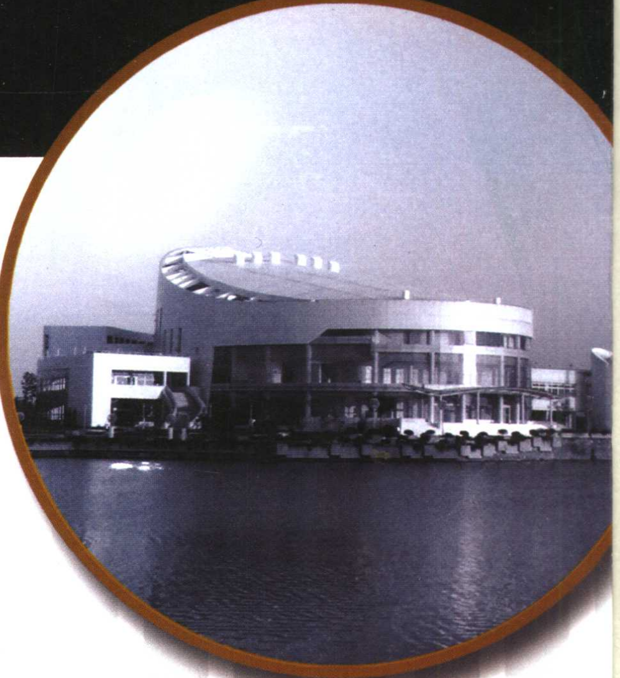
重点大学计算机教材

数值方法

第2版

金一庆　陈越　王冬梅　编著
浙　江　大　学



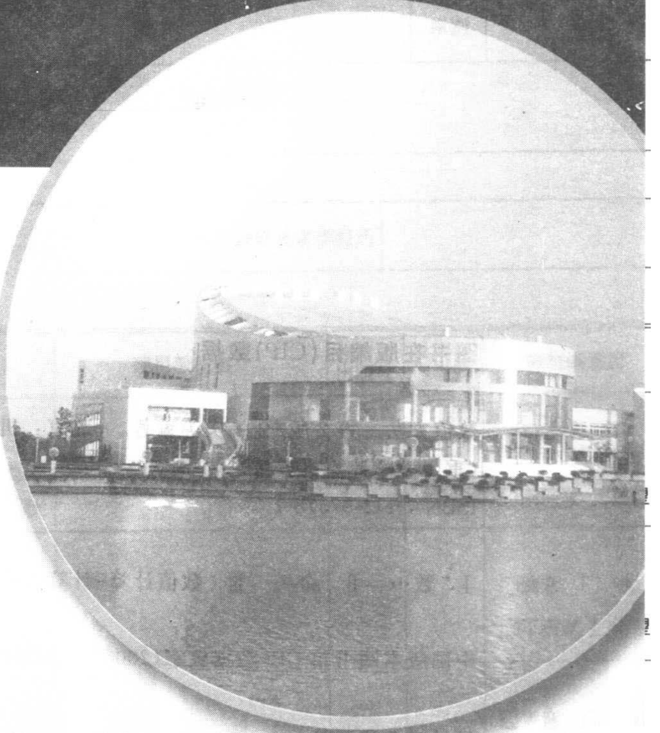
本书为教师配有习题参考答案，
需要的教师可登录华章网站下载。

机械工业出版社
China Machine Press

重点大学计算机教材

数值方法

第2版



金一庆 陈越 王冬梅 编著
浙江大学

机械工业出版社
China Machine Press

本书系统讲解数值方法，作者在第 1 版的基础上进行了较多修改。主要内容包括误差的概念、非线性方程求根方法、线性方程组求解、矩阵的特征值与特征向量的计算、插值、曲线拟合与函数逼近、数值积分方法、常微分方程求解、偏微分方程求解等。书中包含丰富的实例和练习，并且介绍了如何应用 MATLAB 软件完成相关的求解工作。

本书深入浅出，重点突出，适合作为高等院校相关专业的教材，也适合工程技术人员参考。

图书在版编目(CIP)数据

数值方法　第 2 版/金一庆等编著. -北京：机械工业出版社，2000.2
(重点大学计算机教材)
ISBN 7-111-07578-1

Ⅰ. 数…　Ⅱ. 金…　Ⅲ. 数值计算-计算方法　Ⅳ. O241

中国版本图书馆 CIP 数据核字(2000)第 10352 号

机械工业出版社(北京市西城区百万庄大街 22 号　邮政编码　100037)
责任编辑：朱　劼
北京京北制版印刷厂印刷 · 新华书店北京发行所发行
2007 年 1 月第 2 版第 1 次印刷
184mm×260mm · 19.75 印张
定价：32.00 元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换
本社购书热线：(010)68326294

第2版前言

随着计算机技术的迅速发展，数值方法在工程技术领域中的应用越来越广泛，并且已成为数学与计算机之间的桥梁。要解决工程问题，往往需要处理许多数学模型，这要花费大量的人力和时间，同时许多数学模型无法用解析法得到解。使用数值方法并利用计算机就可以解决这些问题。与解析法不同，数值方法得到的解只能是数值解。也就是说，数值方法必须用实际数据进行运算，得出的结果只能是数，而不是某种表达式。而且，由于离散化造成的误差和计算机的有限位运算造成的舍入误差，使数值方法得到的数值解只能是近似解。由于数值方法讨论的问题是如何把实际数学模型转化为可解数学模型，因此这门课归根结底是一门数学课。

