17/10/14 计算物理

1. 寻找问题的近似值的解法

数值方法是对给定问题的输入数据和所需计算结果之间的关系的一种明确描述

建立数值方法原则：1. 便于计算机实现 2. 计算工作量尽量小
3. 存贮量尽量小 4. 问题的解与其近似解的误差小

2 误差来源

数值计算过程中的误差

舍入误差：计算机存储字节有限

截断误差：求级数和无穷序列的极限

例：$e^{-x}=1-x+\frac{1}{2}x^2-\frac{1}{3!}x^3+\cdots$

离散误差：将连续问题离散化

例：梯形积分

数据误差：初始数据是近似的

例：$\pi=3.14159265\cdots$；实验测量数据

3、绝对误差与相对误差

真值 x，近似值 $\bar{x}$，绝对误差 $x-\bar{x}=e_x$

对绝对误差的大小范围作出估计：$|e_x|=|x-\bar{x}|\le\varepsilon$ 绝对误差界

例：$x=100\,cm$ $\bar{x}=99cm$ $e_x=1cm$ | $e_y=50e_x$

$y=10000\,cm$ $\bar{y}=9950cm$ $e_y=50$ | x: 0.01 相对误差 y: 0.005 相对误差

相对误差：$r_x=\frac{e_x}{x}$ 或 $\frac{x-\bar{x}}{\bar{x}}$

$|r_x|\le\delta$ 相对误差界

4. 舍入误差与有效数字

$\pi=3.14159265\cdots$

$e=2.7182818\cdots$

$\sqrt{2}=1.41421356\cdots$

$\Rightarrow$ $a=\pm a_0a_1\cdots a_m.a_{m+1}\cdots a_na_{n+1}\cdots$

$a_i\ (i=0,1,\cdots)$ 为 0~9，且 $m\ne0$，$a_0\ne0$

四舍五入：取 $(n-m)$ 位小数

$$\bar{a}=\begin{cases}\pm a_0a_1\cdots a_m.a_{m+1}\cdots a_n, & a_{n+1}\le4\\ \pm a_0a_1\cdots a_m.a_{m+1}\cdots a_n+10^{-(n-m)}, & a_{n+1}\ge5\end{cases}$$

此时 $|a-\bar{a}|\le\frac{1}{2}\times10^{-(n-m)}$ $\Rightarrow$ 舍入误差

有效数字：近似数：

$\bar{a}=\pm a_0\cdots a_m.a_{m+1}\cdots a_n$

a_s 是 $\bar{a}$ 的第一位(自左至右)非零数，则自 a_s 到最右，所有数都为有效数字，$\bar{a}$ 是有 $(n+1-s)$ 位有效数字的有效数

例：$\bar{\pi}=3.1416$，$\bar{b}=0.035$，30.45 ~~~

定理：形如数字的近似数 $\bar{a}$ 具有 $n+1-s$ 位有效数字，则其相对误差的估计

$$\left|\frac{a-\bar{a}}{\bar{a}}\right|\le\frac{1}{2a_s}\times10^{-(n-s)}$$ 其中 $a_s\ne0$ 是 $\bar{a}$ 第一位有效数字

证明：1° 若 $s=0$，此时 $a_0\ne0$

$|\bar{a}|\ge a_0\times10^m$

$$\left|\frac{a-\bar{a}}{\bar{a}}\right|\le\frac{1}{2a_0}\cdot\frac{10^{-(n-m)}}{10^m}=\frac{1}{2a_0}\times10^{-n}$$

例1

2°. 若 $s\neq 0$, $a_s\neq 0$ 此时 $m=0$

$$|\bar{a}| \geq a_s \times 10^{-s}$$

$$\left|\frac{a-\bar{a}}{\bar{a}}\right| \leq \frac{1}{a_s\times 10^{-s}} \times \frac{1}{2}\times 10 = \frac{1}{2a_s}\times 10^{-(n-s)}$$

