



WORKING FOR A HEALTHY FUTURE

# HISTORICAL RESEARCH REPORT

Research Report R/69/01  
1969

## The automatic computation of maximum likelihood estimates

Smith, CS



WORLD HEALTH ORGANISATION  
COLLABORATING CENTRE  
FOR OCCUPATIONAL HEALTH

**RESEARCH CONSULTING SERVICES**

Multi-disciplinary specialists in Occupational and Environmental Health and Hygiene

[www.iom-world.org](http://www.iom-world.org)





**The automatic computation of maximum likelihood estimates**

Smith, CS

This document is a facsimile of an original copy of the report, which has been scanned as an image, with searchable text. Because the quality of this scanned image is determined by the clarity of the original text pages, there may be variations in the overall appearance of pages within the report.

The scanning of this and the other historical reports in the Research Reports series was funded by a grant from the Wellcome Trust. The IOM's research reports are freely available for download as PDF files from our web site: <http://www.iom-world.org/research/libraryentry.php>



NATIONAL COAL BOARD

PNEUMOCONIOSIS FIELD RESEARCH

THE AUTOMATIC COMPUTATION OF MAXIMUM LIKELIHOOD ESTIMATES

by

C.S. Smith

CONTENTS

SUMMARY

1. INTRODUCTION

2. STANDARD MAXIMISATION METHODS

- 2.1 General Remarks
- 2.2 The Newton-Raphson Method
- 2.3 The Alternating Variables Method
- 2.4 The Method of Steepest Ascent
- 2.5 Rosenbrock's Method

3. DEVELOPMENT OF THE STANDARD METHODS

- 3.1 The Maximum on a Given Line
- 3.2 Step Size in the Steepest Ascent Method
- 3.3 Oscillation About a Ridge
- 3.4 An Adaptation of Rosenbrock's Method

4. A NEW METHOD

- 4.1 Basic Idea
- 4.2 Procedure Adopted
- 4.3 Constraints on Variables
- 4.4 Computer Programme

5. EXAMPLES

6. DISCUSSION AND CONCLUSIONS

7. ACKNOWLEDGMENTS

References

Appendix

Approved by *A. Ashford p.p.*  
Chief Scientist,  
Pneumoconiosis Field Research.

Date 19<sup>th</sup> June 1962

Scientific Department,  
Queensborough House,  
12-18, Albert Embankment,  
LONDON, S.E.1.

## SUMMARY

This report describes the work done in developing a method of computing the maximum of a function of several variables. This is intended for use in maximum likelihood or in minimum chi-square estimation.

The standard methods are described and their defects indicated. A method devised by H.H. Rosenbrock is explained and compared with the new method.

It is concluded that the new method is more suitable for much of the work of P.F.R. A set of instructions for Ferranti Mercury "Autocode" has been written, which will carry out this process, and specifications for its use are given in an Appendix.

# THE AUTOMATIC COMPUTATION OF MAXIMUM LIKELIHOOD ESTIMATES

## 1. INTRODUCTION

Many of the statistical problems of the Pneumoconiosis Field Research consist essentially in estimating the parameters in a mathematical model and testing the "fit" of a set of observations. The estimation is usually done by the method of maximum likelihood, or of minimum chi-square.

The model to be tested leads to expressions for the probabilities of given sets of observations in terms of the data and of a set of parameters. From these probabilities, a likelihood function or a chi-square function can be derived. We look for the values of the parameters which give the greatest value of the likelihood or the least value of chi-square. If maximum likelihood is used, the accuracy of the estimates, expressed by their variances and co-variances, is given by the negative of the inverse of the expected value of the matrix of second derivatives of the likelihood function:

$$[\text{cov}(u,v)] = [-E\left(\frac{\partial^2 L}{\partial u \partial v}\right)]^{-1} \quad (1)$$

The function to be optimised may involve a large number of observations and several unknown parameters. If  $f$  is the function, and  $u_0, \dots, u_{s-1}$  the parameters, we have to solve the equations.

$$\frac{\partial f}{\partial u_i} = 0, \quad i = 0, 1, \dots, (s-1) \quad (2)$$

The function  $f$  is generally too complicated for the equations to be solved directly, or even for the derivatives to be computed from their algebraic expressions, and an iterative method has to be used to approximate successively to the optimum, starting from some initial estimates.

An automatic method for finding the maximum of a given function would therefore be very useful.\* Such a method should be suitable for use with an electronic computer, and should be applicable generally to a variety of functions. The methods already available are discussed below. None of these was felt to be suitable, and several alternative procedures were investigated. A new method has been developed and an Autocode programme has been prepared for a Ferranti Mercury computer.

## 2. STANDARD MAXIMISATION METHODS

### 2.1 General Remarks

There are many fields in which the maximum or minimum value of a function of several variables has to be found. Nevertheless the standard text-book methods are of limited practical application in P.F.R. work and not readily adapted for high-speed computation. The difficulties arise because the "general shape" of the function has to be predicted from its "local" properties, for example the derivatives at a point.

