

高 等 学 校 教 材

计 算 方 法 基 础

王兵团 编著

中 国 铁 道 出 版 社
2 0 0 0 年 · 北 京

(京)新登字 063 号

内 容 简 介

本书着重介绍了科学与工程计算中的基本概念、常用算法及其构造处理方法。书中涉及的各种算法均以实际例题引出,由浅入深,以期读者能顺利进入科学计算领域。书中内容编排科学紧凑,适用面广,尤其适用初次接触科学计算的读者。

全书内容有:非线性方程的求根方法,线性方程组的解法,数据逼近方法,数值积分与微分方法,求矩阵特征值与特征向量的方法及常微分方程初值问题的数值解法等。

本书可作为高等学校非计算数学专业的教材和参考书,也可用于科技人员的自学和参考。

图书在版编目(CIP)数据

计算方法基础/王兵团编著. —北京:中国铁道出版社,2000.9

高等学校教材

ISBN 7-113-03884-0

I. 计… II. 王… III. 计算方法-高等学校-教材 IV. 0241

中国版本图书馆 CIP 数据核字(2000)第 46304 号

书 名:计算方法基础

作 者:王兵团

出版发行:中国铁道出版社(100054,北京市宣武区右安门西街8号)

责任编辑:赵 静

封面设计:陈东山

印 刷:北京市彩桥印刷厂

开 本:787×960 1/16 印张:11 字数:218 千

版 本:2000年9月第1版 2000年9月第1次印刷

印 数:1~2500册

书 号:ISBN 7-113-03884-0/O·83

定 价:15.60元

版权所有 盗印必究

凡购买铁道版的图书,如有缺页、倒页、脱页者,请与本社发行部调换。

前 言

在提倡培养学生创新能力的教学改革大环境下,作为现代科学三大组成部分之一的科学计算正越来越受到重视。科学计算的核心是以计算机为工具、以数学模型为基础的模拟研究,因此,随着社会的进步,从事科学研究的人员必须了解和掌握科学计算的有关知识。

计算方法是科学计算的基础,其中的内容不但可以帮助人们了解科学计算的特点、方法及进行简单的科学计算,而且它还涉及到很多科学研究的方法和现代数学的概念,这对提高科研素质和学习现代数学知识是很有益处的。

本书在选材上主要选择科学计算中常用的有代表性的算法及其概念与处理技术,目的是使读者在学完此书后能对科学计算有基本的了解,并能在思想上建立起科学计算的意识,在所遇到的一些计算问题上少犯错误。书中的每个算法突出构造与分析部分,为有助于理解和应用,所讲的算法都配以相应的例题。根据作者多年讲授计算方法课的科学研究的体会,本书采用了一种新的编写方式,这种方式突出了相关概念和内容的联系与衔接,加入了现代数学知识,层次清晰,强调算法的构造过程与应用,弱化所讲内容的枯燥性,便于自学。

本书涵盖了非计算数学专业计算方法课几乎所有算法,定位于非计算数学专业的本科生和研究生。学习本书要求读者具有高等数学和线性代数方面的一些知识,为便于缺乏这方面知识的读者学习,书中将要用到的高等数学和线性代数的有关概念以附录形式给出。此外,为使读者能尽快使用计算机编程实现算法,书中最后还以附录的形式简单介绍了易学易用的 Matlab 软件。

本书适用面广,内容具有伸缩性。全部讲授需 60 学时,选学部分章节和内容也可满足少学时的教学需要。书中带有“*”号部分主要用于开阔学生知识面,可酌情讲授。

作者衷心感谢北京科技大学刘钦圣教授和北方交通大学陈立成教授、柳金甫教授对本书的仔细审阅及提出的宝贵意见。此外,北方交通大学的张作泉老师、冯国臣老师参与了本书习题部分的编写,刘国忠副教授、金文博士也对本书提出了有益的建议,在此一并向他们表示感谢。

