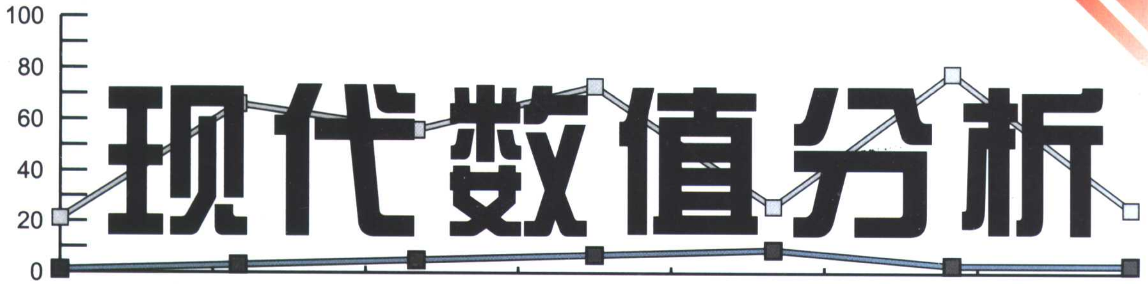
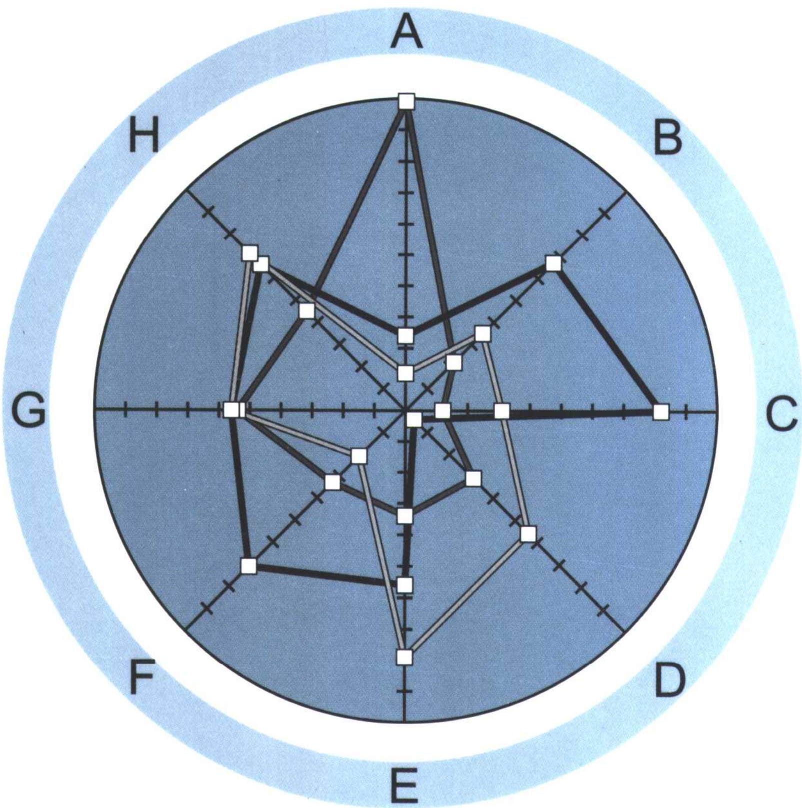
现代数值分析



蔺小林 蒋耀林 编著



国防工业出版社

http://www.ndip.cn

现代数值分析

蘭小林　蒋耀林　编著

国防工業出版社
·北京·

图书在版编目(CIP)数据

现代数值分析/蔺小林．蒋耀林编著．—北京:国防工业出版社,2004.9

ISBN 7-118-03560-2

Ⅰ．现… Ⅱ．①蔺…②蒋… Ⅲ．数值计算 Ⅳ．0241

中国版本图书馆 CIP 数据核字(2004)第 073089 号

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

北京奥隆印刷厂印刷

新华书店经售

*

开本 710×960 1/16 印张 24 457 千字

2004 年 9 月第 1 版 2004 年 9 月北京第 1 次印刷

印数:1—4000 册 定价:28.00 元

内 容 简 介

本书比较全面地介绍了现代科学与工程计算中常用的数值计算方法，介绍了这些数值计算方法的基本理论及其在某些工程技术方面的应用，同时也对这些数值计算方法的计算效果、稳定性、收敛效果、适用范围以及优劣性与特点作了简要的分析。全书共分 9 章，包括数值方法研究内容及误差分析，解线性方程组的直接解方法与迭代方法，解非线性方程(组)的迭代方法，矩阵特征值与特征向量的数值算法，函数插值，函数逼近，数值积分与数值微分，以及常微分方程初值问题与边值问题的数值解法等。本书基本概念清晰，语言叙述通俗易懂，理论分析严谨，结构编排由浅入深，在分析问题时注重启发性，例题选择具有针对性，同时注重实际应用。各章附有一定数量的习题，供读者学习时进行练习。

