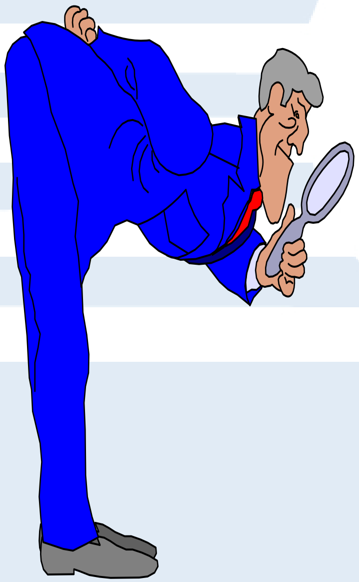
1.2 误差

1.2 误差 /* Error */

学习目标

掌握误差和有效数字、以及算法的数值稳定性等概念；重点是有效数字与相对误差的关系。

1.2.1 误差的来源与分类/* Source & Classification */



误差在我们的日常生活中无处不在，无处不有，如在做热力学实验中，从温度计上读出的温度是23.4度，就不是一个精确的值，而是含有误差的近似值。又如量体裁衣，量与裁的结果都不是精确无误的，都含有误差。

- 从实际问题中抽象出数学模型

  —— 模型误差 /* Modeling Error */

- 通过测量得到模型中参数的值

  —— 观测误差 /* Measurement Error */

- 求近似解 —— 方法误差 (截断误差 Truncation Error)

- 机器字长有限 —— 舍入误差 /* Roundoff Error */

(1). 模型误差

- 用数学方法解决一个具体的实际问题，首先要建立数学模型，这就要对实际问题进行抽象、简化，因而数学模型本身总含有误差，这种误差叫做模型误差
- 数学模型是指那些利用数学语言模拟现实而建立起来的有关量的描述
- 数学模型的准确解与实际问题的真解不同

实际问题的真解 → 数学模型的真解

为减化模型忽略次要因素

定理在特定条件下建立

与实际条件有别

(2). 观测误差

- 在数学模型中通常包含各种各样的参变量，如温度、长度、电压等，这些参数往往是通过观测得到的，因此也带来了误差，这种误差叫观测误差.
- 数学模型中的参数和原始数据，是由观测和试验得到的.
- 由于测量工具的精度、观测方法或客观条件的限制,使数据含有测量误差,这类误差叫做观测误差或数据误差.
- 根据实际情况可以得到误差上下界.
- 数值方法中需要了解观测误差,以便选择合理的数值方法与之适应.

数值运算的一个特点是：

参与运算的数必须是有限位的，而且位数往往是预先规定的（如在计算机高级语言中，单精度实数为6~7位有效数字）。如果运算的数是无限位的或超过规定，那么要用“四舍五入”规则或“截断”规则，将它们处理成规定的位数。

所谓“截断”规则就是：将超过规定位数的部分无条件地去掉。这样 π取4 位小数，就为3.1415。

(3). 截断误差

- 精确公式用近似公式代替时，所产生的误差叫截断误差.

例如， 函数$f(x)$用泰勒(Taylor)多项式

$$p_n(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\cdots+\frac{f^{(n)}(0)}{n!}x^n$$

近似代替，则数值方法的截断误差是

$$R_n(x)=f(x)-p_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1}$$

（ξ介于0与x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，是数值计算中必须考虑的一类误差.

例如，对函数

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \ldots,$$

当 $|x|$ 较小时，我们若用前三项作为 $\sin x$ 的近似值，则截断误差的绝对值不超过 $\frac{|x|^7}{7!}$.

有的计算机是采用"截断"规则的，但大多数计算机是采用"四舍五入"规则处理舍弃位数的。

(4). 舍入误差

- 在数值计算中只能对有限位字长的数值进行运算.
- 需要对参数、中间结果、最终结果作有限位字长的处理工作，这种处理工作称作舍入处理.
- 用有限位数字代替精确数，这种误差叫做舍入误差，是数值计算中必须考虑的一类误差.

