Function minimization without evaluating derivatives—a review

By R. Fletcher*

The efficiency of methods for minimizing functions without evaluating derivatives is considered, with particular regard to three methods recently developed. A set of test functions representative of a wide range of minimization problems is proposed and is used as a basis for comparison.

The problem of minimizing a function $f(\boldsymbol{x})$ of n variables $\boldsymbol{x} = (x_1, x_2, \ldots x_n)$ from a given approximation to the minimum $\boldsymbol{x}_0$, has received considerable attention in recent years. In particular two separate problems can be distinguished—functions for which both the function f and the first derivatives or gradient $\partial f/\partial x_i$ can be evaluated at any given point $\boldsymbol{x}$, and functions for which only f can be evaluated. Although satisfactory methods have been given by Fletcher and Powell (1963), and by Fletcher and Reeves (1964) for solving the first of these problems, the situation with regard to the latter problem is less clear.

Historically it was found that the simplest concepts, those of tabulation, random search, or that of improving each variable in turn, were hopelessly inefficient and often unreliable. Improved methods were soon devised such as the Simplex method of Himsworth, Spendley and Hext (1962), the "pattern search" method of Hooke and Jeeves (1959), and a method due to Rosenbrock (1960). Both the latter methods have been widely used, that of Rosenbrock being probably the most efficient. However, all these methods rely on an *ad hoc* rather than a theoretical approach to the problem. Developments of gradient methods of minimization meanwhile were showing the value of iterative procedures based on properties of a quadratic function. In particular the most efficient methods involved successive linear minimizations along so-called "conjugate directions" generated as the minimization proceeded. An explanation of these terms is given in Fletcher and Reeves (1964).

Two methods involving these concepts have recently been introduced, by Smith (1962) and by Powell (1964). An improvement of Rosenbrock's method to include linear minimizations has also been developed by Davies, Swann and Campey (Swann (1964)). A short description of the basic features of each method is given in the next Section: reference to the source papers should furnish any additional details required. Although these methods represent an advance in the theory of minimization, little is known of how the methods compare amongst themselves for efficiency. This paper sets out to make this comparison, not only for irregular functions designed to prove difficult to minimize, but also for regular functions more likely to occur in practice. The efficiency of the procedures as the number of variables is increased is also of interest. With these aims in mind, a set of test functions is proposed by which this comparison is made. It is hoped that a scheme of test functions of this nature will provide a standard by which the efficiency of future methods can be compared. Certainly conclusions based on results from one or two functions with only a few variables may be somewhat suspect.

The methods

All the methods are iterative and locate the minimum by successive linear minimizations from an initial point $\boldsymbol{x}_0$ along directions $\boldsymbol{p}_i$ generated by the procedure and initially chosen as the co-ordinate directions. That is

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$$

where α_i is chosen so that $f(\boldsymbol{x}_{i+1})$ is a minimum along the direction $\boldsymbol{p}_i$ through the point $\boldsymbol{x}_i$. (This was not strictly true of the original version of Smith's procedure, but is so of an amended procedure developed by the author and used in this comparison.) The methods used to accomplish the linear minimization were similar in each case and are described towards the end of this Section. A short discussion of the convergence criterion used in each case to terminate the iteration is also included here.

(i) *Davies, Swann and Campey (D.S.C.)*

This method is in essence an application of linear minimizations to the Rosenbrock method. Orthonormal directions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots \boldsymbol{p}_n$ are chosen and n linear minimizations are made as above. Vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots \boldsymbol{q}_n$ are then chosen so that

$$\begin{aligned}
\boldsymbol{q}_1 &= \alpha_1 \boldsymbol{p}_1 + \alpha_2 \boldsymbol{p}_2 \ldots + \alpha_n \boldsymbol{p}_n \\
\boldsymbol{q}_2 &= \qquad\quad\;\; \alpha_2 \boldsymbol{p}_2 \ldots + \alpha_n \boldsymbol{p}_n \\
&\;\;\vdots \\
\boldsymbol{q}_n &= \qquad\qquad\qquad\qquad\;\; \alpha_n \boldsymbol{p}_n
\end{aligned}$$

and are orthonormalized by the Schmidt process. These become the new $\boldsymbol{p}_1 \ldots \boldsymbol{p}_n$ for the next iteration. In practice, if the total progress made in each direction in an iteration is less than the step length used in the linear minimization, then this step length is reduced. In this case directions are not replaced, but an extra linear minimization is made along $\boldsymbol{q}_1$. If no progress is made along one particular direction, then this direction is not included in the orthogonalization in order to preserve linear independence.

* *Electronic Computing Laboratory, University of Leeds.*

It is stated by Smith (1962) that when using Rosenbrock's method on quadratic functions, the directions $\boldsymbol{p}_i$ align themselves in the limit along the axes of the function (the eigenvalues of the matrix of second derivatives—a particular case of conjugate directions), and presumably this can be shown for this method. Hence, although the method does not have the property of quadratic convergence, it does have some features in common.

(ii) *Powell's Method*

This method depends upon the properties of conjugate directions defined by a quadratic function. Linearly independent directions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots \boldsymbol{p}_n$ are chosen and a basic iteration (in pseudo-ALGOL) is

$\boldsymbol{y} := \boldsymbol{x}$;
for $i := 1$ **step** 1 **until** n **do** $MIN(i)$;
for $i := 1$ **step** 1 **until** $n - 1$ **do** $\boldsymbol{p}_i := \boldsymbol{p}_{i+1}$;
$\boldsymbol{p}_n := \boldsymbol{y} - \boldsymbol{x}$; $MIN(n)$

where $MIN(i)$ is a procedure which advances the current best approximation $\boldsymbol{x}$, to the minimum along the direction $\boldsymbol{p}_i$. The vector $\boldsymbol{y}$ holds the value of $\boldsymbol{x}$ at the start of the iteration and is usually a minimum along $\boldsymbol{p}_n$. At the end of the iteration the direction $\boldsymbol{p}_{n+1} = \boldsymbol{y} - \boldsymbol{x}$ is then conjugate to $\boldsymbol{p}_n$. The directions are reordered so that $\boldsymbol{p}_i$ is replaced by $\boldsymbol{p}_{i+1}$, and $\boldsymbol{p}_1$ is rejected. After n repetitions of this iteration, the minimum of an n-dimensional quadratic function would be located. In practice, for non-quadratic functions, instability in the form of linear dependence of the $\boldsymbol{p}_i$ can set in. This has resulted in a more sophisticated basic iteration in which rejection of one direction in favour of another is only carried out if it causes an increase in the determinant of the transformation matrix which relates the vectors $\boldsymbol{p}_i$ and any set of conjugate directions $\boldsymbol{\eta}_i$ (both suitably scaled). The implications of this interesting condition are discussed by Powell.

