KE XUE JI SUAN

科学计算

高复先 编

大连海运学院出版社

科　学　计　算

高复先　编

大连海运学院出版社

内 容 提 要

科学计算在科学研究、工程技术和现代管理科学中都有广泛的应用。科学计算需要将数值分析、算法构造与程序设计结合起来，才能有效地在计算机上实现。本书取材注意到常用算法有关概念和理论推导，数值软件结构原理和主要程序设计技术，旨在培养科学计算能力。本书可作为计算机应用专业与管理信息系统专业大学生和研究生的教材，亦可作为计算技术人员和管理人员的工作参考书。

科 学 计 算

高复先 编

责任编辑 刘泰山
封面设计 谢心阳

大连海运学院出版社出版
大连海运学院出版社发行
大连海运学院印刷厂印装

开本：787×1092 1/16 印张：19 字数：420千字
1987年11月第一版 1987年11月第一次印刷
印数：1—2000 定价：3.20元

统一书号：(7501·4) ISBN7—5632—0016—9/G.4

前　言

科学计算是现代数字计算机诞生的直接动因，又是计算机应用的重要领域。

科学计算的基本理论与方法，在形成计算机科学、计算机应用，包括现代管理中的计算机应用人才的知识结构方面的重要性，愈来愈被更多的人所重视；如何培养科学计算能力，已成为计算机教育中的一个重要课题。中国计算机学会教育学组早在一九八一年夏就在大连召开了计算机数学教育专题学术研讨会。会上就计算方法课程教学大纲和教学方法等问题进行了讨论，有两种基本意见：一种意见认为计算机及应用专业的计算方法课程即是数值分析，编程上机留给实习课；另一种意见认为必须把数值分析与程序实现结合起来，在一门课程内实行“结合”教学，培养学生科学计算能力，而分开教学让学生自己去“结合”可能达不到能力培养的目标。讨论结果，会议决定起草两种计算方法教学大纲，一种称为“数值分析型”，一种称为“科学计算型”。本书作者荣幸受托起草科学计算型的计算方法教学大纲。两种大纲都在次年的全国计算机教育学术会议上作为推荐大纲提出，供全国大专院校计算机（及应用）专业参考。本书全面体现了这份大纲，在几年教学中按不同对象对内容进行取舍，例如，对计算机七七、七八届大学生和非计算机专业的工科研究生，我们讲了全书八章所有内容，共120学时，而对计算机八四届和管理信息系统专业八六届大学生只讲前七章的主要部分内容，共80学时。本书也适用于社会上办计算机应用学习班，当然是选取部分内容。从毕业生和研究生参加的有关科研工作情况看，所需的科学计算能力通过本书的学习基本上已达到要求。

在计算机数值软件（包）越来越多的今天，不少人似乎有这样的误解，以为买来软件就可以简单地使用，而无需掌握算法本身方面的推导与组织。其实，使用数值软件（包）决非如此简单，缺乏数值分析基本概念，缺乏离散化算法构造的基本理论与方法，恰恰是没有用好已有的数值软件（包）的主要障碍。因此，不论是工程技术界，还是管理应用领域，继续推广普及科学计算方法基础知识，培养使用计算机进行数值计算的能力，仍是目前我国推广计算机应用的一大课题。