“四舍五入”规则:四舍六入五成双

所谓"四舍五入"规则就是：将超过规定位数的部分按下列原则去掉：

(1)如果舍弃的部分小于保留数的最后一位的单位的1/2，那么保留的数不变。

例如π=3.1415926…，如果取两位小数，那么保留数的最后一位单位是10^{-2}，舍弃部分是0.15926×10^{-2}，小于0.5×10^{-2}，因此取为3.14;

(2)如果舍弃的部分大于所保留数的最后一位单位的1/2，那么将保留数最后一位数字加1。

例如限制π取4位小数，最后一位单位为10^{-4}，但去掉的部分是0.926×10^{-4}，大于0.5×10^{-4}，因此取成3.1416;

(3)如果舍弃的部分恰等于所保留数的最后一位单位的1/2，此时如果保留的数最后一位是奇数，那么加1成偶数；如果保留的数最后一位是偶数，则就不动了。

例如：取2位小数，0.675成0.68，而0.605成0.60。

上述种种误差都会影响计算结果的准确性，因此需要了解与研究误差，在数值计算中将着重研究截断误差、舍入误差，并对它们的传播与积累作出分析.

1.2.2 误差与有效数字 (Error and Significant Digits)

➤ 绝对误差 /* absolute error */ 绝对误差界（限）

定义 1.1 设 x 是某实数的精确值，x_A 是它的一个近似值，则称 $x - x_A$ 为近似值 x_A 的绝对误差.（$\mathbf{x_A}$有时也可记作$\boldsymbol{x}$*）

由于精确值一般是未知的,因而绝对误差不能求出来，但可以根据测量误差或计算情况设法估计出它的取值范围，即误差绝对值的一个上界或称误差限。

定义1.2 设 x 是某实值的精确值，x_A 是它的一个近似值，并可对 x_A 的绝对误差作估计 $|\mathbf{x} - \mathbf{x_A}| \le \boldsymbol{\varepsilon_A}$，则称 ε_A 是 x_A 的绝对误差界(限)。

例1 设π=3.1415926… 近似值π_A=3.14,它的绝对误差是0.0015926…，有

$$|\pi-\pi_A|=0.0015926\ldots\le 0.002=0.2\times 10^{-2}$$

例2 又近似值π_A =3.1416，它的绝对误差是0.0000074…，有

$$|\pi-\pi_A|=0.0000074\ldots\le 0.000008=0.8\times 10^{-5}$$

例3 而近似值π_A =3.1415，它的绝对误差是0.0000926…，有

$$|\pi-\pi_A|==0.0000926\ldots\le 0.0001=0.1\times 10^{-3}$$

可见，绝对误差限ε_A不是唯一的，但ε_A越小越好，绝对误差限都不超过末尾数字的半个单位。

➢相对误差 /* relative error */ 相对误差界（限）

只用绝对误差还不能说明数的近似程度，例如甲打字每100个错一个，乙打字每1000个错一个，他们的误差都是错一个，但显然乙要准确些，这就启发我们除了要看绝对误差外，还必须顾及量的本身。

定义1.3 绝对误差与精确值x的比值 $\dfrac{x-x_A}{x}$ 称为x_A的相对误差。

若 $\left|\dfrac{x-x_A}{x}\right| \le \varepsilon_R$，称$\boldsymbol{\varepsilon_R}$为 x_A 的相对误差限。

当 $\boldsymbol{x=0}$ 时，相对误差没有意义。在实际计算中，精确值 x 往往是不知道的，所以通常把 $(x-x_A)/x_A$ 作为x_A的相对误差。

例4 设 $x^1=1.234$，$x^2=0.002$，$x_A^1=1.233$，$x_A^2=0.001$，估计近似数x^1、x^2的绝对误差与相对误差。

解 $\left|x^1-x_A^1\right|=10^{-3}$，$\left|x^2-x_A^2\right|=10^{-3}$，

但 x_A^1 是 x^1 的一个好的近似，x_A^2 不是 x^2 的好的近似.

$$\left|\varepsilon_R(x^1)\right|=\left|\frac{10^{-3}}{1.234}\right|\approx 0.81\%,\quad \left|\varepsilon_R(x^2)\right|=\left|\frac{10^{-3}}{0.002}\right|=50\%.$$

结论?

