Numerical Analysis

哈尔滨工业大学（深圳）理学院数学学科
杨云云
http://faculty.hitsz.edu.cn/yangyunyun

Teaching Staff

Instructor:

Dr. Yunyun Yang

Associate Professor

Office: G710

E-mail: yangyunyun@hit.edu.cn

Teaching Assistants:

Class 2	Class 3

What is this course about?

- In real world, many problems are usually complicated and often too complex to solve analytically!
- such as
  - the problems of finding solutions for an equation with high order;
  - the problems of finding an integration of some functions;
  - the problems of finding the solutions for differential equations;
  - and so on.

- Turning mathematical models of an engineering problem into numerical results, we need to study the methods how to do this and what are the behaviors of these methods - this is what this course is about.

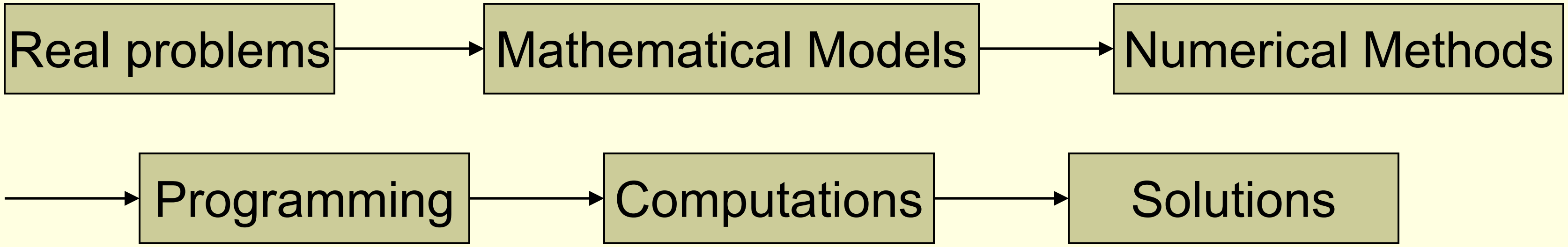
What is this course about?

- Numerical Analysis, as a mathematical course closely related to computer computations, has the following natures:

- It always results in the 4 simple and logical calculations;
- It addresses the efficiency rather than the existence and the uniqueness of the solutions;
- It cares about the skills of getting the solutions;
- It pays attentions on the theoretical study on the convergence, stability and the reliability.

What is this course about?

- Throughout our career as engineers, we will be often facing the task of turning mathematical models of engineering systems into numerical results, which usually has the following steps



- So, it is significant and important for us to study this course.

Grading System

- Class Attendance & Quizzes20%
- Homework & Assignments 20%
- Final Exam 60%

Plagiarism is NOT allowed!

Students can work together in groups to work out how to do the assignment problems but the writing up of the solutions should be each student’s own work.

Reference Books

Textbook:

Numerical Analysis - Mathematics of Scientific Computing (Third Edition) by David Kincaid and Ward Cheney

- Lecture notes, interpretation
- R. L. Burden and J. D. Faires, Numerical Analysis, (seven edition), Higher Education Press, 2002, pp809.
- 李庆扬，王能超，易大义，数值分析（第5版），清华大学出版社 & 施普林格出版社，2008
- 李庆扬，关治，白峰杉，数值计算原理，清华大学出版社，2005
- 刘春凤等编，应用数值分析，冶金工业出版社，2005
- 蔺小林等编著，现代数值分析，国防工业出版社，2004
- 凌永祥等编著，计算方法教程，西安交通大学出版社(第二版)，2005
- 林成森编，数值计算方法(第二版，上、下册)，科学出版社，2005

Main Contents

The contents include mainly two parts:

- Numerical Algebra (Chapters 1,2,3,4)
- Numerical Approximation (Chapters 6,7,8)

- You need to have the basic knowledge of the following subjects:

1. Linear Algebra
2. Calculus
3. Differential Equations

Main Contents

- Chapter 1 Mathematical Preliminaries
- Chapter 2 Computer Arithmetic
- Chapter 3 Solution of Nonlinear Equations
- Chapter 4 Solving Systems of Linear Equations
- Chapter 6 Approximating Functions
- Chapter 7. Numerical Differentiation and integration
- Chapter 8. Numerical Solution of Ordinary Differential Equations

