A method for minimizing a sum of squares of non-linear functions without calculating derivatives

By M. J. D. Powell*

The minimum of a sum of squares can often be found very efficiently by applying a generalization of the least squares method for solving overdetermined linear simultaneous equations. An original method that has comparable convergence but, unlike the classical procedure, does not require any derivatives is described and discussed in this paper. The number of times the individual terms of the sum of squares have to be calculated is approximately proportional to the number of variables. Finding a solution to a set of fifty non-linear equations in fifty unknowns required the left-hand sides of the equations to be worked out fewer than two hundred times.

1. Introduction

Although the problem of minimizing a sum of squares of non-linear functions occurs frequently in curve fitting, in determining physical parameters from experimental observations, and in solving non-linear simultaneous equations, which may be overdetermined, it appears that this field of research is practically neglected by numerical analysts. This is probably because of the undeniable efficacy of the "generalized least squares" method which, although well known, will be described in Section 2. Most users of this procedure are thankful that only first derivatives of the functions are required, particularly as in many applications the eventual convergence is quadratic. However, it will be shown that a procedure, which is similar to the well-known one, can be devised that has the same convergence properties but does not require any derivatives. In essence it approximates the derivatives by differences, but this is done in such a way that only before the initial iteration are substantially more function evaluations required. The new method is described in Section 3.

In Section 4, three important properties of the method are derived. They are (i) that it is unlikely to have converged before the minimum is reached, (ii) that, in a sense, it chooses conjugate directions of search, and (iii) that it yields a ready approximation to the variance–covariance matrix.

Some numerical examples are provided in section 5. They include an illustration that the procedure to be described may be applicable to fitting problems in which the "best fit" is necessarily poor, although the fast convergence depends on the sum of squares tending to zero.

2. The generalized least squares method

It is required to find $x_1, x_2, \ldots, x_n$, ($\mathbf{x}$, say), to minimize

$$F(\mathbf{x}) = \sum_{k=1}^{m} [f^{(k)}(\mathbf{x})]^2, \qquad m \geqslant n. \tag{1}$$

It is hoped that using a superscript to distinguish the m different functions that appear in the sum of squares will not be found confusing—derivatives will be written out explicitly. In addition the notation

$$g_i^{(k)}(\mathbf{x}) = \frac{\partial}{\partial x_i} f^{(k)}(\mathbf{x}) \tag{2}$$

and

$$G_{ij}^{(k)}(\mathbf{x}) = \frac{\partial^2}{\partial x_i \partial x_j} f^{(k)}(\mathbf{x}) \tag{3}$$

will be used.

The method is iterative, and an iteration requires an approximation $\boldsymbol{\xi}$ to the position of the minimum. If the actual minimum is at $\boldsymbol{\xi} + \boldsymbol{\delta}$ then, by differentiating (1),

$$\sum_{k=1}^{m} g_i^{(k)}(\boldsymbol{\xi} + \boldsymbol{\delta}).f^{(k)}(\boldsymbol{\xi} + \boldsymbol{\delta}) = 0; \quad i = 1, 2, \ldots, n. \tag{4}$$

By approximating the left-hand side of (4) by the first two terms of the Taylor series in $\boldsymbol{\delta}$ about $\boldsymbol{\xi}$, equation (5) is obtained.

$$\sum_{k=1}^{m} \left[g_i^{(k)}(\boldsymbol{\xi}).f^{(k)}(\boldsymbol{\xi}) + \sum_{j=1}^{n} \{G_{ij}^{(k)}(\boldsymbol{\xi})f^{(k)}(\boldsymbol{\xi}) + g_i^{(k)}(\boldsymbol{\xi}).g_j^{(k)}(\boldsymbol{\xi})\}\delta_j \right] \approx 0. \tag{5}$$

The least squares method hinges on the further approximation that the term $G_{ij}^{(k)}(\boldsymbol{\xi})f^{(k)}(\boldsymbol{\xi})$ can be ignored. This term is of order δ if $f^{(k)}(\boldsymbol{\xi})$ is zero at the minimum, and it vanishes if $f^{(k)}$ is linear in the variables. In all other cases the convergence of the procedure will be only linear, the correction to $\boldsymbol{\xi}$ being calculated by solving

$$\sum_{j=1}^{n} \left\{ \sum_{k=1}^{m} g_i^{(k)}(\boldsymbol{\xi}) g_j^{(k)}(\boldsymbol{\xi}) \right\} \delta_j = -\sum_{k=1}^{m} g_i^{(k)}(\boldsymbol{\xi}) f^{(k)}(\boldsymbol{\xi}); \quad i = 1, 2, \ldots, n. \tag{6}$$

