nelder65
(The Comp. J. 7(1965) 308-313)

A simplex method for function minimization

By J. A. Nelder and R. Mead†

A method is described for the minimization of a function of n variables, which depends on the comparison of function values at the $(n+1)$ vertices of a general simplex, followed by the replacement of the vertex with the highest value by another point. The simplex adapts itself to the local landscape, and contracts on to the final minimum. The method is shown to be effective and computationally compact. A procedure is given for the estimation of the Hessian matrix in the neighbourhood of the minimum, needed in statistical estimation problems.

Spendley *et al.* (1962) introduced an ingenious idea for tracking optimum operating conditions by evaluating the output from a system at a set of points forming a simplex in the factor-space, and continually forming new simplices by reflecting one point in the hyperplane of the remaining points. This idea is clearly applicable to the problem of minimizing a mathematical function of several variables, as was recognized by these authors. However, they assumed that the relative steps to be made in varying the factors were known, and this makes their strategy rather rigid for general use. In the method to be described the simplex adapts itself to the local landscape, elongating down long inclined planes, changing direction on encountering a valley at an angle, and contracting in the neighbourhood of a minimum. The criterion for stopping the process has been chosen with an eye to its use for statistical problems involving the maximization of a likelihood function, in which the unknown parameters enter non-linearly.

The method

We consider, initially, the minimization of a function of n variables, without constraints. $P_0, P_1, \ldots P_n$ are the $(n+1)$ points in n-dimensional space defining the current "simplex." [The simplex will not, of course, be regular in general.] We write y_i for the function value at P_i, and define

h as the suffix such that $y_h = \max\limits_i (y_i)$ [h for "high"]
and

l as the suffix such that $yl = \min\limits_i (y_i)$ [l for "low"].

Further we define $\bar{P}$ as the centroid of the points with $i \neq h$, and write $[P_iP_j]$ for the distance from P_i to P_j. At each stage in the process P_h is replaced by a new point; three operations are used—*reflection, contraction,* and *expansion.* These are defined as follows: the reflection of P_h is denoted by P^*, and its co-ordinates are defined by the relation

$$P^* = (1+\alpha)\bar{P} - \alpha P_h$$

where α is a positive constant, the *reflection coefficient.* Thus P^* is on the line joining P_h and $\bar{P}$, on the far side of $\bar{P}$ from P_h with $[P^*\bar{P}] = \alpha[P_h\bar{P}]$. If y^* lies between y_h and y_l, then P_h is replaced by P^* and we start again with the new simplex.

If $y^* < y_l$, i.e. if reflection has produced a new minimum, then we expand P^* to P^{**} by the relation

$$P^{**} = \gamma P^* + (1-\gamma)\bar{P}.$$

The *expansion coefficient* γ, which is greater than unity, is the ratio of the distance $[P^{**}\bar{P}]$ to $[P^*\bar{P}]$. If $y^{**} < y_l$ we replace P_h by P^{**} and restart the process; but if $y^{**} > y_l$ then we have a failed expansion, and we replace P_h by P^* before restarting.

If on reflecting P to P^* we find that $y^* > y_i$ for all $i \neq h$, i.e. that replacing P by P^* leaves y^* the maximum, then we define a new P_h to be either the old P_h or P^*, whichever has the lower y value, and form

$$P^{**} = \beta P_h + (1-\beta)\bar{P}.$$

The *contraction coefficient* β lies between 0 and 1 and is the ratio of the distance $[P^{**}\bar{P}]$ to $[P\bar{P}]$. We then accept P^{**} for P_h and restart, unless $y^{**} > \min(y_h, y^*)$, i.e. the contracted point is worse than the better of P_h and P^*. For such a failed contraction we replace all the P_i's by $(P_i + P_l)/2$ and restart the process.

A failed expansion may be thought of as resulting from a lucky foray into a valley (P^*) but at an angle to the valley so that P^{**} is well up on the opposite slope. A failed contraction is much rarer, but can occur when a valley is curved and one point of the simplex is much farther from the valley bottom than the others; contraction may then cause the reflected point to move away from the valley bottom instead of towards it. Further contractions are then useless. The action proposed contracts the simplex towards the lowest point, and will eventually bring all points into the valley. The coefficients α, β, γ give the factor by which the volume of the simplex is changed by the operations of reflection, contraction or expansion respectively. The complete method is given as a flow diagram in Fig. 1.

