THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN INVESTIGATION OF THE APPLICATION OF THE METHOD OF STEEPEST ASCENT IN MEDICAL RESEARCH

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

MILLAGE CLINTON MILLER III

Oklahoma City, Oklahoma

1961

AN INVESTIGATION OF THE APPLICATION OF THE METHOD OF

STEEPEST ASCENT IN MEDICAL RESEARCH

APPROVED BY

John C. Briper

Karl V. Mackey

William W. Schottstaedt

Carl E Marshall

Robert D Morrison

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

I am deeply indebted to Dr. Carl E. Marshall for his encouragement and aid in the preparation of this dissertation and to Dr. John C. Brixey, Dr. Kirk T. Mosley, Dr. Wm. W. Schottstaedt, Dr. McWilson Warren and Dr. R. D. Morrison for their counsel and guidance. Of incalculable aid, too, were Dr. P. W. Goaz and the staff of the Dental Section of the Oklahoma Medical Research Foundation, Dr. J. M. Stephenson of the Department of Pathology, University Hospital, and J. P. Costiloe in the collecting and processing of the data presented.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

AN INVESTIGATION OF THE APPLICATION OF THE METHOD OF STEEPEST ASCENT IN MEDICAL RESEARCH

CHAPTER I

INTRODUCTION

The fundamental objective of many experimental investigations is to characterize and measure the relationships between independent and dependent variables. In particular, the medical research worker may be interested in the optimization of a product, process, or response and the investigation of the relationship among these variables near this optimum. The solution of such a problem is dependent upon the identification of those factors which contribute significantly to the product, process, or response in question and on the selection of the combination of the levels of the independent variables which will optimize the response.

Recognizing this fundamental objective and being cognizant of the great number of interacting variables inherent in medical data and research, it was thought that the optimizing technique known as the method of steepest ascent would be of particular value to the individual interested in medical research. This method was recently developed by Box and Wilson (1951), in an attempt to solve partially the latter aspect of the optimization problem for the chemical industry.

Statement of the Problem

The problem is to investigate and determine the feasibility of a new application of a known statistical tool, the method of steepest ascent in response surface techniques.

In applying this technique, indeed any technique, to a new situation, many complications are bound to arise due to the uniqueness of the data and process being investigated, and due to basic assumptions necessary for the utilization of the particular statistic being employed. It may therefore be necessary to make adaptions in the method, and statistical assumptions about the processes being investigated. Also, if the technique is to be of value in solving optimization problems in the area of medical research, one should be aware of the types of processes that may be optimized and of what usefulness the technique might be.

It is proposed to illustrate the efficacy of the method of steepest ascent to optimize a response and the adjustments necessary to utilize the method by investigating two example problems, one from the field of enzymology and the other from the field of clinical pathology.

CHAPTER II

REVIEW OF LITERATURE

To understand better the problems to be encountered in the application of the method of steepest ascent, it is necessary that one have a clear concept of the optimization problem itself. While there are a variety of methods of presenting the problem, a precise statement of the optimization problem might be:

Suppose there exists a response y which is dependent on the levels of k different quantitative factors, or independent variables, which are subject to precise measurement and control. The problem then is to optimize the response y by the proper selection of a particular combination of the factor levels. The combination of factors which produces this optimal response will be called the optimal factor combination. Mathematically this relationship could be expressed as:

$$y = \phi\ (x_1, x_2, \ldots x_k)$$

where y is the true or hypothetical response that should be obtained in the absence of experimental error, ϕ is the response function and x_j (j = 1,2,...k) are the independent variables.

In experimental work, the response function ϕ is usually unknown and must be approximated. We assume that the response y may be

represented by a general linear regression equation of the form:

$$y = b_o + b_1x_1 + b_2x_2 + \ldots + b_{11}x_1^2 + \ldots + b_{12}x_1x_2 + \ldots + b_{111}x_1^3 + \ldots,$$

where the preceding equation may be interpreted as the Taylor series (Kaplan 1957) expansion of the response function ϕ in the neighborhood of the origin. It is possible within a given region to obtain a satisfactory fit to this expansion by the use of a polynomial equation. Generally these polynomials are of second-order or less (Box 1952a, Box and Hunter 1954, Box and Lucas 1959, Box and Behnken 1959); however, some work has been done with third-order designs (Gardiner, Grandage, and Hader 1959, Debaun 1959).

If N experimental observations are made at strategically chosen points within the region, estimates of the regression coefficients (b's) may be calculated. The conditions for these N observations constitute the experimental design and may be presented as a matrix D, called the design matrix. Hence, the design matrix provides a program for the performance of the N experiments. The elments of the i^{th} row of the design matrix represent the specific levels of the k factors to be used in the i^{th} observation. These elements also represent the spatial coordinates of the i^{th} experimental point in the k-dimensional factor space which when augmented by one dimension, namely, that of the dependent variable or response, constitute a (k+1)-dimensional space in which the response surface lies.

The problem of selecting a ''best'' design has been a difficult and arduously studied one. Generally, however ''best'' refers to that design for which the variances of the estimated regression coefficients are at a minimum.

Box and Hunter (1957) suggest that the ''goodness'' of an experimental design should be judged partly on the precision of the estimates of the regression coefficients and partly on the magnitude of the bias of those estimates. They list these qualities as desirable in the experimental design:

1. The design should estimate the assumed model satisfactorily within the region of interest.

2. The design should have a built-in check on the assumed model.

3. The design should not have an excessively large number of experimental points.

4. The design should be ''blockable!''

5. The design should be easily expanded.

6. The design should have the properties of orthogonality and rotatability, both of which will be discussed later.

Basically there are four different methods generally used for solving the optimization problem. They are the factorial method, the univariate method, the random method, and the steepest ascent method. The reader is referred to Satterthwaite (1959) and Budne (1959) for a discussion of the random method and to Friedman and Savage (1947) for a discussion of the univariate method. A review and explanation of the remaining two methods, the factorial method and method of steepest ascent, follows. The factorial method is reviewed because of its inherent role in the method of steepest ascent; the method of steepest ascent is reviewed since it is fundamental for this dissertation.

The Factorial Method

The factorial method, generally accredited to Fisher (1949) and Yates (1937) is ideally suited for investigating a surface in a preassigned range of values of the independent variables such as in the neighborhood of the maximum. Its adaptability to blocking as well as the ease with which the original design may be augmented by additional observations make it a very useful and frequently employed design. It is of particular value where the experiment is of a non-sequential nature and the factors of the discrete type. However, there are several disadvantages which can, and often do, nullify these advantages. For example, a factorial design requires experimentation to explore regions that may turn out, in view of their results, to be of no value or interest due to their lack of proximity to the maximum. Also, a factorial design frequently investigates a small region comprehensively or a large region superficially. In the former event, a maximum might be missed entirely while in the latter the experimental combinations might be so chosen as to miss the maximum even though one exists within the range being investigated. At least, the factorial method can provide an indication of the direction of the maximum, a fact that the Box technique exploits.

The general technique is to conduct trials at the points of a grid in the factor space. To this end, combinations of factors at specified levels are selected and the response determined for each of these combinations. Next, a regression model, thought to be of sufficient order, is assumed and the regression coefficients determined, generally by the method of least squares (Nielson 1957). This is

followed by the estimation of the conditions for optimal response. These conditions are determined by taking the partial derivatives of the calculated regression equation with respect to the independent variables concerned, setting the partial derivatives equal to zero, and solving the resulting equations simultaneously.

If all the points of the grid are used, the design is said to be a complete factorial. If several determinations at each point are made, it is said to be a replicated factorial. If systematic selection of only a part of the complete factorial points is made, one obtains a fractional factorial or a confounded design (Yates 1937). References for the fractional factorial designs include Finney (1945), Finney (1946), Plackett and Burnam (1946), Kempthorne (1947), Rao (1947), and Davies and Hay (1950).

Of particular interest are the two-level factorial and two-level fractional factorial designs. Two-level designs consist of two levels of each of the independent variables and all of their combinations. If only a part of these combinations is used, one has again a fractional factorial design. Specific advantages of these designs include:

1. First-order effects are determined with maximum accuracy.
2. Specific interaction terms can be isolated.
3. The design is readily augmented to increase precision.
4. The design may be the basis of a ''composite design'' for fitting second-degree surfaces.
5. The adequacy of the model may be checked.
6. These designs have the property of rotatability which enables the researcher to conduct long sequences of experiments in the presence of a trend and yet maintain minimum variance

estimates that are mutually orthogonal, orthogonal to the block effects and orthogonal to the trend.

Orthogonality refers to the case in which the factors in the design matrix are functionally independent. This property is of particular advantage for computational purposes when comparing the effects of the various factors.

The concept of rotatability was first introduced by Box and Hunter (1957) when they were confronted with the selection of a k-dimensional experimental design of order d such that the variance function would be ''spherical''[1]. For the variance function to be ''spherical'', the variances and co-variances of the estimates of the regression coefficients made from the least squares estimate of the truncated Taylor series expansion must be constant on circles, spheres or hyper-spheres about the center of the design. If a design has this property of rotatability, it can be rotated through any angle about the fixed center, and one reasonably expects and obtains a constant quantity of information regardless of direction of orientation. Box and Hunter (1957) prove that a necessary and sufficient condition for a design of order d (d=1,2...) to be rotatable is that the moments of the independent variables be equal through order 2d.

These rotatable designs have the further advantage that replication of the center points provides an estimate of the experimental error and thus provides a basis for testing the lack-of-fit of the model.

Rotatable designs are given by Box and Hunter (1957) for k-dimensional

---

[1] A k-dimensional experimental design of order d is a set of experimental points or observations in the k-dimensional factor space selected so that all the coefficients in the d^{th} degree polynomial can be estimated.

experimental designs of order d (k=2,3...99, d=1,2). Several blocking arrangements as well as confidence regions for the stationary point are also discussed. A more comprehensive discussion of the confidence region for a stationary point and an example of the determination of a confidence region is given by Box and Hunter (1954). Gardiner, Grandage and Hader (1959) expand this concept to third-order rotatable designs and give several examples.

The Method of Steepest Ascent

This method may be thought of as having, or proceeding in, two successive phases. The first is concerned with the location of a near-stationary region; and the second, with the investigation of the response surface in this region.

No knowledge of the form of the function is assumed. However, it is assumed that the function has a unique maximum and that the function is ''smooth'' and continuous. The method is also dependent upon a number of other assumptions. The most important of these are:

1. All factors or independent variables must be measurable quantitative values.
2. The theoretical response is a function of the independent variables.
3. The observed response is a function of the independent variables plus an arbitrary experimental error.
4. The errors are normally and independently distributed with a mean of zero and a variance of one.