Main Contents

- Chapter 1 Mathematical Preliminaries
  - ✓ Taylor's Theorem
  - ✓ Orders of Convergence
- Chapter 2 Computer Arithmetic
  - ✓ Floating-Point Numbers and Roundoff Errors
  - ✓ Absolute and Relative Errors: Loss of Significance
  - ✓ Stable and Unstable Computations: Conditioning
- Chapter 3 Solution of Nonlinear Equations
  - ✓ Bisection Method;
  - ✓ Newton's Method;
  - ✓ Secant Method

Main Contents

- Chapter 4 Solving Systems of Linear Equations
  - ✓ Matrix Algebra: Basic Concepts, Elementary Operation, Equivalent System, Inverse Matrix.
  - ✓ LU Factorization: Easy to Solve System, LU Factorization, LDU Factorization, Doolittle Factorization, Crout’s Factorization, Cholesky Factorization.
  - ✓ Pivoting and Constructing Algorithm: Basic Gaussian Elimination; Pivoting Guassian Elimination with Scaled Pivoting; Diagonal Dominant Matrix; Triangular System.
  - ✓ Norms and the Analysis of Errors: Vector norms, Matrix Norms, Condition Number.
  - ✓ Solution of Equations by Iterative Methods: Jacobi Method, Gauss-Seidel Method, SOR Method

Main Contents

- Chapter 6 Approximating Functions

✓ Polynomial Interpolation: Method of Undetermined Coefficents, The Error of the Interpolation, Lagrange Interpolation Method, Newton Interpolation (Divided Difference), Difference and the Interpolation for Even Spaced Nodes, Hermite Interpolation.

✓ Best Approximation---Least Squares Method: Discrete Least Square Approximation; Orthogonal Polynomial and Least Squares Approximation

Main Contents

- Chapter 7. Numerical Differentiation and integration
- ✓ Numerical Differentiation and Richardson Extrapolation
- ✓ Numerical Integration Based on Interpolation: Newton-Cotes Formula, Trapezoidal Formula, Composite Trapezoidal Formula, Simpson’s Rule, Composite Simpson’s Rule, General Integration Formula;
- ✓ Gaussian Quadrature;
- ✓ Romberg Integration

Main Contents

- Chapter 8. Numerical Solution of Ordinary Differential Equations
  - ✓ The Existence and Uniqueness of Solutions;
  - ✓ Taylor-Series Method: Euler Method, Backward Euler Method, Modified Euler Method;
  - ✓ Runge-Kutta Methods;
  - ✓ Multistep Methods

Chapter 1
Mathematical Preliminaries

----Taylor’s Theorem

----Orders of Convergence

Taylor’s Theorem in various forms is fundamental to many numerical procedures and is an excellent starting point for the study of scientific computing since no advanced mathematical concepts are required.

Taylor's Theorem with Lagrange Remainder

If $f \in C^n[a,b]$ and if $f^{(n+1)}$ exists on the open interval (a,b), then for any points c and x in the closed interval $[a,b]$,

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k + E_n(x) \qquad (1)$$

where, for some point ξ between c and x, the error term is

$$E_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-c)^{n+1}$$

Here "ξ between c and x" means that either $c < \xi < x$ or $x < \xi < c$ depending on the particular values of x and c involved.

An important special case arises when $c = 0$. Equation (1) becomes the **Maclaurin series** for $f(x)$:

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(0)x^k + E_n(x) \qquad (2)$$

where $$E_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi)x^{n+1}$$

We can obtain Taylor series for many important functions such as

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \qquad (-\infty < x < \infty)$$

$$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} \qquad (-\infty < x < \infty)$$

$$\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k \qquad (-1 < x < \infty)$$

$$\frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k \qquad (-1 < x < 1)$$

EXAMPLE: Determine the Taylor series of the following function, using Taylor's Theorem, *f(x)* = ln*x* taking *a* = 1, *b* = 2, and *c* = 1.