我们在编写过程中，力图做到深入浅出地讲解一些重要数值方法的来龙去脉以及公式和算法，通过学习本书，希望读者能掌握这些方法的基本思想和基本技巧，学会对各种方法进行误差分析，并把这些知识融会贯通于编程解题的过程中。本书面向有一定数学基础，并学过一两门高级语言的理工科学生，因此，略去了实际数学模型建立的过程，也删去了用C语言实现的程序。附录A的大部分习题答案省略了过程，希望读者能通过学习，自己完成习题，独立编程解决问题。教授这门课的教师可从华章网站(www.hzbook.com)上下载习题的详解。

本书自2000年出版第1版以来，得到了广大读者的支持与帮助，几经勘误。第2版中增加了许多例题，内容的次序也稍有变动，增加了“解任意线性方程组”的内容和MATLAB的介绍(参见附录B)。线性方程组是数值方法的重要内容，“解任意线性方程组”扩大了方程组可解的范围。MATLAB是科学计算的专用软件，调用内置函数就能解决不少数值方法的问题，也能编程处理一些问题，还有很好的绘图功能。

本书主要由金一庆编写，编写时参照了多年的授课讲义。第9章由陈越编写，并对整本教材进行了补充修改，增加了一些应用实例，将国外优秀的教学思想融入其中。王冬梅参加了第2版的编写工作，编写了附录B。此外，陈顺宝、吕黎明和娄冰也承担了部分工作。由于作者学识有限，本书难免有疏漏之处，敬请广大读者批评指正。

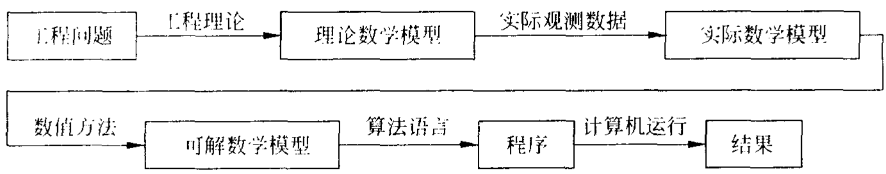
作　者

第 1 版前言

随着电子计算机技术的发展，数值方法在工程技术领域中的应用越来越广泛，且已成为数学与计算机之间的桥梁。

解决工程问题，往往需要处理许多数学模型，这就要花费大量的人力和时间，但是还有不少数学模型无法用解析法得到解(如五次方程就没有求根公式)。使用数值方法并利用计算机就可以克服这些困难。

从工程问题到使用计算机算出结果，要经历的过程如下图所示：



利用计算机来解决数学问题，必须通过算法语言把意图告诉计算机。但算法语言一般只有加、减、乘、除、乘方、开方几种运算，没有微积分运算，更不会有解线性方程组、微分方程等运算，这就需要以数值方法作为中介，用离散化的办法，把机器不认识的数学模型化为机器可以通过算法语言认识的数学模型。这样，许多原来找不到解析解的问题，也可能通过数值方法找到近似解了。

与解析法不同，数值方法得到的解只能是数值解。也就是说，数值方法必须用实际数据进行运算，得出的结果只能是数，而不是某种表达式。而且，由于离散化造成的误差和计算机的有限位运算造成的舍入误差，使数值方法得到的数值解只能是近似解。

由于数值方法讨论的问题是如何把实际数学模型转化为可解数学模型，因此这门课归根结底是一门数学课。本教材采用较少的篇幅，深入浅出地讲述了一些重要数值方法的来龙去脉，希望通过此课程的学习，使读者能掌握这些方法的基本思想和基本技巧，学会方法的误差分析，并把这些知识融会于编程解题的过程中。此教材是为有一定数学基础并学过一两门高级语言的理工科学生写的，因此，略去了实际数学模型建立的过程，也不逐一详细介绍数值方法所对应的算法和框图。

本教材的前 8 章与第 10 章由金一庆依授课讲稿改写，第 9 章由陈越编写，陈越带回了国外的风格，对该教材进行了补充修改，并增加了一些应用实例。为了结合计算机编程，便于自学，附录 A 提供了习题的解答，附录 B 提供了部分 C 语言编写的程序，以供学习时参考。

目　录

第1章 误　差

1.1 误差的来源与分类

从实际工程问题出发，一直到算出问题结果，这其中的每个过程都会产生误差。如果用理论数学模型来描绘物理量，则产生的误差 $|u-u_1|$ 称为模型误差。如果将理论数学模型中的参数用有限位观测数据代入得到实际数学模型，则产生的误差 $|u_1-u_2|$ 称为观测误差或参量误差。

模型中的量和解用如下的字母表示：

u——物理模型客观量。

u_1——理论数学模型的解。

u_2——实际数学模型的解。

u_3——可解数学模型的解。

u_4——有限位运算后得到的结果。

如果用数值方法求近似解，又会出现方法误差(u_2-u_3)。例如，用泰勒展开式求函数 $\sin x$ 的值：

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots$$

从上式看，要求出准确无误的解必须经过无穷步运算来完成，而实际上这是办不到的，只能取式中的有限项，如当 x 趋向 0 时，用 x 近似代替 $\sin x$，这样造成的截断误差约为$\frac{1}{6}x^3$。这时，截断误差就是一种方法误差。

