1

APPROXIMATIONS AND ERRORS IN COMPUTATION

1. Introduction
2. Accuracy of numbers
3. Errors
4. Useful rules for estimating errors
5. Error propagation
6. Error in the approximation of a function
7. Error in a series approximation
8. Order of approximation
9. Growth of error
10. Objective Type of Questions

1.1. INTRODUCTION

The limitations of analytical methods in practical applications have led scientists and engineers to evolve numerical methods. We know that exact methods often fail in drawing plausible inferences from a given set of tabulated data or in finding roots of transcendental equations or in solving non-linear differential equations. There are many more such situations where analytical methods are unable to produce desirable results. Even if analytical solutions are available, these are not amenable to direct numerical interpretation. *The aim of numerical analysis is therefore, to provide constructive methods for obtaining answers to such problems in a numerical form.*

With the advent of high speed computers and increasing demand for numerical solution to various problems, numerical techniques have become indispensible tools in the hands of engineers and scientists.

The input information is rarely exact since it comes from some measurement or the other and the method also introduces further error. As such, the error in the final result may be due to error in the initial data or in the method or both. Our effort will be to minimize these errors, so as to get best possible results. We therefore begin by explaining various kinds of approximations and errors which may occur in a problem and derive some results on error propagation in numerical calculations.

1.2. ACCURACY OF NUMBERS

(1) *Approximate numbers.* There are two types of numbers *exact* and *approximate*. *Exact numbers* are 2, 4, 9, 13, 7/2, 6.45, ... etc. But there are numbers such as 4/3 (= 1.33333 ...), $\sqrt{2}$ (= 1.414213 ...) and π (= 3.141592 ...) which cannot be expressed by a finite number of digits. These may be approximated by numbers 1.3333, 1.4142 and 3.1416 respectively. Such numbers which represent the given numbers to a certain degree of accuracy are called *approximate numbers*.

(2) *Significant figures.* The digits used to express a number are called *significant digits (figures)*. Thus each of the numbers 7845, 3.589, 0.4758 contains four significant figures while the numbers 0.00386, 0.000587 and 0.0000296 contain only three significant figures since zeros only help to fix the position of the decimal point. Similarly the numbers 45000 and 7300.00 have two significant figures only.

(3) *Rounding off.* There are numbers with large number of digits *e.g.*, 22/7 = 3.142857143. In practice, it is desirable to limit such numbers to a manageable number of digits such as 3.14 or 3.143. This process of dropping unwanted digits is called *rounding off*.

(4) Rule *to round off a number to n significant figures* :

(*i*) *Discard all digits to the right of the nth digit*.

(*ii*) *If this discarded number is*

(*a*) *less than half a unit in the nth place, leave the nth digit unchanged* ;

(*b*) *greater than half a unit in the nth place, increase the nth digit by unity* ;

(*c*) *exactly half a unit in the nth place, increase the nth digit by unity if it is odd otherwise leave it unchanged*.

For instance, the following numbers rounded off to three significant figures are :

7.893	to	7.89	3.567	to	3.57
12.865	to	12.9	84767	to	84800
6.4356	to	6.44	5.8254	to	5.82

Also the numbers 6.284359, 9.864651, 12.464762 rounded off to four places of decimal at 6.2844, 9.8646, 12.4648 respectively.

Obs. The numbers thus rounded off to n significant figures (or n decimal places) are said to be *correct to n significant figures (or n decimal places).*

1.3. ERRORS

In any numerical computation, we come across the following types of errors :

(1) *Inherent errors.* Errors which are already present in the statement of a problem before its solution, are called *inherent errors*. Such errors arise either due to the given data being approximate or due to the limitations of mathematical tables, calculators or the digital computer. Inherent errors can be minimized by taking better data or by using high precision computing aids.

(2) *Rounding errors* arise from the process of rounding off the numbers during the computation. Such errors are unavoidable in most of the calculations due to the limitations of the computing aids. Rounding errors can, however, be reduced :

(*i*) by changing the calculation procedure so as to avoid subtraction of nearly equal numbers or division by a small number ;

or (*ii*) by retaining at least one more significant figure at each step than that given in the data and rounding off at the last step.