本书在多次教学和修订过程中，得到校内外许多同志的关心和帮助，在此表示深深谢意。书中错误不当之处，请读者批评指正。

编　者

1987年10月于大连海运学院

目　　录

绪　论……（1）
科学计算过程……（1）
近似数与误差……（1）
1．误差的来源（2）　2．误差的大小与近似数的精度（3）
3．近似数的四则运算（5）　4．数字计算机的字长与固有误差（6）
习题（8）
第一章　插值与逼近……（9）
引　言……（9）
1·1　Lagrange 插值……（10）
1．线性插值（10）　2．抛物线插值（12）　3．Lagrange 插值（13）
4．插值多项式的余项—误差估计（14）　习题1·1（16）
1·2　Newton 插值……（17）
1．差商概念（17）　2．差商的性质（18）　3．Newton插值公式（19）
习题1·2（21）
1·3　等距节点插值……（22）
1．差分概念（22）　2．Newton前插公式（24）　3．Newton后插公式（26）　习题1·3（27）
1·4　样条插值……（27）
1．三次样条插值（27）　2．公式推导（28）　3．计算步骤与程序设计要点（30）　习题1·4（32）
1·5　最小二乘曲线拟合……（33）
1．回归直线（33）　2．回归多项式曲线（34）　3．经验曲线（36）
4．最小二乘曲线拟合的一般理论（38）　习题1·5（41）
1·6　正交多项式与均方逼近……（41）
1．正交多项式概念（41）　2．均方逼近（45）　习题1·6（47）
1·7　通用程序设计……（48）
1．一元三点分段插值通用程序（48）　2．一元两点分段插值通用程序（50）
3．一元 n 点全区间插值（50）　4．回归直线与经验曲线（51）
第二章　数值积分……（52）
引　言……（52）
2·1　Newton—Cotes 求积公式……（53）
1．Newton—Cotes求积的一般公式（53）　2．常用求积公式（54）
3．复合求积公式（56）　习题2·1（57）
2·2　变步长求积算法……（58）

1．变步长梯形公式（58） 2．变步长Simpson公式（59）
3．Romberg算法（60） 习题2·2（63）
2·3 Gauss求积公式……（63）
1．区间[-1，1]上两点Gauss公式（63） 2．Gauss求积的一般理论（64） 3．区间[a，b]上的Gauss求积算法（67） 习题2·3（69）
2·4 二重积分及广义积分的数值方法……（69）
1．矩形区域上的二重积分（69） 2．奇异积分的计算（70）
3．无穷积分的计算（72） 习题2·4（74）
2·5 通用程序设计……（75）
1．定步长Simpson求积（75） 2．变步长求积——Romberg算法（75）
3．自适步长Gauss求积（76） 4．变限多重积分的Gauss求积（79）
第三章 线性代数方程组的数值解法……（84）
引 言……（84）
3·1 消去法……（85）
1．Gauss消去法（85） 2．Gauss列主元消去法（88） 3．Gauss全主元消去法（90） 4．Gauss—Jordan消去法（91） 习题3·1（92）
3·2 矩阵的三角分解……（92）
1．Gauss消去法与系数矩阵的三角分解（93） 2．LU分解的条件（95）
3．LU分解的算法（96） 4．LDV分解（97） 习题3·2（99）
3·3 基于矩阵分解的特殊方程组解法……（99）
1．实对称方程组的Cholesky方法（99） 2．实对称正定方程组的平方根法（101） 3．三对角方程组的追赶法（102） 习题3·3（103）
3·4 消去法的误差分析……（103）
1．向量范数（103） 2．方阵范数（105） 3．消去法的误差分析——摄动理论（107） 4．实用的办法（111） 习题3·4（112）
3·5 线性迭代法……（112）
1．Jacobi迭代法（112） 2．Seidel迭代法（114） 3．迭代法的矩阵表示（115） 4．松弛法（116） 习题3·5（117）
3·6 迭代法的收敛性……（118）
1．基本收敛定理（118） 2．系数矩阵的可约性与对角占优（120）
3．迭代法收敛性判定法（121） 习题3·6（121）
3·7 非线性迭代法……（122）
1．线性方程组求解的等价问题（122） 2．最速下降法（123）
习题3·7（125）
3·8 通用程序设计……（125）
1．Gauss—Jordan列主元消去法（125） 2矩阵求逆与行列式求值（127）
3．实对称方程组的Cholesky方法（128） 4．大型稀疏方程组的解法（128）
第四章 非线性方程和方程组的数值解法……（132）

