EXPERIMENTAL PROCEDURES FOR FINDING OPTIMUM CONDITIONS

by

BASIL GRAHAM FREDERICK SPRINGER

A Thesis submitted for the Degree of

Doctor of Philosophy

in the

University of London

Imperial College, 1968.

ABSTRACT

The purpose of this thesis is to develop and discuss automatic procedures for finding optimum conditions when the response is subject to random error.

Consider a system, which consists of k quantitative factors each of which may be controlled and measured at any given level within a continuous range. In this system the observed response at each combination of factor levels is subject to random variation. The particular aim is to estimate the factor combination at which the true response is optimum.

In Chapter 1 the problem is defined and the principles underlying a good sequential search procedure are discussed.

The literature of existing single factor procedures is reviewed in Chapter 2. New procedures are then discussed and a comparison is made with an existing procedure. The difficulty encountered in determining the behaviour of the procedures analytically is explained and results are obtained by computer simulation.

The merits of existing multi-factor procedures are discussed in Chapter 3, and these ideas are extended to develop two new multi-factor procedures. These new procedures are compared by simulation.

No attempt is made to give elaborate tables from which one can precisely determine the merits of a given procedure for a particular situation. Instead, the procedures are compared for a few simple cases and their potential usefulness discussed.

The description of the new procedures is aided by diagrams. Flow diagrams for all the procedures discussed are given in the Appendix.

ACKNOWLEDGEMENTS

The author is deeply grateful to Professor D.R. Cox for suggesting the topic and for his guidance and encouragement in the preparation of the thesis. Thanks are due to Dr. Agnes Herzberg for her helpful comments and to my wife, Kean, for her patience in typing the thesis.

Thanks are also due to the Commonwealth Scholarship Commission under whose auspices the author pursued the research for the thesis.

TABLE OF CONTENTS

		Page
LIST OF DEFINITIONS	... ...	8
LIST OF FIGURES	... ...	10
LIST OF TABLES	... ...	11
CHAPTER 1 -	INTRODUCTION	12
1.1	The Response Surface	13
1.2	The Problem	15
1.3	The Principles underlying an Initial Stage Sequential Procedure	18
1.4	Deterministic Procedures	23
CHAPTER 2 -	SINGLE FACTOR PROCEDURES	26
2.1	Existing Procedures	27
2.1.1	The Hotelling Procedure	27
2.1.2	The Box-Wilson Procedure	28
2.1.3	The Kiefer-Wolfowitz Procedure ...	28
2.1.4	The Normalized Kiefer-Wolfowitz Procedure	30

CHAPTER 2 (cont'd) Page

2.1.5 The Accelerated Kiefer-Wolfowitz Procedure 31

2.1.6 Discussion 32

2.2 New Procedures 34

2.2.1 The Up-Down Procedure 34

2.2.2 The Generalized Up-Down Procedure ... 44

2.2.3 Results and Discussion 54

CHAPTER 3 - MULTI-FACTOR PROCEDURES 84

3.1 Existing Procedures 85

3.1.1 Non-Sequential Procedures 85

3.1.2 The One-at-a-time Procedure 87

3.1.3 The Box-Wilson Procedure 88

3.1.4 Extensions of the Kiefer-Wolfowitz Procedure 89

3.1.5 The Spendley-Hext-Himsworth Simplex Procedure 90

3.1.6 The Nelder-Mead Simplex Procedure ... 92

3.1.7 Discussion 94

CHAPTER 3 (cont'd) Page

3.2 New Procedures 96

3.2.1 The Generalized Simplex Procedure ... 96

3.2.2 The Duplex Procedure 106

3.2.3 Results and Discussion 117

CHAPTER 4 - CONCLUSION 140

APPENDIX - FLOW DIAGRAMS 144

A Flow Diagram for Routine 3 ... 145

A Flow Diagram for Routine 4 ... 146

A Flow Diagram for Routine 5 ... 147

A Flow Diagram for Routine 6 ... 148

REFERENCES 149

LIST OF DEFINITIONS

	Page
OFC - Optimal factor combination	13
NSR - Near stationary region	17
OFL - Optimal factor level	27
K-W - Kiefer and Wolfowitz	28
Estimator (A), Estimator (B), Estimator (C) — Estimators of the OFL for Routines 1A and 1B	37
Routine 1A - A version of the Up-Down procedure	38
Routine 1B - Another version of the Up-Down procedure	38
Routine 2A - A version of the generalized Up-Down Procedure	51
Routine 2B - Another version of the generalized Up-Down Procedure	53
Routine 3 - A combination of Routine 2B and Routine 1A	72
Routine 4 - The normalized version of the accelerated Kiefer-Wolfowitz procedure	72

DEFINITIONS (cont'd) Page

HOE - Highest observation estimate of the OFC 103

WOE - Weighted observation estimate of the OFC 103

Routine 5- The generalized simplex procedure 106

Routine 6- The duplex procedure 116

LIST OF FIGURES

Figure		Page
1	The Up-Down Procedure	35
2	The Generalized Up-Down Procedure	47
3	The Performances of Routines 1A and 1B on a Quadratic Response Function	67
4	The Response Function $\eta = \exp(-x^2)$	75
5	The Generalized Simplex Procedure	100
6	The Duplex Procedure	111
7	The Pattern of Designs for Routines 5 and 6	119
8	The Performances of Routines 5 and 6 on a Linear Gradient	125

LIST OF TABLES

Table		Page
1	The Values of the Bias for Estimators of Var $\left\{ \hat{\theta}^{A} (R) \right\}$. (Routine 1a)	.42
2	The Estimation of the OFL for a Quadratic Response Function. ($-\beta_{11}/\sigma = 5.0$)	68
3	Values of h/σ associated with the Response Function $\eta = \exp(-x^2)$	76
4	Comparison of Routine 3 and Routine 4	78
5	Comparison of Routine 5 and Routine 6	137

CHAPTER 1

INTRODUCTION

1.1 The Response Surface.

In the field of design and analysis of experiments, one problem sometimes of interest is to determine the set of treatments which yield optimum conditions. It is with this that we shall be concerned.

We consider k quantitative variables (or factors) $x_1, x_2, \ldots, x_k$, each capable of exact measurement and control at any given level over a continuous range of values. The u^{th} combination of factor levels (or the u^{th} experimental unit) is denoted by $(x_{1u}, x_{2u}, \ldots, x_{ku})$ and the k dimensional space of all the experimental units of potential interest is called the experimental region. A finite group of points in the experimental region is called a design.

Associated with the experimental region is a mathematical function η, where $\eta = \eta(x_1, x_2, \ldots, x_k)$, called the response surface and its value for any $(x_{1u}, x_{2u}, \ldots, x_{ku})$ gives the corresponding true response η_u at the u^{th} combination of factor levels.

Suppose that it is desired to investigate the nature of an unknown response surface over the experimental region and in particular to locate the optimal factor combination (OFC), i.e. the combination of factor levels which maximizes (or sometimes minimizes) the response function. For this purpose we make

independent observations of the response at points in the experimental region. We may also wish to determine the behaviour of the function in the neighbourhood of the OFC.

We shall assume that these observations are subject to random variation such that for any given $(x_{1u}, x_{2u}, \ldots, x_{ku})$ the observed response, y_u will be represented by

$$y_u = \eta_u + e_u$$
$$= \eta (x_{1u}, x_{2u}, \ldots, x_{ku}) + e_u ,$$

where e_u is the experimental error of the u^{th} observation. We assume that y_u is normally distributed with expected value η_u and variance σ^2, independent of u. This is called the stochastic model and is to be distinguished from the purely deterministic model when there is no experimental error, i.e. $\sigma = 0$.

The function η depends, of course, also on a number of parameters which determine the shape of the surface and may be a complicated function of these parameters. One way to investigate the shape of the surface in a region of interest and hence to find an estimate of the OFC, is to assume a parametric form and to estimate the parameters. If η is assumed to be continuous with continuous derivatives over a restricted region then it is possible, over this region, to approximate η by a polynomial thus:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{11} x_1^2 + \ldots + \beta_{12} x_1 x_2 + \ldots$$

including terms up to the d^{th} degree, say, where the notation is that of Box and Wilson (1951). Our problem is then reduced to estimating the β's, in this linear function of the parameters, to which problem we can apply standard least squares techniques. There are some cases where prior information may suggest the use of a non-polynomial approximation to the response surface. In such cases it is usually necessary to use iterative techniques to estimate the parameters.

1.2 The Problem.

Most examples in the literature, concerning response surface methodology, are taken from the chemical industry, but the discussion and references in Hill and Hunter (1966) give examples over a wider field. In general the experimenter wants to know the form of the surface near the OFC, to find whether a small change in the optimum level of a factor will seriously affect the optimum true response or to find to what extent the factors affect the response independently. We assume that the experimenter observes only one response per experiment. Box (1954) discusses factor dependence and some of the types of surface which occur in practice. Hill and Hunter (1966) discuss the multiple response problem.

Once the factors to be included in the investigation,

and the experimental region have been decided, it is then wise to make use of any other relevant information which may have accumulated from past experience. Examples of such information are:

(i) A rough estimate of the position of the OFC which will indicate where the search should begin.

(ii) The expected nature of the surface in the experimental region, which will be a warning if peculiar behaviour is expected.

(iii) The magnitude of the experimental error.

The problem is now treated as a statistical one based on the mathematical model described in Section 1.1 and our aim is to obtain an approximation to η in the region of the OFC. We shall assume throughout, without loss of generality, that the OFC maximizes (not minimizes) the response function. We assume also that the function is smooth and has a unique turning value in the experimental region.

There are two main strategies that can be adopted; either the whole of the experimental region can be explored systematically or, starting at some initial estimate of the OFC, the region can be explored sequentially, using a sequence of

designs, basing the current design on information obtained from responses previously observed. Extensive exploration of the whole of the experimental region can be very uneconomical, particularly when the number of factors is large. In general, therefore, sequential exploration is preferable.

The search for optimum conditions can be thought of as taking place in two stages. Particular attention will be given to sequential procedures suitable for the initial stage.

In the initial stage the region of the starting position is explored in search of a direction of ascent. The sequence of designs then moves towards the OFC and finally settles over a region around the OFC. The procedure should be such that it is expected that a low degree polynomial is a good approximation to the unknown function over this region. We shall call such a region a near-stationary region (NSR). An estimate of the OFC is made in the NSR.

In the final stage of the search the NSR is thoroughly investigated using second order designs. With the aid of canonical analysis (Davies, 1963, p. 522), the important features of the surface in the region of the OFC can be presented in a meaningful, easily understood way if a good second degree polynomial approximation has been obtained. This technique is particularly useful if $k \geq 3$. If the second

degree polynomial approximation is unsatisfactory then a third degree polynomial is fitted. References which discuss second and third order designs can be found in Hill and Hunter (1966). However, the increase in the number of parameters to be estimated as d is increased from p to $p+1$, say, is

$$\binom{k+p+1}{p+1} - \binom{k+p}{p} = \binom{k+p}{p+1}$$

which increases rapidly as k increases. Transformations on $x_1, x_2, \ldots, x_k$, Box and Tidwell (1961) or on the observed responses, Box and Cox (1964), can be applied. These transformations may simplify the response function and hence enable us to work with a lower degree polynomial in the transformed variables rather than a higher degree polynomial in the original variables. Alternatively, non-polynomial forms for the surface may be assumed or we may consider a region closer to the estimate of the OPC, over which a good second degree approximation may be obtained.

1.3 The Principles underlying an Initial Stage Sequential Procedure.

In this section the general principles and inherent difficulties involved in developing sequential search

procedures designed for investigation in the initial stage, are discussed.

In cases where the response surface is complex, perhaps partly due to a bad choice of scale of the factors, and even in some simpler cases, constant supervision is required to guide the sequence of designs to the optimum region. It would often be advantageous, therefore, if the procedure could be operated automatically and if it allowed for automatic adaptation of the design as changes in the response surface became evident. Such a feature would minimize the supervision required from step to step and reduce the need to review the procedure in mid-stream to make any changes in the scale of the factors. A procedure capable of doing this is called an automatic procedure. With the computer industry in the process of rapid development, there should soon be readily available the facility for an automatic procedure to be used on time sharing computers. This would require only an operator to read in data and give the necessary output information from step to step.

It is the aim of this work to define such an automatic procedure. Chapter 2 deals with existing and new procedures in the single factor case and Chapter 3 with the more interesting multi-factor case.

To give a detailed account of what our procedure is expected to achieve consider a starting position, i.e. a point $(x_1, x_2, \ldots, x_k)$ in the experimental region, which has been chosen by the experimenter usually on the basis of previous information or from an initial broad survey of the region. This starting point may in general lie in one of three types of region:

(i) A region, distant from the OFC, over which the true response increases slowly.

(ii) A region over which the true response increases rapidly. (Rapidity of increase is conveniently measured by the ratio to σ of the average increase in response for the step size contemplated).

(iii) A region in the neighbourhood of the OFC.

If we relax the assumption of unimodality we admit other types of region.

In any case the function of the procedure at this stage should be to explore the region of the starting point in an attempt to detect regions over which the response is higher. For this purpose it is necessary to estimate a direction of ascent over a region where a first degree polynomial is a good approximation to the response surface. It may not be known over what region, in the vicinity of the starting point, a first

degree polynomial adequately represents the surface. If the initial design is so big that a higher degree polynomial is required to obtain an adequate fit then the procedure should be able to detect this, and correct it,as it should if the design is so small that real effects are confounded with the experimental error and no progress can be made. On relaxing the unimodality assumption, exploration in a NSR might lead to regions of higher response, e.g. if the NSR has a saddle point.

Once we have obtained a direction of ascent our aim is to move in this direction and to choose our next design to search for a further region over which there is higher response. Since it was assumed that the linear approximation to the response surface was valid over the region of the design then it must be assumed that the estimate of the direction of ascent is valid only over the region of the design. Therefore, the next design should be placed adjacent to the first to lessen the risk of making an error in assuming that the direction of ascent is valid over a region larger than it really is. The requirement that the procedure should be operated automatically and should allow for automatic adaptation of the design as changes in the response surface become evident, suggests that the size and location of the next design should be dependent on the information contained in the responses and the design points of the

current design. This requirement implies that, in general, the efficiency of the procedure in achieving optimum conditions is not critically dependent on the initial design and on its scale. In cases where little prior information is available about factor dependence, the property of scale invariance is thus important.

Experimental error, if large, will retard the progress of the sequence of designs and sometimes it may make convergence to the optimum virtually impossible. The procedure should be designed to give the sequence of designs every chance to progress towards the optimum region, in the face of large experimental error, aiming at convergence before considering efficiency.

The procedure should include a stopping rule which would be effective when there is evidence that a NSR has been reached. It is often difficult to give an explicit rule which will stop the sequence of designs at the right time, and it may be necessary after stopping and reviewing the position, to restart and continue the search towards the optimum region.

On reaching a NSR the next step is to choose a sub-region over which a second degree polynomial adequately represents the response surface. In making this choice as

much of the existing information as possible should be used thus reducing the need for further experimentation.

It may often be the case that on its path to the optimum the sequence of designs encounters a boundary of the experimental region, i.e. a constraint on the factor space. Rules for continuing the search on encountering constraints should be considered.

To summarize the important features of an initial stage procedure, we have:

(i) Automatic operation.

(ii) Automatic adaptation to the local nature of the surface.

(iii) Scale invariance.

(iv) Non-sensitivity to large experimental error.

(v) Stopping rule.

(vi) Choice of a region where a second degree polynomial is an adequate fit.

(vii) Behaviour of the procedure on encountering constraints.

1.4 Deterministic Procedures.

In Section 1.1 the distinction was made between the stochastic case and the deterministic case. In this thesis

we deal primarily with sequential stochastic search procedures. Deterministic procedures, however, will not be completely ignored, because in designing stochastic procedures much can be learnt by first considering the problem in deterministic terms and then making the necessary modifications to adapt the procedure for stochastic search. In designing deterministic search procedures the aim is to converge strictly to the OFC but in stochastic search, because of the experimental error, convergence in the strict sense is not possible and the aim is to 'converge' on to a NSR centred about an estimate of the OFC.

In discussing sequential stochastic search procedures comparisons will be made of the performance of the procedure for different values of σ , including $\sigma = 0$, and 'convergence' will be expected to be faster the smaller σ becomes. In general, however, stochastic procedures should not be used in the deterministic case as there will usually be more efficient specially designed deterministic procedures available.

An example of a purely deterministic system, as distinct from a stochastic system with negligible experimental error, is a known but complicated mathematical function which cannot be optimized by the standard methods of the differential calculus and hence a search technique has to be developed. Wilde (1964) describes and compares several suitable search

procedures. Another motivation for work in this field is found in statistical estimation problems where one is often concerned with finding the maximum of the likelihood function when it assumes a complicated form. Nelder and Mead (1965) have developed a procedure with this application in mind. This procedure will be described in Section 3.1.6 and will be useful in developing stochastic procedures in Section 3.2. Further references may be obtained from BOX, M.J. (1966) who compares a number of procedures designed for function minimization.

CHAPTER 2

SINGLE FACTOR PROCEDURES

2.1 Existing Procedures.

In designing optimum seeking procedures it is instructive to begin with the single factor case, i.e. $k = 1$, although single factor problems are not very common in practice nor the difficulties encountered typical of what we expect to find for higher values of k.

2.1.1 The Hotelling Procedure.

Hotelling (1941) was among the first to consider the problem of the experimental determination of the maximum of a response function. Discussing the problem in the single factor case he proposed that experimentation should proceed in three phases:

(i) A speculative and unsystematic phase.

(ii) An intermediate type of exploration.

(iii) A full scale investigation.

A survey of the experimental region is made in phase (i) and a range, within which the optimal factor level (OFL) is expected to lie, is chosen. In phase (ii) a fifth degree polynomial is fitted over this range and σ, β_1, β_{11}, ..., β_{11111} are estimated. Design points, depending on these estimates, are used in phase (iii), to fit a second degree polynomial. The design is such that in fitting the second degree polynomial the cubic

bias is zero and the quartic bias is minimized.

2.1.2 The Box-Wilson Procedure.

Box and Wilson (1951) devised a procedure for the multi-factor case, but Duncan (1959, p. 768) modified this procedure for a single factor. Duncan proposed that observations be made at four levels of the factor, about some starting value, and that a line be fitted to the corresponding responses. Further observations should then be made in the direction indicated by the slope of the fitted line. This process should be repeated until there is evidence that a NSR has been reached, i.e. when the slope becomes close to zero, and then in the final step a second degree polynomial should be fitted to the responses at points in the NSR.

2.1.3 The Kiefer-Wolfowitz (K-W) Procedure.

Kiefer and Wolfowitz (1952) described the following sequential search procedure. If x_n is an estimate of the OFL after n steps then x_{n+1} is determined by the relation

$$x_{n+1} = x_n + (a_n/c_n)(y_{2n} - y_{2n-1}) ,$$

where x_1 is an arbitrary starting level, y_{2n} and y_{2n-1} are the observed responses at $x_n + c_n$ and $x_n - c_n$ respectively and (a_n) and (c_n) are infinite sequences of

positive numbers. The authors proved that x_n converges to θ, the OFL, in probability under the conditions summarized below:

(A) σ^2 is bounded.

(B) η is strictly increasing for $x < \theta$ and strictly decreasing for $x > \theta$.

(C) $c_n \to 0$, $\sum a_n = \infty$, $\sum a_n c_n < \infty$, $\sum a_n^2/c_n^2 < \infty$.

(D) η is reasonably flat in the neighbourhood of θ.

(E) η is not too flat outside the neighbourhood of θ.

(F) η must not rise too steeply in places.

Under these conditions functions like $\eta = \exp(-x^2)$ and $\eta = -x^2$ are excluded, because of their behaviour as x tends to $\pm \infty$. However these functions are included if for some reason experimentation is confined to a finite interval containing the maximum.

Blum (1954a) proved convergence with probability one under weaker conditions, since he found (D) and $\sum a_n c_n < \infty$ in (C) to be unnecessary. Derman (1956) proved convergence in probability under quite restrictive conditions but which allow $\eta = -x^2$. Under some additional conditions he proved that x_n is asymptotically normal. Dvoretsky (1956) proved convergence both in mean square and with probability one, and the regularity conditions are weaker than those previously

used. The conditions are such that $\eta = \exp(-x^2)$ and $\eta = -x^2$ are allowed. Burkholder (1956) gives an alternative proof to Dvoretsky's for convergence with probability one and also proves asymptotic normality. Sacks (1958) obtained more general and complete results than hitherto because of his method of proof. It is shown that in general the sequences (a_n) and (c_n) can be chosen so that x_n is arbitrarily close to $O_p(n^{-\frac{1}{2}})$ without ever attaining it. The strongest conclusion obtained, under certain assumptions with restrictions only on two moments of $y(x)$, is that for $a_n = An^{-1}$, $c_n = (n^{1/6} \log n)^{-1}$ and $2\alpha = -\eta''(\theta)$ then $n^{\frac{1}{2}} c_n (x_n - \theta)$ is asymptotically normal with mean zero and variance $\sigma^2 A^2 (8\alpha A - 1)^{-1}$. This variance is minimized when $A = 1/(4\alpha)$.

Fabian (1967a) shows that the K-W procedure can be modified to increase the speed of convergence, by making more observations at each step. He points out that there are several designs which will achieve this. Fabian (1967b) obtains the conditions for the existence of the best design but gives no practical method of selecting it.

2.1.4 The Normalized Kiefer-Wolfowitz Procedure.

Wilde (1964, p. 178) illustrates some undesirable

behaviour of the K-W procedure, namely the overshooting of the OFL which retards its speed of convergence, when there is asymmetry in the response function near the OFL. Consider the following modification of the K-W procedure. If x_n is an estimate of the OFL after n steps then x_{n+1} is determined by the relation

$$x_{n+1} = x_n + (a_n/c_n) \operatorname{sgn} (y_{2n} - y_{2n-1}) ,$$

where $\operatorname{sgn}(u) \equiv u/|u|$ and $\operatorname{sgn}(0) \equiv 0$. Under the conditions of Dvoretsky (1956) this normalized procedure converges and empirical tests show that is free from the undesirable behaviour discussed above.

2.1.5 The Accelerated Kiefer-Wolfowitz Procedure.

Wilde (1964, p. 179) points out that an ideal stochastic approximation scheme would be such that the step size is large when one is far away from the OFL, shortening the steps as it approached the OFL. The K-W procedure does this for suitably behaved functions.

Kesten (1958) proposed accelerating the K-W procedure by using the number of past reversals in the direction of the search, to indicate whether one is near to the OFL. Kesten shortens the step size only if a change in the direction of the

search is indicated. This procedure converges with probability one under slightly more restrictive assumptions than the original K-W procedure. Kesten requires that $a_{n+1} < a_n$ and that $c_n = c$ is constant, $n = 1, 2, \ldots$.

Wilde describes the normalized version of this procedure, as follows:

If x_n is an estimate of the OFL after n steps then x_{n+1} is determined by the relation

$$x_{n+1} = x_n + (a_m/c) \operatorname{sgn} (y_{2n} - y_{2n-1}) ,$$

where $a_{m+1} < a_m$, $m = 1, 2, \ldots,$ c is constant and where m is increased by one whenever the sequence $\left\{ \operatorname{sgn} (y_{2n} - y_{2n-1}) \right\}$, $n = 1, 2, \ldots,$ changes sign. This procedure behaves in the required way for more general response functions.

2.1.6 Discussion.

Hotelling concentrates on a procedure for the final stage of the search; see Section 2.1.1. This contributes little to finding a suitable initial stage sequential procedure.

For comments on the Box-Wilson procedure see Section 3.1.7.

As stated in Section 2.1.3, a number of proofs of convergence accompanied the development of the K-W procedure.

While the study of the limiting behaviour of a procedure is some guide to its practical behaviour we shall be mainly concerned with performance in finite samples.

The K-W procedure and the modified versions of this procedure all require a predetermined choice of the sequences (a_n) and (c_n), which fix the step size and design size at each step. Possibly, if the step size and design size were determined at each step as functions of the observations, the procedure would be more efficient. For example, if the step size was continually increased over a region where the response was increasing or if the design size was not decreased until no further progress could be made, then the speed of convergence to the optimum region would be improved. To preserve the automatic property of the procedure such a method of choosing (a_n) and (c_n) should be built into the procedure. If c_n tends to zero too quickly there is a danger that the real differences between the pairs of responses used to estimate the gradient at a point, may be confounded with the experimental error, hence retarding the speed of convergence. The accelerated procedure takes account of this by keeping c_n constant, but the speed of convergence might be further improved if c_n is allowed to vary but not get too small.

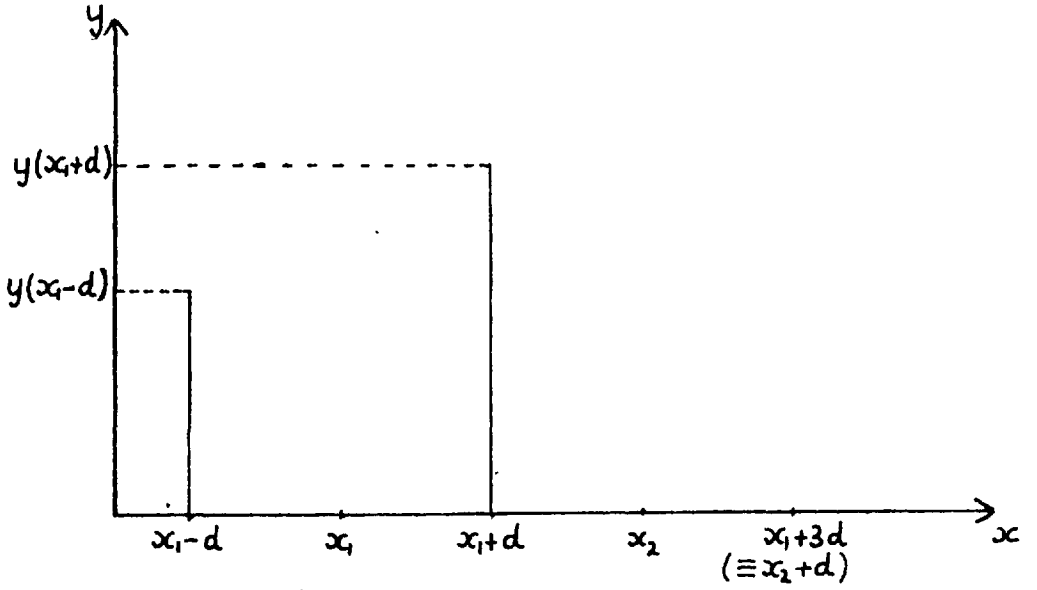
2.2 New Procedures.

