数值分析中的矩阵论

〔美〕A. S. 豪斯霍尔德 著

科学出版社

数值分析中的矩阵论

〔美〕A. S. 豪斯霍尔德 著

孙家昶　孙继广　曹维潞　邓健新 译

科学出版社

1986

内 容 简 介

本书用统一观点总结了数值分析的矩阵理论方面的重要研究成果，是一名著．中译本为原书的前三章，即基础理论部分．第一章讨论一些基本工具，以非常简洁的形式叙述数值分析中常用矩阵变换、分解以及对角化方法等；第二章讨论矩阵的模和极限、矩阵算子的界以及非负矩阵；第三章从矩阵分裂和模的理论的角度来讨论特征值的局部化理论，并给出了值域的一般结果及若干不等式．这三章内容以及相应的练习题对于数值分析工作者具有较大的参考价值．中译本还包括译者给出的练习题的答案．

本书可供数值分析工作者以及高等院校有关专业的教师和学生参考．

A. S. Householder
THE THEORY OF MATRICES IN NUMERICAL ANALYSIS
Blaisdell Publishing Company, 1964

数值分析中的矩阵论
〔美〕A. S. 豪斯霍尔德 著
孙家昶 孙继广 曹维潞 邓健新 译
责任编辑 张鸿林

科学出版社出版
北京朝阳门内大街137号

中国科学院印刷厂印刷
新华书店北京发行所发行 各地新华书店经售

*

1986年4月第一版 开本：787×1092 1/32
1986年4月第一次印刷 印张：6 7/8
印数：0001—6,800 字数：155,000

统一书号：13031·3136
本社书号：4656·13—1

定价：1.65 元

序　　言

本书的目的是选取矩阵理论中的某些材料介绍给读者，这些材料对于发展和评价解线性方程组(包括求逆矩阵)和求特征值的计算方法是极其有用的. 线性不等式的求解和线性规划问题不在直接考虑的范围之内，因为在这些问题中存在特殊的困难，它们在很多方面都具有组合问题的特征，需要用完全不同的方法处理.

书末所列出的文献目录足以表明：本书所讨论的问题不仅会使数值分析家们，而且也会使数学家们很感兴趣. 这一目录是从作者所收集的文献资料中挑选出来的，已收集的文献资料有现在所列的两倍之多，但是仍然是不完全的，而且各种出版物的数量还在急剧地增加. 原因是很清楚的. 在一台有限的数字计算机上，泛函方程或无穷方程组只有在进行有限逼近后才能进行求解，而且通常解非线性方程组的第一步便是线性化. 因此，求解有限线性方程组是全部数学计算的核心. 而且，随着科学与技术的发展，计算机成为更加强有力的工具，需要求解的方程组规模更大，这就要求更为精细和更为有效的处理技巧.

本书的目的不在于提出某些特殊的计算方案和计算框图，也不在于过多地直接涉及运算量的估计和误差分析. 有关这方面的材料可以在别的文献中找到. 本书将在适当的场合提到一些专门的文献资料，这里应该特别注意的是 J. H. Wilkinson 的一些论文以及他将要出版的一些书[1].

---

1) 例如 J. H. Wilkinson, The Algebraic Eigenvalue Problem, Oxford 1965.——译者注

本书第一章提出的一些思想在以后各章都要用到，其中不少是经典的概念，但在其它文献中没有受到足够的重视.第二章叙述了范数理论，这个理论在所有的误差分析以及其它方面起着越来越大的作用. 第三章将范数理论直接应用于各种局部化定理，这些定理对于提供特征值的误差界是重要的.第三章还叙述了某些其它有用的结果.

本书假定读者已经熟悉矩阵代数的一般原理：加法、减法、乘法与求逆；Cayley-Hamilton 定理，特征值与特征向量，以及标准型；向量空间的有关概念，线性相关，秩等等. Householder（1953）简明地介绍了这些理论；较为详细的内容可在 Birkhoff 与 Maclane（1953），Macdubbee（1943）以及许多其它标准的教科书中找到.然而，本书第一章叙述的 Lanczos 算法提供了 Cayley-Hamilton 定理的另一种证明.