(iii) *Smith's Method* (*as modified*)

This method also depends upon the properties of conjugate directions defined by a quadratic function. Orthonormal directions $\boldsymbol{p}_1 \ldots \boldsymbol{p}_n$ are chosen and the basis of an iteration is

$MIN(1)$;
for $i := 2$ **step** 1 **until** n **do**
begin $\boldsymbol{y} := \boldsymbol{x}$; $MIN(i)$;
  for $j := 1$ **step** 1 **until** $i - 1$ **do** $MIN(j)$;
  $\boldsymbol{p}_i := \boldsymbol{y} - \boldsymbol{x}$; $MIN(i)$
end

This differs from Smith's original method in that optimum rather than arbitrary displacements are made. Furthermore, after each iteration, the $\boldsymbol{p}_i$ are rearranged cyclically so that $\boldsymbol{p}_1 = \boldsymbol{p}_n$, and then orthonormalized. This ensures that individual variables do not get unequal treatment, and also saves a linear minimization in subsequent iterations. To locate the minimum of a quadratic function, the method requires $(n-1)(n+4)/2$ linear minimizations as against the original figure of $n(n+1)/2$.

The linear minimization

It is accepted that as none of the methods claims finite convergence for non-quadratic functions, an acceptable estimate of the minimum along a line is given by that of the quadratic passing through three points along the line at which f has been evaluated. In the method of Davies, Swann and Campey, and also that of Smith it is required that these three points should bracket the minimum. Powell only requires that the predicted step should be less than some preassigned maximum. All the methods ensure stable and efficient convergence, however, and differ essentially in details only. Powell, however, introduces the valuable idea that the second derivative of f along a direction can be used whenever future minimizations along that line are made, so reducing the number of function evaluations required. Such a feature can be used in any method where the same direction is used often. All the methods require a step length to be chosen initially for the linear minimization. The results quoted do not involve any optimization of this factor, a natural selection being made in each case.

The convergence criteria

All methods used different convergence criteria, the more stringent the criterion, so the more function evaluations required to satisfy it. Davies, Swann and Campey have the simplest one in which the procedure is terminated when the step-length (automatically reduced during the iteration) becomes less than the accuracy required. This was sufficient in all cases except Powell's function of 4 variables, a particularly stringent test. The natural method of iterating until two estimates agreed to given accuracy, as used in Smith's method, also failed in this case. Powell allowed the user to select, as an alternative to this, another very safe but lengthy procedure.

Test functions

Seven test functions in all were taken. Three are already well known and are designed to prove difficult to minimize. These are

(i) A parabolic valley (Rosenbrock (1960))
$$f = 100\,(x_2 - x_1^2)^2 + (1 - x_1)^2$$
$$\boldsymbol{x}_0 = (-1\cdot 2,\ 1\cdot 0)$$
(ii) A helical valley (Fletcher and Powell (1963))
$$f = 100[(x_3 - 10\,\theta)^2 + (r-1)^2] + x_3^2$$

where $x_1 = r\cos 2\pi\theta$, $x_2 = r\sin 2\pi\theta$, $\boldsymbol{x}_0 = (-1, 0, 0)$ functions of two and three variables, respectively, with steep slopes to a curved valley, and

(iii) A function of 4 variables (Powell (1962))
$$f = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10\,(x_1 - x_4)^4$$
$$\boldsymbol{x}_0 = (3, -1, 0, 1),$$

a function whose matrix of second derivatives becomes singular at the minimum.

A new test function is introduced representative of the more regular type of function also obtained in practice. It has the advantage that the number of variables can be changed readily. A convenient initial approximation x_0 is also readily available for any n. This new function is called "Chebyquad" and is considered here for $n = 2, 4, 6$ and 8. A description of the function together with an ALGOL procedure is given in the Appendix.

The results of minimizing these test functions by each of the three methods are given in **Tables 1–7.** After each iteration the difference between the function and its value at the minimum is given, together with the cumulative number of function evaluations and linear minimizations. In the case of Powell's method the latter figure (n or $n + 1$) was not available explicitly, but was estimated from the number of function evaluations required at each iteration.

Discussion

In order to eliminate the effect of differences in the algorithms for linear minimization, methods have been compared primarily in terms of this factor rather than the number of function evaluations. Graphs of $\log_{10}(f - f_{min})$ against the number of linear minimizations have been prepared for each test function, and together with the tables will provide the basis of the discussion. (see **Figs. 1–7.**)

The situation with regard to Smith's method is fairly clear. When the number of variables n is small (2, 3, 4), then the method is acceptable, although noticeably inferior to the other methods. As n increases, however, the method rapidly becomes unworkable. This situation arises from the fact that at each iteration, many of the linear minimizations are made repeatedly in limited subspaces, permitting of only restricted progress to the minimum. The complete space is thus only covered after $n(n + 1)/2$ linear minimizations ($(n - 1)(n + 4)/2$ as amended) whereas it would seem vital that it is covered at every n minimizations. The cause of inefficiency is the same as that in Powell's (1962) early gradient method for minimization. Some improvement could be made if the second derivative were used to reduce the number of function evaluations required in the linear minimization. This, however, would not remove the basic cause of inefficiency.