证毕

有效数字越多，相对误差越小，精度越高

30.40　相对误差边界 $\frac{1}{6}\times 10^{-3}$　3

30.4　$\frac{1}{6}\times 10^{-3}$

5. 数据误差在算术运算中的传递

$$x=\bar{x}+e_x,\quad y=\bar{y}+e_y$$

计算函数值 $z=f(x,y)$ 产生误差 $\bar{z}=f(\bar{x},\bar{y})$

* 假定绝对误差 e_x, e_y 绝对值很小, $f(x,y)$ 可微, 则 z 的误差

$$e_{\bar{z}} = \bar{z}-\bar{z} = f(x,y) - f(\bar{x},\bar{y})$$

$\cancel{f(x,y)\ f(\bar{x},\bar{y})}$

(1) $e_{\bar{z}} = \left(\frac{\partial f}{\partial x}\right)_{(\bar{x},\bar{y})} e_{\bar{x}} + \left(\frac{\partial f}{\partial x}\right)_{(\bar{x},\bar{y})} e_{\bar{y}}$

(2) $r_{\bar{z}} = \frac{e_{\bar{z}}}{\bar{z}} = \frac{\bar{x}}{\bar{z}}\left(\frac{\partial f}{\partial x}\right)_{(\bar{x},\bar{y})}\frac{e_{\bar{x}}}{\bar{x}} + \frac{\bar{y}}{\bar{z}}\left(\frac{\partial f}{\partial y}\right)_{(\bar{x},\bar{y})}\frac{e_{\bar{y}}}{\bar{y}}$

$$= \frac{\bar{x}}{\bar{z}}\left(\frac{\partial f}{\partial x}\right)_{(\bar{x},\bar{y})} r_{\bar{x}} + \frac{\bar{y}}{\bar{z}}\left(\frac{\partial f}{\partial y}\right)_{(\bar{x},\bar{y})} r_{\bar{y}}$$

同上些关系

f	e^-	r^-
$x\pm y$	$e_{\bar{x}} \pm e_{\bar{y}}$	$\frac{\bar{x}}{\bar{x}\pm\bar{y}} r_{\bar{x}} \pm \frac{\bar{y}}{\bar{x}\pm\bar{y}} r_{\bar{y}}$
xy	$\bar{y}e_{\bar{x}} + \bar{x}e_{\bar{y}}$	$r_{\bar{x}} + r_{\bar{y}}$
x/y	$\frac{\bar{y}e_{\bar{x}} - \bar{x}e_{\bar{y}}}{\bar{y}^2}$	$r_{\bar{x}} - r_{\bar{y}}$

• 避免相减相消

例: 1° $ax^2+bx+c=0$　$a\neq 0$

$$x=\frac{-b+\sqrt{b^2-4ac}}{2a} \to \frac{-2c}{b+\sqrt{b^2-4ac}}$$

2° $1-\cos x \xrightarrow{x\to 0} 2\sin^2\frac{x}{2}$

• 避免小数作分母

6、机器误差

5.1 十进制的小数

$$x=\pm 10^J\sum_{k=1}^{t} d_k 10^{-k}$$　J为整数，$d_1\cdots d_t$ 为 0~9

例：$312.74 = 10^3(3\times10^{-1}+1\times10^{-2}+2\times10^{-3}+7\times10^{-4}+4\times10^{-5})$ $=0.31274\times10^3$

$-0.012 = 10^{-1}(1\times10^{-1}+2\times10^{-2})$　$=-0.12\times10^{-1}$

若记 $a=\sum_{k=1}^{t} d_k 10^{-k}=0.d_1d_2\cdots d_t$

则 $x=\pm a\times 10^J$

↳十进制系统计数法

其中10为十进制系统的基数

二、八、十六 → 2, 8, 16

对于p进制：$x=\pm p^J\sum_{k=1}^{t} d_k p^{-k}=\pm a\times p^J$

a为x的尾数，J为x的阶数

计算机中会规定字长t，以及阶J的范围 $-L\le J\le U$.

定点数：如果J固定　　浮点数：J可变

$5360 = 0.5360\times10^4 = 0.05360\times10^5$

定义：规格化浮点数

要求 $0.1\le a<1$　十进制

$\frac{1}{2}\le a<1$　二进制

对按 $x=\pm p^J\sum_{k=1}^{t} d_k p^{-k}$ 规定的浮点数的全体，且 $x\ne0$ 时，$d_1\ne0$.

称规格化浮点数系

例：二进制 $p=2$　$t=3$　$L=1$　$U=2$.

$$x=\pm a\times 2^J=\pm 2^J(d_1\times2^{-1}+d_2\times2^{-2}+d_3\times2^{-3})$$

~~当~~ 当 $x\ne0$ 时，$d_1=1$

d_1	d_2	d_3	
1	0	0	$\frac{4}{8}$
1	0	1	$\frac{5}{8}$
1	1	0	$\frac{6}{8}$
1	1	1	$\frac{7}{8}$

尾数a：

F中有33个浮点数.