由于计算机只能进行有限位运算，所以在运算过程中会不断地按某种规则进行舍入，这样会产生舍入误差，也称为凑整误差。同样，在函数的运算中，如果自变量有误差，函数也会因此产生误差，这称为误差传播。例如，把观测值看成自变量，运算结果看成函数，那么如果观测值有初始误差，就会造成误差传播。而且，在任何中间步骤产生的误差(如舍入误差)也会引起误差传播。这些误差传播积累在一起，其结果往往是十分可观的。

我们仅研究方法误差、舍入误差以及误差的传播积累，模型误差和观测误差不在本书研究的范围内。以后各章中的数值方法都会讲到方法误差，本章将简要介绍舍入误差及误差的传播与积累。

舍入误差往往会被人们忽视，这必须引起警惕，因为一个正确无误的计算公式，也可能由于舍入误差的传播积累得出极其荒谬的结论。

例 1-1 已知积分 $I_n=\mathrm{e}^{-1}\int_0^1 x^n\mathrm{e}^x\mathrm{d}x$，在 8 位机上求 I_{20}。

解

$$\begin{aligned}I_n&=\mathrm{e}^{-1}\int_0^1 x^n\mathrm{d}\mathrm{e}^x=\mathrm{e}^{-1}x^n\mathrm{e}^x\Big|_0^1-n\mathrm{e}^{-1}\int_0^1\mathrm{e}^x x^{n-1}\mathrm{d}x\\&=1-nI_{n-1}\end{aligned}$$

$I_n=1-nI_{n-1}$是一个精确的递推公式。

$$I_0=\mathrm{e}^{-1}\int_0^1\mathrm{e}^x\mathrm{d}x=\mathrm{e}^{-1}\mathrm{e}^x\Big|_0^1=1-\frac{1}{\mathrm{e}}$$

因为计算机只能接收有限位的数，且 e≐2.718 281 83，所以 I_0 放入机器中成了 I_0^*，这就产

生了初始误差 E_0：

$$0.5\times10^{-9}<|E_0|=|I_0-I_0^*|<0.5\times10^{-8}$$

因为
$$|E_n|=|I_n-I_n^*|=|(1-nI_{n-1})-(1-nI_{n-1}^*)|$$
$$=n|I_{n-1}-I_{n-1}^*|=n(n-1)|I_{n-2}-I_{n-2}^*|$$
$$=\cdots=n!\ |I_0-I_0^*|=n!\ E_0$$
$$E_{20}=|I_{20}-I_{20}^*|=20!\ E_0$$
$$20!\ =\sqrt{20!^2}>(20^{20})^{\frac{1}{2}}=20^{10}$$

所以
$$E_{20}>20^{10}\times0.5\times10^{-9}=2^9\times10=5120$$

如果初始误差很小，但经过某种方法的计算，会使结果的误差相当大，则称此方法(或公式)是不稳定的。显然，以上的递推公式是不稳定的公式。

以下我们粗略地估计例 1-1 中的 I_{20}：

$$I_{20}=e^{-1}\int_0^1 e^x x^{20}\mathrm{d}x$$

因为 $\max\limits_{0\leqslant x\leqslant1} e^x=e$，$\min\limits_{0\leqslant x\leqslant1} e^x=e^0=1$

$$I_{20}<e^0\int_0^1 x^{20}\mathrm{d}x=0.047\ 619\ 04$$

$$I_{20}>e^{-1}\int_0^1 x^{20}\mathrm{d}x=0.017\ 518\ 068$$

取平均值
$$I_{20}=e^{-1}\left(\frac{1+e}{2}\right)\int_0^1 x^{20}\mathrm{d}x=0.032\ 568\ 557$$

所以
$$|E_{20}|<0.015\ 050\ 49$$

这种方法比用计算机递推计算造成的误差还小。

如果用估计得到的 I_{20}，利用 $I_{n-1}=\frac{1}{n}(1-I_n)$公式反向递推来计算 I_{10}：

因为
$$|E_{n-1}|=|I_{n-1}-I_{n-1}^*|=\left|\frac{1}{n}(1-I_n)-\frac{1}{n}(1-I_n^*)\right|=\frac{1}{n}|E_n|$$

$$|E_{n-2}|=\frac{1}{n\cdot n-1}|E_n|$$

$$\cdots$$

所以
$$|E_{10}|=\frac{1}{20\cdot19\cdots\cdot11}|E_{20}|=\frac{10!}{20!}|E_{20}|<0.225\times10^{-13}$$

尽管初始误差较大，但经过递推求得的 I_{10}精度却相当高，远比由 I_0 推出 I_{10}要好得多，反向递推公式是稳定公式。由此可见，方法的好坏对误差传播积累很有关系，不能掉以轻心。■

又如著名的“蝴蝶效应”问题，通俗的说法是“纽约的一只蝴蝶翅膀一拍，风和日丽的北京就刮起台风来了”。实际问题是由大气运动所遵从的一系列物理学定律引发的，并可抽象出一个庞大的复杂方程组。1921 年，气象学家 Richardson 曾试图用这样的方程组作“数值天气预报”，即由初始时刻的风速、压力、温度等数值出发，一步一步地推出以后的天气情况。若用 x_0 表示初始天气变量，则 n 步以后的天气为 $x_{n+1}=f(x_n)(n=0,\ 1,\ 2,\ \cdots)$，其中 f 为非线性函数。在初始时刻，蝴蝶翅膀一拍造成空气的风速或温度在第 1000 位小数上只差一点，而这两个只差一点的初始值推算了 1000 步后，会使初始值中的有效信息全部丧失。这就是计算不稳定性带来的麻烦，可见误差的传播与积累是个不可忽视的问题。

