{bmatrix} 1 & 0 \\ 0 & X_2 \end{bmatrix}$ ,  $T = \begin{bmatrix} \lambda_1 & * \\ 0 & T_2 \end{bmatrix}$ , 则  $X$  可逆,  $T$  上三角, 且  $A = XTX^{-1}$ .

综上, 命题对任意  $n$  成立.

考察特征多项式, 立得  $T$  的对角元素就是  $A$  的  $n$  个特征值. 通过上述证明能看到, 按照特征值的排列选择特征对, 就能使  $T$  的对角元素满足条件.  $\square$

利用上述分解, 能得到 Hamilton-Cayley 定理.



**定理 5.4.7 (Hamilton-Cayley 定理)** 设方阵  $A$  的特征多项式为  $p_A(\lambda)$ , 则  $p_A(A) = O$ .

证. 设  $A$  有分解  $A = XTX^{-1}$ , 其中  $T$  是上三角矩阵. 而  $p_A(A) = Xp_A(T)X^{-1}$ , 只需证明  $p_A(T) = O$ . 若  $\lambda_1, \dots, \lambda_n$  是  $A$  的  $n$  个特征值, 则  $p_A(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$ , 于是  $p_A(T) = (T - \lambda_1 I_n) \cdots (T - \lambda_n I_n)$ . 注意  $T - \lambda_i I_n$  的第  $i$  个对角元素是零, 逐步计算矩阵乘法就得到  $p_A(T) = O$ .  $\square$

上三角矩阵仍然不够简单. 那么对给定矩阵, 与之相似的矩阵中, 最简单的矩阵是什么?

**定理 5.4.8 (Jordan 分解)** 对  $n$  阶方阵  $A$ , 存在可逆矩阵  $X$ , 使得

$$X^{-1}AX = J = \begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_r}(\lambda_r) \end{bmatrix},$$

其中  $J_{n_i}(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{n_i \times n_i}$  是  $n_i$  阶 Jordan 块, 而  $n_1 + \cdots + n_r = n$ , 且除了这些 Jordan 块的排列次序外,  $J$  被  $A$  唯一确定.

定理 5.4.8 的证明超出了本书的要求, 感兴趣的读者可以参考阅读 5.4.7.

分解  $A = XJX^{-1}$  称为  $A$  的 **Jordan 分解**. 定理 5.4.8 中的  $J$  称为  $A$  的 **Jordan 标准形**, 这是  $A$  的一种相似标准形. 另外, 根据问题的需要, 相似标准形不止 Jordan 标准形一种, 还存在其他形式的标准形.

下面通过一个例子来体会 Jordan 分解的用处.

**例 5.4.9 (弹簧振子)** 假设弹簧和振子如图 5.4.1 所示系于左壁, 弹簧的原始长度为  $l$ , 劲度系数为  $k$ , 振子的质量为  $m$ . 系统具有阻尼, 即, 随着振子的运动, 系统将对振子产生与运动方向相反的作用力, 该力大小与速度成正比, 比例系数称为阻尼系数, 记为  $d$ . 通过外力将振子牵引至距左壁  $l + x_0$  处, 待系统静止后, 突然放开振子, 不考虑振子的竖直运动, 振子将如何运动?

所谓如何运动, 就是要描述振子在任意时刻的位置、速度和加速度, 因此需要得到振子的位置、速度和加速度关于时间的函数. 假设振子到侧壁的距离是  $l_1 + x(t)$ , 其中  $x(0) = x_0$ .

记弹簧原长处为原点, 原点右侧为正方向, 振子相对原点的位置为  $x(t)$ . 振子的速度  $v(t)$  是位移 (位置变化量) 的导数, 即  $v(t) = x'(t)$ ,  $v(0) = 0$ . 振子的加速度是二阶

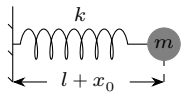


图 5.4.1: 弹簧振子

导数  $a(t) = v'(t) = x''(t)$ . 因此, 只要能求出  $x(t)$ , 就能通过导数得到  $v(t), a(t)$ . 振子所受的拉力  $f(t) = -kx(t)$ , 系统对振子产生的阻尼是  $g(t) = -dv(t)$ . 利用 Newton 第二定律, 我们有  $ma(t) = f(t) + g(t) = -kx(t) - dv(t)$ . 因此得到方程组

$$\begin{bmatrix} x'(t) \\ v'(t) \end{bmatrix} = \begin{bmatrix} v(t) \\ a(t) \end{bmatrix} = \begin{bmatrix} v(t) \\ -\frac{d}{m}v(t) - \frac{k}{m}x(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{d}{m} \end{bmatrix} \begin{bmatrix} x(t) \\ v(t) \end{bmatrix}.$$

直观上, 向量或矩阵的导数, 就是其分量或元素的导数组成的向量或矩阵, 这一点这里不详细说明. 令  $\mathbf{p}(t) = \begin{bmatrix} x(t) \\ v(t) \end{bmatrix}$ ,  $A = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{d}{m} \end{bmatrix}$ , 方程就变为  $\mathbf{p}'(t) = T\mathbf{p}(t)$ .

如果  $T$  可对角化, 则  $T$  有谱分解  $T = Q\Lambda Q^{-1}$ , 其中  $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ . 于是  $\mathbf{p}'(t) = Q\Lambda Q^{-1}\mathbf{p}(t)$ . 令  $\mathbf{q}(t) = Q^{-1}\mathbf{p}(t)$ , 则  $\mathbf{q}'(t) = \Lambda\mathbf{q}(t)$ . 问题通过相似变换化成了两个互不相关的一阶常系数常微分方程  $q_i'(t) = \lambda_i q_i(t)$ . 根据常微分方程知识可知, 其解为  $q_i(t) = C_i e^{\lambda_i t}$ , 其中  $C_i$  由初始值确定. 记  $e^{At} = \text{diag}(e^{\lambda_1 t}, e^{\lambda_2 t})$ ,  $\mathbf{c} = [C_i]$ , 就有  $\mathbf{q}(t) = e^{At}\mathbf{c}$ .

于是  $\mathbf{p}(t) = Qe^{At}\mathbf{c}$ . 注意  $\mathbf{p}(0) = \begin{bmatrix} x(0) \\ v(0) \end{bmatrix} = \begin{bmatrix} x_0 \\ 0 \end{bmatrix} = x_0 \mathbf{e}_1$ , 由此即可解出  $\mathbf{c} = x_0 Q^{-1}\mathbf{e}_1$ . 因此  $x(t) = x_0 \mathbf{e}_1^T Q e^{At} Q^{-1} \mathbf{e}_1$ , 而  $v(t) = x_0 \mathbf{e}_2^T Q e^{At} Q^{-1} \mathbf{e}_1$  或用  $x(t)$  的导数得到.

那么  $T$  何时可对角化? 先来计算  $T$  的特征值,  $T$  的特征多项式为  $\det(\lambda I - T) = \lambda^2 + \frac{d}{m}\lambda + \frac{k}{m} = 0$ , 其判别式为  $\frac{d^2 - 4mk}{m^2}$ .

1. 当  $d > 2\sqrt{mk}$  时,  $T$  有两个不同实特征值  $\lambda_1, \lambda_2$ ,  $T$  在  $\mathbb{R}$  上可对角化. 计算可得

$$Q = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}, \text{ 因此}$$

$$\begin{aligned} x(t) &= x_0 \mathbf{e}_1^T Q e^{At} Q^{-1} \mathbf{e}_1 \\ &= \frac{x_0}{\lambda_2 - \lambda_1} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= x_0 \frac{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}{\lambda_2 - \lambda_1}, \end{aligned}$$

利用  $\lambda_1 \lambda_2 = \frac{k}{m}$ , 有  $v(t) = -x_0 \frac{k}{m} \frac{e^{\lambda_2 t} - e^{\lambda_1 t}}{\lambda_2 - \lambda_1}$ ,  $a(t) = -x_0 \frac{k}{m} \frac{\lambda_2 e^{\lambda_2 t} - \lambda_1 e^{\lambda_1 t}}{\lambda_2 - \lambda_1}$ .

2. 当  $d < 2\sqrt{mk}$  时,  $T$  有一对共轭复特征值  $\lambda_1 = \alpha + \beta i, \lambda_2 = \alpha - \beta i$ ,  $J$  在  $\mathbb{C}$  上可对角化. 类似地,

$$x(t) = x_0 \frac{\lambda_2 e^{\lambda_1 t} - \lambda_1 e^{\lambda_2 t}}{\lambda_2 - \lambda_1} = x_0 \frac{(\alpha - \beta i)e^{(\alpha + \beta i)t} - (\alpha + \beta i)e^{(\alpha - \beta i)t}}{-2\beta i}.$$

利用  $e^{i\beta} = \cos \beta + i \sin \beta$ , 可得  $x(t) = x_0 e^{\alpha t} \left( \cos(\beta t) - \alpha \frac{\sin(\beta t)}{\beta} \right)$ , 而  $\alpha^2 + \beta^2 = \lambda_1 \lambda_2 = \frac{k}{m}$ , 于是  $v(t) = -x_0 \frac{k}{m} e^{\alpha t} \frac{\sin(\beta t)}{\beta}$ ,  $a(t) = -x_0 \frac{k}{m} e^{\alpha t} \left( \alpha \frac{\sin(\beta t)}{\beta} + \cos(\beta t) \right)$ .

3. 当  $d = 2\sqrt{mk}$  时,  $T$  只有一个特征值  $\lambda_1$ , 代数重数为 2, 几何重数为 1, 不可对角化. 根据定理 5.4.8,  $T$  有 Jordan 分解  $T = QJQ^{-1}$ , 其中  $J$  是二阶 Jordan 块.

考虑  $d = d_0 = 2\sqrt{mk}$  的特殊情形, 此时  $\lambda_1 = \sqrt{\frac{k}{m}}$ . 利用  $TQ = QJ = Q \begin{bmatrix} \lambda_1 & 1 \\ & \lambda_1 \end{bmatrix}$ , 可知

$Qe_1$  是  $T$  的特征向量, 而  $Qe_2$  满足  $(T - \lambda_1 I)Qe_2 = Qe_1$ . 计算可得  $Q = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_1 + 1 \end{bmatrix}$ .

另一方面,  $p'(t) = QJQ^{-1}p(t)$ . 令  $q(t) = Q^{-1}p(t)$ , 则  $q'(t) = Jq(t)$ . 问题就通过相似变换转化为了两个一阶常系数常微分方程  $q_1'(t) = \lambda_1 q_1(t) + q_2(t), q_2'(t) = \lambda_1 q_2(t)$ . 后者的解为  $q_2(t) = C_2 e^{\lambda_1 t}$ . 代入前者, 利用常微分方程知识可以解得  $q_1(t) = e^{\lambda_1 t}(C_1 + C_2 t)$ .

因此, 记  $e^{Jt} = \begin{bmatrix} e^{\lambda_1 t} & te^{\lambda_1 t} \\ & e^{\lambda_1 t} \end{bmatrix}$ , 就有  $q(t) = e^{Jt}c$ . 类似地,

$$\begin{aligned} x(t) &= x_0 e_1^T Q e^{Jt} Q^{-1} e_1 \\ &= x_0 \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_1 + 1 \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} & te^{\lambda_1 t} \\ & e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} \lambda_1 + 1 & -1 \\ -\lambda_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = x_0 e^{\lambda_1 t} (1 - \lambda_1 t), \end{aligned}$$

$$v(t) = -x_0 \frac{k}{m} e^{\lambda_1 t} t, \quad a(t) = -x_0 \frac{k}{m} e^{\lambda_1 t} (1 + \lambda_1 t).$$

特殊情形  $d = d_0$  的矩阵  $T$  不可对角化, 从而在计算上引起了很大不便. 然而, 该特殊情形在物理上有明显优点. 考虑系统的机械能  $e(t) = \frac{1}{2} kx(t)^2 + \frac{1}{2} mv(t)^2 = \frac{1}{2} p(t)^T K p(t)$ , 其中  $K = \text{diag}(k, m)$ . 可以验证, 该特殊情形的能量耗散最快. 取不同的阻尼系数  $d = 0.5d_0, d_0, 2d_0$ , 系统的运动如图 5.4.2 所示.  $\odot$

最后, 我们来考虑能否用同一个相似变换将两个矩阵对角化的问题.

**定义 5.4.10 (同时对角化)** 设  $A, B$  是  $n$  阶方阵, 如果存在可逆矩阵  $X$  使得  $X^{-1}AX = A_1$  和  $X^{-1}BX = A_2$  都是对角矩阵, 则称  $A, B$  可以**同时对角化**.

**命题 5.4.11** 对可对角化的  $n$  阶方阵  $A, B$ , 以下叙述等价:

1.  $A, B$  可以同时对角化;

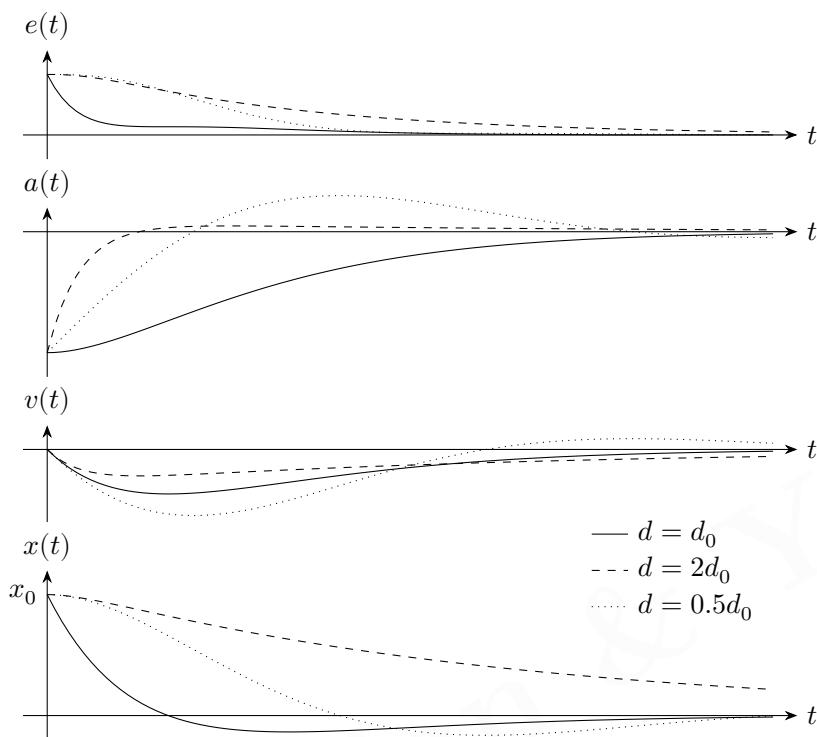


图 5.4.2: 阻尼振动

2. 存在  $n$  个线性无关的向量, 同时是  $A, B$  的特征向量;

3.  $A, B$  可交换, 即  $AB = BA$ .

证. 显然, 1 和 2 等价.

“1  $\Rightarrow$  3”: 设  $X^{-1}AX = \Lambda_1, X^{-1}BX = \Lambda_2$ , 因为对角矩阵可交换, 于是  $AB = XX^{-1}AXX^{-1}BXX^{-1} = X\Lambda_1\Lambda_2X^{-1} = X\Lambda_2\Lambda_1X^{-1} = XX^{-1}BXX^{-1}AXX^{-1} = BA$ .

“3  $\Rightarrow$  1”: 由于  $A$  可对角化, 存在可逆矩阵  $X_1$ , 使得  $X_1^{-1}AX_1 = \tilde{\Lambda}_1$  是对角矩阵.

再利用置换矩阵  $P$ , 可将  $\tilde{\Lambda}_1$  变为如下形式:  $P^{-1}\tilde{\Lambda}_1P = \Lambda_1 = \begin{bmatrix} \lambda_1 I_{n_1} & & \\ & \ddots & \\ & & \lambda_r I_{n_r} \end{bmatrix}$ , 其

中  $\lambda_1, \dots, \lambda_r$  两两不同.

此时  $P^{-1}X_1^{-1}AX_1P = \Lambda_1$ . 令  $\tilde{B} = P^{-1}X_1^{-1}BX_1P$ , 则由  $AB = BA$  可知  $\Lambda_1\tilde{B} = \tilde{B}\Lambda_1$ . 把  $\tilde{B}$  写成与  $\Lambda_1$  对应的分块矩阵  $\tilde{B} = \begin{bmatrix} \tilde{B}_{11} & \cdots & \tilde{B}_{1r} \\ \vdots & & \vdots \\ \tilde{B}_{r1} & \cdots & \tilde{B}_{rr} \end{bmatrix}$ , 由  $\Lambda_1\tilde{B} = \tilde{B}\Lambda_1$  可知

$$(\lambda_i I_{n_i}) \widetilde{B}_{ij} = \widetilde{B}_{ij} (\lambda_j I_{n_j}), (\lambda_i - \lambda_j) \widetilde{B}_{ij} = 0. \text{ 对 } i \neq j, \text{ 有 } \widetilde{B}_{ij} = 0, \text{ 即 } \widetilde{B} = \begin{bmatrix} \widetilde{B}_{11} & & \\ & \ddots & \\ & & \widetilde{B}_{rr} \end{bmatrix}$$

也是分块对角矩阵.

由  $B$  可对角化可得  $\widetilde{B}$  可对角化, 根据命题 5.3.11,  $\widetilde{B}_{ii}$  都可对角化. 设  $X_{ii}^{-1} \widetilde{B}_{ii} X_{ii} = \Lambda_{ii}$ , 其中  $X_{ii}$  可逆,  $\Lambda_{ii}$  对角. 令  $X_2 = \begin{bmatrix} X_{11} & & \\ & \ddots & \\ & & X_{rr} \end{bmatrix}$ , 那么  $X_2$  可逆, 且  $X_2^{-1} \widetilde{B} X_2 =$

$$\begin{bmatrix} \Lambda_{11} & & \\ & \ddots & \\ & & \Lambda_{rr} \end{bmatrix} =: \Lambda_2. \text{ 又有 } X_2^{-1} \Lambda_1 X_2 = \begin{bmatrix} X_{11}^{-1} \lambda_1 I_{n_1} X_{11} & & \\ & \ddots & \\ & & X_{rr}^{-1} \lambda_r I_{n_r} X_{rr} \end{bmatrix} = \Lambda_1,$$

因此

$$X_2^{-1} P^{-1} X_1^{-1} A X_1 P X_2 = \Lambda_1, \quad X_2^{-1} P^{-1} X_1^{-1} B X_1 P X_2 = \Lambda_2,$$

即  $X_1 P X_2$  将  $A, B$  同时对角化. □

**例 5.4.12** 数学物理中的 **Pauli 矩阵** 是三个二阶复矩阵

$$\sigma_1 = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, \sigma_2 = \begin{bmatrix} & -i \\ i & \end{bmatrix}, \sigma_3 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}.$$

容易验证, 三个矩阵的特征多项式都是  $\lambda^2 - 1$ , 都具有两个单特征值  $\pm 1$ , 因此都可对角化. 然而, 直接计算可得

$$\sigma_1 \sigma_2 - \sigma_2 \sigma_1 = 2i\sigma_3, \quad \sigma_2 \sigma_3 - \sigma_3 \sigma_2 = 2i\sigma_1, \quad \sigma_3 \sigma_1 - \sigma_1 \sigma_3 = 2i\sigma_2.$$

即三个矩阵两两不可交换, 因此任意两个都不可同时对角化. ☺

## 习题

**练习 5.4.1** 设  $A, B$  是  $n$  阶方阵, 且满足  $AB = BA$ .

1. 证明, 若  $A$  有  $n$  个不同的特征值, 则  $B$  可对角化.
2. 若  $A$  有代数重数大于 1 的特征值,  $B$  是否一定可对角化?

**练习 5.4.2** 对下列矩阵  $A, B$ , 求  $X$  使得  $A = XBX^{-1}$ .

1.  $A = MN, B = NM$ , 其中  $M, N$  是方阵, 且  $M$  可逆.
2.  $A = \begin{bmatrix} MN & O \\ N & O \end{bmatrix}, B = \begin{bmatrix} O & O \\ N & NM \end{bmatrix}$ , 其中  $M, N$  不必是方阵.
3.  $A = \begin{bmatrix} M & -N \\ N & M \end{bmatrix}, B = \begin{bmatrix} M + iN & O \\ O & M - iN \end{bmatrix}$ , 其中  $M, N$  是方阵.

**练习 5.4.3** 1. 设  $A, B$  是  $n$  阶方阵满足  $AB = BA$ . 若  $A$  有  $n$  个不同的特征值, 证明, 存在次数不超过  $n-1$  的多项式  $f(x)$ , 使得  $B = f(A)$ .

2. 设  $A, B$  是  $n$  阶方阵满足  $AB = BA$ . 若  $A = J_n(\lambda)$ , 证明, 存在次数不超过  $n-1$  的多项式  $f(x)$ , 使得  $B = f(A)$ .

3. 举例说明, 存在  $A, B$  满足  $AB = BA$ , 但不存在多项式  $f(x)$ , 使得  $B = f(A)$ .

**练习 5.4.4** 证明, 任意迹为 0 的方阵相似于一个对角元素全为 0 的方阵.

**练习 5.4.5 (Jordan 链)** 对任意  $n$  阶方阵  $A$ , 一组向量  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , 如果存在  $\lambda \in \mathbb{C}$ , 使得

$$(A - \lambda I)\mathbf{x}_1 = \mathbf{0}, \quad (A - \lambda I)\mathbf{x}_2 = \mathbf{x}_1, \quad \dots, \quad (A - \lambda I)\mathbf{x}_s = \mathbf{x}_{s-1}, \quad \text{而 } (A - \lambda I)\mathbf{y} = \mathbf{x}_s \text{ 无解,}$$

就称其为  $A$  的一个关于  $\lambda$  长度为  $s$  的 **Jordan 链**.

显然,  $\mathbf{x}_1 \in \mathcal{N}(A - \lambda I)$  是特征向量,  $\mathbf{x}_2 \in \mathcal{N}((A - \lambda I)^2), \dots, \mathbf{x}_s \in \mathcal{N}((A - \lambda I)^s)$ , 但  $\mathbf{x}_s \notin \mathcal{R}(A - \lambda I)$ .

求证:

1.  $\mathbf{x}_1, \dots, \mathbf{x}_s$  线性无关.

2. 令  $X_1 = [\mathbf{x}_1 \ \dots \ \mathbf{x}_s]$ , 则  $AX_1 = X_1 J_s(\lambda)$ .

3. 若  $A$  只有一个特征值  $\lambda$ , 且其几何重数为 1, 则  $A$  有一个关于  $\lambda$  的长度为  $n$  的 Jordan 链  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , 且  $A$  相似于  $J_n(\lambda)$ .

**练习 5.4.6** 给定  $m$  阶方阵  $A_1$ ,  $n$  阶方阵  $A_2$  和  $m \times n$  矩阵  $B$ . 证明

1. 如果  $A_1$  和  $A_2$  没有相同的特征值, 关于  $m \times n$  矩阵  $X$  的 Sylvester 方程  $A_1 X - X A_2 = B$  有唯一解.

2. 如果  $A_1$  和  $A_2$  没有相同的特征值, 则存在唯一的矩阵  $X$  满足

$$\begin{bmatrix} I_r & X \\ 0 & I_t \end{bmatrix} \begin{bmatrix} A_1 & B \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} I_r & X \\ 0 & I_t \end{bmatrix}^{-1} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}.$$

3. 对  $n$  阶方阵  $A$ , 存在可逆矩阵  $X$ , 使得

$$X^{-1}AX = \begin{bmatrix} \lambda_1 I + N_1 & & \\ & \ddots & \\ & & \lambda_s I + N_s \end{bmatrix},$$

其中  $N_1, \dots, N_s$  是严格上三角矩阵.

**阅读 5.4.7 (Jordan 分解的证明)** 下面证明定理 5.4.8.

首先考虑严格上三角矩阵的 Jordan 分解, 即对  $n$  阶严格上三角矩阵  $A$ , 存在可逆矩阵  $X$  使得

$$X^{-1}AX = J = \begin{bmatrix} J_{n_1}(0) & & \\ & \ddots & \\ & & J_{n_r}(0) \end{bmatrix},$$

其中  $n_1 + \cdots + n_t = n, n_1 \geq \cdots \geq n_t \geq 1$ .

对  $n$  用数学归纳法.  $n=1$  显然成立, 假设命题对任意不超过  $n-1$  阶的矩阵成立, 下证命题对  $n$  阶矩阵也成立. 对  $A$  分块  $A = \begin{bmatrix} 0 & \mathbf{a}^T \\ & A_1 \end{bmatrix}$ . 根据归纳假设, 存在  $n-1$  阶可逆矩阵  $\tilde{X}_1$ , 使得  $\tilde{X}_1^{-1} A_1 \tilde{X}_1 =$

$\begin{bmatrix} J_{k_1}(0) \\ & J \end{bmatrix}$ , 其中  $J$  是分块对角矩阵, 对角块是  $J_{k_i}(0), i=2, \cdots, s$ , 且  $k_1 \geq \cdots \geq k_s, k_1 + \cdots + k_s =$

$n-1$ . 取  $X_1 = \begin{bmatrix} 1 & \\ & \tilde{X}_1 \end{bmatrix}$ , 于是  $X_1^{-1} A X_1 =: \begin{bmatrix} 0 & \mathbf{a}_1^T & \mathbf{a}_2^T \\ & J_{k_1}(0) & \\ & & J \end{bmatrix}$ . 取  $X_2 = \begin{bmatrix} 1 & \mathbf{a}_1^T J_{k_1}(0)^T & \\ & I_{k-1} & \\ & & I \end{bmatrix}$ , 则

$$X_2^{-1} X_1^{-1} A X_1 X_2 = \begin{bmatrix} 0 & (\mathbf{a}_1^T \mathbf{e}_1) \mathbf{e}_1^T & \mathbf{a}_2^T \\ & J_{k_1}(0) & \\ & & J \end{bmatrix}.$$

若  $\mathbf{a}_1^T \mathbf{e}_1 = 0$ , 则存在置换矩阵  $P$ , 使得  $P^T X_2^{-1} X_1^{-1} A X_1 X_2 P = \begin{bmatrix} J_{k_1}(0) & & \\ & 0 & \mathbf{a}_2^T \\ & & J \end{bmatrix}$ , 对  $\begin{bmatrix} 0 & \mathbf{a}_2^T \\ & J \end{bmatrix}$

使用归纳假设即得结论.

若不然, 先取  $X_3 = \begin{bmatrix} \mathbf{a}_1^T \mathbf{e}_1 & & \\ & I_{k_1} & \\ & & \mathbf{a}_1^T \mathbf{e}_1 I \end{bmatrix}$ , 就有  $X_3^{-1} X_2^{-1} X_1^{-1} A X_1 X_2 X_3 = \begin{bmatrix} J_{k_1+1}(0) & \mathbf{a}_2^T \\ & J \end{bmatrix}$ . 再

取  $X_4 = \prod_{i=1}^{k_1} \begin{bmatrix} I_{k_1+1} & -\mathbf{e}_{i+1} \mathbf{a}_2^T J^{i-1} \\ & I \end{bmatrix}$ , 计算可得  $X_4 X_3^{-1} X_2^{-1} X_1^{-1} A X_1 X_2 X_3 X_4 = \begin{bmatrix} J_{k_1+1}(0) & \\ & J \end{bmatrix}$ ,

也得结论. 综上即有, 命题成立, 即严格上三角矩阵有 Jordan 分解.

次之, 考虑任意矩阵的 Jordan 分解. 根据练习 5.4.6, 存在可逆矩阵  $X_1$ , 使得  $X_1^{-1} A X_1 =$

$$\begin{bmatrix} \lambda_1 I + N_1 & & \\ & \ddots & \\ & & \lambda_s I + N_s \end{bmatrix}. \text{ 其中 } N_1, \cdots, N_s \text{ 是严格上三角矩阵. 再利用严格上三角矩阵的 Jordan}$$

分解, 存在可逆矩阵  $X_2 = \begin{bmatrix} X_{11} & & \\ & \ddots & \\ & & X_{ss} \end{bmatrix}$ , 使得  $X_2^{-1} X_1^{-1} A X_1 X_2 = J$  是分块对角矩阵, 而且每个

对角块都是 Jordan 块. 令  $X = X_1 X_2$ , 就证明了 Jordan 分解存在性.

最后考虑 Jordan 分解的唯一性.  $J$  中相同特征值的 Jordan 块的个数就是该特征值的几何重数, 而

这些 Jordan 块的阶数的和就是该特征值的代数重数. 注意矩阵的特征值、代数重数、几何重数都是相似不变量. 因此只需证明若  $J(\lambda) = \begin{bmatrix} J_{n_1}(\lambda) & & \\ & \ddots & \\ & & J_{n_k}(\lambda) \end{bmatrix}$  和  $\tilde{J}(\lambda) = \begin{bmatrix} J_{m_1}(\lambda) & & \\ & \ddots & \\ & & J_{m_k}(\lambda) \end{bmatrix}$  相似,

其中  $n_1 \geq \cdots \geq n_k \geq 1, m_1 \geq \cdots \geq m_k \geq 1, n_1 + \cdots + n_k = m_1 + \cdots + m_k$ , 则必有  $m_i = n_i, i=1, \cdots, k$ .

若不然, 则存在  $j$ , 使得  $m_i = n_i, i=1, \cdots, j-1$ , 但  $m_j \neq n_j$ . 不妨设  $m_j > n_j$ , 令  $J_1 =$

$$\begin{bmatrix} J_{n_1}(\lambda) & & \\ & \ddots & \\ & & J_{n_{j-1}}(\lambda) \end{bmatrix}, J_2 = \begin{bmatrix} J_{n_j}(\lambda) & & \\ & \ddots & \\ & & J_{n_k}(\lambda) \end{bmatrix}, \tilde{J}_2 = \begin{bmatrix} J_{m_j}(\lambda) & & \\ & \ddots & \\ & & J_{m_k}(\lambda) \end{bmatrix}, \text{ 则 } J(\lambda)^{n_j} =$$

$$\begin{bmatrix} (J_1 - \lambda I)^{n_j} \\ 0 \end{bmatrix}, \tilde{J}(\lambda)^{n_j} = \begin{bmatrix} (J_1 - \lambda I)^{n_j} & \\ & (\tilde{J}_2 - \lambda I)^{n_j} \end{bmatrix},$$
 其中  $(\tilde{J}_2 - \lambda I)^{n_j} \neq 0$ . 但二者相似, 故秩相同, 矛盾.



## 第六章 实对称矩阵

对称矩阵经常出现在实际生产生活中, 前面很多例子也体现了这一点. 那么对称矩阵的特征值和特征向量有何特点? 这是本章的出发点.

### 6.1 实对称矩阵的谱分解

在第 5.3 节中我们得到, 如果复方阵的特征值都半单, 则该矩阵有谱分解. 现在假设实对称矩阵  $A$  有谱分解  $A = X\Lambda X^{-1}$ , 其中  $X$  可逆,  $\Lambda$  对角. 那么  $A^T = (X^{-1})^T \Lambda^T X^T = (X^T)^{-1} \Lambda X^T$ . 而  $A = A^T$ , 因此  $(X^{-1})^T \Lambda X^T = X\Lambda X^{-1}$ . 我们能找到矩阵  $X$ , 使得  $X^T = X^{-1}$  吗? 如果  $X^T = X^{-1}$  且  $X$  是实矩阵, 则  $X$  是正交矩阵. 于是  $A = X\Lambda X^T$ , 该写法本身就暗示了  $A$  对称.

上面的猜想加了不少假设, 我们需要一一验证, 其重点在于分析实对称矩阵的特征值和特征向量.

首先考虑实对称矩阵的特征值.

**命题 6.1.1** 实对称矩阵的特征多项式的根都是实根, 即实对称矩阵在  $\mathbb{C}$  上的特征值都是实数.

证. 对实对称矩阵  $A$ ,  $\lambda$  是其复特征值, 则存在复特征向量  $\mathbf{x} \in \mathbb{C}^n$ , 满足  $A\mathbf{x} = \lambda\mathbf{x}$ . 两端取共轭, 得  $A\bar{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}}$ . 考虑  $\bar{\mathbf{x}}^T A\mathbf{x}$ , 一方面  $\bar{\mathbf{x}}^T A\mathbf{x} = \lambda\bar{\mathbf{x}}^T \mathbf{x}$ ; 另一方面, 由  $A\bar{\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}}$  取转置得  $\bar{\mathbf{x}}^T A = \bar{\lambda}\bar{\mathbf{x}}^T$ , 于是  $\bar{\mathbf{x}}^T A\mathbf{x} = \bar{\lambda}\bar{\mathbf{x}}^T \mathbf{x}$ . 因此  $\lambda\bar{\mathbf{x}}^T \mathbf{x} = \bar{\lambda}\bar{\mathbf{x}}^T \mathbf{x}$ . 若令  $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T$ , 注意  $\mathbf{x} \neq \mathbf{0}$ , 则  $\bar{\mathbf{x}}^T \mathbf{x} = |x_1|^2 + \cdots + |x_n|^2 > 0$ . 因此  $\lambda = \bar{\lambda}$ , 即  $\lambda$  是实数.  $\square$

实对称矩阵的实特征值, 对应的特征向量也是实向量. 由此即可说明谱分解的存在性.

**定理 6.1.2 (实对称矩阵的谱分解)** 对  $n$  阶实对称矩阵  $A$ , 存在  $n$  阶正交矩阵  $Q$  和实对角矩阵  $\Lambda$ , 使得  $A = Q\Lambda Q^T$ .

证. 采用数学归纳法.  $n=1$  时, 显然. 假设任意  $n-1$  阶实对称矩阵都有如上分解.

对  $n$  阶实对称矩阵  $A$ , 根据命题 6.1.1, 总存在实特征对  $(\lambda_1, \mathbf{q}_1)$ , 不妨设  $\|\mathbf{q}_1\|=1$ . 把  $\mathbf{q}_1$  扩充成  $\mathbb{R}^n$  的一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_n$ , 令  $Q_1 = [\mathbf{q}_1 \ Q_{12}] = [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_n]$ . 由于  $\mathbf{q}_1$  与  $Q_{12}$  的列向量都正交,

$$Q_1^T A Q_1 = \begin{bmatrix} \mathbf{q}_1^T A \mathbf{q}_1 & \mathbf{q}_1^T A Q_{12} \\ Q_{12}^T A \mathbf{q}_1 & Q_{12}^T A Q_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 \mathbf{q}_1^T \mathbf{q}_1 & \lambda_1 \mathbf{q}_1^T Q_{12} \\ \lambda_1 Q_{12}^T \mathbf{q}_1 & Q_{12}^T A Q_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & Q_{12}^T A Q_{12} \end{bmatrix}.$$

注意  $Q_{12}^T A Q_{12}$  是  $n-1$  阶实对称矩阵, 根据归纳假设, 存在正交矩阵  $Q_2$  和实对角矩阵  $\Lambda_2$ , 使得  $Q_{12}^T A Q_{12} = Q_2 \Lambda_2 Q_2^T$ . 因此

$$A = Q_1 \begin{bmatrix} \lambda_1 & 0 \\ 0 & Q_2 \Lambda_2 Q_2^T \end{bmatrix} Q_1^T = Q_1 \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}^T Q_1^T.$$

记  $Q = Q_1 \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}$ ,  $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$ , 则  $Q$  正交,  $\Lambda$  实对角, 而  $A = Q \Lambda Q^T$ . 综上, 命题对任意  $n$  成立.  $\square$

分解  $A = Q \Lambda Q^T$  称为对称矩阵  $A$  的谱分解.

如果令  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $Q = [\mathbf{q}_1 \ \dots \ \mathbf{q}_n]$ , 那么

$$A = Q \Lambda Q^T = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \dots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T.$$

这种把矩阵写成  $n$  个秩为一的矩阵的分解往往也称为谱分解.

注意,  $A = Q \Lambda Q^T$ , 当且仅当  $AQ = Q\Lambda$ .  $\Lambda$  的对角元素是  $A$  的特征值,  $Q$  的第  $i$  列就是  $A$  的属于该特征值的特征向量. 属于相同特征值的特征向量, 亦即  $Q$  的对应列, 构成了属于该特征值的特征子空间的一组基. 由  $Q$  的列互相正交, 即得推论 6.1.3.

**推论 6.1.3** 实对称矩阵属于不同特征值的特征向量互相正交.

证二. 对实对称矩阵  $A$ ,  $\lambda_1, \lambda_2$  是两个不同特征值,  $\mathbf{x}_1, \mathbf{x}_2$  是分别属于二者的特征向量. 那么  $\lambda_1 \mathbf{x}_2^T \mathbf{x}_1 = \mathbf{x}_2^T A \mathbf{x}_1 = \mathbf{x}_1^T A \mathbf{x}_2 = \lambda_2 \mathbf{x}_1^T \mathbf{x}_2 = \lambda_2 \mathbf{x}_2^T \mathbf{x}_1$ . 由  $\lambda_1 \neq \lambda_2$ , 必有  $\mathbf{x}_2^T \mathbf{x}_1 = 0$ .  $\square$

对角矩阵  $D$  显然对称, 而  $M^T D M$  也一定对称, 例 1.5.10 中复合映射的矩阵就是这种形式.

**例 6.1.4** 设  $P$  是正交投影矩阵, 即  $P^2 = P^T = P$ . 若  $\lambda$  是  $P$  的特征值,  $P\mathbf{x} = \lambda\mathbf{x}$ , 则  $P^2\mathbf{x} = P(\lambda\mathbf{x}) = \lambda^2\mathbf{x}$ , 因此  $\lambda^2\mathbf{x} = \lambda\mathbf{x}$ ,  $\lambda^2 = \lambda$ , 即  $\lambda = 0, 1$ . 容易证明,  $P\mathbf{x} = \mathbf{x}$  当且仅当  $\mathbf{x} \in \mathcal{R}(P)$ ; 而  $P\mathbf{x} = \mathbf{0}$  当且仅当  $\mathbf{x} \in \mathcal{N}(P)$ . 从这两个子空间各取一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_r$  和  $\mathbf{q}_{r+1}, \dots, \mathbf{q}_n$ , 合并成  $\mathbb{R}^n$  的一组标准正交基. 令  $Q = [\mathbf{q}_1 \ \dots \ \mathbf{q}_n]$ , 则

$$PQ = Q \begin{bmatrix} I_r & \\ & 0 \end{bmatrix} = Q\Lambda. \quad \odot$$

可以看到, 如果  $A = Q\Lambda Q^T$  是其谱分解, 则  $A$  和  $\Lambda$  相似. 和普通的相似关系对比, 区别在于矩阵  $Q$  是正交矩阵. 为此作如下定义.

**定义 6.1.5 (正交相似)** 对实方阵  $A, B$ , 如果存在正交矩阵  $Q$  使得  $Q^T A Q = B$ , 则称  $A$  和  $B$  **正交相似**, 或  $A$  正交相似于  $B$ .

**命题 6.1.6** 实方阵的正交相似关系是等价关系.

证明留给读者.

考虑等价关系的从属概念, 对应与正交相似关系的等价变换称为**正交相似变换**, 即  $T: A \mapsto Q^T A Q$ . 注意正交相似变换不一定只变换实对称矩阵.

再考虑正交相似变换下的不变量. 首先, 因为正交相似变换是一种特殊的相似变换, 因此相似变换下的不变量, 如特征值、代数重数、几何重数, 都是正交相似变换下的不变量. 其次, 对称性也是正交相似变换下的不变量, 即对称矩阵作正交相似变换后得到的矩阵也是对称矩阵, 反对称矩阵作正交相似变换后得到的也是反对称矩阵, 非对称矩阵作正交相似变换后得到的也是非对称矩阵.

定理 6.1.2 说明, 实对称矩阵在正交相似变换下的标准形是对角矩阵. 其他实矩阵的标准形将在第 8 章中阐述.

最后来看一个例子说明实对称矩阵的谱分解的重要性.

**例 6.1.7 (弹簧振子)** 假设四根弹簧和三个振子如图 6.1.1 所示交替依次挂在天花板和地板之间, 弹簧的原始长度为  $l_1, l_2, l_3, l_4$ , 劲度系数分别为  $k_1, k_2, k_3, k_4$ , 振子的质量分别为  $m_1, m_2, m_3$ , 天花板和地板的间距是  $d$ , 而  $d > l_1 + l_2 + l_3 + l_4$ . 系统本处于静止状态, 如果突然将弹簧  $k_4$  剪断, 这些振子将如何运动?

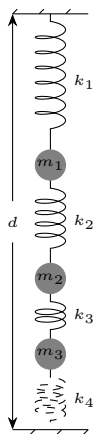


图 6.1.1: 弹簧振子

类似于例 5.4.9, 只需考虑振子的位置关于时间的函数, 分析过程整体与例 5.4.9 类似. 仍然假设三个振子到天花板的距离分别是  $l_1 + x_1, l_1 + l_2 + x_2, l_1 + l_2 + l_3$ , 欲求

$x_1, x_2, x_3$ . 记所求函数为  $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix}$ . 剪断前, 振子的位置在例 1.7.6 已经得出, 即

$\mathbf{x}(0) = (D_{43}^T K_4 D_{43})^{-1} (g M_3 \mathbf{1} - \delta D_{43}^T K_4 \mathbf{e}_4)$ , 其中  $M_3 = \text{diag}(m_1, m_2, m_3)$ .

振子的速度  $\mathbf{v}(t)$  是位移的导数, 即  $\mathbf{v}(t) = \mathbf{x}'(t)$ ,  $\mathbf{v}(0) = \mathbf{0}$ . 振子的加速度是二阶导数  $\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{x}''(t)$ . 振子所受的重力  $\mathbf{g}(t)$ 、振子所受的拉力  $\mathbf{f}(t)$  都与例 1.5.10 相同, 即  $\mathbf{g}(t) = g M_3 \mathbf{1}$ ,  $\mathbf{f}(t) = D_{33}^T K_3 D_{33} \mathbf{x}(t)$ . 利用 Newton 第二定律, 我们有  $\mathbf{g}(t) -$

$\mathbf{f}(t) = \begin{bmatrix} m_1 a_1(t) \\ m_2 a_2(t) \\ m_3 a_3(t) \end{bmatrix} = M_3 \mathbf{a}(t)$ . 亦即  $M_3 \mathbf{a}(t) = g M_3 \mathbf{1} - D_{33}^T K_3 D_{33} \mathbf{x}(t)$ . 因此得到方

程组  $\begin{bmatrix} \mathbf{x}'(t) \\ \mathbf{v}'(t) \end{bmatrix} = \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{a}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{a}_0 - T_0 \mathbf{x}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{a}_0 \end{bmatrix} + \begin{bmatrix} I \\ -T_0 \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{v}(t) \end{bmatrix}$ . 因此, 令

$\mathbf{p}(t) = \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{v}(t) \end{bmatrix}$ ,  $\mathbf{a} = \begin{bmatrix} \mathbf{0} \\ \mathbf{a}_0 \end{bmatrix}$ ,  $T = \begin{bmatrix} I \\ -T_0 \end{bmatrix}$ , 方程就变为  $\mathbf{p}'(t) = \mathbf{a} + T \mathbf{p}(t)$ . 下面来看

$T$  是否有谱分解.

首先考虑  $T_0 = M_3^{-1} D_{33}^T K_3 D_{33}$ , 记  $M_3^{\pm \frac{1}{2}} = \text{diag}(m_1^{\pm \frac{1}{2}}, m_2^{\pm \frac{1}{2}}, m_3^{\pm \frac{1}{2}})$ , 则

$$M_3^{\frac{1}{2}} T_0 M_3^{-\frac{1}{2}} = M_3^{-\frac{1}{2}} D_{33}^T K_3 D_{33} M_3^{-\frac{1}{2}}$$

是对称矩阵, 因此存在谱分解  $Q_0 \Lambda_0 Q_0^T$ . 不难验证  $\Lambda_0$  的对角元素都是正数 (为什么?),

记  $\Lambda_0 = \text{diag}(\mu_1^2, \mu_2^2, \mu_3^2)$ . 因此  $T_0$  有谱分解  $T_0 = M_3^{-\frac{1}{2}} Q_0 \Lambda_0 Q_0^T M_3^{\frac{1}{2}}$ . 先作相似变换

$$\begin{bmatrix} M_3^{-\frac{1}{2}} Q_0 & \\ & M_3^{-\frac{1}{2}} Q_0 \end{bmatrix} T \begin{bmatrix} Q_0^T M_3^{\frac{1}{2}} & \\ & Q_0^T M_3^{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} I \\ -\Lambda_0 \end{bmatrix} =: \tilde{T}. \text{ 再利用置换矩阵 } P =$$

$$P_{35} P_{24} \text{ 作相似变换, 有 } P^T \tilde{T} P = \begin{bmatrix} T_1 & & \\ & T_2 & \\ & & T_3 \end{bmatrix}, \text{ 其中 } T_i = \begin{bmatrix} & 1 \\ \mu_i^2 & \end{bmatrix}, i = 1, 2, 3. \text{ 类似}$$

于例 5.4.9, 每个  $T_i$  都有一对共轭复特征值  $\pm \mu_i i$ , 都可对角化. 因此  $T$  可对角化,  $T$  的特征值为  $\pm \mu_1 i, \pm \mu_2 i, \pm \mu_3 i$ .