In the procedures to be described in the following sections, modifications to the basic K-W procedures are considered. The choice of step size is altered and the number of observations to be made at each step is reduced from two to one.

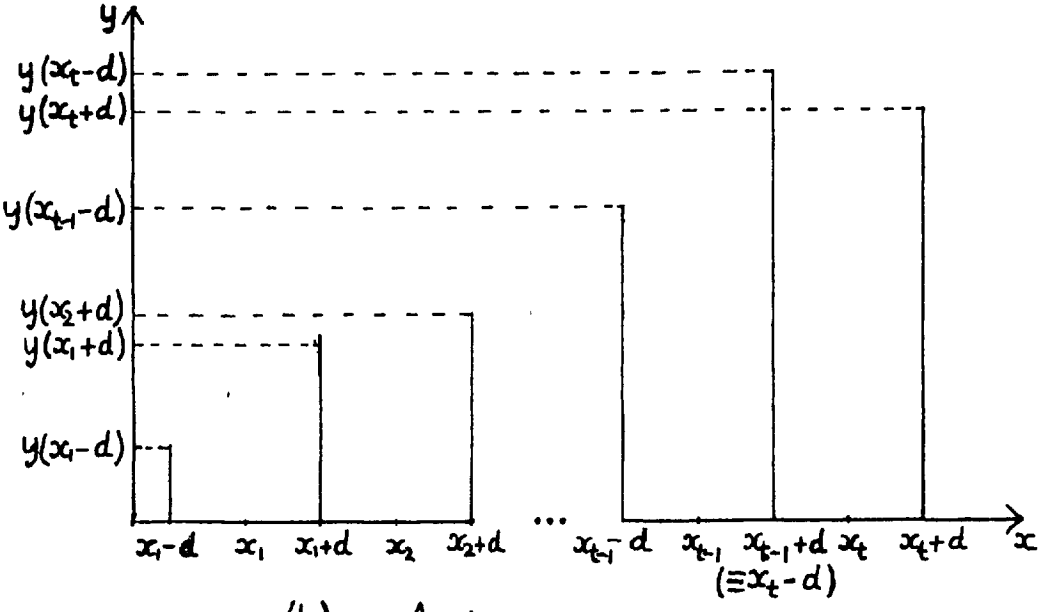
2.2.1 The Up-Down Procedure.

In this procedure observations are made at equally spaced levels. Let x_1 be the starting value. We then make observations $y(x_1 + d)$ and $y(x_1 - d)$, $d > 0$, one at each of the levels $x_1 + d$ and $x_1 - d$ respectively. The initial design size $2d$ is chosen by the experimenter. If $y(x_1 + d) > y(x_1 - d)$ we make the next observation at $x_1 + 3d$ $(\equiv x_2 + d)$, Figure 1a. If $y(x_2 + d) > y(x_1 + d)$ we make the next observation at $x_2 + 3d$ $(\equiv x_3 + d)$ and so on until, for $t \geq 2$, $y(x_t + d) < y(x_t - d)$, where $x_t - d$ $(\equiv x_{t-1} + d)$, Figure 1b. If, initially, $y(x_1 + d) < y(x_1 - d)$ then we make the next observation at $x_1 - 3d$ and proceed similarly. In this case the sequence moves in the direction of the negative x - axis. Throughout this thesis we regard the event $y_a - y_b = 0$ as occurring with a very small probability.

FIGURE 1. THE UP-DOWN PROCEDURE.

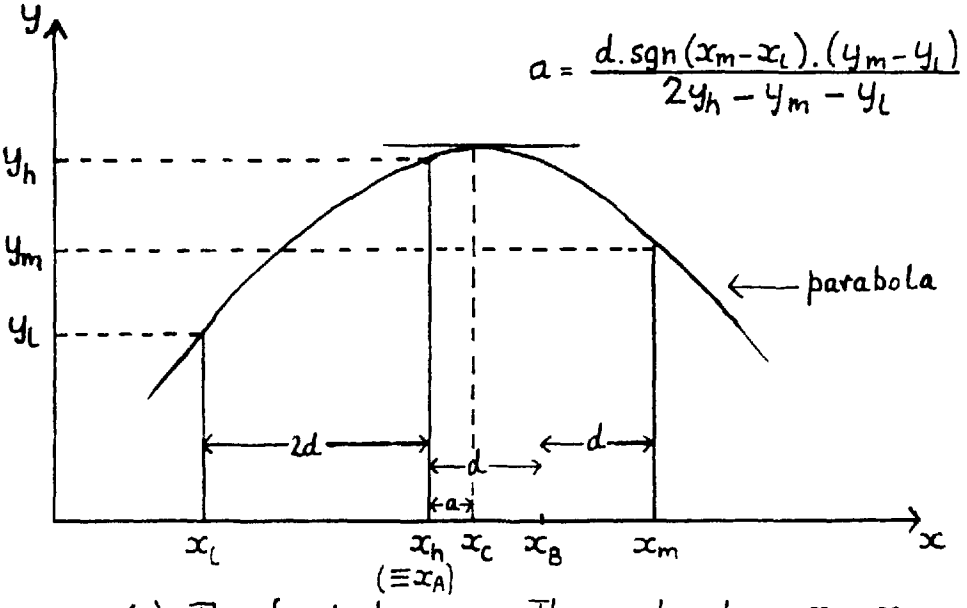


(a) The initial designs.

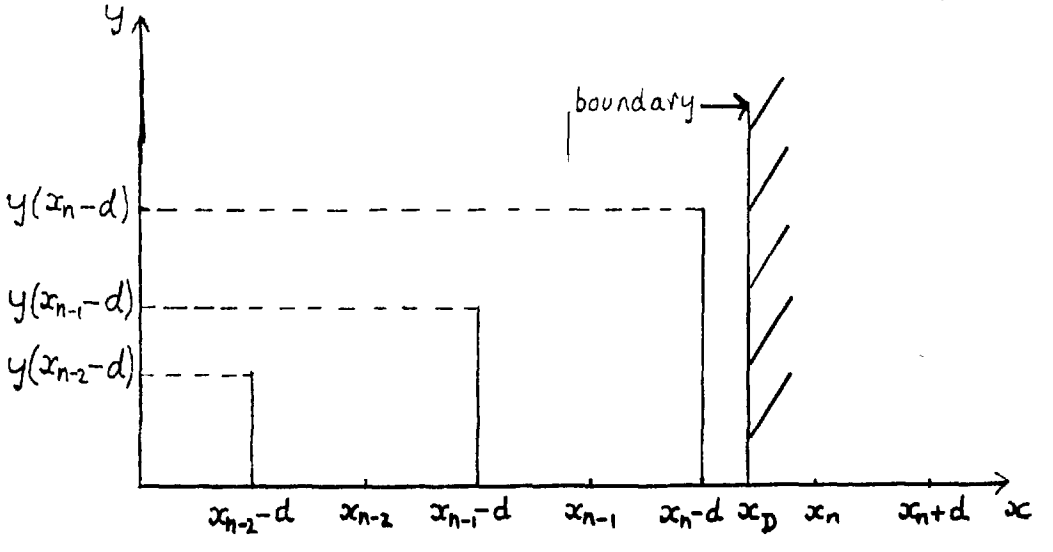


(b) A run.

FIGURE 1 (cont'd). THE UP-DOWN PROCEDURE.



(c) The final designs. The estimates x_A, x_B and x_C correspond to Estimators A, B and C respectively.



(d) A boundary of the experimental region.

When $y(x_t + d) < y(x_t - d)$, Figure 1b, we say that a 'run' has been concluded and estimate the OFL by $\hat{\theta}_1$ to be described below. Before beginning the next run we make a repeat observation at $x_t - d$, the point corresponding to the highest response in the previous run. This guards against a spurious observation at $x_t - d$ indicating a false optimum region. We begin the next run by comparing the new response at $x_t - d$ with the old response at $x_t + d$ and continue as before with x_1 now equal to x_t . We observe that $\mathrm{Cov}(\hat{\theta}_i, \hat{\theta}_j) = 0$, $(j \neq i + 1)$. After R runs we estimate the OFL by

$$\hat{\theta}(R) = (1/R) \sum_{i=1}^{R} \hat{\theta}_i \quad .$$

Suppose we label the last three levels of a run x_l, x_m and x_h, such that the corresponding responses $y_l < y_m < y_h$, Figure 1c . We note that either $x_l < x_h < x_m$, Figure 1c, or $x_m < x_h < x_l$ may occur. Consider the following three estimators for $\hat{\theta}_i$:

Estimator (A) x_h .

Estimator (B) $x_h + d \ \mathrm{sgn}(x_m - x_l)$.

Estimator (C) $x_h + d \ \mathrm{sgn}(x_m - x_l)\ (y_m - y_l)/(2y_h - y_m - y_l)$.

Estimator (A) is the factor level, x_h , corresponding to the

highest observed response in the i^{th} run. Estimator (B) is a refinement of Estimator (A) in that it takes into account the sign of the difference between the observed responses at the neighbouring levels of x_h . Estimator (C) is the factor level corresponding to the maximum of the unique parabola which passes through the points (x_l, y_l) , (x_h, y_h) and (x_m, y_m) , Figure 1c . Estimator (C) is the maximum likelihood estimator of θ given the three observations y_l, y_h and y_m, where

$$y_i \sim N(\beta_{11}x_i^2 + \beta_1 x_i + \beta_0, \sigma^2) \quad (i = l, h \text{ and } m) \ .$$

We shall call this Up-Down procedure, Routine 1A.

In the r^{th} run of Routine 1A, $r \geq 2$, an observation may be made at a level which was observed in the $(r - 1)^{th}$ run. Our procedure could be modified so that, for any given pair of adjacent levels observed in the r^{th} run, we could compare the average of the current response and that response observed (if any) in the $(r - 1)^{th}$ run at each level, in choosing the next design point. If, at a given level, there was no observation made in the $(r - 1)^{th}$ run, then only the current response is used in the r^{th} run. However, as shown in Section 2.2.3, simulation results suggest that there is nothing to be gained by modifying the procedure in this way. We shall call this modified version of the Up-Down procedure, Routine 1B.

Two stopping rules are considered for Routine 1A. The first requires that the procedure be continued for 1 or 2 runs and then an estimate of the OFL made. This can then be followed by phase (ii) and phase (iii) of the Hotelling procedure. For the second, we consider the following sequential stopping rule. If the standard error of $\hat{\theta}(R)$ can be estimated from the dispersion of $\hat{\theta}_i$ $(i = 1, 2, \ldots, R)$, then we continue until the estimate of the standard error of $\hat{\theta}(R)$ is less than some preassigned value. We shall consider Estimator (C) for $\hat{\theta}_i$ since it is shown to be the most efficient in Section 2.2.3. It is possible to obtain an estimate of the standard error of $\hat{\theta}(R)$ under the following assumptions:

(i) The final levels x_l, x_h and x_m of the first run of Routine1A (Figure 1c) are such that $|x_h - \theta| < d$.

(ii) Subsequent runs are confined to these levels.

(iii) The response function η is parabolic in the region of the final designs, i.e.

$$y_i \sim N(\beta_{11}x_i^2 + \beta_1 x_i + \beta_0, \sigma^{-2}) \quad (i = l, m \text{ and } h),$$

where β_{11} (< 0).

Without loss of generality, we let $x_h = 0$, $x_l = -1$, $x_m = 1$ and $d = 0.5$.

Suppose, as in Figure 1b, the sequence of designs moves in the direction of the positive x - axis, for the first run, then x = 1 will be the last point in that run. Let the observations at x = -1, 0 and 1 be y_{11} , y_{12} and y_{13} respectively. $\hat{\theta}_1$ is then calculated. At the end of the second we calculate $\hat{\theta}_2$ from the observations y_{21} , y_{22} and y_{23} $(= y_{13})$ at x = -1, 0 and 1 respectively. The correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ arises because the same response y_{13} at x = 1 appears in each estimate. Let $\mathrm{Cov}(\hat{\theta}_1 , \hat{\theta}_2) = c$. At the end of the third run we find $\hat{\theta}_3$ from the observations $y_{31}(=y_{21})$, y_{32} and y_{33} at x = -1, 0 and 1 respectively. Let $\mathrm{Cov}(\hat{\theta}_2 , \hat{\theta}_3) = c'$. In general, $\mathrm{Cov}(\hat{\theta}_i , \hat{\theta}_{i+1})$ (i = 1, 2, ..., R) will be either c or c' and $\mathrm{Cov}(\hat{\theta}_i , \hat{\theta}_j) = 0$, (j > i+1) .

Consider the function

$$S^2 = \sum_{i=1}^{R} \left\{ \hat{\theta}_i - \hat{\theta}\,(R) \right\}^2 .$$

Now, $$E(S^2) = R \left\{ \mathrm{Var}\ \hat{\theta} - \mathrm{Var}\ \hat{\theta}\,(R) \right\}$$

since $$E(\hat{\theta}) = E\left\{ \hat{\theta}\,(R) \right\} ,$$

where $$E(\hat{\theta}) = E(\hat{\theta}_1) = \ldots = E(\hat{\theta}_R)$$

and $$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}_1) = \ldots = \mathrm{Var}(\hat{\theta}_R) .$$

Now, $$\mathrm{Var}\left\{\hat{\theta}(R)\right\} = (1/R^2)\left\{R\,\mathrm{Var}\hat{\theta} + 2\sum_{i=1}^{R-1}\mathrm{Cov}(\hat{\theta}_i, \hat{\theta}_{i+1})\right\} .$$

If $$K = 2\sum_{i=1}^{R-1}\mathrm{Cov}(\hat{\theta}_i, \hat{\theta}_{i+1})$$

then we get

$$E(S^2) = R(R-1)\ \mathrm{Var}\left\{\hat{\theta}(R)\right\} - K .$$

It can be shown that $c = \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)$ and $c' = \mathrm{Cov}(\hat{\theta}_2, \hat{\theta}_3)$ as defined above are given by

$$c \doteq (\sigma^2/16\overline{\beta_{11}^2})(52\theta^2 + 20\theta + 1)$$

and

$$c' \doteq (\sigma^2/16\overline{\beta_{11}^2})(52\theta^2 + 20\theta + 1) ,$$

where terms of higher degree than σ^2 are neglected. From these we obtain

$$K \doteq (R-1)(\sigma^2/8\overline{\beta_{11}^2})(52\theta^2 + 24\theta + 1) ,$$

if $R\ (>1)$ is odd, and

$$K \doteq (R-1)(\sigma^2/8\overline{\beta_{11}^2})(52\theta^2 + \tfrac{24R-20}{R-1}\theta + 1) ,$$

if R is even, where terms of higher degree than σ^2 are neglected.

For large R we get

$$E(S^2) \doteq R(R-1)\ \mathrm{Var}\left\{\hat{\theta}(R)\right\}$$

which is the corresponding result for uncorrelated $\hat{\theta}_i$'s .

Put $B = \frac{K}{R(R-1)}$, then

$\frac{S^2}{R(R-1)}$ estimates $\mathrm{Var}\left\{\hat{\theta}(R)\right\}$ with bias $-B$.

TABLE 1

The values of the bias for estimates of $\mathrm{Var}\{\hat{\theta}(R)\}$ (Routine 1A).

$R \backslash \theta$	-0.5		0.0		0.5	
	$-\beta_{11}/\sigma$		$-\beta_{11}/\sigma$		$-\beta_{11}/\sigma$	
	1	5	1	5	1	5
2	0.0000	0.0000	0.0625	0.0025	1.7500	0.0700
3	0.0833	0.0033	0.0417	0.0017	1.0833	0.0433
4	0.0424	0.0017	0.0313	0.0013	0.8326	0.0333
5	0.0500	0.0020	0.0250	0.0010	0.6500	0.0260
10	0.0223	0.0009	0.0125	0.0005	0.3277	0.0132
11	0.0227	0.0009	0.0114	0.0005	0.2957	0.0118
21	0.0119	0.0005	0.0059	0.0002	0.1548	0.0062

Table 1 indicates how the bias is reduced as R is increased for given θ and $-\beta_{11}/\sigma$.

If in a given run a design point $x_n + d$, Figure 1d, falls outside the experimental region, for some n, then we allocate a fictitious response to $x_n + d$, which is small compared with any response at $x_n - d$, and hence the run is terminated. The factor level, x_D, at the boundary is taken as the run estimate. We continue as usual by making a repeat observation at $x_n - d$ and then at $x_{n-1} - d$.

If the sequence of designs continually encounters the same boundary, then there is evidence that the OFL is in the neighbourhood of the boundary if not at the boundary itself. In this case further observations should be made in the neighbourhood of the boundary.

If an estimate of the ratio, to σ, of the slope of the curve in the region of the initial design is available, then d should be chosen such that the probability that

$$y(x_1 + d) \gtrless y(x_1 - d)$$

is high, whenever the estimate of the slope indicates that

$$\eta(x_1 + d) \lessgtr \eta(x_1 - d)$$

respectively. In the absence of any initial knowledge Routine 1A or Routine 1B could be operated for a few runs from different starting points and with different values of d to gain some information about the nature of the response function.

Routines 1A and 1B are automatic procedures designed to meet practical needs, however, they do not allow any flexibility in design size. In Section 2.2.2 we shall describe a modified version of Routine 1A which does allow a change in design size depending on the nature of the response function.

Quantitative results for Routines 1A and 1B are given in Section 2.2.3.

2.2.2 The Generalized Up-Down Procedure.

Let x_1 be the starting value. We then make observations, one at each of the levels $x_1 - d_1$ and $x_1 + d_1$, where $d_1 (> 0)$ is chosen by the experimenter. The two levels $x_1 - d_1$ and $x_1 + d_1$ are denoted by x_ℓ and x_h so that the corresponding observed responses satisfy the inequality $y_\ell < y_h$. The level x_ℓ corresponding to the lower response y_ℓ is now replaced by a new point which together with x_h forms the next design. Rules for finding the new point are given below. An observation is made at the new point and this is compared with the observation at x_h; then, the design points of this new design are labelled x_ℓ and x_h appropriately, and the process is repeated.

Let x_n be the estimate of the OFL after $n-1$ steps then x_{n+1} is determined by the relation

$$x_{n+1} = x_n + (d_n + d_{n+1}) \operatorname{sgn}\left\{ y(x_n + d_n) - y(x_n - d_n) \right\} ,$$

where $y(x_n + d_n)$ and $y(x_n - d_n)$ are the observed responses at $x_n + d_n$ and $x_n - d_n$ respectively, $d_n > 0$, and

$$d_{n+1} = \begin{cases} d_n/2 & \text{if the new point is a contracted point,} \\ d_n & \text{if the new point is a reflected point,} \\ 2d_n & \text{if the new point is an expanded point.} \end{cases}$$

The ideas of contraction, reflection and expansion are explained below.

If we consider the n^{th} design and assume that

$$y(x_n + d_n) > y(x_n - d_n) ,$$

then

$$x_n + d_n \equiv x_h \quad \text{and} \quad x_n - d_n \equiv x_l .$$

The new point for the $(n+1)^{th}$ design may be chosen according to the following rules.

Given x_l and x_h the next observation is made at $x_n + 3d_n$ which may be regarded as the reflection of x_l about x_h. We denote $x_n + 3d_n$ by x_r and the observed response at x_i by y_i, for any i.

(a) If $y_r > y_h$, then the next observation is made at the level $x_n + 5d_n$. This level may be regarded as the expansion of x_r and is denoted by x_e.

(i) If $y_e > y_r$, then x_h and x_e form the $(n+1)^{th}$ design (Figure 2a) and are relabelled x_ℓ and x_h respectively (a successful expansion).

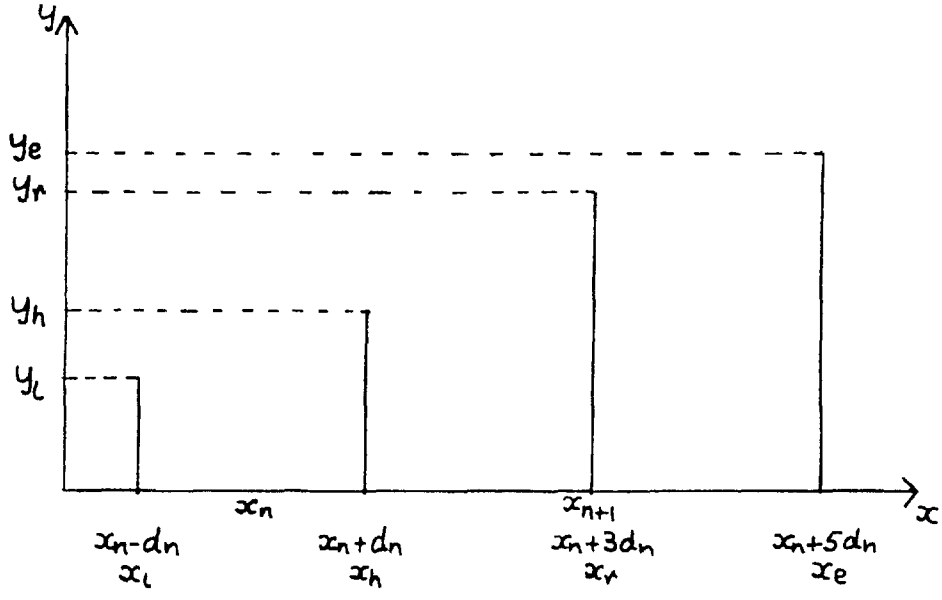
(ii) If $y_e < y_r$, then x_h and x_r form the $(n+1)^{th}$ design (Figures 2b and 2c) and are relabelled x_ℓ and x_h respectively (a failed ~~reflection~~ expansion).

(b) If $y_h > y_r > y_\ell$, then x_h and x_r form the $(n+1)^{th}$ design (Figure 2d) and are relabelled x_h and x_ℓ respectively (a failed reflection).

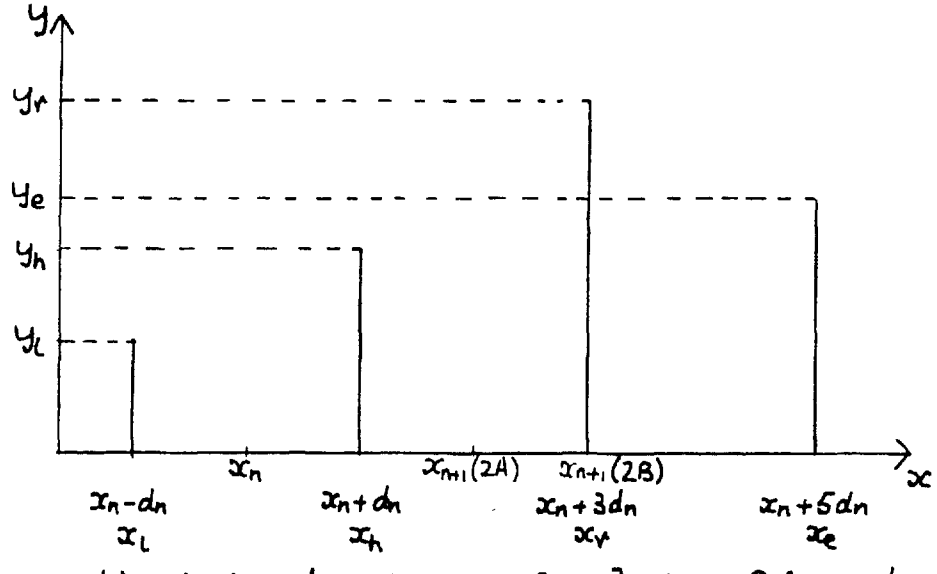
(c) If $y_r < y_\ell$, then the next observation is made at the level $x_n + 2d_n$. This level may be regarded as the contraction of x_r and is denoted by x_c .

(i) If $y_c > y_r$ and $y_c > y_h$, then x_h and x_c form the $(n+1)^{th}$ design (Figure 2e), and are relabelled x_ℓ and x_h respectively (a successful contraction).

FIGURE 2. THE GENERALIZED UP-DOWN PROCEDURE - The n^{th} and $(n+1)^{th}$ designs.

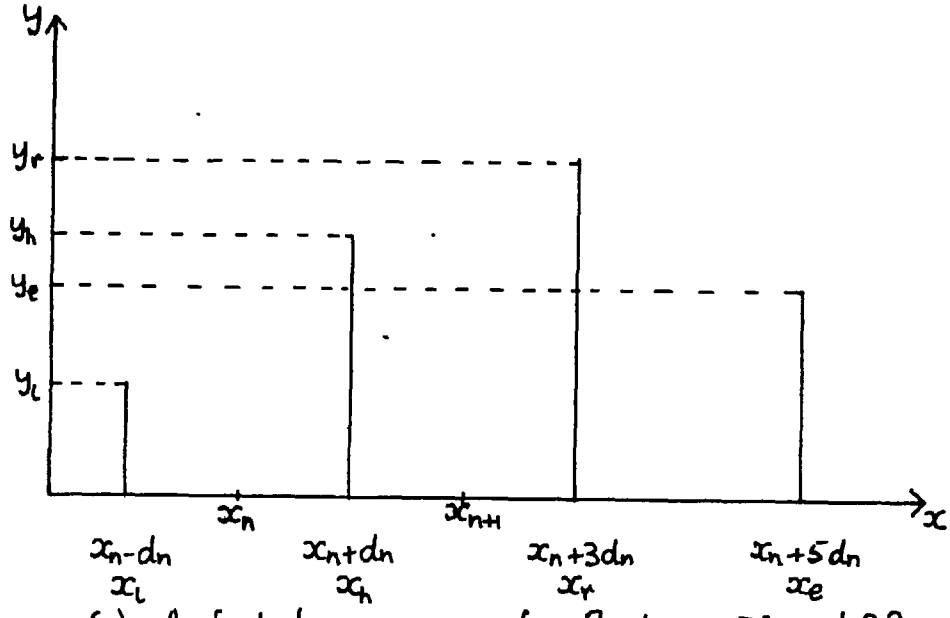


(a) A successful expansion for Routines 2A and 2B.

