# 第1章

# 绪　　论

本章简要介绍数值分析(numerical analysis)的研究对象、内容和特点,误差的基本概念及其与有效数字的关系,并讨论数值计算的误差估计问题.此外,还着重讨论了误差的定性分析以及在数值计算中应遵循的若干原则.本章的知识框架可扫描右侧二维码获取.

1.1

## 1.1　数值分析的研究对象和特点

现代科学技术问题的研究方法大致可分为三种:理论推导、科学实验和科学计算.科学计算就是针对实际问题通过建立数学模型把科学技术问题转化为数学问题,然后对数学问题进行离散化,将其转化为数值问题,最后使用数值计算方法计算出数值问题的数值解,并把所得的数值解作为原科学技术问题解的近似.这一过程可以概括为:


对实际问题应用科学知识和数学理论建立数学模型,这一过程通常属于应用数学的范畴.而根据数学模型提出求解的数值方法、编写程序进行计算、并对结果进行分析,这一过程则是计算数学的任务,也正是数值分析的**研究对象**.因此,数值分析(也称计算方法、数值计算方法)是计算数学的核心内容,也是数学学科的一个重要分支.它主要研究利用计算机解决数学问题的数值计算理论与方法,并研究求解过程中出现的稳定性、收敛性和误差估计等问题.其**经典内容**包括:误差理论、线性与非线性方程(组)的数值解法、插值与拟合、函数逼近、数值微积分、常微分方程数值解等.

随着计算机的迅速发展,数值计算方法的应用已普遍深入到各个科学领域,很多复杂的、大规模的计算问题已成功地在计算机上得到解决,而且新的、有效的数值方法也不断出现.因此,数值分析所阐明的基本原理与方法是从事科学计算不可或缺的工具.

能用计算机计算的"数值问题"是指输入数据(即问题中的自变量与原始数据)与输出数据(结果)之间函数关系的一个确定而无歧义的描述,输入输出数据可用有限维向量表述.根据这种定义,"数学问题"有的是"数值问题",如线性方程组求解;也有不是"数值问题"的"数学问题",如常微分方程

$$\begin{cases}\dfrac{\mathrm{d}y}{\mathrm{d}x}=x^2+y^2,\\ y(0)=0.\end{cases}$$

它不是数值问题,因为输出不是数据而是连续函数 $y=y(x)$. 但只要将连续问题离散化,使输出数据是 $y(x)$ 在求解区间 $[a,b]$ 上的离散点 $x_i(i=1,2,\cdots,n)$ 上的近似值,就是"数值问题". 数值问题可用各种数值计算方法求解,计算方法(或称数值分析)就是研究数值问题的数值计算方法的学科.

计算主要由算子、输入元和输出元组成. 算子可以是简单操作,如算术运算($+,-,\times,\div$)、逻辑运算,也可以是宏操作,如向量运算、数组传输、基本初等函数求值等. 输入元和输出元可分别视为若干变量或向量. 由一个或多个算子组成一个进程,将输入元变换成一个输出元. 面向计算机的算法可分为串行算法和并行算法两类. 只有一个进程的算法适合于串行计算机,称为串行算法. 有两个以上进程的算法适合于并行计算机,称为并行算法. 对于一个给定的数值问题可以有许多不同的算法,它们都能给出近似答案,但所需的计算量和得到的精确程度可能相差很大. 一个面向计算机,有可靠理论分析且计算复杂性好的算法就是一个好算法. 理论分析主要是连续系统的离散化及离散型方程的数值问题求解,它包括误差分析、稳定性、收敛性等基本概念,它刻画了算法的可靠性和准确性. 计算复杂性包括计算时间复杂性和存储空间复杂性两个方面. 在同一规模和精度条件下,计算时间复杂性和存储空间复杂性实际上就是算法中的计算量和存储量分析. 对同一问题的不同算法其计算复杂性可能差别很大.

有些数学问题虽有理论上的准确公式解,但不一定实用. 例如,解 $n$ 阶线性方程组,若依照克莱姆(Cramer)法则用行列式解法要计算 $n+1$ 个 $n$ 阶行列式的值,总共需要 $n!(n-1)(n+1)$ 次乘法. 对 $n=20$ 的线性方程组就需要约 $9.7\times10^{20}$ 次乘法运算,即使用每秒千亿次的计算机也得需要算约 307 年,这是无法实现的. 若用高斯(Guass)列主元消元法则只需约 3060 次乘法运算.

