NUMERICAL METHODS

BOOTH

SECOND EDITION

BUTTERWORTHS SCIENTIFIC PUBLICATIONS

NUMERICAL METHODS

By

ANDREW D. BOOTH, D.Sc.

Reader in Computational Methods,
University of London

SECOND EDITION

LONDON
BUTTERWORTHS SCIENTIFIC PUBLICATIONS
1957

NUMERICAL METHODS

PREFACE

The present book has grown out of the series of lectures given by the author to final honours B.Sc. mathematics students at Birkbeck College, London. The purpose of the course has been, not so much to instruct in the detailed tedium of actual calculation, but rather to give an understanding of the basic principles upon which such analyses rest.

Some existing books on numerical analysis lay much stress upon the detailed form in which a given procedure is to be laid out; it has been my experience that such concentration upon actual numbers obscures the underlying mathematical basis upon which the work rests, and so such tabulations are almost entirely absent from this book. Where they are given, as in Chapter 7, they illustrate the sort of behaviour which will be encountered in a calculation rather than any detailed form of layout.

Were these didactic points the sole reason there would be little justification for a new book on computation; a far more important consideration lies in the growth, during the past decade, of the science and art of programming for an automatic digital calculator. The classical methods of hand calculation are, to a greater or less extent, unsuitable for the modern machines, and only by having a thorough knowledge of the underlying mathematical principles, is the programmer likely to make effective use of the new tools.

At Birkbeck College Computational Laboratory the teaching of numerical methods has been accompanied by the actual use of an automatic calculator, and demonstrations of such things as differencing and the solution of differential equations have been carried out by the machine and not by the student. Perhaps not unnaturally, this has proved more popular than the old method.

A book of this kind must always owe much to the work of previous authors; it is pleasant to acknowledge the help which the author derived from Freeman's 'Actuarial Mathematics' and from the classical 'Calculus of Observations' of Whittaker and Robinson, both of which were practically the only available works during the 1930's. In more recent times the paper on 'Difference and Associated Operators' by W. G. Bickley may be mentioned as having particular influence.

Finally it gives me particular pleasure to acknowledge the help of my wife both in making the book more readable than might otherwise have been the case, and also, in collaboration with J. P. Cleave, B.Sc., for checking the examples given in Chapter 7.

A. D. B.

Fenny Compton
November 1954

AUTHOR'S NOTE

Advantage has been taken of the need for a second edition to correct small errors which were present in the earlier version, and the author wishes to express his thanks to reviewers and others who drew his attention to these.

A. D. B.

Fenny Compton
February 1957

CONTENTS

1
THE NATURE AND PURPOSE OF NUMERICAL ANALYSIS

1.1 HISTORY

ALTHOUGH numerical analysis is considered by some to be a subject of recent origin and development this is not, in fact, so. Dealing, as it does, with the derivation of results in the form of *numbers*, the numerical analyst is really the lineal descendant of the first caveman who enumerated the number of his wives by putting them into one : one correspondence with the fingers of his hand.

Even in its more modern aspects the subject is antique; thus a primary activity of the scientists of Babylon was the construction of mathematical tables. An example is extant, dating from about 2000 B.C., which contains on a tablet the squares of the numbers 1–60. Another tablet records the eclipses going back to 747 B.C., so that astronomical calculation formed a part of the activity of these early numerical analysts.

The ancient Egyptians, too, were energetic numerical analysts. They constructed tables whereby complex fractions could be decomposed into the sum of simpler forms with unit denominators, and invented the method of *false position* (see Chapter 9, section 9.3) for the solution of non-linear algebraic equations.

Passing to the Greek mathematicians we find Archimedes, in about 220 B.C., approximating the value of π and describing it as less than $3\frac{1}{7}$ but greater than $3\frac{10}{71}$. Heron the elder, in about 100 B.C., made use of the iterative process: $\sqrt{a} \sim \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$ which is usually ascribed to Newton, and the Pythagorean school considered the summation of the series $(1 + 2 + 3 \ldots)$. Diophantus, about A.D. 250, apart from his better known work on indeterminate equations, was responsible for a process for the arithmetical solution of quadratic equations.

The Hindus were the creators of our modern arithmetic notation —usually called Arabic—and devised the method of checking the correctness of an arithmetic calculation known as 'casting out nines'.

