化工数值计算与MATLAB

隋志军
杨　榛◎编著
魏永明

華東理工大學出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

化工数值计算与 MATLAB

隋志军　杨　榛　魏永明　编著

華東理工大學出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS
·上海·

图书在版编目(CIP)数据

化工数值计算与 MATLAB/隋志军,杨榛,魏永明编著.
—上海:华东理工大学出版社,2015.2
ISBN 978-7-5628-4111-1

Ⅰ.①化… Ⅱ.①隋…②杨…③魏… Ⅲ.①Matlab 软件—应用—化工计算—数值计算 Ⅳ.①TQ015.9

中国版本图书馆 CIP 数据核字(2014)第 280945 号

内容提要

全书共分为 10 章,第 1 章为 MATLAB 程序设计语言与初等数学运算,第 2 章为矩阵操作与线性方程组求解,第 3 章为非线性方程组求解,第 4 章为插值与拟合,第 5 章为数值微分与数值积分,第 6 章为常微分方程数值解,第 7 章为偏微分方程数值解,第 8 章为概率论与数理统计,第 9 章为数值最优化方法,第 10 章为神经网络。

本书可作为高等院校化学工程、化学工艺及相关专业的本科生教材或研究生参考书,也可供化工科研、工程技术人员参考。

化工数值计算与 MATLAB

编　　著 / 隋志军　杨　榛　魏永明
责任编辑 / 焦婧茹
责任校对 / 李　晔
封面设计 / 肖祥德　裘幼华
出版发行 / 华东理工大学出版社有限公司
地　址:上海市梅陇路 130 号,200237
电　话:(021)64250306(营销部)
(021)64252344(编辑室)
传　真:(021)64252707
网　址:press.ecust.edu.cn
印　　刷 / 上海展强印刷有限公司
开　　本 / 787mm×1092mm　1/16
印　　张 / 21.75
字　　数 / 540 千字
版　　次 / 2015 年 2 月第 1 版
印　　次 / 2015 年 2 月第 1 次
书　　号 / ISBN 978-7-5628-4111-1
定　　价 / 49.00 元

联系我们:电子邮箱 press@ecust.edu.cn
官方微博 e.weibo.com/ecustpress
淘宝官网 http://shop61951206.taobao.com

本书的使用说明

本书适用于化工类及相关专业的本科、研究生和科研人员，特别是数值计算和 MATLAB 为零基础的读者。本书提供了较好的入门知识。对于这部分内容比较熟悉的读者，可以跳过本书的绪论和第 1 章及其他章开始部分的介绍性内容。

本书以数值方法的类型为主线，内容兼顾数值方法、MATLAB 语言和化工计算几个方面。限于篇幅限制，不可能面面俱到，我们的处理方法是：对于数值方法部分只给出各种算法的基本思路，这样可以介绍一些与 MATLAB 函数密切相关的先进算法，从而有助于读者更好地使用这些函数；这部分内容忽略数值计算算法的实现细节及证明等，如果读者对这些内容感兴趣，可以参见相关参考文献；对于 MATLAB 语言，我们详细介绍了与本书计算任务相关的各种细节，对于重要的函数我们均提供详细的使用方法和具体实例。一些复杂的函数，如 ode45、fsolve、lsqcurvefit 等，一般先以简单的纯数学方程为例给出函数的使用方法，然后结合具体的应用实例给出一些使用技巧；对于专业问题的选择，本书例题涵盖了化工原理（分离工程）、化学反应工程、化工热力学（物理化学）、化工设计各方面的内容，这些内容根据求解方法的不同分散于各章中，如表 1 所示。

