NUMERICAL METHODS

BY

R. A. BUCKINGHAM, PH.D. (CAMB.)
Reader in Physics, University College, London

LONDON
SIR ISAAC PITMAN & SONS, LTD.

PREFACE

It will be evident from the contents of this book that its main concern is with the needs of those engaged in scientific computing problems using desk calculators. Nevertheless, one hopes that it will also serve some of those fortunate enough to have access to electronic computers, the advent of which has certainly multiplied many times the quantity of computing being done, but has not made unnecessary a study of methods suitable for hand calculations. The danger is rather that too many may come to use fast machines without adequate knowledge and experience of such methods and the pitfalls of numerical work generally. No doubt the day will come when a handbook of techniques appropriate to electronic computers can be written, but experience is still growing too rapidly to make this other than a hazardous adventure at present.

An attempt has been made to cover a fairly wide range of computing problems, and to blend the old with the new. Many familiar and well-tried methods are therefore to be found here as well as more recent developments. The importance of matrix methods in computing is such that they have been introduced gradually towards the middle of the book, and subsequently stressed; the same applies to processes involving iteration. However there are some notable omissions, such as quadrature formulas of the Gaussian type, and Runge-Kutta methods of integration. If any excuse need be made for this, it is that they are well described elsewhere, and that their value for hand computing is limited. A more serious fault perhaps lies in the emphasis given to methods based on polynomial approximation and comparative neglect of trigonometrical functions; this should some day be remedied.

The fact that the book runs to nearly 600 pages is largely due to the inclusion of many worked examples. It is hoped that these will give readers a reliable impression of the value of the methods described. Apart from this no attempt has been made to provide exercises of a purely numerical kind. These are extremely easy to construct, and those engaged in computing will mostly prefer to get on with their own problems. When practice is wanted it is more profitable to repeat step by step the example given in the text, and to consider whether it has in fact been done in the best way. The exercises following some chapters are often exercises in name only, being rather extensions of the text with possibly significant applications.

In writing the book I have been aided greatly by the published work of British mathematicians, past and present, whose contributions to numerical analysis have indeed been considerable. A similar debt is owed to America, especially to the work of W. E. Milne. To Prof. A. C. Aitken I am grateful among other things for allowing me to include some early numerical examples of his on the solution of polynomial equations. Finally, it is very pleasant to acknowledge that my interest in computation was greatly stimulated by the enthusiasm of the late Dr. Comrie, and sustained by Prof. H. S. W. Massey, whose insistence on the importance of numerical techniques for scientists in general and physicists in particular gave to this volume its origin and purpose.

R. A. BUCKINGHAM

UNIVERSITY COLLEGE
LONDON

ERRATA

1. ***The layout of the table of differences on page* 136 *should be as follows*—**

x	$1/(1+x^2)$	δ	δ^2	δ^3	δ^4	δ^5	δ^6
		+ 38462		− 15917		+ 14050	
0·0	1·000000		− 76924		+ 31834		− 28100
		− 38462		+ 15917		− 14050	
0·2	0·961538		− 61007		+ 17784		− 8892
		− 99469		+ 33701		− 22942	
0·4	0·862069		− 27306		− 5158		+ 14102
		− 126775		+ 28543		− 8840	
0·6	0·735294		+ 1237		− 13998		+ 12103
		− 125538		+ 14545		+ 3263	
0·8	0·609756		+ 15782		− 10735		+ 2236
		− 109756		+ 3810		+ 5499	
1·0	0·500000		+ 19592		− 5236		− 1944
		− 90164		− 1426		+ 3555	
1·2	0·409836		+ 18166		− 1681		− 1937
		− 71998		− 3107		+ 1618	

2. ***In the table at the foot of page* 425 *the words* "check" *and* "Sum" *appearing above the last two columns should read* "Sum check" *and apply only to the last column.***

BOOKS AND TABLES REFERRED TO IN THE TEXT

Books

The abbreviated titles on the left are used in the text.

