

数据科学 与 工程数学 基础

主 编
高胜哲
顾 剑

副主编
潘福臣
屈磊磊
孙 华

清华大学出版社
北 京

内 容 简 介

本书系统介绍了有关数据分析、机器学习和人工智能算法设计过程中数据表示、度量、建模、评价和求解所要用到的最基础的数学知识,全书分为9章,在微积分、线性代数、概率论与数理统计等先修平台课程基础上,进一步讲解线性空间与线性变换、矩阵分析、矩阵分解、矩阵计算问题、概率模型和参数估计、最优化基础知识、最优性条件和对偶理论、优化算法、信息论基础等内容,夯实读者在大数据领域的数学理论基础。

本书语言通俗易懂,侧重知识的理解与应用,既可作为普通本科院校数据科学与大数据技术、人工智能、计算机应用技术等专业的教材,帮助学生掌握专业必备的数学工具,也能为从事大数据、人工智能相关领域教学的教师及初入行业的技术人员提供参考,助力其快速把握数学知识在实际中应用的底层逻辑与运用方法。

版权所有,侵权必究.举报:010-62782989, beiqinquan@tup.tsinghua.edu.cn.

图书在版编目(CIP)数据

数据科学与工程数学基础 / 高胜哲, 顾剑主编. -- 北京: 清华大学出版社, 2026. 5.

ISBN 978-7-302-71465-1

I. TP274;TB11

中国国家版本馆 CIP 数据核字第 2026KW2736 号

责任编辑:刘颖

封面设计:傅瑞学

责任校对:薄军霞

责任印制:刘菲

出版发行:清华大学出版社

网 址: <https://www.tup.com.cn>, <https://www.wqxuetang.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-83470000 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:三河市人民印务有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:15.25

字 数:371千字

版 次:2026年5月第1版

印 次:2026年5月第1次印刷

定 价:49.00元

产品编号:113317-01

前 言

在大数据技术飞速发展的当下,数据已成为驱动社会经济发展的基础性战略资源,其应用的深度与广度持续拓展,贯穿各行各业的数字化转型进程.无论是人工智能、机器学习等前沿技术的突破,还是金融分析、生物信息学等交叉领域的创新,均以扎实的数学理论作为底层支撑,数学已成为解锁大数据价值的核心钥匙.

本书紧扣数据科学与大数据技术及相关专业的教学目标,以“夯实基础、服务应用”为导向,致力于帮助学生构建系统且稳固的数学知识体系,为后续深入学习大数据核心技术、专业进阶课程及开展实践创新筑牢根基.书中聚焦现代数据分析与机器学习所需的核心数学知识,摒弃冗余理论,突出“实用、够用”原则,引导学生精准掌握关键内容,使其能熟练运用数学工具拆解、分析各类数据处理算法的内在逻辑,为专业必修课与选修课的深度学习打下坚实的数学理论基础.

全书共设9章,在充分衔接微积分、线性代数、概率论与数理统计等先修平台课程的基础上,系统讲解线性空间与线性变换、矩阵分析、矩阵分解、矩阵计算问题、概率模型和参数估计、最优化基础知识、最优性条件和对偶理论、优化算法、信息论基础等内容,形成适配应用型人才培养需求的完整知识框架.

由于编者团队学识水平有限,书中难免存在疏漏与不妥之处,恳请各位读者不吝赐教、批评指正,您的宝贵意见将是我们完善教材的重要动力.

编 者

2025年10月于大连

目 录

第 1 章 线性空间与线性变换	1
1.1 线性空间	1
1.1.1 线性空间的定义.....	1
1.1.2 线性空间的简单性质.....	3
1.2 线性空间的基与坐标	3
1.2.1 基、维数与坐标	3
1.2.2 基变换与坐标变换.....	5
1.3 线性子空间	8
1.3.1 线性子空间的定义.....	8
1.3.2 生成子空间.....	9
1.3.3 子空间的交、和与直和.....	10
1.4 线性变换.....	11
1.4.1 线性变换的定义与性质	11
1.4.2 线性变换的运算	12
1.4.3 线性变换的矩阵表示	13
习题 1	15
第 2 章 矩阵分析	18
2.1 内积与范数.....	18
2.1.1 内积	18
2.1.2 向量范数	19
2.1.3 矩阵范数	19
2.1.4 相似性度量	21
2.2 矩阵序列.....	22
2.2.1 矩阵序列的极限	22
2.2.2 收敛矩阵	24
2.3 矩阵级数.....	24
2.3.1 矩阵级数的敛散性	24
2.3.2 矩阵幂级数	25
2.4 矩阵函数.....	27
2.4.1 矩阵函数的定义	27

2.4.2	矩阵函数值的计算	28
2.4.3	常用矩阵函数的性质	30
2.5	矩阵的微分和积分	31
2.5.1	函数矩阵的微分和积分	32
2.5.2	数量函数对矩阵变量的导数	35
2.5.3	矩阵值函数对矩阵变量的导数	36
习题 2	38
第 3 章	矩阵分解	40
3.1	矩阵的三角分解	40
3.1.1	高斯消元法的矩阵形式——三角分解	40
3.1.2	三角分解的紧凑计算格式	44
3.1.3	楚列斯基分解	46
3.2	矩阵的 QR 分解	48
3.2.1	吉文斯矩阵和吉文斯变换	48
3.2.2	QR 分解	51
3.3	矩阵的谱分解	57
3.3.1	单纯矩阵的谱分解	57
3.3.2	正规矩阵的谱分解	61
3.4	矩阵的奇异值分解	62
3.4.1	舒尔引理及正规矩阵的分解	62
3.4.2	奇异值分解	64
习题 3	67
第 4 章	矩阵计算问题	69
4.1	线性方程组的数值解法	69
4.1.1	直接方法	69
4.1.2	迭代方法	70
4.2	最小二乘问题与回归	74
4.2.1	最小二乘法	74
4.2.2	回归问题	75
4.3	矩阵的特征值与特征向量的计算	82
4.3.1	圆盘定理	82
4.3.2	幂法与反幂法	84
4.4	应用举例	87
4.4.1	网页等级排序算法	87
4.4.2	主成分分析	89
习题 4	91

第 5 章 概率模型和参数估计	93
5.1 参数估计	93
5.1.1 矩估计	93
5.1.2 极大似然估计	95
5.1.3 贝叶斯估计	98
5.1.4 点估计的优良性	101
5.1.5 区间估计	103
5.1.6 EM 算法	109
5.2 统计决策理论	111
5.2.1 损失函数和风险函数	112
5.2.2 经验风险最小化和结构风险最小化	114
5.3 概率图模型	115
5.3.1 概率无向图模型	115
5.3.2 贝叶斯网络	117
习题 5	118
第 6 章 最优化基础知识	120
6.1 最优化问题概述	120
6.1.1 最优化问题的基本概念	120
6.1.2 几类重要的最优化问题	121
6.1.3 数据科学中常见的优化问题	123
6.2 凸优化基础知识	126
6.2.1 多元函数微分学基础知识	126
6.2.2 凸集	127
6.2.3 凸函数	128
6.2.4 共轭函数	129
习题 6	130
第 7 章 最优性条件和对偶理论	131
7.1 无约束优化问题的最优性条件	131
7.1.1 无约束优化问题的一阶最优性条件	131
7.1.2 无约束优化问题的二阶最优性条件	131
7.2 约束优化问题的最优性条件	132
7.2.1 紧约束指标集	132
7.2.2 约束优化问题的一阶最优性条件	133
7.3 对偶理论	134
7.3.1 线性规划的对偶问题	134
7.3.2 线性规划的对偶定理	137
习题 7	138

第 8 章 优化算法	139
8.1 无约束优化算法	139
8.1.1 下降算法的基本思想.....	139
8.1.2 一维搜索.....	140
8.1.3 求解无约束优化问题的下降算法.....	143
8.2 约束优化算法	148
8.2.1 可行方向法.....	148
8.2.2 惩罚函数法与障碍函数法.....	151
8.3 机器学习常用优化算法	153
8.3.1 预备知识.....	154
8.3.2 随机梯度下降算法.....	154
8.3.3 自适应梯度算法.....	155
习题 8	156
第 9 章 信息论基础	157
9.1 信息论简介	157
9.1.1 信息论主要内容.....	157
9.1.2 信源的数学模型.....	157
9.2 熵、联合熵、条件熵和互信息	158
9.2.1 熵.....	158
9.2.2 联合熵与条件熵.....	160
9.2.3 互信息.....	161
9.2.4 多个离散型随机变量的互信息.....	162
9.2.5 相对熵与交叉熵.....	163
9.3 微分熵	165
9.3.1 定义.....	165
9.3.2 微分熵与离散熵的关系.....	166
9.3.3 联合微分熵、条件微分熵与互信息	167
9.3.4 微分熵的性质.....	168
习题 9	169
附录 A 微积分基础	170
A.1 函数的极限与连续	170
A.1.1 函数的极限	170
A.1.2 函数极限的性质	171
A.1.3 函数的连续性	171
A.2 导数与微分	172
A.3 不定积分	173
A.3.1 原函数与不定积分的概念	173

A. 3.2	不定积分的性质	173
A. 3.3	不定积分的计算	174
A. 4	定积分	174
A. 4.1	定积分的定义	174
A. 4.2	定积分的性质	175
A. 4.3	定积分的计算	175
A. 4.4	定积分的元素法	176
A. 5	多元函数的微分学	176
A. 5.1	多元函数的极限	176
A. 5.2	多元函数的连续	177
A. 5.3	多元函数的偏导数	177
A. 5.4	多元复合函数的偏导数	178
A. 5.5	多元隐函数的偏导数	178
A. 5.6	多元函数的全微分	179
A. 6	重积分	180
附录 B 线性代数基础		181
B. 1	矩阵的基本概念	181
B. 1.1	矩阵的定义	181
B. 1.2	特殊矩阵	181
B. 2	矩阵的运算	182
B. 2.1	矩阵加法	182
B. 2.2	数乘矩阵	183
B. 2.3	矩阵乘法	183
B. 2.4	矩阵的转置	184
B. 2.5	方阵的行列式	185
B. 2.6	逆矩阵	185
B. 2.7	矩阵的初等变换	186
B. 2.8	矩阵的秩	187
B. 3	线性方程组的一般理论	187
B. 4	特征值和特征向量的概念及性质	189
附录 C 概率论基础		190
C. 1	概率论基本概念	190
C. 1.1	试验、样本空间和事件	190
C. 1.2	概率	190
C. 1.3	条件概率、贝叶斯理论和独立性	191
C. 2	随机变量及概率分布函数	193
C. 2.1	随机变量	193

C. 2. 2	离散型随机变量及其分布	194
C. 2. 3	连续型随机变量及其分布	195
C. 2. 4	多维随机变量及其分布	197
C. 3	随机变量的数字特征	200
C. 3. 1	期望	200
C. 3. 2	方差	201
C. 3. 3	协方差和相关系数	202
C. 3. 4	矩和协方差矩阵	203
C. 4	大数定律与中心极限定理	205
C. 4. 1	概率不等式	205
C. 4. 2	大数定律	207
C. 4. 3	中心极限定理	207
习题提示与答案		209

第 1 章

线性空间与线性变换

线性空间是由集合、数域和线性运算这三个要素共同构成的抽象化概念. 空间向量、矩阵、函数和数列等诸多对象都可以用线性空间的观点统一处理. 线性变换是一个线性空间到自身的保持线性运算(加法与数量乘法)的映射. 线性空间提供“舞台”, 线性变换则是“演员”, 两者结合构成了线性代数的核心框架. 它们不仅是纯数学理论的基础, 更是解决实际问题(如数据降维、密码学)的重要工具, 为现代科学和工程技术的发展提供了强大的数学支撑.

1.1 线性空间

1.1.1 线性空间的定义

在介绍线性空间的定义之前, 我们先看具体的例子. 数域 F 上的 n 维向量空间 F^n , 定义了两个向量的加法和数量乘法:

$$\begin{aligned}(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) &= (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n), \\ k(a_1, a_2, \dots, a_n) &= (ka_1, ka_2, \dots, ka_n), \quad k \in F,\end{aligned}$$

而且这两种运算满足如下的运算规律, 对于任意的 $\alpha, \beta, \gamma \in F^n$ 以及任意 $k, l \in F$, 有

$$\begin{aligned}\alpha + \beta &= \beta + \alpha, \quad (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma), \quad \alpha + 0 = \alpha, \quad \alpha + (-\alpha) = 0, \\ 1\alpha &= \alpha, \quad k(l\alpha) = (kl)\alpha, \quad (k+l)\alpha = k\alpha + l\alpha, \quad k(\alpha + \beta) = k\alpha + k\beta.\end{aligned}$$

再看一个例子, 数域 F 上的一元多项式集合 $F[x]$ 中, 定义了两个多项式的加法和数与多项式的乘法, 同样这两种运算满足如下的运算规律, 对于任意的 $f(x), g(x), h(x) \in F[x]$ 以及任意 $k, l \in F$, 有

$$\begin{aligned}f(x) + g(x) &= g(x) + f(x), \quad (f(x) + g(x)) + h(x) = f(x) + (g(x) + h(x)), \\ f(x) + 0 &= f(x), \quad f(x) + (-f(x)) = 0, \quad 1f(x) = f(x), \quad k(lf(x)) = (kl)f(x), \\ (k+l)f(x) &= kf(x) + lf(x), \quad k(f(x) + g(x)) = kf(x) + kg(x).\end{aligned}$$

定义 1.1 设 V 是一个非空集合, F 是一个数域, 在集合 V 中定义一种代数运算, 称为**加法**, 即对于 V 中任意两个元素 α, β , 在 V 中都存在唯一的一个元素 γ 与它们对应, 称 γ 为 α 与 β 的**和**, 记为 $\gamma = \alpha + \beta$; 在 F 与 V 的元素之间还定义了一种运算, 称为**数量乘法**, 即对于 V 中任意元素 α 与 F 中任意元素 k , 在 V 中都存在唯一的一个元素 δ 与它们对应, 称 δ 为 k 与 α 的**数量乘积**, 记为 $\delta = k\alpha$. 若加法和数量乘法还满足下述 8 条运算律, 则称 V 为数域 F 上的**线性空间**. 对于任意的 $\alpha, \beta, \gamma \in V$ 以及任意 $k, l \in F$, 有

$$(1) \alpha + \beta = \beta + \alpha;$$

- (2) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$;
- (3) V 中存在元素 0 (称为 V 的零元素), 使得对于任意的 $\alpha \in V$, 都有 $\alpha + 0 = \alpha$;
- (4) V 中每一个元素 α , 都存在元素 $\beta \in V$, 使得 $\alpha + \beta = 0$, 称此 β 为 α 的负元素;
- (5) $1\alpha = \alpha$;
- (6) $k(l\alpha) = (kl)\alpha$;
- (7) $(k+l)\alpha = k\alpha + l\alpha$;
- (8) $k(\alpha + \beta) = k\alpha + k\beta$.

注 1.1 (1) 凡满足以上 8 条运算律的加法及数量乘法统称为线性运算.

(2) 线性空间的元素也称为向量, 线性空间也称向量空间, 但这里的向量不一定是有序数组.

(3) 若集合对于定义的加法和数乘运算不封闭, 或者运算封闭但不满足 8 条运算律中的任一条, 则此集合就不能构成线性空间.

例 1.1 数域 F 上的次数小于 n 的多项式的全体, 再添上零多项式组成集合 V , 按多项式的加法和数量乘法构成数域 F 上的一个线性空间, 常用 $F[x]_n$ 表示, 即

$$F[x]_n = \{a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 \mid a_{n-1}, a_{n-2}, \cdots, a_1, a_0 \in F\}.$$

例 1.2 数域 F 上 $m \times n$ 矩阵的全体组成集合 V , 按矩阵的加法和数量乘法, 构成数域 F 上的一个线性空间, 用 $F^{m \times n}$ 表示.

例 1.3 任一数域 F 按照自身的加法与乘法构成一个数域 F 上的线性空间.

例 1.4 全体实函数构成的集合 V , 按函数的加法和数与函数的乘法, 构成实数域 \mathbb{R} 上的一个线性空间.

例 1.5 全体正实数 \mathbb{R}^+ , $\forall a, b \in \mathbb{R}^+$, $\forall k \in \mathbb{R}$, 加法与数量乘法定义为

$$(1) a \oplus b = \log_a b, k \circ a = a^k; \quad (2) a \oplus b = ab, k \circ a = a^k.$$

判断 \mathbb{R}^+ 是否构成实数域 \mathbb{R} 上的线性空间.

解 (1) \mathbb{R}^+ 不构成实数域 \mathbb{R} 上的线性空间. 因为 \oplus 不封闭, 如 $2 \oplus \frac{1}{2} = \log_2 \frac{1}{2} = -1 \notin \mathbb{R}^+$.

(2) \mathbb{R}^+ 构成实数域 \mathbb{R} 上的线性空间. 首先, $\mathbb{R}^+ \neq \emptyset$ 且加法和数量乘法对 \mathbb{R}^+ 是封闭的. 事实上, $\forall a, b \in \mathbb{R}^+$, $a \oplus b = ab \in \mathbb{R}^+$ 且 ab 唯一确定; $\forall a \in \mathbb{R}^+$, $\forall k \in \mathbb{R}$, $k \circ a = a^k \in \mathbb{R}^+$ 且 a^k 唯一确定.

其次, 加法和数量乘法满足以下 8 条运算律:

$$\textcircled{1} \forall a, b \in \mathbb{R}^+, a \oplus b = ab = ba = b \oplus a;$$

$$\textcircled{2} \forall a, b \in \mathbb{R}^+, (a \oplus b) \oplus c = (ab) \oplus c = (ab)c = a(bc) = a \oplus (bc) = a \oplus (b \oplus c);$$

$$\textcircled{3} \forall a \in \mathbb{R}^+, 1 \in \mathbb{R}^+, a \oplus 1 = a1 = a, \text{ 即 } 1 \text{ 是零元};$$

$$\textcircled{4} \forall a \in \mathbb{R}^+, \text{ 存在 } \frac{1}{a} \in \mathbb{R}^+ \text{ 使得 } a \oplus \frac{1}{a} = a \frac{1}{a} = 1, \text{ 即 } a \text{ 的负元素是 } \frac{1}{a};$$

$$\textcircled{5} \forall a \in \mathbb{R}^+, 1 \circ a = a^1 = a;$$

$$\textcircled{6} \forall a \in \mathbb{R}^+, \forall k, l \in \mathbb{R}, k \circ (l \circ a) = k \circ a^l = (a^l)^k = a^{lk} = a^{kl} = (kl) \circ a;$$

$$\textcircled{7} \forall a \in \mathbb{R}^+, \forall k, l \in \mathbb{R}, (k+l) \circ a = a^{k+l} = a^k a^l = a^k \oplus a^l = (k \circ a) \oplus (l \circ a);$$

$$\textcircled{8} \forall a, b \in \mathbb{R}^+, \forall k \in \mathbb{R}, k \circ (a \oplus b) = k \circ (ab) = (ab)^k = a^k b^k = a^k \oplus b^k = (k \circ a) \oplus (k \circ b).$$

综上所述, \mathbb{R}^+ 构成实数域 \mathbb{R} 上的线性空间.

1.1.2 线性空间的简单性质

性质 1.1 线性空间 V 的零元素是唯一的.

证明 假设线性空间 V 有两个零元素 0_1 和 0_2 , 则有 $0_1 = 0_1 + 0_2 = 0_2$. □

性质 1.2 线性空间 V 中任意元素 α 的负元素是唯一的, 记为 $-\alpha$.

证明 假设 α 有两个负元素 β 和 γ , 则 $\alpha + \beta = 0 = \alpha + \gamma$, 所以

$$\beta = \beta + 0 = \beta + (\alpha + \gamma) = (\beta + \alpha) + \gamma = (\alpha + \beta) + \gamma = 0 + \gamma = \gamma. \quad \square$$

性质 1.3 $0\alpha = 0, k0 = 0, (-1)\alpha = -\alpha$.

证明 $0\alpha + \alpha = (0+1)\alpha = \alpha$, 两边加上 $-\alpha$ 即得 $0\alpha = 0$; $k\alpha = k(\alpha + 0) = k\alpha + k0$, 两边加上 $-k\alpha$ 即得 $k0 = 0$; $\alpha + (-1\alpha) = (1\alpha) + (-1\alpha) = (1-1)\alpha = 0\alpha = 0$, 两边加上 $-\alpha$ 即得 $(-1)\alpha = -\alpha$. □

性质 1.4 如果 $k\alpha = 0$, 那么 $k = 0$ 或 $\alpha = 0$.

证明 假设 $k \neq 0$, 则 $\alpha = (k^{-1}k)\alpha = k^{-1}(k\alpha) = k^{-1}0 = 0$. □

1.2 线性空间的基与坐标

如何把线性空间的全体元素表示出来? 这些元素之间的关系又如何呢? 线性空间的构造如何? 线性空间是抽象的, 如何使其元素与具体的数发生联系, 使其能用比较具体的数学式子来表达? 怎样才能便于运算? 线性空间的基与坐标就是为解决这些问题而提出的概念.

1.2.1 基、维数与坐标

定义 1.2 设 V 是数域 F 上的线性空间, $\alpha_1, \alpha_2, \dots, \alpha_r$ 是 V 的 r 个向量. 若存在 r 个不全为零的数 $k_1, k_2, \dots, k_r \in F$, 使得

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_r\alpha_r = 0$$

成立, 则称 $\alpha_1, \alpha_2, \dots, \alpha_r$ 是线性相关的. 否则称 $\alpha_1, \alpha_2, \dots, \alpha_r$ 是线性无关的.

定义 1.3 若线性空间 V 中可找到任意多个线性无关的向量, 则称 V 是无限维线性空间. 若在线性空间 V 中有 n 个线性无关的向量, 但任意 $n+1$ 个向量都是线性相关的, 则称 V 是一个 n 维线性空间, 常记作 $\dim V = n$.

例 1.6 对于任意的正整数 n , 都有 n 个线性无关的向量 $1, x, x^2, \dots, x^{n-1}$, 故所有实系数多项式构成的线性空间 $\mathbb{R}[x]$ 是无限维的.

定义 1.4 在 n 维线性空间 V 中, n 个线性无关的向量称为 V 的一组基.

定义 1.5 设 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是线性空间 V 的一组基且 $\alpha \in V$, 若存在数 $a_1, a_2, \dots, a_n \in F$, 使得 $\alpha = a_1\epsilon_1 + a_2\epsilon_2 + \dots + a_n\epsilon_n$ 成立, 则数组 a_1, a_2, \dots, a_n 称为 α 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 下的坐标, 记为 $(a_1, a_2, \dots, a_n)^T$. 有时也形式地记为

$$\alpha = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

注 1.2 向量 α 的坐标 $(a_1, a_2, \dots, a_n)^T$ 是被向量 α 和基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 唯一确定的, 即 α 在基 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 下的坐标是唯一的. 但是, 在不同基下 α 的坐标一般是不同的.

定理 1.1 若线性空间 V 中的向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 满足

(1) $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关;

(2) 任意 $\beta \in V$, 都存在数 $k_1, k_2, \dots, k_n \in F$ 使得 $\beta = k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n$ (称为 β 可由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出).

则 V 为 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 为 V 的一组基.

证明 因 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, 故 V 的维数至少为 n . 任取 V 中 $n+1$ 个向量 $\beta_1, \beta_2, \dots, \beta_n, \beta_{n+1}$, 由 (2) 可知, $\beta_1, \beta_2, \dots, \beta_n, \beta_{n+1}$ 可由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表出. 若 $\beta_1, \beta_2, \dots, \beta_n, \beta_{n+1}$ 是线性无关的, 则 $n+1 \leq n$, 矛盾. 故 V 中任意 $n+1$ 个向量 $\beta_1, \beta_2, \dots, \beta_n, \beta_{n+1}$ 是线性相关的. 故 V 为 n 维线性空间, $\alpha_1, \alpha_2, \dots, \alpha_n$ 为 V 的一组基. \square

注 1.3 n 维线性空间 V 的基不是唯一的, V 中任意 n 个线性无关的向量都是 V 的一组基.

