矩阵论十讲

Ten Lectures on Matrix Theory

李　乔　张晓东 / 著

中国科学技术大学出版社

矩阵论十讲

Ten Lectures on Matrix Theory

李　乔　张晓东 / 著

中国科学技术大学出版社

内 容 简 介

矩阵的应用日趋深广，大学基础课的内容已很不适应.本书旨在在大学“线性代数”课程的基础上，分10个专题充实和扩大关于矩阵理论的知识.具体内容依次为：方阵函数；矩阵的直积和矩阵方程；复合矩阵和行列式恒等式；酉方阵、Hermite方阵和规范方阵；Hermite方阵的特征值和一般方阵的奇异值；非负元方阵和布尔方阵；矩阵的组合性质；矩阵的广义逆；完全正方阵；图的Laplace方阵.各讲基本独立成章，具备基本的线性代数和分析知识的读者即可读懂.

本书各讲内容可作为多种类型的大学选修课教程，也可以作为关心矩阵理论的教师、学生和科技工作者的自学读物或供查阅的参考书.

图书在版编目(CIP)数据

矩阵论十讲/李乔，张晓东著.—合肥：中国科学技术大学出版社，2015.3
ISBN 978-7-312-03528-9

Ⅰ.矩… Ⅱ.①李… ②张… Ⅲ.矩阵论—研究生—教材 Ⅳ.O151.21

中国版本图书馆CIP数据核字(2015)第052534号

中国科学技术大学出版社出版发行
安徽省合肥市金寨路96号，230026
http://press.ustc.edu.cn
合肥华星印务有限责任公司
全国新华书店经销

开本：710 mm×960 mm　1/16　印张：11　字数：216千
2015年3月第1版　2015年3月第1次印刷
定价：25.00元

前　　言

本书的前身是李乔在1988年出版的《矩阵论八讲》.《矩阵论八讲》的前言阐明该书的宗旨是:"在基本线性代数课程的基础上,充实和扩大关于矩阵理论的有用知识.……还试图使之成为便于了解矩阵理论的某些内容的简明参考书.对读者只要求具备最基本的线性代数知识."

二十多年过去了,大学"线性代数"课程的内容大致依旧,而矩阵的应用则日趋深广.这种状况使得《矩阵论八讲》有更新、充实后再版的需要.现在新版得以真正面世,我们首先要感谢中国科学技术大学出版社的热诚邀约和大力支持.

本书共十讲,前八讲是对1988年版的全面修订和少量增补,由李乔完成;后两讲则是由张晓东新写的.我们在讨论确定了大致内容的同时,也商定这样的合作方式:体例和基本术语全书统一,而前八讲和后两讲则分别按各自的想法和行文风格独立写作,最后由张晓东校阅全书.我们都希望重申《矩阵论八讲》的数学思想,即强调组合学的视角和处理方法,这使得本书与同类著作相比,也许可以说是别具只眼吧.

李　乔　　张晓东
上海交通大学数学系
2014年8月

目　　次

第1讲　方阵函数

1.1　Jordan 标准形温习

本讲都在复数域 $\mathbf{C}$ 上讨论. 矩阵通常用大写、黑体拉丁字母表示. $m\times n$ 矩阵 $\boldsymbol{A}=(a_{ij})$ 表示 $\boldsymbol{A}$ 有 m 行、n 列,且 $\boldsymbol{A}$ 的位于第 i 行、第 j 列的元(简称 $\boldsymbol{A}$ 的 (i,j) 元) $a_{ij}(i=1,\cdots,m;j=1,\cdots,n)$ 都是复数, $n\times n$ 矩阵也称为 n 阶方阵, $n\times 1$ 和 $1\times n$ 矩阵也分别称为 n 维列向量和行向量①. 数 α 称为 n 阶方阵 $\boldsymbol{A}$ 的特征值(eigenvalue)(或特征根),如果有非零的 n 维列向量 $\boldsymbol{x}$,使 $\boldsymbol{Ax}=\alpha\boldsymbol{x}$. 这时,满足这个等式的非零向量 $\boldsymbol{x}$ 称为 $\boldsymbol{A}$ 的属于特征值 α 的特征向量.

特征值可以不依赖于特征向量而直接由方阵来确定. 事实上, n 阶方阵 $\boldsymbol{A}$ 共有 n 个特征值,它们是 λ 的 n 次多项式

$$\chi_{\boldsymbol{A}}(\lambda)=\det(\lambda\boldsymbol{I}_n-\boldsymbol{A})$$

的 n 个根,这里 $\boldsymbol{I}_n$ 是 n 阶单位方阵, $\det(\lambda\boldsymbol{I}_n-\boldsymbol{A})$ 表示方阵 $\lambda\boldsymbol{I}_n-\boldsymbol{A}$ 的行列式. 多项式 $\chi_{\boldsymbol{A}}(\lambda)$ 称为 $\boldsymbol{A}$ 的特征多项式. 如果记 $\chi_{\boldsymbol{A}}(\lambda)=0$ 的 n 个(不一定互不相同)根为 $\lambda_1,\cdots,\lambda_n$,则称重集(multiset) $\{\lambda_1,\cdots,\lambda_n\}$ 为 $\boldsymbol{A}$ 的谱(spectrum),记为 $\mathrm{Spec}\boldsymbol{A}$. $\chi_{\boldsymbol{A}}(\lambda)$ 的一个根(从而是 $\boldsymbol{A}$ 的一个特征值)的重数称为这个特征值的代数重数. 与同一特征值 α 相应的全体特征向量再加上零向量后构成一个线性子空间,称为与特征值 α 相应的特征子空间. 易知,这个特征子空间的维数等于 $n-\mathrm{rank}(\alpha\boldsymbol{I}_n-\boldsymbol{A})$,称为特征值 α 的几何重数.

