# 第1章 数值分析引论

## 1.1 数值分析的作用和内容

**1.1.1 数值分析的地位**

一般来说, 解决科学技术和工程问题的基本步骤如图1-1所示：

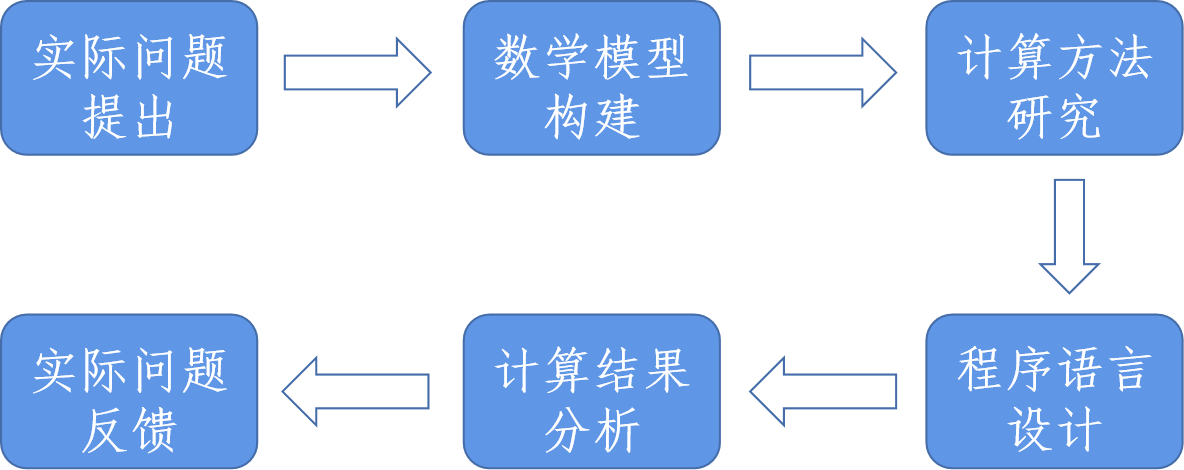


图1-1 问题的解决流程

具体地, 例如, 完成某一地区地形图的绘制通常需要以下的过程：

(1) 用空中航测方法, 在该地区上空连续拍照；

(2) 为形成三维地形图, 建立一个大型超定线性方程组；

(3) 采用最小二乘方法求解该方程组的最小二乘解, 然后再整体平滑；

(4) 形成一个大型程序, 用计算机绘制出地形图.

从以上的步骤和过程可以看出, 数学是解决科学技术问题的基础. 最近半个世纪以来, 由于计算机及科学技术的飞跃发展, 使用计算机采用各种计算方法或手段去解决科学技术和工程中的各类数学问题已成为解决问题的关键环节, 逐步形成了新的计算性交叉学科—科学计算. 目前, 科学计算已与理论研究和科学实验成为现代科学发展的三种主要方法. 计算数学是研究用计算机求解各种数学问题的数值计算方法、过程及其理论的一个数学分支, 是科学计算的基础, 能为科学计算提供了强有力的数学理论支撑. 因此, 作为计算数学学科的基础课程之一, 数值分析, 也越来越受到人们的普遍重视.

**1.1.2 数值分析的主要内容**

数值分析课程主要介绍计算数学中最基本、最常用的数值计算方法及其误差分析、收敛性和稳定性等理论, 具体内容包括以下四个部分：

(1) 数值代数：求解线性和非线性方程的解法；

(2) 插值和数值逼近；

(3) 数值微分和数值积分；

(4) 常微分方程数值解法.

数值分析课程是一门内容丰富, 研究方法深刻, 有自身理论体系的课程, 其突出特点是把数学理论、数值方法和计算机使用紧密结合起来, 既有纯数学高度的抽象性和理论性, 又有实际使用的应用性和可操作性.

## 1.2 误差的来源和基本概念

**1.2.1 误差的来源**

一个量的真实值和它的实测值往往是不相等的, 它们之差称为误差. 误差是无处不在的, 引起误差的原因也有很多.