例 1.7 已知三维几何空间 $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$, $\varepsilon_1 = (1, 0, 0)$, $\varepsilon_2 = (0, 1, 0)$, $\varepsilon_3 = (0, 0, 1)$ 是 \mathbb{R}^3 的一组基; $\alpha_1 = (1, 1, 1)$, $\alpha_2 = (1, 1, 0)$, $\alpha_3 = (1, 0, 0)$ 也是 \mathbb{R}^3 的一组基.

一般地, 向量空间 $F^n = \{(a_1, a_2, \dots, a_n) \mid a_i \in F, i=1, 2, \dots, n\}$ 是 n 维的, 且

$$\varepsilon_1 = (1, 0, \dots, 0), \varepsilon_2 = (0, 1, \dots, 0), \dots, \varepsilon_n = (0, \dots, 0, 1)$$

是 F^n 的一组基, 称为 F^n 的标准基.

例 1.8 证明: (1) 线性空间 $F[x]_n$ 是 n 维的, 并且 $1, x, x^2, \dots, x^{n-1}$ 为 $F[x]_n$ 的一组基;

(2) $1, x-a, (x-a)^2, \dots, (x-a)^{n-1}$ 也为 $F[x]_n$ 的一组基.

证明 (1) 首先, $1, x, x^2, \dots, x^{n-1}$ 是线性无关的.

其次, 对任意多项式 $f(x) = a_0 + a_1x + \dots + a_{n-1}x^{n-1} \in F[x]_n$, 显然 $f(x)$ 可由 $1, x, x^2, \dots, x^{n-1}$ 线性表出. 所以, $1, x, x^2, \dots, x^{n-1}$ 是 $F[x]_n$ 的一组基, 从而, 线性空间 $F[x]_n$ 是 n 维的. 并且此时, $f(x) = a_0 + a_1x + \dots + a_{n-1}x^{n-1}$ 在基 $1, x, x^2, \dots, x^{n-1}$ 下的坐标为 $(a_0, a_1, \dots, a_{n-1})$.

(2) 首先, $1, x-a, (x-a)^2, \dots, (x-a)^{n-1}$ 是线性无关的.

其次, 对任意多项式 $f(x) \in F[x]_n$, 按泰勒展开式有

$$f(x) = f(a) + f'(a)(x-a) + \dots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1},$$

由此可见, $f(x)$ 可由 $1, x-a, (x-a)^2, \dots, (x-a)^{n-1}$ 线性表出. 因此 $1, x-a, (x-a)^2, \dots, (x-a)^{n-1}$ 为 $F[x]_n$ 的一组基. 且 $f(x)$ 在基 $1, x-a, (x-a)^2, \dots, (x-a)^{n-1}$ 下的坐标为 $(f(a), f'(a), \dots, \frac{f^{(n-1)}(a)}{(n-1)!})$. \square

例 1.9 求全体复数的集合 \mathbb{C} 看成复数域 \mathbb{C} 上的线性空间的维数与一组基; 若把 \mathbb{C} 视为实数域 \mathbb{R} 上的线性空间呢?

解 复数域 \mathbb{C} 上的线性空间 \mathbb{C} 是一维的, 数 1 就是它的一组基; 而实数域 \mathbb{R} 上的线性空间 \mathbb{C} 为二维的, 数 1 和 $i = \sqrt{-1}$ 就为它的一组基.

注 1.4 任意数域 F 视为它自身上的线性空间是一维的, 数 1 就是它的一组基.

例 1.10 求数域 F 上的线性空间 $F^{2 \times 2}$ 的维数和一组基.

解 令 $\mathbf{E}_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\mathbf{E}_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\mathbf{E}_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{E}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, 则易验证 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 是线性无关的. 事实上, 由

$$a\mathbf{E}_{11} + b\mathbf{E}_{12} + c\mathbf{E}_{21} + d\mathbf{E}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \text{ 解得, } \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

所以 $a=b=c=d=0$.

对任意矩阵 $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in F^{2 \times 2}$, 有 $\mathbf{A} = a_{11}\mathbf{E}_{11} + a_{12}\mathbf{E}_{12} + a_{21}\mathbf{E}_{21} + a_{22}\mathbf{E}_{22}$. 所以 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 是线性空间 $F^{2 \times 2}$ 的一组基, 并且 $F^{2 \times 2}$ 是 4 维的. 并且, $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ 在基 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 下的坐标为 $(a_{11}, a_{12}, a_{21}, a_{22})^T$.

一般地, 数域 F 上的全体 $m \times n$ 矩阵构成的线性空间 $F^{m \times n}$ 为 $m \times n$ 维的, 令 \mathbf{E}_{ij} 记为第 i 行第 j 列为 1, 其余全为 0 的 $m \times n$ 矩阵, 那么 $\{\mathbf{E}_{ij} \mid i=1, 2, \dots, m; j=1, 2, \dots, n\}$ 为 $F^{m \times n}$ 的一组基. 并且对任意矩阵 $\mathbf{A} = (a_{ij}) \in F^{m \times n}$, 有 $\mathbf{A} = \sum_{i=1}^m \sum_{j=1}^n a_{ij}\mathbf{E}_{ij}$.

例 1.11 已知 $\alpha_1 = (1, 1, 1, 1)^T$, $\alpha_2 = (1, 1, -1, -1)^T$, $\alpha_3 = (1, -1, 1, -1)^T$, $\alpha_4 = (1, -1, -1, 1)^T$ 为线性空间 F^4 的一组基, 求向量 $\xi = (1, 2, 1, 1)^T$ 在基 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 下的坐标.

解 设 $\xi = x_1\alpha_1 + x_2\alpha_2 + x_3\alpha_3 + x_4\alpha_4$, 则有线性方程组

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 1, \\ x_1 + x_2 - x_3 - x_4 = 2, \\ x_1 - x_2 + x_3 - x_4 = 1, \\ x_1 - x_2 - x_3 + x_4 = 1, \end{cases}$$

解得 $x_1 = \frac{5}{4}, x_2 = \frac{1}{4}, x_3 = -\frac{1}{4}, x_4 = -\frac{1}{4}$.

所以向量 $\xi = (1, 2, 1, 1)^T$ 在基 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 下的坐标为 $\left(\frac{5}{4}, \frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}\right)^T$.

1.2.2 基变换与坐标变换

定义 1.6 设 V 为数域 F 上的 n 维线性空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 和 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 是 V 中的两组基, 若

$$\begin{cases} \epsilon'_1 = a_{11}\epsilon_1 + a_{21}\epsilon_2 + \dots + a_{n1}\epsilon_n, \\ \epsilon'_2 = a_{12}\epsilon_1 + a_{22}\epsilon_2 + \dots + a_{n2}\epsilon_n, \\ \vdots \\ \epsilon'_n = a_{1n}\epsilon_1 + a_{2n}\epsilon_2 + \dots + a_{nn}\epsilon_n, \end{cases} \quad (1.1)$$

即

$$(\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n) = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad (1.2)$$

则称矩阵 $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ 为由基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 的过渡矩阵, 称

(1.1)式或(1.2)式为由基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 的基变换公式.

性质 1.5 过渡矩阵都是可逆矩阵. 若由基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 的过渡矩阵为 \mathbf{A} , 则由基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 到基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 的过渡矩阵为 \mathbf{A}^{-1} .

性质 1.6 若由基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 的过渡矩阵为 \mathbf{A} , 由基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 到基 $\epsilon''_1, \epsilon''_2, \dots, \epsilon''_n$ 的过渡矩阵为 \mathbf{B} , 则由基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\epsilon''_1, \epsilon''_2, \dots, \epsilon''_n$ 的过渡矩阵为 \mathbf{AB} .

定义 1.7 已知 V 为数域 F 上的 n 维线性空间, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 和 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 是 V 中的两组基, 且由基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 的过渡矩阵为 \mathbf{A} , 即

$$(\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n) = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)\mathbf{A}. \quad (1.3)$$

设 $\xi \in V$ 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 与 $\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n$ 下的坐标分别为 $(x_1, x_2, \dots, x_n)^T$ 与 $(x'_1, x'_2, \dots, x'_n)^T$, 即

$$\xi = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_n) \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix},$$

则

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{A} \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}, \quad (1.4)$$

或

$$\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \quad (1.5)$$

则称(1.4)式或(1.5)式为向量 ξ 在基变换 (1.3) 下的坐标变换公式.

例 1.12 已知 $\epsilon_1 = (1, 0, \dots, 0), \epsilon_2 = (0, 1, \dots, 0), \dots, \epsilon_n = (0, \dots, 0, 1)$ 与 $\eta_1 = (1, 1, \dots, 1), \eta_2 = (0, 1, \dots, 1), \dots, \eta_n = (0, \dots, 0, 1)$ 是线性空间 F^n 的两组基. 求: (1) 由基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 到基 $\eta_1, \eta_2, \dots, \eta_n$ 的过渡矩阵; (2) 由基 $\eta_1, \eta_2, \dots, \eta_n$ 到基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 的过渡矩阵; (3) 向量 $\alpha = (a_1, a_2, \dots, a_n)^T$ 在基 $\eta_1, \eta_2, \dots, \eta_n$ 下的坐标.

从而

$$\begin{aligned}
 (\mu_1, \mu_2, \mu_3, \mu_4) &= (\eta_1, \eta_2, \eta_3, \eta_4) \begin{pmatrix} 1 & 1 & -1 & -1 \\ 2 & -1 & 2 & -1 \\ -1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2 & 0 & -2 & 1 \\ 1 & 1 & 1 & 3 \\ 0 & 2 & 1 & 1 \\ 1 & 2 & 2 & 2 \end{pmatrix} \\
 &= (\eta_1, \eta_2, \eta_3, \eta_4) \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},
 \end{aligned}$$

所以,由基 $\eta_1, \eta_2, \eta_3, \eta_4$ 到基 $\mu_1, \mu_2, \mu_3, \mu_4$ 的过渡矩阵为

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

1.3 线性子空间

1.3.1 线性子空间的定义

定义 1.8 设 V 为数域 F 上的线性空间,集合 $W \subseteq V (W \neq \emptyset)$,若 W 对于 V 中的两种运算也构成数域 F 上的线性空间,则称 W 为 V 的一个线性子空间,简称子空间.

注 1.5 (1)线性子空间也是数域 F 上一个线性空间,也有基与维数的概念.(2)任一线性子空间的维数不能超过整个空间的维数.(3)设 V 为数域 F 上的线性空间,只含零向量的子集合 $W = \{0\}$ 是 V 的一个线性子空间,被称为 V 的零子空间.线性空间 V 本身也是 V 的一个子空间.这两个子空间有时称为平凡子空间,而其他的子空间称为非平凡子空间.

定理 1.2 设 V 为数域 F 上的线性空间,集合 $W \subseteq V (W \neq \emptyset)$,若 W 对于 V 中的两种运算封闭,即 $\forall \alpha, \beta \in W$,有 $\alpha + \beta \in W$; $\forall \alpha \in W, \forall k \in F$,有 $k\alpha \in W$,则 W 是 V 的一个线性子空间.

例 1.14 设 V 为所有实函数在实数域 \mathbb{R} 上构成的线性空间,则 $\mathbb{R}[x]$ 为 V 的一个线性子空间, $\mathbb{R}[x]_n$ 是 $\mathbb{R}[x]$ 的一个线性子空间.

例 1.15 n 元齐次线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0, \\ \vdots \\ a_{s1}x_1 + a_{s2}x_2 + \cdots + a_{sn}x_n = 0 \end{cases} \quad (1.6)$$

的全部解向量所构成的集合 W ,对于通常的向量加法和数量乘法构成的线性空间是 n 维向量空间 F^n 的一个子空间,称 W 为方程组(1.6)的解空间.

注 1.6 (1)方程组(1.6)的解空间 W 的维数等于 $n - R(\mathbf{A})$,其中 \mathbf{A} 为方程组的系数矩阵, $R(\mathbf{A})$ 为矩阵 \mathbf{A} 的秩;(2)方程组(1.6)的一个基础解系就是解空间 W 的一组基.

例 1.16 判断 F^n 的下列子集哪些是子空间:

$$W_1 = \{(x_1, x_2, \cdots, x_n) \mid x_1 + x_2 + \cdots + x_n = 0, x_i \in F\};$$

$$W_2 = \{(x_1, x_2, \dots, x_n) \mid x_1 + x_2 + \dots + x_n = 1, x_i \in F\};$$

$$W_3 = \{(x_1, x_2, \dots, x_{n-1}, 0) \mid x_i \in F, i = 1, 2, \dots, n-1\}.$$

若为 F^n 的子空间, 求出其维数与一组基.

解 W_1 和 W_3 是 F^n 的子空间, 而 W_2 不是 F^n 的子空间.

事实上, W_1 是 n 元齐次线性方程组

$$x_1 + x_2 + \dots + x_n = 0 \quad (1.7)$$

的解空间, 所以 $\dim(W_1) = n - 1$. 方程组 (1.7) 的一个基础解系 $\eta_1 = (1, -1, 0, \dots, 0)$, $\eta_2 = (1, 0, -1, 0, \dots, 0), \dots, \eta_{n-1} = (1, 0, \dots, 0, -1)$ 是 W_1 的一组基.

而在 W_2 中任取两个向量 $\alpha = (x_1, x_2, \dots, x_n)$ 和 $\beta = (y_1, y_2, \dots, y_n)$, 则

$$\alpha + \beta = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n),$$

但 $(x_1 + y_1) + (x_2 + y_2) + \dots + (x_n + y_n) = (x_1 + x_2 + \dots + x_n) + (y_1 + y_2 + \dots + y_n) = 1 + 1 = 2$, 即 $\alpha + \beta \notin W_2$, 故 W_2 不是 F^n 的子空间.

下证 W_3 是 F^n 的子空间.

首先, 因为 $\mathbf{0} = (0, 0, \dots, 0) \in W_3$, 所以 $W_3 \neq \emptyset$. 其次, 任取 $\alpha = (x_1, x_2, \dots, x_{n-1}, 0) \in W_3, \beta = (y_1, y_2, \dots, y_{n-1}, 0) \in W_3$ 及 $k \in F$, 则 $\alpha + \beta = (x_1 + y_1, x_2 + y_2, \dots, x_{n-1} + y_{n-1}, 0) \in W_3$ 及 $k\alpha = (kx_1, kx_2, \dots, kx_{n-1}, 0) \in W_3$, 故 W_3 是 F^n 的一个子空间, 且 $\dim(W_3) = n - 1, \{\epsilon_i = (0, \dots, 0, 1, 0, \dots, 0) \mid i = 1, 2, \dots, n-1\}$ 是 W_3 的一组基.

1.3.2 生成子空间

定义 1.9 设 V 为数域 F 上的线性空间, $\alpha_1, \alpha_2, \dots, \alpha_r \in V$, 则子空间

$$W = \{k_1\alpha_1 + k_2\alpha_2 + \dots + k_r\alpha_r \mid k_i \in F, i = 1, 2, \dots, r\}$$

称为 V 的由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 生成的子空间, 记为 $L(\alpha_1, \alpha_2, \dots, \alpha_r)$, 称 $\alpha_1, \alpha_2, \dots, \alpha_r$ 为子空间 $L(\alpha_1, \alpha_2, \dots, \alpha_r)$ 的一组生成元.

例 1.17 已知 $\epsilon_1 = (1, 0, \dots, 0), \epsilon_2 = (0, 1, \dots, 0), \dots, \epsilon_n = (0, \dots, 0, 1)$ 是线性空间 F^n 的一组基, 对于任意 $\alpha = (a_1, a_2, \dots, a_n) \in F^n$ 有 $\alpha = a_1\epsilon_1 + a_2\epsilon_2 + \dots + a_n\epsilon_n$, 故 $F^n = L(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$. 事实上, 任一有限维线性空间都可由它的一组基生成.

类似地, 有

$$F[x]_n = L(1, x, x^2, \dots, x^{n-1}) = \{a_0 + a_1x + \dots + a_{n-1}x^{n-1} \mid a_0, a_1, \dots, a_{n-1} \in F\}.$$

注 1.7 (1) 生成子空间 $L(\alpha_1, \alpha_2, \dots, \alpha_r)$ 的维数等于向量组 $\alpha_1, \alpha_2, \dots, \alpha_r$ 的秩, 即

$$\dim(L(\alpha_1, \alpha_2, \dots, \alpha_r)) = R(\alpha_1, \alpha_2, \dots, \alpha_r).$$

(2) 设 $\alpha_1, \alpha_2, \dots, \alpha_r$ 是线性空间 V 中不全为零的一组向量, $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_s} (s \leq r)$ 是它的一个极大无关组, 则 $L(\alpha_1, \alpha_2, \dots, \alpha_r) = L(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_s})$.

定理 1.3 (扩基定理) 设 W 为 n 维线性空间 V 的一个 m 维子空间, $\alpha_1, \alpha_2, \dots, \alpha_m$ 为 W 的一组基, 则这组基向量必定可扩充为 V 的一组基, 即在 V 中必定可找到 $n - m$ 个向量 $\alpha_{m+1}, \alpha_{m+2}, \dots, \alpha_n$ 使得 $\alpha_1, \alpha_2, \dots, \alpha_n$ 为 V 的一组基.

证明 对 $n - m$ 作数学归纳. 首先, 当 $n - m = 0$, 即 $n = m$ 时, $\alpha_1, \alpha_2, \dots, \alpha_m$ 就是 V 的一组基. 其次, 假设当 $n - m = k$ 时结论成立, 下面考虑 $n - m = k + 1$ 的情形.

既然 $\alpha_1, \alpha_2, \dots, \alpha_m$ 还不是 V 的一组基, 并且它还是线性无关的, 那么在 V 中一定存在

一个向量 α_{m+1} 不能被 $\alpha_1, \alpha_2, \dots, \alpha_m$ 线性表出, 因此 $\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}$ 必定线性无关. 由此可知子空间 $L(\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1})$ 是 $m+1$ 维的且 $n - (m+1) = (n-m) - 1 = (k+1) - 1 = k$, 由归纳假设, 子空间 $L(\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1})$ 的基 $\alpha_1, \alpha_2, \dots, \alpha_m, \alpha_{m+1}$ 可以扩充为整个空间 V 的一组基. 因此, 当 $n-m=k+1$ 时结论成立, 由数学归纳法可证定理成立. \square

例 1.18 设 $\alpha_1 = (1, -1, 2, 4), \alpha_2 = (0, 3, 1, 2), \alpha_3 = (3, 0, 7, 14), \alpha_4 = (1, -1, 2, 0), \alpha_5 = (2, 1, 5, 6)$, 求 $L(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ 的维数与一组基.

解 对以 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 为列向量的矩阵 A 作初等行变换

$$A = \begin{pmatrix} 1 & 0 & 3 & 1 & 2 \\ -1 & 3 & 0 & -1 & 1 \\ 2 & 1 & 7 & 2 & 5 \\ 4 & 2 & 14 & 0 & 6 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 3 & 1 & 2 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = B,$$

由 B 知, $\alpha_1, \alpha_2, \alpha_4$ 为 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ 的一个极大线性无关组, 故 $L(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ 的维数为 3, 且 $\alpha_1, \alpha_2, \alpha_4$ 就是 $L(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ 的一组基.

1.3.3 子空间的交、和与直和

本节介绍子空间的两种运算: 子空间的交与和, 以及子空间和的一个重要特殊情形, 即子空间的直和.

定理 1.4 若 V_1, V_2 为线性空间 V 的两个子空间, 则它们的交 $V_1 \cap V_2$ 也为 V 的子空间.

证明 首先, 由 $0 \in V_1, 0 \in V_2$ 知 $0 \in V_1 \cap V_2$, 故 $V_1 \cap V_2$ 是非空的. 其次, 若 $\alpha, \beta \in V_1 \cap V_2$, 即 $\alpha, \beta \in V_1$ 且 $\alpha, \beta \in V_2$, 则 $\alpha + \beta \in V_1$ 且 $\alpha + \beta \in V_2$, 因此 $\alpha + \beta \in V_1 \cap V_2$. 对数量乘法可类似地证明. 所以 $V_1 \cap V_2$ 也为 V 的子空间. \square

定义 1.10 设 V_1, V_2 为线性空间 V 的子空间, 称子集 $\{\alpha \mid \alpha = \alpha_1 + \alpha_2, \alpha_1 \in V_1, \alpha_2 \in V_2\}$ 为 V_1 与 V_2 的和, 并记为 $V_1 + V_2$.

定理 1.5 若 V_1, V_2 为线性空间 V 的两个子空间, 则它们的和 $V_1 + V_2$ 也为 V 的子空间.

证明 首先, 由 $0 \in V_1, 0 \in V_2$ 知 $0 \in V_1 + V_2$, 故 $V_1 + V_2$ 是非空的. 其次, 若 $\alpha, \beta \in V_1 + V_2$, 即 $\alpha = \alpha_1 + \alpha_2$, 且 $\alpha_1 \in V_1, \alpha_2 \in V_2$ 及 $\beta = \beta_1 + \beta_2$, 且 $\beta_1 \in V_1, \beta_2 \in V_2$, 则 $\alpha + \beta = (\alpha_1 + \beta_1) + (\alpha_2 + \beta_2)$. 又因 V_1, V_2 为线性空间 V 的子空间, 故 $\alpha_1 + \beta_1 \in V_1$ 及 $\alpha_2 + \beta_2 \in V_2$, 因此 $\alpha + \beta \in V_1 + V_2$. 类似地, $k\alpha = k\alpha_1 + k\alpha_2 \in V_1 + V_2$. 所以 $V_1 + V_2$ 为 V 的子空间. \square

关于两个子空间的交与和的维数, 有以下的维数定理(证明略).

定理 1.6(维数公式) 若 V_1, V_2 为线性空间 V 的两个子空间, 则

$$\dim(V_1) + \dim(V_2) = \dim(V_1 + V_2) + \dim(V_1 \cap V_2).$$

定义 1.11 设 V_1, V_2 为线性空间 V 的子空间, 若和 $V_1 + V_2$ 中的每个向量 α 的分解式

$$\alpha = \alpha_1 + \alpha_2, \quad \text{其中 } \alpha_1 \in V_1, \alpha_2 \in V_2$$

是唯一的, 则这个和称为直和, 并记为 $V_1 \oplus V_2$.

定理 1.7 和 $V_1 + V_2$ 是直和的充分必要条件为等式 $\alpha_1 + \alpha_2 = 0$, 其中 $\alpha_1 \in V_1, \alpha_2 \in V_2$ 只有在 $\alpha_1 = \alpha_2 = 0$ 时才成立.

证明 定理的条件实际上为零向量的分解式是唯一的, 这个条件显然是必要的, 下面证

明该条件是充分的.

设 $\alpha \in V_1 + V_2$ 有两个分解式, 即 $\alpha = \alpha_1 + \alpha_2$, 且 $\alpha_1 \in V_1, \alpha_2 \in V_2$ 及 $\alpha = \beta_1 + \beta_2$, 且 $\beta_1 \in V_1, \beta_2 \in V_2$, 则 $(\alpha_1 - \beta_1) + (\alpha_2 - \beta_2) = 0$, 其中 $\alpha_1 - \beta_1 \in V_1$ 及 $\alpha_2 - \beta_2 \in V_2$. 由定理条件有 $\alpha_1 - \beta_1 = 0, \alpha_2 - \beta_2 = 0$, 因此 $\alpha_1 = \beta_1, \alpha_2 = \beta_2$. 即 α 的分解式是唯一的. 所以 $V_1 + V_2$ 是直和. \square

推论 和 $V_1 + V_2$ 是直和的充分必要条件为 $V_1 \cap V_2 = \{0\}$.

证明 充分性 设有等式 $\alpha_1 + \alpha_2 = 0, \alpha_1 \in V_1, \alpha_2 \in V_2$, 则 $\alpha_1 = -\alpha_2 \in V_1 \cap V_2$, 由条件得 $\alpha_1 = \alpha_2 = 0$, 由定理 1.7 知 $V_1 + V_2$ 是直和.

必要性 任取 $\alpha \in V_1 \cap V_2$, 那么零向量可表示成 $0 = \alpha + (-\alpha)$, 其中 $\alpha \in V_1, -\alpha \in V_2$. 由直和有 $\alpha = -\alpha = 0$, 即 $V_1 \cap V_2 = \{0\}$. \square

定理 1.8 设 V_1, V_2 为线性空间 V 的子空间, 令 $W = V_1 + V_2$, 则 $W = V_1 \oplus V_2$ 的充分必要条件为 $\dim(W) = \dim(V_1) + \dim(V_2)$.