表 1　本书内容根据求解方法进行大体归纳

问题分类	问题描述	数值计算问题	MATLAB 关键求解函数
化工原理，分离工程			
流体流动	圆管流动阻力	代数运算，第 1 章	
沉降	已知沉降速率求黏度	代数运算，第 1 章	
传热	第二类操作型命题，求冷流体出口温度和流量	非线性方程求解，第 3 章	fzero
萃取	平衡级式分离设备求解，MESH 方程	非线性方程组，第 3 章	fsolve
精馏	MaCabe-Thiele 法求二元精馏的理论板数	插值，第 4 章	interp1
物理化学，化工热力学			
物性计算	比热容计算	代数运算，第 1 章	
流体的热力学性质计算	状态方程（RK）的计算，已知 V、T 求 p	代数运算	—
	状态方程（RK，SRK，PR，范德瓦尔斯）的求解——已知 p、T 求 V	非线性（多项式）方程的求解，第 3 章	fzero roots
	真实气体逸度	数值积分，第 5 章	quadl
	气体的等压温升	数值积分，第 5 章	quadl
相平衡	理想体系的泡点温度与平衡组成	非线性方程，第 3 章	fzero
化学平衡	平衡常数法计算化学平衡组成	非线性方程求解，第 3 章	fsolve

续表

问题分类	问题描述	数值计算问题	MATLAB 关键求解函数
	吉布斯自由能最小计算化学平衡组成	非线性有约束优化，第 9 章	fmincon
化学反应工程			
化学计量学	独立反应数和独立反应的确定	矩阵求秩和线性无关向量组查找，第 2 章	rank
化学反应动力学	反应器的物料平衡	非线性有约束优化，第 9 章	fmincon
	数值微分求近似反应速率	数值微分，第 5 章	
	微分法催化反应动力学参数确定，参数检验	非线性最小二乘回归，第 9 章	lsqnonlin
	积分法催化反应动力学参数确定	非线性最小二乘回归，参数的顺序回归，第 9 章	lsqcurvefit
	催化反应动力学模型判别和参数回归的序贯实验设计	序贯实验设计，全局优化函数，非线性最小二乘法回归，第 9 章	ga lsqcurvefit
反应器模拟	稳定态连续搅拌釜式反应器，二级反应	代数计算，第 1 章	
	瞬态连续搅拌釜式反应器模型	常微分方程初值问题，第 6 章	ode45
	固定床反应器一维拟均相模型的求解	常微分方程初值问题，第 6 章	ode45
	绝热连续搅拌釜式反应器	常微分方程初值问题，第 6 章	ode15s
	平推流反应器的停留时间	常微分方程初值问题，第 6 章	ode45
	考虑颗粒界面梯度的一维非均相模型	代数微分方程，第 6 章	ode15s
	在球形催化剂内某组分的扩散-反应过程	常微分方程边值问题，第 6 章	bvp4c
	固定床反应器一维拟均相轴向分散模型	常微分方程边值问题，第 6 章	bvp4c
	固定床反应器的二维拟均相模型	偏微分方程，第 7 章	pdepe
化工过程开发、设计，化工优化			
稳态流程模拟	简单流程	线性方程组，第 2 章	—
	序贯模块法	代数运算，第 3 章	fsolve
	联立求解法	非线性方程组，第 3 章	fsolve
传递现象			
	一维扩散问题	偏微分方程，第 7 章	pdepe
	瞬态热传导	偏微分方程，第 7 章	pdepe
	水平圆管中流体的瞬态流动	偏微分方程，第 7 章	pdepe
	矩形管内的流动	偏微分方程，第 7 章	pde 工具箱 GUI
	固体传热	偏微分方程，第 7 章	pde 工具箱 GUI