俗称“好坏”、“多少”是相对的

结论 近似数的相对误差是近似数精确度的基本度量，一个近似数 x_A 的相对误差越小，则近似数越精确。

- 绝对误差 $\dfrac{x-x_A}{x}$ 是一个无量纲的数

x一般是未知的，故绝对误差难求。

- 通常将 $\dfrac{x-x_A}{x_A}$ 作为 x_A 的相对误差。

例5　测量一木板长是954cm，问测量的相对误差界是是多大？

解　因为实际问题中所截取的近似数，其绝对误差界一般不超过最小刻度的半个单位，所以当 $x = 954\,cm$ 时，有 $\varepsilon_A = 0.5cm$，其相对误差界为

$$\frac{\varepsilon_A}{|x|} = \frac{0.5}{954} = 0.0005241 \cdots < 0.53\%$$

➢有效数字（significant digits）

1.定义：如果绝对误差限 $|x - x_A| \le 0.5\times 10^{-n}$，则称近似数 x_A 准确到了 n 位小数，该数位到第一个非零数字的所有数位叫做该近似数的有效数位，有效数位上的数字叫做有效数字。

例6 $\pi = 3.1415926535\ \ 897932\cdots\cdots;\quad \pi^* = 3.1415$

问：π^* 有几位有效数字？请证明你的结论。

证明:

$$\because |\pi^* - \pi| = 0.0000926\ldots = 0.926\ldots\times 10^{-4}$$

$$< 0.5\times 10^{-4} = 0.5\times 10^{1-5}$$

$\therefore \pi^*$ 精确到小数点后第 4 位，有5位有效数字.

2.有效数字的等价定义

用十进制科学计数法，记

$$x=\pm(0.a_1\cdots a_n a_{n+1}\cdots)\times 10^k \quad (a_1\neq 0,\ k\text{是整数}),\quad x_A\text{是}\,x\,\text{的}\,a_{n+1}$$

4舍5入得到的近似数，如果

$$|x-x_A|\leq\frac{1}{2}\times 10^{k-n}。$$

则称 x_A 为 x 的具有 n 位有效数字的近似值。

通常在 x 的准确值已知的情况下，若要取有限位数的数字作为近似值，就采用四舍五入得到的近似值，其绝对误差界可以取被保留的最后数位上的半个单位。

例7. 3.142作为 π 的近似值时有几位有效数字

解： $3.141592\ldots = 0.3141592\ldots \times 10^{1}$

$$3.142 = 0.3142 \times 10^{1}$$

$$k = 1$$

$$|\pi - 3.142| = |0.3141592\ldots \times 10^{1} - 0.3142 \times 10^{1}|$$

$$< 0.000041 \times 10^{1} < 0.0005 = \frac{1}{2} \times 10^{-3}$$

$$k - n = 1 - n = -3$$

所以 $n=4$，具有4位有效数字

练习 设 $\pi = 3.1415926 \cdots$

k-n=-2,即 n =3 , 3位有效数字,

$\pi_A = 3.14 = \mathbf{0.314 \times 10^1}$, $k = 1$,

$|\pi - \pi_A| = 0.0015926\cdots$

$= 0.15926\cdots \times 10^{-2} \le \frac{1}{2} \times 10^{-2}$

$\pi_{A_2} = 3.1416 = \mathbf{0.3416 \times 10^1}$, $k = 1$,

k-n=-4 , 即 n =5 , 5位有效数字

$$|\pi - \pi_{A_2}| = 0.0000074\cdots = 0.074\cdots \times 10^{-4} \le \frac{1}{2} \times 10^{-4}$$