由此可见算法研究至关重要,仅提高计算机速度而不改进和选用优良算法是不行的. 此外,许多问题只能依赖数值方法求得近似解,相对于精确解而言的称谓,为精确解的近似,例如:已知方程 $x^2=2\sin x$ 在区间(1,2)内有唯一实根,但无法求得其解析表达式,此时就必须借助数值方法来获取其近似解.

数值分析的核心在于研究求解数值问题的算法. 一个优秀的数值算法通常具备以下4个基本特征:

(1) 面向计算机的可操作性:算法设计须紧密结合计算机的工作原理,确保其由有限且明确的步骤构成. 具体而言,算法的基本操作应仅限于加、减、乘、除四则运算及逻辑判断,从而能够在计算机上直接、高效地实现.

(2) 严密的理论保证:算法必须具备可靠的理论基础. 这包括能够通过可控的步骤无限逼近精确解,并最终满足预设的精度要求. 理论分析须确保算法的数值稳定性、收敛性,并能对其计算误差进行有效的估计与控制.

(3) 良好的计算复杂度:算法应具备优良的计算复杂性. 理想的"时间复杂性"意味着算法执行耗时短、效率高;理想的"空间复杂性"则意味着算法运行时所耗费的存储资源少.

(4) 充分的数值实验验证:任何算法除满足上述理论要求外,还必须接受数值实验的检验. 通过在不同场景和参数下的实际计算,验证算法的稳定性、有效性及效率,从而为其理

论分析提供坚实的实践支撑.

根据“数值分析”课程的特点,学习时首先要注意掌握数值计算方法的基本思想和原理,要注意数值计算方法处理的技巧及其与计算机的结合,要重视算法的误差分析、收敛性和数值稳定性的基本理论;其次,要通过例子,学习使用各种数值计算方法解决实际计算问题;最后,为了掌握本课程的内容,还应做一定数量的理论分析和计算练习.

# 1.2 数值计算的误差

数值计算的结果与实际问题的准确值之间往往存在差异,这种差异称为误差.通常,数值计算的结果总是存在误差的.数值分析的核心任务之一,就是将这些误差控制在允许的范围内,或至少对误差的大小作出可靠的估计.

## 1.2.1 误差分析的重要性

在运用数值计算方法解决实际问题的过程中,会出现各种各样的误差.对这些误差进行分析是十分必要的,否则一个看似合理的计算可能会得出非常离谱的错误结果.

**例 1.1** 计算 $f(x)=e^x$ 在 $x=0$ 处导数的近似值.

**解**:由导数的定义知

$$f'(x)=\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}, \tag{1.1}$$

因此,导数的近似值可以写成

$$f'(x)\approx\frac{f(x+h)-f(x)}{h}. \tag{1.2}$$

由(1.1)式知,$h$ 越小,(1.2)式的近似程度就越高.

事实真是如此吗? $h$ 取不同值,并且在计算过程中其计算值始终保持小数点后 4 位.分别取 $h=0.01,0.001,0.0001,0.00001$,按(1.2)式计算得到:

$$f'(0)\approx\frac{f(h)-f(0)}{h}=\frac{e^{0.01}-e^0}{0.01}\approx\frac{1.0101-1}{0.01}=1.0100,$$

$$f'(0)\approx\frac{f(h)-f(0)}{h}=\frac{e^{0.001}-e^0}{0.001}\approx\frac{1.0010-1}{0.001}=1.0000,$$

$$f'(0)\approx\frac{f(h)-f(0)}{h}=\frac{e^{0.0001}-e^0}{0.0001}\approx\frac{1.0001-1}{0.0001}=1.0000,$$

$$f'(0)\approx\frac{f(h)-f(0)}{h}=\frac{e^{0.00001}-e^0}{0.00001}\approx\frac{1.0000-1}{0.00001}=0.0000.$$

从计算结果看到,当 $h=0.00001$ 时,计算误差反而增大.

从例 1.1 可见,对数值计算中的误差进行分析是十分必要的.

## 1.2.2 误差的来源与分类

科学计算大致遵循以下过程:首先对给定的实际问题建立数学模型,接着收集整理数据以确定模型中的参数,然后选用相应的数值方法,继而编写计算机程序进行数值计算,最终得到计算结果,如图 1.1 所示.

