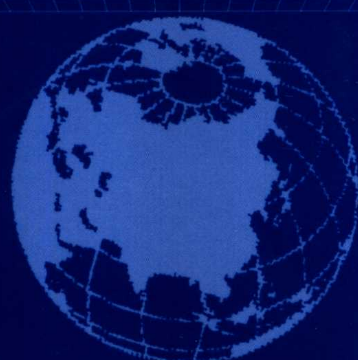
高校计算机教学系列教材

计算方法

李晓红　堵秀凤
张水胜　王延臣　编著



北京航空航天大学出版社

高校计算机教学系列教材

计　算　方　法

李晓红　堵秀凤
张水胜　王延臣　编著

北京航空航天大学出版社

内 容 简 介

本书共分 8 章。内容包括计算方法的特点、任务、研究对象、误差与算法；非线性方程根的数值解法；线性方程组的数值解法；插值法；数值积分；数值微分；常矩阵的特征值与特征向量的数值解法；常微分方程初值问题的数值解法。每章末都配备了适量习题，书末附有习题答案及程序举例。

本书适合数学专业的专、本科，以及计算机等非数学专业作教材，也可供有关方面工程技术人员参考。

图书在版编目(CIP)数据

计算方法/李晓红等编著．—北京：北京航空航天大学出版社，2006.3

ISBN 7-81077-698-3

Ⅰ．计…　Ⅱ．李…　Ⅲ．计算方法—高等学校—教材　Ⅳ．0241

中国版本图书馆 CIP 数据核字(2005)第 117105 号

计算方法

李晓红　堵秀凤
张水胜　王延臣　编著

责任编辑　王文涌

*

北京航空航天大学出版社出版发行

北京市海淀区学院路 37 号(100083)　发行部电话：010-82317024　传真：010-82328026

http://www.buaapress.com.cn　E-mail：bhpress@263.net

北京市松源印刷有限公司印装　各地书店经销

*

开本：787×1092　1/16　印张：13.25　字数：339 千字

2006 年 3 月第 1 版　2006 年 3 月第 1 次印刷　印数：5 000 册

ISBN 7-81077-698-3　定价：20.00 元

总 前 言

科教兴国，教育先行，在全国上下已形成共识。在教育改革过程中，出现了多渠道、多形式、多层次办学的局面。同时，政府逐年加大教育的投入力度。教育发展了，才能有效地提高全民族的文化、科学素质，使我们中华民族屹立于世界民族之林。

计算机科学与技术的发展日新月异，其应用领域迅速扩展，几乎无处不在。社会发展的需求，促使计算机教育生气蓬勃。从普通高校的系统性教学，到远距离的电视、网上教学；从全面讲述，到不同应用领域的、星罗棋布的培训班；从公办的到民办的；从纸介教材到电子教材等等，可以说计算机教学异彩纷呈。要进行教学，就必须有教材。

要对我们这么大的国家和教学形势，在保证国家教学基本要求的前提下，应当提倡教材多样化，才能满足各教学单位的需求，使他们形成各自的办学风格和特色。为此，我们组织北京工业大学、北京航空航天大学、北京理工大学、南开大学、天津工业大学等高校的有丰富教学经验的教师编写了计算机教学系列教材，将陆续与师生见面。

系列教材包括以下各项。

(一) 基础理论：离散数学等。

(二) 技术基础：电路基础与模拟电子技术；数字逻辑基础；计算机组成与体系结构；计算机语言(拼盘，选择使用)，包括C++程序设计基础、Visual Basic程序设计基础、Matlab程序设计基础、Java程序设计基础、Delphi语言基础、汇编语言基础等；数据结构；计算机操作系统基础；计算方法基础；微机与接口技术；数据库技术基础等。

(三) 应用基础：计算机控制技术；网络技术；软件工程；多媒体技术等。

(四) 技术基础扩展：编译原理与编译构造；知识工程——网络计算机环境下的知识处理等。

(五) 应用基础扩展：计算机辅助设计；单片机实用基础；图形、图像处理基础；传感器与测试技术；计算机外设与接口技术等。

本系列教材主要是针对计算机教学编写的，供普通高校、社会民办大学、高等职业学校、业余大学等计算机本科或专科选用。其中一部分也适合非计算机专业本科教学使用。在这些教材的内容简介或前言中对使用范围均作了说明。