(3) *Truncation errors* are caused by using approximate results or on replacing an infinite process by a finite one. If we are using a decimal computer having a fixed word length of 4 digits, rounding off of 13.658 gives 13.66 whereas truncation gives 13.65.

For example, if $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots \infty = X$ (say)

is replaced by $1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} = X'$ (say), then the truncation error is $X - X'$.

Truncation error is a type of algorithm error.

(4) *Absolute, Relative and Percentage errors.* If X is the true value of a quantity and X' is its approximate value, then $| X - X' |$ *i.e.* | Error | is called the *absolute error* E_a.

The *relative error* is defined by $E_r = \left| \frac{X - X'}{X} \right|$ *i.e.* $\frac{|\text{Error}|}{|\text{True value}|}$

and the *percentage error* is $E_p = 100\, E_r = 100 \left| \frac{X - X'}{X} \right|$.

If $\overline{X}$ be such a number that $| X - X' | \leq \overline{X}$, then $\overline{X}$ is an upper limit on the magnitude of absolute error and measures the *absolute accuracy*.

Obs. 1. The relative and percentage errors are independent of the units used while absolute error is expressed in terms of these units.

Obs. 2. *If a number is correct to n decimal places then the error* $= \frac{1}{2} 10^{-n}$.

For example, if the number is 3.1416 correct to 4 decimal places, then the error

$= \frac{1}{2} 10^{-4} = 0.00005$.

1.4. USEFUL RULES FOR ESTIMATING ERRORS

To estimate the errors which creep in when the numbers in a calculation are truncated or rounded off to a certain number of digits, the following rules are useful.

If the approximate value of a number X having n decimal digits is X', then

(1) *Absolute error due to truncation to k digits*

$$= | X - X' | < 10^{n-k}$$

(2) *Absolute error due to rounding off to k digits*

$$= | X - X' | < \frac{1}{2}\, 10^{n-k}$$

(3) *Relative error due to truncation to k digits*

$$= \left| \frac{X - X'}{X} \right| < 10^{1-k}$$

(4) *Relative error due to rounding off to k digits*

$$= \left| \frac{X - X'}{X} \right| < \frac{1}{2}\, 10^{1-k}$$

Obs. 1. *If a number is correct to n significant digits, then the maximum relative error* $\leq \frac{1}{2} 10^{-n}$.

If a number is correct to d decimal places, then the absolute error $\leq \frac{1}{2} 10^{-d}$.

Obs. 2. *If the first significant figure of a number is k and the number is correct to n significant figures, then the relative error* $< 1/(k \times 10^{n-1})$.

Let us verify this result by finding the relative error in the number 864.32 correct to five significant figures.

Here $k = 8$, $n = 5$ and

absolute error $\nmid 0.01 \times \frac{1}{2} = 0.005$.

$$\therefore \quad \text{Relative error} \leq \frac{0.005}{864.32} = \frac{5}{864320} = \frac{1}{2 \times 86432} < \frac{1}{2 \times 80000} = \frac{1}{2 \times 8 \times 10^4}$$

$$< \frac{1}{8 \times 10^4} \quad i.e. \quad \frac{1}{k \times 10^{n-1}}. \qquad \text{Hence the result is verified.}$$

■ **Example 1.1.** *Round off the numbers 865250 and 37.46235 to four significant figures and compute E_a, E_r, E_p in each case.*

Sol. (*i*) Number rounded off to four significant figures = 865200

$$\therefore \quad E_a = | X - X_1 | = | 865250 - 865200 | = 50$$

$$E_r = \left| \frac{X - X_1}{X} \right| = \frac{50}{865250} = 6.71 \times 10^{-5}$$

$$E_p = E_r \times 100 = 6.71 \times 10^{-3}$$

(*ii*) Number rounded off to four significant figures = 37.46

$$\therefore \quad E_a = | X - X_1 | = | 37.46235 - 37.46000 | = 0.00235$$

$$E_r = \left| \frac{X - X_1}{X} \right| = \frac{0.00235}{37.46235} = 6.27 \times 10^{-5}$$

$$E_p = E_r \times 100 = 6.27 \times 10^{-3}.$$

■ **Example 1.2.** *Find the absolute error if the number $X = 0.00545828$ is*

(i) truncated to three decimal digits.