例 1.1　设 $\boldsymbol{A}=\begin{pmatrix}0&1\\0&0\end{pmatrix}$,则

$$\chi_{\boldsymbol{A}}(\lambda)=\lambda^2,\quad \mathrm{Spec}\boldsymbol{A}=\{0,0\},$$

0 的代数重数是 2,几何重数是 1.

① 它们在写法上也不加区别, n 维行向量和 $1\times n$ 矩阵都记为 $(a_1,\cdots,a_n)$.

多项式 $f(\lambda)=a_0\lambda^m+a_1\lambda^{m-1}+\cdots+a_m$ 称为 n 阶方阵 $\boldsymbol{A}$ 的化零多项式，如果 $f(\boldsymbol{A})=a_0\boldsymbol{A}^m+a_1\boldsymbol{A}^{m-1}+\cdots+a_m\boldsymbol{I}_n=\boldsymbol{0}$. $\boldsymbol{A}$ 的次数最小且最高次项系数是 1 的化零多项式称为 $\boldsymbol{A}$ 的最小多项式，记作 $m_{\boldsymbol{A}}(\lambda)$. 对任一方阵 $\boldsymbol{A}$，下述 Cayley-Hamilton 定理保证了 $\boldsymbol{A}$ 一定有化零多项式.

定理 1.1 n 阶方阵 $\boldsymbol{A}$ 的特征多项式是 $\boldsymbol{A}$ 的化零多项式. □

易证 $\boldsymbol{A}$ 的最小多项式 $m_{\boldsymbol{A}}(\lambda)$ 由 $\boldsymbol{A}$ 唯一确定，而且多项式 $f(\lambda)$ 是 $\boldsymbol{A}$ 的化零多项式当且仅当 $m_{\boldsymbol{A}}(\lambda)\mid f(\lambda)$. 两个 n 阶方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 称为相似的，如果有可逆方阵 $\boldsymbol{S}$，使 $\boldsymbol{A}=\boldsymbol{SBS}^{-1}$. 显然，相似关系是一种等价关系. 前面提到的特征多项式、谱、最小多项式、特征值的代数重数和几何重数等等，都是方阵在相似关系下的不变量. 更加深刻的问题是：什么是对于相似关系来说的不变量全系？即由怎样一系列不变量可以确定两个同阶方阵是否相似？同样的问题对于 $m\times n$ 矩阵的相抵关系来说，答案非常简单：矩阵的秩(rank)这一个相抵不变量即已组成相抵关系的不变量全系. 因为两个 $m\times n$ 矩阵相抵当且仅当它们的秩相等. 相似关系当然远不是如此简单，但也早已彻底搞清楚了. 正是 Jordan 标准形简明地展示了方阵相似关系的不变量全系. 在具体讨论之前，再定义两个名词.

设 $\boldsymbol{A}_i(i=1,\cdots,k)$ 是 n_i 阶方阵，称 $n_1+\cdots+n_k$ 阶对角块形方阵

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_1 & & \\ & \ddots & \\ & & \boldsymbol{A}_k\end{pmatrix}=\mathrm{diag}(\boldsymbol{A}_1,\cdots,\boldsymbol{A}_k)$$

为 $\boldsymbol{A}_1,\cdots,\boldsymbol{A}_k$ 的直和，记为

$$\boldsymbol{A}=\boldsymbol{A}_1\oplus\cdots\oplus\boldsymbol{A}_k=\bigoplus_{i=1}^{k}\boldsymbol{A}_i.$$

给定数 α，称 l 阶方阵

$$\boldsymbol{J}_l(\alpha)=\begin{pmatrix}\alpha & 1 & & \boldsymbol{0} \\ & \alpha & 1 & \\ & & \alpha & \ddots & \\ & & & \ddots & 1 \\ \boldsymbol{0} & & & & \alpha\end{pmatrix}$$

为(属于 α 的 l 阶)Jordan 块.

不难得到 $\boldsymbol{J}_l(\alpha)$ 的一些性质：它的特征多项式是 $(\lambda-\alpha)^l$；$\mathrm{Spec}\boldsymbol{J}_l(\alpha)=\{\alpha,\cdots,\alpha\}$；$\alpha$ 的代数重数是 l，几何重数是 1；$\boldsymbol{J}_l(\alpha)$ 的最小多项式也是 $(\lambda-\alpha)^l$.

关键的 Jordan 标准形定理可陈述如下：

定理 1.2 设 n 阶方阵 $\boldsymbol{A}$ 共有 k 个不同的特征值 $\lambda_1,\cdots,\lambda_k$，则：

(1) $\boldsymbol{A}$ 相似于对角块形方阵

$$\bigoplus_{i=1}^{k}(J_{n_{i1}}(\lambda_i)\oplus J_{n_{i2}}(\lambda_i)\oplus\cdots\oplus J_{n_{it_i}}(\lambda_i)),$$

这里 $n_{ij}(i=1,\cdots,k;j=1,\cdots,t_i)$是正整数；

(2) 上述共 $t_1+\cdots+t_k$ 个 Jordan 块 $J_{n_{ij}}(\lambda_i)$是由 A 唯一确定的，它们的直和称为 A 的 Jordan 标准形(Jordan canonical form)；

(3) 两个方阵相似的充要条件是它们有相同的 Jordan 标准形. □

定理 1.2 可以说，是矩阵基础知识领域的一个制高点. 有了 Jordan 标准形后，很多复杂的关系就能理顺了.

