Approximations and Errors

Approximations and Errors

- The major advantage of numerical analysis is that a numerical answer can be obtained even when a problem has no “analytical” solution.

- Although the numerical technique yielded close estimates to the exact analytical solutions, there are errors because the numerical methods involve “approximations”.

Computer based solutions

The major steps involved to solve a given problem using a computer are:

1. Modeling: Setting up a mathematical model, i.e., formulating the problem in mathematical terms, taking into account the type of computer one wants to use.
2. Choosing an appropriate numerical method (algorithm) together with a preliminary error analysis (estimation of error, determination of steps, size etc.)
3. Programming, usually starting with a flowchart showing a block diagram of the procedures to be performed by the computer and then writing, say, a C++ program.
4. Operation or computer execution.
5. Interpretation of results, which may include decisions to rerun if further data are needed.

Accuracy and Precision

- **Accuracy** is related to the closeness to the true value.
- **Precision** is related to the closeness to other estimated values.
- **Bias** refers to systematic deviation of values from the true value.

Significant Figures

Significant figures of a number are those that can be used with confidence.

Rules for identifying sig. figures:

- ***All non-zero digits are considered significant***. For example, 91 has two significant digits (9 and 1), while 123.45 has five significant digits (1, 2, 3, 4 and 5).
- ***Zeros appearing anywhere between two non-zero digits are significant***. Example: 101.12 has five significant digits.
- ***Leading zeros are not significant***. For example, 0.00052 has two significant digits
- ***Trailing zeros are generally considered as significant.*** For example, 12.2300 has six significant digits.

Significant Figures

Scientific Notation

- If it is not clear how many, if any, of zeros are significant. This problem can be solved by using the scientific notation

  $0.0013 = 1.3*10^{-3}$ $0.00130 = 1.30*10^{-3}$

  2 sig. figures 3 sig. figures

- If a number is expressed as $2.55 *10^4$, (3 s.f), then we are only confident about the first three digits. The exact number may be 25500, 25513, 25522.6 , .. etc. So we are not sure about the last two digits nor the fractional part- If any.

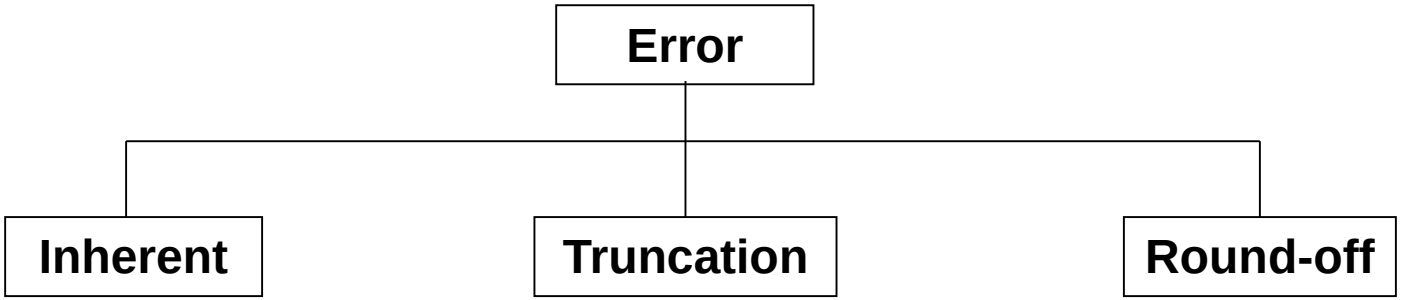
- However, if it is expressed as $2.550 * 10^4$, (4 s.f), then we are confident about the first four digits but uncertain about the last one and the fractional part – if any.

Why measure errors?

- To determine the accuracy of numerical results.
- To develop stopping criteria for iterative algorithms.

Error Definition

- Numerical errors arise from the use of approximations



1. Inherent errors or experimental errors arise due to the assumptions made in the mathematical modeling of problem. It can also arise when the data is obtained from certain physical measurements of the parameters of the problem. i.e., errors arising from measurements.

2. *Truncation errors* are those errors corresponding to the fact that a finite (or infinite) sequence of computational steps necessary to produce an exact result is “truncated” prematurely after a certain number of steps.

3. *Round of errors* are errors arising from the process of rounding off during computation. These are also called *chopping,* i.e. discarding all decimals from some decimals on.

Round-off Errors

- Numbers such as π, e, or $\sqrt{7}$ cannot be expressed by a fixed number of significant figures.
- Computers use a base-2 representation, they cannot precisely represent certain exact base-10 numbers
- Fractional quantities are typically represented in computer using "floating point" form, e.g.,

Example:

π = 3.14159265358 to be stored carrying 7 significant digits.