(ii) rounded off to three decimal digits.

Sol. We have $X = 0.00545828 = 0.545828 \times 10^{-2}$

(*i*) After truncating to three decimal places, its approximate value $X' = 0.545 \times 10^{-2}$

$$\therefore \quad \text{Absolute error} = | X - X' | = 0.000828 \times 10^{-2}$$

$$= 0.828 \times 10^{-5} \quad < 10^{-2-3}$$

This proves rule (1).

(*ii*) After rounding off to three decimal places, its approximate value $X' = 0.546 \times 10^{-2}$

$$\therefore \quad \text{Absolute error} = | X - X' |$$

$$= | 0.545828 - 0.546 | \times 10^{-2}$$

$$= 0.000172 \times 10^{-2} = 0.172 \times 10^{-5}$$

which is $< 0.5 \times 10^{-2-3}$. This proves rule (2).

■ **Example 1.3.** *Find the relative error if the number $X = 0.004997$ is*

(i) truncated to three decimal digits

(ii) rounded off to three decimal digits.

Sol. We have $X = 0.004997 = 0.4997 \times 10^{-2}$

(*i*) After truncating to three decimal places, its approximate value $X' = 0.499 \times 10^{-2}$.

$$\therefore \quad \text{Relative error} = \left| \frac{X - X'}{X} \right| = \left| \frac{0.4997 \times 10^{-2} - 0.499 \times 10^{-2}}{0.4997 \times 10^{-2}} \right|$$

$$= 0.140 \times 10^{-2} < 10^{1-3}$$

This proves rule (3).

(*ii*) After rounding off to three decimal places, the approximate value of the given number

$$X' = 0.500 \times 10^{-2}$$

$$\therefore \quad \text{Relative error} = \left| \frac{X - X'}{X} \right| = \left| \frac{0.4997 \times 10^{-2} - 0.500 \times 10^{-2}}{0.4997 \times 10^{-2}} \right|$$

$$= 0.600 \times 10^{-3} = 0.06 \times 10^{-3+1}$$

which is less than $0.5 \times 10^{-3+1}$. This proves rule (4).

PROBLEMS 1.1

1. Round off the following numbers correct to four significant figures : 3.26425, 35.46735, 4985561, 0.70035, 0.00032217, 18.265101.

2. Round off the number 75462 to four significant digits and then calculate the absolute error and percentage error. (*U.P.T.U., B. Tech., 2004*)

3. If 0.333 is the approximate value of 1/3, find the absolute and relative errors. (*Bhopal, B.E., 2007*)

4. Find the percentage error if 625.483 is approximated to three significant figures.

5. Find the relative error in taking $\pi = 3.141593$ as 22/7. (*V.T.U. MCA, 2007*)

6. The height of an observation tower was estimated to be 47 m, whereas its actual height was 45 m. Calculate the percentage relative error in the measurement.

7. Suppose that you have a 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 the percentage relative error in each case. (*Pune, B. Tech., 2004*)

8. Find the value of e^x using series expansion $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$ for $x = 0.5$ with an absolute error less than 0.005.

9. $\sqrt{29} = 5.385$ and $\sqrt{\pi} = 3.317$ correct to 4 significant figures. Find the relative errors in their sum and difference.

10. Given : $a = 9.00 \pm 0.05$, $b = 0.0356 \pm 0.0002$, $c = 15300 \pm 100$, $d = 62000 \pm 500$. Find the maximum value of absolute error in $a + b + c + d$.

11. Two numbers are 3.5 and 47.279 both of which are correct to the significant figures given. Find their product.

12. Find the absolute error and the relative error in the product of 432.8 and 0.12584 using four digit mantissa. (*Kerala B. Tech., 2003*)

13. The discharge Q over a notch for head H is calculated by the formula $Q = kH^{5/2}$ where k is a given constant. If the head is 75 cm and an error of 0.15 cm is possible in its measurement, estimate the percentage error in computing the discharge.