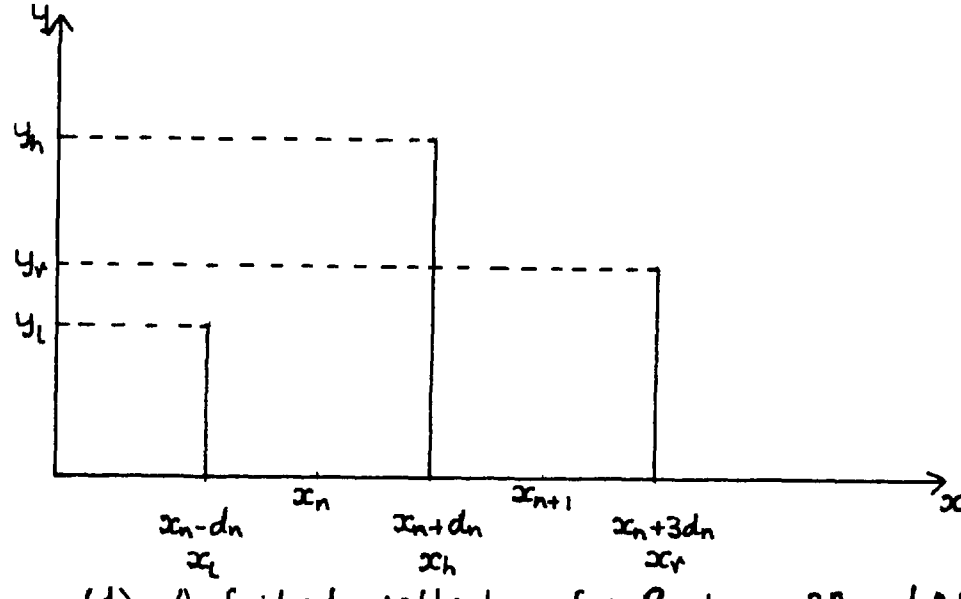


(b) A failed expansion for Routine 2A and a successful expansion for Routine 2B.

FIGURE 2 (cont'd). THE GENERALIZED UP-DOWN PROCEDURE.

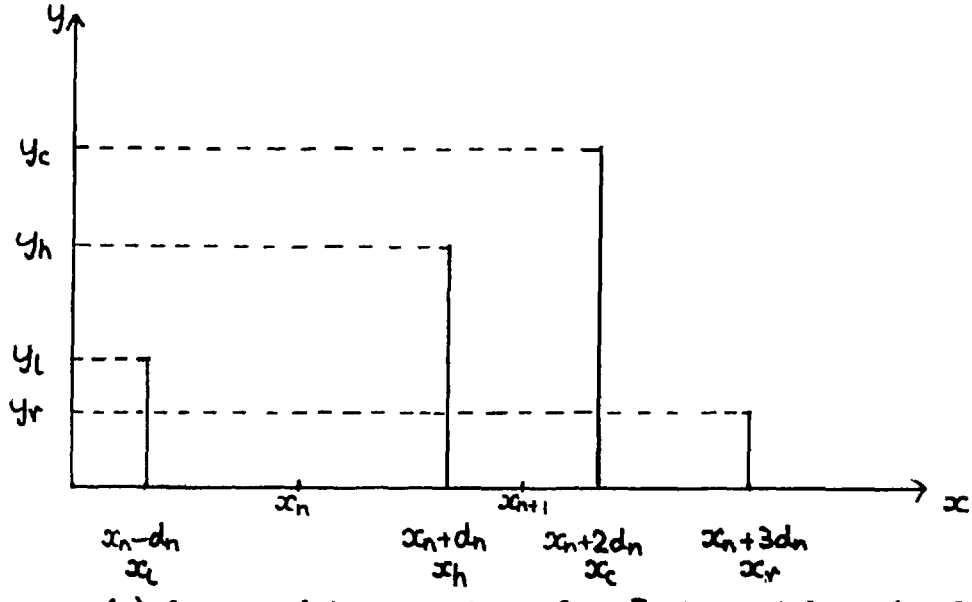


(c) A failed expansion for Routines 2A and 2B.

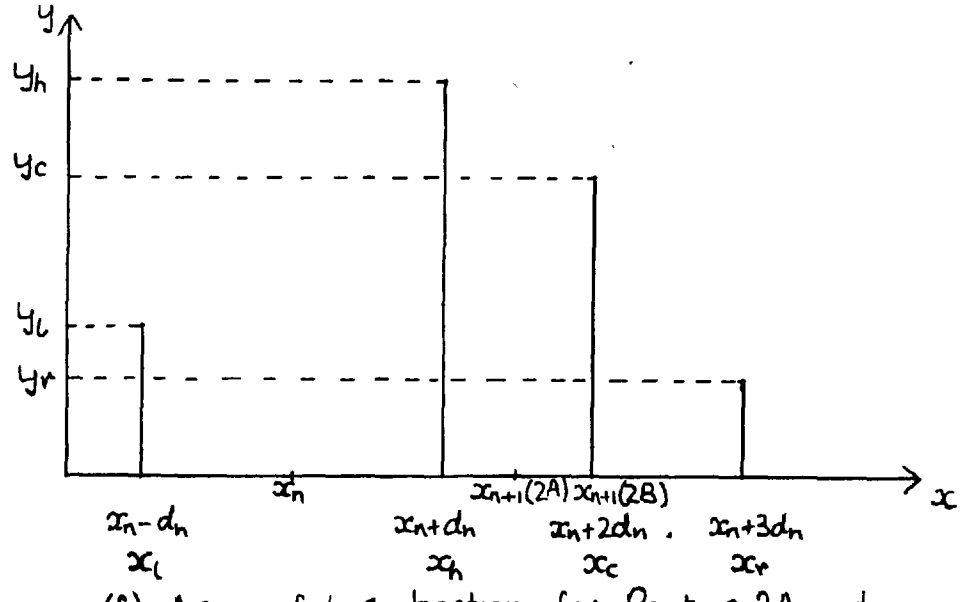


(d) A failed reflection for Routines 2A and 2B.

FIGURE 2 (cont'd). THE GENERALIZED UP-DOWN PROCEDURE.

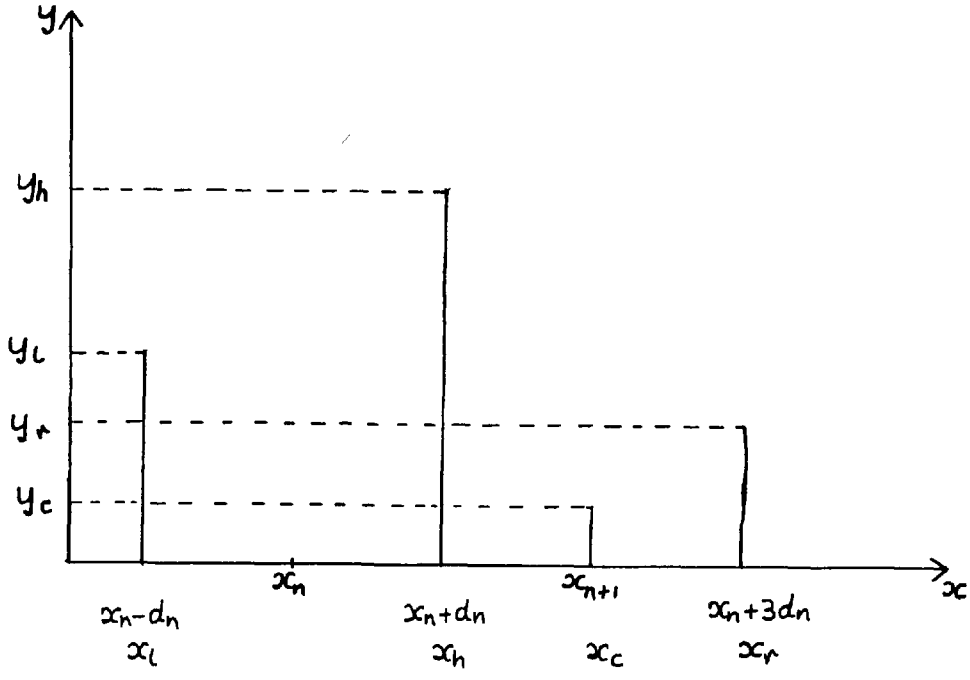


(e) A successful contraction for Routines 2A and 2B.



(f) A successful contraction for Routine 2A and a failed contraction for Routine 2B.

FIGURE 2 (cont'd). THE GENERALIZED UP-DOWN PROCEDURE.



(g) A failed contraction for Routines 2A and 2B.

(ii) If $y_c > y_r$ and $y_c < y_h$, then x_h and x_c form the $(n+1)^{th}$ design (Figure 2f), and are relabelled x_h and x_ℓ respectively (a successful contraction).

(iii) If $y_c < y_r$, then x_h and x_r form the $(n+1)^{th}$ design (Figure 2g) and are relabelled x_h and x_ℓ respectively (a failed contraction).

If, on the other hand, for the n^{th} design

$$y(x_n + d_n) < y(x_n - d_n) ,$$

then the next observation is made at $x_n - 3d_n$ and we proceed similarly.

If the level x_h is in two consecutive designs and is not then due to be replaced, we make a repeat observation at x_h before continuing. This rule applies in cases (b), (c)(ii) and (c)(iii) above. It guards against a spurious observation at x_h indicating a false optimum region.

We shall refer to this procedure as Routine 2A.

We may also consider the following modified procedure. Given x_ℓ and x_h (Figure 2a) for the n^{th} design,

the next observation is made at $x_n + 3d_n$ ($\equiv x_r$) and the new point for the $(n+1)^{th}$ design may be chosen according to the following alternative rules.

(a') If $y_r > y_h$, then the next observation is made at the level $x_n + 5d_n$, denoted by x_e.

(i) If $y_e > y_h$, then x_h and x_e form the $(n+1)^{th}$ design (a successful expansion, Figures 2a and 2b).

(ii) If $y_e < y_h$, then x_h and x_r form the $(n+1)^{th}$ design (a failed expansion, Figure 2c).

(b') If $y_h > y_r > y_l$, then x_h and x_r form the $(n+1)^{th}$ design (a failed reflection, Figure 2d).

(c') If $y_r < y_l$, then the next observation is made at the level $x_n + 2d_n$, denoted by x_c.

(i) If $y_c > y_h$, then x_h and x_c form the $(n+1)^{th}$ design (a successful contraction, Figure 2e).

(ii) If $y_c < y_h$, then x_h and x_r form the $(n+1)^{th}$ design (a failed contraction, Figures 2f and 2g).

If for the n^{th} design

$$y(x_n + d_n) < y(x_n - d_n) \quad ,$$

then the next observation is made at $x_n - 3d_n$ and we proceed similarly.

If the level x_h is in two consecutive designs and is not then due to be replaced, we make a repeat observation at x_h before continuing. This rule applies in cases (b') and (c')(ii) above. It guards against a spurious observation at x_h indicating a false optimum region.

We shall refer to this alternative procedure as Routine 2B.

Unlike Routines 1A and 1B the feature of Routines 2A and 2B to allow an increase in step size and design size makes them suitable as search procedures when the starting level is far from the optimum region. For Routine 2B, case (a') , the probability of an expanded point being chosen in the next design is greater than for Routine 2A, case (a). Also, for Routine 2B, case (c'), the probability of a contracted point being chosen is less than for Routine 2A, case (c) . Thus, Routine 2B will on the average give

rise to a sequence of designs which will climb faster to the optimum than that for Routine 2A. However, since on reaching an optimum region the design size for Routines 2A and 2B will in general be large compared with the initial design size, the design size will have to be contracted to obtain a more precise estimate of the OFL. This contraction will on the average be slower for Routine 2B than for Routine 2A.

In the next section after discussing the quantitative results of Routines 1A, 1B, 2A and 2B, we shall consider amalgamating the best features of these procedures to produce an overall more satisfactory procedure. Discussion of stopping rules and rules for constraints on the experimental region for Routines 2A and 2B is postponed until the next section.

2.2.3 Results and Discussion.

Analytical results for the Up-Down Procedure.

It is in general difficult to obtain analytical results for Routine 1A or Routine 1B. However, some analytical results may be obtained in rather extreme cases and may then be compared with the corresponding results from simulation. Incidentally the agreement found below between

these two sets of results gives some check of the simulation methods used including the subroutine for generating random normal deviates.

Consider a single run of Routine 1A where the levels of the initial design are denoted by x_j and x_{j+1} and where $x_i < x_{i+1}$ for all i. Let x_H be the level at which the highest response in the run is observed. Then, we denote the probability that $x_H = x_k$, say, given x_j, x_{j+1}, by

$$p_{jk} = \begin{cases} p^{+}_{jk} & \text{if} \quad k > j \ , \\ p^{-}_{j+1\,k} & \text{if} \quad k < j+1 \ . \end{cases}$$

Now,

$$p^{+}_{jk} = \text{Prob}\,(y_{j+1} > y_j, \ldots, y_k > y_{k-1}, y_k > y_{k+1}) \ ,$$

which after some manipulation reduces to

$$p^{+}_{jk} = \int_{-a_{j+1}}^{\infty} \cdots \int_{-a_k}^{\infty} \int_{a_{k+1}}^{\infty} \phi(\underline{t}_p, \underline{v}_p)\, d\underline{t}_p \ ,$$

where

$$a_r = (\eta_r - \eta_{r-1})/(\sigma\sqrt{2})$$

and η_r is the true response at x_r. The function

$\phi(\underline{t}_p, \underline{V}_p)$ is the $p(=k+1-j)$ variate normal density with zero means, unit variances and covariance matrix $\underline{V}_p \equiv (v_{mn})$, where $v_{mn} = v_{nm}$ and

$$\begin{aligned} v_{mn} &= 1 && \text{if } m = n \ , \\ &= -1/2 && \text{if } m = n+1 \text{ for } m = 1,2,\ldots, p-2 \ , \\ &= 1/2 && \text{if } m = n+1 \text{ and } m = p-1 \ , \\ &= 0 && \text{otherwise} \ . \end{aligned}$$

Also, $\bar{p}_{j+1\,k}$ is similarly defined.

For $p = 2$, p_{jk} can be calculated from tables of the bivariate normal distribution (National Bureau of Standards, 1959). For $p > 2$, Plackett (1954) gives reduction formulae which, however, rapidly become more tedious to calculate as the number of variates increases. It is thus in general difficult to obtain the transition probabilities p_{jk}.

Consider approximating p_{jk} by p_k, where

$$p_k = \int_{-a_k}^{\infty} \int_{a_{k+1}}^{\infty} \phi(\underline{t}_2, \underline{V}_2) \, d\underline{t}_2 \ .$$

With the aid of bivariate normal tables, p_k can easily be

calculated. This approximation is equivalent to saying that Prob $(x_H = x_k)$ depends only on what happens at the levels x_{k-1}, x_k, x_{k+1} and is thus in general independent of the starting levels of the run. For what type of response function is this approximation plausible?

Consider a parabolic response curve,

$$\eta = \beta_{11}x^2 \quad (\beta_{11} < 0) \ ,$$

where θ, the OFL, is equal to zero. Then

$$a_r = 2d(\beta_{11}/\sigma\sqrt{2})(2x_r - .2d) \ ,$$

where $2d = x_r - x_{r-1}$ for all r . In Routine 1A observations are made at a number of equally spaced levels. Denote the level among these which is nearest to θ by x_N . Then $|x_N - \theta| < d$.

Now $p_k > p_{jk}$ for all j and k and hence

$$\sum_k p_k > \sum_k p_{jk} = 1 \quad \text{for all} \quad j \ .$$

As $-\beta_{11}/\sigma$ increases $|a_r|$ increases, faster for larger $|N-r|$ than for smaller $|N-r|$. Hence $p_{j\,N\pm i}$ tends to $p_{N\pm i}$ for all i and for $i \geq I$, say, $p_{N\pm i}$ is arbitrarily close to zero and may be ignored. Also, as $-\beta_{11}/\sigma$

increases $\sum_k p_k$ tends to $\sum_k p_{jk} = 1$ for all j. Therefore for the approximation to be reasonable $-\beta_{11}/\sigma$ must be large enough that

$$\sum_k p_k - 1 < \delta ,$$

where δ (> 0) is chosen to be small. If possible, we must choose δ large enough so that $p_{N\pm i}$, for $i = 1$ at least, is not too small compared with p_N. For example, from a plot of $-\beta_{11}/\sigma$ against $\sum_k p_k - 1$, it can be shown, for $\delta = 5(10^{-4})$, $x_N = 0.25$, $\theta = 0.0$, $d = 0.5$ and $-\beta_{11}/\sigma = 3.1$, that $p_{N+i} < 10^{-5}$ for $i \geq 2$,

$$p_{N-1} = 0.13670 , \quad p_N = 0.86320 , \quad p_{N+1} = 0.00051$$

and $\sum_k p_k = 1.00041$ correct to five decimal places. We shall assume that the approximation is reasonable under these conditions.

Consider Estimator (A) for Routine 1, i.e. $\hat{\theta}_i \equiv x_H$. Now

$$E\left\{\hat{\theta}(2)\right\} = \sum_l \sum_m \left\{ (x_l + x_m)/2 \right\} p_{jm} p'_{ml} ,$$

where $x_H = x_m$ for the first run and $x_H = x_l$ for the second run and where

$$p'_{ml} = \begin{cases} p_m & \text{if the last two levels of the first run are } x_m, x_{m+1}; \\ p_{m-1,l} & \text{if the last two levels of the first run are } x_{m-1}, x_m . \end{cases}$$

Under the above conditions we can replace p_{jm} and p'_{m_ℓ} by p_m and p_ℓ respectively. Then, it can be shown that

$$E\left\{\hat{\theta}(2)\right\} \doteq (p_{N-1} + p_N + p_{N+1})\ E\left\{\hat{\theta}(1)\right\}$$

and by induction that

$$E\left\{\hat{\theta}(R)\right\} \doteq (p_{N-1} + p_N + p_{N+1})^{R-1}\ E\left\{\hat{\theta}(1)\right\}.$$

It can also be shown that

$$MSE\left\{\hat{\theta}(R)\right\} = (1/R)\left[(p_{N-1} + p_N + p_{N+1})^{R-1}\ MSE\left\{\hat{\theta}(1)\right\} + (R-1)(p_{N-1} + p_N + p_{N+1})^{R-2}\ E\left\{\hat{\theta}(1)\right\}^2\right].$$

Substituting the above values for p_{N-1}, p_N and p_{N+1} we get

$$E\left\{\hat{\theta}(1)\right\} = 0.11395\ ,$$

$$E\left\{\hat{\theta}(10)\right\} = 0.11438\ ,$$

$$MSE\left\{\hat{\theta}(1)\right\} = 0.13160$$

and

$$MSE\left\{\hat{\theta}(10)\right\} = 0.02493\ .$$

Routine 1 was simulated and three samples of 2000 trials and two samples of 3000 trials were made of $\hat{\theta}(1)$ and $\hat{\theta}(10)$. Estimates of the above approximate expectations and mean squared errors were obtained together with their standard errors as follows:

$$\hat{E}\left\{\hat{\theta}(1)\right\} = 0.118327 \pm 0.003111 \quad ,$$

$$\hat{E}\left\{\hat{\theta}(10)\right\} = 0.113466 \pm 0.000991 \quad ,$$

$$\hat{MSE}\left\{\hat{\theta}(1)\right\} = 0.108611 \pm 0.071784$$

and

$$\hat{MSE}\left\{\hat{\theta}(10)\right\} = 0.020703 \pm 0.013784 \quad .$$

From these results there is no reason to reject the hypothesis that the simulation methods are satisfactory.

More results may be obtained if we assume a uniform response function, i.e. η = constant for all x . The results obtained under this assumption will be an approximation to the behaviour of the procedure in regions of slowly increasing response. In this case, $a_r = 0$ for all r and hence

$$p^{+}_{jk} = \Phi^{0}_{p}(\underline{v}_p) \quad ,$$

where $\Phi^{0}_{p}(\underline{V}_p)$ is the p(=k-j+1) variate normal probability integral with integrand $\phi(\underline{t}_p, \underline{V}_p)$, integrated over the positive domain. $\underline{V}_p$ is defined as before and

$$p^{-}_{j+1\,k} = \Phi^{0}_{p}(\underline{V}_p) \quad ,$$

where p = j+1-k+1 . The transition probabilities are only affected by σ through the term a_r . Therefore, since $a_r = 0$, for all r , the behaviour of the procedure is

independent of σ $(\neq 0)$.

Now,

$$\Phi_p^0(\underline{V}_p) = \Phi_{p-1}^0(\underline{W}_{p-1}) - \Phi_p^0(\underline{W}_p) ,$$

where $\underline{W}_p \equiv (w_{mn})$ is the covariance matrix such that w_{mn} differs from v_{mn} only for $n = p-1$, $n = p$, where $w_{p-1\,p} = -1/2$. It can be shown (Kendall and Stuart, Vol. I, 1963, p. 367, Excs. 15.9, 15.10) that

$$\Phi_p^0(\underline{W}_p) = 1/(p+1)!$$

and hence that

$$\Phi_p^0(\underline{V}_p) = (1/p!) - 1/(1+p)! .$$
$$= p/(p+1)! .$$

If we assume that $d = 0.5$ and that x_j , x_{j+1} are equal to -0.5 and 0.5 respectively then, for Estimator (A) of Routine 1A on this uniform response function,

$$E\left\{\hat{\theta}(R)\right\} = 0 , \quad \text{for all } R ,$$

by symmetry. Now

$$\text{Var}\left\{\hat{\theta}(1)\right\} = \sum_{n>j} x_n^2 \, p_{jn}^{+} + \sum_{n<j+1} x_n^2 \, p_{s+1\,n}^{-} .$$

This can be evaluated directly to give

$$\text{Var}\left\{\hat{\theta}(1)\right\} = 1.376872 .$$

It is not easy to evaluate $\mathrm{Var}\left\{\hat{\theta}(2)\right\}$ directly, but an approximation can be obtained if we consider a finite number of levels, say x_{j-15} to x_{j+16}. Then,

$$\sum_{t}\sum_{m} p_{sm}\, p'_{mt} = 1.000004$$

and

$$\mathrm{Var}\left\{\hat{\theta}(2)\right\} = 2.251967 \ .$$

These two examples, then, illustrate the difficulties which arise in attempting to find more general analytical results for Routine 1A and hence for the more complicated Routine 1B.

Analytical results for the Generalized Up-Down Procedure.

Routines 2A and 2B, see Section 2.2.2 , can be written in the form

$$x_{n+1} = x_n + a_n \ ,$$

where a_n, given x_n, is a discrete random variable which may assume the values

$$3d_n/2 \ , \ 2d_n \ , \ 3d_n \ , \ -3d_n \ , \ -2d_n \ \text{ and } \ -3d_n/2$$

with probabilities

$$p_1 \ , \ p_2 \ , \ p_3 \ , \ p_4 \ , \ p_5 \ \text{ and } \ p_6$$

respectively. If we assume a uniform response function,

i.e. η = constant, then

$$p_1 = p_6 \ , \quad p_2 = p_5 \quad \text{and} \quad p_3 = p_4 \ ,$$

by symmetry, and the probabilities are independent of n .

For Routine 2A we have

$$p_1 = \text{Prob}(y_h > y_l > y_r < y_c) \qquad \text{(Figures 2e and 2f)} \ ,$$

$$p_2 = \text{Prob}\left\{ (y_l < y_h < y_r > y_e) \cup (y_l < y_r < y_h) \cup (y_c < y_r < y_l < y_h) \right\}$$

(Figures 2b, 2c, 2d and 2g)

and

$$p_3 = \text{Prob}(y_l < y_h < y_r < y_e) \qquad \text{(Figure 2a)} \ .$$

For the uniform response function,

$$p_1 = 3\ \Phi_3^0(\underline{W}_3) = 1/8 \ ,$$

$$p_2 = \Phi_3^0(\underline{V}_3) + \Phi_2^0(\underline{W}_2) + \Phi_3^0(\underline{W}_3)$$

$$= (1/6 - 1/24) + 1/6 + 1/24$$

$$= 1/3$$

and

$$p_3 = \Phi_3^0(\underline{W}_3) = 1/24 \ .$$

We can now determine the expected behaviour of the procedure, over a uniform response function. This behaviour is independent of σ $(\neq 0)$.

Now,

$$E(a_n \mid x_n) = 0 ,$$

by symmetry, and therefore ,

$$E(x_{n+1} \mid x_n) = x_n .$$

Hence

$$E(x_{n+1}) = E(x_n) = \ldots = x_1 .$$

Also,

$$E(x_{n+1}^2 \mid x_n) = x_n^2 + E(a_n^2 \mid x_n) ,$$

where

$$E(a_n^2 \mid x_n) = (191/48)d_n^2 .$$

Therefore,

$$E(x_{n+1}^2) = E(x_n^2) + (191/48)\, E(d_n^2) .$$

But

$$E(d_{n+1}^2 \mid x_n) = (17/16)d_n^2 ,$$

therefore,

$$E(d_{n+1}^2) = (17/16)\, E(d_n^2) = \ldots = (17/16)^n d_1^2 .$$

We now have

$$E(x_{n+1}^2) = E(x_n^2) + (191/48)(17/16)^n d_1^2$$

$$= x_1^2 + (191/3)\left\{(17/16)^n - 1\right\} d_1^2 .$$

We can also show that

$$E(d_{n+1}) = (23/24)^n d_1 ,$$

where $d_{n+1} > 0$, for all n .

For Routine 2B the probabilities can be shown to be

$$p_1 = 1/24 \ , \ p_2 = 3/8 \text{ and } p_3 = 1/12 \ .$$

Using these values we can obtain

$$E(x_{n+1}) = x_1 \ ,$$
$$E(x_{n+1}^2) = x_1^2 + (75/7)\left\{(23/16)^n - 1\right\} d_1^2 \ ,$$
$$E(d_{n+1}) = (9/8)^n d_1$$

and

$$E(d_{n+1}^2) = (23/16)^n d_1^2 \ .$$

We observe that the expected design size, d_{n+1} , decreases with n for Routine 2A but increases with n for Routine 2B. Also, the dispersion of x_{n+1} increases with n but more rapidly for Routine 2B than for Routine 2A. We thus conclude that Routine 2B is more suitable than Routine 2A, as an exploratory procedure over a very slowly varying response function.

