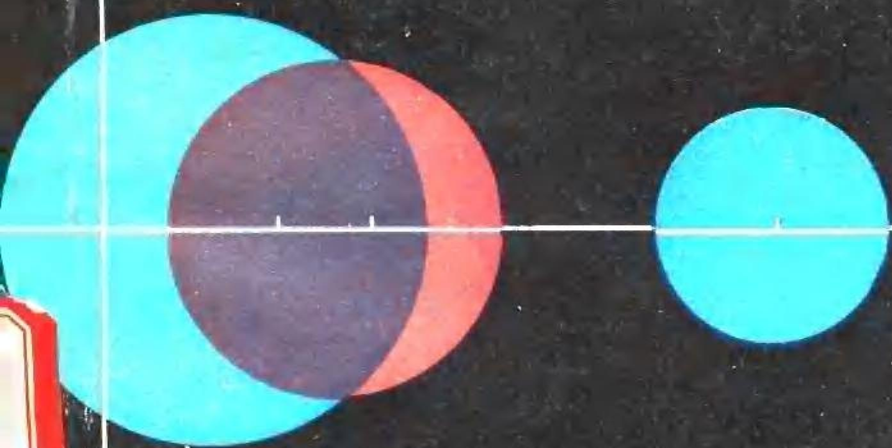
数值分析

〔美〕J·M·奥特加著
张丽君 张乃玲 朱政华译



高 等 教 育 出 版 社

数 值 分 析

〔美〕J. M. 奥特加 著

张丽君 张乃玲 朱政华 译

高等教育出版社

本书根据 James M. Ortega 所著 Numerical Analysis 一书 1973 年第二版译出. 原书是"Computer Science and Applied Mathematics"丛书中的一本，是为计算机科学和数学专业一年级研究生写的，主要介绍数值分析中舍入误差、离散化误差和收敛性误差等方面的一些基本理论.

本书可供计算数学工作者及数值分析研究人员参考，亦可供计算数学专业高年级学生教学参考.

在编辑加工过程中，张德荣先生予以大力帮助，并且提出不少宝贵意见，对此我们表示衷心感谢.

数 值 分 析

〔美〕J. M. 奥特加 著

张丽君 张乃玲 朱政华 译

高等教育出版社出版

新华书店北京发行所发行

北京水利局印刷厂印装

开本 850×1168 1/32 印张 6.25 字数 147,000

1983年6月第1版 1984年7月第1次印刷

印数 00,001—14,100

书号 13010·0895 定价 1.20 元

原　　序

本著作的首要目的是介绍数值分析中舍入误差、离散化误差和收敛性误差(这儿看作是关于序列的收敛性,因而理解为迭代法的收敛性)等三个较重要的问题的某些基本理论结果．事实上,本书是以这些基本问题为主,以线性方程、微分方程等通常的领域为辅而组织的．

在上面提到的三种基本类型的误差分析中，重要和普遍的方面是稳定性的一般概念，它以几种不同的形态出现．数学不稳定性问题,或数值分析中所谓“病态”问题,作为基本误差的背景在第一部分中单独处理．另一方面，涉及特殊方法的数学和数值稳定性结果则分散在本书的其余部分．

作为本书基础的讲义曾在一段时间内在马里兰大学作为计算机科学和数学专业一年级研究生的一学期课程．这样，本书对学习基础差别较大的学生提供了数值方法的理论和分析的常用知识,这些学生可能从事或不从事数值分析的较高级的工作．

但是,本教材也可以给大学生讲授．事实上,所需预备知识都包含在线性代数和高等微积分学两门课程中;而且,我们对特别重要的线性代数的某些部分还在第一章中进行了复习．不言而喻,也假定学生学过数值分析方面的一些大学课程并熟悉诸如解线性方程的 Gauss 消去法，解非线性方程的 Newton 法，解微分方程的 Runge-Kutta 法和多步法等常用的数值方法．但是，书中的材料基本上是自足的，使得虽未在数值分析方面事先工作过然而目的明确且有素养的学生能够掌握这门课程．最后，一些常微分方程的知识是有用的，但不是必备的．

