矩　阵　理　论　教　程

张跃辉

北京大学出版社
2024.7

内 容 提 要

本书共分六章，第一章线性代数概要，除了总结线性代数的主要知识外，增加了矩阵的张量积与求解矛盾方程的最小二乘法(LSM)；第二章线性空间与线性变换，讨论线性空间与线性变换的结构和构造方法，以及有限维线性空间与$\mathbb{F}^n$的关系，线性变换与矩阵的关系，特别强调线性空间的商和张量积以及内积空间与等距变换；第三章 Jordan 标准形与特征值问题，证明了 Schur 三角化定理，利用幂零矩阵给出了 Jordan 标准形定理的快速证明，讨论了特征值估计的盖尔圆盘定理，介绍了特征值与特征向量的一些应用；第四章矩阵分解与应用，介绍几种重要的矩阵分解方法与应用，特别强调了SVD的应用；第五章赋范线性空间与矩阵微积分，介绍了如何在线性空间中引入度量(范数)并由此导入矩阵微积分及其应用；第六章广义逆矩阵与非负矩阵，介绍几种常用的广义逆与非负矩阵的Perron-Froubenius 定理及其应用。

书后附有参考文献和汉英名词索引。

本书是为上海交通大学非数学类研究生写的通用教材，可作为高校理工科高年级本科生以及从事教学、科研等人员的参考用书。

前　言

我校矩阵理论课程使用的教材《矩阵理论与应用》[17]，至今已超过十三个年头了，本书即为更新版本。我校致力于服务全国更多高校，特将本书名定为《矩阵理论教程》。

本书主要参考了R.Horn与C.Johnson的《矩阵分析》([9]，感谢杨奇先生精彩的中译本)。该书内容丰富，深入浅出，是极好的参考书。另一本编者想极力推荐的是美国科学院院士G.Strang的《线性代数及其应用》([13]，感谢侯自新、郑中三、张廷伦等先生出色的中译本)，该书虽然名为“线性代数”，实则包含了国内很多高校为研究生开设的《矩阵理论》的大部分内容；更让编者由衷钦佩的是Strang亲自主讲的配套视频(共34讲，每讲一小时左右)，精彩绝伦。国内许多高校自编《矩阵理论》教材，包括[3, 4, 8, 19]等，也可参考。

对于非数学专业的读者来说, 矩阵理论的重要性更在于其与时俱进的广泛而深刻的应用, 编者特此推荐由三位著名计算机科学家A. Blum、J. Hopcroft、R. Kannan的名著Foundations of Data Science([2])。

国内特别值得推荐的是清华大学张贤达教授的《矩阵分析与应用》([15])，内容广泛，实例丰富，“包括了信号处理、电子、通信、模式识别、神经计算、雷达、图像处理、系统辨识等信息科学与技术的不同学科与领域”。被作者自己定义为“一本矩阵手册”。

另外还有我国已故著名数学家柯召翻译的苏联数学家F. R. Gantmacher(甘特马赫尔)的经典著作《矩阵论》[5](原书为俄文版，英译本分[6]与[7])，美国著名数学家P. Lax的《线性代数及其应用》[11](傅莺莺，沈复兴译)和美国著名数学教育家D.Lay的同名教材[12](刘深泉等译)，已故英国著名数学家、计算机学家J. H. Wilkinson(威尔金森)的名著《代数特征值问题》[14](石钟慈，邓健新译，毛祖范校)，以及张谋成、黎稳的《非负矩阵论》[16]。编者要特别推荐的是一本尚无中译本的好书，即已故美国著名数学家Richard Bellman的“Introduction To Matrix Analysis”[1]，该书包含一个非常具有启发意义的较长前言，值得认真研读。

作者感谢我校研究生院和数学科学学院对本书的大力资助，特别感谢章璞教授的长期支持和鼓励。

编者水平有限，谬误与不当之处必定不少，敬请批评、指正。来信请寄:

200240上海交通大学数学科学学院　张跃辉
邮件：zyh@sjtu.edu.cn

编者2023.12

本 书 导 读

本书需要的预备知识除了高等数学的基本理论外，对线性代数的基础知识有较多要求，读者可参考蒋启芬、马俊编写的《线性代数》([10])。

《矩阵理论》无疑是非数学类研究生最好的数学基础课。首先，《矩阵理论》等于研究生的《线性代数》+《高等数学》，是升华数学能力的基础；其次，《矩阵理论》提供解决大量实际问题的理论框架和思想方法；再次，复杂实际问题的数学模型都是高维的，其最终解决方案必然是线性化，而矩阵是解决高维线性问题的最佳方法；最后，由于人工智能的本质是高维数据的处理+误差估计, 因此矩阵理论是人工智能的核心。

矩阵理论的灵魂是线性空间与线性变换，这两个简单概念源自平面与空间解析几何，是数学思想和抽象思维的最佳体现。学习《矩阵理论》的恰当方法是三管齐下，即“几何+代数+应用”，无几何难以理解矩阵概念，非代数无法透析矩阵本质，唯应用堪可体现矩阵价值。

本书追求数学体系的逻辑性，强调矩阵理论的普适性，注重矩阵应用的前沿性，关切激发读者的探索精神与创新能力，对大量抽象概念赋予了简明的几何意义，精心设计的思考题遍布全书。每章都扼要介绍了若干应用实例，特别在每章习题之后设置了“研究性问题”，展示了中国数学家(包括上海交通大学本科生)的若干最新研究成果并给出相关参考文献。

本书第一章线性代数概要复习线性代数的基本知识，补充了一些由线性代数可以自然引入的概念和结论，比如矩阵的满秩分解、矩阵的张量积和求解矛盾方程的最小二乘法(LSM)等。

第二章线性空间与线性变换是本书的数学基础和核心，阐释了矩阵是线性变换的代数形式，线性变换是矩阵的几何直观，这就使线性代数中的许多疑问豁然开朗：相似的矩阵具有相同的特征值是因为它们是同一线性变换的特征值。对于有限维线性空间，矩阵与线性变换是一回事，这就是所谓“线性代数基本定理”。本章第五节特别指出内积与正定矩阵是一回事，并在第八节应用中简要介绍了广义相对论所使用的非正定内积，希望读者能够借此突破传统束缚，早日创造出自己的新理论。

第三章 Jordan 标准形与特征值问题是矩阵理论的精华，因为矩阵计算的实质是特征值的计算，而矩阵的 Jordan 标准形则从理论上提供了理解矩阵性质、计算矩阵函数、研究矩阵微积分的简便方法。本章对Jordan 标准形的存在性证明采用[9]的直接计算法，其关键在于分块 Schur 三角化定理；本章还讨论了特征值估计的盖尔圆盘定理。本章最后介绍了特征值与特征向量的两个著名应用，即随机矩阵与PangeRank算法。

第四章矩阵分解与应用从著名的 Schur 三角化定理入手，给出矩阵分解的一种框架：即使线性变换$x \mapsto Ax$有较好的几何意义。本章详细介绍了具有广泛应用的谱分解，LU分解， Cholesky 分解，QR分解和奇异值分解(SVD)。矩阵分解是应用最广的矩阵理论和方法，大数据科学可以粗略地看成是SVD+LSM(奇异值分解+最小二乘法)，本章特别给出了SVD的两个应用实例。

第五章赋范线性空间与矩阵微积分是两种强大理论与方法的结合，可视为微积分对矩阵的应用或矩阵理论对微积分的推广。矩阵微积分需要将$\mathbb{R}^n$中的欧氏距离推广为矩阵空间的“范数”，范数理论是众多数学课程共同的基础。利用范数概念即可讨论矩阵函数的极限、连续性以及微积分，特别是矩阵的级数理论。本章包含两个重要应用，即利用范数理论研究矩阵近似和利用矩阵微积分研究线性系统。

第六章广义逆矩阵与非负矩阵，前者是2020年诺贝尔物理学奖得主Penrose的早期发现,现在仍然是矩阵理论研究的前沿；后者在Leontief荣获1973 年诺贝尔经济学奖的主要工作中有举足轻重的作用，其中百年历史的Perron-Frobenius定理依然是矩阵理论最璀璨的明珠。

本书附录为我校 2015-2016 学年《矩阵理论》试题(无答案)，以方便读者检验学习效果。

书末的汉英名词索引提供了本书出现过的术语的汉字词条和相应的(一种)英文对照，按

照数字与字母或汉语拼音顺序排列，以方便读者查阅。

目　录

主要符号表

有时使用同一符号表示不同概念, 如 0 表示数字 0 和矩阵 0, 也表示零空间或零变换.

$\mathbb{R}, \mathbb{C}, \mathbb{Q}, \mathbb{Z}, \mathbb{N}$	实数域, 复数域, 有理数域, 整数(环), 自然数集
i	虚数单位$\sqrt{-1}$
$\mathrm{Re}(\lambda)$	复数λ的实部
$\mathrm{Im}(\lambda)$	复数λ的虚部
$\bar{\lambda}$	复数λ的共轭
$\mathbb{R}^n$	实数域上n维有序数组构成的线性空间
$\mathbb{C}^n$	复数域上n维有序数组构成的线性空间
$\mathbb{F}^n$	数域$\mathbb{F}$上n维有序数组构成的线性空间
$M_n,\ M_n(\mathbb{F})$	数域$\mathbb{F}$上n阶方阵全体构成的线性空间
$\mathbb{R}^{m\times n}$	全体$m\times n$阶实矩阵构成的线性空间
$\mathbb{C}^{m\times n}$	全体$m\times n$阶复矩阵构成的线性空间
$\mathbb{F}^{m\times n}$	数域$\mathbb{F}$上全体$m\times n$阶矩阵构成的线性空间
$C[a,b]$	区间$[a,b]$上全体实变量连续函数构成的线性空间
$\Longleftrightarrow$	充分必要条件
$\forall$	对所有(任意)
$\exists$	存在有
$\square$	证毕
$\partial f(x)$	多项式$f(x)$的次数
A^{-1}	矩阵A的逆矩阵
$A^{\dagger}$	矩阵A的 Moore-Penrose 广义逆矩阵
$A^{(i,j)}$	矩阵A的$\{i,j\}$广义逆
$\mathrm{vec}(A)$	矩阵A的列展开
$\mathrm{rvec}(A)$	矩阵A的行展开
A^T	矩阵(或向量)A的转置
A^*	矩阵(或向量)A的共轭转置
$A>0$	矩阵A为正定矩阵或A为正矩阵即A的所有元素均大于零
$A\geq 0$	矩阵A为半正定矩阵或A为非负矩阵即A的所有元素均非负
$A\otimes B$	矩阵A与B的张量积(也称为 Kronecker 积)
$\mathrm{adj}\,A$	矩阵A的伴随矩阵
$r(A)$	矩阵A的秩
$\mathrm{tr}\,A$	矩阵A的迹
$\sigma(A)$	矩阵A的谱
$\rho(A)$	矩阵A的谱半径
$\lvert A\rvert$	复数A的模或矩阵A的行列式(偶尔也表示A的绝对值矩阵)或集合A的元数
C_n^r	从n个不同元素中取出r个元素的组合数
δ_{ij}	Kronecker 符号, 即$\delta_{ij}=1$如果$i=j$; $\delta_{ij}=0$如果$i\neq j$
$\mathrm{diag}(\lambda_1,...,\lambda_n)$	对角线元为$\lambda_1,...,\lambda_n$的对角矩阵
e_i^T	$(0,...,0,1,0,...,0)$ 第i个分量为 1 的基本行向量
e_j	$(0,...,0,1,0,...,0)^T$ 第j个分量为 1 的基本列向量
E_{ij}	在(i,j)处为 1 其余位置为零的矩阵
H_A	矩阵A的 Hermite 标准形
$I,\ I_m$	单位矩阵, m阶单位矩阵
J	矩阵的 Jordan 标准形
$J_k(\lambda)$	对角线为λ的k阶标准 Jordan 块
$N(A)$	矩阵A的零空间
$R(A)$	矩阵A的列空间(像空间)

$\|A\|$	矩阵A的(矩阵)范数
$\|A\|_1$, $\|A\|_2$, $\|A\|_\infty$	矩阵A的l_p范数$(p=1,2,\infty)$
(x,y)	向量x与向量y的内积
$x\perp y$	向量x与向量y正交(垂直)
$\mathrm{Span}\{\alpha_1,...,a_k\}$	由向量$\alpha_1,...,\alpha_k$生成的子空间, 有时也用$[\alpha_1,...,a_k]$表示
$U\oplus W$ $(U\overset{\perp}{W})$	子空间(或矩阵)U与W的直和(正交直和)
$\sum\limits_{i=1}^{s}\oplus U_i$	子空间(或矩阵)$U_1,\cdots,U_s$的直和
V/U	线性空间V关于子空间U的商空间
1_V	线性空间V上的恒等变换(单位变换)
0_V	线性空间V上的零变换
$\dim V$	线性空间V的维数
$r(\sigma)$	线性变换σ的秩
$\eta(\sigma)$	线性变换σ的零度
σ^*	线性变换σ的伴随变换
$\mathrm{Im}\,(\sigma)$	线性变换σ的像空间
$\mathrm{Ker}\,(\sigma)$	线性变换σ的核空间
$\mathrm{Hom}(V,W)$	由线性空间V到W的线性变换全体构成的集合
V^*	线性空间V的对偶空间
$\mathrm{End}\,V$	由线性空间V到自身的线性变换全体构成的集合
$W^\perp$	子空间W的正交补
V_λ	由对应于特征值λ的特征向量生成的特征子空间
P_U	子空间U上的正交投影变换(矩阵)
$\mathrm{Proj}_U\alpha$	向量α在子空间或向量U上的投影向量
$\sigma\|_U$	线性变换σ在子空间U上的限制
$\sum\limits_{i=1}^{s}\oplus\sigma_i$	线性变换$\sigma_1,\cdots,\sigma_s$的直和
$U\otimes V$	线性空间U与V的张量积
$\sigma\otimes\tau$	线性变换σ与τ的张量积

第一章　线性代数概要

除非特别说明, 所有讨论均假定是在复数域$\mathbb{C}$的某子域$\mathbb{F}$上进行的, 熟知的如$\mathbb{Q}$(有理数域), $\mathbb{R}$(实数域), $\mathbb{C}$(复数域)等, 稍广的有$\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}, \mathbb{Q}(i) = \{a + bi \mid a, b \in \mathbb{Q}\}$, 其中$i = \sqrt{-1}$是虚单位.

不加说明时均默认矩阵是n阶方阵, 其中n是某正整数.

本 章 提 要

本章概括后续章节需要的线性代数知识, 并引入一些自然的概念, 如满秩分解与三角分解, 矩阵的谱等, 新内容只有矩阵的**张量积**. 大多数线性代数的结论均不加证明, 仅利用极端重要的分块矩阵技巧证明了**Sylvester**[1] **不等式**.

本科线性代数基本上由实数域上的线性方程组和矩阵对角化两部分构成.

线性方程组的矩阵形式$Ax = b$特别适合计算机处理, 利用MATLAB等数学软件可以方便快速地判断系数矩阵A与增广矩阵(A, b)的秩是否相等, 从而判断解的存在性.

线性方程组$Ax = b$还可以写为向量形式

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

因此, 线性方程组的解实际上是向量b关于A的列$\alpha_1, \cdots, \alpha_n$的组合系数.

如果$Ax = b$有解, 则其解是其任何一个解(称为特解)与相应齐次线性方程组$Ax = 0$的通解之和. 因此, 齐次线性方程组$Ax = 0$的解的结构是线性方程组理论的核心.

有解的线性方程组称为**相容方程组**, 无解的则称为**矛盾方程组**. 将线性方程组的理论分成两部分是合适的, 即如何表达相容方程组的所有解以及如何处理矛盾方程组.

本科线性代数多采用“基础解系”的任意线性组合来表达齐次相容方程组的所有解(称为“通解”), 而非齐次相容方程组的解可分解为一个特解与相应齐次方程组的通解之和. 这种方法虽然有效, 但弱化了线性方程组的几何意义, 对处理矛盾方程组缺乏启发性, 本教程将利用线性空间的理论予以弥补.

矩阵运算的重点和难点均在于其“乘法”. 矩阵乘法源自向量的线性组合, 本质上是映射的复合运算, 因此满足结合律而不满足交换律(一般称满足结合律的运算为“乘法”, 同时满足结合律与交换律的运算为“加法”). 利用特征值与特征向量, 一些矩阵可以化为对角形, 即存在可逆矩阵P使得$P^{-1}AP = D = \mathrm{diag}\,(\lambda_1, \cdots, \lambda_n)$, 因此

$$A^m = PD^mP^{-1} = P\mathrm{diag}\,(\lambda_1^m, \cdots, \lambda_n^m)P^{-1}.$$

特别地, 实对称矩阵可以正交对角化, 即存在正交矩阵Q使得

$$Q^{-1}AQ = Q^TAQ = \mathrm{diag}\,(\lambda_1, \cdots, \lambda_n).$$

设$x = (x_1, ..., x_n)^T, y = (y_1, ..., y_n)^T$. 则实二次型$f(x) = x^TAx$可以利用坐标变换$x = Qy$化为标准型

$$f = \lambda_1 y_1^2 + + \cdots + \lambda_n y_n^2 = y^T\Lambda y$$

[1]James Joseph Sylvester(1814-1897), 英国数学家, 是19世纪后半叶美国最著名的数学家, 美国历史最悠久的著名数学期刊American Journal of Mathematics的创始人.

其中Λ是对角线元素依次为$\lambda_1, ..., \lambda_n$的对角矩阵.

所以, 本科线性代数本质上是学习矩阵的代数运算, 而研究生的矩阵理论则是理解矩阵的"几何结构"即线性空间理论以及矩阵的广泛应用.

第一节 矩阵运算

1.1.1 线性组合与矩阵乘法

设m, n是正整数. 域$\mathbb{F}$上$m \times n$阶矩阵全体记为$\mathbb{F}^{m\times n}$, 其中零矩阵记为0. 全体n阶可逆矩阵构成的集合记为$GL_n(\mathbb{F})$或GL_n, 其中单位矩阵记为I_n或I(线性代数多用E). 单位矩阵I_n的第i列记为$e_i^{(n)}$或e_i, 称为(第i个)标准向量.

对$A = (a_{ij}) \in \mathbb{F}^{m\times n}$, 将$A$的每个元素改为其共轭元素所得的矩阵称为$A$的共轭矩阵, 记为$\bar{A}$, 即$\bar{A} = (\bar{a}_{ij})_{m\times n}$. A的**共轭转置**矩阵记为A^*(也常用A^H以纪念Hermite[2]), 显然有$A^* = (\bar{A})^T = \overline{A^T}$. 如果$A$是实矩阵, 则$A^* = A^T$.

矩阵A的第i列是Ae_i, 第i行是$e_i^T A$, 所以A按列或按行的分块形式为

$$A = (Ae_1, ..., Ae_n) = \begin{pmatrix} e_1^T A \\ \vdots \\ e_m^T A \end{pmatrix}.$$

特别地, $a_{ij} = e_i^T A e_j$.

矩阵乘法源自向量的线性组合. 设$A \in \mathbb{F}^{m\times n}, x = (x_i) \in \mathbb{F}^n$, 则

$$Ax = x_1(Ae_1) + ... + x_n(Ae_n).$$

第i行第j列元素为1, 其余元素均为0的$m \times n$矩阵称为**基本矩阵**或**矩阵单位**, 记为E_{ij}, $1 \le i \le m$, $1 \le j \le n$. 任意$m \times n$矩阵$A = (a_{ij})$均能唯一地表示成

$$\begin{aligned} A &= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \\ &= a_{11}E_{11} + \cdots + a_{1n}E_{1n} + \cdots + a_{m1}E_{m1} + \cdots + a_{mn}E_{mn} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij}. \end{aligned}$$

利用$E_{ij} = e_i e_j^T$(此处默认$e_i = e_i^{(m)}, e_j^T = (e_j^{(n)})^T$)可得

$$E_{ij}E_{kl} = \delta_{jk}E_{il}, \quad 1 \le i \le m,\ 1 \le j, k \le p,\ 1 \le l \le n,$$

[2] Charles Hermite(1822-1901), 法国著名数学家与教育家.

其中δ_{ij}是Kronecker[3]符号, 其定义为

$$\delta_{jk}=\begin{cases}1, & \text{如果 } j=k;\\ 0, & \text{其他.}\end{cases}$$

于是矩阵的乘法可表为:

$$\begin{aligned}AB&=(\sum_{i=1}^{m}\sum_{j=1}^{p}a_{ij}E_{ij})(\sum_{i=1}^{p}\sum_{j=1}^{n}b_{ij}E_{ij})\\&=\sum_{i,j,k,l}(a_{ij}E_{ij})(b_{kl}E_{kl})=\sum_{i,j,k,l}(a_{ij}b_{kl})(E_{ij}E_{kl})\\&=\sum_{i,j,k,l}(a_{ij}b_{kl})(\delta_{jk}E_{il})=\sum_{i,j,k,l}(\delta_{jk}a_{ij}b_{kl})E_{il}\\&=\sum_{i=1}^{m}\sum_{j=1}^{n}(\sum_{k=1}^{p}a_{ik}b_{kj})E_{ij},\end{aligned}$$

即乘积AB的第i行第j列的元素等于$\displaystyle\sum_{k=1}^{p}a_{ik}b_{kj}$, 这正是矩阵乘法的“左行右列”规则, 是按每个“元素”来做“乘法”运算, 表达了矩阵乘法的几何意义, 因为乘积的每个元素恰好是左(行)右(列)两个向量的内积:

$$A_{m\times p}B_{p\times n}=\begin{pmatrix}e_1^TA\\ \vdots\\ e_m^TA\end{pmatrix}(Be_1,...,Be_n)=\begin{pmatrix}e_1^TABe_1 & \cdots & e_1^TABe_n\\ \vdots & \cdots & \vdots\\ e_m^TABe_1 & \cdots & e_m^TABe_n\end{pmatrix}$$

或简单地

$$AB=(e_i^TABe_j).$$

矩阵的行列分块还可以导出矩阵乘法的“列结构”与“行结构”, 即

$$A_{m\times p}B_{p\times n}=A(Be_1,...,Be_n)=(ABe_1,...,ABe_n)$$

与

$$A_{m\times p}B_{p\times n}=\begin{pmatrix}e_1^TA\\ \vdots\\ e_m^TA\end{pmatrix}B=\begin{pmatrix}e_1^TAB\\ \vdots\\ e_m^TAB\end{pmatrix}.$$

于是两个矩阵的乘积$C=AB$的列向量与行向量的结构为(这是优化的计算机算法):

$$Ce_j=A(Be_j),\quad e_i^TC=(e_i^TA)B.$$

即矩阵AB的第j列是A的列向量的线性组合, 组合系数恰为矩阵B的第j列的相应元素; AB的第i行是B的行向量的线性组合, 组合系数恰为矩阵A的第i行的相应元素.

[3]Leopold Kronecker(1823 - 1891)是著名德国数学家, 逻辑学家, 其名言: God made the integers; all else is the work of man. 符号δ_{ij}是最著名的数学符号之一, 被称为Kronecker delta.

例 1.1.1. 若$AB=0$, 则$A(Be_j)=0$, 故B的每个列向量都是线性方程组$Ax=0$的解向量(特别, 若A是方阵, 则B的每个非零列向量都是A的属于特征值0的特征向量). 同理, 由于$(e_i^TA)B=0$, 故A的每个行向量都是线性方程组$y^TB=0$的解向量.

例 1.1.2. 方程组$Ax=b$有解 $\Longleftrightarrow$ b是系数矩阵A的列的线性组合 $\Longleftrightarrow$ $r(A)=r(A,b)$, 即系数矩阵的秩等于增广矩阵的秩. 特别地, 齐次线性方程组$Ax=0$有非零解 $\Longleftrightarrow$ A的列向量线性相关; 有唯一解(即零解) $\Longleftrightarrow$ A的列向量线性无关.

1.1.2 数字特征与分块矩阵

设$f(x)=a_0+a_1x+\cdots+a_mx^m$是$\mathbb{F}$上的一个多项式. 对$n$阶方阵$A$, 记

$$f(A)=a_0I+a_1A+\cdots+a_mA^m,$$

称为A的多项式. 易知, 同一方阵的两个多项式是可以交换的, 即若

$$g(x)=b_0+b_1x+\cdots+b_lx^l,$$

则$f(A)g(A)=g(A)f(A)$.

如果$\exists k>0, A^k=0$, 则称A是**幂零矩阵**; 对应地, 如果$\exists k>0, A-I$是幂零矩阵, 则称A是**幂幺矩阵**.

例 1.1.3. 最重要的幂零矩阵是

$$J_n=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

矩阵J_n称为n阶**标准幂零矩阵**或n阶**幂零块**或n阶**幂零 Jordan 块**. 矩阵J_n是最简单的秩为$n-1$的n阶严格上三角矩阵, 在矩阵的Jordan标准形理论和微积分运算中起关键性作用. 一般将J_n中元素1所在的斜线称为(第一)上对角线.

方阵$A=(a_{ij})$的**行列式**记为$|A|$(更为通用的记号是$\det A$), 具有性质$|AB|=|A|\,|B|$(此即**“乘法公式”**). 二阶矩阵$A=(\alpha_1,\alpha_2)$的行列式的几何意义是列向量为邻边的平行四边形的有向面积(正、负号意味着从α_1旋转到α_2为逆、顺时针; 三阶矩阵$A=(\alpha_1,\alpha_2,\alpha_3)$的行列式的几何意义是列向量为邻边的平行六面体的有向体积(正、负号意味着$\alpha_1,\alpha_2,\alpha_3$构成右、左手螺旋).

需要特别指出, 大多数线性代数课程将行列式理解为一个数是比较初等的, 研究生阶段应具有新视角和高观点, 即**“行列式是标准交错多重线性函数”**, 具体定义如下:

定义 1.1.1. 设det是$\overbrace{\mathbb{F}^n\times\mathbb{F}^n\times\cdots\mathbb{F}^n}^{n\text{重}}$到$\mathbb{F}$的函数, 对$\forall 1\le i\le n$, 满足下列条件:

1. 多线性:

$$\det(\alpha_1,...,x\alpha+y\beta,...,\alpha_n)=x\det(\alpha_1,...,\alpha,...,\alpha_n)+y\det(\alpha_1,...,\beta,...,\alpha_n);$$

2. 交错性: $\det(\alpha_1,...,\alpha_{i-1},\alpha_i,...,\alpha_n)=-\det(\alpha_1,...,\alpha_i,\alpha_{i-1},...,\alpha_n)$;
3. 标准性: $\det(I)=\det(e_1,...,e_n)=1$.

则称det是一个标准n重交错线性函数. 对$\forall A\in\mathbb{F}^{n\times n}$, 函数值$\det(Ae_1,...,Ae_n)$称为矩阵$A$的**行列式**, 记为$\det(A)$或$\det A$, 即

$$\det(A)=\det(Ae_1,...,Ae_n) \tag{1.1.1}$$

不难证明, 对$\forall A\in\mathbb{F}^{n\times n}$, 均有$\det(A)=|A|$ (留作习题).

将行列式理解为交错多重线性函数有很多益处, 比如由多线性和交错性立即可知奇异矩阵的行列式均为0, 因此只需要处理可逆矩阵的行列式即可. 可逆矩阵都可以写成初等矩阵的乘积, 而初等矩阵的行列式也可以从多线性和交错性立即得到(具体数值由标准性确定). 于是, 行列式的关键是“乘法公式”, 此容易由“矩阵A右乘以初等矩阵P的行列式$\det(AP)=\det(A)\det(P)$”得到, 因为初等矩阵右乘一个矩阵不过是对该矩阵的列向量作简单操作而已(见本章第二节).

方阵A的对角线元素之和称为A的**迹**, 记为$\mathrm{tr}\,A=\sum\limits_{i=1}^{n}a_{ii}$. 迹具有下列基本性质:

命题 1.1.1. 设$A=(a_{ij})$, B均为n阶方阵, α,β是数, 则
(1) $\mathrm{tr}(\alpha A+\beta B)=\alpha\mathrm{tr}\,A+\beta\mathrm{tr}\,B$ (此称为“线性性质”);
(2) $\mathrm{tr}\,(AB)=\mathrm{tr}\,(BA)$ (此仅需AB,BA均有意义即可);
(3) $\mathrm{tr}\,A^*=\mathrm{tr}\,\bar{A}=\overline{\mathrm{tr}\,A}$. 特别地, 对实矩阵有$\mathrm{tr}\,A^T=\mathrm{tr}\,A$;
(4) $\mathrm{tr}(AA^*)=\sum\limits_{i,j=1}^{n}|a_{ij}|^2$. 特别地, $\mathrm{tr}\,(AA^*)=0\iff A=0$.

其中性质(4)是因为AA^*的第j个对角线元素为$e_j^TAA^*e_j=(A^*e_j)^*(A^*e_j)=\sum\limits_{k=1}^{n}|a_{jk}|^2$, 即$A$的第$j$行作为$n$维向量的长度(向量$\alpha$的长度记为$\|\alpha\|$以区别于绝对值)$\|Ae_j\|$的平方, 一般将向量$x$的长度平方即$|x|^2=x^*x$称为$x$的**能量** .

例 1.1.4. 设n阶矩阵A满足对角强优条件(亦称“严格对角占优”条件)

$$|a_{ii}|>\sum_{j\neq i}|a_{ij}|,\ i=1,2,\cdots,n.$$

证明: A可逆.

证 只需证明$Ax=0$只有平凡解. 反设$x=(x_i)$是$Ax=0$的一个非零解, 则存在$m,1\leq m\leq n$使得$|x_m|=\max\{|x_1|,\cdots,|x_n|\}>0$. 因为$\sum\limits_{j=1}^{n}a_{kj}x_j=0,1\leq k\leq n$, 故$|a_{mm}||x_m|\leq\sum_{j\neq m}|a_{mj}||x_j|$, 从而$|a_{mm}|\leq\sum_{j\neq m}|a_{mj}|$, 矛盾! 故$Ax=0$无非零解, 从而$A$可逆. □

矩阵A的另一个重要数字特征是**秩**, 定义为A的所有不为零的子式的最高阶数, 记为$r(A)$. 约定零矩阵的秩是0.

对任意n阶方阵$A=(a_{ij})$, 去掉第i行第j列后所剩余的$n-1$阶方阵的行列式称为元素a_{ij} 的**余子式**, 记为M_{ij}. 而$(-1)^{i+j}M_{ij}$称为元素a_{ij}的**代数余子式**, 记为A_{ij}. n阶方阵

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

称为A的**伴随矩阵**, 记为$\mathrm{adj}\,A$(如果仅限于实矩阵, 也常用符号A^*). 伴随矩阵的重要性在于:

$$A(\mathrm{adj}\,A)=(\mathrm{adj}\,A)A=|A|I. \tag{1.1.2}$$

对n阶方阵而言, “秩为n”(也称为“**满秩**”), “**非奇异**”与“**可逆**”是等价的三个概念. 可逆矩阵的逆矩阵是唯一的, 记为A^{-1}. 逆矩阵具有下述性质:

命题 1.1.2. (1) $A^{-1}=\frac{1}{|A|}\mathrm{adj}\,A$; $|A^{-1}|=|A|^{-1}$; $|\mathrm{adj}\,A|=|A|^{n-1}$;
(2) 若数$\lambda\neq 0$, 则$(\lambda A)^{-1}=\lambda^{-1}A^{-1}$; 特别地, $(A^{-1})^{-1}=A$;
(3) $(A^T)^{-1}=(A^{-1})^T$; $(A^*)^{-1}=(A^{-1})^*$;
(4) $(AB)^{-1}=B^{-1}A^{-1}$;
(5) 设A是$m\times n$矩阵, $P\in GL_m, Q\in GL_n$, 则

$$r(A)=r(PA)=r(AQ)=r(PAQ),$$

此性质简称为“可逆矩阵不改变矩阵的秩”.

矩阵的和与积的秩由下述不等式给出估计.

定理 1.1.1. (1) $r(A+B)\leq r(A)+r(B)$;
(2)(Sylvester不等式) 设A,B分别为$m\times p$, $p\times n$矩阵, 则

$$r(A)+r(B)-p\leq r(AB)\leq \min\{r(A),r(B)\}. \tag{1.1.3}$$

以下利用分块矩阵证明(2)中第1个不等式, 其余不等式均容易证明.

分块初等矩阵是由分块单位矩阵经过一次行(或列)初等变换得到的分块矩阵, 其作用与普通初等矩阵一样. 设P是一个初等分块矩阵, A是一个适当的分块矩阵, 则PA等于对A的行作与P匹配的初等变换. 类似地, 用初等分块矩阵Q右乘A是对A的列作与Q匹配的初等变换.

现考虑下面的矩阵等式

$$\begin{pmatrix} I_m & A \\ 0 & I_p \end{pmatrix}\begin{pmatrix} AB & 0 \\ 0 & I_p \end{pmatrix}\begin{pmatrix} I_n & 0 \\ -B & I_p \end{pmatrix}=\begin{pmatrix} 0 & A \\ -B & I_p \end{pmatrix}.$$

右端矩阵的秩显然不小于

$$r\begin{pmatrix} 0 & A \\ -B & 0 \end{pmatrix}=r(A)+r(B),$$

而左端矩阵的秩恰好是$r(AB)+p$, Sylvester不等式得证.

分块对角矩阵有一种非常简洁的记法, 即

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_s \end{pmatrix} = A_1 \oplus A_2 \oplus \cdots \oplus A_s = \sum_{i=1}^{s} \oplus A_i,$$

这种表示称为矩阵的**直和**, 每个子矩阵A_i称为一个直和项.

例 1.1.5. (分块对角矩阵乘分块矩阵.) 设$A = \sum\limits_{i=1}^{s} \oplus A_i, B = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1t} \\ B_{21} & B_{22} & \cdots & B_{2t} \\ \vdots & \vdots & \cdots & \vdots \\ B_{s1} & B_{s2} & \cdots & B_{st} \end{pmatrix}$. 则

$$AB = \begin{pmatrix} A_1B_{11} & A_1B_{12} & \cdots & A_1B_{1t} \\ A_2B_{21} & A_2B_{22} & \cdots & A_2B_{2t} \\ \vdots & \vdots & \cdots & \vdots \\ A_sB_{s1} & A_sB_{s2} & \cdots & A_sB_{st} \end{pmatrix},$$

即A的第i个直和项乘在B的第i行. 类似地, 分块对角矩阵A右乘分块矩阵B等于A的第i个直和项乘在B的第i列. 这与对角矩阵与一般矩阵乘积的“左行右列”规则完全一致:

$$\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} \begin{pmatrix} \alpha_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \lambda_1 a_{11} & \lambda_1 a_{12} \\ \lambda_2 a_{21} & \lambda_2 a_{22} \end{pmatrix}, \quad \begin{pmatrix} \alpha_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 a_{11} & \lambda_2 a_{12} \\ \lambda_1 a_{21} & \lambda_2 a_{22} \end{pmatrix}.$$

分块矩阵是研究矩阵的强有力工具, 必须熟练掌握.

思考题

1. 讨论任意矩阵的行列式是否有意义?
2. 讨论方阵的平方根是否有意义? 试讨论二阶零矩阵与单位矩阵的平方根.

1.1.3 张量积与线性矩阵方程

矩阵的乘法源自线性组合, 实际上是向量的加法, 本质上是映射的复合. 下面给出一种真正意义的乘法, 即矩阵的**张量积**.

定义 1.1.2. (矩阵的张量积) 设$A = (a_{ij})$与$B = (b_{st})$分别是$m \times n$与$p \times q$ 矩阵. A与B的(左)张量积(矩阵)或Kronecker 积, 记作$A \otimes B$, 是指$mp \times nq$矩阵

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

例 1.1.6. (1) $e_i^{(n)} \otimes (e_j^{(m)})^T = E_{ij}$;

(2) $I_n \otimes I_m = I_{mn}$;

(3) $\begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} ax & ay \\ bx & by \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix}$.

(4) $\begin{pmatrix} a & b \end{pmatrix} \otimes \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax & bx \\ ay & by \end{pmatrix} \neq \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax + by$.

以下列出矩阵张量积的一些性质(证明留作习题).

命题 1.1.3. (1)(结合律) $(A\otimes B)\otimes C=A\otimes(B\otimes C)$.
(2)(分配律1) $(A\otimes B)(C\otimes D)=AC\otimes BD$.
(3)(分配律2) $A\otimes(B\pm C)=A\otimes B\pm A\otimes C$.
(4)(保转置) $(A\otimes B)^T=A^T\otimes B^T$.
(5)(保可逆) $(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$ (故$A\otimes B$可逆 $\Longleftrightarrow$ A与B均可逆).
(6) 设A与B分别为m阶与n阶矩阵, 则$|A\otimes B|=|A|^n|B|^m$.
(7)(保秩与迹) $r(A\otimes B)=r(A)r(B)$; $\operatorname{tr}(A\otimes B)=\operatorname{tr}(A)\operatorname{tr}(B)$.
(8)(保特征值) 设λ与μ分别是m阶矩阵A与n阶矩阵B的特征值, x与y是相应的特征向量, 则$\lambda\mu$是$A\otimes B$的特征值, $\lambda+\mu$是$A\otimes I_m+I_n\otimes B$的特征值, $x\otimes y$是相应的特征向量.

利用矩阵的张量积可以求下面线性矩阵方程的解.

$$A_1XB_1+A_2XB_2+\cdots+A_sXB_s=C \tag{1.1.4}$$

其中$A_i\in M_m(\mathbb{C}),B_i\in M_n(\mathbb{C}),C\in\mathbb{C}^{m\times n}$是已知矩阵, $X\in\mathbb{C}^{m\times n}$为未知矩阵.

定义 1.1.3. 设$A=(a_{ij})_{m\times n}$, 将A的各列依次竖排得到mn维列向量, 称为矩阵A的**列展开**, 记为$\operatorname{vec}(A)$, 即

$$\operatorname{vec}(A)=(a_{11},a_{21},...,a_{m1},a_{12},a_{22},...,a_{m2},......,a_{1n},a_{2n},...,a_{mn})^T.$$

类似地, 矩阵A的**行展开**为:

$$\operatorname{rvec}(A)=(a_{11},a_{12},...,a_{1n},a_{21},a_{22},...,a_{2n},......,a_{m1},a_{m2},...,a_{mn}).$$

由上述定义, 得$\operatorname{rvec}(A^T)=(\operatorname{vec}(A))^T$, $\operatorname{vec}(A^T)=\operatorname{rvec}(A)^T$.

下面的定理列出了矩阵的行(列)展开的一些性质, 证明留作习题.

定理 1.1.2. (1) $\operatorname{rvec}(ABC)=\operatorname{rvec}(B)(A^T\otimes C)$, $\operatorname{vec}(ABC)=(C^T\otimes A)\operatorname{vec}(B)$. 特别地,
$\operatorname{vec}(A_{m\times m}B_{m\times n})=(I_n\otimes A)\operatorname{vec}(B)$, $\operatorname{vec}(B_{m\times n}C_{n\times n})=(C^T\otimes I_m)\operatorname{vec}(B)$.
(2) $\operatorname{vec}(xA+yB)=x\operatorname{vec}(A)+y\operatorname{vec}(B)$, $x,y\in\mathbb{C}$.
(3) $\operatorname{vec}(AC+CB)=(I_n\otimes A+B^T\otimes I_m)\operatorname{vec}(C)$.

现在可以给出线性矩阵方程(1.1.4)的解了.

定理 1.1.3. 矩阵$X\in\mathbb{C}^{m\times n}$是方程(1.1.4)的解 $\Longleftrightarrow$ $\operatorname{vec}(X)$是方程$G\mathbf{x}=\mathbf{c}$的解, 其中$\mathbf{x}=\operatorname{vec}(X)$, $\mathbf{c}=\operatorname{vec}(C)$, $G=\displaystyle\sum_{i=1}^{s}(B_i^T\otimes A_i)$.

证 只需将方程(1.1.4)两端按列展开, 即可得
$\mathbf{c}=\operatorname{vec}(C)=\sum_{i=1}^s\operatorname{vec}(A_iXB_i)=\sum_{i=1}^s(B_i^T\otimes A_i)\operatorname{vec}(X)=G\operatorname{vec}(X)=G\mathbf{x}$. □

推论 1.1.1. 矩阵方程(1.1.4)有解 $\Longleftrightarrow$ $r([G,\mathbf{c}])=r(G)$; 有唯一解 $\Longleftrightarrow$ G是可逆矩阵.

例 1.1.7. 求**Lyapunov**[4]**方程** $AX+XB=C$的解, 其中

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

解

$$G=I_2\otimes A+B^T\otimes I_2=\begin{pmatrix}3&1&0&0\\0&3&0&0\\1&0&3&1\\0&1&0&3\end{pmatrix},$$

$\mathbf{c}=\mathrm{vec}(C)=(-2,3,3,4)^T$. 解方程$G\mathbf{x}=\mathbf{c}$得

$$\mathbf{x}=(-1,1,1,1)^T.$$

因此原矩阵方程的解为

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

思考题

1. 阶数固定时秩为0的矩阵只有1个. 设$1\le s<t$. 讨论秩为s的矩阵与秩为t的矩阵的多寡. 特别地, 当$n\ge 2$时, 讨论n阶可逆矩阵与不可逆矩阵的多寡.

2. 试给出矩阵秩的一种直观意义.

3. 固定正整数$r\le n$.

(1) 记$J(r)$为秩为r的所有n阶矩阵全体. 试问$J(r)$有何“结构”?

(2) 记$K(r)$为秩不超过r的所有n阶矩阵全体. 试问$K(r)$有何“结构”?

4. 试讨论矩阵张量积的直观意义.

第二节 矩阵与线性方程组

1.2.1 相容方程组与矩阵子空间

相容线性方程组$Ax=b$可以化为$A(x-x_0)=0$(x_0是任意解), 而齐次线性方程组$Ax=0$的解集$N(A)=\{x\in\mathbb{F}^n\mid Ax=0\}$有“线性结构”, 即下述

定理 1.2.1. 设α,β是$Ax=0$的两个解, $\lambda\in\mathbb{F}$, 则
(1) $\alpha+\beta$也是$Ax=0$的解.
(2) $\lambda\alpha$也是$Ax=0$的解.
特别地, 如果任意有限个向量$\alpha_1,\cdots,\alpha_s$均是$Ax=0$的解, 则其任意线性组合

$$k_1\alpha_1+\cdots+k_s\alpha_s, k_i\in\mathbb{F}$$

也是解.

定理1.2.1表明$N(A)$关于$\mathbb{F}^n$中的向量加法与数乘均封闭, 具有这种性质的集合称为$\mathbb{F}^n$的“**子空间**”. 本章我们称$\mathbb{F}^n$的所有子空间为“**向量空间**”, 于是$N(A)$是一个向量空间, 称

[4]Aleksandr Mikhailovich Lyapunov(1857-1918), 著名俄罗斯数学家, 力学家, 物理学家, 稳定性理论与动力系统的奠基人. 于其妻病逝当天开枪自杀.

为A的“**零空间**”. 向量空间的一般理论将在第二章“线性空间”中讨论. 易知, $\mathbb{F}^n$中任意有限个向量$\alpha_1,\cdots,\alpha_s$的任意线性组合构成的集合

$$\{\alpha\in\mathbb{F}^n\ |k_1\alpha_1+\cdots+k_s\alpha_s,k_i\in\mathbb{F}\}$$

是向量空间, 称为由$\alpha_1,\cdots,\alpha_s$张成的空间, 记为$\mathrm{Span}\{\alpha_1,\cdots,\alpha_s\}$, $\alpha_1,\cdots,\alpha_s$称为生成元集. $\mathrm{Span}\{\alpha\}$也简记为$\mathbb{F}\alpha$. 向量个数最少的生成元集称为该向量空间的一组**基**. 零空间$N(A)$的基中文线性代数教材一般称为$Ax=0$的基础解系.

向量空间V的基一般不唯一, 但每个基包含的向量个数均相同, 称为该向量空间的**维数**, 记为$\mathrm{dim}V$. 比如, 若$A\in\mathbb{F}^{m\times n}, r=r(A)$. 则$\mathrm{dim}N(A)=n-r$.

除零空间$N(A)$外, 与矩阵A密切相关的还有下述3个向量空间.

(1) A的**列空间**(或像空间, 值域)$R(A)$. 即A的列向量张成的子空间;
(2) A的**行空间**$R(A^T)$: 即A的行向量张成的子空间.
(3) A的**左零(化)空间**$N(A^T)$: 即线性方程组$y^TA=0$或$A^Tx=0$的解空间.

注意, $N(A)$与$R(A^T)$是$\mathbb{F}^n$的子空间; 而$R(A)$与$N(A^T)$是$\mathbb{F}^m$的子空间, 且有

$$\dim N(A)+\dim R(A^T)=n;\qquad \dim N(A^T)+\dim R(A)=m. \tag{1.2.1}$$

于是, 矩阵的秩就是列空间与行空间的维数. 请注意, 行空间中的向量仍然看成是列向量.

计算矩阵四空间的基本方法是所谓**Gauss**[5]**消元法**或初等变换. 初等变换可以通过相应的初等矩阵来实现.

1. 重排变换: 交换第i行(列)与第j行(列); 重排矩阵为: $I-E_{ii}-E_{jj}+E_{ij}+E_{ji}$.
2. 倍乘变换: 给第i行(列)乘以非零数a; 倍乘矩阵为: $I+(a-1)E_{ii}$.
3. 消元变换: 将第j行(第i列)的a倍加到第i行(j列); 消元矩阵为: $I+aE_{ij}$.

矩阵的一次行(列)初等变换相当于左(右)乘相应的初等矩阵. 比如, 设有按列分块矩阵$A=(A_1,A_2,\cdots,A_n)$, 则

$$A(I+(a-1)E_{ii})=A+(a-1)AE_{ii}=(A_1,\cdots,A_{i-1},aA_i,A_{i+1},\cdots,A_n),$$

此即是将A的第i列乘a而其余列均不变的矩阵.

重排矩阵(相应于重排变换的矩阵)的乘积称为**置换矩阵**, 即交换单位矩阵任意行或列而得到的矩阵. n阶置换矩阵的全体记为$\mathcal{P}_n$.

任意置换矩阵$P=(e_{\sigma(1)},\cdots,e_{\sigma(n)})\in\mathcal{P}_n$的列指标$\sigma(1),\cdots,\sigma(n)$是序列$1,2,\cdots,n$的一个排列, 因此$\sigma$可以看成是集合$\underline{n}=\{1,2,\cdots,n\}$到自身的一个双射, 称为$\underline{n}$的一个**置换**, 记为$\sigma=(\sigma(1),\cdots,\sigma(n))$. $\underline{n}$的全体置换记为S_n. S_n按照映射的复合运算(称为“乘法”)构成一个群, 即该乘法满足“结合律”、“单位元”(存在单位元)和“逆运算”(每个元素均有逆元素), 称为$\underline{n}$的**对称群**. S_n中的元素σ称为**奇置换**如果矩阵$(e_{\sigma(1)},\cdots,e_{\sigma(n)})$的行列式为-1, 称为**偶置换**如果矩阵$(e_{\sigma(1)},\cdots,e_{\sigma(n)})$的行列式为1. 奇偶置换的符号分别定义为$-1$与1, $\sigma\in\mathcal{P}_n$的符号记为$sgn(\sigma)$.

显然S_n的元素个数$|S_n|=|\mathcal{P}_n|=n!$, 且若$n\geq 2$, 则奇偶置换的数量均为$\dfrac{n!}{2}$. 置换和置换矩阵可以通过映射$\sigma\mapsto P_\sigma=(e_{\sigma(1)},\cdots,e_{\sigma(n)})$视为同一.

[5]Johann Carl Friedrich Gauss(1777-1855), 德国著名数学家.

例 1.2.1. 设$A=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$. 则$A$是循环矩阵, 也是置换矩阵, 对应的置换为$(2,3,1)$. 设$x=(x_1,x_2,x_3)^T$, 则$Ax=(x_2,x_3,x_1)^T$.

Gauss消元法可将方程组化为同解的最简方程组. 等价地, 初等变换可通过乘以若干初等矩阵化简矩阵, 其中由行初等变换得到的最简形式称为该矩阵的**简化行阶梯形**.

定义 1.2.1. 设$m\times n$矩阵H的秩为r且满足以下条件:
(1) H的后$m-r$行均为0, 前r行均非0且第一个非零元(称为主元)均为1.
(2) 主元的列标随行标严格递增; 即若第k行的主元在第j_k列, 则$j_1<j_2<\cdots<j_r$.
(3) 主元所在列的其余元素均为零, 即第j_k列为标准单位向量e_k.
则称H是简化行阶梯形.

定理 1.2.2. 任一矩阵A都可经过一系列行初等变换化为简化行阶梯形H_A, 且相应的线性方程组同解, 即若$H_A=PA$, 则$Ax=b$与$H_Ax=Pb$同解.

利用定理1.2.2即可**简便计算A的四个子空间**. 首先, 注意到矩阵H_A是简化行阶梯形当且仅当存在置换矩阵Q使得$H_AQ=\begin{pmatrix}I_r&F_{r\times(n-r)}\\0&0\end{pmatrix}$, 其中$F$是自由变量的系数矩阵. 因此有$R(A^T)=R(H_A^T)$, 且$H_A$的前r行(看作列向量)构成$R(A^T)$的一组基. 其次, 由于矩阵$\begin{pmatrix}-F\\I_{n-r}\end{pmatrix}$列满秩, 故其列构成$N(H_AQ)$的一组基, 因此$N(A)=N(H_A)$的一组基恰好是矩阵$Q\begin{pmatrix}-F\\I_{n-r}\end{pmatrix}$的所有列向量. 再次, $H_A=PA$的后$m-r$行均为0, 因此P的后$m-r$行恰好构成$N(A^T)$的一组基. 最后, 由于行初等变换不改变列向量之间的线性关系, 故H_A的主元所在的列对应的A的列向量组恰好是$R(A)$的一组基.

例 1.2.2. 设$i=\sqrt{-1}$是虚数单位, 试求下面复数矩阵A的四个子空间.

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

解 将A化为简化行阶梯形, 并记录相应的可逆矩阵P:

$$(A,I_3)=\begin{pmatrix}1&1&0&1&0&0\\3&3&1&0&1&0\\2&2&i&0&0&1\end{pmatrix}\longrightarrow\begin{pmatrix}1&1&0&1&0&0\\0&0&1&-3&1&0\\0&0&0&3i-2&-i&1\end{pmatrix}=(H_A,P),$$

由此可得$PA=H_A$, 即

$$P=\begin{pmatrix}1&0&0\\-3&1&0\\3i-2&-i&1\end{pmatrix},\ H_A=\begin{pmatrix}1&1&0\\0&0&1\\0&0&0\end{pmatrix}.$$

从而矩阵A的秩$=2$, 而H_A的第1、3两列线性无关, 所以A的这两列也线性无关, 即

$$\begin{aligned}&R(A)=\mathrm{Span}\{Ae_1,Ae_3\}=\mathrm{Span}\{(1,3,2)^T,(0,1,i)^T\};\\&R(A^T)=R(H_A^T)=\mathrm{Span}\{H_A^Te_1,H_A^Te_2\}=\mathrm{Span}\{(1,1,0)^T,(0,0,1)^T\};\\&N(A)=N(H_A)=\mathrm{Span}\{(1,-1,0)^T\}=\mathbb{F}(1,-1,0)^T.\end{aligned}$$

最后计算$N(A^T)$. 矩阵P的最后一行左乘A为零向量, 故它是$xA=0$的解, 即

$$N(A^T)=\mathrm{Span}\{P^Te_3\}=\mathrm{Span}\{(3i-2,-i,1)^T\}=\mathbb{F}(3i-2,-i,1)^T.$$

利用简化行阶梯形还可以得到矩阵的**满秩分解**, 其定义如下.

定义 1.2.2. 如果存在列满秩矩阵L与行满秩矩阵R使得$A=LR$, 则称$A=LR$是A的一个满秩分解.

非0矩阵都存在满秩分解. 满秩分解$A=LR$中的矩阵L与R的秩显然等于A的秩.

例 1.2.3. 设A的秩为r, 则存在可逆矩阵P使得

$$PA=H_A=\begin{pmatrix} H_r \\ 0 \end{pmatrix},$$

其中$r=r(A)=r(H_r)$. 设H_r的第k行的主元所在的列标为j_k, 则$j_1<j_2<\cdots<j_r$, 且H_r的$j_1,j_2,\cdots,j_r$列组成一个r阶单位矩阵, 故A的$j_1,j_2,\cdots,j_r$列线性无关, 且

$$A=(A_{j_1},A_{j_2},\cdots,A_{j_r})H_r. \tag{1.2.2}$$

显然, 公式(1.2.2)给出A的一个满秩分解.

比如, 在例1.2.2中, 由于$PA=H_A=\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, 故$A$的一个满秩分解为

$$A=(Ae_1,Ae_3)\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 3 & 1 \\ 2 & i \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}=LR. \tag{1.2.3}$$

例 1.2.4. 矩阵$A\in\mathbb{F}^{m\times n}$的秩为1当且仅当存在两个非零向量$\alpha\in\mathbb{F}^m,\beta\in\mathbb{F}^n$使得$A=\alpha\beta^T$. 这就是$A$的一个满秩分解. 进一步, 若$A$是方阵, 则还有$A^k=(\beta^T\alpha)^{k-1}A$. 显然$\beta^T\alpha=\alpha\beta^T=\mathrm{tr}A$.

例 1.2.5. 解释实矩阵A的四个子空间的几何意义和相互位置关系.

由矩阵乘法的行列结构, $A\alpha=0$意味着向量α与A的所有行向量的内积均为0, 因此α垂直于A的行空间, 即$\alpha\perp R(A^T)$. 一般, 将垂直于某集合X的所有向量构成的集合称为X的**正交补**, 记为$X^\perp$. 于是, $R(A^T)$包含了所有垂直于$N(A)$的向量, 反之亦然, 即$R(A^T)=N(A)^\perp,R(A^T)^\perp=N(A)$. 类似地, $R(A)=N(A^T)^\perp,R(A)^\perp=N(A^T)$.

注意, 由于两个复数向量α,β的内积需要定义为$\beta^*\alpha$(或者$\alpha^*\beta$), 故若将例1.2.5中的矩阵换为复矩阵, 则有$R(A^*)$与$N(A)$互为正交补, $R(A)$与$N(A^*)$互为正交补.

思考题

1. 非齐次线性方程组的解与其对应的齐次线性方程组的解的几何意义是什么?
2. 利用初等变换将矩阵化为简化行阶梯形的几何意义是什么?
3. 试给出满秩分解的一种直观意义;
3. 平面上$m\geq 2$条直线相交的概率有多大? 空间中$m\geq 2$张平面相交的概率有多大?
4. 试讨论线性方程组$Ax=b$有解与无解的概率.

1.2.2 矛盾方程组与最小二乘法

如果$Ax=b$无解, 即$Ax=b$是一个矛盾方程, 此时$b\notin R(A)$, 即对任意$x\in\mathbb{F}^n$, 误差向量$e=Ax-b$均非0, 因此只能求方程的近似解. 近似解有多种定义, 此处定义满足长度$\|e\|$最小的向量x^0为近似解.

定义 1.2.3. 设$A\in\mathbb{F}^{m\times n}, b\in\mathbb{F}^m$. 如果向量$x^0\in\mathbb{F}^n$使得对任意向量$x\in\mathbb{F}^n$, 均有$\|Ax^0-b\|\le\|Ax-b\|$, 则称$x^0$是方程$Ax=b$的一个**最优解**.

最优解x^0实际上是函数$y=\|Ax-b\|$的最小值点, 其几何意义如下：函数值$y(x^0)$恰好是向量b的终点到“直线”或“平面”(或更一般地“子空间”)$R(A)$的最小距离, 故误差向量$e=Ax-b$与$R(A)$垂直, 由例1.2.5可知, $Ax-b\in R(A)^\perp=N(A^*)$. 于是有下面的

定理 1.2.3. 设$A\in\mathbb{F}^{m\times n}, b\in\mathbb{F}^m$. 则向量$x^0\in\mathbb{F}^n$是方程$Ax=b$的一个**最优解** $\iff$ x^0是方程$A^*Ax=A^*b$的解.

方程$A^*Ax=A^*b$称为原方程$Ax=b$的**正规化方程**. 注意, 正规化方程总是相容的.

定理1.2.3也可以通过高等数学中多元函数的极值理论得到严格证明, 参考第六章“赋范线性空间与矩阵微积分”. 由于函数$y=\|Ax-b\|$与函数

$$y=\|Ax-b\|^2=(Ax-b)^*(Ax-b)=x^*A^*Ax-x^*A^*b-b^*Ax+b^*b \tag{1.2.4}$$

的极值点相同(注意复数x^*A^*b与b^*Ax互为共轭, 故式(1.2.4)中的每一项都是实数), 而首项x^*A^*Ax中的矩阵A^*A是半正定矩阵, 故函数$y=x^*A^*Ax-x^*A^*b-b^*Ax+b^*b$确有最小值, 故最优解也称为**最小二乘解**. 该方法称为高斯最小二乘法(LSM, Least Square Method), 有广泛而重要的应用.

例 1.2.6. 已知变量b是变量x_1, x_2的函数. 现有观测数据为:

b	x_1	x_2
-1	1	-6
2	1	-2
1	1	1
6	1	7

求线性近似公式: $b=a_1x_1+a_2x_2$.

解 将上述数据写为方程组$Ay=\beta$, 其中

$$A=\begin{pmatrix}1&-6\\1&-2\\1&1\\1&7\end{pmatrix},\quad \beta=\begin{pmatrix}-1\\2\\1\\6\end{pmatrix},\quad y=\begin{pmatrix}a_1\\a_2\end{pmatrix}.$$

由于A的两列Ae_1, Ae_2正交, A^TA是对角矩阵$\begin{pmatrix}4&0\\0&90\end{pmatrix}$, 故其正规化方程$A^TAy=A^Tb$为

$$\begin{pmatrix}4&0\\0&90\end{pmatrix}y=\begin{pmatrix}8\\45\end{pmatrix},$$

因此$a_1 = 2, a_2 = 1/2$. 故所求近似公式为$b = 2x_1 + \frac{1}{2}x_2$.

求解正规化方程$A^*Ax = A^*b$的计算量往往巨大, 一个快捷的替代方法是寻找适当的向量$\hat{b}$使得方程$Ax = \hat{b}$的解恰是原方程$Ax = b$的最优解. 显然, $\hat{b}$的几何意义是向量b在列空间$R(A)$上的投影向量, 记作$\mathrm{Proj}_{R(A)}b$. 因此, 方程组$Ax = b$的最优解就是方程组$Ax = \mathrm{Proj}_{R(A)}b$的解.

例 1.2.7. 在$\mathbb{R}^3$中, 任何向量$\alpha = (x, y, z)$在子空间xoy平面上的投影向量是$(x, y, 0)$, 此恰好是向量α在xoy平面的坐标轴$(1, 0, 0)$与$(0, 1, 0)$上的投影向量$(x, 0, 0)$与$(0, y, 0)$之和.

例 1.2.7表明投影向量可以分解为彼此垂直坐标轴上的投影向量之和.

例 1.2.8. 设$0 \neq \alpha \in \mathbb{C}^n$, 则任何向量$\beta \in \mathbb{C}^n$在$\alpha$上的投影向量$\mathrm{Proj}_\alpha \beta = \dfrac{\alpha^*\beta}{\alpha^*\alpha}\alpha$. 特别地, 如果$\alpha$是单位向量, 则$\mathrm{Proj}_\alpha \beta = (\alpha^*\beta)\alpha$.

两个内积为零的向量称为**正交的**. 设$\alpha_1, ..., \alpha_m$是$\mathbb{C}^n$的一组非零向量. 如果$\alpha_j^*\alpha_i = \delta_{ij}, 1 \le i, j \le m$, 即所有向量均为单位向量且彼此正交, 则称该向量组是一个标准正交组. $\mathbb{C}^n$的子空间的由标准正交组构成的基称为**标准正交基**. 以$\mathbb{C}^n$的标准正交组为列构成的矩阵称为**半正交矩阵**; 特别, 以标准正交基为列(或行)构成的矩阵称为**酉矩阵**. 复矩阵U是半正交矩阵的充分必要条件是$U^*U = I$, 是酉矩阵的充分必要条件是

$$U^*U = UU^* = I \tag{1.2.5}$$

实的酉矩阵就是**正交矩阵**. 实矩阵Q是正交矩阵的充分必要条件是

$$Q^TQ = QQ^T = I \tag{1.2.6}$$

求向量$\mathrm{Proj}_{R(A)}b$的关键在于求子空间$R(A)$的一个标准正交基.

标准正交基的概念可以延伸为一组向量两两正交但未必都是单位向量, 此时相应的矩阵也有类似于公式(1.2.6)的如下结论.

命题 1.2.1. 设$A \in \mathbb{F}^{m\times n}, A^*A = D = \mathrm{diag}\,(d_1, d_2, ..., d_n)$ 是对角矩阵, 则A的所有列向量彼此正交, 且$d_j = \|Ae_j\|^2$. 特别地, 如果A的所有列向量均非0, 则$m \ge n, d_i > 0, 1 \le i \le n$, 且$Ae_1/\sqrt{d_1}, Ae_2/\sqrt{d_2}, ..., Ae_n/\sqrt{d_n}$ 构成$\mathbb{C}^m$的一个标准正交组.

1.2.3 正交化方法与投影矩阵

设V是$\mathbb{C}^n$的一个子空间, $\mathcal{B} = \{\alpha_1, \cdots, \alpha_s\}$是$V$的一个基. $\mathcal{B}$的前k个向量张成的子空间称为第k个顺序坐标面, 记为$L_k(\mathcal{B})$, 即

$$L_k(\mathcal{B}) = \mathrm{span}\{\alpha_1, \cdots, \alpha_k\} = \{\sum_{i=1}^{k} x_i\alpha_i | x_i \in \mathbb{C}, 1 \le i \le k\}, 1 \le k \le s.$$

计算与$\mathcal{B}$具有相同顺序坐标面的V的标准正交基的方法称为**Gram-Schmidt** [6] **正交化方法**, 即取

$$\beta_1 = \alpha_1, \quad \gamma_1 = \frac{\beta_1}{\|\beta_1\|}.$$

[6] Jørgen Pedersen Gram(1850-1916), 丹麦著名数学家与精算学家, 1883 年发现 Gram-Schmidt 正交化方法. Erhard Schmidt(1876-1959), 德国数学家, 1907 年发现 Gram-Schmidt 正交化方法. 实际上该方法早在 1836 年即已被 Cauchy 所使用.

归纳地, 假设$\gamma_1, \gamma_2, \cdots, \gamma_{k-1}$已被确定, 则取

$$\beta_k = \alpha_k - [(\gamma_{k-1}^*\alpha_k)\gamma_{k-1} + (\gamma_{k-2}^*\alpha_k)\gamma_{k-2} + \cdots + (\gamma_1^*\alpha_k)\gamma_1], \tag{1.2.7}$$

其中$(\gamma_j^*\alpha_k)\gamma_j$是向量$\alpha_k$在向量$\gamma_j$上的投影向量, 由于$\gamma_j(1 \leq j \leq k-1)$彼此正交, 故

$$(\gamma_{k-1}^*\alpha_k)\gamma_{k-1} + (\gamma_{k-2}^*\alpha_k)\gamma_{k-2} + \cdots + (\gamma_1^*\alpha_k)\gamma_1$$

恰好是向量α_k在第$k-1$个顺序坐标面$L_{k-1}(\mathcal{B})$上的投影向量. 取

$$\gamma_k = \frac{\beta_k}{\|\beta_k\|},$$

则$\mathcal{B}' = \{\gamma_1, \cdots, \gamma_s\}$是$V$的一个标准正交基, 且$\mathcal{B}'$与$\mathcal{B}$具有相同的顺序坐标面, 即

$$L_k(\mathcal{B}) = L_k(\mathcal{B}), \ 1 \leq k \leq s.$$

□

Gram-Schmidt 正交化方法的几何意义如下:

设$\alpha_1, \alpha_2 \in \mathbb{R}^3$线性无关, 则向量$\alpha \in \mathbb{R}^3$在$\alpha_1$, α_2确定的平面上的投影等于它在该平面的一组正交坐标轴上的投影之和. 比如, 公式(1.2.7)中的被减各项恰好是α_k在诸γ_j $(1 \leq j \leq k-1)$上的投影. 注意, 如果β是一个非零向量, α是任意向量, 则向量$\frac{(\alpha,\beta)}{\|\beta\|^2}\beta$恰好是向量$\alpha$在向量$\beta$上的**投影向量**, 一般记为$\mathrm{Proj}_\beta\alpha$. 特别, 若$\beta$是一个单位向量, 则$\mathrm{Proj}_\beta\alpha = (\alpha, \beta)\beta$.

设$A \in \mathbb{C}^{m\times n}$的秩为$r$. 设$\gamma_1, \cdots, \gamma_r$是$R(A)$的一个标准正交基, 令$B = (\gamma_1, \cdots, \gamma_r) \in \mathbb{C}^{m\times r}$. 由于$B$的列是一个标准正交组, 故对任意向量$\beta \in \mathbb{C}^m$, 有

$$\mathrm{Proj}_{R(A)}\beta = (\gamma_1^*\beta)\gamma_1 + (\gamma_2^*\beta)\gamma_2 + \cdots + (\gamma_r^*\beta)\gamma_r = BB^*\beta.$$

即有下述**投影向量公式**

$$\mathrm{Proj}_{R(A)}\beta = BB^*\beta. \tag{1.2.8}$$

因此, 方程$Ax = b$的最优解就是方程$Ax = BB^*b$的解, 其中B的列是$R(A)$的一个标准正交基, 故$B^*B = I_r$.

矩阵BB^*将任何向量β变为其在$R(A) = R(B) = R(BB^*)$上的投影向量. 一般地, 设V是$\mathbb{C}^n$的任意子空间, $P \in \mathbb{C}^{n\times n}$. 如果$\forall\beta \in \mathbb{C}^n$均有$Pb = \mathrm{Proj}_V\beta$, 则称$P$是$V$上的**正交投影矩阵**, 简称**投影矩阵**. 如果P是V上的投影矩阵, 则显然有$V = R(P)$. 由此可知V上的投影矩阵是唯一的.

单位矩阵和零矩阵是最简单的投影矩阵. 矩阵单位E_{ii}都是投影矩阵.

定理 1.2.4. 矩阵$P \in \mathbb{C}^{n\times n}$为投影矩阵 $\iff$ $P^2 = P^* = P$ $\iff$ 存在列满秩矩阵A使得$P = A(A^*A)^{-1}A^*$ $\iff$ 存在半正交矩阵B使得$P = BB^*$.

显然, 定理中的矩阵A, B均满足条件$R(A) = R(B) = R(P)$, 故A的列向量组构成$R(P)$的一组基, B的列向量组构成$R(P)$的一组标准正交基.

一般将满足条件$A^2 = A$的矩阵称为**幂等矩阵**. 因此, 投影矩阵恰好是幂等的Hermite矩阵. 后面将会看到, 幂等矩阵在广义逆矩阵中所起的作用, 相当于单位矩阵在通常的逆矩阵中所起的作用.

例 1.2.9. 设V为由$\alpha_1 = (1,1,0)^T$和$\alpha_2 = (0,1,0)^T$生成的$\mathbb{R}^3$的子空间, 求正交投影矩阵P和向量$\alpha = (1,0,1)^T$在V上的正交投影向量.

解 由定理定理1.2.4可得

$$
\begin{aligned}
X &= (\alpha_1, \alpha_2) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}, \\
X^*X &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \\
(X^*X)^{-1} &= \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.
\end{aligned}
$$

所以

$$
\begin{aligned}
P &= X(X^*X)^{-1}X^* \\
&= \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.
\end{aligned}
$$

由投影向量公式(1.2.8)可知, α在L上的正交投影向量为

$$
P_L\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.
$$

由于V是xoy平面, 实际上无需计算即可"猜到"任意向量$(x,y,z)^T$在V上的正交投影向量是$(x,y,0)^T$.

思考题

1. 正交性概念是通常垂直概念的推广. Gram-Schmidt 正交化方法在立体几何中有何解释?
2. 试给出标准正交基的一个直观解释.
3. 试解释正交矩阵和酉矩阵的几何意义.
4. 方阵A是酉矩阵当且仅当其共轭转置是酉矩阵. 对半正交矩阵的共轭转置是否仍然是半正交矩阵?
5. 试用投影矩阵证明Gram-Schmidt 正交化方法.

第三节 相似对角化与特征值

1.3.1 线性微分方程组与相似对角化

矩阵的相似对角化源自化简线性微分方程组. 为此考察简化爱情模型. 假定爱情仅与时间t有关, 则描述情侣L与Z恋爱关系的最简模型是所谓爱情方程组(Love Equations)

$$
\begin{cases} \dfrac{dL}{dt} = aL + bZ \\[2ex] \dfrac{dZ}{dt} = cL + dZ \end{cases} \tag{1.3.1}
$$

其中$L = L(t), Z = Z(t)$是可微函数, a,b,c,d是常数.

爱情方程组(1.3.1)可以简写为矩阵形式

$$\frac{dx}{dt} = Ax \tag{1.3.2}$$

其中$x = \begin{pmatrix} L(t) \\ Z(t) \end{pmatrix}$, A是系数矩阵$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. 当$A$是对角矩阵即$b = c = 0$时, 方程组的解是显然的. 一般情况下, 需要将$A$化成对角矩阵. 为此作变量替换

$$x = Py \tag{1.3.3}$$

其中P是待定的可逆矩阵. 代入原方程可得

$$P\frac{dy}{dt} = APy \tag{1.3.4}$$

即

$$\frac{dy}{dt} = P^{-1}APy \tag{1.3.5}$$

于是, 只要存在可逆矩阵P使得$P^{-1}AP$是对角矩阵, 则微分线性方程组(1.3.2)即可迎刃而解. 故相似对角化是求解常微分线性方程组的有效途径, 其具体应用将在第五章详细讨论.

思考题

1. 举一个不能相似对角化的矩阵.
2. 如果系数矩阵A不能对角化, 微分线性方程组(1.3.2)如何处理?

1.3.2 特征值与特征向量

设A是n阶矩阵, 若存在对角矩阵$D = \mathrm{diag}\,(\lambda_1, \lambda_2, \cdots, \lambda_n)$与可逆矩阵$P = (\alpha_1, \alpha_2, \cdots, \alpha_n)$使得$A = PDP^{-1}$, 则称$A$与$D$相似, 同时称$A$可以相似对角化或对角化. (一般称矩阵$P^{-1}AP$与矩阵$A$是相似的矩阵.) 如此则有$A^m = PD^mP^{-1}$, 因此$A$的任意高次幂可以较方便地求出.

由于$AP = PD$, 故$A\alpha_i = \lambda_i\alpha_i$, $1 \le i \le n$. 所以α_i是齐次线性方程组$Ax = \lambda_i x$的非零解, 即数λ_i与非零向量α_i均满足齐次线性方程组

$$Ax = \lambda x \tag{1.3.6}$$

因此方程组(1.3.6)的系数矩阵$\lambda I - A$(称为矩阵A的**"特征矩阵"**)不可逆, 故诸λ_i均是多项式方程

$$|\lambda I - A| = 0$$

的根, 称为A的特征值或特征根或本征值, 多项式$|\lambda I - A|$称为A的**特征多项式**. 对A的每个特征值λ, 诸α_i均是线性方程组$(\lambda I - A)x = 0$的非零解, 称为A的属于特征值λ的特征向量, 该齐次线性方程组的解集称为矩阵A的特征值λ的**特征子空间**, 记为V_λ, 其维数(等于$n - r(\lambda I - A)$)称为特征值λ的**几何重数**.

矩阵A的全体特征值称为A的**谱**, 记为$\sigma(A)$, 其中的最大模称为A的**谱半径**, 记为$\rho(A)$, 即

$$\rho(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}.$$

从几何上看, 矩阵A的特征值全部位于以原点为圆心, 谱半径$\rho(A)$为半径的圆盘内.

设$\sigma(A)=\{\lambda_1,\lambda_2,\cdots,\lambda_s\}$, 且

$$|\lambda I-A|=\prod_{i=1}^{s}(\lambda-\lambda_i)^{n_i} \tag{1.3.7}$$

则正整数n_i称为特征值λ_i的**代数重数**.

定理 1.3.1. 设$A\in\mathbb{F}^{n\times n}$. 则

(1) 矩阵的行列式等于其所有特征值的积, 即$|A|=\prod\limits_{i=1}^{s}(\lambda_i)^{n_i}$.

(2) 矩阵的迹等于其所有特征值的和, 即$\mathrm{tr}(A)=\sum\limits_{i=1}^{s}n_i\lambda_i$.

(3) A可逆 $\Longleftrightarrow$ 0不是A的特征值.

(4) 设$f(x)$是任意多项式, λ是A的一个特征值, α是属于λ的特征向量, 则$f(\lambda)$是$f(A)$的一个特征值, α是属于$f(\lambda)$的特征向量.

(5) 设A可逆且其特征多项式为(1.3.7), 则其逆矩阵的特征多项式为

$$|\lambda I-A^{-1}|=\prod_{i=1}^{s}(\lambda-\lambda_i^{-1})^{n_i},$$

且若α是A的属于特征值λ的特征向量, 则α也是A^{-1}的属于特征值λ^{-1}的特征向量.

(6) 任何特征值的几何重数不超过其代数重数.

(7) 相似矩阵具有相同的特征多项式(因此具有相同的特征值).

(8) 属于不同特征值的特征向量线性无关.

(9) A可以对角化 $\Longleftrightarrow$ A有n个线性无关的特征向量.

定理 1.3.2. 一个n阶矩阵A可以对角化 $\Longleftrightarrow$ A的每个特征值的代数重数与几何重数相等. 特别, 若A有n个不同的特征值, 则A可以对角化.

证 由于代数重数之和为n, 故条件的充分性由定理1.3.1的(6)(8)(9)可得. 现证必要性. 设A可以对角化, 于是存在$P=(\alpha_1,\alpha_2,\cdots,\alpha_2)\in GL_n$与对角矩阵

$$D=\mathrm{diag}\,(\lambda_1,\cdots,\lambda_1,\lambda_2,\cdots,\lambda_2,\cdots,\lambda_s,\cdots,\lambda_s)$$

使得$P^{-1}AP=D$, 即

$$(A\alpha_1,A\alpha_2,\cdots,A\alpha_n)=(\lambda_1\alpha_1,\cdots,\lambda_1\alpha_{n_1},\lambda_2\alpha_{n_1+1},\cdots,\lambda_2\alpha_{n_1+n_2},\cdots,\lambda_n\alpha_n),$$

故知$\alpha_1,\cdots,\alpha_{n_1}$是属于$\lambda_1$的特征向量, $\alpha_{n_1+1},\cdots,\alpha_{n_1+n_2}$是属于$\lambda_2$的特征向量, 等等. 因此对每个特征值$\lambda_i$, $1\le i\le n$, 其几何重数至少是其代数重数n_i. 但由定理1.3.1(6)知任何特征值的几何重数最多是其代数重数, 故必相等. □

下面的定理称为**Cayley-Hamilton**[7]**定理**[8].

定理 1.3.3. 设矩阵A的特征多项式为$f(\lambda)$, 则有$f(A)=0$.

---

[7]Arthur Cayley(1821-1895), 著名英国数学家, 英国现代数学的奠基人. William Hamilton(1805-1865), 著名爱尔兰物理学家, 天文学家和数学家, 以其名字命名的Hamilton力学是牛顿力学的革新, 数学上以发现四元数(Quaternions)而闻名.

[8]该定理当$n=2,3$时由Cayley于1853年证明, $n=4$时由Hamilton于1850年左右证明, 但一般形式由Frobenius于1878年证明.

Cayley-Hamilton 定理表明, n阶矩阵A的n次幂可由其较低次幂线性表示, 故对任意自然数m, 有

$$A^m \in \mathrm{Span}\{I,\ A,\ A^2,\ \cdots,\ A^{n-1}\}.$$

换句话说, n阶矩阵A的任意次幂均属于由$I,\ A,\ A^2,\ \cdots,\ A^{n-1}$生成的$M_n(\mathbb{C})$的子空间. 这就提供了一种计算高次幂的降幂算法.

例 1.3.1. 设

$$A = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix},$$

求A^2, A^3, A^4.

解 A的特征多项式为$f(\lambda) = \lambda^2 - 4\lambda + 1$, 所以$A^2 - 4A + I = 0$. 故知

$$A^2 = 4A - I,\ \ A^3 = 4A^2 - A = 15A - 4I,\ \ A^4 = 15A^2 - 4A = 56A - 15I.$$

命题 1.3.1. 设A与B分别是$m \times n$与$n \times m$矩阵, $m \geq n$. 则

$$|\lambda I_m - AB| = \lambda^{m-n} |\lambda I_n - BA|.$$

证 注意下述分块矩阵的恒等式:

$$\begin{pmatrix} I & B \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & 0 \\ A & AB \end{pmatrix} = \begin{pmatrix} BA & BAB \\ A & AB \end{pmatrix} = \begin{pmatrix} BA & 0 \\ A & 0 \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix},$$

因此, 矩阵

$$C_1 = \begin{pmatrix} BA & 0 \\ A & 0 \end{pmatrix} \ \ 与矩阵\ \ C_2 = \begin{pmatrix} 0 & 0 \\ A & AB \end{pmatrix}$$

相似. 而C_1的特征值是BA的特征值加上m个0; C_2的特征值是AB的特征值加上n个0. 现因C_1与C_2的特征值相同, 故AB与BA 的非零特征值相同, 故只相差$m - n$个0. □

上述命题称为**Sylvester降幂公式**或**特征多项式降阶公式**.

例 1.3.2. 设

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

则

$$AB = \begin{pmatrix} 1 & 1 & -2 \\ 3 & 1 & -4 \\ 5 & 1 & -6 \end{pmatrix}.$$

较难直接看出AB的特征值, 但是

$$BA = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix},$$

故无须计算即可知BA的特征值为0,-4. 因此, 由 Sylvester 降幂公式知AB的特征值为$0, 0$,-4. 由于$r(AB) = 2$, 故还知道BA可以对角化, 但AB不能对角化.

例 1.3.3. 设n阶矩阵

$$A=\begin{pmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ b & b & a & \cdots & b \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b & \cdots & b & a & b \\ b & b & \cdots & b & a \end{pmatrix},$$

可逆, 求A的特征多项式、$\det(A)$以及A^{-1}.

解 此题可以直接计算, 也可以利用Sylvester 降幂公式. 设e是全1列向量, $J = ee^T$, 则$A=(a-b)I+bJ$. 故由Sylvester 降幂公式得

$$\begin{aligned}&\det(\lambda I-A)=\det[(\lambda-a+b)I-bJ]\\&=(\lambda-a+b)^{n-1}\det[(\lambda-a+b)-be^Te]\\&=(\lambda-a+b)^{n-1}[\lambda-a+(1-n)b].\end{aligned}$$

所以$\det(A)=(a-b)^{n-1}[a+(n-1)b]$.

又$J^2=nJ$, 所以$(J+xI)[J-(n+x)I]=J^2-nJ-x(n+x)I=-x(n+x)I$. 从而

$$(J+xI)^{-1}=\frac{(n+x)I-J}{x(n+x)}.$$

取$x=\dfrac{a-b}{b}$, 则$A=b(J+xI)$, 所以

$$A^{-1}=b^{-1}(J+xI)^{-1}=b^{-1}(\frac{(n+x)I-J}{x(n+x)})=\frac{[(n-2)b+2a]I-A}{(a-b)[a+(n-1)b]}.$$

例 1.3.4. 设A是秩为1的n阶方阵, 试求其特征多项式.

解 由于A的秩为1, 故它至少有$n-1$个特征值为0, 因此其特征多项式$|\lambda I-A|=\lambda^{n-1}(\lambda-\operatorname{tr}A)$.

当m与n均较小或当$m-n$较大时, Sylvester 降幂公式较为有效, 而Cayley-Hamilton 定理仅仅将n次幂简化为$n-1$次幂和较低次幂的线性组合, 不是理想的快速计算模式. 然而, Cayley-Hamilton 定理提供了一种思路, 即如果有次数很低(设为d)的多项式$g(x)$使得$g(A)=0$, 则A的高次幂就可以通过A的$d-1$次及更低次幂的线性组合求出.

例 1.3.5. 设3阶矩阵A的特征多项式为$f(\lambda)=\lambda^3-5\lambda^2$. 设$g(x)=x^2-5x$满足条件$g(A)=0$, 试求$A^4$.

解 由特征多项式可知$A^3-5A^2=0$, 所以$A^4=25A^2$. 但由$g(A)=0$可知$A^2-5A=0$, 所以$A^4=125A$.

定义 1.3.1. 设A是n阶矩阵, $f(x)$是多项式. 如果$f(A)=0$, 则称$f(x)$是A的**零化多项式**. 次数最低的首一零化多项式称为A的**最小多项式**, 记为$m_A(x)$或$m(x)$.

由于矩阵A的特征多项式是它的零化多项式, 因此最小多项式的次数不超过A的阶数. 容易看出最小多项式存在且唯一. 但一般来说, 寻找矩阵的最小多项式非常困难.

命题 1.3.2. 设$m(x)$是A的最小多项式, $f(x)$是A的任意零化多项式, 则$m(x)|f(x)$. 特别地, $m(x)\,|\,|xI-A|$.

证 设$f(x)=m(x)q(x)+r(x)$, 其中$r(x)$的次数小于$m(x)$的次数或者$r(x)$为0多项式. 则$0=f(A)=m(A)q(A)+r(A)=r(A)$, 即$r(x)$也是$A$的零化多项式. 由于$m(x)$是$A$的最小多项式, 故只有$r(x)=0$, 从而$m(x)\,|\,f(x)$. □

例 1.3.6. 求n阶幂零 Jordan 块J_n的最小多项式.

解 直接计算可知$A^n=0$, 但$A^{n-1}\neq 0$, 因此A的最小多项式为$m(x)=x^n$, 恰好等于A的特征多项式.

例 1.3.7. 试求下列分块矩阵的最小多项式:

$$A=(a)\oplus\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}\oplus\begin{pmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix}.$$

解 A的特征多项式为$f(\lambda)=(\lambda-a)^6$. 从而由命题1.3.2可知其最小多项式$m(x)=(x-a)^k$, $k\le 6$, 故考察$(A-aI)$的乘积即可. 因为

$$A-aI=(0)\oplus\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\oplus\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

即$A-aI$是由三个子块组成的分块对角矩阵, 每一个子块都是幂零块, 因此$m(x)=(x-a)^3$.

例 1.3.8. 设$a\neq b$, 试求下列分块矩阵的最小多项式:

$$A=\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}\oplus\begin{pmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix}\oplus(b)\oplus\begin{pmatrix} b & 1 \\ 0 & b \end{pmatrix}.$$

解 A的特征多项式为$(\lambda-a)^5(\lambda-b)^3$. 所以$A$的最小多项式具有形式$(\lambda-a)^s(\lambda-b)^t$. 考察矩阵$A-aI$与$A-bI$的乘积

$$A-aI=J_2\oplus J_3\oplus B,\quad A-bI=C\oplus(0)\oplus J_2,$$

其中B与C是可逆矩阵(因为$a\neq b$). 由分块对角矩阵的乘法可知, 只要分别求出使$A-aI$与$A-bI$的幂零块为零的次数即可. 这两个次数分别为3与2. 因此, A的最小多项式为$m(x)=(x-a)^3(x-b)^2$.

命题 1.3.3. 设A是任意方阵, $m(x)$是A的最小多项式. 设$\lambda_0\in\mathbb{C}$. 则λ_0是A的特征值$\Longleftrightarrow m(\lambda_0)=0$.

证 充分性是显然的, 因为最小多项式的零点也是特征多项式的零点, 故必为特征值.

反之, 设λ_0是A的特征值, α是相应的特征向量, 则$0=m(A)\alpha=m(\lambda_0)\alpha$. 由于$\alpha\neq 0$, 故只有$m(\lambda_0)=0$. □

例 1.3.9. 矩阵A是幂零矩阵 $\iff$ A的特征值均为0.
这是因为, $A^m=0 \iff$ A的最小多项式为x^k, 其中$k\le m$.

例 1.3.10. 求下列矩阵的最小多项式:

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

解 A的特征多项式为$f(\lambda)=(\lambda-3)(\lambda-2)^3$, 故其最小多项式仅有下述三种可能:

$$(x-3)(x-2);\quad (x-3)(x-2)^2;\quad (x-3)(x-2)^3.$$

直接计算得,

$$(A-3I)(A-2I)\neq 0;\quad (A-3I)(A-2I)^2=0,$$

因此最小多项式为$m(x)=(x-3)(x-2)^2$.

一般地有

命题 1.3.4. 分块对角矩阵的最小多项式等于各个子块的最小多项式的最小公倍式(对照: 分块对角矩阵的特征多项式等于各个子块的特征多项式的乘积).

命题 1.3.5. 相似矩阵具有相同的最小多项式.

证 设A与B相似, 即存在可逆矩阵P使得$B=P^{-1}AP$. 设A与B的最小多项式分别为$m_A(x)$与$m_B(x)$, 则$0=m_B(B)=m_B(P^{-1}AP)=P^{-1}m_B(A)P$, 故$m_B(A)=0$, 即$m_B(x)$是$A$的零化多项式, 于是$m_A(x)\mid m_B(x)$; 同理, $m_B(x)\mid m_A(x)$. 由于它们都是首一多项式, 故$m_A(x)=m_B(x)$. □

下面是Cayley-Hamilton 定理的一个重要应用, 可称为**无公共特征值定理**.

定理 1.3.4. 设矩阵A与B无公共特征值, 则$AX=XB \iff X=0$.

证 记$p(\lambda)=|\lambda I-A|$. 由于A,B无公共特征值, 故$p(A)=0$, 但$p(B)$可逆.

现设$AX=XB$, 则$A^kX=XB^k, \forall k\ge 0$. 于是对任意$g(x)\in\mathbb{F}[x]$, 均有$g(A)X=Xg(B)$. 特别地, $0=p(A)X=Xp(B)$, 故$X=0$. □

思考题

1. 矩阵的特征向量和特征值有何直观意义?

2. 如果同时交换矩阵A与B的相同两行(比如同时交换第1、2行), 所得的矩阵相似, 那么A与B是否相似? 如果既交换1、2两行, 又交换1、2两列, 则又如何?

3. 讨论有重特征值的矩阵与无重特征值的矩阵的多寡?

4. 讨论可对角化的矩阵与不可对角化的矩阵的多寡?

5. 设A,B为同阶方阵, 则由降幂公式知AB与BA有相同的特征多项式, 它们是否相似?

6. 特征多项式与最小多项式的商多项式有何意义?

1.3.3 正交对角化与二次型

实对称矩阵的复数版本是**Hermite 矩阵**, 即满足$H^*=H$的矩阵.

定理 1.3.5. Hermite 矩阵的特征值均为实数, 且属于不同特征值的特征向量彼此正交.

该定理的证明与实对称矩阵的特征值均为实数且可以正交对角化类似, 请读者自行完成.

注意, 复对称矩阵$\begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}$的两个特征值均非实数, 因此复对称矩阵不是实对称矩阵的真正推广.

定理 1.3.6. (Hermite矩阵的酉对角化) Hermite矩阵A可以酉对角化, 即存在酉矩阵U使得$U^*AU = D$是对角矩阵. 特别地, 实对称矩阵可以正交对角化.

证明思路: 由定理1.3.5可知, 存在n维线性空间$\mathbb{C}^n$的一组标准正交基, 使得每个基向量都是该Hermite矩阵的特征向量, 以此组标准正交基为列作成的矩阵就是所需要的酉矩阵.

定理1.3.6为化简一般n元(Hermite)二次型提供了理论依据.

设$x = (x_1, \cdots, x_n)^T$是未定元(即未知数)向量,称下面关于$x_1, \cdots, x_n, \bar{x}_1, \cdots, \bar{x}_n$的复系数二次多项式

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \bar{x}_i x_j$$

为(复) Hermite二次型, 简称二次型. 易知存在唯一的n阶Hermite矩阵$A = (a_{ij})$使得$f(x) = x^*Ax$, 该矩阵A称为二次型$f(x)$的矩阵.

定义 1.3.2. 设$f(x) = x^*Ax$是复二次型, A是Hermite矩阵, 若对任意非零向量$\alpha \in \mathbb{C}^n$, 均有$f(\alpha) = \alpha^*A\alpha > 0$, 则称$f(x)$是正定二次型, A是正定矩阵. 类似地可以定义半正定、负定、半负定二次型, 半正定、负定、半负定矩阵等. 对实二次型和实对称矩阵的定义类似.

定理 1.3.7. 设A是n阶Hermite矩阵, 则下列条件等价:
(1) A是正定的.
(2) $f(x) = x^*Ax$是正定二次型.
(3) A的特征值均为正实数.
(4) 存在$m \times n$阶列满秩矩阵M使得$A = M^*M$.
(5) 存在n阶可逆矩阵M使得$A = M^*M$.
(6) 存在n阶可逆矩阵P使得$P^*AP = I$(即A与I**合同**).
(7) A的所有顺序主子式均为正.

定理 1.3.8. 设$A \in \mathbb{C}^{n\times n}$是半正定矩阵, 则存在唯一的半正定矩阵$P$使得

$$A = P^*P = P^2 \tag{1.3.8}$$

矩阵P称为A的平方根(也有作者称为**绝对值**), 常记为$\sqrt{A}$或$A^{1/2}$.

证 设$r = r(A)$. 由定理1.3.6知存在$\sigma_1 \geq \sigma_r > 0$和对角矩阵$D = \operatorname{diag}(\sigma_1^2, ..., \sigma_r^2, 0, ..., 0)$与酉矩阵$U$使得$A = UDU^*$, 令$P = U\operatorname{diag}(\sigma_1^2, ..., \sigma_r^2, 0, ..., 0)U^*$, 则$P$是半正定矩阵且$A = P^2$.

再证明唯一性. 设P, Q均为秩为r的半正定矩阵使得$A = Q^2 = P^2$. 设酉矩阵W使得$W^*QW = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}$, 其中$\Lambda$是$r$阶正定对角矩阵. 则

$$W^*Q^2W = \begin{pmatrix} \Lambda^2 & 0 \\ 0 & 0 \end{pmatrix} = (W^*PW)^*(W^*PW) \tag{1.3.9}$$

上式表明W^*PW是后$n-r$列均为0, 但由于其半正定矩阵, 故其后$n-r$行也均为0, 即有

$$W^*PW = \begin{pmatrix} R & 0 \\ 0 & 0 \end{pmatrix} \tag{1.3.10}$$

比较(1.3.9)与(1.3.10), 即得$R = D, W^*PW = W^*QW$, 从而$P = Q$. 故矩阵A的平方根是唯一的. □

关于半正定矩阵, 有下述**樊畿**[9]**行列式不等式**(1950年), 证明留作习题.

例 1.3.11. 设A, B均为半正定矩阵, $\alpha, \beta \geq 0, a + \beta = 1$. 则$|\alpha A + \beta B| \geq |A|^\alpha |B|^\beta$.

由于存在酉矩阵U与实对角矩阵$D = \mathrm{diag}\,(\lambda_1, \cdots, \lambda_n)$使得$U^*AU = D$ (故$\lambda_1, \cdots, \lambda_n$是$A$的实特征值), 作坐标变换$x = Uy\,(y = (y_1, y_2, \cdots, y_n)^T)$即可将$f(x)$化简为下面的对角形式:

$$f(x) = (Uy)^*A(Uy) = \lambda_1|y_1|^2 + \lambda_2|y_2|^2 + \cdots + \lambda_n|y_n|^2 \tag{1.3.11}$$

上式具有重要意义, 由此不仅可以判断二次型$f(x)$及其矩阵A的正定性等而且可以确定(实)二次型$f(x)$所确定的几何体$f(x) = 0$或$f(x) = 1$的形状.

关于二次型的研究历史悠久, 其中不乏著名且有趣的理论, 比如下面两条.

Lagrange[10] **四平方和定理.** 对$\forall n \in \mathbb{N}$, $\exists x, y, z, w \in \mathbb{Z}$使得$n = x^2 + y^2 + z^2 + w^2$.

Conway-Schneeberger 15-定理. 设$a \leq b \leq c \leq d$均是正整数. 则对任意正整数n, 存在整数x, y, z, w使得$n = ax^2 + by^2 + cz^2 + dw^2 \iff$ 存在整数x, y, z, w使得$k = ax^2 + by^2 + cz^2 + dw^2, 1 \leq k \leq 15$.

Lagrange 四平方和定理被誉为“初等数学之巅”, Conway-Schneeberger 15-定理简称为“15-定理”, 更是被誉为“二十世纪的勾股定理”, 读者可参考[18].

思考题

1. 矩阵的对角化与高等数学或数学分析中多元二次函数的极值有何关系?

2. 使用非退化坐标变换$x = Py$, 其中P为可逆矩阵, 可将二次型$f(x) = x^*Ax$化简为比公式(1.3.11)更简单的形式. 试比较非退化变换与正交变换对角化二次型的异同.

第四节 应用

随着科学技术日新月异的巨大进步, 线性代数在当今社会的各个领域的作用越来越大. 本节仅作一简要介绍.

1.4.1 图与矩阵

一、无向图

一个**无向图**G可以用其顶点集V与边(也称为“弧(arc)”)集E的二元组(V, E)来定义. 如果$V = \{v_1, \cdots, v_n\}$与E都是有限集, 则图$G = (V, E)$可以用矩阵表示. 设v_i与v_j之间边的数目为a_{ij}, $1 \leq i, j \leq n$, 则称矩阵$A(G) = (a_{ij})_{n \times n}$为图$G$的**邻接矩阵**, 二次型$x^T(2I - A)x$称为图$G$的二次型.

例 1.4.1. 设G是下面图D_4(参见习题49):

[9] Ky Fan(1914-2010), 著名华裔美国数学家.
[10] Joseph Louis Lagrange(1736-1813), 著名法国数学家和天文学家.

$$
\begin{tikzcd}
a \arrow[r, dash] & b \arrow[r, dash] \arrow[d, dash] & c \\
& d &
\end{tikzcd}
$$

图 1.4.1

则其邻接矩阵为(行标列标均为a, b, c, d)

$$
A(G) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.
$$

D_4的二次型为

$$
x^T(2I - A)x = 2x_1^2 + 2x_2^2 + 2x_3^2 + 2x_4^2 - 2x_1x_2 - 2x_2x_3 - 2x_2x_4.
$$

图的邻接矩阵是对称非负整数矩阵. 容易看出, 有限图与对称非负整数矩阵一一对应. 用矩阵表示图, 便于用代数理论与计算机研究较为复杂的图, 揭示图的深层性质, 比如例 1.4.1中图的二次型是正定的, 再比如$e_i^T A(G)^k e_j$表示顶点i到j的长度为k的路径总数, 证明留作习题.

二、有向图

如果将一个无向图的边都赋予方向, 就得到了**有向图**或**“箭图”**, 此时顶点常称为(特别是在计算机、网络和通讯技术领域)**“结点”**, 边常称为**箭向**. 有向图的箭向α的起点记为$s(\alpha)$, 终点记为$t(\alpha)$. 邻接矩阵不能反映边的方向, 因此刻画有向图需要新的矩阵, 这就是**“关联矩阵”**. 有向图G的关联矩阵B的行指标为其箭向, 列指标为其顶点, B的第α行$e_\alpha^T B$定义如下：

$$
B_{\alpha v} = \begin{cases} 1, & \text{如果 } s(\alpha) = v; \\ -1, & \text{如果 } t(\alpha) = v; \\ 0, & \text{其他}. \end{cases}
$$

换句话说, 如果边e的起点为u而终点为v，则$B_{\alpha u} = 1, B_{\alpha v} = -1$, 而若顶点$w$既非$\alpha$的起点, 亦非$\alpha$的终点, 则$B_{\alpha w} = 0$.

例 1.4.2. 设G是例 1.4.1中图D_4的有向图:

$$
\begin{tikzcd}
a \arrow[r, "\alpha_1"] & b \arrow[r, "\alpha_2"] \arrow[d, "\alpha_3"] & c \\
& d &
\end{tikzcd}
$$

图 1.4.2

则其关联矩阵为(行标$\alpha_1, \alpha_2, \alpha_3$, 列标$a, b, c, d$)

$$
B(G) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.
$$

有向图的关联矩阵是一类**差分矩阵**, 比如上例中的矩阵$B(G)$与向量$x=(x_i)$的乘积为

$$B(G)x = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} x_1-x_2 \\ x_2-x_3 \\ x_2-x_4 \end{pmatrix}.$$

有向图的关联矩阵可以用来处理与流量相关的实际问题, 比如著名的Kirchhoff 电流定律即可用所涉电路的有向图的关联矩阵表述为$By=0$.

实际应用中特别重要的一类有向图是所谓**"强连通图"**, 即任意两个顶点之间都存在路径.

1.4.2 随机变量的独立性

概率统计的一个基本概念是随机变量的独立性. 设X与Y是两个密度函数分别为$p(X)$与$p(Y)$的随机变量, 如果

$$p(XY)=p(X)p(Y)$$

则称X与Y是独立的. 设二维离散随机变量(X,Y)的联合密度矩阵为$P=(p_{ij})_{m\times n}$, 关于X和Y的边缘分布分别为$P_X=(p_{i\bullet})_{m\times 1}$与$P_Y=(p_{\bullet j})_{1\times n}$. 则有下述定理(证明留作习题).

定理 1.4.1. 设X与Y是离散随机变量, 二维随机变量(X,Y)的联合密度矩阵为$P=(p_{ij})$. 则下列条件等价.

(1) X与Y相互独立.

(2) 对任意i,j均有$p_{ij}=p_{i\bullet}p_{\bullet j}$.

(3) $P=P_XP_Y^T=P_X\otimes P_Y^T$.

(4) 矩阵$P=(p_{ij})$的秩为1.

流行的概率统计教材仅指出第一个等价, 但第二个与第三个等价明显更加简洁有力高效.

例 1.4.3. 本例取自1999年考研数学一试题. 设随机变量X与Y独立, 填充下表.

