数值计算方法

数值计算中的误差分析

张晓平

目录

1.1 数值计算的对象、任务和特点

现代科学的三大手段

- 理论
- 实验
- 科学计算

1. 1.1 数值计算的对象、任务和特点
   - 研究数值方法的必要性
   - 科学计算的流程
   - 研究对象
   - 研究任务

2. 1.2 误差

3. 1.3 选用和设计计算法应遵循的原则

研究数值方法的必要性

对于线性方程组

$$Ax = b,$$

定理 (Crammer法则)

若A非奇异，则此方程组有唯一解，且

$$x_i = \frac{|A_i|}{|A|}, \quad i = 1, 2, \cdots, n.$$

其中A_i是将A的第i列换为b而得的矩阵。

该结论非常漂亮，它把线性方程组的求解问题归结为计算$n+1$个n阶行列式的计算问题。

研究数值方法的必要性

对于线性方程组

$$Ax = b,$$

定理 (Crammer法则)

若A非奇异，则此方程组有唯一解，且

$$x_i = \frac{|A_i|}{|A|}, \quad i = 1, 2, \cdots, n.$$

其中A_i是将A的第i列换为b而得的矩阵。

该结论非常漂亮，它把线性方程组的求解问题归结为计算$n+1$个n阶行列式的计算问题。

对于行列式的计算

定理 (Laplace展开定理)

若A非奇异，则此方程组有唯一解，且

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}$$

其中A_{ij}是元素a_{ij}的代数余子式。

该方法的运算量大的惊人，以至于完全不能用于实际计算。

对于行列式的计算

定理 (Laplace展开定理)

若A非奇异，则此方程组有唯一解，且

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}$$

其中A_{ij}是元素a_{ij}的代数余子式。

该方法的运算量大的惊人，以至于完全不能用于实际计算。

研究数值方法的必要性

设k阶行列式所需乘法运算的次数为m_k，则

$$m_k = k + km_{k-1},$$

于是有

$$\begin{aligned} m_n &= n + nm_{n-1} \\ &= n + n[(n-1) + (n-1)m_{n-2}] \\ &= \cdots \\ &= n + n(n-1) + n(n-1)(n-2) + \cdots + n(n-1)\cdots 3\cdot 2 \\ &> n! \end{aligned}$$

故用Crammer法则和Laplace展开定理求解一个n阶线性方程组，所需乘法运算的次数就大于

$$(n+1)n! = (n+1)!.$$

研究数值方法的必要性

在一台百亿次的计算机上求解一个25阶线性方程组，则至少需要

$$\frac{26!}{10^{10}\times 3600\times 24\times 365}\approx \frac{4.0329\times 10^{28}}{3.1526\times 10^{17}}\approx 13\text{亿年}$$

而用下章介绍的消去法求解，则需要不到一秒钟。

研究数值方法的必要性

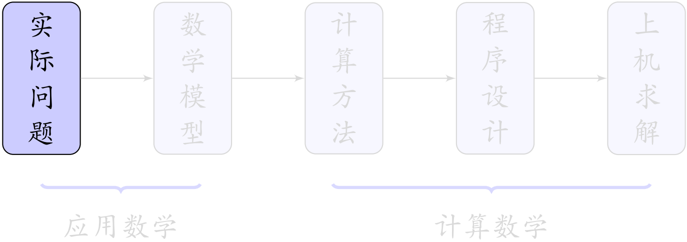
在一台百亿次的计算机上求解一个25阶线性方程组，则至少需要

$$\frac{26!}{10^{10}\times 3600\times 24\times 365}\approx \frac{4.0329\times 10^{28}}{3.1526\times 10^{17}}\approx 13\text{亿年}$$

