21世纪高等院校计算机教材系列

数值计算方法

第2版

●马东升　雷勇军　编

购书可获得增值回报
提供教学用电子教案

机械工业出版社
CHINA MACHINE PRESS

21世纪高等院校计算机教材系列

数值计算方法

第2版

马东升　雷勇军　编

机械工业出版社

本书介绍了计算机上常用的数值计算方法，阐明了数值计算方法的基本理论和实现，讨论了一些数值计算方法的收敛性和稳定性，以及数值计算方法在计算机上实现时的一些问题。内容包括数值计算引论，非线性方程的数值解法，线性代数方程组的数值解法，插值法，曲线拟合的最小二乘法，数值积分和数值微分，常微分方程初值问题的数值解法。各章内容有一定的独立性，可根据需要进行取舍。对各种数值计算方法都配有典型的例题，每章后有较丰富的习题，书末有部分习题参考答案。

本书可作为高等院校工科各专业本科生学习数值分析或计算方法的教材或参考书，也可供从事科学与工程计算的科技人员参考。

图书在版编目（CIP）数据

数值计算方法/马东升，雷勇军编．—2版．—北京：机械工业出版社，2006.9（2008.1重印）

（21世纪高等院校计算机教材系列）

ISBN 978-7-111-08968-1

Ⅰ.数…　Ⅱ.①马…②雷…　Ⅲ.数值计算－计算方法－高等学校－教材　Ⅳ.O241

中国版本图书馆CIP数据核字（2006）第034804号

机械工业出版社（北京市百万庄大街22号　邮政编码100037）
策　　划：胡毓坚　责任编辑：陈振虹　版式设计：张世琴
责任校对：张　媛　责任印制：杨　曦
北京机工印刷厂印刷（北京双新装订有限公司装订）
2008年1月第2版第3次印刷
184mm×260mm · 17.5印张 · 429千字
8 001—12 000册
标准书号：ISBN 978-7-111-08968-1
定价：25.00元

出版说明

计算机技术是一门迅速发展的现代科学技术，它在经济建设与社会发展中，发挥着非常重要的作用。近年来，我国高等院校十分注重人才的培养，大力提倡素质教育、优化知识结构，提倡大学生必须掌握计算机应用技术。为了满足教育的需求，机械工业出版社组织了这套“21 世纪高等院校计算机教材系列”。

在本套系列教材的组织编写过程中，我社聘请了各高等院校相关课程的主讲老师进行了充分的调研和细致的研讨，并针对非计算机专业的课程特点，根据自身的教学经验，总结出知识点、重点和难点，一并纳入到教材中。

本套系列教材定位准确，注重理论教学和实践教学相结合，逻辑性强，层次分明，叙述准确而精炼，图文并茂，习题丰富，非常适合各类高等院校、高等职业学校及相关院校的教学，也可作为各类培训班和自学用书。

参加编写本系列教材的院校包括：清华大学、西安交通大学、上海交通大学、北京交通大学、北京邮电大学、北京化工大学、北京科技大学、山东大学、首都经贸大学、河北大学等。

机械工业出版社

第 2 版前言

本书自 2001 年出版以来，先后重印了 5 次，根据这几年的使用情况，我们对部分内容进行了修订。这次修订在保持原有框架基本不变的前提下，增加了反插法、样条插值，删去了非线性方程组的数值方法，充实了迭代原理、矩阵三角分解的紧凑格式、埃米特插值和分段插值，精简了高斯消去法的计算机实现，充实了最小二乘法并单列为一章。此外，还增加了相当数量、不同难易程度的习题，对部分章节作了文字修饰加工。

感谢北京理工大学李小平教授、国防科技大学唐国金教授对本书修订工作的支持和帮助。感谢几年来使用本书的老师和读者，没有他们的支持和帮助，本书第 2 版的出版是不可能的。

本书附有电子教案，可登录 http \\ : www.cmpbook.com 免费下载。

作者学识有限，谬误之处，敬祈批评指正。

作　者

第1版前言

随着计算机技术与计算数学的发展，在计算机上用数值计算方法进行科学与工程计算已成为与理论分析、科学实验同样重要的科学研究方法。利用计算机去计算各种数学模型的数值计算方法，已成为科学技术人员的必备知识。