The work involved to obtain general analytical results for Routine 2 is unjustified with computer simulation so readily available. To give an example of the work involved consider a linear response function $\eta = \beta_1 x$. For Routine 2A,

$$p_3 = \int_{-a_1}^{\infty} \int_{-a_2}^{\infty} \int_{-a_3}^{\infty} \phi(\underline{t}_3, \underline{W}_3)\, d\underline{t}_3 \quad ,$$

where

$$a_1 = 2\beta_1 d_n/(\sigma\sqrt{2}) = a_2 = a_3 \quad .$$

For fixed values of β_1/σ and d_n, p_3 can be evaluated by a reduction formula (Plackett, 1954) , which involves the numerical integration of a single integral. For p_2 and p_1 numerical integration is also necessary; similarly for Routine 2B. If we consider a parabolic response function the a_r's depend on x_r as well as d_n .

Empirical results for the Up-Down Procedure.

Consider a parabolic response curve

$$\eta = \beta_{11} x^2 \quad , \quad \beta_{11} < 0 \quad .$$

The ratio, β_{11}/σ , is important in controlling the speed of convergence to the optimum region. The dimensions of β_{11}/σ are independent of the dimensions of η . A value of $\beta_{11}/\sigma = \mu$ has the interpretation that for a unit increase in x the rate of change of η decreases by $2\mu\sigma$.

A simulation study was made for Routine 1A and for Routine 1B and the Estimators (A), (B) and (C) are compared.

FIGURE 3. Routines 1A and 1B for a quadratic response function.

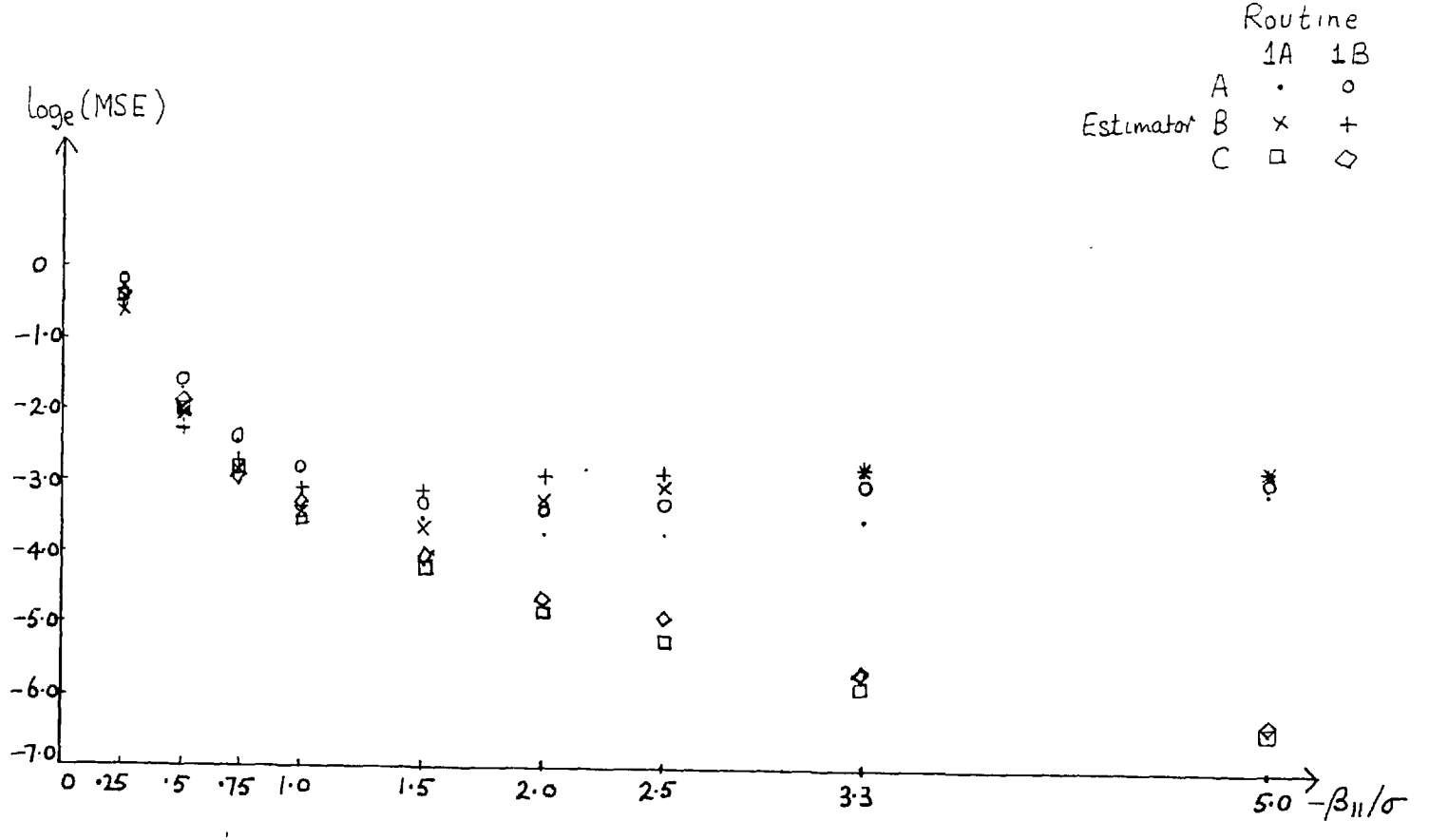


TABLE 2

THE ESTIMATION OF THE OFL FOR
A QUADRATIC RESPONSE FUNCTION ($-\beta_{11}/\sigma = 5.0$)

ROUTINE 1

		ESTIMATOR (A)	ESTIMATOR (B)	ESTIMATOR (C)
ROUTINE 1A	MEAN (S.ERROR)	0.2064 (0.003)	-0.2498 (0.000)	-0.0140 (0.002)
	VAR (S.ERROR)	0.0042 (0.000)	0.0000 (0.000)	0.0014 (0.000)
	MSE (S.ERROR)	0.0468 (0.000)	0.0624 (0.000)	0.0016 (0.000)
	LOG(MSE)	-3.0614	-2.7739	-6.4573
ROUTINE 1B	MEAN (S.ERROR)	0.2376 (0.002)	-0.2496 (0.000)	-0.0069 (0.002)
	VAR (S.ERROR)	0.0018 (0.000)	0.0000 (0.000)	0.0017 (0.000)
	MSE (S.ERROR)	0.0583 (0.000)	0.0017 (0.000)	0.0017 (0.000)
	LOG(MSE)	-2.8422	-2.7745	-6.3778

ROUTINE 2

ROUTINE 2A	MEAN (S.ERROR)	-0.0220 (0.014)
	VAR (S.ERROR)	0.1028 (0.000)
	MSE (S.ERROR)	0.1033 (0.007)
ROUTINE 2B	MEAN (S.ERROR)	-0.5000 (0.000)
	VAR (S.ERROR)	0.0000 (0.000)
	MSE (S.ERROR)	0.2500 (0.000)

The OFL is set at 0.0, the level spacing, 2d, at 1.0, the level nearest the OFL at 0.25 and the starting levels are set at -5.75 and -4.75. For each simulation 500 trials of 20 runs each were made, giving an average sample size between 26 and 28. The mean squared errors of the estimate of the OFL were obtained for a range of β_{11}/σ. A plot of $\log_e(\text{MSE})$ against $-\beta_{11}/\sigma$ is given in Figure 3, and in Table 2 the mean, variance, mean squared error together with their standard errors are given for $-\beta_{11}/\sigma = 5.0$.

In comparing Routines 1A and 1B and Estimators (A), (B) and (C) we shall judge two quantities to be significantly different if their intervals of three standard errors on either side of the estimate of the mean squared error do not overlap. The standard errors of the mean squared error were calculated under the assumption of normality.

Of the three estimators, Estimator (C) is the best for $-\beta_{11}/\sigma$ in the range 1.5 to 5.0. In the range 0.5 to 1.0 there is no difference between Estimators (B) and (C) but both are better than Estimator (A). For $-\beta_{11}/\sigma = 0.01$ (not shown in Figure 3) and $-\beta_{11}/\sigma = 0.25$, there is no real difference between the estimators.

For the estimators which do better Routine 1B does

not give an improved performance over Routine 1A. We may thus conclude that there is nothing to be gained by modifying Routine 1A in the particular way given by Routine 1B.

Of the six combinations Routine 1A, Estimator (C) appears to be uniformly the most satisfactory, over the range of $-\beta_{11}/\sigma$ considered.

Empirical Results for the Generalized Up-Down Procedure.

A simulation study was made for Routine 2A and for Routine 2B under basically the same conditions as for Routines 1A and 1B . The starting levels were at -5.5 and -4.5 , and 500 trials, giving an average sample size between 25 and 26, were made. The mean squared errors of the OFL were obtained for a range of β_{11}/σ.

These mean squared errors decrease as $-\beta_{11}/\sigma$ increases from 0.01 to 5.0 but were very much larger than the corresponding values for Routines 1A and 1B, even for small values of $-\beta_{11}/\sigma$. The mean, variance and mean squared error, together with their standard errors, for $-\beta_{11}/\sigma = 5.0$, are given in Table 2.

We should note that because Routine 2B has a high

probability of a failed contraction (Figures 2f and 2g), the sequence of designs gets stuck on approaching the optimum region, for large $-\beta_{11}/\sigma$. In this case no contraction occurs and the sequence of designs oscillates over three levels in the neighbourhood of the OFL. The mean squared error then tends to the square of the bias as $-\beta_{11}/\sigma$ increases, Table 2 . Routine 2A has a smaller probability of a failed contraction and the mean squared error decreases slowly as $-\beta_{11}/\sigma$ increases.

From qualitative considerations which led to the design of these procedures and from the quantitative results given in this section we may conclude that, for exploration and rapid movement to the regions of higher response, from regions distant from the optimum region, Routine 2B is more successful than either Routine 2A or Routine 1A. However, for rapid convergence to a smaller neighbourhood of the OFL, given a starting point in the optimum region, Routine 1A is much more successful than either Routine 2A or Routine 2B. Therefore, it seems reasonable to consider the following automatic procedure which is a combination of Routine 2B and Routine 1A (using Estimator (C)) .

Let x_1 be the starting value. Then proceed as for Routine 2B, Section 2.2.2, by making two observations at $x_1 - d_1$ and $x_1 + d_1$, where d_1 is chosen by the experimenter. Routine 2B is stopped when a failed contraction occurs (Figures 2f and 2g), as this is some indication that we are near to the optimum region. Routine 2B is also stopped if a design point falls outside the experimental region. Let x_H be the level corresponding to the highest response for the step when Routine 2B is stopped. Then proceed as for Routine 1A, Section 2.2.1, by making two observations at $x_H - d_S$ and $x_H + d_S$, where d_S may be taken equal to d_1 to preserve the automatic property of the procedure or where d_S may be chosen to be consistent with any gain in information by the experimenter since the start of the procedure. We shall refer to this combination of Routine 1A and Routine 2B, as Routine 3. A flow diagram for it is given in the Appendix.

In Section 2.1.5 we described the normalized version of the accelerated K-W procedure, which we shall refer to as Routine 4. A flow diagram for it is given in the Appendix. With Routine 4 as a nucleus, Routine 3 was designed as a consequence of qualitative arguments supported by quantitative results. We shall now show to what extent Routine 3 is an

improvement over Routine 4, apart from its automatic property, by comparing their performances for different simulated experimental conditions. This comparison is not intended as an exhaustive test of the two procedures but should serve as a guide to their relative merits.

Consider the response curve

$$\eta = \exp(-x^2) \quad \text{(Figure 4)} \ .$$

The OFL is at 0.0 . Three starting levels (x_1)

$$S_1 = -2.6 \ , \quad S_2 = -1.1 \quad \text{and} \quad S_3 = 0.1 \ ,$$

two initial level spacings ($2d_1$)

$$D_1 = 0.2 \ , \quad D_2 = 0.5$$

and three error standard deviations (σ)

$$\sigma_1 = 0.001 \ , \quad \sigma_2 = 0.05 \quad \text{and} \quad \sigma_3 = 0.2$$

are used for both procedures; thus giving 18 sets of results for comparison. The estimate of the OFL and its standard error are obtained after 500 trials for each procedure, Table 4.

Routine 3 is stopped after $R_1(=2)$ and $R_2(=10)$ runs of Routine 1A and the mean and the variance of the number of observations used, are recorded.

For Routine 4 a sequence (a_n) is considered such that

$$a_{n+1} = (1/2)a_n \ ,$$

where two initial step sizes (a_1) 0.4 and 0.8 are used. Routine 4 is stopped after a fixed number of observations to facilitate comparison with Routine 3.

If the responses y_1 and y_2, at $x_1 - d_1$ and $x_1 + d_1$, are normally distributed with means η_1 and η_2 respectively, $\eta_2 > \eta_1$, and variance σ^2; then, if $h = \eta_2 - \eta_1 > 2.3\sigma$ the probability that $y_1 > y_2$ is less that 5%. Figure 4 indicates whether the sequence of designs is given a good chance of progressing for a particular combination of x_1, d_1 and σ. For example, given S_1, D_1 and σ_3 then

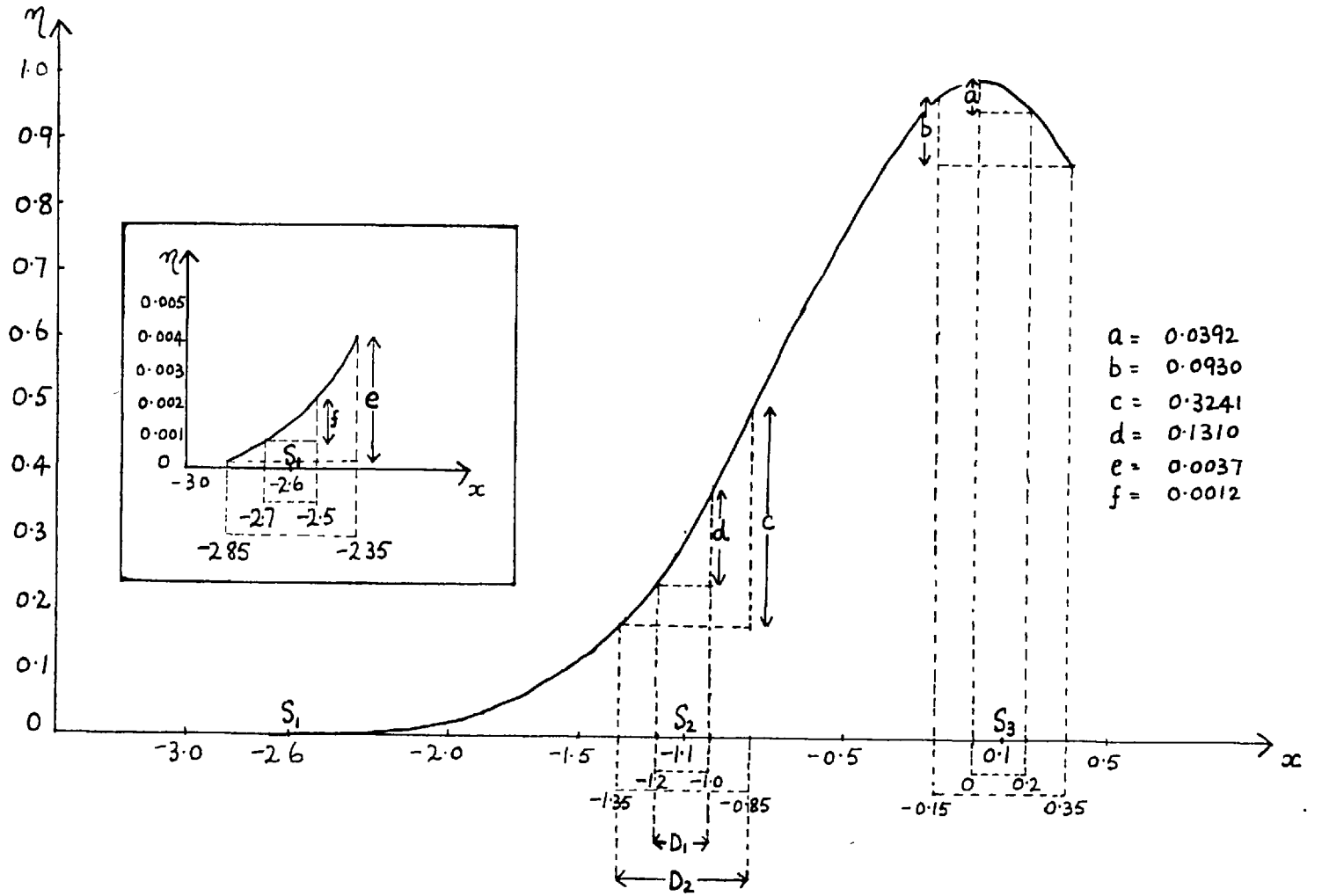
$$h(=0.0012) < 2.3\sigma\,(=0.46) \ ,$$

and the probability that $y_1 > y_2$ is not very small. Then, the sequence of designs experiences great difficulty in moving away from the starting region (Table 4). On the other hand, given S_2, D_1 and σ, then

$$h(=0.1310) > 2.3\sigma\,(=0.0023) \ ,$$

and the probability that $y_1 > y_2$ is very small. Then the sequence of designs easily progresses to the optimum region (Table 4).

FIGURE 4. The response function $\eta = \exp(-x^2)$.



The quantity h/σ for all the combinations of x_1, d_1 and σ which appear in Table 4, are given in Table 3.

TABLE 3

Values of h/σ associated with the response function $\exp(-x^2)$.

		σ_1	σ_2	σ_3
S_1	D_1	1.20	0.024	0.006
	D_2	3.70	0.074	0.019
S_2	D_1	131.00	2.620	0.655
	D_2	324.10	6.482	1.621
S_3	D_1	39.20	0.784	0.196
	D_2	93.00	1.860	0.465

For Routine 3, Table 4 shows how the estimate of the bias is improved as R is increased from 2 to 10. There is little improvement for small σ or small $|x_1|$. The increase in the initial design size in general reduces the mean and variance of the number of observations and hence improves the speed of convergence to the optimum region; it also results in reducing the estimate of the bias.

For Routine 4, given S_1, the larger number of observations in general fails to produce a decrease of the bias.

This is perhaps due to the basic design of Routine 4 where the step size is reduced whenever the search changes direction, and will not necessarily be remedied by choosing a sequence which decreases less rapidly. Also, given S_1, there is nothing to be gained by using the larger initial step size.

On the basis of Table 4, given S_3, there is no real difference in the performances of Routine 3 and Routine 4. Given S_2, Routine 3 is uniformly better particularly so as σ increases. Given S_1, Routine 3 performs much better than Routine 4. We therefore conclude that Routine 3 is an automatic procedure which, overall, is more successful in improving the initial estimate of the OFL than is Routine 4.

Routine 3 and Routine 4 are both based on the sign, and not on the magnitude, of the difference between a pair of responses; hence, even when there is asymmetry in the response curve in the optimum region, we do not expect the behaviour of these procedures to differ very much from that described above.

We have not included in our simulation study the automatic stopping rule or the rules for dealing with constraints on the experimental region.

TABLE 4

COMPARISON OF ROUTINE 3 AND ROUTINE 4

STARTING POINT (x_1) = −2.60 . INITIAL DESIGN SIZE (d_1) = 0.10 .

		ROUTINE 3				ROUTINE 4			
EXPTL. ERROR S.DEV.(σ)	NUM. OF RUNS	NUM. OF OBS. PER TRIAL MEAN	VARIANCE	EST. OF OFL MEAN	S.ERROR	NUM. OF OBS.	INITIAL STEP SIZE(a_1)	EST. OF OFL MEAN	S.ERROR
0.001	2	16.4	6.78	−0.25	0.038	14	0.40	−0.36	0.056
0.001	2	16.4	6.78	−0.25	0.038	14	0.80	−0.55	0.061
0.001	10	33.2	9.53	−0.09	0.021	34	0.40	−0.52	0.055
0.001	10	33.2	9.53	−0.09	0.021	34	0.80	−0.52	0.062
0.050	2	14.9	22.51	−2.23	0.055	14	0.40	−2.35	0.052
0.050	2	14.9	22.51	−2.23	0.055	14	0.80	−2.20	0.084
0.050	10	36.8	29.70	−1.97	0.063	34	0.40	−2.32	0.061
0.050	10	36.8	29.70	−1.97	0.063	34	0.80	−2.06	0.087
0.200	2	14.0	15.75	−2.45	0.043	14	0.40	−2.50	0.041
0.200	2	14.0	15.75	−2.45	0.043	14	0.80	−2.51	0.080
0.200	10	34.8	24.50	−2.44	0.050	34	0.40	−2.55	0.043
0.200	10	34.8	24.50	−2.44	0.050	34	0.80	−2.67	0.084

TABLE 4 (CONT'D)

COMPARISON OF ROUTINE 3 AND ROUTINE 4

STARTING POINT (x_1) = -2.60 . INITIAL DESIGN SIZE (d_1) = 0.25 .

EXPTL. ERROR S.DEV. (σ)	NUM. OF RUNS	NUM. OF OBS. PER TRIAL MEAN	VARIANCE	ROUTINE 3 EST. OF OPT. MEAN	S.ERROR	ROUTINE 4 NUM. OF OBS.	INITIAL STEP SIZE (a_1)	EST. OF OPT. MEAN	S.ERROR
0.001	2	14.0	0.00	-0.01	0.000	14	0.40	0.20	0.004
0.001	2	14.0	0.00	-0.01	0.000	14	0.80	-0.01	0.007
0.001	10	30.0	0.00	-0.01	0.000	30	0.40	-0.00	0.001
0.001	10	30.0	0.00	-0.01	0.000	30	0.80	-0.02	0.002
0.050	2	14.4	11.64	-1.77	0.098	14	0.40	-2.36	0.052
0.050	2	14.4	11.64	-1.77	0.098	14	0.80	-2.15	0.084
0.050	10	22.6	13.42	-1.48	0.102	34	0.40	-2.31	0.061
0.050	10	32.6	13.42	-1.48	0.102	34	0.80	-2.06	0.087
0.200	2	14.5	16.39	-2.24	0.095	14	0.40	-2.42	0.047
0.200	2	14.5	16.39	-2.24	0.095	14	0.80	-2.10	0.080
0.200	10	34.2	13.97	-1.87	0.099	34	0.40	-2.43	0.051
0.200	10	34.2	13.97	-1.87	0.099	34	0.80	-2.40	0.089

TABLE 4 (CONT'D)

COMPARISON OF ROUTINE 3 AND ROUTINE 4

STARTING POINT (x_1) = -1.10 . INITIAL DESIGN SIZE (d_1) = 0.10 .

	ROUTINE 3						ROUTINE 4			
EXPTL. ERROR S.DEV.(σ)	NUM. OF RUNS	NUM. OF OBS. PER TRIAL MEAN	VARIANCE	EST. OF OPT MEAN	S.ERROR	NUM. OF OBS.	INITIAL STEP SIZE(a_1)	EST. OF OPT MEAN	S.ERROR
0.001	2	13.7	0.70	0.00	0.000	14	0.40	0.01	0.002
0.001	2	13.7	0.70	0.00	0.000	14	0.80	0.00	0.000
0.001	10	32.5	3.16	0.00	0.000	34	0.40	0.00	0.000
0.001	10	32.5	3.16	0.00	0.000	34	0.80	0.00	0.000
0.050	2	13.8	0.82	0.02	0.004	14	0.40	-0.05	0.009
0.050	2	13.8	0.82	0.02	0.004	14	0.80	-0.02	0.014
0.050	10	33.2	4.15	0.00	0.002	34	0.40	-0.02	0.007
0.050	10	33.2	4.15	0.00	0.002	34	0.80	0.00	0.010
0.200	2	14.8	11.55	-0.36	0.029	14	0.40	-0.59	0.032
0.200	2	14.8	11.55	-0.36	0.029	14	0.80	-0.59	0.065
0.200	10	34.8	11.47	-0.20	0.021	34	0.40	-0.62	0.036
0.200	10	34.8	11.47	-0.20	0.021	34	0.80	-0.51	0.054

TABLE 4 (CONT'D)

COMPARISON OF ROUTINE 3 AND ROUTINE 4

STARTING POINT (x_1) = -1.10 . INITIAL DESIGN SIZE (d_1) = 0.25 .

	ROUTINE 3						ROUTINE 4			
EXPTL. ERROR S.DEV.(σ)	NUM. OF RUNS	NUM. OF OBS. PER TRIAL MEAN	VARIANCE	EST. OF OFL MEAN	S.ERROR	NUM. OF OBS.	INITIAL STEP SIZE(a_1)	EST. OF OFL MEAN	S.ERROR
0.001	2	11.0	0.00	-0.01	0.000	10	0.40	0.00	0.000
0.001	2	11.0	0.00	-0.01	0.000	10	0.80	-0.10	0.000
0.001	10	27.0	0.00	-0.01	0.000	30	0.40	0.00	0.000
0.001	10	27.0	0.00	-0.01	0.000	30	0.80	0.00	0.000
0.050	2	11.0	0.09	-0.00	0.002	10	0.40	-0.00	0.006
0.050	2	11.0	0.09	-0.00	0.002	10	0.80	-0.07	0.004
0.050	10	27.3	0.57	-0.00	0.001	30	0.40	-0.00	0.002
0.050	10	27.3	0.57	-0.00	0.001	30	0.80	0.00	0.002
0.200	2	12.5	3.68	-0.06	0.018	14	0.40	-0.16	0.019
0.200	2	12.5	3.68	-0.06	0.018	14	0.80	-0.09	0.027
0.200	10	31.7	7.91	-0.06	0.022	30	0.40	-0.12	0.017
0.200	10	31.7	7.91	-0.06	0.022	30	0.80	-0.18	0.038

TABLE 4 (CONT'D)

COMPARISON OF ROUTINE 3 AND ROUTINE 4

STARTING POINT (x_1) = 0.10 . INITIAL DESIGN SIZE (d_1) = 0.10 .