π = 3.141592 chopping

π = 3.141593 rounding

Truncation Errors

- Truncation errors are those that result using approximation in place of an exact mathematical procedure.

$$\frac{dv}{dt} \approx \frac{\Delta v}{\Delta t} = \frac{V\left(t_{i+1}\right) - V\left(t_i\right)}{t_{i+1} - t_i}$$

True Error

- **True error (E_t)** Defined as the difference between the true value in a calculation and the approximate value found using a numerical method etc.

True error (E_t) or Exact value of error
= true value – approximated value

- **True percent relative error ( )** is defined as the ratio between the true error, and the true value.

$$True\ percent\ relative\ error = \varepsilon_t = \frac{True\ error}{True\ value} \times 100\ (\%)$$

$$= \frac{true\ value - approximated\ value}{true\ value} \times 100\ (\%)$$

Example 3.1

Calculation of Errors

Problem Statement. Suppose that you have the task of measuring the lengths of a bridge and a rivet and come up with 9999 and 9 cm, respectively. If the true values are 10,000 and 10 cm, respectively, compute (*a*) the true error and (*b*) the true percent relative error for each case.

Solution.

(a) The error for measuring the bridge is [Eq. (3.2)]

$$E_t = 10{,}000 - 9999 = 1 \text{ cm}$$

and for the rivet it is

$$E_t = 10 - 9 = 1 \text{ cm}$$

Example 3.1

(b) The percent relative error for the bridge is [Eq. (3.3)]

$$\varepsilon_r = \frac{1}{10{,}000} 100\% = 0.01\%$$

and for the rivet it is

$$\varepsilon_r = \frac{1}{10} 100\% = 10\%$$

Thus, although both measurements have an error of 1 cm, the relative error for the rivet is much greater. We would conclude that we have done an adequate job of measuring the bridge, whereas our estimate for the rivet leaves something to be desired.

Approximate Error

- The true error is known only when we deal with functions that can be solved analytically.
- In many applications, a prior true value is rarely available.
- For this situation, an alternative is to calculate an approximation of the error using the best available estimate of the true value as:

Approximate error is defined as the difference between the present approximation and the previous approximation.

Approximate Error (E_a) = Present Approximation – Previous Approximation
and Relative Approximate Error

$$\varepsilon_a = Approximate\ percent\ relative\ error = \frac{Approximate\ error}{approximation} \times 100\ (\%)$$

Approximate Error

- In many numerical methods a present approximation is calculated using previous approximation:

$$\varepsilon_a = \frac{present\ approximation - previous\ approximation}{present\ approximation} \times 100\ (\%)$$

Note:

- The sign of ε_a or ε_t may be positive or negative
- We interested in whether the absolute value is lower than a prespecified tolerance (ε_s), not to the sign of error. Thus, the computation is repeated until (stopping criteria):

$$|\varepsilon_a| < \varepsilon_s$$

Prespecified Error

- We can relate (ε_s) to the number of significant figures in the approximation,

  So, we can assure that the result is correct to at least n significant figures if the following criteria is met:

$$\varepsilon_s = (0.5 \times 10^{2-n})\ \%$$

Example

The exponential function can be computed using Maclaurin series as follows:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

Estimate $e^{0.5}$ using series, add terms until the absolute value of approximate error ε_a fall below a pre-specified error ε_s conforming with three significant figures.

{The exact value of $e^{0.5}=1.648721\ldots$}

- Solution

$$\varepsilon_s = \left(0.5 \times 10^{2-3}\right)\% = 0.05\%$$

- ✓ Using one term: $e^{0.5} = 1$ $\varepsilon_t = \frac{1.648721 - 1.0}{1.648721} 100\% = 39.3$
- ✓ Using two terms:

$$e^{0.5} = 1 + 0.5 = 1.5 \qquad \varepsilon_t = \frac{1.648721 - 1.5}{1.648721} 100\% = 9.02\% \qquad \varepsilon_a = \frac{1.5 - 1.0}{1.5} 100\% = 33.3\%$$

- ✓ Using three terms:

$$e^{0.5} = 1 + 0.5 + \frac{0.5^2}{2!} = 1.625 \qquad \varepsilon_t = \frac{1.648721 - 1.625}{1.648721} 100\% = 1.44\% \qquad \varepsilon_a = \frac{1.625 - 1.0}{1.625} 100\% = 7.69\%$$

Terms	Results	ε_t%	ε_a%
1	1.0	39.3	---
2	1.5	9.02	33.3
3	1.625	1.44	7.69
4	1.645833333	0.175	1.27
5	1.648437500	0.0172	0.158
6	1.648697917	0.00142	0.0158