

再版序言

随着以人工智能为引领的计算机学科的发展，理工科专业对数学的要求日益提升，这种提升不仅体现在研究生阶段，在本科教育中也有明显体现。本书的改版正是为了适应更多非数学专业的开课需要。

从课程内容看，本书面向的是非数学专业线性代数的接续课程。线性代数课程通常包含了矩阵的基本运算（含行列式）、线性方程组解的结构及求解方法、矩阵的特征值与二次型，以及线性空间初步等内容。其接续课程根据不同学科的要求，有各自的内容侧重，大致涉及矩阵函数、欧氏空间中的线性变换、数值代数初步、张量初步等内容。本书的结构正是按照上述内容来组织编排的。

与第一版相比，这次改版的变化体现在以下四个方面。第一，将多项式的基础知识独立为一章。在第一版中，多项式的基础知识压缩在矩阵的 Jordan 标准形一章的前两节，结构局促，舍去了诸如插值多项式、零化多项式等有用的内容，本次改版将多项式的内容分离出来，做了适当扩充，并在后续章节中，体现了新增内容的应用。第二，将矩阵函数一章提前至 Jordan 标准形后。矩阵函数的定义与计算，与矩阵的 Jordan 标准形和矩阵多项式密切相关，而与上一版位于该章之前的线性变换、欧氏空间两章联系不大，这样的调整更适合教学的安排。第三，补充了部分定理、优化了部分例题解答与定理证明。在内积空间一章中，补充了在后续章节中多次提及的正规变换与正规矩阵的相关内容。对第一版中的部分题解与证明，通过使用补充的定理，简化了原有的证明。第四，补充了参考书目和部分习题，更正了第一版中的部分错误。

在这一版本中，仍然保留了矩阵的基础知识作为第 1 章，使学习这门课的研究生可以快速回顾线性代数课程的基础知识。在教学时，可根据学生的实际情况，适当安排课时或选择跳过。由于线性空间和线性变换、欧氏空间在后续章节中有详细讲解，因此在线性代数中涉及的相关内容不再写入第 1 章中。第 2 章介绍多项式的基础知识，其中数的零化多项式为新增，大部分非数学专业的学生虽然未系统学习过多项式的理论，但由于多项式在各类初级课程中大量出现，学生对其有充分的感性认识，在实际教学中亦可在压缩在 2-3 课时内。第 3-8 章是同济大学矩阵论课程的重点内容，这其中对应于数学专业高等代数课程中的 Jordan 标准形、线性空间与线性变换、内积空间，对应于数值代数课程的矩阵分解与矩阵的广义逆，以及包含于常微分方程课程的矩阵函数。与数学专业的课程相比，本书的内容相对简明，在教学中也以讲解计算方法为主，出于理论完整性的考虑，并为满足部分学生的

需要，书中对大部分定理给出了证明。本书的 9-10 章为特征值的估计和张量简介，根据作者的经验，作为 50 课时的课程，可以将其中一章加入教学内容中，如有更多课时，可以考虑讲授全书的内容，并作适当补充。

本书的两个版本，均参考了原同济大学应用数学系所著的《矩阵分析》一书，虽然章节内容相较过去做了较大改动，但仍借鉴了前辈在多年教学实践中积累的经验。在本书第一版出版后，同济大学数学科学学院的濮燕敏副教授、韩建智副教授、宋林亮副教授、郑宁博士等，提出了很多有益的意见。开课过程中，部分学生指出了第一版中的一些错误。在此对所有为本书的出版与改版给予帮助的老师与同学们，致以诚挚的谢意。限于编者的水平，再版中的不当与疏漏仍在所难免，恳请读者批评指正。

前 言

数学是所有理工科的基础，线性代数又是大学数学的基础课程之一。随着近年来计算机的发展，各理工类的行业对从业人员处理海量数据的能力要求越来越高，与计算机科学密切相关的代数学在理工科、甚至管理科学中的所起的作用越来越大。在这一背景下，以矩阵和线性方程组的基础知识为主要内容的线性代数已成为理工科和管理科学相关专业的本科基础课程，而对知识深度、广度要求更高的研究生来说，还需要更多的线性代数知识，为上述专业的研究生开设更深的线性代数课程，已在各大理工院校得到共识。这本书便是为非数学专业研究生所写的。

矩阵作为代数学的基本工具之一，贯穿于整个线性代数理论的始终，因此大部分院校将研究生的线性代数课程命名为矩阵论或矩阵分析。从内容上说，非数学专业的线性代数课程和矩阵论课程，大致与数学专业的高等代数课程内容相近，虽然深度要求稍低，但要突出与理工科、管理学科相关的内容。由于高等代数是一门非常成熟的数学基础课，因此无论哪本教材，包含的大致内容都相差不多。另一方面，多数非数学专业的线性代数教材都出自数学专业的教师之手，由于当前数学下属的各个二级学科分工越来越细，不同方向的教师知识结构差异越来越大，对线性代数的理解也有一定的不同，这导致了不同方向的教师在内容侧重和讲法上，会有一些差异。

这本教材，是在同济大学应用数学系编写的《矩阵分析》（2005 年版）的基础上，参考了国内兄弟院校的相关课程讲义，并结合 10 多年来课堂教学的经验感受编写而成的，讲述了理学、工学、管理学、经济学所需要的常用的代数理论和方法。我们力求兼顾基础理论和应用，培养学生的逻辑思维、抽象思维以及实际应用的能力。全书共分九章，内容包含了方阵的 Jordan 标准形、线性空间与线性变换、Euclid 空间与酉空间、矩阵分析、矩阵分解、矩阵的广义逆、矩阵特征值的估计、张量等内容。为了配合实际教学的需要，各个章节的内容尽量保持独立，以便在教学中可以做适当的取舍，同时也便于不同背景、不同需求的读者自学。同时，在每一章节之后，配备了一定数量的习题，这些习题不仅可以帮助学生巩固本章的知识点，而且有些习题还是本章内容的延伸，从而满足不同程度的学生的需求。

本书由同济大学数学系吴群副教授（第 1、2、5 章）、周矜君副教授（第 3、4、9 章）、殷俊锋教授（第 6、7、8 章）共同编写，最后由周矜君统稿。同济大学数学系拥有一批常年从事理工科数学教育的优秀教师，他们在教学方法上给予了编者大量建议和指导，可以说没有前辈的积累，就不可能有现在的教材，在此向同济大学数学系的前辈们致以深深的谢意。

本书的编者长期从事同济大学工科研究生基础课矩阵论、数值分析、工科数学等课程的课堂教学，了解理工科学生的特点和需要，在内容的选取上，很多启发恰恰来自于与理工科学生的交流，在此也特别感谢所有参与过编者课堂教学的学生们。同济大学出版社的元福军编审，在本书的出版过程中，付出了大量工作，在此向其表示衷心的感谢。

本书选取了一些新的例题，部分内容采用了一些新的讲法，并加入了一些编者认为对某些专业的研究生有需要，但并没有出现在以往教材中的内容，这些改变是编者在教学中的一些新的尝试，有待进一步的实践检验，欢迎同行与读者对教材提出建议和批评。限于编者的水平，本书中错误与疏漏在所难免，恳请读者不吝指正。

目 录

再版序言	i
前 言	iii
第 1 章 矩阵的基础知识	1
1.1 矩阵及其基本运算	1
1.1.1 加法	2
1.1.2 数乘	3
1.1.3 乘法	3
1.1.4 转置	5
1.1.5 行列式	6
1.1.6 逆矩阵	7
1.1.7 迹	8
1.1.8 共轭	8
1.1.9 矩阵的分块	8
1.2 线性方程组	10
1.2.1 初等变换与初等矩阵	11
1.2.2 阶梯型矩阵	13
1.2.3 矩阵的秩和矩阵的等价标准形	14
1.2.4 向量组的线性相关性	16

1.2.5	线性方程组的求解方法与解的结构	17
1.3	相似矩阵	20
1.3.1	方阵的特征值与特征向量	20
1.3.2	矩阵的相似对角化	21
1.3.3	实对称矩阵与正定矩阵	23
	习题 1	24
第 2 章	多项式	28
2.1	数域	28
2.2	一元多项式	29
2.3	最大公因式和最小公倍数	32
2.4	插值多项式与中国剩余定理	36
2.5	因式分解与不可约多项式	39
2.6	多项式的导数和零化多项式	42
2.6.1	多项式的导数	42
2.6.2	零化多项式	43
	习题 2	44
第 3 章	矩阵的 Jordan 标准形	46
3.1	λ -矩阵的标准形	46
3.2	λ -矩阵标准形的计算方法	50
3.3	矩阵的 Jordan 标准形	54
3.4	矩阵的零化多项式和最小多项式	58
	习题 3	63

第 4 章	矩阵函数	66
4.1	矩阵的极限、导数和积分	66
4.2	矩阵幂级数	69
4.3	矩阵函数的计算方法	73
4.4	利用矩阵函数求解线性微分方程组	77
	习题 4	79
第 5 章	线性空间与线性变换	82
5.1	线性空间	82
5.2	线性空间的维数、基与坐标	84
5.3	子空间	87
5.4	线性变换	90
5.5	线性变换的矩阵	92
5.6	与线性变换相关的子空间	95
	习题 5	98
第 6 章	内积空间	103
6.1	实内积与欧氏空间	103
6.2	标准正交基、度量矩阵与正交补空间	105
6.3	复内积与酉空间	109
6.4	正规变换	111
6.5	正交变换	113
6.6	对称变换	117
	习题 6	119

第 7 章 矩阵分解	123
7.1 矩阵的三角分解	123
7.2 正交三角分解	129
7.3 满秩分解	134
7.4 矩阵的谱分解	136
7.5 奇异值分解	138
习题 7	140
第 8 章 广义逆矩阵	143
8.1 概述	143
8.2 广义逆矩阵 A^-	146
8.3 自反广义逆 $A_{\{1,2\}}$	151
8.4 广义逆矩阵 A^+	153
8.5 广义逆矩阵的应用	159
8.5.1 广义逆在解线性方程组中的应用	159
8.5.2 广义逆在解线性最小二乘问题上的应用	161
习题 8	162
第 9 章 特征值的估计	164
9.1 向量的范数	164
9.2 矩阵的范数	166
9.3 特征值与 Rayleigh 商	170
9.4 圆盘定理	172
习题 9	175

第 10 章 张量	177
10.1 张量的物理描述	177
10.2 张量的运算	180
10.3 张量的代数描述	183
习题 10	186
参考文献	188

草稿，

请勿外传

第 1 章 矩阵的基础知识

矩阵分析是以矩阵作为主要研究对象的课程，是大学线性代数课程的直接后续课程。本章的主要目的是回顾和总结一下大学线性代数的基础知识，便于以后章节的学习。本章中的绝大部分定理，只给出陈述，略去的证明可在大学线性代数教材的相应部分找到。

1.1 矩阵及其基本运算

线性代数起源于线性方程组的求解问题。19 世纪 Cayley¹ 引入矩阵的概念后，不仅使得线性方程组的基本理论可以简洁地表达，亦为更深入地讨论线性方程组解的结构和性质指明了方向。本节将回顾矩阵的概念及包括加法、数乘、乘法、转置、行列式、逆矩阵、迹和共轭等在内的基本矩阵运算。

矩阵是由 $m \times n$ 个数、符号或数学表达式，按顺序排成 m 行 n 列的矩形数表，一般再添加上括号表示为如下形式

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

上述矩阵可简写为 $(a_{ij})_{m \times n}$ ，在明确矩阵为 $m \times n$ 型时，也可进一步简写为 (a_{ij}) 。每个 a_{ij} 都称为矩阵的一个元素，更精确地，称为第 i 行第 j 列的元素或 (i, j) -元素。若以 A 记矩阵，则其 (i, j) -元素也可记为 $\text{ent}_{ij}A$ 。

所有元素均为零的矩阵称为零矩阵，记为 $\mathbf{0}$ 。

只有一行的矩阵称为行矩阵或行向量²；只有一列的矩阵称为列矩阵或列向量。单独的一个数视为 1×1 的矩阵。

$n \times n$ 的矩阵称为 n 阶方阵。在 n 阶方阵 $A = (a_{ij})$ 中，自 A 的左上角到右下角的对角线，刚好穿过全体行标与列标相同的元素 $a_{11}, a_{22}, \dots, a_{nn}$ ，这条对角线称为方阵 A 的主对角线，或简称为对角线。如果一个方阵，除主对角线上的元素以外，其余元素都是零，则

¹Cayley (1821–1895)，英国数学家

²向量和矩阵是不同的概念，不宜等同，这里仅仅是沿用国内通行的教材中的名称。

方阵 A 称为**对角矩阵**, 记为 $\text{diag}(a_{11}, a_{22}, \dots, a_{nn})$. 特别, 主对角线上的元素都是 1 的对角矩阵称为 n 阶**单位矩阵**, 记作 E_n , 或简记为 E . 单位阵的 (i, j) -元素记为 δ_{ij} , 称为 **Kronecker 符号**. 主对角线上(下)方元素均为零的矩阵, 称为下(上)**三角矩阵**, 主对角线元素均为 1 的上(下)三角矩阵, 称为**单位上(下)三角矩阵**.

在中文文献中, 通常用加粗的斜体字母表示矩阵、向量, 普通的斜体字母表示数, 本书遵循这一做法.

例 1.1 平面坐标系下, 任何一个点的坐标 (x, y) 可以看作为一个 1×2 型的矩阵; 空间坐标系下, 任何一个点的坐标 (x, y, z) 可以看作为一个 1×3 型的矩阵. 在绝大多数线性代数教材中, 习惯将坐标写为列矩阵, 本书后续都将沿用列矩阵的记法.

例 1.2 某中学某班学生的某次考试的成绩表, 可视为矩阵. 如

	赵	钱	孙	李	周	吴
语文	80	70	85	75	90	95
数学	95	90	80	85	90	85

就可以看作一个 2×6 型的矩阵(只看表除去第一行、第一列的部分). 如果单独看某一科目的成绩, 则语文、数学的成绩表, 可分别看作 1×6 型的行矩阵; 如果单独看某一个人的成绩, 则可看作 2×1 型的列矩阵. 如果成绩以如下等级计算,

	赵	钱	孙	李	周	吴
语文	B	C	B	C	A	A
数学	A	A	B	B	A	B

则该表也可以认为是矩阵, 只不过以一般文字符号作为元素的矩阵不属于线性代数课程的研究范围.

例 1.3 考虑北京、上海、广州三个城市, 某一天的直飞航班情况.

	北京	上海	广州
北京	0	6	7
上海	6	0	8
广州	7	8	0

表中第一列为出发城市, 第一行为抵达城市, 交叉格内为航班架次, 则上表可看作一个 3×3 型的矩阵.

通过上述例子看出, 矩阵在实际问题中有广泛的应用.

1.1.1 加法

定义 1.1 设 $A = (a_{ij})_{m \times n}$, $B = (b_{ij})_{m \times n}$ 为 $m \times n$ 型矩阵, 则矩阵 $(a_{ij} + b_{ij})_{m \times n}$ 称为 A 与 B 的和, 记为 $A + B$; 矩阵 $(-a_{ij})_{m \times n}$ 称为 A 的负矩阵, 记作 $-A$.

注 1.1 本书中但凡出现矩阵求和, 均默认所有矩阵为同型矩阵, 不再另行说明.

定理 1.1 矩阵加法满足以下基本性质:

- (1) 交换律: $A + B = B + A$;
- (2) 结合律: $(A + B) + C = A + (B + C)$;
- (3) 有零元: $A + \mathbf{0} = A$;
- (4) 有负元: $A + (-A) = \mathbf{0}$.

例 1.4 设

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & -3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 3 & 1 \end{pmatrix}$$

求 $A - B$.

解 直接计算, 有

$$A - B = A + (-B) = \begin{pmatrix} 1 & 0 & 3 \\ 2 & -3 & 4 \end{pmatrix} + \begin{pmatrix} -1 & 1 & -2 \\ 0 & -3 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 2 & -6 & 3 \end{pmatrix}.$$

1.1.2 数乘

定义 1.2 设 $A = (a_{ij})_{m \times n}$ 为矩阵, k 为数, 则矩阵 $(ka_{ij})_{m \times n}$ 称为 k 与 A 的数乘, 记作 kA .

定理 1.2 矩阵数乘满足以下基本性质:

- (1) $1A = A$;
- (2) $(kl)A = k(lA)$;
- (3) 分配律 I: $(k + l)A = kA + lA$;
- (4) 分配律 II: $k(A + B) = kA + kB$.

命题 1.3 矩阵的加法、数乘满足以下性质:

- (1) 消去律: 对任何 A_1, A_2, B , 若 $A_1 + B = A_2 + B$, 则 $A_1 = A_2$;
- (2) 对任何 A , $0A = \mathbf{0}$, $(-1)A = -A$;
- (3) 若 $kA = \mathbf{0}$, 则 $k = 0$ 或 $A = \mathbf{0}$.

1.1.3 乘法

定义 1.3 设 $A = (a_{ij})_{m \times s}$ 为 $m \times s$ 的矩阵, $B = (b_{ij})_{s \times n}$ 为 $s \times n$ 的矩阵, 则 $m \times n$ 的矩阵 $C = (c_{ij})_{m \times n}$ 称为 A 与 B 的乘积, 记作 AB , 其中

$$c_{ij} = \sum_{k=1}^s a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{is}b_{sj}.$$

注 1.2 本书中但凡出现矩阵乘积, 均默认相关矩阵满足可做乘法的基本要求, 不再另行说明.

定理 1.4 矩阵乘法满足以下性质:

- (1) 结合律: $(AB)C = A(BC)$;
- (2) $k(AB) = (kA)B = A(kB)$;
- (3) 分配律: $A(B+C) = AB+AC$, $(B+C)A = BA+CA$;
- (4) 有单位元: $EA = AE = A$;
- (5) 零矩阵的性质: $0A = A0 = 0$.

矩阵乘法的定义源于两个线性变换的复合运算, 例如, 线性变换

$$\begin{cases} x_1 = a_{11}y_1 + a_{12}y_2 + a_{13}y_3, \\ x_2 = a_{21}y_1 + a_{22}y_2 + a_{23}y_3, \end{cases}$$

与线性变换

$$\begin{cases} y_1 = b_{11}z_1 + b_{12}z_2, \\ y_2 = b_{21}z_1 + b_{22}z_2, \\ y_3 = b_{31}z_1 + b_{32}z_2, \end{cases}$$

的复合为

$$\begin{cases} x_1 = (a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31})z_1 + (a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32})z_2, \\ x_2 = (a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31})z_1 + (a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32})z_2. \end{cases}$$

若记

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, Z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix},$$

利用矩阵的乘法, 就可以把两个线性变换分别记为 $X = AY$, $Y = BZ$, 它们的复合为 $X = ABZ$.

注 1.3 矩阵的乘法不满足交换律和消去律, 即一般情况下, $AB = BA$ 不成立; 若 $AB = AC$, 并不能得到 $B = C$. 例如, 设

$$A = \begin{pmatrix} 1 & 2 \\ -2 & -4 \end{pmatrix}, B = \begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix},$$

则有

$$AB = 0, \quad BA = \begin{pmatrix} -6 & -12 \\ 3 & 6 \end{pmatrix},$$

此时 $AB \neq BA$, 且 $AB = A0 = 0$, 但 $B \neq 0$ 不成立.

由于矩阵的乘法不满足交换律, 因此 AB 须表述成 A 左乘 B , 或 B 右乘 A .

定义 1.4 对于 n 阶方阵 A 及正整数 k , 定义 $A^k = A^{k-1}A$ (约定 $A^0 = E$), 称为方阵的幂.

例 1.5 设矩阵 $A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, 求 A^2 和 A^n , 这里 n 为正整数.

解 设 $H = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, 直接计算, 得

$$H^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad H^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad H^n = \mathbf{0} \ (n \geq 4).$$

从而

$$A^2 = (E + H)^2 = E + 2H + H^2 = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$A^n = (E + H)^n = E + C_n^1 H + C_n^2 H^2 + C_n^3 H^3 = \begin{pmatrix} 1 & C_n^1 & C_n^2 & C_n^3 \\ 0 & 1 & C_n^1 & C_n^2 \\ 0 & 0 & 1 & C_n^1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \ (n \geq 3).$$

注 1.4 本题中因为 E 和 H 满足乘积的交换律, 因而可以使用二项式定理, 当两个矩阵的乘积不可交换时, 二项式定理对矩阵不再成立.

1.1.4 转置

定义 1.5 设 $A = (a_{ij})_{m \times n}$ 为一个矩阵, 则将 A 一切 (i, j) -元素变为 (j, i) -元素所得的矩阵

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

称为矩阵 A 的转置, 记作 A^T .

定理 1.5 矩阵的转置满足以下性质:

- (1) $(A^T)^T = A$;
- (2) $(A + B)^T = A^T + B^T$;
- (3) $(kA)^T = kA^T$;
- (4) $(AB)^T = B^T A^T$.

定义 1.6 若 A 为实矩阵, 且满足 $A^T = A$, 则称 A 为对称矩阵.

定义 1.7 若 A 为实矩阵, 且满足 $A^T = -A$, 则称 A 为反对称矩阵.

对称矩阵与反对称矩阵的性质, 将在后续章节中阐述.

1.1.5 行列式

定义 1.8 设 $A = (a_{ij})$ 为 n 阶方阵, 则

$$\det A = \sum_{(p_1, p_2, \dots, p_n) \in S_n} \operatorname{sgn}(p_1, p_2, \dots, p_n) a_{1p_1} a_{2p_2} \cdots a_{np_n} \quad (1.1)$$

称为 A 的行列式, 也记作 $|A|$. 其中 S_n 表示由 $\{1, 2, \dots, n\}$ 的所有全排列构成的集合, (p_1, p_2, \dots, p_n) 取遍 S_n 中的所有全排列, $\operatorname{sgn}(p_1, p_2, \dots, p_n)$ 表示该排列的符号.

注 1.5 排列的符号可通过排列的逆序数计算. 一个排列中, 若存在下标 $i < j$, 使得 $p_i > p_j$, 则下标对 (i, j) 称为一个逆序. 当一个排列的逆序数为偶数时, 排列的符号为正, 当逆序数为奇数时, 排列的符号为负.

根据定义不难得到, 当 $n > 2$ 时, 行列式定义式满足以下特征:

- (1) 求和式的每一项都是 n 个处于 A 的不同行、不同列的元素的积, 共有 $n!$ 项;
- (2) 每一项行指标的排列均为 $(1\ 2\ \cdots\ n)$, 其符号由列指标排列的奇偶性所决定;
- (3) 符号为正的项与符号为负的项个数相同.

定理 1.6 矩阵的行列式满足以下性质:

- (1) $\det A^T = \det A$;
- (2) $\det kA = k^n \det A$;
- (3) $\det AB = \det A \det B$;
- (4) 上(下)三角方阵的行列式恰为主对角线上元素的乘积.

上述定理除去 (3) 的证明需要借助矩阵分块的运算技巧, 其他性质均可直接通过行列式的定义证明.

定义 1.8 中的式 (1.8) 称为行列式的完全展开, 另有一种与之等价的定义方式 (式 (1.2), 称为行列式按行列展开), 亦被很多教材采用.

定理 1.7 n 阶方阵 $A = (a_{ij})$ 去掉 a_{ij} 所在的行与列后所得到的 $(n-1)$ 阶方阵的行列式, 称为元素 a_{ij} 的余子式, 记为 M_{ij} , $(-1)^{i+j} M_{ij}$ 称为 a_{ij} 的代数余子式, 记为 A_{ij} . 此时

$$\det A = \sum_{k=1}^n a_{ik} A_{ik} = \sum_{k=1}^n a_{kj} A_{kj}. \quad (1.2)$$

当 $i \neq j$ 时,

$$\sum_{k=1}^n a_{ik}A_{jk} = \sum_{k=1}^n a_{ki}A_{kj} = 0.$$

注 1.6 矩阵的行列式是一个数, 余子式、代数余子式都是行列式, 从而也是数.

定理 1.8 设 $A = (a_{ij})$ 为一个 n 阶方阵, A_{ij} 为元素 a_{ij} 的代数余子式, 矩阵

$$A^* = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

称为 A 的伴随矩阵, 则 $AA^* = A^*A = (\det A)E$.

本定理是定理 1.7 的直接推论.

注 1.7 无论是机器计算还是手工计算, 利用行列式的完全展开或按行列展开计算一般方阵的行列式, 计算量都很大, 不可取. 借助矩阵的初等变换计算行列式, 通常较为简便.

1.1.6 逆矩阵

定义 1.9 设 $A = (a_{ij})$ 为一个 n 阶方阵, 若存在一个 n 阶方阵 B , 使得 $AB = BA = E$, 则称 A 是可逆的或非奇异的 (否则称为不可逆的或奇异的), 并称 B 为 A 的逆矩阵, 记作 A^{-1} .

事实上, 定义中的条件形式上可以减弱.

定理 1.9 对于 n 阶方阵 A, B , 若 $AB = E$, 则 $BA = E$.

下列命题给出了方阵 A 可逆的充分必要条件.

定理 1.10 A 可逆当且仅当 $|A| \neq 0$, 且 $A^{-1} = |A|^{-1}A^*$.

注 1.8 如果利用定理 1.10 计算逆矩阵, 需要计算 $(n^2 + 1)$ 个行列式, 计算量太大, 不可取, 具体计算逆矩阵, 通常要借助矩阵的初等变换.

逆矩阵关于矩阵的数乘、乘法、转置和行列式运算有以下性质.

定理 1.11 设 A, B 为 n 阶可逆矩阵, k 为数, 则

- (1) A^{-1} 是可逆的, 且 $(A^{-1})^{-1} = A$;
- (2) 当 $k \neq 0$ 时, kA 是可逆的, 且 $(kA)^{-1} = k^{-1}A^{-1}$;
- (3) AB 是可逆的, 且 $(AB)^{-1} = B^{-1}A^{-1}$;
- (4) A^T 是可逆的, 且 $(A^T)^{-1} = (A^{-1})^T$;
- (5) $\det A^{-1} = (\det A)^{-1}$.

例 1.6 求二阶方阵 $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 的逆矩阵, 这里 $ad - bc \neq 0$.

解 $|A| = ad - bc \neq 0$, $A^* = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, 故 $A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

1.1.7 迹

定义 1.10 设 $A = (a_{ij})$ 为一个 n 阶方阵, 则 A 的主对角线上元素的和称为 A 的迹, 记作 $\text{tr} A$, 即

$$\text{tr} A = a_{11} + a_{22} + \cdots + a_{nn}.$$

定理 1.12 方阵的迹运算满足以下性质:

- (1) $\text{tr} A + \text{tr} B = \text{tr}(A + B)$;
- (2) $\text{tr}(kA) = k \text{tr} A$;
- (3) $\text{tr}(AB) = \text{tr}(BA)$.

例 1.7 证明对任意的方阵 A, B 都有 $AB - BA \neq E$.

证明 由于 $\text{tr}(AB - BA) = \text{tr}(AB) - \text{tr}(BA) = 0$, 而 $\text{tr} E = n$, 因此对任意的方阵 A, B 都有 $AB - BA \neq E$. 证毕.

例 1.8 设 A 为 $m \times n$ 型的实矩阵, 试证 $A^T A = 0$ 当且仅当 $A = 0$.

证明 充分性. 若 $A = 0$, 显然有 $A^T A = 0$.

必要性. 设 $A = (a_{ij})_{m \times n}$, 则 $\text{tr} A^T A = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 = 0$, 从而 $a_{ij} = 0$ 对一切 i, j 成立,

即 $A = 0$. 证毕.

1.1.8 共轭

定义 1.11 设 $A = (a_{ij})_{m \times n}$ 为一个复矩阵, \bar{a}_{ij} 为元素 a_{ij} 的共轭复数, 则复矩阵 $(\bar{a}_{ij})_{m \times n}$ 称为 A 的共轭矩阵, 记作 \bar{A} .

定理 1.13 矩阵的共轭运算满足以下性质:

- (1) $\overline{A + B} = \bar{A} + \bar{B}$,
- (2) $\overline{kA} = \bar{k} \bar{A}$,
- (3) $\overline{AB} = \bar{A} \bar{B}$.

注 1.9 对复矩阵, 共轭与转置两种运算经常同时进行, 本书中将 \bar{A}^T 简记为 A^H , 在其他文献中也有另外的记法. 若 A 满足 $A^H = A$ (或 $A^H = -A$), 则称 A 为 **Hermite 矩阵** (或 **反 Hermite 矩阵**).

1.1.9 矩阵的分块

将高阶的矩阵分成若干块低阶的矩阵, 是矩阵运算的重要技巧. 在做加法、数乘、乘

法、转置、共轭计算时,只要相关的矩阵运算是有定义的,可将矩阵的子块当成数来对待,得到的结果都是正确的,但对行列式、矩阵求逆、迹的运算,不能简单地将矩阵块处理为数. 具体来说,有以下基本结论.

定理 1.14 (1) 设 $A = (A_{ij})$ 和 $B = (B_{ij})$ 为分块方式完全相同的同型矩阵, 则 $A + B = (A_{ij} + B_{ij})$.

(2) 设 $A = (A_{ij})$, 则 $kA = (kA_{ij})$.

(3) 设 $A = (A_{ij})$ 和 $B = (B_{jk})$, A 按列分块的方式与 B 的按行分块的方式完全相同, 则 $AB = (C_{ik})$, 其中 $C_{ik} = \sum_j A_{ij}B_{jk}$.

(4) 设 $A = (A_{ij})$, 则 $A^T = (A_{ji}^T)$, $A^H = (A_{ji}^H)$.

(5) 设 $A = \text{diag}(A_1, A_2, \dots, A_m)$, 则 $A^{-1} = \text{diag}(A_1^{-1}, A_2^{-1}, \dots, A_m^{-1})$.

(6) 设 A 为分块上(下)三角方阵, A_1, A_2, \dots, A_m 对角线穿过的矩阵块, 则

$$\det A = \det A_1 \det A_2 \cdots \det A_m.$$

例 1.9 设 $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 1 & 5 \end{pmatrix}$, 求 A^{-1} .

解 先将矩阵 A 分块,

$$A = \left(\begin{array}{c|cc} 3 & 0 & 0 \\ \hline 0 & 1 & 2 \\ 0 & 1 & 5 \end{array} \right) = \begin{pmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{pmatrix},$$

因

$$A_1 = 3, A_1^{-1} = \frac{1}{3}, A_2 = \begin{pmatrix} 1 & 2 \\ 1 & 5 \end{pmatrix}, A_2^{-1} = \begin{pmatrix} 5 & -2 \\ -1 & 1 \end{pmatrix},$$

故

$$A^{-1} = \begin{pmatrix} A_1^{-1} & \mathbf{0} \\ \mathbf{0} & A_2^{-1} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & -2 \\ 0 & -1 & 1 \end{pmatrix}.$$

虽然上述定理中,对行列式和矩阵求逆运算,对矩阵的形式有要求,但对一般的矩阵,仍可借助一定的变换,化为适合定理要求的矩阵.

例 1.10 设 A, B, C, D 为 n 阶方阵, 其中 A 是可逆阵, 证明

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| \cdot |D - CA^{-1}B|.$$

$$\begin{pmatrix} E & \mathbf{0} \\ -CA^{-1} & E \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} E & -A^{-1}B \\ \mathbf{0} & E \end{pmatrix} = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & D - CA^{-1}B \end{pmatrix},$$
$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} E & 0 \\ -CA^{-1} & E \end{vmatrix} \begin{vmatrix} A & B \\ C & D \end{vmatrix} \begin{vmatrix} E & -A^{-1}B \\ 0 & E \end{vmatrix} = \begin{vmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{vmatrix} = |A| \cdot |D - CA^{-1}B|.$$

例 1.11 设 A, B 为 n 阶方阵, 证明

$$(2) \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{vmatrix} = |\mathbf{A} + \mathbf{B}| \cdot |\mathbf{A} - \mathbf{B}|.$$

$$\begin{pmatrix} \mathbf{0} & E \\ E & \mathbf{0} \end{pmatrix} \begin{pmatrix} A & \mathbf{0} \\ E & B \end{pmatrix} \begin{pmatrix} E & B \\ \mathbf{0} & -E \end{pmatrix} = \begin{pmatrix} E & \mathbf{0} \\ A & AB \end{pmatrix},$$

(2) 利用分块矩阵可得

$$\begin{pmatrix} \mathbf{0} & E \\ -E & E \end{pmatrix} \begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} E & \mathbf{0} \\ E & E \end{pmatrix} = \begin{pmatrix} A+B & B \\ \mathbf{0} & A-B \end{pmatrix}.$$

上式两边取行列式, 即得结论. 证毕.

1.2 线性方程组

考虑线性方程组

[illegible]

若记 $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$, 方程组 (1.3) 可简写为 $Ax = b$. 线性

方程组的基本问题是: 是否有解、有多少个解、如何求解. 线性代数的理论正是围绕这三个基本问题展开的.

矩阵 A , x 和 b 分别称为线性方程组 (1.3) 的系数矩阵、未知数矩阵和常数项矩阵, 则线性方程组 (1.3) 可记为 $Ax = b$. 分块矩阵

$$(A \ b) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix},$$

称为线性方程组 (1.3) 的增广矩阵. 若将 A 按列分块, 即 $A = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n)$, 其中 $\alpha_j = (a_{1j} \ a_{2j} \ \cdots \ a_{mj})^T$, 则线性方程组 (1.3) 又可表示为

$$x_1 \alpha_1 + x_2 \alpha_2 + \cdots + x_n \alpha_n = b.$$

当系数矩阵为方阵时, 有下面的简明结论.

定理 1.15 (Cramer¹法则) 如果线性方程组 $Ax = b$ 的系数矩阵为方阵, 则线性方程组 $Ax = b$ 存在唯一解的充分必要条件为系数矩阵 A 可逆, 且此时

$$x = A^{-1}b = \frac{1}{\det A} \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1 \ k-1} & b_1 & a_{1 \ k+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2 \ k-1} & b_2 & a_{2 \ k+1} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n \ k-1} & b_n & a_{n \ k+1} & \cdots & a_{nn} \end{vmatrix}.$$

特别, 线性方程组 $Ax = b$ 存在唯一解当且仅当齐次方程 $Ax = 0$ 只有零解.

Cramer 法则虽然给出了系数矩阵为可逆阵时, 线性方程组的求解公式, 但使用该公式计算方程的解, 需要计算 $(n+1)$ 个行列式, 计算量很大. 通常情况下, 求解线性方程组需要借助矩阵的初等变换.

1.2.1 初等变换与初等矩阵

利用矩阵的初等变换实现通过加减消元法求解线性方程组, 是在线性代数理论体系下研究方程的基本方法.

定义 1.12 下列三类变换称为矩阵的初等变换:

- (1) 将第 i, j 两行 (或列) 对换, 记作 $r_i \leftrightarrow r_j$ (或 $c_i \leftrightarrow c_j$);
- (2) 将第 i 行 (或列) 乘非零数 k , 记作 kr_i (或 kc_i);

¹Cramer (1704–1752), 克莱默, 瑞士数学家.

(3) 将第 i 行 (或列) 加上第 j 行 (或列) 的 k 倍, 记作 $r_i + kr_j$ (或 $c_j + kc_i$).

注 1.10 如果考虑某个线性方程组对应的增广矩阵, 则三类初等行变换分别对应了线性方程组的三类同解变形: 将两个方程互换; 将某个方程两边同乘一个非零常数; 将某个方程加另一个方程的常数倍. 其中后两种同解变形正是求解线性方程组的本质、关键的步骤.

三种初等变换对应三类初等矩阵, 分别记为

$$(1) E(i, j) = E - (e_i - e_j)(e_i - e_j)^T = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 0 & 1 & \\ & & & \ddots & \\ & 1 & & 0 & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} \begin{matrix} i \\ ; \\ j \end{matrix}$$

$$(2) E(i(k)) = E + (k-1)e_i e_i^T = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & k & \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \begin{matrix} i \\ ; \end{matrix}$$

$$(3) E(i, j(k)) = E + k e_i e_j^T = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & k & \\ & & & \ddots & \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \begin{matrix} i \\ . \\ j \end{matrix}$$

这里 e_i 为单位矩阵 E 的第 i 列. 对矩阵做一次初等行变换等同于对该矩阵左乘相应的初等矩阵, 而做一次初等列变换等同于该矩阵右乘相应的初等矩阵, 具体表现为

(1) $A \xrightarrow{r_i \leftrightarrow r_j} B$ 当且仅当 $E(i, j)A = B$; $A \xrightarrow{c_i \leftrightarrow c_j} B$ 当且仅当 $AE(i, j) = B$;

(2) $A \xrightarrow{kr_i} B$ 当且仅当 $E(i(k))A = B$; $A \xrightarrow{kc_j} B$ 当且仅当 $AE(i(k)) = B$;

(3) $A \xrightarrow{r_i + kr_j} B$ 当且仅当 $E(i, j(k))A = B$; $A \xrightarrow{c_i + kc_j} B$ 当且仅当 $AE(i, j(k))^T = B$.

方阵的初等变换与方阵的行列式之间满足以下性质.

定理 1.16 设 A 为方阵, 则

- (1) 对 A 施行第一类初等变换, 行列式变为 $-\det A$;
- (2) 对 A 施行第二类初等变换, 行列式变为 $k \det A$;
- (3) 对 A 施行第三类初等变换, 行列式不变.

上述定理可通过行列式的定义直接证明, 也可通过初等变换与初等矩阵的关系得到. 在计算一般矩阵的行列式时, 通常都是利用初等变换将矩阵变为方便计算行列式的矩阵类型(如上三角矩阵), 从而根据初等变换与行列式的关系, 得到原矩阵的行列式.

由初等变换与初等矩阵的关系易见, 矩阵的初等变换都是可逆的, 并且其逆变换是同一类型的初等变换; 相应地, 初等矩阵也都是可逆的, 并且其逆是同一类型的初等矩阵.

定义 1.13 若 n 阶方阵 P 为有限个 $E(i, j)$ 类的初等矩阵的乘积, 即存在有限个 $E(i, j)$ 类的初等矩阵 P_1, P_2, \dots, P_s , 使得 $P = P_1 P_2 \cdots P_s$, 则称 P 为 n 阶置换阵.

显然, 置换阵就是第一类的初等矩阵.

定义 1.14 矩阵 A 经过有限次初等变换变成 B , 则称 A 与 B 等价, 记作 $A \sim B$.

由矩阵等价的定义, 立即得到下面的结论.

定理 1.17 矩阵等价满足以下三条性质.

- (1) 自反性: 对任何矩阵 A , $A \sim A$;
- (2) 对称性: 若 $A \sim B$, 则 $B \sim A$;
- (3) 传递性: 若 $A \sim B$, $B \sim C$; 则 $A \sim C$.

注 1.11 类似矩阵等价一样, 满足自反性、对称性、传递性的“关系”有很多, 例如数的相等关系, 三角形的全等关系、相似关系等. 一般地, 如果某种联系两个数学对象的“关系”(称为二元关系)满足自反性、对称性、传递性, 则统称这种关系为等价关系. 于是, 矩阵的等价, 三角形的全等与相似, 数的相等, 都是等价关系.

定理 1.18 矩阵 A 与 B 等价当且仅当存在可逆阵 P 和 Q , 使 $PAQ = B$.

利用初等矩阵和初等变换的关系, 还可以计算矩阵的逆.

定理 1.19 方阵 A 可逆当且仅当存在有限个初等矩阵 P_1, P_2, \dots, P_s , 使 $AP_1 P_2 \cdots P_s = E$ (或 $P_1 P_2 \cdots P_s A = E$), 此时, $A^{-1} = P_1 P_2 \cdots P_s$.

根据上述定理, 求一个矩阵 A 的逆矩阵, 只需对矩阵 A 和单位矩阵 E 实施相同的初等行变换(或初等列变换, 但不可混合使用初等行、列变换), 当 A 变为单位矩阵时, 相应的 E 就变为 A^{-1} .

1.2.2 阶梯型矩阵

在利用矩阵的初等行变换求解线性方程组的过程中, 行阶梯形矩阵和行最简形矩阵是两个重要的概念.

定义 1.15 一个行阶梯形矩阵是一个满足下列条件的矩阵:

- (1) 若有全零行, 则全零行都在矩阵的最下面;
- (2) 非零行按其第一个非零元素的行标与列标同时增加的规则排列.

矩阵的非零行的第一个非零元素也通常称为这一行的**首非零元素**，简称**首元素**。

一个**行最简形矩阵**是一个满足下列条件的行阶梯形矩阵：

- (1) 非零行的首非零元素都是 1；
- (2) 非零行的首非零元素所在的列的其余元素都是 0。

任意一个矩阵 A 都可以经过有限次初等行变换变成行阶梯形矩阵 A_0 ，以及行最简形矩阵 A_1 。记 A_0 为 A 的行阶梯形矩阵， A_1 为 A 的行最简形矩阵。显然一个矩阵的行阶梯形矩阵不是唯一的，而行最简形矩阵则是唯一的。

1.2.3 矩阵的秩和矩阵的等价标准形

在利用矩阵研究线性方程组的可解性时，矩阵的秩是一个重要的概念。矩阵的秩本质地刻画了方程组中，相互独立的方程的个数。

定义 1.16 矩阵 A 的非零子式的最高阶数，称为矩阵 A 的**秩**，记作¹ $\text{rank } A$ ，并且规定零矩阵的秩为零。

矩阵的秩是矩阵的一个非常重要的特性。

定理 1.20 矩阵 A 的秩有以下的性质：

(1) $\text{rank } A \geq r$ 当且仅当 A 有一个 r 阶非零子式； $\text{rank } A < r$ 当且仅当 A 的所有 r 阶子式全为零；

(2) 若 $A \sim B$ ，则 $\text{rank } A = \text{rank } B$ ；

(3) 若矩阵 A 的行阶梯形矩阵是 A_0 ，则 $\text{rank } A$ 就等于 A_0 中非零行的行数；

(4) $\text{rank } A = \text{rank } A^T$ ；

(5) 若 A 为 $m \times n$ 阵，则 $\text{rank } A \leq \min\{m, n\}$ ；

(6) 若 A, B 是同型矩阵，则 $\text{rank}(A + B) \leq \text{rank } A + \text{rank } B$ ；

(7) 若 A 为 $m \times n$ 阵， B 为 $n \times s$ 阵，则 $\text{rank } AB \leq \min\{\text{rank } A, \text{rank } B\}$ ；

(8) 若 A 为 $m \times n$ 阵， B 为 $n \times s$ 阵，且 $AB = 0$ ，则 $\text{rank } A + \text{rank } B \leq n$ ；

(9) $\text{rank } A^T A = \text{rank } A$ ，其中 A 为实矩阵。

设 $m \times n$ 阵 A 的秩为 r ，则存在可逆阵 P 和 Q ，使得

$$PAQ = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix},$$

则分块矩阵 $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ 称为 A 的**等价标准形**。

一个矩阵 A 总可以经过有限次初等变换变成 A 的等价标准形。

例 1.12 设 n 阶阵 A 的秩为 r ，试证明存在秩为 $(n - r)$ 的 n 阶阵 B ，使得 $AB = 0$ 。

¹部分参考书中将矩阵 A 的秩记为 $r(A)$ ，本书中为了避免与第 9 章中出现的另一符号混淆，未采用这一记法。

证明 由于 $\text{rank } A = r$, 因此存在可逆阵 P 和 Q , 使得

$$PAQ = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix},$$

取 $B = Q \begin{pmatrix} 0 & 0 \\ 0 & E_{n-r} \end{pmatrix}$, 则 $\text{rank } B = n - r$, 于是

$$PAB = PAQ \begin{pmatrix} 0 & 0 \\ 0 & E_{n-r} \end{pmatrix} = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & E_{n-r} \end{pmatrix} = 0,$$

因为 P 为可逆阵, 所以 $AB = 0$, 且 $\text{rank } B = n - r$. 证毕.

例 1.13 设 A 为 $m \times n$ 阵, B 为 $n \times s$ 阵, 试证明

$$\text{rank } AB \geq \text{rank } A + \text{rank } B - n.$$

证明 记 $\text{rank } A = r_1$, $\text{rank } B = r_2$, 则存在可逆阵 P_1, P_2, Q_1, Q_2 使得

$$A = P_1 \begin{pmatrix} E_{r_1} & 0 \\ 0 & 0 \end{pmatrix} Q_1, \quad B = P_2 \begin{pmatrix} E_{r_2} & 0 \\ 0 & 0 \end{pmatrix} Q_2,$$

于是

$$AB = P_1 \begin{pmatrix} E_{r_1} & 0 \\ 0 & 0 \end{pmatrix} Q_1 P_2 \begin{pmatrix} E_{r_2} & 0 \\ 0 & 0 \end{pmatrix} Q_2,$$

记 $Q_1 P_2 = C \begin{pmatrix} C_1 & C_2 \\ C_3 & C_4 \end{pmatrix}$, 其中, C_1 为 $r_1 \times r_2$ 阵, 由

$$\begin{pmatrix} E_{r_1} & 0 \\ 0 & 0 \end{pmatrix} C \begin{pmatrix} E_{r_2} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}$$

可知, $\text{rank } AB = \text{rank } C_1$. 又 C_1 是在 C 中取 r_1 行, r_2 列所得, 由于在 C 中划去一行或一列时, 其秩至多减少 1, 于是

$$\text{rank } C_1 \geq n - (n - r_1) - (n - r_2) = r_1 + r_2 - n,$$

因此 $\text{rank } AB \geq r_1 + r_2 - n$. 用矩阵的初等变换以及阶梯形矩阵和矩阵的秩的概念, 可以得到求解线性方程组的算法. 证毕.

关于线性方程组, 有以下两个重要的结论.

定理 1.21 设 A 为 $m \times n$ 阵, 则

(1) 齐次线性方程组 $Ax = 0$ 有非零解的充分必要条件是 $\text{rank } A < n$.

(2) 非齐次线性方程组 $Ax = b$ 有解的充分必要条件是 $\text{rank } A = \text{rank } (A \ b)$, 并且当 $\text{rank } A = \text{rank } (A \ b) = n$ 时有唯一解, 当 $\text{rank } A = \text{rank } (A \ b) < n$ 时有无穷多解.

注 1.12 将线性方程组的增广矩阵化为行标准形、行最简形, 本质就是将方程组中不独立的方程消掉, 并把独立的方程尽可能地化简. 方程是否有解的条件, 刻画的正是方程组是否存在矛盾的方程, 如果系数矩阵的秩小于增广矩阵的秩, 则方程组中存在矛盾的方程, 此时方程组无解, 反之则不存在矛盾的方程, 方程组有解.

1.2.4 向量组的线性相关性

为了搞清线性方程组解的结构, 需要引入向量组线性相关与无关的概念.

n 维空间中向量组的线性相关性理论是解析几何中的重要内容, 在建立笛卡尔坐标系后, 任何一个 n 维向量可以表示为一个 n 行的列矩阵, 反之任何一个 n 行的列矩阵, 也对应了一个 n 维向量. 进一步, 向量的线性运算对应了列矩阵的线性运算. 于是可以利用列矩阵来研究向量组的线性相关性. 在上述意义下, 可将向量与对应的列矩阵等同.

注 1.13 这里强调欧氏坐标系是必要的. 除笛卡尔坐标系外, 平面上建立坐标系的方式还有极坐标系等, 空间中建立坐标系的方式还有柱坐标系、球坐标系等, 在极坐标系、柱坐标系、球坐标系下, 向量的坐标虽然仍可对应为矩阵, 但向量的线性运算不对应为坐标的线性运算. 从这里可以看出, 列矩阵和向量不能无条件地等同. 何况, 矩阵还可以表示除坐标之外的对象 (如例 1.2 或例 1.3), 此时矩阵的线性运算没有实际意义, 在这种情况下, 矩阵就不宜视作向量.

定义 1.17 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 为向量组, 若存在不全为零的数 k_1, k_2, \dots, k_n , 使得

$$k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n = \mathbf{0},$$

则称向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性相关, 否则称向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关.

定理 1.22 设 $A = (\alpha_1 \alpha_2 \cdots \alpha_n)$ 为向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 组成的分块矩阵, 则下列命题等价:

- (1) 向量 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性相关;
- (2) 齐次线性方程组 $Ax = \mathbf{0}$ 有非零解;
- (3) $\text{rank } A < n$.

推论 1.23 设 $\alpha_1, \alpha_2, \dots, \alpha_n$ 为 $m \times 1$ 的列矩阵, 若 $m < n$, 则 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性相关.

证明 因 $\text{rank } A < \min\{m, n\} \leq m < n$, 故由定理 1.22 立得结论. 证毕.

定义 1.18 设 $\alpha_1, \alpha_2, \dots, \alpha_n, \beta$ 为向量, 若存在数 k_1, k_2, \dots, k_n , 使得

$$\beta = k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n,$$

则称向量 β 可由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表示.

定理 1.24 设 $A = (\alpha_1 \alpha_2 \cdots \alpha_n)$, 则下列命题等价:

- (1) 向量 β 可由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性表示;
- (2) 线性方程组 $Ax = \beta$ 有解;
- (3) $\text{rank } A = \text{rank } (A \beta)$.

注 1.14 按照上述做法, 由 n 个向量构成的向量组, 可以对应为一个 n 列的矩阵. 在本书中, 为叙述简单, 经常将这两者等同起来, 不再额外说明.

对向量组, 也可以引进秩的概念.

定义 1.19 设 $\alpha_1, \alpha_2, \dots, \alpha_r$ 为向量组 S 中的 r 个向量组, 且满足

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

(2) S 中任何向量 β , 均可由 $\alpha_1, \alpha_2, \dots, \alpha_r$ 线性表示, 则称 $\alpha_1, \alpha_2, \dots, \alpha_r$ 为向量组 A 的一个极大线性无关组, r 称为向量组 S 的秩.

注 1.15 当向量组 S 只包含有限个向量时, 此时向量组 S 可等同为一个矩阵 A , 不难证明向量组 S 的极大线性无关组必定存在, 且该向量组的秩就是矩阵 A 的秩. 当向量组 S 包含无穷多个向量时, S 虽然不能再等同为一个矩阵, 但由于本节中所指的向量均为 $n \times 1$ 型矩阵, 而此时任何 $n+1$ 个向量必定线性相关, 可以证明 S 的极大线性无关组也必定存在, 且所有的极大线性无关组包含的向量个数相同.

定义 1.20 设两个向量组 $A: \alpha_1, \alpha_2, \dots, \alpha_s$, $B: \beta_1, \beta_2, \dots, \beta_t$, 若向量组 B 的每个向量都可由向量组 A 线性表示, 则称向量组 B 可由向量组 A 线性表示. 若向量组 A 可由 B 线性表示, 同时向量组 B 可由 A 线性表示, 则称向量组 A 与向量组 B 等价, 记作 $A \sim B$.

向量组的线性相关性有以下的性质.

定理 1.25 (1) 若向量组 A 的一个子组线性相关, 则向量组 A 也线性相关;

(2) 若向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, 向量组 $\alpha_1, \alpha_2, \dots, \alpha_n, \beta$ 线性相关, 则 β 可由向量组 $\alpha_1, \alpha_2, \dots, \alpha_n$ 唯一地线性表示;

(3) 若向量组 $A: \alpha_1, \alpha_2, \dots, \alpha_s$ 可由向量组 $B: \beta_1, \beta_2, \dots, \beta_t$ 线性表示, 则 $\text{rank } A \leq \text{rank } B$.

有一类包含无穷多个向量的集合在线性代数理论中十分重要, 这就是向量空间.

定义 1.21 设 V 是一个向量构成的非空集合, 若对任意 $\alpha, \beta \in V$, 都有 $\alpha + \beta \in V, k\alpha \in V$, 其中 k 为实数, 则称 V 是一个向量空间. V 作为向量组的极大线性无关组, 称为 V 的一组基, V 的秩, 称为 V 的维数, 记作 $\dim V$.

直观上, 0 维、1 维、2 维、3 维实向量空间可以分别看作点、直线、平面和空间.

这里所提及的向量空间, 仅仅是一个狭义的概念, 更一般的向量空间, 将在线性空间与线性变换一章中介绍.

1.2.5 线性方程组的求解方法与解的结构

线性方程组的解集具有鲜明的代数结构和几何意义.

考虑齐次线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0, \\ \dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0. \end{cases} \quad (1.4)$$

记其系数矩阵为 A .

定理 1.26 记齐次线性方程组 $Ax = 0$ 全体解构成的集合为 $N(A)$. 则 $N(A)$ 是一个实向量空间, 称为线性方程组 $Ax = 0$ 的解空间, $N(A)$ 的基称为线性方程组 $Ax = 0$ 的一组基础解系.

解 对系数矩阵 A 作初等变换

$$A = \begin{pmatrix} 1 & -2 & -1 & 2 \\ 2 & -4 & 0 & 8 \\ -2 & 4 & 3 & -2 \end{pmatrix} \xrightarrow[r_3+2r_1]{r_2-2r_1} \begin{pmatrix} 1 & -2 & -1 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 1 & 2 \end{pmatrix} \xrightarrow{r_2-2r_3} \begin{pmatrix} 1 & -2 & -1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix} \xrightarrow[r_1+r_3]{r_2 \leftrightarrow r_3} \begin{pmatrix} 1 & -2 & 0 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

得

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

取 x_1, x_3 作为因变量, x_2, x_4 为自由变量, 得到线性方程组的基础解系为

$$\xi_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} -4 \\ 0 \\ -2 \\ 1 \end{pmatrix},$$

线性方程组的通解为

$$x = k_1 \xi_1 + k_2 \xi_2,$$

其中, k_1, k_2 为任意数.

1.3 相似矩阵

矩阵的相似关系与线性变换密切相关, 本节仅仅回顾与相似矩阵相关的基本概念与结论, 在矩阵的 Jordan 标准形一章中, 将详细介绍相关内容.

1.3.1 方阵的特征值与特征向量

定义 1.22 设 $A = (a_{ij})$ 为 n 阶方阵, 如果参数 λ 使得关于 x 的线性方程组 $Ax = \lambda x$ 存在非零解, 则称 λ 为方阵 A 的**特征值**, x 称为 A 的关于特征值 λ 的**特征向量**.

根据 Cramer 法则 (定理 1.15), 线性方程组 $Ax = \lambda x$ 存在非零解的充分必要条件是 $|\lambda E - A| = 0$. 记 $f(\lambda) = |\lambda E - A|$, 这是一个关于 λ 的 n 次多项式, 称为 A 的**特征多项式**. 将行列式展开, 得到

$$f(\lambda) = \lambda^n - \operatorname{tr} A \lambda^{n-1} + \cdots + (-1)^n \det A \quad (1.7)$$

根据代数学基本定理, $f(\lambda)$ 在复数范围内恰有 n 个根, 设其为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则

$$f(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n). \quad (1.8)$$

将 (1.8) 展开, 并与 (1.7) 比较右端对应项的系数, 即得下面的定理

定理 1.28 设 $\lambda_1, \lambda_2, \dots, \lambda_n$ 为 n 阶方阵 A 的特征值全体, 则

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \operatorname{tr} A, \quad \lambda_1 \lambda_2 \cdots \lambda_n = \det A.$$

关于方阵的特征值, 有下列性质.

定理 1.29 设 λ 是方阵 A 的特征值, 则

- (1) λ 也是 A^T 的特征值;
- (2) 对于任意正整数 m , λ^m 是 A^m 的特征值;
- (3) λ^{-1} 是 A^{-1} 的特征值; $\lambda^{-1}|A|$ 是 A^* 的特征值;
- (4) 设方阵 A 满足 $A^k = 0$, 则称 A 为幂零阵. 幂零阵的全体特征值均为 0;
- (5) 设方阵 A 满足 $A^2 = A$, 则称 A 为幂等阵. 幂等阵的特征值为 0 或 1.

证明 (1) 因为 λ 是 A 的特征值, 所以

$$|\lambda E - A| = 0, \text{ 而 } |\lambda E - A| = |(\lambda E - A)^T| = |\lambda E - A^T| = 0,$$

故 λ 也是 A^T 的特征值.

(2) 因为 λ 是 A 的特征值, 所以有 $Ax = \lambda x$, 将此式的两边左乘 A , 得 $A^2x = A(\lambda x) = \lambda Ax = \lambda^2x$, 再将此式两边左乘 A , 可得 $A^3x = \lambda^3x$, 如此重复进行下去, 可得 $A^mx = \lambda^mx$, 这就证明了 λ^m 是 A^m 的特征值.

(3) 因为 λ 是 A 的特征值, 所以有 $Ax = \lambda x$, 又 A 可逆, 由特征方程 $|\lambda E - A| = 0$, 知 $\lambda \neq 0$. 否则将有 $|A| = 0$, 这与 A 可逆矛盾. 于是得到 $\lambda^{-1}x = A^{-1}x$, 此式表明 λ^{-1} 是 A^{-1} 的特征值. 再由 $Ax = \lambda x$ 得 $A^*Ax = \lambda A^*x$, 于是得 $|A|x = \lambda A^*x$, 从而有 $A^*x = \lambda^{-1}|A|x$, 此式表明 $\lambda^{-1}|A|$ 是 A^* 的特征值.

(4) 设 λ 是 A 的特征值, 所以有 $Ax = \lambda x$, 两边左乘 A^{k-1} 得 $0 = A^kx = \lambda^kx$. 由于 $x \neq 0$, 所以 $\lambda^k = 0$, 故 $\lambda = 0$, 即幂零阵的特征值只能是零.

(5) 设 λ 是 A 的特征值, 则有 $Ax = \lambda x$, 两边左乘 A 得 $A^2x = \lambda Ax$, 于是得 $\lambda x = Ax = A^2x = \lambda^2x$, 故 $\lambda(\lambda - 1)x = 0$, 因此 $\lambda(\lambda - 1) = 0$, 从而 $\lambda = 0$ 或 $\lambda = 1$, 即幂等阵的特征值只能是 1 或 0. 证毕.

定理 1.30 设 $\lambda_1, \dots, \lambda_k$ 是方阵 A 的 k 个不同的特征值, x_1, \dots, x_k 分别是相应于 $\lambda_1, \dots, \lambda_k$ 的特征向量, 则 x_1, \dots, x_k 线性无关.

证明 当 $k = 1$ 时, 因 x_1 非零向量, 结论成立. 用归纳法, 假设结论对 $k = n$ 成立, 则对 $k = n + 1$ 时, 设

$$l_1x_1 + \dots + l_nx_n + l_{n+1}x_{n+1} = 0, \quad (1.9)$$

两边左乘矩阵 A , 由 $Ax_j = \lambda_jx_j$ 得

$$l_1\lambda_1x_1 + \dots + l_n\lambda_nx_n + l_{n+1}\lambda_{n+1}x_{n+1} = 0. \quad (1.10)$$

将式 (1.9) 乘 λ_{n+1} 减式 (1.10) 得

$$l_1(\lambda_{n+1} - \lambda_1)x_1 + \dots + l_n(\lambda_{n+1} - \lambda_n)x_n = 0,$$

由归纳假设 x_1, \dots, x_n 线性无关, 得 $l_j(\lambda_{n+1} - \lambda_j) = 0$ 对一切 $1 \leq j \leq n$ 成立, 而 $\lambda_{n+1} \neq \lambda_j$, 故 $l_j = 0$, 代入式 (1.9) 得到 $l_{n+1} = 0$, 从而 x_1, \dots, x_{n+1} 线性无关. 证毕.

1.3.2 矩阵的相似对角化

若一个矩阵有足够多个线性无关的特征向量, 该矩阵就存在一种非常有力的分解方式

——相似对角化.

定义 1.23 设 A, B 为 n 阶方阵, 若存在可逆矩阵 P 使得 $P^{-1}AP = B$, 则称 A 与 B 相似, 记作 $A \sim B$, 称 P 为过渡矩阵, 称运算 $P^{-1}AP$ 为对 A 作相似变换.

由矩阵相似的定义可立即得到下述结论.

定理 1.31 矩阵的相似关系满足自反性、对称性和传递性, 即矩阵的相似关系是等价关系 (定义见注 1.11).

定理 1.32 矩阵 A 与 B 相似, 则 A 与 B 有相同的特征多项式.

由相似变换的概念, 引出一个自然的问题: 是否存在相似变换, 可将矩阵变为一个简单的形式? 所谓简单的形式, 首先是对角矩阵, 其次是分块对角矩阵.

定义 1.24 若 A 与对角矩阵相似, 则称 A 可以相似对角化.

设 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, 则由 $P^{-1}AP = \Lambda$, 得 $AP = P\Lambda$. 进一步将 P 按列分块, $P = (p_1 \ p_2 \ \cdots \ p_n)$, 则

$$A(p_1 \ p_2 \ \cdots \ p_n) = (p_1 \ p_2 \ \cdots \ p_n) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix} = (\lambda_1 p_1 \ \lambda_2 p_2 \ \cdots \ \lambda_n p_n).$$

因此 $Ap_i = \lambda_i p_i$ ($i = 1, 2, \dots, n$). 也就是说, 若 A 可以对角化, 即存在过渡矩阵 P , 使得 $P^{-1}AP = \Lambda$, 则对角阵 Λ 的对角线上元素是 A 的特征值, $P = (p_1 \ p_2 \ \cdots \ p_n)$ 的列矩阵是 A 的特征向量. 注意到, 上述推理的第一步不可逆, 即由 $AP = P\Lambda$ 到 $P^{-1}AP = \Lambda$ 需要增加矩阵 P 可逆的条件, 即有下述定理.

定理 1.33 矩阵 A 可以相似对角化的充分必要条件是 A 具有 n 个线性无关的特征向量, 此时与 A 相似的对角阵的对角元素恰是 A 的全体特征值.

结合定理 1.30, 立即得到下面的结论.

定理 1.34 n 阶方阵 A 有 n 个两两不同的特征值, 则 A 可以相似对角化.

注 1.19 该定理的等价表述是, 方阵不能对角化, 则特征多项式必定有重根. 但该定理的逆命题不成立, 即可以相似对角化的矩阵, 未必特征值互不相同.

例 1.15 设矩阵 $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & x \end{pmatrix}$ 与 $\Lambda = \begin{pmatrix} 2 & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & -1 \end{pmatrix}$ 相似, 求 x, y 以及过渡矩阵 P , 使

得 $A = P\Lambda P^{-1}$.

解 因为矩阵 A 与 Λ 相似, 所以 $|\lambda E - A| = |\lambda E - \Lambda|$, 即

$$\begin{vmatrix} \lambda - 2 & 0 & 0 \\ 0 & \lambda & -1 \\ 0 & -1 & \lambda - x \end{vmatrix} = \begin{vmatrix} \lambda - 2 & 0 & 0 \\ 0 & \lambda - y & 0 \\ 0 & 0 & \lambda + 1 \end{vmatrix}.$$

或 $(\lambda - 2)(\lambda^2 - x\lambda - 1) = (\lambda - 2)[\lambda^2 + (1 - y)\lambda - y]$. 比较等式两端得 $x = 0, y = 1$. 此时

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

为求 P , 设 $P = (p_1 \ p_2 \ p_3)$, 由 $AP = P\Lambda$ 得

$$(2E - A)p_1 = 0, (E - A)p_2 = 0, (-E - A)p_3 = 0.$$

分别解上述齐次方程组得 $p_1 = (1 \ 0 \ 0)^T$, $p_2 = (0 \ 1 \ 1)^T$, $p_3 = (0 \ 1 \ -1)^T$, 故

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

注 1.20 对不能相似对角化的矩阵, 在矩阵的 Jordan 标准形一章中将证明, 这类矩阵一定相似于某种分块对角阵.

注 1.21 由于矩阵特征值的计算涉及到多项式的求根, 只有在非常特殊的情形下, 多项式的根才能精确计算. 一般情形矩阵特征值的计算, 是数值代数的重要内容之一.

1.3.3 实对称矩阵与正定矩阵

实对称矩阵一定可以相似对角化, 且有下列的结论.

定理 1.35 若 A 为对称矩阵, 则存在矩阵 P 与 Λ , 使得 $A = P\Lambda P^T$, 其中 $PP^T = E$, Λ 为实对角阵.

该定理的证明将在内积空间一章中给出.

注 1.22 满足定理 1.35 中的条件 $PP^T = E$ 的矩阵 P 称为正交矩阵, 关于正交矩阵的更多内容, 将在内积空间一章中给出.

注 1.23 若存在可逆矩阵 P , 使得 $A = PBP^T$, 则称矩阵 A 与 B 合同, 易见矩阵的合同关系也是一种等价关系. 定理 1.35 指出, 实对称矩阵与对角矩阵既相似又合同.

正定矩阵是一类特殊的实对称矩阵, 与空间的度量有密切关系.

定义 1.25 设 A 为 n 阶实对称阵. 对任意的 n 维实非零列向量 x ,

- (1) 若 $x^T Ax > 0$, 则称 A 为正定矩阵;
- (2) 若 $x^T Ax \geq 0$, 则称 A 为半正定矩阵;
- (3) 若 $x^T Ax < 0$, 则称 A 为负定矩阵;
- (4) 若 $x^T Ax \leq 0$, 则称 A 为半负定矩阵.

由定义可直接推得下述结论.

定理 1.36 若 A 为 (半) 正定矩阵, 则 $-A$ 必为 (半) 负定矩阵.

正定 (半正定) 矩阵由以下性质.

定理 1.37 对实对称矩阵 A , 下列条件等价:

- (1) A 为正定 (半正定) 矩阵;

- (2) A 的所有特征值均为正 (非负) 数;
 (3) 存在可逆 (无须可逆) 矩阵 S , 使得 $A = S^T S$;
 (4) 存在正定 (半正定) 矩阵 Q , 使得 $A = Q^2$;
 (5) A 的各阶顺序主子式均为正 (非负) 数.

定理 (1)-(4) 的等价性可以由定理 1.35 得到, 条件 (5) 与 (1) 的等价性称为 Hurwitz 定理, 可利用归纳法证明.

例 1.16 设 A 为正定矩阵, 则 A^{-1} 和 A^* 都是正定矩阵.

证明 设正定阵 A 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n > 0$, A 的行列式 $|A| > 0$, 则 A^{-1} 的特征值为 $\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_n^{-1} > 0$, A^* 的特征值为 $\lambda_1^{-1}|A|, \lambda_2^{-1}|A|, \dots, \lambda_n^{-1}|A| > 0$, 故 A^{-1} 和 A^* 都是正定矩阵. 证毕.

对一般的实对称矩阵, 其正、负特征值的个数分别称为该矩阵的**正惯性指数**和**负惯性指数**, 矩阵的惯性指数, 是合同变换下的不变量, 即下面的结论成立.

定理 1.38 若实对称矩阵 A 与 B 合同, 充分必要条件是 A 与 B 有相同的正惯性指数和负惯性指数.

证明 充分性. 由于矩阵的合同关系是等价关系, 从而只需证明任何实对称矩阵 A 必合同于对角阵 $\text{diag}(E_p, E_q, \mathbf{0})$, 这里 p, q 分别为矩阵 A 的正、负惯性指数. 事实上, 由定理 1.35, 矩阵 A 合同于

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p, -\lambda_{p+1}, \dots, -\lambda_{p+q}, \mathbf{0}),$$

其中 $\lambda_1, \dots, \lambda_{p+q}$ 均为正数. 取

$$P = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{p+q}}, 1, \dots, 1),$$

则

$$\Lambda = P \text{diag}(E_p, E_q, \mathbf{0}) P^T.$$

必要性的证明将在内积空间一章给出.

习题 1

1. 计算下列乘积

$$(1) \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}; \quad (2) \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix};$$

$$(3) \begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}; \quad (4) \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix}.$$

2. 设矩阵 $A = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$, 求 A^k .

3. 如果矩阵的背景如例 1.1、例 1.2 或例 1.3, 则矩阵的加法、数乘、乘法运算是否有相应的意义?

4. 设 A 为 n 阶方阵, α 为 n 维列矩阵, λ 是数, 并且 $A\alpha = \lambda\alpha$, 证明:

(1) $A^2\alpha = \lambda^2\alpha$;

(2) $(A^5 + 2A^3 - 3E)\alpha = (\lambda^5 + 2\lambda^3 - 3)\alpha$.

5. 设 α 为 n 维列矩阵, 证明: $\alpha^T\alpha = 0$ 当且仅当 $\alpha = \mathbf{0}$.

6. 求下列矩阵行列式和逆矩阵

(1) $\begin{pmatrix} 3 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 5 \end{pmatrix}$; (2) $\begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 3 \\ 5 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 \end{pmatrix}$.

7. 在 \mathbf{R}^3 上, 定义向量 $\mathbf{x} = (x_1 \ x_2 \ x_3)^T$, $\mathbf{y} = (y_1 \ y_2 \ y_3)^T$ 的外积如下

$$\mathbf{x} \times \mathbf{y} = \begin{pmatrix} \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} & \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix} & \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \end{pmatrix}^T.$$

(1) 证明: $\mathbf{x} \times \mathbf{y} = -\mathbf{y} \times \mathbf{x}$.

(2) 证明: 对 $k, l \in \mathbf{R}$, $(k\mathbf{x} + l\mathbf{y}) \times \mathbf{z} = k(\mathbf{x} \times \mathbf{z}) + l(\mathbf{y} \times \mathbf{z})$.

(3) 证明: $(\mathbf{x} \times \mathbf{y}) \times \mathbf{z} + (\mathbf{y} \times \mathbf{z}) \times \mathbf{x} + (\mathbf{z} \times \mathbf{x}) \times \mathbf{y} = \mathbf{0}$.

8. 设矩阵 $A = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 2 & 1 & -2 & -1 \\ 1 & -1 & -4 & -2 \end{pmatrix}$, 求可逆矩阵 P , 使得 PA 是行最简形.

9. 问 λ 取什么值时非齐次线性方程组

$$\begin{cases} -2x_1 + x_2 + x_3 = -2 \\ x_1 - 2x_2 + x_3 = \lambda \\ x_1 + x_2 - 2x_3 = \lambda^2 \end{cases}$$

有解? 并求其通解.

10. 设 A 为 $m \times n$ 型矩阵且 $\text{rank } A = n$. 证明: 方程 $BX = \mathbf{0}$ 与 $ABX = \mathbf{0}$ 的解相同.

11. 设向量组 A 线性无关, 向量组 $B: \mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_r$ 能由向量组 $A: \mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_s$ 线性表示

$$(\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_r) = (\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_s)K,$$

其中

$$\mathbf{K} = \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1r} \\ k_{21} & k_{22} & \cdots & k_{2r} \\ \vdots & \vdots & & \vdots \\ k_{s1} & k_{s2} & \cdots & k_{sr} \end{pmatrix}.$$

证明向量组 \mathbf{B} 线性无关的充要条件是 $\text{rank } \mathbf{K} = r$.

12. 设向量组 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ 线性无关, 判断向量组 $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ 是否线性相关.

(1) $\mathbf{b}_1 = \mathbf{a}_1 + \mathbf{a}_2, \mathbf{b}_2 = \mathbf{a}_2 + \mathbf{a}_3, \mathbf{b}_3 = \mathbf{a}_3 + \mathbf{a}_1$;

(2) $\mathbf{b}_1 = \mathbf{a}_1 + 2\mathbf{a}_2, \mathbf{b}_2 = 2\mathbf{a}_2 - 3\mathbf{a}_3, \mathbf{b}_3 = 3\mathbf{a}_3 + \mathbf{a}_1$;

(3) $\mathbf{b}_1 = \mathbf{a}_1 + 3\mathbf{a}_2, \mathbf{b}_2 = \mathbf{a}_2 + 3\mathbf{a}_3, \mathbf{b}_3 = \mathbf{a}_1 - \mathbf{a}_3$.

13. 设向量组 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ 线性无关, 而向量组 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ 线性相关, 并且向量组 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_5$ 线性无关, 求向量组 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4 + \mathbf{a}_5$ 的一个极大线性无关组.