本书可作为高等院校理工科研究生和数学、信息与计算科学、物理、计算机等高年级本科生使用，也可供从事科学与工程计算的科技工作者参考。

前　　言

随着科学技术的飞速发展,科学与工程计算愈来愈显示出其重要性。科学计算、科学实验与科学理论三足鼎立,科学计算成为科学实践的重要手段之一,其应用范围已经渗透到科学活动的各个领域。作为科学与工程计算的数学工具,现代数值分析从20世纪80年代起,就相继成为各高等院校工科硕士研究生学位公共必修课和数学、信息与计算科学、物理、计算机等本科生的专业基础课。

本书比较全面地介绍了现代科学与工程计算中常用的数值计算方法,对这些数值计算方法的基本理论与实际应用进行了较详细的分析,同时简要的分析了这些数值算法的计算效果、稳定性、收敛效果、适用范围以及优劣性与特点。全书共分9章,包括数值方法研究内容及误差分析,解线性方程组的直接方法和迭代方法,解非线性方程(组)的迭代方法,矩阵特征值与特征向量的数值算法,函数插值,函数逼近,数值积分与数值微分,以及常微分方程初值问题与边值问题的数值解法等。

本书的特点是:第一,语言简单明了,叙述通俗易懂。考虑到学习者的知识结构和不同层次,作者在对问题的叙述和分析时,尽量使语言简单明了,通俗易懂,做到理论联系实际。具有高等数学和线性代数知识的学生都可以使用本教材。第二,基本概念清晰,理论分析严谨。在分析问题时注重启发性,例题选择具有针对性,注重实际应用效果。通过内容学习,给学习者建立一条思考问题的清晰思路。第三,取材全面合理,观点较新。本书既对经典的数值方法进行了较全面的介绍,也增加了一些方法的推广和最新发展,以适应不同学习者的需要。此外各章附有一定数量的习题,供读者学习时进行练习,书后附有部分习题的解答或参考答案。

本书全部讲完需要80学时左右,教师可根据学习者的情况及实际学时,有选择地讲解部分内容。

本书第1章、第5章、第9章由蒋耀林执笔,第2章～第4章和第6章～第8章由蔺小林执笔,最后由蒋耀林统一定稿。

本书在出版过程中得到了西安交通大学数学系的大力支持,我们在此深表谢意。由于时间仓促,水平有限,书中难免有错误与疏漏之处,敬请各位同仁批评指正。

作　者

2004年8月

目　录

第1章 引 论

21世纪,信息科学进入飞速发展的时代,使用计算机进行科学计算、数据处理及分析已成为人类科技活动的主要方法之一。熟练地运用计算机进行科学计算,已成为科技工作者的一项基本技能。为了快速、高效地进行数据分析,有必要研究和掌握适用于计算机上使用的数值分析方法。

本章介绍了现代数值分析的研究内容及特点,分析了误差的种类、误差来源和减小误差的一般性原则。

1.1 数值分析研究内容

现代数值分析是应用数学研究的一个重要分支,它是研究如何高效地应用计算机进行科学计算的数值方法及理论的一门科学,是程序设计的重要依据,是对数值结果进行理论分析的基础。

应用计算机进行科学与工程计算、解决实际问题需要经历以下几个主要过程:

(1) 实际问题。在科学研究及工程技术中,各个领域都有许多实际问题,我们需要这些不同领域的专家提出具有明确意义的问题,给出该问题符合本领域所固有的规律或自然法则。

(2) 数学模型。对不同领域专家提出的实际问题,用辩证唯物主义的思想进行分析。在抓住事物主要因素和合理的假设下,运用该领域中的规律,结合数学理论、方法和工具,建立起问题中各种量之间的联系,从而得到完备的数学模型。

(3) 数值分析。对数学模型先从数学理论上进行分析,研究解的存在性、惟一性。只有在满足解的存在惟一性条件下,才能进行数据计算。有些问题可以给出解析解,但在大多数情况下,要对数学模型进行数值计算。把连续模型如何离散化,用什么样的方法进行计算,算法的相容性、收敛性、稳定性等,都是数值分析的研究内容。

