A Method for the Reduction of Empirical Multi-Variable Functions

by C. D. Allen

Summary: A method is given for reducing an empirically given function of many variables to an expression involving functions of each variable singly. The number of functions required is minimized for a given accuracy. A method of altering such expressions without losing accuracy is also described. Some notes are given on programming the problem for a digital computer.

INTRODUCTION

When dealing with functions of several variables whose values have been empirically determined, it is often useful to be able to express the function as an algebraic expression in functions of each variable singly, e.g. in the form

$$\phi(x, y, \ldots) = \sum_{r=1}^{n} \phi_r(x)\psi_r(y) \ldots \qquad (1)$$

The problem thus arises of finding values to be given to the $\phi_r(x)$, etc., which are in some sense optimum values, given a table of values of ϕ. One possible criterion for an optimum set is that the values should satisfy (1) to a specified degree of accuracy, with a value of n as low as possible.

For example, in analogue computers, the direct generation of functions of more than one variable is difficult, the degree of difficulty rising rapidly with the number of independent variables. However, many types of function generator have been described (Korn and Korn, 1956) which will multiply a variable input by an empirical function of a single variable, and can thus be arranged as an analogue of the right-hand side of (1). To economize in equipment, the number of functions required should be as small as possible, and it may be desirable to give the functions special properties, such as that of being of one sign only.

It has also been pointed out to the author that an expression such as (1) may be of use in storing tables of functions of many variables in a digital computer, requiring less storage space than the original table. Here also it is desirable to keep the number of functions as low as possible.

An analytic method is described for determining values of the $\phi_r(x)$, etc., for the given values of the variables, from the given values of ϕ. Functions of one variable are found one at a time, in such a way that the errors in equation (1) are minimized by each successive function. The actual errors remaining may be calculated, or simpler criteria may be used to estimate the degree of approximation attained. The corresponding functions of the remaining variables are then calculated. The whole of the computational work may be performed on a digital computer.

The case of two independent variables is described first, and its extension to three or more variables thereafter.

MATRIX FORMULATION OF THE PROBLEM

We now assume that values of the required function $\phi(x, y)$ of two variables are given for all combinations of values:

$$x = x_p, \qquad p = 1 \ldots \alpha,$$
$$y = y_q, \qquad q = 1 \ldots \beta,$$

of the variables, and that these values are sufficient adequately to define the function over the ranges considered. We then require values of the functions $\phi_r(x)$ and $\psi_r(y)$ such that:

$$\phi(x_p, y_q) = \sum_{r=1}^{n} \phi_r(x_p)\psi_r(y_q), \quad \begin{matrix} p = 1 \ldots \alpha \\ q = 1 \ldots \beta, \end{matrix} \qquad (2)$$

to a satisfactory accuracy.

Now write

$$f_{pq} = \phi(x_p, y_q)$$
$$a_{pr} = \phi_r(x_p)$$
$$b_{qr} = \psi_r(y_q)$$

then (2) becomes

$$f_{pq} = \sum_{r=1}^{n} a_{pr}b_{qr}$$

or

$$F = AB' \qquad (3)$$

where F is a (α, β) matrix of the given values of ϕ, A and B are matrices whose columns are sets of values of the required functions. A and B are to have as few columns as possible.

It should be noted here that the rank of a product of two matrices cannot exceed the rank of either factor (Aitken, 1951), hence the rank of AB' cannot exceed n, while the rank of F may well be the lesser of α and β, particularly if the values of $\phi(x, y)$ are empirically determined. Thus for exactness in equation (3), the number of columns in A and B would need to be large, if a large number of values of x and y were used to specify ϕ in detail. However, it is often found possible to approximate satisfactorily to F with a matrix of much lower rank, thus obtaining a good approximation in equation (3) with few columns in A and B.

THE DETERMINATION OF A

We now show that an additional condition

$$B'B = I$$

may be imposed on B without loss of generality. For considering any two columns ψ_l and ψ_m of B we may write

$$\psi_l = a\psi_m + \psi_{l1}$$

where $$\psi_{l1}' \psi_m = 0,$$

and in terms of the functions of y we shall have

$$\psi_l(y_q) = a\psi_m(y_q) + \psi_{l1}(y_q), \quad q = 1 \ldots \beta,$$

and the appropriate terms of equation (2) become

$$\begin{aligned}\phi_l(x_p)[a\psi_m(y_q) + \psi_{l1}(y_q)] + \phi_m(x_p) \cdot \psi_m(y_q) \\ = \phi_l(x_p) \cdot \psi_{l1}(y_q) + [a\phi_l(x_p) + \phi_m(x_p)]\psi_m(y_q) \\ = \phi_l(x_p) \cdot \psi_{l1}(y_q) + \phi_{m1}(x_p) \cdot \psi_m(y_q)\end{aligned}$$

in which no increase in the number of functions has occurred.