On the basis of function evaluations the most efficient method is certainly that of Powell. This arises from the repeated use of the same directions, permitting the second derivative to be used in the interpolation. Comparing by linear minimizations, the most noticeable feature is the rapid convergence near the minimum. In fact, for functions of a few variables the method compares advantageously with both others (except on Powell's function of 4 variables, on account of the singular behaviour at the minimum). However, as the number of variables increases, the comparison with the D.S.C. method becomes less favourable. Powell himself observes that as the number of variables is increased, there is a tendency for new directions to be chosen less often, and it seems likely that these factors are related. The criterion which determines when a new direction is to be chosen is such that the directions retained never proceed towards linear dependence (as measured by the appropriate determinant). It could be that this criterion is too stringent, and that an alternative should be found. Possibly some lower limit on this determinant could be fixed; new directions to be chosen unless the limit be violated. It would certainly seem that if this point can be solved, and convergence for large numbers of variables can be improved, then Powell's method will be the most powerful for the general solution of the problem.

The method of Davies, Swann and Campey is certainly a simple and effective method of minimization, permitting a convenient choice of convergence criterion, and showing up well with larger numbers of variables. On the other hand, as directions are only used once, the second derivative cannot be used directly to save function evaluations. It is also difficult to suggest any way in which convergence can be appreciably improved. It would seem that the extent to which it is successful as a general method will depend upon what improvements, if any, can be made to Powell's method.

The desirability or not of quadratic convergence has already caused many arguments. Certainly the advantage of the D.S.C. method is most marked when the minimum cannot be represented adequately by a quadratic. However, such situations rarely occur in practice, the only case to my knowledge being at a non-zero minimum, when attempting to solve non-linear equations (see the example in Freudenstein and Roth (1963)). As against this the rate of convergence of Powell's method near the minimum is striking, this being a feature of methods with quadratic convergence. In the region remote from the minimum, complex situations occur, in particular the presence of narrow curving valleys. Methods with quadratic convergence are successful in generating good directions, inasmuch as they take into account the local curvature of the function in these regions. However, the Rosenbrock—D.S.C. approach of attempting to align the directions of search along the axes of the valley is equally valid, and no doubt there are other ways in which this problem can be tackled.

To construct a program in which different methods are used in different regions would be extremely clumsy and would introduce difficulties in the choice of change-over point. In choosing a "best buy" therefore, we want a reliable method which performs efficiently in as many situations as possible. The real question then, lies in the nature of any advantage which the D.S.C. method may have when the number of variables is large. Is this due to disadvantages associated with the general properties of conjugate gradients, or rather to the particular over-stringent criterion used by Powell for accepting new directions? Experience with gradient methods has shown that whilst there are many ways of generating conjugate directions of widely differing efficiency, the best of these has proved to be the best in general. I am inclined to think that this will be the case here.

Acknowledgements

I am very grateful to Dr. D. Davies and Mr. W. H. Swann of I.C.I. Ltd., and equally to Mr. M. J. D. Powell of A.E.R.E. Harwell, for their ready help in providing some of the results quoted in this paper. I am also indebted to Dr. C. M. Reeves who first suggested the test function "Chebyquad".

Appendix

The new function, Chebyquad, arises as follows. With Chebyshev (equal weights) n-point quadrature, the integral $\int_0^1 F(x)dx$ is represented by the sum

$$\frac{1}{n}\sum_{i=1}^{n} F(x_i).$$

The abcissae $x = (x_1, x_2, \ldots x_n)$ in the range $0 \leqslant x_i \leqslant 1$ can be determined from the condition that if F is a polynomial of degree n or less then the above representation is exact. In particular, for arbitrary x, we can define the residual Δ_i as the difference between the integral and the sum when F is a polynomial of degree i. Choosing the shifted Chebyshev polynomial T_i we define

$$\Delta_i(x) = \int_0^1 T_i(x)dx - \frac{1}{n}\sum_{j=1}^{n} T_i(x_j).$$

Then the function

$$f(x) = \sum_{i=1}^{n} (\Delta_i(x))^2$$

has the property that if x is the vector of abcissae, then $f = 0$, otherwise $f > 0$. Hence we can determine the abcissae for any n, by estimating x and minimizing $f(x)$ from this point. This is the basis of the ALGOL procedure CHEBYQUAD given below. Convenient initial estimates of the x_i are at equal intervals in the range, that is $x_i = i/(n + 1)$. In reality the quadrature formula is only accurate (in the sense that a zero minimum of f exists) for $n = 1(1)7$ and 9 (Krylov (1962), page 191). However, the minimization problem is valid for all values of n.

```
procedure CHEBYQUAD (f, x);
real f; array x;
comment It is assumed that CHEBYQUAD is declared
      within the scope of the global identifiers n and count;
begin integer i, j; real delta, Ti plus; Boolean i even;
   real array y, Ti, Ti minus [1 : n];
   delta := 0;
   for j := 1 step 1 until n do
   begin y [j] := 2 × x [j] − 1;
        delta := delta + y[j];
        Ti[j] := y[j]; Ti minus [j] := 1
   end;
   f := delta × delta; i even := false;
   for i := 2 step 1 until n do
   begin i even := not i even;
        delta := 0;
        for j := 1 step 1 until n do
        begin Ti plus := 2 × y[j] × Ti[j] − Ti minus [j];
              delta := delta + Ti plus;
              Ti minus [j] := Ti [j];
              Ti [j] := Ti plus;
        end;
        delta := delta/n−(if i even then − 1/(i × i − 1)
              else 0);
        f := f + delta × delta
   end;
   count := count + 1
end of CHEBYQUAD;
```

References

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

FLETCHER, R., and REEVES, C. M. (1964). "Function minimization by conjugate gradients," *The Computer Journal*, Vol. 7, p. 149.

FREUDENSTEIN, F., and ROTH, B. (1963). "Numerical solution of systems of non-linear equations," *Journal of the A.C.M.*, Vol. 10, p. 550.

HIMSWORTH, F. R., SPENDLEY, W., and HEXT, G. R. (1962). "The sequential application of simplex designs in optimization and evolutionary operation," *Technometrics*, Vol. 4, p. 441.

HOOKE, R., and JEEVES, F. A. (1959). "Direct Search Solution of Numerical and Statistical Problems," *Journal of the A.C.M.*, Vol. 8, p. 212.

KRYLOV, V., trans. STROUD, A. H. (1962). *Approximate calculation of integrals*, London: Macmillan.

POWELL, M. J. D. (1962). "An iterative method for finding stationary values of a function of several variables," *The Computer Journal*, Vol. 5, p. 147.