本系列教材在编写时，注重以下几点：(1) 面对计算机科学与技术动态发展的现实，在内容上应具有前瞻性；(2) 面对学以致用，既有系统的基础知识，又具有应用价值的实用性；(3) 具有科学性、严谨性。另外，力求使有限的版面具有最大的信息量，以使读者得到实惠。

能否实现这些愿望，只有师生在教学实践中评价。我们期望得到师生的批评和指正。

高校计算机教学系列教材编委会

高校计算机教学系列教材编委会成员

前　言

计算方法也称数值数学、数值计算。它是研究代数、微分方程、函数论、统计数学等各类问题的数值解法的一门学科。电子计算机出现后，其技术的发展日新月异，应用领域也在迅速扩大，极大地推动了计算方法的发展。因此，寻找适当的计算方法以适应计算机的计算服务，已成为这门学科的一个重要研究部分。

数值计算领域极其广泛，求解各类问题的计算方法是多种多样的。怎样选择与使用适当的计算方法，怎样估计计算结果的误差，怎样解释计算过程中的异常现象，自然也成为广大科技工作者迫切需要解决的问题。同时了解计算方法的基本知识，掌握运用计算机进行科学计算的方法也已成为当代理、工类大学生必须具备的技能。

本书是为适应理、工类大学的本科生、专科生学习和掌握电子计算机常用的、基本的数值计算方法的需要编写的。内容丰富，重点突出，语言简练，理论分析简明，算法的构造原理介绍详细。

在内容上，主要介绍了电子计算机上常用的各种数值计算的方法，并逐一介绍了计算方法所对应的算法。为了结合计算机编程，本书对主要算法用C语言编程（附录Ⅰ），还对算法的稳定性、收敛性、误差估计、优缺点、运算量等作了适当介绍；还选择了典型例题与习题，在附录Ⅱ中提供了部分习题的答案。

在篇幅上，本书尽量做到少而精，凡是与基本方法、基本观点、基本概念有关的部分，不惜篇幅，尽量阐述清楚，尽量做到有方法、有计算、有实例、有应用、有分析、有习题。

在结构上，本书尽量从方法的完整性着眼，按照问题的提出→方法原理描述及公式→算法→数值算例的方法，符合简单→复杂；特殊→一般的认识规律，避免了复杂化、抽象化，贴近学生基础，便于学习。另外，本书在编写时已注意到各章节的独立性，删掉带*的章节不至于影响其它章节的学习。

在算法描述上，以数学公式加以必要的文字说明，逐一介绍了计算方法所对应的算法。本书还对主要算法用C语言编程（附录Ⅰ），算法简单，便于验证。

总之，希望通过本课程的学习，使读者能够掌握计算方法的基本思想和基本技巧，学会方法的误差分析，并把这些知识融汇于编程解题过程中，同时本书还给读者留有深入思考的余地，有助于培养他们独立思

考、分析问题、解决问题的能力。

在本课程之前应先修微积分、线性代数、常微分方程、一两门高级语言等课程。

全书共8章，第1章主要介绍计算方法的研究对象、任务、特点和计算方法的基本问题：误差与算法。第2,3,4章分别介绍非线性方程根的数值解法、线性方程组的数值解法。第5,6,8章分别介绍了插植法、数值微分、数值积分、常微分方程初值问题的数值解法。第7章介绍了矩阵的特征值与特征向量的数值求法。为了便于验证，在附录Ⅰ中给出了部分例题的C语言程序，同时为了便于自学，在附录Ⅱ中还提供了部分习题的答案。

本书是在多年教学经验的基础上，经深入进行教学研究，参考了国内外有关教材而编写的。全书设计讲授学时数约为54学时，可安排一定学时的上机实习。对于不同的专业，可根据实际需要选取适当的章节讲授。

作　者

2005年6月

目 录

第1章 绪论

本章主要介绍两个基本概念：误差与算法，并简要介绍计算方法的研究对象及特点，提出算法设计中应遵循的原则与注意事项，有利于读者对本课程的主要内容及学习方法有个初步了解.

1.1 计算方法的研究对象与特点

计算方法是数学的一个分支，又称计算数学、数值计算. 它是在解决各种生产实践与科学实验中，提出数值计算问题的过程中逐步形成的一门学科. 在电子计算机技术迅猛发展的今天，掌握数值计算方法，熟练地运用计算机进行科学计算，是每一位科学工作者必须具备的能力.

一般地，利用计算机解决科学问题需要以下几个步骤：

实际问题→数学模型→算法设计→程序设计→上机计算结果.

