线性代数学习心得

林开亮

2023 年 8 月 12 日

目录

序言

自 2010 年起，我在《数学传播》、《高等数学研究》等期刊发表了多篇关于线性代数的教学文章，今将其中10 篇汇集成册，分享给有兴趣的读者。清单如下：

[1] 林开亮，Hua 引理及其应用，《数学传播》，第 34 卷第 3 期（2010 年），39–48.

[2] 林开亮，Hurwitz–Radon 矩阵方程，《数学传播》，第 36 卷第 1 期（2012 年），48–63.

[3] 林开亮，线性代数珍宝十三则，《数学传播》，第 37 卷第 4 期（2013 年），65–83.

[4] 林开亮、陈见柯，Hurwitz 定理的矩阵证明，《高等数学研究》, 2018 年第 1 期，24–27.

[5] 林开亮，解常系数线性微分方程和递推关系的新方法——秦九韶和亥维赛的遗产，《数学传播》，第 43 卷第 2 期（2019 年），63–79.

[6] 林开亮、王兢，求解常系数线性微分方程的代数方法，《内蒙古师范大学学报(自然科学汉文版) 》，第 48 卷第 6 期（2019 年），577–582.

[7] 林开亮，王兢，论 Heaviside 算子法的合理性，《数学研究及评论》2020 年第 1 期，1-16 页

[8] 林开亮，吴艳霞，三维空间向量积与旋度算子的高维推广，《渭南师范学院学报》，2021 年第 2 期，74–79.

[9] 林开亮，赵教练，Hurwitz–Radon 矩阵方程的华罗庚链,《渭南师范学院学报》，2023 年第 2 期, 75–84.

[10] 安金鹏、林开亮、孙亦青，从 $(xI_n - A)$ 的列变换矩阵求 A 的标准型基底，《蛙鸣》第 66 期，2023 年 7 月, 5–14. 英文版见安金鹏老师的主页Construction of Bases for Rational and Jordan Canonical Forms.

除此以外，我还做了10 个线性代数教学的普及报告，清单如下：

[1]从射雕到九章，天津大学 (2017 年，刘云朋教授邀请)、上海交通大学 (2018 年，张跃辉、李吉有、朱佳俊教授邀请)、首都师范大学 (2019 年，刘海涛书记、戎小

春教授邀请)、河南大学 (2019 年，陈敏茹教授邀请)、西安电子科技大学 (2021 年，张哲教授邀请)、山东大学 (2022 年，文洁晶、许光午教授邀请)、宁德师范学院 (2022 年，蒋剑剑教授邀请)、湘潭大学 (2022 年，易年余教授邀请)。

[2]整数与多项式：平行的世界，江苏九章在线（2021 年，江苏省初中数学冬令营，宋书华老师邀请)、吉林师范大学（2022 年，刘鹏飞、程晓亮教授邀请）

[3]整数，2022 年，湘潭大学，易年余教授邀请

[4]多项式，2022 年，湘潭大学，易年余教授邀请

[5]方程术，宁德师范学院 (2022 年，蒋剑剑教授邀请)、湘潭大学 (2022 年，易年余教授邀请）

[6]微分算子与差分算子，2022 年，湘潭大学，易年余教授邀请

[7]向量空间的几何，2022 年，湘潭大学，易年余教授邀请

[8]Hurwitz–Radon 矩阵方程，2022 年，湘潭大学，易年余教授邀请

[9]线代观点看微分方程，宁德师范学院 (2022 年，蒋剑剑教授邀请)、内蒙古大学 (2023 年，颜昭雯教授邀请)

[10]高等代数与中国古代数学，2023 年，湖南大学，孙鹏、李军教授邀请

感谢我的合作者安金鹏、陈见柯、孙亦青、王兢、吴艳霞、赵教练，感谢邀请或推荐我做报告的诸位朋友。这个小册子是我们学术交流与友谊的结晶与见证。

感谢朱富海教授对初稿提出宝贵意见。

——开亮于西农

第一章　华引理及其应用

1.1　华引理的历史

在 1949 年发表的短文 [5] 中, 华罗庚解决了古典射影几何的一个难题. 在证明主要结果的过程中, 华先生用一个巧妙的推理得到一个有趣的命题, 该命题立即被 Jacobson 和 Rickart 在次年联合发表的文章 [8] 中加以引用并表述成以下形式:

引理 1 (华引理) 设 R, R' 是两个具有分配律的代数系 (未必可结合的环), J 是从 R 到 R' 的一个保持加法的映射, 并使得对任意的 $a, b \in R$, 有 $(ab)^J = a^J b^J$ 或 $(ab)^J = b^J a^J$, 则 J 是同态或反同态.

华罗庚在 [5] 中借助上述引理证明了下述定理:

定理 1 (除环的半自同构定理) 设 σ 是除环 K 到自身的一个满射, 满足

$$\sigma(a+b) = \sigma(a) + \sigma(b), \qquad \sigma(aba) = \sigma(a)\sigma(b)\sigma(a), \qquad \sigma(1) = 1.$$

则 σ 是 K 的自同构或反自同构.

1951 年, Jacobson 的 Lectures in Abstract Algebra 第一卷 [9] 出版, 将华罗庚的上述引理收入到环论一章的习题中.

1953 年, E.Artin 在对 Bourbaki 的代数学书评 [1] 中建议在新一版中收入华罗庚的下述漂亮定理到体论部分.

定理 2 设 σ 是除环 D 到 D' 之间的一个映射, 满足 $\sigma(a+b) = \sigma(a)+\sigma(b), \sigma(1) = 1'$, 而且对 $a \neq 0$ 有 $\sigma(a) \neq 0$ 且 $\sigma(a)^{-1} = \sigma(a^{-1})$, 则 σ 是同态或反同态.

Artin 紧接着指出了这个定理与华罗庚的结果之间的联系, 即这里有一个“有趣的 (amusing)”非交换恒等式:

命题 1 在一个除环中, 若 $a \neq 0$ 且 $a \neq b^{-1}$, 则

$$a - (a^{-1} + (b^{-1} - a)^{-1})^{-1} = aba.$$

上述恒等式一般称为华恒等式, 它甚至在 Jordan 代数中也成立. 应该公正地指出, 上述恒等式以及定理 2 是 Artin 首先指出的. 就笔者所知, 华罗庚已发表的论文中并没有出现这个恒等式 (虽然他擅长发现并应用一些恒等式, 如他对 Cartan-Brauer-Hua 定理以及华行列式不等式的证明).

未及 Bourbaki 代数学出版第二版, Artin 的几何代数 [2]1957 年问世, 他在这本书中详细地论证了定理 2 并且给出了它在射影几何上的应用. 这就是下述

定理 3 (一维射影几何的基本定理) 一条直线 K 到其自身的保持 (有限) 调和点列的双射 σ 具有下述形式

$$\sigma(x) = ax^{\tau} + b,$$

这里 $a \neq 0$ 且 τ 是 K 的自同构或反自同构.

直线 K 上的四个点 x_1, x_2, x_3, x_4 构成调和点列, 如果它们的交比 (cross ratio)

$$(x_2 - x_4)^{-1}(x_2 - x_3)(x_1 - x_3)^{-1}(x_1 - x_4)$$

等于 -1.

在 $K = \mathbb{R}$ 为实数域的情形, 这个结果曾为德国数学家 von Staudt (1798–1867) 得到, 于是这个几何定理又被称为 von Staudt-华定理. 在 $K = \mathbb{C}$ 为复数域的情形, 华罗庚对此定理是极为了解的. 这就是下述定理 (参见华罗庚《高等数学引论》第三册)p. 25, 高等教育出版社, 2009):

定理 4 (von-Staudt) 任何一个连续变换如果将复射影直线一一地变为其自己, 并且使得调和点列变为调和点列, 一定是一个广义线性变换.

把这个定理从复数域推广到各种类型的复矩阵空间是华罗庚 1940 年代研究矩阵几何的一个主要动机.

1974 年 Jacobson 的 Basic Algebra 第一卷 [10] 出版, 采纳了 Artin 的建议将上述定理 2 收入到环论一章的习题中. 1977 年, P.M.Cohn 的 Algebra 第二卷 [4] 出版, 由于作者对除环情有独钟, 书中专辟一节讨论了华罗庚的上述定理 2 以及除环中另一个有名的结果 Cartan–Brauer–Hua 定理. 最可喜的是, Cohn 这里 ([4,p.252]) 将华引理的证明完全归结为一个简单的群论事实——一个群不能

写成两个真子群的并, 从而使得华罗庚原先的计算论证变得好懂多了. 这是一个以思想代替计算的典型例子. 我们下边将应用相同的手法解决线性代数和量子力学基础中的两个问题.

1.2 线性代数中的华引理及其应用

我们的目标是证明下述

定理 5 (Birkhoff-Von Neumann) *设 B 是域 F 上向量空间 V 上的双线性型, 使得由 $B(x,y)=0$ 定义的正交关系是对称的, 即 $B(x,y)=0 \Leftrightarrow B(y,x)=0$, 则 B 是对称的或者反对称的.*

这个结果首先出现在 Artin 的专著 [2] 中[1], 后来 Jacobson 又作为定理收录在 [10] 中, 他们的证明是计算性质的. 通过与环中的华引理类比, 作者找到了一个更简单的证明. 只要我们注意到下述看似平凡的结论, 这是华引理在线性代数框架内的一个类比:

引理 2 *设 B 是域 F 上向量空间 V 上的双线性型, 使得对任意的 $x,y\in V$ 或者有 $B(x,y)=B(y,x)$ 或者有 $B(x,y)=-B(y,x)$, 则 B 是对称的或反对称的.*

证明: 对固定的 $x\in V$, 使得 $B(x,y)=B(y,x)$ 成立的 $y\in V$ 构成一个子空间 U_x, 使得 $B(x,y)=-B(y,x)$ 成立的 $y\in V$ 也构成一个子空间 W_x. 由条件有 $V=U_x\cup W_x$. 由于 V 不能写成两个真子空间的并, 于是 $U_x=V$ 或 $W_x=V$. 现在变动 x, 令 $V_1=\{x\in V\,|\,U_x=V\}$, $V_2=\{x\in V\,|\,W_x=V\}$, 于是由条件知, $V=V_1\cup V_2$, 再用一次上述事实, 得到 $V_1=V$ 或 $V_2=V$. 按照 V_1,V_2 的定义, 这分别对应 $B(x,y)=B(y,x)$ 一致成立或 $B(x,y)=-B(y,x)$ 一致成立, 即 B 对称或反对称. ■

下面我们给出定理 5 的基于上述引理的证明.

定理 5 的证明: 对任意的 $x,y\in V$, 令

$$z=B(x,x)y-B(y,x)x,$$

[1]我们这里对定理 5 的表述与 Artin [2] 中表述略有差异, 那里的结论是 (在相同条件下) B 是对称的或交错的. 所谓交错, 就是 $B(x,x)=0$ 对一切 $x\in V$. 但是容易看到这个结论与我们的结论没有区别: 当特征不为 2 时, 反对称的一定是交错的 $B(x,x)=-B(x,x)=0$, 而当特征是 2 时, 反对称型与对称型一致. 按照这个结果, 我们只限于研究对称的与交错的双线性型, 它们对应的几何通常称为正交几何与辛几何.

则

$$B(z,x) = B(x,x)B(y,x) - B(y,x)B(x,x) = 0.$$

从而由已知条件得到

$$B(x,z) = B(x,x)B(x,y) - B(y,x)B(x,x) = 0,$$

也就是

$$B(x,x)(B(x,y) - B(y,x)) = 0. \tag{1}$$

在上式中互换 x, y 有

$$B(y,y)(B(x,y) - B(y,x)) = 0, \tag{2}$$

又在 (1) 中用 $x+y$ 代替 x, 有

$$B(x+y,x+y)(B(x,y) - B(y,x)) = 0. \tag{3}$$

又由极化恒等式得

$$B(x,y) + B(y,x) = B(x+y,x+y) - B(x,x) - B(y,y),$$

联合 $(1),(2),(3)$ 就有

$$(B(x,y) + B(y,x))(B(x,y) - B(y,x)) = 0. \tag{4}$$

从而 $B(x,y) = B(y,x)$ 或 $B(x,y) = -B(y,x)$, 由引理 2 即得结论. ■

读者或许要问: 令 $z = B(x,x)y - B(y,x)x$ 这一步是如何来的? 任何一个熟悉 Schmidt 正交化过程的人都应该能想到这一点. 即为了利用正交的假定, 我们自然应该想到将一个向量沿着另一个向量做正交投影. 用同样的办法, 我们可以证明线性代数中的一个基本结果, 在欧氏空间中, 保持正交关系不变的线性变换一定是某个正交变换的常数倍.[2] 在某种程度上, 下边将要叙述的 Wigner 定理就是这个结果在酉空间上的类似.

1.3 华引理在量子力学中的应用

利用类似的推导, 我们可以证明下述 Hua-Wigner 引理:

[2] 一个更简单的证明是利用平形四边形为棱形当且仅当对角线互相垂直的这一几何事实的代数表达 $(x+y) \perp (x-y) \iff (x,x) = (y,y)$ 构造两个互相正交的向量 $\|y\|x + \|x\|y$ 与 $\|y\|x - \|x\|y$.

引理 3 设 T 是酉空间 V 到自身的一个变换, 使得对任意的 $x, y \in V$ 有 $|(Tx, Ty)| = |(x, y)|$, 且 $\mathrm{Re}(Tx, Ty) = \mathrm{Re}(x, y)$, 则 T 保持加性, 即对任意的 $x, y \in V$ 有 $T(x + y) = Tx + Ty$. 事实上, T 是酉变换或反酉变换[3], 也就是说, T 是保持内积并且是线性的, 或者使内积共轭并且是共轭线性的.

证明: 为证 T 保持加性, 只要证 $z = T(x + y) - Tx - Ty = 0$. 也就是

$$(z, z) = 0$$

注意到 $\mathrm{Re}(z, z) = (z, z)$ 与 T 保实部的条件, 就得到

$$\begin{aligned}(z, z) &= \mathrm{Re}(z, z)\\ &= \mathrm{Re}(T(x + y) - Tx - Ty, T(x + y) - Tx - Ty)\\ &= \mathrm{Re}((x + y) - x - y, (x + y) - x - y)\\ &= 0.\end{aligned}$$

现在根据 T 保持内积的模长与实部推出, 对任意的 $x, y \in V$, 下面两式之一成立:

$$\mathrm{Im}(y, x) = \mathrm{Im}(x, y), \tag{5}$$

$$\mathrm{Im}(y, x) = -\mathrm{Im}(x, y) \tag{6}$$

类似于引理 2 的证明, 在加性条件的保证下我们容易证明, (5) 与 (6) 中的一个对对所有 $x, y \in V$ 一致成立. 从而在各自情形下, 将有

$$(Tx, Ty) = (x, y), \tag{7}$$

$$(Tx, Ty) = \overline{(x, y)}, \tag{8}$$

并且, 对于数乘对应有

$$T(\lambda x) = \lambda Tx, \tag{9}$$

$$T(\lambda x) = \overline{\lambda} x. \tag{10}$$

[3]通常酉变换是指酉空间中满足 $U^*U = UU^* = I$ 的线性变换. 这里我们简单地称 V 上的保持酉内积的线性变换为酉变换, 而反酉变换是指 V 到自身的一个满足以下条件的映射 U:

$$U(x + y) = Ux + Uy, \qquad U(\lambda x) = \bar{\lambda} Ux, \qquad (Ux, Uy) = \overline{(x, y)}.$$

要验证 (9) 与 (10) 只需展开内积 $(T(\lambda x)-\lambda Tx, T(\lambda x)-\lambda Tx)$ 与 $(T(\lambda x)-\overline{\lambda}x, T(\lambda x)-\overline{\lambda}x)$ 即可.■

基于极化恒等式, 可以给出引理 3 的另一个证明. 引理 3 曾出现在 Sharma 和 Almeida 1990 年发表的同一主题的论文 [13] 中, 他们的证明是计算性质的, 大多数物理学家给出的证明也都是计算性质的, 作者在分析 Bergmann 的论文时通过类比华罗庚引理独立地得到了这个引理, 并称之为 Hua-Wigner 引理.

利用这个引理, 可以证明量子力学中的 Wigner 定理. 为叙述这个定理, 我们先给出量子力学的一些基本数学框架.

量子力学中, 一个物理系统的状态由某个酉空间 (通常假定是 Hilbert 空间) 的一条直线 $\langle v\rangle$ 给出, 更确切地说, 其相空间是射影空间 $\mathbb{P}(V):=V-\{0\}/\mathbb{C}^*$, 其元素 (点) 是 非零向量的等价类, 两个非零向量 x, y 等价当且仅当 x, y 线性相关, 我们把 x 所在的等价类记为 $[x]$. 于是关于物理系统的任何有意义的断言从数学上讲都是关于射影空间的点的. 特别的, 两个状态之间的跃迁几率由内积在对应直线上诱导给出 (下边有确切定义). 一个基本的问题是, 决定相空间内这种保持跃迁几率的变换. Wigner 在 1930 年代解决了这个问题, 得到结论说这类变换必定是由酉空间上的酉变换或反酉变换所诱导的射影变换, 这个结果肯定了酉表示在量子力学中的重要性, Sternberg [13] 有附录专门讨论这个定理. Wigner 原先的证明有含糊的地方, 其实那里缺少的正是一个类似于 Hua-Wigner 引理的结果, (见 Weinberg [14] 的讨论, 并与 Wigner [15] 比较), 1960 年代起许多数学物理学家重新严格论证并推广了这个结果, 其中以 Bergmann 的文章 [3] 最具可读性, 这篇文章在极为初等的水平上填补了 Wigner 原先的证明中的漏洞.

为叙述定理, 我们需要指出一些基本的事实. 首先, V 上的线性变换或共轭线性变换 U 可以诱导出 $\mathbb{P}(V)$ 上的一个变换 $[U]$, 因为这两类变换保持线性相关性. 用表达式把这个诱导变换 $[U]$ 写出来就是

$$[U]:\ [x]\mapsto [Ux]$$

其次, V 上的酉内积在 $\mathbb{P}(V)\times\mathbb{P}(V)$ 上可以诱导出一个函数, 即

$$([x],[y])\mapsto \left|\frac{(x,y)^2}{(x,x)(y,y)}\right|$$

它表示跃迁几率 (transition probability), 由 Cauchy 不等式, 它介于 0 和 1 之间.

为便于讨论, 我们引进一个定义, 向量空间 V 到自身的两个映射 U_1, U_2 称为射影等价的, 如果它们在 $\mathbb{P}(V)$ 上诱导出相同的射影变换.

定理 6 (Wigner) : 设 V 是维数大于等于 2 的酉空间, T 是射影酉空间 $\mathbb{P}(V)$ 到自身的一个映射, 且保持跃迁几率不变, 则 T 是 V 上某个酉变换或反酉变换诱导的, 即 $T=[U]$, 其中 U 是 V 上的酉变换或反酉变换.

证明: 对 V 的每个一维子空间 $\langle v\rangle$, 选取其中一个单位向量 v, 并指定一个单位向量 $v'\in T\langle v\rangle$ 与之对应, 对任意单位向量 $w\in\langle v\rangle$, 令 $Tw=\lambda v'$, 其中 $\lambda=(w,v)$ 是 w 关于 v 的 Fourier 系数. 我们以公式 $Tx=\|x\|T\left(\dfrac{x}{\|x\|}\right)$ 将 T 扩张到 $V-\{0\}$, 并令 $T0=0$.

现在立即可以验证, 这个扩充至全空间的变换 (仍记为 T) 保持任意两个向量的内积的模长不变. 由此可以推出 T 保持向量之间的线性相关性, 特别地, 有下面的结果: 若非零向量 x,y 正交, 则存在模长为 1 的两个复数 $\alpha_{x,y},\beta_{x,y}$ 使得 $T(x+y)=\alpha_{x,y}Tx+\beta_{x,y}Ty$. 为看出这一点, 首先注意到, 当 $x\perp y$ 时有 $Tx\perp Ty$. 我们知道,[4] 为证明某向量 z 在两个单位正交向量 $\frac{Tx}{\|Tx\|}$, $\frac{Ty}{\|Ty\|}$ 生成的子空间上, 只需要验证 Parseval 等式

$$\|z\|^2=\left\|\left(z,\frac{Tx}{\|Tx\|}\right)\right\|^2+\left\|\left(z,\frac{Ty}{\|Ty\|}\right)\right\|^2$$

成立. 特别地, 可以验证这个等式对 $z=T(x+y)$ 成立, 而且可以算出 $T(x+y)$ 关于 Tx,Ty 的 Fourier 系数分别是

$$\alpha_{x,y}=\frac{(T(x+y),Tx)}{\|Tx\|^2},\qquad \beta_{x,y}=\frac{(T(x+y),Ty)}{\|Ty\|^2}\tag{$W1$}$$

由此易见它们的模长是 1.

下面从这条性质出发构造一个与 T 射影等价的加性变换 U. 思路是这样的, 由于 T 的上述性质依赖于两个向量正交的条件, 我们先固定一个向量 x_0 并考虑其正交补 $V_0=\langle x_0\rangle^\perp$. 在 V_0 上构造出 U 之后再将它延拓到整个空间 V.

假定函数 $\lambda:V\mapsto\mathbb{C}$ 使得由 $U(x)=\lambda(x)T(x)$ 定义的变换在 V 上是加性的. 对某固定的单位向量 x_0, 考虑 $y\in V_0=\langle x_0\rangle^\perp$, 则由前面的结果有

$$U(x_0+y)=\lambda(x_0+y)T(x_0+y)=\lambda(x_0+y)(\alpha_{x_0,y}Tx_0+\beta_{x_0,y}Ty).\tag{$W2$}$$

另一方面, 因为 U 是加性的, 所以

$$U(x_0+y)=Ux_0+Uy=\lambda(x_0)Tx_0+\lambda(y)Ty,\tag{$W3$}$$

[4] 例如, 参见 Halmos *Finite-Dimensional Vector Spaces*,§64,Theorem1.

比较 $(W2)(W3)$ 两式的右边就有

$$\lambda(x_0+y)=\frac{\lambda(x_0)}{\alpha_{x_0,y}},\qquad \lambda(y)=\frac{\lambda(x_0)}{\alpha_{x_0,y}}\beta_{x_0,y},\tag{$W4$}$$

这就是说 λ 在 V_0 和 V_0 关于 x_0 的平移空间 $\{x_0\}+V_0$ 的值由 $\lambda(x_0)$ 完全确定.

下边我们取定 $\lambda(x_0)=1$ 并用 $(W3)(W4)$ 定义 U 在 V_0 以及 $\{x_0\}+V_0$ 上的作用, 那么容易看到 U 在已定义的集合上与 T 射影等价 (由公式 (5) 可见, $\alpha_{x_0,y}$, $\beta_{x_0,y}$ 模长为 1). 由 Hua-Wigner 引理, 为证明 U 在 V_0 上是加性的, 只要证明 $\mathrm{Re}(Ux,Uy)=\mathrm{Re}(x,y)$. 这一点已经由 Bargmann [3] 给出了精彩的证明, 我们照搬过来. 注意到 U 在定义域上满足以下带限制的加性条件, 即对任意的 $y\in V_0$ 有 $U(x_0+y)=Ux_0+Uy$. 于是对任意的 $x,y\in V_0$ 有

$$|(U(x_0+x),U(x_0+y))|=|(x_0+x,x_0+y)|,$$

也就是

$$|1+(Ux,Uy)|=|1+(x,y)|,$$

即

$$\mathrm{Re}(Ux,Uy)=\mathrm{Re}(x,y),$$

由 Hua-Wigner 引理, U 在 V_0 上是一个酉变换或反酉变换.

由于 $V=\langle x_0\rangle\perp V_0$, 很明显有唯一的方式将 U 线性地或共轭线性地延拓到 V: 对任意的 $x\in V$, 写 $x=\gamma x_0+y$, 并按照 U 在 V_0 上是酉变换还是反酉变换, 分别令 $Ux=\gamma Ux_0+Uy$ 或 $Ux=\overline{\gamma}Ux_0+Uy$. 容易验证, 如此得到的 U 在整个 V 上与 T 射影等价, 而且在整个空间上都是加性的. 再一次利用 Hua-Wigner 引理, 知 U 是 V 上的酉变换或反酉变换. ■

上述证明主要参考了 Bargmann 的文章, 他在文章中还把 Wigner 定理推广到四元数 $\mathbb{H}$ 上的酉空间.

1.4 评论与结语

文献 [11] 将 Wigner 定理与经典的射影几何的基本定理机械地联系起来, 是生搬硬套的一个代表. 经典的射影几何基本定理利用共线性刻画射影变换, 我们引述如下 (参见 [7]):

定理 7 (高维射影几何的基本定理) 设 $V = K^n, K$ 是一个除环, $n \geqslant 3$. T 是 $\mathbb{P}(V)$ 到自身的一个双射, 将任意共线的三点映射为共线的三点, 则 T 具有以下形式

$$T[x] = [A\,\sigma(x)].$$

其中 $A \in GL(n, K)$, σ 是 K 的一个自同构, $\sigma(x)$ 是将 σ 作用到 x 的各个分量上得到的向量.

注: 按定义, 三点 $[x], [y], [z] \in \mathbb{P}(V)$ 共线, 是指向量 $x, y, z \in V$ 线性相关 (从而位于同一个平面, 进而 $[x], [y], [z]$ 位于 $\mathbb{P}(V)$ 的一条直线上).

正如我们上边看到的, Wigner 定理其实与一维射影几何的基本定理——von Staudt–Hua 定理 (定理 3, 定理 4) 的关联更紧密.

据说, Dieudonné 习惯说, 数学家希望因为他们最难的定理而被人们记住, 但是大多数时候, 正是他们最简单的结果在后人中流传. 作者期望, 华罗庚先生的名字至少能因为这个漂亮的引理特别是它的思想及其最简单的应用而被人们记住.

据华罗庚的弟子徐利治 (见 [19, p.198]) 所引述的说法:

> 华罗庚先生跟我讲过, 被引用的研究成果可以分为不同的等级. 第一个等级是在国外的综述性文章或介绍性文章中被引用, 这是一般性引用; 第二个等级是在国外学术专著中被引用, 并且把具体结果写出来; 第三个等级是作为正文被写进教科书中, 这是最高级别的引用.

今天笔者在这里将华先生的这个小结果介绍给读者是一般引用, 鉴于这个思想方法的简单性和美学价值, 笔者期望, 在不多久的明天, 这些内容可以写进本科生线性代数和量子力学的教材中.

参考文献

[1] E. Artin, *Review of Bourbaki's Algebra*, Bull.A.M.S.**59**(1953), 474–479.

[2] E.Artin, *Geometric Algebra,* Interscience, New York, 1957.

[3] V.Bargmann, *Note on Wigner's theorem on symmetry operations,* J.Math,Phys.**5**(1964), 862–868.

[4] P. M. Cohn, *Algebra Vol.2,* London, Wiley, 1977. 251–253.

[5] Loo-Keng Hua, *On the automorphisms of a sfield,* Proc.Nat.Acad.U.S.A.**35** (1949), 386–389.

[6] 华罗庚, 环之准同构及对射影几何的一应用, 中国科学 **1**(1950): 1–6.

[7] 华罗庚, 万哲先, 《典型群》上海: 上海科技出版社, 1963.

[8] N.Jacobson and C.E.Rickart, *Jordan homomorphisms of rings,* Trans.A.M.S.**69** (1950), 479–502.

[9] N.Jacobson, *Lectures in Abstract Algebra, Vol.1,* Van Nostrand, Princeton, 1951.

[10] N.Jacobson, *Basic Algebra, vol.1,* Freeman, San Francisco, 1974.

[11] J.S.Lomont and P.Mendelson, *The Wigner unitary-antiunitary theorem,* Ann.Math.**78**(1963): 548–559.

[12] A. Messian, 《量子力学》(第二卷), 陈学俊、余加莉译, 北京, 科学出版社, 1986.

[13] C.S.Sharma and D.F.Almeida, *A direct proof of Wigner's theorem on maps which preserve transition probabilities between pure states of quantum systems,* Ann.Phys,**197**(1990): 300–309.

[14] B.Simon, *Quantum dynamics: from autmorphim to Hamiltonian,* pages 327-349 in Studies in Mathematical Physics, Essays in Honor of Valentine Bargmann, edited by E.H.Lieb, B.Simon, and A.S.Wightman, Priceton University, 1976.

[15] S.Sternberg, *Group Theory and Physics,* Cambridge University Press, 1994.

[16] V. S. Varadarajan, *Geometry of Quantum Theory* , Van Nostrand, Princeton, 1985.

[17] S. Weinberg, *The Quantum Theory of Fields,* London, Cambridge University Press,1995.

[18] E.P.Wigner, *Group Theory and its Applications to Quantum Mechaics of Atomic Spectra,* Academic Press, New York, 1959.

[19] 徐利治, 《徐利治访谈录》, 湖南教育出版社出版, 2009 年.

第二章　线性代数珍宝十三则

2.1　引言

> 当我做学生的时候, 我们用的矩阵论的教材是 Bocher 的旧得可怜的书 (我认为写得一团糟), 我在这个科目上花的大量时间当中, 我的主要情绪是恼火达到愤怒. ……四、五年以后, 在我已经取得博士学位, 听过了 Von Neumann 讲算子理论以后, 我才真正开始懂得这个科目是讲什么的.
>
> Halmos(哈尔莫斯, 1916–2006), 《我要作数学家》pp.51–52

我相信, Halmos 学习线性代数的经历不是独一无二的, 事实上, 清华大学数学系的周坚教授在给研究生上微分流形这门课的时候说得更直白：

> 第一次学线性代数时我虽然考试成绩不错, 但是听老师说没有人是第一次学线性代数就真正学懂的. 我后来学习微分流形的时候才真正把线性代数搞明白了.

如果笔者没有记错的话, 这是当时他讲到微分流形理论中的 Morse 引理时所发表的感慨, 因为当时需要用到二次型的惯性定理.

对此, 笔者也深有同感, 大学一年级学线性代数的时候似懂非懂的, 只是在别的学科里应用矩阵时才慢慢体会到线性代数的实质与威力.

Halmos 在听了 Von Neumann 的讲座以后不久就动手写了他这一生最有影响的一本书 [7](准确地说, 是 [7] 的前身, 见 [9, pp.126–128]). 根据他自己在 [10, p.550] 的说法, “《有限维向量空间》和《希尔伯特空间问题集》(即参考文献中的 [7] 和 [8]) 或许是他写得最好的书. ”笔者正是从 [7] 中第一次了解到谱定理的极端重要性, 理解到线性代数的重点所在, 并体会到线性代数的乐趣. 这里, 笔者将

这几年来学习线性代数的一点心得与读者一起分享, 也算是作为对 Halmos [7] 的一个回报. 笔者深信, Halmos 必定对 G. H. Hardy 的下述格言深深赞同：

> 美是首要的试金石：丑陋的数学不可能永存.

笔者希望, 这里所选择的一些例子或多或少具有优美的特征.

2.2 AB 与 BA 的特征子空间

一般来说, 两个同阶矩阵的乘积不可交换, 但是有一个基本的事实说, AB 与 BA 最重要的数值不变量——特征值是相同的. 事实上, 我们有这样的结果：设 A, B 是同阶矩阵, 则 AB 与 BA 具有相同的特征多项式. 特别的, 如果 λ 是 AB 的特征值, 具有代数重数 ν, 则 λ 也是 BA 的特征值, 而且具有相同的代数重数 ν. 问题本来可以就此打住, 但是在很多情形我们更感兴趣的是特征值的几何重数而不是代数重数. 因此, 我们或许可以像 Flanders [5] 一样, 转而考虑它们的几何重数是否相同的问题. 即, 对于 AB 的属于特征值 λ 的特征子空间 $E_\lambda(AB) = \{\xi : AB\xi = \lambda\xi\}$ 以及 BA 的属于特征值 λ 的特征子空间 $E_\lambda(BA) = \{\xi : BA\xi = \lambda\xi\}$, 我们要问其维数是否相等？这就引导我们得到下面的结果.

定理 1 设 A, B 分别是 $m \times n$ 和 $n \times m$ 矩阵, $\lambda \neq 0$ 是 AB 的特征值, 则 λ 也是 BA 的特征值, 而且 $\dim E_\lambda(AB) = \dim E_\lambda(BA)$.

证明： 设 $\xi \in E_\lambda(AB)$, 即有 $AB\xi = \lambda\xi$, 两边用 B 同时作用就有 $BA(B\xi) = \lambda(B\xi)$, 即 $B\xi \in E_\lambda(BA)$, 即 $B(E_\lambda(AB)) \subset E_\lambda(BA)$. 类似的, $A(E_\lambda(BA)) \subset E_\lambda(AB)$. 于是, 复合算子 AB 是 $E_\lambda(AB)$ 到自身的一个线性变换, 容易看出这个限制下来的线性变换实际上是以 $\lambda \neq 0$ 为系数的伸缩变换. 因此, 我们立即有 $AB(E_\lambda(AB)) = E_\lambda(AB)$, 注意到 $AB(E_\lambda(AB)) \subset AE_\lambda(BA) \subset E_\lambda(AB)$, 从而 $AE_\lambda(BA) = E_\lambda(AB)$, 由此得到 $\dim E_\lambda(BA) \geqslant \dim E_\lambda(AB)$. 根据对称性有, $\dim E_\lambda(AB) \geqslant \dim E_\lambda(BA)$. 这就证明了 $\dim E_\lambda(AB) = \dim E_\lambda(BA)$. ■

通常对“$\lambda \neq 0$ 是 AB 的特征值, 则 λ 也是 BA 的特征值”这一事实的证明是借助于熟知的行列式等式

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

而上述证明则完全是几何的. 或许, 这一证明很好地佐证了 Artin [1, p.14] 的话：

> 我的经验是, 一个用矩阵进行的证明, 如果你抛开矩阵的话往往可以使这个证明缩短一半. 有时, 这一点是办不到的, 你需要计算一个行列式.

[1] 这本书名字为《几何代数》, 言下之意就是说用几何的方法、从几何的角度来研究代数. 这个观点既是 Halmos [7] 所倡导的观点, 也是本文强调的重点所在.

2.3 保角变换

在欧氏空间中, 如果 $x, y \in \mathbb{R}^n$ 不为零, 则 x, y 之间的夹角可以用内积 $\langle , \rangle$ 定义为

$$\angle(x, y) = \cos^{-1} \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle}\sqrt{\langle y, y\rangle}} \in [0, \pi]$$

根据 Cauchy-Schwarz 不等式, 这个定义是有意义的. $\mathbb{R}^n$ 的可逆线性变换 T 称为保角的, 如果对任意的非零向量 $x, y \in \mathbb{R}^n$ 有 $\angle(Tx, Ty) = \angle(x, y)$. 一个基本的事实是, “保角”就是“保形”(常常称为“共形”), 即保持形状, 其精确的数学含义就是相似. 显然, 每个相似变换都是保角的. 因此, 我们提到的这个基本事实无非就是说其逆命题也成立, 可以认为这是相似的基本定理 (它不是别的, 只是中学平面几何中关于三角形相似的 AAA 定理的本质):

定理 2 *欧几里得空间 $\mathbb{R}^n$ 上的每一个保角变换 T 具有形式 $T = kS$, 其中 k 为非零常数, S 是 $\mathbb{R}^n$ 的正交变换.*

事实上, 我们可以证明下述更强的结论:

定理 3 *设 T 是欧氏空间中的可逆线性变换, 则以下三个条件等价:*

(i) *T 是保角线性变换.*
(ii) *T 保持任意两个向量之间的垂直 (正交) 关系不变.*
(iii) *T 是某个正交变换的非零常数倍.*

这里 (i) $\Rightarrow$ (ii) 与 (iii) $\Rightarrow$ (i) 是显然的, 而我们有两种方法证明 (ii) $\Rightarrow$ (iii). 第一种是借助于 Schmidt 正交化手续, 将任意两个向量变换为两个垂直的向量, 然后利用保持垂直关系做计算. 第二种方法更富技巧性, 是利用下述代数引理来正交化任意两个向量.

引理 1 *在欧氏空间中, $(x + y) \perp (x - y) \Longleftrightarrow \|x\| = \|y\|$.*

于是对任意的 $x, y \in V$, 构造向量 $\|y\|x + \|x\|y$ 与 $\|y\|x - \|x\|y$, 它们互相正交. 然后可以利用假设展开证明, 我们留给有兴趣的读者.

定理 2 与定理 3 其实是非常简单而基本的事实, 但是从一些文献的处理来看, 似乎作者并没有说清楚. 尤其是 Spivak [25, p.4] 习题 8, 作者似乎不知道保角变换可以用相似的概念来精确刻划, 而且中译本译者似乎也被作者弄糊涂了, 他们在 [26, pp.144–147] 所附加的“习题解答与提示”中给出的证明也是不得要领. 其他的作者则或者将这个事实推给线性代数 (例如 [17, pp.15–16]), 或者将相似变换直接拿来作为保角变换的定义 (如 [13, pp.561–562 的附加习题 1])；总而言之, 他们总是假定读者在线性代数中了解了这个基本的事实. 或许, 这个结果有必要写进线性代数的教科书. 正如初中几何的主题是全等和相似, 保角变换应该得到与等距变换同等程度的强调, 而且这两个概念之间的关系应该得到澄清.

引理 1 表达了这样一个熟知的几何事实：平形四边形为菱形当且仅当对角线互相垂直. 注意到, 其对偶 $x \perp y \iff \|x + y\| = \|x - y\|$ 表达的是一个与之对偶的几何事实：平行四边形为矩形当且仅当对角线等长.

2.4 华罗庚引理的再次应用

在 [15] 中笔者类比抽象代数中环同态的华罗庚引理提出了所谓的“线性代数中的华罗庚引理”(见 [15, p.41 引理 1]) :

引理 2 设 B 是域 F 上向量空间 V 上的双线性型, 使得对任意的 $x, y \in V$ 或者有 $B(x, y) = B(y, x)$ 或者有 $B(x, y) = -B(y, x)$, 则 B 是对称的或反对称的.

笔者进一步利用这个引理对线性代数中的下述基本结果 (例如, 见 [30, 第 385 页]) 给出了一个更简单的概念性的证明.

定理 4 (Birkhoff–Von Neumann) 设 B 是域 F 上向量空间 V 上的双线性型, 使得由 $B(x, y) = 0$ 定义的正交关系是对称的, 即 $B(x, y) = 0 \iff B(y, x) = 0$, 则 B 是对称的或者反对称的.

事实上, 笔者后来注意到, 应用同样的想法, 利用数学归纳法, 我们可以将上述结果从双线性映射推广到多线性映射. 即, 事实上我们可以证明下述结果：

定理 5 设 A 是域 F 上向量空间 V 上的 n 重线性函数, 使得如果 $A(x_1, \ldots, x_n) = 0$, 则对任意的置换 $\sigma \in S_n$, 有 $A(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = 0$, 则 A 是对称的或者反对称的.

事实上, 只有 $n=3$ 的情况需要单独证明, 其他情况直接应用归纳假设很容易证明. 对于 $n=3$ 的情况我们有所谓的结辫引理 (见 [2, 第一卷 p.224 页]):

引理 3 设 V 和 W 是特征不等于 2 的域上的向量空间, 且 $A: V\times V\times V\to W$ 是一个三线性映射, 对于前两个变量是对称的, 对于后两个变量是反对称的, 则 $A=0$.

2.5 幂等矩阵

从几何上讲, 最简单的线性变换是投影变换, 它的代数表现就是幂等矩阵 ($P^2=P$). 从代数上讲, 幂等矩阵只有唯一的不变量——秩：两个幂等矩阵相似当且仅当它们的秩相等. 关于这类最简单的矩阵, 我们也可以说出一些不平凡的事实, 以下就是一个例子.

定理 6 设 P,Q 是有限维向量空间 V 上的投影变换, 且 $P-Q$ 与 $Q-P$ 仅有平凡的不动点, 则 P,Q 相似.

证明： 令 $A=Q-P-I, B=P-Q-I$, 则

$$PB=P(P-Q-I)=P^2-PQ-P=-PQ,$$

$$AQ=(Q-P-I)Q=Q^2-PQ-Q=-PQ,$$

于是 $PB=AQ$, 由条件可知 A,B 可逆, 从而 $Q=A^{-1}PB$, 即 Q 与 P 等价, 从而 P,Q 有相同的秩. 由于秩是投影变换在相似下的唯一不变量, 所以 P,Q 相似. ■

当 V 为内积空间且 P,Q 为正交投影变换时, 由于秩也是正交投影变换在正交相似下的唯一不变量, 所以 P,Q 正交相似. 上述定理出现在 J.-H. Wang 与 P. Y. Wu [28] 定理 1.3. 该定理有一个无限维版本, 可见于 Riesz 与 Sz.-Nagy [22, p.268].

定理 7 设 P,Q 是 Hilbert 空间 H 到子空间 V 和 W 上的正交投影, 而且

$$\|P-Q\|<1$$

则 V 与 W 等距, 即存在一个变换 $T: V\to W$, 它是保距的.

2.6 Cochran 定理的代数本质

线性代数中有许多结果都有着代数的本质, 概率统计中的 Cochran 定理就是一个著名的例子. 1950 年代, Chipman 和 Rao 指出, 1930 年代发现的 Cochran 定理实则是一个简单的矩阵定理, 而它所讨论的不过只是幂等矩阵的代数性质. 他们将这个定理表述如下：

定理 8 设 $A_1,\dots,A_k$ 为域 F 上 n 阶矩阵使得 $\sum_i A_i = A$ 是幂等的, 且 $\sum_i \operatorname{rank}(A_i) = \operatorname{rank}(A)$, 则 $A_1,\dots,A_k$ 都是幂等矩阵, 而且当 $i \neq j$ 时, $A_iA_j = 0$.