14. If the number p is correct to 3 significant digits, what will be the maximum relative error ?

1.5. ERROR PROPAGATION

A number of computational steps are carried out for the solution of a problem. It is necessary to understand the way the error propagates with progressive computation.

If the approximate values of two numbers X and Y be X' and Y' respectively, then the absolute error

$$E_{ax} = X - X' \quad \text{and} \quad E_{ay} = Y - Y'$$

(1) *Absolute error in addition operation*

$$X + Y = (X' + E_{ax}) + (Y' + E_{ay})$$
$$= X' + Y' + E_{ax} + E_{ay}$$
$$\therefore \quad | (X + Y) - (X' + Y') | = | E_{ax} + E_{ay} | \leq | E_{ax} | + | E_{ay} |$$

Thus the absolute error in taking $(X' + Y')$ as an approximation to $(X + Y)$ is less than or equal to the sum of the absolute errors in taking X' as an approximation to X and Y' as an approximation to Y.

(2) *Absolute error in subtraction operation*

$$X - Y = (X' + E_{ax}) - (Y' + E_{ay})$$
$$= (X' - Y') + (E_{ax} - E_{ay})$$
$$\therefore \quad | (X - Y) - (X' - Y') | = | E_{ax} - E_{ay} | \leq | E_{ax} | + | E_{xy} |$$

Thus the absolute error in taking $(X' - Y')$ as an approximation to $(X - Y)$ is less than or equal to the sum of the absolute errors in taking X' as an approximation to X and Y' as an approximation to Y.

(3) *Absolute error in multiplication operation*

To find the absolute error E_a in the product of two numbers X and Y, we write

$$E_a = (X + E_{ax})(Y + E_{ay}) - XY$$

where E_{ax} and E_{ay} are the absolute errors in X and Y respectively. Then

$$E_a = XE_{ay} + YE_{ax} + E_{ax}E_{ay}$$

Assuming E_{ax} and E_{ay} are reasonably small so that $E_{ax}\, E_{xy}$ can be ignored.

Thus $E_a = XE_{ay} + YE_{ax}$ approximately.

(4) *Absolute error in division operation*

Similarly the absolute error E_a in the quotient of two numbers X and Y is given by

$$E_a = \frac{X + E_{ax}}{Y + E_{ay}} - \frac{X}{Y} = \frac{Y\,E_{ax} - X\,E_{ay}}{Y(Y + E_{ay})}$$

$$= \frac{YE_{ax} - XE_{ay}}{Y^2\,(1 + E_{ay}/Y)}$$

$$= \frac{YE_{ax} - XE_{ay}}{Y^2}, \text{ assuming } E_{ay}/Y \text{ to be small.}$$

$$= \frac{X}{Y}\left(\frac{E_{ax}}{X} - \frac{E_{ay}}{Y}\right)$$

■ **Example 1.4.** *Find the absolute error and relative error in* $\sqrt{6} + \sqrt{7} + \sqrt{8}$ *correct to 4 significant digits.*

Sol. We have $\sqrt{6} = 2.449$, $\sqrt{7} = 2.646$, $\sqrt{8} = 2.828$

$\therefore \quad S = \sqrt{6} + \sqrt{7} + \sqrt{8} = 7.923.$

Then the absolute error E_a in S, is

$$E_a = 0.0005 + 0.0007 + 0.0004 = 0.0016$$

This shows that S is correct to 3 significant digits only. Therefore, we take $S = 7.92$

Then the relative error E_r is

$$E_r = \frac{0.0016}{7.92} = 0.0002.$$

■ **Example 1.5.** *The area of cross-section of a rod is desired upto 0.2% error. How accurately should the diameter be measured?* *(Pune, B. Tech., 2003)*

Sol. If A is the area and D is the diameter of the rod, then $A = \pi\left(\frac{D}{2}\right)^2 = \frac{\pi}{4} D \cdot D$.

Now error in area A is 0.2% *i.e.,* 0.002 which is due to the error in the product $D \times D$.

