

科学与工程计算

KEXUE YU

廖晓钟 赖汝 编

GONGCHENG JISUAN



国防工业出版社

科学与工程计算

廖晓钟 赖汝 编

国防工业出版社

·北京·

图书在版编目(CIP)数据

科学与工程计算/廖晓钟, 赖汝编. —北京: 国防工业出版社, 2003.1

ISBN 7-118-02954-8

I. 科... II. ①廖...②赖... III. 数值计算—计算方法 IV. 0241

中国版本图书馆 CIP 数据核字(2002)第 071703 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 100044)

涿中印刷厂印刷

新华书店经售

*

开本 787×1092 1/16 印张 19½ 450 千字

2003 年 1 月第 1 版 2003 年 1 月北京第 1 次印刷

印数: 1—3000 册 定价: 27.00 元

(本书如有印装错误, 我社负责调换)

前 言

随着计算机技术和计算数学的发展,在计算机上用数值计算方法进行科学与工程计算的科学计算的应用范围已扩大到许多学科,科学计算成为与理论分析、科学实验同样重要的科学研究方法。科学研究和工程技术中提出的数学问题往往要求数值解,利用计算机求解各种数学模型的数值计算方法已成为广大科学与技术人员的必备知识。许多高等学校普遍将科学与工程计算中最基本的数值计算方法列入必修课或选修课。

本书介绍与现代科学与工程计算有关的基本数值计算方法,阐明了数值计算方法的基本理论和应用,讨论了一些数值方法的收敛性和稳定性,以及在计算机上实现时的一些问题。内容包括绪论、方程求根、解线性代数方程组的直接法、解线性代数方程组的迭代法、插值和最小二乘法、数值积分和数值微分以及常微分方程数值解等7章。各章内容具有相对独立性,可根据需要进行取舍。为便于自学,书中对各种方法都配有丰富的例题和简单的计算实例,每章后面有思考题和习题,最后附有习题参考答案。本书作为一本入门教材,阅读时需具备高等数学和线性代数知识。

本书是在多年教学实践的基础上,参考当前科学与工程计算、数值分析和计算方法教材编写而成的,可作为工科各专业本科生教材也可供科技人员参考。在编写过程中,参考了国务院学位办编写的《计算机科学与技术学科综合水平全国统一考试大纲及指南》中的《数值分析》部分,按《数值分析》大纲和复习指南的顺序进行编写,并采用了相同的符号,选用了其中的部分思考题和考试样卷作为例题或复习题,因此,便于本科生进一步学习和深造,对有关应试者也有参考价值。书末列出了部分参考书目,谨向被本书参考过的列入和未列入参考书目的编著者致以衷心的感谢。

感谢陈杰教授、李庆常教授对本书编写给予的热情关心和支持。本书是在国防工业出版社陈洁主任精心策划和辛勤编辑下出版的,她为本书的出版和保证出版质量起到了关键作用。

限于作者水平,书中缺点和错误之处,敬祈批评指正。

编 者

内 容 简 介

本书介绍现代科学与工程计算中常用的数值计算方法,阐明了数值计算方法的基本理论及其实现方法。内容包括数值计算的误差分析,方程求根,解线性方程组的直接法,解线性方程组的迭代法,插值和最小二乘法,数值积分和微分,常微分方程数值解。选材既考虑基础理论,又注重实用性,并有丰富的典型例题和计算实例,每章后有思考题和适量的习题,书末附有习题参考答案。