这个定理有许多证明, 一个巧妙证明可见 [7, p. 109] 习题 5. 或许 Waterhouse [29] 提供的下述证明是最简单的一个, 因为他发现这个线性代数结果其实有着环论的背景. 他的证明由以下两个引理构成.

引理 4 记 $\mathcal{M} = \mathcal{M}(n,F)$ 为 F 上 n 阶矩阵的集合. 设 $A_1,\dots,A_k \in \mathcal{M}$ 使得 $\sum_{i=1}^k \operatorname{rank}(A_i) = \operatorname{rank}(\sum_{i=1}^k A_i)$, 则

$$\left(\sum_i A_i\right)\mathcal{M} = \bigoplus_i A_i\mathcal{M}$$

证明： 令 $A = \sum_{i=1}^k A_i$. 很明显, $A\mathcal{M} \subset A_1\mathcal{M}+\cdots+A_k\mathcal{M}$. 注意到对任意的 n 阶矩阵 C, 集合 $C\mathcal{M}$ 中的矩阵都是那些每一列都在 C 的值域中的向量构成的矩阵, 因此 $\dim C\mathcal{M} = n\cdot\operatorname{rank}(C)$, 从而有 $\dim A\mathcal{M} = n\cdot\operatorname{rank}(A)$ 以及 $\dim A_i\mathcal{M} = n\cdot\operatorname{rank}(A_i)$. 因此, 条件 $\sum_i \operatorname{rank}(A_i) = \operatorname{rank}(A)$ 意味着 $\dim A\mathcal{M} = \sum \dim A_i\mathcal{M}$, 从而有 $A\mathcal{M} = \bigoplus_i A_i\mathcal{M}$. ■

引理 5 设 R 是一个带单位元的环. e 是 R 中的幂等元, $e = e_1+\cdots+e_k$, 并且满足 $eR = \oplus_{i=1}^k e_iR$, 则 $e_1,\dots,e_k$ 都是幂等元, 而且当 $i \neq j$ 时, $e_ie_j = 0$.

证明： 我们有 $e_i = e_i\cdot 1 \in e_iR \subset eR$, 从而不妨设 $e_i = es_i$. 于是 $e_i = es_i = e^2s_i = e(es_i) = ee_i = (e_1+\cdots+e_k)e_i = \sum_j e_je_i$, 由于 $\oplus_{i=1}^k e_iR$ 是直和分解, 所以 $e_i = e_i\cdot 1 = e_ie_i = e_i^2$, 且对于 $j \neq i$ 有 $e_je_i = 0$. ■

只要将引理 5 应用于环 $R = \mathcal{M} = \mathcal{M}(n,F)$, 并注意到引理 4 的结论使得引理 5 的条件成立, 即可推出定理 8. 或许我们可以借用 J. Sylvester 的一句话[1]来评论 Waterhouse 的这个证明：

[1] 引自 M. Artin, *Algebra*, Pearson Education, Inc., 1991, p. 197. 中译本《代数》p.150, 郭晋云译, 机械工业出版社, 2009 年.

一种奇怪的智力现象, 可以归结为, 证明普遍论点通常比对其中的个别情况作出证明简单得多.

注：目前大陆最为全面的线性代数教材 [23, 259–261] 以例子的形式给出了这个定理, 但显然那里给出的证明过于复杂. 笔者在 2006 年曾写信告诉 [23] 的作者北京大学的丘维声教授 (当时 [23] 的前身收入了这个结果作为练习) 这个问题的上述简单证法, 丘教授在回信中分享了他的两种证明, 而且非常巧合的是, 在那一年的北京大学数学院研究生招生考试的高等代数试卷中出现了这道题.

1978 年，吉林大学谢邦杰教授将上述定理推广到任意的体上, 见 Cochran 定理在任意体上的推广.

2.7 Von Neumann 特征值问题

线性代数中最重要的结果当属谱定理, 特别是实对称矩阵、Hermite 矩阵和酉矩阵的谱定理. 然而, 具体情况下求出矩阵的谱分解就是要求解矩阵的特征值与特征向量, 毫无疑问, 这是线性代数中最核心的问题. 然而, 一般来说, 求解特征值问题并不是一件容易的事. Von Neumann 解过一个有意思的特征值问题 (该问题也收入 L. Lovász [18, p.20 习题 1.29]), 我们介绍如下, 用以挑战有兴趣的读者.

问题：求实二次型

$$\sum_{\mu=1}^{n-1}(x_{\mu+1}-x_\mu)^2 = x_1^2 + 2\sum_{\mu=2}^{n-1} x_\mu^2 + x_n^2 - 2\sum_{\mu=1}^{n-1} x_\mu x_{\mu+1}$$

所对应的实对称矩阵的特征值.

注：每一个 n 元二次型 $f(x) = \sum_{i,j=1}^n a_{ij}x_ix_j$ 对应着唯一的 n 阶对称矩阵 $A=(a_{ij})$, 使得 f 可以写为 $f(x)=x^TAx$, 其中 $x=(x_1,\dots,x_n)^T$.

答案：$\lambda_\mu = 2-2\cos\frac{\mu\pi}{n} = 4\sin^2\frac{\mu\pi}{2n}$, $\quad \mu=0,1,\dots,n-1$.

正如张尧庭在 [31, 第 352–353 页] 所回忆的, 中国近代数学家许宝騄曾经“也考虑过这个问题, 但是没有解决, 因为一个特定矩阵的特征值很难求, 而 *Von Neumann* 却把特征值和特征向量放在一起来求, 这样问题反而解决了, 所以他从这里学到了东西. ”

另一个相近的特征对问题 (同样用差分方程求解) 的结果（一个简单的推导参见这里）如下：

定理: $n\times n$ 矩阵 $T=\sum_{\substack{1\leqslant i,j\leqslant n\\ |i-j|=1}} E_{ij}$ 的特征值是 $\lambda_k = 2\cos\frac{k\pi}{n+1}$, $1\leqslant k\leqslant n$, 对应的特征向量是 $v_k=(\sin\frac{k\pi}{n+1}, \sin\frac{2k\pi}{n+1},\dots,\sin\frac{nk\pi}{n+1})$.

2.8 Schur 特征值问题

矩阵论的另一个有趣的应用是用于计算数论中的 Gauss 和

$$\sum_{s=0}^{n-1} \varepsilon^{s^2}, \qquad (\text{其中 } \varepsilon = e^{i\frac{2\pi}{n}} \text{ 是 } n \text{ 次单位根})$$

Gauss 首先确定了这一求和, 然而他的方法非常巧妙. [2]后来, I. Schur 发现了一种较为简单的方法, 他注意到, 这个 Gauss 和可以表达为一个 n 阶矩阵

$$S = (s_{jk}) = (\varepsilon^{(j-1)(k-1)}), \qquad (j,k = 1,\ldots,n)$$

的迹. Schur 求出了这个矩阵的所有特征值, 从而确定了 Gauss 和. 因为 Schur 的这一贡献, 这一矩阵也被命名为 Schur 矩阵, 而且通常记为 S. 关于 Schur 特征值问题的具体求法, 我们留给有兴趣的读者, 也可以参考 [14, pp.207–212]. 这里我们只给出最后结果如下：S 只有四个特征值, $\sqrt{n}, -\sqrt{n}, i\sqrt{n}, -i\sqrt{n}$, 按照 n 模 4 的不同, 其重数分别为：

(1) 若 $n = 4k+1$, 则对应的重数分别为 $k+1, k, k, k$.
(2) 若 $n = 4k+2$, 则对应的重数分别为 $k+1, k+1, k, k$.
(3) 若 $n = 4k+3$, 则对应的重数分别为 $k+1, k+1, k+1, k$.
(4) 若 $n = 4k+4$, 则对应的重数分别为 $k+2, k+1, k+1, k$.

于是我们最后可以得到 Gauss 的著名公式 (见 [14, 第 191 页])：

定理 9 设 n 是一个正整数, 则

$$\sum_{s=0}^{n-1} e^{i\frac{2\pi}{n}s^2} = \begin{cases} (1+i)\sqrt{n} & \text{当 } n \equiv 0 \pmod 4 \\ \sqrt{n} & \text{当 } n \equiv 1 \pmod 4 \\ 0 & \text{当 } n \equiv 2 \pmod 4 \\ i\sqrt{n} & \text{当 } n \equiv 3 \pmod 4 \end{cases}$$

定理 9 在历史上具有极高的重要性, 它是 Gauss 在 1801 年 5 月记录下的一则日记 (见 [3, p.481]), 他利用这个结果给出了二次互反律的第五个证明 (他一生总共给出了六个证明). 但是, Gauss 为了确定出 Gauss 和的符号, 竟然费了好几年时间, 他在 1805 年 8 月 30 日记录下这样的日记 (见 [3, pp.481–482])：“经过

[2]例如, 可见 T. Nagell, *Introduction to Number Theory*, AMS Chelsea Publishing, 1964, pp. 177–180.

四年多的不懈努力之后, *1801* 年 *5* 月所记录的富有魅力的定理的证明终于大功告成了. ”即便如此, Gauss 最后得到这个结果也是灵感的杰作. 他在 1805 年 9 月 3 日写给 Olbers 的一封信中如是说道:“最后, 只是几天以前, 终于成功了 *(*我想说, 不是由于我苦苦的探索, 而只是由于上帝的恩惠*)*. 就像是闪电轰击的那一刹那, 这个谜解开了; 我以前的知识, 我最后一次尝试的方法、以及成功的原因, 这三者究竟是如何联系起来的, 我自己也未能理出头绪来. ”[3]

注: $U = \frac{1}{\sqrt{n}}S$ 是一个酉矩阵, 它与离散 Fourier 变换有密切的关系, 对此有兴趣的读者请参考 Taussky [27] 中关于 Schur 矩阵的一节的论述.

2.9 Von Neumann 恒等式

Von Neumann 的主要贡献可能是无限维空间的算子代数, 但是他对矩阵论也作出了巨大的贡献. 前面我们已经提到过, 他巧妙地求解了一个特征值问题, 他 1940 年代发现的小结果 (见 [20]) 也可以算得上一颗小珍珠.

定理 10 对于同阶复方阵 A 和 B, 有

$$\|AB^* - B^*A\|^2 - \|AB - BA\|^2 = \mathrm{Tr}[(A^*A - AA^*)(B^*B - BB^*)]$$

其中 $\|A\|^2$ 的定义为

$$\|A\|^2 = \mathrm{Tr}(AA^*)$$

并由此得到结论:

定理 11 若 A, B 之一为正规矩阵[4], 则 A, B 可换蕴含 A, B^* 可换.

注: 定理 11 在无限维也成立, 称为 Fuglede 可换性定理, 见 [6, p.68]. 而且, 事实上我们有更为一般的 Putnam-Fuglede 定理, 见 [8, p.78 问题 152].

2.10 酉矩阵的遍历定理

如果复数 u 满足 $|u| \leqslant 1$, 则平均值

$$a_n = \frac{1}{n}\sum_{k=0}^{n-1} u^k$$

[3] 参见 Gauss, *Gesammelta Werke*, Vol.10, 第 24–25 页. 英文引自 A. Borel, *Mathematics: art and science*, The Mathematical Intelligence, **5**:4 (1983) 9–17. 中译文《数学——艺术与科学》, 江嘉禾译, 《数学译林》第 4 卷 (1985 年) 第 3 期, 243–253.

[4] 矩阵 A 称为正规矩阵, 如果 A 与 A^* 可交换.

构成一个收敛的序列, 这是经典分析中一个简单而有趣的结果. 我们来看一下其证明. 若 $u=1$, 则 $a_n=1$, 所以 $\{a_n\}$ 收敛到 1. 如果 $u\neq 1$, 则

$$a_n=\frac{1}{n}\sum_{k=0}^{n-1}u^k=\frac{1}{n}\frac{1-u^n}{1-u}$$

由于 $|u|\leqslant 1$, 所以 $|1-u^n|\leqslant 1+|u^n|\leqslant 1+1=2$, 从而

$$|a_n|=\frac{1}{n}\frac{|1-u^n|}{|1-u|}\leqslant\frac{1}{n}\frac{2}{|1-u|}$$

所以 $\{a_n\}$ 收敛到 0.

利用酉矩阵的谱定理 (酉矩阵酉等价于一个对角矩阵), 可以将这一结果从复数 (一阶矩阵) 推广到高阶矩阵上, 从而得到下述有名的 Von Neumann 平均遍历定理.

定理 12 *设 U 是有限维内积空间 $\mathcal{H}$ 上的酉变换, 设 W 是 U 的不动点构成的子空间, 则由*

$$A_n=\frac{1}{n}(I+U+\cdots+U^{n-1})$$

定义的序列当 $n\to\infty$ 时收敛到正交投影 P_W.

这里所谓“A_n 收敛到 P_W”是指, 对任意的 $x\in V$, $\|A_nx-P_Wx\|\to 0$, 即泛函分析中的 L^2 范数收敛.

现在 Von Neumann 本人的原始证明已经难得见到了, 因为文献中往往收入的都是 F. Riesz 对这一定理给出的简单而漂亮的证明, 这里我们概述如下, 具体的细节读者可以自行补充, 或者参见 [7], [8] 或 [22].

证明： 设 R 为 $I-U$ 的值域, 则容易证明, 对 $x\in R$, A_nx 收敛到 0. 另一方面, 当 $x\in W$ 时, $A_nx=x$ 收敛到 x. 最后, 注意到 W 和 R 互为正交补. ■

事实上, F. Riesz 的上述证明还可以推广到所谓的压缩算子, 即满足 $\|T\|\leqslant 1$ 的算子, 其要点在于, 下面的引理成立.

引理 6 : *设内积空间 $\mathcal{H}$ 上的算子 T 满足对任意 $x\in\mathcal{H}$, 有 $\|Tx\|\leqslant\|x\|$, 则 T 与 T^* 有相同的不动点.*

对此有兴趣的读者, 我们推荐读者去参看 [8] 或 [22].

另一方面, 在矩阵的情形, 我们有下述最一般的结论 (见 [4, p.560])：

定理 13 矩阵 C 使得

$$\frac{1}{n}(I + C + \cdots + C^{n-1})$$

收敛当且仅当 C 的所有特征值的模长小于等于 1, 并且对于模长为 1 的特征值, 其代数重数等于几何重数.

2.11 辛矩阵的行列式为 1：Siegel 的证明

令域 F 上的 $2n$ 阶矩阵

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

则 F 上满足 $PJP' = J$ 的 $2n$ 阶矩阵 P 称为辛矩阵, F 上的所有 $2n$ 阶辛矩阵构成一个群, 称为辛群, 记为 $Sp(2n, F)$. 辛群与正交群一起构成两类最重要的典型群. 从某种程度上讲, 辛矩阵要比正交矩阵简单 (正如反对称双线性型要比对称双线性型简单), 例如辛矩阵的行列式总是 1. 我们将对 $F = \mathbb{R}$ 的情形证明这一点, 方法来自 C. L. Siegel [25].

定理 14 实辛矩阵的行列式为 1.

证明： 设 $P \in Sp(2n, \mathbb{R})$, 写成与 J 一致的分块

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

于是 P 满足 $AB' = BA', AD' - BC' = I, CD' = DC'$. 注意到, 当 $P \in Sp(2n, \mathbb{R})$ 时, $P' \in Sp(2n, \mathbb{R})$, 从而有 $A'C = C'A$, $A'D - C'B = I$, $B'D = D'B$. 于是

$$\begin{bmatrix} iC' + D' & -iA' - B' \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & iI \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & iC + D \end{bmatrix}$$

两边取行列式有 $(\det P - 1)\det(iC + D) = 0$, 只要证明 $\det(iC + D) \neq 0$. 注意到 $(iC + D)(-iC' + D') = CC' + DD'$, 只要证明 $CC' + DD'$ 可逆. 用反证法, 不然, 假设存在非零向量 $x \in \mathbb{R}^n$ 使得 $(CC' + DD')x = 0$, 由此 (仅在此处用到基域 $F = \mathbb{R}$ 的假定) 推出 $C'x = D'x = 0$, 从而

$$\begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix} \begin{bmatrix} 0 \\ x \end{bmatrix} = \begin{bmatrix} C'x \\ D'x \end{bmatrix} = 0$$

此与 $P' \in Sp(2n, \mathbb{R})$ 可逆矛盾! ■

或许有些奇怪, 一个实数域的问题最终要放到复数域来解决, 然而这并没有什么奇怪的, J. Hadamard 早就告诫过我们:“实数域中两个真理之间的最短路程是通过复数域. ”

2.12 行列式的传递性

M. H. Ingraham[11] 曾注意到, 对某一类特殊的分块矩阵, 其行列式有一简单公式, 我们介绍如下.

设 $M = (A_{ij})_{n\times n}$, 其中每个 A_{ij} 都是域 F 上的 $r \times r$ 的矩阵, 而且这 n^2 个子块矩阵 A_{ij} 彼此交换, 则我们有

$$\det M = \det(\det{}_R M),$$

其中 $\det_R M$ 称为 M 的约化行列式, 是将 $M = (A_{ij})_{n\times n}$ 的各个分块矩阵 A_{ij} 当做普通的数 a_{ij}(这就是约化的含义) 并按照普通行列式的公式来计算得到的结果, 即我们有

$$\det{}_R M = \sum_{\sigma\in S_n} (\mathrm{sgn}(\sigma) A_1\sigma(1)\cdots A_n\sigma(n).$$

因此, $\det_R M$ 也是一个 $r \times r$ 的矩阵, 而 Ingraham 的上述结果断言, 若对该矩阵再计算其行列式, 就得到分块矩阵 M 的行列式. 这一公式不仅可以大大化简满足此类矩阵之行列式计算, 而且可以用来证明域上代数的范数传递性质, 参见 Jacobson[13] 第 7 章第 4 节. 此处我们仅给出当 $F = \mathbb{C}$ 为复数域 (它适用于一切代数闭域, 并且由此可以推出一般情形下的结果) 时的一个证明, 一般情形下的证明可以参见 Jacobson[13] 第 7 章第 4 节或 Ingraham 的短文 [11]. 当 $F = \mathbb{C}$ 时, Jacobson[13] 给出了一个别致的证明, 我们援引如下.

证明: 由于矩阵 A_{ij} 两两可交换, 根据线性代数的一个基本结果 (参见 [12] 第 4 章第 9 节), 存在 $r \times r$ 复可逆矩阵 P 使得所有 A_{ij} 同时上三角化, 即

$$P^{-1}A_{ij}P = B_{ij} = \begin{bmatrix} \lambda_{ij_1} & * & \cdots & * \\ 0 & \lambda_{ij_2} & \ddots & * \\ 0 & \ddots & \ddots & * \\ 0 & \cdots & 0 & \lambda_{ij_r} \end{bmatrix}.$$

因此，若令 Q 为 n 个 P 构成的对角矩阵, 即

$$Q=\begin{bmatrix} P & & & \\ & P & & \\ & & \ddots & \\ & & & P \end{bmatrix},$$

则有

$$Q^{-1}MQ=\begin{bmatrix} B_{11} & B_{21} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nn} \end{bmatrix}=M_1.$$

显然我们有 $\det A=\det(Q^{-1}AQ)=\det M_1$. 另一方面, 由于 $B_{ij}=P^{-1}A_{ij}P$, 所以根据 $\det_R$ 的定义, 显然有

$$\det\nolimits_R M_1=P^{-1}(\det\nolimits_R M)P,$$

从而

$$\det(\det\nolimits_R M)=\det(P\det\nolimits_R M_1P^{-1})=\det(\det\nolimits_R M_1).$$

因此, 我们只需要验证, 对全部由两两可交换的上三角矩阵构成的分块矩阵 M_1 有

$$\det M_1=\det(\det\nolimits_R M_1).$$

由 $\det_R$ 的定义以及上三角矩阵的加法与乘法性质可得

$$\det\nolimits_R M_1=\begin{bmatrix} \det\Lambda_1 & * & \cdots & * \\ 0 & \det\Lambda_2 & \ddots & * \\ 0 & \ddots & \ddots & * \\ 0 & \cdots & 0 & \det\Lambda_r \end{bmatrix}.$$

其中

$$\Lambda_k=\begin{bmatrix} \lambda_{11_k} & \lambda_{21_k} & \cdots & \lambda_{1n_k} \\ \lambda_{21_k} & \lambda_{22_k} & \cdots & \lambda_{2n_k} \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_{n1_k} & \lambda_{n2_k} & \cdots & \lambda_{nn_k} \end{bmatrix},\qquad k=1,2,\ldots,r.$$

从而

$$\det(\det{}_R M_1) = \det\Lambda_1 \det\Lambda_2 \cdots \det\Lambda_r.$$

接下来我们计算 $\det M_1$. 我们对 $nr \times nr$ 矩阵 M_1 做以下行列置换: 对 $i = 1, 2, \ldots, n$ 以及 $j = 1, 2, \ldots$, 将 M_1 的第 $(i-1)r+j$ 行变成第 $(j-1)n+i$ 行, 同时将 M_1 的第 $(i-1)r+j$ 列变成第 $(j-1)n+i$ 列, 这 (整个变换相当于对 M_1 做了一个相似变换) 就给出矩阵

$$TM_1T^{-1} = \begin{bmatrix} \Lambda_1 & * & \cdots & * \\ 0 & \Lambda_2 & \ddots & * \\ 0 & \ddots & \ddots & * \\ 0 & \cdots & 0 & \Lambda_r \end{bmatrix},$$

其中 Λ_k 如前. 因此,

$$\det M_1 = \det(TM_1T^{-1}) = \det\Lambda_1 \det\Lambda_2 \cdots \det\Lambda_r.$$

这就证明了 $\det M_1 = \det(\det_R M_1)$, 进而有 $\det M = \det(\det_R M)$. ■

2.13 线性递推关系 (差分方程) 的稳定性

在我的本科线性代数老师田代军教授讲授的内容中, 令我很着迷的一个例子, 是用矩阵方法来求解 Fabonacci 数列的通项. 更一般地, 可以用矩阵来求解递推关系 $u_{k+1} = Au_k$, 其中 $u_k \in \mathbb{C}^n, A \in \mathcal{M}(n, \mathbb{C})$。特别的, 迭代矩阵 A 的谱决定了该系统的稳定性。我们有下述经典的结果：

定理 15 设 $A \in \mathcal{M}(n, \mathbb{C})$, 其谱半径为 ρ，则我们有下述结论（参见 [23, pp.383-386, 518 and 630]）：

(i) $\lim\limits_{k\to\infty} A^k = 0$(此时 A 称为收敛的) 当且仅当 $\rho < 1$。

(ii) $\lim\limits_{k\to\infty} A^k$ 存在且不等于 0 当且仅当 1 是 A 的半单特征值 (几何重数 = 代数重数) 而且其它特征值模长都小于 1. 且此时 $\lim\limits_{k\to\infty} A^k$ 到 $E_1(A)$ 的投影.

(iii) 若 $\rho(A) = 1$，模长为 1 的特征值仅有 1，且它是单重的，令 ξ 与 η^T 分别是相对应的右特征向量和左特征向量，则 $\lim\limits_{k\to\infty} A^k = \frac{\xi\cdot\eta^T}{\eta^T\cdot\xi}$。

(iv) A^k 有界当且仅当 $\rho \leqslant 1$ 且模长为 1 的特征值（如果有的话）是半单的。

对于具有初值 u_0 的差分方程组 $u_{k+1} = Au_k$ 的解为 $u_k = A^k u_0$，从而推出系统的稳定性结论如下：在情形 (i) 系统是稳定的 (stable), 对任意的初值 u_0，$\lim\limits_{k\to\infty} u_k = 0$；在情形 (ii) 系统是稳定的 (neutrally stable)，而在其它情形 ($\rho > 1$) 系统是不稳定的 (unstable)。类似地，对常系数微分方程之稳定性讨论，有类似结果，此时与几何数列 A^k 对应的，是矩阵指数函数 $\exp(At)$，此处从略。

特别的，如果我们考虑的矩阵 A 是谱半径等于 1 的正矩阵，则根据正矩阵的基本性质 (Perron-Froubenius 定理, 参见 [13] 定理 8.2.7)，此时将满足 (iii)，并且可以算出 $\lim\limits_{k\to\infty} A^k = \frac{\xi\cdot\eta^T}{\eta^T\cdot\xi}$，其中 ξ 与 η^T 分别是 A 所对应的右特征向量和左特征向量 (分别称为 Perron 右向量和 Perron 左向量)。

更进一步，若 A 是行随机的正矩阵，则它满足上述条件，并且 ξ 可取为 $\xi = (1, 1, \ldots, 1)^T$。丁玖等 [3] 注意到，这种特殊情形下的结果，可以用来讨论由重复平均所给出的数列 (或点列) 的极限，他那里的主要结果如下：

定理 16 *设点列 $\{x_n\}_{n=1}^{\infty} \subset \mathbb{C}^m$ 满足递推关系*

$$x_{n+k} = \alpha_1 x_n + \alpha_2 x_{n+1} + \cdots + \alpha_k x_{n+k-1}, \qquad n \geqslant 1,$$

初值为 $x_1, \ldots, x_k$。若 $\alpha_1, \ldots, \alpha_k$ 是正数，且 $\alpha_1 + \cdots + \alpha_k = 1$，则有

$$\lim_{n\to\infty} x_n = \sum_{i=1}^{k} \left(\frac{\sum_{j=1}^{i} \alpha_j}{k + 1 - \sum_{j=1}^{k} j\alpha_j} \right) x_i.$$

这结果是概率论中更新定理 (renewal theorem) 的一个特例，见 [6, p. 254](10.18) 式.

我们利用定理 15 来证明，为此，我们令

$$u_n = \begin{bmatrix} x_n \\ x_{n+1} \\ \vdots \\ x_{n+k-1} \end{bmatrix} \in \mathbb{C}^{km}, \qquad n = 1, 2, \ldots,$$

于是递推关系可以写为 $u_{n+1} = A_m u_n$，其中

$$A_m = \begin{bmatrix} 0 & I_m & 0 & \cdots & 0 \\ 0 & 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_m \\ \alpha_1 I_m & \alpha_2 I_m & \cdots & \cdots & \alpha_k I_m \end{bmatrix}$$

容易看出 $A_m = A_1 \otimes I_m$ 相似于 $\underbrace{A_1 \oplus \cdots \oplus A_1}_{m}$，因此只要证明 $A_1 = A$ 满足定理 15 条件 (iii)，容易看出 A 的特征多项式为

$$f(x) = x^k - \alpha_k x^{k-1} - \cdots - \alpha_2 x - \alpha_k$$

当 $\alpha_1 + \cdots + \alpha_k = 1$ 时，$f(1) = 0$，即 1 是 A 的特征根，不难证明 1 是 A 的谱半径。为证明 1 时是单重的且 $f(x) = 0$ 的其它根都满足 $|x| < 1$，我们需要引用 Ostrowski 的下述结果 (见 [26] 定理 12.2):

定理 17 设 $b_1, \ldots, b_m)$ 是非负实数, 令

$$q(x) = x^m - b_1 x^{m-1} - \cdots - b_m.$$

设 $\gcd\{k \in \{1, \ldots, m\} : b_k > 0\} = 1$，则 $q(x)$ 有唯一的正根 r. 并且 r 是 q 的单根, 且 q 的任何其它根模长小于 r.

如果我们考虑常系数的递推关系组而不只是一个数列，则将建议以下问题：若将递推关系中的系数 α_i 改为 m 阶非负矩阵 A_i，即我们考虑

$$x_{n+k} = A_1 x_n + A_2 x_{n+1} + \cdots + A_k x_{n+k-1}, \qquad n \geqslant 1,$$

并假定 $A = \sum_{i=1}^{k} A_i$ 是行随机矩阵，问 x_n 是否有极限，如果有，可否求得 x_∞。

初步分析：此时如前设 u_n，则递推关系可以 $u_{n+1} = Bu_n$，其中

$$B = \begin{bmatrix} 0 & I_m & 0 & \cdots & 0 \\ 0 & 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_m \\ A_1 & A_2 & \cdots & \cdots & A_k \end{bmatrix}$$

于是我们要做的就是，讨论当 $A_1, \ldots, A_k$ 满足何种条件时，对 B 可以应用推广的 Perron–Frobenius 定理 (丁玖 [3] 的思路)，或者可以保证 B 的不变因子具有简单的结构（如上的思路)?

2.14 Pauli 矩阵

笔者的论文 [16] 实际上是从一个习题开始的, 或许这个小问题是值得了解的[5]

[5] 事实上, 笔者最初对这个问题感兴趣, 也是从一个习题引起的, 见 F. W. Byron and R. W. Fuller, *Mathematics of Classic and Quantum Physics*, Vol.1, Addison-Wesley Publishing Company(1968), p. 139, Ex.27.

练习：设 2 阶复矩阵 $A_1,\dots,A_k$ 满足矩阵方程

$$\begin{cases} A_i^2 = I \\ A_iA_j = -A_jA_i \qquad (i\neq j) \end{cases} \qquad (i,j=1,\dots,k)$$

证明

(1) k 的最大值是 3, 并且这 3 个矩阵可以全部取为 Hermite 矩阵.

(2) 若 $A_1,\dots,A_k$ 都是对称矩阵, 则 k 的最大值是 2.

(3) 若 $A_1,\dots,A_k$ 都是实矩阵, 则 k 的最大值是 2.

注：在理论物理中, 著名的 Pauli 矩阵

$$\sigma_1=\begin{bmatrix}0&1\\1&0\end{bmatrix},\qquad \sigma_2=\begin{bmatrix}0&i\\-i&0\end{bmatrix},\qquad \sigma_3=\begin{bmatrix}1&0\\0&-1\end{bmatrix}$$

满足上述方程. Pauli 矩阵或许是理论物理中出现最为频繁的矩阵, 原因正如费曼所一语道破的:“*Pauli* 的自旋矩阵和算子不是什么新的东西, 而正是 *Hamilton* 的四元数”. [6]复数的重要性在数学中已经得到了广泛的认识, 而四元数在物理学中的重要性则似乎有待进一步挖掘, 杨振宁先生在他的论文选中如是说[7]:

> $SU(2)$ 对称的存在, 一定有一个理由, 因为在最根本的层次上, 造化的安排一定不会是无缘无故的, 这种说法已经不止一次得到应验. 除此之外, 我们期待着的解释, 可能要用到四元数代数, 因为四元数的对称就是 $SU(2)$. 此外, 四元数代数是一种美丽的结构. 虽然它不是交换的. 但是我们知道, 造化选择非交换的代数作为量子力学的语言, 她怎么会拒绝使用这仅有的另一种可能的美妙代数作为她在宇宙万物中建立起来的所有复杂对称性的语言呢?

小结：几何与代数

总之, 为了使大部分学生能最有效地学习线性代数, 我们应该尽可能地强调几何的语言和方法. 几何的语言, 自然是相对于代数的语言而说的. 简单地讲, 就

中译本《物理学中的数学方法》, 熊家炯、曹小平译, 科学出版社, 1984 年. 也参见朱栋培，狄拉克方程和反易自逆矩阵的定理，《中国科学技术大学学报》，1975 年第 2 期。

[6]见 *The Pleasure of Finding Things Out: The Best Short Works of Richard P. Feynman*, edited by Jeffrey Robbins, Perseus Books, 1999. 中译本《发现的乐趣》第 205 页, 张郁乎译, 湖南科学技术出版社, 2007 年.

[7]见杨振宁, *Selected Papers 1945–1980 With Commentary*, San Francisco: Freeman & Co, 1983. 中译文《优雅的四元数》, 收入《杨振宁文集》第 35–38 页, 张奠宙主编, 华东师范大学出版社, 1998 年.

是用线性变换代替矩阵, 用抽象向量代替 n 维列向量. 几何语言的优点是简洁明快, 例如“作用”这个词给人的感觉就是如此. 代数语言的好处是具体清晰, 两个矩阵“相乘”在我们头脑中的图象自然是一系列具体运算的运作. 通常的教科书都过分强调了代数的语言, 这同时也充分暴露了它的诸多缺点. 最大的缺点在于坐标不具有内蕴的含义. 或许可以提一句, 正如陈省身先生所注意到的, 爱因斯坦将狭义相对论推广到广义相对论花了七年之久就是因为他一直摆脱不了坐标的束缚[8]:

> 如果你觉得接受一般的坐标概念有困难, 那么你并不孤单. 爱因斯坦花了七年时间才从狭义相对论过度到广义相对论. 他对之所以延迟了这么久的解释是:“为什么建立广义相对论又用了七年时间呢? 主要原因是: 要摆脱‘坐标必须具有直接的度量意义’这个旧观念是不容易的.”

在很多问题中坐标的选取并不重要, 我们所需要的往往只是一些基本的运算规律, 例如分配律、结合律等. 这时候抽象的几何语言就十分适用了. [9] 例如在内积空间的理论中, 往往采用几何的语言, 而且事实上, 线性代数的几何直观性在这里得到了最大的体现. 特别地, 实对称矩阵的谱定理可以叙述为: 在欧氏空间中, 对给定的对称变换, 存在一组由特征向量构成的标准正交基. 我们正是靠这种几何观点来指引具体的代数运算的, 例如所谓 Schmidt 正交化, 无非就是将第二个向量沿第一个向量作垂线, 一旦指出这一点, Schmidt 正交化公式就很容易理解了.

本文主要强调了线性代数的几何观点, 这是 Von Neumann–Halmos 传统的延续. 用几何的语言叙述的线性代数教科书, 这里向读者推荐 Halmos [7]. 但是应该指出, 线性代数还有许多不同的侧面. 我们这里主要介绍了另外两个不同的观点作为补充, 或许某些读者是感兴趣的.

Taussky 在 [27] 中叙述了她所了解的矩阵论, 这也是一家之言, 特别的, 她在结语部分对读者所提的**几点忠告**很值得重提:

> 当你观察到数的一个有趣性质时, 也许你没有把它当做是 $n \times n$ 矩阵的有趣性质在 $n = 1$ 的情形. 请想一想, $GL(n, F)$ 或 $SL(n, F)$, $GL(n, \mathbb{Z})$ 或 $SL(n, \mathbb{Z})$.
>
> 当你有一对有趣性质的矩阵时, 请研究它们生成的束或代数.

[8] 见陈省身《从三角形到流形》, 收入《陈省身文集》pp. 233–243, 张奠宙主编, 华东师范大学出版社, 2002 年.

[9] 所以, H. Weyl 这么说, “以坐标的形式把数引进几何学, 是一种暴力行为”. 见上一脚注中提到的陈省身的文章第 237 页.

> 当某一个矩阵的行列式被证明是重要的时候, 试从整体上探究一下这个矩阵, 例如, 是否可以作为代数数域的判别式矩阵.
>
> 当某一个一元多项式使你感兴趣时, 试考虑一下以之为特征多项式的矩阵.
>
> 当有人瞧不起矩阵时, 请回想一下对矩阵论作出了重要贡献的大数学家, 比如, Frobenius, Schur, Siegel, Ostrowski, Motzkin, Kac, I. M. Gel'fand 等. [10]

此外, 笔者在 2010 年参加了在北京清华大学召开的第五届世界华人数学家大会, 聆听了 MIT 数学系的 Gilbert Strang 所做的报告, 按照他的说法, Israel Gohberg(1928–2009) 是近一百年里最伟大的线性代数专家, 对此笔者也不甚理解, 或许是因为 Gohberg 的工作更偏向于应用吧![11]

最近笔者就这个问题请教了 Strang 教授, 他答复如下：

> 一个方法是从亚马逊网页上浏览一下他的著作标题. 而且现在有许多奉献给他的书, 还有许多线性代数方面的论文. 他与 Harry Dym 解决的一个小问题如下：如果你有一个三对角的实对阵矩阵, 主对角占优, 如何将它补充成一个实对称矩阵, 使它的行列式取得最大值? 补充之后的实对称矩阵的逆又是什么——它为什么是三对角的?
>
> 也许你可以让学生求解一个 3×3 的例子. [大概是我在通信中没有说清楚自己的学生身份, 以至于让 Strang 教授误以为我是一名教师了！]

致谢

我总结这份材料是响应 2006 年徐泽同学的一个建议, 她认为在毕业时可以给下一届的学弟学妹们留一点什么, 或许对他们有帮助. 这就是本文的初衷.

其次, 我要感谢刘云朋同学, 这里的许多数学点子都蒙他赐教. 感谢首都师大学赵洁、段红伟同学, 作者从与她们的讨论中获益良多.

最后, 我要感谢我的线性代数老师田代军, 他向我推荐了 Jacobson 的经典教科书 [12][13] 以及华罗庚与万哲先合著的《典型群》, 让我体会到学习线性代数的快乐.

特别要感谢审稿人, 对作者的初稿提出了许多有益的建议.

[10] 或许这个名单上还应该加上华罗庚、许宝騄和 Milnor 的名字.

[11] 参见 Harm Bart 等人的回忆文章In memoriam Israel Gohberg: August 23, 1928 - October 12, 2009, *Linear Algebra and its Applications* Volume 433, Issue 5, 15 October 2010, Pages 877–892.

参考文献

[1] E. Artin, *Geometric Algebra*, Interscience, New York, 1957.

[2] M. Berger, *Geometry*, (Corrected fouth printing), Springer, 2009. 中译本《几何》, 陈志杰、周克希译, 科学出版社, 1989 年.

[3] J. Ding and T. Fay, The Perron–Frobenius theorem and limits in geometry, American Mathematics Monthly, 112(2005), 171–175.

[4] G. W. Dunnington, *Gauss: Titan of Science*, Math. Asso. Amer., 2004.

[5] N. Dunford and J.T. Schwartz, *Linear Operators, Part* I, *General Theory*, John Wiley & Sons, 1988.

[6] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 1*, 有中译本《概率论及其应用》, 胡迪鹤译, 人民邮电出版社, 2014 年.

[7] H. Flanders, *Elementary divisors of AB and BA*, Proc. Amer. Math. Soc., **2**(1951), 871-874.

[8] P. R. Halmos, *Hilbert Space and the Theory of Spectral Multiplicity*, Chelsea, 1951.

[9] P. R. Halmos, *Finite-Dimensional Vector Spaces*, Springer, 1974.

[10] P. R. Halmos, *A Hilbert Space Problem Book*, Springer-Verlag, 1967. 中译本《希尔伯特空间问题集》, 林辰译, 上海科学技术出版社, 1984 年.

[11] P. R. Halmos, *I Want to Be a Mathematician*, Springer-Verlag, 1985. 中译本《我要作数学家》, 马元德、沈永欢、胡作玄、赵慧琪译, 江西教育出版社, 1999 年.

[12] P. R. Halmos, *Linear Algebra Problem Book*, Math. Asso. Amer., 1995.

[13] Roger Horn and Charles Johnson, *Matrix Analysis* , Chapter 5, Cambridge University Press, 1985. 有两个中译本：《矩阵分析》，杨奇译，机械工业出版社，2005 年；《矩阵分析》，张明尧、张凡译，机械工业出版社，2014 年。

[14] M. H. Ingraham, *A note on determinants*, Bull. Amer. Math. Soc. 43 (1937), 579–580.

[15] N. Jacobson, *Lectures in Abstract Algebra II. Linear Algebra*, GTM31, 1953. 中译本《抽象代数学卷 2 线性代数》, 黄缘芳译, 科学出版社, 1960 年.

[16] N. Jacobson, *Lectures in Abstract Algebra III. Theory Of Fields And Galois Theory*, GTM31, 1964. 中译本《抽象代数学卷 3 域论及伽罗瓦理论》, 李忠侯、俞曙霞、李世余译, 科学出版社, 1987 年.

[17] N. Jacobson, *Basic Algebra, Vol.1,* Freeman, San Francisco, 1974.

[18] E. Landau, *Elementary Number Theory*, AMS Chelsea Publishing, 1999.

[19] 林开亮, Hua 引理及其应用,《数学传播》第 34 卷 (2010 年) 第 135 期, 39–48.

[20] 林开亮, Hurwitz–Radon 矩阵方程, 《数学传播》第 36 卷 (2012 年) 第 1 期，48–63.

[21] L. H. Loomis and S. Sternberg, *Advanced Calculus*, Addison-Wesley Publishing Company, 1968. 中译本《高等微积分》, 王元、胥鸣伟译, 高等教育出版社, 2006 年.

[22] L. Lovász, *Combinatorial Problems and Exercises*, Second Edition, AMS Chelsea Publishing, 2007.

[23] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

[24] J. Von Neumann, *Distribution of the ratio of the mean square successive difference to the variance*, Ann. Math. Stat., **12**(1941), 367–395. (收入 *Collected Works*, Vol.4, 452–480, 特别的, 见第 456–457 页.)

[25] J. Von Neumann, *Approximative properties of matrices of higher finite order*, Portugaliae Math.**3**(1942), 1–62. (收入 *Collected Works*, Vol.4, 270–331, 特别的, 见第 330 页.)

[26] A.M. Ostrowski, *Solution of Equations and Systems of Equations*, second edition, Academic Press, New York and London, 1966.

[27] R. Remmert, *Theory of Complex Analysis*, Graduate Texts in Mathematics, Reading in Mathematics, V.122, Springer, 1991.

[28] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Dover Publications Inc, New York, 1990.

[29] 丘维声, 《高等代数》(下册), 清华大学出版社, 2010 年.

[30] C. L. Siegel, *Symplectic geometry*, Amer. J. Math., **65**(1943), 1–86.

[31] M. Spivak, *Calculus on Manifold*, Benjamin, New York, 1965. 中译本《流形上的微积分》, 齐民友、路见可译, 人民邮电出版社, 2005 年.