Also we may arrange that

$$\psi_r' \cdot \psi_r = 1$$

by extracting suitable scalar factors from the $\psi_r(y)$ and incorporating them in the $\phi_r(x)$. When these modifications are made we obtain matrices A_1 and B_1 of the same size as before, such that

$$A_1 \cdot B_1' = A \cdot B' = F \tag{4}$$

and $$B_1' \cdot B_1 = I. \tag{5}$$

We now have

$$R = F \cdot F' = A_1 \cdot B_1' \cdot B_1 \cdot A_1' = A_1 \cdot A_1' \tag{6}$$

and it is this equation which we solve for A_1 column by column so as to minimize the residuals

$$R - A_1 \cdot A_1' \tag{7}$$

at each stage.

In equation (6) we have effectively shown that R contains the variation in ϕ due to x, but not that due to y. Thus to obtain accuracy in the final expansion, as much as possible of this variation must be accounted for by the matrix A_1. Thus the residuals (7) may be used as a reliable guide to the errors in the final expansion obtained with a given A_1.

Note that R is symmetric of order α.

The residuals (7), after the first column of A_1 has been found, are

$$R_1 = R - \phi_1 \cdot \phi_1'$$

where we write ϕ_r for the rth column of A_1, or

$$r_{ij}' = r_{ij} - a_{ij} \cdot a_{ji}$$

and, for the sum of the squares of these to be a minimum with variation of the a_{ij}, we require

$$\sum a_{ji}(r_{ij} - a_{i1} \cdot a_{j1}) = 0$$

or, writing $$\lambda_1 = \sum_j a_{ji}^2$$

and $$\sqrt{\lambda_1} \cdot V_1 = \phi_1 \tag{8}$$

where $$V_1' \cdot V_1 = 1 \tag{9}$$

we require $$R \cdot V_1 - \lambda_1 V_1 = 0. \tag{10}$$

Now (9) and (10) are the defining equations for a latent root and vector of R, and the appropriate values of $\phi_1(x)$ are given by (8). The residuals are then

$$R_1 = R - \lambda_1 V_1 V_1'$$

and it is obvious from this that λ_1 must be the greatest latent root of R.

Further functions of x may be found from R_1 in a similar fashion. Now it is well known that a symmetric matrix such as R may be expanded as

$$R = \sum_{r=1}^{p} \lambda_r V_r V_r', \tag{11}$$

thus the latent roots and vectors of R_1 are the remaining latent roots and vectors of R. Thus all the required functions of x are of the form

$$\phi_r = \lambda_r V_r$$

where λ_r and V_r are the latent roots and vectors of R.

For exact equality in (11) we must include all the non-zero latent roots of R, and their number is the rank of R. However, it is usually found that the first few largest roots contribute the majority of this expansion, and these only need be retained. The latent roots of a symmetric matrix are all positive and their sum is the sum of the diagonal elements of the matrix (its trace or spur). Thus, as each root and vector is found, an upper bound to the magnitude of the next largest root may be found, and a decision made as to whether it is worth extracting.

If any doubt remains, the residuals

$$R_r = R - \sum_{s=1}^{r} \lambda_r V_r V_r' \tag{12}$$

may be inspected. These usually bear a similar ratio to the elements of R as the eventual errors in (2) do to the values of ϕ. Alternatively, the actual errors in (2) may be calculated at this stage, as will be explained later.

THE DETERMINATION OF B_1

We have now obtained a (α, p) matrix A_1 approximately satisfying (6), and we wish to find B_1 where

$$A_1 B_1' = F$$

to as high an accuracy as possible. This equation represents a set of $\alpha\beta$ simultaneous equations in the $p\beta$ unknowns b_{ij}. These must therefore be normalized to give

$$A_1' A_1 B_1' = A_1' F \tag{13}$$

which represents β sets of p equations in p unknowns, one set being obtained from each column of F with the corresponding row of B_1. The values of the appropriate functions of y are thus obtained as the columns of B_1.