上述过程中根据数学模型提出求解的数值计算方法，到编写程序上机计算结果的过程，便是计算方法研究的任务，也是计算方法的主要研究对象. 因此说，计算方法是研究怎样利用计算机等计算工具来求解各种数学问题的数值解(近似解)及其理论的一门学科.

一方面，它研究的是数学问题的数值解法(线性方程组的解法、非线性方程的解法、矩阵的特征值的解法)、数值逼近、数值微分、数值积分、微分方程的数值解法等. 它的基本理论与研究方法是建立在纯数学基础之上的，但它不局限于研究数学本身的理论，而是着重研究数学问题求解的数值方法及与此有关的理论，例如方法的收敛性、稳定性、误差分析等等.

另一方面，最明显特点是与计算机科学联系紧密，有很高的技术性. 因此，我们在考虑算法时，往往要注意结合计算机的特点，例如计算速度、存储量、字长等等技术指标，也就是说本课程除了学习理论知识外，还要注意程序设计的可行性与复杂性和上机计算等环节的学习与实践，所以说计算方法是一门与计算机密切结合，实用性很强的数学课程，也是每一个理工类大学生应该学习的一门学科.

1.2 误差

1.2.1 误差的来源

运用数学方法计算出的实际问题中的数值通常是近似的，实际值与近似值之差称之为误差. 误差的来源主要有4种：

1) 模型误差：由实际问题建立数学模型时，往往忽略了很多次要因素，以求简化问题，使问题理论化，由此产生模型误差.

2）观测误差：数学模型中常包含某些参数，需要通过观测确定，这时产生观测误差.

3）截断误差：一般数学问题难以求准确解，常常通过近似代替，化为简单问题. 这种简化带入的误差是数值方法本身引起的误差，称为方法误差或截断误差.

4）舍入误差：计算机只能对较小位的有限位数进行计算，对超过位数的数字需要进行舍入，便会产生舍入误差.

例如，积分 $I=\int_0^1 e^{-x^2}\,\mathrm{d}x$ 可展开幂级数.

$$I=\int_0^1\left(1-x^2+\frac{x^4}{2!}-\frac{x^6}{3!}+\cdots\right)\mathrm{d}x=1-\frac{1}{3}+\frac{1}{5\times 2!}-\frac{1}{7\times 3!}+\cdots$$

若取级数起始若干项部分之和作为积分近似值.

不妨取 $I\approx\int_0^1\left(1-x^2+\frac{x^4}{2!}-\frac{x^6}{3!}\right)\mathrm{d}x=1-\frac{1}{3}+\frac{1}{2!\times 5}-\frac{1}{7\times 3!}$.

则由于从第 5 项后全部舍去了，自然产生误差，这就是截断误差；而在上式中 $1/3=0.333\,33\cdots$，在计算机上应用时，要按要求进行四舍五入成为有限数，这便是舍入误差.

在上述 4 种误差中，建立适宜的数学模型，从而得到易于计算的公式，不属于计算方法的研究内容，即计算方法不考虑模型误差与观测误差. 截断误差与舍入误差是计算方法课程要讨论的对象. 在各式计算中，通过数学方法要推导截断误差公式，它涉及方法的收敛性；舍入误差的产生随机性较大，讨论起来比较困难，它涉及方法的稳定性. 只有既收敛，又稳定的方法才能提供比较可靠的计算结果.

1.2.2 误差与有效数字

1. 误　差

定义 1　设 x^* 为准确值 x 的近似值，则 x 与 x^* 的差 $e=x-x^*$ 称为近似值 x^* 的绝对误差(误差).

一般地，准确值并不知道，因此 e 通常是无法确定的，人们常常不关心 e 值的本身而只关心它的限.

定义 2　若存在正数 η，使 $|e|=|x-x^*|\leqslant\eta$，则称 η 为绝对误差限(误差限)，有时也用 $x^*=x\pm\eta$ 来表示上式，即只要 x^* 落在 $[x-\eta,x+\eta]$ 内，则认为所得 x^* 为满意的，否则不符合要求，η 越小表明近似值 x^* 的精度越高.

例如，$V=220\pm1(\mathrm{V})$，$c=(2.997\,902\pm0.000\,009)\times10^{10}$ cm/s(光速的准确值范围).