1.2 误差与有效数字

为了比较方法的好坏及衡量一个结果的精确程度，我们给出绝对误差、相对误差和有效数字的概念。

某一个量的准确值，称为真值，记作 x，其近似值记为 x^*。

1. 绝对误差

即误差本身的大小。记作

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

其中，e^* 称为近似数 x^* 的绝对误差。

由于 x 的准确值一般无法知道，所以实际上也就无法求得 e^*，不过可以估计 e^* 的大小范围，若

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

则 ε^* 是 e^* 的绝对值的上界，称为绝对误差限。

e^* 可以是正的，也可以是负的，但误差限 ε^* 只能是正的。真正的绝对误差 e^* 是惟一的，而 ε^* 并不惟一。

由 $|x^* - x| \leqslant \varepsilon^*$ 可得 $x^* - \varepsilon^* \leqslant x \leqslant x^* + \varepsilon^*$，工程上常用 $x = x^* \pm \varepsilon^*$ 表示一个数，这种方法明确表示了近似数及绝对误差限。

例 1-2 测量会议室的长和宽，测量结果为会议室长为 30m，宽为 10m，长的误差不超过 5cm，宽的误差不超过 2cm，则长和宽应写成：

$$y(长) = 30 \pm 0.05, \quad x(宽) = 10 \pm 0.02$$ ■

绝对误差及绝对误差限都是有量纲的，且必须与近似数有统一的量纲。

例 1-2 中的长与宽哪一个精度高呢？看上去 0.02<0.05，似乎宽的精度高，其实不然，为了解释该问题，引进了相对误差的概念。

2. 相对误差

即绝对误差与真值之比。记作

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

由于真值 x 无法知道，常取 $e_r^* = \frac{e^*}{x^*}$ 作为定义，这是因为

$$\frac{e^*}{x} - \frac{e^*}{x^*} = \frac{e^*(x^* - x)}{xx^*} = \frac{e^{*2}}{xx^*} = \frac{e^{*2}}{(x^* - e^*)x^*} = \frac{\left(\frac{e^*}{x^*}\right)^2}{1 - \left(\frac{e^*}{x^*}\right)}$$

当绝对误差相对于真值来说是相当小的时候，这两种定义之差是关于 $\frac{e^*}{x^*}$ 的高阶无穷小，所以用 $e_r^* = \frac{e^*}{x^*}$ 作为定义还是合理的。

同样 $e_r^* = \frac{x^* - x}{x^*}$ 中还含有 x，也无法确切计算，由 $|e^*| \leqslant \varepsilon^*$，可得 $\varepsilon_r^* = \left|\frac{\varepsilon^*}{x^*}\right|$ 为相对误差限。

在例 1-2 中，

$$\varepsilon_r^*(x) = \left|\frac{\varepsilon^*(x)}{x^*}\right| = \frac{0.02}{10} = 0.002$$

$$\varepsilon_r^*(y)=\left|\frac{\varepsilon^*(y)}{y^*}\right|=\frac{0.05}{30}=0.0016<0.002$$

因此，测得的长的精度比测得的宽的精度要高一些。

注意 $\varepsilon_r^*(x)$表示宽的相对误差限，$\varepsilon_r^*(y)$表示长的相对误差限。

通常是用相对误差来衡量一个近似数的精度的，但相对误差本身没有量纲，只是绝对误差与近似数的比值。

3. 四舍五入规则

· 舍入后使绝对误差限不超过其末位数的半个单位。

· 若需舍入的部分刚好是末位的半个单位时，要使末位凑成偶数。

例 1-3 对 0.7135、0.7265、0.732 51 分别取三位小数，结果分别为 0.714、0.726、0.733。■

事实上，在程序设计语言中，不一定按第 2 条规则来做。

4. 有效数字和有效数

如果近似值 x^* 的绝对误差限不超过某一位的半个单位，从该位向左数到 x^* 的第一个非零的数字，共有 n 位，则称这 n 位数字为有效数字，并说 x^* 具有 n 位有效数字。

如果表示一个数的数字全是有效数字，称此数为有效数。

例 1-4 0.2300 表示有四位有效数字，0.023 只有两位有效数字。■

如果整数并非全是有效数字，可用浮点数表示，把 71 600 表示成只有 4 位有效数字，可写为 0.7160×10^5。

5. 数的浮点表示

任一个有效数 x^* 均能表示成一个 0.1～1 之间的数与 $\pm10^m$ 的乘积(m 为任何整数)的形式。

$$x^*=\pm10^m(\alpha_1\times10^{-1}+\alpha_2\times10^{-2}+\cdots+\alpha_n\times10^{-n})\quad(\alpha_1\neq0)$$