有效数字的位数不能仅考虑 $|x - x_A| \le \frac{1}{2} \times 10^{k-n}$，还要看 x_A本身

例8 设 $x=8.000033$，考虑 $x_{A_1}=8.0000$，$x_{A_2}=8$

5位有效数字，即 n=5

1位有效数字，即 n =1

最多有5位有效数字

但

$$\text{尽管} \left|x-x_{A_1}\right|=\left|x-x_{A_2}\right|=0.000033=0.33\times10^{-4}\le\frac{1}{2}\times10^{-4},$$

$$\text{因为} \quad \left|x-x_{A_2}\right|=0.000033=0.000033\times10^{1-1}\le\frac{1}{2}\times10^{1-1},$$

显然，近似值的有效数字位数越多，相对误差越小，反之也对。下面，我们给出相对误差界与有效数字的关系。

3. 有效数字与相对误差之间的关系

$x_A = \pm 10^k \times 0.a_1 a_2 \cdots a_i \cdots$

定理 设x 的近似值x_A 有（**1.2.1**）的表达式。

（**1**）如果 x_A 有 n 位有效数字，则

$$\varepsilon_R = \frac{|x - x_A|}{|x_A|} \le \frac{1}{2a_1} \times 10^{1-n}; \qquad (1.2.2)$$

有效数字的位数估计相对误差限

有效数字的位数越多，相对误差限就越小

（**2**）如果

$$\varepsilon_R = \frac{|x - x_A|}{|x_A|} \le \frac{1}{2(a_1+1)} \times 10^{1-n}, \qquad (1.2.3)$$

则x_A 至少具 n 位有效数字。

相对误差限越小，
有效数字的位数就越多

相对误差限估计
有效数字的位数

$$x_A = \pm 10^k \times 0.a_1 a_2 \cdots a_i \cdots$$

定理　设 x 的近似值 x_A 有（**1.2.1**）的表达式。

（**1**）如果 x_A 有 n 位有效数字，则

$$\varepsilon_R = \frac{|x - x_A|}{|x_A|} \le \frac{1}{2a_1} \times 10^{1-n}; \qquad (1.2.2)$$

证　由（**1.2.1**）可得到

$$a_1 \times 10^{k-1} \le |x_A| \le (a_1 + 1) \times 10^{k-1}。 \qquad (1.2.4)$$

所以，当 x_A 有 n 位有效数字时，

$$\frac{|x - x_A|}{|x_A|} \le \frac{0.5 \times 10^{k-n}}{a_1 \times 10^{k-1}} = \frac{1}{2a_1} \times 10^{1-n},$$

即（**1.2.2**）得证。

$$x_A = \pm 10^k \times 0.a_1 a_2 \cdots a_i \cdots$$

定理　设 x 的近似值 x_A 有（1.2.1）的表达式。

（2）如果

$$\varepsilon_R = \frac{|x - x_A|}{|x_A|} \le \frac{1}{2(a_1+1)} \times 10^{1-n}, \tag{1.2.3}$$

则 x_A 至少具 n 位有效数字。

证

由（1.2.3）和（1.2.4）有

$$|x - x_A| = |x_A| \frac{|x - x_A|}{|x_A|} \le (a_1+1) \times 10^{k-1} \frac{1}{2(a_1+1)} \times 10^{1-n} = 0.5 \times 10^{k-n},$$

即说明 x_A 有 n 位有效数字，（2）得证。

例9 取3.14作为π的四舍五入的近似值时，求其相对误差。

解：$3.14=0.314\times10^{1}$　　$a_1=3$　$k=1$

$\because$ 四舍五入的近似值,其各位都是有效数字

$\therefore$ n=3

$$\varepsilon_R=(1/2a_1)\times10^{-(n-1)}$$
$$=(1/2*3)\times10^{-2}=17\%$$

例10 已知近似数x_A有两位有效数字，试求其相对误差限.