由于作者水平有限,书中不当之处在所难免,敬请同行和读者批评指正。

编者
2000 年 5 月

目 录

第 1 章 绪论	1
1.1 为什么要学习计算方法	1
1.2 计算机中的数系与运算特点	3
1.2.1 计算机的数系	3
1.2.2 计算机对数的接收与处理	3
1.3 误差及其相关概念	4
1.3.1 误差的来源	4
1.3.2 误差的定义	5
1.3.3 有效数字	6
1.3.4 和、差、积、商的误差	6
1.3.5 计算机的舍入误差	7
1.4 计算方法研究的对象、内容及发展	8
1.5 计算方法中常用的一些概念	9
1.6 科学计算中值得注意的地方	11
习题一	12
第 2 章 非线性方程的求根方法	14
2.1 引例	14
2.2 问题的描述与基本概念	15
2.3 二分法	16
2.3.1 构造原理	17
2.3.2 误差估计与分析	17
2.3.3 例题	18
2.4 简单迭代法	19
2.4.1 构造原理	20
2.4.2 迭代法分析	20
2.4.3 简单迭代法的误差估计和收敛速度	22
2.4.4 例题	24
2.4.5 迭代法的加速技术和特点	26
2.5 Newton 迭代法	26
2.5.1 构造原理	27
2.5.2 方法分析	27

2.5.3 例题	29
2.6 Newton 迭代法的变形与推广	30
2.6.1 Newton 迭代法的变形	30
2.6.2 Newton 迭代法的推广	31
2.7* 涉及的现代数学概念——不动点与压缩映射	33
简评	33
习题二	34
第 3 章 线性方程组的解法	36
3.1 引例	36
3.2 问题的描述与基本概念	37
3.3 线性方程组的迭代解法	38
3.3.1 Jacobi 迭代、Seidel 迭代及 Sor 法	38
3.3.2 研究迭代法收敛的现代数学概念	41
3.3.3 迭代法的收敛条件与误差估计	44
3.3.4 例题	47
3.4 线性方程组的直接解法	49
3.4.1 Gauss 消元法	49
3.4.2 LU 分解法	53
3.4.3 特殊线性方程组解法	57
3.4.4 例题	60
3.5 线性方程组解对系数的敏感性	62
简评	64
习题三	64
第 4 章 求矩阵特征值与特征向量的方法	67
4.1 引例	67
4.2 问题的描述与基本概念	68
4.3 幂法与反幂法	68
4.3.1 构造原理	69
4.3.2 分析	70
4.3.3 例题	72
4.4 Jacobi 方法	73
4.4.1 构造原理	73
4.4.2 分析	75
4.4.3 例题	76
4.5 QR 方法	77

4.5.1 构造原理	77
4.5.2 分析	78
4.5.3 例题	80
简评	82
习题四	82
第5章 插值与拟合方法	84
5.1 引例	84
5.2 问题的描述与基本概念	85
5.3 插值法	86
5.3.1 Lagrange 插值	87
5.3.2 Newton 插值	89
5.3.3 Hermite 插值	91
5.3.4 分段多项式插值	93
5.3.5 三次样条插值	96
5.3.6 例题	100
5.4 曲线拟合法	105
5.4.1 构造原理	105
5.4.2 分析	106
5.4.3 可用线性最小二乘拟合求解的几个非线性拟合类型	107
5.4.4 曲线拟合法的推广	108
5.4.5 例题	109
5.5* 涉及的现代数学概念: 内积空间与正交	111
简评	112
习题五	113
第6章 数值积分与数值微分方法	115
6.1 引例	115
6.2 问题的描述与基本概念	115
6.3 插值型求积公式	118
6.3.1 Newton-Cotes 求积公式	119
6.3.2 复合求积公式	122
6.3.3 Gauss 求积公式	124
6.3.4 例题	128
6.4 Romberg 求积方法	130
6.4.1 构造原理	131
6.4.2 分析	131

6.4.3 Romberg 求积方法的计算过程	133
6.4.4 例题	133
6.5 数值微分	134
6.5.1 利用 n 次多项式插值函数求数值导数	134
6.5.2 利用三次样条插值函数求数值导数	135
6.6* Monte-Carlo 方法	136
简评	137
习题六	138
第 7 章 常微分方程初值问题数值解法	140
7.1 引例	140
7.2 问题的描述与基本概念	140
7.2.1 问题的描述	140
7.2.2 建立数值解法的思想与方法	141
7.2.3 数值解法的误差、阶与绝对稳定性	142
7.2.4 Euler 方法的有关问题	144
7.3 Runge-Kutta 方法	146
7.3.1 构造原理	147
7.3.2 构造过程	148
7.3.3 Runge-Kutta 方法的阶与级的关系	149
7.3.4 例题	150
7.4 线性多步法	152
7.4.1 构造原理	152
7.4.2 分析	153
7.4.3 例题	155
7.5 步长的自动选取	156
7.6 一阶微分方程组初值问题的数值解法	157
简评	159
习题七	160
附录	162
附录 A 符号与名词注释	162
附录 B Matlab 使用速成	164
参考文献	168