It is generally thought that the technique works best when the errors are small and there is a previous estimate of the error. It is possible to utilize the method when these conditions do not exist, however, by replicating observations and making use of the estimate of the variance obtained by replication of the center point of the central composite design. This, of course, partially defeats the purpose of the method, namely, the reduction in the overall number of experiments necessary to obtain the region desired.

The first phase is a sequential one, much like the univariate method and has as its objective the location of the near-stationary region. It proceeds as follows:

1. Select all the factors thought to be influencing the response. Frequently, the selection of the factors must be modified to include only those factors thought to be especially significant since the selection of all factors might lead to an excessive number of independent variables. It has been shown that the method loses some of its effectiveness under such circumstances (Brooks 1959).
2. Make an initial linear approximation of the response surface in the vicinity of the estimated optimum.
3. Use the lack-of-fit term in the analysis of variance to test whether the linear approximation fits within the limits of experimental error. If it does, proceed to step 4; if not, a second-order model is adopted and the experimental points augmented by additional observations so that the regression coefficients for the second-order model may be estimated.

4. Calculate the path of steepest ascent (Fig. 1 and 2). This path is determined by incrementing the coordinates of the center of the design in the factor space by amounts proportional to the relative size of the regression coefficients as determined in step 1.
5. Make observations along this path at regularly spaced intervals until the observed response differs significantly from the predicted response using the regression equation determined in step 1.
6. Make a second linear approximation of the response surface using the last agreeing predicted and observed response as a center for the design and proceed to steps 3, 4 and 5, etc.
7. Continue this process until it is stopped by the adoption of a second-order model in step 3.

The selection of the independent variables and the subsequent determination of the levels at which the observations are to be made represents a crucial stage in the use of the method of steepest ascent. It has been pointed out, and rightly so, that the entire method is dependent upon this selection and that the size of the regression coefficients, and hence the path of steepest ascent, will vary according to the width of the interval between the factor levels. Two obvious errors are possible through the improper selection of the factor levels. First, if the interval is too large, a maximum may be missed entirely; second, if the interval is too small, the experimental error may ''mask'' any true difference in the response at the two different levels and the subsequent calculation of the path of steepest ascent will therefore be in error. In fact, it may even be in the wrong direction entirely,

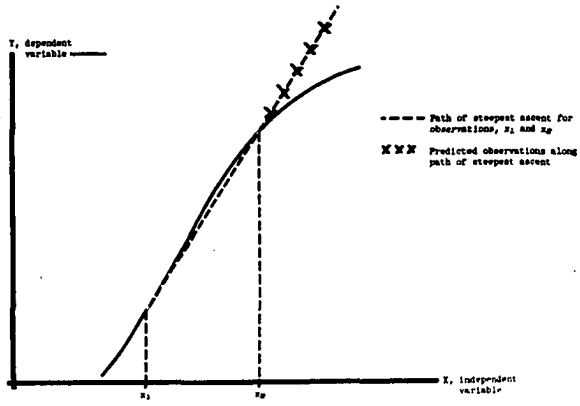


Figure 1. - Response surface for one independent variable, illustration of the path of steepest ascent for observations at x_1 and x_2.

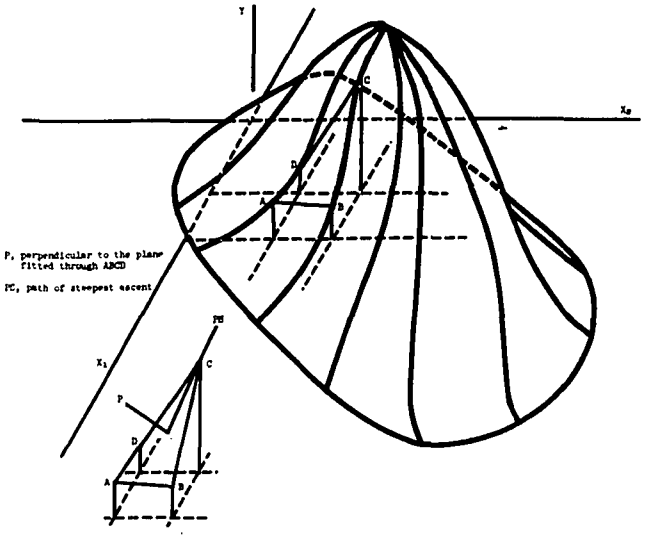


Figure 2. - Response surface for two independent variables, and path of steepest ascent.

a mistake that will generally be corrected with the determination of the next path. In either case, it is possible that the independent factors' levels may be so chosen that the responses are exactly equal and it will appear as if the procedure has reached a plateau. Further experimentation on the axis to complete the central composite design will prove this erroneous. One may then adjust the interval between the levels of the factors and reinitiate the experimental procedure.

David and Arens (1949) discuss the question of the spacing of independent variables and suggest criteria that might be applicable to various situations, Legendre and Tchebysheff spacing being the most carefully considered. Specifically their work is concerned with that situation in which the dependent variable may be observed for a continuous range of values of a single independent variable and the functional relationship is unknown. They also investigate empirically the situation where the true functional relation is a quadratic and the fitted curve linear.

De la Garza (1954) and Kiefer and Wolfowitz (1952) discuss the case where the function is known apart from various parameters.

The second phase of the method of steepest ascent is concerned with the determination of the exact optimal factor combination and with the characterization of the response surface in the vicinity of a near-stationary region. It requires the following;

1. The adoption of the second-order model.
2. The estimation of the regression coefficients.
3. The calculation of the optimal factor combination.
4. The determination of the contour lines for the second-order model.

5. The determination of the canonical forms along with the transformation necessary to obtain them.

In some cases a confidence region for the predicted optimal response (Box and Hunter 1954) and an interpretative study of the response surface are also made (Box and Youle 1955).

Once the regression coefficients of the second-order model have been determined, generally by the method of least squares (Nielsen 1957), standard mathematical techniques may be utilized for finding the maximum or minimum values of the dependent variable. These techniques involve the calculation of the partial derivatives (Kaplan 1957a), setting the partial derivatives equal to zero (Kaplan 1957b) and the subsequent simultaneous solution of these equations to give the values of the independent variables that provide the maximum or minimum of the dependent variable.

Intuitively, this might be expected if one remembers the basic concept of the partial derivative. The partial derivative represents the rate of change in the dependent variable for a given change in the independent variable. By setting the partials equal to zero, one is, in effect, imposing the restriction that the change in the dependent variable with respect to a change in the independent variable be zero. This is the condition that one might expect to find at maxima or minima of the dependent variable.

Next, in the development of a response surface, should be the determination of contour lines (Fig. 3) based on the second-order model. The contour lines represent all of the combinations of the independent variables that will give a fixed level of response of the dependent variable.

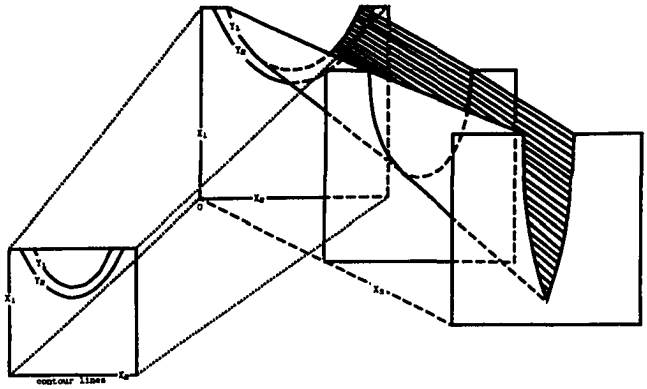


Figure 3. - Response surface for three independent variables, illustrating two levels of the dependent variable.

For example, in the three (k=3) dimensional case, the contour lines are determined by choosing a constant level for the dependent variable and one independent variable, and substituting these values into the experimentally determined second-order equation. One then has a second-order equation in the two remaining independent variables. By selecting various levels of one of these two independent variables and substituting it into the second-order equation, one obtains quadratic equations for the third variable. The solutions of these equations may then be plotted to give contour lines of the response surface (Fig. 3). The contour lines represent all of the combinations of two of the independent variables that will give a fixed level of response of the dependent variable at a given level of the third independent variable. Hence, the contour lines represent the intersection of horizontal planes and the response surface. That is, the horizontal plane will represent a constant level of one of the independent variables and the line of intersection will represent a constant magnitude of the dependent variable or response.

There are several advantages that might be expected from the analysis and study of contour lines. By plotting contour lines it is possible to comprehend features of the response surface which may be exploited in an attempt to further increase or decrease the response in question. It may be possible to optimize a primary response with secondary constraining or restricting conditions. This demand on the optimizing process would obviously lead to a set of conditions which are ''best'' only in a compromising sense provided there is a conflict between the optimal conditions for the primary response and the constraint. If no such conflict exists, there is no constraint.

Careful consideration of the response surface may lead to a better understanding of the mechanism being examined and also indicate the need for additional investigation. Box and Youle (1955) have discussed this latter possibility in some detail and have demonstrated how various characteristics of an operating system may affect contour lines.

The canonical form plays the same role in the second-degree approximations as does the path of steepest ascent in the first-degree approximations. An extensive discussion of the value and use of the canonical form is given by Box and Hunter (Chew 1958) in which changes in the signs of the general form of the canonical equation are illustrated and the relative sizes of the coefficients are examined. Also several 3-dimensional surfaces of second-degree are discussed and the general procedure for a k-dimensional second-degree fitted surface is outlined. This general outline is as follows:

1. The coordinates of the new center and the value of the response at this center are calculated.
2. The canonical form of the equation is determined.
3. The new coordinates are determined in terms of the old coordinates.

Investigating an unknown response function in several independent variables frequently leads an experimenter to a sequence of experiments that falls naturally into blocks. The initial step in such a sequence will be that of approximating the response function with a first-order model.

The first-order model adopting Box's notation is:

$$y = b_o x_o + \ldots + b_k x_k + e.$$

Based upon the results of such a sequence of experiments, decisions may be made concerning the following:

1. Elimination of one or more of the independent variables by comparing the magnitudes of the fitted coefficients.
2. Expansion or reduction of the interval between levels.
3. Lack-of-fit of the model.
4. Prediction of paths to optimal responses.

It is interesting to note that the same class of designs for the first-order approximation are obtained if it is required that the design be such that:

1. The variance of the predicted response $(\hat{y})$ be a minimum.
2. The confidence region for the regression coefficients, b's, be minimum.
3. The variance of $\hat{y}$ be equal for all equidistant points from the origin.

The first-order designs may be visualized geometrically as the k-dimensional space of the vertices of a regular simplex in (k+1) dimensions where the $(k+1)^{th}$ dimension represents the response. If the number of observations is N = k + 1, these designs represent the projection of the (N-1) dimensional simplex into k-dimensional space.