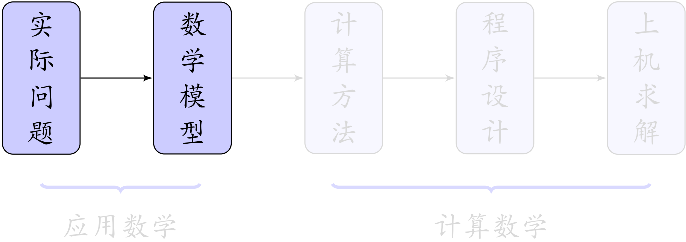
而用下章介绍的消去法求解，则需要不到一秒钟。

科学计算的流程

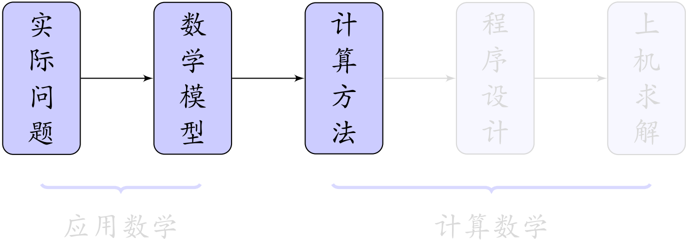
科学计算的流程



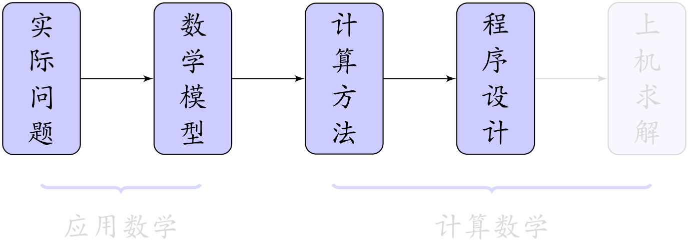
科学计算的流程



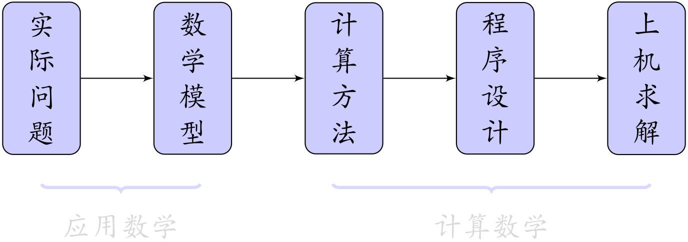
科学计算的流程



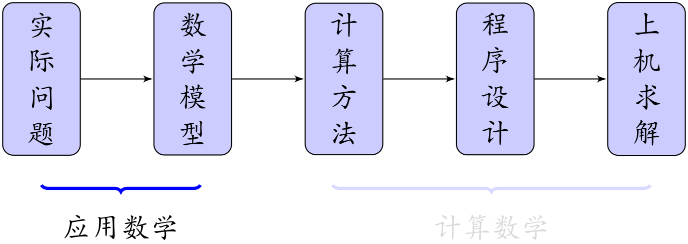
科学计算的流程



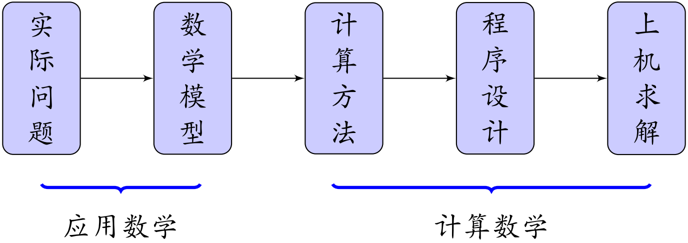
科学计算的流程



科学计算的流程



科学计算的流程



研究对象

- 代数
- 微积分
- 微分方程
- 积分方程
- $\cdots$

1. 1.1 数值计算的对象、任务和特点
   - 研究数值方法的必要性
   - 科学计算的流程
   - 研究对象
   - 研究任务
2. 1.2 误差
3. 1.3 选用和设计计算法应遵循的原则