	ROUTINE 3						ROUTINE 4			
EXPTL. ERROR S.DEV.(σ)	NUM. OF RUNS	NUM. OF OBS. PER TRIAL MEAN	VARIANCE	EST. OF OFL MEAN	S.ERROR	NUM. OF OBS.	INITIAL STEP SIZE(a_1)	EST. OF OFL MEAN	S.ERROR
0.001	2	11.0	3.19	-0.00	0.000	10	0.40	-0.03	0.003
0.001	2	11.0	3.19	-0.00	0.000	10	0.80	0.00	0.000
0.001	10	29.7	5.85	-0.00	0.000	30	0.40	0.00	0.000
0.001	10	29.7	5.85	-0.00	0.000	30	0.80	0.00	0.000
0.050	2	11.4	4.85	0.00	0.004	10	0.40	-0.00	0.007
0.050	2	11.4	4.85	0.00	0.004	10	0.80	0.01	0.013
0.050	10	30.4	7.16	0.00	0.002	30	0.40	-0.01	0.004
0.050	10	30.4	7.16	0.00	0.002	30	0.80	-0.01	0.006
0.200	2	12.1	6.86	0.04	0.011	14	0.40	0.01	0.020
0.200	2	12.1	6.86	0.04	0.011	14	0.80	-0.09	0.050
0.200	10	32.3	10.89	-0.00	0.009	34	0.40	0.03	0.022
0.200	10	32.3	10.89	-0.00	0.009	34	0.80	-0.01	0.044

TABLE 4 (CONT'D)

COMPARISON OF ROUTINE 3 AND ROUTINE 4

STARTING POINT (x_1) = 0.10 . INITIAL DESIGN SIZE (d_1) = 0.25 .

	ROUTINE 3						ROUTINE 4			
EXPT'L. ERROR S.DEV.(σ)	NUM. OF RUNS	NUM. OF OBS. PER TRIAL MEAN	VARIANCE	EST. OF OFL MEAN	S.ERROR	NUM. OF OBS.	INITIAL STEP SIZE(a_1)	EST. OF OFL MEAN	S.ERROR
0.001	2	9.0	0.00	0.01	0.000	10	0.40	-0.03	0.003
0.001	2	9.0	0.00	0.01	0.000	10	0.80	0.00	0.000
0.001	10	25.0	0.00	0.01	0.000	30	0.40	-0.00	0.000
0.001	10	25.0	0.00	0.01	0.000	30	0.80	-0.00	0.000
0.050	2	9.1	0.12	0.00	0.002	10	0.40	-0.02	0.004
0.050	2	9.1	0.12	0.00	0.002	10	0.80	0.00	0.005
0.050	10	25.3	0.52	0.00	0.001	30	0.40	0.00	0.001
0.050	10	25.3	0.52	0.00	0.001	30	0.80	-0.00	0.001
0.200	2	10.6	3.00	0.01	0.010	10	0.40	0.01	0.013
0.200	2	10.6	3.00	0.01	0.010	10	0.80	0.03	0.026
0.200	10	29.5	6.90	0.00	0.009	30	0.40	0.00	0.007
0.200	10	29.5	6.90	0.00	0.009	30	0.80	0.00	0.016

CHAPTER 3

MULTI-FACTOR PROCEDURES

3.1 Existing Procedures.

As the number of factors increases one is faced with the prospect of more complicated response surfaces. The unimodality assumption becomes less plausible. The choice of a starting point for the procedure becomes more important. In addition to the previous requirements for a good procedure, emphasis should be placed on procedures which allow the sequence of designs to continue exploration. Then, important changes in the surface may be detected as soon as possible. This requirement was not so important in the single factor case because of the limited nature of the response curves. We shall now review existing multi-factor procedures, criticizing any shortcomings and indicating directions in which improvements should be considered.

3.1.1 Non-sequential Procedures.

As we have already mentioned in Section 1.2, sequential procedures are preferable to non-sequential procedures. However, two non-sequential procedures, the random method and the factorial method, have been discussed in the literature, (Brooks, 1959). We shall describe them only briefly here.

In the random method the experimental region is sampled randomly either by simple random sampling or by stratified sampling, and the factor combination corresponding to the highest observed response is taken as the estimate of the OFC. In the simple random sampling method it is possible that important regions of high response will not be discovered. Stratified random sampling limits this but as the factor space increases, only dense stratification or good fortune can guarantee the successful use of the random method. However, the method may be suitable for a preliminary survey of the experimental region to determine a starting point for an initial stage procedure.

In the factorial method a grid of points in the form of a factorial design is considered over the experimental region. If this grid is sufficiently dense, this method gives an overall picture of the response surface. However, since our objective is to estimate the OFC, this method - although more systematic than the random method - will expend experimental effort in regions which contribute little relevant information and so is unsatisfactory.

3.1.2 The One-at-a-time Procedure.

This is one of the oldest optimum seeking procedures, proposed formally by Friedman and Savage (1947), and is simple to operate. One factor is varied, the others remaining constant at their starting values, until a conditional maximum is achieved. The second factor is then varied, the others remaining constant at their values for the first conditional maximum, until a second conditional maximum is achieved. This process is continued until the last factor has been varied when the whole operation is repeated starting again with the first factor. When no further progress can be made the OFC is estimated by the factor combination corresponding to the final conditional maximum. A disadvantage of this one-at-a-time procedure is that a point on a rising ridge is often chosen as the estimate of the OFL.

If all the factors affecting the response are considered simultaneously then a smaller number of observations will in general be required to achieve optimum conditions than if the factors were considered one at a time. Procedures considering several factors simultaneously are described in the following sections.

3.1.3 The Box-Wilson Procedure.

Box and Wilson (1951) paved the way for the development of response surface methodology. In particular they developed the following steepest ascent procedure.

From a starting point S, say, it is desired to proceed to another point, P, a specified distance from S, where the gain in response is maximum. Box and Wilson prove that to do this one should move in the direction perpendicular to the local contours at S, i.e. in the direction of steepest ascent. To estimate the direction of steepest ascent, a first order design is used in a region about S, over which a first degree polynomial is assumed to be an adequate representation of the surface. Experiments are made in the direction of steepest ascent until P is found. From P the process is repeated and another direction of steepest ascent is found. The process is continued until the first order effects of the factors, estimated from the current design, are so reduced that they are considered small in comparison with higher order effects. At this stage it is assumed that a NSR has been reached. A second degree polynomial is fitted and an estimate of the OFC is obtained.

3.1.4 Extensions of the Kiefer-Wolfowitz Procedure.

(1) Blum (1954) proposed an extension of the K-W procedure which generates a sequence of designs each consisting of (k+1) points. The n^{th} design consists of the base point $\underline{z}_n$, and one point a distance c_n from $\underline{z}_n$ parallel to each of the k axes representing the k factors, thus forming an asymmetric design. For $k = 2$ the design points are at the vertices of a right-angled triangle. The procedure is then defined by

$$\underline{z}_{n+1} = \underline{z}_n + (a_n/c_n)\,\underline{G} \quad ,$$

where $\underline{G}$ is an estimate of the gradient vector at $\underline{z}_n$. The sequences (a_n) and (c_n) satisfy the same conditions as in Section 2.1.3. This sequence of designs can be shown to converge to the OFC both in mean square and with probability one under no stiffer restrictions than for the K-W procedure described in Section 2.1.3.

Sacks (1958), however, has pointed out that the asymmetry of the design points forces the sequence of designs to concentrate around the wrong value and hence retards the speed of convergence.

(2) Sacks (1958) proposed a symmetric extension of the K-W procedure in which each design consists of $2k$ points symmetrically placed about the base point $\underline{z}_n$. The n^{th} design consists of two points, a distance c_n from $\underline{z}_n$, one on either side of $\underline{z}_n$, parallel to each of the k axes representing the k factors. For $k = 2$ the design points are at the vertices of a square centred about $\underline{z}_n$. The procedure is then defined by

$$\underline{z}_{n+1} = \underline{z}_n + (a_n/c_n)\,\underline{G}^* \quad ,$$

where $\underline{G}^*$ is an estimate of the gradient vector at $\underline{z}_n$. Under the same conditions as in (1) this procedure converges in mean square and with probability one.

3.1.5 The Spendley-Hext-Himsworth Simplex Procedure.

Spendley, Hext and Himsworth (1962) proposed an automatic procedure based on a simplex design which has $(k+1)$ points. There is no loss in generality in assuming the simplex design to be regular. For $k = 2$ the design points for a regular simplex design are at the vertices of an equilateral triangle and for $k = 3$ they are at the vertices of a regular tetrahedron. The initial simplex is centred about the starting point and responses are observed. The design point

corresponding to the lowest observed response is then replaced by its mirror image in the hyperplane containing the remaining points, thus forming a new simplex design. This operation is then repeated and two additional rules are given:

(i) If the mirror image, i.e. the new point for the next design, gives the lowest response in the new design, the point with the second lowest response is determined and is replaced by its mirror image in the hyperplace containing the remaining points. This rule is designed to allow the sequence of designs to explore in the search of higher response instead of oscillating over a limited region.

(ii) If a point has occurred in (k+1) successive simplex designs and is not then eliminated, a repeat observation should be made at this point. This rule is designed to prevent the sequence of designs from being anchored to a point with a spuriously high response. This procedure will eventually cause the sequence of designs to circle continuously around the point corresponding to the highest observed response which can be taken as an estimate of the OFC.

3.1.6 The Nelder-Mead Simplex Procedure.

Nelder and Mead (1965) extended the ideas of Spendley, Hext and Himsworth (1962) and produced a less rigid procedure for use in the deterministic case. A brief description of this procedure is given here and will be followed by a modified version, suitable for stochastic search, in Section 3.2.1.

The initial simplex design is chosen and the suffices h and l are associated with the highest and lowest observed responses respectively. The point P_l is replaced at each stage by a new point which forms a new simplex design with the remaining points. Three operations - reflection, expansion and contraction - which allow the simplex design to adapt itself to the local landscape, are used.

Let P_m be the centroid of all the points not including P_l. Then form P_r by the relation

$$P_r = (1+\alpha)P_m - \alpha P_l ,$$

where $\alpha\ (>0)$ is the reflection coefficient. The design points are chosen as follows:

(a) If $y_l < y_r < y_h$, then take P_r as the new design point.

(b) If $y_r > y_h$, then expand P_r to P_e by the relation

$$P_e = \gamma P_r + (1-\gamma)P_m ,$$

where γ (>1) is the expansion coefficient.

(i) If $y_e > y_h$, then take P_e as the new design point.

(ii) If $y_e < y_h$, then take P_r as the new design point.

(c) If $y_r < y_u$ for all $u(u \neq l)$, then

let $P = P_r$ if $y_r > y_l$ or

let $P = P_l$ if $y_r < y_l$ and

contract P to P_c by the relation

$$P_c = \beta P + (1-\beta)P_m ,$$

where β ($0 < \beta < 1$) is the contraction coefficient.

(i) If $y_c > y_l$, then take P_c as the new design point.

(ii) If $y_c < y_l$, then replace the points P_u, for all u, by $(P_u + P_h)/2$ to form the new design.

Nelder and Mead (1965) show that the simple strategy $\alpha = 1$, $\beta = 0.5$ and $\gamma = 2$ was clearly the best among those tried.

3.1.7 Discussion.

For the Box-Wilson procedure, which has been extensively used in the chemical industry, the general knowledge and judgement of the experimenter are invaluable for its success. It is thus not an automatic procedure but it may be argued that this is on the whole an advantage. However, it is of interest to consider more automatic procedures. The direction of steepest ascent differs depending on the choice of scale of the factors. Hence, the procedure is not scale invariant. The authors admit that this method is only suitable in cases where the experimental error is fairly small.

The Sacks extension of the K-W procedure has an automatically determined step size and design size but there is still the difficulty of choosing beforehand suitable sequences (a_n) and (c_n) . This procedure is not scale invariant.

The Spendley-Hext-Himsworth procedure conforms more to the features of a good procedure, outlined in Section 1.3, than the other procedures. This procedure is operated by a set of automatic rules, is scale invariant but is rather rigid in design and hence does not adapt itself to changes in the response surface. Nelder and Mead remedied this by

introducing rules which change the shape and size of the simplex design depending on the responses observed. In Section 3.2 these ideas are extended in an attempt to describe an automatic procedure suitable for multi-factor stochastic search.

Brooks (1959) gives a comparison of the random method, the factorial method, the one-at-a-time procedure and the Box-Wilson steepest ascent procedure. A few trials were made for two factors over four different response surfaces and the average performances of the procedures were compared. He concludes that the steepest ascent procedure is the best. However, to operate the steepest ascent procedure, one has to choose the step size thus introducing a subjective element into the procedure. It is this lack of objectivity in a procedure that weakens the conclusion in a comparison such as this.

It is of interest therefore to define procedures possessing this objectivity and then comparing them over several surfaces and for a greater number of factors. Two such procedures are described in Sections 3.2.1 and 3.2.2.

3.2 New Procedures.

In describing and comparing the procedures in Sections 3.2.1 and 3.2.2 it will be necessary to consider the 'size' of a design. Box and Wilson (1951) considered the quantity

$$S_t^2 = (1/N) \sum_{u=1}^{N} (x_{tu} - \bar{x}_{t.})^2$$

for a design with N points, where

$$\bar{x}_{t.} = (1/N) \sum_{u=1}^{N} x_{tu} \quad .$$

Here S_t^2 is called the spread associated with the t^{th} factor. If all the spreads S_t^2 for one design are equal to the corresponding S_t^2 for another design then the two designs are said to be of the same 'size'. Other measures are certainly possible, for instance

$$\underset{u}{\text{Max}}(x_{tu}) - \underset{u}{\text{Min}}(x_{tu}) \quad .$$

3.2.1 The Generalized Simplex Procedure.

The procedure now to be described is based on the Spendley-Hext-Himsworth and the Nelder-Mead simplex procedures.

Here, we consider the initial simplex design to be regular. This facilitates the comparison of this procedure with other procedures. The co-ordinates of the (k+1) vertices of a regular simplex design for k factors, having one vertex at the origin, are given in matrix form by a (k+1) x k matrix

$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \\ k+1 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & \cdots & k \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ p & q & q & \cdots & q \\ q & p & q & \cdots & q \\ \vdots & & & & \\ q & q & q & \cdots & p \end{bmatrix} \end{array} ,$$

where

$$p = \left\{ 1/(k\sqrt{2}) \right\} \left(\sqrt{k+1} + k - 1 \right)$$

and

$$q = \left\{ 1/(k\sqrt{2}) \right\} \left(\sqrt{k+1} - 1 \right) .$$

Here the edges of the simplex are unity and the centroid of the design is the point $(a,a,\ldots,a)$, where

$$a = 1/\sqrt{2(k+1)} .$$

For k = 2

$$p = \left\{ 1/(2\sqrt{2}) \right\} \left(\sqrt{3} + 1 \right),$$

$$q = \left\{ 1/(2\sqrt{2}) \right\} \left(\sqrt{3} - 1 \right)$$

and the centroid is at $(1/\sqrt{6}, 1/\sqrt{6})$.

The experience of the experimenter is used in choosing the size of the initial simplex design. At each step of the procedure the point corresponding to the lowest observed response is replaced by a new point to form a new simplex design. The new point is chosen according to the following rules:

Denote the response at the vertex, P_u , of a simplex by y_u . Let

$$y_h = \underset{u}{\text{Max}}(y_u) \ ,$$

$$y_l = \underset{u}{\text{Min}}(y_u)$$

and

$$y_s = \underset{u \neq l}{\text{Min}}(y_u) \ ,$$

and let P_m be the centroid of the points for which $u \neq l$. Form P_r by the relation

$$P_r = 2P_m - P_l \ .$$

(a) If $y_r > y_h$, expand P_r to P_e by the relation

$$P_e = (3P_r - P_m)/2 \ .$$

(i) If $y_e > y_h$, then take P_e as the new point (Figure 5a).

(ii) If $y_e < y_h$, then take P_r as the new point (Figure 5b).

(b) If $y_s < y_r < y_h$, then take P_r as the new point (Figure 5c).

(c) If $y_r < y_s$, contract P_r to P_c by the relation

$$P_c = (P_m + P_r)/2 \ .$$

(i) If $y_c > y_h$, then take P_c as the new point (Figure 5d).

(ii) If $y_c < y_h$, then take P_1 as the new point (Figure 5e).

There are thus five ways in which the replacement for P_l may be obtained. Three of these lead to the choice of P_r and one each to the choice of P_e and P_c. This procedure as we have described it so far differs from the Nelder-Mead procedure in the treatment of a failed contraction, (c)(ii) above. In the presence of error, step size should not be continually reduced as sooner or later real effects will become confounded with error, thus halting the progress of the sequence of designs.

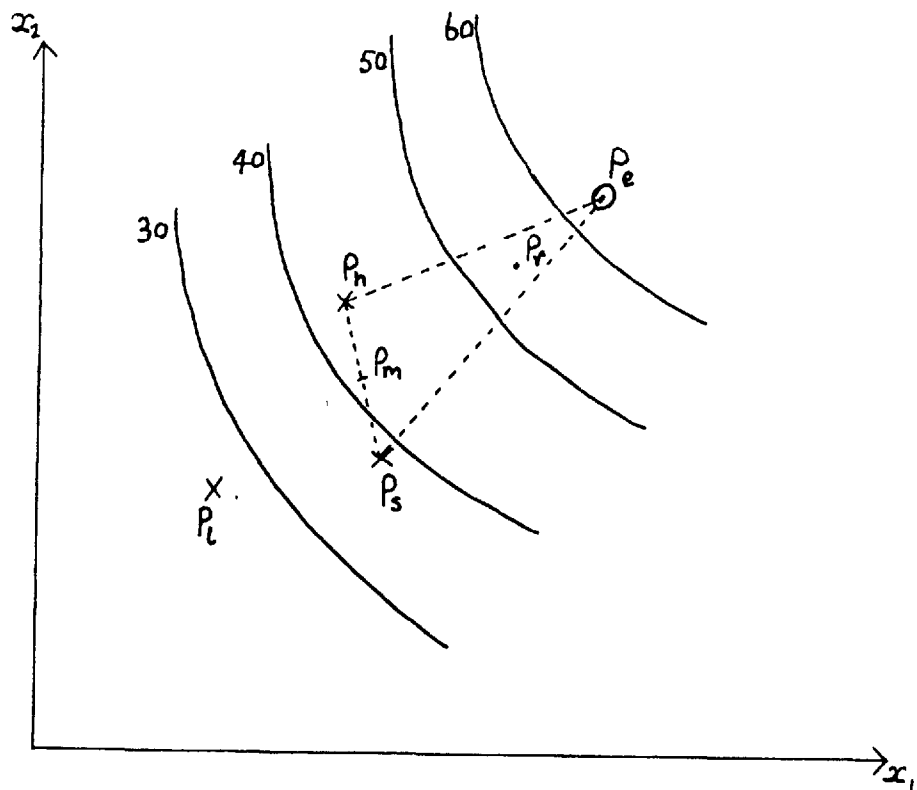
The following special rules are also necessary:

Rule A:

If the design point P_h in the current simplex design has occurred in (k+2) successive simplex designs and is not then replaced, a repeat observation should be made at that point and the procedure continued. This guards against the procedure becoming anchored at a point with a spuriously high response.

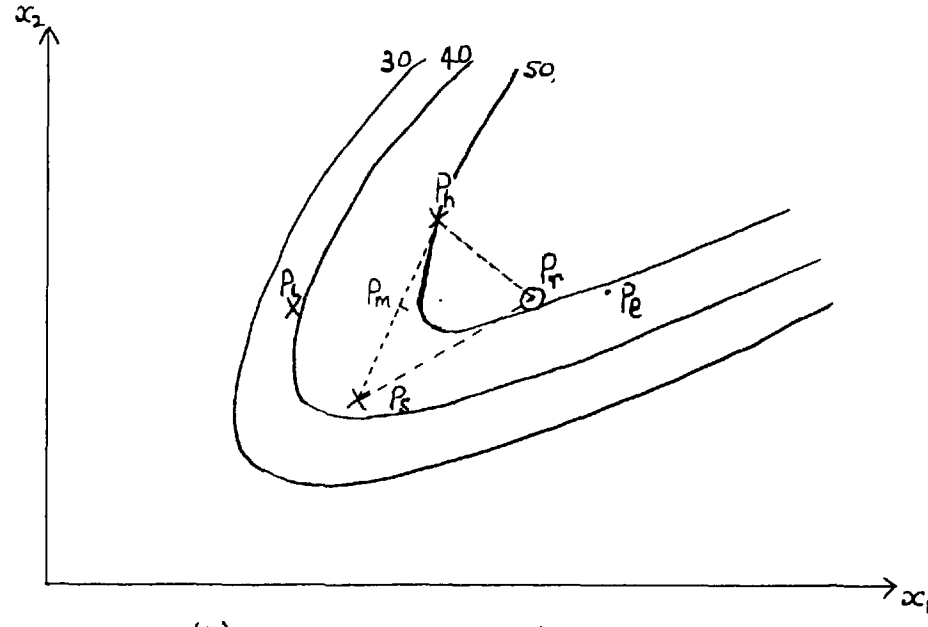
FIGURE 5. THE GENERALIZED SIMPLEX PROCEDURE ($K = 2$).

× denotes a point in the current design.
• denotes a point at which a subsequent observation is made.
⊙ denotes the point chosen as the new point for the next design.
----- joins the points which form the new design.

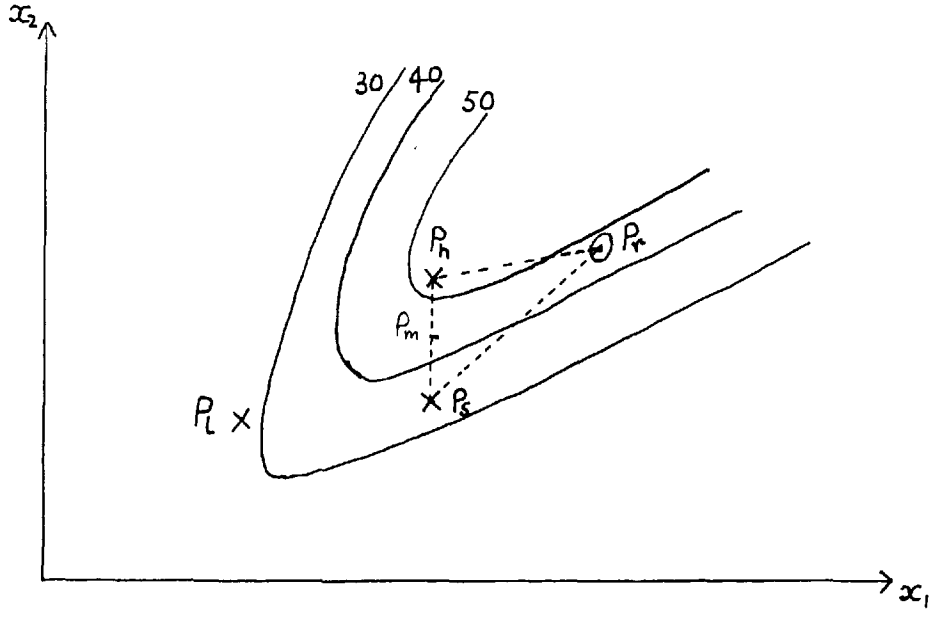


(a) $y_r > y_h$ and $y_e > y_h$.

FIGURE 5 (cont'd). THE GENERALIZED SIMPLEX PROCEDURE (K = 2).

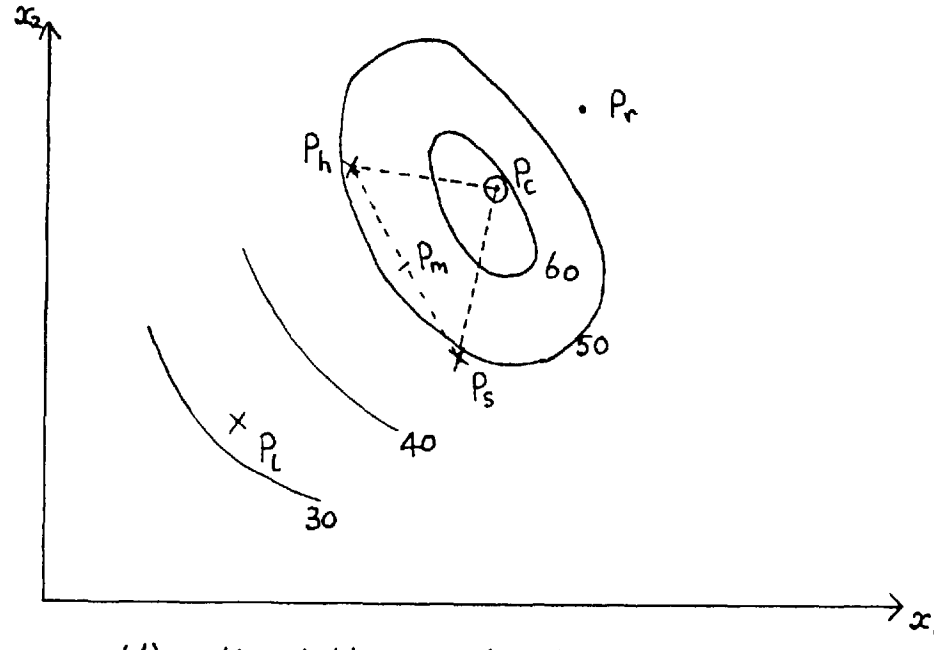


(b) $y_r > y_h > y_e$.

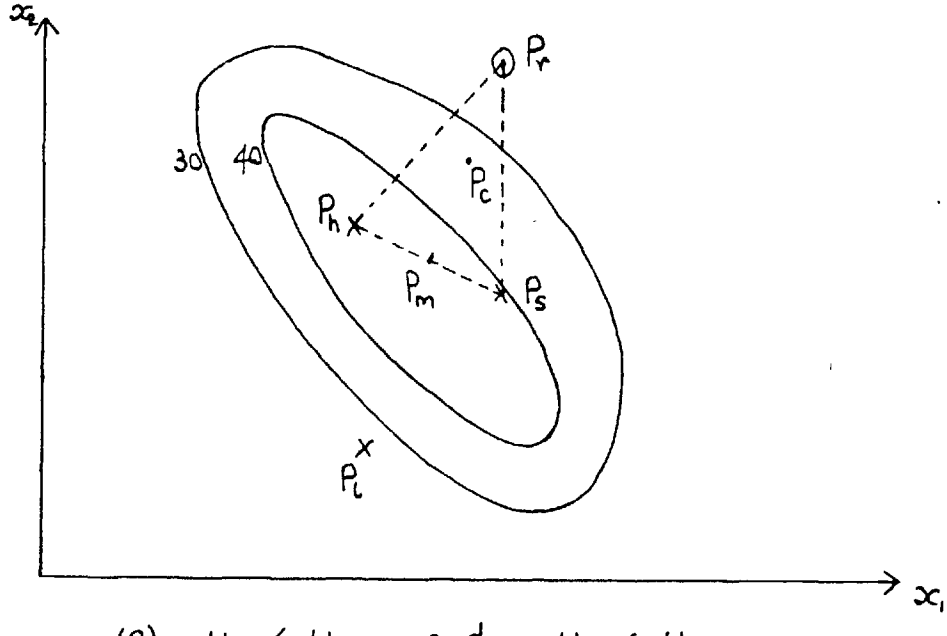


(c) $y_s < y_r < y_h$.