(X,Y)	y_1	y_2	y_3	$p_{i\bullet}$
x_1		$1/8$		
x_2	$1/8$			
$p_{\bullet j}$	$1/6$			1

解 因为$p_{11}=\dfrac{1}{6}-\dfrac{1}{8}=\dfrac{1}{24}$, 故第1列完全确定. 由$X$与$Y$的独立性和定理1.4.1可知, 所有的列(与行)均成比例, 因此第2、4列分别是第1列的3、6倍, 所有数据于是均被确定.

推论 1.4.1. 两个联合密度矩阵的张量积仍然是罗赫密度矩阵. 确切地说, 设$P=P(X_1,X_2)$与$Q=Q(Y_1,Y_2)$分别是二维随机变量(X_1,X_2)与(Y_1,Y_2)的联合密度矩阵, 则存在二维随机变量(U,V)使得其联合密度矩阵为$P(U,V)=P(X_1,X_2)\otimes Q(Y_1,Y_2)$.

1.4.3 曲线拟合

一、线性拟合

线性拟合的最简单例子是给定平面$\mathbb{R}^2$上有限个点$P_i = (x_i, y_i), 1 \le i \le n$, 寻找一条直线$\ell$使得所有点$P_i$到$\ell$的距离的平方和最小. 设$\ell$的方程为

$$y = ax + b \tag{1.4.1}$$

则点P_i到ℓ的距离为

$$d_i = d(P_i, \ell) = \frac{|y_i - ax_i - b|}{\sqrt{1+a^2}} \tag{1.4.2}$$

所以线性拟合等价于下述优化问题

$$min \sum_{i=1}^{n} d_i^2 = min \sum_{i=1}^{n} \frac{(y_i - ax_i - b)^2}{1+a^2} \tag{1.4.3}$$

请对照, 如果利用最小二乘法来求最佳逼近直线$L: y = Ax + B$, 则本质上是下述优化问题

$$min \sum_{i=1}^{n} (y_i - Ax_i - B)^2 \tag{1.4.4}$$

与最小二乘法相比, 求解优化问题(1.4.3)的难度大大增加, 目前仅有零星结果, 简介如下.

定义有限点集$P_i (1 \le i \le n)$的中心为$C = (\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i)$.

定理 1.4.2. 优化问题(1.4.3)的解(即拟合直线)必通过点集中心C.

实际上, 定理1.4.2的结果对任意维数均成立, 证明留作习题.

二、非线性拟合

当观测的数据点明显不接近任何直线或平面时, 仍然用一条直线或平面等去拟合所有数据会产生较大的误差, 这时合适的方法是尝试用更高次的拟合曲线, 比如, 若数据点$(x_1, y_1), \cdots, (x_n, y_n)$ 明显位于某条抛物线之上时, 就应该选择方程

$$y = a_0 + a_1 x + a_2 x^2$$

来逼近数据. 更一般地, 记上面的方程导出的方程组为

$$Y = X\alpha$$

其中X称为**设计矩阵**, α称为参数(系数)向量, y称为观测向量. 为了使误差向量(也称为余差向量)$Y - X\alpha$的长度最小, 相当于求出$Y = X\alpha$的最小二乘解, 即方程组$X^T Y = X^T X\alpha$的解.

例 1.4.4. 试求观测数据$(-2, 0), (-1, 0), (0, 1), (1, 0), (2, 0)$的最小二乘二次拟合曲线.

解 设拟合曲线为$y = a + bx + cx^2$. 将观测数据代入得

$$\begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \tag{1.4.5}$$

记系数矩阵为A, 右边的向量为β. 注意A的前两列与后两列分别正交, 易知

$$A^TA=\begin{pmatrix}5&0&10\\0&10&0\\10&0&34\end{pmatrix}, A^T\beta=\begin{pmatrix}1\\0\\0\end{pmatrix}.$$

故正规化方程为

$$\begin{pmatrix}5&0&10\\0&10&0\\10&0&34\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}=\begin{pmatrix}1\\0\\0\end{pmatrix}.$$

于是$a=\dfrac{34}{70}, b=0,\ c=-\dfrac{1}{7}$. □

然而, 故事并非如此简单. 假如对方程组(1.4.5)先进行消元以简化矩阵, 则可得到如下方程组

$$\begin{pmatrix}0&-2&4\\0&-1&1\\1&0&0\\0&1&1\\0&2&4\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}=\begin{pmatrix}-1\\-1\\1\\-1\\-1\end{pmatrix} \tag{1.4.6}$$

现在再求正规化方程可得(注意方程组(1.4.6)的系数矩阵的列向量彼此正交)

$$\begin{pmatrix}1&0&0\\0&10&0\\0&0&34\end{pmatrix}\begin{pmatrix}a\\b\\c\end{pmatrix}=\begin{pmatrix}1\\0\\-10\end{pmatrix} \tag{1.4.7}$$

因此, $a=1, b=0,\ c=-\dfrac{5}{17}$.

这个不同的解表明, 如果对方程组$Ax=b$两端同乘可逆矩阵P而得到(同解)方程$PAx=Pb$, 则原方程的正规化方程$A^*Ax=A^*b$的解与新方程的正规化方程$A^*P^*PAx=A^*P^*Pb$的解可以千差万别! 下面简单的例子解释了该现象.

例 1.4.5. 试求线性方程组$x=1, x=2, y=1$的最小二乘解.

解 正规化方程为$2x=3, y=1$, 因此最小二乘解是(1.5, 1), 该解从几何上来看是非常明显的. 然而, 所求线性方程组从几何上看是平面上的三条直线$ax=a, bx=2b, cy=c$, 其中a, b, c均非0. 此时正规化方程为$(a^2+b^2)x=a^2+2b^2, c^2y=c^2$, 于是最小二乘解是$(\dfrac{a^2+2b^2}{a^2+b^2}, 1)$. 因此, 最小二乘解可以取到直线$y=1$夹在平行线$x=1, x=2$之间的所有点!

一般地, 如果允许对线性方程组进行同解变形(即左右两端同乘以任意可逆矩阵), 则有下述结果(证明略) [11]

定理 1.4.3. 设有n条平面直线围成n边形($n\geq 3$). 则由该n条直线形成的线性方程组的最小二乘解构成的集合是它们的广义交集(即任意三条直线构成的三角形的内点的并集). 特别地，当所有直线共点时, 广义交集即是其交点. 而当所有直线平行时, 最外侧两条直线所夹的区域的内点均可取做最小二乘解.

[11]此结果由上海交通大学数学科学学院2018级本科生徐志鹏, 孙硕, 杨松坛, 黄大庆在2019年得到.

定理 1.4.3也有三维的版本.

1.4.4 华为核心技术极化码

华为的核心技术之一是所谓"**极化码**",, 其数学理论源自著名论文[12], 此处仅作简单介绍.

设$N = 2^n$. 极化码理论包含一系列二元域$GF(2) = \{0,1\}$(四则运算为$mod 2$)上的线性变换, 其中第一个过程称为"**信道合并**"(Channel Combining), 如下图(取自Arikan[13]原始论文):

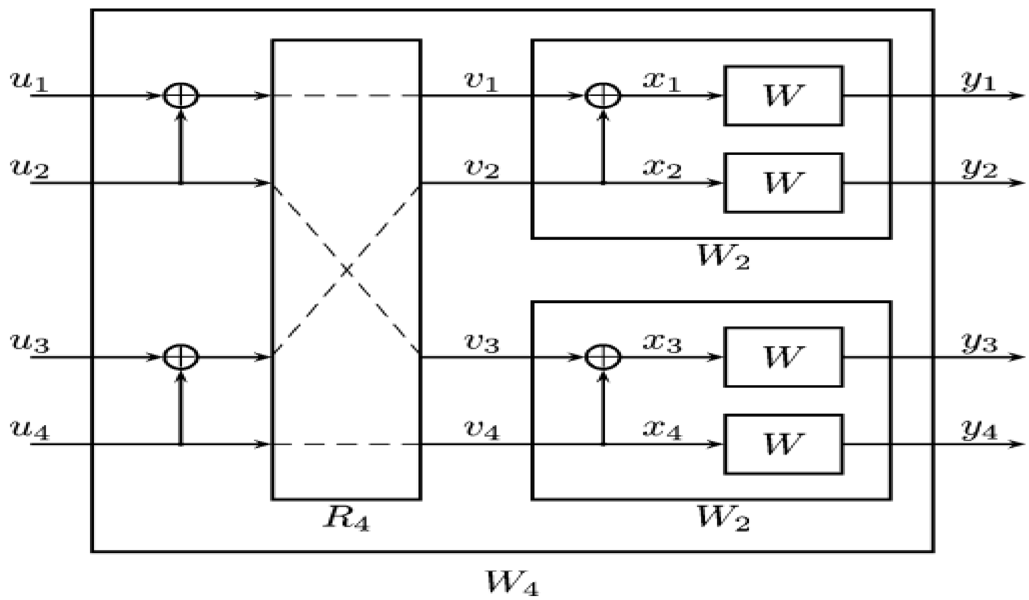

图1.4.3

信道合并实际上是一个线性变换, 其矩阵$G_N \in GF(2)^{N\times N}$称为**生成矩阵**, 关于$G_N$的计算有下述**Arikan 定理**.

定理 1.4.4. 存在置换矩阵B公式使得

$$G_N = BF^{\otimes n} \tag{1.4.8}$$

其中F为比特反转(bit-reversal)矩阵, 即$F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

证 按定义, 生成矩阵可以归纳地定义为

$$G_N = (I_{N/2} \otimes F)R_N(I_2 \otimes G_{N/2}),\ G_1 = I_1 \tag{1.4.9}$$

其中R_N是给定的置换矩阵(例如图1.4.3中的R_4). 按照信道合并的定义(请参阅相关论文), 可得

$$G_N = R_N(I_{N/2} \otimes F)(I_2 \otimes G_{N/2}) \tag{1.4.10}$$

$$G_N = R_N(F \otimes G_{N/2}) \tag{1.4.11}$$

[12]Arikan Erdal, Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Trans. Inform. Theory55(2009), no.7, 3051–3073. 作者注：该文在美国Mathematical Reviews上的评论仅仅照抄了Summary, 说明评论者颇不看好, 华为沙里淘金独具慧眼.

[13]Arikan Erdal(1958-), 土耳其著名数学家, 极化码发明人.

现以$G_{N/2} == R_{N/2}(F\otimes G_{N/4})$代入公式(1.4.11)可得

$$G_N = R_N(I_2\otimes R_{N/2})(F^{\otimes 2}\otimes G_{N/4}) \tag{1.4.12}$$

重复上述步骤, 并注意$G_1 = I_1$以及$R_N, I_2\otimes G_{N/2^k}$均为置换矩阵即可得$G_N = BF^{\otimes n}$. □

习　题　一

1. 计算:

$$(1)\ \begin{pmatrix} \cos x & \sin x \\ -\sin x & \cos x \end{pmatrix}^n;\quad (2)\ \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^n;\quad (3)\ \begin{pmatrix} a & 1 & & & \\ & a & 1 & & \\ & & a & 1 & \\ & & & a & 1 \\ & & & & a \end{pmatrix}^n.$$

2. 证明: (1) 与任意对角矩阵可交换的矩阵必是对角矩阵.

(2) 与任意n阶方阵可交换的矩阵必是纯量矩阵λI.

3. 利用初等变换求$A^{-1}B$及CA^{-1}, 其中

$$A=\begin{pmatrix} 4 & 5 & 0 \\ 2 & 3 & 1 \\ 2 & 7 & -3 \end{pmatrix},\ B=\begin{pmatrix} 4 & 5 & 0 & 10 \\ 2 & 3 & 1 & -1 \\ 2 & 7 & 9 & -3 \end{pmatrix},\ C=\begin{pmatrix} 4 & 5 & 0 \\ 2 & 3 & 1 \\ 2 & 7 & 9 \\ -2 & 3 & 7 \end{pmatrix}.$$

4. 设$A, B\in M_n$, 证明: $\mathrm{adj}\,(AB) = \mathrm{adj}\,(B)\mathrm{adj}\,(A)$.

5. 证明: (1) 对任意矩阵A, 有$r(A^*A) = r(AA^*) = r(A)$.

(2) 对任意n阶矩阵A, 有$r(A^n) = r(A^{n+1})$.

6. 证明: (1) 非零矩阵A的秩为1 $\Longleftrightarrow$ 存在非零列向量α与β使得$A = \alpha\beta^T$.

(2) 设方阵A的秩为1, 则$A^m = (\mathrm{tr}\,A)^{m-1}A$.

7. 设ω是n次本原单位根(可设$\omega = e^{2\pi i/n} = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}$), 试求Fourier [14]矩阵

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{pmatrix}$$

的逆矩阵.

8. 设$A = (a_{ij})_{n\times n}$是可逆的对称实矩阵. 证明: 二次型

$$f(x_1, x_2, \cdots, x_n) = \begin{vmatrix} 0 & x_1 & \cdots & x_n \\ -x_1 & a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ -x_n & a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

的矩阵是A的伴随矩阵.

9. 证明矩阵秩的Frobenius[15] 不等式: $r(AB) + r(BC) \le r(B) + r(ABC)$.

10. 证明许宝騄[16]定理(1953年)：对任意方阵A, 存在对角线元素均为±1的对角矩阵J使得$A+J$可逆.

11. 设所有矩阵均可逆, 证明华罗庚恒等式(1949年): $A - (A^{-1} + (B^{-1} - A)^{-1})^{-1} = ABA$.

---

[14]Joseph Fourier(1768-1830), 著名法国数学家与物理学家.

[15]Ferdinand Georg Frobenius(1849-1917), 德国著名数学家.

[16]P.L.Hsu(1910-1970), 统计学家.

12. 设n阶矩阵A可逆, x与y是n维列向量. 如果$(A+xy^*)^{-1}$可逆, 证明 **Sherman-Morrison**[17] **公式**:

$$(A+xy^*)^{-1}=A^{-1}-\frac{A^{-1}xy^*A^{-1}}{1+y^*A^{-1}x}.$$

13. 设n阶矩阵A可逆, B,C,D分别是$n\times m, m\times n, m\times m$矩阵.证明

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

14. (1) 设矩阵A,C均可逆, 求分块矩阵$\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$的逆矩阵.

(2) 设矩阵A可逆, $D-CA^{-1}B$也可逆,证明分块矩阵$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$可逆并求其逆.

15. 设f是交错n-重线性函数, 证明：对$\forall A\in\mathbb{F}^{n\times n}$, 有$f(A)=f(I)det(A)$. 特别地, $\det(A)=|A|$.

16. 设$\Omega=\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$为$2n$阶分块矩阵. 一个$2n$阶复矩阵$M$称为是**辛矩阵** 如果$M^T\Omega M=\Omega$. 证明:
(1) $2n$阶辛矩阵的全体构成一个群, 即辛矩阵的逆矩阵仍是辛矩阵, 两个辛矩阵的乘积仍是辛矩阵.
(2) 任何辛矩阵的行列式均为1. (提示: 利用分块矩阵.)

17. 求下列各矩阵的满秩分解:

$$(1)\quad A=\begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ 1 & 0 & 2 & 1 \end{pmatrix};\qquad (2)\quad A=\begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}.$$

18. 设

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

求A的四个相关子空间.

19. 证明命题1.1.3.

20. 证明定理1.1.2.

21. (1) 证明任何子空间上的投影矩阵是唯一的.
(2) 证明Afriat[18]定理(1956年)：两个投影矩阵乘积的谱含于闭区间[0, 1].

22. 证明:(1) Hermite矩阵的特征值均为实数, 且属于不同特征值的特征向量彼此正交.
(2) Hermite矩阵A是正定矩阵 $\Longleftrightarrow$ 存在可逆下三角矩阵L使得$A=LL^*$.
(3) 反Hermite矩阵(即$A^*=-A$)的特征值为0或纯虚数, 且属于不同特征值的特征向量彼此正交.
(4) 反Hermite矩阵的秩必为偶数.

23. 证明第三种初等矩阵(即$I+aE_{ij}$, $i\neq j$, $a\neq 0$)彼此相似. 又, 第一种初等矩阵是否彼此相似?

24. 设矩阵A满足方程$A^2-A+2I=0$, 问A可以对角化吗? 为什么? 将本题一般化.

25. 设A是Hermite矩阵. 如果对任意向量x均有$x^*Ax=0$, 则$A=0$.

26. 证明樊畿行列式不等式：设A,B均为半正定矩阵, $\alpha,\beta\geq 0, a+\beta=1$. 则$|\alpha A+\beta B|\geq|A|^\alpha|B|^\beta$.

27. 证明樊畿交换子引理：设$X,Y\in GL_n$. 则存在$A,B\in GL_n$使得$Y=A^{-1}B^{-1}XAB\Leftrightarrow XY^{-1}$是交换子, 即存在$U,V\in GL_n$使得$XY^{-1}=U^{-1}V^{-1}UV$.

[17] Jack Sherman(1927-2007) 和 Winifred Morrison(1910-1961) 均为美国统计学家, 该公式发表于1949年.
[18] S.N.Afriat(1925-2021), 英-加-意数学家、经济学家.

28. 证明樊畿交换子定理：设$X, Y \in GL_n$. 则存在$A_i \in GL_n(1 \leq i \leq 2s+1)$使得$X = \prod_{i=1}^{2s+1} A_i$, $Y = \prod_{i=1}^{2s+1} A_{2s+2-i} \Leftrightarrow XY^{-1}$是若干交换子的乘积.

29. 证明樊畿酉矩阵引理：设X, Y是酉矩阵. 则存在酉矩阵A, B, C使得$X = ABC, Y = CBA \Leftrightarrow |A| = |B|$.

30. 证明实方阵的平方必含非负项.

31. 设$\alpha = (a_1, \cdots, a_n)^T$, $\beta = (b_1, \cdots, b_n)^T, x$为任意常数, $A = xI_n + \alpha\beta^T$.
(1) 直接计算行列式$|A|$;
(2) 利用 Sylvester 降幂公式计算行列式$|A|$;
(3) 利用特征值计算行列式$|A|$.

32. 设A的特征值为0, 1, 对应的特征向量为$(1,2)^T, (2,-1)^T$. 判断A是否为对称矩阵并求A.

33. 求下列矩阵的最小多项式并指出其中可以对角化的矩阵:
(1) $\begin{pmatrix} 3 & 2 \\ 4 & 5 \end{pmatrix}$; (2) $\begin{pmatrix} 2 & 3 & 2 \\ 0 & 5 & 4 \\ 0 & -2 & -1 \end{pmatrix}$; (3) $\begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ 3 & 6 & 1 \end{pmatrix}$; (4) $\begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & 6 & 1 \end{pmatrix}$.

34. 试构造两个同阶矩阵, 使得它们
(1) 具有相同的特征多项式与不同的最小多项式.
(2) 具有相同的最小多项式与不同的特征多项式.
(3) 证明矩阵的最小多项式存在且唯一.

35. 设n阶矩阵A的特征值均为实数. 证明:
(1) A的特征多项式的$n-k$次项的系数等于A的所有k阶主子式之和.
(2) 若A的所有一阶主子式之和与所有二阶主子式之和都等于零, 则A是幂零矩阵.

36. 设n阶矩阵A的主对角元全是1, 且其特征值均为非负数, 证明$|A| \leq 1$.

37. 设$AB = BA$, 证明A与B有公共的特征向量. 该结论的逆命题成立吗?

38. 试判断下面4个矩阵, 哪些是相似的:

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

39. 设

$$A = \begin{pmatrix} 110 & 55 & -164 \\ 42 & 21 & -62 \\ 88 & 44 & -131 \end{pmatrix}.$$

(1) 试直接观察以得出A的特征值;
(2) 由(1)的结果计算A^{2024}.

40. 证明优化问题(1.4.3)的解(即拟合直线)必通过点集中心C, 且此结果对任意高维均成立.

41. 求方程$AX + XB = C$的解, 其中

$$A = \begin{pmatrix} 1 & -1 & 3 \\ -2 & 2 & 2 \\ -1 & -1 & 5 \end{pmatrix}, B = \begin{pmatrix} 1 & -1 & 3 & 0 \\ -2 & 2 & 2 & 0 \\ -1 & -1 & 5 & 0 \\ 0 & 1 & -3 & 1 \end{pmatrix}, C = \begin{pmatrix} 2 & 3 & 1 & 0 \\ 2 & 0 & 1 & 1 \\ 0 & -1 & -2 & 3 \end{pmatrix}.$$

42. 设A, B, C是给定的矩阵. 则下述命题等价：
(1) 矩阵方程$AXB = C$有解.

(2) 存在矩阵M,N使得$C=AM=NB$.

(3) $r(A)=r(A,C), r(B)=r(\begin{pmatrix} B \\ C \end{pmatrix})$.

43. 设x是矩阵A的一个特征向量x, 证明相应于x的特征值为x^*Ax/x^*x(此商称为**Rayleigh**[19] **商**, 是研究特征值的重要工具). 据此研究n元二次型x^*Ax与A的特征值的关系.

44. Vandermonde (范德蒙德)[20] 行列式对应的矩阵称为Vandermonde矩阵. 研究Vandermonde矩阵是否可以对角化.

45. 设$I=I_2$, 试求整数矩阵方程$X^2=\pm I$的所有解. 试一般地讨论方程$X^n=I_n$的解(这样的整数矩阵称为周期矩阵), 其中n为某自然数. (提示: 利用特征多项式.)

46. 设$\sigma: X\mapsto AX+XB$是$M_n(\mathbb{C})$的线性变换. 证明σ是同构 $\Longleftrightarrow$ A和$-B$没有相同的特征值.

47. 设图G的邻接矩阵为

$$A(G)=\begin{pmatrix} 1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{pmatrix},$$

求图G并研究矩阵$A(G)$的谱与图G的关系.

48. 设A是图G的邻接矩阵, 证明$e_i^TA^ke_j$表示顶点i到j的长度为k的路径总数.

49. 设A是有限图$G=(V,E)$的邻接矩阵, 其中$V=\{1,2,...,n\}$. 证明$x^T(2I-A)x$是正定二次型当且仅当下列情形之一成立:

(1) $E=\{(j,j+1), 1\le j\le n-1\}$.

(2) $E=\{(j,j+1), 1\le j\le n-2, (n-2,n)\}$.

(3) $6\le n\le 8, E=\{(j,j+1), 1\le j\le n-2, (3,n)\}$.

上述三种图分别记为A_n, D_n, E_6, E_7, E_8, 统称为邓金图(Dynkin graph).

研究性问题.

方阵A的积和式(permanent)$P(A)$是在行列式的定义中将所有项的负号均取为正号, 即

$$P(A)=\sum_{\underline{i}} a_{1i_1}a_{2i_2}\cdots a_{ni_n}$$

其中$\underline{i}=1i_1i_2\cdots i_n$是$1,2,...,n$的一个排列. A称为完全非负的, 如果A的所有子式均非负. 计算矩阵的积和式是NP-问题. 张晓东[21]等证明了下面的定理(1997年):

设A是完全非负矩阵, 则下列条件等价:

(1) A 的行列式等于其对角元乘积.

(2) A 的积和式等于其对角元乘积.

(3) A 的积和式等于A 的行列式.

相关论文参见[22].

[19] Rayleigh 男爵, 全名 John William Strutt(1842-1919), 著名英国物理学家, 1904年诺贝尔物理学奖得主.

[20] Alexandre-Théophile Vandermonde(1735-1796), 法国数学家, 音乐家和化学家.

[21] Xiaodong Zhang(1965-), 上海交通大学教授.

[22] 张晓东, 杨尚骏, 关于HADAMARD 不等式的注记. 应用数学学报, 20卷2期, 1997, 269-274.

第二章　线性空间与线性变换

本 章 提 要

最简单的数学运算“加法”和“数乘”统称线性运算, 拥有加法和数乘的系统称为**线性空间**, n维向量空间$\mathbb{F}^n$是线性空间的原型和特例.

本章主要讨论有限维线性空间, 其根本结论有三个：

1、有限维线性空间本质上是某个$\mathbb{F}^n$.

2、有限维线性空间的运算本质上是矩阵运算.

3、有限维线性空间可以通过内积引入几何, 近似性等问题可利用正交性解决.

第一节　线性空间

2.1.1　定义与例子

考察n维向量空间$\mathbb{F}^n$可知, 其运算加法“$+$”满足下面四个条件:
(A1) **结合律**: 对任意$\alpha, \beta, \gamma \in \mathbb{F}^n$, $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$.
(A2) **交换律**: 对任意$\alpha, \beta \in \mathbb{F}^n$, $\alpha+\beta=\beta+\alpha$.
(A3) 存在**零向量**: 即存在元素0, 使得对任意$\alpha \in \mathbb{F}^n$, $\alpha+0=\alpha$.
(A4) 存在**负向量**: 对任意$\alpha \in \mathbb{F}^n$, 存在一个向量, 记为$-\alpha$, 使得$\alpha+(-\alpha)=0$.

一般称满足上述条件(A1)的运算称为**“乘法”**(即乘法=结合律), 满足(A1)与(A2)的运算称为“加法”(即乘法=结合律+交换律); 条件(A3)与(A4)实际上保证了加法“$+$”的逆运算减法“$-$”也有意义.

满足上述4个条件的集合V广泛存在, 称为**加群**或**交换群**或**阿贝尔**[23]**群** (无交换律(A2)的系统称为**群**, 此时, 运算称为“乘法”而非“加法”, 运算符号用“$\bullet$”或省略), 记为$(V, +)$, 上下文清楚时, 经常省略$(V, +)$的加法符号“$+$”而简称V是加群. 比如整数集合$\mathbb{Z}$,有理数域$\mathbb{Q}$, 实数域$\mathbb{R}$与复数域$\mathbb{C}$等都是加群(其中加法“$+$”即普通加法).

域$\mathbb{F}$上的$m\times n$阶矩阵的集合$\mathbb{F}^{m\times n}$也是一个加群, 其中加法是矩阵加法. 类似地, $\mathbb{F}$上的多项式全体

$$\mathbb{F}[x]=\{f(x)=a_0+a_1x+...+a_nx^n | a_i\in\mathbb{F}, 0\le i\le n, \forall n\ge 0\}$$

也是一个加群, 其中加法是就是普通多项式加法, 即合并同次项.

数学创新往往始于构造新加法. 著名的黎曼-罗赫定理[24]即始于定义所谓的除子. 设C是一条曲线. 对C上任何有限个点$P_1, ..., P_s$, 称形式和

$$D(P_1,...,P_s)=n_1P_1+\cdots+n_sP_s \ (n_j\in\mathbb{Z}, 1\le j\le s) \tag{2.1.1}$$

是由$P_1, ..., P_s$生成的除子. 所有的除子按照系数的加法显然构成一个加群. 黎曼的这个形式和思想后来被Hilbert发展成“啤酒杯、椅子和石头”, 即若记啤酒杯、椅子和石头分别为B, C, S, 则所有形式和构成的集合

$$H=\{n_1B+n_2C+n_3S | n_j\in\mathbb{Z}, 1\le j\le 3\} \tag{2.1.2}$$

[23] N.H.Abel(1802-1829), 著名挪威数学家.
[24] B. Riemann(1826-1866), 著名德国数学家. G. Roch(1837-1866), 德国数学家.

按照系数的加法构成一个加群.

最著名且非平凡的例子当属源自十七世纪牛顿等, 最终被庞加莱[25]于1901年左右研究成功的**椭圆曲线**加群, 即光滑曲线$y^2 = x^3 + ax + b$上的加法结构, 精彩、深刻且应用广泛, 请读者务必查阅相关文献以理解其大意, 留作习题1.

向量空间$\mathbb{F}^n$的元素α还可以与$\mathbb{F}$中的数a作**"数乘"**得到$\mathbb{F}^n$中的新元素, 记为$a \bullet \alpha$(简记为$a\alpha$). $\mathbb{F}^n$中的数乘满足下列四个条件(以下记$\mathbb{F}^n$为V):

(B1) 数乘的结合律: 设$a, b \in \mathbb{F}$, $\alpha \in V$, 有$a(b\alpha) = (ab)\alpha$.

(B2) 数乘关于加法的分配律: 设$a \in \mathbb{F}$, $\alpha, \beta \in V$, 有$a(\alpha + \beta) = a\alpha + a\beta$.

(B3) 数乘关于数的加法的分配律: 设$a, b \in \mathbb{F}$, $\alpha \in V$, 有$(a + b)\alpha = a\alpha + b\alpha$.

(B4) 数乘的初始条件: $1 \bullet \alpha = \alpha$, 其中$1 \in \mathbb{F}$.

定义 2.1.1. 设$(V, +)$是一个加群, 如果定义了数域$\mathbb{F}$中的数与V中元素的数乘$\bullet$(常省略不写), 则称$(V, +, \bullet)$是域$\mathbb{F}$上的**线性空间**或$\mathbb{F}$-空间, 简称V是线性空间[26].

线性空间的元素称为**向量**(即定义了加法和数乘的元素, 而非既有大小又有长度的量). 当$\mathbb{F} = \mathbb{R}$或$\mathbb{C}$时, V称为实空间或复空间. 一般地, 域$\mathbb{F}$称为线性空间V的**基域**. 本书除特别指明, 所有线性空间的基域均是指域$\mathbb{F}$, 而将$\mathbb{F}$中的元素统称为"数".

注1. 条件(B1)-(B3)保证数乘(两个不同集合间的运算)与所有"运算"(即V中的加法"$+$"、$\mathbb{F}$中的加法"$+$"与数乘)保持和谐. 条件(B4)一是为了使数乘有意义(否则, 令$1 \bullet \alpha = 0$, 则将有$a\alpha = 0$, 任意$a \in \mathbb{F}$), 二是给数乘一个自然的初始规则(读者试研究如下问题: 将条件(B4)改为$1 \bullet \alpha = \beta$将如何?).

注2. 陈建龙[27]等证明了向量空间满足的(A1)-(A4), (B1)-(B4)共8个条件不是独立的, 其中加法的交换律(A2)可由其它7个条件推出, 见文献[28].

注3. 定义2.1.1中V与$\mathbb{F}$各有一个加法, 需注意两个加法"$+$"符号相同, 但含义不同(参考下面的例2.1.2).

以下是几种常见的线性空间.

例 2.1.1. (1) 单元集$\{\alpha\}$是线性空间, 只需定义α是零向量即可, 称为**"零空间"**.

(2) 任何域$\mathbb{F}$按照其自身的加法和乘法(作为数乘)构成$\mathbb{F}$上的线性空间.

(3) 域$\mathbb{F}$上的$m \times n$矩阵全体按矩阵加法和数乘矩阵作成$\mathbb{F}$上的线性空间.

(4) $\mathbb{F}$上的一元多项式$\mathbb{F}[x]$按多项式加法和数乘多项式作成$\mathbb{F}$上的线性空间; $\mathbb{F}[x]$中次数小于n的多项式也构成$\mathbb{F}$上的线性空间, 记为$\mathbb{F}[x]_n = \{f(x) \in \mathbb{F}[x] \mid \partial f(x) < n\}$($\partial f(x)$表示$f(x)$的次数, 零多项式的次数可以约定为$-\infty$).

(5) 设A是一个$m \times n$矩阵, 则齐次线性方程组$Ax = 0$的所有解构成一个线性空间, 即A的零空间$N(A)$.

(6) 设A是任意非空集合, A到$\mathbb{F}$的所有函数(映射)的全体记为$\mathbb{F}^A$. $\mathbb{F}^A$按照普通意义下的函数加法和数乘函数构成一个线性空间, 称为集合A上的**函数空间**.

(7) 闭区间$[a, b]$上的n次可微函数全体按照函数的通常加法和数乘函数构成$\mathbb{R}$上的线性空间, 记为$C^n[a, b]$; 特别地, 全体连续函数构成的线性空间$C^0[a, b]$一般简记为$C[a, b]$.

[25] J.H.Poincare(1854-1912), 著名法国数学家、天文学家、物理学家、哲学家.

[26] 线性空间的一般定义由意大利数学家 Giuseppe Peano 于1888年给出.

[27] Jianlong Chen(1963-), 东南大学教授.

[28] 陈建龙, 张小向, 线性代数中几个疑难问题的处理, 高等数学研究, 第24卷第3期, 2021年, 28-33.

(8) 域$\mathbb{F}$上的所有形式幂级数记为$\mathbb{F}[[x]]$, 其中的加法与数乘都是按分量进行的, 即

$$\sum_{i=0}^{\infty} a_i x^i + \sum_{i=0}^{\infty} b_i x^i = \sum_{i=0}^{\infty} (a_i + b_i) x^i, \quad \lambda \sum_{i=0}^{\infty} a_i x^i = \sum_{i=0}^{\infty} \lambda a_i x^i.$$

则$\mathbb{F}[[x]]$构成$\mathbb{F}$上的线性空间, 称为**形式幂级数空间**.

下面的例子有助于理解“加法”和“数乘”的本质.

例 2.1.2. 设$V = \{$所有正实数$\}, \mathbb{F} = \mathbb{R}$是实数域. 定义$V$中的加法运算为$x \oplus y = xy$ (即通常的实数乘法); 定义V中元素与$\mathbb{F}$中数的数乘运算为$k \bullet x = x^k$(通常的幂运算). 则$(V, \oplus, \bullet)$是实线性空间, 具体验证留作习题.

本章习题3也设计了一个不同寻常的“加法”与“数乘”, 请读者仔细体会.

2.1.2 基、维数、坐标

线性空间的两种运算“加法”和“数乘”(合称为线性运算)的结合是**线性组合**, 即如下的向量:

$$k_1\alpha_1 + \cdots + k_n\alpha_n \tag{2.1.3}$$

其中$\alpha_1, \cdots, \alpha_n$是线性空间$V$中任意向量, $k_1, \cdots, k_n$是$\mathbb{F}$中的数, 称为组合系数. 如果向量α是$\alpha_1, \cdots, \alpha_n$的线性组合, 则称$\alpha$可由$\alpha_1, \cdots, \alpha_n$**线性表示**. 由此产生与$\mathbb{F}^n$中完全类似的若干基本概念, 比如**线性相关**, **线性无关**, **基**, **维数**, **坐标**等. 本章将证明, 一般线性空间与$\mathbb{F}^n$的数学结构是完全相同的, 唯一差别仅仅是元素的名称不同而已.

如果存在不全为0的组合系数 $k_1, \cdots, k_n$使得线性组合(2.1.3)为0向量, 则称$\alpha_1, \cdots, \alpha_n$**线性相关**. 否则称$\alpha_1, \cdots, \alpha_n$**线性无关**. 如果一个向量组的任意有限个向量均是线性无关的, 则该向量组称为线性无关的. 一个线性无关向量组称为是**极大的**, 如果真包含它的任意向量组均是线性相关的. 两个向量组如果可以相互线性表示, 则称它们是**等价的**.

定理 2.1.1. 线性空间中两个有限的线性无关组如果等价, 则它们所含的元素个数相等.

证 设$\alpha_1, \cdots, \alpha_m$与$\beta_1, \cdots, \beta_n$是两个等价的线性无关组. 不妨设$n \geq m$. 由等价的定义知存在$A = (a_{ij}) \in \mathbb{F}^{m\times n}$使得

$$\begin{array}{ccc} \beta_1 & = & a_{11}\alpha_1 + \cdots + a_{m1}\alpha_m \\ \beta_2 & = & a_{12}\alpha_1 + \cdots + a_{m2}\alpha_m \\ \cdots & \cdots & \cdots \\ \beta_n & = & a_{1n}\alpha_1 + \cdots + a_{mn}\alpha_m \end{array} \tag{2.1.4}$$

断言A必列满秩：设$x = (x_1, ..., x_n)^T \in \mathbb{F}^n, Ax = 0$. 则

$$\begin{aligned} x_1\beta_1 + ... + x_n\beta_n &= x_1(a_{11}\alpha_1 + \cdots + a_{m1}\alpha_m) + ... + x_n(a_{1n}\alpha_1 + \cdots + a_{mn}\alpha_m) \\ &= (a_{11}x_1 + + \cdots + a_{1n}x_n)\alpha_1 + ... + (a_{m1}x_1 + + \cdots + a_{mn}x_n)\alpha_m \\ &= (e_1^T Ax)\alpha_1 + (e_2^T Ax)\alpha_2 + ... + (e_m^T Ax)\alpha_m \end{aligned} \tag{2.1.5}$$

由于$Ax = 0$, 故$x_1\beta_1 + ... + x_n\beta_n = 0$. 但$\beta_1, \cdots, \beta_n$线性无关, 故$x_1 = ... = x_n = 0$, 即$A$列满秩, 从而$m \geq n$. 所以$m = n$. □

注1. 由该定理可知, 如果一个线性空间包含一个有限极大线性无关组, 则所有极大线性无关组均有限, 且包含的元素个数相同.

注2. 以后将等式组(2.1.4)记为

$$(\beta_1, \cdots, \beta_n) = (\alpha_1, \cdots, \alpha_m)A \tag{2.1.6}$$

当$\alpha_i, \beta_j \in \mathbb{F}^q$时, 上式即是矩阵的乘法. 一般地, 设$B \in \mathbb{F}^{n\times p}$, 则由(2.1.5)可得

$$(\beta_1, \cdots, \beta_n)B = [(\alpha_1, \cdots, \alpha_m)A]B = (\alpha_1, \cdots, \alpha_m)(AB) \tag{2.1.7}$$

故由等式(2.1.6)定义的运算满足结合律.

定义 2.1.2. 如果线性空间V中存在n个向量构成的极大线性无关组$\alpha_1, \cdots, \alpha_n$, 则称$V$是$n$维线性空间, 该向量组称为$V$的一组**基**, 向量$\alpha_j$称为**基向量**, 非负整数$n$称为$V$的维数, 记作$\dim V$(或更为准确地, $\dim_{\mathbb{F}} V$). 如果不存在这样的有限整数, 则称V是**无限维线性空间**. $n = 0$时, V没有基, 称为平凡(线性)空间或零(线性)空间, 简记为0 (注意: 符号0同时表示数字0, 向量0和矩阵0等不同对象).

推论 2.1.1. n维线性空间中任意n个线性无关的向量均构成一组基, 任意$n+1$个向量必线性相关.

例 2.1.1中线性空间(1)-(5)的基较为明显, (6)中的线性空间$\mathbb{F}^A$的基留作习题, (7)中的线性空间$C^n[a,b]$显然是无限维的, 寻找$C^n[a,b]$的一组基是非常困难的.

下面的定理给出寻找有限维线性空间的基的一种办法.

定理 2.1.2. n维线性空间中任意$r < n$个线性无关的向量均能扩充成一组基.

证 由于$r < n$, 故存在$\alpha_1 \in V$与这r个向量线性无关, 将α_1添到这些向量中去, 如果$n = r+1$, 则这$r+1$个向量构成V的一组基, 或者$r+1 < n$, 重复上述步骤即可. □

设$\alpha_1, \cdots, \alpha_n$是$n$维线性空间$V$的一组基, $\alpha \in V$是任意向量. 则由基的定义可知α可唯一地表为线性组合

$$\alpha = k_1\alpha_1 + \cdots + k_n\alpha_n. \tag{2.1.8}$$

由组合系数 $k_1, \cdots, k_n$ 确定的n元有序数组称为向量α关于基$\alpha_1, \cdots, \alpha_n$的坐标, k_j称为第j个坐标或分量.

本书总将坐标写成列向量, 即$(k_1, \cdots, k_n)^T$. 由公式(2.1.6), 等式(2.1.8)可以写成

$$\alpha = k_1\alpha_1 + + \cdots + k_n\alpha_n = (\alpha_1, \cdots, \alpha_n)x \tag{2.1.9}$$

其中$x = (k_1, \cdots, k_n)^T \in \mathbb{F}^n$.

例 2.1.3. 常微分方程$y'' - 2y' + y = 0$的全体解V是$\mathbb{R}$上的2维线性空间, 其特解$(1-x)e^x$在基e^x, xe^x下的坐标为$(1, -1)^T$, 即$(1-x)e^x = (e^x, xe^x)\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

非零线性空间都有无穷多个基, 不同的实际问题往往需要不同的基来处理, 高等数学或线性代数中熟知的**Lagrange 插值多项式**就是多项式线性空间的一组非常有用的基.

定理 2.1.3. 设$\mathbb{F}[x]_{n+1}$是次数不超过n的多项式空间, $a_i, i=1,2,\cdots,n+1$是不同的数. 对每个$i=1,2,\cdots,n+1$, 令

$$L_i(x)=\frac{(x-a_1)\cdots(x-a_{i-1})(x-a_{i+1})\cdots(x-a_{n+1})}{(a_i-a_1)\cdots(a_i-a_{i-1})(a_i-a_{i+1})\cdots(a_i-a_{n+1})}=\frac{\prod_{j=1}^{n+1}(x-a_j)}{\prod_{j=1,j\neq i}^{n+1}(a_j-a_i)} \tag{2.1.10}$$

则$L_i, i=1,2,\cdots,n+1$构成$\mathbb{F}[x]_{n+1}$的一组基. 故对任意$f(x)\in\mathbb{F}[x]_{n+1}$, 有Lagrange插值公式

$$f(x)=\sum_{i=1}^{n+1}f(a_i)L_i(x). \tag{2.1.11}$$

公式(2.1.10)中的n次多项式$L_i(x)$称为Lagrange插值多项式.

证 只需证明公式(2.1.11)成立, 由此可知$L_i, i=1,2,\cdots,n+1$是$\mathbb{F}[x]_{n+1}$的秩为$n+1$的向量组, 由于$\mathbb{F}[x]_{n+1}$是$n+1$维线性空间, 由定理2.1.2可知其是一组基. 记(2.1.11)右端的多项式为$g(x)$. 由于$L_i(a_j)=\delta_{ij}$(这是Lagrange插值多项式的关键!), 故$g(a_i)=f(a_i), 1\le i\le n+1$, 因此诸$a_i$均为$n$次多项式$f(x)-g(x)$的根. 由于这$n+1$个$a_i$互异, 而$n\ge 1$次多项式方程在复数域只有$n$个根, 故$f(x)-g(x)$只能是0多项式. □

2.1.3 线性空间的同构

表达式(2.1.9)的另一个重要意义是建立了n维线性空间V和$\mathbb{F}^n$之间的天然联系. 取定V的一组基, 对$\forall\alpha\in V$, 设α的坐标为$x\in\mathbb{F}^n$, 定义

$$\sigma:\alpha\mapsto x \tag{2.1.12}$$

容易验证σ是满足下述性质的映射:

(L1) σ是**单映射**(简称**单射**), 即对任意$\alpha,\beta\in V,\sigma(\alpha)=\sigma(\beta)$, 则$\alpha=\beta$.

(L2) σ是**满映射**(简称**满射**), 即对任意$y\in\mathbb{F}^n,\exists\beta\in V$使得$\sigma(\beta)=y$.

(L3) σ是线性的, 即对任意$\alpha,\beta\in V, a,b\in\mathbb{F}$, 有$\sigma(a\alpha+b\beta)=a\sigma(\alpha)+b\sigma(\beta)$.

满足条件(L3)的映射称为**线性映射**或者**线性变换**.

一般将U到自身的线性映射称为线性算子(本书不区分线性变换、线性映射和线性算子), U到V的线性变换全体记为$\mathrm{Hom}_{\mathbb{F}}(U,V)$或简记为$\mathrm{Hom}(U,V)$. 特别, 将$\mathrm{Hom}(V,V)$记为$\mathrm{End}\,V$, 而将$\mathrm{Hom}(V,\mathbb{F})$记为$V^*$, 称为$V$的**对偶空间**或**共轭空间**.

例 2.1.4. 以实数域$\mathbb{R}$为例, 设f是从$\mathbb{R}$到$\mathbb{R}$的线性映射, 则由(L3)知, 对任意$x\in\mathbb{R}, f(x)=xf(1)$, 于是$f(x)=kx$, 故$\mathbb{R}$的对偶空间$\mathrm{Hom}_{\mathbb{R}}(\mathbb{R},\mathbb{R})=\mathbb{R}^*$就是线性函数, 其几何意义是全体通过原点的直线. 现若将f的定义域$\mathbb{R}$变为$\mathbb{R}^2$, 值域仍为$\mathbb{R}$, 则容易得到(请计算!)$f(x,y)=ax+by$, 其中$a=f(1,0)$而$b=f(0,1)$. 这是二元线性函数.

既单又满的映射称为**双射**. 设$\sigma\in\mathrm{Hom}(U,V)$. 则当$\sigma$是单射(或满射)时, 称$\sigma$是单变换 (或满变换). 既单又满的变换称为**同构**. 如果存在同构$\sigma\in\mathrm{Hom}(U,V)$, 则$U$与$V$称为是**同构的线性空间**, 记为$U\simeq V$.

因此, 公式(2.1.12)中的σ是V与$\mathbb{F}^n$之间的一个同构.

简单地说, (L1)与(L2)是说V与$\mathbb{F}^n$作为集合大小相同, 而(L3)则是说V与$\mathbb{F}^n$具有相**同**的结**构**, 即它们的"线性运算"相同. 比如, n-维线性空间V与$\mathbb{F}^n$作为线性空间本质上是一样的!

下面的**"同构定理"**是线性空间分类的基本定理.

定理 2.1.4. 域$\mathbb{F}$上的两个线性空间U与V同构 $\Longleftrightarrow$ $\dim_{\mathbb{F}} U=\dim_{\mathbb{F}} V$.

证明思路. 必要性是显然的, 因为同构变换必然将U的一组基变为V的一组基, 因此U与V的维数相同. 反之, 若U与V的维数相同, 则将U的一组基变为V的一组基的线性变换必然既单又满, 故是同构. □

定理2.1.4是线性空间理论中最重要的定理之一, 它表明, 任何域$\mathbb{F}$(不必是复数域的子域)上的任意n维线性空间V均与$\mathbb{F}^n$同构, 唯一差别仅是元素命名不同, 从而对V的研究可以归结到对$\mathbb{F}^n$的研究, 而后者是熟知的最简单的线性空间.

例 2.1.5. 设$U=\mathbb{F}[x]_n, V=\mathbb{F}^n$(为简便起见, 本例$\mathbb{F}^n$中的向量记为行向量). 因为它们的维数均为$n$, 故由定理2.1.4知$U$与$V$同构. 这就是说, U与V的差别仅是U中的向量被称为"多项式"而已. 考虑U与V的如下同构映射σ:

$$\sigma:\ f(x)=\sum_{i=0}^{n-1} a_i x^i \mapsto (a_0,\, a_1, \cdots, a_{n-1}).$$

则σ不过是一个"重起名字"的操作, 即将U中的向量$f(x)=\sum\limits_{i=0}^{n-1} a_i x^i$ 重新命名为$(a_0,\, a_1, \cdots, a_{n-1})$, 比如, 多项式1与$x$的和现在变为向量$(1,0,\cdots 0)$与$(0,1,0,\cdots,0)$的和, 前者的结果是$1+x$, 后者是$(1,1,0,\cdots,0)$. 所以, U与V的线性运算是一回事. 为什么我们心中的多项式与n元数组有很大差别? 那是因为我们给多项式附加了乘法等"非线性运算"性质, 但这些性质并不能由多项式空间的加法和数乘导出.

例 2.1.6. 设A是非空有限集合, $|A|=n$. 则$\mathbb{F}^A\cong\mathbb{F}^n$; 故函数空间$\mathbb{F}^A$的(线性)结构与我们熟悉的线性空间$\mathbb{F}^n$完全一致.

例 2.1.7. 例2.1.2中的实线性空间$V=\{$全体正实数$\}$是几维的? 注意, 由于V不是$\mathbb{R}$的子空间, 故不能直接推出$\dim\, V\le 1$. 但容易证明映射$x\mapsto \log x$是V到$\mathbb{R}$的一个线性同构($\log x$可以是线性的！), 故由定理2.1.4知$\dim\, V=\dim\,\mathbb{R}=1$. 于是, 正实数的乘法运算就是实数的加法运算, 而正实数的幂就是普通的实数乘法(此时每一个正实数x都有一个新名字$\log x$).

线性代数最著名的教材当属G. Strang[29]的"Introduction to Linear Algebra", 全书几乎只字不提抽象线性空间, 完全聚焦于$\mathbb{R}^n$, 就是因为从同构的观点看, 任何有限维实线性空间的本质与某$\mathbb{R}^n$完全相同.

2.1.4 过渡矩阵与坐标变换

设V是n维线性空间, $\alpha_1,\cdots,\alpha_n$与$\beta_1,\cdots,\beta_n$是V的两组基, 分别称为α-基与β-基. 两组基当然等价, 故由(2.1.9)知存在n阶矩阵$P=(p_{ij})$使得

$$(\beta_1,\cdots,\beta_n)=(\alpha_1,\cdots,\alpha_n)P \tag{2.1.13}$$

[29]G. Strang(1934-), 美国科学院院士, 著名数学家.

矩阵P称为由α-基到β-基的**过渡矩阵**. 容易证明, 过渡矩阵必是可逆矩阵(留作习题). 显然, 由β-基到α-基的过渡矩阵为P^{-1}.

设$\gamma \in V$关于α-基与β-基的坐标分别为$x = (x_1, \cdots, x_n)^T$与$y = (y_1, \cdots, y_n)^T$. 由于$\gamma = y_1\beta_1 + \cdots + y_n\beta_n = (\beta_1, \cdots, \beta_n)y$, 故由公式(2.1.13)(注意利用公式(2.1.9)的结合律)可得

$$\gamma = [(\alpha_1, \alpha_2, \cdots, \alpha_n)P]y = (\alpha_1, \alpha_2, \cdots, \alpha_n)(Py).$$

由坐标的唯一性可知,

$$x = Py, \quad 或 \quad y = P^{-1}x. \tag{2.1.14}$$

公式(2.1.14)称为**坐标变换公式**.

例 2.1.8. 设$V = \mathbb{R}^2$, 取$\alpha_1 = (1, 0)^T$, $\alpha_2 = (0, 1)^T$与$\beta_1 = (1, 1)^T$, $\beta_2 = (-1, 1)^T$, 求由α-基到β-基的过渡矩阵以及$\gamma = (2, 3)^T$在β-基下的坐标.

解 由于α-基是标准基, 易知由α-基到β-基的过渡矩阵为

$$P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = (\beta_1, \beta_2).$$

因此, $\gamma = (2, 3)^T$在β–基下的坐标为$P^{-1}(2, 3)^T = (\dfrac{5}{2}, \dfrac{1}{2})^T$.

例 2.1.9. 求$\forall f(x) \in \mathbb{F}[x]_n$在标准基$1, x, \cdots, x^{n-1}$下的坐标. 向量组$1, x - a, \cdots, (x - a)^{n-1}$也构成$\mathbb{F}[x]_n$的一组基, 求$f(x)$在该组基下的坐标. 求从标准基到第二组基的过渡矩阵.

解 由 Taylor[30] (泰勒)公式可知:

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} x^k,$$

即$f(x)$在标准基下的坐标为

$$(f(0), f'(0), \cdots, \frac{f^{(k)}(0)}{k!}, \cdots, \frac{f^{(n-1)}(0)}{(n-1)!})^T.$$

对任意$a \in \mathbb{F}$, 仍由 Taylor 公式 可知, $f(x)$在第二组基下的坐标为

$$(f(a), f'(a), \cdots, \frac{f^{(k)}(a)}{k!}, \cdots, \frac{f^{(n-1)}(a)}{(n-1)!})^T.$$

由于

$$(x - a)^k = \sum_{r=0}^{k} C_k^r x^r (-a)^{k-r},$$

故从标准基到第二组基的过渡矩阵$P = (p_{ij})$由下式给出:

$$p_{ij} = \begin{cases} C_{j-1}^{i-1}(-a)^{j-i} & i \le j, \\ 0 & i > j. \end{cases}$$

[30]Brook Taylor(1685-1739), 英国数学家.

思考题

1. 线性空间的条件(B4)即$1\bullet\alpha=\alpha$是否可以改为其它形式?

2. 线性空间中, 任何数x与零向量0的数乘是否仍为零向量? 数0与任何向量α的数乘是否仍为零向量?

3. 设$u=u(x,y,z,t)$是未知函数, c是常数, $\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$是**Laplace**[31] **算符**. **波动方程** $\frac{\partial^2 u}{\partial t^2}=c^2\nabla^2 u$的全体解是否构成线性空间? 若$u$与时间$t$无关, 则波动方程变为**Laplace 方程** $\nabla^2 u=0$. 该方程的全体解是否构成线性空间? 总结之.

4. 试给出基与基向量一个直观的解释.

5. 试给出过渡矩阵的一种直观解释.

第二节 子空间、直和分解与分块矩阵

2.2.1 子空间

复数域$\mathbb{C}$是实数域$\mathbb{R}$上的2维线性空间, 实数域作为复数域的子集自身是实数域上的1维线性空间, 这时我们称$\mathbb{R}$是$\mathbb{C}$的**子空间**. 一般地, 有下述定义.

定义 2.2.1. 设V是一个线性空间, U是V的一个非空子集. 如果U本身关于V的加法与数乘作成一个线性空间, 则称U是V的一个线性子空间或子空间.

任何非零线性空间都有两个平凡子空间, 即零子空间$\{0\}$与它自身. 其余的子空间称为**真子空间**. 零子空间常简记为0.

注意子空间必然包含0向量(在平面或空间解析几何中, 就是子空间要过原点).

例 2.2.1. 通常向量空间$\mathbb{R}^2$的所有真子空间即通过原点的直线, 而向量空间$\mathbb{R}^3$的所有真子空间即通过原点的直线与平面.

例 2.2.2. (1) 设$A\in\mathbb{F}^{m\times n}$. 则$A$的零空间和行空间均是向量空间$\mathbb{F}^n$的子空间. 若$m=n$, 则$A$的特征子空间均是向量空间$\mathbb{C}^n$的子空间.

注意子空间的定义强调"关于V的加法与数乘", 比如$\mathbb{R}$是$\mathbb{C}$的子集且本身是线性空间, 因此$\mathbb{R}$是$\mathbb{C}$的实线性子空间, 但$\mathbb{R}$却不是复线性空间$\mathbb{C}$的子空间, 因为$\mathbb{R}$关于$\mathbb{C}$的数乘并不封闭.

例 2.2.3. 设$U=\{(1,y)\in\mathbb{R}^2\,|\,y\in\mathbb{R}\}$, 定义$U$上的加法"$\heartsuit$"与数乘"$\spadesuit$"分别为:

$$(1,y)\heartsuit(1,y')=(1,y+y');\quad a\spadesuit(1,y)=(1,ay).$$

则显然$(U,\heartsuit,\spadesuit)$构成$\mathbb{R}$上的1维线性空间. 注意$U$是一条不过原点的直线, 因此不是通常向量空间$\mathbb{R}^2$的子空间.

例 2.2.4. $\mathbb{F}[x]_n$是多项式空间$\mathbb{F}[x]$的一个n维子空间. 若把实系数多项式也看成闭区间$[a,b]$上的函数, 则$\mathbb{R}[x]$与$\mathbb{R}[x]_n$都是$[a,b]$上的连续函数空间$C[a,b]$的子空间.

由定理2.1.2, 有限维线性空间V的子空间U仍是有限维的, 且U的维数不超过V的维数; 且若二者维数相等, 则它们本身也是相等的.

下述简单易行的子空间判别准则(证明很简单, 留作习题)告诉我们, 一个非空子集是子空间当且仅当它关于加法与数乘均封闭.

[31] Pierre-Simon Laplace(1749-1827), 著名法国数学家, 物理学家, 天文学家, 被称为法国的Newton(牛顿).

定理 2.2.1. (子空间判别法) 设U是线性空间V的一个非空子集. 则U是子空间 $\iff$ 对任意$\lambda \in \mathbb{F}$, $\alpha, \beta \in U$, 有$\alpha + \beta \in U$与$\lambda\alpha \in U$.

例 2.2.5. 实数域$\mathbb{R}$作为有理数域$\mathbb{Q}$上的线性空间, 其有理数子集$\mathbb{Q}$当然是子空间; 但无理数子集不是子空间, 因为加法与数乘均不封闭; 整数子集虽然加法封闭但数乘不封闭, 故也不是子空间; 所有$\sqrt{2}$的有理数倍构成的集合是子空间, 因为加法与数乘皆封闭.

思考题

1. 过原点的二次曲线不是$\mathbb{R}^2$的子空间. 能否构造适当的线性运算使得$y = x^2$构成线性空间?
2. 有理数域$\mathbb{Q}$不是实数域作为实线性空间的子空间. 能否构造适当的线性运算使得$\mathbb{Q}$构成实线性空间?
3. 整数环$\mathbb{Z}$不是$\mathbb{Q}$作为$\mathbb{Q}$-空间的子空间. 能否构造适当的线性运算使得$\mathbb{Z}$构成$\mathbb{Q}$-线性空间?

2.2.2 子空间的运算：交、和与直和

由定理2.2.1容易直接验证子空间的下述性质.

命题 2.2.1. (1) 传递性: 即若U是V的子空间, W是U的子空间, 则W也是V的子空间.

(2) 任意多个(可以无限)子空间的交集仍是子空间, 称为这些子空间的交, 且是含于这些子空间的最大子空间; 特别, 两个子空间U与W的交$U \cap W$仍是子空间.

两个子空间的并集$U \cup W$一般不是子空间(为什么?). 但包含U与W的最小子空间存在. 这样的子空间必须包含所有形如$\alpha + \beta$, $\alpha \in U$, $\beta \in W$的向量. 将全体这样的向量构成的子集合记为$U + W$, 由定理2.2.1可知$U + W$是子空间, 且是包含U与W的最小的子空间, 称为U与W的**和**.

子空间的和的概念可以推广到任意有限多个子空间的情形.

设$U_1, U_2, \cdots, U_s$是线性空间V的子空间, 则集合

$$\{\alpha_1 + \alpha_2 + \cdots + \alpha_s \,|\, \alpha_j \in U_j,\, 1 \leq j \leq s\}$$

也是子空间(为什么?), 称为$U_1, U_2, \cdots, U_s$的和, 记为$U_1 + U_2 + \cdots + U_s$或$\sum\limits_{j=1}^{s} U_j$.

设V是线性空间, $S \subset V$. 称V的包含S的最小子空间为由S生成(或张成)的子空间, 记为$\operatorname{Span} S$, S称为$\operatorname{Span} S$的生成元集. 显然, 当$S = \emptyset$或$S = \{0\}$时, $\operatorname{Span} S = 0$是零维子空间; 若$S = \{\alpha\}$是一元集且$\alpha \neq 0$, 则$\operatorname{Span}\{\alpha\} = \{k\alpha \,|\, k \in \mathbb{F}\}$是一维子空间, 记为$\mathbb{F}\alpha$; 若$S = \{\alpha_1, \alpha_2, \cdots, \alpha_s\}$是有限集, 则记$\operatorname{Span} S = \operatorname{Span}\{\alpha_1, \alpha_2, \cdots, \alpha_s\}$, 此时有

$$\operatorname{Span}\{\alpha_1, \alpha_2, \cdots, \alpha_s\} = \{k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s \,|\, k_j \in \mathbb{F}, 1 \leq j \leq s\}.$$

一般地, 直接验证可知,

$$\operatorname{Span} S = \{k_1\alpha_1 + k_2\alpha_2 + \cdots + k_s\alpha_s \,|\, \alpha_j \in S, k_j \in \mathbb{F}, s \geq 0\},$$

即$\operatorname{Span} S$由S中元素的所有可能的(有限)线性组合构成. 特别地, 两个子空间U与W的和$U + W = \operatorname{Span} U \cup W$.

例 2.2.6. 设k是正整数, $A \in \mathbb{F}^{n\times n}, \beta \in \mathbb{F}^n$. 子空间$\operatorname{Span}\{\beta, A\beta, A^2\beta, ..., A^{k-1}\beta\}$称为$A$的**Krylov子空间**, 相应的矩阵$(\beta, A\beta, A^2\beta, ..., A^{k-1})$称为**Krylov矩阵**, 一般记为$K(A, \beta, k)$. Krylov子空间与Krylov矩阵均具有重要的应用.

容易直接验证子空间的交与和两种运算满足下述性质.

命题 2.2.2. 设X, Y, Z均是V的子空间, 0表示V的零子空间. 则
(1) 交换律：$X \bigcap Y = Y \bigcap X$; $X + Y = Y + X$.
(2) 结合律：$(X \bigcap Y) \bigcap Z = X \bigcap (Y \bigcap Z)$; $(X + Y) + Z = X + (Y + Z)$.
(3) 零元素：$0 \bigcap X = 0$; $0 + X = X$.

这就是说线性空间V的所有子空间构成的集合本身关于交与和均具有“群胚”的数学结构. 实际上, 一个线性空间V的所有子空间构成的集合在交与和两种运算下, 构成一个“格”(即任何两个元素存在**上确界**与**下确界**), 称为V的“子空间格”. 子空间格具有重要的理论意义和广泛的应用价值.

例 2.2.7. 设$V = M_n(\mathbb{R})$是n阶实矩阵构成的线性空间. 令$D = \{$全体对角矩阵$\}$, $U = \{$全体上三角矩阵$\}$, $W = \{$全体下三角矩阵$\}$. 则D, U与W均是V的子空间, 且$U + W = V$; $U \cap W = D$; 但$U \cup W$不是子空间(为什么?). 注意D的维数等于n; U与W的维数均等于$n(n+1)/2$; $U + W = V$的维数等于n^2. 故有

$$(\dim U + \dim W) - \dim (U + W) = \dim (U \cap W). \tag{2.2.1}$$

请比较该公式与普通集合的计数公式: $|A \cup B| = |A| + |B| - |A \cap B|$, 其中$|A|$表示$A$的元素个数.

例 2.2.8. 设$V = \mathbb{R}[x]_n$, $U = \{f(x) \in V \mid f(1) = 0\}$, $W = \{f(x) \in V \mid f(2) = 0\}$. 则$U + W = V$, $U \cap W = \{f(x) \in V \mid f(1) = 0 \text{ 且 } f(2) = 0\}$. 但$U \cup W = \{f(x) \in V \mid f(1) = 0 \text{ 或 } f(2) = 0\}$不是子空间(比如, $x - 1 + x - 2 \notin U \cup W$). 容易计算, $\dim U = n - 1 = \dim W$, $\dim (U \cap W) = n - 2$. 故仍有(2.2.1)式.

定理 2.2.2. (维数公式) 设V是线性空间, U与W是V的两个子空间. 则

$$\dim (U + W) = (\dim U + \dim W) - \dim (U \cap W). \tag{2.2.2}$$

证 证明的思路是**“中心开花”**, 即将$U \cap W$的基扩展成U与W的各一组基. 设$\dim U = s$, $\dim W = t$, $\dim (U \cap W) = r$. 任取$U \cap W$的一组基$\alpha_1, \alpha_2, \cdots, \alpha_r$. 由于$U \cap W$是$U$与$W$的公共子空间, 故$U \cap W$的基是$U$与$W$的线性无关的向量组, 因此可以扩充成$U$或$W$的基. 设

$$\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_{r+1}, \beta_{r+2}, \cdots, \beta_s$$

与

$$\alpha_1, \alpha_2, \cdots, \alpha_r, \gamma_{r+1}, \gamma_{r+2}, \cdots, \gamma_t$$

分别是U与W的基. 则容易验证

$$\alpha_1, \alpha_2, \cdots, \alpha_r, \beta_{r+1}, \beta_{r+2}, \cdots, \beta_s, \gamma_{r+1}, \gamma_{r+2}, \cdots, \gamma_t$$

是$U + W$的一组基. □

维数定理的极端情形是$U \cap W = 0$, 此时我们称和$U + W$是**直和**, 记为$U \oplus W$. 注意：子空间的直和记号与分块矩阵的直和记号$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = A \oplus B$完全一致, 其内在原因将在下一小节阐明.

例 2.2.9. 二维平面$\mathbb{R}^2$是x轴与y轴(均是 1 维子空间)的直和. 类似地, $\mathbb{R}^3$是x轴, yoz平面(这是一个 2 维子空间)的直和.

定理 2.2.3. 设U与W是线性空间V的两个子空间, 则下列命题等价:

(1) $U+W$是直和(即$U\cap W=0$).

(2) 对任意$\alpha\in U+W$, 分解式$\alpha=u+w$, 其中$u\in U,\ w\in W$是唯一的, 即若还有$\alpha=u'+w'$, 则$u=u',\ w=w'$.

(3) 零向量的分解式唯一; 即若$0=u+w,\ u\in U,\ w\in W$, 则$u=w=0$.

(4) $\dim(U+W)=\dim U+\dim W$.

注: 一般将定理2.2.3(3)作为直和的定义.

证 由直和的定义, (1)与(4)是等价的, 而(2)显然蕴涵(3). 故只需证明(3)$\Longrightarrow$(1)$\Longrightarrow$(2).

现设(3)成立, 而$\alpha\in U\cap W$, 则$-\alpha\in U\cap W$, 因此$0=\alpha+(-\alpha)$, 但零向量的分解式唯一, 故$-\alpha=\alpha=0$, 即$U\cap W=0$, 故(1)成立.

如果(1)成立, 而$\alpha\in U+W$有两个分解式$\alpha=u+w=u'+w'$, 其中$u,u'\in U,\ w,w'\in W$, 则$u-u'=w'-w$, 但$u-u'\in U, w'-w\in W$, 从而$u-u'=w'-w\in U\cap W=0$, 即$u=u',w'=w$, 故(2)成立. □

可以定义任意有限多个子空间的直和, 即设$W_1,W_2,\cdots,W_s$是线性空间V的子空间, 满足条件$W_j\cap\sum_{k\neq j}W_k=0, 1\le j\le s,\ 1\le k\le s$, 则定义这些子空间的直和为

$$W_1\oplus W_2\oplus\cdots\oplus W_s=\{w_1+w_2+\cdots+w_{s-1}+w_s|w_i\in W_i,1\le i\le s\}. \tag{2.2.3}$$

类似两个子空间的直和的判定, 有下述多子空间直和的判定定理, 其证明留做习题18.

定理 2.2.4. 设$W_1,W_2,\cdots,W_s$是线性空间V的子空间, 则下列命题等价:

(1) $W_1+W_2+\cdots+W_s$是直和即$W_j\cap\sum_{k\neq j}W_k=0,1\le j\le s,\ 1\le k\le s$.

(2) $\dim(W_1+W_2+\cdots+W_s)=\dim W_1+\dim W_2+\cdots+\dim W_s$.

(3) 任意向量$\alpha\in W_1+W_2+\cdots+W_s$的分解式唯一.

(4) 零向量的分解式唯一.

例 2.2.10. 每个有限维线性空间均可以分解成1维子空间的直和. 这是因为设$\alpha_1,\alpha_2,\cdots,\alpha_n$是$n$维线性空间$V$的一组基, 则$V=\mathbb{F}\alpha_1\oplus\mathbb{F}\alpha_2\oplus\cdots\oplus\mathbb{F}\alpha_n$.

例 2.2.11. n阶矩阵空间是严格上三角矩阵子空间, 严格下三角矩阵子空间与对角矩阵子空间的直和.

设V是线性空间, U是V的一个子空间. 则存在另一个子空间W使得$V=U\oplus W$(对有限维空间, 此仅需将U的一组基扩充成V的一组基即可, 新扩充的部分生成的子空间即是一个W; 对无限维空间, 需要处理"无限"的工具即所谓"选择公理"). W称为U的**补子空间**. 显然U的补子空间一般不是唯一的(什么时候唯一?).

例 2.2.12. 设$V=\mathbb{R}^3$. 则V的1维子空间就是通过原点的所有直线; 2维子空间就是通过原点的所有平面. 任何一个1维子空间的补子空间可以是任意一个不含该1维子空间的2维子空间. 反之亦然.

例 2.2.13. 设$A\in\mathbb{R}^{m\times n}$. 则$\mathbb{R}^n=N(A)\oplus R(A^T),\mathbb{R}^m=N(A^T)\oplus R(A)$. 故列空间与左零空间互补, 行空间与零空间互补.

2.2.3 直和分解与分块矩阵

子空间的直和与矩阵的直和$\begin{pmatrix} A & 0\\ 0 & B\end{pmatrix}=A\oplus B$有密切联系.

定理 2.2.5. 设$A \in \mathbb{F}^{m\times n}, B \in \mathbb{F}^{p\times q}$. 则

(1) $R(A\oplus B) = R(\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}) \oplus R(\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}) = R(\begin{pmatrix} A \\ 0 \end{pmatrix}) \oplus R(\begin{pmatrix} 0 \\ B \end{pmatrix})$.

(2) $N(A\oplus B) = N(\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}) \oplus N(\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}) = N(\begin{pmatrix} A & 0 \end{pmatrix}) \oplus N(\begin{pmatrix} 0 & B \end{pmatrix})$.

(3) 设$C \in \mathbb{F}^{m\times p}, D \in \mathbb{F}^{n\times p}$. 则$N(\begin{pmatrix} C \\ D \end{pmatrix}) = N(C)\bigcap N(D)$.

(4) 设$E \in \mathbb{F}^{m\times p}, F \in \mathbb{F}^{m\times q}$. 则$R(\begin{pmatrix} E & F \end{pmatrix}) = R(E) + R(F)$.

证 (1) $R(A\oplus B) = R(\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}) \subseteq R(\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}) + R(\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}) = R(\begin{pmatrix} A \\ 0 \end{pmatrix}) \oplus R(\begin{pmatrix} 0 \\ B \end{pmatrix})$. 但左右两端子空间的维数均为$r(A)+r(B)$, 故相等.

其余三条的证明类似, 请读者自行验证. □

分块矩阵可以简化计算的本质是降维, 因为相应的线性空间可以分解成真子空间的直和, 从而降低了计算对象的维数.

思考题

1. 实数域$\mathbb{R}$作为实线性空间的所有子空间是什么? 作为有理数域上的线性空间呢?
2. 复数域$\mathbb{C}$作为实线性空间的所有子空间是什么? 作为复数域上的线性空间呢?
3. 两个向量张成的子空间的几何意义是什么?
4. 两个子空间的交, 并与和的几何意义分别是什么?
5. 设$\mathbb{F} = \mathbb{C}$或$\mathbb{R}$. 则$\mathbb{F}$上的n元二次型全体构成$\mathbb{F}$上的线性空间. 全体半正定二次型是否是该线性空间的子空间? 全体不定二次型呢?
6. 设$\mathbb{F} = \mathbb{C}$或$\mathbb{R}$. 则$\mathbb{F}$上的n维双线性型全体构成$\mathbb{F}$上的线性空间. 全体n维对称双线性型是否是该线性空间的子空间?

第三节 线性变换与矩阵

2.3.1 线性变换的定义与例子

线性变换$\sigma : U \longrightarrow V$满足的条件(L3)可分拆为下述两条简单性质:

(T1) (**可加性**) $\sigma(x+y) = \sigma(x) + \sigma(y)$.
(T2) (**齐次性**) $\sigma(ax) = a\sigma(x)$, $a \in \mathbb{F}$.

本节将证明, 如果映射$\tau : \mathbb{F}^n \to \mathbb{F}^m$是线性的, 则必然存在矩阵$A_{m\times n}$使得$\sigma(x) = Ax, \forall x \in \mathbb{F}^n$. 于是, 求向量$b \in \mathbb{F}^m$在$\sigma$下的原像等价于解线性方程组$Ax = b$. 因此, 矩阵$A$和线性变换或线性映射 是同一事物的两种表达.

注 1. 可加性条件(T1)与齐次性条件(T2)等价于"线性叠加原理"

$$\sigma(a_1\alpha_1 + \cdots + a_s\alpha_s) = a_1\sigma(\alpha_1) + \cdots + a_s\sigma(\alpha_s), a_j \in \mathbb{F},\ \alpha_j \in U. \tag{2.3.1}$$

一般称函数或映射f是r-齐次的是指存在固定的常数r使得对任意的x均有$f(kx) = k^r f(x)$. 但如果f还满足可加性条件, 则r-齐次性必是齐次性, 即有$r=1$, 请读者自证.

注 2. 如果$\mathbb{F}$是有理数域$\mathbb{Q}$, 则容易证明可加性蕴涵齐次性. 因为由(T1)可知, 对任意正整数n, m有

$$f(2x) = f(x+x) = 2f(x),\ f(nx) = f(x+(n-1)x) = nf(x).$$

故有

$$f(nx) = f(m \times (\frac{n}{m}x)) = mf(\frac{n}{m}x),$$

即可得

$$f(\frac{n}{m}x) = \frac{n}{m}f(x),$$

即对任意正有理数r有

$$f(rx) = rf(x).$$

但由$f(0) = f(0+0) = f(0) + f(0)$, 故有$f(0) = 0$, 从而$0 = f(x + (-x)) = f(x) + f(-x)$, 即有$f(-x) = -f(x)$, 故知条件T(2)对所有有理数成立.

但是, 即使$\mathbb{F}$是有理数域$\mathbb{Q}$, 齐次性条件(T2)也不蕴涵可加性条件(T1).

一般将满足可加性的映射称为加性映射, 对非有理数域的其它数域$\mathbb{F} \subseteq \mathbb{C}$, 加性映射未必是齐次的, 如下例.

例 2.3.1. 设$\mathbb{F} = \mathbb{Q}(\sqrt{2} = \{a + b\sqrt{2} | a, b \in \mathbb{Q}\,)$, 则数域$\mathbb{F}$是其自身上的1维线性空间. 对任意$\alpha = a + b\sqrt{2} \in \mathbb{F}, a, b \in \mathbb{Q}$, 定义

$$\sigma(\alpha) = \sigma(a + b\sqrt{2}\,) = a,$$

则σ显然是$\mathbb{F}$到自身的一个加性映射, 但σ不满足齐次性条件, 因为

$$\sigma(\sqrt{2} \bullet \sqrt{2}\,) = \sigma(2) = 2 \text{ 而 } \sqrt{2}\sigma(\sqrt{2}\,) = 0.$$

例 2.3.2. 考虑平面坐标系中坐标轴的旋转. 假定坐标轴顺时针旋转了θ角度, 则原坐标系中的点$P(x, y)$将变为$P'(x', y')$. 两种坐标间的关系由所谓**转轴公式**表达:

$$\begin{aligned} x = &\quad x'\cos\theta + y'\sin\theta, \\ y = &\ -x'\sin\theta + y'\cos\theta. \end{aligned} \tag{2.3.2}$$

易知, 如果点P与Q分别被变到点P'与Q', 则点$P + Q$与kP (平面可以定义向量加法与数乘从而构成线性空间!)将被分别变到点$P' + Q'$与kP'! 现将任意一个点$P(x, y)$记成α, 旋转后的点记成$f(\alpha)$, 则有

$$f(\alpha + \beta) = f(\alpha) + f(\beta); \quad f(k\alpha) = kf(\alpha).$$

因此旋转变换是线性变换. 注意, 转轴公式中的两个等式各自表示一个二元线性函数, 即平面的旋转变换是两个二元线性函数.

例 2.3.3. 考察定义在$\mathbb{R}$上的全体无限次可微函数的集合V(这是无限维实线性空间)的求导运算$\partial : f(x) \mapsto f'(x)$, 它是$V$到自身的一个映射. 由高等数学或数学分析,

$$\partial(f + g) = \partial(f) + \partial(g); \ \ \partial(kf) = k\partial(f).$$

故求导运算是线性变换. 类似地, 变上限的定积分$f(x) \mapsto \int_a^x f(x)\,\mathrm{d}\,x$也是线性变换.

例 2.3.4. 考察n阶线性常微分方程

$$y^{(n)} + f_1(x)y^{(n-1)} + f_2(x)y^{(n-2)} + \cdots + f_{n-1}(x)y' + f_n(x)y = 0, \tag{2.3.3}$$

其中$y^{(i)} = \frac{\mathrm{d}^{\,i} y}{\mathrm{d}\,x^i}$, $i = 1, 2, \cdots, n$. 如果将方程的左边记为$L(y)$, 则$L : y \mapsto L(y)$给出全体$\geq n$阶可导函数之集合到全体实函数之集合的映射(二者均是无限维实线性空间), 显然L满足(T1)与(T2), 即是线性变换. 微分方程(2.3.3)的解恰好是映射L的“零点”集$\{y : L(y) = 0\}$.

例 2.3.5. 考虑函数$f(x)$的Laplace变换

$$L(f(x)) = \int_{-\infty}^{+\infty} f(t)e^{-ixt}\,\mathrm{d}\,t.$$

则L是线性变换.

线性变换是最简单的映射, 通常的映射都是非线性变换, 如平面上的极坐标变换

$$\begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} \tag{2.3.4}$$

不是线性变换. 类似地, 空间中的柱面坐标变换与球面坐标变换也不是线性变换.

由定义立即可得线性变换如下的简单性质.

命题 2.3.1. (1) $\sigma(0)=0$; $\sigma(-\alpha)=-\sigma(\alpha)$.
(2) 若$\alpha_1, \alpha_2, \cdots, \alpha_s$线性相关, 则$\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_s)$ 也线性相关.
(3) 若$\sigma(\alpha_1), \sigma(\alpha_2), \cdots, \sigma(\alpha_s)$线性无关, 则$\alpha_1, \alpha_2, \cdots, \alpha_s$也线性无关.

注意性质$\sigma(0)=0$的几何意义就是线性变换保持原点不动, 因此平面或空间解析几何中非恒等的平移变换不是线性变换.

下面的定理给出线性变换的构造.

定理 2.3.1. 设$\alpha_1, \alpha_2, \cdots, \alpha_n$是线性空间$U$的一组基, $\beta_1, \beta_2, \cdots, \beta_n$是线性空间$V$的任意$n$个向量, 则唯一地存在一个线性变换$\sigma$使得$\sigma(\alpha_j)=\beta_j$, $1\le j\le n$.

证 证明思路. 如果线性变换σ满足条件, 则σ必须将U中任意向量$\alpha = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n$映到

$$\sigma(\alpha) = k_1\sigma(\alpha_1) + k_2\sigma(\alpha_2) + \cdots + k_n\sigma(\alpha_n) = k_1\beta_1 + k_2\beta_2 + \cdots + k_n\beta_n.$$

容易验证, 如上定义的映射σ确是一个U到V的满足条件$\sigma(\alpha_j)=\beta_j$, $1\le j\le n$的线性变换, 且是唯一可能满足这些条件的线性变换. □

注 1. 上述证明中构造线性变换σ的方法是基本的, 称为**线性扩展法**, 即确定一组基的像, 再对任意的向量, 定义其像就是相应于基元素的像的线性组合.

注 2. 定理2.3.1告诉我们, 要确定一个定义域为n维线性空间的线性变换, 只需要知道该线性变换在一组基下的像即可. 对照：一个一元n次多项式被其在$n+1$个不同点的值确定.

例 2.3.6. 一个从$\mathbb{R}^n$到$\mathbb{R}$的线性变换σ被其在任意一组基下的值确定. 假设σ在标准基下的值为

$$\sigma(e_i) = a_i,\ 1\le i\le n\,.$$

则对任意$x = (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n$, 有

$$\sigma(x) = \sigma(\sum_{i=1}^n x_i e_i) = \sum_{i=1}^n x_i\sigma(e_i) = \sum_{i=1}^n a_i x_i.$$

因此从$\mathbb{R}^n$到$\mathbb{R}$的线性变换σ恰好就是n元线性函数. 换言之, $\mathbb{R}^n$的对偶空间$(\mathbb{R}^n)^*$就是n元线性函数的集合.

下面介绍一些特殊的线性变换. 设U, V是线性空间.

(1) 零变换: 将V的所有向量均变为0 向量的变换, 称为零变换, 记为0, 即对任意$\alpha \in V$, 有$0(\alpha) = 0$.

(2) **恒等变换**: 将V中任意向量均固定的变换称为恒等变换或单位变换, 记为I_V或简记为I或1; 即对任意$\alpha \in V$, $I(\alpha) = \alpha$. 恒等变换显然是V到自身的同构(称为V的**自同构**).

(3) **位似**: 设$k \in \mathbb{F}$. 将线性空间V的任意向量α变为$k\alpha$的变换σ称为(伸缩)系数为k的位似, 即$\sigma(\alpha) = k\alpha$. 零变换与恒等变换分别是$k = 0$与$k = 1$时的位似. 非零位似均是自同构.

(4) **可逆变换**: 设线性变换$\sigma \in \mathrm{Hom}(U, V)$, 如果存在线性变换$\tau \in \mathrm{Hom}(V, U)$使得对任意$\alpha \in U$, $\beta \in V$, 均有$\tau(\sigma(\alpha)) = \alpha$, $\sigma(\tau(\beta)) = \beta$, 则称$\sigma$是可逆线性变换, τ称为其逆变换. 可以证明, 如果σ的逆变换存在则必定唯一, 其逆记为σ^{-1}, 且$(\sigma^{-1})^{-1} = \sigma$(习题 23).

比如, 任何系数为$k(k \neq 0)$的位似均是可逆的, 且其逆变换是系数为k^{-1}的位似. 又如, 例 2.3.2中的旋转变换就是线性空间$\mathbb{R}^2$的可逆线性变换, 其逆变换就是顺时针旋转θ角度或逆时针旋转$-\theta$角度.

例 2.3.7. 由于$\mathbb{R}$上的线性变换 就是线性函数(即正比例函数), 故$\mathbb{R}$上的可逆线性变换f的逆变换就是f的反函数f^{-1}.

例 2.3.8. 设$A \in M_n(\mathbb{F})$. 考虑$\mathbb{F}^n$上的由A定义的线性变换$\sigma_A : x \mapsto Ax$, 则σ_A可逆 $\Longleftrightarrow$ 矩阵A可逆, 且此时$\sigma^{-1} : x \mapsto A^{-1}x$, 即$(\sigma_A)^{-1} = \sigma_{A^{-1}}$.

例 2.3.9. 设P是m阶可逆矩阵, Q是n阶可逆矩阵, 则

$$X \mapsto PXQ \tag{2.3.5}$$

是矩阵空间$\mathbb{F}^{m\times n}$的一个自同构, 称为相抵变换或等价变换. 任何矩阵A在相抵变换下的像是与A相抵的矩阵.

例 2.3.10. 设P是n阶可逆矩阵, 则

$$X \mapsto P^{-1}XP \tag{2.3.6}$$

是矩阵空间$M_n(\mathbb{F})$的一个自同构, 称为由P诱导的相似变换或共轭变换. 任何矩阵A在相似变换下的像是与A相似的矩阵.

例 2.3.11. 设P是n阶可逆矩阵, 则

$$X \mapsto P^TXP \tag{2.3.7}$$

是矩阵空间$M_n(\mathbb{F})$的一个自同构, 称为由P诱导的相合变换或合同变换. 任何矩阵A在合同变换下的像是与A合同的矩阵. 在线性代数中, 相合变换一般限制为实对称矩阵.

可以说相抵变换、相似变换与相合变换构成了线性代数课程的全部.

2.3.2 线性变换的核与像

设$\sigma \in \mathrm{Hom}(U, V)$, 其“零点”集$\{\alpha \in U \mid \sigma(\alpha) = 0\}$, 称为$\sigma$的核, 记为$\mathrm{Ker}(\sigma)$或$\sigma^{-1}(0)$; 其“函数值”的集合$\{\alpha \in V \mid \exists \beta \in U$使得$\alpha = \sigma(\beta)\}$, 称为$\sigma$的像, 记为$\mathrm{Im}(\sigma)$或$\sigma(U)$. 易证, $\mathrm{Ker}\sigma$与$\mathrm{Im}\sigma$分别是U与V的子空间, 其维数分别记为$\eta(\sigma)$与$r(\sigma)$, 称为σ的**零度**与**秩**.

例 2.3.12. 设$U = \mathbb{F}^n$, $V = \mathbb{F}^m$, A是一个$m \times n$矩阵. 对任意$x \in U$, 定义U到V的线性变换σ为$\sigma(x) = Ax$. 则σ的核就是A的零空间, σ的零度恰好等于$n - r(A)$ (故此数又称为矩阵A的零度); σ的像空间就是A的列空间$R(A)$, σ的秩就是A的秩$r(A)$.

例 2.3.13. 设U是线性空间V的子空间, 定义$\iota: U \longrightarrow V$为$\iota(u) = u$, 则$\iota$是单变换, 常称为**包含映射**. 设$V = U \oplus W$. 对任意$v \in V$, 定义$\pi: V \longrightarrow U$为$\pi(v) = u$, 其中$v = u + w(u \in U, w \in W)$ 是v关于直和$V = U \oplus W$的唯一分解, 则π是满变换, 称为V沿W向U的**投影**. 正交投影显然是投影的一种, 但需要V上事先有“正交”的概念.

下面的命题揭示了单变换与满变换的重要性.

命题 2.3.2. 任意线性变换都可以写成一个满变换和一个单变换的复合.

证 $\sigma \in \mathrm{Hom}(U, V)$. 定义$U$到线性空间$\mathrm{Im}(\sigma)$的映射$\widetilde{\sigma}$为$\widetilde{\sigma}(x) = \sigma(x)$; $\mathrm{Im}(\sigma)$到线性空间V的映射ι为包含映射. 则$\widetilde{\sigma}$与ι分别是满变换与单变换, 且$\sigma = \iota\widetilde{\sigma}$. □

例 2.3.14. 命题2.3.2的矩阵版本恰好是矩阵的满秩分解, 即设$A \in \mathbb{F}^{m\times n}$, 则将线性变换$\mathbb{F}^n \longrightarrow \mathbb{F}^m : x \mapsto Ax$写成一个满变换与一个单变换的复合等价于$A$的满秩分解$A = BC$, 其中$\mathbb{F}^n \longrightarrow R(A) : x \mapsto Cx$是满变换, $R(A) \longrightarrow \mathbb{F}^m$为$y \mapsto By$为单变换. 特别地, $x \mapsto Ax$是单变换 $\Longleftrightarrow$ A列满秩, 是满变换 $\Longleftrightarrow$ A行满秩.

例 2.3.15. 设U, V是$\mathbb{F}$上的线性空间(不必有限维), $\sigma \in \mathrm{Hom}(U, V)$. 定义$U$到线性空间$\mathrm{Im}(\sigma)$的映射$\widetilde{\sigma}$为$\widetilde{\sigma}(x) = \sigma(x)$. 则$\widetilde{\sigma}$是线性变换. 显然, $\widetilde{\sigma}$是满的. 则由一个线性变换σ可以诱导出如下三个线性变换

$$\mathrm{Ker}(\sigma) \xrightarrow{\iota} U \xrightarrow{\widetilde{\sigma}} \mathrm{Im}(\sigma) \xrightarrow{\iota'} V \tag{2.3.8}$$

其中ι与ι'均为包含映射. 请读者给出公式(2.3.8)的矩阵版本.

利用核空间与像空间可以给出单变换与满变换的简洁刻画.

定理 2.3.2. 设U, V是$\mathbb{F}$上的有限维线性空间, $\sigma \in \mathrm{Hom}(U, V)$. 则
(1) σ是单的 $\Longleftrightarrow$ $\mathrm{Ker}(\sigma) = 0$ $\Longleftrightarrow$ 其零度为0.
(2) σ是满的 $\Longleftrightarrow$ $\mathrm{Im}(\sigma) = V$ $\Longleftrightarrow$ 其秩等于$\dim V$.
(3) σ是同构 $\Longleftrightarrow$ σ可逆.
特别地, $\sigma \in \mathrm{End}V$是单的 $\Longleftrightarrow$ σ是满的 $\Longleftrightarrow$ σ可逆.

证 (1)的证明留作习题, (2)是显然的, 此处给出(3)的证明框架. 注意, 可逆线性变换必然既是单映射也是满映射, 即可逆线性变换必是同构. 反过来, 定义同构变换$\sigma \in \mathrm{Hom}(U, V)$的逆变换$\tau$如下: 对任意$\beta \in V$, 由于$\sigma$是满的, 故存在$\alpha \in U$使得$\sigma(\alpha) = \beta$, 令

$$\tau: \beta \mapsto \alpha.$$

由于σ是单的, 故上面的定义是合理的. 验证$\tau\sigma = I_U$, $\sigma\tau = I_V$以及τ是线性变换即可. □

利用定理2.3.2, 往往可将矩阵方程化为线性变换来研究.

定理 2.3.3. 设$A \in \mathbb{F}^{m\times m}, B \in \mathbb{F}^{n\times n}, C \in \mathbb{F}^{m\times n}$. 则矩阵线性方程

$$AX - XB = C \tag{2.3.9}$$

有唯一解当且仅当A与B无公共特征值.

证 定义$\sigma\in \mathrm{End}(\mathbb{F}^{m\times n})$如下：

$$\sigma: X\mapsto AX-XB \tag{2.3.10}$$

断言σ是(自)同构 $\iff$ A与B无公共特征值.

由(第一章)无公共特征值定理可知, 如果A与B无公共特征值, 则$\mathrm{Ker}(\sigma)=0$, 故σ是单变换, 故由定理2.3.2知σ是同构.

反之, 设A与B有公共特征值λ, $A\alpha=\lambda\alpha, B^T\beta=\lambda\beta, \alpha\neq 0, \beta\neq 0$. 计算可得

$$A(a\beta^T)=\lambda\alpha\beta^T=\alpha(\lambda\beta^T)=\alpha(\beta^TB)=(\alpha\beta^T)B.$$

由于$\alpha\beta^T\neq 0$, 故σ不是同构.

矩阵线性方程$AX-XB=C$即$\sigma(X)=C$. 如果A与B无公共特征值, 则σ是自同构, 故$\sigma(X)=C$有且仅有唯一解. 反之, 如果$AX-XB=C$只有唯一解, 则σ是同构, 因此A与B无公共特征值. □

推论 2.3.1. 设$A\in\mathbb{F}^{m\times m}, B\in\mathbb{F}^{n\times n}, C\in\mathbb{F}^{m\times n}$. 如果$A$与$B$无公共特征值, 则矩阵$\begin{pmatrix} A & C\\ 0 & B\end{pmatrix}$与$\begin{pmatrix} A & 0\\ 0 & B\end{pmatrix}$相似.

证 由定理2.3.3, 此时$AX-XB=C$有唯一解M, 则有

$$\begin{pmatrix} I & M\\ 0 & I\end{pmatrix}\begin{pmatrix} A & C\\ 0 & B\end{pmatrix}=\begin{pmatrix} A & 0\\ 0 & B\end{pmatrix}\begin{pmatrix} I & M\\ 0 & I\end{pmatrix},$$

矩阵$\begin{pmatrix} I & M\\ 0 & I\end{pmatrix}$显然可逆, 故$\begin{pmatrix} A & C\\ 0 & B\end{pmatrix}$与$\begin{pmatrix} A & 0\\ 0 & B\end{pmatrix}$相似. □

2.3.3 线性变换的矩阵

设$\alpha_1,\cdots,\alpha_n$是U的一组基, $\alpha_1',\cdots,\alpha_m'$是$V$的一组基, $\sigma\in\mathrm{Hom}(U,V)$. 设

$$(\sigma(\alpha_1),\cdots,\sigma(\alpha_n))=(\alpha_1',\cdots,\alpha_m')A, \tag{2.3.11}$$

其中$A=(a_{ij})\in\mathbb{F}^{m\times n}$称为$\sigma$关于$\alpha$-基和$\alpha'$-基的矩阵. 特别地, 如果$U=V$且取$\alpha$-基等于$\alpha'$-基, 则$\sigma$关于$\alpha$-基和$\alpha$-基的矩阵简称为$\sigma$关于$\alpha$-基的矩阵.

公式(2.3.11)的左端一般简记为$\sigma(\alpha_1,\cdots,\alpha_n)$, 即有

$$\sigma(\alpha_1,\cdots,\alpha_n)=(\alpha_1',\cdots,\alpha_m')A.$$

(上述记号有两个意义, 一是简化记号, 二是强调“线性变换”是作用在“向量”上而非系数上(线性变换的齐次性), 从而系数或坐标或坐标形成的矩阵都可以拿到线性变换的外面.)

公式(2.3.11)确定一组基的像, 由此可确定任何向量的像, 即有下述**“坐标变换定理”**

定理 2.3.4. 设线性变换$\sigma\in\mathrm{Hom}(U,V)$在$\alpha$-基和$\alpha'$-基下的矩阵为$A$, 向量$v\in U$在$\alpha$-基下的坐标为$x$, 则$\sigma(v)$在$\alpha'$-基下的坐标为$A\,x$.

注1. 由定理 2.3.4, 设σ是线性变换, 其在取定基下的矩阵为A, 则σ将坐标为$x=(x_1,x_2,\cdots,x_n)^T$的向量变为坐标为Ax的向量, 而Ax的每一行$a_{j1}x_1+a_{j2}x_2+\cdots+a_{jn}x_n$, $1\le j\le n$恰好是一个n元线性函数. 这就是说, 线性变换本质上是多个多元线性函数.

例 2.3.16. 为计算例 2.3.2的旋转变换σ在标准基$e_1=(1,0)^T, e_2=(0,1)^T$下的矩阵A, 需要计算$\sigma(e_1)$, $\sigma(e_2)$. 直接计算可得, $\sigma(e_1)=(\cos\theta,\sin\theta)^T$, $\sigma(e_2)=(-\sin\theta,\cos\theta)^T$. 故

$$A=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}.$$

一个线性变换在不同基下的矩阵有何联系? 为此, 设

$$\{\alpha_1,\alpha_2,\cdots,\alpha_n\},\quad \{\beta_1,\beta_2,\cdots,\beta_n\}$$

是U的两组基, P是由α-基到β-基的过渡矩阵; 设

$$\{\alpha_1',\alpha_2',\cdots,\alpha_m'\},\quad \{\beta_1',\beta_2',\cdots,\beta_m'\}$$

是V的两组基, Q是由α'-基到β'-基的过渡矩阵. 设$\sigma\in\mathrm{Hom}(U,V)$关于$\alpha$-基和$\alpha'$-基的矩阵为$A$, 关于$\beta$-基和$\beta'$-基的矩阵为$B$. 设$\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)x$, 问$\sigma(\alpha)$关于$\beta'$–基的坐标是什么?

首先, α关于β-基的坐标为$P^{-1}x$; 再按公式(2.3.11), $\sigma(\alpha)$关于α'-基的坐标为Ax, 关于β'-基的坐标为$BP^{-1}x$. 故由坐标变换公式得, $Q^{-1}Ax=BP^{-1}x$. 故$Q^{-1}A=BP^{-1}$, 即$QB=AP$或

$$B=Q^{-1}AP. \tag{2.3.12}$$

注 公式(2.3.12)说明线性变换在α-α'-基下的矩阵与在β-β'-基下的矩阵是等价的; 反过来, 如果两个同阶矩阵A与B是等价的, 即公式(2.3.12)成立, 则可以构造相应的线性变换σ, 使得A与B分别是σ在α-α'-基与β-β'-基下的矩阵, 请读者自行证明之.

现假定$U=V$且α-基等于α'-基, β-基等于β'-基, 则公式(2.3.12)中的矩阵P与Q相等, 即

$$B=P^{-1}AP. \tag{2.3.13}$$

所以相似矩阵是同一线性变换在不同基下的矩阵. 这就是下面的

定理 2.3.5. 设V是n维线性空间, σ是V的一个线性变换. 设σ关于V的两组基的矩阵分别为A与B. 则A与B相似.

由于相似矩阵具有相同的行列式和迹, 可将线性变换σ在任意一组基下的矩阵的行列式和迹定义为σ的行列式与迹, 分别记为$|\,\sigma\,|$与$\mathrm{tr}(\sigma)$.

自然也可以定义线性变换的特征向量、特征值与特征子空间等概念.

设V是$\mathbb{C}$上的线性空间(可以是无限维), $\sigma\in\mathrm{End}V$, $\lambda\in\mathbb{C}$. 如果存在非零向量$\alpha\in V$使得$\sigma(\alpha)=\lambda\alpha$, 则称$\lambda$是$\sigma$的一个特征值, 而$\alpha$是$\sigma$的一个(属于特征值$\lambda$的)特征向量.

如果λ是σ的特征值, 则V的子集合$\mathrm{Ker}(\lambda I-\sigma)=\{x\in V\,|\,\sigma(x)=\lambda x\}$称为$\sigma$的(属于特征值$\lambda$的)特征子空间, 记为$V_\lambda$.

从线性变换的角度看, 特征向量是在线性变换下方向不变的那些非零向量, 而特征值则是特征向量的伸缩系数. 注意线性变换在其每个特征子空间上的限制是一个位似, 因此特征值与特征向量具有简化线性变换的矩阵表示的作用. 确切地说, 如果有限维线性空间V有一组基其基向量全部为σ的特征向量, 则σ在这组基下的矩阵是对角矩阵. 故有下述定理(请读者自证).

定理 2.3.6. 设V是有限维线性空间, $\sigma \in \mathrm{End}V$, A是σ在某组基下的矩阵. 则
(1) A与σ有完全相同的特征值(即重数也一样).
(2) 设σ的不同特征值为$\lambda_1, \cdots, \lambda_s$, 则$A$可以对角化 $\iff V = \sum\limits_{i=1}^{s} \oplus V_{\lambda_i}$.

从线性变换的角度理解特征值与特征向量较为直观. 比如考虑旋转矩阵(设$\theta \neq 0, \pi$)

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

的特征值与特征向量. 如果仅从矩阵的角度出发, 那就只好硬算. 但是, 矩阵A是平面上的旋转, 因此它不改变任何向量的长度, 故其特征值的模只能是1, 特别其实特征值只能是±1. 另一方面, 该旋转改变了所有非零向量的方向, 因此它没有实特征向量. 所以将矩阵的特征值与特征向量问题转化为线性变换的相应问题常常是有效的方法. 比如, 相似矩阵具有相同的特征值是因为这些特征值都是同一个线性变换的特征值.

例 2.3.17. 设σ是平面$\mathbb{R}^2$的旋转$\pi/4$的变换. 则σ在标准基下的矩阵为

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

易知σ在基$(1,0)^T, (1,1)^T$下的矩阵为

$$B = \begin{bmatrix} 0 & -\sqrt{2} \\ \frac{1}{\sqrt{2}} & \sqrt{2} \end{bmatrix}.$$

由于标准基到基$(1,0)^T, (1,1)^T$的过渡矩阵为

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

所以$AP = PB$, 或$B = P^{-1}AP$. 因此$|\sigma| = 1$, $\mathrm{tr}(\sigma) = \sqrt{2}$.

第四节 线性代数基本定理

本节将建立联系线性变换与矩阵的基本定理.