This prediction involves making certain assumptions about the nature of the function, for example, that it is approximately quadratic. Near to the optimum such assumptions may hold approximately but they may be misleading unless accurate initial estimates are available.

/For

\* Obviously, the same method can be used for minima.

For example, a simple iterative method for finding the maximum of a function of a single variable,  $f(u)$ , is Newton's method. If  $u_0$  is an initial estimate, a better estimate may be given by

$$u_1 = u_0 - f'(u_0)/f''(u_0) \quad (3)$$

One way of looking at this process is to consider the parabola  $y = B - A(u - \bar{u})^2$  which touches  $y = f(u)$  at  $u_0$ , having the same first and second derivatives there.

$$-2A(u_0 - \bar{u}) = f'(u_0)$$

$$-2A = f''(u_0)$$

$$\therefore \bar{u} = u_0 - f'(u_0)/f''(u_0) = u_1$$

Thus  $u_1$  is the maximum of the parabola, and should be a better approximation than  $u_0$  to  $u_m$ , the maximum of  $f$ , if the parabola is a good "representation" of  $f(u)$ . Thus :

(a) If  $f''(u)$  does not change very much,  $u_1$  will be closer to  $u_m$ .

However, as Figure 1 indicates, there are other possibilities.

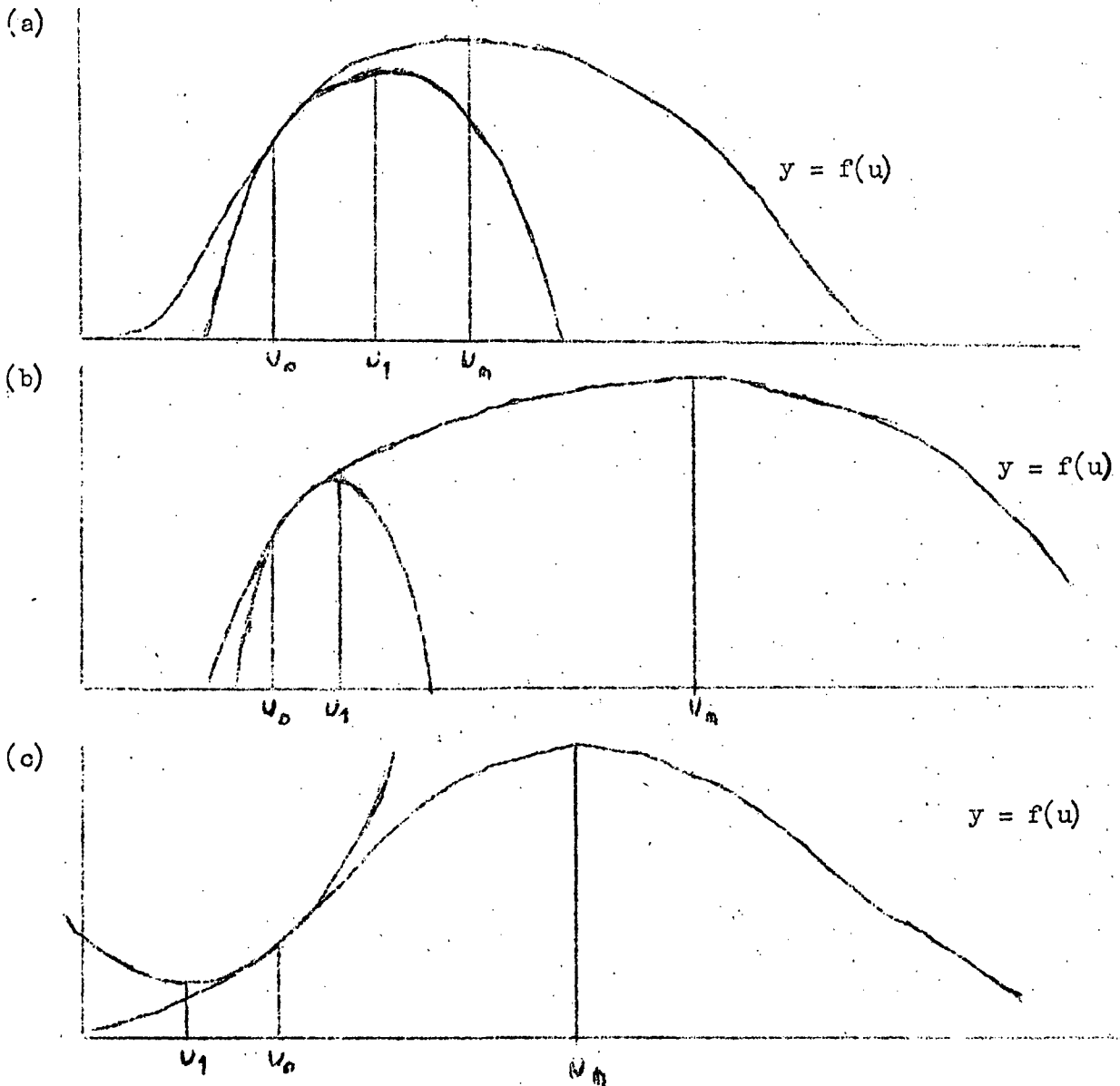


Figure 1

Thus :

- (b) If  $f''(u)$  changes rapidly,  $u_1$  may not be much nearer than  $u_0$ .
- (c) If  $f''(u_0)$  is positive, the parabola will have a minimum and not a maximum, so that  $u_1$  will be further away from the desired value.