引　言……（132）
4·1　根的分离……（132）
1．简单的方法——画图与试探（132）　2．代数方程根模的上下界（133）
3．实系数代数方程实根的分离方法（136）　习题4·1（137）
4·2　实根的加细方法……（138）
1．平分法（138）　2．Newton迭代法（139）　3．线性插值法（141）
习题4·2（141）
4·3　求高次代数方程全部根的劈因子法……（142）
1．Bairstow方法的推导（142）　2．程序设计要点与框图（144）
习题4·3（145）
4·4　非线性方程组的数值解法……（145）
1．Newton迭代法（145）　2．最速下降法（148）　3．Newton迭代法与最速下降法的联合使用（151）　习题4·4（152）
4·5　通用程序设计……（152）
1．用平分法求区间[a、b]内全部单重实根（152）　2．求代数方程全部根（152）　3．求解非线性方程组（153）
第五章　代数特征值问题的数值方法……（158）
引言与予备知识……（158）
5·1　实对称矩阵的Jacobi方法……（159）
1．旋转矩阵（159）　2．Jacobi方法（162）　3．Jacobi方法的计算步骤与程序设计要点（164）　习题5·1（165）
5·2　化实对称矩阵为相似的三对角对称矩阵……（165）
1．反射变换（165）　2．用反射变换化对称阵为相似的三对角对称阵（167）
3．算法的简化（170）　习题5·2（171）
5·3　三对角对称矩阵的特征值……（172）
1．序列$\{p_i(\lambda)\}$的Sturm性质（173）　2．$p_n(\lambda)$根的隔离（175）
3．特征值计算（176）　习题5·3（177）
5·4　实对称矩阵特征值问题的Householder方法……（177）
1．三对角对称阵特征向量的求法（178）　2．对称阵A的特征向量算法（180）
3．Householder方法的推广（183）　习题5·4（184）
5·5　乘幂法与压缩法……（185）
1．乘幂法（185）　2．乘幂法的困难（187）　3．加速收敛（189）
4．压缩法（190）　习题5·5（191）
5·6　通用程序设计——Householder方法……（191）
第六章　常微分方程数值解法……（196）
引　言……（196）
6·1　离散化方法……（197）
1．数值微分法（197）　2．数值积分法（198）　3．Taylor展开法（199）
4．几个基本概念（200）　习题6·1（201）

6・2 Euler 方法的改进……（201）
1．改进Euler法——梯形法则（202） 2．予测校正法（202）
3．一次校正法（204） 习题6・2（205）
6・3 Runge—Kutta 法……（205）
1．一次予测校正算法的精度（205） 2．二阶Runge—Kutta法（206）
3．三阶Runge—Kutta法（207） 4．四阶Runge—Kutta公式（208）
习题6・3（209）
6・4 线性多步法……（209）
1．Adams外推法（210） 2．Adams内插法（212） 3．Adams
外推法与内插法的联合使用（213） 习题6・4（214）
6・5 收敛性与稳定性问题……（214）
1．单步法的收敛性问题（215） 2．标准四阶Runge—Kutta 方法的
收敛性（217） 3．初值问题的稳定性（218） 4．实际保证计
算精度的方法（220） 习题6・5（221）
6・6 常微分方程组初值问题的数值解法……（221）
1．常微分方程组初值问题来源及一般形式（221）
2．数值解法——标准四阶Runge—Kutta法（224）习题6・6（226）
6・7 常微分方程边值问题的差分方法……（226）
1．差分方程的建立（227） 2．极值原理（227） 3．收敛性与误
差估计（228） 4．解差分方程组的追赶法（229）
习题6・7（231）
6・8 通用程序设计……（232）
1．定步长Runge—Kutta方法（232）
2．定步长Adams予测校正法（234） 习题6・8（237）
第七章 偏微分方程的数值解法……（238）
引 言……（238）
7・1 椭圆型方程的差分方法……（238）
1．正方网格的差分格式（239） 2．极值原理和差分方程的可解
性（242） 3．差分方程的收敛性与误差估计（243）习题7・1（246）
7・2 椭圆型差分方程的解法……（246）
1．典型问题的差分方程（246） 2．边界条件的处理（251）
习题7・2（254）
7・3 抛物型方程的差分解法……（255）
1．显式差分格式Ⅰ（256） 2．隐式差分格式Ⅱ（257）
3．六点对称差分格式Ⅲ（258） 习题7・3（259）
7・4 抛物型差分方程的稳定性与收敛性……（259）
1．显式格式的稳定性（260） 2．显式格式的收敛性（261）
3．隐式格式的稳定性与收敛性（263） 习题7・4（264）
7・5 双曲型方程的差分解法……（264）