作者对马里兰大学 Werner Rheinboldt 教授，James Vandergraft 教授，圣地亚哥的加里福尼亚大学 Willian Gragg 教授，特别对康奈大学 Jorge Moré 教授表示感谢．他们提出许多有益的意见和建议，并对 Dawn Shifflett 夫人对原稿的熟练打字表示感谢．

常用记号表

$\boldsymbol{R}^n$	实 n 维空间
$\boldsymbol{C}^n$	复 n 维空间
$\boldsymbol{x}^T$	向量 $\boldsymbol{x}$ 的转置
$\boldsymbol{x}^H$	向量 $\boldsymbol{x}$ 的共轭转置
$L(R^n, R^m)$	从 R^n 到 R^m 的线性算子[若 $m=n$, 记作 $L(R^n)$]
A^{-1}	矩阵 A 的逆
$\det A$	矩阵 A 的行列式
$\operatorname{diag}(a_1, \cdots, a_n)$	对角阵
$\rho(A)$	A 的谱半径
$\boldsymbol{e}_i$	第 i 个坐标向量 $(0, \cdots, 0, 1, 0, \cdots, 0)^T$
$\\| \quad \\|$	任意范数
$\\| \quad \\|_p$	l_p 范数
$K(A)$	条件数 $\\|A\\|\\|A^{-1}\\|$
$K_p(A)$	$\\|A\\|_p\\|A^{-1}\\|_p$
$R^m \times R^n$	笛卡尔[乘]积 $=\{(\boldsymbol{x}, \boldsymbol{y}): \boldsymbol{x} \in R^m, \boldsymbol{y} \in R^n\}$
$F: R^n \to R^m$	定义域在 R^n 中而值域在 R^m 中的映照
$o(h)$	o; $f(h)=o(h^p)$ 意指当 $h \to 0$ 时, $h^{-p}f(h) \to 0$
$O(h)$	O; $f(h)=O(h^p)$ 意指当 $h \to 0$ 时, $h^{-p}f(h)$ 为有限
$\cup$, $\cap$	并集, 交集
∂_i	偏导数
$\forall$	对一切
$\triangle$	证明结束

引 言

曾经把数值分析定义为①："研究用于求得数学问题近似解的方法和过程." 这个定义虽然相当概括，但它正确地指出了数值分析中某些关键性的课题，即近似解（通常不能理所当然地指望得到精确解）；数学问题；方法和过程的研究. 让我们举一特例来说明和讨论这个定义.

考虑二阶非线性微分方程

$$y''(t)=f(y(t)), \quad a\leqslant t\leqslant b, \tag{1}$$

具有边界条件

$$y(a)=\alpha, \quad y(b)=\beta. \tag{2}$$

这里 f 是给定的函数，α, β 是给定的常数．我们假设这个边值问题有唯一解；但是通常对它不存在"闭型"的表示，而必需去考虑一个产生近似解的方法．求近似解的常用方法之一是**有限差分法**．首先，用格点

$$a=t_0<t_1<\cdots<t_{n+1}=b$$

分划区间 $[a, b]$；为简单起见，我们设点 t_i 是等距的，间距是 h，于是 $t_i=a+ih$, $i=0, 1, \cdots, n+1=(b-a)/h$．其次，用标准二阶差商

$$y''(t_i)\doteq\frac{1}{h^2}[y(t_{i+1})-2y(t_i)+y(t_{i-1})], \quad i=1, \cdots, n$$

来逼近 $y''(t_i)$．因此，在(1)中应用这个近似式，就有

$$y(t_{i+1})-2y(t_i)+y(t_{i-1})\doteq h^2f(y(t_i)), \quad i=1, \cdots, n,$$

并且，若求出 n 个（一般是非线性的）方程的方程组

① S. Parter, Comm. ACM 12 (1969), 661—663.

$$y_{i+1}-2y_i+y_{i-1}=h^2f(y_i),\quad i=1,\cdots,n,$$
$$y_0=\alpha,\quad y_{n+1}=\beta \tag{3}$$

的解，谁都希望，解分量 y_i 是(1)在点 t_i 的精确解 $y(t_i)$ 的一个近似.