本书介绍了与现代科学计算有关的数值计算方法，阐明了数值算法的基本理论和方法，讨论了有关数值算法的收敛性和稳定性，以及这些数值算法在计算机上实现时的一些问题。内容包括数值计算的误差分析，非线性方程的数值解法，线性代数方程组的数值解法，插值和拟合，数值积分和数值微分，常微分方程初值问题的数值解法等六章。各章内容具有一定的相对独立性，可根据需要进行取舍。同时对各种算法都配有适当的例题和习题，并附有部分习题答案。本书叙述力求清晰准确，条理分明，概念和方法的引进深入浅出，通俗易懂。阅读本书需具备高等数学和线性代数的基本知识。

北京理工大学在20世纪80年代初将计算方法课定为某些工科专业必修课，本书是在多年教学实践及科学研究成果的基础上，参考当前数值分析和计算方法教材编写而成的。书末列出了部分参考书目，作者谨向参考过的列出和未列出书目的编著者致以衷心的谢意。

感谢胡佑德教授对本书编写给予的热情关心和鼓励。

限于作者水平，书中缺点和错误之处，敬请批评指正。

编　者

目 录

第1章　数值计算引论

本章介绍数值计算方法的研究对象及其主要特点和数值计算中的误差分析。误差分析包括误差的来源、近似数的误差表示、运算误差分析、数值稳定性和减小运算误差。

1.1　数值计算方法

随着计算机的发展与普及，继理论分析、科学实验之后，在计算机上用数值方法进行科学计算已成为科学研究的另一种重要手段。求解各种数学问题的数值计算方法不仅在自然科学得到广泛的应用，而且还渗透到包括生命科学、经济科学和社会科学的许多领域。数值计算方法是应用数学的一个分支，又称数值分析或计算方法，它是研究用数字计算机求解各种数学问题的数值方法及其理论的一门学科，是程序设计和对数值结果进行分析的依据和基础。本书介绍的是在微积分、线性代数和常微分方程等基础数学中最常用的、行之有效的数值方法，内容包括非线性方程的数值解法、数值代数、代数插值、最小二乘法、常微分方程的数值解法等。

应用计算机解决科学计算问题需要经过以下几个主要过程：提出实际问题，建立数学模型，选用或构造数值计算方法，程序设计和上机计算得出数值结果。因此，选用或构造数值计算方法是应用计算机进行科学计算全过程的一个重要环节。

数值计算方法是以数学问题为研究对象，但它不是研究数学本身的理论，而是着重研究求解的数值方法及其相关理论，包括误差分析、收敛性和稳定性等内容。它应具有以下特点：

1）把每个求解的数学问题用计算机所能直接处理的四则运算的有限形式的公式表达出来。

2）每个数值方法要保证收敛性，即数值解能逼近精确解到要求的程度，还要保证数值稳定性。

3）数值方法有良好的计算复杂性，即运算次数要少，所需存储量要小。

将数学模型问题变成数值问题，进而研究求解数值问题的数值方法，并设计行之有效的数值算法，这些内容属于计算方法的范围。

数学问题可以通过离散化、逼近（包括插值、数值微积分等）转化成数值问题。数值问题是指输入数据（即问题中的自变量和原始数据）与输出数据之间函数关系的一个确定无歧义的描述。

在计算机上可执行的求解数值问题的系列计算公式称为数值方法。“计算机上可执行的系列计算公式”是指这一系列计算公式中的运算只有四则运算和逻辑运算等计算机上能够执行的运算。

用计算机上可执行的系列计算公式求解数值问题，具有完整而准确步骤的方法称为数值算法。因此，数值方法是数值算法的核心。

对一个数学问题能有许多不同的算法，而用计算机求各种数学问题的数值算法即是数值计算方法研究的内容。

在数值计算方法中，对许多问题常采用的处理方法有构造性方法、离散化方法、递推化方法、迭代方法、近似替代方法、化整为零方法和外推法等。本书将详细讨论这些方法。

由于数值计算方法研究的对象以及解决问题方法的广泛适用性，现在流行的数学工具软件，如 Maple，Matlab，Mathematica 等，已将其绝大多数内容设计成简单函数，经简单调用，便可得到运行结果。但由于实际问题具体特性的复杂性以及算法自身的适用范围决定了应用中必须选择、设计适合于自己所要解决的特定问题的算法，因而掌握数值计算方法的思想和内容是必须的。