记  $T$  的谱分解为  $T = Q \Lambda Q^{-1}$ , 则  $\mathbf{p}'(t) = \mathbf{a} + Q \Lambda Q^{-1} \mathbf{p}(t)$ . 令  $\mathbf{q}(t) = Q^{-1} \mathbf{p}(t)$ ,  $\mathbf{b} = Q^{-1} \mathbf{a}$ , 则  $\mathbf{q}''(t) = \mathbf{b} + \Lambda \mathbf{q}(t)$ . 问题就转化为六个互不相关的一阶常系数常微分方程  $q_i'(t) = b_i + \lambda_i q_i(t)$ . 根据常微分方程知识可知, 其解是  $q_i(t) = C_i e^{\lambda_i t} - \frac{b_i}{\lambda_i}$ , 其中  $C_i$  由初

始值确定. 记  $e^{\Lambda t} = \text{diag}(e^{\lambda_i t})$ ,  $\mathbf{c} = [C_i]$ , 就有  $\mathbf{q}(t) = e^{\Lambda t} \mathbf{c} - \Lambda^{-1} \mathbf{b}$ . 于是  $\mathbf{p}(t) = Q e^{\Lambda t} \mathbf{c} -$

$T^{-1} \mathbf{a}$ . 注意  $\mathbf{p}(0) = \begin{bmatrix} \mathbf{x}(0) \\ \mathbf{v}(0) \end{bmatrix} = Q \mathbf{c} - T^{-1} \mathbf{a}$ , 由此即可解出  $\mathbf{c} = Q^{-1} \begin{bmatrix} \mathbf{x}(0) - T_0^{-1} \mathbf{a}_0 \\ 0 \end{bmatrix}$ . 因此

$$\begin{bmatrix} \mathbf{x}(t) \\ \mathbf{v}(t) \end{bmatrix} = Qe^{At}Q^{-1} \begin{bmatrix} \mathbf{x}(0) \\ 0 \end{bmatrix} + (I - Qe^{At}Q^{-1}) \begin{bmatrix} T_0^{-1}\mathbf{a}_0 \\ 0 \end{bmatrix}.$$
 注意  $\begin{bmatrix} \mathbf{x}(0) \\ 0 \end{bmatrix}$  表示了例 1.7.6 的状态,  $\begin{bmatrix} T_0^{-1}\mathbf{a}_0 \\ 0 \end{bmatrix}$  表示了例 1.5.10 的状态, 因此振子的运动状态就可以视作二者的加权平均, 权重由时间决定, 而振子就在往复振荡中不断重复这两种状态.  $\odot$

## 习题

**练习 6.1.1** 求下列实对称矩阵的谱分解.

$$\begin{aligned}
 &1. \begin{bmatrix} 2 & -2 & 0 \\ -2 & 1 & -2 \\ 0 & -2 & 0 \end{bmatrix}; \quad 2. \begin{bmatrix} 2 & 2 & -2 \\ 2 & 5 & -4 \\ -2 & -4 & 5 \end{bmatrix}; \quad 3. \begin{bmatrix} 0 & 0 & 4 & 1 \\ 0 & 0 & 1 & 4 \\ 4 & 1 & 0 & 0 \\ 1 & 4 & 0 & 0 \end{bmatrix}; \quad 4. \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.
 \end{aligned}$$

**练习 6.1.2** 设三阶实对称矩阵  $A$  的特征值是  $0, 3, 3$ ,  $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in V_{\lambda=0}$ ,  $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \in V_{\lambda=3}$ , 求  $A$ .

**练习 6.1.3** 设实对称矩阵  $A = \begin{bmatrix} a & 1 & 1 & -1 \\ 1 & a & -1 & 1 \\ 1 & -1 & a & 1 \\ -1 & 1 & 1 & a \end{bmatrix}$ , 有一个单特征值  $-3$ , 求  $a$  的值和  $A$  的谱分解.

**练习 6.1.4** 设三阶实对称矩阵  $A$  的各行元素之和均为  $3$ ,  $\mathbf{a}_1 = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}$ ,  $\mathbf{a}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \in \mathcal{N}(A)$ , 求  $A$  及其谱分解.

**练习 6.1.5** 求下列实对称矩阵  $A$  的谱分解.

- $A$  满足  $A^3 = O$ .
- $A = a_1 \mathbf{x}_1 \mathbf{x}_1^T + a_2 \mathbf{x}_2 \mathbf{x}_2^T$ , 其中  $\mathbf{x}_1, \mathbf{x}_2$  是  $\mathbb{R}^2$  的一组标准正交基,  $a_1, a_2$  为实数.
- $A = \begin{bmatrix} O & M \\ M & O \end{bmatrix}$ , 其中  $M$  是  $n$  阶对称矩阵, 有谱分解  $M = Q\Lambda Q^T$ .

**练习 6.1.6** 设矩阵  $A = \begin{bmatrix} 2 & b \\ 1 & 0 \end{bmatrix}$ . 计算满足下列条件的  $b$  的取值范围.

- $A$  不可逆.
- $A$  可以正交对角化.
- $A$  不可对角化.

**练习 6.1.7** 计算下列矩阵及其特征值. 哪些是对称矩阵? 哪些矩阵的特征值是  $\pm 1$ ? 由此看出正交相似的特殊性.

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

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

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

练习 6.1.8 证明命题 6.1.6.

练习 6.1.9 证明, 两个实对称矩阵正交相似, 当且仅当它们具有相同的特征多项式.

练习 6.1.10 设实对称矩阵  $A$  满足  $A^5 = I_n$ , 证明  $A = I_n$ .

练习 6.1.11 设  $A_1, \dots, A_m$  是  $m$  个两两可交换的实对称矩阵, 证明它们可以同时正交对角化.

练习 6.1.12 当  $n$  充分大时, 矩阵  $A = \begin{bmatrix} 1 & 10^{-n} \\ 0 & 1 + 10^{-n} \end{bmatrix}$  非常接近对称矩阵. 计算其谱分解, 并求两个线性无关的特征向量的夹角.

练习 6.1.13 设  $\lambda_1$  是实对称矩阵  $A$  的最大的特征值. 证明  $A$  的左上角元素  $a_{11} \leq \lambda_1$ .

练习 6.1.14 考虑下列复对称矩阵.

1. 设复对称矩阵  $A = \begin{bmatrix} i & 1 \\ 1 & -i \end{bmatrix}$ , 计算  $A^2$ , 并判断  $A$  是否可对角化.

2. 设复对称矩阵  $F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$ , 计算  $F_4^2$  和  $F_4 \overline{F_4}^T$ ; 根据  $F_4^2$  的特征值、 $F_4$  的迹和

行列式, 求  $F_4$  的所有特征值. 是否存在正交矩阵  $Q$ , 使得  $Q^T F_4 Q$  是对角阵?

这里  $F$  是一个四阶离散 Fourier 变换的表示矩阵, 这类矩阵在信号处理和数值积分中有重要应用.

练习 6.1.15 设实对称矩阵  $S = \begin{bmatrix} O & A \\ A^T & O \end{bmatrix}$ .

1. 证明,  $Sx = \lambda x$ , 当且仅当  $x = \begin{bmatrix} y \\ z \end{bmatrix}$ , 满足  $Az = \lambda y, A^T y = \lambda z$ .

2. 证明, 如果  $\lambda$  是  $S$  的特征值, 则  $-\lambda$  也是  $S$  的特征值.

3. 证明, 如果  $\lambda \neq 0$  是  $S$  的特征值, 则  $\lambda^2$  是  $A^T A$  的特征值, 也是  $AA^T$  的特征值.

4. 证明,  $AA^T$  和  $A^T A$  的非零特征值相同, 且有相同的重数.

5. 分别取  $A = I_2$  或  $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ , 求对应  $S$  的谱分解.

练习 6.1.16 构造一个实方阵  $A$ , 满足  $AA^T = A^T A$  但  $A \neq A^T$ , 并验证  $A$  和  $A^T$  具有相同的特征值和特征向量. 注意, 这里相同的特征向量不意味着对应的特征值相同.

提示: 事实上, 对实方阵  $A$ , 如果  $AA^T = A^T A$ , 则  $A$  和  $A^T$  具有相同的特征值和特征向量.

**练习 6.1.17 (二阶差分矩阵与二阶导数)** 在微积分中,  $f(x)$  的二阶导数  $f''(x) \approx \frac{f(x+\delta)+f(x-\delta)-2f(x)}{2\delta^2}$ . 如果将一个区间等分, 将函数  $f(x)$  在等分点处的取值列成一个向量  $\mathbf{f}$ , 则  $\mathbf{f}$  就是离散化的函数, 计算中可以代替  $f(x)$ . 而  $f(x)$  的二阶导数就可以用  $\frac{1}{2h^2}D\mathbf{f}$  来表示, 其中  $h$  是每个分出的小区间的长度, 而  $D =$

$$D = \begin{bmatrix} * & 1 & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & \ddots & & \\ & & \ddots & \ddots & 1 & \\ & & & 1 & -2 & 1 \\ & & & & 1 & * \end{bmatrix}, \text{ 其中“*”处元素与函数在区间端点处满足的条件有关.}$$

1. 求  $A_3 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$  的谱分解, 并将特征向量写成  $\begin{bmatrix} \sin d \\ \sin 2d \\ \sin 3d \end{bmatrix}$  的形式. 提示: 这是离散化的正弦函数.

2. 求  $B_4 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$  的谱分解, 并将特征向量写成  $\begin{bmatrix} \cos d \\ \cos 3d \\ \cos 5d \\ \cos 7d \end{bmatrix}$  的形式. 提示: 这是离散化的余弦函数.

注意,  $\frac{d^2}{dx^2} \sin kx = -k^2 \sin kx$ ,  $\frac{d^2}{dx^2} \cos kx = -k^2 \cos kx$ . 对离散化的情形, 也有类似结论. 这一点将在第 7 章中得到更多讨论. 在 Fourier 分析中, 将函数写成正弦函数和余弦函数的和有极大的好处; 而在计算上, 将向量写成离散正弦或离散余弦的线性组合, 也有类似的好处. 例如, 图片的 JPEG 压缩就常用  $B_8$  谱分解产生的正交基来进行处理.

**练习 6.1.18** 考虑正方形铁丝上的热扩散问题, 设四个顶点的温度组成向量  $\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix}$ . 在热量扩散

时, 如果一个点的温度低于周围点温度的平均值, 则该点温度升高; 反之则该点温度降低. 假设每经过一个时间单位, 一个点的温度变化与周围点平均温度和该点温度之差成正比, 比例系数为  $k$ .

1. 写出矩阵  $A$ , 使得  $A\mathbf{t}$  表示经过一个单位时间之后四个点的温度.
2. 令  $H$  为 4 阶 Hadamard 矩阵 (见练习 5.3.5), 计算  $AH$ .
3. 求  $A$  的谱分解.
4. 求  $\lim_{t \rightarrow \infty} A^t$ , 它是否与  $k$  有关? 经过足够长的时间后, 最终的温度分布是什么?

**练习 6.1.19** 采集  $n$  个人的三项数据, 例如身高、体重、收入, 组成向量  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \in \mathbb{R}^n$ . 令  $x_1, x_2, x_3$  为三组数据的平均值,  $\mathbf{u}$  为分量全是 1 的向量. 定义向量  $\mathbf{a}_i, \mathbf{a}_j$  的协方差为  $\text{cov}(\mathbf{a}_i, \mathbf{a}_j) = \frac{1}{n}(\mathbf{a}_i - x_i \mathbf{u})^T (\mathbf{a}_j - x_j \mathbf{u})$ . 称  $C = [\text{cov}(\mathbf{a}_i, \mathbf{a}_j)]_{3 \times 3}$  为数据的协方差矩阵.

1. 令  $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix}$ ,  $U$  为所有元素都是 1 的  $n \times 3$  矩阵, 求常数  $k$ , 使得  $kUU^T A = \begin{bmatrix} x_1 \mathbf{u} & x_2 \mathbf{u} & x_3 \mathbf{u} \end{bmatrix}$ .

2. 用  $A, U$  来表示  $C$ , 并证明  $C$  是对称矩阵.
3. 任取  $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^3$ , 证明向量  $A\mathbf{y}_1, A\mathbf{y}_2$  的协方差是  $\mathbf{y}_1^T C \mathbf{y}_2$ .
4. 设  $C$  有谱分解  $C = Q\Lambda Q^T$ , 考虑  $[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3]Q$  的三个列向量对应的三组数据, 证明这三组数据两两协方差为零.

协方差为零的两组数据称为不相关的数据, 利用谱分解从原本相关的数据中得到不相关的数据的线性组合, 是统计学中很重要的一个方法.

## 6.2 正定矩阵

例 5.4.9 中, 弹簧振子的动能  $E = \frac{1}{2}mv^2$ , 弹簧振子的势能  $E = \frac{1}{2}kx^2$ , 而系统总机械能  $E = \frac{1}{2}\mathbf{p}K\mathbf{p}$ . 类似地, 多质点系统的能量常常可用  $\mathbf{x}^T A \mathbf{x}$  表示, 其中  $\mathbf{x} \in \mathbb{R}^n$ . 关于  $\mathbf{x}$  的函数  $\mathbf{x}^T A \mathbf{x}$  常称为**能量函数**. 很多物理系统的能量函数  $\mathbf{x}^T A \mathbf{x}$  具有**正定性**, 即  $\mathbf{x}^T A \mathbf{x} \geq 0$ , 且  $\mathbf{x}^T A \mathbf{x} = 0$  当且仅当  $\mathbf{x} = \mathbf{0}$ . 特别地, 当  $A = I_n$  时, 这就是内积的正定性.

**定义 6.2.1 (正定矩阵)** 给定  $n$  阶实矩阵  $A$ , 如果对任意非零向量  $\mathbf{x} \in \mathbb{R}^n$ , 都有  $\mathbf{x}^T A \mathbf{x} > 0$ , 则称  $A$  **正定**.

若  $D = \text{diag}(d_i)$  是实对角矩阵, 则对任意  $\mathbf{x} = [x_i]$ ,  $\mathbf{x}^T D \mathbf{x} = d_1 x_1^2 + \cdots + d_n x_n^2$ , 因此  $D$  正定当且仅当所有对角元素都是正数. 容易验证, 对任意可逆矩阵  $T$ ,  $A$  正定当且仅当  $T^T A T$  正定.

实对称矩阵  $A$  有谱分解  $A = Q\Lambda Q^T$ , 其中  $\Lambda$  的对角元素是  $A$  的特征值. 回忆矩阵的三角分解. 如果条件合适, 则对称矩阵  $A$  有  $\text{LDL}^T$  分解  $A = \text{LDL}^T$ , 其中  $D$  的对角元素是主元,  $L$  为单位下三角矩阵. 这时,  $A$  正定、 $\Lambda$  正定和  $D$  正定都等价, 这就把解方程组和求特征值这两个矩阵计算的主题联系起来了.

易见,  $\Lambda$  的对角元素都是正数,  $D$  的对角元素也都是正数 (为什么?). 而  $\det(A) = \det(\Lambda) = \det(D) > 0$ . 事实上, 如下命题 6.2.2 说明,  $A$  的第  $i$  个顺序主子阵, 记为  $A_i$ , 的行列式  $\det(A_i) > 0$ .

矩阵  $A$  的第  $i$  个顺序主子阵的行列式称为其第  $i$  个**顺序主子式**.

**命题 6.2.2** 对实对称矩阵  $A$ , 以下叙述等价:

1.  $A$  正定;
2.  $A$  的特征值都是正数;
3. 存在可逆矩阵  $X$ , 使得  $A = T T^T$ ;
4.  $A$  有  $\text{LDL}^T$  分解, 且  $D$  的对角元素都是正数;



5.  $A$  的顺序主子式都是正数;

6.  $A$  的顺序主子阵都正定.

证. 采用轮转证法.

“1  $\Rightarrow$  2”: 对任意特征值  $\lambda$ , 任取对应的特征向量  $\mathbf{x}$ , 则  $\mathbf{x}^T A \mathbf{x} = \lambda \mathbf{x}^T \mathbf{x} > 0$ , 于是  $\lambda > 0$ .

“2  $\Rightarrow$  3”: 对称矩阵  $A$  有谱分解  $A = Q \Lambda Q^T$ , 由于  $\Lambda$  的对角元素是特征值且都是正数, 因此存在对角矩阵  $D$  使得  $D^2 = \Lambda$ . 令  $T = QD$ , 即得结论.

“3  $\Rightarrow$  4”: 设  $T$  的 QR 分解为  $Q\tilde{L}^T$ , 则  $A = \tilde{L}Q^TQ\tilde{L}^T = \tilde{L}\tilde{L}^T$ . 令  $\tilde{L} = L\tilde{D}$ , 其中  $L$  是单位上三角矩阵,  $\tilde{D}$  是对角矩阵, 则有  $A = L\tilde{D}^2L^T$ . 令  $D = \tilde{D}^2$  即得.

“4  $\Rightarrow$  5”: 由  $A$  有 LDL<sup>T</sup> 分解, 按第  $i$  个顺序主子阵对  $A$  分块, 有

$$\begin{bmatrix} A_i & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix} = A = LDL^T = \begin{bmatrix} L_i & \\ & L_{22} \end{bmatrix} \begin{bmatrix} D_i & \\ & D_{22} \end{bmatrix} \begin{bmatrix} L_i & \\ & L_{22} \end{bmatrix}^T,$$

计算即得  $A_i = L_i D_i L_i^T$ , 因此第  $i$  个顺序主子式

$$\det(A_i) = \det(L_i) \det(D_i) \det(L_i^T) = \det(D_i) > 0.$$

“5  $\Rightarrow$  6”: 对  $n$  用数学归纳法. 当  $n = 1$  时, 显然, 现假设命题对任意  $n - 1$  阶实对称矩阵成立. 对  $n$  阶实对称矩阵  $A$ , 对  $A$  分块  $A = \begin{bmatrix} A_{n-1} & \mathbf{a} \\ \mathbf{a}^T & a_{nn} \end{bmatrix}$ , 则  $A_{n-1}$  的顺序主子式都是正数,  $\det(A)$  也是正数. 根据归纳假设,  $A_{n-1}$  的顺序主子阵都正定, 只需再证  $A$  正定. 由于  $A_{n-1}$  正定, 利用 “1  $\Rightarrow$  2  $\Rightarrow$  3”, 即得  $A_{n-1}$  可逆. 做分块矩阵 LU 分解, 如 (1.7.5), 有

$$A = \begin{bmatrix} A_{n-1} & \mathbf{a} \\ \mathbf{a}^T & a_{nn} \end{bmatrix} = \begin{bmatrix} I & \\ \mathbf{a}^T A_{n-1}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{n-1} & \\ & a_{nn} - \mathbf{a}^T A_{n-1}^{-1} \mathbf{a} \end{bmatrix} \begin{bmatrix} I & A_{n-1}^{-1} \mathbf{a} \\ & I \end{bmatrix}.$$

计算行列式即有  $a_{nn} - \mathbf{a}^T A_{n-1}^{-1} \mathbf{a} = \frac{\det(A)}{\det(A_{n-1})} > 0$ . 容易验证  $\begin{bmatrix} A_{n-1} & \\ & a_{nn} - \mathbf{a}^T A_{n-1}^{-1} \mathbf{a} \end{bmatrix}$  正定, 因此  $A$  正定.

“6  $\Rightarrow$  1”: 显然. □

命题 6.2.2 的证明中 “3  $\Rightarrow$  4” 说明, 对实对称正定矩阵  $A$ , 存在上三角矩阵  $L$ , 使得  $A = LL^T$ , 这称为  $A$  的 **Cholesky 分解**.

**定义 6.2.3** 给定  $n$  阶实矩阵  $A$ , 如果对任意非零向量  $\mathbf{x} \in \mathbb{R}^n$ , 都有

1.  $\mathbf{x}^T A \mathbf{x} > 0$ , 则称矩阵  $A$  **正定**, 如前定义;

2.  $\mathbf{x}^T A \mathbf{x} \geq 0$ , 则称矩阵  $A$  半正定;

3.  $\mathbf{x}^T A \mathbf{x} < 0$ , 则称矩阵  $A$  负定;

4.  $\mathbf{x}^T A \mathbf{x} \leq 0$ , 则称矩阵  $A$  半负定;

如果  $A$  不满足以上任何一种条件, 则称  $A$  不定.

显然, 负定矩阵和半负定矩阵能够转化为正定矩阵和半正定矩阵来研究.

对称半正定矩阵具有和对称正定矩阵类似的刻画.

**命题 6.2.4** 对实对称矩阵  $A$ , 以下叙述等价:

1.  $A$  半正定;
2.  $A$  的特征值都是非负数;
3. 存在矩阵  $T$ , 使得  $A = TT^T$ ;
4.  $A$  存在  $LDL^T$  分解, 且  $D$  的对角元素都是非负数.

证明留给读者. 读者还可以思考如下问题: 为什么  $A$  的顺序主子式都是非负数不能说明  $A$  半正定?

最后我们列出不定矩阵的性质.

**命题 6.2.5** 对  $n$  阶实对称矩阵  $A$ , 如果存在  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , 使得  $\mathbf{x}^T A \mathbf{x} > 0, \mathbf{y}^T A \mathbf{y} < 0$ , 则存在非零向量  $\mathbf{z} \in \mathbb{R}^n$ , 使得  $\mathbf{z}^T A \mathbf{z} = 0$ .

证. 拟设  $\mathbf{z} = \mathbf{x} + k\mathbf{y}$ . 考虑  $\mathbf{z}^T A \mathbf{z} = \mathbf{x}^T A \mathbf{x} + 2k\mathbf{x}^T A \mathbf{y} + k^2\mathbf{y}^T A \mathbf{y} = 0$ . 注意判别式  $(2\mathbf{x}^T A \mathbf{y})^2 - 4(\mathbf{x}^T A \mathbf{x})(\mathbf{y}^T A \mathbf{y}) > 0$ , 故方程有实数根. 因此存在  $k \in \mathbb{R}$ , 使得拟设的  $\mathbf{z}$  满足条件.  $\square$

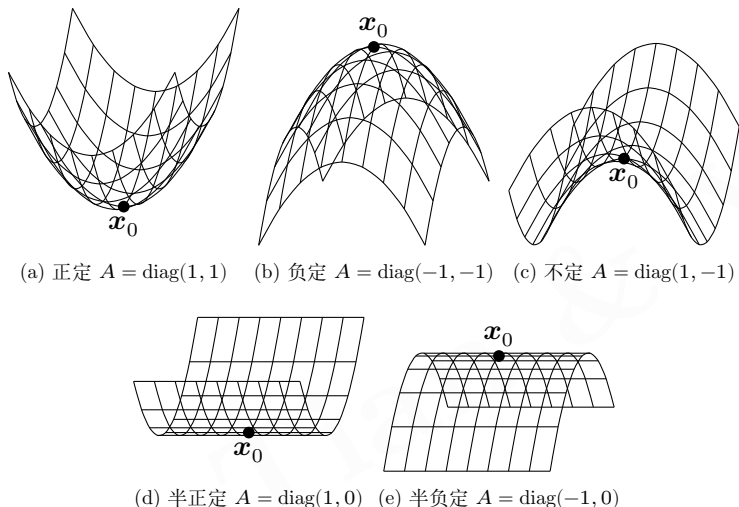
**例 6.2.6 (Hesse 矩阵)** 假定  $n$  元实函数  $f(\mathbf{x})$  在一个单连通区域上有定义,  $\mathbf{x}_0$  是区域内一点. 如果  $f$  三阶连续可微, 则  $f$  在半径为  $\rho$  的小邻域上可以展开

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T H(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\rho^2),$$

其中  $\nabla f(\mathbf{x}_0) = \begin{bmatrix} f'_{x_1} \\ \vdots \\ f'_{x_n} \end{bmatrix}$ ,  $H(\mathbf{x}_0) = \begin{bmatrix} f''_{x_1 x_1} & \cdots & f''_{x_1 x_n} \\ \vdots & & \vdots \\ f''_{x_n x_1} & \cdots & f''_{x_n x_n} \end{bmatrix}$ , 分别称为  $f$  在  $\mathbf{x}_0$  点处的梯

度和 Hesse 矩阵. 如果  $\mathbf{x}_0$  是一个驻点, 即  $\nabla f(\mathbf{x}_0) = \mathbf{0}$ , 则  $f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T H(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\rho^2)$ . 因此, 可通过判断 Hesse 矩阵的正定性来确定  $\mathbf{x}_0$  是否是极值点:

1. 如果  $H(\mathbf{x}_0)$  正定, 则在  $\mathbf{x}_0$  邻域内任意点的函数值都比  $f(\mathbf{x}_0)$  大,  $\mathbf{x}_0$  是极小值点;
2. 如果  $H(\mathbf{x}_0)$  负定, 则在  $\mathbf{x}_0$  邻域内任意点的函数值都比  $f(\mathbf{x}_0)$  小,  $\mathbf{x}_0$  是极大值点;
3. 如果  $H(\mathbf{x}_0)$  不定, 则在  $\mathbf{x}_0$  邻域内点的函数值某些比  $f(\mathbf{x}_0)$  小, 某些比  $f(\mathbf{x}_0)$  大,  $\mathbf{x}_0$  不是极值点;
4. 如果  $H(\mathbf{x}_0)$  半正定或半负定, 则条件不足, 还需要补充其他条件才能判断. ☺

图 6.2.1: Hesse 矩阵与能量函数  $\mathbf{x}^T A \mathbf{x}$ 

最后, 从前述可见, 对实对称矩阵  $A$ ,  $\mathbf{x}^T A \mathbf{x}$  是一个重要的衡量. 为此作如下定义.

**定义 6.2.7 (Rayleigh 商)** 给定实矩阵  $A$  和非零向量  $\mathbf{x} \in \mathbb{R}^n$ , 实数  $\frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$  称为  $\mathbf{x}$  关于  $A$  的 Rayleigh 商.

对实对称矩阵而言, Rayleigh 商和特征值之间存在着重要联系.

**命题 6.2.8** 设对称矩阵  $A$  的特征值为  $\lambda_1 \geq \dots \geq \lambda_n$ , 相应的特征向量为  $\mathbf{q}_1, \dots, \mathbf{q}_n$ , 则

$$\lambda_1 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad \lambda_i = \max_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x} \perp \text{span}(\mathbf{q}_1, \dots, \mathbf{q}_{i-1})}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, i = 2, \dots, n.$$

类似地,

$$\lambda_n = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad \lambda_i = \min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{x} \perp \text{span}(\mathbf{q}_{i+1}, \dots, \mathbf{q}_n)}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, i = 1, \dots, n-1.$$

证. 先说明后者能由前者简单得到. 考察  $-A$ , 注意  $-A$  的特征值是  $-\lambda_n \geq \dots \geq -\lambda_1$ , 由前者就得到  $-\lambda_n = \max_{x \neq 0} \frac{x^T(-A)x}{x^T x}$ , 于是  $\lambda_n = \min_{x \neq 0} \frac{x^T A x}{x^T x}$ . 另一等式类似.

下证  $\lambda_1 = \max_{x \neq 0} \frac{x^T A x}{x^T x}$ . 记  $A = Q \Lambda Q^T$  为  $A$  的谱分解. 则

$$\max_{x \neq 0} \frac{x^T A x}{x^T x} = \max_{x \neq 0} \frac{x^T Q \Lambda Q^T x}{x^T Q Q^T x} = \max_{y=Q^T x \neq 0} \frac{y^T \Lambda y}{y^T y} = \max_{y \neq 0} \frac{\lambda_1 y_1^2 + \dots + \lambda_n y_n^2}{y_1^2 + \dots + y_n^2} = \lambda_1,$$

最后一个等式成立, 是因为  $\lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \leq \lambda_1 y_1^2 + \dots + \lambda_1 y_n^2 = \lambda_1 (y_1^2 + \dots + y_n^2)$ , 而令  $y_1 = 1, y_2 = \dots = y_n = 0$ , 等式成立.

再证  $\lambda_i = \max_{x \perp \text{span}(q_1, \dots, q_{i-1})} \frac{x^T A x}{x^T x}$ . 注意  $x \perp \text{span}(q_1, \dots, q_{i-1}) \Leftrightarrow x \in \text{span}(q_i, \dots, q_n)$ .

则

$$\begin{aligned} \max_{\substack{x \neq 0 \\ x \in \text{span}(q_i, \dots, q_n)}} \frac{x^T A x}{x^T x} &= \max_{\substack{x \neq 0 \\ x \in \text{span}(q_i, \dots, q_n)}} \frac{x^T Q \Lambda Q^T x}{x^T Q Q^T x} \\ &= \max_{\substack{y \neq 0 \\ y \in \text{span}(e_i, \dots, e_n)}} \frac{y^T \Lambda y}{y^T y} \quad (\text{变量替换 } y = Q^T x) \\ &= \max_{y \in \text{span}(e_i, \dots, e_n)} \frac{\lambda_i y_i^2 + \dots + \lambda_n y_n^2}{y_i^2 + \dots + y_n^2} = \lambda_i. \end{aligned}$$

□

## 习题

**练习 6.2.1** 判断下列矩阵是否正定.

1.  $\begin{bmatrix} 5 & 6 \\ 6 & 7 \end{bmatrix}$ .

2.  $\begin{bmatrix} -1 & -2 \\ -2 & -5 \end{bmatrix}$ .

3.  $\begin{bmatrix} 1 & 10 \\ 10 & 100 \end{bmatrix}$ .

4.  $\begin{bmatrix} 1 & 10 \\ 10 & 101 \end{bmatrix}$ .

5.  $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 5 \end{bmatrix}$ .

6.  $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & 5 \end{bmatrix}$ .

**练习 6.2.2** 考虑实矩阵  $S = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix}$ . 求使得下列条件成立的  $b$  的取值范围.

1.  $S$  不正定但是半正定.

2.  $S$  不定.

3.  $S$  半负定.

**练习 6.2.3** 下列实矩阵中未知元素满足什么条件时, 矩阵正定? 半正定?

1.  $\begin{bmatrix} 1 & b \\ b & 9 \end{bmatrix}$ .

2.  $\begin{bmatrix} 2 & 4 \\ 4 & c \end{bmatrix}$ .

3.  $\begin{bmatrix} c & b \\ b & c \end{bmatrix}$ .

4.  $\begin{bmatrix} c & 1 & 1 \\ 1 & c & 1 \\ 1 & 1 & c \end{bmatrix}$ .

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

6.  $\begin{bmatrix} s & -4 & -4 \\ -4 & s & -4 \\ -4 & -4 & s \end{bmatrix}$ .

7.  $\begin{bmatrix} t & 3 & 0 \\ 3 & t & 4 \\ 0 & 4 & t \end{bmatrix}$ .

8.  $\begin{bmatrix} a & a & a \\ a & a+c & a-c \\ a & a-c & a+c \end{bmatrix}$ .

**练习 6.2.4** 设  $A, B$  是  $n$  阶实对称矩阵.

1.

$A$  正定. 证明, 存在可逆矩阵  $T$ , 使得  $T^T A T$  和  $T^T B T$  同时是对角矩阵.

**练习 6.2.5** 设  $A, B$  是  $n$  阶实对称矩阵,  $A, B$  半正定. 证明, 存在可逆矩阵  $T$ , 使得  $T^T A T$  和  $T^T B T$  同时是对角矩阵.

**练习 6.2.6** 设  $A$  对称正定,  $B$  是实矩阵.

1. 证明, 对任意整数,  $A^k$  也正定.
2. 若存在正整数  $r$ , 使得  $A^r B = B A^r$ , 证明  $AB = BA$ .

**练习 6.2.7** 对  $n$  阶实对称矩阵  $A$ , 证明, 当实数  $t$  充分大时,  $tI_n + A$  正定.

**练习 6.2.8** 对  $n$  阶实对称矩阵  $A$ , 证明, 存在正实数  $c$ , 使得对任意  $n$  维列向量  $x$ , 都有  $|x^T A x| \leq c x^T x$ .

**练习 6.2.9 (Hadamard 不等式)** 求证:

1. 如果  $A$  正定, 那么任取列向量  $y$ , 分块矩阵  $\begin{bmatrix} A & y \\ y^T & 0 \end{bmatrix}$  负定;
2. 如果  $A = [a_{ij}]$  正定, 那么  $\det(A) \leq a_{nn} A_{n-1}$ , 其中  $A_{n-1}$  是  $A$  的  $n-1$  阶顺序主子式;
3. 如果  $A$  正定, 那么  $\det(A) \leq a_{11} a_{22} \cdots a_{nn}$ .
4. 如果实矩阵  $T = [t_{ij}]$  可逆, 那么  $\det(T)^2 \leq \prod_{i=1}^n (t_{1i}^2 + t_{2i}^2 + \cdots + t_{ni}^2)$ .

提示: 练习 4.2.28 用不同方法证明了相同结论.

**练习 6.2.10** 证明  $A = \left[ \frac{1}{i+j} \right]_{n \times n}$  正定.

**练习 6.2.11** 设  $A, B$  是实对称矩阵, 如果  $A - B$  正定, 则记作  $A \succ B$ . 求证:

1.  $A \succ B, B \succ C$  可以推出  $A \succ C$ .
2.  $A \succ B$  和  $B \succ A$  不可能同时成立.
3. 对任意实对称矩阵  $A$ , 都存在实数  $k_1, k_2$  使得  $k_1 I_n \succ A \succ k_2 I_n$ .

**练习 6.2.12** 举例说明, 实对称矩阵  $A$  的所有顺序主子式都非负, 但  $A$  并不半正定.

**练习 6.2.13** 证明命题 6.2.4.

**练习 6.2.14** 证明, 实对称矩阵半正定, 当且仅当它的所有主子式都非负.

**练习 6.2.15** 设  $A$  是实对称矩阵, 证明,

1.  $A$  半正定, 当且仅当存在实对称矩阵  $B$ , 使得  $A = B^2$ .
2. 若  $A$  半正定, 则存在唯一的半正定实对称矩阵  $B$ , 使得  $A = B^2$ .

**练习 6.2.16** 证明, 若实对称矩阵对角占优, 且对角元素全为正数, 则该矩阵正定.

**练习 6.2.17** 证明或举出反例:

1. 如果  $A, B$  对称正定, 则  $A + B$  正定.
2. 如果  $A, B$  对称半正定, 则  $A + B$  半正定.
3. 如果  $A, B$  对称不定, 则  $A + B$  不定.
4. 如果  $A$  列满秩,  $B$  对称正定, 则  $A^T B A$  正定.
5. 如果  $S = A^T A$  且  $A$  有 QR 分解  $A = QR$ , 则  $S = R^T R$  是  $S$  的 Cholesky 分解.
6. 如果  $A, B$  正定, 则  $AB$  的特征值都是正数.

**练习 6.2.18** 考虑实对称矩阵  $S = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ . 证明  $S$  正定当且仅当  $A$  及其 Schur 补都正定.

**练习 6.2.19 (非对称矩阵的 Rayleigh 商)** 1. 如下对实矩阵的特征值都是实数的证明, 哪里有问题?

设  $Ax = \lambda x$ , 于是  $x^T Ax = \lambda x^T x$ ,  $\lambda = \frac{x^T Ax}{x^T x}$ . 由于分子、分母都是实数, 特征值  $\lambda$  也是实数.

2. 设  $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ , 计算  $\max_{x \neq 0} \frac{x^T Ax}{x^T x}$ .

3. 设  $A = \begin{bmatrix} 2 & 10 \\ -2 & -2 \end{bmatrix}$ , 计算  $\max_{x \neq 0} \frac{x^T Ax}{x^T x}$ .

4. 对任意实矩阵  $A$  与非零实向量  $x$ , 证明  $\frac{x^T Ax}{x^T x} = \frac{x^T A^T x}{x^T x} = \frac{x^T Bx}{x^T x}$ , 其中  $B = \frac{A+A^T}{2}$  是对称矩阵.

## 6.3 奇异值分解

如果不加特别说明, 本节中的矩阵都是实矩阵.

目前对特征值的讨论都局限于方阵. 对  $m \times n$  矩阵  $A$ , 我们能够考虑特征值吗? 特征值的关键是方程  $Ax = \lambda x$ . 显然, 当  $m \neq n$  时,  $Ax = \lambda x$  不能成立. 然而,  $A$  定义了一个  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射  $A$ , 而  $A^T$  定义了一个  $\mathbb{R}^m$  到  $\mathbb{R}^n$  的线性映射  $A^*$ , 那么  $Ax = \lambda y$ ,  $A^T y = \lambda x$  可能成立吗? 这表示  $x$  被映射  $A$  变成了  $\mathbb{R}^m$  中和  $y$  共线的向量, 而这个向量被映射  $A^*$  变成了和  $x$  共线的向量. 事实上, 注意到两个线性映射的复合  $A^* A$  的表示矩阵是  $A^T A$ , 而需要求解的是方程  $A^T Ax = \lambda^2 x$ . 可见  $\lambda^2$  恰好是对称半正定矩阵  $A^T A$  的特征值.

### 6.3.1 基本性质

**定义 6.3.1 (奇异值)** 给定  $m \times n$  矩阵  $A$ , 如果存在非零向量  $x \in \mathbb{R}^n, y \in \mathbb{R}^m, \sigma \geq 0$ , 使得  $Ax = \sigma y, A^T y = \sigma x$ , 则称  $\sigma$  为  $A$  的一个奇异值,  $x$  为  $A$  的属于  $\sigma$  的一个右奇异向量,  $y$  为  $A$  的属于  $\sigma$  的一个左奇异向量.

容易验证,  $A$  的右奇异向量是  $A^T A$  的特征向量, 类似地,  $A$  的左奇异向量是  $AA^T$  的特征向量, 而  $A$  的奇异值是  $A^T A$  或  $AA^T$  的特征值的算术平方根.

类似于实对称矩阵的谱分解, 我们有定理 6.3.2.

**定理 6.3.2 (奇异值分解)** 给定  $m \times n$  矩阵  $A$ , 存在  $m$  阶正交矩阵  $U$  和  $n$  阶正交矩阵  $V$ , 使得  $A = U\Sigma V^T$ , 其中

$$\Sigma = \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}, \quad \Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}, \quad \sigma_1 \geq \cdots \geq \sigma_r > 0.$$

证. 由于  $A^T A$  对称半正定, 因此有谱分解  $A^T A = V\Lambda V^T$ , 其中  $\Lambda$  的对角元素都是非负数, 并由大到小排列. 设  $\Lambda = \begin{bmatrix} \Sigma_r^2 & \\ & O \end{bmatrix}$ , 其中  $\Sigma_r$  是  $r$  阶方阵. 记  $V = [V_1 \ V_2]$ , 其中  $V_1$  是  $n \times r$  矩阵,  $V_2$  是  $n \times (n-r)$  矩阵. 则  $V_1^T A^T A V_1 = \Sigma_r^2$ ,  $V_2^T A^T A V_2 = O$ , 即  $AV_2 = O$ . 令  $U_1 = AV_1 \Sigma_r^{-1}$ , 则有  $U_1^T U_1 = \Sigma_r^{-1} V_1^T A^T A V_1 \Sigma_r^{-1} = I$ , 即  $U_1$  列正交. 把  $U_1$  补成正交矩阵  $U = [U_1 \ U_2]$  (这相当于把  $U_1$  的列扩充成一组标准正交基), 则有  $U^T A V = \begin{bmatrix} U_1^T A V_1 & U_1^T A V_2 \\ U_2^T A V_1 & U_2^T A V_2 \end{bmatrix} = \begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}$ .  $\square$

分解  $A = U\Sigma V^T$  称为  $A$  的**奇异值分解**, 简称 **SVD**.

由奇异值分解  $A = U\Sigma V^T$ , 可得  $AV = U\Sigma$ ,  $A^T U = V\Sigma^T$ , 这就说明  $\sigma_i$  是  $A$  的奇异值,  $u_i$  是  $A$  的属于  $\sigma_i$  的左奇异向量,  $v_i$  是  $A$  的属于  $\sigma_i$  的右奇异向量.

**例 6.3.3** 计算  $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \\ 0 & 0 \end{bmatrix}$  的一个奇异值分解.

先计算  $A^T A$  的谱分解:

$$A^T A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 45 & \\ & 5 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T =: V D V^T.$$

奇异值是  $\sigma_1 = 3\sqrt{5}, \sigma_2 = \sqrt{5}$ , 记  $\Sigma_r = \begin{bmatrix} 3\sqrt{5} & \\ & \sqrt{5} \end{bmatrix}$ , 则

$$U_1 = AV \Sigma_r^{-1} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{3\sqrt{5}} & \\ & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ 0 & 0 \end{bmatrix}.$$

将  $U_1$  补成正交矩阵  $U = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} & 0 \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ , 令  $\Sigma = \begin{bmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{bmatrix}$ . 于是  $A$  的一个奇异值分解为  $U\Sigma V^T$ , 其中  $U, \Sigma, V$  如上. ☺

从例 6.3.3 的求解可以看出, 在实际计算奇异值分解时, 只需要计算左右奇异向量中的一组, 就可以由定义导出另一组.

对  $A$  的奇异值分解  $A = U\Sigma V^T$ , 记  $U = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_m]$ ,  $V = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_n]$ , 则

1.  $\mathbf{u}_1, \cdots, \mathbf{u}_r$  是  $\mathcal{R}(A)$  的一组标准正交基;
2.  $\mathbf{u}_{r+1}, \cdots, \mathbf{u}_m$  是  $\mathcal{N}(A^T)$  的一组标准正交基;
3.  $\mathbf{v}_1, \cdots, \mathbf{v}_r$  是  $\mathcal{R}(A^T)$  的一组标准正交基;
4.  $\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n$  是  $\mathcal{N}(A)$  的一组标准正交基.

记  $U_r = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r]$ ,  $V_r = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_r]$ , 则

$$A = U_r \Sigma_r V_r^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T,$$

这称为  $A$  的**简化奇异值分解**.

奇异值分解在理论和实际中都有广泛应用.

**例 6.3.4 (广义逆)** 第 3.3 节中讨论过关于  $A$  的正交投影矩阵  $P_A$ , 若  $\mathcal{R}(A)$  的一组基排成矩阵  $B$ , 则  $P_A = B(B^T B)^{-1} B^T$ . 而矩阵  $U_r$  的列是  $\mathcal{R}(A)$  的一组标准正交基, 因此  $P_A = U_r U_r^T$ .

若记矩阵  $A^+ = V_r \Sigma_r^{-1} U_r^T$ , 则  $AA^+ = U_r U_r^T$  是  $\mathcal{R}(A)$  上的正交投影的表示矩阵,  $A^+A = V_r V_r^T$  是  $\mathcal{R}(A^T)$  上的正交投影的表示矩阵. 于是  $I - AA^+$  是  $\mathcal{N}(A^T)$  上的正交投影的表示矩阵,  $I - A^+A$  是  $\mathcal{N}(A)$  上的正交投影的表示矩阵. 矩阵  $A^+$  称为  $A$  的 **Moore-Penrose 广义逆**, 简称**广义逆**. 矩阵的广义逆唯一, 证明留给读者.

广义逆在解方程组时很有用.

当  $A$  可逆时, 方程组有唯一解  $A^{-1}\mathbf{b}$ . 此时,  $A^+$  就是  $A$  的逆  $A^{-1}$ .

当  $A\mathbf{x} = \mathbf{b}$  有不只一个解时,  $A$  不可逆, 但注意到  $\mathbf{b} \in \mathcal{R}(A)$ , 就有  $AA^+\mathbf{b} = \mathbf{b}$ , 即  $A^+\mathbf{b}$  是方程组的一个解. 不仅如此,  $A^+\mathbf{b}$  还是方程组所有解中长度最小的解. 方程组的解集是  $\{A^+\mathbf{b} + \mathbf{x}_0 : \mathbf{x}_0 \in \mathcal{N}(A)\}$ , 而  $\mathbf{x}_0^T A^+\mathbf{b} = ((I - A^+A)\mathbf{x}_0)^T A^+\mathbf{b} = \mathbf{x}_0^T (I - A^+A)A^+\mathbf{b} = 0$ , 因此  $\|A^+\mathbf{b} + \mathbf{x}_0\|^2 = \|A^+\mathbf{b}\|^2 + \|\mathbf{x}_0\|^2 \geq \|A^+\mathbf{b}\|^2$ , 这就说明它是长度最小的解.

再考虑方程组无解的情形, 即最小二乘问题. 由于  $AA^+\mathbf{b}$  是  $\mathbf{b}$  在  $\mathcal{R}(A)$  上的正交投影, 因此  $A^T\mathbf{b}$  是一个最小二乘解. 事实上,  $A^+\mathbf{b}$  也是  $A\mathbf{x} = \mathbf{b}$  的最小二乘解中长度最小的解 (为什么?). ☺



下面来说明奇异值是良态的, 即如果矩阵变化不大, 那么奇异值变化也不大.

**例 6.3.5** 考察 Jordan 块  $J_n = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}$ , 只有一个  $n$  重特征值 0; 有两个奇异值, 是  $n-1$  重的 1 和 1 重的 0.

给  $J_n(0)$  添加一点微小变化  $\tilde{J}_n = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 10^{-n} & & & 0 \end{bmatrix}$ , 有  $n$  个单特征值  $10^{-1}e^{2\pi i \frac{j}{n}}$ ,

$j = 1, \dots, n$ ; 有两个奇异值, 是  $n-1$  重的 1 和 1 重的  $10^{-n}$ .

可以看到, 矩阵添加了一点微小变化, 特征值的变化比矩阵的变化要大的多, 而奇异值的变化则和矩阵的变化相差不大.  $\odot$

对任意矩阵, 要如何描述其变化呢? 当矩阵从  $A$  变成  $\tilde{A}$ , 当然最好能找到一个  $\tilde{A}-A$  的函数来刻画. 我们知道,  $m \times n$  矩阵实质是  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射, 因此只要能描述变化前后任意向量的像的变化, 就能描述矩阵的变化.

### 6.3.2 矩阵范数和低秩逼近

为了把所谓矩阵的变化更好地表述出来, 我们引入如下定义.

**定义 6.3.6 (矩阵的谱范数)** 对任意矩阵  $A$ , 非负数  $\max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$  称为矩阵  $A$  的谱范数, 记为  $\|A\|$ .

于是就可用  $\|\tilde{A} - A\|$  来表述矩阵的变化.

从例 6.3.5 中可以看到, 奇异值的变化总是不超过矩阵的变化, 即对任意奇异值  $\sigma_i$ ,  $|\tilde{\sigma}_i - \sigma_i| \leq \|\tilde{J}_n - J_n\|$ ; 而特征值变化则不满足此点, 即对任意特征值  $\lambda_i$ , 不能得到  $|\tilde{\lambda}_i - \lambda_i| \leq \|\tilde{J}_n - J_n\|$ . 观察可得  $|\tilde{\lambda}_i - \lambda_i| \leq \|\tilde{J}_n - J_n\|^{\frac{1}{n}}$ .