推论 1.1　在定理1.2的记号下，特征值 $\lambda_i(i=1,\cdots,k)$的代数重数是 $n_{i1}+\cdots+n_{it_i}$，几何重数是 t_i，从而一定有：几何重数≤代数重数，而且下列诸性质互相等价：

(1) $\lambda_i(i=1,\cdots,k)$的几何重数和代数重数相等；

(2) A 相似于对角方阵；

(3) A 有 n 个线性无关的特征向量；

(4) k 个特征值$\lambda_1,\cdots,\lambda_k$ 的几何重数之和等于 n. □

推论 1.2　设 $n_{i1}=\max\limits_{1\leqslant j\leqslant t_i}\{n_{ij}\}(i=1,\cdots,k)$，则 A 的最小多项式 $m_A(\lambda)=\prod\limits_{i=1}^{k}(\lambda-\lambda_i)^{n_{i1}}$. 从而 $m_A(\lambda)$的根除重数外与 $\chi_A(\lambda)$的根相同. $m_A(\lambda)=\chi_A(\lambda)$当且仅当 $t_i=1(i=1,\cdots,k)$，即相应于每个特征值 λ_i 只有一个 Jordan 块 $J_{n_{ij}}(\lambda_i)$. □

每个 Jordan 块 $J_{n_{ij}}(\lambda_i)$的特征多项式$(\lambda-\lambda_i)^{n_{ij}}(i=1,\cdots,k;j=1,\cdots,t_i)$称为 A 的一个初等因子. 这里，A 的初等因子只是 A 的 Jordan 标准形中的 Jordan 块的另一种记法. 我们也可以说，方阵的全部初等因子是方阵在相似下的不变量全系. 今后将不固定用一种说法，而是根据表述的简明程度来采用两种说法之一.

下面举两个用 Jordan 标准形解决矩阵问题的例子. 由此可以看到，有了 Jordan 标准形以后，犹如高屋建瓴，使看来难以下手的问题变得容易处理了.

例 1.2　任一方阵 A 和它的转置阵 A^{T} 相似.

设 $SAS^{-1}=J_1\oplus\cdots\oplus J_m$ 是 A 的 Jordan 标准形，则

$$J_1^{\mathrm{T}}\oplus\cdots\oplus J_m^{\mathrm{T}}=(SAS^{-1})^{\mathrm{T}}=(S^{-1})^{\mathrm{T}}A^{\mathrm{T}}S^{\mathrm{T}}=(S^{\mathrm{T}})^{-1}A^{\mathrm{T}}S^{\mathrm{T}}.$$

故只需证明一个 Jordan 块 $J_1=J_l(\lambda_1)$和它的转置阵 J_1^{T} 相似即可.

令 P 是次对角线上元素都是 1、其余元素都是 0 的 l 阶方阵：

$$P = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix},$$

则不难验证 $P^{-1} = P, PJ_1P = J_1^{\mathrm{T}}$. □

例 1.3 记方阵 A 的特征值的最大模为 $\rho(A)$(也称 $\rho(A)$ 为 A 的谱半径),则 $\lim\limits_{k\to\infty} A^k$ 存在(指当 $k\to\infty$ 时,A^k 在每个位置上的元素都有有限极限)的充要条件是 $\rho(A)<1$ 或 $\rho(A)=1$,但 1 是 A 的唯一的模为 1 的特征值,且对应于 1 的初等因子的次数都是 1.

由于 $\lim\limits_{k\to\infty} A^k$ 存在与否以及存在时的极限矩阵都是相似下不变的性质,所以只要讨论当 A 是 Jordan 标准形,或更进一步,当 A 是一个 Jordan 块的情形. 而当 A 是 Jordan 块时,不难直接验证结论成立. □

我们不准备再多举这类例子了. 从根本上可以说,矩阵理论的相当一部分内容都能看作是 Jordan 标准形的应用. 在第 2 讲讨论矩阵方程时,Jordan 标准形的作用是关键的. 本讲对方阵函数的处理,虽然不是从 Jordan 标准形出发的,但从它出发定义和讨论方阵函数也是完全可行的(见 1.4 节).

1.2 方阵函数的定义

本节所面临的问题是:对于复变量 λ 的复数值函数 $f(\lambda)$,怎样自然而又合理地定义 $f(A)$(这里 A 是 n 阶复方阵)?