Note that the matrix of these equations is in general positive definite so that

$$\left[\frac{\partial}{\partial \lambda} F(\boldsymbol{\xi} + \lambda\boldsymbol{\delta})\right]_{\lambda=0} < 0 \tag{7}$$

unless all the derivatives of $F(\mathbf{x})$ at $\boldsymbol{\xi}$ are zero. Therefore, unless $\boldsymbol{\xi}$ happens to be a stationary point of $F(\mathbf{x})$,

[*] *Applied Mathematics Group, Theoretical Physics Division, A.E.R.E., Harwell, Berks.*

extending the iteration to calculate a positive value of λ, λ_m say, which minimizes $F(\boldsymbol{\xi} + \lambda\boldsymbol{\delta})$, provides a theoretical guarantee that the least squares method will converge. Of course $(\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta})$ is chosen as the new approximation to the minimum. The positive semi-definite case is discussed in Section 4.

If the second derivatives $G_{ij}^{(k)}(\boldsymbol{\xi})$ are not zero, the quadratic convergence depends on the functions $f^{(k)}(\boldsymbol{\xi})$ being of the same order as the correction $\boldsymbol{\delta}$. In this case numerical estimates of the derivatives $g_i^{(k)}(\boldsymbol{\xi})$ in (6) that are in error by $\boldsymbol{\delta}$ are acceptable. On the other hand, if the second derivatives are zero, one expects numerical estimates of the derivatives to be exact. It is for these reasons that the procedure to be described has convergence comparable to the generalized least squares method.

3. The procedure without derivatives

The new method is iterative, and at the start of an iteration n linearly independent directions in the space of the variables, $\boldsymbol{d}(1), \boldsymbol{d}(2), \ldots, \boldsymbol{d}(n)$, say, are required together with estimates of the derivatives of the $f^{(k)}$ along the directions. The notation that will be used for the estimated derivative of the kth function along the ith direction is $\gamma^{(k)}(i)$, so

$$\gamma^{(k)}(i) \approx \sum_{j=1}^{n} g_j^{(k)}(\boldsymbol{x}).d_j(i); \quad i = 1, 2, \ldots, n; \qquad k = 1, 2, \ldots, m. \quad (8)$$

To equilibrate the matrix of equations (11), the directions should be scaled so that

$$\sum_{k=1}^{m} [\gamma^{(k)}(i)]^2 = 1; \quad i = 1, 2, \ldots, n. \quad (9)$$

It is intentional that the notation does not allow for the dependence of $\gamma^{(k)}(i)$ on $\boldsymbol{x}$, because the approximation to the derivative is a number which is calculated when $\boldsymbol{d}(i)$ is chosen. If $\boldsymbol{x}$ is changed by $\boldsymbol{\delta}$, the resultant error in $\gamma^{(k)}(i)$ will be of order $\boldsymbol{\delta}$ multiplied by a second derivative term, and it has been pointed out that this can be tolerated.

As in the least squares method, an approximation to the position of the minimum, $\boldsymbol{\xi}$, is required and a correction to it, $\boldsymbol{\delta}$, is calculated. The correction is worked out by substituting the approximate derivatives in (6) so, if

$$\boldsymbol{\delta} = \sum_{i=1}^{n} q(i).\boldsymbol{d}(i) \quad (10)$$

$$\sum_{=1}^{n} \left\{ \sum_{k=1}^{m} \gamma^{(k)}(i)\gamma^{(k)}(j) \right\} q(j) = -\sum_{k=1}^{m} \gamma^{(k)}(i) f^{(k)}(\boldsymbol{\xi}); \qquad i = 1, 2, \ldots, n. \quad (11)$$

It is convenient to define

$$p(i) = -\sum_{k=1}^{m} \gamma^{(k)}(i) f^{(k)}(\boldsymbol{\xi}). \quad (12)$$

As recommended in Section 2, the iteration is extended to find λ_m to minimize $F(\boldsymbol{\xi} + \lambda\boldsymbol{\delta})$ but, because (8) is an approximation, λ_m is not necessarily positive. The procedure described by Powell (1964) is used for finding the minimum along a line and, at the same time, estimates of the derivatives of the functions $f^{(k)}$ in the direction $\boldsymbol{\delta}$ are worked out in the following way.