Frequently, in the application of Box's technique, it is desirable to adopt a central composite rotatable design which can be blocked (Box 1952b, DeBaun 1956, and Box and Hunter 1957). The central composite design consists of N_c points at the vertices of a cube corresponding to a 2^k factorial design or some fraction thereof, with coordinates coded $(\pm 1, \pm 1, \ldots, \pm 1)$, plus $N_a = 2k$ ''axial'' points with coordinates coded $(\pm \alpha, 0, \ldots 0)$, $(0, \pm \alpha, 0, \ldots 0)$, $\ldots (0, \ldots 0, \pm \alpha)$ plus

N_o points at the center of the design with coordinates coded (0,...,0). These sets of points offer an opportunity for blocking. If we let N_{co} and N_{ao} represent the center points associated with the cube and axial points respectively, the requirements for orthogonal blocking and rotatability are:

$$\alpha = \left[\frac{N_c\ (N_a + N_{ao})}{2\ (N_c + N_{co})}\right]^{1/2} \quad \text{and} \quad \alpha = N_c^{1/4} \quad \text{respectively.}$$

To satisfy both requirements we require

$$\frac{N_c^{1/2}}{2} = \frac{(N_c + N_{co})}{(N_a + N_{ao})}.$$

As an example let us examine the k=4 dimensional design which is to be used in Chapter IV. The design matrix consists of the $2^k = 2^4 = 16$ experimental points at the vertices of the cube, with coordinates $(\pm 1, \pm 1, \pm 1, \pm 1)$, 2k = 8 axial points with coordinates $(\pm \alpha, o,o,o)$, $(o, \pm \alpha, o,o), (o,o, \pm\alpha, o), (o,o,o, \pm\alpha)$ where $\alpha = N_c^{1/4} = 16^{1/4} = 2$, for rotatability and with a yet-to-be-determined number of points at the center. If seven points are observed at the center the variance function V(p) = V(1) = V(o) (Box and Draper 1959) and hence one attains a relatively uniform distribution of precision. If rather than seven, we chose $N_o = 6$ to satisfy the equation below:

$$\frac{N_c^{1/2}}{2} = \frac{N_c + N_{co}}{N_a + N_{ao}}, \quad 2 = \frac{16 + N_{co}}{8 + N_{ao}} = \quad 2 = \frac{16 + 4}{8 + 2}$$

orthogonal blocking and rotatability will be attained. The only effect on the variance function will be to decrease slightly the precision near the center of the design.

Now, the 2^4 factorial part of the design may be further divided into two orthogonal blocks. This division can be accomplished by

confounding the block effect with the 4 factor interaction, that is, placing all points with a positive third-order interaction, 4 factors, in one block and all experimental points with a negative third-order interaction into a second block. The four center points associated with the ''cubic'' part of the design are then divided, two points to each block.

Now, the original design of 30 experimental points is divided into 3 blocks of 10 points each, the 8 axial points plus 2 center points, and the two half-replicates plus 2 center points each. The design is orthogonally blocked and rotatable and has nearly uniform precision.

The analysis of variance when blocking is used breaks the residual sum of squares into three parts as shown below.

Analysis of variance

Source	d.f.	Sum of squares
Residual	$N_r + N_o - 1$	R
Blocks	$B - 1$	$SS_b = \sum_{b=1}^{B} N_b(Y_b - Y)^2$
Experimental error + lack-of-fit	N_r	$R - SS_b - SS_e$
Pure error	$N_o - B$	(Pure error)

Where the pure error = the sum of the individual sums of squares for repeated observations at the center of each block,

and B = number of blocks

N_b = number of observations in the b^{th} block

Y_b = mean of the observations in the b^{th} block

N_o = number of observations at the center

R = residual sum of squares.

SS_e = pure error

Y = grand mean

N_r = degrees of freedom for experimental error and lack-of-fit

The applicability of the Box technique in the field of medical research has not been adequately studied and reported in the literature, it is the purpose of this research to investigate this area.

CHAPTER III

APPLICATION OF THE METHOD OF STEEPEST ASCENT TO THE RESPONSE SURFACE OF THE NITRATE-NITRITE REDUCTASE ACTIVITY IN SALIVARY SEDIMENT

The utilization of the method of steepest ascent and the application of the statistical methods for characterizing a response surface appeared especially apropos to the investigation of several factors simultaneously such as the effects of pH, temperature and electron donor concentration on the reduction of nitrate in saliva. If successful, the study should illustrate the application of this experimental methodology to discrete biochemical phenomena.

Experimental Methods

The reduction of nitrate and nitrite in saliva has been described in an earlier report (Goaz and Biswell 1961). Inasmuch as nitrate reduction in the oral cavity may play a role in the energy metabolism and the assimilation of nitrogen by oral bacteria, an investigation of the relationship between nitrate reduction and the process of decay seemed pertinent to them. Although an initial pilot study of forty young adults did show a positive correlation between the capacity of an individual's whole saliva to reduce nitrate and his caries experience, the correlation was not found to be significant.

To evaluate this relationship critically, in a more extensive clinical study, two refinements of the technique of measuring the nitrate-nitrite reductase activity of saliva have been introduced.

In the initial pilot study by Goaz and Biswell (1961) the capacity of whole saliva to reduce nitrate and nitrite was measured. However, it has subsequently been determined that the reaction of whole saliva on nitrate is not only dependent upon the nitrate-nitrite reductase activity that is characteristic of the salivary sediment but also upon the level of electron donor compounds in the supernatant, and possibly on other undisclosed factors in the supernatant. That fluctuations in the composition of the supernatant can precipitate variations in the apparent enzymatic activity of whole saliva has been demonstrated by recombining aliquots of an individual salivary sediment with the supernatants from a number of other salivary specimens and noting the variations in the measured activity of the sediment when associated with different supernatants (Fig. 4). This finding suggested the possibility that the composition of the supernatant and its effect on the activity of the sediment may be more variable than the potential enzymatic level of the sediment; while the nitrate-nitrite reductase activity of the sediment may be more related to the magnitude of decay activity prevailing at the time the salivary specimen was collected. In order to eliminate this possible source of variation on the measured nitrate-reducing activity of a salivary sample, this activity was determined on the salivary sediment plus a standard artificial electron donor, yeast extract (1 mg/ml). Also, it seemed reasonable that the most accurate assay of a salivary sediment's

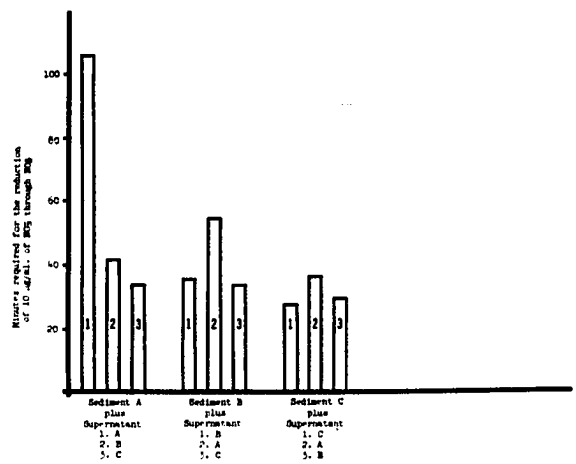


Figure 4. - Effect of supernatant composition on the measurable nitrate-nitrite reductase activity of salivary sediment. Demonstrated by recombining aliquots of an individual salivary sediment with the supernatant from other salivary specimens. (Supernatants and sediments from salivary specimens, A, B and C).

capacity to reduce nitrate, and the most valid for comparative studies, would probably be achieved if the determination was conducted under the optimal conditions for the reactions that the salivary enzyme catalyzes.

Through pilot studies, the effects of three factors, independent variables, on this salivary activity were considered to be important. These were factors that were relatively easily controlled and had marked influence on the apparent enzyme activity; these factors were pH, temperature, and concentration of electron donor, x_1, x_2 and x_3, respectively.

It was on this basis that the present study of the most efficient and effective manner of delineating these optimal conditions for this salivary reaction was undertaken.

The salivary sediment examined in this study was from saliva obtained by paraffin stimulation. It was collected upon arising, before breakfast and before brushing the teeth. The sample was kept refrigerated until it was assayed, except for the interval during which the saliva donor brought the sample to the laboratory.

Upon receipt at the laboratory, the saliva was pooled and depleted of endogenous nitrate and nitrite by incubation at 37° C. It was then centrifuged at 10,000 rpm for 30 minutes and the sediment washed twice in distilled water.

The sediment was resuspended in distilled water to one-tenth the volume of the salivary pool from which it was obtained. This slurry of the particulate fraction, containing the nitrate-nitrite reductase, was then stored at 4° C.

Inasmuch as it was anticipated that the determination of the optimal conditions for this enzyme system would require a number of days, and in veiw of the necessity of making serial determinations on the same stock of enzyme sediment, it was necessary, initially, to make an estimate of the decay or loss of measurable enzyme activity that would result from storage.

The experimental design for the estimation of the loss of salivary activity that occurs during storage was that described by Snedecor (1956) for the analysis of covariance. The covariate here is the duration of storage in days.

Saliva from seven individuals was collected and pooled and the sediment removed and stored at 4^{o} C. Three aliquots (Group A, B, and C) were drawn from this stock of sediment and their capacity to reduce nitrate through nitrite was determined daily for four days (Table 1). The order of measuring the activity of each aliquot was randomized each day to eliminate possible bias in the experimental observations.

The nitrate-reducing capacity of the sediment was made by removing a measured amount, 0.5 ml, of the stock sediment, resuspending in distilled water and diluting to a volume equivalent to 70 percent, 3.5 ml, of the salivary pool that the sediment originally represented. One milliliter of the standard electron donor solution was added and the pH of the system adjusted to the desired level by adding solid monobasic or dibasic potassium phosphate[1]. The buffered enzyme system was then placed in a constant temperature water bath. After a sufficient period to permit temperature equilibrium, 50 μg of nitrate, as KNO_3, in 0.5 ml of aqueous solution were added, and the time required

[1] The solid buffer was used to minimize dilution.