一般说来，找到方程(3)的显式解亦是不可能的，而人们能用某些迭代过程得到一个近似解. 例如，对(3)的最简单的这种迭代过程之一是

$$y_i^{k+1}=y_i^k-\frac{[2y_i^k-y_{i-1}^k-y_{i+1}^k+h^2f(y_i^k)]}{2+h^2f'(y_i^k)},\quad i=1,\cdots,n, \tag{4}$$

其中上标 k 表示迭代次数.（有时，这方法叫 Newton-Jacobi 方法；在第三部分对迭代法给出了比较充分的讨论.）现在，即使迭代收敛到解，即若

$$\lim_{k\to\infty} y_i^k\to y_i,\quad i=1,\cdots,n,$$

我们将在迭代有限次 k 以后被迫停止并取 y_i^k 作为 y_i 的一个近似. 但是，一般说来，不能精确地计算 y_i^k，因为在执行(4)的计算过程中有舍入误差. 因此，譬如说，我们以 $\hat{y}_i^k$ 来结束计算，$\hat{y}_i^k$ 是 y_i 的一个近似值，因此是 $y(t_i)$ 的近似值.

上面给出的例子说明了在数值方法研究中误差的三个主要来源. 给出任意一个"连续问题"（例如，任一个微分方程或积分方程），为了求它的近似解，第一步是化为仅含有限多个变量的"离散模拟". 于是，方程(3)构成关于方程(1)和(2)的一个离散模拟. 离散模拟中的变量不一定是连续解在所选点上的实际函数值，却可能是 Fourier 系数等等. 总之，离散模拟的精确解和原来解法的精确解之间的差称为（整体的）**离散化误差**，数值分析的主要问题是对这个误差给出估计，通常用原来问题的初始数据与离散参数一起表示. 对上面的例子，人们希望知道当 $h\to0$ 时 $y_i=y_i(h)$ 的性态；特别是离散问题的解是否趋于原来问题的解，如果这样，其

速度如何. 在第二部分，我们将得到关于常微分方程初值问题和边值问题离散化误差的估计.

误差的第二个主要来源是把一个无限序列终止于有限项而产生的，这称为**收敛性误差**. 这类误差在很多场合出现(例如无穷级数的求和)，但它在应用迭代过程时，出现得最为普遍. 在上面的例子中，收敛性误差当然是差$|y_i^k-y_i|$. 主要在第三部分，我们将对某些常用迭代方法研究这类误差.

最后，还有舍入误差的问题，它实质上渗透到每个计算中. 第9章将致力于获得用消去法解线性方程组的舍入误差的界限.

上述三种误差来源的基础都是**不稳定性**的一般概念，不稳定性以不同的形态在各种场合出现，但通常不稳定性的意义是"小变化引起大变化". 首先，求解时存在解的稳定性问题；一般地，若问题的初始数据的"小"的变化，仅引起解的"小"的变化，解称为稳定的. 例如，如果(2)的边值α和β的小变化，在(1)的解中仅引起小的变化，那么，我们说解是稳定的(关于边值)，若在解中引起大的变化，那么，说它是不稳定的或者(用数值分析中比较常用的语言)称为**病态**的. 病态问题实质上影响到一大类数学问题，在第一部分，我们将举几个例子，对稳定和不稳定问题进行研究.

其次，还存在离散模拟的稳定性，例如，可能出现这种情况，原来问题的解是稳定的，但离散模拟的解是不稳定的. 正如我们在第5章中将要看到的，尤其对初值问题，这是一个严重的潜在问题. 因为，微分方程的最一般的离散模拟是差分方程[(3)可以作为差分方程来考虑]，这就导致我们在第4章中象考虑微分方程解的稳定性一样去考虑差分方程解的稳定性.

因为任何迭代方法可以看作是一个差分方程，对于差分方程初值问题解的(渐近)稳定性准则精确地对应于迭代过程的(局部)收敛性准则，因此，(在这个特定的意义上)稳定性和研究序列的收

敛性误差的某些问题是等价的. 这种关系将在第三部分进一步加以探讨.