Mohammed ibn Musa Al-Khowarizmi was the first Arab arithmetician and was responsible, around A.D. 820, for the systematization of computational processes. He gave the value $\pi = 62832/20000$ and was active in the preparation of astronomical tables. Abul Wefa

(A.D. 960) devised a method for the computation of tables of sines and gave the value of sin ($\frac{1}{2}°$) correct to nine decimal places; he also used the *tangent* and calculated a table of this function.

Jumping to the seventeenth century, it is interesting to note that Napier's first table of logarithms was produced before the use of exponents was current, and that his 'logarithm' differs from any in current use since:

$$\text{Naperian } \log x = 10^7 \log_e (10^7/x).$$

In 1614 Napier published his *Mirifici logarithmorum canonis descriptio* and, posthumously, his *Mirifici logarithmorum canonis constructio* in 1619. Briggs, only slightly later in 1624, produced his *Arithmetica logarithmica*, which contains the logarithms, to 14 places of the numbers 1–20,000 and 90,000–100,000. Vlacq produced, in 1628, a table which is still fundamental of the 14-place logarithms of the numbers 1–100,000. The first authoritative publication of the logarithms of trigonometric functions was made at about the same time (1620), by Gunter, who invented the words 'cosine' and 'cotangent', and was responsible for the so-called 'Gunter's chain'.

In the nineteenth century there occurred one of the triumphs of numerical analysis, the simultaneous prediction by Adams and le Verrier in 1845, of the existence and position of the planet Neptune. This century saw also the rise and development of automatic calculating machinery, from the crude desk multiplier of Thomas de Colmar to the almost unmodified Brunsviga of the present day, the Hollerith punched card census calculator, and the difference and analytical engines of Charles Babbage.

Not until the end of the 1930's did the fully automatic calculators begin to come into use, and since the late 1940's there has been a revolution and renaissance in numerical analysis. New methods have been developed and problems which could not previously have been contemplated, even for a life's work, are now solved in hours. It is perhaps dangerous to quote examples, but outstanding achievements are the calculation of π and e to more than 2000 decimals, which took, on the *E.N.I.A.C.*,[1] only about 12 h. The demonstration of the primeness of the Mersenne number $2^{1279} - 1$ on *S.W.A.C.* in 13 min. 25 sec., may also be cited as a noteworthy achievement.

1.2 THE TOOLS OF ANALYSIS (HAND)

From the classical standpoint of the individual numerical analyst the tools of computation are:

(1) Tables of formulae
(2) Tables of function values
(3) A desk calculator
(4) Pencil, paper *and rubber*

Few investigations are of such a fundamental nature that they make no use of existing mathematical knowledge; probably the most common table of formulae in use is a list of integrals. Four standard works may be mentioned:

(1) PEIRCE, B.O., 'A short table of integrals,' Ginn, Boston (1929)
(2) DWIGHT, H. B., 'Tables of integrals and other mathematical data,' Macmillan, New York (1934)
(3) DE HAAN, D. B., 'Nouvelles tables d'intégrales définies, reprinted Stechert, New York (1939)
(4) ' Interpolation and allied tables,' H.M. Stationery Office (1956)

The first two volumes are chiefly concerned with indefinite integrals, and the third exclusively with definite integrals. The last booklet contains most of the useful formulae for interpolation.

Tables of function values are almost too numerous to mention. For 4- or 5-figure accuracy there are the classical:

JAHNKE-EMDE, 'Tafeln höherer Funktionen,' Teubner (4th edn.), Leipzig (1948)

EMDE, 'Tafeln elementarer Funktionen,' Teubner, Leipzig (1940) which, besides giving numerical values, contain useful graphs and formulae. Other moderate accuracy collections are:

'Mathematical Tables from the Handbook of Chemistry and Physics,' Chemical Rubber Publishing Co. Cleveland (1946)

DALE, J. B., 'Five-figure Tables of Mathematical Functions,' Arnold, London (1937)

DWIGHT, H. B., 'Mathematical Tables,' McGraw Hill, New York (1941)

More accurate tables (6 or more decimal digits) are:

'Chambers's Seven-figure Mathematical Tables,' Chambers, London (1937)

'Chambers's Six-figure Mathematical Tables' (2 vols.) (Ed. Comrie) London (1948–9)

'Barlow's Tables of Squares, Cubes and Reciprocals' (Ed. Comrie) Spon (1941)

For more specialist tables reference can be made to the monumental:

FLETCHER, A., MILLER, J. C. P., and ROSENHEAD, L., 'Index of Mathematical Tables,' Scientific Computing Service (1946)

Unfortunately this work is now out of print and, to some extent, out of date. It may be supplemented by reference to 'Mathematical Tables and other Aids to Computation' (*M.T.A.C.*) which maintains a cumulative description of new tables.