Collatz, *N.B.D.*	Collatz, L., *Numerische Behändlung von Differential-gleichungen* (Springer, 1951).
Dwyer, *L.C.*	Dwyer, P. S., *Linear Computations* (Wiley, 1951).
Frazer, Duncan and Collar	Frazer, R. A., Duncan, W. F., and Collar, A. R., *Elementary Matrices* (Cambridge, 1938).
Hartree, *N.A.*	Hartree, D. R., *Numerical Analysis* (Oxford, 1952).
Jeffreys, *M.M.P.*	Jeffreys, H. and Jeffreys, B. S., *Methods of Mathematical Physics*, 2nd ed. (Cambridge, 1950).
Milne, *N.C.*	Milne, W. E., *Numerical Calculus* (Princeton, 1949).
Milne, *N.S.D.E.*	Milne, W. E., *Numerical Solution of Differential Equations* (Wiley, 1953).
Mineur, *T.C.N.*	Mineur, H., *Techniques de Calcul Numérique* (Librairie Polytechnique Ch. Béranger, 1952).
Scarborough, *N.M.A.*	Scarborough, *Numerical Mathematical Analysis*, 2nd ed. (Johns Hopkins Univ., 1950).
Steffensen	Steffensen, *Interpolation* (1925, Dover rep. 1950).
Whittaker and Robinson	Whittaker, E. T. and Robinson, C., *Calculus of Observations*, 4th ed. (Blackie, 1944).

Tables, etc.

Barlow	*Barlow's Tables of Squares, Cubes, etc.*, ed. Comrie, 4th ed. (Spon, 1941).
Chambers [4]	*Chamber's Four-figure Mathematical Tables*, ed. Comrie (1947).
Chambers [6]	*Chamber's Six-figure Mathematical Tables*, Vol. 2, "Natural Values," ed. Comrie (1949). *Note.* Vol. 1 contains logarithmic values.
Dale [5]	Dale, J. B., *Five-figure Tables of Mathematical Functions*, 2nd ed. (Arnold, 1949).

I.A.T.	*Interpolation and Allied Tables* (H.M. Stationery Office, rep. 1947).
Jahnke-Emde	Jahnke, P. R. E. and Emde, F., *Funktiontafeln*, 3rd ed. (Dover rep., 1943); also 4th ed. (Teubner, 1948).
M.T.C. [4]	Milne-Thomson and Comrie, *Standard Four-figure Mathematical Tables* (Macmillan, 1931).
Index	Fletcher, A., Miller, J. C. P., and Rosenhead, L., *Index of Mathematical Tables* (Scientific Computing Service, 1946; 2nd ed. in preparation).
Math. Tab., Wash.	*Mathematical Tables and Other Aids to Computation* (sometimes referred to as *M.T.A.C.*; a valuable quarterly journal).

In addition to the above general books of tables, there are some valuable series of more specialized tables; in particular, those sponsored by the British Association (and subsequently by the Royal Society), and by the U.S. National Bureau of Standards (Applied Mathematics Series).

CONTENTS

CHAPTER 1

AN INTRODUCTION TO COMPUTATION

WHAT is the art of computation? This question may for the moment be answered as follows. Most calculations are carried out with numbers which are to some extent approximate, by methods which are often inexact, and their results, expressed in numerical form, are also approximate; yet there is normally only one answer which is correct to a specified degree of accuracy. The art of computation therefore lies in obtaining this correct answer with reasonable certainty, and with the least unnecessary labour.

Like most short definitions, this statement implies a great deal more than is immediately obvious. The emphasis, however, is not merely on attaining the right result, but also on knowing that it is right, and the extent to which it can be relied upon, even though in the process rather more work has been done than seemed essential. A satisfactory end to a computation depends largely on using well-tried methods in solving particular types of problem, chosen both for ease of checking and economy of effort. The development of such methods is the main preoccupation of later chapters; for the present we shall be concerned more with the question of accuracy, which is now seen to have two aspects. First, there is the assessment of errors which are legitimate and unavoidable; secondly, the avoidance of those which are illegitimate, or what is so often necessary, the speedy discovery and removal of mistakes which should never come into being. It is therefore not inappropriate that the first section of this chapter dealing in general terms with the art of computation should be mainly concerned with errors, both genuine and accidental. In it, we hope to discuss satisfactorily what errors are and how they are propagated in simple calculations, and the principles, rules and precautions which all help to eliminate blunders.

In asking for a minimum of effort the definition implies further that the best possible use should be made of auxiliary information, provided for instance by tables, and of such machines as may be available.