最后，作者向从各方面为本书的出版作出贡献的人们表示衷心的感谢.

目 录

第一章　一些基本恒等式与不等式

1.0　目的，记号

本章及下一章叙述一些以后各章要反复使用的基本工具和关系式．我们首先引入一种形式十分简单的矩阵．它是由单位矩阵减去一个秩不超过 1 的矩阵而得到的．这里把任何这种形式的矩阵都称为"初等矩阵"．在求逆矩阵及用相似变换将矩阵化成便于求特征值的形式时，这种形式的矩阵起着十分重要的作用．它们也将直接用于导出某些对于求逆矩阵及解方程组均有重要意义的因式分解定理．

在几何上最小二乘法可以解释为将一个任意的给定向量对某个子空间进行投影．某些迭代方法的每一步，同样也可能解释为投影．因此我们将导出投影算子的一般形式，并且，在构造这种算子时，利用了一些刚才提及的因式分解定理．紧接着的一节，在推导某些经典的行列式、恒等式及不等式的过程中，也要利用这些因式分解定理．在分析与推导一些迭代方法及作某些更一般的理论研究时，都需要运用这些恒等式与不等式．

本章的最后两节，介绍和一个任意矩阵或向量相关联的某些多项式．一类是以 Lanczos 命名的算法所定义的多项式．Lanczos 引入这个算法的原意是作为寻求矩阵的特征值及特征向量的第一个步骤；其实，这个算法还有更多的用处．它在解线性方程组的共轭梯度法中以及在实验数据的多项式最小二乘法拟合中都会出现．此外，为了不破坏后面叙述的

连续性，我们把 Lanczos 算法放在这里叙述. 该算法放在这里叙述的另一个优点是：虽然假定读者熟悉矩阵的 Jordan 标准型，但是从 Lanczos 方法所导出的三对角矩阵出发，可以很直接地导出 Jordan 标准型来. 事实上，只需要证明某一类矩阵实际上可将三对角矩阵变换成 Jordan 标准型即可. 因此，愿意这样做的读者，可以按照这条思路首先作出一个关于任意矩阵的 Jordan 标准型的构造性证明. 甚至 Cayley-Hamilton 定理也可以作为一个简单推论而得到，不必预先假设它成立.

由 Lanczos 算法导出的多项式称为 Lanczos 多项式，它是与另一组称为正交的多项式紧密相联系的，其原因将在本章的最后来说明. 对于 Hermite 矩阵，这两组多项式能互相重合. 也可以用别的方式隐含地定义正交多项式. 在第三章研究矩阵特征值的局部化问题时，也用到正交多项式.

本章使用某些记号，为了叙述上的连续性，这里先将它们列举如下.

除了维数、指数与其它特别指明者外，小写希腊字母表示标量；小写拉丁字母表示列向量；大写希腊字母或拉丁字母表示矩阵. 记号 $C_{n,i}$ 表示从 n 个元素中取 i 个的组合的个数. 上标 H (由 Hermite 一词得来)表示共轭转置，T 表示转置，C 表示共轭，A 表示伴随. 1.3 节中用的上标 I 表示广义逆因为非奇异矩阵的广义逆矩阵就是它的逆矩阵，上标 I 有时也用来表示逆. 一般 a_i 是 A 的第 i 列，α_{ij} 是 A 中位于 (i,j) 的元素，e_i 是单位矩阵 I 的第 i 列，并且

$$e = \Sigma e_i.$$

x 的元素记为 ξ_i，y 的元素记为 η_i.

