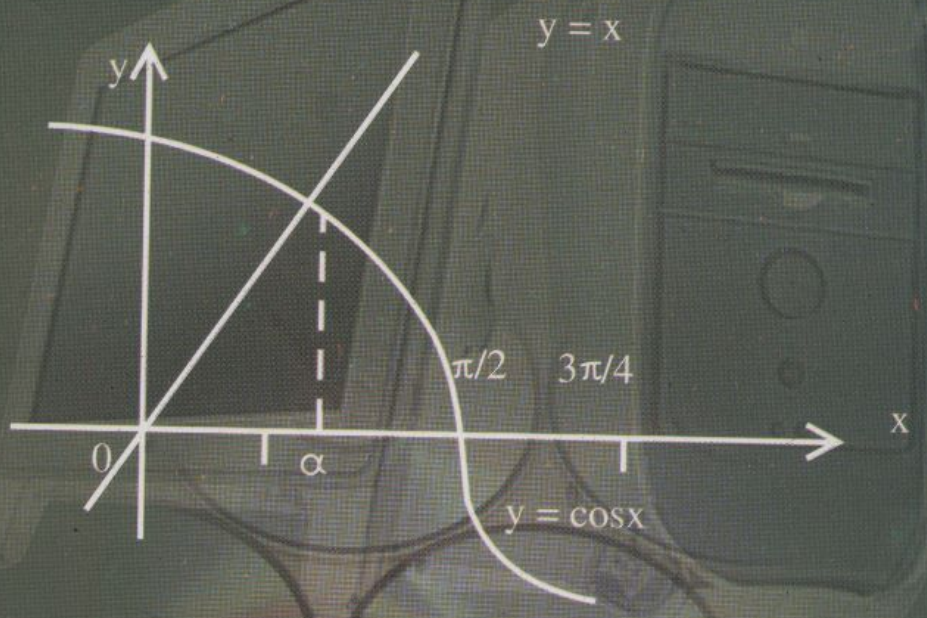
Prof.Dr. S A Bhatti
Mr. N A Bhatti

A First Course in NUMERICAL ANALYSIS With C++



Fifth Edition

Numerical Analysis is a Science

– computation is an art.

Fifth Enlarged Edition

A First Course in NUMERICAL ANALYSIS With C++

Prof. Saeed Akhter Bhatti
Mr. Naeem Akhtar Bhatti

SHAHARYAR PUBLISHERS
AL-FAZAL MARKET, URDU BAZAR, LAHORE.

CAN BE HAD FROM:

A-ONE PUBLISHERS
AL-FAZAL MARKET, URDU BAZAR, LAHORE
Phone No. 37232276 - 37357177 - 37224655
Email: aonepub@hotmail.com / info @aonepublishers.com
Website: aonepublishers.com

Book: A First Course in Numerical Analysis with C++

Author: Prof. Saeed Akhter Bhatti
Mr. Naeem Akhtar Bhatti

First Edition: May, 1990
Second Edition: Mar, 1996
Third Edition: Aug, 1999
Fourth Edition: May, 2002
Fifth Edition: Jan, 2008
Reprint: 2013

Quantity: 1100

Price: **Rs. 400/= (Rupees Four Hundred Only)**
Library Edition: **Rs. 500/= (Rupees Five Hundred Only)**

Published by:
Muhammad Azhar
(Shaharyar Publishers)

Printed at:
Ali Ejaz Printers,
Rattigun Road, Lahore

ISBN: **969-8105-01-8**

In the loving memory of our parents,

Mr and Begum Sana Ullah Sufi

Preface To The Fifth Edition

The Goal of this enlarged edition of our book on **Numerical Analysis** remains the same as for the previous editions: to give a comprehensive and state-of-the-art treatment of all the important aspects of the subject. In this, we have made modifications in all the first eight chapters and added extra problems at the end of each chapter. A new chapter, Chapter 9, based on eigenvalues and eigenvectors has been included. We have tried to cover all the basic and important procedures to compute eigenvalues and eigenvectors of a matrix. This chapter has been written especially on the request of users of the subject in various engineering universities.

We gratefully thank the users and the reviewers of the previous editions who provided valuable suggestions and ideas for the improvement of this book. Their feedback is valuable in our efforts for continuous improving this book. We are also thankful to our various teaching assistants both at BIIT and FUIMCS who checked the references and exercises in many chapters.

The authors would also like to thank Professor Akram Javed, Faculty of Science, University of Engineering and Technology, Taxila, for his many useful comments. We are thankful to Prof. Aftab Ahmad, Director, Institute of Management and Computer Sciences, Foundation University, Rawalpindi, for providing us the necessary infrastructure to complete this project. Thank you all.

The Bhattis
Islamabad,

Preface To The Fourth Edition

The Fourth Edition of this book on numerical analysis is in your hands now. It is geared specifically to the needs and background of our students. During this period, we received several comments from the users. In reviewing their comments, we have made modifications in some chapters of the book to sharpen the reader's understanding of the material presented. The plan of presentation of all chapters has been that of step by step. We start with an elementary method and then proceed to develop this or alternative, more sophisticated methods. The presentation just given is, of course, much over-simplified. In practice, a combination of conventional mathematical analysis and numerical analysis is likely to be used. Proofs of formulas are given where these are reasonably easy to follow but have been omitted in the more difficult cases.

A major change has been made in computer programs that implement the use of numerical methods presented in the book for solving problems. This edition contains computer programs written in C++. They have deliberately been kept as straightforward as possible so that the reader should understand the precise function of every step in each program. While the programs are intended primarily for educational-purposes, they can, of course, be used for solving some simple practical problems. However, for more complex practical problems, they do not offer any guarantee regarding the accuracy, adequacy or completeness of any information herein. Therefore, the user should make use of the excellent software packages now available. Hopefully the reader will appreciate this edition. We recommend them to learn and make more substantial use of their computers. We have benefited much by sitting at the feet of the wise, and we hope that, through this book, it may be possible to transmit a spark from their fire to all our readers. Good luck!

We would like to thank the users and reviewers of the previous editions whose comments and suggestions have enormously proved to be valuable in updating the material of the book. However, comments and suggestions for further improvements to the book and supporting software are welcome and can be communicated to us through the publisher. The authors would also like to express their gratitude to Prof. Akram laved, Dean, Faculty of Science, UET, Taxila, for his many useful comments received to improve the quality of this book and particularly to Dr. Jamil Sarwar, Director, BIIT, Rawalpindi, for providing necessary facilities to accomplish this reviewing exercise.

In closing, we are also grateful to our families for their continued patience and understanding during the review effort.

The Bhattis
Islamabad
May, 2002

Preface To The First Edition

The importance of Numerical Analysis to the scientists and engineers is now widely acknowledged. In the book world, there is no dearth of good books on numerical analysis written by foreign authors but the majority of these books are not available in this country. I have written this book to meet the long-felt need of indigenous students.