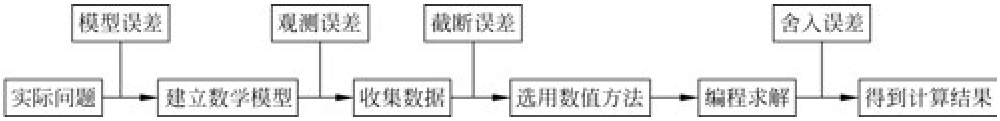


图 1.1 科学计算的过程以及误差的来源

因此,从实际问题到计算结果之间存在着以下几种误差.

**1. 模型误差**

用数学方法去解决一个具体的实际问题,首先要建立数学模型,它是对被描述的实际问题进行抽象和简化并忽略一些次要因素而得到的,因而是一种近似.这种数学模型与实际问题之间出现的误差称为**模型误差**.例如,用

$$s=\frac{1}{2}gt^2 \tag{1.3}$$

(其中 $g$ 是重力加速度)在描述质点自由落体运动时,忽略了空气阻力、风力等次要因素,因此其预测结果将不可避免地引入模型误差.

由于模型误差难以量化,通常在建模时只要对实际问题的抽象和简化合理,这种误差便可忽略不计,因此在“数值分析”的框架内,通常不予专门讨论.

**2. 观测误差**

在建立的数学模型中通常包含各种各样的物理量,如温度、长度、电压、电流等,这些物理量往往是通过观测或实验得到的,由于测量工具的精度、观测方法或客观条件的限制,观测难免存在误差,观测值与真实值之间的误差称为**观测误差**.

例如,在(1.3)式中取重力加速度为 $g\approx 9.8\mathrm{m/s^2}$,则 $g-9.8$ 就是观测误差.

数值分析本身并不讨论观测误差,但了解观测误差的存在与特性,对于在实际计算中选择合理的数值计算方法(即对数据误差稳健的算法)至关重要.

**3. 截断误差**

在计算中常常遇到只有通过无限过程才能得到结果,但实际计算时只能用有限过程来计算.这种用有限过程代替无限过程的误差称为**截断误差**.这种误差是由计算方法本身引起的,因此也称为**方法误差**.

例如,计算 $f(x)=\mathrm{e}^x$ 的值.考虑 $\mathrm{e}^x$ 的泰勒(Taylor)展开式

$$\mathrm{e}^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots,$$

若从第4项后“截断”,则有

$$\mathrm{e}^x\approx 1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}=P_3(x).$$

当 $|x|<1$ 时,其截断误差为

$$R_3(x)=\mathrm{e}^x-P_3(x)=\frac{x^4}{4!}\mathrm{e}^\xi\quad(\xi\text{ 在 }0\text{ 与 }x\text{ 之间}).$$

截断误差的大小直接影响计算结果的精度和计算工作量,是数值计算中必须要考虑的一类误差.

**4. 舍入误差**

在计算中遇到的数据可能位数很多,也可能是无穷小数,如 $\sqrt{2}$.但计算机在运算时只能

对有限位数进行计算.因此,原始数据、中间结果和最后结果都要进行四舍五入,这样产生的误差称为**舍入误差**.例如:用 $\pi^* = 3.14$ 近似代替 $\pi$,则舍入误差为 $\pi - \pi^* = 0.00159\cdots$.少量的舍入误差是微不足道的,但在计算机上完成了千百万次运算之后,舍入误差的累积有时可能是十分惊人的.

在上述误差来源中,模型误差与观测误差是客观存在的,而截断误差和舍入误差则是由计算方法和工具引起的.数值分析的核心任务,正是要讨论截断误差与舍入误差在计算过程中的传播及其对结果的影响,并研究控制它们的方法,从而保证最终结果满足精度要求.

### 1.2.3 误差与有效数字

绝对误差、相对误差以及有效数字是用来描述一个近似值的准确程度的有效指标.

**1. 绝对误差与绝对误差限**

**定义 1.1** 设准确值 $x$ 的一个近似值为 $x^*$,则称 $x - x^*$ 为近似值 $x^*$ 的**绝对误差**,简称**误差**,记为 $e$,即 $e = x - x^*$①.

例如,用 1.414 近似$\sqrt{2}$,其绝对误差为$\sqrt{2} - 1.414 = 1.414213\cdots - 1.414 = 0.000213\cdots$.

通常无法知道准确值 $x$,因而也不可能知道近似值 $x^*$ 的误差 $e$ 的准确值.但可以很容易得到 $e$ 的取值范围.例如,用 1.414 作为$\sqrt{2}$的近似值,其绝对误差的绝对值不会超过 0.0003,因此可以给出其绝对误差的上限.