TABLE 1

Effect of Storage on the Capacity of Salivary Sediment to Reduce Nitrate

Time required (minutes, Y) to reduce 10 μg. NO_3^-/ml through NO_2^-

Aliquot Group	Storage time (days, X)			
	0	1	2	3
A	13	19	22	25
B	14	17	26	23
C	14	18	23	27

Determination of intercept and slope for the regression lines

Equation for the regression line; $Y = a + bX$

$$b = \text{regression coefficient} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

$a = Y$ intercept; $\Sigma Y = na + b\Sigma X$

Aliquot A

$\Sigma X^2 = 14$	$\Sigma XY = 138.0$	$\Sigma Y^2 = 1639.00$
$C^* = 9$	$C = 118.0$	$C = 1560.25$
$\Sigma x^2 = 5$	$\Sigma xy = 19.5$	$\Sigma y^2 = 78.75$

Aliquot B

$\Sigma X^2 = 14$	$\Sigma XY = 138$	$\Sigma Y^2 = 1690$
$C = 9$	$C = 120$	$C = 1600$
$\Sigma x^2 = 5$	$\Sigma xy = 18$	$\Sigma y^2 = 90$

Aliquot C

$\Sigma X^2 = 14$	$\Sigma XY = 145$	$\Sigma Y^2 = 1778$
$C = 9$	$C = 123$	$C = 1681$
$\Sigma x^2 = 5$	$\Sigma xy = 22$	$\Sigma y^2 = 97$

* C = correction factor for ΣX^2 or $\Sigma Y^2 = \frac{(\text{sum of variable})^2}{\text{no. of observations}}$

correction factor for $\Sigma XY = \frac{(\text{sum of X})\ (\text{sum of Y})}{\text{no. of observations}}$

TABLE 1 (Cont).

Effect of Storage on the Capacity of Salivary Sediment to Reduce Nitrate

	Regression Coefficients	
Aliquot	b	a
A	3.90	13.90
B	3.60	14.60
C	4.40	13.90
Mean for the three periods	3.97	14.13

to reduce the added nitrate, through nitrite, to a yet unidentified end-product determined by the spot plate method. The end-point of the spot plate method as described by Goaz and Biswell (1961) is based on the disappearance of the nitrite. These measurements of enzyme activity were made at 37° C, pH 6.4 and using 1 mg/ml of yeast extract as a standard hydrogen donor. Those levels of pH and hydrogen donor had previously been shown by Goaz and Biswell (1961) to be optimal at 37° C using a univariate method consisting of only one round.

The assumptions of this design are that the samples were drawn from a normal population with common variances. Due to the manner in which the aliquots were obtained, it was thought these assumptions were met.

The regression coefficients, b's for Groups A, B and C were determined by the method of least squares (Table 1) and tested for homogeneity. To make this test, the difference between the sum of squares for the common regression and the sum of squares within samples was calculated. This difference measures the difference among the samples' regression coefficients, and its mean square may be compared with the mean square within samples (Table 2). It may be observed that the mean square for the regression for one of the samples (aliquot B) is relatively large. On inspection, the data shows that the determination of nitrate-reducing activity in aliquot B on the second day storage varied markedly. It was felt this variation reflected an experimental error and hence explained the large mean square for

TABLE 2

Analysis of Covariance Effect of Storage on the Nitrate Reducing Activity of Salivary Sediment

Aliquot	d.f.	x^2	xy	y^2	Reg. Coef.	Deviations from Regression		
						d.f.	$y^2-(xy)^2/x^2$	Mean Square
A	3	5	19.5	78.75	3.9	2	2.70	1.35
B	3	5	18.0	90.00	3.6	2	25.20	12.60
C	3	5	22.0	97.00	4.4	2	0.20	0.10
Within						6	28.10	4.683
Reg. Coef.						2	1.63	0.815
Common	9	15	59.5	265.750	3.97	8	29.73	3.716
Adj. Means						2	1.17	0.585
Total	11	15	59.5	266.917		10	30.90	

$$F \text{ (variance ratio)} = \frac{\text{Mean Square of Regression Coefficients}}{\text{Mean Square Within Samples}} = \frac{0.815}{4.683} = 0.174; \text{ d.f.} = 2, 6$$

aliquot B. In any event, the F value (variance ratio)[1] obtained was obviously not significant. Thus, there was a failure to reject the null hypothesis at a greater than 25 percent confidence level, and it was assumed that there was no difference between the slopes of the regression lines for the three groups (A, B and C) (Fig. 5). It was next necessary to test the hypothesis that the population regression lines coincide. This was done by comparing the difference between the total sums of squares and the common sums of squares or the adjusted mean squares with the common mean square. This difference corresponds to the sample differences in elevation.

$$F = \frac{0.585}{3.716} = 0.1574; \quad \text{d. f.} = 2, 8,$$

which lacks significance at the 25 percent confidence level. Hence, the three groups of aliquots may have the same regression lines; at least, there is not a great enough difference among the groups to be detected by samples of this size. Consequently, the data from all three groups was pooled to make the best estimate of the regression equation which was found to be,

$$y = 14.13 + 3.97x. \qquad (3)$$

To characterize the response surface of the nitrate reducing activity of saliva, a stock of sediment was extracted from a salivary pool from forty donors, and stored at 4° C. The levels of the previously selected independent variables, at which the investigation was initiated, are shown in Table 3, along with their coded values. Such

[1] The variance ratio is a statistic which was developed by Snedecor (1956) and named in honor of R. A. Fisher. It is the ratio of two variables which follow the chi-square distribution function divided by their respective degrees of freedom, and forms itself a distribution function called the F or Fisher distribution. It can be, and is, used as a means of making confidence or significance statements.

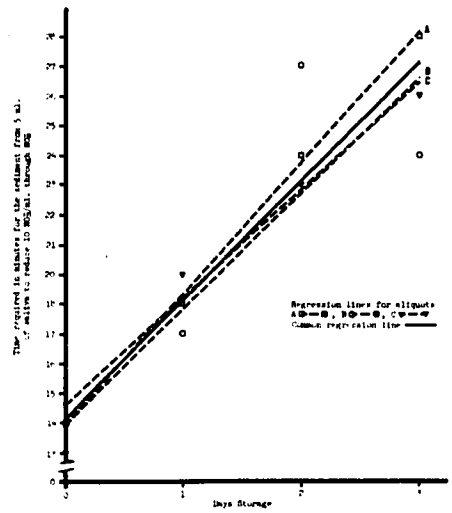


Figure 5. - Regression of enzyme activity on storage time.

TABLE 3

Initial Code

Variables Investigated		Coded Value		
		-1.0	0	+1.0
pH	(x_1)	5.0	5.25	5.5
Temp.[1]	(x_2)	31.0	33.00	35.0
Conc.[2] H+ donor	(x_3)	0.4	0.50	0.6

[1] Degrees centigrade

[2] mg/ml of yeast extract

coding is a linear transformation and simplifies subsequent calculations. Since each of the independent variables was investigated at only two different levels in any experiment; the upper level of x was coded as +1 and the lower level as -1. The relations between the coded and experimental units were selected as follows:

$$x_1 = \frac{\text{pH} - 5.25^1}{0.25}; \quad x_2 = \frac{\text{Temp.} - 33^1}{2}; \quad x_3 = \frac{\text{Conc. H+ Donor} - 0.5^1}{0.1} \quad (4)$$

Assuming the general linear model in three independent variables, $y = b_0x_0 + b_1x_1 + b_2x_2 + b_3x_3 + e$, the first linear approximation of the response surface was made. As one might well expect, this model did not adequately represent a four-dimensional curved surface consisting of the three independent and one dependent variables, however it gave some indication as to the orientation of the surface. To determine the regression coefficients for this model a 2^3 factorial experiment[2] was necessary. The experimental conditions or combinations, the coded values of x_1, x_2 and x_3, and the observed results, expressed as the reaction time in minutes for the reduction of nitrate through nitrite under these conditions, are shown in Table 4.

The values of the regression coefficients in the linear model were then determined by the method of least squares. The method of

---

[1] These levels of pH, temperature, and electron donor concentration were chosen as the level at which to initiate the investigation since preliminary work with this enzyme system seemed to indicate that it would be less than optimal, and yet provide a reasonable area from which to approach the maximum, utilizing the method of steepest ascent.

[2] A 2^3 factorial experiment is the study of three factors, or variables, at two different levels and all the combinations thereof. Hence, a total of eight determinations would be made.

TABLE 4

Initial Attempt to Approximate Response Surface

Experimental combinations of independent variables	X Coded Scale x_o[1]	x_1	x_2	x_3	y Observed Results (reaction time in minutes)
1	1	-1	-1	-1	105
2	1	1	-1	-1	62
3	1	-1	1	-1	72
4	1	1	1	-1	37
5	1	-1	-1	1	87
6	1	1	-1	1	36
7	1	-1	1	1	38
8	1	1	1	1	32

[1] Since b_o occurs in every equation, the dummy variable x_o, which has the value of +1 for every observation in the sample, is introduced for computational purposes.

least squares is to minimize the sum of the squared deviations between the observed y values and the predicted values using the assumed model, that is, one minimizes the following equation:

$$\sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} (y_i - b_o x_o - b_1 x_{1i} - b_2 x_{2i} - b_3 x_{3i}). \qquad (5)$$

This is accomplished by taking the partial derivatives of these squared deviations with respect to b_o, b_1, b_2 and b_3, and equating them to zero. The resulting system of normal equations had the form:

$$b_o\Sigma x_o^2 + b_1\Sigma x_o x_1 + b_2\Sigma x_o x_2 + b_3\Sigma x_o x_3 = \Sigma x_o y$$

$$b_o\Sigma x_1 x_o + b_1\Sigma x_1^2 + b_2\Sigma x_1 x_2 + b_3\Sigma x_1 x_3 = \Sigma x_1 y$$

$$b_o\Sigma x_2 x_o + b_1\Sigma x_2 x_1 + b_2\Sigma x_2^2 + b_3\Sigma x_2 x_3 = \Sigma x_2 y$$

$$b_o\Sigma x_3 x_o + b_1\Sigma x_3 x_1 + b_2\Sigma x_3 x_2 + b_3\Sigma x_3^2 = \Sigma x_3 y$$

The numerical solution of these equations would, in general, be extremely tedious but their solution can be simplified by the use of appropriately coded values as in this study or, by the abbreviated Doolittle technique (Graybill 1961).