1．微分方程的差分近似（265） 2．初始条件与边界条件的处理（265）
3．差分格式的收敛性（267） 4．差分格式的稳定性（267）
习题7·5（268）
7·6 抛物型方程差分解法的通用程序设计……（268）
1．显式格式Ⅰ的通用程序（269） 2．隐式格式Ⅱ的通用程序（269）
3．六点格式Ⅲ的通用程序（269）
第八章 解微分方程的有限元方法……（272）
引 言……（272）
8·1 解常微分方程的有限元方法……（272）
1．等价问题（272） 2．单元分析（275） 3．总体合成与最终求解（279） 习题8·1（281）
8·2 解椭圆型方程的有限元方法……（282）
1．平面区域的三角剖析（282） 2．三角单元上的线性插值函数（283）
3．单元分析（284） 4．总体合成与最终求解（286）
习题8·2（288）
8·3 通用程序设计——解Laplace方程的有限元方法……（289）
1．功能（289） 2．算法（289） 3．Algol过程（290）
4．实例（292）

绪　　论

科学计算过程

现代科学技术提出了越来越多的直接或间接的计算问题。这些计算问题主要是靠数字电子计算机来解决的，成为计算机应用的重要的、基本的方面。随着计算机应用领域的扩大，以及计算机在存储容量、计算速度和计算精度等方面的发展，许多经典的计算方法经受了严格的检验，新的数值方法层出不穷。学习与研究科学计算的方法，不仅是计算数学专业的事，也是计算机应用专业、信息系统和管理科学专业所关心的事。

科学计算过程，是从数学模型的提出到上机计算得出结果的完整过程。图1表明了其中的主要步骤及其相互联系。把实际问题抽象为数学问题，即数学化，是解决问题的基础；而对数值分析来说，则是讨论的出发点。对数学模型作离散化处理，是需要一系列分析和代数等方面的知识的，同时又有数值分析本身的特殊方法与技巧；只有把各种数学模型变为有限次的四则运算过程，即离散化，才能排程序，上机执行。当然程序设计也是有技巧的，但作为解决科学计算问题的程序设计，除了掌握有关程序语言方面的知识外，更重要的是对算法的研究和了解。

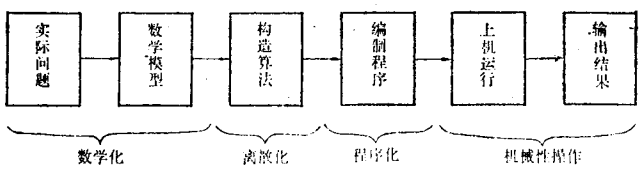

图1

科学计算中的基本问题是：插值与逼近，数值积分，线性代数方程组的数值解法，非线性方程（组）的数值解法，代数特征值问题的数值方法，常微分方程与偏微分方程的数值解法等。这些基本问题的各种算法，一般来说不是从计算经验中归纳出来的，而是从一系列理论分析中构造出来的。本书的中心内容，是介绍用数字电子计算机解决科学计算问题所使用的基本的数学理论与方法，培养科学计算能力。因此，我们既注意必要的理论分析，又注重研究计算步骤，强调理论与实践的统一。我们将以基本的，常用的算法为重点，分析其理论背景，研究方法的建立，并与程序设计相结合，从而为科学计算打下基础。

近似数与误差

科学计算中所处理的一些数据，不论是原始数据，中间结果，还是最终结果，大多数都是近似数。不过，这些近似数的精确度是人们能够而且必须掌握的。这就要研究近似数的误

差来源，估计误差大小的方法，把误差控制在一定范围内的措施。这些就是误差分析的任务，这里我们作一些初步的介绍。

1、误差的来源

观测误差：人们利用各种仪器所观测到的数据，都不可能达到绝对准确的程度，都是带有误差的，这种误差叫做观测误差，例如，用毫米刻度尺来量一根轴的长度，读出数据为98mm，实际上这是误差不超过半毫米的近似数。图2左边表示不足近似数，右边表示过剩近似数，这两种可能性都是客观存在的。

