误差

若干定义

1. 误差的来源：
    1. 模型误差：由计算方法或计算模型的不严格而引入
    2. 观测误差：生产实践中由于测量精度的不足而引入的测量误差
    3. **截断误差**：对于将连续问题/无限问题近似为离散问题/有限问题而进行的“差分”或“舍去高阶无穷小项”等操作而引入的计算误差
    4. 舍入误差：对于利用计算机求解的问题，计算机最长字长限制了我们的有效数字位数，故我们进行舍入而引入的误差
2. **绝对误差和绝对误差限**：约定物理量的真实值记为x，其观测值记为x^*. 那么，我们将值$x - x^*$称为绝对误差$e(x^*)$，将绝对误差之绝对值的**上界**称为绝对误差限$\epsilon(x^*)$(实际上在授课的PPT中，绝对误差限没有被表达为一个对映关系，但为了数学上的对称，将其修改为了函数形式).
这样我们就可以将真实值表达为：
$$x = x^* \pm \epsilon(x^*)$$
3. **相对误差和相对误差限**：约定物理量的真实值记为x，其观测值记为x^*. 那么，我们将值$\frac{e(x^*)}{x}$称为相对误差$e_r(x^*)$. 实际上，绝大多数情况下，物理量的真实值x是不可知的，因此我们可以将相对误差近似地计算为$\frac{e(x^*)}{x^*}$. 同理，相对误差之绝对值的**上界**称为相对误差限$\epsilon_r(x^*)$

给出定义之后，我们可以发现绝对误差限和相对误差限之间的关系：

$$\epsilon_r = \frac{\epsilon}{|x^*|} \tag{1.1}$$

请注意，在上述的描述中使用了“上界”的措辞，这表示了误差限可以远远大于相应误差. 作为基本的数学分析内容，**上界**的概念应与**上确界**区分. **请务必将此概念彻底理解，以便进行下述的数学推导.**

有效数字

定义

兹定义：若$\epsilon(x^*)$ 的绝对误差限可以是 x^* 某一位上数字的半个单位，且该位直到 x^* 的首个非零数字一共有 n 位数字，则称近似值 x^* 有 n 位有效数字.

分析地描述：设置测量值x^*被c进制表示，我们将其从小数点后的第m位称为“最后一位有效数字”**当且仅当**m满足如下条件：

$$|x - x^*| \leqslant \frac{1}{2} \times c^{-m} \tag{1.2}$$

此时x^*具有n位有效数字**当且仅当**其首个非零数字到小数点之后第m位数字中共有n个数字.

对于常用的十进制情况，我们将式(1.2)转写为

$$|x - x^*| \leqslant \frac{1}{2} \times 10^{-m} \tag{1.2'}$$

另一等价定义

将以c进制表示的**测量值**x^*按照数位转写为如下的标准形式:

$$x^* = \pm\overline{0.a_1a_2a_3 \ldots a_n \ldots} \times c^m \ (a_1 \neq 0) \tag{1.3}$$

称x^*具有n位有效数字当且仅当:

$$|x - x^*| \leqslant \frac{1}{2} \times c^{m-n} \tag{1.4}$$

容易证明，上述两种定义是等价的.

有效数字位数与首位数字的关系

定理1.1-1. 若以c进制表示的估计值x^*有n位有效数字，则

$$\epsilon_r(x^*) = \frac{1}{2a_1} \times c^{-n+1} \tag{1.5}$$

可以成立.(注：相对误差限是一个任意大的上界，因此所谓"可以成立"是指其相对误差的上确界不大于式1.5的右值)

定理1.1-1，证明.

由式1.1，

$$\epsilon_r = \frac{\epsilon}{|x^*|} \leqslant \frac{1/2 \times c^{m-n}}{|x^*|} = \frac{1/2 \times c^{m-n}}{\overline{0.a_1a_2a_3 \ldots a_n \ldots} \times c^m} \leqslant \frac{1/2 \times c^{m-n}}{\overline{0.a_1000 \ldots} \times c^m} \leqslant \frac{1}{2a_1} \times c^{-n+1}$$

■

定理1.1-2

若以c进制表示的估计值x^*的相对误差限可以取到:

$$\epsilon_r(x^*) = \frac{1}{2(a_1 + 1)} \times c^{-n+1} \tag{1.6}$$

那么x^*至少具有n位有效数字

定理1.1-2，证明

首先注意，上述的"可以取到"意味着

$$\epsilon_r(x^*) \leqslant \frac{1}{2(a_1+1)} \times c^{-n+1} \tag{1.6’}$$

由式1.1，

$$\epsilon = \epsilon_r|x^*| \leqslant (\frac{1}{2(a_1+1)} \times c^{-n+1})|x^*| \tag{1.7}$$

注意到

$$|x^*| = \overline{0.a_1a_2a_3 \ldots a_n \ldots} \times c^m \ \leqslant (a_1+1) \times c^{m-1} \tag{1.8}$$

将式1.8代入至式1.7中，解出

$$\epsilon \leqslant \frac{1}{2} \times c^{-n+m} \tag{1.4}$$

■

在课间听到了类似于"为什么分母上一个是a_1一个是(a_1+1)呢？"的问题，其实这是一个初等数学内容：我们知道，对于正分数，分母缩小分数则变大，分母变大分数则变小，对于两个定理的证明过程，我们均需保留一个不大于号$\leqslant$，且我们都做了将x^*的小数形式"截断"的操作，因此为保持推导过程中的不等号方向不变(否则推导将难以收敛)，我们就将其近似为了不同的形式.