We cannot be sure that the initial estimates available will be very near to the optimum. We need a method which will be generally applicable to any function which is known to have a true maximum somewhere in the region where it can be computed. The method must therefore be based on a minimum of a priori assumptions.

## 2.2 The Newton-Raphson Method

The well-known multi-dimensional analogue of the method discussed in 2.1, by which a quadratic form having the same first and second derivatives is fitted, is subject to similar limitations. A quadratic form

$$y = B + \sum_i \sum_j A_{ij} (u_i - \bar{u}_i) (u_j - \bar{u}_j) \quad (4)$$

has a maximum at  $(u_i)$  if  $(A_{ij})$  is a negative definite matrix. Thus, if  $\frac{\partial^2 f}{\partial u_i \partial u_j}$  is negative-definite at a point  $(u_i^0)$ , a new point  $(u_i^1)$ , obtained by solving the equations

$$\sum_i \frac{\partial^2 f}{\partial u_i^0 \partial u_j^0} (u_i^0 - u_i^1) = \frac{\partial f}{\partial u_j^0}, \quad j = 0, 1, \dots, (s-1) \quad (5)$$

will be closer to the maximum.

However, unless the initial estimates  $(u_i^0)$  are fairly good, the matrix of second derivatives may be indefinite, positive-definite or ill-conditioned. It may then be necessary to use the third derivatives, in effect fitting a cubic function, to reach a new starting point for the process.

The first and second derivatives have to be found in this process, and a matrix inverted. For large numbers of variables this can be quite slow, especially if the differentiation has to be performed numerically.

## 2.3 The Alternating Variable Method

This method involves the evaluation of the function at a series of points. One variable at a time is altered, keeping the others fixed. Each time, the highest value of the function is found by trial and error. Thus the estimate point moves parallel to each axis in turn, changing direction when the highest point is reached.

The method, which makes few assumptions about the function, has a defect which illustrates another general difficulty of optimisation. This is the problem of the "ridge" or "tunnel". The shape of the surface will usually be such that the contours will not be circular (or hyperspherical for more than two variables) but will be elongated in some direction. Unless this direction happens to be parallel to one of the axes, the process described will zig-zag backwards and forwards across a slightly inclined ridge, each step being very small, and tending to get smaller. (See Figure 2).