本书可作为高等学校工科本科生学习科学与工程计算、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。

目 录

第1章 绪论	1
1.1 科学与工程计算	1
1.2 误差的产生和分类	2
1.3 误差和有效数字	4
1.3.1 绝对误差和绝对误差限	4
1.3.2 有效数字	5
1.3.3 相对误差和相对误差限	7
1.3.4 有效数字与相对误差	8
1.4 运算误差分析	11
1.4.1 函数运算误差	11
1.4.2 算术运算误差	12
1.5 数值稳定性和减小运算误差	13
1.5.1 数值稳定性	13
1.5.2 减小运算误差	15
1.6 数学基础	19
1.6.1 微积分的若干定理	19
1.6.2 高等代数的若干概念和结论	20
思考题	26
习题	26
第2章 方程求根	29
2.1 二分法	29
2.1.1 初始值的搜索	30
2.1.2 区间二分法	30
2.2 迭代法的一般理论	32
2.2.1 不动点迭代	32
2.2.2 迭代法的收敛性	34
2.2.3 迭代法的收敛阶	42
2.2.4 迭代过程的加速	44
2.3 牛顿迭代法	48
2.3.1 牛顿迭代公式	48
2.3.2 局部收敛和收敛阶	51
2.3.3 改进牛顿迭代公式	52

2.4 弦位法	57
2.5 多项式方程求根	59
2.5.1 牛顿法求根	59
2.5.2 劈因子法	61
思考题	66
习题	67
第3章 解线性代数方程组的直接法	69
3.1 高斯消去法	70
3.1.1 基本思想	70
3.1.2 计算步骤	71
3.1.3 使用条件	75
3.1.4 计算量	77
3.1.5 矩阵描述	78
3.2 列主元高斯消去法	82
3.2.1 数值不稳定	82
3.2.2 选主技术	83
3.2.3 计算矩阵行列式	87
3.2.4 高斯-约当消去法	88
3.3 直接三角分解法	91
3.3.1 杜里特尔分解法	91
3.3.2 三对角方程组的追赶法	99
3.3.3 对称正定阵的乔累斯基分解法	103
3.4 向量和矩阵的范数	107
3.4.1 向量的范数	107
3.4.2 矩阵的范数	111
3.5 病态方程组和误差估计	114
3.5.1 病态方程组和矩阵的条件数	114
3.5.2 误差估计	118
思考题	119
习题	119
第4章 解线性代数方程组的迭代法	123
4.1 迭代法	123
4.1.1 迭代格式的构造	123
4.1.2 雅可比迭代	124
4.1.3 高斯-塞德尔迭代	125
4.1.4 松弛法	127
4.1.5 迭代公式的矩阵形式	128
4.2 迭代的收敛性	130
4.2.1 收敛的充分必要条件	130

4.2.2 用迭代矩阵的充分条件	133
4.2.3 用特殊系数矩阵的充分条件	139
4.2.4 松弛法的收敛性	143
思考题	144
习题	144
第5章 插值和最小二乘法	147
5.1 引言	147
5.1.1 多项式插值	147
5.1.2 多项式插值的唯一性	148
5.2 拉格朗日插值	149
5.2.1 线性插值和抛物线插值	149
5.2.2 基函数和拉格朗日插值公式	152
5.2.3 反插值	155
5.2.4 插值余项及误差估计	157
5.2.5 迭代插值	161
5.3 牛顿插值	164
5.3.1 差商及其性质	164
5.3.2 牛顿插值公式	167
5.3.3 差商和导数	170
5.3.4 差分	172
5.3.5 等距节点牛顿插值公式	175
5.4 埃尔米特插值	177
5.4.1 埃尔米特插值多项式的构造	177
5.4.2 埃尔米特插值的唯一性及余项	178
5.4.3 带不完全导数埃尔米特插值多项式	179
5.5 分段插值法	180
5.5.1 高次插值的龙格现象	180
5.5.2 分段线性插值	181
5.5.3 分段三次埃尔米特插值	183
5.6 三次样条	184
5.6.1 三次样条插值	184
5.6.2 三弯矩方程	186
5.7 正交多项式	192
5.7.1 预备知识	192
5.7.2 正交多项式	193
5.7.3 勒让德多项式	194
5.7.4 切比雪夫多项式	197
5.8 最小二乘曲线拟合	198
5.8.1 线性最小二乘拟合	198