四舍五入近似法，有效数位的简单理解

先来看一个例子：

例1.1 以$x_1^* = 3.1415$和$x_2^* = 3.1416$近似$\pi = 3.141592653..$，试分别计算其有效数字.

解 由$\pi - x_1^* = 0.00009265...$得知：

$$1/2 \times 10^{-4} \leqslant |\pi - x_1^*| \leqslant 1/2 \times 10^{-3}$$

故$x_1^* = 3.1415$具有4位有效数字.

同理，

$$1/2 \times 10^{-5} \leqslant |\pi - x_2^*| \leqslant 1/2 \times 10^{-4}$$

故$x_2^* = 3.1416$具有5位有效数字.■

直观地来看，上例中x_1进行了错误的四舍五入近似操作，而x_2进行了正确的四舍五入近似操作，这似乎导致了x_2在上述的定义中具有更多的有效数字. 这种看法不失为一种对于有效数字定义的理解. 结合我们在中学物理中学习到的"估读位数也算有效数字"的结论，以科学哲学的眼光来看，我们可以这样理

解有效数字的定义：在实际生产实践的测量操作中，由于人眼的某种固有属性或者人脑的某种固有价值，我们在判断"测量值是否超过了最小刻度的1/2"有着较强的能力. 因此在数学中，我们将上述的价值从生产中抽离出来，将"是否准确估读至1/2"理性衍生为"误差是否大于某一位上数字的半个单位"，作为了有效数字的判断依据.

误差的传播

兹定义误差传递模式：对于一个对映关系$y = f(X)$，y的误差表示为：

$$e(y^*) = y - y^* = f(X) - f(X^*),\ X \in \mathbb{R}^n \tag{1.9}$$

若f在点X^*附近可微，且我们假定$X \approx X^*$，即可提出误差传递的近似计算方式：

$$e(y^*) = f(X) - f(X^*) \approx df(X^*) \tag{1.10-1}$$

由中值定理：

$$e(y^*) \approx f'(\xi)(X - X^*),\ \xi \in (X, X^*) \tag{1.10-2}$$

我们已经假定$X \approx X^*$，因此：

$$e(y^*) \approx f'(X^*)e(X^*) \tag{1.10-3}$$

若将n维向量X展开为实数形式，则

$$e(y^*) = \sum_{i=1}^{n} \frac{\partial f(x_1^*, x_2^*, ..., x_n^*)}{\partial x_i} e(x_i) \tag{1.10}$$

同理，我们推导相对误差的传递：

$$e_r(y^*) = \frac{e(y^*)}{y^*} \approx \sum_{i=1}^{n} \frac{\partial f(x_1^*, x_2^*, ..., x_n^*)}{\partial x_i} \frac{e(x_i)}{y^*} \tag{1.11-1}$$

为了形式上的对称，我们将$e(x_i)$转写为$x_i^* e_r(x_i^*)$，有

$$e_r(y^*) \approx \sum_{i=1}^{n} \frac{\partial f(x_1^*, x_2^*, ..., x_n^*)}{\partial x_i} \frac{x_i^*}{y^*} e_r(x_i) \tag{1.11-1}$$

计算方法的数值稳定性，误差传播对工程的启发

若在计算过程中，数据误差不增长，则称算法是数值稳定的.

在实际计算中应注意如下可能会使误差急剧增长的操作：

1. 避免两个相近的数相减

$$|e_r(x-y)| = \frac{|e(x)-e(y)|}{|x-y|}$$

在$x \approx y$时，x与y的相减操作可能会得到相对误差非常大的结果.
因此，在当x相当大时，计算如下等式时应进行变换：

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

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

2. 优先计算较小数
   在利用计算机浮点数时，应优先计算较小数. 例：

```
double res1 = 0;
double res2 = 0;
res1 += 1e20;
for (int i = 0; i < INT_MAX; i++) {
    res1 += 1;
    res2 += 1;
}
res2 += 1e20;
```

   在运行上述代码段后，上述两个变量 `res1` 和 `res2` 会得到一个不同的数值，而在数学上，它们理应是相等的.
3. 避免小数除大数

$$\epsilon(\frac{x}{y}) = \frac{ye(x)-xe(y)}{y^2}$$

4. 简化运算次数

2021.10.16
Hautbois