These normal equations may be written in matrix notation as,

$$X'X\,\beta = X'Y,$$

where X is the data matrix augmented by a dummy variable x_o which represents the mean (Table 4). X' is the transpose[1] of the X matrix, β is the coefficient matrix, and Y is the observed matrix (Table 4). The general form of the normal equation in matrix notation is as follows:

---

[1] The transpose of an X matrix is a matrix of the same elements but with its rows and columns interchanged.

$$\underset{X'X}{\begin{bmatrix} \Sigma x_0 & \Sigma x_0x_1 & \Sigma x_0x_2 & \Sigma x_0x_3 \\ \Sigma x_1x_0 & \Sigma x_1^2 & \Sigma x_1x_2 & \Sigma x_1x_3 \\ \Sigma x_2x_0 & \Sigma x_2x_1 & \Sigma x_2^2 & \Sigma x_2x_3 \\ \Sigma x_3x_0 & \Sigma x_3x_1 & \Sigma x_2x_3 & \Sigma x_3^2 \end{bmatrix}} \quad \underset{\beta}{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}} = \underset{X'Y}{\begin{bmatrix} \Sigma x_0y \\ \Sigma x_1y \\ \Sigma x_2y \\ \Sigma x_3y \end{bmatrix}} \qquad (6)$$

Inspection of the X'X matrix above will reveal that it is symmetrical around the diagonal from the upper left to the lower right. By a judicious selection of codes for the independent variables, namely that given in equation (4) the normal equations for this specific experiment reduce to the simply solved equations:

$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} \quad \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 469 \\ -135 \\ -111 \\ -83 \end{bmatrix}$$

Solving these equations by matrix algebra one obtains:

$$b_0 = \frac{469}{8} = 58.625; \quad b_1 = \frac{-135}{8} = -16.875; \quad b_2 = \frac{-111}{8} = -13.875$$

$$b_3 = \frac{-83}{8} = -10.375. \qquad (7)$$

By substituting the regression coefficients, obtained above, into the linear model, one finds:

$$\hat{y} = 58.625 - 16.875x_1 - 13.875x_2 - 10.375x_3 . \qquad (8)$$

The analysis of variance for this initial approximation of the response surface is presented in tabular form in Table 5. This analysis indicates that the lack-of-fit term, 145.12, is significant[1] and following the standard Box technique that a new model which includes second-

---

[1] This significance is based on the fact that the mean square of the ''lack-of-fit'' term is substantially larger than the estimate of the variance that was observed in the study of enzyme activity decay during storage. See Table 5.

TABLE 5

Analysis of Variance for Reaction Times
(first approximation)

Source	d.f.	s.s.	m.s.
1. Mean	1	27,495.125	
2. Linear model	3	4,679.375	1,559.79
3. Lack-of-fit	4	580.500	145.12
4. Error[1]	--	------	3.716
Total	8	32,755.000	

[1] Estimated by common mean square from the analysis of covariance of the effect of storage on enzyme activity.

order terms should be adopted. The adoption of such a model with its subsequent augmentation of experimental observations should either prove to be satisfactory or unsatisfactory. If satisfactory, the method may be pursued as described previously, if unsatisfactory, further exploration is indicated. The loss is thus confined to the determination of the augmented points.

In this case however, since the original selection of the factor levels was sufficiently removed from the optimum, the adoption of a second-order model at this juncture was not deemed warranted. Also, to demonstrate the usefulness of the technique to be employed, the path of steepest ascent was determined along with three expected values along the path. The calculation of these expected values using equation (8) and their observed values are summarized in Table 6. These calculations were used as a verification of the experimental conditions selected for the second approximation of the surface and were initially chosen as appropriate on the basis of visual inspection of a three-dimensional plot of the response surface constructed with the data from the first approximation.

The experimental conditions, results, normal equations, regression coefficients and analysis of variance for the second approximation of the response surface are summarized in Table 7.

The analysis of variance (Table 7) of the second attempt to approximate the response surface indicates that the lack-of-fit term, 0.125, was not significant. Hence, the path of steepest ascent was again calculated (Table 8), and some expected values along the path were determined using the equation given in Table 7. The results of the experimental observations made at these points are also shown in

TABLE 6

Calculations of Initial Path of Steepest Ascent

		x_1 (pH)	x_2 (Temp.)	x_3 (H+ don.)	$\hat{y}^1$ (min.)	y^2 (min.)
1.	Relative change in design units = b_i	-16.8750	-13.8750	-10.3750		
2.	No. of original units = 1 design unit	0.2500	2.0000	0.1000		
3.	Relative change in original units	- 4.2188	-27.7500	- 1.0375		
4.	Change per 0.25 pH units	0.2500	1.6444	0.0615		
	Path of steepest ascent					
5.	Initial levels	5.2500	33.0000	0.5000		
6.	Observations along the path					
	Coded					
	No. x_1 x_2 x_3					
	(1) +1 0.8222 0.615	5.5000	34.6444	0.5615	+23.961	32
	(2) +2 1.6444 1.230	5.7500	36.2888	0.6230	-10.702[3]	19
	(3) +3 2.4666 1.845	6.0000	37.9332	0.6845	-45.366[3]	

1 $\hat{y}$ is the predicted value.

2 y is the observed corrected value.

3 These calculated expected values being negative confirm the suspicion that the linear model was not fitting the response surface. However, further investigation also confirmed the use of the path of steepest ascent, indicating the direction of the optimum (or minimum).

TABLE 7

Second Attempt to Fit the Linear Equation to a 2^3 Factorial Experiment

Data

Experimental combinations of independent variables	Coded Units X Coded Scale x_0	x_1	x_2	x_3	y Observed Results (reaction time in min.) observed (y_o)	corrected (y_c) [1]
1	1	-1	-1	-1	21	17
2	1	1	-1	-1	19	15
3	1	-1	1	-1	18	14
4	1	1	1	-1	15	11
5	1	-1	-1	1	20	16
6	1	1	-1	1	18	14
7	1	-1	1	1	16	12
8	1	1	1	1	14	10

[1] The observed times corrected for storage by use of the previously determined decay curve (Fig. 5), $y_c = y_o - 3.97x$, where x is the number of days the enzyme has been stored.

Relation between coded and experimental units

Variables		Coded Value -1.0	0	+1.0
pH	(x_1)	5.75	6.0	6.25
Temperature	(x_2)	37.00	39.0	41.00
Conc. H+ donor; mg/ml	(x_3)	0.70	0.8	0.90

Normal Equations

$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 109 \\ -9 \\ -15 \\ -5 \end{bmatrix}$$

Solutions: $b_o = 13.625$; $b_1 = -1.125$; $b_2 = -1.875$; $b_3 = -0.625$

Substituting into the linear model: $\hat{y} = 13.625 - 1.125x_1 - 1.875x_2 - 0.625x_3$.

Analysis of variance (second approximation)

Source	d.f.	s.s.	m.s.
1. Mean	1	1,485.125	
2. Linear model	3	41.375	13.792
3. Lack-of-fit	4	0.500	0.125
4. Error[1]	--	------	3.716

[1] Estimated by common mean square from the analysis of covariance of the effect of storage on enzyme activity.

TABLE 8

Calculation of the Second Path of Steepest Ascent

		x_1 (pH)	x_2 (Temp.)	x_3 (H^+ don.)	$\hat{y}^1$ (min.)	y^2 (min.)
1.	Relative change in design units = b_i	-1.125	-1.875	-0.625		
2.	No. of original units = 1 design unit	0.250	2.000	0.100		
3.	Relative change in original units	-0.281	-3.750	-0.062		
4.	Change per 0.25 pH unit	0.250	3.333	0.056		
	Path of steepest ascent					
5.	Initial levels	6.000	39.000	0.800		
6.	Observations along the path					
	Coded					
No.	x_1	x_2	x_3			
(1)	+1	1.666	0.560	6.250	42.333	0.856	+9.0262	10
(2)	+2	3.332	1.120	6.500	45.666	0.912	+4.4275	7
(3)	+3	4.998	1.680	6.750	48.999	0.968	-0.1712	7

[1] $\hat{y}$ is the predicted value

[2] y is the observed corrected value

Table 8. Since the third observation ($x_1 = +3$) along the path of steepest ascent was found to be significantly different from the expected value, the need for a new, third, approximation of the surface was indicated. The code, design matrix, results and necessary computations are given in Table 9.

Using the data from this third set of experiments, the path of steepest ascent was again determined (Table 10), and the predicted values ($\hat{y}$) along this path calculated. Corresponding observed values for the path are also presented in Table 10. Guided by these results and following the above procedure, another set of experiments was conducted, employing the experimental combinations given in Table 11[1]. The results of this fourth approximation are summarized in Table 12.

The analysis of variance indicates the linear model is no longer satisfactory and that the second-order model,

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 , \quad (9)$$

should be adopted. The rationale for this decision is as follows: The lack-of-fit term and the sum of squares for the regression on the linear terms are both relatively small. Also, from the previous experiments it seemed obvious that the investigation was descending the response surface, and the lack of significance of these terms, as well as the magnitude of the regression coefficients, b's, indicated that either a minimum or a plateau had been reached. It was assumed that this was a minimum, and the basic design of the fourth attempt to fit

---

[1] Table 11 also includes the coded levels for the completed central composite rotatable design. At this point, however, only the columns +1 and -1 are pertinent.

TABLE 9

Third Attempt to Fit the Linear Equation
to a 2^3 Factorial Experiment

Relation between coded and experimental units

Variables		Coded Value -1.0	0	+1.0
pH	(x_1)	6.65	6.75	6.85
Temperature	(x_2)	46.00	48.00	50.00
Conc. H+ donor; mg/ml	(x_3)	0.94	0.97	1.00

Code of x's examined

$$x_1 = \frac{\text{pH} - 6.75}{0.1}; \qquad x_2 = \frac{\text{Temp.} - 48}{2}; \qquad x_3 = \frac{\text{H+ donor} - 0.9}{0.03}$$

Data

Experimental combinations of independent variables	Coded Units X Coded Scale x_0	x_1	x_2	x_3	Y Observed Results (reaction time in min.) observed (y_o)	corrected (y_c)
1	1	-1	-1	-1	16	8
2	1	1	-1	-1	15	7
3	1	-1	1	-1	22	14
4	1	1	1	-1	20	12
5	1	-1	-1	1	15	7
6	1	1	-1	1	15	7
7	1	-1	1	1	22	14
8	1	1	1	1	20	12

Normal Equations

$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 81 \\ -5 \\ 23 \\ -1 \end{bmatrix}$$

Solutions

$$b_o = \frac{81}{8} = 10.125; \quad b_1 = \frac{-5}{8} = -0.625; \quad b_2 = \frac{23}{8} = b_3 = \frac{-1}{8} = -0.125$$

TABLE 9 (Cont).

Third Attempt to Fit the Linear Equation

to a 2^3 Factorial Experiment

Substituting into the linear model:

$$\hat{y} = 9 - 0.625x_1 + 2.875x_2 - 0.125x_3$$

Analysis of variance (third approximation)

Source	d.f.	s.s.	m.s.
1. Mean	1	820.125	
2. Linear model	3	69.375	23.125
3. Lack of fit	4	1.500	0.375
4. Error[1]	—	—	3.716
Total	8	891.000	

[1] Estimated by common mean square from the analysis of covariance of the effect of storage on enzyme activity.