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

(*references concluded on p.* 41)

Table 1

A parabolic valley

ITN.	D.S.C.			POWELL			SMITH		
	F	FNS.	L.M.	F	FNS.	L.M.	F	FNS.	L.M.
0	$2{\cdot}4_{10^{1}}$	1	0	$2{\cdot}4_{10^{1}}$	1	0	$2{\cdot}4_{10^{1}}$	1	0
1	$3{\cdot}9_{10^{0}}$	12	3	$4{\cdot}0_{10^{0}}$	13	3	$3{\cdot}8_{10^{0}}$	14	4
2	$2{\cdot}8_{10^{0}}$	21	5	$3{\cdot}3_{10^{0}}$	25	6	$3{\cdot}3_{10^{0}}$	28	7
3	$2{\cdot}3_{10^{0}}$	28	7	$2{\cdot}6_{10^{0}}$	35	9	$2{\cdot}2_{10^{0}}$	44	10
4	$1{\cdot}8_{10^{0}}$	36	9	$1{\cdot}9_{10^{0}}$	46	12	$1{\cdot}2_{10^{0}}$	55	13
5	$1{\cdot}4_{10^{0}}$	43	11	$1{\cdot}1_{10^{0}}$	58	15	$9{\cdot}6_{10^{-1}}$	66	16
6	$1{\cdot}1_{10^{0}}$	50	13	$5{\cdot}6_{10^{-1}}$	73	18	$6{\cdot}0_{10^{-1}}$	77	19
7	$5{\cdot}1_{10^{-1}}$	58	15	$3{\cdot}4_{10^{-1}}$	86	21	$5{\cdot}4_{10^{-1}}$	98	22
8	$5{\cdot}1_{10^{-1}}$	64	17	$1{\cdot}2_{10^{-1}}$	98	24	$4{\cdot}1_{10^{-1}}$	110	25
9	$2{\cdot}0_{10^{-1}}$	80	20	$7{\cdot}4_{10^{-3}}$	113	27	$2{\cdot}2_{10^{-1}}$	119	28
10	$1{\cdot}3_{10^{-1}}$	91	22	$5{\cdot}4_{10^{-3}}$	122	29	$4{\cdot}1_{10^{-2}}$	130	31
11	$8{\cdot}8_{10^{-2}}$	101	24	$8{\cdot}2_{10^{-5}}$	134	32	$4{\cdot}0_{10^{-2}}$	148	34
12	$5{\cdot}3_{10^{-2}}$	112	26	$4{\cdot}4_{10^{-9}}$	145	35	$1{\cdot}2_{10^{-2}}$	164	37
13	$3{\cdot}0_{10^{-2}}$	121	28	$6{\cdot}9_{10^{-12}}$	153	37	$1{\cdot}4_{10^{-3}}$	171	40
14	$1{\cdot}4_{10^{-2}}$	130	30	$1{\cdot}3_{10^{-16}}$	158	39	$5{\cdot}4_{10^{-5}}$	184	43
15	$5{\cdot}6_{10^{-3}}$	139	32				$3{\cdot}1_{10^{-6}}$	195	46
16	$1{\cdot}5_{10^{-3}}$	148	34				$4{\cdot}0_{10^{-8}}$	206	49
17	$1{\cdot}9_{10^{-4}}$	156	36				$4{\cdot}6_{10^{-9}}$	215	52
18	$3{\cdot}3_{10^{-6}}$	163	38				$1{\cdot}4_{10^{-14}}$	223	55
19	$3{\cdot}7_{10^{-7}}$	169	40				$3{\cdot}6_{10^{-16}}$	234	58
20	$7{\cdot}0_{10^{-10}}$	178	43						
21	$1{\cdot}5_{10^{-12}}$	187	46						

Table 2

A helical valley

ITN.	D.S.C.			POWELL			SMITH		
	F	FNS.	L.M.	F	FNS.	L.M.	F	FNS.	L.M.
0	$2{\cdot}5_{10^{3}}$	1	0	$2{\cdot}5_{10^{3}}$	1	0	$2{\cdot}5_{10^{3}}$	1	0
1	$1{\cdot}4_{10^{2}}$	20	4	$1{\cdot}4_{10^{2}}$	29	4	$1{\cdot}4_{10^{1}}$	43	8
2	$1{\cdot}2_{10^{1}}$	33	7	$1{\cdot}1_{10^{1}}$	41	8	$6{\cdot}9_{10^{0}}$	65	15
3	$1{\cdot}1_{10^{1}}$	42	10	$1{\cdot}1_{10^{1}}$	51	12	$4{\cdot}6_{10^{0}}$	92	22
4	$1{\cdot}1_{10^{1}}$	55	14	$6{\cdot}6_{10^{0}}$	63	16	$1{\cdot}8_{10^{0}}$	118	29
5	$9{\cdot}7_{10^{0}}$	69	17	$5{\cdot}2_{10^{0}}$	74	20	$1{\cdot}4_{10^{0}}$	145	36
6	$7{\cdot}6_{10^{0}}$	84	20	$2{\cdot}9_{10^{0}}$	86	24	$6{\cdot}7_{10^{-1}}$	176	43
7	$6{\cdot}3_{10^{0}}$	97	23	$1{\cdot}5_{10^{0}}$	98	28	$7{\cdot}0_{10^{-2}}$	199	50
8	$2{\cdot}4_{10^{0}}$	111	26	$6{\cdot}9_{10^{-1}}$	109	32	$6{\cdot}5_{10^{-3}}$	227	57
9	$1{\cdot}9_{10^{0}}$	125	29	$1{\cdot}5_{10^{-1}}$	121	36	$1{\cdot}5_{10^{-3}}$	250	64
10	$1{\cdot}4_{10^{0}}$	138	32	$8{\cdot}6_{10^{-2}}$	133	40	$2{\cdot}6_{10^{-5}}$	280	71
11	$6{\cdot}9_{10^{-1}}$	152	35	$3{\cdot}3_{10^{-3}}$	146	44	$3{\cdot}2_{10^{-6}}$	303	78
12	$4{\cdot}1_{10^{-1}}$	165	38	$1{\cdot}0_{10^{-5}}$	155	47	$1{\cdot}9_{10^{-8}}$	324	87
13	$9{\cdot}3_{10^{-2}}$	178	41	$1{\cdot}4_{10^{-8}}$	166	50	$1{\cdot}3_{10^{-9}}$	346	94
14	$5{\cdot}6_{10^{-2}}$	188	44	$1{\cdot}8_{10^{-10}}$	173	53	$1{\cdot}5_{10^{-12}}$	365	101
15	$1{\cdot}2_{10^{-2}}$	199	47	$2{\cdot}1_{10^{-12}}$	180	56			
16	$1{\cdot}0_{10^{-4}}$	209	50						
17	$9{\cdot}4_{10^{-7}}$	218	53						
18	$2{\cdot}5_{10^{-8}}$	230	57						
19	$2{\cdot}6_{10^{-10}}$	242	61						
20	$2{\cdot}3_{10^{-12}}$	254	65						
21	$2{\cdot}1_{10^{-14}}$	266	69						