首先定义$\mathrm{Hom}(U,V)$中的加法以及数乘运算. 设$\sigma, \tau \in \mathrm{Hom}(U,V)$, $\alpha \in V$, $k \in \mathbb{F}$, 定义

$$\sigma + \tau : \alpha \mapsto \sigma(\alpha) + \tau(\alpha); \quad k\sigma : \alpha \mapsto k\sigma(\alpha). \tag{2.4.1}$$

显然, 如上(2.4.1)定义的$\sigma + \tau$与$k\sigma$仍是U到V的线性变换, 分别称为σ与τ的和以及数k与σ的数乘. 如此, 我们在$\mathrm{Hom}\,(U,V)$中定义了向量的加法与数乘. 易证, $\mathrm{Hom}\,(U,V)$在上述加法与数乘下, 作成$\mathbb{F}$上的一个线性空间.

下面的定理反映了线性变换与矩阵的本质联系, 被称为**“线性变换基本定理”**.

定理 2.4.1. 设$\dim_{\mathbb{F}} U = n$, $\dim_{\mathbb{F}} V = m$, 则$\mathrm{Hom}(U, V) \cong \mathbb{F}^{m\times n}$. 因此,

$$\dim_{\mathbb{F}} \mathrm{Hom}(U, V) = mn = (\dim_{\mathbb{F}} U)(\dim_{\mathbb{F}} V). \tag{2.4.2}$$

特别地, $U^* = \mathrm{Hom}(U, \mathbb{F}) \cong \mathbb{F}^{1\times n} \cong \mathbb{F}^{n\times 1} \cong \mathbb{F}^n$, 且

$$\bigoplus_{i=1}^{m} U^* \cong \mathrm{Hom}(U, V) \cong V^* \otimes U \tag{2.4.3}$$

证 对$\forall \sigma \in \mathrm{Hom}(U, V)$, 记$A(\sigma)$是$\sigma$在固定基下的矩阵. 定义$\mathrm{Hom}(U, V)$到$\mathbb{F}^{m\times n}$的映射$\psi$为

$$\begin{array}{cccc} \psi: & \mathrm{Hom}(U, V) & \longrightarrow & \mathbb{F}^{m\times n} \\ & \sigma & \longmapsto & A(\sigma) \end{array}$$

容易验证ψ是既单又满的线性变换, 故$\mathrm{Hom}(U, V) \cong \mathbb{F}^{m\times n}$, 从而公式(2.4.2)成立.

公式(2.4.3)中的第一个同构是说$\mathrm{Hom}\,(U, V)$是$\dim V$个对偶空间U^*的直和, 下证第二个同构是“自然的”, 为此设V的一组基为$v_1, ..., v_m$, 而$v_1^*, ..., v_m^*$是相应的V^*的对偶基. 对任意$v = \sum_{i=1}^{m} x_i v_i \in V$, 记$v^* = \sum_{i=1}^{m} x_i v_i^* \in V^*$. 现定义$\Phi : \mathrm{Hom}(U, V) \longrightarrow V^* \otimes U$如下:

$$\Phi : f \mapsto f(u)^* \otimes u, \forall f \in \mathrm{Hom}(U, V), \forall u \in U \tag{2.4.4}$$

Φ显然是线性变换, 以下证明Φ是单的. 设$f \in \mathrm{Ker}(\Phi)$, 则对$\forall u \in U, f(u)^* \otimes u = 0$. 如果$f \neq 0$, 则存在$u \neq 0, f(u) \neq 0$, 从而$f(u)^* \neq 0, f(u)^* \otimes u \neq 0$, 矛盾. 故$\mathrm{Ker}(\Phi) = 0$即$\Phi$是单的, 比较维数即可知$\Phi$是同构. □

线性空间V的对偶空间V^*也是线性空间, 其对偶空间$(V^*)^*$记为V^{**}.

由定理 2.4.1可知, 如果V是有限维的, 则$V \cong V^* \cong V^**$.

例 2.4.1. 设$V = \mathbb{Q}[x]$是有理数域上一元多项式空间, 设$\mathbb{Q}[[x]]$是有理数域上一元形式幂级数空间. 定义

$$\phi : V^* \longrightarrow \mathbb{Q}[[x]], \phi(f) = \sum_{i=0}^{\infty} f(x^i) x^i, \forall f \in V^* \tag{2.4.5}$$

则易证ϕ是同构映射, 因此$V^* \cong \mathbb{Q}[[x]]$. 注意, $\mathbb{Q}[x]$到$\mathbb{Q}[[x]]$的任何映射都不可能是满射, 这是因为$\mathbb{Q}[x]$是可数集合, 而$\mathbb{Q}[[x]]$是不可数集合. 因此$V = \mathbb{Q}[x] \ncong \mathbb{Q}[[x]] = V^*$.

例 2.4.1的结论对任意无限维空间都是成立的, 即若$\dim V = \infty$, 则V^*与V不同构. 关于对偶空间有一个有趣的结论, 即V与V^*之间没有自然的线性变换, 即使是有限维也没有(即任意线性变换都依赖于基), 而V与V^{**}之间则有自然的线性变换.

命题 2.4.1. 设V是线性空间. 定义

$$\psi : V \longrightarrow V^{**}, \phi(v)(f) = f(v), \forall v \in V, \forall f \in V^* \tag{2.4.6}$$

则ψ是单线性变换. 特别地, 若V是有限维的, 则ψ是同构映射.

证 ψ显然是线性变换. 设$v \in V, v \in \mathrm{Ker}(\psi)$. 则$0 = \phi(v) \in V^{**}$, 即$0 = \phi(v)(f) = f(v), \forall f \in V^*$, 故$v \in \bigcap_{f \in V^*} \mathrm{Ker}(f) = 0$. 故$\psi$是单的.

□

线性代数中的"代数"一词一般是指定义了乘法(非数乘)的线性空间. 如果$U = V$, 则可以定义$\mathrm{End}\, V$中两个线性变换σ与τ的乘法

$$\sigma\tau : \alpha \mapsto \sigma(\tau(\alpha)), \quad \forall\, \alpha \in V \tag{2.4.7}$$

由公式(2.4.7)定义的线性变换的乘法实际上是映射的合成, 故满足结合律但不满足交换律. 还容易验证, 该乘法与线性变换的线性运算(即加法与数乘)满足分配律、结合律等等. 这样就在$\mathrm{End}\, V$上建立了具有加法、数乘和乘法三种运算的结构, 称为V的**线性变换代数**.

线性变换σ与自己的乘积$\sigma\sigma$记为σ^2. 归纳地, 对任意自然数k, 可以定义σ的k次幂$\sigma^k = \sigma^{k-1}\sigma$ (为方便记, 规定$\sigma^0 = I$). 继而, 对$\mathbb{F}$上的任意多项式$f(x)$, 可以定义线性变换σ的多项式$f(\sigma)$. 易证, 它是一个可以和σ交换的线性变换, 即$\sigma\, f(\sigma) = f(\sigma)\sigma$.

设n是正整数. 下面的定理被称为**"线性代数基本定理"**, 它说明n维线性空间V的线性变换代数与n阶矩阵代数本质上是一码事.

定理 2.4.2. 设V是$\mathbb{F}$上的n维线性空间, $\alpha_1, \alpha_2, \cdots, \alpha_n$是$V$的一组基. 设$M_n(\mathbb{F})$是$\mathbb{F}$上全体$n$阶矩阵组成的线性空间. 对任意$\sigma \in \mathrm{End}\, V$, 记$A(\sigma)$是$\sigma$在该基下的矩阵. 定义$\mathrm{End}\, V$到$M_n(\mathbb{F})$的映射$\psi$为

$$\begin{array}{cccc} \psi : & \mathrm{End}\, V & \longrightarrow & M_n(\mathbb{F}) \\ & \sigma & \longmapsto & A(\sigma) \end{array}$$

则ψ是一个保持运算(加法, 数乘与乘法)的一一映射, 即满足下列条件:

(1) $A(\sigma + \tau) = A(\sigma) + A(\tau)$.
(2) $A(k\sigma) = kA(\sigma), \forall\, k \in \mathbb{F}$.
(3) $A(\sigma\tau) = A(\sigma)A(\tau)$.
(4) σ可逆 $\Longleftrightarrow$ $A(\sigma)$可逆; 且此时$A(\sigma)^{-1} = A(\sigma^{-1})$.
(5) $A(0) = 0$; $A(I) = I$.

证 此处仅证明ψ是一一映射, 其余基本类似, 留作习题. 由前两条可知ψ是线性变换. 为证ψ是单的, 只需证明$\mathrm{Ker}(\psi) = 0$. 设$\psi(\sigma) = 0$, 即$A(\sigma) = 0$. 对任意$\alpha \in V$, 设x是α在该基下的坐标, 则由定理2.3.4可知$\sigma(\alpha)$的坐标为$A(\sigma)x = 0x = 0$. 于是$\sigma = 0$. 再证ψ是满的. 设B是任意n阶矩阵. 令$\sigma(\alpha) = (\alpha_1, \alpha_2, \cdots, \alpha_n)Bx$. 则显然$\sigma$是$V$的线性变换, 且$\sigma$在该基下的矩阵是$B$, 从而$\psi(\sigma) = B$, 即$\psi$是满的. □

例 2.4.2. 设$\sigma \in \mathrm{End}\mathbb{R}[x]_3$是求导运算, 即$\forall f = a + bx + cx^2 \in \mathbb{R}[x]_3, \sigma(f) = b + 2cx$. 则

$$\sigma((1, x, x^2)) = (0, 1, x) = (1, x, x^2)\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} = (1, x, x^2)A(\sigma). \tag{2.4.8}$$

定理2.4.2说σ与矩阵$A(\sigma) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$本质上是一回事. 比如, $\sigma(f) = (1, x, x^2)[A(\sigma)(a, b, c)^T]$; 再比如, $\sigma^2(f) = 2c = (1, x, x^2)[A(\sigma)^2(a, b, c)^T]$.

华罗庚对线性变换的运算有若干精彩研究成果, 下面一条是被称为“**华罗庚半同态定理**”的矩阵版本(请读者自证).

定理 2.4.3. 设$V=\mathbb{F}^{n\times n}$. 记

$$\mathbb{H}=\{f:V\longrightarrow V\mid f(X+Y)=f(X)+f(Y),\forall X,Y\in V\}.$$

设$f\in\mathbb{H}$. 如果$f(XY)=f(X)f(Y)$ 或$f(XY)=f(Y)f(X),\forall X,Y\in V$, 则$f(XY)=f(X)f(Y),\forall X,Y\in V$ 或$f(XY)=f(Y)f(X),\forall X,Y\in V$.

类似于矩阵的情形, 满足$\sigma^2=\sigma$的线性变换称为**幂等变换**. 满足$\sigma^k=0$的线性变换称为**幂零变换**(且使此式成立的最小自然数称为σ的**幂零指数**). 零变换与恒等变换都是幂等变换, 例 2.4.2中的求导运算是幂零变换.

例 2.4.3. 设$V=\mathbb{F}^n$. 对$\alpha=(x_1,x_2,\cdots,x_n)^T$, 定义

$$\sigma(\alpha)=(x_1,0,\cdots,0)^T;\ \tau(\alpha)=(x_2,x_3,\cdots,x_n,0)^T.$$

则σ与τ分别是V的幂等变换与幂零变换. σ与τ在标准基下的矩阵分别是E_{11}与$E_{12}+...+E_{n,n-1}$(此矩阵是第三章Jordan标准形的核心, 记作J_n), 故τ的幂零指数是n.

例 2.4.4. 零变换在任何基下的矩阵都是零矩阵. 恒等变换在任何基下的矩阵都是单位矩阵. 位似变换在任何基下的矩阵都是同一纯量矩阵. 幂等变换在任何基下的矩阵都是幂等矩阵. 幂零变换在任何基下的矩阵都是幂零矩阵.

我们知道, Hermite幂等矩阵即投影矩阵具有唯一性, 下面的命题表明幂等矩阵也有唯一性.

命题 2.4.2. 设$P,Q\in\mathbb{F}^{n\times n},P^2=P,Q^2=Q,R(P)=R(Q),N(P)=N(Q)$, 则$P=Q$.

证 首先幂等矩阵固定其列空间的所有向量, 故幂等矩阵P的列空间$R(P)$与零空间$N(P)$互补, 这是因为若$x\in R(P)\bigcap N(P)$, 则$x=Px=0$. 因此$\mathbb{F}^n=R(P)\oplus N(P)$. 对$\forall\alpha\in\mathbb{F}^n,\alpha=P\alpha+(I-P)\alpha$, 因为$P\alpha\in R(P)=R(Q),(I-P)\alpha\in N(P)=N(Q)$, 故$Q(P\alpha)=P\alpha,Q[(I-P)\alpha]=0$, 所以$Q\alpha=Q[P\alpha+(I-P)\alpha]=P\alpha$, 从而$Q=P$. □

思考题

1. 是否有可加性与齐次性等价的情形?
2. 平面(即$\mathbb{R}^2$)上的线性变换能否将直线变为抛物线或者椭圆? 能否将抛物线或者椭圆变为直线? 空间(即$\mathbb{R}^3$)中的线性变换能否将平面变为直线?能否将抛物线变为直线或者椭圆?
3. 如何建立空间中的过原点的直线和平面上的过原点的直线之间的同构映射?
4. 解释关于线性变换的公式(2.3.12)与矩阵的等价之间的关系?
5. 以线性变换的观点解释列满秩与行满秩矩阵以及矩阵的满秩分解.
6. 设V是1维线性空间, 则End$V$与Aut$V$是什么? 特别, 什么是End$\mathbb{R}$, Aut$\mathbb{R}$, End$\mathbb{C}$, Aut$\mathbb{C}$?
7. 有限维线性空间上的单线性变换就是满线性变换, 此结论对无限维线性空间成立吗?
8. 如果一个线性变换σ的特征值的模均小于1, σ有何特点?
9. 如果一个线性变换σ有一组正交的特征向量, σ有何特点?

第五节 内积空间、投影与等距变换

本节将在线性空间上通过定义内积而引入长度等概念以讨论几何问题.

2.5.1 内积与长度

定义 2.5.1. 设$\mathbb{F}=\mathbb{C}$或$\mathbb{R}$, V是$\mathbb{F}$上的线性空间. 对$\forall\alpha,\ \beta\in V$, 定义$(\alpha,\beta)\in\mathbb{F}$满足条件

(1) (共轭对称性) $(\alpha,\beta)=\overline{(\beta,\alpha)}$, 其中$\overline{(\beta,\alpha)}$是复数$(\beta,\alpha)$的共轭复数.

(2) (正定性) $(\alpha,\alpha)\geq 0$, 且等号成立 $\iff \alpha=0$.

(3) (双线性) $(a\alpha+b\beta,\gamma)=a(\alpha,\gamma)+b(\beta,\gamma)$, 对任意$\alpha,\beta,\gamma\in V,\ a,b\in\mathbb{F}$成立.

则称V为一个**内积空间**, 数(α,β)称为向量α与β的内积.

注1. 由于内积的共轭对称性, 内积的第三个条件"双线性"仅对第一个变量成立, 而对第二个变量是"共轭双线性"的, 即

$$(\alpha,a\beta+b\gamma)=\bar{a}(\alpha,\beta)+\bar{b}(\alpha,\gamma).$$

若$\mathbb{F}=\mathbb{R}$是实数域, 则内积是对称的, 即$(\alpha,\beta)=(\beta,\alpha)$, 故此时的内积对两个变量均为"双线性"的.

注2. $\mathbb{F}$-线性空间V上的内积实际上是$V\times V$到$\mathbb{F}$的一个函数, 一般记为$(\bullet,\bullet)_V$或$(\bullet,\bullet)$.

有限维实内积空间通常称为欧几里得空间或**欧氏空间**. 复数域上的内积空间称为**酉空间**或复内积空间. 除特别说明, 本书仅限于讨论有限维内积空间.

例 2.5.1. 实数域$\mathbb{R}$上的通常内积即普通乘积: $(a,b)=ab,\forall a,b\in\mathbb{R}$, 通常绝对值由该内积诱导. 容易证明, $\mathbb{R}$上的所有内积均是通常内积的正常数倍.

例 2.5.2. 复数域$\mathbb{C}$上的通常内积为: $(a,b)=\bar{b}a,\forall a,b\in\mathbb{C}$, 复数的模长由该内积诱导. 容易证明, $\mathbb{C}$上的所有内积均是通常内积的正常数倍.

例 2.5.3. 设$V=\mathbb{R}[x]_n$是次数小于n的实系数多项式构成的n维实线性空间, $a<b$是实数. 设$f,g\in V$, 规定

$$(f,g)=\int_a^b f(x)g(x)\,\mathrm{d}\,x, \tag{2.5.1}$$

则(f,g)定义了V上的一个内积. (公式(2.5.1)也是无限维空间$C[a,b]$上的内积.)

例 2.5.4. 设$(\bullet,\bullet)_i$是V上的内积, $a_i>0(i=1,2,...,n)$, 则$\sum\limits_{i=1}^{n}a_i(\bullet,\bullet)_i$也是$V$上的内积.

非负实数(α,α)的算数平方根定义为向量α的**模**(或**范数**, 或长度), 记为

$$\|\alpha\|=\sqrt{(\alpha,\alpha)}.$$

特别, 模为1的向量称为**单位向量**或**标准向量**.

内积与范数具有下述性质.

命题 2.5.1. (1) 齐次性: $(a\alpha,\ b\beta)=a\bar{b}(\alpha,\beta)$.

(2) 可加性: $(\alpha+\beta,\gamma)=(\alpha,\gamma)+(\beta,\gamma)$.

(3) 绝对性: $\|c\alpha\|=|c|\|\alpha\|$.

(4) Cauchy-Schwarz[32]不等式: $|(\alpha\,,\beta)|\leq\|\alpha\|\|\beta\|$且等号成立 $\iff$ α与β线性相关.

[32]Augustin-Louis Cauchy(1789-1857), 法国著名数学家, ϵ-δ-语言即由其发明. Hermann Schwarz(1843-1921), 德国数学家. Cauchy-Schwarz不等式首先由Cauchy于1821年发现, 再由俄国数学家Victor Yakovlevich Bunyakovsky(1804-1889)于1859年重新发现, 后由Schwarz于1888年再次重新发现.

(5) 三角不等式：$\|\alpha+\beta\| \leq \|\alpha\|+\|\beta\|$.

(6) 极化恒等式(又称为“平行四边形恒等式”)：

$$\|x+y\|^2+\|x-y\|^2=2(\|x\|^2+\|y\|^2) \tag{2.5.2}$$

证 仅证明(6). 对$\mathbb{F}=\mathbb{R}$, 取

$$(x,y)=\frac{1}{4}\left(\|x+y\|^2-\|x-y\|^2\right) \tag{2.5.3}$$

对$\mathbb{F}=\mathbb{C}$, 取

$$(x,y)=\frac{1}{4}\left(\|x+y\|^2-\|x-y\|^2+i\|x+iy\|^2-i\|x-iy\|^2\right) \tag{2.5.4}$$

□

平行四边形恒等式的几何意义非常明显, 即平行四边形两对角线的平方和等于四条边的平方和.

由 Cauchy-Schwarz 不等式即可一般地定义“角度”, 即设α与β的夹角为θ, 则

$$\theta=\arccos\frac{|(\alpha\,,\beta)|}{\|\alpha\|\|\beta\|} \tag{2.5.5}$$

此时$0\leq\theta\leq\dfrac{\pi}{2}$. 特别地, 若$\mathbb{F}=\mathbb{R}$, 则此夹角可以进一步定义为

$$\theta=\arccos\frac{(\alpha\,,\beta)}{\|\alpha\|\|\beta\|} \tag{2.5.6}$$

此时有$0\leq\theta\leq\pi$.

内积具有明显的物理意义, 即“恒力作功”, 范数则揭示了内积的几何意义即长度, 比如欧氏空间$\mathbb{R}^n$中的通常内积定义为

$$(\alpha,\beta)=\beta^T\alpha=\alpha^T\beta=(\beta,\alpha);$$

此即是坐标乘积之和; 而$\mathbb{C}^n$中的内积则定义为

$$(\alpha,\beta)=\beta^*\alpha=\overline{(\beta,\alpha)}=\overline{\alpha^*\beta}.$$

在$\mathbb{R}^2$中有

$$(\alpha\,,\beta)=\|\alpha\|\|\beta\|\cos\theta, \tag{2.5.7}$$

其中θ是α与β的夹角. 注意

$$(\alpha-\beta\,,\alpha-\beta)=\|\alpha\|^2-2(\alpha,\,\beta)+\|\beta\|^2. \tag{2.5.8}$$

记$a=\|\alpha\|$, $b=\|\beta\|$, $c=\|\alpha-\beta\|$. 则由等式(2.5.7)与(2.5.8)可得余弦定理:

$$c^2=a^2+b^2-2ab\cos\theta. \tag{2.5.9}$$

而当$\cos\theta=0$即$\theta=\dfrac{\pi}{2}$时, 即得勾股定理.

由定义可知, 二非零向量平行当且仅当其夹角为0或π(当$\mathbb{F}=\mathbb{R}$时), 而二非零向量垂直当且仅当其夹角为$\dfrac{\pi}{2}$.

定义 2.5.2. 两个内积为0的向量α与β称为**正交**的, 记为$\alpha \perp \beta$. 如果一组非零向量两两正交则称为**正交组**. 单位向量构成的正交组称为**标准正交组**.

命题 2.5.2. 正交组必是线性无关的.

证 设$\alpha_1, \alpha_2, \cdots, \alpha_s$是内积空间$V$的一个正交组, $a_1, a_2, \cdots, a_s \in \mathbb{F}$使得

$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_s\alpha_s = 0.$$

则

$$0 = (\alpha_i, a_1\alpha_1 + a_2\alpha_2 + \cdots + a_s\alpha_s) = \sum_{j=0}^{s} a_j(\alpha_i, \alpha_j) \tag{2.5.10}$$

由于$(\alpha_i, \alpha_j) = 0, \forall j \neq i$ 且$(\alpha_i, \alpha_i) = \|\alpha_i\|^2 \neq 0$, 故(2.5.10)蕴含$a_i = 0, \forall i$. 因此, 该向量组是线性无关的. □

注. 正交的几何意义是垂直, 故对内积空间$\mathbb{R}^n$或$\mathbb{C}^n$来说, 命题2.5.2是显然的. 对一般n维内积空间而言, 一个标准而方便的办法是利用线性空间的同构定理将所讨论的空间视为$\mathbb{R}^n$或$\mathbb{C}^n$.

定理 2.5.1. 有限维内积空间必存在标准正交基.

证 设$\alpha_1, \alpha_2, \cdots, \alpha_s$是内积空间$V$的任意一组基, 注意第一章第二节介绍的Gram-Schmidt 正交化方法完全适合于V, 故可求得V的一个标准正交基. 确切地说, 令

$$\beta_1 = \alpha_1\,, \quad \gamma_1 = \frac{\beta_1}{\|\beta_1\|}.$$

则按照归纳法可取

$$\beta_k = \alpha_k - (\alpha_k, \gamma_{k-1})\gamma_{k-1} - (\alpha_k, \gamma_{k-2})\gamma_{k-2} - \cdots - (\alpha_k, \gamma_1)\gamma_1. \tag{2.5.11}$$

再令$\gamma_k = \frac{\beta_k}{\|\beta_k\|}$, 即可求得$V$的一个标准正交$\gamma_1, \gamma_2, \cdots, \gamma_s$基. □

例 2.5.5. 设$V = \mathbb{R}[x]_3, 1, x, x^2$是$V$的一组基. 由例2.5.3可知, 可以定义$\forall f, g \in V$的内积为$(f, g) = \displaystyle\int_{-1}^{1} f(x)g(x)\,\mathrm{d}\,x$. 直接计算可知$1, x, x^2$不是$V$的标准正交基. 根据Gram-Schmidt 正交化方法, 可计算出相应的标准正交基为

$$\frac{1}{\sqrt{2}},\ \frac{\sqrt{6}}{3}x,\ \frac{2\sqrt{10}}{5}(1 - 3x^2).$$

两个向量α, β的**距离**$d(\alpha, \beta)$可以自然地定义为其差向量的长度, 即

$$d(\alpha, \beta) = \|\alpha - \beta\|. \tag{2.5.12}$$

该距离称为由内积$(\bullet, \bullet)$诱导的距离.

容易验证由公式(2.5.12)定义的距离确实满足下面的距离三公理(证明留作习题):

公理1.(正定性)$d(x, y) \geq 0$且等号成立当且仅当$x = y$.

公理2.(对称性)$d(x,y)=d(y,x)$.

公理3.(三角不等式)$d(x,y)\le d(x,z)+d(z,y)$.

我们将在第五章进一步讨论线性空间上的内积与距离的关系.

2.5.2 内积与正定矩阵

本节将揭示内积与正定矩阵(正定二次型)的关系.

设$\alpha_1,\,\alpha_2,\cdots,\,\alpha_n$是$n$维内积空间$V$的一组基, 设$a\,,\beta\in V$的坐标分别是$x=(x_1,\,x_2,\,\cdots,\,x_n)^T$与$y=(y_1,\,y_2,\,\cdots,\,y_n)^T$, 则有

$$(\alpha\,,\beta)=(\sum_{i=1}^n x_i\alpha_i,\sum_{j=1}^n y_j\alpha_j)=\sum_{i=1}^n\sum_{j=1}^n(\alpha_i,\,\alpha_j)x_i\bar{y}_j. \tag{2.5.13}$$

记$(\alpha_i,\,\alpha_j)=a_{ji}$, 则公式(2.5.13)变为

$$(\alpha\,,\beta)=y^*Ax \tag{2.5.14}$$

其中$A=(a_{ij})$是Hermite矩阵. 由定义2.5.1(2)的正定性可知, $A=(a_{ij})$还是正定矩阵, 称为向量组$\alpha_1,,\cdots,\,\alpha_n$的**度量矩阵**或Gram[33]矩阵. 反过来, 如果Hermite矩阵$A=(a_{ij})$正定, 则容易验证公式(2.5.14)确实定义了V的一个内积. 因此有下述定理:

定理 2.5.2. 设V是n维实或复线性空间, 则其上的内积与正定矩阵一一对应. 确切地说, 设$(\alpha\,,\beta)$是V上的二元向量函数, 则(α,β)是内积 $\iff$ 存在正定Hermite矩阵$A=(a_{ij})$使得$(\alpha\,,\beta)=y^*Ax$, 其中x与y分别是α与β在某取定基下的坐标.

由定理2.5.2可知, 任何非平凡实或复线性空间上均可以定义无限多种不同的内积, 我们将在第五章证明有限维实或复线性空间上的任意两个内积本质上是完全相同的.

标准正交基可以使内积的计算大为简化.

设$\alpha_1,\cdots,\,\alpha_n$是$V$的标准正交基, $\alpha,\beta\in V$的坐标分别为$(a_1,\cdots,a_n)^T$与$(b_1,\cdots,b_n)^T$. 则

$$(\alpha,\beta)=(\sum_{i=1}^n a_i\alpha_i,\sum_{j=1}^n b_j\alpha_j=\sum_{i=1}^n a_i\bar{b}_i. \tag{2.5.15}$$

即在标准正交基下, 向量的内积就是对应坐标的(共轭)乘积之和; 这与$\mathbb{R}^n$或$\mathbb{C}^n$的普通内积完全一致(事实上同构定理保证$V\cong\mathbb{F}^n$, 再将$\mathbb{F}^n$中的内积移植到V上即可). 这里的本质是标准正交基的度量矩阵乃是单位矩阵, 因此其确定的二次型没有交叉项且平方项的系数均为1.

对任意$A=(a_{ij})\in\mathbb{F}^{n\times n}$, 一般称

$$y^*Ax=\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_i\bar{y}_j \tag{2.5.16}$$

为矩阵A(对应)的(n维)**共轭双线性型**或干脆**双线性型**(注意, 它对第一个变元x是线性的, 对第二个变元y是共轭线性的). 二次型x^*Ax称为双线性函数y^*Ax的二次型(线性代数中要求$A^*=A$). 按照二次型的正定、半正定等可以定义相应的双线性型的正定、半正定等概念. 于是内积不过是特殊的即正定的双线性型而已.

[33]即Gram-Schmidt正交化方法中的J.P.Gram.

例 2.5.6. 实双线性型$(x,y)=ax_1y_1+bx_1y_2+bx_2y_1+cx_2y_2$的矩阵形式为:

$$(x,y)=(y_1,y_2)\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=y^TGx. \tag{2.5.17}$$

因此该双线性型 定义$\mathbb{R}^2$的一个内积 $\Longleftrightarrow$ 其定义矩阵

$$G=\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

是正定矩阵. 现设$a=1,b=-1,c=2$, 此时G是正定矩阵, 故公式(2.5.17)定义了$\mathbb{R}^2$的一个新内积, 在该内积下有

$$(e_1,e_1)=1,(e_1,e_2)=-1,(e_2,e_2)=2.$$

因此e_1与e_2的长度分别是$1,\sqrt{2}$, 而它们的夹角为$\arccos\dfrac{-1}{\sqrt{2}}=\dfrac{3\pi}{4}$!

内积的另一种理解是下面的"**Riesz**[34] **表示定理**".

定理 2.5.3. 设$(V,(\bullet,\bullet))$是n维内积空间, $\forall f\in V^*,\exists v_0\in V$使得$f(v)=(v,v_0)$.

证 设$v_1,...,v_n$是V的一组标准正交基, 取$v_0=\overline{f(v_1)}v_1+...+\overline{f(v_n)}v_n$. 对$\forall v=\sum\limits_{i=1}^n x_iv_i\in V$, 有

$$f(v)=\sum_{i=1}^n x_if(v_i)$$

而

$$(v,v_0)=\sum_{i=1}^n x_i\overline{\overline{f(v_i)}}=\sum_{i=1}^n x_if(v_i).$$

□

请思考, 定理2.5.3中的向量v_0唯一吗?

前面曾讨论过V与其对偶空间V^*之间不存在自然映射, 但Riesz 表示定理给出有限维内积空间与其对偶空间之间的一个"自然映射".

推论 2.5.1. 设$(V,(\bullet,\bullet))$是n维内积空间, 则映射

$$\varphi:V\longrightarrow V^*,\varphi(v)(u)=(v,u),\forall u\in V \tag{2.5.18}$$

是一个线性同构.

矩阵空间$\mathbb{F}^{m\times n}$上的内积本质上当然与mn维线性空间$\mathbb{F}^{mn}$上的内积一致, 但下面的定理给出的$\mathbb{F}^{m\times n}$上的一种特殊内积融汇了矩阵固有的特殊结构.

[34]F. Riesz(1880-1956), 著名匈牙利数学家, 泛函分析奠基人之一.

定理 2.5.4. 设$V=\mathbb{R}$或$\mathbb{C}$, $V=\mathbb{F}^{m\times n}$. 设$A=(a_{ij}),B=(b_{ij})\in V$. 则

$$(A,B)=\operatorname{tr}(AB^*)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\bar{b}_{ij} \tag{2.5.19}$$

是V的一个内积, 基本矩阵$E_{ij},1\le i\le m,1\le j\le n$是$V$在该内积下的一组标准正交基.

证 由于AA^*是半正定矩阵, 故公式(2.5.19)显然满足内积的正定性和共轭对称性. 另外, 公式(2.5.19)显然对第一个变量是线性的, 对第二个变量是共轭线性的. 故该公式确实定义了$V=\mathbb{F}^{m\times n}$上的一个内积.

最后, 直接计算可知$(E_{ij},E_{kl})=\operatorname{tr}(E_{ij}E_{lk})=\operatorname{tr}(\delta_{jl}E_{ik})=\delta_{jl}\delta_{ik}$. □

公式(2.5.19)定义的内积称为$\mathbb{F}^{m\times n}$上的**标准内积**.

例 2.5.7. 设$\|\bullet\|$是$\mathbb{F}^{m\times n}$的标准内积诱导的长度, 则有$\|A+B\|\le\|A\|+\|B\|$.

定理2.5.4与公式(2.5.19)在第五章第六节“樊畿范数与低秩分解”有重要应用.

例2.5.8. 在$\mathbb{R}^n$或$\mathbb{C}^n$的普通内积下, 度量矩阵的行列式具有明显的几何意义. 设$\alpha_1,\alpha_2,\cdots,\alpha_n\in\mathbb{F}^n$(不必线性无关), 其度量矩阵为

$$G=\begin{pmatrix}(\alpha_1,\alpha_1)&(\alpha_1,\alpha_2)&\cdots&(\alpha_1,\alpha_n)\\(\alpha_2,\alpha_1)&(\alpha_2,\alpha_2)&\cdots&(\alpha_2,\alpha_n)\\\vdots&\vdots&\cdots&\vdots\\(\alpha_n,\alpha_1)&(\alpha_n,\alpha_2)&\cdots&(\alpha_n,\alpha_n)\end{pmatrix}.$$

令$A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$. 由于普通内积$(\alpha_i,\alpha_j)=\alpha_j^*\alpha_i$, 故由矩阵的乘法结构可知

$$G=\begin{pmatrix}\alpha_1^*\\\alpha_2^*\\\vdots\\\alpha_n^*\end{pmatrix}(\alpha_1,\alpha_2,\cdots,\alpha_n)=A^*A.$$

故行列式$|G|=||A||^2$. 由于行列式$|A|$的几何意义是以$\alpha_1,\alpha_2,\cdots,\alpha_n$为邻边的平行多面体的有向体积, 故度量矩阵的行列式是该体积的平方.

比如, 设$\alpha_1,\alpha_2\in\mathbb{R}^2$的夹角为$\theta$, 则可直接计算相应的度量矩阵的行列式如下：

$$D=\begin{vmatrix}(\alpha_1,\alpha_1)&(\alpha_2,\alpha_1)\\(\alpha_1,\alpha_2)&(\alpha_2,\alpha_2)\end{vmatrix}=\|\alpha_1\|^2\|\alpha_2\|^2-(\alpha_1,\alpha_2)^2=\|\alpha_1\|^2\|\alpha_2\|^2-\|\alpha_1\|^2\|\alpha_2\|^2\cos^2\theta=\|\alpha_1\|^2\|\alpha_2\|^2\sin^2\theta,$$

这恰好是以α_1,α_2为邻边的平行四边形的面积的平方.

思考题

1. 将内积的正定性条件去掉将如何? 是否这是无稽之谈?
2. 比较公式(2.5.5)与(2.5.6), 说明为什么不在公式(2.5.5)中将内积的绝对值符号去掉?
3. 正交性概念是通常垂直概念的推广. Gram-Schmidt 正交化方法在立体几何中有何解释?
4. 试给出标准正交基的一个直观解释.

5. 由标准正交基到另一组标准正交基的过渡矩阵有何特点?

6. 设$\mathbb{F}=\mathbb{C}$或$\mathbb{R}$. $\mathbb{F}$上的n元二次型全体是否构成$\mathbb{F}$上的线性空间? n维双线性型全体呢?

7. 试对$\mathbb{F}$上的任意m维向量x与n维向量y, 推广双线性型的概念. 这样的双线性型全体是否构成$\mathbb{F}$上的线性空间?

8. 设$(\bullet,\bullet)_i, i=1,2$是$n$维实线性空间$V$上的两个不同的内积, $\alpha,\beta\in V$. 是否可能$(\alpha,\beta)_1=0$但$(\alpha,\beta)_2\neq 0$? 是否可能$(\alpha,\alpha)_1<(\beta,\beta)_1$但$(\alpha,\alpha)_2>(\beta,\beta)_2$? 一般地, 这两个内积有何关系?

9. 试对n维实线性空间V上的双线性型讨论上题类似的问题?

2.5.3 正交补与投影

内积空间V的任意非空子集X都有唯一一个与之"垂直"的极大子空间, 称为X的**正交补**, 记为$X^\perp$, 即

$$X^\perp=\{\alpha\in V\,|\,(\alpha,\beta)=0,\,\forall\,\beta\in X\}=\{\alpha\in V|(\alpha,X)=0\}.$$

易知$X^\perp$总是V的子空间, 且$0^\perp=V, V^\perp=0$.

例 2.5.9. 在通常内积下, $\mathbb{R}^2$的子空间x轴与y轴互为正交补, $\mathbb{R}^3$的子空间x轴与yoz平面互为正交补.

定理 2.5.5. 设V是n维内积空间, U,W是V的子空间. 则

(1) $U^\perp=\{\alpha\in V\,|\,(\alpha,\beta)=0,\,\forall\,\beta\in U\}$.

(2) $V=U\oplus U^\perp$. 特别, $\dim U^\perp=n-\dim U$.

(3) $(U^\perp)^\perp=U$.

(4) $(U+W)^\perp=U^\perp\bigcap W^\perp$.

(5) $(U\bigcap W)^\perp=U^\perp+W^\perp$.

证 (1) 显然.

(2) 任取U的一个标准正交基$\alpha_1,\cdots,\alpha_s$, 将其扩充为V的一个标准正交基$\alpha_1,\cdots,\alpha_s,\alpha_{s+1},\cdots,\alpha_n$. 则$U^\perp=\mathrm{Span}\{\alpha_{s+1},\alpha_{s+2},\cdots,\alpha_n\}$. 因此$VU\oplus U^\perp, \dim U^\perp=n-\dim U$.

(3) 显然$U\subseteq U^\perp$, 再由(2)比较维数即可.

(4) 由$U,W\subseteq U+W$知$(U+W)^\perp\subseteq U^\perp\cap W^\perp$.

设$v\in U^\perp\cap W^\perp, x\in U+W$. 存在$u\in U, w\in W$使得$x=u+w$.

因此$(v,x)=(v,u+w)=(v,u)+(v,w)=0+0=0$. 即$v\in(U+W)^\perp$.

(5) 由$U^\perp,W^\perp\subseteq(U\cap W)^\perp$知$(U\cap W)^\perp\geq U^\perp+W^\perp$.

设$v\in(U\cap W)^\perp, x\in U^\perp+W^\perp$. 存在$u\in U^\perp\subseteq(U\cap W)^\perp, w\in W^\perp\subseteq(U\cap W)^\perp$使得$x=u+w$.

因此$(v,x)=(v,u+w)=(v,u)+(v,w)=0+0=0$. 即$v\in U^\perp+W^\perp$. □

例 2.5.10. 单个非零向量$v\in V$的正交补记为$v^\perp$. 由定理2.5.5可知$\dim\, v^\perp=\dim\, V-1$, 即$v^\perp$是$V$的一个超平面.

由在第一章, 矩阵$A\in\mathbb{F}^{m\times n}$的子空间$N(A)$与$R(A^*)$在$\mathbb{F}^n$中互补, $R(A)$与$N(A^*)$在$\mathbb{F}^m$中互补. 实际上, 还有下面更强的结论

定理 2.5.6. 设$A\in\mathbb{F}^{m\times n}$. 则$N(A)=R(A^*)^\perp$; $R(A)=N(A^*)^\perp$.

证 只证(1), (2)的证明类似. 设$x \in N(A)$, 即$Ax = 0$, 亦即$A^{(i)}x = 0\,, i = 1, \cdots, m$. 由$\mathbb{C}^n$中的(普通)内积定义, 这就是$\bar{A}$(注意: 不是$A$!)的每个行向量$\overline{A^{(i)}}$与$x$正交. 从而$\bar{A}$的行空间的每个向量也与$x$正交. 故$x \in R(A^*)^{\perp}$. 反之亦然. □

定理2.5.6表明, 在实数域上, 线性方程组$Ax = 0$的解空间恰好是矩阵A的列空间的正交补. 例如, 平面上的直线$ax + by = 0$实际上是子空间$\begin{pmatrix} a \\ b \end{pmatrix}^{\perp}$, 此即通过原点且与法向量$(a, b)^T$垂直的直线.

如果引入记号"$\overset{\perp}{\oplus}$"来表示"**正交直和**", 则定理2.5.6可以写为

$$\mathbb{F}^n = N(A) \overset{\perp}{\oplus} R(A^*), \quad \mathbb{F}^m = N(A^*) \overset{\perp}{\oplus} R(A) \tag{2.5.20}$$

公式(2.5.20)被G. Strang称为"线性代数基本定理", 可见其重要性, 不过因为其仅涉及矩阵及特殊的线性空间, 故称其为"**矩阵基本定理**"或"**四子空间定理**"更为恰当.

例 2.5.11. (Fourier 系数) 设$V = \{f(x) \,|\, f(x) = \sum\limits_{n=0}^{m}(a_n \cos nx + b_n \sin nx), \forall a_n, b_n \in \mathbb{R}\}$. 则$V$是一个$2m+1$维实线性空间. 定义其上的内积为:

$$(f(x), g(x)) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}\,x.$$

则$\frac{1}{\sqrt{2}}$, $\cos\, x$, $\sin\, x$, $\cos\, 2x$, $\sin\, 2x$, $\cdots$, $\cos\, mx$, $\sin\, mx$构成V 的一组标准正交基(这是内积定义中乘以系数$\frac{1}{\pi}$的唯一理由). 对任何非零常函数c有

$$c^{\perp} = 1^{\perp} = \mathrm{Span}\{\, \cos\, x,\, \sin\, x,\, \cos\, 2x,\, \sin\, 2x, \cdots,\, \cos\, mx,\, \sin\, mx\}.$$

我们在数学分析或高等数学课程中已经知道

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos\, nx\,\mathrm{d}\,x, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin\, nx\,\mathrm{d}\,x.$$

实际上, $a_n = (f(x), \cos\, nx)$与$b_n = (f(x), \sin\, nx)$恰好是$f(x)$与该组标准正交基的每个基向量的内积.

下面介绍正交补空间的再求近似解方面的应用. 先引入一个定义.

定义 2.5.3. (最佳近似) 设U是内积空间V的子空间, $\beta \in V$. 如果$\alpha \in U$满足条件:

$$\|\beta - \alpha\| \le \|\beta - \gamma\|, \quad \forall\, \gamma \in U.$$

则称α是β在U中的**最佳近似(向量)**, 常记为$\hat{\beta}$(如果不需要强调子空间U).

因此, β在U中的最佳近似就是U中与β距离最短的向量, 这样的向量存在且唯一(为什么?).

例 2.5.12. 设U是$\mathbb{R}^3$的通过原点的平面, 则任何向量$\alpha = (x,\, y,\, z)$在U上的最佳近似就是α在平面U上的投影向量, 即$\hat{\beta} = \mathrm{Proj}_U\alpha$. 故在$\mathbb{R}^3$中, 任何向量$\alpha = (x,\, y,\, z)$在子空间$x$-轴上的最佳近似为$(x\,, 0,\, 0)$.

下面的定理称为**“最佳近似定理”**.

定理 2.5.7. 设U是内积空间V的子空间, $\beta \in V$, $\alpha \in U$. 则α是β在U上的最佳近似向量 $\iff$ $\beta - \alpha \in U^{\perp}$.

证 对$\forall \gamma \in U$, 因为$\beta - \gamma = (\beta - \alpha) + (\alpha - \gamma)$, 且$\beta - \alpha \in U^{\perp}$, $\alpha - \gamma \in U$, 所以

$$\|\beta - \gamma\|^2 = \|\beta - \alpha\|^2 + \|\alpha - \gamma\|^2 \geq \|\beta - \alpha\|^2.$$

因此, 对任意$\gamma \in U$, 有

$$\|\beta - \alpha\| \leq \|\beta - \gamma\|,$$

即α是β在U上的最佳近似向量. □

由最佳近似定理, 求β在子空间U中的最佳近似向量, 等价于求β在直和分解$V = U \oplus U^{\perp}$下的表达式, 即若$\beta = \alpha + \gamma$, 其中$\alpha \in U$, $\gamma \in U^{\perp}$, 则α就是β在U上的最佳近似向量$\hat{\beta}$.

先考察$U = \mathbb{R}v$是1-维子空间的情形. 设β在U中的最佳近似为$\hat{\beta} = xv$, 则由最佳近似定理, 内积$(\beta - xv, v) = 0$, 故

$$x = \frac{(\beta, v)}{(v, v)} \tag{2.5.21}$$

因此$\hat{\beta} = \frac{(\beta,v)}{(v,v)}v$, 此即$\beta$在$v$上的投影向量$\mathrm{Proj}_v\beta$, 而$\gamma = \beta - \mathrm{Proj}_v\beta$恰是$\beta$关于子空间$U$的正交投影向量.

如果$U = \mathbb{R}v_1 \oplus \mathbb{R}v_2$是2-维子空间, 则

$$\hat{\beta} = x_1v_1 + x_2v_2 \tag{2.5.22}$$

且内积$(\beta - \hat{\beta}, v_i) = 0,\ i = 1, 2$, 即有下面的方程组

$$(\beta - x_1v_1 - x_2v_2, v_i) = 0,\ i = 1, 2 \tag{2.5.23}$$

由于v_1, v_2线性无关, 故方程组(2.5.23)有唯一解, 由此即可求得$\hat{\beta}$. 特别, 若v_1, v_2还是正交的, 则

$$x_1 = \frac{(\beta, v_1)}{(v_1, v_1)},\ x_2 = \frac{(\beta, v_2)}{(v_2, v_2)} \tag{2.5.24}$$

因此$\hat{\beta}$在表达式(2.5.22)中的两个分量x_1v_1与x_2v_2恰好分别是β在v_1与v_2上的投影向量$\mathrm{Proj}_{v_1}\beta$与$\mathrm{Proj}_{v_2}\beta$. 故此时β在子空间U中的最佳近似就是β在U的各个正交基向量的投影向量之和$\mathrm{Proj}_{v_1}\beta + \mathrm{Proj}_{v_2}\beta$. 故有下述计算最佳近似的结论.

命题 2.5.3. 设U是内积空间V的子空间, $\beta \in V, \alpha_1, \cdots, \alpha_s$是$U$的一个标准正交基, 则$\beta$在$U$上的最佳近似向量为

$$\hat{\beta} = (\beta, \alpha_1)\alpha_1 + (\beta, \alpha_2)\alpha_2 + \cdots + (\beta, \alpha_s)\alpha_s. \tag{2.5.25}$$

由命题2.5.3可知内积空间中的向量β在任何子空间$U \neq 0$中的投影向量$\mathrm{Proj}_U\beta$即其最佳近似向量, 就是其在U的一组(标准)正交基的各个基向量上的投影向量之和. 此点实际上完全蕴含于Gram-Schmidt正交化方法之中.

思考题

1. 试用正交分解理论解释勾股定理.

2. 试利用正交分解理论在空间中建立关于面积的勾股定理. 试讨论高维度量(比如体积等)的勾股定理.

3. 在$\mathbb{R}^3$中定义"广义内积"$(x,y)=x_1y_1+x_2y_2-x_3y_3$, 正交性有何变化?

2.5.4 酉矩阵与等距变换

在内积空间中, 可以问线性变换将三角形变成了什么图形? 是否仍然将抛物线变为抛物线? 等等. 此时问题的本质是线性变换是否保持长度和角度.

定义 2.5.4. 设V是内积空间, $\sigma\in\operatorname{End}V$. 如果$\sigma$保持向量间的距离, 即对任意$\alpha,\beta\in V$, 均有$d(\sigma(\alpha),\sigma(\beta))=d(\alpha,\beta)$, 则称$\sigma$是**等距变换** 或**保距变换**.

等距变换显然是单变换,故是可逆变换. 恒等变换是等距变换; $\mathbb{R}^2$中的旋转也是等距变换; 但比例系数的模不等于1的位似变换均非等距变换.

定理 2.5.8. 设V是内积空间, $\sigma\in\operatorname{End}V$. 则$\sigma$是等距变换 $\Longleftrightarrow$ σ保持向量的长度 $\Longleftrightarrow$ σ保持内积.

证 保持距离显然等价于保持长度, 保持内积则蕴含保持长度. 现设σ保持长度, 则

$$\begin{aligned}(\sigma(\alpha+\beta),\,\sigma(\alpha+\beta)) &= (\sigma(\alpha),\sigma(\alpha))+(\sigma(\alpha),\sigma(\beta))+(\sigma(\beta),\sigma(\alpha))+(\sigma(\beta),\sigma(\beta))\\ &= (\alpha,\alpha)+(\sigma(\alpha),\sigma(\beta))+\overline{(\sigma(\alpha),\sigma(\beta))}+(\beta\,,\beta).\end{aligned}$$

而

$$\begin{aligned}(\alpha+\beta,\,\alpha+\beta) &= (\alpha,\,\alpha)+(\alpha\,,\beta)+(\beta,\,\alpha)+(\beta,\,\beta)\\ &= (\alpha,\alpha)+(\alpha,\,\beta)+\overline{(\alpha,\,\beta)}+(\beta\,,\beta).\end{aligned}$$

故

$$(\sigma(\alpha),\,\sigma(\beta))+\overline{(\sigma(\alpha),\,\sigma(\beta))}=(\alpha,\,\beta)+\overline{(\alpha,\,\beta)},$$

即$(\sigma(\alpha),\,\sigma(\beta))$与$(\alpha,\,\beta)$具有相同的实部. 但易证$\operatorname{Re}(\alpha,\,\mathrm{i}\beta)=\operatorname{Im}(\alpha,\,\beta)$, 从而$(\sigma(\alpha),\,\sigma(\beta))$与$(\alpha,\,\beta)$也具有相同的虚部, 因此$(\sigma(\alpha),\,\sigma(\beta))=(\alpha,\,\beta)$, 故定理成立. □

定理 2.5.9. 设V是n维内积空间, $\alpha_1,\alpha_2,\cdots,\alpha_n$是$V$ 的一组标准正交基, $\sigma\in\operatorname{End}V$, A是σ在该组基下的矩阵. 则σ是等距变换 $\Longleftrightarrow$ A是酉矩阵.

证 对任意$\alpha\in V$, 设$\alpha=(\alpha_1,\cdots,\alpha_n)x, x\in\mathbb{C}^n$, 因$\alpha_1,\cdots,\alpha_n$是标准正交基, 故其长度为$\|\alpha\|=\sqrt{(\alpha,\alpha)}=\sqrt{x^*x}$. 设$\sigma(\alpha)=\beta=(\alpha_1,\alpha_2,\cdots,\alpha_n)Ax$, 则$\|\beta\|=\sqrt{(Ax)^*(Ax)}=\sqrt{x^*(A^*A)x}$. 故若$A^*A=I$, 则$\|\sigma(\alpha)\|=\|\beta\|=\sqrt{x^*x}=\|\alpha\|$, 即$\sigma$保持长度, 故由定理2.5.8知$\sigma$是等距变换.

反过来, 设σ是等距变换, 则对任意向量$x\in\mathbb{C}^n$(或$\mathbb{R}^n$), 有

$$(\beta,\beta)=x^*(A^*A)x=(\alpha,\alpha)=x^*x.$$

故$x^*(I-A^*A)x=0,\forall x$, 故$I-A^*A=0$, 即$A^*A=I$. 所以A是酉矩阵. □

需要注意, 定理2.5.9中的条件"标准正交"不可去, 即等距变换在一般基下的矩阵未必是酉矩阵, 见下例.

例 2.5.13. 设σ为实平面上逆时针旋转$\pi/2$, 则它在基$\{e_1, e_1+e_2\}$下的矩阵为

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

这显然不是酉矩阵.

由于定理 2.5.9, 欧氏空间的等距变换又称为**正交变换**；而复内积空间的等距变换也称为**酉变换**.

V的等距变换之积仍是等距变换, 等距变换之逆仍是等距变换. 从而等距变换的全体组成一个群, 称为V的等距变换群. 特别, 欧氏空间的全体正交变换构成正交变换群. 相应地, 全体正交矩阵组成$GL_n(\mathbb{R})$的一个子群, 称为正交矩阵群.

定理 2.5.10. 设V是内积空间, $\sigma \in \operatorname{End} V$. 则$\sigma$是等距变换 $\Longleftrightarrow$ σ将标准正交基变为标准正交基.

证 设$\alpha_1, \cdots, \alpha_n$是$V$的任意一组标准正交基, A是σ在该组基下的矩阵, $\beta_j = \sigma(\alpha_j)$, $1 \leq j \leq n$. 若σ是等距变换, 则由定理 2.5.8, σ保持内积与长度, 故$\beta_1, \cdots, \beta_n$仍是一个标准正交基.

反过来, 如果σ将标准正交基变为标准正交基, 则σ保持所有单位向量的长度, 故也保持所有向量的长度, 因此是等距变换. □

例 2.5.14. 二维正交矩阵的分类. 实直线$\mathbb{R}$上的正交变换只有$\pm I$, 即恒等变换及其负变换(将x变为$-x$). 实平面$\mathbb{R}^2$上的正交变换可以计算如下: 选定标准基e_1, e_2. 设正交变换σ在标准基下的矩阵为A, 则由定理 2.5.10知A是正交矩阵, 于是

$$Q = \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \tag{2.5.26}$$

或

$$P = \begin{pmatrix} c & s \\ s & -c \end{pmatrix} \tag{2.5.27}$$

其中$c^2 + s^2 = 1$. 矩阵Q对应的正交变换是我们熟悉的旋转, 而矩阵P既是正交矩阵又是对称矩阵, 其对应的正交变换是一种反射(对称轴为$y = \dfrac{1-c}{s}x$, 见习题 39), 故称为反射矩阵.

例 2.5.15. 三维正交矩阵的分类. 3阶正交矩阵的特征值或者全是实数, 或者只有一个实数, 且实特征值只能是± 1, 因此必正交相似于下面 6 种矩阵:

$$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix},$$
$$\begin{pmatrix} -1 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & & \\ & \cos\theta & \sin\theta \\ & -\sin\theta & \cos\theta \end{pmatrix}, \quad \begin{pmatrix} -1 & & \\ & \cos\theta & \sin\theta \\ & -\sin\theta & \cos\theta \end{pmatrix}.$$

一般将三维空间保持长度的映射称为刚体运动. 上述结论表明, 三维空间的刚体运动是平移和上述 6 种变换的合成. 如果上面的矩阵有复特征值, 则意味着相应的正交变换是绕一固定轴的旋转, 且旋转的平面与轴正交, 固定轴即是该变换的1-维(实)特征子空间, 而旋转面是相应于复特征值的特征子空间的直和.

例 2.5.16. 设

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

是正交矩阵, 求相应的正交变换的旋转轴与旋转的角度.

解 A的一个特征值为1, 设其对应的特征向量为$\alpha = (x_1, x_2, x_3)^T$, 则由方程组$A\alpha = \alpha$可得$\alpha = (\sqrt{3}, 1, 1)^T$, 故旋转轴为过点$(0,0,0)$与$(\sqrt{3}, 1, 1)$的直线. 为求$A$的旋转角, 可利用矩阵的相似不变量–迹, 即设

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} = Q^{-1}AQ,$$

则$\operatorname{tr} B = \operatorname{tr} A$, 故$1 + 2\cos\theta = \frac{1}{4} - \frac{1}{2} - \frac{1}{2} = -\frac{3}{4}$, 所以$\cos\theta = -\frac{7}{8}$.

现在回答本节开始的几个问题. 保距变换保持长度和内积, 从而保持向量之间的角度, 因此保距变换不改变图形的形状, 比如保距变换将圆仍变为圆, 将三角形仍变为三角形等等. 但一般的线性变换则可能将图形变得面目全非. 比如, 向x轴的投影变换把圆心在原点的单位圆变成了x轴上的线段, 而将抛物线$y = x^2$变为整个x轴. 请读者自行考察其它一些简单图形在线性变换下的形状.

最著名的正交变换当属**Householder**[35]**变换**与**Givens** [36]**旋转**, 分别介绍如下.

例 2.5.17. 设$v \in \mathbb{C}^n$是非零向量, 定义n阶复矩阵H如下:

$$H = I - \frac{2vv^*}{v^*v}. \tag{2.5.28}$$

矩阵H称为**Householder 矩阵**. 显然有$H^* = H, H^2 = I$. 由矩阵H定义的线性变换H_v称为**Householder变换**, 即对任意$x \in \mathbb{C}^n$

$$H_v x = x - \frac{2vv^*}{v^*v}x. \tag{2.5.29}$$

容易看出, Householder变换H_v实际上是关于超平面$v^\perp$的镜面反射(即将$v^\perp$的向量固定, 而将v变为$-v$). 如果将向量x按直和$\mathbb{C}^n = \operatorname{Span}\{v\} \oplus v^\perp$正交分解为

$$x = P_v x + P_v^\perp x,$$

则

$$H_v x = P_v^\perp x - P_v x.$$

如下图.

[35]Alston Scott Householder(1904-1993), 著名美国数值分析学家和生物数学家, 他的开创性研究奠定了计算机科学的基础. 国际数值代数会议自 1996 年改称Householder会议. 曾任美国数学会主席, 美国工业与应用数学会主席和 Association for Computing Machinery 主席.

[36]James Wallace Givens(1910-1993), 美国数学家, 计算机科学先驱. 曾任美国工业与应用数学会主席.

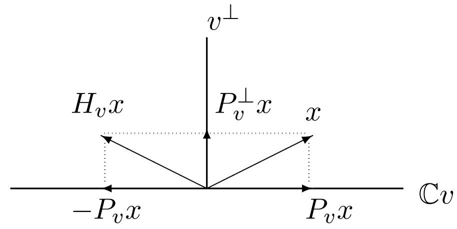

图2.4.1

例 2.5.18. 设$v = e_3 \in \mathbb{R}^3$是标准向量, 则相应的Householder变换H_v为

$$H_v x \;=\; x - 2vv^* x = x - 2E_{33}x = \begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix} x$$

正是关于xoy平面的反射.

例 2.5.19. 设u是n维单位向量, 求n阶实 Householder 矩阵$I - 2uu^T$的特征值、迹和行列式.

解 由 Sylvester 降幂公式,

$$|\lambda I - (I - 2uu^T)| = |(\lambda - 1)I + 2uu^T| = (\lambda - 1)^{n-1}|\lambda - 1 + 2u^T u| = (\lambda - 1)^{n-1}(\lambda + 1).$$

由此知, $\lambda = 1$是$n - 1$重根, 而$\lambda = -1$是1重根. 因此, $\mathrm{tr}(I - 2uu^T) = n - 2$; $|I - 2uu^T| = -1$.

例 2.5.20. 由矩阵(2.5.26)决定的旋转称为**Jacobi**[37]**旋转**或(2-维)**Givens旋转**. 一般地, 设$c, s \in \mathbb{R}, c^2 + s^2 = 1, \theta = \arctan\dfrac{s}{c}$, 则将$n$阶实矩阵

$$G(i,\, j,\, \theta) = I_n - (1 - c)(E_{ii} + E_{jj}) + s(E_{ij} - E_{ji}) \tag{2.5.30}$$

称为**Givens旋转矩阵**. 易知这是一个正交矩阵.

Householder变换与Givens变换都可以用来将矩阵中某些特定的元素变为0, 是矩阵快速计算中不可或缺的方法.

作为本节的结束, 我们讨论实对称矩阵对应的欧氏空间的线性变换—对称变换.

定义 2.5.5. 设σ是欧氏空间V的线性变换. 如果对任意$\alpha, \beta \in V$, 均有

$$(\sigma(\alpha), \beta) = (\alpha, \sigma(\beta)) \tag{2.5.31}$$

则称σ是一个**对称变换**.

位似变换都是对称变换. 反射变换、投影变换也是对称变换. 一般地有

定理 2.5.11. 内积空间V的线性变换σ是对称变换 $\iff$ σ在一组标准正交基下的矩阵是Hermite矩阵.

[37]Carl Gustav Jacob Jacobi(1804-1851), 著名德国数学家, 数学上有著名的 Jacobian Conjecture (雅可比猜想), 见本书第五章.

证 设σ在一组标准正交基$\alpha_1, \cdots, \alpha_n$下的矩阵为$A$. 设$\alpha, \beta \in V$在该标准正交基下的坐标分别为$x, y$.

充分性. 设A是Hermite矩阵. 则$\sigma(\alpha) = (\alpha_1, \cdots, \alpha_n)Ax$, $\sigma(\beta) = (\alpha_1, \cdots, \alpha_n)Ay$, 因此

$$(\sigma(\alpha), \beta) = (Ax, y) = y^*Ax = (A^*y)^*x = (x, Ay) = (\alpha, \sigma(\beta)).$$

故σ是对称变换.

必要性. 设σ是对称变换. 因为

$$(\sigma(\alpha), \beta) = (Ax, y) = y^*Ax; (\alpha, \sigma(\beta)) = (x, Ay) = (Ay)^*x = y^*A^*x.$$

故$y^*Ax = y^*A^*x$对任意$x, y \in \mathbb{F}^n$均成立. 因此$e_i^*Ae_j = e_i^*A^*e_j$对所有$1 \le i, j \le n$均成立, 此即$A = A^*$. □

例 2.5.21. 由二维正交矩阵的分类可知, 如果一个线性变换既是对称变换又是正交变换, 则它必然是反射. 因此, 平面上非反射变换的对称变换不是等距变换, 从而对称变换和我们通常的理解相去甚远, 比如线性变换

$$(x, y)^T \mapsto (x + y, x)^T$$

是平面上的对称变换, 它将单位正方形$0 \le x, y \le 1$变为如下的平行四边形：

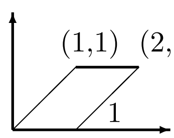

图2.4.2

由对称变换可以自然导出另一类重要的线性变换, 其定义如下.

定义 2.5.6. 设σ, τ是欧氏空间V的两个线性变换. 如果对任意$\alpha, \beta \in V$, 均有

$$(\sigma(\alpha), \beta) = (\alpha, \tau(\beta)) \tag{2.5.32}$$

则称τ是σ的一个**伴随变换**.

对称变换显然是自己的一个伴随变换. 但从定义定义 2.5.6中并不容易看出欧氏空间的任意线性变换都有伴随变换, 这是与以前所有定义的一个重要差别, 所以应该**首先质疑其合理性**. 先考察一个例子.

例 2.5.22. 设σ是复空间$\mathbb{C}^2$上的线性变换

$$\sigma : (x, y)^T \mapsto (a_1x + b_1y, a_2x + b_2y)^T.$$

则容易验证下面的线性变换

$$\tau : (x, y)^T \mapsto (\bar{a}_1x + \bar{a}_2y, \bar{b}_1x + \bar{b}_2y)^T$$

确实是σ的一个伴随变换. 注意到τ在标准基下的矩阵恰好是σ在标准基下的矩阵的共轭转置, 因此, 伴随变换实际上是矩阵共轭转置的几何意义.

一般地有

定理 2.5.12. 设σ是内积空间V的线性变换, 则σ的伴随变换存在且唯一, 记为σ^*.

证 取定V的一组标准正交基, 设σ(在该组基下)的矩阵为A. 由"线性变换基本定理", 存在线性变换τ使其矩阵为A^*. 对任意$\alpha, \beta \in V$, 设其坐标分别为x, y. 则由坐标变换定理即定理2.3.4, $\sigma(\alpha)$的坐标为Ax, $\tau(\beta)$的坐标为A^*y. 因此

$$(\sigma(\alpha), \beta) = y^*(Ax);\ (\alpha, \tau(\beta)) = (A^*y)^*x = y^*Ax.$$

故$(\sigma(\alpha), \beta) = (\alpha, \tau(\beta))$, 即$\tau$是$\sigma$的一个伴随变换.

假设ρ也是σ的伴随变换, 则$(\sigma(\alpha), \beta) = (\alpha, \tau(\beta)) = (\alpha, \rho(\beta)), \forall \alpha, \beta \in V$. 于是, $(\alpha, (\tau - \rho)(\beta)) = 0$. 由$\alpha$的任意性可知$(\tau - \rho)(\beta) = 0$. 再由$\beta$的任意性可知$\tau - \rho = 0$. $\square$

如果$\sigma = \sigma^*$, 则称σ是**自伴随的**或**自伴的**. 比如, 内积空间的对称变换是自伴变换.

由定理2.5.12及其证明可知, 伴随变换(在标准正交基下)的矩阵恰好是原变换的矩阵的共轭转置, 因此有下面的

定理 2.5.13. 设σ, τ是内积空间V的两个线性变换, 它们在取定标准正交基下的矩阵分别为A, B. 则

(1) $(\sigma^*)^* = \sigma$.

(2) $(\sigma + \tau)^* = \sigma^* + \tau^*$.

(3) $(\lambda\sigma)^* = \overline{\lambda}\sigma^*, \forall \lambda \in \mathbb{C}$.

(4) $(\sigma\tau)^* = \tau^*\sigma^*$.

(5) $\tau = \sigma^* \iff B = A^*$.

从内积空间的角度再来看线性变换基本定理和线性代数基本定理, 就可以说"线性变换与矩阵是一码事", 线性变换反映线性空间的几何联系, 矩阵则反映线性空间的代数特征. 比如, 位似将所有向量放大同一倍数, 纯量矩阵则将每个向量的坐标放大同一倍数; 伸缩变换将所有向量沿坐标轴的方向各自伸缩若干倍数, 对角矩阵则将每个向量的坐标分别放缩若干倍数; 可逆变换的核为0, 可逆矩阵的零空间为0; 等距变换保持所有向量长度不变, 酉矩阵乘任何向量的内积等于该向量自己的内积; 等等.

思考题

1. 平面(即$\mathbb{R}^2$)上的非等距线性变换不能保持所有向量的长度, 但可否保持所有角度?

2. 空间(即$\mathbb{R}^3$)中的非等距线性变换能否保持一些向量的长度? 能否将某个半径为1的圆还变为半径为1的圆? 特别, 空间中的幂零变换能否保持一些向量的长度? 幂等变换保持哪些向量的长度?

3. 平面上的反射变换能否由旋转实现? 反过来呢?

4. 对称变换并不保持图形的对称性, 如何为"对称"一词找一个恰当的几何解释?

5. 反对称矩阵对应的线性变换有何特点?

6. 对称变换是否在任何一组基下的矩阵均为Hermite矩阵? 在某组基下的矩阵为Hermite矩阵的线性变换是否一定是对称变换?

7. 定理2.5.13中的"某组标准正交基"是否可以减弱为"某组基"?

第六节 张量积

2.6.1 线性空间的(外)直和

子空间的交与和都是由已知线性空间构造新线性空间的方法, 但这些方法局限在给定空间的内部. 本节我们研究如何由两个给定的$\mathbb{F}$空间U与V构造新的线性空间. 最简单的办法是仿照子空间的直和, 构造所谓的"直和", 其定义如下.

考虑集合U与V的卡氏积(也称笛卡尔积)$U \times V = \{(u,v)\,|\,u \in U, v \in V\}$; 按分量定义$U \times V$的加法与数乘运算, 即对$(u,v),(u_1,v_1),(u_2,v_2) \in U \times V$以及$\lambda \in \mathbb{F}$, 定义:

$$(u_1,v_1)+(u_2,v_2)=(u_1+u_2,\,v_1+v_2), \qquad \lambda(u,v)=(\lambda u,\lambda v). \tag{2.6.1}$$

则容易验证, $U \times V$在如上的加法和数乘下构成线性空间, 称为U与V的直和, 记为$U \oplus V$, U与V分别称为**直和项**.

显然有关于直和的下述基本定理.

定理 2.6.1. (直和空间的基) 设$\alpha_1, \alpha_2, \cdots, \alpha_n$(称为$\alpha$-基)与$\beta_1, \beta_2, \cdots, \beta_m$(称为$\beta$-基)分别是$U$与$V$的一组基, 则向量组

$$(\alpha_i,0),\,(0,\beta_j), 1 \le i \le n,\, 1 \le j \le m \tag{2.6.2}$$

构成$U \oplus V$的一组基, 称为$\alpha \oplus \beta$-基. 特别地,

$$\dim(U \oplus V) = \dim U + \dim V. \tag{2.6.3}$$

注. 如果U与V还是内积空间, 则可以按例 2.5.4定义$U \oplus V$的内积, 即定义

$$((u_1,v_1),(u_2,v_2)) = a(u_1,u_2)_U + b(v_1,v_2)_V \tag{2.6.4}$$

其中a,b是任意给定的正数. 则$U \oplus V$在上述内积下构成内积空间.

一般将子空间的直和称为“内直和”, 而将上面定义的直和称为“外直和”. 但这两种直和本质上是一样的. 这是因为“外直和”$U \oplus V$中的两个直和项U与V可以分别看成是线性空间$W = U \oplus V$的子空间$\{(u,v) \in W\,|\,u \in U,\, v=0\}$与$\{(u,v) \in W\,|\,v \in V,\, u=0\}$. 例如实平面$\mathbb{R}^2$既可以看成是由$x$轴和$y$轴的内直和产生的, 也可以看成是由两个数轴的外直和产生的, 因为这两个数轴可以分别看成是x轴和y轴.

常将外直和$U \oplus V$中的向量(u,v)改记为$u+v$, 这样外直和与内直和就完全统一了. 以下将不区分外直和与内直和.

现考虑直和空间的线性变换. 设$\sigma_i \in \mathrm{Hom}(U_i, V_i)$, $i=1,2$. 则可以定义$U_1 \oplus U_2$到$V_1 \oplus V_2$的映射σ如下(其中$u_i \in U_i$):

$$\sigma(u_1+u_2) = \sigma_1(u_1)+\sigma_2(u_2) \tag{2.6.5}$$

直接验证可知σ是一个线性变换, 称为σ_1与σ_2的直和, 记为$\sigma_1 \oplus \sigma_2$.

关于线性变换的直和有下面的定理:

定理 2.6.2. 设线性变换σ_1与σ_2关于α-α'-基与β-β'-基的矩阵分别为A与B, 则$\sigma_1 \oplus \sigma_2$关于$\alpha \oplus \beta$-$\alpha' \oplus \beta'$-基的矩阵为$A \oplus B$.

证 由于$\sigma_1 \oplus \sigma_2$在$\alpha \oplus \beta$-基上的作用是按照分量进行的, 故

$$\sigma_1 \oplus \sigma_2(\alpha_i) = \sigma_1(\alpha_i), \quad \sigma_1 \oplus \sigma_2(\beta_j) = \sigma_2(\beta_j).$$

因此

$$\sigma_1 \oplus \sigma_2(\alpha \oplus \beta) = (\sigma_1(\alpha_1), \cdots, \sigma_1(\alpha_n), \sigma_2(\beta_1), \cdots, \sigma_2(\beta_m)) = (\alpha \oplus \beta)(A \oplus B).$$

□

例 2.6.1. 设σ, $\tau \in (\mathbb{R}^2)^*$分别为函数$\sigma(x,y)=ax+by$与$\tau(x,y)=cx+dy$. 则$\sigma\oplus\tau\in \mathrm{Hom}(\mathbb{R}^2\oplus\mathbb{R}^2,\mathbb{R}\oplus\mathbb{R})$为:

$$\sigma\oplus\tau(x,\,y,\,u,\,v)=(ax+by,\,cu+dv).$$

进一步, 直接计算可知$\sigma\oplus\tau$关于标准基-标准基的矩阵为

$$C=\begin{pmatrix} a & b & 0 & 0\\ 0 & 0 & c & d\end{pmatrix}$$

注意σ与τ关于标准基的矩阵分别为$A=(a\quad b)$与$B=(c\quad d)$, 所以确有$C=A\oplus B$.

定理2.6.2可以推广到任意有限个线性空间的直和, 即有下面的定理(读者可自行验证).

定理2.6.3. 设$\sigma_i\in\mathrm{Hom}(U_i,\,V_i)$, $1\le\ i\ \le\ n$. 再设σ关于α_i-α_i'-基的矩阵为A_i. 则$\sum\limits_{i=1}^{n}\oplus\sigma_i\in \mathrm{Hom}(\sum\limits_{i=1}^{n}\oplus U_i,\,\sum\limits_{i=1}^{n}\oplus V_i)$关于$\sum\limits_{i=1}^{n}\oplus\alpha_i$-$\sum\limits_{i=1}^{n}\oplus\alpha_i'$-基的矩阵为$\sum\limits_{i=1}^{n}\oplus A_i$.

上面的过程实际上是利用较小空间的线性变换构造较大空间的线性变换. 这个过程可以反过来考虑, 即讨论如何将大空间的线性变换化为小空间的线性变换的直和. 设$V\ =\ U\oplus W$, $\sigma\ \in\ \mathrm{End}V$. 如果$\sigma$分解为$\sigma_1\ \in\ \mathrm{End}U$与$\sigma_2\ \in\ \mathrm{End}W$的直和, 即$\sigma\ =\ \sigma_1\oplus\sigma_2$, 则$\sigma(u+w)=\sigma_1(u)+\sigma_2(w)$. 特别, 对任意$u\in U$, $w\in W$, 有

$$\sigma(u)=\sigma_1(u)\in U,\qquad \sigma(w)=\sigma_2(w)\in W.$$

换句话说, $\sigma(U)\subseteq U$, $\sigma(W)\subseteq W$. 这样的子空间U(与W)称为σ的**不变子空间**. 平凡子空间是任何线性变换的不变子空间. 另外, 核空间与像空间也是不变子空间.

定理 2.6.4. 设$V\ =\ V_1\oplus V_2\oplus\cdots\oplus V_s$, $\sigma\in\mathrm{End}V$. 如果诸V_i均为σ的不变子空间, 则存在$\sigma_i\in\mathrm{End}V_i$, $1\le\ i\ \le s$, 使得

$$\sigma\ =\ \sigma_1\oplus\sigma_2\oplus\cdots\oplus\sigma_s.$$

证 由于诸V_i均为σ的不变子空间, 故可定义诸V_i的映射σ_i如下:

$$\sigma_i(x)=\sigma(x),\,\forall x\in V_i.$$

σ_i显然是线性变换. 由直和的定义知$v\in V$有唯一分解

$$v\ =\ v_1+v_2+\cdots+v_s,\quad v_i\in V_i,$$

因此

$$\begin{aligned}&(\sigma_1\oplus\sigma_2\oplus\cdots\oplus\sigma_s)(v)\\ &=\sigma_1(v_1)+\sigma_2(v_2)+\cdots+\sigma_s(v_s)\\ &=\sigma(v_1)+\sigma(v_2)+\cdots+\sigma(v_s)\\ &=\sigma(v_1+v_2+\cdots+v_s)=\sigma(v),\end{aligned}$$

即$\sigma\ =\ \sigma_1\oplus\sigma_2\oplus\cdots\oplus\sigma_s$. □

一般将上述诸σ_i称为线性变换σ在子空间V_i上的限制, 记为$\sigma|_{V_i}$.

推论 2.6.1. 设$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$, $\sigma \in \mathrm{End}V$. 如果诸V_i均为σ的不变子空间, 则存在V的基, 使得σ在该基下的矩阵是分块对角矩阵.

证 由定理2.6.4, 可设$\sigma = \sigma_1 \oplus \sigma_2 \oplus \cdots \oplus \sigma_s$. 设诸$\sigma_i$在$V_i$的$\alpha_i$-基下的矩阵是$A_i$, 则$\sigma$在$\sum_{i=1}^{s} \oplus \alpha_i$-基下的矩阵为

$$\sum_{i=1}^{s} \oplus A_i.$$

□

例 2.6.2. 设$\sigma \in \mathrm{End}V$. 对任意非零向量v, 投影变换$x \mapsto P_v x$是幂等变换. 对任何幂等变换σ有

$$V = \mathrm{Ker}(\sigma) \oplus \mathrm{Im}(\sigma) \tag{2.6.6}$$

这是因为$v = (v - \sigma(v)) + \sigma(v)$, 其中$v - \sigma(v) \in \mathrm{Ker}(\sigma)$, $\sigma(v) \in \mathrm{Im}(\sigma)$且$\mathrm{Ker}(\sigma) \cap \mathrm{Im}(\sigma) = 0$. 现选取$\mathrm{Ker}(\sigma)$与$\mathrm{Im}(\sigma)$的各一组基, 它们的并构成$V$的一组基, 在这组基下, σ的矩阵为分块矩阵

$$0_{n-r} \oplus I_r,$$

其中r是σ的秩, n是V的维数. 换句话说, σ在$\mathrm{Ker}(\sigma)$上的限制是0变换, 在$\mathrm{Im}(\sigma)$上的限制是恒等变换.

2.6.2 线性空间的张量积

由于两个线性空间的直和中的运算是按分量进行的, 故两个直和项没有相互“融入”, 因此直和空间没有包含任何新的信息. 设想将U与V的基相互融合, 即以

$$(\alpha_i, \beta_j), 1 \le i \le n, 1 \le j \le m \tag{2.6.7}$$

为基构造新的线性空间, 称为U与V的**张量积**[38], 记为$U \otimes V$. 一般, 将$U \otimes V$的基(2.6.7)写成

$$\alpha_i \otimes \beta_j, 1 \le i \le n, 1 \le j \le m \tag{2.6.8}$$

从而$U \otimes V$中的向量形如

$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} \alpha_i \otimes \beta_j. \tag{2.6.9}$$

由上述定义立即可得下述的

定理 2.6.5. 设U, V, W均是有限维线性空间, 则
(1) $\dim(U \otimes V) = (\dim U)(\dim V)$.
(2) $(U \otimes V) \otimes W \cong U \otimes (V \otimes W)$.

所以张量积是线性空间的“乘法”(定理2.6.5的第2条说明张量积满足结合律). 无限维线性空间的张量积可类似定义.

[38]张量(运算)由意大利数学家Gregorio Ricci-Curbastro (1853-1925)提出的, 后经其学生Tullio Levi-Civita (1873—1941)发展而成为众多学科的重要理论和工具(包括相对论).

例 2.6.3. 考察行向量空间$U = \mathbb{F}^{1\times n}$与列向量空间$V = \mathbb{F}^{m\times 1}$ 的张量积. 选取U与V的标准基$e_1^T, \cdots, e_n^T$与$e_1^{(m)}, \cdots, e_m^{(m)}$ (其中$e_i^{(m)}$表示第i个m维标准向量). 则$\mathbb{F}^{1\times n}\otimes\mathbb{F}^{m\times 1}$有一组标准基如下:

$$e_i^T\otimes e_j^{(m)},\ 1\le i\le n,\ 1\le j\le m. \tag{2.6.10}$$

如果将基向量$e_i^T\otimes e_j^{(m)}$对应到基本矩阵E_{ji}, 则可得

$$\mathbb{F}^{1\times n}\otimes\mathbb{F}^{m\times 1}\cong\mathbb{F}^{m\times n}. \tag{2.6.11}$$

例 2.6.4. 证明$\mathbb{F}[x]\otimes\mathbb{F}[y]\cong\mathbb{F}[x,y]$.

证 分别取$\mathbb{F}[x]$与$\mathbb{F}[y]$的标准基

$$1,\ x,\ x^2,\ \cdots,\ x^n,\ \cdots$$

以及

$$1,\ y,\ y^2,\ \cdots,\ y^m,\ \cdots$$

则可得$\mathbb{F}[x]\otimes\mathbb{F}[y]$的一组基如下

$$x^n\otimes y^m,\ m,n\ge 0.$$

将$x^n\otimes y^m$对应到x^ny^m即给出需要的一个同构变换. $\square$

例 2.6.5. 设U,V是$\mathbb{F}$上的两个有限维内积空间, $u_i, i = 1,...,m$与$v_j, j = 1,...,n$分别是U与V的一组标准正交基. 则$U\otimes V$也具有较为自然的内积空间结构, 只需将$u_i\otimes v_j (1\le i\le m, 1\le j\le n)$定义为$U\otimes V$的一组标准正交基即可.

2.6.3 线性变换的张量积

以下我们讨论张量积空间的线性变换. 设$\sigma_i\in \mathrm{Hom}(U_i, V_i)$, $i = 1,2$. 类似于直和的情形, 可以定义$U_1\otimes U_2$到$V_1\otimes V_2$的张量积映射$\sigma_1\otimes\sigma_2$如下(其中$u_i\in U_i$):

$$\sigma_1\otimes\sigma_2(u_1\otimes u_2) = \sigma_1(u_1)\otimes\sigma_2(u_2) \tag{2.6.12}$$

容易验证(习题51), $\sigma_1\otimes\sigma_2$是一个线性变换, 称为σ_1与σ_2的张量积. 需要注意, 张量积空间$U_1\otimes U_2$中的元素的一般形式如式(2.6.9), 因此公式(2.6.12)只定义了$U_1\otimes U_2$中的一部分向量的像, 其余向量的像可由线性扩张得到, 见定理2.3.1.

为了得到线性变换的张量积的矩阵表示, 我们先考察两个例子.

例 2.6.6. 设$\sigma\in(\mathbb{F}^2)^*$, $\tau\in \mathrm{Hom}(\mathbb{F},\mathbb{F}^2)$. 设$\sigma$与$\tau$关于标准基-标准基的矩阵分别为$A = (a\ \ b)$与$B = (x\ \ y)^T$. 则$\sigma\otimes\tau\in\mathrm{Hom}(\mathbb{F}^2\otimes\mathbb{F},\mathbb{F}\otimes\mathbb{F}^2)$满足条件

$$\begin{array}{l} \sigma\otimes\tau(e_1\otimes 1) = \sigma(e_1)\otimes\tau(1) = a\otimes(x,y)^T = ax(1\otimes e_1)+ay(1\otimes e_2) \\ \sigma\otimes\tau(e_2\otimes 1) = \sigma(e_2)\otimes\tau(1) = b\otimes(x,y)^T = bx(1\otimes e_1)+by(1\otimes e_2) \end{array}.$$

因此$\sigma\otimes\tau$在$\{e_1\otimes 1, e_2\otimes 1\}$-$\{1\otimes e_1, 1\otimes e_2\}$-基下的矩阵为$\begin{pmatrix} ax & bx \\ ay & by \end{pmatrix}$. 该矩阵正是矩阵$A = (a\ \ b)$与$B = (x\ \ y)^T$的张量积.

例 2.6.7. 设A, B均为2阶矩阵. 设σ是$\mathbb{F}^{2\times 2}$上的一个线性变换, 其中

$$\sigma:\ X \mapsto \sigma(X) = AXB^T.$$

则直接计算可知, σ关于(按行顺序)标准基E_{11}, E_{12}, E_{21}, E_{22}的矩阵是$A\otimes B$, 关于(按列顺序)标准基E_{11}, E_{21}, E_{12}, E_{22}的矩阵是$B\otimes A$.

一般地, 有下述

定理 2.6.6. 设$\sigma_i \in \mathrm{Hom}(U_i, V_i)$关于$\alpha_i$-$\alpha_i'$基的矩阵分别为$A_i$, $i = 1, 2$. 则$\sigma_1\otimes\sigma_2$关于$\alpha_1\otimes\alpha_2$-$\alpha_1'\otimes\alpha_2'$基的矩阵为$A_1\otimes A_2$.

由公式(2.6.8)定义的线性空间的张量积容易接受但依赖于特定的基而缺少“自然性”(比如比较难以看出不同的基定义的张量积空间之间的关系, 特别是无限维空间), 从而不能反映张量积的本质, 下面我们利用现代数学的深刻概念**“泛性质”**给出张量积的本质定义.

定义 2.6.1. 设U, V, W 是三个向量空间, $\sigma: U\times V \longrightarrow W$称为**双线性映射**是指$\sigma$关于每个分量都是线性的, 即$\forall \alpha, \beta\in U, \gamma, \delta\in V, x, y, a, b\in\mathbb{F}$, 有

$$\sigma(x\alpha+y\beta,\gamma) = x\sigma(\alpha,\gamma)+y\sigma(\beta,\gamma);\ \sigma(\alpha, a\gamma+b\delta) = a\sigma(\alpha,\gamma)+b\sigma(\alpha,\delta).$$

例 2.6.8. 取定$a, b\in\mathbb{R}$. 则从$\mathbb{R}\times\mathbb{R}$到$\mathbb{R}$的映射$\sigma(x, y) = axy$是双线性映射; 从$\mathbb{R}^2\times\mathbb{R}$到$\mathbb{R}$的映射$\tau(\begin{pmatrix} x \\ y \end{pmatrix}, z) = axz + ayz$是双线性映射, 但若$a\neq b$, 则$\tau(\begin{pmatrix} x \\ y \end{pmatrix}, z) = axz + byz$不是双线性映射.

定义 2.6.2. 设U_1, U_2, V 是三个线性空间, $t: U_1\times U_2 \longrightarrow V$ 是双线性映射. 如果对任意线性空间W与任意双线性映射$f: U_1\times U_2\longrightarrow W$, 均存在唯一的线性变换$u: V\longrightarrow W$使得$ut = f$, 称$(V, t)$是$U_1, U_2$的张量积.

定义2.6.2称为张量积的**泛性质**定义, 其中的双线性映射t称为**张量积映射**, 线性变换u称为**泛映射**, 其唯一性正是张量积的本质.

读者可以自行证明该定义与前面利用取定基的张量积的定义是等价的, 但使用泛性质的定义2.6.2不但简洁明了, 显然更加深刻、有力, 反映了当代数学高度概括、高度抽象的特征. 数学最年轻、最富活力的分支范畴论有**“数学就是伴泛”**的名言, 其中的伴是指伴随, 而泛正是泛性质, 泛性质的重要性由此可见一斑.

2.6.4 实线性空间与复线性空间的转化

实数域与复数域上的线性空间可以相互转化. 首先, 设V是复数域$\mathbb{C}$上的线性空间, 则V在原有的加法和数乘运算下, 自动成为实数域$\mathbb{R}$上的线性空间, 差别是维数有变化. 比如, $\mathbb{C}$本身作为复线性空间是1维的, 但作为实线性空间却是2维的, 因为1与i在$\mathbb{C}$上线性相关, 但在$\mathbb{R}$上却是线性无关的.

定理 2.6.7. 设V是复数域$\mathbb{C}$上的n维线性空间, 则V作为实数域上的线性空间的维数是$2n$.

证 设$\alpha_1, ..., \alpha_n$是V在$\mathbb{C}$上的一组基, 断言

$$\alpha_1, ..., \alpha_n, i\alpha_1, ..., i\alpha_n$$

是V在$\mathbb{R}$上的一组基. 设$x_1,...,x_n,y_1,...,y_n$是实数使得

$$x_1\alpha_1+...+x_n\alpha_n+y_1i\alpha_1+...+y_ni\alpha_n=0$$

则

$$(x_1+iy_1)\alpha_1+...+(x_n+iy_n)\alpha_n=0.$$

故

$$x_1+iy_1=...=x_n+iy_n=0$$

因此$\alpha_1,...,\alpha_n,i\alpha_1,...,i\alpha_n$在$\mathbb{R}$上线性无关.

对$\forall v\in V$, 存在复数$z_1=x_1+iy_1,...,z_n=x_n+iy_n$, 其中$x_i,y_i$均为实数, 使得

$$v=z_1\alpha_1+...+z_n=x_1\alpha_1+...+x_n\alpha_n+y_1i\alpha_1+...+y_ni\alpha_n$$

故$\alpha_1,...,\alpha_n,i\alpha_1,...,i\alpha_n$是$V$在$\mathbb{R}$上的一组基. □

既然复线性空间自动是实线性空间, 复线性空间的线性变换自然也自动是实线性空间的线性变换, 再由定理2.6.7即可知

推论 2.6.2. 设V,W是复数域$\mathbb{C}$上的有限维维线性空间, 则

$$\dim_{\mathbb{R}}\mathrm{Hom}_{\mathbb{R}}(V,W)=4\dim_{\mathbb{C}}\mathrm{Hom}_{\mathbb{C}}(\mathrm{V},\mathrm{W})\tag{2.6.13}$$

例 2.6.9. $\dim_{\mathbb{C}}\mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^n,\mathbb{C}^m)=\dim_{\mathbb{R}}\mathbb{C}^{m\times n}=mn$; $\dim_{\mathbb{R}}\mathrm{Hom}_{\mathbb{R}}(\mathbb{C}^n,\mathbb{C}^m)=4mn.$

再来讨论如何将实线性空间转化为复线性空间, 该过程称为“复化”(complexification).

设V是实数域$\mathbb{R}$上的线性空间, 则复数的数乘运算对V没有意义, 因此V不是复数域$\mathbb{C}$上的线性空间. 但复数域$\mathbb{C}$是具有自然的实线性空间结构, 因此$\mathbb{C}\otimes V$是维数为$2\dim V$的实线性空间, 该空间具有“自然的”复线性空间结构. 首先, 线性空间的加群结构与数乘无关, 因此只需考察$\mathbb{C}\otimes V$的数乘“$\bullet$”结构. $\forall c\in\mathbb{C},v\in\mathbb{C}\otimes V$, 由张量积空间的结构知, 存在$c_i\in\mathbb{C},v_i\in V(1\le i\le s)$使得$v=\sum\limits_{i=1}^{s}c_i\otimes v_i$. 因此$c\bullet v$可以“自然地”定义为

$$c\bullet v=c\bullet\sum_{i=1}^{s}c_i\otimes v_i=\sum_{i=1}^{s}(cc_i)\otimes v_i\tag{2.6.14}$$

上式显然是良定义的, 由此得到实线性空间V的“复化空间”$\mathbb{C}\otimes V$, 作为复空间, $\mathbb{C}\otimes V$的维数自然还是$\dim_{\mathbb{R}}V$.

第七节 商空间

直和是线性空间的加法, 张量积是线性空间的乘法, 本节将定义线性空间的“减法”.

2.7.1 等价关系

定义 2.7.1. 设S是非空集合. 称S中任意两个元素间的一个关系“$\sim$”为**等价关系**是指关系“$\sim$”满足下面三个条件:

自反性, 即$a\sim a,\forall a\in S$.

对称性, 即若$a\sim b$, 则$b\sim a,\forall a,b\in S$.

传递性, 即$a\sim b,b\sim c$, 则$a\sim c,\forall a,b,c\in S$.

如果$\sim$是等价关系, 而$a \sim b$, 则称a与b**等价**. 所有等价元素构成的子集称为**等价类**.

比如“$=$”是等价关系, 但“$\leq$”与“$\geq$”均非等价关系; 矩阵的相抵是$\mathbb{F}^{m\times n}$上的等价关系, 相似与相合均为$\mathbb{F}^{n\times n}$上的等价关系; 线性空间的同构也是全体线性空间构成的集合上的等价关系.

定义 2.7.2. 设$A_i, i \in I$是非空集合S的一族子集. 如果

$$A_i \bigcap A_j = \emptyset, \forall i \neq j; \quad S = \bigcup_{i\in I} A_i,$$

则称$A_i, i \in I$是S的一个**划分**.

等价关系与划分是一回事, 因为非常容易证明下面的定理.

定理 2.7.1. 设$S \neq \emptyset$, $A_i, i \in I$是S的一个划分. 定义$a \sim b \iff \exists j \in I, a, b \in S_j$. 则$\sim$是$S$上的等价关系. 反之, 任意给定$S$上的等价关系, 则所有等价类构成$S$的一个划分.

划分或定义等价关系的目的是发现新运算或新结构, 基本方法就是将“等价的元素”看成是相同的. 除法或商运算都是定义在等价类上的, 比如$\dfrac{1}{2}$就是等价类$\{\dfrac{1}{2}, \dfrac{2}{4}, \dfrac{-3}{-6}, ...\}$. 又如, 在整数环$\mathbb{Z}$中, 记全体偶数的集合为$2\mathbb{Z}$. 如果$a - b \in 2\mathbb{Z}$, 则称$a$与$b$等价. 因此所有偶数等价, 每个偶数$n$所在的等价类为$n + 2\mathbb{Z} = 0 + 2\mathbb{Z} = \bar{0}$; 所有奇数等价, 每个奇数$m$的等价类为$m + 2\mathbb{Z} = 1 + 2\mathbb{Z} = \bar{1}$. 因此“商”$\mathbb{Z}/2\mathbb{Z}$实际上是一个2元集合$\{\bar{0}, \bar{1}\}$. 可以自然地定义该集合上的加法运算“$+$”与乘法运算“$\bullet$”(即由$\mathbb{Z}$的运算诱导)如下:

$$\begin{aligned} &\bar{0} + \bar{0} = \bar{0}, \ \ \bar{0} + \bar{1} = \bar{1} + \bar{0} = \bar{1} \\ &\bar{0} \bullet \bar{0} = \bar{0} \bullet \bar{1} = \bar{1} \bullet \bar{0} = \bar{0}, \ \ \bar{1} \bullet \bar{1} = \bar{1} \end{aligned} \tag{2.7.1}$$

容易看出, 由公式(2.7.1)定义的$\mathbb{Z}/2\mathbb{Z}$上的加法与乘法运算就是熟知的“偶数+偶数=偶数, 奇数$\bullet$奇数=奇数”等的简洁表达. 在这样的加法与乘法下, 商$\mathbb{Z}/2\mathbb{Z}$构成了应用中最为重要的一个有限域即二元域$\mathbb{Z}_2$或$\mathbb{F}_2$. 一般, 将$\bar{0}$与$\bar{1}$分别简记为0与1, 即有$\mathbb{F}_2 = \{0, 1\}$.

上述做法可以推广到任意素数p, 即将$2\mathbb{Z}$换成$p\mathbb{Z}$, 即可得到有限域$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

2.7.2 商空间与张量积

设U是线性空间V的子空间. 对任意$\alpha, \beta \in V$, 定义$\alpha \sim \beta \iff \alpha - \beta \in U$. 易知$\sim$是$V$上的一个等价关系, α所在的等价类是V的子集$\{\beta \in V \mid \beta = \alpha + u, u \in U\}$, 一般称为$\alpha$关于$U$的**陪集**, 记为$\alpha + U$. 所有不同的陪集构成的集合记为$V/U$. 定义$V/U$上的加法和数乘为

$$(\alpha_1 + U) + (\alpha_2 + U) = (\alpha_1 + \alpha_2) + U, \ \ a \bullet (\alpha + U) = a\alpha + U. \tag{2.7.2}$$

一般将运算符号“$\bullet$”省略. 直接验证可知$(V/U, +, \bullet)$称为V关于U的**商空间**, 简称V/U是商空间. 常将商空间V/U中的向量$\alpha + U$记为$\bar{\alpha}$.

例 2.7.1. 考虑$V = \mathbb{R}^2$的子空间$U = x$-轴. 对任意向量$\alpha \in \mathbb{R}^2$, 陪集$\alpha + U$就是x-轴平移α所得到的平行于x-轴的一条直线, 因此所有关于x-轴的陪集就是全体平行于x-轴的直线, 它们就是V/U的全体元素(向量). 所以商空间V/U的几何意义就是将平行于x-轴的直线看成一个点, 或者说将它们压缩成y-轴上的点, 于是商空间V/U就是y-轴, 即$V/U \cong y$-轴. 请读者自行建立该同构映射.

类似于例 2.7.1, 线性空间V的子空间U的陪集$\alpha+U$实际上是将U沿向量α“平移”所得, 故几何学中常将点, 直线, 平面等称为**线性流形**或**仿射子空间**. 因此, 线性空间的子空间实际上是过原点的线性流形.

由商空间的定义直接可以验证下面的命题.

命题 2.7.1. (1) 对任意$u\in U$, 有$u+U=0+U=U$, 即U是商空间V/U的 0 向量.
(2) $v+U=\bar{0} \iff v\in U$.
(3) $\alpha+U=\beta+U \iff \alpha-\beta\in U$.

下面的定理说明线性空间的商运算本质上是直和运算的逆运算, 故商空间的“正确”称呼应该是“差空间”.

定理 2.7.2. 设U, W是V的两个子空间. 则
(1) $(U+W)/U\cong W/(U\bigcap W)$.
特别地, 若$V=U\oplus W$, 则$V/U\cong W$, 且若V是有限维的, 则$\dim V/U=\dim V-\dim U$.
(2) $(U+W)/(U\bigcap W)\cong U/(U\bigcap W)\oplus W/(U\bigcap W)$.
(3) 假设U_0, W_0分别是U与W的子空间, 则

$$\frac{U_0+U\bigcap W}{U_0+U\bigcap W_0}\cong\frac{U\bigcap W}{U\bigcap W_0+U_0\bigcap W}\cong\frac{U\bigcap W+W_0}{U_0\bigcap W+W_0} \tag{2.7.3}$$

证 (1) 对$\forall u\in U$, 记$u+(U\bigcap W)=u'$. 构造映射

$$\sigma: U/(U\bigcap W)\longrightarrow (U+W)/W$$

为$\sigma(u')=u+W$. 设$x'=y'\in U/(U\bigcap W)$, 则$x-y\in U\bigcap W$. 因此$x+W=y+W$. 故$\sigma(x')=\sigma(y')$, 所以σ是良定的映射.

其次, 容易验证σ是线性变换.

最后, 由于对$\forall w\in W$, $u+w+W=u+W$, 故σ是满射. 若$\sigma(x')=W$, 则$x+W=W$, 即$x\in W$. 但$x\in U$, 故$x\in U\bigcap W$. 所以σ是单射, 故是同构.

(2)与(3)的证明留作习题. □

注1. 构造商空间之间的映射时, 不可省略的一步是检查定义是否是“**良定的**”(well-defined), 即对同一个等价类的不同代表元, 映射的像是否是相同的.

注2. 定理 2.7.2的(1)与(2)常被称为商空间的“**平行四边形定理**”, 名字源于下图

$$
\begin{tikzcd}
 & U+W \arrow[dl] \arrow[dr] & \\
U \arrow[dr] & & W \arrow[dl] \\
 & U\bigcap W &
\end{tikzcd}
$$

图 2.7.1

注3. 定理2.7.2的(3)是群论中著名的**"蝴蝶引理"**, 请读者自行作图解释其命名.

由上面的证明可以知道商空间的基本结构, 即若$U+W$是直和, 则U的一组基与W的一组基的并集是$U+W=U\oplus W$的一组基, 故W的这组基的每个向量关于U的陪集构成商空间V/U的一组基. 从而有下述所谓(商空间的)**"同构基本定理"**.

推论 2.7.1. 设$\alpha_1,\cdots,\alpha_s$是U的一组基, $\alpha_1,\cdots,\alpha_s,\alpha_{s+1},\cdots,\alpha_n$是$V$的一组基. 则

$$\alpha_{s+1}+U,\cdots,\alpha_n+U$$

是商空间V/U的一组基. 特别地, 设$\sigma\in\mathrm{Hom}(U,V)$. 则$U/\mathrm{Ker}(\sigma)\cong\mathrm{Im}(\sigma)$.

定义 2.7.3. 设V是线性空间, $\sigma\in\mathrm{End}V$. 设U是σ的不变子空间. 定义V/U到自身的映射$\bar{\sigma}$如下:

$$\bar{\sigma}:\ \alpha+U\mapsto\sigma(\alpha)+U. \tag{2.7.4}$$

则$\bar{\sigma}$是线性变换, 称为σ的**诱导映射** 或**商映射**.

