21 世 纪 高 校 教 材

计算方法

——数值分析

袁东锦 编著

南京师范大学出版社

21世纪高校教材

计算方法

——数值分析

袁东锦 编著

南京师范大学出版社

图书在版编目(CIP)数据

计算方法——数值分析／袁东锦编著．—南京：南京师范大学出版社，2004.7

ISBN 7－81101－094－1／O·23

Ⅰ.计…　Ⅱ.袁…　Ⅲ.计算方法—高等学校—教材
Ⅳ.O241

中国版本图书馆 CIP 数据核字（2004）第 078334 号

书　　名	计算方法——数值分析
编　　著	袁东锦
责任编辑	王书贞
出版发行	南京师范大学出版社
地　　址	江苏省南京市宁海路 122 号(邮编:210097)
电　　话	(025)83598077(传真)　83598412(营销部)　83598297(邮购部)
网　　址	http://press.njnu.edu.cn
E－mail	nnuniprs@public1.ptt.js.cn
照　　排	江苏兰斯印务发展有限公司
印　　刷	扬州鑫华印刷有限公司
开　　本	850×1168　1/32
印　　张	11.5
字　　数	287 千
版　　次	2004 年 8 月第 1 版　2004 年 8 月第 1 次印刷
印　　数	1－3 000 册
书　　号	ISBN 7－81101－094－1／O·23
定　　价	22.60 元

内 容 提 要

本书主要介绍各种数值计算方法以及相关的基本概念和理论.内容主要包括误差问题,非线性方程的数值解,插值与逼近,数值微分和数值积分,解线性代数方程组的直接法和迭代法,矩阵的特征值和特征向量,常微分方程初值问题的数值解法以及非线性方程组的迭代解法等.全书对主要基本算法的推导、构造原理、各种方法的收敛性、误差估计等进行了较详细的讨论,内容取材适当,由浅入深,各章均有例题和适量的习题,附录中给出一些数值算例,以供上机实验的读者参考.

本书可作为理工科院校数学与应用数学、信息与计算科学、计算机科学、力学、物理等专业大学生及其他专业研究生“计算方法”(或“数值分析”)课程的教材,也可供科技工作者和工程技术人员参考使用.

前　　言

随着电子计算机的应用日益广泛，科学计算已成为各学科、领域中的一项重要工作．人们在科学研究和生产实践当中会经常碰到数值计算的问题，如何选择与使用适当的数值方法，如何比较、分析数值方法的收敛速度，如何估计、判断计算结果的误差，如何解释数值计算过程中的异常现象等等，要解决这一切诸如此类的问题，都需要通过学习“计算方法”（或“数值分析”）去解决．

当今，在各类理工科院校中凡开设了线性代数（或高等代数）、数学分析（或高等数学）专业的，几乎都要开设此课程，相关专业的研究生更不例外．这是一门内容丰富、思想方法深刻而又有着自身理论体系的课程，我们不应仅从字面上认为该课程只是各种算法的简单罗列，事实上，它所研究的数值计算问题种类繁多，即使同一类问题也因其解法多种多样、为保证方法的收敛性与数值的稳定性而需要进行一系列理论性的分析与比较．因此，很多业内人士皆认为该课程名为“数值分析”（“Numerical Analysis”）更为适宜．

新一轮的教学改革正在深入进行之中，而教材改革是其中一个重要环节，本课程同样面临这一任务．为此，本人集讲授此课程近三十遍的体会和对教学改革的认识，在对扬州大学多个专业所使用的“计算方法”讲义的基础上，同时注意借鉴中外经典、博采各家之长，修改、增删而成此书．

本书方法叙述与理论推导并重；内容取舍合理，章节次序恰当；既注重习题的数量，又考虑其难度的梯次；对各类基本方法给

出了便于编程的算法描述.附录中还给出了一些数值算例,以供读者在进行实验时参考.教材同时兼顾了各类学校不同专业的教学要求与教学时数(一般为一学期 64～72 学时,适当取舍后也适用于 48 学时或 96 学时).

扬州大学的张天平教授、林支桂教授曾多次使用原"计算方法"讲义,并对本书的编写提出了许多宝贵的建设性的意见,本人在此致谢两位教授.

由于本人水平有限,书中疏漏不当之处在所难免,恳切希望各位有识之士不吝赐教、广大读者予以指正.

袁东锦

2004 年 3 月

目　　录

第一章　绪　论

§1　数值计算方法的任务与算法的概念

在解决实际问题的长期过程中，形成了计算方法这门学科. 快速电子计算机的发展，为科学技术中提出的各种数学问题求得数值解答提供了有力的工具. 随着计算机广泛地用于大型的科学计算，数值计算方法已经成为应用数学的一个重要分支.

