ePG Pathshala

Subject: Information Technology

Paper: Numerical Methods

**Module 1 : Introduction to Numerical Methods and Errors**

1. Introduction:

The first step in solving many problems in in Science, engineering, business, economics, biomedicines etc. is to build a mathematical model. This mathematical model is built using several principles and laws of science and engineering. Implementation of this mathematical model leads to formulation of mathematical problems to be solved. The mathematical problems arrived at could be one of the following types, but not limited to only these. Let us have a glimpse at these problems one by one. Our modules in this paper are based on studying numerical methods to solve these problems.

### 2. Overview of different problems discussed in this paper of numerical methods:

2.1. Roots of an equation f(x) = 0

One comes across the problem of finding roots of an equation $f(x) = 0$. In modules 3 to 9, we would be learning different methods for the same, namely, Bisection method, method of False Position, Secant Method, Successive Approximation method, Newton Raphson method and special techniques to find roots of a polynomial as they deserve special attention
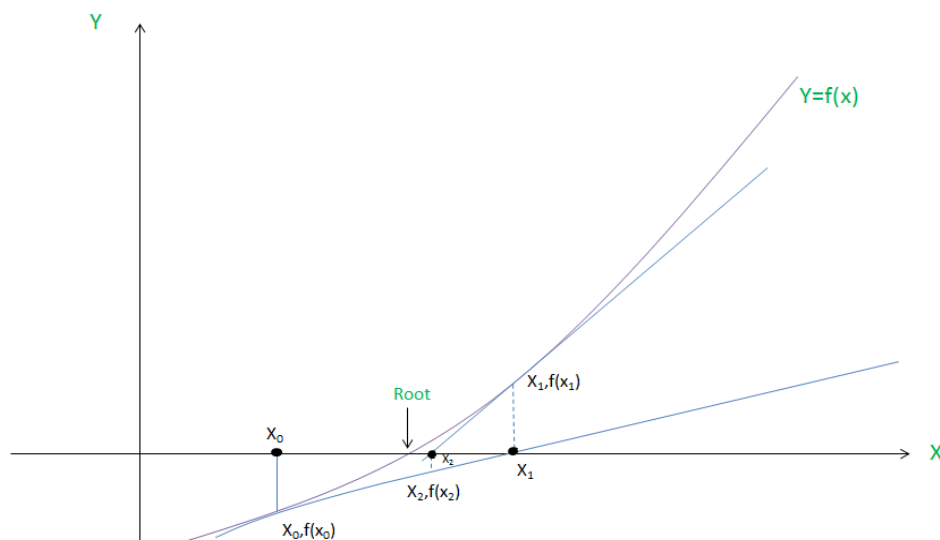
2.2. Interpolation

It turns out that due to many reasons (to be discussed in detail in module 10), function values are available at discrete points and function value is required at an argument, for which the function value is unknown. One of the ways to estimate function value is through interpolation. Interpolation is discussed in modules 10-16 and inverse interpolation in module 17.
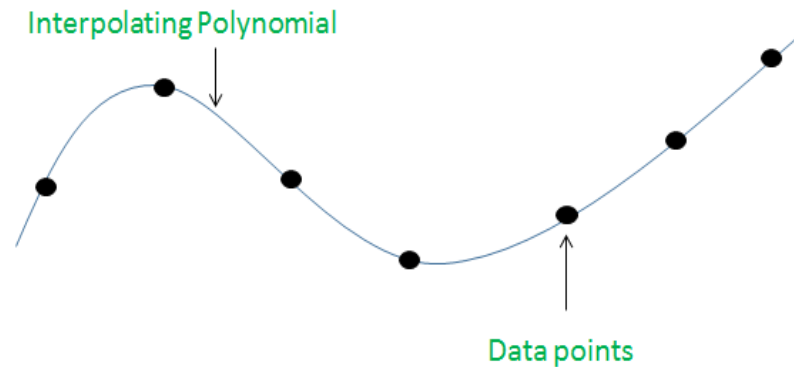