但是，化工计算涉及的内容和体系千差万别，无法一一列举。由于必要的练习是掌握这些内容的必经之路，书中也提供了一些习题，这些习题难度与书中的例题类似；更多的以及更面向实际过程的（当然也更复杂）专业计算问题将以上机实践的形式给出（相关内容正在编写中），目前部分内容可以通过如下网址获取：

http://unilab.ecust.edu.cn/cre/cecomputing

前　言

“计算机化工应用”是化学工程与工艺及相关专业一门重要的专业课程，其目的是强化本专业学生应用计算机解决专业问题的能力。这一课程在华东理工大学已开设30年，当时计算机在化工领域的应用主要是数值计算，这也一直是本门课程的核心内容。在多年的实践中，本课程使用的编程语言从开始的Fortran变化到后来的C语言，但核心内容没有明显变化。在21世纪里，随着计算机软硬件技术的飞速发展，化工领域出现了很多优秀的专业软件，使完成传统化工过程设计的工作大大简化。在这种形势下，本门课程的内容也面临调整的契机。

2006年起，华东理工大学化工学院进行了关于化工专业课程中计算机类课程的改革，决定将MATLAB作为“计算机化工应用”课程的编程语言。MATLAB软件是一款优秀的数值计算软件，利用这一软件有助于学生掌握复杂数学问题的求解方法。随着学生求解复杂模型能力的提高，相关专业课程的内容也可以更加深入并接近实际应用。同时本课程还能培养学生数学思维、兴趣以及加强建模的能力，MATLAB语言的学习对这些能力的提高也很有帮助。

合适的教材对于课程教学目的的实现十分重要。在本课程建设之初，我们调查了国内外相关课程使用教材的情况。在国内，华东理工大学黄华江老师编著的《实用化工计算机模拟——MATLAB在化学工程中的应用》、朱开宏教授编著的《化学反应工程分析例题与习题(MATLAB版)》是最相关的两本书籍，两者都有大量丰富的实例，但美中不足的是不适合零基础学生的学习；其他关于MATLAB的书籍则有的偏重数值分析内容，或者专业方向与本专业相差较远。在国外类似的课程中，则大多数以数值分析的内容为主，如K. J. Beers编著的《Numerical Methods for Chemical Engineering-Application in MATLAB》，A. Contantinides和N. Mostoufi编著的《Numerical Methods for Chemical Engineers with MATLAB Application》等。基于这种情况，我们决定编写一本适合零基础学生学习的教材。

本书在编写过程中充分借鉴了国外高校相关课程的内容，在内容取舍上注重全面提炼化工专业学习过程可能涉及的数值计算内容，较好地平衡了数值计算、计算机语言和专业计算三方面的内容，具有自己的特色。从2008年起，本书的前身——《计算机化工应用讲义》开始使用。根据教学情况2011年我们对该讲义进行了改编。这次我们将进一步对全书进行系统的整编。

本书的内容可以分为两部分，从绪论到第7章内容以模型数值求解为目标；而第8章到

第 10 章则以模型，特别是经验性模型的建立为核心展开。在华东理工大学的教学实践中，第 1～7 章作为“计算机化工应用”的授课内容，是学生的必修内容，采用 24 学时课堂教学和 16 学时上机实践完成；第 8～10 章作为“MATLAB 与化工模拟计算”课程内容，供感兴趣的学生选修。

由于作者水平有限，错漏之处在所难免，请各位读者不吝指正。以下是我们的联系方式。

隋志军：zhjsui@ecust. edu. cn

杨　榛：yangzhen@ecust. edu. cn

魏永明：ymwei@ecust. edu. cn

本书编写分工如下。

隋志军：绪论，第 1、3、6、7、8 章，第 9 章第 6 节；

杨　榛：第 4、5、10 章；

魏永明：第 2 章、第 9 章第 1～5 节；

最后由隋志军统稿。

编者感谢朱开宏教授审阅全书和提出的宝贵意见；感谢国家“973”项目(2012CB720500)的资助。

编　者

目　录

绪　论

1. 数值计算及其在化工中的作用

在化工过程研究与开发过程中经常需要求解各种数学模型，其中绝大多数模型均为非线性问题，如以下几个问题的求解。

问题 1　热力学问题：利用 Redlich-Kwong 方程求解比容

Redlich-Kwong 状态方程是范德瓦尔斯方程的修正：

$$p=\frac{RT}{v-b}-\frac{a}{v(v+b)} \tag{1}$$

现需在已知方程参数 a，b 和体系压力 p 和温度 T 时求比容 v。问题 1 是一个关于 v 的三次方程求解问题。

问题 2　反应工程问题：固定床反应器的模拟

已知在固定床反应器中某反应的反应速率 $-r$ 与转化率 x 的关系为

$$(-r)=\frac{0.12}{15.73+x}\times 15\exp\left(\frac{10000}{805-182x^2}\right) \tag{2}$$