14. 设 $\boldsymbol{\eta}$ 是非齐次线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的一个解, $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_{n-r}$ 是 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 对应的齐次线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{0}$ 的基础解系, 证明向量组 $\boldsymbol{\xi}_1 + \boldsymbol{\eta}, \boldsymbol{\xi}_2 + \boldsymbol{\eta}, \dots, \boldsymbol{\xi}_{n-r} + \boldsymbol{\eta}, \boldsymbol{\eta}$ 线性无关.

15. 设 \mathbf{A} 是 $m \times n$ 矩阵, $\text{rank } \mathbf{A} = r$, 非齐次线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 有解, 并且 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_{n-r+1}$ 是的 $n - r + 1$ 个线性无关的解, 证明 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 的任意解 \mathbf{X} 可表示为 $\mathbf{X} = k_1\boldsymbol{\eta}_1 + k_2\boldsymbol{\eta}_2 + \dots + k_{n-r+1}\boldsymbol{\eta}_{n-r+1}$, 其中 $k_1 + k_2 + \dots + k_{n-r+1} = 1$.

16. 设 n 阶矩阵 $\mathbf{A} = (\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2 \cdots \boldsymbol{\alpha}_n)$, 证明 \mathbf{A} 是正交矩阵当且仅当 $\boldsymbol{\alpha}_i^T \boldsymbol{\alpha}_j = \delta_{ij}$.

17. 判断下列矩阵是否为正交矩阵:

$$(1) \frac{1}{3} \begin{pmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix}; \quad (2) \begin{pmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

18. 设 $n \times 1$ 的矩阵 $\boldsymbol{\alpha} = (1 \ 1 \ \cdots \ 1)^T$, 矩阵 $\mathbf{A} = \boldsymbol{\alpha}\boldsymbol{\alpha}^T$, 求 \mathbf{A} 的特征值以及 n 个线性无关的特征向量.

19. 设 \mathbf{A} 是 3 阶矩阵, 并且 $|\mathbf{A} + 3\mathbf{E}| = |\mathbf{A} - \mathbf{E}| = |3\mathbf{A} - 2\mathbf{E}| = 0$, 求 \mathbf{A} 的特征值以及 n 个线性无关的特征向量.

20. 设 $\lambda_1 = \lambda_2 = 2, \lambda_3 = -1$ 是 3 阶矩阵 \mathbf{A} 的特征值, 并且与这 3 个特征值对应的特征向量依次为

$$\mathbf{p}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{p}_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \mathbf{p}_3 = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}.$$

求矩阵 A .

21. 设 A, B 都是 n 阶正定矩阵.

(1) 证明 $A + B$ 是正定矩阵;

(2) 若 $AB = BA$, 证明 AB 是正定矩阵.

22. 设 A 为 n 阶实对称矩阵,

(1) 证明存在实数 c , 使得当 $b \geq c$ 时, $bE + A$ 是正定矩阵;

(2) 证明存在实数 C_1, C_2 , 使得对任意的 n 维实向量 x , 都有

$$C_1 x^T x \leq x^T A x \leq C_2 x^T x.$$

第2章 多项式

多项式是数学中的一个重要工具，在线性代数中有广泛应用。本章将介绍系数在数域上的一元多项式的基本理论，为后续章节的应用做准备。

2.1 数域

在讨论多项式前，首先介绍数域的概念。

定义 2.1 设 F 是一个数集，若该集合关于加、减、乘、除四则运算都是封闭的，即该集合中任意两个数作合法的四则运算，其结果仍然在该集合中，则该数集 F 称为一个数域。

例 2.1 只有 0 一个元素的单元素集合 $\{0\}$ 是一个数域。注意，不能以 $0 \div 0$ 没有意义，而认为该集合对除法不封闭，因为 0 不可以做除数，所以 $0 \div 0$ 不是一个“合法”的四则运算。

例 2.2 有理数集 Q 、实数集 R 和复数集 C 等等都是数域，但整数集 Z 关于除法不是封闭的，不是数域。

除了上述数域之外，数域的例子还有很多。

例 2.3 集合 $Q(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in R\}$ 是数域。

证明 容易验证，集合 $Q(\sqrt{2})$ 关于加、减、乘三种运算封闭，以下证明关于除法运算的封闭性。任取 $a_1 + b_1\sqrt{2}, a_2 + b_2\sqrt{2} \in Q(\sqrt{2})$ ， a_2, b_2 不全为零，有 $a_2^2 - 2b_2^2 \neq 0$ （否则与 $\sqrt{2}$ 为无理数矛盾），于是

$$\frac{a_1 + b_1\sqrt{2}}{a_2 + b_2\sqrt{2}} = \frac{(a_1 + b_1\sqrt{2})(a_2 - b_2\sqrt{2})}{(a_2 + b_2\sqrt{2})(a_2 - b_2\sqrt{2})} = \frac{a_1a_2 - 2b_1b_2}{a_2^2 - 2b_2^2} + \frac{a_2b_1 - a_1b_2}{a_2^2 - 2b_2^2}\sqrt{2} \in Q(\sqrt{2}).$$

从而 $Q(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in R\}$ 关于四则运算是封闭的，从而是数域。证毕。

注 2.1 不难发现，将上面例子中的 $\sqrt{2}$ 换成其他非完全平方数的平方根，就可得到不同的非平凡的数域。

例 2.4 设 a 和 b 为整数， m 为大于 1 的正整数，若 $a - b$ 可以被 m 整除，则称 a 和 b 关于 m 同余（即被 m 除余数相同），记作 $a \equiv b \pmod{m}$ 。特别，记 $Z_m = \{0, 1, \dots, m-1\}$ ，则对任意 $m > 1$ 和 a ，总存在 $\bar{a} \in Z_m$ ，使得 $a \equiv \bar{a} \pmod{m}$ ，于是 Z_m 关于整数的加法、乘法再取“mod m ”是封闭的，将这一运算称为 Z_m 上的加法和乘法。按照这一定义， $Z_3 = \{0, 1, 2\}$

上的加法、乘法可以定义逆运算（即减法、除法）从而 $\mathbf{Z}_3 = \{0, 1, 2\}$ 是数域。

事实上，由于

$$1 + 2 \equiv 0 \pmod{3}, \quad 2 \times 2 \equiv 1 \pmod{3},$$

从而在 $\mathbf{Z}_3 = \{0, 1, 2\}$ 上， $(-1) \equiv 2$ ， $(-2) \equiv 1$ ， $2^{-1} \equiv 2$ ，于是可以定义减法、除法。

注 2.2 在数域中，四则运算的定义可以与通常意义下的四则运算不同，但加法与减法，乘法与除法必须互为逆运算，且加法、乘法必须满足基本的交换律、结合律和分配律（上例中的加法、乘法满足）。

注 2.3 可以证明，对一切素数 p ， \mathbf{Z}_p 关于上述加法、乘法都存在逆运算，因此都是数域。若 m 非素数，如此定义的加法仍然有逆运算，但并非对所有的非零元素都存在关于乘法运算的倒数，从而乘法不存在逆运算，相应的 \mathbf{Z}_m 就不再是数域。

本书中涉及的数域，大多是复数域和实数域，个别章节提及有理数域。

2.2 一元多项式

本节将介绍数域上的一元多项式的基本概念。下文中， \mathbf{F} 为数域，本节涉及的结论，对一切数域都成立。为便于理解，读者可不失一般性地将 \mathbf{F} 取为有理数域、实数域或复数域。

定义 2.2 设 n 是一个非负整数， x 是一个文字符号， $a_0, a_1, \dots, a_n \in \mathbf{F}$ ， $a_n \neq 0$ ，称

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 \quad (2.1)$$

为数域 \mathbf{F} 上的一个一元多项式， n 称为多项式 $f(x)$ 的次数，记为 $\deg f(x)$ 。 $a_k x^k$ ($1 \leq k \leq n$) 称为这个多项式的 k 次项， a_k 称为 k 次项系数， a_0 称为多项式的 0 次项或常数项。 $a_n x^n$ 与 a_n 分别称为首项和首项系数。特别，0 也是一个多项式，称为零多项式，零多项式的次数规定为 $-\infty$ 。数域 \mathbf{F} 上全体一元多项式构成的集合，记作 $\mathbf{F}[x]$ 。

定义 2.3 对多项式 $f(x)$ 和 $g(x)$ ，若同次项的系数都相等，则称多项式 $f(x)$ 和 $g(x)$ 相等，记为 $f(x) = g(x)$ 。

多项式可以定义加法和乘法运算。为了叙述简洁，用求和号来表示多项式。记

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 = \sum_{k=0}^n a_k x^k,$$

$$g(x) = b_m x^m + b_{m-1} x^{m-1} + \dots + b_0 = \sum_{k=0}^m b_k x^k,$$

并不失一般性地假设 $n \geq m$ ，且约定 $b_n = b_{n-1} = \dots = b_{m+1} = 0$ 。

定义 2.4 多项式 $f(x)$ 和 $g(x)$ 的和与差定义为

$$f(x) \pm g(x) = \sum_{k=1}^n (a_k \pm b_k) x^k$$

特别， $f(x) + 0 = f(x)$ 。

关于多项式加法，下面的命题是显然的。

命题 2.1 若记 $-g(x) = (-b_m)x^m + (-b_{m-1})x^{m-1} + \dots + (-b_0)$ ，则 $f(x) - g(x) = f(x) + (-g(x))$ 。

从这一意义上说, 多项式的减法是加法的逆运算.

定义 2.5 多项式 $f(x)$ 和 $g(x)$ 的乘积定义为

$$f(x)g(x) = \sum_{k=0}^{m+n} c_k x^k,$$

这里 s 项的系数

$$c_k = \sum_{j=0}^k a_j b_{k-j} = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0.$$

特别, $f(x) \cdot 0 = 0$.

易见, 多项式的次数关于加法和乘法满足以下命题.

定理 2.2 约定 0 多项式的次数为 $-\infty$, 且 $-\infty$ 与任何数的和均为 $-\infty$, 则

$$\begin{aligned} \deg(f(x) \pm g(x)) &\leq \max\{\deg f(x), \deg g(x)\}, \\ \deg(f(x)g(x)) &= \deg f(x) + \deg g(x). \end{aligned}$$

与数的运算类似, 多项式的运算也具有下面的一些运算律.

命题 2.3 多项式的运算律:

- (1) 加法交换律: $f(x) + g(x) = g(x) + f(x)$;
- (2) 加法结合律: $(f(x) + g(x)) + h(x) = f(x) + (g(x) + h(x))$;
- (3) 乘法交换律: $f(x)g(x) = g(x)f(x)$;
- (4) 乘法结合律: $(f(x)g(x))h(x) = f(x)(g(x)h(x))$;
- (5) 分配律: $f(x)(g(x) + h(x)) = f(x)g(x) + f(x)h(x)$.

这些运算律都可以由定义直接证明, 这里略去.

例 2.5 证明对多项式加法和乘法, 满足消去律, 即

- (1) 若 $f(x) + g(x) = f(x) + h(x)$, 则 $g(x) = h(x)$;
- (2) 若 $f(x)g(x) = f(x)h(x)$, 且 $f(x) \neq 0$, 则 $g(x) = h(x)$.

证明 (1) 在等式两边同时加上 $-f(x)$, 得

$$-f(x) + f(x) + g(x) = -f(x) + f(x) + h(x),$$

由结合律得

$$(-f(x) + f(x)) + g(x) = (-f(x) + f(x)) + h(x),$$

从而 $g(x) = h(x)$.

(2) 在 $f(x)g(x) = f(x)h(x)$ 两边减去 $f(x)h(x)$, 并利用分配律, 得

$$0 = f(x)g(x) - f(x)h(x) = f(x)(g(x) - h(x)).$$

根据多项式次数与乘法的关系,

$$-\infty = \deg f(x) + \deg(g(x) - h(x)),$$

得 $\deg(g(x) - h(x)) = -\infty$, 即 $g(x) - h(x) = 0$, 再由 (1) 得到, $g(x) = h(x)$. 证毕.

对于两个多项式 $f(x)$ 和 $g(x)$, 通常不能定义二者的除法, 即通常不存在多项式 $h(x)$, 使得 $f(x) = g(x)h(x)$, 但可以类似整数一样定义带余除法.

定理 2.4 (带余除法) 设 $f(x), g(x) \in \mathbf{F}[x]$, 并且 $g(x) \neq 0$, 则存在唯一的 $q(x), r(x) \in \mathbf{F}[x]$, 使得

$$f(x) = q(x)g(x) + r(x), \quad (2.2)$$

其中 $\deg r(x) < \deg g(x)$, $q(x)$ 和 $r(x)$ 分别称为 $g(x)$ 除 $f(x)$ 所得的商式和余式, 并将这种算法称为带余除法.

先通过一个例子, 看一看带余除法的计算.

例 2.6 设 $f(x) = 4x^4 - 2x^3 - 6x^2 + 5x + 9$, $g(x) = 2x^2 - 5x + 4$. 求 $g(x)$ 除 $f(x)$ 的商式 $q(x)$ 和余式 $r(x)$.

解 由

$$\begin{array}{r} 2x^2 - 5x + 4 \overline{) 4x^4 - 2x^3 - 6x^2 + 5x + 9} \\ \underline{4x^4 - 10x^3 + 8x^2} \\ 8x^3 - 14x^2 + 5x + 9 \\ \underline{8x^3 - 20x^2 + 16x} \\ 6x^2 - 11x + 9 \\ \underline{6x^2 - 15x + 12} \\ 4x - 3 \end{array}$$

得出 $q(x) = 2x^2 + 4x + 3$ 和 $r(x) = 4x - 3$.

这一做法与计算整数除法的竖式笔算方法本质相同, 称为长除法. 两种除法唯一的差别是: 在整数的带余除法中, 要求余数必须小于除数, 在多项式的带余除法中, 则要求余式的次数必须小于除式的次数. 在后续内容中, 多项式的次数, 还将多次承担类似的角色.

根据长除法, 不难得到定理 2.4 的证明.

证明 首先证明 $q(x)$ 和 $r(x)$ 的存在性. 当 $\deg f(x) < \deg g(x)$ 时, 取 $q(x) = 0$, $r(x) = f(x)$ 即可. 当 $\deg f(x) \geq \deg g(x)$ 时, 对次数的差 $\deg f(x) - \deg g(x)$ 用归纳法. 假设当 $\deg f(x) - \deg g(x) \leq k$ 时, $q(x)$ 和 $r(x)$ 存在, 那么当 $\deg f(x) - \deg g(x) = k + 1$ 时, 设

$$f(x) = \sum_{k=0}^n a_k x^k, \quad g(x) = \sum_{k=0}^m b_k x^k,$$

取 $f_1(x) = f(x) - \frac{a_n}{b_m} x^{n-m} g(x)$, 则

$$\deg f_1(x) - \deg g(x) \leq \deg f(x) - 1 - \deg g(x) = k,$$

从而由归纳法假设可知, 存在 $q_1(x), r(x)$ 使得

$$f_1(x) = q_1(x)g(x) + r(x),$$

这里 $\deg r(x) < \deg q(x)$. 于是

$$f(x) = f_1(x) + \frac{a_n}{b_m} x^{n-m} g(x) = (\frac{a_n}{b_m} x^{n-m} + q_1(x))g(x) + r(x),$$

记 $q(x) = \frac{a_n}{b_m} x^{n-m} + q_1(x)$, 即有 $q(x), r(x)$ 在 $\deg f(x) - \deg g(x) = k + 1$ 时存在, 归纳法完成.

再证 $q(x)$ 和 $r(x)$ 的唯一性. 设另有多项式 $\tilde{q}(x), \tilde{r}(x)$ 使

$$f(x) = \tilde{q}(x)g(x) + \tilde{r}(x),$$

其中 $\deg \tilde{r}(x) < \deg g(x)$ ，于是

$$q(x)g(x) + r(x) = \tilde{q}(x)g(x) + \tilde{r}(x),$$

即

$$(q(x) - \tilde{q}(x))g(x) = \tilde{r}(x) - r(x),$$

然而 $\deg g(x) > \deg(\tilde{r}(x) - r(x))$ ，所以只能有 $\deg(q(x) - \tilde{q}(x)) = -\infty$ ，故 $q(x) = \tilde{q}(x)$ ，进而 $r(x) = \tilde{r}(x)$ 。证毕。

从这一证明可以看出，要求多项式的系数取值的集合为数域是十分必要的，否则“取 $f_1(x) = f(x) - \frac{a_n}{b_m}x^{n-m}g(x)$ ”这一步将不成立，因为 $\frac{a_n}{b_m}$ 未必在系数取值的集合中。

在带余除法中，余式为零是一种特殊的情形，类似整数的整除关系，有下面的定义。

定义 2.6 设 $f(x), g(x) \in \mathbf{F}[x]$ ，若存在 $q(x) \in \mathbf{F}[x]$ 使得 $f(x) = q(x)g(x)$ ，则称 $g(x)$ 整除 $f(x)$ ，记作 $g(x) \mid f(x)$ ，并称 $g(x)$ 为 $f(x)$ 的因式， $f(x)$ 为 $g(x)$ 的倍式。

因任意一个多项式 $f(x)$ 以及其非零常数倍都一定整除 $f(x)$ 自身（即 $c \cdot f(x) \mid f(x)$ ），零次多项式整除任意一个多项式，于是零次多项式和 $c \cdot f(x)$ 称为 $f(x)$ 的平凡因式。又任意一个非零多项式 $f(x)$ 一定整除零多项式（即 $f(x) \mid 0$ ），于是零多项式和 $cf(x)$ 称为 $f(x)$ 的平凡倍式。

关于整除，有以下性质，证明是容易的，此处略去。

命题 2.5 设 $f(x), g(x), h(x) \in \mathbf{F}[x]$ ，则

- (1) 若 $f(x) \mid g(x)$ ，并且 $g(x) \mid f(x)$ ，则 $f(x) = c \cdot g(x)$ ，其中 c 为非零常数；
- (2) 若 $f(x) \mid g(x)$ ，并且 $g(x) \mid h(x)$ ，则 $f(x) \mid h(x)$ ；
- (3) 若 $f(x) \mid g_j(x)$ ， $u_j(x) \in \mathbf{F}[x]$ ($j = 1, 2$)，则 $f(x) \mid (u_1(x)g_1(x) + u_2(x)g_2(x))$ 。

当除式的次数为 1 时，可以不通过长除法而直接计算余式，即有如下定理。

定理 2.6 (余式定理) 设 $f(x) \in \mathbf{F}[x]$ ， $c \in \mathbf{F}$ ，则 $f(x)$ 除以 $(x - c)$ 的余式为 $f(c)$ 。

证明 设 $f(x) = q(x)(x - c) + r(x)$ ，两边令 $x = c$ 得到 $r(c) = f(c)$ 。而 $\deg r(x) = 0$ ，于是 $r(x)$ 不依赖于 x ，从而 $r(x) = f(c)$ 。证毕。

推论 2.7 设 $f(x) \in \mathbf{F}[x]$ ， $c \in \mathbf{F}$ ，则 $(x - c) \mid f(x)$ ，当且仅当 $f(c) = 0$ 。此时 c 称为多项式 $f(x)$ 的根或零点。

2.3 最大公因式和最小公倍数

类似两个整数有最大公因数和最小公倍数，两个多项式也有最大公因式和最小公倍式，而且相应定理的表述和证明，也都非常相近。

首先介绍两个多项式的最大公因式。

定义 2.7 设 $f(x), g(x) \in \mathbf{F}[x]$ ，若 $d(x) \in \mathbf{F}[x]$ 既是 $f(x)$ 的因式，又是 $g(x)$ 的因式，则称 $d(x)$ 为 $f(x)$ 和 $g(x)$ 的一个公因式。进一步，若 $d(x)$ 是 $f(x)$ 和 $g(x)$ 的公因式，并且 $f(x)$ 和 $g(x)$ 的任意公因式都是 $d(x)$ 的因式，则称 $d(x)$ 为 $f(x)$ 和 $g(x)$ 的一个最大公因式。

从定义可以直接看出, 最大公因式是次数最大的公因式.

在给出这一定义后, 首先需要考虑对任意两个多项式, 它们的最大公因式是否存在. 为此先证明以下引理.

引理 2.8 设 $f(x), g(x), q(x), r(x) \in \mathbf{F}[x]$, 并且

$$f(x) = q(x)g(x) + r(x), \quad (2.3)$$

则 $f(x)$ 和 $g(x)$ 的公因式与 $g(x)$ 和 $r(x)$ 的公因式完全相同.

证明 任取 $g(x)$ 与 $r(x)$ 的公因式 $\varphi(x)$, 由 $\varphi(x) \mid g(x)$ 且 $\varphi(x) \mid r(x)$, 则 $\varphi(x) \mid (q(x)g(x) + r(x))$, 即 $\varphi(x) \mid f(x)$. 从而 $g(x)$ 与 $r(x)$ 的公因式也是 $f(x)$ 和 $g(x)$ 的公因式.

反之, 对任何 $f(x)$ 和 $g(x)$ 的公因式 $\psi(x)$, 因 $\psi(x) \mid f(x)$ 且 $\psi(x) \mid g(x)$, 从而 $\psi(x) \mid (f(x) - q(x)g(x))$, 即 $\psi(x) \mid r(x)$, 故 $f(x)$ 和 $g(x)$ 的公因式也是 $g(x)$ 与 $r(x)$ 的公因式. 证毕.

以下证明最大公因式的存在性.

定理 2.9 设 $f(x), g(x) \in \mathbf{F}[x]$, 则 $f(x)$ 和 $g(x)$ 的最大公因式 $d(x)$ 存在, 且存在 $u(x), v(x) \in \mathbf{F}[x]$, 使得

$$d(x) = u(x)f(x) + v(x)g(x). \quad (2.4)$$

证明 不妨设 $g(x) \neq 0$. 由带余除法, 用 $g(x)$ 除 $f(x)$, 得到商 $q_1(x)$, 余式 $r_1(x)$; 若 $r_1(x) \neq 0$, 就再用 $r_1(x)$ 除 $g(x)$, 得到商 $q_2(x)$, 余式 $r_2(x)$; 又若 $r_2(x) \neq 0$, 就用 $r_2(x)$ 除 $r_1(x)$, 得到商 $q_3(x)$, 余式 $r_3(x)$; 如此依次不断相除下去. 显然, 所得余式的次数就依次不断降低, 即 $\deg g(x) > \deg r_1(x) > \deg r_2(x) > \cdots$. 这样在经过有限次相除之后, 必然有余式为零. 于是得到以下等式

$$\begin{aligned} f(x) &= q_1(x)g(x) + r_1(x), \\ g(x) &= q_2(x)r_1(x) + r_2(x), \\ r_1(x) &= q_3(x)r_2(x) + r_3(x), \\ &\dots\dots\dots \\ r_{s-2}(x) &= q_s(x)r_{s-1}(x) + r_s(x), \\ r_{s-1}(x) &= q_{s+1}(x)r_s(x), \end{aligned}$$

其中 $r_s(x)$ 是 $r_s(x)$ 与 $r_{s-1}(x)$ 的一个最大公因式, 因此 $r_s(x)$ 就是 $r_{s-1}(x)$ 与 $r_{s-2}(x)$ 的一个最大公因式, 这样逐步推上去, 就有 $r_s(x)$ 是 $f(x)$ 和 $g(x)$ 的一个最大公因式.

由上面等式列的倒数第二个可得,

$$r_s(x) = r_{s-2}(x) - q_s(x)r_{s-1}(x),$$

再由倒数第三式可得

$$r_{s-1}(x) = r_{s-3}(x) - q_{s-1}(x)r_{s-2}(x),$$

代入上式消去 $r_{s-1}(x)$ 后就得到

$$r_s(x) = (1 + q_s(x)q_{s-1}(x))r_{s-2}(x) - q_s(x)r_{s-3}(x).$$

用同样的方法逐一消去 $r_{s-k}(x)$ ($k = 2, 3, \dots, s-1$), 即可得到式 (2.4). 证毕.

注 2.4 两个多项式的最大公因式不唯一. 显然, 若 $d(x)$ 是 $f(x)$ 和 $g(x)$ 的最大公因式, 则对任何非零常数 $c \in \mathbf{F}$, $c \cdot d(x)$ 也是 $f(x)$ 和 $g(x)$ 的最大公因式. 反之, 若 $d(x)$ 和 $d_1(x)$ 都

是 $f(x)$ 和 $g(x)$ 的最大公因式, 依据定义, 应有 $d(x) \mid d_1(x)$ 和 $d_1(x) \mid d(x)$, 于是存在非零常数 $c \in \mathbf{F}$, 使得 $d(x) = c \cdot d_1(x)$. 从而可在 $f(x)$ 和 $g(x)$ 的最大公因式中, 可选取首项系数为 1 的多项式, 记为 $\gcd(f(x), g(x))$.

在上述定理的证明过程中, 反复利用带余除法计算两个多项式的最大公因式的算法称为 **辗转相除法**. 这种算法与计算两个数的最大公因数的方法完全相同. 下面是使用这种算法的一个例子.

例 2.7 求多项式 $f(x) = 4x^4 - 2x^3 - 8x^2 + 9x - 3$ 与 $g(x) = 2x^3 - x^2 - 5x + 4$ 求的最大公因式, 及满足等式 (2.4) 的多项式 $u(x)$ 和 $v(x)$.

解 对 $f(x)$ 和 $g(x)$ 做带余除法,

$$\begin{array}{r} 2x \\ 2x^3 - x^2 - 5x + 4 \overline{) 4x^4 - 2x^3 - 8x^2 + 9x - 3} \\ \underline{4x^4 - 2x^3 - 10x^2 + 8x} \\ 2x^2 + x - 3 \end{array}$$

故 $f(x) = 2xg(x) + r_1(x)$, $r_1(x) = 2x^2 + x - 3$. 再对 $g(x)$ 与 $r_1(x)$ 做带余除法,

$$\begin{array}{r} x-1 \\ 2x^2 + x - 3 \overline{) 2x^3 - x^2 - 5x + 4} \\ \underline{2x^3 + x^2 - 3x} \\ -2x^2 - 2x + 4 \\ \underline{-2x^2 - x + 3} \\ -x + 1 \end{array}$$

得 $g(x) = (x-1)r_1(x) + r_2(x)$, $r_2(x) = -x + 1$. 对 $r_1(x)$ 与 $r_2(x)$ 做带余除法,

$$\begin{array}{r} -2x-3 \\ -x+1 \overline{) 2x^2 + x - 3} \\ \underline{2x^2 - 2x} \\ 3x - 3 \\ \underline{3x - 3} \\ 0 \end{array}$$

得 $r_1(x) = (-2x-3)r_2(x)$. 由此得

$$r_2(x) = g(x) - (x-1)r_1(x) = g(x) - (x-1)[f(x) - 2xg(x)],$$

即

$$-r_2(x) = (x-1)f(x) - (2x^2 - 2x + 1)g(x),$$

故 $\gcd(f(x), g(x)) = x-1$, 且

$$x-1 = u(x)f(x) + v(x)g(x),$$

其中 $u(x) = x-1$, $v(x) = -(2x^2 - 2x + 1)$.

定义 2.8 设 $f(x), g(x) \in \mathbf{F}[x]$, 若 $\gcd(f(x), g(x)) = 1$, 则称 $f(x)$ 和 $g(x)$ **互素**.

显然, 如果两个多项式互素, 那么它们除去零次多项式以外没有其他的公因式, 反之亦

然.

定理 2.10 $f(x), g(x) \in \mathbf{F}[x]$ 互素的充分必要条件是存在 $u(x), v(x) \in \mathbf{F}[x]$, 使得

$$u(x)f(x) + v(x)g(x) = 1.$$

证明 必要性是定理 2.9 的直接推论. 现在来证明充分性, 设存在 $u(x)$ 和 $v(x)$ 使得 $u(x)f(x) + v(x)g(x) = 1$, 并且 $\varphi(x)$ 是 $f(x)$ 和 $g(x)$ 的一个最大公因式. 于是就有 $\varphi(x) \mid f(x)$ 和 $\varphi(x) \mid g(x)$, 从而 $\varphi(x) \mid 1$, 即 $f(x)$ 和 $g(x)$ 互素. 证毕.

进而可以证明以下定理.

定理 2.11 若 $\gcd(f(x), g(x)) = 1$, 且 $f(x) \mid g(x)h(x)$, 则 $f(x) \mid h(x)$.

证明 由 $\gcd(f(x), g(x)) = 1$ 可知, 有 $u(x)$ 和 $v(x)$ 使

$$u(x)f(x) + v(x)g(x) = 1.$$

在等式两边乘 $h(x)$, 得

$$u(x)f(x)h(x) + v(x)g(x)h(x) = h(x).$$

因为 $f(x) \mid g(x)h(x)$, 所以 $f(x)$ 整除等式左端, 从而 $f(x) \mid h(x)$. 证毕.

推论 2.12 若 $f_1(x) \mid g(x)$, $f_2(x) \mid g(x)$, 并且 $\gcd(f_1(x), f_2(x)) = 1$, 则 $f_1(x)f_2(x) \mid g(x)$.

证明 设 $g(x) = f_1(x)q(x)$, 因 $f_2(x) \mid g(x)$ 且 $\gcd(f_1(x), f_2(x)) = 1$, 从而 $f_2(x) \mid q(x)$, 故可设 $q(x) = f_2(x)\tilde{q}(x)$, 于是 $g(x) = f_1(x)f_2(x)\tilde{q}(x)$, 即 $f_1(x)f_2(x) \mid g(x)$.

对于有限多个多项式 $f_1(x), f_2(x), \dots, f_s(x)$, 同样可以定义其最大公因式.

定义 2.9 若多项式 $d(x)$ 满足下面的性质:

(1) 对一切 $k = 1, 2, \dots, s$, $d(x) \mid f_k(x)$;

(2) 若 $d_1(x) \mid f_k(x)$ 对一切 $k = 1, 2, \dots, s$ 成立, 那么 $d_1(x) \mid d(x)$,

则多项式 $d(x)$ 称为 $f_1(x), f_2(x), \dots, f_s(x)$ 的一个最大公因式, 其中首项系数为 1 的最大公因式记为 $\gcd(f_1(x), f_2(x), \dots, f_s(x))$.

下面的定理, 提供了计算多个多项式的最大公因式的方法.

定理 2.13 $\gcd(f_1(x), f_2(x), \dots, f_s(x)) = \gcd(f_1(x), \gcd(f_2(x), \dots, f_s(x)))$.

证明 只需证明 $f_1(x), f_2(x), \dots, f_s(x)$ 的公因式与 $f_1(x)$ 和 $\gcd(f_2(x), \dots, f_s(x))$ 的公因式完全相同.

一方面, 若 $\varphi(x)$ 是 $f_1(x), f_2(x), \dots, f_s(x)$ 的公因式, 则 $\varphi(x)$ 显然是 $f_2(x), \dots, f_s(x)$ 的公因式, 从而 $\varphi(x) \mid \gcd(f_2(x), \dots, f_s(x))$, 即 $\varphi(x)$ 是 $f_1(x)$ 和 $\gcd(f_2(x), \dots, f_s(x))$ 的公因式.

另一方面, 若 $\psi(x)$ 是 $f_1(x)$ 和 $\gcd(f_2(x), \dots, f_s(x))$ 的公因式, 则

$$\psi(x) \mid \gcd(f_2(x), \dots, f_s(x)),$$

从而 $\psi(x)$ 也是 $f_2(x), \dots, f_s(x)$ 的公因式. 证毕.

定理 2.9 可以推广到多个多项式的最大公因式的情形, 留作习题.

最后, 简单介绍一下两个多项式的最小公倍式.

定义 2.10 设 $f(x), g(x) \in \mathbf{F}[x]$, 若 $m(x) \in \mathbf{F}[x]$ 既是 $f(x)$ 的倍式, 又是 $g(x)$ 的倍式, 则称 $m(x)$ 为 $f(x)$ 和 $g(x)$ 的一个公倍式. 进一步, 若 $m(x)$ 是 $f(x)$ 和 $g(x)$ 的公倍式, 并且 $f(x)$ 和 $g(x)$ 的任意公倍式都是 $m(x)$ 的倍式, 则称 $m(x)$ 为 $f(x)$ 和 $g(x)$ 的一个最小公倍式.

同样, 最小公倍式是次数最小的公倍式.

最小公倍式的存在性与计算, 可依据下面的定理.

定理 2.14 若 $f(x)$ 和 $g(x)$ 为多项式, 则 $\frac{f(x)g(x)}{\gcd(f(x), g(x))}$ 是 $f(x)$ 和 $g(x)$ 的最小公倍式.

证明 记 $d(x) = \gcd(f(x), g(x))$, $f_1(x) = \frac{f(x)}{d(x)}$, $g_1(x) = \frac{g(x)}{d(x)}$, 易见 $f_1(x)$ 和 $g_1(x)$ 都是多项式, 从而 $\frac{f(x)g(x)}{d(x)} = f_1(x)d(x)g_1(x)$ 是 $f(x)$ 和 $g(x)$ 的公倍式. 于是只需证明, 对任意的 $f(x)$ 和 $g(x)$ 的公倍式 $m(x)$, 有 $f_1(x)d(x)g_1(x) \mid m(x)$ 即可. 选取多项式 $u(x), v(x)$, 使得

$$u(x)f(x) + v(x)g(x) = \gcd(f(x), g(x)),$$

于是

$$u(x)f_1(x) + v(x)g_1(x) = 1,$$

即 $\gcd(f_1(x), g_1(x)) = 1$. 记 $m(x) = f(x)p(x) = g(x)q(x)$, 同除以 $\gcd(f(x), g(x))$ 后得到, $f_1(x)p(x) = g_1(x)q(x)$, 特别 $f_1(x) \mid g_1(x)q(x)$. 由 $f_1(x)$ 与 $g_1(x)$ 互素, 得 $f_1(x) \mid q(x)$, 从而 $f_1(x)d(x)g_1(x) \mid q(x)d(x)g_1(x)$, 即 $f_1(x)d(x)g_1(x) \mid m(x)$. 证毕.

注 2.5 可以证明, 两个多项式 $f(x)$ 和 $g(x)$ 的最小公倍式不是唯一的, 不同的最小公倍式之间必定相差一个非零常数倍, 其中首项系数为 1 的最小公倍式, 记为 $\text{lcm}(f(x), g(x))$.

对多个多项式的最小公倍式, 可类似定义, 这里不再赘述.

注 2.6 通过对比整数和多项式的相关内容, 不难发现两者之间非常相似. 造成这一现象的本质原因是, 在整数和多项式两个集合上, 都可以定义加法、减法、乘法以及带余除法. 整数和多项式中所有的定义、定理都是基于上述四种运算得到的. 可以定义加法、减法、乘法以及带余除法的集合在代数学中称为 Euclid 环 (参见文献 [1]). 研究具有相同代数结构的数学对象的共性, 正是代数学的本质.

2.4 插值多项式与中国剩余定理

本节先考虑一个简单且基本的问题: 设 x_k, y_k ($1 \leq k \leq n$) 为 $2n$ 个数, 且 x_1, x_2, \dots, x_n 互不相同, 求一个多项式 $f(x)$, 使得 $f(x_k) = y_k$ ($1 \leq k \leq n$).

解决这个问题的初等方法是待定系数法, 即设

$$f(x) = a_{n-1}x^{n-1} + \dots + a_1x + a_0,$$

代入条件 $f(x_k) = y_k$ ($1 \leq k \leq n$), 得到一个关于系数 a_0, \dots, a_{n-1} 的 n 元一次方程组, 解这个方程组, 便可得到最终的解. 当 n 较大时, 解方程组需要相当的计算量, 下面给出该问题的一般公式.

为简化表述, 记 $P(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$, 以及 $P_k(x) = \frac{P(x)}{x - x_k}$ (即 $P_k(x)$ 为 $P(x)$

去掉因子 $(x - x_k)$ 后的多项式), 此时有

$$(1) P_k(x_k) \neq 0;$$

(2) 当 $j \neq k$ 时, $P_k(x_j) = 0$.

定理 2.15 (Lagrange¹插值公式) 多项式

$$L(x) = \sum_{j=1}^n \frac{P_j(x)}{P_j(x_j)} y_j = \frac{P_1(x)}{P_1(x_1)} y_1 + \frac{P_2(x)}{P_2(x_2)} y_2 + \cdots + \frac{P_n(x)}{P_n(x_n)} y_n,$$

满足对一切 $1 \leq k \leq n$, $L(x_k) = y_k$.

该定理的证明可通过直接验证得到.

例 2.8 求一个多项式 $f(x)$, 满足 $f(0) = 0$, $f(1) = 3$, $f(-1) = 1$.

解 记 $P(x) = x(x-1)(x+1)$ 以及 $P_1(x) = (x-1)(x+1)$, $P_2(x) = x(x+1)$, $P_3(x) = x(x-1)$, 从而由 Lagrange 插值公式得到

$$f(x) = \frac{P_1(x)}{P_1(0)} \cdot 0 + \frac{P_2(x)}{P_2(1)} \cdot 3 + \frac{P_3(x)}{P_3(-1)} \cdot 1,$$

满足要求. 代入数值并化简, 得到

$$f(x) = \frac{3}{2}x(x+1) + \frac{1}{2}x(x-1) = 2x^2 + x.$$

Lagrange 插值公式, 在代数上蕴含了深刻的内容. 根据余式定理 (定理 2.6), 条件 $f(x_k) = y_k$ 等价于 $(x - x_k)$ 除 $f(x)$ 余式为 y_k . 为了叙述方便, 引进下面的定义.

定义 2.11 对非零多项式 $g(x)$, 若多项式 $f_1(x), f_2(x)$ 除以 $g(x)$ 的余式相同, 则称 $f_1(x), f_2(x)$ 关于 $g(x)$ 同余, 记为 $f_1(x) \equiv f_2(x) \pmod{g(x)}$.

容易得到下面的命题, 证明留作习题.

命题 2.16 (1) 若 $f_1(x) \equiv f_2(x) \pmod{g(x)}$, $f_2(x) \equiv f_3(x) \pmod{g(x)}$, 则 $f_1(x) \equiv f_3(x) \pmod{g(x)}$.

(2) 若 $g(x)$ 除 $f(x)$ 所得的余式为 $r(x)$, 则 $f(x) \equiv r(x) \pmod{g(x)}$.

(3) $f_1(x) \equiv f_2(x) \pmod{g(x)}$ 当且仅当 $g(x) \mid (f_1(x) - f_2(x))$.

利用同余的记法, Lagrange 插值多项式满足的条件可等价表述为

$$L(x) \equiv y_k \pmod{(x - x_k)} \quad (1 \leq k \leq n).$$

现在考虑更一般的问题: 设 $g_1(x), g_2(x), \dots, g_n(x), r_1(x), r_2(x), \dots, r_n(x)$ 是 $2n$ 个多项式, 其中 $g_1(x), g_2(x), \dots, g_n(x)$ 次数大于等于 1 且两两互素, 是否存在多项式 $f(x)$, 满足

$$f(x) \equiv r_k(x) \pmod{g_k(x)} \quad (1 \leq k \leq n). \quad (2.5)$$

这一问题的结论是肯定的, 称为中国剩余定理.

定理 2.17 (中国剩余定理) 设 $g_1(x), g_2(x), \dots, g_n(x), r_1(x), r_2(x), \dots, r_n(x)$ 是 $2n$ 个多项式, 其中 $g_1(x), g_2(x), \dots, g_n(x)$ 次数大于 1 且两两互素, 则存在多项式 $f(x)$, 满足条件 (2.5). 进一步, $\tilde{f}(x)$ 满足条件 (2.5) 当且仅当

$$f(x) \equiv \tilde{f}(x) \pmod{g_1(x)g_2(x) \cdots g_n(x)},$$

即 $f(x)$ 关于 $g_1(x)g_2(x) \cdots g_n(x)$ 的余式唯一. 特别, 若要求

$$\deg f(x) < \deg g_1(x)g_2(x) \cdots g_n(x),$$

¹Lagrange (1736–1813), 拉格朗日, 法国数学家.

则 $f(x)$ 存在且唯一.

中国剩余定理的证明以及 $f(x)$ 的计算方法, 与 Lagrange 插值多项式构造的本质是相同的. 注意到 $P_k(x)$ 为一切因子 $(x - x_j)$ ($j \neq k$) 的乘积, 从而 $\frac{P_k(x)}{P_k(x_k)}$ 是这些因子的公倍式, 又根据余式定理,

$$\frac{P_k(x)}{P_k(x_k)} \equiv 1 \pmod{(x - x_k)},$$

因此为构造中国剩余定理中的 $f(x)$, 只需将 $P_k(x)$ 用 $Q_k(x)$ 取代, y_k 用 $r_k(x)$ 取代即可, 这里多项式 $Q_k(x)$ 满足:

(1) $Q_k(x)$ 为一切因子 $g_j(x)$ ($j \neq k$) 的公倍式;

(2) $Q_k(x) \equiv 1 \pmod{g_k(x)}$.

为得到 $Q_k(x)$, 记 $q_k(x)$ 一切因子 $g_j(x)$ ($j \neq k$) 的乘积, 根据 $g_1(x), g_2(x), \dots, g_n(x)$ 两两互素, 得到 $g_k(x)$ 与 $q_k(x)$ 互素, 从而存在多项式 $u(x), v(x)$, 满足

$$g_k(x)u(x) + q_k(x)v(x) = 1,$$

于是

$$q_k(x)v(x) = -g_k(x)u(x) + 1 \equiv 1 \pmod{g_k(x)}.$$

令 $Q_k(x) = q_k(x)v(x)$, 则

$$f(x) = \sum_{k=1}^n Q_k(x)r_k(x) = Q_1(x)r_1(x) + Q_2(x)r_2(x) + \dots + Q_n(x)r_n(x), \quad (2.6)$$

即为所求.

证明 直接验证可知由式 (2.6) 定义的 $f(x)$ 满足条件 (2.5), 于是 $\tilde{f}(x)$ 满足条件 (2.5) 当且仅当对一切 $1 \leq k \leq n$,

$$g_k(x) \mid (f(x) - \tilde{f}(x)). \quad (2.7)$$

由于 $g_1(x), g_2(x), \dots, g_n(x)$ 两两互素, 条件 (2.7) 等价于 $g_1(x)g_2(x) \cdots g_n(x) \mid (f(x) - \tilde{f}(x))$, 即

$$f(x) \equiv \tilde{f}(x) \pmod{g_1(x)g_2(x) \cdots g_n(x)}.$$

证毕.

了解关于整数的中国剩余定理的读者不难发现, 多项式的中国剩余定理与整数的中国剩余定理, 表述基本相同, 证明的本质完全相同.

例 2.9 计算 $f(x) = x^{20}$ 关于 $(x-2)(x-3)$ 的余式 $r(x)$.

解 根据余式定理知,

$$f(x) \equiv f(2) = 2^{20} \pmod{(x-2)},$$

$$f(x) \equiv f(3) = 3^{20} \pmod{(x-3)},$$

从而根据中国剩余定理以及 Lagrange 插值公式,

$$f(x) \equiv 2^{20} \frac{x-3}{2-3} + 3^{20} \frac{x-2}{3-2} = (3^{20} - 2^{20})x + (3 \cdot 2^{20} - 2 \cdot 3^{20}) \pmod{(x-2)(x-3)}.$$

2.5 因式分解与不可约多项式

本节考虑一元多项式的因式分解.

定义 2.12 设 $p(x)$ 为数域 \mathbf{F} 上的非常值多项式, 若 $p(x)$ 在 $\mathbf{F}[x]$ 上只有平凡因式, 则称 $p(x)$ 为 $\mathbf{F}[x]$ 上的不可约多项式, 否则, 称 $p(x)$ 为 $\mathbf{F}[x]$ 上的可约多项式.

按照定义, 一次多项式总是不可约多项式.

注 2.7 多项式 $p(x)$ 的不可约性与 $p(x)$ 的系数所在的数域密切相关. 例如, $x^2 + 2$ 是 $\mathbf{R}[x]$ 上的不可约多项式, 但是它在 $\mathbf{C}[x]$ 上可以分解成两个一个多项式的乘积

$$x^2 + 2 = (x + \sqrt{2}i)(x - \sqrt{2}i),$$

因而在 $\mathbf{C}[x]$ 上是可约的. 这就说明, 一个多项式是否不可约是依赖于系数域. 但需要注意的是, 两个多项式的最大公因式并不依赖于域的选取, 这是因为辗转相除法并不会因为域的选择不同, 而得到不同的结果.

在 $\mathbf{F}[x]$ 上的不可约多项式 $p(x)$, 在 $\mathbf{F}[x]$ 上的因式只有平凡因式, 即除非零常数和它自身的非零常数倍 $cp(x)$ ($c \neq 0$) 两种因式外, 再无其他因式了. 特别, 有下面的命题成立.

命题 2.18 $\mathbf{F}[x]$ 上的不可约多项式 $p(x)$ 与 $\mathbf{F}[x]$ 上的任一多项式 $f(x)$ 之间只可能有两种关系: $p(x) \mid f(x)$, 或者 $p(x)$ 与 $f(x)$ 互素.

证明 设 $\gcd(p(x), f(x)) = d(x)$, 因 $p(x)$ 在 $\mathbf{F}[x]$ 上的只有平凡因式, 故 $d(x)$ 或者是 1 或者是 $cp(x)$. 当 $d(x) = 1$ 时, $p(x)$ 与 $f(x)$ 互素; 当 $d(x) = cp(x)$ 时, $p(x) \mid f(x)$.

不可约多项式有下述的重要性质.

定理 2.19 设 $p(x)$ 是数域 \mathbf{F} 上的不可约多项式, $f(x)$ 和 $g(x)$ 是数域 \mathbf{F} 上的两个多项式, 若 $p(x) \mid f(x)g(x)$, 则 $p(x) \mid f(x)$ 或者 $p(x) \mid g(x)$ 至少有一个成立.

证明 如果 $p(x) \mid f(x)$, 那么结论已经成立. 如果 $p(x) \nmid f(x)$, 则

$$\gcd(p(x), f(x)) = 1,$$

于是由定理 2.11 立得 $p(x) \mid g(x)$. 证毕.

利用数学归纳法, 不难将上述定理可以推广为以下形式.

定理 2.20 如果不可约多项式 $p(x)$ 整除多项式 $f_1(x), f_2(x), \dots, f_s(x)$ 的乘积, 那么 $p(x)$ 至少整除这些多项式之中的某一个.

下面给出本节的主要结论.

定理 2.21 (因式分解唯一性定理) 系数在数域 \mathbf{F} 上的任意一个非常数多项式 $f(x)$ 都可以分解成系数在数域 \mathbf{F} 上的有限个不可约多项式的乘积, 并且这个分解式在不计非零常数因子和因子次序的意义下是唯一的, 即若有两个分解式

$$f(x) = p_1(x)p_2(x) \cdots p_s(x) = q_1(x)q_2(x) \cdots q_t(x),$$

则必有 $s = t$, 并且适当排列因式的次序后有

$$p_i(x) = c_i q_i(x) \quad (i = 1, 2, \dots, s),$$

其中 $c_i \in \mathbf{F}$ ($i = 1, 2, \dots, s$) 为非零常数.

证明 先证分解式的存在性. 对多项式 $f(x)$ 的次数 $n = \deg f(x)$ 作归纳法. 因为一次多

项式都是不可约的, 所以当 $n = 1$ 时结论成立. 假设当 $\deg f(x) \leq n$ 时结论成立. 现在考虑当 $\deg f(x) = n + 1$ 时的情形, 若此时 $f(x)$ 是不可约多项式, 结论是显然成立. 不妨设 $f(x)$ 不是不可约的, 即有

$$f(x) = f_1(x)f_2(x),$$

其中 $f_1(x), f_2(x)$ 的次数都小于 n . 由归纳法假定 $f_1(x)$ 和 $f_2(x)$ 都可分解成数域 \mathbf{F} 上有限个不可约多项式的乘积. 把 $f_1(x), f_2(x)$ 的分解式合起来就得到 $f(x)$ 的一个分解式. 由归纳法原理知, 结论成立.

再证唯一性. 设 $f(x)$ 存在两种不可约多项式分解

$$f(x) = p_1(x)p_2(x) \cdots p_s(x) = q_1(x)q_2(x) \cdots q_t(x). \quad (2.8)$$

对 s 作归纳法. 当 $s = 1$ 时, $f(x)$ 是不可约多项式, 由定义可得 $s = t = 1$, 并且 $f(x) = p_1(x) = q_1(x)$. 假设当不可约因式的个数为 $s - 1$ 时, 唯一性成立, 则当不可约因式的个数为 s 时, 因 $p_1(x) \mid q_1(x)q_2(x) \cdots q_t(x)$, 则 $p_1(x)$ 必能整除 $q_1(x), \cdots, q_t(x)$ 中的某一个, 不妨设 $p_1(x) \mid q_1(x)$, 因为 $q_1(x)$ 也是不可约多项式, 所以有 $p_1(x) = c_1 q_1(x)$. 根据消去律, 可在 (2.8) 式两边消去 $p_1(x)$, 得到

$$p_2(x) \cdots p_s(x) = c_1^{-1} q_2(x) \cdots q_t(x),$$

上式左边是 $s - 1$ 个不可约多项式的乘积, 由归纳假设得 $s - 1 = t - 1$, 且 $q_2(x), \cdots, q_s(x)$ 经次序重排后有 $p_i(x) = c_i q_i(x) (i = 2, \cdots, s)$. 证毕.

当数域为复数域或实数域时, 有下面的结论.

定理 2.22 $\mathbf{C}[x]$ 上的多项式不可约的充分必要条件是多项式是一次多项式.

该定理的证明需要借助著名的代数学基本定理: 任何一个复系数多项式, 在复数域 \mathbf{C} 上恰有 n 个根.

定理 2.23 $\mathbf{R}[x]$ 上的多项式 $p(x)$ 不可约的充分必要条件是 $p(x)$ 满足下述条件之一:

- (1) $p(x)$ 是一次多项式;
- (2) $p(x)$ 是形如 $c(x^2 + px + q)$ 的二次多项式, 其中 $c, p, q \in \mathbf{R}$ 且 $p^2 - 4q < 0$.

该定理的证明, 只须利用实系数多项式虚根共轭成对, 并结合定理 2.22.

在多项式 $f(x)$ 的分解式中, 可以将每一个不可约因式的首项系数提出来, 使它们成为首项系数为 1 的多项式, 再将相同的不可约因式合并成幂的形式. 于是 $f(x)$ 在数域 \mathbf{F} 上的分解式可写为

$$f(x) = ap_1^{r_1}(x)p_2^{r_2}(x) \cdots p_s^{r_s}(x),$$

其中 a 是 $f(x)$ 的首项系数, $p_1(x), p_2(x), \cdots, p_s(x)$ 是两两不同的并且首项系数为 1 的不可约多项式, 而 r_1, r_2, \cdots, r_s 是正整数. 这种分解式称为多项式 $f(x)$ 在数域 \mathbf{F} 上的标准分解式.

如果已经有了两个多项式在同一个数域 \mathbf{F} 上的标准分解式, 就可以直接写出两个多项式的最大公因式. 多项式 $f(x)$ 和 $g(x)$ 的最大公因式 $d(x)$ 就是那些同时在 $f(x)$ 和 $g(x)$ 的标准分解式中都出现的不可约多项式的方幂的乘积, 所带的方幂就等于它在 $f(x)$ 和 $g(x)$ 中所带方幂较小的一个.

结合复系数和实系数不可约多项式的特征, 复系数和实系数多项式的因式分解定理可以

表示为以下形式.

定理 2.24 (复系数多项式的因式分解定理) 非常数的复系数多项式在复数域 \mathbf{C} 上可唯一地分解成一次因式的乘积.

复系数多项式 $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ 的标准分解式为

$$f(x) = a_n (x - r_1)^{n_1} (x - r_2)^{n_2} \cdots (x - r_k)^{n_k},$$

其中 n_1, n_2, \cdots, n_k 是正整数, 且 $n_1 + n_2 + \cdots + n_k = n$.

定理 2.25 (实系数多项式的因式分解定理) 非常数的实系数多项式在实数域 \mathbf{R} 上可唯一地分解成一次因式和二次不可约因式的乘积.

实系数多项式 $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ 的标准分解式为

$$f(x) = a_n (x - r_1)^{n_1} \cdots (x - r_s)^{n_s} (x^2 + p_1 x + q_1)^{m_1} \cdots (x^2 + p_t x + q_t)^{m_t},$$

其中 $r_1, \cdots, r_s, p_1, \cdots, p_t, q_1, \cdots, q_t$ 都是实数, $n_1, \cdots, n_s, m_1, \cdots, m_t$ 是正整数, 并且 $n_1 + \cdots + n_s + 2m_1 + \cdots + 2m_t = n$, $p_i^2 - 4q_i < 0$ ($i = 1, 2, \cdots, t$).

注 2.8 多项式的因式分解定理并没有给出一个具体的将多项式分解为不可约多项式乘积的可计算的方法. 对不超过 5 次的复系数多项式, 尚可通过求根公式得到所有根的显式表达式, 而对 5 次及以上的多项式, 根据著名的 Galois 理论, 其根通常已不能通过根式计算, 从而无法得到该情形下的因式分解显式表达式.

判定有理系数多项式在 $\mathbf{Q}[x]$ 上是否可约, 难度更大. 通常先提取一个常系数, 将有理系数多项式化为整系数多项式, 即令有理系数多项式 $f(x) = c(a_0 x^n + \cdots + a_{n-1} x + a_n)$, 这里 $a_0, \cdots, a_{n-1}, a_n$ 为整数且他们的最大公因数为 1, 此时有下面的定理.

定理 2.26 (Eisenstein 判别法) 若存在素数 p , 满足以下条件

- (1) $p \nmid a_0$;
- (2) $p \mid a_k$, 对一切 $1 \leq k \leq n$;
- (3) $p^2 \nmid a_n$,

则 $f(x)$ 是 $\mathbf{Q}(x)$ 上的不可约多项式.

注意, Eisenstein 判别法只给出了有理系数多项式不可约的充分条件, 该条件未必是必要的, 该判别法的证明涉及到整系数多项式的部分性质, 这里略去.

例 2.10 证明 $f(x) = x^{p-1} + \cdots + x + 1$ 是 $\mathbf{Q}(x)$ 上的不可约多项式, 其中 p 为素数.

证明 易见, $f(x)$ 不可约与 $f(x+1)$ 不可约等价, 以下证明 $f(x+1)$ 在 $\mathbf{Q}(x)$ 上不可约.

显然 $f(x+1)$ 仍为 $(p-1)$ 次多项式, 且最高次项系数为 1, 从而 p 满足 Eisenstein 判别法的条件 (1). 其次, 由 $f(1) = p$ 得 $f(x+1)$ 的常数项为 p , 从而 p 满足 Eisenstein 判别法的条件 (3). 最后由等比数列求和公式,

$$f(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = \frac{(x+1)^p - 1}{x},$$

故当 $1 \leq k \leq p-2$ 时, k 次项的系数为

$$C_p^{k+1} = \frac{p!}{(p-k-1)!(k+1)!}.$$

根据 $p \mid (p-k-1)!(k+1)!C_p^{k+1}$ 且 p 是素数, 于是 $p \nmid (p-k-1)!(k+1)!$, 故 $p \mid C_p^{k+1}$, 即 p 满足 Eisenstein 判别法的条件 (2). 从而 $f(x+1)$ 在 $\mathbf{Q}(x)$ 上不可约. 证毕.

2.6 多项式的导数和零化多项式

本节将介绍在后续章节中有重要应用的两个概念.

2.6.1 多项式的导数

将多项式看作关于不定元的函数, 则可以对多项式求导数, 根据导数的计算公式易知, 多项式的任意阶导数仍然是多项式.

在本书中, 多项式的导数的应用体现在两处. 其一是判别多项式是否有重根, 有下面的结论.

定理 2.27 多项式 $f(x)$ 在复数域上无重根的充分必要条件是 $\gcd(f(x), f'(x)) = 1$.

该定理的意义在于, 将判断多项式是否存在重根的问题, 转化为计算两个多项式的最大公因式, 前者由于一般次数的多项式的根没有一般的计算方法, 无法直接判别, 而后者可以通过辗转相除法计算得到.

证明 充分性. 设 x_0 为多项式 $f(x)$ 的任一根, 重数为 k , 于是根据唯一分解定理, $f(x) = (x-x_0)^k g(x)$, 这里 $g(x)$ 为多项式且 $g(x_0) \neq 0$, 此时

$$f'(x) = (x-x_0)^k g'(x) + k(x-x_0)^{k-1} g(x) = (x-x_0)^{k-1} ((x-x_0)g'(x) + kg(x)),$$

从而 $(x-x_0)^{k-1}$ 是 $f(x)$ 与 $f'(x)$ 的公因式. 但 $\gcd(f(x), f'(x)) = 1$, 故 $k = 1$.

必要性. 设 x_0 为多项式 $f(x)$ 的任一根, 由于其重数为 1, 故 $f(x) = (x-x_0)g(x)$, 这里 $g(x)$ 为多项式且 $g(x_0) \neq 0$, 由余式定理 (定理 2.6) 知 $(x-x_0) \nmid g(x)$. 又 $f'(x) = (x-x_0)g'(x) + g(x)$, 故 $(x-x_0) \nmid f'(x)$. 由 x_0 的任意性, 得 $f(x)$ 的任何一个不可约因子都不整除 $f'(x)$, 从而 $\gcd(f(x), f'(x)) = 1$. 证毕.

注 2.9 判断两个多项式是否互素, 还可以进一步转化为计算由两个多项式的系数构成的行列式的值是否为零, 该行列式称为两个多项式的结式. 多项式与其导数的结式亦称为多项式的判别式. 这里不再展开, 有兴趣的读者可参考文献 [1].

多项式的导数的另一个应用是计算多项式关于高次因子的余式. 设 $f(x)$ 为 n 次多项式, 考虑 $f(x)$ 关于 $(x-x_0)^m$ 的余式 (这里不妨假设 $m \leq n$, 否则余式极为 $f(x)$ 自身). 当 $m = 1$ 时, 由余式定理知余式为 $f(x_0)$, 以下考虑 $m > 1$ 的情形.

根据微积分中的 Taylor 公式

$$f(x) = \sum_{k=1}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x-x_0)^k,$$

而上式左边为 n 次多项式, 故对一切 $k > n$, $f^{(k)}(x) = 0$, 故上式可化简为

$$f(x) = \sum_{k=1}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k.$$

于是

$$f(x) \equiv \sum_{k=1}^{m-1} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \pmod{(x - x_0)^m}. \quad (2.9)$$

例 2.11 计算 $f(x) = x^{10}$ 关于 $(x - 1)^2$ 的余式.

解 根据式 (2.9), $f(x) = x^{10}$ 关于 $(x - 1)^2$ 的余式

$$r(x) = f'(1)(x - 1) + f(1) = 10(x - 1) + 1 = 10x - 9.$$

结合中国剩余定理 (定理 2.17), 还可以计算更一般的问题.

例 2.12 计算 $f(x) = x^{10}$ 关于 $(x - 1)^2(x + 1)$ 的余式.

解 设 $f(x) \equiv r(x) \pmod{(x - 1)^2(x + 1)}$. 由上题结论以及余式定理知

$$\begin{aligned} r(x) &\equiv f(x) \equiv 10x - 9 \pmod{(x - 1)^2}, \\ r(x) &\equiv f(x) \equiv f(-1) = 1 \pmod{(x + 1)}. \end{aligned} \quad (2.10)$$

根据中国剩余定理, 满足上述同余方程, 且次数低于 3 的多项式 $r(x)$ 唯一存在. 从而设

$$r(x) = a(x - 1)^2 + 10x - 9,$$

此时 $r(x)$ 满足 (2.10) 的第 1 个方程, 只需取适当的 a 使得 $r(x)$ 满足 (2.10) 的第 2 个方程即可. 再次根据余式定理,

$$r(x) \equiv r(-1) = 4a - 19 \pmod{(x + 1)},$$

故 $a = 5$, 即

$$r(x) = 5(x - 1)^2 + 10x - 9 = 5x^2 - 4.$$

注 2.10 利用上述方法计算余式, 某些情形下比使用长除法简便. 在计算软件日益发达的今天, 这种简便可能微不足道, 但当 $f(x)$ 为一般的解析函数 (即无穷幂级数) 时, 长除法将不再适用, 上述方法依然可行.

2.6.2 零化多项式

为介绍零化多项式, 首先看一个例子.

例 2.13 设 $f(x) = 2x^3 - x^2 - 5x + 4$, 计算 $f(\sqrt{2} + 1)$.

本题固然可以将 $x = \sqrt{2} + 1$ 代入多项式直接求值, 但运算时需要将每一个关于 $\sqrt{2} + 1$ 的幂完全展开, 计算量较大, 而当 $f(x)$ 的次数较高时, 这种做法的缺点将更加显著. 以下的做法, 避开了这一困难.

解 令 $x = \sqrt{2} + 1$, 则 $(x - 1)^2 = 2$, 即 $x^2 - 2x - 1 = 0$. 将 $f(x)$ 关于 $(x^2 - 2x - 1)$ 求余式, 得

$$f(x) = q(x)(x^2 - 2x - 1) + (3x + 7),$$

(这里 $q(x) = 2x + 3$, 但对本题是无关紧要的), 从而 $f(\sqrt{2} + 1) = 3(\sqrt{2} + 1) + 7 = 3\sqrt{2} + 10$.

本题中, 利用多项式 $x^2 - 2x - 1$ 在 $\sqrt{2} + 1$ 处取值为零, 将计算 $f(x)$ 在 $\sqrt{2} + 1$ 处取值的问题, 转化为一个次数较低的多项式在 $\sqrt{2} + 1$ 处取值的问题. 一般地, 有如下定义.

定义 2.13 若多项式 $f(x)$ 以 c 为根, 则称 $f(x)$ 为 c 的一个零化多项式.

注 2.11 一个数的零化多项式不一定存在, 若数 c 在 $\mathbf{Q}[x]$ 上存在零化多项式, 则称 c 为代数数, 否则称为超越数. 可以证明 e 和 π 都是超越数 (其证明超出本书范围, 略去).

零化多项式如若存在, 必定不唯一. 显然, 任何一个零化多项式的倍式必定是零化多项式. 但所有的零化多项式中存在次数最小的多项式, 且在相差一个常数因子的意义下唯一. 该结论基于下述定理.

定理 2.28 若 $f(x)$ 与 $g(x)$ 同为 c 的零化多项式, 则 $\gcd(f(x), g(x))$ 也是 c 的零化多项式.

证明 由定理 2.9, 存在多项式 $u(x)$ 与 $v(x)$, 使得

$$u(x)f(x) + v(x)g(x) = \gcd(f(x), g(x)),$$

两边在 c 处取值, 即得结论. 证毕.

由此可以定义一个数的最小多项式.

定义 2.14 数 c 的所有在 $\mathbf{F}[x]$ 上的零化多项式的最大公因式, 称为数 c 的最小多项式.

根据定理 2.28, 最小多项式必定是零化多项式, 且是次数最小的零化多项式.

在例 2.13 中, 寻找数的零化多项式并进而求关于零化多项式的余式, 通常比直接计算多项式在该数处的值简便, 而在选择零化多项式时, 应尽量寻求最小多项式, 以减少后续求值时的计算量.

注 2.12 在后续章节中将看到, 矩阵同样存在零化多项式, 于是例 2.13 的思想方法可以移植到矩阵情形.

习题 2

1. 计算 $g(x)$ 除 $f(x)$ 的商式 $q(x)$ 和余式 $r(x)$.

(1) $f(x) = x^4 - 4x + 5$, $g(x) = x^2 - x + 2$;

(2) $f(x) = x^3 - 3x^2 - x - 1$, $g(x) = 3x^2 - 2x + 1$.

2. 求多项式 $f(x)$ 和 $g(x)$ 的最大公因式和最小公倍式.

(1) $f(x) = x^4 + x^3 + 2x^2 + x + 1$, $g(x) = x^3 + 2x^2 + 2x + 1$;

(2) $f(x) = x^4 - 4x^3 + 1$, $g(x) = x^3 - 3x^2 + 1$.

3. 求多项式 $f(x)$ 和 $g(x)$ 的最大公因式 $\gcd(f(x), g(x))$, 以及满足等式 $u(x)f(x) + v(x)g(x) = \gcd(f(x), g(x))$ 的多项式 $u(x)$ 和 $v(x)$.

(1) $f(x) = x^4 - x^3 - 4x^2 + 4x + 1$, $g(x) = x^2 - x - 1$;

$$(2) f(x) = 4x^4 - 2x^3 - 16x^2 + 5x + 9, g(x) = 2x^3 - x^2 - 5x + 4.$$

4. 求满足下述条件的 Lagrange 插值多项式 $f(x)$.

$$(1) f(0) = 1, f(1) = 3, f(2) = 9;$$

$$(2) f(-2) = -1, f(-1) = 1, f(1) = 1, f(2) = -1.$$

5. 证明命题 2.16 .

6. 若多项式 $f(x), g(x), u(x), v(x)$ 满足 $u(x)f(x) + v(x)g(x) = \gcd(f(x), g(x))$, 证明 $u(x), v(x)$ 互素.

7. 若 $f(x), g(x) \in \mathbf{F}[x]$, 证明存在 $u(x), v(x) \in \mathbf{F}[x]$, 且 $\deg u(x) < \deg g(x), \deg v(x) < \deg f(x)$, 使得 $u(x)f(x) + v(x)g(x) = \gcd(f(x), g(x))$.

8. 若 $f(x)$ 和 $g(x)$ 分别为 m 次与 n 次非零多项式. 证明: 存在次数小于 n 次与 m 次的非零多项式 $u(x)$ 和 $v(x)$, 使得 $u(x)f(x) = v(x)g(x)$ 的充分必要条件是 $f(x)$ 与 $g(x)$ 不是互素的.

9. 设 $g_1(x), g_2(x), g_3(x)$ 两两互素, 证明 $g_1(x)$ 与 $g_2(x)g_3(x)$ 互素.

10. 试给出多个多项式的最小公倍式的定义及计算方法.

11. (1) 若 ω 是有理系数 n 次不可约多项式 $f(x)$ 的一个非有理根, 证明 $\mathbf{Q}(\omega) = \{a_0 + a_1\omega + \cdots + a_{n-1}\omega^{n-1} : a_0, a_1, \cdots, a_{n-1} \in \mathbf{Q}\}$ 是数域.

(2) 证明 $\mathbf{Q}(\sqrt[3]{2}) = \{a + b\sqrt[3]{2} + c\sqrt[3]{4} : a, b, c \in \mathbf{Q}\}$ 是数域.

12. 设 $f(x) = a_mx^m + a_{m-1}x^{m-1} + \cdots + a_0$, $g(x) = b_nx^n + b_{n-1}x^{n-1} + \cdots + b_0$, 分别为 m 次与 n 次非零多项式. 证明: $f(x)$ 与 $g(x)$ 互素的充分必要条件是

$$\left\{ \begin{array}{l} n \text{ 行} \\ m \text{ 行} \end{array} \right\} \left| \begin{array}{cccc} a_m & a_{m-1} & \cdots & a_0 \\ & a_m & a_{m-1} & \cdots & a_0 \\ & & \ddots & & \ddots \\ & & & a_m & a_{m-1} & \cdots & a_0 \\ b_n & b_{n-1} & \cdots & b_0 \\ & b_n & b_{n-1} & \cdots & b_0 \\ & & \ddots & & \ddots \\ & & & b_n & b_{n-1} & \cdots & b_0 \end{array} \right| \neq 0.$$

该 $(m+n)$ 阶行列式称为多项式 f 与 g 的结式.

13. 多项式 $f(x)$ 与其导数多项式 $f'(x)$ 的结式称为多项式 f 的判别式. 证明: 多项式 f 无重根的充分必要条件是 f 的判别式非零.

14. 试计算一元二次多项式的判别式, 比较该判别式与一元二次方程的判别式的异同.

第3章 矩阵的 Jordan 标准形

本章将研究方阵相似意义下的标准形问题.

3.1 λ -矩阵的标准形

为得到矩阵相似意义下的标准形, 需要借助 λ -矩阵这一工具. 首先给出 λ -矩阵的概念.

定义 3.1 元素为 λ 的多项式的矩阵称为 λ -矩阵, 记作 $A(\lambda)$.

本书中为了区分方便, 将元素为数值的矩阵称为数值矩阵. 对数值矩阵 A , 矩阵 $\lambda E - A$ 就是一个 λ -矩阵.

定义 3.2 λ -矩阵 $A(\lambda)$ 的所有元素的最高次数称为 $A(\lambda)$ 的次数, m 次 λ -矩阵 $A(\lambda)$ 可表示为

$$A(\lambda) = \lambda^m A_0 + \lambda^{m-1} A_1 + \cdots + \lambda A_{m-1} + A_m,$$

其中, $A_i (i = 0, 1, \cdots, m)$ 是数值矩阵, 且 $A_0 \neq 0$.

例如

$$\begin{aligned} A(\lambda) &= \begin{pmatrix} \lambda^2 + 3\lambda + 1 & \lambda^3 + \lambda & 2 \\ \lambda + 2 & \lambda + 1 & \lambda^2 - 4 \\ 3 & 2\lambda^2 & 3\lambda^2 \end{pmatrix} \\ &= \lambda^3 \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \lambda^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 2 & 3 \end{pmatrix} + \lambda \begin{pmatrix} 3 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & -4 \\ 3 & 0 & 0 \end{pmatrix}. \end{aligned}$$

于是有, 数值矩阵也就是零次 λ -矩阵. 而 λ -矩阵的加, 减, 乘, 数乘以及行列式等等运算律都与数值矩阵相应的运算律相同.

定义 3.3 λ -矩阵 $A(\lambda)$ 不恒为零的子式的最高阶数称为 $A(\lambda)$ 的秩, 记作 $\text{rank } A(\lambda)$.

例 3.1 设 A 是 n 阶数字方阵, 则 $|\lambda E - A|$ 是 λ 的 n 次多项式. 因此 $\lambda E - A$ 的秩为 n . 也就是说, 不论数字方阵 A 是否满秩, 它的特征矩阵 $\lambda E - A$ 总是满秩的.

定义 3.4 设 $A(\lambda)$ 是 λ -矩阵, 若存在 λ -矩阵 $B(\lambda)$, 使 $A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = E$, 则称 $A(\lambda)$ 是可逆的, 并且 $B(\lambda)$ 称为 $A(\lambda)$ 的逆矩阵, 记作 $A(\lambda)^{-1}$.

显然, 如果 $A(\lambda)$ 是可逆的, 那么 $A(\lambda)$ 的逆矩阵必定唯一. 事实上, 假如 $B_1(\lambda)$ 和 $B_2(\lambda)$ 两个矩阵都是可逆 λ -矩阵 $A(\lambda)$ 的逆矩阵, 那么, $A(\lambda)B_1(\lambda) = B_1(\lambda)A(\lambda) = E$, 以及 $A(\lambda)B_2(\lambda) = B_2(\lambda)A(\lambda) = E$, 于是

$$B_1(\lambda) = B_1(\lambda)[A(\lambda)B_2(\lambda)] = [B_1(\lambda)A(\lambda)]B_2(\lambda) = B_2(\lambda).$$

定理 3.1 λ -矩阵 $A(\lambda)$ 是可逆的充分必要条件是 $|A(\lambda)|$ 为非零常数.

证明 记 $A(\lambda)^*$ 为 $A(\lambda)$ 的伴随矩阵. 若 $|A(\lambda)|$ 为非零常数 c , 则

$$A^{-1}(\lambda) = \frac{1}{|A(\lambda)|} A(\lambda)^* = \frac{1}{c} A(\lambda)^*$$

也是 λ -矩阵, 即 $A(\lambda)$ 可逆. 反之, 若 $A(\lambda)$ 可逆, 即存在 λ -矩阵 $A^{-1}(\lambda)$ 使 $A(\lambda)A^{-1}(\lambda) = E$. 于是 $|A(\lambda)||A^{-1}(\lambda)| = 1$. 而 $|A(\lambda)|$ 与 $|A^{-1}(\lambda)|$ 均是 λ 的多项式, 比较等式两边多项式的次数, 得 $|A(\lambda)|$ 是非零常数. 证毕.