2.3. Least Square Curve fitting

Many times function (values available at discrete points) is to be modeled by a curve whose algebraic equation is known and parameters of the curve are determined by method, known as least square curve fitting. Least square curve approximation and other approximation methods are discussed in modules 18-20. The simplest least square curve fitting is when the curve being fitted is a straight line.
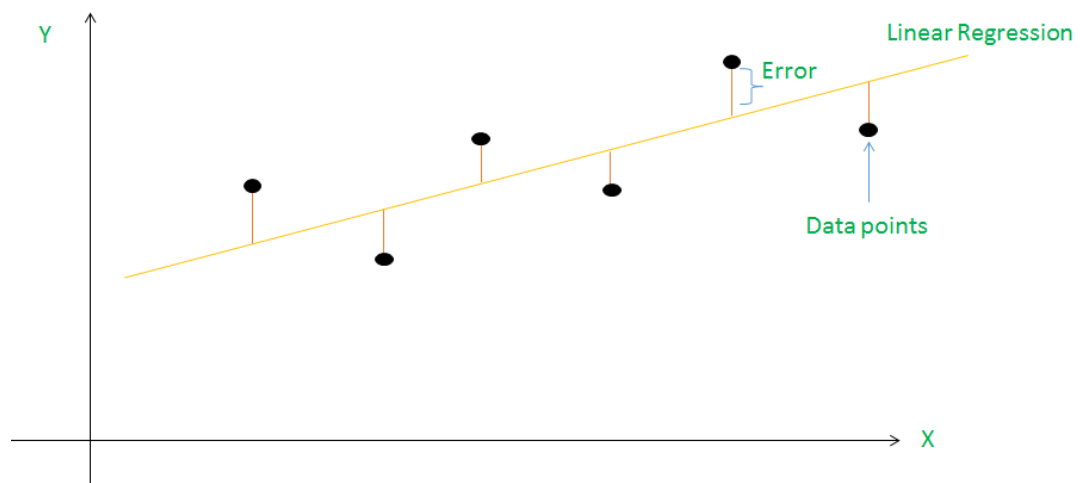
2.4. Numerical Differentiation

In several cases, one looks for the value of derivative of function at a point. The process of estimating derivative of a function at a point, be it tabular or non-tabular is called numerical differentiation. Numerical differentiation is covered in module 21.
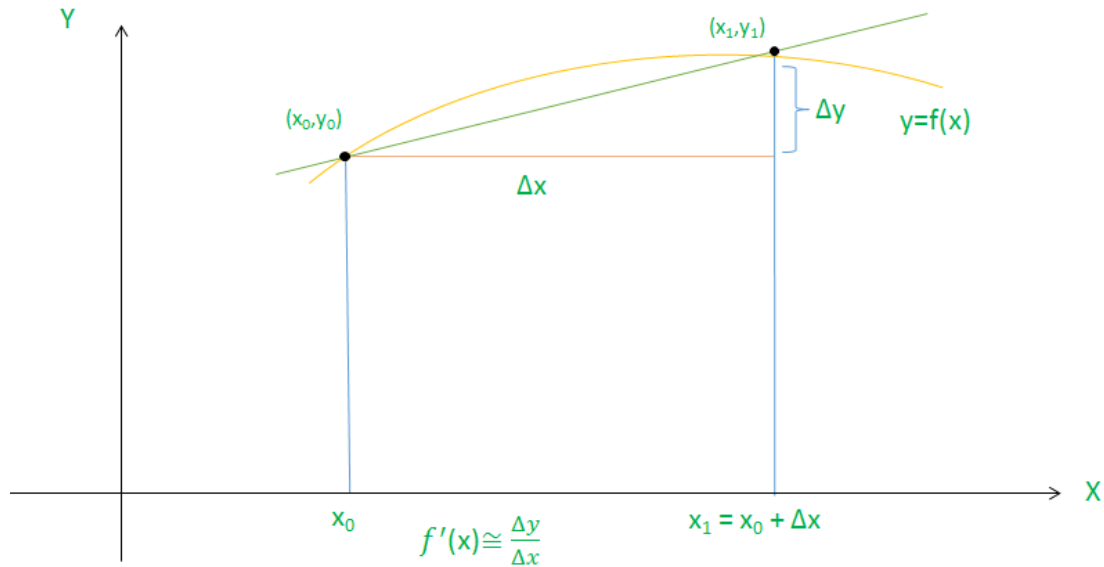

