Numerical Analysis I - Math 4650/CSC4656

I. **MACHINE (computer) and SOLUTION CHARACTERISTICS**

A. **Precision –** the mantissa and exponent length in bits used by a particular machine; for a particular machine, it is the characteristic (mantissa size) of floating point numbers; that is, the number of bits used to represent a floating point number.

B. **Accuracy** – the number of correct digits in the (approximated) solution

C. **Components** -

1. **Machine epsilon (**or **unit roundoff error)–** the smallest floating point number which when added to 1 results in a number different than 1; that is, it is the smallest

$$\varepsilon \quad \text{such that} \quad 1+\varepsilon > 1$$

2. **Machine infinity** – the largest number a particular computer is able to represent

3. **Underflow (**or **machine zero)** – the set of non-zero numbers that are rounded to zero

II. **ERROR CLASSIFICATION**

A. **Modeling Errors** - These are errors associated with the mathematical representation of a system where by its very nature, the mathematical representation is a simplification of the actual system.

B. **Data (**or **input parameters) uncertainty** - These are measurement errors. The situation is actually more complex (see the general discussion at the end). In addition to measurement errors, we have the following.
   1. Measuremen**t** – resolution, granularity
   2. Deterministic
   3. Interval
   4. Probabilistic – probability distribution
   5. Fuzzy – membership function
   6. Possibilistic – possibilistic distribution (constructed from nested sets which means that they are monotonic)

In general, **uncertainty** can be classified as follows:
   1. Fuzzy – lack of sharp boundaries
   2. Ambiguity – one to many mapping
      a. Discord or contradiction (if you are ill, one physician says operate and another says don't operate but go to the beach)
      b. Non-specificity (the person may be innocent)

C. **Truncation Errors** - These are errors incurred by the finite representation of an infinite process.
   1. **Roundoff Errors** - These are the truncation errors associated with storing and performing arithmetic operations on finite representation of numbers.
   2. **Numerical Method Truncation Errors** - Truncation errors associated with numerical methods.
      a. **Iterative Errors** - These are errors incurred by stopping an iterative algorithm after a finite number of steps.
      b. **Function Approximation Errors** - These are errors incurred in approximating a (non-algebraic) function by a finite number of algebraic operations.

c. **Limit Errors** - These are errors associated with computing an approximation to mathematical processes defined as limits; for example, derivatives and integrals.

D. **Solution Errors** – Let x be the exact value and x^* its approximation

1. **Absolute Error** - $| x – x^*|$
2. **Relative Error** - $| x – x^*|/| x|$
3. **Theoretical Error** – errors incurred without regard to its particular implementation; that is, errors incurred using infinite precision

Let x = 2/3 and let $x^* = 0.6667$, then

1. Absolute error is: $| x – x^*| = 0.000033333333$ (on my HP-48)
2. Relative error is: $| x – x^*|/| x| = 0.000033333333/0.66666666667 = 4.99999995E-5$
3. Theoretical error

$$f'(x) = \frac{f(x+h) - f(x)}{h} - f''(c)\frac{h}{2}$$

$-f''(c)\frac{h}{2}$ is the theoretical error; that is, the exact truncation error.

Examples

Roundoff

$$\frac{1}{3} = 0.3333 \;\; 4-digit\ round \;\; = 0.3333 \;\; 4-digit\ chop$$

$$\frac{2}{3} = 0.6667 \;\; 4-digit\ round \;\; = 0.6666 \;\; 4-digit\ chop$$

$$0.3333 \times 0.6667 = 0.2222 \;\; 4-digit\ round$$

$$0.3333 \times 0.6666 = 0.222177778 = 0.2221 \;\; 4-digit\ chop$$

Iterative - If we applying the bisection method to approximate the square root of 2 by solving the equation $f(x)=x^2 - 2 = 0$, we obtain the following converging sequence of iterates:

$$x_0 = 1.5$$
$$x_1 = 1.25$$
$$x_2 = 1.375$$
M
$$x_\infty = \sqrt{2}$$
$$| x_\infty - x_2 | = 0.0392135623\text{K}$$

Then, the (absolute) iterative truncation error associated with x_2 (as an approximation of the square root of 2).

Function Approximation Errors -

$$e^x = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots$$

$$\approx 1 + x + \frac{x^2}{2!}$$

Limit Errors –

$$f'(x) = \frac{f(x+10^{-7}) - f(x)}{10^{-7}} + error$$

$$f'(x) \approx \frac{f(x+10^{-7}) - f(x)}{10^{-7}}$$

$$\left| f'(x) - \frac{f(x+10^{-7}) - f(x)}{10^{-7}} \right| = \textit{truncation error}$$

$$\int_a^b f(x)dx = \frac{b-a}{2}(f(a) + f(b)) + error$$

$$\left| \int_a^b f(x)dx - \frac{b-a}{2}(f(a) + f(b)) \right| = \textit{truncation error}$$

UNCERTAINTY

Imprecision and uncertainty can be considered as two complementary aspects of a single reality, that of <u>imperfect</u> <u>information</u>. There are three ways to look at an item of information:

1. Logical – its structure
2. Set-theoretic – its content
3. Factually – relation of items to real events

A. IMPRECISION – relates to the content of an item of information and its value. Precise items of information have values that cannot sub-divided.

1. Choice of granularity
2. Ambiguity (allied with language)
3. Generality (allied to a process of abstraction)

An item of information can be considered a four-tuple (attribute, object, value, confidence):

1. Attribute – a function that assigns a <u>value</u> to the <u>object</u>
2. Object
3. Value
4. Confidence

B. UNCERTAINTY – relates to the truth, confidence (classified above)

C. VAGUENESS/FUZZINESS - absence of clear boundaries to the set of values attached to the objects it refers to.