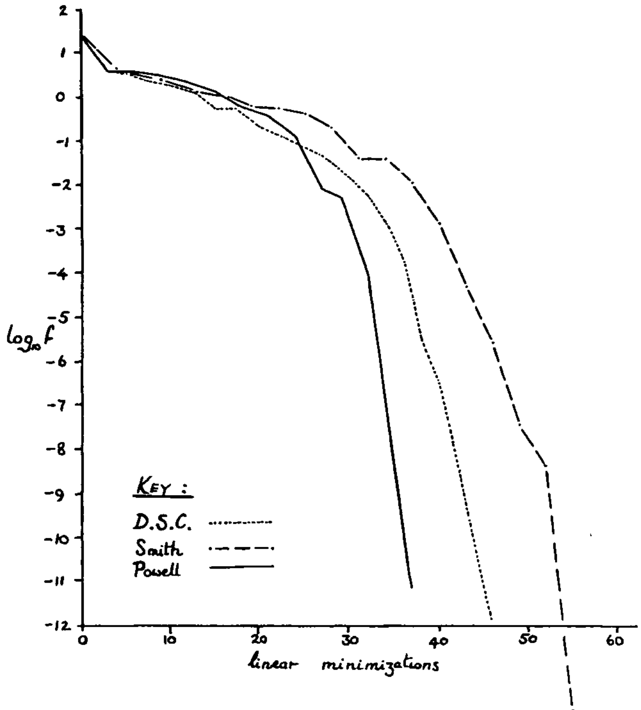

Fig. 1.—A parabolic valley

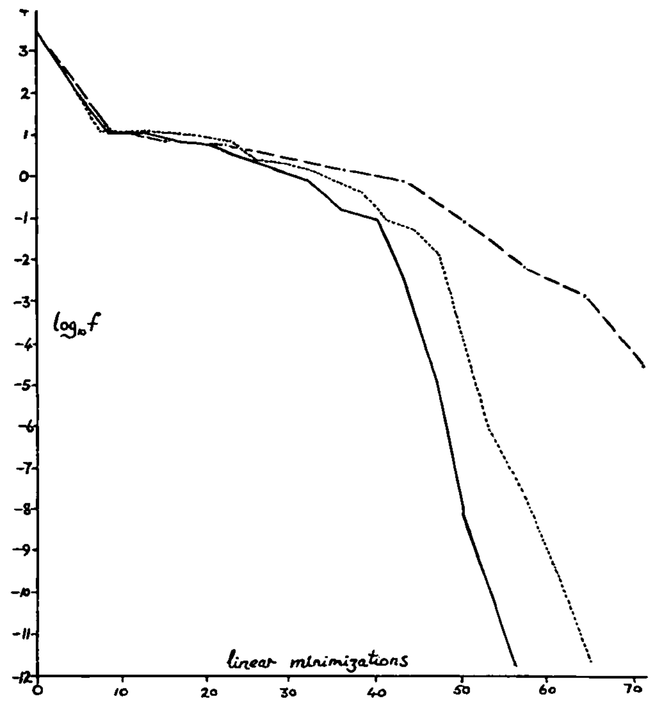

Fig. 2.—A helical valley

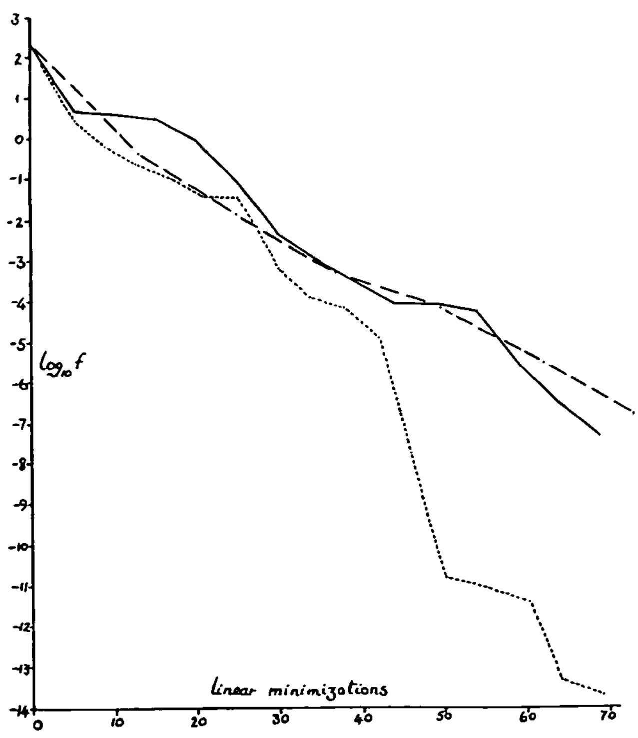

Fig. 3.—Powell's function of 4 variables

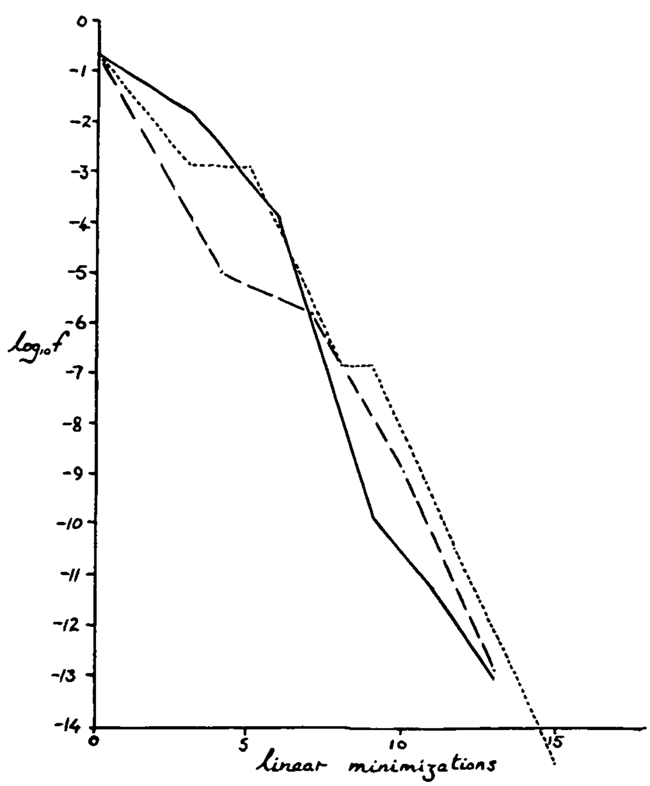

Fig. 4.—Chebyquad $n = 2$

Table 3

Powell's function of 4 variables

ITN.	D.S.C.			POWELL			SMITH		
	F	FNS.	L.M.	F	FNS.	L.M.	F	FNS.	L.M.
0	$2\cdot2_{10^{2}}$	1	0	$2\cdot2_{10^{2}}$	1	0	$2\cdot2_{10^{2}}$	1	0
1	$2\cdot8_{10^{0}}$	27	5	$4\cdot7_{10^{0}}$	26	5	$4\cdot1_{10^{-1}}$	51	13
2	$6\cdot3_{10^{-1}}$	45	9	$3\cdot8_{10^{0}}$	41	10	$1\cdot4_{10^{-2}}$	101	25
3	$2\cdot3_{10^{-1}}$	61	13	$2\cdot9_{10^{0}}$	57	15	$5\cdot0_{10^{-4}}$	144	37
4	$1\cdot0_{10^{-1}}$	75	17	$8\cdot9_{10^{-1}}$	72	20	$8\cdot4_{10^{-5}}$	194	49
5	$3\cdot9_{10^{-2}}$	88	21	$7\cdot4_{10^{-2}}$	86	25	$4\cdot7_{10^{-6}}$	230	61
6	$3\cdot6_{10^{-2}}$	100	25	$4\cdot3_{10^{-3}}$	102	30	$1\cdot7_{10^{-7}}$	278	73
7	$6\cdot2_{10^{-4}}$	124	30	$9\cdot0_{10^{-4}}$	117	35	$8\cdot4_{10^{-8}}$	312	85
8	$1\cdot1_{10^{-4}}$	138	34	$2\cdot9_{10^{-4}}$	126	39	$2\cdot8_{10^{-8}}$	358	97
9	$6\cdot1_{10^{-5}}$	152	38	$8\cdot4_{10^{-5}}$	138	44	$2\cdot4_{10^{-10}}$	406	109
