普通高等院校“十二五”规划教材

数值计算方法

SHUZHI JISUAN FANGFA

蔡锁章 杨明 雷英杰 编著

内 容 简 介

本书在高等理工科院校的高等数学和线性代数知识的基础上，介绍数值计算方法的基本概念、方法和理论，着重介绍工程计算中的常用算法，包括误差理论、方程的近似解法、线性方程组解法、特征值和特征向量的求法、插值法和曲线拟合、数值微分与数值积分、常微分方程数值解法、偏微分方程数值解法等。各章配有适量习题，并附有习题答案。

本书可作为高等工科院校数值计算方法的教材，也可供工程技术人员自学参考。

图书在版编目(CIP)数据

数值计算方法／蔡锁章，杨明，雷英杰编著. —北京：国防工业出版社，2011.12 重印

ISBN 978-7-118-07747-6

Ⅰ. ①数… Ⅱ. ①蔡… ②杨… ③雷… Ⅲ. ①数值计算－计算方法 Ⅳ. ①O241

中国版本图书馆 CIP 数据核字(2011)第 195712 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 100048)

腾飞印务有限公司印刷

新华书店经售

*

开本 787×1092　1/16　**印张** 16　**字数** 390 千字

2011 年 12 月第 2 次印刷　**印数** 5001—10000 册　**定价** 32.00 元

(本书如有印装错误，我社负责调换)

国防书店：(010)68428422　　发行邮购：(010)68414474

发行传真：(010)68411535　　发行业务：(010)68472764

前　言

随着电子技术的飞速发展和科学研究、生产实践的需要，电子计算机的使用日益广泛。作为电子计算机应用的一个重要方面——科学计算技术也在日新月异地迅速发展。科学计算技术是以电子计算机为工具，以数值计算方法为理论依据的一门技术。掌握数值计算方法的基本知识、熟练运用电子计算机进行科学计算，已成为现代科学技术人员必须具备的基础与技能。本书就是为理工科院校开设数值计算方法的课程而编写的教材。

学习本书需具备高等数学、线性代数和算法语言等方面的知识。

本书将介绍数值计算方法的基本概念、方法和理论，着重介绍科学、工程计算中的常用算法，包括误差理论、方程的近似解法、线性方程组解法、特征值和特征向量求法、插值法和曲线拟合、数值微分和数值积分、常微分方程的数值解法、偏微分方程的数值解法等。

每章习题中都有该章主要算法的编程上机题，完成这些习题有助于真正掌握这些算法。

本书由蔡锁章、雷英杰、杨明等共同编写，由蔡锁章教授担任主编。

限于水平，书中难免有不妥和错误之处，恳请读者批评指正。

编著者

2011 年 7 月

目　录

第1章　误差分析与数值计算

1.1　引　言

利用数学方法去描述、研究、解决生产实践和科学研究中的问题时，从模型的建立、参数的测量、算法设计到最终的数值计算，其各个环节都有一定程度的近似，最终也只能得到原问题的近似解。原问题的真解与这个近似解之间的偏差称为误差。提起近似和误差，往往会给人以不严格、不完美、甚至错误的预感。其实，这是一种误解，科学研究中的近似是正常的，误差是不可避免的客观存在。问题在于能否将误差控制到所允许的范围。本节将讨论误差的来源。

1.1.1　误差的来源

误差按照来源可分为以下4类。

1. 模型误差

用数值计算解决科学技术中的具体问题，首先必须建立这个具体问题的数学模型。由于数学模型总是具体问题的一种简化和近似，因而即使能求出数学模型的准确解，也与实际问题的真解有所偏差，这种偏差称为**模型误差**。

2. 观测误差

数学模型中的很多参数（如时间、长度、电压等）是通过实验测量得来的，受测量工具本身的精密程度、实验手段的局限性、环境变化以及操作者的工作态度和能力等因素的影响，测量的结果往往带有一定程度的误差，这类误差称为**观测误差**。