最后，存在着“数值稳定性”问题或在舍入误差影响下的稳定性问题. 在这里所说的意义下，这意味着至少对良态问题，使最后确定的计算机算法具体实现时，一连串算术运算将不会导致舍入误差的灾难性的积累. 例如，没有行交换的 Gauss 消去法过程在上述意义下是一个(潜在的)不稳定方法. 这种类型的稳定性和舍入误差分析是交织在一起的，将在第 9 章中研究.

目　录

第一章　线性代数

在数值分析的许多领域中，最重要的工具是线性代数和矩阵理论．对于在线性代数中提出的计算问题，如求解线性方程组、计算矩阵的特征值和特征向量等等，这句话肯定正确而且十分自然．对于相当大量的其他领域，如非线性方程、微分方程、逼近论等等，这句话也是正确的．在这些领域中，相应的数值方法的分析常常取决于从线性代数引出的结果．在更进一步的工作中，无限维线性代数（泛函分析）起着类似的作用．

总之，线性代数将为本书提供大部分基础．在本章，我们将复习那些对我们最有用的线性代数的许多基本部分．此外，比较特殊的结果还分散在本书的其他各节．

1.1　特征值和标准形

我们用 R^n 表示分量为 $x_1, \cdots, x_n$ 的列向量 $\boldsymbol{x}$ 的实 n 维空间，并用 C^n 表示相应的复空间．对于 $\boldsymbol{x}\in R^n$，$\boldsymbol{x}^T$ 表示其转置，它是行向量 $(x_1, \cdots, x_n)$，若 $\boldsymbol{x}\in C^n$，$\boldsymbol{x}^H=(\bar{x}_1, \cdots, \bar{x}_n)$ 是其共轭转置．

从 R^m 到 R^n 的线性算子的集合——或等价地，如文中所述，实 $n\times m$ 矩阵的集合——记为 $L(R^m, R^n)$，若 $m=n$，简单地记为 $L(R^n)$．类似地，复 $n\times m$ 矩阵的集合记为 $L(C^m, C^n)$，若 $m=n$，记为 $L(C^n)$．矩阵 $A\in L(C^m, C^n)$ 的元素写作 a_{ij}．我们指出，当写作 $A\in L(C^m, C^n)$ 时，决不排除 A 实际上是实的可能性．

若 $A\in L(R^m, R^n)$，则 A^T 表示 A 的转置，如果 A 是复的，则 A^H 表示其共轭转置．对任何 $n\times n$ 矩阵 A，$\det A$ 是 A 的行列式，而 A^{-1} 是 A 的逆．假如 A^{-1} 存在，我们说 A 是非奇的．我们来

回忆下列可逆性的基本定理.

1.1.1 令 $A\in L(C^n)$, 则下列命题是等价的:

(a) A 是非奇的;

(b) $\det A\neq 0$;

(c) 线性方程组 $A\boldsymbol{x}=\boldsymbol{0}$ 只有一个解 $\boldsymbol{x}=\boldsymbol{0}$;

(d) 对任一向量 $\boldsymbol{b}$, 线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 有唯一解;

(e) A 的列(和行)是线性无关的; 即若 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n$ 是 A 的列, 且 $\alpha_1\boldsymbol{u}_1+\cdots+\alpha_n\boldsymbol{u}_n=\boldsymbol{0}$, 则数 α_i 全为零.

最后一个条件(1.1.1(e))又可简单地说成 A 的秩是 n, 通常, 把矩阵的线性无关列(或行)的数目定义为秩.