Solution : The various derivatives are

$$f'(x) = x^{-1},\ f''(x) = -x^{-2},\ f'''(x) = 2x^{-3},\ f^{(4)}(x) = -6x^{-4},$$

and so on. Next, we obtain the general term

$$f^{(k)}(x) = (-1)^{k-1}(k-1)!\,x^{-k} \qquad (k \geq 1)$$

Clearly, at $x = 1$, we have

$$f^{(k)}(1) = (-1)^{k-1}(k-1)! \qquad (k \geq 1)$$

Of course, $f^{(0)}(1) = f(x) = \ln 1 = 0.$

Putting all of this into Taylor's formula (1), gives us

$$\ln x = \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k}(x-1)^k + E_n(x) \qquad (1 \leq x \leq 2)$$

where $$E_n(x) = (-1)^n \frac{1}{n+1} \xi^{-(n+1)} (x-1)^{n+1} \qquad (1 < \xi < x)$$

$E_n(x)$, can be regarded as an error term.

Taylor's Theorem with Integral Remainder

If $f \in C^{n+1}[a,b]$, then for any points c and x in the closed interval $[a,b]$,

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(c)(x-c)^k + R_n(x) \qquad (1)$$

where,

$$R_n(x) = \frac{1}{n!} \int_c^x f^{(n+1)}(t)(x-t)^n \, dt$$

Orders of Convergence

Some special terminology is used to describe the rapidity with which a sequence converges. Let $[x_n]$ be a sequence of real numbers tending to a limit x^*.

We say that the rate of convergence is at least **linear** if there are a constant $c < 1$ and an integer N such that

$$\left|x_{n+1} - x^*\right| \leq c\left|x_n - x^*\right| \quad (n \geq N)$$

We say that the rate of convergence is at least **superlinear** if there exist a sequence ε_n tending to 0 and an integer N such that

$$\left|x_{n+1} - x^*\right| \leq \varepsilon_n\left|x_n - x^*\right| \quad (n \geq N)$$

Orders of Convergence

The convergence is at least **quadratic** if there are a constant C (not necessarily less than 1) and an integer N such that

$$\left|x_{n+1} - x^*\right| \le C\left|x_n - x^*\right|^2 \qquad (n \ge N)$$

In general, if there are positive constants C and p and an integer N such that

$$\left|x_{n+1} - x^*\right| \le C\left|x_n - x^*\right|^p \qquad (n \ge N)$$

we say that the rate of convergence is of **order** p at least.

Orders of Convergence

As an example of a rapidly convergent sequence, consider the one defined recursively by putting

$$\begin{cases} x_1 = 2 \\ x_{n+1} = \frac{1}{2} x_n + \frac{1}{x_n} \quad (n \geq 1) \end{cases}$$

The elements of this sequence are

$$x_1 = 2.00000\ 0$$
$$x_2 = 1.50000\ 0$$
$$x_3 = 1.41666\ 7$$
$$x_4 = 1.41421\ 6$$

The limit is $\sqrt{2} = 1.41421\ 3562\cdots$, and the sequence is converging to its limit with great rapidity. Using double-precision computation, we find evidence that

$$\frac{\left|x_{n+1} - \sqrt{2}\right|}{\left|x_n - \sqrt{2}\right|^2} \leq 0.36$$

Such a condition corresponds to **quadratic convergence**.

Orders of Convergence

Theorem 1 For the iteration $x_{k+1}=\varphi(x_k)$, if $\varphi^{(p)}(x)$ is continuous near x^*, and

$$\varphi'(x^*)=\varphi''(x^*)=\cdots=\varphi^{(p-1)}(x^*)=0,\ \varphi^{(p)}(x^*)\neq 0,$$

then the rate of convergence is of order p near x^*.

Example When using the following iteration to find the root $x^*=\sqrt{3}$ of the equation $x^2-3=0,$ what is the convergence order?

$$x_{k+1}=\frac{1}{2}\left(x_k+\frac{3}{x_k}\right)$$

Exercises

$Ex1.$ Verify the Taylor Expansion:

$$\frac{1}{1+x}=\sum_{k=0}^{\infty}(-1)^k x^k \qquad (-1<x<1)$$

$Ex2.$ If we use the iteration $x_{k+1}=\dfrac{x_k\left(x_k{}^2+3a\right)}{3x_k{}^2+a}$ to calculate $\sqrt{a}$, what is the order of convergence of the sequence $\left[x_k\right]$?