这里的 $|A(\lambda)|$ 是非零常数也就是说 $|A(\lambda)|$ 是零次多项式.

定义 3.5 λ -矩阵 $A(\lambda)$ 的三类初等变换定义如下:

- (1) 对换第 i, j 两行 (或列), 记作 $r_i \leftrightarrow r_j$ (或 $c_i \leftrightarrow c_j$);
- (2) 第 i 行 (或列) 乘非零数 k , 记作 $r_i \times k$ (或 $c_i \times k$);
- (3) 第 i 行 (列) 加上第 j 行 (列) 的 $k(\lambda)$ 倍, 记作 $r_i + k(\lambda)r_j$ (或 $c_i + k(\lambda)c_j$), 其中, $k(\lambda)$ 为 λ 的多项式.

类似于数值矩阵, λ -阵 $A(\lambda)$ 也可定义初等 λ -矩阵.

定义 3.6 三类初等 λ -矩阵定义如下:

- (1) 把单位阵 E 对换第 i, j 两行所得矩阵, 记作 $E(i, j)$;
- (2) 把单位阵 E 的第 i 行乘非零数 k 所得矩阵, 记作 $E(i(k))$;
- (3) 把单位阵 E 的第 i 行加上第 j 行的 $k(\lambda)$ 倍所得矩阵, 记作 $E(i, j(k(\lambda)))$.

对一个 λ -矩阵 $A(\lambda)$ 做一次初等行变换相当于用一个对应的初等 λ -矩阵左乘 $A(\lambda)$, 反过来, 对 $A(\lambda)$ 做一次初等列变换相当于用一个对应的初等 λ -矩阵的转置去右乘 $A(\lambda)$. λ -阵 $A(\lambda)$ 可逆的充分必要条件是 $A(\lambda)$ 为有限个初等 λ -矩阵的乘积.

定义 3.7 若 λ -矩阵 $A(\lambda)$ 经有限次初等变换后变成 $B(\lambda)$, 则称 $A(\lambda)$ 为 $B(\lambda)$ 等价, 记作 $A(\lambda) \rightarrow B(\lambda)$.

由上面的结论立即可得到如下定理.

定理 3.2 设 $A(\lambda)$ 与 $B(\lambda)$ 均是 $m \times n$ 阶 λ -矩阵, 则 $A(\lambda) \rightarrow B(\lambda)$ 的充分必要条件是存在 m 阶可逆 λ -矩阵 $P(\lambda)$ 与 n 阶可逆 λ -矩阵 $Q(\lambda)$, 使 $B(\lambda) = P(\lambda)A(\lambda)Q(\lambda)$.

现在的问题是, 一个 λ -矩阵是否能象数字方阵那样, 经有限次初等变换后能变成一个简单的, 并且唯一确定的标准形矩阵呢? 从下面的引理开始, 逐步讨论.

引理 3.3 设 $A(\lambda) = (a_{ij}(\lambda))$ 为 n 阶 λ -矩阵, $a_{11}(\lambda) \neq 0$, 若 $A(\lambda)$ 中存在一个元素不能被 $a_{11}(\lambda)$ 整除, 则必存在与 $A(\lambda)$ 等价的矩阵 $B(\lambda) = (b_{ij}(\lambda))$ 使得 $\deg b_{11}(\lambda) < \deg a_{11}(\lambda)$.

证明 情形 1. 存在与 $a_{11}(\lambda)$ 同列的某个元素 $a_{i1}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 设 $a_{i1}(\lambda) =$

$q(\lambda)a_{11}(\lambda) + r(\lambda)$, 其中 $\deg r(\lambda) < \deg a_{11}(\lambda)$ 则经过下列两次初等行变换

$$A(\lambda) = \begin{pmatrix} a_{11} & \cdots \\ \vdots & \vdots \\ a_{i1} & \cdots \\ \cdots & \cdots \end{pmatrix} \xrightarrow{r_i - qr_1} \begin{pmatrix} a_{11} & \cdots \\ \vdots & \vdots \\ r & \cdots \\ \cdots & \cdots \end{pmatrix} \xrightarrow{r_1 \leftrightarrow r_i} \begin{pmatrix} r & \cdots \\ \vdots & \vdots \\ a_{11} & \cdots \\ \cdots & \cdots \end{pmatrix},$$

即可实现引理的结论.

情形 2. 存在与 $a_{11}(\lambda)$ 同行的某个元素 $a_{1j}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 则类似情形 1, 只需经过下列两次初等列变换, 即可实现引理的结论.

情形 3. 所有与 $a_{11}(\lambda)$ 所在行或列的元素均可被 $a_{11}(\lambda)$ 整除, 但存在某个 $a_{ij}(\lambda)$ 不能被 a_{11} 整除. 此时设 $a_{i1}(\lambda) = q(\lambda)a_{11}(\lambda)$, 则经过下列两次初等行变换

$$A(\lambda) \xrightarrow{r_i - qr_1} \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & a_{ij} - qa_{1j} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \xrightarrow{r_1 + r_i} \begin{pmatrix} a_{11} & \cdots & (1-q)a_{1j} + a_{ij} & \cdots \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & a_{ij} - qa_{1j} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

此时与 a_{11} 同行的 $(1-q)a_{1j} + a_{ij}$ 不能被 a_{11} 整除, 转化为情形 2.

上述 3 种情形已包含了所有可能, 证毕.

引理 3.4 设 $A(\lambda)$ 为 n 阶非零 λ -矩阵, 则必存在与 $A(\lambda)$ 等价的矩阵 $B(\lambda) = (b_{ij}(\lambda))$ 使得 $b_{11}(\lambda)$ 可整除 $B(\lambda)$ 的所有元素.

证明 设 $A(\lambda)$ 的所有非零元素的最低次数为 n , 对 n 采用归纳法.

当 $n = 0$ 时, 即 $A(\lambda)$ 中存在元素为非零常数, 非零常数当然可整除 $A(\lambda)$ 剩余所有元素. 此时只需通过两次第 1 类初等变换, 即可将该元素换至首行首列, 结论得证.

设 $n \leq k$ 时结论成立, 当 $n = k + 1$ 时, 设 $\deg a_{ij}(\lambda) = n$, 若该元素可以整除 $A(\lambda)$ 的剩余所有元素, 同样只需通过两次第 1 类初等变换, 即可将该元素换至首行首列, 结论得证. 否则先通过两次第 1 类初等变换, 将该元素换至首行首列, 再根据引理 3.3, 存在与 $A(\lambda)$ 等价的矩阵 $B(\lambda)$, 使得其首行首列元素的次数低于 $n + 1$, 由归纳假设, 结论得证.

由上述引理, 可得到最终的结论.

定理 3.5 任意一个秩为 r 的 λ -矩阵 $A(\lambda)$ 都等价于一个分块 λ -矩阵

$$\begin{pmatrix} D(\lambda) & 0 \\ 0 & 0 \end{pmatrix},$$

其中

$$D(\lambda) = \begin{pmatrix} d_1(\lambda) & & & \\ & d_2(\lambda) & & \\ & & \ddots & \\ & & & d_r(\lambda) \end{pmatrix},$$

为 r 阶对角阵, $d_k(\lambda)$ ($1 \leq k \leq r$) 是关于 λ 的首项系数为 1 的多项式, 并且 $d_k(\lambda) \mid d_{k+1}(\lambda)$ ($1 \leq i < r-1$), 称这个分块 λ -矩阵为 $\mathbf{A}(\lambda)$ 的等价标准形, 记作 $\mathbf{I}_r(\lambda)$. $d_k(\lambda)$ 称为 $\mathbf{A}(\lambda)$ 的第 k 个不变因子.

证明 对秩 r 进行归纳. 当 $r=0$ 时, 结论显然成立.

设 $r=k$ 时结论成立. 则当 $r=k+1$ 时, 由引理 3.4 可设 a_{11} 整除 $\mathbf{A}(\lambda)$ 的其余所有元素, 因此可通过有限次第 3 类初等变换, 将 $\mathbf{A}(\lambda)$ 的首行、首列的其他元素均变为零, 得到如下情形

$$\begin{pmatrix} a_{11}(\lambda) & \mathbf{0} \\ \mathbf{0} & \mathbf{B}(\lambda) \end{pmatrix}.$$

此时, $\mathbf{B}(\lambda)$ 的所有元素可被 $a_{11}(\lambda)$ 整除, 且其秩为 $r-1=k$, 故可对 $\mathbf{B}(\lambda)$ 使用归纳假设变为定理结论中的对角形矩阵, 再将 a_{11} 的首项系数变为 1, 定理得证.

注 3.1 该证明过程, 尚未说明 λ -矩阵的等价标准形的唯一性, 下节将指出, 这一标准形是唯一的.

定理 3.5 的证明, 同时给出了将 λ -矩阵 $\mathbf{A}(\lambda)$ 通过初等变换化为标准形的一般步骤.

例 3.2 求 λ -矩阵 $\mathbf{A}(\lambda)$ 的等价标准形.

$$\mathbf{A}(\lambda) = \begin{pmatrix} 1-\lambda & 2\lambda-1 & \lambda \\ \lambda & \lambda^2 & -\lambda \\ 1+\lambda^2 & \lambda^3+\lambda-1 & -\lambda^2 \end{pmatrix}.$$

解 第 1 步, 判断 $\mathbf{A}(\lambda)$ 的 $(1,1)$ 元素是否能整除其他元素. 若能, 则用初等变换把第一行和第一列的其余元素都化为零; 若不能, 则利用引理 3.3 和引理 3.4 证明中的方法降低 $(1,1)$ 元素的次数, 直到变化后的矩阵的 $(1,1)$ 元素能整除其他元素为止.

$$\mathbf{A}(\lambda) = \begin{pmatrix} 1-\lambda & 2\lambda-1 & \lambda \\ \lambda & \lambda^2 & -\lambda \\ 1+\lambda^2 & \lambda^3+\lambda-1 & -\lambda^2 \end{pmatrix} \xrightarrow[r_3-\lambda r_2]{r_1+r_2} \begin{pmatrix} 1 & \lambda^2+2\lambda-1 & 0 \\ \lambda & \lambda^2 & -\lambda \\ 1 & \lambda-1 & 0 \end{pmatrix}.$$

第 2 步, 用初等变换将首行、首列的其余元素都化为零.

$$\mathbf{A}(\lambda) \xrightarrow[r_2-\lambda r_1]{r_3-r_1} \begin{pmatrix} 1 & \lambda^2+2\lambda-1 & 0 \\ 0 & -\lambda^3-\lambda^2+\lambda & -\lambda \\ 0 & -\lambda^2-\lambda & 0 \end{pmatrix} \xrightarrow{c_2-(\lambda^2+2\lambda-1)c_1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\lambda^3-\lambda^2+\lambda & -\lambda \\ 0 & -\lambda^2-\lambda & 0 \end{pmatrix}.$$

对去掉首行首列后的矩阵块重复上述过程, 直到化成标准形矩阵.

$$\mathbf{A}(\lambda) \xrightarrow{r_2-\lambda r_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & -\lambda^2-\lambda & 0 \end{pmatrix} \xrightarrow[r_3-(\lambda-1)r_2]{c_3+c_2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda(\lambda+1) \end{pmatrix} = \mathbf{I}_3(\lambda).$$

3.2 λ -矩阵标准形的计算方法

按定理 3.5 的思想方法, 按部就班地计算 λ -矩阵的等价标准形, 在很多情形下, 不够简便. 本节将给出 λ -矩阵在初等变换下的几个不变量. 通过这些不变量, 可以简化求 λ -矩阵的等价标准形的计算过程, 同时也可证明定理 3.5 中 λ -矩阵的等价标准形的唯一性.

定义 3.8 λ -矩阵 $A(\lambda)$ 的所有 k 阶子式的首项系数为 1 的最大公因式称为 $A(\lambda)$ 的 k 阶行列式因子, 记为 $D_k(\lambda)$.

λ -矩阵的行列式因子, 是初等变换下的不变量. 这是因为, 第 1 类初等变换, 只改变某些子式的符号, 第 2 类初等变换, 只会使某些子式变成原来的常数倍, 第 3 类初等变换, 只会是某些子式叠加了另一些子式的倍式, 这三种变换都不会改变子式的最大公因式. 特别, 有下面的结论.

定理 3.6 设 $D_k(\lambda)$ ($1 \leq k \leq r$) 为 λ -矩阵 $A(\lambda)$ 的 k 阶行列式因子, 则 $A(\lambda)$ 的第 k 个不变因子 $d_k(\lambda) = D_k(\lambda)/D_{k-1}(\lambda)$ ($1 \leq k \leq r$), 这里约定 $D_0(\lambda) = 1$. 特别, λ -矩阵的等价标准形唯一.

证明 设 $A(\lambda)$ 的等价标准形为 $I_r(\lambda) = \begin{pmatrix} D(\lambda) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, 其中

$$D(\lambda) = \begin{pmatrix} d_1(\lambda) & & & \\ & d_2(\lambda) & & \\ & & \ddots & \\ & & & d_r(\lambda) \end{pmatrix}.$$

因 $d_k(\lambda) \mid d_{k+1}(\lambda)$ ($1 \leq i \leq r-1$), 故 $I_r(\lambda)$ 的 k 阶子式的最大公因式为 $d_1(\lambda) \cdots d_k(\lambda)$. 又 $I_r(\lambda)$ 与 $A(\lambda)$ 的各阶行列式因子相同, 故 $D_k(\lambda) = d_1(\lambda) \cdots d_k(\lambda)$. 证毕.

在对 λ -矩阵做初等变换, 求等价标准形的过程中, 一旦观察出行列式因子, 应充分利用, 简化计算过程.

例 3.3 求 $A(\lambda) = \begin{pmatrix} \lambda-1 & 1 & 0 \\ -2 & \lambda-4 & 1 \\ 0 & 0 & \lambda-3 \end{pmatrix}$ 的等价标准形.

解 因 $A(\lambda)$ 的元素中有非零常数, 故 $D_1(\lambda) = 1$. 又 $A(\lambda)$ 有二阶子式 $\begin{vmatrix} 1 & 0 \\ \lambda-4 & 1 \end{vmatrix} = 1$,

为非零常数, 故 $D_2(\lambda) = 1$. 而

$$D_3(\lambda) = \begin{vmatrix} \lambda-1 & 1 & 0 \\ -2 & \lambda-4 & 1 \\ 0 & 0 & \lambda-3 \end{vmatrix} = (\lambda-2)(\lambda-3)^3,$$

故 $d_1(\lambda) = d_2(\lambda) = 1$, $d_3(\lambda) = (\lambda - 2)(\lambda - 3)^2$. 即 $A(\lambda)$ 的等价标准形为

$$I_r(\lambda) = \begin{pmatrix} 1 & & \\ & 1 & \\ & & (\lambda - 2)(\lambda - 3)^2 \end{pmatrix}.$$

除了行列式因子和不变因子, λ -矩阵还有一套等价变换下的不变量——初等因子.

定理 3.7 λ -矩阵 $A(\lambda)$ 的全体不变因子在复数域 \mathbf{C} 上有下面的分解式

$$\begin{aligned} d_1(\lambda) &= (\lambda - a_1)^{l_{11}}(\lambda - a_2)^{l_{21}} \cdots (\lambda - a_q)^{l_{q1}}, \\ d_2(\lambda) &= (\lambda - a_1)^{l_{12}}(\lambda - a_2)^{l_{22}} \cdots (\lambda - a_q)^{l_{q2}}, \\ &\dots\dots\dots \\ d_r(\lambda) &= (\lambda - a_1)^{l_{1r}}(\lambda - a_2)^{l_{2r}} \cdots (\lambda - a_q)^{l_{qr}}, \end{aligned}$$

其中 a_1, a_2, \dots, a_q 两两不同, $0 \leq l_{i1} \leq l_{i2} \leq \dots \leq l_{ir}$ ($i = 1, 2, \dots, q$), 当 $l_{ij} > 0$ 时, 因子 $(\lambda - a_i)^{l_{ij}}$ 称为 $A(\lambda)$ 的初等因子.

根据多项式的唯一分解定理和不变因子 $d_i | d_{i+1}$, 立即得到上述定理的证明.

注 3.2 上述定理中, 当某个 a_i 不是 $d_j(\lambda)$ 的零点时, 相应的 l_{ij} 取为 0. 而对 $j \neq k$, 指数 l_{ij} 和 l_{ik} 有可能相等, 此时 $(\lambda - a_i)^{l_{ij}}$ 和 $(\lambda - a_i)^{l_{ik}}$ 是 $A(\lambda)$ 的两个相同的初等因子时. 对相同的初等因子, 应重复记录, 不能略去.

例 3.4 5 阶 λ -矩阵

$$I_r(\lambda) = \begin{pmatrix} 1 & & & & \\ & \lambda - 2 & & & \\ & & (\lambda - 2)(\lambda + 3) & & \\ & & & (\lambda - 2)^2(\lambda + 3)^2 & \\ & & & & 0 \end{pmatrix}$$

全体初等因子为 $(\lambda - 2)$, $(\lambda - 2)$, $(\lambda + 3)$, $(\lambda - 2)^2$, $(\lambda + 3)^2$.

在不计顺序的前提下, $A(\lambda)$ 的全体初等因子由全体不变因子唯一确定. 反之, 如果已知 $A(\lambda)$ 的全体初等因子和秩, 同样能够确定 $A(\lambda)$ 的全体不变因子.

例 3.5 已知 λ -矩阵 $A(\lambda)$ 的秩 4, 全体初等因子为 $(\lambda - 2)$, $(\lambda - 2)$, $(\lambda - 2)^2$, $(\lambda + 3)$, $(\lambda + 3)^2$, 求 $A(\lambda)$ 的全体不变因子.

解 首先取 $A(\lambda)$ 的全部初等因子 $\lambda - 2$, $\lambda - 2$, $(\lambda - 2)^2$, $\lambda + 3$, $(\lambda + 3)^2$ 的最小公倍式作为 $A(\lambda)$ 的第 4 个不变因子, 即 $d_4(\lambda) = (\lambda - 2)^2(\lambda + 3)^2$, 然后再取剩余的初等因子 $\lambda - 2$, $\lambda - 2$, $\lambda + 3$ 的最小公倍式作为 $A(\lambda)$ 的第 3 个不变因子, 即 $d_3(\lambda) = (\lambda - 2)(\lambda + 3)$. 同理得到 $d_2(\lambda) = \lambda - 2$. 这时 $A(\lambda)$ 的全部初等因子都已经使用过了, 于是剩余的不变因子 $d_1(\lambda) = 1$.

综上所述, 可得出以下定理.

定理 3.8 设 $A(\lambda)$ 与 $B(\lambda)$ 为同型的 λ -矩阵, 则以下条件等价:

(1) $A(\lambda)$ 可经过有限次初等变换变为 $B(\lambda)$ (即 $A(\lambda)$ 与 $B(\lambda)$ 等价);

- (2) $A(\lambda)$ 与 $B(\lambda)$ 有相同的不变因子;
 (3) $A(\lambda)$ 与 $B(\lambda)$ 有相同的行列式因子;
 (4) $A(\lambda)$ 与 $B(\lambda)$ 有相同的秩和初等因子.

下面的指出, 初等因子同样可在初等变换的过程中被提前发现.

引理 3.9 设 2 阶 λ -矩阵

$$A_1 = \begin{pmatrix} (\lambda - a)^{m_2} g_1(\lambda) & 0 \\ 0 & (\lambda - a)^{m_1} g_2(\lambda) \end{pmatrix}, \quad A_2 = \begin{pmatrix} (\lambda - a)^{m_1} g_1(\lambda) & 0 \\ 0 & (\lambda - a)^{m_2} g_2(\lambda) \end{pmatrix},$$

其中 $m_1, m_2 \geq 0$, 且 $g_i(a) \neq 0$ ($i = 1, 2$), 则 A_1 与 A_2 等价.

证明 不妨设 $m_1 \leq m_2$, 因为 $g_i(a) \neq 0$ ($i = 1, 2$), 所以 $(\lambda - a)$ 不是 $g_i(\lambda)$ ($i = 1, 2$) 的因子. 故 A_1 和 A_2 的行列式因子均为

$$D_1(\lambda) = (\lambda - a)^{m_1} \gcd(g_1(\lambda), g_2(\lambda)), \quad D_2(\lambda) = (\lambda - a)^{m_1 + m_2} g_1(\lambda) g_2(\lambda),$$

故 A_1 与 A_2 等价. 证毕.

对一般阶数的对角 λ -矩阵

$$A_1 = \begin{pmatrix} (\lambda - a)^{m_1} g_1(\lambda) & & & \\ & (\lambda - a)^{m_2} g_2(\lambda) & & \\ & & \ddots & \\ & & & (\lambda - a)^{m_r} g_r(\lambda) \end{pmatrix},$$

其中 $g_i(a) \neq 0$ ($1 \leq i \leq r$), 根据引理 3.9, 可通过第一类的初等变换将因子 $(\lambda - a)$ 按次数, 由小到大重新排列, 而保持 $g_i(\lambda)$ 的位置不变, 得到

$$A_2 = \begin{pmatrix} (\lambda - a)^{n_1} g_1(\lambda) & & & \\ & (\lambda - a)^{n_2} g_2(\lambda) & & \\ & & \ddots & \\ & & & (\lambda - a)^{n_r} g_r(\lambda) \end{pmatrix},$$

即此处 $0 \leq n_1 \leq n_2 \leq \cdots \leq n_r$ 是 m_1, m_2, \dots, m_r 的排列.

进一步, 还有下面的结论.

定理 3.10 设 λ -矩阵 $A(\lambda)$ 为分块对角阵

$$A(\lambda) = \begin{pmatrix} A_1(\lambda) & & & \\ & A_2(\lambda) & & \\ & & \ddots & \\ & & & A_s(\lambda) \end{pmatrix},$$

则 $A(\lambda)$ 的每个子块 $A_j(\lambda)$ ($1 \leq j \leq s$) 的每个初等因子都是 $A(\lambda)$ 的初等因子, 并且 $A(\lambda)$ 的每个初等因子必是某个 $A_j(\lambda)$ ($1 \leq j \leq s$) 的初等因子.

证明 不妨设 $A_j(\lambda)$ 的等价标准形为

$$I_j(\lambda) = \begin{pmatrix} d_{1j}(\lambda) & & \\ & \ddots & \\ & & d_{r_jj}(\lambda) \end{pmatrix} \quad (1 \leq j \leq s),$$

其中 $d_{ij}(\lambda)$ ($1 \leq i \leq r_j$) 是 $A_j(\lambda)$ 的不变因子, 并且 $r_1 + r_2 + \cdots + r_s = r$, 于是 $A(\lambda)$ 就等价于对角阵

$$B(\lambda) = \begin{pmatrix} d_{11}(\lambda) & & & \\ & \ddots & & \\ & & d_{r_11}(\lambda) & \\ & & & \ddots & \\ & & & & d_{1s}(\lambda) \\ & & & & & \ddots \\ & & & & & & d_{r_ss}(\lambda) \end{pmatrix}.$$

此时 $B(\lambda)$ 的对角元不一定满足由上到下依次整除的条件, 故不一定是 $A(\lambda)$ 的标准形. 但由引理 3.9 的结论, 可将所有 $A_j(\lambda)$ ($1 \leq j \leq s$) 的全部初等因子按指数从小到大重排, 将 $B(\lambda)$ 等价地变为标准形 $I_r(\lambda)$. 在该过程中, $A_j(\lambda)$ 的初等因子未发生变化, 故 $A_j(\lambda)$ 的初等因子全体构成了 $A(\lambda)$ 的初等因子全体. 证毕.

由定理 3.10, 在用初等变换求 λ -矩阵 $A(\lambda)$ 的标准形的过程中, 一旦 $A(\lambda)$ 变成分块对角阵, 可以先求每个子块的标准形, 通过每个子块的初等因子, 得到 $A(\lambda)$ 的初等因子, 进而得到 $A(\lambda)$ 的标准形.

例 3.6 求 $A(\lambda) = \begin{pmatrix} 1-\lambda & 2\lambda-1 & \lambda \\ \lambda & \lambda^2 & -\lambda \\ 1+\lambda^2 & \lambda^3+\lambda-1 & -\lambda^2 \end{pmatrix}$ 的等价标准形.

解 对 $A(\lambda)$ 做初等变换得

$$\begin{aligned} A(\lambda) &\xrightarrow[r_1+r_2]{r_3-\lambda r_2} \begin{pmatrix} 1 & \lambda^2+2\lambda-1 & 0 \\ \lambda & \lambda^2 & -\lambda \\ 1 & \lambda-1 & 0 \end{pmatrix} \xrightarrow[r_2-\lambda r_1]{r_3-r_1} \begin{pmatrix} 1 & \lambda^2+2\lambda-1 & 0 \\ 0 & -\lambda^3-\lambda^2+\lambda & -\lambda \\ 0 & -\lambda^2-\lambda & 0 \end{pmatrix} \\ &\xrightarrow{c_2-(\lambda^2+2\lambda-1)c_1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\lambda^3-\lambda^2+\lambda & -\lambda \\ 0 & -\lambda^2-\lambda & 0 \end{pmatrix} = \begin{pmatrix} 1 & & \\ & A_1(\lambda) & \end{pmatrix}. \end{aligned}$$

对于 $A_1(\lambda)$, 容易观察出其 1 阶行列式因子 $D_1(\lambda) = \lambda$, 而 $D_2(\lambda) = \lambda^2(\lambda+1)$, 从而其不变因子 λ 和 $\lambda(\lambda+1)$, 从而其初等因子为 $\lambda, \lambda, (\lambda+1)$, 这也是 $A(\lambda)$ 的初等因子. 于是 $A(\lambda)$ 的

等价标准形为

$$I_3(\lambda) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda(\lambda+1) \end{pmatrix}.$$

注 3.3 计算不变因子、行列式因子的过程，本质上只涉及数的四则运算，因此是可精确计算的，但寻求初等因子，涉及到多项式的求根，只有在非常特殊的情形下才可精确计算。

3.3 矩阵的 Jordan 标准形

有了前面的准备，本节将介绍数值矩阵相似的充分必要条件，其中最本质的是如下定理。

定理 3.11 数字方阵 A 与 B 相似的充要条件是 $\lambda E - A$ 与 $\lambda E - B$ 等价。

证明 必要性. 设 A 与 B 相似，即存在可逆阵 P 使得 $P^{-1}AP = B$ ，则有

$$\lambda E - B = \lambda E - P^{-1}AP = P^{-1}(\lambda E - A)P,$$

即 $\lambda E - A$ 与 $\lambda E - B$ 等价。

充分性. 设 $\lambda E - A$ 与 $\lambda E - B$ 等价，则存在可逆的 λ -矩阵 $P(\lambda)$ 和 $Q(\lambda)$ ，使得

$$P(\lambda)(\lambda E - A)Q(\lambda) = \lambda E - B,$$

故有 $P(\lambda)(\lambda E - A) = (\lambda E - B)Q(\lambda)^{-1}$ 。下面先把 $P(\lambda)$ 和 $Q(\lambda)^{-1}$ 分别变形为

$$P(\lambda) = (\lambda E - B)P_1(\lambda) + R_1, \quad Q(\lambda)^{-1} = Q_1(\lambda)(\lambda E - A) + R_2,$$

其中 R_1, R_2 都是数值矩阵。带入前式，再适当变形可以得到

$$(\lambda E - B)[P_1(\lambda) - Q_1(\lambda)](\lambda E - A) = (\lambda E - B)R_2 - R_1(\lambda E - A),$$

比较两端的次数得 $P_1(\lambda) = Q_1(\lambda)$ ，于是 $R_1(\lambda E - A) = (\lambda E - B)R_2$ 。再变形成 $\lambda(R_1 - R_2) = R_1A - BR_2$ ，就有 $R_1 = R_2$ ，故 $R_1A = BR_1$ 。只需再证明 R_1 是可逆阵。

设 $P(\lambda)^{-1} = (\lambda E - B)P_2(\lambda) + R_0$ ，其中 R_0 是只含常数的矩阵，则有

$$P(\lambda)P(\lambda)^{-1} = [(\lambda E - B)P_1(\lambda) + R_1][(\lambda E - B)P_2(\lambda) + R_0]$$

$$= (\lambda E - B)P_1(\lambda)P(\lambda)^{-1} + R_1(\lambda E - B)P_2(\lambda) + R_1R_0,$$

即

$$E = (\lambda E - B)P_1(\lambda)P(\lambda)^{-1} + R_1(\lambda E - B)P_2(\lambda) + R_1R_0.$$

比较两端的次数有 $R_1R_0 = E$ ，即 R_1 是可逆阵，因此 A 与 B 相似。证毕。

结合定理 3.11 以及上一节的结论就有如下推论。

定义 3.9 设 A 为数值方阵，则 λ -矩阵 $\lambda E - A$ 的不变因子和初等因子称为数值矩阵 A 的不变因子和初等因子。

推论 3.12 设 A 与 B 为数值方阵，则下列命题等价：

(1) A 与 B 相似；

- (2) $\lambda E - A$ 与 $\lambda E - B$ 等价;
 (3) A 与 B 具有相同的行列式因子;
 (4) A 与 B 具有相同的不变因子;
 (5) A 与 B 具有相同的初等因子.

证明 只需注意 (5), 不再需要秩的条件, 这是因为任何矩阵的特征多项式必定非零多项式, 故相应的 λ -矩阵必定是满秩的. 证毕.

注 3.4 若将矩阵 B 看作矩阵 A 的相似变换, 则上述推论指出, 矩阵的特征值、行列式因子、初等因子、不变因子, 都是在相似变换下保持不变的, 这就是不变量的观点. 数学的很多研究都是围绕寻求变换下的不变量展开的.

例 3.7 求 n_0 -阶方阵 $J_0 = \begin{pmatrix} \lambda_0 & 1 & & \\ & \lambda_0 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_0 \end{pmatrix}$ 的初等因子.

解 因为

$$\lambda E - J_0 = \begin{pmatrix} \lambda - \lambda_0 & -1 & & \\ & \lambda - \lambda_0 & \ddots & \\ & & \ddots & -1 \\ & & & \lambda - \lambda_0 \end{pmatrix},$$

易见存在 J_0 $(n_0 - 1) - 1$ 子式为非常数, 故行列式因子 $D_1(\lambda) = \cdots = D_{n_0-1}(\lambda) = 1$, 而 $D_{n_0}(\lambda) = (\lambda - \lambda_0)^{n_0}$, 从而 J_0 的不变因子为 $d_1(\lambda) = \cdots = d_{n_0-1}(\lambda) = 1$, $d_{n_0}(\lambda) = (\lambda - \lambda_0)^{n_0}$, 所以 J_0 的初等因子为 $(\lambda - \lambda_0)^{n_0}$.

以下考虑矩阵在复相似变换下的标准形.

定义 3.10 分块矩阵

$$J = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_s \end{pmatrix}$$

称为 Jordan¹ 标准形, 其中

$$J_k = \begin{pmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{pmatrix} \quad (1 \leq k \leq s),$$

¹Jordan (1838–1922), 若当, 法国数学家.

为 n_k 阶上三角矩阵, 称为一个 **Jordan 块**.

根据例 3.7 和定理 3.10 的结论可以得到, Jordan 标准形 J 的初等因子全体为 $(\lambda - \lambda_1)^{n_1}, (\lambda - \lambda_2)^{n_2}, \dots, (\lambda - \lambda_s)^{n_s}$, 结合推论 3.12, 就得到本章的核心定理.

定理 3.13 设 n 阶矩阵 A 的初等因子是 $(\lambda - \lambda_1)^{n_1}, (\lambda - \lambda_2)^{n_2}, \dots, (\lambda - \lambda_s)^{n_s}$, 则得 A 相似于 Jordan 标准形

$$J = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_s \end{pmatrix},$$

其中 J_k 为 n_k 阶 Jordan 块 ($1 \leq k \leq s$), $n_1 + n_2 + \dots + n_s = n$.

特别, J 为对角阵, 当且仅当每个 Jordan 块都是 1 阶阵, 亦即下面的推论.

推论 3.14 n 阶矩阵 A 相似于对角阵的充分必要条件是 A 的初等因子都是一次多项式.

通过下面的例题, 介绍如何计算矩阵 A 的 Jordan 标准形 J , 以及相应的过渡矩阵 P , 使得 $A = PJP^{-1}$.

例 3.8 求方阵

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 4 & -1 \\ 0 & 0 & 3 \end{pmatrix}.$$

的 Jordan 标准形 J , 以及过渡矩阵 P , 使得 $A = PJP^{-1}$.

解 因为

$$\begin{aligned} \lambda E - A &= \begin{pmatrix} \lambda - 1 & 1 & 0 \\ -2 & \lambda - 4 & 1 \\ 0 & 0 & \lambda - 3 \end{pmatrix} \xrightarrow{c_1 \leftrightarrow c_2} \begin{pmatrix} 1 & \lambda - 1 & 0 \\ \lambda - 4 & -2 & 1 \\ 0 & 0 & \lambda - 3 \end{pmatrix} \\ &\xrightarrow{\substack{c_2 - (\lambda - 1)c_1 \\ r_2 - (\lambda - 4)r_1}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\lambda^2 + 5\lambda - 6 & 1 \\ 0 & 0 & \lambda - 3 \end{pmatrix}, \end{aligned}$$

此时可以观察出, 右下矩阵块的 1 阶行列式因子为 1, 2 阶行列式因子为 $(\lambda - 2)(\lambda - 3)^2$, 从而矩阵 A 的初等因子为 $(\lambda - 2), (\lambda - 3)^2$, A 的 Jordan 标准形为

$$J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}.$$

为求过渡矩阵 P , 设 $P = (p_1 \ p_2 \ p_3)$, 这里 p_1, p_2, p_3 均为列矩阵. 由 $A = PJP^{-1}$, 得

$$A(p_1 \ p_2 \ p_3) = (p_1 \ p_2 \ p_3) \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}.$$

即 $Ap_1 = 2p_1, Ap_2 = 3p_2, Ap_3 = p_2 + 3p_3$.

解方程组 $(A - 2E)p_1 = 0$, 得 $p_1 = (1 \ -1 \ 0)^T$.

解方程组 $(A - 3E)p_2 = 0$, 得 $p_2 = (1 \ -2 \ 0)^T$.

解方程组 $(A - 3E)p_3 = p_2$, 得 $p_3 = (0 \ -1 \ 1)^T$.

于是

$$P = (p_1 \ p_2 \ p_3) = \begin{pmatrix} 1 & 1 & 0 \\ -1 & -2 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

当 Jordan 标准形中不同 Jordan 块存在相同的特征值时, 计算过程要复杂一些.

例 3.9 设 $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{pmatrix}$, 求 A 的 Jordan 标准形 J 和过渡矩阵 P 使得 $A = PJP^{-1}$.

解 因为

$$\begin{aligned} \lambda E - A &= \begin{pmatrix} \lambda - 2 & 0 & 0 \\ -1 & \lambda - 1 & -1 \\ -1 & 1 & \lambda - 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -(\lambda - 1) & 1 \\ \lambda - 2 & 0 & 0 \\ -1 & 1 & \lambda - 3 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\lambda - 1)(\lambda - 2) & -(\lambda - 2) \\ 0 & -(\lambda - 2) & \lambda - 2 \end{pmatrix}, \end{aligned}$$

此时可以观察出, 右下矩阵块的 1 阶行列式因子为 $(\lambda - 2)$, 2 阶行列式因子为 $(\lambda - 2)^3$, 故 A 的初等因子为 $(\lambda - 2)$, $(\lambda - 2)^2$, 从而 A 的 Jordan 标准形为

$$J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

为求过渡矩阵 P , 设 $P = (p_1 \ p_2 \ p_3)$, 这里 p_1, p_2, p_3 为列矩阵. 由 $AP = PJ$ 得

$$(Ap_1 \ Ap_2 \ Ap_3) = (2p_1 \ 2p_2 \ p_2 + 2p_3).$$

解方程组 $(A - 2E)X = 0$, 得基础解系 $X_1 = (1 \ 1 \ 0)^T$, $X_2 = (1 \ 0 \ -1)^T$, 因 p_1 和 p_2 是 $(A - 2E)X = 0$ 的 2 个线性无关的解, 并且 p_2 还需要使关于 p_3 的方程 $(A - 2E)p_3 = p_2$ 有

解. 于是设 $p_2 = k_1 X_1 + k_2 X_2$, 则

$$(A - 2E)p_2 = \begin{pmatrix} 0 & 0 & 0 & k_1 + k_2 \\ 1 & -1 & 1 & k_1 \\ 1 & -1 & 1 & -k_2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -1 & 1 & -k_2 \\ 0 & 0 & 0 & k_1 + k_2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

当 $k_1 + k_2 = 0$ 时, $(A - 2E)p_3 = p_2$ 有解, 取 $k_1 = 1, k_2 = -1$, 即 $p_2 = X_1 - X_2 = (0 \ 1 \ 1)^T$, 得 $p_3 = (1 \ 0 \ 0)^T$. 再取 $p_1 = X_1 = (1 \ 1 \ 0)^T$ (只须与 p_2 线性无关即可), 得到

$$P = (p_1 \ p_2 \ p_3) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

注 3.5 一般情形下, 由于矩阵的初等因子并不能精确求出, 因此也就无法精确得到 Jordan 标准形和相应的过渡矩阵, 是否存在有效的数值计算方法, 成为一个重要的问题.

注 3.6 一般情况下, 矩阵的 Jordan 标准形和过渡矩阵都是复矩阵, 因此 Jordan 标准形只解决了矩阵复相似变换下的标准形问题. 如果过渡矩阵限制为实矩阵, 结论要复杂一些, 读者可参考本章习题或文献 [3].

3.4 矩阵的零化多项式和最小多项式

本节介绍矩阵的零化多项式与最小多项式, 与矩阵的 Jordan 标准形有关联, 同时在本书后续的具体问题中有重要应用.

首先给出多项式在矩阵上的取值的定义.

定义 3.11 设多项式 $\varphi(\lambda) = a_0 \lambda^n + a_1 \lambda^{n-1} + \cdots + a_{n-1} \lambda + a_n$, A 为方阵, 定义

$$\varphi(A) = a_0 A^n + a_1 A^{n-1} + \cdots + a_{n-1} A + a_n E,$$

称为多项式 $\varphi(\lambda)$ 在矩阵 A 上的取值.

多项式在矩阵上的取值, 本质即为将多项式的文字符号替换为矩阵, 其中矩阵的零次方约定为单位矩阵.

对取定的矩阵, 有一类多项式非常特殊, 即如下定义.

定义 3.12 设 A 是 n 阶方阵, 若存在多项式 $\varphi(\lambda)$, 使得 $\varphi(A) = \mathbf{0}$, 则称 $\varphi(\lambda)$ 为 A 的一个零化多项式.

例 3.10 设矩阵 $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, 验证 $\varphi(\lambda) = \lambda^3 - 3\lambda^2 + 4$ 是 A 的一个零化多项式.

证明 直接计算, 有

$$A^3 - 3A^2 + 4E = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}^3 - 3 \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}^2 + 4 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{0},$$

于是多项式 $\varphi(\lambda) = \lambda^3 - 3\lambda^2 + 4$ 是 A 的一个零化多项式.

下面的定理指出, 数域 \mathbf{F} 上的方阵, 在 $\mathbf{F}(\lambda)$ 中必定存在零化多项式.

定理 3.15 (Cayley-Hamilton¹) 矩阵 \mathbf{A} 的特征多项式 $f(\lambda) = |\lambda\mathbf{E} - \mathbf{A}|$ 是 \mathbf{A} 的一个零化多项式.

本定理不能简单地将 \mathbf{A} 代入右端, 因为此时右端算出的结果是一个数, 而关于 λ 的多项式 $|\lambda\mathbf{E} - \mathbf{A}|$ 在矩阵 \mathbf{A} 上的取值是矩阵.

证明 设 $f(\lambda) = |\lambda\mathbf{E} - \mathbf{A}| = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$ 是方阵 \mathbf{A} 的特征多项式, $(\lambda\mathbf{E} - \mathbf{A})^*$ 是特征矩阵 $\lambda\mathbf{E} - \mathbf{A}$ 的伴随矩阵. 由于 $(\lambda\mathbf{E} - \mathbf{A})^*$ 的每个元素都是 $\lambda\mathbf{E} - \mathbf{A}$ 的某个元素的代数余子式, 从而至多是 λ 的 $n-1$ 次多项式, 故可设

$$(\lambda\mathbf{E} - \mathbf{A})^* = \mathbf{B}_0\lambda^{n-1} + \mathbf{B}_1\lambda^{n-2} + \cdots + \mathbf{B}_{n-2}\lambda + \mathbf{B}_{n-1},$$

其中 \mathbf{B}_j ($j = 0, 1, \cdots, n-1$) 是 n 阶数字方阵, 于是

$$\begin{aligned} (\lambda\mathbf{E} - \mathbf{A})(\lambda\mathbf{E} - \mathbf{A})^* &= (\lambda\mathbf{E} - \mathbf{A})(\mathbf{B}_0\lambda^{n-1} + \mathbf{B}_1\lambda^{n-2} + \cdots + \mathbf{B}_{n-2}\lambda + \mathbf{B}_{n-1}) \\ &= \mathbf{B}_0\lambda^n + (\mathbf{B}_1 - \mathbf{A}\mathbf{B}_0)\lambda^{n-1} + \cdots + (\mathbf{B}_{n-1} - \mathbf{A}\mathbf{B}_{n-2})\lambda + \mathbf{A}\mathbf{B}_{n-1}. \end{aligned}$$

另一方面,

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

比较这两个等式 λ 的同次乘幂的系数矩阵, 可得

$$\begin{cases} \mathbf{B}_0 = \mathbf{E}, \\ \mathbf{B}_1 - \mathbf{A}\mathbf{B}_0 = a_1\mathbf{E}, \\ \cdots \cdots \cdots \\ \mathbf{B}_{n-1} - \mathbf{A}\mathbf{B}_{n-2} = a_{n-1}\mathbf{E}, \\ -\mathbf{A}\mathbf{B}_{n-1} = a_n\mathbf{E}, \end{cases}$$

分别用 $\mathbf{A}^n, \mathbf{A}^{n-1}, \cdots, \mathbf{A}, \mathbf{E}$ 从上至下依次左乘这些等式的两边后再相加, 得

$$\mathbf{0} = \mathbf{A}^n + a_1\mathbf{A}^{n-1} + \cdots + a_{n-1}\mathbf{A} + a_n\mathbf{E},$$

即 $f(\mathbf{A}) = \mathbf{0}$. 证毕.

类似于数的零化多项式, 同样有下述定理和定义.

定理 3.16 若 $f(x)$ 与 $g(x)$ 同为方阵 \mathbf{A} 的零化多项式, 则 $\gcd(f(x), g(x))$ 也是 \mathbf{A} 的零化多项式.

定义 3.13 方阵 \mathbf{A} 的所有零化多项式的最大公因式, 称为 \mathbf{A} 的最小多项式.

定义 3.14 设 \mathbf{A} 为 n 阶方阵, 则称 \mathbf{A} 的首项系数为 1 的次数最低的零化多项式为 \mathbf{A} 的最小多项式, 记作 $m_{\mathbf{A}}(\lambda)$.

下面给出最小多项式的一些性质.

定理 3.17 设 \mathbf{A} 为 n 阶方阵, $f(\lambda)$ 为方阵 \mathbf{A} 的特征多项式, $m_{\mathbf{A}}(\lambda)$ 为 \mathbf{A} 的最小多项式, 则

- (1) $m_{\mathbf{A}}(\lambda)$ 在相差一个常数因子的意义下唯一;
- (2) \mathbf{A} 的一切特征值必为 $m_{\mathbf{A}}(\lambda)$ 的根.

¹Hamilton (1805–1865), 哈密尔顿, 英国数学家.

证明 (1) 设 $m_1(\lambda)$, $m_2(\lambda)$ 都是 A 的最小多项式, 则 $m_1(\lambda) \mid m_2(\lambda)$, $m_2(\lambda) \mid m_1(\lambda)$, 故存在常数 c , 使得 $m_1(\lambda) = cm_2(\lambda)$.

(2) 设 λ 是 A 的特征值, 则存在 A 的特征向量 α 使得 $A\alpha = \lambda\alpha$, 于是 $m_A(A)\alpha = m_A(\lambda)\alpha$. 而 $m_A(A)\alpha = 0\alpha = 0$, 故 $m_A(\lambda) = 0$.

定理 3.18 若矩阵 A 与 B 相似, 则 A 与 B 的零化多项式相同, 特别最小多项式也相同.

证明 设 $A = PBP^{-1}$, 则对一切多项式 $f(\lambda)$, 有 $f(A) = Pf(B)P^{-1}$, 从而 $f(A) = 0$ 当且仅当 $f(B) = 0$. 证毕.

下面的定理指出, 方阵的最小多项式, 就是该方阵的最后一个不变因子. 为此先证明两个引理.

引理 3.19 对 r -阶方阵

$$H = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix},$$

当 $l < r$ 时,

$$H^l = \begin{pmatrix} \overbrace{0 \cdots 0}^{l \text{ 个}} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{pmatrix},$$

当 $l \geq r$ 时, $H^l = 0$.

证明 当 $l < r$ 时, 用数学归纳法证明. 当 $l = 1$ 时, 结论显然成立. 设当 $l = k - 1$ 时有

$$H^{k-1} = \begin{pmatrix} \overbrace{0 \cdots 0}^{(k-1) \text{ 个}} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{pmatrix}.$$

则

$$H^k = \begin{pmatrix} \overbrace{0 \cdots 0}^{(k-1) \text{ 个}} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} = \begin{pmatrix} \overbrace{0 \cdots 0}^{k \text{ 个}} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{pmatrix},$$

故 $l < r$ 时结论成立. 特别

$$H^{r-1} = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ & \ddots & & 0 \\ & & \ddots & \vdots \\ & & & 0 \end{pmatrix}$$

于是

$$H^r = H^{r-1} \cdot H = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ & \ddots & & 0 \\ & & \ddots & \vdots \\ & & & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} = \mathbf{0},$$

故当 $l \geq r$ 时有 $H^l = \mathbf{0}$. 证毕.

由上述引理立即得到下面的推论.

推论 3.20 r 阶 Jordan 块

$$J = \begin{pmatrix} \lambda_0 & 1 & & \\ & \lambda_0 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_0 \end{pmatrix}$$

的最小多项式 $m_J(\lambda) = (\lambda - \lambda_0)^r$.

证明 由引理 3.19 知 $(\lambda - \lambda_0)^r$ 是 J 的一个零化多项式, 故 $m_J(\lambda) \mid (\lambda - \lambda_0)^r$, 从而 $m_J(\lambda) = (\lambda - \lambda_0)^s$ ($s \leq r$). 再根据引理 3.19, 当 $s < r$ 时, $(J - \lambda_0 E)^s \neq \mathbf{0}$, 得 $s = r$. 证毕.

引理 3.21 分块对角阵

$$A = \begin{pmatrix} A_1 & & \\ & A_2 & \\ & & \ddots \\ & & & A_s \end{pmatrix},$$

的最小多项式是 A_1, A_2, \dots, A_s 的最小多项式的最小公倍式.

证明 先指出一个基本事实: 对任何多项式 $f(\lambda)$,

$$f(A) = \begin{pmatrix} f(A_1) & & \\ & f(A_2) & \\ & & \ddots \\ & & & f(A_s) \end{pmatrix}.$$

由于 $\text{lcm}(m_{A_1}(\lambda), m_{A_2}(\lambda), \dots, m_{A_s}(\lambda))$ 是每一个 $m_{A_j}(\lambda)$ 的倍式, 故是 A_j 的零化多项式, 从而是 A 的零化多项式, 于是

$$m_A(\lambda) \mid \text{lcm}(m_{A_1}(\lambda), m_{A_2}(\lambda), \dots, m_{A_s}(\lambda)).$$

又对一切 j , $m_A(A_j) = \mathbf{0}$, 即 $m_A(\lambda)$ 是 A_j 的零化多项式, 从而是 $m_{A_j}(\lambda)$ 的倍式. 于是 $m_A(\lambda)$ 是 $m_{A_1}(\lambda), m_{A_2}(\lambda), \dots, m_{A_s}(\lambda)$ 的公倍式, 故

$$\text{lcm}(m_{A_1}(\lambda), m_{A_2}(\lambda), \dots, m_{A_s}(\lambda)) \mid m_A(\lambda).$$

证毕.

定理 3.22 n -阶矩阵 A 的最小多项式, 为矩阵 $\lambda E - A$ 的第 n 个不变因子.

证明 根据定理 3.18, 只需考虑 A 的 Jordan 标准形 J 的最小多项式. 再根据引理 3.21, J 的最小多项式为各 Jordan 块的最小多项式的最小公倍式, 这正是矩阵 $\lambda E - A$ 的第 n 个不变因子. 证毕.

定理 3.23 方阵 A 相似于对角阵的充分必要条件是 $\gcd(m_A(\lambda), m'_A(\lambda)) = 1$.

证明 由推论 3.14 可知, 方阵 A 相似于对角阵的充分必要条件是 A 的初等因子都是 λ 的一次多项式, 而 A 的初等因子都是 λ 的一次多项式当且仅当 A 的第 n 个不变因子无重根, 而 A 的第 n 个不变因子恰是 A 的最小多项式. 再根据定理 2.27 立得结论. 证毕.

根据上一节的推论 3.14, 判断一个矩阵是否可相似对角化, 需要证明该矩阵没有超过 1 次的初等因子, 但由于一般的多项式无法精确求根, 因此无法通过因式分解判断初等因子的次数. 而定理 3.23 给出的充分必要条件, 可以计算实现.

例 3.11 设 $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 2 & -1 & 3 \end{pmatrix}$, 判断 A 的 Jordan 标准形是否为对角阵.

解 通过观察易见,

$$\lambda E - A = \begin{pmatrix} \lambda - 2 & -1 & 0 \\ -1 & \lambda - 1 & -1 \\ -2 & 1 & \lambda - 3 \end{pmatrix},$$

存在 1 阶、2 阶子式为 nonzero 常数, 故其最小多项式

$$m_A(\lambda) = \det(\lambda E - A) = \lambda^3 - 6\lambda^2 + 11\lambda - 7.$$

该多项式的零点可以通过 3 次方程的求根公式计算, 运算较繁, 但由于

$$\gcd(m_A(\lambda), m'_A(\lambda)) = \gcd(\lambda^3 - 6\lambda^2 + 11\lambda - 7, 3\lambda^2 - 12\lambda + 11) = 1,$$

故 A 的 Jordan 标准形为对角阵.

类似数的最小多项式, 若两个多项式关于某矩阵的最小多项式同余, 则两个多项式在该矩阵上的值相同. 故计算多项式在矩阵上的取值时, 应充分利用矩阵的最小多项式, 将原多项式降次.

例 3.12 已知 $A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & 3 \end{pmatrix}$, 求 A^{20} .

解 先求 A 的最小多项式. 因为

$$\begin{aligned} \lambda E - A &= \begin{pmatrix} \lambda - 3 & -1 & -1 \\ 1 & \lambda - 1 & 1 \\ -1 & -1 & \lambda - 3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & \lambda - 1 & 1 \\ \lambda - 1 & -1 & -1 \\ -1 & -1 & \lambda - 3 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\lambda^2 + 4\lambda - 4 & -\lambda + 2 \\ 0 & \lambda - 2 & \lambda - 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda - 2 & 0 \\ 0 & 0 & (\lambda - 2)(\lambda - 3) \end{pmatrix}, \end{aligned}$$

所以 A 的最小多项式 $m_A(\lambda) = (\lambda - 2)(\lambda - 3)$. 根据例 2.9,

$$\lambda^{20} \equiv (3^{20} - 2^{20})x + (3 \cdot 2^{20} - 2 \cdot 3^{20}) \pmod{(\lambda - 2)(\lambda - 3)},$$

故

$$\begin{aligned} A^{20} &= (3^{20} - 2^{20})A + (3 \cdot 2^{20} - 2 \cdot 3^{20})E = 3^{20}(A - 2E) - 2^{20}(A - 3E) \\ &= 3^{20} \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{pmatrix} - 2^{20} \begin{pmatrix} 0 & 1 & 1 \\ -1 & -2 & -1 \\ 1 & 1 & 0 \end{pmatrix}. \end{aligned}$$

习题 3

1. 利用初等变换把下列 λ 矩阵化为等价标准形.

$$(1) \begin{pmatrix} \lambda-1 & -\lambda^2 & \lambda^2 \\ \lambda & \lambda & \lambda \\ \lambda^2+1 & \lambda^2 & \lambda^2 \end{pmatrix} \quad (2) \begin{pmatrix} \lambda+1 & \lambda^2+1 & \lambda^2 \\ \lambda-1 & \lambda^2-1 & \lambda \\ 3\lambda-1 & 3\lambda^2-1 & \lambda^2+2\lambda \end{pmatrix}$$

2. 求下列 λ 矩阵的行列式因子、不变因子和初等因子.

$$(1) \begin{pmatrix} \lambda+1 & -2 & 2 \\ 1 & \lambda-2 & 1 \\ -1 & 1 & \lambda-2 \end{pmatrix} \quad (2) \begin{pmatrix} \lambda+1 & -3 & -6 \\ 0 & \lambda-3 & 8 \\ 0 & 2 & \lambda+5 \end{pmatrix}$$

$$(3) \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda-1 & 0 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix} \quad (4) \begin{pmatrix} \lambda & 5 & 1 & 14 \\ 0 & \lambda+1 & 0 & 4 \\ -1 & -1 & \lambda-2 & -6 \\ 0 & -1 & 0 & \lambda-3 \end{pmatrix}$$

3. 设 6 阶 λ 矩阵 $A(\lambda)$ 的秩为 5, 其初等因子是 $\lambda-1$, $\lambda-1$, $(\lambda-2)^3$, $\lambda+2$, $(\lambda+2)^2$, 求 $A(\lambda)$ 的行列式因子、不变因子, 以及 $A(\lambda)$ 的等价标准形.

4. 设 7 阶 λ 矩阵 $A(\lambda)$ 的秩为 6, 其初等因子是 λ , λ , λ^2 , $\lambda-2$, $(\lambda-2)^3$, $(\lambda-2)^3$, 求 $A(\lambda)$ 的行列式因子、不变因子, 以及 $A(\lambda)$ 的等价标准形.

5. 证明: 两个等价的 n 阶 λ 矩阵的行列式只相差一个常数因子.

6. 求下列矩阵的 Jordan 标准形.

$$(1) \begin{pmatrix} 1 & -2 & 2 \\ 1 & -2 & 1 \\ -1 & 1 & -2 \end{pmatrix} \quad (2) \begin{pmatrix} 1 & -3 & 3 \\ -2 & -6 & 13 \\ -1 & -4 & 8 \end{pmatrix} \quad (3) \begin{pmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{pmatrix} \quad (4) \begin{pmatrix} 3 & -4 & 0 & 0 \\ 4 & -5 & -2 & 4 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 2 & -1 \end{pmatrix}$$

7. 证明: 矩阵 A 是幂零阵 (即存在正整数 k , 使得 $A^k = \mathbf{0}$) 当且仅当 A 的特征值都等于零.

8. 设非零矩阵 A 是幂零阵, 证明: A 不相似于对角阵.

9. 求 3 阶幂零阵的全部可能的 Jordan 标准形.

10. 求 3 阶幂等阵 (即满足 $A^2 = A$ 的矩阵 A) 的全部可能的 Jordan 标准形.

11. 设 3 阶矩阵 A 满足 $A^2 = E$, 求 A 的全部可能的 Jordan 标准形.

12. 设 n 阶矩阵 A 满足 $A^2 = E$, 证明: A 相似于对角阵.

13. 设 n 阶矩阵 A 满足 $A^3 - A = 10E$, 证明: A 相似于对角阵.

14. 分别求单位矩阵和零矩阵的最小多项式.

15. 求下列矩阵的最小多项式.

$$(1) \begin{pmatrix} 3 & 0 & 1 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad (2) \begin{pmatrix} 4 & -5 & -6 \\ -2 & 7 & 7 \\ 2 & -5 & -4 \end{pmatrix}$$

16. 设 n 阶方阵 $R = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}$.

(1) 求 R 的不变因子全体.

(2) 计算 Re_j , 这里 $e_j (1 \leq j \leq n)$ 为单位列矩阵.

(3) 设 $f(\lambda) = \lambda^m + b_{m-1}\lambda^{m-1} + \cdots + b_1\lambda + b_0$, 求 $f(R)e_1$.

17. 设 A 为 n 阶方阵, $\varphi(\lambda)$ 为非常值多项式. 证明:

(1) 若 $\varphi(\lambda) \mid m_A(\lambda)$, 则 $\varphi(A)$ 不可逆;

(2) 设 $d(\lambda) = \gcd(\varphi(\lambda), m_A(\lambda))$, 则 $\text{rank } \varphi(A) = \text{rank } d(A)$;

(3) $\varphi(A)$ 可逆当且仅当 $\varphi(\lambda)$ 与 $m_A(\lambda)$ 互素.

18. 设 J_r 为以 λ_0 为特征值的 r 阶 Jordan 块, $\varphi(\lambda)$ 为多项式, 且 $(\lambda - \lambda_0)^r \nmid \varphi(\lambda)$. 证明: 矩阵 $\varphi(J_r)$ 最后一列的元素不全为零.

19. 设 A 为 n 阶非零复方阵, $d = \deg m_A(\lambda)$.

(1) 证明: 对一切 $n \times 1$ 的列矩阵 x , 都有 $x, Ax, \cdots, A^{d-1}x$ 线性相关.

(2) 证明: 对一切正整数 $k < d$, 都存在列矩阵 x , 使得 $x, Ax, \cdots, A^{k-1}x$ 线性无关.

(3) 当 A 为实方阵时, 是否存在实的列矩阵 x , 使得 (2) 成立?

第 4 章 矩阵函数

当矩阵的元素为依赖于某些变量的函数时，关于矩阵的微积分就产生了。本章将从矩阵的极限出发，依次引出矩阵关于变量的导数与积分、矩阵幂级数等概念，并给出矩阵幂级数在求解微分方程中的应用。

4.1 矩阵的极限、导数和积分

类似数列可以定义极限、函数可以定义导数和积分，对矩阵构成的序列、元素为函数的矩阵，同样可以定义极限、导数和积分。

定义 4.1 任给 $m \times n$ 矩阵序列 $\{A_k\}$ ，其中

$$A_k = \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \cdots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \cdots & a_{2n}^{(k)} \\ \vdots & \vdots & & \vdots \\ a_{m1}^{(k)} & a_{m2}^{(k)} & \cdots & a_{mn}^{(k)} \end{pmatrix},$$

如果当 $k \rightarrow \infty$ 时， $m \times n$ 个序列 $\{a_{ij}^{(k)}\}$ 都收敛，且分别收敛于 a_{ij} ，则称矩阵序列 $\{A_k\}$ 收敛于矩阵

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

并称 A 是序列 $\{A_k\}$ 在 $k \rightarrow \infty$ 时的极限，记作

$$\lim_{k \rightarrow \infty} A_k = A.$$

例 4.1 已知

$$A_k = \begin{pmatrix} \frac{1}{k} & \frac{2k^2 - 1}{3k^2 + 4} \\ \left(1 + \frac{1}{k}\right)^k & \cos \frac{1}{k^3} \end{pmatrix},$$

求 $\{A_k\}$ 在 $k \rightarrow \infty$ 时的极限.

解 令

$$A = \lim_{k \rightarrow \infty} A_k,$$

则

$$A = \begin{pmatrix} \lim_{k \rightarrow \infty} \frac{1}{k} & \lim_{k \rightarrow \infty} \frac{2k^2 - 1}{3k^2 + 4} \\ \lim_{k \rightarrow \infty} \left(1 + \frac{1}{k}\right)^k & \lim_{k \rightarrow \infty} \cos \frac{1}{k^3} \end{pmatrix} = \begin{pmatrix} 0 & \frac{2}{3} \\ e & 1 \end{pmatrix}.$$

定理 4.1 已知 A_k, B_k, A, B 为矩阵, a_k, b_k, a, b 为数, $\lim_{k \rightarrow \infty} A_k = A, \lim_{k \rightarrow \infty} B_k = B$,

$\lim_{k \rightarrow \infty} a_k = a, \lim_{k \rightarrow \infty} b_k = b$, 则

$$(1) \lim_{k \rightarrow \infty} (a_k A_k + b_k B_k) = aA + bB;$$

$$(2) \lim_{k \rightarrow \infty} (A_k B_k) = AB.$$

证明 (1) 令矩阵 $C_k = a_k A_k + b_k B_k$, 则 C_k 的第 i 行第 j 列元素为

$$c_{ij}^{(k)} = a_k a_{ij}^{(k)} + b_k b_{ij}^{(k)},$$

于是

$$\lim_{k \rightarrow \infty} c_{ij}^{(k)} = \lim_{k \rightarrow \infty} a_k \lim_{k \rightarrow \infty} a_{ij}^{(k)} + \lim_{k \rightarrow \infty} b_k \lim_{k \rightarrow \infty} b_{ij}^{(k)} = aa_{ij} + bb_{ij},$$

因此

$$\lim_{k \rightarrow \infty} C_k = aA + bB.$$

(2) 令矩阵 $C_k = A_k B_k$, 则 C_k 的第 i 行第 j 列元素为

$$c_{ij}^{(k)} = \sum_{m=1}^n a_{im}^{(k)} b_{mj}^{(k)},$$

于是

$$\lim_{k \rightarrow \infty} c_{ij}^{(k)} = \sum_{m=1}^n \lim_{k \rightarrow \infty} a_{im}^{(k)} \lim_{k \rightarrow \infty} b_{mj}^{(k)},$$

从而

$$\lim_{k \rightarrow \infty} C_k = AB.$$

证毕.

推论 4.2 已知 $P, Q, A_k, A \in \mathbb{C}^{n \times n}$, $\lim_{k \rightarrow \infty} A_k = A$, 则

$$\lim_{k \rightarrow \infty} P A_k Q = P A Q.$$

现在考虑其矩阵元素是实变量 t 的函数的 $m \times n$ 阶矩阵

$$\mathbf{A}(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & & \vdots \\ a_{m1}(t) & a_{m2}(t) & \cdots & a_{mn}(t) \end{pmatrix},$$

其所有元素 $a_{ij}(t)$ 定义在区间 $[a, b]$ 上, 函数矩阵 $\mathbf{A}(t)$ 在 $[a, b]$ 上有界、有极限、连续、可微、可积等概念, 可用其 $m \times n$ 个元素 $a_{ij}(t)$ 同时在 $[a, b]$ 上有界、有极限、连续、可微、可积来定义. 例如

$$\begin{aligned} \frac{d}{dt} \mathbf{A}(t) &= \left[\frac{d}{dt} a_{ij}(t) \right]_{m \times n}; \\ \int \mathbf{A}(t) dt &= \left[\int a_{ij}(t) dt \right]_{m \times n}; \\ \int_a^b \mathbf{A}(t) dt &= \left[\int_a^b a_{ij}(t) dt \right]_{m \times n}. \end{aligned}$$

例 4.2 已知 $\mathbf{A}(t) = \begin{pmatrix} \sin t & 2t^3 \\ 2\sqrt{t} & e^{2t} \end{pmatrix}$, 求 $\frac{d}{dt} \mathbf{A}(t)$.

解 直接计算, 得

$$\frac{d}{dt} \mathbf{A}(t) = \begin{pmatrix} \frac{d}{dt}(\sin t) & \frac{d}{dt}(2t^3) \\ \frac{d}{dt}(2\sqrt{t}) & \frac{d}{dt}(e^{2t}) \end{pmatrix} = \begin{pmatrix} \cos t & 6t^2 \\ \frac{1}{\sqrt{t}} & 2e^{2t} \end{pmatrix}.$$

定理 4.3 (1) 若 $\mathbf{A}(t)$, $\mathbf{B}(t)$ 为同阶可微矩阵, 则

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

(2) 若 $\mathbf{A}(t)$, $\mathbf{B}(t)$ 分别为 $m \times n$, $n \times l$ 阶可微矩阵,

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

(3) 若 $\mathbf{A}(t)$ 与 $\mathbf{A}^{-1}(t)$ 皆可微, 则

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

证明 (1)(2) 易证, 以下证明 (3). 对 $\mathbf{A}(t)\mathbf{A}^{-1}(t) = \mathbf{E}$ 两边求导, 得

$$\left(\frac{d}{dt} \mathbf{A}(t) \right) \mathbf{A}^{-1}(t) + \mathbf{A}(t) \frac{d}{dt} \mathbf{A}^{-1}(t) = \mathbf{0},$$

移项并两边左乘 $\mathbf{A}^{-1}(t)$, 得

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

证毕.

4.2 矩阵幂级数

类似数列的级数，矩阵序列也可以同样的方式定义级数。

定义 4.2 设 $A_k \in \mathbf{C}^{m \times n}$ ，令 $S_N = \sum_{k=0}^N A_k$ ，若矩阵序列 $\{S_N\}$ 收敛到 S ，则称矩阵级数 $\sum_{k=0}^{\infty} A_k$ 收敛，并记 $S = \sum_{k=0}^{\infty} A_k$ ，否则称矩阵级数 $\sum_{k=0}^{\infty} A_k$ 发散。

易见， $\sum_{k=0}^{\infty} A_k$ 收敛的充分必要条件是所有 $m \times n$ 个数值级数 $\sum_{k=0}^{\infty} a_{ij}^k$ 都收敛。

矩阵级数中，矩阵幂级数是最简单且有重要应用的一种，其本质是矩阵多项式的极限。

定义 4.3 设 A 为 n 阶方阵，则称 $\sum_{k=0}^{\infty} a_k A^k$ （约定 $A^0 = E$ ）为关于矩阵 A 的幂级数。若记 $f(\lambda) = \sum_{k=0}^{\infty} a_k \lambda^k$ ，则矩阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 也记为 $f(A)$ 。

一个自然的问题是，矩阵幂级数何时收敛。为此，先不加证明地引用一个定理（参见文献 [5]）。

定理 4.4 对幂级数 $f(\lambda) = \sum_{k=0}^{\infty} a_k \lambda^k$ ，存在收敛半径 $R \in [0, +\infty]$ ，当 $|\lambda| < R$ 时，级数收敛；当 $|\lambda| > R$ 时，级数发散。进一步，幂级数 $f(\lambda)$ 的导数幂级数 $\sum_{k=0}^{\infty} k a_k \lambda^{k-1}$ 的收敛半径，与 $f(\lambda)$ 的收敛半径相同，且在收敛区域内满足 $f'(\lambda) = \sum_{k=0}^{\infty} k a_k \lambda^{k-1}$ 。

幂级数的收敛半径，亦决定了相应矩阵幂级数的敛散性。

引理 4.5 若 $f(z) = \sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 R ，则在收敛区域内 $\sum_{k=s}^{\infty} C_k^s a_k z^{k-s} = \frac{1}{s!} f^{(s)}(z)$ 。

证明 因为 $C_k^s a_k z^{k-s} = \frac{k!}{s!(k-s)!} a_k z^{k-s} = \frac{1}{s!} a_k \frac{d^s}{dz^s} z^k$ ，所以

$$\sum_{k=s}^{\infty} C_k^s a_k z^{k-s} = \frac{1}{s!} \sum_{k=s}^{\infty} a_k \frac{d^s}{dz^s} z^k = \frac{1}{s!} f^{(s)}(z)。$$

证毕。

定理 4.6 设 $f(z) = \sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 R , r 阶 Jordan 块 $J = \lambda E + H$, 其中

$$H = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}.$$

则当 $|\lambda| < R$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} a_k J^k$ 收敛, 且

$$f(J) = \sum_{k=0}^{r-1} \frac{1}{k!} f^{(k)}(\lambda) H^k = \begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!} f''(\lambda) & \cdots & \frac{1}{(r-1)!} f^{(r-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(r-2)!} f^{(r-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{pmatrix}. \quad (4.1)$$

证明 因为单位矩阵 E 与 H 乘法可交换, 于是

$$\begin{aligned} J^k &= (\lambda E + H)^k = \lambda^k E + C_k^1 \lambda^{k-1} H + \cdots + C_k^{k-1} \lambda H^{k-1} + H^k \\ &= \begin{pmatrix} \lambda^k & C_k^1 \lambda^{k-1} & \cdots & \cdots & C_k^{r-1} \lambda^{k-r+1} \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & C_k^1 \lambda^{k-1} \\ & & & & \lambda^k \end{pmatrix}, \end{aligned}$$

这里当 $j > k$ 时, 约定 $C_k^j = 0$, 进而

$$\sum_{k=0}^{\infty} a_k J^k = \begin{pmatrix} \sum_{k=0}^{\infty} a_k \lambda^k & \sum_{k=0}^{\infty} a_k C_k^1 \lambda^{k-1} & \cdots & \cdots & \sum_{k=0}^{\infty} a_k C_k^{r-1} \lambda^{k-r+1} \\ & \sum_{k=0}^{\infty} a_k \lambda^k & \cdots & \cdots & \sum_{k=0}^{\infty} a_k C_k^{r-2} \lambda^{k-r+2} \\ & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ & & & & \sum_{k=0}^{\infty} a_k \lambda^k \end{pmatrix}.$$

从而根据引理 4.5 知, 当 $|\lambda| < R$ 时, 有

$$f(J) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(r-1)!}f^{(r-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(r-2)!}f^{(r-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{pmatrix}.$$

证毕.

定理 4.7 设幂级数 $f(z) = \sum_{k=0}^{\infty} a_k z^k$ 的收敛半径为 R , 方阵 A 的一切特征值的模均小于 R (称方阵 A 的一切特征值的模的最大值为 A 的谱半径, 记为 $\rho(A)$), 则矩阵幂级数 $\sum_{k=0}^{\infty} a_k A^k$ 收敛, 且 $f(A) = P f(J) P^{-1}$, 这里 $A = P J P^{-1}$, J 为 A 的 Jordan 标准形, P 为过渡矩阵.

证明 设 A 的 Jordan 标准形

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_s \end{pmatrix},$$

因为对一切正整数 n ,

$$\sum_{k=0}^n a_k J^n = \begin{pmatrix} \sum_{k=0}^n a_k J_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^n a_k J_s^k \end{pmatrix},$$

故当 A 的谱半径小于 R 时,

$$f(J) = \begin{pmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_s) \end{pmatrix},$$

于是

$$f(A) = \lim_{n \rightarrow \infty} \sum_{k=0}^n a_k A^k = \lim_{n \rightarrow \infty} P \left(\sum_{k=0}^n a_k J^k \right) P^{-1} = P \left(\lim_{n \rightarrow \infty} \sum_{k=0}^n a_k J^k \right) P^{-1} = P f(J) P^{-1}.$$

证毕.

注 4.1 当矩阵谱半径大于幂级数的收敛半径时, $f(\lambda)$ 在至少一个 Jordan 块上的取值是没有定义的, 从而“矩阵谱半径小于幂级数的收敛半径”这一条件基本不能再改进了.

推论 4.8 若幂级数 $f(\lambda)$ 在矩阵 A 上的取值有定义, 矩阵 B 与 A 相似, 则 $f(B)$ 有定义

且与 $f(\mathbf{A})$ 相似.

证明 由定理 4.7 知 $f(\mathbf{A})$ 有定义且与 $f(\mathbf{J})$ 相似, 其中 \mathbf{J} 为 \mathbf{A} 的 Jordan 标准形. 而 \mathbf{B} 与 \mathbf{A} 相似, 从而具有相同的 Jordan 标准形, 得证.