绝对误差有时不足以完全刻画近似值的精度，例如测量 200 m 的长度时，产生的误差为 1 cm，测量 1 m 的长度时产生的误差也是 1 cm，虽然两者的绝对误差相同，都是 1 cm，但由于所测量的长度差 200 倍，显然前一种测量要比后一种测量精度高，这说明要决定一个近似值的精确程度，除了绝对误差之外，还必须考虑本身大小，这就要引入相对误差的概念.

定义 3　绝对误差与精确值之比 $\frac{e}{x}=\frac{x-x^*}{x}$ 称为近似值 x^* 的相对误差.

在上例中，前一种测量的相对误差为 $\frac{1}{20\,000}$，而后一种测量的相对误差为 $\frac{1}{100}$，是前一种的 200 倍.

相对误差是个量纲为 1 的量，常用百分比表示，可正可负. 它在估计近似值运算结果的误

差时，比绝对误差更能反映出误差的特性. 因此，在误差分析中，相对误差比绝对误差更重要.

同绝对误差一样，因为 x 未知，相对误差也仅是一个估计值，人们常常也只关心它是否落在指定范围之内.

定义 4 若存在正数 δ，使 $|e_r|=\frac{|x-x^*|}{|x|}\leqslant\delta$，则称 δ 为相对误差限.

在实际应用中，由于 x 不知道，用上式无法确定 e_r，常取 $\frac{x-x^*}{x^*}$ 为 x^* 的相对误差. 于是用 $|e_r^*|=\frac{|e|}{|x^*|}=\frac{|x-x^*|}{|x^*|}$ 来作为相对误差限的另一定义.

事实上，$\left|\frac{e}{|x^*|}-\frac{e}{|x|}\right|=\frac{e(|x|-|x^*|)}{|x^*\cdot x|}\leqslant\frac{e^2}{|x^*\cdot x|}=\frac{\left(\frac{e}{x}\right)^2}{\left|\frac{x^*}{x}\right|}=o(e_r^2)$.

上式右端是一个高阶无穷小量，可见此时产生的影响非常小，可以忽略不计. 因此，用 e_r^* 来代替 e_r 不致引起明显的误差.

误差的运算性质

微分是描述绝对误差和相对误差的一个有效公式，设 x^* 为 x 的近似值，绝对误差 $e=x^*-x$ 可看作 x 的微分，记作

$$e=x^*-x=\mathrm{d}x.$$

x^* 作为 x 的近似值的相对误差为

$$e_r=\frac{x^*-x}{x}=\frac{\mathrm{d}x}{x}=\mathrm{d}\ln x.$$

它是 x 的对数微分，则有下列性质：

1) 和的绝对误差不超过绝对误差之和

事实上 $(x+y)^*=x^*+y^*$，$|(x+y)-(x+y)^*|=|(x-x^*)+(y-y^*)|\leqslant|x-x^*|+|y-y^*|$.

2) 积的相对误差等于各因子相对误差之和

事实上 $e_r(xy)=\mathrm{d}\ln(xy)=\mathrm{d}(\ln x+\ln y)=\mathrm{d}(\ln x)+\mathrm{d}(\ln y)=e_r(x)+e_r(y)$.

3) 商的相对误差等于各因子相对误差之差

$e_r(x/y)=e_r(x)-e_r(y)$.

2. 有效数字

对于一个近似数，我们常常希望能反映它的准确程度，因此我们引进有效数字的概念.

定义 5 若 x 的某一近似值 x^* 的误差限不超过某位数字的半个单位，则从该位数字起到 x^* 的左边的第一个非零数字为止的所有数字，都称为 x^* 的有效数字.

一般地，设 x^* 的规格化形式为：

$x^*=\pm0.\alpha_1\alpha_2\cdots\alpha_n\times10^m$，其中 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 为 $0,1,\cdots,9$ 中的某个自然数，$\alpha_1\neq0$，n 为正整数，m 为整数，若

$$e=|x-x^*|\leqslant\varepsilon=\frac{1}{2}\times10^{m-n},\ 1\leqslant n\leqslant m.$$

则称 x^* 有 n 位有效数字，或称它精确到 10^{m-n}. 其中，每一位数 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 都是 x^* 的有效数字.

例如，$\pi=3.14159265\cdots$
则 3.14 与 3.1416，3.143 分别有 3 位和 5 位及 3 位有效数字，它们分别精确到 0.01，0.0001，0.01.

近似值 x^* 所有数位上的数不一定都是有效数字，但如果 x^* 是四舍五入得到的近似值，那么可以证明，从 x^* 被保留的最后一位起直到 x^* 最左边的非零数字之间的所有数字都是有效数字.