手算是熟悉数值计算公式，掌握数值计算方法和计算过程的重要一环。尽管手算的例题都很简单，但是其计算过程和步骤与计算机按程序计算的过程和步骤相一致，因此，应该充分重视这一环节。数值计算方法的目的是用计算机解决科学研究和工程实际中的数值计算问题，因此，熟练地在计算机上实现这些数值方法是必备的基本技能。同时，通过上机实际计算，可以对各种数值方法有进一步深入的理解。因此，对手算和上机计算都应给予充分重视。

1.2 误差的来源

在数值计算中，要大量地用数进行运算。这些数可以分成两类，一类是精确地反映实际情况的数，这类数称为精确数、准确数或真值，如教室里有 42 名学生，42 就是准确数。另一类数则不是这样，它们只能近似地反映实际情况，这类数称为近似数或某准确数的近似值。例如，通过测量得到桌子的长度为 956 mm，一般说来，这个测量值 956 是不能精确反映桌子实际长度的近似值。将一个数的准确值与其近似值之差称为误差。近似数是有误差的数。误差在数值计算中是不可避免的，也就是说，在数值算法中，绝大多数情况下是不存在绝对的严格和精确的，在考虑数值算法时应该能够分析误差产生的原因，并能将误差限制在许可的范围之内。

误差的来源，即产生误差的原因是多方面的，可以根据误差产生的原因对误差进行分类，下面介绍工程上最常遇到的四类误差。

定量分析客观事物时，要抓住其主要的本质的方面，忽略其次要因素，建立已知量和未知量之间的数学关系式，即数学模型。因此，这样得到的数学模型只是客观现象的一种近似描述，而这种数学描述上的近似必然会产生误差。建立的数学模型和实际事物的差距称为模型误差或描述误差。

例如，物体在重力作用下自由下落，其下落距离 s 与时间 t 满足自由落体方程

$$s = \frac{1}{2}gt^2$$

其中，g 为重力加速度。该方程就是自由落体的数学模型，它忽略了空气阻力这个因素，因而由此求出的在某一时刻 t 的下落距离 s，必然是近似的、有误差的。

在建立的各种计算公式中，通常会包括一些参数，而这些参数又往往是通过观测或实验得到的，它们与真值之间有一定的差异，这就给计算带来了一定的误差。这种误差称为观测

误差或测量误差。

自由落体方程中的重力加速度 g 和时间 t 就是观测来的。观测值的精度依赖于测量仪器的精密程度和操作仪器的人的素质等。

在数值计算方法中不研究模型误差和观测误差，总是认为数学模型是正确合理地反映了客观实际，只是对求解数学模型时产生的误差进行分析研究，求解数学模型时常遇到的误差是截断误差和舍入误差。

许多数学运算是通过极限过程来定义的，而计算机只能完成有限次的算术运算与逻辑运算。因此，实际应用时，需将解题方案加工成算术运算与逻辑运算的有限序列，即表现为无穷过程的截断，这样无穷过程用有限过程近似引起的误差，即模型的准确解与用数值算法求得的准确解之差称为截断误差或方法误差。例如，数学模型是无穷级数

$$\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0)$$

在实际计算时，只能取前面的有限项(如 n 项)

$$\sum_{k=0}^{n-1} \frac{1}{k!} f^{(k)}(x_0)$$

来代替，这样就舍弃了无穷级数的后半段，因而出现了误差，这种误差就是一种截断误差。对这个数学模型的截断误差是

$$\sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(x_0) - \sum_{k=0}^{n-1} \frac{1}{k!} f^{(k)}(x_0) = \sum_{k=n}^{\infty} \frac{1}{k!} f^{(k)}(x_0)$$

当计算机执行算法时，由于受计算机字长的限制，参加运算的数据总是只能具有有限位的数据，原始数据在机器中的表示可能会产生误差，每一次运算又可能产生新的误差，这种误差称为舍入误差或计算误差。

例如，$\pi = 3.141\,592\,6\cdots$，$\sqrt{2} = 1.414\,213\,56\cdots$，$\frac{1}{3} = 0.333\,3\cdots$，等等，在计算机上表示这些数时只能用有限位小数，如取小数点后四位有效数字，则 $3.141\,6 - \pi = 0.000\,007\,3\cdots$，$1.414\,2 - \sqrt{2} = -0.000\,013\cdots$，$0.333\,3 - \frac{1}{3} = -0.000\,033\cdots$ 就是舍入误差。又比如，计算机作 4 位数乘 4 位数乘法运算，若乘积也只许保留 4 位，通过把第 5 位数字进行"四舍五入"，这时产生的误差就是舍入误差。