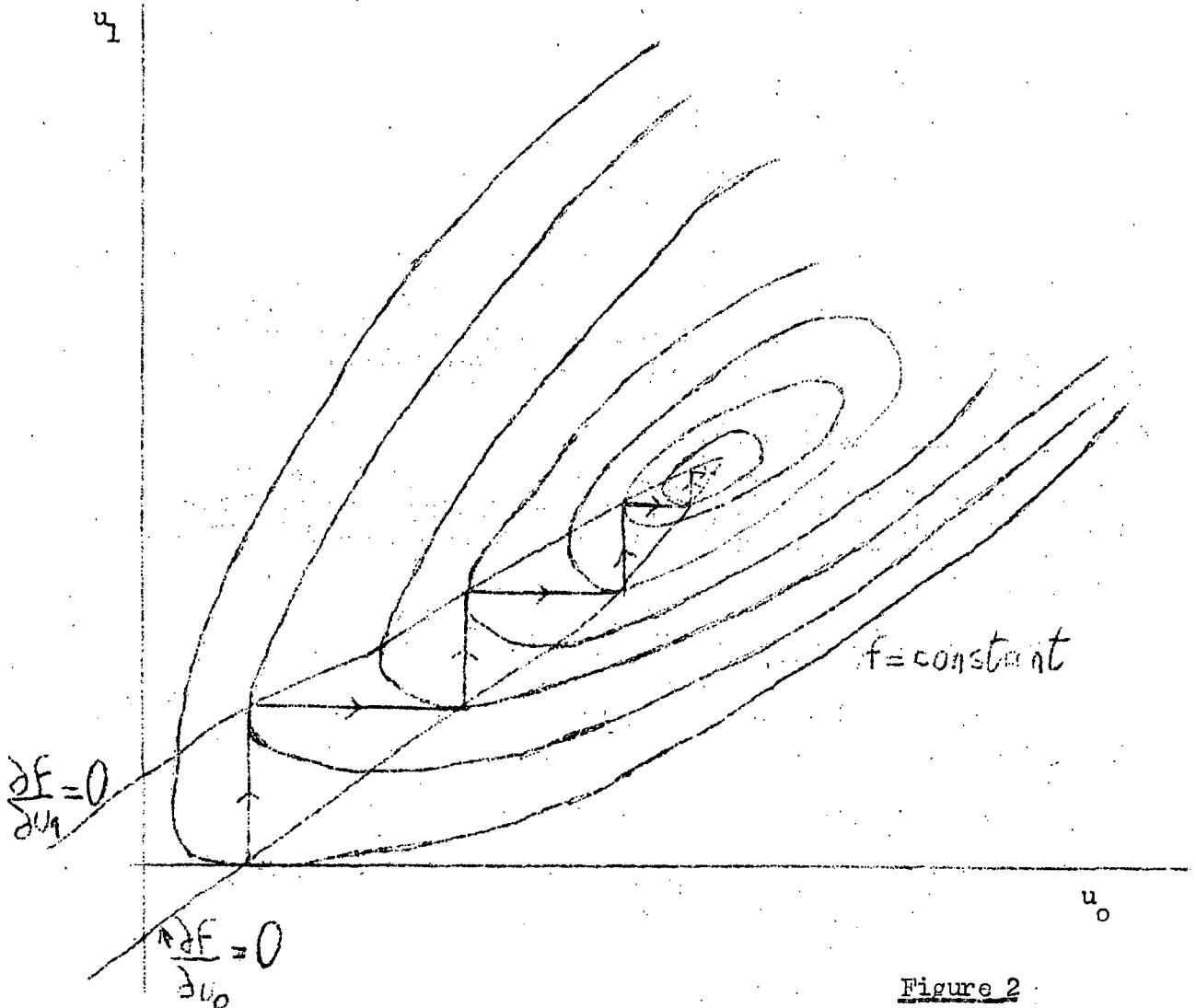


Figure 2

The process will be more inefficient, the more variables are introduced.

Another way of looking at this process is to consider the hypersurfaces given by

$$\frac{\partial f}{\partial u_i} = 0, \quad i = 0, 1, \dots, (s - 1)$$

These intersect at the optimum point. The path taken moves parallel to the axis of  $u_0$  until meeting  $\frac{\partial f}{\partial u_0} = 0$ , then parallel to the axis of  $u_1$  until  $\frac{\partial f}{\partial u_1} = 0$  and so on. If the contours are circular, each hypersurface is respectively normal to its corresponding axis, and the path will then reach its objective in  $s$  steps. Generally, the contours will be elongated and the path will zig-zag between the hypersurfaces with decreasing steps.

#### 2.4 The Method of Steepest Ascent

Another, more sophisticated, method uses the first derivatives of the function to find the direction of the steepest ascent, i.e., the gradient. The variables are altered simultaneously, each in proportion to the corresponding first derivative. New derivatives are found when the highest point along this line has been reached. Rosenbrock (1) has pointed out that for two variables, this process is, in fact, identical with the alternating variables process. For each direction starts by being perpendicular to a contour and ends when it is parallel to a higher one. Successive directions are

/therefore

therefore orthogonal, although for more than two variables, three successive directions may not be mutually orthogonal.

The elongation of the contours will thus lead to oscillation, as in 2.3. It is possible to "damp" this by retreating from the top-most point in each direction. Booth (2) suggests moving nine-tenths of the full distance for four stages and then taking a full step. Even with this refinement, the method is not entirely satisfactory, and in any case requires the evaluation of derivatives.

## 2.5 Rosenbrock's Method

Rosenbrock (1) describes a method which does not involve differentiation, and which is designed to overcome the problem of oscillation. This performs the "alternating variables" process for  $s$  orthogonal directions, re-defining these directions after each complete stage as follows.

Suppose the  $j$ th of the orthogonal directions at a particular stage is given by the unit vector  $(x_{ij}^0)$ ,  $i, j = 0, 1, \dots, (s-1)$ , and that the path taken in the course of this stage involves moving a distance  $v_j$  in the  $j$ th direction. Let

$$y_{ij} = \sum_{k=j}^{s-1} v_k x_{ik}^0, \quad i, j = 0, 1, \dots, (s-1).$$

Then  $(y_{ij})$  is the vector joining the  $j$ th point of the path to the final point at this stage. The orthogonal directions for the next stage are given by

$$z_{i0} = y_{i0}$$

$$x_{i0} = z_{i0} / \sqrt{\left\{ \sum_{k=0}^{s-1} z_{ko}^2 \right\}}$$

$$z_{i1} = y_{i1} - x_{i0} \sum_{k=0}^{s-1} y_{k1} x_{k0}$$

$$x_{i1} = z_{i1} / \sqrt{\left\{ \sum_{k=0}^{s-1} z_{k0}^2 \right\}}$$

$$z_{in} = y_{in} - \sum_{m=0}^{n-1} x_{im} \sum_{k=0}^{s-1} y_{kn} x_{km}$$

$$x_{in} = z_{in} / \sqrt{\left\{ \sum_{k=0}^{s-1} z_{k0}^2 \right\}}$$

$(x_{i0})$  is thus in the direction which yielded the greatest advance on the previous stage,  $(x_{i1})$  the next most successful direction orthogonal to it, and so on. The result is that the directions will tend to line up with the axes of the contours, thus minimising oscillation.