证明 由定理 1.6 有 $\dim(V_1) + \dim(V_2) = \dim(W) + \dim(V_1 \cap V_2)$, 再由定理 1.7 的推论知 $V_1 + V_2$ 是直和的充分必要条件为 $V_1 \cap V_2 = \{0\}$, 这与 $\dim(V_1 \cap V_2) = 0$ 是等价的, 也就与 $\dim(W) = \dim(V_1) + \dim(V_2)$ 等价. \square

1.4 线性变换

1.4.1 线性变换的定义与性质

定义 1.12 设 V 为数域 F 上的线性空间, 若映射 $\sigma: V \rightarrow V$ 满足对于任意 $\alpha, \beta \in V, k \in F$ 有

$$\sigma(\alpha + \beta) = \sigma(\alpha) + \sigma(\beta), \quad \sigma(k\alpha) = k\sigma(\alpha),$$

则称 σ 为线性空间 V 上的线性变换.

注 1.8 (1) 单位变换(恒等变换) $E: V \rightarrow V$ 定义为对于任意 $\alpha \in V, E(\alpha) = \alpha$;

(2) 零变换 $0: V \rightarrow V$ 定义为对于任意 $\alpha \in V, 0(\alpha) = 0$;

(3) 由数 k 决定的数乘变换 $K: V \rightarrow V$ 定义为对于任意 $\alpha \in V, K(\alpha) = k\alpha$.

例 1.19 设线性空间 $V = \mathbb{R}^2$, 把 V 中每一个向量绕坐标原点逆时针旋转 θ 角, 这就是一个线性变换, 用 T_θ 表示, 即 $T_\theta: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$.

例 1.20 设线性空间 $V = F[x]$ 或 $F[x]_n$, 映射 $D: V \rightarrow V$ 定义为 $D(f(x)) = f'(x), \forall f(x) \in V$, 则 D 为线性空间 V 上的线性变换.

例 1.21 记 $C(a, b)$ 为闭区间 $[a, b]$ 上全体实连续函数构成的线性空间, 定义映射 $J: C(a, b) \rightarrow C(a, b), J(f(x)) = \int_a^x f(t) dt, \forall f(x) \in C(a, b)$, 则 J 为线性空间 $C(a, b)$ 上的线性变换.

性质 1.7 设 σ 为线性空间 V 上的线性变换, 则 $\sigma(0) = 0, \sigma(-\alpha) = -\sigma(\alpha)$.

性质 1.8 设 σ 为线性空间 V 上的线性变换, 若 $\beta = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r$, 则

$$\sigma(\beta) = k_1\sigma(\alpha_1) + k_2\sigma(\alpha_2) + \cdots + k_r\sigma(\alpha_r).$$

性质 1.9 设 σ 为线性空间 V 上的线性变换, 若向量组 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 线性相关, 则向量组 $\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_r)$ 也线性相关.

注 1.9 性质 1.9 说明线性变换将线性相关向量组变成线性相关向量组,但是线性变换不一定将线性无关向量组变成线性无关向量组,例如零变换.

1.4.2 线性变换的运算

1. 线性变换的加法

定义 1.13 设 σ, τ 为线性空间 V 上的两个线性变换,定义它们的和 $\sigma + \tau$ 为

$$(\sigma + \tau)(\alpha) = \sigma(\alpha) + \tau(\alpha), \quad \forall \alpha \in V,$$

则 $\sigma + \tau$ 也是 V 的线性变换.

定义 1.14 设 σ 为线性空间 V 上的线性变换,定义变换 $-\sigma$ 为

$$(-\sigma)(\alpha) = -\sigma(\alpha), \quad \forall \alpha \in V,$$

则 $-\sigma$ 也是 V 上的线性变换,被称为 σ 的负变换.

性质 1.10 设 σ, τ, δ 为线性空间 V 上的三个线性变换,则线性变换的加法满足:

- (1) 交换律,即 $\sigma + \tau = \tau + \sigma$;
- (2) 结合律,即 $(\sigma + \tau) + \delta = \sigma + (\tau + \delta)$;
- (3) 具有零元,即 $0 + \sigma = \sigma + 0 = \sigma$,其中 0 为零变换;
- (4) 具有负元,即 $\sigma + (-\sigma) = (-\sigma) + \sigma = 0$.

2. 线性变换的数量乘法

定义 1.15 设 σ 为线性空间 V 上的线性变换, $k \in F$,定义 k 与 σ 的数量乘法 $k\sigma$ 为

$$(k\sigma)(\alpha) = k\sigma(\alpha), \quad \forall \alpha \in V,$$

则 $k\sigma$ 也是 V 的线性变换.

性质 1.11 设 σ, τ 为线性空间 V 上的两个线性变换, $k, l \in F$,则有

- (1) $(kl)\sigma = k(l\sigma)$;
- (2) $(k+l)\sigma = k\sigma + l\sigma$;
- (3) $k(\sigma + \tau) = k\sigma + k\tau$;
- (4) $1\sigma = \sigma$.

注 1.10 线性空间 V 上的全体线性变换所构成的集合,对于线性变换的加法与数量乘法构成数域 F 上的一个线性空间,记为 $L(V)$.

3. 线性变换的乘积

定义 1.16 设 σ, τ 为线性空间 V 上的两个线性变换,定义它们的乘积 $\sigma\tau$ 为

$$(\sigma\tau)(\alpha) = \sigma(\tau(\alpha)), \quad \forall \alpha \in V,$$

则 $\sigma\tau$ 也是 V 上的线性变换.

性质 1.12 设 σ, τ, δ 为线性空间 V 上的三个线性变换, E 为单位变换,则线性变换的乘积满足(1)结合律,即 $(\sigma\tau)\delta = \sigma(\tau\delta)$;

- (2) $E\sigma = \sigma E = \sigma$;
- (3) 一般地, $\sigma\tau \neq \tau\sigma$.

例 1.22 对于线性空间 $\mathbb{R}[x]$ 的两个线性变换

$$D(f(x)) = f'(x), \quad J(f(x)) = \int_0^x f(t) dt,$$

它们的乘积 $(DJ)(f(x)) = D\left(\int_0^x f(t) dt\right) = f(x)$, 即 $DJ = E$, 而 $(JD)(f(x)) = J(f'(x)) = \int_0^x f'(t) dt = f(x) - f(0)$, 所以 $DJ \neq JD$.

注 1.11 (1) 矩阵 A 的第 i 列是 $\sigma(\epsilon_i)$ 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 下的坐标, 它是唯一的, 故线性变换 σ 在一组基下的矩阵是唯一的.

(2) 单位变换在任意一组基下的矩阵皆为单位矩阵; 零变换在任意一组基下的矩阵皆为零矩阵; 数乘变换在任意一组基下的矩阵皆为数量矩阵.

例 1.24 设线性空间 F^3 的线性变换为 $\sigma(x_1, x_2, x_3) = (x_1, x_2, x_1 + x_2)$, 求 σ 在标准基 $\epsilon_1, \epsilon_2, \epsilon_3$ 下的矩阵.

解 因为

$$\begin{cases} \sigma(\epsilon_1) = \sigma(1, 0, 0) = (1, 0, 1) = \epsilon_1 + 0\epsilon_2 + \epsilon_3, \\ \sigma(\epsilon_2) = \sigma(0, 1, 0) = (0, 1, 1) = 0\epsilon_1 + \epsilon_2 + \epsilon_3, \\ \sigma(\epsilon_3) = \sigma(0, 0, 1) = (0, 0, 0) = 0\epsilon_1 + 0\epsilon_2 + 0\epsilon_3, \end{cases}$$

故 σ 在标准基 $\epsilon_1, \epsilon_2, \epsilon_3$ 下的矩阵为 $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$.

例 1.25 设 $\epsilon_1, \epsilon_2, \dots, \epsilon_m$ ($m < n$) 是 n 维线性空间 V 的 m 维子空间 W 的一组基, 把它扩充为 V 的一组基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$, 在线性空间 V 上定义线性变换 σ 为

$$\begin{cases} \sigma(\epsilon_i) = \epsilon_i, & i = 1, 2, \dots, m, \\ \sigma(\epsilon_i) = 0, & i = m + 1, \dots, n, \end{cases}$$

即 $\sigma(\epsilon_1, \epsilon_2, \dots, \epsilon_n) = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \text{diag}(1, \dots, 1, 0, \dots, 0)$, 其中 $\text{diag}(1, \dots, 1, 0, \dots, 0)$ 是主对角线上恰有 m 个 1, $n - m$ 个 0 的对角矩阵, 称这样的线性变换 σ 为线性空间 V 对子空间 W 的投影.

定理 1.10 设 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是数域 F 上线性空间 V 的一组基, 在这组基下, V 的每一个线性变换都与 $F^{n \times n}$ 中的唯一一个矩阵对应, 并且具有以下性质:

- (1) 线性变换的和对应于矩阵的和;
- (2) 线性变换的乘积对应于矩阵的乘积;
- (3) 线性变换的数量乘积对应于矩阵的数量乘积;
- (4) 可逆线性变换与可逆矩阵对应, 且逆变换对应于逆矩阵.

定理 1.11 设线性变换 σ 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 下的矩阵为 A , 向量 $\xi \in V$ 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 下的坐标为 $(x_1, x_2, \dots, x_n)^T$, 像 $\sigma(\xi)$ 在基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 下的坐标为 $(y_1, y_2, \dots, y_n)^T$, 则有

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

定理 1.12 设线性空间 V 上的线性变换 σ 在两组基 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ (I) 和 $\eta_1, \eta_2, \dots, \eta_n$ (II) 下的矩阵分别为 A 和 B , 且从基 (I) 到基 (II) 的过渡矩阵为 X , 则 $B = X^{-1}AX$.

例 1.26 设 ϵ_1, ϵ_2 为二维线性空间 V 的一组基, V 上的线性变换 σ 在这组基下的矩阵为 $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$, η_1, η_2 为 V 的另一组基, 且 $(\eta_1, \eta_2) = (\epsilon_1, \epsilon_2) \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.

求: (1) σ 在 η_1, η_2 下的矩阵 B ; (2) A^k .

解 (1) σ 在基 η_1, η_2 下的矩阵

$$B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

(2) 由于 $B = X^{-1}AX$, 有 $A = XBX^{-1}$, 于是 $A^k = X B^k X^{-1}$. 所以

$$A^k = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^k \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} k+1 & k \\ -k & -k+1 \end{pmatrix}.$$

例 1.27 设 $\eta_1 = (-1, 0, 2)$, $\eta_2 = (0, 1, 1)$, $\eta_3 = (3, -1, 0)$ 为线性空间 F^3 的一组基, 线性变换 σ 定义为

$$\begin{cases} \sigma(\eta_1) = (-5, 0, 3), \\ \sigma(\eta_2) = (0, -1, 6), \\ \sigma(\eta_3) = (-5, -1, 9), \end{cases}$$

求: (1) σ 在标准基 $\epsilon_1, \epsilon_2, \epsilon_3$ 下的矩阵; (2) σ 在基 η_1, η_2, η_3 下的矩阵.

解 (1) 由题可得

$$(\eta_1, \eta_2, \eta_3) = (\epsilon_1, \epsilon_2, \epsilon_3) \begin{pmatrix} -1 & 0 & 3 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix} \stackrel{\text{def}}{=} (\epsilon_1, \epsilon_2, \epsilon_3) X \quad (1.8)$$

以及

$$\sigma(\eta_1, \eta_2, \eta_3) = (\epsilon_1, \epsilon_2, \epsilon_3) \begin{pmatrix} -5 & 0 & -5 \\ 0 & -1 & -1 \\ 3 & 6 & 9 \end{pmatrix}. \quad (1.9)$$

设 σ 在标准基 $\epsilon_1, \epsilon_2, \epsilon_3$ 下的矩阵为 A , 即 $\sigma(\epsilon_1, \epsilon_2, \epsilon_3) = (\epsilon_1, \epsilon_2, \epsilon_3)A$, 故由(1.8)式得

$$\sigma(\eta_1, \eta_2, \eta_3) = \sigma((\epsilon_1, \epsilon_2, \epsilon_3)X) = \sigma(\epsilon_1, \epsilon_2, \epsilon_3)X = (\epsilon_1, \epsilon_2, \epsilon_3)AX. \quad (1.10)$$

对比(1.9)式与(1.10)式, 可得 $AX = \begin{pmatrix} -5 & 0 & -5 \\ 0 & -1 & -1 \\ 3 & 6 & 9 \end{pmatrix}$, 所以

$$A = \begin{pmatrix} -5 & 0 & -5 \\ 0 & -1 & -1 \\ 3 & 6 & 9 \end{pmatrix} X^{-1} = \begin{pmatrix} -5 & 0 & -5 \\ 0 & -1 & -1 \\ 3 & 6 & 9 \end{pmatrix} \begin{pmatrix} -1 & 0 & 3 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix}^{-1} = \frac{1}{7} \begin{pmatrix} -5 & 20 & -20 \\ -4 & -5 & -2 \\ 27 & 18 & 24 \end{pmatrix}.$$

(2) 设 σ 在基 η_1, η_2, η_3 下的矩阵为 B , 则 $B = X^{-1}AX$, 所以

$$B = \begin{pmatrix} -1 & 0 & 3 \\ 0 & 1 & -1 \\ 2 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} -5 & 0 & -5 \\ 0 & -1 & -1 \\ 3 & 6 & 9 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 5 \\ -1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix}.$$

习 题 1

1. 检验下列集合对于所指的线性运算是否构成实数域上的线性空间. 如果是, 给出证明; 如果不是, 说明理由.

(1) 全体 n 阶实对称矩阵组成的集合 V , 对于矩阵的加法和数量乘法;

(2) 平面上全体向量 \mathbb{R}^2 , 对于通常的向量加法, 数量乘法定义为 $k \circ \alpha = \mathbf{0}$;

(3) 平面上全体向量 \mathbb{R}^2 , 对于通常的向量加法, 数量乘法定义为 $k \circ \alpha = \alpha$.

2. 在线性空间 F^4 中, 求由基 $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ 到基 $\eta_1, \eta_2, \eta_3, \eta_4$ 的过渡矩阵, 并求向量 ξ 在所指基下的坐标.

$$(1) \begin{cases} \epsilon_1 = (1, 0, 0, 0), \\ \epsilon_2 = (0, 1, 0, 0), \\ \epsilon_3 = (0, 0, 1, 0), \\ \epsilon_4 = (0, 0, 0, 1), \end{cases} \begin{cases} \eta_1 = (2, 1, -1, 1), \\ \eta_2 = (0, 3, 1, 0), \\ \eta_3 = (5, 3, 2, 1), \\ \eta_4 = (6, 6, 1, 3), \end{cases}$$

$\xi = (1, 1, 0, 1)$ 在基 $\eta_1, \eta_2, \eta_3, \eta_4$ 下的坐标;

$$(2) \begin{cases} \epsilon_1 = (1, 1, 1, 1), \\ \epsilon_2 = (1, 1, -1, -1), \\ \epsilon_3 = (1, -1, 1, -1), \\ \epsilon_4 = (1, -1, -1, 1), \end{cases} \begin{cases} \eta_1 = (1, 1, 0, 1), \\ \eta_2 = (2, 1, 3, 1), \\ \eta_3 = (1, 1, 0, 0), \\ \eta_4 = (0, 1, -1, -1), \end{cases}$$

$\xi = (1, 0, 0, -1)$ 在基 $\eta_1, \eta_2, \eta_3, \eta_4$ 下的坐标.

3. 在线性空间 F^4 中, 求由齐次线性方程组

$$\begin{cases} 3x_1 + 2x_2 - 5x_3 + 4x_4 = 0, \\ 3x_1 - x_2 + 3x_3 - 3x_4 = 0, \\ 3x_1 + 5x_2 - 13x_3 + 11x_4 = 0 \end{cases}$$

确定的解空间的一组基与维数.

4. 设 $\alpha_1 = (2, 1, 3, 1), \alpha_2 = (1, 2, 0, 1), \alpha_3 = (-1, 1, -3, 0), \alpha_4 = (1, 1, 1, 1)$, 求由 $\alpha_i (i=1, 2, 3, 4)$ 生成的子空间的维数与一组基.

5. 设 $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, 集合 $W = \{B \in F^{3 \times 3} \mid AB = BA\}$, 证明 W 是 $F^{3 \times 3}$ 的一个子空间,

并求它的维数与一组基.

6. 检验下列变换是不是线性变换, 如果是, 给出证明; 如果不是, 说明理由.

(1) 在线性空间 \mathbb{R}^3 中, 变换 $\sigma(x_1, x_2, x_3) = (2x_1, x_2, x_2 - x_3)$.

(2) 在线性空间 $F[x]_n$ 中, 变换 $\sigma(f(x)) = f^2(x)$.

(3) 在线性空间 V 中, 变换 $\sigma(\xi) = \xi + \alpha$, 其中 $\alpha \in V$ 为非零固定变量.

(4) 在线性空间 $F^{n \times n}$ 中, 变换 $\sigma(X) = AX$, 其中 $A \in F^{n \times n}$ 为固定的 n 阶矩阵.

(5) 复数域 \mathbb{C} 看成自身上的线性空间, 变换 $\sigma(x) = \bar{x}$, 其中 \bar{x} 表示 x 的共轭复数.

(6) 复数域 \mathbb{C} 看成实数域 \mathbb{R} 上的线性空间, 变换 $\sigma(x) = \bar{x}$, 其中 \bar{x} 表示 x 的共轭复数.

7. 已知线性空间 F^3 的线性变换 σ 在基 $\eta_1 = (-1, 1, 1), \eta_2 = (1, 0, -1), \eta_3 = (0, 1, 1)$

下的矩阵为 $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}$, 求 σ 在基 $\epsilon_1 = (1, 0, 0), \epsilon_2 = (0, 1, 0), \epsilon_3 = (0, 0, 1)$ 下的矩阵.

8. 已知 $E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ 是线性空间 $F^{2 \times 2}$ 的

一组基, 设矩阵 $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, 在 $F^{2 \times 2}$ 中定义线性变换 $\sigma(\mathbf{X}) = \mathbf{A}\mathbf{X}$ 和 $\tau(\mathbf{X}) = \mathbf{X}\mathbf{A}$, 求线性变换 $\sigma + \tau$ 和 $\sigma\tau$ 在基 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 下的矩阵.

9. 给定线性空间 F^3 的两组基:

$$\boldsymbol{\epsilon}_1 = (1, 0, 1), \quad \boldsymbol{\epsilon}_2 = (2, 1, 0), \quad \boldsymbol{\epsilon}_3 = (1, 1, 1), \quad (\text{I})$$

$$\boldsymbol{\eta}_1 = (1, 2, -1), \quad \boldsymbol{\eta}_2 = (2, 2, -1), \quad \boldsymbol{\eta}_3 = (2, -1, -1). \quad (\text{II})$$

定义线性变换 σ 为 $\sigma(\boldsymbol{\epsilon}_i) = \boldsymbol{\eta}_i (i=1, 2, 3)$. 求: (1) 由基 $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3$ 到基 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3$ 的过渡矩阵; (2) σ 在基 $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3$ 下的矩阵; (3) σ 在基 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3$ 下的矩阵.

第 2 章

矩阵分析

矩阵的分析运算是矩阵的重要运算工具之一,在数学的许多分支(如统计学、微分几何)和工程实际(如阵列信号处理、通信系统、雷达、声呐)中有着广泛的应用.

本章首先介绍内积与范数的概念和理论,解决了两个向量或两个矩阵之间的距离问题,之后与高等数学类似地讨论矩阵序列的极限、矩阵级数、矩阵函数及其计算,最后讨论矩阵函数的微积分理论.

2.1 内积与范数

2.1.1 内积

定义 2.1 设 V 是数域 F 上的线性空间,对 V 中任意两个向量 α, β , 定义一个二元实函数,记作 (α, β) . 对于任意 $\alpha, \beta, \gamma \in V$ 及任意的数 $k \in F$, 若 (α, β) 满足

- (1) 对称性, 即 $(\alpha, \beta) = (\beta, \alpha)$;
- (2) 齐次性, 即 $(k\alpha, \beta) = k(\alpha, \beta)$;
- (3) 可加性, 即 $(\alpha + \beta, \gamma) = (\alpha, \gamma) + (\beta, \gamma)$;
- (4) 正定性, 即 $(\alpha, \alpha) \geq 0$, 当且仅当 $\alpha = 0$ 时 $(\alpha, \alpha) = 0$, 则 (α, β) 称为 α 与 β 的内积.

例 2.1 在线性空间 \mathbb{R}^n 中, 对于向量 $\alpha = (a_1, a_2, \dots, a_n)$, $\beta = (b_1, b_2, \dots, b_n)$, 定义 $(\alpha, \beta) = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$ 和 $(\alpha, \beta)' = a_1 b_1 + 2a_2 b_2 + \dots + k a_k b_k + \dots + n a_n b_n$, (α, β) 和 $(\alpha, \beta)'$ 都满足内积定义 2.1 中的 4 条性质, 因此 (α, β) 和 $(\alpha, \beta)'$ 都是内积. 其中, (α, β) 是向量空间经典的内积, 也记为 $\alpha \cdot \beta$.

例 2.2 记 $C(a, b)$ 为闭区间 $[a, b]$ 上全体实连续函数构成的 \mathbb{R} 上的线性空间, 对于函数 $f(x), g(x) \in C(a, b)$, 定义 $(f, g) = \int_a^b f(x)g(x)dx$, 证明 (f, g) 是内积.

证明 对于任意函数 $f(x), g(x), h(x) \in C(a, b)$ 及任意的 $k \in \mathbb{R}$, 由 (f, g) 的定义以及定积分性质可知,

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

$$(kf, g) = \int_a^b kf(x)g(x)dx = k \int_a^b f(x)g(x)dx = k(f, g),$$

$$\begin{aligned}(f + g, h) &= \int_a^b [f(x) + g(x)]h(x)dx = \int_a^b f(x)h(x)dx + \int_a^b g(x)h(x)dx \\ &= (f, h) + (g, h),\end{aligned}$$

$$(f, f) = \int_a^b f^2(x) dx \geq 0, (f, f) = \int_a^b f^2(x) dx = 0 \text{ 当且仅当 } f(x) = 0.$$

故 (f, g) 是内积. □

2.1.2 向量范数

定义 2.2 若对任意向量 $\mathbf{x} \in \mathbb{C}^n$ 都有一个实数 $\|\mathbf{x}\|$ 与之对应, 且满足:

- (1) 非负性, 即当 $\mathbf{x} \neq \mathbf{0}$ 时, $\|\mathbf{x}\| > 0$, 当 $\mathbf{x} = \mathbf{0}$ 时, $\|\mathbf{x}\| = 0$;
- (2) 齐次性, 即对任意数 $\lambda \in \mathbb{C}$, $\|\lambda\mathbf{x}\| = |\lambda| \|\mathbf{x}\|$;
- (3) 三角不等式, 即对任意 $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ 都有 $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, 则称 $\|\mathbf{x}\|$ 为 \mathbb{C}^n

上向量 \mathbf{x} 的范数, 简称向量范数.

定义 2.3 设 $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{C}^n$, 定义向量 \mathbf{x} 的如下范数:

- (1) 向量 1 范数: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$;
- (2) 向量 2 范数(或向量长度): $\|\mathbf{x}\|_2 = \sqrt{(\mathbf{x}, \mathbf{x})} = \sqrt{\sum_{i=1}^n |x_i|^2}$;
- (3) 向量 ∞ 范数: $\|\mathbf{x}\|_\infty = \max_i |x_i|$;
- (4) 向量 p 范数: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \quad (1 \leq p < +\infty)$.

注 2.1 当 $p=1$ 和 $p=2$ 时, 向量 p 范数就是向量 1 范数和 2 范数; 当 $p \rightarrow \infty$ 时, 向量 p 范数就是向量 ∞ 范数.

例 2.3 求向量 $\mathbf{x} = (2i, 0, 4+i, -1, 0)$ 的 1, 2 和 ∞ 范数.