The function values $f^{(k)}(\boldsymbol{\xi} + \lambda_1\boldsymbol{\delta})$ and $f^{(k)}(\boldsymbol{\xi} + \lambda_2\boldsymbol{\delta})$, $k = 1, 2, \ldots, m$, which yield the lowest and next lowest values of $F(\boldsymbol{\xi} + \lambda\boldsymbol{\delta})$ are noted. These are differenced to provide the approximation

$$\frac{\partial}{\partial\lambda} f^{(k)}(\boldsymbol{\xi} + \lambda\boldsymbol{\delta}) \approx \frac{f^{(k)}(\boldsymbol{\xi} + \lambda_1\boldsymbol{\delta}) - f^{(k)}(\boldsymbol{\xi} + \lambda_2\boldsymbol{\delta})}{(\lambda_1 - \lambda_2)} = u^{(k)}(\boldsymbol{\delta}). \quad (13)$$

The approximation is improved by

$$v^{(k)}(\boldsymbol{\delta}) = u^{(k)}(\boldsymbol{\delta}) - \mu f^{(k)}(\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta}) \quad (14)$$

where

$$\mu = \sum_{k=1}^{m} [u^{(k)}(\boldsymbol{\delta}).f^{(k)}(\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta})] / \sum_{k=1}^{m} [f^{(k)}(\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta})]^2 \quad (15)$$

because it is known that the derivative of $F(\boldsymbol{x})$ along $\boldsymbol{\delta}$ at $\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta}$ must be zero. Finally $v^{(k)}(\boldsymbol{\delta})$ and $\boldsymbol{\delta}$ are scaled so that the Euclidean norm of the derivative vector is unity, in accordance with (9).

Of course the derivatives along $\boldsymbol{\delta}$ have been calculated in order that $\boldsymbol{\delta}$ may replace one of $\boldsymbol{d}(1), \boldsymbol{d}(2), \ldots, \boldsymbol{d}(n)$. $\boldsymbol{d}(t)$ is replaced, where t is the integer such that

$$|p(t).q(t)| = \max_{1 \leqslant i \leqslant n} |p(i).q(i)|. \quad (16)$$

$\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta}$ replaces the original value of $\boldsymbol{\xi}$, and then the next iteration may be commenced.

An important point to notice is that, apart from calculating function values, the most laborious stage of the iteration is solving the equations (11). After each iteration just one row and one column of the left-hand side matrix are changed so that, if the inverse of the old matrix is stored, that of the new can be worked out by partitioning in an n^2 rather than an n^3 process. The details of this calculation have been set out very clearly by Rosen (1960).

For the first iteration, $\boldsymbol{d}(1), \boldsymbol{d}(2), \ldots, \boldsymbol{d}(n)$ are chosen to be the coordinate directions. A starting value of $\boldsymbol{\xi}$ has to be provided, and then values of $\gamma^{(k)}(i)$ must be worked out. This calculation requires increments $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ to be specified which will yield reasonable estimates of the first derivatives. They are calculated from

$$\gamma^{(k)}(i) = s_i . \frac{f^{(k)}(\xi_1, \xi_2, \ldots, \xi_{i-1}, \xi_i + \epsilon_i, \xi_{i+1}, \ldots, \xi_n) - f^{(k)}\boldsymbol{\xi})}{\epsilon_i} \quad (17)$$

where s_i is a scaling factor, introduced so that (9) may be satisfied. In accordance with (17), for the first iteration,

$$\boldsymbol{d}(i) = (0, 0, \ldots, 0, s_i, 0, \ldots, 0) \quad (18)$$

the only non-zero element being the ith component.

The obvious criterion for ultimate convergence has been found satisfactory, although it is probably not difficult to construct examples in which it fails. It is to stop iterating when both $\boldsymbol{\delta}$ and $\lambda_m\boldsymbol{\delta}$ have acceptably small components.