[32] O. Taussky, *How I became a torchbearer for matrix theory*, Amer. Math. Monthly, **95** (1988), 801–812. 我是怎样成为矩阵论传人的, 冯慈黄译, 《数学译林》第 9 卷 (1990 年) 第 2 期, 119–129.

[33] J.-H. Wang and P. Y. Wu, *Difference and similarity models of two idempotent operators*, Linear Algebra Appl., 208/209 (1994), 257-282.

[34] W. C. Waterhouse, *Cochran's theorem for rings*, Amer. Math. Monthly, **108** (2001), 58–59.

[35] 姚慕生、吴泉水, 《高等代数学》, 复旦大学出版社, 2008 年.

[36] 张尧庭, 深深的怀念——我所知道的许宝騄先生, 收入《道德文章垂范人间：纪念许宝騄先生百年诞辰》pp.344–356, 北京大学出版社, 2010 年.

第三章　Hurwitz 定理的矩阵证明

3.1　介绍

本文将对德国数学家 Adolf Hurwitz 1898 年得到的下述定理给出一个简单证明.

定理 1 设 $n \geqslant 1$. 则存在 $x_1, \ldots, x_n$ 与 $y_1, \ldots, y_n$ 的实双线性函数 $z_1, \ldots, z_n$ 使得

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = (z_1^2 + \cdots + z_n^2) \tag{1}$$

对一切实变数 $x_1, \ldots, x_n$ 与 $y_1, \ldots, y_n$ 成立当且仅当 $n = 1, 2, 4, 8$.

首先, 在 Hurwitz 之前, 大家已经注意到对 $n = 1, 2, 4, 8$, 存在以下引人注目的公式：

对于 $n = 1$ 等式是平凡的：$x_1^2 y_1^2 = (x_1 y_1)^2$.

对 $n = 2$ 我们有

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1 y_1 - x_2 y_2)^2 + (x_1 y_2 + x_2 y_1)^2 \tag{2}$$

这不过是复数模长的乘积法则的实形式.

类似地, 由四元数范数的可乘性质可以得到由 Euler 1748 年发现的四平方和乘积公式：

$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2 \tag{3}$$

其中 z_1, z_2, z_3, z_4 为

$$z_1 = x_1 y_1 - x_2 y_2 - x_3 y_3 - x_4 y_4, \quad z_2 = x_1 y_2 + x_2 y_1 + x_3 y_4 - x_4 y_3$$

$$z_3 = x_1 y_3 - x_2 y_4 + x_3 y_1 + x_4 y_2, \quad z_4 = x_1 y_4 + x_2 y_3 - x_3 y_2 + x_4 y_1$$

在 $n=8$ 时, 存在一个类似的平方和的乘积公式, 见维基百科链接内容, 最早发现这个公式的丹麦数学家 F. Degen, 那是在 1818 年. Cayley 后来指出, 这个公式可以从八元数 (又称 Cayley 数) 的范数乘法公式给出 Cayley 还进一步考虑了这样的问题, 对于 $n=16$ 或者更一般的 $n=2^k$ 是否存在类似的平方和乘积公式? 或者更一般的, 我们可以提出下列问题:

Hurwitz 问题: 对于哪些自然数 n, 存在 n 个变数的平方和乘积公式?

1898 年, Hurwitz [3] 对这一问题给出了一个漂亮的解答, 这就是上述定理 1. Hurwitz 的证明是如此的简单优美, 以至于 20 多年以后, 美国代数学家 L. E. Dickson 专门撰文 [2] 介绍了 Hurwitz 的证明, 这一经 Dickson 修改后的证明被 Curtis 收录于他的线性代数教科书 [1] 中. 笔者这里将通过借助于 O. Veblen 和 J. Von Neumann [6] 中的一个引理进一步简化 Hurwitz 的证明.

3.2 Hurwitz 问题的矩阵约化

Hurwitz 解决上述平方和问题的关键在于将这一问题转化为一个矩阵问题. 我们现在就遵循 Hurwitz 的思路完成这个转化.

假定存在 n 个变数的平方和公式, 设 z_i 作为 x 和 y 的实双线性函数表达如下:

$$z_i=\sum_{h,j} b_{ihj}x_h y_j, (i,h,j=1,2,\cdots,n) \tag{4}$$

其中 $b_{ihj}\in\mathbb{R}$.

如果对 $h=1,\ldots,n$, 我们令 n 阶矩阵

$$B_h=(b_{ihj}), (i,j=1,\ldots,n)$$

则条件 (1) 将等价于

$$(x_1^2+\cdots+x_n^2)E=(\sum_{h=1}^n B_h x_h)(\sum_{k=1}^n B_k x_k)'=(\sum_{h=1}^n B_h x_h)(\sum_{k=1}^n B_k' x_k)$$

这就得到原始的 Hurwitz 矩阵方程如下:

$$B_h B_k'+B_k B_h'=2\delta_{hk} \tag{5}$$

如果我们令 $A_h=B_n^{-1}B_h$, 这里 $h=1,\cdots,n-1$, 则容易验证 $A_1,\cdots,A_{n-1}$ 满足

$$A_iA_j + A_jA_i = -2\delta_{ij}; \qquad A_i' = -A_i \tag{6}$$

反之, 如果 $\mathcal{M}(n,\mathbb{R})$ 中存在 $n-1$ 个矩阵 $A_1,\ldots,A_{n-1}$ 满足 (6), 则容易验证这 $n-1$ 个矩阵和单位阵一起满足 (5). 所以 Hurwitz 问题就归结为讨论 $\mathcal{M}(n,\mathbb{R})$ 中是否存在 $n-1$ 个矩阵 $A_1,\ldots,A_{n-1}$ 满足 (6)? (Hurwitz 在后来的文章中进一步考虑了这样的问题：设矩阵 $A_1,\ldots,A_k \in,\mathcal{M}(n,\mathbb{R})$ 满足 (6), 则 k 的最大值 $K(n)$ 是多少? 对于这一问题, 也有一个比较简单的解答, 有兴趣的读者可以参见 [4] 或 [5]). Hurwitz 证明了下述

定理 2 $\mathcal{M}(n,\mathbb{R})$ 中存在 $n-1$ 个矩阵 $A_1,\ldots,A_{n-1}$ 满足 (6) 当且仅当 $n = 1,2,4,8$.

由此推出定理 1, 这就解决了平方和的乘积公式的问题.

3.3 Hurwitz 矩阵方程的解的基本性质

为了证明定理 2, 我们需要 Hurwitz 矩阵方程的一些基本性质. 对我们而言, 最关键的性质将是下面的

引理 1 设 $A_1,\ldots,A_k \in \mathcal{M}(n,\mathbb{R})$ 满足

$$A_iA_j + A_jA_i = -2\delta_{ij} \tag{7}$$

则以下 2^{k-1} 个矩阵

$$I,\ A_i(1\leqslant i\leqslant k-1), A_{i_1}A_{i_2}(1\leqslant i_1<i_2\leqslant k-1),\ldots,A_1\cdots A_{k-1} \tag{8}$$

线性无关.

证明：我们先来证明下述观察:

$A_1,\ldots,A_k$ 中任意取 l 个 $(1\leqslant l\leqslant k-1)$ 不同矩阵作成的乘积 $A_{i_1}\cdots A_{i_l}$ 的迹为零.

事实上, 若 l 是奇数, 则对任意一个不同于 $i_1,\ldots,i_l$ 的指标 j 有

$$A_j(A_{i_1}\cdots A_{i_l})A_j^{-1} = (-1)^l(A_{i_1}\cdots A_{i_l})A_jA_j^{-1} = -(A_{i_1}\cdots A_{i_l}),$$

两边取迹就得到

$$\mathrm{Tr}(A_{i_1}\cdots A_{j_l}) = -\mathrm{Tr}(A_{i_1}\cdots A_{j_l})$$

即 $\mathrm{Tr}(A_{i_1}\cdots A_{j_l})=0$.

若 l 是偶数, 则对于 $i_1,\ldots,i_l$ 中的任意一个指标, 比方说 i_1, 有

$$A_{i_1}(A_{i_1}\cdots A_{i_l})A_{i_1}^{-1}=(-1)^{l-1}(A_{i_1}\cdots A_{i_l})A_{i_1}A_{i_1}^{-1}=-(A_{i_1}\cdots A_{j_l}),$$

两边取迹就得到 $\mathrm{Tr}(A_{i_1}\cdots A_{i_l})=0$.

现在我们来证明 (8) 中的矩阵线性无关. 假定存在实数 $c^0,c^i,,c^{i_1i_2},\ldots,c^{12\cdots(k-1)}$ 使得

$$c^0I+a^iA_i+c^{i_1i_2}A_{i_1}A_{i_2}+\cdots+c^{i_1i_2\cdots i_l}A_{i_1}\cdots A_{i_l}+\cdots+c^{12\cdots(k-1)}A_1\cdots A_{k-1}=0,$$

在上式两边同乘以 $A_{i_1}\cdots A_{i_l}$ 并取迹我们得到

$$c^{i_1i_2\cdots i_l}=0.\blacksquare$$

注：这里的引理取自 [6], 它代替了 [1] p. 321–322 的引理.

3.4 定理 2 的证明

现在我们可以给出定理 2 的证明了, 因为前面我们已经给出了 $n=1,2,4,8$ 时的平方和乘积公式, 所以充分性是没有问题的, 下面我们只证明必要性.

证明：不妨设 $n\geqslant 2$. 若 $\mathcal{M}(n,\mathbb{R})$ 中存在 $n-1$ 个矩阵 $A_1,\ldots,A_{n-1}$ 满足 (6), 那么特别的, 它们满足 (7), 于是由引理知, $\mathcal{M}(n,\mathbb{R})$ 中存在 2^{n-2} 个线性无关的矩阵, 从而我们有下列维数控制：

$$2^{n-2}\leqslant n^2$$

另一方面, 奇数阶反对称矩阵不可逆, 所以从 A_1 满足的条件

$$A^2=-I,A'=-A$$

推出 n 一定为偶数.

然而当 $n\geqslant 10$ 时我们可以归纳证明出：

$$2^{n-2}>n^2$$

($2^{n+1-2}=2\cdot 2^{n-2}>2n^2>(n+1)^2$) 所以 n 只能等于 $2,4,6,8$ 之一. 下面我们只需要排除 $n=6$ 的情况.

如果 $A_1,\dots,A_5\in,M(6,\mathbb{R})$ 满足 (6), 我们将证明下面 16 个反对称矩阵

$$A_i,(1\leqslant,i\leqslant 5),\quad A_iA_j(1\leqslant i<j\leqslant 5),\quad ,A_1A_2A_3A_4A_5$$

是线性无关的. 但是我们知道 $M(6,\mathbb{R})$ 中的反对称矩阵构成的子空间的维数为 $\frac{6\times5}{2}=15$, 这就得到了矛盾, 从而否定了 $n=6$ 这种可能.

为了证明上面的 16 个矩阵线性无关, 我们只需要再次应用引理的证明中用到的那个观察：$A_1,\dots,A_5$ 中取任意的不超过 4 个矩阵作成的乘积的迹等于零. 事实上, 设存在实数 c^i,c^{ij},c^{12345} 使得

$$\sum_{i=1}^{6}c^iA_i+\sum_{1\leqslant i<j\leqslant 6}c^{ij}A_iA_j+c^{12345}A_1A_2A_3A_4A_5=0$$

在等式两边同乘以 A_k 得到

$$c^iA_iA_k+c^{ij}A_iA_jA_k+c^{12345}A_1A_2A_3A_4A_5A_k=0$$

两边取迹得到 $c^k=0$. (注意 $A_1A_2A_3A_4A_5A_k$ 在不计正负号的意义下其实只是四项的乘积, 因为反交换性和 $A_k^2=-I$.) 类似的, 可以证明其他各个系数 $c^{ij}=0,c^{12345}=0$. ■

3.5 评注

事实上, 这里我们并不需要将数域限制在实数域 $\mathbb{R}$, 整个推理对复数域 $\mathbb{C}$ 甚至是任意的特征不等于 2 的域 F 都适用. 换言之, 本文的方法可以证明下面的

定理 3 设 F 是一个特征不等于 2 的域, 则 $\mathcal{M}(n,F)$ 中存 $n-1$ 个矩阵 $A_1,\dots,A_{n-1}$ 满足 (6) 当且仅当 $n=1,2,4,8$.

由此可得更一般的 Hurwitz 定理 (我们有理由相信, Dickson 本人已经看出这样的结果成立).

定理 4 设 F 是一个特征不等于 2 的域且 $n\geqslant 1$. 则存在 $x_1,\dots,x_n$ 与 $y_1,\dots,y_n$ 的双线性函数 $z_1,\dots,z_n$ 使得

$$(x_1^2+\cdots+x_n^2)(y_1^2+\cdots+y_n^2)=(z_1^2+\cdots+z_n^2)\tag{3.1}$$

对一切 $x_1,\dots,x_n,y_1,\dots,y_n\in,F$ 成立当且仅当 $n=1,2,4,8$.

从我们的证明来看, 定理 4 的神奇之处不在于必要性方面, 因为上述论证是如此的简单. 定理 4 的神奇之处在于充分性方面, 例如, Euler 发现的四平方和的乘积公式 (3), 这绝不是一个平凡的公式. 希望我们的读者也因此而记住 Euler. 应该认识到, 像 (3) 那样的平方和等式是极其稀有的, 物以稀为贵, 我们应当像珍珠一样收藏.

3.6 拓展

作为对读者的一个考验, 我们在最后提出以下问题供有兴趣的读者思考.

问题：能不能找到 Euler 四平方和乘积公式的一个狭义相对论版本, 换言之, Lorentz 二次型 $Q(X) = x_1^2 + x_2^2 + x_3^2 - x_4^2$ 在 $\mathbb{R}$ 上是否有合成公式？这等价于问, 是否存在 X, Y 的双线性乘积 XY 使得 $Q(XY) = Q(X)Q(Y)$？

参考文献

[1] C. W. Curtis, Linear Algebra, An Introductory Approach, UTM, Springer, 1996.

[2] L. E. Dickson, On quaternions and their generalization and the history of the eight square theorem, Ann. of Math., series2, 20(1919), 155-171.

[3] A. Hurwitz, quadratischen Formen von beliebig vielen Variablen, Nachrichten Ges. der Wiss. Gottingen,1898, 309–316.

[4] 江上鷗, Hurwitz–Radon 问题——等价和约化是处理数学问题的一种基本方法, 《数学通报》, 1964 年 04 期.

[5] 林开亮, Hurwitz–Radon 矩阵方程, 《数学传播》, 36 卷 1 期 (2012), pp. 48-63.

[6] O. Veblen and J. Von Neumann, Geometry of Complex Domain, Princeton Mimeographed Notes, Notes by W. Givens and A. H. Taub, Institute for Advanced Study, 1935–1936.

第四章　向量积与旋度算子从三维到高维的推广

4.1　向量积

众所周知，在三维欧氏空间中，对两个向量 v, w，除了数量积 (也称内积或点乘)$v \cdot w$，还有向量积 (也称叉乘)$v \times w$。而且，数量积的概念是高维欧氏空间的基础，在欧氏空间中度量长度、角度 (特别地，垂直关系) 都依赖于数量积。那么，很自然地，我们可以问下述问题：向量积是否可以从三维空间推广到高维空间呢？对于这个问题，直到 1942 年才由 Beno Eckamann(1917–2008)[1] 首先考虑并完全解决。1983 年，美国著名的代数拓扑学家 William S. Massey(1920–2017) 从中提炼出下述代数定理 (见 [3]):

定理 1 (Massey 定理，只有三维或七维空间存在向量积) 设 $n \geqslant 2$，则当且仅当 $n \in \{3, 7\}$ 时 n 维欧氏空间 $\mathbb{R}^n$ 中可以定义一个满足以下两个条件的双线性向量积 $V \times V \to V : (v, w) \mapsto v \times w$：

(i) $\langle v \times w, v\rangle = \langle v \times w, w\rangle = 0$,

(ii) $\|v \times w\|^2 = \|v\|^2\|w\|^2 - \langle v, w\rangle^2$,

正如 Massey 所指出的，定理 1 是下述著名的 Hurwitz 定理 (参见 [2]) 的推论。

定理 2 当且仅当 $n \in \{1, 2, 4, 8\}$ 时存在 $x_1, \ldots, x_n$ 与 $y_1, \ldots, y_n$ 的实系数双线性函数 $z_1, \ldots, z_n$ 使得

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = (z_1^2 + \cdots + z_n^2)$$

对一切实数值变量 $x_1, \ldots, x_n$ 与 $y_1, \ldots, y_n$ 成立。

定理 2 又称为“1,2,4,8 定理”，它也可以用矩阵来表述 (参见 [2])，即

定理 3 设 $B_1,\ldots,B_k$ 是 n 阶实矩阵，且满足下述 Hurwitz 方程

$$B_iB_j+B_jB_i=-2\delta_{ij}I_n,\qquad B_i^T=-B_i, \tag{1}$$

其中 δ_{ij} 是 Kronecker 符号 (当 $i=j$ 时取值 1，否则取值 0)，I_n 是 n 阶单位阵，B_i^T 表示 B_i 的转置矩阵。则 $k\leqslant n-1$，并且等号成立当且仅当 $n\in\{1,2,4,8\}$。

定理 2 与定理 3 的证明可参见 [2]。考虑到从定理 2 到定理 1 的推导有益于我们理解后面的论证 (从定理 3 推导定理 4)，这里我们给出定理 1 的证明。

证明：先证必要性。考虑空间 $\mathbb{R}^{n+1}=\mathbb{R}\oplus\mathbb{R}^n$。对任意的 $x=(a,v),y=(b,w)\in\mathbb{R}\oplus\mathbb{R}^n=\mathbb{R}^{n+1}$，定义乘法如下:

$$xy=(a,v)(b,w)=(ab-\langle v,w\rangle,aw+bv+v\times w).$$

容易验证，$\mathbb{R}^{n+1}$ 中的上述乘法是双线性的，并且满足

$$\begin{aligned}
\|xy\|^2&=\|(ab-\langle v,w\rangle,aw+bv+v\times w)\|^2\\
&=(ab-\langle v,w\rangle)^2+\langle aw+bv+v\times w,\,aw+bv+v\times w\rangle\\
&=(ab-\langle v,w\rangle)^2+a^2\|w\|^2+b^2\|v\|^2+\|v\times w\|^2+2ab\langle v,w\rangle+2a\langle w,v\times w\rangle+2b\langle v,v\times w\rangle\\
&=a^2b^2+\langle v,w\rangle^2+a^2\|w\|^2+b^2\|v\|^2+\|v\times w\|^2\\
&=a^2b^2+\langle v,w\rangle^2+a^2\|w\|^2+b^2\|v\|^2+\|v\|^2\|w\|^2-\langle v,w\rangle^2\\
&=a^2b^2+a^2\|w\|^2+b^2\|v\|^2+\|v\|^2\|w\|^2\\
&=(a^2+\|v\|^2)(b^2+\|w\|^2)\\
&=\|x\|^2\|y\|^2.
\end{aligned}$$

由定理 2，有 $n+1\in\{2,4,8\}$，从而 $n\in\{1,3,7\}$。

充分性的证明是构造性的，$n=1$ 的情况平凡，$n=3$ 的情况即我们熟知的向量积，$n=7$ 的情况可见维基百科条目seven-dimensional cross product。**证毕**。

定理 1 告诉我们，与数量积可定义于任意维数的空间不同，满足适当条件的向量积仅存在于 3 维与 7 维空间。

4.2 旋度算子

在三维空间中, 有与向量积密切相关的旋度算子。然而，在一般的高等微积分教材中，对高维空间并没有关于关于旋度算子的讨论。这就引出一个问题：是否与

向量积的情况类似, 关于旋度算子存在一个类似于定理 1 的结果？事实上, 7 年前有人在 math.stackexchange 论坛明确提出这样的问题(can-the-curl-operator-be-generalized-to-non-3d?)。针对这一问题，我们得到下述平行于定理 1 的结果：

定理 4 (只有三维或七维空间存在旋度算子) 设 $V = C^2(\mathbb{R}^n)$ 是 $\mathbb{R}^n$ 上的二次连续可微函数空间, 则当且仅当 $n \in \{3, 7\}$ 时 V^n 上存在满足下述三个条件的一阶常值实系数微分算子 $\mathrm{curl} : V^n \to V^n$：

(i) 对任意的 $F \in V^n$ 有

$$\mathrm{div}(\mathrm{curl}F) = 0,$$

这里 div 是作用在 V^n 上的散度算子.

(ii) 对任意的 $f \in V$ 有,

$$\mathrm{curl}(\nabla f) = \mathbf{0},$$

这里 ∇ 是作用在 V 上的梯度算子.

(iii) 对任意的 $F \in V^n$ 有,

$$\mathrm{curl}(\mathrm{curl}F) = \nabla(\mathrm{div}F) - \Delta F,$$

其中 $\Delta F = (\Delta f_1, \Delta f_2, \ldots, \Delta f_n)$, 而 $\Delta f_j = \sum_{i=1}^n \partial_i^2 f_j$ 是 $f_j \in V$ 的拉普拉斯.

证明: 设满足条件 (i)(ii)(iii) 的旋度算子 $\mathrm{curl} : V^n \to V^n$ 具有形式：

$$F = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \mapsto \mathrm{curl}(F) = \begin{pmatrix} \sum_{k,l=1}^n a_1^{k\ell} \partial_k f_\ell \\ \vdots \\ \sum_{k,l=1}^n a_n^{k\ell} \partial_k f_\ell \end{pmatrix},$$

其中 $a_i^{k\ell} \in \mathbb{R}, i, k, \ell = 1, \ldots, n$.

根据条件 (i)，我们有

$$\mathrm{div}(\mathrm{curl}F) = \sum_{i=1}^n \partial_i \left(\sum_{k,\ell=1}^n a_i^{k\ell} \partial_k f_\ell \right) = 0,$$

即

$$\sum_{\ell=1}^n \left(\sum_{i,k=1}^n a_i^{k\ell} \partial_k \partial_i f_\ell \right) = 0.$$

由于各个 f_ℓ 独立, 所以对任意给定的 $\ell=1,\dots,n$ 有

$$\sum_{i,k=1}^{n} a_i^{k\ell}\partial_k\partial_i f_\ell = 0,$$

在上式中令 $f_\ell(x)=x_ix_k$ 有

$$a_i^{k\ell} = -a_k^{i\ell}. \tag{2}$$

再看条件 (ii)，对任意的 $f\in V$ 有,

$$\nabla f = \begin{pmatrix}\partial_1 f\\ \vdots\\ \partial_n f\end{pmatrix} \mapsto \mathrm{curl}(\nabla f) = \begin{pmatrix}\sum_{k,l=1}^n a_1^{k\ell}\partial_k\partial_\ell f\\ \vdots\\ \sum_{k,l=1}^n a_n^{k\ell}\partial_k\partial_\ell f\end{pmatrix} = \begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix},$$

从而对每个 $i=1,\dots,n$ 都有

$$\sum_{k,l=1}^{n} a_i^{k\ell}\partial_k\partial_\ell f = 0,$$

在上式中令 $f(x)=x_\ell x_k$ 有

$$a_i^{k\ell} = -a_i^{\ell k}. \tag{3}$$

现在考虑条件 (iii), 为此我们先计算出

$$\mathrm{curl}(\mathrm{curl}F) = \mathrm{curl}\begin{pmatrix}\sum_{k,l=1}^n a_1^{k\ell}\partial_k f_\ell\\ \vdots\\ \sum_{k,l=1}^n a_n^{k\ell}\partial_k f_\ell\end{pmatrix} = \begin{pmatrix}\sum_{r,s=1}^n a_1^{rs}\partial_r(\sum_{k,\ell=1}^n a_s^{k\ell}\partial_k f_\ell)\\ \vdots\\ \sum_{r,s=1}^n a_1^{rs}\partial_r(\sum_{k,l=1}^n a_s^{k\ell}\partial_k f_\ell)\end{pmatrix} = \begin{pmatrix}\sum_{r,s,k,\ell=1}^n a_1^{rs}a_s^{k\ell}\partial_r\partial_k f_\ell\\ \vdots\\ \sum_{r,s,k,\ell=1}^n a_n^{rs}a_s^{k\ell}\partial_r\partial_k f_\ell\end{pmatrix}$$

而

$$\nabla(\mathrm{div}F)-\Delta F = \begin{pmatrix}\partial_1(\sum_{i=1}^n\partial_i f_i)-\sum_{i=1}^n\partial_i^2 f_1\\ \vdots\\ \partial_n(\sum_{i=1}^n\partial_i f_i)-\sum_{i=1}^n\partial_i^2 f_n\end{pmatrix} = \begin{pmatrix}\sum_{i=1}^n\partial_i\partial_1 f_i-\sum_{i=1}^n\partial_i^2 f_1\\ \vdots\\ \sum_{i=1}^n\partial_i\partial_n f_i-\sum_{i=1}^n\partial_i^2 f_1\end{pmatrix}$$

于是根据条件 (iii) 可知, 对 $j=1,\dots,n$ 有

$$\sum_{r,s,k,\ell=1}^{n} a_j^{rs}a_s^{k\ell}\partial_r\partial_k f_\ell = \sum_{i=1}^{n}\partial_i\partial_j f_i - \sum_{i=1}^{n}\partial_i^2 f_j.$$

注意到各个 f_ℓ 独立, 所以有

$$\sum_{r,s,k=1}^{n} a_j^{rs}a_s^{k\ell}\partial_r\partial_k f_\ell = \begin{cases}(\partial_\ell^2-\sum_{i=1}^n\partial_i^2)f_\ell & \text{若}\ell=j,\\ \partial_j\partial_\ell f_\ell & \text{若}\ell\neq j.\end{cases}$$

令 $f_\ell(x)=x_kx_r$, 就有

$$\sum_{s=1}^{n}a_j^{rs}a_s^{k\ell}+\sum_{s=1}^{n}a_j^{ks}a_s^{r\ell}=\begin{cases}2\delta_{r\ell}\delta_{k\ell}-2\delta_{kr} & \text{若}\ell=j,\\ \delta_{jr}\delta_{k\ell}+\delta_{jk}\delta_{\ell r} & \text{若}\ell\neq j.\end{cases}$$

根据 (3) 与 (2) 有

$$a_s^{k\ell}=-a_s^{\ell k}=a_\ell^{sk},$$

从而上式可以改写为

$$\sum_{s=1}^{n}a_j^{rs}a_\ell^{sk}+\sum_{s=1}^{n}a_j^{ks}a_\ell^{sr}=\begin{cases}2\delta_{r\ell}\delta_{k\ell}-2\delta_{kr} & \text{若}\ell=j,\\ \delta_{jr}\delta_{k\ell}+\delta_{jk}\delta_{\ell r} & \text{若}\ell\neq j.\end{cases}$$

注意到，如果令 A_j 为一个以 a_j^{rs} 为其 r,s 元素的 n 阶矩阵, 则 $\sum_{s=1}^{n}a_j^{rs}a_\ell^{sk}$ 恰好是乘积矩阵 A_jA_ℓ 的第 r,k 元素, 从而上式等价于矩阵等式

$$A_jA_\ell+(A_jA_\ell)^T=\begin{cases}2e_\ell e_\ell^T-2I_n & \text{若}\ell=j,\\ e_je_\ell^T+e_\ell e_j^T & \text{若}\ell\neq j.\end{cases}$$

其中 e_j 是 $\mathbb{R}^n$ 中的第 j 个自然基底 (即仅第 j 个分量取 1, 其它分量均等于 0), I_n 是 n 阶单位矩阵. 注意到由 (3) 可知, A_j 都是反对称矩阵, 因此上式可以化简为

$$A_jA_\ell+A_\ell A_j=\begin{cases}2e_\ell e_\ell^T-2I & \text{若}\ell=j\\ e_je_\ell^T+e_\ell e_j^T & \text{若}\ell\neq j\end{cases}\tag{4}$$

对每个 $j=1,\dots,n$, 构造 $n+1$ 阶矩阵 B_j 如下:

$$B_j=\begin{pmatrix}0 & e_j^T\\ -e_j & A_j\end{pmatrix}\tag{5}$$

下面验证, 这 n 个矩阵 $B_1,\cdots,B_n$ 满足 Hurwitz 矩阵方程 (1) 。由于

$$B_iB_j=\begin{pmatrix}0 & e_i^T\\ -e_i & A_i\end{pmatrix}\begin{pmatrix}0 & e_j^T\\ -e_j & A_j\end{pmatrix}=\begin{pmatrix}-e_i^Te_j & e_i^TA_j\\ -A_ie_j & -e_ie_j^T+A_iA_j\end{pmatrix},$$

从而

$$\begin{aligned}B_iB_j+B_jB_i&=\begin{pmatrix}-e_i^Te_j & e_i^TA_j\\ -A_ie_j & -e_ie_j^T+A_iA_j\end{pmatrix}+\begin{pmatrix}-e_j^Te_i & e_j^TA_i\\ -A_je_i & -e_je_i^T+A_jA_i\end{pmatrix}\\ &=\begin{pmatrix}-(e_i^Te_j+e_j^Te_i) & (e_i^TA_j+e_j^TA_i)\\ -(A_ie_j+A_je_i) & -(e_ie_j^T+e_je_i^T)+(A_iA_j+A_jA_i)\end{pmatrix}.\end{aligned}\tag{6}$$

从而我们的目标是要证明

$$\begin{pmatrix} -(e_i^T e_j + e_j^T e_i) & (e_i^T A_j + e_j^T A_i) \\ -(A_i e_j + A_j e_i) & -(e_i e_j^T + e_j e_i^T) + (A_i A_j + A_j A_i) \end{pmatrix} = \begin{pmatrix} -2\delta_{ij} & \mathbf{0} \\ \mathbf{0} & -2\delta_{ij} I_n \end{pmatrix} \tag{7}$$

容易算出

$$-(e_i^T e_j + e_j^T e_i) = -2\delta_{ij}, \tag{8}$$

而根据 (4) 有

$$-(e_i e_j^T + e_j e_i^T) + (A_i A_j + A_j A_i) = -2\delta_{ij} I_n \tag{9}$$

(8)(9) 两式表明，为证明 (7), 我们只需验证

$$(A_i e_j + A_j e_i) = \mathbf{0}. \tag{10}$$

事实上, 它是 (2)(3) 的推论: 用 e_r^T 左乘向量 $(A_i e_j + A_j e_i)$, 有

$$e_r^T (A_i e_j + A_j e_i) = a_i^{rj} + a_j^{ri} = -a_i^{jr} + a_j^{ir} = 0.$$

由于上式对任意的 $r = 1, \ldots, n$ 成立, 所以 $A_i e_j + A_j e_i = \mathbf{0}$, 即 (10) 成立.

应用定理 3, 就推出 $n + 1 \in \{2, 4, 8\}$, 从而 $n \in \{3, 7\}$. 注意 $n = 1$ 时的旋度算子就是求导算子, 而 $n = 3$ 时即得到经典的旋度算子, 所对应的三个矩阵分别为

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad A_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

至于七维空间中的 7 个反对称矩阵, 我们留给有兴趣的读者来构造. **证毕**。

4.3 与 Taussky–Stiefel 定理的关系

旋度算子与广义 Cauchy–Riemann 算子有密切的关系。

1939 年，Taussky 定义 n 维广义 Cauchy–Riemann 算子如下：由矩阵

$$L = (\ell_{ij}(\partial)), \qquad \text{其中 } \ell_{ij}(\partial) \text{ 是 } \partial_1, \cdots, \partial_n \text{ 的常系数线性组合}, i, j = 1, \ldots, n$$

所定义的作用于 n 维向量函数空间上的一阶微分算子称为 n 维广义 Cauchy–Riemann 算子，如果存在矩阵

$$W = (w_{ij}(\partial)), \qquad \text{其中 } w_{ij}(\partial) \text{ 是 } \partial_1, \cdots, \partial_n \text{ 的常系数线性组合}, i, j = 1, \ldots, n$$

使得 $WL = \Delta I_n$，其中 $\Delta = \sum_{i=1}^{n} \partial_i^2$ 是 Laplace 算子，I_n 是 n 阶单位阵。

例如，经典的 Cauchy–Riemann 算子是

$$L = \begin{pmatrix} \partial_1 & -\partial_2 \\ \partial_2 & \partial_1 \end{pmatrix},$$

注意，由于 $\det(L) = \partial_1^2 + \partial_2^2$，因此符合条件的 M 由

$$M = L^* = \begin{pmatrix} \partial_1 & \partial_2 \\ -\partial_2 & \partial_1 \end{pmatrix}$$

给出。

Taussky[6] 中的主要结果如下：

定理 5 设 $n \geqslant 2$，则 n 维广义 Cauchy–Riemann 算子存在当且仅当 $n = 2, 4, 8$。

Taussky 最初的证明用到下述深刻的拓扑结果：设 $n \geqslant 2$，则 $\mathbb{R}^n$ 可以构成可除代数当且仅当 $n = 2, 4, 8$。当 $n = 4$ 时，著名的 Dirac 算子为广义 Cauchy–Riemann 算子提供了一个光辉的例子。

后来 Stiefel[5] 利用 Hurwitz 矩阵方程化简了 Taussky 的证明，因此定理 5 称作 Taussky–Stiefel 定理。

事实上，上一节的定理 4 是定理 5 的推论。证明如下。

设我们有满足定理 4 条件的旋度算子 curl，现在构造作用于 $n+1$ 维向量函数空间的一阶微分算子 L, W 如下：

$$L = \begin{pmatrix} \partial_0 & \operatorname{div} \\ -\nabla & \partial_0 + \operatorname{curl} \end{pmatrix}, \qquad W = \begin{pmatrix} \partial_0 & -\operatorname{div} \\ \nabla & \partial_0 - \operatorname{curl} \end{pmatrix}$$

其中 $\operatorname{div}, \nabla, \operatorname{curl}$ 均只作用于坐标 $x_1, \ldots, x_n$。则根据 curl 所满足的三条性质 (i)–(iii)，容易验证，有

$$WL = \begin{pmatrix} \partial_0 & -\operatorname{div} \\ \nabla & \partial_0 - \operatorname{curl} \end{pmatrix} \begin{pmatrix} \partial_0 & \operatorname{div} \\ -\nabla & \partial_0 + \operatorname{curl} \end{pmatrix} = \begin{pmatrix} \partial_0^2 + \operatorname{div}\nabla & -\operatorname{div}(\operatorname{curl}) \\ \operatorname{curl}\nabla & \partial_0^2 + (\nabla\operatorname{div} - \operatorname{curl}^2) \end{pmatrix} = \left(\sum_{i=0}^{n} \partial_i^2 \right) I_{n+1}.$$

因此根据定理 5，我们有 $n+1 = 4, 8$，从而 $n = 3, 7$。

4.4 小结与引申

在前两节我们证明了，仅在 3, 7 维空间存在向量积和旋度，我们又知道在三维空间，向量积与旋度之间有密切联系。于是我们提出这样的问题，定理 1 与定

理 4 是否等价？或者它们可否统一？另一方面，我们关于定理 1 与定理 4 的证明都依赖于定理 2 或等价的定理 3，它们表明 2,4,8 是极特殊的数。事实上，在数学中，有些数学对象仅当 $n = 2, 4, 8$ 时存在，实数域上的可除代数 (仅有实数、复数、四元数和八元数) 就是一个典型的例子，更多的例子来自拓扑，见 [1] 中 F. Hirzebruch(1927–2012) 的文章。

参考文献

[1] Heinz-Dieter Ebbinghaus 等编, Numbers, Graduate Texts in Mathematics / Readings in Mathematics v. 123, Springer, 1990.

[2] Beno Eckmann, Stetige Lôsungen linearer Gleichungssysteme, Comm. Math. Helv. 15, 1942/43, 318–339.

[3] 林开亮、陈见柯, Hurwitz 定理的矩阵证明，《高等数学研究》，2018 年第 1 期，24–27.

[4] W. S Massey, Cross products of vectors in higher dimensional Euclidean spaces. *The American Mathematical Monthly*, **90** (1983), 697–701.

[5] E. Stiefel, On Cauchy–Riemann Equations in Higher Dimensions, J. Research Nat. Bur. Standards 48 (1952) 395–398.

[6] O. Taussky, An algebraic property of Laplace's Differential Equation, Quarterly J. Math. (Oxford) 10 (1939) 99–103.

第五章　Hurwitz–Radon 矩阵方程

记号约定

本文中考虑的域 F 的特征不等于 2.

$\mathcal{M}(n,F)$ 表示系数在 F 中的 n 阶方阵的集合, $GL(n,F)$ 表示 $\mathcal{M}(n,F)$ 中的可逆方阵的集合.

$\mathbb{R}$ 和 $\mathbb{C}$ 分别表示实数域和复数域.

A' 表示矩阵 A 的转置. $\mathrm{Tr}(A)$ 表示矩阵 A 的迹.

E 表示单位阵. $i=\sqrt{-1}$ 是虚数单位, 不引起混淆时也作指标使用.

5.1　内容简介

我们将简单介绍一下本文讨论的主要问题和正文的框架结构.

所谓 Hurwitz 矩阵问题, 就是对给定的域 F 以及正整数 n, 求 $\mathcal{M}(n,F)$ 中满足两两反交换并且每个矩阵的平方为 -1 的反对称矩阵集的最大基数. 借助于 Kronecker 符号 δ_{ij}, 这个问题可以表述如下.

设 F 是一个域, 如果 $A_1,\dots,A_k\in\mathcal{M}(n,F)$ 满足以下 Hurwitz 矩阵方程

$$A_iA_j+A_jA_i=-2\delta_{ij};\qquad A_i'=-A_i, \tag{1}$$

则称它们是 $\mathcal{M}(n,F)$ 中的一组 Hurwitz 矩阵.

著名的 Hurwitz 矩阵问题, 就是对给定的 F 以及 n, 求 $\mathcal{M}(n,F)$ 中的一组 Hurwitz 矩阵 $A_1,\dots,A_k$ 的所含的矩阵个数 k 的最大值 $K_F(n)$.

首先我们指出, 这个问题是有意义的. 实际上, 我们有下述

引理 1 *设 $A_1,\dots,A_k\in\mathcal{M}(n,F)$ 满足方程*

$$A_iA_j+A_jA_i=-2\delta_{ij}, \tag{2}$$

则它们是 F-线性无关的.

证明. 假定存在 $c_1, \ldots, c_k \in F$ 使得

$$c_1 A_1 + \cdots + c_k A_k = 0$$

对每一个 $i = 1, \ldots, k$, 等式两边分别左右乘以 A_i, 并将得到的两个式子相加, 利用 (2) 得到

$$c_i = 0.$$

即 $A_1, \ldots, A_k$ 线性无关.

对于 $F = \mathbb{C}$ 及 $F = \mathbb{R}$ 的情形, Hurwitz [6] 和 Radon [14] 分别在 1920 年代前后给出 Hurwitz 矩阵问题的解答, 这就是 Hurwitz–Radon 定理.

定理 1

$$K_{\mathbb{C}}(n) = K_{\mathbb{R}}(n) = K(n) = \begin{cases} 2q & q \equiv 0 \pmod 4 \\ 2q-1 & q \equiv 1 \pmod 4 \\ 2q-1 & q \equiv 2 \pmod 4 \\ 2q+1 & q \equiv 3 \pmod 4 \end{cases}$$

其中 q 满足 $n = 2^q p$, p 是奇数.

这个经典的定理已经有许多证明, 除了 Hurwitz 与 Radon 的原始证明, 还有 Eckmann [2] 的有限群表示论证明, 华罗庚 [4] 的方程链的证明, 李华宗 [**?**] 的利用 Clifford 代数表示论的证明以及 Shapiro [16] 的二次型理论的抽象证明.

Hurwitz 的原始证明的叙述可以参考黄用谳 [21], Radon 的证明的一个转述可见 [7], Eckmann 的证明可见 [2], 李华宗的证明的变体可参看 Lam [9] 与 Prasolov [13].

本文将给出 Hurwitz–Radon 定理的一个比较简单的证明, 这个证明源于 1930 年代 Newman [12] 和 Williamson [20] 的工作, 并且对任意的特征不等于 2 的域都适用. 因为这个证明对 $F = \mathbb{R}$ 为实数域的情形最自然, 所以我们首先考虑这个特殊情况.

遵循 Shapiro [16] 的基本建议, 我们引入一个混合型的 Hurwitz 矩阵方程如

下

$$\begin{cases} A_i' = -A_i \\ B_k' = B_k \\ A_iA_j + A_jA_i = -2\delta_{ij} \qquad (i,j = 1,\dots,s;\ k,l = 1,\dots,t.) \\ B_kB_l + B_lB_k = 2\delta_{kl} \\ A_iB_k + B_kA_i = 0 \end{cases} \tag{3}$$

如果 $A_1,\dots,A_s,B_1,\dots,B_t \in \mathcal{M}(n,F)$ 满足方程 (3), 则称之为 (3) 的一组 (s,t) 型解. 为说话方便, 我们将满足 $A^2 = -E$, $A' = -A$ 的矩阵 A 称为 A 型矩阵, 类似的, 满足 $B^2 = E$, $B' = B$ 的矩阵 B 称为 B 型矩阵. 对混合型方程 (3), 我们引入一个相应的问题: 求 $\mathcal{M}(n,F)$ 中的一组极大的 (s,t) 型解的数偶 (s,t) 的所有可能分布.

通过与 Newman [12] 中的主要结果类比, 我们得到以下定理.