1.3 近似数的误差表示

近似数的误差常用绝对误差、相对误差和有效数字表示。下面介绍这三种表示方法及其相互之间的关系。

1.3.1 绝对误差

定义 1-1 设 x^* 是准确值 x 的一个近似值，则

$$e(x^*) = x - x^* \tag{1-1}$$

称为近似值 x^* 的绝对误差，简称误差。在不致混淆时，$e(x^*)$ 简记为 e^*。

从定义 1-1 可以看出，e^* 可正可负，当 $e^* > 0$ 时，x^* 称为 x 的弱(不足)近似值；当 $e^* < 0$

时，x^* 称为 x 的强（过剩）近似值。$|e^*|$ 的大小标志着 x^* 的精度。一般地，在同一量的不同近似值中，$|e^*|$ 越小，x^* 的精度越高。e^* 是有量纲的。

一般情况下，无法准确地知道绝对误差 e^* 的大小，但根据具体测量或计算的情况，可以事先估计出误差的绝对值不超过某个正数 ε^*，这个正数 ε^* 叫做误差绝对值的上界或误差限。

定义 1-2 如果

$$|e^*| = |x - x^*| \leqslant \varepsilon(x^*) \tag{1-2}$$

则称 $\varepsilon(x^*)$ 为 x^* 近似 x 的绝对误差限，简称误差限（界），用它反映近似数的精度。在不致混淆时 $\varepsilon(x^*)$ 简记为 ε^*。

从定义 1-2 可以看出，ε^* 是一个正数。又因为在任何情况下都有

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

即

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

这就表明准确值 x 在 $[x^* - \varepsilon^*, x^* + \varepsilon^*]$ 区间内，用 $x = x^* \pm \varepsilon^*$ 来表示近似数 x^* 的精确度，或准确值所在的范围。同样有 $-\varepsilon^* \leqslant e^* \leqslant \varepsilon^*$，即 $|e^*|$ 是在 ε^* 的范围内，所以 ε^* 应取得尽可能小。例如，$x = 4.376\,281\,6\cdots$，取近似数 $x^* = 4.376$，则 $x - x^* = 0.000\,281\,6\cdots$，这时

$$|e^*| = 0.000\,281\,6\cdots < 0.000\,3 = 0.3 \times 10^{-3}$$

同样

$$|e^*| = 0.000\,281\,6\cdots < 0.000\,29 = 0.29 \times 10^{-3}$$

显然，0.3×10^{-3} 和 0.29×10^{-3} 都是 $|e^*|$ 的上界，都可以作为近似值 x^* 的绝对误差限，即

$$\varepsilon^* = 0.3 \times 10^{-3} \text{ 或 } \varepsilon^* = 0.29 \times 10^{-3}$$

由此可见，绝对误差限 ε^* 不是惟一的，这是因为一个数的上界不惟一所致。但是 ε^* 越小，x^* 近似真值 x 的程度越好，即 x^* 的精度越高。在实际应用中，往往根据需要对准确值取近似值，按四舍五入原则取近似值是使用最广泛的取近似值的方法。

例 1-1 用一把有毫米刻度的尺子，测量桌子的长度，读出来的值 $x^* = 1\,235$ mm，这是桌子实际长度 x 的一个近似值，由尺子的精度可以知道，这个近似值的误差不会超过 1/2 mm。

$$|x - x^*| = |x - 1\,235 \text{ mm}| \leqslant \frac{1}{2} \text{ mm}$$

$$1\,234.5 \text{ mm} \leqslant x \leqslant 1\,235.5 \text{ mm}$$

这表明真值 x 在 $[1\,234.5, 1\,235.5]$ 区间内，写成

$$x = (1\,235 \pm 0.5) \text{ mm}$$

这里绝对误差限 $\varepsilon^* = 0.5$ mm，即绝对误差限是末位的半个单位。

下面讨论“四舍五入”的绝对误差限。

设 x 为一实数，其十进制表示的标准形式（十进制规格化浮点数形式）

$$x = \pm 0.x_1x_2\cdots \times 10^m$$

其中，m 是整数；$x_1, x_2, \cdots$ 是 $0, 1, \cdots, 9$ 中的任一数，但 $x_1 \neq 0$。若经过四舍五入保留 n 位数字，得到近似值

$$x^* = \begin{cases} \pm 0.x_1x_2\cdots x_n \times 10^m, \text{当 } x_{n+1} \leqslant 4 \text{（四舍）} \\ \pm 0.x_1x_2\cdots x_{n-1}(x_n + 1) \times 10^m, \text{当 } x_{n+1} \geqslant 5 \text{（五入）} \end{cases}$$