We know that if E_a is the absolute error in the product of two numbers X and Y, then

$$E_a = X_{aY}E + YE_{aX}$$

Here $X = Y = D$ and $E_{aX} = E_{aY} = E_D$, therefore

$$E_a = DE_D + DE_D \quad \text{or} \quad 0.002 = 2DE_D$$

Thus $E_D = 0.001/D$ *i.e.,* the error in the diameter should not exceed $0.001\ D^{-1}$.

■ **Example 1.6.** *Find the product of the numbers 3.7 and 52.378 both of which are correct to given significant digits.*

Sol. Since the absolute error is greatest in 3.7, therefore we round off the other number to 3 significant figures *i.e.* 52.4.

$\therefore$ Their product $P = 3.7 \times 52.4 = 193.88 = 1.9388 \times 10^2$

Since the first number contains only two significant figures, therefore retaining only two significant figures in the product, we get

$$P = 1.9 \times 10^2$$

1.6. ERROR IN THE APPROXIMATION OF A FUNCTION

Let $y = f(x_1, x_2)$ be a function of two variables x_1, x_2. If $\delta x_1, \delta x_2$ be the errors in x_1, x_2, then the error δy in y is given by

$$y + \delta y = f(x_1 + \delta x_1, x_2 + \delta x_2)$$

Expanding the right hand side by Taylor's series, we get

$$y + \delta y = f(x_1, x_2) + \left(\frac{\partial f}{\partial x_1}\delta x_1 + \frac{\partial f}{\partial x_2}\delta x_2\right)$$

$$+ \text{terms involving higher powers of } \delta x_1 \text{ and } \delta x_2 \qquad \ldots(i)$$

If the errors $\delta x_1, \delta x_2$ be so small that their squares and higher powers can be neglected, then (i) gives

$$\delta y = \frac{\partial f}{\partial x_1}\delta x_1 + \frac{\partial f}{\partial x_2}\delta x_2 \text{ approximately.}$$

Hence $$\delta y \approx \frac{\partial y}{\partial x_1}\,\delta x_1 + \frac{\partial y}{\partial x_2}\,\delta x_2$$

In general, the error δy in the function $y = f(x_1, x_2, \ldots\ldots, x_n)$ corresponding to the errors δx_i in x_i $(i = 1, 2, \ldots\ldots n)$ is given by

$$\delta y \approx \frac{\partial y}{\partial x_1}\,\delta x_1 + \frac{\partial y}{\partial x_2}\,\delta x_2 + \ldots\ldots + \frac{\partial y}{\partial x_n}\,\delta x_n$$

and the relative error in y is $$E_r = \frac{\delta y}{y} \approx \frac{\partial y}{\partial x_1}\frac{\delta x_1}{y} + \frac{\partial y}{\partial x_2}\frac{\delta x_2}{y} + \ldots\ldots + \frac{\partial y}{\partial x_n}\frac{\delta x_n}{y}\,.$$

Example 1.7. *If $u = 4x^2y^3/z^4$ and errors in x, y, z be 0.001, compute the relative maximum error in u when $x = y = z = 1$.*

Sol. Since $$\frac{\partial u}{\partial x} = \frac{8xy^3}{z^4}, \frac{\partial u}{\partial y} = \frac{12x^2y^2}{z^4}, \frac{\partial u}{\partial z} = -\frac{16x^2y^3}{z^5}$$

$$\therefore \qquad \delta u = \frac{\partial u}{\partial x}\,\delta x + \frac{\partial u}{\partial y}\,\delta y + \frac{\partial u}{\partial z}\,\delta z = \frac{8xy^3}{z^4}\,\delta x + \frac{12x^2y^2}{z^4}\,\delta y - \frac{16x^2y^3}{z^5}\,\delta z$$

Since the errors δx, δy, δz may be positive or negative, we take the absolute values of the terms on the right side, giving

$$(\delta u)_{max} \approx \left|\frac{8xy^3}{z^4}\,\delta x\right| + \left|\frac{12x^2y^2}{z^4}\,\delta y\right| + \left|\frac{16x^2y^3}{z^5}\,\delta z\right|$$

$$= 8(0.001) + 12(0.001) + 16(0.001) = 0.036$$

Hence the maximum relative error $= (\delta u)_{max}/u = 0.036/4 = 0.009$.