第1章 绪 论

本章主要介绍科学计算的特点及与计算方法有关的知识和概念,对了解和学习计算方法,以及今后从事科学计算工作都是很有帮助的。

1.1 为什么要学习计算方法

科学计算是用计算机进行数学计算的,很多人,甚至为数不少的一些科研人员,常常认为只要把涉及到的一些数学公式,用一种计算机语言正确编程,计算机就一定能给出正确的结果,问题是这样简单吗?请看下面的例子。

【例 1.1】求一元二次方程

$$x^2 - (10^9 + 4)x + 4 \times 10^9 = 0$$

的根。

解 常用的一元二次方程求根公式为

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

式中 a, b, c 分别为二次项系数、一次项系数和常数项。若利用这个公式编程并在字长为 8 位的计算机上计算,得到的结果为 $x_1 = 10^9, x_2 = 0$, 易验证本题的两个根为 $x_1 = 10^9, x_2 = 4$, 可见计算机给出的计算结果不对。

【例 1.2】计算数列 $I_n = \int_0^1 \frac{x^n}{x+5} dx$

解

因为

$$\begin{aligned} I_n &= \int_0^1 \frac{x^n + 5x^{n-1} - 5x^{n-1}}{x+5} dx \\ &= \int_0^1 x^{n-1} dx - 5 \int_0^1 \frac{x^{n-1}}{x+5} dx \\ &= \frac{1}{n} - 5I_{n-1} \end{aligned}$$

得到计算 I_n 的递推公式

$$I_n = -5I_{n-1} + \frac{1}{n} \quad n=1, 2, \dots \quad (1.1)$$

由 $I_0 = \int_0^1 \frac{1}{x+5} dx = \ln 6 - \ln 5$ 可依次算出 $I_1, I_2, \dots$ 现在把式(1.1)用计算机编程来计算,会发现当 n 较大时, I_n 的计算结果出现负数,例如在字长为 8 的计算机上编程计算,可出现 $I_{12} = -0.329\ 021\ 10 \times 10^{-2}$,这显然是错误的,因为 I_n 的被积函数在积分区间上是非负的,故总应有 $I_n \geq 0$ 才对。

除了上面的两个例子之外,还有很多计算机可以产生错误结果的例子,在此不再列举了。这些例子说明,并不是把科学计算中涉及的数学公式机械地编程计算就一定能得到正确的结果,这其中有很多值得认识和探讨的问题。当然这里并不是否定计算机在科学计算中的作用,事实上,计算机正是由于可以进行科学计算才发展起来的。用计算机做科学计算时绝大部分情况下都能得到所需要的准确的结果,但若在用计算机做科学计算时对可能出现的问题有所警惕并注意采用合适的方法计算,我们就能少犯或不犯错误。怎样使计算机的计算结果可信和尽量减少出错的情况,在科学计算中是非常重要的,因为错误的计算结果会产生错误的结论或否定原来正确的数学模型,这会给科研工作带来很大的损失。

现在很多科学研究和工程问题的解决都是借助计算机进行的,通常用计算机解决实际问题有如下四个步骤:①建立数学模型;②选择数值方法;③编写程序;④上机计算。

建立数学模型是对实际问题进行分析后,根据其内在规律做出简化假设,并运用适当的数学工具来得到一个适定的数学问题,其表现形式可能是一个方程组、一个函数极小化、一个积分计算式、一个微分方程以及它们的不同组合等。建立数学模型需要一定的专业知识。

选择数值方法是为已建立的数学模型选择一个或几个数值计算方法,以用于编程和计算。这里要考虑的问题是所选择的方法能否达到要求的精度、方法的计算量是否太大、程序能否实现以及方法对数据的微小扰动反应是否敏感等。这些问题正是计算方法所研究和讨论的,了解了如何处理这些问题,就可以最大限度的保证计算机的求解少犯错误。

编写程序是根据所选的数值计算方法,用计算机语言写出源程序。科学计算中常用的计算机语言是 Fortran、C 和 Basic 语言等,当然现在有一些数学软件包可以用来做科学计算,如 Matlab 和 Mathematica 等,它们一般不必事先编写复杂的程序,但对较为复杂的大型问题,由于要考虑计算效率和灵活多样性,用计算机语言编写源程序还是不可缺少的。