例 2.7.2. 设σ是$\mathbb{R}^3$的关于xoy-平面的投影变换. 则σ关于基e_3,e_1,e_2的矩阵为$(0)\oplus I_2$. 设U是z-轴(看成是$\mathbb{R}^3$的子空间), 则σ诱导的商空间V/U的线性变换为$\bar{\sigma}((x,y,z)^T+U)=(x,y,0)^T+U=(x,y,z)^T+U$, $\bar{\sigma}$关于基$\bar{e}_1,\bar{e}_2$的矩阵为I_2. σ关于基e_3,e_1+e_3,e_2+e_3的矩阵为$\begin{pmatrix}0&-1&-1\\0&1&0\\0&0&1\end{pmatrix}$. $\bar{\sigma}$关于基$\overline{e_1+e_3}$, $\overline{e_2+e_3}$的矩阵仍然为I_2(实际上$\bar{\sigma}$是恒等变换).

以下考察诱导映射的矩阵.

定理 2.7.3. 设$A\in M_s(\mathbb{F}),B,C\in M_{n-s}(\mathbb{F})$. 设$V$是有限维线性空间, $\sigma\in\mathrm{End}V$. 设σ在V的两组基$\alpha_1,\cdots,\alpha_s,\alpha_{s+1},\cdots,\alpha_n$ 与$\alpha_1,\cdots,\alpha_s,\beta_{s+1},\cdots,\beta_n$下的矩阵分别为$A\oplus B$与$A\oplus C$. 则$U=\mathrm{Span}\{\alpha_1,\cdots,\alpha_s\}$是$\sigma$的不变子空间, 且$\sigma$的诱导映射$\bar{\sigma}$关于商空间$V/U$的两组基$\bar{\alpha}_{s+1},\cdots,\bar{\alpha}_n$ 与$\bar{\beta}_{s+1},\cdots,\bar{\beta}_n$下的矩阵分别为$B$与$C$.

证 由σ在此两组基下的矩阵为分块对角矩阵可知, 不仅$U=\mathrm{Span}\{\alpha_1,\cdots,\alpha_s\}$是$\sigma$的不变子空间, $\mathrm{Span}\{\alpha_{s+1},\cdots,\alpha_n\}$与$\mathrm{Span}\{\beta_{s+1},\cdots,\beta_n\}$也均是$\sigma$的不变子空间. 由推论 2.7.1可知, $\bar{\alpha}_{s+1},\cdots,\bar{\alpha}_n$ 与$\bar{\beta}_{s+1},\cdots,\bar{\beta}_n$都是$V/U$的基. 由诱导映射$\bar{\sigma}$的定义可得

$$(\bar{\sigma}(\bar{\alpha}_{s+1}),\cdots,\bar{\sigma}(\bar{\alpha}_n))=(\bar{\alpha}_1,\cdots,\bar{\alpha}_s,\bar{\alpha}_{s+1},\cdots,\bar{\alpha}_n)\begin{pmatrix}0\\B\end{pmatrix}.$$

由于$\bar{\alpha}_1=0,\cdots,\bar{\alpha}_s=0$, 故有

$$(\bar{\sigma}(\bar{\alpha}_{s+1}),\cdots,\bar{\sigma}(\bar{\alpha}_n))=(\bar{\alpha}_{s+1},\cdots,\bar{\alpha}_n)B,$$

即$\bar{\sigma}$在基$\bar{\alpha}_{s+1},\cdots,\bar{\alpha}_n$下的矩阵为$B$. 同理知$\bar{\sigma}$在基$\bar{\beta}_{s+1},\cdots,\bar{\beta}_n$下的矩阵$C$. □

由于线性变换在不同基下的矩阵彼此相似, 故由定理 2.7.3可得下述

推论 2.7.2. 设A, B, C均是方阵, 则$A\oplus B$与$A\oplus C$相似 $\iff$ B与C相似.

现在我们可以利用商空间来重新定义线性空间的张量积了.

设U与V是两个线性空间, 将以集合$U\times V$为基的无限维线性空间记为$F(U\times V)$, 称为$U\times V$上的**自由线性空间**, 即

$$F(U\times V)=\{\sum_{i=0}^{n}x_i(u_i,v_i)\,|\,u_i\in U,\,v_i\in V,x_i\in\mathbb{F},n\in\mathbb{N}\} \tag{2.7.5}$$

记线性空间$F(U\times V)$的由下列四种向量生成的子空间为N:

$$(u_1+u_2,v)-(u_1,v)-(u_2,v),(u,v_1+v_2)-(u,v_1)-(u,v_2),x(u,v)-(xu,v),\quad x(u,v)-(u,xv) \tag{2.7.6}$$

定义 2.7.4. 称商空间$F(U\times V)/N$为U与V的张量积(空间).

利用商空间定义的张量积空间$F(U\times V)/N$不依赖于U与V的特定的基, 且子空间N突出了张量积的双线性性质—这是张量积的本质. 但利用商空间来定义张量积的困难也是显而易见的, 不仅涉及到无限维空间$F(U\times V)$, 而且子空间N往往也很难确定.

例 2.7.3. 考察最简单的情形, 即取$U=V=\mathbb{F}$. 固定U与V的各一组基α与β. 则定义 2.7.4中的子空间N由以下三种元素生成:

$$(x_1\alpha+x_2\alpha,y\beta)-(x_1\alpha,y\beta)-(x_2\alpha,y\beta)$$

$$(x\alpha,y_1\beta+y_2\beta)-(x\alpha,y_1\beta)-(x\alpha,y_2\beta)$$

$$z(x\alpha,y\beta)-(zx\alpha,y\beta),\quad z(x\alpha,y\beta)-(x\alpha,zy\beta)$$

因此在商空间$F(U\otimes V)/N$中, 任意向量$(x\alpha,y\beta)+N=xy(\alpha,\beta)+N$, 因此向量$(\alpha,\beta)+N$是商空间$F(U\otimes V)/N$的一组基. 从而商空间$F(U\otimes V)/N$是 1 维线性空间.

定理 2.7.4. 设U与V是两个有限维线性空间, 则$\dim_{\mathbb{F}}F(U\otimes V)/N=\dim_{\mathbb{F}}U\dim_{\mathbb{F}}V$. 特别地, 商空间$F(U\otimes V)/N$与张量积$U\otimes V$同构.

证 取定U与V的各一组基为$\mathcal{B}_U=\{\alpha_1,...,\alpha_m\}$与$\mathcal{B}_W=\{\beta_1,...,\beta_n\}$. 断言

$$\mathcal{B}=(\alpha_i,\beta_j)+N, 1\le i\le m, 1\le j\le n$$

构成商空间$F(U\otimes V)/N$的一组基.

首先, 由公式(2.7.5)知, 自由向量空间$F(U\otimes V)$中任意元素均为有限和$\alpha=\sum\limits_{i=0}^{n}x_i(u_i,v_i), u_i\in U, v_i\in V$, 但每个$u_i,v_j$分别是$\mathcal{B}_U$与$\mathcal{B}_W$的线性组合, 故$\alpha$在商空间中的像$\alpha+N$是$\mathcal{B}$的线性组合.

其次, 设

$$\sum_{i=0}^{m}\sum_{j=0}^{n}x_{ij}[(\alpha_i,\beta_j)+N]=0$$

则

$$\beta=\sum_{i=0}^{m}\sum_{j=0}^{n}x_{ij}(\alpha_i,\beta_j)\in N \tag{2.7.7}$$

检查公式(2.7.6)中N的生成元素, 第一类生成元的第二个分量相同, 第二类生成元的第一个分量相同, 第三、四类生成元的两个分量分别线性相关, 因此(2.7.7)中的元素$\beta\in N \iff x_{ij}=0,\forall i,j$. 从而$\mathcal{B}$是线性无关的.

因此, $\mathcal{B}$是商空间$F(U\otimes V)/N$的一组基, 故其维数$\dim_{\mathbb{F}}F(U\otimes V)/N=\dim_{\mathbb{F}}U\dim_{\mathbb{F}}V$, 从而与张量积$U\otimes V$同构. □

思考题

1. 设$U_i,V_j,1\le i\le n,1\le j\le m$是线性空间. 描述$\mathrm{Hom}(\sum\limits_{i=1}^{n}\oplus U_i,\sum\limits_{j=1}^{m}\oplus V_i)$中的元素的结构, 并以此给出分块矩阵的一个几何解释.

2. $\mathbb{Q}(\sqrt{2})$是有理数域上的2维线性空间. 它与$\mathbb{Q}$及自身的张量积(空间)分别是什么?

3. 复数域$\mathbb{C}$是实数域$\mathbb{R}$上的2维线性空间. 商空间$\mathbb{C}/\mathbb{R}$是什么?

4. 如果V是内积空间, U是V的子空间, 商空间V/U是否有自然的内积?

5. 实多项式空间$\mathbb{R}[x]$与复数域$\mathbb{C}$均是$\mathbb{R}$上的线性空间, 它们的张量积是什么?

6. 设V是线性空间, $\sigma\in\mathrm{End}V$是幂零(幂等, 同构, 等)变换. 设U是σ的不变子空间, 设$\bar{\sigma}$是由σ诱导的V/U上的商变换, 问$\bar{\sigma}$是否也是幂零(幂等, 同构等)变换?

第八节* 应用

2.8.1 广义相对论

以下我们介绍正交补在广义相对论中的应用.

对于实线性空间V, 如果在其内积的定义中去掉正定性而代之以

非退化性：若$(\alpha,v)=0,\forall v\in V$, 则$\alpha=0$.

则称V是一个**Minkowski**[39] **空间**, 而该内积称为**Minkowski 内积**.

内积空间当然是Minkowski空间, 因为内积的正定性蕴涵非退化性. 由于没有正定性, Minkowski空间中的向量与自身的内积(称为该向量的范数)可以为负数. 广义相对论使用的

[39] Hermann Minkowski(1864-1909), 著名俄裔德国数学家, 创造数的几何(geometry of numbers), 为相对论奠定了数学基础, 是Albert Einstein (爱因斯坦)的老师.

是 4-维 Minkowski 空间V, 其一组正交基记为e_0, e_1, e_2, e_3, 但

$$-(e_0)^2 = (e_1)^2 = (e_2)^2 = (e_3)^2 = 1.$$

任意两个向量$\alpha = (x_0, x_1, x_2, x_3)^T$与$\beta = (y_0, y_1, y_2, y_3)^T$的内积为

$$(\alpha, \beta) = -x_0y_0 + x_1y_1 + x_2y_2 + x_3y_3,$$

从而向量$\alpha = (x_0, x_1, x_2, x_3)^T$的范数为

$$\alpha^2 = -x_0^2 + x_1^2 + x_2^2 + x_3^2.$$

根据其范数的符号, 向量被分成三类:

(1) 类空的(spacelike): $\alpha^2 > 0$.

(2) 类时的(timelike): $\alpha^2 < 0$.

(3) 类光的(lightlike): $\alpha^2 = 0$.

所有类光向量的集合称为(一个事件的)**光锥**(light cone).

类似于向量, V的子空间W也被分为上述三类:

(a) 类空的(spacelike): $\alpha^2 > 0, \forall \alpha \in W$.

(b) 类光的(lightlike): $\alpha^2 \geq 0, \forall \alpha \in W$且存在$0 \neq \alpha \in W$使得$\alpha^2 = 0$.

(c) 类时的(timelike): 其它.

命题 2.8.1. (正交性与相对论) 设W是 Minkowski 4-空间V的子空间. 则
(1) W是类时的 $\iff$ $W^\perp$是类空的, 反之亦然.
(2) W是类光的 $\iff$ $W \cap W^\perp \neq 0$ $\iff$ $W^\perp$是类光的.

下面是极有趣的一条:

命题 2.8.2. (平行=垂直) 两个类光向量正交 $\iff$ 它们平行(即成比例).

2.8.2 滤波与移动通信

一. 正交性与移动通信

例 2.8.1. 设$\mathbb{S} = \{\{x_k\} = (\cdots, x_{-2}, x_{-1}, x_0, x_1, x_2, \cdots) \mid x_i \in \mathbb{R}\}$表示双向无穷实数列的集合, 按分量自然定义$\mathbb{S}$中的向量加法和数乘, 则$V$构成一个无限维实线性空间. 工程上常将$\mathbb{S}$称为(离散时间)信号空间.

实践中常用的是信号空间$\mathbb{S}$的一个子空间, 即平方可和子空间

$$l_2 = \{\{x_k\} \in \mathbb{S} \mid \sum_{k=-\infty}^{+\infty} x_k^2 < +\infty\}.$$

信号空间$\mathbb{S}$的另几个常用的子空间是:
(1) 绝对收敛子空间

$$l_1 = \{\{x_k\} \in \mathbb{S} \mid x_k = 0, 若k \leq 0且 \sum_{k=1}^{+\infty} |x_k| < +\infty\};$$

(2) 有限非零序列子空间

$$\bigsqcup_{i=-\infty}^{\infty} \mathbb{R} = \{\{x_k\} \in \mathbb{S} \,|\, x_k \text{ 几乎处处为 } 0\,\} \,(\text{即仅有有限个} x_k \text{非} 0\,)$$

(3) 有界序列子空间

$$l_\infty = \{\{x_k\} \in \mathbb{S} \,|\, x_k = 0, \text{若} k \leq 0, \text{ 且存在整数} C, |x_k| < C, \forall k\}.$$

平方可和子空间l_2是一个无限维内积空间, 其中的内积是自然定义的(见习题 64), 即

$$(\{x_k\}, \{y_k\}\,) = \sum_{k=-\infty}^{+\infty} x_k y_k. \tag{2.8.1}$$

于是两个离散信号$\{x_k\}$与$\{y_k\}$正交 $\Longleftrightarrow \sum\limits_{k=-\infty}^{+\infty} x_k y_k = 0.$

在实际应用中, 所有用户共享整个频率信道, 但不同用户被分配至不同的时区, 并且不同用户的通信信号在时域上没有任何重叠, 即若以$\{x_k\}$与$\{y_k\}$分别表示两个不同用户的通信信号, 则$x_k y_k = 0, \forall k$ (此时由向量$\{x_k\}$与$\{y_k\}$生成的子空间不但相交为0, 而且还是正交的), 因此$\{x_k\}$与$\{y_k\}$正交. 这就是时分多址通信的工作原理. 移动通信中的**频分多址**与**码分多址**的工作原理完全类似. 所以正交性对移动通信起着至为关键的作用.

二. 差分方程与线性滤波器

在数字信号处理中, 线性滤波器可以用如下的n阶线性差分方程来表示(a_i为常数且$a_0 a_n \neq 0$)

$$y_k = a_n x_k + a_{n-1} x_{k+1} + \cdots + a_1 x_{k+n-1} + a_0 x_{k+n} \tag{2.8.2}$$

其中$\{x_k\}$是输入信号而$\{y_k\}$为输出信号. 如果$\{y_k\}$是零序列, 则称该差分方程为齐次的, 此时的解对应于被过滤掉的信号(并被换为零信号).

例 2.8.2. 对 2 阶滤波器$y_k = \frac{1}{2\sqrt{2}} x_k + \frac{1}{2} x_{k+1} + \frac{1}{2\sqrt{2}} x_{k+2}$ 输入信号(来自于连续信号$x = \cos\frac{\pi t}{4}$的整数抽样)

$$\{x_k\} = \{\cdots, x_0 = 1, \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, -1, -\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 1, -\frac{1}{\sqrt{2}}, 0, \cdots\}$$

可得输出信号(具体计算见习题 65)

$$\{y_k\} = \{\cdots, y_0 = \frac{1}{\sqrt{2}}, y_1 = 1, \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, -1, -\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 1, -\frac{1}{\sqrt{2}}, 0, \cdots\}.$$

如果再输入一个新的更高的频率信号$\{u_k\}$(来自于连续信号$x = \cos\frac{3\pi t}{4}$的整数抽样), 则输入信号为(具体计算留作习题)

$$\{u_k\} = \{\cdots, u_0 = 1, -\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, -1, \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, 1, -\frac{1}{\sqrt{2}}, 0, \cdots\}.$$

此时滤波器产生的输出序列是零序列, 因此该滤波器将高频信号过滤掉了(故称为低通滤波器).

命题 2.8.3. n阶齐次线性差分方程的解集是一个n维线性空间. 因此, 被线性滤波器过滤掉的所有信号构成一个线性空间.

研究n阶齐次线性差分方程的一种方法是降阶法, 即化为等价的一阶差分方程, 而每个一阶差分方程可写成如下形式:

$$y_{k+1} = Ay_k, \forall k \tag{2.8.3}$$

其中$y_k \in \mathbb{R}^n$, $A \in M_n(\mathbb{R})$. 比如差分方程(注意$a_0 = 1$)

$$a_n x_k + a_{n-1}x_{k+1} + \cdots + a_1 x_{k+n-1} + x_{k+n} = 0, \forall k$$

可以化为$y_{k+1} = Ay_k\ (\forall k)$, 其中$y_k = (x_k, x_{k+1}, \cdots, x_{k+n-1})^T$, 而

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}.$$

注意, 矩阵A正是多项式$f(x) = a_n + a_{n-1}x + \cdots + a_1 x^{n-1} + x^n$的友矩阵.

如果设$x_k = 0, k < 0$, 则由迭代法可以将(2.8.3)化为

$$y_k = A^k y_0 \tag{2.8.4}$$

其中$y_0 = (x_0, x_1, \cdots, x_{n-1})^T$. 于是矩阵$A$的高次幂的计算成为解决问题的关键. 我们将在下一章继续讨论该问题.

2.8.3 度量与任意矩阵的行列式

设$\mathfrak{S}$是欧氏空间$V = \mathbb{R}^n (n = 2, 3)$中的图形, $A \in \mathbb{R}^{n\times n}$, 问$A(\mathfrak{S}) = \{y \in V \mid y = \sigma(x), \exists x \in \mathfrak{S}\}$与$\mathfrak{S}$的面积或体积的关系.

我们知道, 2阶与3阶行列式的绝对值的几何意义分别是(平面或立体)图形的面积与体积, 从线性变换的观点看, 平面上任何矩形都是单位正方形$\mathfrak{D}_2 = \{(x, y) \in \mathbb{R}^2 | 0 \le x, y \le 1\}$在一个两列正交的矩阵$A$下的像.

例2.8.3. 设$\mathfrak{S}$是以$(0,0), (a,b), (-2b, 2a), (a-2b, 2a+b)$为顶点的矩形, 取$A = \begin{pmatrix} a & -2b \\ b & 2a \end{pmatrix}$, 则$\mathfrak{S} = A\mathfrak{D}_2$. 因此$\mathfrak{S}$的面积$2(a^2+b^2)$恰好是$A$的行列式的绝对值.

类似地, 空间中任何长方体的体积都是一个三列相互正交的矩阵的行列式. 因此, 对平面图形或空间立体进行平行于坐标轴或坐标面的分割即可证明下面的定理.

定理 2.8.1. 记欧氏空间$V = \mathbb{R}^n (n = 2, 3)$中图形$\mathfrak{S}$的面积或体积为$vol(S)$, $A \in \mathbb{R}^{n\times n}$的行列式为$d$. 则$vol(A(\mathfrak{S})) = |d| vol(\mathfrak{S})$.

例 2.8.4. 不使用微积分, 求椭圆$S = \{(x, y) \mid \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \le 1\}$的面积.

这里的本质是任何椭圆一定是单位圆在某线性变换下的像. 实际上本题中的椭圆是单位圆盘在$\alpha \mapsto A\alpha$下的像, 其中$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$. 因此由定理2.8.1知$S$的面积为$|A|\pi = \pi ab$.

如果归纳地定义高维图形的度量, 则定理2.8.1仍然成立.

例 2.8.5. 行列式理论的关键是乘法公式$|AB| = |A||B|$, 其几何意义如下. 设$A, B \in M_n(\mathbb{F})$, 有界图形$G \subset \mathbb{F}^n$的体积为$V \neq 0$. 考察$\mathbb{F}^n$到自身的两个线性变换$\sigma : x \mapsto Ax$与$\tau : x \mapsto Bx$及其复合$\sigma\tau : x \mapsto (AB)x$. 由定理2.8.1知, $\tau(G)$的体积为$|B|V$, $\sigma(\tau(G))$的体积为$|A||B|V$, 而$\sigma\tau(G)$的体积为$|AB|V$. 由于$\sigma(\tau(G)) = \sigma\tau(G)$, 故$|AB|V = |A||B|V$. 所以

$$|AB| = |A||B|.$$

习 题 二

1. 设E是椭圆曲线$y^2 = x^3 + bx + c, a^3 + 27b^2 \neq 0$. 对$E$上任意两点$P, Q$, 定义$P\spadesuit Q$为过点$P, Q$的直线$L$(如果$P = Q$, 则$L$是切线)与$E$的交点关于$x$轴的对称点.

(1) 不考虑L与E第三个交点的存在性, 证明上述运算"$\spadesuit$"满足结合律与交换律, 故是加法.

(2) 查阅相关资料, 证明如果将无穷远点定义为零元, 则$(E, \spadesuit)$构成一个阿贝尔群. 此时$P \in E$的负元是什么?

2. 设$V = \{$ 所有正实数 $\}, \mathbb{F} = \mathbb{R}$是实数域. 定义$V$中的加法运算为$x \oplus y = xy$ (即通常的实数乘法); 定义V中元素与$\mathbb{F}$中数的数乘运算为$k \bullet x = x^k$(通常的幂运算). 证明$(V, \oplus, \bullet)$是实线性空间.

3. 设$V = \{(x, y) \in \mathbb{R}^2 \mid y = e^x, \forall x \in \mathbb{R}\}$. 定义$V$上的加法"$\heartsuit$"与数乘"$\spadesuit$"分别为:

$$(x, y)\heartsuit(x', y') = (x + x', yy'); a\spadesuit(x, y) = (ax, y^a).$$

证明$(V, \heartsuit, \spadesuit)$构成$\mathbb{R}$上的1维线性空间.

4. 设V是向量空间, 证明:

(1) 数0与任何向量的数乘等于0向量.

(2) 任何数与0向量的数乘等于0向量.

(3) 数"-1"与向量$\alpha \in V$的数乘等于$-\alpha$.

(4) 数"-1"与向量$\alpha \in V$的负向量"$-\alpha$"的数乘等于α. (此即"负负得正".)

5. (1) 证明$\mathbb{F}$上的所有多重交错线性函数构成一个线性空间, 并求其一组基.

(2) 设$i = \sqrt{-1}, \mathbb{F} = \mathbb{Q}(i)$是虚二次扩域. 设加群$V = (\mathbb{Q}(i), +)$的加法与$\mathbb{F}$相同. 试定义$V = (\mathbb{Q}(i), +)$的一个与$\mathbb{F}$中的乘法不同的数乘, 使得$V$是$\mathbb{F}$-线性空间.

6. 设A是任意非空集合, A到$\mathbb{F}$的所有函数的集合记为$\mathbb{F}^A$.

(1) 证明$\mathbb{F}^A$按照普通意义下的函数加法和数乘函数构成一个线性空间.

(2) 求$\mathbb{F}^A$的一组基.

7. 设$\alpha_1, \alpha_2, \cdots, \alpha_m$与$\beta_1, \beta_2, \cdots, \beta_n$是两个等价向量组, 且$\beta_1, \beta_2, \cdots, \beta_n$线性无关. 证明: $m \geq n$, 且等号成立当且仅当$\alpha_1, \alpha_2, \cdots, \alpha_m$线性无关.

8. 设V是有限维线性空间, $\alpha_1, ..., \alpha_n$是V的一组基. 设$\sigma : V \longrightarrow \mathbb{F}^n$是将$\forall v \in V$对于到其在$\alpha_1, ..., \alpha_n$下的坐标的映射. 证明:

(L1) σ是单射, 即对任意$\alpha, \beta \in V, \sigma(\alpha) = \sigma(\beta)$, 则$\alpha = \beta$.

(L2) σ是满射, 即对任意$y \in \mathbb{F}^n, \exists\beta \in V$使得$\sigma(\beta) = y$.

(L3) σ是线性的, 即对任意$\alpha, \beta \in V, a, b \in \mathbb{F}$, 有$\sigma(a\alpha + b\beta) = a\sigma(\alpha) + b\sigma(\beta)$.

9. 证明过渡矩阵必是可逆矩阵.

10. 设二次曲线C的方程为$3x^2 + 4xy + 3y^2 = 1$.

(1) 不计算, 判断C是何种二次曲线.

(2) 求C在以$x = y, x = -y$为坐标轴的新坐标系下的方程, 并再次判断C是何种二次曲线.

11. 设U是线性空间V的一个非空子集. 证明U是子空间 $\Longleftrightarrow$ 对任意$\lambda \in \mathbb{F}$, $\alpha, \beta \in U$, 有$\alpha + \beta \in U$与$\lambda\alpha \in U$.

12. 举例说明两个子空间的并集未必是子空间. 请给出两个子空间的并集是子空间的充分必要条件.

13. 设$U_1, U_2, \cdots, U_s$是线性空间V的子空间.

(1) 证明集合
$$\{\alpha_1+\alpha_2+\cdots+\alpha_s \mid \alpha_j \in U_j,\, 1\le j\le s\}$$
也是子空间. 该子空间称为$U_1, U_2, \cdots, U_s$的和, 记为$U_1+U_2+\cdots+U_s$.

(2) 可否将上述定义推广到无限个子空间?

14. 证明或否定：线性空间V的子空间格是分配格, 即对V的子空间X, Y, Z有$X\bigcap(Y+Z) = (X\bigcap Y)+(X\bigcap Z)$.

15. (1) 证明维数公式：
$$\dim(U+W) = (\dim U+\dim W) - \dim(U\cap W).$$

(2) 尝试将维数公式推广至3个子空间并证明或证伪之.

(3) 设V是有限维线性空间. 证明下面的公式并解释之:
$$\dim V = \max\{m \mid 0 = V_0 \subset V_1 \subset \cdots \subset V_{m-1} \subset V_m = V,\ V_i\text{ 是 }V_{i+1}\text{ 的真子空间}\}.$$

16. 试用子空间的直和理论证明并解释下述命题：

(1) 每个方阵都可以唯一地分解成一个对称矩阵与一个反对称矩阵的和.

(2) 每个方阵都可以唯一地分解成一个纯量矩阵与一个迹为0的矩阵的和.

(3) 每个多项式都可以唯一地分解成一个偶多项式与一个奇多项式的和.

(4) 每个实函数都可以唯一地分解成一个偶函数与一个奇函数的和.

(5) 解释数域$\mathbb{F}$上的一元多项式的欧几里德带余除法: 设$f(x), g(x)$是任意两个多项式, 其中$g(x)\neq 0$, 则存在唯一一对多项式$q(x)$与$r(x)$使得
$$f(x) = g(x)q(x)+r(x)$$
其中$r(x)=0$或$\partial\, r(x) < \partial g(x)$.

(6) 试给出一个类似的新命题.

17. 证明定理 2.2.4(多子空间直和的判定).

18. (1) 设$A\in\mathbb{F}^{m\times n}, B\in\mathbb{F}^{p\times q}$. 证明：

$N(A\oplus B) = N(\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix})\oplus N(\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}) = N(\begin{pmatrix} A & 0 \end{pmatrix})\oplus N(\begin{pmatrix} 0 & B \end{pmatrix})$.

(2) 设$C\in\mathbb{F}^{m\times p}, D\in\mathbb{F}^{n\times p}$. 证明$N(\begin{pmatrix} C \\ D \end{pmatrix}) = N(C)\bigcap N(D)$.

(3) 设$E\in\mathbb{F}^{m\times p}, F\in\mathbb{F}^{m\times q}$. 证明$R(\begin{pmatrix} E & F \end{pmatrix}) = R(E)+R(F)$.

19. 设$A\in\mathbb{F}^{n\times n}$是幂零指数为$e$的幂零矩阵. 对任意正整数$k, \forall\beta\in\mathbb{F}^n$, 研究Krylov子空间$\mathrm{Span}\{\beta, A\beta, A^2\beta, ..., A^{k-1}\beta\}$的维数.

20. 设V是线性空间, $W_1, W_2, \cdots, W_s$是V的真子空间. 证明$W_1\cup W_2\cup\cdots\cup W_s\neq V$.(提示: 利用 Vandermonde 行列式或归纳法.)

21. 设V是所有次数小于n的实系数多项式组成的实线性空间, $U=\{f(x)\in V \mid f(1)=0\}$. 证明$U$是$V$的子空间, 并求$V$的一个补空间.

22. 设$U = span\{(1,2,3,6)^T, (4,-1,3,6)^T, (5,1,6,12)^T\}$, $W = span\{(1,-1,1,1)^T, (2,-1,4,5)^T\}$是$\mathbb{R}^4$的两个子空间.

(1) 求$U\cap W$的基.

(2) 扩充$U\cap W$的基, 使其成为U的基.

(3) 扩充$U\cap W$的基, 使其成为W的基.

(4) 求$U+W$的基.

23. 设A, B分别是$n\times m$, $m\times p$矩阵, V是齐次线性方程组$xAB=0$的解空间. 证明$U=\{y=xA \mid x\in V\}$是$\mathbb{F}^m$的子空间, 并求U的维数.

24. 证明**替换定理**：设$X=\{x_1,...,x_n\}$与$Y=\{y_1,...,y_n\}$是V的两组基. 则对X中任意$k\leq n$元子集X_1, 均存在Y中的k元子集Y_1, 使得$(X-X_1)\bigcup Y_1$也是V的一组基.

25. 设$\sigma\in \mathrm{Hom}(U,V)$. 证明$\mathrm{Ker}\sigma$与$\mathrm{Im}\sigma$分别是$U$与$V$的子空间.

26. 设U,V是线性空间, 映射$\sigma:U\longrightarrow V$满足条件$\sigma(\alpha+b)=\sigma(a)+\sigma(\beta),\sigma(a\alpha)=a^r\sigma(\alpha)$. 证明$r=1$.

27. 设U,V是$\mathbb{F}$上的线性空间, $\sigma\in \mathrm{Hom}(U,V)$. 证明：$\sigma$是单的 $\iff$ $\mathrm{Ker}(\sigma)=0$. 28. 设$\sigma\in \mathrm{Hom}(U,V)$. 证明, 如果$\sigma$的逆变换存在则必定唯一, 且$(\sigma^{-1})^{-1}=\sigma$.

29. 计算2维实线性空间$\mathbb{C}$的所有自同构.

30. 设$\sigma\in \mathrm{End}V$在V的某组基下的矩阵为A. 证明或否定: 存在$\tau\in \mathrm{End}V$, $\tau\neq\sigma$, 使得τ在另一组基下的矩阵也是A.

31. 设V是数域$\mathbb{F}$上的n阶矩阵全体, σ是将V中任意元素的严格下三角部分变为0的映射. 判断σ是否为V的线性变换. 若是, 求其核与像; 并任选V的一组基, 求σ在该组基下的矩阵.

32. 证明有理数域$\mathbb{Q}$到自身的加性函数必是齐次函数, 但反之不对.

33. (1) 设f是定义在实数域上的加性函数. 证明: 如果f是连续的, 则它一定是齐次的, 从而是线性变换.
(2) 试将(1)中的结论推广到一般情形.

34. Aupetit[40]定理(1988年). 设$\sigma\in \mathrm{End}V$. 设$\forall\alpha\in V$, $\alpha,\sigma(\alpha),...,\sigma^k(\alpha)$均线性相关, 证明$I,\sigma,...,\sigma^k$线性相关.

35. 36. 分别求导数运算∂: $f(x)\mapsto f'(x)$在标准基$1, x, x^2, \cdots, x^{n-1}$ 与基$1, (x-a), (x-a)^2, \cdots, (x-a)^{n-1}$下的矩阵. 问$\partial$的行列式与迹是多少? 解释之.

37. 设$\sigma,\tau\in \mathrm{End}V$.
(1) 设V的维数有限. 证明$tr(\sigma\tau-\tau\sigma)=0$.
(2) 设V的维数无限. 求适当的σ与τ使得$\sigma\tau-\tau\sigma=I$. (此即Heisenberg测不准原理.)

38. (1) 设$\sigma,\tau\in \mathrm{End}V$分别是线性空间$V$的同构变换和幂零变换, 证明$\sigma+\tau$是$V$的同构变换.
(2) 设A,D是可逆矩阵, B,C是幂零矩阵, 证明分块矩阵$\begin{pmatrix} A & B\\ C & D\end{pmatrix}$可逆.

39. 设V是数域$\mathbb{F}$上的n阶矩阵全体, A是V中一个固定元素, P是V中一个固定的可逆矩阵, σ是左乘A的映射, τ是左乘P逆右乘P的映射. 判断σ与τ是否为V的线性变换. 若是, 求其核与像. 并任选V的一组基, 计算σ与τ在该组基下的矩阵.

40. 设$\sigma,\tau\in \mathrm{End}V,\sigma\tau=\tau\sigma$.
(1) 证明σ与τ有公共的特征向量.
(2) 利用(1)证明：若矩阵A,B可换, 则A与B有公共的特征向量.

41. 证明迹函数$\mathrm{tr}:X\mapsto \mathrm{tr}(X)$是线性空间$M_n(\mathbb{F})$到$\mathbb{F}$的满足性质$\sigma(XY)=\sigma(YX)$以及$\sigma(I)=n$的唯一线性变换.

42. 验证: 若$(\alpha,\beta)_1$与$(\alpha,\beta)_2$是欧氏空间V的两个不同的内积, 则$(\alpha,\beta)=(\alpha,\beta)_1+(\alpha,\beta)_2$也是$V$的一个内积. 试创造一种新办法再构造$V$的一种内积.

43. 设V是n维内积空间, U是V的子空间. 令$W=\{\alpha\in V\,|\,(\alpha,\beta)=0,\forall\beta\in U\}$. 证明$W$是$V$的子空间且$V=U\oplus W$.

44. 设$V=\mathbb{R}[x]_n$, 其上的内积为

$$(f(x),g(x))=\int_0^1 f(x)g(x)\,\mathrm{d}\,x.$$

设$U=\{f(x)\in V\,|\,f(0)=0\}$.

[40]B. Aupetit, 加拿大数学家.

(1) 证明U是V的一个$n-1$维子空间, 并求U的一组基.
(2) 当$n=3$时, 求U的正交补$U^{\perp}$.

45. 证明命题2.5.3.

46. 试任意构造维数大于5的一个线性空间V以及V的一个线性映射σ, 使得σ的核的维数等于5. 进一步, 试将V改造成内积空间, 求$\mathrm{Im}\sigma$的正交补空间. 再构造一个线性变换τ, 使得$\mathrm{Ker}\tau=\mathrm{Im}\sigma$, $\mathrm{Im}\tau=\mathrm{Ker}\sigma$.

47. 设α_0是欧氏空间V中的单位向量, $\sigma(\alpha)=\alpha-2(\alpha,\alpha_0)\alpha_0$, $\alpha\in V$.
(1) 证明: σ是正交变换.
(2) 求子空间α_0与$\alpha_0^{\perp}$上的投影变换的表达式.

48. 证明: 欧氏空间V的线性变换σ是反对称变换(即$(\sigma(\alpha),\beta)=-(\alpha,\sigma(\beta))$) $\Longleftrightarrow$ σ在V的标准正交基下的矩阵是反对称矩阵.

49. 设σ是实平面$\mathbb{R}^2$上的线性变换, 其关于标准基的矩阵为

$$P=\begin{pmatrix} c & s \\ s & -c \end{pmatrix}$$

其中$c^2+s^2=1$. 证明σ是反射变换, 并计算其对称轴.

50. 证明例2.5.15关于三维正交矩阵的结论. (提示: 利用三维正交矩阵有1个或3个实特征值并结合二维正交矩阵的分类.)

51. 设$\sigma\in\mathrm{End}\mathbb{R}^2$. 记单位正方形$S=\{(x,y):0\leq x,y\leq 1\}$在$\sigma$下的图形为$G=\sigma(S)=\{\sigma(x,y):(x,y)\in S\}$. 回答下列问题:
(1) 列出G所有可能的形状.
(2) 如果G仍为正方形, σ应满足什么条件?
(3) 如果σ可逆, 则G是什么形状?

52. (1) 设$\sigma\in\mathrm{End}\mathbb{R}^2$. 设$C$是一个二次曲线(即抛物线, 椭圆或双曲线). 计算$\sigma(C)$所有可能的形状(可设C的方程均为标准方程).
(2) 设P是一个平面n次代数曲线(即C的方程是n次多项式), 计算$\sigma(Q)$所有可能的形状.
(3) 分别对指数函数, 对数函数, 三角函数研究其曲线在σ下的图像;
(4) 设$\sigma\in\mathrm{End}\mathbb{R}^3$. 设$Q$是一个二次曲面. 计算$\sigma(Q)$所有可能的形状(可设$Q$的方程均为标准方程).

53. 证明 Householder 变换H_v是关于超平面$v^{\perp}$的反射, 从而是正交变换. 试画出三维 Householder 变换的示意图.

54. 证明 Givens 旋转矩阵G是正交矩阵. 对任意向量$x=(x_1,x_2,\cdots,x_n)^T$, 计算Gx的各个分量. 设x是单位向量, 讨论如何重复使用若干 Givens 旋转矩阵将x变为标准向量e_1.

55. 证明: 函数$f(x)$的 Fourier 级数中的系数a_n, b_n $(n>0)$恰好是$f(x)$与诸基向量$\cos nx$, $\sin nx$的内积.

56. 证明Djokovic[41]定理(1971年)：设$V=V_1\oplus...\oplus V_k$, 其中V_i均非0. 设$P_i:V\longrightarrow V$是V_i上的正交投影变换, $A=P_1+...+P_k$. 则$0<|A|\leq 1$且$|A|=1$当且仅当$V_i\perp V_j, i\neq j$.

57. (1) 如何在酉空间中定义 Hermite 矩阵对应的 Hermite 变换? 导出 Hermite 变换的一个判断准则.
(2) 在酉空间中定义伴随变换与自伴变换, 并导出伴随变换的基本性质(参考定理2.5.13).

58. 证明：一个幂等矩阵A是自伴的当且仅当$N(A)\perp R(A)$.

59. 证明集合V/U按公式(2.7.2)定义的加法和数乘作成线性空间.

60. 验证公式(2.7.4)定义的诱导映射$\bar{\sigma}$是线性变换. 条件"U是σ的不变子空间"可以去掉吗？

61. (1) 设$A\in M_s(\mathbb{F}), B, C\in M_{n-s}(\mathbb{F})$. 设$V$是有限维线性空间, $\sigma\in\mathrm{End}V$. 设σ在V的两组基

$$\alpha_1,\cdots,\alpha_s,\alpha_{s+1},\cdots,\alpha_n \text{与} \alpha_1,\cdots,\alpha_s,\beta_{s+1},\cdots,\beta_n$$

[41]Dragomir Z. Djokovic(1937-), 南斯拉夫-加拿大数学家.

下的矩阵分别为$\begin{pmatrix} A & X \\ 0 & B \end{pmatrix}$与$\begin{pmatrix} A & Y \\ 0 & C \end{pmatrix}$. 证明$U = \mathrm{Span}\{\alpha_1, \cdots, \alpha_s\}$是$\sigma$的不变子空间, 且$\sigma$的诱导映射$\bar{\sigma}$关于商空间$V/U$的两组基$\bar{\alpha}_{s+1}, \cdots, \bar{\alpha}_n$与$\bar{\beta}_{s+1}, \cdots, \bar{\beta}_n$下的矩阵分别为$B$与$C$. 并据此推广推论 2.7.2.

(2) 设A, B, C均是方阵. 不用线性变换直接证明：$A \oplus B$与$A \oplus C$相似 $\iff$ B与C相似.

62. 推广例 2.6.7, 即设A, B均为n阶矩阵. 设σ是$\mathbb{F}^{n\times n}$上的一个线性变换, 其中

$$\sigma:\ X \mapsto \sigma(X) = AXB^T.$$

证明σ关于标准基$E_{11}, \cdots, E_{1n}, E_{21}, \cdots, E_{2n}, E_{n1}, \cdots, E_{nn}$ 的矩阵是$B \otimes A$.

63. 设σ是$\mathbb{C}$到自身的线性变换, 其定义为

$$\sigma:\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto A\begin{pmatrix} x \\ y \end{pmatrix},$$

其中

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

将$(x\,,y)^T$记为普通复数$x + y\mathrm{i}$, 证明$\sigma((x, y)^T) = (a + b\mathrm{i})(x + y\mathrm{i})$. 请解释之.

64. 设$V = M_2(\mathbb{C}), U = M_2(\mathbb{R})$. 记$\mathrm{i}U = \{A \in V \mid A = \mathrm{i}X, \exists X \in U\}$. 研究(实)线性空间$U \oplus (\mathrm{i}U)$, V/U以及$(V/U) \otimes \mathbb{C}$的结构.

65. 设$\sigma \in \mathrm{Hom}(U, V)$. 证明商空间$U/\mathrm{Ker}\sigma$与$\mathrm{Im}\sigma$同构并建立相应的一个同构映射.

66. (1) 复数域$\mathbb{C}$是实数域$\mathbb{R}$上的2维线性空间. 是否存在$\mathbb{C}$上的一个内积, 使得i 与$1 + \mathrm{i}$成为$\mathbb{C}$的一组标准正交基, 为什么?

(2) 试构造实线性空间$\mathbb{R}^3$上的一个内积, 使得向量组$e_1, e_1+e_2, e_1+e_2+e_3$是一组标准正交基. 问此时$e_2$与$e_3$的长度各是多少? 它们的夹角又是多少?

(3) 试尽可能一般性地(1)与(2)中的问题.

67. 设n维欧氏空间V的一组基为$\alpha_1, \alpha_2, \cdots, \alpha_n$. 证明：存在正定矩阵$C$, 使得由

$$(\beta_1, \beta_2, \cdots, \beta_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n)C$$

确定的向量组$\beta_1, \beta_2, \cdots, \beta_n$是$V$的一个标准正交基.

68. 设A是**反对称**实矩阵(即$A^T = -A$), 证明:

(1) A的秩是偶数.

(2) A的特征值为 0 或纯虚数.

(3) 设$\alpha + \beta\mathrm{i}$是A的属于一个非零特征值的特征向量, 其中α, β均为实向量, 则α与β正交.

69. 设$f(x, y)$是$\mathbb{C}^n$上的对称**双线性函数** (即$f(x, y) = f(y, x)$且关于两个变元x与y均是线性的).

(1) 给出$f(x, y)$的一般表达式.

(2) 证明方程$f(x, x) = 0$总有非零解.

(3) 设$f(x, y)$非退化(即若$f(x, \alpha) = 0, \forall x \in \mathbb{C}^n$, 则$\alpha = 0$), 证明存在线性无关的向量$\alpha, \beta$使得$f(\alpha, \beta) = f(\beta, \alpha) = 0$且$f(\alpha, \beta) = 1$.

70. 设U, W是线性空间V的两个子空间.

(1) 证明$(U + W)/W$与$U/(U \bigcap W)$是同构的线性空间.

(2) 试建立(1)中的不依赖于任何基的一个同构映射.

71. (双线性函数与张量积) 设$A \in \mathbb{F}^{m\times n}$. 对任意$x \in \mathbb{F}^m, y \in \mathbb{F}^m$, 定义

$$f(x, y) = x^T Ay.$$

则$f(x, y)$称为$\mathbb{F}$上的一个$m \times n$维的双线性函数, A称为该双线性函数的矩阵. $\mathbb{F}$上的$m \times n$维双线性函数的全体记为$\mathcal{B}(m, n)$.

(1) 证明$\mathcal{B}(m,n)$按照普通加法与数乘运算构成$\mathbb{F}$上的线性空间.

(2) 计算$\mathcal{B}(m,n)$的维数与一组基.

(3) 建立$\mathcal{B}(m,n)$与$\mathrm{Hom}_{\mathbb{F}}(\mathbb{F}^m\otimes\mathbb{F}^n,\mathbb{F})=(\mathbb{F}^m\otimes\mathbb{F}^n)^*$之间的一个同构映射. 解释之.

72. 设U与W是线性空间V的两个子空间, $\alpha,\beta\in V$. 证明:

(1) $\alpha+U=P_{U^\perp}(\alpha)+U$, 其中$P_{U^\perp}$是$U$的正交补空间$U^\perp$的投影变换.

(2) $(\alpha+U)+(\beta+W)=(\alpha+\beta)+(U+W)$.

(3) $(\alpha+U)\cap(\beta+W)\neq\emptyset \iff \alpha-\beta\in U+W$.

73. 设U,W是V的两个子空间. 证明$(U+W)/(U\bigcap W)\cong U/(U\bigcap W)\oplus W/(U\bigcap W)$. 进一步, 设$U_0,W_0$分别是$U$与$W$的子空间, 证明**蝴蝶引理**

$$\frac{U_0+U\bigcap W}{U_0+U\bigcap W_0}\cong\frac{U\bigcap W}{U\bigcap W_0+U_0\bigcap W}\cong\frac{U\bigcap W+W_0}{U_0\bigcap W+W_0}.$$

74. 设$A=\begin{pmatrix}1&-1\\-1&1\end{pmatrix}$. 定义$\mathbb{R}^2$上的二元(向量)函数$<x,y>$如下:

$$<x,y>=x^TAy.$$

此二元函数与普通内积的差别是什么? 以此二元函数为基础, 建立相应的长度, 角度等概念, 研究其中的正交与平行的定理.

75. 设$V=\mathbb{R}^{n\times n}$. 记$\mathbb{H}=\{f:V\longrightarrow V\mid f(X+Y)=f(X)+f(Y),\forall X,Y\in V\}$.

(1) 设$f\in\mathbb{H}$. 如果$f(XY)=f(X)f(Y)$ 或$f(XY)=f(Y)f(X),\forall X,Y\in V$. 证明华罗庚定理:

$f(XY)=f(X)f(Y),\forall X,Y\in V$ 或$f(XY)=f(Y)f(X),\forall X,Y\in V$.

(2) 上述华罗庚定理是否对$V=\{A\in\mathbb{R}^{n\times n}\mid e_i^TAe_j=0,\forall 1\leq j<i\leq n\}$仍然成立? 为什么?

研究性问题.

对于任意矩阵$A\in\mathbb{F}^{m\times n},m\neq n$, 从几何层面看, A的行列式$|A|$仍然有意义, 即将其视为以A的列向量为邻边的平行多面体的(有向)体积, 比如$m\geq 1,n=1$时, A是一个列向量, 行列式$det(A)$可以是A的某种有向度量; 从代数层面看, 我们在第一章曾经指出, 行列式的本质是“标准交错多重线性函数”, 这样的函数对任意$m\times n$ 矩阵都有意义, 即实际上可以对任意矩阵定义其行列式如下.

定义 2.8.1. 设$k,x,y\in\mathbb{F},\alpha,\beta,\alpha_i\in\mathbb{F}^m,1\leq i\leq n,\sigma\in S_n$. 则存在唯一的函数$f:(\mathbb{F}^m)^n\longrightarrow\mathbb{F}$ 满足下述条件:

(1) $f(\alpha_1,\cdots,\alpha_{i-1},x\alpha+y\beta,a_{i+1},\cdots,\alpha_n)$
$=xf(\alpha_1,\cdots,\alpha_{i-1},\alpha,\alpha_{i+1},\cdots,\alpha_n)+yf(\alpha_1,\cdots,\alpha_{i-1},\beta,\alpha_{i+1},\cdots,\alpha_n)$.

(2) $f(\alpha_1,\cdots,\alpha_i,\alpha_{i+1},\cdots,\alpha_n)=-f(\alpha_1,\cdots,\alpha_{i+1},\alpha_i,\cdots,\alpha_n)$.

(3) $f(e_{\sigma(1)},\cdots,e_{\sigma(n)})=(-1)^{sgn(\sigma)}$, 如果 $m\geq n$.

(4) $f(\alpha_1,\cdots,\alpha_n)=0$, 如果 $m<n$.

显然, 当$m=n$时, 定义2.8.1与方阵的普通行列式完全相同; 当$m\neq n$时, 可以证明, 如果函数f满足定义2.8.1中的前三个条件且满足乘法公式$f(AB)=f(A)f(B),A\in\mathbb{F}^{m\times p},B\in\mathbb{F}^{p\times n}$, 则对$m<n$的所有矩阵$X$必然有$f(X)=0$, 即定

义 2.8.1中的第四个条件自动成立[42]. 与方阵不同的是, 当$m \neq n$时, 满足定义 2.8.1的函数不是唯一的.

[42]该定义及其相关性质由上海交通大学2018级本科生陆雪松、毛松涛和王子兴于2020年给出.

第三章　Jordan标准形与特征值问题

本 章 提 要

在第一章第三节(特征值)中, 我们看到求解微分方程组$x' = Ax$常常需要作变量代换$x = PY$, 从而产生与A相似的矩阵$P^{-1}AP$以及计算A^m等矩阵理论的基本问题. 如果A可以(相似)对角化, 即$P^{-1}AP = D$为对角矩阵, 则$A^m = PD^mP^{-1}$. 本章的主要结论有二：

1、Schur[43]酉三角化引理. 任何复方阵A都可以酉三角化, 即存在酉矩阵U使得U^*AU是上三角矩阵.

2、任何复方阵A都相似于Jordan[44]标准形J, 而J^m非常容易计算.

历史上, Jacobi 1837 年证明了**Jacobi 定理**，即任何复矩阵均可以(相似)三角化. 1910年, Schur进一步证明了上述酉三角化引理. 这是矩阵理论最强大的基础工具, 请读者务必深刻理解、熟练掌握.

第一节　Schur 三角化定理

3.1.1　Schur 酉三角化定理

存在不能相似对角化的矩阵, 但每个方阵都可以三角化, 这就是Schur 酉三角化定理.

定理 3.1.1. 设$A \in \mathbb{C}^{n\times n}$, 则存在酉矩阵$U$与上三角矩阵$B$, 使得

$$U^*AU = B \tag{3.1.1}$$

证　对矩阵A的阶数n施行数学归纳法. 当$n = 1$时, 结论显然成立. 假设结论对阶数小于n的矩阵都成立, 下证结论对n阶矩阵也成立.

设$\lambda_1 \in \mathbb{C}$为A的一个特征值, 相应的单位特征向量为α_1, 将α_1扩充成$\mathbb{C}^n$的一组标准正交基$\alpha_1, \alpha_2, \cdots, \alpha_n$, 则$U_1 = (\alpha_1, \alpha_2, \cdots, \alpha_n)$是一个酉阵. 因为$A\alpha_1 = \lambda\alpha_1, A\alpha_2, \cdots, A\alpha_n$为基$\alpha_1, \alpha_2, \cdots, \alpha_n$的线性组合, 所以

$$AU_1 = A(\alpha_1, \alpha_2, \cdots, \alpha_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n)\begin{pmatrix} \lambda_1 & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & c_{22} & c_{23} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c_{n2} & c_{n3} & \cdots & c_{nn} \end{pmatrix},$$

其中$c_{ij} \in \mathbb{C}$. 令

$$C_1 = \begin{pmatrix} c_{22} & c_{23} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n2} & c_{n3} & \cdots & c_{nn} \end{pmatrix}.$$

[43]Issai Schur (1875-1941), 著名德国数学家, Frobenius的学生.

[44]Marie Ennemond Camille Jordan (1838-1922), 法国著名数学家, 数学上还有著名的Jordan Curve Theorem. 第 25593 号小行星 camillejordan 即以其名字命名.

则

$$AU_1 = U_1 \begin{pmatrix} \lambda_1 & c_{12}\ c_{13} \cdots c_{1n} \\ 0 & \\ \vdots & C_1 \\ 0 & \end{pmatrix}.$$

由于C_1为$n-1$阶矩阵, 由归纳假设, 存在$n-1$阶酉矩阵U_2使

$$U_2^* C_1 U_2 = B_1 = \begin{pmatrix} b_{22} & b_{23} & \cdots & b_{2n} \\ & b_{33} & \cdots & b_{3n} \\ & & \ddots & \vdots \\ & & & b_{nn} \end{pmatrix}$$

为上三角矩阵. 令

$$U = U_1 \begin{pmatrix} 1 & \\ & U_2 \end{pmatrix},$$

则

$$\begin{aligned} U^* A U &= \begin{pmatrix} 1 & \\ & U_2^* \end{pmatrix} U_1^* A U_1 \begin{pmatrix} 1 & \\ & U_2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & \\ & U_2^* \end{pmatrix} \begin{pmatrix} \lambda_1 & c_{12}\ c_{13} \cdots c_{1n} \\ 0 & \\ \vdots & C_1 \\ 0 & \end{pmatrix} \begin{pmatrix} 1 & \\ & U_2 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & d_{12} & d_{13} & \cdots & d_{1n} \\ & b_{22} & b_{23} & \cdots & b_{2n} \\ & & b_{33} & \cdots & b_{3n} \\ & & & \ddots & \vdots \\ & & & & b_{nn} \end{pmatrix}. \end{aligned}$$

定理得证. □

公式(3.1.1)可变形为

$$A = UBU^* \tag{3.1.2}$$

一般称公式(3.1.2)为矩阵A的 **Schur 分解**.

从线性变换的角度看, Schur 分解将任意线性变换的作用分解为较为简单的三步, 即$Ax = (U(B(U^* x)))$, 其中第一步是旋转($y = U^* x$), 第二步是三角变换($z = By$), 第三步是第一步旋转的反转(Uz). 旋转是最基本最简单的线性变换, 三角变换较之一般的线性变换大为简单, 因此, Schur酉三角化定理大大简化了矩阵的相关计算, 是矩阵理论中最重要的结果之一.

3.1.2 Schur 分块三角化定理

从Schur酉三角化定理出发, 还可以进一步使用相似变换得到更为简单的分块对角矩阵, 即每个对角块是对角元素均相同三角矩阵, 称这样的矩阵为**"纯三角矩阵"**, 此即下述**"分块 Schur 三角化定理**.

定理 3.1.2. 设n阶复矩阵A的特征多项式$|\lambda I - A| = \prod\limits_{i=1}^{s}(\lambda-\lambda_i)^{n_i}$, 其中$\sigma(A) = \{\lambda_i \,|\, 1 \le i \le s\}$, $\sum\limits_{i=1}^{s} n_i = n$. 则存在可逆矩阵$P$使得

$$P^{-1}AP = A_1 \oplus A_2 \oplus \cdots \oplus A_s,$$

其中A_i是特征值均为λ_i的n_i阶上三角矩阵.

证 由Schur酉三角化定理可设A是上三角矩阵. 对n作归纳. $n=2$时, A或者对角线元素相同, 或者A相似于对角矩阵, 定理自然成立. 假设$n=k$时定理成立, 考虑$n=k+1$的情形. 记$A = \begin{pmatrix} A_1 & \alpha \\ 0 & a_{nn} \end{pmatrix}$. 按假设, $A_1 = B_1 \oplus B_2 \oplus \cdots \oplus B_q$是对角元素互不相同的纯上三角矩阵的直和. 如果$a_{nn}$与其它对角元素均不相同, 则由第二章推论2.3.1可知A相似于$A_1 \oplus (a_{nn})$. 如果a_{nn}与其中某个纯三角矩阵的对角元素相同, 则由于矩阵的直和项交换次序保持相似性, 故可以假定A_1的最后一个直和项的对角元素就是a_{nn}, 于是$A = \begin{pmatrix} C & X \\ 0 & M \end{pmatrix}$, 其中$M$是对角元素为$a_{nn}$的纯上三角矩阵, $C = B_1 \oplus B_2 \oplus \cdots \oplus B_{q-1}$的特征值与$M$的特征值互异. 故再由第二章推论2.3.1即可知$A = \begin{pmatrix} C & X \\ 0 & M \end{pmatrix}$相似于$\begin{pmatrix} C & 0 \\ 0 & M \end{pmatrix}$. □

分块Schur三角化定理显然比酉三角化定理更加强大, 比如读者可以由其简单地证明Cayley-Hamilton定理. 此处证明矩阵对角化的**最小多项式准则**.

定理 3.1.3. 矩阵A可以对角化 $\Longleftrightarrow$ A的最小多项式没有重根.

证 对角矩阵的最小多项式没有重根, 所以必要性是显然的. 下证充分性.

设A的最小多项式$m(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_s)$无重根. 需要证明分块Schur三角化定理中的每个对角块都是对角矩阵. 由于相似矩阵具有相同的对角化性质, 故由分块Schur三角化定理可设矩阵

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_s,$$

其中A_i是特征值均为λ_i的n_i阶上三角矩阵, $\sum\limits_{i=1}^{s} n_i = n$. 由于诸$\lambda_i$互不相同, A的最小多项式是每个直和项A_i的最小多项式的乘积. 然而, 纯三角矩阵的最小多项式无重根当且仅当其是对角矩阵, 故每个A_i都是对角矩阵, 从而A是对角矩阵. □

推论 3.1.1. 设A为方阵, $f(x)$是无重因式的多项式. 若$f(A)=0$, 则A可以对角化.

例 3.1.1. 幂等矩阵与对合矩阵均可以对角化.

这是因为幂等矩阵A满足$A^2 = A$, 故其最小多项式$m(x) \,|\, (x^2 - x)$, 故无重根; 而对合矩阵A满足$A^2 = I$, 从而其最小多项式$m(x) \,|\, (x^2-1)$, 也无重根. 因此, 这两类矩阵均可以对角化.

3.1.3 酉三对角化定理

酉三角化往往成本过高而难以实际应用, 因此常常退而求其次而使用酉三角化定理先将矩阵化简为上**Hessenberg矩阵**, 即满足条件$e_i^T H e_j = 0$, 其中$i \ge j+2$的矩阵H. 和上三角矩阵相比, Hessenberg矩阵多了一条非0的下对角线. 下Hessenberg矩阵可以类似定义.

后文中Hessenberg矩阵均指上Hessenberg矩阵.

例 3.1.2. $A = \begin{pmatrix} a & b & c \\ d & e & f \\ 0 & g & h \end{pmatrix}$, $B = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ 0 & 0 & v & w \\ 0 & 0 & x & y \end{pmatrix}$都是Hessenberg矩阵.

显然, 三角矩阵都是Hessenberg矩阵, 故由Schur酉三角化定理直接可得

定理 3.1.4. 设$A \in \mathbb{C}^{n\times n}$, 则存在酉矩阵$U$与Hessenberg矩阵$H$, 使得

$$U^*AU = H \tag{3.1.3}$$

将公式(3.1.5)改写为如下形式

$$A = UHU^* \tag{3.1.4}$$

即得A的**Hessenberg分解**.

Hessenberg分解显然比Schur分解更容易获得, 可以证明(留作习题), 仅需要若干Householder变换, 即可将任意矩阵化为Hessenberg矩阵, 确切地说, 有下述

命题 3.1.1. 设$A \in \mathbb{C}^{n\times n}$, 则存在Householder矩阵$H_1, ..., H_s$与Hessenberg矩阵$H$, 使得

$$H_s...H_1AH_1...H_s = H \tag{3.1.5}$$

一个Hessenberg矩阵$H \in \mathbb{C}^{n\times n}$称为**不可约的**, 如果$e_{i+1}^THe_i \neq 0, 1 \leq i \leq n-1$, 即下对角线元素均非0. 不难看出, 不可约Hessenberg矩阵的特征值都是单的.

例 3.1.3. 在例 3.1.2中, 矩阵B不是不可约的; 如果$dg \neq 0$, 则矩阵A是不可约的.

定理3.1.5. 设$A = UHU^*$是矩阵$A \in \mathbb{C}^{n\times n}$的Hessenberg分解, 其中$U$是酉矩阵, H是Hessenberg矩阵. 则$H = U^*AU$不可约 $\Longleftrightarrow$ 矩阵$B = U^*K(A, u_1, n)$是可逆上三角矩阵, 其中$K(A, u_1, n)$是Krylov矩阵, $u_1 = Ue_1$是矩阵U的第一列.

证 首先, 由$u_1 = Ue_1$可得$U^*u_1 = e_1, U^*(UH^jU^*u_1) = H^je_1$, 故

$$B = U^*K(A, u_1, n) = U^*(u_1, Au_1, ..., A^{n-1}u_1) = (e_1, He_1, ..., H^{n-1}e_1) \tag{3.1.6}$$

因此B的第j个对角元素为

$$e_j^TBe_j = \prod_{j=1}^{i-1} e_{j+1}^THe_j \tag{3.1.7}$$

由于H是Hessenberg矩阵, 显然B是上三角矩阵. 进一步, $H = U^*AU$不可约$\Leftrightarrow$ $e_{j+1}^THe_j \neq 0 (1 \leq j \leq n-1) \Leftrightarrow e_k^TBe_k \neq 0 \Leftrightarrow B$可逆. □

Hessenberg矩阵的一个重要特例是**三对角矩阵**, 即除中心三条对角线外的元素均为0, 或满足条件$e_i^TAe_j = 0, \forall i \leq j-2$或$i \geq j+2$的矩阵$A$.

例 3.1.4. $A = \begin{pmatrix} a & b & 0 \\ c & d & e \\ 0 & f & g \end{pmatrix}$, $B = \begin{pmatrix} a & b & 0 & 0 \\ c & d & e & 0 \\ 0 & 0 & f & g \\ 0 & 0 & x & y \end{pmatrix}$都是三对角矩阵.

由定理3.1.4与定理3.1.5容易证明(留作习题)下面的**三对角分解定理**.

定理 3.1.6. 设$A \in \mathbb{C}^{n\times n}, A^* = A$, 则存在酉矩阵$U$与Hermite三对角矩阵$T$, 使得

$$U^*AU = T \text{ 或 } A = UTU^* \tag{3.1.8}$$

特别地, $T = U^*AU$不可约 $\iff$ 矩阵$B = U^*K(A, u_1, n)$是可逆上三角矩阵, 其中$K(A, u_1, n)$是Krylov矩阵, $u_1 = Ue_1$是矩阵U的第一列.

定理3.1.6保证了Hermite矩阵可以通过正交相似变换化为三对角矩阵, 而三对角矩阵的非零元素最多只有$3n-2$个(对称三对角矩阵甚至只有$2n-1$个), 非常便于存储与计算.

思考题
1. 实数域上的Schur三角化定理成立吗, 即每个实方阵是否可以正交三角化?
2. 是否每个矩阵都可以分块酉三角化, 即分块Schur三角化定理中的可逆矩阵是否可以加强为酉矩阵?

第二节 Jordan标准形

分块Schur三角化定理每个矩阵均相似于纯三角矩阵的直和, 但纯三角矩阵可以写成

$$A = \lambda I + N$$

的形式, 其中N是严格三角矩阵. 因为

$$P^{-1}AP = \lambda I + P^{-1}NP,$$

故只需研究严格上三角矩阵的相似标准形即可.

3.2.1 幂零矩阵的Jordan标准形

例 3.2.1. 设A是n阶严格上三角矩阵且$r(A) = n-1$, 则A与J_n相似.

证 将A写成分块矩阵的形式有$A = \begin{pmatrix} 0 & B \\ 0 & \mathbf{0}^T \end{pmatrix}$, 其中$B$是$n-1$阶上三角矩阵. 因$r(B) = r(A) = n-1$知$B$可逆, 故$B$的主对角元素均非0. 因此$A$相似于第一条上对角线均为1的严格上三角矩阵. 注意 **"将第i行的a倍加到第1行对应的相似变换是将第1列的$-a$倍加到第i列"**, 而第一列是0向量, 从而A进一步相似于第1行为e_2^T而其余各行不变的矩阵. 最后使用归纳法即可. □

例 3.2.2. 设$A = \begin{pmatrix} 0 & J_3 \\ 0 & \mathbf{0}^T \end{pmatrix}$是4阶严格上三角矩阵, 注意$A$不与$B = \begin{pmatrix} J_3 & 0 \\ \mathbf{0}^T & 0 \end{pmatrix}$相似, 而与$C = J_2 \oplus J_2$相似.

容易验证幂零Jordan块J_n的下述性质.

命题 3.2.1. (1) $J_n^TJ_n = (0) \oplus I_{n-1}$, $J_nJ_n^T = I_{n-1} \oplus (0)$.
(2) $J_ne_i = e_{i-1}$, 其中e_i是标准向量, 约定e_0是0向量.
(3) 对任意$x \in \mathbb{C}^n$, $(I_n - J_n^TJ_n)x = (x^Te_1)e_1$.

例 3.2.3. 设$\alpha \in \mathbb{C}^k$, 严格上三角矩阵$A = \begin{pmatrix} 0 & \alpha^T \\ 0 & J_k \end{pmatrix}$, 则

(1) 若$\alpha^T e_1 = 0$(此即α^T的第一个分量为0), 则A与J_k相似.
(2) 若$\alpha^T e_1 \neq 0$, 则A与J_{k+1}相似.

证 (1) 由命题3.2.1, 可以得出下述等式

$$\begin{pmatrix} 1 & -\alpha^T J_k^T \\ 0 & I_k \end{pmatrix} \begin{pmatrix} 0 & \alpha^T \\ 0 & J_k \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & J_k \end{pmatrix} \begin{pmatrix} I & -\alpha^T J_k^T \\ 0 & I_k \end{pmatrix} \tag{3.2.1}$$

(2) 此是例3.2.3的直接推论, 也可由命题3.2.1得出下述等式

$$\begin{pmatrix} 1 & -\alpha^T J_k^T \\ 0 & I_k \end{pmatrix} \begin{pmatrix} 0 & \alpha^T \\ 0 & J_k \end{pmatrix} = \begin{pmatrix} 0 & (\alpha^T e_1) e_1^T \\ 0 & J_k \end{pmatrix} \begin{pmatrix} I & -\alpha^T J_k^T \\ 0 & I_k \end{pmatrix}. \tag{3.2.2}$$

但$\alpha^T e_1 \neq 0$, 故矩阵$\begin{pmatrix} 0 & (\alpha^T e_1) e_1^T \\ 0 & J_k \end{pmatrix}$与$J_{k+1}$相似. □

引理 3.2.1. 设$k_1 \geq k_2 \geq 1, \alpha_i \in \mathbb{C}^{k_i}, i = 1, 2$, 严格上三角矩阵$A = \begin{pmatrix} 0 & \alpha_1^T & \alpha_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J_{k_2} \end{pmatrix}$. 则

(1) 若$\alpha_1^T e_1 = \alpha_2^T e_1 = 0$, 则$A$与$J_{k_1} \oplus J_{k_2} \oplus (0)$相似.
(2) 若$\alpha_1^T e_1 = 0, \alpha_2^T e_1 \neq 0$, 则$A$与$J_{k_1} \oplus J_{k_2+1}$相似.
(3) 若$\alpha_1^T e_1 \neq 0$, 则A与$J_{k_1+1} \oplus J_{k_2}$相似.

证 (1) 由例3.2.3可知, 若$\alpha_1^T e_1 = 0$, 则A相似于$\begin{pmatrix} 0 & 0 & \alpha_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J_{k_2} \end{pmatrix}$, 从而相似于矩阵(交换前两列与前两行是相似变换)

$$\begin{pmatrix} J_{k_1} & 0 & 0 \\ 0 & 0 & \alpha_2^T \\ 0 & 0 & J_{k_2} \end{pmatrix} = J_{k_1} \oplus \begin{pmatrix} 0 & \alpha_2^T \\ 0 & J_{k_2} \end{pmatrix} \tag{3.2.3}$$

再由$\alpha_2^T e_1 = 0$知$\begin{pmatrix} 0 & \alpha_2^T \\ 0 & J_{k_2} \end{pmatrix}$相似于$J_{k_2} \oplus (0)$.

(2) 此是公式(3.2.3)与例3.2.3(2)的直接推论.

(3) 若$\alpha_1^T e_1 \neq 0, \alpha_2^T e_1 = 0$, 则$A$与$\begin{pmatrix} 0 & \alpha_1^T & 0 \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J_{k_2} \end{pmatrix}$相似, 从而与$J_{k_1+1} \oplus J_{k_2}$相似.

若$\alpha_1^T e_1 \neq 0, \alpha_2^T e_1 \neq 0$, 则$A$与$\begin{pmatrix} J_{k_1+1} & E_{11} \\ 0 & J_{k_2} \end{pmatrix}$相似(此处$E_{11}$是矩阵单位), 此时将第2列的$-1$倍加到第$k_1+2$列可将第一行$E_{11}$中的1消去, 相应的相似变换需要将$k_1+2$行的1倍加到第2行, 故$A$与矩阵$\begin{pmatrix} J_{k_1+1} & E_{22} \\ 0 & J_{k_2} \end{pmatrix}$相似; 进一步与$\begin{pmatrix} J_{k_1+1} & E_{33} \\ 0 & J_{k_2} \end{pmatrix}$相似, 等等. 注意到第$k_1+k_2+1$行为0以及$k_1 \geq k_2$可知, 上述相似过程的最终矩阵为$\begin{pmatrix} J_{k_1+1} & 0 \\ 0 & J_{k_2} \end{pmatrix}$, 即$A$与$J_{k_1+1} \oplus J_{k_2}$相似. □

该引理有很好的启发作用, 请读者务必手算完整过程以理解该证明所包含的算法. 由此即可证明**"幂零矩阵的Jordan标准形定理"**[45].

定理 3.2.1. 设A是n阶幂零矩阵. 则存在可逆矩阵P和正整数$n_1 \geq n_2 \geq \cdots \geq n_m$, $n_1 + n_2 + \cdots + n_m = n$, 使得

$$P^{-1}AP = J_{n_1} \oplus J_{n_2} \oplus \cdots \oplus J_{n_m}. \tag{3.2.4}$$

上式右端的矩阵称为**幂零 Jordan 矩阵**, 它是矩阵A的Jordan标准形, 记为J_A或J. 不计诸n_j的次序与大小, A的Jordan标准形是唯一的.

证 首先, 公式(3.2.4)右端每个直和因子的顺序不影响矩阵的相似性, 因此条件

$$n_1 \geq n_2 \geq \cdots \geq n_m,\ n_1 + n_2 + \cdots + n_m = n$$

总是可以满足的, 故只需证明公式(3.2.4)及唯一性.

其次, 由Schur三角化定理可设A是严格上三角矩阵. 对n作归纳. 当$n = 1$时, 定理显然成立. 假设对阶数$< n$的所有严格上三角矩阵, 定理成立. 现将A写成分块矩阵

$$\begin{pmatrix} 0 & \alpha^T \\ 0 & A_1 \end{pmatrix},$$

其中$\alpha \in \mathbb{C}^{n-1}$而$A_1$是$n-1$阶严格上三角矩阵. 由归纳假设, A_1相似于形如公式(3.2.4)的下述矩阵

$$A_2 = J_{k_1} \oplus J_{k_2} \oplus \cdots \oplus J_{k_s},$$

其中$k_1 \geq k_2 \geq \cdots \geq k_s$, $k_1 + k_2 + \cdots + k_s = n-1$. 令

$$J = J_{k_2} \oplus \cdots \oplus J_{k_s}$$

则$A_2 = J_{k_1} \oplus J$. 且存在其中$\beta \in \mathbb{C}^{k_1}, \gamma \in \mathbb{C}^{n-1-k_1}$使得$A$进一步相似于矩阵

$$A_3 = \begin{pmatrix} 0 & \beta^T & \gamma^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix}.$$

现若$\beta^T e_1 = 0$, 则A_3相似于$A_4 = \begin{pmatrix} 0 & 0 & \gamma^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix}$, 从而相似于矩阵$\begin{pmatrix} J_{k_1} & 0 & 0 \\ 0 & 0 & \gamma^T \\ 0 & 0 & J \end{pmatrix}$. 再由归纳假设可知严格上三角矩阵$\begin{pmatrix} 0 & \gamma^T \\ 0 & J \end{pmatrix}$相似于一个幂零Jordan矩阵$J'$. 因此$A$相似于$J_{k_1} \oplus J'$, 定理成立.

如果数$\beta^T e_1 \neq 0$, 则按例3.2.3(2)知A_3相似于

$$A_5 = \begin{pmatrix} J_{k_1+1} & e_1\gamma^T \\ 0 & J \end{pmatrix} \tag{3.2.5}$$

[45]该定理及其一般形式即本节定理3.2.2.

由命题3.2.1(2)知, $J_{k_1+1}e_{i+1}=e_i$, 故

$$\begin{pmatrix} I & e_2\gamma^T \\ 0 & I \end{pmatrix}\begin{pmatrix} J_{k_1+1} & e_1\gamma^T \\ 0 & J \end{pmatrix}\begin{pmatrix} I & -e_2\gamma^T \\ 0 & I \end{pmatrix}=\begin{pmatrix} J_{k_1+1} & e_2\gamma^T J \\ 0 & J \end{pmatrix}. \tag{3.2.6}$$

注意上式右端矩阵的右上角与原矩阵相比, 由$e_1\gamma^T$变为$e_2\gamma^T J$, 因此重复上述步骤, 可使式(3.2.6)右端矩阵的右上角变为$e_{N+1}\gamma^T J^N$, 其中N为任意正整数. 由$k_1 \geq k_i$可知$J^{k_1}=0$, 故A_5相似于矩阵$J_{k_1+1}\oplus J$, 而这是一个幂零Jordan矩阵.

再证唯一性. 仍对n作归纳. $n=1$时是显然的. 假设对阶数$<n$的严格上三角矩阵唯一性成立, 考虑n阶矩阵A的两个Jordan标准形

$$J=J_{n_1}\oplus J_{n_2}\oplus\cdots\oplus J_{n_s}$$

与

$$\tilde{J}=J_{m_1}\oplus J_{m_2}\oplus\cdots\oplus J_{m_t}$$

其中$n_1\geq n_2\geq\cdots\geq n_s\geq 1$, $m_1\geq m_2\geq\cdots\geq m_t\geq 1$. 则$n_1=m_1$, 否则, 不妨设$n_1>m_1$, 则$J^{m_1}\neq 0$而$\tilde{J}^{m_1}=0$, 矛盾! 现由第二章推论2.7.2知$J_{n_2}\oplus\cdots\oplus J_{n_s}$与$J_{m_2}\oplus\cdots\oplus J_{m_t}$相似, 故由归纳假设知$s=t$且$n_2=m_2,\cdots,n_s=m_s$. □

推论 3.2.1. 设n阶幂零矩阵A的Jordan标准形J同公式(3.2.4), 其幂零指数为e. 则

(1) J中Jordan块的最大阶数等于幂零指数, 即$e=\max\{n_i\,|\,1\leq i\leq m\}$.

(2) J中Jordan块的个数m等于A的零度, 等于A的线性无关的特征向量的个数.

(3) 记J中k阶Jordan块的个数为ℓ_k, A^k的秩与零度分别为$r_k,\eta_k=n-r_k$, $0\leq k\leq m$. 则

$$\ell_k=2\eta_k-\eta_{k-1}-\eta_{k+1}=r_{k-1}+r_{k+1}-2r_k,\quad 1\leq k\leq m. \tag{3.2.7}$$

证 (1) 设A的幂零指数为e, 其Jordan标准形为J, J的Jordan块的阶数的最大值为f. 则因$A^e=0$知$J^e=0$, 从而$f\leq e$. 同理, 由$J^f=0$可知$A^f=0$, 故$e\leq f$.

(2) 注意A的每个Jordan块的零度均为1, 故所有Jordan块的个数恰好是A的零度.

(3) 易知

$$r(J_p^i)=\begin{cases} 0, & \text{如果 } i\geq p; \\ p-i, & \text{如果 } i<p. \end{cases}$$

因此

$$r_i=r(A^i)=r(J^i)=\sum_{p=1}^{e}\ell_p r(J_p^i)=\sum_{p=i+1}^{e}\ell_p(p-i).$$

故(约定$r_0=r(A^0)=r(I)=n$)

$$r_{k-1}-r_k=\sum_{p=k}^{e}\ell_p,\forall\ 1\leq k\leq e.$$

再由$\eta(A^k)=n-r(A^k)$即可得公式(3.2.7). □

例 3.2.4. 验证下列矩阵为幂零矩阵, 并求其Jordan标准形:

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

解 直接计算知$A^2=0$. 故A的幂零指数为2. 由推论3.2.1可知, A的Jordan标准形J的最大Jordan块为2阶的. 由于$r(A)=2$, 故A的零度为2, 即有2个块, 所以$J=J_2\oplus J_2$.

例 3.2.5. 试证下列矩阵为幂零矩阵, 并求其Jordan标准形:

$$A=\begin{pmatrix} 2 & 6 & 5 & 0 & 0 & 4 & 0 & 6 \\ 1 & 3 & 2 & 0 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -2 & 0 & -1 & -4 & 0 & -1 \\ 2 & 5 & 4 & 0 & 0 & 4 & 0 & 5 \\ -1 & -3 & -2 & 0 & 0 & -2 & 0 & -3 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & -3 & -2 & 0 & 0 & -2 & 0 & -3 \end{pmatrix}.$$

解 直接计算可知$A^3=0$. 因此A为幂零矩阵, 且幂零指数为3. 再计算可得$r(A)=5$. 故由推论3.2.1可知, A的Jordan标准形J共有3块, 且最大Jordan块为3阶的. 因此其它两块必为一个3阶块和一个2阶块, 故知$J=J_3\oplus J_3\oplus J_2$.

由推论3.2.1即可得到下述**幂零矩阵相似准则**

推论 3.2.2. 设M与N是两个n阶幂零矩阵. 则M与N相似 $\Longleftrightarrow$ $r(M^k)=r(N^k), \forall k\geq 1$.

3.2.2 Jordan 标准形定理

例 3.2.6. 设A是秩为1的矩阵, 则A或者幂零, 或者特征值为$tr(A)$. 故A的Jordan标准形为:

(1) $\operatorname{diag}(\operatorname{tr}A, 0, \cdots, 0)$ (若 $\operatorname{tr}A\neq 0$), 或

(2) $\begin{pmatrix} 0 & 1 & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}$ (若$\operatorname{tr}A=0$, 此时A幂零, 且幂零指数为2).

定义 3.2.1. 矩阵λI_n+J_n称为n阶λ-Jordan块, 记为$J_n(\lambda)$.

一般矩阵的Jordan标准形定理是分块Schur三角化定理和幂零矩阵Jordan标准形定理的推论.

定理 3.2.2. 设A是n阶复矩阵, 则存在可逆矩阵P和正整数$n_1\geq n_2\geq\cdots\geq n_m$, $n_1+n_2+\cdots+n_m=n$, 使得

$$P^{-1}AP=J_{n_1}(\lambda_1)\oplus J_{n_2}(\lambda_2)\oplus\cdots\oplus J_{n_m}(\lambda_m). \tag{3.2.8}$$

上式右端的矩阵称为**Jordan矩阵**, 它是矩阵A的Jordan标准形. 不计诸n_j的次序与大小, A的Jordan标准形是唯一的.

证 由分块Schur三角化定理, 存在可逆矩阵P与n_i阶纯上三角矩阵A_i使得

$$P^{-1}AP=A_1\oplus A_2\oplus\cdots\oplus A_s.$$

由于$A_i=\lambda_i+N_i$, 其中N_i是严格上三角矩阵, 故由幂零矩阵的 Jordan 标准形定理, 知每个A_i均是幂零Jordan块的直和, 从而定理成立. □

注1. 公式(3.2.8)中的诸特征值λ_i可能相同.

注2. 定理 3.2.2仅对复数域上的矩阵成立, 因为实数域上的矩阵可能没有实特征值, 故其相似标准形具有其它形式.

推论 3.2.3. 设n阶矩阵A的 Jordan 标准形J同公式(3.2.8). 设μ为A的一个特征值, 记$(A-\mu I)^k$的秩为r_k, 零度为η_k, J中对角线元素为μ的k 阶 Jordan 块的个数为ℓ_k, 则

(1) η_1等于J中对角线元素为μ的 Jordan 块的个数;

(2) $\ell_1=2\eta_1-\eta_2$, $\ell_k=2\eta_k-\eta_{k-1}-\eta_{k+1}=r_{k-1}+r_{k+1}-2r_k$, $k\geq 2$.

例 3.2.7. 设$A\in\mathbb{F}^{n\times n}$, 则对$\forall m\geq n, r(A^m)=r(A^n)$. 特别地, $N(A^n)=N(A^{n+1}$.

证 设A的Jordan标准形$J=F\oplus N$, 其中F是所有非零特征值的Jordan块的直和, N是所有幂零Jordan块的直和. 设N中幂零块的最大阶数为$q\leq n$, 则$r(A^q)=r(J^q)=r(F^q\oplus 0)=r(F^q)$. 注意$F$可逆, 故对所有$m\geq q$均有$r(A^m)=r(F^q)=r(A^q)$. □

例 3.2.8. 方阵A可以对角化 $\iff$ A的 Jordan 标准形是对角矩阵.

例 3.2.9. 设A是例 3.2.4中的矩阵, 求可逆矩阵P, 使得$P^{-1}AP=J$.

解 设$P=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$, 可得

$$A(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)(J_2\oplus\ J_2)=(0,\alpha_1,0,\alpha_3).$$

由此可得P的各个列向量应满足的方程组分别为

$$A\alpha_1=0,\qquad A\alpha_2=\alpha_1,\qquad A\alpha_3=0,\qquad A\alpha_4=\alpha_3. \tag{3.2.9}$$

注意到四个方程组的系数矩阵都相同, 可以采用下面的方式统一求解:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & \vdots & b_1 \\ 0 & -1 & -1 & -1 & \vdots & b_2 \\ 1 & 1 & 0 & 1 & \vdots & b_3 \\ -1 & 0 & 1 & 0 & \vdots & b_4 \end{pmatrix}\to\begin{pmatrix} 1 & 0 & -1 & 0 & \vdots & b_1+b_2 \\ 0 & 1 & 1 & 1 & \vdots & -b_2 \\ 0 & 0 & 0 & 0 & \vdots & -b_1+b_3 \\ 0 & 0 & 0 & 0 & \vdots & b_1+b_2+b_4 \end{pmatrix}$$

由此可知要使方程组$Ax=\beta$有解, 向量$\beta=(b_1,b_2,b_3,b_4)^T$要满足

$$b_1=b_3,\qquad b_1+b_2+b_4=0.$$

解方程组

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

得$\alpha_1=(1,-1,1,0)^T$, $\alpha_3=(0,1,0,-1)^T$. 这两个向量都满足$Ax=\beta$的相容性条件. 解$Ax=\alpha_1$, 即

$$\begin{cases} x_1 \quad -x_3 \qquad =0, \\ \qquad x_2+x_3+x_4=1, \end{cases}$$

得$\alpha_2 = (1,0,1,0)^T$. 解$Ax = \alpha_3$, 即

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

得$\alpha_4 = (1,0,0,-1)^T$. 因此,

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

方程(3.2.9)中的向量α_1与α_3显然是属于特征值0的特征向量, 而向量α_2与α_4不是特征向量, 它们实际上满足方程(请验证!)

$$A^2 x = 0. \tag{3.2.10}$$

矩阵的高次幂难以直接计算, 但Jordan块的高次幂则有简单的计算公式.

定理 3.2.3. 设$J = J_n(\lambda)$是特征值为λ的n阶 Jordan 块, 则(其中约定, 若$s > m$, 则$C_m^s = 0$)

$$J^k = \begin{pmatrix} \lambda^k & C_k^1\lambda^{k-1} & C_k^2\lambda^{k-2} & \cdots & C_k^{n-1}\lambda^{k-n+1} \\ & \lambda^k & C_k^1\lambda^{k-1} & \cdots & C_k^{n-2}\lambda^{k-n+2} \\ & & \lambda^k & \cdots & C_k^{n-3}\lambda^{k-n+3} \\ & & & \ddots & \vdots \\ & & & & \lambda^k \end{pmatrix}. \tag{3.2.11}$$

证 记$N = J_n(0)$是n阶幂零Jordan块, 则$J = J_n(\lambda) = \lambda I + N$. 注意到$N^n = 0$, 故由二项式定理可知

$J^k = (\lambda I + N)^k = \sum_{m=0}^k C_k^m \lambda^{k-m} N^m$

$= \lambda^k I + C_k^1\lambda^{k-1}N + C_k^2\lambda^{k-2}N^2 + \cdots + C_k^{n-1}\lambda^{k-n+1}N^{n-1}$.

这就是定理中的公式.

由定理3.2.3, 对任意方阵A, 只要其特征值可以求出, 则可利用A的Jordan标准形来计算A的高次幂, 即设$A = P^{-1}JP$, 则$A^k = P^{-1}J^kP$. 然而, 一般而言, 基本上不可能求出高阶矩阵的全部特征值, 所以矩阵的特征值计算是矩阵理论的研究热点. 另外, 即使是幂零矩阵, 定理3.2.1的证明中给出的算法也是不能付诸实践的, 因为该算法是不稳定的(主要是因为Jordan标准形不是矩阵元素的连续函数). 参考下面的例子.

例 3.2.10. 设$A_t = \begin{pmatrix} t & 1 \\ 0 & 0 \end{pmatrix}$. 则当$t \neq 0$时, 其Jordan标准形为对角矩阵$D = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}$. 但当$t = 0$时, 其Jordan标准形为$A = J_2$. 因此当$t \to 0$时, A_t的Jordan标准形并不收敛于其极限的Jordan标准形.

尽管如此, Jordan标准形仍然是极其重要的, 因为它不仅具有非常简单的形式, 而且包含了相关矩阵的几乎所有信息, 比如秩, 特征值, 线性无关的特征向量的个数, 特征子空间的维数等等.

3.2.3 中国剩余定理与Jordan-Chevalley分解

由Jordan标准形定理可以证明下面的定理, 一般称为**Jordan-Chevalley**[46]**分解定理**.

定理 3.2.4. 设矩阵$A \in \mathbb{C}^{n\times n}$. 则存在唯一一对矩阵$D, N$使得$A = D + N$, 其中$D$为可对角化矩阵, N是幂零矩阵, 且$DN = ND$.

证明用到被誉为东方明珠的**"中国剩余定理"**. 中国剩余定理又称**秦九韶**[47]**定理**, 秦九韶在1247年完成的名著《数书九章》中称其为"大衍求一术", 比Euler[48]和Gauss的相同结果早500年以上. 中国剩余定理(Chinese Remainder Theorem)由著名德国数学家Cantor [49]命名, 其内容如下.

定理 3.2.5. 若 $q_1(x), \ldots, q_s(x)$ 为两两互素的多项式, $f_1(x), \ldots, f_s(x)$ 为任意多项式, 则下列多项式线性同余方程组

$$\begin{cases} f(x) \equiv f_1(x)(\mathrm{mod}\ q_1(x)); \\ \cdots\cdots\cdots\cdots \\ f(x) \equiv f_s(x)(\mathrm{mod}\ q_s(x)). \end{cases} \tag{3.2.12}$$

有解.

中国剩余定理完全解决了一次不定方程(组)的求解问题, 至今仍然有广泛深入的应用, 其整数版本的证明中学生即可完成, 有兴趣的读者请自行完成多项式版本的证明.

Jordan-Chevalley 分解定理的证明. 存在性. 设$J = D + N$是A的Jordan标准形, 其中D与N分别是J的对角部分与严格上三角部分. 则直接验证可知$DN = ND$. 由于$A = PJP^{-1} = PDP^{-1} + PNP^{-1}$, 显然$PDP^{-1}$与$PNP^{-1}$分别是可对角化矩阵与幂零矩阵, 并且$(PDP^{-1})(PNP^{-1}) = PDNP^{-1} = PNDP^{-1} = (PNP^{-1})(PDP^{-1})$.

唯一性. 可证更强的结论, 即D是A的多项式(从而N也是). 设$det(xI - D) = p(x) = \prod_i (x - \lambda_i)^{n_i}, \sum_i n_i = n$.

断言：$p(x) = det(xI - A)$. 设$v \in N(\lambda_i I - D), \forall i$, 则$Av = Dv + Nv = \lambda_i v + Nv$, 故$(A - \lambda_i I)v = Nv$. 因此必有

$$(A - \lambda_i I)^n v = N^n v = 0.$$

即$N(\lambda_i I - D) \subseteq N(\lambda_i I - A)^n$, 即$v$是$A$的广义特征向量. 再由$\sum_i n_i = n$知$det(\lambda I - A) = \prod_i (\lambda - \lambda_i I)^{n_i}$.

记$p_i(x) = (x - \lambda_i)^{n_i}, \forall i$. 由中国剩余定理, 存在多项式$f(x) \in \mathbb{C}[x]$满足同余方程

$$f(x) \equiv \lambda_i (mod p_i(x)), f(x) \equiv 0(mod x).$$

断言：$f(A) = D$.

只需证明对$v \in N(\lambda_i I - D), \forall i$, 则$f(A)v = Dv = \lambda_i v$即可. 由于$f(x) - \lambda_i \equiv 0(mod p_i(x)), f(x) - \lambda_i = g_i(x)(x - \lambda_i)^{n_i}$, 故

$$(f(A) - \lambda_i)v = g_i(A)(A - \lambda_i v)^{n_i} v = 0.$$

因此$f(A) = D, g(A) = A - f(A) = N$. □

[46]Claude Chevalley(1909-1984), 著名法国-美国数学家.
[47]秦九韶(1208-约1268), 著名南宋数学家, 被欧美史学家称为那个时代最伟大的数学家之一.
[48]Leonhard Euler(1707-1783), 欧拉, 著名瑞士数学家, 物理学家. 欧拉公式$e^{i\pi} + 1 = 0$被称为世界第一公式.
[49]Gerog Cantor(1845-1918), 著名德国数学家, 无限理论的奠基人.

例 3.2.11. 设$A=\begin{pmatrix}1&1\\0&0\end{pmatrix}$. 注意$A=\begin{pmatrix}1&0\\0&0\end{pmatrix}+\begin{pmatrix}0&1\\0&0\end{pmatrix}$不是$A$的Jordan-Chevalley 分解, 因为二矩阵不可换. 由于A本身可以对角化, 因此其Jordan-Chevalley 分解就是$A=A+0$.

例 3.2.12. 设$A=\begin{pmatrix}1&1&0\\0&1&1\\0&0&0\end{pmatrix}$. 注意$A=\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}+\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}$不是$A$的Jordan-Chevalley 分解. 由于$A$的两个特征值分别为1, 0, 相应的特征向量为$(1,0,0)^T,(1,-1,1)^T$, 相应于特征值1的广义特征向量为$(0,1,0)^T$. 故取$P=\begin{pmatrix}1&0&1\\0&1&-1\\0&0&1\end{pmatrix}$, 则由定理3.2.4可知

$$A=P(\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}+\begin{pmatrix}0&1&0\\0&0&0\\0&0&0\end{pmatrix})P^{-1}=P\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}P^{-1}+P\begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix}P^{-1}$$

是A的Jordan-Chevalley 分解, 即

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

一个线性变换如果其在某组基下的矩阵可以对角化, 则称该线性变换是**半单的**. 故可以对角化的矩阵也称为半单矩阵(部分作者用单纯矩阵(simple matrix)). 请读者自行证明下述乘法版本的Jordan-Chevalley 分解定理.

推论 3.2.4. 设A是可逆矩阵. 则存在半单矩阵S与幂幺矩阵U使得$A=SU=US$.

第三节 Jordan 标准形的其它理论

矩阵A的Jordan标准形可以认为是矩阵在相似变换下的最简形式. 前两节讨论的方法是纯计算, 历史上还有众多计算Jordan标准形的理论, 本节简介其中比较著名的两个(略去证明).

3.3.1 λ-矩阵

一般把未定元为λ而系数属于域$\mathbb{F}$的多项式全体记为$\mathbb{F}[\lambda]$, 称为$\mathbb{F}$上的一元多项式环. 所谓λ-矩阵即是指元素取自$\mathbb{F}[\lambda]$的矩阵, 一般记为$A(\lambda)$. 为简单起见, 本小节讨论的所有矩阵均假定是n阶的, 即属于$\mathbb{F}[\lambda]^{n\times n}$. 显然, $\mathbb{F}^{n\times n}\subseteq\mathbb{F}[\lambda]^{n\times n}$. 需要特别强调的是, 与普通矩阵相比, 行列式非零(与满秩等价)虽然是可逆的必要条件, 但不再是可逆性的必要条件. 比如λ作为1阶矩阵, 是满秩的但不可逆.

例 3.3.1. 任何矩阵$A\in\mathbb{F}^{n\times n}$的特征矩阵$\lambda I-A$的秩均为$n$. 特征矩阵均不可逆.

λ-矩阵的初等变换与通常矩阵的初等变换前两种(即重排变换与重排矩阵、倍乘变换与倍乘矩阵)完全相同, 但第三种消元变换与消元矩阵略有差异, 即消元矩阵可以使用多项式, 因此其消元矩阵为：$I+p(\lambda)E_{ij},i\neq j,\forall p(\lambda)\in\mathbb{F}[\lambda]$.

定理 3.3.1. 设矩阵$A(\lambda)\in\mathbb{F}[\lambda]^{n\times n}$. 则
(1) $A(\lambda)$可逆当且仅当其行列式为非零常数.

(2) $A(\lambda)$的秩为r可逆当且仅当其(在初等变换下)等价于对角矩阵

$$\operatorname{diag}(d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda), 0, \cdots, 0) \tag{3.3.1}$$

其中$d_i(\lambda)$均为首一多项式且$d_i(\lambda)|d_{i+1}(\lambda), 1 \le i \le r-1$.

公式(3.3.1)中的矩阵称为$A(\lambda)$的**Smith 标准形**[50], $d_i(\lambda)$称为$A(\lambda)$的**不变因子**.

$A(\lambda)$的每个正次数的不变因子分解为不同的首一一次多项式的幂的乘积, 这些一次多项式的幂合称为矩阵$A(\lambda)$的**初等因子**.

推论 3.3.1. 每个 Jordan 块仅有唯一的不变因子, 该不变因子就是它的唯一初等因子.

定理 3.3.2. 设$A, B \in \mathbb{F}^{n\times n}$. 则
(1) $\lambda I - A$与$\lambda I - B$等价 $\Longleftrightarrow$ 它们有相同的不变因子.
(2) 矩阵A与B相似 $\Longleftrightarrow$ $\lambda I - A$与$\lambda I - B$等价.
(3) 矩阵A与B相似 $\Longleftrightarrow$ 特征矩阵$\lambda I - A$与$\lambda I - B$有相同的不变因子或有相同的初等因子.
(4) 任何矩阵A一定相似于一个 Jordan 标准形.

例 3.3.2. 将下列 λ-矩阵化为Smith 标准形:

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

解 $$A(\lambda) = \begin{pmatrix} 1-\lambda & 2\lambda-1 & \lambda \\ \lambda & \lambda^2 & -\lambda \\ 1+\lambda^2 & \lambda^2+\lambda-1 & -\lambda^2 \end{pmatrix} \to \begin{pmatrix} 1 & 2\lambda-1 & \lambda \\ 0 & \lambda^2 & -\lambda \\ 1 & \lambda^2+\lambda-1 & -\lambda^2 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 2\lambda-1 & \lambda \\ 0 & \lambda^2 & -\lambda \\ 0 & \lambda^2-\lambda & -\lambda^2-\lambda \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda^2 & \lambda \\ 0 & \lambda^2-\lambda & \lambda^2+\lambda \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & \lambda^2 \\ 0 & \lambda^2+\lambda & \lambda^2-\lambda \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & \lambda^2+\lambda & -\lambda^3-\lambda \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^3+\lambda \end{pmatrix}.$$

结合推论 3.3.1与定理 3.3.2, 如果能够得到得到矩阵A的所有不变因子或初等因子, 则可求得A的Jordan标准形.

例 3.3.3. 利用l-矩阵求下列矩阵的 Jordan 标准形:

$$A = \begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix}.$$

[50]Henry John Stephen Smith (1826-1883), 爱尔兰数学家.

解 首先求 $\lambda I - A$ 的初等因子.

$$\lambda I - A = \begin{pmatrix} \lambda-3 & 0 & -8 \\ -3 & \lambda+1 & -6 \\ 2 & 0 & \lambda+5 \end{pmatrix} \to \begin{pmatrix} \lambda-3 & 0 & -8 \\ -3 & \lambda+1 & -6 \\ -1 & \lambda+1 & \lambda-1 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & -\lambda-1 & -\lambda+1 \\ -3 & \lambda+1 & -6 \\ \lambda-3 & 0 & -8 \end{pmatrix} \to \begin{pmatrix} 1 & -\lambda-1 & -\lambda+1 \\ 0 & -2\lambda-2 & -3\lambda-3 \\ 0 & (\lambda+1)(\lambda-3) & (\lambda+1)(\lambda-5) \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda+1 & 0 \\ 0 & 0 & (\lambda+1)^2 \end{pmatrix}.$$

所以 $\lambda I - A$ 的初等因子是 $(\lambda+1)^2$, $\lambda+1$. 因而 A 的 Jordan 标准形为:

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

3.3.2 广义特征子空间

从线性变换的角度看, 矩阵A与其Jordan标准形J是同一个线性变换$\sigma : \mathbb{C}^n \longrightarrow \mathbb{C}^n$在不同基下的矩阵, 因此如果能求出$\mathbb{C}^n$的一组适当的基, 则可以使$\sigma$在该基下的矩阵恰好是$J$.

定义 3.3.1. 设λ是$A \in \mathbb{C}^{n\times n}$的特征值, 称零空间$N((A-\lambda I)^n)$是$A$的(属于特征值$\lambda$的) **广义特征子空间**或**根子空间**, 记为E_λ, 其中的非零向量称为A的(属于特征值λ的)**广义特征向量**.

例 3.3.4. Jordan块$J_n(\lambda)$的特征子空间$V_\lambda = \mathbb{C}e_1$, 但其广义特征子空间$E_\lambda = \mathbb{C}$.

例 3.3.5. 设λ是$A \in \mathbb{C}^{n\times n}$的特征值, 则广义特征子空间$E_\lambda$是所有方程$(A-\lambda I)^k x = 0$的解空间的并集.

定理3.3.3. 设$|\lambda I - A| = \prod\limits_{i=1}^{s}(\lambda-\lambda_i)^{n_i}$是$n$阶矩阵$A$的特征多项式, $\sigma(A) = \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, g_i为λ_i的几何重数. 则

(1) 广义特征子空间$E_{\lambda_i} = \{x \in \mathbb{C}^n \,|\, (A-\lambda I)^{n_i}x = 0\}$.

(2) $\dim_{\mathbb{C}} E_{\lambda_i} = n_i$, 从而$\mathbb{C}^n = \sum\limits_{\lambda\in\sigma(A)} \oplus E_\lambda$.

(3) 存在$\alpha_j \in E_{\lambda_i}, 1 \le j \le g_i$, 使得$\cup_{1\le j\le g_i}\{\alpha_j, (A-\lambda_i I)\alpha_j, \cdots, (A-\lambda_i I)^{m_j-1}\alpha_j\}$构成$E_{\lambda_i}$的一组基 (称为由诸向量$\alpha_j$生成的循环基), 从而$E_{\lambda_i}$是$A$的不变子空间.