A final point concerns the criterion used for halting the procedure. The criterion adopted is somewhat different from that used by Powell (1964) in that it is concerned with the variation in the y values over the

† *National Vegetable Research Station, Wellesbourne, Warwick.*

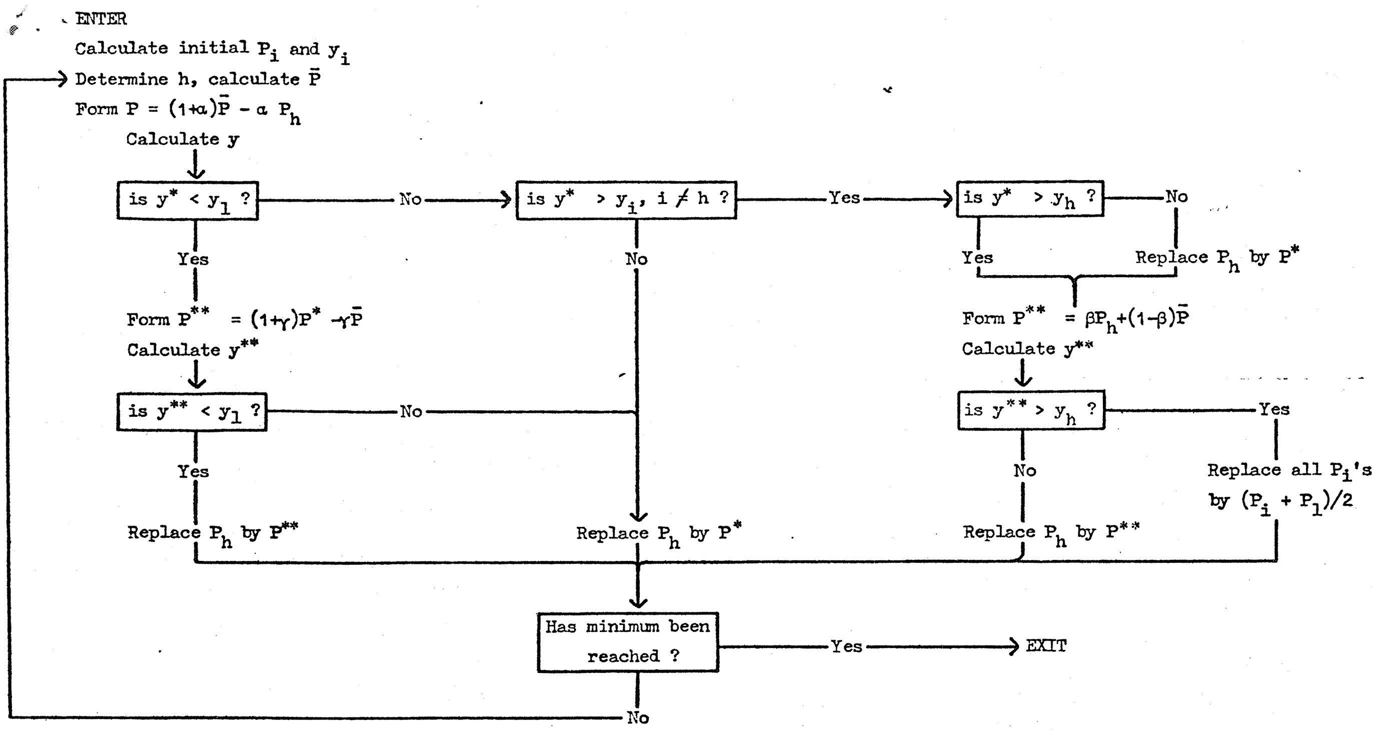
ENTER
Calculate initial P_i and y_i
Determine h, calculate $\bar{P}$
Form $P = (1+\alpha)\bar{P} - \alpha P_h$
Calculate y
is $y^* < y_l$?
No
is $y^* > y_i$, $i \neq h$?
Yes
is $y^* > y_h$?
No
Yes
No
Yes
Replace P_h by P^*
Form $P^{**} = (1+\gamma)P^* - \gamma\bar{P}$
Calculate y^{**}
Form $P^{**} = \beta P_h + (1-\beta)\bar{P}$
Calculate y^{**}
is $y^{**} < y_l$?
No
is $y^{**} > y_h$?
Yes
Yes
No
Replace all P_i's by $(P_i + P_l)/2$
Replace P_h by P^{**}
Replace P_h by P^*
Replace P_h by P^{**}
Has minimum been reached ?
Yes
EXIT
No