上机计算是将已知的原始数据输入计算机,计算机按程序指令进行计算,以得到所需的结果。若结果不满意,应检查一下所选数值方法是否合理或编程是否考虑不周,还不行,应检查数学模型的合理性。通常可先用一些其他数据检验程序的质量或对错,以确保上机计算的结果可信。

实际问题的解决有定性和定量之分,其中定量的比定性的更具说服力。例如,某公司投资一个项目能否赚钱是定性问题,而能赚多少钱是定量问题,显然,公司应更关心定量问题。定量的科研结果是要经科学计算得出的。在数学模型正确的前提下,计算的结果取决于相应的数值方法的选择,这里也可看出计算方法的重要性。

1.2 计算机中的数系与运算特点

1.2.1 计算机的数系

计算方法是根据来自实际问题的数学模型在计算机上求解的方法,因此,应该了解计算机是如何进行数字运算的,这有助于构造和分析各种数值方法。数字运算主要是实数运算,而实数集是稠密的无限集。任何一个非零实数可表示为

$$x = \pm 10^c \times 0.a_1a_2a_3\cdots \quad (1.2)$$

式中 $a_i \in \{0, 1, 2, \dots, 9\}$, c 为整数。由式(1.2)表示的 x 称为十进制浮点数,一般地,可定义 β 进制浮点数为

$$x = \pm \beta^c \times 0.a_1a_2a_3\cdots \quad a_i \in \{0, 1, \dots, \beta-1\}$$

但在计算机中,由于机器本身的限制,一般实数被表示为

$$x = \pm \beta^c \times 0.a_1a_2\cdots a_t \quad a_i \in \{0, 1, \dots, \beta-1\} \quad (1.3)$$

式中 t 是正整数,是计算机的字长, c 为整数,满足 $L \leq c \leq U$, L 和 U 为固定整数,对不同的计算机, t , L 和 U 是不同的。一般 β 取为 2, 8, 10 和 16。

由式(1.3)表示的数 x 称为 t 位 β 进制浮点数,其中 c 称为阶码, $0.a_1a_2\cdots a_t$ 称为尾数,这样一些数的全体 $F(\beta, t, L, U) = \{\pm \beta^c \times 0.a_1a_2\cdots a_t \mid a_i \in \{0, 1, \dots, \beta-1\}, L \leq c \leq U\}$ 称为机器数系,它是计算机进行实数运算所使用的数系。显然,这样的数系是有限的离散集,从总体看,其中数的分布不是均匀的,但从局部看,阶数相同的数又等距地分布在实数轴的某一段上。计算机使用的数系中有绝对值最大和最小的非零数 M 和 m ,如数系 $F(10, 4, -99, 99)$ 中, $M = \pm 10^{99} \times 0.9999$, $m = \pm 10^{-99} \times 0.0001$,若一个非零实数的绝对值大于 M ,则计算机产生上溢错误,若其绝对值小于 m ,则计算机产生下溢错误。上溢时,计算机中断程序处理,下溢时,计算机将此数用零表示继续执行程序。无论是上溢,还是下溢,都称为溢出错误。通常,计算机把尾数为 0 且阶数最小的数表示数零。

1.2.2 计算机对数的接收与处理

计算机是怎样表示一般实数呢?

设非零实数 x 是计算机接收的数, 则计算机对其的处理为:

(1) 若 $x \in F(\beta, t, L, U)$, 则原样接收 x ;

(2) 若 $x \notin F(\beta, t, L, U)$ 但 $m \leq |x| \leq M$, 则用 $F(\beta, t, L, U)$ 中最接近 x 的数 $fl(x)$ 表示并记录 x , 以便后面处理。

计算机是怎样做数学运算的呢? 首先计算机只能做加减乘除运算, 两个数在计算机中参与运算按如下方式进行:

(1) 加减法: 先对阶, 后运算, 再舍入。

(2) 乘除法: 先运算, 再舍入。

例如, 某一计算机中的数系 $F(10, 4, -90, 90)$ 中有两个数 $x_1 = 0.2337 \times 10^{-1}$ 和 $x_2 = 0.3364 \times 10^2$ 则有

$$\begin{aligned}
 fl(x_1 + x_2) &= fl(0.2337 \times 10^{-1} + 0.3364 \times 10^2) \\
 &\stackrel{\text{对阶}}{=} fl(0.0002337 \times 10^2 + 0.3364 \times 10^2) \\
 &\stackrel{\text{运算}}{=} fl(0.3366337 \times 10^2) \\
 &\stackrel{\text{舍入}}{=} 0.3366 \times 10^2 \\
 fl(x_1 x_2) &= fl(0.2337 \times 10^{-1} \times 0.3364 \times 10^2) \\
 &\stackrel{\text{运算}}{=} fl(0.7861668 \times 10^0) \\
 &\stackrel{\text{舍入}}{=} 0.7862 \times 10^0 = 0.7862
 \end{aligned}$$

上面计算机接收和运算的特点, 常使得数学上很完美的公式在计算机上编程计算时, 却得不到正确的结果, 但只要我们注意到计算机的这些特点, 就可以用科学的方法解决这个问题。

在计算机的数系 $F(\beta, t, L, U)$ 中, 把尾数的第一位 $a_1 \neq 0$ 的数称为规格化的浮点数。规格化浮点数表示一个实数, 具有表示形式唯一和精度高的特点, 但并不是 $F(\beta, t, L, U)$ 中每个数都可以用规格化浮点数表示, 例如数 $0.00\cdots 1 \times \beta^L$, 就不能表示为规格化浮点数。

1.3 误差及其相关概念

误差是科学计算经常要考虑的问题之一, 很多经计算机运算得出的不正确结果, 多半是由误差引起的。准确值与近似值的差异就是误差, 误差无处不在, 但并不可怕, 只要能对误差进行合适的处理和控制, 就可以有效地解决问题。

1.3.1 误差的来源

误差可以来自很多场合, 但其来源主要有 4 个方面: ①模型误差 (也称描述误差);

②观测误差(也称数据误差);③截断误差(也称方法误差);④舍入误差(也称计算误差)。

模型误差是在建立数学模型时,由于忽略了一些次要因素而产生的误差,它是数学建模阶段要考虑的误差,不是计算方法可以解决的;观测误差是在采集原始数据时,由仪器的精度或其他客观因素产生的误差,它也不是计算方法能解决的问题,截断误差是对参与计算的数学公式做简化处理后所产生的误差,例如要计算函数 e^x 在某点的值,由于 e^x 的展开式

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots$$

有无穷项,而计算机不能做无限运算,因此要计算 e^x ,可用其展开式的前 $n+1$ 项代替无穷项来计算 e^x ,即取近似公式

$$e^x \approx 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$$

来计算 e^x 的值,这样产生的误差就是截断误差。由于科学计算经常要把一些数学函数变成计算机易于处理的形式,这样总会产生截断误差,因此,截断误差是计算方法主要研究的误差,很多好的数值计算方法都是巧妙地处理截断误差得出的。舍入误差是计算机因数系不全由接收和运算数据的舍入处理引起的误差,它也是计算方法关注的内容,本章例 1.1 和例 1.2 的错误就是舍入误差造成的。

1.3.2 误差定义

定义 1 设 x^* 为准确值, x 是 x^* 的一个近似值,称

$$e = x^* - x$$

为近似值 x 的绝对误差,简称误差。

由于准确值 x^* 通常是未知的,故误差 e 一般是不可知的,为此引入误差限来对误差进行估计。

定义 2 称满足

$$|e| = |x^* - x| \leq \epsilon$$

的正数 ϵ 为近似值 x 的误差限。

误差限一般是可以求出的,例如用具有毫米刻度的皮尺去测量某个物件的长度,测得的数据与物件的实际长度不会超过半个毫米,这半个毫米就是该物件长度的误差限。误差限 ϵ 给出了准确值 x^* 的所在范围为 $x - \epsilon \leq x^* \leq x + \epsilon$,显然误差限 ϵ 越小,近似值越精确。绝对误差不能反映近似值 x 的近似程度,例如某个量的准确值 $x_1^* = 1\,000$,其近似值 $x_1 = 999$,另一个量的准确值 $x_2^* = 10$,相应近似值 $x_2 = 9$ 这两个量的绝对误差都是 1,但显然 x_1 的近似程度比 x_2 好。为反映这种近似程度,需引入如下相对误差。

