Introduction

Numerical Analysis is concerned with the design and analysis of algorithms for solving mathematical problems that arise in many fields, especially science and engineering. *Scientific computing* could be regarded as a combination of modeling, visualization, and numerical analysis. In particular, *Numerical Methods*, the foundation of *Scientific computing* deals with quantities that are continuous, as opposed to discrete which occurred in most other parts of computer science.

- A Review of Calculus:

  $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

  Taylor's Theorem

  Orders of Convergence: Big *Oh* and little *oh*

  Truncation and Rounding

- Types of Errors

  Absolute Error and Relative Error

  Data Error and Computational Error

  Truncation Error and Rounding Error

  Sensitity and Condition Number

  Stability and Accuracy

- Computer Arithmetic

  Floating-Point Numbers

  Representation and Normalization

  Machine Precision

  Summing a series $\sum_{n=1}^{\infty} \frac{1}{n}$

  Quadratic formula (Solving $ax^2 + bx + c = 0$)

- Mathematical Software and Environments

  Reliability, Accuracy, Portability

  Robustness, Efficiency, Maintainability

  Usability, Applicability

  Matlab, IMSL (PV-Wave), Netlib, NAG

- Historical Notes from Mathematicians

  Newton (1642∼1727), Euler (1707∼1783), Lagrange (1736∼1813)

  Laplace (1749∼1827), Legendre (1752∼1833), Gauss (1777∼1855)

  Cauchy (1789∼1857), Jacobi (1804∼1851), Adams (1819∼1892),

  Chebyshev (1821∼1894), Hermite (1822∼1901), Laugerre (1834∼1886)

A Review of Calculus

Show that

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad \forall\ a,b>0$$

where

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx, \quad a>0$$

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx, \quad a,b>0$$

[*Hint:*] let $u = x/(x+y)$, $v = x+y$, then $x = uv$, $y = v - uv = (1-u)v$, then the Jacobian $J(x,y) = v$. Furthermore, show that $\Gamma(a)\Gamma(b) = \Gamma(a+b)B(a,b)$.

Taylor's Theorem: If $f \in C^n[a,b]$ and $f^{(n+1)}$ exists on the open interval (a,b), then for any $c \in (a,b)$ and $x \in [a,b]$,

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

where ξ is between c and x. When $c=0$, the above expansion is called a Maclaurin series for $f(x)$.

Examples

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

$$cosx = \sum_{k=0}^{\infty}(-1)^k \frac{x^{2k}}{(2k)!}, \quad -\infty < x < \infty$$

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

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

$$e^{-x} = \sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k!}, \quad x \in R$$

$$\sqrt{1+2x} = 1 + x + \sum_{k=2}^{\infty} \alpha_k x^k, \quad -\frac{1}{2} < x < \frac{1}{2}$$

Orders of Convergence
Big Oh and little oh

$$limit_{n\to\infty}\frac{n+1}{n}=1$$

$$limit_{n\to\infty}(1+\frac{1}{n})^n=e\approx 2.71828$$

Definition: Let $\{x_n\}$ and $\{\alpha_n\}$ be two different sequences, we write

$$x_n=O(\alpha_n)$$

if $\exists$ constants C and n_0 such that $|x_n|\leq C|\alpha_n|\ \forall\ n\geq n_0$. Here we say that x_n is "big oh" of α_n. On the other hand,

$$x_n=o(\alpha_n)$$

if $\exists\ \epsilon_n\to 0$ such that $|x_n|\leq\epsilon_n|\alpha_n|$. Here we say that x_n is "little oh" of α_n.

Examples

$$\frac{n}{n^2+1}=O(\frac{1}{n})$$

$$\frac{1}{nln(n)}=o(\frac{1}{n})$$

$$\frac{1}{n}=o(\frac{1}{ln(n)})$$

$$\frac{5}{n}+e^{-n}=O(\frac{1}{n})$$

$$e^{-n}=o(\frac{1}{n^2})$$

$$sin(x)=x-\frac{x^3}{6}+O(x^5)$$

$$cos(x)=1-\frac{x^2}{2}+O(x^4)$$

Types of Errors

Let y be the true value, and let $\hat{y}$ be the approximate value, then

$$\begin{aligned} Absolute\ error &= \hat{y} - y \\ Relative\ error &= \frac{\hat{y} - y}{y} \end{aligned}$$