$J=-1$	$J=0$	$J=1$	$J=2$

以及0

6.2. 浮点运算与舍入误差.

近似 十进制: $x = \pm a \times 10^J$

$a = 0.d_1 \cdots d_t d_{t+1} \cdots d_n$

$$\bar{a} = \begin{cases} 0.d_1 \cdots d_t & 0 \le d_{t+1} \le 4 \\ 0.d_1 \cdots d_t + 10^{-t} & d_{t+1} \le 5 \end{cases}$$

$x_R = \pm \bar{a} \times 10^J$

$\because \left|\frac{\bar{x}_R - x}{x}\right| = \left|\frac{\bar{a}-a}{a}\right|$

$\because \quad a \ge 10^{-1} \leftarrow$ 规范化

$\therefore \left|\frac{\bar{a}-a}{a}\right| \le \frac{1}{2a} \times 10^{-t} = \frac{1}{2} \times 10^{-t+1} = 5 \times 10^{-t}$

即 $\left|\frac{x_R - x}{x}\right| \le 5 \times 10^{-t}$.

若令 $\frac{x_R - x}{x} = \varepsilon$

$x_R = x(1+\varepsilon) \qquad |\varepsilon| \le 5 \times 10^{-t}$.

二进制: $x = \pm a \times 2^J \qquad \bar{a} = \begin{cases} 0.d_1 \cdots d_t & ; d_{t+1} = 0 \\ 0.d \cdots d_t + 2^{-t} & ; d_{t+1} = 1 \end{cases}$

$x_R = \pm \bar{a} \times 2^J$

$x_R = x(1+\varepsilon) \qquad |\varepsilon| \le 2^{-t}$

试证: 对于"只舍不入" 断位法

十: $|\varepsilon| \le 10^{-t}$

二: $|\varepsilon| \le 2^{1-t}$.

浮点数运算: ⊕ ⊖

例 舍入法 $t=5 \quad p=10$, 阶小的移 $x = 0.21062 \times 10^{-5} \quad y = 0.12345 \times 10^{-3}$

$x \oplus y = 0.0021 \times 10^{-3} + 0.12345 \times 10^{-3}$

$= 0.12556 \times 10^{-3}$

例 断位法: $p=10,\ t=3,\ L=U=5$. $0.0438, 0.0693, 13.2$ 做加法

法一: $(0.438 \times 10^{-1} \oplus 0.693 \times 10^{-1}) \oplus 0.132 \times 10^2$

$= 0.113 \times 10^0 \oplus 0.132 \times 10^2$

$= 0.001 \times 10^2 \oplus 0.132 \times 10^2$

$= 0.133 \times 10^2$

法二: $(0.132 \times 10^2 \oplus 0.693 \times 10^{-1}) \oplus 0.438 \times 10^{-1}$

$= 0.132 \times 10^2 \oplus 0.438 \times 10^{-1}$

$= 0.132 \times 10^2$ 运算规律不成立

例: 舍入法: $p=10, \quad t=4 \qquad x = 0.12378 \quad y = 0.12362$

$x_R \ominus y_R = 0.1237 \times 10^0 \ominus 0.1236 \times 10^0$

$= 0.1000 \times 10^{-3}$

x_R 相对误差 6.467×10^{-4}

y_R 相对误差 1.618×10^{-4} $\Rightarrow$ 相减相消

$x_R \ominus y_R$ 相对误差: $\frac{0.00006}{0.0001} = 0.6$

· 考察浮点运算舍入误差.