Example 1.8. *Find the relative error in the function $y = ax_1^{m_1}\, x_2^{m_2} \ldots\ldots x_n^{m_n}$.*

Sol. We have $\log y = \log a + m_1 \log x_1 + m_2 \log x_2 + \ldots + m_n \log x_n$

$$\therefore \qquad \frac{1}{y}\frac{\partial y}{\partial x_1} = \frac{m_1}{x_1}, \frac{1}{y}\frac{\partial y}{\partial x_2} = \frac{m_2}{x_2} \text{ etc.}$$

Hence $$E_r \approx \frac{\partial y}{\partial x_1}\frac{\delta x_1}{y} + \frac{\partial y}{\partial x_2}\frac{\delta x_2}{y} + \ldots. + \frac{\partial y}{\partial x_n}\frac{\delta x_n}{y} = m_1\frac{\delta x_1}{x_1} + m_2\frac{\delta x_2}{x_2} + \ldots + m_n\frac{\delta x_n}{x_n}$$

Since the errors δx_1, δx_2, ..., δx_n may be positive or negative, we take the absolute values of the terms on the right side. This gives :

$$(E_r)_{max} \le m_1\left|\frac{\delta x_1}{x_1}\right| + m_2\left|\frac{\delta x_2}{x_2}\right| + \ldots + m_n\left|\frac{\delta x_n}{x_n}\right|$$

Cor. Taking $a = 1$, $m_1 = m_2 = \ldots\ldots = m_n = 1$, we have

$$y = x_1x_2 \ldots\ldots x_n.$$

then $$E_r \approx \frac{\delta x_1}{x_1} + \frac{\delta x_2}{x_2} + \ldots\ldots + \frac{\delta x_n}{x_n}$$

Thus *the relative error of a product of n numbers is approximately equal to the algebraic sum of their relative errors.*

1.7. ERROR IN A SERIES APPROXIMATION

We know that the Taylor's series for $f(x)$ at $x = a$ with a remainder after n terms is

$$f(x) = f(a + \overline{x - a}) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a) + \ldots + \frac{(x-a)^{n-1}}{(n-1)!} f^{n-1}(a) + R_n(x)$$

where $R_n(x) = \dfrac{(x-a)^n}{n!} f^n(\theta)$, $a < \theta < x$.

If the series is convergent, $R_n(x) \to 0$ as $n \to \infty$ and hence if $f(x)$ is approximated by the first n terms of this series, then the maximum error will be given by the remainder term $R_n(x)$. On the other hand, if the accuracy required in a series approximation is preassigned, then we can find n, the number of terms which would yield the desired accuracy.

■ **Example 1.9.** *Find the number of terms of the exponential series such that their sum gives the value of e^x correct to six decimal places at $x = 1$.*

Sol. We have $$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\ldots + \frac{x^{n-1}}{(n-1)!} + R_n(x) \qquad \ldots(i)$$

where $R_n(x) = \dfrac{x^n}{n!} e^\theta$, $0 < \theta < x$.

$\therefore$ Maximum absolute error (at $\theta = x$) $= \dfrac{x^n}{n!} e^x$ and the maximum relative error $= \dfrac{x^n}{n!}$

Hence $(E_r)_{max}$ at $x = 1$ is $\dfrac{1}{n!}$.

For a six decimal accuracy at $x = 1$, we have

$$\frac{1}{n!} < \frac{1}{2} 10^{-6}, \quad i.e. \quad n! > 2 \times 10^6 \qquad \text{which gives } n = 10.$$

Thus we need 10 terms of the series (i) in order that its sum is correct to 6 decimal places.

■ **Example 1.10.** *The function $f(x) = \tan^{-1} x$ can be expanded as*

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots\ldots + (-1)^{n-1} \frac{x^{2n-1}}{2n-1} + \ldots\ldots,$$

find n such that the series determine $\tan^{-1} x$ correct to eight significant digits at $x = 1$.