(1) 模型误差

在建立数学模型过程中, 不可能将所有因素均考虑, 必然要进行必要的简化并引入合理的假设, 这就带来了与实际问题的误差. 这种数学模型与实际问题之间出现的误差称为模型误差.

(2) 观测误差

在给出的数学模型中往往涉及一些根据观测得到的物理量, 如长度、温度、速度和电压等, 显然观测不可避免地会产生误差（插入题目1：判断题）. 这种由观测而产生的误差就称为观测误差.

(3) 截断误差

在计算中常常会遇到只有通过无限过程才能得到最终结果, 但实际计算时只能采用有限过程. 这种有限过程代替无限过程产生的误差, 称为截断误差. 例如, 可微函数可进行如下泰勒展开



若用展开式的前项



来近似代替, 就会产生截断误差.

(4) 舍入误差

在计算中遇到的数据可能位数很多, 但由于计算机的字长有限, 原始数据在计算机上表示时, 一般需要进行四舍五入, 由此产生的误差称为舍入误差（插入题目2：判断题）. 例如, 用1.414近似代替, 产生的误差



就是舍入误差.

由误差来源的分析可知：误差是不可避免的, 要求绝对精确和绝对严格是办不到的（插入题目3：思考题）. 进一步地, 前两种误差是客观存在的, 后两种误差是由计算方法所引起的. 研究计算结果的误差是否满足精度要求就是误差估计问题. 本课程是研究数学问题的数值算法, 因此只涉及算法的截断误差和舍入误差. （插入题目4：选择题）

**1.2.2 误差的基本概念**

(1) 绝对误差和绝对误差限

设是精确值, 是它的一个近似值, 称是近似值的**绝对误差**, 简称误差. 注意, 绝对误差不是误差的绝对值, 它可以是正的值, 也可以是负值. 注意到精确值虽然是客观存在的, 但在实际计算中很难事先得到, 故误差是无法准确计算的, 但可以估计出它的一个上界（当然, 总是希望这个上界越小越好）, 即. 这里称为近似值的误差限, 即有. 例如, 用毫米刻度的米尺去测量一长度为的物体, 读出的刻度为, 但由于观测误差的存在, 是的近似值, 它的误差限一般为该尺最小刻度的半个单位0.5mm, 即有mm.

(2) 相对误差和相对误差限

绝对误差的大小还不能完全表示近似值的好坏, 还应考虑精确值本身的大小和量级. 我们把近似值的绝对误差和精确值的比值, 即, 称之为近似值的**相对误差**, 记作. 相对误差是个相对数, 是无量纲的. 若相对误差有估计, 则称为相对误差限, 即

.

下面的例子就说明了绝对误差的局限性和给出相对误差的必要性.

**例1.1** 计算如表1-1所示数值的绝对误差和相对误差.

表1-1 误差对比表

|  |  |  |  |
| --- | --- | --- | --- |
| 精确值 | 近似值 |  |  |
| 10 | 9 | 1 | 0.1 |
| 100 | 99 | 1 | 0.01 |

精确值10和100的近似值分别为9和99. 显然, 近似值9和99的绝对误差都是为1. 因此, 无法通过绝对误差来判断9和99近似它们精确值的好坏. 进一步地, 近似值9和99的相对误差分别为0.1和0.01, 从而可知近似值99近似精确值100的准确程度要好. （插入题目5：思考题）

众所周知, 在实际计算中, 精确值往往是不知道的, 人们通常取



和



分别作为近似值的相对误差和相对误差限.

(3) 有效数字

在前面的讨论中, 不难发现, 可以通过误差限的大小来刻画近似值逼近精确值的准确程度. 但是在实际计算时, 若把这个误差限也参与到具体的计算中, 会给计算造成困扰, 也不切实际. 因此希望所表示的近似值本身就能显示出它的准确程度, 这就需要引入有效数字的概念.