3. 截断误差

理论上的精确值往往要求用无限次的运算才能获得，但计算机只能进行有限多次的运算。所以人们在构造数值计算方法求解数学模型时，只能通过有限多次的数值运算去近似其准确解。这种由数值方法得到的近似解与数学模型的准确解之间的误差称为**截断误差**。

例如，由泰勒公式，函数$f(x)$可表示为

$$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\cdots+\frac{f^{(n)}(0)}{n!}x^n+\frac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1}$$

式中：$\theta\in(0,1)$。

为了简化计算，当$|x|$接近于0时，去掉上式中的最后一项，得近似公式

$$f(x)\approx f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\cdots+\frac{f^{(n)}(0)}{n!}x^n$$

用此近似公式计算$f(x)$产生的误差就是截断误差。

4. 舍入误差

由于计算机只能对有限位的数字进行存储和运算，每个超出计算机字长的数据都要经过

四舍五入进行取舍或截断方法处理，由此引起的误差称为**舍入误差**。例如，在十位十进制的限制下，会出现

$$1 \div 3 = 0.3333333333$$

$$(1.000002)^2 - 1.000004 = 0$$

两个结果都不准确，后者的准确结果应是 4×10^{-12}，这里所产生的误差就是舍入误差。

1.1.2 误差理论在数值计算中的作用

例 1.1.1 建立 $I_n = \int_0^1 \frac{x^n}{x+5}\mathrm{d}x$ 的递推公式，并求当 $n = 0,1,2,\cdots,20$ 时，I_n 的值。

解 由

$$I_0 = \int_0^1 \frac{\mathrm{d}x}{x+5} = \ln 6 - \ln 5$$

$$I_n + 5I_{n-1} = \int_0^1 \frac{x^n + 5x^{n-1}}{x+5}\mathrm{d}x = \int_0^1 x^{n-1}\mathrm{d}x = \frac{1}{n}, n = 1,2,\cdots \tag{1-1}$$

得

$$\begin{cases} I_0 = \ln 6 - \ln 5 \\ I_n = \dfrac{1}{n} - 5I_{n-1}, n = 1,2,\cdots \end{cases} \tag{1-2}$$

按递推公式(1-2)计算的结果见表 1-1 第二列(箭头表示递推方向)。

表 1-1

n	按公式(1-2)计算得到的 I_n	按公式(1-5)计算得到的 I_n
0	0.1823215568	0.1823215568
1	0.0883922160	0.0883922160
2	0.0580389200	0.0580389198
3	0.0431387333	0.0431387341
4	0.0343063334	0.0343063296
5	0.0284683333	0.0284683522
6	0.0243250004	0.0243249055
7	0.0212321408	0.0212326152
8	0.0188392962	0.0188369242
9	0.0169146301	0.0169264899
10	0.0154268495	0.0153675505
11	0.0137748437	0.0140713383
12	0.0144591150	0.0129766419
13	0.0046275017	0.0120398676
14	0.0482910628	0.0112292335
15	-0.1747886472	0.0105204991
16	0.9364432358	0.0098975045
17	-4.623392649	0.0093360067
18	23.17251880	0.0088755221
19	-115.8099624	0.0082539683
20	579.0998122	0.0087301587

又由 $0<I_n<I_{n-1}$ 得

$$5I_{n-1}<I_n+5I_{n-1}<6I_{n-1} \tag{1-3}$$

将式(1-1)代入式(1-3)得不等式

$$0<\frac{1}{6n}<I_{n-1}<\frac{1}{5n} \tag{1-4}$$

由此可见，I_n 的值随 n 的不断增加而趋近于零。而在表 1-1 中，$I_{15}<0$，从 $n=15$ 开始，I_n 的值正负相间且绝对值不趋于零，而是不断递增。理论分析与计算结果严重不符。有必要寻求新的计算公式。