ERRORS DUE TO INSUFFICIENT FUNCTIONS

It is now possible to see how the errors in equation (2) resulting from ignoring some of the latent roots of R may be calculated at that stage. The matrix A_1 there obtained is

$$A_1 = [V_1\sqrt{\lambda_1}, \quad V_2\sqrt{\lambda_2}, \ldots V_p\sqrt{\lambda_p}]$$

and by (13) $\qquad B_1 = (A_1'A_1)^{-1}A_1'F.$

Since the V_r are latent vectors of a matrix we have

$$V_r'V_s = 0 \qquad r \neq s$$
$$= 1 \qquad r = s \text{ by 9}$$

and $\qquad (A_1'A_1) = [\lambda_1, \lambda_2, \ldots \lambda_p]$

a diagonal matrix.* Thus we obtain

$$B_1 = \begin{bmatrix} \lambda_1^{-\frac{1}{2}}V_1' \\ \lambda_2^{-\frac{1}{2}}V_2' \\ \ldots \\ \lambda_p^{-\frac{1}{2}}V_p' \end{bmatrix} F$$

and $$F - A_1B_1' = \left(I - \sum_{r=1}^{p} V_rV_r'\right) . F. \qquad (14)$$

Thus, when the V_r are known, these errors may be calculated. In practice, determination of the errors in full from this equation is laborious, and one of the simpler methods of estimation described above is normally preferable. Equation (14) may, however, be used to sample the errors if this is thought necessary.

TRANSFORMATION OF THE SOLUTION

If T is any (p, p) matrix having an inverse T^{-1}, we may write

$$A_1B_1' = (A_1T^{-1})(TB_1').$$

Thus, if A_1 and B_1 give a solution,

$$A_2 = A_1T^{-1} \qquad (15)$$

and $$B_2 = (TB_1')' \qquad (16)$$

give an equally accurate solution with the same number of functions. This fact may be used in certain cases to modify the functions in a desirable manner.

In this connection it is advantageous to consider the values of the $\phi_r(y)$ as being the co-ordinates in a p-dimensional space of β points on a locus Z, with the parametric equations

$$z_r = \phi_r(y).$$

Equation (16) thus represents some transformation of the co-ordinates z_r in this space, T being the matrix of the transformation. Two-dimensional projections of Z may be obtained by plotting one function against another, and may show that desirable properties of the functions (such as being constant, or being of one sign only) may be achieved by a suitable rotation of the axes.

* This fact may be used as a check on the calculation.

TABLE 1

F VALUES OF GIVEN FUNCTION

y \ x	1	2	3	4	5	6
1	980	780	670	618	468	348
2	822	628	537	484	348	240
3	680	494	408	362	238	138
4	558	372	286	248	124	36
5	430	248	166	126	8	−70

When the degree of rotation has been selected, the appropriate matrix T may be assembled. If three or more functions have been obtained, several such transformations may be made successively.

EXAMPLE

In the example to be considered, a function of three variables, x, y, and z, was to be analysed. A preliminary plot of the function showed that the variation with z at constant x and y was linear, indicating that we could expand the function in the form

$$\phi(x, y, z) = \phi_a(x, y) + z \, . \, \phi_b(x, y)$$

where $\phi_b(x, y)$ is the slope of ϕ against z, which could be read from the graphs, and $\phi_a(x, y)$ is the value of ϕ at $z = 0$. The analysis of ϕ_a and ϕ_b was then undertaken.

The values of $\phi_a(x, y)$ are shown in Table 1, arranged to form the matrix F of this example. Since there are fewer values of y than of x, it was decided to eliminate x first, so the matrix $R = FF'$ was calculated as in Table 2. The entries have been multiplied by 10^{-5}: this arbitrary factor will be recovered automatically in the solution of the equations for B_1.

Two latent roots and vectors were then found from this matrix; these, and the corresponding functions of y, are shown in Table 3. At this stage, since the trace of R is 6·4953, the sum of the remaining latent roots is 0·0001, and these may be ignored. Thus the function values in Table 3 are taken as the matrix A_1, and the six sets of normalized equations

$$A_1'A_1B_1' = A_1'F$$

solved for B_1, which is shown in Table 4. Note that the factor of 10^{-5} has been recovered in these functions.