(U.P.T.U., B. Tech. 2007)

Sol. If we retain n terms in the expansion of $\tan^{-1} x$, then $(n+1)$th term

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

$$= \frac{(-1)^n}{2n+1} \quad \text{for } x = 1.$$

To determine $\tan^{-1}(1)$ correct to eight significant digits accuracy $\left|\frac{(-1)^n}{2n+1}\right| < \frac{1}{2}\times 10^{-8}$

i.e., $2n + 1 > 2 \times 10^8$ or $n > 10^8 - \frac{1}{2}$

Hence $n = 10^8 + 1$.

1.8. ORDER OF APPROXIMATION

We often replace a function $f(h)$ with its approximation $\phi(h)$ and the error bound is known to be $\mu(h^n)$, n being a positive integer so that

$$| f(h) - \phi(h) | \leq \mu | h^n | \qquad \text{for sufficiently small } h.$$

Then we say that $\phi(h)$ *approximates f(h) with order of approximation* $O(h^n)$ and write $f(h) = \phi(h) + O(h^n)$.

For instance, $\frac{1}{1-h} = 1 + h + h^2 + h^3 + h^4 + h^5 + ...$

is written as $\frac{1}{1-h} = 1 + h + h^2 + h^3 + O(h^4)$ ...(*i*)

to the 4th order of approximation.

Similarly $\cos(h) = 1 - \frac{h^2}{2!} + \frac{h^4}{4!} - \frac{h^6}{6!} + \frac{h^8}{8!} - .....$

to the 6th order of approximation becomes

$$\cos(h) = 1 - \frac{h^2}{2!} + \frac{h^4}{4!} + O(h^6) \qquad ...(ii)$$

The sum of (*i*) and (*ii*) gives

$$(1-h)^{-1} + \cos(h) = 2 + h + \frac{h^2}{2!} + h^3 + O(h)^4 + \frac{h^4}{4!}O(h^6) \qquad ...(iii)$$

Since $O(h^4) + \frac{h^4}{4!} = O(h^4)$ and $O(h^4) + O(h^6) = O(h^4)$

$\therefore$ (*iii*) takes the form $(1-h)^{-1} + \cos(h) = 2 + h + \frac{h^2}{2} + h^3 + O(h^4)$, which is of the 4th order of approximation.

Similarly the product of (*i*) and (*ii*) yields

$$(1-h)^{-1}\cos(h) = (1 + h + h^2 + h^3)\left(1 - \frac{h^2}{2!} + \frac{h^4}{4!}\right) + (1 + h + h^2 + h^3)\,O(h^6)$$
$$+ \left(1 - \frac{h^2}{2!} + \frac{h^4}{4!}\right)O(h^4) + O(h^4)\,O(h^6)$$

$$= 1 + h + \frac{h^2}{2} + \frac{h^3}{2} - \frac{11h^4}{24} + \frac{11}{24}h^5 + \frac{h^6}{24} + \frac{h^7}{24} + O(h^4) + O(h^6) + O(h^4)\,O(h^6) \qquad ...(iv)$$

Since $O(h^4)\, O(h^6) = O(h^{10})$

and $$-\frac{11h^4}{24} + \frac{11}{24}h^5 + \frac{h^6}{24} + \frac{h^7}{24} + O(h^4) + O(h^6) + O(h^{10}) = O(h^4)$$

$\therefore$ (iv) is reduced to $(1-h)^{-1}\cos(h) = 1 + h + \frac{h^2}{2} + \frac{h^3}{2} + O(h^4)$, which is of the 4th order of approximation.

1.9. GROWTH OF ERROR

Let $e(n)$ represent the growth of error after n steps of a computation process.

If $|\, e(n)\,| \sim n\,\varepsilon$, *we say that the growth of error is* **linear.**

If $|\, e(n)\,| \sim \delta^n\,\varepsilon$, *we say that the growth of error is* **exponential**.

If $\delta > 1$, the exponential error grows indefinitely as $n \to \infty$, and

if $0 < \delta < 1$, the exponential error decreases to zero as $n \to \infty$.