The main feature of the present text is to introduce numerical methods – covering the syllabi of various universities, colleges and other institutes, where this subject is being taught as a first course. In writing such an elementary book, I have inevitably been confronted by the problem of selection of material, which covers to a great extent the syllabi of the concerned institutes. Naturally, some will disagree with me over this choice of selection. I respect their prerogative. However, I shall be relieved if it is felt that the topics included do provide a reasonably solid background to the student's training and one from which he can easily proceed to further advanced courses in the subject.

The book is designed for a one-semester course in numerical analysis and consists of eight chapters. Each chapter includes a large number of thoroughly explained examples and problems of various complexity. These problems are very necessary and the students should work them out carefully. Each question has been designed to test the student's understanding of a particular formula. The answers of these problems are given at the end of the book. Proofs of formulas are included only where these are reasonably easy to follow, but the formulas are mentioned without proofs in the more difficult cases. It has been tried to keep the explanation straightforward and practically-oriented. The minimum prerequisite for using this book is elementary calculus (including some exposure to series and partial derivatives), linear algebra (determinant an matrices) and differential equations. It is also assumed that the student has taken a programming course in one of the computer languages. Fortran 77, which continues to be an excellent computer language for a wide variety of mathematical problems, is used in this book. Computer programs are given at appropriate places in the text.

No book emerges fully formed from an author's forehead. I would like to acknowledge the inspiration and encouragement I received from my colleagues and the help of many students who worked with early versions of the manuscript and checked exercise solutions and text examples.

The responsibility for any errors, omissions or lack of clarity naturally remains with me. I would appreciate having any such omissions, oversights or needed corrections called to my attention so that they can be implemented for improving the quality of this book. I would also like to thank Mr. Ghulam Shabir Qureshi and Syed Akbar Shah for their help in turning rough drafts into a beautifully prepared final manuscript.

I would like to express my gratitude to the National Book Council of Pakistam for accepting the manuscript of this book under the Creative Writer's Scheme. I also wishn to thank the anonymous referees who reviewed the manuscript.

Above all, I wish to thank my family, without whose encouragement, patiemce and sacrifice this book would not have been completed.

Saeed A. Bhatt
Islamabad
May, 1990

Contents

Chapter 1

Error Analysis

1.1 INTRODUCTION TO NUMERICAL ANALYSIS

When a mathematical problem can be solved analytically, its solution may be exact, but more frequently, there may not be a known method of obtaining its solution. For example, it is rather difficult to solve the following integral analytically:

$$\int_0^t \frac{e^{-x^2}\,dx}{\sqrt{1-x^2}}; \quad -1 \le t \le 1.$$

Many more such examples can be cited for which solutions by analytical means are either impossible or may be so complex that they are quite unsuitable for practical purposes. In this situation, the only way of obtaining an idea of the behaviour of a solution is to approximate the problem in such a manner that the number representing the solution can be produced. The process of obtaining a solution is to reduce the original problem to a repetition of the same step or series of steps so that the computations become automatic. Such a process is called a **numerical method** and the derivation and analysis of such methods lie within the discipline of **numerical analysis**. Thus, the subject of numerical analysis is concerned with the derivation, analysis and implementation of methods for obtaining reliable numerical answers to complex mathematical problems. In other words, numerical analysis is the subject concerned with the construction, analysis, and use of algorithms for the numerical solution of mathematical problems to given degree of numerical accuracy.

Numerical methods provide estimates that are very close to the exact analytical solutions; obviously, an error is introduced into the computation. It is important to understand that an error here does not mean a human error, such as a blunder or mistake or oversight but rather a discrepancy between the exact and approximate (computed) values. Such errors are likely to arise in all methods described in this book. In fact, numerical analysis is a vehicle to study errors in computations. It is not a static discipline. The continuous change in this field is to devise algorithms, which are both fast and accurate. These algorithms may become obsolete and may be replaced by more powerful algorithms as computer capability increases or as new techniques are developed. It is necessary to point out from personal experience that the best test of whether one understands a method is not to carry out a hand calculation (although this can be useful in early stages of attempting to understand the logic), but to program the method in a specific programming language, like BASIC, FORTRAN, PASCAL, C, C++ and JAVA

and run it on a computer. We all know that computers are ideally suited to handle tedious computations with high speed, accuracy and without ever making mistakes. Hence, the use of numerical method for the analysis, simulation and design of scientific and engineering processes and systems has been increasing at a rapid rate. This course is introduced to better prepare future scientists and engineers in understanding the fundamentals of numerical methods, especially their applications, limitations and potentials.

Although good computer programming skills can enhance the study of numerical analysis, actually writing programs are not always necessary. Numerical analysis is so important that extensive commercial software packages are available. For example, IMSL (International Mathematical and Statistical Library). It has several routines for numerical methods written in FORTRAN and C++. Some other packages are LAPACK (Linear Algebra Package) written in FORTRAN 77, LINPACK, EISPACK, Mathematica, Derive, Maple, MathCad, MathLab, MacSyma NUMERICOMP, etc. In addition a set of books, **Numerical Recipes**, lists and discusses numerical analysis programs in a variety of computer languages. However, one special feature of most of these programs is their ability to carry out many operations with exact arithmetic; an interesting example is to see the value of π displayed to 100 dp.

1.2 DEFINITION OF AN ERROR

The knowledge we have of the physical world is obtained by doing experiments and making measurements. It is important to understand how to express such data and how to analyze and draw meaningful conclusions from it. In doing this it is crucial to understand that all measurements of physical quantities are subject to uncertainties. It is never possible to measure anything exactly. It is good, of course, to make the error as small as possible but it is always there. And in order to draw valid conclusions the error must be indicated and dealt with properly. Take the measurement of a person's height as an example. Assuming that his height has been determined to be 5′ 8″, how accurate is our result?

Well, the height of a person depends on how straight he stands, whether he just got up (most people are slightly taller when getting up from a long rest in horizontal position), whether he has his shoes on, and how long his hair is and how it is made up. These inaccuracies could all be called **errors of definition**. A quantity such as height is not exactly defined without specifying many other circumstances. Even if you could precisely specify the "circumstances", your result would still have an error associated with it. The scale you are using is of limited accuracy; when you read the scale, you may have to estimate a fraction between the marks on the scale, etc. If the result of a measurement is to have meaning it cannot consist of the measured value alone. An indication of how accurate the result is must be included also. Indeed, typically more effort is required to determine the error or uncertainty in a measurement than to perform the measurement itself. Error, then, has to do with uncertainty in measurements that nothing can be done about. If a measurement is repeated, the values obtained will differ and none of the results can be preferred over the others. Although it is not possible to do anything about such error, it can be characterized. For instance, the repeated

measurements may cluster tightly together or they may spread widely. This pattern can be analyzed systematically.