解 按定义, $\|\mathbf{x}\|_1 = |2i| + |0| + |4+i| + |-1| + |0| = 2 + \sqrt{17} + 1 = 3 + \sqrt{17}$;

$$\|\mathbf{x}\|_2 = \sqrt{|2i|^2 + |0|^2 + |4+i|^2 + |-1|^2 + |0|^2} = \sqrt{4+17+1} = \sqrt{22};$$

$$\|\mathbf{x}\|_\infty = \max\{|2i|, |0|, |4+i|, |-1|, |0|\}$$

$$= \max\{2, 0, \sqrt{17}, 1, 0\} = \sqrt{17}.$$

2.1.3 矩阵范数

矩阵范数为矩阵的“大小”或“重要性”提供一个严格、可计算的量化标准, 解决了矩阵作为多维数组无法直接比较“大小”的问题.

定义 2.4 设 $\mathbf{A} \in \mathbb{C}^{m \times n}$ 都有一个实数 $\|\mathbf{A}\|$ 与之对应, 且满足:

- (1) 非负性, 即当 $\mathbf{A} \neq \mathbf{0}$ 时, $\|\mathbf{A}\| > 0$, 当 $\mathbf{A} = \mathbf{0}$ 时, $\|\mathbf{A}\| = 0$;
- (2) 齐次性, 即对任意数 $\lambda \in \mathbb{C}$, $\|\lambda\mathbf{A}\| = |\lambda| \|\mathbf{A}\|$;
- (3) 三角不等式, 即对任意 $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times n}$ 都有 $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$;
- (4) 相容性, 即当矩阵乘积有意义时, 有 $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$, 则称 $\|\mathbf{A}\|$ 为矩阵 \mathbf{A}

的范数, 简称矩阵范数.

以下列出常用的 7 种矩阵范数.

定义 2.5 对于任意矩阵 $\mathbf{A} = (a_{ij})_{m \times n} \in \mathbb{C}^{m \times n}$, 定义矩阵 \mathbf{A} 的如下范数:

(1) 矩阵 m_1 范数: $\|A\|_{m_1} = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$;

(2) 矩阵 F 范数: $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$;

(3) 矩阵 M 范数(或最大范数): $\|A\|_M = \max\{m, n\} \cdot \max_{i,j} |a_{ij}|$;

(4) 矩阵 G 范数(或几何平均范数): $\|A\|_G = \sqrt{mn} \max_{i,j} |a_{ij}|$;

(5) 矩阵 1 范数(或列和范数): $\|A\|_1 = \max_j \sum_{i=1}^m |a_{ij}|$;

(6) 矩阵 ∞ 范数(或行和范数): $\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$;

(7) 矩阵 2 范数(或谱范数): $\|A\|_2 = \sqrt{\mathbf{A}^H \mathbf{A}}$ 的最大特征值.

注 2.2 对于任意矩阵 $A = (a_{ij})_{m \times n} \in \mathbb{C}^{m \times n}$ 和 $B = (b_{ij})_{n \times s} \in \mathbb{C}^{n \times s}$, 记 $AB = (c_{ij})_{m \times s}$, 其中 $c_{ij} = \sum_{t=1}^n a_{it} b_{tj}$, 由三角不等式可知 $|c_{ij}| = \left| \sum_{t=1}^n a_{it} b_{tj} \right| \leq \sum_{t=1}^n |a_{it}| \cdot |b_{tj}|$, 因此, 有

$$\begin{aligned} \|AB\|_{m_1} &= \sum_{i=1}^m \sum_{j=1}^s |c_{ij}| \leq \sum_{i=1}^m \sum_{j=1}^s \sum_{t=1}^n |a_{it}| \cdot |b_{tj}| \leq \left(\sum_{i=1}^m \sum_{t=1}^n |a_{it}| \right) \left(\sum_{t=1}^n \sum_{j=1}^s |b_{tj}| \right) \\ &= \|A\|_{m_1} \|B\|_{m_1}, \end{aligned}$$

$$\begin{aligned} \|AB\|_M &= \max\{m, s\} \cdot \max_{i,j} |c_{ij}| \leq \max\{m, s\} \cdot \max_{i,j} \sum_{t=1}^n |a_{it}| \cdot |b_{tj}| \\ &= \max\{m, s\} \cdot \sum_{t=1}^n \max_{i,j} (|a_{it}| \cdot |b_{tj}|) \\ &\leq \max\{m, s\} \cdot n \cdot \max_{i,j} |a_{ij}| \cdot \max_{i,j} |b_{ij}| \\ &\leq (\max\{m, n\} \cdot \max_{i,j} |a_{ij}|) (\max\{n, s\} \cdot \max_{i,j} |b_{ij}|) \\ &= \|A\|_M \|B\|_M, \end{aligned}$$

$$\begin{aligned} \|AB\|_G &= \sqrt{ms} \cdot \max_{i,j} |c_{ij}| \leq \sqrt{ms} \cdot \max_{i,j} \sum_{t=1}^n |a_{it}| \cdot |b_{tj}| \\ &\leq \sqrt{ms} \cdot n \cdot \max_{i,j} |a_{ij}| \cdot \max_{i,j} |b_{ij}| \\ &= (\sqrt{mn} \max_{i,j} |a_{ij}|) (\sqrt{ns} \max_{i,j} |b_{ij}|) = \|A\|_G \|B\|_G. \end{aligned}$$

例 2.4 求 $A = \begin{pmatrix} 1 & 1+i & 4 \\ 4i & 5 & 2 \\ 0 & 1-3i & 2-2i \end{pmatrix}$ 的 m_1, F, M, G , 以及列和与行和范数.

解 $\|A\|_{m_1} = |1| + |1+i| + |4| + |4i| + |5| + |2| + |0| + |1-3i| + |2-2i|$
 $= 16 + 3\sqrt{2} + \sqrt{10}$;

$$\|A\|_F = \sqrt{|1|^2 + |1+i|^2 + |4|^2 + |4i|^2 + |5|^2 + |2|^2 + |0|^2 + |1-3i|^2 + |2-2i|^2} = \sqrt{82};$$

$$\|A\|_M = \max\{3, 3\} \times \max\{|1|, |1+i|, |4|, |4i|, |5|, |2|, |0|, |1-3i|, |2-2i|\} \\ = 3 \times 5 = 15;$$

$$\|A\|_G = \sqrt{3 \times 3} \cdot \max\{|1|, |1+i|, |4|, |4i|, |5|, |2|, |0|, |1-3i|, |2-2i|\} \\ = 3 \times 5 = 15;$$

$$\text{列和范数 } \|A\|_1 = \max\{|1| + |4i| + |0|, |1+i| + |5| + |1-3i|, |4| + |2| + |2-2i|\} \\ = \max\{5, 5 + \sqrt{2} + \sqrt{10}, 6 + 2\sqrt{2}\} = 5 + \sqrt{2} + \sqrt{10};$$

$$\text{行和范数 } \|A\|_\infty = \max\{|1| + |1+i| + |4|, |4i| + |5| + |2|, |0| + |1-3i| + |2-2i|\} \\ = \max\{5 + \sqrt{2}, 11, \sqrt{10} + 2\sqrt{2}\} = 11.$$

本小节最后给出矩阵范数的等价性定义及定理, 定理留给读者自证.

定义 2.6 设 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 是任意两种矩阵范数, 若存在常数 C_1 和 C_2 , 使得对任意 $m \times n$ 矩阵 A 都有 $C_1 \|A\|_\alpha \leq \|A\|_\beta \leq C_2 \|A\|_\alpha$, 则称矩阵范数 $\|\cdot\|_\alpha$ 和 $\|\cdot\|_\beta$ 范数等价.

定理 2.1 任意两种矩阵范数都等价.

2.1.4 相似性度量

聚类和分类是数据分析的重要技术. 聚类是把大数据集聚为 N 类子集, 并且每个子集 (目标类) 的数据都具有共同或者相似的特征. 分类则是将一个数据映射到某个已知目标类别中. 相似性度量是聚类与分类算法中一个很重要的数学工具.

非概率相关的相似性度量是不依赖概率分布、仅通过数据本身的数值特征或结构特征直接计算相似度的方法. 常见的非概率相关的相似性度量主要有两类: 基于向量距离的度量和基于数据关联的度量.

下面主要讨论基于向量距离 (通过向量范数定义) 的度量. 值得注意的是, 向量范数是构建向量距离的核心工具, 但并非所有的向量距离都必须由范数生成. 此外注意向量距离与向量长度的区别: 向量长度是单个向量自身的属性, 而向量距离是两个向量之间的属性. 当其中一个向量为零向量时, 两个向量的欧氏距离就等于另一个向量的长度.

定义 2.7 设 $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, 定义向量 x 与 y 的

$$(1) \text{ 欧氏距离为 } \text{dist}(x, y) = \|x - y\|_2 = \sqrt{\sum_{i=1}^n |x_i - y_i|^2};$$

$$(2) \text{ 曼哈顿距离为 } \text{dist}(x, y) = \|x - y\|_1 = \sum_{i=1}^n |x_i - y_i|;$$

$$(3) \text{ 切比雪夫距离为 } \text{dist}(x, y) = \|x - y\|_\infty = \max_i |x_i - y_i|;$$

(4) 闵可夫斯基距离为

$$\text{dist}(x, y) = \|x - y\|_p = \left(\sum_{i=1}^n |x_i - y_i|^p\right)^{1/p}, \quad 1 \leq p < +\infty.$$

定义 2.8 设 $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, 向量 x 与 y 之间夹角的余弦值称为 x 与 y 的余弦相似度, 记为 $\text{sim}_{\cos}(x, y)$, 即

$$\text{sim}_{\cos}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}.$$

例 2.5 已知两个样本特征向量为 $\mathbf{x} = (0, -1, 2, 0, 2)$, $\mathbf{y} = (-1, -3, 5, 0, 7)$, 求 \mathbf{x} 与 \mathbf{y} 的欧氏距离、曼哈顿距离、切比雪夫距离以及余弦相似度.

解 因为 $\mathbf{x} - \mathbf{y} = (1, 2, -3, 0, -5)$, 所以 \mathbf{x} 与 \mathbf{y} 的欧氏距离为

$$\|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} = \sqrt{1^2 + 2^2 + (-3)^2 + 0^2 + (-5)^2} = \sqrt{39};$$

曼哈顿距离为 $\|\mathbf{x} - \mathbf{y}\|_1 = \sum_{i=1}^n |x_i - y_i| = |1| + |2| + |-3| + |0| + |-5| = 11$;

切比雪夫距离为 $\|\mathbf{x} - \mathbf{y}\|_{\infty} = \max_i |x_i - y_i| = \max\{|1|, |2|, |-3|, |0|, |-5|\} = 5$;

余弦相似度为

$$\begin{aligned} \text{sim}_{\cos}(\mathbf{x}, \mathbf{y}) &= \frac{0 \times (-1) + (-1) \times (-3) + 2 \times 5 + 0 \times 0 + 2 \times 7}{\sqrt{[0^2 + (-1)^2 + 2^2 + 0^2 + 2^2]} \times \sqrt{[(-1)^2 + (-3)^2 + 5^2 + 0^2 + 7^2]}} \\ &= \frac{3\sqrt{21}}{14}. \end{aligned}$$

2.2 矩阵序列

2.2.1 矩阵序列的极限

定义 2.9 设有 $m \times n$ 矩阵序列 $\{\mathbf{A}^{(k)}\}_{k=0}^{\infty}$, 其中 $\mathbf{A}^{(k)} = (a_{ij}^{(k)})_{m \times n}$. 若 $m \times n$ 个序列 $\{a_{ij}^{(k)}\}$ 都收敛, 即 $\lim_{k \rightarrow \infty} a_{ij}^{(k)} = a_{ij}$ ($i=1, 2, \dots, m; j=1, 2, \dots, n$), 则称矩阵序列 $\{\mathbf{A}^{(k)}\}$ 收敛于矩阵 $\mathbf{A} = (a_{ij})_{m \times n}$, \mathbf{A} 称为矩阵序列 $\{\mathbf{A}^{(k)}\}$ 的极限, 记作 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{A}$ 或 $\mathbf{A}^{(k)} \rightarrow \mathbf{A}$ ($k \rightarrow \infty$). 若矩阵序列 $\{\mathbf{A}^{(k)}\}$ 不收敛, 则称 $\{\mathbf{A}^{(k)}\}$ 发散.

例 2.6 设矩阵序列 $\mathbf{A}^{(k)} = \begin{pmatrix} \left(1 + \frac{1}{1+k}\right)^{1+k} & \arctan k \\ 1 & \frac{5}{1+k} \end{pmatrix}$, 求 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)}$.

解 记 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{A}$, 则 $\mathbf{A} = \begin{pmatrix} \lim_{k \rightarrow \infty} \left(1 + \frac{1}{1+k}\right)^{1+k} & \lim_{k \rightarrow \infty} \arctan k \\ 1 & \lim_{k \rightarrow \infty} \frac{5}{1+k} \end{pmatrix} = \begin{pmatrix} e & \frac{\pi}{2} \\ 1 & 0 \end{pmatrix}$.

注 2.3 从定义 2.9 可知, 一个 $m \times n$ 矩阵序列收敛相当于 $m \times n$ 个序列同时收敛. 当数值 $m \times n$ 较大时, 讨论 $m \times n$ 个序列的收敛性是较为烦琐的, 下面的定理是用矩阵范数来研究矩阵序列的收敛性.

定理 2.2 设有 $m \times n$ 矩阵序列 $\{\mathbf{A}^{(k)}\}_{k=0}^{\infty}$ 和矩阵 \mathbf{A} , 则 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{A}$ 的充分必要条件为

$$\lim_{k \rightarrow \infty} \| \mathbf{A}^{(k)} - \mathbf{A} \| = 0,$$

其中 $\| \cdot \|$ 是任意一种矩阵范数.

证明 由定理 2.1, 即矩阵范数的等价性, 只需证明对矩阵的 G 范数结论成立即可. 因为

$$| a_{ij}^{(k)} - a_{ij} | \leq \sqrt{mn} \max_{i,j} | a_{ij}^{(k)} - a_{ij} | = \| \mathbf{A}^{(k)} - \mathbf{A} \|_G \leq \sqrt{mn} \sum_{i=1}^m \sum_{j=1}^n | a_{ij}^{(k)} - a_{ij} |,$$

所以 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{A}$ 的充分必要条件为 $\lim_{k \rightarrow \infty} \| \mathbf{A}^{(k)} - \mathbf{A} \|_G = 0$. \square

推论 设有 $m \times n$ 矩阵序列 $\{ \mathbf{A}^{(k)} \}_{k=0}^{\infty}$ 和矩阵 \mathbf{A} , 且 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{A}$, 则

$$\lim_{k \rightarrow \infty} \| \mathbf{A}^{(k)} \| = \| \mathbf{A} \|,$$

其中 $\| \cdot \|$ 是任意一种矩阵范数.

注 2.4 此推论的逆命题不成立. 例如矩阵序列 $\mathbf{A}^{(k)} = \begin{pmatrix} (-1)^k & 2 \\ 0 & 1/(1+k) \end{pmatrix}$ 不收敛, 但

$$\lim_{k \rightarrow \infty} \| \mathbf{A}^{(k)} \|_{m_1} = \lim_{k \rightarrow \infty} \sum_{i=1}^m \sum_{j=1}^n | a_{ij}^{(k)} | = 3.$$

矩阵序列收敛的性质, 有许多与数列收敛的性质相类似.

定理 2.3 设 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{A}$, $\lim_{k \rightarrow \infty} \mathbf{B}^{(k)} = \mathbf{B}$, 其中 $\mathbf{A}^{(k)}, \mathbf{B}^{(k)}, \mathbf{A}, \mathbf{B}$ 为适当的矩阵, α, β 为数, 则

- (1) $\lim_{k \rightarrow \infty} (\alpha \mathbf{A}^{(k)} + \beta \mathbf{B}^{(k)}) = \alpha \mathbf{A} + \beta \mathbf{B}$;
- (2) $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} \mathbf{B}^{(k)} = \mathbf{A} \mathbf{B}$;
- (3) 当 $\mathbf{A}^{(k)}$ 和 \mathbf{A} 均可逆时, $\lim_{k \rightarrow \infty} (\mathbf{A}^{(k)})^{-1} = \mathbf{A}^{-1}$.

证明 对任一矩阵范数 $\| \cdot \|$.

(1) 当 $k \rightarrow \infty$ 时, 有

$$\| (\alpha \mathbf{A}^{(k)} + \beta \mathbf{B}^{(k)}) - (\alpha \mathbf{A} + \beta \mathbf{B}) \| \leq |\alpha| \| \mathbf{A}^{(k)} - \mathbf{A} \| + |\beta| \| \mathbf{B}^{(k)} - \mathbf{B} \| \rightarrow 0,$$

因此 $\lim_{k \rightarrow \infty} (\alpha \mathbf{A}^{(k)} + \beta \mathbf{B}^{(k)}) = \alpha \mathbf{A} + \beta \mathbf{B}$.

(2) 当 $k \rightarrow \infty$ 时, 依据矩阵范数满足的三角不等式与相容性, 有

$$\begin{aligned} \| \mathbf{A}^{(k)} \mathbf{B}^{(k)} - \mathbf{A} \mathbf{B} \| &= \| \mathbf{A}^{(k)} \mathbf{B}^{(k)} - \mathbf{A}^{(k)} \mathbf{B} + \mathbf{A}^{(k)} \mathbf{B} - \mathbf{A} \mathbf{B} \| \\ &\leq \| \mathbf{A}^{(k)} \| \| \mathbf{B}^{(k)} - \mathbf{B} \| + \| \mathbf{A}^{(k)} - \mathbf{A} \| \| \mathbf{B} \| \rightarrow 0, \end{aligned}$$

因此 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} \mathbf{B}^{(k)} = \mathbf{A} \mathbf{B}$.

(3) 记 $|\mathbf{A}|$ 为矩阵 \mathbf{A} 的行列式, \mathbf{A}^* 为矩阵 \mathbf{A} 的伴随矩阵. 由伴随矩阵定义知 $(\mathbf{A}^{(k)})^* = (\mathbf{A}_{ji}^{(k)})$, 其中 $\mathbf{A}_{ji}^{(k)}$ 是矩阵 $\mathbf{A}^{(k)}$ 中元素 $a_{ij}^{(k)}$ 的代数余子式. 由数列极限的四则运算法则得 $\lim_{k \rightarrow \infty} \mathbf{A}_{ij}^{(k)} = \mathbf{A}_{ij}$, 其中 \mathbf{A}_{ij} 是矩阵 \mathbf{A} 中元素 a_{ij} 的代数余子式, 故 $\lim_{k \rightarrow \infty} (\mathbf{A}^{(k)})^* = \mathbf{A}^*$. 同理

$$\lim_{k \rightarrow \infty} |\mathbf{A}^{(k)}| = |\mathbf{A}|, \mathbf{A}^{(k)} \text{ 和 } \mathbf{A} \text{ 可逆, 故 } \lim_{k \rightarrow \infty} (\mathbf{A}^{(k)})^{-1} = \lim_{k \rightarrow \infty} \frac{(\mathbf{A}^{(k)})^*}{|\mathbf{A}^{(k)}|} = \frac{\mathbf{A}^*}{|\mathbf{A}|} = \mathbf{A}^{-1}. \quad \square$$

注 2.5 定理 2.3 中的结论(3)中条件“ $\mathbf{A}^{(k)}$ 和 \mathbf{A} 均可逆”是必不可少的, 即使所有的 $\mathbf{A}^{(k)}$ 都可逆也不能保证 \mathbf{A} 一定可逆. 例如, 矩阵序列 $\mathbf{A}^{(k)} = \begin{pmatrix} 1 & 1 \\ 0 & 1/(1+k) \end{pmatrix}$ 满足所有的 $\mathbf{A}^{(k)}$ 都可

逆, 即 $(\mathbf{A}^{(k)})^{-1} = \begin{pmatrix} 1 & -(1+k) \\ 0 & (1+k) \end{pmatrix} (k=0, 1, 2, \dots)$, 但矩阵 $\mathbf{A} = \lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ 是不可逆的.

2.2.2 收敛矩阵

在矩阵序列中, 我们常考虑的是由一个方阵的幂所构成的序列. 对于这样的矩阵序列, 有以下的概念、收敛定理及其推论.

定义 2.10 设 \mathbf{A} 为 n 阶方阵, 如果 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{0}$, 那么称 \mathbf{A} 为收敛矩阵.

定理 2.4 设 \mathbf{A} 为 n 阶方阵, 则 \mathbf{A} 为收敛矩阵的充分必要条件为 $\rho(\mathbf{A}) < 1$, 其中 $\rho(\mathbf{A})$ 为矩阵 \mathbf{A} 特征值的模的最大值, 称为矩阵 \mathbf{A} 的谱半径.

性质 2.1 设 \mathbf{A} 为 n 阶方阵, 则对于任一矩阵范数 $\|\cdot\|$ 都有 $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$.

推论 设 \mathbf{A} 为 n 阶方阵, 若对于某一矩阵范数 $\|\cdot\|$ 有 $\|\mathbf{A}\| < 1$, 则 \mathbf{A} 为收敛矩阵.

例 2.7 判断下列矩阵是否为收敛矩阵:

$$(1) \mathbf{A} = \begin{pmatrix} 1 & 1/2 \\ -1/2 & 0 \end{pmatrix}; \quad (2) \mathbf{A} = \begin{pmatrix} 0.5 & 0.6 \\ 0.3 & 0.1 \end{pmatrix}.$$

解 (1) 经计算可得矩阵 \mathbf{A} 的特征值为 $\lambda_1 = \lambda_2 = 1/2$, 因此 $\rho(\mathbf{A}) = 1/2 < 1$, 故矩阵 \mathbf{A} 为收敛矩阵.

(2) 因为矩阵 \mathbf{A} 的列和范数 $\|\mathbf{A}\|_1 = 0.8 < 1$, 所以矩阵 \mathbf{A} 为收敛矩阵.

2.3 矩阵级数

在建立矩阵分析的理论时, 特别着重讨论矩阵级数, 尤其是矩阵的幂级数, 因为它是建立矩阵函数的依据.

2.3.1 矩阵级数的敛散性

定义 2.11 设由 $m \times n$ 矩阵序列 $\{\mathbf{A}^{(k)}\}_{k=0}^{\infty}$ 构成的无穷和 $\mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \dots + \mathbf{A}^{(k)} + \dots$ 称为矩阵级数, 记为 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$, 即 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)} = \mathbf{A}^{(0)} + \mathbf{A}^{(1)} + \dots + \mathbf{A}^{(k)} + \dots$. 对于任一正整数 N ,

称 $\mathbf{S}^{(N)} = \sum_{k=0}^N \mathbf{A}^{(k)}$ 为矩阵级数的部分和. 若矩阵序列 $\{\mathbf{S}^{(N)}\}$ 收敛且 $\lim_{N \rightarrow \infty} \mathbf{S}^{(N)} = \mathbf{S}$, 则称矩阵

级数 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$ 收敛且有和 \mathbf{S} , 记 $\mathbf{S} = \sum_{k=0}^{\infty} \mathbf{A}^{(k)}$. 不收敛的矩阵级数称为发散的.

注 2.6 (1) 若记 $\mathbf{A}^{(k)} = (a_{ij}^{(k)})_{m \times n}$, $\mathbf{S} = (s_{ij})_{m \times n}$, 则 $\mathbf{S} = \sum_{k=0}^{\infty} \mathbf{A}^{(k)}$ 相当于

$$\sum_{k=0}^{\infty} a_{ij}^{(k)} = s_{ij}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n,$$

即 $m \times n$ 个数项级数都收敛.

(2) 若 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$ 收敛, 则 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)} = \mathbf{0}$.

例 2.8 设有矩阵序列 $\{\mathbf{A}^{(k)}\}_{k=0}^{\infty}$, 其中 $\mathbf{A}^{(k)} = \begin{pmatrix} 0 & 1/(1+k)(3+k) \\ 1/(-2)^k & 0 \end{pmatrix}$, 讨论矩

矩阵级数 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$ 的敛散性.

解 计算可得

$$\begin{aligned} \mathbf{S}^{(N)} &= \sum_{k=0}^N \mathbf{A}^{(k)} = \begin{pmatrix} 0 & \sum_{k=0}^N \frac{1}{(1+k)(3+k)} \\ \sum_{k=0}^N \frac{1}{(-2)^k} & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \frac{3}{4} - \frac{1}{2} \left(\frac{1}{N+2} + \frac{1}{N+3} \right) \\ \frac{2}{3} \left[1 - \frac{1}{(-2)^{N+1}} \right] & 0 \end{pmatrix}, \end{aligned}$$

因此 $\mathbf{S} = \lim_{N \rightarrow \infty} \mathbf{S}^{(N)} = \begin{pmatrix} 0 & 3/4 \\ 2/3 & 0 \end{pmatrix}$, 矩阵级数 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$ 收敛, 且其和为 \mathbf{S} .