当 $f(\lambda)$ 是多项式函数时,我们显然有这种定义:设 $f(\lambda) = a_0\lambda^l + a_1\lambda^{l-1} + \cdots + a_l$,自然定义 $f(A) = a_0A^l + a_1A^{l-1} + \cdots + a_lI_n$. 这时,若有另一多项式 $g(\lambda)$,则易知

$$f(A) = g(A) \iff m_A(\lambda) \mid (f(\lambda) - g(\lambda)),$$

这里 $m_A(\lambda)$ 是 A 的最小多项式. 若记

$$m_A(\lambda) = (\lambda - \lambda_1)^{m_1}\cdots(\lambda - \lambda_k)^{m_k},$$
$$\deg m_A(\lambda) = m_1 + \cdots + m_k = m,$$

则

$$m_A(\lambda) \mid (f(\lambda) - g(\lambda)) \iff f^{(j)}(\lambda_i) = g^{(j)}(\lambda_i)$$
$$(i = 1,\cdots,k; j = 0,1,\cdots,m_i - 1),$$

这里 $f^{(j)}(\lambda_i)$ 是多项式函数 $f(\lambda)$ 的 j 次导数在 $\lambda = \lambda_i$ 时的值. 所以,对多项式 $f(\lambda)$ 来说,$f(A)$ 是由总共 $m = m_1 + \cdots + m_k$ 个值

$$f(\lambda_i), f^{(1)}(\lambda_i), \cdots, f^{(m_i-1)}(\lambda_i) \quad (i = 1, \cdots, k)$$

唯一确定的.如果下面 m 个等式

$$f^{(j)}(\lambda_i) = g^{(j)}(\lambda_i) \quad (i = 1, \cdots, k; j = 0, 1, \cdots, m_i - 1)$$

成立,这 m 个等可简记为$f(\Lambda_A) = g(\Lambda_A)$,则对多项式 $f(\lambda)$和 $g(\lambda)$,有

$$f(\boldsymbol{A}) = g(\boldsymbol{A}) \quad \Leftrightarrow \quad f(\Lambda_A) = g(\Lambda_A).$$

现在来对一般的 $f(\lambda)$定义 $f(\boldsymbol{A})$.我们总假设 $f(\lambda)$在各 λ_i 处有直到$m_i - 1$阶的导数——简称为 $f(\lambda)$在 Λ_A 上有意义.

定义 1.1　对函数 $f(\lambda)$,如果有多项式 $p(\lambda)$使 $p(\Lambda_A) = f(\Lambda_A)$,则定义 $f(\boldsymbol{A}) = p(\boldsymbol{A})$,并称 $p(\lambda)$是 $f(\boldsymbol{A})$的定义多项式.

显然,上述定义是明确的:$f(\boldsymbol{A})$与可能有多种选择的多项式 $p(\lambda)$无关.不很明显的是,这个定义是否真有意义,即所要求的多项式 $p(\lambda)$是否一定存在?回答是肯定的:使 $p(\Lambda_A) = f(\Lambda_A)$的多项式一定存在,而且次数小于 m 的这种多项式是唯一的.这就是著名的 Lagrange-Sylvester 插值多项式.当 $m_1 = \cdots = m_k = 1$ 时,它就是 Lagrange 插值多项式:

$$p(\lambda) = \sum_{i=1}^{k} \frac{(\lambda - \lambda_1)\cdots(\lambda - \lambda_{i-1})(\lambda - \lambda_{i+1})\cdots(\lambda - \lambda_k)}{(\lambda_i - \lambda_1)\cdots(\lambda_i - \lambda_{i-1})(\lambda_i - \lambda_{i+1})\cdots(\lambda_i - \lambda_k)} f(\lambda_i),$$

一般情形的 Lagrange-Sylvester 插值多项式尽管看似较繁,但在归并 $f(\lambda)$在各 λ_i 处的各阶导数后可以写成

$$p(\lambda) = \sum_{i=1}^{k} [f(\lambda_i)\varphi_{i1}(\lambda) + f^{(1)}(\lambda_i)\varphi_{i2}(\lambda) + \cdots + f^{(m_i-1)}(\lambda_i)\varphi_{im_i}(\lambda)],$$

其中 $\deg\varphi_{ij}(\lambda) < m$,$\varphi_{ij}(\lambda)$与 $f(\lambda)$无关,而由 $m_A(\lambda)$确定.所以

$$f(\boldsymbol{A}) = p(\boldsymbol{A}) = \sum_{i=1}^{k} [f(\lambda_i)\varphi_{i1}(\boldsymbol{A}) + \cdots + f^{(m_i-1)}(\lambda_i)\varphi_{im_i}(\boldsymbol{A})].$$

若记 $\varphi_{ij}(\boldsymbol{A}) = \boldsymbol{Z}_{ij}$,则 n 阶方阵 $\boldsymbol{Z}_{ij}$是由 $\boldsymbol{A}$ 确定的,与 $f(\lambda)$无关.这个表示式对具体求 $f(\boldsymbol{A})$来说,有时会带来方便,因为我们可以选取适当的函数作为 $f(\lambda)$,并通过这些函数来确定 $\boldsymbol{Z}_{ij}$,从而最终求得 $f(\boldsymbol{A})$.在下一节,我们会通过另一条途径再次得到这个表示式.

注　$f(\boldsymbol{A})$的定义多项式 $p(\lambda)$既取决于函数 f,又取决于方阵 $\boldsymbol{A}$,因此 $f(\boldsymbol{A})$和 $f(\boldsymbol{B})$一般有不同的定义多项式.

1.3　方阵函数的其他等价定义

首先注意到 1.2 节中所定义的方阵函数有如下两个基本性质:

(1) $\boldsymbol{A}=\boldsymbol{SBS}^{-1}\Rightarrow f(\boldsymbol{A})=\boldsymbol{S}f(\boldsymbol{B})\boldsymbol{S}^{-1}$;

(2) $f(\boldsymbol{A}_1\oplus\cdots+\oplus\boldsymbol{A}_l)=f(\boldsymbol{A}_1)\oplus\cdots\oplus f(\boldsymbol{A}_l)=p(\Lambda_{\boldsymbol{B}})$.