All measurements, however, carefully and scientifically performed are subject to errors. Errors once committed contaminate subsequent results. **Errors analysis** is the study and evaluation of these errors; its main functions are to estimate the errors and suggest ways to eliminate or minimize them. Investigations of error propagation are, of course, particularly important in connection with iterative processes and computations where each value depends on its predecessors. Examples of such problems are in linear systems of equations, ordinary and partial differential equations. Since the study of errors is central to numerical analysis, we shall discuss it at length.

An error in a numerical computation is the difference between the actual value of a quantity and its computed (approximate) value. If x represents the computed value of a quantity, the actual value for which is x*, then the difference,

$$E = x^* - x \qquad \ldots \qquad (1.1)$$

is called the **error of approximation.**

1.3 SOURCES OF ERRORS

A numerical method for solving a given problem will, in general, involve an error of one or several types. Although different sources initiate the error, they all cause the same effect: **diversion from the exact answer.** Some errors are small and may be neglected, while others may be devastating if overlooked. In all cases, error analysis must accompany the computational scheme, whenever possible.

The main sources of error are as follow:

- Gross errors,
- Round errors,
- Truncation errors.

1.3.1 Gross Errors

Although gross errors are not directly concerned with most of the numerical methods discussed in this book, they can sometimes have great impact on the success of modeling efforts. Thus, they always be kept in mind when applying numerical techniques in context of real-world problems.

The gross errors are either caused by human mistakes or by the computer. Such mistakes are trivial, with better or no effect on the accuracy of the calculation, or they may be so serious as to render the calculated results quite wrong. A few examples of these errors are as follows:

i) Misreading or misquoting the figures, particularly in the interchanges of adjacent digits,

ii) Use of inaccurate mathematical formula (algorithm) to solve a particular problem, and

iii) Use of inaccurate data.

These errors are not very serious and can be avoided, if enough care is taken in using proper numerical analysis techniques. We shall primarily concern ourselves with the latter types of errors.

1.3.2 Rounding Errors

When a numerical method is actually run on a digital computer after transcription to computer program form a kind of error called **round-off error** is introduced.

The error introduced by rounding-off numbers to a limited number of decimal places is called the **rounding error**. In simple words, the error in the result that is caused by rounding is called round-off error. For example, it would be impracticable to mention the distance between two points on the earth as 15.2967 metres. It would be more reasonable if it were to be round to the nearest whole number, i.e., 15 metres. Thus, the error introduced by rounding is 0.2967 metres. Another example is the value of π = 3.1415926353 and may be meaningfully rounded-off to 3.1416 or 3.142.

Rounding-off errors play an important role in numerical analysis. In order to obtain a smaller error as a result of rounding-off, we may apply the following rules when performing manual calculations (these rules are not normally applied when performing extensive computer calculations).

Suppose we are given a number and we want to round it to the first decimal place. We discard all digits after the first decimal place and proceed as follows:

(a) If the first discarded digit is less than 5, the previous digit is unchanged. For example, the number 56.44, when rounded to the first decimal place, then it becomes 56.4.

(b) If the discarded digit is greater that 5, the previous digit is increased by 1. For example, the number 56.46, when rounded to first decimal, it becomes 56.5.

(c) If the discarded digit is exactly 5, the previous digit is unchanged, if it is even and is increased by 1, if it is odd. For example, the number 56.45, becomes 56.4 and 56.75 becomes 56.8.

However, the most commonly used rule (we are familiar with) for rounding-off the numbers is: **"if the discarded digit exceeds or equals 5, we add 1 to the last retained digit"**.

Analysis of the round-off error present in the final result of a numerical computation, usually termed the **accumulated** rounded-off error, is difficult, particularly when the algorithm used is of some complexity. Except in very simple cases, the accumulated error is not simply the sum of the **local** round-off error, that is, errors resulting from individual rounding or truncating operations. The local error at any stage of the calculation is propagated throughout the remaining part of the computation. In order to establish a round-off error bound, we must assume the worst possible outcome for the result of each arithmetic operation and follow the preparation of all such errors throughout the remaining calculations.

1.3.3 Truncation Errors

Truncation is defined as the replacement of one infinite series (or iterative process) by another with fewer terms. The error arising from this approximation is called the **truncation errors**. We shall devote considerable attention to truncation errors associated with the numerical methods discussed in this book. Because when different numerical methods are compared, we usually consider the truncation errors first.

In analyzing errors arising from the truncation of series, several types of series expansions can be considered. These include (but are not limited to) the following:

- Binomial expansion,
- Infinite geometric progression,
- Taylor/MacLaurin series.

In order to understand better the properties of truncation error, we turn to a mathematical formulation that is used commonly in numerical analysis for expressing functions in an approximate fashion – **the Taylor series.**

For example, the Taylor series expansion of f(x) about some chosen point x is defined by

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!}f''(x_0) + \frac{(x - x_0)^3}{3!}f'''(x_0)$$

$$+ \ldots + \frac{(x - x_0)^i}{i!}f^{(i)}(x_0) + \ldots + \frac{(x - x_0)^n}{n!}f^n(x_0) + R_n \qquad \ldots (1.2)$$

where R_n is the **remainder term** (error caused by truncating terms) that is included to account for all terms (n + 1) to infinity and is given by:

$$R_n = \frac{(x - x_0)^{n+1}}{(n+1)!}f^{(n+1)}(Z) \qquad \ldots (1.3)$$

where the subscript n connotes that this is the remainder for the nth-order approximation and Z is some value of x that lies somewhere between x_0 and x, i.e., $x_0 \le Z \le x$.

It is often convenient to simplify Taylor series by defining a step size $h = x - x_0$ and expressing (1.2) and (1.3) as,

$$f(x) = f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \frac{h^3}{3!}f'''(x_0) + \ldots +$$

$$\frac{h^n}{n!}f^{(n)}(x_0) + R_n \qquad \ldots (1.4)$$

where

$$R_n = \frac{h^{n+1}}{(n+1)!}f^{(n+1)}(Z) \qquad \ldots (1.5)$$

At this stage, it is sufficient to say that the remainder term provides an estimate of the maximum absolute error. We have devoted Section 1.7 to the computation of such errors.

1.4 SOME DEFINITIONS

Before proceeding further, let us define the following terms:

- Significant digits (figures),
- Precision and accuracy,
- Absolute, relative and percentage errors.

1.4.1 Significant Digits

In considering rounding errors, it is necessary to be precise in the usage of approximate digits. A **significant digit** in an approximate-number is a digit, which gives reliable information about the size of the number. In other words, a **significant digit** is used to express accuracy, that is, how many digits in the number have meaning. Whenever we use a number in a computation, we must have awareness that it can be used with confidence. To be significant, the last digit contained should be accurate within half a unit in the last decimal place. For example, if an approximate number α is equal to 1.23 and the value of α lies in the interval $1.225 \leq \alpha \leq 1.235$, then α is said to have three significant digits.