由式(1-4)知

$$\frac{1}{6\times 21}<I_{20}<\frac{1}{5\times 21}$$

取不等式两边的平均值作为 I_{20} 的近似值

$$I_{20}\approx\left(\frac{1}{6\times 21}+\frac{1}{5\times 21}\right)\times\frac{1}{2}=0.0087301587$$

于是得新的递推公式

$$\begin{cases} I_{20}\approx 0.0087301587 \\ I_{n-1}=-\dfrac{1}{5}I_n+\dfrac{1}{5n} \end{cases},n=20,19,\cdots,1 \tag{1-5}$$

由 I_{20} 为起始值，按式(1-5)中第 2 式逐次递推的计算结果见表 1-1 第三列。

比较两种计算结果，I_0 是一样的。在递推公式(1-2)中，当 n 越大时，I_n 的值越不可靠；而在递推公式(1-5)中，尽管粗略地取 $I_{20}=0.0873015867$，但按递推方向算下去，基本符合 I_n 的特性，最后求得的 I_0 又很准确。这是为什么呢?

在式(1-2)中，理论上 $I_1=-5I_0+1$，其中 $I_0=\ln 6-\ln 5$，但计算机只能存贮有限位小数，所以实际参与计算的是 I_0 的近似值 $\hat{I}_0=0.1823215568$。记 $I_0-\hat{I}_0=\varepsilon$，则

$$I_1-\hat{I}_1=-5(I_0-\hat{I}_0)=-5\varepsilon$$

同理 $I_n-\hat{I}_n=-5(I_{n-1}-\hat{I}_{n-1})=(-5)^2(I_{n-2}-\hat{I}_{n-2})=\cdots=(-5)^n(I_0-\hat{I}_0)=(-5)^n\varepsilon$

尽管 ε 非常小，但误差随传播逐步扩大，$\hat{I}_n$ 与 I_n 的误差为 $(-5)^n\varepsilon$，当 n 较大时，其值就可能很大，因此按递推公式(1-2)进行计算的数值结果很不可靠。

在公式(1-5)中

$$I_0-\hat{I}_0=\frac{1}{-5}(I_1-\hat{I}_1)=\frac{1}{(-5)^2}(I_2-\hat{I}_2)=\cdots=\frac{1}{(-5)^n}(I_n-\hat{I}_n)$$

尽管 $\hat{I}_{20}$ 粗略地取 $\frac{1}{2}\left(\frac{1}{6\times 21}+\frac{1}{5\times 21}\right)$，但因误差随传播逐步缩小，故按递推公式(1-5)计算的数值结果是可靠的。

例 1.1.2 求方程 $\lambda^2+(\alpha+\beta)\lambda+10^9=0$ 的根，其中 $\alpha=-10^9,\beta=-1$。

解 显然

$$\lambda_{1,2}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

其中

$$-b=-(\alpha+\beta)=10^9+1=0.1\times10^{10}+0.0000000001\times10^{10}$$

式中:0.1×10^{10}为10^9的浮点表示;$0.0000000001\times10^{10}$为按$10^{10}$对阶后的1的浮点表示。

若在电子计算机上进行单精度计算时,取数只能取到小数点后第8位,这时$\beta=1$在计算中不起作用,于是有

$$-b\approx-a=10^9$$

类似地,有

$$b^2-4ac\approx b^2$$

$$\sqrt{b^2-4ac}\approx|b|=10^9$$

得

$$\lambda_1=\frac{10^9+10^9}{2}=10^9$$

$$\lambda_2=\frac{10^9-10^9}{2}=0$$

由初等数学可知

$$\lambda^2+(\alpha+\beta)\lambda+10^9=\lambda^2-(10^9+1)\lambda+10^9=(\lambda-1)(\lambda-10^9)$$

