数值计算误差初步理解及简单分析

刘 平¹,吕海兵²,张 峰¹,胡文礼²

(1.湖北省基础地理信息中心,湖北 武汉 430071; 2.中南勘察设计院(湖北)有限责任公司,湖北 武汉 430071)

摘 要:介绍了数值计算中的误差类别,用数学表达式举例,说明了截断误差,根据 IEEE754 标准用数学表达式说明了舍

入误差;并对截断误差与舍入误差进行了简单的分析;最后给出了数值计算中应注意的事项。

关键词:数值计算;计算误差;初步理解;简单分析

中图分类号: O241 文献标志码: B



1 数值计算及数值计算误差

研究如何运用计算机去获得数学问题的数值解的 理论和方法称为数值计算方法。对那些在经典数学中, 用解析方法在理论上已作出解的存在,但要实际求出它 的解析解又十分困难, 甚至是不可能的这类数学问题, 数值计算方法就显得不可缺少, 同时又十分有效。

如果 x 为某个量的真值, X 为采用一定的数学模 型和确定的方法,通过计算机程序计算得到的近似值, 则 x-X 称为数值计算误差。数值计算误差可理解为人 们用来描述数值计算中近似的精确程度,是科学计算 中的一个十分重要的概念。

2 数值计算误差来源

2.1 模型误差(描述误差)

用计算机解决科学计算问题首先要建立数学模型。 数学模型是指那些利用数学语言模拟现实而建立起来 的有关量的描述,而模型往往是对被描述的实际问题 进行抽象、简化而得到的,因而往往也是近似的,亦 即两者之间存在有误差,在数值计算中,把数学模型 与实际问题之间出现的这种误差称为模型误差或称为 描述误差。

2.2 观测误差

在数学模型中,往往还有一些根据观测或实验得 到的物理量或几何量,如温度、长度、电压等,由于 测量工具的精度、观测方法和客观环境条件的限制等 因素的影响,使得观测或实验得到的参数也包含误差, 这种由观测产生的误差在数值计算中称为观测误差。 如,当用布尔沙-沃尔夫七参数坐标系转换模型作坐标 系转换时,所采用的重合点的坐标值(在此模型中就 是观测值)含有观测误差。

2.3 截断误差(方法误差)

理论上的精确值往往要求用无限的过程才能求出, 而实际运算中往往只能完成有限项或有限步运算,因 此要将有些需用极限或无穷过程进行的运算有限化 即 对无穷过程进行截断,这样产生的误差称为截断误差 (有限逼近无限时的误差)。截断误差是由计算方法本 身所带来的,因而也称为方法误差。如采用级数展开 式中舍去部分引起的误差(在高斯投影计算中,其子 午线弧长的计算就是采用级数展开的前有限项)、取收 敛级数的前有限项而舍其后项以及以离散量近似代替 连续量而引起的误差等。

文章编号: 1672-4623 (2011) 03-0156-03

截断误差是数值计算方法中要研究的主要误差 很 多好的数值计算方法都是巧妙地处理截断误差得出的。 截断误差的大小直接影响计算结果的精度和计算工作 量,是数值计算中必须考虑的一类误差。

2.4 舍入误差

在进行数值计算时,必须要求参与计算的数的长 度有限,否则计算无法进行。在利用计算机进行计算 时,由于计算机字长有限,因此也同样要求数的长度 有限。而在实际计算中,大量的数的长度是无限的。 如,无理数、除不尽的分数等,因此必须用与它们比 较接近的数来表示它们,即对参与计算的原始数据、中 间计算结果、最终计算结果数据都必须做舍入处理,由 此产生的误差称为舍入误差。

舍入误差在每一步计算中可能微不足道,但在整 个的计算过程中,它会不断积累和传播,会造成最终 计算结果产生较大误差。因此,舍入误差是数值计算 中必须考虑的一类误差。