In considering significant digits, the following rules are generally used for number written in the conventional form:

a) Leading zeros are not significant.

b) Following zeros that appear after the decimal point are significant.

c) Following zeros that appear before the decimal point may or may not be significant, as more information is required to decide.

d) The significant digits in a number do not depend on the position of the decimal point in the number.

The above rules are illustrated by the following examples:

i) The number .0002025 has only four significant digits. The leading zeros are not significant.

ii) The number .00202570 has six significant digits. The following zeros should not be written unless it is significant.

iii) The number 2025000 may have four, five, six or seven significant digits depending upon the situation. The conventional form of writing number is somewhat ambiguous in this instance.

iv) The number 12546 and .12546 both contain five significant digits.

Note: The simplest way of reducing the number of significant digits in the representation of a number is merely to ignore the unwanted digits. This procedure,

known as **chopping**, was used by many early computers. A more common and better procedure is **rounding**, which involves adding 5 to the first unwanted digit and then chopping. For example, π chopped to four decimal places, is 3.1415, but it is 3.1416 when rounded; the representation of 3.1416 is correct to five significant digits (5S). The error involved in the reduction of the number of digits is called **round-off error**. Since π is 3.14159..., we note that chopping has introduced much more round-off error than rounding.

1.4.2 Precision and Accuracy

The meaning of the terms: precision and accuracy are often confused.

Precision is the number of digits in which a number is expressed or an answer given irrespective of the correctness of these digits. For example, if we are using a four-figure logarithmic table to perform calculations, our final answer will seldom be correct to four figures because of the accumulation of round-off errors.

Accuracy, on the other hand, is the number of digits to which an answer is correct. Accuracy can be quoted in either of the following ways:

i) to a given number of decimal places (abbreviated to **dp** throughout this book), or,

ii) to a given number of significant figures (abbreviated to **sf**).

Suppose, the result of a calculation is obtained as 65.5432, then the answer has a precision of 4 dp. If we know that the last two digits are unreliable, then the result may be rounded to 65.54 to achieve an accuracy of 2 dp or 4 sf. When statements about precision are made, the units involved need to be expressed. Thus, the quantity 6.474 kg is accurate to 4 sf, but precise to the nearest .001 kg; also the quantity is precise to the nearest .01 metre, but accurate to 1 sf.

Numerical methods should be sufficiently accurate (or unbiased) to meet the requirements of a particular scientific problem and they also should be precise enough. We now discuss the errors in performing numerical computations.

1.4.3 Absolute, Relative and Percentage Errors

The accuracy of any computation is always of great importance. There are two common ways to express (measure) the size or error in a computed result by **absolute error** and **relative error**. Let us define them one by one.

Absolute Error

We use the term **absolute error** (abbreviated to AE) to denote the actual value of a quantity less its rounded (approximate) value. If x and x* are respective by the rounded and actual values of a quantity, then the absolute error is defined by,

$$AE = |x^* - x| \qquad \ldots (1.6)$$

For example, if $x^* = 4.83$ and $x = 4.832$, then,

$$AE = |4.83 - 4.832| = .002$$

Generally, if a number is correct to n dp, it has a rounding error:

$$AE \leq \frac{1}{2} \times 10^{-n}.$$

Relative Error

Relative error (abbreviated to RE) is the ratio of the absolute error to the absolute actual value of a quantity.

$$\text{Thus,} \quad RE = \frac{AE}{|x^*|}; \; x^* \neq 0. \qquad \ldots (1.7)$$

If the actual value is not known, the relative error is defined by,

$$RE = \frac{AE}{|x|}; \; x \neq 0. \qquad \ldots (1.8)$$

As a measure of accuracy, relative error is more precise and meaningful than the absolute error, this is particularly so when the actual value is either very small or very large. The size of AE depends on the units used, whereas RE is a dimensionless quantity.

From the above example, $RE = \frac{.002}{4.83} = .00041.$

A decimal number correct to n significant-digits has:

$$RE \leq 5 \times 10^{-n}.$$

Percentage Error

Relative error expressed in percentage is called the **percentage error** (abbreviated by PE) and is defined by,

$$PE = 100 \times RE \qquad \ldots (1.9)$$

From the above example, $PE = 100 \times .00041 = .041\%$.

It is also called **probable error.**

In order to investigate the effect of total error in a method, we often compute an **error bound** which is a limit on how large and small the error can be.

1.5 EFFECT OF ROUNDING ERRORS IN ARITHMETIC OPERATIONS

In this section, we shall derive formulas for AE, and RE, for each of the fundamental operations of arithmetic, namely, addition, subtraction, multiplication and division, etc. Idea of error bound will also be introduced.

1.51 Error Accumulation in Addition

Let x_1 and x_2 be two approximate numbers and z be their sum. Then,

$$z = x_1 + x_2 \qquad \ldots (1.10)$$

Let e_1, e_2 and e_z be the errors in x_1, x_2 and z respectively.

Thus, we may add (or subtract) the errors from respective number:

$$z - e_z = (x_1 - e_1) + (x_2 - e_2)$$

$$= (x_1 + x_2) - (e_1 + e_2)$$

From (1.10), we have,

$$e_z = e_1 + e_2$$

Thus, the error simply add. So, the absolute error of two approximate numbers is given below:

$$AE = |e_z| \leq |e_1| + |e_2| \qquad \ldots (1.11)$$

The above proof can be extended to the sum of any given number of factors, i.e.,

$$AE = |e_z| \leq |e_1| + |e_2| + \ldots + |e_n| \qquad \ldots (1.12)$$

Hence, the absolute error of the sum of n approximate numbers does not exceed the sum of the absolute errors of the numbers.

The relative error is calculated using the following relation:

$$RE = \frac{\text{Absolute Error}}{\text{Absolute sum of the given number}}$$

$$= \frac{AE}{|z|} \qquad \ldots (1.13)$$

1.52 Error Accumulation in Subtraction

Let $z = x_1 - x_2$, where $x_1 > x_2$. $\ldots (1.14)$

As before, $z - e_z = (x_1 - e_1) - (x_2 - e_2)$

$$= (x_1 - x_2) - (e_1 - e_2)$$

From (1.14), we have,

$$e_z = e_1 - e_2$$

$$AE = |e_z| \leq |e_1| + |e_2| \qquad \ldots (1.15)$$

which is same as (1.11).

Hence, the absolute error of a difference between two numbers is the sum of the absolute errors of the given numbers. This formula can also be extended to any number of factors. Thus the formula for the addition of numbers and subtraction of numbers are the same.