因此$\lambda_1=10^9$,$\lambda_2=1$。

为什么这种算法会出错呢?这是因为忽略了一次项系数$(\alpha+\beta)$中的β和整个常数项c,实际是求解了方程

$$\lambda^2+\alpha\lambda=0$$

结果当然是错的。电子计算机在运算过程中,由于加减法运算时要对阶,在小数的阶数向大数的阶数对齐的过程中,大数"吃掉"了小数,α"吃掉"了β,使$b=a$,b^2"吃掉"了$4ac$,使常数项c的作用被忽略,导致计算λ_2时失败。在计算中大数"吃掉"小数,在某种情况下是允许的,如本例中计算λ_1;在某种情况下又不允许,如本例中计算λ_2。

为了避免以上情况,并考虑到在分子部分有可能出现两个相近数相减而导致有效数位严重损失的不利情况,在电子计算机上求

$$a\lambda^2+b\lambda+c=0$$

的根,是按下列步骤进行的(退化情况$a=0$,$b=0$另考虑):

$$\begin{cases}\lambda_1=\dfrac{-b-\operatorname{sign}(b)\sqrt{b^2-4ac}}{2a}\\[2ex]\lambda_2=\dfrac{c}{a\lambda_1}\end{cases}$$

式中:$\operatorname{sign}(b)$为符号函数,其定义为

$$\operatorname{sign}(b)=\begin{cases}1, & b>0\\0, & b=0\\-1, & b<0\end{cases}$$

按照上述算法，重新计算例 1.1.2 中的 λ_2，得

$$\lambda_2 = \frac{c}{\lambda_1} = 1$$

这个值是精确的。

例 1.1.3 给定 $g(x) = 10^7(1 - \cos x)$，试用 4 位数学用表求 $g(2°)$ 的近似值。

解 以下给出两种解法：

(1) 因为 $\cos 2° \approx 0.9994$，所以

$$g(2°) = 10^7(1 - \cos 2°) \approx 10^7(1 - 0.9994) = 6000$$

(2) 因为 $\sin 1° \approx 0.0175$，利用 $\cos 2° = 1 - 2\sin^2 1°$ 计算，有

$$g(2°) = 10^7(1 - \cos 2°) = 2 \times 10^7(\sin 1°) \approx 2 \times 10^7(0.0175)^2 = 6125$$