Let $f : R \to R$. Suppose that we must work with inexact input $\hat{x}$ and we can compute only an approprixation to the function $\hat{f}$. Then, we have

$$Total\ error \ = \ \hat{f}(\hat{x}) - f(x) \ = \ [\hat{f}(\hat{x}) - f(\hat{x})] \ + \ [f(\hat{x}) - f(x)]$$

$$Computation\ error \ = \hat{f}(\hat{x}) - f(\hat{x})$$

$$Propagated\ data\ error \ = f(\hat{x}) - f(x)$$

♣ Computing $sin(\pi/8)$

Use $\hat{f}(x) = x$ instead of $f(x) = sin(x)$ and use $\hat{x} = 3/8$ to approximate $x = \pi/8$. By calculator, we have

$$f(\pi/8) = sin(\pi/8) \approx 0.3827 \ \ and \ \ f(3/8) \approx 0.3663$$

On the other hand,

$$\hat{f}(\hat{x}) - f(\hat{x}) = \hat{x} - sin(3/8) = 0.3750 - 0.3663 = 0.0087$$

$$f(\hat{x}) - f(x) = 0.3663 - 0.3827 = -0.0164$$

So the total error is

$$\hat{f}(\hat{x}) - f(x) \ = \ [\hat{f}(\hat{x}) - f(\hat{x})] \ + \ [f(\hat{x}) - f(x)] \ = \ 0.0087 + (-0.0164) = -0.077$$

♣ $3.14159 \approx 3.1415$ by truncation, $3.14159 \approx 3.1416$ by round-off.

♣ $f(x+h) = f(x) + f'(x)h + f''(\theta)h^2/2$ for some $\theta \in [x, x+h]$

Condition Number

Suppose that we want to compute $y = f(x)$, where $f : R \to R$, but we obtain instead an approximate value $\hat{y}$. The discrepancy $\Delta y = \hat{y} - y$ is called the *forward error*. On the other hand, how much data error in the initial input would be required to explain all of the error in the final computed result? The quantity $\Delta x = \hat{x} - x$, where $f(\hat{x}) = \hat{y}$, is called *backward error*.

♣ Consider $f(x) = \sqrt{x}$, as an approximation to $y = \sqrt{2}$, $\hat{y} = 1.4$ has an absolute forward error $|\Delta y| = |1.4 - 1.4142\cdots|$. The backward error $|\Delta x| = |\hat{x} - x| = |(1.4)^2 - 2| = 0.04$. So, $|\Delta y|/y \approx 1\%$ and $|\Delta x|/x \approx 2\%$.

A problem is said to be insensitive, or *well-conditioned*, if a given relative change in the input data causes a reasonably commensurate relative change in the solution. Otherwise, it is called sensitive, or *ill-conditioned*.

$$\textit{Condition number} \;=\; \frac{|\Delta y/y|}{|\Delta x/x|} \;=\; \frac{|[f(\hat{x}) - f(x)]/f(x)|}{|[\hat{x} - x]/x|}$$

If f is continuously differentiable around x, then

$$\textit{Condition number} \;=\; |\frac{x f'(x)}{f(x)}|$$

□ Implement and discuss the following problems.

(1) $\sum_{n=1}^{\infty} \frac{1}{n}$ (Prob. 1.8 on p.45, Heath)

(2) Solve $x^3 + ax^2 + bx + c = 0$ (Prob. 1.11 on p.46, Heath)

(3) Evaluate $\|\mathbf{x}\|_2 = (\sum_{i=1}^{n} x_i^2)^{1/2}$ (Prob. 1.15 on p.47, Heath)

Miscellaneous Examples

(1) $A = \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$

(2) $\Gamma(\frac{1}{2}) = \int_0^\infty e^{-x} x^{-1/2} dx = \sqrt{\pi}$

(3) If $x_{n+2} = x_{n+1} + x_n, \quad for\ n \geq 0, \quad x_0 = 1,\ x_1 = 1$, then $x_n = \frac{1}{\sqrt{5}} \left[\frac{1+\sqrt{5}}{2}\right]^{n+1} - \frac{1}{\sqrt{5}} \left[\frac{1-\sqrt{5}}{2}\right]^{n+1}$, for $n \geq 0$.

(4) $x = (-1)^s q \times 2^m$, where $q = (1.f)_2$, $m = e - 127$, and $e = 0,\ or\ 255$ are reserved for $\pm 0,\ \pm\infty$.