5.8.2 可线性化模型的最小二乘拟合	202
5.8.3 多项式拟合	204
5.8.4 用正交多项式作最小二乘拟合	206
思考题	207
习题	208
第 6 章 数值积分和数值微分	211
6.1 引言	212
6.1.1 数值积分	212
6.1.2 代数精确度	213
6.1.3 插值求积公式	215
6.1.4 待定系数法	218
6.2 梯形公式和辛卜生公式	221
6.2.1 牛顿-柯特斯公式	221
6.2.2 低阶求积公式的代数精度	226
6.2.3 低阶求积公式的余项	228
6.2.4 求积公式的稳定性	232
6.2.5 复化求积法	233
6.3 外推原理和龙贝格算法	237
6.3.1 变步长梯形公式	237
6.3.2 外推原理	239
6.3.3 龙贝格算法	241
6.4 高斯型求积公式	243
6.4.1 引言	243
6.4.2 高斯-勒让德求积公式	246
6.4.3 带权的高斯型求积公式	249
6.4.4 高斯-切比雪夫求积公式	251
6.4.5 数值稳定性	252
6.5 数值微分	252
6.5.1 差商求导	252
6.5.2 插值求导	254
思考题	257
习题	257
第 7 章 常微分方程数值解	260
7.1 欧拉法和改进欧拉法	261
7.1.1 欧拉法	261
7.1.2 局部截断误差和阶	264
7.1.3 隐式欧拉法和两步法	264
7.1.4 梯形法	266
7.1.5 改进欧拉法	266

7.2 龙格-库塔法	269
7.2.1 引言	269
7.2.2 龙格-库塔法的基本思想	269
7.2.3 二阶龙格-库塔法的推导	271
7.2.4 经典龙格-库塔法	274
7.3 误差和稳定性	278
7.3.1 误差和方法的阶	278
7.3.2 单步法的稳定性	280
7.4 线性多步法	282
7.4.1 引言	282
7.4.2 亚当斯法	284
7.4.3 预报-校正公式	287
7.5 微分方程组和高阶微分方程	289
思考题	292
习题	293
习题参考答案	296
参考文献	302

第1章 绪 论

本章介绍科学与工程计算中常用的数值计算方法的含义及其特点和数值运算中的误差分析。误差分析包括误差的产生和分类、误差和有效数字、运算误差分析、数值稳定性和减小运算误差。本章最后给出了常用的数学基础知识。

1.1 科学与工程计算

随着计算机的发展与普及,继理论分析、科学实验之后,在计算机上用数值计算方法进行科学与工程计算的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,因此,数值计算的内容相当丰富。科学与工程计算中常用的数值计算方法又称数值分析或计算方法。数值计算方法是应用数学的一个分支,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、常微分方程等基础数学中最常用的、行之有效的数值方法,并通过典型例子阐明构造数值计算方法的基本思想和技巧,引出相应方法的步骤。

应用计算机解决科学研究与工程设计中的科学计算问题时,先由有关专业领域提出实际问题,然后或由理论根据各专业学科的规律,利用各种定律、原理、规则、公式等分析和推演出其数学模型,或由实验根据运行特性及数据识别出实际问题的数学模型,对实际问题进行定量的描述。这里的数学模型是用来描述实际问题中各种物理量间定量关系的数学结构。在科学与工程计算中常用到的数学模型有非线性方程、线性代数方程组、插值和最小二乘法、积分和微分、常微分方程等。最后,根据各种数学模型选用或构造相应数值方法,并进行程序设计,上机计算得出实际问题的数值结果。因此,选用或构造数值方法是应用计算机进行科学与工程计算全过程中重要的一环。

数值计算方法是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计算方法及其相关理论,包括误差分析、收敛性和稳定性等内容,它的任务是面向计算机,能提供在计算机上实际可行,达到要求精度,理论分析可靠,计算复杂性好的各种数值方法。