We envisage the computer as presiding over (*a*) one or more calculating machines on which the majority of arithmetical operations are performed and numbers are temporarily stored, (*b*) all the necessary tables, representing a storehouse of numerical knowledge on which the computer can call at will, (*c*) a sheet of paper containing the preliminary data of the problem, to which are added the essential parts of the calculation as it proceeds, and the final result. The computer's task is then so to organize the transference of numbers between the three sets of equipment under his control that the solution of the problem is achieved as rapidly and as surely as possible.

With a view to filling in some of the details of this general picture, the second and third sections of this chapter discuss some of the questions raised by the use of tables, and the role of calculating machines, particularly desk calculators.

Errors

A few elementary definitions are necessary. If y' is the approximate value of a quantity whose exact value is denoted by y, practice differs as to whether the error of y' should be defined by $y' - y$ or $y - y'$. The distinction is often irrelevant, and we shall often use the term error in the broad sense when sign does not matter. When it does matter, we shall try to adhere to the following definitions—

Absolute error $\qquad e = y' - y$

Absolute correction or remainder $r = y - y'$

Thus to obtain the correct value, the error should be subtracted or the correction added to the approximate value. The term *deviation* may sometimes be used instead of absolute error, particularly when this represents, not the difference between y' and y itself, but between y' and the best estimate of y that happens to be available.

Although errors are often expressed in absolute measure, which will involve the dimensions, if any, of the quantity concerned, a better guide to the merit of an approximation is often given by its relative error

$$\delta = \left|\frac{y' - y}{y}\right| \simeq \left|\frac{y' - y}{y'}\right|$$

or by the *percentage error*, which is 100δ. Here it is usually only magnitude which matters, and relative errors are obviously dimensionless. In particular the measurement of a physical quantity, e.g. length, mass, time, is regarded as more accurate the smaller the relative error. The term tends to lose any significance, however, if δ exceeds, say 0·5, and it may then be preferable to define the error logarithmically, e.g.

$$\text{logarithmic error} = \log_{10} |y'/y|$$

A logarithmic error of $\pm$ 1, i.e. a factor of 10, is loosely described as "an order of magnitude." We shall be concerned mainly with relative errors.

So much for definitions which do not involve any consideration of the nature or origin of the errors concerned. When we consider the numbers which provide the starting-point for calculation, however, it is important to distinguish two types of error which they may contain. These are

(*a*) rounding errors,

(*b*) experimental or statistical errors.

We shall discuss rounding errors first.

1.1. Rounding Errors

The usual method of rounding off a number to a desired number of decimals is illustrated by successive reductions of π—

3·1415926535...
3·141593
3·14159
3·1416
3·142
3·14

The rule for rounding k decimals may be summarized as follows—

Rule A

$(k+1)$th decimal	kth decimal
0, 1, 2, 3, 4	unchanged
6, 7, 8, 9	increased by 1

If this rule is correctly applied, the absolute rounding error will not exceed $\pm$ 0·5 $\times$ 10^{-k}.

The example above does not involve the critical case represented by the number

1234·5000...

which might be rounded up to 1235 or down to 1234 without violating the limits $\pm$ 0·5 for the rounding error. To dismiss the critical case thus, however, is to be a trifle unrealistic. The computer rarely knows a large number of digits beyond the one which is being rounded; often he knows only two or three, or perhaps only one, and the last known digit may itself be uncertain to 1 or 2 units, because of previous rounding and other causes.

A typical instance is the rounding of a $(k + 1)$-decimal table to k decimals. Of the final digits to be dropped, roughly one-tenth will be 5's, and of these about one-half would involve rounding up and one-half rounding down if further decimals were available. Since the computer does not know which to round up and which to round down, a convention is necessary which preserves the random distribution of rounding errors between positive and negative values. The usual convention may be expressed as a rule supplementary to rule A already given—

Rule B

When the $(k + 1)$th decimal is 5, the kth decimal is increased by 1 when odd, and unchanged when even.

This involves an increase in the maximum possible rounding error. Thus if 1234·5 can in fact represent any number between 1234·40... and 1234·60..., the effect of rounding to 1234 may introduce a rounding error of magnitude 0·6.

When numbers are rounded off to k decimals from $(k + 2)$ decimals, the last of which is uncertain to 1 unit, a rounding error of $\pm$ 0·51 of the kth decimal should not be exceeded in the critical case, which is now ten times less frequent.