**定义1.1** 如果近似值的误差限是某一数位上的半个单位, 且从该位直到的第一个非零数字为止, 一共具有个数位, 则称具有位**有效数字**（插入题目6：判断题）.

**例1.2** 现有的三个近似值, , , 问它们分别具有几位有效数字？

**解：**取一个相对准确的, 如取. 注意, 因为是无理数, 在选取该相对准确值时, 要使该值的小数点后的位数稍多于所考虑的近似值中小数点的位数.

现考虑近似的有效数字情况, 则

,

误差小于小数点后第2位的半个单位. 现在对近似值的小数点后第2位开始数, 一直到的第1个非零数字3, 共有3个数位, 则根据有效数字的定义可以, 具有3位有效数字；

当时, 则, 误差中的“7”已超过了它前一位的半个单位, 因此要对作四舍五入为, 这就小于了小数点后第4位的半个单位, 故可知：具有5位有效数字；

当时, 则. 这就小于了小数点后第3位的半个单位, 所以具有4位有效数字.

事实上, 是对小数点后的第5位进行了四舍五入, 而是对小数点后的第5位进行了直接截断. 从而可知, 在对数据进行近似处理时, 四舍五入比直接截断更合理.

进一步地, 也可以借助科学记数法来判断一个近似值的有效数字. 设精确值的近似值, 即

,

其中, 是0到9中的一个数字, , 为正整数. 若具有位有效数字, 则根据有效数字的定义, 有

.

因此, 只要把的绝对误差限表示成以上形式, 中的即为的有效数字.

（插入题目7：思考题）

**例1.3** 设, 下列

, , 

都是的近似值. 试求 , , 的有效数字.

**解：** 将进行改写

, , ,

则有

,

,

,

因此, 有4位有效数字, 有4位有效数字, 有5位有效数字.

**定理1.1** 设近似值, 有**位有效数字, . 则其相对误差限.

**证明：**, 故, 那么 .

此定理说明, 相对误差限也是由有效数字所决定. 一个近似值的有效数字越多, 其相对误差也越小. （插入题目8：判断题）

**定理1.2** 设近似值的相对误差限不大于, 则它至少有位有效数字.

**证明：**.

**例1.4** 重力加速度常数, , 两者均有3位有效数字.  , , 后者的绝对误差大. 而由定理1.2, 相对误差限分别为和, 两者相等, 与量纲的选取无关.

**例1.5** 用4位浮点数计算.

**解：** , 结果只有1位有效数字, 有效数字大量损失, 造成相对误差扩大. 这是由两个比较接近的数相减造成的. 而



结果仍然有4位有效数字. 这说明在设计算法时, 应尽量避免两个比较接近的数相减.

**例1.6** 设是的近似值. 试确定至少有几位有效数字, 才能使相对误差范围不超过0.01%.

**解：** 假设至少有*n*位有效数字, 由知, 根据定理1.1有

.

由上式可得. 因此, 即至少有4位有效数字. （插入题目9：选择题）（插入题目10：思考题）

## 1.3 数值计算中的若干准则

计算机在进行数值运算时, 几乎每步运算都会产生舍入误差. 然而对于一个实际计算问题, 往往要进行成千上万次运算, 当然不可能每步运算都作误差分析, 而是转化为数据误差对计算结果影响的分析. 因此, 在设计数值计算中的算法时, 必需要考虑算法在运算过程中能控制好误差, 确保计算结果的精度.

**1.3.1 算法的数值稳定性**

**定义1.2 设有**一个算法, 如果初始数据有小的误差, 仅使最终计算结果产生小的误差, 则称该算法是数值稳定的；否则就称此算法是数值不稳定的.

**例1.7** 建立计算积分



的递推公式, 并研究其误差传播.

**解：** 



, 

及



从而得到计算的递推关系：

 .

在具体计算时, 由于是无理数, 可取具有6位有效数字的近似值. 设表示的近似值, 则实际计算公式为：

  (1.3.1)

由上式计算得到的结果见表1-2.

表1-2 递推序列的计算值

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 0.0883900 | 0.0580500 | 0.0430833 | 0.0345835 | 0.0270825 |
|  |  |  |  |  |
| 0.0312542 |  | 0.192070 |  | 4.34620 |

由于且单调递减（插入题目1：思考题）, 上面的计算结果显然有误差. 因为, , . 现在来分析计算结果发生错误的原因. 事实上, 记, (), 则可推得

,

即

.