证 (1) 由 $\boldsymbol{A}=\boldsymbol{SBS}^{-1}$ 可知 $m_{\boldsymbol{A}}(\lambda)=m_{\boldsymbol{B}}(\lambda)$,从而 $\Lambda_{\boldsymbol{A}}=\Lambda_{\boldsymbol{B}}, f(\Lambda_{\boldsymbol{A}})=f(\Lambda_{\boldsymbol{B}})$. 于是 $f(\Lambda_{\boldsymbol{A}})=p(\Lambda_{\boldsymbol{A}})\Rightarrow f(\Lambda_{\boldsymbol{B}})=p(\Lambda_{\boldsymbol{B}}), f(\boldsymbol{A})=p(\boldsymbol{A})=\boldsymbol{S}p(\boldsymbol{B})\boldsymbol{S}^{-1}=\boldsymbol{S}f(\boldsymbol{B})\boldsymbol{S}^{-1}$.

(2) 由于 $m_{\boldsymbol{A}}(\lambda)$一定是每个 $\boldsymbol{A}_i$ 的化零多项式,所以 $m_{\boldsymbol{A}_i}(\lambda)\mid m_{\boldsymbol{A}}(\lambda)\Rightarrow\Lambda_{\boldsymbol{A}_i}\subseteq\Lambda_{\boldsymbol{A}}$. 从而对 $f(\boldsymbol{A})$的任一定义多项式 $p(\lambda)$来说,$p(\boldsymbol{A}_1\oplus\cdots\oplus\boldsymbol{A}_l)=p(\boldsymbol{A}_1)\oplus\cdots+\oplus p(\boldsymbol{A}_l)$显然成立,故(2)成立.

基于(1)和(2)以及 Jordan 标准形定理,只要能定义 $f(\boldsymbol{J})$,就能定义 $f(\boldsymbol{A})$,这里 $\boldsymbol{J}$ 是一个 Jordan 块. 记

$$\boldsymbol{J}=\boldsymbol{J}_l(\alpha)=\alpha\boldsymbol{I}+\boldsymbol{H},\quad \boldsymbol{H}=\boldsymbol{J}_l(0),$$

则 $m_{\boldsymbol{J}}(\lambda)=(\lambda-\alpha)^l$. 从而多项式

$$p(\lambda)=f(\alpha)+\frac{f'(\alpha)}{1!}(\lambda-\alpha)+\cdots+\frac{f^{(l-1)}(\alpha)}{(l-1)!}(\lambda-\alpha)^{l-1}$$

满足 $p(\Lambda_{\boldsymbol{J}})=f(\Lambda_{\boldsymbol{J}})$,即可取 $p(\lambda)$作为 $f(\boldsymbol{J})$的定义多项式. 因此

$$f(\boldsymbol{J})=p(\boldsymbol{J})=\sum_{i=1}^{l-1}\frac{f^{(i)}(\alpha)}{i!}(\boldsymbol{J}-\alpha\boldsymbol{I})=\sum_{i=1}^{l-1}\frac{f^{(i)}(\alpha)}{i!}\boldsymbol{H}^i.$$

(3) 设 $\boldsymbol{J}=\boldsymbol{J}_l(\alpha)$,则

$$f(\boldsymbol{J})=\begin{pmatrix} f(\alpha) & \dfrac{f'(\alpha)}{1!} & \cdots & \dfrac{f^{(l-1)}(\alpha)}{(l-1)!} \\ & f(\alpha) & \ddots & \vdots \\ & \boldsymbol{0} & \ddots & \dfrac{f'(\alpha)}{1!} \\ & & & f(\alpha) \end{pmatrix}.$$

结合性质(1)~(3),可以给出矩阵函数 $f(\boldsymbol{A})$的另一种定义:设 $\boldsymbol{SAS}^{-1}=\boldsymbol{J}_1\oplus\cdots\oplus\boldsymbol{J}_m$ 是 $\boldsymbol{A}$ 的 Jordan 标准形,则 $f(\boldsymbol{A})=\boldsymbol{S}^{-1}(f(\boldsymbol{J}_1)\oplus\cdots\oplus f(\boldsymbol{J}_m))\boldsymbol{S}$.

这个定义和1.2节中的定义当然应该是等价的. 事实上,它们各有特点. 在1.2节的定义中,我们明确地知道,对函数 $f(\lambda)$及矩阵 $\boldsymbol{A}$,有多项式 $p(\lambda)$,使 $f(\boldsymbol{A})=p(\boldsymbol{A})$. 这个事实很重要. 而现在所说的通过相似及 Jordan 标准形的定义,虽然对于计算 $f(\boldsymbol{A})$并不更方便(因为即使能求出 Jordan 标准形,但将原方阵化到标准形的变换方阵也不易求得),但它明确地表明,$f(\boldsymbol{A})$一定相似于一些特殊结构的上三角方阵 $f(\boldsymbol{J})$的直和,而且这些上三角方阵可以直接通过 $f(\Lambda_{\boldsymbol{A}})$表示出来. 这个事实很有用. 例如,由此可以得到1.2节最后的表示式:

$$f(\boldsymbol{A})=\sum_{i=1}^{k}[f(\lambda_i)\boldsymbol{Z}_{i1}+\cdots+f^{(m_i-1)}(\lambda_i)\boldsymbol{Z}_{im_i}].$$

在 1.5 节中，我们还将根据它求出 $f(\boldsymbol{A})$ 的全部初等因子，这里先提出它的一个简单而又常用的推论.

