Lecture 1: Numerical Analysis (UMA011)

Dr. Meenu Rani

School of Mathematics
TIET, Patiala
Punjab-India

General Information

Website:

https://sites.google.com/view/uma007numericalanalysis/home

Books:

1. Richard L. Burden, J. Douglas Faires, and Annette M. Burden, Numerical Analysis, 10th edition, 2015.
2. K. Atkinson and W. Han, Elementary Numerical Analysis, 3rd edition, John Willey and sons, 2004.
3. Brian Bradie, A Friendly Introduction to Numerical Analysis, Pearson Publishers, 2006.
4. Steven C. Chapra and Raymond P. Canale, Numerical Methods for Engineers, McGraw-Hill Higher Education; 6th edition, 2010.

Introduction

A major advantage for numerical technique is that a numerical answer can be obtained even when a problem has no analytical solution. However, result from numerical analysis is an approximation, in general, which can be made as accurate as desired. For example to find the approximate values of π, $\sqrt{2}$ etc.

When presented with a problem that cannot be solved directly, they try to replace it with a nearby problem that can be solved more easily. Examples are the use of interpolation in developing numerical integration methods and root-finding methods.

$x-3=0$

$x=3$

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

$x^3-3x^2+2x+1=0$

$x=1$ ✓

$(x-1)$

$a_4x^4+a_3x^3+a_2x^2+a_1x+a_0=0$

$2.999 \ldots$

$a_{100}x^{100}+a_{99}x^{99} - \ldots - a_1x+a_0=0$ ✓

$f(x) \approx p(x)$ ✓

Error Analysis

Floating point representation of numbers

Let x be any real no. then any real no. can be represented as infinite sequence of the digits

$$x = (0.a_1 a_2 a_3 \cdots a_n a_{n+1} \cdots)$$

$$\frac{1}{2} = 0.500000 \cdots$$

$$\frac{8}{3} = 2.66666 \cdots \cdots 66 \cdots$$

n- bit computer $(-2^{n-1}, 2^{n-1}-1)$

$(-2^{31}, 2^{31}-1)$

$$fl(x) = 0.a_1 a_2 \cdots a_n$$

$$x = \underbrace{(0.a_1 a_2 \cdots a_n a_{n+1} \cdots)}_{\text{Mantissa}} {}_{10} \times 10^{e} \rightarrow \text{exponent}$$

10 → base

$$fl(x) = (0.a_1 a_2 \cdots a_n)_{10} \times 10^{e}$$

for e.g.

$$42.965 = 4\times 10^{1} + 2\times 10^{0} + 9\times 10^{-1} + 6\times 10^{-2} + 5\times 10^{-3}$$

$$= 10^{2}\left(\frac{4}{10} + \frac{2}{10^{2}} + \frac{9}{10^{3}} + \frac{6}{10^{4}} + \frac{5}{10^{5}}\right)$$

$$= (0.42965)_{10} \times 10^{2}$$

$$-0.00234 = -\left(2\times 10^{-3} + 3\times 10^{-4} + 4\times 10^{-5}\right)$$

$$= -10^{-2}\left(0.234\right)_{10}$$

$$= -(0.234)_{10} \times 10^{-2} =$$

$$0.2666 \times 10^{1} = 0.02666 \times 10^{2}$$

not unique

representation

Error Analysis

Normal form

A non-zero floating point number is in the normal form if the value of mantiss lies in $(-1, -0.1]$ or $[0.1, 1)$

$(0.a_1 a_2 \cdots a_n)_{10} \times 10^e$

$0 \leq a_i \leq 9$, $i = 2 \cdots n$, $a_i \in \mathbb{Z}$

$a_1 \geq 1$

There are m, M s.t $-m \leq e \leq M$

Error Analysis

Overflow and Underflow

An overflow is obtained when a number is too large to fit into floating point system in use ie $e > M$

$\frac{8}{3} = 2.6666 - \mid -6$

An underflow is obtained when a no. is too small to fit into floating pt system in use ie $e < -m$

-0.00000000002

Error Analysis

Rounding and Chopping

Let x be any exact real number and $fl(x)$ be the approximation to exact no. x.