**定义 1.2** 设准确值 $x$ 的一个近似值为 $x^*$,称 $x^*$ 的绝对误差的绝对值的上限为 $x^*$ 的**绝对误差限**,简称**误差限**,记为 $\varepsilon$,即 $|e| = |x - x^*| \leqslant \varepsilon$.

显然,如果 $\varepsilon$ 是准确值 $x$ 的近似值 $x^*$ 的绝对误差限,那么 $x$ 仅位于区间 $[x^* - \varepsilon, x^* + \varepsilon]$ 上.在工程技术中常用 $x = x^* \pm \varepsilon$ 表示.例如,用毫米刻度的直尺测量某一长度为 $x$ 的物体,测量其长度的近似值为 $x^* = 45\text{mm}$,由于直尺以毫米为刻度,其误差不超过 0.5mm,即 $x = 45 \pm 0.5\text{mm}$.

**例 1.2** 已知 $\pi = 3.1415926\cdots$,取其近似值 $\pi^* = 3.14$,计算 $e(\pi^*)$ 和 $\varepsilon(\pi^*)$.

**解**:依据绝对误差和绝对误差限的定义,可知

$$e(\pi^*) = \pi - \pi^* = 3.1415926\cdots - 3.14 = 0.0015926\cdots,$$

$$|e(\pi^*)| = 0.0015926\cdots \leqslant 0.002 = \varepsilon(\pi^*).$$

**2. 相对误差与相对误差限**

在许多情况下,绝对误差限并不能完全刻画一个近似值的精确程度.例如,比较

$$x = 10 \pm 1 \quad 和 \quad y = 1000 \pm 10$$

两种情况.从绝对误差限来看,$y^*$ 的绝对误差限是 $x^*$ 的绝对误差限的 10 倍;但从实际情况来看,$y^*$ 的精确程度要高于 $x^*$ 的精确程度.因此,一个近似值的精确程度不仅与绝对误差限有关,而且还与其准确值 $x$ 的大小有关.由此,给出相对误差的定义.

---

① 为明确表达:

- 绝对误差 $e = x - x^*$ 是近似值 $x^*$ 的函数(可严格记为 $e(x^*)$),本书在不引起混淆时简记为 $e$.
- 相对误差 $e_r = e/x$(或 $e/x^*$)同理是 $x^*$ 的函数,可严格记为 $e_r(x^*)$,简记为 $e_r$.
- 绝对误差限 $\varepsilon$ 是一个已知的正数,满足 $|x - x^*| \leqslant \varepsilon$.虽然该上界与 $x^*$ 有关,但通常直接以常数 $\varepsilon$ 表示.
- 相对误差限 $\varepsilon_r$ 类似,是满足 $|e_r| \leqslant \varepsilon_r$ 的已知正数,也以常数形式出现.

当需要区分多个近似值时,可使用下标区分.

**定义 1.3**　设准确值 $x$ 的一个近似值为 $x^*$，则称 $\frac{e}{x}$ 为近似值 $x^*$ 的**相对误差**，记为 $e_r$，即

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

在实际计算中，由于准确值 $x$ 未知，通常取

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

作为近似值 $x^*$ 的相对误差，条件是 $\frac{e}{x^*}$ 较小，此时

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

是 $e_r$ 的平方项级，故可忽略不计.

相对误差可正可负，与绝对误差限的概念类似，引入相对误差限的概念.

**定义 1.4**　设准确值 $x$ 的一个近似值为 $x^*$，称 $x^*$ 的相对误差 $e_r$ 的绝对值上限为 $x^*$ 的**相对误差限**，记为 $\varepsilon_r$，即 $|e_r|\leqslant\varepsilon_r$.

相对误差和相对误差限都是无量纲的数，常用百分数表示相对误差限. 例如，$x=10\pm1$ 和 $y=1000\pm10$ 的近似值 $x^*=10$ 和 $y^*=1000$ 的相对误差限分别为

$$\varepsilon_r(x^*)=\frac{1}{10}=10\%\quad 和\quad \varepsilon_r(y^*)=\frac{10}{1000}=1\%.$$

可见，$x^*$ 的相对误差限 $\varepsilon_r(x^*)$ 是 $y^*$ 的相对误差限 $\varepsilon_r(y^*)$ 的 10 倍，这意味着 $y^*$ 的精度比 $x^*$ 的精度高得多.