命题 1.1　设 $\mathrm{Spec}\boldsymbol{A}=\{\lambda_1,\cdots,\lambda_n\}$，则 $\mathrm{Spec}f(\boldsymbol{A})=\{f(\lambda_1),\cdots,f(\lambda_n)\}$. □

最后再讲一种方阵函数的等价定义——通过方阵幂级数定义. 矩阵函数最初就是在求解常系数线性微分方程组时以这种方式引入的.

定理 1.3　设复函数 $f(\lambda)$ 在开圆域 $|\lambda-\lambda_0|<r$ 内解析，即 $f(\lambda)$ 在此开圆域内可展开成幂级数

$$f(\lambda)=\sum_{k=0}^{\infty}a_k(\lambda-\lambda_0)^k,$$

则只要方阵 $\boldsymbol{A}$ 的所有特征值都在这个开圆域内，就有

$$f(\boldsymbol{A})=\sum_{k=0}^{\infty}a_k(\boldsymbol{A}-\lambda_0\boldsymbol{I})^k.$$

证　由性质(1)和(2)，我们只要证明当 $\boldsymbol{A}$ 是 Jordan 块时定理成立即可. 设 $\boldsymbol{J}=\boldsymbol{J}_l(\alpha)=\alpha\boldsymbol{I}+\boldsymbol{H}$，$\alpha$ 是 $\boldsymbol{J}$ 的唯一特征值. 因为 $f(\lambda)$ 在 $\lambda=\alpha$ 的一个邻域内解析，故 $f(\lambda)$ 也可在此邻域内展开成 $\lambda-\alpha$ 的幂级数

$$f(\lambda)=\sum_{k=0}^{\infty}b_k(\lambda-\alpha)^k,\quad 其中\ b_k=\frac{f^{(k)}(\alpha)}{k!}.$$

由于在 α 的一个邻域内，

$$\sum_{k=0}^{\infty}a_k(\lambda-\lambda_0)^k=\sum_{k=0}^{\infty}b_k(\lambda-\alpha)^k,$$

故矩阵等式

$$\sum_{k=0}^{\infty}a_k(\boldsymbol{J}-\lambda_0\boldsymbol{I})^k=\sum_{k=0}^{\infty}b_k(\boldsymbol{J}-\alpha\boldsymbol{I})^k$$

也成立. 因为上式两边都是有限和的极限，而绝对收敛级数又可任意重排各项的次序，所以根据性质(3)以及 $\boldsymbol{H}^l=\boldsymbol{0}$，即得

$$\begin{aligned}f(\boldsymbol{J})&=\sum_{k=0}^{l-1}\frac{f^{(k)}(\alpha)}{k!}\boldsymbol{H}^k=\sum_{k=0}^{l-1}b_k\boldsymbol{H}^k=\sum_{k=0}^{\infty}b_k\boldsymbol{H}^k,\\&=\sum_{k=0}^{\infty}b_k(\boldsymbol{J}-\alpha\boldsymbol{I})^k=\sum_{k=0}^{\infty}a_k(\boldsymbol{J}-\lambda_0\boldsymbol{I})^k.\end{aligned}$$ □

对解析函数来说，定理 1.3 中的矩阵幂级数表示式可以作为 $f(\boldsymbol{A})$ 的定义，最常用的一些矩阵函数均属此列. 例如，可以有下面的表示式：

$$\mathrm{e}^{\lambda}=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!},\quad|\lambda|<+\infty,\quad \mathrm{e}^{\boldsymbol{A}}=\sum_{k=0}^{\infty}\frac{\boldsymbol{A}^k}{k!}\quad(\boldsymbol{A}\ 任意);$$

$$\ln(1+\lambda)=\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\lambda^k,\quad|\lambda|<1,$$

$$\ln(\boldsymbol{I}+\boldsymbol{A})=\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\boldsymbol{A}^{k},\quad p(\boldsymbol{A})<1;$$

$$(1+\lambda)^{a}=\sum_{k=0}^{\infty}\frac{a(a-1)\cdots(a-k+1)}{k!}\lambda^{k},\quad |\lambda|<1\quad(a\text{ 是实数}),$$

$$(\boldsymbol{I}+\boldsymbol{A})^{a}=\sum_{k=0}^{\infty}\frac{a(a-1)\cdots(a-k+1)}{k!}\boldsymbol{A}^{k},\quad p(\boldsymbol{A})<1;$$

$$\ln\lambda=\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}(\lambda-1)^{k},\quad |\lambda-1|<1,$$

$$\ln\boldsymbol{A}=\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}(\boldsymbol{A}-\boldsymbol{I})^{k}.$$

其中 $\boldsymbol{A}$ 的每个特征值 λ_j 都满足 $|\lambda_j-1|<1$.