例如：304，0.3040 都具有三位有效数字，但要注意 0.0304 和 0.030400 就不同了. 前者仅具有三位有效数字，精确到 0.0001；而后者则具有五位有效数字，精确到 0.000001. 可见，有效数字尾部的零不可随意消去，以免损失精度.

近似值 x^* 的有效数字与它的绝对误差密切相关，有效数位越多，误差越小；误差越小，有效数字也越多. 不但如此，还可以从有效数字求出其相对误差限.

定理 1 设近似值 $x^*=0.\alpha_1\alpha_2\cdots\alpha_n\times10^m$ 有 n 位有效数字，则其相对误差限

$$e_r^*\leqslant\frac{1}{2\alpha_1}\times10^{-(n-1)}.$$

证明：由 x^* 具有 n 位有效数字，则绝对误差限 $|x-x^*|\leqslant\frac{1}{2}\times10^{m-n}$，$|x^*|=10^m\times0.\alpha_1\alpha_2\cdots\alpha_n=10^{m-1}\times\alpha_1\alpha_2\cdots\alpha_n$.
由于 $\alpha_1\leqslant\alpha_1\alpha_2\cdots\alpha_n\leqslant\alpha_1+1$，$\alpha_1\times10^{m-1}\leqslant|x^*|\leqslant(\alpha_1+1)\times10^{m-1}$.
所以相对误差限为

$$|e_r^*|=\left|\frac{x^*-x}{x^*}\right|\leqslant\frac{\frac{1}{2}\times10^{m-n}}{\alpha_1\times10^{m-1}}\leqslant10^{-(n-1)}\times\frac{1}{2\alpha_1}.$$

定理 1 不仅反映了有效数字与其相对误差之间的关系，并且也说明了有效数字的位数，充分反映了近似值的相对精确度.

定理 1 的逆定理也是成立的.

定理 2 设近似值 $x^*=\pm0.\alpha_1\alpha_2\cdots\alpha_n\times10^m$ 的相对误差限不大于 $\frac{1}{2(\alpha_1+1)}\times10^{-n+1}$，则它至少有 n 位有效数字.

证明：$|x^*|\leqslant(\alpha_1+1)\times10^{m-1}$，$|e|=|x^*||e_r^*|\leqslant(\alpha_1+1)\times10^{m-1}\times\frac{1}{2(\alpha_1+1)}\times10^{-(n-1)}=\frac{1}{2}\times10^{m-n}$，即 x^* 至少有 n 位有效数字.

例 1 为了使积分 $I=\int_0^1 e^{-x^2}dx$ 的近似值 I^* 具有相对误差限不超过 0.1%，问至少取几位有效数字？

解：

$$I=\int_0^1\left(1-x^2+\frac{x^4}{2!}-\frac{x^6}{3!}+\cdots\right)dx$$
$$=1-\frac{1}{3}+\frac{1}{5\times2!}-\frac{1}{7\times3!}+\cdots$$

令

$$\frac{1}{(2k+1)\times k!}<0.0005，得 k\geqslant6.$$

$$I^* = 1 - \frac{1}{3} + \frac{1}{5\times 2!} - \frac{1}{7\times 3!} + \frac{1}{9\times 4!} - \frac{1}{11\times 5!}$$

$$\approx 1 - 0.3333 + 0.1 - 0.0238 + 0.0046 - 0.0008 = 0.7468.$$

I 的第一位数字 $\alpha_1=7$.

$|e_r(I)| \leqslant \left|\frac{1}{2\alpha_1}\times 10^{-n+1}\right| = \frac{1}{14}\times 10^{-n+1} \leqslant 0.001, 10^{-n+1} \leqslant 0.014$,所以 $n\geqslant 3$.

至少 3 位有效数字 $I^*=0.747$.

1.3 数值计算的原则

1.3.1 算 法

广义上说算法是指为解决某一问题而采取的方法和步骤.计算机的算法可分为两大类:数值运算的算法与非数值运算的算法.本书所指的是前者,数值计算的算法是指由加减乘除等算术运算、逻辑运算、顺序的规定求某类数学问题的数值解的步骤.它可用图框、算法语言、数学公式和文字说明、算法程序等来描述.本书采用数学公式与文字说明来描述算法,符合学生的理解习惯,易于学生接受.能否制定和使用优越的算法是科学计算成败的关键.本书所有的算法程序采用 C 语言编写,因为 C 语言简洁,使用方便,运算符丰富,具有现代化语言的各种数据结构.程序执行效率高,表达与运算能力强,应用广泛.