(4) 由每个广义特征子空间的循环基构成的$\mathbb{C}^n$的基称为关于A的Jordan 基, 即

$$\mathbb{C}^n = \sum_{\lambda\in\sigma(A)} \oplus E_\lambda \tag{3.3.2}$$

(5) 线性变换$\sigma : x \mapsto Ax$在Jordan 基下的矩阵是其 Jordan 标准形, 即令P是由Jordan基的基向量构成的矩阵, 则$P^{-1}AP = J$.

该定理中的(2)也称为**"谱分解定理"**.

从本章第二节已得到的Jordan标准形定理的角度看, 定理 3.3.3几乎是显然的, 因为A的各广义特征子空间的维数与其Jordan标准形的相应广义特征子空间完全一致.

例 3.3.6. 求下列矩阵的 Jordan 标准形J与Jordan基:

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

解 易知$|\lambda I-A|=(\lambda-1)^4$, 故$A$的 Jordan 标准形$J$的对角元素均为1. 再计算$A-I$的幂零度为3, 故最大子块是3阶的, 因此

$$J=J_3(1)\oplus(1).$$

设$P^{-1}AP=J$. 令$P=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$, 则由$AP=PJ$得

$$A(\alpha_1,\alpha_2,\alpha_3,\alpha_4)=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)(J_3(1)\oplus(1))=(\alpha_1,\alpha_1+\alpha_2,\alpha_2+\alpha_3,\alpha_4).$$

于是得到四个方程组:

$$A\alpha_1=\alpha_1,\quad A\alpha_2=\alpha_1+\alpha_2,\quad A\alpha_3=\alpha_2+\alpha_3,\quad A\alpha_4=\alpha_4,$$

即

$$(I-A)\alpha_1=0,\quad (I-A)\alpha_2=-\alpha_1,\quad (I-A)\alpha_3=-\alpha_2,\quad (I-A)\alpha_4=0.$$

解之得$\alpha_1=(0,0,1,-1)^T$, $\alpha_2=(1,-2,0,0)^T$, $\alpha_3=(-1,3,0,0)^T$, $\alpha_4=(-1,2,1,0)^T$.

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

矩阵P的列向量即构成$\mathbb{C}^n$的一组关于A的Jordan基.

例 3.3.7. 求下列矩阵的 Jordan 标准形:

$$A=\begin{pmatrix}3&-1&0&0&0&0&0&0\\5&-3&-1&0&0&0&0&0\\-8&7&3&0&0&0&0&0\\8&0&2&-1&1&0&0&0\\5&-11&-6&0&0&1&0&0\\4&8&5&0&-1&-2&0&0\\6&13&8&-4&-1&-4&2&1\\14&-12&-4&-3&-1&1&0&2\end{pmatrix}.$$

解 容易看出A是一个分块下三角矩阵. 因此A的特征多项式为

$$\begin{aligned}|\lambda I-A|&=\begin{vmatrix}\lambda-3&1&0\\-5&\lambda+3&1\\8&-7&\lambda-3\end{vmatrix}\cdot\begin{vmatrix}\lambda+1&-1&0\\0&\lambda&-1\\0&1&\lambda+2\end{vmatrix}\cdot\begin{vmatrix}\lambda-2&-1\\0&\lambda-2\end{vmatrix}\\&=(\lambda-1)^3(\lambda+1)^3(\lambda-2)^2.\end{aligned}$$

注意3重特征值1均来自左上角的主子矩阵A_1(前三行前三列), 3重特征值-1均来自第4-6行与列构成的主子矩阵A_2, 而两重特征值2则来自后两行两列构成的主子矩阵A_3. 因此由分块Schur三角化定理知存在严格上三角矩阵N_i使得A相似于分块对角矩阵

$$B = (I_3 + N_1) \oplus (-I_3 + N_2) \oplus (2I_2 + N_3),$$

其中N_i相似于$A_i - \lambda_i I$. 由于$A_1 - I$的零度为1, 故N_1相似于J_3, 即对角元素为1的Jordan块只有1块. 类似地, $-A-I$与$A-2I$的零度都为1. 故N_2与N_3分别相似于J_3与J_2. 因此A的Jordan标准形为

$$J = J_3(1) \oplus J_3(-1) \oplus J_2(2).$$

Jordan基的计算只需求解相应的线性方程组, 请读者自行完成.

思考题

1. 两个矩阵的和与积的Jordan标准形是否等于它们的Jordan标准形的和与积?
2. 如果P与Q均为Jordan标准形中的变换矩阵, 它们之间有何关系?
3. 设$A \in M_n(\mathbb{Q})$. 是否存在可逆矩阵$P \in M_n(\mathbb{Q})$使得$P^{-1}AP$为Jordan标准形? 将$\mathbb{Q}$换成$\mathbb{R}$又如何?
4. 分块矩阵$\begin{pmatrix} A & B \\ B & A \end{pmatrix}$的Jordan标准形与$A, B$的Jordan标准形有何关系? 特征值有何联系? 特别讨论$A = 0$与$A = B$的情形.
5. 仿照幂零矩阵相似的判别准则(即推论3.2.2)给出两个同阶矩阵相似的判别准则. 是否能够判断该准则与幂零矩阵相似的判别准则哪个更有意义?

第四节 特征值估计

3.4.1 盖尔圆定理

前面已经指出, 高阶矩阵的特征值计算非常困难, 实际上精确计算高阶矩阵的特征值一般是不可能的(1825年左右, Abel[51]与Galois[52] 等人证明了5次及以上的代数方程无公式解). 不过, 众多实际问题往往只需要确定特征值是否满足一定精度, 比如线性系统的稳定性仅需要所有特征值的实部均小于0, 而矩阵级数$\sum\limits_{i=0}^{\infty} a_i A^i$的敛散性只需知道$A$的特征值的模是否均小于1即可(详见第五章). 因此更为实用且合理的办法是对特征值的精度作估计.

设$A = (a_{ij})$是n阶复数矩阵. 称集合

$$D_i(A) = \{x \in \mathbb{C} \mid |x - a_{ii}| \le \sum_{j \ne i} |a_{ij}|\} \quad (i = 1, 2, \cdots, n)$$

为矩阵A的第i个圆盘. 将$\sum_{j \ne i} |a_{ij}|$记为$R_i(A)$, 称为A的**去心绝对行和**, 则矩阵A的第i个圆盘又可写成

$$D_i(A) = \{x \in \mathbb{C} \mid |x - a_{ii}| \le R_i(A)\}.$$

一般地, 称A的所有圆盘的并形成的区域$\cup_{i=1}^n D_i(A)$为A的(关于行的) **Gerschgorin**[53] **区域**, 记为$G(A)$, 而将$G(A)$中的每一个圆盘称为**Gerschgorin 圆盘**或**盖尔圆盘**, 这些圆盘的边界称为Gerschgorin圆或盖尔圆.

---

[51] Niels Henrik (1802-1829), 著名挪威数学家.

[52] Évariste Galois(1811-1832), 天才的法国数学家, 群论及有限域的创始人.

[53] Semyon Aranovich Gershgorin(1901-1933), 前苏联(现白俄罗斯)数学家.

例 3.4.1. 求A的圆盘, 其中

$$A=\begin{pmatrix} 0 & 1 & -1 \\ 1 & 2 & 0 \\ \mathrm{i} & 1 & 2\mathrm{i} \end{pmatrix}.$$

解 $D_1(A)=\{x\,|\,|x|\le 2\},\ D_2(A)=\{x\,|\,|x-2|\le 1\},\ D_3(A)=\{x\,|\,|x-2\mathrm{i}|\le 2\}.$

例 3.4.2. 如果矩阵$A=(a_{ij})$的对角元素均为非负实数, 则A与其绝对值矩阵$|A|=(|a_{ij}|)$具有完全相同的盖尔圆盘和盖尔区域.

下面的定理称为**"盖尔圆盘定理"**(Gerschgorin, 1931 年).

定理 3.4.1. 设$A\in\mathbb{C}^{n\times n},\lambda\in\sigma(A)$, 则下列不等式之一成立:

$$|\lambda-a_{ii}|\le\sum_{j\ne i}^{n}|a_{ij}|,\ i=1,2,\cdots,n.$$

等价地, $\sigma(A)\subseteq G(A)$, 即A的每个特征值都落在A的某个圆盘之内.

证 设λ是A的一个特征值, $\alpha=(x_1,\cdots,x_n)^T$是它的一个特征向量. 则$A\alpha=\lambda\alpha$, 即

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=\lambda x_1, \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=\lambda x_2, \\ \qquad\vdots \\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n=\lambda x_n. \end{cases} \qquad (*)$$

令 $\max\{|x_1|,\cdots,|x_n|\}=|x_m|$, 则由$x\ne 0$知$|x_m|>0$. 将方程组(*)中的第$m$个方程改写为

$$(\lambda-a_{mm})x_m=\sum_{j\ne m}a_{mj}x_j.$$

两边取模得

$$|\lambda-a_{mm}||x_m|\le\sum_{j\ne m}|a_{mj}||x_j|\le|x_m|\sum_{j\ne m}|a_{mj}|,$$

所以

$$|\lambda-a_{mm}|\le\sum_{j\ne m}|a_{mj}|.$$

即特征值λ落在第m个圆盘内. □

注 对A^T应用圆盘定理, 可以得到A的关于列的相应结果.

例 3.4.3. 求A的圆盘并估计其特征值, 其中

$$A=\begin{pmatrix} 10 & 1 & 2 & 3 \\ 5 & 30 & 1 & 2 \\ 10 & 3 & -10 & 5 \\ 2 & 3 & 1 & -40 \end{pmatrix}.$$

解　矩阵A的圆盘为

$$\begin{aligned}
|\lambda-10| &\leq 1+2+3 &= 6,\\
|\lambda-30| &\leq 1+5+2 &= 8,\\
|\lambda+10| &\leq 10+3+5 &= 18,\\
|\lambda+40| &\leq 2+3+1 &= 6.
\end{aligned}$$

如下面的示意图:

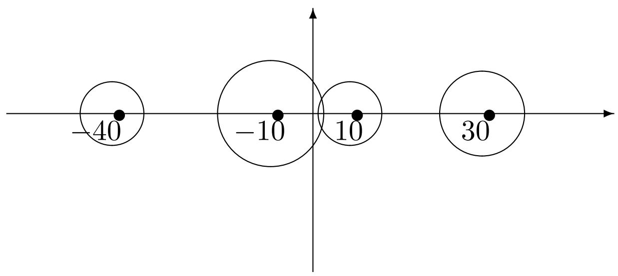

图3.4.1

因此, 四个特征值均落在四个圆盘之中.

例 3.4.4.　设n阶矩阵A满足对角强优条件

$$|a_{ii}| > \sum_{j\neq i} |a_{ij}|,\ i=1,2,\cdots,n,$$

证明: $|A|\neq 0$.

证　本结论在第一章中是利用$Ax=0$只有平凡解证明的, 现在只需证明0不是A的特征值即可. 题设条件说明A的每个圆盘的半径均小于圆心到原点的距离, 因此A的盖尔区域不包含原点, 从而由圆盘定理知A的特征值均非零, 从而A的行列式不等于0. □

圆盘定理说明每个特征值必然落在一个圆盘内, 但并未指明落在哪个圆盘内. 因此, 有些圆盘可能不含特征值, 而另一些则可能包含多个特征值.

例 3.4.5.　求A的圆盘, 其中

$$A=\begin{pmatrix} 10 & -8 \\ 5 & 0 \end{pmatrix}.$$

解　直接计算可知, 矩阵A的特征值为

$$\lambda_1 = 5+\sqrt{15}\mathrm{i},\ \lambda_2 = 5-\sqrt{15}\mathrm{i}.$$

而A的圆盘为

$$|\lambda-10|\leq 8,\quad |\lambda|\leq 5.$$

如下图:

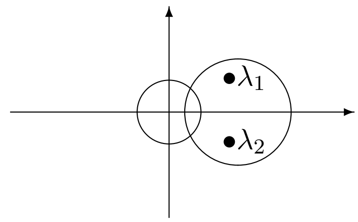

图3.4.2

因此, 两个特征值全部落在圆盘$|\lambda-10|\le 8$内, 而在圆盘$|\lambda|\le 5$之外!

实际上有下述**"精细圆盘定理"**.

命题 3.4.1. 设C是盖尔区域的一个由k个圆盘组成的连通分支, 则C恰好含有k个特征值.

证 盖尔区域的连通分支是指其最大的连通部分. 以下的证明基于一个基本事实, 即复系数多项式的根是其系数的连续函数. 记$A=(a_{ij})$的对角元素构成的对角矩阵为D, 令$B=A-D$. 考察矩阵$A(\varepsilon)=D+\varepsilon B$, 其中$0\le\varepsilon\le 1$. 显然$A(0)=D$, $A(1)=A$. 由于$A(\varepsilon)$的特征多项式的系数是ε的多项式, 因此$A(\varepsilon)$的特征值是ε的连续函数. 由圆盘定理, 对任意ε, 矩阵$A(\varepsilon)$的特征值均位于以a_{ii}为圆心, 半径为$\varepsilon R_i(A)$的圆盘之内. 当ε从0连续地变到1时, 特征值也连续地变化.

不妨设连通分支C由前k个圆盘组成, 注意它与A的其它$n-k$个圆盘是分离的. 因此对于每个$\varepsilon\in[0,1]$, 矩阵$A(\varepsilon)$的盖尔区域也具有相同的性质. 但当$\varepsilon=0$时, 特征值是$a_{11},a_{22},\cdots,a_{nn}$, 且前$k$个落在相应于前$k$个圆盘的区域中, 而其余$n-k$个特征值则落在该区域之外. 因此, 对所有的$\varepsilon\in[0,1]$, 该结论依然成立. 特别, 当$\varepsilon=1$时, 结论也成立. $\square$

比如在例3.4.3中, A的圆盘共有4个, 它们共构成3个连通部分(见例3.4.3的图): (1) $|\lambda+40|\le 6$; (2) $|\lambda+10|\le 18$; $|\lambda-10|\le 6$; (3) $|\lambda-30|\le 8$. 因此可以断言第二个连通部分含有两个特征值, 而在第一、三个由单个圆盘组成的连通部分各含有一个特征值. 进一步, 由于A是实数矩阵, 其特征多项式的复数根两两共轭, 而第一、三个圆盘是自共轭的(即属于该圆盘的复数的共轭仍属于该圆盘), 因此它们所包含的唯一的特征值必然是实特征值.

推论 3.4.1. 设n阶矩阵A的主对角线元素均为实数, A的特征多项式是实系数多项式. 若A的每个圆盘均与其余圆盘分离, 则A的特征值均为实数.

一般来说, 直接使用圆盘定理得到的特征值估计往往较为粗糙. 比如矩阵

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

的两个盖尔圆盘分别为$|x-1|\le 4$与$x=2$, 利用圆盘定理只能知道A的两个特征值都落在圆盘$|x-1|\le 4$中, 这个估计当然太差了. 为了得到更为有效的估计, 需要将每个圆盘的半径适当缩小, 以使每个圆盘只包含A的一个特征值. 由于相似矩阵的特征值相同, 故可以先对矩阵A施行适当的相似变换, 再应用圆盘定理. 一般选择对角矩阵作相似变换, 即令$P=D=\operatorname{diag}(d_1,d_2,\cdots,d_n)$, 其中$d_i>0$. 于是$B=D^{-1}AD=(a_{ij}\frac{d_j}{d_i})$. 显然$B$的对角元素与$A$的对角元素完全相同, 因此这样的相似变换(称为对角相似变换) 只改变盖尔圆盘的半径而

不改变它们的中心. (d_j的选取原则是: 欲使第j个圆盘缩小, 可取$d_j > 1$, 而其余的d_k取为1, 此时B的其余圆盘相对放大; 反之, 欲使第j个圆盘放大, 可取$d_j < 1$, 而其余的d_k取为1, 此时B的其余圆盘相对缩小.) 利用这个技巧, 再研究前面的矩阵A, 则得到

$$B = \begin{pmatrix} d_1^{-1} & 0 \\ 0 & d_2^{-1} \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} = \begin{pmatrix} 1 & \dfrac{4d_2}{d_1} \\ 0 & 2 \end{pmatrix},$$

因此B的两个盖尔圆盘分别为$|x-1| \le \dfrac{4d_2}{d_1}$与$x = 2$, 由于可以将数$\dfrac{4d_2}{d_1}$ 取为任意正实数, 因此两个特征值必然为1与2.

例 3.4.6. 估计矩阵A的特征值, 其中

$$A = \begin{pmatrix} 0.9 & 0.01 & 0.11 \\ 0.01 & 0.8 & 0.12 \\ 0.01 & 0.02 & 0.4 \end{pmatrix}.$$

解 由圆盘定理可知A的特征值在下列圆盘之中: $|\lambda - 0.9| \le 0.12$; $|\lambda - 0.8| \le 0.13$; $|\lambda - 0.4| \le 0.03$. 所以第一、二个圆盘构成一个连通部分, 而第三个圆盘单独构成一个连通部分. 这样, 有两个特征值不能分离. 但若作相似变换, $D^{-1}AD = B$, 其中D是对角矩阵$\mathrm{diag}\,(1, 1, \frac{1}{10})$, 则

$$B = \begin{pmatrix} 0.9 & 0.01 & 0.011 \\ 0.01 & 0.8 & 0.012 \\ 0.1 & 0.2 & 0.4 \end{pmatrix}.$$

故B的圆盘为

$$|\lambda - 0.9| \le 0.021; \quad |\lambda - 0.8| \le 0.022; \quad |\lambda - 0.4| \le 0.3.$$

从而B的三个圆盘都是孤立的, 因而每个圆盘中都有一个特征值. 由于A与B相似, 从而有相同的特征值, 故A的特征值分别在上述三个圆盘中, 且按推论 3.4.1, 它们都是实数.

例 3.4.7. 证明矩阵A至少有两个实特征值, 其中

$$A = \begin{pmatrix} 7 & 1 & 2 & -1 \\ 0 & 8 & 1 & 1 \\ -1 & 0 & 5 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

证 A的四个圆盘为

$$D_1 = |\lambda - 7| \le 4; \quad D_2 = |\lambda - 8| \le 2; \quad D_3 = |\lambda - 5| \le 1; \quad D_4 = |\lambda - 1| \le 1.$$

直接计算可知, 四个圆盘构成两个连通部分, 分别为$G_1 = D_1 \bigcup D_2 \bigcup D_3$与$G_2 = D_4$. 因此$G_2$包含唯一的特征值, 该特征值只能与自己共轭, 故为实数. 因此, 含在G_1中的三个特征值必有一个是实数. 从而A至少有两个实特征值.

利用圆盘定理可以得到矩阵谱半径的如下估计.

命题 3.4.2. 设$A = (a_{ij})$是n阶复数矩阵. 令$\nu = \max\limits_{1 \le k \le n} \sum_{j=1}^{n} |a_{kj}|$ (此即矩阵行的元素的绝对值之和的最大者), 则$\rho(A) \le \nu$.

证 设λ_0是A的一个特征值, 由圆盘定理可知, 一定存在k使得$|\lambda_0 - a_{kk}| \le \sum_{j=1,j\neq k}^n |a_{kj}|$成立. 所以$|\lambda_0| \le |a_{kk}| + \sum_{j=1,j\neq k}^n |a_{kj}| = \sum_{j=1}^n |a_{kj}| \le \nu$. 因此, 任何特征值的模都不超过$\nu$, 所以$\rho(A) \le \nu$. □

注意A与A^T有相同的特征值, 若记矩阵A的列的绝对值之和的最大值为ν', 即$\nu' = \max\limits_{1\le j\le n} \sum_{k=1}^n |a_{kj}|$, 对$A^T$使用上述命题可得

命题 3.4.3. $\rho(A) \le \min\{\nu, \nu'\}$.

盖尔圆盘定理有众多新的拓展方法, 以下介绍其中两种(均略去证明).

类似于去心绝对行和, 定义矩阵$A = (a_{ij})$的去心绝对列和如下:

$$C_i(A) = \sum_{j\neq i} |a_{ji}|,\ 1 \le\ i \le n.$$

定理 3.4.2. (Ostrowski[54] 圆盘定理) 设$\lambda \in \sigma(A)$, $\alpha \in [0,1]$. 则存在$1 \le i \le n$使得

$$|\lambda - a_{ii}| \le R_i(A)^\alpha C_i(A)^{1-\alpha}.$$

盖尔圆盘定理是Ostrowski 圆盘定理当$\alpha = 1$与0时的特例, 读者可以尝试其它类型的推广, 比如可以进一步证明精细的Ostrowski 圆盘定理.

例 3.4.8. 矩阵$A = \begin{pmatrix} 1 & 4 \\ 1 & 6 \end{pmatrix}$的两个盖尔圆盘$|x-1| \le 4$与$|x-6| \le 1$相切, 故由盖尔圆盘定理得出的估计较为粗糙. 利用Ostrowski 圆盘定理可取$\alpha = \frac{1}{2}$, 则得两个分离的圆盘分别为$|x-1| \le 2$与$|x-6| \le 2$, 因此由精细的Ostrowski 圆盘定理可知这两个圆盘各包含A的一个特征值, 故一个实特征值位于-1与3之间, 而另一个实特征值位于 -1 与 7 之间.

定理 3.4.3. (Brauer[55] 定理) 设$\lambda \in \sigma(A)$, 则存在$1 \le i \neq j \le n$使得

$$|\lambda - a_{ii}||\lambda - a_{jj}| \le R_i(A)R_j(A).$$

(上式表达的几何图形称为Cassini[56] 卵形.)

定理3.4.3在下述意义下是盖尔圆盘定理的最佳推广, 即它不能进一步推广至三个或更多盖尔圆盘, 请读者自行验证.

3.4.2 特征值的极值与不等式

本节讨论Hermite矩阵的特征值. 设A是,Hermite 矩阵, 则其特征值都是实数, 故可将其特征值从大到小排为$\lambda_1(A) \ge ... \ge \lambda_n(A)$.

定理 3.4.4. 设A是 Hermite 矩阵, 其全部特征值为$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$, 则

$$\lambda_1 = \max_{0\neq x\in\mathbb{C}^n} \frac{x^*Ax}{x^*x} = \max_{x^*x=1} x^*Ax, \quad \lambda_n = \min_{0\neq x\in\mathbb{C}^n} \frac{x^*Ax}{x^*x} = \min_{x^*x=1} x^*Ax.$$

[54]Alexander Markowich (1893-1986), 著名俄裔德国瑞士数学家.

[55]Alfred Brauer(1894-1985), 德裔美籍数学家, 是著名德裔美籍数学家 Richard Brauer 的哥哥.

[56]Giovanni Domenico Cassini(1625-1712), 意大利法国天文学家, 地心说的代表之一, Cassini 卵形即是其描述太阳绕地球旋转的轨迹.

证 由$A^* = A$知存在$\mathbb{C}^n$的一组由A的特征向量$\alpha_1, ..., \alpha_n$组成的标准正交基, 且不妨设$A\alpha_k = \lambda_k\alpha_k, \forall k$. $\forall x \in \mathbb{C}^n$, 设$x = \sum_{k=1}^{n} x_k\alpha_k$, 则

$$x^*Ax = (\sum_{k=1}^{n} x_k\alpha_k)^*(A\sum_{k=1}^{n} x_k\lambda_k\alpha_k) = \sum_{k=1}^{n} \bar{x}_k x_k\lambda_k(\alpha_k)^*\alpha_k = \sum_{k=1}^{n} \lambda_k|x_k|^2.$$

由于$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 故

$$\lambda_1 x^*x = \lambda_1\sum_{k=1}^{n}|x_k|^2 \geq x^*Ax \geq \lambda_n\sum_{k=1}^{n}|x_k|^2 = \lambda_n x^*x.$$

所以

$$\lambda_n = \min_{0\neq x\in\mathbb{C}^n}\frac{x^*Ax}{x^*x} = \min_{x^*x=1} x^*Ax, \quad \lambda_1 = \max_{\|x\|=1}\frac{x^*Ax}{x^*x} = \max_{x^*x=1} x^*Ax.$$

□

下面精彩的定理称为"**Weyl**[57] **不等式**"或者"**单调定理**".

定理 3.4.5. 设A, B均为Hermite矩阵, 且B是半正定矩阵, 则

$$\lambda_j(A) + \lambda_j(B) \leq \lambda_{j+k+1}(A+B) \tag{3.4.1}$$

证 以下证明取自[58].

首先注意, 如果$S_1, ..., S_k$是n-维线性空间V的子空间, 则

$$\dim(S_1 + ... + S_k) > n(k-1) \Rightarrow \dim(\bigcap_{i=1}^{k} S_i) > 0 \tag{3.4.2}$$

其次, 设$H_1, ..., H_k$均为n-阶Hermite矩阵, 且$H_1 + ... + H_k$是半负定的, 则对所有$1 \leq p_1, ..., p_k \leq n$使得$p_1 + ... + p_k > n(k-1)$, 均有

$$\lambda_{p_1}(H_1) + ... + \lambda_{p_k}(H_k) \leq 0 \tag{3.4.3}$$

这是因为, 设S_i是由H_i的属于特征值$\lambda_1, \lambda_2, ..., \lambda_{p_i}$的特征向量张成的子空间, 则

$$\sum_{i=1}^{k}\dim(S_i) = \sum_{i=1}^{k}(n - p_i + 1) = nk + k - (p_1 + ... + p_k) > n(k-1).$$

故由(3.4.2)知存在单位向量$\alpha \in S_i$.

注意, $\lambda_{p_i}(H_i)$是H_i在S_i中的最小特征值, 所以$\lambda_{p_i}(H_i) \leq \alpha^*H_i\alpha, 1 \leq i \leq k$. 由于$H_1 + ... + H_k$是半负定的, 故

$$\sum_{i=1}^{k}\lambda_{p_i}(H_i) \leq \sum_{i=1}^{k}\alpha^*H_i\alpha = \alpha^*(\sum_{i=1}^{k}H_i)\alpha \leq 0.$$

---

[57]H. Weyl(1885-1955), 著名德国数学家、物理学家、哲学家, 被誉为"闯入物理学世界的数学大象".

[58]Steve Fisk, A note on Weyl's inequality, Amer. Math. Monthly 104 (1997), no. 3, 257–258.

现取$H_1 = A, H_2 = B, H_3 = -(A+B), p_1 = j, p_2 = k, p_3 = j+k+1$, 并注意对任意Hermite矩阵$C$均有$\lambda_s(-C) = -\lambda_{n+1-s}(C)$，即可得到Weyl 不等式(3.4.1).

下面将证明Hermite矩阵特征值的著名结果, 即“Courant-Fischer 极大-极小-定理”. 为此, 将特征值由小到大排列较为方便, 故采用记号

$$\lambda_1^A \le \lambda_2^A \le ... \le \lambda_n^A.$$

Courant-Fischer 极大-极小-定理. 对$1 \le k \le n$, 有

$$\lambda_k^A = \min_{S^k} \max\{x^*Ax | x \in S^k, x^*x = 1\} = \max_{S^{k-1}} \min\{x^*Ax | x \perp S^{k-1}, x^*x = 1\}.$$

证 下述证明取自[59].

设$A\alpha_i = \lambda_i^A \alpha_i, 1 \le i \le n; B\beta_j = \lambda_j^B \beta_j, 1 \le j \le m, \alpha_i^*\alpha_j = \beta_i\beta_j = \delta_{ij}$.

记$\gamma_i = \begin{pmatrix} \alpha_i \\ 0 \end{pmatrix} \in \mathbb{F}^n, 1 \le i \le m$.

取$S_1 = \mathrm{Span}\{\alpha_k, ..., \alpha_n\}, S_2 = \mathrm{Span}\{\gamma_1, ..., \gamma_k\}$. 则$\dim(S_1) = n-k+1, \dim(S_2) = k$, 故存在$x \in S_1 \bigcap S_2, x^*x = 1$. 因此

$$\lambda_k^A \le x^*Ax \le \lambda_k^B \tag{3.4.4}$$

对矩阵$-A$使用上式即可得$\lambda_k^B \le \lambda_{k+n-m}^A$, 所以Cauchy 交错定理成立.

为证Courant-Fischer 极大-极小-定理, 只需将上面的S_2改为任何k-维子空间S^k, 则仍有$x \in S_1 \bigcap S_2, x^*x = 1, x^*Ax \ge \lambda_k^A$.

另一方面, 对任何单位向量$\alpha \in \mathrm{Span}\{\alpha_1, ..., \alpha_k\}$, 均有$\alpha^*A\alpha \le \lambda_k^A$以及$\alpha_k^*A\alpha^k = \lambda_k^A$, 故Courant-Fischer 极大-极小-定理中的第一个等号成立.

为证第二个等式, 取$S_1 = \mathrm{Span}\{\alpha_1, ..., \alpha_k\}, S_2 = (S^{k-1})^\perp$即可. □

思考题

1. 设A是二阶实对称矩阵, 试用微分学研究函数$\dfrac{x^*Ax}{x^*x}$的极值, 并解释最大特征值与最小特征值的意义.
2. 设A是二阶实对称矩阵, 试用微分学研究矩阵A的第二大特征值的意义.

第五节 应用

矩阵计算的实质是特征值的计算, 尽管Jordan标准形不是矩阵计算的实用方法, 但它是理解矩阵性质、计算矩阵函数、研究矩阵微积分的最为重要的理论工具, 从而具有广泛而深刻的应用. 本节讨论两个重要的应用, 即**随机矩阵**和**PageRank算法**.

3.5.1 随机矩阵

定义 3.5.1. 矩阵$A \in \mathbb{R}^{n\times n}$称为**非负矩阵**如果其元素均非负. 一个非负矩阵称为“**行随机矩阵**”如果其所有行和均为1. 类似地, 非负矩阵称为“**列随机矩阵**”如果其所有列和均为1. 既是行随机矩阵又是列随机矩阵就称为“**双随机矩阵**”或简称为随机矩阵.

[59] YASUHIKO IKEBE, TOSHIYUKI INAGAKI, and SADAAKI MIYAMOTO, The Monotonicity Theorem, Cauchy's Interlace Theorem, and the Courant-Fischer Theorem, The American Mathematical Monthly V.94(1987)4, 352-354.

单位矩阵是最简单的随机矩阵. 分量之和为1的非负列(行)向量是列(行)随机向量.

命题 3.5.1. 设A, B是两个n-阶双随机矩阵, 1_n是全1向量, $J = 1_n 1_n^T$是全1矩阵. 则
(1) 1是A的特征值, 1_n是1的特征向量且$\rho(A) = 1$.
(2) 设$0 \le \alpha \le 1$, 则$\alpha A + (1-\alpha)B$是双随机矩阵.
(3) $AJ = JA = J$.
(4) A是$m \le (n-1)^2 + 1$个置换矩阵的凸线性组合, 即存在$\alpha_i \ge 0, \sum_{i=1}^m \alpha_i = 1$使得

$$A = \sum_{i=1}^m \alpha_i P_i \tag{3.5.1}$$

其中P_i均为置换矩阵. (此条称为**Birkhoff**[60]**-von Neumann**[61]**定理.**)

证 第一条是圆盘定理的直接推论; 第二、三条由随机矩阵的定义可直接验证; 现证第四条Birkhoff-von Neumann定理, 关键是将$\mathbb{R}^{n\times n}$视为$\mathbb{R}^{n^2}$(同构的观点), 这样所有双随机矩阵构成的集合$\mathfrak{S}$就是满足约束条件

$$\sum_{j=1}^n x_{ij} = 1, \quad \sum_{i=1}^n x_{ij} = 1, \quad 0 \le x_{ij} \le 1 \tag{3.5.2}$$

的所有点$X = (x_{ij})$构成的$\mathbb{R}^{n^2}$中的有界凸多面体.

以下证明有界凸多面体$\mathfrak{S}$的顶点(extreme point, 即非$\mathfrak{S}$中任何线段的中点)坐标均为整数(从而均为0或1), 故顶点都是置换矩阵. 等价地, 只需证明所有非置换矩阵的双随机矩阵X都是$\mathfrak{S}$中某线段的中点即可(如此X将不能成为顶点). 非置换矩阵的双随机矩阵X的特征是存在元素$x_{a_1b_1}$满足条件$0 < x_{a_1b_1} < 1$. 由约束条件(3.5.2), 可知存在下标b_2使得$x_{a_1b_2}$满足条件$0 < x_{a_1b_2} < 1$. 同理, 对$x_{a_1b_2}$存在下标a_2使得$x_{a_2b_2}$满足条件$0 < x_{a_2b_2} < 1$. 如此继续, 至到出现重复下标则停止. 现假定第一次出现重复下标时的步数最少, 则再由约束条件(3.5.2)可知, 第一个重复下标必然是a_1b_1.

断言第一个重复下标必然是$a_kb_k(= a_1b_1)$, 即第一个重复下标出现时的步数必为偶数步. 否则, $a_kb_{k+1} = a_1b_1$, 则$a_kb_{k+1}, a_1b_1, a_1b_2$均在同一行(第$a_1$行), 于是删除$a_1b_2$而从$a_2b_2$开始可以得到更少的步数, 于前面的假设矛盾.

现令$\varepsilon_0 = \min\{x_{a_jb_j}, 1 - x_{a_jb_j}\}, \varepsilon = 0.5\varepsilon_0$. 构造矩阵$X^+(\varepsilon)$使得其第$x_{a_jb_j}$项增加$\varepsilon$, 同时其第$x_{a_jb_{j+}}$项减少$\varepsilon$, 如此可以保证$X^+(\varepsilon) \in \mathfrak{S}$. 类似地, 构造矩阵$X^-(\varepsilon)$使得其第$x_{a_jb_j}$项减少$\varepsilon$, 同时其第$x_{a_jb_{j+}}$项增加$\varepsilon$, 自然也有$X^-(\varepsilon) \in \mathfrak{S}$. 显然, 以$X^+(\varepsilon)$和$X^-(\varepsilon)$为端点的线段完整落在$\mathfrak{S}$中.

现在$X = \dfrac{X^+(\varepsilon) + X^-(\varepsilon)}{2}$, 故$X$是$\mathfrak{S}$中线段的中点, 故必不是顶点. 从而$\mathfrak{S}$的顶点均为置换矩阵. 因此任何双随机矩阵必然是置换矩阵的凸线性组合.

最后, 由于约束条件(3.5.2)中的线性方程共有$2n$个, 故相应的齐次方程的解空间的维数最多为$n^2 - 2n$, 因此非齐次方程的解张成的子空间的维数不超过$n^2 - 2n + 1 = (n-1)^2$, 因此$m \le (n-1)^2 + 1$个. □

[60]G. Birkhoff(1911-1996), 著名美国数学家.
[61]J. von Neumann(1903 1957), 著名匈牙利-美国数学家, 计算机之父.

基于命题3.5.1中的优良性质, 双随机矩阵甚至在量子计算中已经得到了深刻应用, 比如读者可参考文献[62]. 以下简介行随机矩阵在2012年诺贝尔经济学奖主题**"稳定婚姻"**中的应用.

设有N男N女适婚人群, 在一夫一妻制的前提下, 每个人均须对每个异性严格排序, 即将自己最钟情的异性排在第一位, 次钟情的异性排在第二位, 如此等等. 自然希望最后的结局满足下列要求:

SM1. 每个人都有配偶.

SM2. 所有N对匹配构成稳定婚姻, 即任何非夫妇的一对异性不会出现"私奔"现象.

何时会出现"私奔"现象? 如果男子M_1对妻子W_1的排名在女子W_2之后, 同时女子W_2对丈夫M_2的排名在男子M_1之后, 则男子M_1和女子W_2将"必然"私奔.

所以稳定婚姻的目标是在人人有配偶的前提下杜绝私奔现象.

问题是: 稳定婚姻是否存在呢? 请研究下面的例子[63].

例 3.5.1. 设有3男A、B、C, 3女1、2、3, 他们各自对异性的钟情程度如下表(钟情的程度按顺序严格递减):

A(1, 2, 3)	······	1(C, A, B)
B(3, 1, 2)	······	2(B, C, A)
C(2, 3, 1)	······	3(A, B, C)

试给出最好的婚姻匹配.

显然(A, 1), (B, 3), (C, 2)就是一个稳定婚姻. 读者不难发现其它的稳定婚姻匹配.

Gale[64]-Shapley[65]定理. 稳定婚姻必定存在.

代替证明, 我们特别指出稳定婚姻问题可以转化为行随机矩阵. 为此, 首先将N个男子对N个女子的排序赋予开区间(0, 1)中的一个数字. 确切地说, 以$s_{ij} \in (0,1)$表示男子i对女子j的钟情程度, $s_{ij} > s_{ik}$等价于男子i对女子j的钟情程度高于女子k. 按一夫一妻制的要求可以对s_{ij}进一步赋予下述约束条件:

$$\sum_{j=1}^{N} s_{ij} = 1, 0 < s_{ij} < 1, s_{ij} - s_{ik} = 0 \iff j = k \tag{3.5.3}$$

容易看出, S是一个行随机矩阵, 从而稳定婚姻等价于研究$N \times N$的行随机矩阵$S = (s_{ij})$.

3.5.2 PageRank算法

Google曾经独步天下的核心技术PageRank算法大约是使用数学理论最简单的著名算法了, 其数学模型基于两条基本常识:

流量常识. 入链越多, 网页越重要.

[62] Mariella, Nicola; Zhuk, Sergiy, A doubly stochastic matrices-based approach to optimal qubit routing, Quantum Inf. Process. 22 (2023), no. 7, Paper No. 264.

[63] 请参考张跃辉, 李吉有, 朱佳俊著《数学的天空》, 原文见D. Gale and L. S. Shapley: "College Admissions and the Stability of Marriage", American Mathematical Monthly 69, 9-14, 1962.

[64] D. Gale(1921-2008), 美国数学家、工程师.

[65] L. S. Shapley(1923-), 美国数学家、经济学家, 2012年诺贝尔经济学奖得主.

流向常识. 重要网页指向重要网页.

在Brin和Page的原始论文[66] 中, 该算法只有一行公式

$$PR(A) = (1-d) + d\left(\frac{PR(T1)}{C(T1)} + \cdots + \frac{PR(Tn)}{C(Tn)}\right) \tag{3.5.4}$$

其中A是一个网页, $T1, ..., Tn$是指向A的网页, 参数$d \in (0,1)$是阻尼系数(damping factor)或衰减系数, 在文中取为0.85, $C(A)$ 是从A发出的链接总数, $PR(A)$就是A的PageRank, 所有网页的PageRank之和为1. 任何网页的PangeRank值介于0与1之间, 其值越大, 该网页越重要, 搜索引擎中的排序越靠前.

PageRank算法与初始值无关、与搜索主题无关,因此有普适的数学模型,称为**"PageRank问题"**.

定义 3.5.2. 设P列随机矩阵, v是列随机向量, $0 < \alpha < 1$是参数. 线性方程组

$$(I - \alpha P)x = (1-\alpha)v \tag{3.5.5}$$

称为PageRank 方程组, 其解称为PageRank 向量, 其求解问题称为PageRank 问题.

下面的定理说明PageRank 向量的重要性, 即求解PageRank 问题的算法总有解且收敛于PageRank 向量.

定理 3.5.1. 设P列随机矩阵, v是列随机向量, e是全1向量, $0 < \alpha < 1$. 则PageRank 问题(3.5.5)存在唯一解, 且PageRank 向量x满足方程

$$(\alpha P + (1-\alpha)ve^T)x = x \tag{3.5.6}$$

证 由于P是列随机矩阵, $0 < \alpha < 1$, 故PageRank 问题(3.5.5)的系数矩阵$(I - \alpha P)$的特征值均非0, 因此可逆, 故PageRank 问题存在唯一解. 由命题3.5.1知(3.5.6)中的系数矩阵$\alpha P + (1-\alpha)ve^T$仍然是列随机矩阵, 故1是其特征值, 于是取$x$是属于特征值1的满足条件$e^T v = 1$的一个特征向量, 则容易验证$x$就是PageRank 问题(3.5.5)的唯一解. □

习 题 三

1. 详细证明分块Schur三角化定理.

2. (1) 证明不等于零的幂零矩阵一定不相似于对角矩阵.
(2) 设A具有唯一特征值但A不是对角矩阵. 证明A一定不相似于对角矩阵.

3. 设$\alpha \neq \beta$, 试求矩阵

$$\begin{pmatrix} \alpha & a & b & c \\ 0 & \alpha & d & e \\ 0 & 0 & \beta & x \\ 0 & 0 & 0 & \beta \end{pmatrix}$$

的相似分块对角矩阵.

[66]Sergey Brin, Lawrence Page, The anatomy of a large-scaleh ypertextual Web searche ngine, Computer Networks and ISDN Systems 30 (1998) 107- 117. Sergey Brin(1973-), 美籍俄裔企业家, Google公司联合创始人之一. Lawrence Page(1973-), 美国企业家, Google公司联合创始人之一.

4. (1) 举例说明 Schur 三角化定理在实数域上不成立.

(2) 证明实数域上的 Schur 三角化定理: 设$A \in M_n(\mathbb{R})$, 则存在正交矩阵Q使得

$$Q^T AQ = \begin{pmatrix} A_1 & & * & \\ & A_2 & & \\ & & \ddots & \\ & 0 & & A_k \end{pmatrix}$$

其中每个A_i或者是1阶实矩阵或者是形如$\begin{pmatrix} a_i & b_i \\ -b_i & a_i \end{pmatrix}$的2阶实矩阵($b_i \neq 0$).

(提示: 首先,如果λ是A的非实数特征值, $Ax = \lambda x$, 则$A\bar{x} = \bar{\lambda}\bar{x}$, 由此可知$x$与$\bar{x}$线性无关, 进而Re$x$与Im$x$线性无关, 将其正交化后构造正交矩阵, 再利用归纳法.)

5. 不用 Jordan 标准形定理, 直接计算A的 Jordan 标准形, 其中.

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

6. 证明命题 3.1.1, 即任何矩阵均可通过一系列Householder矩阵而相似于Hessenberg矩阵.

7. 证明Hermite矩阵均酉相似于三对角Hermite矩阵. 讨论相应的矩阵分解的唯一性.

8. (1) 证明幂零矩阵的相似准则: 设M与N是两个n阶幂零矩阵. 则M与N相似 $\iff r(M^k) = r(N^k), \forall k \geq 1$.

(2) A是幂零矩阵当且仅当$tr(A^k) = 0, \forall k \geq 1$.

9. 已知$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, 线性空间$V = \{X = (x_{ij}) \in M_2(\mathbb{R}) \,|\, \mathrm{tr}\, X = 0\}$的线性变换$\sigma$为$\sigma(X) = B^T X - X^T B$, $X \in V$. 试求V的一个基, 使得σ在该基下的矩阵尽可能简单.

10. 求下列矩阵的 Jordan 标准形与变换矩阵P使$P^{-1}AP = J$:

$$(1) \begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}; \quad (2) \begin{pmatrix} 2 & 6 & -15 \\ 1 & 1 & -5 \\ 1 & 2 & -6 \end{pmatrix}; \quad (3) \begin{pmatrix} 9 & -6 & -2 \\ 18 & -12 & -3 \\ 18 & -9 & -6 \end{pmatrix}$$

$$(4) \begin{pmatrix} -4 & 2 & 10 \\ -4 & 3 & 7 \\ -3 & 1 & 7 \end{pmatrix}; \quad (5) \begin{pmatrix} 3 & 2 & 1 \\ 0 & 4 & 0 \\ -1 & 2 & 5 \end{pmatrix}; \quad (6) \begin{pmatrix} 2 & 2 & 2 & 1 \\ -1 & -1 & -3 & -2 \\ 1 & 2 & 5 & 3 \\ -1 & -2 & -4 & -2 \end{pmatrix}.$$

11. 设$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$.

(1) 求A与A^k的 Jordan 标准形.

(2) A的Jordan-Chevalley 分解是什么?

(3) 求$\begin{pmatrix} A & A \\ A & A \end{pmatrix}$的的 Jordan 标准形.

12. 设$A \in \mathbb{F}^{n \times n}$, 集合$A = \{X \in \mathbb{F}^{n \times n} | AX = XA\}$称为$A$的交换子.

(1) 证明$C(A)$是$\mathbb{F}^{n \times n}$的子空间.

(2) 设$A = J_n(\lambda)$, 计算$C(A)$的维数.

13. 利用矩阵的 Jordan 标准形定理证明 **Fitting**[67] **引理**(对照第二章公式(2.6.6)):
设V为n维线性空间, $\sigma \in \mathrm{End}V$, 则$V = \mathrm{Im}(\sigma^n) \oplus \mathrm{Ker}(\sigma^n)$.

[67]Hans Fitting(1906-1938), 德国数学家.

14. 试求矩阵A的特征多项式与最小多项式相同的一个充分必要条件.

15. 证明 **Hadamard**[68] **不等式**: 对任意n阶矩阵$A=(a_{ij})$有$|A|\le\prod\limits_{i=1}^{n}(\sum\limits_{i=1}^{n}|a_{ij}|^2)^{1/2}$. 并由此证明:

(1) 若A是正定矩阵, 则$|A|\le\prod\limits_{i=1}^{n}a_{ii}$;.

indent (2) 设C是非负实数, 若$|a_{ij}|\le C, 1\le i,\ j\le n$, 则$|A|\le C^n n^{n/2}$.

(3) 设$a_{ij}=\pm1, 1\le i,\ j\le n$, 则由(2)可知$|A|\le n^{n/2}$. 如果等号成立, 则称$A$是一个 **Hadamard 矩阵**. 证明$A$是 **Hadamard 矩阵** $\iff$ $A^TA=nI_n$ $\iff$ A的列两两正交.

16. 设 $A\in\mathbb{C}^{n\times n}$. 证明华罗庚定理(1944年):

(1) 存在可逆对称矩阵S使得$AS=SA^T$.

(2) **对称矩阵定理**: 存在对称矩阵B,C使得$A=BC$, 即任何矩阵都是对称矩阵的乘积.

17. 证明Djokovic 定理(1967年)：设A可逆. 则A与A^{-1}相似当且仅当A是两个对合矩阵的乘积.

18. 设$p(\lambda)=(-1)^n[\lambda^n-a_{n-1}\lambda^{n-1}-a_{n-2}\lambda^{n-2}-\cdots-a_1\lambda-a_0]$. 称矩阵

$$C=\begin{pmatrix} a_{n-1} & a_{n-2} & \cdots & a_1 & a_0\\ 1 & 0 & \cdots & 0 & 0\\ 0 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & 1 & 0\end{pmatrix}$$

为多项式$p(\lambda)$的友矩阵. 设n阶矩阵A的特征多项式为$(-1)^np(\lambda)$.

(1) 计算C的特征多项式.

(2) 当$n=2$时, 证明：A与C相似当且仅当A的最小多项式等于其特征多项式.

(3) 试将(2)中的结论推广到一般情形. (提示: 查阅 Frobenius 标准形与有理标准形.)

19. 设a是复常数, $V=\{e^{ax}f(x)\,|\,f(x)\in\mathbb{C}_n[x]\}$是$n$维复线性空间.

(1) 证明求导运算$\partial:\alpha\mapsto\frac{\mathrm{d}\,\alpha}{\mathrm{d}\,x}$是$V$上的线性变换.

(2) 求∂在某组基下的矩阵的 Jordan 标准形.

20. (酉矩阵与离散 Fourier 变换) 设σ是$\mathbb{C}^n$的循环位移变换, 即$\sigma((x_1,x_2,\cdots,x_n)^T)=(x_2,x_3,\cdots,x_n,x_1)^T$. 证明:

(1) σ的特征值恰好为方程$\lambda^n=1$的所有根$\lambda_j=e^{\frac{2\pi\mathrm{i}}{n}j}, 1\le\ j\le\ n$.

(2) σ的属于特征值λ_j的特征向量为$\alpha_j=(\lambda_j,\lambda_j^2,\cdots,\lambda_j^n)^T$, 且$\|\alpha_j\|=\sqrt{n}$.

(3)$\alpha_1,\cdots,\alpha_n$是$\mathbb{C}^n$的一组正交基.

(4) 任何向量$x=(x_1,x_2,\cdots,x_n)^T$均是σ的特征向量α_j的线性组合$x=\sum\limits_{j=1}^{n}a_j\alpha_j$, 即$x_k=\sum\limits_{j=1}^{n}a_je^{\frac{2\pi\mathrm{i}}{n}j}$.

(5) 上面的系数$a_j=(x,\alpha_j)/n=\frac{1}{n}\sum\limits_{k=1}^{n}x_ke^{-\frac{2\pi\mathrm{i}}{n}jk}$;

(6) 研究σ与第一章习题 7 中的 Fourier 矩阵的关系, 并由此再求该矩阵的逆.

21. 设$A\in\mathbb{C}^{n\times n},\lambda\in\sigma(A)$. 证明广义特征子空间$E_\lambda$的维数等于$\lambda$的代数重数.

22. 求下列矩阵的盖尔圆盘并讨论其特征值的范围与性质.

(1) $\begin{pmatrix} 9 & 1 & 1\\ 1 & i & 1\\ 1 & 1 & 3\end{pmatrix}$; (2) $\begin{pmatrix} 20 & 3 & 1\\ 2 & 10 & 2\\ 8 & 1 & 0\end{pmatrix}$.

23. 设矩阵

$$A=\begin{pmatrix} 7 & -16 & 8\\ -16 & 7 & -8\\ 8 & -8 & -5\end{pmatrix}.$$

[68]J. S. Hadamard(1865-1963), 著名法国数学家, 组合学中有著名的 **Hadamard 猜想**: 对每个正整数k, 均存在$4k$阶的 Hadamard 矩阵.

(1) 求A的盖尔圆盘并利用对角相似变换改进之;

(2) 利用特征多项式计算A的特征值并与(1)比较.

24. 证明 Hilbert[69] 矩阵

$$A=\begin{pmatrix} 2 & \frac{1}{2} & \frac{1}{2^2} & \cdots & \frac{1}{2^{n-1}} \\ \frac{2}{3} & 4 & \frac{2}{3^2} & \cdots & \frac{2}{3^{n-1}} \\ \frac{3}{4} & \frac{3}{4^2} & 6 & \cdots & \frac{3}{4^{n-1}} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{n}{n+1} & \frac{n}{(n+1)^2} & \frac{n}{(n+1)^3} & \cdots & 2n \end{pmatrix}$$

可以对角化, 且A的特征值都是实数.

25. 分别利用盖尔圆盘定理和 Ostrowski 圆盘定理估计下面矩阵的谱半径:

$$A=\begin{pmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 1.2 & -0.6 & -0.2 & -3.6 \end{pmatrix}.$$

26. 证明$\sigma(A)\subseteq G(A)\cap G(A^T)$.

27. 设矩阵A酉相似于矩阵

$$\begin{pmatrix} 3.05 & -0.06 & 0.02 \\ -0.06 & -6.91 & 0.07 \\ 0.02 & 0.07 & 8.44 \end{pmatrix}.$$

试估计A的特征值.

28. 证明$\sigma(A)=\cap_P G(P^{-1}AP)$, 其中$P$取遍所有可逆矩阵. 如果将$P$限定为可逆对角矩阵如何?

29. 设$A=(a_{ij})$有s行严格对角占优, 证明$r(A)\geq s$.

30. 设n阶矩阵$A=(a_{ij})$至少有$n-1$行严格对角占优, 且剩余一行的对角元素非零, 证明A可逆.

31. 设$A=(a_{ij})$是严格对角占优矩阵, $D=\mathrm{diag}\,(a_{11},\cdots,a_{nn})$. 证明$D$可逆且$\rho(I-D^{-1}A)<1$.

32. 研究下面的矩阵, 说明 Brauer 定理不能推广到三个盖尔圆盘的方程相乘的情形.

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

33. 试利用 Ostrowski 圆盘定理和 Brauer 定理各给出一个矩阵可逆的充分条件.

34. 设数列$a_0=0,a_1=1,a_2,\cdots,a_n,\cdots$满足条件$a_{n+1}=xa_{n-1}+a_n, n\geq 1$, 试求$a_n$的通项公式, 其中$x$为实参数. (当$x=1$时, 此数列即为 **Fibonacci[70] 数列** .)

35. 设A是n阶矩阵, 称满足条件$y^TA=\lambda y^T$的向量y为A的属于特征值λ的左特征向量. 证明: A的相应于特征值λ的左特征向量与相应于特征值μ的特征向量正交($\lambda\neq\mu$).

36. (无限维线性空间的线性变换的特征值与特征向量) 无限次可导的实函数全体构成一个无限维实线性空间, 记为C^∞. 定义C^∞上的线性变换$\partial=\frac{\mathrm{d}}{\mathrm{d}x}$:

$$\partial:\quad f(x)\mapsto f'(x).$$

试求∂的谱$\sigma(\partial)$与特征向量. 比较你的结论与有限维线性空间的相应结论.

---

[69] David Hilbert(1862-1943), 著名德国数学家, 对数学与物理的众多分支有杰出贡献, 他于 1900 年提出的数学的 23 个问题几乎确定了此后一个多世纪世界数学的整个发展方向. Hilbert 的名言: We must know, we shall know.

[70] Leonardo Pisano 或 Leonardo Bonacci(1170-1250), 意大利数学家, 被称为中世纪最具天赋的西方数学家.

下面的37-40题展示了如何利用广义特征子空间来得到矩阵的Jordan标准形, 其中设$|\lambda I - A| = \prod_{i=1}^{s}(\lambda - \lambda_i)^{n_i}$是$n$阶矩阵$A$的特征多项式, $\sigma(A) = \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, g_i为λ_i的几何重数.

37. 证明广义特征子空间$E_{\lambda_i} = \{x \in \mathbb{C}^n \mid (A - \lambda I)^{n_i} x = 0\}$.

38. 证明$\dim_{\mathbb{C}} E_{\lambda_i} = n_i$, 从而$\mathbb{C}^n = \sum_{\lambda \in \sigma(A)} \oplus E_\lambda$.(此即“谱定理”.)

39. 证明存在$\alpha_j \in E_{\lambda_i}, 1 \le j \le g_i$, 使得$\cup_{1 \le j \le g_i}\{\alpha_j, (A - \lambda_i I)\alpha_j, \cdots, (A - \lambda_i I)^{m_j - 1}\alpha_j\}$构成$E_{\lambda_i}$的一组基(称为由诸向量$\alpha_j$生成的循环基), 从而$E_{\lambda_i}$是$A$的不变子空间.

40. 由每个广义特征子空间的循环基构成的$\mathbb{C}^n$的基称为Jordan基. 证明A在$\mathbb{C}^n$的Jordan基下的矩阵是其Jordan标准形(即将A看成是线性变换$x \mapsto Ax$).

下面的41-45题是证明Jordan标准形存在与唯一性的λ-矩阵方法, 为简单起见, 所有矩阵均假定是n阶的. λ-矩阵的初等变换与通常的线性变换类似(倍加变换可以使用多项式).

41. 证明任何秩为r的λ-矩阵$A(\lambda)$均(在初等变换下)等价于下面的对角矩阵(称为$A(\lambda)$的**Smith标准形**[71])

$$\operatorname{diag}(d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda), 0, \cdots, 0),$$

其中$d_i(\lambda)$均为首一多项式且$d_i(\lambda) | d_{i+1}(\lambda), 1 \le i \le r-1$. 这些$d_i(\lambda)$称为$A(\lambda)$的**不变因子**.

42. 将$A(\lambda)$的每个正次数的不变因子分解为不同的首一一次多项式的幂的乘积, 这些一次多项式的幂合称为矩阵$A(\lambda)$的**初等因子**. 证明每个Jordan块仅有唯一的不变因子, 从而这个不变因子就是它的唯一初等因子.

43. 证明$\lambda I - A$与$\lambda I - B$等价 $\Longleftrightarrow$ 它们有相同的不变因子.

44. 证明矩阵A与B相似 $\Longleftrightarrow$ $\lambda I - A$与$\lambda I - B$等价.

45. 证明任何矩阵A一定相似于一个Jordan标准形.

研究性问题.

记相似于$A \in \mathbb{C}^{n\times n}$的矩阵全体为$\mathcal{S}(A)$. 记$\delta(A)$为$A$的非对角元素中非0元素的个数. A的Jordan标准形$J(A)$含有很多个0, 但$J(A)$是否是$\mathcal{S}(A)$中含零元素最多的矩阵, 即

$$\delta(J(A)) = \min_{B \in \mathcal{S}(A)} \delta(B)?$$

答案是否定的, 詹兴致[72]与合作者证明了下面的

定理. $\delta(J(A)) \le \delta(B), \forall B \in \mathcal{S}(A)$.

相关论文请参见[73].

[71]Henry John Stephen Smith (1826-1883), 爱尔兰数学家.

[72]X. Z. Zhan(1965-), 华东师范大学教授.

[73]R. Bruadi, P. Pei, X. Z. Zhan, An extremal sparsity property of the Jordan canonical form. Linear Algebra Appl. 429 (2008), no. 10, 2367–2372.

第四章　矩阵分解与应用

本 章 提 要

本章集中介绍几种常用的矩阵分解, 包括谱分解和Schur三角分解, Cholesky分解与LU分解, 正交三角分解(又称为QR分解), 奇异值分解(即SVD, singular value decompostion)以及非负矩阵分解(即NMF, Non-Negative Matrix Factorization).

第一节　正规矩阵与谱分解

4.1.1　正规矩阵

由Schur三角化定理, 任何矩阵都可以酉三角化, 但2阶幂零Jordan块显然不能酉对角化.

定理 4.1.1. 设$A \in \mathbb{C}^{n\times n}$. 则下述条件等价：
(1) A可以酉对角化;
(2) A有n个两两正交的特征向量;
(3) $AA^* = A^*A$.

证　(1)与(2)显然是等价的. 由Schur三角化定理, 存在酉矩阵U使得$U^*AU = T$为上三角矩阵. 显然$AA^* = A^*A \iff TT^* = T^*T$. 因此不妨设$A$是上三角矩阵.

(1)⇒(3). 设A可以酉对角化, 则存在酉矩阵U使得$U^*AU = D$是对角矩阵, 因此

$$AA^* = (UDU^*)(UD^*U^*) = UDD^*U^* = (UD^*U^*)(UDU^*) = A^*A.$$

(3)⇒(1). 记$A = (a_{ij})$, 其中$a_{ij} = 0, i > j$. 因为$AA^* = A^*A$, 故两端矩阵具有相同的对角元素, 因此

$$a_{11}\overline{a}_{11} + a_{12}\overline{a}_{12} + \cdots + a_{1n}\overline{a}_{1n} = \overline{a}_{11}a_{11},$$

故知A的第一行的非对角元素均为0. 于是由归纳法即可知A是对角矩阵. □

定义 4.1.1. 设$A \in \mathbb{C}^{n\times n}$, 若$AA^* = A^*A$, 则称$A$为正规矩阵.

由定理4.1.1可知, 正规矩阵就是可以酉对角化的矩阵. 实对称矩阵, 实反对称矩阵, 正交矩阵, Hermite矩阵, 反Hermite矩阵, 酉矩阵等都是正规矩阵. 另外, 若A为正规矩阵, 则与A酉相似的矩阵仍为正规矩阵.

例 4.1.1. 矩阵$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$是正规矩阵.

例 4.1.2. 复对称矩阵一般不是正规矩阵, 比如$\begin{pmatrix} 2i & 1 \\ 1 & 0 \end{pmatrix}$不是正规矩阵.

例 4.1.3. 三角矩阵是正规的当且仅当其为对角矩阵. 因此矩阵$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$可以相似对角化, 但不能酉对角化.

推论 4.1.1. 正规矩阵属于不同特征值的特征向量相互正交.

定理 4.1.2. 设$A \in \mathbb{C}^{n\times n}$是复矩阵, $\lambda_1, \lambda_2, \cdots, \lambda_n$为$A$的$n$个特征值. 则

(1) (**Schur 不等式**) $\sum\limits_{i=1}^{n} |\lambda_i|^2 \le \mathrm{tr}(A^*A) = \sum\limits_{i,j=1}^{n} |a_{ij}|^2$.

(2) A为正规矩阵 $\iff$ $\sum\limits_{i=1}^{n} |\lambda_i|^2 = \mathrm{tr}(A^*A) = \sum\limits_{i,j=1}^{n} |a_{ij}|^2$.

证 只须证明(1). 由Schur 酉三角化定理可知存在酉矩阵U使得$U^*AU = B$是上三角矩阵, B与A有相同的特征值$\lambda_1, \lambda_2, \cdots, \lambda_n$, 不妨设$B$的对角线元素依次为$\lambda_1, \lambda_2, \cdots, \lambda_n$, 则$B^*$的对角线元素依次为$\bar{\lambda_1}, \cdots, \bar{\lambda_n}$. 注意$A^*A$与$B^*B$相似, 故$\mathrm{tr}(B^*B) = \mathrm{tr}(A^*A)$. 由于

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \mathrm{tr}(A^*A) = \mathrm{tr}(B^*B) = \sum_{i=1}^{n} |\lambda_i|^2 + \text{非对角元素的绝对值的平方},$$

故$\sum\limits_{i=1}^{n} |\lambda_i|^2 \le \mathrm{tr}(A^*A) = \sum\limits_{i,j=1}^{n} |a_{ij}|^2$. □

例 4.1.4. 设A为正规矩阵且幂零, 则$A = 0$.

例 4.1.5. 实矩阵$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$是正规矩阵, 但不能正交对角化.

为了更好地描述实正规矩阵, 我们引入以下的定义.

定义 4.1.2. 设a与b是实数且$b \neq 0$. 则称2阶实矩阵

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{4.1.1}$$

为一个 **Schur 型**.

注. 1831年, Gauss 将复数看作平面上的点, 并得出了模长为1的复数相当于平面上的旋转变换这一结论. 式(4.1.1)中的Schur 型正是复数$a + b\mathrm{i}$的矩阵表示, 源自 Cayley(1845 年). 易知, 一个Schur 型的特征值恰好是复数$a \pm b\mathrm{i}$.

例 4.1.6. 记式(4.1.1)中的Schur 型为A. 由于$b \neq 0$, A具有非实特征值$a \pm b\mathrm{i}$, 且酉相似于对角矩阵$(a + b\mathrm{i}) \oplus (a - b\mathrm{i})$. $A^*A = AA^* = (a^2 + b^2) \oplus (a^2 + b^2)$. 因此每个Schur 型都是正规矩阵, 但不能正交对角化.

实际上在第三章习题4 中, 我们已经看到了实正规矩阵在正交变换下的最简形式如下

定理 4.1.3. 设A是n阶实矩阵, 则A是正规矩阵 $\iff$ 存在正交矩阵Q使得

$$Q^TAQ = A_1 \oplus A_2 \oplus \cdots \oplus A_s \tag{4.1.2}$$

其中每个A_i或者是1阶实矩阵, 或者是一个Schur 型.

例 4.1.7. 2阶实正规矩阵A的几何意义可由定理4.1.3解释. A正交相似于对角矩阵或一个Schur 型$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, 其中$b \neq 0$, 而2阶正交矩阵是旋转变换与反射变换的复合, 对角矩阵则是

在两个正交方向的伸缩; 一个Schur型是系数为$\sqrt{a^2+b^2}$的位似与一个旋转的合成, 因此2阶实正规矩阵对应的平面线性变换或者是在两个正交方向的伸缩, 或者是一个旋转与一个位似的合成. 一般将正规矩阵对应的线性变换称为**正规变换**, 于是平面上的可逆线性变换σ是正规变换 $\Longleftrightarrow$ σ将某个正方形伸缩为矩形或者将所有正方形均变为正方形.

思考题
1. 复对称矩阵是否是正规矩阵?
2. 正规矩阵的和与积是否为正规矩阵?
3. 相似变换是否保持矩阵的正规性?
4. 试讨论正规矩阵与非正规矩阵的多寡?
5. 讨论2阶与3阶实对称矩阵的特征值(包括零)的几何意义.

4.1.2 正规矩阵的谱分解

设A为正规矩阵, 则由定理4.1.1知, 存在酉矩阵U使得$U^*AU = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 因而$A = U\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U^*$. 令$U = (\alpha_1, \alpha_2, \cdots, \alpha_n)$, 则

$$\begin{aligned} A &= (\alpha_1, \alpha_2, \cdots, \alpha_n)\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}\begin{pmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_n^* \end{pmatrix} \\ &= \lambda_1\alpha_1\alpha_1^* + \lambda_2\alpha_2\alpha_2^* + \cdots + \lambda_n\alpha_n\alpha_n^*. \end{aligned} \tag{4.1.3}$$

公式(4.1.3)称为正规矩阵A的**谱分解**或**特征(值)分解**. 将公式(4.1.3)按系数合并同类项, 并去掉0特征值对应的项, 则公式(4.1.3)化为

$$A = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_s P_s, \tag{4.1.4}$$

其中λ_1, $\lambda_2, \cdots, \lambda_s$为$A$的互不相同的非零特征值. 由于

$$\begin{aligned} &(\alpha_i\alpha_i^*)^* = \alpha_i\alpha_i^*, \qquad 1 \le i \le n, \\ &(\alpha_i\alpha_i^*)(\alpha_j\alpha_j^*) = 0, \qquad 1 \le i \ne j \le n, \\ &(\alpha_i\alpha_i^*)^2 = \alpha_i\alpha_i^*, \qquad 1 \le i \le n, \end{aligned}$$

所以

$$P_i^* = P_i, \quad P_i^2 = P_i, \quad P_iP_j = 0, \quad 1 \le i \ne j \le s. \tag{4.1.5}$$

这样, 所有的P_i都是投影矩阵.

例 4.1.8. 已知正规矩阵

$$A = \begin{pmatrix} 8 & 4 & -1 \\ 4 & -7 & 4 \\ -1 & 4 & 8 \end{pmatrix},$$

求A的谱分解.

解

$$|\lambda I - A| = \begin{vmatrix} \lambda-8 & -4 & 1 \\ -4 & \lambda+7 & -4 \\ 1 & -4 & \lambda-8 \end{vmatrix} = (\lambda-9)^2(\lambda+9).$$

所以A的特征值为$\lambda_1 = 9$(二重), $\lambda_2 = -9$. 相应的相互正交的单位特征向量为:

$$\alpha_1 = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \quad \alpha_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \alpha_3 = \frac{1}{3\sqrt{2}}\begin{pmatrix} 1 \\ -4 \\ 1 \end{pmatrix}.$$

则

$$\begin{aligned} P_1 = \alpha_1\alpha_1^* + \alpha_2\alpha_2^* &= \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}) + \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}) \\ &= \frac{1}{9}\begin{pmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \frac{1}{18}\begin{pmatrix} 17 & 4 & -1 \\ 4 & 2 & 4 \\ -1 & 4 & 17 \end{pmatrix} \\ P_2 = \alpha_3\alpha_3^* &= \frac{1}{3\sqrt{2}}\begin{pmatrix} 1 \\ -4 \\ 1 \end{pmatrix}(\frac{\sqrt{2}}{6}, -\frac{2\sqrt{2}}{3}, \frac{\sqrt{2}}{6}) = \frac{1}{18}\begin{pmatrix} 1 & -4 & 1 \\ -4 & 16 & -4 \\ 1 & -4 & 1 \end{pmatrix}. \end{aligned}$$

所以A的谱分解为

$$A = 9P_1 - 9P_2.$$

例 4.1.9. 谱分解有明显的几何意义. 如果2阶实正规矩阵A有两个相同的特征值λ, 则$A = \lambda I$就是它的谱分解. 如果A有两个不同的特征值λ_1与λ_2, 则其谱分解为$A = \lambda_1 P_1 + \lambda_2 P_2$. 因此有

$$A\alpha = \lambda_1 P_1 \alpha + \lambda_2 P_2 \alpha, \forall \alpha \in \mathbb{R}^2. \tag{4.1.6}$$

计算内积可得$(P_1\alpha, P_2\alpha) = (P_1\alpha)^T P_2 \alpha = \alpha^T P_1^T P_2 \alpha = 0$, 所以$\lambda_1 P_1 \alpha$与$\lambda_2 P_2 \alpha$是正交的向量. 所以公式(4.1.6)将$A\alpha$分解成了两个正交向量的和. 因此, 二维正规矩阵的谱分解实际上是平面正交投影变换的推广. 对任意n阶正规矩阵的谱分解公式(4.1.4)有类似的解释.

例 4.1.10. 设$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. 则$A$的谱分解为

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \mathrm{i}\begin{pmatrix} \frac{1}{2} & -\frac{\mathrm{i}}{2} \\ \frac{\mathrm{i}}{2} & \frac{1}{2} \end{pmatrix} - \mathrm{i}\begin{pmatrix} \frac{1}{2} & \frac{\mathrm{i}}{2} \\ -\frac{\mathrm{i}}{2} & \frac{1}{2} \end{pmatrix},$$

于是

$$A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathrm{i}\begin{pmatrix} \dfrac{x_1 - \mathrm{i}x_2}{2} \\ \dfrac{\mathrm{i}x_1 + x_2}{2} \end{pmatrix} - \mathrm{i}\begin{pmatrix} \dfrac{x_1 + \mathrm{i}x_2}{2} \\ -\dfrac{\mathrm{i}x_1 + x_2}{2} \end{pmatrix},$$

容易看出, 上式右端的两个复向量是正交的.

例 4.1.11. 设$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, 求$A^TA$与$AA^T$的谱分解.

解　直接计算可知

$$A^TA=\begin{pmatrix}1&1&0\\1&2&1\\0&1&1\end{pmatrix},\qquad |\lambda I-A^TA|=\lambda(\lambda-1)(\lambda-3).$$

只需计算非零特征值的特征向量, 得: 属于特征值1与3的特征向量(必定正交)分别为$(1,0,-1)^T$与$(1,2,1)^T$. 因此A^TA的谱分解为

$$A^TA=1\bullet[\frac{1}{2}\begin{pmatrix}1&0&-1\\0&0&0\\-1&0&1\end{pmatrix}]+3\bullet[\frac{1}{6}\begin{pmatrix}1&2&1\\2&4&2\\1&2&1\end{pmatrix}].$$

类似地, 可得AA^T的谱分解为

$$AA^T=1\bullet[\frac{1}{2}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}]+3\bullet[\frac{1}{2}\begin{pmatrix}1&1\\1&1\end{pmatrix}]. \tag{4.1.7}$$

例 4.1.12.　如果A是可逆Hermite矩阵, 则可以利用A的谱分解来求其逆矩阵. 设A的谱分解为

$$A=\sum_{i=1}^{n}\lambda_i\alpha_i\alpha_i^*,$$

则(证明见习题20)

$$A^{-1}=\sum_{i=1}^{n}\frac{1}{\lambda_i}\alpha_i\alpha_i^*. \tag{4.1.8}$$

比如例4.1.11中的矩阵AA^T是可逆对称的, 故由其谱分解公式(4.1.7)及公式(4.1.8)可知其逆矩阵为

$$(AA^T)^{-1}=1\bullet[\frac{1}{2}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}]+\frac{1}{3}\bullet[\frac{1}{2}\begin{pmatrix}1&1\\1&1\end{pmatrix}]=\frac{1}{3}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}.$$

4.1.3　单纯矩阵的谱分解

可对角化的矩阵A也称为单纯矩阵, 也具有类似于正规矩阵的谱分解.

定理4.1.4. (谱分解定理) 设A为一个n阶可对角化矩阵, A的谱为$\sigma(A)=\{\lambda_1,\lambda_2,\cdots,\lambda_s\}$, 其中$\lambda_i$的重数为$k_i$. 则存在唯一一组$s$个$n$阶方阵$P_1,P_2,\cdots,P_s$满足:

$$(1)\ A=\sum_{i=1}^{s}\lambda_iP_i;\qquad (2)\ P_i^2=P_i;\qquad (3)\ P_iP_j=0(\,i\neq j\,);$$
$$(4)\ \sum_{i=1}^{s}P_i=I;\qquad (5)\ r(P_i)=k_i.$$

P_i称为矩阵A的**成分矩阵**或**主幂等矩阵**.

证　设A有s个不同的特征值$\lambda_1,\lambda_2,\cdots,\lambda_s$, 其重数分别为$k_1,k_2,\cdots,k_s$. 设$\alpha_{i1},\alpha_{i2},\cdots,\alpha_{i,k_i}$为对应于$\lambda_i$的线性无关的特征向量. 令

$$P=(\alpha_{11},\alpha_{12},\cdots,\alpha_{1,k_1},\alpha_{21},\alpha_{22},\cdots,\alpha_{2,k_2},\cdots,\alpha_{s1},\alpha_{s2},\cdots,\alpha_{s,k_s}),$$

则

$$P^{-1}AP=\operatorname{diag}(\overbrace{\lambda_1,\cdots,\lambda_1}^{k_1},\overbrace{\lambda_2,\cdots,\lambda_2}^{k_2},,\cdots,\overbrace{\lambda_s,\cdots,\lambda_s}^{k_s}),$$

即

$$A=P\operatorname{diag}(\overbrace{\lambda_1,\cdots,\lambda_1}^{k_1},\overbrace{\lambda_2,\cdots,\lambda_2}^{k_2},,\cdots,\overbrace{\lambda_s,\cdots,\lambda_s}^{k_s})P^{-1}.$$

令

$$P^{-1}=\begin{pmatrix}\beta_{11}^T\\ \vdots\\ \beta_{1,k_1}^T\\ \vdots\\ \beta_{s1}^T\\ \vdots\\ \beta_{s,k_s}^T\end{pmatrix}.$$

则

$$A=\sum_{i=1}^{s}\lambda_i(\alpha_{i1}\beta_{i1}^T+\alpha_{i2}\beta_{i2}^T+\cdots+\alpha_{i,k_i}\beta_{i,k_i}^T) \tag{4.1.9}$$

令

$$P_i=\alpha_{i1}\beta_{i1}^T+\alpha_{i2}\beta_{i2}^T+\cdots+\alpha_{i,k_i}\beta_{i,k_i}^T,\quad i=1,2,\cdots,s. \tag{4.1.10}$$

则

$$A=\lambda_1P_1+\lambda_2P_2+\cdots+\lambda_sP_s. \tag{4.1.11}$$

由于$PP^{-1}=I$, 所以

$$\sum_{i=1}^{s}(\alpha_{i1}\beta_{i1}^T+\alpha_{i2}\beta_{i2}^T+\cdots+\alpha_{i,k_i}\beta_{i,k_i}^T)=I,$$

即$\sum_{i=1}^s P_i=I$. 又因$P^{-1}P=I$, 所以$\beta_{ij}^T\alpha_{kl}=\delta_{ik}\delta_{jl}$. 因此$P_iP_j=0,\forall i\neq j$, 且$(\alpha_{ij}\beta_{ij}^T)^2=\alpha_{ij}\beta_{ij}^T$, 因此$P_i^2=P_i$. 由于

$$A^T=(P^{-1})^T\operatorname{diag}(\overbrace{\lambda_1,\cdots,\lambda_1}^{k_1},\overbrace{\lambda_2,\cdots,\lambda_2}^{k_2},,\cdots,\overbrace{\lambda_s,\cdots,\lambda_s}^{k_s})P^T,$$

所以,

$$A^T(P^{-1})^T=(P^{-1})^T\operatorname{diag}(\overbrace{\lambda_1,\cdots,\lambda_1}^{k_1},\overbrace{\lambda_2,\cdots,\lambda_2}^{k_2},,\cdots,\overbrace{\lambda_s,\cdots,\lambda_s}^{k_s}),$$

由此可知$(P^{-1})^T$的各列为A^T的对应于特征值$\lambda_1,\cdots,\lambda_1,\lambda_2,\cdots,\lambda_2,\cdots,\lambda_s,\cdots,\lambda_s$的线性无关的特征向量. 而

$$(P^{-1})^T=(\beta_{11},\cdots,\beta_{1,k_1},\beta_{21},\cdots,\beta_{2,k_2},\cdots,\beta_{s1},\cdots,\beta_{s,k_s}),$$

因此

$$A^T\beta_{ij}=\lambda_i\beta_{ij},\quad\text{或}\quad\beta_{ij}^TA=\lambda_i\beta_{ij}^T,\quad i=1,2,\cdots,s;\ \ j=1,2,\cdots,k_i. \tag{4.1.12}$$

因$\sum_{i=1}^{s} P_i = I$, 所以

$$n = r(I) \leq \sum_{i=1}^{s} r(P_i) \leq k_1 + k_2 + \cdots + k_s = n,$$

其中k_i为λ_i的重数, $i = 1, 2, \cdots, s$, 而上式的第二个不等式由式(4.1.10)推出. 因此$r(P_i) = k_i$.

唯一性由幂等矩阵的唯一性定理(第二章命题2.4.2)即得. $\square$

注1. 与正规矩阵相比, 单纯矩阵的成分矩阵未必是Hermite矩阵. 因此, 分解$Ax = \lambda_1 P_1 x + \lambda_2 P_2 x + \cdots + \lambda_s P_s x$中的诸向量$P_i x$未必是正交的.

注2. 证明中的向量β_{ij}^T被称为矩阵A的**左特征向量**(相应地, α_{ij}被称为A的**右特征向量**).

例 4.1.13. 求下述矩阵的谱分解:

$$A = \begin{pmatrix} -1 & 3 & -1 \\ -3 & 5 & -1 \\ -3 & 3 & 1 \end{pmatrix}.$$

解

$$|\lambda I - A| = \begin{vmatrix} \lambda+1 & -3 & 1 \\ 3 & \lambda-5 & 1 \\ 3 & -3 & \lambda-1 \end{vmatrix} = (\lambda-1)(\lambda-2)^2.$$

所以A有特征值$\lambda_1 = 1$, $\lambda_2 = 2$(二重). 通过解齐次线性方程组$(\lambda_i I - A)x = 0$, 可得对应于λ_1, λ_2的线性无关的特征向量分别为

$$\alpha_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}; \quad \alpha_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}.$$

令$P = (\alpha_1, \alpha_2, \alpha_3)$, 则可求出

$$P^{-1} = \begin{pmatrix} 3 & -3 & 1 \\ -3 & 4 & -1 \\ -1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \beta_1^T \\ \beta_2^T \\ \beta_3^T \end{pmatrix}.$$

因此,

$$P_1 = \alpha_1 \beta_1^T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}(3, -3, 1) = \begin{pmatrix} 3 & -3 & 1 \\ 3 & -3 & 1 \\ 3 & -3 & 1 \end{pmatrix},$$

$$\begin{aligned} P_2 = \alpha_2 \beta_2^T + \alpha_3 \beta_3^T &= \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}(-3, 4, -1) + \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}(-1, 1, 0) \\ &= \begin{pmatrix} -3 & 4 & -1 \\ -3 & 4 & -1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ -3 & 3 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 3 & -1 \\ -3 & 4 & -1 \\ -3 & 3 & 0 \end{pmatrix}. \end{aligned}$$

故A的谱分解为$A = P_1 + 2P_2$. 注意, 矩阵P_1与P_2的第一列不正交, 故向量$P_1 e_1$与$P_2 e_1$ 不正交.

推论 4.1.2. 设$A=\sum_{i=1}^{s}\lambda_i P_i$是单纯矩阵$A$的谱分解, 则

$$A^m=\sum_{i=1}^{s}\lambda_i^m P_i \tag{4.1.13}$$

从而对任意多项式$f(x)$有$f(A)=\sum_{i=1}^{s}f(\lambda_i)P_i$.

例 4.1.14. 设A如例 4.1.13, 求$e^A=\sum_{n=0}^{\infty}\frac{1}{n!}A^n$.

解 利用定理 4.1.4的(2),(3)和(4), 我们有$A^n=(P_1+2P_2)^n=P_1+2^nP_2, \forall\, n\geq 0$. 从而

$$e^A=\sum_{n=0}^{\infty}\frac{1}{n!}A^n=eP_1+e^2P_2.$$

思考题

1. 试讨论非正规矩阵的谱分解的几何意义.
2. 两个n阶矩阵A与B何时满足条件$AB=BA=0$?
3. 研究单纯矩阵的谱分解, 说明为什么不定义非单纯矩阵的谱分解.

第二节 三角分解与 Cholesky 分解

4.2.1 三角分解

如果线性方程组$Ax=b$的系数矩阵A是可逆上(下)三角矩阵, 则利用反向(顺向)代入法可快速求解. 一般地, 如果A能分解成下三角矩阵与上三角矩阵之积, 即$A=LU$, 其中L与U分别为下三角矩阵与上三角矩阵, 则令$y=Ux$即可将原方程化为$Ly=b$与$Ux=y$两个简单的线性方程组. 注意, 如果这样的分解是存在的, 则显然可以将L(或U)的对角元素均变为1(这样的矩阵称为单位三角矩阵).

定义 4.2.1. 设A是n阶矩阵, 如果存在上三角矩阵U与单位下三角矩阵L使得

$$A=LU \tag{4.2.1}$$

则称A有三角分解或LU分解[74], (4.2.1)称为A的一个**三角分解**或 **LU分解**.

例 4.2.1. 可逆矩阵$A=\begin{pmatrix}0&1\\1&0\end{pmatrix}$不存在三角分解.

三角分解一旦存在, 则必定唯一, 即有下面的定理.

定理 4.2.1. 设A是n阶可逆矩阵, 且有三角分解$A=LU$. 则该分解是唯一的, 且$|A|=|U|=u_{11}u_{22}\cdots u_{nn}$.

[74] LU分解由著名英国数学家、逻辑学家、密码专家、计算机先驱 Alan Mathison Turing (图灵)于1948年提出.

证 设$A=LU=L'U'$是A的两个三角分解, 则由于A可逆, 故所有的矩阵均可逆, 而且$L^{-1}L'=U(U')^{-1}$既是下三角矩阵又是上三角矩阵, 故是对角矩阵. 但L^{-1}与L'均是单位下三角矩阵, 故知对角矩阵$L^{-1}L'$是单位矩阵. 因此$L=L',U=U'$. □

例 4.2.2. 显然不能期望奇异矩阵三角分解的唯一性, 比如$\begin{pmatrix}0&0\\0&1\end{pmatrix}=\begin{pmatrix}1&0\\a&1\end{pmatrix}\begin{pmatrix}0&0\\0&1\end{pmatrix},\forall a.$

消元法的本质就是系数矩阵的三角分解. 容易看出, 只要适当交换矩阵的行, 则三角分解必定存在, 因此消元法总是可以进行的.

定理 4.2.2. 设$A\in\mathbb{F}^{m\times n}$, 则存在置换矩阵$P$使得$PA$有三角分解$PA=LU$.

证 设$r(A)=r>0$. 则存在置换矩阵P使得PA的前r个顺序主子式均非0, 因此存在可逆下三角矩阵矩阵B使得$BPA=\begin{pmatrix}I_r&F\\0&0\end{pmatrix}$. 故$PA=B^{-1}\begin{pmatrix}I_r&F\\0&0\end{pmatrix}$. 注意$B^{-1}$仍然是可逆下三角矩阵, 将其主对角线元素提出得到单位下三角矩阵L, 则有$PA=LU$, □

显然, 如果将(非单位)三角矩阵U的对角线归一化, 则有下面的结论.

推论 4.2.1. 设n阶矩阵A可逆. 则A存在三角分解 $\iff$ A的所有顺序主子式均非0. 此时, 唯一地存在一对单位下三角矩阵L'和单位上三角矩阵U'与对角矩阵D, 使得$A=L'DU'$, 其中D与A具有完全相同的顺序主子式.

使用LU分解求解n元线性方程组$Ax=b$的大约需要$\dfrac{n^3}{3}$次乘法与$\dfrac{n^3}{3}$次减法. 但如果系数矩阵包含较多0, 比如三对角矩阵, 则LU分解将只需要大约$3n$次乘法与减法.

4.2.2 Cholesky 分解

由定理 4.2.2的证明可知, 正定矩阵一定存在三角分解, 实际上有下述更强的结论.

定理 4.2.3. 正定矩阵A必有三角分解$A=LU$, 且存在唯一的对角元素均为正的下三角矩阵G使得
$$A=GG^* \tag{4.2.2}$$

证 对A的阶数n作归纳. 当$n=1$时定理显然成立. 因为A正定, 故$a_{11}>0$, 因此令

$$G_1=\begin{pmatrix}1&0&0&\cdots&0\\-\dfrac{a_{21}}{a_{11}}&1&0&\cdots&0\\-\dfrac{a_{31}}{a_{11}}&0&1&\cdots&0\\\cdots&\cdots&\cdots&\cdots&\\-\dfrac{a_{n1}}{a_{11}}&0&0&\cdots&1\end{pmatrix}$$

即可得

$$G_1AG_1^*=\begin{pmatrix}a_{11}&0\\0&A_1\end{pmatrix},$$

由于左端正定, 故A_1是$n-1$阶正定矩阵. 由归纳假设即可知存在下三角矩阵G_2使得

$$G_2AG_2^* = \mathrm{diag}\,(a_1, a_2, \cdots, a_n).$$

由于A正定, 从而诸$a_i > 0$. 因此

$$A = G_2^{-1}\mathrm{diag}\,(a_1, \cdots, a_n)(G_2^{-1})^* = G_2^{-1}\mathrm{diag}\,(\sqrt{a_1}, \cdots, \sqrt{a_n})\mathrm{diag}\,(\sqrt{a_1}, \cdots, \sqrt{a_n})(G_2^{-1})^*,$$

故知$A = GG^*$, 其中$G = G_2^{-1}\mathrm{diag}\,(\sqrt{a_1}, \cdots, \sqrt{a_n})$是下三角矩阵.

唯一性. 设$A = GG^* = HH^*$, 其中G, H均为对角元素为正的下三角矩阵, 则

$$H^{-1}G = H^*(G^{-1})^* = (G^{-1}H)^* \tag{4.2.3}$$

注意上式左端为对角元素为正的下三角矩阵且等于$(G^{-1}H)^{-1}$, 而右端是对角元素为正的上三角矩阵, 因此$G^{-1}H$是对角正交矩阵, 故对角元素只能是±1, 因此必为单位矩阵, 即$G = H$. □

公式(4.2.2)称为矩阵A的**Cholesky[75]分解**, 矩阵G称为**Cholesky 三角**.

显然, 可以将(4.2.2)中的矩阵G加强为单位下三角矩阵, 而将(4.2.2)化为

$$A = GDG^* \tag{4.2.4}$$

其中D为正定对角矩阵.

例 4.2.3. 求正定矩阵$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$的 Cholesky 分解.

方法一. 由定理 4.2.3的证明可知, 可以利用三角矩阵将A化为正定的对角矩阵, 即令

$$G_1 = \begin{pmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{pmatrix},$$

则

$$G_1AG_1^T = \begin{pmatrix} 2 & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

因此

$$G = G_1^{-1}\begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{\dfrac{3}{2}} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 \\ \dfrac{\sqrt{2}}{2} & \sqrt{\dfrac{3}{2}} \end{pmatrix}.$$

方法二. 设G为对角元素均为正的下三角矩阵, 直接比较$A = GG^T$的两端可知

$$a_{11} = g_{11}^2, a_{12} = g_{11}g_{21}, a_{22} = g_{21}^2 + g_{22}^2,$$

于是

$$g_{11} = \sqrt{2}, g_{21} = \frac{\sqrt{2}}{2}, g_{22} = \sqrt{\frac{3}{2}}.$$

[75] André-Louis Cholesky(1875-1918), 法国大地测量学家.

方法二称为**Cholesky 算法**. 一般地, 设A为n阶实正定矩阵, 比较$A=GG^T$的两端可得

$$a_{ij}=\sum_{k=1}^{j}g_{ik}g_{jk}, i\geq j$$

因此

$$g_{jj}g_{ij}=a_{ij}-\sum_{k=1}^{j-1}g_{ik}g_{jk} \tag{4.2.5}$$

记上式右端为$v(i)$, 则可由(按列)递推的办法求出诸$v(i)$. 特别地, 在公式(4.2.5)中令$i=j$可得$g_{jj}^2=v(j)$, 因此

$$g_{ij}=\frac{v(i)}{g_{jj}}=\frac{v(i)}{\sqrt{v(j)}} \tag{4.2.6}$$

公式(4.2.5)与(4.2.6)合称为Cholesky算法或**平方根法**，因为Cholesky三角矩阵G可以看作是A的平方根, 记作$G=\sqrt{A}$.

例 4.2.4. 设正定矩阵A的Cholesky分解为$A=GG^*$, 则其逆矩阵A^{-1}可由下式求得

$$A^{-1}=(G^{-1})^*G^{-1} \tag{4.2.7}$$

Cholesky分解充分利用了正定矩阵的对称性, 其效率较LU分解大致提高了一倍, 即大约$\dfrac{n^3}{6}$次乘法与减法.

思考题

1. 如果一个矩阵有LU分解, 它是否一定有UL(即上三角在左,下三角在右)分解?
2. 设一个矩阵既有LU分解也有UL分解, 试比较正定矩阵的这两种分解在计算上的差异?
3. 半正定矩阵有无类似的Cholesky分解? 负定矩阵和不定矩阵呢?
4. 如果去掉对角元素均为正的条件, 正定矩阵的Cholesky分解是否具有唯一性?

第三节 QR 分解

利用Gram-Schmidt正交化方法可以得到任意可逆矩阵的正交三角分解.

定理 4.3.1. 设$A\in\mathbb{C}^{n\times n}$可逆, 则存在唯一的酉矩阵$Q$和对角线元素均为正的上三角矩阵$R$满足

$$A=QR. \tag{4.3.1}$$

证 设$\alpha_j=Ae_j, j=1,\cdots,n$. 由$A$可逆知$\alpha_1,\cdots,\alpha_n$构成$\mathbb{C}^n$的一组基. 回忆Gram-Schmidt正交化方法可将$\alpha_1,\cdots,\alpha_n$化成一个正交组$\eta_1,\eta_2,\cdots,\eta_n$:

$$\begin{cases}\eta_1=\alpha_1,\\ \cdots\cdots\cdots\cdots\\ \eta_n=\alpha_n-\dfrac{(\alpha_n,\eta_1)}{(\eta_1,\eta_1)}\eta_1-\cdots-\dfrac{(\alpha_n,\eta_{n-1})}{(\eta_{n-1},\eta_{n-1})}\eta_{n-1}.\end{cases}$$

于是

$$\begin{cases} \alpha_1 = \eta_1, \\ \cdots\cdots\cdots\cdots \\ \alpha_n = \dfrac{(\alpha_n, \eta_1)}{(\eta_1, \eta_1)}\eta_1 + \cdots + \dfrac{(\alpha_n, \eta_{n-1})}{(\eta_{n-1}, \eta_{n-1})}\eta_{n-1} + \eta_n. \end{cases}$$

令$\beta_i = \dfrac{1}{\sqrt{(\eta_i, \eta_i)}}\eta_i$, 则$Q = (\beta_1, \cdots, \beta_n)$为酉阵, 且

$$\begin{cases} \alpha_1 = \sqrt{(\eta_1, \eta_1)}\beta_1, \\ \cdots\cdots\cdots\cdots \\ \alpha_n = (\alpha_n, \beta_1)\beta_1 + \cdots + (\alpha_n, \beta_{n-1})\beta_{n-1} + \sqrt{(\eta_n, \eta_n)}\beta_n. \end{cases}$$

因而$A = QR$, 其中$Q = (\beta_1, \cdots, \beta_n)$, 以及

$$R = \begin{pmatrix} \sqrt{(\eta_1, \eta_1)} & (\alpha_2, \beta_1) & \cdots & (\alpha_{n-1}, \beta_1) & (\alpha_n, \beta_1) \\ & \sqrt{(\eta_2, \eta_2)} & \cdots & (\alpha_{n-1}, \beta_2) & (\alpha_n, \beta_2) \\ & & \ddots & \vdots & \vdots \\ & & & \sqrt{(\eta_{n-1}, \eta_{n-1})} & (\alpha_n, \beta_{n-1}) \\ & & & & \sqrt{(\eta_n, \eta_n)} \end{pmatrix}.$$

设$A = Q_1R_1$为另一分解, 其中Q_1为酉矩阵, R_1为对角线元素为正的上三角矩阵. 则从$QR = Q_1R_1$得$RR_1^{-1} = Q^*Q_1$为上三角酉阵, 故为正规矩阵, 从而RR_1^{-1}为对角矩阵, 且对角线元素均为正. 这样的酉阵必须是单位矩阵(因酉阵的特征值都是模为1的). 因此$Q = Q_1$, $R = R_1$, 唯一性得证. □

公式(4.3.1)称为矩阵A的**正交三角分解**, 也叫QR分解. 当A为实满秩矩阵时, 上面定理中的Q为正交矩阵.

推论 4.3.1. 设可逆矩阵A的正交三角分解为$A = QR$, 则上三角矩阵R的行列式的模等于矩阵A的行列式的模.

实际上列满秩矩阵均存在类似的正交三角分解, 只需将正交矩阵推广为**列正交矩阵**, 即满足$Q^*Q = I_r$的矩阵Q. 列正交矩阵显然是列满秩的, 因其列向量组是标准正交组. 类似可以定义行正交矩阵, 则酉矩阵既是列正交矩阵又是行正交矩阵. 列正交矩阵与行正交矩阵常统称为**半正交矩阵**. 下面定理的证明与定理4.3.1类似, 留作习题.

定理 4.3.2. 设$A \in \mathbb{C}^{n\times r}$, 且$A$是列满秩的, 则

$$A = QR,$$

其中$Q \in \mathbb{C}^{n\times r}$的$r$个列向量构成一组标准正交向量组, $R \in \mathbb{C}^{r\times r}$为对角线元素大于零的上三角矩阵. 此分解是唯一的.

定理4.3.2可以作如下的变化, 即将Q取为n阶酉矩阵, 而将R取为$n \times r$阶准上三角矩阵, 即$R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$, 其中$r$阶上三角矩阵$R_1$具有正对角元素. 通常称此种分解为矩阵$A$的正交三角分解, 而将定理4.3.2中的分解称为矩阵A的**薄QR分解**.

例 4.3.1. 设列满秩矩阵A的正交三角分解为$A = QR$, 则$A^*A = R^*R$, 因此下三角矩阵R^*恰好是正定矩阵A^*A的 Cholesky 三角.

例 4.3.2. 设$\alpha, \beta \in \mathbb{F}^n$线性无关, 则一元线性方程组$\alpha x = \beta$显然无解, 但由定理 4.3.2, α的正交三角分解为

$$\alpha = (\frac{\alpha}{\sqrt{\alpha^*\alpha}})\sqrt{\alpha^*\alpha} = QR \tag{4.3.2}$$

因此原方程两端同乘以$R^{-1}Q^* = \dfrac{1}{\sqrt{\alpha^*\alpha}}(\dfrac{\alpha^*}{\sqrt{\alpha^*\alpha}})$得

$$x = \frac{\alpha^*\beta}{\alpha^*\alpha} \tag{4.3.3}$$

这就是原方程组的最小二乘解.

思考题

1. 可逆矩阵是否存在“三角正交分解”即“$A = RQ$”, 其中R, Q同正交三角分解? 又, 能否将上三角矩阵变为下三角矩阵?
2. 对行满秩矩阵如何定义正交三角分解?
3. 对不可逆矩阵能否定义类似的分解?
4. 由$U^*U = I$是否可以推出$UU^* = I$?

第四节 奇异值分解与极分解

本节介绍**奇异值分解**及由其导出的**极分解**. 奇异值分解是 Beltrami[76] 1873 年与Jordan(即Jordan标准形的Jordan)1874年独立发现的, 在当今科技领域具有极其重要的应用.

4.4.1 奇异值分解

设$m \geq n$, 符号$\operatorname{diag}(d_1, \cdots, d_n)_{m\times n}$表示下面的$m \times n$阶对角矩阵

$$\begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & d_n \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

类似地, 若$m \leq n$, 则以符号$\operatorname{diag}(d_1, \cdots, d_m)_{m\times n}$表示$m \times n$阶对角矩阵.

定理 4.4.1. 设非零矩阵$A = (a_{ij}) \in \mathbb{C}^{m\times n}$, 则存在$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$和酉阵$U \in \mathbb{C}^{m\times m}$与$V \in \mathbb{C}^{n\times n}$, 使得

$$A = UDV^* \tag{4.4.1}$$

其中$D = \operatorname{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0)_{m\times n}$.

公式(4.4.1)称为矩阵A的奇异值分解, 简称为SVD, 而$\sigma_1, \cdots, \sigma_r, 0, \cdots, 0$(共$\min\{m, n\}$个) 称为$A$的**奇异值**.

---

[76]Eugenio Beltrami(1835-1900), 意大利数学家.

证　不妨设$m \ge n$. 对任意矩阵A, A^*A总是半正定的, 因此可设$\sigma(A^*A) = \{\sigma_1^2, \cdots, \sigma_n^2\}$, 其中$\sigma_1 \ge \cdots \ge \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_n$. 还可设$\alpha_1, \cdots, \alpha_n$是由$A^*A$的特征向量构成的$\mathbb{C}^n$的一组标准正交基. 令$V = (\alpha_1, \cdots, \alpha_n) \in \mathbb{C}^{n\times n}$, 则$V$是酉矩阵且

$$V^*(A^*A)V = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_n^2) \tag{4.4.2}$$

此即$(AV)^*(AV) = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_r^2, 0, ..., 0)$. 从而$AV$的列向量彼此正交, 且

$$(AVe_i)^*(AVe_i) = \begin{cases} \sigma_i^2, & 1 \le i \le r \\ 0, & r+1 \le i \le n \end{cases} \tag{4.4.3}$$

从而$u_1 = \dfrac{AVe_1}{\sigma_1}, ..., u_r = \dfrac{AVe_r}{\sigma_r}$是一个标准正交组. 任取列空间$R(AV)$的正交补$R(AV)^\perp = N(V^*A^*)$的一组标准正交基$u_{r+1}, ..., u_m$, 并令

$$U = (u_1, \cdots, u_r, u_{r+1}, \cdots, u_m).$$

则U是酉矩阵且

$$AV = (u_1, ..., u_r, 0, ..., 0) = (u_1, \cdots, u_r, u_{r+1}, \cdots, u_m)\operatorname{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0) = UD.$$

故

$$A = UDV^* \tag{4.4.4}$$

□

显然, 实矩阵的奇异值分解中的酉矩阵都可以取为正交矩阵.

一般, 将矩阵A的非零奇异值从大到小记为$\sigma_1(A) > \sigma_2(A) > ... > \sigma_k(A)$, 其中最大奇异值$\sigma_1(A)$与最小奇异值$\sigma_k(A)(> 0)$也经常分别记为$\sigma_{\max}(A)$与$\sigma_{\min}(A)$.