then $x = (0.a_1 a_2 \cdots a_n a_{n+1} \cdots)_{10} \times 10^e$

$$fl(x) = (0.a_1 a_2 \cdots a_n)_{10} \times 10^e$$

by chopping after n digits $fl(x) = (0.a_1 a_2 \cdots a_n)_{10} \times 10^e$

by rounding after n digits

$$fl(x) = \begin{cases} (0.a_1 a_2 \cdots a_n)_{10} \times 10^e, & 0 \le a_{n+1} < 5 \\ (0.a_1 a_2 \cdots a_n + 1)_{10} \times 10^e, & 5 \le a_{n+1} \le 9 \end{cases}$$

$$fl(x) = \begin{cases} (0.a_1 a_2 \cdots a_n)_{10} \times 10^e & 0 \le a_{n+1} < 5 \\ \left[(0.a_1 a_2 \cdots a_n)_{10} + (0.00 \cdots \underset{n\text{th place}}{1})\right] \times 10^e & ; 5 \le a_{n+1} \le 9 \end{cases}$$

Exact no. $x = \frac{6}{7} = 0.85714285714$

By chopping with 2 digits $fl(x) = 0.85$ ✓

By rounding with 2 digits $fl(x) = 0.86$ ✓

If $a_{n+1} = 5$

Case I and 5 is followed by non-zero numbers

ie $x = 0.a_1 a_2 \text{---} a_n 5 a_{n+2} a_{n+3} \text{----}$

$a_{n+2} \neq 0$

then $fl(x) = 0.a_1 a_2 \text{---} a_n + 1$

Case II 5 is followed by zero.

ie $x = 0.a_1 a_2 \text{---} a_n 5 0 \text{---}$

then
$$fl(x) = \begin{cases} 0.a_1 a_2 \text{---} a_n & \text{if } a_n \text{ is even} \\ 0.a_1 a_2 \text{---} a_n + 1 & \text{if } a_n \text{ is odd.} \end{cases}$$

Errors in the Numerical Approximation

Absolute error and Relative error

A.E. Let x be an exact no. and $fl(x)$ be the approximation to x then A.E. is $|x - fl(x)|$ ✓

R.E. then R.E. is $\dfrac{|x - fl(x)|}{|x|} = \dfrac{A.E.}{|x|}$

Error Analysis

Examples:

1. Compute the absolute error and relative error in approximations of $\sqrt{2}$ by 1.414.

Solution :-

let $x = \sqrt{2} = 1.41421356237$

$x^* = 1.414$

✓ A.E. $= |x - x^*| = 0.00021356237$

✓ R.E. $= 0.000151014$

Error Analysis

Examples:

2. Find the largest interval in which $fl(x)$ must lie to approximate π with relative error at most 10^{-5} for each value of x.

Solution

Let $x = \pi$

$fl(x) = ?$

$fl(x) = (\, , \,) = ?$

$R.E \le 10^{-5}$

$$\frac{|x - fl(x)|}{|x|} \le 10^{-5}$$

$\Rightarrow |\pi - fl(x)| \le \pi \times 10^{-5}$

$-\pi \times 10^{-5} \le \pi - fl(x) \le \pi \times 10^{-5}$

$-\pi - \pi \times 10^{-5} \le -fl(x) \le -\pi + \pi \times 10^{-5}$

$$-(-\pi - \pi \times 10^{-5}) \geq fl(x) \geq -(-\pi + \pi \times 10^{-5})$$

$$\pi - \pi \times 10^{-5} \leq fl(x) \leq \pi + \pi \times 10^{-5}$$

$$fl(x) \in [\pi - \pi \times 10^{-5}, \pi + \pi \times 10^{-5}]$$

$$[3.14156123766, 3.14162406952]$$

Error Analysis

Exercise:

1. Compute the absolute error and relative error in approximations of x by x^*, where $x = \pi$ and $x^* = 22/7$. ✓
2. Find the largest interval in which $fl(x)$ must lie to approximate $\sqrt{2}$ with relative error at most 10^{-4} for each value of x.