2.5. Numerical Integration

As derivative is to be estimated, similarly, integral of function is required to be estimated over an interval, with the help of function values at the discrete set of points. (Integrand expression may be available or not available) Numerical Integration methods come to our rescue. Numerical Integration methods are elaborated in modules 22-26
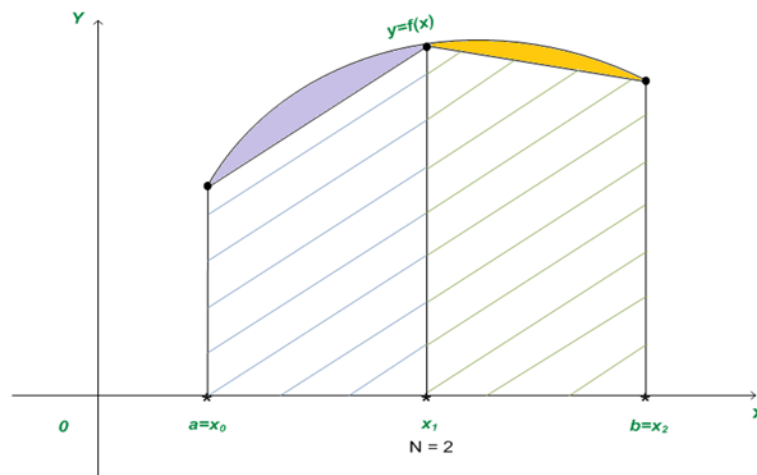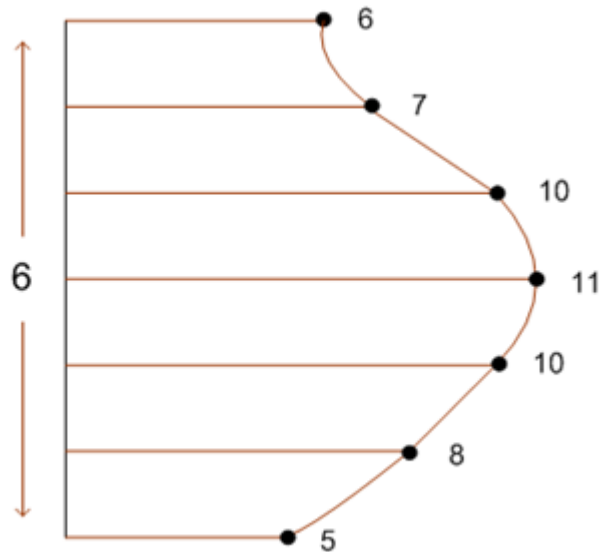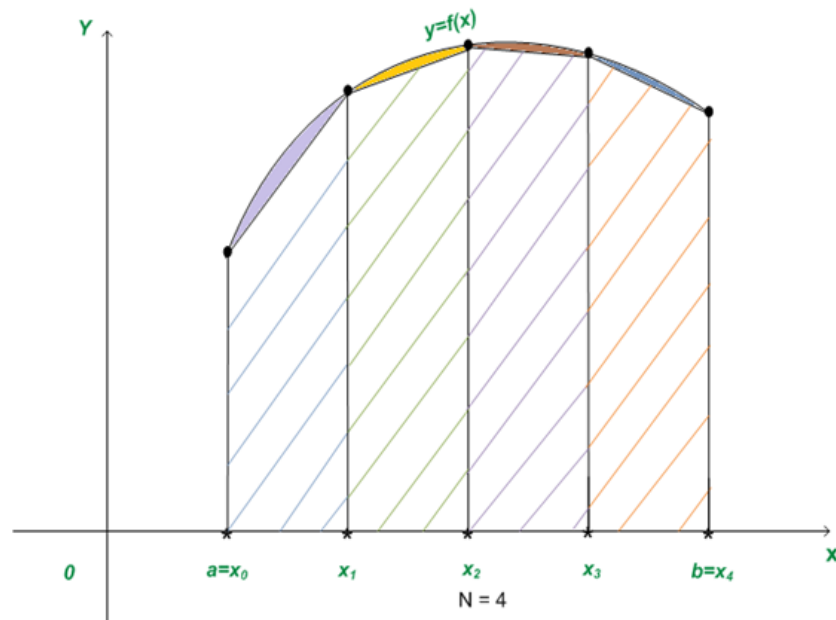
2.6. Solution of Simultaneous Equations