矩阵 $|A|$ 的元素是 $|\alpha_{ij}|$，向量 $|x|$ 的元素是 $|\xi_i|$，即为 A 与 x 的元素本身的绝对值；

$$x \leqslant y \longleftrightarrow \xi_i \leqslant \eta_i \quad 对每个\ i;$$

$$A \leqslant B \longleftrightarrow \alpha_{ij} \leqslant \beta_{ij} \quad \text{对每个 } i \text{ 与 } j.$$

单位矩阵 I 与矩阵

$$J = (e_2, e_3, \cdots, e_n, 0),$$
$$K = J + J^{\mathrm{T}}$$

是一些特殊的矩阵. A 的行列式记为 $\delta(A)$, A 的迹记为 $\tau(A)$. $A\begin{pmatrix} i, j, \cdots \\ k, l, \cdots \end{pmatrix}$是由 A 的第 i 行、第 j 列、…以及第 k 行、第 l 列、…的相应元素组成的行列式. $\lambda_i(A)$ 是 A 的特征值(经常称为 A 的根),常常将它们按如下次序编号:

$$|\lambda_1(A)| \geqslant \cdots \geqslant |\lambda_n(A)|;$$

记号 $\lambda = \lambda(A)$ 表示 λ 是 A 的特征值; $\sigma_i(A) \geqslant 0$ 为 A 的奇异值,其中

$$\sigma_i^2(A) = \lambda_i(A^{\mathrm{H}}A);$$

A 的谱半径是

$$\rho(A) = \max_i |\lambda_i(A)|.$$

如果 $\lambda = \lambda(A)$, 那么当 $x \neq 0$, 并且等式

$$Ax = \lambda x$$

成立时称 x 属于 λ. 在这种情况下, x 是一个特征向量. $y \neq 0$ 为属于 λ 的主向量是指对某个 ν 有

$$(A - \lambda I)^{\nu} y = 0,$$

并且称 ν 为该主向量的次数. 因此,特征向量是一次的主向量.

1.1 初等矩阵

形如

$$E(u, v, \sigma) = I - \sigma u v^{\mathrm{H}} \tag{1}$$

的矩阵称为初等矩阵. 因为

(2) $$E(u,v;\sigma)\ E(u,v;\tau)=E(u,v;\sigma+\tau-\sigma\tau v^{\mathrm{H}}u)$$
由此推出，当
$$\sigma^{-1}+\tau^{-1}=v^{\mathrm{H}}u$$
时，有
(3) $$E(u,v;\sigma)=E^{-1}(u,v;\tau).$$
按元素展开容易验证其行列式值为
(4) $$\delta[E(u,v;\sigma)]=1-\sigma v^{\mathrm{H}}u.$$

假设 (P,Q) 是一个 n 阶非奇异矩阵，并且 $P^{\mathrm{H}}Q=0$. 因而 P 与 Q 的列是相互正交并且互补的子空间的基底. 于是，对于满足 $v^{\mathrm{H}}a\neq 0$ 的任意向量 a 与 v，可以选取 σ 与 u，使得对于任一向量 b，如果
(5) $$Ea=b,$$
则必有
(6) $$P^{\mathrm{H}}b=P^{\mathrm{H}}a,\qquad Q^{\mathrm{H}}b=0.$$
这就是说，a 通过 E 变换成向量 b，b 与 Q 的空间正交，并且在 P 的空间中与 a 具有相同的投影. 为看出这点，只要用 P^{H} 与 Q^{H} 分别乘以(5)，利用(6)，并注意到 u 满足方程组
$$P^{\mathrm{H}}u=0$$
$$Q^{\mathrm{H}}u=Q^{\mathrm{H}}a/(\sigma v^{\mathrm{H}}a).$$
由假定这个方程组系数矩阵 $(P,Q)^{\mathrm{H}}$ 是非奇异的.

矩阵
(7) $$L_i(l_i)=E(l_i,e_i;1),\ e_j^{\mathrm{T}}l_i=0,\ j\leqslant i$$
与它的转置都称为初等三角形矩阵，且
(8) $$L_i^{-1}(l_i)=L_i(-l_i),\delta(L_i)=1.$$

矩阵
(9) $$I_{ij}=E(e_i-e_j,\ e_i-e_j;1)$$
称为初等置换矩阵. 显然有
(10) $$I_{ii}=I,\ \delta(I_{ij})=-1,\ i\neq j.$$

一个矩阵左乘 $I_{ij}(i\neq j)$ 的效果是第 i 行与第 j 行对换，右乘 I_{ij} 则是交换 i 列与 j 列.