定理 4.7 给出了矩阵幂级数的收敛条件和计算方法. 在诸多幂级数中,

$$\mathrm{e}^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!} + \cdots,$$

$$\sin z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots,$$

$$\cos z = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} z^{2k} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \cdots,$$

为三种最为常见的初等函数. 众所周知, 上述三个幂级数在整个复平面上收敛, 于是对一切方阵 $\mathbf{A} \in \mathbf{C}^{n \times n}$, $\mathrm{e}^{\mathbf{A}}$, $\sin \mathbf{A}$ 和 $\cos \mathbf{A}$ 有定义, 特别

$$\mathrm{e}^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} = \mathbf{E} + \mathbf{A} + \frac{1}{2!} \mathbf{A}^2 + \cdots + \frac{1}{n!} \mathbf{A}^n + \cdots. \quad (4.2)$$

由于三角函数可根据 Euler¹ 公式

$$\cos z = \frac{\mathrm{e}^{\mathrm{i}z} + \mathrm{e}^{-\mathrm{i}z}}{2}, \quad \sin z = \frac{\mathrm{e}^{\mathrm{i}z} - \mathrm{e}^{-\mathrm{i}z}}{2\mathrm{i}},$$

表示为指数函数 (参见参考文献 [5]), 故以下只介绍矩阵指数函数的性质和计算.

例 4.3 计算 $\mathrm{e}^{\mathbf{0}}$.

解 $\mathbf{0}$ 可看作特征值全为零的对角矩阵, 故

$$\mathrm{e}^{\mathbf{0}} = \begin{pmatrix} \mathrm{e}^0 & & & \\ & \mathrm{e}^0 & & \\ & & \ddots & \\ & & & \mathrm{e}^0 \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} = \mathbf{E}.$$

证毕.

定理 4.9 若 n 阶方阵 \mathbf{A} 与 \mathbf{B} 满足 $\mathbf{AB} = \mathbf{BA}$, 则 $\mathrm{e}^{\mathbf{A}}\mathrm{e}^{\mathbf{B}} = \mathrm{e}^{\mathbf{B}}\mathrm{e}^{\mathbf{A}} = \mathrm{e}^{\mathbf{A}+\mathbf{B}}$.

证明 对一切正整数 k , 因为 $\mathbf{AB} = \mathbf{BA}$, 所以

$$\frac{1}{k!}(\mathbf{A} + \mathbf{B})^k = \frac{1}{k!} \sum_{m=0}^k \frac{k!}{m!(k-m)!} \mathbf{A}^m \mathbf{B}^{k-m} = \sum_{m=0}^k \frac{\mathbf{A}^m}{m!} \frac{\mathbf{B}^{k-m}}{(k-m)!},$$

从而

$$\mathrm{e}^{\mathbf{A}+\mathbf{B}} = \sum_{k=0}^{\infty} \frac{(\mathbf{A} + \mathbf{B})^k}{k!} = \sum_{k=0}^{\infty} \sum_{m=0}^k \frac{\mathbf{A}^m}{m!} \frac{\mathbf{B}^{k-m}}{(k-m)!}. \quad (4.3)$$

¹Euler (1707–1783), 欧拉, 瑞士数学家.

令 $n = k - m$, n 将取遍所有非负整数. 对每一个固定的 n , 式 (4.3) 右端级数中包含

$$\sum_{m=0}^{\infty} \frac{A^m}{m!} B^n = e^A B^n, \quad (4.4)$$

于是式 (4.3) 右端可进一步写为

$$\sum_{k=0}^{\infty} \sum_{m=0}^k \frac{A^m}{m!} \frac{B^{k-m}}{(k-m)!} = \sum_{n=0}^{\infty} e^A \frac{B^n}{n!} = e^A \sum_{n=0}^{\infty} \frac{B^n}{n!} = e^A e^B, \quad (4.5)$$

即 $e^{A+B} = e^A e^B$. 交换 A 和 B 的位置, 得到 $e^{B+A} = e^B e^A$, 而 $e^{A+B} = e^{B+A}$ 故 $e^{A+B} = e^B e^A$. 证毕.

注 4.2 上面的证明中, 式 (4.4) 与 (4.5) 用到了交换级数项的求和次序, 这对幂级数在收敛区域内是可行的, 但在幂级数的收敛区域外, 或对其他的级数, 并不都能随意交换求和次序.

定理 4.10 对一切 n 阶方阵 A , e^A 可逆, 且 $(e^A)^{-1} = e^{-A}$.

证明 令 $B = -A$, 利用定理 4.9 并结合例 4.3 得

$$e^{-A} e^A = e^{-A+A} = e^{-0} = E.$$

即 $e^{-A} = (e^A)^{-1}$. 证毕.

定理 4.11 $\det e^A = e^{\operatorname{tr} A}$.

证明 设 $A = PAP^{-1}$, 其中 $J = \operatorname{diag}(J_1, \dots, J_s)$ 为 A 的 Jordan 标准形, 其中每个 Jordan 块 J_s 的特征值为 λ_s , 阶数为 n_s ($1 \leq s \leq r$). 由定理 4.7、推论 4.8 和定理 4.6 得,

$$\det e^A = \det e^J = \det e^{J_1} \cdots \det e^{J_r} = e^{n_1 \lambda_1 + \cdots + n_r \lambda_r}.$$

注意到 $n_1 \lambda_1 + \cdots + n_r \lambda_r$ 为 A 的所有特征值的和, 故 $n_1 \lambda_1 + \cdots + n_r \lambda_r = \operatorname{tr} A$. 证毕.

定理 4.12 矩阵 A 与 e^{At} 乘法可交换, 且 $\frac{d}{dt} e^{At} = e^{At} A = A e^{At}$.

证明 根据矩阵指数函数的定义式 (4.2),

$$e^{At} = \sum_{k=0}^{\infty} \frac{1}{k!} (At)^k = \sum_{k=0}^{\infty} \frac{1}{k!} A^k t^k.$$

故矩阵 e^{At} 中的每个元素都是关于 t 的收敛幂级数, 从而可以关于 t 逐项求导, 即

$$\frac{d}{dt} e^{At} = \sum_{k=0}^{\infty} \frac{1}{k!} A^k \frac{d}{dt} t^k = \sum_{k=0}^{\infty} \frac{1}{k!} A^k k t^{k-1} = \sum_{k=0}^{\infty} \frac{1}{k!} A^{k+1} t^k = e^{At} A = A e^{At}.$$

证毕.

上述定理与数值指数函数的性质十分类似, 由此说明将式 (4.2) 作为矩阵指数函数的定义式, 是合理的.

4.3 矩阵函数的计算方法

本节介绍计算矩阵函数的计算方法.

直接根据定理 4.7, 可以得到矩阵函数的第 1 种计算方法, 共分 3 个步骤:

第 1 步, 先计算 A 的 Jordan 标准形 $J = \text{diag}(J_1, J_2, \dots, J_s)$, 过渡矩阵 P 及其逆矩阵 P^{-1} ;

第 2 步, 根据定理 4.6, 对每个 Jordan 块计算 $f(J_k)$;

第 3 步, 根据下式得到

$$f(A) = Pf(J)P^{-1} = P\text{diag}(f(J_1), f(J_2), \dots, f(J_s))P^{-1}.$$

例 4.4 设 $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{pmatrix}$, 求矩阵函数 A^{20} , e^A , $\sin A$, e^{At} .

解 由例 3.9 知 A 的 Jordan 标准形、过渡矩阵为

$$J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

进而求得过渡矩阵的逆矩阵

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

记 $f_1(A) = A^{20}$, $f_2(A) = e^{At}$. 由于 J 的 Jordan 块的最高阶数为 2, 于是先求出

$$f_1(x) = x^{20}, f_1'(x) = 20x^{19}; f_2(x) = e^{xt}, f_2'(x) = te^{xt},$$

代入式 (4.1) 得到

$$\begin{aligned} f_1(J) &= J^{20} = \begin{pmatrix} 2^{20} & 0 & 0 \\ 0 & 2^{20} & 20 \times 2^{19} \\ 0 & 0 & 2^{20} \end{pmatrix} = 2^{20} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{pmatrix}, \\ f_2(J) &= e^{Jt} = \begin{pmatrix} e^{2t} & 0 & 0 \\ 0 & e^{2t} & te^{2t} \\ 0 & 0 & e^{2t} \end{pmatrix} = e^{2t} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

再由 $f(A) = Pf(J)P^{-1}$, 得

$$\begin{aligned} A^{20} &= Pf_1(J)P^{-1} = 2^{20} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix} = 2^{20} \begin{pmatrix} 1 & 0 & 0 \\ 10 & -9 & 10 \\ 10 & -10 & 11 \end{pmatrix}, \\ e^{At} &= Pf_2(J)P^{-1} = e^{2t} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix} = e^{2t} \begin{pmatrix} 1 & 0 & 0 \\ t & 1-t & t \\ t & -t & 1+t \end{pmatrix}. \end{aligned}$$

在 e^{At} 中令 $t = 1$, 得到

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

利用 Euler 公式, 并在 e^{At} 中令 $t = \pm i$, 得到

$$\sin A = \frac{e^{iA} - e^{-iA}}{2i} = \begin{pmatrix} \sin 2 & 0 & 0 \\ \cos 2 & \sin 2 - \cos 2 & \cos 2 \\ \cos 2 & -\cos 2 & \sin 2 + \cos 2 \end{pmatrix}.$$

上述根据定理 4.7 来计算 $f(A)$ 的方法, 需要计算过渡矩阵 P 及其逆 P^{-1} , 当 Jordan 块结构比较复杂时 (例如本例中不同的 Jordan 块有相同的特征值), 过渡矩阵计算过程较繁. 以下介绍的第二种算法, 有效地克服了这一困难.

第二种计算方法基于下述定理.

定理 4.13 设 $f(\lambda)$ 和 $g(\lambda)$ 为幂级数, A 为 n 阶方阵, $\lambda_1, \lambda_2, \dots, \lambda_t$ 是矩阵 A 的全体互不相同的特征值, A 的最小多项式为 $m_A(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_t)^{m_t}$, 记 $m = \deg m_A(\lambda)$. 若 $f(A)$ 和 $g(A)$ 均有定义, 且

$$\begin{aligned} f(\lambda_1) &= g(\lambda_1), f'(\lambda_1) = g'(\lambda_1), \dots, f^{(m_1-1)}(\lambda_1) = g^{(m_1-1)}(\lambda_1), \\ f(\lambda_2) &= g(\lambda_2), f'(\lambda_2) = g'(\lambda_2), \dots, f^{(m_2-1)}(\lambda_2) = g^{(m_2-1)}(\lambda_2), \\ &\dots \dots \dots \\ f(\lambda_t) &= g(\lambda_t), f'(\lambda_t) = g'(\lambda_t), \dots, f^{(m_t-1)}(\lambda_t) = g^{(m_t-1)}(\lambda_t), \end{aligned} \quad (4.6)$$

(上述 m 个方程左右两端的值分别称为 $f(\lambda)$ 与 $g(\lambda)$ 关于矩阵 A 的谱上的值), 则 $f(A) = g(A)$.

证明 由定理 4.7 立即得到 $f(J) = g(J)$, 其中 J 是 A 的 Jordan 标准形, 进而 $f(A) = Pf(J)P^{-1} = Pg(J)P^{-1} = g(A)$. 证毕.

对给定的矩阵 A 和幂级数 $f(\lambda)$, 当 $g(\lambda)$ 可取为某个多项式 $p(\lambda)$ 时, 矩阵多项式 $p(A)$ 可直接计算, 从而计算一般矩阵幂级数 $f(A)$, 可转化为寻求满足条件 (4.6) 的多项式 $p(\lambda)$ 的问题. 以下定理指出, 这样的多项式 $p(\lambda)$ 必定存在.

定理 4.14 满足条件 (4.6), 且次数低于 m 的多项式存在且唯一.

证明 考虑同余方程组

$$p(\lambda) \equiv f(\lambda_k) + f'(\lambda_k)(\lambda - \lambda_k) + \cdots + \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!}(\lambda - \lambda_k)^{m_k-1} \pmod{(\lambda - \lambda_k)^{m_k}} \quad (1 \leq k \leq t),$$

由中国剩余定理 (定理 2.17) 知, 该同余方程的解, 关于模 $m_A(\lambda)$ 存在唯一. 以下说明条件 (4.6) 与上述同余方程等价. 事实上, 考虑 $p(\lambda)$ 在 λ_k 点的 Taylor¹ 展开式, 两者的等价性立得. 证毕.

注 4.3 当 $f(\lambda)$ 为多项式时, $p(\lambda)$ 就是 $f(\lambda)$ 关于 $m_A(\lambda)$ 的余式, 上述定理本质上说明,

¹Taylor (1685–1731), 泰勒, 英国数学家.

对一般的幂级数 $f(\lambda)$ ，也可以定义 $f(\lambda)$ 关于一个多项式的余式. 将求 $f(A)$ 转化为求 $p(A)$ 的做法，本质上与多项式、矩阵的 Jordan 标准形两章中，利用零化多项式简化计算多项式的值的思想，完全相同.

例 4.5 设矩阵 $A = \begin{pmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{pmatrix}$ ，求 e^{At} .

解 先求 A 的最小多项式 $m_A(\lambda)$. 由

$$\begin{aligned} \lambda E - A &= \begin{pmatrix} \lambda-2 & -1 & -4 \\ 0 & \lambda-2 & 0 \\ 0 & -3 & \lambda-1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & (\lambda-2) & -4 \\ -(\lambda-2) & 0 & 0 \\ 3 & 0 & \lambda-1 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\lambda-2)^2 & -4(\lambda-2) \\ 0 & -3(\lambda-2) & \lambda+11 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-2)^2(\lambda-1) \end{pmatrix}, \end{aligned}$$

得 $m_A(\lambda) = (\lambda-1)(\lambda-2)^2$ ，于是满足条件 (4.6) 的多项式 $p(\lambda)$ 至多为 2 次. 此时 $f(\lambda) = e^{At}$ ，故

$$f(1) = e^t, f(2) = e^{2t}, f'(2) = te^{2t},$$

从而可设

$$p(\lambda) = e^{2t} + te^{2t}(\lambda-2) + a_2(\lambda-2)^2,$$

根据余式定理 (定理 2.6)，有 $p(1) = f(1)$ ，解得 $a_2 = e^t - e^{2t} + te^{2t}$ ，于是

$$p(\lambda) = (4e^t - 3e^{2t} + 2te^{2t}) + (-4e^t + 4e^{2t} - 3te^{2t})\lambda + (e^t - e^{2t} + te^{2t})\lambda^2.$$

从而

$$\begin{aligned} e^{At} &= p(A) = (4e^t - 3e^{2t} + 2te^{2t})E + (-4e^t + 4e^{2t} - 3te^{2t})A + (e^t - e^{2t} + te^{2t})A^2 \\ &= \begin{pmatrix} e^{2t} & 12e^t - 12e^{2t} + 13te^{2t} & -4e^t + 4e^{2t} \\ 0 & e^{2t} & 0 \\ 0 & -3e^t + 3e^{2t} & e^t \end{pmatrix}. \end{aligned}$$

通过例 4.5，矩阵函数 $f(A)$ 的第二种计算方法可总结为三个步骤：

第 1 步，求出矩阵 A 的最小多项式 $m_A(\lambda)$ 的所有根以及根的重数；

第 2 步，求出满足条件 (4.6) 的多项式 $p(\lambda)$ ；

第 3 步，计算 $p(A)$ ，从而得到 $f(A)$.

第二种方法不需要计算过渡矩阵 P ，尤其当不同的 Jordan 块有相同的特征值时，最小多项式的次数将低于矩阵 A 的阶数，从而减小了求 $p(\lambda)$ 的计算量. 而且，具有相同特征值的 Jordan 块越多，利用第一种方法，计算特征向量的难度越大，而利用第二种方法， $p(\lambda)$ 次数反而越低，优越性就愈加明显.

无论哪一种方法，都需要计算矩阵的特征值，因此矩阵特征值的计算，仍然是无法避开

的问题. 相关内容将在特征值的估计一章中介绍.

4.4 利用矩阵函数求解线性微分方程组

本节讨论如何用矩阵函数表示一阶线性常系数微分方程组的解.

考虑如下的一阶线性常系数常微分方程组

[illegible]

若记

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}, \quad F(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{pmatrix}.$$

则方程组 (4.7) 可写成如下矩阵形式

$$X'(t) = AX(t) + F(t), \quad (4.8)$$

利用矩阵函数的知识, 并参考常微分方程的求解技巧, 可得求解方程组 (4.8) 的如下定理.

定理 4.15 一阶线性常系数常微分方程组 (4.8) 的解为

$$X(t) = e^{At}X(0) + \int_0^t e^{A(t-\tau)}F(\tau) \, d\tau \quad (4.9)$$

证明 将方程组 (4.8) 的两边左乘 e^{-At} , 得

$$e^{-At}X'(t) = e^{-At}AX(t) + e^{-At}F(t).$$

根据定理 4.12, 上式可化为

$$\frac{d}{dt}(e^{-At}X(t)) = e^{-At}F(t) ,$$

两边对 t 在 $[0, t]$ 上积分, 得

$$e^{-At}X(t) - X(0) = \int_0^t e^{-A\tau}F(\tau) \, d\tau \, ,$$

上式两边左乘 e^{At} 即得结论. 证毕.

注 4.4 当 $F(t) = 0$ 时, 此时的方程组称为齐次的, 否则称为非齐次的. 若记初值 $X(0) = C$, 则 $X(t) = e^{At}C$ 称为齐次线性方程组 $X'(t) = AX(t)$ 的通解. 这一通解表达式与 $X(t)$ 为数值函数时, 相应的常微分方程的通解表达式完全相同, 由此再次看出矩阵指数函数定义的合理性.

例 4.6 求方程组

$$\begin{cases} x_1'(t) = 3x_1(t) + 8x_3(t), \\ x_2'(t) = 3x_1(t) - x_2(t) + 6x_3(t), \\ x_3'(t) = -2x_1(t) - 5x_3(t), \end{cases} \quad (4.10)$$

满足初值条件 $x_1(0) = 1$, $x_2(0) = 2$, $x_3(0) = -3$ 的解.

解 令 $X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}$, $A = \begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix}$, $X(0) = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$, 则由定理 4.15 知, $X(t) =$

$e^{At}X(0)$. 为此, 计算 e^{At} .

$$\begin{aligned} \lambda E - A &= \begin{pmatrix} \lambda - 3 & 0 & -8 \\ -3 & \lambda + 1 & -6 \\ 2 & 0 & \lambda + 5 \end{pmatrix} \rightarrow \begin{pmatrix} \lambda - 3 & 0 & -8 \\ 1 & -(\lambda + 1) & -(\lambda - 1) \\ 2 & 0 & \lambda + 5 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & -(\lambda + 1) & -(\lambda - 1) \\ \lambda - 3 & 0 & -8 \\ 2 & 0 & \lambda + 5 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\lambda - 3)(\lambda + 1) & (\lambda - 5)(\lambda + 1) \\ 0 & 2(\lambda + 1) & 3(\lambda + 1) \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda + 1 & 0 \\ 0 & 0 & (\lambda + 1)^2 \end{pmatrix}, \end{aligned}$$

故 A 的最小多项式为 $m_A(\lambda) = (\lambda + 1)^2$. 令

$$p(\lambda) = e^{-t} + te^{-t}(\lambda + 1) = te^{-t}\lambda + (e^{-t} + te^{-t}),$$

则 $p(-1) = e^{-t}|_{t=-1}$, $p'(-1) = \frac{d}{dt}e^{-t}|_{t=-1}$, 故

$$\begin{aligned} X(t) &= e^{At}X(0) = p(A)X(0) = (te^{-t}A + (e^{-t} + te^{-t})E)X(0) \\ &= te^{-t} \begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} + (e^{-t} + te^{-t}) \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} = \begin{pmatrix} e^{-t} - 20te^{-t} \\ 2e^{-t} - 15te^{-t} \\ -3e^{-t} + 10te^{-t} \end{pmatrix}. \end{aligned}$$

例 4.7 求微分方程组 $X'(t) = AX(t) + F(t)$, 满足初始条件 $X(0) = (1 \ 0 \ 0)^T$ 的解, 其中

$$A = \begin{pmatrix} -6 & 1 & 0 \\ -11 & 0 & 1 \\ -6 & 0 & 0 \end{pmatrix}, F(t) = \begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix}.$$

解 因为 $|\lambda E - A| = (\lambda + 1)(\lambda + 2)(\lambda + 3)$, 故可求得

$$J = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix}, P = \begin{pmatrix} 1 & 1 & 1 \\ 5 & 4 & 3 \\ 6 & 3 & 2 \end{pmatrix}, P^{-1} = \frac{1}{2} \begin{pmatrix} 1 & -1 & 1 \\ -8 & 4 & -2 \\ 9 & -3 & 1 \end{pmatrix},$$

其中 J 为 A 的 Jordan 标准形, P 为过渡矩阵. 从而

$$\begin{aligned} X(t) &= e^{At} X(0) + \int_0^t e^{A(t-\tau)} F(\tau) d\tau \\ &= P \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-2t} & 0 \\ 0 & 0 & e^{-3t} \end{pmatrix} P^{-1} X(0) + P \int_0^t \begin{pmatrix} e^{-(t-\tau)} & 0 & 0 \\ 0 & e^{-2(t-\tau)} & 0 \\ 0 & 0 & e^{-3(t-\tau)} \end{pmatrix} P^{-1} F(\tau) d\tau \\ &= \begin{pmatrix} \frac{3}{2}e^{-t} - 5e^{-2t} + \frac{25}{6}e^{-3t} + \frac{1}{3} \\ \frac{15}{2}e^{-t} - 20e^{-2t} + \frac{25}{2}e^{-3t} \\ 9e^{-t} - 15e^{-2t} + \frac{25}{3}e^{-3t} - \frac{7}{3} \end{pmatrix}. \end{aligned}$$

习题 4

1. 设函数矩阵 $A(t) = \begin{pmatrix} \sin t & -e^t & t \\ \cos t & e^t & t^2 \\ 1 & 0 & 0 \end{pmatrix}$, 试求 $\frac{d}{dt}A(t)$, $\left| \frac{d}{dt}A(t) \right|$, $\lim_{t \rightarrow 0} A(t)$.

2. 设函数矩阵 $A(t) = \begin{pmatrix} e^{2t} & te^t & 1 \\ e^{-t} & 2e^{2t} & 0 \\ 3t & 0 & 0 \end{pmatrix}$, 试求 $\int A(t) dt$, $\int_0^t A(t) dt$.

3. 判断级数 $\sum_{n=0}^{\infty} \frac{1}{10^n} \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix}^n$ 是否收敛. 如果收敛, 计算出结果.

4. 已知矩阵 $A = \begin{pmatrix} 0 & 1 \\ -2 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 0 & -1 \\ 4 & 4 \end{pmatrix}$, 试求 e^A , e^B .

5. 已知矩阵 $A = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}$, 试证 $e^{At} = \begin{pmatrix} \cos \theta t & -\sin \theta t \\ \sin \theta t & \cos \theta t \end{pmatrix}$.

6. 设矩阵 $A = \begin{pmatrix} \sigma & -\theta \\ \theta & \sigma \end{pmatrix}$, 利用上题结果求 e^A .

7. 设 $A = \begin{pmatrix} 9 & -6 & -7 \\ -1 & -1 & 1 \\ 10 & -6 & -8 \end{pmatrix}$, 求 e^{2At} .

8. 设 $A = \begin{pmatrix} 3 & 1 & -3 \\ -7 & -2 & 9 \\ -2 & -1 & 4 \end{pmatrix}$, 求 e^{At} .

9. 已知 $A = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & -2 \end{pmatrix}$, 试求 $\cos A$, $\sin B$, e^{Bt} .

10. 已知 $A = \begin{pmatrix} 0 & 1 \\ 0 & -2 \end{pmatrix}$, $B = \begin{pmatrix} -2 & 1 & 1 \\ 0 & 2 & 0 \\ -4 & 1 & 3 \end{pmatrix}$, 试求 e^{At} , e^{Bt} , $\sin Bt$.

11. 求常系数线性齐次微分方程组

$$\begin{cases} x_1'(t) = -7x_1 - 7x_2 + 5x_3, \\ x_2'(t) = -8x_1 - 8x_2 - 5x_3, \\ x_3'(t) = -5x_2, \end{cases}$$

满足初始条件 $x_1(0) = 3$, $x_2(0) = -2$, $x_3(0) = 1$ 的解.

12. 求常系数线性微分方程组

$$\begin{cases} x_1'(t) = x_1(t) - x_2(t), \\ x_2'(t) = 4x_1(t) - 3x_2(t) + 1, \end{cases}$$

满足初始条件 $x_1(0) = 1$, $x_2(0) = 2$ 的解.

13. 求微分方程组 $X'(t) = AX(t)$ 的通解, 其中 $A = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & -1 & 1 \end{pmatrix}$.

14. 求微分方程组 $X'(t) = AX(t) + F(t)$ 的通解, 其中 $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$, $F(t) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

15. 设 A 和 B 为方阵, 令

$$M(s, t) = \frac{d}{dt} \frac{d}{ds} e^{sA} e^{tB} e^{-sA} e^{-tB}.$$

(1) 证明 $M(0, 0) = AB - BA$.

(2) 若 $AB - BA = 0$, 证明 $M(s, t) = 0$, 并由此证明定理 4.9.

16. 设 A 为方阵, $B(t) = e^{At}$.

(1) 若 $\operatorname{tr} A = 0$, 证明对一切 $t \in \mathbf{R}$, $\det B(t) = 1$.

(2) 若 $\bar{A} + A^T = 0$, 证明对一切 $t \in \mathbf{R}$, $\bar{B}(t)B(t)^T = E$.

第 5 章 线性空间与线性变换

包括数、向量、矩阵、多项式、函数等在内的很多数学对象，加法和数乘都是定义在这些数学对象上的最基本的两种运算（把它们统称为线性运算），且定义在这些数学对象上的很多高级运算，都保持线性运算，即与线性运算可以交换次序。本章将在一个统一的框架内研究这些具有共同特点的数学对象。需要指出的是，将具备一定共性的不同数学对象放在一个框架下研究其共同的性质，正是代数学的重要思想和特点。

5.1 线性空间

在定义矩阵的加法与数乘运算时，发现同一类型的矩阵全体关于这两种运算满足以下 8 条性质：

- (1) 加法交换律： $A + B = B + A$ ；
- (2) 加法结合律： $(A + B) + C = A + (B + C)$ ；
- (3) 对任何矩阵 A ，有 $0 + A = A$ ，这里 0 为零矩阵；
- (4) 对任何矩阵 A ， $A + (-A) = 0$ ；
- (5) $1A = A$ ；
- (6) 对任何数 k, l ，任何矩阵 A ，有 $k(lA) = (kl)A$ ；
- (7) 分配律 I：对任何数 k, l ，任何矩阵 A ，有 $(k + l)A = kA + lA$ ；
- (8) 分配律 II：对任何数 k ，任何矩阵 A, B ，有 $k(A + B) = kA + kB$ 。

回顾诸如向量、多项式、函数等数学对象，发现对这些数学对象，也可以定义加法和数乘运算，且这两种运算也有类似的 8 条性质。将这些共性抽象出来，就得到线性空间的概念。

定义 5.1 设 V 为一个非空集合， F 是一个数域。对 V 中任意两个元素 v_1, v_2 ，存在唯一的 V 中元素与它们对应，称为 v_1 与 v_2 的和，记为 $v_1 + v_2$ ，即在 V 上定义了加法运算。对 F 中任意元素 k 和 V 中任意元素 v ，存在唯一的 V 中元素与它们对应，称为 k 与 v 的积，记为 kv ，即在 V 上定义了数乘运算。这两种运算满足以下 8 条性质：

- (1) 加法交换律：对任何 $v_1, v_2 \in V$ ，有 $v_1 + v_2 = v_2 + v_1$ ；
- (2) 加法结合律：对任何 $v_1, v_2, v_3 \in V$ ，有 $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$ ；

- (3) 存在 $\mathbf{0} \in V$ (称为**零元素**), 使得对任何 $\mathbf{v} \in V$, 有 $\mathbf{0} + \mathbf{v} = \mathbf{v}$;
- (4) 对任何 $\mathbf{v} \in V$, 存在 $-\mathbf{v} \in V$ (称为 **\mathbf{v} 的负元素**), 使得 $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$;
- (5) 对任何 $\mathbf{v} \in V$, $1\mathbf{v} = \mathbf{v}$;
- (6) 对任何 $k, l \in \mathbf{F}$, 任何 $\mathbf{v} \in V$, 有 $k(l\mathbf{v}) = (kl)\mathbf{v}$;
- (7) 分配律 I: 对任何 $k, l \in \mathbf{F}$, 任何 $\mathbf{v} \in V$, 有 $(k+l)\mathbf{v} = k\mathbf{v} + l\mathbf{v}$;
- (8) 分配律 II: 对任何 $k \in \mathbf{F}$, 任何 $\mathbf{v}_1, \mathbf{v}_2 \in V$, 有 $k(\mathbf{v}_1 + \mathbf{v}_2) = k\mathbf{v}_1 + k\mathbf{v}_2$.

此时称 V 为数域 \mathbf{F} 上的**线性空间**或**向量空间**, V 中元素称为**向量**.

在本书中, 向量用加粗的斜体字母表示, 数用通常的斜体字母表示.

实数域或复数域上的线性空间, 在具体问题中用得最频繁, 分别简称为**实线性空间**或**复线性空间**.

以下是几个线性空间的例子.

例 5.1 平面 (或空间) 中全体向量所成的集合, 关于向量的加法和数乘构成实线性空间, 该空间中的零元素是零向量. 一般的线性空间正是这种狭义的向量空间的推广.

例 5.2 全体 $m \times n$ 的实矩阵所成的集合 $\mathbf{R}^{m \times n}$, 关于矩阵加法和数乘构成实线性空间, 该空间中的零元素是零矩阵. 当 $n = 1$ 时, 该空间称为**列矩阵空间**, 简记为 \mathbf{R}^m . 特别地, 实数域 \mathbf{R} 可以看作 1×1 的实矩阵全体构成的实线性空间.

例 5.3 全体系数在数域 \mathbf{F} 上的多项式所成的集合 $\mathbf{F}[x]$, 关于多项式的加法和数乘构成数域 \mathbf{F} 上的线性空间, 该空间中的零元素是零多项式.

例 5.4 定义在 $[a, b]$ 上的实值连续函数全体所成的集合 $C([a, b])$, 关于函数的加法与数乘构成实线性空间, 其零元素为零常值函数.

零元素、负元素、加法、数乘形式上可能完全不同于传统意义上的零、负元素、加法、数乘, 比如下面的例子.

例 5.5 在全体正实数所成的集合 \mathbf{R}^+ 上, 定义加法为 $\mathbf{a} \oplus \mathbf{b} = \mathbf{ab}$, 定义数乘为 $k \cdot \mathbf{a} = \mathbf{a}^k$, 则 \mathbf{R}^+ 关于加法 \oplus 和数乘 \cdot 构成实线性空间, 该空间中的零元素是 1, \mathbf{a} 的负元素为 \mathbf{a}^{-1} .

通过上述例子发现, 线性空间中的元素, 在不同的例子中可以是不同的数学对象, 加法、数乘也就具有相应不同的定义. 从这个意义上说, 零元素、负元素、加法、数乘只是借用了矩阵范畴中的相应称谓, 它们在线性空间中, 扮演了与矩阵集合中零元素、负元素、加法、数乘相同的角色.

线性空间与集合是两个不同的概念, 线性空间包含了四个要素: 集合 V 、数域 \mathbf{F} 、加法和数乘. 即使同一个集合, 如果数域取得不同, 或者加法、数乘的定义方式不同, 就是不同的线性空间, 比如下面的例子.

例 5.6 在全体复数所成的集合 \mathbf{C} 上, 数域 \mathbf{F} 取为复数域, 取加法和数乘分别为复数的加法和乘法, 则 \mathbf{C} 构成复线性空间. 现将集合仍取为全体复数集 \mathbf{C} , 数域取为实数域, 则容易验证 \mathbf{C} 构成实线性空间. 这两个线性空间的本质差别, 将在下节给出.

线性空间中的元素, 由 8 条基本性质, 可以推出下述直观上很自然的性质.

命题 5.1 设 V 为域 \mathbf{F} 上的线性空间, 则

- (1) V 上的零元素是唯一的, 对任何 $v \in V$, v 的负元素是唯一的;
 (2) 消去律: 对任何 $v_1, v_2, w \in V$, 若 $v_1 + w = v_2 + w$, 则 $v_1 = v_2$;
 (3) 对任何 $v \in V$, $0v = \mathbf{0}$, $(-1)v = -v$;
 (4) 若 $kv = \mathbf{0}$, 则 $k = 0$ 或 $v = \mathbf{0}$.

证明 (1) 设 $\mathbf{0}$ 和 $\mathbf{0}'$ 都是零元素, 则 $\mathbf{0} \stackrel{3}{=} \mathbf{0} + \mathbf{0}' \stackrel{3}{=} \mathbf{0}'$ (等号上面的数字 3 表示这一步是根据线性空间的第 3 条基本性质, 下同), 于是零元素唯一.

$$(2) v_1 \stackrel{3,4}{=} v_1 + (w + (-w)) \stackrel{2}{=} (v_1 + w) + (-w) = (v_2 + w) + (-w) \stackrel{2}{=} v_2 + (w + (-w)) \stackrel{3,4}{=} v_2.$$

(3) 因 $0v + v \stackrel{5}{=} 0v + 1v \stackrel{7}{=} (0+1)v = 1v \stackrel{5}{=} v \stackrel{3}{=} \mathbf{0} + v$, 根据消去律得 $0v = \mathbf{0}$. 而 $v + (-1)v \stackrel{7}{=} (1-1)v = 0v = \mathbf{0} \stackrel{4}{=} v + (-v)$, 再根据消去律得 $(-1)v = -v$.

(4) 只须证当 $k \neq 0$ 时必有 $v = \mathbf{0}$. 事实上, 此时 $\mathbf{0} = k^{-1}\mathbf{0} = k^{-1}kv \stackrel{6}{=} (k^{-1}k)v = 1v \stackrel{5}{=} v$.
 证毕.

5.2 线性空间的维数、基与坐标

通过上一节的例子发现, 很多不同的数学集合上, 都可以定义线性空间结构. 本节将指出, 所有的有限维线性空间, 本质上都可以看作列矩阵线性空间.

首先给出一般线性空间中线性相关的定义.

定义 5.2 设 V 为域 \mathbf{F} 上的线性空间, v_1, \dots, v_m 为 V 中的 m 个元素, 若存在 \mathbf{F} 中不全为零的数 l_1, \dots, l_m , 使得

$$l_1v_1 + l_2v_2 + \dots + l_mv_m = \mathbf{0},$$

则称 v_1, \dots, v_m 线性相关, 否则称 v_1, \dots, v_m 线性无关.

当 V 取为矩阵空间时, 这一定义与矩阵空间中的相应定义完全一致.

有了线性相关的定义, 就可以进一步定义向量组的极大线性无关组与秩.

定义 5.3 设 V 为域 \mathbf{F} 上的线性空间, W 为 V 的非空子集. 若 $\{w_1, \dots, w_n\} \subset W$ 且满足

- (1) w_1, \dots, w_n 线性无关,
 (2) 对任何 $w \in W$, 均有 w, w_1, \dots, w_n 线性相关,

则 $\{w_1, \dots, w_n\}$ 称为 W 的一个极大线性无关组, n 称为向量组 W 的秩. 特别, 若 W 中只含零向量, 约定 W 的极大线性无关组为空集, 秩为 0.

对该定义, 有两个问题需要进一步说明. 首先, 对任何向量组, 其极大线性无关组必定存在. 其次, 秩的定义不依赖于极大线性无关组的选取. 对有限向量组 (即 W 为有限集时), 有下述定理.

定理 5.2 对有限向量组 W , 极大线性无关组必定存在, 且任意极大线性无关组所含向量个数相同.

证明 W 为零向量组时, 该定理自然成立, 故可设 W 为非零向量组, 首先证明极大线性无关组的存在性.

任取 W 中的非零向量 w_1 , 令 $S_1 = \{w_1\}$, 则 S_1 作为向量组是一个线性无关组. 若对任何 W 中向量 w , w, w_1 均线性相关, 则 S_1 即为极大线性无关组, 否则, 必存在 W 中向量 w_2 , 使得 w_1, w_2 线性无关, 令 $S_2 = \{w_1, w_2\}$.

若 S_2 已为 W 的极大线性无关组, 则存在性得证, 否则必存在 W 中向量 w_3 , 使得 $S_3 = \{w_1, w_2, w_3\}$ 为线性无关组. 由于 W 中只有有限个元素, 因此必定存在有限数 n , 使得依此法构造的 $S_n = \{w_1, w_2, \dots, w_n\}$ 为 W 的极大线性无关组.

以下证明 n 不随极大线性无关组的选取而改变. 用反证法, 设 v_1, \dots, v_m 也是 W 的一个极大线性无关组, 且 $m > n$. 因对任何 j , 有 v_j, w_1, \dots, w_n 线性相关, 于是存在不全为零的数 k, l_{1j}, \dots, l_{nj} , 使得

$$kv_j + l_{1j}w_1 + l_{2j}w_2 + \dots + l_{nj}w_n = 0.$$

此时易知 $k \neq 0$ (否则 w_1, \dots, w_n 将线性相关), 于是可总取 $k = -1$, 即

$$l_{1j}w_1 + l_{2j}w_2 + \dots + l_{nj}w_n = v_j.$$

令 $A = (l_{ij})_{n \times m}$, 考虑矩阵方程 $AX = 0$. 一方面, $\text{rank} A \leq n < m$, 因此 $AX = 0$ 存在非零解, 设为 x_1, \dots, x_m . 另一方面

$$x_1v_1 + x_2v_2 + \dots + x_mv_m = \left(\sum_{k=1}^m l_{1k}x_k\right)w_1 + \left(\sum_{k=1}^m l_{2k}x_k\right)w_2 + \dots + \left(\sum_{k=1}^m l_{nk}x_k\right)w_n = 0,$$

这与 v_1, \dots, v_m 线性无关矛盾, 于是 $m \leq n$.

同理 $n \leq m$, 从而 $m = n$. 证毕.

向量组的秩有很好的几何意义.

例 5.7 考虑空间向量全体构成的实线性空间 V , $W = \{v_1, v_2, v_3\}$ 为非零向量组, 则 W 的秩为 1 当且仅当 v_1, v_2, v_3 共线; W 的秩为 2 当且仅当 v_1, v_2, v_3 共面且不共线; W 的秩为 3 当且仅当 v_1, v_2, v_3 不共面.

当向量组包含无穷多个向量时, 有限的极大线性无关组可能不存在.

例 5.8 $\mathbf{R}[x]$ 中的向量组 $\{1, x, x^2, \dots\}$, 其任意有限子集作为向量组都是线性无关的.

注 5.1 本书对极大线性无关组的定义, 要求极大线性无关组中只能包含有限个向量, 是出于简单化的需要, 如去掉这一限制, 相关定义将涉及到集合论中的部分概念, 超出了本书的讨论范围.

定义 5.4 设 V 为域 \mathbf{F} 上的线性空间, 若 V 作为向量组, 存在有限的极大线性无关组 v_1, \dots, v_n , 则称 V 为有限维线性空间, n 为 V 的维数, 记为 $\dim V = n$, v_1, \dots, v_n 称为 V 的一组基. 进一步, 对任何 $w \in V$, 存在唯一的 $x_1, \dots, x_n \in \mathbf{F}$, 使得

$$w = x_1v_1 + \dots + x_nv_n, \quad (5.1)$$

记 $\text{crd}(w: v_1, \dots, v_n) = (x_1 \dots x_n)^T$, 称为向量 w 在基 v_1, \dots, v_n 下的坐标.

n 维线性空间在选定一组基后, 任意一个向量可以唯一对应为 \mathbf{F}^n 中的一个列矩阵. 反之任意一个 \mathbf{F}^n 中的列矩阵, 由式 (5.1) 也唯一确定了一个 V 中的向量. 因此选定一组基, 构造了有限维线性空间 V 到列矩阵空间 \mathbf{F}^n 的一一对应. 不仅如此, 这一对应还是保持线性运

算的, 即对任意 $\mathbf{w}_1, \mathbf{w}_2 \in V, l_1, l_2 \in \mathbf{F}$, 有

$$\text{crd}(l_1 \mathbf{w}_1 + l_2 \mathbf{w}_2 : \mathbf{v}_1, \cdots, \mathbf{v}_n) = l_1 \text{crd}(\mathbf{w}_1 : \mathbf{v}_1, \cdots, \mathbf{v}_n) + l_2 \text{crd}(\mathbf{w}_2 : \mathbf{v}_1, \cdots, \mathbf{v}_n),$$

于是一切 n 维线性空间上的线性运算都可对应为 \mathbf{F}^n 上的线性运算.

例 5.9 全体 $m \times n$ 的实矩阵构成的实线性空间 $\mathbf{R}^{m \times n}$, 可选择 $\mathbf{E}_{ij} (1 \leq i \leq m, 1 \leq j \leq n)$

作为该空间的一组基, 于是 $\mathbf{R}^{m \times n}$ 是 $m \times n$ 维实线性空间. 当 $m = n = 2$ 时, 矩阵 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 在

基 $\mathbf{E}_{11}, \mathbf{E}_{12}, \mathbf{E}_{21}, \mathbf{E}_{22}$ 下的坐标为 $(a \ b \ c \ d)^T$.

需要注意的是, 基向量的排列次序不能随意改变, 不同的排列次序, 对应了不同的基, 同一向量的坐标也就随之改变. 在本例中, 如果将基向量的次序排列为 $\mathbf{E}_{11}, \mathbf{E}_{21}, \mathbf{E}_{12}, \mathbf{E}_{22}$, 那相应的坐标应改为 $(a \ c \ b \ d)^T$.

例 5.10 平面中全体向量构成的实线性空间, 任意两个不共线的向量, 都可以作为该空间的一组基, 于是平面向量空间的维数为 2; 空间中全体向量构成的实线性空间, 任意三个不共面的向量, 都可以作为该空间的一组基, 于是空间向量空间的维数为 3.

例 5.11 全体次数小于 n 的实系数多项式所成的集合 $\mathbf{R}_n[x]$, 关于多项式的加法和数乘构成实线性空间, $1, x, \cdots, x^{n-1}$ 为该空间的一组基, 于是 $\dim \mathbf{R}_n[x] = n$.

例 5.12 全体正实数所成的集合 \mathbf{R}^+ , 按例 5.5 中定义的加法和数乘构成实线性空间, 任何不等于 1 的正实数都可以作为它的一组基, 从而 $\dim \mathbf{R}^+ = 1$.

上节指出, 同一个集合, 选取不同的数域, 将得到不同的线性空间, 这两个线性空间本质的差别是维数.

例 5.13 复数集 \mathbf{C} , 作为复线性空间, 是 1 维的, 任何一个非零复数都可作为该线性空间的一组基, 然而若将 \mathbf{C} 看作实线性空间, 则是 2 维的, 此时 $1, i$ 是 \mathbf{C} 的一组基.

在线性空间中选定一组基, 其本质就是解析几何中建立坐标系的思想. 在初等几何课程中, 用纯几何方法研究平面或空间中的点线面的关系是非常困难的, 但一旦在平面或空间建立了坐标系, 所有的点和向量就对应了一个坐标, 点线面的很多性质都转化为坐标的代数运算, 于是很多问题得到了简化, 在一般的线性空间中也是同样的道理.

通过上述例子可以看出, 对一切线性空间, 选定基后, 其中的元素的坐标都是列矩阵, 该线性空间的加法、数乘运算都可以转化为列矩阵的线性运算. 但很多线性空间上还有其他运算, 例如 2 阶方阵就有乘法、行列式等运算, 这些运算与列矩阵的运算没有自然的对应, 这一事实说明 $\mathbf{R}^{2 \times 2}$ 有比线性空间更丰富的内容, 前面提到的很多线性空间亦是如此.

刻画一个线性空间, 只需要指出其任何一组基就够了, 有下述显而易见的定理.

定理 5.3 设 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 是线性空间 V 的一组基, 则 $V = \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$.

线性空间基的选取是不唯一的, 就像平面坐标系的建立方式有无穷多种一样, 在不同的基下, 同一向量的坐标通常不会相同, 在具体应用中, 同一向量在不同基下坐标之间的关系非常重要.

设 $\mathbf{u}_1, \cdots, \mathbf{u}_n$ 和 $\mathbf{v}_1, \cdots, \mathbf{v}_n$ 都是 V 的基, 向量 \mathbf{w} 在这两组基下的坐标分别为 $(x_1 \cdots x_n)^T$

和 $(y_1 \cdots y_n)^T$, 于是

$$\boldsymbol{w} = x_1 \boldsymbol{u}_1 + \cdots + x_n \boldsymbol{u}_n = y_1 \boldsymbol{v}_1 + \cdots + y_n \boldsymbol{v}_n.$$

要建立两组坐标之间的关系, 只需将 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 用 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$ 线性表示即可.

为此, 设

$$\boldsymbol{v}_j = t_{1j} \boldsymbol{u}_1 + \cdots + t_{nj} \boldsymbol{u}_n = \sum_{i=1}^n t_{ij} \boldsymbol{u}_i, \quad (j = 1, \cdots, n),$$

即

$$\text{crd}(\boldsymbol{v}_j : \boldsymbol{u}_1, \cdots, \boldsymbol{u}_n) = (t_{1j} \cdots t_{nj})^T,$$

从而

$$\boldsymbol{w} = \sum_{j=1}^n y_j \boldsymbol{v}_j = \sum_{j=1}^n y_j \left(\sum_{i=1}^n t_{ij} \boldsymbol{u}_i \right) = \sum_{j=1}^n \sum_{i=1}^n t_{ij} y_j \boldsymbol{u}_i = \sum_{i=1}^n \left(\sum_{j=1}^n t_{ij} y_j \right) \boldsymbol{u}_i,$$

由坐标的唯一性得,

$$x_i = \sum_{j=1}^n t_{ij} y_j.$$

若记矩阵

$$\boldsymbol{T}_{uv} = (t_{ij})_{n \times n} = (\text{crd}(\boldsymbol{v}_1 : \boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), \cdots, \text{crd}(\boldsymbol{v}_n : \boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)),$$

则有

$$\text{crd}(\boldsymbol{w} : \boldsymbol{u}_1, \cdots, \boldsymbol{u}_n) = \boldsymbol{T}_{uv} \text{crd}(\boldsymbol{w} : \boldsymbol{v}_1, \cdots, \boldsymbol{v}_n). \quad (5.2)$$

矩阵 \boldsymbol{T}_{uv} 称为由基 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$ 到基 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 的过渡矩阵, 式 (5.2) 称为坐标变换公式.

关于过渡矩阵, 有如下定理, 证明从略.

定理 5.4 设 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$, $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 和 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n$ 分别为线性空间 V 的三组基, 则

- (1) $\boldsymbol{T}_{uu} = \boldsymbol{E}$;
- (2) $\boldsymbol{T}_{uv} \boldsymbol{T}_{vu} = \boldsymbol{E}$;
- (3) $\boldsymbol{T}_{uv} \boldsymbol{T}_{vw} = \boldsymbol{T}_{uw}$.

5.3 子空间

类似于集合有子集合, 线性空间也有子空间的概念.

定义 5.5 设 V 是域 \mathbf{F} 上的线性空间, W 是 V 的非空子集合, 若 W 关于 V 上的加法和数乘构成线性空间, 则 W 称为 V 的子空间.

例 5.14 对任何线性空间 V , V 和 $\{\mathbf{0}\}$ 总是 V 的子空间, 称为平凡子空间.

例 5.15 平面向量全体可以看作空间向量全体的子空间.

例 5.16 定义在区间 $[a, b]$ 上的无穷次连续可微函数全体所成的集合 $C^\infty([a, b])$ 可以看作 $C([a, b])$ 的子空间.

例 5.17 将多项式看作定义在 $[a, b]$ 上的函数, 则 $\mathbf{R}[x]$ 可以看作 $C^\infty([a, b])$ 的子空间.

例 5.18 例 5.5 中的 \mathbf{R}^+ 不能看作 \mathbf{R} 的子空间, 因为两个空间上的加法、数乘定义不一致.

在验证子空间时, 并不需要对子集合逐一验证线性空间的 8 条基本性质, 有如下定理.

定理 5.5 设 V 是域 \mathbf{F} 上的线性空间, W 是 V 的非空子集合, 若 W 关于 V 上的加法和数乘封闭, 即对任何 $w, w_1, w_2 \in W, k \in \mathbf{F}$, 都有 $w_1 + w_2, kw \in W$, 则 W 为 V 的子空间.

证明 若对任何 $w, w_1, w_2 \in W, k \in \mathbf{F}$, 都有 $w_1 + w_2, kw \in W$, 则 V 上的加法、数乘可看作 W 上的加法、数乘, 从而 W 上的加法、数乘满足与 V 上相同的性质, 于是 W 可以看作 V 的子空间. 证毕.

由该定理, 容易验证下面的例子.

例 5.19 设 V 为 \mathbf{F} 上的线性空间, $v_1, \dots, v_m \in V$ 为 m 个给定的向量, 则

$$\text{span}\{v_1, \dots, v_m\} = \{k_1 v_1 + \dots + k_m v_m : k_1, \dots, k_m \in \mathbf{F}\}$$

关于 V 的加法、数乘封闭, 从而为 V 的子空间, 称为由 v_1, \dots, v_m 生成的子空间. 特别, 当 $V = \mathbf{F}^{n \times 1}$ 时, 记矩阵 $A = (v_1 \ v_2 \ \dots \ v_m)$, 则上述子空间也记为 $\mathcal{R}(A)$.

关于子空间与全空间基的关系, 先给出一个很有用的基的扩充定理.

定理 5.6 设 V 为 n 维线性空间, W 为其 k 维子空间 ($k < n$), 设 w_1, \dots, w_k 为 W 的一组基, 则存在 $w_{k+1}, \dots, w_n \in V$, 使得 w_1, \dots, w_n 为 V 的一组基.

证明 取 $w_{k+1} \notin W$, 则 w_1, \dots, w_k, w_{k+1} 线性无关, 若 $W_1 = \text{span}\{w_1, \dots, w_k, w_{k+1}\} = V$, 则 w_1, \dots, w_k, w_{k+1} 为 V 的一组基. 否则再取 $w_{k+2} \notin W_1$, 则 $w_1, \dots, w_k, w_{k+1}, w_{k+2}$ 线性无关. 因 V 为有限维的, 因此必经过有限步后, 可得 w_1, \dots, w_n 为 V 的一组基. 证毕.

对同一个线性空间 V 的两个子空间 V_1 与 V_2 , 有以下两种重要的运算.

定理 5.7 集合 $V_1 \cap V_2$ 和 $V_1 + V_2 = \{v_1 + v_2 : v_1 \in V_1, v_2 \in V_2\}$ 均为 V 的子空间, 分别称为 V_1 与 V_2 的交空间和和空间.

注 5.2 集合 $V_1 \cup V_2$ 通常不是 V 的子空间, 例如在 2 维欧氏空间 \mathbf{R}^2 中, x -轴和 y -轴都是全平面的子空间, 但其并集显然不构成子空间.

和空间和交空间的维数, 满足下面的维数公式.

定理 5.8 设 V_1, V_2 是线性空间 V 的有限维子空间, 则 $\dim V_1 + \dim V_2 = \dim(V_1 \cap V_2) + \dim(V_1 + V_2)$.

证明 取 $V_1 \cap V_2$ 的基 u_1, \dots, u_k , 将其分别扩充为 V_1 的基 $u_1, \dots, u_k, v_1, \dots, v_m$ 和 V_2 的基 $u_1, \dots, u_k, w_1, \dots, w_n$, 以下证明 $u_1, \dots, u_k, v_1, \dots, v_m, w_1, \dots, w_n$ 构成 $V_1 + V_2$ 的基, 于是定理得证.

任何 $V_1 + V_2$ 中的向量可以写为 $u_1, \dots, u_k, v_1, \dots, v_m, w_1, \dots, w_n$ 的线性组合是显然的, 从而只需证明这组向量线性无关即可. 设存在一组数 $x_1, \dots, x_k, y_1, \dots, y_m$ 以及 z_1, \dots, z_n , 使得

$$x_1 u_1 + \dots + x_k u_k + y_1 v_1 + \dots + y_m v_m + z_1 w_1 + \dots + z_n w_n = \mathbf{0},$$

于是

$$x_1 \mathbf{u}_1 + \cdots + x_k \mathbf{u}_k + y_1 \mathbf{v}_1 + \cdots + y_m \mathbf{v}_m = -(z_1 \mathbf{w}_1 + \cdots + z_n \mathbf{w}_n).$$

注意到上式左端的向量在 V_1 中, 右端的向量在 V_2 中, 于是左端向量也在 V_2 中, 从而 $y_1 = \cdots = y_m = 0$, 同理 $z_1 = \cdots = z_n = 0$. 于是

$$x_1 \mathbf{u}_1 + \cdots + x_k \mathbf{u}_k = \mathbf{0},$$

从而 $x_1 = \cdots = x_k = 0$, 故 $\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{w}_1, \cdots, \mathbf{w}_n$ 线性无关. 证毕.

对两个子空间的和而言, 有一种情形是十分特殊的.

定义 5.6 设 V_1, V_2 是线性空间 V 的子空间, 若 $V_1 \cap V_2 = \{\mathbf{0}\}$, 则 V_1, V_2 的和称为直和, 其和空间记为 $V_1 \oplus V_2$.

关于直和的判别, 有下述定理.

定理 5.9 设 V_1, V_2 是线性空间 V 的子空间, 则以下条件等价:

1. V_1 与 V_2 的和是直和;
2. 对任何 $\mathbf{v} \in V_1 + V_2$, 存在唯一分解式 $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, 其中 $\mathbf{v}_1 \in V_1, \mathbf{v}_2 \in V_2$;
3. 若 $\mathbf{0} = \mathbf{v}_1 + \mathbf{v}_2$, 其中 $\mathbf{v}_1 \in V_1, \mathbf{v}_2 \in V_2$, 则 $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$;
4. 对 V_1 的一组基 $\mathbf{w}_1, \cdots, \mathbf{w}_k$ 和 V_2 的一组基 $\mathbf{w}_{k+1}, \cdots, \mathbf{w}_n$, $\mathbf{w}_1, \cdots, \mathbf{w}_n$ 恰好构成 $V_1 + V_2$ 的一组基;
5. $\dim V_1 + \dim V_2 = \dim(V_1 + V_2)$.

证明 设 1 成立, 即 $V_1 \cap V_2 = \{\mathbf{0}\}$. 此时若有 $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}'_1 + \mathbf{v}'_2$, 其中 $\mathbf{v}_1, \mathbf{v}'_1 \in V_1, \mathbf{v}_2, \mathbf{v}'_2 \in V_2$, 则 $\mathbf{v}_1 - \mathbf{v}'_1 = \mathbf{v}'_2 - \mathbf{v}_2$. 另一方面 $\mathbf{v}_1 - \mathbf{v}'_1 \in V_1, \mathbf{v}'_2 - \mathbf{v}_2 \in V_2$, 从而 $\mathbf{v}_1 - \mathbf{v}'_1 = \mathbf{v}'_2 - \mathbf{v}_2 \in V_1 \cap V_2 = \{\mathbf{0}\}$, 即 $\mathbf{v}_1 = \mathbf{v}'_1, \mathbf{v}'_2 = \mathbf{v}_2$, 于是分解是唯一的, 即 1 可推出 2.

由 2 可推出 3 显然.

设 3 成立. 并设 $\mathbf{w}_1, \cdots, \mathbf{w}_k$ 和 $\mathbf{w}_{k+1}, \cdots, \mathbf{w}_n$ 分别为 V_1 与 V_2 的一组基, 将证明 $\mathbf{w}_1, \cdots, \mathbf{w}_n$ 线性无关, 从而构成 $V_1 + V_2$ 的一组基. 事实上, 若

$$l_1 \mathbf{w}_1 + \cdots + l_k \mathbf{w}_k + l_{k+1} \mathbf{w}_{k+1} + \cdots + l_n \mathbf{w}_n = \mathbf{0},$$

则因 $l_1 \mathbf{w}_1 + \cdots + l_k \mathbf{w}_k \in V_1, l_{k+1} \mathbf{w}_{k+1} + \cdots + l_n \mathbf{w}_n \in V_2$, 故

$$l_1 \mathbf{w}_1 + \cdots + l_k \mathbf{w}_k = l_{k+1} \mathbf{w}_{k+1} + \cdots + l_n \mathbf{w}_n = \mathbf{0},$$

从而

$$l_1 = \cdots = l_k = l_{k+1} = \cdots = l_n = 0,$$

于是 $\mathbf{w}_1, \cdots, \mathbf{w}_n$ 线性无关, 即 3 可推出 4.

直接根据维数的定义可知 4 可推出 5.

直接根据维数公式可知 5 可推出 1.

综上, 定理中 5 个条件等价. 证毕.

类似地, 可以定义多个子空间的和与直和.

定义 5.7 设 V_1, \cdots, V_m 均为 n 维线性空间 V 的子空间, 则

$$V_1 + \cdots + V_m = \{\mathbf{v}_1 + \cdots + \mathbf{v}_m : \mathbf{v}_i \in V_i, i = 1, \cdots, m\}$$

构成 V 的子空间, 称为 V_1, \dots, V_m 的**和空间**. 若对任何指标集 $I_1 = \{i_1, \dots, i_k\}$ 和 $I_2 = \{i_{k+1}, \dots, i_l\}$, 满足 $I_1 \cup I_2 \subset \{1, \dots, n\}$, 都有 $V_{i_1} + \dots + V_{i_k}$ 与 $V_{i_1} + \dots + V_{i_l}$ 的和是直和, 则称 V_1, \dots, V_m 的和为**直和**, 并记和空间为 $V_1 \oplus \dots \oplus V_m$.

注 5.3 对多个子空间来说, 即使两两之间的和都是直和, 其整体的和也未必是直和.

对于多个子空间的直和判别, 也有类似的结论.

定理 5.10 设 V_1, \dots, V_m 是线性空间 V 的子空间, 则以下陈述等价:

1. V_1, \dots, V_m 的和是直和;
2. 对任何 $\mathbf{v} \in V_1 + \dots + V_m$, 存在唯一分解式 $\mathbf{v} = \mathbf{v}_1 + \dots + \mathbf{v}_m$, 其中 $\mathbf{v}_i \in V_i, i = 1, \dots, m$;
3. 若 $\mathbf{0} = \mathbf{v}_1 + \dots + \mathbf{v}_m$, 其中 $\mathbf{v}_i \in V_i, i = 1, \dots, m$, 则 $\mathbf{v}_1 = \dots = \mathbf{v}_m = \mathbf{0}$;
4. 存在 V_i 的一组基 $\mathbf{w}_{k_{i-1}+1}, \dots, \mathbf{w}_{k_i}$, 这里 $i = 1, \dots, m, k_0 = 0, k_m = n$, 使得 $\mathbf{w}_1, \dots, \mathbf{w}_n$ 恰好构成 $V_1 + \dots + V_m$ 的一组基;
5. $\dim V_1 + \dots + \dim V_m = \dim(V_1 + \dots + V_m)$.

证明留作习题.

5.4 线性变换

在初等代数课程中, 实数到实数的映射 (即函数) 是一个重要的概念和研究对象, 这一概念在高等代数中的对应, 就是线性空间到其自身的映射.

在所有的函数中, 正比例函数 $f(x) = kx$ 是最简单的一类函数. 正比例函数具有以下性质: 对一切 $x, y, k \in \mathbf{R}$, 满足

$$f(x+y) = f(x) + f(y), f(kx) = kf(x).$$

正比例函数在线性空间上的对应正是本节将介绍的线性变换.

定义 5.8 设 V 为 \mathbf{F} 上的线性空间, \mathcal{A} 为 V 到 V 的映射, 满足

$$(1) \text{ 对任何 } \mathbf{v}, \mathbf{w} \in V, \mathcal{A}(\mathbf{v} + \mathbf{w}) = \mathcal{A}\mathbf{v} + \mathcal{A}\mathbf{w},$$

$$(2) \text{ 对任何 } \mathbf{v} \in V, k \in \mathbf{F}, \mathcal{A}(k\mathbf{v}) = k\mathcal{A}\mathbf{v},$$

则 \mathcal{A} 称为 V 上的一个**线性变换**.

本书中, 线性变换用花体字母表示.

以下是几个线性变换的例子.

例 5.20 V 为线性空间, 则 V 上的映射

$$\begin{array}{ccc} \mathcal{O} : V \rightarrow V & & \text{id} : V \rightarrow V \\ & \text{和} & \\ \mathbf{v} \mapsto \mathbf{0} & & \mathbf{v} \mapsto \mathbf{v} \end{array}$$

均为 V 上的线性变换, 分别称为**零变换**和**恒同变换**.

例 5.21 设 A 为 n 阶方阵, 对 \mathbf{R}^n 上的列矩阵 \mathbf{v} , 定义映射 $\mathcal{A}\mathbf{v} = A\mathbf{v}$, 则 \mathcal{A} 为 \mathbf{R}^n 上的线性变换.

注 5.4 例 5.21 中线性变换的定义式与 \mathbf{R} 上的正比例函数具有相似的形式, 下节将指出, 任意的线性变换都可以转化为这一形式.

例 5.22 空间向量全体组成的 3 维线性空间 \mathbf{R}^3 , \mathcal{A} 定义为将向量向 xOy 平面投影, 则 \mathcal{A} 为 \mathbf{R}^3 上的线性变换.

例 5.23 域 \mathbf{F} 上次数小于 n 的多项式空间 $\mathbf{F}[x]_n$ 上定义映射 $\frac{d}{dx}(f(x)) = \frac{df}{dx}(x)$, 则 $\frac{d}{dx}$ 为 $\mathbf{F}[x]_n$ 上的线性变换.

例 5.24 $[a, b]$ 上无穷次连续可微函数全体构成的线性空间 $C^\infty([a, b])$, 其上定义映射 $\frac{d}{dx} : f(x) \mapsto f'(x)$ 与 $\int_a^x : f(x) \mapsto \int_0^x f(x) dx$, 则 $\frac{d}{dx}$ 与 \int_0^x 均为 $C^\infty([a, b])$ 上的线性变换.

证明上述例子为线性空间是容易的. 下面的例子直观上也很明显.

例 5.25 平面向量全体组成的 2 维线性空间 \mathbf{R}^2 , \mathcal{A} 定义为将向量逆时针旋转 θ 角, \mathcal{B} 定义为将向量关于 x 轴反射, 则 \mathcal{A}, \mathcal{B} 均为 \mathbf{R}^2 上的线性变换.

对线性变换, 可以定义运算, 进而得到更多的线性变换.

定理 5.11 设 V 为 \mathbf{F} 上的线性空间, \mathcal{A}, \mathcal{B} 为 V 上的线性变换, $k \in \mathbf{F}$, 定义 $\mathcal{A} + \mathcal{B}$ 和 $k\mathcal{A}$ 为 V 上的映射, 其在向量 $v \in V$ 上的作用为

$$(\mathcal{A} + \mathcal{B})v = \mathcal{A}v + \mathcal{B}v, \quad (k\mathcal{A})v = k\mathcal{A}v,$$

则 $\mathcal{A} + \mathcal{B}$ 和 $k\mathcal{A}$ 也是 V 上的线性变换. 进一步, 若记 $\text{End}(V)$ 为 V 上一切线性变换的全体, 则 $\text{End}(V)$ 在上述加法、数乘定义下, 构成 \mathbf{F} 上的线性空间.

证明 首先证明 $\mathcal{A} + \mathcal{B}$ 和 $k\mathcal{A}$ 是 V 上的线性变换. 任取 $v_1, v_2 \in V$, 有

$$\begin{aligned} (\mathcal{A} + \mathcal{B})(v_1 + v_2) &= \mathcal{A}(v_1 + v_2) + \mathcal{B}(v_1 + v_2) \\ &= \mathcal{A}v_1 + \mathcal{A}v_2 + \mathcal{B}v_1 + \mathcal{B}v_2 = (\mathcal{A}v_1 + \mathcal{B}v_1) + (\mathcal{A}v_2 + \mathcal{B}v_2) \\ &= (\mathcal{A} + \mathcal{B})v_1 + (\mathcal{A} + \mathcal{B})v_2, \end{aligned}$$

即 $\mathcal{A} + \mathcal{B}$ 保持加法, 类似可证 $\mathcal{A} + \mathcal{B}$ 保持数乘, 过程略, 从而 $\mathcal{A} + \mathcal{B}$ 是 V 上的线性变换. 类似可证 $k\mathcal{A}$ 是 V 上的线性变换, 过程略.

其次证明线性变换关于加法、数乘运算满足线性空间的 8 个条件. 先证明加法交换律 $\mathcal{A} + \mathcal{B} = \mathcal{B} + \mathcal{A}$ 成立, 只须证对任何 $v \in V$, 有 $(\mathcal{A} + \mathcal{B})v = (\mathcal{B} + \mathcal{A})v$. 事实上

$$(\mathcal{A} + \mathcal{B})v = \mathcal{A}v + \mathcal{B}v = \mathcal{B}v + \mathcal{A}v = (\mathcal{B} + \mathcal{A})v.$$

该空间的零元素为零变换, 连同其他 6 条性质的证明都是类似的, 从略. 证毕.

线性变换作为映射, 还有自然的复合运算, 通常将复合运算记为乘法.

定理 5.12 设 V 为 \mathbf{F} 上的线性空间, \mathcal{A}, \mathcal{B} 为 V 上的线性变换, 定义 $\mathcal{A}\mathcal{B}$ 为 V 上的映射, 其在向量 $v \in V$ 上的作用为

$$(\mathcal{A}\mathcal{B})v = \mathcal{A}(\mathcal{B}v),$$

则 $\mathcal{A}\mathcal{B}$ 也是 V 上的线性变换.

证明与上一定理类似, 这里略去.

基于线性变换的乘法运算, 可以进一步定义线性变换的幂与多项式.

定义 5.9 设 V 为 \mathbf{F} 上的线性空间, \mathcal{A} 为 V 上的线性变换, n 为正整数, 定义

$$\mathcal{A}^n = \underbrace{\mathcal{A} \cdots \mathcal{A}}_{n \text{ 个}}.$$

对任意 $\mathbf{F}[x]$ 中的多项式 $f(x) = a_0 + a_1x + \cdots + a_nx^n$, 定义

$$f(\mathcal{A}) = a_0\text{id} + a_1\mathcal{A} + \cdots + a_n\mathcal{A}^n.$$

根据定理 5.11 和定理 5.12, 线性变换的幂与多项式都是线性变换.

若一个线性变换作为映射是可逆的, 可以定义其逆映射, 有如下定理.

定理 5.13 可逆线性变换的逆映射也是线性变换.

证明 设 \mathcal{A} 为 V 上的可逆线性变换, 记 \mathcal{A}^{-1} 为其逆映射. 以下验证 \mathcal{A}^{-1} 为线性变换. 任取 $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w} \in V, k \in \mathbf{F}$, 因

$$\begin{aligned}\mathbf{w}_1 + \mathbf{w}_2 &= \mathcal{A}\mathcal{A}^{-1}\mathbf{w}_1 + \mathcal{A}\mathcal{A}^{-1}\mathbf{w}_2 = \mathcal{A}(\mathcal{A}^{-1}\mathbf{w}_1 + \mathcal{A}^{-1}\mathbf{w}_2), \\ k\mathbf{w} &= k\mathcal{A}\mathcal{A}^{-1}\mathbf{w} = \mathcal{A}(k\mathcal{A}^{-1}\mathbf{w}),\end{aligned}$$

以 \mathcal{A}^{-1} 作用于上式两边, 得到

$$\begin{aligned}\mathcal{A}^{-1}(\mathbf{w}_1 + \mathbf{w}_2) &= \mathcal{A}^{-1}\mathcal{A}(\mathcal{A}^{-1}\mathbf{w}_1 + \mathcal{A}^{-1}\mathbf{w}_2) = \mathcal{A}^{-1}\mathbf{w}_1 + \mathcal{A}^{-1}\mathbf{w}_2, \\ \mathcal{A}^{-1}(k\mathbf{w}) &= \mathcal{A}^{-1}\mathcal{A}(k\mathcal{A}^{-1}\mathbf{w}) = k\mathcal{A}^{-1}\mathbf{w}.\end{aligned}$$

证毕.

5.5 线性变换的矩阵

在线性空间中选定一组基后, 向量的线性运算就可以转化为向量坐标的线性运算. 利用这一思想, 线性变换同样可以通过向量坐标来描述.

设 V 为 n 维线性空间, $\mathbf{v}_1, \dots, \mathbf{v}_n$ 为 V 的一组基, \mathcal{A} 为 V 上的线性变换, 任取 V 上的向量 \mathbf{v} , 考虑 \mathbf{v} 和 $\mathcal{A}\mathbf{v}$ 的坐标之间的关系. 设

$$\text{crd}(\mathbf{v} : \mathbf{v}_1, \dots, \mathbf{v}_n) = (x_1 \cdots x_n)^T,$$

即

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n,$$

于是

$$\mathcal{A}\mathbf{v} = x_1\mathcal{A}\mathbf{v}_1 + \cdots + x_n\mathcal{A}\mathbf{v}_n,$$

从而

$$\text{crd}(\mathcal{A}\mathbf{v} : \mathbf{v}_1, \dots, \mathbf{v}_n) = x_1\text{crd}(\mathcal{A}\mathbf{v}_1 : \mathbf{v}_1, \dots, \mathbf{v}_n) + \cdots + x_n\text{crd}(\mathcal{A}\mathbf{v}_n : \mathbf{v}_1, \dots, \mathbf{v}_n).$$

令 $\mathbf{A} = (\text{crd}(\mathcal{A}\mathbf{v}_1 : \mathbf{v}_1, \dots, \mathbf{v}_n), \dots, \text{crd}(\mathcal{A}\mathbf{v}_n : \mathbf{v}_1, \dots, \mathbf{v}_n))$, 则

$$\text{crd}(\mathcal{A}\mathbf{v} : \mathbf{v}_1, \dots, \mathbf{v}_n) = \mathbf{A}\text{crd}(\mathbf{v} : \mathbf{v}_1, \dots, \mathbf{v}_n). \quad (5.3)$$

公式 (5.3) 给出了 \mathbf{v} 和 $\mathcal{A}\mathbf{v}$ 的坐标之间的关系, 这一关系与例 5.21 中线性变换的形式完全相同, 即对任何线性空间上的线性变换, 选定线性空间的一组基之后, 向量的像和向量自身坐

标之间满足形如式 (5.3) 的简单关系, 式 (5.3) 中的 A 称为线性变换 \mathcal{A} 在基 v_1, \dots, v_n 下的矩阵.

下面给出上一节给出的线性变换在某组基下的矩阵.

例 5.26 V 上的零变换和恒同变换在任何基下的矩阵均为零矩阵和单位矩阵.

例 5.27 选择 e_1, \dots, e_n 为 \mathbf{R}^n 的基, \mathbf{R}^n 上线性变换 $\mathcal{A}v = Av$ 的矩阵恰为 A .

例 5.28 选择 e_1, e_2 为 \mathbf{R}^2 的基, 线性变换 \mathcal{A} 为将向量逆时针旋转 θ 角, \mathcal{B} 为将向量关于 x 轴反射, 则 \mathcal{A}, \mathcal{B} 在 e_1, e_2 下的矩阵分别为

$$A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

例 5.29 选取 e_1, e_2, e_3 为 \mathbf{R}^3 的一组基, \mathcal{A} 为将向量向 xOy 平面投影, 则 \mathcal{A} 在 e_1, e_2, e_3 下的矩阵

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

事实上, 任选 xOy 平面上的线性无关的量 v_1, v_2 , \mathcal{A} 在 v_1, v_2, e_3 下的矩阵均为上述 A .