如上观察对任意矩阵都成立吗? 事实上, 对任意矩阵  $A$  和  $\tilde{A}$ ,

$$|\tilde{\sigma}_i - \sigma_i| \leq \|\tilde{A} - A\|, \quad |\tilde{\lambda}_i - \lambda_i| \leq C_A \|\tilde{A} - A\|^{\frac{1}{n}},$$

其中  $C_A$  是一个与  $A, \tilde{A}$  有关的常数<sup>1</sup>. 这就说明, 奇异值是良态的, 而特征值不是良态的.

下面给出谱范数的一些基本性质.

<sup>1</sup>证明这一结论需要较深入的矩阵论知识, 这里不会展开, 感兴趣的读者请查阅矩阵分析的教材或专著.

**命题 6.3.7** 矩阵的谱范数满足:

1.  $\|A\| \geq 0$ , 且  $\|A\| = 0$  当且仅当  $A = O$ ;
2.  $\|kA\| = |k|\|A\|$ ;
3.  $\|A + B\| \leq \|A\| + \|B\|$ ;
4.  $\|AB\| \leq \|A\|\|B\|$ ;
5. 如果  $U, V$  正交, 则  $\|UAV^T\| = \|A\|$ .

证明留给读者.

**命题 6.3.8** 对任意矩阵  $A$ , 矩阵的谱范数  $\|A\|$  等于  $A$  的最大奇异值.

证. 利用命题 6.2.8,

$$\|A\|^2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T A^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_{\max}(A^T A) = \sigma_{\max}(A)^2,$$

其中  $\lambda_{\max}(A^T A)$  是  $A^T A$  的最大特征值, 亦即  $A$  的最大奇异值  $\sigma_{\max}(A)$  的平方.  $\square$

下面来分析矩阵逼近问题. 所谓矩阵逼近问题, 就是给定一个矩阵和一个矩阵集合, 试图在该集合中寻找一个和给定矩阵最接近的矩阵.

通常会设定给定集合是满足某种条件的矩阵, 然后用逼近矩阵来代替原矩阵, 参与其他计算, 如此往往会带来一些便利.

**定理 6.3.9 (低秩逼近)** 设  $m \times n$  矩阵  $A$  的简化奇异值分解为  $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$ . 则对任意正整数  $k < r$ ,  $A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$  是  $A$  的最佳秩  $k$  逼近, 即对任意秩不超过  $k$  的  $m \times n$  矩阵  $B$ , 都有  $\|A - A_k\| \leq \|A - B\|$ , 亦即

$$\|A - A_k\| = \min_{\text{rank}(B) \leq k} \|A - B\|.$$

证. 根据命题 6.3.8,  $\|A - A_k\| = \|\sigma_{k+1} \mathbf{u}_{k+1} \mathbf{v}_{k+1}^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T\| = \sigma_{k+1}$ . 只需再证  $\|A - B\| \geq \sigma_{k+1}$  对任意秩不超过  $k$  的矩阵  $B$  都成立.

由  $\mathcal{N}(B) = \mathcal{R}(B^T)^\perp$ , 可得  $\dim \mathcal{N}(B) = n - \dim \mathcal{R}(B^T) \geq n - k$ . 记  $V_{k+1} = [\mathbf{v}_1 \cdots \mathbf{v}_{k+1}]$ , 则存在非零向量  $\mathbf{x} \in \mathcal{N}(B) \cap \mathcal{R}(V_{k+1})$  且  $\|\mathbf{x}\| = 1$ . 令  $\mathbf{x} = V_{k+1} \hat{\mathbf{x}}$ . 记  $A = U \Sigma V^T = U_r \Sigma_r V_r^T$  为  $A$  的奇异值分解和简化奇异值分解. 则

$$\begin{aligned} \|A - B\| &\geq \|(A - B)\mathbf{x}\| = \|A\mathbf{x}\| = \|U \Sigma V^T V_{k+1} \hat{\mathbf{x}}\| \\ &= \left\| \begin{bmatrix} \Sigma_k & \\ & \sigma_{k+1} \\ & & * \end{bmatrix} \begin{bmatrix} I_k \\ 1 \\ \mathbf{0} \end{bmatrix} \hat{\mathbf{x}} \right\| = \left\| \begin{bmatrix} \Sigma_k & \\ & \sigma_{k+1} \end{bmatrix} \begin{bmatrix} I_k \\ 1 \end{bmatrix} \hat{\mathbf{x}} \right\| \geq \sigma_{k+1}. \quad \square \end{aligned}$$

定理 6.3.9 中的  $A_k$  常称为  $A$  的截断 SVD.

下面我们来看低秩逼近的实际应用.

**例 6.3.10 (图像处理)** 计算机在存储图像时, 往往是在照片上打上均匀的正方形网格, 并只存储网格点的信息, 这些网格点称为像素. 为简单起见, 这里只考虑灰度图像. 灰度图像用 0 表示黑色, 1 表示白色, 二者之间的数字表示不同深浅的灰色. 那么所有像素的灰度值按位置排列就得到一个矩阵, 换言之, 我们就可以用矩阵来表示具有相同像素数的图像.

来源于现实生活的图像, 产生的矩阵一般都低秩. 可以如此理解: 理想情形下, 来源于现实生活的图像总是可以划分为若干区域, 每个区域内部的灰度、颜色等都比较单纯或者变化比较平滑, 而区域之间则有很大的不同, 从而形成了图像上明显的分界线. 每个区域由于变化比较平滑, 因而对应的矩阵低秩, 例如, 区域上的数据按照格点坐标的线性函数来变化, 则其秩应当不超过 2. 而图像相当于这些区域的总和, 因此图片对应的矩阵的秩不会超过这些区域对应的矩阵的秩的和, 于是也低秩.

因此, 可以合理地对图像对应的矩阵做低秩逼近, 这相当于舍去了少量误差 (图像处理中称之为噪声).

现在考虑对图像进行压缩, 即适当减少图像占用的存储. 原来一幅  $m \times n$  像素的图像, 对应产生  $m \times n$  矩阵  $A$ , 需要存储  $mn$  个元素. 如果对  $A$  做秩  $k$  逼近, 设  $A$  的截断 SVD 为  $A_k = U_k \Sigma_k V_k^T$ , 则只需存储  $mk + k + kn = k(m + n + 1)$  个元素. 当  $m, n$  很大时, 后者所需存储显然会远小于前者. 图 6.3.1 是一幅  $256 \times 256$  像素的图像的原图以及对相应矩阵分别做秩 1, 2, 4, 8, 16, 32, 64, 127 逼近得到的图像. 八幅压缩后的图像的压缩比, 即所占存储与原图所占存储之比  $\frac{k(m+n+1)}{mn}$ , 分别约为 0.78%, 1.57%, 3.13%, 6.26%, 12.52%, 25.05%, 50.10%, 99.41%. 可见, 适当选择  $k$ , 就能在图像细节和压缩比之间取得平衡. ☺

### 6.3.3 主成分分析

最后考虑一个统计模型. 假设我们测量了某类对象的多个具体属性的值, 并希望从中找到起决定性作用的某些属性的组合. 例如体检中测量了一系列生化指标, 我们希望知道何种指标的数值的组合就能够标识某种疾病, 例如糖尿病. 最简单的办法是测量大量糖尿病人的生化指标, 从中发现特殊的指标数值的组合. 如果确实能够发现这种组合, 那么只要某人的生化指标出现了这种组合, 就可以合理怀疑此人可能患有糖尿病.

现在把上述思考数学化. 假设已测得  $d$  维空间上的随机变量的  $n$  个样本  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , 这些样本往往聚集在如直线、平面等低维集合附近. 不妨假设这些样本在任意方向上的均值都是 0, 不然则通过平移数据来获得. 那么这个低维集合可以认为是一个低维子空

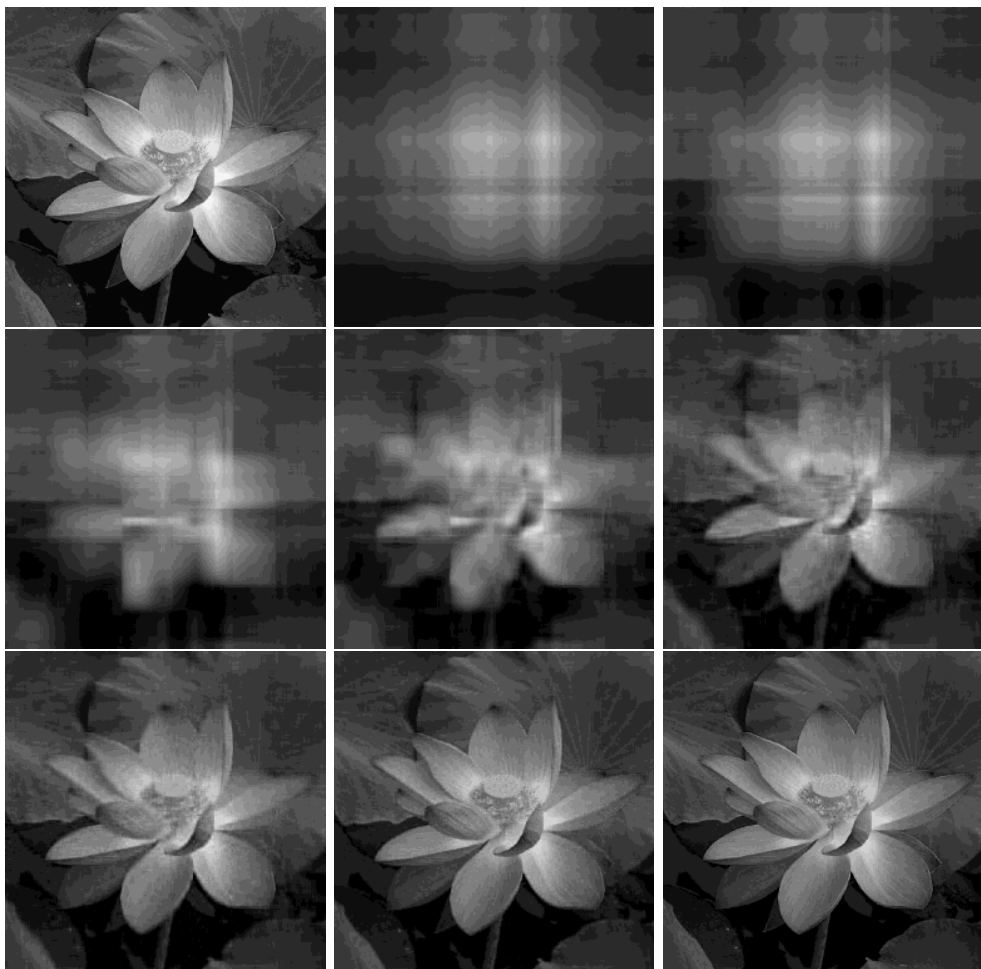


图 6.3.1: 荷花

间. 只要能得到这个低维子空间的一组基, 就能近似地得到这个随机变量. 这组基就称为随机变量的**主成分**, 这种模型称为**主成分分析**.

那么, 如何计算这个子空间的一组基? 记该低维子空间为  $\mathcal{M}$ , 则每个数据点  $\mathbf{a}_i$  到  $\mathcal{M}$  的正交投影是  $\mathbf{P}_{\mathcal{M}}\mathbf{a}_i$ . 而所有数据点到  $\mathcal{M}$  的距离的平方和, 就是  $\sum_{i=1}^n \|\mathbf{a}_i - \mathbf{P}_{\mathcal{M}}\mathbf{a}_i\|^2$ . 记  $\mathcal{M}$  的一组标准正交基组成的矩阵为  $\mathbf{Q}$ , 又记数据  $\mathbf{a}_i$  作为列组成的矩阵为  $\mathbf{A}$ , 则这个平方和是

$$\begin{aligned} \sum_{i=1}^n \|\mathbf{a}_i - \mathbf{P}_{\mathcal{M}}\mathbf{a}_i\|^2 &= \sum_{i=1}^n \|(I - \mathbf{Q}\mathbf{Q}^T)\mathbf{a}_i\|^2 \\ &= \sum_{i=1}^n \mathbf{a}_i^T (I - \mathbf{Q}\mathbf{Q}^T) \mathbf{a}_i \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \text{trace}((I - QQ^T)\mathbf{a}_i\mathbf{a}_i^T) \\
&= \text{trace}\left((I - QQ^T)\sum_{i=1}^n \mathbf{a}_i\mathbf{a}_i^T\right) \\
&= \text{trace}((I - QQ^T)AA^T) \\
&= \text{trace}(AA^T) - \text{trace}(Q^TAA^TQ).
\end{aligned}$$

如果我们希望这个主成分空间越准确越好,则需要找一个列正交矩阵  $Q$ , 使得此平方和最小, 亦即使得  $\text{trace}(Q^TAA^TQ)$  最大.

在  $\mathcal{M}$  的维数  $m = 1$  时, 最大值就是  $AA^T$  的最大特征值, 亦即  $A$  的最大奇异值的平方, 而主成分就是  $A$  的属于最大奇异值的左奇异向量. 当  $\mathcal{M}$  的维数  $m > 1$  时, 最大值就是  $AA^T$  的最大  $m$  个特征值之和, 而主成分就是  $A$  的属于最大  $m$  个奇异值的左奇异向量, 这可由如下定理 6.3.11 得到.

**定理 6.3.11** 设  $m \times n$  矩阵  $A$  的简化奇异值分解为  $A = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T$ . 则对任意正整数  $k < r$ ,  $U_k = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_k]$  满足

$$\text{trace}(U_k^TAA^TU_k) = \max_{\substack{Q \text{ 列正交} \\ \text{rank}(Q) \leq k}} \text{trace}(Q^TAA^TQ).$$

证. 只需证明  $\text{trace}(U_k^TAA^TU_k) = \max_{\substack{Q \text{ 列正交} \\ \text{rank}(Q) = k}} \text{trace}(Q^TAA^TQ)$ , 结合  $\text{trace}(U_1^TAA^TU_1) \leq \cdots \leq \text{trace}(U_k^TAA^TU_k)$ , 可得结论.

利用数学归纳法. 命题 6.2.8 直接得到  $k = 1$  的情形. 现假设该命题对  $k-1$  情形成立. 设  $W = [W_0 \ \mathbf{w}_k]$  是一个最大解, 即  $\text{trace}(W^TAA^TW) = \max_{\substack{Q \text{ 列正交} \\ \text{rank}(Q) = k}} \text{trace}(Q^TAA^TQ)$ .

不妨假设  $U_{k-1}^T\mathbf{w}_k = 0$ .

这是因为, 对任意正交矩阵  $T$ ,  $\text{trace}(T^TW^TAA^TWT) = \text{trace}(W^TAA^TW)$ , 故  $W$  可选其列空间  $\mathcal{R}(W)$  的任意一组基; 因此, 取  $\mathbf{u}_i$  在  $\mathcal{R}(W)$  上的正交投影的单位化向量, 不难得到与这些向量都正交的  $\mathbf{w}_k$ , 该向量就满足  $U_{k-1}^T\mathbf{w}_k = 0$  (为什么?).

现在  $U_{k-1}^T\mathbf{w}_k = 0$ . 根据归纳假设,  $\text{trace}(W_0^TAA^TW_0) \leq \text{trace}(U_{k-1}^TAA^TU_{k-1})$ . 根据命题 6.2.8,  $\mathbf{w}_k^TAA^T\mathbf{w}_k \leq \mathbf{u}_k^TAA^T\mathbf{u}_k$ . 因此,

$$\begin{aligned}
\text{trace}(W^TAA^TW) &= \text{trace}(W_0^TAA^TW_0) + \mathbf{w}_k^TAA^T\mathbf{w}_k \\
&\leq \text{trace}(U_{k-1}^TAA^TU_{k-1}) + \mathbf{u}_k^TAA^T\mathbf{u}_k = \text{trace}(U_k^TAA^TU_k).
\end{aligned}$$

这就说明  $\text{trace}(U_k^TAA^TU_k) = \max_{\substack{Q \text{ 列正交} \\ \text{rank}(Q) = k}} \text{trace}(Q^TAA^TQ)$ . □

由定理 6.3.11 可见, 主成分分析本质上也是低秩逼近.

为了更简洁地描述上述过程, 我们引入如下定义.

**定义 6.3.12 (Frobenius 范数)** 对任意矩阵  $A = [a_{ij}]_{m \times n}$ , 非负数  $\sqrt{\text{trace}(A^T A)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$  称为矩阵  $A$  的 **Frobenius 范数**, 记为  $\|A\|_F$ .

为区分起见, 定义 6.3.6 中的范数又称为**谱范数**.

下面给出 Frobenius 范数的一些基本性质.

**命题 6.3.13** 矩阵的 Frobenius 范数满足:

1.  $\|A\|_F \geq 0$ , 且  $\|A\|_F = 0$  当且仅当  $A = O$ ;
2.  $\|kA\|_F = |k|\|A\|_F$ ;
3.  $\|A + B\|_F \leq \|A\|_F + \|B\|_F$ ;
4.  $\|AB\|_F \leq \|A\|\|B\|_F, \|AB\|_F \leq \|A\|_F\|B\|$ ;
5. 如果  $U, V$  正交, 则  $\|UAV^T\|_F = \|A\|_F$ .

**命题 6.3.14** 对任意矩阵  $A$ , 其 Frobenius 范数平方  $\|A\|_F^2$  等于  $A$  的所有奇异值的平方和. 因此,  $\|A\|_F \geq \|A\|$ .

证明留给读者. 注意命题 6.3.13 第 4 条不等式右端两个范数不同, 不等式  $\|AB\|_F \leq \|A\|_F\|B\|_F$  当然也成立, 但不如它精细.

**定理 6.3.15 (低秩逼近)** 设  $m \times n$  矩阵  $A$  的简化奇异值分解为  $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$ . 则对任意正整数  $k < r$ ,  $A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$  是  $A$  在 Frobenius 范数意义下的最佳秩  $k$  逼近, 即对任意秩不超过  $k$  的  $m \times n$  矩阵  $B$ , 都有  $\|A - A_k\|_F \leq \|A - B\|_F$ , 亦即

$$\|A - A_k\|_F = \min_{\text{rank}(B) \leq k} \|A - B\|_F.$$

证. 设  $B_0$  是满足  $\|A - B_0\|_F = \min_{\text{rank}(B) \leq k} \|A - B\|_F$  的秩为  $k' \leq k$  的  $m \times n$  矩阵, 而  $B_0 = QR$  是其简化 QR 分解. 则  $QQ^T B_0 = B_0$ , 因此

$$\|A - B_0\|_F^2 = \|A - QQ^T B_0\|_F^2 = \|A - QQ^T A\|_F^2 + \|QQ^T A - QQ^T B_0\|_F^2 \geq \|A - QQ^T A\|_F^2.$$

但  $\|A - QQ^T A\|_F^2 = \|A\|_F^2 - \|Q^T A\|_F^2 = \|A\|_F^2 - \text{trace}(Q^T A A^T Q)$ . 记  $A_k = U_k \Sigma_k V_k^T$ , 根据定理 6.3.11,  $\text{trace}(Q^T A A^T Q) \leq \text{trace}(U_k^T A A^T U_k) = \sum_{j=1}^{k'} \sigma_j^2$ . 因此

$$\|A - QQ^T A\|_F^2 \geq \sum_{j=1}^r \sigma_j^2 - \sum_{j=1}^{k'} \sigma_j^2 = \sum_{j=k'+1}^r \sigma_j^2 \geq \sum_{j=k+1}^r \sigma_j^2 = \|A - A_k\|_F^2.$$

由此立得. □

定理 6.3.15 给出了截断 SVD 也是低秩逼近的最优解的另一描述.

## 习题

练习 6.3.1 求下列矩阵的奇异值分解.

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

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

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

4.  $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ .

5.  $\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix}$ .

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

7.  $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ .

8.  $\begin{bmatrix} A & O \\ O & O \end{bmatrix}$ , 其中  $A$  有奇异值分解  $A = U\Sigma V^T$ .

练习 6.3.2 矩阵  $A$  的 QR 分解  $A = QR$ , 且  $R$  有奇异值分解  $R = U\Sigma V^T$ , 求  $A$  的奇异值分解.

练习 6.3.3 设  $A$  的奇异值分解为  $A = U\Sigma V^T$ , 求矩阵  $\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}$  的谱分解.

练习 6.3.4 设矩阵  $A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ , 考虑单位圆  $C = \{\mathbf{v} \in \mathbb{R}^2 \mid \|\mathbf{v}\| = 1\}$  及其在  $A$  对应的线性变换  $A$  下的像  $A(C) = \{A\mathbf{v} \in \mathbb{R}^2 \mid \|\mathbf{v}\| = 1\}$ .

1. 设  $\mathbf{w} \in A(C)$ , 证明  $\mathbf{w}^T(AA^T)^{-1}\mathbf{w} = 1$ .
2. 求  $A$  的奇异值分解  $A = U\Sigma V^T$ .
3. 注意  $V, U$  为二阶正交矩阵, 对应的线性变换是旋转或反射, 而  $\Sigma$  是对角矩阵, 对应伸缩变换. 从几何上看, 曲线  $V^T(C), \Sigma V^T(C), U\Sigma V^T(C)$  分别是什么形状?

练习 6.3.5 设矩阵  $A$  的奇异值分解是  $A = U\Sigma V^T$ .

1. 证明  $AA^T = U(\Sigma\Sigma^T)U^T, A^TA = V(\Sigma^T\Sigma)V^T$  分别是这两个对称矩阵的谱分解, 并得到  $AA^T$  和  $A^TA$  的非零特征值相同.
2. 对任意  $A$  的奇异值  $\sigma \neq 0$ , 设  $\mathbf{v}$  和  $\mathbf{w}$  分别是  $A^TA$  和  $AA^T$  的属于  $\sigma^2$  的特征向量, 证明  $A\mathbf{v}$  和  $A^T\mathbf{w}$  分别是  $AA^T$  和  $A^TA$  的属于  $\sigma^2$  的特征向量.

练习 6.3.6 1. 将  $\frac{3x^2+2xy+3y^2}{x^2+y^2}$  表示成对称矩阵  $S$  的 Rayleigh 商, 并通过 Rayleigh 商求表达式的最大值和最小值.

2. 将  $\frac{(x+4y)^2}{x^2+y^2}$  表示成矩阵  $S$  的 Rayleigh 商, 并通过 Rayleigh 商求表达式的最大值和最小值.

练习 6.3.7 证明命题 6.3.7.

练习 6.3.8 证明命题 6.3.13.

**练习 6.3.9** 证明命题 6.3.14.

**练习 6.3.10 (樊畿迹定理)** 对任意  $n$  阶对称矩阵  $A \in \mathbb{R}^{n \times n}$ , 假设特征值为  $\lambda_1 \geq \dots \geq \lambda_n$ , 对应特征向量为  $\mathbf{u}_1, \dots, \mathbf{u}_n$ , 则  $\max_{\substack{n \times m \text{ 矩阵 } Q: \\ Q^T Q = I}} \text{trace}(Q^T A Q) = \sum_{i=1}^m \lambda_i$ , 且  $Q = [\mathbf{u}_1 \ \dots \ \mathbf{u}_m]$  时取得最大值.

**练习 6.3.11 (极分解)** 对  $n$  阶方阵  $A$ , 存在正交矩阵  $Q$  和对称半正定矩阵  $S$ , 使得  $A = QS$ .

分解式  $A = QS$  称为  $A$  的**极分解**. 容易看到,  $A = S_1 Q_1$ , 即方阵分解为对称半正定矩阵和正交矩阵的乘积, 也存在.

**练习 6.3.12** 证明矩阵的广义逆唯一.

**练习 6.3.13** 证明矩阵任意特征值的绝对值不大于其最大的奇异值.

**练习 6.3.14** 证明或者举出反例.

1.  $n$  阶方阵  $A$  为正交矩阵当且仅当它有  $n$  个奇异值, 值都是 1.
2. 如果  $n$  阶方阵有  $n$  个奇异值, 则所有奇异值的乘积等于所有特征值的乘积.
3. 假设  $n$  阶方阵  $A$  的 SVD 分解为  $A = U \Sigma V^T$ , 而  $A + I_n$  的 SVD 分解为  $A + I_n = U(\Sigma + I_n)V^T$ . 证明  $A$  是对称矩阵.
4. 如果  $n$  阶方阵  $A$  的  $n$  个奇异值就是它的  $n$  个特征值, 则  $A$  是对称矩阵.

**练习 6.3.15 (低秩逼近与数据拟合)** 考虑平面上的点  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \end{bmatrix}, \begin{bmatrix} 4 \\ 20 \end{bmatrix}$ . 现在想找到一条直线, 使得点到直线距离的平方和最小. 令  $A = \begin{bmatrix} 0 & 1 & 3 & 4 \\ 0 & 8 & 8 & 20 \end{bmatrix}$ .

1. 找到函数  $f(a, b, c)$ , 使得  $f(a, b, c)$  就是每个点到直线  $ax + by + c = 0$  的距离的平方和.
2. 假设  $a, b$  已知, 用导数证明此时最好的  $c$  是  $-[a \ b] \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ , 这里  $x_0, y_0$  代表四个点的  $x$  坐标平均值和  $y$  坐标平均值. 提示: 这意味着欲求直线应为  $a(x - x_0) + b(y - y_0) + c = 0$ .
3. 对  $A$  进行“中心化”, 即对每一行所有元素减去该行的平均值, 得到  $B$ . 此时对  $B$  的列对应的四个点来说, 最佳直线为何应该经过原点?  $A, B$  对应的最佳直线为何一定平行?
4. 计算  $B$  的最佳秩 1 逼近.
5. 计算欲求直线.
6. 思考: 如果考虑的不是平面上的点, 而是  $n$  维空间中的点, 问题如何处理?

**练习 6.3.16** 考虑子空间  $\mathcal{M}, \mathcal{N}$ , 其对应的正交投影矩阵为  $P, Q$ . 我们想要研究矩阵  $H = P(P + Q)^+ Q + Q(P + Q)^+ P$ .

1. 计算  $(P + Q)(P + Q)^+$  的列空间和零空间, 该矩阵是否是一个正交投影矩阵?
2. 计算  $(P + Q)^+(P + Q)$  的列空间和零空间, 该矩阵是否是一个正交投影矩阵? 和前一矩阵有何关联?
3. 证明  $Q(P + Q)^+(P + Q) = Q$ ,  $(P + Q)(P + Q)^+ Q = Q$ .



4. 证明  $H = 2P(P+Q)^+Q = 2Q(P+Q)^+P$ .
5. 假设  $T$  是  $\mathcal{M} \cap \mathcal{N}$  上的正交投影矩阵, 证明  $HP = HQ = HT = H$ .
6. 证明  $HT = T$ .

注意  $PT = QT = T$ , 再利用第 1 条, 可得  $H = HT = T$ , 由此即得  $\mathcal{M} \cap \mathcal{N}$  的正交投影矩阵的表达式.

Liang, Tian & Yang

# 第七章 线性空间和线性映射

第 1 至 6 章主要研究了线性空间  $\mathbb{R}^n$  中的向量、子空间以及其上的线性变换. 本章和第 8 章将把之前的概念推广到一个很大的范围, 从而使得线性代数不再仅仅局限于向量和矩阵, 使得线性代数被不同领域的工作者广泛运用在实际生产生活中.

## 7.1 线性空间

求解线性方程组的计算过程只使用了加减乘除这四种四则运算. 有理数四则运算(除 0 除外)的结果还是有理数, 而自然数或整数却不是这样. 为此引入一个新的概念.

**定义 7.1.1 (数域)** 给定  $\mathbb{C}$  的子集  $\mathbb{F}$ , 如果  $\mathbb{F}$  中至少包含一个非零复数, 且  $\mathbb{F}$  对复数的加减乘除四则运算封闭, 即对任意  $a, b \in \mathbb{F}$ , 都有  $a + b, a - b, ab \in \mathbb{F}$ , 且当  $b \neq 0$  时  $\frac{a}{b} \in \mathbb{F}$ , 则称  $\mathbb{F}$  是一个**数域**.

可以验证,  $\mathbb{Q}, \mathbb{R}, \mathbb{C}$  都是数域, 而  $\mathbb{N}, \mathbb{Z}$  不是数域. 数域  $\mathbb{F}$  上的线性方程组的解(如果存在)也在数域  $\mathbb{F}$  上.

我们在第 1 章中为向量和矩阵定义了加法和数乘两种线性运算, 并发现二者的线性运算满足相同的八条运算法则. 因此, 向量和矩阵在某种意义上相同, 因为二者定义的运算及其运算法则相同.

一般地, 我们作如下定义.

**定义 7.1.2 (线性空间)** 给定非空集合  $\mathcal{V}$  和数域  $\mathbb{F}$ , 如果  $\mathcal{V}$  上定义了**加法运算**  $+: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ ,  $\mathcal{V}$  的元素和  $\mathbb{F}$  中的数定义了**数乘运算**  $\cdot: \mathbb{F} \times \mathcal{V} \rightarrow \mathcal{V}$ , 且这两种运算满足如下八条运算法则:

1. 加法结合律: 对任意  $a, b, c \in \mathcal{V}$ ,  $(a + b) + c = a + (b + c)$ ;
2. 加法交换律: 对任意  $a, b \in \mathcal{V}$ ,  $a + b = b + a$ ;
3. 零元素: 存在元素  $0 \in \mathcal{V}$ , 对任意  $a \in \mathcal{V}$ ,  $a + 0 = a$ , 其中  $0$  称为**零元素**;

4. 负元素: 对任意  $a \in \mathcal{V}$ , 存在  $b \in \mathcal{V}$ , 满足  $a + b = 0$ , 称它为  $a$  的负元素, 记为  $-a$ ;
5. 单位数: 对任意  $a \in \mathcal{V}$ ,  $1a = a$ ;
6. 数乘结合律: 对任意  $a \in \mathcal{V}, k, l \in \mathbb{F}$ ,  $(kl)a = k(la)$ ;
7. 数乘对数的分配律: 对任意  $a \in \mathcal{V}, k, l \in \mathbb{F}$ ,  $(k + l)a = ka + la$ ;
8. 数乘对向量的分配律: 对任意  $a, b \in \mathcal{V}, k \in \mathbb{F}$ ,  $k(a + b) = ka + kb$ ;

则称  $\mathcal{V}$  是  $\mathbb{F}$  上的向量空间或线性空间, 其中的元素可以称为向量, 零元素和负元素可以称为零向量和负向量.

减法可以自然地定义:  $a - b = a + (-b)$ .

在定义 7.1.2 下, 线性空间中的向量不再局限于  $\mathbb{R}^m$  中的向量, 而是满足运算法则的抽象对象. 因此一些基本性质不再不证自明.

**命题 7.1.3** 在线性空间中,

1. 零向量唯一;
2. 任意向量的负向量唯一;
3. 加法消去律, 即由  $a + b = a + c$  可以推出  $b = c$ ;
4. 可以移项, 即由  $a + b = c$  可以推出  $a = c - b$ ;
5.  $0a = 0, (-1)a = -a, k0 = 0$ ;
6. 可以约系数, 即由  $ka = b, k \neq 0$  可以推出  $a = \frac{1}{k}b$ .

证. 第 1 条: 设有两个零向量  $0, \tilde{0}$ , 则  $0 = 0 + \tilde{0} = \tilde{0} + 0 = \tilde{0}$ .

第 2 条: 设向量  $a$  有两个负向量  $b, c$ , 则  $b = b + 0 = b + (a + c) = (b + a) + c = 0 + c = c$ .

第 3 条:  $b = 0 + b = (-a + a) + b = -a + (a + b) = -a + (a + c) = (-a + a) + c = c$ .

第 4 条:  $a = a + 0 = a + (b - b) = (a + b) - b = c - b$ .

第 5 条:  $0a + 0 = 0a = (0 + 0)a = 0a + 0a$ , 由消去律得  $0a = 0$ ;  $(-1)a + a = (-1 + 1)a = 0 = -a + a$ , 由消去律得  $(-1)a = -a$ ; 对任意  $a$ ,  $k0 + ka = k(0 + a) = ka$ , 因此  $k0$  也是一个零向量, 由零向量唯一得  $k0 = 0$ .

第 6 条:  $a = (\frac{1}{k}k)a = \frac{1}{k}(ka) = \frac{1}{k}b$ . □

在考察一个集合是否线性空间时, 除了运算及运算法则外, 还要注意两个隐含条件:

1. 集合非空;
2. 加法和数乘封闭.

下面列举一些常见的线性空间.

**例 7.1.4** 1. 只有一个零元素的集合构成数域  $\mathbb{F}$  上的线性空间, 记为  $\{0\}$ .

2. 数域的扩张: 实数集  $\mathbb{R}$  是有理数集  $\mathbb{Q}$  上的线性空间, 加法和数乘运算就是实数的运算. 注意, 这与  $\mathbb{R}$  作为  $\mathbb{R}$  上的线性空间不同.
3. 几何空间: 考虑三维几何空间中的所有向量 (即有向线段), 在向量的加法和数乘下构成  $\mathbb{R}$  上的线性空间. 注意, 未设定坐标系前, 该线性空间与  $\mathbb{R}^3$  **并不相同**. ☺

**例 7.1.5** 1. 坐标向量空间及其子集:

- (a)  $m$  维向量的全体  $\mathbb{F}^m$ , 加法和数乘运算由定义 1.2.4 给出;
- (b)  $\mathbb{F}$  中数组成的矩阵诱导的坐标向量空间的子集, 如列空间  $\mathcal{R}(A)$ 、零空间  $\mathcal{N}(A)$ .

2. 矩阵空间及其子集:

- (a)  $m \times n$  矩阵的全体, 记为  $\mathbb{F}^{m \times n}$ , 加法和数乘运算由定义 1.5.3 给出;
- (b) 矩阵空间的子集:  $n$  阶上 (下) 三角矩阵的全体;  $n$  阶对角矩阵的全体;  $n$  阶 (反) 对称矩阵的全体.

3. 函数空间及其子集:

- (a) 定义域为  $D$  的实值函数  $f: D \rightarrow \mathbb{R}$  的全体构成  $\mathbb{R}$  上的线性空间, 其中加法是函数的加法, 数乘是常数和函数的乘法, 称为函数空间.
- (b) 定义域相同的实值连续函数的全体也构成  $\mathbb{R}$  上的线性空间, 称为连续函数空间, 记为  $C(D)$ ;
- (c) 定义域相同的实值无穷次可导函数的全体也构成  $\mathbb{R}$  上的线性空间, 称为光滑函数空间, 记为  $C^\infty(D)$ ;
- (d) 实系数多项式的全体也构成  $\mathbb{R}$  上的线性空间, 称为多项式空间, 记为  $\mathbb{R}[x]$ ;
- (e) 次数小于  $n$  的实系数多项式的全体添上零多项式也构成  $\mathbb{R}$  上的线性空间, 记为  $\mathbb{R}[x]_n$ . ☺

这三类线性空间是我们主要研究的对象. 可以看到, 构造新的线性空间的一个常用方法是考察已知线性空间的子集. 满足一定条件时, 子集也可以构成线性空间.

**定义 7.1.6 (子空间)** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{V}$  及其非空子集  $\mathcal{M}$ . 如果  $\mathcal{M}$  关于  $\mathcal{V}$  上的加法和数乘也构成线性空间, 则称  $\mathcal{M}$  是  $\mathcal{V}$  的子空间.

**命题 7.1.7** 线性空间  $\mathcal{V}$  的非空子集  $\mathcal{M}$  是一个子空间, 当且仅当它对加法和数乘封闭.

证. “ $\Rightarrow$ ”: 显然. “ $\Leftarrow$ ”: 八条运算法则中只需验证零向量和负向量的存在性. 由于  $\mathcal{M}$  非空, 则存在  $\mathbf{a} \in \mathcal{M}$ . 根据数乘的封闭性,  $\mathbf{0} = 0\mathbf{a} \in \mathcal{M}$ ,  $-\mathbf{a} = (-1)\mathbf{a} \in \mathcal{M}$ .  $\square$

由命题 7.1.7 可知, 子空间的定义 7.1.6 与  $\mathbb{R}^n$  的子空间的定义 2.1.3 一致.

**例 7.1.8** 回顾例 7.1.5 中的子空间.

1. 对任意  $A \in \mathbb{F}^{m \times n}$ ,  $\mathcal{R}(A)$  是  $\mathbb{F}^m$  的子空间,  $\mathcal{N}(A)$  是  $\mathbb{F}^n$  的子空间.

2. 矩阵空间的子空间的包含链:

$$\{n \text{ 阶对角矩阵的全体} \} \subseteq \{n \text{ 阶上(下)三角矩阵的全体} \} \subseteq \{n \text{ 阶方阵的全体} \};$$

$$\{n \text{ 阶(反)对称矩阵的全体} \} \subseteq \{n \text{ 阶方阵的全体} \}.$$

3. 函数空间的子空间的包含链:

$$\text{多项式空间} \subseteq \text{光滑函数空间} \subseteq \text{连续函数空间} \subseteq \text{函数空间}.$$

☺

**例 7.1.9** 1.  $n$  阶可逆矩阵的全体不是线性空间, 更不是  $\mathbb{F}^{n \times n}$  的子空间, 因为它对矩阵的加法和数乘运算不封闭.

2. 次数等于  $n-1$  的实系数多项式的全体不是线性空间, 更不是  $\mathbb{R}[x]_n$  的子空间, 因为它对多项式的加法运算不封闭. 例如,  $(x^{n-1} + x^{n-2}) + (-x^{n-1}) = x^{n-2}$ , 其次数不是  $n-1$ .  $\odot$

下面来考虑子空间的运算. 子空间作为线性空间的子集, 自然有子集的运算, 如交、并等.

**定义 7.1.10 (子空间的交)** 给定线性空间  $\mathcal{V}$  的两个子空间  $\mathcal{M}_1, \mathcal{M}_2$ , 集合  $\mathcal{M}_1 \cap \mathcal{M}_2$  是  $\mathcal{V}$  的子空间, 称为子空间  $\mathcal{M}_1$  与  $\mathcal{M}_2$  的交.

容易发现, 两个子空间的并通常不是一个子空间. 例如, 平面上的两个坐标轴的并集, 显然对向量的加法运算不封闭. 为了使得对线性运算封闭, 可以考虑两个子空间中向量的所有线性组合构成的集合.

**定义 7.1.11 (子空间的和)** 给定线性空间  $\mathcal{V}$  的两个子空间  $\mathcal{M}_1, \mathcal{M}_2$ , 集合

$$\mathcal{M}_1 + \mathcal{M}_2 := \{\mathbf{m} + \mathbf{n} \mid \mathbf{m} \in \mathcal{M}_1, \mathbf{n} \in \mathcal{M}_2\}$$

是  $\mathcal{V}$  的子空间, 称为子空间  $\mathcal{M}_1$  与  $\mathcal{M}_2$  的**和**.

注意, 子空间的交与和两个运算的前提是, 它们都是同一个线性空间的子空间.

**例 7.1.12** 在坐标向量空间  $\mathbb{R}^3$  中, 设  $\mathcal{M}_1, \mathcal{M}_2$  是两个不同的二维子空间. 几何上, 二者是两个不重合的过原点的平面. 此时,  $\mathcal{M}_1 \cap \mathcal{M}_2$  是过原点的一条直线, 是一维子空间; 而  $\mathcal{M}_1 + \mathcal{M}_2$  是整个  $\mathbb{R}^3$ .

设  $\mathcal{L}_1, \mathcal{L}_2$  是两个不同的一维子空间, 即两条不重合的过原点的直线, 则  $\mathcal{L}_1 \cap \mathcal{L}_2$  是原点, 即零维子空间; 而  $\mathcal{L}_1 + \mathcal{L}_2$  是由两条直线生成的平面.

注意,  $\mathcal{M}_1 \cup \mathcal{M}_2, \mathcal{L}_1 \cup \mathcal{L}_2$  都不是子空间. ☺

**例 7.1.13** 矩阵的列空间的和容易描述. 设  $A, B$  分别是  $m \times n$  矩阵和  $m \times p$  矩阵, 则

$$\mathcal{R}(A) + \mathcal{R}(B) = \mathcal{R}(C),$$

其中  $C = \begin{bmatrix} A & B \end{bmatrix}$  是  $m \times (n+p)$  矩阵. 三个列空间都是  $\mathbb{F}^m$  的子空间.

矩阵的零空间的交也容易描述. 设  $A, B$  分别是  $m \times n$  矩阵和  $p \times n$  矩阵, 二者的零空间  $\mathcal{N}(A), \mathcal{N}(B)$  分别是齐次方程组  $A\mathbf{x} = \mathbf{0}, B\mathbf{x} = \mathbf{0}$  的解空间. 因此, 两个解空间的交集就是联立方程组  $D\mathbf{x} = \mathbf{0}$  的解空间, 其中  $D = \begin{bmatrix} A \\ B \end{bmatrix}$  是  $(m+p) \times n$  矩阵. 因此,

$$\mathcal{N}(A) \cap \mathcal{N}(B) = \mathcal{N}(D),$$

三个零空间都是  $\mathbb{F}^n$  的子空间. ☺

**例 7.1.14** 考虑线性空间  $\mathbb{F}^{n \times n}$ , 设  $\mathcal{U}, \mathcal{L}$  分别是由所有上/下三角矩阵构成的子空间, 则  $\mathcal{U} \cap \mathcal{L}$  是所有对角矩阵构成的子空间. 而  $\mathcal{U} + \mathcal{L} = \mathbb{F}^{n \times n}$ , 因为任意方阵显然能分解为上三角矩阵和下三角矩阵的和, 记  $A = U + L$ . 注意, 这个分解式并不唯一, 因为对任意非零对角矩阵  $D$ ,  $A = (U + D) + (L - D)$  是一个不同的分解式. ☺

例 7.1.14 引出一个自然的问题:  $\mathcal{M}_1 + \mathcal{M}_2$  中的向量分解为  $\mathcal{M}_1$  和  $\mathcal{M}_2$  中的向量之和, 分解方式何时唯一? 为此引入如下定义.

**定义 7.1.15 (子空间的直和)** 给定线性空间  $\mathcal{V}$  的两个子空间  $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2$ . 如果  $\mathcal{M}$  的任意向量  $\mathbf{m}$  的分解式

$$\mathbf{m} = \mathbf{m}_1 + \mathbf{m}_2, \quad \mathbf{m}_1 \in \mathcal{M}_1, \mathbf{m}_2 \in \mathcal{M}_2,$$

唯一, 则称  $\mathcal{M}$  为  $\mathcal{M}_1$  与  $\mathcal{M}_2$  的**直和**, 也称  $\mathcal{M}_1 + \mathcal{M}_2$  是直和, 记作  $\mathcal{M} = \mathcal{M}_1 \oplus \mathcal{M}_2$ .

子空间的和是直和, 要求其中每一个向量的分解式都唯一, 这看起来很苛刻. 然而事实上, 只验证零向量的分解式唯一即可.

**定理 7.1.16** 对线性空间  $\mathcal{V}$  的两个有限维子空间  $\mathcal{M}_1, \mathcal{M}_2$ , 以下叙述等价:

1.  $\mathcal{M}_1 + \mathcal{M}_2$  是直和;
2. 零向量有唯一的分解式, 即  $\mathbf{0} = \mathbf{m}_1 + \mathbf{m}_2, \mathbf{m}_1 \in \mathcal{M}_1, \mathbf{m}_2 \in \mathcal{M}_2$ , 推出  $\mathbf{m}_1 = \mathbf{m}_2 = \mathbf{0}$ ;
3.  $\mathcal{M}_1 \cap \mathcal{M}_2 = \{\mathbf{0}\}$ .

证. 采用轮转证法.

$1 \Rightarrow 2$ : 显然, 因为零向量是  $\mathcal{M}_1 + \mathcal{M}_2$  中的一个向量.

$2 \Rightarrow 3$ : 对任意  $\mathbf{m} \in \mathcal{M}_1 \cap \mathcal{M}_2$ , 则  $\mathbf{m} \in \mathcal{M}_1, -\mathbf{m} \in \mathcal{M}_2$ , 于是零向量有分解式  $\mathbf{0} = \mathbf{m} + (-\mathbf{m})$ , 由第 2 条即有  $\mathbf{m} = \mathbf{0}$ .

$3 \Rightarrow 1$ : 根据定义验证. 对任意  $\mathbf{m} \in \mathcal{M}$ , 有两个分解式

$$\mathbf{m} = \mathbf{m}_1 + \mathbf{m}_2 = \mathbf{m}'_1 + \mathbf{m}'_2.$$

于是

$$\mathbf{m}_1 - \mathbf{m}'_1 = \mathbf{m}'_2 - \mathbf{m}_2.$$

注意, 等式左端  $(\mathbf{m}_1 - \mathbf{m}'_1) \in \mathcal{M}_1$ , 等式右端  $(\mathbf{m}'_2 - \mathbf{m}_2) \in \mathcal{M}_2$ . 因此这个向量属于  $\mathcal{M}_1 \cap \mathcal{M}_2$ , 由第 3 条, 只能是零向量. 因此  $\mathbf{m}_1 = \mathbf{m}'_1, \mathbf{m}_2 = \mathbf{m}'_2$ , 即分解式唯一.  $\square$

**例 7.1.17** 考虑线性空间  $\mathbb{F}^{n \times n}$ , 设  $\mathcal{M}_1, \mathcal{M}_2$  分别是由所有对称/反对称矩阵构成的子空间, 则  $\mathcal{M}_1 \cap \mathcal{M}_2 = \{\mathbf{0}\}$ ,  $\mathcal{M}_1 \oplus \mathcal{M}_2$  是直和. 任意方阵都能分解成对称矩阵和反对称矩阵的和:  $A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$ . 于是,  $\mathcal{M}_1 \oplus \mathcal{M}_2 = \mathbb{F}^{n \times n}$ . 因此, 上述分解式唯一.  $\odot$

## 习题

**练习 7.1.1** 在所有正实数构成的集合  $\mathbb{R}^+$  上, 定义加法和数乘运算:

$$a \oplus b := ab, \quad k \odot a := a^k, \quad \forall a, b \in \mathbb{R}^+, k \in \mathbb{R}.$$

判断  $\mathbb{R}^+$  对这两个运算是否构成  $\mathbb{R}$  上的线性空间.