Example 1 If the numbers 0.3062 – 0.25026 + 2.51392 are rounded, estimate the maximum absolute and relative errors. Find also the range in which the true answer lies.

Solution Let $x_1 = 0.3062$, $x_2 = 0.25026$ and $x_3 = 2.51392$.

Thus, $z = x_1 - x_2 + x_3 = 2.56986$.

Let e_1, e_2 and e_3 be the errors in x_1, x_2 and x_3, respectively. Thus, the absolute errors in the respective numbers are as follows:

$$|e_1| \le \frac{1}{2} \times 10^{-4}$$

$$|e_2| \le \frac{1}{2} \times 10^{-5}$$

$$|e_3| \le \frac{1}{2} \times 10^{-5}$$

$$AE = |e_1| + |e_2| + |e_3|$$

$$\le \frac{1}{2} \times 10^{-4} + \frac{1}{2} \times 10^{-5} + \frac{1}{2} \times 10^{-5} = 0.6 \times 10^{-4}$$

$$RE = \frac{AE}{z} = \frac{0.6 \times 10^{-4}}{2.56986} = 0.2335 \times 10^{-4}.$$

The result lies in the range $z \pm AE$:

$$2.56986 \pm 0.6 \times 10^{-4}$$

or $$2.56980 \le z \le 2.56992.$$

The answer may be rounded meaningfully to 2.57, which is correct to 3 sf (2 dp).

1.5.3 Error Accumulation in Multiplication

Suppose, we want to multiply two approximate numbers, x_1 and x_2.

Let $z = x_1 \cdot x_2$. ... (1.16)

As before, $z - e_z = (x_1 - e_1)(x_2 - e_2)$

$$= x_1 x_2 - x_1 e_2 - x_2 e_1 + e_1 e_2$$

Since e_1 and e_2 are small quantities, their product is still smaller and hence may be neglected. Thus,

$$z - e_z = x_1 x_2 - x_1 e_2 - x_2 e_1$$

From (1.16), we have,

$$e_z = x_1 e_2 + x_2 e_1 \quad \text{... (1.17)}$$

Dividing (1.17) by z, we get

$$RE = \left|\frac{e_z}{z}\right| \le \left|\frac{e_1}{x_1}\right| + \left|\frac{e_2}{x_2}\right| \qquad \ldots (1.18)$$

Hence, the relative error modulus of the product of two numbers does not exceed the sum of the relative error moduli of the given numbers.

Example 2 If the given numbers are rounded, estimate the relative and absolute errors of the product, $4.0643 \times .37487$. Find also the range in which the product lies.

Solution Let $x_1 = 4.0643$ and $x_2 = .37487$.

$$z = x_1 \cdot x_2 = 1.5236.$$

$$|e_1| \le \frac{1}{2}\times 10^{-4}; \quad |e_2| \le \frac{1}{2}\times 10^{-5}$$

Relative error, $RE = \left|\frac{e_1}{x_1}\right| + \left|\frac{e_2}{x_2}\right|$

$$< \frac{\frac{1}{2}\times 10^{-4}}{4.0643} + \frac{\frac{1}{2}\times 10^{-5}}{.37487}$$

$$< .2564 \times 10^{-4}$$

Absolute Error, $AE = RE \times z$

$$= .2564 \times 10^{-4} \times 1.5263 = .39 \times 10^{-4}.$$

Thus, the product leis in the range : $z \pm AE$

$$1.5263 \pm .39 \times 10^{-4}$$

or $1.523561 \le z \le 1.523639$

or 1.524 correct to 4 sf (or 3 dp).

Example 3 The values of x_1 and x_2 have been estimated as follows:

$$x_1 = 4.57 + e_1 \text{ and } x_2 = 8.48 + e_2$$

where $|e_1| < .35$ and $|e_2| < .82$. Find the range in which the product of x_1 and x_2 lies.

Solution

Upper Limit:

$$\begin{aligned} x_1 x_2 &= (4.57 + e_1)(8.48 + e_2) \\ &\leq (4.57 + |e_1|)(8.48 + |e_2|) \\ &< (4.57 + .35)(8.48 + .82) \\ &< (4.92)(9.30) = 45.76. \end{aligned}$$

Lower Limit:

$$\begin{aligned} x_1 \cdot x_2 &= (4.57 + e_1)(8.48 + e_2) \\ &\geq (4.57 - |e_1|)(8.48 - |e_2|) \\ &> (4.57 - .35)(8.48 - .82) \\ &> (4.22)(7.66) = 32.33. \end{aligned}$$

So, the product lies in the range, 32.33 to 45.76.

1.5.4 Error Accumulation in Division

Given two rounded numbers, x_1 and x_2.

Then $z = \dfrac{x_1}{x_2}$; $x_2 \neq 0$. ... (1.19)

As before, $z - e_z = \dfrac{x_1 - e_1}{x_2 - e_2}$

$$= \frac{x_1\left(1 - \dfrac{e_1}{x_1}\right)}{x_2\left(1 - \dfrac{e_2}{x_2}\right)}$$

$$= \frac{x_1}{x_2}\left(1 - \frac{e_1}{x_1}\right)\left(1 - \frac{e_2}{x_2}\right)^{-1}$$

Expanding with the help of binomial theorem and ignoring the product of errors being small, we have,

$$z - e_z = \frac{x_1}{x_2}\left(1-\frac{e_1}{x_1}\right)\left(1-\frac{e_2}{x_2}\right)$$

$$= \left(\frac{x_1}{x_2}-\frac{e_1}{x_2}\right)\left(1+\frac{e_2}{x_2}\right)$$

$$= \frac{x_1}{x_2} - \frac{e_1}{x_2} + \frac{x_1 e_2}{x_2^2}$$

$$e_z = \frac{e_1}{x_2} - \frac{x_1 e_2}{x_2^2} \qquad \ldots (1.20)$$

Dividing (1.20) by z, we get,

$$\frac{e_z}{z} = \left(\frac{\frac{e_1}{x_2} - \frac{x_1 . e_2}{x_2^2}}{\frac{x_1}{x_2}}\right)$$

$$= \frac{x_2}{x_1}\left(\frac{e_1}{x_2} - \frac{x_1 . e_2}{x_2^2}\right)$$

$$= \frac{e_1}{x_1} - \frac{e_2}{x_2}$$

$$RE = \left|\frac{e_z}{z}\right| \leq \left|\frac{e_1}{x_1}\right| + \left|\frac{e_2}{x_2}\right| \qquad \ldots (1.21)$$

Thus, the relative error of a quotient of two terms is equivalent to the sum of the relative error moduli of the dividend and divisor.

Example 4 Given the data, $\frac{4.0643}{37.487}$, estimate the following quantities:

a) the relative error,

b) the maximum absolute error, and

c) the range in which the quotient lies.