例 5.30 选取 $1, x, \dots, x^{n-1}$ 为 $\mathbf{F}[x]_n$ 的一组基, 其上线性变换 $\frac{d}{dx}(f(x)) = \frac{df}{dx}(x)$, 则 $\frac{d}{dx}$ 在这组基下的矩阵为

$$A = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & n-1 \\ & & & & 0 \end{pmatrix}.$$

类似向量的坐标保持向量的线性运算, 线性变换的矩阵保持线性变换的加法、数乘和复合运算, 有以下定理.

定理 5.14 设 V 为 \mathbf{F} 上的线性空间, \mathcal{A}, \mathcal{B} 为 V 上的线性变换, $k \in \mathbf{F}$, A, B 分别为 \mathcal{A}, \mathcal{B} 在基 v_1, \dots, v_n 下的矩阵, 则 $\mathcal{A} + \mathcal{B}$, $k\mathcal{A}$ 和 $\mathcal{A}\mathcal{B}$ 在基 v_1, \dots, v_n 下的矩阵分别为 $A + B$, kA 和 AB .

通过这一定理可以看出, 线性变换加法、数乘和乘法, 分别对应了矩阵的相应运算.

推论 5.15 若线性空间 V 上的线性变换 \mathcal{A} 是可逆的, \mathcal{A} 在基 v_1, \dots, v_n 下的矩阵为 A , 则 \mathcal{A} 的逆变换在基 v_1, \dots, v_n 下的矩阵为 A^{-1} .

证明 只须注意到 $\mathcal{A}^{-1}\mathcal{A} = \text{id}$, 而恒同变换在任何基下的矩阵均为单位阵, 根据定理 5.14 即可得证.

与向量在不同基下有不同的坐标一样, 线性变换在不同的基下, 也有不同的矩阵, 不同的矩阵之间也满足一定的关系, 有以下定理.

定理 5.16 设 \mathcal{A} 是线性空间 V 上的线性变换, v_1, \dots, v_n 和 w_1, \dots, w_n 为 V 的两组基, T 为 v_1, \dots, v_n 到 w_1, \dots, w_n 的过渡矩阵, A_v 与 A_w 分别为 \mathcal{A} 在 v_1, \dots, v_n 和 w_1, \dots, w_n 下的矩阵, 则 A_v 与 A_w 相似, 且

$$A_v = T^{-1} A_w T.$$

证明 任取 V 中向量 v , 由式 (5.3),

$$\text{crd}(\mathcal{A}v : v_1, \dots, v_n) = A_v \text{crd}(v : v_1, \dots, v_n), \quad (5.4)$$

将坐标变换公式

$$\text{crd}(\mathcal{A}v : v_1, \dots, v_n) = T \text{crd}(\mathcal{A}v : w_1, \dots, w_n),$$

$$\text{crd}(v : v_1, \dots, v_n) = T \text{crd}(v : w_1, \dots, w_n),$$

代入式 (5.4) 有

$$T \text{crd}(\mathcal{A}v : w_1, \dots, w_n) = A_v T \text{crd}(v : w_1, \dots, w_n),$$

即

$$\text{crd}(\mathcal{A}v : w_1, \dots, w_n) = T^{-1} A_v T \text{crd}(v : w_1, \dots, w_n).$$

另一方面,

$$\text{crd}(\mathcal{A}v : w_1, \dots, w_n) = A_w \text{crd}(v : w_1, \dots, w_n),$$

由 v 的任意性, $A_v = T^{-1} A_w T$. 证毕.

由定理 5.16 知, 线性变换在不同基下的矩阵相似, 而相似的矩阵有相同的行列式、迹、不变因子、特征多项式、零化多项式和最小多项式, 故对线性变换也可以定义行列式、迹、不变因子、特征多项式、零化多项式和最小多项式等概念.

例 5.31 在实线性空间 $\mathbf{R}^{n \times n}$ 上, 定义映射 \mathcal{T} 为 $A \mapsto A^T$, 易证其为线性变换. 一方面, 注意到对一切矩阵, $(A^T)^T = A$, 即 $\mathcal{T}^2 = \text{id}$, 故 $\lambda^2 - 1$ 是 \mathcal{T} 的一个零化多项式. 又因为 $\lambda^2 - 1$ 的任何一个非平凡因式都不是零化多项式, 从而 $\lambda^2 - 1$ 还是该线性变换的最小多项式. 另一方面, 该结论也可以通过线性变换的矩阵得到. 选择 E_{kk} ($1 \leq k \leq n$), $E_{ij} + E_{ji}$, $E_{ij} - E_{ji}$ ($1 \leq i < j \leq n$) 为全空间的基 (基的证明留作习题), 则该变换在这组基下的矩阵为

$$T = \text{diag} \left(E_{\frac{n(n+1)}{2}}, E_{\frac{n(n-1)}{2}} \right),$$

可以由此计算出 $\lambda^2 - 1$ 是 T 的最小多项式.

另有一些关于矩阵的概念, 例如特征值、初等因子、Jordan 标准形等, 虽然在矩阵的相似变换下保持不变, 但依赖于数域的选择, 对线性变换定义相关概念时, 与不依赖于数域选择的概念会有所不同.

定义 5.10 设 \mathcal{A} 为域 \mathbf{F} 上线性空间 V 上的线性变换, 若对 $\lambda \in \mathbf{F}$, 存在非零向量 $v \in V$, 使得 $\mathcal{A}v = \lambda v$, 则称 λ 为线性变换 \mathcal{A} 的**特征值**, 非零向量 v 称为 \mathcal{A} 关于特征值 λ 的**特征向量**.

线性变换的特征值与矩阵的特征值的联系如下.

定理 5.17 设 \mathcal{A} 为域 \mathbf{F} 上线性空间 V 上的线性变换, v_1, \dots, v_n 为 V 的任意一组基,

A 为 \mathcal{A} 在基 v_1, \dots, v_n 下的矩阵, 则线性变换 \mathcal{A} 的特征值就是矩阵 A 在数域 F 上的特征值.

定理 5.17 中, 要求矩阵 A 的特征值须在数域 F 上是必要的, 当数域不为 C 时, 矩阵的特征值未必是线性变换的特征值, 例如下面的例子.

例 5.32 考虑实线性空间 R^2 上的旋转变换, 若旋转角度不为 π 的整数倍, 则任意非零向量旋转后, 都不会与该向量共线 (即线性相关), 故这样的旋转变换没有特征值. 这一结论也可通过定理 5.17 得到. 事实上, 旋转变换相应的矩阵

$$A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

的两个特征值分别为 $e^{\pm i\theta}$, 都不是实数.

5.6 与线性变换相关的子空间

线性空间结合具体的线性变换, 可得到一些特殊的子空间, 这类子空间在研究线性变换时特别重要.

首先是线性变换的核空间和像空间.

定义 5.11 设 V 为 F 上的线性空间, \mathcal{A} 为 V 上的线性变换. 集合

$$\ker \mathcal{A} = \{v \in V : \mathcal{A}v = 0\}$$

称为线性变换 \mathcal{A} 的核, 集合

$$\operatorname{Im} \mathcal{A} = \{\mathcal{A}v : v \in V\}$$

称为线性变换 \mathcal{A} 的像.

线性空间的像与核, 与函数的值域和零点对应的概念, 而且可以证明, 像与核都是子空间.

定理 5.18 设 V 为 F 上的线性空间, \mathcal{A} 为 V 上的线性变换, 则

- (1) $\ker \mathcal{A}$ 和 $\operatorname{Im} \mathcal{A}$ 都是 V 的子空间;
- (2) 若 v_1, \dots, v_n 为 V 的一组基, 则 $\operatorname{Im} \mathcal{A} = \operatorname{span}\{\mathcal{A}v_1, \dots, \mathcal{A}v_n\}$;
- (3) $\dim \ker \mathcal{A} + \dim \operatorname{Im} \mathcal{A} = \dim V$.

证明 (1) 设 $u_1, u_2 \in \ker \mathcal{A}$, $l_1, l_2 \in F$, 则 $\mathcal{A}u_1 = \mathcal{A}u_2 = 0$, 于是

$$\mathcal{A}(l_1u_1 + l_2u_2) = l_1\mathcal{A}u_1 + l_2\mathcal{A}u_2 = 0,$$

从而 $l_1u_1 + l_2u_2 \in \ker \mathcal{A}$, 即 $\ker \mathcal{A}$ 是 V 的子空间. 设 $w_1, w_2 \in \operatorname{Im} \mathcal{A}$, 则存在 $v_1, v_2 \in V$, 使得 $w_1 = \mathcal{A}v_1, w_2 = \mathcal{A}v_2$, 从而

$$l_1w_1 + l_2w_2 = l_1\mathcal{A}v_1 + l_2\mathcal{A}v_2 = \mathcal{A}(l_1v_1 + l_2v_2),$$

于是 $l_1w_1 + l_2w_2 \in \operatorname{Im} \mathcal{A}$.

(2) $\operatorname{span}\{\mathcal{A}v_1, \dots, \mathcal{A}v_n\} \subset \operatorname{Im} \mathcal{A}$ 是显然的, 从而只需证 $\operatorname{Im} \mathcal{A} \subset \operatorname{span}\{\mathcal{A}v_1, \dots, \mathcal{A}v_n\}$ 就可得到 $\operatorname{Im} \mathcal{A} = \operatorname{span}\{\mathcal{A}v_1, \dots, \mathcal{A}v_n\}$. 事实上, 任取 $w \in \operatorname{Im} \mathcal{A}$, 存在 $v \in V$, 使得 $w = \mathcal{A}v$.

因 v_1, \dots, v_n 为 V 的一组基, 于是存在 $l_1, \dots, l_n \in \mathbf{F}$, 使得 $v = l_1 v_1 + \dots + l_n v_n$, 从而

$$w = \mathcal{A}v = \mathcal{A}(l_1 v_1 + \dots + l_n v_n) = l_1 \mathcal{A}v_1 + \dots + l_n \mathcal{A}v_n,$$

即 $w \in \text{span}\{\mathcal{A}v_1, \dots, \mathcal{A}v_n\}$, 亦即 $\text{Im } \mathcal{A} \subset \text{span}\{\mathcal{A}v_1, \dots, \mathcal{A}v_n\}$.

(3) 取 u_1, \dots, u_k 为 $\ker \mathcal{A}$ 的一组基, 并将其扩充为 V 的基 $u_1, \dots, u_k, u_{k+1}, \dots, u_n$, 以下证 $\mathcal{A}u_{k+1}, \dots, \mathcal{A}u_n$ 构成 $\text{Im } \mathcal{A}$ 的一组基, 从而 $\dim \ker \mathcal{A} + \dim \text{Im } \mathcal{A} = \dim V$. 因为

$$\text{Im } \mathcal{A} = \text{span}\{\mathcal{A}u_1, \dots, \mathcal{A}u_k, \mathcal{A}u_{k+1}, \dots, \mathcal{A}u_n\} = \text{span}\{\mathcal{A}u_{k+1}, \dots, \mathcal{A}u_n\},$$

故只需证明 $\mathcal{A}u_{k+1}, \dots, \mathcal{A}u_n$ 线性无关. 设存在 $l_{k+1}, \dots, l_n \in \mathbf{F}$, 使得

$$0 = l_{k+1} \mathcal{A}u_{k+1} + \dots + l_n \mathcal{A}u_n = \mathcal{A}(l_{k+1} u_{k+1} + \dots + l_n u_n),$$

从而 $l_{k+1} u_{k+1} + \dots + l_n u_n \in \ker \mathcal{A}$, 于是 $l_{k+1} = \dots = l_n = 0$, 即 $\mathcal{A}u_{k+1}, \dots, \mathcal{A}u_n$ 线性无关. 证毕.

利用线性变换的矩阵, 结合定理 5.18, 可具体计算出线性变换核空间与像空间的基. 设线性变换的矩阵为 A , 则线性变换的核空间就是齐次线性方程 $Ax = 0$ 的解空间, 于是该方程的一个基础解系, 就是核空间的一组基. 根据定理 5.18, $\text{Im } \mathcal{A} = \text{span}\{\mathcal{A}v_1, \dots, \mathcal{A}v_n\}$, 而 A 的每一列正是 $\mathcal{A}v_1, \dots, \mathcal{A}v_n$ 的坐标, 从而为求像空间的一组基, 只须求出 A 作为列向量组的极大线性无关组即可.

另一类与线性空间相关的重要子空间为线性变换的不变子空间.

定义 5.12 设 \mathcal{A} 为线性空间 V 上的线性变换, W 为 V 的子空间, 若对任何 $w \in W$, 都有 $\mathcal{A}w \in W$, 则称 W 为 \mathcal{A} 的不变子空间. 此时 \mathcal{A} 也可看作 W 上的线性变换, 称为 \mathcal{A} 在 W 上的限制.

下面给出几个不变子空间的例子.

例 5.33 对任何线性空间 V 上的线性变换 \mathcal{A} , 零空间和 V 总是 \mathcal{A} 的不变子空间. 这种不变子空间是平凡的, 没有特别的价值.

例 5.34 设 λ 为线性空间 V 上的线性变换 \mathcal{A} 的特征值, p 为对应的一个特征向量, 则 $\text{span}\{p\}$ 是 \mathcal{A} 的不变子空间. 进一步, 记

$$E_\lambda(\mathcal{A}) = \{v \in V : \mathcal{A}(v) = \lambda v\},$$

则 $E_\lambda(\mathcal{A})$ 构成 V 的子空间, 且为 \mathcal{A} 的不变子空间, 称为 λ 对应的特征子空间.

例 5.35 设 λ 为线性空间 V 上的线性变换 \mathcal{A} 的特征值, 定义

$$R_\lambda(\mathcal{A}) = \{v \in V : \text{存在正整数 } k, \text{ 使得 } (\lambda \text{id} - \mathcal{A})^k(v) = 0\},$$

则 $R_\lambda(\mathcal{A})$ 构成 V 的子空间, 且为 \mathcal{A} 的不变子空间, 称为 λ 对应的根子空间.

$E_\lambda(\mathcal{A})$ 和 $R_\lambda(\mathcal{A})$ 为子空间的证明是容易的, 留作习题.

因任何矩阵在复数域上必定存在特征值, 因此可以借助特征向量构造出不变子空间.

定理 5.19 任何复线性空间上的线性变换, 必定存在一维的不变子空间.

证明 任取该线性变换的一个特征向量 p , 则 $\text{span}\{p\}$ 是 1 维不变子空间. 证毕.

对实线性空间上的线性变换, 其特征值未必存在, 但只要空间维数大于 3, 仍存在低维的非平凡的不变子空间.

定理 5.20 任何实线性空间上的线性变换, 必定存在 1 维或 2 维的不变子空间.

证明 不妨设实线性空间为 \mathbf{R}^n 且 $n \geq 2$, 线性变换为 $X \mapsto AX$. 若矩阵 A 存在实特征值, p 为从属于该特征值的特征向量, 则子空间 $\text{span}\{p\}$ 就是一个一维不变子空间. 若矩阵 A 无实特征值, 则设 $\lambda = a + ib$ 为 A 的一个复特征值, p 为相应的特征向量 (注意此时 p 是一个复的列矩阵), 对 $Ap = \lambda p$ 两边取共轭, 并注意到 $A = \bar{A}$, 得到 $A\bar{p} = \bar{\lambda}\bar{p}$, 即 $\bar{\lambda}$ 也是 A 的特征值, \bar{p} 是相应的特征向量. 此时

$$\text{Re } p = \frac{1}{2}(p + \bar{p}), \quad \text{Im } p = \frac{1}{2i}(p - \bar{p})$$

均为实矩阵, 即为 \mathbf{R}^n 中的向量, 以下证明 $W = \text{span}\{\text{Re } p, \text{Im } p\}$ 为一个二维的不变子空间. 事实上

$$\begin{aligned} A \text{Re } p &= \frac{1}{2}(Ap + A\bar{p}) = \frac{1}{2}((a + ib)p + (a - ib)\bar{p}) = a \text{Re } p - b \text{Im } p \in W, \\ A \text{Im } p &= \frac{1}{2i}(Ap - A\bar{p}) = \frac{1}{2i}((a + ib)p - (a - ib)\bar{p}) = b \text{Re } p + a \text{Im } p \in W \end{aligned}$$

而 $\text{Re } p, \text{Im } p$ 线性无关 (否则 p 与 \bar{p} 线性相关, 这与定理 1.30 矛盾), 从而 W 是 2 维的. 证毕.

注 5.5 由上面两个定理, 读者可以体会到数域不同, 对线性变换的影响. 决定这种差别的根本原因是, 在复数域上, 任何多项式必定有根 (由此称复数域为代数封闭域), 而在实数域上则不然.

利用不变子空间, 可以简化线性变换的矩阵.

定理 5.21 设 \mathcal{A} 为线性空间 V 上的线性变换, V 有子空间直和分解 $V = W_1 \oplus W_2$, w_1, \dots, w_k 和 w_{k+1}, \dots, w_n 分别为子空间 W_1, W_2 的基, 若 W_1 为 \mathcal{A} 的不变子空间, 则 \mathcal{A} 在 w_1, \dots, w_n 下的矩阵为形如

$$\begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$$

的分块上三角矩阵, 这里 A_1, A_2 分别为 k 阶和 $(n-k)$ 阶方阵. 进一步, W_2 为 \mathcal{A} 的不变子空间, 当且仅当 $B = 0$, 即 \mathcal{A} 在 w_1, \dots, w_n 下的矩阵为分块对角阵.

基于上面的定理, 再利用归纳法, 不难得到下面的结论.

定理 5.22 设 \mathcal{A} 为 n 维线性空间 V 上的线性变换, V 有子空间直和分解 $V = W_1 \oplus \dots \oplus W_m$, 则存在 V 的一组基, 使得 \mathcal{A} 在这组基下的矩阵为形如 $\text{diag}(A_1, \dots, A_m)$ 的分块对角阵, 这里 A_i 为 $\dim W_i$ 阶方阵 ($i = 1, \dots, m$). 特别地, 若 V 有 n 个线性无关的特征向量 v_1, \dots, v_n , 则 \mathcal{A} 在基 v_1, \dots, v_n 下的矩阵为对角阵, 此时称 \mathcal{A} 可对角化.

因一切复矩阵都复相似于矩阵的 Jordan 标准形, 因此对复线性空间上的线性变换, 总可以找到一组基, 使得该线性变换在这组基下的矩阵恰为 Jordan 标准形, 但对一般域 (例如实数域) 上的线性空间, 就没有这么好的结果.

习题 5

1. 在 \mathbf{R}^4 中取基

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

(1) 验证 $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \\ -1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 3 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 5 \\ 3 \\ 2 \\ 1 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} 6 \\ 6 \\ 1 \\ 3 \end{pmatrix}$, 也是一组基.

(2) 求由 $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$ 到 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ 的过渡矩阵.

(3) 求向量 $(1, 3, -1, 2)^T$ 在基 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ 下的坐标.

2. 在 $\mathbf{R}^{2 \times 2}$ 中, 取

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{E}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{E}_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \mathbf{E}_4 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

(1) 证明: $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4$ 构成 $\mathbf{R}^{2 \times 2}$ 的一组基.

(2) 已知 $\mathbf{R}^{2 \times 2}$ 中元素 \mathbf{A} 在基 $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4$ 下的坐标为 $(1 \ 2 \ 3 \ 4)^T$, 求 \mathbf{A} .

(3) 求 $\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ 在基 $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4$ 下的坐标.

3. 在 \mathbf{R}^3 中, 取

$$\mathbf{F}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{F}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{F}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

(1) 证明: $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ 构成 \mathbf{R}^3 的一组基.

(2) 已知 \mathbf{R}^3 中元素 \mathbf{A} 在基 $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ 下的坐标为 $(1, 2, 3)^T$, 求 \mathbf{A} .

(3) 求 $\mathbf{B} = (1, 2, 3)^T$ 在基 $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ 下的坐标.

4. 验证

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 3 \\ 7 \\ 10 \end{pmatrix},$$

与

$$w_1 = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}, w_2 = \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}, w_3 = \begin{pmatrix} 1 \\ 1 \\ -6 \end{pmatrix},$$

都可作为 \mathbf{R}^3 的基, 并求出 v_1, v_2, v_3 到 w_1, w_2, w_3 的过渡矩阵.

5. 证明定理 5.3 和定理 5.4 .

6. 设 v_1, v_2, v_3 为 \mathbf{R}^3 中的非零向量

(1) 给出 $\text{span}\{v_1\}$ 的几何解释?

(2) 给出 $\text{span}\{v_1, v_2\}$ 的几何解释?

(3) 给出 $\text{span}\{v_1, v_2, v_3\}$ 的几何解释?

(4) 若 v_1, v_2 线性无关, 则 $\text{span}\{v_1, v_2\}$ 与 $\text{span}\{v_3\}$ 的和空间与交空间的几何解释是什么?

7. 设 V_1, V_2 均为 \mathbf{R}^n 的二维子空间.

(1) 当 $n = 3$ 时, 给出 $V_1 \cap V_2$ 与 $V_1 + V_2$ 可能的维数以及各自的几何解释.

(2) 当 $n = 4$ 时, 给出 $V_1 \cap V_2$ 与 $V_1 + V_2$ 可能的维数以及各自的几何解释.

8. 证明定理 5.7 .

9. 设 V_1, V_2 为线性空间 V 的子空间, 试问 $V_1 \cup V_2$ 是否为 V 的子空间?

10. 设 V_1, V_2, V_3 为线性空间 V 的子空间, 且 $V_1 \cap V_2 \cap V_3 = \{\mathbf{0}\}$, 试问 $V_1 + V_2 + V_3$ 是否为直和?

11. 证明定理 5.10 .

12. 设 V_1, V_2, \dots, V_n 为线性空间 V 的子空间, 举例说明, 即使 V_1, V_2, \dots, V_n 两两的交空间均为零空间, 其和 $V_1 + V_2 + \dots + V_n$ 也未必是直和.

13. 设 $\mathbf{R}^{2 \times 2}$ 为所有 2 阶实方阵关于矩阵的加法和数乘构成的实线性空间, 在 $\mathbf{R}^{2 \times 2}$ 上定义变换 \mathcal{T} 如下: 对任意 $A \in \mathbf{R}^{2 \times 2}$

$$\mathcal{T}(A) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} A \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix}$$

(1) 证明: \mathcal{T} 是 $\mathbf{R}^{2 \times 2}$ 上的一个线性变换;

(2) 求 \mathcal{T} 在 $\mathbf{R}^{2 \times 2}$ 的基

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

下的矩阵.

14. 考虑关于函数的集合 $V = \{(a_2x^2 + a_1x + a_0)e^x : a_0, a_1, a_2 \in \mathbf{R}\}$.

(1) 证明该集合关于函数的线性运算构成 3 维实线性空间.

(2) 证明求导算子 $D : f \mapsto f'$ 为 V 上的线性变换, 并给出 D 在基 $\alpha_1 = x^2e^x$, $\alpha_2 = xe^x$, $\alpha_3 = e^x$ 下的矩阵.

15. 考虑关于函数的集合 $V = \{a \cos x + b \sin x + c : a, b, c \in \mathbf{R}\}$.

(1) 证明该集合关于函数的线性运算构成 3 维实线性空间, 其中

$$W = \{a \cos x + b \sin x : a, b \in \mathbf{R}\}$$

为其 2 维子空间.

(2) 考虑定义在该集合上关于变量 x 的以下运算——求导、在区间 $[-\pi, \pi]$ 的定积分、在原点求值, 则上述哪些变换是 V 上的线性变换? 对线性变换, 给出其在某组基下的矩阵.

(3) 对 (2) 中得到的线性变换, 哪些以 W 为不变子空间?

(4) 对 (2) 中得到的线性变换, 是否存在一组基, 使得其在这组基下的矩阵为对角阵?

16. 将上一题中的所有三角函数, 改为对应的双曲函数 (即 $\cos x$ 和 $\sin x$ 分别改为 $\cosh x$ 和 $\sinh x$), 所有的问题, 结论如何?

17. 设 \mathcal{A} 是线性空间 V 上的线性变换, 且 $\text{Im } \mathcal{A}^2 = \text{Im } \mathcal{A}$, 则 $\mathcal{A}^2 = \mathcal{A}$ 是否成立? 说明理由.

18. 设 \mathcal{A} 是线性空间 V 上的线性变换, 且 $V = \ker \mathcal{A} \oplus \text{Im } \mathcal{A}$, 证明 $\text{Im } \mathcal{A}^2 = \text{Im } \mathcal{A}$. 举例说明一般情况下 $\ker \mathcal{A}$ 和 $\text{Im } \mathcal{A}$ 不构成直和关系?

19. 定义映射 $\mathcal{T} : \mathbf{R}^{2 \times 2} \rightarrow \mathbf{R}^{2 \times 2}$ 为

$$\mathcal{T}(A) = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} A, \quad A \in \mathbf{R}^{2 \times 2}.$$

(1) 证明: \mathcal{T} 是 $\mathbf{R}^{2 \times 2}$ 上的线性变换.

(2) 求 \mathcal{T} 在基

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, E_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, E_4 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

下的矩阵.

(3) 已知 $\mathbf{R}^{2 \times 2}$ 中元素 A 在基 E_1, E_2, E_3, E_4 下的坐标为 $(1 \ 2 \ 3 \ 4)^T$, 求 $\mathcal{T}(A)$.

(4) 求 $\ker \mathcal{T}$ 和 $\text{Im } \mathcal{T}$.

(5) 求 \mathcal{T} 的不变因子和最小多项式.

(6) 是否存在一组基, 使得 \mathcal{T} 在这组基下的矩阵为对角阵? 如存在, 求出这组基和相应的对角阵.

20. 记 $\mathfrak{sl}(n, \mathbf{R})$ 为数域 \mathbf{R} 上全体迹为 0 的 n 阶方阵构成的集合.

- (1) 证明 $\mathfrak{sl}(n, \mathbf{R})$ 为线性空间, 并求其维数和一组基.
- (2) 对方阵 A, B , 定义 $[A, B] = AB - BA$, 证明对一切 A, B , 有 $[A, B] \in \mathfrak{sl}(n, \mathbf{R})$.
- (3) 取定矩阵 $A \in \mathfrak{sl}(n, \mathbf{R})$, 在 $\mathfrak{sl}(n, \mathbf{R})$ 上定义映射 ad_A 为 $\text{ad}_A(M) = [A, M]$, 证明: ad_A 是 $\mathfrak{sl}(n, \mathbf{R})$ 上的线性变换.
- (4) 记 $\mathfrak{so}(n, \mathbf{R})$ 为所有反对称实矩阵构成的集合, 证明 $\mathfrak{so}(n, \mathbf{R})$ 为 $\mathfrak{sl}(n, \mathbf{R})$ 的子空间, 并求其维数和一组基.
- (5) 若 $A, B \in \mathfrak{so}(n, \mathbf{R})$, 证明 $[A, B] \in \mathfrak{so}(n, \mathbf{R})$.
- (6) 取定矩阵 $A \in \mathfrak{so}(n, \mathbf{R})$, 证明: ad_A 是 $\mathfrak{so}(n, \mathbf{R})$ 上的线性变换.
- (7) 取 $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, 分别求 ad_A 作为 $\mathfrak{so}(n, \mathbf{R})$ 上的线性变换和作为 $\mathfrak{sl}(n, \mathbf{R})$ 上的线性变换的一个矩阵.

21. 复数集 \mathbf{C} 上的共轭变换 $z \mapsto \bar{z}$ 是否是 \mathbf{C} 作为复线性空间上的线性变换? 是否是 \mathbf{C} 作为实线性空间上的线性变换?

22. 设矩阵 A 可以相似对角化, 证明: A 可以表示成矩阵 P_1, \dots, P_n 的线性组合, 其中 P_1, \dots, P_n 满足

- (1) 对一切 i , 有 $P_i^2 = P_i$;
- (2) 对一切 $i \neq j$, 有 $P_i P_j = 0$;
- (3) $E = P_1 + \dots + P_n$.

给出具体的构造方法, 并讨论该分解的唯一性.

23. 已知 \mathcal{A} 为线性空间 V 上的线性变换, $v \in V$, $k \geq 1$ 为正整数, 满足 $\mathcal{A}^k v = 0$ 且 $\mathcal{A}^{k-1} v \neq 0$.

- (1) 证明: $v, \mathcal{A}v, \dots, \mathcal{A}^{k-1}v$ 线性无关, 特别 $k \leq \dim V$.
- (2) 证明: $W = \text{span}\{v, \mathcal{A}v, \dots, \mathcal{A}^{k-1}v\}$ 为 \mathcal{A} 的不变子空间.
- (3) 求 \mathcal{A} 在 W 上的限制 $\mathcal{A}|_W$ 在基 $v, \mathcal{A}v, \dots, \mathcal{A}^{k-1}v$ 下的矩阵.

24. 已知 \mathcal{A} 为线性空间 V 上的线性变换, $\alpha_i \in V$, $k_i \geq 1$ 为正整数, 其中 $i = 1, 2, \dots, s$, 满足 $(\mathcal{A} - \lambda_i \text{id})^{k_i} \alpha_i = 0$ 且 $(\mathcal{A} - \lambda_i \text{id})^{k_i-1} \alpha_i \neq 0$, 并记

$$W_i = \text{span}\{\alpha_i, (\mathcal{A} - \lambda_i \text{id})\alpha_i, \dots, (\mathcal{A} - \lambda_i \text{id})^{k_i-1} \alpha_i\}.$$

- (1) 证明: 若 $\lambda_i \neq \lambda_j$, 则 $W_i \cap W_j = \{0\}$.
- (2) 若 $V = W_1 \oplus W_2 \oplus \dots \oplus W_s$, 求 V 的一组基, 使得 \mathcal{A} 在该组基下的矩阵恰为 Jordan 标准形.

25. 设 V 为线性空间, W 为其子空间. 对任何 $u \in V$, 定义集合

$$[u] = \{v \in V : v - u \in W\},$$

称为一个等价类, 其中 u 称为该等价类的一个代表元. 证明以下命题:

(1) $[u] = [0]$ 当且仅当 $u \in W$;

(2) 若 $u_1, u_2 \in V$, 则 $[u_1] = [u_2]$ 当且仅当 $u_1 - u_2 \in W$;

(3) 对任何 $u, u_1, u_2 \in V, l \in \mathbf{R}$, 定义 $[u_1] + [u_2] = [u_1 + u_2]$, $l[u] = [lu]$, 试问如此定义是否合理 (即该定义是否依赖于等价类代表元的选取);

(4) 定义 $V/W = \{[u] : u \in V\}$, 则 V/W 为线性空间 (称为 V 关于 W 的商空间);

(5) 试求出 V/W 的一组基, 并以此证明 $\dim V/W = \dim V - \dim W$;

(6) 若 \mathcal{A} 为 V 上的线性变换, 对一切 $[u] \in V/W$, 定义

$$\tilde{\mathcal{A}}[u] = [\mathcal{A}u],$$

证明该定义是合理的, 且 $\tilde{\mathcal{A}}$ 是 V/W 上的线性变换 (称为 \mathcal{A} 在商空间 V/W 上的诱导).

第6章 内积空间

在线性空间中，向量的运算只涉及到加法和数乘，而在很多实际问题中，向量除了线性运算，还有其他的属性（例如向量的长度、两个向量之间的夹角等），因此需要对线性空间赋予更丰富的内容。

6.1 实内积与欧氏空间

在解析几何中，两个向量之间可以定义内积运算。以空间向量为例，设 $\mathbf{x} = (x_1 \ x_2 \ x_3)^T$, $\mathbf{y} = (y_1 \ y_2 \ y_3)^T$ ， \mathbf{x} 与 \mathbf{y} 的内积 (\mathbf{x}, \mathbf{y}) 定义为

$$(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_2 y_2 + x_3 y_3. \quad (6.1)$$

下面将这种内积推广到一般的实线性空间。

定义 6.1 设 V 为实线性空间，对 V 中任意两个向量 \mathbf{x} 和 \mathbf{y} ，定义运算 $(\mathbf{x}, \mathbf{y}) \in \mathbf{R}$ ，并满足以下条件

(1) 正定性：对任何 $\mathbf{x} \in V$ ， $(\mathbf{x}, \mathbf{x}) \geq 0$ ，等号成立当且仅当 $\mathbf{x} = \mathbf{0}$ ；

(2) 对称性：对任何 $\mathbf{x}, \mathbf{y} \in V$ ， $(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})$ ；

(3) 双线性性：对任何 $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ ， $k, l \in \mathbf{R}$ ， $(k\mathbf{x} + l\mathbf{y}, \mathbf{z}) = k(\mathbf{x}, \mathbf{z}) + l(\mathbf{y}, \mathbf{z})$ ，

则 (\mathbf{x}, \mathbf{y}) 称为 \mathbf{x} 和 \mathbf{y} 的内积，定义了内积运算的实线性空间 V 称为实内积空间，有限维的实内积空间也称为 **Euclid¹ 空间**或欧氏空间。

易见，解析几何中的内积 (6.1) 是一种特殊的内积。除此之外，在常见的实线性空间上，还有如下的内积定义方式。

例 6.1 在 \mathbf{R}^n 中，对 $\mathbf{x} = (x_1 \ \cdots \ x_n)^T$ 与 $\mathbf{y} = (y_1 \ \cdots \ y_n)^T$ ，定义

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i.$$

容易验证上述定义满足内积的条件，称为 \mathbf{R}^n 上的**标准内积**，通常情况下， \mathbf{R}^n 上的内积，都默认为该标准内积。

例 6.2 设 \mathbf{G} 为 n 阶正定矩阵，在对任意 $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ 上定义 $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{G} \mathbf{y}$ ，可以验证上述定义满足内积的条件，于是在这种内积下， \mathbf{R}^n 也构成欧氏空间，当 \mathbf{A} 为单位矩阵时，这

¹Euclid（约公元前330—公元前275），欧几里得，古希腊数学家。

里的内积就是标准内积.

例 6.3 对实线性空间 $C([a, b])$ 上的任意函数 f, g , 定义

$$(f, g) = \int_a^b f(x)g(x)dx, \quad (6.2)$$

可以验证上述定义满足内积的条件, 于是在上述内积定义下, $C([a, b])$ 是实内积空间, 但该空间是无限维的.

通过内积, 可以进一步定义向量的“长度”(在一般的内积空间中, 习惯上称其为范数).

定义 6.2 设 (\cdot, \cdot) 为实线性空间 V 上的内积, $\sqrt{(x, x)}$ 称为向量 x 的范数, 记为 $\|x\|$.

根据内积的三条性质, 可以推出 Cauchy 不等式、范数的基本性质、平行四边形恒等式和极化恒等式.

定理 6.1 (Cauchy¹不等式) 设 (\cdot, \cdot) 为实线性空间 V 上的内积, 则对一切 $x, y \in V$, 有

$$|(x, y)| \leq \|x\| \cdot \|y\|,$$

等号成立当且仅当 x, y 线性相关.

证明 任取实数 k 与向量 x, y , 不妨设 y 为非零向量, 有

$$0 \leq (x + ky, x + ky) = (x, x) + 2k(x, y) + k^2(y, y).$$

将上式右端看作关于 k 的二次函数, 则该函数的判别式非正, 即

$$4(x, y)^2 - 4(x, x)(y, y) \leq 0,$$

从而 $|(x, y)| \leq \|x\| \cdot \|y\|$. 若等号成立, 则上述关于 k 的二次函数存在实零点, 因此 $x + ky = 0$, 即 x, y 线性相关. 证毕.

定理 6.2 (范数的基本性质) 设 V 是实内积空间, $\|\cdot\|$ 为由内积定义的范数, 则

(1) 正定性: 对任意 $x \in V$, $\|x\| \geq 0$, 等号成立当且仅当 $x = 0$;

(2) 齐次性: 对任意 $k \in \mathbf{R}$, $x \in V$, $\|kx\| = |k| \cdot \|x\|$;

(3) 三角不等式: 对任意 $x, y \in V$, $\|x\| + \|y\| \geq \|x + y\|$, 等号成立当且仅当 x 与 y 线性相关.

证明 性质 1 和 2 都可以直接根据内积性质和范数定义得到, 只需证明三角不等式. 事实上, 对任意 $x, y \in V$,

$$\|x + y\|^2 = (x + y, x + y) = \|x\|^2 + 2(x, y) + \|y\|^2 \leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2,$$

于是三角不等式成立, 上式最后一个不等号利用了内积的 Cauchy 不等式. 证毕.

定理 6.3 设 (\cdot, \cdot) 为实线性空间 V 上的内积, $\|\cdot\|$ 为由内积定义的范数, 则

(1) 平行四边形恒等式: 对任意 $x, y \in V$, $2(\|x\|^2 + \|y\|^2) = \|x + y\|^2 + \|x - y\|^2$.

(2) 极化恒等式: 对任意 $x, y \in V$, $4(x, y) = \|x + y\|^2 - \|x - y\|^2$.

该定理的证明可以直接通过内积与范数的关系得到, 留作习题.

注 6.1 上述定理在 \mathbf{R}^2 和 \mathbf{R}^3 中, 恰为初等几何中的基本定理, 于是这些定理可以看作

¹Cauchy (1789–1857), 柯西, 法国数学家.

这些初等几何基本定理在一般实内积空间上的推广. 从而对不同形式的内积, 上述等式或不等式便派生出多种不同的形式.

根据 Cauchy 不等式, 可以定义两个向量之间的夹角.

定义 6.3 实内积空间 V 中任意两个向量 \mathbf{x}, \mathbf{y} 的夹角定义为

$$\arccos \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}.$$

特别地, 当 $(\mathbf{x}, \mathbf{y}) = 0$ 时, 称向量 \mathbf{x} 与 \mathbf{y} 正交.

在 2 维和 3 维欧氏空间中, 如此定义的向量夹角, 与几何上两个向量的夹角一致. 特别, 当两个非零向量正交时, 这两个向量必定不共线, 从而线性无关. 一般地, 有如下定理.

定理 6.4 设 $\mathbf{v}_1, \dots, \mathbf{v}_m$ 为实内积空间 V 中的一组两两正交的非零向量, 则 $\mathbf{v}_1, \dots, \mathbf{v}_m$ 必定线性无关.

证明 设 $l_1\mathbf{v}_1 + \dots + l_n\mathbf{v}_n = \mathbf{0}$, 等式两边与 \mathbf{v}_i 做内积, 有

$$0 = \sum_{j=1}^n l_j(\mathbf{v}_i, \mathbf{v}_j) = l_i(\mathbf{v}_i, \mathbf{v}_i),$$

由 \mathbf{v}_i 非零知 $l_i = 0$, 由 i 的任意性, 定理得证. 证毕.

6.2 标准正交基、度量矩阵与正交补空间

在 2 维与 3 维欧氏空间建立坐标系时, 正交坐标基 (即直角坐标系) 因在计算中有很多便利之处, 成为应用最多的一种坐标系. 本节首先将这一想法推广到一般的欧氏空间.

定义 6.4 若两两正交的向量组 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 为欧氏空间 V 的一组基, 则称 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 为 V 的一组正交基. 进一步, 若对一切 i , $\|\mathbf{e}_i\| = 1$, 则称这一组基为标准正交基.

下面的结论是欧氏空间的基本定理.

定理 6.5 (Gram-Schmidt) 对任何欧氏空间, 标准正交基必定存在.

证明 任取欧氏空间的一组基 $\mathbf{v}_1, \dots, \mathbf{v}_n$, 令

$$\begin{aligned} \mathbf{f}_1 &= \mathbf{v}_1, \\ \mathbf{f}_2 &= \mathbf{v}_2 - \frac{(\mathbf{v}_2, \mathbf{f}_1)}{(\mathbf{f}_1, \mathbf{f}_1)} \mathbf{f}_1, \\ &\dots\dots\dots \\ \mathbf{f}_n &= \mathbf{v}_n - \frac{(\mathbf{v}_n, \mathbf{f}_1)}{(\mathbf{f}_1, \mathbf{f}_1)} \mathbf{f}_1 - \frac{(\mathbf{v}_n, \mathbf{f}_2)}{(\mathbf{f}_2, \mathbf{f}_2)} \mathbf{f}_2 - \dots - \frac{(\mathbf{v}_n, \mathbf{f}_{n-1})}{(\mathbf{f}_{n-1}, \mathbf{f}_{n-1})} \mathbf{f}_{n-1}. \end{aligned} \tag{6.3}$$

直接验证可得, $\mathbf{f}_1, \dots, \mathbf{f}_n$ 是非零的正交向量组, 故 $\mathbf{f}_1, \dots, \mathbf{f}_n$ 是欧氏空间的一组正交基.

再令 $\mathbf{e}_i = \frac{\mathbf{f}_i}{\|\mathbf{f}_i\|}$ ($i = 1, \dots, n$), 则 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 是欧氏空间的一组标准正交基 (长度为 1 的向量也称为单位向量, 最后一步称为向量的单位化). 证毕.

注 6.2 由任何一组基生成正交基的方法 (式 (6.3)) 称为 **Schmidt 正交化法**, 这一方法

有明显的几何意义, 向量 $\frac{(v, f_j)}{(f_j, f_j)} f_j$ 恰为 v 在 f_j 方向上的正交投影, Schmidt 正交化法就是将向量, 减去它在正交基向量上的正交投影的和 (实际就是在正交基向量张成的子空间上的正交投影), 得到新的正交基向量.

由该定理不难得到下面的推论, 证明留作习题.

推论 6.6 设 v_1, \dots, v_n 为欧氏空间 \mathbf{R}^n 的一组基, 则存在 \mathbf{R}^n 一组标准正交基 e_1, \dots, e_n , 使得 v_1, \dots, v_n 到 e_1, \dots, e_n 的过渡矩阵为上三角矩阵.

选择正交基或标准正交基作为欧氏空间的基是很方便的, 例如向量在正交基下的坐标分量可以通过正交投影直接得到.

定理 6.7 设 f_1, \dots, f_n 是欧氏空间 V 的一组正交基, 则对任何 $v \in V$, 有 $v = \sum_{i=1}^n \frac{(v, f_i)}{(f_i, f_i)} f_i$.

证明 设 $v = x_1 f_1 + \dots + x_n f_n$, 等式两边与 f_i 做内积, 有

$$(v, f_i) = \sum_{j=1}^n x_j (f_j, f_i) = x_i (f_i, f_i).$$

得证. 证毕.

进一步, 如果选择欧氏空间的一组标准正交基, 则空间上的内积与坐标的关系, 恰为标准内积的定义式.

定理 6.8 设 e_1, \dots, e_n 为欧氏空间 V 的一组标准正交基, 向量 $x, y \in V$ 在这组基下的坐标分别为 $(x_1 \dots x_n)^T$ 和 $(y_1 \dots y_n)^T$, 则

$$(x, y) = \sum_{i=1}^n x_i y_i.$$

证明 直接计算有

$$(x, y) = \left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j (e_i, e_j) = \sum_{i=1}^n x_i y_i.$$

证毕.

上述定理指出, 所有的欧氏空间上的内积本质上都是 \mathbf{R}^n 上的标准内积. 以下定理进一步指出, 在一般坐标系下, 内积必定可写成例 6.2 的形式.

定理 6.9 设 f_1, \dots, f_n 是欧氏空间 V 的一组基, 对任何 $x, y \in V$, 其坐标 $X = \text{crd}(x : f_1, \dots, f_n) = (x_1 \dots x_n)^T$, $Y = \text{crd}(y : f_1, \dots, f_n) = (y_1 \dots y_n)^T$, 则 $(x, y) = X^T G Y$, 这里 $G = ((f_i, f_j))_{1 \leq i, j \leq n}$ 是正定矩阵, 称为基 f_1, \dots, f_n 下的度量矩阵.

证明 直接计算有

$$(x, y) = \left(\sum_{i=1}^n x_i f_i, \sum_{j=1}^n y_j f_j \right) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j (f_i, f_j) = X^T G Y.$$

G 的正定性可由内积的基本性质立即得到. 证毕.

类似同一线性变换在不同坐标系下的矩阵具有相似关系, 同一个内积在不同坐标系下的

度量矩阵具有合同关系.

定理 6.10 设 v_1, \dots, v_n 和 w_1, \dots, w_n 分别是欧氏空间 V 的两组基, 这两组基下的度量矩阵分别为 G 和 \tilde{G} , 设 P 为由 v_1, \dots, v_n 到 w_1, \dots, w_n 的过渡矩阵, 则 $\tilde{G} = P^T G P$.

证明 任取 $x, y \in V$, 设

$$X = \text{crd}(x : v_1, \dots, v_n), \quad Y = \text{crd}(y : v_1, \dots, v_n),$$

$$\tilde{X} = \text{crd}(x : w_1, \dots, w_n), \quad \tilde{Y} = \text{crd}(y : w_1, \dots, w_n),$$

则 $X = P\tilde{X}, Y = P\tilde{Y}$, 于是

$$(x, y) = X^T G Y = (P\tilde{X})^T G (P\tilde{Y}) = \tilde{X}^T (P^T G P) \tilde{Y}.$$

另一方面 $(x, y) = \tilde{X}^T \tilde{G} \tilde{Y}$, 由 x, y 的任意性得 $\tilde{G} = P^T G P$. 证毕.

该定理有下面的直接推论.

推论 6.11 若 P 为欧氏空间一组标准正交基到另一组标准正交基的过渡矩阵, 则 $PP^T = P^T P = E$, 这样的矩阵称为正交矩阵.

证明 只需注意到, 此时两组基下的度量矩阵都为单位阵即可. 证毕.

结合标准正交基的存在性定理 6.3, 又有如下推论.

推论 6.12 所有正定矩阵都与单位矩阵合同.

对无穷维内积空间, 同样可以引进正交基, 下面是一个著名的例子.

例 6.4 在 $C([- \pi, \pi])$ 上按式 (6.2) 定义内积, 则 $1, \cos x, \sin x, \dots, \cos nx, \sin nx, \dots$ 是一组正交向量, 这正是 **Fourier**¹ 级数的代数背景. 根据 Fourier 级数的理论, 任何一个定义在 $[- \pi, \pi]$ 上的连续函数 $f(x)$ (事实上, 还可以进一步放宽为平方可积函数), 都可以写为

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx), \quad (6.4)$$

这里

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx = \frac{(f(x), 1)}{(1, 1)}, \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx = \frac{(f(x), \cos nx)}{(\cos nx, \cos nx)}, \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx = \frac{(f(x), \sin nx)}{(\sin nx, \sin nx)}. \end{aligned}$$

根据式 (6.4), $1, \cos x, \sin x, \dots, \cos nx, \sin nx, \dots$ 就是 $C([- \pi, \pi])$ 的一组正交基, 无穷维线性空间区别于有限维线性空间的地方在于, 如何定义向量的无穷级数 (注意式 (6.4) 右端并不能简单地看作逐点的数项级数, 熟悉微积分的读者不难发现, 当 $f(\pi) \neq f(-\pi)$ 时, 右端级数的值与左端函数值并不相等), 限于本书的范围, 此处不再展开.

一个空间的标准正交基同样不是唯一的, 如同 2 维欧氏空间中将任意一个直角坐标系旋转后仍为直角坐标系一样.

¹Fourier (1768–1830), 傅里叶, 法国数学家、物理学家.

通过标准正交基, 可以构造欧氏空间的一种重要的子空间直和分解.

定义 6.5 设 W 是欧氏空间 V 的子空间, 记 W^\perp 为一切与 W 中向量都正交的向量所成的集合, 称为 W 在 V 中的正交补空间.

定理 6.13 设 W 是欧氏空间 V 的子空间, 则 V 有直和分解 $V = W \oplus W^\perp$.

证明 设 w_1, \dots, w_k 为 W 的标准正交基, 将其扩充为 V 的标准正交基 w_1, \dots, w_n (只需先扩充为 V 的基, 再 Schmidt 正交化即可), 令 $W' = \text{span}\{w_{k+1}, \dots, w_n\}$, 则 $V = W \oplus W'$. 以下证明 $W' = W^\perp$.

因对一切 $i \in \{k+1, \dots, n\}$, w_i 与一切 W 中元素正交, 故 $w_i \in W^\perp$, 从而 $W' \subset W^\perp$. 而对一切 $w' \in W^\perp$, 设 $w' = l_1 w_1 + \dots + l_n w_n$, 且 w' 与 W 中一切向量都正交, 特别必与 w_i ($1 \leq i \leq k$) 正交, 从而 $0 = (w', w_i) = l_i$, 故 $w' = l_{k+1} w_{k+1} + \dots + l_n w_n \in W'$, 即 $W^\perp \subset W'$. 证毕.

通过正交补空间, 可以得到如下重要结论.

定理 6.14 设 W 是欧氏空间 V 的子空间, v 为 V 中任意取定的向量, 则定义在 W 上的函数 $d(w) = \|w - v\|$ 必定可唯一地取到最小值, 最小值取到当且仅当 $v - w \in W^\perp$.

证明 取 v 在直和分解 $V = W \oplus W^\perp$ 下的唯一分解 $v = v_1 + v_2$, 这里 $v_1 \in W$, $v_2 \in W^\perp$, 于是

$$\begin{aligned} \|v - w\|^2 &= ((v_1 - w) + v_2, (v_1 - w) + v_2) \\ &= \|v_1 - w\|^2 + \|v_2\|^2 + 2(v_1 - w, v_2) = \|v_1 - w\|^2 + \|v_2\|^2 \geq \|v_2\|^2, \end{aligned} \quad (6.5)$$

等号成立当且仅当 $v_1 = w$, 即 w 是唯一的, $w = v_1$ 当且仅当 $v - w \in W^\perp$. 证毕.

从几何上看, 这个定理就是“点到平面(或直线)的距离垂线段最短”在一般内积空间上的推广. 式 (6.5) 中的

$$\|v - w\|^2 = \|v_1 - w\|^2 + \|v_2\|^2$$

就是一般欧氏空间中的勾股定理.

定理 6.15 若欧氏空间中的向量 v, w 正交, 则 $\|v \pm w\|^2 = \|v\|^2 + \|w\|^2$.

将本节的结论与线性变换相结合, 可得到下面的结论.

定理 6.16 设 \mathcal{A} 为 n 维欧氏空间上的线性变换, 则存在标准正交基, 使得 \mathcal{A} 在这组基下的矩阵为如下分块下三角阵

$$R = \begin{pmatrix} A_{11} & & & & \\ \vdots & \ddots & & & \\ A_{k1} & \cdots & A_{kk} & & \\ A_{k+1\ 1} & \cdots & A_{k+1\ k} & \lambda_{2k+1\ 2k+1} & \\ \vdots & & & \vdots & \ddots \\ A_{n1} & \cdots & A_{nk} & \lambda_{n\ 2k+1} & \cdots & \lambda_{nn} \end{pmatrix}. \quad (6.6)$$

这里所有的 λ_{ij} 均为数, 且 $\lambda_{2k+1\ 2k+1}, \dots, \lambda_{nn}$ 为 \mathcal{A} 的全体实特征值; 所有的 A_{ij} 均为矩阵, 列数均为 2, 当 $1 \leq i \leq k$ 时, 行数为 2, 当 $i > k$ 时, 行数为 1.

证明 对 n 做归纳. 当 $n = 1$ 时, 显然成立. 设 $n \leq m$ 时, 结论成立.

当 $n \leq m + 1$ 时, 若 \mathcal{A} 存在实特征值, 记为 λ_{m+1} , 取 \mathbf{e}_{m+1} 为相应的单位特征向量, 记 $W_1 = \text{span}\{\mathbf{e}_{m+1}\}$; 若 \mathcal{A} 无实特征值, 则根据定理 5.20 取其 2 维的不变子空间 W_2 , 并取其标准正交基 $\mathbf{e}_1, \mathbf{e}_2$. 无论哪种情形, 均将其扩充为全空间的标准正交基 $\mathbf{e}_1, \dots, \mathbf{e}_{m+1}$, 此时根据定理 5.21, \mathcal{A} 在这组基下的矩阵分别为

$$\begin{pmatrix} B_1 & \\ * & \lambda_{m+1} \end{pmatrix} \text{ 或 } \begin{pmatrix} A_{11} & \\ * & B_2 \end{pmatrix}.$$

无论哪种情形, 均将矩阵 B_i ($i = 1, 2$) 看作 W_i^\perp 上的线性变换, 根据归纳假设, 可适当选择 W_i^\perp 的标准正交基, 使其矩阵形如 (6.6). 证毕.

注 6.3 如果将定理 6.16 中, 基向量的次序倒置, 则相应的 \mathbf{R} 将变成分块上三角矩阵.

推论 6.17 任何实方阵 \mathbf{A} 均正交相似于形如 (6.6) 的分块下三角矩阵 \mathbf{R} 或形如其转置形式的分块上三角矩阵, 即存在正交矩阵 \mathbf{P} , 使得 $\mathbf{A} = \mathbf{P}^T \mathbf{R} \mathbf{P}$.

证明 只需在 \mathbf{R}^n 上考虑线性变换 $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$, 则由定理 6.16, 存在一组标准正交基, 在这组基下, 该线性变换的矩阵为 \mathbf{R} , 而此时的过渡矩阵为正交矩阵. 证毕.

注 6.4 虽然任何矩阵都相似于 Jordan 标准形, 而 Jordan 标准形为上三角矩阵, 但定理 6.16 和推论 6.30 不能作为 Jordan 标准形定理的直接推论. 首先, 矩阵相似变换到 Jordan 标准形的过渡矩阵通常是复矩阵, 而这里考虑的是实线性空间, 过渡矩阵只能是实矩阵. 其次, 定理 6.16 和推论 6.30 要求的过渡矩阵须为正交矩阵, 一般的过渡矩阵未必为正交矩阵.

6.3 复内积与酉空间

在复线性空间上, 同样可以定义内积, 这在其他学科同样也有很多应用. 复内积和实内积的大部分性质是相同的, 只有个别地方不同, 不同之处应当引起特别的关注.

定义 6.6 设 V 为复线性空间, 对 V 中任意两个向量 \mathbf{x}, \mathbf{y} , 定义运算 $(\mathbf{x}, \mathbf{y}) \in \mathbf{C}$, 并满足以下条件

(1) 正定性: 对任何 $\mathbf{x} \in V$, $(\mathbf{x}, \mathbf{x}) \geq 0$, 等号成立当且仅当 $\mathbf{x} = \mathbf{0}$;

(2) 共轭对称性: 对任何 $\mathbf{x}, \mathbf{y} \in V$, $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})}$;

(3) 共轭双线性性: 对任何 $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $k, l \in \mathbf{C}$, $(k\mathbf{x} + l\mathbf{y}, \mathbf{z}) = k(\mathbf{x}, \mathbf{z}) + l(\mathbf{y}, \mathbf{z})$,

则 (\cdot, \cdot) 称为复线性空间 V 上的一个内积, V 称为复内积空间, 有限维的复内积空间也称为酉空间.

例 6.5 在 \mathbf{C}^n 中, 对 $\mathbf{x} = (x_1 \cdots x_n)^T$ 与 $\mathbf{y} = (y_1 \cdots y_n)^T$, 定义

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \bar{\mathbf{y}} = \sum_{i=1}^n x_i \bar{y}_i.$$

容易验证上述定义满足内积的条件, 称为 \mathbf{C}^n 上的标准酉内积.

注 6.5 标准酉内积定义中, 对第二个向量的坐标分量加复共轭是必要的, 因为在复数域上, 完全平方式并不是正定的, 模平方才是正定的.

通过复内积, 同样可以定义复向量的范数.

定义 6.7 设 (\cdot, \cdot) 为复线性空间 V 上的内积, $\sqrt{(x, x)}$ 称为由内积定义的向量 x 的范数, 记为 $\|x\|$.

复内积空间上的 Cauchy 不等式、范数基本性质与实内积空间上相应结论是相同的, 但极化恒等式要做修正.

定理 6.18 (Cauchy 不等式) 设 (\cdot, \cdot) 为实线性空间 V 上的内积, 则对一切 $x, y \in V$, 有 $|(x, y)| \leq \|x\| \cdot \|y\|$, 等号成立当且仅当 x, y 线性相关.

注 6.6 由 Cauchy 不等式, 可以定义两个非零复向量的夹角为

$$\arccos \frac{|(x, y)|}{\|x\| \cdot \|y\|}.$$

与两个非零实向量夹角定义方式不同, 由于分子上的内积不一定是实数, 因此须加绝对值, 从而复向量的夹角落在区间 $[0, \frac{\pi}{2}]$ 上.

定理 6.19 (范数的基本性质) 设 V 是复内积空间, $\|\cdot\|$ 为由内积定义的范数, 则

(1) 正定性: 对任意 $x \in V$, $\|x\| \geq 0$, 等号成立当且仅当 $x = 0$;

(2) 齐次性: 对任意 $k \in \mathbb{C}$, $x \in V$, $\|kx\| = |k| \cdot \|x\|$;

(3) 三角不等式: 对任意 $x, y \in V$, $\|x\| + \|y\| \geq \|x + y\|$.

定理 6.20 (极化恒等式) 设 (\cdot, \cdot) 为复线性空间 V 上的内积, $\|\cdot\|$ 为由内积定义的范数, 则对任意 $x, y \in V$:

$$4(x, y) = \|x + y\|^2 - \|x - y\|^2 + \|x + iy\|^2 - \|x - iy\|^2.$$

欧氏空间中的大部分结论, 在酉空间都有平行的推广, 例如下面两个定理, 证明与欧氏空间的相应结论类似, 这里略去.

定理 6.21 酉空间上必定存在标准正交基, 在标准正交基下, 酉内积的定义为 \mathbb{C}^n 上的标准酉内积.

定理 6.22 n 阶复方阵 H 称为 **Hermite¹矩阵**, 如果 $H^T = \bar{H}$. 若 n 阶 Hermite 矩阵 H 满足对一切非零的 $z \in \mathbb{C}^n$, 有 $z^T H \bar{z} > 0$, 则称 H 为**正定的**. 任何一组基 z_1, z_2, \dots, z_n 下, 酉空间上的度量矩阵 $((z_i, z_j))_{1 \leq i, j \leq n}$ 必为正定 Hermite 矩阵.

定理 6.23 设 v_1, \dots, v_n 和 w_1, \dots, w_n 分别是酉空间 V 的两组基, 这两组基下的度量矩阵分别为 G 和 \tilde{G} , 设 T 为由 v_1, \dots, v_n 到 w_1, \dots, w_n 的过渡矩阵, 则 $\tilde{G} = T^T G \bar{T}$.

推论 6.24 若 n 阶复方阵 U 满足 $U^T \bar{U} = E$, 这样的矩阵称为**酉矩阵**. 酉空间上, 标准正交基到标准正交基的过渡矩阵必为酉矩阵.

推论 6.25 一切正定 Hermite 矩阵必定与单位矩阵合同. 这里, 两个复矩阵 A 和 B 合同定义为存在可逆矩阵 Q , 使得 $A = Q^T B \bar{Q}$.

¹Hermite (1822–1901), 埃尔米特, 法国数学家.

定理 6.26 设 W 为酉空间 V 的子空间, 定义

$$W^\perp = \{v \in V : \text{对一切 } w \in W, \text{ 有 } (v, w) = 0\},$$

称为 W 的正交补空间, 则 $V = W \oplus W^\perp$.

定理 6.27 设 W 是酉空间 V 的子空间, v 为 V 中任意取定的向量, 则定义在 W 上的函数 $d(w) = \|w - v\|$ 必定可唯一地取到最小值, 最小值取到当且仅当 $v - w \in W^\perp$.

定理 6.28 若酉空间中的向量 v, w 正交, 则 $\|v \pm w\|^2 = \|v\|^2 + \|w\|^2$.

定理 6.29 设 \mathcal{A} 为酉空间上的线性变换, 则存在标准正交基, 使得 \mathcal{A} 在这组基下的矩阵为上(下)三角阵.

推论 6.30 任何复方阵 A 均酉相似于上(下)三角矩阵 R , 即存在酉矩阵 U , 使得 $A = U^H R U$.

定理 6.29 的结论比对应的定理 6.16 更强, 证明方法甚至更加简明. 因为复线性变换总存在特征值, 从而必可分离出 1 维的不变子空间. 该定理也可以借助 Jordan 标准形得到, 见本章习题.

6.4 正规变换

线性空间配备内积后, 与内积存在某些特殊关系的线性变换, 就会显现出某些特别的性质. 其中, 正规变换就是最有代表性的一类线性变换.

为介绍正规变换, 首先介绍共轭变换的概念.

定理 6.31 设 \mathcal{A} 为酉空间 V 上的线性变换, 则存在唯一的线性变换 \mathcal{A}^* , 使得对一切 $x, y \in V$, 有 $(\mathcal{A}x, y) = (x, \mathcal{A}^*y)$, 此时 \mathcal{A}^* 称为 \mathcal{A} 的共轭变换. 进一步, 在标准正交基下, \mathcal{A} 与 \mathcal{A}^* 对应的矩阵互为共轭转置.

证明 首先证明, 共轭变换如果存在, 必然唯一. 若不然, 设 $\mathcal{A}^*, \mathcal{A}_1^*$ 都是 \mathcal{A} 的共轭变换且不相同, 从而存在 $v \in V$, 使得 $w = \mathcal{A}^*v - \mathcal{A}_1^*v \neq 0$. 然而

$$0 = (\mathcal{A}w, v) - (\mathcal{A}w, v) = (w, \mathcal{A}^*v) - (w, \mathcal{A}_1^*v) = (w, \mathcal{A}^*v - \mathcal{A}_1^*v) = (w, w),$$

与 $w \neq 0$ 矛盾, 从而共轭变换的唯一性得证.

以下证明共轭变换的存在性. 取 V 的一组标准正交基, 记 A 为 \mathcal{A} 在这组基下的矩阵, 则

$$(\mathcal{A}x, y) = (A \text{crd } x)^T \overline{\text{crd } y} = (\text{crd } x)^T A^H \overline{\text{crd } y}. \quad (6.7)$$

从而定义 \mathcal{A}^* 在这组标准正交基下的的矩阵为 A^H , 则 \mathcal{A}^* 为 \mathcal{A} 的共轭变换.

任取 V 的一组标准正交基, 易见式 (6.7) 也是成立的, 从而在标准正交基下, \mathcal{A} 与 \mathcal{A}^* 对应的矩阵互为共轭转置. 证毕.

注 6.7 如定理中的酉空间改为欧氏空间, 则只需将结论中的共轭转置改为转置即可, 证明只须去掉所有的共轭运算即可. 本节后续的定理也是如此.

明确了共轭变换的定义后, 正规变换的定义可立即给出.

定义 6.8 若内积空间上的线性变换 \mathcal{A} , 与其共轭变换 \mathcal{A}^* 可交换, 即 $\mathcal{A}\mathcal{A}^* = \mathcal{A}^*\mathcal{A}$,

则该变换称为正规变换.

根据定理 6.31 立即得到下面的结论.

定理 6.32 设 A 为内积空间上的正规变换 \mathcal{A} 在标准正交基下的矩阵, 则 $AA^H = A^H A$, 这样的矩阵 A 称为正规矩阵.

正规矩阵有两种常见的特例.

定理 6.33 酉矩阵和 Hermite 矩阵都是正规矩阵, 特别正交矩阵和实对称矩阵都是实正规矩阵.

正规矩阵有一个重要性质, 该性质直接关联了正规变换的性质.

引理 6.34 分块上三角的正规矩阵, 必定为分块对角矩阵. 特别, 上三角的正规矩阵, 必定是对角矩阵.

证明 设 $A = \begin{pmatrix} A_1 & B \\ \mathbf{0} & A_2 \end{pmatrix}$ 为正规矩阵, 故 $AA^H = A^H A$, 即

$$\begin{pmatrix} A_1 A_1^H + BB^H & BA_2^H \\ A_2 B^H & A_2 A_2^H \end{pmatrix} = \begin{pmatrix} A_1^H A_1 & A_1^H B \\ B^H A_1 & B^H B + A_2^H A_2 \end{pmatrix},$$

从而 $A_1 A_1^H + BB^H = A_1^H A_1$. 两边取迹, 得到 $\text{tr } BB^H = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 = 0$, 故 $B = \mathbf{0}$. 证毕.

定理 6.35 若 W 是酉空间 V 上的正规变换 \mathcal{A} 的不变子空间, 则 W^\perp 也是的不变子空间.

证明 任取 W 一组标准正交基, 并将其扩充为 V 的标准正交基, 则根据定理 5.21 可设 \mathcal{A} 在这组基下的矩阵

$$A = \begin{pmatrix} A_1 & B \\ \mathbf{0} & A_2 \end{pmatrix}.$$

为证 W^\perp 也是 \mathcal{A} 的不变子空间, 只须证明 $B = \mathbf{0}$. 由于 \mathcal{A} 是正规变换, 根据定理 6.32, A 是正规矩阵, 再根据引理 6.34 得到 $B = \mathbf{0}$. 证毕.

定理 6.36 对任何正规变换, 必定存在一组标准正交基, 使其在这组基下的矩阵为对角矩阵.

证明 由定理 6.29, 必定存在一组标准正交基, 使得该变换在这组基下的矩阵为上三角矩阵. 而正规变换在标准正交基下的矩阵又是正规矩阵, 由引理 6.34 知, 该矩阵只能是对角矩阵. 证毕.

注 6.8 如果在欧氏空间上考虑正规变换, 则该结论需要修正为分块对角矩阵, 且每个对角块的阶数不超过 2 (见本章习题). 其原因仍然是数域不同而导致矩阵标准形的不同.

定理 6.37 矩阵为正规矩阵的充分必要条件是該矩阵酉相似于对角矩阵.

证明 必要性是定理 6.36 的直接推论. 对充分性, 只须设 $A = U^H \Lambda U$, 这里 U 和 Λ 分别为酉矩阵和对角阵, 则

$$AA^H = U^H \Lambda U U^H \Lambda^H U = U^H \Lambda \Lambda^H U = U^H \Lambda^H \Lambda U = U^H \Lambda^H U U^H \Lambda U = A^H A.$$

证毕.

作为特殊的正规矩阵, 酉矩阵和 Hermite 矩阵还有很多特别的性质, 这里先列举两条关于特征值的结论, 其他结论将在后续小节中介绍.

定理 6.38 酉矩阵的特征值必为单位复数 $e^{i\theta}$ ($\theta \in \mathbf{R}$).

证明 设 U 为酉矩阵, λ 为其特征值, X 为其特征向量, 以下证明 $|\lambda| = 1$. 考虑酉空间上的通常内积 (UX, UX) . 一方面,

$$(UX, UX) = (UX)^T \overline{UX} = X^T U^T \bar{U} X = X^T \bar{X} = \|X\|^2,$$

另一方面,

$$(UX, UX) = (\lambda X, \lambda X) = |\lambda|^2 \|X\|^2,$$

从而 $|\lambda| = 1$, 即 λ 为单位复数. 证毕.

定理 6.39 Hermite 矩阵的特征值必为实数.

证明 设 H 为 Hermite 矩阵, λ 为其特征值, X 为其特征向量, 以下证明 $\lambda = \bar{\lambda}$. 考虑酉空间上的通常内积 (HX, X) . 一方面,

$$(HX, X) = (\lambda X, X) = \lambda \|X\|^2,$$

另一方面,

$$(HX, X) = (HX)^T \bar{X} = X^T H^T \bar{X} = X^T \overline{HX} = X^T \bar{\lambda} X = \bar{\lambda} \|X\|^2.$$

从而 $\lambda = \bar{\lambda}$, 即 λ 为实数. 证毕.

注 6.9 正交矩阵和实对称矩阵, 分别可看作实的酉矩阵和 Hermite 矩阵, 因此上面的结论, 正交矩阵和实对称矩阵也各自成立.

6.5 正交变换

本节介绍欧氏空间上保内积的线性变换——正交变换.

定义 6.9 设 \mathcal{A} 为欧氏空间 V 上的线性变换, 满足对任何 $v \in V$, 有 $\|\mathcal{A}v\| = \|v\|$, 则 \mathcal{A} 称为 V 上的正交变换.

由定义, 正交变换是保持向量范数不变的线性变换, 事实上正交变换不仅保持向量范数, 还保持向量内积, 于是也是保持向量夹角及空间中任何两点之间的距离.

定理 6.40 设 \mathcal{A} 是欧氏空间 V 上的正交变换, 则对任何 $x, y \in V$, 有 $(\mathcal{A}x, \mathcal{A}y) = (x, y)$.

证明 由极化恒等式 $4(x, y) = \|x + y\|^2 - \|x - y\|^2$, 有

$$\begin{aligned} (\mathcal{A}x, \mathcal{A}y) &= \frac{1}{4}(\|\mathcal{A}x + \mathcal{A}y\|^2 - \|\mathcal{A}x - \mathcal{A}y\|^2) \\ &= \frac{1}{4}(\|\mathcal{A}(x + y)\|^2 - \|\mathcal{A}(x - y)\|^2) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2) = (x, y). \end{aligned}$$

证毕.

注 6.10 上述定理指出, 内积以及范数, 是欧氏空间正交变换下的不变量.

推论 6.41 欧氏空间上的线性变换是正交变换, 当且仅当它将欧氏空间的标准正交基变

为标准正交基, 当且仅当它在标准正交基下的矩阵为正交矩阵.

证明 设 \mathcal{A} 是 n 维欧氏空间 V 上的正交变换, e_1, \dots, e_n 是 V 的一组标准正交基, 则 $\mathcal{A}e_1, \dots, \mathcal{A}e_n$ 也是 V 的一组两两正交的单位向量组, 于是也是 V 上的标准正交基. 又 \mathcal{A} 在基 e_1, \dots, e_n 下的矩阵为

$$A = (\text{crd}(\mathcal{A}e_1 : e_1, \dots, e_n), \dots, \text{crd}(\mathcal{A}e_n : e_1, \dots, e_n)),$$

在标准正交基下, 内积与坐标的关系为 \mathbf{R}^n 上的标准内积关系, 于是对一切 $1 \leq i, j \leq n$,

$$\text{crd}(\mathcal{A}e_i : e_1, \dots, e_n)^T \text{crd}(\mathcal{A}e_j : e_1, \dots, e_n) = (e_i, e_j) = \delta_{ij},$$

从而 $A^T A = E$, 即 A 为正交矩阵.

反之, 若线性变换 \mathcal{A} 在标准正交基下的矩阵为正交矩阵 A , 则 \mathcal{A} 将标准正交基变为标准正交基, 且对任何向量 x , 设其坐标为 X , 有

$$\|\mathcal{A}x\|^2 = (\mathcal{A}x, \mathcal{A}x) = (AX)^T(AX) = X^T A^T A X = X^T X = (x, x),$$

从而 \mathcal{A} 为正交变换. 证毕.

注 6.11 欧氏空间上的正交变换, 在酉空间的对应为酉变换, 即对一切 $v \in V$, 有 $\|\mathcal{A}v\| = \|v\|$ 成立的线性变换. 可以类似地证明, 酉变换保持向量的夹角和范数; 酉变换将酉空间上的一组标准正交基, 变为标准正交基; 酉变换在标准正交基下的矩阵为酉矩阵.

因正交矩阵的行列式必为 ± 1 , 于是可对正交变换以行列式的值进行分类.

定义 6.10 行列式为 1 的正交变换 (矩阵) 称为**第一类正交变换 (矩阵)**, 行列式为 -1 的正交变换 (矩阵) 称为**第二类正交变换 (矩阵)**. 全体 n 阶正交矩阵构成的集合记为 $O(n)$, 全体 n 阶第一类正交矩阵构成的集合记为 $SO(n)$.

定理 6.42 (第一类) 正交变换是可逆的, 且其逆变换也是 (第一类) 正交变换. (第一类) 正交变换的复合是 (第一类) 正交变换.

只需利用 (第一类) 正交矩阵的逆和乘积都是 (第一类) 正交矩阵即可.

例 6.6 \mathbf{R}^2 上的旋转与反射变换都是正交变换, 其中旋转变换是第一类的, 反射变换是第二类的.

\mathbf{R}^2 上的反射可以推广到高维欧氏空间.