Fig. 1.—Flow diagram

simplex rather than with changes in the x's. The form chosen is to compare the "standard error" of the y's in the form $\sqrt{\{\Sigma(y_i - \bar{y})^2/n\}}$ with a pre-set value, and to stop when it falls below this value. The success of the criterion depends on the simplex not becoming too small in relation to the curvature of the surface until the final minimum is reached. The reasoning behind the criterion is that in statistical problems where one is concerned with finding the minimum of a negative likelihood surface (or of a sum-of-squares surface) the curvature near the minimum gives the information available on the unknown parameters. If the curvature is slight the sampling variance of the estimates will be large so there is no sense in finding the co-ordinates of the minimum very accurately, while if the curvature is marked there is justification for pinning down the minimum more exactly.

Constraints on the volume to be searched

If, for example, one of the x_i must be non-negative in a minimization problem, then our method may be adapted in one of two ways. The scale of the x concerned can be transformed, e.g., by using the logarithm, so that negative values are excluded, or the function can be modified to take a large positive value for all negative x. In the latter case any trespassing by the simplex over the border will be followed automatically by contraction moves which will eventually keep it inside. In either case an actual minimum with $x = 0$ would be inaccessible in general, though arbitrarily close approaches could be made to it. Clearly either technique can deal with individual limitations on the range of any number of x's. Constraints involving more than one x can be included using the second technique provided that an initial simplex can be found inside the permitted region, from which to start the process. Linear constraints that reduce the dimensionality of the field of search can be included by choosing the initial simplex to satisfy the constraints and reducing the dimensions accordingly. Thus to minimize $y = f(x_1, x_2, x_3)$ subject to $x_1 + x_2 + x_3 = X$, we could choose an initial simplex with vertices $(X, 0, 0)$, $(0, X, 0)$, and $(0, 0, X)$, treating the search as being in two dimensions. In particular, any x_i may be held constant by setting its value to that constant for all vertices of the initial simplex.

Results

Three functions, all of which have been used before for testing minimization procedures, were used to test the method. The functions, all of which have a minimum of zero, were:

(1) Rosenbrock's parabolic valley (Rosenbrock (1960))
$y = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$, starting point $(-1{\cdot}2, 1)$.

(2) Powell's quartic function (Powell (1962))
$y = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$,
starting point $(3, -1, 0, 1)$.

(3) Fletcher and Powell's helical valley (Fletcher and Powell (1963))

$$y = 100[x_3 - 10\theta(x_1, x_2)]^2 + [\sqrt{(x_1^2 + x_2^2)} - 1]^2 + x_3^2$$

where
$$\begin{aligned} 2\pi\theta(x_1, x_2) &= \arctan (x_2/x_1),\ x_1 > 0 \\ &= \pi + \arctan (x_2/x_1),\ x_1 < 0 \end{aligned}$$

starting point $(-1, 0, 0)$.

The stopping criterion used was $\sqrt{\{\Sigma(y_i - \bar{y})^2/n\}} < 10^{-8}$. The function value at the centroid of the final simplex usually deviated from the true minimum by less than 10^{-8}; a sample of runs gave $2\cdot5 \times 10^{-9}$ as the geometric mean of this deviation. A difficulty encountered in testing the procedure was that the size and orientation of the initial simplex had an effect on the speed of convergence and consequently several initial step-lengths and several arrangements of the initial simplex were used for each trial (the arrangements consisted of two forms of the initial simplex, regular, as in Spendley *et al.*'s original method, and axial from the starting point, combined with several orientations). The first set of trials investigated the different strategies (values of α, β and γ) of the simplex method of minimization, and the second compared the results for the best strategy with those of Powell (1964), which are among the best previously obtained.