数值计算方法常采用如下一些处理方法。

①构造性方法。数值方法中许多问题的存在性的证明都是以“构造性”方法为基础,所谓用构造性方法证明数学问题的存在性,就是具体地把该问题的计算公式构造出来。这种方法不仅证明了问题的存在性,而且给出了具体的计算公式,使需要证明其存在性的问题通过数学运算过程来完成,以便于编程上机进行求解。

②离散化方法。离散化就是把连续变量问题转化为求离散变量的问题,以便应用计算机求解。例如,对定积分进行离散求和、微分方程离散成差分方程等。

③递推化方法。递推是把一个复杂过程归结为简单过程的多次重复,以便于编程上机计算。

④迭代法。迭代法是一种逐次逼近的方法,即用某个固定公式反复计算一个数值过程,使之得出满足一定精度要求的数值解。在实际应用和理论上,迭代法都是一种重要的方法,在处理线性问题 and 非线性问题时,迭代法都获得了广泛的应用。

⑤近似替代方法。把一个无穷过程转化成一个满足一定误差要求的有限过程,用计算机进行有限次运算得出无穷过程的近似结果。

⑥以直代曲法。这也是一种近似替代方法,它是将非线性问题进行线性化处理,也就是在局部范围内用直线代替曲线,或进一步用多项式逼近复杂函数。

⑦化整为零的处理方法。这种方法是将整体分割成若干小部分处理,在每个小部分可以求出高精度的近似值,以此得到整体的近似值,从而使整体的误差减小,达到提高精度的目的。

⑧外推法。这是对依赖于步长的数值计算公式快速提高逼近程度的一种有效方法,在求数值积分和数值微分时都有重要的应用。

以下将通过对各种数值方法的分析,深入阐明这些处理方法的原理和功能及其在数值计算中的作用。

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,在学习过程中应该充分重视这一环节。学习本课程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,这就要求在学习本课程时,对手算和上机实习都应给予充分的重视。

1.2 误差的产生和分类

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,如某教室里有 42 名学生,数 42 就是准确数;另一类数则不是这样,它们只是近似地反映实际情况,这类数称为近似数或称某准确数的近似值,如从测量得到桌子的长度为 950mm,一般说来,这个测量值 950 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,绝大多数情况下不存在绝对的严格和精确,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 4 类误差。

在定量分析客观事物时,总是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的距离称为模型误差或称描述误差。

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研究分析,求解数学模型时常遇到的误差是截断误差和舍入误差。

许多数学模型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方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(例如 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。对这个问题,其截断误差是

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

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 $\pi = 3.1415926\cdots$, $\sqrt{2} = 3.141421356\cdots$, $\frac{1}{3} = 0.3333\cdots$ 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 4 位有效数字,则 $3.1416 - \pi = 0.0000073\cdots$, $1.4142 - \sqrt{2} = -0.000013\cdots$, $0.3333 - \frac{1}{3} = -0.000033\cdots$ 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 位数字进入“四舍五入”,这时产生的误差就是舍入误差。

总括起来,误差类型一般有模型误差、观测误差、截断误差和舍入误差。但在数值计算方法中主要讨论的是截断误差和舍入误差。

1.3 误差和有效数字

近似数的误差常用绝对误差、有效数字和相对误差表示。下面介绍这3种表示方法及其互相之间的关系。

1.3.1 绝对误差和绝对误差限

设 x^* 是准确值 x 的一个近似值, 定义准确值 x 与其近似值 x^* 之差

$$e(x^*) = x - x^* \quad (1.1)$$

为近似值 x^* 的绝对误差, 简称误差。 $e(x^*)$ 又简记为 e^* 。

从这个定义可以看出, 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^* , 把这个正数 ϵ^* 叫做误差绝对值的上界或称误差限。