初等 **Hermite** 矩阵(或初等酉阵)是形如

$$(11)\qquad H(w)=E(w,w;2),\ w^{\mathrm{H}}w=1$$

的矩阵，这样的矩阵既是 Hermite 阵，同时又是酉阵，也是对合阵:

$$(12)\qquad H=H^{\mathrm{H}}=H^{-1},\ \delta(H)=-1.$$

这个变换把空间对于与 w 正交的超平面进行反演. 如果 a 与 b 是长度相等的向量，$a^{\mathrm{H}}a=b^{\mathrm{H}}b$，且 $a^{\mathrm{H}}b=b^{\mathrm{H}}a$ (即为实数)，则总可能选取 w 使

$$Ha=b.$$

事实上，若 $a^{\mathrm{H}}b$ 为实数，令

$$2\mu w=a-b,$$

通过把 $a-b$ 规范化求得实数 2μ，可以验证 $\mu=w^{\mathrm{H}}a=-w^{\mathrm{H}}b$. 这意味着反演是对 a，b 之间的角平分超平面进行的. 类似地，可以固定 a 的某些分量，如果向量 b 落在补子空间中，这个反演在该补子空间中进行.

1.2 某些因式分解

应用正交关系(1.1.7)可直接验证

$$(1)\qquad \begin{aligned}L&=L_1(l_1)\,L_2(l_2)\,L_3(l_3)\cdots\\&=I-l_1e_1^{\mathrm{T}}-l_2e_2^{\mathrm{T}}-l_3e_3^{\mathrm{T}}-\cdots.\end{aligned}$$

因为任何一个单位下三角形矩阵(对角线元素均等于 1)可写成上式右边所示的形式，由此可知任何一个单位下三角形矩阵总可分解成(1)的形式. 因为下三角形矩阵的转置是上三角形矩阵，所以立刻得到单位上三角形矩阵的分解. 单位上三角形矩阵又可写成

$$R=I-r_2e_2^{\mathrm{T}}-r_3e_3^{\mathrm{T}}-\cdots e_i^{\mathrm{T}}r_j=0,\ i\geqslant j.$$

因此另一种分解是

$$R = \cdots(I - r_3e_3^{\mathrm{T}})(I - r_2e_2^{\mathrm{T}}).$$

取转置，得到单位下三角形矩阵的另一种分解．任何一个非奇异的三角形矩阵可以写成一个对角矩阵与一个单位三角形矩阵的乘积．因此，任何一个非奇异的下三角形矩阵可分解为一系列初等下三角形矩阵与一个对角矩阵的乘积．

假设

(2) $$A = A^{(0)} = (a_1^{(0)}, a_2^{(0)}, \cdots, a_m^{(0)}) \neq 0$$

是任意 n 行、m 列的矩阵．我们提出某些算法把 A 表示成如下形式：

(3) $$A = MP,$$

其中 M 是有 p 个线性无关列的矩阵，P 是有 p 个线性无关行的矩阵，p 是 A 的秩．

假设 j_1 是满足 $a_{j_1}^{(0)} \neq 0$ 的最小下标，并且假设 $e_i^{\mathrm{T}} a_{j_1}^{(0)} \neq 0$．这就是说，如果在 i_1 位置上的元素不为零，则在 $I_{1i_1}A^{(0)}$ 中的位置 $(1, j_1)$ 上的元素也不为零，并且对于某一个 L_1^{-1}（可能是单位矩阵），矩阵

$$A^{(1)} = L_1^{-1} I_{1i_1} A^{(0)}$$

的前 $j_1 - 1$ 列为零，而在第 j_1 列第一个元素以下为零．

在 $A^{(1)}$ 中每一行的第一个元素以下可能全为零，在这种情况下，这个算法已经完成．否则，如果 $n = 2$，则取 $A^{(2)} = A^{(1)}$，算法也已完成．如果 $n \neq 2$，则挑出在第一个元素下面含有非零元素的第一列，假定这是 $a_{j_2}^{(1)}$，那么对于某个 i_2，