其中 α_1 是第一位不是零的数字，x^* 有 n 位有效数字，分别为 α_1，α_2，…，α_n。

6. 有效数字与相对误差的关系

近似数的相对误差与有效数字的多少有着密切的关系。

因为
$$x^*=\pm10^m(\alpha_1\times10^{-1}+\alpha_2\times10^{-2}+\cdots+\alpha_n\times10^{-n})\quad(\alpha_1\neq0)$$

$$\varepsilon^*(x)=0.5\times10^{m-n}$$

所以
$$\alpha_1\times10^{m-1}\leqslant|x^*|<(\alpha_1+1)\times10^{m-1}$$

$$\varepsilon_r^*(x)=\frac{\varepsilon^*(x)}{|x^*|}\leqslant\frac{0.5\times10^{m-n}}{\alpha_1\times10^{m-1}}=\frac{1}{2\alpha_1}\times10^{-n+1}$$

以上结果可归纳成下面的定理。

定理 1-1 若 x^* 具有 n 位有效数字，则其相对误差限为

$$\varepsilon_r^*(x)=\frac{1}{2\alpha_1}\times10^{-(n-1)}$$

其中 α_1 为 x^* 的第一个有效数字。

反之也归纳出定理 1-2。

定理 1-2 若 x^* 的相对误差为 $|e_r^*(x)|\leqslant\frac{1}{2(\alpha_1+1)}\times10^{-(n-1)}$，则 x^* 具有 n 位有效数字。

因为
$$\varepsilon_r^*(x)=\frac{1}{2(\alpha_1+1)}\times10^{-(n-1)}$$

$$\varepsilon^*(x)=|\varepsilon_r^*(x)\cdot x^*|<\frac{1}{2(\alpha_1+1)}\times 10^{-(n-1)}\times(\alpha_1+1)\times 10^{m-1}$$

所以
$$\varepsilon^*(x)<0.5\times 10^{m-n}$$
x^* 至少有 n 位有效数字。

由上所述，有效数字的多少也反映了一个数的精度。例如，0.23×10^1、0.230×10^4 和 2300 从数值上来说几乎一样大，但精度却不同，有效数字多的精度高。0.023 00 与 2300 的绝对误差限不同，但由于有效数字一样多，相对误差一样，因而精度是相同的。

1.3　函数的误差估计

对于函数 $y=f(x)$，$x\in[a, b]$，如果自变量 x 被近似值 x^* 代替，那么函数值 $f(x)$ 被 $f(x^*)$ 代替，将如何估计 $f(x^*)$ 的精度呢？由微分中值定理可知：

$$e^*(f(x))=f(x^*)-f(x)=f'(\xi)e^*(x)\quad(\text{其中 }\xi\text{ 在 }x\text{ 与 }x^*\text{ 之间})$$
$$|e^*(f(x))|=|f'(\xi)|\,|e^*(x)|\approx|f'(x^*)|\,|e^*(x)|$$

$|f'(x^*)|$ 可看成是函数的绝对误差关于自变量的绝对误差的放大数，称为绝对误差条件数。如果条件数较小，函数的绝对误差可以控制在某个范围内，则称函数 $f(x)$ 为好条件的，如果 $|f'(x^*)|$ 在某点 x_0 的值很大，则称函数 $f(x)$ 在 x_0 这点在绝对误差意义下是坏条件的。

因为
$$|e_r^*(f(x))|=\left|\frac{e^*(f(x))}{f(x^*)}\right|=\left|\frac{f(x^*)-f(x)}{f(x^*)}\right|$$
$$=\left|\frac{f(x^*)-f(x)}{x^*-x}\right|\left|\frac{x^*-x}{x^*}\right|\left|\frac{x^*}{f(x^*)}\right|$$
$$\approx|f'(x^*)|\,|e_r^*(x)|\left|\frac{x^*}{f(x^*)}\right|=\left|\frac{x^*f'(x^*)}{f(x^*)}\right|\,|e_r^*(x)|$$

所以称 $\left|\dfrac{x^*f'(x^*)}{f(x^*)}\right|$ 为相对误差条件数。

对于多元函数 $y=f(x_1, x_2, \cdots, x_n)$，有全微分公式

$$\mathrm{d}y=\frac{\partial f}{\partial x_1}\mathrm{d}x_1+\frac{\partial f}{\partial x_2}\mathrm{d}x_2+\cdots+\frac{\partial f}{\partial x_n}\mathrm{d}x_n$$

如果以增量代替微分，即 $\Delta y\approx\mathrm{d}y$，$\Delta x_i\approx\mathrm{d}x_i$，此处 $\Delta y=y^*-y=e^*(y)$，$\Delta x_i=x_i^*-x_i=e^*(x_i)$，可得近似公式：

$$e^*(y)\approx\left.\frac{\partial f}{\partial x_1}\right|_{(x_1^*,x_2^*,\cdots,x_n^*)}e^*(x_1)+\left.\frac{\partial f}{\partial x_2}\right|_{(x_1^*,x_2^*,\cdots,x_n^*)}e^*(x_2)+\cdots+\left.\frac{\partial f}{\partial x_n}\right|_{(x_1^*,x_2^*,\cdots,x_n^*)}e^*(x_n)$$
$$|e^*(y)|\approx|f_{x_1}e^*(x_1)+f_{x_2}e^*(x_2)+\cdots+f_{x_n}e^*(x_n)|$$