例 1.4 由于

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}=\frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=0}^{\infty}\frac{\boldsymbol{A}^{k}t^{k}}{k!}=\sum_{k=1}^{\infty}\frac{\boldsymbol{A}^{k}t^{k-1}}{(k-1)!}=\boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t},$$

所以向量函数 $\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{a}$(这里 $\boldsymbol{a}$ 是常数列向量,$\boldsymbol{x}(t)$是变量 t 的函数列向量)是常系数线性齐次微分方程组

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t}=a_{11}x_1+\cdots+a_{1n}x_n,\\ \cdots,\\ \dfrac{\mathrm{d}x_n}{\mathrm{d}t}=a_{n1}x_1+\cdots+a_{nn}x_n\end{cases}$$

$\left(\text{或记为}\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x},\text{这里 }\boldsymbol{A}=\begin{pmatrix}a_{11}&\cdots&a_{1n}\\ \vdots&&\vdots\\ a_{n1}&\cdots&a_{nn}\end{pmatrix},\boldsymbol{x}(t)=\begin{pmatrix}x_1(t)\\ \vdots\\ x_n(t)\end{pmatrix}\right)$,在初始条件 $\boldsymbol{x}(0)=\boldsymbol{a}$ 下的唯一解.这和 $n=1$ 时的常微分方程初值问题

$$\frac{\mathrm{d}x}{\mathrm{d}t}=ax,\quad x(0)=a_0$$

的唯一解 $x=a_0\mathrm{e}^{at}$ 完全一样.类似地,可以解非齐次方程组的初值问题

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{f}(t),\quad \boldsymbol{x}(0)=\boldsymbol{\alpha}.$$

这时,可令解

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{\alpha}(t),$$

代入原方程组后再确定向量函数 $\boldsymbol{\alpha}(t)$.这也和 $n=1$ 时非齐次方程的常数变易法一样.

1.4　方阵函数的性质

命题 1.2　设 $G(u_1,\cdots,u_l)$ 是 l 元多项式，函数 $f_1(\lambda),\cdots,f_l(\lambda)$ 都在 Λ_A 上有意义．若记 $g(\lambda)=G(f_1(\lambda),\cdots,f_l(\lambda))$，则有 $g(\boldsymbol{A})=G(f_1(\boldsymbol{A}),\cdots,f_l(\boldsymbol{A}))$．特别地，若 $g(\Lambda_A)=0$，则 $G(f_1(\boldsymbol{A}),\cdots,f_l(\boldsymbol{A}))=\boldsymbol{0}$．

（这个结论虽在意料之中，但不是不证自明的．因为按照定义，$g(\boldsymbol{A})$ 应等于某一多项式方阵函数 $p(\boldsymbol{A})$，而多项式 $p(\lambda)$ 却要根据 $g(\lambda)=G(f_1(\lambda),\cdots,f_l(\lambda))$ 来确定．）

证　取多项式 $r_i(\lambda)(i=1,\cdots,l)$，$r_i(\lambda)$ 和 $f_l(\lambda)$ 在 Λ_A 上相等，从而 $g(\lambda)=G(f_1(\lambda),\cdots,f_l(\lambda))$ 和 $G(r_1(\lambda),\cdots,r_l(\lambda))$ 在 Λ_A 上也相等．但后者是 λ 的多项式，所以

$$g(\boldsymbol{A})=G(r_1(\boldsymbol{A}),\cdots,r_l(\boldsymbol{A}))=G(f_1(\boldsymbol{A}),\cdots,f_l(\boldsymbol{A})).\qquad\square$$

命题 1.3　设复合函数 $g(\lambda)=h(f(\lambda))$ 在 Λ_A 上有意义，则 $g(\boldsymbol{A})=h(f(\boldsymbol{A}))$，即 $g(\boldsymbol{A})=h(\boldsymbol{B})$，$\boldsymbol{B}=f(\boldsymbol{A})$．

证　记 $m_A(\lambda)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$ 是 $\boldsymbol{A}$ 的最小多项式，再记 $f(\lambda_i)=\mu_i$ $(i=1,\cdots,s)$，则 $\varphi(\mu)=(\mu-\mu_1)^{m_1}\cdots(\mu-\mu_s)^{m_s}$ 是 $\boldsymbol{B}=f(\boldsymbol{A})$ 的化零多项式（因为函数 $\varphi(f(\lambda))=(f(\lambda)-f(\lambda_1))^{m_1}\cdots(f(\lambda)-f(\lambda_s))^{m_s}$ 在 Λ_A 上为零．由命题 1.2 可知 $\varphi(f(\boldsymbol{A}))=\boldsymbol{0}$，从而 $\boldsymbol{B}$ 的最小多项式必形如

$$m_B(\mu)=(\mu-\mu_1)^{m_1'}\cdots(\mu-\mu_s)^{m_s'},\quad 0\leqslant m_i'\leqslant m_i\quad(i=1,\cdots,s).$$

对函数 $h(\mu)$，取多项式 $r(\mu)$，使

$$r^{(j)}(\mu_i)=h^{(j)}(\mu_i)\quad(i=1,\cdots,s;j=0,1,\cdots,m_i-1).$$

则当然有 $r(\boldsymbol{B})=h(\boldsymbol{B})$．

根据复合函数的求导公式，$g(\lambda)=h(f(\lambda))$ 和 $r(f(\lambda))$ 在 Λ_A 上相等．现在应用命题 1.2 于 $g(\lambda)-r(f(\lambda))$（$r(u_1)+r(u_2)$ 是 u_1,u_2 的多项式！），即得 $g(\boldsymbol{A})-r(f(\boldsymbol{A}))=\boldsymbol{0}$，即

$$g(\boldsymbol{A})=r(f(\boldsymbol{A}))=r(\boldsymbol{B})=h(\boldsymbol{B}).\qquad\square$$