定理 2 记 $n = 2^qp$, 其中 p 是奇数. 则方程 (3) 在 $\mathcal{M}(n,\mathbb{R})$ 中的任意一组解 $A_1,\dots,A_s,B_1,\dots,B_t$ 满足 $s+t \leqslant 2q+1$; 并且, 存在一组使得 $s+t = 2q+1$ 的 (s,t) 型解当且仅当 $t \in [0,2q+1]$ 满足 $t \equiv q+1 \pmod 4$.

这是本文的主要结果, 我们将在 §2 介绍其证明思想 (我们借鉴了 Von Neumann-Veblen [19] 的处理), 详细的证明在 §3 给出. 在 §4 将指出, 由定理 2 可以非常容易地推出定理 1 的 Radon 部分 $K_{\mathbb{R}}(n) = K(n)$, 这就给出了 Radon 定理的一个简单证明. 在 §5 将指出, 定理 2 可以推广到任意的域 (定理 3), 并由此得到 Hurwitz–Radon 定理在任意域上的推广 (定理 4); 并且指出, 我们的方法可以处理更简单的方程 (2), 由此得到 Lam [9] 一书给出的一些结果 (定理 5) 的直接证明. 在 §6 我们将给出定理 1 的一个著名推论 (定理 7) 及其在几何与代数中的两个应用 (定理 6 和定理 8). 在 §7 我们介绍一下 Williamson [20] 对定理 2 的一个有趣推广.

5.2 定理 2 的证明思想

这一节我们假定考虑的矩阵都是实矩阵.

为理解下一节将要给出的定理 2 的证明, 读者需要具备线性代数的一些基本经验, 特别是关于分块矩阵的乘法运算和实对称矩阵的谱定理 *(实对称矩阵正交相似于对角阵)* 和反对称矩阵的谱定理 (见 Kaplansky [8]).

证明定理 2 的基本想法是数学归纳法, 把高阶的情形归结为低阶的情形. 下边的引理提供了约化的可能. 首先注意到, 我们考虑的方程 (3) 在正交相似变换下是不变的: 若

$$A_1, \ldots, A_s, B_1, \ldots, B_t$$

满足方程 (3) , 则对任意的同阶正交矩阵 P,

$$PA_1P^{-1}, \ldots, PA_sP^{-1}, PB_1P^{-1}, \ldots, PB_tP^{-1}$$

也满足 (3).

引理 2 (i) 设两个 n 阶对称矩阵 B_1, B_2 满足

$$B_1^2 = B_2^2 = E, \qquad B_1B_2 = -B_2B_1.$$

则 $n = 2m$ 且存在正交矩阵 P 使得

$$PB_1P^{-1} = \widetilde{B_1} = \begin{bmatrix} E & 0 \\ 0 & -E \end{bmatrix} \qquad \text{且} \qquad PB_2P^{-1} = \widetilde{B_2} = \begin{bmatrix} 0 & E \\ E & 0 \end{bmatrix}.$$

(ii) 设 $A, B \in \mathcal{M}(n, \mathbb{R})$, A 是反对称的, B 是对称的, 并且满足

$$A^2 = -E, \quad B^2 = E, \quad AB = -BA.$$

则 $n = 2m$ 且存在正交矩阵 P 使得

$$PAP^{-1} = \widetilde{A} = \begin{bmatrix} 0 & E \\ -E & 0 \end{bmatrix} \qquad \text{且} \qquad PBP^{-1} = \widetilde{B} = \begin{bmatrix} E & 0 \\ 0 & -E \end{bmatrix}.$$

证明. 对于 (i), 首先我们注意到

$$\mathrm{Tr}(B_1) = 0$$

这是对等式

$$B_1 = -B_2B_1B_2^{-1}$$

两边取迹的结果. 于是根据实对称矩阵的谱定理, 存在正交矩阵 P_1 使得

$$P_1B_1P_1^{-1} = \begin{bmatrix} E & 0 \\ 0 & -E \end{bmatrix} = \widetilde{B_1}$$

这里我们用到 $B_1^2 = E$ 以及 $\mathrm{Tr}(B_1) = 0$ 的事实. 现在 $P_1B_2P_1^{-1}$ 与 $\widetilde{B_1}$ 反交换, 由此容易求得 $P_1B_2P_1^{-1}$ 具有形式

$$P_1B_2P_1^{-1} = \begin{bmatrix} 0 & Y \\ Z & 0 \end{bmatrix}$$

进一步, 从 $(P_1B_2P_1^{-1})^2 = E$ 推出 Y, Z 满足 $YZ = E$. 从而 $P_1B_2P_1^{-1}$ 具有形式

$$P_1B_2P_1^{-1} = \begin{bmatrix} 0 & Y \\ Y^{-1} & 0 \end{bmatrix}$$

其中 Y 是正交矩阵.

我们现在找一个正交矩阵 P_2 使得

$$P_2\widetilde{B_1}P_2^{-1} = \widetilde{B_1}, \quad P_2(P_1B_2P_1^{-1})P_2^{-1} = \widetilde{B_2}.$$

容易求出这样一个矩阵

$$P_2 = \begin{bmatrix} E & 0 \\ 0 & Y \end{bmatrix}$$

从而, 如果我们令 $P = P_2P_1$, 则得到 (i).

对于 (ii), 若令

$$B_1 = B, \quad B_2 = BA,$$

则 B_1, B_2 满足 (i) 的条件, 从而

$$PBP^{-1} = PB_1P^{-1} = \widetilde{B_1} = \widetilde{B}, \quad PB_2P^{-1} = \widetilde{B_2}$$

$$PAP^{-1} = P(B_1B_2)P^{-1} = (PB_1P^{-1})(PB_2P^{-1}) = \widetilde{B_1}\widetilde{B_2} = \widetilde{A}.$$

这就完成了引理 2 的证明.

在展开对定理 2 的证明之前, 我们先给出两点说明.

根据引理 2, 若 (s, t) 型解 $A_1, \ldots, A_s, B_1, \ldots, B_t$ 中有两个矩阵是 B 型的, 不妨设 $B_1 = \widetilde{B_1}, B_2 = \widetilde{B_2}$, 于是与它们同时反交换的矩阵 X 有形式

$$X = \begin{bmatrix} 0 & Y \\ -Y & 0 \end{bmatrix}$$

并且容易验证, X 是 A 型或 B 型的当且仅当对应的 Y 是 B 型或 A 型的. 进一步, 两个这样的矩阵 X_1, X_2 反交换当且仅当与之对应的 Y_1, Y_2 反交换. 于

是, $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型解 $(t \geqslant 2)$ 当且仅当 $M(n/2,\mathbb{R})$ 中存在一组 $(t-2,s)$ 型解.

类似地, 若 $A_1,\dots,A_s,B_1,\dots,B_t$ 中有一个 A 型和一个 B 型的, 则不妨设

$$A_1=\widetilde{A},\quad B_1=\widetilde{B},$$

于是与它们同时反交换的矩阵 Z 具有形式

$$Z=\begin{bmatrix}0 & W\\ W & 0\end{bmatrix},$$

并且 Z 是 A 型或 B 型的当且仅当对应的 W 是 A 型或 B 型的. 进一步, 两个这样的矩阵 Z_1,Z_2 反交换当且仅当对应的 W_1,W_2 反交换. 于是, $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型解 $(s\geqslant 1,t\geqslant 1)$ 当且仅当 $M(n/2,\mathbb{R})$ 中存在一组 $(s-1,t-1)$ 型解.

其次, 我们要介绍由 Newman [12] 引入的一个技巧, 它可以把一组 (s,t) 型解 ($s\geqslant 4$) 转换为一组 $(s-4,t+4)$ 型解. 这个转换如下给出.

设

$$A_1,\dots,A_s,B_1,\dots,B_t$$

是 (3) 的一组解, 其中 $s\geqslant 4$. 令

$$\widetilde{A_i}=A_{i+4},\quad (i=1,\dots,s-4);\qquad \widetilde{B_j}=B_j\quad (j=1,\dots,t),$$

以及

$$\widetilde{B}_{t+1}=A_2A_3A_4,\quad \widetilde{B}_{t+2}=A_1A_3A_4,\quad \widetilde{B}_{t+3}=A_1A_2A_4,\quad \widetilde{B}_{t+4}=A_1A_2A_3$$

容易验证 $\widetilde{A_1},\dots,\widetilde{A}_{s-4},\widetilde{B_1},\dots,\widetilde{B}_{t+4}$ 是 (3) 的一组 $(s-4,t+4)$ 型解. 类似的转换 (A,B 的位置互换) 可以把一组 (s,t) 型解 ($t\geqslant 4$) 转换成一组 $(s+4,t-4)$ 型解. 这个事实我们称之为 Newman 技巧, 它对定理 2 中出现的模 4 条件给出了一个合理的解释, 由此也解释了 Hurwitz–Radon 定理中出现的模 4 条件.

为便于叙述, 我们把上边提到的各个约化结果分别表述成以下 3 个引理.

引理 3 $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型解 $(t\geqslant 2)$ 满足 (3) 当且仅当 $M(n/2,\mathbb{R})$ 中存在一组 $(t-2,s)$ 型解满足 (3).

引理 4 $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型解 $(s\geqslant 1,t\geqslant 1)$ 满足 (3) 当且仅当 $M(n/2,\mathbb{R})$ 存在一组 $(s-1,t-1)$ 型解满足 (3).

引理 5 $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型解 $(s\geqslant 4)$ 满足 (3) 当且仅当存在一组 $(s-4,t+4)$ 型解满足 (3); $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型解 $(t\geqslant 4)$ 满足 (3) 当且仅当存在一组 $(s+4,t-4)$ 型解满足 (3).

5.3 定理 2 的证明

本节我们给出定理 2 的证明.

证明. 对 $n=2^q p$ 的 2 指数 q 用数学归纳法.

第一步. $q=0$ 的情况. 此时 $n=p$ 是奇数.

首先, $s=0$. 否则将存在 p 阶的反对称矩阵 A 满足 $A^2=-E$, 这与奇数阶反对称矩阵不可逆矛盾.

其次, $t\leqslant 1$. 这是引理 2(i) 的结论. 又, 单位阵是 B 型阵, 所以 s 可以取到最大值 1.

于是我们证明了, $q=0$ 时方程 (3) 仅存在 $(0,1)$ 型的极大解.

第二步. $q=1$ 的情况.

我们先证明, 对 $M(2p,\mathbb{R})$ 的任意一组 (s,t) 型解, 必定有 $s\leqslant 1$ 且 $t\leqslant 2$, 从而 $s+t\leqslant 3$.

若 $t\geqslant 3$, 由引理 3, $M(p,\mathbb{R})$ 中存在一组 $(1,0)$ 型解, 这与 $q=0$ 的结果矛盾, 所以 $t\leqslant 2$.

下面证明 $s\leqslant 1$. 用反证法. 假定 $s\geqslant 2$. 由于 $A_1^2=-E$, 根据实反对称矩阵的谱定理, 存在正交矩阵 P 使得

$$PA_1P^{-1}=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix}=\widetilde{A_1}.$$

由于反对称矩阵 PA_2P^{-1} 与 $\widetilde{A_1}$ 反交换, 于是 PA_2P^{-1} 具有形式

$$PA_2P^{-1}=\begin{bmatrix}X & Y\\ Y & -X\end{bmatrix}$$

其中 $X,Y\in M(p,\mathbb{R})$ 皆为反对称矩阵. 如果 A_2 进一步满足 $A_2=-E$, 则 X,Y 将满足

$$X^2+Y^2=-E,\quad XY=YX.$$

这两个式子可以拼成一个紧凑的式子

$$(X+iY)(X-iY)=-E$$

这个式子表明 p 阶 (复) 反对称矩阵 $X+iY$ 与 $X-iY$ 可逆. 矛盾!

另一方面, 根据引理 4, 我们可以构造出一组 $(1,2)$ 型极大解如下:

$$A_1=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix},\qquad B_1=\begin{bmatrix}E & 0\\ 0 & -E\end{bmatrix},\qquad B_2=\begin{bmatrix}0 & E\\ E & 0\end{bmatrix}$$

第三步. 我们将对 $q\geqslant 2$ 归纳证明, $\mathcal{M}(n,\mathbb{R})$ 中任意一组 (s,t) 型解满足 $s+t\leqslant 2q+1$ 而且等号可以成立, 对于 $n=2^q p$.

一方面, 我们利用引理 4 可以从 $M(p,\mathbb{R})$ 中的 $(0,1)$ 型解归纳构造出 $M(2^q p,\mathbb{R})$ 中的一组 $(q,q+1)$ 型解. 例如, $n=2p$ 时我们有上述 $(1,2)$ 型解. 一般的, 设

$$A_1^{(q-1)},\dots,A_{q-1}^{(q-1)},B_1^{(q-1)},\dots,B_q^{(q-1)}$$

为 $M(2^{q-1}p,\mathbb{R})$ 中的一组 $(q-1,q)$ 型解, 则

$$A_i^{(q)}=\begin{bmatrix}0 & A_i^{(q-1)}\\ A_i^{(q-1)} & 0\end{bmatrix}\quad(i=1,\dots,q-1),\qquad A_q^{(q)}=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix},$$
$$B_k^{(q)}=\begin{bmatrix}0 & B_k^{(q-1)}\\ B_k^{(q-1)} & 0\end{bmatrix}\quad(k=1,\dots,q),\qquad B_{q+1}^{(q)}=\begin{bmatrix}E & 0\\ 0 & -E\end{bmatrix}$$

是 $M(2^q p,\mathbb{R})$ 中的一组 $(q,q+1)$ 型解.

下面我们将说明, $\mathcal{M}(n,\mathbb{R})$ 中任何一组 (s,t) 解都满足 $s+t\leqslant 2q+1$.

这是因为 $\mathcal{M}(n,\mathbb{R})$ 中任何一组 (s,t) 解都可以通过引理 3-5 化归为 $M(n/2,\mathbb{R})$ 中的一组个数为 $s+t-2$ 的解 (从而利用归纳假设即可完成证明): 如果 $t\geqslant 2$ 应用引理 3; 如果 $s\geqslant 1,t\geqslant 1$ 应用引理 4; 如果 $t=0$, 应用引理 5.

作为例子, 我们考虑最后一种情况. 用反证法. 假定存在一组 $(s,0)$ 型解使得 $s\geqslant 2q+2$, 由于 $q\geqslant 2$ 所以 $s\geqslant 2\times 2+2=6$, 利用引理 5 可以得到一组 $(s-4,4)$ 型解, 从而化归为第一种情况.

最后, 我们归纳证明, $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型的极大解当且仅当 $t\in[0,2q+1]$ 满足 $t\equiv q+1\pmod 4$.

充分性. 将表明对 $[0,2q+1]$ 中任意的满足 $t\equiv q+1\pmod 4$ 的自然数 t, $\mathcal{M}(n,\mathbb{R})$ 中存在一组 $(2q+1-t,t)$ 型解. 注意到总是存在一组 $(q,q+1)$ 型解, 从这组解出发, 通过若干次 Newman 转换, 可以得到一组 $(2q+1-t,t)$ 型解, 这由同余条件 $t\equiv q+1\pmod 4$ 所保证.

必要性. 设 $n=2^q p$, 其中 $q\geqslant 2$, 如果 $\mathcal{M}(n,\mathbb{R})$ 中存在一组 (s,t) 型的极大解, 根据 Newman 转换, 可以不妨设 $s\geqslant 1,t\geqslant 1$, 于是由引理 4, $M(n/2,\mathbb{R})$ 中存

在一组 $(s-1,t-1)$ 型的极大解, 由 $q-1$ 时的归纳假设, $t-1\equiv q \pmod 4$, 也就是 $t\equiv q+1 \pmod 4$.

5.4 Hurwitz–Radon 定理的证明

这一节我们将从定理 2 推出 Hurwitz–Radon 定理的 Radon 部分.

证明. 首先注意到 (1) 相当于纯 A 型的方程 (3), 根据定理 2, 总有 $K_{\mathbb{R}}(n)\leqslant 2q+1$. 下边我们要确定 $K_{\mathbb{R}}(n)$ 的精确值.

基本的想法是, 对给定的 $n=2^q p$, 从 (3) 的所有可能的 (s,t) 型极大解中选出一组使得 A 型矩阵最多的解, 再看能不能扩充这组解. 注意到, $s+t=2q+1$ 为定值, 所以 s 取得最大等价于 t 取得最小. 下边我们分情况讨论.

(i) 若 $q\equiv 3 \pmod 4$, 则 (3) 的一组 (s,t) 型极大解满足 $t\equiv q+1\equiv 0 \pmod 4$, 取 $t=0$, 此时 $s=2q+1$ 达到最大, 换言之, $K_{\mathbb{R}}(n)=2q+1$.

(ii) 若 $q\equiv 0 \pmod 4$, 则 (3) 的一组 (s,t) 型极大解满足 $t\equiv q+1\equiv 1 \pmod 4$, 取 $t=1$, 此时 $s=2q$, 于是 $K_{\mathbb{R}}(n)\geqslant 2q$. 事实上此时等号成立: $K_{\mathbb{R}}(n)=2q$. 我们用反证法来说明这一点. 假设 $K_{\mathbb{R}}(n)=2q+1$, 这就意味着 (3) 存在一组 $(2q+1,0)$ 型的极大解, 根据定理 2, 这当且仅当 $0\equiv q+1 \pmod 4$, 即 $q\equiv 3 \pmod 4$, 矛盾.

(iii) 若 $q\equiv 1 \pmod 4$, 则 (3) 的一组 (s,t) 型极大解满足 $t\equiv q+1\equiv 2 \pmod 4$, 取 $t=2$, 此时 $s=2q-1$, 于是 $K_{\mathbb{R}}(n)\geqslant 2q-1$. 我们断言此时 $K_{\mathbb{R}}(n)=2q-1$. 用反证法. 假定存在 $2q$ 个矩阵 $A_1,\dots,A_{2q}$ 满足 (1), 则令

$$A_{2q+1}=A_1\cdots A_{2q},$$

则容易看到 A_{2q+1} 是一个 A 型矩阵, 且与 $A_1,\dots,A_{2q}$ 反交换, 于是我们得到 (3) 的一组 $(2q+1,0)$ 型的极大解, 根据定理 2, 这当且仅当 $q\equiv 3 \pmod 4$, 这与条件矛盾.

(iv) 若 $q\equiv 2 \pmod 4$, 则 (3) 的一组 (s,t) 型极大解满足 $t\equiv q+1\equiv 3 \pmod 4$, 取 $t=3$, 此时 $s=2q-2$, 于是 $K_{\mathbb{R}}(n)\geqslant 2q-2$. 设 $A_1,\dots,A_{2q-2}$ 满足 (1), 则我们可以添加矩阵

$$A_{2q-1}=A_1\cdots A_{2q-2}$$

得到一组满足 (1) 的 $2q-1$ 个矩阵. 于是进一步有 $K_{\mathbb{R}}(n) \geqslant 2q-1$, 我们断言 $K_{\mathbb{R}}(n) = 2q-1$. 用反证法来说明这一点, 假设 $A_1, \ldots, A_{2q}$ 满足 (1), 则

$$B_1 = A_1 \cdots A_{2q}$$

与 $A_1, \ldots, A_{2q}$ 一起给出 (3) 的一组 $(2q, 1)$ 型的极大解, 根据定理 2, 这当且仅当 $1 \equiv q+1 \pmod 4$, 即 $q \equiv 0 \pmod 4$, 矛盾.

综上, 我们确定出 $K_{\mathbb{R}}(n)$ 的值如下:

$$K_{\mathbb{R}}(n) = K(n) = \begin{cases} 2q & q \equiv 0 \pmod 4 \\ 2q-1 & q \equiv 1 \pmod 4 \\ 2q-1 & q \equiv 2 \pmod 4 \\ 2q+1 & q \equiv 3 \pmod 4 \end{cases}$$

推论 1 $K_{\mathbb{R}}(n) \leqslant n-1$, 等号成立当且仅当 $n = 1, 3, 7$.

因为这个推论非常重要, 我们将给出一个独立的证明, 该证明源于 Hurwitz [5].Hurwitz 用另一种等价的方式来描述这个结果, 他的表述如下:

定理 3 (Hurwitz, 1898) 设 $n \geqslant 1$. 若存在 $x_1, \ldots, x_n$ 与 $y_1, \ldots, y_n$ 的实系数双线性函数 $z_1, \ldots, z_n$ 使得

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = (z_1^2 + \cdots + z_n^2) \tag{5.1}$$

对一切复变量 $x_1, \ldots, x_n$ 与 $y_1, \ldots, y_n$ 成立, 则 $n = 1, 2, 4, 8$.

Hurwitz 正是在研究这种平方和的乘积问题时导出著名的 Hurwitz 矩阵方程. 与平方和有关的历史可以参考 Taussky [17] 的精彩论述.

5.5 Newman 定理与 Hurwitz 定理到任意域的推广

本节我们将说明, 定理 1 与定理 2 都可以推广到任意的特征不等于 2 的域.

如上一节所看到的, 从定理 2 推导定理 1 只涉及到纯粹的加减乘除四则基本运算, 而完全不需要用到关于域 F 的任何特性, 所以我们只需证明定理 2 对任意的域成立.

为此我们需要把引理 2-5 推广到任意的域上, 其中引理 2 的推广是最基本的, 一旦引理 2 得证, 引理 3-5 将水到渠成.

我们首先考虑 F 为代数封闭域的情况. 此时, 下述矩阵定理——似乎首先为 Albert 发现, 见 [1] 定理 27——是关键的 (证明可见 Kaplansky [8]).

引理 6 设 F 是特征不等于 2 的代数封闭域. $A_1,\dots,A_k$ 与 $B_1,\dots,B_k$ 同是 F 上的 n 阶对称矩阵或反对称矩阵, 若存在矩阵 $P\in GL(n,F)$ 使得

$$PA_iP^{-1}=B_i,\quad (i=1,\dots,k).$$

则存在矩阵 $Q\in GL(n,F)$, $QQ'=E$ 使得

$$QA_iQ^{-1}=B_i,\quad (i=1,\dots,k).$$

为利用引理 6 推导代数闭域情形的引理 2, 我们需要下述更一般的结果.

引理 7 (i) 设 $B_1,B_2\in\mathcal{M}(n,F)$ 满足

$$B_1^2=B_2^2=E,\qquad B_1B_2=-B_2B_1.$$

则 $n=2m$ 且存在 $P\in GL(n,F)$ 使得

$$PB_1P^{-1}=\widetilde{B_1}=\begin{bmatrix}E & 0\\ 0 & -E\end{bmatrix}\qquad 且\qquad PB_2P^{-1}=\widetilde{B_2}=\begin{bmatrix}0 & E\\ E & 0\end{bmatrix}.$$

(ii) 设 $A,B\in\mathcal{M}(n,F)$ 满足

$$A^2=-E,\quad B^2=E,\quad AB=-BA$$

则 $n=2m$ 且 $P\in GL(n,F)$ 使得

$$PAP^{-1}=\widetilde{A}=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix}\qquad 且\qquad PBP^{-1}=\widetilde{B}=\begin{bmatrix}E & 0\\ 0 & -E\end{bmatrix}.$$

引理 7 的证明几乎平行于引理 2 的证明, 我们留给感兴趣的读者.

从引理 6 和引理 7 可以很容易推出下述引理 8, 它就是引理 2 在代数封闭域上的对应结果.

引理 8 设 F 是一个特征不等于 2 的代数闭域.

(i) 设 $B_1, B_2 \in \mathcal{M}(n, F)$ 都是对称矩阵且满足

$$B_1^2 = B_2^2 = E, \qquad B_1 B_2 = -B_2 B_1.$$

则 $n = 2m$ 且存在 $Q \in GL(n, F)$, $QQ' = E$ 使得

$$QB_1Q^{-1} = \widetilde{B_1} = \begin{bmatrix} E & 0 \\ 0 & -E \end{bmatrix} \qquad 且 \qquad QB_2Q^{-1} = \widetilde{B_2} = \begin{bmatrix} 0 & E \\ E & 0 \end{bmatrix}.$$

(ii) 设 $A, B \in \mathcal{M}(n, F)$, A 是反对称的, B 是对称的, 并且满足

$$A^2 = -E, \quad B^2 = E, \quad AB = -BA$$

则 $n = 2m$ 且存在 $Q \in GL(n, F)$, $QQ' = E$ 使得

$$QAQ^{-1} = \widetilde{A} = \begin{bmatrix} 0 & E \\ -E & 0 \end{bmatrix} \qquad 且 \qquad QBQ^{-1} = \widetilde{B} = \begin{bmatrix} E & 0 \\ 0 & -E \end{bmatrix}.$$

由此可以证明引理 3-5 在代数闭域上成立, 从而得到定理 2 在代数封闭域上的推广.

现在考虑一般的情况. 假定 F 是任意一个特征不等于 2 的域, $\overline{F}$ 为其代数闭包. 设 $A_1, \ldots, A_s, B_1, \ldots, B_t$ 是 $\mathcal{M}(n, F)$ 中的一组 (s, t) 型解, 它自然是 $M(n, \overline{F})$ 中的一组 (s, t) 型解, 所以根据代数封闭域 $\overline{F}$ 上的定理 2, 可以推出 $s+t \leqslant 2q+1$, 而且如果 $s + t = 2q + 1$, 则 $t \equiv q + 1 \pmod 4$. 为证明当 $t \in [0, 2q + 1]$ 满足同余条件 $t \equiv q + 1 \pmod 4$ 时 $\mathcal{M}(n, F)$ 存在一组 $(2q + 1 - t, t)$ 型解, 只注意定理 2 的证明中给出的那组 $(q, q + 1)$ 型极大解矩阵 (4) 本质上是由 $0, 1, -1$ 三个数经加减乘除生成的, 所以在任何一个特征不等于 2 的域 F 上都有定义. 这就证明了当 $t = q + 1$ 时存在一组 $(q, q + 1)$ 型解, 再注意到 Newman 转换不依赖于域 F 的任何特性, 所以由此可以衍生出所有满足同余条件 $t \equiv q + 1 \pmod 4$ 的 $(2q + 1 - t, t)$ 型解. 这样我们就得到了定理 2 的下述推广:

定理 4 设域 F 的特征不等于 2. 记 $n = 2^q p$ 其中 p 是奇数. 则方程 (3) 在 $\mathcal{M}(n, F)$ 中的任意一组解 $A_1, \ldots, A_s, B_1, \ldots, B_t$ 满足 $s + t \leqslant 2q + 1$; 并且, 存在一组使得 $s + t = 2q + 1$ 的 (s, t) 型解当且仅当 $t \in [0, 2q + 1]$ 满足 $t \equiv q + 1 \pmod 4$.

由此, 我们得到 Hurwitz–Radon 定理的最一般的版本. [1]

[1] 很抱歉, 作者尚不清楚这个一般结果最初由谁发现.

定理 5 设 F 是一个特征不等于 2 的域, 若 $A_1, \ldots, A_k \in \mathcal{M}(n, F)$ 满足方程 (1), 则 k 有最大值, 并且其最大值 $K_F(n)$ 如下给出:

$$K_F(n) = K(n) = \begin{cases} 2q & q \equiv 0 \pmod 4 \\ 2q-1 & q \equiv 1 \pmod 4 \\ 2q-1 & q \equiv 2 \pmod 4 \\ 2q+1 & q \equiv 3 \pmod 4 \end{cases}$$

其中 q 满足 $n = 2^q p$, p 是奇数.

最后, 作者想就本文提供的这个证明思想做一个小结: 我们没有直接考虑方程 (1) 而是考虑了混合型方程 (3), 然后从中提取出 (1) 的信息. 我们得到的教益是: 如果一个对象有对偶, 那么连同它的对偶一起来考虑会看得更清楚. 这就好比粒子物理中的 Bose-Fermi 对应的哲学一样. 同样的想法可以应用于方程 (2), 对应的结果由下述定理给出.

定理 6 设 F 是一个特征不等于 2 的域, 记 $n = 2^q p$, p 为奇数. 若 $A_1, \ldots, A_k \in \mathcal{M}(n, F)$ 满足 (2), 则 k 有最大值 $E_F(n)$, 且 $E_F(n)$ 如下给出:

(i) 若 -1 是 F 中的平方数, 则 $E_F(n) = 2q + 1$;

(ii) 若 -1 不是 F 中的平方数, 但可以写成 F 中两个平方数的和, 则

$$E_F(n) = \begin{cases} 2q & q \equiv 0 \pmod 2 \\ 2q+1 & q \equiv 1 \pmod 2 \end{cases};$$

(iii) 若 -1 不能写成 F 中两个平方数的和, 则

$$E_F(n) = K(n) = \begin{cases} 2q & q \equiv 0 \pmod 4 \\ 2q-1 & q \equiv 1 \pmod 4 \\ 2q-1 & q \equiv 2 \pmod 4 \\ 2q+1 & q \equiv 3 \pmod 4 \end{cases}$$

定理 5 本质上就是 Lam [9, pp.125–126] 对应的定理 4.4, 4.6, 4.8, 那里是作为 Clifford 代数理论的副产品给出的.

5.6 Hurwitz–Radon 定理的一些应用

在本节我们将介绍 Hurwitz–Radon 定理在几何和代数方面的两个应用. 应该指出, Hurwitz 矩阵方程与代数和几何中的许多问题关联密切, 有兴趣的读者可以参考 Eckmann [3]

特别值得介绍的是 1983 年 Massey [11] 给出的下述关于欧氏空间的向量积的基本结果:

定理 7 设 $n \geqslant 2$ 维欧氏空间 $\mathbb{R}^n$ 中定义了一个双线性的向量积 $V \times V \to V$: $(x, y) \mapsto x \times y$ 满足以下条件:

(i) $\langle x \times y, x\rangle = \langle x \times y, y\rangle = 0.$
(ii) $\|x \times y\|^2 = \|x\|^2\|y\|^2 - \langle x, y\rangle^2.$

则 $n = 3$ 或 $n = 7.$

定理 7 可以从定理 6 得到, 具体推导可见 Prasolov [13].

致谢

作者在本文写作过程中从清华大学刘云朋同学处得到许多教益; 还要感谢天津大学田代军老师、田长亮、郑景锐、徐泽、张雅轩同学以及南开大学白承铭老师、刘会同学与首都师范大学方复全老师、李克正老师、费少明老师和许权、陈见柯、段红伟、赵洁等同学的鼓励和帮助.

参考文献

[1] A. Albert, Symmetric and alternate matrices in an arbitrary fields, *Trans. A. M.S.*, **43**(1938), 193–228.

[2] B.Eckmann, Gruppentheorectischer Beweis des Satzes von Hurwitz–Radon uber die Komposition quadratischer Formen, *Comment.Math.Helv.*,**15**(1942), 358–366.

[3] B.Eckmann, Topology, Algebra, Analysis–Relations and missing links, *Notices A.M.S.*, **46**(1999), 520–527.

[4] Loo-Keng Hua, Geometries of Matrices II. Study of involutions in the geometry of symmetric matrices, *Trans.A.M.S.***61**(1947), 193–228.

[5] A.Hurwitz, quadratischen Formen von beliebig vielen Variablen, *Nachrichten Ges. der Wiss. Gottingen* 1898, 309–316.

[6] A.Hurwitz, *Ü*ber die Komposition der quadratischen Formen, *Math.Ann.* **88**(1923), 1–25.

[7] 江上鷗, Hurwitz–Radon 问题——等价和约化是处理数学问题的一种基本方法, 《数学通报》, 1964 年 04 期.

[8] I.Kaplansky, Linear Algabra and Geometry–A Second Course, Allyn and Bacon, 1969.

[9] T.Y.Lam, Introduction to Quadratic Forms over Fields, Graduate Studies in Mathematics, Vol.**67**, American Mathematical Society, Providence, Rhode Island, 2005.

[10] H.C.Lee, Sur le theoreme de Hurwitz–Radon pour la composition des formes quadratiques, *Comment.Math.Helv.***21)**(1948),261–269.

[11] M.S.Massey, Cross products of vectors in higher-dimensional Euclidean spaces. *Amer. Math. Monthly* **90 (no.10)**(1983), 697–701.

[12] M.A.H.Newman, Note on an algebraic theorem of Eddington, *Journal London Math.Soc.***7**(1932),93–99; Corrigenda, 272.

[13] V.V.Prasolov, Problems and Theorems in Linear Algebra, Translation of Mathematical monographs, Vol.134, American Mathematical Society, 1994.

[14] J.Radon, Lineare Scharen orthogonaler Matrizen, *Abhandlungen aus dem mathematischen Seminar der Hambergischen Univerdität***1**(1922), 1–14.

[15] J.P.Serre, Linear Representations of Finite Groups, GTM Vol.42, Springer.

[16] D.B.Shapiro, Compositions of Quadratic Forms, de Gruyter Expositions in Mathematics **33**, 2000.

[17] O.Taussky, Sums of Squares,*The American Mathematical Monthly*, **77 No.8**,(1970), 805-830.

[18] J.A.Tyrrell and J.G.Semple, Generalized Clifford Parallelism, Cambridge University Press, 1971.

[19] O.Veblen and J.Von Neumann, Geometry of Complex Domain, Princeton Mimeographed Notes, Notes by W.Givens and A.H.Taub, Institute for Advanced Study, 1935–1936.

[20] J. Williamson, Sets of semi-commutative matrices, I.*Proc.Edinburgh Math. Soc. (Series 2)***3** (1933): 179-188. II.*Proc. Edinburgh Math. Soc.(Series 2)***3** (1933): 231-240.

[21] Yung-Chow Wong, Isoclinic n-Planes in $2n$-Spaces, Clifford Parallels in Elliptic $(2n-1)$-Spaces, and Hurwitz Matrix Equations, *Memoirs A.M.S.***no.41**, 1961 (second printing with corrections and minor changes, 1971.)

第六章　Hurwitz–Radon 矩阵方程的华罗庚链

6.1　记号约定

$M(n,F)$ 表示系数在域 F 中的 n 阶方阵的集合.

$GL(n,F)$ 表示 $M(n,F)$ 中的可逆方阵的集合.

对于 $A \in M(n,\mathbb{F})$, A^T 表示其转置. 对于 $A \in M(n,\mathbb{C})$, $A^* = (\overline{A})^T$ 表示其共轭转置.

E 表示单位阵.

记

$$O(n,F) = \{A \in M(n,F) \,|\, A^T A = E\}$$

称为 F 上的 n 级正交群. 矩阵 A 称为正交阵. 当 $F = \mathbb{R}$ 时, $O(n,\mathbb{R})$ 通常简记为 $O(n)$.

$\mathbb{C}$ 上的 n 级酉群定义如下

$$U(n) = \{T \in M(n,\mathbb{C}) \,|\, T^* T = E\}$$

令

$$J = \begin{bmatrix} 0 & E \\ -E & 0 \end{bmatrix}.$$

记

$$Sp(n,F) = \{P \in M(n,F) \,|\, P^T J P = J\}$$

称为 F 上的 n 阶辛群. P 称为 n 级辛矩阵. 注意到 J 本身是一个辛矩阵, 称为标准辛矩阵.

令为 n 级酉辛群为

$$Sp(n) = Sp(n, \mathbb{C}) \cap U(n)$$

$i = \sqrt{-1}$ 是虚数单位, 不引起混淆时也作指标使用.

6.2 内容简介

设 F 是一个域, 如果 $A_1, \cdots, A_k \in M(n, F)$ 满足以下 Hurwitz 矩阵方程

$$A_i A_j + A_j A_i = -2\delta_{ij} E; \qquad A_i^T = -A_i \tag{6.1}$$

则称它们是 $M(n, F)$ 中的一组 Hurwitz 矩阵. 这里 δ_{ij} 为 Kronecker 符号, 当 $i = j$ 时取值 1, 否则取值 0.

著名的 Hurwitz 矩阵问题, 就是对给定的 F 以及任意的正整数 n, 求 $M(n, F)$ 中的一组 Hurwitz 矩阵 $A_1, \cdots, A_k$ 的所含的矩阵个数 k 的最大值 $K_F(n)$.

有必要首先指出, 这个问题是有意义的. 实际上, 容易证明 (例如, 见 [9] 引理 1), $M(n, F)$ 中的任意一组 Hurwitz 矩阵是线性无关的, 从而 $K_F(n) \leqslant n^2$.

对于 $F = \mathbb{C}$ 及 $\mathbb{R}$ 的情形, Hurwitz [6] 和 Radon [13] 分别在 1920 年前后给出 Hurwitz 矩阵问题的解答, 这就是 Hurwitz-Radon 定理.

定理 1

$$K_{\mathbb{C}}(n) = K_{\mathbb{R}}(n) = \begin{cases} 2q & q \equiv 0 \pmod 4 \\ 2q-1 & q \equiv 1 \pmod 4 \\ 2q-1 & q \equiv 2 \pmod 4 \\ 2q+1 & q \equiv 3 \pmod 4 \end{cases}$$

其中 q 满足 $n = 2^q p$, p 是奇数.

这个经典的定理已经有许多证明, 除了 Hurwitz 和 Radon 的原始证明 (可分别参见 [15] 与 [7]) 以外, 还有 Eckmann [2] 的有限群表示论证明, 李华宗的利用 Clifford 代数表示论的证明 (参看 [12]).

作者之一曾在 [9] 中给出了 Hurwitz-Radon 定理的一个简单证明, 那个证明很好地解释了定理中的模 4 周期性, 并且可以推广到任意的特征不等于 2 的域 F.

本文将介绍 Hurwitz-Radon 定理的另一个比较简单的证明, 它基于华罗庚 1947 年的工作 [5], 但由于某些原因这个优美的证明被忽略了.

6.3 华罗庚链

对于 $F=\mathbb{C}$ 的情形, 华罗庚 [4] 对 Hurwitz 定理曾经给出一个漂亮的证明. 我们介绍如下.

从这一节开始, 如果不作特殊说明, 将假定所有谈及的矩阵都是复矩阵.

华罗庚在 1947 年发表的矩阵几何的文章 [4] 中独立地得到了 Hurwitz 矩阵方程, 并将它化归为其他两个平行的方程, 但由于运算疏忽, 导致他得出了错误的结论. 但他的思路是对的, 遵循其方法, 我们可得到下述结果, 我们称之为华罗庚链 (参见 [10]). 叙述如下:

定理 2: 设 (1) 和下述三个方程

$$B_iB_j+B_jB_i=2\delta_{ij}E, \qquad B_i^T=B_i \tag{6.2}$$

$$C_iC_j+C_jC_i=-2\delta_{ij}E, \qquad C_i^TJC_i=J \tag{6.3}$$

$$D_iD_j+D_jD_i=2\delta_{ij}E, \qquad D_i^TJD_i=J \tag{6.4}$$

在 $M(n,\mathbb{C})$ 中的极大解个数分别为 $K(n),\Sigma(n),P(n),X(n)$, 则有

$$K(n)=X(n/2)+2 \tag{6.5}$$

$$X(n)=P(n/2)+2 \tag{6.6}$$

$$P(n)=\Sigma(n/2)+2 \tag{6.7}$$

$$\Sigma(n)=K(n/2)+2 \tag{6.8}$$

其中 (5)(6) 中要求 n 被 4 整除, (7)(8) 中要求 n 为偶数. 并且, 对于奇数 p, 我们有以下初始值:

$$K(2p)=1, \quad X(2p)=1, \quad P(2p)=3, \quad \Sigma(2p)=2. \tag{6.9}$$

为方便叙述, 我们称满足 $A^2=-E$ 的反对称矩阵 A 为 A 型矩阵, 满足 $B^2=E$ 的对称矩阵 B 称为 B 型矩阵, 满足 $C^2=-E$ 的辛矩阵 C 为 C 型矩阵, 满足 $D^2=E$ 的辛矩阵 D 称为 D 型矩阵.

我们把上述定理的内容分解为以下四个引理.

引理 1: 设两个 n 阶反对称矩阵 A_1,A_2 满足

$$A_1^2=A_2^2=-E, \qquad A_1A_2=-A_2A_1.$$

则 $n=4l$ 且存在 $P\in O(n,\mathbb{C})$ 使得

$$PA_1P^{-1}=\widetilde{A_1}=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix}\qquad 且\qquad PA_2P^{-1}=\widetilde{A_2}=\begin{bmatrix}J & 0\\ 0 & -J\end{bmatrix}$$

与 $\widetilde{A_1},\widetilde{A_2}$ 同时反交换的 A 型矩阵 A 有形式

$$A=\begin{bmatrix}X & Y\\ Y & -X\end{bmatrix}\tag{6.10}$$

其中 X,Y 是反对称矩阵, 并且满足

$$X^2+Y^2=-E,\quad XY=YX,\quad JX=-XJ,\quad JY=YJ$$

对于满足上述关系的一对 X,Y, 令

$$D=J(iX-Y)\tag{6.11}$$

则 D 是 D 型矩阵. 进一步, 两个如上形式的 A 型矩阵 A_j,A_k 反交换当且仅当对应的 D 型矩阵 D_j,D_k 反交换. 于是 (5) 成立. 对奇数 p, 有 $K(2p)=1$.

引理 2: 设两个 n 阶辛矩阵 D_1,D_2 满足

$$D_1^2=D_2^2=E,\qquad D_1D_2=-D_2D_1.$$

则 $n=4l$ 且存在 $P\in Sp(n,\mathbb{C})$ 使得

$$PD_1P^{-1}=\widetilde{D_1}=\begin{bmatrix}0 & iJ\\ iJ & 0\end{bmatrix}\qquad 且\qquad PD_2P^{-1}=\widetilde{D_2}=\begin{bmatrix}0 & -J\\ J & 0\end{bmatrix}$$

与 $\widetilde{D_1},\widetilde{D_2}$ 同时反交换的 D 型矩阵 D 有形式

$$D=\begin{bmatrix}iC & 0\\ 0 & iJCJ\end{bmatrix}$$

其中 C 是 C 型矩阵. 进一步, 两个如上形状的 D 型矩阵 D_j,D_k 反交换当且仅当对应的 C 型矩阵 C_j,C_k 反交换. 于是 (6) 成立. 对奇数 p, 有 $X(2p)=1$.