由奇异值分解定理的证明可知,

$$Av_i = \begin{cases} \sigma_i u_i, & 1 \le i \le r \\ 0, & r+1 \le i \le n \end{cases} \tag{4.4.5}$$

与

$$u_i^*A = \begin{cases} \sigma_i v_i^*, & 1 \le i \le r \\ 0, & r+1 \le i \le m \end{cases} \tag{4.4.6}$$

因此 矩阵V与U的列向量分别称为矩阵A的**右奇异向量**和**左奇异向量**, 而V与U分别称为A的右奇异向量矩阵和左奇异向量矩阵.

例 4.4.1.　正规矩阵的奇异值是其特征值的模. 酉阵的奇异值均为1. 半正定矩阵的特征值与奇异值相同.

奇异值分解的计算方法.　设$m \ge n$, 则可按公式(4.4.5)与(4.4.6)先求A^*A的一组标准正交特征向量v_i, 然后计算$\sigma_i u_i = Av_i, 1 \le i \le r$, 其余向量$u_i(r+1 \le i \le m)$可由任取$\operatorname{Span}\{u_1, ..., u_r\}^\perp = R(AV)^\perp = N(V^*A^*)$的一组标准正交基即可(显然不唯一).

例 4.4.2.　求任意非0向量$\alpha \in \mathbb{F}^n$的奇异值分解.

解 此时$\alpha^*\alpha$是正数, 故$\alpha^*\alpha$的特征值就是自己, 相应的特征向量取1即可. 于是α的奇异值是$\sqrt{\alpha^*\alpha}$, 相应的右奇异向量是$v_1 = 1 \in \mathbb{F}^1 = \mathbb{F}$, 因此$\alpha$的奇异值分解为

$$\alpha = (\sqrt{\alpha^*\alpha}\alpha, Q) \bullet (\sqrt{\alpha^*\alpha}, 0, ..., 0)^T \bullet 1 \tag{4.4.7}$$

其中Q是$n-1$维子空间$\alpha^\perp$的一组标准正交基构成的列正交矩阵.

例 4.4.3. 求矩阵的奇异值分解:

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

解 先求A^*A和AA^*中阶数较小的矩阵的特征值. 因为

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

故AA^*的特征多项式为

$$|\lambda I - AA^*| = \begin{vmatrix} \lambda-2 & 1 \\ 1 & \lambda-2 \end{vmatrix} = (\lambda-1)(\lambda-3).$$

因此AA^*的特征值为$\lambda_1 = 1$, $\lambda_2 = 3$. 它们相应的单位特征向量分别为

$$\beta_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \beta_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

则A^*A的三个特征值分别为$\lambda_1 = 1$, $\lambda_2 = 3$, $\lambda_3 = 0$, 且属于它们的特征向量分别为

$$\alpha_1 = \frac{1}{\sqrt{\lambda_1}}A^*\beta_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \alpha_2 = \frac{1}{\sqrt{\lambda_2}}A^*\beta_2 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \ \alpha_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix},$$

其中α_3可由Hermite矩阵的属于不同特征值的特征向量彼此正交得到. 令

$$V = (\alpha_1, \alpha_2, \alpha_3), \quad U = (\beta_1, \beta_2), \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{pmatrix},$$

则

$$A = UDV^* = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{pmatrix}.$$

例 4.4.4. 正规矩阵的谱分解当然好于奇异值分解. 设A是正规矩阵, A的谱分解为$A = U^*\Lambda U$, 其中$\Lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 则存在对角酉矩阵$W$使得

$$D = \Lambda W^* = \mathrm{diag}\,(|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|),$$

于是得$A = U^*\Lambda W^*WU = U^*DWU = U^*DV$, 其中矩阵$V = WU$是酉矩阵, 即$A$的一个奇异值分解为$A = U^*DV$.

奇异值分解具有明显的几何意义. 将$A_{m\times n}=UDV^*$看作是从$\mathbb{C}^n$到$\mathbb{C}^m$的线性变换, 则该线性变换首先(在$\mathbb{C}^n$内)将向量x做一旋转而得到向量V^*x, 然后再(将$\mathbb{C}^n$中的向量V^*x)沿前$r=r(A)$个坐标做伸缩(其余坐标变为0)而得到($\mathbb{C}^m$的) 向量DV^*x, 最后再(在$\mathbb{C}^m$内)做一旋转而得到向量UDV^*x. 比较正规矩阵的谱分解可知, 此处的两次旋转一般不是互逆的.

例 4.4.5. 研究单位圆$S^1: x^2+y^2=1$在矩阵$A=\begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix}$作用下的变化. 由于$A^TA=\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$的特征值为2, 8, 易得$A$的奇异值分解为

$$A=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

因此, 矩阵A在单位圆S^1上的作用被分解为三步: 第一步, 旋转, 这不会改变S^1; 第二步, 伸缩, S^1变为椭圆, 第三步, 再次旋转(本次的旋转使原来的坐标轴互换).

从线性变换的角度看, 设$A=UDV^*$是A的一个奇异值分解, 则$AV=UD$的两端分别是同一组向量在两个不同的直角坐标系中的表达, 左右两端的标准正交基分别为V与U的列向量组. 此处的要点是, 只需分别取U与V的前r个列向量即可得到矩阵A的完全信息. 确切地说, 将A的奇异值分解(4.4.1)展开可得下面的所谓矩阵A的**并向量分解**或奇异值分解展开

$$A=\sum_{i=1}^{n}\sigma_i u_i v_i^* \tag{4.4.8}$$

由于$\sigma_j=0,\ j>r$, 所以公式(4.4.8)可以写为

$$A=\sum_{i=1}^{r}\sigma_i u_i v_i^*=U_rD_rV_r^* \tag{4.4.9}$$

其中U_r, V_r分别是U, V的前r列构成的矩阵, 而$D_r=\operatorname{diag}(\sigma_1,\cdots,\sigma_r)$是正定对角矩阵. 上式称为矩阵$A$的**截尾奇异值分解**或**薄奇异值分解**. 公式(4.4.8)与(4.4.9)在数据科学中具有重要的应用.

由奇异值分解定理立即可得

推论 4.4.1. 设$A=(a_{ij})_{m\times n}$的秩为r, 则AA^*与A^*A有完全相同的非零特征值(相同的按重数计算)$\sigma_1^2,\sigma_2^2,\cdots,\sigma_r^2$, 其中$\sigma_1,\sigma_2,\cdots,\sigma_r$是$A$的非零奇异值. 因此, 矩阵的非0奇异值的个数等于其秩, 且A与A^*具有相同的(非零)奇异值.

例 4.4.6. 设可逆矩阵A的奇异值分解为$A=UDV^*$, 则其逆的的奇异值分解为$A^{-1}=VD^{-1}U^*$. 因此, 若A的奇异值为$\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n>0$, 则$A^{-1}$的奇异值为$1/\sigma_n\geq 1/\sigma_{n-1}\geq\cdots\geq 1/\sigma_1>0$. 矩阵计算中的重要概念—矩阵$A$的(谱) **条件数**, 记为$Cond(A)$, 可以表示为

$$Cond(A)=\sigma_1(A)/\sigma_n(A). \tag{4.4.10}$$

下面的命题概括了奇异值分解的性质, 所有的证明均比较简单, 留作习题.

命题 4.4.1. 设$A = UDV^*$是$m \times n$矩阵A的一个奇异值分解, $r = r(A)$, 则
(1) 酉矩阵U的前r列是A的列空间的一组标准正交基.
(2) 酉矩阵V的前r列是A^*的列空间的一组标准正交基.
(3) U的后$m - r$列是A^*的零空间的一组标准正交基.
(4) V的后$n - r$列是A的零空间的一组标准正交基.
特别地, A的列空间$R(A)$上的投影矩阵$P_{R(A)} = \sum_{i=1}^{r} UE_{ii}U^*$, 其中$E_{ii} = e_i e_i^T$是矩阵单位.

奇异值与特征值的关系如下.

命题 4.4.2. 设λ是n阶矩阵A的一个特征值, 则$\sigma_{\max}(A) \geq |\lambda| \geq \sigma_{\min}(A)$. 换言之, 矩阵的最大奇异值与最小奇异值是其特征值的模的上下界.

证 设λ是A的一个特征值, α是相应的特征向量. 则

$$A\alpha = \lambda\alpha, \alpha^* A^* = \bar{\lambda}\alpha^*, \alpha^* A^* A\alpha = |\lambda|^2 \alpha^*\alpha.$$

所以

$$|\lambda|^2 = \frac{\alpha^* A^* A\alpha}{\alpha^*\alpha}.$$

由于σ_1^2, σ_n^2恰好是半正定矩阵A^*A的最大和最小特征值, 故由第三章定理3.4.4可知

$$\sigma_n^2 \leq |\lambda|^2 = \frac{\alpha^* A^* A\alpha}{\alpha^*\alpha} \leq \sigma_1^2.$$

□

例 4.4.7. 设$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, 则$A$的特征值为1, 但奇异值为$\sqrt{(3 \pm \sqrt{5})/2}$.

命题 4.4.3. 设λ是n阶Jordan块$J = J_n(\lambda)$的特征值, $\sigma_1 \geq \cdots \geq \cdots \geq \sigma_n > 0$为$J$的奇异值. 则

$$\sigma_n \geq \frac{|\lambda|^n}{(1 + |\lambda|)^{n-1}}.$$

证 直接计算可得

$$A = J^* J = \begin{pmatrix} |\lambda|^2 & \overline{\lambda} & 0 & \cdots & 0 \\ \lambda & 1 + |\lambda|^2 & 0 & \cdots & 0 \\ 0 & \lambda & 1 + |\lambda|^2 & \overline{\lambda} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \lambda & 1 + |\lambda|^2 \end{pmatrix}.$$

由于J^*J的特征值为$\sigma_1^2 \geq \cdots \geq \sigma_n^2$, 利用盖尔圆定理可知

$$0 < \sigma_i^2 \leq 1 + |\lambda|^2 + 2|\lambda|, i = 1, 2, ..., n.$$

于是

$$\sigma_k^2 = \frac{\prod\limits_{i=1}^{k} \sigma_i^2}{\prod\limits_{i=1}^{k-1} \sigma_i^2} = \frac{|A|}{\prod\limits_{i=1}^{k-1} \sigma_i^2} \geq \frac{|\lambda|^{2k}}{(1 + |\lambda|^2 + 2|\lambda|)^{k-1}},$$

故命题成立. □

例 4.4.8. 设$A \in \mathbb{C}^{m\times n}$, 则$\mathrm{tr}(A^*A) = \sum_{i=1}^{r} \sigma_i^2$. 请对照 Schur 不等式.

例 4.4.9. (奇异值与奇异矩阵) 矩阵A列满秩 $\iff$ A的奇异值均非0. 特别地, 方阵A 非奇异 $\iff$ A的奇异值均非0.

定义 4.4.1. 设$0 \le \varepsilon < 1$. 设矩阵A的奇异值为$\sigma_1 \ge ... \ge \sigma_n$. 正整数$k$称为$A$的$\varepsilon$-**数值秩**如果条件$\sigma_{k+1} \le \varepsilon\sigma_1, \sigma_k > \varepsilon\sigma_1$成立. A的数值秩记为$r_\varepsilon(A)$.

例 4.4.10. 设$A \in \mathbb{F}^{1000\times 1000}$是Hilbert矩阵, 即$A_{ij} = \dfrac{1}{i+j-1}$, 则$r(A) = 1000, r_\varepsilon = 28, \varepsilon = 10^{-15}$. 这就是说, 虽然$A$的1000个奇异值均为正, 但第28大的奇异值已经非常接近0了.

定义 4.4.1中的正数ε称为**"容忍度"**. 显然如果$\varepsilon = 0$, 则A的数值秩与其秩相同, 因此数值秩是秩的真正推广. 矩阵的数值秩是奇异值减小速度的度量, 具有广泛应用. 一般来说, 矩阵的奇异值减小的速度非常快, 这是用截尾奇异值分解近似原始矩阵的理论基础.

4.4.2 极分解

由奇异值分解可得到矩阵的另一种有趣分解—**极分解**.

定理 4.4.2. 设$A \in \mathbb{C}^{n\times n}$, 则存在酉矩阵$U$和唯一的半正定矩阵$P$使得

$$A = PU \tag{4.4.11}$$

上式称为矩阵A的极分解. 矩阵P与U分别称为A的 Hermite 因子与酉因子. 若A可逆, 则酉矩阵U也是唯一的.

证 设$A = U_1DV^*$是A的奇异值分解, 令$P = U_1DU_1^*$, $U = U_1V^*$, 则P是半正定矩阵, U是酉矩阵, 故有极分解$A = PU$.

再证明唯一性. 设$r = r(A)$. 设$A = PU = QV$是可逆A的两个极分解, 则P, Q均为秩为r的半正定矩阵, $Q = PUV^*, Q^2 = QQ^* = P^2$. 故半正定矩阵$Q$与$P$互为平方根. 由于半正定矩阵的平方根是唯一的, 故A的 Hermite 因子是唯一的.

若矩阵A可逆, 则极分解中的 Hermite 因子唯一且可逆, 故酉矩阵U也是唯一的. □

历史上, 极分解是Autonne[77]于1902年发现的, 惊奇的是他并未使用奇异值分解, 而是用极分解导出了奇异值分解！有兴趣的读者可以验证, 直接证明极分解与直接证明奇异值分解的难度恰好旗鼓相当.

矩阵A的极分解(4.4.11)中的半正定矩阵P是唯一满足条件$P^2 = AA^*$的矩阵, 正是矩阵AA^*的平方根, 也常记为$\sqrt{A}$或$A^{\frac{1}{2}}$. 显然, 对任意正整数k, 均可以定义方阵A的k次方根.

例 4.4.11. 极分解的几何意义可以用复数的极形式$z = re^{\mathrm{i}\theta}$来解释. 因为若记$r = |P|, e^{\mathrm{i}\theta} = |U|$, 则行列式$|A| = re^{\mathrm{i}\theta}$恰好是复数$|A|$的极分解. 由于半正定矩阵是正规矩阵, 故矩阵的极分解的几何意义是先旋转然后再沿着一组正交的方向做伸缩, 而复数的极分解的几何意义恰好是旋转角度θ, 再伸缩r倍. 所以, 矩阵的 Hermite 因子可以看成是该矩阵的"长度", 而其酉因子则是其"方向".

[77]L. Autonne(1859-1916), 法国数学家.

由一个矩阵的极分解可以判断其正规性, 即有下述命题.

命题 4.4.4. 设$A=PU$是矩阵A的极分解, 则A是正规矩阵 $\iff PU=UP$.

证 充分性. 设$A=PU=UP$, 则$A^*A=(PU)^*PU=(UP)^*UP=P^*U^*UP=PUU^*P^*=AA^*$.

必要性. 设$A=PU$是正规矩阵. 则$AA^*=P^2=U^*P^2U$. 由于半正定矩阵P^2的半正定平方根为P, 半正定矩阵U^*P^2U的半正定平方根为U^*PU，而平方根唯一，故$P=U^*PU$.

思考题

1. 矩阵的奇异值分解不唯一, 但是否可以确定到某种程度?
2. 能否将极分解中的顺序改变? 即是否存在酉矩阵U和半正定矩阵P使得$A=UP$?
3. 不是方阵的矩阵可否定义极分解? 唯一性如何?
4. 何种矩阵具有唯一的奇异值? 试刻画之.
5. 可否以满足条件$B^2=A$的矩阵B来定义$\sqrt{A}$? 更一般地, 可否以满足条件$B^m=A$的矩阵B来定义$A^{1/m}$?

第五节 应用

本节介绍矩阵分解的若干应用. 为简单起见, 所有矩阵均假定是实数矩阵.

4.5.1 QR 分解的应用

正交三角分解的重要性无论如何都不会高估, 此处近介绍两个应用.

一、最小二乘解

QR分解是求解线性方程组最高效的方法之一. 直接求解线性方程组$Ax=b$一般来说是不可能的, 但如果$A=QR$是一个三角分解, 则该线性方程组可以使用倒代换快速求解.

定理 4.5.1. 设列满秩矩阵$A\in\mathbb{R}^{m\times n}(m\geq n)$的正交三角分解为$A=QR$, 其中$Q\in\mathbb{R}^{m\times n}$为列正交矩阵, $R\in\mathbb{R}^{n\times n}$为上三角矩阵. 则线性方程组$Ax=b$的最小二乘解是$\hat{x}=R^{-1}Q^Tb$.

证 将$A=QR$代入$Ax=b$的正规化方程$A^TAx=A^Tb$可得$R^TQ^TQRx=R^TQ^Tb$, 由于R可逆, $Q^TQ=I_n$, 故正规化方程即是

$$Rx=Q^Tb \tag{4.5.1}$$

从而$\hat{x}=R^{-1}Q^Tb$是最小二乘解.

例 4.5.1. 用QR分解解线性方程组$AX=b$, 其中

$$A=\begin{pmatrix}1&1&3\\2&1&2\\2&-5&8\\-1&3&-7\end{pmatrix},\quad b=\begin{pmatrix}-1\\1\\-11\\9\end{pmatrix}.$$

解 将A的列向量组α_1, α_2, α_3正交化得

$$\begin{cases} \eta_1 = \alpha_1 = (1,2,2,-1)^T, \\ \eta_2 = \alpha_2 - \dfrac{(\alpha_2,\eta_1)}{(\eta_1,\eta_1)}\eta_1 = (1,1,-5,3)^T - \dfrac{-10}{10}(1,2,2,-1)^T = (2,3,-3,2)^T, \\ \eta_3 = \alpha_3 - \dfrac{(\alpha_3,\eta_1)}{(\eta_1,\eta_1)}\eta_1 - \dfrac{(\alpha_3,\eta_2)}{(\eta_2,\eta_2)}\eta_2 \\ \quad = (3,2,8,-7)^T - \dfrac{30}{10}(1,2,2,-1)^T - \dfrac{-26}{26}(2,3,-3,2)^T = (2,-1,-1,-2)^T. \end{cases}$$

再单位化得

$$\beta_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 2 \\ 2 \\ -1 \end{pmatrix}, \quad \beta_2 = \frac{1}{\sqrt{26}}\begin{pmatrix} 2 \\ 3 \\ -3 \\ 2 \end{pmatrix}, \quad \beta_3 = \frac{1}{\sqrt{10}}\begin{pmatrix} 2 \\ -1 \\ -1 \\ -2 \end{pmatrix}.$$

则$Q = (\beta_1,\beta_2,\beta_3)$. 由于$Q^TQ = I_3$, 所以由$A = QR$可得

$$R = Q^TQR = Q^TA = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{2}{\sqrt{10}} & \frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{2}{\sqrt{26}} & \frac{3}{\sqrt{26}} & -\frac{3}{\sqrt{26}} & \frac{2}{\sqrt{26}} \\ \frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & -\frac{2}{\sqrt{10}} \end{pmatrix}\begin{pmatrix} 1 & 1 & 3 \\ 2 & 1 & 2 \\ 2 & -5 & 8 \\ -1 & 3 & -7 \end{pmatrix}$$

$$= \begin{pmatrix} \sqrt{10} & -\sqrt{10} & 3\sqrt{10} \\ & \sqrt{26} & -\sqrt{26} \\ & & \sqrt{10} \end{pmatrix}.$$

所以

$$R^{-1} = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{26}} & -\frac{2}{\sqrt{10}} \\ & \frac{1}{\sqrt{26}} & \frac{1}{\sqrt{10}} \\ & & \frac{1}{\sqrt{10}} \end{pmatrix}.$$

因此

$$x = R^{-1}Q^Tb = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{26}} & -\frac{2}{\sqrt{10}} \\ & \frac{1}{\sqrt{26}} & \frac{1}{\sqrt{10}} \\ & & \frac{1}{\sqrt{10}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{2}{\sqrt{10}} & \frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{2}{\sqrt{26}} & \frac{3}{\sqrt{26}} & -\frac{3}{\sqrt{26}} & \frac{2}{\sqrt{26}} \\ \frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & -\frac{2}{\sqrt{10}} \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ -11 \\ 9 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{26}} & -\frac{2}{\sqrt{10}} \\ & \frac{1}{\sqrt{26}} & \frac{1}{\sqrt{10}} \\ & & \frac{1}{\sqrt{10}} \end{pmatrix}\begin{pmatrix} -2\sqrt{10} \\ 2\sqrt{26} \\ -\sqrt{10} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

注. 由于$A = QR$, 故若A可逆, 则矩阵$R^{-1}Q^T$正是A^{-1}. 一般情形下, $R^{-1}Q^T$是所谓**广义逆**, 将在第六章进一步讨论. 利用广义逆的概念, 定理4.5.1即是说$Ax = b$的最小二乘解就是$x = A^{\dagger}b$.

二、QR算法

计算矩阵特征值的重要方法—QR算法[78] 的基础是矩阵的QR分解.

设$A = A_0 = Q_0R_0$是n阶矩阵A的QR分解. 归纳地定义$A_1 = R_0Q_0 = Q_0^*AQ_0, A_{m+1} = R_mQ_m$. 如果$A$的特征值均不相同(回忆：几乎所有的方阵都是这样的矩阵!), 则矩阵序列

$$A_0, A_1, \cdots, A_k, \cdots \tag{4.5.2}$$

收敛到一个上三角矩阵R. 由于序列(4.5.2)中的每个矩阵均与A酉相似, 因此上三角矩阵R的对角元素就是A的全部特征值, 证明留作习题.

注. 一般来说, QR分解高效但成本较大, 所以通常先使用Householder矩阵将矩阵化为Hessenberg矩阵后再做QR分解, 如此可将运算次数从$O(n^3)$降低至$O(n^2)$.

4.5.2 奇异值分解的应用

一、奇异值分解与最小二乘解

定理 4.5.2. 设矩阵$A \in \mathbb{R}^{m\times n}(m \geq n)$的秩为$r > 0$. 设$A = UDV^T$是$A$的奇异值分解, 其中$U, V$为正交矩阵, $D = \mathrm{diag}\,(\sigma_1, ..., \sigma_r, 0, ..., 0) \in \mathbb{R}^{m\times n}, \sigma_1 \geq ... \geq \sigma_r > 0$. 则线性方程组$Ax = b$的最小二乘解是$\hat{x} = VD^{\dagger}U^Tb$, 其中$D^{\dagger} = \mathrm{diag}\,(\sigma_1^{-1}, ..., \sigma_r^{-1}, 0, ..., 0)$.

证 记$\Sigma = \mathrm{diag}\,(\sigma_1, ..., \sigma_r)$. 将$A = UDV^T$代入$Ax = b$的正规化方程$A^TAx = A^Tb$可得$VD^TU^TUDV^Tx = VDU^Tb$, 即$D^TDV^Tx = DU^Tb$. 展开得

$$\begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} V^Tx = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} U^Tb \tag{4.5.3}$$

去掉后面$m - r$个平凡方程得

$$(\Sigma^2,\ 0)V^Tx = (\Sigma,\ 0)U^Tb \tag{4.5.4}$$

由于Σ可逆, 作变量替换$x = V(\Sigma^{-2},\ 0)^Ty$可得

$$(\Sigma^2,\ 0)V^TV(\Sigma^{-2},\ 0)^Ty = (\Sigma,\ 0)U^Tb \tag{4.5.5}$$

即

$$y = (\Sigma,\ 0)U^Tb \tag{4.5.6}$$

故$\hat{x} = V(\Sigma^{-2},\ 0)^Ty = V(\Sigma^{-2},\ 0)^T(\Sigma,\ 0)U^Tb = VD^{\dagger}U^Tb$是最小二乘解.

注. 矩阵$D^{\dagger} = \mathrm{diag}\,(\sigma_1^{-1}, ..., \sigma_r^{-1}, 0, ..., 0)$称为$D$的**广义逆**, 将在第六章详细讨论. 与定理4.5.1类似, 定理4.5.2是说$Ax = b$的最小二乘解就是$x = A^{\dagger}b$, 可以看出, 奇异值分解是求广义逆的有效途径之一.

二、奇异值分解与图像压缩

[78] QR算法被称为世界十大算法之一, 源自英国数学家J. G. F. Francis的论文: The QR transformation: a unitary analogue to the LR transformation. I. Comput. J. 4 (1961/62), 265 – 271.

通讯公司需要传输大量数字照片. 以华为手机Mate30为例, 一张照片的最大分辨率为4000×3000(像素), 即包含高达12000000个数据. 将用户的所有类似数据不加处理而全部传输会产生高昂成本, 所以传输之前须对原始数据进行压缩, 一个有效途径就是奇异值分解.

设传输的原始数据为$A\in\mathbb{R}^{m\times n}$. 设$A=UDV^T$是$A$的一个奇异值分解, 其中对角矩阵$D=\mathrm{diag}\,(\sigma_1,\cdots,\sigma_n),\sigma_1\geq\cdots\geq\sigma_r>0$. 对任意$1\leq k\leq r$, 取

$$A_k=\sum_{i=1}^{k}\sigma_i(Ue_i)(Ve_i)^T \tag{4.5.7}$$

则$r(A_k)=k$. 以A_k近似A就是以A的前k个大奇异值进行图像传输, 总共需要传输奇异值$\sigma_1,\cdots,\sigma_k$以及相对应的左右奇异向量Ue_i与$Ve_i(1\leq i\leq k)$, 即实际传输$k+mk+nk=k(m+n+1)$个数据, 而不是原来的mn个数据. 比值$\dfrac{mn}{k(m+n+1)}$称为图像的压缩比(其倒数$\times$100%称为数据压缩率).

显然, k越大数据保真度越高, k越小传输效率越高. 实际应用可根据需要适当选择k以获得满意的保真度和传输效率. 比如, 如果照片的第201个奇异值已经较小, 则可以选择$k=200$, 于是仅需要传输$200\times(4000+3000+1)=1400200$个数据, 不到原始数据的12%, 图像压缩比为$12000000\div 1400200\approx 8.57$. 一般来说, 前200个大奇异值能够恢复95%以上原始数据, 足以满足保真度的普通要求.

习　题　四

1. 判断下列矩阵能否酉对角化, 如能, 则求一个酉矩阵U, 使U^*AU为对角形:

$$(1)\ A=\begin{pmatrix}-1&\mathrm{i}&0\\-\mathrm{i}&0&-\mathrm{i}\\0&\mathrm{i}&-1\end{pmatrix};\ (2)\ A=\begin{pmatrix}0&\mathrm{i}&1\\-\mathrm{i}&0&0\\1&0&0\end{pmatrix};\ (3)\ A=\begin{pmatrix}\mathrm{i}&\mathrm{i}&0\\\mathrm{i}&1&0\\0&\mathrm{i}&\mathrm{i}\end{pmatrix}.$$

2. 设A是正规矩阵. 证明或证伪下述命题:
(1) $N(A)=N(A^*)$.
(2) $N(A)=N(A^2)$.
(3) $R(A)=R(A^*)$.
(4) $R(A)=R(A^2)$.

3. 证明两个正规矩阵相似当且仅当它们的特征多项式相同当且仅当它们的极小多项式相同.

4. 设A是正规矩阵, 证明A的任何正整数次幂也是正规矩阵, 但反之不对.

5. 设A是n阶正规矩阵, x是任意复数. 证明
(1) $A-xI$也是正规矩阵.
(2) 对于任何向量x, 向量Ax与A^*x的长度相同.
(3) A的任一特征向量都是A^*的特征向量.
(4) Cullen[79]定理(1965年)：设B是半定矩阵. 则AB是正规矩阵当且仅当$AB=BA$.

6. 设A是正规矩阵, 证明
(1) A是Hermite矩阵 $\iff$ A的特征值全为实数.
(2) A是酉阵 $\iff$ A的特征值的模都是1.
(3) A是幂等阵 $\iff$ A的特征值只能是0与1.
(4) 若A的全部特征值为$\lambda_1,\ \lambda_2,...,\ \lambda_n$, 则$AA^*$与$A^*A$的全部特征值为$|\lambda_1|^2,\ |\lambda_2|^2,...,\ |\lambda_n|^2$. 此结论对非正规矩阵成立吗?

[79]C.G.Cullen(1936-), 美国数学家.

7. 设A是正规矩阵, 证明

(1) 若A是幂等阵, 则A是Hermite矩阵.

(2) 若$A^3 = A^2$, 则$A^2 = A$.

(3) 若A又是Hermite阵, 而且也是一个周期阵(即$A^k = I, k > 0$), 则A是对合阵(即$A^2 = I$).

8. 设$A \in \mathbb{C}^{n\times n}$. 证明:

(1) A是正规矩阵当且仅当A的特征向量均为A^*的特征向量.

(2) A是正规矩阵当且仅当存在多项式$f(x)$使得$A^* = f(A)$.

9. 樊畿正规矩阵引理: 设$X \in \mathbb{C}^{n\times n}$是正规矩阵且$|X| = 1$. 则存在可逆正规矩阵$A, B$使得$X = A^{-1}B^{-1}AB$.

10. 设A, B均为Hermite矩阵. 证明$A + iB$是正规矩阵当且仅当$AB = BA$.

11. 设A可逆且$X^TAX = A$. 证明X是两个对合矩阵的乘积. 特别地, 正交矩阵是两个对合矩阵的乘积.

12. 设P, Q各为m阶及n阶方阵, 证明: 若$m + n$阶方阵$A = \begin{pmatrix} P & B \\ 0 & Q \end{pmatrix}$是酉矩阵, 则$P$, Q也酉矩阵, 且B是零矩阵.

13. 证明**Sylvester 惯性定律**, 即两个Hermite矩阵合同 $\Longleftrightarrow$ 它们具有相同的惯性指标, 即相同的正负特征值(因此0特征值)的个数.

16. (1) 证明例4.1.7关于平面正规变换的结论.

(2) 计算2阶实正规矩阵$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$将哪些正方形变为了矩形?

(3) 证明矩阵$\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$是非正规矩阵, 说明它不能将任何正方形变为矩形.

(4) 试给出3阶实正规矩阵的几何意义.

17. 若3×3矩阵S表示一个反射, 则存在一个正交矩阵C, 使得 $C^{-1}SC = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. 当$S = \frac{1}{9}\begin{pmatrix} 7 & 4 & -4 \\ 4 & 1 & 8 \\ -4 & 8 & 1 \end{pmatrix}$时, 求这样的矩阵$C$.

18. 求习题1中所有单纯矩阵的谱分解.

19. 设单纯矩阵A仅有一个非零特征值λ, 求A的谱分解.

20. 证明例4.1.12并写出其实数形式.

21. 设$A = LU$可逆, 其中L与U分别为下三角矩阵与上三角矩阵, 证明存在单位下三角矩阵L'与上三角矩阵U'使得$A = L'U'$.

22. 证明矩阵$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$不存在三角分解.

23. 试给出正定Hermite矩阵的Cholesky分解定理.

24. 证明定理4.3.2.

25. 证明矩阵序列(4.5.2)中的每一个矩阵均与$A = A_0$酉相似, 并且当A的特征值均不相同时, 该序列收敛于一个与A酉相似的上三角矩阵. 如果A有重特征值, 此结论还成立吗?

26. 证明定理4.3.2中的存在性.

27. 设$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 12 \\ 6 \\ 8 \end{pmatrix}$.

(1) 求$R(A)$的标准正交基.

(2) 写出A的QR分解.

(3) 求$Ax=b$的最小二乘解.

(4) 证明$u_1=(0,1,0)^T$, $u_2=(\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}})^T$,也是$R(A)$的标准正交基, 其中$R(A)$为$A$的列空间.

28. 求下列矩阵的QR分解:

$$(1)\quad A=\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix};\quad (2)\quad A=\begin{pmatrix} 2 & 2 & 1 \\ 0 & 2 & 2 \\ 2 & 1 & 2 \end{pmatrix};(3)\quad A=\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

29. 设$A,B\in\mathbb{C}^{m\times n}$. 证明矩阵分解引理: $A^*A=B^*B\Longleftrightarrow$存在酉矩阵$U$使得$B=UA$.

30. 计算28题中各矩阵的奇异值分解和相应的四个子空间.

31. 设$A\in\mathbb{C}^{m\times n}$, 证明:

$$\sigma_{\min}(A)=\min\{(x^*A^*Ax)^{1/2},x^*x=1\},\qquad \sigma_{\max}(A)=\max\{(x^*A^*Ax)^{1/2}:x^*x=1\}.$$

32. 设变换$\sigma:\sigma x=x-a(x,w)w,\forall x\in\mathbb{R}^n$, 其中$w$为长度为1的向量, 问$a$取何值时, σ为正交变换? 如果w是任意向量, 你的结论又如何?

33. 设A的奇异值分解为$A=U\Sigma V^*$, 求$B=\begin{pmatrix} A & A \\ A & A \end{pmatrix}$与$C=\begin{pmatrix} A \\ A \end{pmatrix}$的奇异值分解.

34. 设A的奇异值是$\sigma_1,...,\sigma_n$, 证明$\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix}$的特征值是$\sigma_1,...,\sigma_n,-\sigma_1,...,-\sigma_n$.

35. 证明矩阵的极分解的唯一性.

36. 证明推论4.4.1.

37. 证明任意n阶矩阵A均可表示成$A=Pe^{\mathrm{i}H}$, 其中P是半正定矩阵, H是Hermite矩阵. 该分解唯一吗?

38. 证明命题4.4.4.

39. 试对任意矩阵定义其极分解, 并由此计算任意向量$x\in\mathbb{C}^n$的极分解.

40. 证明命题4.4.1, 并由此计算$A=\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix}$的四个子空间.

41. 证明命题4.4.2.

42. 证明例4.4.8.

43. 设$x,y,\alpha,\beta\in\mathbb{C}^n$, 且$x^*y=\alpha^*\beta=0$. 设$A=x\alpha^*+y\beta^*$, 求$A$的F-范数.

44. 证明矩阵A可以对角化$\Longleftrightarrow$存在Hermite正定矩阵P使得$P^{-1}AP$是正规矩阵.

45. 证明半正定矩阵的半正定平方根是唯一的.

46. 设A是正定矩阵. 则$A+A^{-1}-2I$也是正定矩阵.

47. 设V为n维线性空间. 证明：V中任意线性变换必可表示为一个可逆变换与一个幂等变换的乘积. (提示: 利用奇异值分解.) 48. 研究正交三角分解, 谱分解, 极分解和奇异值分解之间的关系.

49. 设$A,B\in\mathbb{C}^{m\times n}$. 证明: $x\in N(A)\cap N(B)\Longleftrightarrow\begin{pmatrix} A \\ B \end{pmatrix}x=0$.

50. 证明奇异值的极大极小定理: 设$A\in\mathbb{C}^{m\times n}$的奇异值为$\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n$, 则:

$$\sigma_k=\min_{\substack{w_i\in\mathbb{C}^n\\1\leq i\leq n-k}}\ \max_{\substack{0\neq x\perp w_i\\1\leq i\leq n-k}}\frac{\|Ax\|_2}{\|x\|_2}=\max_{\substack{w_i\in\mathbb{C}^n\\1\leq i\leq n-k}}\ \min_{\substack{0\neq x\perp w_i\\1\leq i\leq n-k}}\frac{\|Ax\|_2}{\|x\|_2}.$$

特别地,

$$\sigma_{\max}=\sigma_1=\max_{x\neq 0,x\in\mathbb{C}^n}\frac{\|Ax\|_2}{\|x\|_2}=\max_{x^*x=1}\|Ax\|_2,$$

$$\sigma_{\min}=\sigma_n=\min_{x\neq 0,x\in\mathbb{C}^n}\frac{\|Ax\|_2}{\|x\|_2}=\min_{x^*x=1}\|Ax\|_2.$$

51. 证明: 对任意同阶矩阵A,B均有$\sigma_{\max}(A+B)\le\sigma_{\max}(A)+\sigma_{\max}(B)$.

52. (矩阵的低秩近似)设矩阵$A_{m\times n}$的秩为r, 其奇异值分解为$A=UDV^*,U=(u_1,\cdots,u_m),V=(v_1,\cdots,v_n)$. 对任意$k<r$, 定义矩阵

$$A_k=\sum_{i=1}^{k}\sigma_i u_i v_i^*,k<r.$$

证明:

$$\min_{r(B)=k}\|A-B\|_1=\|A-A_k\|_1=\sigma_{k+1},k<r$$

以及

$$\min_{r(B)=k}\|A-B\|_F^2=\|A-A_k\|_F^2=\sigma_{k+1}^2+\sigma_{k+1}^2+\cdots+\sigma_r^2.$$

53. 利用上题的结果, 证明利用截尾奇异值分解压缩数据的合理性.

54. (同时奇异值分解) 设A,B是两个$m\times n$矩阵. 证明存在酉矩阵U,V以及对角矩阵D,Λ使得$A=UDV^*$, $B=U\Lambda V^*\iff A^*B$与AB^*均是正规矩阵. 该结论对三个或更多的矩阵成立吗?

55. **樊畿奇异值不等式**. 设$A,B\in\mathbb{C}^{m\times n},q=\min\{m,n\}$, 则对$1\le i,j\le q,i+j\le q+1$, 有
(1) $\sigma_{i+j-1}(A+B)\le\sigma_i(A)+\sigma_j(B)$, $\sigma_{i+j-1}(AB^*)\le\sigma_i(A)\sigma_j(B)$;
(2) $|\sigma_i(A+B)-\sigma_i(A)|\le\sigma_1 B$, $\sigma_i(AB^*)\le\sigma_i(A)\sigma_1(B)$.

研究性问题.

矩阵方程的求解往往非常困难, 中国数学家在该领域的研究处于世界第一方阵, 以下介绍王卿文[80]及其团队的一个精巧结果, 为此先引入几个记号.

设$A\in\mathbb{C}^{n\times n}$, 如果$A$半正定, 则记$A\ge 0$. 对$M,N\in\mathbb{C}^{n\times n}$, 下面的矩阵方程是一类特殊的随机Riccati方程:

$$X+M^*X^{-1}M-N^*X^{-1}N=I \tag{4.5.8}$$

定理. 设$\psi(\lambda)=I+\lambda M+\lambda^{-1}M^*$. 如果$\psi$是正则的(即$|\psi(\lambda)|\not\equiv 0$)且$\psi(\lambda)\ge 0,\forall|\lambda|=1$, 则方程(4.5.8)存在正定解.

相关论文参见[81].

[80]Qingwen Wang(1964-), 上海大学教授.

[81]Xue-Feng Duan, Qing-Wen Wang, Chun-Mei Li, Positive definite solution of a class of nonlinear matrix equation, Linear and Multilinear Algebra. 62(2014), no. 6, 839-852.

第五章　赋范线性空间与矩阵微积分

本 章 提 要

极限概念本质上属于几何学或拓扑学范畴, 本章建立线性空间上的范数(=长度或距离), 由此即可研究矩阵无穷级数的敛散性, 进而研究矩阵函数的微积分. 本章的主要结论如下：

1、有限维线性空间的范数都是等价的

2、矩阵无穷级数的敛散性本质上是数项级数的敛散性

3、矩阵函数可以用来求解常系数线性微分方程组

本章总假定$\mathbb{F} = \mathbb{R}$或$\mathbb{C}$.

第一节　赋范线性空间

5.1.1　赋范线性空间

定义 5.1.1.　设$V \neq 0$是$\mathbb{F}$上的线性空间. 如果V上的实函数$\|\bullet\|$满足下列性质:
(1) 正定性: 对$\forall x \in V$, $\|x\| \geq 0$且$\|x\| = 0 \iff x = 0$.
(2) 齐次性: 对$\forall k \in \mathbb{F}$, $x \in V$, 有

$$\|kx\| = |k|\,\|x\|.$$

(3) 三角不等式: 对$\forall x, y \in V$, 有

$$\|x + y\| \leq \|x\| + \|y\|,$$

则称V是**赋范线性空间**或赋范空间, 记为$(V, \|\cdot\|)$(常省略$\|\bullet\|$), 非负实数$\|x\|$称为向量x的向量范数或范数.

当涉及多个赋范线性空间时, 将以$\|\bullet\|_V$表示V上的范数或干脆省略.

由齐次性立即可知: $\|-x\| = \|x\|$.

例 5.1.1.　(1) 内积空间是自然的赋范线性空间, 只需取$\|x\| = \sqrt{(x,x)}$即可.
(2) 1-维赋范线性空间$(V, \|\cdot\|)$具有自然的内积空间结构, 只需取$(x,x) = \|x\|^2$(由第二章, (x,y)随之唯一确定). 这就是说, 对1-维而言, 赋范线性空间与内积空间是一回事. 比如对实数域$\mathbb{R}$而言, 任何范数都是普通绝对值的$a\,(a > 0)$倍. 请读者给出复数域$\mathbb{C}$上的所有范数.
(3) 设$(V, \|\cdot\|)$是赋范线性空间, 则V的子空间U沿用V的范数也是赋范线性空间.
(4) 任何赋范线性空间$(V, \|\cdot\|)$都具有无穷多种范数, 比如对$\forall \alpha > 0, \alpha \neq 1, \forall 0 \neq x \in V$, 定义$\|x\|_\alpha = \alpha\|x\|$即可得到$V$上的一种新范数.
(5) 根据第二章内积诱导的范数的极化恒等式可知, 赋范线性空间未必是内积空间, 比如下例中的l_p范数由内积诱导$\iff p = 2$. 一般地, 是否每个线性空间均可赋范有趣且复杂, 请有兴趣的读者查阅相关材料.

例 5.1.2.　设$V = \mathbb{C}^n$或$\mathbb{R}^n$, 则下列实值函数都是$\mathbb{C}^n$(或$\mathbb{R}^n$)上的范数, 因而V为赋范空间:

(1) $\|x\|_\infty = \max_{1\le j\le n} |x_j|$ (**最大范数或l_∞范数或∞-范数**).

(2) $\|x\|_1 = \sum_{j=1}^{n} |x_j|$ (**和范数**或l_1**范数**或1-**范数**).

(3) $\|x\|_2 = \left(\sum_{j=1}^{n} |x_j|^2\right)^{1/2}$ (**欧几里得范数**或l_2**范数**).

(4) $\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$, $p \ge 1$ (**Hölder 范数** 或l_p**范数**或p-**范数**).

显然, 当$n=1$时, 例 5.1.2中的所有范数都变成$\mathbb{C}$或$\mathbb{R}$上的普通范数(模或绝对值).

容易看出, $\|\cdot\|_1$与$\|\cdot\|_2$为Hölder 范数中取$p=1$与$p=2$ 的情形, 而$\|\cdot\|_\infty$是Hölder 范数当$p\to\infty$的极限情形. 直接验证可知(见习题3), $\|\cdot\|_\infty$满足定义5.1.1的3个条件, 故为V上的范数. l_p范数($p\ge 1$)显然满足定义中的条件(1)和(2). 为验证条件(3), 只需应用下列**Minkowski 不等式**(证明见习题4):

$$\left(\sum_{j=1}^{n} |x_j+y_j|^p\right)^{1/p} \le \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} + \left(\sum_{j=1}^{n} |y_j|^p\right)^{1/p}, \quad p\ge 1. \tag{5.1.1}$$

注1. 定义5.1.1中的字母"l"是序列空间或Lebesgue [82] 空间的统称. 确切地说, 数域$\mathbb{F}$上的所有绝对收敛的无穷数列构成的线性空间称为l^1空间, 绝对平方收敛的无穷数列构成的线性空间称为l^2空间, 所有有界无穷数列构成的线性空间称为l^∞空间, 以及l^p空间等等(见下面的例5.1.6). 类似地, 可以定义绝对可积函数空间L^1, 绝对平方可积函数空间L^2, 以及L^p, L^∞等等.

注2. 当$0<p<1$时, l^p范数仍然满足范数的前两个条件, 但不满足三角不等式, 见习题5.

注3. 对于$\mathbb{C}$或$\mathbb{R}$上一般n维线性空间V, 可以通过取V的一组基, 然后像例5.1.2中一样定义V的范数.

注4. 常将1-范数称为Manhattan (曼哈顿)-度量, 因为在赋范线性空间中可以由范数自然定义距离, 即

$$d(x,y) = \|x-y\|.$$

请读者画出两点间的距离的示意图. 如果连接两点间的最短曲线称为线段, 请问各种范数下的线段是什么?

例 5.1.3. 下面的图从左至右依次展示了1-范数, 普通范数(欧几里得范数) 和∞-范数下的平面上的单位圆(1维单位球面):

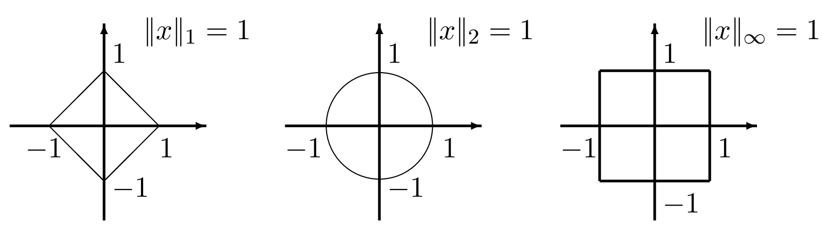

图5.1.1

[82] Henri Léon Lebesgue(1875-1941), 法国数学家, 数学上有著名的Lebesgue积分.

从此例可以看出, 1-范数和∞-范数下的单位圆按通常意义都是正方形! 一般地, 将赋范线性空间V中范数为1的向量的集合称为单位球面, 范数小于等于1的向量的集合称为单位球. 如果单位球面是多面体, 则称该范数是多面的. 因此1-范数和∞-范数是多面的.

请思考,在1-范数和∞-范数下的单位圆中的四个角是直角吗? 此时的角度与我们的常识一致吗? 还有别的多面范数吗?

例 5.1.3还表明, 与内积空间不同, 赋范线性空间中每个向量的范数不必被一组基向量的范数确定, 因此范数较内积远为复杂. 特别地, 等距同构等概念不适合赋范线性空间.

矩阵空间$\mathbb{F}^{n\times n}$是特殊的线性空间, 因为向量可以做乘法, 故其范数有特殊性.

例 5.1.4. 设$\mathbb{F}^{n\times n}$为数域$\mathbb{F}$上n阶矩阵所构成的线性空间, 对$\forall A=(a_{ij})_{n\times n}$, 定义

$$\|A\|_F=\sqrt{\operatorname{tr}(A^*A)}=\left(\sum_{i,j=1}^{n}|a_{ij}|^2\right)^{1/2}=\sqrt{\sigma_1^2+...+\sigma_n^2}\quad \text{(Frobenius 范数或 F-范数)}$$

$$\|A\|_2=\sqrt{\rho(A^*A)}=\sigma_1(=A\text{的最大奇异值})\quad \text{(矩阵2-范数)}$$

$$\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|\quad \text{极大列和范数}$$

$$\|A\|_\infty=\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|\quad \text{极大行和范数}$$

则不难验证(见习题6), 它们都是$\mathbb{F}^{n\times n}$中的范数, 因而$\mathbb{F}^{n\times n}$成为赋范线性空间.

注. 由内积的极化恒等式可知, 矩阵的F-范数是内积诱导的范数, 此时的内积恰为第二章定理2.5.4定义的内积, 即

$$(A,B)=tr(B^*A) \tag{5.1.2}$$

注意, 例 5.1.4中的1-范数与∞-范数与例 5.1.2中的l_1-范数与l_∞-范数并不一致, 其理由见后.

例 5.1.5. 对$(x,y)\in\mathbb{R}^2$, 规定$\|(x,y)\|=\sqrt{x^2+2y^2-2xy}$, 则由命题 5.1.1可知这确是$\mathbb{R}^2$中的一个范数(为什么?). 此时单位圆方程为$x^2+2y^2-2xy=1$, 以欧几里得度量看, 该"单位圆"是非圆的椭圆.

下面介绍著名的**樊畿范数**(又称为**核范数**).

设$A\in\mathbb{C}^{m\times n}$, $\sigma_1\ge...\ge\sigma_q$是$A$的奇异值, $1\le k\le q$. 则

$$\|A\|_{(k)}=\sum_{i=1}^{k}\sigma_i \tag{5.1.3}$$

是一个范数, 称为矩阵A的第k个**樊畿范数**. A的所有特征值之和即第q个樊畿范数$\|A\|_{(q)}$又称为**核范数**, 记为$\|A\|_*$, 即

$$\|A\|_*=\sum_{i=1}^{q}\sigma_i \tag{5.1.4}$$

所有的樊畿范数显然满足范数的正定性与齐次性条件, 但证明$\|A\|_{(k)}$满足三角不等式比较困难, 读者可借助矩阵空间$\mathbb{C}^{m\times n}$上的标准内积(5.1.2)自行证明. 樊畿范数显然是酉不变范数.

注意, A的第一个樊畿范数是A的最大奇异值, 故恰为A的2-范数, 即

$$\|A\|_{(1)} = \sqrt{\rho(A^*A)} = \|A\|_2 \tag{5.1.5}$$

特别地, 当$A^* = A$即A是Hermite矩阵时, A的第一个樊畿范数$\|A\|_{(1)} = \rho(A)$, 即谱范数.

由于A的奇异值正好是半正定矩阵AA^*的特征值的算术平方根, 因此A的核范数恰为

$$\|A\|_* = \sum_{i=1}^{q} \sigma_i = \sum_{i=1}^{q} \sqrt{\lambda_i(A^*A)} \tag{5.1.6}$$

其中$\lambda_i(A^*A)$是半正定矩阵A^*A的特征值. 注意核范数与F-范数的差别.

容易证明下面的命题是由已知范数构造新范数的一种简便方法.

命题 5.1.1. (构造新范数) 设$\|\cdot\|_\alpha$为$\mathbb{F}^m$($\mathbb{F} = \mathbb{C}$或$\mathbb{R}$)上的一种向量范数, $A \in \mathbb{F}^{m\times n}$是列满秩的矩阵. 对$\forall\, x = (x_1, x_2, \cdots, x_n)^T \in \mathbb{F}^n$, 定义

$$\|x\|_\beta = \|Ax\|_\alpha,$$

则$\|\cdot\|_\beta$为$\mathbb{F}^n$的范数.

请注意, 定义 5.1.1适用于无限维线性空间, 以下是几个常见的例子.

例 5.1.6. 设$\mathbb{S} = \{\{x_k\} = (\cdots, x_{-2}, x_{-1}, x_0, x_1, x_2, \cdots) \,|\, x_i \in \mathbb{R}\}$是离散时间信号空间(参见第二章第八节), $\mathbb{S}$的若干子空间如下：

(1) 平方可和子空间

$$l_2 = \{\{x_k\} \in \mathbb{S} \,|\, \sum_{k=-\infty}^{+\infty} x_k^2 < +\infty\}.$$

l_2中向量$x = \{x_k\}$的范数定义为

$$\|x\| = \sum_{k=-\infty}^{+\infty} x_k^2 \tag{5.1.7}$$

(2) 绝对收敛子空间

$$l_1 = \{\{x_k\} \in \mathbb{S} \,|\, x_k = 0, \text{若} k \le 0 \text{且} \sum_{k=1}^{+\infty} |x_k| < +\infty\}.$$

l_1中向量$x = \{x_k\}$的范数定义为

$$\|x\| = \sum_{k=-\infty}^{+\infty} |x_k| \tag{5.1.8}$$

(3) 有限非零序列子空间

$$\bigsqcup_{i=-\infty}^{\infty} \mathbb{R} = \{\{x_k\} \in \mathbb{S} \,|\, x_k \text{ 几乎处处为} 0\,\} \,(\text{即仅有有限个} x_k \text{非} 0\,).$$

有限非零子空间 $\bigsqcup_{i=-\infty}^{\infty} \mathbb{R}$ 是l_1与l_2的公共子空间, 其上的范数可以继承l_1或l_2的范数.

(4) 有界序列子空间

$$l_\infty = \{\{x_k\} \in \mathbb{S} \mid x_k = 0, 若k \le 0, 且存在整数C, |x_k| < C, \forall k\}.$$

l_∞中向量$x = \{x_k\}$的范数定义为

$$\|x\| = \max_k |x_k| \tag{5.1.9}$$

不难验证, 公式(5.1.7)定义的平方可和子空间l_2上的范数恰好是由其上内积诱导的(见第二章第八节), 该内积如下

$$(\{x_k\}, \{y_k\}) = \sum_{k=-\infty}^{+\infty} x_k y_k. \tag{5.1.10}$$

赋范线性空间的积与商是否还是(自然的)赋范线性空间? 下面两个定理给出肯定的回答.

定理 5.1.1. 设U, V是两个赋范线性空间. 对任意$(u, v) \in U \times V$, 定义

$$\|(u, v)\| = \|u\|_U + \|v\|_V \tag{5.1.11}$$

则$(U \times V, \|\bullet\|)$是赋范线性空间.

证 正定性与齐次性都是显然的. 三角不等式的证明如下(以下略去范数的下标):

$$\|(u, v) + (u', v')\| = \|(u + u', v + v')\| = \|u + u'\| + \|v + v'\|$$

$$\le \|u\| + \|u'\| + \|v\| + \|v'\| = \|(u, v)\| + \|(u', v')\|.$$

□

按照Minkowski 不等式(5.1.1), 可以将公式(5.1.11)修改为(其中$p > 1$)

$$\|(u, v)\| = ((\|u\|_U)^p + (\|v\|_V)^p)^{\frac{1}{p}} \tag{5.1.12}$$

这实际上也是一种Hölder 范数.

请思考, 是否可将公式(5.1.11)中的加法改为乘法?

定理 5.1.2. 设V是赋范线性空间, U是V的子空间. 对任意$v + U \in V/U$, 定义

$$\|v + U\| = \inf_{x \in v+U} \|x\|_V \tag{5.1.13}$$

(此处inf是下确界.) 则$(V/U, \|\bullet\|)$是赋范线性空间.

证 因为$\|v+U\|=\inf\limits_{x\in v+U}\|x\|\geq 0$且等号成立$\iff x=0\in v+U \iff v+U=U$, 故正定性成立.

注意到$x\in v+U \iff kx\in kv+U$, 故

$$\|kv+U\|=\inf_{x\in kv+U}\|x\|=\inf_{x\in v+U}\|kx\|=|k|\inf_{x\in v+U}\|x\|=|k|\|v+U\|,$$

故齐次性成立.

再证三角不等式如下(为简明起见, 略去范数的下标):

$$\|(v+U)+(v'+U)\|=\inf_{x\in v+v'+U}\|x\|=\inf_{u,u'\in U}\|v+v'+u+u'\|$$

$$\leq \inf_{u\in U}\|v+u\|+\inf_{u'\in U}\|v'+u'\|=\|v+U\|+\|v'+U\|.$$

□

在例 5.1.2的(1)与例 5.1.6的(4)中, 范数都是取 max , 而在定理 5.1.2中, 商空间V/U的范数却是取inf, 相当于取极小值. 这是因为商空间V/U的每个元素都是子空间U的陪集, 因此一般不存在极大值, 但存在极小值, 几何上达到范数极小值的点就是陪集$v+U$中距离原点最近的点.

例 5.1.7. 设$V=\mathbb{R}^2$是普通欧氏空间(即范数为欧几里得范数). 设$U=\{k(1,1)|k\in\mathbb{R}\}$. 对任意$v=(v_1,v_2)\in V$, 公式(5.1.13)定义的$v+U\in V/U$的范数为

$$\|v+U\|=\inf_{(x,y)\in v+U}\|(x,y)\|_V=\inf_{k\in\mathbb{R}}\|(v_1,v_2)+k(1,1)\|_V=\frac{|v_1-v_2|}{\sqrt{2}}.$$

此即原点到直线$v+U$的距离.

如何定义赋范线性空间的张量积上的范数比较复杂, 读者可以自行查阅相关文献.

5.1.2 单位球与范数的等价

赋范线性空间中的**单位球**或单位球面具有重要意义. 单位球(面)在几何上相当于实数轴上的单位闭区间(其端点), 相当于平面上的单位圆盘(单位圆周), 以及空间中的单位球(面), 因此都是有界闭集(或更精确地, 紧集). 由于连续函数在有界闭集上一定有最大值和最小值, 故研究赋范线性空间上的连续函数或算子的一个重要技巧就是设法将函数的定义域限制或转移到单位球或单位球面上.

例 5.1.8. 按矩阵的 1-范数或∞-范数, 基本矩阵E_{ij}和置换矩阵都在单位球面上. 按照 F-范数, 所有基本矩阵仍在单位球面上, 但任何$n\geq 2$阶置换矩阵的范数均为$\sqrt{n}$, 从而都不在单位球面上.

本节最后将证明“有限维线性空间中的范数均等价”(定理 5.1.4), 而按同构定理, 对任何n维赋范线性空间V, 总可以定义$\mathbb{F}^n$的范数使得存在同构$\sigma\in \mathrm{Hom}(V,\mathbb{F}^n)$, 使得$\|\sigma(x)\|_{\mathbb{F}^n}=\|x\|_V$. 因此, 以下仅限于讨论$\mathbb{R}^n$中的单位球$B^n=\{x\in\mathbb{R}^n|\|x\|=1\}$与单位球面$S^{n-1}=\{x\in\mathbb{R}^n|\|x\|\leq 1\}$. 在拓扑学中, 单位球面$S^{n-1}$是$n$-维单位球$B^n$的边界, 因此是$n-1$维曲面.

记单位球B^n与单位球面S^{n-1}的体积与表面积为$V(B^n)$与$A(S^{n-1})$.

由多重积分学可得(或参见《数学的天空》[83])前几个单位球的表面积与体积如下.

$$
\begin{aligned}
&A(S^1) = 2\pi, && V(B^2) = \pi;\\
&A(S^2) = 4\pi, && V(B^3) = \frac{4}{3}\pi;\\
&A(S^3) = 2\pi^2, && V(B^4) = \frac{1}{2}\pi^2;\\
&A(S^4) = \frac{8}{3}\pi^2, && V(B^5) = \frac{8}{15}\pi^2;\\
&A(S^5) = \pi^3, && V(B^6) = \frac{1}{6}\pi^3;\\
&A(S^6) = \frac{16}{15}\pi^3, && V(B^7) = \frac{16}{105}\pi^3;\\
&A(S^7) = \frac{1}{3}\pi^4, && V(B^8) = \frac{1}{24}\pi^4;\\
&A(S^8) = \frac{32}{105}\pi^4, && V(B^9) = \frac{32}{945}\pi^4;\\
&\ldots\ldots
\end{aligned}
\tag{5.1.14}
$$

单位圆 S^1 是边长为 2 的正方形的内接圆, 它们的容积比为 $\pi:4$; 单位球面 S^2 是棱长为 2 的正方体的内接球, 此时的容积比为 $\pi:6$; 一般地, 当 $n\to\infty$ 时, S^n 与棱长为 2 的 n-维(超)正方体的容积比趋于 0. 换句话说, 单位球面在其外接(超)正方体内所占的空间随着维数增大而迅速减少并无限趋近于0. 请注意, 棱长为 2 的 n-维(超)正方体的体积为 2^n, 表面积为 $n2^n$, 均趋向于 ∞.

思考题.

1. 是否存在正整数N满足$V(B^N)\geq V(B^n), \forall n\geq 1$. 如果存在, 请问$N=$?
2. 当 $n\to\infty$ 时, 极限 $\lim\limits_{n\to\infty} A(S^n)=$?

命题 5.1.2. 设$n\geq 1$. 则

$$
A(S^n) = (n+1)V(B^{n+1}) \tag{5.1.15}
$$

$$
V(B^n) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} = \begin{cases} \dfrac{(2\pi)^{n/2}}{(n-2)!!}, & \text{如果 } n \text{ 是偶数} \\ \dfrac{2(2\pi)^{(n-1)/2}}{(n-2)!!}, & \text{如果 } n \text{ 是奇数} \end{cases} \tag{5.1.16}
$$

其中的$\Gamma(s)=\displaystyle\int_0^\infty e^{-x}x^{s-1}dx$是$\Gamma$-函数.

命题5.1.2表明, 当$n\to\infty$时, n-维单位球的表面积和体积都趋于0. 这个事实和公式(5.1.14)以及我们的直觉相差甚远, 著名的**庞加莱猜想**[84]的核心就是理解高维球面.

定理 5.1.3. 设$S\subseteq\mathbb{R}^n, a\in\mathbb{R}$是常数. 记$aS=\{ax|x\in S\}$. 则$aS$的体积$V(aS)$与$S$的体积$V(S)$的关系为$V(aS)=a^nV(S)$. 特别地, 对任意$\varepsilon>0$, $V((1-\varepsilon)B^n)=(1-\varepsilon)^nV(B^n)$.

[83]张跃辉, 李吉有, 朱佳俊著, 北京大学出版社, 2017年7月第一版.

[84]Jules Henri Poincaré(1854-1912), 法国数学家、物理学家、工程师、哲学家, 被称为数学界的最后一个全才. 法兰西学院院士(法国文人的最高荣誉), 法国科学院院长.

证　证明与第二章第八节2.8.3类似, 只需将S划分为充分小的(超)立方体即可. $\square$

定理5.1.3表明, 高维空间几何体的体积也集中于表面附近, 比如, 当$n \geq 500$时, 99%的体积集中在直径最外端1%的外壳上. 与此类似, 高维球的体积集中在任何赤道面附近的薄片. 这两个事实常用来作近似计算.

定义 5.1.2.　设常数$C \geq 1$, $\|\cdot\|_\alpha$与$\|\cdot\|_\beta$是线性空间V上的两种范数. 若对$\forall\, x \in V$, 都有

$$\frac{1}{C} \leq \frac{\|x\|_\alpha}{\|x\|_\beta} \leq C \tag{5.1.17}$$

则称$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$是等价的.

例 5.1.9.　在$\mathbb{R}^2$中, 1-范数, 欧几里得范数和∞-范数都是等价的, 因为

$$\frac{\|x\|_1}{2} \leq \|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq 2\|x\|_\infty.$$

由定义可以直接证明下面的结论.

命题 5.1.3.　设$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$是线性空间V上的两种范数. 则下列命题等价:
(1) $\|\cdot\|_\alpha$与$\|\cdot\|_\beta$等价.
(2) 存在正数C_1与C_2, 使得对$\forall\, x \in V$, 都有

$$\|x\|_\alpha \leq C_1\|x\|_\beta, \qquad \|x\|_\beta \leq C_2\|x\|_\alpha. \tag{5.1.18}$$

(3) $\|\cdot\|_\alpha$与$\|\cdot\|_\beta$具有相同的敛散性, 即设$x_1, x_2, \cdots, x_n, \cdots$是$V$中的(向量)序列, v是V中某给定向量. 则

$$\lim_{n\to\infty} \|x_n - v\|_\alpha = 0 \iff \lim_{n\to\infty} \|x_n - v\|_\beta = 0.$$

命题5.1.3中(3)说明等价的范数定义相同的极限.

引理 5.1.1.　有限维线性空间的范数是向量坐标的连续函数.

证　设V是n维线性空间, $\alpha_1, \alpha_2, \cdots, \alpha_n$是$V$的一组基, 则对$\forall\, x \in V$, x可唯一地表示成

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

显然V中任何一种范数$\|\cdot\|$都是坐标$x_1, x_2, \cdots, x_n$的函数, 故记

$$\|x\| = \varphi(x_1, x_2, \cdots, x_n).$$

要证对$\forall\, y = y_1\alpha_1 + y_2\alpha_2 + \cdots + y_n\alpha_n \in V$, 如果每个$|x_i - y_i| \to 0$, 则$|\varphi(x_1, x_2, \cdots, x_n) - \varphi(y_1, y_2, \cdots, y_n)| \to 0$. 由三角不等式知

$$\begin{aligned}
&|\varphi(x_1, x_2, \cdots, x_n) - \varphi(y_1, y_2, \cdots, y_n)| = \\
&= |\,\|x\| - \|y\|\,| \leq \|x - y\| = \|\sum_{i=1}^n (x_i - y_i)\alpha_i\| \\
&\leq \sum_{i=1}^n |x_i - y_i|\,\|\alpha_i\| \leq \sqrt{\sum_{i=1}^n \|\alpha_i\|^2}\sqrt{\sum_{i=1}^n |x_i - y_i|^2} \to 0
\end{aligned}$$

所以$\|x\|$是坐标$x_1, x_2, \cdots, x_n$的连续函数. $\square$

定理 5.1.4. 有限维线性空间中的任何两种范数都等价.

证 设$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$是V中任意两种范数. 当$x=0$时, 不等式(5.1.17)式显然成立. 设$x\neq 0$, 则$\|x\|_\beta\neq 0$. 由引理 5.1.1, $\|x\|_\alpha$, $\|x\|_\beta$都是$x_1, x_2, \cdots, x_n$的连续正函数, 因此

$$f(x_1,x_2,\cdots,x_n)=\frac{\|x\|_\alpha}{\|x\|_\beta}$$

也是$x_1, x_2, \cdots, x_n$的连续函数. 考虑在$\|\cdot\|_\beta$下的单位球面$S=\{x\in V\mid \|x\|_\beta=1\}$. 由于$S$为有界闭集, 且$S$上的点均不为零, 因此$f$在$S$上连续. 根据多元连续函数的性质, f在S上有最大值C_2与最小值$C_1>0$. 由于对$\forall\, x\neq 0$, 都有$x/\|x\|_\beta\in S$, 从而

$$C_1\leq \|\frac{x}{\|x\|_\beta}\|_\alpha=\frac{\|x\|_\alpha}{\|x\|_\beta}\leq C_2.$$

即有$C_1\|x\|_\beta\leq\|x\|_\alpha\leq C_2\|x\|_\beta$. □

注. 在无限维线性空间中, 两个向量范数未必等价, 见下例.

例 5.1.10. 设$V=C[0,1]$是闭区间$[0,1]$上全体连续函数构成的无限维实线性空间. 则

$$\|f\|_\infty=\max_{0\leq x\leq 1}|f(x)| \tag{5.1.19}$$

与

$$\|f\|_1=\int_0^1|f(x)|dx \tag{5.1.20}$$

均是V中的范数. 考虑V中的函数列(请画草图以便于理解)f_n, 其中$f_1=1$, 而对每个$n\geq 2$,

$$f_n=\begin{cases} 2nx, & 0\leq x\leq \frac{1}{2n};\\ -2nx+2, & \frac{1}{2n}<x\leq\frac{1}{n};\\ 0, & x>\frac{1}{n}.\end{cases}$$

则易知$\lim\limits_{n\to\infty}\|f_n\|_1=0$. 但$\lim\limits_{n\to\infty}\|f_n\|_\infty=1$, 故由命题 5.1.3知此二范数不等价.

第二节 赋范矩阵空间

在例 5.1.4中将$m\times n$矩阵看成mn维向量而定义了矩阵的几个范数. 但方阵还有自身的特点, 即矩阵乘法, 因此有必要考虑$\|AB\|$与$\|A\|\times\|B\|$的关系. 由于存在$A\neq 0, B\neq 0$但$AB=0$, 故$\|AB\|\geq\|A\|\times\|B\|$是不可能的, 故有下面的定义.

定义 5.2.1. 设$\|\!|\cdot|\!\|$为$\mathbb{F}^{n\times n}$上一个非负的实函数, 若$\|\!|\cdot|\!\|$是$\mathbb{F}^{n\times n}$的向量范数, 且对$\forall\, A,B\in\mathbb{F}^{n\times n}$, 有

$$(\text{次乘性})\qquad \|\!|AB|\!\|\leq\|\!|A|\!\|\,\|\!|B|\!\| \tag{5.2.1}$$

则称$\|\!|\cdot|\!\|$为$\mathbb{F}^{n\times n}$上的一个**矩阵范数**.

例 5.2.1. 容易验证, 例5.1.4中定义的范数均满足次乘性条件(5.2.1), 因此均是矩阵范数. 此处仅对F-范数验证, 其余见习题11. 设$A=(a_{ij})_{n\times n}$, $B=(b_{ij})_{n\times n}$. 则

$$
\begin{aligned}
\|AB\|_F &= \sqrt{\sum_{i,j=1}^{n}|\sum_{k=1}^{n}a_{ik}b_{kj}|^2} \leq \sqrt{\sum_{i,j=1}^{n}(\sum_{k=1}^{n}|a_{ik}|^2)(\sum_{k=1}^{n}|b_{kj}|^2)} \\
&= \sqrt{\sum_{i,j=1}^{n}|a_{ij}|^2}\sqrt{\sum_{i,j=1}^{n}|b_{ij}|^2} = \|A\|_F\,\|B\|_F,
\end{aligned}
$$

即不等式(5.2.1)成立. □

今后把矩阵的1-范数, F-范数和∞-范数改记为矩阵范数$\|\!|\cdot|\!\|_1, \|\!|\cdot|\!\|_F, \|\!|\cdot|\!\|_\infty$等等.

例 5.2.2. 由次乘性可知, $\|\!|A^k|\!\| \leq \|\!|A|\!\|^k$. 因此, 如果$\|\!|A|\!\|<1$, 则数列$\{\|\!|A^k|\!\|\}$收敛于0.

例 5.2.3. 在任何矩阵范数$\|\!|\cdot|\!\|$之下, 次乘性保证单位矩阵的范数$\|\!|I|\!\| \geq 1$. 这是因为$\|\!|I|\!\| = \|\!|I^2|\!\| \leq \|\!|I|\!\|^2$. 注意, 如果没有次乘性, 则单位矩阵的范数可以为任何正数!(为什么?)

定义 5.2.2. 设$\|\!|\cdot|\!\|$与$\|\cdot\|$分别是$\mathbb{F}^{n\times n}$及$\mathbb{F}^n$的矩阵范数和向量范数. 若对$\forall\, A\in\mathbb{F}^{n\times n}$及$x\in\mathbb{F}^n$, 都有

$$
\|Ax\| \leq \|\!|A|\!\|\,\|x\| \tag{5.2.2}
$$

则称矩阵范数$\|\!|\cdot|\!\|$与范数$\|\cdot\|$是相容的.

定理 5.2.1. 设$(V,\|\cdot\|)$是n维赋范线性空间, $A\in\mathbb{F}^{n\times n}$. 则

$$
\|\!|A|\!\| = \sup_{0\neq x\in\mathbb{F}^n}\frac{\|Ax\|}{\|x\|} \tag{5.2.3}
$$

定义了一个与范数$\|\cdot\|$相容的矩阵范数, 称为由$\|\cdot\|$诱导的矩阵范数或**算子范数**.

证 齐次性、相容性与正定性都是显然的.
注意

$$
\sup_{0\neq x\in\mathbb{F}^n}\frac{\|Ax\|}{\|x\|} = \sup_{0\neq x\in\mathbb{F}^n}\|A\frac{x}{\|x\|}\| = \sup_{\|x\|=1}\|Ax\|.
$$

因赋范线性空间$(\mathbb{F}^n,\|\cdot\|)$的单位闭球或单位球面皆为有界闭集, 而$\|Ax\|$为x的连续函数, 故在单位球或单位球面上取得最大值, 所以上式中的“sup”可以换为“max”, 即

$$
\|\!|A|\!\| = \max_{\|x\|=1}\|Ax\| \tag{5.2.4}
$$

对$\forall\, A,B\in\mathbb{F}^{n\times n}$, 有

$$
\begin{aligned}
\|\!|A+B|\!\| &= \max_{\|x\|=1}\|(A+B)x\| \leq \max_{\|x\|=1}(\|Ax\|+\|Bx\|) \\
&\leq \max_{\|x\|=1}\|Ax\| + \max_{\|x\|=1}\|Bx\| = \|\!|A|\!\| + \|\!|B|\!\|.
\end{aligned}
$$

所以三角不等式成立. 设$\|ABx\|$在单位球面的最大值点为y, 即$\|\!|AB|\!\| = \|ABy\|$. 如果$\|By\| = 0$, 即$By = 0$, 则显然$|\!|\!|AB|\!|\!| = 0 \le |\!|\!|A|\!|\!|\,|\!|\!|B|\!|\!|$, 即次乘性成立; 如果$\|By\| \ne 0$, 则

$$
\begin{aligned}
|\!|\!|AB|\!|\!| &= \frac{\|ABy\|}{\|By\|} \cdot \|By\| \le \max_{\substack{0\ne x\in\mathbb{F}^n \\ \|Bx\|\ne 0}} \frac{\|A(Bx)\|}{\|Bx\|} \cdot \max_{\|x\|=1} \|Bx\| \\
&\le \sup_{0\ne x\in\mathbb{F}^n} \frac{\|Ax\|}{\|x\|} \cdot |\!|\!|B|\!|\!| = |\!|\!|A|\!|\!|\,|\!|\!|B|\!|\!|.
\end{aligned}
$$

即得次乘性. □

注1. 可以证明(留作习题), 例 5.2.1中给出的矩阵的1-范数恰好是向量的1-范数的诱导范数, 而矩阵的∞-范数恰好是向量的∞-范数的诱导范数(这是我们将矩阵的1-范数和∞-范数没有仿照向量的相应范数定义的原因, 见注2和例 5.1.4后的注).

注2. 如果仿照向量的1-范数与∞-范数定义矩阵的1-范数和∞-范数, 则它们不与是向量的1-范数与∞-范数诱导的范数, 实际上, 这样定义的∞-范数不是矩阵范数, 请读者自行证明之.

推论 5.2.1. (1) $\mathbb{F}^{n\times n}$上任意两种矩阵范数均等价.
(2) 对于$\mathbb{F}^n$上每种向量范数, 都存在$\mathbb{F}^{n\times n}$上与它相容的矩阵范数.
(3) 对于$\mathbb{F}^{n\times n}$上每种矩阵范数, 都存在$\mathbb{F}^n$上与它相容的向量范数.

证 只需证明(3). 设$|\!|\!|\cdot|\!|\!|$是$\mathbb{F}^{n\times n}$上的一种矩阵范数, 对每个$x \in \mathbb{F}^n$, 定义

$$
\|x\| = |\!|\!|xJ^T|\!|\!| \tag{5.2.5}
$$

其中$J = \sum_{i=1}^n e_i$是全1向量矩阵(实际上$xJ^T = (x, x, \cdots, x)$是每列均为x的矩阵). 则公式(5.2.5)定义了一个与矩阵范数$|\!|\!|\cdot|\!|\!|$相容的向量范数, 这是因为对任意$A \in \mathbb{F}^{n\times n}$, 有

$$
\|Ax\| = |\!|\!|A(xJ^T)|\!|\!| = |\!|\!|A(xJ^T)|\!|\!| \le |\!|\!|A|\!|\!| \cdot |\!|\!|xJ^T|\!|\!| = |\!|\!|A|\!|\!| \cdot \|x\|
$$

. □

例 5.2.4. 公式(5.2.5)定义的范数称为矩阵范数$|\!|\!|\cdot|\!|\!|$ 诱导的向量范数. 矩阵的极大列和范数$|\!|\!|\cdot|\!|\!|_1$诱导的向量范数恰好是向量的1-范数! 请思考: 矩阵的∞-范数与F-范数诱导的向量范数是否分别为∞-范数与2-范数?

定义 5.2.3. 设U, V是任意维实或复赋范线性空间, $\sigma \in \mathrm{Hom}(U, V)$. 定义

$$
\|\sigma\| = \sup_{0\ne x\in U} \frac{\|\sigma(x)\|}{\|x\|} = \max_{\|x\|=1} \|\sigma(x)\| \tag{5.2.6}
$$

则上式是线性空间$\mathrm{Hom}(U, V)$上的与U中范数相容的向量范数, 即$\|\sigma(x)\| \le \|\sigma\|\|x\|$. 如果$\|\sigma\|$是常数, 则称$\sigma$是**有界算子**.

注1. 如果$U = V$, 则容易证明该范数还是矩阵范数, 即满足次乘性$\|\sigma\tau\| \le \|\sigma\|\|\tau\|$. 一般线性变换的范数均指由公式(5.2.6)定义的范数, 它实际上是两个赋范线性空间U与V的范数诱导的算子范数.

注2. 如果U是有限维的, 则由公式(5.2.6)定义的算子范数$\|\sigma\|$一定有界(为什么?), 此时称σ是**有界算子**(变换). 但若U是无限维的, 则此范数可能是无穷大, 此时称σ是无界算子.

例 5.2.5. 内积空间的非零正交投影变换的范数均为1.

有了线性变换的范数概念, 就可以仿照函数的连续性来定义线性变换的连续性了.

定义 5.2.4. 设U, V是实或复赋范线性空间(有限维或无限维), $\sigma \in \mathrm{Hom}(U, V)$. 如果对任意给定的正数$\varepsilon > 0, \exists \delta > 0$, 使得对任意$x, y \in U$, $\|x - y\| < \delta$, 均有

$$\|\sigma(x) - \sigma(y)\| < \varepsilon,$$

则称σ是连续的.

引理 5.2.1. 赋范线性空间的线性变换是连续的 $\Longleftrightarrow$ 它在一点处连续 $\Longleftrightarrow$ 它在0点连续.

证 设$\sigma \in \mathrm{Hom}(U, V)$. 只需证明$\sigma$在0点连续则处处连续. 设$\sigma$在0处连续, 则对任意$u \in U$, 有

$$\|\sigma(u + \Delta u) - \sigma(u)\| = \|\sigma(u + \Delta u - u)\| = \|\sigma(\Delta u)\| \to 0(\Delta u \to u_0),$$

故σ在u处连续. □

定理 5.2.2. 赋范线性空间的线性变换是连续的 $\Longleftrightarrow$ 它是有界的.

证 设$\sigma \in \mathrm{Hom}(U, V)$. 由$\|\sigma(u)\| \leq \|\sigma\|\|u\|$以及引理 5.2.1即知有界性蕴含连续性.

反过来, 设σ连续但无界, 则存在序列$u_n \in U$使得

$$\sigma(u_n) \geq n\|u_n\|.$$

因此

$$\sigma(\frac{u_n}{n\|u_n\|}) \geq 1 \tag{5.2.7}$$

由于$\dfrac{u_n}{n\|u_n\|} \to 0$而$\sigma$连续, 故(5.2.7)与引理 5.2.1矛盾. 所以, σ必有界. □

推论 5.2.2. 有限维赋范线性空间的线性变换均是连续的.

实际上, 我们知道有限维线性空间的一个线性变换由若干个线性函数组成, 而线性函数都连续, 所以有限维线性空间的线性变换的连续性是自然的.

无限维赋范线性空间总存在无界线性变换. 比如, 设$V = \mathbb{R}[x]$的范数由公式(5.1.19)定义(见例 5.1.10), 则$\sigma : \sum\limits_{n\geq 0} a_n x^n \mapsto \sum\limits_{n\geq 0} n a_n$是$V$上的线性泛函. 注意$\sigma \in \mathrm{Hom}(V, \mathbb{R})$无界(为什么?)从而不连续.

由于有限维赋范线性空间上的范数都是等价的, 故任意范数导出的极限理论与欧几里得范数的相应理论完全相同, 由此导出的微积分理论与高等数学的基本内容完全重合. 然而, 无限维赋范线性空间上存在不等价的范数, 故其上的数学理论自然更加丰富多彩, 憾其与矩阵理论联系不深, 强烈建议读者一探究竟, 可参考[85].

思考题

1. 在$\mathbb{R}^2$中, 中心在原点的非等边矩形是否可以是单位圆? 中心在原点的正三角形与双曲线呢?
2. 三角不等式中的等号何时成立? 是否存在范数使得三角不等式总是等式?

[85]张恭庆, 林源渠, 泛函分析讲义(第二版)(上), 北京大学出版社, 2021年05月.

3. 两个范数的乘积是否仍是范数?

4. 内积可以诱导范数. 哪些p-范数可以诱导内积, 哪些不能?

5. 矩阵A与其共轭转置A^*的矩阵范数有何联系? 可逆矩阵与其逆矩阵的矩阵范数有何联系? 线性变换与其伴随变换的范数有何联系?

6. 矩阵范数中次乘性的等号何时成立? 是否存在矩阵范数使得次乘性中的等号永远成立?

7. 试定义一种不同于本书的矩阵范数?

8. 能否在赋范线性空间中定义合理的角度? 研究 1-范数和∞-范数的单位圆中的几个角, 它们是直角吗?

9. 设V是任意无限维赋范线性空间. 试构造V上一个不连续的线性变换.

第三节 矩阵级数与矩阵函数

5.3.1 矩阵的幂收敛

向量与矩阵序列敛散性的理论与数学分析或高等数学完全类似, 请读者注意比较.

定义 5.3.1. 设$(V, \|\cdot\|)$是n维赋范线性空间, $x_1, x_2, \cdots, x_k, \cdots$是$V$的一个向量序列, α是V的一个固定向量. 如果

$$\lim_{k\to\infty} \|x_k - \alpha\| = 0$$

则称向量序列$x_1, x_2, \cdots, x_k, \cdots$在$\|\cdot\|$收敛, 且$\alpha$是该序列的极限, 记作

$$\lim_{k\to\infty} x_k = \alpha \quad \text{或} \quad x_k \to \alpha.$$

不收敛的向量序列称为发散的.

例 5.3.1. 设$V = \mathbb{R}^n$是普通赋范线性空间(取欧几里得范数), 则定义 5.3.1中的敛散性与高等数学中的相应概念完全一致.

普通赋范线性空间$V = \mathbb{F}^n$中的敛散性可以具体地描述如下(证明见习题 25):

引理 5.3.1. 设$\{x^{(k)}\}$是$\mathbb{F}^n$中的向量序列, $x^{(k)} = (x_1^{(k)}, \cdots, x_n^{(k)})$, $k = 1, 2, \cdots$. 设$x = (x_1, \cdots, x_n)$是$\mathbb{F}^n$中的固定向量. 则$\lim\limits_{k\to\infty} x^{(k)} = x \iff \lim\limits_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \cdots, n.$

引理 5.3.1表明, 向量序列的极限实际上是坐标序列的极限, 也就是说向量序列收敛 $\iff$ 该序列的向量的每个坐标构成的数列(共有n个)均收敛. 所以, 向量序列的收敛可以称为**"按坐标收敛"**. 另外, 此定义显然可以推广到所有无穷数列构成的无限维空间.

定理 5.3.1. 设$\{x^{(k)}\}$为$\mathbb{F}^n$的向量序列, $\|\cdot\|$为$\mathbb{F}^n$中的任一向量范数, 则

$$\lim_{k\to\infty} x^{(k)} = x \iff \lim_{k\to\infty} \|x^{(k)} - x\| = 0.$$

证 由向量范数的等价性, 定理中的结论只要对一种向量范数成立, 则对任何一种向量范数都成立. 故就向量范数$\|\cdot\|_\infty$来证明即可, 因此定理成立.

注. 定理 5.3.1可以等价地描述为向量序列按坐标收敛于向量x $\iff$ 它**按范数收敛**于x.

将n阶矩阵看作n^2维向量即可定义矩阵序列的敛散性. 为方便读者, 我们将其单独列出.

定义 5.3.2. 设$A_m = (a_{ij}^{(m)})_{n\times n} \in \mathbb{F}^{n\times n}$, $m = 1, 2, \cdots$. 若对任意的i, j, 有$\lim\limits_{m\to\infty} a_{ij}^{(m)} = a_{ij}$, $1 \le i, j \le n$, 则称矩阵序列$\{A_m\}$收敛于矩阵$A = (a_{ij})_{n\times n}$, 记为$\lim\limits_{m\to\infty} A_m = A$. 否则, 称$\{A_m\}$为发散的.

上述定义可称作矩阵序列**按元素收敛**或按坐标收敛.

由定理 5.3.1可知, 矩阵序列按坐标收敛与按范数收敛等价, 即有下述推论

推论 5.3.1. 设$\{A_m\}$为$\mathbb{F}^{n\times n}$的矩阵序列, $\|\cdot\|$为$\mathbb{F}^{n\times n}$的任意向量范数, 则

$$\lim_{m\to\infty} A_m = A \iff \lim_{m\to\infty} \|A_m - A\| = 0.$$

注. 推论 5.3.1中的“向量范数”不必是“矩阵范数”.

例 5.3.2. 设$A_m = \begin{pmatrix} m\sin\frac{1}{m} & (1-\frac{1}{m})^{2m} \\ 1 & m^2(1-\cos\frac{1}{m}) \end{pmatrix}$, 则$\lim\limits_{m\to\infty} A_m = \begin{pmatrix} 1 & e^{-2} \\ 1 & \frac{1}{2} \end{pmatrix}$.

下面是矩阵序列极限的四则运算法则, 证明是直接的.

性质 5.3.1. (1) 若$A_m \to A$, $B_m \to B$, $\{a_m\} \to a$, $\{b_m\} \to b$, a_m, b_m, a, $b \in \mathbb{F}$, 则

$$\lim_{m\to\infty} (a_m A_m + b_m B_m) = aA + bB, \qquad \lim_{m\to\infty} A_m B_m = AB.$$

特别地, 若P是可逆矩阵, 则

$$\lim_{m\to\infty} P^{-1} A_m P = P^{-1} A P.$$

(2) 若$\lim\limits_{m\to\infty} A_m = A$, 则对$\mathbb{F}^{n\times n}$中任意范数$\|\cdot\|$, $\|A_m\|$有界.

(3) 若$A_m \to A$, 且A_m^{-1}及A^{-1}存在, 则

$$\lim_{m\to\infty} A_m^{-1} = A^{-1}.$$

定义 5.3.3. 设$A \in \mathbb{F}^{n\times n}$, 则由$A$的幂可得一矩阵序列:

$$I,\ A,\ A^2,\ A^3, \cdots \tag{5.3.1}$$

若矩阵序列(5.3.1)收敛, 则称矩阵A**幂收敛**.

由矩阵序列极限的四则运算法则立即可得下述幂收敛的等价条件.

命题 5.3.1. 设矩阵A与B相似, 则A幂收敛 $\iff$ B幂收敛. 特别地, 矩阵A幂收敛 $\iff$ 其 Jordan 标准形幂收敛.

例 5.3.3. 对角矩阵幂收敛 $\iff$ 其对角元素为 1 或绝对值小于 1.

一般地有下面的定理.

定理 5.3.2. 设$A \in \mathbb{C}^{n\times n}$. 则$A$幂收敛 $\iff$ A的任一特征值λ满足: $|\lambda| \le 1$, 并且, 若$|\lambda| = 1$, 则$\lambda = 1$且对角线元素为1的 Jordan 块都是一阶的.

证 A幂收敛 $\iff$ 其Jordan标准形J幂收敛 $\iff$ J的每个Jordan块幂收敛, 而每个$k \ge 2$阶Jordan块幂收敛 $\iff$ 对角元素的绝对值小于 1. □

注. “对角线元素为1的 Jordan 块都是一阶的”的含义是特征值 1 的代数重数等于几何重数. 等价地, 最小多项式中的因式$x-1$的次数为 1.

下面介绍两个矩阵谱半径与矩阵的范数(不必是矩阵范数)之间的两个基本关系.

定理 5.3.3. 设$A \in \mathbb{C}^{n\times n}$, $\|\cdot\|$为$\mathbb{C}^{n\times n}$的任意一种范数, 则$\rho(A) \le \|A\|$.

证 作矩阵
$$B = \frac{1}{\|A\| + \varepsilon} A,$$
其中ε为任意正实数. 则
$$\|B\| = \frac{1}{\|A\| + \varepsilon}\|A\| < 1.$$
于是, $\lim\limits_{m\to\infty}\|B^m\| = 0$, 由推论 5.3.1, $\lim\limits_{m\to\infty} B^m = 0$. 于是由定理 5.3.2, B的所有特征值的模都小于或等于 1. 即
$$\frac{1}{\|A\| + \varepsilon}\,|\lambda\,| \le 1,$$
其中λ为A的任一特征值. 于是, $|\lambda| \le \|A\| + \varepsilon$. 因为$\varepsilon$为任意正实数, 所以$\rho(A) \le \|A\|$. □

例 5.3.4. 设两个离散随机变量的联合分布矩阵是方阵, 则该矩阵幂收敛. 这是因为该矩阵的所有元素之和为1, 因此其 1-范数必定小于 1 或者该矩阵仅有一个元素为 1 其余均为 0, 因此也幂收敛.

由定理 5.3.3只是说明矩阵的谱半径不超过矩阵的任何范数, 实际上谱半径与矩阵的任意范数有更为深刻的联系, 此即下述著名的"**Gelfand**[86] **谱半径定理**".

定理 5.3.4. 设$A \in \mathbb{C}^{n\times n}$, $\|\cdot\|$为$\mathbb{C}^{n\times n}$的任意一种范数, 则
$$\rho(A) = \lim_{k\to\infty}\|A^k\|^{1/k} \tag{5.3.2}$$

证 仿定理 5.3.3的证明, 对$\forall \varepsilon > 0$, 作矩阵
$$\widehat{A} = \frac{1}{\rho(A) + \varepsilon} A,$$
则
$$\rho(\widehat{A}) < 1.$$
故存在正数$c > 0$, 使得$\rho(\widehat{A}) = 1 - c$. 因此$\forall k \ge 1, \rho(\widehat{A}^k) = (1-c)^k$, 从而有$\rho(\widehat{A}^k) \to 0$. 故再由推论 5.3.1, 知$\|\widehat{A}^k\| \to 0$, 从而对充分大的$k$, 有$\|\widehat{A}^k\| < 1$, 即$\|A^k\| < (\rho(A)+\varepsilon)^k$, 因此
$$\|A^k\|^{1/k} < \rho(A) + \varepsilon, k \to \infty \tag{5.3.3}$$

反过来, 作矩阵
$$\widetilde{A} = \frac{1}{\rho(A) - \varepsilon} A,$$
其中$\varepsilon \ne \rho(A)$为任意正实数. 则
$$\rho(\widetilde{A}) > 1.$$

[86]I. Gelfand(1913-2009), 苏-美著名数学家, 首届Wolf奖得主.

故存在正数$d > 0$, 使得$\rho(\widetilde{A}) = 1 + d$. 因此$\forall k \geq 1, \rho(\widetilde{A}^k) = (1+d)^k$, 从而$\rho(\widetilde{A}^k) \to \infty$, 因此$\|A^k\| > (\rho(A) - \varepsilon)^k$, 即

$$\|A^k\|^{1/k} > \rho(A) - \varepsilon, k \to \infty \tag{5.3.4}$$

结合公式(5.3.3)与(5.3.4), 即得$\rho(A) = \lim\limits_{k\to\infty} \|A^k\|^{1/k}$. □

例 5.3.5. 非零幂零矩阵的谱半径为0, 但其范数可以是任意大的正数. 此并不与Gelfand公式矛盾, 因为对充分大的$k, A^k = 0$.

Gelfand 谱半径定理将在第六章证明著名的Perron-Frobenius 定理时发挥重要的作用.

思考题

设$\lim\limits_{k\to\infty} A_k = A$. 如果$A_k$均为正定矩阵(正规矩阵、可逆矩阵、幂零矩阵、幂等矩阵), A是否也是正定矩阵(正规矩阵、可逆矩阵、幂零矩阵、幂等矩阵)?

5.3.2 矩阵级数

定义 5.3.4. 设$\{A_k \mid k=0,1,2,\cdots\}$为一个矩阵序列. 对$n \geq 0$, 令$S_n = \sum\limits_{k=0}^{n} A_k$. 若存在矩阵$S$使得$\lim\limits_{n\to\infty} S_n = S$, 则称矩阵级数$\sum\limits_{k=0}^{\infty} A_k$收敛于$S$, 记作$\sum\limits_{k=0}^{\infty} A_k = S$. 否则, 称$\sum\limits_{k=0}^{\infty} A_k$发散.

矩阵级数有着与普通数项级数相类似的性质.

性质 5.3.2. (1) 若$\sum\limits_{k=0}^{\infty} A_k$收敛, 则$\lim\limits_{k\to\infty} A_k = 0$.

(2) 若$\sum\limits_{k=0}^{\infty} A_k = A$, $\sum\limits_{k=0}^{\infty} B_k = B, a \in \mathbb{F}$, 则

$$\sum_{k=0}^{\infty}(A_k + B_k) = A + B, \quad \sum_{k=0}^{\infty} aA_k = aA.$$

矩阵 Jordan 块的幂级数求和的基本定理如下.

定理 5.3.5. 设J为对角线元素为λ的n阶 Jordan 块, $f(t) = \sum\limits_{k=0}^{\infty} a_k t^k$是收敛半径为$r$的幂级数. 则当$|\lambda| < r$时, 矩阵幂级数$\sum\limits_{k=0}^{\infty} a_k J^k$是收敛的, 且其和为矩阵

$$\begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2}f''(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{pmatrix}.$$

证 令$S_m=\sum\limits_{k=0}^{m}a_kJ^k$, $S_m(\lambda)=\sum\limits_{k=0}^{m}a_k\lambda^k$. 由第三章定理3.2.3(Jordan块的高次幂)知,

$$J^k=\begin{pmatrix}\lambda^k & C_k^1\lambda^{k-1} & C_k^2\lambda^{k-2} & \cdots & C_k^{n-1}\lambda^{k-n+1}\\ & \lambda^k & C_k^1\lambda^{k-1} & \cdots & C_k^{n-2}\lambda^{k-n+2}\\ & & \ddots & \ddots & \vdots\\ & & & \ddots & C_k^1\lambda^{k-1}\\ & & & & \lambda^k\end{pmatrix},$$

其中, 若$s>k$, 则$C_k^s=0$. 所以

$$S_m=\begin{pmatrix}\sum\limits_{k=0}^{m}a_k\lambda^k & \sum\limits_{k=0}^{m}a_kC_k^1\lambda^{k-1} & \sum\limits_{k=0}^{m}a_kC_k^2\lambda^{k-2} & \cdots & \sum\limits_{k=0}^{m}a_kC_k^{n-1}\lambda^{k-n+1}\\ & \sum\limits_{k=0}^{m}a_k\lambda^k & \sum\limits_{k=0}^{m}a_kC_k^1\lambda^{k-1} & \cdots & \sum\limits_{k=0}^{m}a_kC_k^{n-2}\lambda^{k-n+2}\\ & & \ddots & \ddots & \vdots\\ & & & \ddots & \sum\limits_{k=0}^{m}a_kC_k^1\lambda^{k-1}\\ & & & & \sum\limits_{k=0}^{m}a_k\lambda^k\end{pmatrix}$$

$$=\begin{pmatrix}S_m(\lambda) & S_m'(\lambda) & \frac{1}{2!}S_m''(\lambda) & \cdots & \frac{1}{(n-1)!}S_m^{(n-1)}(\lambda)\\ & S_m(\lambda) & S_m'(\lambda) & \cdots & \frac{1}{(n-2)!}S_m^{(n-2)}(\lambda)\\ & & \ddots & \ddots & \vdots\\ & & & \ddots & S_m'(\lambda)\\ & & & & S_m(\lambda)\end{pmatrix}.$$

因$S_m(t)=\sum\limits_{k=0}^{m}a_kt^k$的收敛半径为$r$, 且$|\lambda|<r$, 所以$S_m(\lambda),S_m'(\lambda),S_m''(\lambda),\cdots,S_m^{(n-1)}(\lambda)$ 皆收敛, 且$\lim\limits_{m\to\infty}S_m(\lambda)=f(\lambda)$, $\lim\limits_{m\to\infty}S_m^{(k)}(\lambda)=f^{(k)}(\lambda)$, $k=1,2,\cdots,n-1$. 因此,

$$\lim_{m\to\infty}S_m=\sum_{k=0}^{\infty}a_kJ^k=\begin{pmatrix}f(\lambda) & f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda)\\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda)\\ & & \ddots & \ddots & \vdots\\ & & & \ddots & f'(\lambda)\\ & & & & f(\lambda)\end{pmatrix}.$$

例 5.3.6. 设$f(t)=\sum\limits_{k=0}^{\infty}(\frac{t}{3})^k$, 求$f(J)$, 其中

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

解 $f(t)=(1-\frac{t}{3})^{-1}$, 其收敛半径为3. 因$J$的特征值2落在$f(t)$的收敛域内, 所以$f(J)=\sum\limits_{k=0}^{\infty}(\frac{J}{3})^k$是收敛的且

$$f(J)=\begin{pmatrix} f(2) & f'(2) & \frac{1}{2!}f''(2) & \frac{1}{3!}f'''(2) \\ & f(2) & f'(2) & \frac{1}{2!}f''(2) \\ & & f(2) & f'(2) \\ & & & f(2) \end{pmatrix}.$$

因

$$f'(t)=\frac{1}{3}(1-\frac{t}{3})^{-2},\ f''(t)=\frac{2}{9}(1-\frac{t}{3})^{-3},\ f'''(t)=\frac{2}{9}(1-\frac{t}{3})^{-4},$$

所以

$$f(2)=3,\quad f'(2)=3,\quad f''(2)=6,\quad f'''(2)=18.$$

因此

$$f(J)=\begin{pmatrix} 3 & 3 & 3 & 3 \\ & 3 & 3 & 3 \\ & & 3 & 3 \\ & & & 3 \end{pmatrix}.$$

由定理5.3.5即可得到矩阵幂级数求和的基本定理.

定理 5.3.6. (Lagrange-Sylvester 定理) 设$f(t)=\sum\limits_{k=0}^{\infty}a_kt^k$, 它的收敛半径为$r$. 设矩阵$A$的Jordan标准形为

$$J=\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{pmatrix},\quad J_i=\begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix},\quad 1\le i\le s,$$

其变换矩阵为P, 即$A=PJP^{-1}$. 若对所有的$i=1,2,\cdots,s$, 都有$|\lambda_i|<r$, 则矩阵幂级

数$\sum_{k=0}^{\infty} a_k A^k$收敛, 其和为

$$f(A)=\sum_{k=0}^{\infty} a_k A^k = P\begin{pmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_s) \end{pmatrix} P^{-1},$$

其中

$$f(J_i)=\begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(n_i-2)!}f^{(n_i-2)}(\lambda_i) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{pmatrix}, \quad i=1,2,\cdots,s.$$

推论 5.3.2. 设$S_m=\sum_{k=0}^{m} A^k$, 则$\{S_m\}$收敛$\Longleftrightarrow \lim_{k\to\infty} A^k=0$.

推论 5.3.3. (Neumann[87] 引理) 设矩阵A的某个矩阵范数小于1, 则A幂收敛, $I-A$可逆且

$$(I-A)^{-1}=I+A+A^2+\cdots+A^m+\cdots=\sum_{k=0}^{\infty} A^k. \tag{5.3.5}$$

例 5.3.7. 已知$f(t)=2-t+2t^3$, 求$f(A)$, 其中

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

解 由于$f(t)$次数较低, 读者可以尝试直接计算本题. 但注意多项式是最简单的幂级数, 因此利用定理5.3.6更为简洁. 因$f(1)=3$, $f'(1)=5$, $f(2)=16$, $f'(2)=23$, $f''(2)=24$, 所以

$$f(A)=\begin{pmatrix} 3 & 5 & & & \\ & 3 & & & \\ & & 16 & 23 & 12 \\ & & & 16 & 23 \\ & & & & 16 \end{pmatrix}.$$

思考题

1. 若$\lim_{n\to\infty} A_nB_n$存在, 是否$\lim_{n\to\infty} A_n$, $\lim_{n\to\infty} B_n$一定存在? 为什么?