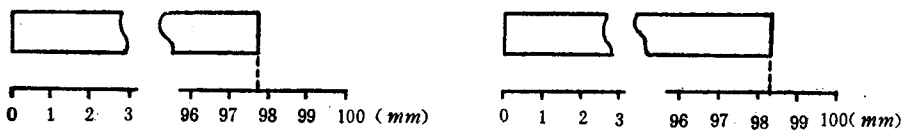

图2

截断误差：将连续函数作离散化处理的方法之一，是展成收敛的无穷级数；而实际计算时，总是用到前面有限几项，抛掉后面的无穷个项，这种误差叫做截断误差。例如

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

当x很小时，常用x代替$\sin x$，这样，截断误差接近于$\frac{x^3}{6}$。最常使用的级数展开法是Taylor级数，有

定理（Taylor定理）设$f(x)$，$f'(x)$，$f''(x)$，……$f^{(n)}(x)$在区间$[a, b]$上存在且连续，$f^{(n+1)}(x)$在区间(a, b)上存在，记Taylor多项式为

$$P_n(x) = f(x_0) + \frac{(x-x_0)}{1!}f'(x_0) + \frac{(x-x_0)^2}{2!}f''(x_0)$$

$$+ \cdots\cdots + \frac{(x-x_0)^n}{n!}f^{(n)}(x_0), \qquad (0\cdot 1)$$

其中$x_0 \in [a, b]$，则对任何$x \in [a, b]$，在x_0与x之间存在ξ_x（依赖于x），使得

$$R_n(x) = f(x) - P_n(x) = \frac{(x-x_0)^{n+1}}{(n+1)!}f^{(n+1)}(\xi_x) \qquad (0\cdot 2)$$

如果令$x = x_0 + h$，即有

$$f(x) = f(x_0) + \frac{h}{1!}f'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots\cdots$$

$$+ \frac{h^n}{n!}f^{(n)}(x_0) + \frac{h^{n+1}}{(n+1)!}f^{(n+1)}(x_0+\theta h) \qquad (0\cdot 3)$$

其中$0 < \theta < 1$。

这个定理在数学分析中已证明过，其中$R_n(x)$就是所谓余项或截断误差项，用它容易得出上面例子中的估计。

舍入误差：在用有限小数来近似无限小数时，或者用较少位数的小数来代替较多位数的

小数时，对某一小数位以后的数字要作舍入处理，才得到一个近似数，这样产生的误差叫舍入误差。例如，

$\pi = 3.1415926\cdots\cdots$

用去尾法得到四位近似数是3.141（亏）

用**收尾法**得到四位近似数是3.142（盈）

用四舍五入法得到四位近似数是3.142

2、误差的大小与近似数的精度

一般来说，用近似数误差的大小来刻划近似数的精确程度。最常使用的概念有

绝对误差：设 x^* 为准确数，x 为近似数，则称

$$E = x^* - x$$

为近似数 x 的误差；而

$$|E| = |x^* - x| = |x - x^*|$$

称为近似数 x 的绝对误差。

一般情况下，求近似数的绝对误差是不现实的，因为并不知道准确数是多少。但常常可以得到估计式

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

则称正数 $\varepsilon(x)$ 为近似数 x 的**绝对误差限**。绝对误差限这一概念，实际上反映了准确数 x^* 的范围

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

或

$$x^* = x \pm \varepsilon(x)$$

显然，提高近似数 x 的精度，就在于尽量减小 $\varepsilon(x)$ 的值。为简单计，人们常常把绝对误差限就说成是绝对误差。

例 1 用去尾法和收尾法得到的 π 的四位近似数 3.141和3.142的绝对误差限，都可取为 0.001，而用四舍五入法得到的四位近似数 3.142 的绝对误差限，可取为 0.0005。一般，用去尾法和收尾法得到的近似数，绝对误差限为近似数末位的一个单位；用四舍五入法得到的近似数，绝对误差限为近似数末位的半个单位。