研究任务

- 算法设计

  快速、可靠

- 理论分析

  算法的收敛性、稳定性以及误差分析

- 复杂度分析

  计算时间最短、所需内存最少

研究任务

- 算法设计

  快速、可靠

- 理论分析

  算法的收敛性、稳定性以及误差分析

- 复杂度分析

  计算时间最短、所需内存最少

研究任务

- 算法设计

  快速、可靠

- 理论分析

  算法的收敛性、稳定性以及误差分析

- 复杂度分析

  计算时间最短、所需内存最少

误差来源与分类

- 模型误差
- 观测误差
- 截断误差
- 舍入误差

误差来源与分类

模型误差

数学模型只是复杂客观现象的一种近似，它与实际问题总会存在一定误差

观测误差

由于测量精度和手段的限制，观测或实验得来的物理量总会与实际量之间存在误差

误差来源与分类

模型误差

数学模型只是复杂客观现象的一种近似，它与实际问题总会存在一定误差

观测误差

由于测量精度和手段的限制，观测或实验得来的物理量总会与实际量之间存在误差

截断误差

数学模型的精确解与由数值方法求出的近似解之间的误差.

$$e^x \approx 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{10}}{10!}$$

$$R_{10}(x) = \frac{\xi^{11}}{11!}$$

误差来源与分类

舍入误差

由于计算机的字长有限，进行数值计算的过程中，对计算得到的中间结果数据要使用“四舍五入”或其他规则取近似值，因而使计算过程有误差。

1990年2月25日，海湾战争期间，在沙特阿拉伯宰赫兰的爱国者导弹防御系统因浮点数舍入错误而失效，该系统的计算机精度仅有24位，存在0.0001%的计时误差，所以有效时间阈值是20个小时。当系统运行100个小时以后，已经积累了0.3422秒的误差。这个错误导致导弹系统不断地自我循环，而不能正确地瞄准目标。结果未能拦截一枚伊拉克飞毛腿导弹，飞毛腿导弹在军营中爆炸，造成28名美国陆军死亡。

舍入误差

由于计算机的字长有限，进行数值计算的过程中，对计算得到的中间结果数据要使用“四舍五入”或其他规则取近似值，因而使计算过程有误差。

1990年2月25日，海湾战争期间，在沙特阿拉伯宰赫兰的爱国者导弹防御系统因浮点数舍入错误而失效，该系统的计算机精度仅有24位，存在0.0001%的计时误差，所以有效时间阙值是20个小时。当系统运行100个小时以后，已经积累了0.3422秒的误差。这个错误导致导弹系统不断地自我循环，而不能正确地瞄准目标。结果未能拦截一枚伊拉克飞毛腿导弹，飞毛腿导弹在军营中爆炸，造成28名美国陆军死亡。

1 1.1 数值计算的对象、任务和特点

2 1.2 误差

- 误差来源与分类
- 误差与有效数字
  - 绝对误差与绝对误差限
  - 相对误差与相对误差限
  - 有效数字
  - 有效数字与绝对误差、相对误差

3 1.3 选用和设计计算法应遵循的原则

绝对误差与绝对误差限

定义 (绝对误差与绝对误差限)

设某个量的精确值为x，其近似值为$x^\star$，则称

$$E(x) = x - x^\star$$

为近似值$x^\star$的绝对误差，简称误差。若存在$\eta > 0$，使得

$$|E(x) = |x - x^\star| \leq \eta$$

则称η为近似值$x^\star$的绝对误差限，简称误差限或精度。

η越小，表示近似值$x^\star$的精度越高。

$$x^\star - \eta \leq x \leq x^\star + \eta$$

或

$$x = x^\star \pm \eta$$

绝对误差与绝对误差限

定义 (绝对误差与绝对误差限)

