数值代数

蔡大用 编著

SHU ZHI DAI SHU

清华大学出版社

数 值 代 数

蔡 大 用

清 华 大 学 出 版 社

内 容 提 要

本书共分三部分，第一部分矩阵理论，主要是补充大学课程中这方面知识的不足;第二部分古典迭代法；第三部分为投影类方法和半迭代法．内容取自作者近年来研究成果和有关杂志上最新发表的文章．

本书可作为计算数学有关专业的大学生和研究生教材，和有关人员的参考书．

数 值 代 数

蔡 大 用

*

清华大学出版社出版

北京 清华园

中国科学院印刷厂排版印刷

新华书店北京发行所发行

*

开本：850×1168 1/32 印张：7 字数：186 千字

1987 年 9 月第 1 版 1987年 9 月第 1 次印刷

印数：00001～10000

统一书号：平装15235・305 定价：平装 1.40 元

精装15235・313 精装 3.40 元

前　言

可以毫不夸张地说，相当一部分计算数学问题最终都要化成求解 $AX=b$ 这样的方程组，其中矩阵 A 可以是非奇异的也可能是奇异的，方程可以是超定的也可以是不定的，可以是相容的也可以是不相容的．看来这个问题似乎很简单，但在实际工作中人们往往不能成功地解决它．这个现象促使我们想把这个问题的研究成果和现状作一个拾一漏万的介绍．

此外，从作者的教学工作中感到大学生毕业后为了独立开展研究工作或者解决实际问题，还都需要在矩阵论和数值代数方面补充一些必要的知识，因此想写这本小书，为这些同志提供一些方便．

这本书是根据作者在中国科学院研究生院和清华大学，先后为计算数学专业研究生和高年级大学生讲课所用讲义改写而成．

全书共分三部分，第一部分介绍有关的矩阵理论．这些材料在一般大学的教程中是没有的，其目的在于补充矩阵论方面知识的不足．第二部分介绍了迭代法的古典理论，而第三部分介绍了投影类方法．

这三部分彼此之间没有十分紧密的联系，对于熟悉矩阵理论的读者，可以直接阅读第二、三部分．同样，仅对某一种方法感兴趣的读者，也可以直接阅读有关章节．

本书部分内容曾以油印本形式试用过，在试用过程中胡显承、顾丽珍等同志阅读了某些章节，提出了很有价值的意见并给作者以鼓励，在此表示衷心的感谢．此外，很多大学生和研究生在学习本课程中对本书原稿中的疏漏给以订正，并提出了很多宝贵意见，在此对他们一并致谢．

目　　录

第一部分　矩阵理论

第二部分 古典迭代方法

第三部分 投影类方法和半迭代法一般理论

第一部分　矩阵理论

这部分是计算数学中常要用到的有关矩阵理论和技术的某些课题，主要用来补充大学生在线性代数和计算方法课程中这方面知识的不足.

这部分内容共包含五章. 第一章是有关矩阵分解的知识；第二章介绍了有关非负矩阵的几个主要定理；第三章讨论了矩阵特征值的估计和摄动理论；第四章主要叙述了 Perose-Moore 意义下的广义逆矩阵 A^+ 和有关最小二乘问题的基本知识；第五章扼要地叙述了稀疏矩阵技术的基本知识和常用的排序方法.

各章之间没有十分紧密的联系，对某一课题感兴趣的读者可以抽出其中一章阅读，碰到少量的符号和预备知识上的困难时，再稍微往前翻阅一下有关部分就可以了.

第一章 矩阵分解定理

本章用较短的篇幅介绍几个矩阵分解定理．在后续章节中要不止一次地用到其中某些结果．当然，这些定理不仅仅是为后面章节作准备，它们在不同的应用领域中都有独立意义．

§1 QR 分解和 Schur 定理

矩阵的 QR 分解在近代计算数学中起着十分重要的作用，它是计算矩阵特征值以及求解线性代数方程的一个重要工具． 而 Schur 定理是很多重要定理证明的出发点． 所以我们首先介绍这两个结果．