运用计算机解决科学计算问题需要经历以下几个主要的步骤：得出实际问题→建立数学模型→选用数值计算方法→程序设计→上机计算得出数值结果(或归结为"五化"：实际问题数学化、数学问题计算化、计算问题公式化、公式问题程序化、上机计算自动化). "计算方法"的任务是研究适用于科学计算的数值计算方法及有关的数学理论，它是程序设计和对数值结果进行分析的依据和基础，是用计算机进行科学计算全过程的一个重要环节.

众所周知，电子计算机具有极高的运算速度，但它只能根据给定的指令完成加、减、乘、除等算术运算和一些逻辑运算. 因此，要使用计算机来求解各种数学问题，诸如方程求根、微分和积分的计算、求解大型线性方程组、微分方程的求解等等，必须把求解过程归结为按一定的规则进行一系列四则算术运算. 计算机只能机械地执行人们所给定的指令，而不会主动地思维，去进行创造性的工作. 交给计算机执行的解题方法的每一步骤都必须加以准确的规

定.我们把对数学问题的解法归结为有加、减、乘、除等基本运算并有确定的运算顺序的完整而准确的描述,称为**数值算法**或简称**算法**."计算方法"就是以数学问题为对象,研究各种数值算法及其有关理论的一门学科.当然,计算方法也有许多自身的特点,解决实际问题时,应当根据问题的要求、计算工具的性能,选择良好的算法,以较高的效率,得出有用的结果.

§2 浮 点 数

1 定点数

设 r 为大于 1 的整数,a_i 为 $0,1,\cdots,r-1$ 中的某一个.位数有限的 r 进制正数可以写成

$$x \triangleq a_{l-1}a_{l-2}\cdots a_0.a_{-1}a_{-2}\cdots a_{-m} \tag{1}$$

x 有 l 位整数,m 位小数.因为进位制的基数是 r,所以

$$x = a_{l-1}\times r^{l-1} + a_{l-2}\times r^{l-2} + \cdots + a_0\times r^0 + a_{-1}\times r^{-1} + a_{-2}\times r^{-2} + \cdots + a_{-m}\times r^{-m} \tag{1'}$$

当 $l=4,m=4,r=10$ 时,

$$109.312,\quad 0.437\,5,\quad 4\,236$$

分别表示为

$$0\,109.312\,0,\quad 0\,000.437\,5,\quad 4\,236.000\,0$$

这种把小数点永远固定在指定位置上,位数有限的数,称为**定点数**.当 $l=m=4$ 而 $r=10$ 时,八位定点非零数中,绝对值最小和最大的数分别为

$$\pm 0\,000.000\,1,\quad \pm 9\,999.999\,9$$

由此可见,定点数所能表示的数的范围非常小.

2 浮点数

设 s 是十进制数,p 是十进制正负整数或零.数 x 可以用 s 和 10^p 的乘积来表示,即

$$x \triangleq s \times 10^{p} \tag{2}$$

再设 s 的整数部分等于零，即 s 满足条件

$$-1 < s < 1 \tag{3}$$

则 x 的小数点位置主要由整数 p 决定. 即使用以表示 s 的位数和用以表示 p 的位数之和不大，也能表示绝对值相当大的数以及绝对值相当小的数. 形如(2)而满足条件(3)的十进制数 x，称为**十进制浮点数**. s 和 p 分别称为浮点数 x 的**尾数**和**阶数**. 如果尾数的小数位数等于有限正整数 t，则把 x 称为 t **位浮点数**.

此外，如果还要求尾数 s 小数点后第一位数字不等于零，也就是要求尾数 s 满足条件

$$10^{-1} \leqslant |s| < 1 \tag{4}$$

则形如(2)而满足条件(4)的浮点数称为**十进制规格化浮点数**. 例如数

$$0.004\,012, \quad 0.321\,7, \quad 284.5$$

的规格化浮点数分别为

$$0.401\,2 \times 10^{-2}, \quad 0.321\,7 \times 10^{0}, \quad 0.284\,5 \times 10^{3}$$

只要 $x \neq 0$，则 x 一定可以表示为规格化浮点数.

上面定义了十进制浮点数和十进制规格化浮点数. 设用任何大于 1 的整数作为数制的基数，并设 s 是 r 进制数，p 是 r 进制正负整数或零，则形如

$$x \triangleq s \times r^{p} \tag{2$'$}$$

并满足条件(3)的数称为 r **进制浮点数**，如果再让 s 的小数点后第一位数字不等于零，即 s 满足条件

$$r^{-1} \leqslant |s| < 1 \tag{4$'$}$$

则形如(2′)并满足条件(4′)的 r 进制数 x 称为 r **进制规格化浮点数**.