解：$3.14=0.314\times10^1$ $a_1=3$ $m=1$

$\because$ 四舍五入的近似值,其各位都是有效数字

$\therefore$ n=3

$$\varepsilon_R=(1/2a_1)\times10^{-(n-1)}=(1/2*3)\times10^{-2}=17\%$$

x_A的第一位有效数字a_1没有给出，可进行如下讨论:

当 $a_1=1$ $\varepsilon_R=1/2a_1\times10^{-1}=1/2*1\times10^{-1}=5\%$

$a_1=9$ $\varepsilon_R=1/2a_1\times10^{-1}=1/2*9\times10^{-1}=0.56\%$

取 $a_1=1$ 时相对误差为最大，即 5%

例11　　已知近似数x_A的相对误差界为0.3%，问x_A至少有几位有效数字？

解　设x_A有n位有效数字，由于x_A的第一个有效数a_1没有具体给定，而我们知道a_1一定是1，2，···，9中的一个，由

$$\frac{|x-x_A|}{|x_A|}\le\frac{3}{1000}<\frac{1}{2\times10^2}=\frac{1}{2(9+1)}\times10^{-1}$$

故由（1.2.3）式知$n=2$，即x_A至少有2位有效数字。

注意:

已知有效数字, 求相对误差用公式

$$\varepsilon_R \le \frac{1}{2a_1} \times 10^{-(n-1)}$$

已知相对误差, 求具有几位有效数字公式

$$\varepsilon_R \le \frac{1}{2(x_1+1)} \times 10^{-(n-1)}$$

1.2.3 函数求值的误差估计

对一元函数 $f(x)$，自变量 x 的一个近似值为 x_A，以 $f(x_A)$ 近似 $f(x)$，其误差界记作 $\varepsilon(f(x_A))$。

若 $f(x)$ 具有2阶连续导数，由*Taylor*展开式

$$f(x)-f(x_A)$$

$$=f'(x_A)(x-x_A)+\frac{f''(\xi)}{2!}(x-x_A)^2$$

$$(\xi \text{介于} x, x_A \text{之间})$$

取绝对值得

$$|f(x)-f(x_A)| \le |f'(x_A)|\varepsilon(x_A)+\frac{f''(\xi)}{2!}\varepsilon^2(x_A)$$

假定 $f'(x_A)$ 与 $f''(x_A)$ 的比值不太大，可忽略 $\varepsilon(x_A)$ 的高阶项，可得计算函数的误差限

$$\varepsilon(f(x_A)) \approx |f'(x_A)|\varepsilon(x_A)$$

对n元函数 $f(x_1, x_2, \ldots, x_n)$，自变量 $x_1, x_2, \ldots x_n$ 的近似值分别为 $x_{1A}, x_{2A}, \cdots, x_{nA}$，则有

$$f(x_1, x_2, \ldots, x_n) - f(x_{1A}, x_{2A}, \ldots, x_{nA}) \approx \sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}\right)_A (x_k - x_{kA}),$$

其中 $\left(\frac{\partial f}{\partial x_k}\right)_A = \frac{\partial}{\partial x_k} f(x_{1A}, x_{2A}, \ldots, x_{nA})$。可以得到函数值的一个近似误差界：

$$\varepsilon(f(x_{1A}, x_{2A}, \ldots, x_{nA})) \approx \sum_{k=1}^{n} \left| \left(\frac{\partial f}{\partial x_k}\right)_A \right| \varepsilon(x_{kA})。$$

特别地，对 $f(x_1, x_2) = x_1 \pm x_2$ 有

$$\varepsilon(x_{1A} \pm x_{2A}) = \varepsilon(x_{1A}) + \varepsilon(x_{2A})。$$

同样，可以得到

$$\varepsilon(x_{1A} x_{2A}) \approx |x_{1A}| \varepsilon(x_{2A}) + |x_{2A}| \varepsilon(x_{1A}),$$

$$\varepsilon\left(\frac{x_{1A}}{x_{2A}}\right) \approx \frac{|x_{1A}| \varepsilon(x_{2A}) + |x_{2A}| \varepsilon(x_{1A})}{|x_{2A}|^2}, \quad x_{2A} \neq 0。$$