Solution Let $x_1 = 4.0643$ and $x_2 = 37.487$.

$$z = \frac{x_1}{x_2} = 0.1084$$

$$|e_1| \le \frac{1}{2}\times 10^{-4}$$

$$|e_2| \le \frac{1}{2}\times 10^{-3}$$

a) $$RE = \left|\frac{e_1}{x_1}\right| + \left|\frac{e_2}{x_2}\right|$$

$$\le \frac{\frac{1}{2}\times 10^{-4}}{4.0643} + \frac{\frac{1}{2}\times 10^{-3}}{37.487}$$

$$= .2564 \times 10^{-4}$$

b) $$AE = RE \times z$$

$$= .2564 \times 10^{-4} \times 0.1084 = 0.0287 \times 10^{-4}$$

e) The quotient lies in range, $z \pm AE$

$$0.1084 \pm 0.0287 \times 10^{-4}$$

or 0.1083972 to 0.1084028

or 0.108 correct to 3 dp.

1.5.5 Errors of Powers and Roots

Let $z = x^n$, where n is the power and denotes an integral or a fractional quantity.

As before, $z - e_z = (x - e_1)^n = x^n\left(1 - \frac{e_1}{x}\right)^n$.

Expanding the right side by the binomial theorem, and neglecting higher powers of $\frac{e_1}{x}$, we get

$$z - e_z = x^n\left(1 - n\frac{e_1}{x}\right)$$

$$= x^n - n e_1 x^{n-1}$$

Therefore, $e_z = n e_1 x^{n-1}$. ... (1.22)

Dividing (1.22) by z, we get

$$\frac{e_z}{z} = n\, e_1 \frac{x^{n-1}}{x^n}$$

$$= \frac{n\, e_1}{x}$$

$$RE = \left|\frac{e_z}{z}\right| \le |n| \cdot \left|\frac{e_1}{x}\right| \qquad \ldots (1.23)$$

Thus, the relative error modulus of a factor raised to a power is the product of the modulus of power and the relative error of the factor.

Example 5 Given $\sqrt{48.424}$, determine the maximum absolute error, relative error and the range in which the answer lies.

Solution Let $x = 48.425;\ n = \frac{1}{2}$.

Let $z = \sqrt{x} = \sqrt{48.424} = 6.959$

Therefore, $|e_1| \le \frac{1}{2} \times 10^{-3}$.

$$RE = \frac{1}{2} \times \frac{10^{-3}}{2} \times \frac{1}{48.425}$$

$$= .005 \times 10^{-3}$$

$$AE = RE \times z = 0.005 \times 10^{-3} \times 6.959 = .035 \times 10^{-3}$$

The correct value of z lies in the range, $z \pm AE$, i.e.,

$$6.959 \pm .035 \times 10^{-3}.$$

Example 6 Evaluate $\sqrt{6.2343 \times \frac{.82135}{2.7268}}$, and find the minimum transmitted error if the given numbers are rounded.

Solution Let $x_1 = 6.2343,\ x_2 = .82137$ and $x_3 = 2.7268$.

$$z = \sqrt{\frac{x_1 x_2}{x_3}} = 1.37035$$

$$n = \frac{1}{2};\ |e_1| = |e_3| \le \frac{1}{2}\times 10^{-4};\ |e_2| \le \frac{1}{2}\times 10^{-5}$$

$$RE \le \frac{1}{2}\left\{\frac{\frac{1}{2}\times 10^{-4}}{6.2343} + \frac{\frac{1}{2}\times 10^{-5}}{.82137} + \frac{\frac{1}{2}\times 10^{-4}}{2.7268}\right\}$$

$$= .16222 \times 10^{-4}$$

$$AE = RE \times z = .16222 \times 10^{-4} \times 1.37035 = 2.223 \times 10^{-5}$$

So, the answer lies in the range $z \pm AE$

$$1.37035 \pm 2.223 \times 10^{-5}$$

or 1.37033 to 1.37037

Thus, the answer, correct to 3 sf, is 1.37.

1.5.6 Error in Function Evaluation

Let $z = f(x)$.

As before, $z + e_z = f(x + e_1)$.

Using Taylor series expansion and neglecting higher powers of e_1, being small, we have,

$$Z + e_z = f(x) + e_1 f'(x)$$

or $e_z \hat{=} e_1 f'(x)$

Therefore, $AE = |e_z| \le |e_1 f'(x)|$. ... (1.24)

Dividing (1.24) by z, we get,

$$RE = \left|\frac{e_z}{z}\right| \le \left|e_1 \frac{f'(x)}{f(x)}\right| \quad \text{... (1.25)}$$

The formula can be extended to any given number of factors, for example, $z = f(x_1, x_2, \ldots, x_n)$:

$$RE = \left|\frac{e_z}{z}\right| \le \left|e_1 \frac{\delta f}{\delta x_1}\right| + \left|e_1 \cdot \frac{\delta f}{\delta x_2}\right| + \ldots + \left|e_n \frac{\delta f}{\delta x_n}\right|$$

$$\le \sum_{i=1}^{n} \left|e_i \frac{\delta f}{\delta x_i}\right| \quad \text{... (1.26)}$$

where $\frac{\delta f}{\delta x_i}$ are partial derivatives with respect to x_i, for i = 1, 2, ..., n.

Example 7 Estimate the absolute and relative errors if (i) $f(x) = e^x$ and (ii) $f(x) = \sin x$, for $x = 0.2345$, where x is rounded.

Solution

(i) $f(x) = e^x = e^{.2345} = 1.2643$

$f'(x) = e^x = 1.2643$

$$|e_1| \le \frac{1}{2} \times 10^{-4}$$

$$RE \le \left| e_1 \frac{f'(x)}{f(x)} \right| \le \frac{1}{2} \times 10^{-4} \times \frac{1.2643}{1.2643} = \frac{1}{2} \times 10^{-4}$$

$$AE = RE \times f(x) = \frac{1}{2} \times 10^{-4} \times 1.2643 = 0.00006$$

(ii) $f(x) = \sin(x) = \sin(0.2345) = .0041$

$f'(x) = \cos(x) = \cos(0.2345) = .9999$

$$|e_1| \le \frac{1}{2} \times 10^{-4}$$

$$RE \le \frac{1}{2} \times 10^{-4} \times \frac{.9999}{.0041} = 121.94 \times 10^{-4}$$

$$AE = RE \times z = 121.94 \times 10^{-4} \times 0.0041 = 4.9995 \times 10^{-5}$$

Note: The given angle in the trigonometric should be reported in radians. If the given angle, say θ, is in degrees, it should be converted to radians as:

$$\theta \text{ in degrees} = \frac{\theta \pi}{180} = \frac{\theta}{57.3} = \theta \times 0.0174 \text{ radians.}$$