因为在计算有误差, 由上式可以看出误差在经过步传播后的误差扩大了倍, 所以递推式(1.3.1)是不稳定的.

现在使用下标从大到小的次序来进行递推, 即有递推公式为

 (1.3.2)

利用上式只要算出的近似值, 就可以算出其它值. 类似于上面的推导, 可得递推公式(1.3.2)的误差传播关系, 即有

, ,

或



即每递推一步误差没有增加, 反而缩小到原值的, 所以递推式(1.3.2)是稳定的.

下面来计算的近似值. 由积分第一中值定理得



从而有



可取



因此, 有



该近似值的绝对误差限可估计为



（插入题目2：选择题）

计算结果如表1-3所示：

表1-3 递推关系生成的序列值

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 0.0166667 | 0.0188889 | 0.02112222 | 0.0243027 | 0.0284679 |
|  |  |  |  |  |
| 0.0343064 | 0.0431387 | 0.0580389 | 0.0883922 | 0.1823216 |

这里计算得到的的近似值, 该近似值已具有6位有效数字. 由此例可以看出：算法的好坏对计算结果有很大的影响, 一定要设计出数值稳定的算法用于实际计算. （插入题目3：选择题）（插入题目4：判断题）

**1.3.2 问题本身的性态**

对一个数值问题本身如果输入数据有微小扰动（即误差）, 引起输出数据（即问题解）相对误差很大, 这就是病态问题. 例如, 计算函数值时, 若有扰动, 其相对误差为, 函数值的相对误差为, 相对误差比值

 （1.3.3）

称为计算函数值问题的条件数, 自变量相对误差一般不会太大, 如果条件数很大, 将引起函数值相对误差很大, 出现这种情况的问题就是病态问题.

例如, 取, 则有, 它表示相对误差可能放大倍, 如, 有, , 若取自变量相对误差为2%, 函数值相对误差为24%, 这时问题可以认为是病态的, 一般情况下, 条件数就认为是病态, 越大病态越严重.（插入题目5：选择题）

**例1.8** 求解线性方程组

 （1.3.4）

**解** 当时, 系数行列式为零, 方程无解, 但当时解为当时, 若输入数据有微小扰动（误差）, 则解的误差很大. 例如, 取, 则解；如果有误差0.001, 取, 则解, 误差很大, 表明此时线性方程组（1.3.4）是病态的.实际上, 由是的函数, 利用（1.3.3）式可求得



当时, 表明条件数很大, 故问题是病态的.

注意病态问题不是计算方法引起的, 是数值问题本身固有的, 因此, 对数值问题首先要分清问题是否病态, 对病态问题就必须采取相应的特殊方法以减少误差危害. （插入题目6：判断题）

**1.3.3 简化计算步骤, 减少运算次数**

同样一个计算问题, 如果能减少运算次数, 不但可以节省计算机的计算时间, 而且还能减少舍入误差, 这是数值计算必须遵守的原则, 也是数值计算所需研究的重要内容. （插入题目7：判断题）

**例1.9** 计算的值.

**解** 若将的值逐个相乘, 那么需做21次乘法, 但若写成



其中, 只要做7次乘法就可以了（插入题目8：思考题）.

又如计算多项式



的值, 若直接计算再逐项相加, 总共需做



次乘法和次加法, 但若将前项提出, 则有



于是括号内次多项式, 对它再施行同样手续, 又有



对内层括号内次多项式再施行上述同样手续, 又得一个次多项式, 这样每作一步, 最内层得多项式就降低1次, 最终可将多项式表示为如下嵌套形式：



利用此式结构上的特点, 从里往外一层层地计算, 设



得递推公式



于是.此即秦九韶算法（秦九韶是宋代数学家, 此法由他最早提出, 国外称此法为Hornor法, 其实Hornor比秦九韶晚了五六百年）. 按此法求的值只需作次乘法和次加法, 计算量少, 由于此公式是递推公式, 因此极易编制程序.