设某个量的精确值为x，其近似值为$x^\star$，则称

$$E(x) = x - x^\star$$

为近似值$x^\star$的绝对误差，简称误差。若存在$\eta > 0$，使得

$$|E(x) = |x - x^\star| \leq \eta$$

则称η为近似值$x^\star$的绝对误差限，简称误差限或精度。

η越小，表示近似值$x^\star$的精度越高。

$$x^\star - \eta \leq x \leq x^\star + \eta$$

或

$$x = x^\star \pm \eta$$

绝对误差与绝对误差限

例

用毫米刻度的直尺量一长度为x的物体，测得其近似值为$x^\star = 84mm$。

因直尺以mm为刻度，其误差不超过$0.5mm$，即有

$$|x - 84| \leq 0.5(mm)$$

或

$$x = 84 \pm 0.5(mm)$$

绝对误差与绝对误差限

例

用毫米刻度的直尺量一长度为x的物体，测得其近似值为$x^\star = 84mm$。

因直尺以mm为刻度，其误差不超过$0.5mm$，即有

$$|x - 84| \leq 0.5(mm)$$

或

$$x = 84 \pm 0.5(mm)$$

相对误差与相对误差限

例

测量$100m$和$10m$的两个长度，若它们的绝对误差均为$1cm$，显然前者的测量更为精确。

由此可见，决定一个量的近似值的精确度，除了绝对误差外，还必须考虑该量本身的大小，为此引入相对误差的概念。

相对误差与相对误差限

例

测量$100m$和$10m$的两个长度，若它们的绝对误差均为$1cm$，显然前者的测量更为精确。

由此可见，决定一个量的近似值的精确度，除了绝对误差外，还必须考虑该量本身的大小，为此引入相对误差的概念。

相对误差与相对误差限

近似值$x^\star$的相对误差是绝对误差与精确值之比，即

$$E_r(x) = \frac{E(x)}{x} = \frac{x - x^\star}{x}.$$

实际中由于精确值x一般无法知道，故往往取

$$E_r^\star(x) = \frac{E(x)}{x^\star} = \frac{x - x^\star}{x^\star}.$$

若存在$\delta > 0$，使得

$$|E_r(x)| = |\frac{x - x^\star}{x^\star}| \leq \delta,$$

则称δ为近似值$x^\star$的相对误差限。

相对误差与相对误差限

例

当$|x-x^{\star}|\leq 1cm$时，测量*100m*物体时的相对误差为

$$|E_r(x)| = \frac{1}{10000} = 0.01\%,$$

测量*10m*物体时的相对误差为

$$|E_r(x)| = \frac{1}{1000} = 0.1\%.$$

有效数字

定义 (有效数字)

若近似值$x^\star$的绝对误差限是某一位的半个单位，就称其精确到这一位， 且从该位直到$x^\star$的第一位非零数字共有n位，则称近似值$x^\star$有n位有效数字。

358.467	358.47
0.00427511	0.0042751
8.000034	8.0000
8.000034×10^3	8000.0

有效数字

定义 (有效数字)

若近似值$x^\star$的绝对误差限是某一位的半个单位，就称其精确到这一位，且从该位直到$x^\star$的第一位非零数字共有n位，则称近似值$x^\star$有n位有效数字。

358.467	358.47
0.00427511	0.0042751
8.000034	8.0000
8.000034×10^3	8000.0

有效数字

任何一个实数x经四舍五入后得到的近似值$x^\star$都可写成

$$x^\star = \pm(\alpha_1 \times 10^{-1} + \alpha_2 \times 10^{-2} + \cdots + \alpha_n \times 10^{-n}) \times 10^m.$$

当其绝对误差限满足

$$|x - x^\star| \leq \frac{1}{2} \times 10^{m-n}$$

时，则称近似值$x^\star$具有n位有效数字，其中m为整数，α_1为1到9中的一个数字，$\alpha_1, \cdots, \alpha_n$是0到9中的数字。