Numerous desk calculators are now available but none can be said to possess such excellence as to satisfy all felt wants and to outshine the others. Our experience recommends the Brunsviga and Madas machines in the hand-operated range, and the Marchant, Madas and Mercedes-Euclid amongst those electrically operated. We shall not attempt any description of the means of using these machines, since a short time with an experienced operator *and* the machine will do more than many pages of words in this direction.

Regarding pencil and paper, we may remark that foolscap paper ruled with feint ¼ in. squares seems convenient for most general purposes. Pencils should be soft, B is suitable, and the rubber should not be one degraded by age and use!

1.3 THE TOOLS OF ANALYSIS (AUTOMATIC)

It is hoped that amongst the readers of this book there will be many who have access to one of the automatic digital calculators which become daily more easily available. To these fortunates we would address the following words: do not throw aside the classical tools described in section 1.2. Few problems are of such a nature as to be immediately suitable for an automatic machine and the well-tried aids will almost always be necessary during the period of problem preparation.

Fortunately most of the available automatic machines have 'codes' which are very similar, so that an operator used to one machine can readily apply himself to another. Nevertheless the efficient use of a particular machine will only result from a detailed knowledge of its mathematical structure, and this should always be acquired at the earliest possible moment.

1.4 PRECISION, ACCURACY AND ERRORS

In planning a calculation the three factors detailed in the heading must always be considered. First, in any calculation using data obtained by physical measurement, the inherent precision of the data itself must be examined. If no figures for experimental errors are presented and these are not easily obtainable, a knowledge of the experimental technique may give a clue. Measurements of length are rarely accurate to better than 1/10 per cent, measurements of weight often attain 1/10,000 per cent. Electrical measurements are frequently of precision as low as 5 per cent. These circumstances should be taken into account at the planning stage, and a rough working rule is to calculate to two places more than those given by the data.

The accuracy of the calculation (excluding errors of the careless type) will depend on the numerical process involved. Additions and subtractions neither increase nor decrease the precision of the data; multiplications and divisions, however, lead to round off procedures and thus to an overall decrease in accuracy. Hand calculations are seldom of such length as to cause trouble from the growth of round off errors, but with automatic calculators the situation is different. Thus a typical matrix inversion may lead to over 10,000 multiplications, and this, in turn, to a rounding error which has a probable value of the order of 100 units in the last place.

Errors are of two main types, mathematical and human. The former may result from the use of approximations, which are inevitable consequences of the use of discrete processes to represent continuous ones. The latter class of error should be avoidable, at least in the long run, by the provision of adequate checks.

In manual calculation it has long been a platitude that a person should never check his own work and that, if possible, the same method should not be used. This has tended to become forgotten in connection with the use of automatic digital calculators, and it is frequently suggested that because the machine has produced the same answer twice in succession it must be correct. Unfortunately most machines suffer at times from 'pattern sensitivity', that is they will work with complete accuracy on all numbers except *one*. Under these circumstances the same *wrong* answer can be produced as often as required, and the only valid check is a completely different computing routine.

In practice it is often possible to check the results of a long series of calculations by some completely external means, such as differencing (see Chapter 3, section 3.2), or, alternatively, from a knowledge of the observations with which the calculations are intended to agree. When this is not so (with an automatic digital calculator) a good plan is to repeat the calculation after an interval of several days, since few, if any, of these machines survive such a period without adjustment, and adjustment almost always varies any pattern sensitivity which may be present.

Formulae may be used in different ways and with differing resultant accuracy. Thus, we may calculate

$$\sin\theta \simeq \theta - \theta^3/6 + \theta^5/120 \qquad \ldots.(1.4.1)$$

in two ways. In the first the terms are formed separately and then added. In this event the round off *may* be equal to 1·5 in the last place. On the other hand, by forming:

$$[(\tfrac{1}{120}\theta^2 - \tfrac{1}{6})\theta^2 + 1]\theta \qquad \ldots.(1.4.2)$$

the greatest error will be 0·5 in the last place. This example is also instructive in that it illustrates an efficient means of calculating a polynomial. Thus a direct calculation of equation 1.4.1 involves two divisions, four multiplications and two additions or subtractions, whereas in equation 1.4.2 the multiplications are reduced to three.