由柯西不等式 $|(a, b)|\leqslant|a|\,|b|$（其中 a，b 为向量）得：

$$|e^*(y)|\leqslant\sqrt{f_{x_1}^2+f_{x_2}^2+\cdots+f_{x_n}^2}\cdot\sqrt{e^{*2}(x_1)+e^{*2}(x_2)+\cdots+e^{*2}(x_n)}$$

所以称 $\sqrt{f_{x_1}^2+f_{x_2}^2+\cdots+f_{x_n}^2}$ 为绝对误差条件数，其中 f_{x_i} 表示函数 $f(x)$ 关于 x_i 的在 $(x_1, x_2, \cdots, x_n)$ 点上的偏导数值。

同样，因为

$$|e_r^*(y)|=\left|\frac{e^*(y)}{f^*}\right|=\left|\frac{f_{x_1}e^*(x_1)}{f^*}+\frac{f_{x_2}e^*(x_2)}{f^*}+\cdots+\frac{f_{x_n}e^*(x_n)}{f^*}\right|$$

$$=\left|\frac{x_1^*f_{x_1}}{f^*}\cdot\frac{e^*(x_1)}{x_1^*}+\frac{x_2^*f_{x_2}}{f^*}\cdot\frac{e^*(x_2)}{x_2^*}+\cdots+\frac{x_n^*f_{x_n}}{f^*}\cdot\frac{e^*(x_n)}{x_n^*}\right|$$

$$\leqslant\sqrt{\left(\frac{x_1^*f_{x_1}}{f^*}\right)^2+\left(\frac{x_2^*f_{x_2}}{f^*}\right)^2+\cdots+\left(\frac{x_n^*f_{x_n}}{f^*}\right)^2}$$

$$\cdot\sqrt{e_r^{*2}(x_1)+e_r^{*2}(x_2)+\cdots+e_r^{*2}(x_n)}$$

所以称 $\sqrt{\left(\frac{x_1^*f_{x_1}}{f^*}\right)^2+\left(\frac{x_2^*f_{x_2}}{f^*}\right)^2+\cdots+\left(\frac{x_n^*f_{x_n}}{f^*}\right)^2}$ 为相对误差条件数，其中 f^* 表示函数 $f(x_1^*,x_2^*,\cdots,x_n^*)$。当自变量为一个时，正是 $y=f(x)$时的情况。

例 1-5 要使$\sqrt{20}$的近似值的相对误差限小于 0.1%，要取几位有效数字？若求$\sqrt{B}$，$B\approx20$，B需取几位有效数字，才能保证$\sqrt{B}$近似值的相对误差小于 0.1%？

解 1)
$$\sqrt{20}\doteq4.472,\ \alpha_1=4$$

由定理 1-1 知
$$\varepsilon_r^*=\frac{1}{2\times4}\times10^{-(n-1)}<0.001=10^{-3}$$

n 取 4 能满足上式，应取 4 位有效数字。

2)设 $y=\sqrt{B}$。

方法一：

$$y'=\frac{1}{2\sqrt{B}},\qquad y'\mid_{B=20}=\frac{1}{2\sqrt{20}}$$

$$\text{相对误差条件数}=\frac{B^*y'\mid_{B=20}}{y^*}=\frac{20\frac{1}{2\sqrt{20}}}{\sqrt{20}}=\frac{1}{2}$$

$$\varepsilon_r^*(y)\doteq\frac{1}{2}\varepsilon_r^*(B)$$

$$\varepsilon_r^*(B)=2\varepsilon_r^*(y)=2\times10^{-3}$$

注意到 B 的 α_1 有可能是 1，也可能是 2，应考虑最坏的情况，取 $\alpha_1=1$。要寻找一个 n，使

$$\varepsilon_r^*(B)=\frac{1}{2\times1}\times10^{-(n-1)}<2\times10^{-3}$$

成立。可知当 n 取 4 时能满足上式，所以 B 应取 4 位有效数字。

方法二：可以利用绝对误差条件数来解答。欲使$\sqrt{B}$近似值相对误差限小于 10^{-3}，则应使

$$|e^*(y)|=|y^*e_r^*(y)|\leqslant|y^*||e_r^*(y)|=\sqrt{20}\times10^{-3}$$

且
$$|e^*(y)|\doteq y'\mid_{B=20}|e^*(B)|$$

为了满足题设，希望 $|e^*(B)|\leqslant\frac{1}{y'\mid_{B=20}}|e^*(y)|=2\sqrt{20}\cdot\sqrt{20}\times10^{-3}=40\times10^{-3}$。

可见若 B 取 4 位有效数字($B\doteq20$)，则 $\varepsilon^*(B)=0.5\times10^{-2}<0.4\times10^{-1}$能保证$\sqrt{B}$近似值的相对误差限小于 10^{-3}。 ■

1.4 近似数的四则运算及数值计算中需注意的几个问题

近似数的四则运算可以看成是多元函数的特殊情况。

1. 加、减运算

$$y=x_1+x_2=f(x_1,x_2)$$

$$f_{x_1}=\frac{\partial f}{\partial x_1}=1,\quad f_{x_2}=\frac{\partial f}{\partial x_2}=1$$

$$\mathrm{d}y=\mathrm{d}x_1+\mathrm{d}x_2$$

$$|e^*(y)|=|e^*(x_1)+e^*(x_2)|\leqslant|e^*(x_1)|+|e^*(x_2)|$$

$$\leqslant\varepsilon^*(x_1)+\varepsilon^*(x_2)$$

这个结果表明和差的绝对误差不超过两数的绝对误差限之和，且

$$|e^*(y)|\leqslant\sqrt{1^2+1^2}\sqrt{e^{*2}(x_1)+e^{*2}(x_2)}=\sqrt{2}\sqrt{e^{*2}(x_1)+e^{*2}(x_2)}$$