引理 3: 设两个 n 阶辛矩阵 C_1,C_2 满足

$$C_1^2=C_2^2=-E,\qquad C_1C_2=-C_2C_1.$$

则 $n=2m$ 且存在 $P\in Sp(n,\mathbb{C})$ 使得

$$PC_1P^{-1}=\widetilde{C_1}=\begin{bmatrix}iE & 0\\ 0 & -iE\end{bmatrix}\qquad 且\qquad PC_2P^{-1}=\widetilde{C_2}=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix}$$

与 $\widetilde{C_1},\widetilde{C_2}$ 同时反交换的 C 型矩阵 C 有形式

$$C=\begin{bmatrix}0 & iB\\ iB & 0\end{bmatrix}$$

其中 B 是 B 型矩阵. 进一步, 两个如上形状的 C 型矩阵 C_j,C_k 反交换当且仅当对应的 B 型矩阵 B_j,B_k 反交换. 于是 (7) 成立. 对奇数 p, 有 $P(2p)=3$.

引理 4: 设两个 n 阶对称矩阵 B_1,B_2 满足

$$B_1^2=B_2^2=E,\quad B_1B_2=-B_2B_1.$$

则 $n=2m$ 且存在 $P\in O(n,\mathbb{C})$ 使得

$$PB_1P^{-1}=\widetilde{B_1}=\begin{bmatrix}E & 0\\ 0 & -E\end{bmatrix}\qquad 且\qquad PB_1P^{-1}=\widetilde{B_2}=\begin{bmatrix}0 & E\\ E & 0\end{bmatrix}$$

与 $\widetilde{B_1},\widetilde{B_2}$ 同时反交换的 B 型矩阵 B 有形式

$$B=\begin{bmatrix}0 & A\\ -A & 0\end{bmatrix}$$

其中 A 是 A 型矩阵. 进一步, 两个如上形状的 B 型矩阵 B_j,B_k 反交换当且仅当对应的 A 型矩阵 A_j,A_k 反交换. 于是 (8) 成立. 对奇数 p, 有 $\Sigma(2p)=2$.

6.4 华罗庚链的证明

我们将给出引理 1 的详细证明, 引理 $2-4$ 的证明相对简单, 作为例子, 我们勾勒出引理 2 的证明.

在展开证明之前, 我们需要两个矩阵引理. 注意到方程 (1), (2) 在正交相似下不变, (3), (4) 在辛相似变换下不变.

引理 5:

(i) 设 $\{A_\alpha\}$, $\{B_\alpha\}$ 是 $M(n,\mathbb{C})$ 中的对称矩阵集或反对称矩阵集. 若存在 $P\in GL(n,\mathbb{C})$ 使得

$$PA_\alpha P^{-1}=B_\alpha,\qquad \forall\alpha.$$

则存在 $Q \in O(n,\mathbb{C})$ 使得

$$QA_\alpha Q^{-1} = B_\alpha, \qquad \forall\alpha.$$

(ii) 设 $\{A_\alpha\}, \{B_\alpha\}$ 是 $M(n,\mathbb{C})$ 中的辛矩阵集. 若存在 $P \in GL(n,\mathbb{C})$ 使得

$$PA_\alpha P^{-1} = B_\alpha, \qquad \forall\alpha.$$

则存在 $Q \in Sp(n,\mathbb{C})$ 使得

$$QA_\alpha Q^{-1} = B_\alpha, \qquad \forall\alpha.$$

对于单个 (反) 对称矩阵的情形, 这是 Kaplansky [8] 中的标准定理, 对于单个辛矩阵的情况, 见该书 p.81 练习 4. 容易看到, 那里的证明对矩阵集也成立, 我们这里给出 (i) 的证明.

证明: 只证必要性. 设 A_α, B_α 是 $\mathbb{C}$ 上的同阶对称矩阵, 且存在矩阵 P 使得 $P^{-1}A_\alpha P = B_\alpha$. 设 P 的代数极分解 (见 [4]) 为 $P = ST$, 则 S 为复对称矩阵, T 为复正交矩阵, 而且 S 是 P 的多项式. 我们将证明 $T^{-1}A_\alpha T = B_\alpha$. 首先, 由 $PA_\alpha P^{-1} = B_\alpha$ 以及 $PA_\alpha^T P = B_\alpha^T$ 有

$$A_\alpha P = PB_\alpha, \qquad A_\alpha^T P = PB_\alpha^T$$

由第二个等式有 $P^T A_\alpha = B_\alpha P^T$, 结合第一个等式有

$$(PP^T)A_\alpha = P(P^T A_\alpha) = P(B_\alpha P^T) = (PB_\alpha)P^T = (A_\alpha P)P^T = A_\alpha(PP^T)$$

即 PP^T 与 A_α 交换. 于是推出 S 与 A_α 交换, 从而

$$T^{-1}A_\alpha T = P^{-1}SA_\alpha S^{-1}P = P^{-1}A_\alpha SS^{-1}P = P^{-1}A_\alpha P = B_\alpha$$

证毕.

下边的引理对一般的域 F 也成立, 见 [9] 中引理 7, 此处略去证明.

引理 6 设 $B_1, B_2 \in M(n,\mathbb{C})$ 满足

$$B_1^2 = B_2^2 = E, \qquad B_1B_2 = -B_2B_1.$$

则 $n = 2m$ 且存在 $P \in GL(n,\mathbb{C})$ 使得

$$PB_1P^{-1} = \widetilde{B_1} = \begin{bmatrix} E & 0 \\ 0 & -E \end{bmatrix} \qquad \text{且} \qquad PB_2P^{-1} = \widetilde{B_2} = \begin{bmatrix} 0 & E \\ E & 0 \end{bmatrix}.$$

引理 1 的证明: 我们首先证明 $K(2p)=1$, 对奇数 p. 用反证法. 假定 $K(2p)\geqslant 2$, 设 $A_1,A_2\in GL(2p,\mathbb{C})$ 满足 $A_1^2=A_2^2=_E$ 且 $A_1A_2=-A_2A_1$. 由于 $A_1^2=-E$, 容易推出 存在矩阵 $P\in GL(2p,\mathbb{C})$ 使得

$$PA_1P^{-1}=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix}=\widetilde{A_1}.$$

由引理 5 (i), 存在矩阵 $Q\in O(2p\mathbb{C})$ 使得 $QA_1Q^{-1}=\widetilde{A_1}$. 由于反对称矩阵 QA_2Q^{-1} 与 $\widetilde{A_1}$ 反交换, 于是 QA_2Q^{-1} 具有形式

$$QA_2Q^{-1}=\begin{bmatrix}X & Y\\ Y & -X\end{bmatrix}$$

其中 $X,Y\in M(p,\mathbb{C})$ 皆为反对称矩阵. 由 $A_2=-E$ 推出 X,Y 满足

$$X^2+Y^2=-E,\quad XY=YX.$$

这两个式子可以拼成一个紧凑的式子

$$(X+iY)(X-iY)=-E$$

这个式子表明 p 阶反对称矩阵 $X+iY$ 与 $X-iY$ 可逆. 矛盾!

由此我们证明了 $K(2p)=1$. 又根据引理的条件, 必定有 $n=4l$, 且

令 $B_1=-iA_1$, $B_2=-iA_2$. 于是, B_1,B_2 满足引理 6 的条件, 从而存在 $P_1\in GL(n,\mathbb{C})$ 使得

$$P_1A_1P_1^{-1}=i\widetilde{B_1}=\begin{bmatrix}iE & 0\\ 0 & -iE\end{bmatrix}\qquad 且\qquad P_1A_2P_1^{-1}=i\widetilde{B_2}=\begin{bmatrix}0 & iE\\ iE & 0\end{bmatrix}.$$

类似的, 注意到 $\widetilde{A_1},\widetilde{A_2}$ 满足引理 1 的条件, 所以如上一样可以推出, 存在 $P_2\in GL(n,\mathbb{C})$ 使得

$$P_2\widetilde{A_1}P_2^{-1}=\begin{bmatrix}iE & 0\\ 0 & -iE\end{bmatrix}\qquad 且\qquad P_2\widetilde{A_2}P_2^{-1}=\begin{bmatrix}0 & iE\\ iE & 0\end{bmatrix}.$$

由此看出, $\{A_1,A_2\}$ 与 $\{\widetilde{A_1},\widetilde{A_2}\}$ 满足引理 5 (i) 的条件, 于是 A_1,A_2 同时正交相似于 $\widetilde{A_1},\widetilde{A_2}$, 不妨设 $A_1=\widetilde{A_1},A_2=\widetilde{A_2}$. 容易求出与 $\widetilde{A_1},\widetilde{A_2}$ 同时反交换的反对称矩阵 A 具有形式

$$A=\begin{bmatrix}X & Y\\ Y & -X\end{bmatrix}$$

其中 X,Y 是反对称矩阵, 并且满足

$$JX=-XJ,\quad JY=YJ$$

如果 A 进一步满足 $A^2=-E$, 则 X,Y 进一步满足

$$X^2+Y^2=-E,\quad XY=YX.$$

由此推出 $(iX+Y)=(iX-Y)^{-1}$.

令

$$D=J(iX-Y)=-(iX+Y)J=-(iX-Y)^{-1}J$$

则容易验证 D 是辛矩阵且满足 $D^2=E$, 换言之, D 是 D 型矩阵.

现在设 A_j,A_k 是两个与 $\widetilde{A_1},\widetilde{A_2}$ 同时反交换的 A 型矩阵, 则 A_j,A_k 分别具有形式

$$A_j=\begin{bmatrix}X_j & Y_j\\ Y_j & -X_j\end{bmatrix},\qquad A_k=\begin{bmatrix}X_k & Y_k\\ Y_k & -X_k\end{bmatrix}$$

与 A_j, A_k 对应的 D 型矩阵分别记为

$$\begin{aligned}D_j&=J(\sqrt{-1}X_j-Y_j)=-(\sqrt{-1}X_j+Y_j)J\\ D_k&=J(\sqrt{-1}X_k-Y_k)=-(\sqrt{-1}X_k+Y_k)J\end{aligned}$$

计算 A_j,A_k 的反交换子得到

$$\begin{aligned}A_jA_k+A_kA_j&=\begin{bmatrix}X_j & Y_j\\ Y_j & -X_j\end{bmatrix}\begin{bmatrix}X_k & Y_k\\ Y_k & -X_k\end{bmatrix}+\begin{bmatrix}X_k & Y_k\\ Y_k & -X_k\end{bmatrix}\begin{bmatrix}X_j & Y_j\\ Y_j & -X_j\end{bmatrix}\\ &=\begin{bmatrix}X_jX_k+Y_jY_k+X_kX_j+Y_kY_j & X_jY_k-Y_jX_k+X_kY_j-Y_kX_j\\ -(X_jY_k-Y_jX_k+X_kY_j-Y_kX_j) & X_jX_k+Y_jY_k+X_kX_j+Y_kY_j\end{bmatrix}\\ &=\begin{bmatrix}S_{jk} & T_{jk}\\ -T_{jk} & S_{jk}\end{bmatrix}\end{aligned}$$

计算 D_j,D_k 的反交换子有：

$$\begin{aligned}D_jD_k+D_kD_j&=-(\sqrt{-1}X_j+Y_j)J\cdot J(\sqrt{-1}X_k-Y_k)-(\sqrt{-1}X_k+Y_k)J\cdot J(\sqrt{-1}X_j-Y_j)\\ &=(\sqrt{-1}X_j+Y_j)(\sqrt{-1}X_k-Y_k)+(\sqrt{-1}X_k+Y_k)(\sqrt{-1}X_j-Y_j)\\ &=-(X_jX_k+Y_jY_k+X_kX_j+Y_kY_j)-\sqrt{-1}(X_jY_k-Y_jX_k+X_kY_j-Y_kX_j)\\ &=-S_{jk}-\sqrt{-1}T_{jk}\end{aligned}$$

比较两式容易推出 (在推导 $\Leftarrow$ 时注意到 S_{jk} 与 T_{jk} 分别是对称的与反对称的)

$$A_jA_k + A_kA_j = 0 \iff D_jD_k + D_kD_j = 0$$

于是, 如果 $4l$ 阶矩阵 $\widetilde{A_1}, \widetilde{A_2}, A_3, \cdots, A_r$ 满足 (1), 则 $A_3, \cdots, A_r$ 给出 $r-2$ 个 $2l$ 阶矩阵 $D_3, \cdots, D_r$ 满足 (4).

反之, 设 $2l$ 阶矩阵 $D_1, \cdots, D_r$ 满足 (4). 对每一个 $k = 1, \cdots, r$, 从方程

$$\sqrt{-1}X_k - Y_k = JD_k, \qquad \sqrt{-1}X_k + Y_k = -D_kJ$$

解得

$$X_k = \frac{JD_k - D_kJ}{2\sqrt{-1}}, \qquad Y_k = -\frac{D_kJ + JD_k}{2}.$$

然后以这些矩阵偶 X_k, Y_k 构造 $4l$ 阶矩阵

$$A_k = \begin{bmatrix} X_k & Y_k \\ Y_k & -X_k \end{bmatrix}, \qquad (k = 1, \cdots, r)$$

容易验证 $A_1, \cdots, A_r$ 是 A 型矩阵, 而且与 $\widetilde{A_1}, \widetilde{A_2}$ 一起满足 (1). 于是 (5) 成立. **证毕.**

引理 2 的证明: 令 $A = D_1D_2$, 则 $A^2 = -E$, 根据引理 5 (ii), 不妨设

$$A = \begin{bmatrix} iE & 0 \\ 0 & -iE \end{bmatrix}$$

注意到辛矩阵 D_2 与 A 反交换, 于是推出 D_2 具有形式

$$D_2 = \begin{bmatrix} 0 & T \\ -T^{-1} & 0 \end{bmatrix}$$

其中 $T \in M(n/2, \mathbb{C})$ 是可逆反对称矩阵, 于是 $n/2 = 2l$ 是偶数, 且由反对称矩阵的相合理论知, 存在 $Z \in GL(n/2, \mathbb{C})$ 使得

$$ZTZ^T = iJ.$$

令

$$P = \begin{bmatrix} Z & 0 \\ 0 & (Z^T)^{-1} \end{bmatrix}$$

则容易验证 $P \in Sp(n/2, \mathbb{C})$, 而且

$$PAP^{-1} = A, \quad PD_2P^{-1} = \widetilde{D_2}.$$

于是

$$PD_1P^{-1} = P(AD_2)P^{-1} = PAP^{-1} \cdot PD_2P^{-1} = A\widetilde{D_2} = \widetilde{D_1}.$$

接下来的证明是直接的. 从略. **证毕.**

定理 2 的证明. 这是引理 $1-4$ 的全部内容. **证毕.**

6.5 Hurwitz 定理的华罗庚证明

这一节我们从华罗庚链 (5) – (8) 导出 $\mathbb{C}$ 上的 Hurwitz 定理.

由 (5) – (8), 我们有递推关系

$$\lambda(16n) = \lambda(n) + 8$$

对 $\lambda(n) = K(n), \Sigma(n)$ 成立, 而对 $\lambda(n) = P(n), X(n)$ 当 n 是偶数时成立.

特别的, $K(n)$ 满足周期关系

$$K(16n) = K(n) + 8 \tag{6.12}$$

根据递推关系 (12), 我们只需对 $q = 0, 1, 2, 3$ 求出 $K(n)$ 的值就可以确定 $K(n)$ 的所有值. 下面我们逐一确定.

由于奇数阶反对称矩阵不可逆, 所以不存在奇数阶的 A 型矩阵, 于是

$$K(p) = 0$$

根据定理 2,

$$K(2p) = 1$$

又根据 (5), 以及 (5), (6), 有

$$\begin{aligned} K(4p) &= X(2p) + 2 = 1 + 2 = 3, \\ K(8p) &= X(4p) + 2 = P(2p) + 4 = 3 + 4 = 7. \end{aligned}$$

于是对于 $n = 2^q\, p$, 其中 p 是奇数, $q = 4a + b$, $b = 0, 1, 2, 3$, 我们有

$$K(n) = K(2^{4a+b}\, p) = K(16^a\, 2^b\, p) = 8a + K(2^b\, p) = \begin{cases} 8a + 0 & b = 0 \\ 8a + 1 & b = 1 \\ 8a + 3 & b = 2 \\ 8a + 7 & b = 3 \end{cases}$$

容易看到, 此处给出的 $K(n)$ 与定理 1 中给出的 $K_{\mathbb{C}}(n)$ 的表达式一致. 这就证明了 $\mathbb{C}$ 上的 Hurwitz 定理.

注 1: $K(n)$ 是华罗庚采用的记号, 通常的文献中沿袭 Radon 的记号 $\rho(n)$, 定义如下: 设 $n = 2^q p$, 其中 p 是奇数, $q = 4a + b$, $a \geqslant 0$, $b = 0, 1, 2, 3$, 则

$$\rho(n) = 8a + 2^b$$

可以直接验证, $\rho(n) = K(n) + 1$. 实际上, $\rho(n)$ 是下述原始的 Hurwitz 方程

$$A_i A_j^T + A_j A_i^T = 2\delta_{ij} E$$

在 $M_n(F)$($F = \mathbb{C}$ 或 $\mathbb{R}$) 中的极大解个数, 这就是 Hurwitz-Radon 定理的经典表述, 见 [13].

注 2: 这个证明从更深的层面上解释了 $K_{\mathbb{C}}(n)$ 运算式中的模 4 周期性. 而且, 我们还可以得到 $\Sigma(n), P(n), X(n)$ 的运算式. 例如, $P(n)$ 的运算式如下:

$$P(n) = \begin{cases} 2q - 1 & q \equiv 0 \pmod 4 \\ 2q + 1 & q \equiv 1 \pmod 4 \\ 2q & q \equiv 2 \pmod 4 \\ 2q - 1 & q \equiv 3 \pmod 4 \end{cases}$$

6.6 酉化的华罗庚矩阵方程组与 Radon 定理

前面我们已经确定出 $K_{\mathbb{C}}(n)$ 的值, 根据有限群表示一个基本结果, 我们可以直接推出 $K_{\mathbb{R}}(n) = K_{\mathbb{C}}(n)$, 见 [9]. 但是, 事实上我们可以通过稍加修改华罗庚链而得到 Radon 定理的一个平行证明.

注意到线性代数里的一个基本事实, 一个矩阵满足下述三个条件的两个必定同时满足第三个:

$$A^T = -A, \quad A^2 = -E, \quad AA^T = E$$

由此可知, (1) 中的每个矩阵 A_i 必定也是正交矩阵, 所以矩阵代数 $M(n, F)$ 中的 Hurwitz 矩阵方程等价于正交群 $O(n, F)$ 的下述方程

$$A_i^2 = -E, \qquad A_i A_j = -A_j A_i, \qquad i \neq j$$

Hurwitz-Radon 定理告诉我们上述方程在复正交群 $O(n, \mathbb{C})$ 和实正交群 $O(n, \mathbb{R})$ 中的解的最大个数一致. 注意到

$$O(n, \mathbb{C}) \cap U(n) = O(n).$$

这就启发我们考虑带酉限制的华罗庚矩阵方程组, 即要求方程 (1), (2), (3), (4) 中的每个矩阵 A_i, B_i, C_i, D_i 都是酉矩阵. 或许可以期望, 此时华罗庚链 (5) – (8) 仍然保持, 于是作为推论, 就得到 Hurwitz 定理的酉化对应: Radon 定理. 事实表明, 这个期望是合理的.

为了突出群论的视角, 我们将定理 2 的表述稍微改变一下. 注意到, 按照定义, 辛矩阵的酉限制就是酉矩阵.

定理 3(酉化的华罗庚链): 设下述四个方程

$$A_i^2 = -E, \qquad A_iA_j = -A_jA_i,\ (i \neq j), \qquad A_i \in O(n) \tag{6.13}$$

$$B_i^2 = E, \qquad B_iB_j = -B_jB_i,\ (i \neq j), \qquad B_i \in O(n) \tag{6.14}$$

$$C_i^2 = -E, \qquad C_iC_j = -C_jC_i,\ (i \neq j), \qquad C_i \in Sp(n) \tag{6.15}$$

$$D_i^2 = E, \qquad D_iD_j = -D_jD_i,\ (i \neq j), \qquad D_i \in Sp(n) \tag{6.16}$$

的极大解个数分别为 $\widetilde{K}(n), \widetilde{\Sigma}(n), \widetilde{P}(n), \widetilde{X}(n)$, 则有

$$\widetilde{K}(n) = \widetilde{X}(n/2) + 2 \tag{6.17}$$

$$\widetilde{X}(n) = \widetilde{P}(n/2) + 2 \tag{6.18}$$

$$\widetilde{P}(n) = \widetilde{\Sigma}(n/2) + 2 \tag{6.19}$$

$$\widetilde{\Sigma}(n) = \widetilde{K}(n/2) + 2 \tag{6.20}$$

其中 (15)(16) 中要求 n 被 4 整除, (17)(18) 中要求 n 为偶数. 并且, 对 $n = 2p$, 其中 p 为奇数, 我们有以下初始值:

$$\widetilde{K}(2p) = 1, \quad \widetilde{X}(2p) = 1, \quad \widetilde{P}(2p) = 3, \quad \widetilde{\Sigma}(2p) = 2 \tag{6.21}$$

作为定理 2 与定理 3 的推论, 对于 (11) – (14), 我们得到对 $\lambda = K, X, P, \Sigma$ 有,

$$\widetilde{\lambda}(n) = \lambda(n)$$

特别的, 我们得到 Radon 定理: $\widetilde{K}(n) = K(n)$

定理 3 的证明需要将引理 $1-4$ 酉化, 我们只要在相应的引理 1-4 的证明中将各个相似变换或相合变换的矩阵加强为酉矩阵. 作为例子, 我们这里给出引理 1 和引理 2 的酉化对应如下.

引理 7: 设 $A_1, A_2 \in O(n)$ 满足

$$A_1^2 = A_2^2 = -E, \qquad A_1A_2 = -A_2A_1.$$

则 $n=4l$ 且存在 $P\in O(n)$ 使得

$$PA_1P^{-1}=\widetilde{A_1}=\begin{bmatrix}0 & E\\ -E & 0\end{bmatrix}\qquad 且\qquad PA_2P^{-1}=\widetilde{A_2}=\begin{bmatrix}J & 0\\ 0 & -J\end{bmatrix}$$

设 $A\in O(n)$ 与 $\widetilde{A_1},\widetilde{A_2}$ 同时反交换且满足 $A^2=-E$, 则 A 具有形式

$$A=\begin{bmatrix}X & Y\\ Y & -X\end{bmatrix}$$

其中 $X,Y\in M(n/2,\mathbb{R})$ 是反对称矩阵, 并且满足

$$X^2+Y^2=-E,\quad XY=YX,\quad JX=-XJ,\quad JY=YJ$$

对于满足上述关系的一对 $X,Y\in M(n/2,\mathbb{R})$, 令

$$D=J(iX-Y)$$

则 $D\in Sp(n/2)$ 且 $D^2=E$. 进一步, 两个如上形状的 A 型矩阵 A_j,A_k 反交换当且仅当对应的 D 型矩阵 D_j,D_k 反交换. 于是 (15) 成立.

引理 8: 设 $D_1,D_2\in Sp(n)$ 满足

$$D_1^2=D_2^2=E,\qquad D_1D_2=-D_2D_1.$$

则 $n=4l$ 且存在 $P\in Sp(n)$ 使得

$$PD_1P^{-1}=\widetilde{D_1}=\begin{bmatrix}0 & iJ\\ iJ & 0\end{bmatrix}\qquad 且\qquad PD_2P^{-1}=\widetilde{D_2}=\begin{bmatrix}0 & -J\\ J & 0\end{bmatrix}$$

设 $D\in Sp(n)$ 与 $\widetilde{D_1},\widetilde{D_2}$ 同时反交换且 $D^2=E$, 则 D 具有形式

$$D=\begin{bmatrix}iC & 0\\ 0 & iJCJ\end{bmatrix}$$

其中 $C\in Sp(n)$ 且 $C^2=-E$. 进一步, 两个如上形状的 D 型矩阵 D_j,D_k 反交换当且仅当对应的 C 型矩阵 C_j,C_k 反交换. 于是 (16) 成立.

为证明引理 7 和引理 8, 我们需要以下四个技术性的矩阵引理.

引理 9: 设 $\{A_\alpha\}$, $\{B_\alpha\}$ 是 $M(n,\mathbb{R})$ 中的矩阵集, 若存在 $P\in GL(n,\mathbb{C})$ 使得

$$PA_\alpha P^{-1}=B_\alpha,\qquad \forall\alpha.$$

则存在 $Q \in GL(n,\mathbb{R})$ 使得

$$QA_\alpha Q^{-1} = B_\alpha, \qquad \forall\alpha.$$

下面的证明引自 Halmos [3].

证明: 令 $P = S + iT$, 其中 S,T 为实矩阵, 则对每个指标 α,

$$PA_\alpha P^{-1} = B_\alpha$$

给出

$$PA_\alpha = B_\alpha P.$$

从而有

$$SA_\alpha = B_\alpha S, \qquad TA_\alpha = B_\alpha T \qquad \forall\alpha.$$

于是对任意的 $\lambda \in \mathbb{C}$ 有

$$(S+\lambda T)A_\alpha = B_\alpha (S+\lambda T) \qquad \forall\alpha.$$

选取 $\lambda_0 \in \mathbb{R}$ 使得 $Q = (S+\lambda_0 T)$ 可逆 (这等价于选取 $\lambda_0 \in \mathbb{R}$ 使得多项式 $f(\lambda) = \det(S+\lambda T)$ 在 $\lambda = \lambda_0$ 处的取值不等于零, 这一点是容易办到的, 因为 $f(\lambda) \neq 0$, 事实上 $f(i) = \det(S+iT) = \det(P) \neq 0$.), 则 $Q \in GL(n,\mathbb{R})$ 满足引理所述性质. **证毕.**

引理 10(酉辛矩阵的谱定理): 设 $U \in Sp(n)$, 则存在 $T \in Sp(n)$ 使得

$$TUT^{-1} = \mathrm{diag}[\lambda_1, \cdots, \lambda_m, \overline{\lambda_1}, \cdots, \overline{\lambda_m}].$$

此处 $m = n/2$, 且 $\lambda_1, \cdots, \lambda_m$ 是模为 1 的复数.

证明可见 Admas [1]. 它是酉矩阵的谱定理在酉辛矩阵的平行结果.

下面两个结果常常见诸矩阵论的标准教材, 如 [4]. 事实上, 华罗庚早年也独立地发现了引理 11.

引理 11(Takagi 分解): 设 A 为 n 阶复对称矩阵, 则存在 $U \in U(n)$ 使得

$$U'AU = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$$

其中 $\sigma_1, \cdots, \sigma_n$ 为 AA^* 的特征值的非负平方根.

引理 12: 设 A 为 n 阶复反对称矩阵, 则存在 $U \in U(n)$ 使得

$$U'AU = \mathrm{diag}(\sigma_1 J, \cdots, \sigma_k J, 0)$$

其中 $\sigma_1, \cdots, \sigma_k$ 为 AA^* 的正特征值的正平方根, J 为 2 阶标准辛矩阵.

注: 有必要指出, 华罗庚链 (5)–(8) 在 $\mathbb{R}$ 上不成立. 原因在于, 引理 5 (ii) 在 $\mathbb{R}$ 上不成立. 例如, 对于 2 阶标准辛矩阵 J, J 与其转置矩阵 $J' = -J$ 是实相似的, 但是直接计算表明, 它们不是实辛相似的. 这是我们不能直接实化华罗庚链的原因所在. 我们在标题中所谓的“酉化”, 事实上就是酉限制, 在拓扑上相当于紧致化, 将非紧李群 $O(n, \mathbb{C})$ 与 $Sp(n, \mathbb{C})$ 中的紧致部分提取出来. 这一技巧为 H.Weyl 首创, 并且命名为“酉技巧”(unitary trick), 见 [14]p.173.

6.7 历史评述

事实上, 华罗庚 [4] 并没有得到正确的华罗庚链, 这一点首先为黄用谞 [15] (p.69) 指出. 华罗庚在引理 1(见 [4]p221 (75) 给出的是 $K(n) = K(n/2) + 1$) 的证明中出现了失误. 根据笔者的分析发现, 导致这一错误的是根源是 [4]p224 (94) 式下边的连等式, 其中第 3 个等号不成立.

华罗庚本应该从文献中获知 Hurwitz 定理的, 因为 1930 年代出版的 MacDuffee [11] 在两个不同的地方分别提到了 Hurwitz 和 Radon 的结果. 华罗庚显然是翻过这本书的, 他不仅在同一篇文章 [4](p206 脚注) 中引用过此书, 而且早在 1944 年发表的文章中就多次引用过.

MacDuffee [11] p. 97 只是简单提到了 Hurwitz [6] 中的工作, 并没有给出他得到的具体结果. 事实上, Hurwitz 对定理 1 的原始表述比较复杂 (见黄用谞 [15]):

定理 3(Hurwitz): 存在 p 个矩阵 $A_1, \cdots, A_p \in M(n, \mathbb{C})$ 满足

$$A_i A_j^T + A_j A_i^T = 2\delta_{ij} E$$

当且仅当 p 与 n 满足以下关系:

(i) 对 $p = 2r + 1$,

$$n = \begin{cases} \mu \cdot 2^r & r \equiv 0, 3 \pmod 4 \\ \mu \cdot 2^{r+1} & r \equiv 1, 2 \pmod 4 \end{cases}$$

(ii) 对 $p = 2r + 2$,

$$n = \begin{cases} \mu \cdot 2^r & r \equiv 3 \pmod 4 \\ \mu \cdot 2^{r+1} & r \equiv 0, 1, 2 \pmod 4 \end{cases}$$

其中 μ 是一个正奇数.

相对而言, Radon [13] 的结果在 MacDuffee [11] p. 80 有详细表述.

定理 4(Radon): 存在 k 个矩阵 $A_1, \cdots, A_k \in M(n, \mathbb{R})$, 使得对任意的实数 $x_1, \cdots, x_k$, $x_1^2 + \cdots + x_k^2 = 1$, 矩阵 $x_1A_1 + \cdots + x_kA_k$ 为正交矩阵, 当且仅当 $k \leqslant \rho(n)$. 此处 $\rho(n)$ 如第五节注 1 定义.

正如等周定理有两种等价的表述一样, Radon 定理与 Hurwitz 定理根本就是同一个定理的两种不同表述, 这两个定理合在一起就是说, Hurwitz 问题在 $\mathbb{C}$ 和 $\mathbb{R}$ 上有相同的解. 正是认识到这一点, Eckmann [3] 给出了 Hurwitz-Radon 定理的一个统一的基于有限群表示论的证明.

现在, Hurwitz-Radon 定理已经作为练习被吸收到线性代数的习题集 [12] 中, 但是那里的解法远不如华罗庚先生给出的解法漂亮, 华先生的这个证明完全可以与 Hurwitz 和 Radon 原来的解法相媲美. 将 Radon [13] 的原始证明跟华罗庚的证明放在一起比较是极为有趣的. Radon 的证明中产生了 8 个矩阵方程, 其中 3 个是 $\mathbb{R}$ 上的, 2 个是 $\mathbb{C}$ 上的, 还有 3 个是四元数体 $\mathbb{H}$ 上的. 关于 Radon 的详细论证请读者参考 [7] 的转述.

参考文献

[1] J.F.Adams, Lectures on Lie Groups, The University of Chicago Press, 1969.

[2] B.Eckmann, Gruppentheorectischer Beweis des Satzes von Hurwitz-Radon uber die Komposition quadratischer Formen, *Comment.Math.Helv.*,**15**(1942), 358–366.

[3] P.R.Halmos, Linear Algebra Problem Book, The Mathematical Association of America, 1995.

[4] R.A.Horn and C.R.Johnson, Matrix Analysis, 中译本《矩阵分析》, 杨奇译, 北京, 机械工业出版社, 2005.

[5] Loo-Keng Hua, Geometries of Matrices II. Study of involutions in the geometry of symmetric matrices, *Trans.A.M.S.***61**(1947), 193–228.

[6] A.Hurwitz, Über die Komposition der quadratischen Formen, *Math.Ann.* **88**(1923), 1–25.

[7] 江上鸥, Hurwitz-Radon 问题——等价和约化是处理数学问题的一种基本方法, 《数学通报》, 1964 年 04 期.

[8] I.Kaplansky, Linear Algabra and Geometry–A Second Course, Allyn and Bacon, 1969.

[9] 林开亮, Hurwitz-Radon 矩阵方程,《数学传播》, 36 卷 1 期 (2012), pp. 48-63.

[10] Kai-Liang Lin, Hurwitz–Radon's symplectic analogy and Hua's cyclic recurrence relation, *Electronic Journal of Linear Algebra*, **26**(2013), 858–872.

[11] C. C. MacDuffee, The Theory of Matrices, Berlin, Springer, 1933.

[12] V.V.Prasolov, Problems and Theorems in Linear Algebra, Translation of Mathematical monographs, Vol.134, American Mathematical Society, 1994.

[13] J.Radon, Lineare Scharen orthogonaler Matrizen, *Abhandlungen aus dem mathematischen Seminar der Hambergischen Univerdität***1**(1922), 1–14.

[14] H.Weyl, Classical Groups, Their Invariants and Representations, Princeton University Press, 1946.

[15] Yung-Chow Wong, Isoclinic n-Planes in $2n$-Spaces, Clifford Parallels in Elliptic $(2n-1)$-Spaces, and Hurwitz Matrix Equations, *Memoirs A.M.S.***no.41**, 1961.

第七章　解常系数线性微分方程和递推关系的新方法——秦九韶和亥维赛的遗产

7.1　引言：亥维赛的妙招

众所周知, 非齐次的常系数线性微分方程往往可以通过待定系数法求解, 这是通常教科书上介绍的方法, 如 [2].

记得从前我在给大一学生上微积分讲到常系数线性微分方程时, 常有爱动脑筋的学生问我："为什么要用待定系数法求解, 这个方法从天而降, 让我感觉自己很蠢. 数学太深奥了!" 确实, 作为老师, 我讲待定系数法也很难受. 数学是讲道理的, 我按照书上的讲当然没错, 可是好像也说不出什么道理. 我都不记得当时我是怎么回答他的! 直到有一天我看到亥维赛 (Oliver Heaviside, 1850–1925) 的一个妙招, 我才觉得终于得救了, 下一次讲到常系数线性微分方程时我一定要分享给学生!

我是从彭罗斯 (Rogers Penrose) 的一本书 [10, pp. 493–494] 中学到这个妙招的, 摘引如下：

> 亥维赛的洞见是, 微分算子通常可以如普通数一样处理, 这个事实对求解某些类型的微分方程非常有用. 我们来看一个例子, 考虑微分方程
>
> $$y + \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = x^5. \tag{1}$$
>
> 我们想要求出一个特解. 亥维赛的方法是, 像对待普通数那样对待微分算子 $\mathrm{d}/\mathrm{d}x$. 为让它看起来更合"情理", 我们用一个单独的字母 D 来

表示这个算子

$$D = \frac{\mathrm{d}}{\mathrm{d}x}. \tag{2}$$

两次作用 D 所得到的平方算子 D^2 表示两次微分, 即二阶求导算子 $\mathrm{d}^2/\mathrm{d}x^2$; D^3 表示三阶求导算子 $\mathrm{d}^3/\mathrm{d}x^3$; 依次类推, 于是我们的方程成为 $y + D^2y = x^5$, 我们可以表示为

$$(1 + D^2)y = x^5. \tag{3}$$

我们可以通过"除以系数 $1+D^2$"而形式地"解"方程, 从而得到解 $y = (1 + D^2)^{-1}x^5$. 将 $(1 + D^2)^{-1}$ 展开成"D 的幂级数":

$$(1 + D^2)^{-1} = 1 - D^2 + D^4 - D^6 + \cdots, \tag{4}$$

从而我们求出一个 (正确的!) 特解:

$$\begin{aligned} y =& (1 + D^2)^{-1}x^5 \\ =& (1 - D^2 + D^4 - D^6 + \cdots)x^5 \\ =& (1 - D^2 + D^4)x^5 \qquad (\text{注意 } D^6x^5 = 0, \text{ 等等}) \\ =& x^5 - (x^5)'' + (x^5)'''' \\ =& x^5 - 20x^3 + 120x \end{aligned} \tag{5}$$

如果注意恰当的规则, 那么可以使得这个形式步骤无懈可击——不过, 亥维赛在首次运用它时却遭到了强烈的反对.

7.2 寻找关键点

据加州大学伯克利大学的数学教授伍鸿熙讲 ([17]), 他从数学大师陈省身那里学到的最重要的一点, 就是

> 从整体而非局部地去把握问题, 不考虑复杂的技术细节, 即抓住关键点.
> ……他 [陈省身先生] 对待问题的人生哲学, 就是每件事情都有一个关键点. 如果你抓住了这个关键点, 那么剩下的迟早会解决.

回到上述亥维赛解方程的巧妙方法, 成功的关键, 又在哪里呢?

一个初步的分析指出, 成功的关键在于: 方程 (1) 右边的非齐次项 x^5 是一个多项式, 而微分算子 D 的高次幂作用在上面等于 0, 因此 (4) 右边给出的 D 的级数作用于 x^5 后, 约化为一个有限截断 (即 $1 - D^2 + D^4$) 的作用, 从而变无限为有限, 化分析为代数!

7.3　隐藏的本质

可以设想, 亥维赛的方法当初之所以会遭到反对, 最主要的一点, 是无穷和 (4) 不好接受. [1]正如彭罗斯所指出的, 这是可以严格化的. [2] 但实际上, 有一个更简单的方法, 完全不涉及无穷和!

我们只要观察到:

$$(1+D^2)(1-D^2+D^4)=1+D^6$$

这意味着, 在次数小于 6 的复系数多项式函数空间 $\mathbb{C}_6[x]$ 上 (有 $D^6=0$), 其实有算子乘积

$$(1+D^2)(1-D^2+D^4)=1+D^6=1,$$

即, $(1-D^2+D^4)$ 是 $(1+D^2)$ 的右逆, 为简单起见, 我们仍然记为 $(1+D^2)^{-1}$! 从而, 对任意的 $f(x)\in\mathbb{C}_6[x]$, 可以直接写出方程 $(1+D^2)y=f(x)$ 的解为

$$y=(1+D^2)^{-1}f(x)=(1-D^2+D^4)f(x).$$

特别的, 对 $f(x)=x^5$, 我们即得到原来方程 $(1+D^2)y=x^5$ 的解

$$y=(1-D^2+D^4)x^5=x^5-20x^3+120x.$$

现在我们看出, 为了求解方程 $(1+D^2)y=x^5$, 我们真正需要的, 只是求出 $1+D^2$ 在包含 x^5 的某个函数空间 (这里就是 $\mathbb{C}_6[x]$) 上的右逆! 亥维赛的方法, 相当于通过对 $(1+D^2)^{-1}$ 的形式幂级数 (4) 取有限截断而得到这个右逆. 但通过进一步的分析, 我们发现有更简单的方法.

7.4　化微分方程为代数方程

为了说明这个更一般的方法, 我们不妨假设所考虑的, 是一个更一般的方程

$$P(D)y=x^5, \tag{6}$$

[1]因此, 通常的微分方程教科书并不介绍这个妙招. 经典的北大教材 [2] 第一版倒是介绍了, 但是作为加星号的一节 (" 算子法和拉普拉斯变换法简介"). 也许是反响不好, 在第二版中, 这一节又被删掉了!

[2]对此, 包括维纳 (Norbert Wiener) 在内的许多数学家都作出了贡献, 其关键词是 operational calculus, 有兴趣的读者可以进一步了解.

其中 $P = P(x)$ 是一个复系数多项式. 根据前面的分析容易看出, 为了求出方程 (6) 的一个特解, 我们只要求出 $P(D)$ 在 $\mathbb{C}_6[x]$ 上的一个右逆. 由于 D 在 $\mathbb{C}_6[x]$ 上的极小多项式是 x^6, 所以我们只要求出满足同余条件[3]

$$P(x)U(x) \equiv 1 \pmod{x^6}, \tag{7}$$

的一个复系数多项式 $U = U(x)$, 即可得到 $P(D)$ 在 $\mathbb{C}_6[x]$ 上的一个右逆 $U(D)$. 这是因为, 在 $P(x)U(x) - 1 = x^6Q(x)$ 中令 $x = D$ 即可得到

$$P(D)U(D) - I = D^6Q(D) = 0Q(D) = 0$$

即 $P(D)U(D) = I$.

因此, 我们要做的, 就是求解多项式同余方程 (7). 幸运的是, 这种类型的问题是古人早已解决的. 特别是, 当我们用整数代替多项式时, 相应的问题即整数的同余方程[4]

$$ax \equiv 1 \pmod{b}, \tag{8}$$

其中 $a, b > 0$ 是给定的非零整数, x 是待求的整数.

为帮助读者更好地掌握多项式同余方程 (7) 的解法, 我们下面先回顾一下整数同余方程 (8) 的解法.

7.5 解整数同余方程的求一术

对整数同余方程 (8), 我们有经典的解法, 而且至少有三种等价的表述：分别是欧几里得算法、连分数法与求一术. 这里我们介绍的是第三种：求一术.