设 x 是对 x^* 作四舍五入得到的 k 位小数，则有

$$|x^* - x| \leqslant \frac{1}{2} \times 10^{-k} \qquad (0 \cdot 4)$$

反之，具有这样的绝对误差限的近似数 x，就说它准确到第 k 位小数（值得注意的是，这是近似数，而不是准确数）。

相对误差：近似数的误差与准确数之比

$$\frac{x^* - x}{x^*}$$

称为近似数的相对误差，通常写成百分数的形式，所以又称**百分误差**。实际上使用的，是近似数的绝对误差限与近似数的绝对值之比，

$$\frac{|x^*-x|}{|x|}\leqslant\frac{\varepsilon(x)}{|x|}=\delta(x) \qquad (0\cdot 5)$$

称为近似数的**相对误差限**。

有效数字：四舍五入得到的近似数，从左边第一个非零数字起，直到最后一个数字，都叫做有效数字。一般，若近似数 x 写成十幂和的形式

$$x=\pm 10^m(\alpha_1+\alpha_2\times 10^{-1}+\alpha_3\times 10^{-2}+\cdots\cdots$$
$$+\alpha_n\times 10^{-(n-1)}),\ \text{其中}\alpha_1\neq 0$$

其绝对误差限

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

而相对误差限

$$\frac{\varepsilon(x)}{|x|}\leqslant\frac{\frac{1}{2}10^{m-n+1}}{10^m}=\frac{1}{2}\times 10^{-n+1}=\delta(x)$$

则称近似数**具有 n 位有效数字**，α_1，$\alpha_2\cdots\cdots\alpha_n$ 分别叫做 第一位，第二位，……，第 n 位有效数字。

例 2　用四舍五入法得到 π 的近似值

$3.14=10^0(3+1\times 10^{-1}+4\times 10^{-2})$

绝对误差限为 $0.005=\frac{1}{2}\times 10^{-2}$

$m-n+1=0-n+1=-2$

$n=3$

有 3 位有效数字

让我们通过下面的例子来考察绝对误差、相对误差和有效数字三者之间的关系。

例 3　四舍五入得到

的近似数	绝对误差	相对误差	有效数字
120	$\frac{1}{2}\times 10^0$	0.42%	3 位
12.0	$\frac{1}{2}\times 10^{-1}$	0.42%	3 位
1.20	$\frac{1}{2}\times 10^{-2}$	0.42%	3 位
0.120	$\frac{1}{2}\times 10^{-3}$	0.42%	3 位
0.0120	$\frac{1}{2}\times 10^{-4}$	0.42%	3 位

可见近似数左边第一个非 0 数字左边的 0，并不影响相对误差和有效数字，仅影响绝对误差；但是右边的 0 不仅影响绝对误差，也要影响相对误差和有效数字。一般来说，**绝对误差与小数位数有关；相对误差与有效数字位数有关**。绝对误差、相对误差和有效数字都能刻划近似数的精度。

3、近似数的四则运算

利用微分学可简单地得出下述规则：

（i）两个近似数之和（差）的绝对误差等于两个近似数的绝对误差之和，即

$$\varepsilon(x+y)=\varepsilon(x)+\varepsilon(y)$$

$$\varepsilon(x-y)=\varepsilon(x)+\varepsilon(y)$$

（ii）两个近似数积（商）的相对误差等于两个近似数的相对误差之和，即

$$\delta(x\cdot y)=\delta(x)+\delta(y)$$

$$\delta(x/y)=\delta(x)+\delta(y)$$

例4 已知近似数285.35，186.87，58.43，4.96都准确到末位数字（即四舍五入得到的，或者说绝对误差都是$\frac{1}{2}\times10^{-2}$），求这些近似数之和，并估计结果的精度。

解：

```
  285.35
  186.87
   58.43
+) 4.96
-------
  535.61
    ↑
```

按通常准确数加法求出结果为535.61，第二位小数含有误差，因为由规则（i）可知和的绝对误差为0.02。若舍入成535.6，结果绝对误差的保守估计是

$$0.02+0.01=0.03$$

所以，这是四位有效数字（或准确到十分位）的近似数。

例5 求近似数3.150950，15.426463，568.3758，7684.3876之和，其中前三个准确到最末一位数字，最后一个准确到第三位小数（该位上可能有半个单位的误差）。