**练习 7.1.2** 令  $\omega = \frac{-1+\sqrt{3}i}{2}$ ,  $\mathbb{Q}[\omega] = \{a + b\omega \mid a, b \in \mathbb{Q}\}$ .

1. 证明  $\mathbb{Q}[\omega]$  关于数的加法和数乘构成  $\mathbb{Q}$  上的一个线性空间.
2. 证明子集  $\mathbb{Q}$  和  $\mathcal{M} = \{b\omega \mid b \in \mathbb{Q}\}$  都是  $\mathbb{Q}[\omega]$  的子空间. 并求二者的交与和.



3. 判断  $\mathbb{Q}[\omega]$  是否是数域.

**练习 7.1.3** 把复数域  $\mathbb{C}$  看作有理数域  $\mathbb{Q}$  上的线性空间, 子集  $\mathbb{R}$  是否是子空间?

**练习 7.1.4** 设  $\mathbb{Q}[i] = \{a + bi \mid a, b \in \mathbb{Q}\}$ .

1. 证明  $\mathbb{Q}[i]$  关于数的加法和数乘构成  $\mathbb{Q}$  上的一个线性空间.

2. 证明  $\mathbb{Q}[i]$  是数域.

3. 把复数域  $\mathbb{C}$  看作有理数域  $\mathbb{Q}[i]$  上的线性空间, 子集  $\mathbb{R}$  是否是子空间?

**练习 7.1.5** 设  $\mathcal{V}$  是以 0 为极限的实数序列全体:  $\mathcal{V} = \left\{ \{a_n\} \mid \lim_{n \rightarrow \infty} a_n = 0 \right\}$ . 定义加法和数乘分别为:

$$\{a_n\} + \{b_n\} = \{a_n + b_n\}; \quad k\{a_n\} = \{ka_n\}, k \in \mathbb{R}.$$

证明  $\mathcal{V}$  是  $\mathbb{R}$  上的线性空间.

**练习 7.1.6** 设  $\mathcal{M}_1 = \text{span}(\mathbf{a}_1, \mathbf{a}_2), \mathcal{M}_2 = \text{span}(\mathbf{b}_1, \mathbf{b}_2)$ , 其中

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \\ -3 \end{bmatrix}, \mathbf{b}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ -2 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -3 \end{bmatrix}.$$

分别求  $\mathcal{M}_1 + \mathcal{M}_2, \mathcal{M}_1 \cap \mathcal{M}_2$  的一组基和维数.

**练习 7.1.7** 设  $\mathbb{F}_0^{n \times n}$  是矩阵空间  $\mathbb{F}^{n \times n}$  中所有迹为零的矩阵构成的子集.

1. 证明  $\mathbb{F}_0^{n \times n}$  是  $\mathbb{F}^{n \times n}$  的子空间.

2. 求子空间  $\mathbb{F}_0^{n \times n}$  和  $\text{span}(I_n)$  的交与和.

**练习 7.1.8** 对  $n$  阶方阵  $A$ , 令  $P(A) = \{f(A) \mid f(x) \in \mathbb{F}[x]\}$ . 证明  $P(A)$  关于矩阵的加法和数乘构成  $\mathbb{F}$  上的线性空间.

**练习 7.1.9** 对  $n$  阶方阵  $A$ , 令  $\text{Com}(A) = \{n \text{ 阶方阵 } B \mid AB = BA\}$ .

1. 证明,  $\text{Com}(A)$  是  $\mathbb{F}^{n \times n}$  的子空间.

2. 证明, 对任意  $B, C \in \text{Com}(A)$ , 都有  $BC \in \text{Com}(A)$ ; 由此证明对任意多项式  $f(x)$ , 都有  $f(A) \in \text{Com}(A)$ .

**练习 7.1.10** 1. 可逆矩阵的全体不是  $\mathbb{F}^{n \times n}$  的子空间.

2. 奇异矩阵的全体不是  $\mathbb{F}^{n \times n}$  的子空间.

## 7.2 基和维数

本节的概念都是第 2 章对应概念的直接推广, 结论及其证明也基本一致, 因此这里将略去大部分证明, 请读者参考第 2.1 至 2.2 节.

我们首先给出线性空间中向量的线性组合、线性无关/相关等概念.

**定义 7.2.1 (线性组合、生成、线性无关)** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{V}$  内的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  和数  $k_1, \dots, k_n \in \mathbb{F}$ , 称向量  $k_1\mathbf{a}_1 + \dots + k_n\mathbf{a}_n$  是向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  的一个**线性组合**.

若向量  $\mathbf{b}$  和向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  的一个线性组合相等, 则称  $\mathbf{b}$  可以被向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  **线性表示**.

若向量组  $\mathbf{b}_1, \dots, \mathbf{b}_m$  中的每一个向量都可以被向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性表示, 则称向量组  $\mathbf{b}_1, \dots, \mathbf{b}_m$  可以被向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  **线性表示**.

向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  的线性组合的全体构成  $\mathcal{V}$  的一个子空间, 称为该向量组**生成**的子空间, 记作  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ .

如果存在  $\mathbb{F}$  内的  $n$  个不全为 0 的数  $k_1, \dots, k_n$ , 使得  $k_1\mathbf{a}_1 + \dots + k_n\mathbf{a}_n = \mathbf{0}$ , 则称向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  **线性相关**.

如果由  $k_1\mathbf{a}_1 + \dots + k_n\mathbf{a}_n = \mathbf{0}$  必定推出  $k_1 = \dots = k_n = 0$ , 则称向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$  **线性无关**.

在坐标向量空间  $\mathbb{F}^n$  中, 向量组的线性关系能够通过线性方程组判断. 对一般的线性空间, 则只能根据其上的线性运算具体分析.

**例 7.2.2** 实数集  $\mathbb{R}$  作为  $\mathbb{R}$  上的线性空间, 其中的向量组  $1, \pi$  线性相关, 因为  $(\pi) \cdot 1 + (-1) \cdot \pi = 0$ .

实数集  $\mathbb{R}$  作为  $\mathbb{Q}$  上的线性空间, 其中的向量组  $1, \pi$  线性无关, 因为  $\pi$  是无理数, 对任意不全为零的有理数  $k_1, k_2$ ,  $k_1 \cdot 1 + k_2 \cdot \pi \neq 0$ .  $\odot$

**例 7.2.3** 连续函数空间  $C(\mathbb{R})$  中的向量组  $\sin x, \sin 2x, \sin 3x$  线性无关.

考察方程  $k_1 \sin x + k_2 \sin 2x + k_3 \sin 3x = 0$ . 注意等式右端是零函数, 即函数空间中的零向量. 两函数相等是指函数值处处相等. 选择三个值  $\frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}$  代入, 有

$$\begin{cases} k_1 - k_3 = 0, \\ \frac{\sqrt{3}}{2}k_1 + \frac{\sqrt{3}}{2}k_2 = 0, \\ \frac{\sqrt{2}}{2}k_1 + k_2 + \frac{\sqrt{2}}{2}k_3 = 0. \end{cases}$$

这一方程组只有零解  $k_1 = k_2 = k_3 = 0$ , 因此向量组线性无关.  $\odot$

**例 7.2.4** 给定相异实数  $\lambda_1, \lambda_2$ , 光滑函数空间  $C^\infty(\mathbb{R})$  中的向量组  $e^{\lambda_1 x}, e^{\lambda_2 x}$  线性无关.

考察方程  $k_1 e^{\lambda_1 x} + k_2 e^{\lambda_2 x} = 0$ . 可以采用与例 7.2.3 相同的方法得到结论. 下面给出另一种方法.

由于函数等式成立, 故对其两端求导数等式也成立. 两端对  $x$  求导数, 得  $k_1 \lambda_1 e^{\lambda_1 x} + k_2 \lambda_2 e^{\lambda_2 x} = 0$ . 由此得到关于函数的矩阵方程:

$$\begin{bmatrix} e^{\lambda_1 x} & e^{\lambda_2 x} \\ \lambda_1 e^{\lambda_1 x} & \lambda_2 e^{\lambda_2 x} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

系数矩阵的行列式是

$$\begin{vmatrix} e^{\lambda_1 x} & e^{\lambda_2 x} \\ \lambda_1 e^{\lambda_1 x} & \lambda_2 e^{\lambda_2 x} \end{vmatrix} = e^{(\lambda_1 + \lambda_2)x} \begin{vmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{vmatrix} = e^{(\lambda_1 + \lambda_2)x} (\lambda_2 - \lambda_1) \neq 0.$$

对任意实数  $x$ , 系数矩阵都可逆, 因此方程组只有零解  $k_1 = k_2 = 0$ .

利用同一方法能够证明, 对任意相异实数  $\lambda_1, \dots, \lambda_n$ , 向量组  $e^{\lambda_1 x}, \dots, e^{\lambda_n x}$  线性无关. ☺

综合线性表示和线性无关两个概念有如下定义.

**定义 7.2.5 (极大线性无关部分组)** 给定线性空间  $\mathcal{V}$  中的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_s$ , 如果其部分组  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  满足:

1.  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  线性无关;
2.  $\mathbf{a}_1, \dots, \mathbf{a}_s$  可以被  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  线性表示;

则称  $\mathbf{a}_{i_1}, \dots, \mathbf{a}_{i_r}$  是  $\mathbf{a}_1, \dots, \mathbf{a}_s$  的一个**极大线性无关部分组**.

极大线性无关部分组仍然可以利用筛选法构造得到, 从而证明其存在性. 关键仍然是如下线性表示与线性无关之间的关系.

**命题 7.2.6** 如果向量组  $\mathbf{a}_1, \dots, \mathbf{a}_s$  线性无关, 则对任意向量  $\mathbf{b}$ , 有

1. 向量组  $\mathbf{a}_1, \dots, \mathbf{a}_s, \mathbf{b}$  线性相关当且仅当  $\mathbf{b}$  可以被向量组  $\mathbf{a}_1, \dots, \mathbf{a}_s$  线性表示;
2.  $\mathbf{b}$  不能被向量组  $\mathbf{a}_1, \dots, \mathbf{a}_s$  线性表示当且仅当  $\mathbf{a}_1, \dots, \mathbf{a}_s, \mathbf{b}$  线性无关.

**例 7.2.7** 假设物体放在斜面上, 一个人用力推着这个物体匀速往下走, 如图 7.2.1 所示.

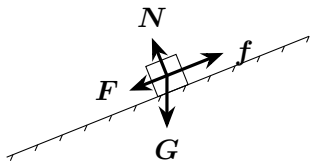


图 7.2.1: 受力分析

与三维几何空间类似, 空间中的力在力的合成和倍数下构成  $\mathbb{R}$  上的线性空间. 该物体受到四个力的作用: 人的推力  $\mathbf{F}$ , 地面的摩擦力  $\mathbf{f}$ , 重力  $\mathbf{G}$ , 以及地面的支持力  $\mathbf{N}$ . 考察这四个向量组成的向量组, 其极大线性无关组可以利用筛选法得到. 先看  $\mathbf{F}$ , 它不

为零, 保留. 再看  $f$ , 它与  $F$  方向相反, 可以被  $F$  线性表示, 因此去掉  $f$ . 再看  $G$ , 重力方向竖直向下, 而  $F$  沿着斜面的方向, 二者不共线, 因此  $F, G$  线性无关. 最后看  $N$ , 由力的分解可知, 它可以被  $F, G$  线性表示, 因此  $F, G, N$  线性相关, 去掉  $N$ . 于是极大线性无关部分是  $F, G$ .

不难发现, 在这四个向量中任取两个, 只要不是  $F, f$  这一对, 都是原向量组的极大线性无关部分组. ☺

如果  $b$  可以被  $a_1, \dots, a_s$  线性表示, 则存在数  $k_1, \dots, k_s$ , 使得  $b = k_1 a_1 + \dots + k_s a_s$ . 仿照矩阵和向量的乘法, 这可以写作  $b = (a_1, \dots, a_s) \begin{bmatrix} k_1 \\ \vdots \\ k_s \end{bmatrix}$ . 注意这并不是真正的矩阵和向量的乘法, 而只是借用了记号来表示线性组合<sup>1</sup>. 类似地, 如果  $b_1, \dots, b_t$  可以被  $a_1, \dots, a_s$  线性表示, 仿照矩阵乘法, 也可以写作

$$(b_1, \dots, b_t) = (a_1, \dots, a_s) \begin{bmatrix} k_{11} & \cdots & k_{1t} \\ \vdots & & \vdots \\ k_{s1} & \cdots & k_{st} \end{bmatrix},$$

注意这也不是真正的矩阵乘法, 也只是借用了记号来表示线性组合. 利用这种表示方法, 根据系数矩阵的性质, 容易证明命题 2.2.10 的结论对一般的线性空间都成立. 由此立得如下结论.

**命题 7.2.8** 如果向量组  $a_1, \dots, a_s$  和  $b_1, \dots, b_t$  可以互相线性表示, 且两个向量组分别线性无关, 则  $s = t$ .

如果两个向量组可以互相线性表示, 则称二者**线性等价**. 一个向量组的任意两个极大线性无关部分组线性等价. 因此, 二者中的向量个数相同.

**定义 7.2.9 (秩)** 一个向量组  $S$  的任意一个极大线性无关部分组中向量的个数称为这个向量组的**秩**, 记为  $\text{rank}(S)$ . 一个只包含零向量的向量组的秩定义为零.

类似地, 我们有如下关于线性空间的最重要的定义.

**定义 7.2.10 (基、维数)** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{V}$ . 如果  $\mathcal{V}$  中存在一个线性无关的向量组,  $\mathcal{V}$  中的任意向量都可以被它线性表示, 则称该向量组为  $\mathcal{V}$  的一组**基**.

如果  $\mathcal{V}$  中存在  $n$  个向量组成的一组基, 则称  $\mathcal{V}$  为  $n$  **维线性空间**, 又称  $\mathcal{V}$  的**维数**是  $n$ , 记为  $\dim \mathcal{V} = n$ .

如果  $\mathcal{V}$  中存在任意多个线性无关的向量, 则称其为**无限维线性空间**; 反之, 则称为**有限维线性空间**.

<sup>1</sup> 为了提示与坐标向量空间中的向量的区别, 本书给该向量组外加圆括号.

单由零向量组成的线性空间  $\{0\}$ , 其维数定义为 0.

定义 7.2.10 与  $\mathbb{R}^n$  的子空间的基的定义 2.1.12 一致. 特别地,  $\mathbb{F}^n$  的子空间都是有限维线性空间.

**例 7.2.11** 回顾例 7.1.4.

三维几何空间中, 任意不共面的三个向量都是一组基, 因此几何空间的维数是 3.

复数集  $\mathbb{C}$  作为  $\mathbb{R}$  上的线性空间,  $1, i$  是一组基, 维数是 2.

复数集  $\mathbb{C}$  作为  $\mathbb{C}$  上的线性空间, 任意非零复数是一组基, 维数是 1.

实数集  $\mathbb{R}$  作为  $\mathbb{Q}$  上的线性空间, 由对任意  $n$ , 向量组  $1, \pi, \dots, \pi^n$  在  $\mathbb{Q}$  上线性无关这一事实<sup>2</sup>可知,  $\mathbb{R}$  是  $\mathbb{Q}$  上的无限维线性空间.  $\odot$

**例 7.2.12** 矩阵空间  $\mathbb{F}^{m \times n}$  中, 对任意的  $1 \leq i \leq m, 1 \leq j \leq n$ , 设  $E_{ij}$  是  $(i, j)$  元为 1, 其余元素都是 0 的矩阵. 根据定义直接验证, 这  $mn$  个向量  $E_{ij}, 1 \leq i \leq m, 1 \leq j \leq n$  是  $\mathbb{F}^{m \times n}$  的一组基. 因此,  $\dim \mathbb{F}^{m \times n} = mn$ .  $\odot$

**例 7.2.13** 线性空间  $\mathbb{R}[x]_n$  中,  $1, x, x^2, \dots, x^{n-1}$  线性无关. 该命题有多种证法.

证一. 设一组数  $k_0, k_1, \dots, k_{n-1}$  使得  $k_0 + k_1x + \dots + k_{n-1}x^{n-1} = 0$ . 注意等式右端是零多项式. 选择  $n$  个两两不同的数  $\lambda_1, \dots, \lambda_n$  代入, 可以得到齐次线性方程组

$$\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{bmatrix} \begin{bmatrix} k_0 \\ k_1 \\ \vdots \\ k_{n-1} \end{bmatrix} = \mathbf{0}.$$

系数矩阵是 Vandermonde 矩阵, 由于  $\lambda_1, \dots, \lambda_n$  两两不同, 系数矩阵可逆, 方程组只有零解. 故  $1, x, \dots, x^{n-1}$  线性无关.  $\square$

证二. 设一组数  $k_0, k_1, \dots, k_{n-1}$  使得  $k_0 + k_1x + \dots + k_{n-1}x^{n-1} = 0$ . 注意等式右端是零多项式. 由于函数等式成立, 故对其两端求导数等式也成立, 即  $k_1 + 2k_2x + \dots + (n-1)k_{n-1}x^{n-2} = 0$ . 类似地求  $n-1$  次导数, 我们可得齐次线性方程组

$$\begin{bmatrix} 1 & x & x^2 & \cdots & x^{n-1} \\ 0 & 1 & 2x & \ddots & \vdots \\ 0 & 0 & 2 & \ddots & \frac{(n-1)!}{2}x^2 \\ \vdots & \vdots & \ddots & \ddots & (n-1)!x \\ 0 & 0 & \cdots & 0 & (n-1)! \end{bmatrix} \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ \vdots \\ k_{n-1} \end{bmatrix} = \mathbf{0}.$$

显然方程组只有零解. 故  $1, x, \dots, x^{n-1}$  线性无关.  $\square$

<sup>2</sup>这是因为  $\pi$  是超越数, 感兴趣的读者请查阅数论教材或专著.

证三. 利用数学归纳法.  $n=1$  时, 显然线性无关. 现假设  $1, x, \dots, x^{n-2}$  线性无关. 若  $1, x, \dots, x^{n-2}, x^{n-1}$  线性相关, 则  $x^{n-1}$  可以被  $1, x, \dots, x^{n-2}$  线性表示. 这意味着存在  $k_0, k_1, \dots, k_{n-2}$  使得  $x^{n-1} = k_0 + k_1x + \dots + k_{n-2}x^{n-2}$ . 将其看作一元  $n-1$  次方程, 则它至多有  $n-1$  个不同的根. 选择这些根之外的数  $\lambda$ , 则  $\lambda^{n-1} \neq k_0 + k_1\lambda + \dots + k_{n-2}\lambda^{n-2}$ . 矛盾. 故  $1, x, \dots, x^{n-1}$  也线性无关. 综上所述, 对任意  $n$ , 结论都成立.  $\square$

向量组  $1, x, x^2, \dots, x^{n-1}$  显然可以线性生成  $\mathbb{R}[x]_n$ , 因此它是一组基,  $\dim \mathbb{R}[x]_n = n$ .

由于对任意  $n$ ,  $1, x, x^2, \dots, x^{n-1}$  线性无关, 多项式空间  $\mathbb{R}[x]$  是无限维线性空间. 而其中由  $1, x, x^2, \dots, x^{n-1}$  线性生成的子空间正是  $\mathbb{R}[x]_n$ .  $\odot$

如下结论与关于  $\mathbb{R}^n$  的子空间的结论类似, 见命题 2.2.17 和定理 2.2.14.

**命题 7.2.14** 对  $n$  维线性空间  $\mathcal{V}$ , 给定其中含有  $n$  个向量的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_n$ .

1. 如果  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性无关, 则  $\mathbf{a}_1, \dots, \mathbf{a}_n$  是  $\mathcal{V}$  的一组基;
2. 如果  $\mathcal{V} = \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$ , 则  $\mathbf{a}_1, \dots, \mathbf{a}_n$  是  $\mathcal{V}$  的一组基.

事实上,  $\dim \mathcal{V} = n$ ;  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性无关;  $\mathbf{a}_1, \dots, \mathbf{a}_n$  线性生成  $\mathcal{V}$ : 这三个条件中的任意两个都可以推出另外一个, 因此都可以作为基的判定条件.

**命题 7.2.15** 有限维线性空间  $\mathcal{V}$  中任意  $r$  个线性无关的向量  $\mathbf{a}_1, \dots, \mathbf{a}_r$  都可以扩充成  $\mathcal{V}$  的一组基.

证. 设  $\dim \mathcal{V} = n$ . 显然  $r \leq n$ , 不然与维数定义矛盾. 若  $r = n$ , 则该向量组已经是一组基. 下面关注  $r < n$  的情形. 此时该向量组不是一组基, 则  $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_r) \neq \mathcal{V}$ , 故存在  $\mathbf{a}_{r+1} \in \mathcal{V}$ , 且  $\mathbf{a}_{r+1} \notin \text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_r\}$ . 此时  $\mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{a}_{r+1}$  线性无关, 因为若线性相关, 则根据命题 7.2.6,  $\mathbf{a}_{r+1}$  能被  $\mathbf{a}_1, \dots, \mathbf{a}_r$  线性表示, 即  $\mathbf{a}_{r+1} \in \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_r)$ , 矛盾.

现在已有  $r+1$  个线性无关的向量. 重复上述步骤  $n-r$  次, 得到  $n$  个线性无关的向量  $\mathbf{a}_1, \dots, \mathbf{a}_n$ . 因为  $\dim \mathcal{V} = n$ , 由命题 7.2.14 可知, 该向量组是  $\mathcal{V}$  的一组基.  $\square$

**推论 7.2.16** 给定有限维线性空间  $\mathcal{V}$  的子空间  $\mathcal{M}$ , 则  $\mathcal{M}$  的任意一组基都可以扩充成  $\mathcal{V}$  的一组基. 因此  $\dim \mathcal{M} \leq \dim \mathcal{V}$ .

下面来看子空间的运算与维数之间的关系.

给定有限维线性空间  $\mathcal{V}$  的两个子空间  $\mathcal{M}_1, \mathcal{M}_2$ , 根据定理 2.2.14 和推论 7.2.16 中基扩张“从小到大”的原则, 先取  $\mathcal{M}_1 \cap \mathcal{M}_2$  的一组基

$$R: \mathbf{a}_1, \dots, \mathbf{a}_r,$$

将其分别扩充成  $\mathcal{M}_1$  和  $\mathcal{M}_2$  的一组基  $S$  和  $T$ :

$$S: \mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{b}_1, \dots, \mathbf{b}_s,$$

$$T: \mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{c}_1, \dots, \mathbf{c}_t.$$

此时,  $R = S \cap T$ . 那么  $S \cup T$  是什么, 是  $\mathcal{M}_1 + \mathcal{M}_2$  的一组基吗?

**定理 7.2.17** 结合如上记号, 向量组  $S \cup T: \mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{b}_1, \dots, \mathbf{b}_s, \mathbf{c}_1, \dots, \mathbf{c}_t$  是  $\mathcal{M}_1 + \mathcal{M}_2$  的一组基. 特别地, 如下维数公式成立:

$$\dim(\mathcal{M}_1 + \mathcal{M}_2) = \dim \mathcal{M}_1 + \dim \mathcal{M}_2 - \dim \mathcal{M}_1 \cap \mathcal{M}_2.$$

证. 根据基的定义, 只需验证  $S \cup T$  线性生成  $\mathcal{M}_1 + \mathcal{M}_2$ , 而且线性无关.

首先,  $\mathcal{M}_1 + \mathcal{M}_2$  中的任意向量都有分解式  $\mathbf{m} + \mathbf{n}$ , 其中  $\mathbf{m} \in \mathcal{M}_1, \mathbf{n} \in \mathcal{M}_2$ . 二者分别可以被向量组  $S$  和  $T$  线性表示, 因此  $\mathbf{m} + \mathbf{n}$  可以被  $S \cup T$  线性表示. 亦即,  $\mathcal{M}_1 + \mathcal{M}_2 \subseteq \text{span}(S \cup T)$ . 因此  $\mathcal{M}_1 + \mathcal{M}_2 = \text{span}(S \cup T)$ .

其次, 证明向量组  $S \cup T$  线性无关. 考虑方程

$$k_1 \mathbf{a}_1 + \dots + k_r \mathbf{a}_r + l_1 \mathbf{b}_1 + \dots + l_s \mathbf{b}_s + m_1 \mathbf{c}_1 + \dots + m_t \mathbf{c}_t = \mathbf{0}. \quad (7.2.1)$$

目标是证明所有系数都是零. 移项可得

$$k_1 \mathbf{a}_1 + \dots + k_r \mathbf{a}_r + l_1 \mathbf{b}_1 + \dots + l_s \mathbf{b}_s = -(m_1 \mathbf{c}_1 + \dots + m_t \mathbf{c}_t).$$

等式左端向量在  $\mathcal{M}_1$  内, 右端向量在  $\mathcal{M}_2$  内, 所以该向量在  $\mathcal{M}_1 \cap \mathcal{M}_2$  内. 因此, 它可以被向量组  $R: \mathbf{a}_1, \dots, \mathbf{a}_r$  线性表示:

$$-(m_1 \mathbf{c}_1 + \dots + m_t \mathbf{c}_t) = n_1 \mathbf{a}_1 + \dots + n_r \mathbf{a}_r.$$

移项即得

$$n_1 \mathbf{a}_1 + \dots + n_r \mathbf{a}_r + m_1 \mathbf{c}_1 + \dots + m_t \mathbf{c}_t = \mathbf{0}.$$

而向量组  $T: \mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{c}_1, \dots, \mathbf{c}_t$  线性无关, 因此  $n_1 = \dots = n_r = m_1 = \dots = m_t = 0$ . 代入 (7.2.1), 有

$$k_1 \mathbf{a}_1 + \dots + k_r \mathbf{a}_r + l_1 \mathbf{b}_1 + \dots + l_s \mathbf{b}_s = \mathbf{0}.$$

向量组  $S: \mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{b}_1, \dots, \mathbf{b}_s$  线性无关, 因此  $k_1 = \dots = k_r = l_1 = \dots = l_s = 0$ . 这说明 (7.2.1) 中所有系数都是零, 故向量组  $S \cup T$  线性无关.

对四个子空间的基中向量计数, 立得维数公式. □

从定理 7.2.17 中的维数公式可得如下结论.

**推论 7.2.18** 给定线性空间  $\mathcal{V}$  的两个有限维子空间  $\mathcal{M}_1, \mathcal{M}_2$ ,

1.  $\mathcal{M}_1 + \mathcal{M}_2$  是直和, 当且仅当  $\dim(\mathcal{M}_1 + \mathcal{M}_2) = \dim \mathcal{M}_1 + \dim \mathcal{M}_2$ .
2. 若  $\mathcal{M} = \mathcal{M}_1 \oplus \mathcal{M}_2$ , 则  $\mathcal{M}_1$  与  $\mathcal{M}_2$  各取一组基, 其并集就是  $\mathcal{M}$  的一组基.

证. 第 1 条: 利用定理 7.1.16 和维数公式.

第 2 条: 易见, 两组基的并集可以线性表示  $\mathcal{M}_1$  中的任意向量. 再由维数公式和命题 2.2.17 可知, 它是一组基.  $\square$

**例 7.2.19** 例 7.2.12 给出了矩阵空间  $\mathbb{F}^{n \times n}$  的一组基  $E_{ij}$ , 下面给出几个子空间的基.

1. 考虑例 7.1.14 中的子空间  $\mathcal{U}, \mathcal{L}$ . 易得  $E_{ii}, 1 \leq i \leq n$  是  $\mathcal{U} \cap \mathcal{L}$  的一组基; 由此扩充成  $\mathcal{U}$  的一组基  $E_{ij}, 1 \leq i \leq j \leq n$ , 和  $\mathcal{L}$  的一组基  $E_{ij}, 1 \leq j < i \leq n$ . 于是,  $\mathcal{U} + \mathcal{L}$  的一组基是  $E_{ij}, 1 \leq i, j \leq n$ , 这得到  $\mathcal{U} + \mathcal{L} = \mathbb{F}^{n \times n}$ . 此时的维数公式是  $n^2 = \frac{n^2+n}{2} + \frac{n^2+n}{2} - n$ .
2. 考虑例 7.1.17 中的子空间  $\mathcal{M}_1, \mathcal{M}_2$ . 易得  $E_{ij} + E_{ji}, E_{kk}, 1 \leq i < j \leq n, 1 \leq k \leq n$  是  $\mathcal{M}_1$  的一组基,  $E_{ij} - E_{ji}, 1 \leq i < j \leq n$  是  $\mathcal{M}_2$  的一组基. 两组基的并集是  $\mathbb{F}^{n \times n}$  的一组基.  $\odot$

## 习题

**练习 7.2.1** 把数域  $\mathbb{F}$  看作自身上的线性空间, 求它的一组基和维数.

**练习 7.2.2** 求练习 7.1.1 中线性空间  $\mathbb{R}$  的一组基和维数.

**练习 7.2.3** 在练习 7.1.2 中的线性空间  $\mathbb{Q}[\omega] = \{a + b\omega \mid a, b \in \mathbb{Q}\}$  内,

1. 求下列向量组的秩:  $S_1: \frac{1}{2}, 3, -7, S_2: 1, \omega, \omega^2, \omega^3, S_3: \omega, \bar{\omega}, \sqrt{3}$ .
2. 求  $\mathbb{Q}[\omega]$  的一组基和维数.

**练习 7.2.4** 判断练习 7.1.5 中线性空间的维数是否有限.

**练习 7.2.5** 判断  $\mathbb{R}$  上的线性空间  $C[-\pi, \pi]$  内的下列向量组是否线性相关, 并求其秩.

1.  $\cos^2 x, \sin^2 x.$
2.  $\cos^2 x, \sin^2 x, 1.$
3.  $\cos 2x, \sin 2x.$
4.  $1, \sin x, \sin 2x, \dots, \sin nx.$
5.  $1, \sin x, \sin^2 x, \dots, \sin^n x.$

**练习 7.2.6** 考虑练习 7.1.9 中的线性空间  $\text{Com}(A)$ , 对下列  $A$  求  $\text{Com}(A)$  的一组基.

1.  $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$
2.  $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$
3.  $\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}_{n \times n}.$
4.  $\text{diag}(a_i)$ , 其中  $a_i$  各不相同.
5.  $\text{diag}(a_i).$



**练习 7.2.7** 给定  $\mathbb{F}$  中两两不相等的数  $a_1, \dots, a_n$ .

1. 在线性空间  $\mathbb{F}[x]_n$  中, 令

$$f_i(x) = (x - a_1) \cdots \widehat{(x - a_i)} \cdots (x - a_n), \quad i = 1, \dots, n,$$

其中  $\widehat{(x - a_i)}$  表示不含该项. 证明,  $f_1(x), \dots, f_n(x)$  是  $\mathbb{F}[x]_n$  的一组基.

2. 设  $b_1, \dots, b_n$  是  $\mathbb{F}$  中任意  $n$  个数, 找出  $f(x) \in \mathbb{F}[x]_n$ , 使得  $f(a_i) = b_i, i = 1, \dots, n$ .

**练习 7.2.8** 证明,  $n$  维线性空间中任意多于  $n$  个的向量都线性相关.

**练习 7.2.9** 考虑练习 7.1.8 中的线性空间  $P(A)$ .

1. 判断其维数是否有限.
2. 证明存在次数不大于  $n^2$  的多项式  $f(x) \in \mathbb{F}[x]$ , 使得  $f(A) = O$ .
3. 令  $A = \text{diag}(1, \omega, \omega^2)$ , 其中  $\omega = \frac{-1 + \sqrt{3}i}{2}$ , 求  $P(A)$  的维数和一组基.

**练习 7.2.10** 设  $\mathcal{M}$  是  $\mathbb{R}^n$  的子空间,  $\mathcal{M}^\perp$  是其正交补空间, 证明,  $\mathbb{R}^n = \mathcal{M} \oplus \mathcal{M}^\perp$ .

**练习 7.2.11** 证明, 练习 7.1.7 中的  $\mathbb{F}^{n \times n} = \mathbb{F}_0^{n \times n} \oplus \text{span}(I_n)$ .

**练习 7.2.12** 证明连续函数空间的子集  $\text{span}(f(x) = k_0 + k_1 \cos x + k_2 \cos 2x \mid f(0) = 0)$  是一个子空间, 并求一组基.

**练习 7.2.13** 给定  $A = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$  和  $\mathbb{R}^{3 \times 3}$  上的线性变换  $f: X \mapsto AX$ . 分别求  $\mathcal{N}(f)$  和  $\mathcal{R}(f)$  的维数和一组基.

**练习 7.2.14** 设  $\mathcal{M}_1, \mathcal{M}_2$  是  $\mathbb{R}^n$  的两个子空间, 且  $\mathcal{M}_1 \subseteq \mathcal{M}_2$ . 证明, 如果  $\dim \mathcal{M}_1 = \dim \mathcal{M}_2$ , 则  $\mathcal{M}_1 = \mathcal{M}_2$ .

## 7.3 线性映射

上节我们推广了线性空间的概念, 本节将推广线性映射的概念.

**定义 7.3.1 (线性映射)** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{U}, \mathcal{V}$ , 如果从  $\mathcal{U}$  到  $\mathcal{V}$  的映射  $f$  满足

1. 对任意  $\mathbf{a}, \mathbf{b} \in \mathcal{U}$ , 有  $f(\mathbf{a} + \mathbf{b}) = f(\mathbf{a}) + f(\mathbf{b})$ ;
2. 对任意  $\mathbf{a} \in \mathcal{U}, k \in \mathbb{F}$ , 有  $f(k\mathbf{a}) = kf(\mathbf{a})$ ,

则称其为  $\mathcal{U}$  到  $\mathcal{V}$  的**线性映射**,  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射的全体记作  $\text{Hom}(\mathcal{U}, \mathcal{V})$ .

特别地, 对任意线性映射  $f$ , 都有  $f(\mathbf{0}_{\mathcal{U}}) = \mathbf{0}_{\mathcal{V}}$ . 易见定义 7.3.1 中的两个条件和如下条件等价:

对任意  $\mathbf{a}, \mathbf{b} \in \mathcal{U}, k, l \in \mathbb{F}$ , 有  $f(k\mathbf{a} + l\mathbf{b}) = kf(\mathbf{a}) + lf(\mathbf{b})$ .

**例 7.3.2** 下面列出几个常见的线性映射.

1. 设  $A \in \mathbb{F}^{m \times n}$ , 左乘矩阵  $A$  定义了一个线性映射:

$$\begin{aligned} \mathbf{L}_A: \mathbb{F}^{n \times p} &\rightarrow \mathbb{F}^{m \times p}, \\ X &\mapsto AX. \end{aligned}$$

类似地, 右乘矩阵  $A$  也定义了一个线性映射:

$$\begin{aligned} \mathbf{R}_A: \mathbb{F}^{l \times m} &\rightarrow \mathbb{F}^{l \times n}, \\ X &\mapsto XA. \end{aligned}$$

2. 转置:

$$\begin{aligned} \cdot^T: \mathbb{F}^{m \times n} &\rightarrow \mathbb{F}^{n \times m}, \\ A &\mapsto A^T. \end{aligned}$$

3. 取迹:

$$\begin{aligned} \text{trace}: \mathbb{F}^{n \times n} &\rightarrow \mathbb{F}, \\ A &\mapsto \text{trace}(A). \end{aligned}$$

4. 函数在若干给定点的取值:

$$\begin{aligned} C(\mathbb{R}) &\rightarrow \mathbb{R}^n, \\ f &\mapsto \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}. \end{aligned}$$

☺

与  $\mathbb{R}^n$  到  $\mathbb{R}^m$  的线性映射类似, 也可以考虑推广的线性映射的运算.

**定义 7.3.3 (线性运算)** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{U}, \mathcal{V}$ ,  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射全体是  $\text{Hom}(\mathcal{U}, \mathcal{V})$ . 规定

1.  $\text{Hom}(\mathcal{U}, \mathcal{V})$  上的**加法**: 给定  $f, g \in \text{Hom}(\mathcal{U}, \mathcal{V})$ , 定义

$$\begin{aligned} f + g: \mathcal{U} &\rightarrow \mathcal{V}, \\ \mathbf{x} &\mapsto f(\mathbf{x}) + g(\mathbf{x}), \end{aligned}$$

2.  $\text{Hom}(\mathcal{U}, \mathcal{V})$  上的**数乘**: 给定  $f \in \text{Hom}(\mathcal{U}, \mathcal{V})$ , 定义

$$\begin{aligned} kf: \mathcal{U} &\rightarrow \mathcal{V}, \\ \mathbf{x} &\mapsto kf(\mathbf{x}). \end{aligned}$$

**命题 7.3.4** 集合  $\text{Hom}(\mathcal{U}, \mathcal{V})$  关于加法和数乘两种运算构成线性空间.

证明留给读者.

**定义 7.3.5 (乘法 (复合))** 给定数域  $\mathbb{F}$  线性空间  $\mathcal{U}, \mathcal{V}, \mathcal{W}$ , 若  $f \in \text{Hom}(\mathcal{U}, \mathcal{V}), g \in \text{Hom}(\mathcal{V}, \mathcal{W})$ , 则定义  $f$  与  $g$  的**复合**为

$$\begin{aligned} g \circ f: \mathcal{U} &\rightarrow \mathcal{W}, \\ x &\mapsto g(f(x)), \end{aligned}$$

$f$  与  $g$  的复合运算又称为  $g$  与  $f$  的**乘法**, 记为  $gf$ .

线性映射的乘积 (复合) 仍然是线性映射. 从已知的线性映射出发, 通过线性运算和乘法, 能得到更多的线性映射. 容易验证, 线性映射的运算法则和矩阵的运算法则类似.

我们知道, 矩阵能够给出坐标向量空间之间的线性映射, 它有零空间和列空间. 如下是对应于线性映射的类似概念.

**定义 7.3.6 (核、像集)** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{U}, \mathcal{V}$ , 以及  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$ . 则集合  $\mathcal{N}(f) := \{\mathbf{a} \in \mathcal{U} \mid f(\mathbf{a}) = \mathbf{0}\}$ , 称为线性映射  $f$  的**核**; 集合  $\mathcal{R}(f) := \{f(\mathbf{a}) \mid \mathbf{a} \in \mathcal{U}\}$ , 称为线性映射  $f$  的**像集**.

类似于矩阵与对应子空间的关系, 也通过核和像集来分析线性映射与对应集合的关系.

**命题 7.3.7** 对  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$ ,

1.  $\mathcal{N}(f)$  是  $\mathcal{U}$  的子空间, 而且  $f$  是单射当且仅当  $\mathcal{N}(f) = \{\mathbf{0}\}$ ;
2.  $\mathcal{R}(f)$  是  $\mathcal{V}$  的子空间, 而且  $f$  是满射当且仅当  $\mathcal{R}(f) = \mathcal{V}$ .

证明留给读者.

**定义 7.3.8 (线性变换)** 线性空间  $\mathcal{U}$  到自身的线性映射称为  $\mathcal{U}$  上的**线性变换**.

注意,  $\mathcal{U}$  上任意两个线性变换都可以复合, 得到的还是  $\mathcal{U}$  上的线性变换.

类似于  $\mathbb{F}^n$  上的线性变换有特征值和特征向量, 这两个概念也可以推广到一般线性空间的线性变换上.

**定义 7.3.9 (特征值)** 给定  $\mathbb{F}$  上的线性空间  $\mathcal{U}$ , 以及其上的线性变换  $f$ . 如果对  $\lambda \in \mathbb{F}$ , 存在非零向量  $\mathbf{x} \in \mathcal{U}$ , 使得  $f(\mathbf{x}) = \lambda\mathbf{x}$ , 则称  $\lambda$  为线性变换  $f$  的一个**特征值**, 而称非零向量  $\mathbf{x}$  为  $f$  的一个属于特征值  $\lambda$  的**特征向量**.

二元组  $(\lambda, \mathbf{x})$  常称为线性变换  $f$  的一个**特征对**.

易见, 非零向量  $\boldsymbol{x}$  为  $f$  的属于特征值  $\lambda$  的特征向量, 当且仅当  $\boldsymbol{x} \in \mathcal{N}(\lambda \boldsymbol{I} - f)$ . 子空间  $\mathcal{N}(\lambda \boldsymbol{I} - f)$  称为  $f$  的属于特征值  $\lambda$  的**特征子空间**, 通常记为  $\mathcal{U}_\lambda$ . 不难证明如下与命题 5.3.3 类似的结论, 留给读者.

**命题 7.3.10** 对  $\mathcal{U}$  上的线性变换  $f$ , 属于不同特征值的特征向量线性无关.

**例 7.3.11** 1. 设  $A \in \mathbb{F}^{n \times n}$ , 取左乘矩阵和右乘矩阵映射的差:

$$\begin{aligned} \boldsymbol{L}_A - \boldsymbol{R}_A: \mathbb{F}^{n \times n} &\rightarrow \mathbb{F}^{n \times n}, \\ X &\mapsto AX - XA, \end{aligned}$$

定义了  $\mathbb{F}^{n \times n}$  上的线性变换. 其核  $\mathcal{N}(\boldsymbol{L}_A - \boldsymbol{R}_A) = \{X \in \mathbb{F}^{n \times n} \mid AX = XA\}$  是  $\mathbb{F}^{n \times n}$  的子空间, 包含了所有与  $A$  交换的  $n$  阶方阵.

2. 恒同变换和转置的线性组合定义了  $\mathbb{F}^{n \times n}$  上的线性变换:

$$\begin{aligned} \boldsymbol{S}: \mathbb{F}^{n \times n} &\rightarrow \mathbb{F}^{n \times n}, \\ A &\mapsto \frac{1}{2}(A + A^T). \end{aligned}$$

核  $\mathcal{N}(\boldsymbol{S})$  是全体  $n$  阶反对称矩阵构成的子空间; 像集  $\mathcal{R}(\boldsymbol{S})$  是全体  $n$  阶对称矩阵构成的子空间.

注意,  $\boldsymbol{S}^2 = \boldsymbol{S}$ , 因此其特征值只能是 1, 0 (为什么?). 两个特征子空间

$$\begin{aligned} \mathcal{U}_{\lambda=1} &= \{A \in \mathbb{F}^{n \times n} \mid \boldsymbol{S}(A) = A\} = \mathcal{R}(\boldsymbol{S}), \\ \mathcal{U}_{\lambda=0} &= \{A \in \mathbb{F}^{n \times n} \mid \boldsymbol{S}(A) = O\} = \mathcal{N}(\boldsymbol{S}). \end{aligned}$$

3. 在  $\mathbb{R}$  上的光滑函数空间  $C^\infty(\mathbb{R})$  上, 求导运算定义了线性变换:

$$\begin{aligned} \boldsymbol{D} = \frac{d}{dx}: C^\infty(\mathbb{R}) &\rightarrow C^\infty(\mathbb{R}), \\ f(x) &\mapsto f'(x), \end{aligned}$$

称为求导算子. 核  $\mathcal{N}(\boldsymbol{D})$  是常数函数的全体构成的线性空间.

再看  $\boldsymbol{D}$  的特征值和特征向量. 对任意常数  $\lambda \in \mathbb{R}$ , 考虑线性变换:

$$\begin{aligned} \lambda \boldsymbol{I} - \boldsymbol{D}: C^\infty(\mathbb{R}) &\rightarrow C^\infty(\mathbb{R}), \\ f(x) &\mapsto \lambda f(x) - f'(x), \end{aligned}$$

容易验证, 指数函数  $e^{\lambda x} \in \mathcal{N}(\lambda \boldsymbol{I} - \boldsymbol{D})$ . 事实上, 利用常微分方程知识,  $(\lambda \boldsymbol{I} - \boldsymbol{D})f = \lambda f - f' = 0$  的解只能是  $f(x) = ke^{\lambda x}$ . 因此  $\mathcal{N}(\lambda \boldsymbol{I} - \boldsymbol{D})$  是一维线性空间, 而  $e^{\lambda x}$  是一组基.

4. 求导算子和自身的乘积  $D^2: f(x) \mapsto f''(x)$  是  $C^\infty(\mathbb{R})$  上的线性变换. 设  $\lambda \in \mathbb{R}$ , 考虑线性变换  $\lambda I - D^2: C^\infty(\mathbb{R}) \rightarrow C^\infty(\mathbb{R})$ .

利用常微分方程知识, 可得  $\mathcal{N}(\lambda I - D^2)$  是二维线性空间, 而一组基是

$$\begin{cases} e^{\sqrt{\lambda}x}, & e^{-\sqrt{\lambda}x}, & \lambda > 0; \\ \cos(\sqrt{-\lambda}x), & \sin(\sqrt{-\lambda}x), & \lambda < 0; \\ 1, x, & & \lambda = 0. \end{cases}$$

☺

下面来看一种特殊的线性映射.

**定义 7.3.12 (线性空间的同构)** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{U}, \mathcal{V}$ , 如果存在  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$  是双射, 则称  $\mathcal{U}$  和  $\mathcal{V}$  **同构**, 称  $f$  为  $\mathcal{U}$  到  $\mathcal{V}$  的**同构映射**.

特别地,  $\mathcal{U}$  到  $\mathcal{U}$  的同构映射称为  $\mathcal{U}$  上的**自同构**.

**例 7.3.13** 定义  $\mathcal{V} := \left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} : a, b \in \mathbb{R} \right\}$ . 易见  $\mathcal{V}$  是  $\mathbb{R}^{2 \times 2}$  的子空间. 可以验证, 映射

$$\begin{aligned} f: \quad \mathbb{C} &\rightarrow \mathcal{V}, \\ a + bi &\mapsto \begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \end{aligned}$$

是一个同构映射.

☺

**例 7.3.14** 给定可逆矩阵  $T \in \mathbb{F}^{n \times n}$ , 那么映射

$$\begin{aligned} f: \quad \mathbb{F}^n &\rightarrow \mathbb{F}^n, \\ \mathbf{x} &\mapsto T\mathbf{x}, \end{aligned}$$

是一个  $\mathbb{F}^n$  上的自同构.

☺

同构映射是双射, 因此它有逆映射.