4. Properties of the method

For this discussion, it is convenient to introduce a notation that has purposely been avoided so far. It is to consider the numbers $f^{(1)}(\boldsymbol{\xi}), f^{(2)}(\boldsymbol{\xi}), \ldots, f^{(m)}(\boldsymbol{\xi})$ as the components of a vector $\boldsymbol{f}(\boldsymbol{\xi})$. Similarly, $\boldsymbol{\gamma}(i)$ will represent $\gamma^{(1)}(i), \gamma^{(2)}(i), \ldots, \gamma^{(m)}(i)$. Therefore, the effect of equation (14) is to ensure that

$$\boldsymbol{f}(\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta}).\boldsymbol{v}(\boldsymbol{\delta}) = 0 \qquad (19)$$

and a necessary condition for $\boldsymbol{\xi}$ to be an approximate minimum of $F(\boldsymbol{x})$ is that

$$\boldsymbol{f}(\boldsymbol{\xi}).\boldsymbol{\gamma}(i) \approx 0; \quad i = 1, 2, \ldots, n. \qquad (20)$$

From (19), it may be seen that after an iteration

$$\boldsymbol{f}(\boldsymbol{\xi}).\boldsymbol{\gamma}(t) = 0 \qquad (21)$$

so at least one of the equations (20) is satisfied. In consequence, on the next iteration, $p(t)$ is zero and, by (16), the direction just defined cannot be replaced, until an iteration is started from a point different from the current $\boldsymbol{\xi}$. Because (16) also ensures that $q(t)$ is not zero, the directions $\boldsymbol{d}(1), \boldsymbol{d}(2), \ldots, \boldsymbol{d}(n)$ remain linearly independent, and these two facts are practically always sufficient to ensure that the procedure will not "stick."

The words "practically always" have been chosen because the possibility that $|p(i).q(i)|$ is zero for all i has not been considered. Unless the matrix of equations (11) is positive semi-definite, it can only occur if all the numbers $p(i)$ are zero, which is the condition for convergence (20). Otherwise, in the semi-definite case, the numbers $q(i)$ are not defined by the equations, and will be calculated to be the components of an eigenvector of the matrix, the eigenvalue being zero. Along such an eigenvector the derivative of each $f^{(k)}(\boldsymbol{x})$ is predicted to be zero, and a search along $\boldsymbol{\delta}$ will correct the predictions. If the derivatives of the individual functions are in fact zero, the position of the minimum is poorly determined, and more functions are required to define it.

The linear approximations behind the least squares method are such that, to first order in $\boldsymbol{\delta}$, equations (20) are satisfied at the predicted position of the minimum, $\boldsymbol{\xi} + \boldsymbol{\delta}$. Therefore, if $\boldsymbol{f}(\boldsymbol{\xi}).\boldsymbol{\gamma}(r) = O(\delta^2)$,

$$\boldsymbol{f}(\boldsymbol{\xi} + \boldsymbol{\delta}).\boldsymbol{\gamma}(r) = O(\delta^2). \qquad (22)$$

Hence, provided λ_m is of order unity,

$$\boldsymbol{f}(\boldsymbol{\xi} + \lambda_m\boldsymbol{\delta}).\boldsymbol{\gamma}(r) = O(\delta^2) \qquad (23)$$

and

$$\boldsymbol{v}(\boldsymbol{\delta}).\boldsymbol{\gamma}(r) = O(\delta). \qquad (24)$$

These equations show that the derivative vector of a new direction is, to order δ, orthogonal to the derivative vectors of those directions that satisfy (20) at the initial point of the iteration. Therefore, remembering (9), the left-hand side matrix of equations (11) tends to the unit matrix. Furthermore, by (8), the successive directions $\boldsymbol{d}$ that are chosen tend to be mutually conjugate (Powell, 1964) with respect to the matrix

$$\Gamma_{ij} = \sum_{k=1}^{m} g_i^{(k)}(\boldsymbol{x})g_j^{(k)}(\boldsymbol{x}). \qquad (25)$$