Chapter 2

Computer Arithmetic

2.0 Introduction
2.1 Floating-Point Numbers and Roundoff Errors
2.2 Absolute and Relative Errors: Loss of Significance
2.3 Stable and Unstable Computations: Conditioning

2.1 Floating-Point Numbers and Roundoff Errors

Most computers deal with real numbers in the binary number system, in contrast to the decimal number system that humans prefer to use. The binary system uses 2 as the base in the same way that the decimal system uses 10. We recall first how our familiar number representation works. When a real number such as 427.325, in the decimal system, is written out in more detail, we have:

$$427.325 = 4\times 10^{2} + 2\times 10^{1} + 7\times 10^{0} + 3\times 10^{-1} + 2\times 10^{-2} + 5\times 10^{-3}$$

The expression on the right contains powers of 10 together with the digits $0,1,2,3,4,5,6,7,8,9.$

$$(1001.11101)_2 = 1\times 2^{3} + 0\times 2^{2} + 0\times 2^{1} + 1\times 2^{0} + 1\times 2^{-1} + 1\times 2^{-2} + 1\times 2^{-3} + 0\times 2^{-4} + 1\times 2^{-5}$$

This is the same real number as 9.90625 in decimal notation.

2.1 Floating-Point Numbers and Roundoff Errors

- Rounding Concept

Computer is not able to operate using real numbers expressed with more than a fixed number of digits. The word length of the computer places a restriction on the precision with which real numbers can be represented. Rounding is an important concept in scientific computing. Consider a positive decimal number x of the form $0.\bullet\bullet\bullet\ldots\bullet\bullet\bullet$ with m digits to the right of the decimal point. One rounds x to n decimal places $(n < m)$ in a manner that depends on the value of the $(n+1)st$ digit. If this digit is 0, 1, 2, 3 or 4, then the nth digit is not changed and all following digits are discarded. If it is 5, 6, 7, 8, or 9, then the nth digit is increased by one unit and the remaining digits are discarded. The situation with 5 as the $(n+1)st$ digit can be handled in a variety of ways .

2.1 Floating-Point Numbers and Roundoff Errors

- Normalized Scientific Notation

In the decimal system, any real number can be expressed in **normalized scientific notation**. This means that the decimal point is shifted and appropriate powers of 10 are supplied so that all the digits are to the right of the decimal point and the first digit displayed is not 0. Examples are

$$732.5051 = 0.7325051 \times 10^{3}$$

$$-0.005612 = -0.5612 \times 10^{-2}$$

In general, a nonzero real number x can be represented in the form

$$x = \pm r \times 10^{n}$$

where r is a number in the range $\frac{1}{10} \leq r < 1$ and n is an integer (positive, negative, or zero). Of course, if $x = 0$, then $r = 0$; we can adjust n so that r lies in the given range.

2.2 Absolute and Relative Errors:

- Absolute and Relative Errors

When a real number x is approximated by another number x^*, the error is $x - x^*$. The absolute error is

$$\left| x - x^* \right|$$

and the relative error is

$$\left| \frac{x - x^*}{x} \right|$$

In scientific measurements, it is almost always the relative error that is significant. Information about the absolute error is usually of little use in the absence of knowledge about the magnitude of the quantity being measured. (An error of only 1 meter in determinig the distance from Jupiter to Earth would be quite remarkable, but you would not want a surgeon to make such an error in an incision!)

2.2 Absolute and Relative Errors

- errors due to roundoff

EXAMPLE: Let us use a decimal machine operating with five decimal digits in its floating-point number system, and determine the relative error in adding , subtracting , multiplying , and dividing the two machine numbers

$$x = 0.31426 \times 10^3 \quad y = 0.92577 \times 10^5$$

Solution: Using a double-length accumulator for the intermediate results, we have

$$x + y = 0.9289126000 \times 10^5$$

$$x - y = -0.9226274000 \times 10^5$$

$$x * y = 0.2909324802 \times 10^8$$

$$x \div y \approx 0.3394579647 \times 10^{-2}$$

(continued)

2.2 Absolute and Relative Errors

- errors due to roundoff

The computer with five decimal digits stores these in rounded form as

$$fl(x+y) = 0.92891 \times 10^5$$

$$fl(x-y) = -0.92263 \times 10^5$$

$$fl(x*y) = 0.29093 \times 10^8$$

$$fl(x \div y) = 0.33946 \times 10^{-2}$$

The relative errors in these results are

$$2.8 \times 10^{-6}, \quad 2.8 \times 10^{-6}, \quad 8.5 \times 10^{-6}, \quad 6.0 \times 10^{-6},$$

respectively.