3 计算机中的数系

任一计算机只能用有限的位数来表示浮点数的尾数和阶数.

设进位制为 r,阶数 p 满足条件

$$-m \leqslant p \leqslant M \tag{5}$$

其中 m,M 为正整数,它们主要由计算机用多少位数来表示阶数而决定. 如果尾数的小数位数为 t(t 一般比 m 和 M 的位数大若干倍),则计算机的数系由一切阶数满足(5)的 t 位 r 进制浮点数的集合 F 组成.

当 $r=10,t=4,m=M=99$ 时,

$$-0.000\,1\times 10^{-99},\quad 0.000\,1\times 10^{-99}$$

是数系中绝对值最小的非零数,而

$$-0.999\,9\times 10^{99},\quad 0.999\,9\times 10^{99}$$

分别是此数系中的最小数和最大数. 若计算的中间结果超出了上述范围,则称为**溢出**.

当 $10^{-1}\leqslant s<s+10^{-4}<1,-m\leqslant p\leqslant M$ 时,

$$s\times 10^{p} \text{ 和} (s+10^{-4})\times 10^{p}$$

之间的任何数都不属于上述计算机的数系. 一切满足上列条件的两数之差都等于 10^{-4+p}. 由此可见,在计算机数系 F 中,数的个数有限,数系中的每一个数都是有理数. 从整体看,数系中的数分布很不均匀;从局部看,阶数相等的数,又以相等的距离分布在数轴的某一段上.

在计算机中,常用尾数等于零而阶数最小的数来表示零. 例如在上述计算机数系中,用 0×10^{-99} 来表示常数零. 零不能化为规格化浮点数.

使用一定位数的浮点数进行数值计算,可以满足一般实际问题的要求. 所以,虽然计算机数系由一些残缺不全、分布不匀的数组成,但用以解决实际问题却很有效.

§3 误差问题

1 误差的来源

用近似方法解决科学技术问题时，一般有误差，其来源有下列四种.

1° 数学描述和实际问题之间的误差. 用数学模型描述实际问题时，往往抓住主要因素，略去一些次要因素，将实际问题理想化以后，才进行数学概括. 这种描述，虽然相当好地反映了实际情况，但也有误差.

2° 观测误差. 数值问题的原始数据，一般由观测或实验获得. 观测结果和这些数量的实际大小总有误差. 这种误差，称为**观测误差**.

3° 截断误差. 实际计算只能用有限次运算来完成，理论上的精确值往往要求用无限的过程才能求出. 例如，已知 $x>0$，求 e^{-x} 时，由表达式

$$e^{-x}=1-x+\frac{1}{2}x^2-\frac{1}{6}x^3+\cdots \tag{1}$$

取部分和

$$E(x)\triangleq 1-x+\frac{1}{2}x^2-\frac{1}{6}x^3 \tag{2}$$

作为 e^{-x} 的近似值，有

$$e^{-x}-E(x)=\frac{1}{24}e^{-\xi}x^4 \tag{3}$$

其中 ξ 介于零和 x 之间. 这样求出的近似值 $E(x)$ 和精确值 e^{-x} 之间存在着误差. 这种类型的误差，称为**截断误差**.

4° 舍入误差. 计算机数系是有限集，不仅无理数 e、π 等不属于计算机数系，一些有理数，如 1/3 也不属于计算机数系. 常常用计算机数系中和它们比较接近的数来表示它们. 由此产生的误差，

称为**舍入误差**.

观测误差和原始数据的舍入误差,就其来源说,有所不同,而就其对计算结果的影响看,完全一样.数学描述和实际问题之间的误差,往往是计算工作者不能独立解决的,甚至是尚待研究的课题.基于这些原因,在计算方法课程中所涉及的误差,一般指舍入误差(包括初始数据的误差)和截断误差.我们所要做的就是:讨论它们在计算过程中的传播和对计算结果的影响;研究控制它们的影响以保证最终结果有足够的精度;既希望解决数值问题的算法简便而有效,又想使最终结果准确而可靠.

2 绝对误差和相对误差

假设某一数 x 的近似值是 x^*(经常在一数的右上角加星号"*"表示此数的近似值).由

$$\varepsilon(x) \triangleq x^* - x \tag{4}$$

定义的数 $\varepsilon(x)$称为近似数 x^* 的**绝对误差**,简称**误差**.

准确值 x 一般是未知的,因而 $\varepsilon(x)$也是未知的.但往往可以估计出绝对误差的上限,即可以求出一正数 η,使

$$|\varepsilon(x)| \leqslant \eta \tag{5}$$

满足上式的 η 称为 x^* 的**绝对误差限**.