$$e_2^{\mathrm{T}} I_{2i_2} a_{j_2}^{(1)} \neq 0.$$

因此对于某个 L_2^{-1}，矩阵

$$A^{(2)} = L_2^{-1} I_{2i_2} A^{(1)}$$

的前 $j_1 - 1$ 列全为零，这以后的 $j_2 - j_1$ 列的第一元素以下全为零，紧接着的一列的第二个元素以下全为零．照此办理，

直至得到 $A^{(p)}$，其中或 $p=n$，或第 p 行之下全是零. 结果，A 可表示成以下形式：

$$A = M'A^{(p)},$$

其中

$$M' = I_{1i_1}L_1I_{2i_2}L_2\cdots,$$

并且是非奇异矩阵，$A^{(p)}$ 的前 p 行线性无关，其余各行（如果有的话）全为零. 在后一种情况下，如果丢掉 $A^{(p)}$ 的 $n-p$ 个零行，划去 M' 的后 $n-p$ 列，这个乘积仍然成立. 把所得到的这两个矩阵分别表示为 P 和 M，这就得到所要求的形式(3). 显然，A 本身的秩是 p.

若 A 是方阵并且是非奇异的，有一种可能是取

$$i_1=1,\ i_2=2,\ i_3=3,\cdots$$

这是说，在任何阶段都不需要进行置换. 在这种情况下，M 是单位下三角形矩阵，P 是上三角形矩阵. **当一个非奇异阵 A 能表示成一个单位下三角形矩阵与一个上三角形矩阵的乘积时，这种形式的分解是唯一的.** 事实上，我们在 1.4 节中将求出这两个矩阵的元素的唯一行列式表示. 然而由假设，A 是非奇异的并且

$$A = M_1P_1 = M_2P_2.$$

其中 M_1 与 M_2 是单位下三角形矩阵，P_1 与 P_2 是上三角形矩阵，那么

$$M_2^{-1}M_1 = P_2P_1^{-1}.$$

但上式左边的乘积是单位下三角形矩阵，右边的乘积是上三角形矩阵，因此它们都是对角矩阵，并且左边乘积中的对角线元素均等于 1. 因此，每个乘积只能是 I，故 $M_1=M_2$，$P_1=P_2$.

当 A 是方阵时，作为计算 $\delta(A)$ 的一个方法，这种三角化的方法是 Chió 方法. Gauss 消去法是求 A 的逆的第一步（由三角形矩阵求逆来实现）. 但是，除非自始至终采取精确的算

术运算(没有舍入误差或截断误差)，这种方法的这个简单形式不能作为一种实用算法，它不能使我们达到上述的任何一个目的．有关计算的细节在后面讨论．

初等 Hermite 矩阵能用于另一种分解．再次考虑 n 行、m 列的任意矩阵

$$A = A^{(0)} = (a_1^{(0)}, a_2^{(0)}, \cdots, a_m^{(0)}) \neq 0.$$

设 $a_{j_1}^{(0)}$ 是第一个非零列，形成 $H_1 = H(w_1)$，使$H_1 a_{j_1}^{(0)}$ 是 e_1 的某个倍数．因此设

$$A^{(1)} = H_1 A^{(0)}.$$

注意，这里用不着置换．如果 $A^{(1)}$ 中第一行下面有非零行，设 $a_{j_2}^{(1)}$ 是第一个与 e_1 正交的有非零分量的列，形成 $H_2 = H(w_2)$，使 $a_{j_2}^{(1)}$ 与 e_1 正交的分量反演成 e_2 的一个倍数．如果 $n = 2$，则 $H_2 = I$．继续下去，构成矩阵 $A^{(p)}$，使得

$$A = H A^{(p)},$$

$$H = H_1 H_2 H_3 \cdots,$$

并且 $p = n$ 或者 $A^{(p)}$ 的每一行的第 p 个元素以下都是零．这个矩阵H当然不是 Hermite 矩阵，然而是酉阵．若 $p < n$，则丢掉 $A^{(p)}$ 中的全零行与H中相应的列，得到的矩阵分别记为R与W，于是有

