Approximations and Errors in Computation

P. Sam Johnson

February 3, 2020

Overview

The study of error is a central problem of numerical analysis. Most numerical methods give answers that are only approximations to the desired true solution.

The input information given in form of a tabulated data, 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.

In the lecture, we discuss various kinds of approximations and errors in numerical calculations.

Inherent Error

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 aide.

Rounding Error

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 :

   (a) by changing the calculation procedure so as to avoid substraction of nearly equal numbers or division by a small number, (or)

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

Truncation Error

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

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = S(\text{say})$$

is replaced by

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} = S'(\text{say}),$$

then the truncation error is $S - S'$.

Absolute, Relative and Percentage Errors

If S is the true value of a quantity and S' is its approximate value, then the absolute value of $S - S'$,

$$E_a = |S - S'|,$$

is called the **absolute error**.

The **relative error** is defined by

$$E_r = \frac{|S - S'|}{|S|},$$

and the **percentage error** is

$$E_p = 100E_r.$$

Absolute, Relative and Percentage Errors

If a number if correct to n decimal places, then the absolute error is **less than** $\frac{1}{2}10^{-n}$.

If $X = 0.51$ and is correct to 2 decimal places, then $E_x = 0.005$, and the percentage accuracy is given by

$$\frac{0.0005}{0.51} x100 = 0.98\%.$$

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

Observations

If X is having n decimal digits and X' is its approximate value, then

(a) absolute error due to truncation to k digits: $|X - X'| < 10^{-n-k}$.

(b) absolute error due to rounding off to k digits: $|X - X'| < \frac{1}{2}10^{-n-k}$.

(c) relative error due to truncation to k digits: $|\frac{X-X'}{X}| < 10^{1-k}$.

(d) relative error due to rounding off to k digits: $|\frac{X-X'}{X}| < \frac{1}{2}10^{1-k}$.

Error in Addition / Subtraction of Numbers

Let $f = f(x_1, x_2, \ldots, x_n)$ be a function of n-variables x_i $(i = 1, 2, \ldots, n)$.

Let E_{x_i} be a error in x_i for each i. Let E_y be the corresponding error in y.

Suppose $y \equiv \pm x_1 \pm x_2 \pm \cdots \pm x_n$. Then $E_y = \pm E_{x_1} \pm E_{x_2} \pm \cdots \pm E_{x_n}$ and hence the absolute error is

$$|E_y| \leq \sum_{i=1}^{n} |E_{x_i}|.$$

The relative error is

$$\left|\frac{E_y}{y}\right| \leq \sum_{i=1}^{n} \left|\frac{E_{x_i}}{y}\right|.$$

Error in Addition / Subtraction of Numbers

While adding up several numbers of different absolute accuracies, the following procedure may be adopted.

1. Isolate the number the greatest absolute error,
2. Round-off all other numbers retaining in them one digit more than in the isolated number,
3. Add up, and
4. Round-off the sum by discarding one digit.

Error in Product / Quotient of Numbers

Suppose $y \equiv x_1 x_2$. Then $E_y = (x_1 + E_{x_1})(x_2 + E_{x_2}) - x_1 x_2$.

The absolute error is

$$E_y = x_1 E_{x_2} + x_2 E_{x_1} - E_{x_1} E_{x_2}.$$

Hence

$$E_y \approx x_1 E_{x_2} + x_2 E_{x_1}.$$

The absolute error in the quotient x_1/x_2 is given by

$$\frac{x_1 + E_{x_1}}{x_2 + E_{x_2}} \approx \frac{x_1}{x_2}\left(\frac{E_{x_1}}{x_1} - \frac{E_{x_2}}{x_2}\right).$$

Exercises

Exercises

1. *Round off the numbers* 865250 *and* 37.46235 *to* 4 *significant figure and compute* E_a, E_r, E_p *in each case.*
2. *Find the absolute error if the number* $X = 0.00545828$ *is*
   (a) *truncated to* 3 *decimal digits,*
   (b) *rounded off to* 3 *decimal digits.*
3. *The discharge* Q *over a notch for head* H *is calculated by the formula*

$$A = kH^{3/2}$$

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

Exercises

Exercises

5. *If the number p is correct to 3 significant digits, what will be the maximum relative error?*
6. *Three approximate values of the number $1/3$ are given as $0.3, 0.33$ and 0.34. Which of these three values is the best approximation?*
7. *Find the relative error of the number 8.6 if both of its digits are correct.*
8. *Evaluate the sum $S = \sqrt{3} + \sqrt{5} + \sqrt{7}$ to 4 significant digits and find its absolute and relative errors.*
9. *Find the difference $\sqrt{6.37} - \sqrt{6.36}$ to three significant figures.*
10. *Two numbers are given as 2.5 and 48.289, both of which being correct to the significant figures given. Find their product.*
11. *Calculate the value of $\sqrt{102} - \sqrt{101}$ correct to four significant figures.*

Exercises

Exercises

12. *Explain the term 'round-off error' and round-off the following numbers to two decimal places:*

$$48.21416,\ 2.3742,\ 52.275,\ 2.375,\ 2.385,\ 81.255.$$

13. *Round-off the following numbers to four significant figures:*

$$38.46235,\ 0.70029,\ 0.0022218,\ 19.235101,\ 2.36425.$$

14. *If* $p = 3c^6 - 6c^2$, *find the percentage error in* p *at* $c = 1$, *if the error in* c *is* 0.05.

15. *Find the absolute error in the sum of the numbers*

$$105.6,\ 27.38,\ 5.63,\ 0.1467,\ 0.000523,\ 208.05,\ 0.0235,\ 0.432,\ 0.0467,$$

where each number is correct to the digits given.