The last important property of the method is that it provides a ready approximation to the least squares variance–covariance matrix, $\boldsymbol{H}$ say, which, by definition, is the inverse of $\boldsymbol{\Gamma}$. It will now be proved that

$$H_{ij} \approx \sum_{r} d_i(r)d_j(r). \qquad (26)$$

For the proof, it is most convenient to use a matrix notation, and the following definitions are employed:

$$B_{ij} = g_i^{(j)}(\boldsymbol{x}); \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m,$$
$$C_{ij} = \gamma^{(j)}(i); \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, m$$

and

$$D_{ij} = d_i(j); \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, n.$$

Therefore, from (25),

$$\boldsymbol{\Gamma} = \boldsymbol{B}\boldsymbol{B}^T \qquad (27)$$

and it is required to prove that

$$\boldsymbol{\Gamma}^{-1} \approx \boldsymbol{D}\boldsymbol{D}^T. \qquad (28)$$

From (8),

$$\boldsymbol{C} \approx \boldsymbol{D}^T\boldsymbol{B} \qquad (29)$$

and it has just been observed that

$$\boldsymbol{C}\boldsymbol{C}^T \approx \boldsymbol{I} \qquad (30)$$

so

$$\boldsymbol{D}^T\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{D} \approx \boldsymbol{I}. \qquad (31)$$

$\boldsymbol{D}$ has an inverse because the directions $\boldsymbol{d}(i)$ are linearly independent, therefore, from (27) and (31),

$$\boldsymbol{\Gamma} \approx (\boldsymbol{D}^T)^{-1}\boldsymbol{D}^{-1} = (\boldsymbol{D}\boldsymbol{D}^T)^{-1} \qquad (32)$$

which proves (28).

Note that, particularly if the procedure converges in less than n iterations, (30) may be a very crude approximation. This is usually acceptable, because not often are the elements of a variance–covariance matrix required to high accuracy. If they are required more precisely, the following formula, derived from (27) and (29), may be used

$$\boldsymbol{\Gamma}^{-1} \approx \boldsymbol{D}(\boldsymbol{C}\boldsymbol{C}^T)^{-1}\boldsymbol{D}^T. \qquad (33)$$

The elements of $(\boldsymbol{C}\boldsymbol{C}^T)^{-1}$ will have been calculated, if the recommendation of partitioning in Section 3 is heeded.

5. Numerical examples

The method was tested using the trigonometrical equations introduced by Fletcher and Powell (1963).

Table 1

Number of function values to solve equations (34)

n	METHOD			
	LST. SQS.	NEW	DAVIDON	CONJ. DIRN.
3	6	19	—	61
3	5	18	—	84
5	5	24	19	104
5	10	24	23	103
10	5	38	36	329
10	8	34	29	369
20	6	46	89	1519
20	9	65	84	2206
30	12	75	86	—
30	10	61	92	—
50	—	173	169	—
50	—	101	119	—

They are

$$\sum_{j=1}^{n} A_{kj} \sin x_j + B_{kj} \cos x_j = E_k; \quad k = 1, 2, \ldots, m \qquad (34)$$

and a solution is obtained by defining

$$f^{(k)}(x) = \sum_{j=1}^{n} A_{kj} \sin x_j + B_{kj} \cos x_j - E_k. \qquad (35)$$

The elements of A and B are random integers between -100 and $+100$, and the components of the initial value of ξ differ from those of a known solution by up to $\pm 0{\cdot}1\pi$. No difficulties were encountered in obtaining the expected answer from the initial approximation. The number of function values required to calculate $x_1, x_2, \ldots, x_n$ to accuracy 0·0001 is given in Table 1 for values of n ranging from 3 to 50. m was chosen to be equal to n so that a comparison could be made with the results of Fletcher and Powell (1963) and Powell (1964). The number of function values required by the least squares method, as described in Section 2, is also tabulated.

In comparing the columns of the table, it must be remembered that each time function values are required by the least squares method, or by Davidon's method, all the first derivatives must be provided as well. Also, the methods of the last two columns are designed to minimize a general function, and take only $F(x)$ into account when choosing the directions of search. As well as the labour of calculating first derivatives, a further factor may be significant. It is that the number of administrative operations in an iteration of each of the last three methods is of order n^2, while solving equations (6) is an n^3 process.