定义 3 称 $e_r = \frac{e}{x^*} = \frac{x^* - x}{x^*}$ 为近似值 x 的相对误差。

相对误差的绝对值越小,近似程度越高,例如,前面两个量 x_1 和 x_2 ,它们相对误差分别为 $e_{1r} = 0.001$ 和 $e_{2r} = 0.1$,因为 $|e_{1r}| < |e_{2r}|$,所以 x_1 比 x_2 逼近程度好。同样,由于准确数 x^* 通常是未知的,导致相对误差不能计算出来,因而产生估计相对误差的相对误差限概念。

定义 4 称满足

$$|e_r| = \left| \frac{x^* - x}{x^*} \right| = \left| \frac{e}{x^*} \right| \leq \epsilon_r$$

的正数 ϵ_r 为 x 的相对误差限。

相对误差限不如绝对误差限容易得到,实用中常借助绝对误差限来求之,即取 $\epsilon_r = \epsilon/|x|$ 。

1.3.3 有效数字

科学计算中常用有效数字来估计和处理误差,有效数字容易计算且与误差有密切关系。

定义 5 设近似数 $x = \pm 10^k \times 0.a_1a_2\cdots a_n$ 或 $x = \pm 10^k \times 0.a_1a_2\cdots$, 这里 $a_1 \neq 0, a_i \in \{0, 1, 2, \cdots, 9\}$, k 为整数,如果有

$$|e| = |x^* - x| \leq 0.5 \times 10^{k-n}$$

则称 x 为 x^* 的具有 n 为有效数字的近似值,简称 x 具有 n 位有效数字。由定义可知,有效数字越多,绝对误差越小,这也是科学计算中要尽可能多保留有效数字的原因。

【例 1.3】 考虑 $\pi = 3.141\,592\,6\cdots$ 的近似值 $x_1 = 3.14$ 和 $x_2 = 3.141$ 的有效数字。

解 $x_1 = 10^1 \times 0.314$, $x_2 = 10^1 \times 0.3141$, $k = 1$

因为 $|\pi - x_1| = 0.001\,592\,6\cdots = 10^{-2} \times 0.159\,26\cdots < 0.5 \times 10^{-2}$

所以有 $k - n = -2, n = 3$, 即 x_1 有 3 位有效数字。

因为 $|\pi - x_2| = 0.000\,592\,6\cdots = 10^{-3} \times 0.592\,6\cdots < 0.5 \times 10^{-2}$

所以可知 x_2 有 3 位有效数字。

如果十进制准确数 x^* 经过四舍五入得到近似数 x , 则 x 的有效数字位为将 x 写为规格化浮点数后尾数的位数,例如 $x^* = 0.003\,45$, 四舍五入得 $x = 0.003\,5 = 0.35 \times 10^{-2}$ 可知 x 有 2 位有效数字。

1.3.4 和、差、积、商的误差

计算机的数值运算主要是加、减、乘、除四则运算,带有误差的数经过四则运算后误

差怎样变化,用微分关系可以简单描述。注意到准确数 x^* 与其近似数通常很接近,其差可认为是较小的增量,可以把差看作微分,由此可得误差与微分的近似关系式

$$e = x^* - x = dx$$

$$e_r = \frac{e}{x^*} \approx \frac{dx}{x} = d \ln x$$

这用文字描述就是 x 的微分 dx 表示 x 的绝对误差, $\ln x$ 的微分表示 x 的相对误差。利用这两个关系式及微分运算可以得到一系列有关四则运算的误差结果,例如:

由 $d(x \pm y) = dx \pm dy$ 可得两数之和(差)的误差等于两数的误差之和(差);

由 $d \ln(xy) = d \ln x + d \ln y$ 可得两数之积的相对误差等于两数相对误差之和等。

一般地,设多元函数 $u = f(x_1, x_2, \dots, x_n)$ 有绝对误差关系:

$$du = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i$$

【例 1.4】 考查 $y = x^n$ 的相对误差关系。

解 因为 $\ln y = n \ln x$

所以 $d \ln y = n d \ln x$

由微分与误差的关系可知函数 x^n 的相对误差为自变量 x 的相对误差的 n 倍,即 n 个数相乘后其结果是自变量的相对误差扩大 n 倍。

用微分关系描述四则运算的误差是粗略的,但所得出的结论基本上与严格用定义推导相同。

1.3.5 计算机的舍入误差

设计算机的数系为 $F(\beta, t, L, U)$, 某数

$$x = \pm \beta^t \times 0.a_1 a_2 \cdots a_i \cdots, a_1 \neq 0 \quad x \in F(\beta, t, L, U)$$