一个算法, 如果在运算过程中舍入误差在一定条 件下能够得到控制,或者舍入误差的增长不影响产生 可靠的结果, 则称该算法是数值稳定的, 否则称其为数 值不稳定。

收稿日期:2011-05-03

3 截断误差和舍入误差举例说明

模型误差和观测误差是用计算机通过科学计算来解决实际问题所必然会产生的,属于固有误差。而在计算过程中所引起的误差叫做计算误差,包括截断误差和舍入误差。在数值计算中着重研究的是计算误差(截断误差和舍入误差),并对它们的传播与积累规律做出分析。

3.1 截断误差举例说明

在大地测量及其他测绘工程所采用的数学模型中, 普遍会做 sin x 的计算,一般是采用 sin x 的无穷级数来 计算,即:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots + (-1)^n \frac{x^{2^{n+1}}}{(2n+1)!} + \dots$$

如,若要计算 $\sin x$ 的值, $x \in \left(0, \frac{\pi}{4}\right)$,我们只能在适当的地方"截断",使计算量不太大,而精度又能满足要求。如,计算 $\sin 0.5$,如果取 n=3,则有:

 $\sin 0.5 \approx 0.5 - \frac{0.5^3}{3!} + \frac{0.5^5}{5!} - \frac{0.5^7}{7!} = 0.479625$,据泰勒余

项公式,它的误差应为:

$$R = (-1)^9 \frac{\xi^9}{9!} , \xi \in \left(0, \frac{\pi}{4}\right) ,$$
$$\left|R\right| \le \frac{\left(\pi / 4\right)^9}{362880} = 3.13 \times 10^{-7} .$$

例如,用泰勒 (Taylor) 多项式

$$P_n(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n ,$$
 近似代替函数 f(x)

计算方法的截断误差是:

 $R_n(x) = f(x) - P_n(x) = \frac{f^{(n)}(\xi)}{(n+1)!} x^{n+1}$, (ξ介于 0 与 x 之间)。一般地,通过数学方法可推导出截断误差限的计算公式。

3.2 舍入误差举例说明及简要分析

一般地,当近似值 x*的绝对误差限是某一位上的半个单位时,我们称其"准确"到这一位,且从该位起直到前面第一位非零数字为止的所有数字都称为有效数字。例如, π 3.141 592 6...精确到小数点后 4位的近似值为 3.141 6(ϵ (x) 0.5 \times 10?4),它具有 5位有效数字;精确到小数点后 5位的近似值为 3.141 59(ϵ (x) 0.5 \times 10-5),它具有 6位有效数字。根据有效数字可以算出近似数的绝对误差限;有效数字的位数越多,其绝对误差限就越小。有效数字的位数反映了近似值的相对精确度。

一个十进制实数的标准表示式为:

 $x = \pm 0.a_1a_2...a_na_{n+1}... \times 10^m, a_1 0$,做"四舍五入"得:

舍入误差:

$$|e| = |x^* - x| \le 0.00 \cdot \cdot \cdot 0.00 = \frac{1}{2} \times 10^{m-n}$$

又如,采用圆周率的近似值 3.14159,则所产生的误差是 $R = \pi - 3.14159 = 0.0000026535...$

为了了解计算机中二进制计算产生舍入误差的原理,首先简单谈谈计算机是如何表示十进制实数的。计算机中采用有限二进制位表示实数,根据实数的小数点位置在机器中固定与否,对实数有定点表示与浮点表示之分。由于浮点表示有其较大优越性,大多数计算机都用浮点表示实数。在计算机各种程序语言中,浮点数常用单精度浮点数或双精度浮点数表示(扩展双精度浮点数较少使用)。浮点运算很少是精确的。虽然一些数字(比如 0.5)可以精确地表示为二进制小数(因为 0.5 等于 2-1),但其他一些数字(比如 0.1)就不能表示为精确的二进制小数,事实上(0.1)10=(0.000 110 011...)2,是无限循环小数。即用有限位,按 IEEE 754 浮点数表示法是无法精确表示无限循环小数的。因此,浮点运算可能导致舍入误差。其实,在 0.1 至 0.9 这 9 个数值中,只有 0.5 能精确表示成二进制数。