由定义 1.2 与定义 1.4，可以得到绝对误差限与相对误差限之间的关系为

$$\varepsilon_r=\frac{\varepsilon}{|x^*|}.$$

**例 1.3**　已知 $\pi=3.1415926\cdots$，取其近似值 $\pi^*=3.14$，计算 $e_r(\pi^*)$ 和 $\varepsilon_r(\pi^*)$.

**解**：依据相对误差和相对误差限的定义，可知

$$e_r(\pi^*)=\frac{\pi-\pi^*}{\pi^*}=\frac{e(\pi^*)}{\pi^*}=\frac{0.0015926\cdots}{3.14}=0.0005071\cdots,$$

$$|e_r(\pi^*)|=0.0005071\cdots\leqslant0.0006=\varepsilon_r(\pi^*).$$

**3. 有效数字与舍入误差**

在实际计算中，经常按四舍五入原则取近似值. 例如

$$\sqrt{200}=14.142\cdots\approx14.1,\quad e(14.1)=\sqrt{200}-14.1=0.042\cdots,$$

$$\lg2=0.30102\cdots\approx0.301,\quad e(0.301)=\lg2-0.301=0.00002\cdots,$$

$$\mathrm{e}^{-5}=0.0067379\cdots\approx0.00674,\quad e(0.00674)=\mathrm{e}^{-5}-0.00674=-0.0000020\cdots.$$

它们的绝对误差的绝对值都不会超过末位数字的半个单位，即

$$|\sqrt{200}-14.1|<\frac{1}{2}\times10^{-1},\quad \varepsilon(14.1)=\frac{1}{2}\times10^{-1},$$

$$|\lg2-0.301|<\frac{1}{2}\times10^{-3},\quad \varepsilon(0.301)=\frac{1}{2}\times10^{-3},$$

$$|\mathrm{e}^{-5}-0.00674|<\frac{1}{2}\times10^{-5},\quad \varepsilon(0.00674)=\frac{1}{2}\times10^{-5}.$$

若一个近似值是由准确值经四舍五入到某一位而得到,则其绝对误差限不超过该位上的半个单位.例如,若 $x^*=-1.15$ 和 $y^*=4.1200$ 是分别由准确值 $x$ 和 $y$ 经过四舍五入而得到的近似值,则

$$\varepsilon(x^*)=0.005,\quad \varepsilon_r(x^*)=\frac{0.005}{1.15}\approx 0.43\%,$$

$$\varepsilon(y^*)=0.00005,\quad \varepsilon_r(y^*)=\frac{0.00005}{4.1200}\approx 0.0012\%.$$

由此,给出近似值的有效数字定义如下.

**定义 1.5** 设准确值 $x$ 的一个近似值为 $x^*$,若近似值 $x^*$ 的绝对误差限是其某一位的半个单位,且该位到 $x^*$ 的左边第一位非零数字共有 $n$ 位,则称近似值 $x^*$ 有 $n$ 位有效数字.

例如,$\pi=3.1415926\cdots$的近似值 $\pi^*=3.142$,由于绝对误差

$$|e(\pi^*)|=0.0004073\cdots\leqslant 0.0005=\frac{1}{2}\times 10^{-3},$$

于是,绝对误差限 $\varepsilon(\pi^*)=0.5\times 10^{-3}$.因此,$\pi$ 的近似值 $\pi^*=3.142$ 有 3,1,4 和 2 共 4 位有效数字.

再如,若近似值 $a=1.23$,$b=-0.0725$,$c=1.86\times 10^{-5}$ 的绝对误差限都是 0.005,则 $a$ 有 3 位有效数字 1,2 和 3,$b$ 有 1 位有效数字 7,$c$ 没有有效数字.

一般地,准确值 $x$ 的有 $n$ 位有效数字的近似值 $x^*$ 总是可以表示为

$$\begin{aligned} x^* &= \pm 0.a_1a_2\cdots a_k\cdots\times 10^m \\ &= \pm(a_1\times 10^{-1}+a_2\times 10^{-2}+\cdots+a_k\times 10^{-k}+\cdots)\times 10^m, \end{aligned} \tag{1.4}$$

其中,$m$ 为整数,$k\geqslant n$,$a_1,a_2,\cdots$均为 0～9 之间的数,且 $a_1\neq 0$,并满足关系式

$$|x-x^*|\leqslant\frac{1}{2}\times 10^{m-n}=\varepsilon(x^*), \tag{1.5}$$

则由定义 1.5 知,近似值 $x^*$ 具有 $n$ 位有效数字 $a_1,a_2,\cdots,a_n$.在 $m$ 相同的情况下,$n$ 越大则 $10^{m-n}$ 越小,故有效数字的位数越多,绝对误差限越小,近似值的精度越高.