You are quite familiar with simultaneous linear equations and have solved them in school days using elimination techniques. In real life situations, one faces simultaneous linear as well nonlinear equations in large number of variables to be solved, which is tedious using analytical methods. Solution of simultaneous equations, determination of eigen values and eigen vectors of matrices (matrices are associated with coefficients occurring in equations) has been presented in modules 27-32
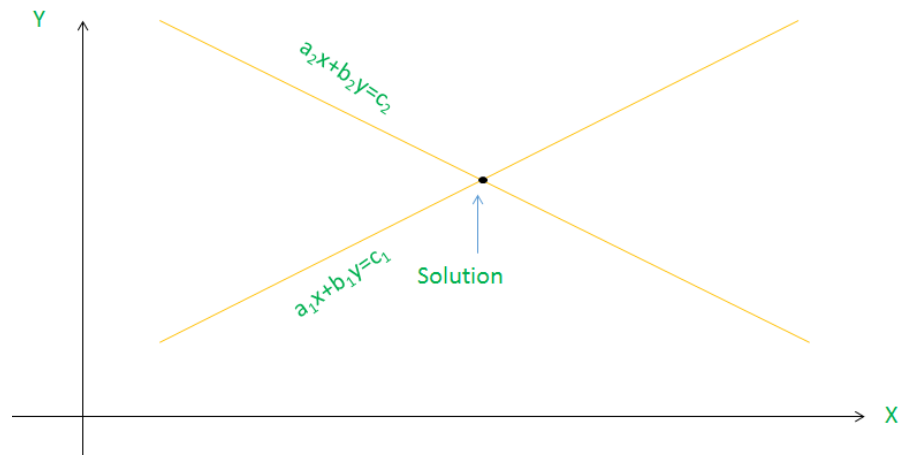
2.7. Solution of Differential Equations

Many scientific and other problems on mathematical modeling give rise to differential equations, whose solution is to be determined. Methods for finding numerical solutions are discussed in modules 33-35.
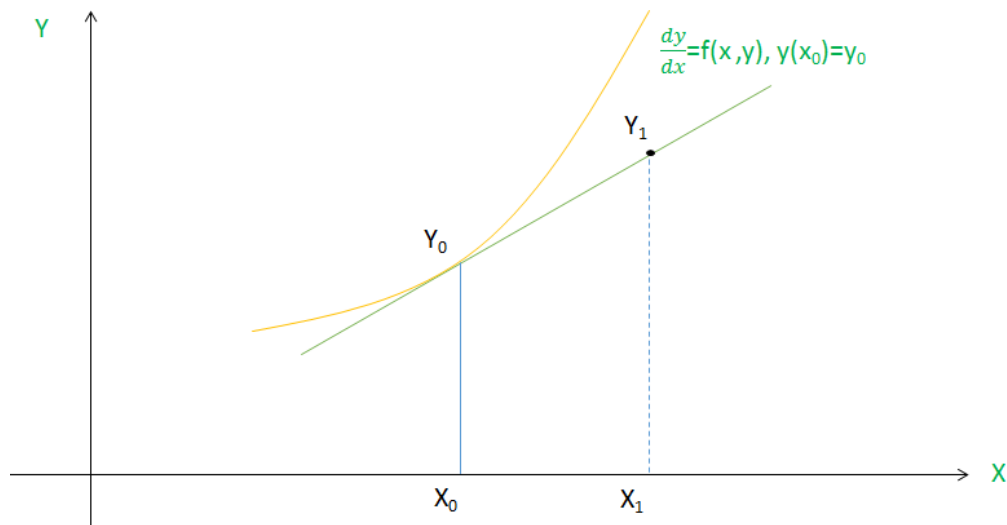
3. Why numerical methods?