用同一本数学用表，都计算到小数点后 4 位，为什么答案不一致？这是由于解法(1)中，两个相近数“1”和“0.9994”相减，从而使有效数位减少所致。

例 1.1.1 ~ 例 1.1.3 都是由于误差处理不恰当而造成种种谬误。因此，学习计算方法之前，首先学习误差理论是必不可少的。

1.2 绝对误差与相对误差、有效数字

1.2.1 绝对误差与相对误差

设 x 为原来的数或要测量的真值，x^* 为 x 的近似值或是测得的数值，记

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

称 $E(x^*)$ 为近似数 x^* 的**绝对误差**。

由于 x 的准确值无法得到，因此 $E(x^*)$ 也是无法得到的，如果能估计其绝对值的范围为

$$|E(x^*)| = |x - x^*| \leqslant \Delta$$

则称 Δ 为近似数 x^* 的**绝对误差限**。

例 1.2.1 $\pi = 3.14159265358\cdots$，若取 $\pi^* = 3.14159$ 作为 π 的近似值，于是

$$|E(\pi^*)| \leqslant 0.000003$$

则 $\Delta = 0.000003$ 就可以作为用 3.14159 近似表示 π 的绝对误差限。

例 1.2.2 用毫米刻度的米尺测量不超过 1m 的长度 x，如果长度接近于某毫米刻度 x^*，x^* 就作为 x 的近似值，显然有

$$|E(x^*)| \leqslant \frac{1}{2} \times 1\text{mm} = 0.5\text{mm}$$

则近似值 x^* 的绝对误差限可取为 0.5mm。

当然，我们在估计绝对误差限 Δ 时总希望尽可能定得小些，估计得越精确越好。

有了绝对误差限就可以知道 x(准确值)的范围

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

即 x 落在区间 $[x^* - \Delta, x^* + \Delta]$ 内。在应用上，常采用如下写法来刻画 x^* 的精度：

$$x = x^* \pm \Delta$$

例如 $\pi = 3.14159 \pm 0.000003$。

绝对误差限不能完全表示近似值近似程度的好坏。例如

$$x = 10 \pm 1$$

$$y = 1000 \pm 5$$

虽然 x 的绝对误差限比 y 的小，但显然 1000 作为 y 的近似值要比 10 作为 x 的近似值近似程度好。

记
$$R(x^*) = \frac{E(x^*)}{x^*} = \frac{x - x^*}{x^*}$$

称 $R(x^*)$ 为近似数 x^* 的**相对误差**。

显然，$R(x^*)$ 的准确值也是无法得到的。若

$$|R(x^*)| = \left|\frac{E(x^*)}{x^*}\right| \leqslant \delta$$

则称 δ 为近似数 x^* 的相对误差限。

绝对误差和绝对误差限是有量纲的量，而相对误差和相对误差限是没有量纲的量。

1.2.2 有效数字

设有一数 x，经过四舍五入得其近似值

$$x^* = \pm(x_1 + x_2 \times 10^{-1} + x_3 \times 10^{-2} + \cdots + x_n \times 10^{-n+1}) \times 10^m \tag{1-6}$$

或写成

$$x^* = \pm(x_1 \cdot x_2 x_3 \cdots, x_n) \times 10^m$$

式中：m 为整数；$x_1, x_2, x_3, \cdots, x_n$ 分别是 $0, 1, 2, \cdots, 9$ 中的一个数字，但 $x_1 \neq 0$。由四舍五入的规则知 x^* 的绝对误差满足不等式

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

绝对误差限取

$$\Delta = \frac{1}{2} \times 10^{m-n+1}$$

此时，称 x^* 作为 x 的近似值具有 n 位**有效数字**（或准确数字）。例如

$$\pi = 3.14159265\cdots$$

则近似值 3.14159 的绝对误差限为

$$\Delta = \frac{1}{2} \times 10^{-5}$$

它有 6 位有效数字；近似值 3.1416 的绝对误差限为

$$\Delta = \frac{1}{2} \times 10^{-4}$$

它有 5 位有效数字;近似值 3.14 的绝对误差限为

$$\Delta = \frac{1}{2} \times 10^{-2}$$

它有 3 位有效数字。

通常对写出的具有有限位数字的数,从其左面第 1 个不为零的数字起,到它最右边的一位数字,都认为是有效数字。

如近似数 0.0053、0.123、123.4,依次有 2、3、4 位有效数字,它们的绝对误差限依次取 0.00005、0.0005、0.05,当 0.0053 的绝对误差限为 0.000005 时,就把它记成 0.00530,以示区别,当然其有效数字有 3 位。

下面讨论有效数字与相对误差之间的关系。

定理 1.2.1 若 x^* 具有 n 位有效数字,则其相对误差满足不等式

$$|R(x^*)| \leqslant \frac{1}{2x_1} \times 10^{-(n-1)}$$

其中 x_1 是 x^* 的第 1 位有效数字。

证明 因为 $|x^*| \geqslant x_1 \times 10^m$,所以

$$|R(x^*)| = \left|\frac{E(x^*)}{x^*}\right| \leqslant \frac{\frac{1}{2} \times 10^{m-n+1}}{x_1 \times 10^m} = \frac{1}{2x_1} \times 10^{-(n-1)}$$