例13　设有长为 l ,宽为 d 的某场地。现测得 l 的近似值 $l_A=120$M,d 的近似值 d_A =90M，并已知它们的差界为 $|l-l_A|\le 0.2m, |d-d_A|\le 0.2m.$ 试估计该场地面积 $s = ld$ 的误差界和相对误差界。

解　这里 $\varepsilon(l_A)=0.2$，$\varepsilon(d_A)=0.2$，并且有

$$\frac{\partial S}{\partial l}=d,\ \frac{\partial S}{\partial d}=l，\ S_A=l_Ad_A=10800m^2。$$

于是有误差界

$$\varepsilon(S_A)\approx 120\times 0.2+90\times 0.2=42m^2$$

相对误差界

$$\varepsilon_r(S_A)=\frac{\varepsilon(S_A)}{l_Ad_A}\approx\frac{42}{10800}=0.39\%。$$

例14 设有三个近似数

$$a = 2.31,\quad b = 1.93,\quad c = 2.24,$$

它们都有三位有效数字。试计算 $p = a + bc$ 的误差界，并问 p 的计算结果能有几位有效数字？

解 $p = 2.31 + 1.93 \times 2.24 = 6.6332$。 于是有误差界

$$\begin{aligned}\varepsilon(p) &= \varepsilon(a) + \varepsilon(bc)\\ &\approx \varepsilon(a) + |b|\varepsilon(c) + |c|\varepsilon(b)\\ &= 0.005 + 0.005(1.93 + 2.24) = 0.02585\end{aligned}$$

相对误差界

$$\varepsilon_r(p) = \frac{\varepsilon(p)}{|p|} \approx \frac{0.02586}{6.6332} \approx 0.39\%。$$

因为$\varepsilon(p) \approx 0.02585 < 0.05$，所以$p = 6.6332$ 能有两位有效数字。

1.2.4 计算机中数的表示和舍入误差

$x_A = \pm 10^k \times 0.a_1 a_2 \cdots a_i \cdots$

任意一个非零实数用（1.2.1）表示，是规格化的十进制科学记数方法。在计算机中通常采用二进制的数系（或其变形的十六进制等），并且表示成与十进制类似的规格化形式，即浮点形式

$$\pm 2^m \times 0.\beta_1 \beta_2 \ldots \beta_t,$$

这里整数m称为阶码，用二进制表示为 $m = \pm a_1 a_2 \ldots a_s$，$a_j = 0$ 或 1（$j = 1,2,\ldots$，s），S是阶的位数。小数 $0.\beta_1 \beta_2 \ldots \beta_t$ 称为尾数，其中 $\beta_1 = 1$，$\beta_j = 0$，或 $1(j = 2,3,\cdots,t)$，t是尾数部位的位数。S和t与具体的机器有关。

由于计算机的字长总是有限位的，所以计算机所能表示的数系是一个特殊的离散集合，此集合的数称为机器数。

十进制输入计算机时转换成二进制，并对 t 位后面的数做舍入处理，使得尾数为 t 位，因此一般都有舍入误差。两个二进制数作算术运算时，对计算结果也要作类似的舍入处理，使得尾数为 t 位，从而也有舍入误差。

在实现计算时，计算的最后结果与计算的精确解之间的误差，从根本上说是由机器的舍入误差造成的，包括输入数据和算术运算的舍入误差。因此有必要对计算机数的浮点表示方法和舍入误差有一个初步的了解。有时为了分析某一个计算方法可能出现的误差现象，为了适应人们的习惯，我们会采用十进制实数系统进行误差分析。