PROBLEMS 1.2

1. Find the smaller root of the equation $x^2 - 400x + 1 = 0$, correct to 4 decimal places.

2. If $r = h(4h^5 - 5)$, find the percentage error in r at $h = 1$, if the error in h is 0.04.
(W.B.T.U. B. Tech., 2005)

3. If $R = 10\, x^3y^2z^2$ and errors in x, y, z are 0.03, 0.01, 0.02 respectively at $x = 3, y = 1, z = 2$. Calculate the absolute error and % relative error in evaluating R ?

4. If $R = 4xy^2/z^3$ and errors in x, y, z be 0.001, show that the maximum relative error at $x = y = z = 1$ is 0.006.

5. If $V = \frac{1}{2}\left(\frac{r^2}{h} + h\right)$ and the error in V is at the most 0.4%, find the percentage error allowable in r and h when $r = 5.1$ cm and $h = 5.8$ cm.

6. Find the value of $I = \int_0^{0.8} \frac{\sin x}{x}\, dx$ correct to 4 decimal places.

7. Using the series $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots\ldots$, evaluate $\sin 25°$ with an accuracy of 0.001.

8. Determine the number of terms required in the series for $\log(1 + x)$ to evaluate $\log 1.2$ correct to six decimal places.

9. Use the series $\log_e\left(\frac{1+x}{1-x}\right) = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \ldots\ldots\right)$ to compute the value of $\log(1.2)$ correct to seven decimal places and find the number of terms retained.
(U.P.T.U., B.Tech., 2003)

10. Find the order of approximation for the sum and product of the following expansions :

$$e^h = 1 + h + \frac{h^2}{2} + \frac{h^3}{3!} + O(h^4) \quad \text{and} \quad \cos(h) = 1 - \frac{h^2}{2!} + \frac{h^4}{4!} + O(h^6).$$

11. Given the expansions :

$$\sin(t) = t - \frac{t^3}{3!} + \frac{t^5}{5!} + O(t^7) \quad \text{and} \quad \cos(t) = 1 - \frac{t^2}{2!} + \frac{t^3}{4!} + O(t^6)$$

Determine the order of approximation for their sum and product.

1.10. OBJECTIVE TYPE OF QUESTIONS

PROBLEMS 1.3

Select the correct answer or fill up the blanks in the following questions :

1. If x is the true value of a quantity and x_1 is its approximate value, then the relative error is

(a) $| x_1 - x |/x_1$ (b) $| x - x_1 |/x$

(c) $| x_1/x |$ (d) $x/| x_1 - x |$.

2. The relative error in the number 834.12 correct to five significant figures is ..... .

3. If a number is rounded to k decimal places, then the absolute error is

(a) $\frac{1}{2}\ 10^{k-1}$ (b) $\frac{1}{2} 10^{-k}$

(c) $\frac{1}{3}\ 10^{k}$ (d) $\frac{1}{4}\ 10^{-k}$.

4. If π is taken = 3.14 in place of 3.14156, then the relative error is .....

5. Given $x = 1.2$, $y = 25.6$ and $z = 4.5$, then the relative error in evaluating $w = x^2 + y/z$ is ..... .

6. Round off values of 43.38256, 0.0326457 and 0.2537623 to four significant digits are ..... .

7. Round relative maximum error in $3x^2y/z$ when $\delta x = \delta y = \delta z = 0.001$ at $x = y = z = 1$ is ..... .

8. If both the digits of the number 8.6 are correct, then the relative error is ..... .

9. If a number is correct to n significant digits, then the relative error is

(a) $\frac{1}{2}\ 10^{n}$ (b) $\frac{1}{2} 10^{n-1}$

(c) $\leq \frac{1}{2}\ 10^{-n}$ (d) $< \frac{1}{2}\ 10^{n-1}$.

10. If $(\sqrt{3} + \sqrt{5} + \sqrt{7})$ is rounded to four significant digits, then the absolute error is ..... .

11. $(\sqrt{102} - \sqrt{101})$ correct to three significant figures is ..... .

12. Approximate values of 1/3 are given as 0.3, 0.33 and 0.34. Out of these the best approximation is ..... .

13. The relative error if $\frac{2}{3}$ is approximated to 0.667, is ..... . *(U.P.T.C., MCA, 2009)*

14. If the first significant digit of a number is p and the number is correct to n significant digits, then the relative error is ..... .