1.6 NUMERICAL CANCELLATION

Accuracy may result in loss when two nearly equal numbers are subtracted. For example, the two numbers 9.4157233 and 9.4157227 are each accurate to 8 sf, yet their difference (0.0000006) is accurate to only 1 sf. Thus, care should be taken to avoid such subtractions where possible. This phenomenon is also called **subtractive cancellation.**

Case 1: We take first an example of evaluating roots of the quadratic equation, $ax^2 + bx + c = 0$, where $a \neq 0$. The roots are given by the formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

If 4ac is very very small as compared with b^2, the two quantities b and $\sqrt{b^2-4ac}$ will be nearly equal and one of the roots will be subject to a large error, thus resulting in a considerable loss of significance. It may lead to uncertainty in deciding whether the roots are real or complex. To avoid this situation we modify the formula in the following way:

If $b > 0$, the bigger root will be computed as, $x_1 = \dfrac{-b+\sqrt{b^2-4ac}}{2a}$ and the smaller root will be computed without loss of significance, suing the following:

$$x_1 x_2 = \frac{c}{a}; \quad x_2 = \frac{c}{a x_1}$$

Determining the smaller root by this method is far superior from the point of view of numerical analysis.

We shall illustrate the method by means of the following example.

Example 8 Find the roots of the equation, $x^2 - 40.12x + 1.3 = 0$, correct to 4 sf (The coefficients are exact).

Solution Let the roots be x_1 and x_2. Using the usual quadratic formula, the roots are: $x_1 = 40.087571$, $x_2 = 0.032429$. The larger root x_1 is given to 8 sf, whereas the smaller root x_2 to 5 sf. Thus, there is a loss of 3 sf. The second root is comparatively inaccurate. If the larger root, $x_1 = 40.09$ to 4 sf, then the smaller root, $x_2 = \dfrac{c}{a x_1} = 0.03243$ to 4 sf. Thus, a comparable number of significant figures can be given here as for the larger toot.

Another way to improve the quadratic formula is to calculate the roots with the following formulas:

i) $x_1 = \dfrac{-2c}{b+\sqrt{b^2-4ac}}$ and

ii) $x_2 = \dfrac{-2c}{b-\sqrt{b^2-4ac}}$.

In the cases when $|b| \approx \sqrt{b^2-4ac}$, we should proceed with caution to avoid loss of precision due to catastrophic cancellation. If $b > 0$, then x_1 should be computed with the formula (ii) and x_2 should be computed with formula (i). However, if $b < 0$, then x_1 should be computed using formula (i) and x_2 should be computed using formula (ii).

Case 2: Another example to illustrate the avoidance of loss of significance is as follows:

Example 9 Compare the results of computing f(500) and g(500) using six digits and rounding. The functions are as follows:

i) $f(x) = x\left[\sqrt{x+1} - \sqrt{x}\right]$ and

ii) $g(x) = \dfrac{x}{\sqrt{x+1} + \sqrt{x}}$.

Solution

i) $f(x) = x\left[\sqrt{x+1} - \sqrt{x}\right]$

$f(500) = 500\left[\sqrt{500+1} - \sqrt{500}\right]$

$= 500\,[22.3830 - 22.3607]$

$= 500 \times 0.0223 = 11.1500$

ii) $g(x) = \dfrac{x}{\sqrt{x+1} + \sqrt{x}}$

$g(500) = \dfrac{500}{\sqrt{500+1} + \sqrt{500}}$

$= \dfrac{500}{44.7437} = 11.1748$

The function g(x) is algebraically equivalent to f(x) as shown below:

$$f(x) = \left[x\sqrt{x+1} - \sqrt{x}\right] \times \frac{\sqrt{x+1} + \sqrt{x}}{\sqrt{x+1} + \sqrt{x}}$$

$$= \frac{x}{\sqrt{x+1} + \sqrt{x}}$$

The answer g(500) = 11.1748 involves less error and is the same as that obtained by rounding the true answer 11.174753 ... to six digits.

1.7 EVALUATION OF FUNCTIONS BY SERIES EXPANSION AND ESTIMATION OF ERRORS

This section deals with the problems of finding values of trigonometric, logarithmic, exponential and other functions by means of series expansion and also estmating errors, which arise when the series are truncated.

We confine our attention to the **Taylor series**, which is considered to be the foundation of numerical analysis. The series is commonly used in deriving several numerical methods.

Let f(x) be a function that is infinitely differentiable on an interval I containing the numbers x_0 and x. Then, for each positive integer n, the value of f(x) at x is given by:

$$f(x) = f(x_0) + (x - x_0) f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \frac{(x - x_0)^3}{3!} f'''(x_0)$$

$$+ \ldots + \frac{(x - x_0)^n}{n!} f^n(x_0) + R_n(x, x_0) \qquad \ldots (1.27)$$

where $R_n(x, x_0)$ is the is the remainder term and is included to account for all terms from (n + 1) to infinity.

$$R_n(x, x_0) = \int_0^x \frac{(x - t)^n}{n!} f^{(n+1)}(t) dt$$

$$= \frac{(x - x_0)^{n+1}}{(n+1)!} f^{(n+1)}(Z) \qquad \ldots (1.28)$$

for some unknown number Z which lies between x and x_0.

For a convergent series, $R_n(x, x_0)$ tends to zero as $n \to \infty$, i.e.,

$$\lim_{n \to \infty} R_n(x, x_0) = 0,$$

It follows that $f(x) = \sum_{n=0}^{\infty} \left(\frac{(x - x_0)^n}{n!} f^{(n)}(x_0) \right)$ ... (1.29)

The right hand side of (1.29) is called a **Taylor series representation** for f(x). It is a power of $(x - x_0)$ because the coefficients $\frac{f^{(n)}(x_0)}{n!}$ are constant — that is, they do not depend on x. The quantity Z in the remainder term is unknown and is difficult to calculate it. Nevertheless, we know the range in which Z lies. If we approximate f(x) in (1.27) or (1.29) by the first n term of the series, then the maximum error introduced in this series is given by the remainder term (1.28). Conversely, if the accuracy required is known before hand, then it would be possible to find the number of terms n such that the finite series give the required accuracy.

MacLaurin's Series

When $x_0 = 0$, in the Taylor series, we get MacLaurin's series and MacLaurin's polynomials. From (1.27), we get,

$$f(x) = f(0, x_0) = f(0) + f'(0) + \frac{x^2}{2!} f''(0) + \frac{x^3}{3!} f'''(0)$$

$$+ \ldots + \frac{x^n}{n!} f^{(n)}(0) + R_n(x) \qquad \ldots (1.30)$$

where $R_n(x) = \frac{x^{n+1}}{(n+1)!} f^{(n+1)}(Z)$ and $0 \le Z \le x$. ... (1.31)

Further examples of error analyses will be introduced in later chapters.