$fl(x+y)$ ······ 代表舍入规则的结果

上例 $x+y=0.0021062\times10^{-3}+0.12345\times10^{-3}$

$=0.1255562\times10^{-3}$

$fl(x+y)=0.12556\times10^{-3}$

定理:

$$fl(x+y)=(x+y)(1+\varepsilon_1)$$
$$fl(x-y)=(x-y)(1+\varepsilon_2)$$
$$fl(xy)=(xy)(1+\varepsilon_3)$$
$$fl(x/y)=(x/y)(1+\varepsilon_4)$$

其中 $\varepsilon_i\le eps \quad i=1,2,3,4$

$$eps=\begin{cases}5\times10^{-t} & (\text{十进制})\\ 2^{-t} & (\text{二进制})\end{cases}$$

$$fl(x+y+z)=fl(fl(x+y)+z)$$
$$=(x+y)(1+\varepsilon_1)(1+\varepsilon_2)+z(1+\varepsilon_2)$$

$$fl\left(\sum_{i=1}^{3}x_i\times y_i\right)=fl(x_1\times y_1+x_2\times y_2+x_3\times y_3)$$
$$=(x_1\times y_1)(1+\varepsilon_1)(1+\varepsilon_4)(1+\varepsilon_5)+(x_2\times y_2)(1+\varepsilon_2)(1+\varepsilon_4)(1+\varepsilon_5)$$
$$+(x_3\times y_3)(1+\varepsilon_3)(1+\varepsilon_5)$$

其中 $|\varepsilon_i|\le eps$, $\prod(1+\varepsilon_i)$

引理: 若 $|\varepsilon_i|\le eps$ 且 $n\cdot eps\le0.01$

则 $1-n\cdot eps\le\prod_{i=1}^{n}(1+\varepsilon_i)\le1+1.01n\cdot eps$

其中 $eps=\begin{cases}5\times10^{-t}\\ 2^{-t}\end{cases}$

可设为 $\prod_{i=1}^{n}(1+\varepsilon_i)=1+1.01\cdot\theta\cdot eps$, $|\theta|\le1$

证明: 由假设 $|\varepsilon_i|\le eps$

$$(1-eps)^n\le\prod_{i=1}^{n}(1+\varepsilon_i)\le(1+eps)^n$$

对 $(1-x)^n$ 作泰勒展开

$$(1-x)^n=1-nx+\frac{n(n-1)}{2}(1-\theta x)^{n-2}$$
$$\ge1-nx$$

即 $1-neps\le(1-eps)^n\le\prod_{i=1}^{n}(1+\varepsilon_i)$ 得证.

又由于 $e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots$

$$=1+x+\frac{x}{2}\cdot x\left(1+\frac{x}{3}+\frac{2x^2}{4!}+\cdots\right)$$
$$\le1+x+\frac{x}{2}\cdot x\cdot e^x$$

当 $0\le x\le0.01$ 时,

$$1+x\le e^x\le1+x+\frac{0.01}{2}xe^{0.01}\le1+1.01x$$
$$(1+eps)^n\le e^{n\cdot eps}\le1+1.01\cdot n\cdot eps$$

误证

定理2 若 $n\cdot eps \le 0.01$

$$fl\left(\sum_{i=1}^{n} x_i \times y_i\right) = x_1 y_1 (1+1.01 n\theta_1 eps) + \sum_{i=2}^{n} x_i y_i [1+1.01(n+2-i)\theta_i eps]$$

其中 $|\theta_i| \le 1$

上述误差分析过程，将初始数据的实际浮点运算归结为初始近似数据的精确运算，从而将计算过程的误差归结为初始数据的误差，$\boxed{\text{向后误差分析法}}$

$$fl(x+y) = (x+y)(1+\varepsilon) = x(1+\varepsilon) + y(1+\varepsilon)$$

如果是直接估计计算结果与真确值之间的误差，$\boxed{\text{向前误差分析法}}$

$$|fl(x+y) - (x+y)|.$$

二、非线性方程的数值方法

1. 迭代法

$$f(x)=0$$

$$x_0 \;\text{——}\; x_k \rightarrow \lim_{k\to\infty} x_k = p$$

$\sqrt{3}$:

$$x_k = \left(x_{k-1} + \frac{3}{x_{k-1}}\right)/2$$

大范围收敛：从任何可取的初始出发都保证收敛

局部收敛：初值要充分接近于所求的根

2. 区间半分法

$f(x)=0$ 介值定理 二分法

$f(x)$ 在 $[a,b]$ 上连续，一旦 $f(a)f(b)<0$

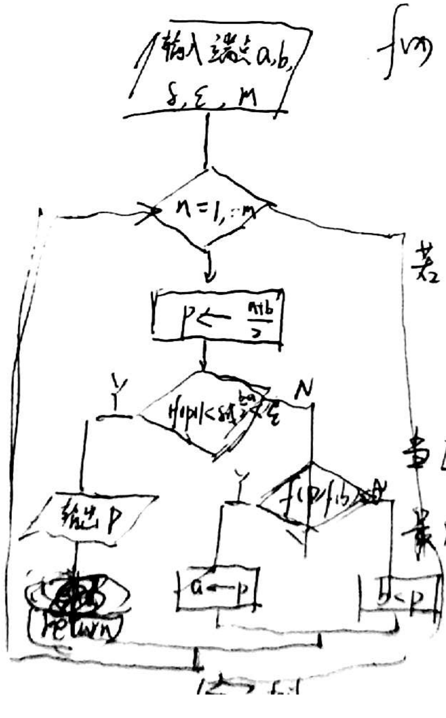
设 $(a,b)=(a_1,b_1)$，p_1 为 (a_1,b_1) 中点

$$p_1 = \frac{a_1+b_1}{2}$$

若 $|f(p_1)|<\delta$ 则为根近似

$$f(p_1)f(b_1) \begin{cases} <0 \rightarrow a_2=p_1,\ b_2=b_1 \\ >0 \rightarrow a_2=a_1,\ b_2=p_1 \end{cases}$$

当区间中点函数值绝对值小于误差容限 δ 或区间长小于容限 ε 时终止，最后区间中点作为 $f(x)=0$ 一个近似解



讨论

1° 收敛 $f(x)=0$ 一个根 p，大范围收敛，

得到 $p_1, p_2, \cdots, p_n, \cdots$

则有 $$|p_n-p| \le \frac{1}{2^n}(b-a)$$

这是先验的误差界 令 ε 为预先的绝对误差容限.

要求 $|p_n-p|<\varepsilon$，只要 $\frac{1}{2^n}(b-a)<\varepsilon$.

$$2^n > \frac{(b-a)}{\varepsilon}$$

两边取对数

$$n > \frac{\lg \frac{b-a}{\varepsilon}}{\lg 2} \longrightarrow$$ 迭代终止准则 第n步停止

2° $|f(p_n)|<n$. 对某些其它函数，存在缺点

$f(p_n)\to 0$，但 p_n 与 p 相差很大

例：$f(x)=(x-1)^{10} \to p=1$

令 $p_n = 1+\frac{1}{n}$

当 $n>1$ 时 $|f(p_n)| < \frac{1}{2^{10}} < 10^{-3}$

但要 $|p_n-p|<10^{-3}$，则 $n>1000$

∴ 另一个较好的终止准则是 $\frac{p_n-p_{n-1}}{p_n} < \varepsilon$

对于1°例：

$f(x)=x^3-x-1 \qquad f(1)f(2)<0$

$f'(x)=3x^2-1>0$，只有一根

要求 p 的近似值绝对误差不超过 10^{-4}

$$|p_n-p| \le \frac{b-a}{2^n} = \frac{1}{2^n}$$

$$\frac{1}{2^n} < 10^{-4} \to n > \frac{\lg 10^4}{\lg 2} \approx 13.3 \to n=14$$

∴ 二分法，误差下降不快，但安全，简单

3. 不动点迭代

$$f(x)=0 \Rightarrow x=g(x)$$

例：$f(x)=x^3+2x^2-4=0$

(1) $x = x-x^3-2x^2+4$

(2) $x=g_2(x)=\left[2\left(\frac{2}{x}-x\right)\right]^{1/2}$

(3) $x=g_3(x)=\left(2-\frac{x^3}{2}\right)^{1/2}$

(4) $x=g_4(x)=2\left(\frac{1}{2+x}\right)^{1/2}$

(5) $x=g_5(x)=x-\frac{x^3+2x^2-4}{3x^2+4x}$

$x_k = g(x_{k-1})$ 不动点迭代法 Picard迭代 $\lim\limits_{k\to\infty} x_k = p$

定理1　假设$g(x)$为定义在$[a,b]$上的实函数

(I)　$g(x)\in[a,b]$，$\forall x\in[a,b]$

(II)　Lipschitz条件，且Lipschitz常数$L<1$。

即　存在正常数$L<1$，使$|g(x)-g(y)|\le L|x-y|$，$\forall x,y\in[a,b]$

那么对任意初始值$x_0\in[a,b]$，由Picard迭代产生的序列都收敛于g的唯一不动点p，并且有误差估计式：

$$|e_k|\le\frac{L^k}{1-L}|x_1-x_0|$$

其中　$e_k=x_k-p$。