TABLE 2

THE MATRIX $R = FF'$

y	1	2	3	4	5
1	2·7398	2·2007	1·7082	1·2524	0·7833
2	2·2007	1·7714	1·3794	1·0177	0·6453
3	1·7082	1·3794	1·0796	0·8042	0·5205
4	1·2524	1·0177	0·8042	0·6097	0·4094
5	0·7833	0·6453	0·5205	0·4094	0·2948

TABLE 3

DERIVATION OF $\psi_r(y)$

$\lambda_1 = 6{\cdot}4052$	$\lambda_2 = 0{\cdot}0900$	y	$\psi_1(y)$	$\psi_2(y)$
V_1 0·6521	V_2 −0·4178	1	1·650	−0·1254
0·5256	−0·1343	2	1·330	−0·0403
0·4102	0·1405	3	1·038	0·0422
0·3039	0·4498	4	0·769	0·1349
0·1944	0·7651	5	0·492	0·2296

TABLE 4

VALUES OF THE $\phi_r(x)$ . . . B_1

x	$\phi_1(x)$	$\phi_2(x)$
1	633·6	518·4
2	475·3	53·9
3	397·5	−130·4
4	358·0	−214·8
5	247·0	−489·8
6	160·9	−651·9

TABLE 5

FIRST ERRORS $F - A_1 . B_1'$

y \ x	1	2	3	4	5	6
1	0·5	−2·5	2·2	−0·4	1·0	−0·8
2	−0·2	2·0	−0·1	0·8	0·2	0·3
3	−0·4	1·6	−0·9	0·5	−2·3	1·5
4	−0·8	0·8	2·1	−1·7	−0·1	−0·2
5	0·8	−1·8	−0·4	0·8	1·1	−0·5

TABLE 6

TRANSFORMED FUNCTIONS

x	$\phi_1'(x)$	$\phi_2'(x)$	y	$\psi_1'(y)$	$\psi_2'(y)$
1	393·3	718·0	1	1·577	0·5018
2	420·5	228·0	2	1·248	0·4609
3	417·4	28·0	3	0·947	0·4280
4	412·4	−65·1	4	0·662	0·4131
5	412·5	−361·6	5	0·370	0·3972
6	393·4	−544·2			
mean	408				

TABLE 7

ERRORS FROM TRANSFORMED FUNCTIONS

y \ x	1	2	3	4	5	6
1	21	−26	−15	−10	−14	14
2	22	−13	−10	−3	−16	13
3	15	−9	−8	−1	−8	14
4	9	−10	−5	−5	−7	6
5	6	7	−4	0	−5	0

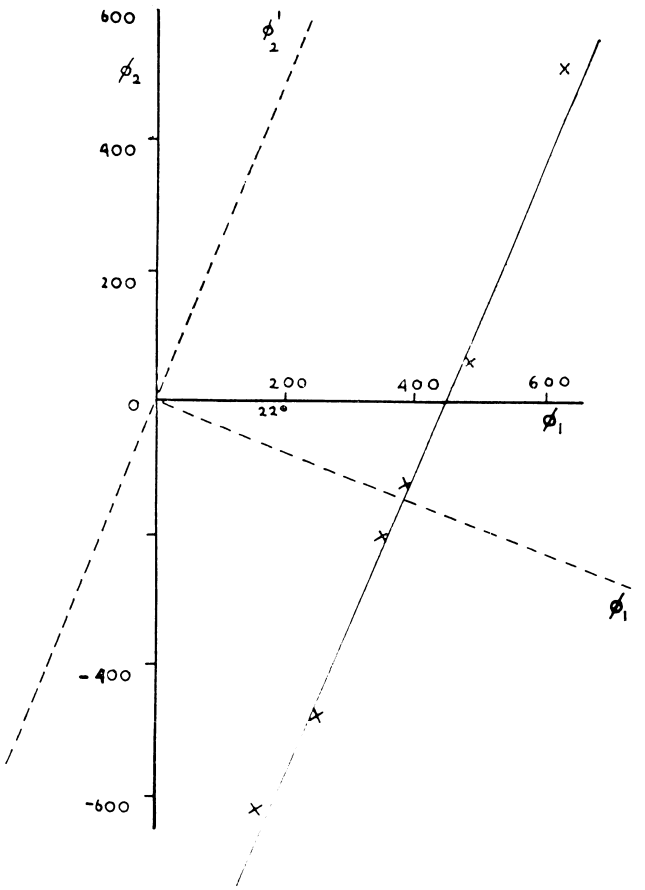

FIG. 1.—Transformation of the functions.

In fact any scalar multiple of the functions of x, such as the latent vectors themselves, could have been used for A_1.