Many times analytical methods to solve a given problem may not exist at all (not known so far), or are too laborious and complex to apply. In many situations, information (Data) available does not admit applicability of direct analytic methods. Like, if function is not known and only values are available at discrete arguments, analytical methods are of no use. Finally, analytic methods exist but are quite time consuming due to huge data/complex functions involved. As a result, numerical methods are of great rescue and results are obtainable to the desired accuracy in many situations.

5

4. How are numerical methods different from analytical methods?

Let us understand some characteristics of Numerical Methods. Numerical methods are applicable to numerical problems, that is, the problem to be solved by any numerical method has no scope of containing arbitrary parameters. Solution cannot be determined in terms of parameters. For example, no numerical method would be applicable for finding root of a quadratic equation $ax^2 + bx + c = 0$ as the equation contains parameters a, b, c . Nevertheless, any equation with known coefficients, for example, $2x^2 - 3x + 5 = 0$ would be solvable by an appropriate numerical method. In short, numerical answer to a numerical problem is obtained under numerical method. If the same type of problem with different data set is to be solved, the entire method is to be reapplied. Numerical methods act like algorithms and given problem acts like data set. It is likely, that there would be many methods to solve a given type of problem, all justified by mathematical analytical theory and to select the most suitable method for the occasion requires considerable skill.

Secondly, analytical methods boast of giving exact answer whereas numerical methods can only ensure approximate answer. But, in many of numerical methods, good approximate answer to desired accuracy can be obtained. Also, the computations are fast, many of the methods being iterative in nature are the best candidate to be implemented on computer. One can expect the answer to complex problem within time limit. Some methods are direct methods also. The following diagram gives basic nature of iterative processes applied in numerical methods.