Example 10 Obtain a second degree polynomial approximation to the function $f(x) = \sqrt{x+1}$, using Taylor series about $x_0 = 0$. Calculate the truncation error for $x = 0.4$.

Solution

$$f(x) = (1+x)^{\frac{1}{2}}; \qquad f(0) = 1$$

$$f'(x) = \frac{1}{2}(1+x)^{-\frac{1}{2}}; \qquad f'(0) = \frac{1}{2}$$

$$f''(x) = -\frac{1}{4}(1+x)^{-\frac{3}{2}}; \qquad f''(0) = -\frac{1}{4}$$

$$f'''(x) = \frac{3}{8}(1+x)^{-\frac{5}{2}}; \qquad f'''(0) = \frac{3}{8}$$

From (1.29), we have,

$$f(x) = f(x_0) + (x - x_0) f'(0) + \frac{(x - x_0)^2}{2!} f''(x_0)$$

$$= f(0) + x f'(0) + \frac{x^2}{2} f''(0)$$

$$= 1 + \frac{x}{2} + \frac{x^2}{2} \times -\frac{1}{4}$$

$$= 1 + \frac{x}{2} - \frac{x^2}{8}$$

The truncation error (alternately, absolute error) can be calculated from the remainder term,

$$R(x) \le \frac{x^2}{3!} f'''(0) = \frac{x^3}{8} \times \frac{3}{8} = \frac{x^3}{16}$$

Therefore, $R(.4) \le \frac{0.4^3}{16} = 0.0042.$

Example 11 MacLaurin's series for e^x is given by,

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

where $0 < Z < x$. Determine the number of terms of this series such that their sum gives the value of e^x correct to 8 dp where $x = 1$.

Solution The remainder term is given by,

$$R(x) \leq \left| \frac{x^n}{n!} e^z \right|, \text{ at } Z = x.$$

The maximum relative error,

$$RE = \frac{\text{Absolute error}}{\text{Actual value}} = \left| \frac{x^n}{n!} e^z \times \frac{1}{e^z} \right| = \frac{x^n}{n!}$$

$$= \frac{1}{n!} \text{ at } x = 1.$$

For an accuracy of 8 dp, we have to add n terms such that

$$RE = \frac{1}{n!} < \frac{1}{2} \times 10^{-8}$$

or $n! = 2 \times 10^8 = 200000000$

If we take $n = 11$, $n! = 11! = 39916800$ and for $n = 12$, $12! = 479001600$. It is clear from above that about 12 terms of the series will be required to get an accuracy of 8 dp.

PROBLEMS

1. Find the absolute and relative errors in each of the following cases (all numbers are rounded).

   (a) $187.2 + 93.5$

   (b) 0.281×3.7148

   (c) $\sqrt{28.315}$

   (d) $\sqrt{\dfrac{6.2342 \times .82137}{27.268}}$

   (e) $2.3\,(4.18 - 3.24)$

   (f) $\dfrac{1.3384 - 2.038}{4.577}$

   (g) Evaluate the following as accurately as possible, assuming all values to be rounded.

   i) $8.24 + 5.33$

   ii) $124.53 - 124.52$

   iii) 4.27×3.13

iv) $9.48 \times 0.513 - 6.72$

v) $0.25 \times 2.84/0.64$

vi) $1.73 - 2.16 + 0.08 + 1.00 - 223 - 0.97 + 3.02$

2. If $x = 1.0$ and $y = 2.5$ round-up numbers, find the maximum absolute error involved in evaluating:

(a) $x + y$; (b) $\frac{x}{y}$; (c) $x^2 + xy + y^2$.

3. If two numbers x and y are in error by 1.0 and 0.5 respectively and the value of x is 10 and that of y is 6, state the intervals within exact values of x, y, $x - y$ and $x + y$ lie.

4. (a) Two parameters u and v have been estimated as follows:

$u = 2.5 + e_1$

$v = 4.5 + e_2$

where $|e_1| < 0.2$ and $|e_2| < 0.4$. Find bounds on the values of the product and quotient of u and v.

(b) The length and breadth of a rectangle are given by 2.52 cm and 1.78 cm respectively. Find the range in which its area lies, giving the answer to as many dp as are meaningful.

(c) Determine the largest relative error in a calculation of the cross-sectional area of a wire from a measurement on its diameter D, where $D = 0.825 \pm 0.002$ cm.

$$\left(\text{Area} = \pi\frac{D^2}{4}\right)$$

5. (a) Suppose 1.414 is considered to be an approximation of $\sqrt{2}$. Find the absolute and relative errors due to this choice.

(b) If $u = 0.1$ and $v = 0.01$ are rounded numbers, calculate the maximum absolute error in $\frac{u}{v}$.

(c) Determine the maximum relative error where p_1 is calculated from the relation: $p_1 u_2^n = p_2 u_2^n$; where $n = 1.4$. The maximum relative errors of u_1, u_2 and p_2 are 0.75%, 0.75% and 2.0% respectively.

(d) Obtain the range of values within which lies the exact value of

$$2.7654 + 3.8006 - \frac{15.178}{0.9876},$$ if all numbers are rounded off.

6. Obtain correctly rounded off answers for each of the following (all quantities are assumed rounded to the number of digits shown):

(i) $\cos 18°$, (ii) $\sin 0.18$, (iii) e^x for $x = 7.765$, (iv) $\ln x$ for $x = 1.377$.

(b) Rearranging the series speeds up the convergence:

$$\frac{\pi}{8} = \frac{1}{1\times3} + \frac{1}{5\times7} + \frac{1}{9\times11} \cdots \qquad \ldots (1)$$

Write a computer program in C++ to compute π using this series instead.

(c) Use the Taylor series;

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} \cdots$$

to write a program to compute cos x, correct to 4 dp, (x being in radians). See how many terms are required to achieve 4-figure agreement with the library function cos ( ).

13. (a) For small x, show that

i) $x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots$ is better than $e^x - 1$.

ii) $\frac{x^3}{6} - \frac{x^5}{120} + \ldots$ is better than $x - \sin x$.

iii) $\frac{x}{2} - \frac{x^2}{8} + \ldots$ is better than $1 - \sqrt{1-x}$.

(b) For value of v in the neighbourhood of $\frac{\pi}{2}$, show that $2 \sin^2 (\frac{\pi}{2} - v)\, z$ is better than $1 - \sin v$.

14. Use Taylor's theorem to estimate the truncation error in each of the following approximation formulas, when the step size h is small:

a) $$f'\left(x + \frac{h}{2}\right) = \frac{f(x+h) - f(x)}{h}$$

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

c) $$f'(x) = \frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + (x-2h)}{12h}$$

15. Derive the Taylor series approximate

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \ldots + \frac{(-1)^{n-1}}{n}x^n$$

stating clearly the form of the error term. How might it be bounded?