求一术是南宋数学家秦九韶的发明, 他命名为大衍求一术[5], 是所谓“大衍总数术”的关键一步. 清代数学家黄宗宪后来进一步简化了秦九韶的方法, 我们现在介绍的, 就是这个简化的版本.

秦九韶–黄宗宪的方法可用矩阵表述. 首先写出一个 2×2 的矩阵

$$\begin{bmatrix} a & 1 \\ b & 0 \end{bmatrix}$$

其中第一列 a, b 都是源自方程 (8), 而第二列的两个元素 $1, 0$ 则是我们添加进来的 (用于探测未知数 x).

[3]这个条件等价于 $x^6 \mid P(x)U(x) - 1$, 即 x^6 整除 $P(x)U(x) - 1$.

[4]这个条件等价于 $b \mid ax - 1$, 即 b 整除 $ax - 1$.

[5]至于” 大衍” 的意思, 陈省身在 [1] 中将它解释为“推广”.

求一术算法如下 (见 [15] 或 [11])：对第一列的数 a, b 用带余除法 (较大的数除以较小的数), 设得到的商为 q, 则较大的数那一行减去较小数的那一行对应元素的 q 倍；于是新得到的矩阵的第一列两个元素替换为第一次带余除法的除数与余数, 重复之前的操作, 直到某一步带余除法得到的余数为 1(算法结束, 用红色标记), 此时 1 的正右方的数, 即为所求的 x(用蓝色示意).

这个算法的理论基础可见最后一节的练习 1.

作为例子, 我们用秦九韶–黄宗宪的方法来求

$$5x \equiv 1 \pmod 7, \tag{9}$$

的一个解.

解：求一术步骤如下：

$$\begin{bmatrix} 5 & 1 \\ 7 & 0 \end{bmatrix} \xrightarrow[\text{下行减去上行的1倍}]{7=1\cdot 5+2} \begin{bmatrix} 5 & 1 \\ 2 & -1 \end{bmatrix} \xrightarrow[\text{上行减去下行的2倍}]{5=2\cdot 2+1} \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}$$

根据求一术, x 在 1 的右边, 即 $x = 3$. 这是很容易验证的：

$$5 \cdot 3 = 15 \equiv 1 \pmod 7.$$

当然, 你或许以为我是把问题搞复杂了, 你甚至在一开始就试出来 $x = 3$ 是一个解. 然而, 正如吴文俊先生多次强调的 (例如, 见 [16]), 中国古代数学讲的是一种算法. 简单的例子你用技巧可以解决, 但如果换成一个稍微复杂的例子, 如解方程

$$250x \equiv 1 \pmod{2017},$$

你可能就无计可施了！[6]

可以看出, 上述求一术的基础是：

整数的带余除法：设 a 和 b 是两个整数, 其中 $b > 0$, 则存在唯一的整数 q 和 r 使得

$$a = qb + r,$$

其中 r 满足 $0 \leqslant r < b$.

[6]这让我们回想起著名数学家、数学教育家波利亚 (George Pólya) 的一句名言 ([9]“传统的数学教授”一节):“方法与技巧之差别何在？方法乃放诸四海而皆准之技巧. (What is the difference between method and device? A method is a device which you used twice.)” 中国古代数学, 注重的是方法 (“法”“术”同义) 而非技巧. 读者若想领教秦九韶–黄宗宪的求一术的威力, 不妨用上面的方程 $250x \equiv 1 \pmod{2017}$ 一试！

7.6 解多项式同余方程的求一术

注意到, 类似的, 对多项式我们也有带余除法：

多项式的带余除法：设 $a(x)$ 和 $b(x)$ 是两个多项式, 其中 $b(x)$ 次数大于等于 1, 则存在唯一的多项式 $q(x)$ 和 $r(x)$ 使得

$$a(x) = q(x)b(x) + r(x),$$

其中 $r(x)$ 满足 $\deg r(x) < \deg b(x)$, 这里 $\deg r(x), \deg b(x)$ 分别表示多项式 $r(x), b(x)$ 的次数 (degree, 缩写为 deg).

因此, 同样有求解一般多项式同余方程 (假定其中 $a(x), b(x)$ 互质, 这是方程有解的充分必要条件)

$$a(x)u(x) \equiv 1 \pmod{b(x)}, \tag{10}$$

的求一术：

首先写出一个 2 行 2 列的阵

$$\begin{bmatrix} a(x) & 1 \\ b(x) & 0 \end{bmatrix}$$

对第一列的两个多项式 $a(x), b(x)$ 用带余除法 (次数高的多项式除以次数低的), 设得到的商为 $q(x)$, 则次数高的那一行减去次数低的那一行对应元素的 $q(x)$ 倍；于是新得到的矩阵的第一列两个元素替换为第一次带余除法的除式与余式, 重复之前的操作, 直到某一步带余除法得到的余式为某个非零常数 c(算法结束), 此时 c 的正右方的多项式除以 c, 即为所求的 $u(x)$.

评注：这里用非零常数 c 取代了整数同余方程求一术中的目标“1”. 可以这么理解, 非零常数是多项式环中的单位 (可逆元), 而 ± 1 也是整数环中的单位. (在整数的情况, 由于我们限定了余数大于等于 0, 所以把 -1 的情况就排除了. 如果我们限定余数是关于 0 最对称的一组剩余系, 求一术就成为了“求正负一术”.) 所以求一术的目标“得一”(因而又名“得一术”), 实则是“得单位”. 因此, 最恰当的称谓既不是“求一术”, 也不是“得一术”, 而是“求逆术”. 不论是整数还是多项式的情况, 我们所讨论的方程 (8)(9) 实际上就是求一个给定元素 (在某个商环中) 的逆.

作为例子, 我们来看一开头彭罗斯所给出的微分方程 (3) 所确定的多项式同余方程：

$$(x^2+1)u(x) \equiv 1 \pmod{x^6}$$

解：求一术步骤如下：

$$\begin{bmatrix} x^2+1 & 1 \\ x^6 & 0 \end{bmatrix} \xrightarrow[\text{下行减去上行的}(x^4-x^2+1)\text{倍}]{x^6=(x^4-x^2+1)\cdot(x^2+1)-1} \begin{bmatrix} x^2+1 & 1 \\ -1 & -(x^4-x^2+1) \end{bmatrix}$$

根据求一术, $u(x)$ 等于 -1 的右边的多项式除以 -1, 即

$$u(x) = \frac{-(x^4-x^2+1)}{-1} = x^4-x^2+1$$

这是很容易验证的：

$$(x^2+1)\cdot(x^4-x^2+1) = x^6+1 \equiv 1 \pmod{x^6}.$$

这样我们就求得了 D^2+1 在 V_5 上的逆为 $u(D) = D^4-D^2+1$, 这与我们前面用无穷幂级数取有限截断得到的结果一致. 但从方法上讲, 解同余方程的方法更切中要害 (打蛇打七寸！), 从而也更容易理解.

7.7　非齐次项为多项式的情形

通过将微分方程转化为多项式的同余方程, 现在我们可以得到求解常系数线性微分方程

$$P(D)y = f(x), \tag{11}$$

(其中 P, f 是任意的复系数多项式) 之特解的新方法.

设 f 的次数为 m, 则 $f \in \mathbb{C}_{m+1}[x]$(其中 $\mathbb{C}_{m+1}[x]$ 表示次数小于 $m+1$ 的复系数多项式空间), 且 D 在 $\mathbb{C}_{m+1}[x]$ 上的极小多项式 x^{m+1}. 因此 (11) 所对应的代数同余方程为

$$P(x)U(x) \equiv 1 \pmod{x^{m+1}}, \tag{12}$$

当且仅当 $P(x)$ 与 x^{m+1} 互质——显然, 这等价于 $P(0) \neq 0$, 即 P 的常数项非零——时, 我们可以求出满足方程 (12) 的 $U(x)$, 从而 $U(D)$ 是 $P(D)$ 的一个右逆 (实际上是真正的逆), 由此立即得到方程 (11) 的一个特解 $y = U(D)f(x)$, 注意, 这个解是多项式, 而且次数与 f 的次数相等, 即为 m.

当 $P(0)=0$ 时怎么办呢？我们先看一个最特殊的例子, P 以 0 为唯一的零点, 如 $P(x)=x^k$, 从而我们要解的方程即

$$D^k y = f(x).$$

很明显, 通过逐次积分, 可以得到这个方程的通解.

在一般情况, 若 $P(0)=0$, 我们对 $P(x)$ 作分解

$$P(x) = x^k Q(x)$$

其中 $Q(0)\neq 0$. 从而原方程 (11) 的一个特解, 可以通过以下两步得到：第一, 用前述方法求出方程 $Q(D)z=f(x)$ 的一个特解 z, 这是一个次数等于 m 的多项式；第二, 求出方程 $D^k y = z$ 的一个特解 y, 通过每一次积分取平凡常数 (即 0), 我们可以保证 $y=x^k w(x)$, 其中 $w(x)$ 是一个 m 次多项式. 事实上, 若写

$$z = a_m \frac{x^m}{m!} + a_{m-1}\frac{x^{m-1}}{(m-1)!} + \cdots + a_0,$$

则

$$\begin{aligned} y &= a_m \frac{x^{k+m}}{(k+m)!} + a_{m-1}\frac{x^{k+m-1}}{(k+m-1)!} + \cdots + a_0 \frac{x^k}{k!} \\ &= x^k\Big(a_m \frac{x^m}{(k+m)!} + a_{m-1}\frac{x^{m-1}}{(k+m-1)!} + \cdots + a_0\frac{1}{k!}\Big) \end{aligned}$$

是方程 $D^k y = z$ 的一个特解.

评注：现在完全可以理解, 为什么在待定系数法中, 我们可以假设具有所述形式的特解了. 然而, 说明这一点绝非我们的本意. 我们的目标, 说得宏大一些, 就是要把待定系数法取而代之！当然, 目前我们只解决了非齐次项为多项式的情形, 但容易看到, 只要非齐次项可以被某个多项式 $Q(D)$ 零化, 那么同样可以“先除之而后快”. 于是问题是：究竟哪些函数 f 会满足某个多项式方程 $Q(D)f=0$ 呢? 这将我们引向下一节.

7.8 D 的广义特征函数与齐次方程的通解

为考虑以上问题, 我们先换个视角, 重新提问为：对给定的多项式 Q, 被 $Q(D)$ 零化的函数是什么样的. 换言之, 我们要求 $Q(D)$ 的零化子空间 (又称核空间). 对于这个问题, 线性代数中早已给出回答.

定理 1 设 A 是复向量空间 V 的线性变换, 设 $f(x)$ 是复系数多项式, 假定 $f(x) = (x-\lambda_1)^{m_1}\cdots(x-\lambda_s)^{m_s}$, 则有以下对应的零化子空间分解

$$\ker f(A) = \ker(A-\lambda_1)^{m_1} \oplus \cdots \oplus \ker(A-\lambda_s)^{m_s},$$

其中 $\ker B = \{v \in V, Bv = 0\}$ 表示线性算子 B 的核空间.

根据这一结果, 我们只要考虑 $Q(D) = (D-\lambda)^m$ 的特殊情形, 即我们要解方程

$$(D-\lambda)^m y = 0, \tag{13}$$

当 $m = 1$ 时, 我们得到微分算子 D 的特征方程, $(D-\lambda)y = 0$, 通常写作

$$Dy = \lambda y, \tag{14}$$

对一般的 m, (13) 可谓 D 的广义特征方程.

众所周知, (14) 的通解为 $y = C\mathrm{e}^{\lambda x}$, 其中 C 为任意常数. (据此可以理解, 在微积分中何以指数函数如此重要, 因为它是微分算子的特征函数！) 现在我们要求广义特征方程 (13) 的通解 (可谓之“广义特征函数”), 一个最简便的方法（参见 [6, p. 224]）是借助 指数平移定理 (Exponential Shift Theorem)：

定理 2 对任意的光滑函数 z 与多项式 P，有

$$P(D)\mathrm{e}^{\lambda x}z = \mathrm{e}^{\lambda x}P(D+\lambda)z. \tag{15}$$

证明：我们先证明 (15) 对 $P(D) = D$ 成立。事实上，根据 Leibniz 求导法则，我们有

$$\begin{aligned} D(\mathrm{e}^{\lambda x}z) &= D(\mathrm{e}^{\lambda z})v + \mathrm{e}^{\lambda x}Dz \\ &= \lambda\mathrm{e}^{\lambda x}v + \mathrm{e}^{\lambda x}Dz \\ &= \mathrm{e}^{\lambda x}(D+\lambda)z \end{aligned}$$

于是用数学归纳法，我们可以证明，对任意的正整数 n 有，

$$D^n(\mathrm{e}^{\lambda x}z) = \mathrm{e}^{\lambda x}(D+\lambda)^n z.$$

由此不难推出 (15) 对一切多项式 $P(D)$ 成立。**证毕**.

定理 3 广义特征方程 (13) 的通解为

$$y = \mathrm{e}^{\lambda x}(c_0 + c_1 x + \cdots + c_{m-1}x^{m-1}), \tag{16}$$

其中 $c_0, c_1, \ldots, c_{m-1}$ 为任意常数。

证明：根据指数平移定理 (15)，在变量替换 $y = \mathrm{e}^{\lambda x}z$ 之下，方程 (13) 等价于 $D^m z = 0$。容易推出，后者的通解为

$$y = c_0 + c_1 x + \cdots + c_{m-1}x^{m-1},$$

其中 $c_0, c_1, \ldots, c_{m-1}$ 为任意常数。于是，广义特征方程 (12) 的通解 $y = \mathrm{e}^{\lambda x}z$ 具有 (13) 的形式。**证毕**.

我们将形如 (16) 的函数称为拟多项式.[7] 于是, 根据前面的叙述, 我们可知, 上一节的方法可以推广到非齐次项为拟多项式的线性组合, 下一节再举例说明.

从上面的定理, 我们立即得到常系数高阶线性齐次方程的下述基本结果（参见 [2, p. 198] 定理 6.6）：

定理 4 设复系数多项式 $P(x)$ 有分解

$$P(x) = (x - \lambda_1)^{m_1} \ldots (x - \lambda_s)^{m_s},$$

其中 $\lambda_1, \ldots, \lambda_s$ 两两互异。则微分方程

$$P(D)\, u = 0$$

的通解为

$$u = \sum_{k=1}^{s} \mathrm{e}^{\lambda_k x} p_k(x),$$

其中 $p_k(x)$ 为次数小于 m_k 的复系数多项式。

7.9 非齐次项为拟多项式的情形

现在我们转向非齐次项为拟多项式的情形, 即考虑方程

$$Ly = P(D)y = \mathrm{e}^{\lambda x} f(x), \tag{17}$$

[7] 换言之, 形如 $\mathrm{e}^{\lambda x} f(x)$(其中 $f(x)$ 为多项式) 的函数, 称为拟多项式. 当 $\lambda = 0$ 时, 我们得到真正的多项式 $f(x)$.

其中 P, f 是任意的复系数多项式, λ 是一个给定的 (复) 常数. 我们的目标, 仍然是求方程 (17) 的一个特解.

为此, 只要做变量替换 $y = \mathrm{e}^{\lambda x} z$, 关于 y 的方程 (17) 等价于关于 z 的方程

$$P(D + \lambda)z = f, \tag{18}$$

这个问题我们在前一节就已经解决.

7.10　解常系数线性递推关系的新方法

不难想见, 在离散情形, 我们可以得到求解常系数线性递推关系的新方法.

齐次情形

设 $\ell = \{x = (x_0, x_1, \ldots, x_n, \ldots),\, x_i \in \mathbb{C}\}$ 为所有的复数值数列的集合.

令 T 表示作用在 ℓ 上的右平移算子：

$$T : x = (x_0, x_1, \ldots, x_n, \ldots) \mapsto (x_1, \ldots, x_n, \ldots) = Tx, \tag{19}$$

于是, 一个 d 阶常系数齐次递推关系

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_d x_{n-d} = 0, \qquad (n \geqslant d), \tag{20}$$

可以写成下述形式：

$$P(T)x_n = 0, \qquad (n \geqslant d), \tag{21}$$

其中 $P(T) = T^d - a_1 T^{d-1} - \cdots - a_d$.

根据定理 3, 为了求解方程 (21), 假定我们有特征多项式分解

$$p(t) = t^d - a_1 t^{d-1} - \cdots - a_d = (t - \lambda_1)^{m_1} \cdots (t - \lambda_s)^{m_s}$$

我们只需要求解对应的广义特征方程, 即

$$(T - \lambda)^m x_n = 0, \qquad (n \geqslant m), \tag{22}$$

对此, 我们有下述平行于定理 3 的结果：

定理 5 : 设 λ 为常数, 则 $(T - \lambda)^m x_n = 0\,(n \geqslant m + 1)$ 的通解为

$$x_n = \lambda^n (C_1 n^{m-1} + C_2 n^{m-2} + \cdots + C_m), \qquad (n \geqslant m) \tag{23}$$

其中 $C_1, \ldots, C_m$ 为任意常数.

证明： 若 $\lambda = 0$, 则此时我们要解的方程为 $T^m x = 0$. 根据定义容易看出, $T^m x = 0$ 当且仅当 $x_m = x_{m+1} = \cdots = 0$.

若 $\lambda \neq 0$, 令 $x_n = \lambda^n y_n$, 则容易看出 (参见文末最后一节练习 2), y_n 满足方程 $(\lambda T - \lambda)^m y_n = 0$, 即 $\lambda^m (T-1)^m y_n = 0$, 由于 $\lambda \neq 0$, 它进一步等价于

$$(T-1)^m y_n = 0.$$

注意到 $T - \lambda = T - 1 = \Delta$ 是向前差分算子 (即 $\Delta x_n = x_{n+1} - x_n$), 上述方程化简为 $\Delta^m x = 0$. 根据朱世杰招差公式 (参见 [8, 定理 3]), 我们推出 $y_n = y(n)$ 是一个次数不超过 $m-1$ 的多项式. 从而 $x_n = \lambda^n y_n$ 具有形式 (23). ■

至此, 常系数齐次递推关系的求解得到了圆满的解决.

非齐次情形

现在我们转向非齐次情形, 即我们要求解方程

$$P(T)x(n) = \lambda^n f(n), \qquad (n \geqslant d), \tag{24}$$

其中 λ 是一个非零常数, P, f 是复系数多项式, 且 P 的次数为 d.

先考虑 $\lambda = 1$ 的特殊情况, 即方程

$$P(T)x(n) = f(n), \qquad (n \geqslant d), \tag{25}$$

其中 P, f 是多项式, 次数分别为 d, m.

根据上一节的分析, 我们只要求出 $f(n)$ 生成的一个 T-不变子空间. 由于 $T = \Delta + 1$, 不难确定 $f(n)$ 的一个 Δ-不变子空间从而也是 T-不变子空间不变子空间为

$$\mathbb{C}_{m+1} = \{(x_0, \ldots, x_n, \ldots) \,|\, x_n = g(n),\ g\text{为次数不超过 } m \text{ 的复系数多项式}\},$$

即以次数不超过 m 的多项式为通项公式的数列. 由于差分算子 Δ 在 $\mathbb{C}_{m+1}$ 上的极小多项式为 t^{m+1}, 所以右平移算子 $T = \Delta + 1$ 在 $\mathbb{C}_{m+1}$ 上的极小多项式为 $(t-1)^{m+1}$. 因此对方程 (25) 的讨论分为以下两种情况：

(i) 若 $P(1) \neq 0$, 这意味着多项式 $P(t)$ 与 $(t-1)^{m+1}$ 互质, 从而用求一术可以解出一个多项式 $U(t)$, 使得

$$P(t)U(t) \equiv 1 \pmod{(t-1)^{m+1}}$$

于是 (25) 的解为 $x = U(T)f$. 注意到, 对任意的正整数 k 有, $T^k f(n) = f(n+k)$, 由此不难看出, 解 $x(n) = U(T)f(n)$ 是一个 m 次多项式.

(ii) 不然有 $P(t) = (t-1)^k Q(t)$, 其中 $Q(1) \neq 0$. 从而 $P(T) = (T-1)^k Q(T) = \Delta^k Q(T)$, 于是我们可以分两步求解 (25). 第一步, 用前述方法求出方程 $Q(T)z = f$ 的一个特解 $z = z(n)$, 这是一个 m 次多项式；第二步, 对第一步得到的 z, 求出方程 $\Delta^k x = z$ 的一个特解 x, 我们可以保证 $x(n) = [n(n-1)\cdots(n-k+1)]w(n)$, 其中 $w(n)$ 是一个 m 次多项式. 事实上, 根据反复应用杨辉恒等式的差分形式 (参见 [8, p. 71] 第 (16) 式), 我们有：若 $z(n)$ 的朱世杰招差公式 (参见 [8, p. 67] 定理 3) 为:

$$z(n) = a_m \binom{n}{m} + a_{m-1}\binom{n}{m-1} + \cdots + a_0,$$

则

$$\begin{aligned} x(n) &= a_m \binom{n}{k+m} + a_{m-1}\binom{n}{k+m-1} + \cdots + a_0\binom{n}{k} \\ &= [n(n-1)\cdots(n-k+1)] \cdot \sum_{s=0}^{m} \left[\frac{a_s s!}{(k+s)!} \binom{n-k}{s} \right] \end{aligned}$$

是方程 $\Delta^k x = z$ 的一个特解.

现在我们转向一般情形 (24) 的讨论. 根据文末练习 3 的结果, 借助变数替换 $x_n = \lambda^n y_n$, 方程 (24) 等价于关于 y 的方程

$$P(\lambda T)y = f, \tag{26}$$

而这个方程我们在上面已经解决了!

评注

评注 1：不难想到, 我们也有一个平行于定理 4 的类似结果, 此处从略. 这样的结果限定了我们这个方法的适用范围, 即只能应用于非齐次项为形如 (23) 的函数的线性组合的情形.

评注 2：在通常的教科书 (如 [16][12]) 中, 对于常系数线性递推关系的求解, 作者一般优先介绍母函数法, 当然也有其它方法, 如华罗庚先生在 [17] 第四章讲了包括母函数法在内的四个方法, 但正如华先生指出的：“虽然我们讲了四个方法, 但实质上并没有太多的差异, 讲来讲去绕不过分项分数 (注：即我们所谓的部分分式分解) 这一关. ”

当然, 也有作者为了绕开“部分分式分解”, 就采取“待定系数法”, 如高德纳 (Donald E. Knuth) 及其合作者就在书中说 [18, p. 17]：

> 对处理所有的常系数线性递推关系, 部分分式分解完全能够胜任. 然而, 为简单起见, 我们将介绍一个不同的方法, 这在许多旧一点的文献中都可以找见. 这个方法是基于试探解, 类似于微分方程的求解. 在某些例子中, 这个方法能够最快提供答案, 但其规则看起来像黑魔法, 为了理解这种“掐指一算”为何奏效, 疑惑的读者终究要回到作为其基础的部分分式理论.

我们要指出, 事实上本文介绍的新方法, 不仅可以避开待定系数法和母函数法, 甚至可以帮助我们重新理解作为其公共基础的部分分式理论. 限于篇幅, 此处我们不再展开.

应用举例

例 1.(取自中译本 [16, p. 189]) 求递推关系 $x_n = 5x_{n-1} - 6x_{n-2} + 7^n$, $(n \geqslant 2)$ 的一个特解.

解：令 T 为右平移算子, 则上述方程可写成 (注意, 右边非齐次项多出一个 7^2)

$$P(T)x_n = 7^2 7^n$$

其中 $P(T) = T^2 - 5T + 6$. 令 $x_n = 7^n z_n$, 则上述方程等价于

$$Q(T)z_n = 7^2 = 49$$

其中 $Q(T) = P(7T) = 49T^2 - 35T + 6$. 由于 $Q(1) = 49 - 35 + 6 = 20 \neq 0$, 所以我们只要求出一个多项式 $U(T)$ 使得

$$Q(T)U(T) \equiv 1 \mod (T-1).$$

当然我们可以用求一术, 但此处有更简单的方法. 因为上述条件等价于

$$Q(1)U(1) - 1 = 0$$

从而 $U(1) = \frac{1}{P(1)} = \frac{1}{20}$, 因此 $U(T) = \frac{1}{20}$ 是同余方程的一个特解, 从而 $z_n = U(T)49 = \frac{49}{20}$ 是 $Q(T)z_n = 49$ 的一个特解, $x_n = \frac{49}{20}7^n$ 是原方程的一个特解.

例 2(取自 [16, p. 341]). 求递推关系 $x_n = x_{n-1} + 2x_{n-2} + (-1)^n$, $(n \geqslant 2)$ 的通解.

解：分三步.

第一步, 求出一个特解. 令 T 为右平移算子, 则上述方程等价于

$$P(T)x_n = (-1)^{n+2} = (-1)^n$$

其中 $P(T) = T^2 - T - 2 = (T+1)(T-2)$. 令 $x_n = (-1)^n z_n$, 则上述方程等价于 z_n 的下述方程

$$Q(T)z_n = 1$$

其中 $Q(T) = P(-T) = (T-1)(T+2) = \Delta(T+2)$. 因此我们先求出 y_n 使得 $(T+2)y_n = 1$, 只要取常数列 $y_n = \frac{1}{3}$. 进一步求解 $\Delta z_n = y_n = \frac{1}{3}$, 即得 $Q(T)z_n = 1$ 的一个特解 $z_n = \frac{1}{3}n$. 从而原方程的一个特解为 $x_n = \frac{1}{3}n(-1)^n$.

第二步, 求出对应齐次方程的通解. 因为特征方程为 $P(T) = T^2 - T - 2 = (T+1)(T-2)$, 根据定理 3, 我们只要分别求出 $(T+1)x_n = 0$ 与 $(T-2)x_n = 0$ 的通解, 并作迭加, 根据定理 $2'$, 结果为

$$C_1(-1)^n + C_2 2^n, \qquad \text{其中 } C_1, C_2 \text{ 为任意复常数}$$

第三步, 根据线性原理, 得到非齐次方程的通解公式为：

$$\frac{1}{3}n(-1)^n + C_1(-1)^n + C_2 2^n, \qquad \text{其中 } C_1, C_2 \text{ 为任意复常数}$$

7.11　练习

1. 证明以下结果, 并说明这是求一术解整数同余方程的理论依据 (提示：左乘即行变换).

命题：设二阶整数矩阵

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

(i) 若二阶整数矩阵 B 使得

$$BA = \begin{bmatrix} 1 & u \\ * & * \end{bmatrix}$$

则 $x = u$ 满足同余方程 $ax \equiv 1 \pmod{b}$.

(ii) 若二阶整数矩阵 C 使得

$$CA = \begin{bmatrix} * & * \\ 1 & u \end{bmatrix}$$

则 $x = u$ 满足同余方程 $ax \equiv 1 \pmod{b}$.

2. 设 T 是作用在数列空间上的右平移算子, 证明, 对任意的多项式 P 与常数 λ, 有 $P(T)\lambda^n = \lambda^n P(\lambda T)$.

3. 利用本文介绍的方法, 求出斐波拉契数列 (Fibonacci sequence)

$$x_n = x_{n-1} + x_{n-2} (n \geqslant 2); \qquad x_0 = x_1 = 1$$

的通项公式. (读者可以参见 [14, p. 36])

致谢：三年前, 笔者第一次从史特格兹 (Steven Strogatz) 的科普著作 [14] 中见到从线性代数的观点来求解斐波拉契数列的通项公式, 令人耳目一新. 在本文的写作过程中, 笔者从与天津大学理学院刘云朋教授、中央民族大学理学院王兢教授、麻省工学院 (MIT) 数学系斯特朗 (Gilbert Strang) 教授、威斯康辛大学米尔沃基分校许光午教授、劳伦斯伯克利国家实验室邵美悦博士的讨论中受益良多, 特表感谢！笔者对当初追问待定系数法道理何在的那些学生深表感谢, 在某种程度上, 这篇文章首先是笔者对他们的一个答复.

参考文献

[1] 陈省身, 中国算学之过去和现在, 《科学》第 25 卷第 5、6 期, 上海, 1941 年.

[2] 丁同仁, 李承治. 常微分方程教程 [第 2 版]. 北京：北京大学出版社. 2004.

[3] Ronald Graham, Donald Knuth, and Oren Patashnik, *Concrete Mathematics* (2nd ed.). Reading, MA: Addison-Wesley Professional. 1994. 张明尧、张凡 (译). 具体数学. 北京：人民邮电出版社, 2013.

[4] Daniel H. Greene, Donald E. Knuth, *Mathematics for the Analysis of Algorithms* (3rd ed.), Birkhäuser, 1990.

[5] 龚昇, 张德健. 线性代数五讲——第五讲：向量空间在线性算子下的分解. 数学传播季刊, 32(2), 34–53, 2008.

[6] Kenneth Hoffman, Ray Kunze, *Linear Algebra* (2nd ed.), Prentice-Hall, 1971.

[7] 华罗庚. 高等数学引论 (第四册). 北京：科学出版社. 1998.

[8] 林开亮. 微积分之前奏 (或变奏)：高阶等差数列的求和. 数学传播季刊, 41(1), 61–79, 2017.

[9] George Pólya, *How to Solve It*(with foreword by John Horton Conway and added exercises), Princeton University Press, 2004. 涂泓、冯承天 (译). 怎样解题. 上海：上海科技教育出版社, 2011.

[10] Rogers Penrose, *The Road to Reality*, New York: Vintage Books, 2004.

[11] 钱宝琮. 中国数学史. 北京：科学出版社. 1964.

[12] Kenneth H. Rosen, *Discrete Mathematics and its Applications*(Seventh Edition), McGraw-Hill. 徐六通, 杨娟, 吴斌 (译), 陈琼 (改编). 离散数学及其应用 [本科教学版]. 北京：机械工业出版社, 2017.

[13] Gilbert Strang, Nice function, manuscript. 见其个人主页 http://www-math.mit.edu/ gs/

[14] Steven Strogatz, *The Calculus of Friendship: What a Teacher and a Student Learned about Life While Corresponding about Math*, Princeton University Press, 2009. 有三个中译本：李晓东 (译). 心中有数的人生. 沈阳, 万卷出版公司, 2010；蔡承志 (译). 学微积分, 也学人生. 台北, 远流出版公司, 2011；李晓东 (译). 微积分的人生哲学. 北京, 中国财政经济出版社, 2022.

[15] 许光午, 李宝. 大衍求一术的算法意义与分析. 2016.

[16] 吴文俊. 邓若鸿、吴天骄 (访问整理). 走自己的路——吴文俊口述自传. 长沙：湖南教育出版社. 2015.

[17] 伍鸿熙. 陈省身的伯克莱岁月. 收入纪念陈省身先生文集, 丘成桐等主编, 杭州, 浙江大学出版社, 2005.

第八章　求解常系数线性微分方程和差分方程的代数方法

8.1　引言

对非齐次常系数线性微分方程的特解的求法，通常教科书会介绍待定系数法 (也称为比较系数法，见 [2]) 或拉普拉斯 (Laplace) 变换法 (见 [2]) 或算子法（见 [3, 4]）。第一种方法虽然初等但略显繁琐，而后两种方法则比较微妙，学生理解起来往往有困难。

本文揭示了算子法的代数本质，从而建议了一种更简单的解法，将微分方程与差分方程的求解，转化为多项式的同余方程的求解，而后一问题可用秦九韶的“求一术”方便地求解。“求一术”是矩阵变换，因此这方法对于学习过线性代数的学生来说，是比较容易理解和掌握的。为推广到更一般情况，我们只需换元并利用指数平移定理。我们进一步指出，利用指数平移定理，可以对齐次方程的通解给出简单的推导。

8.2　对算子法的观察

我们先看彭罗斯 (Rogers Penrose) 的数学物理通俗名著 [5, pp. 493–494] 中的一个例子，在那里要求微分方程

$$u'' + u = x^5 \tag{8.1}$$

的一个特解。通过引入微分算子

$$D = \frac{\mathrm{d}}{\mathrm{d}x}, \tag{8.2}$$

方程 (1) 可以写成

$$(D^2+1)u = x^5. \tag{8.3}$$

于是对 (3) 式两边施以算子

$$\left(D^2+1\right)^{-1} = \left(1+D^2\right)^{-1} = 1 - D^2 + D^4 - D^6 + \cdots, \tag{8.4}$$

并注意到 x^5 的 6 阶以上的导数都等于 0, 就有

$$\begin{aligned} u &= (D^2+1)^{-1}x^5 \\ &= (1 - D^2 + D^4 - D^6 + \cdots)x^5 \\ &= (1 - D^2 + D^4)x^5 \\ &= x^5 - 20x^3 + 120x. \end{aligned} \tag{8.5}$$

这个巧妙的方法称为*算子法* (operational calculus method)，归功于英国自学成才的数学家、物理学家、电气工程师亥维赛（Oliver Heaviside，1850-1925）。其主要思想，就是直接对微分算子（如这里的 D^2+1）求逆。这相当于，在线性代数中，为求解线性方程 $Lu = f$，两边用 L 的逆 L^{-1} 作用，得到 $u = L^{-1}f$。在线性方程组的情况，这是很容易办到的，例如我们可以用矩阵变换的技巧求矩阵的逆；但在微分方程的情况（如上面的例子），微分算子取逆往往涉及无穷级数（如 (4) 式），这并不好理解。亥维赛当初提出这个方法，也一度遭到数学家的拒绝。

正所谓“不管白猫黑猫，能抓住老鼠的就是好猫”, 我们不妨来验证一下, (5) 所给出的解是否满足方程 (3)。为此将 $u = (1 - D^2 + D^4)x^5$ 代入 (3) 就有

$$\begin{aligned} (D^2+1)u &= (D^2+1)(1 - D^2 + D^4)x^5 \\ &= (1 + D^6)x^5 \\ &= x^5. \end{aligned}$$

没问题!

从前数学家在使用 Heaviside 算子法时，往往都会有验证这一步，因为中间有些步骤不严格。但事实上，我们可以完全去掉中间不严格的步骤，因为正如我们将要表明的，问题的*关键*在于（在上个例子中）：多项式 $(1 - D^2 + D^4)$ 恰好满足多项式同余方程

$$(D^2+1)(1 - D^2 + D^4) \equiv 1 \pmod{D^6}.$$

在整个验证过程中，我们只用到这条关键的性质。

这个观察将我们直接引向对一般方程的讨论。

8.3 非齐次项为多项式的情形

设给定常系数线性微分算子

$$L = P(D), \tag{8.6}$$

其中 $P(D) = D^n + a_1 D^{n-1} + \cdots + a_n$ 是一个多项式，D 是微分算子。为简单起见，我们假设 D 所作用的函数空间，是定义在整个数轴 $\mathbb{R}$ 上的复数值光滑函数空间 (当然, 我们完全可以确定 $P(D)$ 的定义域)。

我们先考虑右边为多项式的非齐次方程，即：

$$P(D)u = f, \tag{8.7}$$

其中 $f = f(x)$ 是复系数多项式, 次数为 m。

我们有下述：

观察 设 f 是次数为 m 的多项式，则 $D^{m+1}f = 0$.

假设我们能够求出某多项式 $U(x)$ 使得

$$P(x)U(x) \equiv 1 \pmod{x^{m+1}}, \tag{8.8}$$

即

$$P(x)U(x) - 1 = q(x)x^{m+1},$$

其中 $q(x)$ 为多项式。将 $x = D$ 代入上式，就有算子等式

$$P(D)U(D) - 1 = q(D)D^{m+1}.$$

上式两边作用在 f 上,（注意 $D^{m+1}f = 0$）就有

$$P(D)U(D)f - f = q(D)D^{m+1}f = q(D)0 = 0,$$

移项即有

$$P(D)U(D)f = f,$$

这就意味着 $u = U(D)f$ 是方程 (7) 的一个特解。

以上观察使得我们将微分方程 (7) 的一个特解之求解，转化为多项式同余方程 (8) 的求解。下面我们就来讨论 (8) 的求解。

容易看出，方程 (8) 可解的一个充要条件是 $P(x)$ 与 x^{m+1} 互素，这等价于 $P(0) \neq 0$。并且此时 (8) 的求解可以用多种方法求得，如求两个多项式的最大公因子的欧几里得算法，这里我们介绍一个等价的方法，即南宋数学家秦九韶 (1208–1268) 在《数书九章》中提出的“求一术”，见 [5, 6, 7]。

求一术是用于求解整数同余方程 $ax \equiv 1 \pmod{b}$，而这里所需的是其多项式版本。这个方法的基础可以表述为下述矩阵引理。

引理 1 设 $P(x), Q(x)$ 是非零多项式，令

$$A = \begin{bmatrix} P(x) & 1 \\ Q(x) & 0 \end{bmatrix}.$$

(i) 若存在以多项式为元素的二阶矩阵 B 使得

$$BA = \begin{bmatrix} c & R(x) \\ * & * \end{bmatrix},$$

其中 c 为非零常数,则 $U(x) = \frac{R(x)}{c}$ 满足同余方程 $P(x)U(x) \equiv 1 \pmod{Q(x)}$。

(ii) 若存在以多项式为元素的二阶矩阵 C 使得

$$CA = \begin{bmatrix} * & * \\ c & R(x) \end{bmatrix},$$

其中 c 为非零常数,则 $U(x) = \frac{R(x)}{c}$ 满足同余方程 $P(x)U(x) \equiv 1 \pmod{Q(x)}$。

证明：只证明 (i)，(ii) 类似可证。设

$$B = \begin{bmatrix} S(x) & T(x) \\ * & * \end{bmatrix},$$

则计算可得

$$\begin{aligned} BA &= \begin{bmatrix} S(x) & T(x) \\ * & * \end{bmatrix} \begin{bmatrix} P(x) & 1 \\ Q(x) & 0 \end{bmatrix} \\ &= \begin{bmatrix} P(x)S(x) + T(x)Q(x) & S(x) \\ * & * \end{bmatrix} \end{aligned}$$

于是，根据假设，我们有

$$\begin{bmatrix} P(x)S(x) + T(x)Q(x) & S(x) \\ * & * \end{bmatrix} = \begin{bmatrix} c & R(x) \\ * & * \end{bmatrix},$$

从而有

$$P(x)S(x)+T(x)Q(x)=c,$$

以及

$$S(x)=R(x).$$

将 $S(x)=R(x)$ 代入前面的式子, 就有

$$P(x)R(x)+T(x)Q(x)=c.$$

因为 c 为非零常数，等式两边除以 c，就得到

$$P(x)\frac{R(x)}{c}+\frac{T(x)}{c}Q(x)=1.$$

这意味着 $U(x)=\frac{R(x)}{c}$ 满足同余方程

$$P(x)U(x)\equiv 1 \pmod{Q(x)}.$$

证毕.

基于上述引理，我们可以写出求解同余方程

$$P(x)U(x)\equiv 1 \pmod{Q(x)}$$

的求一术如下 (注意, 对矩阵 A 左乘一个矩阵相当于对 A 做行变换，初等矩阵对应着初等变换)：

算法 1 (求解多项式同余方程的求一术) 首先写出

$$A=\begin{bmatrix}P(x) & 1\\ Q(x) & 0\end{bmatrix}.$$

然后对第一列的两个多项式用带余除法（用次数高的多项式除以次数低的），设得到的商为 $q(x)$，则次数高的那一行减去次数低的那一行对应元素的 $q(x)$ 倍；于是新得到的矩阵的第一列两个元素替换为第一次带余除法的除式与余式，重复之前的操作（第一列两个元素辗转相除决定出一系列行变换），直到某一步带余除法得到的余式为某非零常数 c（算法结束），此时用 c 的旁边的多项式 (即引理 1 中的 $R(x)$) 除以 c，即为所求的 $U(x)$。

例 1 用“求一术”求解同余方程

$$(x^2+1)U(x)\equiv 1 \pmod{x^6}.$$

解：

$$A=\begin{bmatrix} x^2+1 & 1 \\ x^6 & 0 \end{bmatrix}$$
$$\xrightarrow[r_2-(x^4-x^2+1)r_1]{x^6=(x^4-x^2+1)\cdot(x^2+1)-1} \begin{bmatrix} x^2+1 & 1 \\ -1 & -(x^4-x^2+1) \end{bmatrix}$$

[文字解释：对矩阵 A 的第一列的两个元素做带余除法，有 $x^6=(x^4-x^2+1)(x^2+1)-1$，从而对矩阵 A 做初等行变换：第二行 (r_2) 减去第一行 (r_1) 的 (x^4-x^2+1) 倍，于是得到第二个矩阵。]

注意到第一列已经出现了非零常数 -1，于是

$$U(x)=\frac{-(x^4-x^2+1)}{-1}=\left(x^4-x^2+1\right).$$

由此，立即可以求出 $u''+u=x^5$ 的一个特解为

$$u=U(D)\,x^5=\left(D^4-D^2+1\right)x^5=120x-20x^3+x^5,$$

这与之前截断无穷级数得到的结果一致，见 (5)。

上述求一术仅仅解决了非齐次项 f 为多项式，且微分算子的多项式 $P(x)$ 的常数项不等于 0 的情况。若 $P(0)=0$，则我们可以提取因子 $P(x)=x^kQ(x)$，其中 $Q(0)\neq 0$。于是方程 (7) 可以分为两步求解：

- 第一步，用求一术求解方程

$$Q(D)v=f, \tag{8.9}$$

  求出 v。

- 第二步，直接对 v 累次积分求解关于 u 的方程

$$D^ku=v. \tag{8.10}$$

至此，我们对非齐次项 f 为多项式的情况就讨论完了。

8.4 非齐次项为拟多项式的情形

现在我们转向 f 为拟多项式（即某个指数函数与多项式的乘积）情况下。我们将指出，通过一个简单的变换，拟多项式情况可以转化为已经解决的多项式情况。

现在设我们要讨论的方程是

$$P(D)u = \mathrm{e}^{\lambda x} f(x), \tag{8.11}$$

其中 $P(D)$ 是 D 的多项式，$f(x)$ 为一个 m 次多项式，λ 是常数。

我们需要的一个基本结果，是指数平移定理 (Exponential Shift Theorem)：

引理 2 对任意的光滑函数 v 与多项式 P，有

$$P(D)\mathrm{e}^{\lambda x} v = \mathrm{e}^{\lambda x} P(D+\lambda) v. \tag{$*$}$$

证明：我们先证明 $(*)$ 对 $P(D) = D$ 成立。事实上，根据 Leibniz 求导法则，我们有

$$\begin{aligned} D(\mathrm{e}^{\lambda x} v) &= D(\mathrm{e}^{\lambda x}) v + \mathrm{e}^{\lambda x} Dv \\ &= \lambda \mathrm{e}^{\lambda x} v + \mathrm{e}^{\lambda x} Dv \\ &= \mathrm{e}^{\lambda x} (D+\lambda) v \end{aligned}$$

于是用数学归纳法，我们可以证明，对任意的正整数 n 有，

$$D^n(\mathrm{e}^{\lambda x} v) = \mathrm{e}^{\lambda x} (D+\lambda)^n v.$$

由此不难推出 $(*)$ 对一切多项式 $P(D)$ 成立。**证毕**.