利用矩阵级数收敛的定义与高等数学中级数的性质, 可以得到以下矩阵级数类似的性质.

定理 2.5 设有收敛的矩阵级数 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)} = \mathbf{A}$, $\sum_{k=0}^{\infty} \mathbf{B}^{(k)} = \mathbf{B}$, 其中 $\mathbf{A}^{(k)}, \mathbf{B}^{(k)}, \mathbf{A}, \mathbf{B}, \mathbf{P}, \mathbf{Q}$

为适当的矩阵, λ 为数, 则: (1) $\sum_{k=0}^{\infty} (\mathbf{A}^{(k)} \pm \mathbf{B}^{(k)}) = \mathbf{A} \pm \mathbf{B}$; (2) $\sum_{k=0}^{\infty} \lambda \mathbf{A}^{(k)} = \lambda \mathbf{A}$; (3) 矩阵级数

$\sum_{k=0}^{\infty} \mathbf{P} \mathbf{A}^{(k)} \mathbf{Q}$ 收敛且其和为 $\mathbf{P} \mathbf{A} \mathbf{Q}$.

证明 只证(3). 若 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)} = \mathbf{A}$, 记 $\mathbf{S}^{(N)} = \sum_{k=0}^N \mathbf{A}^{(k)}$, 故 $\lim_{N \rightarrow \infty} \mathbf{S}^{(N)} = \mathbf{A}$. 从而

$$\lim_{N \rightarrow \infty} \sum_{k=0}^N \mathbf{P} \mathbf{A}^{(k)} \mathbf{Q} = \mathbf{P} \left(\lim_{N \rightarrow \infty} \sum_{k=0}^N \mathbf{A}^{(k)} \right) \mathbf{Q} = \mathbf{P} \mathbf{A} \mathbf{Q}. \quad \square$$

定义 2.12 设矩阵级数 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$, 其中 $\mathbf{A}^{(k)} = (a_{ij}^{(k)})_{m \times n}$. 若 $m \times n$ 个数项级数 $\sum_{k=0}^{\infty} a_{ij}^{(k)}$

($i=1, 2, \dots, m; j=1, 2, \dots, n$) 都绝对收敛, 即 $\sum_{k=0}^{\infty} |a_{ij}^{(k)}|$ 都收敛, 则称矩阵级数 $\sum_{k=0}^{\infty} \mathbf{A}^{(k)}$ 绝对收敛.

注 2.7 利用数项级数绝对收敛性质可知, 绝对收敛的矩阵级数必定收敛, 并且调换其项的顺序所得的矩阵级数依然收敛, 并且其和不变.

2.3.2 矩阵幂级数

定义 2.13 设 \mathbf{A} 为 n 阶方阵, $a_k \in \mathbb{C} (k=0, 1, 2, \dots)$, 称矩阵级数

$$\sum_{k=0}^{\infty} a_k \mathbf{A}^k = a_0 \mathbf{E} + a_1 \mathbf{A} + a_2 \mathbf{A}^2 + \dots + a_k \mathbf{A}^k + \dots$$

为矩阵 \mathbf{A} 的幂级数.

因为数项级数 $\sum_{k=0}^{\infty} \|a_k \mathbf{A}^k\|$ 的每一项都不大于数项级数 $\sum_{k=0}^{\infty} |a_k| \|\mathbf{A}\|^k$ 的对应项, 因

此有如下定理.

定理 2.6 若正项级数 $\sum_{k=0}^{\infty} |a_k| \|\mathbf{A}\|^k$ 收敛, 则矩阵幂级数 $\sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 绝对收敛, 其中 $\|\cdot\|$ 是任意一种矩阵范数.

推论 若矩阵 \mathbf{A} 的某一种范数在幂级数 $\sum_{k=0}^{\infty} a_k z^k = a_0 + a_1 z + a_2 z^2 + \cdots + a_k z^k + \cdots$ 的收敛域内, 则矩阵幂级数 $\sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 绝对收敛.

例 2.9 设 $\mathbf{A} = \begin{pmatrix} 0.1 & 0.3 & 0.5 \\ 0.2 & 0.2 & 0.1 \\ 0.4 & 0.1 & 0.4 \end{pmatrix}$, 证明: 矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k = \mathbf{E} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^k + \cdots$

绝对收敛.

证明 因为幂级数 $1+z+z^2+\cdots+z^k+\cdots$ 的收敛半径为 1, 矩阵的行和范数 $\|\mathbf{A}\|_{\infty} = \max_i \left\{ \sum_{j=1}^n |a_{ij}| \right\} = 0.9 < 1$, 故矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 绝对收敛. \square

定理 2.7 设 \mathbf{A} 为 n 阶方阵, 幂级数 $\sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 r , 则:

(1) 当 $\rho(\mathbf{A}) < r$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 绝对收敛;

(2) 当 $\rho(\mathbf{A}) > r$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 发散.

例 2.10 判断矩阵幂级数 $\sum_{k=0}^{\infty} \frac{k}{6^k} \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix}^k$ 的敛散性.

解 令 $\mathbf{A} = \frac{1}{6} \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix}$, 可以算得 $\rho(\mathbf{A}) = \frac{5}{6}$. 而幂级数 $\sum_{k=0}^{\infty} k z^k$ 的收敛半径为 $r = 1$,

故由 $\rho(\mathbf{A}) < r$ 知矩阵幂级数 $\sum_{k=0}^{\infty} k \mathbf{A}^k$ 绝对收敛.

最后考虑一个特殊的矩阵幂级数.

定理 2.8 设 \mathbf{A} 为 n 阶方阵, 矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 绝对收敛的充分必要条件是 $\rho(\mathbf{A}) < 1$, 且该级数的和为 $(\mathbf{E} - \mathbf{A})^{-1}$.

证明 因为幂级数 $\sum_{k=0}^{\infty} z^k$ 的收敛半径为 1, 当 $\rho(\mathbf{A}) < 1$ 时, 由定理 2.7 知矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 绝对收敛. 反之, 若矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 绝对收敛, 则 $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{0}$, 即 \mathbf{A} 为收敛矩阵, 由定理 2.4 知 $\rho(\mathbf{A}) < 1$.

最后计算矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 的和. 当 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 绝对收敛时, $\rho(\mathbf{A}) < 1$, 因此 $\mathbf{E} - \mathbf{A}$ 可逆. 记

$\mathbf{S}^{(N)} = \sum_{k=0}^N \mathbf{A}^k$, 则 $\mathbf{S}^{(N)}(\mathbf{E} - \mathbf{A}) = (\mathbf{E} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^N)(\mathbf{E} - \mathbf{A}) = \mathbf{E} - \mathbf{A}^{N+1}$, 所以

$\mathbf{S}^{(N)} = (\mathbf{E} - \mathbf{A})^{-1} - \mathbf{A}^{N+1}(\mathbf{E} - \mathbf{A})^{-1}$, 故 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 的和为 $\mathbf{S} = \lim_{N \rightarrow \infty} \mathbf{S}^{(N)} = (\mathbf{E} - \mathbf{A})^{-1}$. \square

例 2.11 设 $\mathbf{A} = \begin{pmatrix} 0.1 & 0.7 \\ 0.3 & 0.6 \end{pmatrix}$, 求矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 的和.

解 因矩阵 \mathbf{A} 的行和范数 $\|\mathbf{A}\|_{\infty} = \max_i \left(\sum_{j=1}^n |a_{ij}| \right) = 0.9 < 1$, 故矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$

绝对收敛并且和为 $\sum_{k=0}^{\infty} \mathbf{A}^k = (\mathbf{E} - \mathbf{A})^{-1} = \frac{1}{0.15} \begin{pmatrix} 0.4 & 0.7 \\ 0.3 & 0.9 \end{pmatrix}$.

2.4 矩阵函数

矩阵函数的概念与通常的函数概念类似, 矩阵函数的自变量与因变量均为矩阵. 这一节我们通过收敛的矩阵级数来定义矩阵函数, 并讨论常用的矩阵函数的性质以及矩阵函数值的计算.

2.4.1 矩阵函数的定义

定义 2.14 设幂级数 $\sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 r , 且当 $|z| < r$ 时, 幂级数收敛于函数 $f(z)$, 即

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \quad |z| < r.$$

若 n 阶方阵 \mathbf{A} 满足 $\rho(\mathbf{A}) < r$, 则称收敛的矩阵幂级数 $\sum_{k=0}^{\infty} a_k \mathbf{A}^k$ 的和为矩阵函数, 记为 $f(\mathbf{A})$, 即

$$f(\mathbf{A}) = \sum_{k=0}^{\infty} a_k \mathbf{A}^k.$$

依据上述定义, 我们可以得到一些在形式上与高等数学中的函数类似的矩阵函数. 例如下面我们熟知的函数的幂级数展开式(其中 r 为收敛半径)

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad r = +\infty,$$

$$\sin z = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{(2k+1)!}, \quad r = +\infty,$$

$$\cos z = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{(2k)!}, \quad r = +\infty,$$

$$(1-z)^{-1} = \sum_{k=0}^{\infty} z^k, \quad r = 1,$$

$$\ln(1+z) = \sum_{k=0}^{\infty} \frac{(-1)^k z^{k+1}}{k+1}, \quad r=1.$$

相应地有矩阵函数

$$\begin{aligned} e^{\mathbf{A}} &= \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}, \quad \forall \mathbf{A} \in \mathbb{C}^{n \times n}, \\ \sin \mathbf{A} &= \sum_{k=0}^{\infty} \frac{(-1)^k \mathbf{A}^{2k+1}}{(2k+1)!}, \quad \forall \mathbf{A} \in \mathbb{C}^{n \times n}, \\ \cos \mathbf{A} &= \sum_{k=0}^{\infty} \frac{(-1)^k \mathbf{A}^{2k}}{(2k)!}, \quad \forall \mathbf{A} \in \mathbb{C}^{n \times n}, \\ (\mathbf{E} - \mathbf{A})^{-1} &= \sum_{k=0}^{\infty} \mathbf{A}^k, \quad \rho(\mathbf{A}) < 1, \\ \ln(\mathbf{E} + \mathbf{A}) &= \sum_{k=0}^{\infty} \frac{(-1)^k \mathbf{A}^{k+1}}{k+1}, \quad \rho(\mathbf{A}) < 1. \end{aligned}$$

称 $e^{\mathbf{A}}$ 为矩阵指数函数, $\sin \mathbf{A}$ 为矩阵正弦函数, $\cos \mathbf{A}$ 为矩阵余弦函数.

2.4.2 矩阵函数值的计算

与普通函数值一样, 当给定自变量矩阵 \mathbf{A} 时, 如何求矩阵函数 $f(\mathbf{A})$ 的值呢? 如果不研究切实可行的算法, 矩阵函数值的计算是相当复杂的. 研究如何方便地计算矩阵函数值是非常有意义的. 这里介绍两种常用的算法.

1. 利用哈密顿-凯莱定理

哈密顿-凯莱(Hamilton-Cayley)定理描述了矩阵和其特征多项式之间的关系, 是矩阵理论的一个重要定理.

定理 2.9(哈密顿-凯莱定理) 设 \mathbf{A} 为 n 阶方阵, 其特征多项式为

$$p(\lambda) = |\lambda \mathbf{E} - \mathbf{A}| = \lambda^n + a_1 \lambda^{n-1} + a_2 \lambda^{n-2} + \cdots + a_{n-1} \lambda + a_n,$$

则 $p(\mathbf{A}) = \mathbf{0}$.

利用哈密顿-凯莱定理找到矩阵方幂之间的关系, 然后化简矩阵幂级数求出矩阵函数的值. 下面通过两个例子来说明.

例 2.12 设 $\mathbf{A} = \begin{pmatrix} 0 & 2 \\ -1 & 0 \end{pmatrix}$, 求 $e^{\mathbf{A}}$.

解 计算得矩阵 \mathbf{A} 的特征多项式 $p(\lambda) = |\lambda \mathbf{E} - \mathbf{A}| = \lambda^2 + 2$. 由哈密顿-凯莱定理得 $\mathbf{A}^2 + 2\mathbf{E} = \mathbf{0}$, 因而 $\mathbf{A}^2 = -2\mathbf{E}$, $\mathbf{A}^3 = -2\mathbf{A}$, $\mathbf{A}^4 = (-2)^2 \mathbf{E}$, $\mathbf{A}^5 = (-2)^2 \mathbf{A}$, \cdots , 即 $\mathbf{A}^{2k} = (-2)^k \mathbf{E}$, $\mathbf{A}^{2k+1} = (-2)^k \mathbf{A}$, $k=1, 2, \cdots$. 故

$$\begin{aligned} e^{\mathbf{A}} &= \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} = \mathbf{E} \sum_{k=0}^{\infty} \frac{(-1)^k \sqrt{2}^{2k}}{(2k)!} + \mathbf{A} \frac{1}{\sqrt{2}} \sum_{k=0}^{\infty} \frac{(-1)^k \sqrt{2}^{2k+1}}{(2k+1)!} \\ &= (\cos \sqrt{2}) \mathbf{E} + \frac{1}{\sqrt{2}} (\sin \sqrt{2}) \mathbf{A} = \begin{pmatrix} \cos \sqrt{2} & \sqrt{2} \sin \sqrt{2} \\ -\frac{1}{\sqrt{2}} \sin \sqrt{2} & \cos \sqrt{2} \end{pmatrix}. \end{aligned}$$

例 2.13 设三阶方阵 \mathbf{A} 的特征值为 $\pi, -\pi, 0$, 求 $\sin \mathbf{A}$ 和 $\cos \mathbf{A}$.

解 矩阵 \mathbf{A} 的特征多项式 $p(\lambda) = |\lambda \mathbf{E} - \mathbf{A}| = (\lambda - \pi)(\lambda + \pi)\lambda = \lambda^3 - \pi^2 \lambda$. 由哈密顿-凯莱定理得 $\mathbf{A}^3 - \pi^2 \mathbf{A} = \mathbf{0}$, 故 $\mathbf{A}^3 = \pi^2 \mathbf{A}, \mathbf{A}^4 = \pi^2 \mathbf{A}^2, \mathbf{A}^5 = \pi^4 \mathbf{A}, \mathbf{A}^6 = \pi^4 \mathbf{A}^2, \dots$, 即 $\mathbf{A}^{2k+1} = \pi^{2k} \mathbf{A}$ ($k=1, 2, \dots$), $\mathbf{A}^{2k} = \pi^{2k-2} \mathbf{A}^2$ ($k=2, 3, \dots$). 故

$$\begin{aligned} \sin \mathbf{A} &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \mathbf{A}^{2k+1} \\ &= \mathbf{A} + \sum_{k=1}^{\infty} \frac{(-1)^k}{(2k+1)!} \pi^{2k} \mathbf{A} = \mathbf{A} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \pi^{2k} \\ &= \frac{\mathbf{A}}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \pi^{2k+1} = \frac{\mathbf{A}}{\pi} \sin \pi = \mathbf{0}; \\ \cos \mathbf{A} &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} \mathbf{A}^{2k} = \mathbf{E} - \frac{\mathbf{A}^2}{2!} + \sum_{k=2}^{\infty} \frac{(-1)^k}{(2k)!} \pi^{2k-2} \mathbf{A}^2 = \mathbf{E} + \mathbf{A}^2 \sum_{k=1}^{\infty} \frac{(-1)^k}{(2k)!} \pi^{2k-2} \\ &= \mathbf{E} - \frac{\mathbf{A}^2}{\pi^2} + \frac{\mathbf{A}^2}{\pi^2} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} \pi^{2k} = \mathbf{E} - \frac{\mathbf{A}^2}{\pi^2} + \frac{\mathbf{A}^2}{\pi^2} \cos \pi = \mathbf{E} - \frac{2}{\pi^2} \mathbf{A}^2. \end{aligned}$$

2. 待定系数法

设 \mathbf{A} 为 n 阶方阵, 其特征多项式为

$$p(\lambda) = |\lambda \mathbf{E} - \mathbf{A}| = (\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_t)^{r_t},$$

其中, $\lambda_1, \lambda_2, \dots, \lambda_t$ 为 \mathbf{A} 的全部互异特征值, 且 $r_1 + r_2 + \cdots + r_t = n$.

为计算矩阵函数 $f(\mathbf{A}) = \sum_{k=0}^{\infty} a_k \mathbf{A}^k$, 记 $f(\lambda) = \sum_{k=0}^{\infty} a_k \lambda^k$, 则按多项式除法将 $f(\lambda)$ 改写为

$$f(\lambda) = q(\lambda)p(\lambda) + r(\lambda),$$

其中 $q(\lambda)$ 为 λ 的幂级数, $r(\lambda)$ 为次数不超过 $n-1$ 的 λ 的多项式, 设为

$$r(\lambda) = b_{n-1} \lambda^{n-1} + b_{n-2} \lambda^{n-2} + \cdots + b_1 \lambda + b_0.$$

由哈密顿-凯莱定理知 $p(\mathbf{A}) = \mathbf{0}$, 于是有

$$f(\mathbf{A}) = q(\mathbf{A})p(\mathbf{A}) + r(\mathbf{A}) = b_{n-1} \mathbf{A}^{n-1} + b_{n-2} \mathbf{A}^{n-2} + \cdots + b_1 \mathbf{A} + b_0 \mathbf{E}.$$

注意到

$$p^{(s)}(\lambda_i) = 0, \quad i = 1, 2, \dots, t, \quad s = 0, 1, \dots, r_i - 1,$$

故将等式 $f(\lambda) = q(\lambda)p(\lambda) + r(\lambda)$ 两边对 λ 求导得

$$\left. \frac{d^s}{d\lambda^s} f(\lambda) \right|_{\lambda=\lambda_i} = \left. \frac{d^s}{d\lambda^s} r(\lambda) \right|_{\lambda=\lambda_i}, \quad i = 1, 2, \dots, t, \quad s = 0, 1, \dots, r_i - 1.$$

求解上述以 $b_{n-1}, b_{n-2}, \dots, b_1, b_0$ 为未知量的线性方程组即可算出 $f(\mathbf{A})$.

例 2.14 设 $\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ -4 & -3 & 0 \\ -1 & 2 & 1 \end{pmatrix}$, 求 $e^{\mathbf{A}}$ 和 $\sin \mathbf{A}$.

解 \mathbf{A} 的特征多项式 $p(\lambda) = |\lambda \mathbf{E} - \mathbf{A}| = (\lambda + 2)(\lambda - 1)^2$. 设 $r(\lambda) = b_2 \lambda^2 + b_1 \lambda + b_0$.

(1) 当 $f(\mathbf{A}) = e^{\mathbf{A}}$ 时, 记 $f(\lambda) = e^\lambda$, 则

$$\begin{cases} r(1) = b_2 + b_1 + b_0 = e = f(1), \\ r'(1) = 2b_2 + b_1 = e = f'(1), \\ r(-2) = 4b_2 - 2b_1 + b_0 = e^{-2} = f(-2), \end{cases}$$

解得

$$b_2 = b_0 = \frac{1}{9}(e^{-2} + 2e), \quad b_1 = \frac{1}{9}(5e - 2e^{-2}).$$

因此

$$e^{\mathbf{A}} = b_2 \mathbf{A}^2 + b_1 \mathbf{A} + b_0 \mathbf{E} = \frac{1}{9} \begin{pmatrix} 12e - 3e^{-2} & 3e - 3e^{-2} & 0 \\ 12e^{-2} - 12e & 12e^{-2} - 3e & 0 \\ -27e - 9e^{-2} & -9e^{-2} & 9e \end{pmatrix}.$$

(2) 当 $f(\mathbf{A}) = \sin \mathbf{A}$ 时, 记 $f(\lambda) = \sin \lambda$, 则

$$\begin{cases} r(1) = b_2 + b_1 + b_0 = \sin 1 = f(1), \\ r'(1) = 2b_2 + b_1 = \cos 1 = f'(1), \\ r(-2) = 4b_2 - 2b_1 + b_0 = -\sin 2 = f(-2), \end{cases}$$

解得 $b_2 = \frac{1}{9}(-\sin 2 - \sin 1 + 3\cos 1)$, $b_1 = \frac{1}{9}(2\sin 1 + 3\cos 1 + 2\sin 2)$, $b_0 = \frac{1}{9}(8\sin 1 - 6\cos 1 - \sin 2)$.

故

$$\sin \mathbf{A} = b_2 \mathbf{A}^2 + b_1 \mathbf{A} + b_0 \mathbf{E} = \frac{1}{9} \begin{pmatrix} 12\sin 1 + 3\sin 2 & 3\sin 1 + 3\sin 2 & 0 \\ -12\sin 2 - 12\sin 1 & -12\sin 2 - 3\sin 1 & 0 \\ 9\sin 1 - 36\cos 1 + 9\sin 2 & 9\sin 1 - 9\cos 1 + 9\sin 2 & 9\sin 1 \end{pmatrix}.$$

2.4.3 常用矩阵函数的性质

常用的矩阵函数有 $e^{\mathbf{A}}$, $\sin \mathbf{A}$ 和 $\cos \mathbf{A}$, 它们具有一些与普通指数函数和三角函数类似的性质, 但注意矩阵乘法不满足交换律, 因此有些性质又与普通的指数函数和三角函数不同.

定理 2.10 设 \mathbf{A} 为 n 阶方阵, $i = \sqrt{-1}$, 则

(1) $\sin(-\mathbf{A}) = -\sin \mathbf{A}$; (2) $\cos(-\mathbf{A}) = \cos \mathbf{A}$; (3) $e^{i\mathbf{A}} = \cos \mathbf{A} + i\sin \mathbf{A}$;

(4) $\cos \mathbf{A} = \frac{1}{2}(e^{i\mathbf{A}} + e^{-i\mathbf{A}})$; (5) $\sin \mathbf{A} = \frac{1}{2i}(e^{i\mathbf{A}} - e^{-i\mathbf{A}})$.

证明 由 $\sin \mathbf{A}$ 和 $\cos \mathbf{A}$ 的矩阵幂级数形式直接得到 $\sin(-\mathbf{A}) = -\sin \mathbf{A}$ 和 $\cos(-\mathbf{A}) = \cos \mathbf{A}$.

$$e^{i\mathbf{A}} = \sum_{k=0}^{\infty} \frac{i^k \mathbf{A}^k}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k \mathbf{A}^{2k}}{(2k)!} + i \sum_{k=0}^{\infty} \frac{(-1)^k \mathbf{A}^{2k+1}}{(2k+1)!} = \cos \mathbf{A} + i\sin \mathbf{A},$$

$$e^{-i\mathbf{A}} = \cos(-\mathbf{A}) + i\sin(-\mathbf{A}) = \cos \mathbf{A} - i\sin \mathbf{A},$$

从而 $\cos \mathbf{A} = \frac{1}{2}(e^{i\mathbf{A}} + e^{-i\mathbf{A}})$, $\sin \mathbf{A} = \frac{1}{2i}(e^{i\mathbf{A}} - e^{-i\mathbf{A}})$. □

定理 2.11 设 \mathbf{A}, \mathbf{B} 均为 n 阶方阵, 且 $\mathbf{AB} = \mathbf{BA}$, 则

(1) $e^{\mathbf{A}+\mathbf{B}} = e^{\mathbf{A}} e^{\mathbf{B}} = e^{\mathbf{B}} e^{\mathbf{A}}$;

(2) $\sin(\mathbf{A} + \mathbf{B}) = \sin \mathbf{A} \cos \mathbf{B} + \cos \mathbf{A} \sin \mathbf{B}$;

(3) $\cos(\mathbf{A} + \mathbf{B}) = \cos \mathbf{A} \cos \mathbf{B} - \sin \mathbf{A} \sin \mathbf{B}$.