定理 6.43 设 v 为 \mathbf{R}^n 中给定的单位向量, 则一切与 v 正交的向量构成 $(n-1)$ 维子空间 W (n 维线性空间的 $(n-1)$ 维子空间也称为一个**超平面**), v 称为该超平面 W 的单位法向量, 线性变换 $\mathcal{A}x = x - 2(x, v)v$ 是一个第二类正交变换, 称为关于超平面 W 的**镜面反射**. 进一步, 设 v 在某标准正交基下的坐标为 N , 则 \mathcal{A} 在该标准正交基下的矩阵为 $E - 2NN^T$, 这一类矩阵称为 **Householder 矩阵**.

证明 易见 $W = \text{span}\{v\}^\perp$, 故 $\dim W = n - 1$. 对一切 $x \in V$, 有

$$\begin{aligned} \|\mathcal{A}x\|^2 &= (\mathcal{A}x, \mathcal{A}x) = (x - 2(x, v)v, x - 2(x, v)v) \\ &= (x, x) + 4(x, v)^2(v, v) - 4(x, v)(x, v) = (x, x) = \|x\|^2, \end{aligned}$$

故 \mathcal{A} 是正交变换. 取 \mathbf{R}^n 的一组标准正交基, 设 x 与 v 在这组基下的坐标分别为 X 与 N , 则 $\mathcal{A}x$ 在这组基下的坐标为

$$X - 2(x, v)N = X - 2(N^T X)N = X - 2N(N^T X) = (E - 2NN^T)X,$$

即 \mathcal{A} 在这组标准正交基下的矩阵为 $E - 2NN^T$.

最后证 \mathcal{A} 是第二类正交变换, 即证明 \mathcal{A} 的行列式为 -1 . 取 w_1, \dots, w_{n-1} 为 w 的一组基, 于是 w_1, \dots, w_{n-1}, v 是 \mathbf{R}^n 的一组基. 注意到

$$\mathcal{A}w_i = w_i - 2(w_i, v)v = w_i = 1 \cdot w_i, \text{ 对一切 } 1 \leq i \leq n-1,$$

$$\mathcal{A}v = v - 2(v, v)v = -v = (-1) \cdot v,$$

从而 \mathcal{A} 在基 w_1, \dots, w_{n-1}, v 下的矩阵为 $\text{diag}(1, \dots, 1, -1)$, 于是其行列式为 -1 . 证毕.

上一段证明蕴含了 w_1, \dots, w_{n-1}, v 是 \mathcal{A} 的特征向量, 其对应特征值为 $1, \dots, 1, -1$.

在应用中, \mathbf{R}^2 和 \mathbf{R}^3 上的正交变换是最常见的, 下面给出这两个空间上正交变换的标准形. 在以下定理的证明中, 将多次用到定理 6.35 和定理 6.38.

定理 6.44 对 \mathbf{R}^2 上的正交变换, 必定存在一组标准正交基, 使得该正交变换在这组基下的矩阵为

$$P_1 = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad \text{或} \quad P_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

证明 设 \mathcal{A} 为 \mathbf{R}^2 上的正交变换, 按其行列式, 分两种情况讨论.

情形1. \mathcal{A} 有两个实特征值 λ_1, λ_2 , 取 e_1 为 λ_1 对应的单位特征向量, 由 \mathcal{A} 为正规变换, 故 $\text{span}\{e_1\}^\perp$ 是 \mathcal{A} 的不变子空间, 取其中的单位向量 e_2 , 则 e_2 必定是 λ_2 对应的特征向量. 于是 \mathcal{A} 在标准正交基 e_1, e_2 下的矩阵为 P_2 或 P_1 对应 θ 为 0 或 π 的情形.

情形2. \mathcal{A} 有一对互为共轭的复特征值, 此时 \mathcal{A} 的行列式为 1 . 任一组标准正交基 e_1, e_2 , 设 \mathcal{A} 在这组基下的矩阵为

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

因 A 须为正交矩阵, 故 $a^2 + c^2 = b^2 + d^2 = 1$, $ab + cd = 0$, 于是可设 $a = \cos \theta_1$, $c = \sin \theta_1$, $b = \cos \theta_2$, $d = \sin \theta_2$, 从而

$$0 = \cos \theta_1 \cos \theta_2 + \sin \theta_1 \sin \theta_2 = \cos(\theta_1 - \theta_2).$$

再由 $ad - bc = 1$, 得

$$1 = \cos \theta_1 \sin \theta_2 - \sin \theta_1 \cos \theta_2 = \sin(\theta_2 - \theta_1).$$

于是 $\theta_2 = \theta_1 + \frac{\pi}{2}$, 代入 A 即得 $A = P_1$. 证毕.

定理 6.45 对 \mathbf{R}^3 上的正交变换, 必定存在一组标准正交基, 使得该正交变换在这组基下的矩阵为

$$Q_1 = \begin{pmatrix} 1 & & \\ & \cos \theta & -\sin \theta \\ & \sin \theta & \cos \theta \end{pmatrix} \quad \text{或} \quad Q_2 = \begin{pmatrix} -1 & & \\ & \cos \theta & -\sin \theta \\ & \sin \theta & \cos \theta \end{pmatrix}.$$

证明 因 3 阶实矩阵的特征多项式为 3 次多项式, 故由连续函数的介值定理知其必有一个实零点, 即 \mathbf{R}^3 上线性变换至少有一个实特征值 λ_1 . 另外两个特征值, 或者均为实数, 或

者为互为共轭的一对复数.

当三个特征值均为实数时, 必定存在两个同号, 此时均可归结为 Jordan 标准形为 Q_1 或 Q_2 , 当 $\theta = 0$ 或 π 的情形, 相应的过渡矩阵必可取为正交矩阵.

当存在一对互为共轭的复特征值时, 设 e_1 为实特征值 λ_1 对应的单位特征向量. 令 $W = \text{span}\{e_1\}^\perp$, 则 W 是 \mathcal{A} 的不变子空间, 故 \mathcal{A} 限制在 W 上可看作 W 上的正交变换, 且行列式为 1, 于是存在 W 的一组基 w_1, w_2 , 则 \mathcal{A} 在基 e_1, w_1, w_2 下的矩阵为

$$A = \begin{pmatrix} \lambda_1 & \\ & A_1 \end{pmatrix}.$$

其中 A_1 为定理 6.44 中的 P_1 , 此时也均可归结为 Q_1 或 Q_2 之一. 证毕.

利用 2 阶和 3 阶正交矩阵的相似标准形, 结合归纳法, 可以给出一般的 n 阶正交矩阵的相似标准形.

定理 6.46 对 \mathbf{R}^n 上的正交变换 \mathcal{A} , 必定存在一组标准正交基, 使得该正交变换在这组基下的矩阵为

$$\text{diag}(R_1, \dots, R_k, \lambda_1, \dots, \lambda_{n-2k}), \quad (6.8)$$

这里 $0 \leq k \leq n$ 为确定的整数, $\lambda_j = \pm 1$ ($1 \leq j \leq n - 2k$),

$$R_i = \begin{pmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{pmatrix} \quad (1 \leq i \leq k).$$

证明 由于 $n = 2, 3$ 时结论成立, 归纳假设当维数小于 n 时结论成立. 对 n 维情形 ($n \geq 4$), 根据定理 5.20, \mathcal{A} 必定存在 1 维或 2 维的不变子空间. 取 W 为 \mathcal{A} 的维数最小的非平凡不变子空间, 则 W^\perp 也是 \mathcal{A} 的不变子空间, 则根据归纳假设, 存在 W 和 W^\perp 的标准正交基使得 \mathcal{A} 在其上的限制的矩阵分别为式 (6.8) 中情形. 而 $\mathbf{R}^n = W \oplus W^\perp$, 于是适当调整基向量的次序, 可得 \mathcal{A} 在 \mathbf{R}^n 上的矩阵为式 (6.8) 中情形. 证毕.

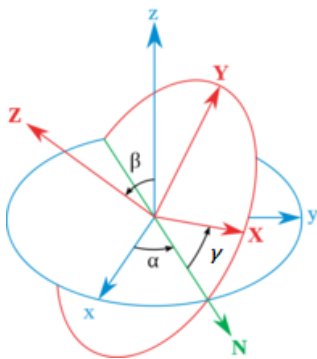
推论 6.47 任何 n 阶正交矩阵必定相似且合同于分块对角阵 (6.8).

借助上面的定理, 立即得到在物理、化学、工程中都有广泛应用的正交矩阵的 Euler 角表示, 证明留作习题.

定理 6.48 (Euler 角) 任何一个第一类正交矩阵 A , 一定可以分解为如下形式

$$A = \begin{pmatrix} \cos \gamma & -\sin \gamma \\ \sin \gamma & \cos \gamma \end{pmatrix} \begin{pmatrix} 1 & \\ & \cos \beta & -\sin \beta \\ & \sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}, \quad (6.9)$$

这里 α, β, γ 为实数, 称为正交矩阵 A 的 Euler 角.



6.6 对称变换

本节介绍欧氏空间上另一种特殊的线性变换——对称变换.

定义 6.11 设 \mathcal{A} 为欧氏空间 V 上的线性变换, 满足对任何 $x, y \in V$, 有 $(\mathcal{A}x, y) = (x, \mathcal{A}y)$, 则 \mathcal{A} 称为对称变换.

类似正交变换, 有下面的定理.

定理 6.49 欧氏空间上的线性变换是对称变换, 当且仅当它在标准正交基下的矩阵为实对称矩阵.

证明 设 \mathcal{A} 为欧氏空间 V 上的对称变换, e_1, \dots, e_n 为 V 的一组标准正交基, A 为 \mathcal{A} 在这组基下的矩阵. 因

$$\text{ent}_{ij} A = (\mathcal{A}e_j, e_i) = (e_j, \mathcal{A}e_i) = (\mathcal{A}e_i, e_j) = \text{ent}_{ji} A,$$

故 A 为对称矩阵.

反之, 设 A 为 \mathcal{A} 在标准正交基 e_1, \dots, e_n 下的矩阵, 任取 $x, y \in V$, 以 X, Y 记它们在基 e_1, \dots, e_n 下的坐标, 则由 A 是对称的, 有

$$(\mathcal{A}x, y) = (AX)^T Y = X^T A^T Y = X^T (AY) = (x, \mathcal{A}y),$$

从而 \mathcal{A} 是对称变换. 证毕.

定理 6.50 欧氏空间上对称变换从属于不同特征值的特征向量必定正交.

证明 设 λ_1, λ_2 为 V 上对称变换 \mathcal{A} 的两个不同的特征值, e_1, e_2 为相应的特征向量, 则

$$\lambda_1(e_1, e_2) = (\lambda_1 e_1, e_2) = (\mathcal{A}e_1, e_2) = (e_1, \mathcal{A}e_2) = (e_1, \lambda_2 e_2) = \lambda_2(e_1, e_2),$$

从而 $(e_1, e_2) = 0$. 证毕.

对称变换同样有非常简单的标准形.

定理 6.51 任何欧氏空间上的对称变换, 都存在一组标准正交基, 使其在这组基下的矩阵为对角阵, 即任何实对称矩阵必正交相似于对角阵.

证明 设 \mathcal{A} 为欧氏空间 V 上的对称变换, 当 $\dim V = 1$, 结论是显然的. 以下对 $\dim V$ 做归纳, 设 $\dim V < n$ 时结论是成立的. 当 $\dim V = n$ 时, 设 λ_1 为 \mathcal{A} 的实特征值, e_1 为相应的单位特征向量, 考虑 $W = \text{span}\{e_1\}^\perp$, 由定理 6.35 为 \mathcal{A} 的不变子空间. 由归纳假设, 存

在 W 的标准正交基 $\mathbf{e}_2, \dots, \mathbf{e}_n$, 使得 \mathcal{A} 限制在 W 上的矩阵为对角阵, 从而 \mathcal{A} 在 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 下的矩阵为对角阵. 证毕.

对称矩阵 Jordan 标准形的一个直接应用是化简二次型.

例 6.7 考虑 n 元实函数

$$f(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j.$$

不失一般性, 设 $a_{ij} = a_{ji}$. 当该函数只含有完全平方项时, 即对一切 $i \neq j$, 系数 $a_{ij} = 0$ 时, 该函数的分析性质 (比如函数图像如何, 函数临界值的性质如何等) 是容易得出的, 对一般情形, 希望能将其化为只含有完全平方项的二次型. 容易想到的方法是配方, 但配方换元对应的坐标变换不一定是正交变换, 于是换元后的函数图像与原函数的图像形状不一定相同, 因此希望寻求一个正交坐标变换将其化为只有完全平方项的二次型.

为此, 令 $\mathbf{A} = (a_{ij})$ 为实对称矩阵, $\mathbf{x} = (x_1 \cdots x_n)^T$, 则 $f(x_1, \dots, x_n) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, 根据上面的定理, 存在正交矩阵 \mathbf{P} , 使得 $\mathbf{A} = \mathbf{P}^T \mathbf{J} \mathbf{P}$, 这里 $\mathbf{J} = \text{diag}(\mu_1, \dots, \mu_n)$ 为实对角阵, 每个对角元为 \mathbf{A} 的特征值, 于是若令 $\mathbf{y} = (y_1 \cdots y_n)^T = \mathbf{P} \mathbf{x}$, 则

$$f(x_1, \dots, x_n) = \mathbf{y}^T \mathbf{J} \mathbf{y} = \mu_1 y_1^2 + \dots + \mu_n y_n^2,$$

且该坐标变换是正交变换.

当 \mathbf{P} 只是一般的可逆矩阵的时候, 若存在对角阵 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, 使得 $\mathbf{A} = \mathbf{P}^T \mathbf{\Lambda} \mathbf{P}$, 此时 \mathbf{A} 和 $\mathbf{\Lambda}$ 只是合同关系, $\mathbf{\Lambda}$ 的对角元通常不再是 \mathbf{A} 的特征值, 但仍有一些不变量存在, 这正是第 1 章中的定理 1.38.

定理 6.52 若实对称矩阵 \mathbf{A} 与 \mathbf{B} 合同, 充分必要条件是 \mathbf{A} 与 \mathbf{B} 有相同的正惯性指数和负惯性指数.

必要性的证明已在第 1 章中给出, 以下证明充分性. 为此先证明下述引理.

引理 6.53 设 W 为 \mathbf{R}^n 的子空间, \mathbf{A} 为对称矩阵, $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ 为 \mathbf{R}^n 的二次型. 若对任何非零向量 $\mathbf{x} \in W$, 都有 $f(\mathbf{x}) > 0$ (< 0), 就称 f 限制在子空间 W 上为正 (负) 定的, 则子空间 W 的最大维数为 \mathbf{A} 的正 (负) 惯性指数.

证明 设 $\lambda_1, \dots, \lambda_p$ 为 \mathbf{A} 的正特征值全体, $\mathbf{e}_1, \dots, \mathbf{e}_p$ 为相应的单位特征向量, 则 f 限制在子空间 $W_+ = \text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_p\}$ 是正定的. 事实上, 对任何 $\mathbf{x} = x_1 \mathbf{e}_1 + \dots + x_p \mathbf{e}_p$, x_1, \dots, x_p 不全为零, 则

$$f(\mathbf{x}) = \lambda_1 x_1^2 + \dots + \lambda_p x_p^2 > 0.$$

于是子空间 W 的最大维数至少为 p .

以下证明当 $\dim W > p$ 时, f 限制在子空间 W 上不可能是正定的. 由于 W_+^\perp 是 \mathbf{A} 的全体非正特征值对应的特征向量张成的子空间, 从而对一切 $\mathbf{x} \in W_+^\perp$, $f(\mathbf{x}) \leq 0$. 考虑 $W \cap W_+^\perp$, 显然该交空间不可能只包含零元素, 否则 $W + W_+^\perp$ 将是直和, 但

$$\dim W + \dim W_+^\perp > p + (n - p) = n,$$

这是不可能的. 于是存在非零向量 $\mathbf{x} \in W \cap W_+^\perp$, 从而 $f(\mathbf{x}) \leq 0$, 即 f 限制在子空间 W 上不

是正定的.

同理可证另一半结论. 证毕.

定理 6.52 的充分性证明 设 $A = P^T B P$, P 为可逆矩阵, 令 $y = P x$, 考虑二次型 $f(x) = x^T A x$ 和 $g(y) = y^T B y$. 因 $f(x) = g(y)$, 故 $f(x)$ 和 $g(y)$ 可看作同一个二次型在不同坐标系下的表示, 于是应有相同的正定子空间和负定子空间. 从而根据引理 6.53, A 与 B 有相同个数的正特征值和负特征值. 证毕.

注 6.12 欧氏空间上的对称变换, 在酉空间上的对应为 **Hermite 变换**, 即对一切 $x, y \in V$, 满足 $(\mathcal{A}x, y) = (x, \mathcal{A}y)$ 的线性变换. 类似可证, Hermite 变换在标准正交基下的矩阵为 Hermite 矩阵, 对任何 Hermite 变换, 均存在标准正交基使其在这组基下的矩阵为实对角矩阵.

习题 6

1. 证明内积的平行四边形恒等式和极化恒等式 (定理 6.3).
2. 求 \mathbf{R}^3 上的一组标准正交基 v_1, v_2, v_3 , 其中 v_1 与向量 $(1, 1, 1)^T$ 线性相关.
3. 已知 \mathbf{R}^3 中的向量 $v = (1, 2, -1)^T$, $w = (1, 1, 0)^T$, 求一个与 v, w 都正交的单位向量.
4. 设 A, B, C, D 为实数.

(1) 从几何的观点解释 \mathbf{R}^2 中的直线方程为 $Ax + By = C$, 其中 A, B 不全为零.

(2) 从几何的观点解释 \mathbf{R}^3 中的平面方程为 $Ax + By + Cz = D$, 其中 A, B, C 不全为零.

(3) 从几何的观点解释二元、三元线性方程组解的结构定理.

5. 在 $[-\pi, \pi]$ 上全体连续函数构成的线性空间 $C([-\pi, \pi])$ 上定义内积为

$$(f, g) = \int_{-\pi}^{\pi} f(x)g(x) dx.$$

证明: $1, \cos x, \sin x, \dots, \cos mx, \sin mx, \dots$ 为该内积空间上的一个正交组.

6. 在一元多项式函数构成的线性空间 $\mathbf{R}[x]$ 上定义内积为

$$(f, g) = \int_{-1}^1 f(x)g(x) dx.$$

试求 $1, x, x^2, x^3$ 的 Schmidt 正交化.

7. 证明推论 6.6.

8. 在 \mathbf{R}^3 上, 定义向量 $\mathbf{x} = (x_1 \ x_2 \ x_3)^T$, $\mathbf{y} = (y_1 \ y_2 \ y_3)^T$ 的外积如下

$$\mathbf{x} \times \mathbf{y} = \begin{pmatrix} \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} & \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix} & \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \end{pmatrix}^T.$$

$$(1) \text{ 证明: 若 } \mathbf{z} = (z_1 \ z_2 \ z_3)^T \in \mathbf{R}^3, \text{ 则 } (\mathbf{x} \times \mathbf{y}, \mathbf{z}) = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}.$$

(2) 证明: $(\mathbf{x} \times \mathbf{y}, \mathbf{x}) = (\mathbf{x} \times \mathbf{y}, \mathbf{y}) = 0$.

(3) 证明: $\|\mathbf{x} \times \mathbf{y}\| = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \sin \theta$, 其中 θ 为 \mathbf{x} 与 \mathbf{y} 的夹角.

(4) 试给出 $\|\mathbf{x} \times \mathbf{y}\|$ 的几何意义.

(5) 试给出 3 阶、2 阶行列式的几何意义.

(6) 外积是否满足结合律, 即 $(\mathbf{x} \times \mathbf{y}) \times \mathbf{z} = \mathbf{x} \times (\mathbf{y} \times \mathbf{z})$ 是否一定成立?

9. 设 $\mathbf{A}, \mathbf{B} \in \mathbf{O}(n)$ (即为 n 阶正交矩阵).

(1) 当 $n = 2$ 时, 求出 \mathbf{A} 的所有可能的 Jordan 标准形, 并解释其几何意义.

(2) 当 $n = 2$ 时, 证明 $\mathbf{AB} = \mathbf{BA}$, 并解释其几何意义.

(3) 当 $n = 3$ 时, 求出 \mathbf{A} 的所有可能的 Jordan 标准形, 并解释其几何意义.

(4) 当 $n = 3$ 时, 是否有 $\mathbf{AB} = \mathbf{BA}$?

10. 证明关于 Euler 角的定理 6.48.

11. 设 $\mathbf{v} \in \mathbf{R}^n$, $\mathbf{A} = \mathbf{v}\mathbf{v}^T$, 求 \mathbf{A} 的正交对角化.

$$12. \text{ 设 } \mathbf{A} = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}.$$

(1) 计算 \mathbf{A} 的实特征值和相应的特征向量.

(2) 证明 \mathbf{A} 正交相似于 $\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\omega \\ 0 & \omega & 0 \end{pmatrix}$ 型矩阵, 并求出 ω 的值.

(3) 计算 $\mathbf{e}^{\mathbf{A}t}$, 并指出 (1) (2) 所求的特征向量和 ω 的几何意义.

13. 证明: 实反对称矩阵 ($\mathbf{A}^T = -\mathbf{A}$) 与反 Hermite 矩阵 ($\mathbf{A}^H = -\mathbf{A}$) 的特征值为纯虚数.

14. 证明下述结论:

(1) 酉矩阵必定酉相似于对角元为单位复数的对角矩阵;

(2) Hermite 矩阵必定酉相似于实对角矩阵;

(3) 反 Hermite 矩阵必定酉相似于对角元为纯虚数的对角矩阵.

15. 设 V 为 Euclid 空间, \mathcal{A} 为 V 上的对称变换, 若对一切非零向量 $\mathbf{v} \in V$, 均有 $(\mathcal{A}\mathbf{v}, \mathbf{v}) > 0$, 这样的对称变换称为**正定的**. 求证: 正定的对称变换在标准正交基下的矩阵为正定矩阵.

16. 试给出一个既不是对称变换, 也不是正交变换的正规变换.

17. 设 A 为 n 阶实正规矩阵, 则 A 必定正交相似于分块对角阵

$$\text{diag}(\rho_1 \mathbf{R}_1, \cdots, \rho_k \mathbf{R}_k, \lambda_1, \cdots, \lambda_{n-2k}),$$

这里 $0 \leq k \leq n$ 为某个确定的整数, $\lambda_j \in \mathbf{R} (1 \leq j \leq n-2k)$, $\rho_i > 0$,

$$\mathbf{R}_i = \begin{pmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{pmatrix} \quad (1 \leq i \leq k).$$

18. 设 A 为 n 阶反对称矩阵, 证明: 上题中的分块对角阵可进一步简化为 $\lambda_j = 0 (1 \leq j \leq$

$$n-2k), \rho_i > 0, \mathbf{R}_i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (1 \leq i \leq k).$$

19. 设线性空间 V 上的线性变换 \mathcal{A} , 在基 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ 下的矩阵恰为 Jordan 标准形, 若将 $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ 做 Schmidt 正交化得到标准正交基 $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$, 则 \mathcal{A} 在这组标准正交基下的矩阵为何种类型, 由此是否可得到定理 6.29?

20. 设矩阵 $A = (a_{ij})_{n \times n} \in \mathbf{C}^{n \times n}$ 为正规矩阵, $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为 A 的 n 个特征值, 证明 $\sum_{i=1}^n |\lambda_i|^2 = \sum_{i,j=1}^n |a_{ij}|^2$. 若 A 不是正规的, 结论将如何修正?

21. 设 V 为 Euclid 空间, \mathcal{P} 为 V 上的线性变换, 且对任何 $\mathbf{v} \in V$, 均有 $\mathbf{v} - \mathcal{P}\mathbf{v} \in (\text{Im } \mathcal{P})^\perp$, 这样的线性变换 V 上的**投影变换**.

(1) 证明: 对任何 V 上的投影变换 \mathcal{P} , 有 $\ker \mathcal{P} = (\text{Im } \mathcal{P})^\perp$. 该命题的逆命题是否成立?

(2) 证明: 线性变换 \mathcal{P} 为 V 上的投影变换的充分必要条件是 \mathcal{P} 是对称变换且满足 $\mathcal{P}^2 = \mathcal{P}$.

(3) 设 $\mathcal{P}_1, \mathcal{P}_2$ 均为 V 上的投影变换, 求证: $\mathcal{P}_1 + \mathcal{P}_2$ 为投影变换当且仅当 $\mathcal{P}_1 \mathcal{P}_2 = \mathcal{O}$, 当且仅当 $\text{Im } \mathcal{P}_1 \perp \text{Im } \mathcal{P}_2$.

(4) 设 $\mathcal{P}_1, \mathcal{P}_2$ 均为 V 上的投影变换, 求证: $\mathcal{P}_1 - \mathcal{P}_2$ 为投影变换当且仅当 $\text{Im } \mathcal{P}_1 \supset \text{Im } \mathcal{P}_2$.

(5) 设 $\mathcal{P}_1, \mathcal{P}_2$ 均为 V 上的投影变换, 求证: $\mathcal{P}_1 \mathcal{P}_2$ 为投影变换当且仅当 $\mathcal{P}_1 \mathcal{P}_2 = \mathcal{P}_2 \mathcal{P}_1$, 且此时有 $\text{Im } \mathcal{P}_1 \mathcal{P}_2 = \text{Im } \mathcal{P}_1 \cap \text{Im } \mathcal{P}_2$.

(6) 设 \mathcal{A} 为 V 上的对称变换, 证明: \mathcal{A} 可表示为一组投影变换的线性组合, 且该组投影变换的像空间的直和恰为 V .

22. 设 \mathcal{A} 为 n 维复线性空间上的正规变换, 证明: $\mathcal{A} = \sum_{k=1}^n \lambda_k \mathcal{P}_k$, 这里 λ_k 为 \mathcal{A} 的特征值, \mathcal{P}_k 为投影变换 ($1 \leq k \leq n$). 若 \mathcal{A} 为 n 维实线性空间上的对称变换, 结论是否仍然成立?
23. 设 A 为 $m \times n$ 的非零矩阵, 证明:

(1) 若 A 为实矩阵, 则存在正交矩阵 U, V , 使得 $A = UDV$, 这里

$$D = \begin{pmatrix} \lambda_1 & & & 0 \\ & \ddots & & \\ & & \lambda_r & \\ 0 & & & 0 \end{pmatrix},$$

其中 $\lambda_1 > \lambda_2 > \cdots > \lambda_r > 0$ 称为 A 的奇异值, $r = \text{rank } A$;

(2) 若 A 为复矩阵, 则存在酉矩阵 U, V , 使得 $A = UDV$, 这里 D 同 (1) 中定义.

24. 设 A 为 n 阶实对称矩阵, $b \in \mathbf{R}^n, c \in \mathbf{R}, X = (x_1 \cdots x_n)^T, Y = (X^T \ 1)^T$, 记

$$f(X) = Y^T \begin{pmatrix} A & b \\ b^T & c \end{pmatrix} Y,$$

考虑方程 $f(X) = 0$.

- (1) 当 $n = 2$ 时, 方程 $f(X) = 0$ 称为一个二次曲线, 试给出所有可能的二次曲线.
- (2) 当 $n = 3$ 时, 方程 $f(X) = 0$ 称为一个二次曲面, 试给出所有可能的二次曲面.
25. 设 B 为方阵.
- (1) 若 $\det B = 1$, 是否存在方阵 A , 使得 $B = e^A$? 如果存在, 是否必有 $\text{tr } A = 0$? 这样的 A 是否唯一?
- (2) 若 B 为酉矩阵, 是否存在方阵 A , 使得 $B = e^A$? 如果存在, 是否必有 A 为反 Hermite 矩阵? 这样的 A 是否唯一?

第 7 章 矩阵分解

矩阵分解对矩阵理论及近代计算数学的发展起了关键作用. 所谓矩阵分解, 就是将一个矩阵写成结构比较简单的或者性质比较熟悉的另一些矩阵的乘积. 在前面的章节中, 实质上已经提及了一些矩阵分解, 例如

- (1) 方阵 A 相似于 Jordan 标准形 J , 即存在可逆阵 P , 使得 $A = PJP^{-1}$;
- (2) 正规矩阵 A 酉相似于对角阵 Λ , 即存在酉阵 U , 使得 $A = U\Lambda U^H$;
- (3) 方阵 A 酉相似于上三角阵 R , 即存在酉阵 U , 使得 $A = URU^H$.

本章将依此介绍矩阵的 LU 三角分解、QR 分解、满秩分解、谱分解以及奇异值分解. 所有这些分解都在数值代数和最优化问题中扮演着十分重要的角色.

7.1 矩阵的三角分解

矩阵中有一类矩阵比较特殊, 称之为三角矩阵, 其中对角线以下全为零的称为上三角矩阵, 对角线以上全为零的称为下三角矩阵. 特别, 对角线元素均为 1 的上(下)三角阵称为单位上(下)三角矩阵, 除主对角线外均为零的称为对角矩阵. 上三角矩阵和下三角矩阵统称三角矩阵.

定义 7.1 给定矩阵 $A \in \mathbf{R}^{n \times n}$, 若存在下三角矩阵 $L \in \mathbf{R}^{n \times n}$ 和上三角矩阵 $U \in \mathbf{R}^{n \times n}$ 使得 $A = LU$, 这种分解称为矩阵的三角分解或 LU 分解.

LU 分解与 Gauss¹消去法解线性方程组密切相关. 首先回顾这一众所周知的过程. 考虑 n 元线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n. \end{cases}$$

或记为 $Ax = b$, 其中 $A = (a_{ij})_{1 \leq i, j \leq n}$, $x = (x_1 \ x_2 \ \cdots \ x_n)^T$, $b = (b_1 \ b_2 \ \cdots \ b_n)^T$. Gauss 消去法的基本思想是利用矩阵的初等行变换化系数矩阵 A 为上三角矩阵, 再用回代法求解上三角线性方程组, 具体步骤如下.

¹Gauss (1777–1855), 高斯, 德国数学家.

先记矩阵 $\mathbf{A}^{(1)} = \mathbf{A}$ ，向量 $\mathbf{b}^{(1)} = \mathbf{b}$ 。第1步，如果 $a_{11}^{(1)} \neq 0$ ，可对 $i = 2, 3, \dots, n$ 作如下的运算，用数 $-a_{i1}^{(1)}/a_{11}^{(1)}$ 依次乘以方程组的第一行，并加到第 i 行上去，即可将 $\mathbf{A}^{(1)}$ 第1列上，从第2到第 n 行的元素化为零，即

$$\begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(2)} \\ \vdots \\ b_n^{(2)} \end{pmatrix}.$$

此时，新元素 $a_{ij}^{(2)}$ 的计算公式为

$$a_{ij}^{(2)} = a_{ij}^{(1)} - l_{i1}a_{1j}^{(1)}, \quad i, j = 2, 3, \dots, n,$$

其中 $l_{i1} = \frac{a_{i1}^{(1)}}{a_{11}^{(1)}}$ 。若记

$$\mathbf{L}_1 = \begin{pmatrix} 1 & & & \\ -l_{21} & 1 & & \\ \vdots & & \ddots & \\ -l_{n1} & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & & & \\ \frac{a_{21}^{(1)}}{a_{11}^{(1)}} & 1 & & \\ \vdots & & \ddots & \\ \frac{a_{n1}^{(1)}}{a_{11}^{(1)}} & & & 1 \end{pmatrix},$$

于是有

$$\mathbf{L}_1 \mathbf{A}^{(1)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix} = \mathbf{A}^{(2)}$$

如此继续下去，经过 $n-1$ 步后，则可将矩阵 \mathbf{A} 变为上三角矩阵

$$\mathbf{L}_{n-1} \cdots \mathbf{L}_2 \mathbf{L}_1 \mathbf{A}^{(1)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{nn}^{(n)} \end{pmatrix} = \mathbf{A}^{(n)},$$

其中

$$L_i = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & -l_{i+1\ i} & 1 & \\ & & \vdots & & \ddots \\ & & -l_{n\ i} & & 1 \end{pmatrix}, \quad i = 1, 2, \dots, n-1.$$

上述对 \mathbf{A} 的元素进行的消元过程便是 Gauss 消元的全过程. 显然, Gauss 消元过程能够进行到底当且仅当每一步的主元素 $a_{11}^{(1)}, a_{22}^{(2)}, \dots, a_{n-1\ n-1}^{(n-1)}$ 都不为零. 怎样判断 \mathbf{A} 的前 $n-1$ 个主元素是否为零呢?

由于每一步初等变换不改变行列式的值, 矩阵 \mathbf{A} 的前 $n-1$ 个顺序主子式应该满足下列关系

$$\begin{aligned} a_{11} &= a_{11}^{(1)} \\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} &= \begin{vmatrix} a_{11}^{(1)} & a_{12}^{(1)} \\ 0 & a_{22}^{(2)} \end{vmatrix} = a_{11}^{(1)} a_{22}^{(2)} \\ &\dots\dots\dots \\ \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1\ n-1} \\ a_{21} & a_{22} & & a_{2\ n-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n-1\ 1} & a_{n-1\ 2} & \dots & a_{n-1\ n-1} \end{vmatrix} &= \begin{vmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & & a_{2n}^{(2)} \\ & & \ddots & \vdots \\ & & & a_{n-1\ n-1}^{(n-1)} \end{vmatrix} = a_{11}^{(1)} a_{22}^{(2)} \dots a_{n-1\ n-1}^{(n-1)}, \end{aligned}$$

即矩阵 \mathbf{A} 的前 $n-1$ 个顺序主子式都不为零的情况下, 有每一步的主元素 $a_{11}^{(1)}, a_{22}^{(2)}, \dots, a_{n-1\ n-1}^{(n-1)}$ 都不为零.

综上所述, 可得到如下定理.

定理 7.1 给定矩阵 $\mathbf{A} \in \mathbf{R}^{n \times n}$, Gauss 消元过程能够进行到底的充分必要条件是 \mathbf{A} 的各阶顺序主子阵均不为零.

通过简单的计算, 可知

$$L_i^{-1} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & l_{i+1\ i} & 1 & \\ & & \vdots & & \ddots \\ & & l_{n\ i} & & 1 \end{pmatrix}, \quad i = 1, 2, \dots, n-1.$$

从而在 Gauss 消去过程中有

$$\mathbf{A} = L_1^{-1} L_2^{-1} \dots L_{n-1}^{-1} \mathbf{A}^{(n)}.$$

记 $L = L_1^{-1}L_2^{-1}\cdots L_{n-1}^{-1}$ 是单位下三角矩阵, $U = A^{(n)}$ 是一个上三角矩阵, 则有 $A = LU$. 这里 L 是单位下三角矩阵, U 是上三角矩阵, 这种矩阵分解称为 **Doolittle 分解**.

由 Gauss 消去法过程可知如下矩阵三角分解的存在定理.

定理 7.2 给定矩阵 $A \in \mathbf{R}^{n \times n}$, 存在单位下三角阵 $L \in \mathbf{R}^{n \times n}$ 和可逆上三角阵 $U \in \mathbf{R}^{n \times n}$, 使得 $A = LU$ 的充分必要条件是 A 的各阶顺序主子阵均不为零.

例 7.1 已知 $A = \begin{pmatrix} 2 & 3 & 4 \\ -2 & 0 & 2 \\ 2 & -3 & -3 \end{pmatrix}$, 计算矩阵 A 的 Doolittle 分解.

解 对矩阵 A 依次用 $r_i + kr_j$ 类型初等变换消去下三角非零元素, 其中 $j < i$. 于是, 对于第 1 列

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad L_1 A = \begin{pmatrix} 2 & 3 & 4 \\ 0 & 3 & 6 \\ 0 & -6 & -7 \end{pmatrix},$$

对于第 2 列

$$L_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}, \quad L_2 L_1 A = \begin{pmatrix} 2 & 3 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{pmatrix},$$

故 $A = LU$, 其中

$$L = L_1^{-1}L_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 2 & 3 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{pmatrix}.$$

矩阵的三角分解未必存在, 譬如 $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 就不存在三角分解. 矩阵的三角分解一般来说也不唯一. 譬如在上面的例子中, 有

$$\begin{aligned} \begin{pmatrix} 2 & 3 & 4 \\ -2 & 0 & 2 \\ 2 & -3 & -3 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 3 & 6 \\ 0 & 0 & 5 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 3/2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ -2 & 3 & 0 \\ 2 & -6 & 5 \end{pmatrix} \begin{pmatrix} 1 & 3/2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

矩阵的三角分解除除了 Doolittle 分解, 还有以下两种变形:

(1) **Crout 分解**: $A = LU$, 这里 L 是下三角矩阵, U 是单位上三角矩阵;

(2) **LDU 分解**: $A = LDU$, 这里 L 是单位下三角矩阵, D 是对角矩阵, U 是单位上三角矩阵.

Doolittle 分解、Crout 分解、LDU 分解统称为三角分解或 LU 分解. 如果不作特殊说明, 一般所说的 LU 分解就是指 Doolittle 分解.

矩阵的三角分解具有唯一性, 有下述定理.

定理 7.3 若 A 为 n 阶矩阵, 且所有顺序主子式均不等于零, 则 A 可唯一地分解为一个单位下三角矩阵 L 与一个上三角矩阵 U 的乘积.

证明 Doolittle 分解的存在性上面已经给出了, 下面来证明它的唯一性. 不妨假设矩阵 A 有两种 Doolittle 分解

$$A = L_1 U_1 = L_2 U_2, \quad (7.1)$$

其中, L_1 和 L_2 为单位下三角矩阵, U_1 和 U_2 为上三角矩阵. 由于 A 非奇异, 则 U_1 和 U_2 也是非奇异矩阵, 于是由式 (7.1) 可得:

$$L_2^{-1} L_1 = U_2 U_1^{-1}. \quad (7.2)$$

由于单位下三角矩阵的逆矩阵仍是单位下三角矩阵, 单位下三角矩阵与单位下三角矩阵的乘积仍是单位下三角矩阵, 且上三角矩阵的逆矩阵仍是上三角矩阵, 上三角矩阵与上三角矩阵的乘积仍是上三角矩阵. 这样, 式 (7.2) 的左边为单位下三角矩阵, 而右边为上三角矩阵, 所以必有

$$L_2^{-1} L_1 = U_2 U_1^{-1} = E.$$

即 $L_1 = L_2$, $U_1 = U_2$, 唯一性得证. 证毕.

同理, 存在性和唯一性定理也适用于另外两种三角分解的变形.

定理 7.4 如果矩阵 A 的所有顺序主子式均不等于零, 则有

(1) A 有唯一的 LDU 分解: $A = LDU$.

(2) A 有唯一的 Crout 分解: $A = LU$.

除了用初等变换法, 矩阵的三角分解也可以通过待定系数法来完成计算.

例 7.2 已知矩阵 $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \\ 1 & 16 & 81 & 256 \end{pmatrix}$, 计算 A 的 Doolittle 分解.

解 假设

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \\ 1 & 16 & 81 & 256 \end{pmatrix} = \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ l_{31} & l_{32} & 1 & \\ l_{41} & l_{42} & l_{43} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ & u_{22} & u_{23} & u_{24} \\ & & u_{33} & u_{34} \\ & & & u_{44} \end{pmatrix}.$$

根据待定系数依次计算, 可得

$$L = \begin{pmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 3 & 1 & \\ 1 & 7 & 6 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 2 & 3 & 4 \\ & 2 & 6 & 12 \\ & & 6 & 24 \\ & & & 24 \end{pmatrix}.$$

当 A 为对称正定矩阵时, 它的所有顺序主子式都大于零, 故由定理 7.4 可知存在唯一的 LDU 分解. 由于对称正定的特殊性, 还可以得到一个性质更好的三角分解.

定理 7.5 若 $A \in \mathbf{R}^{n \times n}$ 为对称正定矩阵, 则存在唯一的对角元素均为正的下三角矩阵 G , 使得 $A = GG^T$, 这样的分解称为对称正定矩阵的 **Cholesky 分解**.

证明 根据定理 7.4, 由于 A 是对称正定矩阵, 所以存在唯一的 LDU 分解, 即 $A = LDU$, 其中, L 是单位下三角矩阵, D 是非奇异的对角矩阵, U 是单位上三角矩阵.

由 A 的对称性可得 $LDU = U^T D L^T$, 按照分解的唯一性可得 $L = U^T$, 从而得到 $A = LDL^T$.

设 $D = \text{diag}(d_1, d_2, \dots, d_n)$, $d_i \neq 0$, $i = 1, 2, \dots, n$. 下面进一步来证明 D 的对角元素均为正数, 即 $d_i > 0$.

由于 L 是单位下三角矩阵, 所以 L^T 是单位上三角矩阵, 当然也是非奇异矩阵. 故对于单位坐标向量 $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T$, 存在非零向量 x_i , 使得

$$L^T x_i = e_i \quad i = 1, 2, \dots, n.$$

另外, $x_i^T A x_i = x_i^T (LDL^T) x_i = (L^T x_i)^T D (L^T x_i) = e_i^T D e_i = d_i$. 由于 A 是对称正定矩阵, 则有 $x_i^T A x_i > 0$, 从而 $d_i > 0$, $i = 1, 2, \dots, n$. 这就证明了 D 的对角元素都为正数.

记 $D^{1/2} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n})$, 则有

$$A = LDL^T = LD^{1/2} D^{1/2} L^T = (LD^{1/2})(LD^{1/2})^T.$$

如果记 $G = LD^{1/2}$, 其中, L 是对角元素均大于零的下三角矩阵, 则有

$$A = GG^T.$$

容易证明, 这个三角分解也是唯一的. 证毕.

例 7.3 计算矩阵 $A = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 2 & -3 \\ -2 & -3 & 14 \end{pmatrix}$ 的 Cholesky 分解.

解 利用待定系数法. 设 A 的 Cholesky 矩阵分解 $A = GG^T$, 其中

$$G = \begin{pmatrix} g_{11} & 0 & 0 \\ g_{21} & g_{22} & 0 \\ g_{31} & g_{32} & g_{33} \end{pmatrix}.$$

根据计算,

$$g_{11} = \sqrt{4} = 2, \quad g_{21} = 2/2 = 1, \quad g_{31} = -2/2 = 1,$$

$$g_{22} = \sqrt{2-1} = 1, \quad g_{32} = -2,$$

$$g_{33} = \sqrt{14-1-4} = 3.$$

由此可得

$$\mathbf{G} = \begin{pmatrix} 2 & & \\ 1 & 1 & \\ -1 & -2 & 3 \end{pmatrix}.$$

7.2 正交三角分解

在介绍矩阵的正交三角分解之前，首先回顾一下几类特殊的正交矩阵。

1. 置换矩阵

将单位矩阵的任意两行（列）交换得到的矩阵，称为置换矩阵。譬如，将单位矩阵的第 i 行和第 j 行交换，得到置换矩阵

$$\mathbf{P}_{ij} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & \\ & & & \ddots & & \\ & & 1 & & 0 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ i \\ \\ j \\ \\ \end{matrix}.$$

易见，任意个置换矩阵的乘积仍然是置换矩阵。

2. 旋转矩阵（Givens 变换）

对 2 为情形，对于某个角度 θ ，

$$\mathbf{G} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

是一个正交阵。记 $\mathbf{w} = (x, y)^T = (r \cos \phi, r \sin \phi)^T$ 为二维平面中的一个向量，则

$$\mathbf{G}\mathbf{w} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix} = \begin{pmatrix} r \cos(\theta + \phi) \\ r \sin(\theta + \phi) \end{pmatrix},$$

即 $\mathbf{G}\mathbf{w}$ 表示将向量 \mathbf{w} 逆时针旋转 θ 角所得到的向量（如图 7.1 所示）。

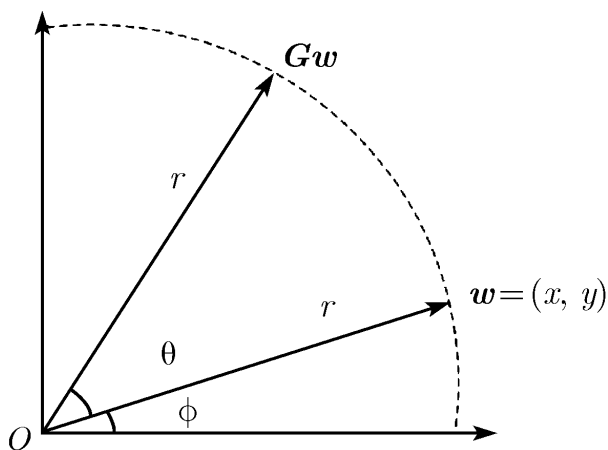


图 7.1: Givens 变换

推广到 $n \times n$ 的情形, 形如

$$G(i, j, \theta) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \cos \theta & -\sin \theta & \\ & & \sin \theta & \cos \theta & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ i \\ j \\ \\ \end{matrix}$$

的矩阵称为 **Givens 矩阵**或**旋转矩阵**, 相应的线性变换称为 **Givens 变换**或**旋转变换**, 其中 θ 为旋转的角度. 显然, $G(i, j, \theta)$ 也是正交矩阵.

若 $x \in \mathbf{R}^n$, $y = G(i, j, \theta)x$, 则 y 的分量为

$$\begin{cases} y_i = cx_i + sx_j \\ y_j = -sx_i + cx_j \\ y_k = x_k \end{cases} \quad k \neq i, k \neq j.$$

如果要使 $y_j = 0$, 只要选择 θ 满足

$$c = \cos \theta = \frac{x_i}{\sqrt{(x_i^2 + x_j^2)}}, \quad s = \sin \theta = \frac{x_j}{\sqrt{(x_i^2 + x_j^2)}}$$

即可.

例 7.4 用 Givens 变换将上 Hessenberg 型矩阵

$$A = \begin{pmatrix} 4.8 & 2.56 & 2.528 \\ 3.6 & 4.92 & 3.296 \\ 0 & 1.8 & 1.84 \\ 0 & 0 & 0.6 \end{pmatrix}$$

化为上三角矩阵.

解 首先, 为了消去 A 中 $(2, 1)$ 元, 构造 Givens 变换 $G(1, 2, \theta)$, 其中

$$\cos \theta = \frac{4.8}{\sqrt{(4.8^2 + 3.6^2)}} = 0.8 \quad \sin \theta = \frac{3.6}{\sqrt{(4.8^2 + 3.6^2)}} = 0.6.$$

从而

$$A_1 = G(1, 2, \theta)A = \begin{pmatrix} 6 & 5 & 4 \\ 0 & 2.4 & 1.12 \\ 0 & 1.8 & 1.84 \\ 0 & 0 & 0.6 \end{pmatrix}.$$

其次, 消去 A_1 中 $(3, 2)$ 元. 为此, 构造 Givens 变换 $G(2, 3, \theta)$, 其中 $\cos \theta = 0.8$, $\sin \theta = 0.6$. 从而

$$A_2 = G(2, 3, \theta)A_1 = \begin{pmatrix} 6 & 5 & 4 \\ 0 & 3 & 2 \\ 0 & 0 & 0.8 \\ 0 & 0 & 0.6 \end{pmatrix}.$$

最后, 消去 A_2 中 $(4, 3)$ 元. 为此, 构造 Givens 变换 $G(3, 4, \theta)$, 其中 $\cos \theta = 0.8$, $\sin \theta = 0.6$. 从而上三角矩阵 R 为

$$R = G(3, 4, \theta)A_2 = \begin{pmatrix} 6 & 5 & 4 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

3. 反射矩阵 (Householder 变换)

设 $w \in \mathbf{R}^n$, 且 $\|w\| = 1$, 则

$$P = E - 2ww^T$$

称为 **Householder** 矩阵, 相应的线性变换称为 **Householder** 变换. Householder 矩阵有如下性质:

(1) $P^T = P$, 即 P 是实对称矩阵;

(2) P 是正交矩阵 (定理 6.43).

(3) 如图 7.2, 设 w 是 \mathbf{R}^3 上的一个单位向量, 并设 S 为过原点且与 w 垂直的平面, 则一切 $v \in \mathbf{R}^3$ 可分解成 $v = v_1 + v_2$, 其中 $v_1 \in S$, $v_2 \perp S$. 不难验证 $Pv_1 = v_1$, $Pv_2 = -v_2$, 所以

$$Pv = v_1 - v_2.$$

这样, v 经变换后的象 Pv 是 v 关于 S 对称的向量. 所以, Householder 变换又称镜面反射变

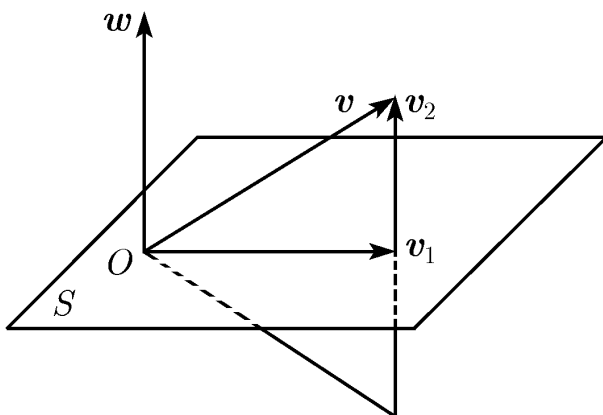


图 7.2: Householder 变换

换, Householder 矩阵也称初等反射矩阵.

一个重要的应用是对 $\mathbf{x} \neq 0$, 求 Householder 矩阵 \mathbf{P} , 使得

$$\mathbf{P}\mathbf{x} = k\mathbf{e}_1$$

其中, $\mathbf{e}_1 = (1 \ 0 \ \cdots \ 0)^T$. 由正交矩阵的性质可知 $\|\mathbf{P}\mathbf{x}\| = \|k\mathbf{e}_1\| = \|\mathbf{x}\|$, 即 $k = \pm\|\mathbf{x}\|$. 由上面所讨论的 \mathbf{P} 的构造, 有

$$\mathbf{u} = \mathbf{x} - k\mathbf{e}_1, \quad \mathbf{w} = \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

设 $\mathbf{x} = (x_1 \ \cdots \ x_n)^T$, 为了使 $\mathbf{x} - k\mathbf{e}_1$ 计算时不损失有效数位, 取

$$k = -\text{sign}(x_1)\|\mathbf{x}\|, \quad \text{sgn}(x_1) = \begin{cases} 1 & \text{当 } x_1 \geq 0, \\ -1 & \text{当 } x_1 < 0, \end{cases}$$

则

$$\mathbf{u} = (x_1 + \text{sgn}(x_1)\|\mathbf{x}\|, x_2, \cdots, x_n)^T.$$

从而

$$\mathbf{P} = \mathbf{E} - \beta\mathbf{u}\mathbf{u}^T,$$

其中

$$\beta = 2(\|\mathbf{u}\|^2)^{-1} = 2(\|\mathbf{x}\|(\|\mathbf{x}\| + |x_1|))^{-1}.$$

例 7.5 已知 $\mathbf{x} = (3 \ 5 \ 1 \ 1)^T$, 求 Householder 矩阵 \mathbf{P} , 使得 $\mathbf{P}\mathbf{x} = -6\mathbf{e}_1$, 其中 $\|\mathbf{x}\| = 6$.

解 取 $k = -6$, $\mathbf{u} = \mathbf{x} - k\mathbf{e}_1 = (9 \ 5 \ 1 \ 1)^T$, $\|\mathbf{u}\|^2 = 108$, $\beta = \frac{1}{54}$, 则

$$\mathbf{P} = \mathbf{E} - \beta\mathbf{u}\mathbf{u}^T = \frac{1}{54} \begin{pmatrix} -27 & -45 & -9 & -9 \\ -45 & 29 & -5 & -5 \\ -9 & -5 & 53 & -1 \\ -9 & -5 & -1 & 53 \end{pmatrix}.$$

定义 7.2 给定矩阵 $\mathbf{A} \in \mathbf{R}^{n \times n}$, 若存在正交矩阵 $\mathbf{Q} \in \mathbf{O}(n)$ 和上三角矩阵 $\mathbf{R} \in \mathbf{R}^{n \times n}$ 使得 $\mathbf{A} = \mathbf{Q}\mathbf{R}$, 这种分解称为矩阵的一个正交三角分解或 QR 分解.

下面给出正交三角分解的存在唯一性定理.

定理 7.6 设 $A \in \mathbf{R}^{n \times n}$, 则存在正交阵 Q , 使得 $A = QR$, 其中 R 为上三角阵.

证明 给出构造性证明. 首先, 考虑 A 的第一列 $a_1 = (a_{11} \ a_{21} \ \cdots \ a_{n1})^T$, 可找到 Householder 矩阵 P_1 , 使得 $P_1 a_1$ 的元素除了第 1 个以外都为零.

同理, 找到 P_2 使得 $P_2 P_1 A$ 的第 2 列对角元以下元素为 0, 而第一列对角元以下元素与 $P_1 A$ 一样是 0. 依次这样下去, 可以得到

$$P_{n-1} P_{n-2} \cdots P_1 A = R,$$

其中 R 为上三角矩阵, $Q^T = P_{n-1} P_{n-2} \cdots P_1$ 为正交阵. 证毕.

该定理保证了 A 可分解为 $A = QR$, 若 A 非奇异, 则 R 也非奇异. 如果不规定 R 的对角元为正, 则分解不是唯一的.

定理 7.7 设 $A \in \mathbf{R}^{n \times n}$, 且 A 非奇异, 则存在正交阵 Q 与上三角阵 R , 使得

$$A = QR,$$

且当 R 的对角元均为正时, 分解是唯一的.

另外, 注意到

$$A^T A = (QR)^T QR = R^T Q^T QR = R^T R,$$

所以, 矩阵 A 正交三角分解中的 R 恰好是矩阵 $A^T A$ 的 Cholesky 分解中的上三角矩阵.

注 7.1 除了用 Householder 变换和 Givens 变换, 还可以用 Schmidt 正交化计算矩阵 A 的正交三角分解 (见推论 6.6).

例 7.6 用 Schmidt 正交化过程计算矩阵 A 的正交三角分解

$$A = \begin{pmatrix} 2 & -2 & -1 \\ 2 & 7 & 2 \\ 1 & 8 & 7 \end{pmatrix}.$$

解 令矩阵 $A = (\alpha_1 \ \alpha_2 \ \alpha_3)$, 用 Schmidt 正交化过程寻找正交矩阵 $Q = (q_1 \ q_2 \ q_3)$ 和一个上三角矩阵 R , 使得

$$(\alpha_1 \ \alpha_2 \ \alpha_3) = (q_1 \ q_2 \ q_3) \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{pmatrix},$$

即

$$\alpha_1 = r_{11} q_1, \quad (7.3)$$

$$\alpha_2 = r_{12} q_1 + r_{22} q_2, \quad (7.4)$$

$$\alpha_3 = r_{13} q_1 + r_{23} q_2 + r_{33} q_3. \quad (7.5)$$

经计算, 由式 (7.3) 得

$$r_{11} = \|\alpha_1\| = \sqrt{2^2 + 2^2 + 1^2} = 3,$$

$$q_1 = \frac{\alpha_1}{r_{11}} = \frac{1}{3} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}.$$

由式 (7.4) 有

$$r_{12} = (\alpha_2, q_1) = 6,$$

$$\bar{q}_2 = \alpha_2 - r_{12}q_1 = \begin{pmatrix} -2 \\ 7 \\ 8 \end{pmatrix} - 6 \cdot \frac{1}{3} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -6 \\ 3 \\ 6 \end{pmatrix},$$

$$r_{22} = \|\bar{q}_2\| = \sqrt{(-6)^2 + 3^2 + 6^2} = 9,$$

$$q_2 = \frac{\bar{q}_2}{r_{22}} = \frac{1}{3} \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix}.$$

由式 (7.5) 有

$$r_{13} = (\alpha_3, q_1) = 3,$$

$$r_{23} = (\alpha_3, q_2) = 6,$$

$$\bar{q}_3 = \alpha_3 - r_{13}q_1 - r_{23}q_2 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix},$$

$$r_{33} = \|\bar{q}_3\| = \sqrt{1^2 + (-2)^2 + 2^2} = 3,$$

$$q_3 = \frac{\bar{q}_3}{r_{33}} = \frac{1}{3} \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}.$$

因此

$$Q = \frac{1}{3} \begin{pmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{pmatrix}, \quad R = \begin{pmatrix} 3 & 6 & 3 \\ 0 & 9 & 6 \\ 0 & 0 & 3 \end{pmatrix}.$$

QR 分解是计算特征值的有力工具, 也可用于其他矩阵计算问题, 包括解方程组 $Ax = b$. 这只要令 $y = Q^T b$, 再解上三角方程组 $Rx = y$. 这个计算过程是稳定的, 也不必选主元, 但是计算量比 Gauss 消元法将近大一倍.

7.3 满秩分解

定义 7.3 设 $A \in \mathbf{R}^{m \times n}$. 当 $\text{rank } A = m$ 时, 称 A 为行满秩矩阵; 若 $\text{rank } A = n$ 时, 称 A

为列满秩矩阵.

如果将矩阵看作行(列)向量组, A 为行(列)满秩矩阵, 当且仅当矩阵 A 的行(列)向量组线性无关.

定理 7.8 设 $A \in \mathbf{R}^{m \times n}$ 且 $\text{rank } A = r \leq \min\{m, n\}$, 则可将 A 作满秩分解

$$A = CD,$$

其中 $C \in \mathbf{R}^{m \times r}$, $D \in \mathbf{R}^{r \times n}$, 且 $\text{rank } C = \text{rank } D = r$.

证明 因为 $\text{rank } A = r$, 所以存在 m 阶可逆阵 P 和 n 阶置换阵 Q , 使得

$$A = P \begin{pmatrix} E_r & B \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^T.$$

令 $P = (P_1 \ P_2)$, 其中 P_1 是 $m \times r$ 列满秩阵, 这样

$$A = (P_1 \ P_2) \begin{pmatrix} E_r & B \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q^T = (P_1 \ P_1 B) Q^T = P_1 (E_r \ B) Q^T,$$

显然, $(E_r \ B) Q^T$ 是行满秩的 $r \times n$ 阵. 令 $C = P_1$, $D = (E_r \ B) Q^T$, 即得所证.

设 $C = (c_1 \ c_2 \ \cdots \ c_r)$, $D = (d_1 \ d_2 \ \cdots \ d_r)^T$, 则

$$A = (c_1 \ c_2 \ \cdots \ c_r) \begin{pmatrix} d_1^T \\ d_2^T \\ \vdots \\ d_r^T \end{pmatrix} = \sum_{i=1}^r c_i d_i^T,$$

这也是 A 的满秩分解的一种表示形式.

下面是求矩阵 A 的满秩分解的例子.

例 7.7 设 $A = \begin{pmatrix} 2 & 1 & 6 & 1 & 0 \\ 3 & 2 & 10 & 1 & 0 \\ 2 & 3 & 10 & -1 & 3 \\ 4 & 4 & 16 & 0 & 1 \end{pmatrix}$, 求矩阵 A 的满秩分解.

解 设 $A = (\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4 \ \alpha_5)$, 用行初等变换把矩阵 A 化为行最简形矩阵

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

该矩阵是 3×5 行满秩阵, 记 $B = (\beta_1 \ \beta_2 \ \beta_3 \ \beta_4 \ \beta_5)$. 显然 $\beta_1, \beta_2, \beta_5$ 线性无关, 且 $\beta_3 = 2\beta_1 + 2\beta_2, \beta_4 = \beta_1 - \beta_2$. 由于行初等变换保持矩阵列向量组的线性组合关系, 因此 $\alpha_1, \alpha_2, \alpha_5$

线性无关, 且 $\alpha_3 = 2\alpha_1 + 2\alpha_2$, $\alpha_4 = \alpha_1 - \alpha_2$. 取

$$C = (\alpha_1 \ \alpha_2 \ \alpha_5) = \begin{pmatrix} 2 & 1 & 0 \\ 3 & 2 & 0 \\ 2 & 3 & 3 \\ 4 & 4 & 1 \end{pmatrix},$$

显然 C 是 4×3 列满秩阵, 且满足

$$CB = (\alpha_1 \ \alpha_2 \ 2\alpha_1 + 2\alpha_2 \ \alpha_1 - \alpha_2 \ \alpha_5) = (\alpha_1 \ \alpha_2 \ \alpha_5 \ \alpha_4 \ \alpha_5) = A.$$

注 7.2 一般地, 矩阵 A 的行最简形矩阵可能存在全零行, 则行满秩矩阵 D 须取为行最简形 B 去除全零行后得到的矩阵, 列满秩矩阵 C 仍取为矩阵 A 的对应 C 的各行非零首元所在列.

进一步, 还有如下的定理.

定理 7.9 (正交满秩分解定理) 设 A 是 $m \times n$ 阶实矩阵, A 的秩为 r , 则存在 $m \times r$ 列正交矩阵 Q 和行满秩的 $r \times n$ 阵 R , 使 $A = QR$. 其中, Q 列正交的含意为 $Q^T Q = E_r$.

证明 由定理 7.8 知存在列满秩的 $m \times r$ 阵 C 和行满秩的 $r \times n$ 阵 D , 使 $A = CD$. 于是 $C^T C$ 是秩为 r 的 r 阶对称方阵, 且易证 $C^T C$ 是正定阵. 这样存在 r 阶对称正定阵 S , 使 $C^T C = S^2$, 且 $(CS^{-1})^T (CS^{-1}) = S^{-1} C^T C S^{-1} = S^{-1} S^2 S^{-1} = E_r$. 记 $Q = CS^{-1}$, $R = SD$, 则 Q 是 $m \times r$ 列正交阵且 R 是行满秩的 $r \times n$ 阵, 于是有

$$QR = CS^{-1}SD = CD = A.$$

证毕.

注 7.3 矩阵的满秩分解的几何意义是, 将 A 作为列向量组, 用一个线性无关的列向量组 C 分别线性表示每一列, 而 D 的每一列恰是 A 的相应列在线性无关列向量组 C 下的坐标. 按照这一意义, C 可取为 A 的极大线性无关组, 这正是例 7.7 的做法. 若进一步将 C 的列做 Schmidt 正交化, 就得到一个列正交矩阵 Q , 进而得到正交满秩分解. 由此还可以得到满秩分解不是唯一的. 事实上, 任意将 C 的某列乘以一个非零常数, 或任意交换 C 的列的次序, 都可找到相应的行满秩矩阵 D .

7.4 矩阵的谱分解

矩阵的谱分解是对可对角化的矩阵的一种分解.

定义 7.4 若矩阵 $A \in \mathbf{R}^{n \times n}$ 可相似对角化, 即存在可逆矩阵 $P \in \mathbf{R}^{n \times n}$ 和对角矩阵 $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \in \mathbf{R}^{n \times n}$ 使得 $A = P\Lambda P^{-1}$, 这种分解称为矩阵 A 的谱分解. 其中特征值 $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ 也称为矩阵 A 的谱.

设 $P = (\alpha_1 \ \alpha_2 \ \dots \ \alpha_n)$, $P^{-1} = (\beta_1 \ \beta_2 \ \dots \ \beta_n)^T$, 则 $\alpha_1, \alpha_2, \dots, \alpha_n$ 线性无关, $\beta_1, \beta_2, \dots, \beta_n$ 也线性无关, 且

$$A\alpha_k = \lambda_k \alpha_k, \quad A^T \beta_k = \lambda_k \beta_k \quad (1 \leq k \leq n),$$

于是

$$A = PAP^{-1} = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_n^T \end{pmatrix} = \sum_{k=1}^n \lambda_k \alpha_k \beta_k^T. \quad (7.6)$$

式 (7.6) 为矩阵 A 的谱分解的另一种表达形式. 如果记 $A_k = \alpha_k \beta_k^T$, 则式 (7.6) 可写为

$A = \sum_{k=1}^n \lambda_k A_k$, 其中 A_i 满足以下性质:

$$(1) A_k^2 = A_k \ (k = 1, 2, \cdots, n);$$

$$(2) A_j A_k = \mathbf{0} \ (j \neq k);$$

$$(3) \sum_{k=1}^n A_k = E.$$

事实上,

$$P^{-1}P = \begin{pmatrix} \beta_1^T \alpha_1 & \cdots & \beta_1^T \alpha_n \\ \vdots & & \vdots \\ \beta_n^T \alpha_1 & \cdots & \beta_n^T \alpha_n \end{pmatrix} = E, \quad PP^{-1} = \alpha_1 \beta_1^T + \cdots + \alpha_n \beta_n^T = E.$$

于是有

$$\beta_k^T \alpha_k = 1, \quad \beta_j^T \alpha_k = 0 \ (j \neq k), \quad \sum_{k=1}^n \alpha_k \beta_k^T = E.$$

再结合 $A_k = \alpha_k \beta_k^T$ 就得到上面三个性质.

注 7.4 在 6.4 节介绍过的正规矩阵必定可以相似对角化, 于是存在谱分解. 而作为正规矩阵的特例, 实对称矩阵、实反对称矩阵、正交矩阵、Hermite 矩阵、反 Hermite 矩阵、酉矩阵等, 也都存在谱分解. 进一步, 对正规矩阵, 过渡矩阵 P 可以选为酉矩阵, 由此得到的 A_k 还是 Hermite 矩阵, 此时的 A_k 称为投影矩阵. 从线性变换的观点看, A_k 恰为到特征子空间 $\text{span } \alpha_k$ 的正交投影变换的矩阵.

例 7.8 设 $A = \begin{pmatrix} 4 & -6 & 0 \\ 2 & -3 & 0 \\ -2 & 3 & 2 \end{pmatrix}$, 求 A 的谱分解.

解 先求 A 的特征值和特征向量.

$$|\lambda E - A| = \begin{vmatrix} \lambda - 4 & 6 & 0 \\ -2 & \lambda + 3 & 0 \\ 2 & -3 & \lambda - 2 \end{vmatrix} = \lambda(\lambda - 2)(\lambda - 1),$$

因此, A 有 3 个不同的特征值 $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 2$, 故 A 可对角化, 从而 A 的谱分解一

定存在. 容易求出它们对应的特征向量为

$$\mathbf{P}_1 = (3 \ 2 \ 0)^T, \quad \mathbf{P}_2 = (2 \ 1 \ 1)^T, \quad \mathbf{P}_3 = (0 \ 0 \ 1)^T.$$

令

$$\mathbf{P} = (\mathbf{P}_1 \ \mathbf{P}_2 \ \mathbf{P}_3) = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

显然 \mathbf{P} 可逆, 且易求得

$$\mathbf{P}^{-1} = \begin{pmatrix} \beta_1^T \\ \beta_2^T \\ \beta_3^T \end{pmatrix} = \begin{pmatrix} -1 & 2 & 0 \\ 2 & -3 & 0 \\ -2 & 3 & 1 \end{pmatrix}.$$

于是有

$$\mathbf{A} = \mathbf{P} \begin{pmatrix} 0 & & \\ & 1 & \\ & & 2 \end{pmatrix} \mathbf{P}^{-1} = (\mathbf{P}_1 \ \mathbf{P}_2 \ \mathbf{P}_3) \begin{pmatrix} 0 & & \\ & 1 & \\ & & 2 \end{pmatrix} \begin{pmatrix} \beta_1^T \\ \beta_2^T \\ \beta_3^T \end{pmatrix} = \mathbf{P}_2 \beta_2^T + 2 \mathbf{P}_3 \beta_3^T,$$

这就是 \mathbf{A} 的谱分解. 证毕.

注 7.5 矩阵的谱分解还可以进一步推广到 Hilbert 空间 (一类非常有代表性的内积空间) 的正规线性变换上, 量子力学中所提及谱, 正是这一推广的重要应用之一. 在无穷维的内积空间中, 正规线性变换的谱比特征值意义更加广泛, 这里不再展开.

7.5 奇异值分解

正规矩阵可以酉相似对角化, 因此其对应的线性变换具有优良的性质. 非正规矩阵, 甚至非方阵, 能否有类似地分解呢? 本节介绍的奇异值分解就是酉相似对角化对一般矩阵的推广.

定义 7.5 设 $\mathbf{A} \in \mathbf{R}^{m \times n}$, 半正定矩阵 $\mathbf{A}^T \mathbf{A}$ 的 n 个特征值记为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 因 $\lambda_i \geq 0$, 故可令 $\sigma_i = \sqrt{\lambda_i}$ ($i = 1, 2, \dots, n$), 称为矩阵 \mathbf{A} 的奇异值.

定理 7.10 (奇异值分解定理) 设矩阵 $\mathbf{A} \in \mathbf{R}^{m \times n}$ 的奇异值中有 r 个不等于零, 记为 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$. 它们构成的 r 阶对角阵记为 $\mathbf{D} = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_r\}$. 它们构成的 r 阶对角阵记为 $\mathbf{D} = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_r\}$. 令 $m \times n$ 阶矩阵 Σ 具有如下分块形式:

$$\Sigma = \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

则存在正交矩阵 $\mathbf{U} \in \mathbf{R}^{m \times m}$, $\mathbf{V} \in \mathbf{R}^{n \times n}$, 使 $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$, 称为 \mathbf{A} 的奇异值分解.

证明 因为 $A^T A$ 是 n 阶半正定矩阵, 必存在 n 阶正交矩阵 V 使

$$V^T(A^T A)V = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} = \begin{pmatrix} D^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

上式右端为 $n \times n$ 阶矩阵. 将 V 分块, 写成

$$V = (V_1 \ V_2),$$

其中 $V_1 \in \mathbf{R}^{n \times r}$, $V_2 \in \mathbf{R}^{n \times (n-r)}$. 因为 V 是正交阵, 所以 $V_1^T V_1 = E_r$, $V_1^T V_2 = \mathbf{0}$, 由

$$\begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} A^T A (V_1 \ V_2) = \begin{pmatrix} D^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

得

$$V_1^T A^T A V_1 = D^2,$$

$$V_2^T A^T A V_2 = \mathbf{0},$$

故 $AV_2 = \mathbf{0}$. 而

$$\begin{aligned} A &= AVV^T = A(V_1 \ V_2) \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = AV_1 V_1^T + AV_2 V_2^T = AV_1 V_1^T \\ &= AV_1 D^{-1} D V_1^T = U_1 D V_1^T, \end{aligned}$$

其中 $U_1 = AV_1 D^{-1}$, 且

$$U_1^T U_1 = D^{-1} V_1^T A^T A V_1 D^{-1} = D^{-1} D^2 D^{-1} = E_r.$$

将 U_1 扩张成正交矩阵 $U = (U_1 \ U_2)$, 则有

$$U \Sigma V^T = (U_1 \ U_2) \begin{pmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = (U_1 D \ \mathbf{0}) \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U_1 D V_1^T = A.$$

证毕.

定理 7.10 的证明同时给出了奇异值分解的计算方法.

例 7.9 求矩阵 $A = \begin{pmatrix} 1 & 1 \\ 1 & -2 \\ 2 & 1 \end{pmatrix}$ 的奇异值分解.

解 因为 $A^T A = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$ 的特征值为 $\lambda_1 = 7$, $\lambda_2 = 5$, 故 A 的奇异值为 $\sigma_1 = \sqrt{7}$, $\sigma_2 =$

$\sqrt{5}$. 又因为 $\mathbf{A}^T \mathbf{A}$ 的正交单位特征向量为

$$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

于是可令

$$\mathbf{D} = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \end{pmatrix}, \mathbf{V} = \mathbf{V}^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\mathbf{U}_1 = \mathbf{A} \mathbf{V}_1 \mathbf{D}^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{7}} & 0 \\ 0 & \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{14}} & 0 \\ -\frac{1}{\sqrt{14}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{10}} \end{pmatrix},$$

解线性方程组

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

得通解为

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = k \begin{pmatrix} 5 \\ 1 \\ 3 \end{pmatrix},$$

取 $k = \frac{1}{\sqrt{35}}$, 从而

$$\mathbf{U} = \begin{pmatrix} \frac{2}{\sqrt{14}} & 0 & \frac{5}{\sqrt{35}} \\ -\frac{1}{\sqrt{14}} & \frac{3}{\sqrt{10}} & \frac{3}{\sqrt{35}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{35}} \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}.$$

此时 $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ 即为 \mathbf{A} 的奇异值分解.