An initial trial with function (1) used all combinations of $\alpha = \frac{1}{2}, \frac{2}{3}, 1$; $\beta = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}$; $\gamma = 2, 3, 4$; and initial step-lengths $\frac{1}{2}$, 1, and 2. The main result was that the lower values of α and β gave generally slower convergence. In a second trial with function (1) six main strategies $\alpha = 1$, $\beta = \frac{1}{2}$ or $\frac{1}{3}$, and $\gamma = 2, 3, 4$ with two additional strategies $(\frac{2}{3}, \frac{1}{4}, 3)$ and $(\frac{2}{3}, \frac{1}{3}, 4)$ were all tried for three initial step-lengths $\frac{1}{2}$, 1, 2, and eight arrangements of the initial simplex. This trial showed that the additional strategies were more variable in performance and, on the average, slower to converge than the six main strategies. These six strategies gave very similar results, as was shown by an analysis of variance of the number of function evaluations, the mean square for strategies being 197 compared with a residual mean square (after removing size and orientation effects) of 463.

Using function (2) five strategies were tried: $\alpha = 1$, $\beta = \frac{1}{2}$, $\gamma = 2$, 3 or 4, and $\alpha = 1$, $\beta = \frac{1}{3}$ or $\frac{1}{4}$, $\gamma = 3$. The latter two strategies gave very variable results, as is indicated in Table 1 which shows the mean and minimum numbers of evaluations required for convergence over eight arrangements.

Analyzing the results for the best three strategies gave a standard error of difference between the strategy mean numbers of $\pm 7\cdot0$ (based on 70 degrees of freedom), showing that the results for the three strategies are all significantly different at the 95% level.

Using function (3) for the same trial confirmed these conclusions, the last two strategies converging to false minima on several occasions.

It was clear, therefore, that the simple strategy $\alpha = 1$, $\beta = \frac{1}{2}$, $\gamma = 2$ was the best, and these values were used in comparing the method with Powell's (1964) technique.

Table 1

Mean and minimum numbers of evaluations for function

MEAN NUMBER

STEP-LENGTH	STRATEGY				
	$(1, \frac{1}{2}, 2)$	$(1, \frac{1}{2}, 3)$	$(1, \frac{1}{2}, 4)$	$(1, \frac{1}{3}, 2)$	$(1, \frac{1}{4}, 2)$
0·25	225	234	282	255	344
0·5	210	234	254	300	335
1	216	229	260	283	343
2	216	239	250	264	253
4	226	241	251	249	347
Mean	219	235	259	270	322

MINIMUM NUMBER

STEP-LENGTH	$(1, \frac{1}{2}, 2)$	$(1, \frac{1}{2}, 3)$	$(1, \frac{1}{2}, 4)$	$(1, \frac{1}{3}, 2)$	$(1, \frac{1}{4}, 2)$
0·25	191	205	241	172	189
0·5	181	181	190	190	233
1	180	191	210	150	216
2	180	194	201	174	117
4	184	170	217	152	217
Mean	183	188	212	168	194

Comparison with Powell's results

For this comparison we used the same convergence criterion; smaller values of the criterion would not be justifiable when it is remembered that most functions met with in practice are likely to have a rounding-off error of this order. We used all three functions with a range of initial step-lengths 0·1(0·1)1(0·2)3 and eight initial simplex arrangements. Table 2 shows the results of these trials.

Apart from the smaller step-lengths the effect of step-length on number of evaluations required is not very large and it is reasonable to use the mean number of evaluations over all the unasterisked step-lengths for comparison with other methods. The mean values obtained are 144 evaluations for function (1), 216 for function (2), and 228 for function (3). Powell's results for functions (1) and (2) give the evaluations required to reach our mean final value of $2\cdot5 \times 10^{-9}$ as 150 and 235 respectively.

An EMA version of Powell's method was written and for function (3) it was found that the initial step-length had a considerable effect on the speed of convergence. Furthermore the optimum initial step-lengths for the two methods were different. For initial step-lengths in the range 0·1 to 0·95, Powell's method gave numbers of function values between 177 and 375 (sample of six).