且满足 $m < |x| < M$, m 及 M 是 $F(\beta, t, L, U)$ 中绝对值最小及最大的正数,则计算机经舍入处理后以数 $fl(x)$ 接收,即 $fl(x) = \pm \beta^t \times \tilde{a}$

$$\tilde{a} = \begin{cases} 0.a_1 a_2 \cdots a_i & \text{当 } 0 \leq a_{i+1} < \beta/2 \\ 0.a_1 a_2 \cdots a_i + \beta^{-i} & \text{当 } a_{i+1} \geq \beta/2 \end{cases}$$

因此计算机对 x 的舍入绝对误差满足

$$|e| = |x - fl(x)| \leq 0.5 \times \beta^{-i}$$

舍入相对误差满足

$$|e_r| = \frac{|x - fl(x)|}{|x|} \leq \frac{0.5 \times \beta^{-i}}{0.1 \times \beta^t} = 0.5 \times \beta^{t-i} \quad (1.4)$$

由式(1.4)可知计算机对任何实数的舍入相对误差限与实数本身无关,只与计算机字长 t 有关的一个数 $0.5 \times \beta^{t-i}$ 有关,因此通常定义数 $eps = 0.5 \times \beta^{t-i}$ 为计算机的精度,给定一台计算机后,其舍入误差的精度也就给定了,任何运算要想得到比计算机精度更小的

误差,需要做特别处理才行。由于计算机的精度只与字长有关,计算机字长 t 越大,其精度越高,这也是在数值计算中,为了提高计算结果的精度,有些数值要用双字长处理的原因,双字长数据也称为双精度数。

1.4 计算方法研究的对象、内容及发展

计算方法是计算数学的范畴,有时也称它为计算数学、数值方法、数值分析等,其研究对象是各种数学问题的数值方法设计、分析,有关的数学理论和软件实现,它是一个数学分支。

计算方法的内容可分为两大方面:①连续系统的离散化;②离散型方程的数值求解。

连续系统的离散化是指怎样将连续的问题化为离散的问题来求解,因为计算机的数集是有限的离散集,不能直接处理连续的量。如求定积分 $\int_a^b f(x)dx$ 是连续系统问题,计算方法常用公式

$$\int_a^b f(x)dx \approx \sum_{k=1}^n A_k f(x_k)$$

来将定积分离散化,以求出定积分的近似值。至于离散型方程的数值求解是指将来自实际问题的微分方程等经离散化后,怎样求解对应的离散化方程问题。例如微分方程的定解问题

$$\begin{cases} y' = f(x, y) \\ y|_{x=a} = y_a \end{cases} \quad a \leq x \leq b$$

可离散化为

$$y_{m+1} = y_m + hf(x_m, y_m)$$

其中 $y_m \approx y(x_m)$, $h = (b-a)/n$, $x_m = a + mh$ $m=0, 1, 2, \dots, n$

由 $y_0 = y_a$, 可依次求出 $y_1, y_2, \dots, y_n$, 以求得原问题的近似解。

考查计算方法中的数值求解公式,会看到其中大部分内容都是在这两大方面所做的研究得出的。

计算方法是一门最古老的数学,因为它只涉及到数的四则运算,但它又是一门很年轻的数学,因为它是随计算机的诞生、发展和广泛应用而逐步发展形成的,其理论还涉及到现代数学的理论和内容。

谈到数学,数值和变量及其对问题描述的准确性是基础,而计算方法则是以近似为基础,它只通过来自实际问题的近似数据处理,就能解决实际问题,这使得它与数学建模一起成为当今科学技术中最有用的数学研究领域。一个好的侦探,可以通过点滴的蛛丝马迹找到罪犯,那么计算方法在求解实际问题方面可以与之相提并论。利用计算数学

发展起来的数学学科有很多,如数值逼近、计算几何、计算概率及最优化等。此外,它与其他学科相结合也产生其他一些边缘学科,如计算力学、计算物理、计算生物学及计算经济学等。随着计算机的发展和人们对计算方法认识的深入,计算方法也在不断发展着,如:传统的计算方法是根据计算机串行处理指令的特点来构造的,它可称为**串行计算方法**或**串行算法**,现代的计算方法是面向并行计算机处理指令的特点来构造算法的,它称为**并行算法**。并行算法具有计算速度快和效率高的优点。目前改造已有的有效串行算法使之适用于并行计算机以及直接设计高效率的并行算法,是当今计算数学中最活跃的新领域。