2. 设A, B均幂收敛, $A+B, AB$幂收敛吗?

[87]Carl Gottfried Neumann (1832-1925), 德国数学家, 等式(5.3.5)中的算子幂级数称为Neumann 级数.

5.3.3 矩阵函数

熟知下面三个幂级数定义的复变函数($z \in \mathbb{C}$)

$$e^z = \sum_{m=0}^{\infty} \frac{z^m}{m!},$$

$$\sin z = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{z^{2m-1}}{(2m-1)!},$$

$$\cos z = 1 + \sum_{m=1}^{\infty} (-1)^m \frac{z^{2m}}{(2m)!}.$$

在复平面上都是收敛的. 故由 Lagrange-Sylvester 定理知下列矩阵幂级数

$$\sum_{m=0}^{\infty} \frac{A^m}{m!}, \quad \sum_{m=1}^{\infty} (-1)^{m-1} \frac{A^{2m-1}}{(2m-1)!}, \quad I + \sum_{m=1}^{\infty} (-1)^m \frac{A^{2m}}{(2m)!},$$

都收敛. 它们的和(矩阵)分别用记号e^A, $\sin A$, $\cos A$来表示, 并分别称为方阵A的指数函数, 正弦函数及余弦函数. 同样地, 由

$$\ln(1+z) = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{z^m}{m}, \quad |z| < 1,$$

$$(1+z)^a = 1 + \sum_{m=1}^{\infty} \frac{a(a-1)\cdots(a-m+1)}{m!} z^m, \quad |z| < 1, \quad a \text{ 为任意实数},$$

可定义方阵函数

$$\ln(I+A) = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{A^m}{m}, \quad \rho(A) < 1,$$

$$(I+A)^a = I + \sum_{m=1}^{\infty} \frac{a(a-1)\cdots(a-m+1)}{m!} A^m, \quad \rho(A) < 1,$$

这里$\rho(A)$为A的谱半径.

例 5.3.8. $e^{\lambda I} = e^{\lambda} I$, $\sin(\lambda I) = (\sin \lambda) I$, $\cos(\lambda I) = (\cos \lambda) I$.

根据 Lagrange-Sylvester 定理, 我们将矩阵函数e^A与$\sin A$在n阶 Jordan块处的函数值写成以下命题, 以方便读者, 其余几个矩阵函数可类似计算, 此处略去.

命题 5.3.2. 设J是特征值为λ的n阶 Jordan块, 则

$$e^J = e^\lambda \begin{pmatrix} 1 & 1 & \frac{1}{2!} & \frac{1}{3!} & \cdots & \frac{1}{(n-1)!} \\ & 1 & 1 & \frac{1}{2!} & \cdots & \frac{1}{(n-2)!} \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & 1 & 1 & \frac{1}{2!} \\ & & & & 1 & 1 \\ & & & & & 1 \end{pmatrix} \tag{5.3.6}$$

$$\sin J = \begin{pmatrix} \sin\lambda & \cos\lambda & -\frac{\sin\lambda}{2!} & -\frac{\cos\lambda}{3!} & \cdots & \frac{\sin[((n-1)\pi)/2+\lambda]}{(n-1)!} \\ & \sin\lambda & \cos\lambda & -\frac{\sin\lambda}{2!} & \cdots & \frac{\sin[((n-2)\pi)/2+\lambda]}{(n-2)!} \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & \sin\lambda & \cos\lambda & -\frac{\sin\lambda}{2!} \\ & & & & \sin\lambda & \cos\lambda \\ & & & & & \sin\lambda \end{pmatrix} \tag{5.3.7}$$

例 5.3.9. 设$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, 求$e^A$, $\sin A$, $\cos A$.

解 $\sigma(A) = \{1, 2\}$. 因此A与对角矩阵$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$相似. 计算对应的特征向量为

$$\alpha_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

令$P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, 则$A = P\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}P^{-1}$. 于是

$$\begin{aligned} e^A &= P\begin{pmatrix} e & 0 \\ 0 & e^2 \end{pmatrix}P^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e & 0 \\ 0 & e^2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e & -e+e^2 \\ 0 & e^2 \end{pmatrix}, \\ \sin A &= P\begin{pmatrix} \sin 1 & 0 \\ 0 & \sin 2 \end{pmatrix}P^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \sin 1 & 0 \\ 0 & \sin 2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \sin 1 & -\sin 1 + \sin 2 \\ 0 & \sin 2 \end{pmatrix}, \\ \cos A &= P\begin{pmatrix} \cos 1 & 0 \\ 0 & \cos 2 \end{pmatrix}P^{-1} = \begin{pmatrix} \cos 1 & -\cos 1 + \cos 2 \\ 0 & \cos 2 \end{pmatrix}. \end{aligned}$$

例 5.3.10. 设n阶矩阵A的特征值为$\lambda_1, \lambda_2, \cdots, \lambda_n$. 则$e^A$, $\sin A$, $\cos A$的特征值分别为$e^{\lambda_1}, e^{\lambda_2}, \cdots, e^{\lambda_n}$; $\sin\lambda_1, \sin\lambda_2, \cdots, \sin\lambda_n$; $\cos\lambda_1, \cos\lambda_2, \cdots, \cos\lambda_n$.

单纯矩阵的幂级数可由其谱分解方便地得到, 立即可得下述定理

定理 5.3.7. 设单纯矩阵A的谱分解为

$$A = \sum_{i=1}^{s} \lambda_i P_i.$$

设幂级数$f(t)$的收敛半径$r > \rho(A)$. 则

$$f(A) = \sum_{i=1}^{s} f(\lambda_i) P_i \tag{5.3.8}$$

例 5.3.11. 已知正规矩阵(第四章例4.1.8)

$$A = \begin{pmatrix} 8 & 4 & -1 \\ 4 & -7 & 4 \\ -1 & 4 & 8 \end{pmatrix}$$

的谱分解为$A = 9P_1 - 9P_2$, 因此$\sin A = (\sin 9)(P_1 - P_2)$.

下面给出指数函数e^A的一些基本性质.

命题 5.3.3. (1) 若$AB = BA$, 则$e^A e^B = e^{A+B} = e^B e^A$.
(2) $(e^A)^{-1} = e^{-A}$.
(3) $|\, e^A \,| = e^{\operatorname{tr} A}$.

证 (1) 因$AB = BA$, 所以二项式定理成立, 故有

$$(A+B)^m = \sum_{k=0}^{m} C_m^k A^{m-k} B^k.$$

所以

$$\begin{aligned} e^{A+B} &= \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k=0}^{m} C_m^k A^{m-k} B^k \\ &= \sum_{m=0}^{\infty} \sum_{k=0}^{m} \frac{1}{k!(m-k)!} A^{m-k} B^k \\ &= \left(\sum_{k=0}^{\infty} \frac{A^k}{k!} \right) \left(\sum_{j=0}^{\infty} \frac{B^j}{j!} \right) = e^A e^B. \end{aligned}$$

(2) 在(1)中令$B = -A$, 则得$e^A e^{-A} = I$, 所以

$$(e^A)^{-1} = e^{-A}.$$

(3) 设A的特征值为$\lambda_1, ..., \lambda_n$, 则$e^A$ 的特征值为$e^{\lambda_1}, ..., e^{\lambda_n}$, 故$|e^A| = e^{\lambda_1} \cdots e^{\lambda_n} = e^{\operatorname{tr} A}$. □

对于矩阵正弦函数和余弦函数, 容易验证如下结论.

命题 5.3.4. (1) (Euler 公式) $e^{\mathrm{i}A} = \cos A + \mathrm{i}\sin A$,

$$\cos A = \frac{1}{2}(e^{\mathrm{i}A} + e^{-\mathrm{i}A}),$$

$$\sin A = \frac{1}{2\mathrm{i}}(e^{\mathrm{i}A} - e^{-\mathrm{i}A}),$$

$$\cos(-A) = \cos A, \quad \sin(-A) = -\sin A.$$

(2) 若$AB = BA$, 则

$$\cos(A+B) = \cos A\cos B - \sin A\sin B,$$

$$\sin(A+B) = \sin A\cos B + \cos A\sin B.$$

注. 上述公式与数字的情形略有差异, 即矩阵乘积的公式往往需要交换性方能成立. 请参考本节的思考题.

思考题

1. 设$f(x)$是一个幂级数, 则$f(A)f(B) = f(B)f(A)$成立的可能性如何? 比如, $e^Ae^B = e^Be^A$成立的可能性有多大? 一般地, 如何比较与A可交换的矩阵的数量(当然是无穷多个)和与A不可交换的矩阵的数量?

2. 试举例说明矩阵e^Ae^B, e^Be^A与e^{A+B}可以两两不等. 又, 如果$e^Ae^B = e^Be^A$, 是否有$e^Ae^B = e^{A+B}$?

3. 矩阵的勾股定理是否成立, 即是否有$\cos^2 A + \sin^2 A = I$?

第四节 一元矩阵函数的微积分

5.4.1 定义与基本性质

定义 5.4.1. 若$m\times n$矩阵A的每个元素a_{ij}都是变量t的函数, 则称A为函数矩阵或矩阵函数, 记为$A(t) = (a_{ij}(t))_{m\times n}$.

我们将仿照高等数学来定义矩阵函数的相应概念, 也将略去大部分性质的证明.

定义 5.4.2. 设$A(t) = (a_{ij}(t))_{m\times n}$. 若对$\forall\, a_{ij}(t)$, $1\le i\le m$, $1\le j\le n$, 都有$\lim\limits_{t\to t_0} a_{ij}(t) = a_{ij}$(其中$a_{ij}\in\mathbb{C}$), 则称函数矩阵$A(t)$在$t_0$点处的极限为$A = (a_{ij})_{m\times n}$.

性质 5.4.1. 设$\lim\limits_{t\to t_0} A(t) = A, \lim\limits_{t\to t_0} B(t) = B$.

(1) 若$A(t)$, $B(t)$是同类型的矩阵, 则

$$\lim_{t\to t_0}(A(t) + B(t)) = A + B.$$

(2) 若$A(t)$, $B(t)$分别为$m\times n$, $n\times s$矩阵, 则

$$\lim_{t\to t_0}(A(t)B(t)) = \lim_{t\to t_0} A(t)\lim_{t\to t_0} B(t) = AB.$$

(3) 设k为常数, 则

$$\lim_{t\to t_0}(kA(t)) = kA = k\lim_{t\to t_0} A(t).$$

类似地, 可以定义函数矩阵$A(t) = (a_{ij}(t))_{m\times n}$在一点或某一区间内连续, 可微和可积. 另外, 设$A\in\mathbb{C}^{n\times n}$, t是未知数, 习惯上将矩阵函数tA记为At.

若$A(t)$可微, 其导数定义如下:

$$A'(t) = (a'_{ij}(t))_{m\times n}.$$

若$A(t)$可积, 定义:

$$\int_a^b A(t)\,\mathrm{d}\,t = \left(\int_a^b a_{ij}(t)\,\mathrm{d}\,t\right)_{m\times n}.$$

例 5.4.1. 设$N_{n\times n}$是幂零Jordan块, 则

$$e^{Nt} = \sum_{k=0}^{\infty} \frac{t^k N^k}{k!} = \sum_{k=0}^{n-1} \frac{t^k N^k}{k!}.$$

因此

$$\frac{\mathrm{d}e^{Nt}}{\mathrm{d}t} = Ne^{Nt} \tag{5.4.1}$$

性质 5.4.2. (1) $(aA(t)+bB(t))' = aA'(t)+bB'(t)$; $(A(t)\oplus B(t))' = A'(t)\oplus B'(t)$.
(2) $(A(t)B(t))' = A'(t)B(t)+A(t)B'(t)$; $(A(t)\otimes B(t))' = A'(t)\otimes B(t)+A(t)\otimes B'(t)$.
(3) $\left(\int_a^t A(s)\mathrm{d}\,s\right)' = A(t)$; $\int_a^t A'(s)\mathrm{d}\,s = A(t)-A(a)$.
(4) $\int_a^t BA(s)\mathrm{d}\,s = B\int_a^t A(s)\mathrm{d}\,s$, $\int_a^t A(s)B\mathrm{d}\,s = (\int_a^t A(s)\mathrm{d}\,s)B$, B为常数矩阵.

命题 5.4.1. (1) $\dfrac{\mathrm{d}e^{At}}{\mathrm{d}t} = Ae^{At}$;
(2) $\dfrac{\mathrm{d}\sin At}{\mathrm{d}t} = A\cos At$, $\dfrac{\mathrm{d}\cos At}{\mathrm{d}t} = -A\sin At$.
特别地, $\dfrac{\mathrm{d}e^{At}}{\mathrm{d}t}|_{t=0} = A$; $\dfrac{\mathrm{d}\sin At}{\mathrm{d}t}|_{t=0} = A$.

证 只证(1), 其余类似. 不妨设$A = J_n(\lambda) = \lambda I + N$. 由命题5.3.3(1)知, $e^{Jt} = e^{\lambda t}e^{Nt}$, 故

$$\frac{\mathrm{d}e^{Jt}}{\mathrm{d}t} = \lambda e^{\lambda t}e^{Nt} + e^{\lambda t}\frac{\mathrm{d}e^{Nt}}{\mathrm{d}t}.$$

再由公式(5.4.1)即可. □

推论 5.4.1. 对任意方阵A有

$$\int_{t_0}^t Ae^{As}\mathrm{d}s = e^{At} - e^{At_0} \tag{5.4.2}$$

$$\int_{t_0}^t A\sin As\mathrm{d}s = \cos At_0 - \cos At \tag{5.4.3}$$

$$\int_{t_0}^t A\cos As\mathrm{d}s = \sin At - \sin At_0 \tag{5.4.4}$$

例 5.4.2. 设向量 $x = x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T$ 及对称矩阵$A = A(t) = (A_{ij}(t))_{n\times n}$都可微, 求二次型$x^TAx$关于变量$t$的导数.

解

$$\begin{aligned}(x^TAx)' &= (x^T)'Ax + x^T(Ax)' \\ &= (x^T)'Ax + x^TA'x + x^TAx'.\end{aligned}$$

又

$$((x^T)'Ax)^T = x^TA^Tx' = x^TAx'.$$

而$(x^T)'Ax$为一阶矩阵, 所以$(x^T)'Ax = x^TAx'$. 于是, 我们得到

$$(x^TAx)' = x^TA'x + 2x^TAx'.$$

进一步, 如果假定对称矩阵A是常数矩阵, 则$A' = 0$, 于是二次型x^TAx(这是变量t的二次函数)的极值点满足条件$x^TAx' = 0$. 比如, 设$x = x(t) = (1-t, 1+t)^T, A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, 则二次型$x^TAx = a(1-t)^2 + 2b(1-t)(1+t) + c(1+t)^2 = f(t)$, 故

$$f'(t) = 2(a-2b+c)t - 2a + 2c.$$

而

$$x^TAx' = (1-t, 1+t)\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = (a-2b+c)t - a + c.$$

因此$x^TAx' = 0$确为二次型x^TAx的极值点满足的条件.

例 5.4.3. 设N是3阶幂零 Jordan 块, 则

$$\int_0^t e^{Ns}\mathrm{d}s = \begin{pmatrix} t & \frac{t^2}{2!} & \frac{t^3}{3!} \\ & t & \frac{t^2}{2!} \\ & & t \end{pmatrix}.$$

思考题

1. 公式$(A(t)^2)' = 2A(t)A'(t)$正确吗?
2. 设$A(t)$可逆, 如何计算$(A(t)^{-1})'$? 又$A'(t)$是否可逆?
3. 设$A(t)$是正交矩阵, 问$A'(t)$还是正交矩阵吗?
4. 例 5.4.3表明, 即使A不可逆, 积分$\int_{t_0}^t e^{As}\mathrm{d}s$仍然有意义. 应如何计算?

5.4.2 一元矩阵函数的计算

本节将利用 Lagrange-Sylvester 定理讨论矩阵函数及其微积分的计算.

定理 5.4.1. (1) 设

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}_{s\times s}, \quad 则 \quad e^{Jt} = e^{\lambda t}\begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{s-1}}{(s-1)!} \\ & 1 & t & \cdots & \frac{t^{s-2}}{(s-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & t \\ & & & & 1 \end{pmatrix}.$$

(2) 设$A=PJP^{-1}$且A的 Jordan 标准形为

$$J=J_1\oplus J_2\oplus\cdots\oplus J_s,$$

则

$$e^{At}=P(e^{J_1t}\oplus e^{J_2t}\oplus\cdots\oplus e^{J_st})P^{-1}.$$

证 只证(1)即可. 因为$J=\lambda I+N$, 其中

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

所以

$$\begin{aligned} e^{Jt} &= e^{\lambda tI+tN}=e^{\lambda t}e^{tN}=e^{\lambda t}\sum_{k=0}^{\infty}\frac{(tN)^k}{k!} \\ &= e^{\lambda t}\sum_{k=0}^{s-1}\frac{t^kN^k}{k!}. \end{aligned}$$

由于N^k仅第k条上对角线的元素为1, 其余元素均为0, 所以定理成立. □

例 5.4.4. 求e^{Jt}, 其中

$$J=\begin{pmatrix} J_1 & \\ & J_2 \end{pmatrix},\quad J_1=\begin{pmatrix} a & 1 & \\ & a & 1 \\ & & a \end{pmatrix},\quad J_2=\begin{pmatrix} b & 1 \\ & b \end{pmatrix}.$$

解 由定理 5.4.1, 有

$$e^{J_1t}=e^{at}\begin{pmatrix} 1 & t & \dfrac{t^2}{2!} \\ & 1 & t \\ & & 1 \end{pmatrix},\qquad e^{J_2t}=e^{bt}\begin{pmatrix} 1 & t \\ & 1 \end{pmatrix}.$$

所以,

$$e^{Jt}=\begin{pmatrix} e^{J_1t} & \\ & e^{J_2t} \end{pmatrix}=\begin{pmatrix} e^{at} & te^{at} & \dfrac{t^2e^{at}}{2!} & & \\ & e^{at} & te^{at} & & \\ & & e^{at} & & \\ & & & e^{bt} & te^{bt} \\ & & & & e^{bt} \end{pmatrix}.$$

设A为n阶方阵. 若已知A的最小多项式的次数为m, 则矩阵A的任何次幂都可由$I,\ A,\cdots,A^{m-1}$的线性组合表示. 因此, 由矩阵幂级数定义的矩阵函数$f(A)$, 实际上是一个次数不超过$m-1$的A的多项式$g(A)$. 这个多项式$g(t)$有什么特点呢? 回忆两个n次多项式相等$\iff$它们在$n+1$个点处的值相同, 因此为确定$g(A)$, 只需要研究一些特殊点处的值即可. 根据 Lagrange-Sylvester 定理, 只需要研究$g(t)$以及$g(t)$的适当阶数的导数在A的所有特征值处的值即可. 因此有下述定义.

定义 5.4.3. 设方阵A的最小多项式为

$$m(\lambda)=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_s)^{k_s}, \tag{5.4.5}$$

其中$\lambda_1,\ \lambda_2,\cdots,\lambda_s$互不相同. 对$f(t)=\sum_{k=0}^{\infty}a_kt^k$, 称

$$\begin{array}{ll}\{ & f(\lambda_1),\ f'(\lambda_1),\cdots,\ f^{(k_1-1)}(\lambda_1),\\ & f(\lambda_2),\ f'(\lambda_2),\cdots,\ f^{(k_2-1)}(\lambda_2),\\ & \cdots\cdots\cdots\cdots,\\ & f(\lambda_s),\ f'(\lambda_s),\cdots,\ f^{(k_s-1)}(\lambda_s)\quad \}\end{array} \tag{5.4.6}$$

为函数$f(t)$在矩阵A的谱上的数值.

定理5.4.2. (Sylvester 矩阵定理) 设$f(t)=\sum\limits_{k=0}^{\infty}a_kt^k$, $g(t)=\sum\limits_{k=0}^{\infty}b_kt^k$. 则$f(A)=g(A)\iff f(t)$与$g(t)$在 A的谱上的数值相等, 即, $f(\lambda_i)=g(\lambda_i)$, $f'(\lambda_i)=g'(\lambda_i),\cdots,f^{(k_i-1)}(\lambda_i)=g^{(k_i-1)}(\lambda_i)$, $i=1,2,\cdots,s$.

证 设A的 Jordan 标准形为$P^{-1}AP=J=J_1\oplus\cdots\oplus J_s$, 则由$f(A)=g(A)\iff f(J)=g(J)\iff f(J_i)=g(J_i),\forall i$, 再对照 Lagrange-Sylvester 定理中的每个块即可, 详见习题 40. □

推论 5.4.2. 设矩阵A可对角化, 则$f(A)=g(A)\iff f(\lambda)=g(\lambda),\forall\lambda\in\sigma(A)$.

例 5.4.5. 已知矩阵A的最小多项式是$m(\lambda)=\lambda^2(\lambda-\pi)$. 证明$\sin A=A-\dfrac{1}{\pi}A^2$.

证 设$f(t)=\sin t$, $g(t)=t-\dfrac{1}{\pi}t^2$. 由定理 5.4.2, 只要证明$f(t)$与$g(t)$在A的谱上的数值相同, 则$f(A)=g(A)$. 直接计算可得,

$$\begin{aligned}&f(0)=\sin 0=0=g(0),\quad f'(0)=\cos 0=1=g'(0),\\ &f(\pi)=\sin\pi=0=g(\pi).\end{aligned}$$

因此$\sin A=f(A)=g(A)=A-\dfrac{1}{\pi}A^2$. □

例 5.4.6. 求e^{At}, 其中

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

解 A的特征多项式为

$$|\lambda I-A|=\begin{vmatrix}\lambda-2&-1&-4\\0&\lambda-2&0\\0&-3&\lambda-1\end{vmatrix}=(\lambda-1)(\lambda-2)^2.$$

直接验证便知, $(x-1)(x-2)$不是A的零化多项式, 所以A的最小多项式为

$$m(x)=(x-1)(x-2)^2.$$

令$f(\lambda)=e^{\lambda t}$(在这里把t看成常数), 则$e^{At}=f(A)$. 设$g(\lambda)=a_0+a_1\lambda+a_2\lambda^2$(其中$a_0$, a_1, a_2为待定系数). 要求$f(\lambda)$与$g(\lambda)$在A的谱上的数值相等, 即

$$\begin{cases} e^t = f(1) = g(1) = a_0 + a_1 + a_2, \\ e^{2t} = f(2) = g(2) = a_0 + 2a_1 + 4a_2, \\ te^{2t} = f'(2) = g'(2) = a_1 + 4a_2. \end{cases}$$

解上述方程得

$$a_0 = 4e^t - 3e^{2t} + 2te^{2t},\ a_1 = -4e^t + 4e^{2t} - 3te^{2t},\ a_2 = e^t - e^{2t} + te^{2t}.$$

因而得到

$$e^{At} = \begin{pmatrix} e^{2t} & 12e^t - 12e^{2t} + 13te^{2t} & -4e^t + 4e^{2t} \\ 0 & e^{2t} & 0 \\ 0 & -3e^t + 3e^{2t} & e^t \end{pmatrix}.$$

上述例子采用的是待定系数法. 下面介绍一种实际应用中经常采用的Lagrange插值法.

命题 5.4.2. 设方阵A的最小多项式$m(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_m)$无重根, $f(t)$为任一收敛半径$r>\rho(A)$的幂级数. 则

$$f(A) = \sum_{i=1}^{m} f(\lambda_i) L_i(A),$$

其中$L_i(x)$, $i=1,2,\cdots,m$是Lagrange插值多项式(见第二章定理2.1.3).

例 5.4.7. 设$A=\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, 求$e^{At}$.

解 因为$|\lambda I - A| = \begin{vmatrix} \lambda-1 & -2 \\ -2 & \lambda-1 \end{vmatrix} = (\lambda+1)(\lambda-3)$, 所以$A$的特征值为$\lambda_1=-1$, $\lambda_2=3$, 因而其最小多项式为

$$m(x) = (x+1)(x-3).$$

令$f(\lambda)=e^{\lambda t}$, 则按Lagrange插值公式, 有

$$f(A) = f(\lambda_1)L_1(A) + f(\lambda_2)L_2(A),$$

其中

$$L_1(A) = \frac{A-\lambda_2 I}{\lambda_1-\lambda_2} = -\frac{1}{4}\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix},\quad L_2(A) = \frac{A-\lambda_1 I}{\lambda_2-\lambda_1} = \frac{1}{4}\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.$$

又$f(\lambda_1)=e^{-t}$, $f(\lambda_2)=e^{3t}$. 所以

$$\begin{aligned} e^{At} = f(A) &= e^{-t}\cdot(-\frac{1}{4})\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} + e^{3t}\frac{1}{4}\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} e^{-t}+e^{3t} & -e^{-t}+e^{3t} \\ -e^{-t}+e^{3t} & e^{-t}+e^{3t} \end{pmatrix}. \end{aligned}$$

如果A的最小多项式有重根, 则Lagrange插值公式变为如下的Lagrange-Sylvester插值公式, 证明比较繁琐, 略去.

命题 5.4.3. (Lagrange-Sylvester 插值公式) 设n阶方阵A的最小多项式为

$$m(\lambda)=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_s)^{k_s},$$

其中$\sum_{i=1}^{s}k_i=m\leq n$. 则

$$f(A)=\sum_{i=1}^{s}\varphi_i(A)\left(a_{i1}I+a_{i2}(A-\lambda_iI)+\cdots+a_{ik_i}(A-\lambda_iI)^{k_i-1}\right),$$

其中

$$\varphi_i(A)=(A-\lambda_1I)^{k_1}\cdots(A-\lambda_{i-1}I)^{k_{i-1}}(A-\lambda_{i+1}I)^{k_{i+1}}\cdots(A-\lambda_sI)^{k_s},\quad 1\leq i\leq s,$$

$$a_{ij}=\frac{1}{(j-1)!}\frac{\mathrm{d}^{j-1}}{\mathrm{d}\,\lambda^{j-1}}\left((\lambda-\lambda_i)^{k_i}\frac{f(\lambda)}{m(\lambda)}\right)\Bigg|_{\lambda=\lambda_i},\quad 1\leq i\leq s;\ j=1,2,\cdots,k_i.$$

第五节 多元矩阵函数的导数

本节限定在实数域$\mathbb{R}$上.

5.5.1 多元矩阵函数的导数与Jacobian行列式

一元矩阵函数的导数定义本质上是一元函数的导数, 多个变量的情形显然更加重要, 比如若例 5.4.2中的向量x与矩阵A中的自变量均为二元的, 如何计算二次型x^TAx的导数? 为此, 先考察下面的例子.

例 5.5.1. 若$A\in\mathbb{R}^{n\times n}$, 求实二次型$x^TAx$在单位球面$x^Tx=1$上的最大值与最小值.

作 Lagrange 辅助函数$L=x^TAx+\lambda(1-x^Tx)$, 其梯度$\nabla(L)$的第i个分量为

$$2\sum_{j=1}^{n}a_{ij}x_j-2\lambda x_i.$$

因此L的极大值点需要满足条件$2\sum\limits_{j=1}^{n}a_{ij}x_j-2\lambda x_i=0, 1\leq i\leq n$, 此恰好是$Ax=\lambda x$(Lagrange 乘数因子$\lambda$是特征值!). 因此, 二次型$x^TAx$在单位球面$x^Tx=1$上的最大值与最小值均在单位特征向量处取得, 并且最大值与最小值就是最大特征值与最小特征值(极大极小原理)!

例 5.5.2. 设$f(x)=ax_1+bx_2+cx_3=\begin{pmatrix}a & b & c\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$, 则其梯度向量

$$\nabla(f)=(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\frac{\partial f}{\partial x_3})=\begin{pmatrix}a & b & c\end{pmatrix}.$$

上面两个例子中函数L或$f(x)=f(x_1,x_2,\cdots,x_n)$的梯度

$$\nabla(f)=(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n})\tag{5.5.1}$$

给出了一个多元函数的“向量值”导数, 即若$f(x)=f(x_1,x_2,\cdots,x_n)$是一个$n$元函数, 其梯度可以看作是其导数$\dfrac{\partial f}{\partial x}$, 于是我们引入以下定义.

定义 5.5.1. 设$f(x) = f(x_1, x_2, \cdots, x_n)$是$n$元可微函数. 称$f$的梯度向量$\nabla(f)$为其**导数向量**即下面的行函数向量

$$\frac{\partial f}{\partial x} = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}) \tag{5.5.2}$$

更一般地, 需要定义多元函数矩阵的导数. 为简单起见, 假定所涉及的函数在整个空间上都是可微的.

定义 5.5.2. 设$f = (f_1, f_2, \cdots, f_m)^T : \mathbb{F}^n \longrightarrow \mathbb{F}^m$是一个$n$元可微映射(写成列向量), 即每个$n$元函数$f_i, 1 \leq i \leq m$均为可微函数, 则映射$f$的导数定义为如下的$m \times n$矩阵

$$\frac{\partial(f_1, f_2, \cdots, f_m)}{\partial(x_1, x_2, \cdots, x_n)} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \tag{5.5.3}$$

矩阵$\frac{\partial(f_1, f_2, \cdots, f_m)}{\partial(x_1, x_2, \cdots, x_n)}$称为映射(或函数向量) f的Jacobian矩阵, 常简记为$J(f)$. 如果$J(f)$是方阵, 则其行列式称为Jacobian行列式.

注1. 如果映射f是行向量$(f_1, f_2, \cdots, f_m)$, 则定义其导数为公式(5.5.3)中矩阵的转置. 今后, 我们将把任何函数或映射f的导数记为$J(f)$, 而$\nabla(f)$将意味着f是一个多元函数.

注2. 部分作者将梯度向量看成是列向量, 另一部分(如本书作者)则将其看成是行向量, 因此用行向量记号还是用列向量记号常常导致混乱, 务必特别仔细. 沿用高等数学的习惯, 本书将梯度向量看成是行向量, 行向量记号的优点见下例.

例 5.5.3. 设$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $f : X^T = (x, y)^T \mapsto (a_1x + b_1y, a_2x + b_2y)^T = AX$是$\mathbb{F}^2$上的线性变换, 则其导数为

$$J(f) = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = A.$$

相当于$\frac{\mathrm{d}(AX)}{\mathrm{d}X} = A$, 与高等数学一致. 一般地, 如果$f = (f_1, f_2, \cdots, f_m)^T : \mathbb{F}^n \longrightarrow \mathbb{F}^m$是一个线性变换, 设$f$在标准基-标准基下的矩阵为$A_{m\times n}$, 则$f$的导数(矩阵)正是$A$本身.

例 5.5.4. 设x是n元(未知)列向量, A是$m \times n$阶常数矩阵, 则

(1)$J(x) = \frac{\partial x}{\partial x} = I_n = \frac{\partial x^T}{\partial x} = J(x^T)$;

(2)$J(Ax) = \frac{\partial Ax}{\partial x} = A,\ \ J(x^TA^T) = \frac{\partial x^TA^T}{\partial x} = A^T$.

例 5.5.5. 设$f(x) = f(x_1, x_2, \cdots, x_n)$是$n$元二次可微函数, 则其导数(=梯度)$J(f) = \frac{\partial f}{\partial x}$的

导数$J(J(f))$称为f的Hessian [88] 矩阵, 记为$H(f)$, 即有

$$H(f)=J(J(f))=\begin{pmatrix}\dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_2\partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n\partial x_1}\\ \dfrac{\partial^2 f}{\partial x_1\partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_n\partial x_2}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\partial^2 f}{\partial x_1\partial x_n} & \dfrac{\partial^2 f}{\partial x_2\partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}\end{pmatrix} \tag{5.5.4}$$

由多元微分学可知, 如果f是二次连续可微的, 则其Hessian矩阵$H(f)$是对称矩阵. 函数$f(x)$的Hessian矩阵$H(f)$可以视为$f(x)$的二阶导数.

例 5.5.6. 设$f(x,y)=x^2+2xy+y^2-x$, 则其导数为$J(f)=(2x+2y-1,2x+2y)$, 其Hessian矩阵为

$$H(f)=\begin{pmatrix}2&2\\2&2\end{pmatrix}.$$

例 5.5.7. 设$f=f(x,y)=(x^2+2xy+y^2-x,x^2+2xy+y^2+y)^T$是二元多项式映射(即$f$的两个分量$f_1,f_2$均为二元多项式), 则其Jacobian矩阵$J(f)$为

$$J(f)=\begin{pmatrix}2x+2y-1 & 2x+2y\\ 2x+2y & 2x+2y+1\end{pmatrix},$$

易知其Jacobian行列式$|J(f)|=1$. 未解决的著名猜想— **Jacobian 猜想**(1939 年)是说如果一个$\mathbb{C}^n$到其自身的多项式映射f的Jacobian行列式$|J(f)|=1$, 则f一定有逆(必定也是多项式映射). 线性变换是最简单的多项式映射, Jacobian猜想对线性映射成立. 目前二次多项式映射的Jacobian猜想已由S.Wang[89] 于1980年所证明. 比如, 本题中的映射是可逆的, 请读者自行验证.

多元矩阵函数的导数的四则运算法则与普通函数的导数法则基本类似(注意区分行向量与列向量), 我们列出其中如下部分.

命题 5.5.1. 设f,g均为n元向量x的可微函数, $h(x)$是n元行映射, $p(x),q(x)$是m维行映射, $a,b\in\mathbb{F}$, 则

(1) (线性法则)$J(af+bg)=aJ(f)+bJ(g)$.

(2) (乘法公式)$J(pq^T)=pJ(q^T)+qJ(p^T)$.

(3) (除法公式)$J(f/g)=(gJ(f)-fJ(g))/g^2$.

(4) (链法则) $J(f(h))=\dfrac{\partial f(h)}{\partial h}J(h^T)$.

证 我们只证明链法则(4). 记$x=(x_1,x_2,\cdots,x_n)^T,h=(h_1,h_2,\cdots,h_n)$. 则由多元函数的链法则可知

$$\frac{\partial f(h(x))}{\partial x}=\nabla(f(h_1,h_2,\cdots,h_n))=(\sum_{i=1}^n\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_1},\sum_{i=1}^n\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_2},\cdots,\sum_{i=1}^n\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_n}).$$

[88] Ludwig Otto Hesse(1811-1874), 德国数学家, Carl Gustav Jacob Jacobi 的学生.

[89] Stuart Sui-Sheng Wang(王穗生, 1946-), 华裔美籍数学家, 现美国Oakland大学教授.

另一方面, 由

$$\frac{\partial f(h)}{\partial h}\frac{\partial h^T}{\partial x}=(\frac{\partial f}{\partial h_1},\frac{\partial f}{\partial h_2},\cdots,\frac{\partial f}{\partial h_n})J(h^T)=(\sum_{i=1}^n\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_1},\sum_{i=1}^n\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_2},\cdots,\sum_{i=1}^n\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_n}).$$

故链法则成立. □

注. 通常将多元复合函数$f(h)$中的中间变量h写成行向量的形式即$h=(h_1,h_2,\cdots,h_n)$, 因此要求链法则中的第二个因子是h^T的导数而不是h的导数.

例 5.5.8. 设$f=f(u,v)$是二元函数, $g(x,y)=(x^2,ax+y^2)$为二元映射. 则$f(g)=f(x^2,ax+y^2)$, 由链法则可得

$$J(f(g))=\frac{\partial f(g)}{\partial(x,y)}=J(f)\frac{\partial g^T}{\partial(x,y)}=(\frac{\partial f}{\partial u},\frac{\partial f}{\partial v})\begin{pmatrix}2x & 0\\ a & 2y\end{pmatrix}=(2x\frac{\partial f}{\partial u}+a\frac{\partial f}{\partial v},2y\frac{\partial f}{\partial v}),$$

此与多元微分学的结果一致.

注. 如果$f=(f_1,f_2,\cdots,f_m)^T$与$g=(g_1,g_2,\cdots,g_m)^T$是两个m维列映射, 则乘法公式$J(f^Tg)$的表达应该与命题5.5.1(2)中的乘法公式略有差别但本质一致, 验证留作习题.

现在可以得到求二次型x^Tx的导数的正确公式了:

$$\frac{\partial x^Tx}{\partial x}=x^T\frac{\partial x}{\partial x}+x^T\frac{\partial x^T}{\partial x}=2x^T \tag{5.5.5}$$

再来讨论例5.5.1中的条件极值问题. 直接对Lagrange辅助函数L关于向量x求导可得

$$J(L)=J(x^TAx)+J(\lambda(1-x^Tx))=x^TJ(Ax)+(Ax)^TJ(x^T)-\lambda J(x^Tx)=2x^TA-2\lambda x^T,$$

因此其极值点必须满足条件$x^TA=\lambda x^T$, 即x^T是A的左特征向量或x是A^T的右特征向量.

总结. 一元矩阵函数的导数运算法则与一元函数的导数运算法则完全一致, 多元矩阵函数的导数本质上是对自变量向量求导(注意本书始终对列向量x求导, 而没有讨论对行向量x^T求导), 因此其运算法则与函数的导数运算法则有较大差别.

例 5.5.9. 之前一直在对向量$x=(x_1,x_2,\cdots,x_n)^T$求导, 假如对向量$y=(x_1,x_1+x_2,\cdots,x_1+x_n)^T$求导会得到什么结果? 这相当于对复合函数$f(g)$求导, 其中$f$是$n$元函数, 而$y=g(x)=(x_1,x_1+x_2,\cdots,x_1+x_n)$是$n$元映射(这是一个坐标变换). 具体计算留作习题.

行列式和迹都是从$\mathbb{F}^{n\times n}$到$\mathbb{F}$的多元(n^2元)函数, 现在研究它们的导数. $\mathbb{F}^{n\times n}$中的未知矩阵可以写成$X=(x_{ij})$, 将其行列式$|X|$改记为$D(X)$, 其迹仍记为$\mathrm{tr}(X)$或$\mathrm{tr}X$, 称为**迹函数**. 由于$D(X)$与$\mathrm{tr}(X)$关于诸变量x_{ij}的导数$\nabla(D(X))=J(D(X))$与$\mathrm{tr}(X)$均为n^2维的行向量, 不便表示, 因此将它们都排列成如下的矩阵形式

$$J(D(X))=\frac{\partial D(X)}{\partial X}=\begin{pmatrix}\frac{\partial D(X)}{\partial x_{11}} & \frac{\partial D(X)}{\partial x_{12}} & \cdots & \frac{\partial D(X)}{\partial x_{1n}}\\ \frac{\partial D(X)}{\partial x_{21}} & \frac{\partial D(X)}{\partial x_{22}} & \cdots & \frac{\partial D(X)}{\partial x_{2n}}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial D(X)}{\partial x_{n1}} & \frac{\partial D(X)}{\partial x_{n2}} & \cdots & \frac{\partial D(X)}{\partial x_{nn}}\end{pmatrix} \tag{5.5.6}$$

与

$$J(\mathrm{tr}(X)) = \frac{\partial \mathrm{tr}(X)}{\partial X} == \begin{pmatrix} \frac{\partial \mathrm{tr}(X)}{\partial x_{11}} & \frac{\partial \mathrm{tr}(X)}{\partial x_{12}} & \cdots & \frac{\partial \mathrm{tr}(X)}{\partial x_{1n}} \\ \frac{\partial \mathrm{tr}(X)}{\partial x_{21}} & \frac{\partial \mathrm{tr}(X)}{\partial x_{22}} & \cdots & \frac{\partial \mathrm{tr}(X)}{\partial x_{2n}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \mathrm{tr}(X)}{\partial x_{n1}} & \frac{\partial \mathrm{tr}(X)}{\partial x_{n2}} & \cdots & \frac{\partial \mathrm{tr}(X)}{\partial x_{nn}} \end{pmatrix} \tag{5.5.7}$$

一般将式(5.5.6)与式(5.5.7)分别称为行列式与迹函数的梯度矩阵.

例 5.5.10. 设$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$, 则

$$D(X) = x_{11}x_{22} - x_{21}x_{12}, \quad \mathrm{tr}(X) = x_{11} + x_{22}.$$

因此

$$J(D(X)) = \begin{pmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{pmatrix} = (\mathrm{adj}X)^T \quad J(\mathrm{tr}(X)) = I.$$

容易计算例 5.5.10的结论一般性成立, 即有下面的命题

命题 5.5.2. 设X是n阶矩阵, 则其行列式的梯度矩阵是其伴随矩阵的转置, 而迹函数的导数是单位矩阵, 即

$$J(D(X)) = (\mathrm{adj}X)^T, \qquad J(\mathrm{tr}(X)) = I_n \tag{5.5.8}$$

以上对矩阵的行列式与迹函数$\mathrm{tr}(\cdot)$的考察相对简单, 更有趣的问题是研究矩阵AXB(及类似矩阵)的迹, 其中A, B为给定的矩阵, 即求$J(\mathrm{tr}(AXB)) = \dfrac{\partial \mathrm{tr}(AXB)}{\partial X}$, 请读者自行讨论.

例 5.5.11. 设$x = r\sin\theta\cos\phi, y = r\sin\theta\sin\phi, z = r\cos\theta$. 则

$$J = \frac{\partial(x,y,z)}{\partial(r,\theta,\phi)} = \begin{pmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ r\cos\theta\cos\phi & r\cos\theta\sin\phi & -r\sin\theta \\ -r\sin\theta\sin\phi & r\sin\theta\cos\phi & 0 \end{pmatrix},$$

因此$|J| = r^2\sin\theta$正是球面坐标下的积分系数.

例 5.5.11具有一般性. 假定对n元函数$f(x)$的重积分$\displaystyle\int_V f(x)\mathrm{d}x$进行换元

$$x_i = g_i(y), \quad i = 1, ..., n.$$

注意所谓微元$\mathrm{d}x = \mathrm{d}x_1 \cdots \mathrm{d}x_n$实际上是以向量$\mathrm{d}x_1 e_1, \cdots, \mathrm{d}x_n e_n$为“邻边”的“超矩形”的“超面积”(请对照二重积分), 而向量

$$\overrightarrow{\mathrm{d}x} = (\mathrm{d}x_1, \cdots, \mathrm{d}x_n)^T = \frac{\partial x}{\partial y}\overrightarrow{\mathrm{d}y} = J(\mathrm{d}y_1, \cdots, \mathrm{d}y_n)^T,$$

故由第二章第七节“度量与任意矩阵的行列式”可知, 以向量$\mathrm{d}x_1e_1,\cdots,\mathrm{d}x_ne_n$为“邻边”的“超矩形”在新坐标系里的“超面积”恰好是($||J||$表示绝对值, 因为重积分没有方向)

$$\mathrm{d}x=\mathrm{d}x_1\cdots\mathrm{d}x_n=||J||\mathrm{d}y_1\cdots\mathrm{d}y_n=||J||\mathrm{d}y.$$

由此可以简便地计算下面的多重积分(其中$A\in\mathbb{R}^{n\times n}$是正定矩阵):

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^TAx}\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

由于A正定, 故存在可逆矩阵P使得$A=P^TP$. 令$x=P^{-1}y$, 则

$$e^{-\frac{1}{2}x^TAx}=e^{-\frac{1}{2}y^Ty}=e^{-\frac{1}{2}\sum_{i=1}^n y_i^2},\quad |A|=|P|^2,\quad \mathrm{d}x=|P|^{-1}\mathrm{d}y=|A|^{-\frac{1}{2}}\mathrm{d}y.$$

故

$$\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^TAx}\mathrm{d}x_1\cdots\mathrm{d}x_n=\prod_{i=1}^n\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{y_i^2}{2}}\mathrm{d}y_i=|A|^{-\frac{1}{2}}(2\pi)^{\frac{n}{2}}. \tag{5.5.9}$$

公式(5.5.9)称为**Aitken 积分公式**.

设n元函数$f(x)=f(x_1,x_2,\cdots,x_n)$的导数为$J(f)$, 我们已经看到$f(x)$的二阶导数即其Hessian矩阵$H(f)$. 自然应该将矩阵函数$H(f)$关于向量$(x_1,x_2,\cdots,x_n)$的导数定义为函数$f(x)$关于向量$(x_1,x_2,\cdots,x_n)$的三阶导数. 那么, 我们应该如何定义$n$阶矩阵函数$H(f)$关于向量$(x_1,x_2,\cdots,x_n)$的导数呢? 更一般地, 设$f(X)$是$p\times q$阶矩阵$X$的$m\times n$阶矩阵函数, 如何定义$f(X)$关于$X$的导数呢? 仿照向量函数对向量的导数, 一个可行的办法是将矩阵看作向量, 即利用矩阵的向量化变换vec, 然后按照向量函数对向量的导数计算导数(矩阵)即可.

定义 5.5.3. 设$f:\mathbb{F}^{p\times q}\longrightarrow\mathbb{F}^{m\times n}$是一个$p\times q$元矩阵函数, 即

$$f(X)=\begin{pmatrix} f_{11}(X) & f_{12}(X) & \cdots & f_{1n}(X)\\ f_{21}(X) & f_{22}(X) & \cdots & f_{2n}(X)\\ \cdots & \cdots & \cdots & \cdots\\ f_{m1}(X) & f_{m2}(X) & \cdots & f_{mn}(X)\end{pmatrix},$$

其中$X=(x_{ij})_{p\times q}$是自变量矩阵, 则$f(X)$的导数定义为$mp\times nq$阶矩阵

$$J(f)=\frac{\partial f(X)}{\partial X}=\frac{\partial\mathrm{vec}(f(X))}{\partial\mathrm{vec}X} \tag{5.5.10}$$

例 5.5.12. 设$f:\mathbb{F}^{2\times 2}\longrightarrow\mathbb{F}^{2\times 2}$是$2\times 2$元恒等矩阵函数, 即$f(X)=X$, 则

$$J(f)=\frac{\partial\mathrm{vec}(f(X))}{\partial\mathrm{vec}X}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}=I_4!$$

这是自然的, 因为我们总是将自变(向)量x视为列向量(即使n元函数$f(x_1,x_2,\cdots,x_n)$中的自变量写成了行的形式, 我们依然认为$f(x)$是列向量$x=(x_1,x_2,\cdots,x_n)^T$的函数, 这等于将函数$f(x)$看作是从$\mathbb{R}^n$到$\mathbb{R}$的映射), 而$\partial\mathrm{vec}X$规定自变量是按照列顺序排列的.

容易证明下述命题阐述的矩阵函数的导数的性质.

命题 5.5.3. (1) 设$f=f(x)=(f_{ij}(x))_{m\times p}$与$g=g(x)=(g_{ij}(x))_{p\times q}$均为$n$元矩阵函数, 则

$$J(f(x)g(x))=(g(x)^T\otimes I_m)J(f(x))+(I_q\otimes f(x))J(g(x)) \tag{5.5.11}$$

特别地,

$$J(xx^T)=x\otimes I_n+I_n\otimes x \tag{5.5.12}$$

(2) 设$A\in\mathbb{C}^{m\times n},B\in\mathbb{C}^{p\times q},X\in\mathbb{C}^{n\times p}$, 则

$$J(AXB)=\frac{\partial AXB}{\partial X}=B^T\otimes A;J((AXB)^T)=\frac{\partial B^TX^TA^T}{\partial X}=B\otimes A^T=(J(AXB))^T.$$

(3) 设$A\in\mathbb{C}^{m\times n},X\in\mathbb{C}^{n\times m}$, 则

$$\frac{\partial \mathrm{tr}(AX)}{\partial X}=\frac{\partial \mathrm{tr}(XA)}{\partial X}=A^T.$$

(4) 设$A\in\mathbb{C}^{n\times n},X\in\mathbb{C}^{m\times n}$, 则

$$\frac{\partial \mathrm{tr}(XAX^T)}{\partial X}=X(A+A^T).$$

(5) 设$\alpha,\beta\in\mathbb{C}^{n\times 1},X\in\mathbb{C}^{m\times n}$, 则

$$\frac{\partial(\alpha^TX^TX\beta)}{\partial X}=X(\alpha\beta^T+\beta\alpha^T).$$

(6) 设$\alpha,\beta\in\mathbb{C}^{m\times 1},X\in\mathbb{C}^{m\times n}$, 则

$$\frac{\partial(\alpha^TXX^T\beta)}{\partial X}=(\alpha\beta^T+\beta\alpha^T)X.$$

(7) 设$X\in\mathbb{C}^{m\times n},f=f(X)=(f_{ij}(X))_{p\times q}$与$g=g(X)=(g_{ij}(X))_{m\times n}$是两个函数矩阵, 则

$$\frac{\partial(f(g(X)))}{\partial X}=\frac{\partial(f(g))}{\partial g}\frac{\partial(g)}{\partial X}.$$

下述定理利用矩阵函数的导数求解极值问题, 略去证明, 请读者参考[15]第五章.

定理 5.5.1. 设$X\in\mathbb{R}^{m\times n}$, $f(X)$是关于矩阵变量X的二阶连续可导函数.

(1) 若X_0是$f(X)$的一个极值点, 则$J(f)(X_0)=\dfrac{\partial f(X)}{\partial X}|_{X=X_0}=0$; 换句话说, $f(X)$的极值点均满足方程

$$J(f)=0\text{ 或 }\frac{\partial f(X)}{\partial X}=0.$$

(2) 若$f(X)$还是关于矩阵变量X的凸函数, 则$f(X)$的极值点必是最值点.

(3) 若X_0是$f(X)$在约束条件$c(X)=0$下的一个条件极值点, $L(X,\lambda)=f(X)-\lambda c(X)$是 Lagrange 辅助函数. 则存在数$\lambda_0\in\mathbb{R}$, 使得

$$J(L)_X(X_0,\lambda_0)=\frac{\partial L(X,\lambda)}{\partial X}|_{(X_0,\lambda_0)}=0,\quad c(X_0)=0.$$

思考题

1. 设$\alpha \in \mathbb{C}^{m\times 1}, \beta \in \mathbb{C}^{n\times 1}$, 则$J(\alpha^T X\beta) = \dfrac{\partial \alpha^T X\beta}{\partial X} =?$ $J(\alpha^T X^T\beta) = \dfrac{\partial \alpha^T X^T\beta}{\partial X} =?$

2. 设X是方阵, $J(X^2) =?$

3. 如果定义$J(X) = \dfrac{\partial X}{\partial \mathrm{rvec} X}$, 将得到何种结果? 是否还有其它方式?

4. 试比较隐函数存在定理与Jacobian猜想.

5.5.2 多元矩阵函数导数的应用

考虑K个用户的移动通信的CDMA系统, 设第k个用户的信号幅值为A_k, 特定时刻发射的比特信号为$b_k\,(=\pm 1)$. 则描述该系统的一个简化数学模型为:

$$y = RAb + n \tag{5.5.13}$$

其中y是通信基站的接收信号向量, K阶对称矩阵R是(用户扩频波形的)自相关矩阵, $b = (b_1, \cdots, b_k)^T$, K阶列向量n是方差为σ^2的噪声向量(一般假定n是加性白Gauss噪声且与用户信号不相关). 问题: 如何设计第k个用户的检测器m_k(即使用$\hat{b}_k = \mathrm{sgn}(m_k^T y)$作为第$k$个用户发射信号的估计), 使得所有$K$个用户的检测器

$$M = (m_1, \cdots, m_K) \tag{5.5.14}$$

具有**最小均方误差**? 即使得以矩阵M为自变量的目标函数

$$J(M) = E\{\|b - My\|_2^2\} \tag{5.5.15}$$

最小化, 其中$\|\bullet\|$是向量的欧几里得范数.

利用数学期望与矩阵的迹的性质可以证明(此处略去)目标函数$J(M)$具有下述表达

$$J(M) = \mathrm{tr}(I) + \mathrm{tr}(M(RA^2R + \sigma^2 R)M^T) - \mathrm{tr}(ARM^T) - \mathrm{tr}(MRA) \tag{5.5.16}$$

由定理 5.5.1, 矩阵函数$J(M)$的最小值点M必须满足方程

$$\frac{\partial J(M)}{\partial M} = 0 \tag{5.5.17}$$

由命题 5.5.3可知(注意R是对称矩阵)

$$\frac{\partial J(M)}{\partial M} = 2M(RA^2R + \sigma^2 R) - 2AR \tag{5.5.18}$$

于是方程(5.5.17)的解M必须满足方程

$$M(RA^2R + \sigma^2 R) = AR \tag{5.5.19}$$

特别, 若R可逆, 则可得下述唯一的最优盲多用户检测器

$$M = A(RA^2 + \sigma^2 I)^{-1}.$$

上述方法称为**盲多用户检测**(英文缩写为BMUD), 是以矩阵为自变量的函数的微分学的前沿应用之一.

第六节 应用I：范数与矩阵近似

除特别指明外, 本节中的矩阵总指$\mathbb{F}^{m\times n}, q = \min\{m, n\}$. 矩阵$A$称为**"低秩矩阵"**如果$r(A) < \dfrac{q}{2}$. 实践中的低秩矩阵一般指$r(A)$远远小于$\dfrac{q}{2}$.

$\mathbb{F}^{m\times n}$上的范数$\|\bullet\|$称为**酉不变范数**, 是指对任意酉矩阵$U \in \mathbb{F}^{m\times m}, V \in \mathbb{F}^{n\times n}$和任意矩阵$X$均有$\|UXV\| = \|X\|$.

比如Frobenius范数和2-范数均是酉不变范数.

5.6.1 低秩近似与Schmidt-Eckart -Young -Mirsky定理

设有带噪图片$A+E$, 其中E为噪音, 如何找到原始图片A的最佳近似呢? 图像降噪的数学本质是矩阵的低秩分解.

对任意矩阵A, 是否存在指定秩的最佳近似矩阵? 下述著名的Schmidt-Eckart-Young-Mirsky 定理[90]肯定回答了该问题.

定理 5.6.1. 设$A \in \mathbb{F}^{m\times n}, 1 \le k \le r = r(A), \sigma_1 \ge ... \ge \sigma_r > \sigma_{r+1} = 0 = ... = \sigma_q$是$A$的奇异值, $v_1, ..., v_r, v_{r+1}, ..., v_n$是相应的单位正交右奇异向量. 设$A_k$是$A$的第$k$个截尾奇异值分解, 即

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^* \tag{5.6.1}$$

则A_k是酉不变范数意义下的A的最佳秩k-近似, 即对任何秩不超过k的矩阵B以及任意酉不变范数$\|\bullet\|$, 均有

$$\|A - A_k\| \le \|A - B\| \tag{5.6.2}$$

证 为简单起见, 此处仅给出2-范数的证明, 对任意酉不变范数的证明可参考[91].

设$r(B) \le k$, 则其零空间的维数$\dim N(B) \ge n-k$. 取$V_{k+1} = \mathrm{Span}\{v_1, ..., v_{k+1}\}$. 则$\exists x \in N(B)\bigcap V_{k+1}, x^*x = 1$. 于是$Bx = 0, x^*v_j = v_j^*x = 0, k+2 \le j \le n$. 故有

$$\|A-B\|_2^2 \ge \|(A-B)x\|_2^2 = \|Ax\|_2^2$$

$$= \sum_{i=1}^{n} \sigma_i^2 (u_i v_i^* x)^* (u_i v_i^* x) = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^* x)^* (v_i^* x) \ge \sigma_{k+1}^2.$$

另一方面, 容易计算$\|A-B\|_2^2 + \sigma_{k+1}^2$. 故定理成立. □

5.6.2 核范数与矩阵完备化

矩阵完备化(又称矩阵填充、矩阵补全等)是根据已知的矩阵元素恢复缺失的矩阵元素的方法.

矩阵完备化的典型例子是Netfix 大奖赛问题, 该问题的目标是根据以往消费者的评价为潜在消费者提供推荐列表. 容易看出, 该问题可以归结为矩阵完备化问题, 其数学模型如下.

---

[90]该定理被广泛称为Eckart-Young 定理, 1907年由俄-德数学家Erhard Schmidt(1876-1959)首次发现, 1936年美国物理学家Carl Henry Eckart(1902-1973)和Gale Young(1912-1990)证明F-范数下的矩阵版本, 1960年俄-英数学家Leon Mirsky(1918-1983)对酉不变范数证明了该定理.

[91]Chi-Kwong Li and Gilbert Strang, An elementary proof of Mirsky's low rank approximation theorem, Electron. J. Linear Algebra36(2020), 694–697.

设$M \in \mathbb{R}^{m\times n}, m_{ij}, ij \in \Omega$是$M$的已知元素. 矩阵完备化问题的目标是求下述问题的解:

$$\min_{x_{ij}=m_{ij}, ij\in\Omega} rank(X) \tag{5.6.3}$$

可以将M的未知元素全部取为0而得到一个新矩阵$M^\Omega = (m_{ij}^\Omega)$如下:

$$m_{ij}^\Omega = \begin{cases} m_{ij} & ij \in \Omega \\ 0 & ij \notin \Omega \end{cases} \tag{5.6.4}$$

则矩阵完备化问题(5.6.3)可以重新写为

$$\min_{X^\Omega = M^\Omega} rank(X) \tag{5.6.5}$$

公式(5.6.5)的求解是NP问题, 一般减低求解复杂度使用所谓**“凸松弛”**(convex relaxation)法, 为此需要回忆**核范数**的概念.

定义 5.6.1. 矩阵A的所有奇异值之和称为A的核范数, 记为$\|A\|_*$. 由于A的奇异值是A^*A的特征值的算术平方根, 故

$$\|A\|_* = \mathrm{tr}(A^*A) \tag{5.6.6}$$

核范数显然满足范数的齐次性和正定性两个条件, 但它还满足三角不等式不易证明, 请读者参考樊畿的原始论文[92], 其中证明了下面更强的结论.

定理 5.6.2. 设$A \in \mathbb{F}^{m\times n}, k \le q$. 记$\|A\|_{(k)}$为$A$的前$k$个大奇异值的和, 即

$$\|A\|_{(k)} = \sigma_1 + ... + \sigma_k \tag{5.6.7}$$

则$\|A\|_{(k)}$是范数.

公式(5.6.7)定义的范数称为A的(第k个)**樊畿范数**, 核范数恰好是第q个樊畿范数.

利用核范数可以将矩阵完备化问题(5.6.3)转化为复杂度降低的下述问题

$$\min_{X^\Omega = M^\Omega} \|X\|_* \tag{5.6.8}$$

第七节 应用II：矩阵函数的导数与常微分方程组

本节我们介绍矩阵函数的微积分在求解常微分方程组和现代控制理论等方面的应用.

5.7.1 线性常微分方程组

设有一个多输入-多输出连续时间系统, 如下图所示:

[92]K. Fan. Maximum properties and inequalities for the eigenvalues of completely continuous operators. Proc. Nat. Acad. Sci., U.S.A., 37:760-766, 1951.

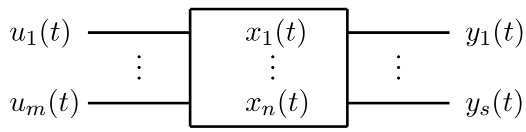

图 5.7.1

其中$u_1(t),\cdots,u_m(t)$是m个输入, $y_1(t),\cdots,y_s(t)$是s个输出, $x_1(t),\cdots,x_n(t)$是系统的n个状态变量. 记

$$x=x(t)=(x_1(t),\cdots,x_n(t))^T, u=u(t)=(u_1(t),\cdots,u_m(t))^T, y=y(t)=(y_1(t),\cdots,y_s(t))^T,$$

分别称为系统的状态向量, 输入向量 (或控制向量), 输出向量(或观测向量). 描述该系统的一种数学模型为所谓**状态方程**

$$\begin{cases} x'(t)=Ax(t)+Bu(t) \\ y(t)=Cx(t)+Du(t) \end{cases} \tag{5.7.1}$$

其中$A\in\mathbb{C}^{n\times n}$称为系数矩阵或**系统矩阵**, $B\in\mathbb{C}^{n\times m}$为输入矩阵(或**控制矩阵**), $C\in\mathbb{C}^{s\times n}$为输出矩阵(或**观测矩阵**), $D\in\mathbb{C}^{s\times m}$为联系矩阵. 由于联系矩阵$D$仅涉及信号的静态传递, 而与系统状态变量无关, 故可取$D=0$. 我们仅限于讨论A,B,C均为常数矩阵的系统, 这样的系统称为**定常线性系统**, 一般简称为系统(A,B,C). 方程组(5.7.1)中的第一个方程称为系统的状态方程, 第二个方程称为系统的**输出方程**.

由线性微分方程理论可知, 系统在某一时刻t_0的状态决定整个系统在任何时刻$t\geq t_0$的状态. 即若给定初始条件$x(t_0)=(x_1(t_0),\cdots,x_n(t_0))$, 则方程组(5.7.1)的解完全确定. 此称为定解问题, 其解决过程分两步, 即先求方程组(5.7.1)对应的齐次方程组的解, 再利用常数变易法或其他办法求方程组(5.7.1)的一个特解.

例 5.7.1. 当$n=1$时, 定解问题涉及的齐次微分方程就是普通的线性常微分方程$x'(t)=ax, x(t)|_{t=t_0}=x(t_0)$. 利用分离变量法直接积分即可得该定解问题的解为$x=x(t_0)e^{a(t-t_0)}$. 此时相应的非齐分方程为$x'(t)=ax+f(t), x(t)|_{t=t_0}=x(t_0)$, 其中$f(t)$为已知函数. 利用常数变易法可求得其解为

$$x=x(t_0)e^{a(t-t_0)}+\int_{t_0}^{t}e^{a(t-s)}f(s)\,\mathrm{d}\,s.$$

一般地, 定解问题涉及的齐次方程组为

$$\begin{cases} x'(t)=Ax \\ x(t)|_{t=t_0}=x(t_0), \end{cases} \tag{5.7.2}$$

其中$x=x(t)=(x_1(t),x_2(t),\cdots,x_n(t))^T$为未知向量函数, $x(t_0)=(x_1(t_0),x_2(t_0),\cdots,x_n(t_0))^T$为给定的$n$维向量, A为给定的n阶常数方阵.

当$n\geq 2$时, 方程(5.7.2)无法分离变量, 例 5.7.1的方法失效, 但可利用 Taylor 级数求解如下:

由$x'(t)=Ax$可知

$$x''(t)=A^2x,\cdots,x^{(k)}(t)=A^kx,\cdots \tag{5.7.3}$$

故

$$x'(t_0)=Ax(t_0), x''(t_0)=A^2x(t_0),\cdots,x^{(k)}(t_0)=A^kx(t_0),\cdots \tag{5.7.4}$$

将$x(t)$在$t=t_0$处展开成 Taylor 级数可得

$$\begin{aligned} x &= x(t_0)+(t-t_0)x'(t_0)+\frac{(t-t_0)^2}{2!}x''(t_0)+\cdots+\frac{(t-t_0)^k}{k!}x^{(k)}(t_0)+\cdots \\ &= x(t_0)+(t-t_0)Ax(t_0)+\frac{(t-t_0)^2}{2!}A^2x(t_0)+\cdots+\frac{(t-t_0)^k}{k!}A^kx(t_0)+\cdots \\ &= e^{A(t-t_0)}x(t_0) \end{aligned} \tag{5.7.5}$$

由公式(5.7.5)给出的解还是唯一的(见习题 55), 故有下述定理

定理 5.7.1. 定解问题(5.7.2)有唯一解, 且其解为

$$x(t)=e^{A(t-t_0)}x(t_0).$$

例 5.7.2. 求常系数线性齐次微分方程组

$$\begin{cases} x'(t)=Ax(t), \\ x(t)|_{t=0}=x(0), \end{cases}$$

的解, 其中

$$A=\begin{pmatrix} 2 & 2 & -1 \\ -1 & -1 & 1 \\ -1 & -2 & 2 \end{pmatrix},\ x(0)=\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

解 由定理 5.7.1, 方程组的解为$x(t)=e^{At}x(0)$. 下面求e^{At}. 直接计算可得A的 Jordan 标准形及变换矩阵分别为

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

从而,

$$\begin{aligned} x(t) &= e^{At}x(0)=Pe^{Jt}P^{-1}x(0) \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} e^t & 0 & 0 \\ 0 & e^t & te^t \\ 0 & 0 & e^t \end{pmatrix}\begin{pmatrix} 0 & -1 & 1 \\ 0 & -1 & 0 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 2t+1 \\ -2t+1 \\ -2t+1 \end{pmatrix}e^t. \end{aligned}$$

设$A=(a_{ij})_{n\times n}$与$B=(b_{ij})_{n\times m}$是常数矩阵, $u(t)=(u_1(t),u_2(t),...,u_m(t))^T$ 是已知向量函数, $x(t)=(x_1(t),x_2(t),...,x_n(t))^T$是未知向量函数. 下面考虑定解问题所涉及的线性常系数非齐次微分方程组

$$\begin{cases} x'(t)=Ax(t)+Bu(t) \\ x(t)|_{t=t_0}=x(t_0) \end{cases} \tag{5.7.6}$$

由线性微分方程组的理论可知, 方程组(5.7.6)的通解是其相应的齐次方程组的通解加上它自身的一个特解, 而该特解可以通过 Lagrange 常数变易法求得. 此处介绍另一个方法即积分因子法, 即给方程组(5.7.6)的两端同乘以e^{-At}, 得

$$e^{-At}\frac{\mathrm{d}\,x(t)}{\mathrm{d}\,t}-e^{-At}Ax(t)=e^{-At}Bu(t). \tag{5.7.7}$$

注意上式的左端为$(e^{-At}x(t))'$, 因而有下述定理

定理 5.7.2. 方程组(5.7.6)有唯一解, 且其解为

$$x(t)=e^{A(t-t_0)}x(t_0)+\int_{t_0}^{t}e^{A(t-s)}Bu(s)\,\mathrm{d}\,s.$$

注. 积分因子法等价于作变量代换$x=e^{At}y$或$y=e^{-At}x$.

例 5.7.3. 求定解问题

$$\begin{cases} x'(t)=Ax(t)+u(t) \\ x(t)|_{t=0}=(1,1,1)^T \end{cases}$$

的解, 其中

$$A=\begin{pmatrix} 3 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 2 \end{pmatrix},\ u(t)=(0,0,e^{2t})^T.$$

解 由定理 5.7.2, 方程组的解为

$$x(t)=e^{At}x(0)+\int_0^t e^{A(t-s)}u(s)\,\mathrm{d}\,s.$$

首先求e^{At}. 由于

$$|\lambda I-A|=\begin{vmatrix} \lambda-3 & 1 & -1 \\ -2 & \lambda & 1 \\ -1 & 1 & \lambda-2 \end{vmatrix}=\lambda(\lambda-2)(\lambda-3).$$

故A有 3 个不同的特征值, 因此A与对角矩阵相似. 与特征值$\lambda_1=0$, $\lambda_2=2$, $\lambda_3=3$相应的三个线性无关的特征向量为

$$\alpha_1=(1,5,2)^T,\quad \alpha_2=(1,1,0)^T,\quad \alpha_3=(2,1,1)^T.$$

于是

$$P=\begin{pmatrix} 1 & 1 & 2 \\ 5 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix},\quad P^{-1}=-\frac{1}{6}\begin{pmatrix} 1 & -1 & -1 \\ -3 & -3 & 9 \\ -2 & 2 & -4 \end{pmatrix}.$$

因而

$$
\begin{aligned}
e^{At}x(0) &= P\begin{pmatrix} 1 & & \\ & e^{2t} & \\ & & e^{3t} \end{pmatrix}P^{-1}x(0) \\
&= \begin{pmatrix} 1 & 1 & 2 \\ 5 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & & \\ & e^{2t} & \\ & & e^{3t} \end{pmatrix}\cdot(-\frac{1}{6})\begin{pmatrix} 1 & -1 & -1 \\ -3 & -3 & 9 \\ -2 & 2 & -4 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
&= -\frac{1}{6}\begin{pmatrix} -1+3e^{2t}-8e^{3t} \\ -5+3e^{2t}-4e^{3t} \\ -2-4e^{3t} \end{pmatrix}.
\end{aligned}
$$

下面计算积分$\int_0^t e^{A(t-s)}u(s)\,\mathrm{d}\,s$. 直接计算得

$$
e^{A(t-s)}u(s) = -\frac{1}{6}\begin{pmatrix} -e^{2s}+9e^{2t}-8e^{3t-s} \\ -5e^{2s}+9e^{2t}-4e^{3t-s} \\ -2e^{2s}-4e^{3t-s} \end{pmatrix}.
$$

$$
\int_0^t e^{A(t-s)}u(s)\,\mathrm{d}\,s = -\frac{1}{6}\begin{pmatrix} \frac{1}{2}+(9t+\frac{15}{2})e^{2t}-8e^{3t} \\ \frac{5}{2}+(9t+\frac{3}{2})e^{2t}-4e^{3t} \\ 1+3e^{2t}-4e^{3t} \end{pmatrix}.
$$

因此方程组的解为

$$
x(t) = -\frac{1}{6}\begin{pmatrix} -\frac{1}{2}+(9t+\frac{21}{2})e^{2t}-16e^{3t} \\ -\frac{5}{2}+(9t+\frac{9}{2})e^{2t}-8e^{3t} \\ -1+3e^{2t}-8e^{3t} \end{pmatrix}.
$$

高阶常系数齐次微分方程的定解问题可以转化为线性微分方程组来求解. 考虑下面的定解问题:

$$
\begin{cases} y^{(n)}+a_1y^{(n-1)}+a_2y^{(n-2)}+\cdots+a_ny=0, \\ y^{(i)}(t)|_{t=0}=y_0^{(i)}, \qquad 0\le i\le n-1. \end{cases} \tag{5.7.8}
$$

令

$$
\begin{cases} x_1=y, \\ x_2=y'=x_1', \\ x_3=y^{(2)}=x_2', \\ \cdots\cdots\cdots \\ x_n=y^{(n-1)}=x_{n-1}', \end{cases}
$$

则

$$\begin{cases} x_1' & = & x_2, \\ x_2' & = & x_3, \\ \cdots & \cdots & \cdots \\ x_{n-1}' & = & x_n, \\ x_n' & = & -a_n x_1 - a_{n-1} x_2 - \cdots - a_1 x_n. \end{cases}$$

令

$$x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T,\ x(0) = (x_1(0), x_2(0), \cdots, x_n(0))^T = (y_0, y_0', \cdots, y_0^{(n-1)})^T.$$

则高阶常微分方程(5.7.8)等价于一阶微分方程组

$$\begin{cases} x'(t) = Ax(t) \\ x(t)|_{t=0} = x(0) \end{cases} \tag{5.7.9}$$

其中

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}. \tag{5.7.10}$$

可以看出, A恰好是方程(5.7.8)的特征多项式$r^n + a_1 r^{n-1} + a_2 r^{n-2} + \cdots + a_n$的友矩阵. 由定理 5.7.1, 方程组(5.7.9)的解为$x(t) = e^{At}x(0)$. 而方程(5.7.8)的解是方程组(5.7.9)的解的第一个分量, 所以定解问题(5.7.8)的解为

$$\begin{aligned} y &= (1, 0, \cdots, 0)x(t) = (1, 0, \cdots, 0)e^{At}x(0) \\ &= (1, 0, \cdots, 0)e^{At} \begin{pmatrix} y_0 \\ y_0' \\ \vdots \\ y_0^{(n-1)} \end{pmatrix}. \end{aligned}$$

相应于方程(5.7.8)的n阶常系数非齐次线性方程的定解问题:

$$\begin{cases} y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \cdots + a_n y = u(t), \\ y^{(i)}(t)|_{t=0} = y_0^{(i)}, \quad 0 \le i \le n-1, \end{cases} \tag{5.7.11}$$

可作相应的讨论. 容易推出定解问题(5.7.11)的解是方程组

$$\begin{cases} x'(t) = Ax(t) + Bu(t), \\ x(t)|_{t=0} = x(0), \end{cases}$$

的解的第一个分量, 其中

$$
\begin{aligned}
&x(t)=(x_1(t),x_2(t),\cdots,x_n(t))^T=(y,y',\cdots,y^{(n-1)})^T,\\
&x(0)=(x_1(0),x_2(0),\cdots,x_n(0))^T=(y_0,y_0',\cdots,y_0^{(n-1)})^T,\\
&A=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & 1\\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix},\qquad B=\begin{pmatrix} 0\\ 0\\ \vdots\\ 0\\ 1\end{pmatrix}.
\end{aligned}
$$

由定理5.7.2知, 定解问题(5.7.11)的解为

$$
y(t)=(1,0,\cdots,0)\left(e^{At}x(0)+\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right).
$$

例 5.7.4. 求下述三阶常微分方程的解$y(t)$:

$$
\begin{cases} y^{(3)}-3y^{(2)}-6y'+8y=u(t),\\ (y(0),y'(0),y''(0))=(1,0,1). \end{cases}
$$

解 令$x(t)=(x_1(t),x_2(t),x_3(t))^T$, 其中$x_1(t)=y(t)$, $x_2(t)=y'(t)$, $x_3(t)=y''(t)$. 则$x(0)=(1,0,1)^T$.

由前面的讨论知,

$$
y(t)=(1,0,0)\left(e^{At}x(0)+\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right),
$$

其中

$$
A=\begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & 1\\ -8 & 6 & 3\end{pmatrix},\qquad B=\begin{pmatrix} 0\\ 0\\ 1\end{pmatrix}.
$$

A的特征多项式为$|\lambda I-A|=\lambda^3-3\lambda^2-6\lambda+8=(\lambda-1)(\lambda+2)(\lambda-4)$. 所以$A$的Jordan标准形为

$$
J=\begin{pmatrix} 1 & & \\ & -2 & \\ & & 4\end{pmatrix},\quad \text{其变换矩阵为}\quad P=\begin{pmatrix} 1 & 1 & 1\\ 1 & -2 & 4\\ 1 & 4 & 16\end{pmatrix}.
$$

因此

$$
\begin{aligned}
e^{At}&=P\begin{pmatrix} e^t & & \\ & e^{-2t} & \\ & & e^{4t}\end{pmatrix}P^{-1}\\
&=\begin{pmatrix} 1 & 1 & 1\\ 1 & -2 & 4\\ 1 & 4 & 16\end{pmatrix}\begin{pmatrix} e^t & & \\ & e^{-2t} & \\ & & e^{4t}\end{pmatrix}\cdot\frac{1}{18}\begin{pmatrix} 16 & 4 & -2\\ 4 & -5 & 1\\ -2 & 1 & 1\end{pmatrix}\\
&=\frac{1}{18}\begin{pmatrix} 16e^t+4e^{-2t}-2e^{4t} & 4e^t-5e^{-2t}+e^{4t} & -2e^t+e^{-2t}+e^{4t}\\ 16e^t-8e^{-2t}-8e^{4t} & 4e^t+10e^{-2t}+4e^{4t} & -2e^t-2e^{-2t}+4e^{4t}\\ 16e^t+16e^{-2t}-32e^{4t} & 4e^t-20e^{-2t}+16e^{4t} & -2e^t+4e^{-2t}+16e^{4t}\end{pmatrix},
\end{aligned}
$$

$$e^{A(t-s)}Bu(s)=\frac{1}{18}u(s)\begin{pmatrix}-2e^{t-s}+e^{-2(t-s)}+e^{4(t-s)}\\-2e^{t-s}-2e^{-2(t-s)}+4e^{4(t-s)}\\-2e^{t-s}+4e^{-2(t-s)}+16e^{4(t-s)}\end{pmatrix}.$$

所以

$$\begin{aligned}y(t)&=(1,0,0)\left(e^{At}x(0)+\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right)\\&=\frac{1}{18}(14e^t+5e^{-2t}-e^{4t})-\frac{1}{9}e^t\int_0^t e^{-s}u(s)\,\mathrm{d}\,s+\frac{1}{18}e^{-2t}\int_0^t e^{2s}u(s)\,\mathrm{d}\,s\\&\quad+\frac{2}{9}e^{4t}\int_0^t e^{-4s}u(s)\,\mathrm{d}\,s.\end{aligned}$$

注1. 如果仅求n阶齐次方程(5.7.8)的解, 则不必使用本节介绍的办法, 因为一旦知道特征方程的根, 则方程(5.7.8)的通解就已经知道了.

注2. 实际求解定常线性系统(5.7.1)的常用方法是Laplace变换.

5.7.2 线性系统的可控性与可测性

本节利用上节的理论研究定常线性系统(5.7.1)的可控性与可观测性.

控制定常系统(5.7.1)的关键是了解系统的状态$x(t)$. 但一般情况下很难直接测量$x(t)$ (比如人造卫星系统或深水探测仪等), 而必须通过观测向量$y(t)$反过来判断$x(t)$. 能否通过观测向量$y(t)$确定出系统的全部状态, 这就是所谓**可观测性问题**. 掌握了系统的状态后, 能否控制它使其达到预期目的, 便是所谓**可控制性问题**.

定义 5.7.1. 对一个定常线性系统, 若在有限时间区间$[0,t_1]$内存在输入$u(t)$ $(0\le t\le t_1)$, 能使系统从初始状态$x(0)$转移到$x(t)=0$, 则称此状态$x(0)$是可控的. 若任意初始状态$x(0)$都是可控的, 则称该系统是**完全可控的**或**完全能控的**.

注. 可以证明, “完全能控”等价于对任意初始状态$x(0)$, 均存在控制信号使得系统在有限时间内变成任意事先指定的状态$x(t)$(而不仅仅是使系统状态变为0).

例 5.7.5. 设$n=1$, 则定常系统实际上是一个一阶线性微分方程$x'(t)=ax(t)+bu(t)$. 显然该系统完全可控 $\Longleftrightarrow b\neq 0$.

例 5.7.6. 设定常系统的状态方程为

$$x'(t)=\begin{pmatrix}\lambda_1&0\\0&\lambda_2\end{pmatrix}x(t)+\begin{pmatrix}a\\b\end{pmatrix}u(t),$$

则该系统完全能控 $\Longleftrightarrow$ $ab\neq 0$且$\lambda_1\neq\lambda_2$.

定理 5.7.3. 系统(A,B,C)完全能控 $\Longleftrightarrow$ n阶Hermite矩阵

$$W(0,t_1)=\int_0^{t_1}e^{-tA}BB^*e^{-tA^*}\,\mathrm{d}\,t \tag{5.7.12}$$

为非奇异矩阵.

证 充分性: 设$W(0,t_1)$非奇异. 由上一节的定理 5.7.2知,

$$x(t_1)=e^{At_1}x(0)+\int_0^{t_1}e^{A(t_1-t)}Bu(t)\,\mathrm{d}\,t. \tag{5.7.13}$$

令

$$u(t)=-B^*e^{-A^*t}W(0,t_1)^{-1}x(0). \tag{5.7.14}$$

将$u(t)$代入(5.7.13)得

$$\begin{aligned}
x(t_1)&=e^{At_1}x(0)-\int_0^{t_1}e^{A(t_1-t)}BB^*e^{-A^*t}W(0,t_1)^{-1}x(0)\,\mathrm{d}\,t\\
&=e^{At_1}x(0)-e^{At_1}\left(\int_0^{t_1}e^{-At}BB^*e^{-A^*t}\,\mathrm{d}\,t\right)W(0,t_1)^{-1}x(0)\\
&=e^{At_1}x(0)-e^{At_1}W(0,t_1)W(0,t_1)^{-1}x(0)=0.
\end{aligned}$$

这说明在控制输入$u(t)$作用下, 能使系统$x(0)$转移到$x(t_1)=0$. 由于$x(0)$的任意性, 因此系统是完全能控的.

必要性: 设有向量$\alpha=(a_1,a_2,\cdots,a_n)^T$满足

$$W(0,t_1)\alpha=0. \tag{5.7.15}$$

则$\alpha^*W(0,t_1)\alpha=0$. 因而

$$\int_0^{t_1}\alpha^*(e^{-At}BB^*e^{-A^*t})\alpha\,\mathrm{d}\,t=0.$$

因此

$$\alpha^*e^{-At}B=0,\quad 0\le t\le t_1. \tag{5.7.16}$$

由于系统是完全能控的, 所以存在某个输入$u(t)$使得$x(t_1)=0$. 故由公式(5.7.13), 有

$$e^{At_1}x(0)+\int_0^{t_1}e^{A(t_1-t)}Bu(t)\,\mathrm{d}\,t=0.$$

所以

$$x(0)=-e^{-At_1}\int_0^{t_1}e^{A(t_1-t)}Bu(t)\,\mathrm{d}\,t=-\int_0^{t_1}e^{-At}Bu(t)\,\mathrm{d}\,t.$$

因而由等式(5.7.16),

$$\alpha^*x(0)=-\int_0^{t_1}\alpha^*e^{-At}Bu(t)\,\mathrm{d}\,t=0.$$

由$x(0)$的任意性知$\alpha=0$, 与假设$\alpha\neq 0$矛盾. 故$W(0,t_1)$是非奇异的. □

由于定理 5.7.3涉及到矩阵函数的积分, 所以在实际应用中常使用下面的简单判别准则.

定理 5.7.4. 系统(A,B,C)完全能控 $\Longleftrightarrow$ 矩阵

$$W=(B,AB,A^2B,\cdots,A^{n-1}B)$$

的秩为n. 一般将$n\times nm$阶矩阵W称为**能控性矩阵**.

证 系统(A,B,C)完全能控 $\Longleftrightarrow$ 对任意$x(0)$, 关于未知向量$u(t)$的方程(5.7.13)(其中$x(t_1)=0$) 有解, 即方程

$$\int_0^{t_1} e^{A(t_1-t)}Bu(t)\,\mathrm{d}\,t = -e^{At_1}x(0) \tag{5.7.17}$$

对于任意$x(0)$总有解. 因为矩阵e^{At_1}总可逆, 故方程(5.7.17)可以化简为

$$\int_0^{t_1} e^{-At}Bu(t)\,\mathrm{d}\,t = -x(0) \tag{5.7.18}$$

由 Cayley-Hamilton 定理可知, n阶矩阵e^{-At}可表示为

$$e^{-At}=\sum_{i=0}^{n-1}a_i(t)A^i,$$

因此方程(5.7.18)又可写为

$$\sum_{i=0}^{n-1}(A^iB)\int_0^{t_1}(a_i(t)u(t))\,\mathrm{d}\,t = -x(0) \tag{5.7.19}$$

记

$$z_i(t)^T=\int_0^{t_1}a_i(t)u(t)\,\mathrm{d}\,t,\ 0\le i\le n-1.$$

令$z=(z_0(t),z_1(t),\cdots,z_{n-1}(t))^T$, 则等式(5.7.19)为

$$(B,AB,A^2B,\cdots,A^{n-1}B)z=-x(0) \tag{5.7.20}$$

于是系统(A,B,C)完全能控 $\Longleftrightarrow$ 方程组(5.7.20)的系数矩阵的秩是n. □

思考. 上面证明中的最后一句话为什么成立? 注意此时向量z的维数并不是未知数的个数, 那么, 方程(5.7.20)中未知数的个数是多少呢?

由定理 5.7.4可知, 如果控制矩阵B是行满秩的, 则系统一定是可控的.

为了更好地理解连续型的定常线性系统的可控性(以及可测性), 以下我们对照离散型的定常线性系统. 类似于连续型的定常系统(5.7.1), 离散型的定常系统的一个简单数学模型是如下的(矩阵)差分方程:

$$\begin{cases} x(k+1)=Ax(k)+Bu(k), & k\ge 0\\ y(k)=Cx(k), & k\ge 0\end{cases} \tag{5.7.21}$$

其中$x(k),u(k),y(k)$分别看作是连续型的定常线性系统的状态变量, 控制信号和观测信号的离散化, 而离散系统(5.7.21)中的两个方程可以分别看作是连续型的定常线性系统的状态方程与输出方程的离散化. 离散系统(5.7.21)也简记为(A,B,C). 利用迭代法可知, 方程(5.7.21)可以写为

$$x(k)=A^kx(0)+\sum_{i=0}^{k-1}A^{k-1-i}Bu(i),k\ge 1 \tag{5.7.22}$$

特别地,

$$x(n)=A^nx(0)+(B,AB,\cdots,A^{n-1}B)(u(n-1),u(n-2),\cdots,u(0))^T \tag{5.7.23}$$

上式正是定理 5.7.4中能控性矩阵的来历. 显然, 离散型的定常系统(A,B,C)与连续型的定常系统(A,B,C)具有完全相同的可控制性.

例 5.7.7. 已知

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

试判断系统(A,B,C)是否完全能控.

解 由条件知, 系统(A,B,C)具有三个状态变量, 两个输入. 下面用定理5.7.4来判断.

$$W=\begin{pmatrix} 2 & 1 & 1 & 1 & -5 & 0 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ -1 & -1 & -1 & -2 & 0 & -1 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & -5 & -1 & -13 & -4 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ 0 & -1 & 2 & -1 & 4 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 0 & 1 & -5 & -1 & -13 & -4 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ 0 & 0 & -3 & -2 & -9 & -3 \end{pmatrix}.$$

所以W的秩为3, 由此知系统是完全能控的.

例 5.7.8. 如果定常系统是单输入系统, 则控制矩阵B为n维列向量, 因此系统的能控性矩阵W是一个n阶方阵. 由定理5.7.4, 此时系统可控 $\Longleftrightarrow$ 能控性矩阵W是可逆矩阵. 这就给出可控性的一个直观解释: 即控制信号可以使系统的任何两个状态相互转化, 换言之, 系统的变化是一个可逆的过程.

如果B的列数大于1, 则能控性矩阵W具有较多的列, 因此利用W的秩判断系统的可控性时, 应注意避免做过多无效计算, 可以先计算矩阵B的秩, 然后再计算矩阵(B,AB)的秩, 等等, 因为计算A的高次幂往往较计算较低次幂大为困难.

下面我们讨论定常系统的可观测性.

定义 5.7.2. 对于一个定常线性系统, 若在有限时间区间$[0,t_1]$内能通过观测系统的输出$y(t)$而唯一地确定初始状态$x(0)$, 则称此系统是**完全能观测** 或**完全可观测**的或完全可测的.

注意上面定义中的输出$y(t)$是指时间区间$[0,t_1]$内的所有输出, 因此所谓连续型的定常系统的可观测性是用一个时间段内的所有(当然是无限多个)观测信号(即系统的反馈)来反推系统的一个(初始)状态. 这是合理的, 因为观测矩阵C一般来说是不可逆的, 因此不能指望用一个反馈信号就能确定系统的初始状态.

例 5.7.9. 若观测矩阵C是可逆矩阵, 则定常系统显然是完全可观测的.

例 5.7.10. 设定常系统的系统方程为

$$\begin{cases} x'(t)=\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}x(t)+Bu(t) \\ y(t)=(a,b)x(t) \end{cases},$$

则该系统完全可测 $\Longleftrightarrow$ $ab\neq 0$且$\lambda_1\neq\lambda_2$. 证明留作习题.

定理 5.7.5. 系统(A,B,C)完全能观测 $\Longleftrightarrow$ n阶 Hermite 矩阵

$$M(0,t_1)=\int_0^{t_1} e^{A^*t}C^*Ce^{At}\,\mathrm{d}\,t$$

为非奇异矩阵.

证 充分性: $y(t)=Cx(t)=Ce^{At}x(0)+C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s$. 所以

$$Ce^{At}x(0)=y(t)-C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s.$$

上式两边左乘$e^{A^*t}C^*$, 并从 0 到t_1积分, 得

$$M(0,t_1)x(0)=\int_0^{t_1} e^{A^*t}C^*\left(y(t)-C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right)\mathrm{d}\,t.$$

由于$M(0,t_1)$ 非奇异, 所以

$$x(0)=M(0,t_1)^{-1}\int_0^{t_1} e^{A^*t}C^*\left(y(t)-C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right)\mathrm{d}\,t.$$

这说明$x(0)$是唯一确定的, 因而系统是完全能观测的.

必要性: 设系统是完全能观测的. 则满足

$$y(t)=Ce^{At}x(0)+C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s \tag{5.7.24}$$

的$x(0)$是唯一确定的. 若$M(0,t_1)$奇异, 则存在非零的$\alpha\in\mathbb{C}^n$满足

$$M(0,t_1)\alpha=0.$$

因而$\alpha^*M(0,t_1)\alpha=0$. 于是

$$\int_0^{t_1}\alpha^*e^{A^*t}C^*Ce^{At}\alpha\,\mathrm{d}\,t=0.$$

由此得到

$$y(t)=Ce^{At}(x(0)+\alpha)+C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s.$$

这说明若$x(0)$满足等式(5.7.24), 则$x(0)+\alpha$也满足(5.7.24). 因而, 由$\alpha\neq 0$知, $x(0)$ 不是唯一确定的, 矛盾. 所以$M(0,t_1)$非奇异. □

类似于可控性的情形, 离散型的定常系统(A,B,C)与连续型的定常系统(A,B,C)具有完全相同的可测性的下列简单判别准则, 证明留作习题.

定理 5.7.6. 系统(A,B,C)完全能观测 $\Longleftrightarrow$ 矩阵

$$M=\begin{pmatrix} C\\ CA\\ \vdots\\ CA^{n-1}\end{pmatrix}$$

的秩为n. 矩阵M称为能观测性矩阵.

例 5.7.11. 设某系统的状态方程与输出方程为

$$\begin{cases} x'(t) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} x(t) + \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} u(t), \\ \\ y(t) = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} x(t). \end{cases}$$

试判断这系统的能控性与能观测性.

解 由于控制矩阵B是可逆的, 所以系统是完全能控的.

系统的能观测性矩阵为

$$M = \begin{pmatrix} C \\ CA \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 0 & 2 \\ 0 & 2 \end{pmatrix}.$$

所以M的秩为2, 说明系统是完全能观测的.

本节最后, 我们讨论定常系统的状态方程的化简问题.

定常系统的状态变量可以有不同的选取, 状态变量的所有线性组合构成一个线性空间. 因此, 选择适当的状态变量实际上就是选择该线性空间一组适当的基. 利用不同的状态变量之间的线性关系可以有效地简化系统的系统方程.

设定常系统的系统矩阵A的Jordan标准形为$J = PAP^{-1}$, 其中P为变换矩阵. 现令$z(t) = P^{-1}x(t)$或$x(t) = Pz(t)$, 则定常系统(A, B, C)可以简化为

$$\begin{cases} z'(t) = Jz(t) + P^{-1}Bu(t) \\ \\ y(t) = CPz(t) \end{cases} \tag{5.7.25}$$

利用简化的系统方程(5.7.25)可以较为方便地判断系统的测控性, 见下面的例子.

例 5.7.12. 对单输入定常系统(A, B, C), 如果A的特征值均不相同, 则系统可控 $\Longleftrightarrow$ 控制矩阵$P^{-1}B$无零行, 而系统可测 $\Longleftrightarrow$ 观测矩阵CP无零列, 详见习题63与66.

思考题

1. 在例5.7.12中, 能否用A为单纯矩阵代替条件“A的特征值均不相同”?
2. 定常系统(A, B, C)与其简化的系统(5.7.25)的可控性与可测性是否完全一致?

习　题　五

1. 设$\|\cdot\|$是酉空间$\mathbb{C}^n$的向量范数, 证明向量范数的下列基本性质:

(1) 零向量的范数为零.

(2) 当x是非零向量时: $\left\| \dfrac{x}{\|x\|} \right\| = 1$.

(3) $\|-x\| = \|x\|$.

(4) $|\,\|x\| - \|y\|\,| \le \|x - y\|$.

2. 证明: 若$x \in \mathbb{C}^n$, 则

(1) $\|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2$.

(2) $\|x\|_\infty \le \|x\|_1 \le n\|x\|_\infty$.

(3) $\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$.

3. (1) 试构造$\mathbb{R}^2$上的一个向量范数, 使得该范数不是任何p-范数.

(2) 画出你构造的范数的单位圆.

(3) 试对$\mathbb{R}^3$做(1)与(2), 并比较你的单位球与1-范数和∞-范数的单位球.

(4) 证明当$0 < p < 1$时, l_p范数仍然满足向量范数的前两个条件, 但不满足三角不等式. 在平面上画出$p = 1/2, 3/2$时的单位圆, 并就$p < 1$与$p \ge 1$的一般情形作比较.

4. 证明 Minkowski 不等式(5.1.1).

5. 证明樊畿范数是酉不变范数.

6. 验证例 5.1.4.

7. 证明命题 5.1.1中的正定性与齐次性.

8. 证明命题 5.1.3.

9. 证明例 5.1.10中的两个范数不等价.

10. 证明赋范线性空间中的单位球S是凸集, 即若$x, y \in S$, 则$\alpha x + \beta y \in S$也属于单位球, 其中α, β 为正数且$\alpha + \beta = 1$.

11. 验证矩阵的极大列和范数与极大行和范数均满足次乘性.

12. 设矩阵A的 F-范数等于a, U是酉矩阵, 问AU与UA的 F-范数各是多少? 请总结你的计算.

13. 证明矩阵的 1-范数, 2-范数和∞-范数分别是向量的 1-范数, 2-范数和∞-范数的诱导范数(因此与之相容).

14. 证明:(1) 矩阵仿照向量的 1-范数是矩阵范数, 且与 1-范数相容, 但不是 1-范数的诱导范数, 试求向量的 1-范数诱导的矩阵范数.

(2) 矩阵仿照向量的∞-范数是向量范数但不是矩阵范数.

15. 证明公式(5.2.4).

16. (1) 证明$\|A\|_2 = (\rho(A^*A))^{1/2}$定义了一个矩阵范数, 称为$A$的**谱范数** .

(2) 试求一个与矩阵的谱范数相容的向量范数.

(3) 证明若A是正规矩阵, 则A的谱范数就是其谱半径$\rho(A)$.

(4) 设V是由全体 Hermite 矩阵构成的复线性空间, 证明谱半径是V上的向量范数. 该范数是矩阵范数吗?

17. 试构造两种矩阵范数使得一个矩阵A的两种范数分别为 2 与$1/3$. 能否使所有非零矩阵的两种范数之积等于 1?

18. (1) 证明向量范数的代数性质: 有限种向量范数的任意正线性组合仍是向量范数.

(2) 设$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$是两种向量范数或矩阵范数, $p > 0$. 判断

$$[(\|\cdot\|_\alpha)^p + (\|\cdot\|_\beta)^p\,]^{1/p}$$

是否为向量范数或矩阵范数?

(3) 判断矩阵范数是否有与向量范数相同的代数性质(1)?

19. 利用特征值的定义直接证明矩阵A的谱半径不超过矩阵A的任何一种矩阵范数. 此结论可以换成矩阵的任何一种向量范数吗?

20. 证明公式(5.2.5)定义了一个与矩阵范数$\|\cdot\|$相容的向量范数.

21. 设T为正交矩阵, 又$A \in \mathbb{R}^{n\times n}$. 证明:

(1) $\|T\|_2 = 1$.

(2) $\|A\|_2 = \|TA\|_2$.

(3) 试解释上面的两个结果.

22. 设A, B为n阶矩阵, 其中A可逆而B不可逆, 设$\|\cdot\|$是任何一种矩阵范数. 定义A的条件数$Cond(A) =$

$\|A\|\|A^{-1}\|$. 证明: $\|A-B\| \geq 1/\|A^{-1}\|$. 解释这个结果.

23. (奇异值与矩阵的范数) 设$A \in \mathbb{C}^{m\times n}$, $\sigma_1, \sigma_2, \cdots, \sigma_n$是$A$的全部奇异值. 证明:
(1) $Cond(A) = \sigma_1(A)/\sigma_n(A)$, 其中$\sigma_1(A)$与$\sigma_n(A)$分别是$A$的最大和最小奇异值.(参考第四章例4.4.6.)
(2) $\|A\|_F = \left(\sum_{i=1}^{r} \sigma_i^2\right)^{1/2} = (\mathrm{tr}(A^*A))^{1/2}$.
(3) $\|A\|_2 = \sigma_{\max}(A)$.

24. (1) 证明定理5.2.2.
(2) 设U, V是任意赋范线性空间(不必有限维), $\sigma \in \mathrm{Hom}(U, V)$. 证明: σ连续 $\Longleftrightarrow$ σ有界.

25. 证明引理5.3.1与定理5.3.1.

26. 设$A_k = \begin{pmatrix} \dfrac{1}{k^2} & \dfrac{k^2+k}{k^2+1} \\ 2 & (1-\dfrac{2}{k})^k \end{pmatrix}$, 求$\lim\limits_{k\to\infty} A_k$.

27. 证明Cullen定理(1965年)：$\lim\limits_{n\to\infty} A^n = 0$当且仅当存在正定矩阵$P$使得$P - A^*PA$是正定矩阵.

28. 若$\lim\limits_{n\to\infty} A^n = B$, 则$B$为幂等矩阵.

29. 证明命题5.3.1.

30. 设$A = \begin{pmatrix} 2 & -\dfrac{1}{2} \\ 2 & 0 \end{pmatrix}$, 求$\sum\limits_{k=0}^{\infty} \dfrac{A^k}{2^k}$.

31. 设$A = \begin{pmatrix} -0.6 & 1 & 0.8 \\ 0 & 0.2 & 0 \\ -0.6 & 1 & 0.8 \end{pmatrix}$. 试判断$A$是否幂收敛.

32. (1) 设$A = \begin{pmatrix} 0 & 0 \\ 1 & -2 \end{pmatrix}$, 求$e^A$, $\sin A$, $\cos A$.
(2) 已知$J = \begin{pmatrix} -2 & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 2 \end{pmatrix}$, 求$e^J$, $\sin J$, $\cos J$.

33. (1) 证明定理5.3.7.
(2) 利用(1) 求第四章习题1中所有正规矩阵的指数函数, 正弦函数和余弦函数.

34. 证明命题5.3.4与命题5.4.1(2).

35. 对下列方阵A, 求矩阵函数e^{At}:

$$(1)A = \begin{pmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix}, \quad (2)A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & -12 & -6 \end{pmatrix}, \quad (3)A = \begin{pmatrix} -2 & 1 & 3 \\ 0 & -3 & 0 \\ 0 & 2 & -2 \end{pmatrix}.$$

36. 求下列两类矩阵的矩阵函数: $\cos A$, $\sin A$, e^A:
(1) A为幂等矩阵.
(2) A为对合矩阵(即$A^2 = I$).