即绝对误差条件数是$\sqrt{2}$，而相对误差条件数为

$$\sqrt{\left(\frac{x_1^*}{x_1^*+x_2^*}\right)^2+\left(\frac{x_2^*}{x_1^*+x_2^*}\right)^2}=\frac{\sqrt{x_1^{*2}+x_2^{*2}}}{x_1^*+x_2^*}$$

当$x_1^*+x_2^*=0$时，条件数为∞，达到相对误差坏条件，因此，要避免两个绝对值差不多的数相互抵消，否则对运算结果的精度影响很大。

2. 乘、除运算

$$y=x_1\cdot x_2,\quad y=x_1/x_2$$

两边取对数
$$\ln y=\ln x_1\pm\ln x_2$$

两边取微分
$$\mathrm{d}\ln y=\mathrm{d}\ln x_1\pm\mathrm{d}\ln x_2$$

因为
$$\mathrm{d}\ln x=\frac{1}{x}\mathrm{d}x\doteq\frac{e^*(x)}{x}=e_r^*(x)$$

$$e_r^*(y)=e_r^*(x_1)\pm e_r^*(x_2)$$

所以
$$|e_r^*(y)|=|e_r^*(x_1)|+|e_r^*(x_2)|\leqslant\varepsilon_r^*(x_1)+\varepsilon_r^*(x_2)$$

即乘除法运算中，积(或商)的相对误差不超过两数相对误差限之和，且注意到

$$y=\frac{x_1}{x_2}$$

$$\mathrm{d}y=\frac{x_2\mathrm{d}x_1-x_1\mathrm{d}x_2}{x_2^2}=\frac{\mathrm{d}x_1}{x_2}-\frac{x_1}{x_2^2}\mathrm{d}x_2$$

$$|e^*(y)|=\left|\frac{1}{x_2}e^*(x_1)-\frac{x_1}{x_2^2}e^*(x_2)\right|$$

$$\leqslant\frac{1}{|x_2|}|e^*(x_1)|+\left|\frac{x_1}{x_2^2}\right||e^*(x_2)|$$

若x_2绝对值很小，有可能导致绝对误差很大，因此，要避免除数接近于0。

在加减法运算中，还要注意不要把绝对值差异很大的数做加减法，下面是一个例子。

例 1-6　求$x^2+(\alpha+\beta)x+10^9=0$的根，其中$\alpha=-10^9$，$\beta=-1$。

解　用二次方程求根公式$x=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a}$，这里$a=1$，$b=\alpha+\beta=-0.1\times10^{10}-0.000\,000\,000\,1\times10^{10}=-0.1\times10^{10}$(设用八位机运算)，$c=10^9$，可得

$$
\begin{aligned}
b^2-4ac &= 10^{18}-4\times1\times10^{9}\\
&=0.1\times10^{19}-0.000\ 000\ 000\ 4\times10^{19}\\
&=0.1\times10^{19}=10^{18}
\end{aligned}
$$

可推导出
$$x_{1,2}=\frac{10^9\pm10^9}{2}=\begin{cases}10^9\\0\end{cases}$$

而此方程的根应为 $x_1=10^9$，$x_2=1$。计算机所得结果中的一个根结果很好，另一个根结果很糟，原因是进行加减法时要对阶，大数吃掉了小数，而且又出现了两个大小相同的数相减，丧失了大量有效数字，因此 x_2 的精度很差。但若换一种方法，利用韦达定理：

$$
\begin{cases}x_1+x_2=-\dfrac{b}{a}\\[2ex] x_1x_2=\dfrac{c}{a}\end{cases}
\Rightarrow
\begin{cases}x_1=\dfrac{-b-\mathrm{sign}(b)\sqrt{b^2-4ac}}{2a}=10^9\\[2ex] x_2=\dfrac{c}{ax_1}=\dfrac{10^9}{10^9}=1\end{cases}
$$

其中 $\mathrm{sign}(x)=\begin{cases}1, & x>0\\ -1, & x<0\end{cases}$ 称为符号函数。这个结果就很好了。此方程避免了两个差不多大的数相抵消，也大大减少了大数吃小数产生的影响。■

从舍入误差的积累方面考虑，减少运算次数，简化步骤，也会减少误差的积累并提高运算速度。如

$$x^{255}=x\cdot x^2\cdot x^4\cdot x^8\cdot x^{16}\cdot x^{32}\cdot x^{64}\cdot x^{128}$$