反应器的物料衡算方程为

$$\frac{\mathrm{d}x}{\mathrm{d}w}=\frac{(-r)}{9.65} \tag{3}$$

其中 w 为反应器中催化剂的质量，已知反应器进口处转化率 $x=0$，现需求解催化剂质量为 10 kg 时，转化率可以达到多少？这是一个常微分方程求解问题。

问题 3　分离工程问题：二元间歇精馏

对于物质 1 和 2 的混合物进行间歇精馏，液相残余量 L 与组分 2 物质的量 x_2 的关系可以由以下关系式表达：

$$\frac{\mathrm{d}L}{\mathrm{d}x_2}=\frac{L}{x_2(k_2-1)} \tag{4}$$

式中，k_2 是组分 2 的汽液平衡关系。该分离体系可认为是理性体系，k_2 可以通过下式计算：

$$k_i=p_i/p \tag{5}$$

其中 p_i 是组分 i 的蒸气压，而 p 为系统总压。在系统温度为 T 时，组分 i 的蒸气压可采用

三参数(A, B, C)的 Antoine 方程描述：

$$p_i = 10^{\left(A-\frac{B}{T+C}\right)} \tag{6}$$

分离过程中系统温度处于泡点温度，此温度可以由以下隐式关系式确定：

$$k_1 x_1 + k_2 x_2 = 1 \tag{7}$$

现需在 1.2 atm① 的系统压力下采用间歇精馏分离苯(组分 1)和甲苯(组分 2)，初始时液相中含苯 60 mol，甲苯 40 mol。试计算甲苯含量为 80%(质量分数)时，液相残余量为多少？已知苯的 Antoine 方程参数为 $A_1 = 6.90565$, $B_1 = 1211.033$, $C_1 = 220.79$；甲苯为 $A_2 = 6.95464$, $B_2 = 1344.8$, $C_2 = 219.482$，此时计算所需 p 的单位为 mmHg, T 的单位为℃。

这一问题要求解在 $x_2 = 0.8$ 时的液相残余量 L，可以通过求解式(4)的常微分方程获得，但应当注意此间歇精馏过程中液相组成一直在不断变化，因此 k_i 也在不断变化，k_2 和 x_2 的关系可以由式(5)和式(6)组成的非线性方程组表示。由此可见，这一问题的求解是一个常微分方程和非线性方程联合的求解过程。

问题 4　传递过程问题：一维热传导

求解以下一维热传导方程：

$$\frac{\partial u}{\partial t} = \pi^{-2}\frac{\partial^2 u}{\partial x^2},\ x \in [0, 1],\ t \geqslant 0 \tag{8}$$

已知边界条件：

$$u(0, t) = 0 \tag{9}$$

$$\pi \cdot \mathrm{e}^{-t} + \frac{\partial u}{\partial x}(1, t) = 0 \tag{10}$$

初始条件：

$$u(x, 0) = \sin(\pi x) \tag{11}$$

方程(8)涉及 u 对时间 t 和坐标 x 的偏导数，这是一个偏微分方程的求解问题。

与问题 1～问题 4 类似的数学求解问题在化学工程领域经常可以遇到，但是利用此前学习的初等、高等数学知识很难或者无法获得它们的解。这时我们需要的是一种全新的知识：数值计算。这是一名化学工程师必须掌握的专业技能。