FIGURE 5 (cont'd). THE GENERALIZED SIMPLEX PROCEDURE (K=2).



(d) $y_r < y_s$ and $y_c > y_h$.



(e) $y_r < y_s$ and $y_c < y_h$.

Rule B (Stopping Rule):

If the point at which the repeat observation was made occurs in the next (k+1) designs, including the one containing the repeat observation, and is not then replaced, the procedure is stopped, as this is evidence that the optimum region has been found. We may estimate the OFC by the point corresponding to the highest observed response in the final design. We shall refer to this as the highest observation estimate (HOE). Alternatively, we may take as an estimate of the OFC a weighted average, W, of the points in the final design given by

$$W = \frac{\sum_{u=1}^{k+1} \left\{ y_u - (y - R) \right\} P_u}{\sum_{u=1}^{k+1} \left\{ y_u - (y - R) \right\}} ,$$

where $R = y_h - y$. We shall refer to this as the weighted observations estimate (WOE).

Rule C:

At any stage of the procedure if the response at the new point of a new design is the lowest for the new design (this is usually an indication that the sequence of designs has reached the vicinity of a NSR), then measure the spreads,

S_t^2 , t=1,2,...,k, for the current design. If any S_t^2 for the current design is greater than α(chosen by the experimenter) times the corresponding S_t^2 for the initial simplex design, then the procedure is restarted. To restart the procedure consider a new simplex design, the same size and orientation as the initial simplex design and centred about the point P_h in the current design, thus preserving the automatic property of the procedure. However, the size of this new simplex design may be chosen to be consistent with any gain in information by the experimenter since the start of the procedure. This rule helps to accelerate the search if the design size becomes too large. In choosing α the experimenter may be guided by the fact that, if α is too small the procedure will be restarted continually and hence the action of the stopping rule will be delayed; also, if α is too large then the procedure may not be restarted when it should be and further design points will be chosen which contribute little information about the NSR. Either case may result in an unneccessarily large number of observations being made.

Rule D:

If, however, all the S_t^2 of the current design are less than or equal to the corresponding S_t^2 of the initial design, then the procedure is continued by reflecting the point corresponding to the second lowest response in the hyperplane containing the remaining points, to obtain a new P_r . This rule prevents the sequence of designs from oscillating over a limited region and allows further exploration.

Rule E:

If P_r falls outside the experimental region then choose P_r as the new point for the new design and allocate to it (since no observation can be made) the lowest response in the new design. The procedure is then continued as in rule D. This rule eventually forces the sequence of designs to return to the experimental region.

If P_e falls outside the experimental region then P_r is taken as the new point for the new design and the procedure is continued. We assume that the boundary of the experimental region is such that there is no question of P_c falling outside.

We shall refer to this generalized simplex procedure

as Routine 5. It is automatic, scale invariant and because of the rules to change the size and shape of the simplex design, the sequence of designs is adapted to local changes in the response surface. Contraction occurs with small probability and this reduces the risk of the sequence of designs getting stuck because of large experimental error.

Now, in Routine 5 the search for higher response takes place in only one new direction at each step. This limited power to explore may increase the number of steps necessary to attain a given goal. In the next section we describe a procedure which is designed to give greater exploration at each step.

3.2.2 The Duplex Procedure.

Consider a design in k dimensions consisting of $2k$ points, where the points are divided into k pairs. The two members of each pair are called mates and the pairs are chosen such that the lines joining pairs of mates are concurrent at O, say, and that O lies in between each pair of mates. This design can be divided into two parts - k points in each part - by a $(k-1)$ dimensional hyperplane passing through O but not through

any of the design points. It is from this feature of the design that it derives the name, duplex design. The duplex procedure consists of a sequence of duplex designs formed by rules to be described below.

Here, we consider the initial duplex design to be regular. This facilitates the comparison of this procedure with other procedures. The coordinates of a regular duplex design centred at the origin and such that the line joining a pair of mates is parallel to an axis, are given by the $(2k \times k)$ matrix

$$
D = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & \cdots & k \end{array} \\ \left[\begin{array}{ccccc} d & 0 & 0 & \cdots & 0 \\ -d & 0 & 0 & \cdots & 0 \\ 0 & d & 0 & \cdots & 0 \\ 0 & -d & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & d \\ 0 & 0 & 0 & \cdots & -d \end{array}\right] \end{array} \begin{array}{l} \left.\begin{array}{l} \\ \\ \end{array}\right\} 1^{st} \text{ pair of mates} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} 2^{nd} \text{ pair of mates} \\ \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} k^{th} \text{ pair of mates} \end{array}
$$

where the k pairs of mates are mutually orthogonal and $2d$ is the distance between the mates of each pair. For $k = 2$, the four points of a regular duplex are at the

vertices of a square, and for $k = 3$, the six points of a regular duplex are at the mid-points of the faces of a cube. The experience of the experimenter is used in choosing the size of the initial duplex design. However, the procedure is designed so that the behaviour of the sequence of designs does not depend critically on the initial design size.

Let P_{hu}, $P_{\ell u}$ $(u = 1,2,\ldots,k)$ represent the k pairs of mates such that $y_{hu} > y_{\ell u}$ $(u = 1,2,\ldots,k)$. The next duplex design is formed by replacing the k points $P_{\ell u}$ $(u = 1,2,\ldots,k)$ by k new points, which must all lie on the other side of the (k-1) dimensional hyperplane containing P_{hu} $(u = 1,2,\ldots,k)$. This new duplex design will not in general be regular.

From the following rules k new points can be chosen which, together with P_{hu} $(u = 1,2,\ldots,k)$ conform to the requirements for a duplex design. Let

$$y_{max} = \underset{u}{\text{Max}}(y_{hu})$$

and

$$y_{min} = \underset{u}{\text{Min}}(y_{\ell u}) \quad .$$

Then, W_h and W_ℓ can be formed by the relations

$$W_h = \frac{\sum_{u=1}^{k} (y_{hu} - y_{min})\, P_{hu}}{\sum_{u=1}^{k} (y_{hu} - y_{min})}$$

and

$$W_\ell = \frac{\sum_{u=1}^{k} (y_{max} - y_{\ell u})\, P_{\ell u}}{\sum_{u=1}^{k} (y_{max} - y_{\ell u})} \quad .$$

The point W_h is a weighted average biased towards the P_{hu} with the higher responses, while the point W_ℓ is a weighted average biased towards the $P_{\ell u}$ with the lower responses. The line joining W_ℓ to W_h thus gives a good indication of the direction of ascent.

Now, form the point C by the relation

$$C = (3W_h - W)/2 \quad , \qquad \text{Figure 6a.}$$

The point C will be the centre of the new design. Reflect P_{hu} about C and obtain P_{ru} by the relation

$$P_{ru} = (1+\delta_u)C - \delta_u P_{hu} \qquad (u = 1,2,\ldots,k) \quad ,$$

where

$$\delta_u = \frac{\sum_{v=1}^{k} \overline{P_{hv}C}}{k\,\overline{P_{hu}C}} = \frac{\nu}{\overline{P_{hu}C}}$$

and $\overline{AB}$ means the distance from A to B. Thus $\delta_u > 0$ and $\overline{CP}_{ru} = \nu$, where ν is constant for all u. We define δ_u in this way to prevent the duplex design from becoming degenerate after a few steps of the procedure. The new points are chosen as follows:

(a) If $y_{ru} > y_{hu}$, expand P_{ru} to P_{eu} by the relation

$$P_{eu} = (3P_{ru} - C)/2 , \quad (u = 1,2,\ldots,k) .$$

(i) If $y_{eu} > y_{hu}$, then take P_{eu} to be the new point (Figure 6b).

(ii) If $y_{eu} < y_{hu}$, then take P_{ru} to be the new point (Figure 6b).

(b) If $y_{ru} < y_{hu}$, contract P_{ru} to P_{cu} by the relation

$$P_{cu} = (C + P_{ru})/2 , \quad (u = 1,2,\ldots,k) .$$

(i) If $y_{cu} > y_{hu}$, then take P_{cu} to be the new point (Figure 6c).

(ii) If $y_{cu} < y_{hu}$, then take P_{ru} to be the new point (Figure 6c).

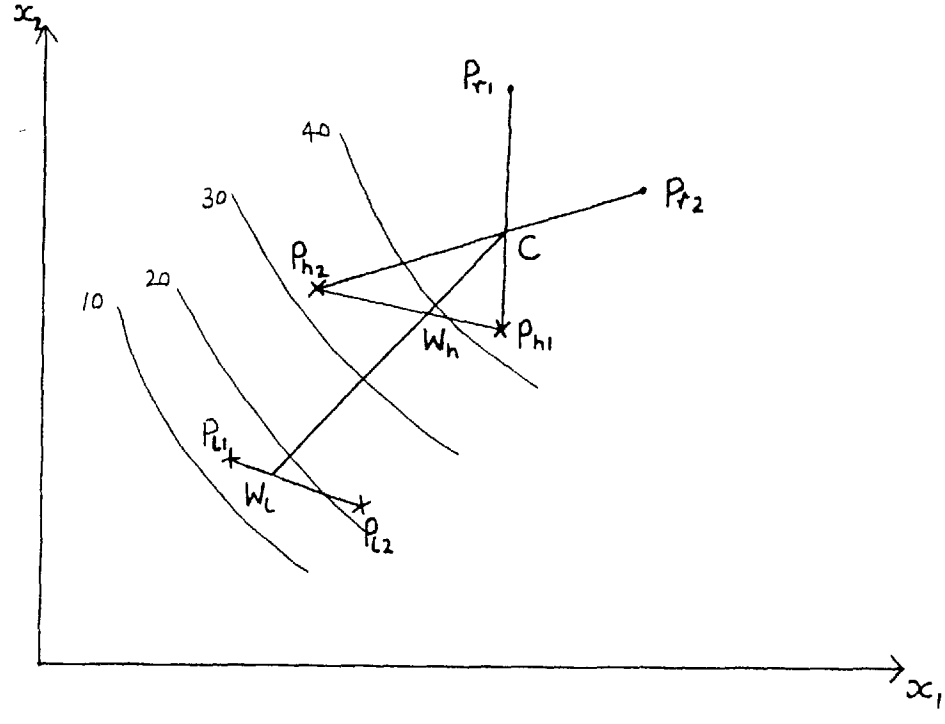
The following additional rules are necessary for the successful operation of the procedure.

Rule A:

If the point P_{max} in the current duplex design has occurred in (k+1) successive duplex designs and is not then replaced, make a repeat observation at that point. This guards against the sequence of designs becoming anchored at a false optimum.

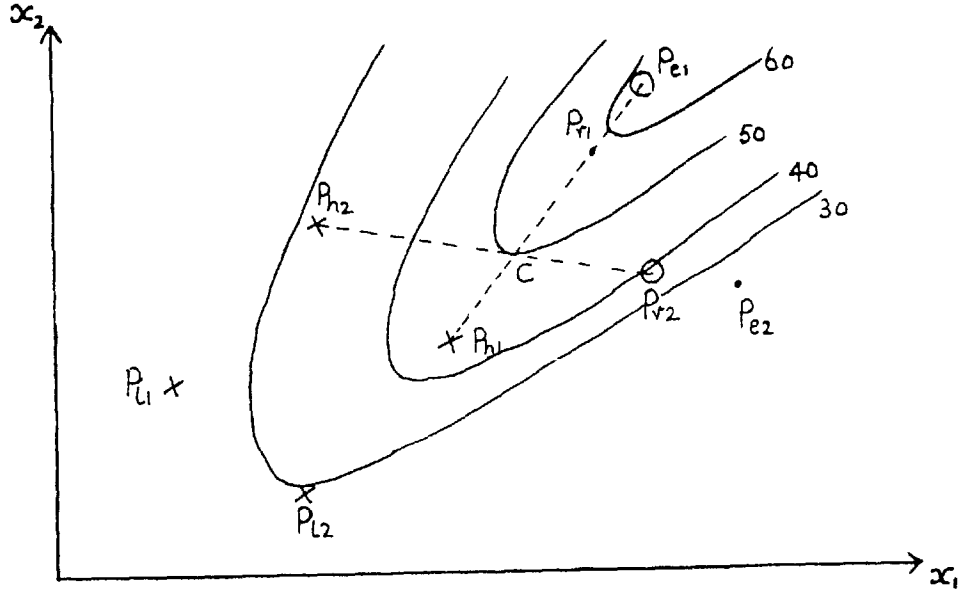
FIGURE 6. THE DUPLEX PROCEDURE. (K=2).

x denotes a point in the current design.
• denotes a point at which a subsequent observation is made.
⊙ denotes a new point for the next design
------ joins pairs of mates in the new design.

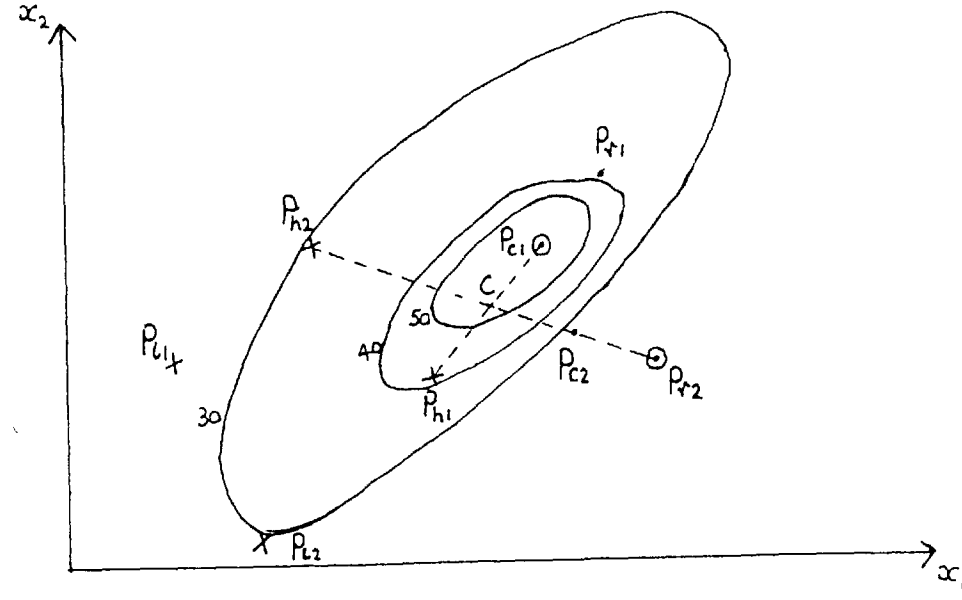


(a) The construction of P_{r1} and P_{r2} from the current design

FIGURE 6 (cont'd). THE DUPLEX PROCEDURE. (K=2).



(b) $y_{r1} > y_{h1}$, $y_{e1} > y_{h1}$ and $y_{r2} > y_{h2}$, $y_{e2} < y_{h2}$.



(c) $y_{r1} < y_{h1}$, $y_{c1} > y_{h1}$ and $y_{r2} < y_{h2}$, $y_{c2} < y_{h2}$.

Rule B (Stopping rule):

If the point at which the repeat observation was taken occurs in the next k designs, including the one containing the repeat observation, and is not then replaced, stop the procedure, as this is evidence that the optimum region has been found. We may estimate the OFC by the point corresponding to the highest observed response in the final design. Alternatively, we may take as an estimate of the OFC a weighted average, W, of the points in the final design given by

$$W = \frac{\sum_{u=1}^{2k} \left\{ y_u - (y_{min} - R) \right\} P_u}{\sum_{u=1}^{2k} \left\{ y_u - (y_{min} - R) \right\}} ,$$

where $R = y_{max} - y_{min}$. We shall refer to this as the weighted observations estimate.

Rule C:

At any stage of the procedure if the responses at <u>all</u> the new points in the new design are lower than the responses of their respective mates in the new design (which can occur only if (b)(ii) above is satisfied for all u ,

and which is an indication that the sequence of designs has reached the vicinity of a NSR) , then measure the spreads, S_t^2 , $t = 1,2,\ldots,k$, of the current design. If any S_t^2 of the current design is greater than α (chosen by the experimenter) times the corresponding S_t^2 of the initial regular duplex design, then the procedure is restarted by making observations at the design points of a new duplex design. This design should have the same size and orientation as the initial duplex design and should be centred at P_{max} in the current design, thus preserving the automatic property of the procedure. However, the size of this new duplex design may be chosen to be consistent with any gain in information by the experimenter since the start of the procedure. This rule helps to accelerate the search if the design size becomes too large. In choosing α the experimenter may be guided by the fact that, if α is too small the procedure will be restarted continually and hence the action of the stopping rule will be delayed; also, if α is too large then the procedure may not be restarted when it should be and further design points will be chosen which contribute little information about the NSR. Either case may result in an unnecessarily large number of observations being made.

Rule D:

If, however, all the S_t^2 of the current design are less than or equal to the corresponding S_t^2 of the initial design then C_l and C_h are formed by the relations

$$C_h = (\sum_{u \neq s} P_{hu} + P_{ls})/k \quad ,$$

where $y_{hs} = \underset{u}{\text{Min}}(y_{hu})$, and

$$C_l = (\sum_{u \neq s} P_{lu} + P_{hs})/k \quad .$$

The line joining C_l to C_h is thus an indication of the direction in which the next design is formed. This rule prevents the sequence of designs from oscillating over a limited region and allows continual exploration.

Rule E:

If for a given u, P_{ru} falls outside the experimental region then choose P_{ru} to be the new point and allocate to P_{ru} the response y_{ru} , where $y_{ru} < y_{hu}$. If for all u, $y_{ru} < y_{hu}$, and at least one P_{ru} falls outside the experimental region, then the procedure is continued by applying rule D. Otherwise, continue the procedure in the normal manner.

If P_{eu} falls outside the experimental region then

P_{ru} is chosen as the new point and the procedure is continued. We assume that the boundary of the experimental region is such that there is no question of P_c falling outside.

We shall refer to this duplex procedure as Routine 6. It is automatic, scale invariant and the sequence of designs is adapted to changes in the response surface. The rules are such that the design is not allowed to get very small. The new feature of this procedure is that a search for regions of higher response is made in several directions at one step of the procedure.

For Routines 5 and 6, on reaching an optimum region, the sequence of designs will tend to 'circle' around it. Thus, in general, at the end of the sequence of designs there will be a number of design points, scattered about the optimum region. These points arise in the natural operation of the procedure whose primary aim is to produce a sequence of designs which will converge to an optimum region. The information obtained in the responses at these points can then be used to fit a polynomial and thus to obtain a more comprehensive estimate of the optimum conditions. Therefore, in general, no experimentation additional to that

involved in the natural running of the procedure is necessary for the final stage of the search for optimum conditions. A more orthodox procedure for the final stage of the search is to use a conventional response surface design in the region of the estimate of the OFC. For this, further experimentation will be necessary.

3.2.3 Results and Discussion.

The difficulties experienced in obtaining analytical results for the single factor case, are amplified in the multi-factor case; consequently, no analytical results have been obtained. The quantitative results presented in this section arise from the simulation of Routine 5 and Routine 6.

First, we compare the basic behaviour of Routines 5 and 6 for a few simple cases in the absence of experimental error ($\sigma = 0$) . Then, we investigate the performance of the procedures on a linear gradient as σ varies. Finally, Routines 5 and 6 are compared on unimodal surfaces as σ varies.

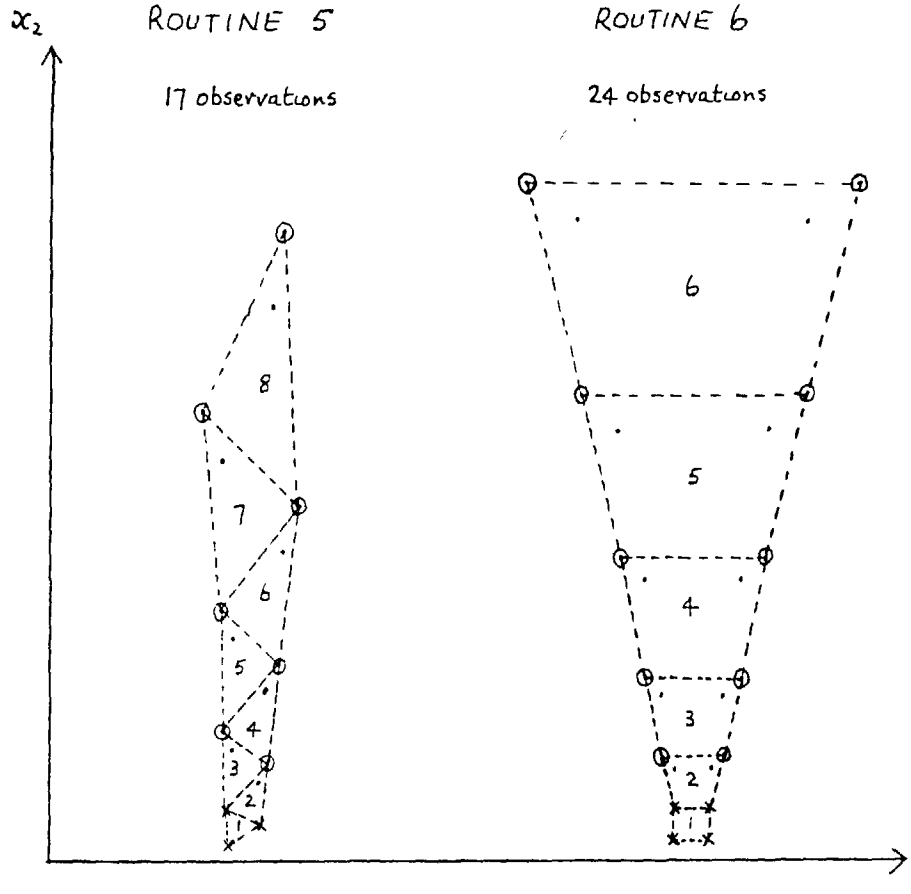
We assume that $k = 2$, $\sigma = 0$ and that the

response surface has the form of a linear gradient. The contours of such a response function and the realization of the sequences of designs for Routines 5 and 6, for a given orientation of the initial designs, are given in Figure 7a. The initial designs are of the same size (see Section 3.2). The orientation of the initial designs may have a slight effect on the direction in which the ensuing sequence climbs up the slope, but not on the characteristic pattern of the sequence of designs.

It is evident, from Figure 7a, that Routine 5 generates a sequence of simplex designs, for which the search for higher response is made in one direction for one design and in another direction for the following design. In contrast, for Routine 6 the search for higher response is made in two directions simultaneously. (For $k \geq 2$, the search is made in k directions simultaneously for Routine 6 but still only in one direction at a time for Routine 5.). Routine 6 gives rise to a more balanced search and the size of the duplex designs expands more rapidly as the sequences of designs progress, than for Routine 5. Because the designs expand as the sequences progress, the gain in response per observation increases as the number of observations made increases.

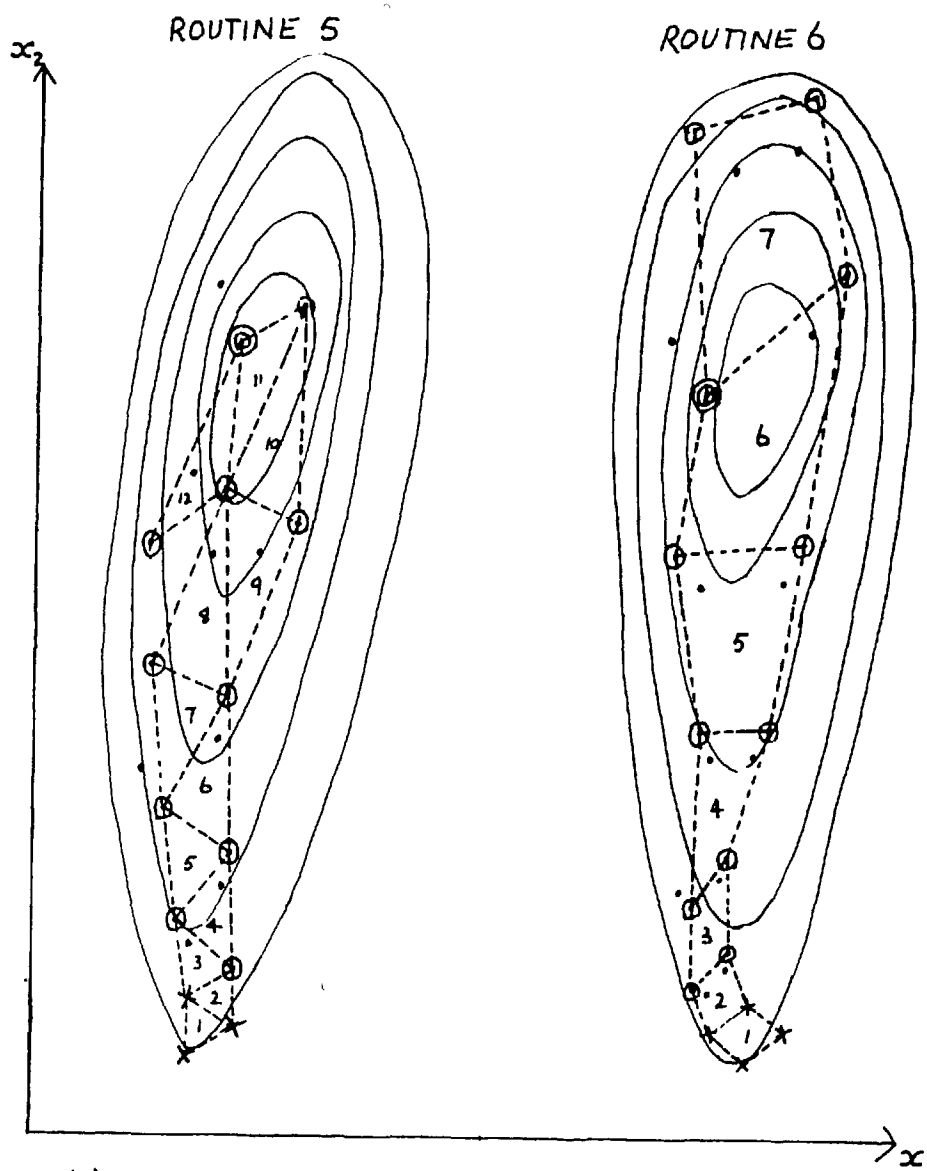
FIGURE 7 THE PATTERN OF DESIGNS FOR ROUTINES 5 AND 6. (K=2).

x denotes a point in the initial design
• denotes a point at which a subsequent observation is made.
⊙ denotes a new point for a new design.
----- joins the design points for each design



(a) An illustration of the characteristic pattern of the sequences of designs on a linear gradient

FIGURE 7. (cont'd). THE PATTERN OF DESIGNS FOR ROUTINES 5 AND 6. (K=2).