根据 IEEE 754 标准,单精度浮点数和双精度浮点数,分别占 32 位字长和 64 位字长。为提高数据的表示精度,IEEE 754 标准规定,当尾数的值不为 0 时,尾数域的最高有效位应为 1,否则以修改阶码同时采取左右移小数点的办法,使其变成一个最高位为 1 的标准形式尾数,这种形式的表示方法称为浮点数的规格化表示。对应的浮点数称为规格化浮点数。因为经规格化后的规格化数中的尾数第一位总为 1,所以在计算机中尾数部分默认了第一位是 1。因而,单精度格式的 23 位尾数实际上表示了 24 位有效数字,同理,双精度格式的 52 位尾数实际上表示了 53 位有效数字,这进一步提高了浮点数的表示精度。

据 IEEE754 标准格式表示的 32 位浮点数中, S 是 浮点数的符号位,占 1 位(0表示正数,1表示负数); M 是尾数,占 23 位,用小数表示,小数点放在尾数域的最前面; E 是阶码占 8 位,阶符采用隐含方式,即采用移码方式来表示正负指数。移码方法对 2 个指数大小的比较和对阶操作都比较方便,因为阶码域值大者其指数值也大。采用这种方式时,将浮点数的指数 真值 e 变成阶码 E 时,应将指数 e 加上一个固定的偏移值 127 (01111111),即 E = e + 127。

			31	30		23 22		0
32位浮点数			S		E		М	
	63	62			52	51		0
64位浮点数	S			E			M	

同理,64 位浮点数中,其 11 位阶码(符)采用隐含方式,即也采用移码方式来表示正负指数。将浮点数的指数真值 e 变成阶码 E 时,应将指数 e 加上一个固定的偏移值(即,偏置常数 2^{n-1} -1,n 为阶码的位数),即, $E=e+2^{n-1}$ -1=e+1 023。

以下根据 IEEE754 标准简单分析二进制浮点数在 计算机中的表示。

用十六进制编辑软件WinHex打开一个记录有浮点数据的二进制文件,其数据记录区域前8个字节是: B3 C5 48 09 D4 94 5C 40,展开为二进制是:

40 5C 94 D4 09 48 C5 B3 0100 0000 0101 1100 1001 0100 1101 0100 0000 1001 0100 1000 1100 0101 1011 0011

则有:S=0,E=10000000101(二)=1 × 210+1 × 22+1 × 20=1029(十)

M=1100 1001 0100 1101 0100 0000 1001 0100 1000 1100 0101 1011 0011

1. M=1.1100 1001 0100 1101 0100 0000 1001 0100 1000 1100 0101 1011 0011

根据公式 $X = (-1)^s \times (1. M) \times 2^{E-1023}$

e = E - 1023

则有:e = 1029 - 1023=6

X=1 × 1.1100 1001 0100 1101 0100 0000 1001 0100 1000 1100 0101 1011 0011 × 26 (十)

=1. C94D40948C5B3 (+\(\phi\)) **x** 64 (+)

到此,先将 1. C94D40948C5B3 (+六) 化为十进制数据,

1. C94D40948C5B3 (+六) = (1+0. C94D40948C5B3)

=1+C94D40948C5B3/1613

=1+3541338130269619/4503599627370496

=1.7863350260416666070995006521116

所以,X=1.7863350260416666070995006521116 × 64

=114.32544166666666285436804173514

而 114.32544166666666285436804173514 用单精度 浮点数表示时,在计算机中是: A0 A6 E4 42,展开为 二进制数是:

42 E4 A6 A0 0100 0010 1110 0100 1010 0110 1010 0000 则 S=0 $E=10000101=1 \times 27+1 \times 22+1 \times 20=133$ (十)

M=110 0100 1010 0110 1010 0000

1. M=1.110 0100 1010 0110 1010 0000

根据公式 $X = (-1)^s \times (1. M) \times 2^{E-127}$

e = E - 127

则有: e = 133 - 127=6

 $X=1 \times (1.110\ 0100\ 1010\ 0110\ 1010\ 0000) \times 26$

=1. C94D40 (+\(\pi\)) **x** 64 (+)