**命题 7.3.15** 给定数域  $\mathbb{F}$  上的线性空间  $\mathcal{U}, \mathcal{V}, \mathcal{W}$ , 以及  $\mathcal{U}$  到  $\mathcal{V}$  的同构映射  $f$ ,  $\mathcal{V}$  到  $\mathcal{W}$  的同构映射  $g$ , 则

1.  $g \circ f$  是  $\mathcal{U}$  到  $\mathcal{W}$  的同构映射;
2.  $f^{-1}$  是  $\mathcal{V}$  到  $\mathcal{U}$  的同构映射.

证. 第 1 条: 显然.

第2条: 显然  $f^{-1}$  是双射, 只需验证它是线性映射. 由于  $f$  是线性映射, 因此对任意  $x_1, x_2 \in \mathcal{V}$ ,  $f(k_1 f^{-1}(x_1) + k_2 f^{-1}(x_2)) = k_1 f(f^{-1}(x_1)) + k_2 f(f^{-1}(x_2)) = k_1 x_1 + k_2 x_2$ . 由于  $f$  是双射, 有  $k_1 f^{-1}(x_1) + k_2 f^{-1}(x_2) = f^{-1}(k_1 x_1 + k_2 x_2)$ . 因此  $f^{-1}$  是线性映射.  $\square$

利用命题 7.3.15 可得如下结论.

**命题 7.3.16** 线性空间的同构关系是等价关系.

那么, 这个等价关系的不变量和标准形是什么呢? 下节将分析有限维线性空间的情形.

## 习题

**练习 7.3.1** 证明命题 7.3.4.

**练习 7.3.2** 证明命题 7.3.7.

**练习 7.3.3** 证明命题 7.3.10.

**练习 7.3.4** 给定  $a \in \mathbb{F}$ , 判断下面定义的  $\mathbb{F}[x]$  上的变换  $T_a$  是否是线性变换:

$$T_a(f(x)) = f(x+a), \quad \forall f(x) \in \mathbb{F}[x].$$

**练习 7.3.5** 在光滑函数空间  $C^\infty(\mathbb{R})$  上定义变换:  $A(f(x)) = (f'(x))^2$ . 判断  $A$  是否是线性变换.

**练习 7.3.6** 考虑练习 7.1.1 中的实线性空间  $\mathbb{R}^+$ , 给定  $a > 0$ , 判断映射

$$\begin{aligned} \log_a: \mathbb{R}^+ &\rightarrow \mathbb{R}, \\ x &\mapsto \log_a x, \end{aligned}$$

是否是线性映射. 如果是, 进一步分析当  $a$  取何值时, 该映射是同构.

**练习 7.3.7** 计算例 7.3.2 中线性映射的核与像集, 并求二者的维数.

**练习 7.3.8** 定义  $\mathbb{F}[x]$  上的变换:  $A(f(x)) = xf(x), \forall f(x) \in \mathbb{F}[x]$ .

1. 证明  $A$  是  $\mathbb{F}[x]$  上的一个线性变换.
2. 设  $D$  是求导算子, 证明  $DA - AD = I$ .

**练习 7.3.9** 设  $f$  是线性空间  $\mathcal{V}$  上的线性变换,  $a \in \mathcal{V}$ . 证明, 如果存在正整数  $m$ , 使得  $f^{m-1}(a) \neq 0, f^m(a) = 0$ , 则  $a, f(a), \dots, f^{m-1}(a)$  线性无关. 由此推出  $\dim \mathcal{V} \geq m$ .

**练习 7.3.10** 设线性空间  $\mathcal{V}$  有直和分解:  $\mathcal{V} = \mathcal{M}_1 \oplus \mathcal{M}_2$ . 任取  $a \in \mathcal{V}$ , 都有唯一的分解式:  $a = a_1 + a_2$ , 其中  $a_1 \in \mathcal{M}_1, a_2 \in \mathcal{M}_2$ . 定义  $\mathcal{V}$  上的变换:

$$P_{\mathcal{M}_1}(a) = a_1, \quad P_{\mathcal{M}_2}(a) = a_2.$$

1. 证明,  $P_{\mathcal{M}_1}, P_{\mathcal{M}_2}$  都是  $\mathcal{V}$  上的线性变换. 称  $P_{\mathcal{M}_1}$  为平行于  $P_{\mathcal{M}_1}$  在  $P_{\mathcal{M}_2}$  上的投影变换.
2. 证明,  $\mathcal{N}(P_{\mathcal{M}_1}) = \mathcal{M}_2, \mathcal{R}(P_{\mathcal{M}_1}) = \mathcal{M}_1$ .
3. 证明,  $P_{\mathcal{M}_1}^2 = P_{\mathcal{M}_1}, P_{\mathcal{M}_1} + P_{\mathcal{M}_2} = I, P_{\mathcal{M}_1} P_{\mathcal{M}_2} = O$ .
4. 分别求  $P_{\mathcal{M}_1}, P_{\mathcal{M}_2}$  的特征值和特征子空间.

**练习 7.3.11** 考虑例 7.3.11 中的线性变换  $\mathbb{F}^{n \times n}$  上的线性变换  $S$ , 试问它是否是投影变换.

**练习 7.3.12** 证明矩阵空间  $\mathbb{F}^{m \times n}$  与  $\mathbb{F}^{mn}$  同构.

**练习 7.3.13** 证明多项式空间  $\mathbb{F}[x]_n$  与  $\mathbb{F}^n$  同构.

**练习 7.3.14** 证明练习 7.1.2 中的线性空间  $\mathbb{Q}[\omega]$  与练习 7.1.4 中的线性空间  $\mathbb{Q}[i]$  同构.

## 7.4 向量的坐标表示

为简化讨论, 如果不加特别说明, 本节和下节所涉及的线性空间都是有限维线性空间.

在有限维线性空间  $\mathcal{V}$  中取一组基, 则  $\mathcal{V}$  中任意向量都可以被这组基线性表示. 另一方面, 类似于命题 2.1.15, 容易证明如下结论, 留给读者.

**命题 7.4.1** 向量组  $a_1, \dots, a_s$  线性无关, 如果  $b$  可以被其线性表示, 则表示法唯一.

因此向量与表示法之间有一个一一对应, 为此将表示法定义如下.

**定义 7.4.2 (坐标)** 对数域  $\mathbb{F}$  上的  $n$  维线性空间  $\mathcal{V}$ , 设  $e_1, \dots, e_n$  是它的一组基. 那么对任意向量  $x \in \mathcal{V}$ , 都有  $x$  可以被这组基线性表示且表示法唯一, 不妨写为  $x = x_1 e_1 + \dots + x_n e_n$ . 有序数组  $x_1, \dots, x_n$  称为向量  $x$  在基  $e_1, \dots, e_n$  下的坐标.

为书写简便, 我们把它写作  $x = (e_1, \dots, e_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ . 注意这并不是真正的矩阵乘法,

而只是借用了记号来表示线性组合<sup>3</sup>. 由表示法的唯一性, 可以定义映射:

$$\sigma = \sigma_{e_1, \dots, e_n} : \mathcal{V} \rightarrow \mathbb{F}^n,$$

$$x \mapsto \hat{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

<sup>3</sup>与前面注记类似, 为了与坐标向量空间中的向量相区别, 本书给该向量组外加圆括号.

映射  $\sigma_{e_1, \dots, e_n}$  就是把  $\mathcal{V}$  中向量映射成它在一组基  $e_1, \dots, e_n$  下的坐标组成的  $n$  维向量.

注意, 坐标表示写成  $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$  后, 基  $e_1, \dots, e_n$  中的向量就不再能随便改换顺序, 因为把

基向量改换顺序将引入不同的映射  $\sigma$ , 而坐标表示也不同. 换言之, 改换顺序的基被认为是不同的基.

**定理 7.4.3** 映射  $\sigma_{e_1, \dots, e_n} : \mathcal{V} \rightarrow \mathbb{F}^n$  是同构映射.

证. 如果  $\mathbf{x} = x_1 e_1 + \dots + x_n e_n, \mathbf{y} = y_1 e_1 + \dots + y_n e_n$ , 则  $\mathbf{x} + \mathbf{y} = (x_1 + y_1)e_1 + \dots + (x_n + y_n)e_n, k\mathbf{x} = (kx_1)e_1 + \dots + (kx_n)e_n$ . 因此,  $\sigma(\mathbf{x} + \mathbf{y}) = \sigma(\mathbf{x}) + \sigma(\mathbf{y}), \sigma(k\mathbf{x}) = k\sigma(\mathbf{x})$ . 于是  $\sigma$  是线性映射. 根据基的定义,  $\sigma$  是双射.  $\square$

同构映射  $\sigma$  的逆映射

$$\sigma_{e_1, \dots, e_n}^{-1} : \mathbb{F}^n \rightarrow \mathcal{V},$$

$$\widehat{\mathbf{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto \mathbf{x} = (e_1, \dots, e_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

就是在形式上推广的矩阵乘法. 它在形式上满足分配律:

$$(e_1, \dots, e_n)(\widehat{\mathbf{x}} + \widehat{\mathbf{y}}) = (e_1, \dots, e_n)\widehat{\mathbf{x}} + (e_1, \dots, e_n)\widehat{\mathbf{y}},$$

$$(e_1, \dots, e_n)(k\widehat{\mathbf{x}}) = k(e_1, \dots, e_n)\widehat{\mathbf{x}}.$$

由定理 7.4.3 可知,  $\mathbb{F}$  上的任意  $n$  维线性空间都同构: 因为它们都和  $\mathbb{F}^n$  同构, 而同构又是等价关系. 线性空间的同构这个等价关系的唯一不变量就是维数, 而标准形是  $\mathbb{F}^n$ . 例如, 例 7.3.13 中的两个线性空间都和  $\mathbb{R}^2$  同构, 因此也互相同构.

同构映射能够保持线性运算, 从而保持一切只与线性运算有关的性质. 因为, 常常可以把一般向量空间中的问题转化到  $\mathbb{F}^n$  上, 例如如下结论.

**命题 7.4.4** 在数域  $\mathbb{F}$  上  $n$  维线性空间  $\mathcal{V}$  中取定一组基  $e_1, \dots, e_n$ , 设  $\mathcal{V}$  内向量组  $\mathbf{a}_1, \dots, \mathbf{a}_m$  在这组基下的坐标表示为  $\mathbf{a}_i = (e_1, \dots, e_n)\widehat{\mathbf{a}}_i, i = 1, \dots, m$ , 则  $\mathbf{a}_1, \dots, \mathbf{a}_m$  在  $\mathcal{V}$  内线性无关, 当且仅当  $\widehat{\mathbf{a}}_1, \dots, \widehat{\mathbf{a}}_m$  在  $\mathbb{F}^n$  内线性无关.

证.  $\mathbf{a}_1, \dots, \mathbf{a}_m$  在  $\mathcal{V}$  内线性相关  $\Leftrightarrow$  存在  $\mathbb{F}$  内不全为 0 的数  $k_1, \dots, k_m$ , 使得  $k_1 \mathbf{a}_1 + \dots + k_m \mathbf{a}_m = \mathbf{0} \Leftrightarrow$  存在  $\mathbb{F}$  内不全为 0 的数  $k_1, \dots, k_m$ , 使得  $(e_1, \dots, e_n)(k_1 \widehat{\mathbf{a}}_1 + \dots + k_m \widehat{\mathbf{a}}_m) = \mathbf{0}$   
 $\xLeftrightarrow[\sigma \text{ 是双射}] \text{ 存在 } \mathbb{F} \text{ 内不全为 } 0 \text{ 的数 } k_1, \dots, k_m, \text{ 使得 } k_1 \widehat{\mathbf{a}}_1 + \dots + k_m \widehat{\mathbf{a}}_m = \mathbf{0} \Leftrightarrow \widehat{\mathbf{a}}_1, \dots, \widehat{\mathbf{a}}_m$   
 在  $\mathbb{F}^n$  内线性相关.  $\square$



**例 7.4.5** 1. 考虑例 7.2.12 中矩阵空间  $\mathbb{F}^{2 \times 2}$  的一组基

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

矩阵  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}E_{11} + a_{12}E_{12} + a_{21}E_{21} + a_{22}E_{22}$ , 即  $A$  在这组基下的坐

标表示为  $A = (E_{11}, E_{12}, E_{21}, E_{22}) \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}$ , 故  $A$  在这组基下的坐标是  $\mathbf{a} = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}$ .

再考虑例 7.2.19 中线性空间  $\mathbb{F}^{2 \times 2}$  的一组基

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

矩阵  $A$  在这组基下的坐标表示为  $A = (E_{11}, E_{22}, E_{12} + E_{21}, E_{12} - E_{21})\tilde{\mathbf{a}}$ , 其中

$\tilde{\mathbf{a}} = \begin{bmatrix} a_{11} \\ a_{22} \\ \frac{a_{12}+a_{21}}{2} \\ \frac{a_{12}-a_{21}}{2} \end{bmatrix}$  是  $A$  在这组基下的坐标.

2. 由例 7.2.13 可知,  $1, x, \dots, x^{n-1}$  是多项式空间  $\mathbb{R}[x]_n$  的一组基. 由二项式展开  $x^k = (x - x_0 + x_0)^k = \sum_{i=0}^k C_k^i x_0^i (x - x_0)^{k-i}$  知,  $1, x, \dots, x^{n-1}$  可以被  $1, x - x_0, \dots, (x - x_0)^{n-1}$  线性表示, 因此后者的秩不小于前者的秩, 这说明  $1, x - x_0, \dots, (x - x_0)^{n-1}$  也是一组基.

对多项式  $f = a_0 + a_1x + \dots + a_{n-1}x^{n-1}$ , 它在基  $1, x, \dots, x^{n-1}$  下的坐标表示为

$f = (1, x, \dots, x^{n-1})\mathbf{f}$ , 其中  $\mathbf{f} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix}$  是  $f$  在这组基下的坐标. 另一方面, 根

据 Taylor 展开式, 它在基  $1, x - x_0, \dots, (x - x_0)^{n-1}$  下的坐标表示为  $f = (1, x -$

$x_0, \dots, (x - x_0)^{n-1})\tilde{\mathbf{f}}$ , 其中  $\tilde{\mathbf{f}} = \begin{bmatrix} f(x_0) \\ f'(x_0) \\ \vdots \\ \frac{f^{(n-1)}(x_0)}{(n-1)!} \end{bmatrix}$  是  $f$  在这组基下的坐标.  $\odot$

可以看到, 同一个向量在不同基下的坐标表示不同. 那么这些坐标表示有什么不同?

既然这些坐标表示对应同一个向量, 那么它们总会有些共同点, 应该如何把这些坐标表示联系起来?

首先来看两组基之间的关系. 给定  $n$  维线性空间的两组基  $e_1, \dots, e_n$  和  $t_1, \dots, t_n$ , 每个  $t_i$  都能被  $e_1, \dots, e_n$  线性表示, 设

$$\begin{aligned} t_1 &= t_{11}e_1 + \dots + t_{n1}e_n, \\ &\vdots \\ t_n &= t_{1n}e_1 + \dots + t_{nn}e_n, \end{aligned}$$

那么就可以形式地写成

$$(t_1, \dots, t_n) = (e_1, \dots, e_n) \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} =: (e_1, \dots, e_n)T,$$

其中  $T$  称为从基  $e_1, \dots, e_n$  到基  $t_1, \dots, t_n$  的过渡矩阵.

**命题 7.4.6** 给定数域  $\mathbb{F}$  上  $n$  维线性空间  $\mathcal{V}$  的一组基  $e_1, \dots, e_n$ , 和  $\mathbb{F}$  上  $n$  阶方阵  $T$ . 令  $(t_1, \dots, t_n) = (e_1, \dots, e_n)T$ , 则有:

1. 如果  $t_1, \dots, t_n$  是一组基, 则  $T$  可逆;
2. 如果  $T$  可逆, 则  $t_1, \dots, t_n$  是一组基.

证. 设  $T = [\hat{t}_1 \ \cdots \ \hat{t}_n]$ , 则  $t_i = (e_1, \dots, e_n)\hat{t}_i$ . 因此:  $t_1, \dots, t_n$  是一组基  $\Leftrightarrow t_1, \dots, t_n$  线性无关  $\xLeftrightarrow{\text{命题 7.4.4}} \hat{t}_1, \dots, \hat{t}_n$  线性无关  $\Leftrightarrow T$  满秩, 即  $T$  可逆.  $\square$

**例 7.4.7** 考虑例 7.4.5.

$$1. \text{ 易得 } (E_{11}, E_{22}, E_{12} + E_{21}, E_{12} - E_{21}) = (E_{11}, E_{12}, E_{21}, E_{22}) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \text{ 因}$$

此从  $E_{11}, E_{12}, E_{21}, E_{22}$  到  $E_{11}, E_{22}, E_{12} + E_{21}, E_{12} - E_{21}$  的过渡矩阵就是该四阶矩阵.

2. 易得  $(1, x - x_0, \dots, (x - x_0)^{n-1}) = (1, x, \dots, x^{n-1})T$ , 其中

$$T = \begin{bmatrix} 1 & -x_0 & x_0^2 & \cdots & (-x_0)^{n-1} \\ & 1 & -2x_0 & \cdots & (n-1)(-x_0)^{n-2} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & -(n-1)x_0 \\ & & & & 1 \end{bmatrix},$$

因此从  $1, x, \dots, x^{n-1}$  到  $1, x - x_0, \dots, (x - x_0)^{n-1}$  的过渡矩阵为上三角矩阵  $T$ .  $\odot$

其次来看同一个向量在两组基下坐标表示的变化规律. 设  $\mathbf{a} \in \mathcal{V}$  在一组基  $\mathbf{e}_1, \dots, \mathbf{e}_n$  下的坐标为  $x_1, \dots, x_n$ , 在另一组基  $\mathbf{t}_1, \dots, \mathbf{t}_n$  下的坐标表示为  $y_1, \dots, y_n$ , 那么坐标表示就是

$$\mathbf{a} = (\mathbf{e}_1, \dots, \mathbf{e}_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} =: (\mathbf{e}_1, \dots, \mathbf{e}_n) \mathbf{x}, \quad \mathbf{a} = (\mathbf{t}_1, \dots, \mathbf{t}_n) \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} =: (\mathbf{t}_1, \dots, \mathbf{t}_n) \mathbf{y}.$$

设两组基之间的过渡矩阵为  $T$ , 即  $(\mathbf{t}_1, \dots, \mathbf{t}_n) = (\mathbf{e}_1, \dots, \mathbf{e}_n)T$ , 则

$$\mathbf{a} = (\mathbf{t}_1, \dots, \mathbf{t}_n) \mathbf{y} = ((\mathbf{e}_1, \dots, \mathbf{e}_n)T) \mathbf{y} = (\mathbf{e}_1, \dots, \mathbf{e}_n)(T\mathbf{y}). \quad (7.4.1)$$

而  $\mathbf{a}$  在基  $\mathbf{e}_1, \dots, \mathbf{e}_n$  下的表示法唯一, 因此  $\mathbf{x} = T\mathbf{y}$ . 可见, 只要知道了向量在一组基下的坐标和这组基到另一组基的过渡矩阵, 就能得到该向量在另一组基下的坐标.

注意, (7.4.1) 还提示了坐标表示在形式上满足乘法结合律.

**例 7.4.8** 考虑例 7.4.5 和 7.4.7.

1. 容易验证

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{22} \\ \frac{a_{12}+a_{21}}{2} \\ \frac{a_{12}-a_{21}}{2} \end{bmatrix}.$$

2. 可以验证

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-2} \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & -x_0 & x_0^2 & \cdots & (-x_0)^{n-1} \\ & 1 & -2x_0 & \cdots & (n-1)(-x_0)^{n-2} \\ & & \ddots & \ddots & \vdots \\ & & & 1 & -(n-1)x_0 \\ & & & & 1 \end{bmatrix} \begin{bmatrix} f(x_0) \\ f'(x_0) \\ \vdots \\ \frac{1}{(n-2)!} f^{(n-2)}(x_0) \\ \frac{1}{(n-1)!} f^{(n-1)}(x_0) \end{bmatrix}.$$

这与上述坐标与过渡矩阵的关系相吻合.  $\odot$

上述内容总结如下:

1. 基变换公式:  $(\mathbf{t}_1, \dots, \mathbf{t}_n) = (\mathbf{e}_1, \dots, \mathbf{e}_n)T$ , 其中  $T$  可逆;
2. 坐标变换公式: 若  $\mathbf{a} = (\mathbf{t}_1, \dots, \mathbf{t}_n)\mathbf{y} = (\mathbf{e}_1, \dots, \mathbf{e}_n)\mathbf{x}$ , 则  $\mathbf{x} = T\mathbf{y}$ .

最后讨论  $\mathbb{F}^n$  中的基变换. 线性空间  $\mathbb{F}^n$  的一组基总能写成由其作为列向量组成的矩阵. 如果一组基组成矩阵  $A$ , 另一组基组成矩阵  $B$ , 两组基的过渡矩阵记为  $T$ , 那么就有  $B = AT$ , 因此  $T = A^{-1}B$ . 因此, 只要给定了两组基, 就能利用初等行变换求出二者的过渡矩阵:

$$\begin{bmatrix} A & B \end{bmatrix} \xrightarrow{\text{Gauss-Jordan}} \begin{bmatrix} I & A^{-1}B \end{bmatrix} = \begin{bmatrix} I & T \end{bmatrix},$$

即用相同的初等行变换作用在  $A, B$  上, 当  $A$  化成  $I$  时,  $B$  就化成了  $T$ .

即使  $\mathcal{V}$  只是  $\mathbb{F}^n$  的子空间, 且  $\dim \mathcal{V} = m < n$ , 也仍然能利用初等行变换, 求出过渡矩阵:

$$\begin{bmatrix} A & B \end{bmatrix} \xrightarrow{\text{Gauss-Jordan}} \begin{bmatrix} I & T \\ O & O \end{bmatrix},$$

即用相同的初等行变换作用在  $A, B$  上, 当  $A$  化成行简化阶梯形时,  $B$  化成的矩阵的前  $m$  行就是过渡矩阵  $T$  (为什么该矩阵的后  $n - m$  行都是零行?). 事实上, 易得  $T = A^+B = (A^T A)^{-1} A^T B$ .

## 习题

**练习 7.4.1** 证明命题 7.4.1.

**练习 7.4.2** 求  $\mathbb{F}^4$  中由基  $e_1, e_2, e_3, e_4$  到基  $t_1, t_2, t_3, t_4$  的过渡矩阵, 并分别求向量  $a$  在两组基下的坐标.

$$1. \quad e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, e_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix};$$

$$t_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 1 \end{bmatrix}, t_2 = \begin{bmatrix} 0 \\ 3 \\ 1 \\ 0 \end{bmatrix}, t_3 = \begin{bmatrix} 5 \\ 3 \\ 2 \\ 1 \end{bmatrix}, t_4 = \begin{bmatrix} 6 \\ 6 \\ 1 \\ 3 \end{bmatrix}; \quad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}.$$

$$2. \quad e_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, e_3 = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 1 \end{bmatrix}, e_4 = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix};$$

$$t_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, t_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 2 \end{bmatrix}, t_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \\ 2 \end{bmatrix}, t_4 = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 2 \end{bmatrix}; \quad a = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

**练习 7.4.3** 考虑函数空间的子空间  $\text{span}(\sin^2 x, \cos^2 x)$ .

1. 证明  $\sin^2 x, \cos^2 x$  和  $1, \cos 2x$  分别是子空间的一组基.
2. 分别求从  $\sin^2 x, \cos^2 x$  到  $1, \cos 2x$ , 和从  $1, \cos 2x$  到  $\sin^2 x, \cos^2 x$  的过渡矩阵.
3. 分别求  $1$  和  $\sin^2 x$  在两组基下的坐标.

**练习 7.4.4** 考虑练习 7.2.7 中的线性空间  $\mathbb{F}[x]_n$ , 考虑  $n=3$  的情形.

1. 给定  $\mathbb{F}$  中两两不等的数  $a_1, a_2, a_3$ , 求由基  $1, x, x^2$  到基  $f_1(x), f_2(x), f_3(x)$  的过渡矩阵.
2. 考虑练习 7.2.7 中的  $f(x)$ , 给定  $\mathbb{F}$  中任意三个数  $b_1, b_2, b_3$ , 求  $f(x)$  在两组基下的坐标.

**练习 7.4.5** 矩阵空间  $\mathbb{F}^{2 \times 2}$  有两组基  $e_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, e_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ , 和  $t_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, t_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, t_3 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, t_4 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ . 求从基  $e_1, e_2, e_3, e_4$  到基  $t_1, t_2, t_3, t_4$  的过渡矩阵.

**练习 7.4.6** 设  $e_1, e_2, e_3, e_4$  是线性空间  $\mathcal{V}$  的一组基, 求下列向量组的一个极大线性无关部分组:

$$\begin{aligned} t_1 &= e_1 + e_2 + e_3 + 3e_4, & t_2 &= -e_1 - 3e_2 + 5e_3 + e_4, \\ t_3 &= 3e_1 + 2e_2 - e_3 + 4e_4, & t_4 &= -2e_1 - 6e_2 + 10e_3 + 2e_4. \end{aligned}$$

**练习 7.4.7** 设  $e_1, e_2, \dots, e_n$  是线性空间  $\mathcal{V}$  的一组基.

1. 判断  $t_1 = e_1, t_2 = e_1 + e_2, \dots, t_n = e_1 + e_2 + \dots + e_n$  是否也是  $\mathcal{V}$  的一组基.
2. 判断  $t_1 = e_1 + e_2, t_2 = e_2 + e_3, \dots, t_n = e_n + e_1$  是否也是  $\mathcal{V}$  的一组基.

**练习 7.4.8** 设  $a_1, \dots, a_n$  是  $\mathbb{F}$  中两两不等的数,  $e_1, e_2, \dots, e_n$  是线性空间  $\mathcal{V}$  的一组基, 令  $t_i = e_1 + a_i e_2 + \dots + a_i^{n-1} e_n, i = 1, \dots, n$ . 证明  $t_1, t_2, \dots, t_n$  也是  $\mathcal{V}$  的一组基.

**练习 7.4.9** 设 (I):  $e_1, \dots, e_n$ 、(II):  $t_1, \dots, t_n$  和 (III):  $s_1, \dots, s_n$  是线性空间  $\mathcal{V}$  的三组基, 如果从 (I) 到 (II) 的过渡矩阵是  $P$ , 从 (II) 到 (III) 的过渡矩阵是  $Q$ , 证明,

1. 从 (II) 到 (I) 的过渡矩阵是  $P^{-1}$ .
2. 从 (I) 到 (III) 的过渡矩阵是  $PQ$ .

## 7.5 线性映射的矩阵表示

给定  $\mathbb{F}$  上  $n$  维线性空间  $\mathcal{V}$ , 只要选定一组基, 任意向量就都有坐标表示, 这坐标构成了一个  $\mathbb{F}^n$  中的向量, 而且  $\mathcal{V}$  中向量的线性运算就等价于  $\mathbb{F}^n$  中向量的线性运算, 即  $\mathcal{V}$  和  $\mathbb{F}^n$  同构.

第 1 章已经说明,  $\mathbb{F}^n$  到  $\mathbb{F}^m$  的线性映射能够用  $\mathbb{F}$  上  $m \times n$  矩阵来表示. 那么, 对任意线性映射, 在定义域和陪域分别取一组基后, 是否也有与向量的坐标类似的表示? 具体地,  $n$  维线性空间到  $m$  维线性空间的线性映射是否也可由  $m \times n$  矩阵刻画?

首先来说明只要知道基在线性映射下的像, 线性映射就已经唯一确定.

**命题 7.5.1** 给定数域  $\mathbb{F}$  上有限维线性空间  $\mathcal{U}, \mathcal{V}$ , 而  $\mathbf{e}_1, \dots, \mathbf{e}_n$  是  $\mathcal{U}$  的一组基, 则

1. 任意  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$  由  $\mathcal{U}$  的基  $\mathbf{e}_1, \dots, \mathbf{e}_n$  的像唯一确定, 亦即, 如果又有  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $g$  使得  $g(\mathbf{e}_i) = f(\mathbf{e}_i), i = 1, \dots, n$ , 则对任意  $\mathbf{a} \in \mathcal{U}$ ,  $g(\mathbf{a}) = f(\mathbf{a})$ , 即  $g = f$ ;
2. 对任意  $\mathcal{V}$  中  $n$  个向量  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , 必存在唯一的  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$ , 使得  $f(\mathbf{e}_i) = \mathbf{a}_i, i = 1, \dots, n$ .

证. 第 1 条: 对任意  $\mathbf{a} \in \mathcal{U}$ , 如果  $\mathbf{a} = a_1\mathbf{e}_1 + \dots + a_n\mathbf{e}_n$ , 那么  $f(\mathbf{a}) = a_1f(\mathbf{e}_1) + \dots + a_nf(\mathbf{e}_n) = a_1g(\mathbf{e}_1) + \dots + a_ng(\mathbf{e}_n) = g(\mathbf{a})$ .

第 2 条: 先定义映射

$$\begin{aligned} f: \quad \mathcal{U} &\rightarrow \mathcal{V}, \\ x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n &\mapsto x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n. \end{aligned}$$

只需证明它是线性映射, 因为显然  $f(\mathbf{e}_i) = \mathbf{a}_i, i = 1, \dots, n$ , 而唯一性由第 1 条保证. 对向量  $\mathbf{x} = x_1\mathbf{e}_1 + \dots + x_n\mathbf{e}_n, \mathbf{y} = y_1\mathbf{e}_1 + \dots + y_n\mathbf{e}_n$ ,

$$\begin{aligned} f(k\mathbf{x} + l\mathbf{y}) &= f((kx_1 + ly_1)\mathbf{e}_1 + \dots + (kx_n + ly_n)\mathbf{e}_n) \\ &= (kx_1 + ly_1)\mathbf{a}_1 + \dots + (kx_n + ly_n)\mathbf{a}_n \\ &= k(x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n) + l(y_1\mathbf{a}_1 + \dots + y_n\mathbf{a}_n) \\ &= kf(\mathbf{x}) + lf(\mathbf{y}). \end{aligned}$$

□

根据命题 7.5.1,  $f$  被  $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$  唯一确定. 设  $\mathbf{i}_1, \dots, \mathbf{i}_m$  是  $\mathcal{V}$  的一组基, 则  $f(\mathbf{e}_i)$  可以被  $\mathbf{i}_1, \dots, \mathbf{i}_m$  线性表示. 设

$$\begin{aligned} f(\mathbf{e}_1) &= f_{11}\mathbf{i}_1 + \dots + f_{m1}\mathbf{i}_m, \\ &\vdots \\ f(\mathbf{e}_n) &= f_{1n}\mathbf{i}_1 + \dots + f_{mn}\mathbf{i}_m, \end{aligned}$$

令

$$F = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & & \vdots \\ f_{m1} & \cdots & f_{mn} \end{bmatrix},$$

则可形式上写成

$$(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)) = (\mathbf{i}_1, \dots, \mathbf{i}_m)F.$$

矩阵  $F$  称为线性映射  $f$  在两组给定基下的 (表示) 矩阵. 为简便起见, 引入记号

$$f(e_1, \dots, e_n) := (f(e_1), \dots, f(e_n)),$$

因此  $f(e_1, \dots, e_n) = (i_1, \dots, i_m)F$ .

**例 7.5.2** 下面列出一些线性映射的矩阵, 其中部分来自例 7.3.2.

1. 矩阵  $E_{j1}, E_{j2}, j = 1, \dots, n$  是  $\mathbb{F}^{n \times 2}$  的一组基,  $F_{i1}, F_{i2}, i = 1, \dots, m$  是  $\mathbb{F}^{m \times 2}$  的一组基. 设  $A = [a_{ij}] \in \mathbb{F}^{m \times n}$ , 并记  $\mathbf{a}_j$  为其第  $j$  列, 考虑线性映射

$$\begin{aligned} \mathbf{L}_A: \mathbb{F}^{n \times 2} &\rightarrow \mathbb{F}^{m \times 2}, \\ X &\mapsto AX. \end{aligned}$$

则基向量  $E_{j1}$  的像为

$$\begin{aligned} \mathbf{L}_A(E_{j1}) &= AE_{j1} = A \begin{bmatrix} e_j & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_j & \mathbf{0} \end{bmatrix} = a_{1j}F_{11} + \dots + a_{mj}F_{m1} \\ &= (F_{11}, \dots, F_{m1}, F_{12}, \dots, F_{m2}) \begin{bmatrix} \mathbf{a}_j \\ \mathbf{0} \end{bmatrix}, \end{aligned}$$

即基向量  $E_{j1}$  的像在基  $F_{11}, \dots, F_{m1}, F_{12}, \dots, F_{m2}$  下的坐标为  $\begin{bmatrix} \mathbf{a}_j \\ \mathbf{0} \end{bmatrix}$ . 类似地, 则基

向量  $E_{j2}$  的像在基  $F_{11}, \dots, F_{m1}, F_{12}, \dots, F_{m2}$  下的坐标为  $\begin{bmatrix} \mathbf{0} \\ \mathbf{a}_j \end{bmatrix}$ . 因此线性映射  $\mathbf{L}_A$  在这两组基下的矩阵表示为

$$\begin{aligned} \mathbf{L}_A(E_{11}, \dots, E_{n1}, E_{12}, \dots, E_{n2}) &= (\mathbf{L}_A(E_{11}), \dots, \mathbf{L}_A(E_{n1}), \mathbf{L}_A(E_{12}), \dots, \mathbf{L}_A(E_{n2})) \\ &= (F_{11}, \dots, F_{m1}, F_{12}, \dots, F_{m2}) \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix} \\ &= (F_{11}, \dots, F_{m1}, F_{12}, \dots, F_{m2}) \begin{bmatrix} A & \\ & A \end{bmatrix}, \end{aligned}$$

即  $\mathbf{L}_A$  在这两组基下的矩阵为  $L_A = \begin{bmatrix} A & \\ & A \end{bmatrix}$ .

2. 矩阵  $E_{1j}, E_{2j}, j = 1, \dots, n$  是  $\mathbb{F}^{2 \times n}$  的一组基,  $F_{1i}, F_{2i}, i = 1, \dots, m$  是  $\mathbb{F}^{2 \times m}$  的一组基. 设  $A = [a_{ij}] \in \mathbb{F}^{m \times n}$ , 并记  $\tilde{\mathbf{a}}_i^T$  为其第  $i$  行, 考虑线性映射

$$\begin{aligned} \mathbf{R}_A: \mathbb{F}^{2 \times m} &\rightarrow \mathbb{F}^{2 \times n}, \\ X &\mapsto XA. \end{aligned}$$

则基向量  $F_{1i}$  的像为

$$\begin{aligned} \mathbf{R}_A(F_{1i}) &= F_{1i}A = \begin{bmatrix} \mathbf{e}_i^T & \mathbf{0} \end{bmatrix} A = \begin{bmatrix} \tilde{\mathbf{a}}_i^T & \mathbf{0} \end{bmatrix} = a_{i1}E_{11} + \cdots + a_{in}E_{1n} \\ &= (E_{11}, \cdots, E_{1n}, E_{21}, \cdots, E_{2n}) \begin{bmatrix} \tilde{\mathbf{a}}_i \\ \mathbf{0} \end{bmatrix}, \end{aligned}$$

即基向量  $F_{1i}$  的像在基  $E_{11}, \cdots, E_{1n}, E_{21}, \cdots, E_{2n}$  下的坐标为  $\begin{bmatrix} \tilde{\mathbf{a}}_i \\ \mathbf{0} \end{bmatrix}$ . 类似地, 则基

向量  $F_{2i}$  的像在基  $E_{11}, \cdots, E_{1n}, E_{21}, \cdots, E_{2n}$  下的坐标为  $\begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{a}}_i \end{bmatrix}$ . 因此线性映射  $\mathbf{R}_A$  在这两组基下的矩阵表示为

$$\begin{aligned} \mathbf{R}_A(F_{11}, \cdots, F_{1m}, F_{21}, \cdots, F_{2m}) &= (\mathbf{R}_A(F_{11}), \cdots, \mathbf{L}_A(F_{1m}), \mathbf{L}_A(F_{21}), \cdots, \mathbf{L}_A(F_{2m})) \\ &= (E_{11}, \cdots, E_{1m}, E_{21}, \cdots, E_{2n}) \begin{bmatrix} \tilde{\mathbf{a}}_1 & \cdots & \tilde{\mathbf{a}}_n & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{a}}_1 & \cdots & \tilde{\mathbf{a}}_n \end{bmatrix} \\ &= (E_{11}, \cdots, E_{1n}, E_{21}, \cdots, E_{2n}) \begin{bmatrix} A^T & \\ & A^T \end{bmatrix}, \end{aligned}$$

即  $\mathbf{R}_A$  在这两组基下的矩阵为  $R_A = \begin{bmatrix} A^T & \\ & A^T \end{bmatrix}$ .

3. 在光滑函数空间的子空间  $\mathcal{V} = \text{span}(1, \sin x, \cos x)$  上考虑线性映射  $\mathbf{D}: f \mapsto f'$ . 首先易得  $1, \sin x, \cos x$  是一组基. 注意到  $\mathbf{D}1 = 0, \mathbf{D} \sin x = \cos x, \mathbf{D} \cos x = -\sin x$ , 可知  $\mathbf{D}$  可被定义为  $\mathcal{V}$  上的线性变换.

$$\text{于是 } \mathbf{D}(1, \sin x, \cos x) = (\mathbf{D}1, \mathbf{D} \sin x, \mathbf{D} \cos x) = (1, \sin x, \cos x) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \text{ 线}$$

性变换  $\mathbf{D}$  在给定基下的矩阵就是该三阶矩阵. ☺

定义映射

$$\begin{aligned} \sigma = \sigma_{\mathbf{e}_1, \cdots, \mathbf{e}_n; \mathbf{i}_1, \cdots, \mathbf{i}_m} : \text{Hom}(\mathcal{U}, \mathcal{V}) &\rightarrow \mathbb{F}^{m \times n}, \\ f &\mapsto F, \end{aligned}$$

其中,  $F$  是  $f \in \text{Hom}(\mathcal{U}, \mathcal{V})$  在给定基下的矩阵. 类似于线性空间与  $\mathbb{F}^n$  的同构, 可有如下结论.



**定理 7.5.3** 映射  $\sigma = \sigma_{e_1, \dots, e_n; i_1, \dots, i_m}$  是  $\text{Hom}(\mathcal{U}, \mathcal{V})$  到  $\mathbb{F}^{m \times n}$  的同构映射. 特别地,  $\dim \text{Hom}(\mathcal{U}, \mathcal{V}) = \dim \mathbb{F}^{m \times n} = mn$ .

证. 首先说明  $\sigma$  是双射. 命题 7.5.1 第 1 条说明不同线性映射的矩阵不同, 即  $\sigma$  是单射; 命题 7.5.1 第 2 条说明对任意矩阵都可以找到一个线性映射使其矩阵就是给定矩阵, 即  $\sigma$  是满射. 其次说明  $\sigma$  是线性映射. 设  $\sigma(f) = F, \sigma(g) = G$ , 则计算有

$$\begin{aligned} (kf + lg)(e_1, \dots, e_n) &= (kf(e_1) + lg(e_1), \dots, kf(e_n) + lg(e_n)) \\ &= k(f(e_1), \dots, f(e_n)) + l(g(e_1), \dots, g(e_n)) \\ &= k(i_1, \dots, i_m)F + l(i_1, \dots, i_m)G \\ &= (i_1, \dots, i_m)(kF + lG), \end{aligned}$$

这就说明  $\sigma(kf + lg) = kF + lG = k\sigma(f) + l\sigma(g)$ . □

设  $a \in \mathcal{U}$ , 它的坐标表示为  $a = (e_1, \dots, e_n) \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = (e_1, \dots, e_n)\hat{a}$ , 则

$$\begin{aligned} f((e_1, \dots, e_n)\hat{a}) &= f(a) = f(a_1e_1 + \dots + a_ne_n) \\ &= a_1f(e_1) + \dots + a_nf(e_n) \\ &= (f(e_1), \dots, f(e_n))\hat{a} = (f(e_1, \dots, e_n))\hat{a}. \end{aligned}$$

这说明上式形式上满足结合律. 类似地, 对任意  $n \times p$  矩阵  $A$ , 也可以得到形式上的结合律:

$$f((e_1, \dots, e_n)A) = (f(e_1, \dots, e_n))A.$$

利用上述形式上的计算, 可以得到复合映射在给定基下的矩阵.

**命题 7.5.4** 给定线性空间  $\mathcal{U}, \mathcal{V}, \mathcal{W}$ , 分别取定一组基  $e_1, \dots, e_n; i_1, \dots, i_m; j_1, \dots, j_l$ . 若  $f \in \text{Hom}(\mathcal{U}, \mathcal{V}), g \in \text{Hom}(\mathcal{V}, \mathcal{W})$ , 则

$$\sigma_{e_1, \dots, e_n; j_1, \dots, j_l}(gf) = \sigma_{i_1, \dots, i_m; j_1, \dots, j_l}(g)\sigma_{e_1, \dots, e_n; i_1, \dots, i_m}(f).$$

证. 根据定义域的不同, 我们可以放心地把不同的  $\sigma$  的下标省略, 而信息可以从自变量中得出. 设  $\sigma(f) = F, \sigma(g) = G$ , 那么  $f(e_1, \dots, e_n) = (i_1, \dots, i_m)F, g(i_1, \dots, i_m) = (j_1, \dots, j_l)G$ , 于是

$$\begin{aligned} (gf)(e_1, \dots, e_n) &= (gf(e_1), \dots, gf(e_n)) \\ &= g(f(e_1), \dots, f(e_n)) \end{aligned}$$

$$\begin{aligned}
&= g((i_1, \dots, i_m)F) \\
&= (g(i_1, \dots, i_m))F \\
&= (j_1, \dots, j_l)G)F \\
&= (j_1, \dots, j_l)(GF),
\end{aligned}$$

这就说明  $\sigma(gf) = GF = \sigma(g)\sigma(f)$ . □

定理 7.5.3 和命题 7.5.4 说明, 同构映射  $\sigma: \text{Hom}(\mathcal{U}, \mathcal{V}) \rightarrow \mathbb{F}^{m \times n}$  不仅能够保持线性运算, 在不同的空间之间使用合适的同构映射, 还能够保持乘法运算, 从而保持一切只与线性运算和乘法有关的性质. 简单应用如下, 证明留给读者.

**命题 7.5.5** 给定线性空间  $\mathcal{U}, \mathcal{V}$  及各自一组基  $e_1, \dots, e_n; i_1, \dots, i_m$ . 设  $f \in \text{Hom}(\mathcal{U}, \mathcal{V})$ , 而  $F$  是  $f$  在给定基下的矩阵, 则

$$\sigma_{e_1, \dots, e_n}: \mathcal{N}(f) \rightarrow \mathcal{N}(F), \quad \sigma_{i_1, \dots, i_m}: \mathcal{R}(f) \rightarrow \mathcal{R}(F),$$

都是线性同构. 特别地,  $\dim \mathcal{N}(f) = \dim \mathcal{N}(F)$ ,  $\dim \mathcal{R}(f) = \dim \mathcal{R}(F)$ , 因此,  $\dim \mathcal{N}(f) + \dim \mathcal{R}(f) = \dim \mathcal{U}$ .

下面讨论线性映射的矩阵在基变换下的变化规律.

**命题 7.5.6** 给定数域  $\mathbb{F}$  上线性空间  $\mathcal{U}, \mathcal{V}$ , 和  $\mathcal{U}$  的两组基  $e_1, \dots, e_n; t_1, \dots, t_n$  与  $\mathcal{V}$  的两组基  $i_1, \dots, i_m; s_1, \dots, s_m$ . 记二者的过渡矩阵分别为  $T, S$ , 即

$$(t_1, \dots, t_n) = (e_1, \dots, e_n)T, \quad (s_1, \dots, s_m) = (i_1, \dots, i_m)S.$$

如果  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$  在基  $e_1, \dots, e_n; i_1, \dots, i_m$  下的矩阵为  $F$ , 则该映射在基  $t_1, \dots, t_n; s_1, \dots, s_m$  下的矩阵为  $S^{-1}FT$ .

证. 设  $f$  在后两组基下的矩阵为  $\tilde{F}$ . 由定义,

$$f(e_1, \dots, e_n) = (i_1, \dots, i_m)F, \quad f(t_1, \dots, t_n) = (s_1, \dots, s_m)\tilde{F}.$$

代入过渡矩阵, 有  $f((e_1, \dots, e_n)T) = ((i_1, \dots, i_m)S)\tilde{F}$ . 利用形式上的结合律, 有

$$\begin{aligned}
(i_1, \dots, i_m)(S\tilde{F}) &= ((i_1, \dots, i_m)S)\tilde{F} = f((e_1, \dots, e_n)T) \\
&= (f(e_1, \dots, e_n))T \\
&= ((i_1, \dots, i_m)F)T = (i_1, \dots, i_m)(FT).
\end{aligned}$$

由  $i_1, \dots, i_m$  线性无关可得  $S\tilde{F} = FT$ . 而过渡矩阵可逆, 因此  $\tilde{F} = S^{-1}FT$ . □

命题 7.5.6 还能按如下方式理解: 对  $\mathbf{a} \in \mathcal{U}$ , 如果它在基  $\mathbf{t}_1, \dots, \mathbf{t}_n$  下的坐标是  $\hat{\mathbf{a}}$ , 则在基  $\mathbf{e}_1, \dots, \mathbf{e}_n$  下的坐标是  $T\hat{\mathbf{a}}$ ; 对像  $f(\mathbf{a})$ , 在基  $\mathbf{i}_1, \dots, \mathbf{i}_m$  下的坐标是  $FT\hat{\mathbf{a}}$ , 在基  $\mathbf{s}_1, \dots, \mathbf{s}_m$  下的坐标就是  $S^{-1}FT\hat{\mathbf{a}}$ . 因此  $f$  在这组基下的矩阵就是  $S^{-1}FT$ .