TABLE 10

Calculation for the Third Path of Steepest Ascent

		x_1	x_2	x_3	$\hat{y}^1$ (min.)	y^2 (min.)
1.	Relative change in design units = b_i	-0.6250	2.875	-0.1250		
2.	No. of original units = 1 design unit	0.1000	2.000	0.0300		
3.	Relative change in original units	-0.0625	5.750	-0.0037		
4.	Change per 2^{o} C	-0.0217	2.000	-0.0013		
	Path of steepest ascent					
5.	Initial levels	6.7500	48.000	0.9700		
6.	Observations along the path					

No.	x_1	x_2	x_3					
(1)	0.217	-1	0.0433	6.7717	46.000	0.9713	+7.1089	8
(2)	0.434	-2	0.0866	6.7935	44.000	0.9726	+4.0929	7
(3)	0.651	-3	0.1299	6.8152	42.000	0.9739	-0.0644	11

[1] $\hat{y}$ is the predicted value

[2] y is the observed corrected value

TABLE 11

Relation Between Coded and Experimental Units

Variables		Coded Value -1.682	-1	0	+1	+1.682
pH	(x_1)	6.6818	6.76	6.85	6.95	7.0182
Temperature	(x_2)	42.6360	44.00	46.00	48.00	49.3640
Conc. H+ donor; mg/ml	(x_3)	0.9195	0.94	0.97	1.00	1.0205

TABLE 11

Relation Between Coded and Experimental Units

Variables		Coded Value -1.682	-1	0	+1	+1.682
pH	(x_1)	6.6818	6.76	6.85	6.95	7.0182
Temperature	(x_2)	42.6360	44.00	46.00	48.00	49.3640
Conc. H+ donor; mg/ml	(x_3)	0.9195	0.94	0.97	1.00	1.0205

TABLE 12

Fourth Attempt to Fit the Linear Equation
to a 2^3 Factorial Experiment

Relation between coded and experimental units

See Table 11

Code of x's examined

$$x_1 = \frac{\text{pH } -6.85}{0.1}; \qquad x_2 = \frac{\text{Temp.}-46}{2}; \qquad x_3 = \frac{\text{H+ donor}-0.97}{0.03}$$

Data

Experimental combinations of independent variables	Coded Units X Coded Scale x_o	x_1	x_2	x_3	y Observed Results (reaction time in min.) observed (y_o)	corrected (y_c)[1]
1	1	-1	-1	-1	16	8
2	1	1	-1	-1	14	6
3	1	-1	1	-1	16	8
4	1	1	1	-1	17	9
5	1	-1	-1	1	15	7
6	1	1	-1	1	14	6
7	1	-1	1	1	16	8
8	1	1	1	1	17	9

[1] Corrected as in Table 7.

Normal Equations

$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 61 \\ -1 \\ 7 \\ -1 \end{bmatrix}$$

Solutions

$$b_o = 7.625; \quad b_1 = -0.125; \quad b_2 = 0.875; \quad b_3 = -0.125$$

Substituting into the linear model:

$$\hat{y} = 7.625 - 0.125x_1 + 0.875x_2 - 0.125x_3$$

TABLE 12 (Cont.)

Analysis of variance (fourth approximation)

Source	d.f.	s.s.	m.s.
1. Mean	1	4651.125	
2. Linear model	3	6.375	2.125
3. Lack of fit	4	3.500	0.875
4. Error[1]	—	—	3.716
Total	8	4750.000	

[1] Estimated by common mean square from the analysis of covariance of the effect of storage on enzyme activity.

the surface was augmented with axial and center observations. These observations substantiated the impression that the minimum, and not a horizontal inflection point, had been reached, for the axial observations all gave equal or longer times than the original 2^3 points (Table 13).

The design used for this augmentation was that suggested by Box and Wilson (1951), a central composite rotatable design. This design calls for augmenting the original 2^3 points at the center and on each axis 1.6818 experimental units from the origin. This distance from the origin is chosen to preserve rotatability. The experimental combinations are given in Table 13, and their spatial relationships are illustrated in Figure 6.

The normal equations, the X'X and the X'Y matrices for this specific experiment are shown in Table 14, along with the $(X'X)^{-1}$ matrix[1], the solutions for the regression coefficients, b's, the second-order equation, and the analysis of variance.

To find the coordinates of the minimum, in coded form, the partial derivatives of the second-order equation are taken with respect to x_1, x_2 and x_3. If:

$$\hat{y} = 6.7061 - 0.3195x_1 + 1.0051x_2 - 0.0732x_3 - 0.1278x_1^2 + 1.9932x_2^2 - 0.4813x_3^2 + 0.625x_1x_2 + 0.125x_1x_3 + 0.125x_2x_3$$

then:

$$\frac{\partial y}{\partial x_1} = - 0.3195 - 0.2556x_1 + 0.625x_2 + 0.125x_3 \qquad (11)$$

---

[1] The $(X'X)^{-1}$ matrix is the inverse of the X'X matrix.

TABLE 13

Central Composite Rotatable Design for a 2^3 Factorial Experiment

Experimental Combinations of Independent Variables	Coded Units X Coded Scale x_o	x_1	x_2	x_3	y Observed Results (reaction time in min.) observed (y_o)	corrected (y_c)[1]
Original 2^3 points						
1	1	-1	-1	-1	16	8
2	1	1	-1	-1	14	6
3	1	-1	1	-1	16	8
4	1	1	1	-1	17	9
5	1	-1	-1	1	15	7
6	1	1	-1	1	14	6
7	1	-1	1	1	16	8
8	1	1	1	1	17	9
Axial points						
9	1	-1.6818^2(6.68)		0	20	8
10	1	+1.6818(7.02)		0	18	6
11	1	0	-1.6818^2(42.6)	0	23	11
12	1	0	+1.6818(49.4)	0	27	15
13	1	0	0	-1.6818^2(0.92)	18	6
14	1	0	0	+1.6818(1.02)	18	6

[1] The observed times corrected for storage by use of the previously determined decay curve (Fig. 5), $y_c=y_o - 3.97x$, where x is the number of days the enzyme has been stored. In the above table x=2 for the original 8 points and x=3 for the axial and center points.

[2] The linear distance of these axial points from the origin is determined such that the design is rotatable.

TABLE 13 (Cont.)

Central Composite Rotatable Design for a 2^3 Factorial Experiment

Experimental Combinations of Independent Variables	Coded Units X Coded Scale x_o	x_1	x_2	x_3	y Observed Results (reaction time in min.) observed (y_o)	corrected (y_c)[1]
Center points						
15	1	0^2 (6.85)	0^2 (46)	0^2 (0.97)	18	6
16	1	0	0	0	20	8
17	1	0	0	0	18	6
18	1	0	0	0	19	7
19	1	0	0	0	19	7
20	1	0	0	0	18	6

[1] The observed times corrected for storage by use of the previously determined decay curve (Fig. 5), $y_c = y_o - 3.97x$, where x is the number of days the enzyme has been stored. In the above table x =2 for the original 8 points and x =3 for the axial and center points.

[2] The duplication of the center point given this particular design a built-in estimate of variance.

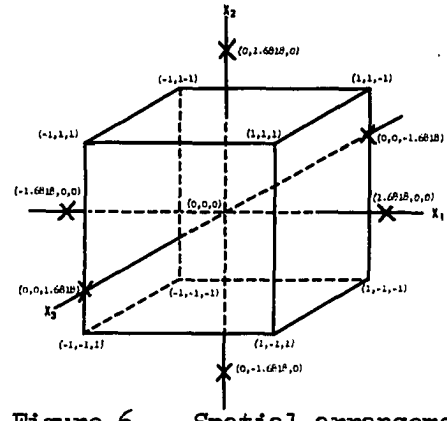


Figure 6. - Spatial arrangement of the coded experimental combinations, illustrating a central, composite, rotatable design for three factors (independent variables).

TABLE 14

Composite Design

X'X Matrix

20	0	0	0	13.656	13.656	13.656	0	0	0
0	13.656	0	0	0	0	0	0	0	0
0	0	13.656	0	0	0	0	0	0	0
0	0	0	13.656	0	0	0	0	0	0
13.656	0	0	0	24	8	8	0	0	0
13.656	0	0	0	8	24	8	0	0	0
13.656	0	0	0	8	8	24	0	0	0
0	0	0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	0	8	0
0	0	0	0	0	0	0	0	0	8

X'Y

153.0000
-4.3636
13.7272
-1.0000
100.5920
134.5280
94.9360
5.0000
1.0000
1.0000

$(X'X)^{-1}$ (1)

0.166338	0	0	0	-0.056791	-0.056791	-0.056791	0	0	0
0	0.073224	0	0	0	0	0	0	0	0
0	0	0.073224	0	0	0	0	0	0	0
0	0	0	0.073224	0	0	0	0	0	0
-0.056791	0	0	0	0.069389	0.006889	0.006889	0	0	0
-0.056791	0	0	0	0.006889	0.069389	0.006889	0	0	0
-0.056791	0	0	0	0.006889	0.006889	0.069389	0	0	0
0	0	0	0	0	0	0	0.125	0	0
0	0	0	0	0	0	0	0	0.125	0
0	0	0	0	0	0	0	0	0	0.125

(1) The $(X'X)^{-1}$ Matrix is the inverse of the X'X Matrix.

TABLE 14 (Cont.)

The solutions for the b's from the least squares equations of the $(X'X)^{-1}$ Matrix

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_{11} \\ b_{22} \\ b_{33} \\ b_{12} \\ b_{13} \\ b_{23} \end{bmatrix} = \begin{bmatrix} 6.7061 \\ -0.3195 \\ 1.0051 \\ -0.0732 \\ -0.1278 \\ 1.9932 \\ -0.4813 \\ 0.6250 \\ 0.1250 \\ 0.1250 \end{bmatrix} \qquad \begin{matrix} 2S_{bi}{}^{*} \\ \begin{bmatrix} 0.3063 \\ 0.4420 \\ 0.4420 \\ 0.4420 \\ 0.4302 \\ 0.4302 \\ 0.4302 \\ 0.5774 \\ 0.5774 \\ 0.5774 \end{bmatrix} \end{matrix}$$

* $S_{bi} = (V(b_i))^{1/2}$ = (diagonals of the inverse Matrix times replication error)$^{1/2}$

Substituting into the second-order equation:

$$y = 6.7061 - 0.3195x_1 + 1.0051x_2 - 0.0732x_3 - 0.127x_1^2 + 1.9932x_2^2 - 0.4813x_3^2 + 0.6250x_1x_2 + 0.1250x_1x_3 + 0.1250x_2x_3$$

Analysis of variance

Source	d.f.	s.s.	m.s.
1. Mean	1	1170.450000	
2. First-order terms	3	14.607129	4.86904
3. Second-order terms	6	68.560576	11.42676
4. Lack-of-fit	5	6.048300	1.20966
5. Replication error	5	3.334000	0.66681
Total	20	1263.000000	

$$\frac{\partial y}{\partial x_2} = 1.0051 + 0.625x_1 + 3.9864x_2 + 0.125x_3$$

$$\frac{\partial y}{\partial x_3} = -0.0732 + 0.125x_1 + 0.125x_2 + 0.9626x_3$$

Setting these partial derivatives equal to zero and solving using the abbreviated Doolittle technique, it was found that

$$x_1 = 1.3319$$

$$x_2 = -0.0187$$

$$x_3 = 0.2640$$

Decoding by use of the following equations,

$$x_1 = \frac{x_1' - 6.85}{0.10} = \text{pH}$$

$$x_2 = \frac{x_2' - 46}{2} = \text{Temperature}$$

$$x_3 = \frac{x_3' - 0.97}{0.10} = \text{Conc. of H+ donor,}$$

the optimal conditions were found to be,

x_1 (pH) = 6.98

x_2 (Temp) = 45.96° C

x_3(H+ donor) = (0.98 mg/ml), where the primes are dropped.