于是，若令 $u = \mathrm{e}^{\lambda x} v$，则根据指数平移定理 $(*)$，(11) 式可以化为关于 v 的方程

$$P\left(D+\lambda\right) v = f,$$

而这个方程是我们之前就已经解决了的。

注解 这个方法之所以有效，主要是因为拟多项式 $\mathrm{e}^{\lambda x} f(x)$ 在 D 的某一个有限维不变子空间内。另一方面，斯特朗 (Gilbert Strang)[8] 的一个结果表明，D 的任意有限维不变子空间都是拟多项式空间。这就表明，该方法本质上不能进一步推广到其他情形。

注解 将 Heaviside 算子法严格化的常用方法是基于 *Laplace* 变换, 但正如在 MIT 讲授了多年微分方程的 *Gian-Carlo Rota* 教授在《我希望讲授微分方程之前就学到的十个教训》[9] 一文中指出的：

第九个教训: *Laplace* 变换的恰当引入

通常我们是在考虑常系数线性微分方程的初值问题时引入 Laplace 变换，然而这种动机非常勉强：对 *Laplace* 变换取逆并非一件容易的事，而且初值问题也可以用其它方法求解。我不知道如何以恰当的动机地引入 Laplace 变换……

8.5 推导齐次方程通解的简便方法

上一节所介绍的换元技巧，还有一个漂亮的理论应用，即可以用来求解最基本的齐次方程

$$(D-\lambda)^m u = 0. \tag{8.12}$$

注意到，对 $m=1$ 的情况，(12) 就是特征方程

$$Du = \lambda u.$$

因此 (12) 可称之为广义特征方程。众所周知，特征方程的通解为 $u = c_0 \mathrm{e}^{\lambda x}$（其中 c_0 为任意常数）。

下面我们将证明一个更一般的结论：

定理 1 广义特征方程 (12) 的通解为

$$y = \mathrm{e}^{\lambda x}(c_0 + c_1 x + \cdots + c_{m-1}x^{m-1}), \tag{8.13}$$

其中 $c_0, c_1, \ldots, c_{m-1}$ 为任意常数。

证明：根据指数平移定理 $(*)$，在变量替换 $u = \mathrm{e}^{\lambda x} v$ 之下，方程 (12) 等价于 $D^m v = 0$。容易推出，后者的通解为

$$v = c_0 + c_1 x + \cdots + c_{m-1}x^{m-1},$$

其中 $c_0, c_1, \ldots, c_{m-1}$ 为任意常数。于是，广义特征方程 (12) 的通解 $u = \mathrm{e}^{\lambda x} v$ 具有 (13) 的形式。**证毕**.

为推出常系数线性齐次方程的基本结果，我们需要线性代数中的下述基本结果。

定理 2 设 A 是域 F 上的向量空间 V 的线性变换，设 $P(x)$ 是 F 上的多项式，并且在 $F[x]$ 中有因式分解

$$P(x) = P_1(x) \cdots P_s(x),$$

其中 $P_1(x),\ldots,P_s(x)$ 两两互素。则有下述向量空间的直和分解：

$$\mathrm{Ker}(P(A)) = \mathrm{Ker}(P_1(A)) \oplus \cdots \oplus \mathrm{Ker}(P_s(A)),$$

其中 $\mathrm{Ker}(P(A)), \mathrm{Ker}(P_1(A)), \ldots, \mathrm{Ker}(P_s(A))$ 分别表示对应算子 $P(A), P_1(A), \ldots, P_s(A)$ 的核空间。

这个结果是复向量空间关于线性变换的根子空间 *(*即广义特征向量子空间*)* 分解（参见 [10, p. 309] 定理 12）的基石，考虑到其重要性与简单性，我们给出一个证明 (在高等代数之外的课程，如微分方程、离散数学中，也许很有必要复习一下这个基本定理)。

证明: 容易看到, 我们只需证明 $s = 2$ 的情况. 于是我们设 $P(x) = P_1(x)P_2(x)$，其中 $P_1(x)$ 与 $P_2(X)$ 最大公因式等于 1. 根据 Bezout 等式, 存在多项式 $Q_1(x), Q_2(x)$ 使得

$$P_1(x)Q_1(x) + P(x)Q_2(x) = 1.$$

在上式中令 $x = A$，就有

$$P_1(A)Q_1(A) + P_2(A)Q_2(A) = 1,$$

其中 1 表示 V 上恒同变换. 对每个满足 $P(A)v = 0$ 的向量 v，我们可以写

$$v = v_1 + v_2,$$

其中 $v_1 = P_2(A)Q_2(A)v, \quad v_2 = P_1(A)Q_1(A)v$. 则有

$$\begin{aligned} P_1(A)v_1 &= P_1(A)(P_2(A)Q_2(A)v) \\ &= Q_2(A)(P_1(A)P_2(A))v \\ &= Q_2(A)P(A)v = 0, \end{aligned}$$

类似的,

$$\begin{aligned} P_2(A)v_2 &= P_2(A)(P_1(A)Q_1(A)v) \\ &= Q_1(A)(P_2(A)P_1(A))v \\ &= Q_1(A)P(A)v = 0. \end{aligned}$$

令一方面, 若 v 满足 $P_1(A)v = 0$ 且 $P_2(A)v = 0$，则

$$v = Q_1(A)(P_1(A)v) + Q_2(A)(P_2(A)v) = 0.$$

证毕.

现在定理 2 中令 V 为 $\mathbb{R}$ 上复值光滑函数空间，取 A 为作用在其上的微分算子 D，并结合定理 1，即可得到常系数高阶线性齐次方程的下述基本结果（参见文献 [11, p. 176] 推论 4 或文献 [12, p. 198] 第 198 页定理 6.6）：

定理 3 设复系数多项式 $P(x)$ 有分解

$$P(x) = (x - \lambda_1)^{m_1} \dots (x - \lambda_s)^{m_s},$$

其中 $\lambda_1, \dots, \lambda_s$ 两两互异。则微分方程

$$P(D)\, u = 0$$

的通解为

$$u = \sum_{k=1}^{s} \mathrm{e}^{\lambda_k x} p_k(x),$$

其中 $p_k(x)$ 为次数小于 m_k 的复系数多项式。

8.6 举例

以下我们对 3，4，5 节所讨论的情况各举一二例，解释这些方法的具体应用。为与现行教材中常用的记号一致，我们改用 $y = y(x)$ 表示未知函数，其它记号也做了适当调整。下面的例 2 与例 3 取自 [2] 第 4 章例 7 与例 9, 有兴趣的读者可以比较这里的解法与书上的解法。

例 2 求方程

$$y'' - 2y' - 3y = 3x + 1$$

的一个特解。

解：由于右边的非齐次项是一次多项式，所以只需解同余方程

$$(x^2 - 2x - 3)u(x) \equiv 1 \pmod{x^2}.$$

以下用求一术来解此同余方程

$$\begin{aligned}
&\begin{bmatrix} x^2-2x-3 & 1 \\ x^2 & 0 \end{bmatrix} \\
\xrightarrow[r_1-1r_2]{x^2-2x-3=1\cdot x^2+(-2x-3)} &\begin{bmatrix} -2x-3 & 1 \\ x^2 & 0 \end{bmatrix} \\
\xrightarrow[r_2+(\frac{1}{2}x)r_1]{x^2=-\frac{1}{2}x\cdot(-2x-3)+(-\frac{3}{2}x)} &\begin{bmatrix} -2x-3 & 1 \\ -\frac{3}{2}x & \frac{1}{2}x \end{bmatrix} \\
\xrightarrow[r_1-\frac{4}{3}r_2]{-2x-3=\frac{4}{3}\cdot(-\frac{3}{2}x)+(-3)} &\begin{bmatrix} -3 & 1-\frac{2}{3}x \\ -\frac{3}{2}x & \frac{1}{2}x \end{bmatrix}.
\end{aligned}$$

于是，我们得到

$$u(x)=\frac{1-\frac{2}{3}x}{-3}=\frac{2}{9}x-\frac{1}{3}.$$

因此，原微分方程的一个特解为

$$\begin{aligned}
y &= u\left(D\right)\left(3x+1\right) \\
&= \left(\frac{2}{9}D-\frac{1}{3}\right)\left(3x+1\right) \\
&= -x+\frac{1}{3}.
\end{aligned}$$

例 3 求方程

$$y'''+3y''+3y'+y=\mathrm{e}^{-x}(x-5)$$

的一个特解。

解：这是右边为拟多项式的情况。如上文所述，作变量替换 $y=\mathrm{e}^{-x}z$ ，方程转化为 $z'''=x-5$ ，可取特解

$$z=\frac{1}{24}x^4-\frac{5}{6}x^3,$$

从而得到原方程的一个特解

$$y=\mathrm{e}^{-x}\left(\frac{1}{24}x^4-\frac{5}{6}x^3\right).$$

例 4 求方程

$$y'''+3y''-4y=0$$

的通解。

解：令 $P(D)=D^3+3D^2-4$，则原方程为 $P(D)\,y=0$。不难算出有因式分解

$$P(D)=(D-1)(D+2)^2,$$

令

$$P_1(D)=D-1,\qquad P_2(D)=(D+2)^2.$$

则根据定理 1 有，

$$\operatorname{Ker}P_1(D)=c_1\mathrm{e}^{x},\qquad \operatorname{Ker}P_2(D)=(c_2+c_3x)\,\mathrm{e}^{-2x}$$

从而根据定理 2, 原方程的通解为

$$y=c_1\mathrm{e}^{x}+(c_2+c_3x)\,\mathrm{e}^{-2x},$$

其中 c_1,c_2,c_3 为任意常数.

例 5 求方程 $y''+y=0$ 的通解。

解: 令 $P(D)=D^2+1$，则原方程为 $P(D)\,y=0$。不难算出有因式分解

$$P(D)=(D+i)(D-i).$$

其中 i 是虚数单位。由定理 1，$(D\pm i)$ 的核空间分别由 e^{-ix} 与 e^{ix} 生成, 进而由定理 2, $P(D)$ 的核空间由 e^{-ix} 和 e^{ix} 生成, 换言之, 方程 $y''+y=0$ 的通解为

$$y=c_1\mathrm{e}^{-ix}+c_2\mathrm{e}^{ix},$$

其中 c_1,c_2 为常数.

注. 对于例 5 所讨论的方程 $y''+y=0$, 通常教材中给出的通解是正弦函数 $\sin x$ 与余弦函数 $\cos x$ 的线性组合，而我们这里给出的表达式是虚指数函数 e^{-ix} 与 e^{ix} 的线性组合。两者其实等价，因为我们有著名的 *Euler* 公式:

$$\mathrm{e}^{ix}=\cos x+i\sin x.$$

8.7 差分方程的情况

不难发现，本文讨论的方法，也适用于求解常系数线性差分方程（在离散数学中，也称为递推关系）。我们留给有兴趣的读者自行探究，对此，读者可以参考 [13] 或 [11] 第 25.5 节或 [14] 第 5.3 节。此处我们仅满足于给出一个对照表，它扩充了我们在 [15] 中给出的离散–连续字典 (摘引如下)。

差分方程	微分方程
数列 $u(n)$	函数 $u(x)$
差分算子 Δ	微分算子 D
若 $\Delta^{m+1}u=0$，则 u 是 m 阶以下等差数列	若 $D^{m+1}u=0$，则 u 是 m 次以下多项式函数
朱世杰招差公式	*Taylor* 展开公式
$P(E)u=f$, 其中 $\Delta^{m+1}f=0$ $P(E)U(E)\equiv 1 \pmod{\Delta^{m+1}}$ 并令 $u=U(E)f$	$P(D)u=f$, 其中 $D^{m+1}f=0$ $P(D)U(D)\equiv 1 \pmod{D^{m+1}}$ 并令 $u=U(D)f$
$P(E)v(n)=\lambda^n f(n)$, 令 $v(n)=\lambda^n u(n)$, 则有 $P(\lambda E)u=f$	$P(D)v(x)=\mathrm{e}^{\lambda x}f(x)$, 令 $v(x)=\mathrm{e}^{\lambda x}u(x)$, 则有 $P(D+\lambda)u=f$
$P(E)\lambda^n=\lambda^n P(\lambda E)$	$P(D)\mathrm{e}^{\lambda x}=\mathrm{e}^{\lambda x}P(D+\lambda)$

注：$E=\Delta+1$ 是右平移算子，也可称为递推算子。

在通常的教科书 (如 [16]) 中, 对于常系数线性递推关系的求解, 作者一般优先介绍母函数法. 当然也有其它方法, 如华罗庚先生在 [17] 第四章讲了包括母函数法在内的四个方法, 但正如华先生指出的：“虽然我们讲了四个方法, 但实质上并没有太多的差异, 讲来讲去绕不过部分分式分解这一关. ”

有作者为了绕开“部分分式分解”而采取“待定系数法”, 但其缺点也是很显著的, 诚如格林 (Daniel H. Greene) 和高德纳 (Donald E. Knuth) 在 [18, p. 17] 中所说：

> 对处理所有的常系数线性递推关系, 部分分式分解完全能够胜任. 然而, 为简单起见, 我们将介绍一个不同的方法, 这在许多旧一点的文献中都可以找见. 它是基于试探解, 类似于微分方程的求解. 在某些例子中, 该方法能够迅速给出答案, 但其规则看起来像黑魔法 *(black magic)*, 为了理解这种“掐指一算 *(rules of thumb)*”为何奏效, 疑惑的读者终究要回到作为其基础的部分分式理论.

我们要指出, 事实上本文介绍的代数方法, 不仅可以避开待定系数法和母函数法, 还可以帮助我们重新理解作为其公共基础的部分分式理论. 限于篇幅, 此处我们不再展开. 在数论中有一类似问题, 即将分数化为分母更小的分数之和, 对此可参见 [19]。

在历史上, 微分方程与差分方程的理论一直平行发展, 例如 Heaviside 算子法的先驱布尔 (George Boole, 1815–1864) 在 1859–1860 年出版了《微分方程通论》[20] 和《差分方程通论》，都曾作为剑桥大学的教材。更让人惊讶的是, 布尔在 [21, p. 108] 中写道：

> 只是在写作本书时我才获悉 Lobatto 先生的杰作 *Théorie des Caractéristiques*, 它于 1837 年在阿姆斯特丹出版. 该书包含了本节的内容, 以及微分方程中的类似定理, 一言一概之, 它包含了自那一两年以后在英格兰重现发现的关于常系数线性微分方程的整个理论, 该理论发表在《剑桥数学杂志》(Cambridge Mathematical Journal) 的前两卷. 每一个英国数学将欣喜地看到我对 Lobatto 先生的公正.

这里布尔提到的 Lobatto 先生是荷兰数学家 *Rehuel Lobatto* (1797–1866). 蒙好友吴帆老师告知：Lobatto 之著作《特征理论》的完整名称，翻译过来是《论数学分析中所用的特征理论》，其中的主题“特征”是指微分算子与差分算子。作者在书中提出，可以将“特征”与函数剥离开来独立考虑，可以考虑“特征”本身的多项式函数、甚至更一般的解析函数。

华东师范大学刘治国教授曾评论说: “Boole 对算子演算有非常大的贡献，他发现了 Ramanujan 主定理的雏形。可惜的是，他没有用很多精彩的例子来诠释他的公式，以至于并没有引起人们的关注。”

8.8 总结

一. 与通常教科书上给出的解释相比，本文的阐述避免了算子法的形式化讨论与对形式解的验证，突出了算子法所隐含的代数本质，特别是它与中国古代数学“求一术”的联系，因此更容易理解。本文所建议的求解常系数线性微分方程与差分方程的代数方法，将线性代数的基本结果应用到微分方程的求解，并吸收了中国传统数学的杰出成就，可供各位同仁在高等代数、线性代数、常微分方程、离散数学课程中作为相关章节的课外阅读材料介绍给学生。

二. 本文所建议的新思路有一个平行的数论版本，即著名的 RSA 解密算法，参见 [22, 23]。基本的事实在于，不论是带线性变换（微分算子或差分算子）的线性空间，还是有限生成的 Abel 群，它们都是欧几里得整环上的模，因此可用欧几里得算法（求一术）。

三. 基于整数矩阵的 Smith 标准型，求一术可以推广到求解丢番图线性方程

组（参见 [24]），类似的，基于多项式矩阵的 Smith 标准型，这里的方法也可以推广到求解常系数线性微分方程组与常系数线性递推关系。这一点已经为 Bourbaki 写进他们的实变函数著作 [25]，I. Gohberg, P. Lancaster, L. Rodman[26] 则同时考虑了微分方程与差分方程 (见章节 $S1$)。整数环 $\mathbb{Z}$ 上的矩阵方程与常系数线性微分方程组之间的这种平行关系，是国内的某些微分方程著作 (参见 [3, 29, 30]) 未曾认识到的，他们曾用域上线性方程组的解法 (如消元法) 来解微分方程组。

四. 在理论物理中常有一些拟设 (ansatz), 如 Bethe ansatz, 这其实就是某种待定系数法, 因此一个自然的问题是, 这里的代数视角可否用于理解种种拟设? 也许更重要的一个问题, 是如何发展费曼 1951 年提出的算子法，他在 [31] 中曾坦承“这超出我的能力”。

8.9 后记

本文的写作缘起于我们的教学经历。在我们讲授微积分和常微分方程课程时，通常是没有任何先兆地介绍求特解的待定系数法，常常给学生一种从天而降不讲道理的感觉，于是总有爱思考而不服气学生追问，待定系数法道理何在？这个问题让我们困惑了很久。我们本人也对待定系数法不太满意，记得我们自己当学生时，就记不住种种情况下该怎么设未知函数的形式。(最近我们碰到以前教过的学生，正准备考研，临考了仍然担心记不住。)

于是我们一直在考虑，有没有更好的办法，不用死记硬背。直到作者之一读到了彭罗斯 (Rogers Penrose) 的著作 [5]，了解到 Heaviside 的算子法。很快，我们就认识到，这个方法可以严格化，而且归结为一个可以用秦九韶“求一术”解决的典型问题。

应该指出，我们此处应用求一术，乃深受吴文俊先生 *(1919–2017)* 的启迪。吴先生对中国古代数学的真知灼见 (参见 [27, 28]) 令我们受益匪浅，本文即是一个明证。今年恰是吴先生诞辰一百周年，我们谨以此文作为纪念。吴先生曾用笔名“顾今用”(见 [27]), 寓意“古为今用”，作者期望这一工作能得到他的认同。

同时这更是对当初追问我们的那些学生 (谢谢你们) 的一个答复，希望你们满意。相信你们懂配图的意思：讲数学要“以理服人”。这正是我们共同的初心。

致谢

感谢 MIT 数学系 *Gilbert Strang* 教授鼓励并指引参考文献 [8]，感谢学友吴帆老师提供关于 Lobatto 的资料，感谢华东师范大学刘治国教授分享他对 Boole 的评论、感谢北京大学李承治教授、北京师范大学刘玉明教授、上海大学李常品教授、北京市朝阳区教育研究中心张浩博士、渭南师范学院赵教练教授、同济大学朱善军同学、西北农林科技大学刘帅教授、杨变霞教授、赵子轩同学对初稿提出宝贵建议。

本文作者林开亮得到国家自然科学基金数学专项天元基金的支持，王兢得到中央民族大学 *2016* 年科研奖励基金的支持，特表感谢！

参考文献

[1] 王高雄等编. 常微分方程（第三版）[M]. 北京: 高等教育出版社,2006:144–155.

[2] 王柔怀, 伍卓群. 常微分方程讲义 [M]. 北京: 人民教育出版社,1964:122–138.

[3] 徐利治等编. 大学数学解题法诠释 [M]. 合肥: 安徽教育出版社,1999:585–588.

[4] Rogers Penrose. The Road to Reality[M].New York:Vintage Books,2004:493–494.

[5] 钱宝琮. 中国数学史 [M]. 北京: 科学出版社, 1964:206–209.

[6] 吴文俊. 从《数书九章》看中国传统数学构造性与机械化特色. 收入秦九韶与《数书九章》[M]. 北京: 北京师范大学出版社, 1987: 73-88.

[7] 林开亮. 求一术与方程术. 数立方网站.

[8] Gilbert Strang, Nice function, manuscript. 见其个人主页.

[9] Gian-Carlo Rota, Ten lessons I wish I had learned before I started teaching differential equations, Invited address delivered at the meeting of the Mathematical Association of America, Simmons College, 1997. (中译文将发表)

[10] 北京大学数学系几何与代数教研室前代数小组编. 高等代数（第四版）[M]. 北京: 高等教育出版社, 2013:309.

[11] V. I. Arnold, 常微分方程 [M]. 沈家骐, 周宝熙, 卢亭鹤译. 北京: 科学出版社, 2001:172–178.

[12] 丁同仁, 李承治. 常微分方程教程 [M]. 北京: 高等教育出版社,2004:197–203.

[13] 林开亮. 解常系数线性微分方程和递推关系的新方法——秦九韶和亥维赛的遗产[J]. 数学传播,2019(2):63–79.

[14] Gilbert Strang, 线性代数及其应用 [M]. 侯自新, 郑仲三, 张延伦译. 天津: 南开大学出版社,1990:215–239.

[15] 林开亮, 从《射雕英雄传》到《四元玉鉴》, 好玩的数学.

[16] Ronald Graham, Donald Knuth, and Oren Patashnik, *Concrete Mathematics* (2nd ed.). Reading, Massachusetts: Addison-Wesley. 1994. 张明尧、张凡 (译). 具体数学. 北京：人民邮电出版社, 2013.

[17] 华罗庚. 高等数学引论 (第四册). 北京：科学出版社. 1998.

[18] Daniel H. Greene, Donald E. Knuth, *Mathematics for the Analysis of Algorithms* (3rd ed.), Birkhäuser, 1990.

[19] 林开亮, 从射雕到九章——在天大理学院物理系的通俗报告, 好玩的数学.

[20] George Boole, *A Treatise on Differential Equations*, Cambridge: Macmillan. 1859.

[21] George Boole, *A Treatise on the Calculus of Finite Differences*, Cambridge: Macmillan. 1860.

[22] 林开亮. RSA 解密算法, 数立方网站.

[23] J. H. Silverman,《数论概论》（第 4 版）[M]. 孙智伟等译. 北京: 机械工业出版社,2016:78–80.

[24] 林开亮, 从《射雕英雄传》到《九章算术注》, 好玩的数学.

[25] N. Bourbaki, *Functions of a Real Variable: Elementary Theory*, Springer, 2004: 194–198.

[26] I. Gohberg, P. Lancaster, L. Rodman, *Matrix Polynomials*, SIAM, 2009.

[27] 顾今用. 中国古代数学对世界文化的伟大贡献 [J]. 数学学报,1975(1):18–23.

[28] 李文林. 古为今用的典范——吴文俊教授的中国数学史研究 [J]. 北京教育学院学报,2001(2):1–5.

[29] 姜福德, 关于用消元法解常系数线性微分方程组的问题, 工科数学, 1995 年第 1 期.

[30] 路玉梅、冯依虎, 运用矩阵消元法求解常系数线性微分方程组, 陇东学院学报, 2015 年第 5 期.

[31] Richard Feynman, An operator calculus having applications in quantum electrodynamics. Phys. Rev. 84, 108– 128 (1951).

第九章　论亥维赛算子法的合理性

9.1　回顾

如 [1] 所述，我们可将常系数线性微分/差分方程的求解，归结为一个多项式同余方程，而后者可以用秦九韶的“求一术”来求解。作为回顾，我们仍然选取彭罗斯(Rogers Penrose) 在其数学物理通俗名著 [2, pp. 493–494] 中给出的例子。

问题 1 求微分方程

$$u'' + u = x^5 \tag{9.1}$$

的一个特解。

解 1 通过引入微分算子

$$D = \frac{\mathrm{d}}{\mathrm{d}\,x}, \tag{9.2}$$

方程 (1) 可以写成

$$(D^2 + 1)u = x^5. \tag{9.3}$$

于是对 (3) 式两边施以算子

$$\begin{aligned} \left(D^2+1\right)^{-1} &= \left(1+D^2\right)^{-1} \\ &= 1 - D^2 + D^4 - D^6 + \cdots, \end{aligned} \tag{9.4}$$

并注意到 x^5 的 6 阶以上的导数都等于 0, 就有

$$\begin{aligned} u &= (D^2+1)^{-1}x^5 \\ &= (1 - D^2 + D^4 - D^6 + \cdots)x^5 \\ &= (1 - D^2 + D^4)x^5 \\ &= x^5 - 20x^3 + 120x. \end{aligned} \tag{9.5}$$

这个巧妙的方法称为算子法(operational calculus method)，归功于英国自学成才的数学家、物理学家、电气工程师亥维赛（Oliver Heaviside，1850–1925)。

亥维赛当初提出这个方法，一度遭到数学家的拒绝，因为它不太严格。比如，上述解法中其实隐含地使用了以下算子等式 (参见 (4) 式)

$$\boxed{(1+D)^{-1} = 1 - D + D^2 - D^3 + \cdots .}$$

如果将 D 替换为一个模长小于 1 的复数 z，那么上式毫无问题（这无非就是几何级数公式)；可是现在 D 是一个算子，右边的无穷级数的含义是要考虑的。亥维赛本人当然了解这些，不过他对严格化的问题并不感兴趣。从他的下述辩护可以看出，物理学家或工程师与数学家有不同的价值观 (引自网页)：

> 有两种数学, 严格的 (Rigorous) 与物理的 (Physical). 前一种数学很狭窄 (Narrow), 后一种数学大胆而宽阔 (Bold and Broad). 若非要停下来表述严格的证明，大多数的物理化数学进展 (physico-mathematical inquiries) 也将停下来。难道只是因为我不完全理解消化的机制，我就要绝食吗?

后来，数学家用其他途径将 Heaviside 算子法严格化，其中最典型的一种方案是用Laplace 变换(有点难)。

在 [1] 中，我们指出，其实上述 Heaviside 解法之核心在于，在 (5) 所给出的一连串等式中，我们真正需要的，是第三个

$$u = (1 - D^2 + D^4)x^5. \tag{9.6}$$

其中的算子多项式 $(1 - D^2 + D^4)$ 恰好满足多项式同余方程

$$(D^2 + 1)(1 - D^2 + D^4) \equiv 1 \pmod{D^6}. \tag{9.7}$$

这个关键性质，使得 (6) 所给出函数恰好满足 (3)。我们验证如下：

$$\begin{aligned}(D^2+1)u &= (D^2+1)(1-D^2+D^4)x^5\\ &= (1+D^6)x^5\\ &= x^5.\end{aligned}$$

从这个观察出发，我们可以得到关于求解常系数线性微分方程

$$P(D)u = f \tag{9.8}$$

的特解的一个算法，如下：

算法 1 给定微分方程 (8)，其中 $P(D)$ 是微分算子 D 的复（实）系数多项式，$f = f(x)$ 是 m 次复（实）系数多项式，$u = u(x)$ 是未知函数。则可按以下步骤求出 (8) 的一个特解 u。

(i) 用秦九韶“求一术”求出多项式同余方程

$$P(x)U(x) \equiv 1 \pmod{x^{m+1}} \tag{9.9}$$

的一个解 $U(x)$.

(ii) 令 $u = U(D)f$ 并化简，即得 (8) 的一个特解。

为完整起见，我们再将 [1] 中介绍的求解多项式同余方程

$$P(x)U(x) \equiv 1 \pmod{Q(x)} \tag{9.10}$$

的秦九韶“求一术”完善如下：

算法 2 (秦九韶“求一术”) 给定同余方程 (10)，其中 $P(x), Q(x)$ 是已知的多项式，且 $Q(x)$ 不是常数，$U(x)$ 是未知多项式。则可按以下步骤求解方程 (10)。首先写出

$$A = \begin{bmatrix} P(x) & 1 \\ Q(x) & 0 \end{bmatrix}.$$

然后对第一列的两个多项式用带余除法（用次数高的多项式除以次数低的），设得到的商为 $q(x)$，则次数高的那一行减去次数低的那一行对应元素的 $q(x)$ 倍；于是新得到的矩阵的第一列两个元素替换为第一次带余除法的除式与余式，重复之前的操作（第一列两个元素辗转相除决定出各个行变换），直到第一列某个数变成常数 c（算法结束），此时有下述结论：

(i) 同余方程 (10) 有解当且仅当 $c \neq 0$;

(ii) 当 $c \neq 0$ 时，用与 c 右边的多项式除以 c，就给出方程 (10) 的一个解 $U(x)$.

例 1 用秦九韶“求一术”求解同余方程

$$(x^2 + 1)U(x) \equiv 1 \pmod{x^6}. \tag{9.11}$$

解 2

$$A=\begin{bmatrix} x^2+1 & 1 \\ x^6 & 0 \end{bmatrix}$$

$$\xrightarrow[r_2-(x^4-x^2+1)r_1]{x^6=(x^4-x^2+1)\cdot(x^2+1)-1} \begin{bmatrix} x^2+1 & 1 \\ -1 & -(x^4-x^2+1) \end{bmatrix}$$

[文字解释：对矩阵 A 的第一列的两个元素做带余除法，有 $x^6=(x^4-x^2+1)(x^2+1)-1$，从而对矩阵 A 做初等行变换：第二行 (r_2) 减去第一行 (r_1) 的 (x^4-x^2+1) 倍，于是得到第二个矩阵。]

现在第一列已经出现了非零常数 -1，于是方程有解

$$U(x)=\frac{-(x^4-x^2+1)}{-1}=\left(x^4-x^2+1\right).$$

注意，例 1 给出的结果与 (7) 式吻合。这就给出了一种新的途径直接得到 (6)，从而回避了连等式 (5) 中第二行所涉及的无穷级数 (4)。

从某种意义上说，我们在 [1] 中的观点，是以秦九韶的立场来看 Heaviside，剥掉其不严格的分析皮相，直取其代数灵魂。这对习惯了代数 (或数论) 的读者来说是容易理解的，不过它没有说明 Heaviside 算子法为什么行得通——因为它本质上是以秦九韶朴素的代数取代了 Heaviside 绚丽的分析。

接下来我们设想，如果 Heaviside 本人看到我们的上述处理，他可能会如何重新解释他的工作。可以想见，他不大可能容忍我们将他那自然而天才的想法（见 (5) 式）直接抹去，那是要化作永恒的灵光一闪。

9.2 对 Heaviside 工作的重新解释

基于我们前面的理解，Heaviside 算子法之所以成功，在于它以一种神奇的方式求解了多项式同余方程

$$P(x)U(x)\equiv 1 \pmod{x^{m+1}}. \tag{9}$$

例如，在问题 1 中，Heaviside 算子法最核心的地方在于，求出了同余方程 (11)，即

$$(x^2+1)U(x)\equiv 1 \pmod{x^6}.$$

其绝妙之处在于，Heaviside 是用一种超越（相对于代数）的手法求出来的 (参见 (4) 式)：

$$\begin{aligned}\left(x^2+1\right)^{-1} &= \left(1+x^2\right)^{-1} \\ &= 1-x^2+x^4-x^6+\cdots \\ &\equiv 1-x^2+x^4 \pmod{x^6}.\end{aligned} \tag{9.12}$$

问题在于，上述做法现在看来仍然是形式的，我们还要说明它的合理性。而解决问题的关键，恰恰就在“形式”二字：我们要将 (12) 式中的幂级数

$$\left(1+x^2\right)^{-1} = 1-x^2+x^4-x^6+\cdots \tag{9.13}$$

理解为形式幂级数(formal power series)。

为突出这里的幂级数是形式幂级数（以区别于通常的函数幂级数），我们将通常表示变量 x 换成字母 X（“未定元”，根据Bourbaki的历史注记，在 1882 年，德国数学家Kronecker就完全清楚了“未定元”只是其代数的一个基元而不是分析意义下的变量。此前Gauss和Galois也有此认识。），正如我们在以形式的观点考察多项式时所做的那样。我们通常将复数域 $\mathbb{C}$ 上的多项式全体记为 $\mathbb{C}[X]$，而将 $\mathbb{C}$ 上的形式幂级数全体记为 $\mathbb{C}[[X]]$，其中的元素形如

$$\sum_{k=0}^{\infty} a_k X^k, \quad \text{其中 } a_k \in \mathbb{C},\ k=0,1,2,\ldots. \tag{9.14}$$

与 $\mathbb{C}[X]$ 一样，$\mathbb{C}[[X]]$ 中的形式幂级数也有加法、乘法，并满足通常的运算法则（如分配律），用代数的术语来说，它们都是环(ring)。加法的定义是显然的，而乘法的定义如同多项式那样，即通过分配律相乘再合并同次项得到。按照这个定义，不难验证，在 $\mathbb{C}[[X]]$ 中成立以下等式：

$$\boxed{(1-X)(1+X+X^2+X^3+\cdots)=1.} \tag{9.15}$$

由此不难推出，在 $\mathbb{C}[[X]]$ 中，也有

$$(1+X^2)(1-X^2+X^4-X^6+\cdots)=1, \tag{9.16}$$

这就是 (13) 式的等价表述。

根据 (16) 式，我们可以推出 (12) 式的结论: 在多项式环 $\mathbb{C}[X]$中成立

$$(X^2+1)(1-X^2+X^4) \equiv 1 \pmod{X^6}. \tag{9.17}$$

原因如下：

记 $1-X^2+X^4=H(X)$ 是形式幂级数

$$(X^2+1)^{-1}=(1-X^2+X^4-X^6+X^8-X^{10}+\cdots)$$

的前 6 项截断 (头部)，$-X^6+X^8-X^{10}+\cdots=T(X)$ 是其余项 (尾部)。我们有

$$(1-X^2+X^4-X^6+\cdots)=H(X)+T(X),$$

代入 (16) 式，有

$$(X^2+1)(H(X)+T(X))=1$$

根据分配律就有

$$(X^2+1)H(X)+(X^2+1)T(X)=1$$

注意到

$$T(X)=-X^6+X^8-X^{10}+\cdots=-X^6(1-X^2+X^4-\cdots)$$

代入上式就有

$$\begin{aligned}
&(X^2+1)H(X)-1\\
=&-(X^2+1)T(X)\\
=&-(X^2+1)\cdot\big(-X^6(1-X^2+X^4-\cdots)\big)\\
=&X^6\cdot\big((X^2+1)(1-X^2+X^4-\cdots)\big)\\
=&X^6\cdot 1\quad(\text{根据 (16)式})
\end{aligned}$$

注意到这实际上是多项式环 $\mathbb{C}[X]$中的等式，并且由此立即得到 (17) 式。

以上分析就证明了第一节所介绍的 Heaviside 算子法的合理性，这个分析具有一般性，作为秦九韶“求一术”的对比，我们也将这个结论提炼成一个算法。

算法 3 (“形式幂级数”法解同余方程) 给定同余方程

$$P(X)U(X)\equiv 1\pmod{X^m}. \tag{9.18}$$

其中 $P(X)\in\mathbb{C}[X]$ 是已知的多项式且 $P(0)\neq 0$，$U(X)$ 是未知多项式。则可按以下步骤求出方程的一个解。

(i) 根据定义用“待定系数法”求出 $P(X)$ 在 $\mathbb{C}[[X]]$ 中的逆 $P(X)^{-1}$ 的前 m 项，记为 $H_m(X)$，即

$$P(X)^{-1}=\left(\sum_{k=0}^{m-1}b_kX^k\right)+\cdots=H_m(X)+T_m(X).$$

(ii) 令 $U(X)=H_m(X)=\sum_{k=0}^{m-1}b_kX^k$，则 $U(X)$ 是 (18) 的一个解。

注解 $a_0\neq 0$ 是形式幂级数 (14) 可逆的充要条件。因此算法中的条件 $P(0)\neq 0$ 保证了多项式 $P(X)$ 在 $\mathbb{C}[[X]]$ 中可逆。而且，其逆可以用待定系数法求出，这只需要解一连串的一元一次方程。

证明：设 $P(X)$ 的次数为 d，则根据定义我们有

$$P(X)\left(H_m(X)+T_m(X)\right)=1$$

注意 $T_m(X)=X^mR_m(X)$，代入上式并展开就有

$$P(X)H_m(X)+P(X)X^mR_m(X)=1$$

移项就有

$$P(X)H_m(X)-1=X^m\cdot\left(P(X)R_m(X)\right)$$

注意等式左边是一个次数不超过 $d+m-1$ 次的多项式，根据形式幂级数相等的定义，右边必定也是一个多项式，从而 $P(X)R_m(X)$ 是次数不超过 $d-1$ 次的多项式，于是上式其实是多项式环 $\mathbb{C}[X]$ 中的等式。进而根据多项式同余的定义，我们有

$$P(X)H_m(X)\equiv 1 \pmod{X^m}. \qquad \textbf{证毕}$$

如果用上述“形式幂级数法”代替算法 1 中的秦九韶“求一术”(即算法 2)，那么新得到的算法就可以解释Heaviside算子法的合理性了。不过，在求解过程中，我们并不需要用到形式幂级数逆的完整表达式

$$(X^2+1)^{-1}=\sum_{k=0}^{\infty}(-1)^kX^{2k}, \tag{9.19}$$

而只需要算出其前 6 项截断

$$(X^2+1)^{-1}=1+0X-X^2+0X^3+X^4+0X^5+\cdots,$$

而这本质上只需要依次求解 6 个一元一次方程，得出各个系数。

可以想见，作为工程师的 Heaviside 不会认同我们将他的天才方法解释为求解一系列的一元一次方程。他根本就是利用了逆的整个表达式，如 (19)。那么他

是如何得到诸如此类的形式幂级数表达式的呢？也许，一个合乎情理的解释是：Heaviside 没有区分形式幂级数等式 (19) 与函数幂级数等式

$$(x^2+1)^{-1}=\sum_{k=0}^{\infty}(-1)^k x^{2k}, \tag{9.20}$$

其中 x 暂且视为一个模长小于 1 的数。

那么问题来了：Heaviside 不作区分是不是合理的呢？更具体地说，是不是一个函数幂级数等式 (如 (20)) 自然就蕴含了一个相应的形式幂级数等式 (如 (19))?