证明 (1) 由于 $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, 故只需证明 $e^{\mathbf{A} + \mathbf{B}} = e^{\mathbf{A}} e^{\mathbf{B}}$.

$$\begin{aligned} e^{\mathbf{A}} e^{\mathbf{B}} &= \left(\sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} \right) \left(\sum_{k=0}^{\infty} \frac{\mathbf{B}^k}{k!} \right) = \mathbf{E} + (\mathbf{A} + \mathbf{B}) + \frac{1}{2!} (\mathbf{A}^2 + 2\mathbf{A}\mathbf{B} + \mathbf{B}^2) + \dots \\ &= \sum_{k=0}^{\infty} \frac{(\mathbf{A} + \mathbf{B})^k}{k!} = e^{\mathbf{A} + \mathbf{B}}. \end{aligned}$$

$$\begin{aligned} (2) \sin(\mathbf{A} + \mathbf{B}) &= \frac{1}{2i} (e^{i(\mathbf{A} + \mathbf{B})} - e^{-i(\mathbf{A} + \mathbf{B})}) = \frac{1}{2i} (e^{i\mathbf{A}} e^{i\mathbf{B}} - e^{-i\mathbf{A}} e^{-i\mathbf{B}}) \\ &= \frac{1}{2i} (e^{i\mathbf{A}} - e^{-i\mathbf{A}}) \cdot \frac{1}{2} (e^{i\mathbf{B}} + e^{-i\mathbf{B}}) + \frac{1}{2} (e^{i\mathbf{A}} + e^{-i\mathbf{A}}) \cdot \frac{1}{2i} (e^{i\mathbf{B}} - e^{-i\mathbf{B}}) \\ &= \sin\mathbf{A} \cos\mathbf{B} + \cos\mathbf{A} \sin\mathbf{B}. \end{aligned}$$

(3) 同理可证.

特别地, 当取 $\mathbf{A} = \mathbf{B}$ 时, 有下面的结论. □

推论 设 \mathbf{A} 为 n 阶方阵, 则 (1) $\cos 2\mathbf{A} = \cos^2 \mathbf{A} - \sin^2 \mathbf{A}$; (2) $\sin 2\mathbf{A} = 2 \sin \mathbf{A} \cos \mathbf{A}$.

注 2.8 当 $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$ 时, $e^{\mathbf{A} + \mathbf{B}} = e^{\mathbf{A}} e^{\mathbf{B}}$ 或 $e^{\mathbf{A} + \mathbf{B}} = e^{\mathbf{B}} e^{\mathbf{A}}$ 不成立. 如 $\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{B} =$

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \text{ 则 } \mathbf{A} + \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \mathbf{A}\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{B}\mathbf{A}, \text{ 且}$$

$$e^{\mathbf{A}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad e^{\mathbf{B}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad e^{\mathbf{A} + \mathbf{B}} = \frac{1}{2} \begin{pmatrix} e + e^{-1} & e - e^{-1} \\ e - e^{-1} & e + e^{-1} \end{pmatrix},$$

$$\text{因此 } e^{\mathbf{A}} e^{\mathbf{B}} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \neq \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = e^{\mathbf{B}} e^{\mathbf{A}}, e^{\mathbf{A} + \mathbf{B}} \neq e^{\mathbf{A}} e^{\mathbf{B}}, e^{\mathbf{A} + \mathbf{B}} \neq e^{\mathbf{B}} e^{\mathbf{A}}.$$

引理 2.1 设 \mathbf{A} 为 n 阶方阵, $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 \mathbf{A} 的 n 个特征值, 则矩阵函数 $f(\mathbf{A})$ 的 n 个特征值为 $f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)$.

定理 2.12 设 \mathbf{A} 为 n 阶方阵, $\text{tr}\mathbf{A}$ 为矩阵 \mathbf{A} 的迹, 即 $\text{tr}\mathbf{A} = \sum_{i=0}^n a_{ii}$, 则

$$(1) |e^{\mathbf{A}}| = e^{\text{tr}\mathbf{A}}; \quad (2) (e^{\mathbf{A}})^{-1} = e^{-\mathbf{A}}.$$

证明 (1) 设矩阵 \mathbf{A} 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则由引理 2.1 可知, $e^{\mathbf{A}}$ 的特征值为 $e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}$. 从而 $|e^{\mathbf{A}}| = e^{\lambda_1} e^{\lambda_2} \dots e^{\lambda_n} = e^{\lambda_1 + \lambda_2 + \dots + \lambda_n} = e^{\text{tr}\mathbf{A}}$.

(2) 因 $|e^{\mathbf{A}}| = e^{\text{tr}\mathbf{A}} \neq 0$, 故 $e^{\mathbf{A}}$ 可逆. 又 $e^{\mathbf{A}} e^{-\mathbf{A}} = e^{\mathbf{A} - \mathbf{A}} = e^{\mathbf{0}} = \mathbf{E}$, 故 $(e^{\mathbf{A}})^{-1} = e^{-\mathbf{A}}$. □

注 2.9 对于任意的 n 阶方阵 \mathbf{A} , $e^{\mathbf{A}}$ 总是可逆的, 但 $\sin\mathbf{A}$ 和 $\cos\mathbf{A}$ 却不一定可逆. 例如对于矩阵 $\mathbf{A} = \begin{pmatrix} \pi & 0 \\ 0 & \pi/2 \end{pmatrix}$, 有 $\sin\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $\cos\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$, $\sin\mathbf{A}$ 和 $\cos\mathbf{A}$ 都不可逆.

2.5 矩阵的微分和积分

在研究线性微分方程组的问题时, 通过对以函数为元素的矩阵引进微分和积分运算, 来简化问题的表述与求解过程; 同样在研究优化问题(如最小二乘问题)时, 需要处理数量函数对矩阵变量的导数, 以及矩阵值函数对矩阵变量的导数. 这里研究的微积分, 从本质上说

只不过是矩阵重新描述的高等数学中多元微积分的若干结果.

2.5.1 函数矩阵的微分和积分

定义 2.15 以变量 t 的函数为元素的矩阵 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$ 称为函数矩阵. 若 $t \in [a, b]$, 则称 $\mathbf{A}(t)$ 定义在 $[a, b]$ 上; 若每个 $a_{ij}(t)$ 在 $[a, b]$ 上连续、可微、可积, 则称 $\mathbf{A}(t)$ 在 $[a, b]$ 上连续、可微、可积. 当 $\mathbf{A}(t)$ 可微时, 导数定义为 $\mathbf{A}'(t) = (a'_{ij}(t))_{m \times n}$ 或 $\frac{d}{dt}\mathbf{A}(t) = \left(\frac{d}{dt}a_{ij}(t)\right)_{m \times n}$. 当 $\mathbf{A}(t)$ 在 $[a, b]$ 上可积时, 积分定义为 $\int_a^b \mathbf{A}(t) dt = \left(\int_a^b a_{ij}(t) dt\right)_{m \times n}$.

注 2.10 函数矩阵的导数本身还是一个函数矩阵, 还可以继续求导, 因此可以定义函数矩阵的高阶导数, 即

$$\frac{d^k}{dt^k}\mathbf{A}(t) = \frac{d}{dt}\left(\frac{d^{k-1}}{dt^{k-1}}\mathbf{A}(t)\right) = \left(\frac{d^k}{dt^k}a_{ij}(t)\right)_{m \times n}.$$

例 2.15 设函数矩阵 $\mathbf{A}(t) = \begin{pmatrix} 1-t^3 & e^{2t} & 0 \\ 2 & t & 3^t \\ \sin t & -1 & \cos t \end{pmatrix}$, 则

$$\mathbf{A}'(t) = \begin{pmatrix} (1-t^3)' & (e^{2t})' & (0)' \\ (2)' & (t)' & (3^t)' \\ (\sin t)' & (-1)' & (\cos t)' \end{pmatrix} = \begin{pmatrix} -3t^2 & 2e^{2t} & 0 \\ 0 & 1 & 3^t \ln 3 \\ \cos t & 0 & -\sin t \end{pmatrix},$$

$$\int_0^1 \mathbf{A}(t) dt = \begin{pmatrix} \int_0^1 (1-t^3) dt & \int_0^1 e^{2t} dt & 0 \\ \int_0^1 2 dt & \int_0^1 t dt & \int_0^1 3^t dt \\ \int_0^1 \sin t dt & -\int_0^1 dt & \int_0^1 \cos t dt \end{pmatrix} = \begin{pmatrix} 3/4 & (e^2 - 1)/2 & 0 \\ 2 & 1/2 & 2/\ln 3 \\ 1 - \cos 1 & -1 & \sin 1 \end{pmatrix}.$$

函数矩阵的导数运算有如下的性质.

定理 2.13 设 $\mathbf{A}(t)$ 和 $\mathbf{B}(t)$ 为适当阶的可微函数矩阵, $\lambda(t)$ 为可微函数, 则

(1) $\mathbf{A}(t)$ 为常数矩阵的充分必要条件是 $\mathbf{A}'(t) = \mathbf{0}$;

(2) $\frac{d}{dt}(\mathbf{A}(t) \pm \mathbf{B}(t)) = \frac{d}{dt}\mathbf{A}(t) \pm \frac{d}{dt}\mathbf{B}(t)$;

(3) $\frac{d}{dt}(\lambda(t)\mathbf{A}(t)) = \left(\frac{d}{dt}\lambda(t)\right)\mathbf{A}(t) + \lambda(t)\frac{d}{dt}\mathbf{A}(t)$;

(4) $\frac{d}{dt}(\mathbf{A}(t)\mathbf{B}(t)) = \left(\frac{d}{dt}\mathbf{A}(t)\right)\mathbf{B}(t) + \mathbf{A}(t)\frac{d}{dt}\mathbf{B}(t)$;

(5) 若 $\mathbf{A}^{-1}(t)$ 是可微函数矩阵, 有 $\frac{d}{dt}\mathbf{A}^{-1}(t) = -\mathbf{A}^{-1}(t)\left(\frac{d}{dt}\mathbf{A}(t)\right)\mathbf{A}^{-1}(t)$;

(6) 若 $t = f(x)$ 是可微函数, 有 $\frac{d}{dx}\mathbf{A}(x) = f'(x)\frac{d}{dt}\mathbf{A}(t)$.

证明 只证结论(4)和结论(5).

(4) 设 $\mathbf{A}(t) = (a_{ij}(t))_{m \times n}$, $\mathbf{B}(t) = (b_{ij}(t))_{n \times s}$, 则

$$\begin{aligned} \frac{d}{dt}(\mathbf{A}(t)\mathbf{B}(t)) &= \frac{d}{dt} \left(\sum_{k=1}^n a_{ik}(t)b_{kj}(t) \right)_{m \times s} \\ &= \left(\sum_{k=1}^n \left(\frac{d}{dt} a_{ik}(t) \right) b_{kj}(t) + \sum_{k=1}^n a_{ik}(t) \left(\frac{d}{dt} b_{kj}(t) \right) \right)_{m \times s} \\ &= \left(\sum_{k=1}^n \left(\frac{d}{dt} a_{ik}(t) \right) b_{kj}(t) \right)_{m \times s} + \left(\sum_{k=1}^n a_{ik}(t) \left(\frac{d}{dt} b_{kj}(t) \right) \right)_{m \times s} \\ &= \left(\frac{d}{dt} \mathbf{A}(t) \right) \mathbf{B}(t) + \mathbf{A}(t) \frac{d}{dt} \mathbf{B}(t). \end{aligned}$$

(5) 因为 $\mathbf{A}(t)\mathbf{A}^{-1}(t) = \mathbf{E}$, 两边对 t 求导得 $\left(\frac{d}{dt} \mathbf{A}(t) \right) \mathbf{A}^{-1}(t) + \mathbf{A}(t) \frac{d}{dt} \mathbf{A}^{-1}(t) = \mathbf{0}$, 因此, $\frac{d}{dt} \mathbf{A}^{-1}(t) = -\mathbf{A}^{-1}(t) \left(\frac{d}{dt} \mathbf{A}(t) \right) \mathbf{A}^{-1}(t)$. □

例 2.16 设 $\mathbf{A}(t) = \begin{pmatrix} t^2 & \sin t \\ e^{-2t} & 1 \end{pmatrix}$, 计算 $\frac{d^2}{dt^2} \mathbf{A}(t)$ 和 $\frac{d}{dt} \mathbf{A}^{-1}(t)$.

解 $\frac{d}{dt} \mathbf{A}(t) = \begin{pmatrix} 2t & \cos t \\ -2e^{-2t} & 0 \end{pmatrix}$, $\frac{d^2}{dt^2} \mathbf{A}(t) = \begin{pmatrix} 2 & -\sin t \\ 4e^{-2t} & 0 \end{pmatrix}$.

下面计算 $\frac{d}{dt} \mathbf{A}^{-1}(t)$, 需要先计算 $\mathbf{A}^{-1}(t)$.

$$\mathbf{A}^{-1}(t) = \frac{1}{\begin{vmatrix} t^2 & \sin t \\ e^{-2t} & 1 \end{vmatrix}} \begin{pmatrix} 1 & -\sin t \\ -e^{-2t} & t^2 \end{pmatrix} = \frac{1}{t^2 - e^{-2t} \sin t} \begin{pmatrix} 1 & -\sin t \\ -e^{-2t} & t^2 \end{pmatrix},$$

所以

$$\begin{aligned} \frac{d}{dt} \mathbf{A}^{-1}(t) &= -\mathbf{A}^{-1}(t) \left(\frac{d}{dt} \mathbf{A}(t) \right) \mathbf{A}^{-1}(t) \\ &= -\frac{1}{(t^2 - e^{-2t} \sin t)^2} \begin{pmatrix} 1 & -\sin t \\ -e^{-2t} & t^2 \end{pmatrix} \begin{pmatrix} 2t & \cos t \\ -2e^{-2t} & 0 \end{pmatrix} \begin{pmatrix} 1 & -\sin t \\ -e^{-2t} & t^2 \end{pmatrix} \\ &= -\frac{1}{(t^2 - e^{-2t} \sin t)^2} \begin{pmatrix} 2t + e^{-2t}(2\sin t - \cos t) & -2t \sin t - 2e^{-2t} \sin^2 t + t^2 \cos t \\ -2t(1+t)e^{-2t} + e^{-4t} \cos t & e^{-2t}[2t(1+t)\sin t - t^2 \cos t] \end{pmatrix}. \end{aligned}$$

例 2.17 设 \mathbf{A} 为 n 阶方阵, $\mathbf{B}(t)$ 为 n 阶函数方阵, 则

(1) $\frac{d}{dt} e^{\mathbf{A}t} = \mathbf{A} e^{\mathbf{A}t} = e^{\mathbf{A}t} \mathbf{A}$; (2) $\frac{d}{dt} \cos(\mathbf{A}t) = -\mathbf{A} \sin(\mathbf{A}t) = -(\sin(\mathbf{A}t)) \mathbf{A}$;

(3) $\frac{d}{dt} \sin(\mathbf{A}t) = \mathbf{A} \cos(\mathbf{A}t) = (\cos(\mathbf{A}t)) \mathbf{A}$; (4) $\frac{d}{dt} \text{tr} \mathbf{B}(t) = \text{tr} \frac{d\mathbf{B}(t)}{dt}$.

证明 只证结论(1)和结论(4), 而结论(2)和结论(3)的证明与结论(1)类似, 故不赘述.

(1) 因为 $(e^{\mathbf{A}t})_{ij} = \sum_{k=0}^{\infty} \frac{t^k}{k!} (\mathbf{A}^k)_{ij}$, 式子右端是 t 的幂级数, 无论 t 取何值, 它都是收敛的. 所以逐项求导得 $\frac{d}{dt} (e^{\mathbf{A}t})_{ij} = \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} (\mathbf{A}^k)_{ij}$, 即

$$\frac{d}{dt} e^{At} = \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} \mathbf{A}^k = \begin{cases} \mathbf{A} \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} \mathbf{A}^{k-1} = \mathbf{A} e^{At}, \\ \left(\sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} \mathbf{A}^{k-1} \right) \mathbf{A} = e^{At} \mathbf{A}. \end{cases}$$

(4) 因为 $\operatorname{tr} \mathbf{B}(t) = \sum_{i=0}^n b_{ii}(t)$, 所以 $\frac{d}{dt} \operatorname{tr} \mathbf{B}(t) = \sum_{i=0}^n \frac{d}{dt} b_{ii}(t)$, 而

$$\frac{d}{dt} \mathbf{B}(t) = \begin{bmatrix} \frac{d}{dt} b_{11}(t) & \cdots & \frac{d}{dt} b_{1n}(t) \\ \vdots & & \vdots \\ \frac{d}{dt} b_{n1}(t) & \cdots & \frac{d}{dt} b_{nn}(t) \end{bmatrix},$$

所以 $\operatorname{tr} \frac{d\mathbf{B}(t)}{dt} = \sum_{i=0}^n \frac{d}{dt} b_{ii}(t) = \frac{d}{dt} \operatorname{tr} \mathbf{B}(t)$. □

定理 2.14 (行列式微分) 设 $\mathbf{A}(t)$ 为 n 阶函数方阵, 则它的行列式微分为

$$\frac{d}{dt} |\mathbf{A}(t)| = \operatorname{tr} \left[(\mathbf{A}(t))^* \frac{d\mathbf{A}(t)}{dt} \right],$$

其中 $(\mathbf{A}(t))^*$ 为 $\mathbf{A}(t)$ 的伴随矩阵, 特别当 $\mathbf{A}(t)$ 可逆时, 有 $(\mathbf{A}(t))^* = |\mathbf{A}(t)| \mathbf{A}^{-1}(t)$, 得

$$\frac{d}{dt} |\mathbf{A}(t)| = |\mathbf{A}(t)| \operatorname{tr} \left(\mathbf{A}^{-1}(t) \frac{d\mathbf{A}(t)}{dt} \right).$$

例 2.18 设 $\mathbf{A}(t) = \begin{pmatrix} t & 1 & 0 \\ 2t & 3 & t^2 \\ 0 & e^t & 1 \end{pmatrix}$, 求 $\frac{d}{dt} |\mathbf{A}(t)|$.

解 首先计算伴随矩阵 $(\mathbf{A}(t))^*$, $(\mathbf{A}(t))^* = \begin{pmatrix} 3-t^2 e^t & -1 & t^2 \\ -2t & t & -t^3 \\ 2t e^t & -t e^t & t \end{pmatrix}$, 其次计算矩阵

的导数 $\frac{d\mathbf{A}(t)}{dt} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 2t \\ 0 & e^t & 0 \end{pmatrix}$, 再计算乘积 $(\mathbf{A}(t))^* \frac{d\mathbf{A}(t)}{dt} = \begin{pmatrix} 1-t^2 e^t & t^2 e^t & -2t \\ 0 & -t^3 e^t & 2t^2 \\ 0 & t e^t & -2t^2 e^t \end{pmatrix}$,

最后按公式得 $\frac{d}{dt} |\mathbf{A}(t)| = \operatorname{tr} \left[(\mathbf{A}(t))^* \frac{d\mathbf{A}(t)}{dt} \right] = 1 - 3t^2 e^t - t^3 e^t$.

函数矩阵的积分运算有如下的性质.

定理 2.15 设 $\mathbf{A}(t)$ 和 $\mathbf{B}(t)$ 为区间 $[a, b]$ 上适当阶的可积函数矩阵, \mathbf{A} 和 \mathbf{B} 为适当阶的常量矩阵, λ 为数, 则

$$(1) \int_a^b (\mathbf{A}(t) \pm \mathbf{B}(t)) dt = \int_a^b \mathbf{A}(t) dt \pm \int_a^b \mathbf{B}(t) dt;$$

$$(2) \int_a^b \lambda \mathbf{A}(t) dt = \lambda \int_a^b \mathbf{A}(t) dt;$$

$$(3) \int_a^b \mathbf{A}(t) \mathbf{B} dt = \left(\int_a^b \mathbf{A}(t) dt \right) \mathbf{B}; \int_a^b \mathbf{A} \mathbf{B}(t) dt = \mathbf{A} \left(\int_a^b \mathbf{B}(t) dt \right);$$

(4) 若 $\mathbf{A}(t)$ 在区间 $[a, b]$ 上连续, 对任意 $t \in (a, b)$ 有 $\frac{d}{dt} \int_a^t \mathbf{A}(\tau) d\tau = \mathbf{A}(t)$;

(5) 若 $\mathbf{A}(t)$ 在区间 $[a, b]$ 上连续可微, 有 $\int_a^b \mathbf{A}'(t) dt = \mathbf{A}(b) - \mathbf{A}(a)$.

2.5.2 数量函数对矩阵变量的导数

在场论中, 数量函数 $f(x, y, z)$ 的梯度定义为 $\text{grad}f(x, y, z) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)^T$, 它可理解为数量函数 $f(x, y, z)$ 对向量 $(x, y, z)^T$ 的导数. 下面将这个概念推广到一般情况.

定义 2.16 设变量矩阵 $\mathbf{X} = (x_{ij})_{m \times n}$, 函数 $f(\mathbf{X})$ 是以矩阵 \mathbf{X} 为自变量的数量函数, 即为 $m \times n$ 元函数. 若 $\frac{\partial f}{\partial x_{ij}} (i=1, 2, \dots, m; j=1, 2, \dots, n)$ 都存在, 则规定数量函数 $f(\mathbf{X})$ 对矩阵变量 \mathbf{X} 的导数为

$$\frac{df}{d\mathbf{X}} = \left(\frac{\partial f}{\partial x_{ij}} \right)_{m \times n} = \begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{pmatrix}.$$

特别地, 数量函数 $f(\mathbf{x})$ 对向量变量 \mathbf{x} 的导数为 $\frac{df}{d\mathbf{x}} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)^T$.

例 2.19 设 $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$ 为常量向量, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 为变量向量, 求数量函数 $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} = \mathbf{x}^T \mathbf{a}$ 对 \mathbf{x} 的导数 $\frac{df}{d\mathbf{x}}$.

解 因为 $f(\mathbf{x}) = \sum_{i=1}^n a_i x_i = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$, $\frac{\partial f}{\partial x_i} = a_i (i=1, 2, \dots, n)$, 故 $\frac{df}{d\mathbf{x}} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)^T = (a_1, a_2, \dots, a_n)^T = \mathbf{a}$.

例 2.20 已知 $\mathbf{x} = (x_1(t), x_2(t), \dots, x_n(t))^T$, 证明数量函数 $f(\mathbf{x})$ 对变量 t 的导数为

$$\frac{df}{dt} = \left(\frac{df}{d\mathbf{x}} \right)^T \frac{d\mathbf{x}}{dt}.$$

证明 利用偏导数的链式法则得

$$\begin{aligned} \frac{df}{dt} &= \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n} \frac{dx_n}{dt} \\ &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \dots, \frac{dx_n}{dt} \right)^T = \left(\frac{df}{d\mathbf{x}} \right)^T \frac{d\mathbf{x}}{dt}. \quad \square \end{aligned}$$

例 2.21 设 $\mathbf{A} = (a_{ij})_{m \times n}$ 为常量矩阵, $\mathbf{X} = (x_{ij})_{n \times m}$ 为变量矩阵, 求数量函数 $f(\mathbf{X}) = \text{tr}(\mathbf{A}\mathbf{X})$ 对 \mathbf{X} 的导数 $\frac{df}{d\mathbf{X}}$.

解 因为 $\mathbf{A}\mathbf{X} = \left(\sum_{k=1}^n a_{ik} x_{kj} \right)_{m \times m}$, 所以 $f(\mathbf{X}) = \text{tr}(\mathbf{A}\mathbf{X}) = \sum_{s=1}^m \sum_{k=1}^n a_{sk} x_{ks}$, 故 $\frac{\partial f}{\partial x_{ij}} = a_{ji}$,

进而 $\frac{df}{d\mathbf{X}} = \left(\frac{\partial f}{\partial x_{ij}} \right)_{n \times m} = (a_{ji})_{n \times m} = \mathbf{A}^T$.

例 2.22 设 $\mathbf{A} = (a_{ij})_{n \times n}$ 为常量方阵, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 为变量向量. 证明二次型 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 对 \mathbf{x} 的导数为 $\frac{d\mathbf{x}^T \mathbf{A} \mathbf{x}}{d\mathbf{x}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}$.