37. 设函数矩阵$A(t) = \begin{pmatrix} \sin t & \cos t & t \\ \dfrac{\sin t}{t} & e^t & t^2 \\ 1 & 0 & t^3 \end{pmatrix}$, 其中$t \neq 0$. 计算$\lim\limits_{t\to 0} A(t)$, $\dfrac{\mathrm{d}}{\mathrm{d}\,t}A(t)$, $\dfrac{\mathrm{d}^2}{\mathrm{d}\,t^2}A(t)$.

38. 设函数矩阵$A(t) = \begin{pmatrix} e^{2t} & te^t & t^2 \\ e^{-t} & 2e^{2t} & 0 \\ 3t & 0 & 0 \end{pmatrix}$, 计算$\int_0^1 A(t)\,\mathrm{d}\,t$和$\dfrac{\mathrm{d}}{\mathrm{d}\,t}\int_0^{t^2} A(s)\,\mathrm{d}\,s$.

39. 证明:(1) 若A为实反对称矩阵, 则e^A为正交矩阵.
(2) 若A为Hermite阵, 则$e^{\mathrm{i}A}$为酉矩阵.

40. 详细证明定理5.4.2.

41. 证明Lagrange插值公式并利用线性空间的直和分解理论解释之.

42. (1) 设$J_n(\lambda)$是一个n阶Jordan块, 求$\sin Jt$, $\cos Jt$.
(2) 对任意n阶矩阵A, 导出$\sin At$与$\cos At$的一般表达式.

43. 设$A=\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$, 求$e^{At}$, $\sin At$, $\cos At$.

44. 设$A=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, 求$e^{At}$, $\sin At$, $\cos At$.

45. 证明命题5.4.3即Lagrange-Sylvester插值公式.(提示: 研究商$\dfrac{f(x)}{m_A(x)}$或利用极限研究无重根的Lagrange插值公式.)

46. 设N是n阶幂零块, 验证$\dfrac{\mathrm{d}e^{Nt}}{\mathrm{d}t}=Ne^{Nt}$并计算$\displaystyle\int_0^t e^{Ns}\mathrm{d}s$.

47. (1) 设$A=\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, 计算积分$\displaystyle\int_0^t e^{As}\mathrm{d}s$.
(2) 设$A=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, 计算$e^A$与$e^{At}$.
(3) 设$A^2=A$, 计算e^{At}与$\displaystyle\int_0^t e^{As}\mathrm{d}s$.

48. 设$A^2-A+2I=0$, 计算e^{At}与$\displaystyle\int_0^t e^{As}\mathrm{d}s$.

49. (1) 证明Jacobian猜想对线性映射成立.
(2) 证明Jacobian猜想的逆命题成立.
(3) 试求例5.5.7中多项式映射的逆映射. 试对次数不超过2(即每个分量的次数不超过2)的二元多项式映射证明Jacobian猜想. 如何将你的证明推广到n元?

50. 如果$f=(f_1(x),\cdots,f_n(x))^T$与$g=(g_1(x),\cdots,g_n(x))^T$是两个$n$元列映射, 计算$J(f^Tg)$并与乘法公式(命题5.5.1(2))比较.

51. 设$x=(x_1,x_2,\cdots,x_n)^T$, A是n阶矩阵, 求n元函数$(\|x-Ax\|_2)^2$的导数与Hessian矩阵.

52. 计算例5.5.9的复合函数的导数并与多元函数的链法则比较.

53. 证明命题5.5.2.

54. (1) 证明命题5.5.3, 并以此推出$J(\mathrm{tr}(AX))$与$J(\mathrm{tr}(XB))$.
(2) 验证公式(5.5.18).

55. 计算一般线性变换的导数, 证明例5.5.3的结论.

56. 验证公式(5.7.5)给出定解问题(5.7.2)的解且该解是唯一的.

57. 求下列微分方程组的通解：

$$(1)x'(t)=\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}x(t);\qquad (2)x'(t)=\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}x(t).$$

58. 求下列微分方程组$x'(t)=Ax(t)$满足初始条件$x(0)$的解：

$$(1)\ A=\begin{pmatrix} 1 & 12 \\ 3 & 1 \end{pmatrix}, x(0)=\begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad (2)\ A=\begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \\ 2 & 0 & 1 \end{pmatrix}, x(0)=\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

59. (1) 求解微分方程组

$$x'(t)=\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & -2 \\ 3 & 2 & 1 \end{pmatrix}x(t)+\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}e^{2t}.$$

(2) 求$x'(t)=Ax(t)+Bu(t)$满足初始条件$x(0)$的解, 其中

$$A=\begin{pmatrix} -6 & 1 & 0 \\ -11 & 0 & 1 \\ -6 & 0 & 0 \end{pmatrix},\ B=\begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix},\ u(t)=1,\ x(0)=\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

60. 求方程$y'''+6y''+11y'+6y=e^{-t}$满足$y(0)=y'(0)=y''(0)=0$的解.

61. (1) 证明微分方程$x'(t)=Ax(t)+\gamma e^{at}$有形如$x(t)=\beta e^{at}$的解 $\iff$ $(\alpha I-A)\beta=\gamma$, 其中β, γ都是n维向量, $a\in\mathbb{C}$.

(2) 解$x'(t)=Ax(t)+e^{2t}C$, 其中

$$A=\begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix},\ C=\begin{pmatrix} -1 \\ -1 \end{pmatrix},\ x(0)=\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

63. (1) 设定常系统的系统矩阵A是对角矩阵, 试给出该系统可控性的一个判断准则.

(2) 设定常系统的系统矩阵A是一个 Jordan 块, 试给出该系统可控性的一个判断准则.

(3) 设单输入定常系统的系统矩阵A是 Frobenius 标准形(即(5.7.10)中的矩阵), 而其控制矩阵B为标准向量e_n, 证明该系统是可控的. 这样的定常系统称为可控标准形.

64. 根据你在 63 题中给出的判断准则, 研究下面几个定常系统的可控性.

(1) $A=\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, $B=(1,1)^T$; (2)$A=\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$, $B=(1,0)^T$;

(3) $A=\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, $B=(c,d)^T$; (4)$A=\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$, $B=(c,d)^T$.

67. (1) 设定常系统的系统矩阵A是对角矩阵, 试给出该系统可测性的一个判断准则.

(2) 设定常系统的系统矩阵A是一个 Jordan 块, 试给出该系统可测性的一个判断准则.

(3) 设定常系统的系统矩阵A是 Frobenius 标准形的转置矩阵, 而其输出矩阵C为标准行向量e_n^T, 证明该系统是可测的. 这样的定常系统称为可测标准形.

68. 根据你在 67 题中给出的判断准则, 研究下面几个定常系统的可测性.

(1) $A=\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$, $C=(c,d)^T$; (2)$A=\begin{pmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix}$, $C=(c,d,f)^T$.

研究性问题.

中国数学家对Jacobi猜想的研究处于世界领先水平, 下面仅介绍严丹[93]的一个研究成果：

定理. **设**$A\in\mathbb{C}^{n\times n}, X=(x_i)\in\mathbb{C}^n$**. 如果多项式映射**$F(X)=X+(AX)^{*3}:=(x_1+(a_{11}x_1+\cdots+a_{1n}x_n)^3,\cdots,x_n+(a_{n1}x_1+\cdots+a_{nn}x_n)^3)^T$**满足条件**$tr[J((AX)^{*3})]=0$**,**

[93]Dan Yan(1987-), 湖南师范大学教授.

则$r(A) \le \dfrac{n+\delta}{2}$，其中$\delta$是$A$的对角线上零的个数.

相关论文参见[94].

[94] Dan Yan, A note on the Jacobian Conjecture. Linear Algebra and its Applications, 435(2011): 2110-2113.

第六章　广义逆矩阵与非负矩阵

本 章 提 要

本章介绍广义逆矩阵与不可约非负矩阵的相关理论. 广义逆矩阵是逆矩阵的推广, 不可约非负矩阵是正矩阵(元素均为正数)的推广. 主要结果如下：

1. **Moore-Penrose定理.** 任意$A \in \mathbb{F}^{m\times n}$均有唯一的广义逆矩阵$A^\dagger$, 使得

$$AA^\dagger A = A, A^\dagger AA^\dagger == A^\dagger, (AA^\dagger)^* = AA^\dagger, (A^\dagger A)^* = A^\dagger A.$$

2. **Perron-Froubenius定理.** 任意不可约非负矩阵$A \in \mathbb{R}^{n\times n}$的谱半径是单特征值, 且存在分量均为正的特征向量.

第一节　广义逆矩阵

6.1.1　Moore-Penrose 广义逆

1920 年, Moore[95]提出了$m \times n$矩阵的**广义逆矩阵**的概念. 若$A \in \mathbb{C}^{m\times n}$, Moore 意义下的广义逆为满足

$$AX = P_{R(A)} \quad 与 \quad XA = P_{R(X)} \tag{6.1.1}$$

的矩阵$X \in \mathbb{C}^{n\times m}$, 其中$P_L$表示子空间$L \subseteq \mathbb{C}^n$上的投影矩阵. 上世纪四十年代, 曾远荣[96]将矩阵的广义逆推广到一般内积空间上的算子(未必线性)广义逆, 并将其归为16类, 因此被视为算子广义逆理论的奠基人. 1955 年, Penrose[97]给出了广义逆矩阵的下述定义: 如果矩阵$X \in \mathbb{C}^{n\times m}$满足下述方程组

$$\begin{array}{ll} (1) & AXA = A \\ (2) & XAX = X \\ (3) & (AX)^* = AX \\ (4) & (XA)^* = XA \end{array} \tag{6.1.2}$$

则称X是矩阵$A \in \mathbb{C}^{m\times n}$的广义逆矩阵.

方程组(6.1.1)与(6.1.2)分别称为**Moore 方程组** 与 **Penrose 方程组**. 广义逆矩阵一直是矩阵理论的研究热点, 并广泛应用于数理统计, 多元分析, 最优化理论, 控制论, 网络理论等众多学科.

通常, 将满足 Penrose 方程组中等式$i_1, \cdots, i_j$的矩阵X称为矩阵A的$\{i_1, \cdots, i_j\}$-逆, 比如满足第一及第三个等式的矩阵X称为A的$\{1,3\}$-逆, 记为$A^{(1,3)}$; 而矩阵A的 **Moore-Penrose 广义逆**$A^\dagger = A^{(1,2,3,4)}$. 因此, 一个矩阵$A$共有(至少) 15 种广义逆, 研究最深、应用最广的当属 Moore-Penrose 广义逆$A^\dagger$以及$\{1\}$-逆 $A^{(1)}$, 通常记为A^-. 本节主要介绍$m \times n$矩阵的 Moore-Penrose 广义逆$A^\dagger$, $\{1\}$-逆A^-, $\{1,3\}$-逆与$\{1,4\}$-逆, 以及这几类广义逆矩阵的一些应用.

定义 6.1.1. 设矩阵$A \in \mathbb{C}^{m\times n}$, 若矩阵$X \in \mathbb{C}^{n\times m}$满足 Penrose 方程组(6.1.2), 则称X为A的一个 Penrose 广义逆(矩阵).

---

[95]Eliakim Hastings Moore(1862-1932), 美国数学家, 曾任美国数学会主席.

[96]Tseng Yuan-Yung(1903-1994), 中国数学家, 广义逆的奠基人.

[97]Sir Roger Penrose(1931-), 著名英国数学家, 物理学家, 哲学家. 2020 年诺贝尔物理学奖得主. 与 Stephen Hawking (霍金)合作证明了广义相对论的奇点存在性, 对黑洞理论有重要贡献.

显然, 若A为非奇异矩阵, 则A^{-1}是A的一个Penrose广义逆. 任意$m\times n$阶零矩阵0的一个Penrose广义逆是$n\times m$阶0矩阵.

定义 6.1.2. 设矩阵$A\in\mathbb{C}^{m\times n}$. 若矩阵$X\in\mathbb{C}^{n\times m}$满足Moore方程组(6.1.1), 则称$X$为$A$的一个Moore广义逆矩阵.

单位矩阵I_n是整个空间$\mathbb{C}^n$上的投影变换, 故当A可逆时, Moore广义逆矩阵与Penrose广义逆矩阵是相同的. 为了一般地证明这个结论, 我们先讨论Penrose广义逆矩阵的基本性质.

定理 6.1.1. 对任意$A\in\mathbb{C}^{m\times n}$, A的Penrose广义逆存在并且唯一, 记为$A^\dagger$.

证 若A为零矩阵, 可取X也为零矩阵. 现设$A\neq 0$. 则A有奇异值分解:

$$A=U\begin{pmatrix}\sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0\end{pmatrix}_{m\times n}V^*,\tag{6.1.3}$$

其中U, V分别为n阶和m阶酉矩阵, r为A的秩. 令

$$X=V\begin{pmatrix}\sigma_1^{-1} & & & & & \\ & \ddots & & & & \\ & & \sigma_r^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0\end{pmatrix}_{n\times m}U^*,\tag{6.1.4}$$

则X显然满足Penrose方程. 所以$A^\dagger$总是存在的.

设X与Y均是A的Penrose广义逆, 则对X与Y重复利用Penrose方程可得

$$X=XAX=XX^*A^*=XX^*A^*Y^*A^*=XAY=XAA^*Y^*Y=A^*Y^*Y=YAY=Y.$$

因此Penrose广义逆是唯一的. □

注. Penrose广义逆$A^\dagger$是Penrose的原始记号, 也是国际通行的记号, 国内常将$A^\dagger$记为A^+并称为“加号逆”.

例 6.1.1. 任意非零向量x的Penrose广义逆为$\dfrac{x^*}{x^*x}$. 单位向量x的Penrose广义逆为x^*.

例6.1.2. 矩阵$A=\begin{pmatrix}1&0\\0&0\end{pmatrix}$的Penrose广义逆为自身. 矩阵$B=\begin{pmatrix}0&1&0\\0&0&0\end{pmatrix}$的Penrose广义逆为$B^T$.

例 6.1.3. n阶幂零Jordan块J_n的Penrose广义逆为J_n^T.

容易证明Penrose广义逆具有下述性质, 证明留作习题.

命题 6.1.1. 矩阵$A^\dagger A$与$AA^\dagger$均为投影矩阵, 且

$$R(A^\dagger) = R(A^*), \qquad N(A^\dagger) = N(A^*).$$

定理 6.1.2. 任意矩阵A的Moore广义逆矩阵与Penrose广义逆矩阵相等. 因此Moore广义逆矩阵的定义与Penrose广义逆矩阵的定义等价. 通常将$A^\dagger$称为Moore-Penrose广义逆或**伪逆**.

证 若X是A的Moore广义逆矩阵, 则AX显然满足(6.1.2), 故X是A的Penrose广义逆矩阵.

反之, 若X是A的Penrose广义逆矩阵. Penrose方程(1)与(3)表明AX是幂等的Hermite矩阵, 故是列空间$R(A)$的投影矩阵. 同理, Penrose方程(2)与(4)表示XA是幂等的Hermite矩阵, 因此XA是列空间$R(X)$的投影矩阵, 从而X满足Moore方程组(6.1.1). □

注1. 用广义逆矩阵, 矩阵A的列空间上的投影矩阵恰好是$AA^\dagger$, 任意向量β在矩阵A的列空间$R(A)$上的投影向量$\mathrm{Proj}_{R(A)}b$恰好是$AA^\dagger\beta$.

注2. 普通逆矩阵是Moore-Penrose广义逆矩阵的一种特例, 故Moore-Penrose广义逆矩阵并不具备普通逆矩阵的一些性质, 如下例.

例 6.1.4. 设$A = (1,0), B = (1,1)^T$, 则$(AB)^\dagger = 1$而$B^\dagger A^\dagger = 1/2$, 因此$(AB)^\dagger \neq B^\dagger A^\dagger$.

为了进一步讨论$A^\dagger$的性质, 我们引入以下符号

$$\lambda^\dagger = \begin{cases} \lambda^{-1}, & \lambda \neq 0 \\ 0, & \lambda = 0. \end{cases} \tag{6.1.5}$$

下面的命题汇总了Moore-Penrose广义逆矩阵$A^\dagger$的一些性质, 证明均较为直接, 见习题12. 请读者比较这些性质与普通逆矩阵的类似性质.

命题 6.1.2. 对任意矩阵A, 有

(1) $(A^\dagger)^\dagger = A$.

(2) $(A^*)^\dagger = (A^\dagger)^*$; $(A^T)^\dagger = (A^\dagger)^T$.

(3) $(\lambda A)^\dagger = \lambda^\dagger A^\dagger$.

(4) $\mathrm{diag}\,(\lambda_1, \cdots, \lambda_n)^\dagger = \mathrm{diag}\,(\lambda_1^\dagger, \cdots, \lambda_n^\dagger)$.

(5) $A^\dagger A A^* = A^*$, $A^* A A^\dagger = A^*$; 如果A还是方阵, 则还有$AA^\dagger A^* = A^*$, $A^* A^\dagger A = A^*$.

(6) $(A^*A)^\dagger = A^\dagger (A^\dagger)^*$.

(7) $A^\dagger = (A^*A)^\dagger A^*$.

(8) 设$A = B + C, B^*C = BC^* = 0$, 则$A^\dagger = B^\dagger + C^\dagger$.

(9) $r(A) = r(A^\dagger) = r(A^\dagger A) = \mathrm{tr}(A^\dagger A)$.

按照矩阵与线性变换的对应关系, 我们可以讨论线性变换的Moore-Penrose广义逆变换, 此只需将Penrose方程组与Moore方程组中的矩阵理解成线性变换即可. 于是$0 \in \mathrm{Hom}(U,V)$的Moore-Penrose广义逆变换为$0 \in \mathrm{Hom}(V,U)$.

例 6.1.5. 平面$\mathbb{R}^2$上的移位变换$\sigma : (x,y)^T \mapsto (y,0)^T$是不可逆变换, 其Moore-Penrose广义逆变换为$\sigma^\dagger : (x,y)^T \mapsto (0,x)^T$, 这是因为$\sigma\sigma^\dagger : (x,y)^T \mapsto (x,0)^T$正是$\mathrm{Im}(\sigma)$上的正交投影变换, 而$\sigma^\dagger\sigma : (x,y)^T \mapsto (0,y)^T$正是$\mathrm{Im}(\sigma^\dagger)$上的正交投影变换.

例 6.1.6. 设$V=\mathbb{F}[x]_n$, 考虑V上的求导变换$\partial: f(x)\mapsto f'(x)$, 我们知道$\partial$不是可逆线性变换, 但只要定义$V$上的内积, 则其Moore-Penrose广义逆变换就唯一地存在. 比如, 设$1,x,x^2,\cdots,x^{n-1}$是V的一组标准正交基, 则∂的Moore-Penrose广义逆变换为(见习题13)

$$\partial^\dagger(x^{i-1})=\begin{cases} x^i/i, & 1\le i\le n-1\\ 0, & i=n.\end{cases}$$

思考题

1. 2×1矩阵与1×2矩阵的广义逆矩阵的几何意义是什么?
2. 两个n阶矩阵A与B何时满足条件$AB=BA=0$?
3. 设P,Q是两个可逆矩阵, 等式$(PAQ)^\dagger=Q^{-1}A^\dagger P^{-1}$成立吗?

6.1.2 Moore-Penrose 广义逆的计算

本节我们讨论Moore-Penrose广义逆矩阵$A^\dagger$的计算. 首先, 由定理6.1.1的证明可知(见公式(6.1.3)与(6.1.4)), 利用奇异值分解可以计算$A^\dagger$.

定理 6.1.3. ($A^\dagger$的SVD算法) 设A的奇异值分解为$A=UDV^*$, 则$A^\dagger=VD^\dagger U^*$.

例 6.1.7. 用奇异值分解求$A^\dagger$, 其中

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

解 A的奇异值分解为

$$A=UDV^*=\begin{pmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}&0\\ \frac{1}{\sqrt2}&-\frac{1}{\sqrt2}&0\\0&0&1\end{pmatrix}\begin{pmatrix}2&0\\0&0\\0&0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\ \frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\end{pmatrix},$$

所以

$$A^\dagger=\begin{pmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\\ \frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\end{pmatrix}\begin{pmatrix}\frac12&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt2}&\frac{1}{\sqrt2}&0\\ \frac{1}{\sqrt2}&-\frac{1}{\sqrt2}&0\\0&0&1\end{pmatrix}=\frac14\begin{pmatrix}1&1&0\\1&1&0\end{pmatrix}.$$

利用矩阵的满秩分解也可以计算$A^\dagger$, 即有下述定理, 证明留作习题.

定理 6.1.4. ($A^\dagger$的满秩算法) (1) 设A为列满秩矩阵, 则$A^\dagger=(A^*A)^{-1}A^*$;
(2) 设A为行满秩矩阵, 则$A^\dagger=A^*(AA^*)^{-1}$;
(3) 设$A=LR\in\mathbb{C}^{m\times n}$的秩为$r$, 其中$L$为列满秩矩阵, R为行满秩矩阵. 则

$$A^\dagger=R^\dagger L^\dagger=R^*(RR^*)^{-1}(L^*L)^{-1}L^* \tag{6.1.6}$$

例 6.1.8. 设$A=\alpha\beta^*$是秩为1的矩阵, 则

$$A^\dagger=\frac{A^*}{\|\alpha\|_2^2\,\|\beta\|_2^2} \tag{6.1.7}$$

重新考察例 6.1.7, 因为$A=(1,1,0)^T(1,1)$, 由公式(6.1.7)可得$A^\dagger=(1/4)A^T$.

例 6.1.9. 求$A^\dagger$, 其中

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

解 A的满秩分解为

$$A=LR=\begin{pmatrix}1&0\\2&1\\0&1\end{pmatrix}\begin{pmatrix}1&0&1\\0&1&1\end{pmatrix},$$

则

$$RR^*=\begin{pmatrix}1&0&1\\0&1&1\end{pmatrix}\begin{pmatrix}1&0\\0&1\\1&1\end{pmatrix}=\begin{pmatrix}2&1\\1&2\end{pmatrix},$$

$$(RR^*)^{-1}=\frac{1}{3}\begin{pmatrix}2&-1\\-1&2\end{pmatrix},$$

$$L^*L=\begin{pmatrix}1&2&0\\0&1&1\end{pmatrix}\begin{pmatrix}1&0\\2&1\\0&1\end{pmatrix}=\begin{pmatrix}5&2\\2&2\end{pmatrix},$$

$$(L^*L)^{-1}=\frac{1}{6}\begin{pmatrix}2&-2\\-2&5\end{pmatrix}.$$

所以

$$\begin{aligned}A^\dagger&=R^*(RR^*)^{-1}(L^*L)^{-1}L^*\\&=\frac{1}{18}\begin{pmatrix}1&0\\0&1\\1&1\end{pmatrix}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}\begin{pmatrix}2&-2\\-2&5\end{pmatrix}\begin{pmatrix}1&2&0\\0&1&1\end{pmatrix}\\&=\frac{1}{6}\begin{pmatrix}2&1&-3\\-2&0&4\\0&1&1\end{pmatrix}.\end{aligned}$$

如果矩阵A是列满秩或行满秩矩阵, 则$A^\dagger$也可由A的正交三角分解求出.

定理 6.1.5. ($A^\dagger$的QR算法) 设列满秩矩阵A有正交三角分解$A=QR$, 其中Q的列向量为单位正交向量组, R为非奇异的上三角矩阵. 则

$$A^\dagger=R^{-1}Q^*. \tag{6.1.8}$$

由命题6.1.2(7)可知, $A^\dagger$可由$(A^*A)^\dagger$与A^*的乘积得到. 因此我们讨论一般Hermite矩阵的Moore-Penrose广义逆. 设A的互不相同的非零奇异值为$\sigma_1,\cdots,\sigma_s,s\geq 1,\sigma_{s+1}=0$, 则Hermite矩阵$A^*A$的谱分解为

$$A^*A=\sigma_1^2P_1+\cdots+\sigma_s^2P_s,$$

由正规矩阵的谱分解定理以及命题6.1.2(8)可得

$$(A^*A)^\dagger = \sigma_1^{-2}P_1 + \cdots + \sigma_s^{-2}P_s \tag{6.1.9}$$

设

$$\phi(x) = \prod_{i=1}^{s+1}(x - \sigma_i^2), \qquad \phi_i(x) = \prod_{j\neq i}(x - \sigma_j^2) \tag{6.1.10}$$

则由于$P_i^* = P_i, P_iP_j = \delta_{ij}P_i$以及$\sum\limits_{i=1}^{s+1} P_i = I$, 可得(细节见习题17)

$$\phi_j(A^*A) = \phi_j(\sigma_j^2)P_j, 1 \le j \le s \tag{6.1.11}$$

因此

$$P_j = \frac{\phi_j(A^*A)}{\phi_j(\sigma_j^2)}, 1 \le j \le s \tag{6.1.12}$$

将上式代入公式(6.1.9)可得下述计算$A^\dagger$的**Lagrange-Sylvester 展开公式**.

定理 6.1.6. (Lagrange-Sylvester 展开公式)

$$A^\dagger = \sum_{i=1}^{s}\sigma_i^{-2}\frac{\prod\limits_{j\neq i}(A^*A - \sigma_j^2 I)}{\prod\limits_{j\neq i}(\sigma_i^2 - \sigma_j^2)}A^* \tag{6.1.13}$$

例 6.1.10. 设$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, 求$A^\dagger$.

解 由第四章例4.1.11可知

$$A^TA = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}, |\lambda I - A^TA| = \lambda(\lambda-1)(\lambda-3).$$

故A的非零奇异值为$\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$. 故由公式(6.1.13)可得

$$\begin{aligned} A^\dagger &= [\tfrac{1}{3}(\tfrac{A^TA-I}{3-1}) + \tfrac{A^TA-3I}{1-3}]A^T \\ &= (\tfrac{4}{3}I - \tfrac{1}{3}A^TA)A^T = \tfrac{1}{3}\begin{pmatrix} 2 & -1 \\ 1 & 1 \\ -1 & 2 \end{pmatrix}. \end{aligned}$$

注. 实际计算例6.1.10中的$A^\dagger$时, 利用定理6.1.4(2)较为简捷.

思考题

1. 公式(6.1.6)的几何意义是什么?
2. 列满秩矩阵与行满秩矩阵的Moore-Penrose 广义逆的几何意义是什么?
3. 利用谱分解计算 Moore-Penrose 广义逆的几何意义是什么?

第二节 其它各类广义逆

通常, 将满足 Penrose 方程组中等式$i_1, \cdots, i_j$的矩阵X称为矩阵A的$\{i_1, \cdots, i_j\}$-逆, 比如满足第一及第三个等式的矩阵X称为A的$\{1,3\}$-逆, 记为$A^{(1,3)}$; 而矩阵A的 **Moore-Penrose 广义逆**$A^{\dagger} = A^{(1,2,3,4)}$. 因此, 一个矩阵$A$共有(至少) 15 种广义逆, 本节讨论$\{1\}$-逆 $A^{(1)}$与$\{1,k\}$-逆 $A^{(1,k)}, 2 \le k \le 4$.

6.2.1 $\{1\}$-逆

Moore 方程组实际上是用特殊的线性变换来定义矩阵的广义逆. 如果从线性方程组的角度讨论矩阵的广义逆, 则可得到下面的定义.

定义 6.2.1. 设A是$m \times n$矩阵. 一个$n \times m$矩阵G称为A的一个 **{1}-广义逆矩阵** , 若对任意给定的m维向量b, 只要方程组$Ax = b$有解, 则$x = Gb$也一定是解.

注. 国内常将 **{1}-广义逆矩阵**称为“减号”逆, 记为A^-.

例 6.2.1. 设$A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, 则$\begin{pmatrix} 1 \\ 0 \end{pmatrix}^{\dagger} = (1, 0)$. 由定义 6.2.1, A^-应满足条件$AA^-b = b$, 于是可取$A^- = (1, y)$, 其中y是任意常数.

下面的定理用 Penrose 方程组中的一个方程来刻画 $\{1\}$-广义逆矩阵.

定理 6.2.1. $n \times m$矩阵G是$m \times n$矩阵A的一个 $\{1\}$-广义逆矩阵 $\Longleftrightarrow \quad AGA = A$.

证 “$\Rightarrow$”: 对任意的$z \in \mathbb{C}^n$, $b = Az$是m维向量, 且z为$Ax = b$的解. 因而$x = Gb$也是一个解, 即$AGb = b$. 于是

$$AGAz = Az.$$

由于z的任意性, 可知$AGA = A$.

“$\Leftarrow$”: 若$AGA = A$, 设$Ax = b$有解, 则

$$AGb = AGAx = Ax = b.$$

所以$x = Gb$也是解. 由定义, G是A的一个 $\{1\}$-广义逆矩阵. □

显然$A^{\dagger}$是A的一个 $\{1\}$-广义逆矩阵. 由例 6.2.1可知, 一般情况下, A^-不唯一. 因此我们用符号$A\{1\}$表示矩阵A的所有 $\{1\}$-广义逆矩阵, 即

$$A\{1\} = \{X \mid AXA = A\} \tag{6.2.1}$$

例如, $0_{m \times n}\{1\} = \mathbb{C}^{n \times m}$.

例 6.2.2. 若A为可逆矩阵, 则方程$AXA = A$只有唯一解A^{-1}, 故此时A的 $\{1\}$-广义逆矩阵唯一, 即$A\{1\} = \{A^{-1} = A^{\dagger}\}$.

例 6.2.3. 由定理 6.2.1可知, 任何一个$n \times m$阶矩阵都是$0_{m \times n}$的一个 $\{1\}$-广义逆矩阵! (请对照, $0^{\dagger}_{m \times n} = 0_{n \times m}$.) 故知 $\{1\}$-广义逆矩阵不是对称的, 即若G是矩阵A的一个 $\{1\}$-广义逆矩阵, 则A未必是G的一个 $\{1\}$-广义逆矩阵. 因此公式$(A^-)^- = A$一般不成立.

例 6.2.4. 设$A \in \mathbb{C}^{m\times n}$. 矩阵方程$AX = I$的解称为矩阵$A$的**右逆** . 类似地, 矩阵方程$XA = I$的解称为矩阵$A$的**左逆**. 显然, 左逆与右逆可能均不存在, 也可能只存在一个. 但由定理6.2.1可知, 左逆与右逆均是A的 $\{1\}$-广义逆矩阵. 左逆与右逆的其它性质见习题19.

为了得到$\{1\}$-广义逆矩阵的一般形式, 需要$\{1\}$-广义逆矩阵的以下性质, 证明见习题20.

命题 6.2.1. 设P, Q为非奇异矩阵, 则$Q^{-1}A^{-}P^{-1} \in (PAQ)\{1\}$, 即$Q^{-1}A^{-}P^{-1}$是$PAQ$ 的一个$\{1\}$-广义逆矩阵.

命题6.2.1可以粗略地解释为$(PAQ)^{-} = Q^{-1}A^{-}P^{-1}$, 请注意此等式不清晰, 因为左右两端既可以表示一个矩阵, 也可以表示一个集合. 但这个记法确实简单且不致引起混乱, 所以常用A^{-}或$A^{(1)}, A^{(1,2)}$等表示一个$\{1\}$-逆或一个$\{1,2\}$-逆. 需要指出, 尽管命题6.2.1成立, 但正如$A^{\dagger}$的情形, $(AB)^{-} = B^{-}A^{-}$一般不成立, 见例6.1.4.

定理 6.2.2. 设$A \in \mathbb{C}^{m\times n}, P \in \mathbb{C}^{m\times m}$与$Q \in \mathbb{C}^{n\times n}$可逆且满足

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

则

$$M = Q\begin{pmatrix} I_r & X \\ Y & Z \end{pmatrix}P, \tag{6.2.2}$$

是A的一个$\{1\}$-广义逆矩阵, 且A的任意一个$\{1\}$-广义逆矩阵都可以写成(6.2.2)的形式, 其中$X \in \mathbb{C}^{r\times(m-r)}$, $Y \in \mathbb{C}^{(n-r)\times r}$, $Z \in \mathbb{C}^{(n-r)\times(m-r)}$是任意的.

证 直接计算可得

$$\begin{aligned} AMA &= P^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1}Q\begin{pmatrix} I_r & X \\ Y & Z \end{pmatrix}PP^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1} \\ &= P^{-1}\begin{pmatrix} I_r & X \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1} \\ &= P^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1} = A. \end{aligned}$$

所以M为A的一个$\{1\}$-广义逆矩阵. 反之, 设A^{-}为A的任意一个$\{1\}$-广义逆矩阵. 由于P, Q可逆, 因此可令

$$A^{-} = Q\begin{pmatrix} W & X \\ Y & Z \end{pmatrix}P.$$

由于$A = AA^{-}A$, 即

$$\begin{aligned} P^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1} &= P^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1}Q\begin{pmatrix} W & X \\ Y & Z \end{pmatrix}PP^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1} \\ &= P^{-1}\begin{pmatrix} W & X \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q^{-1} \\ &= P^{-1}\begin{pmatrix} W & 0 \\ 0 & 0 \end{pmatrix}Q^{-1}. \end{aligned}$$

因此$W=I_r$, 即A^-具有公式(6.2.2)的形式. $\square$

下面的命题罗列了 {1}-广义逆矩阵的一些基本性质, 证明均较简单, 见习题21.

命题 6.2.2. 若$A\in\mathbb{C}^{m\times n}$, 则
(1) $(A^-)^*\in A^*\{1\}$, 即$(A^-)^*$是A^*的一个 {1}-广义逆矩阵.
(2) $r(A)\le r(A^-)$.
(3) A可逆 $\Longleftrightarrow$ $A\{1\}$是一元集合, 即$A\{1\}=\{A^{-1}\}$.
(4) $\lambda^\dagger A^-\in(\lambda A)\{1\}$, 即$\lambda^\dagger A^-$是$\lambda A$的一个 {1}-广义逆矩阵.
(5) AA^-与A^-A都是幂等矩阵.
(6) $R(AA^-)=R(A),\ N(A^-A)=N(A)$.

注意, 命题6.2.2的(2)仅给出了矩阵A的 {1}-广义逆矩阵的秩的一个下限$r(A)$, 读者可尝试由定理6.2.2得出一个上限.

例 6.2.5. 试求表示成公式(6.2.2)形式的广义逆矩阵A^-, 其中

$$A=\begin{pmatrix}0&1\\-1&-4\\1&5\end{pmatrix}.$$

解

$$\begin{pmatrix}A&I_3\\I_2&0\end{pmatrix}=\begin{pmatrix}0&1&1&0&0\\-1&-4&0&1&0\\1&5&0&0&1\\1&0&0&0&0\\0&1&0&0&0\end{pmatrix}\to\begin{pmatrix}0&1&1&0&0\\-1&0&4&1&0\\1&0&-5&0&1\\1&0&0&0&0\\0&1&0&0&0\end{pmatrix}$$

$$\to\begin{pmatrix}0&1&1&0&0\\1&0&-4&-1&0\\0&0&-1&0&1\\1&0&0&0&0\\0&1&0&0&0\end{pmatrix}\to\begin{pmatrix}1&0&1&0&0\\0&1&-4&-1&0\\0&0&-1&0&1\\0&1&0&0&0\\1&0&0&0&0\end{pmatrix}$$

所以

$$P=\begin{pmatrix}1&0&0\\-4&-1&0\\-1&0&1\end{pmatrix},\qquad Q=\begin{pmatrix}0&1\\1&0\end{pmatrix}.$$

于是A的 {1}-广义逆矩阵为

$$A^-=Q\begin{pmatrix}I_2&X\end{pmatrix}P,$$

其中$X\in\mathbb{C}^{2\times1}$是任意矩阵.

以下我们从矩阵方程的观点讨论矩阵的 {1}-广义逆矩阵. 由定理6.2.1和公式(6.2.1), 矩阵A的全体 {1}-广义逆矩阵恰好是矩阵方程$AXA=A$的解集, 而由线性方程组的一般理论, 该方程的通解是相应的线性齐次矩阵方程$AXA=0$的通解与它自身的一个特解的和. 以下, 我们将矩阵方程$AXA=A$称为**对称线性矩阵方程**, 而$AXA=0$称为**对称线性齐次矩阵方程** .

定理 6.2.3. (对称线性齐次矩阵方程的解)设$A \in \mathbb{C}^{m\times n}, r(A) = r$. 则
(1) 齐次矩阵方程$AXA = 0$的解空间是$\mathbb{C}^{n\times m}$的一个$mn - r^2$维子空间.
(2) 齐次矩阵方程$AXA = 0$的通解为

$$X = Y - A^\dagger AYAA^\dagger, \forall Y \in \mathbb{C}^{n\times m} \tag{6.2.3}$$

证 (1) 考察$\mathbb{C}^{n\times m}$到$\mathbb{C}^{m\times n}$的如下线性变换

$$\sigma : X \mapsto AXA, \forall X \in \mathbb{C}^{n\times m}$$

则齐次方程$AXA = 0$的解空间恰好是$\mathrm{Ker}(\sigma)$. 根据矩阵的张量积的性质, σ在按列顺序的标准基下的矩阵为$A^T \otimes A$, 因此

$$\dim \mathrm{Ker}(\sigma) = \dim N(A^T \otimes A) = mn - r^2,$$

此处用到$r(A \otimes B) = r(A)r(B)$.

(2) 公式(6.2.3)中的矩阵显然是方程$AXA = 0$的解. 现设Y是方程$AXA = 0$的一个解. 则$AYA = 0$, 故$A^\dagger AYAA^\dagger = 0$. 于是$Y = Y - 0 = Y - A^\dagger AYAA^\dagger$即为公式(6.2.3) 中的形式, 所以公式(6.2.3)确是方程$AXA = 0$的通解. □

从定理 6.2.3的证明可知, 若A^-是A的任意一个 {1}-广义逆矩阵, 则$AXA = 0$的通解为

$$X = Y - A^- AYAA^-, \forall Y \in \mathbb{C}^{n\times m} \tag{6.2.4}$$

例 6.2.6. 设$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, 则$A^\dagger = (1/2)\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, 故$AXA = 0$的通解为

$$X = Y - A^\dagger AYAA^\dagger = \begin{pmatrix} a & b \\ -a & c \end{pmatrix},$$

其中a, b, c是任意常数. 因此$AXA = 0$的解空间是 3 维的.

由对称线性齐次矩阵方程的通解结构(公式(6.2.4)), 立即可得下述

定理 6.2.4. (Rao[98] 定理) 设$A \in \mathbb{C}^{m\times n}$, A^-是A的任意一个 {1}-广义逆矩阵, 则

$$A\{1\} = \{A^- + Y - A^- AYAA^- \mid \forall Y \in \mathbb{C}^{n\times m}\} \tag{6.2.5}$$

例 6.2.7. 设$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, 则由例 6.2.6及定理 6.2.4可知, 对称矩阵方程$AXA = A$的通解为

$$X = A^\dagger + \begin{pmatrix} x & y \\ -x & z \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} a & b \\ -a & c \end{pmatrix},$$

其中a, b, c是任意常数.

[98]Calyampudi Radhakrishna Rao(1920-2023), 印度著名统计学家, 被誉为印度历史上最著名的十位科学家之一, 第三世界科学院创始人之一, 现为美国 Pennsylvania 州立大学教授.

对称矩阵方程$AXA=A$还可以写成$A(XA-I_n)=0$与$(AX-I_m)A=0$两种等价形式. 分别令

$$Y=XA-I_n,\quad Z=AX-I_m,$$

可得两个齐次矩阵方程

$$AY=0,\qquad ZA=0.$$

由上述矩阵方程组可以得出对称线性矩阵方程$AXA=A$通解的如下形式, 证明留作习题.

定理 6.2.5. 设$A\in\mathbb{C}^{m\times n}$, $A^-\in\mathbb{C}^{n\times m}$为$A$的一个$\{1\}$-广义逆矩阵, 则方程$AXA=A$的通解为

$$X=A^-+Y(I_m-AA^-)+(I_n-A^-A)Z, \forall Y, Z\in\mathbb{C}^{n\times m} \tag{6.2.6}$$

下面我们介绍左逆与$\{1\}$-广义逆矩阵之间的一个关系.

定理 6.2.6. 设$A\in\mathbb{C}^{m\times n}$, 则
(1) A列满秩 $\iff$ $A^-A=I_n$; 即矩阵A的$\{1\}$-广义逆矩阵是左逆 $\iff$ A列满秩.
(2) A行满秩 $\iff$ $AA^-=I_m$; 即矩阵A的$\{1\}$-广义逆矩阵是右逆 $\iff$ A行满秩.

证 我们只证(1), (2)的证明类似.
"$\Rightarrow$": 设A列满秩, 则存在非奇异矩阵$P\in\mathbb{C}^{m\times m}$, $Q\in\mathbb{C}^{n\times n}$使

$$PAQ=\begin{pmatrix} I_n \\ 0 \end{pmatrix}.$$

由定理6.2.1, A^-具有形式$A^-=Q(I_n\quad X)P$. 所以

$$A^-A=Q(I_n\quad X)PP^{-1}\begin{pmatrix} I_n \\ 0 \end{pmatrix}Q^{-1}=QI_nQ^{-1}=I_n.$$

"$\Rightarrow$": 若$A^-A=I_n$, 则显然A是列满秩的. □

本节最后, 我们讨论一个重要的问题, 即A^-A与单位矩阵究竟有多大差别?

定理 6.2.7. 设$A\in\mathbb{C}^{m\times n}, r(A)=r$, A^-是A的一个$\{1\}$-广义逆矩阵, 则$r(I_n-A^-A)=n-r$.

证 设

$$PAQ=\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{m\times n},$$

其中P, Q分别为m阶和n阶非奇异矩阵, 则

$$A^-=Q\begin{pmatrix} I_r & X \\ Y & Z \end{pmatrix}_{n\times m}P.$$

所以

$$\begin{aligned} I_n-A^-A&=I_n-Q\begin{pmatrix} I_r & X \\ Y & Z \end{pmatrix}_{n\times m}PP^{-1}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{m\times n}Q^{-1} \\ &=I_n-Q\begin{pmatrix} I_r & 0 \\ Y & 0 \end{pmatrix}_{n\times n}Q^{-1}=Q\begin{pmatrix} 0 & 0 \\ -Y & I_{n-r} \end{pmatrix}Q^{-1}. \end{aligned}$$

因此$r(I_n - A^- A) = n - r$. $\square$

定理 6.2.7表明, A的秩越大, 则$A^- A$与单位矩阵的差距就越小, 特别, 当A可逆时, $A^- A$就是单位矩阵.

思考题

1. 零矩阵的 $\{1\}$-广义逆矩阵是所有矩阵, 是否还有别的矩阵的 $\{1\}$-广义逆矩阵是所有矩阵?
2. 不可逆的方阵可否有可逆的 $\{1\}$-广义逆矩阵?
3. $A^- A$与AA^-的几何意义是什么?
4. 试给出矩阵A的 $\{1\}$-广义逆矩阵的秩的一个上限?

6.2.2 $\{(1,k)\}$-逆, $k = 2, 3, 4$

定义 6.2.2. 如果矩阵X满足 Penrose 方程组的前两个方程, 即

$$AXA = A, \qquad XAX = X$$

则称X是矩阵A的一个 $\{1,2\}$**-逆**或**自反逆**.

例 6.2.8. 零矩阵的$\{1,2\}$-逆是唯一的, 就是其转置. 任意可逆矩阵的$\{1,2\}$-逆是唯一的, 就是其逆矩阵. 矩阵$(1,0)$的$\{1,2\}$-逆是$(1,\ x)^T$, 其中x为任意常数.

因此矩阵的$\{1,2\}$-逆一般不唯一, 我们用符号$A\{1,2\}$表示矩阵A的全体$\{1,2\}$-逆构成的集合. 显然,

$$A\{1,2\} \subseteq A\{1\}.$$

例 6.2.9. 设$A = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \end{pmatrix}$, 则下列矩阵

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, D = (1/6)A^T$$

依次是A的$\{1\}$-逆, $\{1,2\}$-逆和 Moore-Penrose 逆, 但B, C均不是A的 Moore-Penrose 逆, 而B也不是A的$\{1,2\}$-逆. 因此$A\{1,2\} \subseteq A\{1\}$通常是真包含.

下面的定理给出了任意矩阵在初等变换下的$\{1,2\}$-逆的一般形式, 证明见习题 23.

定理 6.2.8. 设$A \in \mathbb{C}^{m\times n}, r(A) = r$, $P \in \mathbb{C}^{m\times m}$, $Q \in \mathbb{C}^{n\times n}$是可逆矩阵使得

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

则

$$A\{1,2\} = \{Q\begin{pmatrix} I_r & B \\ C & CB \end{pmatrix}P \,|\, \forall B \in \mathbb{C}^{r\times(n-r)}, C \in \mathbb{C}^{(m-r)\times r}\} \tag{6.2.7}$$

命题 6.2.3. 设$A \in \mathbb{C}^{m\times n}$, 则

(1) $A\{1,2\} = \{X_1 A X_2 \,|\, X_1, X_2 \in A\{1\}\}$.

(2) $\forall X \in A\{1,2\}, r(X) = r(A)$.

(3) 设P, Q为适当阶数的可逆矩阵, 则$(PAQ)^{(1,2)} = Q^{-1}A^{(1,2)}P^{-1}$.

定义 6.2.3. 如果矩阵X满足 Penrose 方程组的第一及第三个方程, 即

$$AXA = A, \qquad (AX)^* = AX$$

则称X是矩阵A的一个 $\{1,3\}$**-逆**.

我们以符号$A\{1,3\}$表示矩阵A的全体$\{1,3\}$-逆构成的集合. 显然

$$A\{1,3\} \subseteq A\{1\}.$$

例 6.2.10. 任何矩阵$A \in \mathbb{C}^{n\times m}$都是$m\times n$阶零矩阵的$\{1,3\}$-逆. 任意可逆矩阵的$\{1,3\}$-逆是唯一的, 就是其逆矩阵. 矩阵$(1,0)$的$\{1,3\}$-逆是$(1,x)^T$, 其中$x$为任意常数, 即此时有$A\{1,3\} = A\{1,2\}$. 但矩阵$(1,0)^T$的$\{1,3\}$-逆是唯一的(为什么?), 等于$(1\ \ 0) = (1,0)^{T\dagger}$.

定理 6.2.9. 设B是$A \in \mathbb{C}^{m\times n}$的一个$\{1,3\}$-逆, 则

$$A\{1,3\} = \{B + (I_n - BA)Y \mid \forall Y \in \mathbb{C}^{n\times m}\} \tag{6.2.8}$$

证 由于$A\{1,3\} \subseteq A\{1\}$, 而由定理6.2.4(对称线性矩阵方程的解)我们知道$X = B + (I_n - BA)Y$满足方程$AXA = A$. 再由$A[B + (I_n - BA)Y] = AB$以及$B \in A\{1,3\}$可知AX还是 Hermite 的. □

例 6.2.11. 设$A = \alpha\beta^*$是秩为 1 的矩阵, 则

$$A\{1,3\} = \{\frac{A^*}{\alpha^*\alpha\,\beta^*\beta} + (I - \frac{\beta\beta^*}{\beta^*\beta})Y \mid \forall\, Y\} \tag{6.2.9}$$

下面的命题列出了$\{1,3\}$-逆的一些性质, 证明见习题 28.

命题 6.2.4. 设$A^{(1,3)}$是$A \in \mathbb{C}^{m\times n}$的任意一个$\{1,3\}$-逆, 则
(1) $A\{1,3\} = \{X \mid AX = AA^{(1,3)}\}$.
(2) $AA^{(1,3)} = AA^\dagger = P_{R(A)}$; 特别地, $AA^{(1,3)}$是幂等矩阵.
(3) $A^{(1,3)}A$是幂等矩阵.
(4) $I_m \otimes A^{(1,3)} \in (I_m \otimes A)\{1,3\}$.

定义 6.2.4. 如果矩阵X满足 Penrose 方程组的第一及第四个方程, 即

$$AXA = A, \qquad (XA)^* = XA$$

则称X是矩阵A的一个 $\{1,4\}$**-逆**.

类似于前面几种情形, 我们以符号$A\{1,4\}$表示矩阵A的全体$\{1,4\}$-逆构成的集合. 显然

$$A\{1,4\} \subseteq A\{1\}.$$

由定义可知, 矩阵A的$\{1,4\}$-逆与其$\{1,3\}$-逆的差别是前者满足方程$(XA)^* = XA$而后者满足方程$(AX)^* = AX$, 因此它们具有较为相近的性质. 我们仅将矩阵A的$\{1,4\}$-逆的有关结论列出, 所有的证明均见习题 30-32, 请读者参照$\{1,3\}$-逆的相关结论与证明.

例 6.2.12. 任何矩阵$A \in \mathbb{C}^{n\times m}$都是$m\times n$阶零矩阵的$\{1,4\}$-逆. 任意可逆矩阵的$\{1,4\}$-逆是唯一的, 就是其逆矩阵. 矩阵$(1,0)$的$\{1,4\}$-逆是唯一的(为什么?), 等于$(1,0)^T$, 即此时有$A^{(1,4)} = A^\dagger$. 矩阵$(1,0)^T$的$\{1,4\}$-逆是$(1,x)^T$, 其中$x$为任意常数, 即此时有$A\{1,4\} = A\{1,2\}$. 请读者比较例 6.2.10.

定理 6.2.10. 设B是$A \in \mathbb{C}^{m\times n}$的一个$\{1,4\}$-逆, 则

$$A\{1,4\} = \{B + Y(I_m - AB) \,|\, \forall Y \in \mathbb{C}^{n\times m}\} \tag{6.2.10}$$

例 6.2.13. 设$A = \alpha\beta^*$是秩为 1 的矩阵, 则

$$A\{1,4\} = \{\frac{A^*}{\alpha^*\alpha\,\beta^*\beta} + Y(I - \frac{\alpha\alpha^*}{\alpha^*\alpha}) \,|\, \forall Y\} \tag{6.2.11}$$

思考题

1. 除了零矩阵与可逆矩阵外, 是否还有别的矩阵的$\{1,2\}$-逆是唯一的?
2. Hermite 矩阵的$\{1,2\}$-逆一定是 Hermite 的吗?
3. 不可逆矩阵的$\{1,3\}$-逆与$\{1,4\}$-逆一定是不可逆的吗?
4. 矩阵的$\{1,2\}$-逆, $\{1,3\}$-逆, $\{1,4\}$-逆的几何意义是什么?
5. 何时$A\{1,i\} = A\{1,j\}, 1 \le i \ne j \le 4$?

第三节 非负矩阵与Perron-Frobenius定理

6.3.1 Perron-Frobenius定理

将单位矩阵作任意次行(或列)交换所得的矩阵称为**置换矩阵**. 如果矩阵A的所有元素均是正数, 则称A是**正矩阵**, 记作$A > 0$(注意此符号亦用作正定矩阵的表示). 设A是一个n阶非负矩阵(记作$A \ge 0$), 如果存在置换矩阵P使得$P^TAP = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$, 其中$B, D$均为方阵, 则称$A$是**可约矩阵**, 否则称$A$为**不可约(非负)矩阵**. 比如正矩阵都是不可约的, 矩阵$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$是可约的, 而$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$是不可约的. 对非负不可约矩阵有下述刻画:

定理 6.3.1. 设A是n阶非负矩阵, 则A是不可约矩阵 $\iff$ $(I+A)^{n-1} > 0$.

由此定理即可知矩阵$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$是不可约的. 但该刻画需要计算$n$阶矩阵的$n-1$次幂, 不尽人意. 一个比较好的充分必要条件如下

定理 6.3.2. 设$A = \begin{pmatrix} a & b^T \\ c & D \end{pmatrix}$是$n$阶非负矩阵, 其中$b, c \in \mathbb{R}^{n-1}$均非0, $D \in \mathbb{R}^{(n-1)\times(n-1)}$. 则$A$是不可约矩阵 $\iff$ $D + cb^T$不可约.

定理的证明请参考文献[99], 作者还指出相应于定理 6.3.2的算法复杂度约为$O(\frac{n^3}{3})$, 好于已知算法. 由定理 6.3.2立即可得下例的结论.

[99] 何秀丽, 永学荣, 矩阵不可约的一个新的充要条件及判定矩阵不可约性的一个新算法, 新疆大学学报(自然科学版), V. 15(2)1998, 15-17

例 6.3.1. 设A是非负三对角矩阵, 且三条对角元素均非0, 则A不可约.

定理 6.3.3. (**Perron**[100]**-Froubenius 定理**[101]) 设A是n阶不可约非负矩阵, 则
(1) A的谱半径$\rho(A)$是A的一个单特征值.
(2) A有一个属于特征值$\rho(A)$的正特征向量.
(3) A的其余特征值λ均无正特征向量.

证 为降低证明难度, 下面假定A是正矩阵.

先证A的谱半径为$\rho = \rho(A)$是A的特征值. 注意$\rho^{-1}A$的谱半径为1, 故不妨设A的谱半径为1. 因此存在$\lambda \in \sigma(A)$使得$|\lambda| = 1$.

设$0 \neq x = (x_i) \in \mathbb{C}^n$是$\lambda$的特征向量, 则$Ax = \lambda x$. 记$\alpha = (|x_i|) \in \mathbb{C}^n$, 则$\alpha \geq 0$. 逐个考察$A\alpha$的每一行

$$e_i^T A\alpha = \sum_{j=1}^n a_{ij}|x_j| \geq |\sum_{j=1}^n a_{ij}x_j| = |e_i^T Ax| = |\lambda x_i| = |x_i|, \quad 1 \leq i \leq n.$$

所以$A\alpha \geq \alpha$.

如果$A\alpha = \alpha$, 则1是A的特征值, 而α是属于1的非负特征向量. 且由于$\alpha \neq 0$而$A > 0$, 因此$A\alpha > 0$, 故$\alpha > 0$. 如此定理前两条除单特征值外均被证明.

如果$A\alpha \neq \alpha$, 则$\beta = A\alpha - \alpha \geq 0$, 故$A\beta > 0$, 即$A^2\alpha > A\alpha > 0$. 故存在$\varepsilon > 0$使得$A^2\alpha > (1+\varepsilon)A\alpha$. 于是对任意$k \geq 1$, 有

$$A^{k+1}\alpha > (1+\varepsilon)^k A\alpha.$$

因此$\|A^k\| \geq (1+\varepsilon)^k$, 即$(\|A^k\|)^{1/k} \geq 1+\varepsilon$,

根据Gelfand 谱半径定理(见第五章第三节定理5.3.4)可知

$$\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k} \geq 1+\varepsilon,$$

此与$\rho(A) = 1$矛盾. 因此必有$A\alpha = \alpha$, 即谱半径是A的特征值且特征向量α是正向量.

欲证谱半径1是单根, 首先易证1的几何重数是1, 这是因为如果1还有另一个与α线性无关的特征向量β, 则不妨设β是实特征向量(因为β的实部和虚部都是属于1的特征向量). 如此总存在正数$c > 0$使得$\alpha - c\beta \geq 0$且$\alpha - c\beta$至少有一个分量等于0. 但$\alpha - c\beta = A(\alpha - c\beta) > 0$, 矛盾! 故谱半径的几何重数必然是1.

由于A^T也是正矩阵, 故存在正向量$\gamma \in \mathbb{R}^n$使得$\gamma^T A = \gamma^T$. 注意$\gamma^\perp = \{x \in \mathbb{R}^n | \gamma^T x = 0\}$是$A$的$n-1$维不变子空间且$\alpha \notin \gamma^\perp$ (因两个皆正), 故$\mathbb{R}^n$有A的不变子空间的直和分解

$$\mathbb{R}^n = \gamma^\perp \oplus \mathbb{R}\alpha.$$

故存在可逆矩阵P使得

$$P^{-1}AP = \begin{pmatrix} B & 0 \\ 0^T & 1 \end{pmatrix}.$$

[100]Oskar Perron(1880-1975), 德国数学家, 数学上有著名的**Perron 悖论**, 即若假定有最大的自然数, 则该自然数必然等于1!(这说明假定存在性是很危险的.)

[101] Perron 于 1907 年对正矩阵, Frobenius 于 1912 年对非负矩阵证明了该定理.

现若1的代数重数大于1, 则1也是$n-1$阶矩阵B的特征值. 将B视为$\gamma^{\perp}$到自身的线性变换, 则存在$0\neq\delta\in\gamma^{\perp}$是$B$的属于特征值1的特征向量, 于是$\delta$与$\alpha$是$A$的两个线性无关的特征向量, 与$A$的几何重数为1矛盾!

反证(3)[102]. 设某个正特征向量$y=(y_i)>0$ 属于某特征值$\lambda\neq 1$. 由$Ay=\lambda y$ 可知λ 为正实数. 又$\lambda=|\lambda|\leq 1$, 故$0<\lambda<1$. 设$p=\max\limits_i\dfrac{|x_i|}{y_i}$, 则$py-\alpha\geq 0$. 由$A>0$得$0<A(py-\alpha)=\lambda py-\alpha$, 即$\lambda py_i>|x_i|,\forall i$. 故 $\max\limits_i\dfrac{|x_i|}{y_i}<\lambda p<p$. 矛盾! □

Perron-Froubenius 定理是非负矩阵理论中最精彩的部分, 证明虽然略显复杂, 但鉴于其巨大的理论价值、广泛应用和美学内涵, 非常值得认真领会、仔细品味.

Perron-Froubenius 定理中的谱半径$\rho(A)$称为A的 **Perron 根**, 而属于 Perron 根的正特征向量称为矩阵A的 **Perron 向量**. 由 Perron-Froubenius 定理, 不可约非负矩阵的 Perron 向量是该矩阵唯一的正特征向量.

6.3.2 应用：投入产出模型与商品定价理论

本小节介绍1973 年诺贝尔经济学奖主题, 即著名的**Leontief[103]投入产出模型**和**商品定价理论**. 所有结论均不加证明, 请读者查阅相关文献.

Leontief 投入产出模型

按Leontief理论, 一个国家的经济体系可分为n个生产商品(含服务) 的部门, 以及若干仅消费但不生产的部门. 设$x\in\mathbb{R}^n$是每一部门一年的产出, 称为**产出向量**; $d\in\mathbb{R}^n$是各种非生产部门所需求的商品, 称为**最终需求向量**.

显然x与d的元素均非负, 称为非负向量. 类似地可以定义**非负矩阵**. Leontief 投入产出模型(又称生产方程)为

$$\text{(Leontief 投入产出模型)}\qquad x=Cx+d \tag{6.3.1}$$

其中$C\in M_n(\mathbb{R})$称为**消耗矩阵**, Cx称为**中间需求**.

定理 6.3.4. (Leontief 投入产出定理)设消耗矩阵C和最终需求向量d均非负, C的各列之和均小于1. 则$I-C$可逆, 产出向量$x=(I-C)^{-1}d$非负且是方程(6.3.1)的唯一解.

由于C各列之和均小于1, 故$I-C$实际上是一个**M-矩阵** (即形如$sI-B$的矩阵, 其中B为非负矩阵而$s\geq\rho(B)$), 因此$(I-C)^{-1}$是非负矩阵. 实际上, $(I-C)^{-1}$的第j列表示当第j个部门的最终需求增加1单位时, 各部门需要增加的产出的数量.

Leontief 生产方程的对偶形式是**价格方程**

$$\text{(Leontief 价格方程)}\qquad p=C^Tp+v \tag{6.3.2}$$

其中p称为**价格向量**, 它表示各部门产出的单位价格, v称为**增值向量**, 它表示每单位产出的附加值. 经济中最重要的指标国内生产总值(GDP)可以表示为

$$GDP=p^Td=v^Tx. \tag{6.3.3}$$

[102]此条的证明由上海交通大学数学科学学院2023级本科生朱韬提供.

[103]Wassily Wassilyovich Leontief(1905-1999), 著名俄裔美籍经济学家, 1973 年获得诺贝尔经济学奖. 他的三个博士研究生 Paul Samuelson, Robert Solow 与 Vernon Lomax Smith 分别于 1970, 1987 和 2002 年获得诺贝尔经济学奖.

商品定价理论

理论经济学中的商品定价问题与Perron根与Perron向量密切相关.

考虑n个工厂的封闭经济系统$\mathcal{E}$(即自产自销, 自给自足的理想经济体系), 每个工厂生产一种商品. 设描述该经济系统的状态矩阵为$A=(a_{ij})_{n\times n}$, 其中a_{ij}表示第i个工厂购买第j个工厂商品的百分比. 显然有

$$a_{ij}\geq 0, \qquad \sum_{i=1}^{n} a_{ij}=1,\ j=1,2,\cdots,n. \tag{6.3.4}$$

因此A是一个非负矩阵且其每个列和均为1(这样的矩阵称为**随机矩阵**, 见第三章第五节). 故1是A的一个特征值, 且由盖尔圆盘定理可知A的其余特征值的模均小于1. 进一步, 可以假设A是不可约的. 现在的问题是: 有无可能制定一个"公平的"价格体系使得每个工厂均能收支平衡? 为解决此问题, 设P_i表示第i个工厂的总收入, 并记$P=(P_1,\cdots,P_n)^T$(向量P称为**定价向量** 或**定价结构**). 如果定价结构P使得每个工厂保持收支平衡, 则称P是一个**均衡定价结构** .

定理 6.3.5. (Leontief 定价定理[104]) 设A是封闭经济系统$\mathcal{E}$的状态矩阵, P是一个n元正向量. 则P是一个均衡定价结构 $\Longleftrightarrow$ P是A的一个Perron向量.

证 由于a_{ij}是第i个工厂购买第j个工厂商品的百分比, 故第i个工厂的总支出为$\sum\limits_{j=1}^{n} a_{ij}P_j$. 因此$P$是一个均衡定价结构 $\Longleftrightarrow$

$$P_i=\sum_{j=1}^{n} a_{ij}P_j, \quad i=1,2,\cdots,n. \tag{6.3.5}$$

即$AP=P$. □

当然, 一个封闭的经济系统是不可能持续发展的, 也就是说, 必须要出售部分商品给外部方能使系统"盈利". 设第i个工厂的盈利为$\gamma_i\geq 0$, 即有

$$\gamma_i=P_i-\sum_{j=1}^{n} a_{ij}P_j, \quad i=1,2,\cdots,n. \tag{6.3.6}$$

记$\gamma=(\gamma_1,\cdots,\gamma_n)^T$. 则$\gamma$是一个非负向量, 且$\gamma\neq 0$. 将公式(6.3.6)写成矩阵形式有

$$(I-A)P=\gamma. \tag{6.3.7}$$

使上述方程对任何非负向量γ都有解的一个充分必要条件是$\rho(A)<1$. 此时矩阵A的列和至少有一个小于1, 即至少有一个工厂向外部出售了商品.

现考虑一个简单情形, 即每个工厂的毛利率都相同(这看起来比较"公平"), 设为$r>0$. 于是$\gamma_i=rP_i$, 因此方程(6.3.7)变为

$$(I-A)P=rP. \tag{6.3.8}$$

即

$$AP=\lambda P, \quad 其中\ \lambda=1-r,$$

[104]该定理出自Leontief的著名论文"Input-Output Economics", Scientific Amrican,October 1951, pp. 15-21.

于是定价结构P仍是A的一个正特征向量, 因此由Perron-Froubenius定理可知P必然是A的一个Perron向量! 所以, 无论是一个均衡定价的封闭经济系统还是一个具有正盈利的开放经济系统, 其定价结构必然是其系统矩阵的一个正特征向量即Perron向量, 只不过后者的状态矩阵的谱半径即Perron根小于1.

第四节 广义逆矩阵的应用

本节介绍广义逆矩阵的若干应用.

6.4.1 线性方程组的解的符号表示

首先介绍广义逆矩阵表示线性方程组与矩阵方程$AXB=C$的解的理论.

回忆线性方程组$Ax=b$称为相容的如果其至少存在一个解, 称为不相容的或矛盾的, 如果其没有解. 相容方程组又分为确定方程组(即恰好有一组解)和超定方程组(即有无穷多组解). 对于超定方程组常常需要求出最小范数解, 而对于矛盾方程组则需要求出最小二乘解. 如果最小二乘解是无限的, 则还需要求出范数最小的最小二乘解.

我们先利用$\{1\}$-广义逆矩阵给出一般齐次线性方程组的通解.

定理 6.4.1. 设A^-为$m\times n$矩阵A的一个$\{1\}$-广义逆矩阵, 则$Ax=0$的通解为

$$x=(I_n-A^-A)z, \tag{6.4.1}$$

其中z是任意的n维列向量.

证 由$AA^-A=A$知$(I_n-A^-A)z$为$Ax=0$的解. 设A的秩为r, 则$Ax=0$的解空间为$n-r$维的. 而$L=\{(I_n-A^-A)z\,|\,z\text{任意}\}$是解空间的子空间, 且$L$就是矩阵$(I_n-A^-A)$的列空间, 故由定理6.2.7知其维数为$r(I_n-A^-A)=n-r$. 定理得证. □

若$Ax=b$是相容的, 则由1-逆的定义知A^-b是$Ax=b$的一个特解. 故有下面的

定理 6.4.2. 设A^-为$m\times n$矩阵A的一个$\{1\}$-广义逆矩阵, 则当方程组$Ax=b$有解时, 其通解可表示为

$$x=A^-b+(I_n-A^-A)z, \tag{6.4.2}$$

其中z是任意的n维列向量.

定义 6.4.1. 相容方程组$Ax=b$满足$\|x\|_2=\sqrt{x^*x}$取最小值的解x称为最小范数解.

显然, 为求最小范数解只需使解的范数平方x^*x最小即可.

定理 6.4.3. 设$Ax=b$为相容方程组.

(1) 设$G\in A\{1,4\}$, 方程$Ax=b$的通解 为

$$x=Gb+(I-GA)z, \forall\, z \tag{6.4.3}$$

(2) 设$G\in A\{1\}$, 则$x=Gb$是方程$Ax=b$的最小范数解(即$\|Gb\|_2\le\|x\|_2, \forall\, x, Ax=b$) $\Longleftrightarrow$ $G\in A\{1,4\}$. 因此, 矩阵的$\{1,4\}$-逆也称为**最小范数逆**.

证 显然$y=Gb$为方程组的解. 设x_0为$Ax=b$的一个特解. 由定理6.4.2, $Ax=b$的通解为

$$x=Gb+(I_n-A^-A)z.$$

我们有

$$\begin{aligned}\|x\|_2^2 = x^*x &= (A^-b+(I_n-A^-A)z)^*(A^-b+(I_n-A^-A)z)\\ &= (A^-b)^*A^-b+((I_n-A^-A)z)^*(I_n-A^-A)z\\ &\quad +(A^-b)^*(I_n-A^-A)z+z^*(I_n-A^-A)^*A^-b.\end{aligned}$$

因

$$\begin{aligned}(A^-b)^*(I_n-A^-A)z &= (A^-Ax_0)^*(I_n-A^-A)z = x_0^*(A^-A)^*(I_n-A^-A)z\\ &= x_0^*A^-A(I_n-A^-A)z = x_0^*(A^-A-A^-AA^-A)z=0.\\ z^*(I_n-A^-A)^*A^-b &= z^*(I_n-A^-A)A^-Ax_0 = z^*(A^-A-A^-AA^-A)x_0=0.\end{aligned}$$

所以

$$\|x\|_2^2 = (A^-b)^*A^-b+((I_n-A^-A)z)^*(I_n-A^-A)z \ge (A^-b)^*A^-b.$$

由此便知$y=A^-b$为方程组$Ax=b$的最小范数解. □

因A的 Moore-Penrose 逆$A^\dagger$一定是A的一个$\{1\}$-广义逆矩阵A^-, 所以若方程组$Ax=b$有解, 则$y=A^\dagger b$一定是方程组的最小范数解.

定理 6.4.4. 设$Ax=b$为一矛盾方程. 则$x=Gb$是方程$Ax=b$的最小二乘解 $\Longleftrightarrow G\in A\{1,3\}$. 特别地, 方程$Ax=b$的最小二乘解 为

$$x = Gb+(I-GA)y, \forall y \tag{6.4.4}$$

因此矩阵的$\{1,3\}$-逆也称为**最小二乘逆**.

证 由第二章的讨论我们知道, $Ax=b$的最小二乘解为方程组

$$A^*Ax=A^*b$$

的解. 因此需证明$A^*A(A^-b)=A^*b$. 设

$$PAQ=\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}_{m\times n},$$

其中P, Q分别为m阶和n阶非奇异矩阵. 则

$$\begin{aligned}A^- &= Q\begin{pmatrix} I_r & X\\ Y & Z\end{pmatrix}_{n\times m}P,\\ AA^- &= P^{-1}\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}_{m\times n}Q^{-1}Q\begin{pmatrix} I_r & X\\ Y & Z\end{pmatrix}_{n\times m}P = P^{-1}\begin{pmatrix} I_r & X\\ 0 & 0\end{pmatrix}_{m\times m}P.\end{aligned}$$

因$(AA^-)^*=AA^-$, 所以

$$P^*\begin{pmatrix} I_r & 0\\ X^* & 0\end{pmatrix}(P^{-1})^* = P^{-1}\begin{pmatrix} I_r & X\\ 0 & 0\end{pmatrix}P.$$

即

$$\begin{pmatrix} I_r & 0\\ X^* & 0\end{pmatrix}(P^{-1})^*P^{-1} = (P^{-1})^*P^{-1}\begin{pmatrix} I_r & X\\ 0 & 0\end{pmatrix}.$$

因此

$$\begin{aligned}A^*Ay=A^*AA^-b&=(Q^{-1})^*\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}_{n\times m}(P^{-1})^*P^{-1}\begin{pmatrix} I_r & X\\ 0 & 0\end{pmatrix}_{m\times m}Pb\\ &=(Q^{-1})^*\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}_{n\times m}\begin{pmatrix} I_r & 0\\ X^* & 0\end{pmatrix}_{m\times m}(P^{-1})^*P^{-1}Pb\\ &=(Q^{-1})^*\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}_{n\times m}(P^{-1})^*b=A^*b.\end{aligned}$$

由此推出$y=A^-b$为$Ax=b$的最小二乘解. □

推论 6.4.1. 设$A\in\mathbb{C}^{m\times n},G\in\mathbb{C}^{n\times m}$, 则$G$是矩阵$A$的一个$\{1,3\}$-逆 $\Longleftrightarrow$ G是矩阵方程$AX=I_m$的最小二乘解, 即

$$\|AG-I_m\|_F=\min_{X\in\mathbb{C}^{n\times m}}\|AX-I_m\|_F.$$

证 由矩阵的张量积的性质可知, 矩阵方程$AX=I_m$可化为

$$(I_m\otimes A)\mathrm{vec}(X)=\mathrm{vec}(I_m) \tag{6.4.5}$$

由于矩阵X的F-范数显然等于其列展开$\mathrm{vec}(X)$的l_2范数, 故X是矩阵方程$AX=I_m$的最小二乘解 $\Longleftrightarrow$ $\mathrm{vec}(X)$是方程组(6.4.5)的最小二乘解.

由定理 6.4.4, $\mathrm{vec}(X)$是上述方程组的最小二乘解 $\Longleftrightarrow$

$$\mathrm{vec}(X)=(I_m\otimes A)^{(1,3)}\mathrm{vec}(I_m)+(I_{mn}-(I_m\otimes A)^{(1,3)}(I_m\otimes A))y,\forall\, y$$

因此

$$\mathrm{vec}(X)=(I_m\otimes A^{(1,3)})\mathrm{vec}(I_m)+(I_{mn}-(I_m\otimes A^{(1,3)}A)y,\forall\, y$$

此即

$$X=A^{(1,3)}+(I_n-A^{(1,3)}A)Y,\forall\, Y\in\mathbb{C}^{n\times m},$$

$\Longleftrightarrow$ X是A的$\{1,3\}$-逆. □

定理 6.4.5. 设$A\in\mathbb{C}^{m\times n}$, 且$Ax=b$为不相容方程组, 则$Ax=b$必有唯一的最小范数的最小二乘解, 这个解即为$y=A^\dagger b$.

证 由$A^\dagger$的定义及定理 6.4.4我们知道, $y=A^\dagger b$为$Ax=b$的一个最小二乘解. 设x为$Ax=b$的任意一个最小二乘解, 则x为$A^*Ax=A^*b$的解, 因而$x=A^\dagger b+z$, 其中z 为$A^*Ax=0$的解. 因$A^*Ax=0$与$Ax=0$的解空间是一样的, 所以z也是$Ax=0$的解, 即$Az=0$. 由于

$$x^*x=(A^\dagger b+z)^*(A^\dagger b+z)=(A^\dagger b)^*(A^\dagger b)+z^*z+z^*A^\dagger b+(A^\dagger b)^*z,$$

及

$$\begin{aligned}&z^*A^\dagger b=z^*A^\dagger AA^\dagger b=z^*(A^\dagger A)^*A^\dagger b=(A^\dagger Az)^*A^\dagger b=0,\\ &(A^\dagger b)^*z=(A^\dagger AA^\dagger b)^*z=(A^\dagger b)^*(A^\dagger A)^*z=(A^\dagger b)^*(A^\dagger Az)=0,\end{aligned}$$

所以

$$x^*x=(A^\dagger b)^*(A^\dagger b)+z^*z\geq(A^\dagger b)^*(A^\dagger b). \tag{6.4.6}$$

由此推出$y = A^\dagger b$为$Ax = b$的最小二乘解中范数最小的.

另外, 由式(6.4.6)知, $x^*x = (A^\dagger b)^*(A^\dagger b) \iff z = 0$, 即$x = A^\dagger b$, 所以最小范数的最小二乘解是唯一的. □

关于矩阵方程$AXB = C$的解, 有下述结论(证明留作习题)

定理 6.4.6. (Penrose 定理) 矩阵方程$AXB = C$有解 $\iff$ $AA^\dagger CB^\dagger B = C$; 且此时的全体解为$X = A^\dagger CB^\dagger + Y - A^\dagger AYBB^\dagger$.

由于向量是矩阵的特例, 本节关于线性方程组的几个结论实际上均是定理6.4.6的推论(Penrose 的原始论文正是这样做的).

6.4.2 网络流量矩阵估算

广义逆矩阵理论对求解线性方程组(无论是对未知向量还是对未知矩阵)非常便捷有效, 因此也可以用来研究线性模型的应用问题(可以有非线性的约束). 下面介绍广义逆矩阵在网络流量矩阵估算方面的一个应用.

设$\mathcal{N}$为一具有有限节点的网络系统. 一个**源节点**(origin node) 与一个**目的节点**(destination node)称为一个**OD 对**. 设$\mathcal{N}$共有n个 OD 对. 设第k个 OD 对的流量为x_k(设该 OD 对的源节点为i, 目的节点为j, 则x_k即为从i到j的流量). 向量$X = (x_1, \cdots, x_n)^T$称为$\mathcal{N}$的**流量矩阵**(实际上是一个向量). 设$\mathcal{N}$共有m个链接, 第i个链接的链接数为y_i. 向量$Y = (y_1, \cdots, y_m)^T$称为$\mathcal{N}$的**链接向量** 或链接负载. 令

$$a_{ij} = \begin{cases} 1, & \text{如果第 } j \text{ 个 OD 对的流量经过第 } i \text{ 个链接}, \\ 0, & \text{其它}. \end{cases}$$

则矩阵$A = (a_{ij})_{m\times n}$称为网络$\mathcal{N}$的**路由(信息)矩阵**. 路由矩阵, 流量矩阵和链接向量之间的关系如下:

$$AX = Y \tag{6.4.7}$$

通常网络$\mathcal{N}$的路由矩阵和链接向量较易得到, 因此主要的问题是估计流量矩阵X. 一般而言, 网络中的 OD 对数量n远大于链接总数m, 因此方程(6.4.7)是超定方程, 故有无穷多组解. 所以需要寻求满足一定条件的解$\hat{X}$. 比如可以使用方程(6.4.7)的最小范数解$\hat{X}$来估计网络的最小流量X, 此即求解约束问题

$$\begin{cases} \min X^T X \\ AX = Y \end{cases} \tag{6.4.8}$$

由定理 6.4.3(2), 约束问题(6.4.8)的解为$\hat{X} = A^\dagger Y$或$\hat{X} = A^{(1,4)}Y$.

思考题

1. 利用广义逆矩阵如何刻画方程组$Ax = b$的相容性?
2. 方程$Ax = b$的最小范数解是否唯一? 几何意义是什么?
3. 利用矩阵的张量积与广义逆(定理 6.4.6)求解矩阵方程$AXB = C$有何异同?

习 题 六

1. 设$\mathbb{R}^3$的子空间L由向量由$\alpha = (1,1,0)^T$ 和$\beta = (1,1,1)^T$生成, 求M上的投影矩阵P_M和向量$x = (2,3,1)^T$在M上的投影.

2. 设P_1, P_2均为投影矩阵, 证明:

(1) $P = P_1 + P_2$是投影矩阵 $\iff$ $P_1P_2 = P_2P_1 = 0$.

(2) $P = P_1 - P_2$是投影矩阵 $\iff$ $P_1P_2 = P_2P_1 = P_2$.
(3) 试将(1)推广至任意有限个投影矩阵的和.

3. 设$A \in \mathbb{C}^{n\times n}$是幂等矩阵. 证明下述条件等价:
(1) A是投影矩阵.
(2) $N(A) \perp R(A)$.
(3) $x^*A^*Ax \le x^*x, \forall x \in \mathbb{C}^n$.

4. 证明遍历定理(ergodic theorem)：设$A \in \mathbb{C}^{n\times n}$是酉矩阵, P是$N(I-A)$上的投影矩阵. 证明

$$\lim_{k\to\infty}\sum_{i=1}^{k} A^i x = Px, \forall x \in \mathbb{C}^n.$$

5. 证明$\begin{pmatrix} A \\ 0 \end{pmatrix}^{\dagger} = (A^{\dagger}, 0)$.

6. 证明命题6.1.1.

7. 设$A \in \mathbb{C}^{m\times n}$, 又$U \in \mathbb{C}^{m\times m}$和$V \in \mathbb{C}^{n\times n}$均为酉矩阵. 证明$(UAV)^{\dagger} = V^*A^{\dagger}U^*$.

8. 设H为幂等Hermite矩阵, 证明$H^{\dagger} = H$.

9. 证明$A^{\dagger} = A \iff A^2$为幂等Hermite矩阵且$r(A^2) = r(A)$.

10. 证明: 若A是正规矩阵, 则$A^{\dagger}A = AA^{\dagger}$, 且$(A^n)^{\dagger} = (A^{\dagger})^n$, 其中$n$为正整数.

11. 计算基本矩阵E_{ij}的Moore-Penrose广义逆和$\{1\}$-广义逆矩阵.

12. 证明命题6.1.2.

13. 如果$\mathbb{F}[x]_3$中的内积定义为$(f,g) = \int\limits_0^1 f(x)g(x)\,\mathrm{d}\,x$, 计算求导变换$\partial$的Moore-Penrose广义逆$\partial^{\dagger}$.

14. 设$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$.
(1) 当$b = (1,1,1,1)^T$时, 方程组$Ax = b$是否相容?
(2) 当$b = (1,0,1,0)^T$时, 方程组$Ax = b$是否相容?
若方程组相容, 求其通解和最小范数解; 若方程组不相容, 求其最小范数的最小二乘解.

15. (1) 设$r(BC) = r(B)$. 证明存在矩阵D使$B = BCD$, 且$C(BC)^-$是B的一个$\{1\}$-广义逆矩阵.
(2) 设$r(BC) = r(C)$. 证明存在矩阵D使$C = DBC$, 且$(BC)^-B$是C的一个$\{1\}$-广义逆矩阵.

16. 证明线性方程组$Ax = b$有解$\iff$ $AA^{\dagger}b = b$. 这里$A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$.

17. (1) 设A是$m\times n$矩阵, B是$m\times r$矩阵, 则等式$AA^-B = B \iff$存在矩阵D使$B = AD$;
(2) 设A是$m\times n$矩阵, B是$r\times m$矩阵, 则等式$BA^-A = B \iff$存在矩阵D使$B = DA$.

18. 证明定理6.1.4.

19. 详细证明定理6.1.6.

20. 计算下列矩阵的Moore-Penrose广义逆和$\{1\}$-广义逆矩阵, 并验证所得的结果.

$$(1)\begin{pmatrix} 1 & 0 & 2 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix};\quad (2)\begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 2 & 0 & 2 & 1 \end{pmatrix}.$$

21. 证明: (1) 如果矩阵A的左逆唯一, 则A必是可逆矩阵, 于是左逆等于右逆;
(2) 设矩阵A存在左逆但不唯一, 则A有无穷多个左逆. 类似地, 如果存在两个右逆, 则必存在无穷多个右逆.

22. 证明命题6.2.1.

23. 证明命题6.2.2.

24. 证明定理6.2.5.

25. 证明: $(AB)^\dagger = B^\dagger A^\dagger \iff A^\dagger ABB^*A^* = BB^*A^*$与$BB^\dagger A^*AB = A^*AB$同时成立.

26. 证明定理6.2.8.

27. 证明命题6.2.3.

28. 计算下列矩阵的$\{1,2\}$-逆:

$$(1)\quad A=\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix};\qquad (2)\quad A=\begin{pmatrix} \mathrm{i} & 0 \\ 1 & \mathrm{i} \\ 0 & 1 \end{pmatrix}.$$

29. 计算下列矩阵的$\{1,3\}$-逆:

$$(1)\quad A=\begin{pmatrix} 1 & -1 & 0 & -1 \\ 0 & 0 & 2 & 2 \\ -1 & 1 & 0 & 1 \end{pmatrix};\qquad (2)\quad A=\begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & -2 \\ -1 & 1 & 1 \\ 3 & -1 & -3 \end{pmatrix}.$$

30. 证明命题6.2.4.

31. 计算下列矩阵的$\{1,4\}$-逆:

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

32. (1) 哪些矩阵的$\{1,2\}$-逆等于它的转置矩阵?
(2) 哪些矩阵的$\{1,4\}$-逆等于它的转置矩阵?

33. 试求一个与书中公式形式不同的计算秩为1的矩阵的各种广义逆的公式.

34. 不可逆的方阵可否有可逆的$\{1,2\}$-逆或$\{1,3\}$-逆或$\{1,4\}$-逆?

35. 哪些不可逆的方阵有唯一的$\{1,2\}$-逆或$\{1,3\}$-逆或$\{1,4\}$-逆?

36. 是否存在矩阵其$\{1,2\}$-逆或$\{1,3\}$-逆或$\{1,4\}$-逆不唯一但只有有限个?

37. 设正规矩阵A仅有一个非零特征值λ.
(1) 证明$A^\dagger = \lambda^{-2}A$.
(2) 试求A的$\{1,2\}$-逆, $\{1,3\}$-逆及$\{1,4\}$-逆的表达式.
(3) 根据(1)与(2)计算矩阵$\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$的各种广义逆.

38. 设L, M是$\mathbb{C}^n$的子空间, P_X是子空间X上的投影矩阵. 证明:
(1) $P_{L+M} = (P_L+P_M)(P_L+P_M)^\dagger = (P_L+P_M)^\dagger(P_L+P_M)$.
(2) $P_{L\cap M} = 2P_L(P_L+P_M)^\dagger P_M = 2P_M(P_L+P_M)^\dagger P_L$.

39. 证明: $A^\dagger = A^{(1,4)}AA^{(1,3)}$.

40. 取A_1, A_2分别为第18题的(1)和(2), 并设$b_1=(1,1,0,1)^T$, $b_2=(1,1,2)^T$. 分别求出方程组$A_1x=b_1$和$A_2x=b_2$的通解.

41. 设$A=\begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 0 \end{pmatrix}$, $b=\begin{pmatrix} 2 \\ -1 \end{pmatrix}$. 求$Ax=b$的最小范数解.

42. 已知$A=\begin{pmatrix} 1 & 2 & -1 \\ -1 & -2 & 1 \end{pmatrix}$, $b=\begin{pmatrix} 2 \\ 0 \end{pmatrix}$. 求矛盾方程组$Ax=b$的最小二乘解.

43. 设$A\in\mathbb{C}^{m\times n},G\in\mathbb{C}^{n\times m}$, 则$G$是矩阵$A$的一个$\{1,4\}$-逆 $\Longleftrightarrow$ G是矩阵方程$XA=I_n$的最小二乘解, 即

$$\|GA-I_n\|_F=\min_{X\in\mathbb{C}^{n\times m}}\|XA-I_n\|_F.$$

44. 确定矩阵方程矩阵方程$AXB=0$的通解, 并以此证明定理6.4.6.

45. 判断矩阵方程$AXB=C$是否有解, 有解时求其解, 其中

(1) $A=\begin{pmatrix}1&1\\1&0\\0&1\end{pmatrix}$; $B=\begin{pmatrix}1&2&-1\\-1&1&1\end{pmatrix}$; $C=\begin{pmatrix}1&-1&1\\2&-2&2\\-1&1&-1\end{pmatrix}$;

(2) $A=\begin{pmatrix}1&-1&0&-1\\0&0&2&2\\-1&1&0&1\end{pmatrix}$; $B=\begin{pmatrix}1&0&2\\1&0&2\\1&0&2\end{pmatrix}$; $C=\begin{pmatrix}1&0&1\\1&0&1\\0&2&1\end{pmatrix}$.

46. 相容方程组$Ax=a$的通解$x=A^\dagger a+(I-A^\dagger A)y(\forall y)$还可以表示为$A^\dagger a+N(A)$的陪集形式. 证明:
(1) 这个表示是**正交表示**, 即向量$A^\dagger b$与向量$(I-A^\dagger A)y$正交, $\forall y$.
(2) 方程组$Ax=a$与$Bx=b$有公共解 $\Longleftrightarrow$ $A^\dagger a-B^\dagger b\in N(A)+N(B)$.
(3) 设方程组$Ax=a$与$Bx=b$有公共解. 试用陪集形式表示其解.

47. 设A,B,C,D均为n阶矩阵, 且矩阵方程$AX=B$与$XC=D$均有解. 证明:
(1) 两个方程有公共解 $\Longleftrightarrow$ $AD=BC$.
(2) 设两个方程有公共解. 试利用广义逆矩阵表示它们的公共通解.(提示: 可先研究齐次方程.)

48. 证明约束优化问题$\min\{x^Tx\},Ax=b$具有唯一解, 并求该解.

49. 证明约束优化问题$\min\{\mathrm{tr}(X^TX)-2\mathrm{tr}(X)\},XA=0$的解为$\hat{X}=I-AA^\dagger$.

50. 设U与W是线性空间V的两个子空间, $\alpha,\beta\in V$. 设$(\alpha+U)\cap(\beta+W)\neq\emptyset$. 证明:
(1) $(\alpha+U)\cap(\beta+W)=\alpha+P_U(P_U+P_W)^\dagger(\beta-\alpha)+(U\cap W)$.
(2) $(\alpha+U)\cap(\beta+W)=\alpha+(P_{U^\perp}+P_{W^\perp})^\dagger P_{W^\perp}(\beta-\alpha)+(U\cap W)$.
(3) $(\alpha+U)\cap(\beta+W)=\alpha+(I-P_WP_U)^\dagger P_{W^\perp}(\beta-\alpha)+(U\cap W)$.
(提示: 参考第二章习题75.)

51. 判断下面的矩阵是否为不可约矩阵?

$$(1)\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix},\qquad(2)\begin{pmatrix}0&1&0\\0&0&1\\1&0&1\end{pmatrix},\qquad(3)\begin{pmatrix}0&1&0\\1&1&1\\0&1&0\end{pmatrix}.$$

52. 证明定理6.3.1.

53. 设A是非负矩阵, 证明方程$(I-A)P=\gamma$对任何非负向量γ总有正向量解P $\Longleftrightarrow$ $\rho(A)<1$.(提示: 如果$(I-A)^{-1}\geq 0$, 则$\rho(A)<1$.)

54. 设P是非负矩阵, e是全1向量. P称为亚列随机矩阵(column sub-stochastic matrix)如果$e^TP\leq e^T$(按分量比较). 设v是非负向量, $0<\alpha<1$是参数. 线性方程组

$$(I-\alpha P)x=v \tag{6.4.9}$$

称为伪PageRank 方程组, 其解称为伪PageRank 向量, 其求解问题称为伪PageRank 问题.
(1) 证明伪PageRank 问题的解存在且唯一.
(2) 研究伪PageRank 问题与PageRank 问题的关系.

55. 证明定理6.3.4.

56. 证明(6.3.3)中的第二个等式.

57. 设(6.3.1)中的n阶消耗矩阵C的列和均小于1, 设Δx为满足最终需求Δd 的产出向量.

(1) 证明: 如果最终需求由d变为$d + \Delta d$, 则新的产出向量必须为$x + \Delta x$.

(2) 设$\Delta d = e_1$是第一个标准向量, 证明Δx恰好是矩阵$(I - C)^{-1}$的第一列. 解释这个结果.

研究性问题.

中国数学家对广义逆的研究历史悠久, 其中陈建龙及其团队的研究成果处于世界领先水平, 此处仅举一例:

定理. 设A, X 为复方阵, $k \geq 0$, 则下列条件等价:

(1) $XA^{k+1} = A^k$, **且**$rank(X) \leq rank(Y)$**, 其中**Y **为任意满足**$YA^{k+1} = A^k$ **的矩阵.**

(2) $XAX = X, R(X) = R(A^k)$**.**

(3) $XA^{k+1} = A^k, AX^2 = X$.

相关论文参见[105].

[105] C. Wu, J.L. Chen, Minimal rank weak Drazin inverses: a class of outer inverses with prescribed range. Electronic Journal of Linear Algebra, 2023, 39: 1-16.

附　　录

上海交通大学 2015-2016 学年第一学期《矩阵理论》试题

一. 单项选择题(每题 3 分, 共 15 分)

1. 设 $V=\mathbb{R}[x]_{2016}$ 是次数小于 2016 的实多项式构成的实线性空间. 设 $n\geq 0, f^{(n)}(x)$ 表示 $f(x)\in V$ 的 n 阶导数, $f^{(0)}(x)=f(x)$. 给定 V 的两个子空间 U,W 如下:

$$U=\{f(x)\in V\mid f^{(n)}(0)=0, n\leq 1949\},\ W=\{g(x)\in V\mid g(x)=x^{1896}(x-1)^{60}h(x), \forall h(x)\in V\}.$$

则 V 的子空间 $U+W$ 的维数 $\dim(U+W)=($　　$)$.

(A) 118　　(B) 119　　(C) 120　　(D) 121

2. 设 A 是 $m\times n$ 阶非零复矩阵, $R(A), N(A)$ 分别表示 A 的列空间与零空间. 设 $A=LR$ 是 A 的一个满秩分解. 考虑下述 8 个等式:

$$R(A)=R(L),\quad R(A^*)=R(L^*),\quad R(A)=R(R),\quad R(A^*)=R(R^*),$$
$$N(A)=N(L),\quad N(A^*)=N(L^*),\quad N(A)=N(R),\quad N(A^*)=N(R^*).$$

则上述等式恒成立的个数为(　　).

(A) 2　　(B) 4　　(C) 6　　(D) 8

3. 设两个 5 阶复矩阵 A 与 B 的最小多项式分别为 $x^3(x-1)$ 与 $x^2(x-1)^2$, 则矩阵 $\begin{pmatrix} 2A-B & B-A \\ 2A-2B & 2B-A \end{pmatrix}$ 的 Jordan 标准形所含 Jordan 块的个数为(　　).

(A) 5　　(B) 6　　(C) 7　　(D) 8

4. 设 A 为 n 阶正规矩阵, $\|\bullet\|_F$ 是矩阵的 F-范数, 则(　　).

(A) $\|A^2\|_F=\|A\|_F^2$　　(B) $\|A^2\|_F=\|A^*A\|_F$

(C) $\|A\|_F=\sup\limits_{x\neq 0}\dfrac{x^*Ax}{x^*x}$　　(D) $\|A\|_F^2=\sup\limits_{x\neq 0}\dfrac{x^*A^*Ax}{x^*x}$

5. 设 A 是 $m\times n$ 阶复矩阵, $A^\dagger$ 是 A 的 Moore-Penrose 广义逆. 考虑下述 4 个等式:

$$A^*AA^\dagger=A^*,\qquad A^\dagger AA^*=A^*,\qquad (A^*A)^\dagger A^*=A^\dagger,\qquad (A^*A)^\dagger A^*=A^*.$$

则上述等式恒成立的个数为(　　).

(A) 0　　(B) 1　　(C) 2　　(D) 3

二. 填空题(每题 3 分, 共 15 分)

6. 设 σ 是 $\mathbb{R}^2$ 上的线性变换, $e_1=(1,0,)^T, e_2=(0,1)^T, \sigma(e_1)=e_1, \sigma(e_1+e_2)=2e_1$, 则 σ 关于基 e_1+e_2, e_1-e_2 的矩阵为(　　　　).

7. 设 $e_1=(1,0,0)^T, e_2=(0,1,0)^T, A=(e_1,e_1)$. 则 $Ax=e_2$ 的最优解为(　　　　).

8. 设 $A=\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, 则 $\cos^2(At)-\sin^2(At)=($　　　　$)$.

9. 设 A 是秩为 2 的 3 阶投影矩阵, $B=\sum\limits_{n=0}^{\infty}\dfrac{(I-A)^n}{3^n}$, 则 e^{2B} 的Jordan标准型为(　　　　).

10. 设 $\alpha,\beta\in\mathbb{C}^n\ (n\geq 2)$, $\|\bullet\|_2$ 是向量的 2-范数(即欧几里德范数), $\|\alpha\|_2=2, \|\beta\|_2=1, \alpha^*\beta=1$. 则矩阵 $\alpha\beta^*+\beta\alpha^*$ 的 Moore-Penrose 广义逆为(　　　　).

三. 计算题与证明题 (11-14 题每题 15 分, 15题 10 分, 共 70 分)

11. 设

$$U=\{(x,y,z,w)^T\in\mathbb{R}^4\,|\,x+y+z+w=0\},\ W=\{(x,y,z,w)^T\in\mathbb{R}^4\,|\,x-y+z-w=0\}$$

是通常欧氏空间 $\mathbb{R}^4$ 的两个子空间. 设 I 是 $\mathbb{R}^4$ 上的恒等变换.

(1) 求 U 与 $U\bigcap W$ 的正交补 $(U\bigcap W)^\perp$ 的各一组标准正交基.

(2) 试求出 $\mathbb{R}^4$ 上的所有正交变换 σ 使得线性变换 $I-\sigma$ 的核 $Ker(I-\sigma)=U$.

12. 设 $n\geq 2$, $x=(x_1,x_2,\cdots,x_n)^T\in\mathbb{C}^n$. 定义线性变换 $\sigma:\mathbb{C}^n\to\mathbb{C}^n$ 如下:

$$\sigma(x)=(x_2,x_3,...,x_n,x_1)^T.$$

设 σ 在标准基 $e_1,e_2,...,e_n$ 下的矩阵为 A, 其中 $e_i\,(1\leq i\leq n)$ 为 n 阶单位矩阵的第 i 列.

(1) 求 A.

(2) 求 σ 的特征值与特征向量.

(3) 求 A 的逆矩阵 A^{-1} 的谱分解(请写出乘法形式与加法形式).

13. 设 $A=\begin{pmatrix}2&2&-1\\-1&-1&1\\-1&-2&2\end{pmatrix}$.

(1) 求 A 的 Jordan 标准形 J.

(2) 计算 e^{At}.

(3) 设 $x(0)=(1,0,0)^T$. 求定解问题 $x'(t)=Ax(t)$ 的解.

14. 已知 n 阶Hermite矩阵 A 的秩为 r, 其谱分解为 $A=UDU^*$, 其中 U 为酉矩阵, $D=\mathrm{diag}\,(a_1,\cdots,a_r,0,\cdots,0)$ 是对角矩阵. 记 I 为 n 阶单位矩阵.

(1) 求 A 的一个满秩分解.

(2) 判断矩阵 e^A 是否存在三角分解(即 LU 分解)? 请说明理由.

(3) 求分块矩阵 $M=\begin{pmatrix}A&A\\A&A\end{pmatrix}$ 的奇异值分解.

15. 设 A 为 n 阶复矩阵.

(1) 证明: 存在酉矩阵 U 和半正定矩阵 P, 使得$A=UP$. (此分解称为 A 的极分解.)

(2) 给出 U 与 P 唯一的充分必要条件.

主 要 参 考 书 目

[1] R. Bellman, Introduction To Matrix Analysis, Society For Industrial Mathematics , 1987.

[2] A. Blum, J. Hopcroft, R. Kannan, Foundations of Data Science, Cambridge University Press, 2020.

[3] 陈公宁, 矩阵理论与应用, 高等教育出版社, 1990年.

[4] 程云鹏, 张凯院, 徐仲, 矩阵论, 西北工业大学出版社, 1999年.

[5] 甘特马赫尔, 矩阵论, 柯召译, 高等教育出版社, 1955年.

[6] F. R. Gantmacher, The Theory of Matrices, vol. I. 英译者: K. A. Hirsch. Chelsea Publishing Company, New York 68, 1959.

[7] F. R. Gantmacher, Applications of the Theory of Matrices, 英译者: J.L. Brenner, W. Bushaw, S. Evanusa. New York, Interscience, 1959.

[8] 黄有度, 狄成恩, 朱士信, 矩阵论及其应用, 中国科学技术大学出版社, 1995年.

[9] R. A. Horn, C. R. Johnson, 矩阵分析, 杨奇译, 机械工业出版社, 2005年.

[10] 蒋启芬, 马俊, 线性代数, 机械工业出版社, 2020年.

[11] P. Lax, 线性代数及其应用, 傅莺莺, 沈复兴译, 人民邮电出版社, 2009年.

[12] D. C. Lay, 线性代数及其应用, 刘深泉, 洪毅, 马东魁, 郭国雄, 刘勇平译, 机械工业出版社, 2009年.

[13] G. Strang, 线性代数及其应用, 侯自新, 郑中三, 张廷伦译, 南开大学出版社, 1990年.

[14] J.H.威尔金森, 代数特征值问题, 石钟慈，邓健新译, 毛祖范校, 科学出版社, 2001年.

[15] 张贤达, 矩阵分析与应用, 清华大学出版社, 2005年.

[16] 张谋成, 黎稳, 非负矩阵论, 广东高等教育出版社, 1995年.

[17] 张跃辉, 矩阵理论与应用, 科学出版社, 2011年.

[18] 张跃辉, 李吉有, 朱佳俊, 数学的天空, 北京大学出版社, 2017年.

[19] 周杰, 矩阵分析及应用, 四川大学出版社, 2008年.

索引