有效数字与绝对误差、相对误差

(1) 若某数x的近似值$x^\star$有n位有效数字，则此近似值$x^\star$的绝对误差限为

$$|x - x^\star| \leq \frac{1}{2} \times 10^{m-n}.$$

当m一定时，有效数字位数n越多，其绝对误差限越小。

(1) 若某数x的近似值$x^\star$有n位有效数字，则此近似值$x^\star$的绝对误差限为

$$|x - x^\star| \le \frac{1}{2} \times 10^{m-n}.$$

当m一定时，有效数字位数n越多，其绝对误差限越小。

(2) 若近似值$x^\star$具有n位有效数字，则其相对误差限为

$$E_r^\star(x) \le \frac{1}{2\alpha_1} \times 10^{-(n-1)}.$$

反之，若$x^\star$的相对误差限满足

$$E_r^\star(x) \le \frac{1}{2(\alpha_1+1)} \times 10^{-(n-1)}.$$

则$x^\star$至少具有n位有效数字。

有效数字的位数越多，相对误差限就越小。

(2) 若近似值$x^\star$具有n位有效数字，则其相对误差限为

$$E_r^\star(x) \leq \frac{1}{2\alpha_1} \times 10^{-(n-1)}.$$

反之，若$x^\star$的相对误差限满足

$$E_r^\star(x) \leq \frac{1}{2(\alpha_1+1)} \times 10^{-(n-1)}.$$

则$x^\star$至少具有n位有效数字。

有效数字的位数越多，相对误差限就越小。

一、选用数值稳定的计算公式，控制舍入误差的传播

若算法不稳定，则数值计算的结果就会严重背离数学模型的真实结果。因此在选择数值计算公式来进行近似计算时，应特别注意选用那些在计算过程中不会导致误差迅速增长的计算公式。

例

计算积分

$$I_n = e^{-1}\int_0^1 x^n e^x\,dx,\quad n = 0, 1, 2, \cdots$$

算法 (1)

$$\begin{cases} I_n = 1 - nI_{n-1}, \\ I_0 = 1 - e^{-1} \approx 0.6321. \end{cases}$$

例

计算积分

$$I_n = e^{-1} \int_0^1 x^n e^x \, dx, \quad n = 0, 1, 2, \cdots$$

算法 (1)

$$\begin{cases} I_n = 1 - nI_{n-1}, \\ I_0 = 1 - e^{-1} \approx 0.6321. \end{cases}$$

matlab代码

```
t0 = 0.6321;
for i = 1:9
  fprintf('%10.5f␣', t0);
  if(mod(i,3)==0)
    fprintf('\n');
  end
  t1 = 1 - i * t0;
  t0 = t1;
end
```

运行结果

```
0.63210    0.36790    0.26420
0.20740    0.17040    0.14800
0.11200    0.21600   -0.72800
```

matlab代码

```
t0 = 0.6321;
for i = 1:9
  fprintf('%10.5f␣', t0);
  if(mod(i,3)==0)
    fprintf('\n');
  end
  t1 = 1 - i * t0;
  t0 = t1;
end
```

运行结果

```
0.63210     0.36790     0.26420
0.20740     0.17040     0.14800
0.11200     0.21600    -0.72800
```

由

$$0 < I_n < e^{-1} \max_{0\le x\le 1}(e^x) \int_0^1 x^n \, dx = \frac{1}{n+1}$$

知

$$I_7 < \frac{1}{8} = 0.1250, \quad I_8 < \frac{1}{9} \approx 0.1111,$$

原因：

I_0本身有不超过0.5×10^{-4}的舍入误差，此误差在运算中传播、积累很快，传播到I_7和I_8时，该误差已放大了7与8倍，从而使得I_7和I_8的结果面目全非。

由

$$0 < I_n < e^{-1} \max_{0 \le x \le 1}(e^x) \int_0^1 x^n \, dx = \frac{1}{n+1}$$

知

$$I_7 < \frac{1}{8} = 0.1250, \quad I_8 < \frac{1}{9} \approx 0.1111,$$

原因：

I_0本身有不超过0.5×10^{-4}的舍入误差，此误差在运算中传播、积累很快，传播到I_7和I_8时，该误差已放大了7与8倍，从而使得I_7和I_8的结果面目全非。

算法 (2)

$$I_{n-1} = \frac{1}{n}(1 - I_n),$$

由

$$I_n > e^{-1} \min_{0 \le x \le 1} (e^x) \int_0^1 x^n \, dx = \frac{e^{-1}}{n+1}$$

知

$$\frac{e^{-1}}{n+1} < I_n < \frac{1}{n+1}.$$

$$I_7 \approx 0.1124.$$

算法 (2)

$$I_{n-1} = \frac{1}{n}(1 - I_n),$$

由

$$I_n > e^{-1} \min_{0 \le x \le 1} (e^x) \int_0^1 x^n \, dx = \frac{e^{-1}}{n+1}$$