若 $A\in L(C^n)$, 如果一个(实或复)数 λ 和一个非零向量 $\boldsymbol{x}$ 满足

$$A\boldsymbol{x}=\lambda\boldsymbol{x}, \tag{1}$$

则把 λ、$\boldsymbol{x}$ 分别称为 A 的特征值和特征向量. 由定理 1.1.1 可得 λ 为特征值的充要条件是

$$\det(A-\lambda I)=0. \tag{2}$$

这是 A 的特征方程. (I 总是单位阵)因此, A 正好有 n 个(不一定不相同)特征值, 即(2)的 n 个根. 这 n 个特征值 $\lambda_1, \cdots, \lambda_n$ 的集合叫做 A 的谱, 且

$$\rho(A)=\max_{1\leqslant i\leqslant n}|\lambda_i| \tag{3}$$

是 A 的谱半径.

通常特征值难以计算,但是有一类重要的矩阵,它们的特征值可由观察得到. 这些矩阵是(上或下)三角阵:

$$A=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ 0 & & a_{nn} \end{bmatrix}, \quad A=\begin{bmatrix} a_{11} & & 0 \\ \vdots & \ddots & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

很清楚, 三角阵的特征值正好是主对角线元素. 三角阵的一

个重要的特殊情况是对角阵

$$D=\begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{bmatrix},$$

我们通常用 $D=\mathrm{diag}(d_1, \cdots, d_n)$ 来表示.

对于一般矩阵 $A\in L(C^n)$, 其特征值可能有任意的分布; 但对常用而且重要的某些类矩阵, 特征值却被约束在复平面的某一部分内. 例如, $A\in L(R^n)$, 如果 $A^T=A^{-1}$, A 叫做**正交阵**; $A\in L(C^n)$, 如果 $A^H=A^{-1}$, A 叫做**酉阵**, 这两类矩阵中任一类矩阵的特征值的模都等于 1(习题 1.1.1). 类似地, $A\in L(R^n)$, 如果 $A^T=A$, A 叫做**对称阵**; $A\in L(C^n)$, 如果 $A^H=A$, A 叫做 **Hermite 阵**. 对称阵和 Hermite 阵的任一特征值全是实的(习题 1.1.2). 此外, 矩阵 $A\in L(R^n)$, 若

$$\boldsymbol{x}^T A\boldsymbol{x}\geqslant 0, \quad \boldsymbol{x}\in R^n, \tag{4}$$

则叫做**半正定**的, 如果当 $\boldsymbol{x}\neq\boldsymbol{0}$ 时, (4) 中不等式严格成立, 叫做**正定**的. 对复矩阵, 用 $\boldsymbol{x}^H$ 代替 $\boldsymbol{x}^T$ 有类似的定义. 一个对称半正定阵或 Hermite 半正定阵 A 的特征值全是非负的, 如果 A 是正定的, 则特征值全是正的(习题 1.1.4).

在矩阵论中, 最重要的运算之一是相似变换. 对两个矩阵 A, $B\in L(C^n)$, 若存在一个非奇阵 $P\in L(C^n)$ 使

$$B=P^{-1}AP, \tag{5}$$

则称 A, B 为**相似**的.

考虑变量的线性变换自然引出这一运算; 即考虑方程

$$\boldsymbol{y}=A\boldsymbol{x} \tag{6}$$

和变量变换 $\quad \hat{\boldsymbol{y}}=P^{-1}\boldsymbol{y}, \quad \hat{\boldsymbol{x}}=P^{-1}\boldsymbol{x},$

那么, 关于变量 $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, 方程(6)化为

$$\hat{\boldsymbol{y}}=B\hat{\boldsymbol{x}}.$$

其中 B 由(5)给出.

下面给出相似变换的一条基本性质.

1.1.2 若 $A, B\in L(C^n)$ 是相似的,则 A 和 B 有相同的特征值.

证明 我们来证明 A 和 B 的特征多项式是相同的. 从行列式的乘法规则立即得出

$$\begin{aligned}\det(A-\lambda I)&=\det(P^{-1}P)\det(A-\lambda I)\\&=\det P^{-1}\det(A-\lambda I)\det P\\&=\det(P^{-1}AP-\lambda I).\end{aligned}\qquad \triangle$$

显然,对 1.1.2 的另一种证明是, $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 成立的充要条件是 $P^{-1}AP\boldsymbol{y}=\lambda\boldsymbol{y}$, 其中 $\boldsymbol{y}=P^{-1}\boldsymbol{x}\neq\boldsymbol{0}$.