**例 1.4** 写出$\frac{1}{27}$的具有 1 位、2 位、3 位和 4 位有效数字的近似值.

**解**:由于$\frac{1}{27}=0.037037037\cdots$,则按照定义得到 1 位、2 位、3 位和 4 位有效数字的近似值分别为 0.04,0.037,0.0370 和 0.03704.

**注意**:0.037 与 0.0370 两者之间的差别,前者具有两位有效数字,其误差不超过$\frac{1}{2}\times 10^{-3}$;而后者具有三位有效数字,其误差不超过$\frac{1}{2}\times 10^{-4}$.

**例 1.5** 对重力加速度常数 $g$,若以 $\mathrm{m/s^2}$ 为单位,则 $g\approx 9.80\mathrm{m/s^2}$;若以 $\mathrm{km/s^2}$ 为单位,则 $g\approx 0.00980\mathrm{km/s^2}$.它们都具有 3 位有效数字,因为按第一种写法,有

$$|g-9.80|\leqslant\frac{1}{2}\times 10^{-2}(\mathrm{m/s^2}).$$

根据(1.5)式,这里 $m=1$,$n=3$;按第二种写法,有

$$|g-0.00980|\leqslant\frac{1}{2}\times 10^{-5}(\mathrm{km/s^2}).$$

这里 $m=-2,n=3$. 它们虽然写法不同,但都具有 3 位有效数字. 至于绝对误差限,由于单位不同结果也不同,$\varepsilon_1=\frac{1}{2}\times10^{-2}(\mathrm{m/s^2})$,$\varepsilon_2=\frac{1}{2}\times10^{-5}(\mathrm{km/s^2})$. 而相对误差限是相同的,因为

$$\varepsilon_r=\frac{\frac{1}{2}\times10^{-2}}{9.80}=\frac{\frac{1}{2}\times10^{-5}}{0.00980}.$$

例 1.5 表明有效数字的位数与小数点后有多少位数无关.

定义 1.5 给出了绝对误差限与有效数字之间的关系,下面给出相对误差限与有效数字之间的关系.

**定理 1.1** 设准确值 $x$ 的近似值 $x^*$ 表示为(1.4)式,如果 $x^*$ 具有 $n$ 位有效数字,则其相对误差限为

$$\varepsilon_r(x^*)=\frac{1}{2a_1}\times10^{-(n-1)}. \tag{1.6}$$

反之,若

$$\varepsilon_r(x^*)=\frac{1}{2(a_1+1)}\times10^{-(n-1)}, \tag{1.7}$$

则 $x^*$ 至少具有 $n$ 位有效数字.

**证明**:令准确值 $x$ 的近似值 $x^*=\pm0.a_1a_2\cdots a_k\cdots\times10^m$,其中,$m$ 为整数,$k\geqslant n$,$a_1,a_2,\cdots$均为 0~9 之间的数,且 $a_1\neq0$,则

$$a_1\times10^{m-1}\leqslant|x^*|\leqslant(a_1+1)\times10^{m-1}.$$

若近似值 $x^*$ 具有 $n$ 位有效数字,则有

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

于是

$$|e_r|=\frac{|x-x^*|}{|x^*|}\leqslant\frac{\frac{1}{2}\times10^{m-n}}{a_1\times10^{m-1}}=\frac{1}{2a_1}\times10^{-(n-1)}=\varepsilon_r(x^*),$$

即(1.6)式成立.

若近似值 $x^*$ 的相对误差限满足(1.7)式,即

$$\varepsilon_r(x^*)=\frac{1}{2(a_1+1)}\times10^{-(n-1)},$$

则依据相对误差与绝对误差之间的关系,有

$$|e|=|e_r||x^*|\leqslant\frac{1}{2(a_1+1)}\times10^{-(n-1)}\times(a_1+1)\times10^{m-1}=\frac{1}{2}\times10^{m-n}=\varepsilon(x^*),$$

故近似值 $x^*$ 至少具有 $n$ 位有效数字. □

综上所述,关于有效数字需注意如下几点:

(1) 凡是由准确值 $x$ 经过四舍五入得到的近似值 $x^*$,则从 $x^*$ 末位数字到左边第一位非零数字间的所有数字都是有效数字. 例如,经过四舍五入得到的近似值 0.0301400 有 6 位有效数字.