把 254 次乘法化为 14 次乘法。

习题

1. 序列$\{y_n\}$满足递推关系 $y_{n+1}=100.01y_n-y_{n-1}$，取 $y_0=1$，$y_1=0.01$ 及 $\bar{y}_0=1+10^{-6}$，$\bar{y}_1=0.01$，试分别计算 y_5，$\bar{y}_5$，从而说明递推公式对于计算是不稳定的。
2. 取 $y_0=28$，按递推公式 $y_n=y_{n-1}-\frac{1}{100}\sqrt{783}$，计算 y_{100}，若取 $\sqrt{783}\approx27.982$(五位有效数字)，试问计算 y_{100}将有多大误差？y_{100}中尚留有几位有效数字？
3. 设函数 $f(x)=\ln(x-\sqrt{x^2-1})$，求 $f(30)$的值。若开平方用六位函数表，则求对数时误差有多大？若改用另一等价公式 $\ln(x-\sqrt{x^2-1})=-\ln(x+\sqrt{x^2-1})$计算，求对数时误差有多大？
4. 下列各数都按有效数字给出，试估计 f 的绝对误差限和相对误差限。

   1) $f=\sin[(3.14)(2.685)]$

   2) $f=(1.56)^{3.414}$
5. 计算 $f=(\sqrt{2}-1)^6$，取$\sqrt{2}\doteq1.4$，利用下列等式计算，哪一个得到的结果最好？为什么？

   1) $\dfrac{1}{(\sqrt{2}+1)^6}$，　2) $(3-2\sqrt{2})^3$，　3) $\dfrac{1}{(3+2\sqrt{2})^3}$，　4) $99-70\sqrt{2}$。
6. 怎样计算下列各式才可减小误差？

   1) $\sqrt{1+x}-\sqrt{x}$

   2) $\ln x_1-\ln x_2$　$(x_1\approx x_2)$

   3) $\dfrac{1-\cos x}{\sin x}$　(x 在 0 附近)

4) $\sin(x+\varepsilon)-\sin x$

5) $\int_N^{N+1}\frac{dx}{1+x^2}$

7. 求方程 $x^2-56x+1=0$ 的两个根，要使它们具有四位有效数字，$\Delta=\sqrt{b^2-4ac}$至少要取几位有效数字？如果利用韦达定理，Δ 又该取多少位有效数字呢？
8. 求近似数 x^* 的幂 x^{*n}的相对误差公式。
9. 设 $y=\lg x$，证明 x^* 的常用对数的绝对误差不超过 x^* 的相对误差的一半，反之，y^* 的反对数的相对误差不超过 y^* 的绝对误差的 2.5 倍。
10. 简化求多项式 $y=a_nx^n+a_{n-1}x^{n-1}+\cdots+a_1x+a_0$ 的值的运算式。

本章小结

归纳以上的内容，在数值计算中需要注意以下几个问题：

1)避免两个相近的数相减(代数和接近于 0)。

2)避免除数绝对值太小。

3)避免绝对值相差很大的两个数进行加减运算。

4)尽可能减少运算步骤。

5)算法或公式要稳定。

第 2 章　非线性方程求根

在人口增长模型中，假设某一地区人口的数量随时间连续增长，其增长率与人口总数成正比，即$\frac{dN(t)}{dt}=\lambda N(t)$，其中 $N(t)$表示该地区在 t 时刻的人口总数，λ 为人口增长率常数。该微分方程的解为 $N(t)=N_0 e^{\lambda t}$，其中 N_0 为该地区的初始人口总数。但上述模型没有考虑到有外部移民迁入的情况，若允许移民以某常速率 v 进入该地区，则微分方程是$\frac{dN(t)}{dt}=\lambda N(t)+v$，其解为 $N(t)=N_0 e^{\lambda t}+\frac{v}{\lambda}(e^{\lambda t}-1)$。

现假设某地区有 100 万初始人口，第一年内有 43.5 万移民迁入，第一年末总计人口达 156.4 万。根据上述数据推算该地区的增长率常数 λ，即要求解方程：

$$156.4=100e^{\lambda}+\frac{43.5}{\lambda}(e^{\lambda}-1)$$

直接通过代数方法求解以上方程显然是不可能的。在这一章里，就将讨论如何计算这一类问题的近似解的方法。

首先复习几个定义和重要的结论。

定义 2-1　如果有一个数 x^*，使 $f(x^*)=0$，则称 x^* 为方程 $f(x)=0$ 的根，或称 x^* 为函数 $f(x)$的零点。

定义 2-2　如果 $f(x)=(x-x^*)^m g(x)$，其中 m 为正整数，$g(x)$的分母不含因子$(x-x^*)$，且 $g(x^*)\neq 0$，则称 x^* 为 $f(x)$的 m 重零点，x^* 为 $f(x)=0$ 的 m 重根。当 $m=1$ 时，称 x^* 为单根。

定义 2-3　若 $f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_n(a_0\neq 0)$，则称 $f(x)=0$ 为 n 次代数方程。当 $n=1$ 时，此方程称为线性方程。

当 $f(x)=0$ 不是代数方程时，就称为超越方程。

非线性方程包括 $n\neq 1$ 的代数方程(高次方程)及超越方程。

下面列出几个重要的结论：

1)在$[a, b]$上连续的函数 $f(x)$，若 $f(a)f(b)<0$，则存在一实数 $x^*\in(a, b)$，满足 $f(x^*)=0$。

2)$f(x)=0$ 在(a, b)上有根，且 $f'(x)$在(a, b)中不变号，且不为 0，则 $f(x)=0$ 在(a, b)上只有惟一根。

3)n 次代数方程在复数域上恰有 n 个根(一个 r 重根算 r 个根)，则实系数代数方程的复根会成对出现。

4)高于四次的代数方程没有求根公式。