(5) Write a program to solve the quadratic equation and test the values of coefficients as listed in Table 1 (*Heath*).

$$ax^2 + bx + c = 0$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad or \quad x = \frac{2c}{-b \mp \sqrt{b^2 - 4ac}}$$

Problem	a	b	c
(A)	6	5	-4
(B)	0	1	1
(C)	1	-10^5	1
(D)	1	-4	3.999999
(E)	10^{-155}	-10^{155}	10^{155}
(F)	6×10^{154}	5×10^{154}	-4×10^{154}

Table 1: **Miscellaneous Coefficients of Quadratic Equations.**

Vector Norms

Definition: A vector norm on R^n is a function

$$\tau \;:\; R^n \;\rightarrow\; R^+ = \{x \geq 0 |\; x \in R\}$$

that satisfies

(1) $\tau(\mathbf{x}) > 0 \;\; \forall\; \mathbf{x} \neq \mathbf{0},\; \tau(\mathbf{0}) = 0$

(2) $\tau(c\mathbf{x}) = |c|\tau(\mathbf{x}) \;\forall\; c \in R,\; \mathbf{x} \in R^n$

(3) $\tau(\mathbf{x} + \mathbf{y}) \leq \tau(\mathbf{x}) + \tau(\mathbf{y}) \;\forall\; \mathbf{x}, \mathbf{y} \in R^n$

Hölder norm (p-norm) $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p} \;\; for\; p \geq 1.$

(p=1) $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$ (Mahattan or City-block distance)

(p=2) $\|\mathbf{x}\|_2 = (\sum_{i=1}^n |x_i|^2)^{1/2}$ (Euclidean distance)

(p=∞) $\|\mathbf{x}\|_\infty = max_{1 \leq i \leq n}\{|x_i|\}$ (∞-norm)

Matrix Norms

Definition: A matrix norm on $R^{m\times n}$ is a function

$$\tau \;:\; R^{m\times n} \;\rightarrow\; R^{+} = \{x \geq 0 |\; x \in R\}$$

that satisfies

(1) $\tau(A) > 0 \;\; \forall\, A \neq O$, $\tau(O) = 0$

(2) $\tau(cA) = |c|\tau(A) \;\forall\, c \in R,\;\, A \in R^{m\times n}$

(3) $\tau(A+B) \leq \tau(A) + \tau(B) \;\; \forall\, A, B \in R^{m\times n}$

Consistency Property: $\tau(AB) \leq \tau(A)\tau(B) \;\; \forall\, A, B$

(a) $\tau(A) = max\{|a_{ij}| \mid 1 \leq i \leq m,\; 1 \leq j \leq n\}$

(b) $\|A\|_F = \left[\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right]^{1/2}$ (Fröbenius norm)

Subordinate Matrix Norm: $\|A\| = max_{\|\mathbf{x}\|\neq \mathbf{0}}\{\|A\mathbf{x}\|/\|\mathbf{x}\|\}$

(1) If $A \in R^{m\times n}$, then $\|A\|_1 = max_{1\leq j\leq n}\left(\sum_{i=1}^{m}|a_{ij}|\right)$

(2) If $A \in R^{m\times n}$, then $\|A\|_\infty = max_{1\leq i\leq m}\left(\sum_{j=1}^{n}|a_{ij}|\right)$

(3) Let $A \in R^{n\times n}$ be real symmetric, then $\|A\|_2 = max_{1\leq i\leq n}|\lambda_i|$, where $\lambda_i \in \lambda(A)$

Problems Solved by Matlab

Let A, B, H, **x**, **y**, **u**, **b** be matrices and vectors defined below, and $H = I - 2\mathbf{u}\mathbf{u}^t$

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 6 & 0 \\ -2 & 7 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} -3 & 1 & 0 \\ 1 & -3 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ 2 \\ -5 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

1. Let A=LU=QR, find L, U; Q, R.

2. Find determinants and inverses of matrices A, B, and H.

3. Solve $A\mathbf{x} = \mathbf{b}$, how to find the number of floating-point operations are required?

4. Find the ranks of matrices A, B, and H.

5. Find the characteristic polynomials of matrices A and B.

6. Find 1-norm, 2-norm, and ∞-norm of matrices A, B, and H.

7. Find the eigenvalues/eigenvectors of matrices A and B.

8. Find matrices U and V such that $U^{-1}AU$ and $V^{-1}BV$ are diagonal matrices.

9. Find the singular values and singular vectors of matrices A and B.

10. Randomly generate a 4×4 matrix C with $0 \leq C(i,j) \leq 9$.