由  $\tilde{F} = S^{-1}FT$  可知, 同一个线性映射在不同基下的矩阵相抵. 其逆命题也成立.

命题 7.5.7 给定数域  $\mathbb{F}$  上线性空间  $\mathcal{U}, \mathcal{V}$  及各自一组基  $\mathbf{e}_1, \dots, \mathbf{e}_n; \mathbf{i}_1, \dots, \mathbf{i}_m$ . 如果  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$  在其下的矩阵是  $F$ , 则对任意与  $F$  相抵的矩阵  $\tilde{F}$ , 都存在  $\mathcal{U}, \mathcal{V}$  各自一组基, 使得  $f$  在其下的矩阵是  $\tilde{F}$ .

证. 由于  $\tilde{F}$  与  $F$  相抵, 根据命题 2.3.15, 存在可逆矩阵  $S, T$ , 使得  $\tilde{F} = S^{-1}FT$ . 令

$$(\mathbf{t}_1, \dots, \mathbf{t}_n) = (\mathbf{e}_1, \dots, \mathbf{e}_n)T, \quad (\mathbf{s}_1, \dots, \mathbf{s}_m) = (\mathbf{i}_1, \dots, \mathbf{i}_m)S,$$

则  $f$  在基  $\mathbf{t}_1, \dots, \mathbf{t}_n; \mathbf{s}_1, \dots, \mathbf{s}_m$  下的矩阵为  $\tilde{F} = S^{-1}FT$ .  $\square$

由此可得如下推论.

命题 7.5.8 给定数域  $\mathbb{F}$  上线性空间  $\mathcal{U}, \mathcal{V}$ , 对任意  $f \in \text{Hom}(\mathcal{U}, \mathcal{V})$ , 都存在  $\mathcal{U}, \mathcal{V}$  的一组基, 使得  $f$  在其下的矩阵是  $D_r = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ , 其中  $r = \dim \mathcal{R}(f)$ .

证一. 矩阵的相抵是等价关系, 任意秩为  $r$  的矩阵的相抵标准形是  $D_r$ , 利用命题 7.5.7 立得.  $\square$

证二. 由于  $\dim \mathcal{N}(f) = \dim \mathcal{U} - \dim \mathcal{R}(f) = n - r$ , 取  $\mathcal{N}(f)$  的一组基  $\mathbf{e}_{r+1}, \dots, \mathbf{e}_n$ , 扩充成  $\mathcal{U}$  的一组基  $\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{e}_{r+1}, \dots, \mathbf{e}_n$ .

可证  $f(\mathbf{e}_1), \dots, f(\mathbf{e}_r)$  线性无关. 事实上, 设  $k_1 f(\mathbf{e}_1) + \dots + k_r f(\mathbf{e}_r) = \mathbf{0}$ , 则  $f(k_1 \mathbf{e}_1 + \dots + k_r \mathbf{e}_r) = \mathbf{0}$ , 即  $k_1 \mathbf{e}_1 + \dots + k_r \mathbf{e}_r \in \mathcal{N}(f)$ . 而  $\mathbf{e}_{r+1}, \dots, \mathbf{e}_n$  是  $\mathcal{N}(f)$  的一组基, 存在  $k_{r+1}, \dots, k_n \in \mathbb{F}$ , 使得  $k_1 \mathbf{e}_1 + \dots + k_r \mathbf{e}_r = k_{r+1} \mathbf{e}_{r+1} + \dots + k_n \mathbf{e}_n$ . 但  $\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{e}_{r+1}, \dots, \mathbf{e}_n$  线性无关, 因此  $k_1 = \dots = k_r = k_{r+1} = \dots = k_n = 0$ .

设  $\mathbf{i}_1 = f(\mathbf{e}_1), \dots, \mathbf{i}_r = f(\mathbf{e}_r)$ , 将其扩充成  $\mathcal{V}$  的一组基  $\mathbf{i}_1, \dots, \mathbf{i}_r, \mathbf{i}_{r+1}, \dots, \mathbf{i}_m$ . 直接计算可得,

$$f(\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{e}_{r+1}, \dots, \mathbf{e}_n) = (\mathbf{i}_1, \dots, \mathbf{i}_r, \mathbf{0}, \dots, \mathbf{0}) = (\mathbf{i}_1, \dots, \mathbf{i}_r, \mathbf{i}_{r+1}, \dots, \mathbf{i}_m) \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

即其矩阵为  $D_r$ .  $\square$

最后, 来考虑线性变换在基下的坐标表示. 设  $f$  是  $\mathcal{V}$  上的线性变换,  $\mathbf{e}_1, \dots, \mathbf{e}_n$  是  $\mathcal{V}$  的一组基, 则

$$f(\mathbf{e}_1, \dots, \mathbf{e}_n) = (\mathbf{e}_1, \dots, \mathbf{e}_n)F,$$

其中  $n$  阶方阵  $F$  称为线性映射  $f$  在给定基下的 (表示) 矩阵.

注意, 线性映射在基下的矩阵需要在定义域和陪域各取一组基, 而线性变换的矩阵在定义域和陪域取的是同一组基.

**例 7.5.9** 例 7.5.2 第 3 条给出了一个线性变换的矩阵的例子, 下面再列出一些线性变换的矩阵, 其中部分来自例 7.3.11.

1. 设  $A = [a_{ij}] \in \mathbb{F}^{2 \times 2}$ ,  $\mathbb{F}^{2 \times 2}$  上的线性变换

$$C_A = L_A - R_A: \mathbb{F}^{2 \times 2} \rightarrow \mathbb{F}^{2 \times 2}, \\ X \mapsto AX - XA,$$

在基  $E_{11}, E_{21}, E_{12}, E_{22}$  下的矩阵表示为

$$C_A(E_{11}, E_{21}, E_{12}, E_{22}) = (E_{11}, E_{21}, E_{12}, E_{22})C_A, \\ \text{其中 } C_A = \begin{bmatrix} 0 & a_{12} & -a_{21} & 0 \\ a_{21} & a_{22} - a_{11} & 0 & -a_{21} \\ -a_{12} & 0 & a_{11} - a_{22} & a_{12} \\ 0 & -a_{12} & a_{21} & 0 \end{bmatrix} \text{ 是 } C_A \text{ 在这组基下的矩阵.}$$

2.  $\mathbb{F}^{2 \times 2}$  上的线性变换

$$S: \mathbb{F}^{2 \times 2} \rightarrow \mathbb{F}^{2 \times 2}, \\ A \mapsto \frac{1}{2}(A + A^T).$$

在基  $E_{11}, E_{21}, E_{12}, E_{22}$  下的矩阵表示为

$$S(E_{11}, E_{21}, E_{12}, E_{22}) = (E_{11}, E_{21}, E_{12}, E_{22})S, \\ \text{其中 } S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ 是 } S \text{ 在这组基下的矩阵.} \quad \odot$$

线性变换的特征值问题, 通过矩阵表示可以转化为矩阵的特征值问题.

**命题 7.5.10** 给定  $\mathbb{F}$  上的线性空间  $\mathcal{V}$  及其一组基  $e_1, \dots, e_n$ , 又设  $f$  是其上的线性变换, 在给定基下的矩阵是  $F$ . 对  $\lambda \in \mathbb{F}, x \in \mathcal{V}$ , 若  $\hat{x} \in \mathbb{F}^n$  是  $x$  在该组基下的坐标, 则  $(\lambda, x)$  是  $f$  的特征对, 当且仅当  $(\lambda, \hat{x})$  是  $F$  的特征对, 即

$$f(x) = \lambda x \Leftrightarrow F\hat{x} = \lambda\hat{x}.$$

**例 7.5.11** 考虑例 7.5.9 .

1. 线性映射  $C_A$  在给定基下的矩阵是  $C_A$ . 计算可得,  $C_A$  的特征多项式为  $\lambda^2[\lambda^2 - \text{trace}(A)^2 + 4\det(A)]$ . 设  $A$  的两个特征值为  $\lambda_1, \lambda_2$ , 则上面特征多项式为  $\lambda^2[\lambda^2 - (\lambda_1 - \lambda_2)^2]$ .

(a) 若  $\lambda_1 \neq \lambda_2$ , 容易验证  $\text{rank}(C_A) = 2$ , 因此  $\dim \mathcal{N}(C_A) = 2$ .

(b) 若  $\lambda_1 = \lambda_2$ , 但  $a_{12}, a_{21}, a_{11} - a_{22}$  至少有一个不为零, 则  $\dim \mathcal{N}(C_A) = 4 - \text{rank}(C_A) = 2$ .

(c) 若  $a_{12} = a_{21} = a_{11} - a_{22} = 0$ , 则  $\dim \mathcal{N}(C_A) = \text{rank}(C_A) = 4$ , 此时  $A = a_{11}I_2$  而  $C_A = \mathbf{O}$  是零映射.

不考虑  $A$  是数量矩阵的情形, 通过计算属于 0 的特征向量, 还能得到核  $\mathcal{N}(L_A - R_A) = \{X \in \mathbb{F}^{2 \times 2} \mid AX = XA\}$  的一组基, 由两个矩阵 ( $\mathbb{F}^{2 \times 2}$  中的向量) 组成, 这比例 7.3.11 又深入了一步.

2. 线性映射  $S$  在给定基下的矩阵是  $S$ . 而  $S$  是对称矩阵, 可对角化,  $S$  的四个线性

无关的特征向量为: 属于 0 的  $\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$ ; 属于 1 的  $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$ . 因此, 核和像

集分别为

$$\mathcal{N}(S) = \text{span} \left( \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right), \quad \mathcal{R}(S) = \text{span} \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right).$$

前者是全体反对称矩阵构成的子空间, 后者是全体对称矩阵构成的子空间, 这与例 7.3.11 的结论相一致. ☺

对线性空间  $\mathcal{V}$  上的线性变换  $f$ , 先取  $\mathcal{V}$  的一组基  $e_1, \dots, e_n$ , 由此得到  $f$  在该组基下的矩阵  $F$ . 如果  $F$  可对角化, 即存在可逆矩阵  $T$ , 使得  $T^{-1}FT = \Lambda$  是对角矩阵, 令  $(t_1, \dots, t_n) = (e_1, \dots, e_n)T$ , 则  $f$  在基  $t_1, \dots, t_n$  下的矩阵就是  $\Lambda$ , 即  $t_1, \dots, t_n$  都是  $f$  的特征向量. 这暗示了如下结论, 证明留给读者.

**命题 7.5.12** 设  $\mathbb{F}$  上  $n$  维线性空间  $\mathcal{V}$ ,  $f$  是其上的线性变换, 它在某组基下的矩阵是  $F$ , 则  $F$  可对角化当且仅当  $f$  有  $n$  个线性无关的特征向量.

下面讨论线性变换的矩阵在基变换下的变化规律.

**命题 7.5.13** 给定数域  $\mathbb{F}$  上线性空间  $\mathcal{V}$  和它的两组基  $e_1, \dots, e_n; t_1, \dots, t_n$ . 记过渡矩阵为  $T$ , 即  $(t_1, \dots, t_n) = (e_1, \dots, e_n)T$ . 如果  $\mathcal{V}$  上线性变换  $f$  在基  $e_1, \dots, e_n$  下的矩阵为  $F$ , 则  $f$  在基  $t_1, \dots, t_n$  下的矩阵为  $T^{-1}FT$ .

证明与命题 7.5.6 的证明类似, 留给读者. **线性空间做基变换, 其上线性变换的矩阵就做相似变换.** 事实上, 有如下结论.

**命题 7.5.14** 数域  $\mathbb{F}$  上两个  $n$  阶方阵  $A, B$  相似, 当且仅当  $A, B$  是  $n$  维线性空间  $\mathcal{V}$  上某个线性变换在两组基下的矩阵.

证. “ $\Leftarrow$ ”: 由命题 7.5.13 立得.

“ $\Rightarrow$ ”: 设  $B = T^{-1}AT$ , 其中  $T$  可逆. 根据定理 7.5.3, 能构造  $\mathcal{V}$  上线性变换  $f$  使得它在  $\mathcal{V}$  的某组基下的矩阵为  $A$ , 记该组基为  $e_1, \dots, e_n$ . 令  $(t_1, \dots, t_n) = (e_1, \dots, e_n)T$ , 由于  $T$  可逆, 根据命题 7.4.6,  $t_1, \dots, t_n$  是  $V$  的一组基. 根据命题 7.5.13,  $f$  在这组基下的矩阵就是  $T^{-1}AT = B$ .  $\square$

命题 7.5.14 说明, 把  $n$  阶矩阵  $A$  对角化, 本质上就是为  $\mathbb{F}^n$  找到一组基, 使得  $A$  对应的线性变换在这组基下的矩阵是对角矩阵.

最后, 给出求解定常齐次常微分方程的求解方法, 来体会线性变换的矩阵表示.

**例 7.5.15 (定常齐次线性常微分方程)** 所谓定常齐次线性常微分方程, 就是具有如下形式的方程:

$$\frac{d^n u}{dt^n} + a_{n-1} \frac{d^{n-1} u}{dt^{n-1}} + \dots + a_1 \frac{du}{dt} + a_0 u = 0.$$

类似于第 5.2 节最后对多项式的友矩阵的构造或例 5.4.9 和 6.1.7, 令  $v = \begin{bmatrix} u \\ \frac{du}{dt} \\ \vdots \\ \frac{d^{n-1}u}{dt^{n-1}} \end{bmatrix}$ , 则

$\frac{dv}{dt} = C_p v$ , 其中, 而  $C_p$  是多项式  $t^n + a_{n-1}t^{n-1} + \dots + a_1t + a_0$  的友矩阵. 问题就转化为了关于向量的常微分方程. 含有自变量的向量的微分, 就是逐项微分得到的向量, 这里不做更多说明.

下面考虑关于向量的常微分方程

$$\frac{du(t)}{dt} = Au(t).$$

如果  $u$  是一个 1 维向量, 即数, 那么我们已经知道常微分方程的解是  $u(t) = e^{At}u(0)$ . 如果  $u(t)$  不是 1 维向量, 该如何解决?

观察方程. 事实上, 这一方程意味着, 对向量  $u(t)$  作求导这一线性变换, 相当于用矩阵  $A$  左乘这一线性变换, 记为  $L_A$ . 只要能找到一组基, 使得这两个线性变换在这组

基下的矩阵很简单, 问题就容易了. 如果矩阵  $A$  可对角化, 则存在一组基, 使得线性变换  $L_A$  在这组基下的矩阵是对角矩阵. 记其谱分解为  $A = X\Lambda X^{-1}$ , 则过渡矩阵是  $X^{-1}$ , 于是  $\mathbf{u}$  在这组基下的坐标就是  $\widehat{\mathbf{u}} = X\mathbf{u}$ . 方程就变为

$$\frac{d\widehat{\mathbf{u}}(t)}{dt} = \Lambda\widehat{\mathbf{u}}(t),$$

该方程可直接得到  $n$  个 1 维向量的常微分方程. 记  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ , 解就是  $\widehat{\mathbf{u}}(t) = \text{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t})\widehat{\mathbf{u}}(0) =: e^{\Lambda t}\widehat{\mathbf{u}}(0)$ . 再把基换回, 就有  $\mathbf{u}(t) = Xe^{\Lambda t}X^{-1}\mathbf{u}(0)$ .

通过以上分析可见, 如果  $A$  可对角化, 且特征值为  $\lambda_1, \dots, \lambda_n$ , 那么解是  $e^{\lambda_1 t}, \dots, e^{\lambda_n t}$  的线性组合. 然而, 可对角化的  $A$  的特征值还有可能为复数, 这一情形也需要特殊处理. 如果存在复特征值  $\lambda_i = \alpha + i\beta \notin \mathbb{R}$ , 则  $\bar{\lambda}_i$  也是特征值, 容易验证

$$\text{span}\{e^{\lambda_i t}, e^{\bar{\lambda}_i t}\} = \text{span}\{e^{\alpha t} \cos(\beta t), e^{\alpha t} \sin(\beta t)\},$$

解空间的基向量可以换为后两个实函数.

1. 如果  $\lambda_i \in \mathbb{R}, \lambda_i > 0$ , 则随着时间  $t$  的增长,  $e^{\lambda_i t} \rightarrow \infty$ ;
2. 如果  $\lambda_i \in \mathbb{R}, \lambda_i = 0$ , 则随着时间  $t$  的增长,  $e^{\lambda_i t} = 1$  一直成立;
3. 如果  $\lambda_i \in \mathbb{R}, \lambda_i < 0$ , 则随着时间  $t$  的增长,  $e^{\lambda_i t} \rightarrow 0$ ;
4. 如果  $\lambda_i \notin \mathbb{R}, \Re \lambda_i > 0$ , 则随着时间  $t$  的增长,  $e^{\lambda_i t}$  发散到  $\infty$ ;
5. 如果  $\lambda_i \notin \mathbb{R}, \Re \lambda_i = 0$ , 则随着时间  $t$  的增长,  $e^{\lambda_i t}$  发散, 但总在单位圆上;
6. 如果  $\lambda_i \notin \mathbb{R}, \Re \lambda_i < 0$ , 则随着时间  $t$  的增长,  $e^{\lambda_i t} \rightarrow 0$ .

因此, 如果我们希望方程的解趋于 0, 那么矩阵  $A$  的特征值的实部必须小于 0. 为此引入如下定义: 特征值的实部全为负数的矩阵称为**稳定矩阵**. 可以证明, 实矩阵  $A$  是稳定矩阵, 当且仅当  $A$  负定 (不一定对称), 即对任意非零  $\mathbf{x} \in \mathbb{R}^n$ , 都有  $\mathbf{x}^T A \mathbf{x} < 0$ .

当  $A$  不可对角化时,  $A$  有 Jordan 分解  $A = XJX^{-1}$ , 其中  $J$  是分块对角矩阵, 对角块

是 Jordan 块. 为 Jordan 块  $J_m(\lambda)$  定义指数函数  $e^{J_m(\lambda)t} = \begin{bmatrix} e^{\lambda t} & te^{\lambda t} & \cdots & \frac{t^{m-1}}{(m-1)!}e^{\lambda t} \\ & \ddots & \ddots & \vdots \\ & & e^{\lambda t} & te^{\lambda t} \\ & & & e^{\lambda t} \end{bmatrix},$

然后类推定义,  $e^{Jt}$  是分块对角矩阵, 对角块是  $J$  的对角块的指数函数. 上述基变换仍然有效, 因此方程的解为  $\mathbf{u}(t) = Xe^{Jt}X^{-1}\mathbf{u}(0)$ .

如何用  $A$  来表示解? 显然, 如果定义  $A$  的指数函数  $e^{At} = Xe^{Jt}X^{-1}$ , 则解的形式就和 1 维情形完全相同  $\mathbf{u}(t) = e^{At}\mathbf{u}(0)$ . 定义的难点在于: 谱分解或 Jordan 分解不唯

一, 方阵的指数函数唯一吗? 即如何良好定义方阵的指数函数. 这里不对此展开, 感兴趣的读者可以参考阅读 7.5.17.

综上, 常微分方程  $\frac{d\mathbf{u}(t)}{dt} = A\mathbf{u}(t)$  的解就是  $\mathbf{u}(t) = e^{At}\mathbf{u}(0)$ . ☺

## 习题

练习 7.5.1 证明命题 7.5.5.

练习 7.5.2 证明命题 7.5.12.

练习 7.5.3 证明命题 7.5.13.

练习 7.5.4 设  $A$  是  $\mathbb{F}^3$  上的一个线性变换:

$$A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ x_3 - x_2 \\ x_2 - x_3 \end{bmatrix}.$$

求  $A$  在标准基下的矩阵, 并分别求  $\mathcal{N}(A)$  和  $\mathcal{R}(A)$  的一组基和维数.

练习 7.5.5 设  $\mathcal{V} = \text{span}(f_1, f_2)$  是函数空间的子空间, 其中  $f_1 = e^{ax} \cos bx$ ,  $f_2 = e^{ax} \sin bx$ . 证明求导算子  $D$  是  $\mathcal{V}$  上的线性变换, 并求其在基  $f_1, f_2$  下的矩阵.

练习 7.5.6 设  $\mathbb{F}^{2 \times 2}$  上的线性变换  $L_A: X \mapsto AX$ , 其中  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . 求  $L_A$  在基  $E_{11}, E_{12}, E_{21}, E_{22}$  下的矩阵.

练习 7.5.7 设  $\mathcal{V}$  是所有 2 阶对称矩阵构成的线性空间,  $f$  是其上的线性变换:  $f(X) = A^T X A$ , 其中  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . 求  $f$  在基  $E_{11}, E_{22}, E_{12} + E_{21}$  下的矩阵.

练习 7.5.8 设  $A$  是数域  $\mathbb{F}$  上的  $n$  维线性空间  $\mathcal{V}$  上的一个线性变换, 证明, 存在  $\mathbb{F}[x]$  中一个次数不超过  $n^2$  的多项式  $f(x)$ , 使得  $f(A) = O$ .

练习 7.5.9 设  $\mathcal{V}$  是数域  $\mathbb{F}$  上的  $n$  维线性空间,  $\mathbb{F}$  可以看作自身上的线性空间, 而  $\mathcal{V}$  到  $\mathbb{F}$  的线性映射称为  $\mathcal{V}$  上的线性函数. 令  $\mathcal{V}^* = \text{Hom}(\mathcal{V}, \mathbb{F})$ , 称为  $\mathcal{V}$  的对偶空间. 证明,  $\mathcal{V}^*$  和  $\mathcal{V}$  同构.

练习 7.5.10 求导算子  $D$  定义了多项式空间  $\mathbb{F}[x]_n$  上的线性变换, 给定  $\mathbb{F}[x]_n$  的两组基  $1, x, \dots, x^{n-1}$  和  $1, x, \dots, \frac{1}{(n-1)!}x^{n-1}$ .

1. 求两组基之间的过渡矩阵.
2. 求  $D$  在两组基下的矩阵.
3. 通过过渡矩阵验证两个表示矩阵是相似的.
4. 是否存在一组基, 使得  $D$  在该组基下的矩阵是对角矩阵?



**练习 7.5.11** 设  $\mathcal{V}$  是  $n$  维线性空间,  $f$  是其上的线性变换, 又设存在向量  $\alpha \in \mathcal{V}$ , 使得  $f^{n-1}(\alpha) \neq \mathbf{0}$ , 且  $f^n(\alpha) = \mathbf{0}$ . 证明,  $\mathcal{V}$  存在一组基, 使得  $f$  在该组基下的矩阵是

$$J_n = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}.$$

**练习 7.5.12** 考虑练习 7.5.11 中的方阵  $J_n$ , 判断  $J_n$  与  $J_n^T$  是否相似.

**练习 7.5.13** 已知  $\mathbb{F}^3$  上的线性变换  $f$  在标准基  $e_1, e_2, e_3$  下的矩阵是  $A = \begin{bmatrix} 15 & -11 & 5 \\ 20 & -15 & 8 \\ 8 & -7 & 6 \end{bmatrix}$ . 设

$t_1 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, t_2 = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}, t_3 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$  是  $\mathbb{F}^3$  的另一组基, 求  $f$  在这组基下的矩阵.

**练习 7.5.14** 设 3 维线性空间  $\mathcal{V}$  有一组基  $e_1, e_2, e_3$ , 其上的线性变换  $f$  在该组基下的矩阵是

$$A = \begin{bmatrix} 2 & 3 & 2 \\ 1 & 8 & 2 \\ -2 & -14 & -3 \end{bmatrix}.$$

1. 求  $f$  的全部特征值和特征向量.
2. 判断是否存在一组基, 使得  $f$  在该组基下的表示矩阵是对角矩阵. 如果存在, 写出这组基及对应的对角矩阵.

**练习 7.5.15** 设 4 维线性空间  $\mathcal{V}$  有一组基  $e_1, e_2, e_3, e_4$ , 其上的线性变换  $f$  在该组基下的矩阵是

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

1. 求  $f$  的全部特征值和特征向量.
2. 判断是否存在一组基, 使得  $f$  在该组基下的表示矩阵是对角矩阵. 如果存在, 写出这组基及对应的对角矩阵.

**练习 7.5.16** 设  $B = \begin{bmatrix} -1 & -1 \\ 2 & 1 \end{bmatrix}$ , 在  $\mathbb{F}^{2 \times 2}$  中定义如下变换:

$$f(X) = B^{-1}XB, \quad \forall X \in \mathbb{F}^{2 \times 2}.$$

1. 证明  $f$  是线性变换.
2. 求  $f$  的全部特征值和特征向量.

**阅读 7.5.17 (矩阵函数)** 首先是例 7.5.15 中的指数函数的良定义问题. 最简单的做法是利用幂级数. 在收敛半径内,  $e^{\lambda t} = \sum_{k=0}^{\infty} \frac{1}{k!} (\lambda t)^k = 1 + \lambda t + \frac{1}{2} \lambda^2 t^2 + \dots$ . 因此

$$\begin{aligned} X \operatorname{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t}) X^{-1} &= X \operatorname{diag} \left( \sum_k \frac{1}{k!} (\lambda_1 t)^k, \dots, \sum_k \frac{1}{k!} (\lambda_n t)^k \right) X^{-1} \\ &= \sum_k \frac{1}{k!} t^k X \operatorname{diag}(\lambda_1^k, \dots, \lambda_n^k) X^{-1} \\ &= \sum_k \frac{1}{k!} t^k X A^k X^{-1} = \sum_k \frac{1}{k!} t^k A^k. \end{aligned}$$

这就得到了一个  $A$  的级数.

类似地, 利用上述办法, 在一定条件下, 我们可以定义任意数量函数的矩阵版本. 对解析函数<sup>4</sup>  $f(z)$ , 在其收敛半径内总能写成幂级数  $f(z) = \sum_k \frac{f^{(k)}(z_0)}{k!} (z - z_0)^k$ . 定义矩阵函数为  $f(A) = \sum_k \frac{f^{(k)}(z_0)}{k!} (A - z_0 I)^k$ . 通过细致的分析讨论可知, 只要  $A$  的特征值全都在这一级数的收敛半径内, 矩阵级数就收敛, 从而矩阵函数有良定义.

矩阵函数的特征值, 就是对应数量函数在矩阵的特征值处的值; 矩阵函数的特征向量, 就是矩阵本身的特征向量.

<sup>4</sup>一个函数解析, 是指该函数在定义域内的每一点的邻域内的 Taylor 级数都收敛. 注意这个条件比函数无穷次可微强.

# 第八章 内积空间

第 3 章讲述了  $\mathbb{R}^n$  上的内积, 并利用内积导出了长度、距离、正交等概念. 本章将为  $\mathbb{R}$  或  $\mathbb{C}$  上的一般线性空间定义内积, 并同样定义相应的导出概念.

## 8.1 欧氏空间

**定义 8.1.1 (内积)** 给定  $\mathbb{R}$  上的线性空间  $\mathcal{V}$ , 如果  $\mathcal{V}$  上定义的二元函数  $\langle \cdot, \cdot \rangle \rightarrow \mathbb{R}$ , 满足:

1. 对称性: 对任意  $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ ,  $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$ ;
2. 双线性性: 对任意  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}_1, \mathbf{b}_2 \in \mathcal{V}, k_1, k_2 \in \mathbb{R}$ ,

$$\langle k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2, \mathbf{b}_1 \rangle = k_1 \langle \mathbf{a}_1, \mathbf{b}_1 \rangle + k_2 \langle \mathbf{a}_2, \mathbf{b}_1 \rangle,$$

$$\langle \mathbf{a}_1, k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 \rangle = k_1 \langle \mathbf{a}_1, \mathbf{b}_1 \rangle + k_2 \langle \mathbf{a}_1, \mathbf{b}_2 \rangle;$$

3. 正定性: 对任意  $\mathbf{a} \in \mathcal{V}$ ,  $\langle \mathbf{a}, \mathbf{a} \rangle \geq 0$ , 且  $\langle \mathbf{a}, \mathbf{a} \rangle = 0$  当且仅当  $\mathbf{a} = \mathbf{0}$ ;

则称二元函数  $\langle \cdot, \cdot \rangle$  是  $\mathcal{V}$  上的一个**内积**, 称具有内积的线性空间  $\mathcal{V}$  为一个**实内积空间**或**Euclid 空间**, 简称**欧氏空间**.

欧氏空间的**维数**定义为其作为线性空间的维数.

由对称性可知, 双线性性中的两个条件, 即对第一变量线性和对第二变量线性二者等价.

易见, 为  $\mathbb{R}$  上的线性空间添加一个内积运算, 就得到了欧氏空间.

**例 8.1.2** 1. 坐标向量空间  $\mathbb{R}^n$  关于内积  $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}^T \mathbf{a}$  构成欧氏空间, 这个内积常称为线性空间  $\mathbb{R}^n$  上的**标准内积**. 不加额外说明时, 欧氏空间  $\mathbb{R}^n$  上的内积就是标准内积.

对任意  $D$  是一个对角元素全为正数的对角矩阵,  $\mathbb{R}^n$  关于如下二元函数构成欧氏空间:  $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}^T D \mathbf{a}$ .

一般地,  $\mathbb{R}^n$  上的二元函数  $\mathbf{b}^T A \mathbf{a}$  是一个内积, 当且仅当  $A$  对称正定 (为什么?). 可见, 同一线性空间上可以添加不同的内积运算.

2. 矩阵空间  $\mathbb{R}^{m \times n}$  关于如下二元函数构成欧氏空间:  $\langle A, B \rangle = \text{trace}(B^T A)$ . 请读者自己验证. 这个内积常称为线性空间  $\mathbb{R}^{m \times n}$  上的**标准内积**.
3. 区间  $[a, b]$  上的连续函数构成的线性空间  $C[a, b]$ , 如下二元函数定义了其上的一个内积:  $\langle f, g \rangle = \int_a^b f(x)g(x) dx$ . 事实上, 内积定义中的对称性、双线性性显然满足, 而正定性, 显然有  $\langle f, f \rangle = \int_a^b (f(x))^2 dx \geq 0$ ; 由定积分的知识可知,  $\int_a^b (f(x))^2 dx = 0$ , 当且仅当连续函数  $f(x)$  是零函数, 即  $f(x)$  是零向量. ☺

类似于第 3.1 节中对欧氏空间  $\mathbb{R}^n$  的讨论, 在一般的欧氏空间中, 也能从内积导出长度、距离、正交等概念.

向量  $\mathbf{a}$  的**长度**或**范数**定义为  $\|\mathbf{a}\| := \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}$ , 而向量  $\mathbf{a}, \mathbf{b}$  的**距离**定义为  $\|\mathbf{a} - \mathbf{b}\|$ .

**例 8.1.3** 1. 欧氏空间  $\mathbb{R}^{m \times n}$  中向量  $A$  的长度, 就是矩阵的 Frobenius 范数.

2. 欧氏空间  $C[a, b]$  中向量  $f$  的长度, 就是  $\sqrt{\int_a^b (f(x))^2 dx}$ . ☺

**定理 8.1.4 (Cauchy-Schwarz 不等式)** 对实内积空间中的任意向量  $\mathbf{a}, \mathbf{b}$ , 都有

$$|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\| \|\mathbf{b}\|.$$

等号成立当且仅当  $\mathbf{a}, \mathbf{b}$  共线.

证. 根据内积的正定性, 对任意实数  $t$ ,  $(\mathbf{a} + t\mathbf{b})^T (\mathbf{a} + t\mathbf{b}) = \mathbf{a}^T \mathbf{a} + 2t\mathbf{a}^T \mathbf{b} + t^2 \mathbf{b}^T \mathbf{b} \geq 0$ , 而这意味着判别式  $(2\mathbf{a}^T \mathbf{b})^2 - 4(\mathbf{a}^T \mathbf{a})(\mathbf{b}^T \mathbf{b}) \leq 0$ , 即得结论. □

**例 8.1.5** 将不同欧氏空间上的 Cauchy-Schwarz 不等式<sup>1</sup>具体写出, 可得多种不等式.

1. 欧氏空间  $\mathbb{R}^{m \times n}$ :  $|\text{trace}(A^T B)| \leq \|A\|_F \|B\|_F$ .

2. 欧氏空间  $C[a, b]$ :  $\left( \int_a^b f(x)g(x) dx \right)^2 \leq \int_a^b (f(x))^2 dx \cdot \int_a^b (g(x))^2 dx$ . ☺

立得如下推论.

<sup>1</sup>Cauchy-Schwarz 不等式, 是对一系列内积与范数关系的不等式的统称. 事实上, 关于坐标向量的不等式 (即定理 3.1.4) 是 Cauchy 的工作, 常直称 **Cauchy 不等式**. 而关于积分的不等式由 Bunyakovsky 和 Schwarz 各自独立发现, 也称为 **Cauchy-Bunyakovsky-Schwarz 不等式**. 后来这类不等式都统称如此.

**推论 8.1.6 (三角不等式)** 对欧氏空间中的任意向量  $\mathbf{a}, \mathbf{b}$ , 都有

$$\|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|.$$

等号成立当且仅当  $\mathbf{a}, \mathbf{b}$  共线.

如果向量  $\mathbf{a}, \mathbf{b}$  的内积  $\langle \mathbf{a}, \mathbf{b} \rangle = 0$ , 则称二者**正交**, 记为  $\mathbf{a} \perp \mathbf{b}$ .

由两两正交的非零向量组成的向量组称为**正交向量组**. 如果正交向量组中的向量都是单位向量, 则称其为**正交单位向量组**. 正交向量组线性无关, 证明留给读者.

**定义 8.1.7 (标准正交基)** 在  $n$  维欧氏空间  $\mathcal{V}$  中,  $n$  个向量组成的正交向量组一定是  $\mathcal{V}$  的一组基, 称为  $\mathcal{V}$  的一组**正交基**; 由  $n$  个向量组成的正交单位向量组一定是  $\mathcal{V}$  的一组基, 称为  $\mathcal{V}$  的一组**标准正交基**.

**例 8.1.8** 具有标准内积的欧氏空间  $\mathbb{R}^{m \times n}$  中, 设  $E_{ij}$  是  $(i, j)$  元素为 1, 其余元素为 0 的矩阵. 容易验证,  $E_{ij}, 1 \leq i \leq m, 1 \leq j \leq n$  是一组标准正交基.

事实上, 若  $\mathbf{u}_1, \dots, \mathbf{u}_m$  和  $\mathbf{v}_1, \dots, \mathbf{v}_n$  分别为  $\mathbb{R}^m$  和  $\mathbb{R}^n$  的一组标准正交基, 则  $\mathbf{u}_i \mathbf{v}_j^T, i = 1, \dots, m, j = 1, \dots, n$  是  $\mathbb{R}^{m \times n}$  的一组标准正交基. ☺

**例 8.1.9** 连续函数空间  $C[-\pi, \pi]$  中,  $1, \sin x, \cos x, \sin 2x, \cos 2x, \dots, \sin nx, \cos nx, \dots$  是正交向量组:

$$\begin{aligned} \int_{-\pi}^{\pi} \sin mx \cos nx \, dx &= 0, & m = 1, 2, \dots, n = 0, 1, 2, \dots \\ \int_{-\pi}^{\pi} \cos mx \cos nx \, dx &= \pi \delta_{mn}, & m = 1, 2, \dots, n = 0, 1, 2, \dots \\ \int_{-\pi}^{\pi} \sin mx \sin nx \, dx &= \pi \delta_{mn}, & m = 1, 2, \dots, n = 1, 2, \dots \\ \int_{-\pi}^{\pi} 1 \, dx &= 2\pi. \end{aligned}$$

其中

$$\delta_{mn} = \begin{cases} 1, & \text{若 } m = n, \\ 0, & \text{若 } m \neq n, \end{cases}$$

称为 **Kronecker 符号**. ☺

欧氏空间  $\mathbb{R}^n$  中的 Gram-Schmidt 正交化方法, 可以直接推广到一般的欧氏空间. 给定欧氏空间  $\mathcal{V}$  中线性无关的向量组  $\mathbf{a}_1, \dots, \mathbf{a}_r$ , 利用 Gram-Schmidt 正交化方法能够得到与之线性等价的正交向量组, 具体操作如下:

$$\tilde{\mathbf{q}}_1 = \mathbf{a}_1,$$

$$\begin{aligned}\tilde{q}_2 &= a_2 - \frac{\langle a_2, \tilde{q}_1 \rangle}{\langle \tilde{q}_1, \tilde{q}_1 \rangle} \tilde{q}_1, \\ &\vdots \\ \tilde{q}_r &= a_r - \frac{\langle a_r, \tilde{q}_1 \rangle}{\langle \tilde{q}_1, \tilde{q}_1 \rangle} \tilde{q}_1 - \cdots - \frac{\langle a_r, \tilde{q}_{r-1} \rangle}{\langle \tilde{q}_{r-1}, \tilde{q}_{r-1} \rangle} \tilde{q}_{r-1}.\end{aligned}$$

把以上正交向量组中的每个向量都单位化, 就可以得到正交单位向量组:  $q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$ . 用矩阵乘法表示如下:

$$(a_1, \dots, a_r) = (q_1, \dots, q_r)R, \quad (8.1.1)$$

其中  $R$  是上三角矩阵. 注意, 这和矩阵的  $QR$  分解并不完全相同, 因为此时向量组  $a_1, \dots, a_r$  和  $q_1, \dots, q_r$  在一般的线性空间中, 这仅仅是用来表示线性组合的形式上的记号.

当  $a_1, \dots, a_r$  是欧氏空间  $\mathcal{V}$  的一组基时, 通过 Gram-Schmidt 正交化可以得到一组标准正交基.

**定理 8.1.10** 任意有限维欧氏空间都存在一组正交基, 从而存在一组标准正交基.

此时, (8.1.1) 中的上三角矩阵  $R$  是从标准正交基  $q_1, \dots, q_r$  到基  $a_1, \dots, a_r$  的过渡矩阵.

**例 8.1.11** 多项式空间  $\mathbb{R}[x]$  上的二元函数  $\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx$  定义了其上的一个内积. 单项式  $1, x, x^2, \dots, x^n$  是其中的向量组, 但不是正交向量组. 为得到正交向量组, 利用 Gram-Schmidt 正交化方法,

$$\begin{aligned}\varphi_0(x) &= 1; \\ \varphi_1(x) &= x - \frac{\langle \varphi_0, x \rangle}{\langle \varphi_0, \varphi_0 \rangle} \varphi_0 = x - \frac{\int_{-1}^1 1x dx}{\int_{-1}^1 dx} = x; \\ \varphi_2(x) &= x^2 - \frac{\langle \varphi_0, x^2 \rangle}{\langle \varphi_0, \varphi_0 \rangle} \varphi_0 - \frac{\langle \varphi_1, x^2 \rangle}{\langle \varphi_1, \varphi_1 \rangle} \varphi_1 = x^2 - \frac{1}{3}; \\ \varphi_3(x) &= x^3 - \frac{\langle \varphi_0, x^3 \rangle}{\langle \varphi_0, \varphi_0 \rangle} \varphi_0 - \frac{\langle \varphi_1, x^3 \rangle}{\langle \varphi_1, \varphi_1 \rangle} \varphi_1 - \frac{\langle \varphi_2, x^3 \rangle}{\langle \varphi_2, \varphi_2 \rangle} \varphi_2 = x^3 - \frac{3}{5}x.\end{aligned}$$

通过数乘能得到多项式

$$P_n(x) = \frac{(2n)!}{2^n(n!)^2} \varphi_n(x), \quad n = 0, 1, \dots,$$

使得  $P_n(1) = 1$ , 称  $P_n$  为 **Legendre 多项式**. 显然,  $P_0, P_1, \dots, P_{n-1}$  是  $\mathbb{R}_n[x]$  的一组正交基. 可以验证如下递推关系和通项公式:

$$P_{n+1}(x) = \frac{2n+1}{n+1} x P_n(x) - \frac{n}{n+1} P_{n-1}(x),$$

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n.$$

Legendre 多项式在偏微分方程和数学物理中有重要应用.

更一般地, 对任意正函数 (陪域是正实数的函数)  $\rho(x)$ , 都可以定义内积

$$\langle f, g \rangle = \int_{-1}^1 f(x) \rho(x) g(x) dx.$$

当  $\rho(x) = \frac{1}{\sqrt{1-x^2}}$  时, 在具有该内积的欧氏空间中对向量  $1, x, x^2, \dots, x^n, \dots$  做 Gram-Schmidt 正交化:

$$\psi_0(x) = 1;$$

$$\psi_1(x) = x - \frac{\langle \psi_0, x \rangle}{\langle \psi_0, \psi_0 \rangle} \psi_0 = x - \frac{\int_{-1}^1 \frac{1x}{\sqrt{1-x^2}} dx}{\int_{-1}^1 \frac{1}{\sqrt{1-x^2}} dx} = x;$$

$$\psi_2(x) = x^2 - \frac{\langle \psi_0, x^2 \rangle}{\langle \psi_0, \psi_0 \rangle} \psi_0 - \frac{\langle \psi_1, x^2 \rangle}{\langle \psi_1, \psi_1 \rangle} \psi_1 = x^2 - \frac{1}{2};$$

$$\psi_3(x) = x^3 - \frac{\langle \psi_0, x^3 \rangle}{\langle \psi_0, \psi_0 \rangle} \psi_0 - \frac{\langle \psi_1, x^3 \rangle}{\langle \psi_1, \psi_1 \rangle} \psi_1 - \frac{\langle \psi_2, x^3 \rangle}{\langle \psi_2, \psi_2 \rangle} \psi_2 = x^3 - \frac{3}{4}x.$$

通过数乘能得到多项式

$$T_n(x) = \frac{1}{2^{n-1}} \psi_n(x), \quad n = 0, 1, \dots,$$

使得  $T_n(1) = 1$ , 称  $T_n$  为 **Chebyshev 多项式**. 可以验证如下递推关系和通项公式:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x),$$

$$T_n(x) = \cos(n \arccos(x)).$$

Chebyshev 多项式在函数逼近和数值分析中有重要应用. ☺

基扩充定理对欧氏空间也有直接推广.

**命题 8.1.12** 有限维欧氏空间中任意正交单位向量组都可以扩充成一组标准正交基.

证. 先把正交单位向量组  $\mathbf{q}_1, \dots, \mathbf{q}_r$  扩充成一组基  $\mathbf{q}_1, \dots, \mathbf{q}_r, \mathbf{a}_{r+1}, \dots, \mathbf{a}_n$ , 再对其应用 Gram-Schmidt 正交化方法, 即得一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_r, \mathbf{q}_{r+1}, \dots, \mathbf{q}_n$ . 注意, 在 Gram-Schmidt 正交化过程中, 前  $r$  个向量保持不变. □

下面讨论有限维欧氏空间上内积与坐标的关系.

给定欧氏空间  $\mathcal{V}$  的一组基  $\mathbf{a}_1, \dots, \mathbf{a}_n$ , 对向量  $\mathbf{x} = x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n$  和向量  $\mathbf{y} = y_1\mathbf{a}_1 + \dots + y_n\mathbf{a}_n$ , 二者内积是

$$\begin{aligned}\langle \mathbf{x}, \mathbf{y} \rangle &= \left\langle \sum_{i=1}^n x_i \mathbf{a}_i, \sum_{j=1}^n y_j \mathbf{a}_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n x_i y_j \langle \mathbf{a}_i, \mathbf{a}_j \rangle \\ &= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}^T \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_1, \mathbf{a}_n \rangle \\ \vdots & & \vdots \\ \langle \mathbf{a}_n, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_n \rangle \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.\end{aligned}$$

矩阵

$$G := \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_1, \mathbf{a}_n \rangle \\ \vdots & & \vdots \\ \langle \mathbf{a}_n, \mathbf{a}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{a}_n \rangle \end{bmatrix},$$

称为内积在基  $\mathbf{a}_1, \dots, \mathbf{a}_n$  下的 **Gram 矩阵** 或者 **度量矩阵**.

因此,  $\langle \mathbf{x}, \mathbf{y} \rangle = \hat{\mathbf{x}}^T G \hat{\mathbf{y}}$ , 其中  $\hat{\mathbf{x}}, \hat{\mathbf{y}}$  分别是两个向量在基  $\mathbf{a}_1, \dots, \mathbf{a}_n$  下的坐标. 易见, 基决定了矩阵  $G$ , 而矩阵  $G$  决定了内积的计算方式. 特别地,  $G = I_n$ , 当且仅当基  $\mathbf{a}_1, \dots, \mathbf{a}_n$  是标准正交基. 此时,  $\langle \mathbf{x}, \mathbf{y} \rangle = \hat{\mathbf{x}}^T \hat{\mathbf{y}} = \langle \hat{\mathbf{x}}, \hat{\mathbf{y}} \rangle$ , 其中后一内积是  $\mathbb{R}^n$  上的标准内积. 可以看到, 标准正交基在计算内积时很方便.

根据内积定义, 度量矩阵是对称正定矩阵. 度量矩阵在基变换下有如下变换规律.

**命题 8.1.13** 给定  $n$  维欧氏空间  $\mathcal{V}$  及其两组基  $\mathbf{e}_1, \dots, \mathbf{e}_n; \mathbf{t}_1, \dots, \mathbf{t}_n$ . 记过渡矩阵为  $T$ , 即  $(\mathbf{t}_1, \dots, \mathbf{t}_n) = (\mathbf{e}_1, \dots, \mathbf{e}_n)T$ . 如果内积在基  $\mathbf{e}_1, \dots, \mathbf{e}_n$  下的度量矩阵为  $G$ , 则该内积在基  $\mathbf{t}_1, \dots, \mathbf{t}_n$  下的度量矩阵为  $T^T G T$ .