奇异值分解在统计学、信号处理、图像压缩和人工智能等实际工程中有着十分广泛和非常重要的应用. 有兴趣的读者可以进一步阅读相关文献和资料.

习题 7

1. 计算下列矩阵的 Doolittle 分解, Crout 三角分解和 LDU 三角分解.

$$(1) \begin{pmatrix} 2 & 4 & 6 \\ 2 & 7 & 12 \\ -2 & -10 & -13 \end{pmatrix}, \quad (2) \begin{pmatrix} 4 & 8 & 0 \\ 4 & 11 & 6 \\ -6 & -12 & 10 \end{pmatrix}.$$

2. 计算下列矩阵的 Cholesky 分解.

$$(1) \begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -3 \\ -1 & -3 & 6 \end{pmatrix}, \quad (2) \begin{pmatrix} 4 & 4 & -6 \\ 4 & 5 & -6 \\ -6 & -6 & 13 \end{pmatrix}.$$

3. 计算矩阵 $\begin{pmatrix} 2 & 4 & 6 & 8 \\ 4 & 12 & 20 & 16 \\ 3 & 10 & 20 & 18 \\ 1 & -4 & -9 & 14 \end{pmatrix}$ 的 Doolittle 分解.

4. 计算矩阵 $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 8 & 10 & 2 \\ 3 & 10 & 14 & 6 \\ 4 & 2 & 6 & 29 \end{pmatrix}$ 的 Cholesky 分解.

5. 计算下列矩阵的满秩分解.

$$(1) \begin{pmatrix} 1 & 2 & 3 & 3 \\ 4 & 5 & 9 & 6 \\ 7 & 8 & 15 & 9 \\ 2 & 5 & 7 & 8 \end{pmatrix}, \quad (2) \begin{pmatrix} 1 & 3 & -3 & 4 \\ 3 & 5 & -5 & 8 \\ 6 & -1 & 1 & 5 \\ 8 & -6 & 6 & 2 \end{pmatrix}.$$

6. 计算下列矩阵的谱分解.

$$(1) \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix}, \quad (2) \begin{pmatrix} 6 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 7 \end{pmatrix}.$$

7. 计算下列矩阵的 QR 分解.

$$(1) \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}, \quad (2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad (3) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (4) \begin{pmatrix} 2 & 2 & 1 \\ 0 & 2 & 2 \\ 2 & 1 & 2 \end{pmatrix}.$$

8. 计算下列矩阵的奇异值分解.

$$(1) \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}, \quad (2) \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

9. 证明：对任意实（复）非退化方阵 \mathbf{A} ，存在唯一的正交（酉）矩阵 \mathbf{Q} 和正定矩阵 \mathbf{H}_1 和 \mathbf{H}_2 ，使得 $\mathbf{A} = \mathbf{QH}_1 = \mathbf{H}_2\mathbf{Q}$ ，该分解称为矩阵的极分解。若去掉矩阵的非退化条件，结论将如何修正？
10. 证明：对任何正定矩阵 \mathbf{H} ，存在唯一的正定矩阵 \mathbf{S} ，使得 $\mathbf{H} = \mathbf{S}^2$ 。若将正定矩阵改为半正定矩阵，结论如何？

第 8 章 广义逆矩阵

在线性代数里，对于线性方程组 $Ax = b$ ，如果方阵 A 是非奇异的，则存在唯一的 A 的逆矩阵 A^{-1} ，满足等式

$$AA^{-1} = A^{-1}A = E,$$

从而该线性方程组具有唯一的解 $x = A^{-1}b$ 。

逆矩阵具有许多良好的性质和用途。遗憾的是，在工程和应用数学的许多领域中，遇到奇异方阵和非方阵。因此，对于一般的矩阵，也需要构造一种具有通常逆矩阵的若干性质的矩阵服务于实际应用，这就是本章将要介绍的广义逆（也称为伪逆）。

广义逆的出现，解决了一般的线性方程组的求解问题。广义逆可以很方便地表达若干矩阵方程和最小二乘问题的解。广义逆矩阵不仅在工程上得到广泛的应用，而且还应用于数理统计、多元分析、最优化理论、控制论、网络理论等众多学科，是一种重要的数学工具。

本章将着重介绍这类广义逆矩阵的概念、性质和计算方法，以及在求解矩阵方程和最小二乘问题中的应用。

8.1 概述

引进广义逆的动机，是为把逆矩阵推广到奇异方阵和非方阵。这一概念由 Moore 在 1920 年首先明确提出。

定义 8.1 给定矩阵 $A \in \mathbf{R}^{m \times n}$ ，若存在矩阵 $X \in \mathbf{R}^{n \times m}$ 满足

$$AX = P_{\mathcal{R}(A)} \quad XA = P_{\mathcal{R}(X)}$$

则称 X 为 A 的一个广义逆矩阵。这里 $P_{\mathcal{R}(A)}$ 表示在子空间 $\mathcal{R}(A)$ 上的正交投影矩阵（ $\mathcal{R}(A)$ 表示由矩阵 A 的作为列矩阵向量组生成的线性子空间，见例 5.19）。

Moore 凭借他天才的几何直觉，利用正交投影算子来定义广义逆，但由于这类定义较为抽象且不便进行有效运算，所以在之后的 30 年并未引起人们的注意。直到 1955 年，Penrose 以更直接明确的代数形式给出了 Moore 广义逆矩阵的定义。

定义 8.2 给定矩阵 $A \in \mathbf{R}^{m \times n}$ ，若存在矩阵 $X \in \mathbf{R}^{n \times m}$ 全部（或部分）满足下面四个方程的

(1) $AXA = A$;

$$(2) XAX = X;$$

$$(3) (AX)^T = AX;$$

$$(4) (XA)^T = XA,$$

则称 X 为 A 的一个广义逆矩阵.

Penrose 用 4 个方程 (又称为 **Penrose-Moore 方程组**) 重新定义了广义逆, 证明了这种广义逆的存在唯一性, 并建立了广义逆矩阵与线性方程组 $Ax = b$ 的解的联系. 自此, 广义逆的研究开始蓬勃发展, 成为矩阵理论研究的热点.

在实际应用中, 为了不同的目的, 可以定义不同意义的广义逆, 即也可研究满足 Penrose-Moore 方程中的部分方程的矩阵. 根据定义 8.2 中广义逆矩阵满足的条件不同, 广义逆矩阵共有 $C_4^1 + C_4^2 + C_4^3 + C_4^4 = 15$ 种不同的定义.

如果一个矩阵 $X \in \mathbf{R}^{n \times m}$ 满足四个方程中的第 i 和 j 个方程, 则称 X 为 $A\{i, j\}$ 的广义逆矩阵, 记为 $X \in A^{(i, j)}$. 譬如, 若 X 只满足第一个条件, 则 $X \in A\{1\}$, 此时称 X 为 A 的一个 $\{1\}$ 逆 (又称为 g 逆或者减号逆), 记作 A^- 或 $A^{(1)}$; 若 X 满足第一个和第二个条件, 则 $X \in A\{1, 2\}$, 此时称 X 为 A 的一个 $\{1, 2\}$ 逆 (又称为自反广义逆), 记作 $A^{(1, 2)}$; 若 X 满足所有四个条件, 则 $X \in A\{1, 2, 3, 4\}$, 此时称 X 为 A 的一个 $\{1, 2, 3, 4\}$ 逆 (又称为加号逆或伪逆), 记作 A^+ 或 $A^{(1, 2, 3, 4)}$.

本章主要介绍减号逆, 自反广义逆和加号逆的理论性质、计算方法以应用, 其它广义逆矩阵的性质可以类似地得到.

下面的定理给出 Penrose-Moore 方程的一个等价形式.

定理 8.1 设 $A \in \mathbf{R}^{m \times n}$, 矩阵 $X \in \mathbf{R}^{n \times m}$ 且满足四个 Penrose-Moore 方程, 则 Penrose-Moore 方程与下面的条件等价,

$$\begin{cases} XAA^T = A^T, \\ XX^TA^T = X. \end{cases} \quad (8.1)$$

证明 首先证明条件 $AXA = A$, $(XA)^T = XA$ 与 $XAA^T = A^T$ 等价. 用 X^T 右乘式 (8.1) 的第一式, 得到

$$XAA^TX^T = A^TX^T,$$

即

$$(XA)(XA)^T = (XA)^T.$$

两边同时转置, 得

$$(XA)(XA)^T = XA,$$

由此得

$$(XA)^T = XA.$$

将此式代入式 (8.1) 的第一式, 得到

$$(XA)^TA^T = A^T,$$

两边同时共轭转置, 有

$$AXA = A.$$

反之, Penrose-Moore 方程组的第一和第三个方程可推出式 (8.1) 的第一式. 事实上, 将第三个方程代入第一个方程, 有

$$AXA^T = A.$$

两边取共轭转置, 得

$$XAA^T = A^T.$$

同理可证 Penrose-Moore 方程中的 $XAX = X$, $(AX)^T = AX$ 与 $XX^T A^T = X$ 等价. 证毕.

下面先看几个广义逆的例子.

例 8.1 已知 $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, 验证 B 和 C 均为 A 的减号

逆矩阵.

解 直接计算, 有

$$ABA = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = A,$$

$$ACA = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} = A.$$

由上例可见, A 的减号逆矩阵不唯一.

例 8.2 已知 $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, 验证 A 的加号逆矩阵为自身.

解 令 $X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, 因 $AXA = XAX = AX = XA = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, 从而容易验证 Penrose-

Moore 方程组四个条件都满足.

后续章节将证明, A 的加号逆矩阵唯一.

例 8.3 已知 $B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, 验证 B 加号逆矩阵为 B^T .

解 直接计算, 得

$$BB^T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B^T B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

都是实对称矩阵, 因此 Penrose-Moore 方程组第三个和第四个条件已经满足. 代入计算可知

$$BB^T B = B, \quad B^T BB^T = B^T$$

也满足, 因此 B^T 为 B 的加号逆矩阵.

8.2 广义逆矩阵 A^-

本节介绍减号逆矩阵的性质及其计算方法.

定义 8.3 给定矩阵 $A \in \mathbf{R}^{m \times n}$, 若存在矩阵 $X \in \mathbf{R}^{n \times m}$, 使 $AXA = A$, 则称 X 为 A 的一个 $\{1\}$ -逆, 也称为减号逆或 g 逆, 记作 $A^{(1)}$ 或 A^- . 矩阵 A 的所有 $\{1\}$ -逆的全体记为 $A\{1\}$.

减号逆满足下面的一般结论.

定理 8.2 设 $A \in \mathbf{R}^{m \times n}$, 广义逆矩阵 $A^- \in A\{1\}$ 具有如下性质:

- (1) $(A^-)^T = (A^T)^-$;
- (2) 任取 $\lambda \in \mathbf{R}$, $\lambda^+ A^- \in (\lambda A)\{1\}$, 其中

$$\lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0, \\ 0, & \lambda = 0; \end{cases}$$

- (3) $\text{rank } A^- \geq \text{rank } A$;
- (4) AA^- 和 A^-A 都是幂等矩阵, 且 $\text{rank } AA^- = \text{rank } A^-A = \text{rank } A$;
- (5) 设矩阵 P, Q 可逆, 则 $Q^{-1}A^-P^{-1} \in (PAQ)\{1\}$.

证明 (1) 因为 $A^T(A^-)^T A^T = (AA^-A)^T = A^T$, 所以 $(A^-)^T \in A^T\{1\}$.

(2) 若 $\lambda = 0$, 由定义知零矩阵就是零矩阵的一个 $\{1\}$ 逆. 若 $\lambda \neq 0$, 则 $A = AA^-A$. 于是 $\lambda A = (\lambda A)(\lambda^{-1}A^-)(\lambda A)$, 故 $\lambda^+ A^- \in (\lambda A)\{1\}$.

(3) 由 $AA^-A = A$, 根据两个矩阵之积的秩小于等于这两个矩阵中任一矩阵的秩, 可推出

$$\text{rank } A = \text{rank } AA^-A \leq \text{rank } AA^- \leq \text{rank } A^-.$$

(4) 因为

$$\begin{aligned} (A^-A)^2 &= A^-AA^-A = A^-(AA^-A) = A^-A, \\ (AA^-)^2 &= AA^-AA^- = (AA^-A)A^- = AA^-, \end{aligned}$$

因此, A^-A 和 AA^- 是幂等矩阵. 又因为

$$\text{rank } A = \text{rank } AA^-A \leq \text{rank } AA^- \leq \text{rank } A,$$

故上式中只能等号成立, 即 $\text{rank } A = \text{rank } AA^-$. 同理可证 $\text{rank } A = \text{rank } A^-A$.

(5) 由 $AA^-A = A$ 可推得 $(PAQ)(Q^{-1}A^-P^{-1})(PAQ) = PAQ$, 因此 $Q^{-1}A^-P^{-1} \in (PAQ)\{1\}$. 证毕.

下面研究减号逆的计算方法. 先看一个特殊的例子.

例 8.4 设 $A \in \mathbf{R}^{m \times n}$, 且 A 可写成如下分块矩阵 $A = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$, 其中 E_r 是 r 阶方阵. 利

用定义求 $A\{1\}$.

解 设 $X \in A\{1\}$, 则 X 是 $n \times m$ 矩阵, 将 X 做分块,

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix},$$

其中, $X_{11} \in \mathbf{R}^{r \times r}$, $X_{12} \in \mathbf{R}^{r \times (m-r)}$, $X_{21} \in \mathbf{R}^{(n-r) \times r}$, $X_{22} \in \mathbf{R}^{(n-r) \times (m-r)}$, 于是

$$AXA = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} X_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

由 $AXA = A$ 知 $X_{11} = E_r$, 即 $A\{1\}$ 中的任意一个矩阵可写成

$$X = \begin{pmatrix} E_r & X_{12} \\ X_{21} & X_{22} \end{pmatrix},$$

其中

$$X_{12} \in \mathbf{R}^{r \times (m-r)}, \quad X_{21} \in \mathbf{R}^{(n-r) \times r}, \quad X_{22} \in \mathbf{R}^{(n-r) \times (m-r)},$$

为任意矩阵.

下面的定理给出了一般矩阵全体减号逆的表达式.

定理 8.3 设 $A \in \mathbf{R}^{m \times n}$, $\text{rank } A = r$, $P \in \mathbf{R}^{m \times m}$, $Q \in \mathbf{R}^{n \times n}$, P 和 Q 可逆, 且

$$PAQ = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix},$$

则 $A\{1\}$ 中任一矩阵可写成

$$Q \begin{pmatrix} E_r & X_{12} \\ X_{21} & X_{22} \end{pmatrix} P,$$

其中 $X_{12} \in \mathbf{R}^{r \times (n-r)}$, $X_{21} \in \mathbf{R}^{(m-r) \times r}$, $X_{22} \in \mathbf{R}^{(m-r) \times (n-r)}$ 为任意矩阵.

证明 由例 8.4 及定理 8.2 的性质 (5) 即可推出本定理. 证毕.

定理 8.3 给出了求 A^- 的一种计算方法. 这种方法须先求出可逆矩阵 P 和 Q , 使 PAQ 成为标准形. 为此可以先构造分块矩阵 B , 使

$$B = \begin{pmatrix} A & E_m \\ E_n & 0 \end{pmatrix}.$$

用行和列初等变换把 B 中的 A 化成如下形式的标准形

$$\tilde{A} = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix},$$

同时, E_n 化成了 Q , E_m 化成了 P , 即

$$\begin{pmatrix} P & 0 \\ 0 & E_n \end{pmatrix} \begin{pmatrix} A & E_m \\ E_n & 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & E_m \end{pmatrix} = \begin{pmatrix} \tilde{A} & P \\ Q & 0 \end{pmatrix},$$

故

$$PAQ = \tilde{A} = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix},$$

于是 $A\{1\}$ 中的矩阵可写成

$$X = Q \begin{pmatrix} E_r & X_{12} \\ X_{21} & X_{22} \end{pmatrix} P.$$

定理 8.3 表明 A^- 是存在的, 即 $A\{1\}$ 是非空集合. 由于 X_{12} , X_{21} , X_{22} 中的元素可任取, 故当 A 不是可逆方阵时, A^- 不唯一.

例 8.5 已知矩阵 $A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{pmatrix}$, 求 A 的广义逆 $A\{1\}$.

解 对分块矩阵 B 做初等变换, 得

$$B = \begin{pmatrix} 1 & 0 & -1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 2 & 2 & 0 & 1 & 0 \\ -1 & 4 & 5 & 3 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & & & \\ 0 & 1 & 0 & 0 & & \mathbf{0} & \\ 0 & 0 & 1 & 0 & & & \\ 0 & 0 & 0 & 1 & & & \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & 0 & 1 & -1 & & & \\ 0 & 1 & -1 & -1 & & \mathbf{0} & \\ 0 & 0 & 1 & 0 & & & \\ 0 & 0 & 0 & 1 & & & \end{pmatrix},$$

于是

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 1 & -2 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

因此, A 的任一个 $\{1\}$ 逆可写成

$$X = Q \begin{pmatrix} 1 & 0 & x_1 \\ 0 & 1 & x_2 \\ y_{11} & y_{12} & z_1 \\ y_{21} & y_{22} & z_2 \end{pmatrix} P,$$

其中, $x_i, y_{ij}, z_j (i=1,2; j=1,2)$ 为任意实数.

若取 $x_i = y_{ij} = z_j = 0 (i=1,2; j=1,2)$, 则得到 A 的一个具体的 $\{1\}$ -逆

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

如果 $x_i, y_{ij}, z_j (i=1,2; j=1,2)$ 取为其他数, 就可以得到另一个 A^- .

下述定理给出已知某一个 $\{1\}$ 逆后所有 A^- 的一般表达式.

定理 8.4 设 $A \in \mathbf{R}^{m \times n}$, $A^- \in A\{1\}$ 是 A 的某一个 $\{1\}$ 逆, 则

(1) $X = A^- + U - A^-AUA A^- \in A\{1\}$, 矩阵 $U \in \mathbf{R}^{n \times m}$ 是任意实矩阵,

(2) $X = A^- + V(E_m - AA^-) + (E_n - A^-A)U$, 其中 $U, V \in \mathbf{R}^{n \times m}$,

也是 A 的某一个 $\{1\}$ -逆, 且 $A\{1\}$ 中任何一个矩阵都可以表示成上述形式.

证明 先证明 (1) (2) 中的 $X \in A\{1\}$. 分别将上述 X 代入 AXA 验证得,

$$\begin{aligned} & A(A^- + U - A^-AUA A^-)A \\ &= AA^-A + AUA - (AA^-A)U(AA^-A) \\ &= A + AUA - AUA = A, \\ & A(A^- + V(E_m - AA^-) + (E_n - A^-A)U)A \\ &= AA^-A + AV(E_m - AA^-)A + A(E_n - A^-A)UA \\ &= AA^-A + AV(A - AA^-A) + (A - AA^-A)UA = A, \end{aligned}$$

即 $X \in A\{1\}$.

再设任给 $X \in A\{1\}$, 则

$$A(X - A^-)A = AXA - AA^-A = A - A = 0.$$

令 $U = B - A^-$, 则 $AUA = 0$, 于是 $A^-AUA A^- = 0$, 故

$$X = A^- + U - A^-AUA A^-.$$

取

$$V = X - A^-, \quad U = XAA^-,$$

则

$$X = A^- + V(E_m - AA^-) + (E_n - A^-A)U.$$

故 $A\{1\}$ 中任何一个矩阵都可以表示成上述形式. 证毕.

除了使用定理 8.3 的方法外, 还可以利用矩阵的满秩分解来求 A^- .

对于行满秩矩阵, 有如下结论.

定理 8.5 设 $A \in \mathbf{R}^{m \times n}$ ($m \leq n$) 为行满秩矩阵, 则存在 $B \in \mathbf{R}^{n \times m}$, 使得 $AB = E$, 称 B 为 A 的右逆, 记为 A_R^{-1} .

证明 根据定理 1.20, AA^T 是 m 阶可逆方阵, 故

$$(AA^T)(AA^T)^{-1} = (AA^T)^{-1}(AA^T) = E,$$

从而 A 的右逆可取为 $A_R^{-1} = A^T(AA^T)^{-1}$.

例 8.6 设矩阵 $A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$, 求 A 的右逆 A_R^{-1} .

解 由 $\text{rank } A = 2$, 知 A 为行满秩矩阵, 所以

$$A_R^{-1} = A^T(AA^T)^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ -1 & 2 \end{pmatrix} \left(\begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ -1 & 2 \end{pmatrix} \right)^{-1} = \frac{1}{14} \begin{pmatrix} 5 & 4 \\ 6 & 2 \\ 3 & 8 \end{pmatrix}.$$

相应地, 对于列满秩矩阵, 有如下结论.

定理 8.6 设 $A \in \mathbf{R}^{m \times n}$ ($m \geq n$) 是列满秩矩阵, 则存在 $B \in \mathbf{R}^{n \times m}$, 使得 $BA = E$, 称 B 为 A 的左逆, 记为 A_L^{-1} .

证明 注意到方阵 $A^T A$ 可逆, 故

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

从而 A 的左逆可取为

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

例 8.7 求矩阵 $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix}$ 的左逆 A_L^{-1} .

解 由 $\text{rank } A = 2$, 知 A 为列满秩矩阵, 所以

$$A_L^{-1} = (A^T A)^{-1} A^T = \left(\begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{pmatrix} = \frac{1}{11} \begin{pmatrix} -4 & 7 & 1 \\ 7 & -4 & 1 \end{pmatrix}.$$

注 8.1 矩阵 A 的左逆与右逆一般不能同时存在, 只有当 $m = n$ 时的可逆方阵, 左逆和右逆才能同时存在. 一般 $m \times n$ 矩阵的左逆和右逆也不唯一, 只有是满秩方阵的时候才是唯一的. 可以验证, 行满秩 $m \times n$ 矩阵 A 的右逆一般表达式为

$$B = VA^T(AVA^T)^{-1},$$

其中, V 是使等式 $\text{rank } AVA^T = \text{rank } A = m$ 成立的任意 n 阶方阵. 列满秩 $m \times n$ 矩阵 A 的左逆一般表达式为

$$B = (A^T UA)^{-1} A^T U,$$

其中, U 是使等式 $\text{rank } (A^T UA) = \text{rank } A = n$ 成立的任意 m 阶方阵.

定理 8.7 设 A 为 $m \times n$ 矩阵, 其满秩分解为 $A = BC$, 其中 B 为 $m \times r$ 矩阵, C 为 $r \times n$ 矩阵, 且 $\text{rank } A = \text{rank } B = \text{rank } C = r \leq \min\{m, n\}$. 若 B 的一个左逆为 B_L^{-1} , C 的一个右逆 C_R^{-1} , 则

$$A^- = C_R^{-1} B_L^{-1},$$

特别, A^- 可取为

$$A^- = C^T (CC^T)^{-1} (B^T B)^{-1} B^T.$$

证明 将上式代入 Penrose-Moore 方程, 有

$$AXA = AC_R^{-1} B_L^{-1} A = BCC_R^{-1} B_L^{-1} BC = BC = A.$$

证毕.

例 8.8 用满秩分解求矩阵 $A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 3 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ 的一个广义逆 A^- .

解 因为

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 3 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

由前面两个例子可知,

$$A^- = C_R^{-1} B_L^{-1} = \frac{1}{14} \begin{pmatrix} 5 & 4 \\ 6 & 2 \\ 3 & 8 \end{pmatrix} \frac{1}{11} \begin{pmatrix} -4 & 7 & 1 \\ 7 & -4 & 1 \end{pmatrix} = \frac{1}{154} \begin{pmatrix} 8 & 19 & 9 \\ -10 & 34 & 8 \\ 44 & -11 & 11 \end{pmatrix}.$$

8.3 自反广义逆 $A\{1,2\}$

本节介绍自反广义逆 $A\{1,2\}$ 的概念和性质, 以及与减号逆 $A\{1\}$ 之间的关系.

定义 8.4 设 $A \in \mathbf{R}^{m \times n}$, 若 $X \in \mathbf{R}^{n \times m}$ 且满足方程

$$AXA = A,$$

$$XAX = X,$$

则 $X \in A\{1,2\}$. 此时称 X 为 A 的一个 $\{1,2\}$ -逆或自反广义逆, 记为 $X = A^{(1,2)}$. 矩阵 A 的所有自反广义逆的集合记为 $A\{1,2\}$.

根据定义可直接得到, 若 X 是 A 的自反广义逆, 则 A 也是 X 的自反广义逆, 这就是自反的含义.

定理 8.8 任何矩阵 $A \in \mathbf{R}^{m \times n}$ 都有自反广义逆.

证明 如果 $A = \mathbf{0}$, 则 $X = \mathbf{0}$, 显然就是 A 的自反广义逆. 如果 $A \neq \mathbf{0}$, $\text{rank } A = r$, 则存在可逆矩阵 P 和 Q , 使

$$PAQ = \begin{pmatrix} E_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

直接验证可知矩阵

$$X = Q \begin{pmatrix} E_r & W \\ V & VW \end{pmatrix} P,$$

是 A 的自反广义逆, 其中, $W \in \mathbf{R}^{r \times (m-r)}$, $V \in \mathbf{R}^{(n-r) \times r}$ 是任意矩阵. 证毕.

定理 8.8 说明自反广义逆存在, 但通常不唯一.

例 8.9 求矩阵 $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}$ 的自反广义逆.

解 由初等矩阵变换可知,

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

故

$$A^{(1,2)} = Q \begin{pmatrix} 1 & 0 & w_1 \\ 0 & 1 & w_2 \\ v_1 & v_2 & v_1 w_1 + v_2 w_2 \end{pmatrix} P,$$

其中 $w_i, v_i (i=1, 2)$ 为任意常数.

定理 8.9 对任何 $A \in \mathbf{R}^{m \times n}$, 若 $Y, Z \in A\{1\}$, 则 $X = YAZ \in A\{1, 2\}$.

证明 直接验证, 又

$$AXA = (AYA)ZA = AZA = A,$$

$$XAX = Y(AZA)YAZ = Y(AYA)Z = YAZ = X,$$

因此, X 是 A 的 $\{1, 2\}$ -逆. 证毕.

引理 8.10 设 $A \in \mathbf{R}^{m \times n}$, $X \in \mathbf{R}^{n \times m}$, 若

$$\mathcal{R}(XA) = \mathcal{R}(X),$$

则存在 $Y \in \mathbf{R}^{n \times m}$, 使 $XAY = X$.

证明 令 q_1, q_2, \dots, q_m 是 \mathbf{R}^m 的一组基,

$$r_i = Xq_i, \quad i = 1, 2, \dots, m,$$

则 $r_i \in \mathcal{R}(X)$.

因为 $\mathcal{R}(XA) = \mathcal{R}(X)$, 所以 $r_i \in \mathcal{R}(XA)$, 即存在 $p_i \in \mathbf{R}^n$, 使得

$$r_i = XAp_i, \quad i = 1, 2, \dots, m.$$

记

$$P = (p_1, p_2, \dots, p_m) \in \mathbf{R}^{n \times m},$$

$$Q = (q_1, q_2, \dots, q_m) \in \mathbf{R}^{m \times m},$$

显然 Q 可逆. 由 $Xq_i = XAp_i$ 知 $XQ = XAP$, 即

$$X = XAPQ^{-1}.$$

令 $Y = PQ^{-1}$, 于是 $X = XAY$. 证毕.

定理 8.11 设 $A \in \mathbf{R}^{m \times n}$, $\text{rank } A = r$, $X \in \mathbf{R}^{n \times m}$, 且 $X \in A\{1\}$, 则 $X \in A\{1, 2\}$ 的充分必要条件是 $\text{rank } X = \text{rank } A$.

证明 必要性. 因为 $X \in A\{1, 2\}$, 所以 $AXA = A$, $XAX = X$. 由 $AXA = A$ 可得

$$\text{rank } A = \text{rank } AXA \leq \text{rank } AX \leq \text{rank } X;$$

同理, 根据 $XAX = X$ 可得 $\text{rank } X \leq \text{rank } A$. 从而 $\text{rank } X = \text{rank } A$.

充分性. 由定理 8.2 的 (4) 知

$$\text{rank } XA = \text{rank } A,$$

结合 $\text{rank } X = \text{rank } A$ ，得

$$\text{rank } XA = \text{rank } X,$$

即 $\mathcal{R}(XA)$ 与 $\mathcal{R}(X)$ 的维数相等，而 $\mathcal{R}(XA) \subset \mathcal{R}(X)$ ，故 $\mathcal{R}(XA) = \mathcal{R}(X)$ 。由引理 8.10 知，存在 $Y \in \mathbf{R}^{n \times m}$ ，使 $XAY = X$ ，左乘 A 得

$$AX = AXAY = (AXA)Y = AY,$$

所以

$$XAX = XAY = X, \quad X \in A\{1, 2\}$$

证毕。

8.4 广义逆矩阵 A^+

加号逆 A^+ 是应用最广泛的一类矩阵的广义逆，本节介绍它的定义、性质和计算方法。首先，加号逆存在且唯一。

定理 8.12 任给矩阵 $A \in \mathbf{R}^{m \times n}$ ， A^+ 存在且唯一。

证明 先证明广义逆的存在性。当 $\text{rank } A = 0$ ，即 $A = \mathbf{0}$ 为零矩阵时，容易验证 $X = \mathbf{0}$ 是 A 的一个 A^+ 逆。当 $\text{rank } A = r \neq 0$ 时，取 A 的一个满秩分解为 $A = BC$ ，令

$$X = C^T(CC^T)^{-1}(B^TB)^{-1}B^T, \quad (8.2)$$

则

$$\begin{aligned} AXA &= BCC^T(CC^T)^{-1}(B^TB)^{-1}B^TBC = BC = A, \\ XAX &= C^T(CC^T)^{-1}(B^TB)^{-1}B^TBCC^T(CC^T)^{-1}(B^TB)^{-1}B^T \\ &= C^T(CC^T)^{-1}(B^TB)^{-1}B^T = X, \\ (AX)^T &= (BCC^T(CC^T)^{-1}(B^TB)^{-1}B^T)^T = (B(B^TB)^{-1}B^T)^T = B((B^TB)^{-1})^TB^T \\ &= B(B^TB)^{-1}B^T = AX, \\ (XA)^T &= (C^T(CC^T)^{-1}(B^TB)^{-1}B^TBC)^T = (C^T(CC^T)^{-1}C)^T = C^T((CC^T)^{-1})^TC \\ &= C^T(CC^T)^{-1}C = XA \end{aligned}$$

所以 X 是 A 的一个 A^+ 逆。

再证明广义逆的唯一性。设 X_1 和 X_2 都是 A 的 A^+ 逆，由定理 8.1 和 Penrose-Moore 方程组可知

$$\begin{aligned} X_1 &= X_1X_1^TA^T = X_1X_1^T(AX_2A)^T = X_1X_1^TA^TX_2^TA^T \\ &= X_1(AX_1)^T(AX_2)^T = X_1AX_1AX_2 = X_1AX_2 \\ &= X_1AX_2AX_2 = X_1A(X_2A)^TX_2 = X_1AA^TX_2^TX_2 \\ &= A^TX_2^TX_2 = (X_2A)^TX_2 = X_2AX_2 = X_2. \end{aligned}$$

证毕。

广义逆矩阵 A^+ 有如下运算性质。

定理 8.13 (1) $(A^+)^+ = A$;

(2) $(A^T)^+ = (A^+)^T$;

(3) 对 $\lambda \in \mathbf{R}$, $(\lambda A)^+ = \lambda^+ A^+$, 其中 λ^+ 约定为

$$\lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0, \\ 0, & \lambda = 0; \end{cases}$$

(4) $\text{rank } A^+ = \text{rank } A^+ A = \text{rank } A A^+ = \text{rank } A$;

(5) $A^+ = (A^T A)^+ A^T = A^T (A A^T)^+$;

(6) $\mathcal{R}(A A^+) = \mathcal{R}(A)$, $\mathcal{R}(A^+ A) = \mathcal{R}(A^+) = \mathcal{R}(A^T)$;

(7) $(A^T A)^+ = A^+ (A^T)^+$, $(A A^T)^+ = (A^T)^+ A^+$;

(8) A 为实对称方阵时, $A A^+ = A^+ A$.

(9) $A = A A^T (A^+)^T = (A^+)^T A^T A$, $A^T = A^T A A^+ = A^+ A A^T$.

证明 性质 (1)~(4) 由定义容易证明, 故仅证明性质 (5)~(8).

(5) 设 $A = BC$ 为 A 的满秩分解, 由定理 8.12 得

$$A^+ = C^T (C C^T)^{-1} (B^T B)^{-1} B^T,$$

于是

$$A^T A = C^T B^T B C = (C^T B^T B) C.$$

此式可视为 $A^T A$ 的一个满秩分解, 从而有

$$\begin{aligned} (A^T A)^+ &= C^T (C C^T)^{-1} ((C^T B^T)^T (C^T B^T B))^{-1} (C^T B^T B)^T \\ &= C^T (C C^T)^{-1} (B^T B C C^T B^T B)^{-1} B^T B C \\ &= C^T (C C^T)^{-1} (B^T B^{-1} (C C^T)^{-1} (B^T B)^{-1} (B^T B) C \\ &= C^T (C C^T)^{-1} (B^T B)^{-1} (C C^T)^{-1} C, \end{aligned}$$

于是

$$\begin{aligned} (A^T A)^+ A^T &= C^T (C C^T)^{-1} (B^T B)^{-1} (C C^T)^{-1} C C^T B^T \\ &= C^T (C C^T)^{-1} (B^T B)^{-1} B^T = A^+. \end{aligned}$$

同理可证 $A^+ = A^T (A A^T)^+$.

(6) 因 $\mathcal{R}(A A^+) \subset \mathcal{R}(A)$, $\mathcal{R}(A^+ A) \subset \mathcal{R}(A^+)$, 以及

$$\mathcal{R}(A^+) = \mathcal{R}(A^T (A A^T)^+) \subset \mathcal{R}(A^T),$$

结合性质 (4), 得 $\mathcal{R}(A A^+) = \mathcal{R}(A)$, $\mathcal{R}(A^+ A) = \mathcal{R}(A^+) = \mathcal{R}(A^T)$.

(7) 由性质 (5) 得

$$A^+ (A^T)^+ = (A^T A)^+ A^T A (A^T A)^+ = (A^T A)^+,$$

$$(A^T)^+ A^+ = (A A^T)^+ A A^T (A A^T)^+ = (A A^T)^+.$$

(8) $A^+ A = (A^+ A)^T = A^T (A^T)^+ = A A^+$.

(9) 首先根据 Penrose-Moore 方程组, 得

$$A = A A^+ A = A (A^+ A)^T = A A^T (A^+)^T,$$

$$A^T = A^T (A^T)^+ A^T = A^T ((A^T)^+ A^T)^T = A^T A A^+,$$

再对两式两边取转置, 即得 $A = (A^+)^T A^T A$, $A^T = A^+ A A^T$. 证毕.

对给定矩阵 A , 如何求出 A^+ 是研究广义逆问题的一个重点. 本节将具体讨论 A^+ 的各种计算方法.

先考虑当 A 为某些特殊矩阵时, A^+ 的计算方法, 有如下结论.

定理 8.14 (1) A 是满秩 n 阶方阵, 则 $A^+ = A^{-1}$;

(2) A 是对角矩阵, 即 $A = \text{diag}\{d_1, d_2, \dots, d_n\}$, 则

$$A^+ = \text{diag}\{d_1^+, d_2^+, \dots, d_n^+\};$$

(3) A 是行满秩矩阵时, $A^+ = A^T(AA^T)^{-1}$;

(4) A 是列满秩矩阵时, $A^+ = (A^T A)^{-1} A^T$;

(5) A 是分块对角矩阵, $A = \text{diag}\{A_1, A_2, \dots, A_i\}$, 则

$$A^+ = \text{diag}\{A_1^+, A_2^+, \dots, A_i^+\};$$

(6) A 是实对称矩阵, $A = Q\Lambda Q^T$ 为其正交合同对角化, 则

$$A^+ = Q\Lambda^+ Q^T.$$

证明 (1) (2) (5) (6) 可直接验证. 对 (3), 注意到 AA^T 是可逆矩阵, 由定理 8.13 性质 (5) 可推知结论. 对 (4), 同理可证. 证毕.

例 8.10 已知 $A = \begin{pmatrix} 4 & & \\ & 2 & \\ & & 0 \end{pmatrix}$, 则 $A^+ = \begin{pmatrix} \frac{1}{4} & & \\ & \frac{1}{2} & \\ & & 0 \end{pmatrix}$.

例 8.11 已知 $A = (a_1, a_2, \dots, a_n)$ 为非零行矩阵, 求 A^+ .

解 A 可看作是行满秩矩阵 (其行数与秩都是 1). 因为

$$AA^T = (a_1, a_2, \dots, a_n) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1^2 + a_2^2 + \dots + a_n^2 \neq 0,$$

于是

$$A^+ = A^T(AA^T)^{-1} = \frac{1}{a_1^2 + a_2^2 + \dots + a_n^2} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

例 8.12 已知 $A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ 为非零列矩阵, 求 A^+ .

解 A 可看作是列满秩矩阵, 于是

$$A^T A = (a_1, a_2, \dots, a_n) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a_1^2 + a_2^2 + \dots + a_n^2 \neq 0,$$

$$A^+ = (A^T A)^{-1} A^T = \frac{1}{a_1^2 + a_2^2 + \dots + a_n^2} (a_1, a_2, \dots, a_n).$$

对列满秩矩阵 A , 由定理 7.6, 可将 A 的 n 列标准正交化, 得 $A = QR$, 其中 Q 为列正交矩阵. 于是

$$\begin{aligned} A^+ &= (A^T A)^{-1} A^T = ((QR)^T QR)^{-1} (QR)^T \\ &= (R^T Q^T QR)^{-1} R^T Q^T = (R^T R)^{-1} R^T Q^T \\ &= R^{-1} (R^T)^{-1} R^T Q^T = R^{-1} Q^T. \end{aligned}$$

例 8.13 已知 $A = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & 1 \end{pmatrix}$, 求 A^+ .

解 把 A 的两列标准正交化, 得

$$Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 2 & -1 \end{pmatrix}, \quad R = \sqrt{5} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

于是

$$A^+ = R^{-1} Q^T = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & -1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} -1 & 0 & 3 \\ 2 & 0 & -1 \end{pmatrix}.$$

例 8.14 已知矩阵 $A = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$, 求 A^+ .

解 容易求得 A 的特征值为 $\lambda_1 = 0$, $\lambda_2 = 5$, 相应的特征向量为

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

标准正交化后得

$$\mathbf{x}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \quad \mathbf{x}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

于是

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}, \quad Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad A^+ = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}.$$

从而

$$A^+ = QA^+Q^T = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} = \frac{1}{25} \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}.$$

当 A 为一般矩阵时, 有三种计算 A^+ 的方法. 在具体应用中, 读者可根据矩阵 A 的实际情况, 选择适当的方法计算.

方法 1 通过满秩分解, 利用公式 (8.2) 计算 A^+ .

例 8.15 设矩阵 $A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 2 & -2 \end{pmatrix}$, 求 A^+ .

解 A 的满秩分解为 $A = BC$, 其中

$$B = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad C = (1 \ -1), \quad CC^T = (1 \ -1) \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 2,$$

$$B^TB = (1 \ 1 \ 2) \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = 6, \quad (CC^T)^{-1} = \frac{1}{2}, \quad (B^TB)^{-1} = \frac{1}{6}.$$

于是

$$A^+ = C^T(CC^T)^{-1}(B^TB)^{-1}B^T = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \times \frac{1}{2} \times \frac{1}{6} \times (1 \ 1 \ 2) = \frac{1}{12} \begin{pmatrix} 1 & 1 & 2 \\ -1 & -1 & -2 \end{pmatrix}.$$

综上可以看出, 此时 A^+ 的计算方法与 A^- 的计算方法相同, 这样方便记忆.

方法 2 利用广义逆 A^+ 的如下性质

$$A^+ = (A^TA)^+A^T = A^T(AA^T)^+,$$

和对称矩阵 A^TA 或者 AA^T 的谱分解计算 A^+ .

设 $A \in \mathbf{R}^{m \times n}$, 于是 A^TA 是正定或半正定矩阵, 故存在正交矩阵 Q , 使

$$Q^TA^TAQ = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\},$$

其中 $\lambda_i \geq 0$ ($i = 1, 2, \dots, n$) 是 A^TA 的特征值, 于是

$$A^TA = QAQ^T.$$

由实对称矩阵广义逆的计算方法知

$$(A^TA)^+ = QA^+Q^T,$$

故

$$A^+ = (A^TA)^+A^T = QA^+Q^TA^T.$$

例 8.16 已知矩阵 $A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 2 & -2 \end{pmatrix}$, 求 A^+ .

解 直接计算, 得

$$\begin{aligned} A^T A &= \begin{pmatrix} 6 & -6 \\ -6 & 6 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \\ (A^T A)^+ &= \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{24} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \\ A^+ &= (A^T A)^+ A^T = \frac{1}{24} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ -1 & -1 & -2 \end{pmatrix} = \frac{1}{12} \begin{pmatrix} 1 & 1 & 2 \\ -1 & -1 & -2 \end{pmatrix}. \end{aligned}$$

注 8.2 当 $m < n$ 时, 矩阵 AA^T 的阶数小于矩阵 $A^T A$ 的阶数, 为了使运算简单, 建议采用如下计算公式:

$$A^+ = A^T (AA^T)^+ = A^T Q \Lambda^+ Q^T,$$

其中, Λ 是由矩阵 AA^T 的特征值构成的对角阵, Q 是把 AA^T 化成 Λ 的正交矩阵.

例 8.17 设 $A = \begin{pmatrix} 1 & 1 & -1 \\ -2 & -2 & 2 \end{pmatrix}$, 求 A^+ .

解 直接计算, 得

$$AA^T = \begin{pmatrix} 1 & 1 & -1 \\ -2 & -2 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & -2 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & -6 \\ -6 & 12 \end{pmatrix},$$

进而求得 $|\lambda E - AA^T| = \lambda(\lambda - 15)$, 故 AA^T 的特征值为 $\lambda_1 = 0$, $\lambda_2 = 15$, 它们对应的单位特征向量分别是

$$x_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad x_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

于是

$$Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 0 & 0 \\ 0 & 15 \end{pmatrix}, \quad \Lambda^+ = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{15} \end{pmatrix}.$$

所以

$$A^+ = A^T Q \Lambda^+ Q^T = \begin{pmatrix} 1 & -2 \\ 1 & -2 \\ -1 & 2 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{15} \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix} = \frac{1}{15} \begin{pmatrix} 1 & -2 \\ 1 & -2 \\ -1 & 2 \end{pmatrix}.$$

方法3 利用矩阵 A 的奇异值分解计算 A^+ .

定理 8.15 已知矩阵 A 的奇异值分解为 $A = U\Sigma V^H$, 其中 $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$, U 和 V 是酉矩阵,

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m \times n},$$

D 是由矩阵 A 的非零奇异值构成的对角阵, 则 $A^+ = V\Sigma^+U^T$.

证明 直接验证即得结论. 证毕.

当矩阵 A 的阶数比较高时, 求广义逆 A^+ 通常要借助计算机采用迭代法来求解, 对这方面的知识有兴趣的读者可以参考相关文献.

8.5 广义逆矩阵的应用

前面讨论了广义逆的定义、性质和计算方法. 有了广义逆, 对非齐次线性方程组 $Ax = b$ 的求解就可以更广了. 如果存在 x 使方程组满足, 则称方程组相容; 否则称为不相容方程组. 本节将分两部分来探讨广义逆的应用.

相容方程组中, 范数 (即由内积诱导的范数) 最小的解 x_0 称为极小范数解. 将会讨论极小范数解是唯一的.

不相容方程组中, 取 x_0 使 $\|Ax_0 - b\| = \min\{\|Ax - b\| : x \in \mathbb{R}^n\}$, 则称 x_0 为最小二乘解. 最小二乘解不唯一, 而其中范数最小的一个称为极小范数最小二乘解, 或极小最小二乘解. 极小范数最小二乘解是否一定存在? 是否唯一呢? 这将是讨论的问题.

8.5.1 广义逆在解线性方程组中的应用

定理 8.16 设 $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times q}$, $D \in \mathbb{R}^{m \times p}$ 为已知矩阵, $X \in \mathbb{R}^{n \times p}$ 为未知矩阵, 则

(1) 矩阵方程 $AXB = D$ 有解的充分必要条件为

$$AA^+DB^+B = D; \quad (8.3)$$

(2) 当 $AXB = D$ 有解时 A^+DB^+ 是方程的一个特解, 其通解为

$$X = A^+DB^+ + Y - A^+AYBB^+, \quad (8.4)$$

其中 $Y \in \mathbb{R}^{n \times p}$ 为任意矩阵.

证明 (1) 充分性. 已知 $AA^+DB^+B = D$, 取 $X_0 = A^+DB^+$, 则 X_0 是矩阵方程 $AXB = D$ 的解.

必要性. 若 $AXB = D$ 有解, 记 X_0 是矩阵方程 $AXB = D$ 的一个解, 则

$$D = AX_0B = AA^+AX_0BB^+B = AA^+DB^+B,$$

即式 (8.3) 成立.

(2) 因 $AXB = D$ 有解时 A^+DB^+ 是一个特解. 在等式 (8.4) 的左边乘 A , 右边乘 B , 得

$$AXB = AA^+DB^+B + AYB - AA^+AYBB^+B = D.$$

这说明任取矩阵 $Y \in \mathbf{R}^{n \times p}$, 式 (8.4) 右端所表示的矩阵 X 都是矩阵方程 $AXB = D$ 的解.

反之, 若 X_0 是矩阵方程 $AXB = D$ 的一个解, 即 $AX_0B = D$, 于是

$$X_0 = A^+DB^+ + X_0 - A^+DB^+ = A^+DB^+ + X_0 - A^+AX_0BB^+,$$

把 X_0 视作 Y , 则 X_0 可以表示成式 (8.4) 右端的形式. 证毕.

在定理 8.16, 中令 $p = q = 1$, $B = 1$, $D = b$, 即可得到本下面的结论.

定理 8.17 设矩阵 $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, 则

(1) 线性方程组 $Ax = b$ 有解的充分必要条件是 $AA^+b = b$;

(2) 方程组有解时, A^+b 为它的一个特解, 方程组的通解为

$$x = A^+b + (E - A^+A)y,$$

其中, $y \in \mathbf{R}^n$ 为任意向量.

以下考虑 A^+b 的在标准欧氏空间中的几何特性 (本节后续涉及的几何问题, 均在标准欧氏空间中研究). 为此, 先证明一个引理, 相关符号的意义见定理 1.26 和例 5.19.

引理 8.18 设矩阵 $A \in \mathbf{R}^{m \times n}$, 则 $\mathcal{R}(A^T)^\perp = \mathcal{N}(A)$, 从而 $\mathbf{R}^n = \mathcal{R}(A^T) \oplus \mathcal{N}(A)$.

证明 若 $y \in \mathcal{R}(A^T)^\perp$, 则对一切 $x \in \mathbf{R}^m$, $A^Tx \in \mathcal{R}(A^T)$, 有 $0 = (y, A^Tx) = (Ay, x)$. 由 x 的任意性知 $Ay = 0$, 即 $y \in \mathcal{N}(A)$. 因此 $\mathcal{R}(A^T)^\perp \subset \mathcal{N}(A)$.

若 $y \in \mathcal{N}(A)$, 则 $Ay = 0$, 于是对一切 $x \in \mathbf{R}^m$, $0 = (Ay, x) = (y, A^Tx)$, 从而 $y \in \mathcal{R}(A^T)^\perp$, 因此 $\mathcal{N}(A) \subset \mathcal{R}(A^T)^\perp$. 证毕.

定理 8.19 设矩阵 $A \in \mathbf{R}^{m \times n}$, 方程组 $Ax = b$ 有解, 则 $x_0 = A^+b$ 为 $Ax = b$ 的唯一属于 $\mathcal{R}(A^T)$ 的解, 且是全部解中长度最小的解, 即 x 为 $Ax = b$ 的任一解时, $\|x_0\| \leq \|x\|$.

证明 由定理 8.13 的性质 (6) 知 $\mathcal{R}(A^+) = \mathcal{R}(A^T)$, 显然 $x_0 = A^+b \in \mathcal{R}(A^+)$, 故 $x_0 \in \mathcal{R}(A^T)$.

设 x 为 $Ax = b$ 的任一解. 记 $y = x - x_0$, 则 $Ay = 0$, 所以 $y \in \mathcal{N}(A) = \mathcal{R}(A^T)^\perp$. 由勾股定理 (定理 6.15) 得

$$\|x\|^2 = \|x_0 + y\|^2 = \|x_0\|^2 + \|y\|^2 \geq \|x_0\|^2,$$

而 $x = x_0 + y$, 由分解的唯一性得 x_0 是唯一属于 $\mathcal{R}(A^T)$ 的解. 证毕.

例 8.18 已知线性方程组 $Ax = b$ 中,

$$A = \begin{pmatrix} 1 & 1 & -1 \\ -2 & -2 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 2 \end{pmatrix},$$

求该方程组的最小长度解和通解.

解 由例 8.17 知

$$A^+ = \frac{1}{15} \begin{pmatrix} 1 & -2 \\ 1 & -2 \\ -1 & 2 \end{pmatrix},$$

故最小长度解为

$$\mathbf{x}_0 = \mathbf{A}^+ \mathbf{b} = \frac{1}{15} \begin{pmatrix} 1 & -2 \\ 1 & -2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \frac{1}{15} \begin{pmatrix} -5 \\ -5 \\ 5 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix},$$

因为

$$\mathbf{A}^+ \mathbf{A} = \frac{1}{15} \begin{pmatrix} 1 & -2 \\ 1 & -2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ -2 & -2 & 2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix},$$

所以通解为

$$\mathbf{x} = \mathbf{A}^+ \mathbf{b} + (\mathbf{E} - \mathbf{A}^+ \mathbf{A}) \mathbf{y} = \frac{1}{3} \begin{pmatrix} -1 + 2y_1 - y_2 + y_3 \\ -1 - y_1 + 2y_2 + y_3 \\ 1 + y_1 + y_2 + 2y_3 \end{pmatrix}.$$

8.5.2 广义逆在解线性最小二乘问题上的应用

设矩阵 $\mathbf{A} \in \mathbf{R}^{m \times n}$, 向量 $\mathbf{b} \in \mathbf{R}^m$, 线性方程组 $\mathbf{Ax} = \mathbf{b}$ 当且仅当 $\mathbf{b} \in \mathcal{R}(\mathbf{A})$ 时有解. 定理 8.17 给出了方程组 $\mathbf{Ax} = \mathbf{b}$ 有解的充分必要条件和通解的表达式. 当 $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$, 不存在可使 $\mathbf{Ax} = \mathbf{b}$ 成立的 \mathbf{x} , 此时 $\mathbf{Ax} = \mathbf{b}$ 称为矛盾方程组,

给定了矩阵 \mathbf{A} 和向量 \mathbf{b} , 既然对一切 $\mathbf{x} \in \mathbf{R}^n$, 残差向量 $\mathbf{b} - \mathbf{Ax}$ 都不等于零, 很自然地希望求一个 $\mathbf{x}_* \in \mathbf{R}^n$, 使得残差达到最小, 即

$$\|\mathbf{b} - \mathbf{Ax}_*\| = \min_{\mathbf{x} \in \mathbf{R}^n} \|\mathbf{b} - \mathbf{Ax}\|, \quad (8.5)$$

此时将满足式 (8.5) 的向量 \mathbf{x}_* 称为方程组 $\mathbf{Ax} = \mathbf{b}$ 的最小二乘解.

定理 8.20 \mathbf{x}_* 为线性方程组 $\mathbf{Ax} = \mathbf{b}$ 的最小二乘解的充分必要条件是 \mathbf{x}_* 为方程组

$$\mathbf{A}^T (\mathbf{Ax} - \mathbf{b}) = \mathbf{0} \quad (8.6)$$

的解. 通常称方程组 (8.6) 为法方程组或正规方程组.

证明 由于 $\mathcal{R}(\mathbf{A})$ 是子空间, 故定理 6.14, 式 (8.5) 成立当且仅当 $\mathbf{Ax}_* - \mathbf{b} \in \mathcal{R}(\mathbf{A})^\perp$, 再根据引理 8.18, $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T)$, 而方程组 (8.6) 正是 $\mathbf{Ax}_* - \mathbf{b} \in \mathcal{N}(\mathbf{A}^T)$ 的定义. 证毕.

注 8.3 法方程组 (8.6) 的解必定存在. 事实上, 根据定理 8.13 的性质 (5) 与 (8) 得

$$(\mathbf{A}^T \mathbf{A})(\mathbf{A}^T \mathbf{A})^+ (\mathbf{A}^T \mathbf{b}) = \mathbf{A}^T \mathbf{A} ((\mathbf{A}^T \mathbf{A})^+ \mathbf{A}^T) \mathbf{b} = \mathbf{A}^+ \mathbf{A} \mathbf{A}^+ \mathbf{b} = \mathbf{A}^T \mathbf{b},$$

故根据定理 8.17, 方程组 $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$ 的解存在. 从而, 定理 6.14 蕴含了最小二乘解的存在性.

从理论上讲, 可以利用定理 8.16 来计算方程组 $\mathbf{Ax} = \mathbf{b}$ 的最小二乘解, 但在实际计算时, 由于舍入误差的积累, $\mathbf{A}^T \mathbf{A}$ 的秩可能与理论上的秩不一致, 致使正规方程组无解, 或者所得解不是 $\mathbf{Ax} = \mathbf{b}$ 的最小二乘解. 下面的定理指出最小二乘解, 可以借助广义逆来计算, 从而克服了这一困难.

定理 8.21 x 是 $Ax = b$ 的最小二乘解的充分必要条件为 x 是 $Ax = AA^+b$ 的解, 且

$$\min_{x \in \mathbb{R}^n} \|b - Ax\| = \|(E - AA^+)b\|.$$

证明 由定理 8.17 易知方程 $Ax = AA^+b$ 恒有解. 以下证明, $\|b - Ax\|$ 的最小值为 $\|(E - AA^+)b\|$, 且最小值取到当且仅当 x 是 $Ax = AA^+b$ 的解, 从而定理结论成立. 事实上, 对任意 $x \in \mathbb{R}^n$,

$$b - Ax = (b - AA^+b) + (AA^+b - Ax) = (E - AA^+)b + (AA^+b - Ax),$$

其中, $AA^+b - Ax \in \mathcal{R}(A)$. 又因 $A^T = A^T AA^+$ (定理 8.13 性质 (9)), 故 $(E - AA^+)b \in \mathcal{N}(A^T)$. 从而由引理 8.18 及勾股定理 (定理 6.15) 得

$$\|b - Ax\|^2 = \|(E - AA^+)b\|^2 + \|AA^+b - Ax\|^2 \geq \|(E - AA^+)b\|^2,$$

等号成立当且仅当 $Ax = AA^+b$. 证毕.

定理 8.22 $Ax = b$ 的最小二乘解的通解为

$$x = A^+b + (E - A^+A)y,$$

其中 $y \in \mathbb{R}^n$ 为任意向量. 特别, $x_0 = A^+b$ 是线性方程组 $Ax = b$ 的唯一属于 $\mathcal{R}(A^T)$ 的最小二乘解, 且是长度最小的最小二乘解.

证明 显然 $x_0 = A^+b$ 满足方程组 $Ax = AA^+b$, 由定理 8.21 知 $x_0 = A^+b$ 是 $Ax = b$ 的一个最小二乘解. 由定理 8.17 知 $Ax = AA^+b$ 的通解为

$$x = A^+(AA^+b) + (E - A^+A)y = A^+b + (E - A^+A)y.$$

由定理 8.19 得到后续结论. 证毕.

综上所述, 向量 A^+b 给出了方程组 $Ax = b$ 各种意义下的解: 当 $Ax = b$ 有解时, A^+b 要么是它的唯一解, 要么是唯一的长度最小的最小二乘解; 当 $Ax = b$ 为矛盾方程组 (无解) 时, A^+b 要么是它的唯一的长度最小的最小二乘解, 要么是唯一的长度最小的最小二乘解. 因此, 广义逆 A^+ 无论在理论上, 还是在方法上, 都完美地解决了方程组 $Ax = b$ 的求解问题.

习题 8

1. 设 $A \in \mathbb{R}^{m \times n}$, A 的第 (i, j) -元素 a_{ij} 为 1, 其余元素全为零, 试问 $A\{1\}$ 是怎样的一类矩阵?
2. 设

$$A = \begin{pmatrix} 2 & 3 & 1 & -1 \\ 5 & 8 & 0 & 1 \\ 1 & 2 & -2 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 3 & -1 \\ 4 & 5 & 6 & 2 \\ 7 & 8 & 10 & 7 \\ 2 & 1 & 1 & 6 \end{pmatrix},$$

求 $A\{1\}$, $A\{1, 2\}$, $B\{1\}$, $B\{1, 2\}$.

3. 求下列矩阵的广义逆 A^+ .

$$(1) \mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & 0 \end{pmatrix}; \quad (2) \mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ 1 & 0 & 0 \\ -1 & -1 & 2 \\ -1 & 0 & 0 \end{pmatrix};$$

$$(3) \mathbf{A} = \begin{pmatrix} -2 & 0 & 0 & -2 \\ 1 & 2 & -4 & 3 \\ 2 & -1 & 2 & 1 \\ 0 & 2 & -4 & 2 \end{pmatrix}; \quad (4) \mathbf{A} = \begin{pmatrix} x_1 & 2x_1 & -x_1 \\ -x_2 & -2x_2 & x_2 \end{pmatrix}.$$

4. 证明: $\mathbf{AGA} = \mathbf{A}$ 的充要条件是 $\mathbf{A}^T \mathbf{AGA} = \mathbf{A}^T \mathbf{A}$.

5. 证明以下等式:

$$(1) \mathbf{A} = \mathbf{AA}^T(\mathbf{A}^+)^T = (\mathbf{A}^+)^T \mathbf{A}^T \mathbf{A},$$

$$(2) \mathbf{A}^T = \mathbf{A}^T \mathbf{AA}^+ = \mathbf{A}^+ \mathbf{AA}^T.$$

6. 用定义验证广义逆 \mathbf{A}^+ 的性质 (1).

7. 设 $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\text{rank } \mathbf{A} = 1$, 试证 $\mathbf{A}^+ = \frac{1}{\|\mathbf{A}\|_F^2} \mathbf{A}^T$, 其中

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}.$$

8. 若 \mathbf{A} 是正规矩阵, 试证 $\mathbf{A}^+ \mathbf{A} = \mathbf{AA}^+$.

9. 验证线性方程组 $\mathbf{Ax} = \mathbf{b}$ 有解, 并求其通解和最小长度解, 其中 $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 2 & 4 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix}$.

10. 验证下列线性方程组 $\mathbf{Ax} = \mathbf{b}$ 为矛盾方程组, 求其最小二乘解的通解和最小长度最小二乘解.

$$(1) \mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 0 \\ 2 & 4 \end{pmatrix} \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad (2) \mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ -3 & -6 & 3 \end{pmatrix} \mathbf{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad (3) \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 3 \end{pmatrix} \mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

11. 当 $\mathbf{Ax} = \mathbf{b}$ 为矛盾方程组 (无解) 时, 研究在何种条件下, 该方程组存在唯一的最小二乘解. 给出充分必要条件.

12. 若对复矩阵考虑广义逆, 本章的定义、定理要如何修正?

第9章 特征值的估计

矩阵的特征值和特征向量不仅在数学上非常重要的理论意义，在物理学、图像处理、网络搜索引擎和信息压缩传感中也具有非常广泛的实用价值。但是，计算大型矩阵的全部特征值的代价非常昂贵。能否在计算之前利用矩阵元素或者矩阵性质对矩阵特征值有一个比较好的估计和判断？这正是本章的主要目的。

为了比较长度，人们定义了尺、寸和米等长度标准；为了比较面积，人们定义了亩、平方米和公顷等面积标准。为了比较向量和矩阵的大小，需要引入一个标准来衡量向量和矩阵大小，这就是范数。本章首先介绍向量范数和矩阵范数的概念，然后给出矩阵特征值和矩阵元素之间的一些关系和性质，介绍估计矩阵特征值的一些基本工具，譬如 Rayleigh 商和圆盘定理等，最后给出谱半径的一些估计。

9.1 向量的范数

本节首先给出向量范数的定义。

定义 9.1 设映射 $\|\cdot\|: \mathbf{C}^n \rightarrow \mathbf{R}$, $\mathbf{x} \mapsto \|\mathbf{x}\|$, 满足以下条件:

- (1) 正定性: 对任何 $\mathbf{x} \in \mathbf{C}^n$, $\|\mathbf{x}\| \geq 0$, 等号当且仅当 $\mathbf{x} = \mathbf{0}$ 时成立;
- (2) 齐次性: 对任何 $k \in \mathbf{C}$, $\mathbf{x} \in \mathbf{C}^n$ 都有 $\|k\mathbf{x}\| = |k| \cdot \|\mathbf{x}\|$;
- (3) 三角不等式: 对任何 $\mathbf{x}, \mathbf{y} \in \mathbf{C}^n$ 都有 $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$,

则称 $\|\cdot\|$ 为 \mathbf{C}^n 上的一个范数。

注 9.1 在内积空间一章中，定理 6.2 和定理 6.19 分别指出，由实内积或复内积都可以诱导定义范数，除此之外，还有一些其他的范数。不同于内积诱导的范数，后面给出的一些范数，不一定满足平行四边形恒等式。

例 9.1 设 $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)^T$ 是 \mathbf{C}^n 的任意一个向量，定义

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^H \mathbf{x}} = \sqrt{\sum_{i=1}^n |x_i|^2},$$

则 $\|\mathbf{x}\|$ 是由 \mathbf{C}^n 上的内积诱导的范数（详见内积空间一章，复内积与酉空间一节），称为向量的 2-范数，记为 $\|\mathbf{x}\|_2$ 。譬如， $\|(3 \ 4i)^T\|_2 = \sqrt{3^2 + 4^2} = 5$ 。当 \mathbf{x} 为实向量时， $\|\mathbf{x}\|_2$ 就是欧几里得空间里的向量长度。

例 9.2 设 $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)^T$ 是 \mathbf{C}^n 的任意一个向量, 定义

$$\|\mathbf{x}\| = \max_i |x_i|,$$

则 $\|\mathbf{x}\|$ 也是 \mathbf{C}^n 中的一种范数, 称为 ∞ -范数, 记为 $\|\mathbf{x}\|_\infty$.

证明 (1)(2) 易证. 对 (3), 设 $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)^T$, $\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_n)^T$, 则

$$\|\mathbf{x} + \mathbf{y}\| = \max_i |x_i + y_i| \leq \max_i |x_i| + \max_i |y_i| = \|\mathbf{x}\| + \|\mathbf{y}\|.$$

证毕.

例 9.3 设 $\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_n)^T$ 是一个任意一个向量, 定义

$$\|\mathbf{x}\| = \sum_{i=1}^n |x_i|.$$

则 $\|\mathbf{x}\|$ 也是 \mathbf{C}^n 中的一种向量范数, 称为 1-范数, 记作 $\|\mathbf{x}\|_1$.

证明 (1)(2) 显然成立. 因为

$$\|\mathbf{x} + \mathbf{y}\| = \sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n |a_i| + \sum_{i=1}^n |b_i| = \|\mathbf{x}\| + \|\mathbf{y}\|,$$

故条件 (3) 也满足. 证毕.

例 9.4 设 $\mathbf{x} = (1 \ 2 \ 2 \ 4)^T$ 是 \mathbf{R}^4 的一个向量, 计算 $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_\infty$.

解 直接计算, 有

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| = 1 + 2 + 2 + 4 = 9,$$

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2} = \sqrt{1^2 + 2^2 + 2^2 + 4^2} = 5,$$

$$\|\mathbf{x}\|_\infty = \max_i |x_i| = \max\{1 \ 2 \ 2 \ 4\} = 4.$$

向量范数除上述三种外, 还有其他的定义方式. 可以验证, 对任何 $p > 1$,

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}},$$

都是 \mathbf{C}^n 中的范数, 称为 p -范数. 另外, 当 \mathbf{A} 是正定矩阵时, $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{A} \mathbf{x}}$ 也是一种向量范数, 通常称为能量范数, 记作 $\|\mathbf{x}\|_{\mathbf{A}}$.

正如长度单位米、尺和寸之间的转换关系一样, 向量的范数也有类似的转换关系. 譬如, 对任意 $\mathbf{x} \in \mathbf{C}^n$, 有

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n} \|\mathbf{x}\|_\infty;$$

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n} \|\mathbf{x}\|_2;$$

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n \|\mathbf{x}\|_\infty.$$

上述不等式称为嵌入不等式, 凡满足嵌入不等式的两个向量范数称为是等价的. 下面的定理指出, \mathbf{C}^n 上任何两种不同的向量范数, 都是等价的.

定理 9.1 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 \mathbf{C}^n 上任意两种范数, 则存在正数 $C_2 > C_1 > 0$, 使对任意 $\mathbf{x} \in \mathbf{C}^n$ 都有

$$C_1 \|\mathbf{x}\|_\beta \leq \|\mathbf{x}\|_\alpha \leq C_2 \|\mathbf{x}\|_\beta. \quad (9.1)$$

当定理中的范数为上述例子中的范数时, 可按如下方式给出简单的证明.

证明 考虑函数 $f(\mathbf{x}) = \frac{\|\mathbf{x}\|_\alpha}{\|\mathbf{x}\|_\beta}$, 则该函数为连续函数, 且在在有界闭集 $S_\alpha = \{\mathbf{x} : \|\mathbf{x}\|_\alpha = 1\}$

(称为 $\|\cdot\|$ 下的单位球面) 上存在最大值和最小值, 分别记为 C_1 和 C_2 , 则不等式 (9.1) 在 S_α 上成立. 而对一般的非零向量 \mathbf{x} , 记 $\lambda = \|\mathbf{x}\|_\alpha^{-1}$, 则 $\lambda\mathbf{x} \in S_\alpha$, 于是

$$C_1 \|\lambda\mathbf{x}\|_\beta \leq \|\lambda\mathbf{x}\|_\alpha \leq C_2 \|\lambda\mathbf{x}\|_\beta.$$

利用范数的齐次性可将上式中的 λ 约去, 即得结论. 证毕.

对一般的范数, 上述定理的证明涉及到在一般空间中连续函数的定义及性质, 远远超出本课程内容, 这里不再展开.

9.2 矩阵的范数

定义 9.2 设映射 $\|\cdot\| : \mathbf{C}^{n \times n} \rightarrow \mathbf{R}$, $A \mapsto \|A\|$, 满足以下条件:

- (1) 正定性: $\|A\| \geq 0$, 等号当且仅当 $A = \mathbf{0}$;
- (2) 齐次性: 任给 $k \in \mathbf{C}$, $A \in \mathbf{C}^{n \times n}$, 都有 $\|kA\| = |k| \cdot \|A\|$;
- (3) 三角不等式: 任给 $A, B \in \mathbf{C}^{n \times n}$, 都有 $\|A + B\| \leq \|A\| + \|B\|$;
- (4) 任给矩阵 $A, B \in \mathbf{C}^{n \times n}$, 都有 $\|AB\| \leq \|A\| \cdot \|B\|$,

则称该映射是矩阵 A 的一个范数, 记作 $\|A\|$.

与向量范数定义相比较, 前三个性质只是向量范数定义的推广, 而第四个性质则是和矩阵乘法相容性的要求. 与向量范数类似, 在 $\mathbf{C}^{n \times n}$ 中同样可规定多种方阵范数.

定理 9.2 设方阵 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$ 是任意矩阵, 定义

$$\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2},$$

则 $\|A\|$ 是 $\mathbf{C}^{n \times n}$ 中的一种矩阵范数, 称为 **Frobenius 范数** (或 **F-范数**), 记为 $\|A\|_F$.

证明 容易证明 $\|A\|_F$ 满足定义 9.2 中的 (1) (2) 和 (3), 以下验证 (4). 记 A 的第 i 个行向量为 α_i , 任取列矩阵 \mathbf{x} , 有

$$A\mathbf{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \mathbf{x} = \begin{pmatrix} \alpha_1 \mathbf{x} \\ \alpha_2 \mathbf{x} \\ \vdots \\ \alpha_n \mathbf{x} \end{pmatrix},$$

所以

$$\|A\mathbf{x}\|^2 = \sum_{k=1}^n |\alpha_k \mathbf{x}|^2 \leq \sum_{k=1}^n \|\alpha_k\|_2^2 \|\mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2 \|A\|_F^2.$$

记矩阵 B 的第 j 列为 β_j , 则 $AB = (A\beta_1 \ A\beta_2 \ \cdots \ A\beta_n)$, 于是

$$\|AB\|_F^2 = \sum_{j=1}^n \|A\beta_j\|_2^2 \leq \sum_{j=1}^n \|A\|_F^2 \|\beta_j\|_2^2 = \|A\|_F^2 \sum_{j=1}^n \|\beta_j\|_2^2 = \|A\|_F^2 \|B\|_F^2.$$

故 $\|A\|_F$ 满足定义 9.2 中的 (4). 证毕.

由于在大多数与特征值和误差估计有关的问题中, 矩阵和向量会同时参与讨论, 譬如矩阵向量乘积, 为了在向量范数与方阵范数之间建立某种联系, 引入如下定义.

定理 9.3 任给方阵 $A \in \mathbb{C}^{n \times n}$, 向量 $\mathbf{x} \in \mathbb{C}^n$, 若向量范数 $\|\mathbf{x}\|$ 与方阵范数 $\|A\|$ 之间满足不等式

$$\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|,$$

则称此方阵范数 $\|A\|$ 与向量范数 $\|\mathbf{x}\|$ 相容.

由定理 9.2 的证明可知, $\|A\|_F$ 与 $\|\mathbf{x}\|_2$ 是相容的.

为后续研究需要, 再引进一种与向量范数相容的矩阵范数.

定理 9.4 任给方阵 $A \in \mathbb{C}^{n \times n}$, 向量 $\mathbf{x} \in \mathbb{C}^n$, 记

$$\|A\|_p = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p, \quad (9.2)$$

则 $\|A\|_p$ 是一种矩阵范数, 且与向量范数 $\|\mathbf{x}\|_p$ 相容, 称为由向量范数 $\|\mathbf{x}\|_p$ 诱导的矩阵范数, 也称为诱导范数或算子范数.

证明 由于 $\|A\mathbf{x}\|_p$ 是关于 \mathbf{x} 的连续函数, $S_p = \{\mathbf{x} : \|\mathbf{x}\|_p = 1\}$ 是有界闭集, 故对任何矩阵 A , 定义式 (9.2) 中的最大值必定可以取到. 对任意的非零向量 $\mathbf{x} \in \mathbb{C}^n$, 令 $\lambda^{-1} = \|\mathbf{x}\|_p$, 则 $\|\lambda\mathbf{x}\|_p = 1$, 从而

$$\|A\mathbf{x}\|_p = \|\lambda^{-1} A(\lambda\mathbf{x})\|_p = \|\mathbf{x}\|_p \|A(\lambda\mathbf{x})\|_p \leq \|\mathbf{x}\|_p \|A\|_p,$$

即 $\|A\|_p$ 与向量范数 $\|\mathbf{x}\|_p$ 相容.

以下证明 $\|A\|_p$ 满足矩阵范数的 4 个条件.