定义 1.1.1 设 u, v 是两个 n 维复(列)向量， σ 为一复数，

$$E(u, v, \sigma) \equiv (I - \sigma uv^*)$$

称为**初等矩阵**．其中 I 为 $n \times n$ 的单位矩阵． v 可以看成 $n \times 1$ 的复矩阵，则 v^* 为其转置共轭矩阵．

初等矩阵有下面两个简单性质：

(1) $E(u, v, \sigma)E(u, v, \tau)$

$$= E(u, v, \sigma + \tau - \sigma\tau v^*u) \tag{1.1.1}$$

显然，如果 $\sigma^{-1} + \tau^{-1} = v^*u$，则

$E(u, v, \sigma) = E^{-1}(u, v, \tau)$

(2) $\det E(u, v, \sigma) = 1 - \sigma v^*u$ (1.1.2)

很多数值代数中常用的矩阵都可以用初等矩阵表示出来．下面我们给出几个例子．

例 1.1.1 初等交换矩阵 $I_{i,j}$ 是由单位矩阵 I 交换其第 i, j

两行后得到的矩阵. 则:

$$I_{i,j} = E(e_i - e_j, e_i - e_j, 1) \tag{1.1.3}$$

其中 $e_i = [\underbrace{0, 0, \cdots, 1}_{i\text{ 列}}, 0, \cdots]^T$, 而上标 T 代表矩阵的转置.

例 1.1.2 初等三角矩阵.

$$L_i \equiv \begin{array}{c} \left[\begin{array}{ccccc} 1 & & & & \\ & \ddots & & 0 & \\ & & 1 & \cdots\cdots & \\ & & l_{i+1,i} & \ddots & \\ & 0 & \vdots & 0 & \\ & & l_{n,i} & & 1 \end{array}\right] \cdots i\text{ 行} \\ i\text{ 列} \end{array}$$

这是在高斯消去法中消去第 i 列下三角部分元素时用到的矩阵. 不难直接验证:

$$L_i = E(l_i, e_i, -1) \tag{1.1.4}$$

其中 $l_i = [0, 0, \cdots 0, l_{i+1,i}, \cdots l_{n,i}]^T$

例 1.1.3 在求对称实矩阵的全部特征值的 Jacobi 方法中要用到如下形式的旋转矩阵:

$$R = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

其中 θ 为一实数, 它代表旋转的角度. 旋转矩阵 R 可以表示为两个初等矩阵之积,即:

令

$$w_1 = \begin{bmatrix} \sin\dfrac{\theta}{2} \\ \cos\dfrac{\theta}{2} \end{bmatrix}, \quad w_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

则可以验证

$$R = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = E(w_1, w_1, 2)E(w_2, w_2, 2) \tag{1.1.5}$$

在 (1.1.5) 式中用到的两个特殊的初等矩阵在后续讨论中占有重要地位,为了突出其重要性,我们给出

定义 1.1.2 设w为一复向量，且 $(w, w)=1$，这里$(\cdot, \cdot)$是普通意义下的内积．则 $E(w, w, 2)$ 称为初等 Hermit 矩阵．有些文献中也称之为 Hausholder 矩阵．

设H为一初等 Hermit 矩阵，则下面两个简单性质可以直接从(1.1.1) 和 (1.1.2) 式得出：

$$H=H^*=H^{-1}, \text{ 且 } \det H=-1. \tag{1.1.6}$$

下面的引理是讨论 QR 分解的基础．

引理 A 设a是一n维复向量，则存在一个$n\times n$的初等 Hermit 矩阵 H，使得 $Ha=\alpha e_1$．其中 α 为一复数，$e_1=[1,0,\cdots,0]^T$．

证明

我们显然摈除 $a=0$ 和 $a=\alpha e_1$ 的平凡情况．

令所求矩阵 $H=I-2ww^*$，其中 w 是一个待定复向量，如果我们能够求出 w，满足引理要求也就证明了引理．

因为 $Ha=\alpha e_1$，H 为初等 Hermit 矩阵，从 (1.1.6) 第一式可知它是一个酉阵，对向量 a 施行酉变换之后它的欧氏范数应保持不变，所以