数值计算(数值分析、计算方法、科学计算)采用有效而合理的近似简化求解过程，最终获得所求解问题在求解域中固定点的数值。作为应用数学的一个分支，数值计算在其诞生之初被很多数学家视为“异类”，但随着数值计算在众多实际问题取得重大的应用成果，它也越来越被人们所重视。时至今日，数值计算已成为当今科学研究的三种基本手段之一，是计算数学、计算机科学和其他工程学科相结合的产物，并随着计算机的普及和各门类科学技术的迅速发展日益受到人们的重视。对于化学工程师的任务而言，利用计算机进行数值计算和模拟，将是实验研究的有益补充和拓展。实际上，计算机已使化学工程师设计和

① 1 atm=101325 Pa。

分析过程的方法发生了革命性的变化，化学工程师已有能力解决更加复杂的计算问题。

2. 误差

数值计算与数学分析最大的不同在于它并不追求结果的完美。在进行数值计算时，误差是难以避免的。因此问题不是试图消除误差，而是要把误差控制在一定的范围内。

1）关于误差的几个概念

误差虽然不可避免，但人们总是希望计算结果能足够准确，这就需要估计误差。设以 x 代替数 x^* 的近似值，误差 $x-x^*$ 的具体数值是无法确定的，只能根据测量工具或计算过程设法估算出它的取值范围，即误差绝对值的一个上界。

$$|x-x^*|\leqslant\varepsilon \tag{12}$$

这种上界 ε 称作近似值 x 的绝对误差限，简称误差限，或称精度。

近似值 x 的绝对误差还不足以刻画它的精度，例如，测量 1000 m 的长度时发生了 1 cm 的误差，同测量 1 m 时发生了 1 cm 的误差，两者的含义是大有区别的，可见要刻画近似值的精度，除了要参考绝对误差的大小之外，还应当考查这个值的大小，这就需要进一步引进相对误差的概念。仍以 x 代表 x^* 的近似值，若

$$\frac{|x-x^*|}{|x^*|}\leqslant\varepsilon \tag{13}$$

则称 ε 为近似数 x 的相对误差限。

要将一个位数很多的数表示成一定的位数，通常用四舍五入的办法，如 $\pi=3.14159265\cdots$ 可表示为 3.14，3.1416 等。如果近似值 x 的误差限是它的某一位的半个单位，我们就说它“准确”到这一位，并且从这一位起到前面第一个非零数字为止的所有数字均称为有效数字。具体地说，对于 x^* 的近似值（规范化格式），有

$$x=\pm 0.a_1a_2\cdots a_n\times 10^m \tag{14}$$

其中 $a_1a_2\cdots a_n$ 是 0 到 9 之间的自然数，$a_1\neq 0$。

如果误差

$$|x-x^*|\leqslant\frac{1}{2}\times 10^{m-p},\ 1\leqslant p\leqslant n \tag{15}$$

则称近似值 x 有 p 位有效数字，或称 x“准确”到第 p 位。按照这种说法，π 的近似值3.14和3.1416分别有 3 位和 5 位有效数字。例如以 3.14 代替 π 的值，有：

$$|3.14-3.1415926\cdots|\approx 0.0015926<0.5\times 10^{-2} \tag{16}$$

其中 $m=1$，$m-p=-2$，所以 $p=3$，即 3.14 具有 3 位有效数字。

2）误差来源

一般来讲，在对一个问题进行近似和求解过程中产生的误差可以分为以下几类。

（1）模型误差

在把实际物理模型抽象成一个数学模型时产生。例如，计算一个球在初速度为 0 的情况下从 10 m 高空落地时的速度。最简单的情况，可以忽略空气阻力，球在下落过程中遵循牛顿第二定律，则有

$$v=\sqrt{2g\Delta s} \tag{17}$$

其中 g 为重力加速度；Δs 为位移。

这一模型没有考虑空气阻力的影响，这就是模型的误差。但是，对于本例计算而言这一误差的影响很小，因此模型是适合的。不过这类误差会限制数学模型在某些情况下的应用，例如上式并不适合计算一个从 10000 m 高空坠落的球的落地速度。