Table 2

Mean number of evaluations required for convergence for different step-lengths for functions (1), (2) and (3)

STEP-LENGTH	FUNCTION (1)	FUNCTION (2)	FUNCTION (3)
0·1	165*	262*	310*
0·2	161*	220	256
0·3	159*	214	276
0·4	162*	215	256
0·5	143	210	230
0·6	142	206	268
0·7	151	210	264
0·8	150	231	231
0·9	157	223	236
1·0	147	216	213
1·2	145	217	183
1·4	145	208	194
1·6	154	217	188
1·8	138	225	213
2·0	144	216	207
2·2	130†	197	206
2·4	128	224	224
2·6	134	223	209
2·8	135	241	233
3·0	148	191	234

* See text.
† Omitting arrangements where the true minimum was a vertex of the initial simplex.

There appears, therefore, to be little difference between the two methods to this degree of accuracy as far as our evidence goes. However, there are indications that, with less severe convergence criteria, greater differences would have been found. This is suggested by the progress of the two methods on three typical runs, as shown in Table 3; for all three functions our method establishes an initial advantage over Powell's. For function (2) this advantage is maintained approximately at the same level for the remainder of the run; for functions (1) and (3) the advantage is further increased up to the final stage when Powell's method rapidly closes the gap.

Effect of number of variables

To investigate the effect of increasing the number of variables, the sum of fourth powers was used as a test function, starting at (1, 1, 1, . . ., 1). For functions with 2 to 10 variables it was found that the relationship between the number of variables k and the mean number of evaluations N for convergence (using a final value of approximately $2{\cdot}5 \times 10^{-9}$) was well described by

$$N = 3{\cdot}16(k+1)^{2{\cdot}11}.$$

Discussion

The method presented in this paper differs from most previously put forward in being based neither on gradients (first-order derivatives) nor on quadratic forms (second-order derivatives). As can be seen in Fig. 1, the algebraic operations on the x_i are all linear (e.g. as in determining P from P_h and $\bar{P}$), while those on the y_i are concerned with finding the maximum or minimum of a finite set of quantities. This latter type of operation is non-linear and accounts for the ability of the method to locate a minimum with arbitrary accuracy. The remark of Spendley *et al.* (1962) that "Continued application of the simplex procedure, with progressively reduced step size, is inherently as self-defeating as any other linear technique," is thus somewhat misleading, since the use made of the simplex at any stage is not to estimate parameters in a regression equation (in which case only a linear model could be fitted) but to guide the direction of the next move.

Out method is highly opportunist, in that the least possible information is used at each stage and no account is kept of past positions. No assumptions are made about the surface except that it is continuous and has a unique minimum in the area of the search. It might thus be expected that when the curvature of the landscape (as measured by the Hessian matrix of second-order partial derivatives) is changing rapidly, the present method will do well when compared with methods which depend on arguments applicable to quadratic forms. Conversely in the neighbourhood of a minimum, when the Hessian matrix is relatively stable, it may do worse. This expectation is borne out by the results obtained on the test functions. However, the positions of minima are often needed with only limited accuracy, so that final rapid convergence is not essential. This is especially true in statistical problems where the surface may be, e.g., a sum of squares of residuals, and the position of the minimum is a function of the random errors in the observations. An important property of our method is that it will converge even when the initial simplex straddles two or more valleys, a property which is not shared by, e.g., Powell's method.

A general problem encountered by all minimization methods is that of false convergence at a point other than the minimum. This difficulty has been found in using the simplex method on a four-dimensional surface having a long, curved, valley with extremely steep sides; along the valley bottom the function varies considerably compared with the accuracy to which the minimum function value is required. On meeting the valley, the simplex may need to undergo a drastic change in both size and shape, in the course of which the variation between the function values at the simplex vertices may become small even when compared with a convergence criterion of one-tenth of the required accuracy. Merely refining the convergence criterion would often involve needless evaluations, and in other more extreme cases could still be ineffective.

Table 3

Values of the three functions on 3 typical runs, for simplex method and Powell's (1964) method