证明 因 $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i,j=1}^n a_{ij} x_i x_j$, 故

$$\begin{aligned} \frac{d\mathbf{x}^T \mathbf{A} \mathbf{x}}{d\mathbf{x}} &= \left(\frac{\partial}{\partial x_1} \sum_{i,j=1}^n a_{ij} x_i x_j, \dots, \frac{\partial}{\partial x_n} \sum_{i,j=1}^n a_{ij} x_i x_j \right)^T \\ &= \begin{pmatrix} 2a_{11}x_1 + (a_{12} + a_{21})x_2 + \dots + (a_{1n} + a_{n1})x_n \\ \vdots \\ (a_{n1} + a_{1n})x_1 + (a_{n2} + a_{2n})x_2 + \dots + 2a_{nn}x_n \end{pmatrix} \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{pmatrix} + \begin{pmatrix} a_{11}x_1 + a_{21}x_2 + \dots + a_{n1}x_n \\ \vdots \\ a_{1n}x_1 + a_{2n}x_2 + \dots + a_{mn}x_n \end{pmatrix} \\ &= \mathbf{A} \mathbf{x} + \mathbf{A}^T \mathbf{x} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}. \end{aligned}$$

□

注 2.11 (1) 当 \mathbf{A} 是实对称矩阵时, 二次型 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 对 \mathbf{x} 的导数为 $\frac{d\mathbf{x}^T \mathbf{A} \mathbf{x}}{d\mathbf{x}} = 2\mathbf{A} \mathbf{x}$;

(2) 当 $\mathbf{A} = \mathbf{E}$ 时, 二次型 $\mathbf{x}^T \mathbf{A} \mathbf{x} = x_1^2 + x_2^2 + \dots + x_n^2$ 对 \mathbf{x} 的导数为 $\frac{d\mathbf{x}^T \mathbf{A} \mathbf{x}}{d\mathbf{x}} = 2\mathbf{x}$.

例 2.23 设 $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$ 为常量向量, $\mathbf{X} = (x_{ij})_{n \times n}$ 为对称实变量方阵, 求数量函数 $f(\mathbf{X}) = \mathbf{a}^T \mathbf{X} \mathbf{a}$ 对 \mathbf{X} 的导数 $\frac{df}{d\mathbf{X}}$.

解 $\frac{df}{d\mathbf{X}} = \frac{d\mathbf{a}^T \mathbf{X} \mathbf{a}}{d\mathbf{X}} = \left(\frac{\partial}{\partial x_{ij}} \sum_{i,j=1}^n x_{ij} a_i a_j \right)_{n \times n} = (a_i a_j)_{n \times n} = \mathbf{a} \mathbf{a}^T$.

2.5.3 矩阵值函数对矩阵变量的导数

定义 2.17 设矩阵 $\mathbf{F}(\mathbf{X}) = (f_{ij}(\mathbf{X}))_{p \times q}$, 其中 $f_{ij}(\mathbf{X}) (i=1, 2, \dots, p; j=1, 2, \dots, q)$ 都是矩阵变量 $\mathbf{X} = (x_{ij})_{m \times n}$ 的数量函数, 则称 $\mathbf{F}(\mathbf{X})$ 为矩阵值函数, 规定 $\mathbf{F}(\mathbf{X})$ 对矩阵变量 \mathbf{X} 的导数为

$$\frac{d\mathbf{F}}{d\mathbf{X}} = \left(\frac{\partial \mathbf{F}}{\partial x_{ij}} \right)_{m \times n} = \begin{pmatrix} \frac{\partial \mathbf{F}}{\partial x_{11}} & \dots & \frac{\partial \mathbf{F}}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial \mathbf{F}}{\partial x_{m1}} & \dots & \frac{\partial \mathbf{F}}{\partial x_{mn}} \end{pmatrix},$$

其中

$$\frac{\partial \mathbf{F}}{\partial x_{ij}} = \begin{pmatrix} \frac{\partial f_{11}}{\partial x_{ij}} & \cdots & \frac{\partial f_{1q}}{\partial x_{ij}} \\ \vdots & & \vdots \\ \frac{\partial f_{p1}}{\partial x_{ij}} & \cdots & \frac{\partial f_{pq}}{\partial x_{ij}} \end{pmatrix},$$

即最后结果为 $pm \times qn$ 矩阵.

由于向量是特殊的矩阵,因此定义 2.17 包括向量值函数对于向量变量的导数、向量值函数对于矩阵变量的导数以及矩阵值函数对于向量变量的导数等.

例 2.24 设 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 为向量变量,求 $\frac{d\mathbf{x}^T}{d\mathbf{x}}$ 和 $\frac{d\mathbf{x}}{d\mathbf{x}^T}$.

解 按定义可得

$$\frac{d\mathbf{x}^T}{d\mathbf{x}} = \begin{pmatrix} \frac{\partial \mathbf{x}^T}{\partial x_1} \\ \frac{\partial \mathbf{x}^T}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{x}^T}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \mathbf{E}_n.$$

同理可得

$$\frac{d\mathbf{x}}{d\mathbf{x}^T} = \begin{pmatrix} \frac{\partial \mathbf{x}}{\partial x_1} & \frac{\partial \mathbf{x}}{\partial x_2} & \cdots & \frac{\partial \mathbf{x}}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \mathbf{E}_n.$$

例 2.25 设 $\mathbf{a} = (a_1, a_2, a_3)^T$ 为常量向量, $\mathbf{X} = (x_{ij})_{2 \times 3}$ 为矩阵变量,求 $\frac{d(\mathbf{Xa})^T}{d\mathbf{X}}$ 和 $\frac{d(\mathbf{Xa})}{d\mathbf{X}}$.

解 因 $\mathbf{Xa} = \begin{pmatrix} x_{11}a_1 + x_{12}a_2 + x_{13}a_3 \\ x_{21}a_1 + x_{22}a_2 + x_{23}a_3 \end{pmatrix}$,

$$(\mathbf{Xa})^T = (x_{11}a_1 + x_{12}a_2 + x_{13}a_3, x_{21}a_1 + x_{22}a_2 + x_{23}a_3),$$

所以

$$\frac{d(\mathbf{Xa})^T}{d\mathbf{X}} = \begin{pmatrix} \frac{\partial (\mathbf{Xa})^T}{\partial x_{11}} & \frac{\partial (\mathbf{Xa})^T}{\partial x_{12}} & \frac{\partial (\mathbf{Xa})^T}{\partial x_{13}} \\ \frac{\partial (\mathbf{Xa})^T}{\partial x_{21}} & \frac{\partial (\mathbf{Xa})^T}{\partial x_{22}} & \frac{\partial (\mathbf{Xa})^T}{\partial x_{23}} \end{pmatrix} = \begin{pmatrix} a_1 & 0 & a_2 & 0 & a_3 & 0 \\ 0 & a_1 & 0 & a_2 & 0 & a_3 \end{pmatrix},$$

$$\frac{d(\mathbf{X}\mathbf{a})}{d\mathbf{X}} = \begin{pmatrix} \frac{\partial(\mathbf{X}\mathbf{a})}{\partial x_{11}} & \frac{\partial(\mathbf{X}\mathbf{a})}{\partial x_{12}} & \frac{\partial(\mathbf{X}\mathbf{a})}{\partial x_{13}} \\ \frac{\partial(\mathbf{X}\mathbf{a})}{\partial x_{21}} & \frac{\partial(\mathbf{X}\mathbf{a})}{\partial x_{22}} & \frac{\partial(\mathbf{X}\mathbf{a})}{\partial x_{23}} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ a_1 & a_2 & a_3 \end{pmatrix}.$$

例 2.26 设 $f(\mathbf{x})$ 为向量 $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ 的函数, 而 $x_i = x_i(\mathbf{u}) (i=1, 2, \dots, n)$,

$\mathbf{u} = (u_1, u_2, \dots, u_m)^\top$. 证明 $\frac{df}{d\mathbf{u}} = \frac{d\mathbf{x}^\top}{d\mathbf{u}} \frac{df}{d\mathbf{x}}$.

$$\begin{aligned} \text{证明 } \frac{df}{d\mathbf{u}} &= \begin{pmatrix} \frac{\partial f}{\partial u_1} \\ \frac{\partial f}{\partial u_2} \\ \vdots \\ \frac{\partial f}{\partial u_m} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial u_1} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial u_1} + \dots + \frac{\partial f}{\partial x_n} \frac{\partial x_n}{\partial u_1} \\ \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial u_2} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial u_2} + \dots + \frac{\partial f}{\partial x_n} \frac{\partial x_n}{\partial u_2} \\ \vdots \\ \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial u_m} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial u_m} + \dots + \frac{\partial f}{\partial x_n} \frac{\partial x_n}{\partial u_m} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_2}{\partial u_1} & \dots & \frac{\partial x_n}{\partial u_1} \\ \frac{\partial x_1}{\partial u_2} & \frac{\partial x_2}{\partial u_2} & \dots & \frac{\partial x_n}{\partial u_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_1}{\partial u_m} & \frac{\partial x_2}{\partial u_m} & \dots & \frac{\partial x_n}{\partial u_m} \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} = \frac{d\mathbf{x}^\top}{d\mathbf{u}} \frac{df}{d\mathbf{x}}. \quad \square \end{aligned}$$

习 题 2

- (1) 求向量 $(3-2i, 5, i, -2, 0)$ 的 2 范数; (2) 求矩阵 $\begin{pmatrix} 1-i & 4 & 0 \\ 3i & 5+4i & 0 \\ 0 & 1 & 2 \end{pmatrix}$ 的 F 范数.
- 设两个样本特征向量为 $\mathbf{x} = (1, 2, 0, 1)$ 和 $\mathbf{y} = (0, 3, -1, 2)$, 求 \mathbf{x} 和 \mathbf{y} 的余弦相似度.
- 设矩阵序列 $\mathbf{A}^{(k)} = \begin{pmatrix} e^{\frac{2}{k}} & \ln\left(1+\frac{1}{k}\right) \\ \arctan \frac{1}{k} & \frac{k}{k^2+k+2} \end{pmatrix}$, 求 $\lim_{k \rightarrow \infty} \mathbf{A}^{(k)}$.
- 判断下列矩阵是否为收敛矩阵:
 - $\mathbf{A} = \begin{pmatrix} 1/6 & -4/3 \\ -1/3 & 1/6 \end{pmatrix}$;
 - $\mathbf{A} = \begin{pmatrix} 0.3 & 0.2 & 0.3 \\ 0.5 & 0.1 & 0.3 \\ 0.4 & 0.1 & 0.2 \end{pmatrix}$.
- 设有矩阵序列 $\{\mathbf{A}^{(k)}\}_{k=0}^\infty$, 讨论矩阵级数 $\sum_{k=0}^\infty \mathbf{A}^{(k)}$ 的敛散性, 其中

$$\mathbf{A}^{(k)} = \begin{pmatrix} \frac{1}{(5k+1)(5k+6)} & \frac{1}{2^k} + \frac{1}{3^k} \\ 0 & \frac{2k-1}{2^k} \end{pmatrix}.$$

6. 设 $\mathbf{A} = \begin{pmatrix} 0.1 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}$, 判断矩阵幂级数 $\sum_{k=0}^{\infty} \mathbf{A}^k$ 的敛散性, 若收敛求其和.

7. 设 $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, 求 e^{-A^t} .

8. 设 $\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, 求 $\sin \mathbf{A}$.

9. 设 $\mathbf{A} = \begin{pmatrix} 0 & 3 \\ 2 & 0 \end{pmatrix}$, 求 $\cos(\mathbf{A}t)$.

10. 设 $\mathbf{A}(t) = \begin{pmatrix} e^{1-t} & \frac{1}{t+1} \\ \sin t + \cos t & 2t^3 - 1 \end{pmatrix}$, 计算 $\frac{d}{dt} \mathbf{A}(t)$ 和 $\frac{d^2}{dt^2} \mathbf{A}(t)$.

11. 设 $\mathbf{A}(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$, 计算 $\frac{d}{dt} |\mathbf{A}(t)|$, $\left| \frac{d}{dt} \mathbf{A}(t) \right|$ 和 $\frac{d}{dt} \mathbf{A}^{-1}(t)$.

12. 设 $\mathbf{A}(t) = \begin{pmatrix} te^{-t} & t \\ 2t^2 - 1 & 2e^t \end{pmatrix}$, 计算 $\int_0^1 \mathbf{A}(t) dt$.

13. 设 \mathbf{A} 为 $m \times n$ 常量矩阵, \mathbf{X} 为 $n \times m$ 变量矩阵. 证明

$$\frac{d}{d\mathbf{X}} (\text{tr}(\mathbf{A}\mathbf{X})) = \frac{d}{d\mathbf{X}} (\text{tr}(\mathbf{X}^T \mathbf{A}^T)) = \mathbf{A}^T.$$

14. 设 \mathbf{A} 为 $n \times m$ 常量矩阵, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ 为变量向量, 且 $\mathbf{F}(\mathbf{x}) = \mathbf{x}^T \mathbf{A}$, 求 $\frac{d\mathbf{F}}{d\mathbf{x}}$.

15. 设 $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$, $f(\mathbf{X}) = x_{11} + x_{12}^2 + x_{13}^3$, 求 $\frac{df}{d\mathbf{X}}$.

第3章

矩阵分解

矩阵分解作为矩阵理论中非常重要的一部分,是指将一个矩阵分解成一些特殊类型矩阵的乘积(或和)的形式.一方面矩阵的分解式的特殊形式能明显地反映出原矩阵的某些数值特征,如矩阵的秩、行列式、特征值及奇异值等;另一方面矩阵分解的方法与过程往往提供了某些有效的数值计算方法和理论分析的依据.本章将介绍几种常用的矩阵分解形式.

3.1 矩阵的三角分解

三角矩阵的计算,如求行列式、逆矩阵及解线性方程组等,都是很方便的,因此首先研究是否可以将矩阵分解成一些三角矩阵的乘积.

3.1.1 高斯消元法的矩阵形式——三角分解

消元法的基本思想是利用矩阵的初等行变换化矩阵 \mathbf{A} 为上三角矩阵.

定义 3.1 对角线上的元素均为 1 的下三角矩阵称为单位下三角矩阵.

对于 n 阶方阵 \mathbf{A} ,不妨令 $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$.

可按如下步骤分解:(不妨设 $a_{11} \neq 0$,以下类似)

首先将 $-\frac{a_{i1}}{a_{11}}$ 乘以第一行加到第 i 行上 ($i = 2, 3, \dots, n$),可把 \mathbf{A} 变形为 $\mathbf{A}_1 =$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(1)} & \cdots & a_{nn}^{(1)} \end{pmatrix}, \text{这相当于用单位下三角矩阵 } \mathbf{P}_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ -\frac{a_{n1}}{a_{11}} & \cdots & 0 & 1 \end{pmatrix} \text{左乘矩阵}$$

\mathbf{A} 而得到,再将 $-\frac{a_{i2}^{(1)}}{a_{22}^{(1)}}$ 乘以 \mathbf{A}_1 的第二行加到第 i 行上 ($i = 3, 4, \dots, n$),可把 \mathbf{A}_1 变形为

$$\mathbf{A}_2 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a_{n3}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix},$$

这相当于用单位下三角矩阵

$$\mathbf{P}_2 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & -\frac{a_{32}^{(1)}}{a_{22}^{(1)}} & 1 & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & -\frac{a_{n2}^{(1)}}{a_{22}^{(1)}} & 0 & \cdots & 1 \end{pmatrix}$$

左乘以 \mathbf{A}_1 而得到. 如此一直进行下去直到第 $n-1$ 步, \mathbf{A} 就被化为一个上三角矩阵 \mathbf{A}_{n-1} , 记为

$$\mathbf{U} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{pmatrix}, \text{ 此时 } \mathbf{P}_{n-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & \vdots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \vdots & \ddots & 1 & 0 \\ 0 & 0 & \cdots & -\frac{a_{n,n-1}^{(n-2)}}{a_{n-1,n-1}^{(n-2)}} & 1 \end{pmatrix}.$$

上述对 \mathbf{A} 进行的一系列行初等变换, 相当于用矩阵 $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_{n-1}$ 依次左乘 \mathbf{A} , 即 $\mathbf{P}_{n-1}\mathbf{P}_{n-2}\cdots\mathbf{P}_2\mathbf{P}_1\mathbf{A}=\mathbf{U}$. 由于 $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_{n-1}$ 为具有特殊形式的单位下三角矩阵, 所以它们的逆矩阵都存在, 并且这些逆矩阵及其乘积也都是单位下三角矩阵. 因此 $\mathbf{P}_1^{-1}\mathbf{P}_2^{-1}\cdots\mathbf{P}_{n-1}^{-1}\mathbf{P}_{n-1}\cdots\mathbf{P}_2\mathbf{P}_1\mathbf{A}=\mathbf{P}_1^{-1}\mathbf{P}_2^{-1}\cdots\mathbf{P}_{n-1}^{-1}\mathbf{U}$, 即 $\mathbf{A}=\mathbf{P}_1^{-1}\mathbf{P}_2^{-1}\cdots\mathbf{P}_{n-1}^{-1}\mathbf{U}$.

最后令 $\mathbf{L}=\mathbf{P}_1^{-1}\mathbf{P}_2^{-1}\cdots\mathbf{P}_{n-1}^{-1}$, 则 $\mathbf{A}=\mathbf{LU}$, 其中 \mathbf{L} 为单位下三角矩阵, \mathbf{U} 为上三角矩阵.

定义 3.2 设 \mathbf{A} 是 n 阶矩阵, 若有下三角矩阵 \mathbf{L} 和上三角矩阵 \mathbf{U} 使得 $\mathbf{A}=\mathbf{LU}$, 则称 \mathbf{A} 能作三角分解, 并且称 $\mathbf{A}=\mathbf{LU}$ 为 \mathbf{A} 的三角分解或 LU 分解.

定义 3.3 若 \mathbf{A} 的三角分解 $\mathbf{A}=\mathbf{LU}$ 中, \mathbf{L} 为单位下三角矩阵, \mathbf{U} 为上三角矩阵, 此时的三角分解称为杜利特尔(Doolittle)分解; 若 \mathbf{L} 为下三角矩阵, \mathbf{U} 为单位上三角矩阵, 则称此三角分解为克劳特(Crout)分解.

矩阵的三角分解是不唯一的, 杜利特尔分解和克劳特分解就是两种不同的分解方式. 因为若 \mathbf{A} 的三角分解为 $\mathbf{A}=\mathbf{LU}$, 设 \mathbf{D} 是非奇异的任意对角矩阵, 则 $\mathbf{A}=\tilde{\mathbf{L}}\tilde{\mathbf{U}}$ 也是 \mathbf{A} 的三角分解, 其中 $\tilde{\mathbf{L}}=\mathbf{LD}$ 是下三角矩阵, $\tilde{\mathbf{U}}=\mathbf{D}^{-1}\mathbf{U}$ 是上三角矩阵, 由于 \mathbf{D} 的任意性, \mathbf{A} 的三角分解有无穷多种.

关于三角分解的存在性有如下一些结论.

定理 3.1 (LU 分解定理) 设 \mathbf{A} 是 n 阶非奇异矩阵, 则 \mathbf{A} 可以作三角分解的充分必要

条件是 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$, 其中 $\Delta_k = \det \mathbf{A}_k$ 为 \mathbf{A} 的顺序主子式, 而 \mathbf{A}_k 为 \mathbf{A} 的顺序主子阵.

证明 必要性. 已知 \mathbf{A} 可以作三角分解, 即 $\mathbf{A} = \mathbf{L}\mathbf{U}$, 其中

$$\mathbf{L} = (l_{ij})_{n \times n}, \quad l_{ij} = 0, i < j; \quad \mathbf{U} = (u_{ij})_{n \times n}, \quad u_{ij} = 0, i > j.$$

将 $\mathbf{A}, \mathbf{L}, \mathbf{U}$ 进行分块, 得

$$\begin{pmatrix} \mathbf{A}_k & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_k & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{U}_k & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{pmatrix},$$

这里 $\mathbf{A}_k, \mathbf{L}_k, \mathbf{U}_k$ 分别是 $\mathbf{A}, \mathbf{L}, \mathbf{U}$ 的 k 阶顺序主子阵, 且 $\mathbf{L}_k, \mathbf{U}_k$ 分别是下三角矩阵和上三角矩阵. 由矩阵的分块乘法运算, 得

$$\mathbf{A}_k = \mathbf{L}_k \mathbf{U}_k, \quad k = 1, 2, \dots, n.$$

由于

$$\det \mathbf{A} = \det \mathbf{L} \det \mathbf{U} = l_{11} l_{22} \cdots l_{nn} u_{11} u_{22} \cdots u_{nn} \neq 0,$$

所以

$$\Delta_k = \det \mathbf{A}_k = \det \mathbf{L}_k \det \mathbf{U}_k = l_{11} \cdots l_{kk} u_{11} \cdots u_{kk} \neq 0, \quad k = 1, 2, \dots, n-1.$$

充分性. 对阶数 n 用数学归纳法证明. 当 $n=1$ 时, $\mathbf{A}_1 = (a_{11}) = (1)(a_{11})$, 结论成立. 设对 $n=k$ 结论成立, 即 $\mathbf{A}_k = \mathbf{L}_k \mathbf{U}_k$, 其中 $\mathbf{L}_k, \mathbf{U}_k$ 分别是下三角矩阵和上三角矩阵, 且由 $\Delta_k = \det \mathbf{A}_k = \det \mathbf{L}_k \det \mathbf{U}_k \neq 0$ 知, $\mathbf{L}_k, \mathbf{U}_k$ 均可逆. 则当 $n=k$ 时, 有

$$\mathbf{A}_{k+1} = \begin{pmatrix} \mathbf{A}_k & \mathbf{c}_k \\ \mathbf{u}_k^T & a_{k+1, k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_k & \mathbf{0} \\ \mathbf{u}_k^T \mathbf{U}_k^{-1} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{U}_k & \mathbf{L}_k^{-1} \mathbf{c}_k \\ \mathbf{0}^T & a_{k+1, k+1} - \mathbf{u}_k^T \mathbf{U}_k^{-1} \mathbf{L}_k^{-1} \mathbf{c}_k \end{pmatrix},$$

其中 $\mathbf{c}_k = (a_{1, k+1}, \dots, a_{k, k+1})^T, \mathbf{u}_k^T = (a_{k+1, 1}, \dots, a_{k+1, k})$. 故由归纳假设知 \mathbf{A} 可以作三角分解. \square

需要说明的是, 在定理 3.1 中, 单纯从 LU 分解的角度来讲, 不需要“非奇异”这个条件, 只需要 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$ 即可, 请读者自行证明. 在利用矩阵分解求解线性方程组时, “非奇异”可以保证方程组解的唯一性.

定理 3.1 说明, 并不是每个可逆矩阵都可以作三角分解, 如矩阵 $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 就不能作三角分解.

定理 3.2 设 $\mathbf{A} \in \mathbb{C}^{n \times n}, R(\mathbf{A}) = r$, 且 \mathbf{A} 的前 r 个顺序主子式均不为零, 即 $\Delta_k \neq 0 (k=1, 2, \dots, r)$, 则 \mathbf{A} 可以作三角分解.

证明 由定理 3.1 知, 矩阵 \mathbf{A} 的 r 阶顺序主子阵 \mathbf{A}_r 可以作三角分解, 即 $\mathbf{A}_r = \mathbf{L}_r \mathbf{U}_r$, 且 \mathbf{L}_r 和 \mathbf{U}_r 分别是可逆的下三角矩阵和上三角矩阵, 将矩阵 \mathbf{A} 分块为

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_r & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}.$$

由于 $R(\mathbf{A}_r) = R(\mathbf{A}) = r$, 所以 \mathbf{A} 的后 $n-r$ 行可由前 r 行线性表示, 即存在矩阵 $\mathbf{B} \in \mathbb{C}^{(n-r) \times n}$, 使得 $\mathbf{A}_{21} = \mathbf{B}\mathbf{A}_r, \mathbf{A}_{22} = \mathbf{B}\mathbf{A}_{12}$, 从而

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_r & \mathbf{A}_{12} \\ \mathbf{B}\mathbf{A}_r & \mathbf{B}\mathbf{A}_{12} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_r & \mathbf{0} \\ \mathbf{B}\mathbf{L}_r & \mathbf{E}_{n-r} \end{pmatrix} \begin{pmatrix} \mathbf{R}_r & \mathbf{L}_r^{-1} \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

即得到 \mathbf{A} 的一种三角分解. \square

该定理的条件是充分的, 如矩阵 $\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$ 的秩为 1, 不满足定理的条件, 但 $\mathbf{A} = \mathbf{A}\mathbf{E} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$ 等, 都是 \mathbf{A} 的三角分解.