模型误差不是数值计算的研究对象。

(2) 截断误差

许多数学运算(如微分、积分及无穷级数求和等)是通过极限过程定义的，然而计算机上只能完成有限次的算术运算(如加、减、乘、除等)和逻辑运算，因此需要将求解方案加工成有限次算术运算与逻辑运算序列。这种加工常常表现为无穷过程的截断，由此产生的误差通常称作截断误差。

例如：指数函数 e^x 可展开为幂级数形式

$$e^x=1+x+\frac{x^2}{2!}+\cdots+\frac{x^n}{n!}+\cdots \tag{18}$$

但用计算机求解时，不能计算右端无穷多项的和，而只能截取有限项计算

$$S_n(x)=1+x+\frac{x^2}{2!}+\cdots+\frac{x^n}{n!} \tag{19}$$

这样计算部分和 $S_n(x)$ 作为 e^x 的值必然会有误差，根据泰勒余项定理，其截断误差为

$$e^x-S_n(x)=\frac{x^{n+1}}{(n+1)!}e^{\theta x},\ 0<\theta<1 \tag{20}$$

(3) 舍入误差

计算过程中所用的数据位数可能很多，甚至是无穷小数，然而受机器字长的限制，用机器代码表示的数据必须舍入成一定的位数，这就会引起舍入误差。每一步的舍入误差是微不足道的，但经过计算过程的传播和积累，舍入误差甚至可能会“淹没”所要的真解。

例题 1 在 MATLAB 的命令窗口中输入以下语句：

```
>>format long
>>a=4/3
>>b=a-1
>>c=3*b
>>e=1-c
```

注：以上语句中的“>>”为 MATLAB 默认提示符，无需输入。

以上语句的输出结果为

```
a=
   1.333333333333333
b=
   0.333333333333333
c=
```

```
    1.000000000000000
e=
    2.220446049250313e-016
```

以上程序计算的是1−3*(4/3−1),结果应该为0,但实际上运行结果不是。这是由于在执行除法语句时产生了舍入误差。这种误差是由计算机存储浮点数的存储空间有限决定的。

3) 浮点数与浮点运算

(1) 浮点数

由于计算机资源的有限,在计算机上只能表示有限的实数,这些数被称为浮点数。1985年以后的计算机都是用IEEE标准的浮点运算体系,目前的标准是IEEE Std 754™—2008。在这种体系中,非零浮点数是规范化的,可以表示为

$$x = (-1)^s \cdot (d_0 d_1 d_2 \cdots d_n)_2 \cdot 2^e \tag{21}$$

其中 s 是符号位,表示 x 为正或负数;d_0 默认为1,$d_1 d_2 \cdots d_n$ 称为数 x 的尾数,在二进制中,d_i 只能等于0或1;n 是尾数的长度(或精度);e是指数。在常用的双精度体系中,一个数采用64 bit表示,其中符号位占1 bit,n 占52 bit,指数e占11 bit。双精度浮点数的最大指数 e_{max} 为1023,最小指数 $e_{min}=1-e_{max}=-1022$。

采用这种方法,计算机可以表示的最大实数是

$$real_{max} = (2-2^{-52}) \times 2^{1023} \approx 1.8 \times 10^{308} \tag{22}$$

可以表示的最小正实数是

$$real_{min} = 2^{-1022} \approx 2.2 \times 10^{-308} \tag{23}$$

由此可见,浮点数表示的实数是有范围的。超过这一范围的实数,都被称为无穷(常用Inf表示)。不仅如此,浮点数只能表示有限个实数,浮点数之间的间隔随着数的增大而增加。在双精度浮点数中,与1最近的浮点数与1之间的差值为

$$eps = 2^{-52} \approx 2.22 \times 10^{-16} \tag{24}$$

这一差值通常也被称为机器精度。例题1中的计算结果正好等于这个值。

以下一段MATLAB程序的含义是将变量 a,b 赋值为1,只要 a 和 b 相加之和不等于 a,变量 b 便会被除以2,然后继续判断 a,b 之和是否等于 a。

```
>> a=1; b=1;
>> while a+b~=a;
    b=b/2;
    end
```