1.5 计算方法中常用的一些概念

(1) 数值问题

由一组已知数据(输入数据),求出一组结果数据(输出数据),使得这两组数据之间满足预先指定的某种关系的问题。称为**数值问题**。

(2) 数值解

经过计算机的计算求出的解或由数值计算公式得出的解称为**数值解**。一般数值解是近似解。

(3) 算法

由给定的已知量,经过有限次四则运算及规定的运算顺序,求出所关心的未知量的数值解,这样所构成的整个计算步骤,称为**算法**。

计算方法的目的是任务,主要是研究和构造算法。

(4) 计算量

一个算法所需要的乘法和除法总次数称为**计算量**,常用 N 表示。计算量的单位为 flop,表示完成一次浮点数乘或除法所需要的时间。算法的计算量可以衡量算法的优劣,因为它体现着算法的计算效率,通常算法的计算量越小,则算法的计算效率越高,因而该算法也就越好。

注:由于计算机做加减法要比乘除法快得多,故在一般情况下算法的计算量可以不考虑加减法的计算时间。

【例 1.5】 设 A, B, C 分别为 $10 \times 20, 20 \times 50, 50 \times 1$ 的矩阵,计算 $D = ABC$ 就有如下不同的算法和计算量:

算法 1 $D = (AB)C$ 计算量 $N = 10\ 500\text{flop}$

算法 2 $D = A(BC)$ 计算量 $N = 1\ 200\text{flop}$

显然算法 2 的计算量比算法 1 小,因而算法 2 比算法 1 要好。

(5) 病态问题

因初始数据的微小变化,导致计算结果的剧烈变化的问题称为病态问题。

病态问题也称坏条件问题,这类问题通常是问题本身固有的,例如,线性方程组

$$\begin{cases} x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 = \frac{11}{6} \\ \frac{1}{2}x_1 + \frac{1}{3}x_2 + \frac{1}{4}x_3 = \frac{13}{12} \\ \frac{1}{3}x_1 + \frac{1}{4}x_2 + \frac{1}{5}x_3 = \frac{47}{60} \end{cases} \quad (1.5)$$

的准确解为 $x_1 = x_2 = x_3 = 1$, 当把它的系数都舍入成两位有效数字产生小的扰动后, 原方程组变为

$$\begin{cases} x_1 + 0.50x_2 + 0.33x_3 = 1.8 \\ 0.50x_1 + 0.33x_2 + 0.25x_3 = 1.1 \\ 0.33x_1 + 0.25x_2 + 0.20x_3 = 0.78 \end{cases}$$

这个方程组的准确解为 $x_1 = -6.222\cdots, x_2 = 38.25, x_3 = -33.65$, 此解与扰动前的解完全不同了。方程组(1.5)的求解问题就是病态问题。

求解病态问题应该特别注意, 因为实际问题的数据都是近似的或经计算机计算要对输入数据做舍入处理, 这都会引起原始数据的扰动, 若所求解的正好是个病态问题, 则采用通常算法计算就会出现很隐蔽的错误, 导致不良的后果。病态问题在函数计算、方程求根及方程组求解中都是存在的, 它的计算或求解应用专门的方法或将其转化为非病态问题来解决。

(6) 良态问题

初始数据的微小变化只引起计算结果的微小变化的计算问题称为良态问题。例如方程组

$$\begin{cases} 2x_1 - x_2 = 6 \\ x_1 + 2x_2 = -2 \end{cases}$$

的准确解为 $x_1 = 2, x_2 = -2$, 对其常数项部分做微小扰动变为

$$\begin{cases} 2x_1 - x_2 = 6 \\ x_1 + 2x_2 = -2.005 \end{cases}$$

这样得到的方程组准确解为 $x_1 = 1.999, x_2 = -2.002$ 这两组解之间的差别是不大的。计算数学中主要研究良态问题的数值解法。

(7) 数值稳定算法

在计算过程中产生的舍入误差能被控制在一定的范围内, 且对最后的结果影响不大的算法称为数值稳定算法。不是数值稳定的算法称为数值不稳定算法。

【例 1.6】 设计算机的数系为 $F(10, 4, L, U)$, 今有一批数据: $x_1 = 0.5055 \times 10^4, x_2 = x_3 = \cdots = x_{11} = 0.4500$, 试求其和 $S = \sum_{i=1}^{11} x_i$ 。