第一章　误差理论

在研究算法的同时，必须注重误差分析，使建立起来的算法科学有效.

一、内容要点

1．误差

（1）误差的来源

（2）绝对误差、相对误差和有效数字

①绝对误差

假设某一量的准确值为 x，其近似值为 x^*，则称 $\varepsilon(x)=x-x^*$ 为近似数 x^*的绝对误差.

②绝对误差限

如果 $|\varepsilon(x)|=|x-x^*|\le\eta$，则称 η 为近似值 x^*的绝对误差限.

③相对误差

称 $\varepsilon_r(x)=\dfrac{\varepsilon(x)}{x}=\dfrac{x-x^*}{x}$ 为近似值 x^*的相对误差.

在实际问题中常取 $\varepsilon_r^*(x)=\dfrac{\varepsilon(x)}{x^*}=\dfrac{x-x^*}{x^*}$ 为近似值 x^*的相对误差.

④相对误差限

如果 $|\varepsilon_r(x)|\le\delta$，则称 δ 为近似值 x^*的相对误差限.

⑤有效数字

设准确值 x 的近似值 x^*可表示为

$$x^*=\pm(a_1+a_2\times10^{-1}+a_3\times10^{-2}+\cdots+a_n\times10^{-n+1})\times10^m$$

其中 m 是整数，$a_1,a_2,\cdots,a_n$ 是 0 到 9 之间的一个数字，且 $a_1\ne0$，如果 $|x-x^*|\le\frac{1}{2}\times10^{m-n+1}$，则称近似值 x^*具有 n 位有效数字.

⑥有效数字与相对误差的联系

定理 1　若近似数 $x^*=\pm(a_1+a_2\times10^{-1}+a_3\times10^{-2}+\cdots+a_n\times10^{-n+1})\times10^m$ 具有 n 位有效数字，则其相对误差满足

$$\left|\varepsilon_r^*(x)\right| \le \frac{1}{2a_1}\times 10^{-n+1}$$

定理 2　若近似数 $x^* = \pm(a_1 + a_2\times 10^{-1} + a_3\times 10^{-2} + \cdots + a_n\times 10^{-n+1})\times 10^m$ 的相对误差满足

$$\left|\varepsilon_r^*(x)\right| \le \frac{1}{2(a_1+1)}\times 10^{-n+1}$$

则 x^*具有 n 位有效数字.

2．数值计算中应注意的一些问题

（1）要使用数值稳定的算法；

（2）要避免两个相近的数相减；

（3）要避免除数的绝对值远小于被除数的绝对值；

（4）要防止大数“吃掉”小数的现象；

（5）注意简化运算步骤，减少运算次数.

二、典型分析与解题方法

例 1　问 3.142，3.141，$\dfrac{22}{7}$ 分别作为 π 的近似值各具有几位有效数字？

分析　利用有效数字的概念可直接得出.

解　$\pi = 3.141\,592\,65\cdots$

记 $x_1 = 3.142,\quad x_2 = 3.141,\quad x_3 = \dfrac{22}{7}$

由　$\pi - x_1 = 3.141\,59\cdots - 3.142 = -0.000\,40\cdots$　知

$$\frac{1}{2}\times 10^{-4} < \left|\pi - x_1\right| \le \frac{1}{2}\times 10^{-3}$$

因而 x_1 具有 4 位有效数字

由　$\pi - x_2 = 3.141\,59\cdots - 3.141 = 0.000\,59\cdots$　知

$$\frac{1}{2}\times 10^{-3} < \left|\pi - x_2\right| \le \frac{1}{2}\times 10^{-2}$$

因而 x_2 具有 3 位有效数字.

由　$\pi - \dfrac{22}{7} = 3.141\,59\cdots - 3.142\,85\cdots = -0.001\,26\cdots$　知

$$\frac{1}{2}\times 10^{-3}<\left|\pi-\frac{22}{7}\right|\le\frac{1}{2}\times 10^{-2}$$

因而 x_3 具有 3 位有效数字.

例 2 已知近似数 x^*有两位有效数字，试求其相对误差限.

分析 本题显然应利用有效数字与相对误差的关系.

解 利用有效数字与相对误差的关系.这里 n=2，a_1 是 1 到 9 之间的数字.

$$\left|\varepsilon_r^*(x)\right|=\frac{\left|x-x^*\right|}{\left|x^*\right|}\le\frac{1}{2a_1}\times 10^{-n+1}\le\frac{1}{2\times 1}\times 10^{-2+1}=5\%$$

例 3 已知近似数的相对误差限为 0.3%，问 x^*至少有几位有效数字？

分析 利用有效数字与相对误差的关系.