10	$1\cdot2_{10^{-5}}$	165	42	$8\cdot2_{10^{-5}}$	148	49	$4\cdot3_{10^{-11}}$	448	121
11	$1\cdot3_{10^{-8}}$	180	46	$5\cdot0_{10^{-5}}$	161	54	$3\cdot7_{10^{-11}}$	497	133
12	$1\cdot6_{10^{-11}}$	192	50	$2\cdot8_{10^{-6}}$	177	59	$3\cdot7_{10^{-11}}$	533	145
13	$8\cdot6_{10^{-12}}$	207	55	$2\cdot9_{10^{-7}}$	195	64			
14	$3\cdot9_{10^{-12}}$	223	60	$4\cdot3_{10^{-8}}$	208	69			
15	$5\cdot0_{10^{-14}}$	238	64	$2\cdot0_{10^{-8}}$	219	74			
16	$2\cdot1_{10^{-14}}$	253	69	$5\cdot3_{10^{-9}}$	235	79			

Tables for Chebyquad

Table 4—Chebyquad $n = 2$

ITN.	D.S.C.			POWELL			SMITH		
	F	FNS.	L.M.	F	FNS.	L.M.	F	FNS.	L.M.
0	$2\cdot0_{10^{-1}}$	1	0	$2\cdot0_{10^{-1}}$	1	0	$2\cdot0_{10^{-1}}$	1	0
1	$1\cdot3_{10^{-3}}$	13	3	$1\cdot3_{10^{-2}}$	12	3	$9\cdot7_{10^{-6}}$	17	4
2	$1\cdot3_{10^{-3}}$	19	5	$1\cdot1_{10^{-4}}$	21	6	$1\cdot6_{10^{-6}}$	30	7
3	$1\cdot5_{10^{-7}}$	29	8	$1\cdot6_{10^{-10}}$	30	9	$1\cdot1_{10^{-9}}$	43	10
4	$1\cdot5_{10^{-7}}$	32	9	$5\cdot8_{10^{-12}}$	36	11	$1\cdot5_{10^{-13}}$	51	13
5	$1\cdot4_{10^{-11}}$	41	12	$8\cdot6_{10^{-14}}$	41	13			
6	$1\cdot6_{10^{-15}}$	50	15						
7	$1\cdot6_{10^{-19}}$	59	18						