知

$$\frac{e^{-1}}{n+1} < I_n < \frac{1}{n+1}.$$

$$I_7 \approx 0.1124.$$

matlab代码

```
t0 = 0.1124;
for i = 7:-1:0
  fprintf('%10.5f␣', t0);
  if(mod(7-i+1,3)==0)
    fprintf('\n');
  end
  t1 = (1 - t0) / i;
  t0 = t1;
end
```

运行结果

```
0.11240     0.12680     0.14553
0.17089     0.20728     0.26424
0.36788     0.63212
```

matlab代码

```
t0 = 0.1124;
for i = 7:-1:0
  fprintf('%10.5f␣', t0);
  if(mod(7-i+1,3)==0)
    fprintf('\n');
  end
  t1 = (1 - t0) / i;
  t0 = t1;
end
```

运行结果

```
0.11240    0.12680    0.14553
0.17089    0.20728    0.26424
0.36788    0.63212
```

定义 (数值稳定)

在数值计算中，误差不会增长的计算格式称为是数值稳定的，否则就是不稳定的。

二、尽量简化计算步骤以便减少运算次数

节省计算量，提高计算速度，简化逻辑结构，减少误差积累。

二、尽量简化计算步骤以便减少运算次数

节省计算量，提高计算速度，简化逻辑结构，减少误差积累。

例

计算多项式

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

- 逐项计算

  共需

  $$1 + 2 + \cdots + (n-1) + n = \frac{1}{2} n(n+1)$$

  次乘法和n次加法

- 秦九韶算法

  $$\begin{cases} u_0 = a_n, \\ u_k = u_{k-1} x + a_{n-k}, \quad k = 1, 2, \cdots, n. \end{cases}$$

  共需n次乘法和n次加法。

例

计算多项式

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

- 逐项计算

  共需

  $$1 + 2 + \cdots + (n-1) + n = \frac{1}{2} n(n+1)$$

  次乘法和n次加法

- 秦九韶算法

  $$\begin{cases} u_0 = a_n, \\ u_k = u_{k-1} x + a_{n-k}, \quad k = 1, 2, \cdots, n. \end{cases}$$

  共需n次乘法和n次加法。

三、避免两个相近的数相减

数值计算中，两个相近的数相减会造成有效数字的严重丢失

处理办法：

- 因式分解
- 分子分母有理化
- 三角函数恒等式
- Taylor展开式
- $\cdots$

三、避免两个相近的数相减

数值计算中，两个相近的数相减会造成有效数字的严重丢失

处理办法：

- 因式分解
- 分子分母有理化
- 三角函数恒等式
- Taylor展开式
- $\cdots$

三、避免两个相近的数相减

数值计算中，两个相近的数相减会造成有效数字的严重丢失

处理办法：

- 因式分解
- 分子分母有理化
- 三角函数恒等式
- Taylor展开式
- $\cdots$

例

计算（取4位有效数字）

$$\sqrt{x+1}-\sqrt{x} \quad (x=1000)$$

- 直接计算
$$\sqrt{1001}-\sqrt{1000}\approx 31.64-31.62=0.02$$
只有一个有效数字，损失了三位有效数字
- 分子有理化
$$\sqrt{x+1}-\sqrt{x}=\frac{1}{\sqrt{x+1}+\sqrt{x}}\approx 0.01581$$
没有损失有效数字