(2) 相对误差与相对误差限都是无量纲的,而绝对误差与绝对误差限都是有量纲的.

(3) 有效数字相同的两个近似值,其绝对误差不一定相同.例如,设近似值 $x_1=12345$ 和 $x_2=12.345$ 均有 5 位有效数字,由于 $|e(x_1)|\leqslant 0.5$, $|e(x_2)|\leqslant 0.0005$,故其绝对误差显然是不一样的.

(4) 由(1.5)式和(1.6)式得到:同一个准确值的不同近似值,有效数字的位数越多,其绝对误差、相对误差、绝对误差限和相对误差限都越小,即增加有效数字的位数可以减少舍入误差对计算的影响.

(5) 在有效数字的概念下,2.130012 的两个近似值 2.13 和 2.1300 的写法是有区别的,前者有 3 位有效数字,后者有 5 位有效数字.

(6) 把任何数乘以 $10^m$($m$ 为整数)不影响其有效数字的位数.

(7) 准确值的有效数字可看作有无限多位.

**例 1.6** 为使$\sqrt{200}$的近似值的相对误差不超过 0.1%,问要取几位有效数字.

**解**:由(1.7)式,只需求出满足

$$\frac{1}{2(a_1+1)}\times 10^{-(n-1)}\leqslant 0.1\%$$

的 $n$.显然,近似值的第一位有效数字为 $a_1=1$.因此,由$\frac{1}{4}\times 10^{-(n-1)}\leqslant 0.1\%$,可得 $n\geqslant 4$.

**例 1.7** 已知近似值 $x^*$ 有两位有效数字,试求其相对误差限.

**解**:由(1.6)式可知,$|e_r(x^*)|\leqslant\frac{1}{2a_1}\times 10^{-(n-1)}=\frac{1}{2a_1}\times 10^{-1}$.

当 $a_1=1$ 时,有 $|e_r(x^*)|\leqslant\frac{1}{2a_1}\times 10^{-(n-1)}=\frac{1}{2}\times 10^{-1}=0.05$;

当 $a_1=9$ 时,有 $|e_r(x^*)|\leqslant\frac{1}{2a_1}\times 10^{-(n-1)}=\frac{1}{18}\times 10^{-1}=0.0055\cdots\leqslant 0.05$;

因此,保守估计 $\varepsilon_r(x^*)=0.05$.

### 1.2.4 数值计算中的误差传播

数值计算中误差产生与传播的情况非常复杂,参与运算的数据往往都是近似值,它们都有误差.这些数据的误差在多次运算中又会进行传播,使计算结果产生一定的误差,这就是误差的**传播问题**.下面介绍利用函数的泰勒公式来估计误差的一种常用方法.

设一元函数 $f(x)$ 具有二阶导数,自变量 $x$ 的一个近似值 $x^*$,$f(x)$ 的近似值为 $f(x^*)$.用 $f(x)$ 在 $x^*$ 处的泰勒展开估计误差,可得

$$|f(x)-f(x^*)|\leqslant|f'(x^*)(x-x^*)|+\frac{1}{2}|f''(\xi)(x-x^*)^2|,\tag{1.8}$$

其中 $\xi$ 在 $x$ 和 $x^*$ 之间.如果 $f'(x^*)\neq 0$,且比值 $|f''(\xi)|/|f'(x^*)|$ 较小,而 $|x-x^*|$ 很小,则 $f(x^*)$ 的近似误差限与近似相对误差限分别为

$$\varepsilon(f(x^*))\approx|f'(x^*)|\varepsilon(x^*),\quad \varepsilon_r(f(x^*))\approx\left|\frac{f'(x^*)}{f(x^*)}\right|\varepsilon(x^*).\tag{1.9}$$

如果 $f'(x^*)=f''(x^*)=\cdots=f^{(k-1)}(x^*)=0$,$f^{(k)}(x^*)\neq 0$,且比值

$$|f^{(k+1)}(\xi)|/|f^{(k)}(x^*)|$$

较小，则 $f(x^*)$ 的近似误差限与近似相对误差限分别为

$$\varepsilon(f(x^*)) \approx \frac{|f^{(k)}(x^*)|}{k!}(\varepsilon(x^*))^k,$$

$$\varepsilon_r(f(x^*)) \approx \frac{1}{k!}\left|\frac{f^{(k)}(x^*)}{f(x^*)}\right|(\varepsilon(x^*))^k. \tag{1.10}$$