对于一般实数运算而言,这个程序永远不会停止,而在本例中,程序在运行一定次数以后便会结束,并返回 b 的值1.1102e−016=eps/2。这种情况的出现就是由于浮点数是有限个的,1+eps是离1最近的实数,1与1+eps/2之间的实数在计算机中被认为是1,因此当 b=eps/2时,$a+b=a$,程序终止运行。

(2) 浮点运算

由于计算机只能表示有限个实数，因此实际算术运算时可能引起舍入误差，有些情况下实数的运算法则也不再适用于浮点数的代数运算。

例题 2 假定使用一台十进制计算机，它表示的浮点数具有 4 位尾数和 1 位指数，超过计算机存储位数的数字均被舍去，试分别计算以下表达式的值。

(1) $0.1557\times10^{1}+0.4381\times10^{-1}$；

(2) $250.209-250.100$；

(3) 136.3×0.06423

解：

(1) 当两个浮点数相加时，需要对指数较小数的尾数进行调整，使两个数的指数相同，以便对齐小数点，这一过程也被称为对阶，然后对应位置的尾数进行相加。本例两个数相加时，第二个数首先进行对阶，即：$0.4381\times10^{-1}\rightarrow0.004381\times10^{1}$，然后进行相加，中间结果为 0.160081×10^{1}，由于计算机只有 4 位尾数，因此最终结果为 0.1600×10^{1}，可见第二个数中的最后两位数字在计算过程中丢失，这就是有效数位的丢失。

(2) 首先将以上两个数表示为浮点数，分别为 0.250209×10^{3} 和 0.25010×10^{3}，相减并舍去多余数位后，结果为 0.1000。可见这一减法造成了很大的误差。实际上，将两个几乎相等的数相减而丢失的有效数字是数值方法中舍入误差的最大来源。

(3) 乘法和除法比加减法更为直接。乘法运算时只需指数相加，尾数相乘，然后对结果进行归一化和舍去处理。除法则为指数相减，尾数相除，然后进行归一化和舍去处理。大多数计算机用双倍长度的寄存器来保留中间结果，对于本例有：$0.1363\times10^{3}\times0.6423\times10^{-1}$，结果为 $0.08754549\times10^{2}\rightarrow0.8754\times10^{1}$，进行舍去处理后可得结果为 0.8754×10^{1}。

特殊情况：

对于浮点数运算，加法和乘法运算交换律仍然适用，但是其结合律和分配律已不再适用。当上溢和下溢(数值超过浮点数可以表示最大和最小实数时)的情况发生时，计算结果等于无穷，结合律便不再适用了。例如 $a=1.0e+308$，$b=1.1e+308$，$c=-1.001e+308$，用下面两种方式分别进行运算，可以得到：$a+(b+c)=1.0990e+308$，$(a+b)+c=\mathrm{Inf}$。无穷与有限大非零实数之间的算术运算结果均为无穷。有限大实数除以 0 的运算结果也为无穷。

最后需要注意的一个问题是：对于数学含义不明确的表达形式，如 0/0、∞/∞、(+Inf)+(−Inf)、0 * Inf，遇到这类表达形式，将会给出提示信息 NaN(not a number，非数)，对于 NaN 通常的数值计算规则并不适用，任何与 NaN 进行的运算结果均为 NaN。

例题 3 以下浮点数运算采用 IEEE 双精度格式，试计算其结果。

(1) $(1+1\times10^{-16})-1$；　　(2) $\dfrac{1}{(1+1\times10^{-16})-1}$；

(3) $1.7\times10^{308}-1.8\times10^{308}$；　　(4) $0\times(1.8\times10^{308}+0.5\times10^{308})$