(b) An illustration of the changes in shape of the designs corresponding to changes in the response function.

◎ denotes the point with the highest observed response, which is also a design point

For the particular realizations in Figure 7a, after 8 designs and 17 observations, Routine 5 achieves a gain in response approximately equal to the gain in response for Routine 6, after 6 designs and 24 observations.

We now assume that $k = 2$, $\sigma = 0$ and that the response functions are represented by the contours in Figure 7b. The sequences of designs follow the ridge to the optimum region, more closely for Routine 5 than for Routine 6. The greater scatter of points for Routine 6 indicates that some effort is spent in searching for other possible regions of higher response.

We shall now consider the relative performance of Routines 5 and 6 on the linear gradient

$$\eta = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

in the presence of error. The behaviour of the sequence of designs on this linear gradient will indicate their relative abilities to climb more general response surfaces, over which the response is increasing reasonably uniformly. Later, we shall consider the ability of a sequence of designs to converge to the optimum of a response surface.

Routines 5 and 6 are compared under the following conditions:

(i) The values of k are 2, 3, 4 and 5.

(ii) For each k, the sets β_i (i = 1,2,...,k) were chosen arbitrarily, but subject to the constraint

$$\sum \beta_i^2 = \beta^2 = 1 \quad ,$$

and are

k	β_1	β_2	β_3	β_4	β_5
2	0.5000	0.8660			
3	0.2774	0.4804	0.8321		
4	0.1581	0.2739	0.4743	0.8216	
5	0.0909	0.1575	0.2727	0.4724	0.8182 .

(iii) The initial designs for the two procedures are chosen to be of the same size (see Section 3.2). It can be shown that by this criterion the initial design points for the procedures all lie on a k dimensional sphere of radius r. The initial designs are regular and are centred at the centre of this sphere, which is at the origin.

(iv) For each k , r is equal to 0.1. Since β = 1, the maximum difference in true response between two points on the sphere is 0.2. These two points lie at the extremities of the diameter in the direction of steepest ascent.

(v) The values of σ , chosen relative to this maximum difference, are 0.001, 0.025, 0.050, 0.075, 0.100, 0.200 and 0.250.

(vi) The orientations of the initial designs are as presented in Sections 3.2.1 and 3.2.2. It is assumed that the orientation of the initial design has little effect on the 'long-run' behaviour of the sequence of designs.

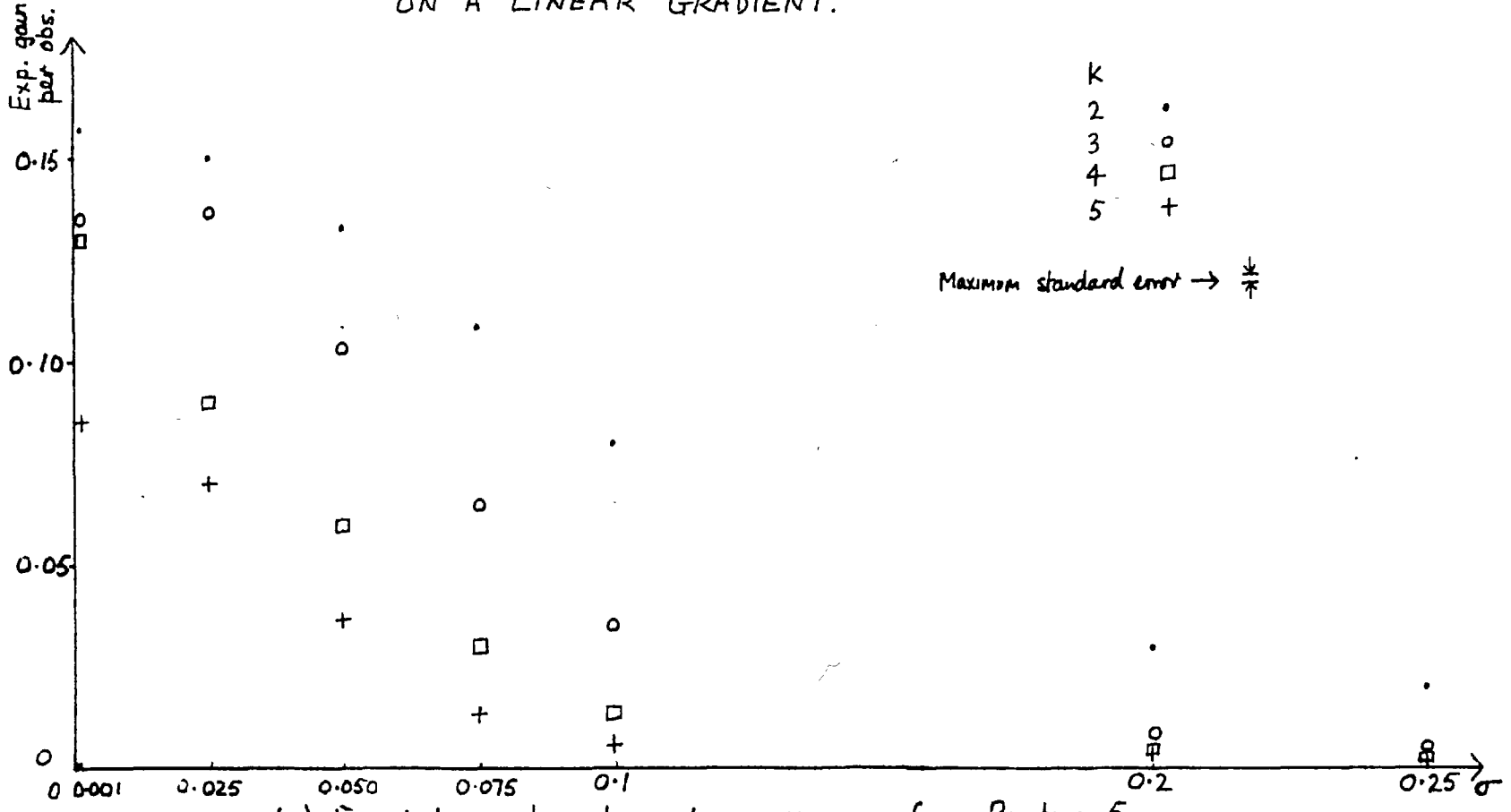
(vii) For each k and σ Routine 6 is run until 5 designs are completed; thus 10k observations are made. For k = 2,3,4 and 5, Routine 5 is run until 9, 14, 18 and 23 designs, respectively, are completed. These numbers were thought to be such that the number of observations made would be approximately the same as for Routine 6, at least for small σ . However, as σ increases, for a given k, and as k increases, the number of observations

observed from the simulation were less than expected. The discrepancy was as much as 20% of the number expected, for $k = 5$ and $\sigma = 0.25$. This is due to one of the rules for choosing a new design point in Routine 5. This rule is illustrated in Figure 5c and requires only one observation, instead of two, to be made before choosing a new design. The choice of the new point by this rule occurred more frequently than was expected.

A simulation study of 500 trials is done for each procedure and the expected gain in response was estimated. The expected gain in response per observation is plotted against σ in Figures 8a and 8b, for Routines 5 and 6 respectively. The expected gain per observation, for large k and σ, for Routine 5 in Figure 8a, should be inflated slightly before comparing Routine 5 with Routine 6, but it is clear that the following general pattern exists.

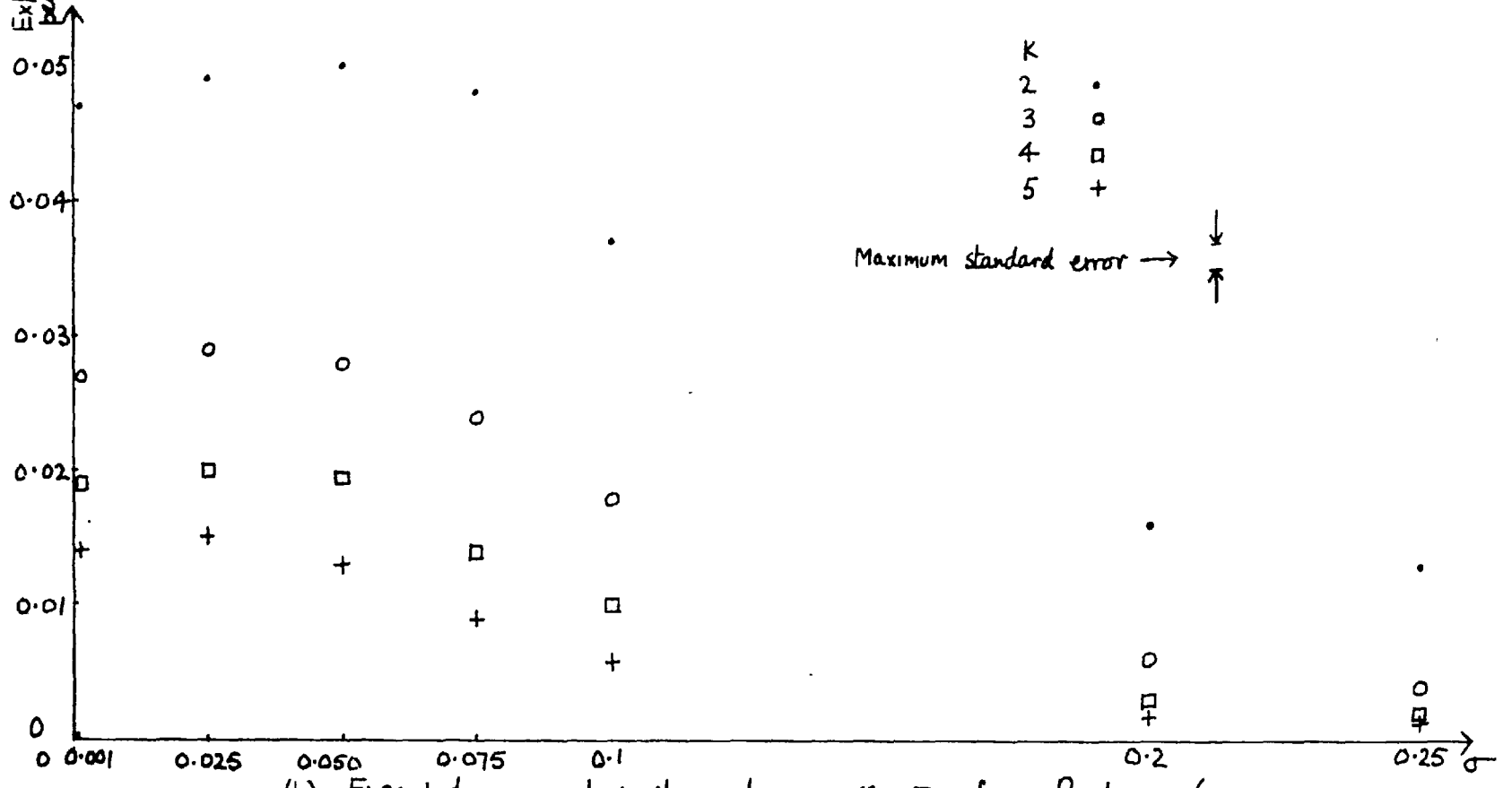
For the higher values of k and σ there is not much difference in performance between the two procedures, but otherwise Routine 5 is superior. The expected gain per

FIGURE 8. THE PERFORMANCES OF ROUTINES 5 AND 6 ON A LINEAR GRADIENT.



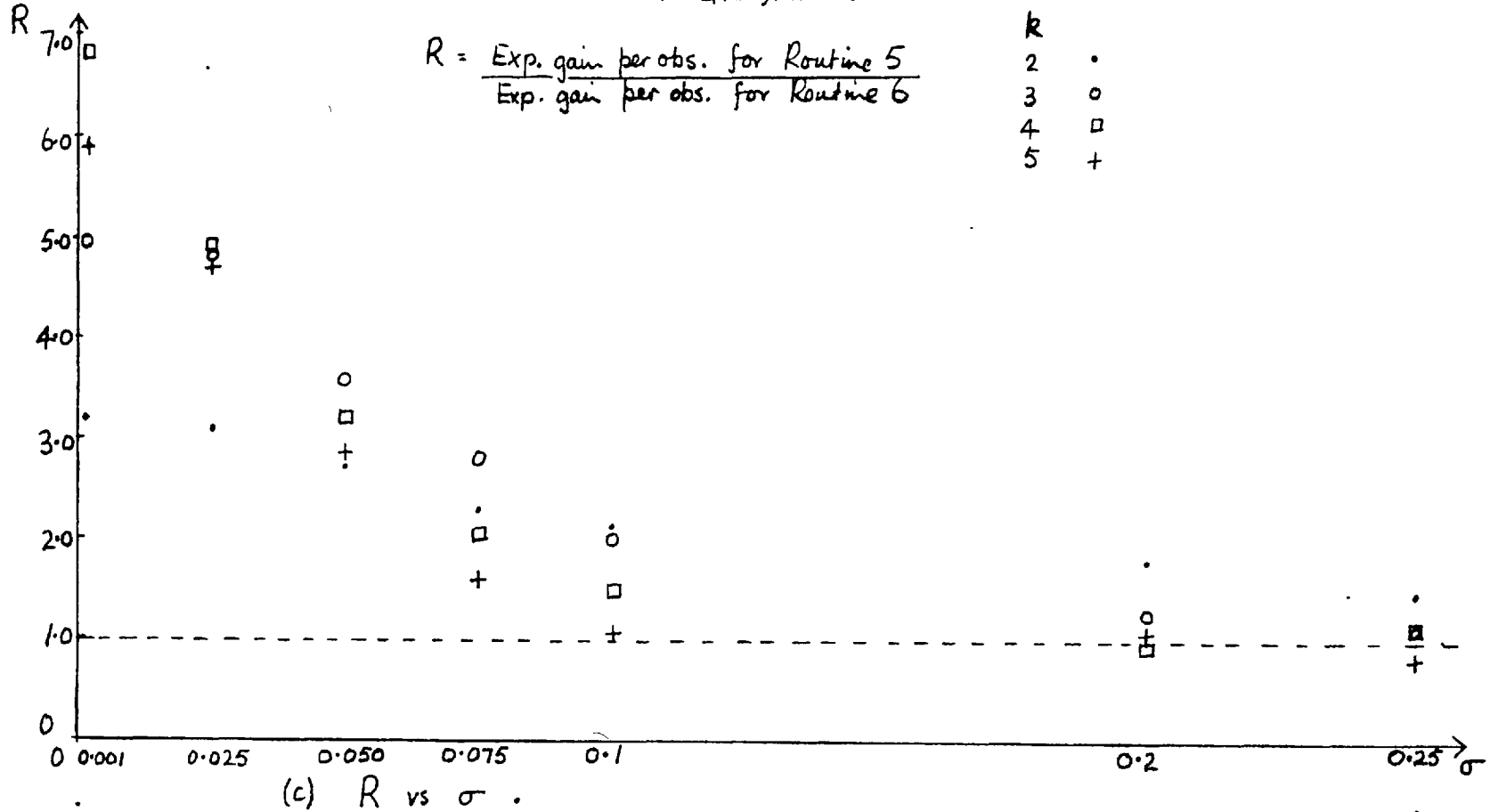
(a) Expected gain per observation vs σ for Routine 5.

FIGURE 8 (cont'd) THE PERFORMANCES OF ROUTINES 5 AND 6 ON A LINEAR GRADIENT



(b) Expected gain per observation vs σ for Routine 6.

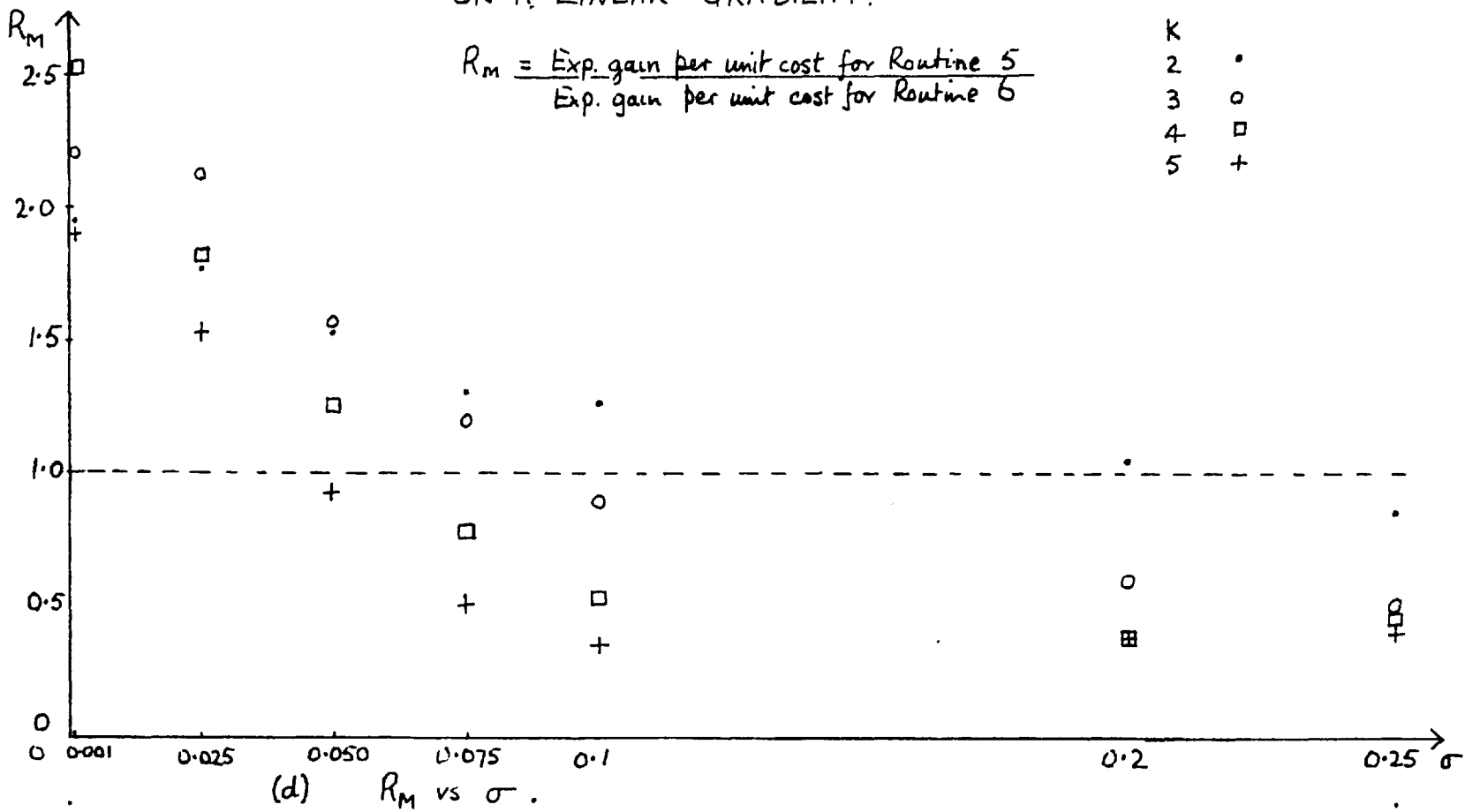
FIGURE 8 (cont'd) THE PERFORMANCES OF ROUTINES 5 AND 6 ON A LINEAR GRADIENT.



(c) R vs σ .

FIGURE 8 (cont'd) THE PERFORMANCES OF ROUTINES 5 AND 6 ON A LINEAR GRADIENT.

R_M = Exp. gain per unit cost for Routine 5 / Exp. gain per unit cost for Routine 6



(d) R_M vs σ.

observation decreases at the rate of $1/\sigma$, approximately, for both procedures. This relationship does not hold for small σ , especially for Routine 6. Under the given conditions the expected gain per observation decreases approximately at the rate 1/k.

In Figure 8c the ratio of the expected gain per observation for Routine 5 to the expected gain per observation for Routine 6, is plotted against σ . This ratio tends to fall off more rapidly, as σ increases, the larger k becomes. This indicates that as k increases in the range of k considered the superiority of Routine 5 diminishes rapidly, as σ increases, settling around unity for the larger σ .

Over the range of k and σ considered above, the results, based on the criterion of expected gain per observation, indicate that Routine 5 is generally to be preferred to Routine 6 for a response surface which approximates to a linear gradient. However, the comparison of these procedures, judged by some other criterion, may be of interest.

Routines 5 and 6 are automatic procedures which both have the same sequential pattern, i.e. design points are chosen (stage 1), responses are observed (stage 2) and

then stage 1 is repeated and so on. For Routine 5, after the initial design, only one point is chosen at stage 1 and one response observed at stage 2 for each k. For Routine 6, however, after the initial design, k points are chosen at stage 1 and the k corresponding responses are observed at stage 2.

Now, suppose there is a cost associated with each combination of stage 1 and stage 2 in addition to the cost per observation, and this cost is M($>$ 0) times the cost per observation. Then, it may be more economical to use a procedure which chooses several points than to use a procedure which chooses one point, at each stage 1, because fewer stages will be required for the same total number of observations. Under these circumstances, Routine 6 is to be preferred to Routine 5 for large k and σ^{-}, if the comparison is based on the expected gain per unit cost. As M increases Routine 6 is better even for smaller k and σ^{-}. In Figure 8d the ratio of the expected gain per unit cost for Routine 5 to the expected gain per unit cost for Routine 6 is given for M = 6.

As an example of a situation in which Routine 6 may be preferable to Routine 5, consider an experimenter who has easy access to a telephone link with a time sharing computer.

He is exploring a response surface in search of optimum conditions. His technique is to observe the responses from an initial design and then to telephone the computer to feed in this information, whereupon the computer gives him the next set of points at which observations are to be made. This process is then repeated until a stopping rule comes into action. Now, every time contact is made with the computer a basic cost is incurred which is distinct from the cost of processing the responses which are fed in. Because of this basic cost and also because of any difficulty in setting up his experiment, it will generally, expecially if k is large, be more economical to make k observations instead of one observation at each stage.

We now go on to consider the performances of Routine 5 and Routine 6 over unimodal surfaces. Here we can observe the effect of the supplementary rules for Routines 5 and 6 which are given in Sections 3.2.1 and 3.2.2 respectively. Particularly we observe the effects of Rules B(Stopping Rule) and C. Rule E which deals with constraints on the experimental region is not included in the simulation.

To test the ability of the sequences of designs to converge to the optimum region, we consider the following

unimodal response surface::

$$\eta = 1 - \sum_{i=1}^{k} (1 - x_i)^2$$

which has an optimum true response of unity at $x_i = 1$ $(i=1,2,\ldots,k)$. In addition, for $k = 2$, the procedures are tested on a more complicated surface, a sharp rising ridge (Brooks, 1959), which is given by

$$\eta = x_1^2 \exp\left\{1 - x_1^2 - 20.25(x_1 - x_2)^2\right\}$$

and which has an optimum true response of unity at $x_1 = x_2 = 1$. No attempt has yet been made to test the procedures on more complicated surfaces.

Routines 5 and 6 are compared under the following conditions:

(i) The values of k are 2,3,4 and 5.

(ii) For the first surface and for each k, a starting point is chosen such that the true response at that point is 0.5. The gradient, β , at that point is $\sqrt{2}$ for all k. The starting values $x_S = x_i$ $(i = 1,2,\ldots,k)$ are

k	2	3	4	5
x_S	0.5000	0.5918	0.6465	0.6838

For the second surface two starting values are used. The one (0.5, 0.5) is along the ridge and the other (1.00, 0.75) is nearer to the OFC but is away from the ridge and the gradient there is greater. The true response at both points is approximately 0.5.

(iii) The initial designs for the two procedures are chosen to be of the same size (see Section 3.2). The initial spreads $S^2 = S_t^2$ (t = 1,2,...k) are

k	2	3	4	5
S^2	0.0050	0.0033	0.0025	0.0020

.

The initial design points for the two procedures all lie on a k dimensional sphere of radius r. The initial designs are regular and are centred at the centre of this sphere, which is at x_s .

(iv) For each k, r is equal to 0.1. Since $\beta = \sqrt{2}$, the maximum difference in true response between two points on the sphere is 0.2828. These two points lie at the extremeties of the diameter in the direction of steepest ascent.

(v) The values of σ chosen relative to this maximum difference are

0.050 , 0.125 and 0.200 .

(vi) The orientations of the initial designs are as presented in Sections 3.2.1 and 3.2.2. The orientation of the initial design has little or no effect on the performance of the procedures, as the shape of the design is altered to be consistent with changes in the surface.

(vii) For each k and σ Routines 5 and 6 are run until the stopping rule acts. However, the effectiveness of the stopping rule depends to some extent on the choice of α (Rule C, Sections 3.2.1 and 3.2.2) . If α is too small the action of the stopping rule is delayed and the procedure may even continue indefinitely. In this case, if the above stopping rule fails to act after a reasonable number of observations, the procedure is stopped.

A simulation study of 500 trials was made for Routines 5 and 6 and the OFC is estimated by the WOE and the HOE.

The criterion of effectiveness is the true response at the estimate of the OFC. In general, the value of α is increased until the number of 'failures' in the 500 trials, i.e. the number of times the stopping rule fails to act before 200 observations have been expended, is less than 10% of the number of trials. This criterion for choosing is relaxed for the larger values of k because of the large amount of computer time involved. The means and standard errors of the true responses at the estimates of the OFC and of the number of observations expended are calculated only from those trials for which the stopping rule acted.

The results of the simulations are given in Table 5 and are discussed below. In general the true response at the WOE differs little from the true response at the HOE. Only the results for the true response at the HOE are presented in Table 5.