In the light of this it is interesting to examine a device commonly used in printing tables*; this is the addition of 5 in the place following the last to be retained, as in the following table of sin x (to be rounded to 4 decimals)—

x	sin x		
0·10	0·0998 33		0·0998 **83**
0·11	0·1097 78	modified	0·1098 **28**
0·12	0·1197 12	to	0·1197 **62**
0·13	0·1296 34		0·1296 **84**

* It is also used in automatic digital machines.

In the modified table the figures to the right of the gap can merely be removed, and the result is a table correctly rounded according to rule *A*. All critical cases, however, are rounded *up*, thus violating rule *B*. When two or more decimals are dropped this rarely matters, because it is so infrequent, but when dropping one decimal only, about 5 per cent of values would be improperly rounded and so give an appreciable excess of positive rounding errors. If these were extensively combined by, say addition, the result might be an unwelcome accumulation of errors.

It is sometimes an advantage to limit rounding errors more closely than is possible by the standard method of rounding. We mention two notations which achieve this to some extent.

(*a*) Single-dot Notation

In this, the last digit to be kept is rounded up only if the figures neglected lie between 0·750... and 0·999..., in units of the last decimal; if these figures lie between 0·250... and 0·750..., the last digit is followed by a dot or some other symbol. Thus 1·23456 might be denoted by—

$$1{\cdot}234^{\cdot} \quad \text{or} \quad 1{\cdot}234^{+} \quad \text{or} \quad 1{\cdot}234_{5}$$

Further examples are given by successive approximations to π—

$3{\cdot}1415926^{\cdot}$	$3{\cdot}1416$
$3{\cdot}141592^{\cdot}$	$3{\cdot}141^{\cdot}$
$3{\cdot}14159^{\cdot}$	$3{\cdot}14$

Addition and subtraction of numbers thus rounded proceed on the assumption that $^{\cdot}$ is equivalent to 1/2.

(*b*) High- and Low-dot Notation

Rounding is carried out by the standard method, but *in addition*, a high dot is added if the neglected figures lie between 1/6 and 1/2 (in units of the last decimal), or a low dot if these figures lie between 1/2 and 5/6.

Thus, using π again as an example, we have—

$3{\cdot}14159265^{\cdot}$	$3{\cdot}1416$
$3{\cdot}1415927.$	$3{\cdot}142.$
$3{\cdot}141593.$	$3{\cdot}14$
$3{\cdot}14159^{\cdot}$	

Addition and subtraction are carried out as if a high dot were equivalent to + 1/3, and a low dot to − 1/3; thus

$$2^{\cdot} + 2. = 4 \qquad 2^{\cdot} + 2^{\cdot} = 5. \qquad 3 \times 2^{\cdot} = 7$$

Obviously method (*a*) limits the rounding error to $\pm$ 1/4 and (*b*) limits it to $\pm$ 1/6, excluding critical cases. An advantage of (*b*), when used in printed tables (e.g. *M.T.C.*), is that the dots can be ignored if the table is accurate enough without using the extra information which they provide.

Let us return for a moment to the definitions of absolute and relative error. When the errors of numbers arise mainly from rounding, as in tables of mathematical functions, a good measure of their absolute accuracy is provided by the number of correct decimal places. A much cruder measure of relative accuracy is given by the number of significant figures. Thus the numbers 0·000101 and 0·000999, both of which have the same absolute accuracy afforded by 6 decimals, and 3 significant figures, may yet have relative errors differing by an order of magnitude. It is common practice to use nD or nS to denote that a number is specified with n correct decimal or significant figures respectively.

1.2. Experimental Errors

The numerical description of experimental errors is usually less simple than that of rounding errors. One of the properties commonly associated with rounding errors, say in a standard mathematical table containing k decimals, is that they are uniformly distributed between extreme values $\pm 0{\cdot}5 \times 10^{-k}$.

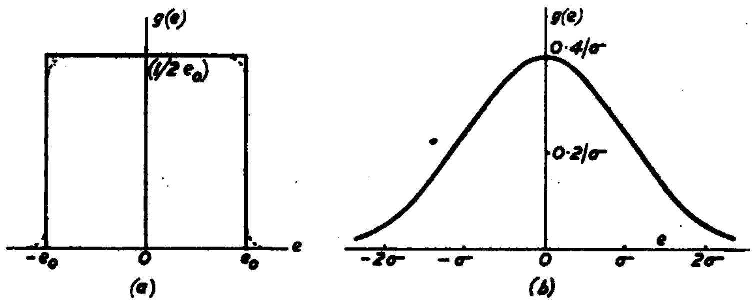

FIG. 1.1. PROBABILITY DISTRIBUTIONS

(*a*) Rectangular distribution of rounding errors.
(*b*) Distribution of experimental errors. The probability of an error between e and $e + de$ is $g(e)\,de$.