(1) 当 $A \neq \mathbf{0}$ 时候, 必可找到 $\mathbf{y} \neq \mathbf{0}$ 使 $A\mathbf{y} \neq \mathbf{0}$, 故

$$\|A\|_p = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \geq \frac{\|A\mathbf{y}\|_p}{\|\mathbf{y}\|_p} > 0.$$

(2) 对任意的数 $k \in \mathbb{C}$, 有

$$\|kA\|_p = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|kA\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = |k| \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = |k| \cdot \|A\|_p.$$

(3) 对任意的 $A, B \in \mathbb{C}^{n \times n}$, 必有 $\mathbf{z}_0 \in \mathbb{C}^n$, 且 $\|\mathbf{z}_0\|_p = 1$ 使得

$$\|A + B\|_p = \|(A + B)\mathbf{z}_0\|_p \leq \|A\mathbf{z}_0\|_p + \|B\mathbf{z}_0\|_p \leq \|A\|_p + \|B\|_p.$$

(4) 对任意的 $A, B \in \mathbb{C}^{n \times n}$, 必有 $\mathbf{z}_0 \in \mathbb{C}^n$, 且 $\|\mathbf{z}_0\|_p = 1$ 使

$$\|AB\|_p = \|AB\mathbf{z}_0\|_p \leq \|A\|_p \cdot \|B\mathbf{z}_0\|_p \leq \|A\|_p \|B\|_p.$$

证毕.

定义式 (9.2) 中, 诱导范数是以函数最大值的形式给出的, 不便直接计算. 以下给出由向量范数 $\|\mathbf{x}\|_\infty$, $\|\mathbf{x}\|_1$ 和 $\|\mathbf{x}\|_2$ 分别诱导的矩阵范数, 关于矩阵元素的计算表达式.

定理 9.5 设方阵 $\mathbf{A} = (a_{ij}) \in \mathbf{C}^{n \times n}$, $\|\mathbf{A}\|_\infty$ 是向量范数 $\|\mathbf{x}\|_\infty$ 诱导的矩阵范数, 则 $\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$, 即 $\|\mathbf{A}\|_\infty$ 恰为 \mathbf{A} 的各行元素绝对值之和中的最大值.

证明 一方面, 根据定义有

$$\|\mathbf{A}\|_\infty = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|_\infty = \max_{\|\mathbf{x}\|=1} \max_i \sum_{j=1}^n |a_{ij}x_j| \leq \max_i \sum_{j=1}^n |a_{ij}|.$$

另一方面, 设

$$\sum_{j=1}^n |a_{kj}| = \max_i \sum_{j=1}^n |a_{ij}|,$$

取 $\mathbf{y} = (\text{sign}(a_{k1}) \text{sign}(a_{k2}) \cdots \text{sign}(a_{kn}))^T$, 则 $\|\mathbf{y}\|_\infty = 1$, 且

$$\|\mathbf{A}\|_\infty = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|_\infty \geq \|\mathbf{A}\mathbf{y}\|_\infty = \sum_{j=1}^n |a_{kj}| = \max_i \sum_{j=1}^n |a_{ij}|.$$

证毕.

定理 9.6 设方阵 $\mathbf{A} = (a_{ij}) \in \mathbf{C}^{n \times n}$, $\|\mathbf{A}\|_1$ 是向量范数 $\|\mathbf{x}\|_1$ 诱导的矩阵范数, 则 $\|\mathbf{A}\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$, 即 $\|\mathbf{A}\|_1$ 恰为 \mathbf{A} 的各列元素绝对值之和中的最大值.

证明 一方面, 设 $\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_n)^T$, 满足 $\|\mathbf{y}\|_1 = 1$ 且 $\max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|_1 = \|\mathbf{A}\mathbf{y}\|_1$, 则

$$\begin{aligned} \|\mathbf{A}\|_1 &= \|\mathbf{A}\mathbf{y}\|_1 = \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij}y_j \right| \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \cdot |y_j| \\ &= \sum_{j=1}^n \sum_{i=1}^n |a_{ij}| \cdot |y_j| = \sum_{j=1}^n |y_j| \sum_{i=1}^n |a_{ij}| \leq \max_i \sum_{j=1}^n |a_{ij}|. \end{aligned}$$

另一方面, 取单位向量 \mathbf{e}_j (见定义 1.12), 则 $\|\mathbf{e}_j\|_1 = 1$, 且

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|_1 \geq \|\mathbf{A}\mathbf{e}_j\|_1 = \sum_{i=1}^n |a_{ij}|,$$

故 $\|\mathbf{A}\|_1 \geq \max_j \sum_{i=1}^n |a_{ij}|$. 证毕.

定义 9.3 设方阵 $\mathbf{A} = (a_{ij}) \in \mathbf{C}^{n \times n}$ 是任意矩阵, 称 $\rho(\mathbf{A}) = \max_i |\lambda_i|$ 为 \mathbf{A} 的谱半径.

定理 9.7 设方阵 $\mathbf{A} = (a_{ij}) \in \mathbf{C}^{n \times n}$, $\|\mathbf{A}\|_2$ 是向量范数 $\|\mathbf{x}\|_2$ 诱导的矩阵范数, 则 $\|\mathbf{A}\|_2 = \sqrt{\rho(\mathbf{A}^H \mathbf{A})}$.

证明 因为 $0 \leq \|Ax\|_2^2 = (Ax, Ax) = (A^H Ax, x)$ 对一切 $x \in \mathbf{C}^n$ 成立, 所以 $A^H A$ 是 Hermite 半正定阵, 其特征值为非负实数, 故可设 $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$ 为 $A^H A$ 的 n 个特征值, x_1, x_2, \dots, x_n 为相应的两两正交的单位特征向量, 从而构成 \mathbf{C}^n 的一组标准正交基.

以下证明 $\|A\|_2 = \sqrt{\lambda_1}$. 一方面,

$$\|A\|_2^2 = \max_{\|x\|_2=1} \|Ax\|_2^2 \geq (A^H Ax_1, x_1) = \lambda_1(x_1, x_1) = \lambda_1.$$

故 $\|A\|_2 \geq \sqrt{\lambda_1}$. 另一方面, 对任何单位向量 x , x 可表示为

$$x = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n,$$

这里 $c_1^2 + c_2^2 + \cdots + c_n^2 = 1$. 从而

$$\begin{aligned} \|Ax\|_2^2 &= (A^H Ax, x) = \left(\sum_{j=1}^n c_j A^H Ax_j, \sum_{j=1}^n c_j x_j \right) = \left(\sum_{j=1}^n \lambda_j c_j x_j, \sum_{j=1}^n c_j x_j \right) \\ &= \lambda_1 c_1^2 + \lambda_2 c_2^2 + \cdots + \lambda_n c_n^2 \leq \lambda_1 c_1^2 + \lambda_1 c_2^2 + \cdots + \lambda_1 c_n^2 = \lambda_1, \end{aligned}$$

于是 $\|A\|_2 \leq \sqrt{\lambda_1}$. 证毕.

例 9.5 设 $A = \begin{pmatrix} 2 & -2 \\ -1 & 4 \end{pmatrix}$, 计算 A 的各种范数.

解 根据定理 9.2 得 $\|A\|_F = 5$. 根据定理 9.6 和定理 9.5 得 $\|A\|_1 = 6$. 再由 $A^H A = \begin{pmatrix} 5 & -8 \\ -8 & 20 \end{pmatrix}$ 的特征值为 $\frac{25}{2} \pm \frac{\sqrt{481}}{2}$, 于是 $\|A\|_2 \approx 4.84$.

矩阵 A 的 2-范数是 $A^T A$ 的最大特征值, 计算起来工作量比较大, 为了计算方便, 通常使用另一种与向量范数 $\|x\|_2$ 相容的矩阵范数 $\|A\|_F$.

注 9.2 并非所有的矩阵范数都是某个向量范数诱导的范数 (即算子范数). 可以证明, F-范数不是算子范数. 后续章节讨论特征值和误差分析时用到的范数都是算子范数, 且与它对应的向量范数也是相容的.

定理 9.8 如果 $\|B\| < 1$, 则 $E \pm B$ 为非奇异矩阵, 且

$$\|E \pm B\| \leq \frac{1}{1 - \|B\|},$$

其中 $\|\cdot\|$ 是矩阵的算子范数.

证明 注意到, 单位矩阵的算子范数均为 1, 这是因为

$$\|E\|_p = \max_{x \neq 0} \frac{\|Ex\|_p}{\|x\|_p} = 1.$$

以下用反证法证明 $E \pm B$ 的非奇异性. 若 $|E - B| = 0$, 则 $(E - B)x = 0$ 有非零解, 即存在 y , 使得 $By = y$, 于是 $\frac{\|By\|}{\|y\|} = 1$, 进而 $\|B\| \geq 1$, 与条件 $\|B\| < 1$ 矛盾.

又由 $(E - B)(E - B)^{-1} = E$, 得

$$(E - B)^{-1} = E + B(E - B)^{-1},$$

从而

$$\|(E - B)^{-1}\| \leq \|E\| + \|B\| \cdot \|(E - B)^{-1}\|,$$

即

$$\|(E - B)^{-1}\| \leq \frac{1}{1 - \|B\|}.$$

证毕.

与向量的等价范数定理 9.1 类似, 矩阵范数也有相应的等价范数定理, 证明略去.

定理 9.9 对于 $\mathbf{C}^{n \times n}$ 中任意两种矩阵范数 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$, 必存在 $C_2 > C_1 > 0$, 使对一切 $A \in \mathbf{C}^{n \times n}$, 有 $C_1 \|A\|_\alpha \leq \|A\|_\beta \leq C_2 \|A\|_\alpha$.

9.3 特征值与 Rayleigh 商

对一般阶数的矩阵 A , 计算其特征值不是一件容易的事. 本节将介绍的定理给出了特征值与矩阵的关系.

在第一章方阵的特征值一节就给出了下面的定理.

定理 9.10 设方阵 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$ 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则

$$(1) \lambda_1 + \lambda_2 + \dots + \lambda_n = \operatorname{tr} A;$$

$$(2) \lambda_1 \lambda_2 \cdots \lambda_n = \det A.$$

对特征值取绝对值后, 相应的和有下列的估计.

定理 9.11 设方阵 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$ 的特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 则

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2,$$

其中, 等号当且仅当 A 为正规矩阵时成立.

证明 由推论 6.30, A 酉相似于上三角阵 $R = (r_{ij})$, 即存在酉矩阵 U , 使得 $U^H A U = R$, 从而 $U^H A A^H U = R R^H$, 即 $A A^H$ 与 $R R^H$ 酉相似, 特别, 它们的特征值相同. 由定理 9.10, 有

$$\operatorname{tr} A A^H = \operatorname{tr} R R^H,$$

于是 R 的主对角元就是 A 的特征值, 因此

$$\sum_{i=1}^n |\lambda_i|^2 = \sum_{i=1}^n |r_{ii}|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |r_{ij}|^2 = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2.$$

上述不等式中的等号当且仅当 $r_{ij} = 0$ ($i \neq j$) 时成立, 此时矩阵 A 酉相似于对角阵, 即 A 为正规矩阵. 证毕.

由定理 9.11 的证明过程可得下面的推论.

推论 9.12 设方阵 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$ 的全体特征值为 $\lambda_1, \lambda_2, \dots, \lambda_n$, 且 $|\lambda_1| \geq |\lambda_2| \geq \dots \geq$

$|\lambda_n|$, 则

$$n|\lambda_n|^2 \leq \sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \leq n^2 \max_{1 \leq i, j \leq n} |a_{ij}|^2.$$

特征值与矩阵的关系还可以通过 Rayleigh 商来刻画.

定义 9.4 设 A 是 n 阶实对称矩阵, $\mathbf{x} \in \mathbf{R}^n$ 且 $\mathbf{x} \neq \mathbf{0}$, 则实数

$$f(\mathbf{x}) = \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

称为矩阵 A 的 **Rayleigh 商**.

利用内积的符号, Rayleigh 商又可写为

$$f(\mathbf{x}) = \frac{(A\mathbf{x}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}.$$

显然, 如果 \mathbf{x} 是 A 的特征向量, 则 $f(\mathbf{x})$ 就是对应的特征值. 下面讨论如何利用 Rayleigh 商来估计 A 的特征值的范围.

定理 9.13 设 A 是 n 阶实对称矩阵, λ_1 是 A 的最大特征值, λ_n 是 A 的最小特征值, 则 $\lambda_n \leq f(\mathbf{x}) \leq \lambda_1$.

证明 设 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 为 A 的 n 个特征值, $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ 为相应的两两正交的单位特征向量, 从而构成 \mathbf{R}^n 的一组标准正交基. 对任何 $\mathbf{x} \in \mathbf{R}^n$, \mathbf{x} 可表示为

$$\mathbf{x} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \cdots + c_n \mathbf{x}_n,$$

从而

$$(A\mathbf{x}, \mathbf{x}) = \left(\sum_{j=1}^n c_j A\mathbf{x}_j, \sum_{j=1}^n c_j \mathbf{x}_j \right) = \left(\sum_{j=1}^n \lambda_j c_j \mathbf{x}_j, \sum_{j=1}^n c_j \mathbf{x}_j \right) = \sum_{j=1}^n \lambda_j c_j^2,$$

$$(\mathbf{x}, \mathbf{x}) = \left(\sum_{j=1}^n c_j \mathbf{x}_j, \sum_{j=1}^n c_j \mathbf{x}_j \right) = \left(\sum_{j=1}^n c_j \mathbf{x}_j, \sum_{j=1}^n c_j \mathbf{x}_j \right) = \sum_{j=1}^n c_j^2.$$

再根据 $\lambda_n \sum_{j=1}^n c_j^2 \leq \sum_{j=1}^n \lambda_j c_j^2 \leq \lambda_1 \sum_{j=1}^n c_j^2$, 即得结论. 证毕.

设 $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{x} \neq \mathbf{0}$, $k \neq 0$, 则

$$f(k\mathbf{x}) = \frac{(k\mathbf{x})^T A(k\mathbf{x})}{(k\mathbf{x})^T (k\mathbf{x})} = \frac{k^2 \mathbf{x}^T A \mathbf{x}}{k^2 \mathbf{x}^T \mathbf{x}} = \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = f(\mathbf{x}).$$

这表明矩阵 A 的 Rayleigh 商对于 \mathbf{x} 和 $k\mathbf{x}$ 是相同的. 因此下面只讨论 Rayleigh 商在单位球面 $\|\mathbf{x}\|_2 = 1$ 上的情形.

定理 9.14 设 A 是 n 阶实对称矩阵, λ_1 是 A 的最大特征值, 则

$$\lambda_1 = \max_{\|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x}).$$

证明 由定理 9.13 知 $\lambda_1 \geq \max_{\|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x})$. 在定理 9.13 的证明中取 $\mathbf{x} = \mathbf{x}_1$, 得 $(A\mathbf{x}, \mathbf{x}) = \lambda_1$,

故 $\lambda_1 \leq \max_{\|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x})$. 证毕.

定理 9.15 设 A 是 n 阶实对称矩阵, λ_n 是 A 的最小特征值, 则

$$\lambda_n = \min_{\|x\|=1} (Ax, x).$$

与定理 9.14 的证明类似, 只须在定理 9.13 的证明中取 $x = x_n$ 即可.

定理 9.16 设 A 是 n 阶实对称矩阵, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ 为 A 的 n 个特征值, x_1, x_2, \cdots, x_n 为相应的两两正交的单位特征向量, 记 $W_k = \text{span}\{x_1, \cdots, x_{k-1}\}^\perp$, 则 $\lambda_k = \max_{x \in W_k, \|x\|_2=1} (Ax, x)$,

当且仅当 x 是属于 λ_k 的单位特征向量时达到最大值.

证明 当 x 是属于 λ_k 的单位特征向量时, $(Ax, x) = \lambda_k$ 是明显的. 注意到, 当 $x \in W_k$ 时, $x = l_k x_k + \cdots + l_n x_n$, 从而

$$(Ax, x) = \sum_{j=k}^n \lambda_j c_j^2 \leq \lambda_k \sum_{j=k}^n c_j^2,$$

等号成立当且仅当 $\lambda_j < \lambda_k$ 时, $c_j = 0$. 证毕.

定理 9.17 设矩阵 A 和 B 的最大特征值分别为 $\lambda_{\max}(A)$ 和 $\lambda_{\max}(B)$, 最小特征值分别为 $\lambda_{\min}(A)$ 和 $\lambda_{\min}(B)$, 矩阵 $A+B$ 的特征值为 $\lambda_i(A+B)$, $i = 1, 2, \cdots, n$, 则 $\lambda_{\min}(A) + \lambda_{\min}(B) \leq \lambda_i(A+B) \leq \lambda_{\max}(A) + \lambda_{\max}(B)$.

证明 设对应于 $\lambda_i(A+B)$ 的特征向量是 x_i , 则有

$$\begin{aligned} \lambda_i(A+B) &= \frac{x_i^T (A+B)x_i}{x_i^T x_i} = \frac{x_i^T A x_i}{x_i^T x_i} + \frac{x_i^T B x_i}{x_i^T x_i} \\ &\leq \max_{x \neq 0} \frac{x^T A x}{x^T x} + \max_{x \neq 0} \frac{x^T B x}{x^T x} = \lambda_{\max}(A) + \lambda_{\max}(B). \end{aligned}$$

同理可证 $\lambda_i(A+B) \geq \lambda_{\min}(A) + \lambda_{\min}(B)$. 证毕.

9.4 圆盘定理

设矩阵 $A \in \mathbb{C}^{n \times n}$, 于是矩阵 A 有 n 个特征值, 它们是复平面上的 n 个点. 本节将介绍的圆盘定理指出了这 n 个点所处的大致位置.

定理 9.18 (圆盘定理) 设 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$, 则 A 的特征值 λ 位于复平面上诸圆盘

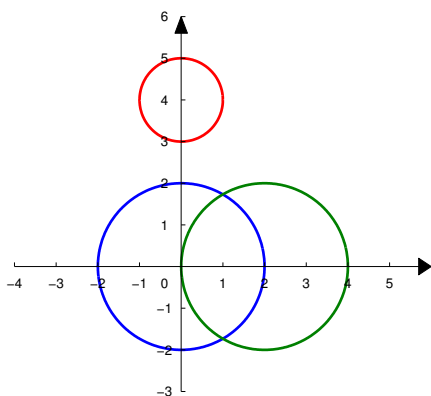
$$D_i = \left\{ \lambda : |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| = \sum_{j=1}^n |a_{ij}| - |a_{ii}| \right\} \quad (i = 1, 2, \cdots, n)$$

的并集.

若记 $\rho_i = \sum_{j \neq i} |a_{ij}|$, 则上述圆盘 D_i 是复平面上以 a_{ii} 为圆心、以 ρ_i 为半径的圆盘, 称为

矩阵 A 的 **Gail 圆盘**. 圆盘定理表明矩阵 A 的任何一个特征值均在 n 个 Gail 圆盘的某一个之中.

证明 设矩阵 A 对应于特征值 λ 的特征向量为 $x = (x_1 \ x_2 \ \cdots \ x_n)^T$. 若 $|x_i| = \max_j |x_j| > 0$,

图 9.1: 矩阵 A 的三个 Gail 圆盘

则由 $Ax = \lambda x$ 可推得 $\sum_{j=1}^n a_{ij}x_j = \lambda x_i$, 即 $(\lambda - a_{ii})x_i = \sum_{j \neq i}^n a_{ij}x_j$. 于是

$$|\lambda - a_{ii}| \cdot |x_i| = \left| \sum_{j \neq i} a_{ij}x_j \right| \leq \sum_{j \neq i} |a_{ij}| \cdot |x_j| \leq \rho_i |x_i|.$$

从而 $|\lambda - a_{ii}| \leq \sum_{j \neq i}^n |a_{ij}|$, 即 $\lambda \in D_i (i = 1, 2, \dots, n)$. 证毕.

注 9.3 对角阵的每一个 Gail 圆盘都退化为一个点.

例 9.6 设 $A = \begin{pmatrix} 0 & 1 & -1 \\ 1.3 & 2 & -0.7 \\ 0.5 & 0.5i & 4i \end{pmatrix}$, 求 A 的 Gail 圆盘.

解 直接计算得 $\rho_1 = 1 + 1 = 2$, $\rho_2 = 1.3 + 0.7 = 2$, $\rho_3 = 0.5 + 0.5 = 1$, 故 A 的三个 Gail 圆盘分别为

$$|\lambda| \leq 2, |\lambda - 2| \leq 2, |\lambda - 4i| \leq 1,$$

如图 9.1 所示, 由圆盘定理知 A 的特征值必在这三个圆盘之中.

圆盘定理给出了 A 的特征值的一个估计, 但它的不足是没有指出特征值位于哪一个圆盘中, 也没有说是否每一个圆盘中必有一个特征值. 下面的定理对这类问题作了进一步的讨论.

定理 9.19 设 G_1, \dots, G_k 是 $\bigcup_{j=1}^n D_j$ 的连通分支全体, 其中 D_j 是 A 的 Gail 圆盘, 连通

分支 G_i 由 n_i 个 Gail 圆盘的并集构成的连通域 ($i = 1, 2, \dots, k$). 则在每个 D_i 中恰有 n_i 个特征值.

证明 令 $B = \text{diag}\{a_{11}, a_{22}, \dots, a_{nn}\}$, $C = A - B$. 考虑矩阵函数 $A(t) = B + tC$, 则 $A(0) = B$, $A(1) = A$. 对任何 t , $A(t)$ 的 Gail 圆盘圆心不变 (均为 A 的对角元), 当 t 从 0 连续变到 1 时, $A(t)$ 的每个 Gail 圆都从一个点连续变大到 A 的 Gail 圆盘. 可以证明, $A(t)$ 的每一个特征值 $\lambda_j(t)$ 同样连续依赖于 t , 因此 $\lambda_j(t)$ 的轨迹为 D_i 内的一条连续曲线. 证毕.

推论 9.20 若方阵 $A \in \mathbf{C}^{n \times n}$ 的 n 个 Gail 圆盘互不相交, 则 A 相似于对角阵.

证明 此时每个 Gail 圆盘内恰有一个 A 的特征值, 于是 A 的特征值互不相同. 证毕.

推论 9.21 若方阵 $A \in \mathbf{R}^{n \times n}$ 的 n 个 Gail 圆盘互不相交, 则 A 的特征值均为实数.

证明 A 的特征多项式是实系数多项式, 故若有复根必共轭成对出现. 由于实方阵 A 的 n 个 Gail 圆盘圆心都在实轴上, 且若有一复特征值位于某一 Gail 圆盘内, 则其共轭的复特征值也必位于同一 Gail 圆盘中, 与定理 9.19 矛盾. 证毕.

定义 9.5 设矩阵 $A = (a_{ij}) \in \mathbf{R}^{n \times n}$, 如果对每一个 i , 都有 $|a_{ii}| > \rho_i$, 则称矩阵 A 为严格对角占优矩阵.

定理 9.22 若 A 是严格对角占优矩阵, 且 A 的所有主对角元为正实数, 则 A 的所有特征值具有正实部.

证明 设 $\lambda = a + bi$ 为 A 的特征值, 由圆盘定理, 至少存在一个 k , 使

$$|\lambda - a_{kk}| = |a + bi - a_{kk}| \leq \rho_k < a_{kk},$$

于是 $|a - a_{kk}| \leq \rho_k < a_{kk}$, 因此 $a > 0$. 证毕.

在矩阵 A 的特征值估计中, 谱半径十分重要. 只要估计了谱半径的界, 也就相当于估计了特征值的上界. 下面是一些关于特殊矩阵谱半径的定理.

定理 9.23 设矩阵 $A \in \mathbf{C}^{n \times n}$, $\rho(A) \leq \|A\|$.

证明 设 λ 为特征值, x 为相应的特征向量, 则 $|\lambda| \cdot \|x\| = \|\lambda x\| = \|Ax\| \leq \|A\| \cdot \|x\|$, 故 $|\lambda| \leq \|A\|$, 由 λ 的任意性, 知 $\rho(A) \leq \|A\|$. 证毕.

推论 9.24 设 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$, 则 $\rho(A) \leq \min\{\|A\|_1, \|A\|_\infty\}$.

定理 9.25 设 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$, b_1, b_2, \dots, b_n 是 n 个非零正实数, 令

$$q_1 = \max_i \sum_{j=1}^n \frac{b_j |a_{ij}|}{b_i}, \quad q_2 = \max_j \sum_{i=1}^n \frac{b_j |a_{ij}|}{b_i},$$

则 $\rho(A) \leq \min\{q_1, q_2\}$.

证明 令矩阵 $B = \text{diag}\{b_1, b_2, \dots, b_n\}$, 则 B 是可逆矩阵, 且

$$B^{-1}AB = \begin{pmatrix} a_{11} & \frac{b_2 a_{12}}{b_1} & \cdots & \frac{b_n a_{1n}}{b_1} \\ \frac{b_1 a_{21}}{b_2} & a_{22} & \cdots & \frac{b_n a_{2n}}{b_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b_1 a_{n1}}{b_n} & \frac{b_2 a_{n2}}{b_n} & \cdots & a_{nn} \end{pmatrix}.$$

因为相似矩阵有相同的特征值，从而有相同的谱半径，而

$$\|B^{-1}AB\|_{\infty} = q_1, \quad \|B^{-1}AB\|_1 = q_2,$$

从而 $\rho(A) = \rho(B^{-1}AB) \leq \min\{q_1, q_2\}$. 证毕.

定理 9.26 设矩阵 A 是 n 阶正规矩阵，则 $\rho(A) = \|A\|_2$.

证明 由正规矩阵的谱分解 (定义 7.4)，有 $A = U\Lambda U^H$ ，其中 U 为酉矩阵， $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ ，从而

$$A^H A = U \text{diag}(|\lambda_1|^2, |\lambda_2|^2, \dots, |\lambda_n|^2) U^H.$$

于是

$$\rho(A) = \max_j |\lambda_j|, \quad \rho(A^T A) = \max_j |\lambda_j|^2 = (\rho(A))^2,$$

结合定理 9.7 得到 $\rho(A) = \sqrt{\rho(A^H A)} = \sqrt{\|A\|_2^2} = \|A\|_2$. 证毕.

定理 9.27 (Frobenius) 设方阵 $A = (a_{ij}) \in \mathbf{R}^{n \times n}$ 是非负矩阵 (即对所有的 i, j 均有 $a_{ij} \geq 0$ ，且不等号至少对一个元素成立)，记

$$r = \min_i \sum_{j=1}^n a_{ij}, \quad R = \max_i \sum_{j=1}^n a_{ij},$$

则 $r \leq \rho(A) \leq R$.

Frobenius 定理不仅给出了非负矩阵谱半径的上界估计，还给出了下界的估计，证明较繁，从略.

习题 9

1. 设向量 $x \in \mathbf{R}^n$ ，求证

$$\|x\|_{\infty} \leq \|x\|_1 \leq n\|x\|_{\infty},$$

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2,$$

$$\|x\|_{\infty} \leq \|x\|_2 \leq \sqrt{n}\|x\|_{\infty}.$$

2. 已知 $A = \begin{pmatrix} -1 & 0 & 2 \\ 3 & 5 & -1 \\ 7 & 1 & 2 \end{pmatrix}$ ，计算 $\|A\|_1$ ， $\|A\|_2$ ， $\|A\|_F$ 和 $\|A\|_{\infty}$.

3. 设矩阵 $A \in \mathbf{C}^{n \times n}$ ，为酉矩阵，试证明 $\|A\|_2 = 1$ ， $\|A\|_F = \sqrt{n}$.

4. 设向量 $A \in \mathbf{R}^{n \times n}$ ，求证

$$(1) \frac{1}{n}\|A\|_{\infty} \leq \|A\|_1 \leq n\|A\|_{\infty}; \quad (2) \frac{1}{\sqrt{n}}\|A\|_{\infty} \leq \|A\|_2 \leq \sqrt{n}\|A\|_{\infty}.$$

5. 用圆盘定理估计矩阵 $A = \begin{pmatrix} i & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.2 & 0.4 \\ 1 & 0.4 & -1 & 0.1 \\ 0.3 & -0.6 & 0.1 & -2 \end{pmatrix}$ 的特征值分布范围, 并且在复平面上

作图表示.

6. 证明矩阵 $A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$ 的谱半径 $\rho(A) < 1$.

7. 证明矩阵 $A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$ 的谱半径 $\rho(A) = 1$.

8. 设 P 是正交矩阵, $A = \text{diag}\{a_1, a_2, \dots, a_n\}$ 为实对角矩阵, $m = \min_{1 \leq i \leq n} |a_i|$, $M = \max_{1 \leq i \leq n} |a_i|$.

试证明 PA 的特征值 μ 满足不等式 $m \leq |\mu| \leq M$.

第 10 章 张量

在力学和物理学的很多领域（例如连续介质力学，相对论等），广泛存在着一类比向量更加广泛的物理量——张量。张量不依赖于坐标系的选取，在不同坐标系下的表示形式与向量有类似之处。这些概念的数学本质，是有限维线性空间上的多重线性函数。本章将简要介绍张量的基本概念和运算。

10.1 张量的物理描述

在力学和物理学中，很多物理量是定义在某个 n 维线性空间 V 上的。在 V 上的物理量中，最简单的是标量，标量是一个与坐标系选取无关的常数，例如功、能量、频率、电流强度等。

比标量稍许复杂的物理量是向量，例如力、位移、速度等。向量同样是一个独立于坐标系存在的一个客观的量，但在不同坐标系下，同一个向量可能有不同的表达式。设 \mathbf{v} 为 V 上的向量， $\mathbf{e}_1, \dots, \mathbf{e}_n$ 为 V 上的一组基， \mathbf{v} 在这组基下的坐标记为 $(x^1, \dots, x^n)^T$ （在本章中，用上标来记坐标分量，下面会看到这种记法的优点），则

$$\mathbf{v} = \sum_{i=1}^n x^i \mathbf{e}_i.$$

再设 $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ 为 V 上的另一组基， \mathbf{v} 在这组基下的坐标记为 $(\tilde{x}^1, \dots, \tilde{x}^n)^T$ ，于是

$$\mathbf{v} = \sum_{j=1}^n \tilde{x}^j \tilde{\mathbf{e}}_j.$$

设 (α_j^i) 为 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 到 $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ 的过渡矩阵（上标表示过渡矩阵的行指标，下标表示列指标），于是

$$\mathbf{v} = \sum_{j=1}^n \tilde{x}^j \tilde{\mathbf{e}}_j = \sum_{j=1}^n \tilde{x}^j \alpha_j^i \mathbf{e}_i. \quad (10.1)$$

对比式 (10.1) 和式 (10.1) 中 \mathbf{e}_i 的系数，有

$$x^i = \sum_{j=1}^n \tilde{x}^j \alpha_j^i. \quad (10.2)$$

将矩阵的行指标写为上标的方式, 与线性代数的表示习惯不一致, 容易记错, 但借助微积分中导数的记号, 可以将过渡矩阵中的元素改写为

$$\alpha_j^i = \frac{\partial x^i}{\partial \tilde{x}^j},$$

此时过渡矩阵的行列指标刚好分处分子和分母上, 代替了上下标的效果, 此时式 (10.2) 可以改写为

$$x^i = \sum_{j=1}^n \tilde{x}^j \frac{\partial x^i}{\partial \tilde{x}^j}. \quad (10.3)$$

根据复合函数求导的链式法则, 有

$$\delta_j^i = \sum_{k=1}^n \frac{\partial x^i}{\partial \tilde{x}^k} \frac{\partial \tilde{x}^k}{\partial x^j},$$

(这里 δ_j^i 为 Kronecker 符号), 亦即 $(\frac{\partial \tilde{x}^i}{\partial x^j})$ 为 $(\frac{\partial x^i}{\partial \tilde{x}^j})$ 的逆矩阵, $(\frac{\partial \tilde{x}^i}{\partial x^j})$ 为 $\tilde{e}_1, \dots, \tilde{e}_n$ 到 e_1, \dots, e_n 的过渡矩阵. 本章中的过渡矩阵均采用偏导数形式的记法.

例 10.1 向量可以看成具备下述特点的一个物理量:

- (1) 在每一个给定的坐标系下, 向量对应了 n 个分量构成的数组;
- (2) 在不同坐标系下, 同一个向量对应的分量之间满足式 (10.3) 中的关系.

这种描述方法在一定程度上避开代数的概念, 深受早期的物理学家欢迎. 随着物理学的发展, 物理学家又陆续发现了一些新的物理量, 具备与向量类似的特点, 如下面两个例子.

例 10.2 设 G 为 V 上的一个欧氏度量, 在基 e_1, \dots, e_n 和基 $\tilde{e}_1, \dots, \tilde{e}_n$ 下对应的度量矩阵分别为 (g_{ij}) 和 (\tilde{g}_{ij}) , 这两个矩阵元素之间满足

$$\tilde{g}_{ij} = \sum_{k=1}^n \sum_{l=1}^n g_{kl} \frac{\partial x^k}{\partial \tilde{x}^i} \frac{\partial x^l}{\partial \tilde{x}^j} \quad (10.4)$$

按照物理学家的观点, V 上的欧氏度量 G 可以看成这样一个物理量:

- (1) 在每一个给定的坐标系下, G 对应了 $n \times n$ 个分量构成的数组;
- (2) 在不同坐标系下, 同一个 G 对应的 $n \times n$ 个分量之间满足式 (10.4) 中的关系.

例 10.3 设 \mathcal{A} 为 V 上的一个线性变换, 在基 e_1, \dots, e_n 和基 $\tilde{e}_1, \dots, \tilde{e}_n$ 下对应的矩阵分别为 (a_j^i) 和 (\tilde{a}_j^i) , 这两个矩阵元素之间满足

$$\tilde{a}_j^i = \sum_{k=1}^n \sum_{l=1}^n a_l^k \frac{\partial \tilde{x}^i}{\partial x^k} \frac{\partial x^l}{\partial \tilde{x}^j} \quad (10.5)$$

按照物理学家的观点, 线性变换 \mathcal{A} 可以看成这样一个物理量:

- (1) 在每一个给定的坐标系下, \mathcal{A} 对应了一个 $n \times n$ 个分量构成的数组;
- (2) 在不同坐标系下, 同一个 \mathcal{A} 对应的 $n \times n$ 个分量之间满足式 (10.5) 中的关系.

对比上述三个不同的物理量, 有两个细微的不同: 其一是 v 对应的数组只有一个上指标, G 对应数组有两个下指标, 而 \mathcal{A} 对应的数组有一个上指标和一个下指标; 其二是数组满足的关系式中, 有的出现的是过渡矩阵中的元素, 有的出现的是过渡矩阵逆阵中的元素,

有的过渡矩阵和其逆阵中的元素同时出现. 仔细检查这二个差别, 可以发现这个看似不易记忆的关系中, 存在这样一个规律:

(1) 在每一个求和式中, 重复出现的指标必定是求和指标;

(2) 求和指标恰好在上下指标各出现一次 (这里, 将在分母上的指标看作下指标);

(3) 重复出现的指标必定出现在同一个坐标系下的物理量上 (即在上述求和式中, 相同指标的物理量要么都有波浪号, 要么都没有波浪号).

物理学家在这种情况下, 经常省略求和记号, 即约定凡是重复出现的上下指标, 默认对它们由 1 到 n 进行求和, 这种记法称为 **Einstein 求和约定**, 在这一章的后续部分, 均将采用这一记法.

将上述观点推广到一般情形, 就得到张量的概念.

定义 10.1 设 T 为 n 维线性空间 V 上的一个物理量, 满足以下两个条件:

(1) 对 V 上任给的一组基 e_1, \dots, e_n , 存在唯一与 T 对应的一个多维数组 $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$, 这里所有的上下指标均取遍 $1, \dots, n$ 中的整数;

(2) 对 V 的任意两组基 e_1, \dots, e_n 和 $\tilde{e}_1, \dots, \tilde{e}_n$, $(x^1, \dots, x^n)^T$ 和 $(\tilde{x}^1, \dots, \tilde{x}^n)^T$ 分别为这两组基下的坐标分量, T 对应的多维数组 $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$ 和 $(\tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r})$ 满足

$$\tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r} = t_{l_1 \dots l_s}^{k_1 \dots k_r} \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}}, \quad (10.6)$$

则称 T 或 $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$ 为 V 上的一个 (r, s) -型张量, 数组 $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$ 称为 T 在基 e_1, \dots, e_n 下的分量, r 称为 T 的逆变阶数, s 称为 T 的协变阶数. 特别的, 标量也称为一个 $(0, 0)$ -型张量.

根据定义, 立即得到, n 维线性空间上的一个 V 上的 (r, s) -型张量, 有 n^{r+s} 个分量. 特别, V 上的向量是一个 1 阶逆变张量, V 上的度量是一个 2 阶协变张量, V 上的线性变换是一个 $(1, 1)$ -型张量, 从这个意义上说, 标量、向量、欧氏度量、线性变换都是特殊的张量.

这里提到了张量在给定基下的分量, 这与向量在给定基下的坐标有一定的相似之处, 将在张量的代数描述中给出这两者之间的联系.

从这一定义看, 要构造线性空间上的一个张量, 需要给出其在每组基下对应的分量, 然而一个线性空间有无穷多组基, 不可能通过穷举每一组基上的分量来构造一个张量, 下面的定理告诉, 只需在任何给定的一组基下指定一个分量, 就可以构造一个张量.

定理 10.1 设 e_1, \dots, e_n 为线性空间 V 上给定的一组基, $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$ 是一个给定的数组, $\tilde{e}_1, \dots, \tilde{e}_n$ 为 V 的任意一组基, $(x^1, \dots, x^n)^T$ 和 $(\tilde{x}^1, \dots, \tilde{x}^n)^T$ 分别为这两组基的坐标分量, 则数组

$$\tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r} = t_{l_1 \dots l_s}^{k_1 \dots k_r} \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}}$$

定义了 V 上的一个张量 T .

证明 由定义方式可知, 从给定基 e_1, \dots, e_n 到任意一组基 $\tilde{e}_1, \dots, \tilde{e}_n$, 数组 $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$ 与 $(\tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r})$ 之间, 满足式 (10.6), 因此只需证对任意两组基 $\tilde{e}_1, \dots, \tilde{e}_n$ 和 $\hat{e}_1, \dots, \hat{e}_n$, 式 (10.6) 成立即可. 记 T 在这两组基下的分量分别为 $(\tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r})$ 和 $(\hat{t}_{j_1 \dots j_s}^{i_1 \dots i_r})$, 这两组基下的坐标分量分别

为 $(\bar{x}^1, \dots, \bar{x}^n)^T$ 和 $(\bar{x}^1, \dots, \bar{x}^n)^T$ ，于是由定义

$$\begin{aligned}\tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r} &= t_{l_1 \dots l_s}^{k_1 \dots k_r} \frac{\partial \bar{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \bar{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \bar{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \bar{x}^{j_s}}, \\ \tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r} &= t_{l_1 \dots l_s}^{k_1 \dots k_r} \frac{\partial \bar{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \bar{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \bar{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \bar{x}^{j_s}},\end{aligned}$$

从而

$$t_{j_1 \dots j_s}^{i_1 \dots i_r} = \tilde{t}_{l_1 \dots l_s}^{k_1 \dots k_r} \frac{\partial x^{i_1}}{\partial \bar{x}^{k_1}} \cdots \frac{\partial x^{i_r}}{\partial \bar{x}^{k_r}} \frac{\partial \bar{x}^{l_1}}{\partial x^{j_1}} \cdots \frac{\partial \bar{x}^{l_s}}{\partial x^{j_s}},$$

于是

$$\begin{aligned}\tilde{t}_{j_1 \dots j_s}^{i_1 \dots i_r} &= \tilde{t}_{l'_1 \dots l'_s}^{k'_1 \dots k'_r} \frac{\partial x^{k_1}}{\partial \bar{x}^{k'_1}} \cdots \frac{\partial x^{k_r}}{\partial \bar{x}^{k'_r}} \frac{\partial \bar{x}^{l'_1}}{\partial x^{l_1}} \cdots \frac{\partial \bar{x}^{l'_s}}{\partial x^{l_s}} \frac{\partial \bar{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \bar{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \bar{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \bar{x}^{j_s}} \\ &= \tilde{t}_{l'_1 \dots l'_s}^{k'_1 \dots k'_r} \frac{\partial \bar{x}^{i_1}}{\partial \bar{x}^{k'_1}} \cdots \frac{\partial \bar{x}^{i_r}}{\partial \bar{x}^{k'_r}} \frac{\partial \bar{x}^{l'_1}}{\partial \bar{x}^{j_1}} \cdots \frac{\partial \bar{x}^{l'_s}}{\partial \bar{x}^{j_s}},\end{aligned}$$

证毕.

上述定理揭示了张量和向量之间的一个相似之处：描述一个向量，只需选择一个坐标系，写出它的坐标就可以了，而无需给出其在所有坐标系下的坐标，对于张量，同样只要在某一个坐标系下写出其所有分量就够了。

注 10.1 的确存在一些物理量，虽然在任一坐标系下也对应了一个多维数组，但不同坐标系对应的数组之间并不满足式 (10.6)，这样的量就不是张量，因此不能将多维数组简单等同于张量，只有满足定义 10.1 的多维数组才是张量。

10.2 张量的运算

张量之间可以定义很多种运算，在定义张量运算之前，先要给出两个张量相等的定义。

定义 10.2 线性空间 V 上的两个张量 \mathbf{A} 和 \mathbf{B} 称为相等的，如果在任意一组基下，其相同指标的分量都分别相等。

由定理 10.1，证明张量相等，只需在其中一组基下验证分量相等即可。

同种类型的张量之间可以定义加法，张量和数之间可以定义数乘。

定理 10.2 设 \mathbf{A} 和 \mathbf{B} 同为线性空间 V 上的 (r, s) -型张量，在基 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 下的分量分别为 $(a_{j_1 \dots j_s}^{i_1 \dots i_r})$ 和 $(b_{j_1 \dots j_s}^{i_1 \dots i_r})$ ，则数组 $(a_{j_1 \dots j_s}^{i_1 \dots i_r} + b_{j_1 \dots j_s}^{i_1 \dots i_r})$ 和 $(k \cdot a_{j_1 \dots j_s}^{i_1 \dots i_r})$ 也定义了 V 上的一个 (r, s) -型张量（这里 k 为常数），分别记为 $\mathbf{A} + \mathbf{B}$ 和 $k\mathbf{A}$ ，其在 V 的任何一组基下的分量也恰为 \mathbf{A} 与 \mathbf{B} 在该组基下分量的和与数乘。

证明 任取 V 的一组基 $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ ，并设 \mathbf{A} 和 \mathbf{B} 在这组基下的分量为 $(\tilde{a}_{j_1 \dots j_s}^{i_1 \dots i_r})$ 和 $(\tilde{b}_{j_1 \dots j_s}^{i_1 \dots i_r})$ ，

根据张量的定义,

$$\begin{aligned}\tilde{a}_{j_1 \cdots j_s}^{i_1 \cdots i_r} &= a_{l_1 \cdots l_s}^{k_1 \cdots k_r} \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}}, \\ \tilde{b}_{j_1 \cdots j_s}^{i_1 \cdots i_r} &= b_{l_1 \cdots l_s}^{k_1 \cdots k_r} \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}},\end{aligned}$$

于是

$$\begin{aligned}\tilde{a}_{j_1 \cdots j_s}^{i_1 \cdots i_r} + \tilde{b}_{j_1 \cdots j_s}^{i_1 \cdots i_r} &= (a_{l_1 \cdots l_s}^{k_1 \cdots k_r} + b_{l_1 \cdots l_s}^{k_1 \cdots k_r}) \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}}, \\ k\tilde{a}_{j_1 \cdots j_s}^{i_1 \cdots i_r} &= ka_{l_1 \cdots l_s}^{k_1 \cdots k_r} \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}}.\end{aligned}$$

证毕.

推论 10.3 n 维线性空间 V 上全体 (r, s) -型张量在上述加法、数乘运算下, 构成 n^{r+s} 维线性空间, 记作 $T^{r,s}(V)$.

$T^{r,s}(V)$ 是线性空间是定理 10.2 的直接推论, 关于维数的证明将放在下节.

同一线性空间上, 任意两个张量可以作张量积.

定理 10.4 设 A 和 B 分别为 V 上的 (r_1, s_1) -型和 (r_2, s_2) -型张量, 其在基 $\tilde{e}_1, \cdots, \tilde{e}_n$ 下的分量分别为 $(a_{j_1 \cdots j_{s_1}}^{i_1 \cdots i_{r_1}})$ 和 $(b_{j_{s_1+1} \cdots j_{s_1+s_2}}^{i_{r_1+1} \cdots i_{r_1+r_2}})$, 则 $(a_{j_1 \cdots j_{s_1}}^{i_1 \cdots i_{r_1}} \cdot b_{j_{s_1+1} \cdots j_{s_1+s_2}}^{i_{r_1+1} \cdots i_{r_1+r_2}})$ 定义了 V 上的一个 (r_1+r_2, s_1+s_2) -型张量, 称为 A 与 B 的张量积, 记为 $A \otimes B$, 其在任何一组基下的分量也恰为 A 与 B 在该组基下分量的乘积.

证明 任取 V 的一组基 $\tilde{e}_1, \cdots, \tilde{e}_n$, 并设 A 和 B 在这组基下的分量为 $(a_{j_1 \cdots j_{s_1}}^{i_1 \cdots i_{r_1}})$ 和 $(b_{j_{s_1+1} \cdots j_{s_1+s_2}}^{i_{r_1+1} \cdots i_{r_1+r_2}})$, 根据张量的定义,

$$\begin{aligned}\tilde{a}_{j_1 \cdots j_{s_1}}^{i_1 \cdots i_{r_1}} &= a_{l_1 \cdots l_{s_1}}^{k_1 \cdots k_{r_1}} \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_{r_1}}}{\partial x^{k_{r_1}}} \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_{s_1}}}{\partial \tilde{x}^{j_{s_1}}}, \\ \tilde{b}_{j_{s_1+1} \cdots j_{s_1+s_2}}^{i_{r_1+1} \cdots i_{r_1+r_2}} &= b_{l_{s_1+1} \cdots l_{s_1+s_2}}^{k_{r_1+1} \cdots k_{r_1+r_2}} \frac{\partial \tilde{x}^{i_{r_1+1}}}{\partial x^{k_{r_1+1}}} \cdots \frac{\partial \tilde{x}^{i_{r_1+r_2}}}{\partial x^{k_{r_1+r_2}}} \frac{\partial x^{l_{s_1+1}}}{\partial \tilde{x}^{j_{s_1+1}}} \cdots \frac{\partial x^{l_{s_1+s_2}}}{\partial \tilde{x}^{j_{s_1+s_2}}},\end{aligned}$$

从而

$$\begin{aligned}\tilde{a}_{j_1 \cdots j_{s_1}}^{i_1 \cdots i_{r_1}} \tilde{b}_{j_{s_1+1} \cdots j_{s_1+s_2}}^{i_{r_1+1} \cdots i_{r_1+r_2}} &= a_{l_1 \cdots l_{s_1}}^{k_1 \cdots k_{r_1}} b_{l_{s_1+1} \cdots l_{s_1+s_2}}^{k_{r_1+1} \cdots k_{r_1+r_2}} \frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_{r_1}}}{\partial x^{k_{r_1}}} \frac{\partial \tilde{x}^{i_{r_1+1}}}{\partial x^{k_{r_1+1}}} \cdots \frac{\partial \tilde{x}^{i_{r_1+r_2}}}{\partial x^{k_{r_1+r_2}}} \\ &\quad \cdot \frac{\partial x^{l_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{l_{s_1}}}{\partial \tilde{x}^{j_{s_1}}} \frac{\partial x^{l_{s_1+1}}}{\partial \tilde{x}^{j_{s_1+1}}} \cdots \frac{\partial x^{l_{s_1+s_2}}}{\partial \tilde{x}^{j_{s_1+s_2}}}.\end{aligned}$$

证毕.

注 10.2 通过这两个定理, 发现, 张量线性运算与张量积的分量定义式, 与坐标系的选取无关, 即在任何一组坐标系下, 定义式的形式是相同的. 这是张量运算最根本的特点.

推论 10.5 张量积运算满足以下性质:

- (1) $k(A \otimes B) = (kA) \otimes B = A \otimes (kB)$, 这里 k 为常数;
- (2) 结合律: $(A \otimes B) \otimes C = A \otimes (B \otimes C)$;
- (3) 分配律: $(A + B) \otimes C = A \otimes C + B \otimes C$.

注 10.3 虽然 $A \otimes B$ 和 $B \otimes A$ 作为张量的类型相同, 其分量也是各自分量的乘积, 但两者通常并不相等, 即张量积的交换律不成立. 这是因为, 虽然分量乘积是可以交换的, 但交换后对应的张量分量指标却是不同的. 例如 A 和 B 都是 $(0, 2)$ -型张量, 其在某组基下的分量分别为 $a^1 = 1, a^2 = 2, b^1 = 1, b^2 = 3$, 记 $C = A \otimes B, D = B \otimes A$, 则根据定义, C 的四个分量分别为

$$c^{11} = a^1 b^1 = 1, c^{12} = a^1 b^2 = 3, c^{21} = a^2 b^1 = 2, c^{22} = a^2 b^2 = 6,$$

而 D 的四个分量分别为

$$d^{11} = b^1 a^1 = 1, d^{12} = b^1 a^2 = 2, d^{21} = b^2 a^1 = 3, d^{22} = b^2 a^2 = 6,$$

于是 $C \neq D$.

注 10.4 两个不为常数的张量的积, 类型必定不同于原先的两个张量, 因此这一定义与矩阵的乘法不同. 例如对两个线性变换, 其张量积是一个 $(2, 2)$ -型张量, 而其作为线性变换的乘积仍然是一个线性变换, 即 $(1, 1)$ -型张量.

对协变阶数和逆变阶数均不为零的张量, 还可以定义“缩并”运算.

定理 10.6 设 T 为线性空间 V 上的 (r, s) -型张量, 在基 e_1, \dots, e_n 下的分量为 $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$, 则 $(t_{m j_2 \dots j_s}^{m i_2 \dots i_r})$ 定义了 V 上的一个 $(r-1, s-1)$ -型张量 (这里对指标 m 采用 Einstein 求和约定), 记为 $C_{11}(T)$, 称为张量的缩并, 缩并运算的分量定义式与坐标系选取无关.

证明 只需证明 $C_{11}(T)$ 的两组分量间满足式 (10.6). 事实上,

$$\begin{aligned} \tilde{t}_{m j_2 \dots j_s}^{m i_2 \dots i_r} &= t_{l_1 l_2 \dots l_s}^{k_1 k_2 \dots k_r} \frac{\partial \tilde{x}^m}{\partial x^{k_1}} \frac{\partial \tilde{x}^{i_2}}{\partial x^{k_2}} \dots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_1}}{\partial \tilde{x}^m} \frac{\partial x^{l_2}}{\partial \tilde{x}^{j_2}} \dots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}} \\ &= t_{l_1 l_2 \dots l_s}^{k_1 k_2 \dots k_r} \delta_{k_1}^{l_1} \frac{\partial \tilde{x}^{i_2}}{\partial x^{k_2}} \dots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_2}}{\partial \tilde{x}^{j_2}} \dots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}} = t_{k_1 l_2 \dots l_s}^{k_1 k_2 \dots k_r} \frac{\partial \tilde{x}^{i_2}}{\partial x^{k_2}} \dots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}} \frac{\partial x^{l_2}}{\partial \tilde{x}^{j_2}} \dots \frac{\partial x^{l_s}}{\partial \tilde{x}^{j_s}}. \end{aligned}$$

证毕.

$C_{11}(T)$ 是通过对分量的第一个上标和第一个下标对等求和得到的, 同理, 对任意一对上下标对等求和, 也可以得到一个 $(r-1, s-1)$ -型张量, 记为 $C_{\mu\nu}(T)$. 总之, 张量分量的任何一对上下标都可以对等求和做缩并, 缩并降低了张量的阶数. 特别, V 上的一个线性变换可以看作一个 $(1, 1)$ -型张量, 对它缩并后得到一个数, 这个数恰好是线性变换的迹, 这的确是一个不依赖于坐标系的量.

对张量的一对上标 (或下标) 对等求和后, 不再是张量, 验证留作习题.

借助缩并运算, 可以定义张量指标的提升与降低. 首先证明下面的定理.

定理 10.7 设 G 是线性空间 V 上的一个 $(0, 2)$ -型张量, 在基 e_1, \dots, e_n 下的分量分别为 (g_{ij}) , 假定该分量作为 n 阶方阵是对称并可逆的, 于是 G 在所有基下的分量作为 n 阶方阵都是对称并可逆的, 且该组逆矩阵, 记为 (g^{ij}) , 定义了 V 上的一个 $(2, 0)$ -型张量.

证明 任取一组基 $\tilde{e}_1, \dots, \tilde{e}_n$, 设 G 在这组基下的分量为 (\tilde{g}_{ij}) , 根据张量的定义,

$$\tilde{g}_{ij} = g_{kl} \frac{\partial x^k}{\partial \tilde{x}^i} \frac{\partial x^l}{\partial \tilde{x}^j} = \tilde{g}_{ji}$$

即 (\tilde{g}_{ij}) 作为矩阵是对称的. 又因为

$$(\tilde{g}^{ij}) = (\tilde{g}_{ij})^{-1} = (g_{kl} \frac{\partial x^k}{\partial \tilde{x}^i} \frac{\partial x^l}{\partial \tilde{x}^j})^{-1} = [(\frac{\partial x^k}{\partial \tilde{x}^i})^T (g_{kl}) (\frac{\partial x^l}{\partial \tilde{x}^j})]^{-1} = (g^{kl} \frac{\partial \tilde{x}^i}{\partial x^k} \frac{\partial \tilde{x}^j}{\partial x^l}),$$

故 (\tilde{g}^{ij}) 定义了 V 上的一个 $(2,0)$ -型张量. 根据对称矩阵的逆矩阵必定对称, 知 (\tilde{g}^{ij}) 作为矩阵也是对称的. 证毕.

特别的, Euclid 空间上度量的逆是一个 $(2,0)$ -型张量.

定理 10.8 设 G 是 V 上的一个给定的 $(0,2)$ -型张量, 其在基 e_1, \dots, e_n 下的分量 (g_{ij}) 作为 n 阶方阵是对称并可逆的, 则对 V 上的任意 (r,s) -型张量 T , 在基 e_1, \dots, e_n 下的分量为 $(t_{j_1 \dots j_s}^{i_1 \dots i_r})$, 数组 $(g_{ik} t_{j_1 \dots j_s}^{ki_2 \dots i_r})$ 和 $(g^{ik} t_{kj_2 \dots j_s}^{i_1 \dots i_r})$ 分别定义了 V 上的一个 $(r-1, s+1)$ -型和 $(r+1, s-1)$ -型张量, 分别称为张量指标的**提升**与**下降**. 张量指标提升与下降的分量定义式, 与坐标系的选取无关.

该定理是定理 10.6 和定理 10.4 的直接推论.

同样道理, 也可以对排在后面的指标作提升或降低, 不再赘述. 在物理学中, 通常都是利用空间的度量矩阵对张量坐标进行提升和降低.

在具体应用中, 很多张量具备一定的对称性, 有下述定义.

定义 10.3 若 T 在 e_1, \dots, e_n 下的分量为 (t_{ij}) , 且满足 $t_{ij} = t_{ji}$, 则称 T 为 2 阶对称张量; 若 T 在 e_1, \dots, e_n 下的分量为 (t_{ij}) , 且满足 $t_{ij} = -t_{ji}$, 则称 T 为 2 阶反对称张量.

对多指标张量, 同样也有对称与反对称的概念.

定义 10.4 若 T 在 e_1, \dots, e_n 下的分量为 $(t_{i_1 \dots i_n})$, 且对任何 $1, \dots, n$ 的全排列 (i_i, \dots, i_n) , 有 $t_{i_1 \dots i_n} = t_{1 \dots n}$, 则称 T 为 n 阶对称张量; 若在同样前提条件下, 有 $t_{i_1 \dots i_n} = \text{sgn}(i_1, \dots, i_n) t_{1 \dots n}$, 则称 T 为 n 阶反对称张量, 这里 $\text{sgn}(i_1, \dots, i_n)$ 是排列的符号.

对于对称 (反对称) 张量, 有下面的定理.

定理 10.9 若 T 为线性空间 V 上的对称 (反对称) 张量, 则 T 在任何一组基下的分量都是对称 (反对称) 的.

证明 不妨只证明二阶对称张量的情形. 设 (t_{ij}) 为对称张量 T 在 (e_1, \dots, e_n) 下的分量, 满足 $t_{ij} = t_{ji}$. 任取 V 的一组基 $\tilde{e}_1, \dots, \tilde{e}_n$, 则 T 在这组基下的分量满足关系

$$\tilde{t}_{ij} = t_{kl} \frac{\partial x^k}{\partial \tilde{x}^i} \frac{\partial x^l}{\partial \tilde{x}^j},$$

于是 $\tilde{t}_{ji} = t_{kl} \frac{\partial x^k}{\partial \tilde{x}^j} \frac{\partial x^l}{\partial \tilde{x}^i} = t_{lk} \frac{\partial x^l}{\partial \tilde{x}^i} \frac{\partial x^k}{\partial \tilde{x}^j} = \tilde{t}_{ij}$. 证毕.

类似地, 还可以定义仅对部分指标为对称或反对称的张量, 这里不再赘述.

10.3 张量的代数描述

设 v 为 V 上的向量, e_1, \dots, e_n 为 V 上的一组基, v 在这组基下的坐标记为 $(x^1, \dots, x^n)^T$,

因

$$\mathbf{v} = \sum_{i=1}^n x^i \mathbf{e}_i,$$

于是称 x^i 是向量 \mathbf{v} 在 \mathbf{e}_i 上的分量. 对于张量, 在前两节中也提到了分量这一名词, 本节将解释这个分量是何种意义上的分量. 事实上, V 上的全体 (r, s) -型张量构成一个 n^{r+s} 维的线性空间, 上节中所提到的张量的分量, 就是在这一空间上某组基下的分量. 为了解释这一事实, 首先考虑线性空间 V 上的线性函数.

定义 10.5 设 V 是域 \mathbf{F} 上线性空间, $f: V \rightarrow \mathbf{F}$ 为定义在 V 上的函数, 并对任何 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v} \in V, k \in \mathbf{F}$, 满足

$$f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2), \quad f(k\mathbf{v}) = kf(\mathbf{v}),$$

则 f 称为 V 上的一个线性函数.

定理 10.10 n 维线性空间 V 上的线性函数全体关于函数的加法和数乘构成 n 维线性空间, 记为 V^* , 称为 V 的对偶空间.

证明 任取 V 上的向量 $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2$, V 上的线性函数 f, f_1, f_2 , 以及数 k, l , 因为

$$\begin{aligned} (f_1 + f_2)(\mathbf{v}_1 + \mathbf{v}_2) &= f_1(\mathbf{v}_1 + \mathbf{v}_2) + f_2(\mathbf{v}_1 + \mathbf{v}_2) \\ &= f_1(\mathbf{v}_1) + f_1(\mathbf{v}_2) + f_2(\mathbf{v}_1) + f_2(\mathbf{v}_2) = (f_1 + f_2)(\mathbf{v}_1) + (f_1 + f_2)(\mathbf{v}_2); \\ (f_1 + f_2)(k\mathbf{v}) &= f_1(k\mathbf{v}) + f_2(k\mathbf{v}) = kf_1(\mathbf{v}) + kf_2(\mathbf{v}) = k(f_1 + f_2)(\mathbf{v}); \\ (lf)(\mathbf{v}_1 + \mathbf{v}_2) &= l \cdot f(\mathbf{v}_1 + \mathbf{v}_2) = l(f(\mathbf{v}_1) + f(\mathbf{v}_2)) = (lf)(\mathbf{v}_1) + (lf)(\mathbf{v}_2); \\ (lf)(k\mathbf{v}) &= l \cdot f(k\mathbf{v}) = l \cdot kf(\mathbf{v}) = k(lf)(\mathbf{v}), \end{aligned}$$

故 $f_1 + f_2, lf$ 都是 V 上的线性函数, 从而 V 上的线性函数全体构成线性空间.

设 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 为 V 的一组基, 容易证明, V 上的线性函数 f 由 f 在基 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 上的取值唯一确定.

定义 V 上的一组线性函数 $\mathbf{e}^{*1}, \dots, \mathbf{e}^{*n}$ 为

$$\mathbf{e}^{*i}(\mathbf{e}_j) = \delta_j^i, \quad (10.7)$$

则 $f = f(\mathbf{e}_i)\mathbf{e}^{*i}$ (这里采用 Einstein 求和约定, 下同). 从而 $V^* = \text{span}\{\mathbf{e}^{*1}, \dots, \mathbf{e}^{*n}\}$. 另一方面, 若存在 $l_1, \dots, l_n \in \mathbf{F}$ 使得 $l_i \mathbf{e}^{*i} = 0$, 则

$$0 = l_i \mathbf{e}^{*i}(\mathbf{e}_j) = l_i \delta_j^i = l_j,$$

从而 $\mathbf{e}^{*1}, \dots, \mathbf{e}^{*n}$ 线性无关, 于是 $\mathbf{e}^{*1}, \dots, \mathbf{e}^{*n}$ 构成 V^* 的一组基. 证毕.

定义 10.6 设 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 为域 \mathbf{F} 上线性空间 V 的一组基, 如式 (10.7) 定义的 $\mathbf{e}^{*1}, \dots, \mathbf{e}^{*n}$ 称为 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 的对偶基, 于是对 V^* 中任一元素 f (即 V 上的任一线性函数), $(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))^T$ 为 f 在对偶基下的坐标.

下面指出, V^* 中的元素可与 V 上的 1 阶协变张量建立一一对应, 于是在这一对应意义下, V^* 中的元素可以等同于 V 上的 1 阶协变张量. 任取 $f \in V^*$ 以及 V 上的一组基 $\mathbf{e}_1, \dots, \mathbf{e}_n$, 构造对应

$$f \mapsto (f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))^T,$$

这一对应满足 1 阶协变张量的定义. 事实上, 一方面, 设 $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ 为 V 的另一组基, $(\frac{\partial x^i}{\partial \tilde{x}^j})$ 为 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 到 $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ 的过渡矩阵, 则有

$$f(\tilde{\mathbf{e}}_j) = f(\frac{\partial x^k}{\partial \tilde{x}^j} \mathbf{e}_k) = f(\mathbf{e}_k) \frac{\partial x^k}{\partial \tilde{x}^j},$$

即 $(f(\mathbf{e}_1), \dots, f(\mathbf{e}_n))^T$ 定义了一个 1 阶协变张量. 另一方面, 任取 V 上的一个 1 阶协变张量 \mathbf{T} , 设其在基 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 下的分量为 $(y_1 \cdots y_n)^T$, 构造对应

$$(y_1 \cdots y_n)^T \mapsto y_i \mathbf{e}^{*i},$$

需说明这一对应不依赖于基的选取. 设 $\tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_n$ 为 V 的另一组基, $\tilde{\mathbf{e}}^{*1}, \dots, \tilde{\mathbf{e}}^{*n}$ 为其对偶基, \mathbf{T} 在这组基下的分量为 $(\tilde{y}_1, \dots, \tilde{y}_n)^T$, 因

$$\tilde{y}_j = y_k \frac{\partial x^k}{\partial \tilde{x}^j},$$

从而对任意 i ,

$$(\tilde{y}_j \tilde{\mathbf{e}}^{*j})(\mathbf{e}_i) = y_k \frac{\partial x^k}{\partial \tilde{x}^j} \tilde{\mathbf{e}}^{*j}(\mathbf{e}_i) = y_k \frac{\partial x^k}{\partial \tilde{x}^j} \tilde{\mathbf{e}}^{*j}(\frac{\partial \tilde{x}^l}{\partial x^i} \tilde{\mathbf{e}}_l) = y_k \frac{\partial x^k}{\partial \tilde{x}^j} \frac{\partial \tilde{x}^l}{\partial x^i} \delta_l^j = y_k \frac{\partial x^k}{\partial \tilde{x}^j} \frac{\partial \tilde{x}^j}{\partial x^i} = y_k \delta_i^k = y_i,$$

即

$$\tilde{y}_j \tilde{\mathbf{e}}^{*j} = y_i \mathbf{e}^{*i}.$$

容易发现上面的两个对应是互逆的, 于是 V^* 中的元素可与 V 上的 1 阶协变张量建立一一对应.

进一步还可以发现, 一个 1 阶协变张量在 V 的基 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 下的分量, 其实就是其对应的线性函数, 作为线性空间 V^* 中的元素, 在对偶基 $\mathbf{e}^{*1}, \dots, \mathbf{e}^{*n}$ 下的坐标, 于是找到了 1 阶协变张量的分量和线性空间中元素的坐标之间的对应关系.

下面借助 V^* 中的元素, 构造一般的 $(0, s)$ -型张量. 考虑定义在 V 上的 s 元线性函数 $\mathbf{T}(\mathbf{v}_1, \dots, \mathbf{v}_s)$ (即 \mathbf{T} 每一组 $(\mathbf{v}_1, \dots, \mathbf{v}_s)$ 映射为一个数, 且 \mathbf{T} 关于每一个 \mathbf{v}_k 都是线性的), 首先在对任意给定的 V 的一组基 $\mathbf{e}_1, \dots, \mathbf{e}_n$, \mathbf{T} 可对应到一个数组 $(\mathbf{T}(\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_s}))$, 这里指标 j_1, \dots, j_s 均取遍 $1, \dots, n$ 中所有整数, 类似地可以证明, 这一对应满足 $(0, s)$ -型张量的定义. 反之对于任何一个 V 上的 $(0, s)$ -型张量, 设其在基 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 上的分量为 $(x_{j_1 \dots j_s})$, 定义 V 上的 s 元函数

$$\mathbf{T}(\mathbf{v}_1, \dots, \mathbf{v}_s) = x_{j_1 \dots j_s} \mathbf{e}^{*j_1}(\mathbf{v}_1) \cdots \mathbf{e}^{*j_s}(\mathbf{v}_s),$$

容易验证这是 V 上的一个 s 元线性函数, 并且这两个对应是互逆的. 于是 V 上的 $(0, s)$ -型张量可以等同于 V 上的一个 s 元线性函数. 特别地, 将上述 \mathbf{T} 记作

$$\mathbf{T} = x_{j_1 \dots j_s} \mathbf{e}^{*j_1} \otimes \cdots \otimes \mathbf{e}^{*j_s}.$$

对 V 上的 $(r, 0)$ -型张量, 可以借助 V 上的元素构造. 事实上 V 也可以看做 V^* 的对偶空间 (证明作为习题), 于是类似前面的构造, 可以将 V 上的 s 阶逆变张量等同于 V^* 上的 r 元线性函数.

对 V 上的 (r, s) -型张量, 则可以将其等同于一个定义在 V^* 和 V 上的 $(r+s)$ -元线性函数, 其中前 r 个自变量定义在 V^* 中, 后 s 个自变量定义在 V 中, 证明略去. 特别地, 若 V

上的 (r, s) -型张量在 V 上的分量为 $x_{i_1 \dots i_s}^{k_1 \dots k_r}$, 则相应的多重线性函数记为

$$T = x_{j_1 \dots j_s}^{i_1 \dots i_r} e_{i_1} \otimes \dots \otimes e_{i_r} \otimes e^{*j_1} \otimes \dots \otimes e^{*j_s}.$$

从这个表示可以看出, V 上的 (r, s) -型张量构成一个 n^{r+s} 维线性空间,

$$e_{i_1} \otimes \dots \otimes e_{i_r} \otimes e^{*j_1} \otimes \dots \otimes e^{*j_s} \quad (1 \leq i_1, \dots, i_r, j_1, \dots, j_s \leq n) \quad (10.8)$$

构成该空间的一组基, 张量 T 在 e_1, \dots, e_n 下的分量, 恰恰就是 T 作为线性空间中的元素, 在基式 (10.8) 下的坐标.

上一节提到对称张量. 可以看作一个对称的多重线性函数.

定理 10.11 设 T 为一个 V 上的 $(0, n)$ -型张量, 若对一切 $v_1, \dots, v_n \in V$, 以及任何 $1, \dots, n$ 的全排列 (i_1, \dots, i_n) , 都有

$$T(v_{i_1}, \dots, v_{i_n}) = T(v_1, \dots, v_n),$$

则 T 为一个 n 阶对称协变张量; 若在同样前提条件下, 有

$$T(v_{i_1}, \dots, v_{i_n}) = \operatorname{sgn}(i_1, \dots, i_n) T(v_1, \dots, v_n),$$

则 T 为一个 n 阶反对称协变张量, 这里 $\operatorname{sgn}(i_1, \dots, i_n)$ 是排列的符号.

习题 10

1. 设 $(g_{\mu\nu}) = \operatorname{diag}(-1, 1, 1, 1)$, 将 Einstein 求和式 $g_{\mu\nu} x^\mu x^\nu$ 展开为连加式.
2. 计算 $\delta_j^i A_{jk}$.
3. 计算 $A^{jm} \frac{\partial x^i}{\partial \bar{x}^k} \frac{\partial \bar{x}^k}{\partial x^j}$.
4. 证明域 \mathbf{F} 上的 n 维线性空间 V 上的线性函数的全体, 关于函数的加法和数乘, 构成 \mathbf{F} 上的 n 维线性空间.
5. 设 \mathbf{R}^3 上的线性函数 f 在基 e_1, e_2, e_3 下的表示为 $f(v) = ax + by + cz$, 其中 $(x \ y \ z)^T$ 为向量 v 在这组基下的坐标, 求 f 在对偶基 e^{*1}, e^{*2}, e^{*3} 下的坐标.
6. 设 $T = (t_{ij})$ 为 n 阶方阵, 定义 \mathbf{R}^n 上的双线性函数在基 e_1, e_2, \dots, e_n 下的形式为 $f(X, Y) = X^T T Y$, 求 f 在对偶基下的表示.
7. 设 V 是一个 n 维线性空间, \mathcal{A} 是 V 上的一个线性变换, f 是 V 上的任意一个线性函数 (即 $f \in V^*$), 对任何 $v \in V$, 定义 $g(v) = f(\mathcal{A}(v))$, 证明: g 也是 V 上的一个线性函数, 从而 \mathcal{A} 诱导了 V^* 上的一个线性变换 \mathcal{A}^* . 试找出 \mathcal{A} 在某组基下的矩阵与 \mathcal{A}^* 在对偶基下的矩阵之间的关系.
8. 设 V, W 是两个有限维线性空间, \mathcal{A} 是 V 到 W 的一个线性映射 (即映射 \mathcal{A} 保持加法和数乘运算), f 是 W 上的任意一个线性函数 (即 $f \in W^*$), 对任何 $v \in V$, 定义 $g(v) = f(\mathcal{A}(v))$, 证明: g 也是 V 上的一个线性函数, 从而 \mathcal{A} 诱导了 W^* 到 V^* 上的一个线性映射 \mathcal{A}^* .

9. 设 V 是一个有限维线性空间, V^* 是 V 的对偶空间, 证明: V 也是 V^* 的对偶空间.
10. 证明商定理: 设 A 与 P 为线性空间 V 上的两个物理量, 任取 V 的一组基, 设 A, P 在这组基下的分量表示分别为 (A^μ) 和 $(P_{\alpha\beta\gamma})$, 若 (A^μ) 和 $(A^\mu P_{\mu\beta\gamma})$ 都是张量, 则 $(P_{\alpha\beta\gamma})$ 也是一个张量.

参考文献

- [1] 范德瓦尔登. 代数学 I [M]. 北京: 科学出版社, 2015.
- [2] 孟道骥. 高等代数与解析几何 [M]. 北京: 科学出版社, 1998.
- [3] 许以超. 线性代数与矩阵论 [M], 第二版. 北京: 高等教育出版社, 2008.
- [4] 同济大学应用数学系. 矩阵分析 [M]. 上海: 同济大学出版社, 2005.
- [5] 周矜君、韩静、狄艳媚. 复变函数与积分变换 [M], 第二版. 上海: 同济大学出版社, 2020.
- [6] Leonid P Lebedev , Michael J Cloud , Victor A Eremeyev . Tensor Analysis with Applications in Mechanics [M] . World Scientific, 2010.