到此,先将1. C94D40 (+六) 化为十进制数据,

 $1.C94D40_{(+\div)} = 1+0.C94D40_{(+\div)} = 1+C94D40_{(+\div)}/166$

=1+ 0.786334991455078125

= 1.786334991455078125 (+)

所以: X=1.786 334 991 455 078 125 (+) × 64 (+)

=114.325 439 453 125。对比双精度值: 114.325 441 666 666 662 854 368 041 735 14 (+)

可知,同一个实数分别用单、双精度数表示时,由于 采用的字长差异,精度就出现差异。

上面介绍了浮点数在计算机中表示是存在舍入误 差的。以下简单谈谈规格化浮点数加减运算原理:设 有 2 个二进制浮点数分别为:A=MA × 2EA,B=MB × 2EB, 基于这 2 个浮点数的加减运算的过程是: A、B 都做规格化处理。 对阶。对阶的目的是使两操 作数 A 和 B 的阶码相等,以提供尾数相加减的可能性。 阶码的比较是通过两阶码的减法来实现的,统一取大 的阶码;小阶码的尾数按两阶码的差值决定右移的数 量。首先应求出两数阶码 EA 和 EB 之差,即ΔE=EA-EB。若 Δ E =0,表示两数阶码相等,即EA=EB;若 Δ E > 0,表示EA > EB;若ΔE < 0,表示EA < EB。当 EA EB 时,要通过尾数的移位来改变 EA 或 EB,使 EA=EB。对阶的规则是:小阶向大阶看齐。即阶码小 的数的尾数右移,每右移一位,阶码加1。 运算。在完成对阶操作后,就可进行尾数求和运算了, 无论是加法还是减法运算,都可以转换为加法操作。 4 运算结果的尾数规格化(结果规格化处理视情况不同 有不同的处理方法,因篇幅所限这里省略。)。

在程序设计中,当定义 float a, b, c; (即定义 a、b、c 都是单精度浮点数)时,如果 a=0.65, b=0.6,那么 c=a-b 经过运算后会等于 0.05 吗?经程序计算,实际得 c=0.049 999 95。就是说运算过程产生了舍入误差。因为,计算机在做 2 个浮点数相加或相减时,首先是做"对价",就是首先把小数点的位置对齐,而浮点数的小数点实际位置取决于阶码的大小,因此,对齐 2 个数的小数点,实际上就是使 2 个数(下转第 161 页)

同时补充完善土地他项权利的登记信息。

2.5 国土资源管理其他方面的应用

第二次土地调查成果采用了高新技术,调查成果质量高,各类土地分布明晰,数据准确客观,真实反映各区域土地利用现状,为编制各级土地利用规划提供详实的基础数据和图件,是各级土地规划的依据,使基本农田保护范围落到实处、各级政府更好地履行保护耕地的职能。

按照第二次土地调查成果的土地分类,全面掌握本地区所需整理、复垦土地分布位置、面积,为科学制定整理复垦计划,有步骤、有计划的开展整理、复垦工作提供了科学依据。

3 结 语

土地调查成果是以地块为基础的土地及其附着物的权属、数量、质量、位置和利用现状的基本信息的集合,它不仅是土地管理工作的成果之一,而且是全面、统一、依法、科学管理土地时不可缺少的依据。武汉市第二次土地调查查清了土地的数量、分布、权属、

利用现状等情况,应用好其成果对规范国土资源管理、实现土地监管方式从"以数管地"到"图数一体管地"的重大转变、更好地满足国土管理、促进社会经济发展具有十分重要的意义。

参考文献

- [1] TD/T 1014-2007 .第二次土地调查技术规程[S]
- [2] 孙炜. 我国第二次土地调查成果应用设想 [J].科技信息, 2009 (23):1174
- [3] 刘小刚. 第二次土地调查及其成果应用设想 [J].现代经济信息, 2009(17):346-347
- [4] 邓炯. 河南省第二次土地调查成果应用分析 [J].河南科技, 2010(19):34-35
- [5] 王履华,高权忠. 第二次土地调查数据成果应用模式研究[J]. 测绘与空间地理信息,2010(6):117-119,123
- [6] 王继亮.邹峥嵘 第二次土地调查中一些问题的探讨 [J.北京 测绘,2009(02): 68-69, 62
- [7] 贾洪武; 王兆强 第二次土地调查成果应用设想 [J].中国高新技术企业, 2008, (20):183