定理 3.3 (LDU 分解定理) 设 \mathbf{A} 是 n 阶非奇异矩阵, 则存在唯一的单位下三角矩阵 \mathbf{L} , 对角矩阵 $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$ 和单位上三角矩阵 \mathbf{U} , 使得

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}$$

的充分必要条件是 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$. 此时对角矩阵 $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$ 的元素满足

$$d_1 = a_{11}, \quad d_k = \frac{\Delta_k}{\Delta_{k-1}}, \quad k=2, 3, \dots, n.$$

证明 由定理 3.1, \mathbf{A} 可作三角分解 $\mathbf{A} = \mathbf{L}\mathbf{U}$ 的充要条件是 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$, 其中 $\mathbf{L} = (l_{ij})_{n \times n} (l_{ij} = 0, i < j)$, $\mathbf{U} = (u_{ij})_{n \times n} (u_{ij} = 0, i > j)$. 由于 \mathbf{A} 是非奇异矩阵, 故 \mathbf{L} 和 \mathbf{U} 均可逆, 所以主对角线元素均非零.

记 $\mathbf{D}_L = \text{diag}(l_{11}, l_{22}, \dots, l_{nn})$, $\mathbf{D}_U = \text{diag}(u_{11}, u_{22}, \dots, u_{nn})$, 由 \mathbf{L} 和 \mathbf{U} 可逆可知 \mathbf{D}_L 和 \mathbf{D}_U 也可逆, 从而 $\mathbf{A} = \mathbf{L}\mathbf{U} = (\mathbf{L}\mathbf{D}_L^{-1})\mathbf{D}_L\mathbf{D}_U(\mathbf{D}_U^{-1}\mathbf{U})$.

这是 \mathbf{A} 的 LDU 分解.

再证唯一性. 设 \mathbf{A} 有两个 LDU 分解, 即

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U} = \tilde{\mathbf{L}}\tilde{\mathbf{D}}\tilde{\mathbf{U}},$$

于是

$$\tilde{\mathbf{L}}^{-1}\mathbf{L} = \tilde{\mathbf{D}}\tilde{\mathbf{U}}\mathbf{U}^{-1}\mathbf{D}^{-1}.$$

上式中左侧为单位下三角矩阵, 右侧是上三角矩阵, 故等式两边都应该是单位矩阵, 即有 $\tilde{\mathbf{L}}^{-1}\mathbf{L} = \mathbf{E}$, $\tilde{\mathbf{D}}\tilde{\mathbf{U}}\mathbf{U}^{-1}\mathbf{D}^{-1} = \mathbf{E}$, 从而 $\mathbf{L} = \tilde{\mathbf{L}}$, $\tilde{\mathbf{U}}\mathbf{U}^{-1} = \tilde{\mathbf{D}}^{-1}\mathbf{D}$. 又由于 $\tilde{\mathbf{U}}\mathbf{U}^{-1}$ 是单位上三角矩阵, 所以 $\tilde{\mathbf{U}}\mathbf{U}^{-1} = \mathbf{E}$, $\tilde{\mathbf{D}}^{-1}\mathbf{D} = \mathbf{E}$, 故 $\mathbf{U} = \tilde{\mathbf{U}}$, $\mathbf{D} = \tilde{\mathbf{D}}$. 故 \mathbf{A} 的 LDU 分解是唯一的.

将 $\mathbf{A}, \mathbf{L}, \mathbf{D}, \mathbf{U}$ 进行分块, 得

$$\begin{pmatrix} \mathbf{A}_k & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_k & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{D}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{U}_k & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{pmatrix},$$

其中, $\mathbf{A}_k, \mathbf{L}_k, \mathbf{D}_k, \mathbf{U}_k$ 分别是 $\mathbf{A}, \mathbf{L}, \mathbf{D}, \mathbf{U}$ 的 k 阶顺序主子阵, 则有

$$\mathbf{A}_k = \mathbf{L}_k\mathbf{D}_k\mathbf{U}_k, \quad k=1, 2, \dots, n.$$

根据 $\Delta_k = \det \mathbf{A}_k = \det \mathbf{L}_k \det \mathbf{D}_k \det \mathbf{U}_k = d_1 d_2 \cdots d_k$, 得

$$d_1 = a_{11}, \quad d_k = \frac{\Delta_k}{\Delta_{k-1}}, \quad k=2, 3, \dots, n. \quad \square$$

推论 设 \mathbf{A} 是 n 阶矩阵, 则 \mathbf{A} 可以唯一地进行杜利特尔分解和克劳特分解的充分必要条件是 \mathbf{A} 的顺序主子式 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$.

注 3.1 与定理 3.1 相同, 定理 3.3 的结论同样可以适当放宽, 即当 $\mathbf{A} \in \mathbb{C}^{n \times n}$ (不要求 \mathbf{A} 可逆时), \mathbf{A} 有唯一 LDU 分解的充分必要条件是 $\Delta_k \neq 0 (k=1, 2, \dots, n-1)$. 证明略去.

注 3.2 矩阵 \mathbf{A} 的 LU 分解与 LDU 分解都需要假设 \mathbf{A} 的前 $n-1$ 阶顺序主子式非零.

如果这个条件不满足,可以给 \mathbf{A} 左(或右)乘以置换矩阵 \mathbf{P} (以 n 阶单位矩阵的 n 个列向量为列作成的 n 阶矩阵),就把 \mathbf{A} 的行(或列)的次序重新排列使之满足这个条件,从而有如下的行交换的矩阵分解定理.

定理 3.4 设 \mathbf{A} 是 n 阶非奇异矩阵,则存在置换矩阵 \mathbf{P} ,使得 \mathbf{PA} 的 n 个顺序主子式均非零,且有 $\mathbf{PA} = \tilde{\mathbf{L}}\tilde{\mathbf{U}} = \mathbf{LDU}$,其中, \mathbf{L} 为单位下三角矩阵, $\tilde{\mathbf{U}}$ 为上三角矩阵, \mathbf{U} 为单位上三角矩阵, \mathbf{D} 为对角矩阵.

证明 应用高斯消元法可得到上三角矩阵 $\tilde{\mathbf{U}}$,设 $\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_r$ 是对应于消去变换的初等变换矩阵, $\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_r$ 是对应于行交换的初等变换矩阵,则有

$$\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_r\mathbf{C}_1\mathbf{C}_2\cdots\mathbf{C}_r\mathbf{A} = \tilde{\mathbf{U}}.$$

令 $\mathbf{P} = \mathbf{C}_1\mathbf{C}_2\cdots\mathbf{C}_r$, 有

$$\mathbf{PA} = (\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_r)^{-1}\tilde{\mathbf{U}} = \mathbf{L}\tilde{\mathbf{U}},$$

其中 $\mathbf{L} = (\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_r)^{-1}$. 由于每一个 $\mathbf{E}_i (i=1, 2, \dots, r)$ 都是单位下三角矩阵,因此 \mathbf{L} 也是单位下三角矩阵. 由于 \mathbf{PA} 的 n 个顺序主子式均非零,由定理 3.3 可得 $\mathbf{PA} = \mathbf{LDU}$. \square

3.1.2 三角分解的紧凑计算格式

现在阐述直接计算三角分解的方法. 以下总假设矩阵 \mathbf{A} 为非奇异矩阵,且 \mathbf{A} 可以作三角分解,即 \mathbf{A} 的所有顺序主子式不为零.

由 \mathbf{A} 的杜利特尔分解 $\mathbf{A} = \mathbf{LU}$, 得

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{nn} \end{pmatrix}.$$

第 1 步: \mathbf{L} 的第 1 行依次乘以 \mathbf{U} 的第 1, 2, \dots , n 列, 得

$$u_{1j} = a_{1j}, \quad j = 1, 2, \dots, n.$$

\mathbf{L} 的第 2, 3, \dots , n 行依次乘以 \mathbf{U} 的第 1 列, 得

$$l_{i1} = \frac{a_{i1}}{u_{11}}, \quad i = 2, 3, \dots, n,$$

故得

$$\begin{cases} u_{1j} = a_{1j}, & j = 1, 2, \dots, n; \\ l_{i1} = \frac{a_{i1}}{u_{11}}, & i = 2, 3, \dots, n. \end{cases}$$

第 2 步: \mathbf{L} 的第 2 行依次乘 \mathbf{U} 的第 2, 3, \dots , n 列, 得

$$l_{21}u_{1j} + u_{2j} = a_{2j}, \quad j = 2, 3, \dots, n,$$

...

最后可得

$$\begin{cases} u_{kj} = a_{kj} - \sum_{t=1}^{k-1} l_{kt} u_{tj}, & j = k, k+1, \dots, n, \quad k = 2, 3, \dots, n; \\ l_{ik} = \left(a_{kj} - \sum_{t=1}^{k-1} l_{it} u_{tk} \right) / u_{kk}, & i = k+1, \dots, n, \quad k = 2, 3, \dots, n. \end{cases}$$

与上面的推导类似, 可得到克劳特分解的紧凑计算格式如下:

$$\begin{cases} l_{i1} = a_{i1}, & i = 1, 2, \dots, n; \\ u_{1j} = \frac{a_{1j}}{l_{11}}, & j = 2, 3, \dots, n; \\ l_{ik} = a_{ik} - \sum_{t=1}^{k-1} l_{it} u_{tk}, & i = k, k+1, \dots, n, \quad k = 2, 3, \dots, n; \\ u_{kj} = \left(a_{kj} - \sum_{t=1}^{k-1} l_{kt} u_{tj} \right) / l_{kk}, & j = k+1, \dots, n, \quad k = 2, 3, \dots, n. \end{cases}$$

例 3.1 求矩阵 $\mathbf{A} = \begin{pmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 3 \end{pmatrix}$ 的杜利特尔分解和克劳特分解.

解 令 $\mathbf{A} = \begin{pmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}$, 代入杜利特尔分解的紧凑

计算格式得

$$\begin{aligned} u_{11} = 2, \quad u_{12} = -1, \quad u_{13} = 3, \quad l_{21} = \frac{1}{2}, \quad l_{31} = 1, \quad u_{22} = \frac{5}{2}, \\ u_{23} = -\frac{1}{2}, \quad l_{32} = 2, \quad u_{33} = 1, \end{aligned}$$

所以

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}.$$

故矩阵 \mathbf{A} 的杜利特尔分解为

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 3 \\ 0 & \frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}.$$

而由克劳特分解的紧凑计算格式得

$$\begin{aligned} l_{11} = 2, \quad l_{21} = 1, \quad l_{31} = 2, \quad u_{12} = -\frac{1}{2}, \quad u_{13} = \frac{3}{2}, \quad l_{22} = \frac{5}{2}, \quad l_{32} = 5, \\ u_{23} = -\frac{1}{5}, \quad l_{33} = 1, \end{aligned}$$

故 \mathbf{A} 的克劳特分解为

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & \frac{5}{2} & 0 \\ 2 & 5 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{3}{2} \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 1 \end{pmatrix}.$$

例 3.2 求解线性方程组

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 2, \\ 2x_1 + 2x_2 + x_3 = 3, \\ 3x_1 + 4x_2 + 2x_3 = 4. \end{cases}$$

解 线性方程组可写为 $\mathbf{Ax} = \mathbf{b}$, 其中

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 4 & 2 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}.$$

首先对 \mathbf{A} 进行三角分解, 得到

$$\mathbf{A} = \mathbf{LU} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -2 & -5 \\ 0 & 0 & -2 \end{pmatrix}.$$

于是线性方程组变为 $(\mathbf{LU})\mathbf{x} = \mathbf{b}$ 或 $\mathbf{L}(\mathbf{U}\mathbf{x}) = \mathbf{b}$.

令 $\mathbf{U}\mathbf{x} = \mathbf{y}$, 则有 $\mathbf{L}\mathbf{y} = \mathbf{b}$, 其中 $\mathbf{y} = (y_1, y_2, y_3)^\top$, 于是有

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \quad \text{即} \begin{cases} y_1 = 2, \\ 2y_1 + y_2 = 3, \\ 3y_1 + y_2 + y_3 = 4, \end{cases} \quad \text{容易求得} \begin{cases} y_1 = 2, \\ y_2 = -1, \\ y_3 = -1. \end{cases}$$

$$\text{再解} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -2 & -5 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \quad \text{即} \begin{cases} x_1 + 2x_2 + 3x_3 = 2, \\ -2x_2 - 5x_3 = -1, \\ -2x_3 = -1, \end{cases} \quad \text{可求得} \begin{cases} x_1 = 2, \\ x_2 = -\frac{3}{4}, \\ x_3 = \frac{1}{2}. \end{cases}$$

此即为原方程组的解.

3.1.3 楚列斯基分解

当 \mathbf{A} 是对称正定矩阵时, 则可以使三角分解的计算量大为减少, 大约是前述的杜利特尔分解工作量的一半.

定理 3.5 若 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 对称正定, 则存在一个实的非奇异下三角矩阵 \mathbf{L} , 使 \mathbf{A} 分解为 $\mathbf{A} = \mathbf{LL}^\top$, 这种分解称为楚列斯基 (Cholesky) 分解, 也称为平方根分解. 如果限定 \mathbf{L} 的对角线为正时, 这种分解是唯一的.

证明 因为 \mathbf{A} 是对称正定的, 所以由线性代数知识知, 它的各阶顺序主子式 $\Delta_i > 0$ ($i = 1, 2, \dots, n$), 再由定理 3.3 知, \mathbf{A} 可唯一地分解为 $\mathbf{A} = \tilde{\mathbf{L}}\mathbf{D}\tilde{\mathbf{U}}$, 其中 $\tilde{\mathbf{L}}$ 为单位下三角矩阵,

$\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$ 为对角矩阵, $\tilde{\mathbf{U}}$ 为单位上三角矩阵.

又因为 $\mathbf{A}^T = \tilde{\mathbf{U}}^T \mathbf{D} \tilde{\mathbf{L}}^T$, 根据 $\mathbf{A} = \mathbf{A}^T$, 所以有 $\tilde{\mathbf{L}} \mathbf{D} \tilde{\mathbf{U}} = \tilde{\mathbf{U}}^T \mathbf{D} \tilde{\mathbf{L}}^T$.

由 $\mathbf{A} = \tilde{\mathbf{L}} \mathbf{D} \tilde{\mathbf{U}}$ 及分解的唯一性, 从而 $\tilde{\mathbf{L}} = \tilde{\mathbf{U}}^T$, 将其代入 $\mathbf{A} = \tilde{\mathbf{L}} \mathbf{D} \tilde{\mathbf{U}}$, 有 $\mathbf{A} = \tilde{\mathbf{U}}^T \mathbf{D} \tilde{\mathbf{U}}$.

因为 $|\tilde{\mathbf{U}}| = 1 \neq 0$, 故 $\tilde{\mathbf{U}}$ 可逆, 由 $\mathbf{A} = \tilde{\mathbf{U}}^T \mathbf{D} \tilde{\mathbf{U}}$ 可知

$$\mathbf{D} = (\tilde{\mathbf{U}}^T)^{-1} \mathbf{A} \tilde{\mathbf{U}}^{-1} = (\tilde{\mathbf{U}}^{-1})^T \mathbf{A} \tilde{\mathbf{U}}^{-1}.$$

由线性代数知, \mathbf{A} 与 \mathbf{D} 是相似的, 而 \mathbf{A} 是对称正定矩阵, 则 \mathbf{D} 也是对称正定矩阵(事实上, 对于任取的非零列向量 $\mathbf{x} \in \mathbb{R}^n$, 显然 $\tilde{\mathbf{U}}^{-1} \mathbf{x} \neq \mathbf{0}$, 二次型 $\mathbf{x}^T \mathbf{D} \mathbf{x} = \mathbf{x}^T (\tilde{\mathbf{U}}^{-1})^T \mathbf{A} \tilde{\mathbf{U}}^{-1} \mathbf{x} = (\tilde{\mathbf{U}}^{-1} \mathbf{x})^T \mathbf{A} (\tilde{\mathbf{U}}^{-1} \mathbf{x}) > 0$, 故 \mathbf{D} 为正定对称矩阵). 既然对角矩阵 \mathbf{D} 对称正定, 那么它的所有一阶主子式, 即主对角线元素 $d_i > 0 (i = 1, 2, \dots, n)$. 令

$$\mathbf{D}^{\frac{1}{2}} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n}),$$

则有唯一的表达式

$$\mathbf{A} = \tilde{\mathbf{L}} \mathbf{D} \tilde{\mathbf{U}} = \tilde{\mathbf{L}} \mathbf{D}^{\frac{1}{2}} \mathbf{D}^{\frac{1}{2}} \tilde{\mathbf{L}}^T = (\tilde{\mathbf{L}} \mathbf{D}^{\frac{1}{2}}) (\tilde{\mathbf{L}} \mathbf{D}^{\frac{1}{2}})^T = \mathbf{L} \mathbf{L}^T,$$

其中 $\mathbf{L} = \tilde{\mathbf{L}} \mathbf{D}^{\frac{1}{2}}$ 是对角线元素全为正数 $\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n}$ 的下三角矩阵. □

设 $\mathbf{A} = (a_{ij})_{n \times n}$, $\mathbf{L} = (l_{ij})_{n \times n} (l_{ij} = 0, i < j)$, 则由 $\mathbf{A} = \mathbf{L} \mathbf{L}^T$, 得

$$a_{ij} = \sum_{k=1}^{j-1} l_{ik} l_{jk} + l_{ij} l_{jj}, \quad i = j, j+1, \dots, n.$$

从而得到对称正定矩阵 \mathbf{A} 的楚列斯基分解的紧凑计算格式如下:

当 $i = j$ 时, 有 $l_{jj} = \left(a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2 \right)^{1/2} (j = 1, 2, \dots, n)$;

当 $i > j$ 时, 有 $l_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk}}{l_{jj}} (i = j+1, \dots, n)$.

注 3.3 当 $j = 1$ 时, 有 $l_{11} = \sqrt{a_{11}}, l_{i1} = \frac{a_{i1}}{l_{11}}$, 对 $j = 2, 3, \dots, n$, 由上面两式逐列求得 \mathbf{L} 的元素 l_{ij} , 即得 \mathbf{A} 的楚列斯基分解.

例 3.3 已知矩阵 $\mathbf{A} = \begin{pmatrix} 5 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 1 \end{pmatrix}$, 求 \mathbf{A} 的楚列斯基分解.

解 先验证矩阵为对称正定, 对称是显然的. 又 $\Delta_1 = 5 > 0, \Delta_2 = \begin{vmatrix} 5 & -2 \\ -2 & 3 \end{vmatrix} = 11 > 0, \Delta_3 = \det \mathbf{A} = 6 > 0$, 故 \mathbf{A} 对称正定. 可用楚列斯基分解, 由 \mathbf{A} 的楚列斯基分解的紧凑计算格式可得

$$l_{11} = \sqrt{5}, \quad l_{21} = -\frac{2}{\sqrt{5}}, \quad l_{31} = 0, \quad l_{22} = \sqrt{\frac{11}{5}}, \quad l_{32} = -\sqrt{\frac{5}{11}}, \quad l_{33} = \sqrt{\frac{6}{11}}.$$

于是

(2) 对 $T_{12}x$ 再构造 $T_{13}(c, s)$: $c = \frac{\sqrt{\xi_1^2 + \xi_2^2}}{\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}}, s = \frac{\xi_3}{\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}}$, 则

$$T_{13}T_{12}x = (\sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}, 0, 0, \xi_4, \dots, \xi_n)^T.$$

(3) 以此类推, 构造

$$T_{1k}(c, s): c = \frac{\sqrt{\xi_1^2 + \dots + \xi_{k-1}^2}}{\sqrt{\xi_1^2 + \xi_2^2 + \dots + \xi_k^2}}, s = \frac{\xi_k}{\sqrt{\xi_1^2 + \xi_2^2 + \dots + \xi_k^2}}, \quad k = 2, 3, \dots, n,$$

则

$$T_{1k}(T_{1,k-1} \cdots T_{13}T_{12}x) = (\sqrt{\xi_1^2 + \xi_2^2 + \dots + \xi_k^2}, 0, 0, \dots, 0, \xi_{k+1}, \dots, \xi_n)^T.$$

直至 $k=n$. 令 $T = T_{1n}T_{1,n-1} \cdots T_{12}$, 则有

$$Tx = (\sqrt{\xi_1^2 + \xi_2^2 + \dots + \xi_n^2}, 0, 0, \dots, 0)^T = \|x\| e_1.$$

再考虑 $\xi_1=0$ 的情形.

若 $\xi_1 = \xi_2 = \dots = \xi_{k-1} = 0, \xi_k \neq 0 (1 < k \leq n)$, 则从第一个不为零的 ξ_k 开始运用上述方法即可. \square

推论 对于任何非零列向量 $x \in \mathbb{R}^n$ 及任何单位列向量 $z (\|z\| = 1)$, 均存在着有限个吉文斯矩阵的乘积 T , 使 $Tx = \|x\| z$.

证明 由定理 3.6, 对 x 存在有限个吉文斯矩阵 $T_{12}^{(1)}, T_{13}^{(1)}, \dots, T_{1n}^{(1)}$ 的乘积 $T^{(1)} = T_{1n}^{(1)} \cdots T_{13}^{(1)} T_{12}^{(1)}$, 使 $T^{(1)}x = \|x\| e_1$.

对 z 同理存在有限个吉文斯矩阵 $T_{12}^{(2)}, T_{13}^{(2)}, \dots, T_{1n}^{(2)}$ 的乘积 $T^{(2)} = T_{1n}^{(2)} \cdots T_{13}^{(2)} T_{12}^{(2)}$, 使 $T^{(2)}z = \|z\| e_1 = e_1$. 于是 $T^{(1)}x = \|x\| e_1 = \|x\| T^{(2)}z = T^{(2)}(\|x\| z)$, 由此得 $(T^{(2)})^{-1}T^{(1)}x = \|x\| z$, 即

$$(T_{1n}^{(2)} \cdots T_{13}^{(2)} T_{12}^{(2)})^{-1} (T_{1n}^{(1)} \cdots T_{13}^{(1)} T_{12}^{(1)})x = \|x\| z,$$

其中

$$\begin{aligned} (T_{1n}^{(2)} \cdots T_{13}^{(2)} T_{12}^{(2)})^{-1} (T_{1n}^{(1)} \cdots T_{13}^{(1)} T_{12}^{(1)}) &= (T_{12}^{(2)})^{-1} (T_{13}^{(2)})^{-1} \cdots (T_{1n}^{(2)})^{-1} T_{1n}^{(1)} \cdots T_{13}^{(1)} T_{12}^{(1)} \\ &= (T_{12}^{(2)})^T (T_{13}^{(2)})^T \cdots (T_{1n}^{(2)})^T T_{1n}^{(1)} \cdots T_{13}^{(1)} T_{12}^{(1)} \end{aligned}$$

为有限个吉文斯矩阵的乘积. \square

例 3.4 用吉文斯变换将向量 $x = (1, -3, 5)^T$ 变换为与 $e_1 = (1, 0, 0)^T$ 同方向.

解 对 x 构造 $T_{12}(c, s)$: $c = \frac{1}{\sqrt{10}}, s = \frac{-3}{\sqrt{10}}$, 则 $T_{12}x = (\sqrt{10}, 0, 5)^T$; 对 $T_{12}x$ 构造

$T_{13}(c, s)$: $c = \frac{\sqrt{10}}{\sqrt{35}}, s = \frac{5}{\sqrt{35}}$, 则 $T_{13}(T_{12}x) = (\sqrt{35}, 0, 0)^T$. 于是

$$T = T_{13}T_{12} = \begin{pmatrix} \frac{\sqrt{10}}{\sqrt{35}} & 0 & \frac{5}{\sqrt{35}} \\ 0 & 1 & 0 \\ -\frac{5}{\sqrt{35}} & 0 & \frac{\sqrt{10}}{\sqrt{35}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} & 0 \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{35}} & -\frac{3}{\sqrt{35}} & \frac{5}{\sqrt{35}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} & 0 \\ -\frac{5}{\sqrt{350}} & \frac{15}{\sqrt{350}} & \frac{10}{\sqrt{35}} \end{pmatrix},$$

$$Tx = \sqrt{35} e_1.$$