$$|\alpha|^2=\|a\|_2^2 \tag{1.1.7}$$

从而

$$\alpha=\|a\|_2 e^{i\phi} \tag{1.1.8}$$

其中ϕ是α的幅角，目前它仍然是个待定值．又从

$$Ha=(I-2ww^*)a=\alpha e_1$$

我们再考虑到 (1.1.8) 式，则有

$$2ww^*a=a-\|a\|_2 e^{i\phi}e_1 \tag{1.1.9}$$

用 $2a^*$ 左乘上式两端，则有

$$4(a^*w)(w^*a)=2(a^*a)-2\|a\|_2 e^{i\phi}\bar{a}_1 e_1$$

即

$$4|w^*a|^2=2\|a\|_2^2-2\|a\|_2 e^{i\phi}\bar{a}_1 \tag{1.1.10}$$

a_1 是向量 a 的第一个支量．不妨先假设 $a_1\neq 0$．从 (1.1.10) 式

可以看出 $e^{i\phi}\bar{a}_1$ 应为实数，即 ϕ 应取值，使得

$$e^{i\phi}\bar{a}_1 = \pm |a_1| \tag{1.1.11}$$

从 (1.1.10) 和 (1.1.11) 式可得

$$2|w^*a| = \sqrt{2\|a\|_2^2 \pm 2\|a\|_2|a_1|} \equiv q \tag{1.1.12}$$

所以

$$2w^*a = qe^{i\phi} \equiv \gamma \tag{1.1.13}$$

其中 γ 为一复数，ϕ 为其幅角，目前依然是个待定值．把 (1.1.13) 式代入 (1.1.9) 式就有，

$$w = \frac{1}{\gamma}(a - \|a\|_2 e^{i\phi} e_1) \tag{1.1.14}$$

可以直接验证 w 的欧氏范数为 1，从而 H 是一个初等 Hermit 矩阵．而从 H 的求解过程可知引理成立． □

下面我们可以很容易地给出这节的主要定理．

定理 1.1.1 设 A 为 $n \times n$ 的复矩阵，则存在有酉阵 Q 和上三角阵 R，使得

$$A = QR.$$

证明 令 $A = [a_1^{(0)}, a_2^{(0)}, \cdots, a_n^{(0)}]$ 其中 $a_i^{(0)}$ 是 A 的第 i 个列向量．根据引理A 存在有初等 Hermit 阵 $H(w_1)$ 使得

$$H(w_1)a_1^{(0)} = \alpha_1 e_1 \tag{1.1.15}$$

记 $A_1 = H(w_1)A$，则 A_1 有如下形状：

$$A_1 = \begin{bmatrix} \alpha_1 & \beta_2 & \cdots & \beta_n \\ 0 & & & \\ \vdots & a_2^{(1)} & \cdots & a_n^{(1)} \\ 0 & & & \end{bmatrix}$$

其中 $a_i^{(1)}$ $i = 2, \cdots, n$ 都是 $n-1$ 维的复向量．

再根据引理 A，存在初等 Hermit 阵 $\widetilde{H}(w_2)$ 使得

$$\widetilde{H}(w_2)a_2^{(1)} = \alpha_2 e_2$$

其中 $e_2 = [1, 0, \cdots, 0]^T$，但它是 $n-1$ 维的单位向量．令

$$H(w_2) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \widetilde{H}(w_2) & \\ 0 & & & \end{bmatrix}$$

则

$$A_2 = H(w_2)H(w_1)A = \begin{bmatrix} \alpha_1 & \beta_1 & \beta_2 & \cdots & \beta_n \\ 0 & \alpha_2 & \gamma_3 & \cdots & \gamma_n \\ \vdots & 0 & & & \\ \vdots & \vdots & a_3^{(2)} & \cdots & a_n^{(2)} \\ 0 & 0 & & & \end{bmatrix}$$

其中 $a_i^{(2)}$ $i = 3, \cdots, n$ 是 $n-2$ 维的向量. 依这种办法进行 $n-1$ 步，则有

$$A_{n-1} = H(w_{n-1})H(w_{n-2})\cdots H(w_1)A$$

$$= \begin{bmatrix} \alpha_1 & & & \\ & \alpha_2 & \gamma_{i,j} & \\ & & \ddots & \\ & 0 & & \alpha_n \end{bmatrix} = R \qquad (1.1.16)$$

显然，R 为一上三角阵.

从 (1.1.16) 式，可得

$$A = H(w_1)\cdots H(w_{n-1})R = QR$$

其中 $Q = H(w_1)H(w_2)\cdots H(w_{n-1})$，为 $n-1$ 个酉阵之积，故 Q 为一酉阵. □

从引理 A 的证明过程中可以看出，向量 w 的取法并不唯一，因此矩阵 A 的 QR 分解一般说来是不唯一的. 为了明瞭这些分解式之间到底有多大差别，我们有

定理 1.1.2 设 A 为一 $n \times n$ 的非奇异矩阵，则在 $A = QR$ 分解中，Q 的每一列和 R 的每一行最多可以差一个 $e^{i\phi}$ 形式的因子.