The method can also be thought of as follows. The path taken moves in a direction  $(x_{i0})$  until meeting the hypersurface given by

$$\sum_{i=0}^{s-1} x_{i0} \frac{\partial f}{\partial u_i} = 0. \quad \text{There are } s \text{ hypersurfaces like this and they}$$

intersect at the maximum point. The next step is in the direction

$(x_{i1})$

$(x_{i1})$  and ends with  $\sum_{i=0}^{s-1} x_{i1} \frac{\partial f}{\partial u_i} = 0$ . This step will move out of the hypersurface  $\sum_{i=0}^{s-1} x_{i0} \frac{\partial f}{\partial u_i} = 0$  unless the latter is orthogonal to the direction  $(x_{i0})$ , and the condition for this to be true is

$$x_{k0} = \lambda \sum_{i=0}^{s-1} x_{i0} \frac{\partial^2 f}{\partial u_i \partial u_k}, \quad k = 0, 1, \dots, (s-1)$$

Rosenbrock's method changes the orthogonal directions  $(x_{ij})$  until they satisfy the conditions  $x_{kj} = \lambda_j \sum_{i=0}^{s-1} x_{ij} \frac{\partial^2 f}{\partial u_i \partial u_k}$ , i.e., until the  $(x_{ij})$  become the eigen-vectors of the matrix  $\frac{\partial^2 f}{\partial u_i \partial u_i}$  or the principle axes of the quadratic which touches the contour.

The distance to be moved in a given direction is determined by Rosenbrock as follows. A step  $v$  is made and the function evaluated at the new point. If it is less than the previous value,  $v$  is replaced by  $-0.5 v$  and the process repeated from the original point. When the new value is greater than or equal to the old one,  $v$  is replaced by  $3 v$  and the process repeated until a point is reached which gives a lower value of the function. The best point is then taken as the starting point for the next direction.

This method makes no assumptions about the function. It is particularly suitable for adaptation when the variables are limited to lie within a given region. On the other hand, quite a large amount of "organisation" is involved in calculating the directions in which to proceed.

### 3. DEVELOPMENT OF THE STANDARD METHODS

Using Mercury Autocode, several examples of the methods described were examined. Various modifications in them were tried out, some of which led to the method finally adopted.

#### 3.1 The Maximum on a Given Line

The methods described in 2.3, 2.4 and 2.5 all involve estimating the best point along a given line. The procedure eventually adopted was as follows.

Let  $(u_i^0)$  be the starting point and  $f_0 = f(u_i^0)$ . If  $(\delta_i)$  is a vector along the given line, evaluate

$$f_1 = f(u_i^0 + \delta_i) \quad \text{and}$$

$$f_2 = f(u_i^0 + \frac{1}{2} \delta_i)$$

If  $f_2 \geq f_1$  and  $f_2 > f_0$ , the optimum is estimated as the apex of the parabola through the three values. Thus the estimate would be

$$(u_i^0 + \delta_i \left[ \frac{1}{2} + \frac{f_0 - f_1}{4(f_2 - f_0 - f_1)} \right])$$

If  $f_0 \geq f_2$  or  $f_1 > f_2$  or  $f_0 = f_1 = f_2$ , and  $f_1 \geq f_0$ ,  $(\delta_i)$  is

/replaced

replaced by  $(2 \delta_i)$ , and the process repeated. If  $f_0 > f_1$ ,  $(u_i^0)$  is replaced by  $(u_i^0 + \delta_i)$  and  $(\delta_i)$  by  $(-\delta_i)$ . Thus the line is doubled in length until the required configuration is reached. Eventually an estimate will be obtained, unless there are constraints on the variables. This question is discussed below. A limit is also placed on the number of times the "stretching" procedure is repeated, in case  $f$  is constant along a given line,

### 3.2 Step Size in the Steepest Ascent Method

The length of the vector  $(\delta_i)$  to be used in the method of steepest ascent (2.4), can be made to depend on the slope at the initial point. If  $\delta_i = v \frac{\partial f}{\partial u_i} / \sqrt{\left(\frac{\partial f}{\partial u_i}\right)^2}$ , then the "vertical" rise along the tangent from  $(u_i^0)$  to  $(u_i^0 + \delta_i)$  is  $\sum \delta_i \frac{\partial f}{\partial u_i} = v$ .

The expected change in  $f$  from  $(u_i^0)$  to the new estimate will thus be of the order of  $v$ .  $v$  can then be made to diminish in the course of the iteration.

### 3.3 Oscillation About a Ridge

The approach of Rosenbrock may be combined with the steepest ascent method in order to avoid the oscillation due to elongation of the contours mentioned in 2.3.

The vector sum of two such oscillating steps will tend to be parallel to the axis of the ridge. Two stages of steepest ascent may thus be followed by a step along the direction of the total movement so obtained.

The optimum point on this line is then used to start the next pair of steepest ascent stages.