(4) 算法设计。算法设计在用计算机进行科学计算过程中起到了非常重要的作用。一个收敛快、精度高的好算法,有时可以比飞速发展的计算机硬件更有实用价值。

例如，在很多科学技术中都要求解线性代数方程组，而我们在线性代数课程中介绍过求解线性方程组的理论和精确解求法，例如用克莱姆法则可以十分整齐简洁地给出一个非奇异线性代数方程组解的表达式。若要在计算机上按克莱姆法则求解一个含有 n 个未知量 n 个方程的非齐次线性方程组，需要计算 $n+1$ 个 n 阶行列式的值，在不计加减运算情况下，总共要进行 $n!\ (n+1)+n$ 次乘除运算，其计算量是非常惊人的。例如 $n=20$，则 $20!\ (20+1)+20 \approx 9.7 \times 10^{20}$，即使使用每秒运算 1 亿次的计算机进行计算也需要大约 30 万年的时间。而用我们在第二章介绍的高斯消去法进行计算，大约需要 3060 次乘除法运算，并且 n 越大，相差就越大。这个例子表明，算法的好坏对提高计算能力起着非常重要的作用。在 1955 年—1975 年的 20 年间，计算机的计算速度提高了数千倍，而同一时间解决一定规模的椭圆型偏微分方程的数值计算方法的效率提高了约 100 万倍。

从程序设计的观点来看，算法是由一个或多个进程组成，每个进程精确地描述了按一定顺序执行的有限序列，所有进程能够同时执行并且协调地在有限个步骤内完成一个给定问题的求解。若算法含有一个进程，则称其为串行算法，否则称为并行算法。若算法在算术运算(加减乘除)过程中占据了总时间的绝大部分，这种算法就称为数值型算法，否则称为非数值型算法。

算法在保证可靠性的前提下有优劣之分。算法的可靠性主要包括：收敛性、稳定性、误差估计等。算法的优劣主要表现在时间复杂度(计算机运行时间)、空间复杂度(占据计算机储存空间以及逻辑复杂度(影响程序开发的周期以及维护的难易程度)。

(5) 软件实现。即根据算法编写科学计算软件。由于数值分析所研究对象以及解决问题的广泛适用性，目前已有的软件如：Maple、Matlab、Mathematica 等，基本上已将数值分析的主要内容设计成简单函数，只要调用运行便可得到结果。在实际工作中，由于所面临的问题具有明确的特征，其复杂性有时已超出书本所述例证的范围，因而有必要深入掌握数值分析的基本思想方法和具体内容。

现代数值分析将具体问题分解成一系列子问题进行系统研究。主要有：线性代数方程组求解、非线性方程(组)求解、矩阵的特征值与特征向量的计算、函数插值、函数逼近、数值积分与数值微分以及微分方程数值解法。

1.2 误差基础知识

对数学问题进行数值求解，求得的结果一般情况下都含有误差，即所得结果多数情况下都是近似值。对这些结果的误差进行分析和估计是数值计算过程中的主要内容。通过对它们的研究可以确切地知道误差的性态和误差的范围。

1.2.1 误差的来源

数值计算中的数可分为两类:一类是精确反映实际情况的数,这些数称为**精确数**、**准确数**或**真值**。如某学校五年级一班有 48 名学生,数字 48 是精确数。另一类数只能近似地反映实际情况,这类数称为近似数或某精确数的**近似值**。如今天早晨的温度是 14℃,一般说来 14℃ 不能精确地反映今天早晨的精确温度,它只是一个近似数值。数的准确值与其近似值之差称为**误差**。在数值的计算过程中误差是不可避免的,在大多数情况下不存在严格的精确数。因此,分析误差产生的原因,把误差限制在允许的范围内是非常必要的。

误差的来源即误差产生的原因,一般来说源于固有误差和计算误差。而固有误差包括模型误差和观测误差,计算误差包括截断误差和舍入误差,下面分别介绍。

模型误差是在建立数学模型时产生的误差。这是因为在定量地分析客观事物时,总是抓住主要矛盾,忽略次要矛盾,从而建立本质已知量与未知量之间的数量关系,即建立起数学模型。因此,这个数学模型与实际的客观事物之间存在一定差距,这种差距就是误差,即模型误差。

观测误差是用物理工具测量数据时由观察而得到一个量(或称为观测值),此量与实际精确值之间有差距。观测误差有时也称为测量误差或参数误差。观测值的精度依赖于测量仪器的精密程度和操作仪器的人等因素。