四舍时的绝对误差

$$|x-x^*| = (0.x_1x_2\cdots x_nx_{n+1}\cdots - 0.x_1x_2\cdots x_n)\times 10^m$$
$$\leqslant (0.x_1x_2\cdots x_n499\cdots - 0.x_1x_2\cdots x_n)\times 10^m$$
$$= 10^m \times 0.\underbrace{0\cdots0}_{n个0}499\cdots \leqslant \frac{1}{2}\times 10^{m-n}$$

五入时的绝对误差

$$|x-x^*| = (0.x_1x_2\cdots x_{n-1}(x_n+1) - 0.x_1x_2\cdots x_n\cdots)\times 10^m$$
$$= (0.\underbrace{0\cdots0}_{n-1个0}1 - 0.\underbrace{0\cdots0}_{n个0}x_{n+1}\cdots)\times 10^m$$
$$\leqslant 10^{m-n}(1-0.x_{n+1})$$

由于此时 $x_{n+1}\geqslant 5$,所以 $1-0.x_{n+1}\leqslant \frac{1}{2}$,有

$$|x-x^*| \leqslant \frac{1}{2}\times 10^{m-n} \tag{1-3}$$

所以,四舍五入得到的近似数的绝对误差限是其末位的半个单位,即

$$\varepsilon^* = \frac{1}{2}\times 10^{m-n}$$

例 1-2 圆周率 $\pi=3.141\ 59\cdots$,用四舍五入取小数点后 4 位时,近似值为 3.141 6,此时 $m=1, n=5, m-n=1-5=-4$,故绝对误差限 $\varepsilon^*=\frac{1}{2}\times 10^{-4}$。同样,取小数点后 2 位时,近似值为 3.14,其绝对误差限 $\varepsilon^*=\frac{1}{2}\times 10^{-2}$。

对于用四舍五入取得的近似值,专门定义有效数字来描述它。

1.3.2 相对误差

在同一量的近似值中,绝对误差小的,精度高,但是,绝对误差不能比较不同条件下的精度。例如,测量 10 mm 误差是 1 mm,测量 1 m 误差是 2 mm,后者比前者绝对误差大,但可以看出在精度上后者比前者情况好,这是因为一个量的近似值的精度不仅与绝对误差有关,还与该量本身的大小有关,为此引入相对误差的概念。

定义 1-3 相对误差是近似数 x^* 的绝对误差 e^* 与准确值 x 的比值,即

$$\frac{e^*}{x} = \frac{x-x^*}{x}, \quad 其中\ x\neq 0 \tag{1-4}$$

相对误差说明了近似数 x^* 的绝对误差 e^* 与 x 本身比较所占的比例,它反映了一个近似数的准确程度,相对误差越小,精度就越高。但由于真值 x 总是不知道的,因此在实际问题中,常取相对误差

$$e_r(x^*) = \frac{e^*}{x^*} = \frac{x-x^*}{x^*} \tag{1-5}$$

并将 $e_r(x^*)$ 简记为 e_r^*。

当 $|e_r^*| = \left|\frac{e^*}{x^*}\right|$ 较小时,有

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

是 e_r^* 的平方级，故可忽略不计，实际问题中按式(1-5)取相对误差是合理的。

上例中 10 mm 时误差为 1 mm，1 m 时误差为 2 mm，其相对误差分别为 0.1，0.002，前者绝对误差小，后者相对误差小，后者精度比前者高。

在实际计算中，由于 e^* 和 x 都不能准确地求得，因此相对误差 e_r^* 也不可能准确地得到，于是也像绝对误差一样，只能估计相对误差的范围。

相对误差可正可负，其绝对值的上界取为相对误差限，因为 ε^* 是 x^* 的绝对误差限，则 $\dfrac{\varepsilon^*}{|x^*|}$ 是 x^* 的相对误差限，即有如下定义。