证明 设 A 有两种不同的 QR 分解式，即

$$A = Q_1R_1 = Q_2R_2 \qquad (1.1.17)$$

因为 A 是非奇异的，故 R_1，R_2 也都是非奇异的. 从 (1.1.17) 式我们有

$$Q_2^*Q_1 = R_2R_1^{-1} \qquad (1.1.18)$$

上式左端为一酉阵，故 $R_2R_1^{-1}$ 必为一酉阵. 又 R_1，R_2 都是上三角阵，故 $R_2R_1^{-1}$ 也是上三角阵. 从而 $R_2R_1^{-1}$ 为一上三角形的酉

阵. 我们可以断言,矩阵 $R_2R_1^{-1}$ 必为一对角线阵，而且主对角线上所有元素的模必为 1. 即:

$$Q_2^*Q_1 = R_2R_1^{-1} = D \tag{1.1.19}$$

其中 $D = \mathrm{diag}\{e^{i\phi_k}\}$

故

$$\begin{aligned} Q_1 &= Q_2D & R_2 &= DR_1 \\ \text{或}\quad Q_2 &= Q_1D^{-1} & R_1 &= D^{-1}R_2 \end{aligned} \tag{1.1.20}$$

(1.1.20) 式证明了我们的结论. □

下面的两个推论是十分有用的.

推论 1 设 A 为一 $n \times n$ 的实对称矩阵,则必存在正交阵 Q，使得 $Q^TAQ = S$，其中 S 为一对称三对角线矩阵.

推论 2 设 A 为一 $n \times n$ 复矩阵,则存在有酉阵 Q，使得

$$Q^*AQ = \begin{bmatrix} * & & & & \\ * & * & & & * \\ & * & \ddots & & \\ & & \ddots & \ddots & * \\ 0 & & & & * \end{bmatrix}$$

上式右端的矩阵下次对角线以下的所有元素均为零，这种矩阵称为 Hessenberg 阵.

上述两个推论表明，在一定的条件下矩阵 A 可以正交相似或酉相似于一个特殊类型的矩阵. 它们之所以非常重要是因为三对角线矩阵和 Hessenberg 矩阵在计算机中处理起来非常方便. 更重要的是，根据引理 A，这些相似变换都可以通过**有限次**的算术运算实现.

下面我们再介绍一个在理论上十分重要的定理.

定理 1.1.3 (Schur 定理) 设 A 为 $n \times n$ 的复阵,则存在一个酉阵 Q，使得

$$A = QTQ^*$$

其中 T 为上三角形矩阵.

证明 设 A 的一个特征值为 λ_1,它相应的特征向量为 $x_1 \neq 0$.

根据引理 A，存在初等 Hermit 矩阵 H_1，使得

$$H_1 x_1 = \alpha_1 e_1 \tag{1.1.21}$$

所以

$$H_1 A H_1^{-1} H_1 x_1 = \lambda_1 H_1 x_1 = \lambda_1 \alpha_1 e_1 \tag{1.1.22}$$

而且 $\alpha_1 \neq 0$，$H_1 = H_1^{-1}$

因此，(1.1.22) 式可写成

$$H_1 A H_1 \alpha_1 e_1 = \lambda_1 \alpha_1 e_1 \tag{1.1.23}$$

从上式两端消去 α_1，则有

$$H_1 A H_1 e_1 = \lambda_1 e_1 \tag{1.1.24}$$

从 (1.1.24) 式可以看出

$$H_1 A H_1 = \begin{bmatrix} \lambda_1 & * & * & \cdots & * \\ 0 & & & & \\ \vdots & & & A_2 & \\ 0 & & & & \end{bmatrix}$$

其中 A_2 为一 $(n-1) \times (n-1)$ 的矩阵． 对 A_2 重复上述讨论，依次重复 $n-1$ 次，我们就有

$$H_{n-1} \cdots H_1 A H_1 \cdots H_{n-1} = \begin{bmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \ddots \\ & & & & \lambda_n \end{bmatrix} = T$$