They all are less than 10^{-5}.

2.2 Absolute and Relative Errors:

- Loss of Significance

Although roundoff errors are inevitable and difficult to control, other types of errors in computation are under our control. In the subject of numerical analysis there are errors of various kinds. Here we shall take up one type of error that is often the result of careless programming. To see the sort of situation in which a large relative error can occur, for example, we subtract the two numbers

$$x = 0.3721478693, \quad y = 0.3720230572$$
$$x - y = 0.0001248121$$

we compute these terms up to five digits

$$fl(x) = 0.37215, \; fl(y) = 0.37202, \;\; fl(x) - fl(y) = 0.00013$$

The relative error is then very large:

$$\left|\frac{x - y - \left[fl(x) - fl(y)\right]}{x - y}\right| = \left|\frac{0.0001248121 - 0.00013}{0.0001248121}\right| \approx 4\%$$

2.2 Absolute and Relative Errors:

- Substraction of Nearly Equal Quantities

EXAMPLE: The assignment statement

$$y \leftarrow \sqrt{x^2+1}-1$$

involves subtractive cancellation and loss of significance for small values of x. How can we avoid this trouble ?

Solution : Rewrite the function in this way

$$y = \left(\sqrt{x^2+1}-1\right)\left(\frac{\sqrt{x^2+1}+1}{\sqrt{x^2+1}+1}\right) = \frac{x^2}{\sqrt{x^2+1}+1}$$

Thus, the difficulty is avoided by reprogramming with a different assignment statement

$$y \leftarrow \frac{x^2}{\left(\sqrt{x^2+1}+1\right)}$$

2.2 Absolute and Relative Errors:

- Substraction of Nearly Equal Quantities

Example: we use the same computer to calculate the solution for

$$x^2 - 18x + 1 = 0$$

The two roots of this equation is

$$x_1 = 9 - \sqrt{80}, \quad x_2 = 9 + \sqrt{80}$$

with this computer, $\sqrt{80} = 0.8944 \times 10^1$, then we have solutions

$$x_1 = 0.5600 \times 10^{-1}, \quad x_2 = 0.1794 \times 10^2$$

However, if we use formula $x_1 \times x_2 = 1$ we have

$$x_1 = \frac{1}{9 + \sqrt{80}} = 0.5574 \times 10^{-1}$$

The latter is better since it avoids the two close number to substract.

2.3 Stable and Unstable Computations: Conditioning

- We introduce another theme that occurs repeatedly in numerical analysis: the distinction between numerical processes that are **stable** and **those that are not**.

- Closely related are the concepts of **well conditioned** problems and **ill conditioned** problems.

2.3 Stable and Unstable Computations: Conditioning

- Numerical Instability

A numerical process is unstable if small errors made at one stage of the process are magnified in subsequent stages and seriously degrade the accuracy of the overall calculation. The following example helps to explain this concept. Consider the sequence of real numbers defined inductively by

$$\begin{cases} x_0 = 1 \qquad x_1 = \dfrac{1}{3} \\ x_{n+1} = \dfrac{13}{3} x_n - \dfrac{4}{3} x_{n-1} \end{cases} \qquad (\mathrm{n} \geq 1) \qquad\qquad (1)$$

It is easily seen that this recurrence relation generates the sequence

$$x_n = \left(\frac{1}{3}\right)^n \qquad\qquad (2)$$

Equation (2) is obviously true for $n = 0$ and $n = 1$. If its validity is granted for $n \leq m$, then its validity for $n = m + 1$ follows from

2.3 Stable and Unstable Computations: Conditioning

- Numerical Instability

$$x_{m+1} = \frac{13}{3} x_m - \frac{4}{3} x_{m-1} = \frac{13}{3}\left(\frac{1}{3}\right)^m - \frac{4}{3}\left(\frac{1}{3}\right)^{m-1}$$

$$= \left(\frac{1}{3}\right)^{m-1} [\frac{13}{9} - \frac{4}{3}] = \left(\frac{1}{3}\right)^{m+1}$$