### 3.4 An Adaptation of Rosenbrock's Method

Some experiment with the methods referred to above suggested a procedure in which the direction of the path at each stage is the vector sum of the previous  $s$  steps. This is equivalent to Rosenbrock's method, without the directions of successive steps being necessarily orthogonal. It therefore is much simpler to carry out. This procedure did result in a direction being found, parallel to the ridge, along which considerable progress could be made. However, the subsequent steps all had large components in this direction, and would consequently tend to be nearly parallel to each other. The process then had to be restarted.

This method has some relation to a procedure for solving simultaneous equations discussed in a paper by Kincaid (3). This approximates to the solution through a sequence of linear approximations along the edges of polyhedra which converge to a point.

## 4. A NEW METHOD

### 4.1 Basic Idea

With the exception of the Newton-Raphson method, none of the processes discussed will reach the maximum without several iterations, even if the function is a quadratic. Rosenbrock reports on a comparison of his method with other results quoted by Booth in finding the minimum of

$$(u_0 + 2u_1 - 7)^2 + (2u_0 + u_1 - 5)^2$$

/starting

starting from  $u_0 = u_1 = 0$ . The required answer is  $u_0 = 1$ ,  $u_1 = 3$ .

In 5 stages, the steepest descent method reached  $u_0 = 1.04$ ,  $u_1 = 2.97$ , while Rosenbrock's method gave  $u_0 = 1.017$ ,  $u_1 = 2.991$ .

A procedure which arrived at the true optimum in one stage if the function were quadratic would appear to have some advantage.

Consider a negative-definite quadratic form. The maximum point lies at the centre of a set of hyper-ellipsoidal contour surfaces. A set of parallel lines will touch these contours at points lying in the hyper-plane conjugate to the lines with respect to the hyper-ellipsoids, and this hyper-plane will pass through the central point.

Thus if  $f = B + \sum \sum A_{ij} (u_i - \bar{u}_i) (u_j - \bar{u}_j)$ , a line in the direction  $(w_i)$  will touch a contour at a point for which  $\sum \sum A_{ij} (u_i - \bar{u}_i) w_j = 0$ . (See Figure 3.)

This suggests a procedure for finding the maximum. Take two lines parallel to a direction  $(w_i^0)$ , and find the optimum point on each. Call the direction of the line joining them  $(w_i^1)$ . The best point on this line,  $B_1$ , is the optimum for the whole plane.

Now choose a line parallel to  $(w_i^1)$  in another plane and find the maximum point  $A_1$  on it. Moving from  $A_1$  in the direction  $(w_i^1)$  will reach the optimum point in the second plane,  $A_2$ . The line  $A_2 B_1$ , direction  $(w_i^2)$ , will pass through the optimum in the whole solid.

The maximum of an  $s$ -dimensional quadratic will thus be reached in  $\frac{1}{2}s(s+1)$  steps. This will also apply to any function whose contours form a family of similar hyper-ellipsoids.

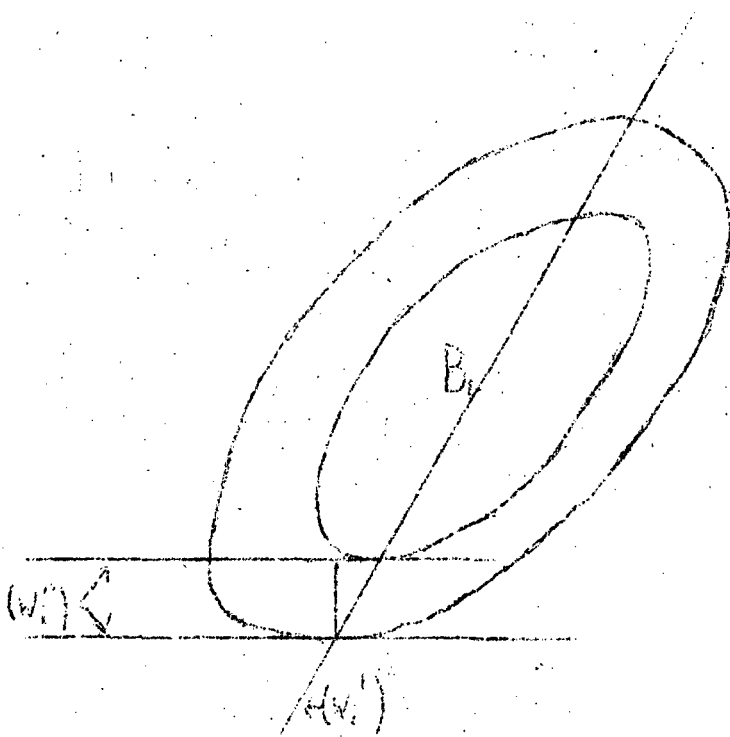


Figure 3

In general, the points where lines in a direction  $(w_i)$  touch the contours:  $f(u_0 \dots u_{s-1}) = \text{constant}$  will satisfy:

$$\sum_i w_i \frac{\partial f}{\partial u_i} = 0$$

The normal to this locus will be in the direction:  $(\sum_i w_i \frac{\partial^2 f}{\partial u_i \partial u_j})$ .