其次,我们证明关于对称矩阵的一个重要结果.

1.1.3 令 $A\in L(R^n)$ 是对称的, 则存在实的正交阵 P, 使 P^TAP 是对角阵.

证明 令 $\lambda_1, \cdots, \lambda_n$ 是 A 的特征值, $\boldsymbol{x}_1$ 是对应于 λ_1 的一个特征向量,且 $\boldsymbol{x}_1^T\boldsymbol{x}_1=1$; 因为 λ_i 是实的(习题 1.1.2), 我们可以假定 $\boldsymbol{x}_1$ 是实向量. 令 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{n-1}$ 是 $n-1$ 个互相正交的向量, 且皆和 $\boldsymbol{x}_1$ 正交, 并且 $\boldsymbol{u}_i^T\boldsymbol{u}_i=1$, $i=1, \cdots, n-1$. 其次, 定义 $n\times(n-1)$矩阵 $U_1=(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{n-1})$, 它的各列是 $\boldsymbol{u}_i$, 与此相应再定义第一列为 $\boldsymbol{x}_1$ 的 $n\times n$ 矩阵 $P_1=(\boldsymbol{x}_1, U_1)$. 则正交性条件就意味着 P_1 是正交的且 $U_1^T\boldsymbol{x}_1=\boldsymbol{0}$. 因此

$$P_1^TAP_1=\begin{bmatrix}\boldsymbol{x}_1^T\\ U_1^T\end{bmatrix}(\lambda_1\boldsymbol{x}_1,\ AU_1)=\begin{bmatrix}\lambda_1 & 0\\ 0 & U_1^TAU_1\end{bmatrix}.$$

因为

$$\begin{aligned}\det(A-\lambda I)&=\det(P_1^TAP_1-\lambda I)\\&=(\lambda_1-\lambda)\det(U_1^TAU_1-\lambda I),\end{aligned}$$

$(n-1)\times(n-1)$ 对称矩阵 $A_2=U_1^TAU_1$ 有特征值 $\lambda_2, \cdots, \lambda_n$. 所

以可以对 A_2 作同样的变换，且可以断定存在 $(n-1)\times(n-1)$ 正交矩阵 Q_2，使

$$Q_2^T A_2 Q_2 = \begin{bmatrix} \lambda_2 & 0 \\ 0 & A_3 \end{bmatrix},$$

其中 A_3 是 $(n-2)\times(n-2)$ 对称矩阵，它的特征值是 $\lambda_3, \cdots, \lambda_n$. 于是

$$P_2^T P_1^T A P_1 P_2 = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & A_3 \end{bmatrix},$$

其中 P_2 是正交矩阵

$$P_2 = \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}.$$

这样继续下去，显然可以作出正交矩阵序列 $P_1, \cdots, P_n$，使

$$P_n^T \cdots P_1^T A P_1 \cdots P_n = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

而矩阵的积 $P = P_1 \cdots P_n$ 也是正交的. △

我们指出，如果 P 取酉阵，那么对 Hermite 阵有同样的结果.

令 $$D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$$

是 1.1.3 中的对角阵，其中 λ_i 是 A 的特征值. 令 $\boldsymbol{p}_i$, $i=1, \cdots, n$, 表示 P 的各列，则对

$$AP = PD \tag{7}$$

按列相等，得出

$$A\boldsymbol{p}_i = \lambda_i \boldsymbol{p}_i, \quad i=1, \cdots, n; \tag{8}$$

即 P 的各列是 A 的特征向量. 因此 1.1.3 的一个等价的形式为：实对称矩阵有 n 个正交的特征向量.

这一结果的几何解释如下. 考虑关系式

$$\boldsymbol{x}^T A \boldsymbol{x} = c = \text{常数}, \tag{9}$$

这是 R^n 中的一个"二次曲线"方程. 例如，如果 A 是正定的，(9) 就是一个椭球方程. 现在作变量变换 $\hat{\boldsymbol{x}} = P^T \boldsymbol{x}$，这里 P 是 1.1.3