定义 如果

$$|e^*| = |x - x^*| \leq \epsilon(x^*) \quad (1.2)$$

则称 $\epsilon(x^*)$ 为 x^* 近似 x 的绝对误差限, 实用中简称误差、误差限、绝对误差, 用它反映近似数的精度。 $\epsilon(x^*)$ 又可简记为 ϵ^* 。

从上述定义可以看出, ϵ^* 是一个正数。又因为在任何情况下, 都有

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

即

$$x^* - \epsilon^* \leq x \leq x^* + \epsilon^*$$

这就表明准确值 x 在 $[x^* - \epsilon^*, x^* + \epsilon^*]$ 这个区间内, 用 $x = x^* \pm \epsilon^*$ 来表示近似数 x^* 的精确度, 或准确值所在的范围。同样有 $-\epsilon^* \leq e^* \leq \epsilon^*$, 即 $|e^*|$ 是在 ϵ^* 的范围内, 所以 ϵ^* 应取得尽可能小, 例如 $x = 4.3762816\cdots$, 取近似数 $x^* = 4.376$, 则 $x - x^* = 0.0002816\cdots$, 有

$$|e^*| = 0.0002816 < 0.0003 = 0.3 \times 10^{-3}$$

同样

$$|e^*| = 0.0002816 < 0.00029 = 0.29 \times 10^{-3}$$

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

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

由此可见, 绝对误差限 ϵ^* 不是唯一的, 这是因为一个数的上界不唯一所致。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 用一把有毫米刻度的尺子测量桌子的长度,读出的值 $x^* = 1235\text{mm}$ 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 $1/2\text{mm}$,即

$$|x - x^*| = |x - 1235\text{mm}| \leq \frac{1}{2}\text{mm}$$

$$1234.5\text{mm} \leq x \leq 1235.5\text{mm}$$

这表明真值 x 在 $[1234.5, 1235.5]$ 这个区间内,写成

$$x = (1235 \pm 0.5)\text{mm}$$

这里绝对误差 ϵ^* 是末位的半个单位。

下面讨论“四舍五入”的误差限。

设 x 为一实数,其十进制表示的标准形式为

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

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

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

四舍时的误差限

$$\begin{aligned} |x - x^*| &= (0.x_1x_2\cdots x_nx_{n+1}\cdots - 0.x_1x_2\cdots x_n) \times 10^m \leq \\ & (0.x_1x_2\cdots x_n499\cdots - 0.x_1x_2\cdots x_n) \times 10^m = \\ & 10^m \times 0.\underbrace{0\cdots 0}_{n\uparrow 0}499\cdots \leq \frac{1}{2} \times 10^{m-n} \end{aligned}$$

五入时的误差限

$$\begin{aligned} |x - x^*| &= \{0.x_1x_2\cdots x_{n-1}(x_n + 1) - 0.x_1x_2\cdots x_n\cdots\} \times 10^m = \\ & (0.\underbrace{0\cdots 01}_{n-1\uparrow 0} - 0.\underbrace{0\cdots 0}_{n\uparrow 0}x_{n+1}\cdots) \times 10^m \leq 10^{m-n}(1 - 0.x_{n+1}) \end{aligned}$$

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

$$|x - x^*| \leq \frac{1}{2} \times 10^{m-n} \quad (1.3)$$

四舍五入的误差限是末位的半个单位。

例 圆周率 $\pi = 3.14159\cdots$, 用四舍五入取小数点后 4 位时, 近似值为 3.1416, 此时 $m = 1, n = 5, m - n = 1 - 5 = -4$, 故绝对误差(限) $\epsilon^* = \frac{1}{2} \times 10^{-4}$ 。同样取小数点后

2 位时, 近似值为 3.14, 其绝对误差 $\epsilon^* = \frac{1}{2} \times 10^{-2}$ 。

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 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 设数 x 的近似值 $x^* = \pm 0.x_1x_2\cdots x_n \times 10^m$, 其中 x_i 是 0 到 9 之间的任一个数, 但 $x_1 \neq 0, i = 1, 2, 3, \cdots, n$ 是正整数, m 是整数, 若

$$|x - x^*| \leq \frac{1}{2} \times 10^{m-n} \quad (1.4)$$

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, $x_1, x_2, \dots, x_n$ 是 x^* 的有效数字。