Considerations such as those mentioned above should always precede any numerical calculation and might be multiplied indefinitely. Some pointers will be given at appropriate places in the text, but pencil and paper analysis can only be learned by long practice and, in our experience, no two expert analysts agree upon the best detailed layout for any particular case. For this reason the details will be left to the reader and such numerical examples as appear are illustrations of such things as convergence, rather than of computational layout.

REFERENCE

(1) BOOTH, A. D., and BOOTH, K. H. V., 'Automatic Digital Calculators,' p. 14, Butterworths, London, 2nd edn. (1956)

2

TABULATIONS AND DIFFERENCES

2.1 THE NATURE OF TABULATED FUNCTIONS

THE differential calculus had its origin in a consideration of the mode of variation of a function $y = f(x)$ with the augument x. In the process of defining the differential coefficient, $\frac{dy}{dx}$, it is necessary to consider the limit of a finite difference ratio:

$$f(x + \delta x) - f(x) / (x + \delta x) - x$$

as $\delta x \rightarrow 0$. When a function is represented by a set of numerical values contained in a table, it is natural to consider the analogues of the differentials dy and dx which can be deduced immediately from the tabular values. Suppose that a function u_x is defined by a table:

x	u_x
0	u_0
1	u_1
2	u_2
⋮	
n	u_n

Then, corresponding to dx, we have $(1 - 0)$, $(2 - 1)$, $(n - \{n - 1\})$ all of which are equal to unity. And corresponding to dy we have the differences $(u_1 - u_0)$, $(u_2 - u_1)$, . . . $(u_n - u_{n-1})$.

Since the interval of tabulation (*i.e.* 1 in this case) is by no means always unity, it is customary to represent this by the symbol $\Delta(x)$, and in a like manner: $u_{m+1} - u_m$ is represented by $\Delta(u_m)$.

Now just as it is possible to proceed to differential coefficients higher than the first, so can the differences of a tabulated function be extended thus:

x	$f(x)$	Δf	$\Delta^2 f$	$\Delta^3 f$
0	u_0	$\Delta u_0 = u_1 - u_0$	$\Delta^2 u_0 = \Delta u_1 - \Delta u_0 = u_2 - 2u_1 + u_0$	$\Delta^3 u_0 = \Delta^2 u_1 - \Delta^2 u_0 = u_3 - 3u_2 + 3u_1 - u_0$
1	u_1	$\Delta u_1 = u_2 - u_1$	$\Delta^2 u_1 = \Delta u_2 - \Delta u_1 = u_3 - 2u_2 - u_1$	$\Delta^3 u_1 = \Delta^2 u_2 - \Delta^2 u_1 = u_4 - 3u_3 + 3u_2 - u_1$
2	u_2	$\Delta u_2 = u_3 - u_2$	*etc.*	*etc.*
3	u_3	$\Delta u_3 = u_4 - u_3$		
4	u_4	$\Delta u_4 = u_5 - u_4$		
5	u_5	$\Delta u_5 = u_6 - u_5$		

The differences on the first line, namely, Δu_0, $\Delta^2 u_0$, $\Delta^3 u_0$ *etc.* are referred to as leading differences.

2.2 SOME ACTUAL TABLES

Before proceeding to a consideration of the uses to which differences may be put, it is instructive to consider some existing mathematical tables and their relation to differences.

As a first example consider the following section of a typical schoolboy's 4-place logarithm table:

											Proportional				*Parts*				
x	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934		*etc.*							*etc.*					

This is a typical example of what may be called a 'two dimensional' table in that, to each lattice point of a two dimensional co-ordinate system is assigned a functional value. Such tables are usually only available for the most elementary functions and at a very low level of precision. First consider the differences of the function at the finest interval available in the table, namely ·001.

x	$f(x)$	Δ	Δ^2
100	0000	0043	0
101	0043	0043	− 1
102	0086	0042	0
103	0128	0042	
104	0170		

It is intuitively clear that, since the first difference is sensibly constant, a linear relationship exists between the function and its argument in the intervals between tabulated values. It follows that, at any rate approximately, if the value of $f(x)$ at a non-tabulated value is required no great error will result from assuming:

$$f(x_0 + \delta) = f(x_0) + \delta . [f(x_1) - f(x_0)]$$

or

$$f(x_0 + \delta) = f(x_0) + \delta . \Delta [f(x_0)]$$

where δ is a *proportion* of the interval $(x_1 - x_0)$, and in any case the error will not exceed unity [*i.e.* the value of $\Delta^2\{f(x_0)\}$]. On the other hand, consider the table of 'proportional parts'. These purport to contain the values of $\delta\Delta[(f(x_0)]$ *etc.* for $\delta = \cdot 0001, \cdot 0002 \ldots \cdot 0009$. But since:

$$\Delta f(100) = 0043$$
$$\Delta f(109) = 0040$$

it is evident that the proportional parts can only be approximate and that the value at $\delta = \cdot 0009$ should be 39 for $x = 1009$ and 36 for $x = 1099$. More honest table makers would mark these proportional parts to indicate that they are not accurate over the whole range of argument which they cover.