由定理 1.2.1 可知,有效数字位越多,相对误差限就越小。

定理 1.2.2 形如式(1-3)的数 x^*,若其相对误差 $R(x^*)$ 满足不等式

$$|R(x^*)| \leqslant \frac{1}{2(x_1+1)} \times 10^{-(n-1)}$$

则 x^* 至少有 n 位有效数字。

证明

$$|E(x^*)| = |R(x^*)| \cdot |x^*|$$

$$|R(x^*)| \leqslant \frac{1}{2(x_1+1)} \times 10^{-(n-1)}$$

$$|x^*| < (x_1+1) \times 10^m$$

因此

$$|E(x^*)| < \frac{1}{2(x_1+1)} \times 10^{-(n-1)} \times (x_1+1) \times 10^m$$

即

$$|E(x^*)| < \frac{1}{2} \times 10^{m-n+1}$$

所以 x^* 至少具有 n 位有效数字。

证毕。

由以上两个定理看出,有效数字的位数可以刻画出近似数的近似程度。

1.3 近似数的简单算术运算

1.3.1 近似数的加法

设有 k 个近似数 $x_i^* > 0 (i = 1, 2, \cdots, k)$,记

$$x^* = \sum_{i=1}^{k} x_i^*$$

的绝对误差为 $E(x^*)$,则

$$E(x^*) = \sum_{i=1}^{k} x_i - \sum_{i=1}^{k} x_i^* = \sum_{i=1}^{k} (x_i - x_i^*) = \sum_{i=1}^{k} E(x_i^*)$$

$$|E(x^*)| \leqslant \sum_{i=1}^{k} |E(x_i^*)|$$

故得以下结论:

(1) 和的绝对误差等于各项绝对误差之和。

(2) 和的绝对误差限不超过各项的绝对误差限之和。

又因为

$$R(x^*) = \frac{E(x^*)}{\sum_{i=1}^{k} x_i^*} = \frac{\sum_{i=1}^{k} E(x_i^*)}{\sum_{i=1}^{k} x_i^*} = \sum_{i=1}^{k} \left[R(x_i^*) \frac{x_i^*}{\sum_{i=1}^{k} x_i^*} \right] \quad (x^* = \sum_{i=1}^{k} x_i^* \neq 0)$$

所以有

$$\min R(x_i^*) \leqslant R(x^*) \leqslant \max R(x_i^*)$$

即有结论:和的相对误差介于各项的相对误差中最大者和最小者之间。

1.3.2 近似数的乘法

记

$$E(x_1^*) = x_1 - x_1^* = \mathrm{d}x_1^*$$

$$E(x_2^*) = x_2 - x_2^* = \mathrm{d}x_2^*$$

$$E(x_1^* x_2^*) = x_1 x_2 - x_1^* x_2^*$$

而

$$\begin{aligned} x_1 x_2 - x_1^* x_2^* &= x_1^* (x_2 - x_2^*) + x_2^* (x_1 - x_1^*) + (x_1 - x_1^*)(x_2 - x_2^*) \\ &= x_1^* \mathrm{d}x_2^* + x_2^* \mathrm{d}x_1^* + \mathrm{d}x_1^* \mathrm{d}x_2^* \end{aligned}$$

略去 $\mathrm{d}x_1^* \mathrm{d}x_2^*$ 这一项,得

$$E(x_1^* x_2^*) \approx x_1^* \mathrm{d}x_2^* + x_2^* \mathrm{d}x_1^* = x_1^* E(x_2^*) + x_2^* E(x_1^*)$$

于是

$$R(x_1^* x_2^*) = \frac{E(x_1^* x_2^*)}{x_1^* x_2^*} \approx \frac{E(x_1^*)}{x_1^*} + \frac{E(x_2^*)}{x_2^*} = R(x_1^*) + R(x_2^*)$$

$$|R(x_1^* x_2^*)| \leqslant |R(x_1^*)| + |R(x_2^*)|$$

即积的相对误差限不超过各因子的相对误差限之和。

1.3.3 近似数的除法

记 $x^* = \frac{x_1^*}{x_2^*}$,则