定义 1-4

$$|e_r^*| = \left|\frac{e^*}{x^*}\right| = \left|\frac{x - x^*}{x^*}\right| \leqslant \frac{\varepsilon^*}{|x^*|} = \varepsilon_r(x^*) \tag{1-6}$$

$\varepsilon_r(x^*)$ 称为相对误差限，在实际计算中用作相对误差，所以相对误差一般是指 $\varepsilon_r(x^*)$，$\varepsilon_r(x^*)$ 可简记为 ε_r^*。显然相对误差是无量纲的，通常用百分数表示。

由定义 1-4 可知，相对误差限可由绝对误差限求出，反之，绝对误差限也可由相对误差限求出，即

$$\varepsilon^* = |x^*|\varepsilon_r$$

例 1-3 光速 $c^* = (2.997\ 925 \pm 0.000\ 001) \times 10^{10}$ cm/s，其相对误差限 $\varepsilon_r^* = \dfrac{\varepsilon^*}{|c^*|} = \dfrac{0.000\ 001}{2.997\ 925} = 3.34 \times 10^{7}$，其中 $c^* = 2.997\ 925 \times 10^{10}$ cm/s 是目前光速公认值(测量值)。

例 1-4 取 3.14 作为 π 的四舍五入的近似值时，试求其相对误差限。

解 四舍五入的近似值 $x^* = 3.14$ 的绝对误差限为 $\varepsilon^* = \dfrac{1}{2} \times 10^{-2}$，则其相对误差限

$$\varepsilon_r^* = \frac{\varepsilon^*}{|x^*|} = \frac{\frac{1}{2} \times 10^{-2}}{3.14} = 0.159\%$$

1.3.3 有效数字

在表示一个近似数时，常常用到“有效数字”，有效数字是由绝对误差决定的。

定义 1-5 设数 x 的近似值 $x^* = 0.x_1x_2\cdots x_n \times 10^m$，其中 x_i 为 0～9 之间的任意数，但 $x_1 \neq 0$，$i = 1,2,3,\cdots,n$ 为正整数，m 为整数，若

$$|x - x^*| \leqslant \frac{1}{2} \times 10^{m-n} \tag{1-7}$$

则称 x^* 为 x 的具有 n 位有效数字的近似值，x^* 准确到第 n 位，$x_1x_2\cdots x_n$ 是 x^* 的有效数字。

例 1-5 $\pi = 3.14\ 592\cdots$，当取 3.142 作为其近似值时

$$|\pi - 3.142| = 0.000\ 407\cdots < 0.000\ 5 = \frac{1}{2} \times 10^{-3}$$

即 $m - n = -3$，$m = 1$，$n = 4$，所以 3.142 作为 π 的近似值有 4 位有效数字。

当取 3.141 作为 π 的近似值时

$$|\pi - 3.141| = 0.000\ 59\cdots < 0.005 = \frac{1}{2} \times 10^{-2}$$

即 $m-n=-2, m=1, n=3$，所以 3.141 作为 π 的近似值时有 3 位有效数字，不具有 4 位有效数字，3.14 是有效数字，千分位的 1 不是有效数字。

如果近似数 x^* 的误差限是某一位的半个单位，由该位到 x^* 的第一位非零数字一共有 n 位，x^* 就有 n 位有效数字，也就是说准确到该位。若用四舍五入法取准确值的前 n 位作为近似值 x^*，则 x^* 有 n 位有效数字，其中每一位数字都叫 x^* 的有效数字。但是，如果 x^* 准确到某位数字，将这位数字以后的数字进行四舍五入（可简称舍入）则不一定得到有效数字。例如，4.145 作为 4.144 的近似值是准确到百分位，若再四舍五入得到 4.15，其最后一位便不是有效数字了，4.15 只有两位有效数字。

例 1-6 四舍五入得到的近似值 0.020 3 有 3 位有效数字，其绝对误差 $\varepsilon^*=\frac{1}{2}\times10^{-4}$。

又如 3.142 是 π 的四舍五入的近似值，所以有 4 位有效数字。3.141 不是 π 的四舍五入的近似值，所以不具有 4 位有效数字。

关于有效数字还可指出以下几点：

1）若用四舍五入取准确值的前 n 位 x^* 作为近似值，则 x^* 必有 n 个有效数字。

例如，$\pi=3.141\ 592\ 6\cdots$，取 3.14 作为近似值，则有 3 位有效数字，取 3.142 作为近似值，则有 4 位有效数字。

2）有效数字位数相同的两个近似数，绝对误差不一定相同。

例如，设 $x_1^*=12\ 345$，$x_2^*=12.345$ 二者均有 5 位有效数字，前者的绝对误差为 $\frac{1}{2}\times1$，后者的绝对误差为 $\frac{1}{2}\times10^{-3}$。