如果 $f$ 为多元函数，自变量为 $x_1,x_2,\cdots,x_n$，其近似值分别为 $x_1^*,x_2^*,\cdots,x_n^*$，类似于一元函数，可用多元函数 $f(x_1,x_2,\cdots,x_n)$ 在点 $(x_1^*,x_2^*,\cdots,x_n^*)$ 的泰勒展开，取一个近似的绝对误差限为

$$\varepsilon(f(x_1^*,x_2^*,\cdots,x_n^*)) \approx \sum_{i=1}^{n}\left|\frac{\partial f(x_1^*,x_2^*,\cdots,x_n^*)}{\partial x_i}\right|\varepsilon(x_i^*), \tag{1.11}$$

相对误差限为

$$\begin{aligned}\varepsilon_r(f(x_1^*,x_2^*,\cdots,x_n^*)) &\approx \sum_{i=1}^{n}\left|\frac{\partial f(x_1^*,x_2^*,\cdots,x_n^*)}{\partial x_i}\right|\frac{\varepsilon(x_i^*)}{|f(x_1^*,x_2^*,\cdots,x_n^*)|}\\ &\approx \sum_{i=1}^{n}\left|\frac{\partial f(x_1^*,x_2^*,\cdots,x_n^*)}{\partial x_i}\right|\frac{|x_i^*|}{|f(x_1^*,x_2^*,\cdots,x_n^*)|}\varepsilon_r(x_i^*),\end{aligned} \tag{1.12}$$

其中

$$\left|\frac{\partial f(x_1^*,x_2^*,\cdots,x_n^*)}{\partial x_i}\right| \quad 和 \quad \left|\frac{\partial f(x_1^*,x_2^*,\cdots,x_n^*)}{\partial x_i}\right|\frac{|x_i^*|}{|f(x_1^*,x_2^*,\cdots,x_n^*)|}$$

分别为各个近似值 $x_i^*(i=1,2,\cdots,n)$ 对函数近似值 $f(x_1^*,x_2^*,\cdots,x_n^*)$ 的绝对误差限和相对误差限的增长因子，即绝对误差限 $\varepsilon(x_i^*)$ 和相对误差限 $\varepsilon_r(x_i^*)$ 经过传播后增大或缩小的倍数. 如果

$$\left|\frac{\partial f(x_1^*,x_2^*,\cdots,x_n^*)}{\partial x_i}\right| = 0, \quad i=1,2,\cdots,n,$$

则要使用泰勒公式中的高阶项.

设 $x^*$ 和 $y^*$ 分别是准确值 $x$ 和 $y$ 的近似值，利用(1.11)式可得 $x^*$ 和 $y^*$ 的四则运算绝对误差估计

$$\varepsilon(x^* \pm y^*) \approx \varepsilon(x^*) + \varepsilon(y^*), \tag{1.13}$$

$$\varepsilon(x^* y^*) \approx |x^*|\varepsilon(y^*) + |y^*|\varepsilon(x^*), \tag{1.14}$$

$$\varepsilon\left(\frac{x^*}{y^*}\right) \approx \frac{|x^*|\varepsilon(y^*) + |y^*|\varepsilon(x^*)}{|y^*|^2}, \quad y^* \neq 0. \tag{1.15}$$

**例 1.8** 设测得某矩形场地长 $a$ 的值 $a^*=120\text{m}$，宽 $b$ 的值 $b^*=80\text{m}$，并且已知 $\varepsilon(a^*)=0.2\text{m}$，$\varepsilon(b^*)=0.1\text{m}$，试计算该矩形场地的面积 $S=ab$ 的绝对误差限和相对误差限.

**解**：因为 $S=ab$，$\dfrac{\partial S}{\partial a}=b$，$\dfrac{\partial S}{\partial b}=a$，故由(1.14)式知，绝对误差限为

$$\varepsilon(S^*) \approx |b^*|\varepsilon(a^*) + |a^*|\varepsilon(b^*) = 80\times 0.2 + 120 \times 0.1 = 28(\text{m}^2).$$

相对误差限为

$$\varepsilon_r(S^*) \approx \frac{\varepsilon(S^*)}{|S^*|} = \frac{28}{120\times 80} \approx 0.2917\%.$$