算法不止一种,优越的算法是指稳定的,占有内存小,误差不超过指定的误差限,运算量小,程序简单,不溢出,若运用迭代法必须是收敛的,收敛速度越快越好等,这是选择算法时应遵循的原则.

1.3.2 设计算法时应注意的事宜

为了使所使用的算法是优越的算法,在设计算法时应注意的事宜如下.

1. 减少运算次数

减少运算次数的目的是减少误差的积累和节省机时.

例如,计算 $3x^3+2x^2+x+1$ 时,采用直接运算要用 3 次加法、6 次乘法,而改变算法:$[(3x+2)x+1]x+1$,则仅需 3 次加法,3 次乘法.

再例如,求 $\ln 2$,用泰勒展开式

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \qquad x\in(-1,1]$$

取 $x=1, \ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{(-1)^{n+1}}{n} + \cdots$

改变算法:

$$\ln\frac{x+1}{1-x} = \ln(x+1) - \ln(1-x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

$$- \left(-x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots\right) = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots + \frac{x^{2n+1}}{2n+1} + \cdots\right), |x|<1.$$

取 $x=\frac{1}{3}, \ln 2 = 2\times\frac{1}{3}\left[1+\frac{1}{3}\times\left(\frac{1}{3}\right)^2+\frac{1}{5}\times\left(\frac{1}{3}\right)^4+\cdots\right]$.

利用第一式计算时取前十万项，可以精确到 10^{-5}，利用第二式计算时取前 9 项，可以精确到 10^{-10}.

2. 防止相近的两个数相减

防止相近的两个数相减是为了避免损失大量有效数字.

例如，设 $x=32.107\ 809$，$y=32.105\ 783$ 为准确数. $x^*=32.107\ 768$，$y^*=32.105\ 765$ 为它们的近似数，分别具有 6 位有效数字.

$$x-y=0.020\ 26, x^*-y^*=0.002\ 003.$$

x^*-y^* 作为 $x-y$ 的近似解仅具有 2 位有效数字，损失了 4 位有效数字. 有时，通过改变算法可以避免相近数相减.

例如 求方程 $x^2-18x+1=0$ 的根，保留 4 位有效数字.

解：用公式法解得

$$x_1=\frac{18+\sqrt{18^2-4}}{2}=9+\sqrt{80}\approx 17.94,$$

$$x_2=\frac{18-\sqrt{18^2-4}}{2}=9-\sqrt{80}\approx 9.000-8.944=0.056.$$

可见第二个根仅有两位有效数字，精度较差.

若第三个根改用韦达定理计算.

$x_2=\dfrac{1}{x_1}\approx 0.055\ 74$，第二个根也具有 4 位有效数字了.

3. 防止数字的溢出

例如，假设计算机的计算范围 $|F|\leqslant 10^{20}$，求 $\sqrt{a^2+b^2}$.

取 $a=2\times 10^{14}$，$b=3\times 10^{12}$.

若直接计算：$a^2=(2\times 10^{14})^2=4\times 10^{28}$，溢出了.

改变算法为：$a\sqrt{\left(\dfrac{b}{a}\right)^2+1}$ 就不会溢出了.

4. 防止小数被大数“吃掉”

计算机在进行计算时，要把所运算的数都写成绝对值小于 1 而阶码相同的数. 如 $a=10^8+1$ 必须改写成

$$a=0.1\times 10^9+0.000\ 000\ 001\times 10^9$$

如果计算机只能表示 8 位小数，则算出 $a=0.1\times 10^9$，大数“吃掉”小数，这种情形有时会产生很大误差.

例如，求方程 $x^2-(10^9+1)x+10^9=0$ 的根，应用公式法解：$x_{1,2}=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a}$

而 $\quad b^2-4ac=(10^9+1)^2-4\times 1\times 10^9=10^{18}+1-2\times 10^9\approx 0.1\times 10^{19}=10^{18}.$

所以 $$x_1=\frac{10^9+10^9}{2}=10^9, x_2=\frac{10^9-10^9}{2}=0.$$

而此方程的根实际上应为 $x_1=10^9$，$x_2=1$. x_2 与以上运算结果相差悬殊. 造成此问题的原因是大数“吃掉”小数，当然本题也可以改变算法来提高精度.