顺便指出,固有误差不是数值分析研究的内容。

截断误差是数学模型本身的准确解与用数值方法而求得的解的差距,截断误差也称为方法误差。这种误差常常是由于用有限过程来逼近无穷过程而产生的误差,如要计算数项级数

$$1+\frac{1}{2!}+\frac{1}{3!}+\cdots+\frac{1}{n!}+\cdots=\sum_{n=1}^{\infty}\frac{1}{n!}$$

的值,我们用计算机计算时,用前 n 项(有限项)和

$$1+\frac{1}{2!}+\frac{1}{3!}+\cdots+\frac{1}{n!}=\sum_{k=1}^{n}\frac{1}{k!}$$

来代替。这样就舍弃了数项级数后边无穷多项,因而产生误差,它的截断误差为

$$\sum_{k=n+1}^{\infty}\frac{1}{k!}$$

舍入误差是计算机中参加运算的数据与原始数据之间的差距。这是因为计算机在执行算法时,由于受到计算机字长的限制,参加运算的数据只能具有有限位,原始数据在机器中表示时可能会产生误差,当然每次运算后又可能会产生新的误差,这些误差都是舍入误差。如 $\pi=3.14159265\cdots$, $e=2.7182818284590\cdots$, $\sqrt{3}=$

1.73205080757…等都不可能用全部小数位参加计算机运算,必须取有限位,即进行“四舍五入”,从而参加计算机运算的数与这些实际的数之间有误差,这就是舍入误差。

在数值分析中主要研究截断误差和舍入误差。

1.2.2 绝对误差和相对误差

设 x^* 是精确值 x 的一个近似值,则称

$$e(x^*) = x - x^*$$

为近似值 x^* 的**绝对误差**,简称**误差**。

显然 $e(x^*)$可正也可负,当 $e(x^*)>0$ 时,x^* 称为 x 的弱(不足)近似值;当 $e(x^*)<0$ 时,x^* 称为 x 的强(过剩)近似值。$|e(x^*)|$的大小标志着 x^* 的精度。一般地,在同一量的不同近似值中,$|e(x^*)|$越小,x^* 的精度越高。

绝对误差依赖于量纲,一般无法准确地知道绝对误差 $e(x^*)$的大小。但可以根据测量与计算的情况,事先给出误差绝对值的一个范围,也就是去估计$|e(x^*)|$的上界。

若令

$$|e(x^*)| = |x - x^*| \leqslant \varepsilon$$

则称 ε 为x^*的**绝对误差界**,简称**误差界**。

通常情况下,我们有

$$|x - x^*| \leqslant \varepsilon$$

即

$$x^* - \varepsilon \leqslant x \leqslant x^* + \varepsilon$$

这表明精确值 x 在区间$[x^* - \varepsilon, x^* + \varepsilon]$内,用 $x = x^* \pm \varepsilon$ 来表示近似数x^*的精确度,或准确值所在的范围。

注意,绝对误差的大小,在许多情况下还不能完全刻画一个近似值的精确程度,如两个数

$$x = 10 \pm 0.1,\ y = 10^{20} \pm 10^6$$

此处 y 的绝对误差是x 绝对误差的 10^7 倍,但是我们不能就据此断言近似值 $x^* = 10$ 一定比近似值 $y^* = 10^{20}$的精确度高。若考虑到精确值本身的大小,在 10^{20}内相差 10^6 显然比在 10 内相差 0.1 要更精确些。这表示一个数的近似值的精确度,不仅与绝对误差有关,还与这个数本身有关。

设 x^* 是精确值 x 的一个近似值,称

$$e_r(x^*) = \frac{e(x^*)}{x} = \frac{x - x^*}{x}$$

为近似值 x^* 的**相对误差**。

相对误差说明了近似数 x^* 的绝对误差 $e(x^*)$与 x 本身比较所占的比例，它反映了一个近似数的准确程度，相对误差越小，精确度就越高。相对误差是一个无量纲的量，通常用百分数表示。与绝对误差一样，我们只能给出相对误差绝对值的一个上界。当

$$|e_r(x^*)| \leqslant \varepsilon_r$$

时，则称 ε_r 为近似值 x^* 的**相对误差界**。

在实际应用中，由于真值 x 总是未知的，因此常取相对误差为

$$e_r(x^*) = \frac{x - x^*}{x^*}$$

这是因为

$$\frac{x - x^*}{x^*} - \frac{x - x^*}{x} = \frac{(x - x^*)^2}{x^* x} = \left[\frac{e(x^*)}{x^*}\right]^2 \left[\frac{1}{1 + \dfrac{e(x^*)}{x^*}}\right]$$