Nature of Iterative Method
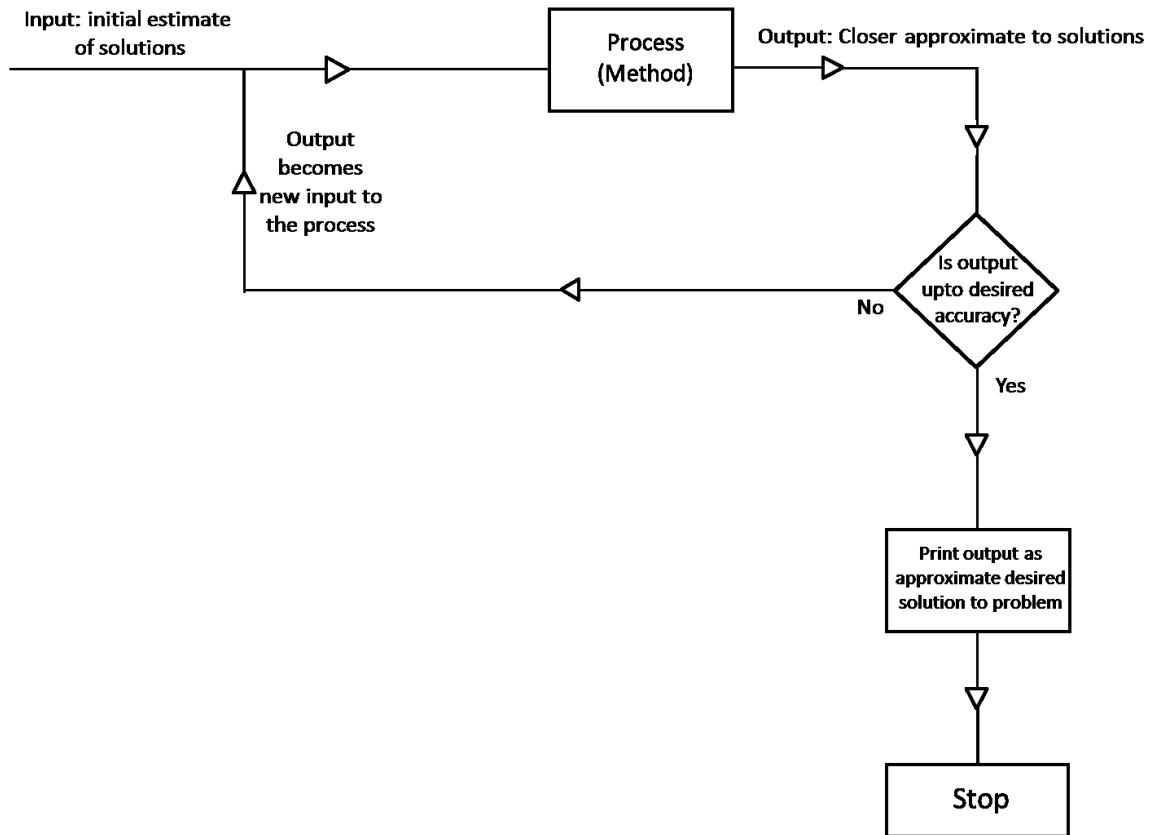
5. Quantification of Errors

Before the problem arrives to a numerical analyst, it undergoes different stages and errors may creep in. Also during computations, errors arise due to several reasons. Different sources of errors and types of error are discussed in our next module 2. Here we shall learn different measures of error. Before that, we also need to understand the difference between accuracy and precision.