Next consider the differences over a larger interval (01).

x	$f(x)$	Δ	Δ^2	Δ^3
10	0000	0414	− 0036	+ 0005
11	0414	0378	− 0031	+ 0006
12	0792	0347	− 0025	+ 0003
13	1139	0322	− 0022	
14	1461	0300		
15	1761			

It is seen that the first differences are no longer constant so that an attempt to evaluate $f(103)$, say, by the 'proportional part' technique would not be justified (it would lead to $\cdot 3 \times 0414 = 0124$ which does not agree at all well with the actual value 0128), and a more sophisticated technique would be needed to make full use of the 4-place tabular accuracy.

The virtue of the simple table, described above, lies in the ease with which it may be used; when more accurate values are required, however, the two dimensional layout is no longer possible. Thus, to continue further the logarithm table to an accuracy of 7 decimal places, and to have direct reading of a 7-place argument would require a volume of some 2000 pages!

To overcome this difficulty, the onus of calculation of intermediate values is placed upon the user, and the interval of tabulation is so chosen as to make possible the linear interpolation process discussed above. (Linear interpolation is the technical term for the 'proportional part' technique just examined). An example from a 7-decimal place logarithm tabulation is the following:

x	*log x*	*Diff.*
268 11	428 3130	162
12	3292	162
13	3454	162
14	3616	162
15	3778	162
16	3940	162
17	4102	162
18	4264	162
19	4426	162
20	4588	162

At a later position in the compendium from which this example is taken is to be found a table of proportional differences corresponding to all differences encountered in the main table, which in this case range from 434 to 43. A section of this table is:

Diff.	*Prop.* 1	2	3	4	5	6	7	8	9
162	16	32	49	65	81	97	113	130	146
163	16	33	49	65	82	98	114	130	147
164	16	33	49	66	82	98	115	131	148
165	17	33	50	66	83	99	116	132	149

Thus, to find log 2·68143, the user of the table has to form the sum:

$$\begin{aligned}\log 2{\cdot}68\,143 &= \log 2{\cdot}6814 + {\cdot}3 \times {\cdot}000\,0162\\ &= {\cdot}428\,3616 + {\cdot}000\,0049\\ &= {\cdot}428\,3665\end{aligned}$$

If greater precision is required, and careful consideration is needed to justify it in any particular case, the function difference must be actually multiplied by the proportional part. Thus:

$$\begin{aligned}\log 2{\cdot}681\,432 &= {\cdot}428\,3616 + {\cdot}32 \times {\cdot}000\,0162\\ &= {\cdot}428\,3668\end{aligned}$$

It should be noticed that when the ultimate accuracy is required, as in this case, there is a possibility of an error of $\pm$ 1 in the last place due to round off in the original table construction. This effect is eliminated in the so-called 'critical tables' in which the range of x for which $f(x)$ has a given value is specified. These tables are, however, comparatively rare and are unlikely to come the way of the student.

Our final example is taken from a recent[1] table of high precision (15 decimal places) for the trigonometrical function sin $x°$

$x°$	*sin x*	δ^2
17·60	0·30236 98907 50445	− 92 10714
·61	·30253 62492 99766	92 15781
·62	·30270 25986 33306	92 20848
·63	·30286 89387 45998	92 25916
·64	·30303 52696 32774	92 30982
17·65	0·30320 15912 88568	− 92 36048
·66	*etc.*	

It will be noticed that, to this precision, no attempt could be made to subdivide at an interval sufficiently small to make possible linear interpolation. Instead second central differences (see section 3.4 *infra*) have been given and these make possible a reasonable interpolation process (that of Everett, section 3.4, equation 3.4.8) for the evaluation of intermediate values.

REFERENCE

(1) Table of Sines and Cosines to 15 decimal places at hundredths of a degree. *U.S. Nat. Bur. Stand.* Applied Mathematics Series. No. 5, Washington (1949)