当

$$\left|\frac{e(x^*)}{x^*}\right| \leqslant \frac{1}{2}$$

时，有

$$\left|1 + \frac{e(x^*)}{x^*}\right| \geqslant 1 - \left|\frac{e(x^*)}{x^*}\right| \geqslant \frac{1}{2}$$

从而

$$\left|\frac{x - x^*}{x^*} - \frac{x - x^*}{x}\right| \leqslant 2\left[\frac{e(x^*)}{x^*}\right]^2$$

当$\left[\dfrac{e(x^*)}{x^*}\right]$很小时，$\dfrac{x - x^*}{x^*}$与$\dfrac{x - x^*}{x}$的差是$\dfrac{e(x^*)}{x^*}$的平方数量级，故可以忽略不计。因此，在实际计算中，可以取

$$e_r(x^*) = \frac{x - x^*}{x^*}$$

1.2.3 有效数字

我们在表示一个近似数时，为了能反映它的精确度，经常用到“有效数字”的概念。

若 x 的某一近似值 x^* 的绝对误差界是某一位的半个单位，则从这一位起直到左边第一个非零数字为止的所有数字都称为 x^* 的**有效数字**。

具体地说，对于数 x 经四舍五入后得到它的近似值为

$$x^* = 0.x_1x_2\cdots x_n \times 10^m$$

其中 x_i 是 0~9 之间的任一个数,但 $x_1 \neq 0, i=1,2,\cdots n$ 为正整数,m 为整数,若

$$|x - x^*| \leqslant \frac{1}{2} \times 10^{m-n}$$

即 x^* 是 x 具有 n 位有效数字的近似值,或称 x^* 精确到第 n 位,$x_1, x_2, \cdots, x_n$ 都是 x 的有效数字。

如 $\pi = 3.14159265\cdots$,当取 3.142 作为 π 的近似值时

$$|\pi - 3.142| = 0.00040735 < 0.0005 = \frac{1}{2} \times 10^{-3}$$

即 $m-n=-3, m=1, n=4$。所以 3.142 作为 π 的近似值有 4 位有效数字。当取 3.141 作为 π 的近似值时

$$|\pi - 3.141| = 0.00059265 < 0.005 = \frac{1}{2} \times 10^{-2}$$

即 $m-n=-2, m=1, n=3$。所以 3.141 作为 π 的近似值时有 3 位有效数字,不具有 4 位有效数字,3.14 是有效数字,千分位的 1 不是有效数字。

如果近似数 x^* 的误差限是某一位的半个单位,由该位到 x^* 的第一位非零数字一共有 n 位,x^* 就有 n 位有效数字,也就是说准确到该位。

若用四舍五入法取准确值的前 n 位作为近似值 x^*,则 x^* 有 n 位有效数字,其中每一位数字也都为 x^* 的有效数字。但是,如果 x^* 准确到某位数字,将这位数字以后的数字进行四舍五入则不一定得到有效数字。

例如 5.145 作为 5.138 的近似值是准确到百分位,若再四舍五入得到 5.15,其最后一位便不是有效数字了,5.15 只有两位有效数字。

关于有效数字还要指出以下几点。

(1) 用四舍五入法取准确值的前 n 位 x^* 作为 x 的近似值,则 x^* 必有 n 个有效数字。

如 $e = 2.7182818284590\cdots$,取 2.718 作为 e 的近似值有 4 位有效数字,取 2.718282 作为 e 的近似值就有 7 位有效数字。

(2) 有效数字位数相同的两个近似数,绝对误差不一定相同。

如 $x_1^* = 10706, x_2^* = 11.104$ 二者都有 5 位有效数字。前者绝对误差为 $\frac{1}{2} \times 1$,后者绝对误差为 $\frac{1}{2} \times 10^{-3}$。

(3) 把任何数乘以 $10^p (p=0, \pm 1, \pm 2, \cdots)$ 等于移动该数的小数点,它并不影响其有效数字的位数。

如 $g = 9.80\text{m/s}^2$ 具有 3 位有效数字,而 $g = 0.00980 \times 10^3\text{m/s}^2$ 也有 3 位有效数字,但是 9.8m/s^2 与 9.80m/s^2 的有效数字是不同的,9.8m/s^2 有两位有效数字,9.80m/s^2 有 3 位有效数字。因此,0.1,0.10,0.100 等含义是不同的。