NUMBER OF EVALUATION	FUNCTION					
	(1)		(2)		(3)	
	Simplex	Powell	Simplex	Powell	Simplex	Powell
1	$2\cdot 4 \times 10^{1}$	$2\cdot 4 \times 10^{1}$	$2\cdot 2 \times 10^{2}$	$2\cdot 2 \times 10^{2}$	$2\cdot 5 \times 10^{3}$	$2\cdot 5 \times 10^{3}$
20	$2\cdot 6 \times 10^{0}$	$3\cdot 6 \times 10^{0}$	$1\cdot 2 \times 10^{1}$	$3\cdot 1 \times 10^{1}$	$1\cdot 9 \times 10^{0}$	$1\cdot 9 \times 10^{2}$
40	$9\cdot 7 \times 10^{-1}$	$2\cdot 2 \times 10^{0}$	$3\cdot 0 \times 10^{-1}$	$4\cdot 2 \times 10^{0}$	$1\cdot 2 \times 10^{0}$	$3\cdot 2 \times 10^{1}$
60	$3\cdot 8 \times 10^{-1}$	$1\cdot 1 \times 10^{0}$	$8\cdot 7 \times 10^{-2}$	$1\cdot 5 \times 10^{0}$	$7\cdot 6 \times 10^{-2}$	$7\cdot 7 \times 10^{0}$
80	$8\cdot 1 \times 10^{-2}$	$4\cdot 6 \times 10^{-1}$	$1\cdot 7 \times 10^{-2}$	$2\cdot 3 \times 10^{-1}$	$3\cdot 0 \times 10^{-2}$	$2\cdot 0 \times 10^{0}$
100	$1\cdot 3 \times 10^{-3}$	$1\cdot 0 \times 10^{-1}$	$1\cdot 2 \times 10^{-3}$	$6\cdot 4 \times 10^{-3}$	$2\cdot 0 \times 10^{-2}$	$7\cdot 6 \times 10^{-1}$
120	$8\cdot 4 \times 10^{-6}$	$1\cdot 7 \times 10^{-2}$	$6\cdot 7 \times 10^{-5}$	$1\cdot 1 \times 10^{-3}$	$4\cdot 4 \times 10^{-3}$	$3\cdot 1 \times 10^{-1}$
140	$7\cdot 7 \times 10^{-9}$	$1\cdot 0 \times 10^{-4}$	$3\cdot 9 \times 10^{-6}$	$1\cdot 3 \times 10^{-4}$	$1\cdot 3 \times 10^{-4}$	$2\cdot 3 \times 10^{-1}$
150	$3\cdot 3 \times 10^{-10}$	$1\cdot 2 \times 10^{-9}$				
160			$1\cdot 2 \times 10^{-6}$	$2\cdot 1 \times 10^{-5}$	$1\cdot 0 \times 10^{-6}$	$6\cdot 6 \times 10^{-2}$
180			$5\cdot 9 \times 10^{-7}$	$2\cdot 1 \times 10^{-6}$	$1\cdot 7 \times 10^{-7}$	$3\cdot 6 \times 10^{-3}$
200			$2\cdot 6 \times 10^{-8}$	$1\cdot 3 \times 10^{-7}$	$3\cdot 7 \times 10^{-8}$	$1\cdot 1 \times 10^{-4}$
220			$9\cdot 3 \times 10^{-10}$	$1\cdot 6 \times 10^{-8}$	$2\cdot 7 \times 10^{-10}$	$7\cdot 2 \times 10^{-9}$

The values for Powell's method were obtained by logarithmic interpolation of the function values at the end of each iteration. Data for functions (1) and (2) from Powell (1964), data for function (3) from our EMA program of his method.

Powell (1964) suggested a more complex convergence criterion, for this general problem, based on perturbing the first minimum found and repeating the method to find a second minimum, followed by exploration along the line joining the two. An alternative technique, more suited to our convergence criterion in terms of function variation, is to continue after the first convergence for a prescribed number of evaluations, to test for convergence again and, if the second test proves successful, to compare the two "converged" function values. Only if these values are sufficiently close is convergence allowed.

The simplex method is computationally compact; on the Orion computer the basic routine (without final printing) contains less than 350 instructions, and the great majority of orders are additions and subtractions or simple logical orders. There are few multiplications, and no divisions at all except on entering and leaving the routine.

Copies of the routine, written in Extended Mercury Autocode, are available from the authors.

Appendix

The Hessian matrix at the minimum

The minimization method proposed, being independent of the properties of quadratic forms, does not yield any estimate of the Hessian matrix of second derivatives at the minimum. This matrix is, of course, the information matrix in statistical problems when the function being minimized is minus the log. likelihood, and its inverse is the sample variance–covariance matrix of the estimates. A convenient way of utilizing a quadratic surface to estimate the minimum when the simplex is close to that minimum was given by Spendley *et al.* (1962) and their method can be readily extended to give the required variance–covariance matrix of the estimates.