Accuracy is a measure of how close the estimated answer of the problem is to true solution, the closer the answer to true value, more accurate it is. Precision is, how closely values agree to each other.
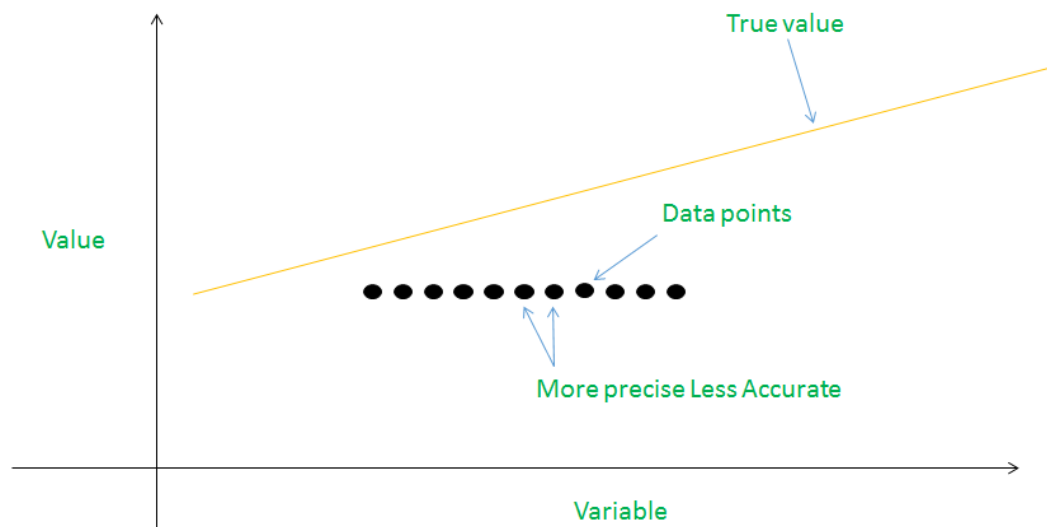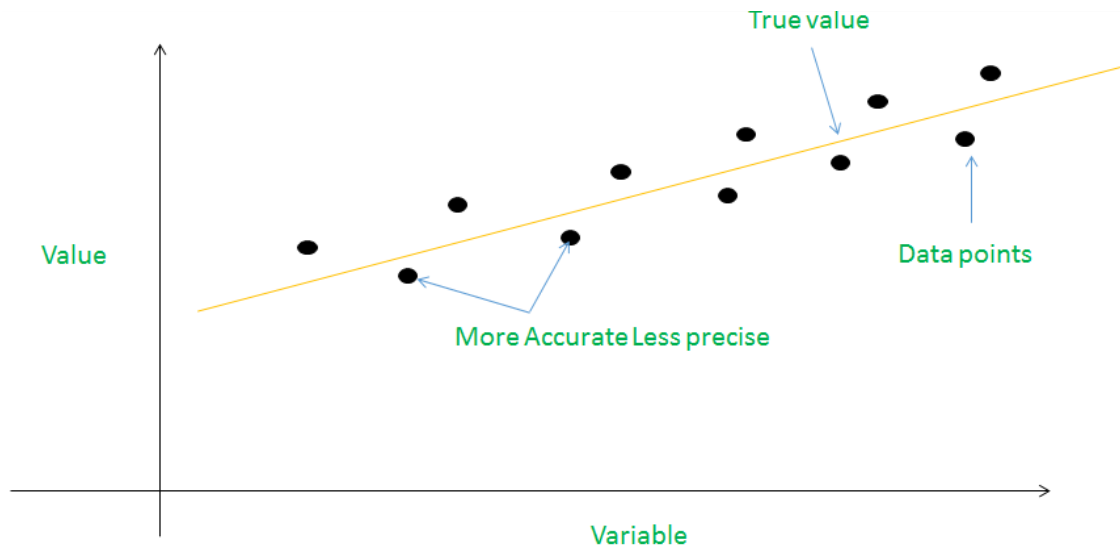
5.1. True Error

True error is defined as difference of true value (exact answer) and approximate answer. Thus, True Error: E_t = *True Value – Approximate Value*. For example, suppose, true value of a quantity = 7.893 and approximate value = 7.672 then

E_t= *True Value – Approximate Value*

$$= 7.893 - 7.672$$

$$= 0.321$$

Note that, True error can be positive or negative.

If true value = 7.893 as above, but approximate value = 7.975, then

$$E_t = 7.893 - 7.975 = -0.082$$

5.2. Absoluter (true) Error

Usually one is interested in the magnitude of error. One is interested in, "How much is the error?" So, we have, Absolute (true) error given by

$$Absolute\ (true)\ error: |E_t| = |True\ Value - Approximate\ Value|$$

5.3. Relative Error

Absolute error may not be desirable in each and every case, as it is measure of amount of error and does not take into account of order of value. It needs to be normalized. This leads to definition of Relative Error as

$$Relative\ Error\quad = \quad \frac{E_t}{True\ Value}$$

$$= \quad \frac{True\ value - Approximate\ value}{True\ Value}$$

5.4. Percentage Error

Multiplication of relative error by 100 gives % error denoted by e_t

$$e_t = \frac{True\ value - Approximate\ value}{True\ Value} * 100\%$$

And therefore, absolute relative error and absolute % error are given by

$$\left|\frac{True\ value - Approximate\ value}{True\ Value}\right| \quad and \quad |e_t| = \left|\frac{True\ Error}{True\ Value}\right| * 100\% \text{ ; respectively.}$$