证. 设内积在后一组基下的度量矩阵为  $\tilde{G}$ . 由度量矩阵定义, 对任意向量

$$\mathbf{x} = (\mathbf{e}_1, \dots, \mathbf{e}_n)\hat{\mathbf{x}} = (\mathbf{t}_1, \dots, \mathbf{t}_n)\tilde{\mathbf{x}}, \quad \mathbf{y} = (\mathbf{e}_1, \dots, \mathbf{e}_n)\hat{\mathbf{y}} = (\mathbf{t}_1, \dots, \mathbf{t}_n)\tilde{\mathbf{y}},$$

有

$$\begin{aligned}\langle \mathbf{x}, \mathbf{y} \rangle &= \langle (\mathbf{e}_1, \dots, \mathbf{e}_n)\hat{\mathbf{x}}, (\mathbf{e}_1, \dots, \mathbf{e}_n)\hat{\mathbf{y}} \rangle = \hat{\mathbf{x}}^T G \hat{\mathbf{y}} \\ &= \langle (\mathbf{t}_1, \dots, \mathbf{t}_n)\tilde{\mathbf{x}}, (\mathbf{t}_1, \dots, \mathbf{t}_n)\tilde{\mathbf{y}} \rangle = \tilde{\mathbf{x}}^T \tilde{G} \tilde{\mathbf{y}},\end{aligned}$$

代入过渡矩阵, 利用形式上的结合律, 有

$$\begin{aligned}\tilde{\mathbf{x}}^T \tilde{G} \tilde{\mathbf{y}} &= \langle (\mathbf{t}_1, \dots, \mathbf{t}_n)\tilde{\mathbf{x}}, (\mathbf{t}_1, \dots, \mathbf{t}_n)\tilde{\mathbf{y}} \rangle \\ &= \langle (\mathbf{e}_1, \dots, \mathbf{e}_n)T\tilde{\mathbf{x}}, (\mathbf{e}_1, \dots, \mathbf{e}_n)T\tilde{\mathbf{y}} \rangle \\ &= (T\tilde{\mathbf{x}})^T G (T\tilde{\mathbf{y}})\end{aligned}$$



$$= \tilde{x}^T T^T G T \tilde{y}.$$

由向量的任意性, 将  $x, y$  取遍  $e_1, \dots, e_n$ , 即知  $\tilde{G} = T^T G T$ .  $\square$

特别地, 两组标准正交基之间的过渡矩阵是正交矩阵.

**命题 8.1.14** 给定  $n$  维欧氏空间  $\mathcal{V}$  中一组标准正交基  $e_1, \dots, e_n$  和  $\mathbb{R}$  上  $n$  阶方阵  $T$ . 令  $(t_1, \dots, t_n) = (e_1, \dots, e_n)T$ , 则  $t_1, \dots, t_n$  是一组标准正交基, 当且仅当  $T$  是正交矩阵.

证. “ $\Rightarrow$ ”: 若两组基都是标准正交基, 则内积在两组基下的度量矩阵都是  $I_n$ . 根据命题 8.1.13, 就有  $I_n = T^T I_n T = T^T T$ , 亦即  $T$  正交.

“ $\Leftarrow$ ”: 若  $T$  正交, 则  $T$  可逆, 因此  $t_1, \dots, t_n$  也是一组基, 内积在这组基下的度量矩阵是  $T^T I_n T = T^T T = I$ , 因此  $t_1, \dots, t_n$  是标准正交基.  $\square$

最后, 来说明第 3 章中的正交补、正交投影等概念可以推广到一般的欧氏空间. 由于结论及其证明基本与第 3.3 节一致, 这里将略去大部分证明.

**命题 8.1.15** 给定欧氏空间  $\mathcal{V}$ , 设  $\mathcal{M}$  是线性空间  $\mathcal{V}$  的子空间, 则  $\mathcal{M}$  关于  $\mathcal{V}$  上的内积也构成欧氏空间.

证明留给读者.

**定义 8.1.16 (正交补)** 给定欧氏空间  $\mathcal{V}$  的子空间  $\mathcal{M}$ , 称集合

$$\mathcal{M}^\perp := \{a \in \mathcal{V} : a \perp b, \forall b \in \mathcal{M}\}$$

为  $\mathcal{M}$  的正交补.

容易验证, 正交补  $\mathcal{M}^\perp$  是  $\mathcal{V}$  的子空间. 正交补具有如下性质.

**命题 8.1.17** 给定欧氏空间  $\mathcal{V}$  的有限子空间  $\mathcal{M}$ , 则

1.  $\mathcal{M} \oplus \mathcal{M}^\perp = \mathcal{V}$ ;
2.  $\dim \mathcal{M}^\perp = n - \dim \mathcal{M}$ ;
3.  $(\mathcal{M}^\perp)^\perp = \mathcal{M}$ .

证. 只证明第 1 条, 其余留给读者. 对任意  $a \in \mathcal{M} \cap \mathcal{M}^\perp$ ,  $\langle a, a \rangle = 0$ , 立得  $a = 0$ , 因此  $\mathcal{M} \cap \mathcal{M}^\perp = \{0\}$ , 即  $\mathcal{M} + \mathcal{M}^\perp$  是直和. 取  $\mathcal{M}$  的一组标准正交基  $q_1, \dots, q_r$ , 根据基扩充定理, 将其扩充成  $\mathcal{V}$  的一组标准正交基  $q_1, \dots, q_r, q_{r+1}, \dots, q_n$ . 易知  $q_{r+1}, \dots, q_n \in \mathcal{M}^\perp$ , 因此  $\mathcal{M} \oplus \mathcal{M}^\perp = \mathcal{V}$ .  $\square$

**例 8.1.18** 具有标准内积的欧氏空间  $\mathbb{R}^{n \times n}$  中, 设  $\mathcal{M}$  是由对称矩阵构成的子空间, 下面计算  $\mathcal{M}^\perp$ . 对任意对称矩阵  $A$ , 反对称矩阵  $B$ ,

$$\begin{aligned}\langle A, B \rangle &= \text{trace}(B^T A) = -\text{trace}(BA) \\ &= -\text{trace}(AB) = -\text{trace}(A^T B) = -\langle B, A \rangle = -\langle A, B \rangle,\end{aligned}$$

因而  $\langle A, B \rangle = 0$ . 于是, 所有反对称矩阵都属于  $\mathcal{M}^\perp$ . 由维数公式可知,

$$\dim \mathcal{M}^\perp = n^2 - \dim \mathcal{M} = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}.$$

由所有反对称矩阵构成的子空间的维数是  $\frac{n(n-1)}{2}$ , 因此它就是  $\mathcal{M}^\perp$ . ☺

给定有限维欧氏空间  $\mathcal{V}$  的子空间  $\mathcal{M}$ , 根据直和分解  $\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^\perp$  可知, 对任意  $\mathbf{a} \in \mathcal{V}$ , 都有唯一的分解  $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$ , 其中  $\mathbf{a}_1 \in \mathcal{M}, \mathbf{a}_2 \in \mathcal{M}^\perp$ . 因此

$$P_{\mathcal{M}}: \mathbf{a} \mapsto \mathbf{a}_1,$$

是  $\mathcal{V}$  上的线性变换.

**定义 8.1.19** 给定有限维欧氏空间  $\mathcal{V}$  的子空间  $\mathcal{M}$ , 其上线性变换  $P_{\mathcal{M}}$  称为  $\mathcal{V}$  到子空间  $\mathcal{M}$  上的正交投影 (变换), 而  $\mathbf{a}_1 = P_{\mathcal{M}}(\mathbf{a})$  称为向量  $\mathbf{a}$  在  $\mathcal{M}$  上的正交投影.

利用正交投影可得最小距离.

**命题 8.1.20** 给定有限维欧氏空间  $\mathcal{V}$  的子空间  $\mathcal{M}$  和向量  $\mathbf{a} \in \mathcal{V}$ , 而  $\mathbf{a}_1 = P_{\mathcal{M}}(\mathbf{a})$  为  $\mathbf{a}$  在  $\mathcal{M}$  上的正交投影, 则  $\|\mathbf{a} - \mathbf{a}_1\| = \min_{\mathbf{x} \in \mathcal{M}} \|\mathbf{a} - \mathbf{x}\|$ .

**例 8.1.21 (主成分分析)** 回顾第 6.3.3 节. 测量数据  $\mathbf{a}_1, \dots, \mathbf{a}_n$  作为列组成矩阵  $A$ . 测量数据在  $m$  维子空间  $\mathcal{M}$  附近, 该模型就是试图计算该子空间, 确切来讲是一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_m$ . 设这组标准正交基组成矩阵  $Q$ . 主成分分析告诉我们, 设  $A$  的截断 SVD 为  $U_m \Sigma_m V_m^T$ , 则矩阵  $Q = U_m$  使得  $\text{trace}(Q^T A A^T Q) = \|Q^T A\|_F^2$  最大.

现在换用欧氏空间上的语言, 希望找到矩阵  $Q$ , 使得  $(A^T Q, A^T Q)$ , 即  $A^T Q$  在欧氏空间  $\mathbb{R}^{n \times m}$  中范数的平方最大. 这不利于使用正交投影. 我们稍作变形, 注意

$$\text{trace}(Q^T A A^T Q) = \text{trace}(Q Q^T A A^T) = (A A^T, Q Q^T),$$

问题就转换成了  $Q Q^T$  与  $A A^T$  在欧氏空间  $\mathbb{R}^{d \times d}$  中的内积最大, 这等价于  $Q Q^T - A A^T$  在该空间中范数最小. 于是

$$\min_{\substack{Q \text{ 列正交} \\ \text{rank}(Q)=m}} \|Q Q^T - A A^T\|_F = \|U_m U_m^T - A A^T\|_F.$$

**例 8.1.22 (信号处理)** 在信号处理中, 以时间为自变量的函数常称为信号. 不难发现, 物理中的波也是一种信号. 中学物理提起过, 最简单的波是简谐波, 其函数表达式为  $A \sin(\omega t + \varphi)$ , 其中  $A$  是振幅,  $\omega$  是角频率,  $\varphi$  是初始相位. 例 8.1.9 说明, 振幅为 1、角频率是整数、初始相位是 0 或  $\frac{\pi}{2}$  的简谐波是空间中的正交向量组. 人工信号的发射和接收, 出于各种考虑, 一般只限于有限频率有限相位内的波. 不妨假设只能接收到角频率为  $0, 1, 2$ , 相位为  $0, \frac{\pi}{2}$  的波. 因此, 任意信号  $f$ , 在如此接收端, 所接收到的信号  $\hat{f}$ , 只能是原始信号在  $\text{span}(1, \sin x, \cos x, \sin 2x, \cos 2x)$  这一子空间上的正交投影. (注意, 任意欧氏空间都有到其有限维子空间的正交投影, 参见阅读 8.1.4.)

不难计算

$$\begin{aligned} \hat{f} = & \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) dt \right) + \left( \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin x dt \right) \sin x + \left( \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos x dt \right) \cos x + \\ & \left( \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin 2x dt \right) \sin 2x + \left( \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos 2x dt \right) \cos 2x. \end{aligned}$$

这就是在接收端得到的信号.

这一思想稍加推广, 就能得到著名的 **Fourier 变换**, 广泛应用于数学物理、偏微分方程、调和分析以及众多应用领域.  $\odot$

## 习题

**练习 8.1.1** 证明正交向量组线性无关.

**练习 8.1.2** 证明命题 8.1.15.

**练习 8.1.3** 证明命题 8.1.17 第 2-3 条.

**阅读 8.1.4 (无限维欧氏空间中的正交投影)** 对欧氏空间  $\mathcal{V}$  的有限维子空间  $\mathcal{M}$ , 命题 8.1.17 第 1 条也成立:  $\mathcal{M} \oplus \mathcal{M}^\perp = \mathcal{V}$ .

事实上, 对任意  $\mathbf{a} \in \mathcal{M} \cap \mathcal{M}^\perp$ ,  $\langle \mathbf{a}, \mathbf{a} \rangle = 0$ , 于是  $\mathbf{a} = \mathbf{0}$ , 这就说明  $\mathcal{M} \cap \mathcal{M}^\perp = \{\mathbf{0}\}$ , 即  $\mathcal{M} + \mathcal{M}^\perp$  是直和. 根据定理 8.1.10,  $\mathcal{M}$  存在一组标准正交基, 记为  $\mathbf{q}_1, \dots, \mathbf{q}_r$ . 对任意  $\mathbf{b} \in \mathcal{V}$ , 定义向量  $\mathbf{b}_1 := \langle \mathbf{b}, \mathbf{q}_1 \rangle \mathbf{q}_1 - \dots - \langle \mathbf{b}, \mathbf{q}_r \rangle \mathbf{q}_r \in \mathcal{M}$ ,  $\mathbf{b}_2 := \mathbf{b} - \mathbf{b}_1$ . 易见  $\langle \mathbf{q}_i, \mathbf{b}_2 \rangle = 0$ , 于是  $\mathbf{b}_2 \perp \mathcal{M}$ , 即得  $\mathbf{b}_2 \in \mathcal{M}^\perp$ . 因此  $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2 \in \mathcal{M} \oplus \mathcal{M}^\perp$ . 这就说明  $\mathcal{M} \oplus \mathcal{M}^\perp = \mathcal{V}$ .

定义 8.1.19 显然可以推广成无限维欧氏空间到其有限维子空间上的正交投影.

**练习 8.1.5** 计算例 8.1.11 中 Legendre 多项式作为多项式空间中向量的范数, 并说明该向量组不是正交单位向量组.

**练习 8.1.6** 在  $\mathbb{R}^2$  中, 对任意  $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ ,  $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ , 定义二元函数

$$f(\mathbf{a}, \mathbf{b}) = a_1 b_1 - a_1 b_2 - a_2 b_1 + 2a_2 b_2.$$

判断  $f(\mathbf{a}, \mathbf{b})$  是否是  $\mathbb{R}^2$  上的一个内积.

**练习 8.1.7** 证明,  $\mathbb{R}^n$  上的二元函数  $\mathbf{b}^T A \mathbf{a}$  定义了一个内积, 当且仅当  $A$  是对称正定矩阵.

**练习 8.1.8** 证明, 矩阵空间  $\mathbb{R}^{m \times n}$  关于如下二元函数构成欧氏空间:  $\langle A, B \rangle = \text{trace}(B^T A)$ .

**练习 8.1.9** 设 3 维欧氏空间  $\mathcal{V}$  的一组基是  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ , 它的度量矩阵是  $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 10 & -2 \\ 1 & -2 & 2 \end{bmatrix}$ . 求  $\mathcal{V}$  的一组标准正交基.

**练习 8.1.10** 设  $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$  是 3 维欧氏空间  $\mathcal{V}$  的一组标准正交基, 令

$$\mathbf{b}_1 = \frac{1}{3}(2\mathbf{q}_1 - \mathbf{q}_2 + 2\mathbf{q}_3), \quad \mathbf{b}_2 = \frac{1}{3}(2\mathbf{q}_1 + 2\mathbf{q}_2 - \mathbf{q}_3), \quad \mathbf{b}_3 = \frac{1}{3}(\mathbf{q}_1 - 2\mathbf{q}_2 - 2\mathbf{q}_3).$$

证明,  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  也是  $\mathcal{V}$  的一组标准正交基.

**练习 8.1.11** 设  $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4, \mathbf{q}_5$  是 5 维欧氏空间  $\mathcal{V}$  的一组标准正交基, 令

$$\mathbf{a}_1 = \mathbf{q}_1 + \mathbf{q}_5, \quad \mathbf{a}_2 = \mathbf{q}_1 - \mathbf{q}_2 + \mathbf{q}_4, \quad \mathbf{a}_3 = 2\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3.$$

求  $\text{span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$  的一组标准正交基.

**练习 8.1.12** 设  $\mathcal{M}$  是  $n$  维欧氏空间  $\mathcal{V}$  的子空间, 在  $\mathcal{M}$  中取一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_r$ . 证明,  $\mathcal{V}$  中任意向量  $\mathbf{a}$  在  $\mathcal{M}$  上的正交投影  $\mathbf{a}_1 = \sum_{i=1}^r \langle \mathbf{a}, \mathbf{q}_i \rangle \mathbf{q}_i$ .

**练习 8.1.13** 设  $\mathcal{M}$  是欧氏空间  $\mathcal{V}$  的子空间,  $\mathbf{P}_{\mathcal{M}}$  是  $\mathcal{V}$  到  $\mathcal{M}$  上的正交投影. 证明,

$$\langle \mathbf{P}_{\mathcal{M}}(\mathbf{a}), \mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{P}_{\mathcal{M}}(\mathbf{b}) \rangle, \quad \forall \mathbf{a}, \mathbf{b} \in \mathcal{V}.$$

**练习 8.1.14** 设  $f$  是欧氏空间  $\mathcal{V}$  内的线性函数, 证明, 存在唯一的固定向量  $\mathbf{b} \in \mathcal{V}$ , 使得对任意  $\mathbf{a} \in \mathcal{V}$ , 都有  $f(\mathbf{a}) = \langle \mathbf{a}, \mathbf{b} \rangle$ .

## 8.2 欧氏空间上的线性映射

首先讨论欧氏空间之间的线性映射.

**定义 8.2.1 (伴随映射)** 给定欧氏空间  $\mathcal{U}, \mathcal{V}$ , 内积分别为  $\langle \cdot, \cdot \rangle_{\mathcal{U}}, \langle \cdot, \cdot \rangle_{\mathcal{V}}$ . 如果线性映射  $f \in \text{Hom}(\mathcal{U}, \mathcal{V}), g \in \text{Hom}(\mathcal{V}, \mathcal{U})$  满足对任意  $\mathbf{x} \in \mathcal{U}, \mathbf{y} \in \mathcal{V}$ , 都有  $\langle f(\mathbf{x}), \mathbf{y} \rangle_{\mathcal{V}} = \langle \mathbf{x}, g(\mathbf{y}) \rangle_{\mathcal{U}}$  成立, 则称  $g$  是  $f$  的共轭映射或伴随映射, 记为  $g = f^*$ .

特别地, 如果  $\mathcal{U} = \mathcal{V}$ , 则  $f^*$  又称为  $f$  的伴随变换.

伴随映射满足:  $(f^*)^* = f$ ; 若  $g = f^*$ , 则  $f = g^*$ ;  $(gf)^* = f^*g^*$ . 请读者自己验证.

**命题 8.2.2** 当  $\mathcal{U} = \mathbb{R}^n, \mathcal{V} = \mathbb{R}^m$  时, 由  $A \in \mathbb{R}^{m \times n}$  确定的线性映射  $\mathbf{A}: \mathbf{x} \mapsto A\mathbf{x}$  的伴随映射是  $\mathbf{A}^*: \mathbf{y} \mapsto A^T \mathbf{y}$ .

证. 设伴随映射由矩阵  $B \in \mathbb{R}^{n \times m}$  确定, 则  $\langle Ax, y \rangle = \langle x, By \rangle$ . 由内积定义, 对任意  $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ ,  $y^T Ax = (By)^T x = y^T B^T x$ , 由此即得  $A = B^T, B = A^T$ .  $\square$

因此, 伴随映射是转置矩阵的推广. 一般地, 有如下结论.

**命题 8.2.3** 给定欧氏空间  $\mathcal{U}, \mathcal{V}$  及各自一组标准正交基  $u_1, \dots, u_n; v_1, \dots, v_m$ . 如果线性映射  $f: \mathcal{U} \rightarrow \mathcal{V}$  在两组基下的矩阵是  $F$ , 则伴随映射  $f^*$  在两组基下的矩阵是  $F^T$ .

证. 设  $f$  的矩阵为  $F = [f_{ij}]_{m \times n}$ ,  $f^*$  的矩阵为  $G = [g_{ji}]_{n \times m}$ . 由伴随映射的定义,  $\langle f(u_i), v_j \rangle = \langle u_i, f^*(v_j) \rangle$ . 另一方面,

$$\begin{aligned}\langle f(u_i), v_j \rangle &= \langle f_{1i}v_1 + \dots + f_{mi}v_m, v_j \rangle = f_{ji}, \\ \langle u_i, f^*(v_j) \rangle &= \langle u_i, g_{1j}u_1 + \dots + g_{nj}u_n \rangle = g_{ij}.\end{aligned}$$

因此  $f_{ji} = g_{ij}$ , 即  $G = F^T$ .  $\square$

注意, 当给定基不是标准正交基时, 伴随映射的矩阵并不一定是原矩阵的转置.

其次讨论欧氏空间上的线性映射的矩阵在标准正交基的基变换下的变化规律. 在一般情形下, 线性映射的矩阵在基变换下的规律见命题 7.5.6. 当给定基都取作标准正交基时, 过渡矩阵是正交矩阵 (参见命题 8.1.14), 立得如下结论.

**命题 8.2.4** 给定欧氏空间  $\mathcal{U}, \mathcal{V}$ , 和  $\mathcal{U}$  的两组基  $e_1, \dots, e_n; v_1, \dots, v_n$  与  $\mathcal{V}$  的两组基  $i_1, \dots, i_m; u_1, \dots, u_m$ . 记二者的过渡矩阵分别为  $V, U$ , 即

$$(v_1, \dots, v_n) = (e_1, \dots, e_n)V, \quad (u_1, \dots, u_m) = (i_1, \dots, i_m)U.$$

如果  $\mathcal{U}$  到  $\mathcal{V}$  的线性映射  $f$  在基  $e_1, \dots, e_n; i_1, \dots, i_m$  下的矩阵为  $F$ , 则该映射在基  $v_1, \dots, v_n; u_1, \dots, u_m$  下的矩阵为  $U^T F V$ .

命题 8.2.4 说明, 计算  $m \times n$  矩阵  $A$  奇异值分解, 本质上就是为  $\mathbb{R}^m$  和  $\mathbb{R}^n$  各找一组标准正交基, 使得  $A$  对应的线性映射在给定基下的矩阵具有如下形式  $\begin{bmatrix} \Sigma_r & O \\ O & O \end{bmatrix}$ , 其中  $\Sigma$  是正定的对角矩阵.

当考虑线性变换, 即  $\mathcal{U} = \mathcal{V}$  时, 直接有如下推论.

**命题 8.2.5** 给定欧氏空间  $\mathcal{V}$  及其两组标准正交基  $e_1, \dots, e_n; q_1, \dots, q_n$ , 记过渡矩阵为  $Q$ , 即  $(q_1, \dots, q_n) = (e_1, \dots, e_n)Q$ . 如果  $\mathcal{V}$  上线性变换  $f$  在基  $e_1, \dots, e_n$  下的矩阵为  $F$ , 那么  $f$  在基  $q_1, \dots, q_n$  下的矩阵为  $Q^{-1} F Q = Q^T F Q$ .

这说明, 欧氏空间做标准正交基的基变换, 其上线性变换的矩阵就做正交相似变换. 事实上, 有如下类似于命题 7.5.14 的结论.

**命题 8.2.6** 数域  $\mathbb{R}$  上两个  $n$  阶方阵  $A, B$  正交相似, 当且仅当  $A, B$  是  $n$  维欧氏空间  $\mathcal{V}$  上某个线性变换在两组标准正交基下的矩阵.

**命题 8.2.6** 说明, 把  $n$  阶对称矩阵  $A$  用正交矩阵对角化, 本质上就是为  $\mathbb{R}^n$  找到一组标准正交基, 使得  $A$  对应的线性变换在这组基下的矩阵是对角矩阵.

下面重点讨论欧氏空间上的两类特殊的线性变换. 先来看对称变换.

**定义 8.2.7 (对称变换)** 给定欧氏空间  $\mathcal{V}$  中的线性变换  $f$ , 如果对任意  $x, y \in \mathcal{V}$ , 都有  $\langle f(x), y \rangle = \langle x, f(y) \rangle$ , 则称  $f$  为  $\mathcal{V}$  上的一个**自伴变换**或**对称变换**.

**命题 8.2.8** 给定  $n$  维欧氏空间  $\mathcal{V}$  上线性变换  $f$ ,

1.  $f$  是对称变换, 当且仅当  $f = f^*$ .
2.  $f$  是对称变换, 当且仅当  $f$  在任意标准正交基下的矩阵都是对称矩阵.

证. 第 1 条: 充分性显然, 下证必要性. 由伴随变换的定义, 对任意  $x, y \in \mathcal{V}$ , 都有  $\langle x, f^*(y) \rangle = \langle x, f(y) \rangle$ . 取定  $\mathcal{V}$  的一组标准正交基  $v_1, \dots, v_n$ , 则

$$f^*(y) = \langle v_1, f^*(y) \rangle v_1 + \dots + \langle v_n, f^*(y) \rangle v_n = \langle v_1, f(y) \rangle v_1 + \dots + \langle v_n, f(y) \rangle v_n = f(y).$$

因此  $f^* = f$ .

第 2 条: 设  $f$  在某组标准正交基下的矩阵为  $A$ , 由命题 8.2.3 可知,  $f^*$  的矩阵是  $A^T$ . 利用线性映射与矩阵之间的一一对应 (定理 7.5.3), 即有  $f = f^* \Leftrightarrow A = A^T$ , 立得结论.  $\square$

**例 8.2.9** 1. 具有标准内积的欧氏空间  $\mathbb{R}^{n \times n}$  上, 取转置这一线性变换  $f: A \mapsto A^T$  是对称变换. 先计算其伴随变换. 对任意  $A, B \in \mathbb{R}^{n \times n}$ ,

$$\langle f(A), B \rangle = \text{trace}(B^T A^T) = \text{trace}(AB) = \text{trace}(BA) = \langle A, B^T \rangle,$$

因此  $f^*(B) = B^T$ , 即  $f = f^*$ .

2. 周期函数组成的集合  $C_T^\infty(\mathbb{R}) = \{f \in C^\infty(\mathbb{R}) \mid f(x) = f(x+T), \forall x \in \mathbb{R}\}$  是具有内积  $\langle f, g \rangle = \int_0^T f(x)g(x) dx$  的欧氏空间 (为什么?). 考虑其上的求导算子

$$\begin{aligned} D: C_T^\infty(\mathbb{R}) &\rightarrow C_T^\infty(\mathbb{R}), \\ f(x) &\mapsto f'(x). \end{aligned}$$

计算其伴随变换. 对任意  $f, g \in C_T^\infty(\mathbb{R})$ ,

$$\langle D(f), g \rangle = \int_0^T f'(x)g(x) dx = [f(x)g(x)] \Big|_0^T - \int_0^T f(x)g'(x) dx = \langle f, -D(g) \rangle,$$

因此  $D^* = -D$ ,  $D$  不是对称变换.

3. 光滑函数空间  $C^\infty[a, b]$  是具有内积  $\langle f, g \rangle = \int_a^b f(x)g(x) dx$  的欧氏空间, 它有一个子空间  $C_0^\infty[a, b] = \{f \in C^\infty[a, b] \mid f(a) = f(b) = 0\}$  (为什么是子空间?). 而  $\mathcal{V}_{a,b} = \{f \in C_0^\infty[a, b] \mid f^{(2k)} \in C_0^\infty[a, b], k = 1, 2, \dots\}$  是  $C_0^\infty[a, b]$  的子空间 (为什么?). 考虑其上的二阶求导算子

$$\begin{aligned} D^2: \mathcal{V}_{a,b} &\rightarrow \mathcal{V}_{a,b}, \\ f(x) &\mapsto f''(x). \end{aligned}$$

利用分部积分可得, 其伴随变换是  $(D^2)^* = D^2$ , 立得  $D^2$  是对称变换.  $\odot$

**例 8.2.10** 给定  $n$  维欧氏空间  $\mathcal{V}$  的子空间  $\mathcal{M}$ , 不难验证正交投影  $P_{\mathcal{M}}$  是  $\mathcal{V}$  上的对称变换 (练习 8.1.13). 分别取  $\mathcal{M}$  和  $\mathcal{M}^\perp$  的一组标准正交基, 其并集就是  $\mathcal{V}$  的一组标准正交基, 则  $P_{\mathcal{M}}$  在该组基下的矩阵是对角矩阵  $\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$ , 其中  $r = \dim \mathcal{M}$ . 这显然是对称矩阵.  $\odot$

给定线性变换, 通过换基寻找其矩阵最简单的形式, 就是计算其矩阵在正交相似这个等价关系下的标准形. 实对称矩阵正交相似于对角矩阵, 因此有如下结论.

**定理 8.2.11** 给定  $n$  维欧氏空间  $\mathcal{V}$ , 对其上任意对称变换  $f$ , 都存在  $\mathcal{V}$  的一组标准正交基, 使得  $f$  在该组基下的矩阵是对角矩阵.

**例 8.2.12** 考虑例 8.2.9 中的对称变换的矩阵.

1. 欧氏空间  $\mathbb{R}^{n \times n}$  上的对称变换  $f: A \mapsto A^T$ , 其特征值, 由于  $f^2 = \text{id}$ , 因此  $f$  的特征值只能是  $\pm 1$ . 其特征向量,  $f(A) = A$ , 当且仅当  $A$  是对称矩阵;  $f(A) = -A$ , 当且仅当  $A$  是反对称矩阵. 由例 8.1.18 可知, 由对称矩阵和反对称矩阵构成的特征子空间互为正交补. 在两个特征子空间中分别取一组标准正交基, 合并成  $\mathbb{R}^{n \times n}$  的一组标准正交基, 则对称变换  $f$  在该组基下的矩阵是对角矩阵. 例如, 取  $n = 2$ , 则对称变换  $f$  在基  $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$  下的矩阵就是  $\text{diag}(1, 1, 1, -1)$ .
2. 欧氏空间  $\mathcal{V}_{-\pi, \pi}$  上的对称变换  $D^2: f(x) \mapsto f''(x)$ , 易得  $D^2(\sin nx) = -n^2 \sin nx$ , 即  $-n^2$  是  $D^2$  的特征值, 而  $\sin nx$  是属于  $-n^2$  的特征向量. 由例 8.1.9 可知,  $\{\sin nx \mid n \geq 1\}$  是正交向量组. 因此, 子空间  $\text{span}(\sin x, \sin 2x, \dots, \sin nx)$  上的线性变换  $D^2$ , 在标准正交基  $\frac{1}{\sqrt{\pi}} \sin x, \frac{1}{\sqrt{\pi}} \sin 2x, \dots, \frac{1}{\sqrt{\pi}} \sin nx$  下的矩阵是对角矩阵  $\text{diag}(-1, -4, \dots, -n^2)$ .

事实上, 数学分析知识说明,  $\{\frac{1}{\sqrt{\pi}} \sin nx \mid n \geq 1\}$  是无穷维欧氏空间  $\mathcal{V}_{-\pi, \pi}$  的一组标准正交基, 而对称变换  $D^2$  的所有特征值是  $-1, -4, \dots, -n^2, \dots$ .  $\odot$

再来看正交变换.

**定义 8.2.13 (正交变换)** 给定欧氏空间  $\mathcal{V}$  中的线性变换  $f$ , 如果对任意  $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ , 都有  $\langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$ , 则称  $f$  为  $\mathcal{V}$  上的一个**正交变换**.

正交变换有如下等价描述, 这可以看作针对正交矩阵的命题 3.2.4 的推广, 证明留给读者.

**命题 8.2.14** 给定  $n$  维欧氏空间  $\mathcal{V}$  上线性变换  $f$ , 以下叙述等价:

1.  $f$  是正交变换, 即, 对任意  $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ , 都有  $\langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$ ;
2.  $f$  为保距变换, 即, 对任意  $\mathbf{x} \in \mathcal{V}$ , 都有  $\|f(\mathbf{x})\| = \|\mathbf{x}\|$ ;
3.  $f$  把  $\mathcal{V}$  的标准正交基映射成标准正交基.

如下结论类似于命题 8.2.8, 证明也类似, 留给读者.

**命题 8.2.15** 给定  $n$  维欧氏空间  $\mathcal{V}$  上线性变换  $f$ ,

1.  $f$  是正交变换, 当且仅当  $ff^* = f^*f = \text{id}_{\mathcal{V}}$ ;
2.  $f$  是正交变换, 当且仅当  $f$  在任意标准正交基下的矩阵都是正交矩阵.

然后讨论正交变换的矩阵通过换基能得到的最简单的形式, 即正交矩阵在正交相似这个等价关系下的标准形. 先考虑正交矩阵的特征值和特征向量.

**命题 8.2.16** 给定正交矩阵  $Q$ , 如果  $\lambda \in \mathbb{C}$  是其特征值, 则  $|\lambda| = 1$ . 特别地, 正交矩阵的实特征值只能是  $\pm 1$ .

证. 设  $Q\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \in \mathbb{C}^n$ , 取共轭转置得  $\bar{\mathbf{x}}^T Q^T = \bar{\lambda} \bar{\mathbf{x}}^T$ , 因此

$$\bar{\mathbf{x}}^T \mathbf{x} = \bar{\mathbf{x}}^T Q^T Q \mathbf{x} = \bar{\lambda} \lambda \bar{\mathbf{x}}^T \mathbf{x} = |\lambda|^2 \bar{\mathbf{x}}^T \mathbf{x}.$$

而  $\mathbf{x} \neq \mathbf{0}$ ,  $\bar{\mathbf{x}}^T \mathbf{x} = |x_1|^2 + \cdots + |x_n|^2 > 0$ , 因此  $|\lambda| = 1$ . □

**命题 8.2.17** 给定正交矩阵  $Q$ , 如果  $\lambda = \cos \theta + i \sin \theta$  是其特征值, 其中  $\theta$  不是平角的倍数, 而  $\mathbf{z} = \mathbf{x} - i\mathbf{y}$  是属于  $\lambda$  的特征向量, 其中  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , 则  $\|\mathbf{x}\| = \|\mathbf{y}\| \neq 0, \mathbf{x}^T \mathbf{y} = 0$ .

证. 由条件知  $Q\mathbf{z} = \lambda\mathbf{z}$ , 取转置可得  $\mathbf{z}^T Q^T = \lambda \mathbf{z}^T$ , 因此  $\mathbf{z}^T \mathbf{z} = \mathbf{z}^T Q^T Q \mathbf{x} = \lambda^2 \mathbf{z}^T \mathbf{z}$ . 而  $\lambda^2 \neq 1$ , 故  $\mathbf{z}^T \mathbf{z} = 0$ , 即  $(\mathbf{x} - i\mathbf{y})^T (\mathbf{x} - i\mathbf{y}) = \mathbf{x}^T \mathbf{x} - \mathbf{y}^T \mathbf{y} + 2i\mathbf{x}^T \mathbf{y} = 0$ , 立得结论. □



事实上, 由于  $Q(x - yi) = (\cos \theta + i \sin \theta)(x - yi)$ , 比较实部和虚部可得:

$$Qx = \cos \theta x + \sin \theta y, \quad Qy = -\sin \theta x + \cos \theta y.$$

根据命题 8.2.17,  $x, y$  正交且长度相等, 因此单位化后得到的两个正交的单位向量  $x_1 = \frac{x}{\|x\|}, x_2 = \frac{y}{\|y\|}$ , 满足

$$Q \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (8.2.1)$$

这意味着, 在二维子空间  $\text{span}(x_1, x_2)$  上, 线性变换  $Q: x \mapsto Qx$  是一个旋转.

下面就能得到正交矩阵的正交相似标准形.

**定理 8.2.18 (正交矩阵的实 Schur 分解)** 对  $n$  阶正交矩阵  $Q$ , 存在  $n$  阶正交矩阵  $X$ , 使得

$$X^T Q X = J = \begin{bmatrix} I_{n_1} & & & & \\ & -I_{n_2} & & & \\ & & \cos \theta_1 & -\sin \theta_1 & \\ & & \sin \theta_1 & \cos \theta_1 & \\ & & & \ddots & \\ & & & & \cos \theta_s & -\sin \theta_s \\ & & & & \sin \theta_s & \cos \theta_s \end{bmatrix}, \quad (8.2.2)$$

其中  $\theta_i$  不是平角的倍数, 即  $J$  是分块对角矩阵, 其中对角块或者是一阶方阵 1, 或者是一阶方阵  $-1$ , 或者是二阶旋转矩阵. 这种分块对角矩阵称为正交矩阵  $Q$  的实相似标准形, 且除了这些对角块的排列次序外,  $J$  被  $Q$  唯一确定.

证. 采用数学归纳法. 当  $n = 1$  时, 显然. 假设命题对  $1, \dots, n-1$  阶正交矩阵成立, 考虑  $n$  阶正交矩阵.

如果  $Q$  有实特征值  $\lambda = \pm 1$ , 则存在单位向量  $x_1$ , 满足  $Qx_1 = \lambda x_1$ . 将  $x_1$  扩充成  $\mathbb{R}^n$  的一组标准正交基  $x_1, x_2, \dots, x_n$ , 记  $X_1 = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} x_1 & X_{12} \end{bmatrix}$ , 则

$$QX_1 = Q \begin{bmatrix} x_1 & X_{12} \end{bmatrix} = \begin{bmatrix} x_1 & X_{12} \end{bmatrix} \begin{bmatrix} \lambda & y^T \\ \mathbf{0} & Q_2 \end{bmatrix} = X_1 Q_1.$$

由  $Q$  正交,  $X_1$  正交可知,  $Q_1 = \begin{bmatrix} \lambda & y^T \\ \mathbf{0} & Q_2 \end{bmatrix}$  也正交. 因此  $y^T = \mathbf{0}^T$ , 且  $Q_1 = \begin{bmatrix} \lambda & \\ & Q_2 \end{bmatrix}$ ,  $Q_2$  是  $n-1$  阶正交矩阵. 由归纳假设, 存在  $n-1$  阶正交矩阵  $X_2$ , 使得  $X_2^{-1} Q_2 X_2 = J_2$

为满足条件的分块对角矩阵. 于是

$$Q_1 \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix} = \begin{bmatrix} \lambda & \\ & Q_2 \end{bmatrix} \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix} = \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix} \begin{bmatrix} \lambda & \\ & J_2 \end{bmatrix}.$$

令  $X = X_1 \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix}$ ,  $J = \begin{bmatrix} \lambda & \\ & J_2 \end{bmatrix}$ , 则

$$QX = QX_1 \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix} = X_1 Q_1 \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix} = X_1 \begin{bmatrix} 1 & \\ & X_2 \end{bmatrix} J = XJ,$$

$J$  满足条件, 而  $X$  是正交矩阵.

如果  $Q$  有非实数特征值  $\cos \theta + i \sin \theta$ , 由命题 8.2.17 可知, 存在两个单位正交向量  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ , 满足 (8.2.1). 将  $\mathbf{x}_1, \mathbf{x}_2$  扩充成  $\mathbb{R}^n$  的一组标准正交基, 重复前一情形的证明, 可得结论.

唯一性可由特征多项式得到. □

从定理 8.2.18 中容易看到, 正交矩阵  $Q$  有分解  $Q = XJX^T$ , 称为  $Q$  的实 Schur 分解.

再回到正交变换, 有如下推论.

**定理 8.2.19** 给定  $n$  维欧氏空间  $\mathcal{V}$ , 对其上任意正交变换  $f$ , 都存在  $\mathcal{V}$  的一组标准正交基, 使得  $f$  在该组基下的矩阵形如 (8.2.2) 中的  $J$ .

最后, 我们来讨论正交矩阵的几何性质.

注 3.2.11, 即如下推论, 可以利用正交矩阵的实 Schur 分解来证明.

**推论 8.2.20** 任意  $n$  阶正交矩阵  $Q$  都可以分解成不超过  $n$  个反射矩阵的乘积.

证. 先说明反射矩阵经过正交相似变换得到的矩阵还是反射. 反射矩阵  $H_v = I_n - 2vv^T$  满足  $H_v v = -v$ , 对任意正交矩阵  $X$ ,  $(X^T H_v X)(X^T v) = X^T H_v v = -X^T v$ , 因此  $X^T H_v X = H_w$ , 其中  $w = X^T v$ .

一阶正交矩阵只能是  $\pm 1$ . 由例 3.2.5 可知, 二阶正交矩阵或者是旋转, 或者是反射; 且任意旋转都可以写成两个反射的乘积.

根据定理 8.2.18, 设  $Q$  的相似标准形是  $J$ . 由于  $J$  是分块对角矩阵, 易知有分解  $J = J_1 \cdots J_{n_2} K_1 \cdots K_s$ , 其中  $J_i$  是保留  $J$  中对角线上的第  $i$  个  $-1$ , 而其余对角块都变成一阶或二阶单位矩阵的分块对角矩阵,  $K_j$  是保留  $J$  中对角线上的第  $j$  个二阶旋转矩阵, 而其余对角块都变成一阶或二阶单位矩阵的分块对角矩阵. 由二阶正交矩阵的讨论可知, 每个  $K_j$  都可以写成两个反射的乘积; 而每个  $J_i$  都是反射矩阵. 因此  $J$  可以写成  $l = n_2 + 2s$  个反射的乘积:  $J = J_1 \cdots J_l$ . 由  $n_1 + l = n$  可知  $l \leq n$ . 根据  $X^T Q X = J$ ,  $Q = X J X^T = (X J_1 X^T) \cdots (X J_l X^T)$ , 其中每个  $X J_i X^T$  都是反射矩阵. □

正交矩阵的实 Schur 分解还能用来讨论正交矩阵的几何分类.

**例 8.2.21 (二阶正交矩阵)** 根据定理 8.2.18, 任意二阶正交矩阵正交相似且仅相似于下列矩阵之一 (其中  $\theta \neq k\pi$ ):

$$I_2 = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}, H_0 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}, -I_2 = \begin{bmatrix} -1 & \\ & -1 \end{bmatrix}, R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

这说明, 任意  $\mathbb{R}^2$  上的保距变换, 通过选择适当的直角坐标系, 其矩阵在四者中必居其一. 其中  $I_2$  是恒同, 也可以理解为转角为  $2k\pi$  的旋转;  $H_0$  是以  $x$  轴为反射轴的反射;  $-I_2$  是中心对称, 也可以理解为转角为  $(2k+1)\pi$  的旋转;  $R_\theta$  是转角为  $\theta$  的旋转. 综上,  $\mathbb{R}^2$  上的保距变换, 只能是旋转或反射, 这与例 3.2.5 一致.

另外, 不难看到  $\det(I_2) = \det(-I_2) = \det(R_\theta) = 1, \det(H_0) = -1$ , 因此旋转就是变积系数为 1 的保距变换, 反射就是变积系数为  $-1$  的保距变换.

几何上, 常常不将恒同称为旋转, 这时  $\mathbb{R}^2$  上的保距变换除恒同外有旋转、反射两种. ☺

**例 8.2.22 (三阶正交矩阵)** 根据定理 8.2.18, 任意三阶正交矩阵正交相似且仅相似于下列矩阵之一 (其中  $\theta \neq k\pi$ ):

$$I_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}, H_z = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix}, G_{k\pi; y, z} = \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix}, -I_3 = \begin{bmatrix} -1 & & \\ & -1 & \\ & & -1 \end{bmatrix},$$

$$G_{\theta; y, z} = \begin{bmatrix} 1 & & \\ & \cos \theta & -\sin \theta \\ & \sin \theta & \cos \theta \end{bmatrix}, H_x G_{\theta; y, z} = \begin{bmatrix} -1 & & \\ & \cos \theta & -\sin \theta \\ & \sin \theta & \cos \theta \end{bmatrix}.$$

这说明, 任意  $\mathbb{R}^3$  上的保距变换, 通过选择适当的直角坐标系, 其矩阵在六者中必居其一. 其中  $I_3$  是恒同;  $H_z$  是以  $z$  轴为法向 (以  $xOy$  平面为反射面) 的反射;  $I_3, G_{k\pi; y, z}, G_{\theta; y, z}$  分别是以  $x$  轴为转轴、转角为  $k\pi, \theta$  的旋转;  $-I_3 = H_x G_{k\pi; y, z}, H_x G_{\theta; y, z}$  是以  $x$  轴为转轴的旋转与以  $x$  轴为法向的反射的复合.

一般地, 以某直线为转轴的旋转与以同一直线为法向的反射的复合变换称为**瑕旋转**, 如图 8.2.1 所示. 反射是特殊的瑕旋转.

综上,  $\mathbb{R}^3$  上的保距变换, 只能是旋转或瑕旋转. 观察行列式可知, 旋转就是变积系数为 1 的保距变换, 瑕旋转就是变积系数为  $-1$  的保距变换.

几何上, 常常不将恒同称为旋转, 类似地, 反射也不称为瑕旋转, 这时,  $\mathbb{R}^3$  上的保距变换除恒同外有旋转、反射、瑕旋转三种. ☺

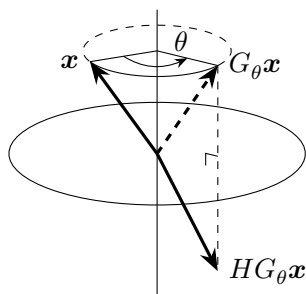


图 8.2.1: 瑕旋转

## 习题

**练习 8.2.1 (线性变换的范数)** 类似于定义 6.3.6, 欧氏空间  $\mathcal{V}$  上向量的范数可以自然地诱导出  $\mathcal{V}$  上线性变换的范数, 即  $\inf_{x \neq 0} \frac{\|f(x)\|}{\|x\|}$  称为  $f$  的范数, 记为  $\|f\|$ . 注意  $\|f\| = +\infty$  有可能成立.

对练习 7.3.8 中的线性变换  $D$  和  $A$ , 证明  $\|D\|\|A\| \geq \frac{1}{2}$ .

这是量子力学中 Heisenberg 不确定性原理的一个简化模型.

**练习 8.2.2** 证明命题 8.2.14.

**练习 8.2.3** 证明命题 8.2.15.

**练习 8.2.4** 求下列正交矩阵的正交相似标准形:

$$1. \frac{1}{3} \begin{bmatrix} 2 & -1 & 2 \\ 2 & 2 & -1 \\ -1 & 2 & 2 \end{bmatrix}.$$

$$2. \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix}.$$

**练习 8.2.5** 设  $q$  是  $n$  维欧氏空间  $\mathcal{V}$  中的单位向量,  $P$  是  $\mathcal{M}$  向  $\text{span}(q)$  上的正交投影. 令  $Q = I - 2P$  是  $\mathcal{V}$  上的线性变换. 求证:

$$1. P(a) = \langle a, q \rangle q.$$

2.  $Q$  是正交变换, 且其正交相似标准形是  $\text{diag}(-1, 1, \dots, 1)$ . 称  $Q$  为关于超平面  $\text{span}(q)^\perp$  的反射.

**练习 8.2.6** 设  $Q$  是  $n$  维欧氏空间  $\mathcal{V}$  上的正交变换,  $\lambda_0 = 1$  是一个特征值, 对应的特征子空间  $\mathcal{V}_{\lambda_0=1}$  的维数是  $n-1$ . 证明,  $Q$  是反射.