Table 2

Number of functions to minimize a sum of squares that does not tend to zero

n	δ			
	0	0·1	1	10
3	21	29	31	29
3	16	16	14	21
5	17	17	37	33
5	20	20	29	34
10	29	26	47	78
10	26	29	47	86
20	41	42	118	175
20	35	36	88	93
30	47	59	77	134
30	46	47	106	206

Table 3

The procedure applied to Rosenbrock's function

ITERATION	x_1	x_2	$f^{(1)}$	$f^{(2)}$
0	−1·2000	1·0000	−4·4000	2·2000
1	−1·0770	0·7294	−4·3053	2·0770
2	−0·9759	0·5294	−4·2304	1·9759
3	−0·4205	0·0701	−1·0669	1·4205
4	−0·4270	0·0765	−1·0577	1·4270
5	−0·3573	0·0181	−1·0949	1·3573
6	−0·4232	0·1697	−0·0941	1·4232
7	−0·1620	−0·0048	−0·3110	1·1620
8	0·0380	−0·0419	−0·4336	0·9620
9	0·4193	0·1554	−0·2045	0·5807
10	0·4089	0·1641	−0·0307	0·5911
11	0·6089	0·3465	−0·2420	0·3911
12	0·6770	0·4309	−0·2747	0·3230
13	0·7602	0·5854	0·0741	0·2398
14	0·8207	0·6675	−0·0605	0·1793
15	0·8725	0·7514	−0·0992	0·1275
16	0·9747	0·9514	0·0132	0·0253
17	0·9841	0·9678	−0·0062	0·0159
18	0·9919	0·9827	−0·0110	0·0081
19	0·9986	0·9973	0·0009	0·0014
20	1·0000	1·0000	−0·0001	0·0000

An experiment was tried to find out the effect of applying the method of this paper to a sum of squares of non-linear functions which do not all tend to zero at the minimum. The individual functions are again defined by (35), but m is chosen to be equal to $2n$. The initial value of ξ is chosen as before but, before commencing the iterations, all the values of E_k are changed by random numbers between $-\delta$ and $+\delta$. Again the position of the minimum is found to accuracy 0·0001; the required number of function values is given in Table 2.

The table shows that the method can be very effective even if the individual functions do not tend to zero at the minimum. The number of function values quoted for $\delta = 0$ is less than the corresponding number in Table 1, because the minimum is better determined in the experiment, as there are twice as many functions as variables.

This paper would not be complete without the example showing the effect of the procedure on Rosenbrock's (1960) minimization problem

$$f^{(1)} = 10(x_2 - x_1^2), f^{(2)} = 1 - x_1. \qquad (36)$$

Because there are only two variables, the results of each iteration are given in Table 3. The total number of function values required is 70, and during the iterations the progress can best be described as "lively." Once the corner of the parabolic valley has been turned, the variables increase monotonically to their final values, the eventual convergence being particularly impressive.

As well as being tried on the examples presented, the procedure has been used to solve a number of practical problems at A.E.R.E. It has proved thoroughly successful, and it is particularly encouraging that there appears to be no tendency for the method to become less efficient as the number of variables is increased.

References

Davidon, W. C. (1959). "Variable metric method for minimization," A.E.C. Research and Development Report, ANL-5990 (Rev.).

Fletcher, R., and Powell, M. J. D. (1963). "A rapidly convergent descent method for minimization," *The Computer Journal*, Vol. 6, p. 163.

Powell, M. J. D. (1964). "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, Vol. 7, p. 155.

Rosen, J. B. (1960). "The gradient projection method for nonlinear programming. Part I, Linear Constraints," *Journal of S.I.A.M.*, Vol. 8, p. 181.

Rosenbrock, H. H. (1960). "An automatic method for finding the greatest or least value of a function," *The Computer Journal*, Vol. 3, p. 175.

THE COMPUTER JOURNAL

Published Quarterly by

The British Computer Society, Finsbury Court, Finsbury Pavement, LONDON, E.C.2, England.

Subscription price per volume £3 10s. (U.S. $10). Single Copies £1.

All inquiries should be sent to the Assistant Secretary at the above address.

HONORARY EDITORS

For scientific and engineering papers:

E. N. Mutch, c/o The University Mathematical Laboratory, Corn Exchange Street, CAMBRIDGE.

Associate Editor: R. M. Needham.

For business applications:

H. W. Gearing, c/o The Metal Box Company Limited, Woodside, WORCESTER.

Associate Editor: L. R. Crawley.