例如,用(2)表示的 $E(x)$作为 e^{-x} 的近似值,当 $0 \leqslant x \leqslant 1$ 时,由(3)式可得

$$|E(x) - e^{-x}| \leqslant \frac{1}{24}$$

有时也用

$$x = x^* \pm \eta \tag{6}$$

来表示(5)式.例如 $x = 0.310\,6 \pm 0.001\,4$ 指的是

$$0.309\,2 \leqslant x \leqslant 0.312\,0$$

绝对误差不足以刻画近似数的精确程度.例如,测量飞机机翼长度时,发生 1 毫米的误差,和测量飞机机翼厚度时,发生 1 毫米

的误差，大不一样．要决定一个近似值的精确程度，除了绝对误差以外，还必须考虑此数本身的大小．这就需要引进相对误差的概念．

绝对误差和准确值的比，即

$$\varepsilon_r(x) \triangleq \frac{\varepsilon(x)}{x} = \frac{x^* - x}{x} \tag{7}$$

称为 x^* 的**相对误差**．(7)式描述了绝对误差和相对误差的关系．

准确值一般是未知的，因而 $\varepsilon_r(x)$ 一般也是未知的．但往往可以估计出相对误差的上限，即可以求出一正数 δ，使得

$$|\varepsilon_r(x)| \leqslant \delta \tag{8}$$

满足(8)式的 δ 称为 x^* 的**相对误差限**．

绝对误差和用 x 表示物理量时使用的单位有关，而相对误差则和单位无关．

例如，以

$$x^* \triangleq 0.872\,7 \times 10^{-90}, \quad y^* \triangleq 0.872\,7 \times 10^{90}$$

分别作为

$$x \triangleq 0.872\,6 \times 10^{-90}, \quad y \triangleq 0.872\,6 \times 10^{90}$$

的近似值，其绝对误差和相对误差分别为

$$\varepsilon(x) = 0.000\,1 \times 10^{-90}, \quad \varepsilon(y) = 0.000\,1 \times 10^{90}$$

$$\varepsilon_r(x) = \frac{0.000\,1}{0.872\,6}, \quad \varepsilon_r(y) = \frac{0.000\,1}{0.872\,6}$$

3 有效数字

为了给出一种近似数的表示法，使之既能表示其大小，又能表示其精确程度，我们来引进有效数字的概念．在计算中常按四舍五入原则得到数 x 的前几位近似值 x^*，例如，设

$$x = \pi = 3.141\,592\,6\cdots$$

经四舍五入，若取三位有效数字得

$$x^* = 3.14, \quad \varepsilon(x) \leqslant 0.002$$

若取五位有效数字得

$$x^* = 3.141\,6,\quad \varepsilon(x) \leqslant 0.000\,008$$

它们的绝对误差都不超过末位数字的半个单位，即

$$|\pi - 3.14| \leqslant \frac{1}{2} \times 10^{-2}$$

$$|\pi - 3.141\,6| \leqslant \frac{1}{2} \times 10^{-4}$$

如果近似值 x^* 的绝对误差限是某一位的半个单位，从该位到 x^* 的第一位非零数字共有 n 位数，我们就说 x^* 有 n 位有效数字.

用严格的数学语言还可以作如下定义：

将数 x 的近似值 x^* 写成

$$x^* = \pm(a_1 \times 10^{-1} + a_2 \times 10^{-2} + \cdots + a_n \times 10^{-n}) \times 10^m \tag{9}$$

若其绝对误差限满足

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

则称近似数 x^* 具有 n **位有效数字**，其中 $a_1, a_2, \cdots, a_n$ 都是 x^* 的有效数字. 这里 n 是正整数，m 是整数，$a_1, a_2, \cdots, a_n$ 都是 0 到 9 中的一个数字，且假定 $a_1 \neq 0$.

根据定义易知 π 的近似值 3.14 具有三位有效数字，而近似值 3.141 6 具有 5 位有效数字. 注意 0.002 3 和 0.002 300 的有效位数是不同的，前者具有两位有效数字，其绝对误差不超过 $\frac{1}{2} \times 10^{-4}$，后者具有四位有效数字，其绝对误差不超过 $\frac{1}{2} \times 10^{-6}$.

以上叙述说明了有效数字与绝对误差之间的关系. 形如(9)的数，若具有 n 位有效数字，由(10)式表明：当 m 一定时，有效位数越多，则绝对误差越小. 对于同一个数 x 的不同的近似值而言，有效位数越多的近似值，其相对误差越小.