若采用计算器计算或手算也是极方便的. 我们把按照降幂排列的系数写在第一行, 把欲求某点之值及写在第二行, 第三行为一、二两行相应值之和, 最后得到的即为所求之值（如表1-4所示）.

表1-4 计算的步骤表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**例1.10** 已知, 用秦九韶算法求.

**解** 如表1-5所示：

表1-5 的求值过程表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 8 | 0 | 4 | 0 | -9 | 1 |
|  |  | 24 | 72 | 228 | 684 | 2025 |
|  | 8 | 24 | 76 | 228 | 675 |  |

可得.

**1.3.4 数值计算中的一些其它注意事项**

(1) 避免除数的绝对值远小于被除数的绝对值

, 当时, 舍入误差会扩大. 事实上, 设和, 且分别是的近似值, 则



现取, 故有,



从而有,



**例1.11** 近似值****的舍入误差均为, 而, 则的舍入误差为：. 很小的数作除数有时还会造成计算机的溢出而停机.

(2) 防止大数“吃”小数

大数“吃”小数是指一个绝对值较大的数加上一个绝对值较小的数时, 计算结果仍等于这个绝对值较大的数. 例如： 用8位字长的计算机计算时, 若采用的过程来计算, 则有

,

这里还是等于, 这个就像被大数“吃”掉了一样, 这种现象就叫做：大数“吃”小数. 如按, 就没有被吃掉. 这也是构造算法时要注意的问题, 尽量要使数量级相同或相差不大的数优先相加减. （插入题目9：判断题）

(3) 计算速度和存储量

计算速度是设计一个算法时必须要考虑的因素. 如果一个算法的计算速度太慢, 会导致解决问题的时间成本变长, 算法因此会没有实际使用价值. 例如, 用线性代数中众所周知的克莱姆法则, 来求解一个20阶线性方程组, 需要进行次乘除法运算, 如用每秒1亿次乘除法运算的计算机来计算, 需要耗时约30万年. 但如用Gauss消元法（见3.2节）, 则只需3000次乘除法运算.

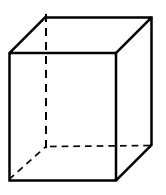
另外, 在计算大型问题时, 算法的存储量必须要有所考虑. 存储量大的算法会极大地耗费计算机的内存, 甚至超出计算机的内存, 从而会出现“死机”.

（插入题目10：判断题）

## 习题

1. 求具有四位有效数字的近似值, 并估计绝对误差和相对误差.
2. 取近似值, 那么有多少位有效数字？
3. 已知准确值与其有位有效数字的近似值的绝对误差限, 试求其近似值.
4. 下列各都是经过四舍五入得到的近似数, 即误差限不超过最后一位的半个单位, 试分别指出其绝对误差限、相对误差限以及有效数字位数.

, , , , ,

1. 为使的近似值的相对误差为0.5%, 试问至少有几位有效数字？
2. 已知, 请用秦九韶算法求.
3. 计算球体的表面积, 为使的相对误差不超过0.2%, 求半径的相对误差的允许范围.
4. 已知, 是经过四舍五入后得到的近似值, 问, 有几位有效数字.
5. 设, 的相对误差为, 求的绝对误差和相对误差.
6. 测得某长方体（见右图）高的值为, 底面是边长为的正方形, . 已知, , 求长方体表面积的绝对误差限与相对误差限.
7. 序列满足：

, , 

取, 计算到时的误差有多大？计算过程是否稳定？

1. 对于一元二次方程, 如果具有4位有效数字, 求其具有4位有效数字的根.
2. 设, , , , 已知, 分别是, 的具有6位有效数字的近似值. 计算, 现有下面两种算法：

第一种算法：

第二种算法：

试分析上述两种算法所得结果的有效数字.

1. 已知多项式, 用秦九韶算法求时多项式的值；当取的近似值时, 求该多项式的绝对误差并判断该函数问题是否病态.
2. 对于积分, 分析递推式是否稳定.