解 a_1 是 1 到 9 间的数字.

$$\varepsilon_r^*(x)=0.3\%=\frac{3}{1000}<\frac{1}{2\times 10^2}=\frac{1}{2\times(9+1)}\times 10^{-1}\le\frac{1}{2(a_1+1)}\times 10^{-1}$$

设 x^*具有 n 位有效数字，令 $-n+1=-1$，则 n=2，从而 x^*至少具有 2 位有效数字.

例 4 计算 sin1.2，问要取几位有效数字才能保证相对误差限不大于 0.01%.

分析 本题应利用有效数字与相对误差的关系.

解 设取 n 位有效数字.由 sin1.2=0.93…，故 a_1=9

$$\varepsilon_r^*(x)=\frac{\left|x-x^*\right|}{\left|x^*\right|}\le\frac{1}{2a_1}\times 10^{-n+1}\le 0.01\%=10^{-4}$$

解不等式 $\frac{1}{2a_1}\times 10^{-n+1}\le 10^{-4}$ 知取 n=4 即可满足要求.

例 5 计算 $\frac{1}{759}-\frac{1}{760}$，视已知数为精确值，用 4 位浮点数计算.

解 $\frac{1}{759}-\frac{1}{760}=0.1318\times 10^{-2}-0.1316\times 10^{-2}=0.2\times 10^{-5}$

结果只和一位有效数字，有效数字大量损失，造成相对误差的扩大，若通分后再计算：

$$\frac{1}{759}-\frac{1}{770}=\frac{1}{759\times 760}=\frac{1}{0.5768\times 10^6}=0.1734\times 10^{-5}$$

就得到 4 位有效数字的结果.

此例说明，在数值计算中，要特别注意两相近数作减法运算时，有效数字常会严重损失，遇到这种情况，一般采取两种办法：第一，应多保留几位有效数字；第二，将算式恒等变形，然后再进行计算.例如，当 x 接近于 0，计算 $\frac{1-\cos x}{\sin x}$ 时，应先把算式变形为

$$\frac{1-\cos x}{\sin x}=\frac{1-\cos^2 x}{\sin x(1+\cos x)}=\frac{\sin x}{1+\cos x}$$ 再计算.

又例如，当 x 充分大时，应作变换

$$\sqrt{1+x}-\sqrt{x}=\frac{1}{\sqrt{1+x}+\sqrt{x}}$$

$$\frac{1}{x}-\frac{1}{x+1}=\frac{1}{x(x+1)}$$

例 6　计算 $a=(\sqrt{2}-1)^6$，取 $\sqrt{2}\approx 1.4$，采用下列等式计算：

（1）$\frac{1}{(\sqrt{2}+1)^6}$；

（2）$99-70\sqrt{2}$；

（3）$(3-2\sqrt{2})^3$；

（4）$\frac{1}{(3+2\sqrt{2})^3}$.

问哪一个得到的结果最好？

解　显然 $a=(\sqrt{2}-1)^6=\frac{(\sqrt{2}-1)^6(\sqrt{2}+1)^6}{(\sqrt{2}+1)^6}=\frac{1}{(\sqrt{2}+1)^6}$

$$(\sqrt{2}-1)^6=[(\sqrt{2}-1)^2]^3=(3-2\sqrt{2})^3=99-70\sqrt{2}$$

$$(\sqrt{2}-1)^6=\frac{1}{(\sqrt{2}+1)^6}=\frac{1}{[(\sqrt{2}+1)^2]^3}=\frac{1}{(3+2\sqrt{2})^3}$$

所以 $(1)\equiv(2)\equiv(3)\equiv(4)$，这 4 个算式是恒等的，但当取 $\sqrt{2}\approx 1.4$ 计算时，因为（2），（3）都涉及到两个相近数相减，使有效数字丢失，而（1）在分母算式上的乘

幂数比算式（4）大，所以算式（4）最好，事实上，当取 $\sqrt{2} \approx 1.4$ 时，有 $|\Delta x| < 0.015$，

再由 $f(x)$ 的误差 $|f(x+\Delta x) - f(x)| \approx |f'(1.4)||\Delta x|$ 也可直接估计出每个算式的误差，

显然，算式（4）误差最小.

具体计算可得：

（1）$\dfrac{1}{(\sqrt{2}+1)^6} \approx 5.2 \times 10^{-3}$；

（2）$99 - 70\sqrt{2} \approx 1.0$；

（3）$(3-2\sqrt{2})^3 \approx 8.0 \times 10^{-3}$；

（4）$\dfrac{1}{(3+2\sqrt{2})^3} \approx 5.1 \times 10^{-3}$.