解： 因为精度最低的近似数准确到第三位小数，所以其他近似数都舍入成4位小数；最后再把结果舍入到两位小数。

```
   3.1510
  15.4265
 568.3758
+) 7684.3876
          ↑
-----------
8271.3409
    ↑
```

结果取8271.34，其绝对误差的保守估计是0.005，具有6位有效数字。值得注意的是，如果不对各个加数作必要的舍入处理，就要作多余的无用的工作；那样得出的结果8271.340813的后四个数字，根据上面的误差分析，是毫无意义的。

根据规则（ii）推导出来的近似数乘法，用有效数字来考虑其精度是十分方便的：两个有效数位不同的近似数相乘，结果有效数位比较低者少一位；须将有效数位较高者舍入成比较低者多一位的有效数字，乘得的结果再舍入成比乘数有效位低者少一位。

例6 求有效数2.01612与3.124之积。

解： $2.01612\approx2.0161$

$$2.0161\times3.124=6.2982964\approx6.30$$

两个近似数相除，办法同上。

例7 计算$31.7/\sqrt{3}$其中31.7是三位有效数。

解： 由于被除数具有3位有效数字，除数取到4位有效数字即可；商有两位有效数字。

$$31.7\div\sqrt{3}=31.7\div1.732050807\cdots$$

$$\approx31.7\div1.732$$

$$=18.30254041\cdots\approx18$$

事实上，掌握了近似数四则运算的规则，不论是初始数据，还是中间数据，都应随时应用有关规则，而不去作无益的工作。这一点对于混合运算尤为重要，一般来说，每个中间步骤都分别按上述规则取**最大可能的精度**。

注意：以上关于近似数计算结果的精度，都是按最保守的情况估计的。比如100个四舍五入的近似数相加，每个近似数的绝对误差是 $\frac{1}{2}\times 10^0$，按规则（i）结果的绝对误差是 $100\times\frac{1}{2}\times 10^0=50$，就是说结果至少准确到百位数。实际上这100个数有的舍，有的入，舍入相消，使得结果的绝对误差远小于50。运用概率统计方法可证明，有99.9%的把握说，结果的绝对误差不超过9.53，即可靠到十位数。

4、数字计算机的字长与固有误差

数字计算机的存储单元的二进制位都是有限的，这个位数称为**字长**。目前常见是16位或32位。因此，计算机表示的数的精度受到限制，存在着固有误差。由于计算机通过软件或硬件实现二进制数与十进制数的转换，所以我们仅就十进制数来讨论计算机的固有误差。设某计算机十进制字长是t位，即能表示 t 个十进制数位。

定点运算：加减法运算，对于机器表数范围之内的数不产生误差，而在超出机器表数范围，机器出现“溢出”而停机。为克服溢出停机，有相应的程序设计技术，如改变算法，或引进比例因子。乘除法运算，两个 t 位数相乘，得到 $2t$ 位的乘积，即双倍字长的数，如小数点在最前面，舍入成单字长的数，这就产生至多是 $\frac{1}{2}\times 10^{-t}$ 的绝对误差，即积的有效数位最多可达 t 位。如小数点在最后，积是 $2t$ 位整数，不能舍入，可用两个存储单元来存放，因而乘积称为双字长数。除法有类似情况，商的位数超过单个字长，若小数点在最左边，结果有 $\frac{1}{2}\times 10^{-t}$ 的舍入误差，两倍字长的整数商也得用两个存储单元存放。