If there is only one maximum, the repetition of the  $\frac{1}{2}s(s+1)$  steps described above will converge to it.

The other processes described require only  $s$  steps at each stage. However, if  $s$  is not very large, this should be compensated by a reduction in the number of iterations needed by the new procedure.

#### 4.2 Procedure Adopted

The implementation of this method involves choosing (a) the direction  $(w_i^0)$ , (b) the "displacements" from the initial line in this direction to the second line, from the initial plane to the new plane and so on, and (c) the increment  $(\delta_i)$  used in finding the optimum point along a line; (see 3.1).

The simplest procedure seemed to be to take  $(w_i^0)$  parallel to the axis of  $u_i$ . The displacement could then be obtained by adding a quantity  $w$  successively to  $u_1, u_2$  and so on.

Thus only the first co-ordinate of  $(w_i^0)$  is used, only the first two of  $(w_i^1)$  and so on.

If a complete cycle has led to a decrease in the value of the function, the process re-starts from the best point obtained so far and  $w$  is decreased. When  $w$  falls below a prescribed value, the process ends.

#### 4.3 Constraints on Variables

Rosenbrock's programme deals with constraints on the variables by modifying the function in a narrow boundary region. This complex procedure seems to be effective for quite complicated boundaries.

The method adopted in the new programme is much simpler. It will work for any convex region (i.e., one such that any point on the line joining two admissible points will be admissible) and is quite effective if the constraints are linear. It may not lead to the true optimum if this lies on a boundary which is concave.

The procedure is as follows. When finding the maximum along any line, each new point is tested to see if it is outside the permitted region, unless it lies between two already tested points. If a new point is found to be inadmissible, successive tests are made at equally spaced points on the line, moving back towards the last found admissible point.

The first point found inside the boundary is then taken as the estimate of the optimum.

#### 4.4 Computer Programme

A sequence of instructions in Mercury Autocode has been prepared embodying the procedure described. It occupies 429 registers, just over half a chapter.

The chapter may be completed by writing a sequence which will calculate the function of  $u_0, u_1 \dots u_{s-1}$  and leave it in the

location  $f$ . Further instructions are required to specify the limits on the variables and to print the results.

If query printing is not suppressed, intermediate results will be printed, showing the values of  $w$  if it is reduced,  $u_0, u_1 \dots u_{s-1}$  and  $f$ .

Before entering the maximising sequence, the initial values of  $u_0, u_1 \dots u_{s-1}, s$  and  $w$  must be specified. The quantity  $x'$  must also be given, such that the process terminates when  $w$  is reduced below this value.

Details of the sequence are given in the Appendix.

5. EXAMPLES

In most applications, the evaluation of the function will take much more time than the "organisation" of the maximising process. The number of times the function has to be evaluated in order to reach the optimum is thus the best criterion to use in comparing optimisation methods.

Two examples given by Rosenbrock were used to make such a comparison with the new method.

(a) Rosenbrock describes results obtained in minimising the function:

$$u(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

In our notation, this is equivalent to maximising :

$$f = -100(u_1 - u_0^2)^2 - (1 - u_0)^2.$$

There are no constraints on the variables.

This function has a very steep "ridge" along the parabola:  $u_1 = u_0^2$ . This is thus a most unfavourable example for either the Rosenbrock method or the new method described in Section 4. The optimum is achieved at  $u_0 = u_1 = 1, f = 0$ .

In the form of his method finally selected Rosenbrock took 21 stages (one incomplete), involving 200 evaluations of the function, to reach the value:  $u_0 = 0.995, u_1 = 0.991, f = 0.000022$ , starting from  $u_0 = -1.2, u_1 = 1$  and moving round the "ridge".

The new method was slightly inferior to this, if the number of evaluations is used as a criterion. After seventeen stages requiring 223 evaluations, the results were

$$u_0 = 0.99561, u_1 = 0.99079, f = 0.000039$$

The other methods quoted by Rosenbrock prove far less efficient in this example.

(b) To illustrate the imposition of limits on the variables Rosenbrock uses the problem of finding the rectangular solid with maximum volume whose length and combined length and girth are restricted. In our notation, this means maximising  $f = u_0 u_1 u_2$  where

$$\begin{aligned} 0 &\leq u_0 \leq 42 \\ 0 &\leq u_1 \leq 42 \\ 0 &\leq u_2 \leq 42 \\ 0 &\leq u_0 + 2u_1 + 2u_2 \leq 72 \end{aligned}$$

In fact, the maximum value is  $24 \times 12 \times 12 = 3456$ , and the first three constraints are inoperative.

Starting at  $u_0 = u_1 = u_2 = 10$ , Rosenbrock's method reached  $23.9873 \times 12.0026 \times 12.0017 = 3455.4$  after 600 evaluations.

The new method in 10 stages and 239 evaluations gave  $23.801 \times 12.003 \times 12.096 = 3455.8$ .