(4) $$A = WR, \quad W^{\mathrm{H}} W = I,$$

其中R有 p 个线性无关行，W有 p 个线性无关列．

若A是非奇异的，则W是酉阵，R是上三角形矩阵．假设在这种情况下以任何方式获得了两种分解

$$A = W_1 R_1 = W_2 R_2,$$

其中 R_1 与 R_2 都是上三角形矩阵，W_1 与 W_2 都是酉阵．那么

$$W_2^{\mathrm{H}} W_1 = R_2 R_1^{-1},$$

因此 $R_2 R_1^{-1}$ 既是酉阵又是上三角形矩阵．容易验证，这样的矩阵必定是对角矩阵(见练习 4)，并且

$$W_2^{\mathrm{H}} W_1 = R_2 R_1^{-1} = D$$

是酉阵也是对角矩阵，因此

$$|D| = I,$$

这就是说，对角元是单位圆上的点．于是

$$W = W_1 = W_2 D, \quad R_2 = DR_1 = DR,$$
$$A = WR = (WD^{-1})(DR).$$

因此，当A是非奇异的并且按(4)分解时，W的每一列与R的每一行除了差数量因子外唯一确定，这个数量是单位圆上的一个点．

如果A本身是酉阵，则R是对角矩阵．因此，每一个酉阵都可以表示成一系列初等 Hermite 矩阵与一个对角线元素在单位圆上的对角矩阵的乘积．

我们还可以用其它方法导出分解式(4)．假设A的各列线性无关．W的第一列是A的规范化的第一列．(注意，引入一个在单位圆上的点作为数量因子不影响规范化．)选取W的第二列是某个单位向量，使得A的第二列a_2是W前二列的线性组合．选取W的第三列是某个单位向量，使得a_3是W前三列的线性组合．容易验证，当这个算法完成时，所需要的分解也就得到了．这个方法称为 **Schmidt** 正交化，是大家所熟悉的．

1.3 投影和广义逆

设矩阵F的各列线性无关．向量Fx是由F的所张成的空间中的一个向量．如果$a - Fx$与F正交：

$$F^{\mathrm{H}}(a - Fx) = 0,$$

那么Fx是一向量a在此空间中的正交投影．因此只要

(1) $$x = (F^{\mathrm{H}}F)^{-1} F^{\mathrm{H}} a,$$

Fx就是a在F的空间中的正交投影，因为F的各列线性无

关，所以逆矩阵 $(F^{H}F)^{-1}$ 必定存在(从 1.4 节有关矩阵的定理直接推出)．因此

(2) $$P_F = F(F^{H}F)^{-1}F^{H}$$

称为对于 F 的空间的投影子，这是因为对任意的 a，$P_F a$ 都是 a 在 F 的空间上的正交投影．它是投影子，并且不只是某一个投影子，因为如果 M 是非奇异的，FM 代替 F，得到的是同一个 P_F．这一点可以直接验证．注意到 P_F 是 Hermite 矩阵，而且

(3) $$P_F^2 = P_F,$$

即 P_F 还是幂等阵．此外

$$(I - P_F)^2 = I - P_F,\ (I - P_F)P_F = 0,$$

而且 $I - P_F$ 是 F 的补子空间的投影子．

任何一个 Hermite 幂等阵 P 都是一个投影子．为了证明这一点，设 P 的秩为 p，并且把 p 表示成如下形式：

$$P = FR^{H},$$

其中 F 与 R 均有 p 个线性无关列．因为 P 是 Hermite 矩阵，并且是幂等的，所以有

$$P = FR^{H} = RF^{H} = FR^{H}RF^{H}.$$

因此

$$F^{H}FR^{H} = F^{H}FR^{H}RF^{H},$$

又因为 $F^{H}F$ 是非奇异的，所以

$$R^{H} = R^{H}RF^{H}.$$

因此

$$R^{H}R = R^{H}R\ F^{H}FR^{H}R.$$

又因为 $R^{H}R$ 是非奇异的，所以

$$F^{H}FR^{H}R = I,$$

因此

$$P = FR^{H}RF^{H} = F(F^{H}F)^{-1}F^{H} = P_F,$$