Substituting these values in the second-order equation (10) it was found that the minimum predicted value for the reduction time was,

$$\hat{y} = 6.0106 \text{ minutes.}$$

To verify or test empirically the accuracy and correctness of the conclusion that this set of conditions was optimal, several observations were made in the surrounding region. The results of these

tests are shown in Table 15 and demonstrate an increase in the reaction time for any deviation from the optimal conditions as determined by the method of steepest ascent.

Using equation (10) to approximate the response surface, contour lines were determined following the procedure described above. These contour lines are presented in Figures 7 and 8. Nine levels of the dependent variable were calculated, however, for clarity of representation only six levels of response are illustrated.

The coordinates for the independent variables x_1 and x_2, were computed for these nine levels of the dependent variable at nine selected levels of x_3 (the independent variable held constant at each level). The levels of x_3 were selected such that there would be a constant increment over the range of interest. These coordinates for the three dimensional contours of the surface were calculated, using a previously written program[1] for the IBM 650 computer, and were printed on the IBM 407 tabulator.

To help visualize the interrelations of the independent variables and their response, a model of the surface was constructed. This construction was accomplished by connecting the contour lines of equal magnitude of the response from one x_3 - level to another. To aid in distinguishing one level of response from another, different colors were used to represent the various levels of response (Figure 9). The model is a four-dimensional protrayal of a saddle-shaped surface, where the fourth dimension, response, is represented by the various colors.

---

[1] Obtained from the University of North Carolina Department of Statistics: Coordinates and Plotting Cards for Five Variable Second-Degree Model. 06.1.004, 1, 2 AG/10-57.

TABLE 15

The Effect on Salivary NO_3-NO_2 Reducing Capacity by Altering the Experimental Conditions from the Optimal

Saliva Sample		x_1(pH)	x_2(Temp. C)	x_3(Conc. H^+ donor; mg/ml)	Reaction time in min.
A	[1]	6.98	46.0	0.980	11
		6.73	44.0	0.985	13
		7.02	44.2	0.955	12
		6.72	48.2	0.940	12
		6.99	49.6	1.000	13
B	[1]	6.98	46.0	0.980	12
		6.73	44.0	0.985	14
		7.02	44.2	0.955	13
		6.72	48.2	0.940	13
		6.99	49.6	1.000	13

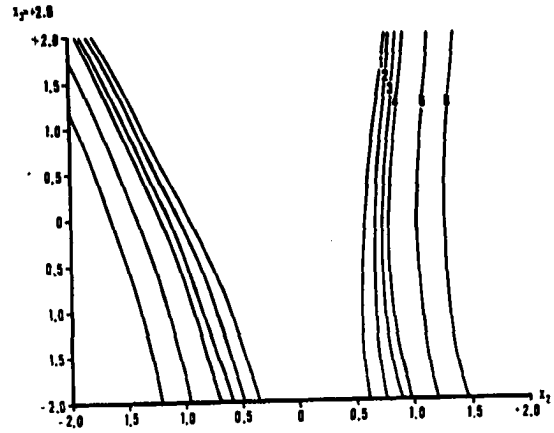
[1] Optimal conditions as determined by the method of steepest ascent

TABLE 15 (Cont).

The Effect on Salivary NO_3-NO_2 Reducing Capacity by Altering the Experimental Conditions from the Optimal

Saliva Sample	x_1 (pH)	x_2 (Temp. °C)	x_3 (Conc. H^+ donor; mg/ml)	Reaction time in min.
C[1]	6.98	46.0	0.980	10
	6.88	44.4	0.970	11
	6.75	44.4	0.955	11
	6.72	48.0	0.985	12
	6.99	49.4	0.985	11

[1] Optimal conditions as determined by the method of steepest ascent



Contour Lines for Salivary Nitrate Nitrite Reductase

X_1 = pH X_2 = Temperature (degrees C.) $\Delta\hat{y}_{0.05} \doteq 0.300.5$ minutes

X_3 = H^+ donor concentration (mg. yeast extract per ml.)

$\hat{y}_1$ = 1 = 6.0106 minutes (minimum)

$\hat{y}_2$ = 2 = 6.3110 minutes (105 per cent minimum)

$\hat{y}_3$ = 3 = 6.6117 minutes (110 per cent minimum)

$\hat{y}_4$ = 4 = 6.9122 minutes (115 per cent minimum)

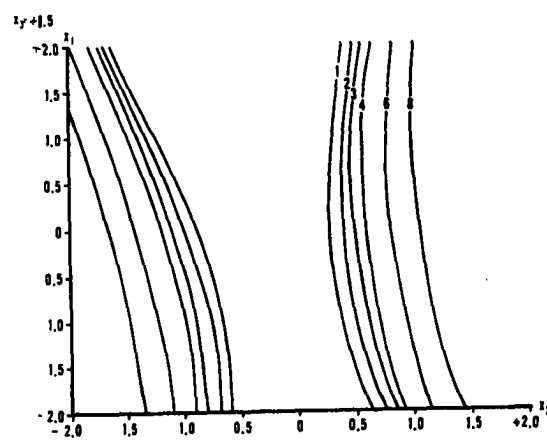
$\hat{y}_5$ = 5 = 7.2127 minutes (120 per cent minimum)

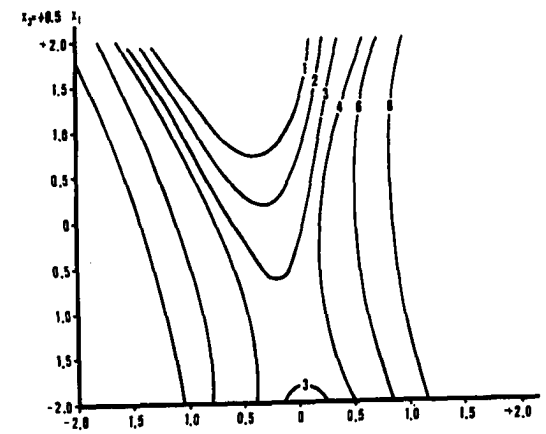
$\hat{y}_6$ = 6 = 7.6138 minutes (130 per cent minimum)

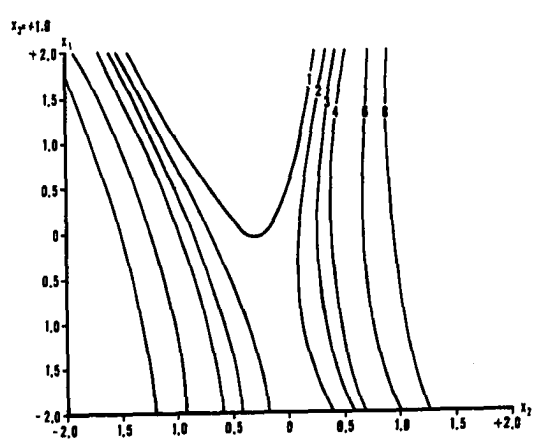
$\hat{y}_7$ = 7 = 8.4148 minutes (140 per cent minimum)

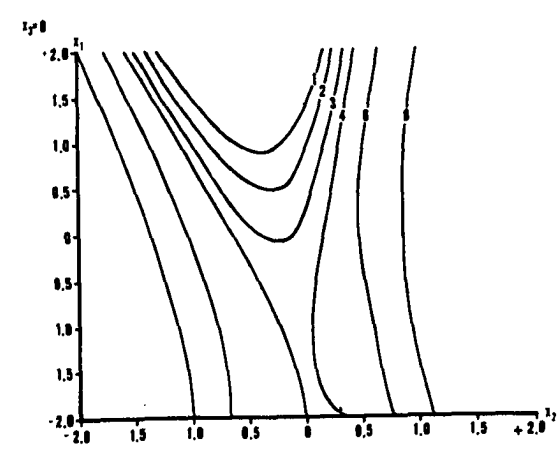
$\hat{y}_8$ = 8 = 9.0159 minutes (150 per cent minimum)

$\hat{y}_9$ = 9 = 9.9175 minutes (165 per cent minimum)