## 6. DISCUSSIONS AND CONCLUSIONS

The relationship between steepest ascent, Rosenbrock's method and the new method is best viewed in connection with the way in which the direction of change of the variables is determined at each stage.

The steepest ascent method takes account only of the slope of the tangent plane at the starting point. It is therefore only dependent on the variation of the function in the immediate vicinity.

Both Rosenbrock's method and the new method 'survey' a larger region. Whereas the former tends to move in the direction of the longest principal axis, the system of contours in that region, the latter aims at the centre.

It is difficult to generalise about the efficiency of optimisation procedures, when the function to be optimised is completely unspecified. Rosenbrock's approach will deal with the most general type of constraints on the variables, but requires a rather complicated process to do so.

The new method, although it will need more evaluations per cycle, will take fewer cycles to reach the optimum (a) if the function is similar to a quadratic in the region concerned or (b) if the constraints are linear.

Rosenbrock's method is undoubtedly superior to all those previously described for general use. However, it is fairly lengthy both in time of operation and machine space.

The new method will be more suitable for many of P.F.R.'s needs, unless the number of parameters is very large. This is because the constraints on our parameters are generally linear and in any case the maxima we require are not usually going to lie on the boundary.

Further improvement in both methods is undoubtedly possible and will be attempted when more experience in their use has been obtained.

The new method has already been employed on a number of problems. These include the estimation of the parameters in a model for the overlap of dust particles on thermal precipitator slides; the fitting of size distributions of dust particles; and the analysis of radiological classification of pneumoconiosis.

## 7. ACKNOWLEDGEMENTS

I should like to acknowledge the valuable assistance of Miss F. Rastall and Miss J. Buckingham.

THE MERCURY AUTOCODE SEQUENCE "MAXIMIZE"

The sequence finds the value of  $u_0, u_1 \dots u_{s-1}$  which make  $f(u_0, u_1 \dots u_{s-1})$  a maximum. It occupies 429 registers.

Initial values of  $u_0, u_1 \dots u_{s-1}$ , and  $s, w$  and  $x'$  must be set before entry.  $x'$  will be of the order of the accuracy required in the  $u$ 's.  $w$  should initially be about 1000 times as great.

The sequence uses :

$$f_0, f_1, f_2, f_3$$

$$u_0, u_1 \dots u_{\frac{1}{2}(s+3)-1}$$

$$v_0, v_1 \dots v_{s-1}$$

$$w_0, w_1 \dots w_{s-1}$$

$$p, q, r, s, t, u, w, x, x', f.$$

Labels 101 to 120.

Any of these quantities may be changed by the sequence, except  $s, x$  and  $x'$ .

Directives for  $u, v, w$  must be provided.

Directive  $f \rightarrow 3$  is included in the sequence.

Auxiliary sequences:

A sequence labelled 1) must evaluate the function

$$f = f(u_0, u_1 \dots u_{s-1})$$

and is followed by "return".

A sequence labelled 3) tests the values of  $u_0, u_1 \dots u_{s-1}$  (without changing them) to see if they are admissible. If they are, jump to "return", otherwise  $r = 1$ , then "return". A sequence labelled 2) will be jumped to when the process terminates.  $u_0, u_1 \dots u_{s-1}$  and  $f$  will contain the appropriate optimum values.

Enter the sequence at label 101).

Query Printing

If query printing is not suppressed, intermediate results will be printed as follows :

Index  $r$   $r = 1, 2 \dots s$

New value of  $w$ , if  $w$  has been reduced.

$u_0$

$u_1$

$\vdots$

$u_{r-1}$

} first  $r$  co-ordinates of new point. ( $u_r \dots u_{s-1}$  remain unchanged.)

Value of function.

#### REFERENCES

- (1) Rosenbrock, H. H. "An automatic method for finding the greatest or least value of a function".  
Computing Journal, October, 1960.
- (2) Booth, A.D. "Numerical Methods", p. 98.
- (3) Kincaid, W.M. "A two point method for the numerical solution of systems of simultaneous equation".  
Quarterly of Applied Mathematics, XVIII, No. 4  
January, 1961, p. 313.

**HEAD OFFICE:**

Research Avenue North,  
Riccarton,  
Edinburgh, EH14 4AP,  
United Kingdom  
Telephone: +44 (0)870 850 5131  
Facsimile: +44 (0)870 850 5132

Tapton Park Innovation Centre,  
Brimington Road, Tapton,  
Chesterfield, Derbyshire, S41 0TZ,  
United Kingdom  
Telephone: +44 (0)1246 557866  
Facsimile: +44 (0)1246 551212

Research House Business Centre,  
Fraser Road,  
Perivale, Middlesex, UB6 7AQ,  
United Kingdom  
Telephone: +44 (0)208 537 3491/2  
Facsimile: +44 (0)208 537 3493

Brookside Business Park,  
Cold Meece,  
Stone, Staffs, ST15 0RZ,  
United Kingdom  
Telephone: +44 (0)1785 764810  
Facsimile: +44 (0)1785 764811

**Email: [iom@iom-world.org](mailto:iom@iom-world.org)**