Illustration 1:

Let true value be 20 meters, absolute true error = 1 cm (approximate value is 20.01 meters)

$$|\text{Relative error}| = \left|\frac{True\ value - Approximate\ value}{True\ Value}\right|$$

$$= \left|\frac{1}{2000}\right|$$

$$= 0.0005 = 0.05\%$$

Whereas, if true value = 1 cm and absolute true error = 1 cm (approximate value 2 cm) then

$$|\text{Relative error}| = \left|\frac{True\ value - Approximate\ value}{True\ Value}\right|$$

$$= \left|\frac{1}{1}\right|$$

$$= 1 = 100\%$$

5.5. Approximate errors

While solving numerical problems, (class room discussion for explaining is different) , usually true solution would be known in only limited cases, like while testing an application/software/ program's performance , program would be tested through test cases as input . If exact answer of a problem is known, why would one apply numerical method for the same? So, for all practical reasons, true answer is not known. Thus measures of errors need to be modified. Methods being iterative in nature, what can be better than current estimate as a substitute for true value and previous estimate as approximate value. As a result, Absolute approximate error is given by

$$|\text{Approximate error}| = |\text{Current estimate} - \text{Previous estimate}|; \text{ and}$$

$$\text{Approximate relative error} = \left| \frac{Approximate\ error}{Current\ Estimate} \right|$$

$$\text{Approximate \% relative error} = \left| \frac{Approximate\ error}{Current\ Estimate} \right| * 100\%, \quad \text{denoted as } |e_a|$$

Illustration 2:

Following is the table of iterations being performed for finding roots of an equation $f(x) = 0$ by Bisection method (Module 3). The column containing c_k's are successive estimates for the root. Let us calculate approximate % error in some iterations.

Iteration No.	a	f(a)	b	f(b)	C_k	$f(c_k)$
0	0	-1	1	1	0.5	-0.375
1	0.5	-0.375	1	0.17188	0.75	0.1719
2	0.5	-0.375	0.75	0.17188	0.625	-0.1309
3	0.625	-0.1309	0.75	0.01245	0.6875	0.0125
4	0.625	-0.1309	0.6875	0.01245	0.65625	-0.0611
5	0.65625	-0.0611	0.6875	0.01245	0.67188	-0.0248
6	0.67188	-0.0248	0.6875	0.01245	0.67969	-0.0063
7	0.67969	-0.0063	0.6875	0.01245	0.6836	0.0031

Iteration No.	a	f(a)	b	f(b)	C_k	$f(c_k)$
8	0.67969	-0.0063	0.6836	0.00305	0.68165	-0.0016
9	0.68165	-0.0016	0.6836	0.00305	0.68263	0.0007
10	0.68165	-0.0016	0.68263	0.00072	0.68214	-0.0005
11	0.68214	-0.0005	0.68263	0.00072	0.68239	0.0001
12	0.68214	-0.0005	0.68239	0.00015	0.68227	-0.0001
13	0.68227	-0.0001	0.68239	0.00015	0.68233	0

Iteration - I

$$e_a = \frac{0.75 - 0.5}{0.75} * 100 = 33\%$$

Iteration - II

$$= \frac{0.625 - 0.75}{0.625} * 100 = 20\%$$

Iteration – last but one

$$= \frac{0.68227 - 0.68239}{0.68227} * 100 = 0.01758\%$$

Iteration – last

$$= \frac{0.68233 - 0.68227}{0.68233} * 100 = 0.00879\%$$

5.6. Stopping Criterion

There are many stopping criterions (achieved the desired accuracy, satisfied with the answer obtained). Two of them are:

   (a) Absolute approximate error < Pre assigned tolerance say ϵ , where $\epsilon > 0$
   (b) Absolute Relative Error < ϵ

If one is interested in obtaining answer correct to certain number of significant digits say m, then

$$|e_a| \leq \frac{1}{2} * 10^{2-m}$$

****