Routines 5 and 6 for $\sigma = 0.05$ are in general successful in achieving the gain in response necessary to attain optimum conditions. For larger σ this gain diminishes and possibly further modifications to the procedures may improve this gain in response. We note that although the gain in response is less as σ increases the number of

observations does not have a similar trend.

Fewer observations are required for Routine 5 than for Routine 6. However, the designs for Routine 6 expand more than those for Routine 5 and in general a larger value of α will be required for Routine 6 to achieve the same small proportion of 'failures' as for Routine 5.

If the simplex designs in Routine 5 get badly positioned on a ridge the sequence of designs may get stuck. This happens on the sharp rising ridge for the starting value (0.5, 0.5), Table 5. Routine 6, however, because of its greater power to search is unlikely to get stuck. This indeed suggests that for more complicated surfaces than we have considered and for larger k, Routine 6 may have some real advantages over Routine 5.

These procedures, then, illustrate the ideas which are useful in designing an automatic sequential procedure and the limited results presented, indicate their potential use.

TABLE 5

COMPARISON OF ROUTINE 5 AND ROUTINE 6

K=2. RESPONSE SURFACE WITH RISING RIDGE.

STARTING POINT =0.5000, 0.5000.

	ROUTINE 5(α = 10)					ROUTINE 6(α = 15)				
EXPTL. ERROR S.DEV.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.
0		13		0.6201			25		0.9519	
0.050	0	15.6	0.3	0.6504	0.01	5	33.3	0.4	0.8941	0.01
0.125	0	18.6	0.4	0.6295	0.01	7	32.9	0.7	0.7600	0.01
0.200	2	21.4	0.6	0.5797	0.01	22	35.5	0.8	0.6571	0.01

K=2. RESPONSE SURFACE WITH RISING RIDGE.

STARTING POINT =1.0000, 0.7500.

	ROUTINE 5(α = 6)					ROUTINE 6(α = 15)				
EXPTL. ERROR S.DEV.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.
0		26		0.9559			29		0.9111	
0.050	0	20.0	0.3	0.8985	0.00	0	26.2	0.2	0.8719	0.00
0.125	2	18.9	0.3	0.8600	0.00	1	25.9	0.4	0.8360	0.01
0.200	2	20.3	0.5	0.8217	0.01	6	27.2	0.5	0.8136	0.01

TABLE 5 (CONT'D)

COMPARISON OF ROUTINE 5 AND ROUTINE 6

K=2. RESPONSE SURFACE WITH SPHERICAL CONTOURS.

STARTING POINT =0.5000, 0.5000.

	ROUTINE 5(α = 10)					ROUTINE 6(α = 12)				
EXPTL. ERROR S.DEV.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.
0		21		0.9945			29		0.9416	
0.050	2	24.2	0.5	0.9617	0.00	7	36.0	0.6	0.9541	0.00
0.125	7	30.7	0.7	0.8886	0.01	26	43.1	0.9	0.8652	0.01
0.200	7	31.1	0.8	0.7825	0.01	41	42.1	0.9	0.7708	0.01

K=3. RESPONSE SURFACE WITH SPHERICAL CONTOURS.

STARTING POINT =0.5918, 0.5918, 0.5918.

	ROUTINE 5(α = 8)					ROUTINE 6(α = 10)				
EXPTL. ERROR S.DEV.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT MEAN	HOE S.ERR.
0		27		0.9808			79		0.9914	
0.050	60	41.4	0.9	0.9639	0.00	306	70.4	1.1	0.9592	0.00
0.125	93	41.5	1.1	0.8275	0.00	350	61.3	2.3	0.7527	0.02
0.200	109	40.9	1.1	0.7057	0.01	325	68.5	1.3	0.7592	0.01

TABLE 5 (CONT'D)

COMPARISON OF ROUTINE 5 AND ROUTINE 6

K=4. RESPONSE SURFACE WITH SPHERICAL CONTOURS.

STARTING POINT =0.6465, 0.6465, 0.6465, 0.6465.

	ROUTINE 5(α = 40)					ROUTINE 6(α =100)				
EXPTL. ERROR S.DEV.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT HOE MEAN	S.ERR.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT HOE MEAN	S.ERR.
0		41		0.9920			299		0.9996	
0.050	56	61.9	1.9	0.9593	0.00	291	154.9	4.6	0.8766	0.01
0.125	66	59.1	1.9	0.7744	0.01	305	178.6	4.6	0.8824	0.01
0.200	85	62.1	2.2	0.6577	0.01	295	154.4	4.4	0.7360	0.01

K=5. RESPONSE SURFACE WITH SPHERICAL CONTOURS.

STARTING POINT =0.6838, 0.6838, 0.6838, 0.6838, 0.6838.

	ROUTINE 5(α = 30)					ROUTINE 6(α =150)				
EXPTL. ERROR S.DEV.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT HOE MEAN	S.ERR.	NUM. OF FAILURES	NUM. OF OBS. PER TRIAL MEAN	S.ERR.	TRUE RESP. AT HOE MEAN	S.ERR.
0		41		0.9793			261		0.9998	
0.050	156	51.9	1.1	0.9471	0.00	380	182.2	8.4	0.7863	0.02
0.125	174	48.4	1.2	0.7093	0.01	385	176.3	11.9	0.6657	0.03
0.200	199	44.8	1.9	0.5869	0.01	370	221.2	6.6	0.6654	0.01

CHAPTER 4

CONCLUSION

The aim of the thesis is to design initial stage procedures to meet the requirements of Chapter 1. Three new procedures have been developed. Routine 3 for the single factor case in Chapter 2 and Routines 5 and 6 for the multi-factor case in Chapter 3. These procedures possess most of the important features discussed in Section 1.3. They are all automatically operated. They generate sequences of designs which adapt to the local nature of the surface and hence a bad choice of scale of the factors is less likely to retard the convergence of the sequence of designs to the optimum region (this is more important in the multi-factor case). They all have automatically operated stopping rules which are good for small error. Further modifications to the procedures are necessary to improve their performance for large error. No attempt has yet been made to choose a region where a second degree polynomial is a good fit. This could be incorporated in a final stage procedure. No quantitative results have been obtained concerning the rules for the procedures on encountering constraints.

Much further work needs to be done, in particular, on the extension of an automatic initial stage procedure to include an automatic final stage procedure. In the final

stage a second degree polynomial could be fitted to the observations arising from a run of an initial stage procedure. It may be argued that it is more efficient for estimating the regression coefficients, to use a conventional response surface design. But, an advantage of an automatic initial and final stage procedure is that an estimate of the OFC is obtained simultaneously with a set of design points (a 'random' design) to which the surface can be fitted. Also, if the object is not so much to estimate precisely certain individual regression coefficients as to estimate the OFC, the response at the OFC or the difference in the responses at two points, then any slight loss in the precision of these estimates, due to this 'random' design could probably be tolerated if the elimination of further experimentation were a great saving.

Automatic procedures will be most useful in situations where the response is quickly obtained. As an example of this consider a statistical problem where it is not possible to determine the distribution of an estimator of some quantity analytically. It may not even be possible to determine the expected value of this distribution asymptotically. In such a situation one may consider simulating the problem on a

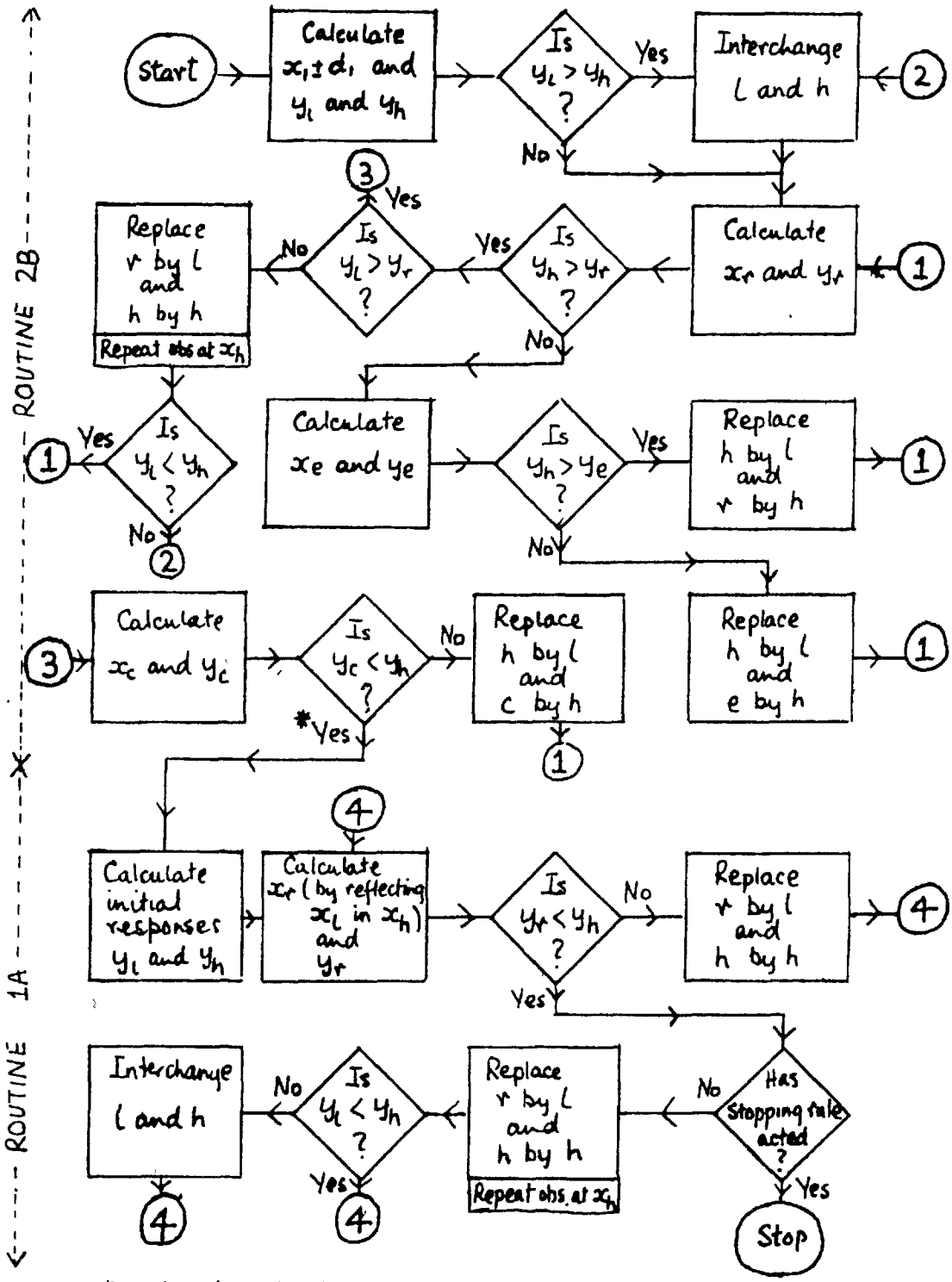
computer, estimating the expected value by taking a large number of samples. This estimate we may regard as the response and it will be subject to error. Now suppose in addition the expected value depends on a number of parameters and one is interested in optimizing this expected value over the parameter space. We may regard these parameters as our factors. This is an ideal case for an automatic sequential procedure.

Another use of an automatic procedure is in evolutionary operation where it is desired to determine continually optimum conditions as time varies. The procedures developed in the thesis have not been designed with this particular application in mind and it is conceivable that slight modifications may be necessary.

APPENDIX

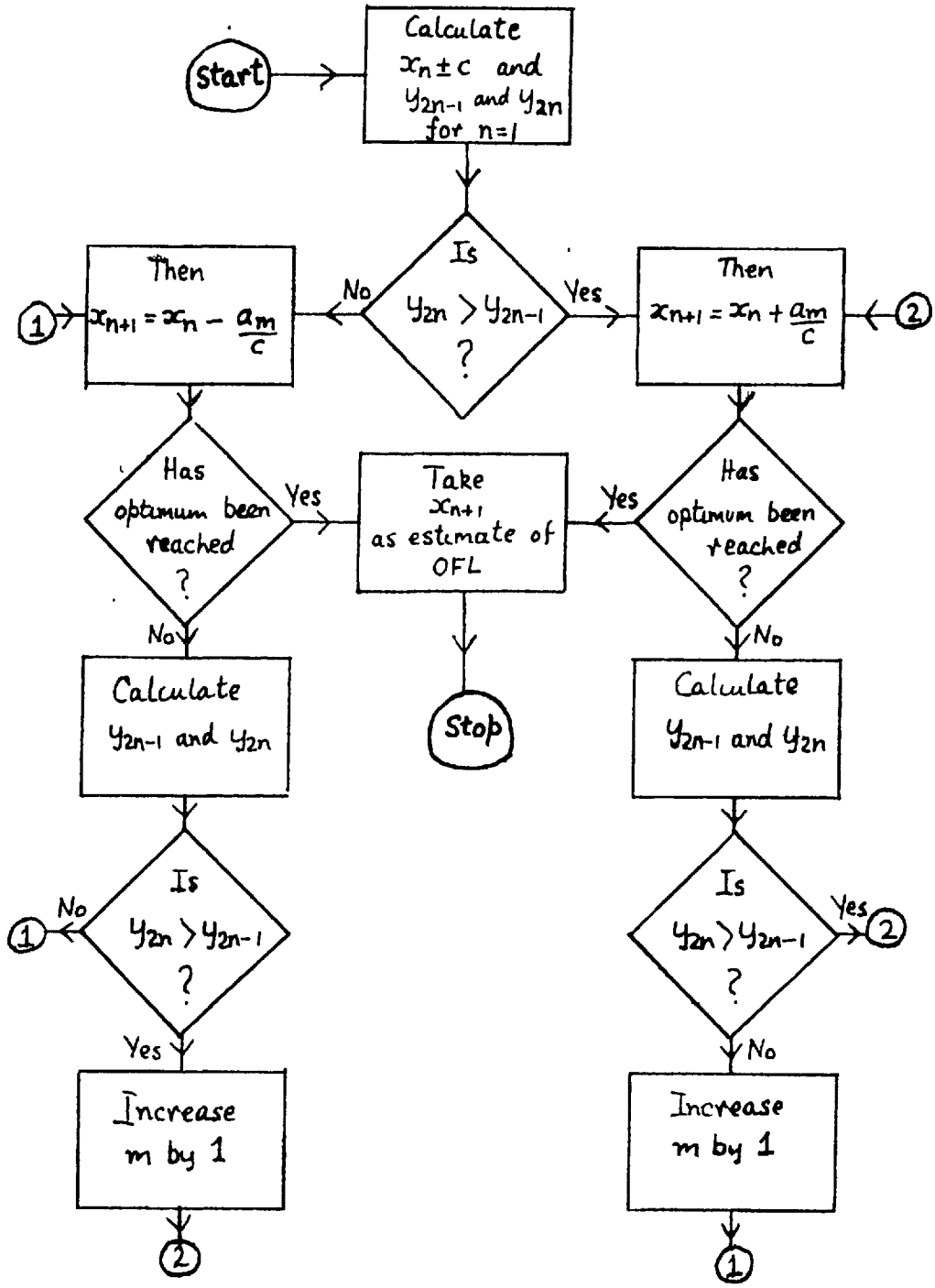
FLOW DIAGRAMS

A FLOW DIAGRAM FOR ROUTINE 3

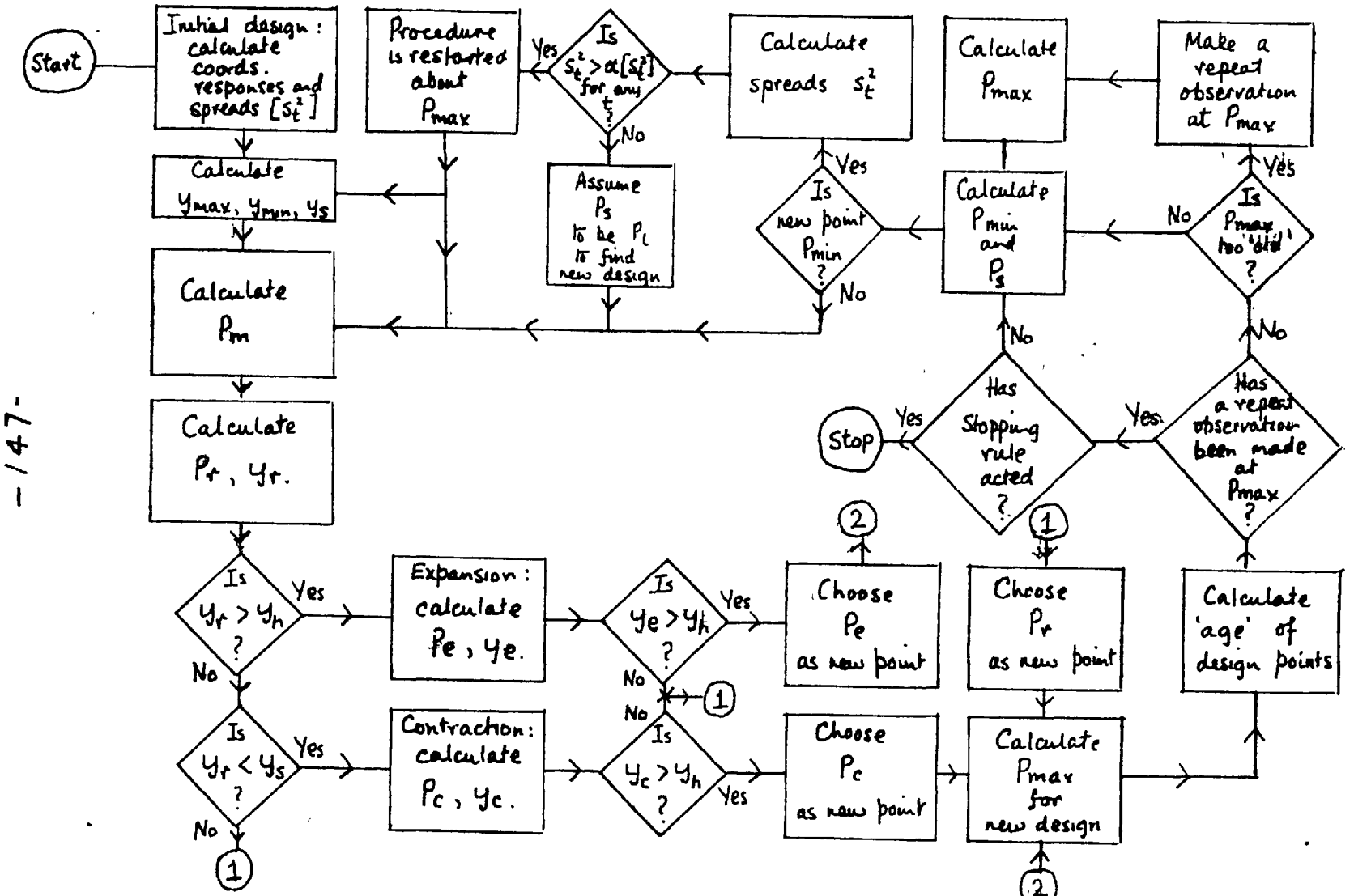


* A failed contraction

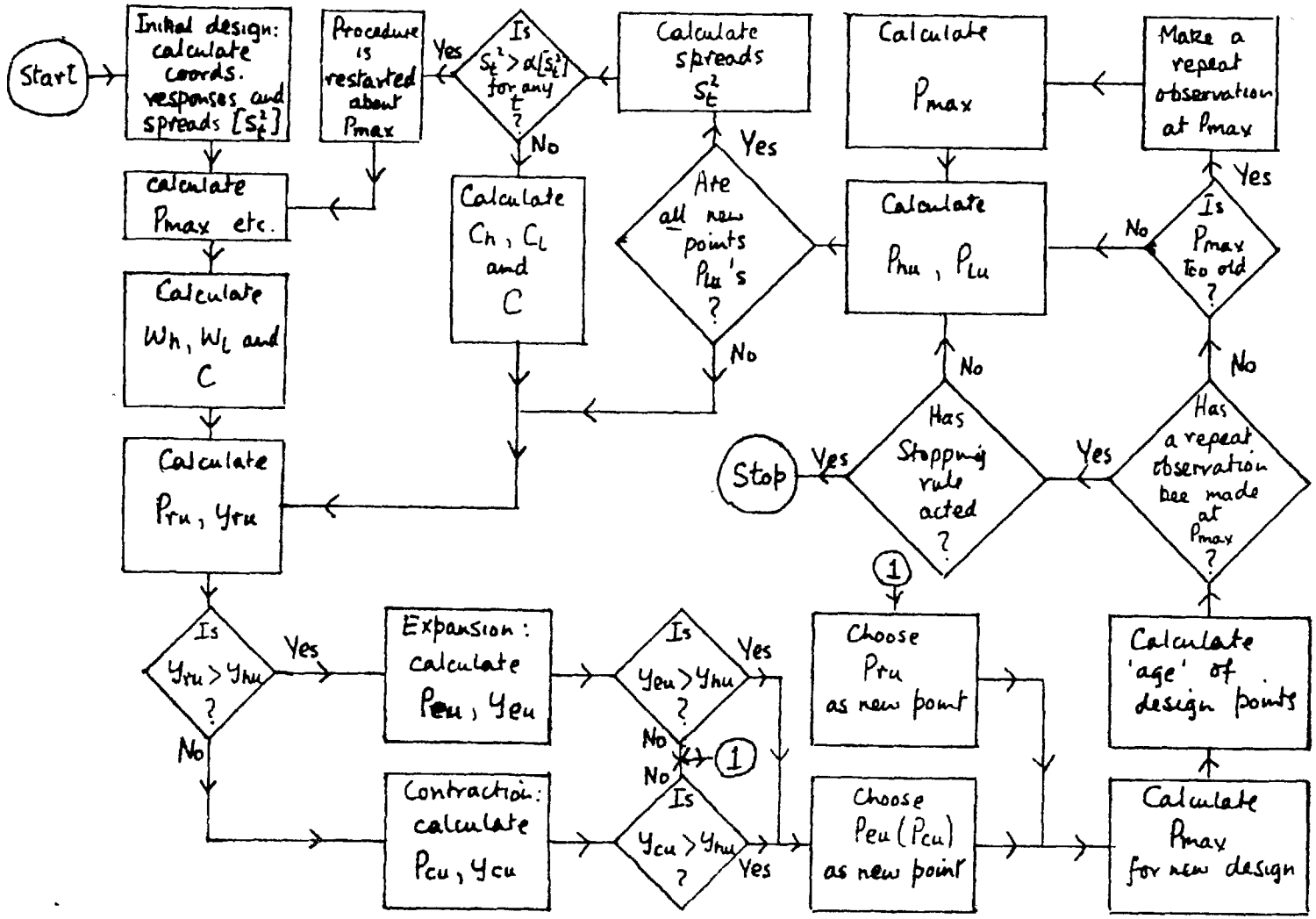
A FLOW DIAGRAM FOR ROUTINE 4



A FLOW DIAGRAM FOR ROUTINE 5



A FLOW DIAGRAM FOR ROUTINE 6



REFERENCES

BLUM, J.R. (1954a). Approximation methods which converge with probability one. Ann. Math. Statist. 25, 382 - 386.

BLUM, J.R. (1954b). Multidimensional stochastic approximation methods. Ann. Math. Statist. 25, 737 - 744.

BOX, G.E.P. (1954). The exploration and exploitation of response surfaces: some general considerations and examples. Biometrics 10, 16 - 60.

BOX, G.E.P. and COX, D.R. (1964). An analysis of transformations (with discussion). J.R. Statist. Soc. B 26, 211 - 252.

BOX, G.E.P. and TIDWELL, P.W. (1962). Transformations of the independent variables. Technometrics 4, 531 - 550.

BOX, G.E.P. and WILSON, K.B. (1951). On the experimental attainment of optimum conditions (with discussion). J.R. Statist. Soc. B 13, 1 - 45.

BOX, M.J. (1966). A comparison of several current optimization methods and the use of transformations in constraint problems. Computer J. 9, 67 - 77.

BROOKS, S.H. (1959). A comparison of optimum seeking methods. Op. Res. 7, 430 - 457.

BURKHOLDER, D.L. (1956). On a class of stochastic approximation processes. Ann. Math. Statist. 27, 1044 - 1059.

DAVIES, O.L. (1963). Design and analysis of industrial experiments. (2nd edition). London: Oliver and Boyd.

DERMAN, C. (1956). Stochastic approximation. Ann. Math. Statist. 27, 879 - 886.

DUNCAN, A.J. (1959). Quality control and industrial statistics. (Revised edition). Illinois: Irwin.

DVORETSKY, A. (1956). On stochastic approximation. Proc. 3rd. Berkeley Symp. Math. Statist. Probab. Vol 1, 39 - 56.

FABIAN, V. (1967). Stochastic approximation of minima with improved asymptotic speed. Ann. Math. Statist. 38, 191 - 200.

FABIAN, V. (1967). Stochastic approximation of minima with improved asymptotic speed, II. Unpublished technical report. Michigan State University.

FRIEDMAN, M. and SAVAGE, L.J. (1947). Selected techniques of statistical analysis. New York: McGraw-Hill.

HILL, W.J. and HUNTER, W.G. (1966). A review of response surface methodology: A literature survey. Technometrics 8, 571 - 590.

HOTELLING, H. (1941). Experimental determination of the maximum of a function. Ann. Math. Statist. 12, 20 - 45.

KENDALL, M.G. and STUART, A. (1963). The advanced theory of statistics. Vol. 1. (2nd Edition). London: Griffin.

KESTEN, H. (1958). Accelerated stochastic approximation. Ann. Math. Statist. 29, 41 - 59.

KIEFER, J. and WOLFOWITZ, J. (1952). Stochastic estimation of the maximum of a regression function. Ann. Math. Statist. 23, 462 - 466.

NATIONAL BUREAU OF STANDARDS (1959). Tables of the bivariate normal distribution function and related functions. Applied Mathematics Series 50. Washington: G.P.O.

NELDER, J.A. and MEAD, R. (1965). A simplex method of function minimization. Computer J. 7, 308 - 313.

PLACKETT, R.L. (1954). A reduction formula for normal multivariate integrals. Biometrika 41, 351 - 360.

SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. Ann. Math. Statist. 29, 373 - 405.

SPENDLEY, W., HEXT, G.R. and HIMSWORTH, F.R. (1962). Sequential application of simplex designs in optimisation and evolutionary operation. Technometrics, 4, 441 - 461.

WILDE, D.J. (1964). Optimum seeking methods. Englewood Cliffs: Prentice Hall.