第一作者简介:徐勇,城市规划工程师,研究方向为土地规划管理及实际应用。

(上接第 158 页)的阶码相等,这个过程称为对阶。因为浮点十进制值通常没有完全等同的二进制表示形式,从而可能会造成一些精度丢失。且一些浮点运算可能会产生意外的结果。如在 c 语言中,如果你定义 double a, b;,则如果用 if (a==b) 比较 2 个浮点数是否相等是不正确的选择,最好的方法是利用一个精确的阈值。这个阈值和作比较的浮点数值大小有关。比如,写成 if $(fabs\ (a-b) <= epsilon*fabs\ (a))$ (epsilon* km km)一个选定的值来控制"接近度",且你要确保 a 不会为 0。),对于有编程经验的人,这些都会有体会。

4 数值计算中建议注意的事项

- 1) 应选用稳定性好的算法,以控制误差的传播;
- 2)注意简化计算步骤和公式,设法减少运算次数;
- 3)避免除数绝对值远远小于被除数绝对值的除法 计算;避免大数做乘数;
- 4) 防止参与运算的数在数量级上相差悬殊,造成"大数吃小数";
- 5) 采用一定的变化方法去避免相近两数相减,以避免造成有效数字损失,从而导致产生较大绝对误差和相对误差;

6) 防止出现"机器零"和"数据溢出"。如果一个浮点数的尾数全为 0,则不论其阶码为何值,计算机在处理时都把这种浮点数当作零看待;如果一个浮点数的阶码小于它所表示范围的最小值,则不论其尾数为何值,计算机在处理时都把这种浮点数当作零看待。在计算机中,当要表示的数据超出计算机所使用的数据的表示范围时,则产生数据溢出。

参考文献

- [1] 崔国华,许如初编著.计算方法[M].北京 :电子工业出版社,2005
- [2] 李庆扬 ,王能超 ,易大义编.数值分析[M].北京 :清华大学出版社 2001
- [3] 王玉龙.计算机导论[M].北京:电子工业出版社,1997
- [4] 俸远祯等编. 计算机组成原理与汇编语言程序设计 [M]. 北京:电子工业出版社,1997
- [5] IEEE Standard 754 for Binary Floating-Point Arithmetic pdf 文 档)http://www.cs.berkeley.edu/~wkahan/ieee754status/ IEEE754.PDF
- [6] 数值计算指南 Sun? Studio 11,2005 年 11 月(pdf 文档)
- [7] 孔祥元等编写.大地测量学基础[M].武汉 武汉大学出版社, 2006

第一作者简介:刘平,高级工程师,主要从事地理信息技术相关管理工作。

Abstract Urgent monitoring of earthquake disasters need thematic mapps of remote sensing. According to the distribution of the earthquake and the secondary disasters, classifications and coding of disaster information were analyzed. With the national basic scale topographic maps and thematic map symbols of the type, size, color etc separately defined at different scales, it was produced that thematic symbol library of the corresponding earthquake and secondary disasters. Meanwhile, standard for thematic maps of the disaster and the secondary disasters was formed. The templates met the cartography requirements of GIS applications, which can provide accurate disaster information and affected body quickly, and improved the mapping efficiency. It is useful to save time for the rescue and assure the earthquake relief work can be undertaken smoothly.

Key words earthquake, the secondary disasters, symbol database, mapping templates (Page:153)

Preliminary Understanding and Simple Analysis of Numerical Calculation Errors by LIU Ping

Abstract This paper introduced the classification of numerical calculation errors, and used math expressions as examples to illustrate truncation error. According to IEEE754 criterion, this paper used math expressions to explain rounding error. Then this paper simply analyzed the truncation error and rounding error. At last, this paper pointed out attentions in the numerical calculation.