其中 * 代表可能是非零的项．令

$$Q = H_1 H_2 \cdots H_{n-1}.$$

它显然是酉阵，于是得到

$$A = QTQ^* \qquad \square$$

值得指出的是，Schur 定理仅仅指出了矩阵 Q 的存在性，它的证明并不是构造性的，而是以事先知道矩阵 A 的特征值和特征向量为条件．所以它不能成为一种直接用来求矩阵特征值的算法．

在结束本节之前我们再叙述一个有趣的定理．

定理 1.1.4 设 A 为一 $n \times n$ 的复阵，则存在有非奇异矩阵 T，使得

$$T^{-1}AT=\begin{bmatrix}\lambda_1 & b_{12} & \cdots & b_{1n}\\ & \lambda_2 & & \\ & & \ddots & b_{i,j}\\ & 0 & & \ddots \\ & & & \lambda_n\end{bmatrix}$$

而且 $\sum\limits_{i,j}|b_{i,j}|<\varepsilon$，其中 ε 是事先给定的任一正数。

证明 根据 Schur 定理，存在有酉阵 Ω 使得

$$\Omega^{-1}A\Omega=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \rho_{i,j}\\ & & \ddots & \\ & 0 & & \ddots \\ & & & \lambda_n\end{bmatrix}$$

令 $F=\mathrm{diag}\{r,r^2,\cdots r^n\}$，其中 r 为一非零常数，且取 $T=\Omega F$，则

$$T^{-1}AT=F^{-1}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \rho_{i,j}\\ & & \ddots & \\ & 0 & & \ddots \\ & & & \lambda_n\end{bmatrix}F$$

$$=\begin{bmatrix}\lambda_1 & r\rho_{1,2} & r^2\rho_{1,3} & \cdots & r^{n-1}\rho_{1,n}\\ & \lambda_2 & r^2\rho_{2,3} & \cdots & r^{n-1}\rho_{2,n}\\ & & \ddots & & \\ & 0 & & \ddots & \\ & & & & \lambda_n\end{bmatrix}$$

$$=\begin{bmatrix}\lambda_1 & b_{1,2} & b_{1,3} & \cdots & b_{1,n}\\ & \lambda_2 & b_{2,3} & \cdots & b_{2,n}\\ & & \ddots & & \\ & 0 & & \ddots & \\ & & & & \lambda_n\end{bmatrix} \qquad (1.1.25)$$

当给定 $\varepsilon>0$ 后，适当地选取 r，总可以使得

$$\sum_{i,j}|b_{i,j}|<\varepsilon$$

成立. □

乍一看来，如果我们令 $\varepsilon\to 0$，似乎可以得出：任何一个矩阵都可以相似于对角矩阵. 显然这个结论是荒谬的，错在何处？我们留给读者思考.

§2 正规矩阵和可交换矩阵的分解

计算数学中人们常常关心的是通过相似变换可以把一个矩阵变换成何种更简单的形状. 本节我们就介绍这方面两个重要的定理.

定义 1.2.1 设 A 为 $n\times n$ 的复矩阵，且满足

$$A^*A=AA^* \tag{1.2.1}$$

则称 A 为一**正规矩阵**.

定理 1.2.1 设 A 为一正规矩阵，则存在有酉阵 T，使得

$$A=TDT^* \tag{1.2.3}$$

其中 $D=\mathrm{diag}\{\lambda_1,\lambda_2,\cdots\lambda_n\}$

证明 根据 Schur 定理，存在酉阵 Q，使得

$$A=QTQ^* \tag{1.2.4}$$

其中 T 为上三角阵. 则

$$A^*=QT^*Q^* \tag{1.2.5}$$

由 (1.2.4), (1.2.5) 及 (1.2.1) 式可得

$$TT^*=T^*T \tag{1.2.6}$$

考虑到 T 是上三角阵，(1.2.6) 式表明 T 必为对角阵. □

显然，如果一个矩阵酉相似于一个对角矩阵，它必然是正规的. 因此，我们可以说：**一个矩阵酉相似于对角阵的充分必要条件为它是一个正规阵.** 这个论断是众所熟知的有关 Hermit 阵酉相似于对角矩阵这一结论的推广.

下面我们介绍两个对称矩阵 A 和 B “同时”酉相似于对角阵的