解：

(1) 因为 1×10^{-16} 小于 eps，在计算过程中被舍掉，因此，计算结果应为 0。

(2) 分母的运算结果同(1)，1/0 的结果为 Inf。

(3) 因为 1.8×10^{308} 超过计算机可以表示的最大实数，被视为+Inf，一个有限实数减正无穷的结果为−Inf。

(4) 同(3)，1.8×10^{308}被视为 Inf，$1.8\times10^{308}+0.5\times10^{308}$的结果为 Inf，而 $0\times$Inf 的结果为非数，NaN。

4) 误差的传递

误差可以在计算过程传递。假定计算结果 Y 与独立的初始数据 x_1^*，x_2^*，…，x_n^* 存在以下函数关系：

$$Y=f(x_1^*,x_2^*,\cdots,x_n^*) \tag{25}$$

x_1，x_2，…，x_n 是 x_1^*，x_2^*，…，x_n^* 的近似值，在每个 x_i 处的绝对误差 $e^*(x_i)=x_i^*-x_i$ 的绝对值都很小，多元函数 f 在点 $x=(x_1,x_2,\cdots,x_n)$ 处可微，则 $Y=f(x_1,x_2,\cdots,x_n)$ 的绝对误差为

$$e^*(Y)=Y^*-Y\approx\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)_x e^*(x_i) \tag{26}$$

相对误差为

$$e_r^*(Y)\approx\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}\right)_x\frac{x_i\cdot e_r^*(x_i)}{Y} \tag{27}$$

可以利用以上两式来估计误差。特别地对于和、差、积、商的误差估计有：

绝对误差

$$|e^*(x\pm y)|\approx|e^*(x)\pm e^*(y)|\leqslant|e^*(x)|+|e^*(y)| \tag{28}$$

$$|e^*(xy)|\approx|y\cdot e^*(x)+x\cdot e^*(y)|\leqslant y\cdot|e^*(x)|+x\cdot|e^*(y)| \tag{29}$$

$$\left|e^*\left(\frac{x}{y}\right)\right|\approx\left|\frac{1}{y}e^*(x)-\frac{x}{y^2}e^*(y)\right|\leqslant\frac{1}{y}|e^*(x)|+\frac{x}{y^2}|e^*(y)| \tag{30}$$

相对误差

$$|e_r(x\pm y)|\approx\left|\frac{x}{x\pm y}e_r(x)+\frac{y}{x\pm y}e_r(y)\right|\leqslant\left|\frac{x}{x\pm y}e_r(x)\right|+\left|\frac{y}{x\pm y}e_r(y)\right| \tag{31}$$

$$|e_r(x\cdot y)|\approx|e_r(x)+e_r(y)|\leqslant|e_r(x)|+|e_r(y)| \tag{32}$$

$$\left|e_r\left(\frac{x}{y}\right)\right|\approx|e_r(x)-e_r(y)|\leqslant|e_r(x)|+|e_r(y)| \tag{32}$$

例题 4　某人采用称量瓶测量某种液体密度，已知称量瓶的准确体积 V 为 50.00 mL，其绝对误差为 0.05 mL；瓶中液体的质量 m 为 48.00 g，其称量绝对误差为 0.02 g，液体密度等于 m/V。采用这种方法测得液体密度的绝对误差限和相对误差限分别是多少？

解：

绝对误差：$e^*\left(\frac{m}{V}\right)\leqslant\frac{1}{V}|e^*(m)|+\frac{m}{V^2}|e^*(V)|$

$$=\frac{1}{50}\times0.02+\frac{48}{50^2}\times0.05=0.00136\ \text{g/mL}$$

相对误差：$\left|e_r\left(\frac{m}{V}\right)\right|\leqslant|e_r(m)|+|e_r(V)|=\frac{0.05}{50}+\frac{0.02}{48}=0.14\%$