浮点运算：将每个非零的数都表为**浮点规格化形式**

$$a\times 10^p$$

其中

$$0.1\leqslant|a|<1,$$

p 是整指数

这样，在计算机存储单元中可用一对数

$$(\pm a,\ p)$$

表示一个浮点数。如 $27.3=(0.273,2)$，$-27.3=(-0.273,2)$

若 a 最多 t 位，$-N\leqslant p\leqslant N$，则浮点数的最大正数是

$$\underbrace{0.99\cdots\cdots 9}_{t\text{位}}\times 10^N<10^N$$

最小正数是

$$\underbrace{0.10\cdots\cdots 0}_{t\text{位}}\times 10^{-N}=10^{-(N+1)}$$

因此浮点数范围是

$$10^{-(N+1)}\leqslant|x|<10^N, \qquad (0\cdot 6)$$

目前许多计算机 $t\approx 10$，$N\approx 77$，这是相当大的范围了，一般的程序很少发生溢出。

设标准浮点数 $x^* = b \times 10^p$，p 在允许范围内。如果舍入成 t 位近似数 $x = a \times 10^p$，即

$$|b-a| \leqslant \frac{1}{2} \times 10^{-t}$$

这里 $0.1 \leqslant |a| < 1$，$0.1 \leqslant |b| < 1$，则绝对误差限为

$$|x^* - x| = |b-a| \times 10^p \leqslant \frac{1}{2} \times 10^{-t} \times 10^p$$

$$= \frac{1}{2} \times 10^{-t+p} = \varepsilon$$

因为

$$|x^*| = |b| \times 10^p \geqslant 0.1 \times 10^p = 10^{p-1}$$

所以

$$\left| \frac{x^* - x}{x^*} \right| \leqslant \frac{\frac{1}{2} \times 10^{-t+p}}{10^{p-1}}$$

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

$$= \delta \qquad (0 \cdot 7)$$

这就是机器固有误差表达式。绝对误差限与浮点数的阶数 p 有关，而相对误差限只与机器字长有关；换言之，机器的浮点数有效数字位数是一定的，但绝对误差可因 p 的不同而不同。

浮点数四则运算的误差是由浮点近似数的误差造成的。

记正浮点数 x 的浮点近似数为 $fl(x)$，则（0.7）式变为

$$\left| \frac{x - fl(x)}{x} \right| \leqslant \delta = \frac{1}{2} \times 10^{-t+1} \qquad (0 \cdot 8)$$

于是

$$\begin{aligned} &\text{(i)} \quad fl(x) = x(1 \pm \delta) \\ &\text{(ii)} \quad fl(x \pm y) = (x \pm y)(1 \pm \delta) \\ &\text{(iii)} \quad fl(xy) = (xy)(1 \pm \delta) \\ &\text{(iv)} \quad fl(x/y) = (x/y)(1 \pm \delta) \end{aligned} \qquad (0 \cdot 9)$$

其中 $\delta = \frac{1}{2} \times 10^{-t+1}$ 是机器固有的相对误差限

对于（i），由

$$-\delta \leqslant \frac{x - fl(x)}{x} \leqslant \delta$$

得出

$$(1-\delta)x \leqslant fl(x) \leqslant (1+\delta)x$$

亦即

$$fl(x) = x(1 \pm \delta)$$

其意义是：对于具有 t 位（十进制）字长的机器来说，浮点规格化数 x 舍入到 t 位浮点近似数 $fl(x)$，其范围在 $x(1-\delta)$ 到 $x(1+\delta)$ 之间。由此立即得出（ii）～（iv），是浮点数四则运算的误差估计式。

使用数字计算机进行计算，具体问题中数据的误差（如本节第1、2段），计算过程中的误差（如本节第3段）以及机器字长的固有误差（如本节第4段）是交织在一起的，这些都应作具体分析，才能对计算结果的精度有所把握。

习　　题

1、举例说明误差的来源及其各种表示方法。

2、要使计算结果具有5位有效数字，相对误差限应规定多少？一般的提法如何？

3、某微机的BASIC系统的浮点数$t=6$，$N=38$，若π舍入成3.14159，试用(0.9 (i))式求δ，$1+\delta$，$1-\delta$，并解释它们的意义。

4、利用Taylor展开公式 (0.3) 计算

① $\sin(0.5)$，使截断误差不超过10^{-5}。

② $\cos(0.5)$，计算至第三项，估计截断误差。

5、如果用Taylor展开公式(0.3)计算编制正弦函数表，自变量区间是从0°到45°，步长为2′，具有5位小数的函数值，那么展开式取几项才能保证精度？