例 1.5　对任一 n 阶方阵 $\boldsymbol{X}$，若记 $\mathrm{Spec}\boldsymbol{X}=\{x_1,\cdots,x_n\}$，则 $\det(\mathrm{e}^{\boldsymbol{X}})=\mathrm{e}^{x_1}\cdots\cdot\mathrm{e}^{x_n}=\mathrm{e}^{\mathrm{tr}\boldsymbol{X}}\neq 0$，即 $\mathrm{e}^{\boldsymbol{X}}$ 一定可逆．反之，设有可逆方阵 $\boldsymbol{A}$，是否存在 $\boldsymbol{X}$，使 $\mathrm{e}^{\boldsymbol{X}}=\boldsymbol{A}$？

解　回答是肯定的．因为 $\boldsymbol{A}$ 是可逆的，所以在复平面上一定有以 $\boldsymbol{A}$ 的所有特征值为内点，但不含原点的单连通区域，在这个区域上可以确定一个（单值）解析函数 $\ln\lambda$，使 $\mathrm{e}^{\ln\lambda}=\lambda$ 成立．由命题 1.3 即得 $\mathrm{e}^{\ln\boldsymbol{A}}=\boldsymbol{A}$，从而方阵 $\boldsymbol{X}=\ln\boldsymbol{A}$ 即为所求．$\square$

例 1.6 对可逆方阵 $\boldsymbol{A}$ 和非零数 p,必有 $\boldsymbol{X}$ 满足 $\boldsymbol{X}^p=\boldsymbol{A}$.

解 $\boldsymbol{X}$ 的存在性可以像例 1.5 一样,从 $(\lambda^{\frac{1}{p}})^p=\lambda$ 得出 $(\boldsymbol{A}^{\frac{1}{p}})^p=\boldsymbol{A}$. 也可以直接从例 1.5 得知,一定存在 $\boldsymbol{B}$,使 $\mathrm{e}^{\boldsymbol{B}}=\boldsymbol{A}$. 因而 $\mathrm{e}^{\frac{1}{p}\boldsymbol{B}}$ 满足 $(\mathrm{e}^{\frac{1}{p}\boldsymbol{B}})^p=\mathrm{e}^{\boldsymbol{B}}=\boldsymbol{A}$. □

上述讨论使人有理由认为:在一个变量为 λ 的通常的函数关系式中,将 λ 代之以方阵 $\boldsymbol{A}$ 后所得的方阵函数关系式仍成立. 但值得指出的是:正像矩阵运算中常常遇到的情形那样——当同时出现多个矩阵时,矩阵乘法的不可交换性将可能引起问题.

例 1.7 当 $\boldsymbol{AB}\neq\boldsymbol{BA}$ 时,$\mathrm{e}^{\boldsymbol{A}+\boldsymbol{B}}$,$\mathrm{e}^{\boldsymbol{A}}\mathrm{e}^{\boldsymbol{B}}$,$\mathrm{e}^{\boldsymbol{B}}\mathrm{e}^{\boldsymbol{A}}$ 可以互不相同.

解 令

$$\boldsymbol{A}=\begin{pmatrix}1&0\\0&0\end{pmatrix},\quad \boldsymbol{B}=\begin{pmatrix}1&1\\0&0\end{pmatrix},$$

则

$$\mathrm{e}^{\boldsymbol{A}}=\begin{pmatrix}\mathrm{e}&0\\0&1\end{pmatrix},\quad \mathrm{e}^{\boldsymbol{B}}=\begin{pmatrix}\mathrm{e}&\mathrm{e}\\0&1\end{pmatrix},\quad \boldsymbol{A}+\boldsymbol{B}=\begin{pmatrix}2&1\\0&0\end{pmatrix},$$

从而

$$\mathrm{e}^{\boldsymbol{A}+\boldsymbol{B}}=\begin{pmatrix}\mathrm{e}^2&\frac{1}{2}\mathrm{e}^2\\0&1\end{pmatrix},\quad \mathrm{e}^{\boldsymbol{A}}\mathrm{e}^{\boldsymbol{B}}=\begin{pmatrix}\mathrm{e}^2&\mathrm{e}^2\\0&1\end{pmatrix},\quad \mathrm{e}^{\boldsymbol{B}}\mathrm{e}^{\boldsymbol{A}}=\begin{pmatrix}\mathrm{e}^2&\mathrm{e}\\0&1\end{pmatrix}. \quad □$$

1.5 矩阵函数的初等因子

这是本讲要讨论的最后一个问题:对给定的函数 $f(\lambda)$ 和方阵 $\boldsymbol{A}$,怎样由 $\boldsymbol{A}$ 的全体初等因子(或 Jordan 标准形)来确定 $f(\boldsymbol{A})$ 的全体初等因子?

由 1.3 节的性质(3)可知,对 $\boldsymbol{A}$ 的一个 Jordan 块 $\boldsymbol{J}_p(\lambda_0)$,$f(\boldsymbol{J}_p(\lambda_0))$ 是所讨论的 n 阶上三角特殊方阵,即

$$f(\boldsymbol{J}_p(\lambda_0))=\begin{pmatrix}f(\lambda_0)&\dfrac{f'(\lambda_0)}{1!}&\cdots&\dfrac{f^{(p-1)}(\lambda_0)}{(p-1)!}\\ &f(\lambda_0)&\ddots&\vdots\\ &\boldsymbol{0}&\ddots&\dfrac{f'(\lambda_0)}{1!}\\ &&&f(\lambda_0)\end{pmatrix},$$