令人惊讶的是，回答是肯定的。而且，颇让人意外的是，这个事实曾被广泛应用，但我没能找到一个正式的表述。这里我们引用 MIT 数学教授Richard Stanley在其名著《计数组合学》第一卷 [3, p. 5] 中的说法：

> 例 1.15 对一个一般原理 (general principle) 给出了一个简单的示例。不正式地说 (informally speaking), 这个一般原理断言：如果我们有一个关于幂级数的等式，当它作为函数 (从而变量是模长充分小的复数) 时成立，那么当这个等式作为形式幂级数的等式仍然成立，只要等式中用到的运算对形式幂级数都有定义。对我们来说，此处给出这个原理的一个精确形式 (precise form) 将是不必要的一般化 (pandetic)，因为读者在任何特殊情形下检验我们对幂级数的操作的形式有效性方面将不会有困难。我们将用贯穿本节的几个例子诠释这个观点 (contention)。

Stanley 这里提到的例 1.15 是 (15) 式的下述推广：

$$(1-\alpha X)(1+\alpha X+\alpha^2X^2+\alpha^3X^3+\cdots)=1. \tag{9.21}$$

Stanley 接下来提到的诸多例子中，有两个值得了解。

例 2 在 $\mathbb{C}[[X]]$ 中成立以下等式

$$\left(\sum_{n=0}^{\infty}\frac{X^n}{n!}\right)\left(\sum_{n=0}^{\infty}\frac{(-1)^nX^n}{n!}\right)=1. \tag{9.22}$$

其函数化身是我们熟悉的指数函数等式 $e^xe^{-x}=1$。

例 3 套用二项展开式，我们可以对复数 λ 定义形式幂级数

$$(1+X)^{\lambda}=\sum_{k=0}^{\infty}\binom{\lambda}{k}X^k, \tag{9.23}$$

其中 $\binom{\lambda}{k} = \dfrac{\lambda(\lambda-1)\cdots(\lambda-k+1)}{k!}$。注意当 $\lambda = n$ 是自然数时，$(1+X)^\lambda = (1+X)^n$ 是多项式。特别地，对 $n=0$ 有 $(1+X)^0 = 1$。而对 $\lambda = -1$，有

$$\boxed{(1+X)^{-1} = 1 - X + X^2 - X^3 + \cdots.}$$

幂函数等式

$$(1+x)^\lambda (1+x)^\mu = (1+x)^{\lambda+\mu}$$

可以给出形式幂级数等式

$$(1+X)^\lambda (1+X)^\mu = (1+X)^{\lambda+\mu}. \tag{9.24}$$

注意，在 (24) 中令 $\lambda = 1, \mu = -1$ 就得到

$$(1+X)(1 - X + X^2 - X^3 + \cdots) = 1, \tag{9.25}$$

这只是 (15) 式的一个变形。

利用形式幂级数 (23) 的一个推广，原则上我们可以给出多项式同余方程

$$P(X)U(X) \equiv 1 \pmod{X^m}. \tag{18}$$

的一个公式解。为此，我们只要利用下述定义。

对满足 $F(0) = 0$ 的 $F(X) \in \mathbb{C}[x]$（实际上，只要 $F(x) \in \mathbb{C}[[X]]$），以及复数 λ，我们令

$$(1+F(X))^\lambda = \sum_{k=0}^{\infty} \binom{\lambda}{k} F(X)^k. \tag{9.26}$$

若同余方程 (18) 中的多项式 $P(x)$ 满足 $P(0) \neq 0$，不失一般性，可以假设 $P(0) = 1$, 从而 $P(x)$ 可以写成这样的形式

$$P(x) = 1 + F(x),$$

其中 $F(x) \in \mathbb{C}[X]$ 且显然有 $F(0) = 0$，于是根据定义 (26)，就有

$$\begin{aligned} P(x)^{-1} &= (1+F(X))^{-1} \\ &= 1 - F(X) + F(X)^2 - F(X)^3 + \cdots \end{aligned}$$

结合算法 3，我们可以给出以下结论：

定理 1 设 $P(x)=1+F(x)$ 是常数项等于 1 的多项式，则同余方程

$$P(X)U(X)\equiv 1 \pmod{X^m}. \tag{18}$$

的一个解 $U(X)$ 可以如下给出：

$$U(X)=1-F(X)+F(X)^2-\cdots+(-1)^{m-1}F(X)^{m-1} \tag{9.27}$$

具体计算时，$F(X)^k$ 的展开式中只要取次数小于 m 的项。作为例子，我们来计算当

$$F(X)=X+X^2=X(1+X)$$

且 $m=6$ 的简单情况。此时我们要计算的是

$$\begin{aligned}
U(X)=&1-F(X)+F(X)^2-F(X)^3+F(X)^4-F(X)^5\\
=&1-(X+X^2)+X^2(1+X)^2-X^3(1+X)^3\\
&+X^4(1+X)^4-X^5(1+X)^5\\
=&1-X(1+X)+X^2(1+2X+X^2)-X^3(1+3X+3X^2+[X^3])\\
&+X^4(1+4X+[X^2])-X^5(1+[X])\\
=&1-(X+X^2)+(X^2+2X^3+X^4)-(X^3+3X^4+3X^5+[X^6])\\
&+(X^4+4X^5+[X^6])-(X^5+[X^6])\\
=&1-X+X^3-X^4+[X^6].
\end{aligned}$$

其中 $[X^k]$ 表示次数大于等于 k 的项。

从而根据定理 1，我们得到同余方程

$$(X^2+X+1)U(X)\equiv 1 \pmod{X^6}$$

的一个特解为

$$U(X)=1-X+X^3-X^4.$$

平行于问题 1，利用上述结果我们可以求出微分方程

$$(D^2+D+1)u=f \tag{9.28}$$

的一个特解，其中 f 是任意一个 5 次多项式，比如 $f=f(x)=x^5$. 为此，只要令

$$u=U(D)f=(1-D+D^3-D^4)f.$$

当然，事实上，我们有更简单的方法求 (28) 的一个特解。为此注意到 $(D^2+D+1)(D-1)=D^3-1$，从而若令 $u=(D-1)v$，则只需 v 满足方程

$$(D^2+D+1)(D-1)v=f,$$

即

$$(D^3-1)v=f.$$

对于一个 5 次多项式 f，以上方程显然有解

$$v=-(D^3+1)f.$$

从而方程 (28) 有解

$$\begin{aligned} u&=(D-1)v\\ &=(D-1)\left(-(D^3+1)f\right)\\ &=\left((1-D)(D^3+1)\right)f\\ &=(1-D+D^3-D^4)f. \end{aligned}$$

这与之前得到的解一致。不难看出，同样的技巧可用于求下述方程的一个特解 (其中 f 是一个多项式):

$$(D^n+D^{n-1}+\cdots+D+1)u=f.$$

9.3 整数同余方程的幂级数解法

现在对于多项式同余方程

$$P(X)U(X)\equiv 1 \pmod{X^m}. \tag{18}$$

我们有两种解法，一种是秦九韶“求一术”(算法 2)，一种是形式幂级数解法 (算法 3)。回想起秦九韶“求一术”原本是求解整数同余方程

$$au\equiv 1 \pmod{b}, \tag{9.29}$$

自然地，我们要问，对这类整数同余方程，是否也有平行于算法 3 的解法呢？回答是肯定的，不过正如我们在算法 3 中为求解多项式环 $\mathbb{C}[X]$ 中的同余方程 (18) 需要深入到更广阔的形式幂级数环 $\mathbb{C}[[X]]$，为了在整数环 $\mathbb{Z}$ 中求解形如

$$au\equiv 1 \pmod{p^m}, \quad \text{其中 } p \text{ 是素数} \tag{9.30}$$

的同余方程，我们也需要进入一个更广阔的天地，那就是p-进整数(p-adic integers)的世界。在历史上，德国数学家亨泽尔 (Kurt Hensel, 1861–1941)正是通过与形式幂级数类比 (这是一个伟大类比中的一部分，推而广之，是Langlands 纲领)，在 1897 年引入了 p-进整数。简单地说，一个 p-进整数与形式幂级数 (14) 类似：

$$\sum_{k=0}^{\infty} a_k p^k, \quad \text{其中 } a_k \in \{0,1,\dots,p-1\},\ k=0,1,2,\dots. \tag{9.31}$$

注意到若取有限和，就得到正整数。所有的 p-进整数构成一个环，记为 $\mathbb{Z}_p$。注意，其中的加法与乘法比形式幂级数要复杂一些，因为涉及“进位”。然而，一个平行于 (15) 的基本关系式在 $\mathbb{Z}_p$ 中仍然成立:

$$\boxed{(1-p)(1+p+p^2+p^3+\cdots)=1.} \tag{9.32}$$

据此可以说明，同余方程

$$(1-p)u \equiv 1 \pmod{p^m}$$

的一个解是 $u_m = 1+p+\cdots+p^{m-1}$。例如，取 $p=5$，则对 $m=2$，可以看到 $u_2 = 1+5 = 6$ 满足

$$(1-p)u_2 = -4\cdot 6 = -24 \equiv 1 \pmod{5^2};$$

对 $m=3$，$u_3 = 1+5+5^2 = 31$ 满足

$$(1-p)u_3 = -4\cdot 31 = -124 \equiv 1 \pmod{5^3}.$$

当然，一般结论是显然的：

$$(1-p)(1+p+\cdots+p^{m-1}) = 1-p^m \equiv 1 \pmod{p^m}.$$

推而演之，我们有一套平行于形式幂级数的理论。特别地，也有算法 3 与定理 1 的平行结果。鉴于笔者也是门外汉，我们就不详细展开。对 p-进数有兴趣的读者，可阅读专门的著作。

我们乐于跟读者分享日本数学家加藤和也、黑川信重、斋藤毅在《数论 1: Fermat 的梦想》[4, p. 58–59] 中对这个奇妙的 p-进数世界的感知 (想必你很少有机会在数学书中读到如此富有诗意的文字，三位作者简直是“来自星星的你”)：

> p-进数最初由亨泽尔在 1900 年左右引进。在数学的历史长河中，一个数就是指一个实数，只是在不久以前我们才意识到存在一个 p-进数的世界. 这就好比一个只是在白天仰望天空的人突然看到了星光闪耀的夜空。p-进数与实数的数学风景截然不同。就好比是夜空的一颗星星，p-进数散发出“素数 p 的光辉”。我们白天无法看到它，是因为太阳——或者说实数——喷发着“实数的光亮”。正如夜空中有数不尽的点点繁星，对每个素数 p，都有一个 p-进数的星球。每颗星星之于太阳，就正如每个 p-进数世界之于实数世界。正如在夜晚我们可以对遥远天空中的物体看得更清楚，透过 p-进数，我们开始看到广袤深邃的数学宇宙。

据说，德国数学家Helmut Hasse (1898–1979)正是在逛书店时偶然发现了亨泽尔的一本数论书，为其中所介绍 p-进数而着迷，从而决定追随亨泽尔学习，最终成为一位大数学家。特别值得一提的是，1920 年 10 月，22 岁的 Hasse 发现了著名的“局部–整体原则”(local-global principle)，这成为他次年提交的博士论文主题。今年恰好是 Hasse 原理诞辰 100 周年。

9.4 形式幂级数的世界

也许，比 p-进数的世界更好理解的，是形式幂级数的世界。但即便是在这个更加熟悉的世界，也极少有人能够在其中游刃有余。一个显著的例外，是 100 年前英年早逝的印度数学家Ramanujan(1887–1920)。

他的合作者、英国数学家G. H. Hardy曾写道 [5]：

> 最令人惊奇的是，他 [Ramanujan] 对于代数公式、无穷级数变换的洞察力。在这方面，我敢说我从未遇到与他旗鼓相当的人，我只能将他与 Euler 或 Jacobi 相比……

Hardy 还说：“Ramanujan 是他那个时代中最伟大的形式主义者 (formalist)。”也许 Heaviside 可以视为 Ramanujan 的先驱。Ramanujan 被誉为“The man who knew infinity”(这是关于他的一部传记的书名，中译本译作《知无涯者》)，我们可以追问：百年以后，“The men who followed Ramanujan”都有谁？

9.5 结语

无论如何，这都是一件很奇妙的事情，形式幂级数的世界与函数幂级数的世界之间好像有一面魔镜。更奇妙的是，大多数人可能都没有意识到，自己一直待在 x 的世界里，而从未注意到镜中还有 X 的世界。[1]

致谢

感谢华东师范大学刘治国教授分享他对 Richard Stanley 所提及关于形式幂级数与函数幂级数的一般原理的评论，感谢首都师范大学李克正教授、北京市朝阳区教育研究中心张浩博士、中国矿业大学张汉雄教授、中国传媒大学陈见柯教授、东南大学丘成桐中心的张超博士、上海交通大学李吉有教授、渭南师范学院赵教练教授对初稿提出宝贵建议。张浩博士还对算法 3 给出了一个基于商环、同态等抽象代数基本概念的证明，李克正教授与张超博士对它则给出了基于完备局部环理论的理解，此处均未收入。作者之一感谢浙江工商大学洪海波教授与他反复讨论下述相关问题：对特征 0 的环中的幂零元 x(即满足 $x^m = 0$)，如何定义其指数函数 $\exp x$，又如何求 $\exp x$ 的逆。从某种意义上说，本文就是对这类问题的一个回应。

感谢审稿人对初稿提出诸多有益的建议。

[1] 想起一件小事，从前用的高等代数教材，有一章标题叫 λ-矩阵，也许改成 X-矩阵会与第一章的多项式观念更统一。

参考文献

[1] 林开亮、王兢，求解常系数线性微分方程和差分方程的代数方法，《内蒙古师范大学学报》2019 年第 6 期.

[2] Rogers Penrose, *The Road to Reality*. New York: Vintage Books, 2004.

[3] Richard P. Stanley, *Enumerative Combinatorics: Volume 1.* Cambridge: Cambridge University Press, 1997. 有中译本《计数组合学 (第一卷)》。

[4] Kazuya Kato, Nobushige Kurokawa, Takeshi Saito, *Number Theory 1: Fermat's Dream*. American Mathematical Soc., 2000. 有中译本。

[5] G. H. Hardy, 印度数学家 Ramanujan, 收入《一个数学家的辩白》，李文林等编译，大连理工大学出版社，2014 年。这里有英文原文。

第十章　从 $(xI_n - A)$ 的列变换矩阵求 A 的标准型基底

摘要

对有限维向量空间 V 上的线性变换 T, 我们提出了构造 V 的基的新算法, 使得 T 在该基下的矩阵表示为 Jordan 标准型和有理标准型. 这个算法给出的基与一个经典算法给出的基吻合. 然而, 此处的算法看起来更直接有效.

10.1　引言

设 V 是域 F 上的有限维向量空间, T 是 V 上的线性变换. 众所周知, 求 T 的有理标准型和 Jordan 标准型[1] 的问题, 归结为不变因子的计算. 而不变因子可以通过初等变换计算. 令 A 是 T 在 V 的一组基 $\{v_1, \dots, v_n\}$ 下的矩阵, 这里 $n = \dim V$. 对多项式矩阵 $xI_n - A$ 做初等行列变换, 可以得到 Smith 标准型. 依据熟知的算法, 可以求出可逆多项式矩阵 $P, Q \in \mathrm{GL}_n(F[x])$ 使得 $xI_n - A$ 化为 Smith 标准型:

$$P(xI_n - A)Q = \mathrm{diag}(1, \dots, 1, d_1, \dots, d_r). \tag{10.1}$$

其中 $d_1, \dots, d_r \in F[x]$ 是不变因子, 满足 $d_1 | d_1 | \cdots | d_r$. 有了不变因子, 就可以读出 A 的有理标准型. 将各个不变因子完全分解, 就可以读出 A 的 Jordan 标准型.

至此, 尚未给出 V 的一组基, 使得 T 在该基下的矩阵表示为有理标准型或 Jordan 标准型. 为此目的, 需要做更多的工作. 文献中已经给出了一些算法, 例如 [1, 3, 4] 描述了一种经典的算法, 从 $F[x]$-模的观点来看, 这也许是最自然的算法.

本文旨在给出一种新算法来求这些基, 其出发点依然是多项式矩阵 $xI_n - A$ 的 Smith 标准型. 与上述算法不同, 我们主要用到通过行列初等变换将 $xI_n - A$

[1] 但凡涉及 Jordan 标准型, 我们总假定 T 的特征多项式在 F 上完全分裂.

对角化的列变换矩阵.

10.2 Jordan 标准型的新算法

我们首先考虑 Jordan 标准型, 从而假定各个不变因子 $d_1, \dots, d_r$ 在 F 上完全分裂.

为简单起见, 通过对 V 中向量取基底 $\{v_1, \dots, v_n\}$ 下的坐标, 将 V 等同于 F^n. 于是 T 等同于 F^n 上由 A 的左乘给出的线性变换 L_A. 我们的构造将用到出现在 (10.1)中的多项式矩阵 Q 的各个列向量. 于是我们记

$$Q = (*, \dots, *, \xi_1, \dots, \xi_r), \qquad \xi_i \in F[x]^n. \tag{10.2}$$

注意到 Q 可以通过对增广矩阵 $\begin{pmatrix} xI_n - A \\ I_n \end{pmatrix}$ 做初等行列变换得到, 其中行变换只对矩阵 $xI_n - A$ 的 n 行做. 当 $xI_n - A$ 化成 Smith 标准型时, I_n 就变成 Q.

对 A 的每个特征值 $\lambda \in \mathrm{Spec}(A)$, 令它在 $d_1, \dots, d_r$ 中的重数依次为 $m_1(\lambda), \dots, m_r(\lambda)$. 于是我们有

$$d_i = \prod_{\lambda \in \mathrm{Spec}(A)} (x - \lambda)^{m_i(\lambda)}, \qquad 1 \leqslant i \leqslant r, \tag{10.3}$$

且

$$\chi_A(x) = \det(xI_n - A) = \prod_{i=1}^{r} d_i = \prod_{\lambda \in \mathrm{Spec}(A)} (x - \lambda)^{m(\lambda)}, \tag{10.4}$$

其中

$$m(\lambda) = m_1(\lambda) + \cdots + m_r(\lambda) \tag{10.5}$$

为特征值 λ 的代数重数.

为给出一组基, 使得它可以给出 Jordan 标准型, 我们先引入一些记号.

- 对多项式 $f \in F[x]$ 以及 $c \in F$ 与整数 $j \geqslant 0$, 令 $\langle f, (x-c)^j \rangle$ 表示 f 中关于 $x - c$ 的展开式中 $(x-c)^j$ 的系数.
- 对一个多项式列向量 $\xi = (f_1, \dots, f_n)^t \in F[x]^n$ (上标 t 表示转置) 以及 $c \in F$ 与整数 $j \geqslant 0$, 记 $\langle \xi, (x-c)^j \rangle = (\langle f_1, (x-c)^j \rangle, \dots, \langle f_n, (x-c)^j \rangle)^t$.

我们对 Jordan 基的算法如下.

定理 18 设 $P, Q \in \mathrm{GL}_n(F[x])$ 满足 (10.1), 将 Q 记为 (10.2), 假定给出了分解 (10.3) . 对每个 $\lambda \in \mathrm{Spec}(A)$ 以及每个使得 $m_i(\lambda) \neq 0$ 的 $i \in \{1, \ldots, r\}$, 令

$$\alpha_{ij}(\lambda) = \langle \xi_i, (x-\lambda)^j \rangle \in F^n, \qquad 0 \leqslant j \leqslant m_i(\lambda) - 1. \tag{10.6}$$

则以下断言成立:

(i) 向量

$$\{\alpha_{ij}(\lambda) : 0 \leqslant j \leqslant m_i(\lambda) - 1\} \tag{10.7}$$

线性无关, 并生成 F^n 的一个 L_A-不变子空间 $W_i(\lambda)$. 限制线性变换 $L_A|_{W_i(\lambda)}$ 在 $W_i(\lambda)$ 的基底 (10.7) 下的矩阵对应于特征值 λ 的 m_i 阶 Jordan 块:

$$J_{m_i}(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}. \tag{10.8}$$

(ii) F^n 是子空间 $W_i(\lambda)$ 的直和. 从而向量

$$\{\alpha_{ij}(\lambda) : \lambda \in \mathrm{Spec}(A), m_i(\lambda) \neq 0, 0 \leqslant j \leqslant m_i(\lambda) - 1\} \tag{10.9}$$

构成 F^n 的一组基, L_A 在这组基下为 Jordan 标准型.

评论 1 (i) 若 F 的特征为 0, 则向量 $\alpha_{ij}(\lambda)$ 可以如下求出: $\alpha_{ij}(\lambda) = \frac{1}{j!}\frac{\mathrm{d}^j}{\mathrm{d}x^j}\xi_i(x)\big|_{x=\lambda}$.

(ii) 显然, 若 M 由 (10.9)中的向量拼成的矩阵, 则它是使得 $M^{-1}AM$ 为 Jordan 标准型的过渡矩阵.

定理 2.1 的证明 (i) 的证明. 根据 (10.1) 以及 (10.2), 我们有

$$P(xI_n - A)\xi_i = (0, \ldots, d_i, \ldots, 0)^t. \tag{10.10}$$

现在 $(x-\lambda)^{m_i(\lambda)} \mid d_i$, 从而 $(x-\lambda)^{m_i(\lambda)} \mid (0, \ldots, d_i, \ldots, 0)^t$(在整除每一个分量的意义下) , 即有 $(x-\lambda)^{m_i(\lambda)} \mid P(xI_n - A)\xi_i$, 进而 $(x-\lambda)^{m_i(\lambda)} \mid P^{-1}P(xI_n - A)\xi_i = (xI_n - A)\xi_i$. 这就意味着 $(xI_n - A)\xi_i$ 关于 $(x-\lambda)$ 的展开式中, 低于 $m_i(\lambda)$ 次的系数都等于 0. 注意, 根据定义, 我们有

$$\xi_i \equiv \sum_{j=0}^{m_i(\lambda)-1} \alpha_{ij}(\lambda)(x-\lambda)^j \pmod{(x-\lambda)^{m_i}}. \tag{10.11}$$

而

$$(xI_n - A) = -(A - \lambda I_n) + I_n(x - \lambda), \tag{10.12}$$

从而不难算出

$$\langle (xI_n - A)\xi_i, (x-\lambda)^j \rangle = \begin{cases} -(A - \lambda I_n)\alpha_{i0}(\lambda), & j = 0 \\ \alpha_{i,j-1}(\lambda) - (A - \lambda I_n)\alpha_{ij}(\lambda), & 1 \leqslant j \leqslant m_i(\lambda) - 1 \end{cases} \tag{10.13}$$

这就推出

$$\begin{cases} (A - \lambda I_n)\alpha_{i0}(\lambda) = 0, \\ (A - \lambda I_n)\alpha_{ij}(\lambda) = \alpha_{i,j-1}(\lambda), \qquad 1 \leqslant j \leqslant m_i(\lambda) - 1 \end{cases} \tag{10.14}$$

为说明 $\alpha_{i0}(\lambda) \neq 0$, 我们注意到

$$\alpha_{i0}(\lambda) = \xi_i(\lambda) \tag{10.15}$$

而 $\xi_i(\lambda)$ 是可逆矩阵 $Q(\lambda)$ 的第 $n - r + i$ 列.

由此易得 (i) 中结论, 特别地, $L_A|_{W_i(\lambda)}$ 在基底 (10.7) 下的矩阵是 $m_i(\lambda)$ 阶 Jordan 块(10.8).[2]

为证明 (ii). 对每个特征值 λ, 令

$$W(\lambda) = \sum_{i\,:\,d_i(\lambda)=0} W_i(\lambda) \tag{10.16}$$

则显然有

$$W(\lambda) \subset \mathrm{Ker}(A - \lambda)^{m(\lambda)} = V(\lambda). \tag{10.17}$$

由根子空间分解定理[3], 我们有

$$V = \bigoplus_{\lambda} V(\lambda) \tag{10.18}$$

从而为证明结论, 我们只要证明 (10.16) 中的和是直和.

为此, 设

$$\sum_{i\,:\,d_i(\lambda)=0} c_i w_i = 0, \tag{10.19}$$

[2]在许多教材 (例如 [2, p. 236]) 中, Jordan 块不是写成(10.8), 而是写成它的转置. 从我们的处理来看, 似乎这里的形式更自然.

[3]参见 [2, p. 211] 定理 12. 注意此处我们其实进一步证明了 $\dim V(\lambda) = m(\lambda)$.

其中 $w_i \in W_i(\lambda), c_i \in F$.

设 w_i 被 $(A - \lambda I_n)$ 零化的指数为 q_i, 即

$$(A - \lambda I_n)^{q_i} w_i = 0, \qquad (A - \lambda I_n)^{q_i - 1} w_i \neq 0 \tag{10.20}$$

注意, $q_i > 1$, 并且 $(A - \lambda I_n)^{q_i - 1} w_i$ 是 $\alpha_{i0}(\lambda)$ 的一个非零倍数. 不妨设 $q_1 \geqslant q_2 \cdots \geqslant q_s > 1 = q_{s+1} = \cdots = q_r$. 将 (10.19) 两边以 $(A - \lambda I_n)^{q_1 - 1}$ 作用, 我们得到 $\alpha_{i0}(\lambda)$ 之间的一个线性关系. 然而我们注意到 $\alpha_{i0}(\lambda) = \xi_i(\lambda)$ 是可逆矩阵 $Q(\lambda)$ 的各个列, 从而线性无关. 于是我们得到零化指数为 q_1 的各个向量的系数都等于 0. 从而 (10.19) 中这些项扔掉, 得到一个化简的线性关系. 对这个化简的关系用 $(A - \lambda I_n)^{q_2 - 1}$ 作用, 得到进一步化简. 最终得到所有系数等于 0.

于是 (2.15) 是直和, 从而有

$$\bigoplus_{\lambda \in \mathrm{Spec}(A)} W(\lambda) = \bigoplus_{\lambda \in \mathrm{Spec}(A)} \bigoplus_{i\,:\, d_i(\lambda) = 0} W_i(\lambda).$$

由于右边子空间的维数之和显然是 $\sum_{\lambda \in \mathrm{Spec}(A)} m(\lambda) = \dim V$, 所以它就是全空间 V. 至此, **定理 2.1 证毕**.

评论 2

- 从证明可以看出, 为求出 A 的 Jordan 标准型, 我们真正需要的, 只是将 $(xI_n - A)$ 对角化, 而并不需要将它化成 Smith 标准型 (参见例 2.2).
- 不必假定 A 的特征多项式在基域 F 上完全分裂, 本算法的一个变形可以求出 A 在 F 上的全部特征值与特征向量. 结论是: A 在 F 上的特征值恰好是各个 d_i 在 F 上的根, 并且对应的特征向量由 ξ_i 在这个根处的取值给出 (见 [6]).
- 上述证明稍加推广, 即可给出任意域上的方阵 A 的广义 Jordan 标准型的基底.

例 1 设

$$A = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix},$$

我们依照上述算法来求 A 的 Jordan 标准型以及对应的 Jordan 基.

$$\begin{pmatrix} x & -1 & -1 & 1 \\ -1 & x & 1 & -1 \\ -1 & 1 & x & -1 \\ 1 & -1 & -1 & x \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & -1 & x \\ -1 & x & 1 & -1 \\ -1 & 1 & x & -1 \\ x & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & x-1 & 0 & x-1 \\ -1 & 0 & x-1 & x-1 \\ x & x-1 & x-1 & 1-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & x-1 \\ 0 & 0 & x-1 & x-1 \\ 0 & x-1 & x-1 & 1-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & x-1 \\ 0 & x-1 & x-1 & 2-x-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x-1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & x-1 \\ 0 & 0 & x-1 & 2-x-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & -x-1 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & 0 \\ 0 & 0 & x-1 & 3-2x-x^2 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & -x-2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x-1 & 0 & 0 \\ 0 & 0 & x-1 & 0 \\ 0 & 0 & 0 & (x-1)(x+3) \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & -x-2 \end{pmatrix}$$

从而 A 的不变因子为

$$d_1(x) = d_2(x) = x-1, \quad d_3(x) = (x-1)(x+3).$$

$$Q = (*, \xi_1, \xi_2, \xi_3) = \begin{pmatrix} * & 0 & 0 & 1 \\ * & 1 & 0 & -1 \\ * & 0 & 1 & -1 \\ * & 1 & 1 & -x-2 \end{pmatrix}.$$

A 的谱为 $\mathrm{Spec}(A) = \{1, -3\}$.

对应于特征值 $\lambda = 1$, 我们有

$$\alpha_{10}(1) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad \alpha_{20}(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad \alpha_{30}(1) = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -x-2 \end{pmatrix}_{x=1} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -3 \end{pmatrix}.$$

对应于特征值 $\lambda = -3$, 我们有

$$\alpha_{30}(-3) = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -x-2 \end{pmatrix}_{x=-3} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

从而令

$$M = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 \\ 1 & 1 & -3 & 1 \end{pmatrix},$$

就有

$$M^{-1}AM = \mathrm{diag}(1, 1, 1, -3).$$

例 2 (对照 [2]pp.310–312 例 2) 设

$$A = \begin{pmatrix} 0 & -1 & 2 \\ 3 & 8 & -14 \\ 3 & 6 & -10 \end{pmatrix}$$

求 M, 使得 $M^{-1}AM$ 为 Jordan 标准形.

$$\begin{pmatrix} x & 1 & -2 \\ -3 & x-8 & 14 \\ -3 & -6 & x+10 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & x & -2 \\ x-8 & -3 & 14 \\ -6 & -3 & x+10 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ x-8 & -x^2+8x-3 & 2x-2 \\ -6 & 6x-3 & x-2 \\ 0 & 1 & 0 \\ 1 & -x & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & -x^2+8x-3 & 2x-2 \\ 0 & 6x-3 & x-2 \\ 0 & 1 & 0 \\ 1 & -x & 2 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & -x^2+2x & x \\ 0 & 6x-3 & x-2 \\ 0 & 1 & 0 \\ 1 & -x & 2 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & x & -x^2+2x \\ 0 & x-2 & 6x-3 \\ 0 & 0 & 1 \\ 1 & 2 & -x \\ 0 & 1 & 0 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -x^2-4x+3 \\ 0 & x-2 & 6x-3 \\ 0 & 0 & 1 \\ 1 & 2 & -x \\ 0 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -x^2-4x+3 \\ 0 & 0 & x(x+1)^2 \\ 0 & 0 & 1 \\ 1 & 2 & -x \\ 0 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \boxed{2x(x+1)^2} \\ 0 & 0 & 2 \\ 1 & 2 & 2x^2+6x-6 \\ 0 & 1 & x^2+4x-3 \end{pmatrix}$$

于是 A 仅有唯一的不变因子 $d(x)=x(x+1)^2$(在此情形我们略去代表不变因子的下标 i), 显然 A 的特征值为 $0,-1$.

对应于 $\lambda=0$ 有

$$\alpha_0(0)=\begin{pmatrix} 2 \\ 2x^2+6x-6 \\ x^2+4x-3 \end{pmatrix}_{x=0}=\begin{pmatrix} 2 \\ -6 \\ -3 \end{pmatrix}.$$

对应于 $\lambda=-1$ 有

$$\alpha_0(-1)=\begin{pmatrix} 2 \\ 2x^2+6x-6 \\ x^2+4x-3 \end{pmatrix}_{x=-1}=\begin{pmatrix} 2 \\ -10 \\ -6 \end{pmatrix},\quad \alpha_1(-1)=\left(\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix} 2 \\ 2x^2+6x-6 \\ x^2+4x-3 \end{pmatrix}\right)_{x=-1}=\begin{pmatrix} 0 \\ 4x+6 \\ 2x+4 \end{pmatrix}_{x=-1}=\begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}.$$

从而, 令

$$M = \begin{pmatrix} \alpha_0(0) & \alpha_0(-1) & \alpha_1(-1) \end{pmatrix} = \begin{pmatrix} 2 & 2 & 0 \\ -6 & -10 & 2 \\ -3 & -6 & 2 \end{pmatrix}$$

就有

$$M^{-1}AM = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}.$$

10.3 有理标准型的新算法

我们给出求有理标准型的一个平行算法如下 (平行的证明此处从略, 一个间接的证明见下一节).

定理 19 设 $P, Q \in \mathrm{GL}_n(F[x])$ 满足 (10.1), 记 Q 如 (10.2). 对 $1 \leqslant i \leqslant r$, 令 $\deg d_i = n_i$, 递归定义以下多项式向量

$$\xi_{i0} = \xi_i, \qquad \xi_{ij} = x\xi_{i,j-1} - d_i\langle \xi_{i,j-1}, x^{n_i-1}\rangle, \quad 1 \leqslant j \leqslant n_i - 1. \tag{10.21}$$

并令

$$\beta_{ij} = \langle \xi_{ij}, x^{n_i-1}\rangle, \qquad 0 \leqslant j \leqslant n_i - 1. \tag{10.22}$$

则下述断言成立:

(i) 向量

$$\{\beta_{ij}, \quad 0 \leqslant j \leqslant n_i - 1\} \tag{10.23}$$

线性无关, 并生成 F^n 的一个 L_A-不变子空间 W_i. 限制线性变换 $L_A|_{W_i}$ 在 W_i 的基底 (10.23) 下的矩阵是 d_i 的友阵.

(ii) F^n 是子空间 W_i 的直和. 从而向量

$$\{\beta_{ij} : 1 \leqslant i \leqslant r, 0 \leqslant j \leqslant n_i - 1\} \tag{10.24}$$

构成 F^n 的一组基, L_A 在这组基下为有理标准型.

评论 3 如同 Jordan 标准型的情况, 将 (10.24) 中向量拼接构成的矩阵 N 是使得 $N^{-1}AN$ 为有理标准型的过渡矩阵.

例 3 设

$$A=\begin{pmatrix}0 & -1 & 2\\ 3 & 8 & -14\\ 3 & 6 & -10\end{pmatrix}$$

求 N, 使得 $N^{-1}AN$ 为有理标准形. 如前, 我们得到 $d=x(x+1)^2=x^3+2x^2+x$,

$$\xi=\begin{pmatrix}1\\ x^2+3x-3\\ \frac{1}{2}x^2+2x-\frac{3}{2}\end{pmatrix}.$$

于是我们有

$$\xi_0=\xi=\begin{pmatrix}1\\ x^2+3x-3\\ \frac{1}{2}x^2+2x-\frac{3}{2}\end{pmatrix},\quad \beta_0=\begin{pmatrix}0\\ 1\\ \frac{1}{2}\end{pmatrix},$$

$$\xi_1=x\xi_0-d\beta_0=x\begin{pmatrix}1\\ x^2+3x-3\\ \frac{1}{2}x^2+2x-\frac{3}{2}\end{pmatrix}-(x^3+2x^2+x)\begin{pmatrix}0\\ 1\\ \frac{1}{2}\end{pmatrix}=\begin{pmatrix}x\\ x^2-4x\\ x^2-2x\end{pmatrix},\quad \beta_1=\begin{pmatrix}0\\ 1\\ 1\end{pmatrix},$$

$$\xi_2=x\xi_1-d\beta_1=x\begin{pmatrix}x\\ x^2-4x\\ x^2-2x\end{pmatrix}-(x^3+2x^2+x)\begin{pmatrix}0\\ 1\\ 1\end{pmatrix}=\begin{pmatrix}x^2\\ -6x^2-x\\ -4x^2-2x\end{pmatrix},\quad \beta_2=\begin{pmatrix}1\\ -6\\ -4\end{pmatrix}.$$

于是, 令

$$N=\begin{pmatrix}\beta_0 & \beta_1 & \beta_2\end{pmatrix}=\begin{pmatrix}0 & 0 & 1\\ 1 & 1 & -6\\ \frac{1}{2} & 1 & -4\end{pmatrix},$$

就有

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

10.4 与经典算法的比较

我们先回顾一下 [1, 3, 4] 中给出的经典算法.

不同于本文所给出的算法, 经典算法是利用出现在 (10.1)中的矩阵 P 的逆. 令

$$P^{-1} = (*, \ldots, *, \eta_1, \ldots, \eta_r), \qquad \eta_i \in F[x]^n.$$

我们将 F^n 视为 $F[x]$-模, $f \in F[x]$ 与 $\alpha \in F^n$ 的乘法由 $f\alpha = f(A)\alpha$ 给出. 考虑自由 $F[x]$-模 $F[x]^n$ 与唯一的 $F[x]$-模同态 $\phi : F[x]^n \to F^n$, 它在 F^n 上的限制是恒同. 则

$$\mathrm{Ker}(\phi) = \{(xI_n - A)\zeta : \zeta \in F[x]^n\}.$$

经典算法如下:

- 我们有循环分解 $F^n = \bigoplus_{i=1}^r F[x]\phi(\eta_i)$, 且 $\phi(\eta_i)$ 的零化子是 d_i. 于是, 对每个 $1 \leqslant i \leqslant r$, 向量

  $$\{x^j\phi(\eta_i) : 0 \leqslant j \leqslant n_i - 1\} \tag{10.25}$$

  构成 $F[x]\phi(\eta_i)$ 的一组基, 并且 L_A 的限制在这组基下的矩阵为 d_i 的友阵.

- 对每个 $1 \leqslant i \leqslant r$, 给定分解 (10.3), 有本原分解 $F[x]\phi(\eta_i) = \bigoplus_{\lambda:d_i(\lambda)=0} F[x]\beta_i(\lambda)$, 其中 $\beta_i(\lambda) = \frac{d_i}{(x-\lambda)^{m_i(\lambda)}}\phi(\eta_i)$ 有零化子 $(x-\lambda)^{m_i(\lambda)}$. 于是向量

  $$\left\{\frac{d_i}{(x-\lambda)^{j+1}}\phi(\eta_i) : 0 \leqslant j \leqslant m_i(\lambda) - 1\right\} \tag{10.26}$$

  构成 $F[x]\beta_i(\lambda)$ 的一组基, 并且 L_A 限制在这组基下的矩阵是一个 $m_i(\lambda)$ 阶 Jordan 块, 对应于特征值 λ.

此算法中, 耗时较多的步骤在于向量 $\phi(\eta_i)$ 的计算, 而它是基于 P^{-1} 的计算. 在 [1, 3] 中 P^{-1} 的计算是通过对 I_n 同步记录施加在 $(xI_n - A)$ 上的行变换的逆而得到. 为从 η_i 计算 $\phi(\eta_i)$, 还需要计算多项式分量 η_i 在矩阵 A 的值.

可以证明, 此处 (10.9) 和 (10.24) 所给出的基底与经典算法给出的基底一致. 我们简要说明如下.

我们首先说明, 对 $1 \leqslant i \leqslant r$ 与 d_i 的根 λ, (10.7)中的向量与 (10.26)中的向量一致, 即

$$\alpha_{ij}(\lambda) = \frac{d_i}{(x-\lambda)^{j+1}}\phi(\eta_i), \qquad 0 \leqslant j \leqslant m_i(\lambda) - 1. \tag{10.27}$$

对固定的指标 i, j , 令

$$\xi_i(\lambda) = (x-\lambda)^{j+1}\zeta_{ij}(\lambda) + \rho_{ij}(\lambda),$$

其中 $\zeta_{ij}(\lambda), \rho_{ij}(\lambda) \in F[x]^n$, 且 $\deg \rho_{ij}(\lambda) \leqslant j$. 则由 (10.6) 可得 $\langle \rho_{ij}(\lambda), (x-\lambda)^j \rangle = \alpha_{ij}(\lambda)$. 由(10.1), 我们有

$$d_i\eta_i = (xI_n - A)\xi_i = (x-\lambda)^{j+1}(xI_n - A)\zeta_{ij}(\lambda) + (xI_n - A)\rho_{ij}(\lambda). \tag{10.28}$$

由于 $(x-\lambda)^{j+1}$ 整除 d_i, 由 (10.28) 得到, $(x-\lambda)^{j+1}$ 整除 $(xI_n - A)\rho_{ij}(\lambda)$ 的各个分量, 这些分量的次数至多为 $j+1$. 另一方面,

$$\langle (xI_n - A)\rho_{ij}(\lambda), (x-\lambda)^{j+1} \rangle = \langle \rho_{ij}(\lambda), (x-\lambda)^{j+1} \rangle = \alpha_{ij}(\lambda).$$

从而必定有 $(xI_n - A)\rho_{ij}(\lambda) = (x-\lambda)^{j+1}\alpha_{ij}(\lambda)$. 将它代入 (10.28)即得

$$\frac{d_i}{(x-\lambda)^{j+1}}\eta_i = (xI_n - A)\zeta_{ij}(\lambda) + \alpha_{ij}(\lambda).$$

将同态 ϕ 作用在上式两边, 并利用 $(xI_n - A)\zeta_{ij}(\lambda) \in \mathrm{Ker}(\phi)$, 即得 (10.27).

类似地, 我们证明, 对每个 $1 \leqslant i \leqslant r$, (10.23) 中的向量与 (10.25) 中的向量一致, 即

$$\langle \xi_{ij}, x^{n_i-1} \rangle = x^j\phi(\eta_i), \qquad 0 \leqslant j \leqslant n_i - 1. \tag{10.29}$$

由(10.21)我们有 $d_i \mid \xi_{ij} - x^j\xi_i$, 从而我们可以写 $x^j\xi_i = d_i\zeta_{ij} + \xi_{ij}$, 其中 $\zeta_{ij} \in F[x]^n$. 由 (10.1) 有

$$x^j d_i\eta_i = x^j(xI_n - A)\xi_i = d_i(xI_n - A)\zeta_{ij} + (xI_n - A)\xi_{ij}. \tag{10.30}$$

这就推出 d_i 整除 $(xI_n - A)\xi_{ij}$ 的每个分量, 而这些分量的次数不超过 n_i. 另一方面,

$$\langle (xI_n - A)\xi_{ij}, x^{n_i} \rangle = \langle \xi_{ij}, x^{n_i-1} \rangle$$

于是 $(xI_n - A)\xi_{ij} = d_i\langle \xi_{ij}, x^{n_i-1} \rangle$. 将它代入 (10.30), 我们得到

$$x^j\eta_i = (xI_n - A)\zeta_{ij} + \langle \xi_{ij}, x^{n_i-1} \rangle.$$

将同态 ϕ 作用在上式两边, 并注意到 $(xI_n - A)\zeta_{ij} \in \mathrm{Ker}(\phi)$, 即得 (10.29).

10.5 结语

最后, 我们分享一段 Knuth[5] 的评论, 希望本文介绍的算法应用于课堂教学:

> For three years I taught a sophomore course in abstract algebra, for mathematics majors at Caltech, and the most difficult topic was always the study of "Jordan canonical form" for matrices. The third year I tried a new approach, by looking at the subject algorithmically, and suddenly it became quite clear. The same thing happened with the discussion of finite groups defined by generators and relations; and in another course, with the reduction theory of binary quadratic forms. By presenting the subject in terms of algorithms, the purpose and meaning of the mathematical theorems became transparent.

致谢

本文第二作者感谢上海应用技术大学陈浦胤老师和宁德师范学院蒋剑剑老师与他讨论交流.

参考文献

[1] Adkins, W., Weintraub, S. (1992). *Algebra: An Approach via Module Theory.* New York: Springer-Verlag.

[2] 北京大学数学系前代数小组, 王萼芳、石生明修订,《高等代数》, 第五版, 北京：高等教育出版社. 2019.

[3] Dummit, D., Foote, R. (2004). *Abstract Algebra,* 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.

[4] Jacobson, N. (1985). *Basic Algebra I,* 2nd ed. New York: W. H. Freeman and Company.

[5] Knuth, D. E. (1974). Computer science and its relation to mathematics, Amer. Math. Monthly 81: 4, 323–343, DOI: 10.1080/00029890.1974.11993556

[6] 乐小英, 利用矩阵的初等变换对矩阵特征值与特征向量同步求解的再讨论, 《江西电力职工大学学报》, 第 10 卷第 1 期, 1997 年 3 月, 16–18.