**练习 8.2.7** 证明伴随映射的基本结论:

$$1. \text{ 若 } g = f^*, \text{ 则 } f = g^*.$$

$$2. (g \circ f)^* = f^* \circ g^*.$$

**练习 8.2.8** 欧氏空间  $\mathcal{V}$  上的线性变换  $f$ , 如果满足  $ff^* = f^*f$ , 则称  $f$  是一个正规变换. 验证对称变换和正交变换都是正规变换.

实矩阵  $A$ , 如果满足  $AA^T = A^T A$ , 则称  $A$  是一个正规矩阵. 验证对称、反对称、正交矩阵都是正规矩阵.

## 8.3 酉空间

如何为复数域  $\mathbb{C}$  上的线性空间  $\mathcal{V}$  定义内积呢?

从  $\mathcal{V}$  到  $\mathbb{C}$  二元函数  $\langle \cdot, \cdot \rangle$  不可能同时满足对称性、双线性性和正定性. 例如, 利用双线性性,  $\langle \mathbf{i}\mathbf{a}, \mathbf{i}\mathbf{a} \rangle = \mathbf{i}^2 \langle \mathbf{a}, \mathbf{a} \rangle = -\langle \mathbf{a}, \mathbf{a} \rangle$ ; 当  $\mathbf{a} \neq \mathbf{0}$  时, 这与正定性矛盾. 看来为定义  $\mathbb{C}$  上线性空间的内积, 我们不得不放弃一些性质. 注意双线性性事实上只需要对第一变量线性. 线性性非常必要 (不然线性代数无以应用). 正定性也非常必要, 这是能够从内积得到长度和距离的必要条件. 放松对称性, 有如下定义.

**定义 8.3.1 (内积)** 给定  $\mathbb{C}$  上的线性空间  $\mathcal{V}$ . 如果  $\mathcal{V}$  上定义的二元函数  $\langle \cdot, \cdot \rangle \rightarrow \mathbb{C}$ , 满足:

1. 共轭对称性: 对任意  $\mathbf{a}, \mathbf{b} \in \mathcal{V}$ ,  $\langle \mathbf{a}, \mathbf{b} \rangle = \overline{\langle \mathbf{b}, \mathbf{a} \rangle}$ ;
2. 线性性和共轭线性性: 对任意  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}_1, \mathbf{b}_2 \in \mathcal{V}, k_1, k_2 \in \mathbb{C}$ ,

$$\langle k_1 \mathbf{a}_1 + k_2 \mathbf{a}_2, \mathbf{b}_1 \rangle = k_1 \langle \mathbf{a}_1, \mathbf{b}_1 \rangle + k_2 \langle \mathbf{a}_2, \mathbf{b}_1 \rangle,$$

$$\langle \mathbf{a}_1, k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 \rangle = \overline{k_1} \langle \mathbf{a}_1, \mathbf{b}_1 \rangle + \overline{k_2} \langle \mathbf{a}_1, \mathbf{b}_2 \rangle;$$

3. 正定性: 对任意  $\mathbf{a} \in \mathcal{V}$ ,  $\langle \mathbf{a}, \mathbf{a} \rangle$  是非负实数, 且  $\langle \mathbf{a}, \mathbf{a} \rangle = 0$  当且仅当  $\mathbf{a} = \mathbf{0}$ ;

那么称二元函数  $\langle \cdot, \cdot \rangle$  是  $\mathcal{V}$  上的一个**内积**, 称具有内积的线性空间  $\mathcal{V}$  为一个**复内积空间**或**酉空间**.

酉空间的**维数**定义为其作为线性空间的维数.

酉空间与欧氏空间主要区别在于共轭对称性. 唯有如此, 才能保证  $\langle \mathbf{a}, \mathbf{a} \rangle$  是实数, 因此可进一步要求正定性. 同样, 由于共轭对称性, 复内积仅仅对第一变量线性, 而对第二变量共轭线性.

**例 8.3.2** 1. 坐标向量空间  $\mathbb{C}^n$  关于内积  $\langle \mathbf{a}, \mathbf{b} \rangle = \overline{\mathbf{b}}^T \mathbf{a}$  构成欧氏空间, 这个内积常称为线性空间  $\mathbb{C}^n$  上的**标准内积**. 不加额外说明时, 酉空间  $\mathbb{C}^n$  指的就是标准内积.

设  $D$  是一个对角元素全为正数的对角矩阵,  $\mathbb{C}^n$  关于如下二元函数构成欧氏空间:  $\langle \mathbf{a}, \mathbf{b} \rangle = \overline{\mathbf{b}}^T D \mathbf{a}$ .

2. 矩阵空间  $\mathbb{C}^{m \times n}$  关于如下二元函数构成欧氏空间:  $\langle A, B \rangle = \text{trace}(\overline{B}^T A)$ .

3. 区间  $[a, b]$  上的复值连续函数构成的线性空间, 如下二元函数定义了其上的一个内积<sup>2</sup>:  $\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx$ . ☺

<sup>2</sup> 对一个复值函数  $f(x) = f_1(x) + \mathbf{i}f_2(x)$ , 其中  $f_1(x), f_2(x)$  是实值函数, 其积分定义为  $\int_a^b f(x) dx = \int_a^b f_1(x) dx + \mathbf{i} \int_a^b f_2(x) dx$ , 即实部虚部分别积分. 类似地, 对复值函数求导数, 也是实部虚部分别求导数, 即  $f'(x) = f'_1(x) + \mathbf{i}f'_2(x)$ .

为书写简便, 对任意向量  $\mathbf{a}$  或矩阵  $A$ , 记其共轭转置为  $\mathbf{a}^H = \overline{\mathbf{a}}^T$  或  $A^H = \overline{A}^T$ .

有了正定性条件, 就能类似地定义长度和距离等概念, 并有如下结论.

**定理 8.3.3 (Cauchy-Schwarz 不等式)** 在酉空间  $\mathcal{V}$  中, 对任意向量  $\mathbf{a}, \mathbf{b}$ , 都有

$$|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\| \|\mathbf{b}\|.$$

等号成立当且仅当  $\mathbf{a}, \mathbf{b}$  共线.

注意, 角度的概念不能直接地推广, 因为两个向量的内积通常是复数. 尽管没有角度的概念, 当  $\langle \mathbf{a}, \mathbf{b} \rangle = 0$  时, 仍称向量  $\mathbf{a}, \mathbf{b}$  **正交**. 由此自然地导出标准正交基、正交补、正交投影等概念. 特别地, Gram-Schmidt 正交化方法仍适用. 注意, 此时内积仅仅满足共轭对称性. 例如, 注意区别  $\langle \mathbf{a}_2, \tilde{\mathbf{q}}_1 \rangle$  和  $\langle \tilde{\mathbf{q}}_1, \mathbf{a}_2 \rangle$ .

下面讨论酉空间上的线性映射和线性变换. 一些平行的命题和定理也在此列出, 证明则和欧氏空间的相应结论几近相同, 略去.

伴随映射的概念与欧氏空间上的伴随映射 (定义 8.2.1) 相同, 即: 如果对任意  $\mathbf{x}, \mathbf{y}$ ,  $\langle f(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, g(\mathbf{y}) \rangle$ , 则称  $g$  是  $f$  的**共轭映射**或**伴随映射**, 记为  $g = f^*$ .

**命题 8.3.4** 当  $\mathcal{U} = \mathbb{C}^n, \mathcal{V} = \mathbb{C}^m$  时, 由  $A \in \mathbb{C}^{m \times n}$  确定的映射  $\mathbf{A}: \mathbf{x} \mapsto A\mathbf{x}$  的伴随映射是  $\mathbf{A}^*: \mathbf{y} \mapsto A^H \mathbf{y}$ .

证. 设伴随映射由矩阵  $B \in \mathbb{C}^{n \times m}$  确定, 则  $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, B\mathbf{y} \rangle$ . 由内积定义, 对任意  $\mathbf{x} \in \mathbb{C}^n, \mathbf{y} \in \mathbb{C}^m$ ,  $\mathbf{y}^H A\mathbf{x} = (B\mathbf{y})^H \mathbf{x} = \mathbf{y}^H B^H \mathbf{x}$ , 由此即得  $A = B^H, B = A^H$ .  $\square$

对伴随映射的矩阵, 有如下结论.

**命题 8.3.5** 给定酉空间  $\mathcal{U}, \mathcal{V}$  及各自一组标准正交基  $\mathbf{u}_1, \dots, \mathbf{u}_n; \mathbf{v}_1, \dots, \mathbf{v}_m$ . 如果线性映射  $f: \mathcal{U} \rightarrow \mathcal{V}$  在两组基下的矩阵是  $F$ , 则伴随映射  $f^*$  在两组基下的矩阵是  $F^H$ .

**定义 8.3.6 (自伴变换、酉变换、正规变换)** 给定酉空间  $\mathcal{V}$  中的线性变换  $f$ ,

1. 如果对任意  $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ , 都有  $\langle f(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, f(\mathbf{y}) \rangle$  (等价于  $f = f^*$ ), 则称  $f$  是  $\mathcal{V}$  上的一个**自伴变换**;
2. 如果对任意  $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ , 都有  $\langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$  (等价于  $f^* f = f f^* = \text{id}_{\mathcal{V}}$ ), 则称  $f$  是  $\mathcal{V}$  上的一个**酉变换**;
3. 如果  $f f^* = f^* f$ , 则称  $f$  是  $\mathcal{V}$  上的一个**正规变换**.

**定义 8.3.7 (Hermite 矩阵、酉矩阵、正规矩阵)** 给定  $n$  阶复矩阵  $A$ ,

1. 如果  $A^H = A$ , 则称  $A$  为  $n$  阶 **Hermite 矩阵**;

2. 如果  $A^H A = A A^H = I_n$ , 则称  $A$  为  $n$  阶酉矩阵;
3. 如果  $A A^H = A^H A$ , 则称  $A$  为  $n$  阶正规矩阵.

易见, Hermite 矩阵是实对称矩阵的推广, 而酉矩阵是正交矩阵的推广

容易验证, 酉变换、自伴随变换都是正规变换; 酉矩阵、Hermite 矩阵都是正规矩阵. 如下结论是欧氏空间上的对应结论在酉空间上的推广.

**命题 8.3.8** 给定  $n$  维酉空间上线性变换  $f$ ,

1.  $f$  是自伴变换, 当且仅当  $f$  在任意标准正交基下的矩阵都是 Hermite 矩阵;
2.  $f$  是酉变换, 当且仅当  $f$  在任意标准正交基下的矩阵都是酉矩阵;
3.  $f$  是正规变换, 当且仅当  $f$  在任意标准正交基下的矩阵都是正规矩阵.

再看酉空间上的基变换与线性变换的矩阵在基变换下的变化规律.

**命题 8.3.9** 给定  $n$  维酉空间  $\mathcal{V}$  中一组标准正交基  $e_1, \dots, e_n$  和  $\mathbb{C}$  上  $n$  阶方阵  $U$ . 令  $(u_1, \dots, u_n) = (e_1, \dots, e_n)U$ , 则  $u_1, \dots, u_n$  是一组标准正交基, 当且仅当  $U$  是酉矩阵.

**命题 8.3.10** 矩阵  $U$  是酉矩阵, 当且仅当  $U$  的 (行) 列向量组构成  $\mathbb{C}^n$  的一组标准正交基.

**命题 8.3.11** 给定酉空间  $\mathcal{V}$  中及其两组标准正交基  $e_1, \dots, e_n; u_1, \dots, u_n$ , 设过渡矩阵为  $U$ , 即  $(u_1, \dots, u_n) = (e_1, \dots, e_n)U$ . 如果  $\mathcal{V}$  上线性变换  $f$  在基  $e_1, \dots, e_n$  下的矩阵为  $F$ , 那么  $f$  在基  $u_1, \dots, u_n$  下的矩阵为  $U^{-1}FU = U^H F U$ .

对复方阵  $A, B$ , 如果存在酉矩阵  $U$  使得  $U^H A U = B$ , 则称  $A$  和  $B$  酉相似, 或  $A$  酉相似于  $B$ . 复方阵的酉相似关系是等价关系, 比相似关系更细致. 不难发现, 酉空间上线性变换在标准正交基下矩阵的标准形问题, 等价于矩阵酉相似变换下的标准形问题.

类似于命题 5.4.6 和定理 8.2.19, 先利用酉相似变换尽量简化矩阵.

**定理 8.3.12 (Schur 分解)** 对  $n$  阶方阵  $A$ , 存在酉矩阵  $U$ , 使得  $U^H A U = T$  是上三角矩阵, 且  $T$  的对角元素是  $A$  的  $n$  个特征值 (计重数). 进一步地, 通过选择特定的  $U$ , 能够令  $T$  的对角元素是  $A$  的特征值的任意排列.

证. 设复方阵为  $A$ , 对阶数  $n$  应用数学归纳法. 当  $n = 1$  时, 显然. 现假设命题对任意  $n - 1$  阶方阵成立.

根据代数学基本定理,  $A$  有  $n$  个特征值. 任取  $A$  的一个特征对  $(\lambda_1, \mathbf{x}_1)$ , 其中  $\mathbf{u}_1$  是单位向量. 将其扩充成  $\mathbb{C}^n$  的一组标准正交基  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ . 记  $U_1 = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n] = [\mathbf{x}_1 \ X_{12}]$ , 则

$$AU_1 = A[\mathbf{x}_1 \ X_{12}] = [\mathbf{x}_1 \ X_{12}] \begin{bmatrix} \lambda & * \\ \mathbf{0} & A_2 \end{bmatrix} = U_1 \begin{bmatrix} \lambda & * \\ \mathbf{0} & A_2 \end{bmatrix}.$$

其中  $U_1$  是酉矩阵. 因此,  $A$  酉相似于分块上三角矩阵  $A_1 = \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & A_2 \end{bmatrix}$ . 而  $A_2$  是  $n-1$  阶方阵, 根据归纳假设, 存在酉矩阵  $U_2$ , 使得  $U_2^H A_2 U_2 = T_2$  是上三角矩阵. 因此,

$$A = U_1 \begin{bmatrix} \lambda_1 & * \\ \mathbf{0} & U_2 T_2 U_2^H \end{bmatrix} U_1^H = U_1 \begin{bmatrix} 1 & \\ & U_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ & T_2 \end{bmatrix} \begin{bmatrix} 1 & \\ & U_2 \end{bmatrix}^H = T.$$

令  $U = U_1 \begin{bmatrix} 1 & \\ & U_2 \end{bmatrix}$ ,  $T = \begin{bmatrix} \lambda_1 & * \\ & T_2 \end{bmatrix}$ , 则  $U$  酉,  $T$  上三角, 且  $A = UTU^H$ .

综上, 命题对任意  $n$  成立.

考察特征多项式, 立得  $T$  的对角元素就是  $A$  的  $n$  个特征值. 通过上述证明能看到, 按照特征值的排列选择特征对, 就能使  $T$  的对角元素满足条件.  $\square$

分解  $A = UTU^H$ , 称为  $A$  的 **Schur 分解**.

Schur 分解在数值计算时很重要. 在第 5.2 节结尾处曾提到利用 Francis 算法计算矩阵的特征值和特征向量以及多项式的根. 事实上, Francis 算法就是在计算矩阵的 Schur 分解.

下面讨论 Hermite 矩阵、酉矩阵和正规矩阵在酉相似变换下的标准形.

**命题 8.3.13** 1. Hermite 矩阵的特征值都是实数;

2. 酉矩阵的特征值都是绝对值为 1 的复数.

**定理 8.3.14 (Hermite 矩阵的谱分解)** 对任意 Hermite 矩阵  $A$ , 都存在酉矩阵  $U$  和实对角矩阵  $\Lambda$ , 使得  $A = U\Lambda U^H$ .

分解  $A = U\Lambda U^H$  称为 Hermite 矩阵  $A$  的**谱分解**.

事实上, 正规矩阵都可以用酉矩阵对角化.

**定理 8.3.15** 复矩阵  $A$  是正规矩阵, 当且仅当  $A$  酉相似于对角矩阵, 即存在酉矩阵  $U$ , 使得  $U^H A U = \Lambda$  是对角矩阵.



证. “ $\Leftarrow$ ”: 如果存在酉矩阵  $U$ , 使得  $U^H A U = \Lambda$  是对角矩阵, 则  $A = U \Lambda U^H, A^H = U \Lambda^H U^H$ , 因而  $A A^H = U \Lambda \Lambda^H U^H = U \Lambda^H \Lambda U^H = A^H A$ , 即  $A$  正规.

“ $\Rightarrow$ ”: 正规矩阵  $A$  有 Schur 分解  $A = U T U^H$ , 易知  $T$  也是正规矩阵. 只需再证上三角的正规矩阵是对角矩阵. 对  $n$  用数学归纳法. 设  $T = \begin{bmatrix} \lambda_1 & \mathbf{x}^H \\ & T_2 \end{bmatrix}$ , 由  $T^H T = T T^H$  可知

$$\begin{bmatrix} \bar{\lambda}_1 & \\ \mathbf{x} & T_2^H \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{x}^H \\ & T_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{x}^H \\ & T_2 \end{bmatrix} \begin{bmatrix} \bar{\lambda}_1 & \\ \mathbf{x} & T_2^H \end{bmatrix}.$$

比较左上角元素可得  $\bar{\lambda}_1 \lambda_1 + \mathbf{x}^H \mathbf{x} = \bar{\lambda}_1 \lambda_1$ , 因此  $\mathbf{x}^H \mathbf{x} = 0$ , 即  $\mathbf{x} = \mathbf{0}$ . 因此  $T = \begin{bmatrix} \lambda_1 & \\ & T_2 \end{bmatrix}$ . 再比较右下角的  $n-1$  阶矩阵可得  $T_2 T_2^H = T_2^H T_2$ , 即  $T_2$  是正规矩阵. 根据归纳假设, 上三角矩阵  $T_2$  是对角矩阵. 因此  $T$  也是对角矩阵.  $\square$

酉矩阵是正规矩阵, 直接应用定理 8.3.15, 可得如下推论.

**推论 8.3.16** 任意酉矩阵酉相似于对角矩阵  $\Lambda$ , 其中  $\Lambda$  对角元素的绝对值都是 1.

实正交矩阵是酉矩阵, 由推论 8.3.16 可以得到实 Schur 分解 (定理 8.2.18) 的又一证明, 细节留给读者.

最后, 我们把一系列平行概念列在表 8.3.1 中, 对应概念的定义、性质及其证明都类似, 不再赘述.

欧氏空间		酉空间	
转置	$A^T$	共轭转置	$A^H$
$\mathbb{R}^n$ 内积	$\mathbf{b}^T \mathbf{a}$	$\mathbb{C}^n$ 内积	$\mathbf{b}^H \mathbf{a}$
正交	$\mathbf{b}^T \mathbf{a} = 0$	<sup>3</sup>	$\mathbf{b}^H \mathbf{a} = 0$
正交矩阵	$A^T A = I_n$	酉矩阵	$A^H A = I_n$
列正交矩阵	$A^T A = I_k$	=	$A^H A = I_k$
对称矩阵	$A = A^T$	Hermite 矩阵	$A = A^H$
对称矩阵谱分解	$A = Q \Lambda Q^T$	Hermite 矩阵 =	$A = Q \Lambda Q^H$
Rayleigh 商	$\mathbf{x}^T A \mathbf{x}$	=	$\mathbf{x}^H A \mathbf{x}$
正定	$\mathbf{x}^T A \mathbf{x} > 0, \forall \mathbf{x}$	=	$\mathbf{x}^H A \mathbf{x} > 0, \forall \mathbf{x}$
奇异值分解	$A = U \Sigma V^T$	=	$A = U \Sigma V^H$
对称变换	$f = f^*$	=	$f = f^*$
正交变换	$f^* f = \text{id}_V$	=	$f^* f = \text{id}_V$

<sup>3</sup>“=” 表示概念名称相同.

正规变换	$f^*f = ff^*$	=	$f^*f = ff^*$
------	---------------	---	---------------

表 8.3.1: 欧氏空间与酉空间概念的对照

注意, 由于复矩阵有  $n$  个复特征值, 很多结论要比实矩阵更简明. 例如, 并非每个实矩阵都正交相似于实上三角矩阵. 正如定理 8.2.18 和推论 8.3.16 的对比, 实矩阵可以有实 Schur 分解, 参见阅读 8.3.1.

## 习题

**阅读 8.3.1 (实 Schur 分解)** 给定  $\mathbb{R}$  上  $n$  阶方阵  $A$ , 则存在正交矩阵  $Q$ , 使得

$$Q^T A Q = T = \begin{bmatrix} T_{11} & T_{12} & \cdots & T_{1r} \\ & T_{22} & \ddots & \vdots \\ & & \ddots & T_{r-1,r} \\ & & & T_{rr} \end{bmatrix},$$

其中  $T_{ii}$  或者是 1 阶实矩阵, 或者是 2 阶实矩阵且其特征多项式的两个零点是一对共轭复数. 并且通过对  $Q$  的选取, 可以实现对对角块  $T_{ii}$  的任意排列.

分解  $A = Q T Q^T$  称为  $A$  的 **实 Schur 分解**.

证. 采用数学归纳法.  $n = 1$  时, 显然. 假设任意不超过  $n - 1$  阶实矩阵都有如上分解, 观察  $n$  阶实矩阵.

如果  $A$  存在实特征值  $\lambda_1$ , 则存在实特征向量  $\mathbf{q}_1$ ,  $\|\mathbf{q}_1\| = 1$ . 把  $\mathbf{q}_1$  扩充成  $\mathbb{R}^n$  的一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_n$ . 记  $Q_1 = [\mathbf{q}_1 \quad Q_{12}] = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n]$ . 于是

$$Q_1^T A Q_1 = \begin{bmatrix} \mathbf{q}_1^T A \mathbf{q}_1 & \mathbf{q}_1^T A Q_{12} \\ Q_{12}^T A \mathbf{q}_1 & Q_{12}^T A Q_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 \mathbf{q}_1^T \mathbf{q}_1 & \mathbf{q}_1^T A Q_{12} \\ \lambda_1 Q_{12}^T \mathbf{q}_1 & Q_{12}^T A Q_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{q}_1^T A Q_{12} \\ 0 & Q_{12}^T A Q_{12} \end{bmatrix}.$$

注意  $Q_{12}^T A Q_{12}$  是  $n - 1$  阶实矩阵, 根据归纳假设, 存在正交矩阵  $Q_2$  和分块上三角矩阵  $T_2$ , 使得  $Q_{12}^T A Q_{12} = Q_2 T_2 Q_2^T$ , 其中  $T_2$  的对角块只有如命题所列的两种. 因此

$$A = Q_1 \begin{bmatrix} \lambda & * \\ 0 & Q_2 A_2 Q_2^T \end{bmatrix} Q_1^T = Q_1 \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}^T Q_1^T.$$

记  $Q = Q_1 \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}$ ,  $T = \begin{bmatrix} \lambda_1 & * \\ 0 & T_2 \end{bmatrix}$ , 则  $Q$  是正交矩阵,  $A$  是满足条件的块上三角矩阵, 而  $A = Q A Q^T$ .

如果  $A$  的特征多项式有虚数零点  $\alpha + i\beta$ ,  $\beta \neq 0$ , 事实上它是  $A$  看作  $\mathbb{C}$  上矩阵时的特征值, 记属于它的一个特征向量为  $\mathbf{x} + i\mathbf{y}$ . 由  $A(\mathbf{x} + i\mathbf{y}) = (\alpha + i\beta)(\mathbf{x} + i\mathbf{y})$  可知,  $A(\mathbf{x} - i\mathbf{y}) = (\alpha - i\beta)(\mathbf{x} - i\mathbf{y})$ , 即  $(\alpha - i\beta, \mathbf{x} - i\mathbf{y})$  也是  $A$  的一个特征对. 易知  $\mathbf{x}, \mathbf{y}$  线性无关, 且  $A \begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$ . 令  $\begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix} = [\mathbf{q}_1 \quad \mathbf{q}_2] R_1$  是简化 QR 分解. 将  $\mathbf{q}_1, \mathbf{q}_2$  扩充成一组标准正交基  $\mathbf{q}_1, \dots, \mathbf{q}_n$ . 记  $Q_1 = [\mathbf{q}_{11} \quad Q_{12}]$ , 其

中  $Q_{11} = [\mathbf{q}_1 \quad \mathbf{q}_2], Q_{12} = [\mathbf{q}_3 \quad \cdots \quad \mathbf{q}_n]$ . 那么  $AQ_{11} = A[\mathbf{x} \quad \mathbf{y}]R^{-1} = [\mathbf{x} \quad \mathbf{y}]\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}R^{-1} = Q_{11}R\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}R^{-1} =: Q_{11}T_{11}$ . 而

$$Q_1^T A Q_1 = \begin{bmatrix} Q_{11}^T A Q_{11} & Q_{11}^T A Q_{12} \\ Q_{12}^T A Q_{11} & Q_{12}^T A Q_{12} \end{bmatrix} = \begin{bmatrix} Q_{11}^T Q_{11} T_{11} & Q_{11}^T A Q_{12} \\ Q_{12}^T Q_{11} T_{11} & Q_{12}^T A Q_{12} \end{bmatrix} = \begin{bmatrix} T_{11} & Q_{11}^T A Q_{12} \\ 0 & Q_{12}^T A Q_{12} \end{bmatrix}.$$

类似前一种情形, 利用  $n-2$  时的归纳假设, 即得结论.

综上, 原命题对任意  $n$  成立.  $\square$

练习 8.2.8 中定义了实正规矩阵, 即如果  $AA^T = A^T A$ , 则  $A$  正规. 类似于定理 8.3.15, 有如下结论: 实矩阵  $A$  是正规矩阵, 当且仅当存在正交矩阵  $Q$  和块对角矩阵  $T$ , 使得  $Q^T A Q = T$ , 其中  $T$  的对角块或者是 1 阶实矩阵, 或者是形如  $\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$  的 2 阶实矩阵.

由此立得定理 8.2.18.

类似于 Schur 分解, 针对实矩阵的特征值问题, Francis 算法就是在计算矩阵的实 Schur 分解.

**练习 8.3.2** 在  $\mathbb{C}[x]_n$  中定义二元函数  $\langle f, g \rangle = \sum_{k=1}^n f(k)\overline{g(k)}$ .

1. 证明, 它定义了  $\mathbb{C}[x]_n$  上的一个内积.
2. 当  $n=3$  时, 求它的一组标准正交基.

**练习 8.3.3** 在具有标准内积的酉空间  $\mathbb{C}^3$  中, 设  $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ 0 \\ i \end{bmatrix}$ , 求与之线性等价的一个正交向量组.

**练习 8.3.4** 证明命题 8.3.9

**练习 8.3.5** 证明命题 8.3.10.

**练习 8.3.6** 证明, 酉矩阵的行列式的绝对值为 1; 酉矩阵的特征值的绝对值为 1.

**练习 8.3.7** 设  $f$  是酉空间  $\mathcal{V}$  上的 Hermite 变换, 证明, 对任意  $\mathbf{a} \in \mathcal{V}$ ,  $\langle f(\mathbf{a}), \mathbf{a} \rangle$  都是实数.

**练习 8.3.8** 如果矩阵的共轭转置等于其自身的负矩阵, 那么该矩阵称为**反 Hermite 矩阵**或**斜 Hermite 矩阵**. 证明, 反 Hermite 矩阵是正规矩阵, 且其特征值都是纯虚数.

**练习 8.3.9** 如果一个 Hermite 矩阵的特征值都是正实数, 则称其为**正定 Hermite 矩阵**. 证明, 二元函数  $\overline{\mathbf{b}}^T A \mathbf{a}$  定义了  $\mathbb{C}^n$  上的一个内积, 当且仅当  $A$  是正定 Hermite 矩阵.

**练习 8.3.10** 利用酉矩阵的酉相似标准形证明定理 8.2.18.

Liang, Tian & Yang

# 名词索引

$n$  维线性空间,  $\rightarrow$  维数

伴随变换 adjoint (transformation),  
286

伴随矩阵 adjugate matrix, 173

伴随映射 adjoint (mapping), 286, 296

半单 semi-simple, 194

半负定 negative semidefinite, 220

半正定 positive semidefinite, 220

倍乘变换 elementary transformation  
of type 2, 27

倍乘矩阵 elementary matrix of  
type 2, 55

倍加变换 elementary transformation  
of type 3, 27

倍加矩阵 elementary matrix of  
type 3, 55

变换 transformation, 2

变积系数, 158

标准基 standard basis, 90

标准内积 standard inner product,  
277, 278, 295

标准形 canonical form, 61

标准正交基 orthonormal basis, 124,  
279

标准坐标向量 standard unit vector,

15

标准坐标向量 standard unit vectors,  
10

标准坐标向量组 standard unit  
vectors,  $\rightarrow$  标准坐标向量

表示矩阵 matrix, 17

不变量 invariant, 61

不定 indefinite, 220

Cauchy-Bunyakovsky-Schwarz 不等式  
Cauchy-Bunyakovsky-Schwarz  
inequality,  $\rightarrow$  Cauchy-Schwarz  
不等式

Cauchy-Schwarz 不等式  
Cauchy-Schwarz inequality,  
123, 278, 296

Cauchy 不等式 Cauchy inequality,  $\rightarrow$   
Cauchy-Schwarz 不等式

Chebyshev 多项式 Chebyshev  
polynomials, 281

Cholesky 分解 Cholesky factorization,  
219

Cramer 法则 Cramer's rule, 173

采样 sampling, 147

叉积 cross product, 14

插值 interpolation, 165

- 超平面 hyperplane, 36
- 乘法 multiplication, 253
- 初等变换 elementary transformation, 26
- 初等矩阵 elementary matrix, 54
- 初等行(列)变换 elementary transformation, 27
- 垂直 perpendicular, 123
- 错切变换 shearing, 11
- 代数学基本定理 fundamental theorem of algebra, 182
- 代数余子式 cofactor, 170
- 代数重数 algebraic multiplicity, 187
- 单 simple, 187
- 单射 injection, 1
- 单位化(向量) unitization, 123
- 单位矩阵 identity matrix,  $\rightarrow$  恒同矩阵
- 单位上(下)三角矩阵 unit upper(lower) triangular matrix, 21
- 单位向量 unit vector, 123
- 导出方程组 induced equations, 113
- 等价变换 equivalence transformation, 61
- 等价关系 equivalence relation, 61
- 等价类 equivalence class, 61
- 点积 dot product,  $\rightarrow$  内积
- 点空间 point space, 36
- 定义域 domain, 1
- 度量矩阵 Gramian matrix, 282
- 对称变换 symmetric transformation, 288
- 对称矩阵 symmetric matrix, 41
- 对称性 symmetry, 122
- 对换 transposition, 174
- 对换变换 elementary transformation of type 1, 27
- 对换变换 transposition, 11
- 对换矩阵 elementary matrix of type 1, 54
- 对角矩阵 diagonal matrix, 21
- 对角占优矩阵 diagonally dominant matrix, 59
- 对偶空间 dual space, 274
- Euclid 空间 Euclidean space,  $\rightarrow$  欧氏空间
- Euler 公式 Euler formula, 115
- 二维 Haar 小波基 Haar wavelet bases, 130
- Fourier 变换 Fourier transform, 285
- Fredholm 二择一定理 Fredholm Alternative Theorem, 119
- Frobenius 范数 Frobenius norm, 232
- 法向量 normal vector, 36, 126
- 樊畿迹定理 Fan trace theorem, 234
- 反 Hermite 矩阵 skew-Hermitian matrix, 301
- 反对称矩阵 skew-symmetric matrix, 41
- 反射变换 reflection, 10
- 范数 norm, 278, 294
- 方阵 square matrix, 19
- 仿射变换 affine transformation, 75
- 非零解 nonzero solution, 32
- 非零行 nonzero row, 28
- 非平凡 nontrivial, 86
- 非平凡解 nontrivial solution, 32

非奇异矩阵 nonsingular matrix,  $\rightarrow$  可逆矩阵

非齐次 inhomogeneous, 32

分块对角矩阵 block diagonal matrix, 67

分量 entry, 4

复合 composition, 2, 253

复内积空间 complex inner product space,  $\rightarrow$  酉空间

负定 negative definite, 220

负矩阵 opposite matrix, 38

负向量 opposite vector, 6

负元素, 238

Gauss-Jordan 消元法 Gauss-Jordan elimination, 30

Gauss 消元法 Gauss elimination, 30

Gershgorin 圆盘定理 Gershgorin disc theorem, 187

Givens 变换 Givens transformation, 132

Gram-Schmidt 正交化 Gram-Schmidt process, 125

Gram 矩阵 Gram matrix, 140

Gram 矩阵 Gramian matrix,  $\rightarrow$  度量矩阵

共轭 conjugate, 183

共轭映射 conjugate mapping, 286, 296

共线 colinear, 97

勾股定理 Pythagoras' theorem, 123

关联矩阵 incident matrix, 90

广义逆 generalized inverse,  $\rightarrow$  Moore-Penrose 广义逆

过渡矩阵 transition matrix, 260

Hadamard 不等式 Hadamard inequality, 169, 223

Hadamard 猜想 Hadamard conjecture, 138

Hadamard 矩阵 Hadamard matrix, 138

Hamilton-Cayley 定理 Hamilton-Cayley theorem, 203

Heisenberg 不确定性原理 Heisenberg's Uncertainty Principle, 294

Hermite 矩阵 Hermitian matrix, 296

Hesse 矩阵 Hessian, 220

Hilbert 矩阵 Hilbert matrix, 167

Householder 变换 Householder transformation, 133

和 sum, 94, 241

核 kernel, 253

恒同变换 identity map, 2, 8

恒同矩阵 identity matrix, 21

化零多项式 annihilating polynomial, 190

回代法 backward substitution, 22, 30

Jordan 标准形 Jordan canonical form, 203

Jordan 分解 Jordan decomposition, 203

Jordan 块 Jordan block, 195

Jordan 链 Jordan chain, 208

基 basis, 89, 246

迹 trace, 51, 188

极大线性无关部分组, 95, 245

极分解 polar decomposition, 234

几何重数 geometric multiplicity, 193

- 加法 addition, 237, 252  
 夹角 angle, 123  
 减法 subtraction, 6, 38  
 简化 QR 分解 reduced QR decomposition, 137  
 简化奇异值分解 reduced singular value decomposition, 226  
 交 intersection, 94, 240  
 阶梯形矩阵 echelon matrix, 29  
 截断 SVD truncated SVD, 229  
 解空间 solution space, 87  
 矩阵 matrix, 17, 265, 270  
 距离 distance, 123, 278  
 Kirchhoff 电流定律 Kirchhoff Current Law, 116  
 Kirchhoff 电压定律 Kirchhoff Voltage Law, 91  
 Kronecker 符号 Kronecker symbol, 279  
 可对角化 diagonalizable, 192  
 可交换 commutable, 45  
 可逆 invertable, 2  
 可逆矩阵 invertable matrix, 56  
 亏损 defective, 194  
 LDL<sup>T</sup> 分解 LDL<sup>T</sup> factorization, 79  
 LDU 分解 LDU factorization, 79  
 Legendre 多项式 Legendre polynomials, 280  
 LU 分解 LU factorization, 77  
 离散 Fourier 变换 discretized Fourier transformation, 216  
 连通 connected, 115  
 列 (向量) 空间 column space, 87  
 列满秩 column full rank, 107  
 列向量 column (vector), 17  
 列正交矩阵 column orthogonal matrix, 136  
 邻接矩阵 adjacent matrix, 52  
 零解 zero solution, 32  
 零矩阵 zero matrix, 38  
 零空间 null space, 87  
 零向量 zero vector, 6  
 零行 zero row, 28  
 零映射 zero map, 37  
 零元素 zero vector, 237  
 Moore-Penrose 广义逆 Moore-Penrose generalized inverse, 226  
 满射 surjection, 2  
 满秩 full rank, 107  
 满秩分解 rank decomposition, 111  
 内积 inner product, 18, 122, 277, 295  
 能量函数 energy function, 218  
 逆 (矩阵) inverse, 56  
 逆 (映射) inverse (map), 2  
 Ohm 定律 Ohm's Law, 115  
 欧氏空间 Euclidean space, 277  
 偶排列 even permutation, 174  
 Pascal 矩阵 Pascal matrix, 78  
 Pauli 矩阵 Pauli matrix, 207  
 PLU 分解 PLU factorization, 81  
 排列 permutation,  $\rightarrow$  置换  
 陪域 codomain, 1  
 偏导数 partial derivative, 169  
 平凡 trivial, 86  
 平凡解 trivial solution, 32  
 谱 spectrum, 192  
 谱范数 spectral norm, 227,  $\rightarrow$  范数



- 谱分解 spectral decomposition, 192, 212, 298
- QR 分解 QR decomposition, 134, 137
- 奇排列 odd permutation, 174
- 奇异矩阵 singular matrix, 56
- 奇异值 singular value, 224
- 奇异值分解 singular value decomposition, 225
- 齐次 homogeneous, 32
- 前代法 forward substitution, 22
- Rayleigh 商 Rayleigh quotient, 221
- Riesz 表示定理 Riesz Representation Theorem, 128
- 弱对角占优矩阵 weakly diagonally dominant matrix, 65
- Schur 补 Schur complement, 70
- Schur 分解 Schur decomposition, 298
- Sherman-Morrison 公式  
Sherman-Morrison formula, 71
- SVD SVD, 225
- Sylvester 方程 Sylvester equation, 191
- 三对角矩阵 tridiagonal matrix, 35
- 三角不等式 triangular inequality, 123, 279
- 上三角矩阵 upper triangular matrix, 21
- 伸缩变换 compression, 11
- 生成 span, 87, 244
- 生成向量 generating vectors, 87
- 实 Schur 分解 real Schur decomposition, 292, 300
- 实内积空间 real inner product space,  
→ 欧氏空间
- 数乘 scalar multiplication, 5, 237, 252
- 数据拟合 data fitting, 147
- 数量矩阵 scalar matrix, 47
- 数域 number field, 237
- 双射 bijection, 2
- 双线性性 bilinearity, 122
- 顺序主子式 leading principal minor, 169, 218
- 顺序主子阵 leading principal submatrix, 77
- 随机矩阵 stochastic matrix, Markov matrix, 197
- 随机游走 random walk, 196
- 特解 particular solution, 113
- 特征对 eigenpair, 183, 253
- 特征多项式 characteristic polynomial, 184
- 特征向量 eigenvector, 183, 253
- 特征值 eigenvalue, 183, 253
- 特征子空间 eigenspace, 185, 254
- 梯度 gradient, 220
- 通解 general solution, 28
- 同构 isomorphic, 255
- 同构映射 isomorphism, 255
- 同时对角化 simultaneously diagonalizable, 205
- 投影变换 projection, 11
- Vandermonde 矩阵 Vandermonde matrix, 165
- Vieta 定理 Vieta's formulae, 182
- 维数 dimension, 100, 246, 277, 295
- 稳定矩阵 stable matrix, 273

无限维线性空间 infinite dimensional space, 246

系数矩阵 coefficient matrix, 26

瑕旋转 improper rotation, 293

下三角矩阵 lower triangular matrix, 21

线性变换 linear transformation, 8, 253

线性表示 linear expressed, 15, 95, 244

线性等价 linearly equivalent, 97, 246

线性函数 linear function, 13, 274

线性空间 linear space, 5, 238

线性无关 linearly independent, 88, 244

线性系统 linear system, 1

线性相关 linearly dependent, 88, 244

线性映射 linear map, 7, 251

线性运算 linear operation, 5

线性组合 linear combination, 15, 244

相等 equal, 5, 17

相抵, 108

相抵标准形, 109

相似 similar, 200

相似变换 similar transformation, 201

相似标准形, 202

像 image, 1

像集 range, image (set), 1, 253

向后差分 backward difference (transformation), 40

向后差分矩阵 backward difference matrix, 40

向量 vector, 4, 238

向量加法 vector addition, 5

向量空间 vector space,  $\rightarrow$  线性空间,

238

向前差分 forward difference (transformation), 39

向前差分矩阵 forward difference matrix, 39

斜 Hermite 矩阵 skew-Hermitian matrix,  $\rightarrow$  反 Hermite 矩阵

斜对称矩阵 skew-symmetric matrix,  $\rightarrow$  反对称矩阵

行(向量)空间 row space, 106

行简化阶梯形矩阵 row reduced echelon matrix, 29

行列式 determinant, 159

行列式函数 determinant function, 159

行满秩 row full rank, 107

行向量 row (vector), 17

旋转变换 rotation, 9

严格上(下)三角矩阵 strictly upper(lower) triangular matrix, 21

一般解 general solution,  $\rightarrow$  通解

映射 map, 1

友矩阵 companion matrix, 189

有限维线性空间 finite dimensional space, 246

酉变换 unitary transformation, 296

酉矩阵 unitary matrix, 297

酉空间 unitary space, 295

酉相似 unitarily similar, 297

右奇异向量 right singular vector, 224

余子式 minor, 170

原像 preimage, 1

增广矩阵 augmented matrix, 26

长度 length, 123, 278

- 正定 positive definite, 218, 219, 301  
正定性 positive definiteness, 122  
正规变换 normal transformation, 294, 296  
正规矩阵 normal matrix, 294, 297  
正交 orthogonal, 123, 142, 279, 296  
正交变换 orthogonal transformation, 290  
正交补 orthogonal complement, 142, 283  
正交单位向量组, 279  
正交单位向量组 orthogonal unit vectors, 124  
正交化方法 orthogonalization method, 149  
正交基 orthogonal basis, 124, 279  
正交矩阵 orthogonal matrix, 131  
正交投影 orthogonal projection, 123, 145, 284  
正交投影(变换) orthogonal projection transformation, 145, 284  
正交投影矩阵 orthogonal projection matrix, 145  
正交相似 orthogonally similar, 213  
正交相似变换 orthogonally similar transformation, 213  
正交向量组, 279  
正交向量组 orthogonal vectors, 124  
正则化方法 normal equation method, 148  
值域 range,  $\rightarrow$  像集  
直和 direct sum, 241  
秩 rank, 99, 104, 246  
秩一分解 rank-1 decomposition, 111  
置换 permutation, 174  
置换矩阵 permutation matrix, 59  
重数 multiplicity, 182, 187  
主变量 pivot variables, 28  
主成分 principal component, 230  
主成分分析 principal component analysis, 230  
主列 pivot column, 105  
主元 pivot, 28  
转置 transpose, 40  
自伴变换 self-adjoint transformation, 288, 296  
自同构 automorphism, 255  
自由变量 free variables, 28  
自由列 free column, 105  
子空间 subspace, 86, 240  
最小二乘 least squares, 148  
最小二乘解 LS solution, 148  
左零空间 left nullspace, 114  
左奇异向量 left singular vector, 224  
左相抵, 61  
左相抵标准形, 61  
坐标 coordinate, 4, 257

Liang, Tian & Yang

# 人名表

Bunyakovsky (Буняковский) 布尼雅科夫斯基, 278

Cauchy 柯西, 123, 278, 296

Cayley 凯莱, 202

Chebyshev (Чебышёв) 切比雪夫, 车贝晓夫, 281

Cholesky 乔列斯基, 晓列斯基, 219

Cramer 克莱姆, 173

De Morgan 德·摩根, 151

Euclid (Εὐκλείδης) 欧几里得, 277

Euler 欧拉, 115

Fan 樊畿, 234

Fibonacci 斐波那契, 40, 196

Fourier 傅里叶, 216

Francis 弗朗西斯, 189

Fredholm 弗莱德霍姆, 119

Frobenius 弗罗贝尼乌斯, 232

Gauss (Gauß) 高斯, 30

Gershgorin (Гершгорин) 盖尔施高林, 187

Givens 吉文斯, 132

Gram 格拉姆, 125, 140

Haar 哈尔, 130

Hadamard 阿达马, 138, 169

Hamilton 哈密顿, 202

Heisenberg 海森堡, 海森伯格, 294

Hermite 埃尔米特, 厄密特, 296

Hesse 海塞, 220

Hessenberg 海森伯格, 177

Hooke 胡克, 41

Householder 豪斯霍尔德, 133

Jordan 若尔当, 30, 195, 203

Kirchhoff 基尔霍夫, 91, 116

Kronecker 克罗内克, 279

Legendre 勒让德, 280

Moore 摩尔, 226

Morrison 莫里森, 71

Newton 牛顿, 204, 214

Ohm 欧姆, 115

Pascal 帕斯卡, 78

Pauli 泡利, 207

Penrose 彭罗斯, 226

Pythagoras (Πυθαγόρας ὁ Σάμιος) 毕达哥拉斯, 305

Rayleigh 瑞利, 221

Riesz 里斯, 128

Schmidt 施密特, 125

Schur 舒尔, 70, 292

Schwarz 施瓦茨, 123, 278, 296

Sherman 舍曼, 71

Sylvester 西尔维斯特, 191

Vandermonde 范德蒙, 165

Vieta (Viète) 韦达, 182

# 符号表

$:=, =:$ , 5

$\cdot!$ , 59

$\cdot^*$ , 286

$\cdot^+$ , 226

$\cdot^H$ , 296

$\cdot^T$ , 40

$\circ$ , 2

$\text{diag}(\cdot)$ , 21

$\text{Hom}(\cdot, \cdot)$ , 251

$\text{id}$ , 2

$\|\cdot\|$ , 123, 227, 294

$\|\cdot\|_F$ , 232

$\mathcal{N}(\cdot)$ , 85, 253

$\oplus$ , 241

$\overline{\cdot}$ , 183

$\perp$ , 123, 142

$\Pi$ , 165

$\det(\cdot)$ , 159

$\dim \cdot$ , 100, 246

$\mathcal{R}(\cdot)$ , 85, 253

$\text{rank}(\cdot)$ , 99, 104, 246

$\text{sign}(\cdot)$ , 174

$\text{span}(\cdot)$ , 87

$\Sigma$ , 44

$\mathbb{C}$ , 1

$\mathbb{Q}$ , 1

$\mathbb{R}$ , 1

$\text{trace}(\cdot)$ , 188

$C_a^b$ , 78