If the inductive definition (1) is used to generate the sequence numerically, then some of the computed terms are grossly inaccurate. Here are a few of them, calculated on a 32-bit computer:

$x_0 = 1.0000000$

$x_1 = 0.3333333$ (7 correctly rounded significant digits)

$x_2 = 0.1111112$ (6 correctly rounded significant digits)

$x_3 = 0.0370373$ (5 correctly rounded significant digits)

$x_4 = 0.0123466$ (4 correctly rounded significant digits)

(continued)

2.3 Stable and Unstable Computations: Conditioning

- Numerical Instability

$x_5 = 0.0041187$ (3 correctly rounded significant digits)

$x_6 = 0.0013857$ (2 correctly rounded significant digits)

$x_7 = 0.0005131$ (1 correctly rounded significant digits)

$x_8 = 0.0003757$ (0 correctly rounded significant digits)

$x_9 = 0.0009437$

$x_{10} = 0.0035887$

$x_{11} = 0.0142927$

$x_{12} = 0.0571502$

$x_{13} = 0.2285939$

$x_{14} = 0.9143735$

$x_{15} = 3.657493$ (incorrect with relative error of 10^8)

2.3 Stable and Unstable Computations: Conditioning

- Numerical Instability

This algorithm is therefore unstable. Any error present in x_n is multiplied by $13/3$ in computing x_{n+1}. Hence there is possibility that the error in x_1 can be propagated into x_{15} with a factor of $(13/3)^{14}$. Since the absolute error in x_1 is around 10^{-8} and since $(13/3)^{14}$ is roughly 10^9, the error in x_{15} due solely to the error in x_1 could be as much as 10. In fact, additional roundoff errors occur in computing each of $x_2, x_3, \ldots$, and these errors may also be propagated into x_{15} with various factors of the form $(13/3)^k$.

2.3 Stable and Unstable Computations: Conditioning

- Conditioning

The words condition and conditioning are used informally to indicate how sensitive the solution of a problem may be to small relative changes in the input data.

A problem is ill conditioned if small changes in the data can produce large changes in the output.

For certain types of problems , a condition number can be defined. If that number is large , it indicates an ill-conditioned problem. The condition number depends on the numerical method that we choose for the solution of the problem.

2.3 Stable and Unstable Computations: Conditioning

- Conditioning

Suppose our problem is simply to evaluate a function f at a point x. We ask, if x is perturbed slightly, what is the effect on $f(x)$? If this question refers to absolute error, we can invoke the Mean-Value Theorem and write

$$f(x+h)-f(x)=f'(\xi)h \approx f'(x)h, \quad x<\xi<x+h$$

Thus, if $f'(x)$ is not too large, the effect of the perturbation on $f(x)$ is small. Usually, however, it is the relative error that is of significance in such questions. In perturbing x by the amount h, we have h/x as the relative size of the perturbation. Likewise, when $f(x)$ is perturbed to $f(x+h)$, the relative size of that perturbation is

$$\left|\frac{f(x+h)-f(x)}{f(x)}\right| \Big/ \left|\frac{h}{x}\right| \approx \left|\frac{xf'(x)}{f(x)}\right| = C_p$$

Thus, C_p severs as a **condition number** for this problem. If C_p is large, we call that the problem is ill-conditioned.

2.3 Stable and Unstable Computations: Conditioning

- Conditioning

For example, $f(x) = x^n$, then $C_p \approx n$, it means that the relative error may be magnified n times. If $n = 10$, it then $f(1) = 1$, $f(1.02) \approx 1.24$, let $x = 1$, $h = 0.02$, $h / x = 2\%$, $\left(f(x+h) - f(x) \right) / f(x) = 24\%$, C_p=12, in this case, we can say that the problem is ill-conditioned. Generally, if $C_p \geq 10$, the problem is thought to be ill-conditioned.

Exercises

$Ex1$. The followings are two numbers after round-off. How many digits of significant do they have? Give the error bounds and the relative error bounds of them.

(i) $\mathrm{x}_1^* = 4.7021$ (ii) $\mathrm{x}_2^* = 0.067$

$Ex2$. $x > 0$, the relative error of x is δ, what is the error of $\ln x$?

$Ex3$. Assume that the relative error of x is 2%, what is the relative error of x^n?