Table 5—Chebyquad $n = 4$

ITN.	D.S.C.			POWELL			SMITH		
	F	FNS.	L.M.	F	FNS.	L.M.	F	FNS.	L.M.
0	$7\cdot1_{10^{-2}}$	1	0	$7\cdot1_{10^{-2}}$	1	0	$7\cdot1_{10^{-2}}$	1	0
1	$1\cdot7_{10^{-2}}$	17	5	$1\cdot1_{10^{-2}}$	22	5	$2\cdot8_{10^{-3}}$	54	13
2	$9\cdot0_{10^{-3}}$	29	9	$1\cdot7_{10^{-3}}$	34	10	$8\cdot2_{10^{-6}}$	95	25
3	$6\cdot6_{10^{-4}}$	48	14	$1\cdot8_{10^{-5}}$	47	15	$1\cdot1_{10^{-11}}$	130	37
4	$1\cdot9_{10^{-4}}$	62	18	$8\cdot9_{10^{-7}}$	61	20	$1\cdot8_{10^{-14}}$	164	49
5	$3\cdot1_{10^{-5}}$	85	23	$7\cdot9_{10^{-8}}$	72	24			
6	$7\cdot2_{10^{-6}}$	100	27	$2\cdot6_{10^{-12}}$	82	28			
7	$1\cdot3_{10^{-6}}$	115	31	$4\cdot1_{10^{-14}}$	91	32			
8	$1\cdot2_{10^{-8}}$	127	35						
9	$2\cdot0_{10^{-11}}$	142	40						
10	$2\cdot2_{10^{-14}}$	157	45						

Table 6—Chebyquad $n = 6$

ITN.	D.S.C.			POWELL			SMITH		
	F	FNS.	L.M.	F	FNS.	L.M.	F	FNS.	L.M.
0	$4\cdot6_{10^{-2}}$	1	0	$4\cdot6_{10^{-2}}$	1	0	$4\cdot6_{10^{-2}}$	1	0
1	$2\cdot6_{10^{-2}}$	22	7	$1\cdot9_{10^{-2}}$	27	7	$1\cdot7_{10^{-2}}$	96	26
2	$1\cdot9_{10^{-2}}$	52	14	$9\cdot4_{10^{-3}}$	47	14	$7\cdot3_{10^{-4}}$	192	51
3	$1\cdot0_{10^{-2}}$	78	20	$5\cdot9_{10^{-3}}$	66	21	$6\cdot5_{10^{-4}}$	290	76
4	$4\cdot3_{10^{-3}}$	104	26	$2\cdot4_{10^{-3}}$	88	28	$5\cdot9_{10^{-4}}$	389	101
5	$6\cdot5_{10^{-4}}$	127	32	$1\cdot2_{10^{-3}}$	108	35	$1\cdot2_{10^{-6}}$	431	126
6	$3\cdot5_{10^{-4}}$	145	38	$6\cdot1_{10^{-4}}$	126	42	$6\cdot5_{10^{-9}}$	589	151
7	$2\cdot4_{10^{-4}}$	175	45	$5\cdot3_{10^{-4}}$	142	48	$2\cdot7_{10^{-11}}$	670	176
8	$1\cdot5_{10^{-4}}$	200	51	$2\cdot3_{10^{-4}}$	160	55			
9	$3\cdot8_{10^{-5}}$	226	57	$1\cdot7_{10^{-4}}$	179	62			
10	$3\cdot0_{10^{-6}}$	251	63	$8\cdot1_{10^{-5}}$	196	68			
11	$6\cdot3_{10^{-7}}$	270	69	$8\cdot4_{10^{-6}}$	215	75			
12	$2\cdot6_{10^{-7}}$	288	75	$6\cdot6_{10^{-8}}$	233	82			
13	$9\cdot4_{10^{-8}}$	312	82	$7\cdot7_{10^{-9}}$	247	88			
14	$2\cdot8_{10^{-8}}$	331	88	$1\cdot3_{10^{-9}}$	261	94			
15	$8\cdot2_{10^{-9}}$	349	94	$1\cdot9_{10^{-11}}$	275	100			
16	$5\cdot0_{10^{-9}}$	376	101	$6\cdot8_{10^{-14}}$	288	106			
17	$3\cdot6_{10^{-9}}$	397	107						
18	$2\cdot5_{10^{-9}}$	418	113						
19	$1\cdot8_{10^{-9}}$	437	119						
20	$1\cdot1_{10^{-9}}$	457	125						
21	$4\cdot4_{10^{-10}}$	476	131						
22	$1\cdot9_{10^{-10}}$	495	137						
23	$3\cdot5_{10^{-11}}$	514	143						
24	$3\cdot9_{10^{-12}}$	532	149						