比较可得用第（4）个算式所得的结果更接近于 a.

例 7　求二次方程 $x^2 - (10^9+1)x + 10^9 = 0$ 的根.

解　由于 $x^2 - (10^9+1)x + 10^9 = (x-10^9)(x-1)$，所以方程的两个根分别为

$$x_1 = 10^9,\quad x_2 = 1.$$

但如果应用一般二次方程 $ax^2 + bx + c = 0$ 的求根公式：

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

由于当遇到 $b^2 \gg 4|ac|$ 的情形时，有 $|b| \approx \sqrt{b^2 - 4ac}$，则用上述公式求出的两个根中，总有一个因用了两个相近的近似数相减而严重不可靠.如本例若在能将规格化的数表示到小数点后 8 位的计算机上进行计算，则 $-b = 10^9 + 1 = 0.1 \times 10^{10} + 0.000\,000\,000\,1 \times 10^{10}$，由于第二项最后两位数“01”在机器上表示不出来，故它在上式的计算中不起作用，即在计算机运算时，$-b = 10^9$.

通过类似的分析可得

$$\sqrt{b^2 - 4ac} \approx |b| = 10^9$$

所以，求得的两个根分别为

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \approx \frac{10^9 + 10^9}{2} = 10^9$$

$$x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \approx \frac{10^9 - 10^9}{2} = 0$$

显然，根 x_2 是严重失真的.

为了求得可靠的结果，可以利用根与系数的关系式：$x_1x_2=\dfrac{c}{a}$，在计算机上采用如下公式：

$$\begin{cases} x_1=\dfrac{-b-\operatorname{sgn}(b)\sqrt{b^2-4ac}}{2a} \\ x_2=\dfrac{c}{ax_1} \end{cases}$$

其中，sgn(b)是 b 的符号函数，当 $b\geq 0$ 时 sgn(b)=1；当 b<0 时，sgn(b)=−1.显然，上述求根公式避免了相近数相减的可能性.

例 8　当 N 充分大时，如何计算

$$I=\int_N^{N+1}\frac{1}{1+x^2}\mathrm{d}x$$

分析　函数 $\dfrac{1}{1+x^2}$ 的原函数已知，我们自然考虑用 Newton-Leibniz 公式求这个定积分的值.由于 N 很大，这样会遇到两个相近的数相减，因此，应采用一些变换公式来避免这种情况.

解　若用定积分的 Newton-Leibniz 公式计算此题，有 $\int_N^{N+1}\dfrac{\mathrm{d}x}{1+x^2}=\arctan(N+1)-\arctan N$，则当 N 充分大时，因为 arctan(N+1)和 arctan$N$ 非常接近，两者相减会使有效数字严重丢失，从而影响计算结果的精度，这在数值计算中是要尽是量避免的，但是可以变换计算公式，例如，若令 $\tan\theta_1=N+1,\quad \tan\theta_2=N$，则由 $\tan(\theta_1-\theta_2)=\dfrac{\tan\theta_1-\theta_2}{1+\tan\theta_1\tan\theta_2}=\dfrac{N+1-N}{1+(N+1)N}=\dfrac{1}{1+(N+1)N}$ 得

$$\theta_1-\theta_2=\arctan(N+1)-\arctan N=\arctan\frac{1}{1+(N+1)N}$$

就可以避免两相近数相减引起的有效数字损失，从而得到较精确的结果.所以，当 N 充分大时，用 $\int_N^{N+1}\dfrac{\mathrm{d}x}{1+x^2}=\arctan\dfrac{1}{1+N+N^2}$ 计算积分的值较好.

例 9　计算积分 $I_n=\int_0^1 x^n\mathrm{e}^{x-1}\mathrm{d}x\quad (n=1,2,\cdots)$.

分析　数值计算中应采用数值稳定的算法，因此在建立算法时，应首先考虑它的稳定性.