例

计算（取4位有效数字）

$$\sqrt{x+1}-\sqrt{x} \quad (x=1000)$$

- 直接计算

$$\sqrt{1001}-\sqrt{1000} \approx 31.64-31.62=0.02$$

只有一个有效数字，损失了三位有效数字

- 分子有理化

$$\sqrt{x+1}-\sqrt{x}=\frac{1}{\sqrt{x+1}+\sqrt{x}} \approx 0.01581$$

没有损失有效数字

1.3 选用和设计计算法应遵循的原则

例

计算（取4位有效数字）

$$\sqrt{x+1}-\sqrt{x}\quad (x=1000)$$

- 直接计算

$$\sqrt{1001}-\sqrt{1000}\approx 31.64-31.62=0.02$$

只有一个有效数字，损失了三位有效数字

- 分子有理化

$$\sqrt{x+1}-\sqrt{x}=\frac{1}{\sqrt{x+1}+\sqrt{x}}\approx 0.01581$$

没有损失有效数字

例

计算（取4位有效数字）

$$A = 10^7(1 - \cos 2^\circ) \quad (\cos 2^\circ = 0.9994)$$

- 直接计算

$$A \approx 10^7(1 - 0.9994) = 6 \times 10^3$$

只有一个有效数字

- 三角恒等式

$$\boxed{1 - \cos x = 2\sin^2\frac{x}{2}}$$

$$\begin{aligned} A &= 10^7(1 - \cos 2^\circ) = 2 \times (\sin 1^\circ)^2 \times 10^7 \\ &\approx 2 \times 0.01745^2 \times 10^7 \approx 6.09 \times 10^3 \end{aligned}$$

三位有效数字

例

计算（取4位有效数字）

$$A = 10^7(1 - \cos 2^\circ) \quad (\cos 2^\circ = 0.9994)$$

- 直接计算

$$A \approx 10^7(1 - 0.9994) = 6 \times 10^3$$

只有一个有效数字

- 三角恒等式

$$\boxed{1 - \cos x = 2\sin^2\frac{x}{2}}$$

$$\begin{aligned} A &= 10^7(1 - \cos 2^\circ) = 2 \times (\sin 1^\circ)^2 \times 10^7 \\ &\approx 2 \times 0.01745^2 \times 10^7 \approx 6.09 \times 10^3 \end{aligned}$$

三位有效数字

例

计算（取4位有效数字）

$$A = 10^7(1 - \cos 2^\circ) \quad (\cos 2^\circ = 0.9994)$$

- 直接计算

$$A \approx 10^7(1 - 0.9994) = 6 \times 10^3$$

只有一个有效数字

- 三角恒等式

$$\boxed{1 - \cos x = 2\sin^2\frac{x}{2}}$$

$$\begin{aligned} A &= 10^7(1 - \cos 2^\circ) = 2 \times (\sin 1^\circ)^2 \times 10^7 \\ &\approx 2 \times 0.01745^2 \times 10^7 \approx 6.09 \times 10^3 \end{aligned}$$

三位有效数字

四、绝对值太小的数不宜做除数

数值计算中，用绝对值很小的数作除数，会使商的数量级增加，甚至在计算机中造成“溢出”停机，而且当很小的除数稍有一点误差，会对计算结果影响很大。

例

$$\frac{3.1416}{0.001} = 3141.6$$

$$\frac{3.1416}{0.001 + 0.0001} = 2856$$

四、绝对值太小的数不宜做除数

数值计算中，用绝对值很小的数作除数，会使商的数量级增加，甚至在计算机中造成“溢出”停机，而且当很小的除数稍有一点误差，会对计算结果影响很大。

例

$$\frac{3.1416}{0.001} = 3141.6$$

$$\frac{3.1416}{0.001 + 0.0001} = 2856$$

五、合理安排运算次序，防止“大数吃小数”

例

计算a, b, c的和，其中$a = 10^{12}$, $b = 10$, $c \approx -a$.

- $(a + b) + c$

  结果接近于0

- $(a + c) + b$

  结果接近于10

五、合理安排运算次序，防止“大数吃小数”

例

计算a, b, c的和，其中$a = 10^{12}$, $b = 10$, $c \approx -a$.

- $(a + b) + c$

  结果接近于0

- $(a + c) + b$

  结果接近于10

五、合理安排运算次序，防止“大数吃小数”

例

计算a, b, c的和，其中$a = 10^{12}$, $b = 10$, $c \approx -a$.

- $(a + b) + c$

  结果接近于0

- $(a + c) + b$

  结果接近于10