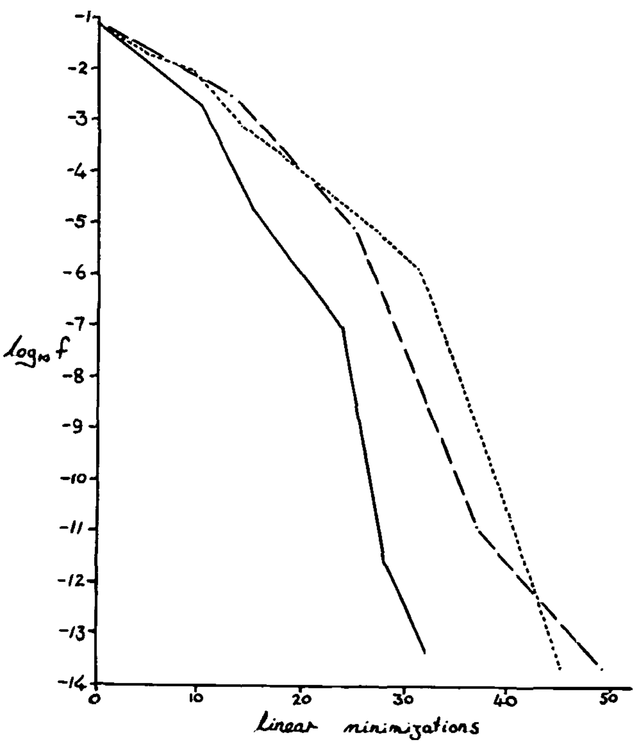

Fig. 5.—Chebyquad $n = 4$

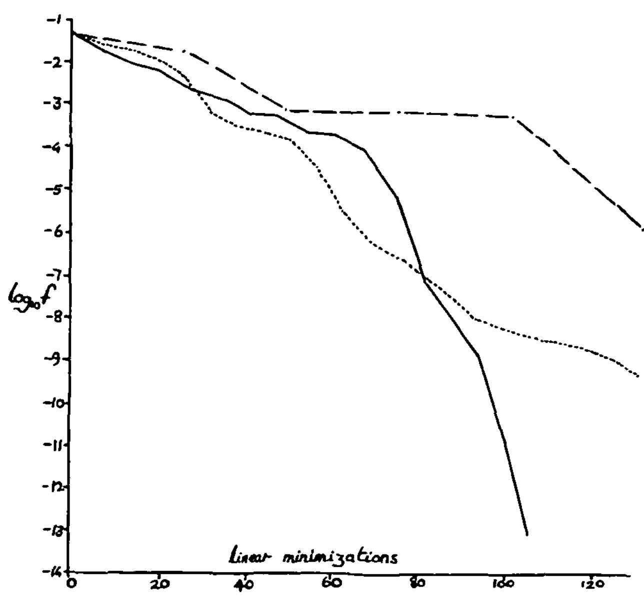

Fig. 6.—Chebyquad $n = 6$

Table 7—Chebyquad $n = 8$

ITN.	D.S.C.			POWELL			SMITH		
	F	FNS.	L.M.	F	FNS.	L.M.	F	FNS.	L.M.
0	$3\cdot5_{10^{-2}}$	1	0	$3\cdot5_{10^{-2}}$	1	0	$3\cdot5_{10^{-2}}$	1	0
1	$2\cdot2_{10^{-2}}$	28	9	$1\cdot2_{10^{-2}}$	29	9	$7\cdot9_{10^{-3}}$	154	43
2	$1\cdot4_{10^{-2}}$	60	18	$8\cdot9_{10^{-3}}$	52	17	$7\cdot0_{10^{-3}}$	319	85
3	$9\cdot6_{10^{-3}}$	90	26	$8\cdot0_{10^{-3}}$	73	25	$6\cdot4_{10^{-3}}$	488	127
4	$8\cdot3_{10^{-3}}$	117	34	$6\cdot6_{10^{-3}}$	91	33	$2\cdot9_{10^{-3}}$	669	169
5	$8\cdot2_{10^{-3}}$	141	42	$2\cdot7_{10^{-3}}$	122	42	$1\cdot2_{10^{-3}}$	840	211
6	$8\cdot0_{10^{-3}}$	178	51	$1\cdot9_{10^{-3}}$	150	51	$1\cdot6_{10^{-4}}$	1002	253
7	$6\cdot8_{10^{-3}}$	223	59	$1\cdot7_{10^{-3}}$	172	59	$2\cdot9_{10^{-5}}$	1177	295
8	$3\cdot4_{10^{-3}}$	266	67	$1\cdot7_{10^{-3}}$	194	67	$6\cdot6_{10^{-6}}$	1327	337
9	$2\cdot4_{10^{-3}}$	306	75	$1\cdot6_{10^{-3}}$	214	75	$7\cdot5_{10^{-7}}$	1482	379
10	$1\cdot1_{10^{-3}}$	348	83	$1\cdot4_{10^{-3}}$	235	83	$3\cdot2_{10^{-8}}$	1652	421
11	$5\cdot7_{10^{-4}}$	386	91	$6\cdot4_{10^{-4}}$	263	92			
12	$1\cdot7_{10^{-4}}$	424	99	$4\cdot1_{10^{-4}}$	292	101	exceeded time limit		
13	$3\cdot4_{10^{-5}}$	454	107	$2\cdot9_{10^{-4}}$	317	110			
14	$1\cdot0_{10^{-5}}$	480	115	$2\cdot0_{10^{-4}}$	339	118			
15	$6\cdot1_{10^{-6}}$	504	123	$4\cdot2_{10^{-5}}$	360	126			
16	$1\cdot7_{10^{-6}}$	544	132	$1\cdot6_{10^{-5}}$	385	135			
17	$7\cdot6_{10^{-7}}$	579	140	$1\cdot1_{10^{-5}}$	409	144			
18	$3\cdot0_{10^{-7}}$	607	148	$6\cdot7_{10^{-6}}$	428	152			
19	$1\cdot2_{10^{-7}}$	634	156	$5\cdot8_{10^{-6}}$	447	160			
20	$5_{10^{-8}}$	659	164	$5\cdot4_{10^{-6}}$	465	168			
21	$1_{10^{-8}}$	684	172	$4\cdot1_{10^{-6}}$	484	176			
22	$5_{10^{-10}}$	708	180	$5\cdot6_{10^{-7}}$	502	184			
23	$1_{10^{-10}}$	739	189	$1\cdot6_{10^{-10}}$	520	192			
24				$5\cdot7_{10^{-13}}$	537	200			

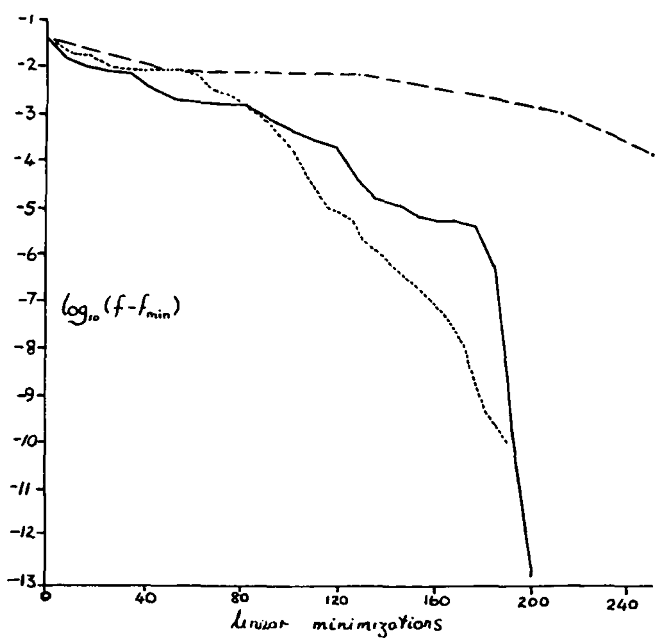

Fig. 7.—Chebyquad $n = 8$

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

SMITH, C. S. (1962). "The automatic computation of maximum likelihood estimates," N.C.B. Scientific Dept. Report S.C. 846/MR/40.

SWANN, W. H. (1964). "Report on the development of a new direct search method of optimization," I.C.I. Ltd., Central Instrument Laboratory Research Note 64/3.