3）把任何数乘以 $10^p(p=0,\pm1,\pm2,\cdots)$ 等于移动该数的小数点，这样并不影响其有效数字的位数。

例如，$g=9.80\ \mathrm{m/s^2}$ 具有 3 位有效数字，而 $g=0.009\ 80\times10^3\ \mathrm{m/s^2}$ 也具有 3 位有效数字。但 $9.8\mathrm{m/s^2}$ 与 $9.80\ \mathrm{m/s^2}$ 的有效数字位数是不同的，前者有两位，后者有 3 位。因此，要注意诸如 0.1，0.10，0.100，…它们的不同含义。

如果整数并非全是有效数字，则可用浮点数表示。如已知近似数 300 000 的绝对误差限不超过 500，即 $\frac{1}{2}\times10^3$，则应把它表示成 $x^*=300\times10^3$ 或 0.300×10^6。若记为 $x^*=300\ 000$，则表示其误差限不超过 $\frac{1}{2}$。这是因为

$$|x-300\times10^3|=|x-0.300\times10^6|\leqslant500=\frac{1}{2}\times10^{6-3}$$

即 $m=6, n=3$，而

$$300\ 000=10^6\times(3\times10^{-1}+0\times10^{-2}+\cdots+0\times10^{-6})$$

且

$$|x-300\ 000|\leqslant\frac{1}{2}\times10^{6-6}$$

即 $m=6, n=6$。

前者有 3 位有效数字，后者有 6 位有效数字。

例 1-7 某地粮食产量为 875 万吨，表示成

$$875\text{ 万吨} = 875 \times 10^4\text{ 吨} = 0.875 \times 10^7\text{ 吨}$$

绝对误差为$\frac{1}{2} \times 10^4$或$\frac{1}{2} \times 10^{-3} \times 10^7$,即$\frac{1}{2}$万吨。而875万吨不能表示成8 750 000吨,因为这时绝对误差为$\frac{1}{2}$吨。

有效数字位数与小数点后有多少位无关。但是具有 n 位有效数字的近似数 x^*,其绝对误差限 $\varepsilon^* = \frac{1}{2} \times 10^{m-n}$,在 m 相同的情况下,n 越大则 ε^* 越小。所以一般来说,近似同一真值的近似数有效数字位数越多,绝对误差越小。

4) 准确值被认为具有无穷多位有效数字。

例如,直角三角形面积 $S = \frac{1}{2}ah = 0.5ah$,其中 a 是底边,h 是高,不能认为公式中用0.5表示$\frac{1}{2}$时,只有一位有效数字。因为0.5是真值,没有误差,$\varepsilon^* = 0$,因此 $n \to \infty$,所以准确值具有无穷多位有效数字。至于底边 a 和高 h 是测量得到的,因此是近似数,应根据测量仪器精度来确定其有效数字的位数。

1.3.4 有效数字与相对误差

根据有效数字与相对误差的概念可以得出二者之间的关系。

定理1-1 若近似数 $x^* = \pm 0.x_1x_2\cdots x_n \times 10^m$ 具有 n 位有效数字,则其相对误差

$$|e_r^*| \leqslant \frac{1}{2x_1} \times 10^{-(n-1)} \tag{1-8}$$

证 由于

$$x^* = \pm 0.x_1x_2\cdots x_n \times 10^m$$

$$|x^*| \geqslant x_1 \times 10^{m-1}$$

又由于 x^* 具有 n 位有效数字,则

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

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

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

实际应用中,可以取

$$\varepsilon_r^* = \frac{1}{2x_1} \times 10^{-(n-1)}$$

由于 n 越大,ε_r^* 越小,所以有效数字位数越多,相对误差就越小。

例1-8 取3.14作为 π 的四舍五入的近似值时,试求其相对误差。

解 四舍五入的近似值3.14,其各位都是有效数字,即 $n = 3$,所以

$$\varepsilon_r^* = \frac{1}{2 \times 3}10^{-(3-1)} = 0.17\%$$