解　利用分部积分法，有

$$\int_0^1 x^n e^{x-1}dx = \int_0^1 x^n de^{x-1} = x^n e^{x-1}\Big|_0^1 - \int_0^1 e^{x-1} n x^{n-1} dx = 1 - n\int_0^1 x^{n-1}e^{x-1}dx$$

得递推公式：

$$I_n = 1 - nI_{n-1} \quad (n = 1, 2, \cdots) \tag{1}$$

$$I_0 = \int_0^1 x^0 e^{x-1}dx = 1 - \frac{1}{e}$$

利用公式（1）计算 I_n，由于初值 I_0 有误差，不妨设求 I_0 的近似值 I_0^* 时有大小为 ε 的误差，即

$$I_0^* = I_0 + \varepsilon$$

则由递推公式（1）得

$$\begin{aligned}
&I_1^* = 1 - I_0^* = 1 - I_0 - \varepsilon = I_1 - \varepsilon \\
&I_2^* = 1 - 2I_1^* = 1 - 2I_1 + 2\varepsilon = I_2 + 2!\varepsilon \\
&I_3^* = 1 - 3I_2^* = 1 - 3I_2 - 3\times 2!\varepsilon = I_3 - 3!\varepsilon \\
&I_4^* = 1 - 4I_3^* = 1 - 4I_3 + 4\times 3!\varepsilon = I_4 + 4!\varepsilon \\
&\vdots \\
&I_n^* = I_n + (-1)^n n!\varepsilon
\end{aligned}$$

显然初始数据的误差 ε 是按 $n!$的倍数增长的，误差传播得很快，例如当 n=10 时，$10! \approx 3.629\times 10^6$，$\left|I_{10}^* - I_{10}\right| = 10!\varepsilon$，这表明 I_{10} 时已把初始误差 ε 扩大了很多倍，从而 I_{10}^* 的误差已把 I_{10} 的真值淹没掉了，计算结果完全失真.

但如果递推公式（1）改成

$$I_{n-1} = \frac{1}{n}(1 - I_n) \quad (n = k, k-1, \cdots, 3, 2)$$

于是，在从后往前计算时，I_n 的误差减少为原来的 $\frac{1}{n}$，所以，若取 n 足够大，误差逐步减小，显然，计算的结果是可靠的.所以，在构造或选择一种算法时，必须考虑到它的数值稳定性问题，数值不稳定的算法是不能使用的.

例 10 为了使计算

$$y = 10 + \frac{3}{x-1} + \frac{4}{(x-1)^2} - \frac{6}{(x-1)^3}$$

的乘除法运算次数尽量地少，应将表达式改写为怎样的形式？

解 设 $t = \frac{1}{x-1}$， $y = 10 + (3 + (4 - 6t)t)t$.

在数值计算中，应注意简化运算步骤，减少运算次数，使计算量尽可能小.

三、综合复习题

1．若 x^*=3587.64 是 x 的具有六位有效数字的近似值，求 x 的绝对误差限.

2．为使 $\sqrt{70}$ 的近似值的相对误差小于 0.1，问查开方表时，要取几位有效数字.

3．利用四位数学用表求 $x=1-\cos 2°$ 的近似值，采用下面等式计算.

（1）$1-\cos 2°$

（2）$2\sin^2 1°$

问哪一个结果较好？

4．求方程 $x^2-56x+1=0$ 的两个根，使它至少具有四位有效数字（已知 $\sqrt{783}\approx 27.982$）.

5．数列 $\{x\}_{n=0}^{\infty}$ 满足递推公式

$$x_n=10x_{n-1}-1,\quad (n=1,2,\cdots)$$

若取 $x_0=\sqrt{2}\approx 1.41$（三位有效数字），问按上述递推公式，从 x_0 计算到 x_{10} 时误差有多大？这个计算过程稳定吗？

6．如果近似值 $x^*=\pm(a_1+a_2\times 10^{-1}+a_3\times 10^{-2}+\cdots+a_n\times 10^{-n+1})\times 10^m$ 的相对误差限小于 $\dfrac{1}{2(a_1+1)}\times 10^{-n+1}$，证明的这个数具有 n 位有效数字.

四、复习题答案

1．0.005

2．取 3 位有效数字

3．（1）有一位有效数字，（2）有二位有效数字，显然（2）式较好.

4．用解二次代数方程 $ax^2+bx+c=0$ 的两个求根公式，有

$$x_1=\frac{-b-\operatorname{sgn}(b)\sqrt{b^2-4ac}}{2a}=\frac{56+\sqrt{56^2-4}}{2}$$

$$=28+\sqrt{783}\approx 55.983$$

$$x_2=\frac{c}{ax_1}=\frac{1}{28+\sqrt{783}}\approx\frac{1}{55.983}\approx 0.017\,863$$

5．$10^{10}\varepsilon<10^{10}\times\dfrac{1}{2}\times 10^{-2}=\dfrac{1}{2}\times 10^{8}$.

显然，误差积累很大，按递推公式 $x_n = 10x_{n-1} - 1$，求 x_{10} 时，会把初始误差 ε 扩大 10^{10} 倍，使计算精度受到严重影响，因此，这个计算过程不稳定.