If the $(n + 1)$ points of the simplex in n dimensions are given by $P_0, P_1, \ldots P_n$, then Spendley *et al.* form the "half-way points" $P_{ij} = (P_i + P_j)/2$, $i \neq j$ and fit a quadratic surface to the combined set of $(n + 1)(n + 2)/2$ points. If the original points of the simplex are used to define a set of oblique axes with co-ordinates x_i, then the points may be taken as

$$\begin{array}{l} (0, 0, 0, \ldots 0) \\ (1, 0, 0, \ldots 0) \\ (0, 1, 0, \ldots 0) \\ \cdot\cdot \quad \cdot\cdot\cdot \quad \cdot\cdot \end{array}$$

and $(0, 0, 0, \ldots 1)$.

If the quadratic approximation to the function in the neighbourhood of the minimum is written as

$$y = a_0 + 2\Sigma a_i x_i + \Sigma b_{ij} x_i x_j,$$

or in vector form as

$$y = a_0 + 2a'x + x'Bx,$$

See p827, Vol 8

then the coefficients are estimated as

$$\begin{aligned} a_0 &= y_0 \\ a_i &= 2y_{0i} - (y_i + 3y_0)/2, & i &= 1, \ldots, n \\ b_{ii} &= 2(y_i + y_0 - 2y_{0i}), & i &= 1, \ldots, n \\ b_{ij} &= 2(y_{ij} + y_0 - y_{0i} - y_{0j}), & i &\neq j, \end{aligned}$$

where y_i is the function value at P_i and y_{ij} that at P_{ij}. The estimated minimum is then given by

$$\boldsymbol{x}_{min} = -\boldsymbol{B}^{-1}\boldsymbol{a}$$

and the information matrix is just $\boldsymbol{B}$.

If $\boldsymbol{p}_i$ denotes the co-ordinates of P_i in the original system, and if $\boldsymbol{Q}$ is the $n \times n$ matrix whose ith column is $\boldsymbol{p}_i - \boldsymbol{p}_0$, then the minimum is estimated to be at

$$\begin{aligned} \boldsymbol{p}_{min} &= \boldsymbol{p}_0 + \boldsymbol{Q}\boldsymbol{x}_{min} \\ &= \boldsymbol{p}_0 - \boldsymbol{Q}\boldsymbol{B}^{-1}\boldsymbol{a}. \end{aligned}$$

The minimum value of the function is estimated to be

$$y_{min} = a_0 - \boldsymbol{a}'\boldsymbol{B}^{-1}\boldsymbol{a}.$$

The information matrix in the original co-ordinate system is given by

$$(\boldsymbol{Q}^{-1})'\boldsymbol{B}\boldsymbol{Q}^{-1}$$

so that the variance–covariance matrix is given by

$$\boldsymbol{Q}\boldsymbol{B}^{-1}\boldsymbol{Q}'.$$

If normal equal-variance independent errors are involved and the sum of squares of residuals is minimized, then this matrix must be multiplied by $2\sigma^2$, where as usual σ^2 would be estimated by $y_{min}/(N - n)$, N being the total number of observations, and n the number of parameters fitted.

In estimating $\boldsymbol{B}$ numerically it is necessary to steer a course between two hazards. In one the simplex is so small that $(y_{ij} + y_0 - y_{0i} - y_{0j})$ consists largely of rounding-off errors incurred in calculating the y's. In the other the simplex is so large that the quadratic approximation is poor, and the b's are correspondingly biased. If the method given in this paper is used, the former hazard will usually be the important one, and it may be necessary to enlarge the final simplex before adding the extra points. A possible way of doing this is to double the distance of each point P_i from the centroid until the corresponding function value exceeds that at the centroid by more than a given constant. The choice of this would depend on the rounding-off error attaching to the evaluation of the function, and would need to be at least 10^3 times that rounding error, if acceptable estimates of the b's were to be obtained.

References

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

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.

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

Spendley, W., Hext, G. R., and Himsworth, F. R. (1962). "Sequential Application of Simplex Designs in Optimisation and Evolutionary Operation," *Technometrics*, Vol. 4, p. 441.

Correspondence

To the Editor,
The Computer Journal.