This implies a probability distribution which is rectangular in shape (see Fig. 1.1 (*a*)) or very slightly rounded if critical cases are significant. It is true that some distributions of rounding errors may be far from uniform, e.g. in a 2- or 3-decimal table of

$n/16$ or $n/3$, where n is an integer; so one must be on the watch for exceptional cases. However, if a quantity is stated to have the value

$$y' \pm e_0$$

and the error arises from rounding, it is generally assumed that the same probability is attached to any error between the limits $\pm e_0$.

When the error is experimental in origin, however, and the same form is used, the significance of e_0 is usually quite different. This is because experimental errors (systematic errors apart), tend to be distributed in a way more resembling the curve in Fig. 1.1 (b), representing the *normal error distribution*, in which

$$g(e) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left(-e^2/2\sigma^2\right)$$

This contains a parameter σ, usually called the *standard deviation*, such that about 5 per cent of errors exceed 2σ in magnitude and slightly less than 0·3 per cent exceed 3σ. On the supposition that errors are distributed according to the normal law, it is possible to state the approximate value of a physical quantity as $y' \pm \sigma$, where σ is estimated in a prescribed way from a set of measurements of the quantity. One warning is necessary: the *probable error*, approximately $0{\cdot}6745\sigma$, is sometimes used in place of σ; so, if confronted with say, a measured length 23·45 $\pm$ 0·21 cm, one must make sure whether 0·21 represents the standard deviation or probable error of the measurements.

1.3. Effects of Simple Calculations on Errors

The result of any computation is affected by

(a) errors in the data, which we have seen may be due either to rounding or to experiment or possibly both;

(b) approximations made during the computation, leading to what may be termed *computational errors* or *truncation errors*. The latter term is used because the kind of approximation which is made involves the use of a finite number of terms of an infinite series, or a finite number of repetitions of an iterative process which should strictly be carried out *ad infinitum*. Rounding errors, too, made during the computation, are essentially computational, and are often such that their cumulative effect is very difficult to assess. On the other hand, a calculation which involves only arithmetical processes, such as addition, subtraction, multiplication, etc., and which can be carried out with virtually complete accuracy, does not contribute to computational errors, but only modifies errors already present in the data. It is the effect of such calculations that we are concerned with at the moment, though some simple computational errors, arising in linear interpolation, will be considered in § 1.6.

Addition and Subtraction

Let
$$Y = a_1y_1 + a_2y_2 + \ . \ . \ . + a_ny_n$$

where the coefficients a_k may be positive or negative. If Y' is an estimate of Y obtained by using approximate values for the y's, e.g. y_1' for y_1, with an error $e_1 = y_1' - y_1$, etc., then the resulting error in Y' is E, where

$$E = a_1e_1 + a_2e_2 + \ . \ . \ . + a_ne_n$$

and hence

$$|E| \leqslant |a_1| \ . \ |e_1| + |a_2| \ . \ |e_2| + \ . \ . \ . + |a_n| \ . \ |e_n| \qquad (1.1)$$

When the magnitudes of the individual errors are limited, (1.1) provides an upper limit to the error involved in Y'. It is easy to see how $|E|$ may be dominated by a single large absolute error, as in the sum of 4-digit numbers shown below. If each number is assumed to have been rounded in the 4th digit, the error of the sum comes almost entirely from 1001; digits to the right of the dotted line have little significance.

23·8	2
12·9	1
0·3	281
1001·	
821·4	
1859·4	

Multiplication and Division

Let
$$Y = y_1^{p_1}y_2^{p_2} \ . \ . \ . \ y_n^{p_n}$$

where the p's may be positive or negative integers, and let an estimate of Y again be made, using y_1' for y_1 and so on. With the assumption that the relative errors of y_1', etc., are small, so that, for instance,

$$(y_1 + e_1)^{p_1} \simeq y_1^{p_1}(1 + p_1e_1/y_1)$$