例 $\pi = 3.141592\cdots$, 当取 3.142 作为其近似值时

$$|\pi - 3.142| = 0.000407\cdots < 0.0005 = \frac{1}{2} \times 10^{-3}$$

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

$$|\pi - 3.141| = 0.00059\cdots < 0.005 = \frac{1}{2} \times 10^{-2}$$

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, 其中每一位数字也都叫 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 4.138 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 四舍五入得到的近似值 0.0203 有 3 位有效数字, 其绝对误差 $\epsilon^* = \frac{1}{2} \times 10^{-4}$ 。

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。3.142 是有效数字, 3.141 不是有效数字。

关于有效数字还可指出以下几点。

① 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, $\pi = 3.1415926\cdots$, 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 0.9460832 和 0.9460834 作为准确值 0.9460831 $\cdots$ 的近似值时, 前者的绝对误差比后者的绝对误差小, 也就是前者的精度好, 后者的精度差, 虽然二者都有 6 位有效数字。因此当用有效数字不足以比较近似数的精度时, 可进一步用绝对误差来衡量。

又如, 设 $x_1^* = 12345, x_2^* = 12.345$ 两者均有 5 位有效数字, 前者的绝对误差为 $\frac{1}{2} \times 1$, 后者的绝对误差为 $\frac{1}{2} \times 10^{-3}$ 。

③ 把任何数乘以 10^p ($p = 0, \pm 1, \pm 2, \cdots$), 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 $g = 9.80\text{m/s}^2$ 具有 3 位有效数字, 而 $g = 0.00980 \times 10^3\text{m/s}^2$ 也具有 3 位有效数字。但 9.8m/s^2 与 9.80m/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, $\cdots$ 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 30 万的绝对误差限

不超过 500 即 $\frac{1}{2} \times 10^3$, 则应把它表示成 $x^* = 300 \times 10^3$ 或 0.300×10^6 。若记为 300000, 则表示其误差限不超过 $\frac{1}{2}$ 。这是因为

$$|x - 300 \times 10^3| = |x - 0.300 \times 10^6| \leq 500 = \frac{1}{2} \times 10^{6-3}$$

即 $m = 6, n = 3$, 而

$$300000 = 10^6 \times (3 \times 10^{-1} + 0 \times 10^{-2} + \cdots + 0 \times 10^{-6})$$

且

$$|x - 300000| \leq \frac{1}{2} \times 10^{6-6}$$

即 $m = 6, n = 6$ 。

前者有 3 位有效数字, 后者有 6 位有效数字。

例 某地粮食产量为 870 万 t, 表示成

$$870 \text{ 万 t} = 870 \times 10^4 \text{ t} = 0.870 \times 10^7 \text{ t}$$

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 t。而 870 万 t 不能表示成 8700000t, 因为这时绝对误差为 $\frac{1}{2}$ t。

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 $\epsilon^* = \frac{1}{2} \times 10^{m-n}$, 在 m 相同的情况下, n 越大, 则 ϵ^* 越小。所以有效数字位数越多, 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, $\epsilon^* = 0$, 因此 $n \rightarrow \infty$, 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。

1.3.3 相对误差和相对误差限

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 10mm 误差是 1mm, 测量 1m 误差是 2mm, 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 相对误差是近似数 x^* 的绝对误差 e^* 与准确值 x 的比值, 即

$$\frac{e^*}{x} = \frac{x - x^*}{x}, \text{ 其中 } x \neq 0$$

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

$$e_r(x^*) = \frac{e^*}{x^*} = \frac{x - x^*}{x^*} \quad (1.5)$$