Key words numerical calculation , calculation errors , preliminary understanding , simple analysis (Page:156)

Application of the Second Land Investigation's Result to Daily Land and Resources Management by XU Yong

Abstract This paper introduced the basal informations, technical route, investigating methodology and investigation's result of the WUHAN Second Time Land Investigation, and discussed the applicating of the Second Time Land Investigation's result based on the daily Land and Resources Management's requirement.

Key words land investigation land management, cadastral management (Page:159)

Method for Prediction of Landslide by Phase Space Reconstruction
by XIONG Tianan

Abstract In view of the nonlinear characteristics of landslide displacement time sequence, this paper introduced the prediction method based on phase space reconstruction and least squares support vector machine (LSSVM). Used Cao's method to determine the embedding dimension, according to mutual information method to compute the best delay time; Then in the phase space, used least squares support vector machine (LSSVM) to establish the forecast model to compared with LSSVM and the neural network predicting mode. The test result show that the model has the high precision, is scientific and feasible. Key words predictable mode; landslide prediction; phase space reconstruction; least squares; neural network (Page:162)

Application of SET 1X Total Station Device to Tunnel Through Survey by CHEN Sansheng

Abstracts Combined with the application in Pusagang tunnel through survey of Yalu Highway of SET 1X total station device, elaborated the adoptive ways and technical measurements of tunnel through survey so as to make sure the accurate perforation of the tunnel.

Key words total station device, Pusagang tunnel, perforation survey, control survey (Page:165)

Evaluation of Uncertainty of GPS Receiver Calibration and Calibration Results by HE Hao

Abstract Based on the actual calibration GPS receiver's experience, through the GPS receiver calibration method of in-depth research, with

practical examples, the calibration process problems encountered were analyzed and investigated

Key words GPS receiver Calibration , measuring error , uncertainty (Page:168)

Establishment of Monitoring Data Processing and Analysis System with Excel VBA Programming by ZHU Xingang

Abstract Excel is popular office software which we often use at work. Excel itself provides a strong secondary development function, VBA, which has powerful programming capabilities. This article described how to use Excel VBA to create a complete deformation monitoring system to realize simple, rapid, accurate and automated monitoring iob.

Key words Excel ,VBA , deformation monitoring data processing and analysis system (Page:170)

Application of Robotic Total Station in Volume of Vertical Metal Cans by ZHU Lianghua

Abstract The application of robotic total station in volume surveying of vertical metal cans was introduced in this paper. The surveying preparation work, fieldwork, data processing and key technology were discussed detailed combining an example.

Key words total station , vertical metal cans , volume surveying (Page:173)

Matlab and Visual C** Mixing Programming to Process the Data of the South Total Station by HU Jiaxing

Abstract During the indoor work data checking, we often compared the point's field work coordinate with its indoor work coordinate. So we could find if there be some points were artificial, or some points had been moved. At the same time, the processed data could direct provide the three dimension coordinate, it was helpful for the next measure work. This article took the South total station instrument for example, based on matlab and Visual C++ mixing programming to generate the directly executable file, which could separate from the MTALAB environment, then gave the code and the images of the result.

Key words MTALAB , hybrid programming , data processing

(Page:175)

Research of Teaching Content in Digital Surveying & Mapping
by MENG Fanchao

Abstract Based on the production units of the engineering survey mapping and technical personnel needs, and for "Digital Mapping" characteristics of the course to a topographic mapping of the main line to explore the teaching content and practice, through the constant adjustment and reform to improve Higher engineering professional teaching quality measurement techniques, enhanced their ability and employability and competitiveness.

Key words Digital Surveying & Mapping , teaching ,research (Page:178)

Strengthening the teaching of surveying error theory for GIS

by ZHAO Dongbao

Abstract Surveying error analysis and treatment is the basic theory which lay a solid foundation for students majored in GIS studying surveying courses, and is also the key to understand surveying principle of various surveying courses. Aiming at the problem that many colleges lacked the uniform plan for surveying error theory course teaching for GIS major students, teaching contents of surveying error theory course for GIS major students were discussed, and teaching schedule was arranged based on characteristic of surveying error theory teaching, and teaching methods of surveying error theory course for GIS major students were listed according to related teaching experience. Key words GIS major surveying error theory ,teaching contents ,teaching method (Page:180)