An impossible program

Sir,

A well-known piece of folk-lore among programmers holds that it is impossible to write a program which can examine any other program and tell, in every case, if it will terminate or get into a closed loop when it is run. I have never actually seen a proof of this in print, and though Alan Turing once gave me a verbal proof (in a railway carriage on the way to a Conference at the NPL in 1953), I unfortunately and promptly forgot the details. This left me with an uneasy feeling that the proof must be long or complicated, but in fact it is so short and simple that it may be of interest to casual readers. The version below uses CPL, but not in any essential way.

Suppose $T[R]$ is a Boolean function taking a routine (or program) R with no formal or free variables as its argument and that for all R, $T[R]$ = **True** if R terminates if run and that $T[R]$ = **False** if R does not terminate. Consider the routine P defined as follows

```
rec routine P
    §L : if T[P] go to L
Return §
```

If $T[P]$ = **True** the routine P will loop, and it will only terminate if $T[P]$ = **False**. In each case $T[P]$ has exactly the wrong value, and this contradiction shows that the function T cannot exist.

Yours faithfully,
C. Strachey.

Churchill College,
Cambridge.

We note that, for any set of representative problems, there is a common set S of basic operations which both machines will perform: the purely arithmetic storage-to-storage operation (excluding pseudo-arithmetics like subscript manipulation and sign change). We now regard the proposed machine P as comprising: Pl, the orthodox register-data flow configuration, together with the microprogram for executing the operations S; P2 the remaining microprogram, which includes the compiler, array manipulation, function execution, and general stack-organization facilities.

Cost

Because the pushdown is microprogrammed, there is no special circuitry in the machine, which may be regarded as typical of machines of the same size. The cost of the proposed philosophy therefore lies wholly within the P2 microprogram, which is in fact almost exactly the same size as that of P1. In a machine in the low-cost scientific field, therefore, the cost is approximately that of doubling the microprogram of a more conventional machine.

Note that the basic FORTRAN processor needs only typewriter input-output.

Performance

Machine X is provided with a single-pass FORTRAN compiler. A set of representative problems was:

(i) hand-coded for machine X,

(ii) written in FORTRAN, and

(iii) compiled on machine X.

After allowing for difference in basic circuit speeds, the execution times for the described machine were, for all problems:

(*a*) greater than the hand-coded solution times, but

(*b*) less than the compiled solution times.

The instruction storage required by machine P was at all times less than that required by either configuration of machine X. The time taken to write the programs in FORTRAN was considerably less than to hand-code them.

Efficiency

For programs which are to be run repeatedly, optimum coding for inner loops becomes increasingly important. If the program is to be restricted to FORTRAN, the P2 microprogram may have to be larger and more sophisticated. Clearly any compiler feature can be implemented in microprogram—at a price. Two factors contribute to the minimization of the P2 microprogram: the close structural resemblance between source and object program, and the fact that the compiler itself can be implemented in a language (chosen by the machine designer) which is itself independent of source and object languages.

The machine is intended for the inexperienced programmer who wishes to express and solve scientific problems, preponderantly "one-off," with a minimum of delay and cost. Time saved in writing and testing programs will more than compensate for any inefficiency in inner-loop processing. Direct communication with the machine will quicken the education process.

Conclusions

There is little doubt as to the usefulness of a machine which provides the facilities described here. Of the several design philosophies conceivable, we have studied the feasibility and practicability of incorporating the FORTRAN processing facilities in the machine circuitry. Some measure has been given of the effect of the proposed philosophy on the cost and performance of the basic computer unit. It is felt that any depreciation here is more than offset by the improved efficiency and operational facilities of the overall system.

Acknowledgements

The design and a complete simulation of the machine described formed part of an exploratory machine design study under Mr. C. E. Owen, IBM World Trade Laboratories (Great Britain) Ltd. The authors wish to thank the laboratory management for permission to publish this paper.

Errata

A simplex method for function minimization by J. Nelder and R. Mead.

The following alterations are required in the appendix to the above paper which was published in Vol. 7, p. 308. The two expressions for the information matrix should have a factor of 2 attached, and that for the variance-covariance matrix a factor of $\frac{1}{2}$. The authors are grateful to Dr. Rodes Trautman for drawing their attention to this slip.