The errors in the expansion thus obtained, viz.

$$\phi_1(x)\psi_1(y) + \phi_2(x)\psi_2(y) - \phi_a(x, y)$$

are shown in Table 5. As can be seen, the maximum error is less than 0·3 per cent of the range of ϕ_a.

A plot of $\phi_1(x)$ against $\phi_2(x)$ is shown as Fig. 1, which shows that the curve is very nearly linear. This figure indicates that a rotation of the axes through 22° in a clockwise direction would make $\phi_1(x)$ a constant. To perform this operation, the matrix B_1' must be premultiplied by the transformation matrix

$$T = \begin{bmatrix} \cos 22^\circ & -\sin 22^\circ \\ \sin 22^\circ & \cos 22^\circ \end{bmatrix}$$

and A_1 must be post-multiplied by T^{-1}. Since T is a rotation matrix whose determinant is unity, its inverse is the same as its transpose, so we post-multiply A_1 by T'.

The results of these operations are shown in Table 6 where it can be seen that $\phi_1'(x)$ varies but little, so a mean value can be taken.

The errors

$$408\psi_1'(y) + \phi_2'(x)\psi_2'(y) - \phi_a(x, y)$$

are shown in Table 7, and are still within 2–3% of the range of ϕ_a.

FUNCTION OF THREE OR MORE VARIABLES

The example above has shown one method of dealing with a function of three variables in the case where the variation of ϕ with one of the variables is linear. Generally this is not so, but the method may readily be extended to cover this case as follows.

The function $\phi(x, y, z)$ is assumed to be defined by its values for certain combination of values of x, y, and z, and nothing is assumed concerning the continuity of the function. We may therefore treat the possible combinations of values of any two variables (y and z, say) as though they were a set of values of a single variable. By eliminating the variation with this single variable first, the case may be treated by the same technique as for two variables. Functions of the variable x will thus be determined first, and the corresponding functions obtained secondly will be functions of y and z jointly in the expansion

$$\phi(x, y, z) = \Sigma \phi_r(x)\psi_r(y, z).$$

These functions of y and z may then be analysed further.

The method may similarly be extended to treat functions of any number of variables.

USE OF DIGITAL COMPUTER

The calculations required in the above method are all of a matrix nature, and therefore repetitive. An electronic computer may therefore be used with advantage. The calculation may be performed in four stages:

(*a*) Calculation of R.
(*b*) Determination of λ_r and V_r.
(*c*) Calculation of B_1 and errors $F - A_1 . B_1'$.
(*d*) Calculation of $A_1 . T^{-1}$ and $T . B'$.

Stages (*a*) and (*c*) are greatly simplified if a matrix interpretive scheme, such as that designed for Pegasus (Hunt, 1956), is available. Stage (*b*) is nowadays a standard library program on most computers.

Stages (*a*) and (*b*) may usually be run concurrently. Stage (*c*) may also be run immediately after stage (*b*) provided that an inspection of the latent roots is sufficient to enable one to decide on the number of functions to be retained (this is normally the case). In this event, it is helpful if the latent roots program is compatible with the matrix interpretive program, in the sense that both programs may be held in the machine simultaneously, and either may use the results of the other without output and re-input. A scheme for setting parameters into a matrix program is also required, so that the number of functions chosen may be set on hand-switches and the program for part (*c*) modified accordingly. Stage (*d*) can only be completed after inspection of plots of one function against another, and selection of a suitable matrix T.

If a satisfactory number of points is used to define ϕ in detail, a moderately large store is required, particularly for functions of more than two variables. The 4,000-word store of Pegasus has been found just adequate for functions of three variables.

ACKNOWLEDGEMENTS

Acknowledgements are due to the management of Air Trainers Link Ltd. for permission to publish this article, to Miss E. M. Whalley for the computing work during the development of the method, and to many of the staff of the Ferranti London Computer Centre for their advice and assistance in programming the problem.

REFERENCES

AITKEN, A. C. (1951). *Determinants and Matrices*, 7th ed. Edinburgh: Oliver and Boyd.
HUNT, P. M. (1956). "The Electronic Digital Computer in Aircraft Structural Analysis," *Aircraft Engineering*, Vol. 28, p. 70.
KORN, G. A., and T. M. (1956). *Electronic Analogue Computers*, 2nd ed. New York: McGraw-Hill Book Co.

Index and Binding of Volume 1

An Index to Volume 1 of The Computer Journal is being prepared and will be available in April 1959. An announcement regarding the binding of members' and subscribers' copies will also be made in April 1959.