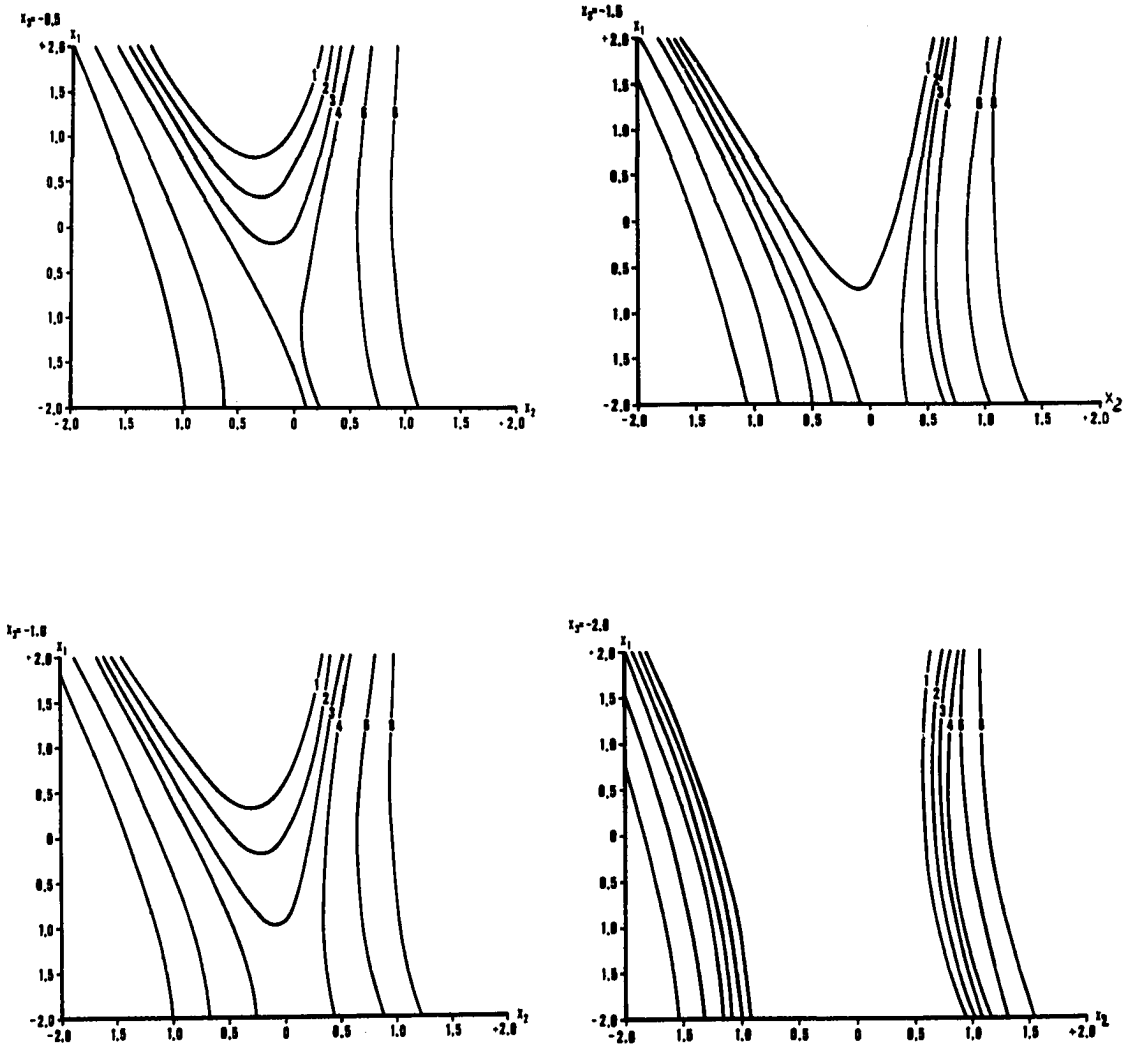






Contour Lines

Figure 7.



Contour Lines

Figure 8.

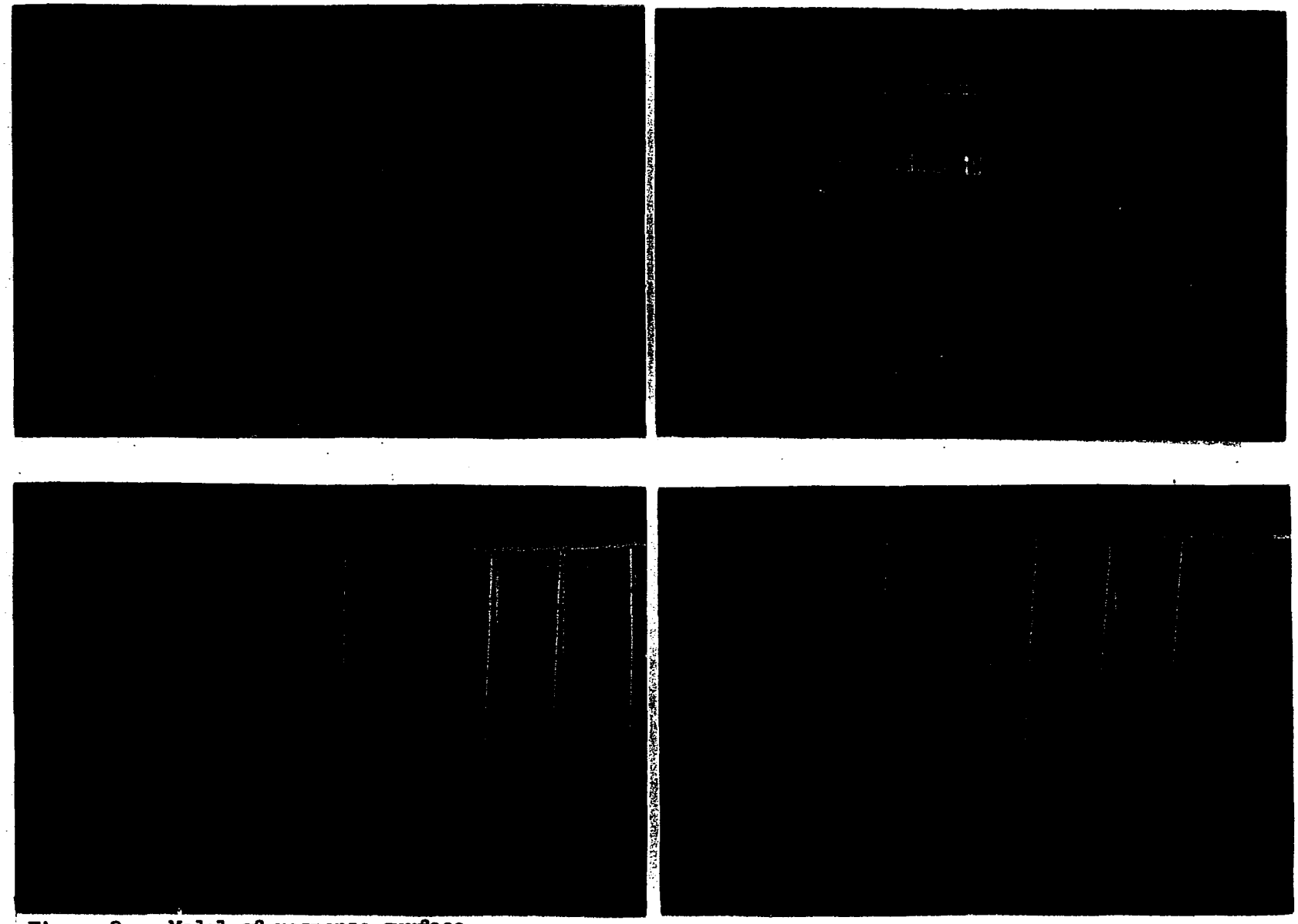


Figure 9. - Model of response surface.

Examination of this response surface reveals that the three independent variables studied did indeed influence the enzyme activity. More careful scrutiny shows that temperature contributes the predominating influence above pH 6.0 and electron donor concentration of 0.7 μg/ml. Also, by use of this response surface one may estimate the predicted response of various combinations of the independent variable levels.

CHAPTER IV

THE APPLICATION OF THE METHOD OF STEEPEST ASCENT TO THE DETERMINATION OF TOTAL SERUM PROTEIN

The second application of the method of steepest ascent was drawn from the field of clinical pathology and is concerned with the refinement of a diagnostic serum test used by physicians.

The fractionation and determination of the individual serum proteins has long been considered of great pragmatic value. It permits studies of the chemical nature and the physiological function of each protein in health and disease. Of particular interest in medicine are the A/G ratio (albumin-globulin ratio) and total serum protein. Hence it is desirable that one not only be able to determine the total serum protein (TSP) present but that methods for separating the plasma proteins be available. These fractionating processes are generally based upon (1) differences in physical properties (solubility in H_2O, salt solutions, etc.); or (2) rates of sedimentation in an ultracentrifuge; or (3) rate of electrophoretic migration. In the study of serum protein which is being reported here, the difference in the solubility of albumin and of globulin in ehtyl alcohol was used. The TSP was determined first, and then the albumin. The amount of globulin was obtained by subtraction. It is assumed the globulins and albumin

account for all but a very insignificant part-quantitatively speaking-of the total serum protein.

Hoping to reduce the variation in the determination of the total serum protein and its subsequent effect on the A/G ratio, it was decided to determine the optimal laboratory procedure for the Biuret method, Ferro, Patrick, Ham and Ana Bell (1961), and to investigate the response surface near the optimum. Such an investigation should allow the diagnostic tool to be made more sensitive to small changes. Then by use of the contour lines, it might be possible in future studies to retain optimal or near optimal conditions for the TSP determination and yet optimize the procedure for determining the amount of albumin present and hence the A/G ratio.

Experimental Method

Careful study of the literature and laboratory technique led to the selection of the following four independent variables:

x_1 = reaction temperature in C^o.

x_2 = ml of Na_2SO_4

x_3 = ml of Biuret Reagent

x_4 = reaction time in minutes.

The concentration of Na_2SO_4 and the concentration of the Biuret Reagent were held constant.

Since the individual conducting the research project was unable to make a prediction based on his personal experience as to possible interaction effects that might exist among the four variables being studied, it was decided to run a complete 2^4 factorial design for the initial approximation to the surface. If the higher order interactions

proved to be non-significant, then a fractional factorial consisting of a 1/2 replication of the 2^4 design was to be adopted for the future approximations of the surface and determinations of the path of steepest ascent. Therefore, the assumed initial model was:

$y = b_0x_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_1x_2 + b_6x_1x_3 + b_7x_1x_4 + b_8x_2x_3 + b_9x_2x_4 + b_{10}x_3x_4 + b_{11}x_1x_2x_3 + b_{12}x_1x_2x_4 + b_{13}x_1x_3x_4 + b_{14}x_2x_3x_4 + b_{15}x_1x_2x_3x_4 + e$, where y was the predicted amount of total serum protein.

The results of epxerimentation will be presented in tabular form whenever possible and the (X'X) and the $(X'X)^{-1}$ matrices omitted in order to conserve space.

Table 16 shows the original selection of the levels of the various factors and their respective codes. The design matrix and results of the first set of experiments are tabulated in Table 17.

The least squares normal equations, $X'X\,\beta = X'Y$, were solved for the regression coefficients. Using these estimates of the regression coefficients and substituting into the assumed model above, it was found that

$\hat{y} = 8.4238 - .0104x_1 - .3114x_2 + .0970x_3 - .0078x_4 + .0330x_1x_2 + .0712x_1x_3 - .0414x_1x_4 - .1368x_2x_3 + .1332x_2x_4 - .0412x_3x_4 + .0095X_1x_2x_3 - .0660x_1x_2x_4 - .0072x_1x_3x_4 + .0509x_2x_3x_4 - .0096x_1x_2x_3x_4$.

The (X'Y) and (β) matrices are given in Table 18, and the analysis of variance presented in Table 19.

Having approximated the regression coefficients, the first path of steepest ascent was determined (Table 20). Proceeding along the path, observations were made at those combinations indicated in Table 20.

TABLE 16

Factors in Coded and Original Units for First Linear Approximation

Factors	Code	
	-1	-1
X_1 (Reaction temperature ^oC)	27.0	30.0
X_2 (ml of Na_2SO_4)	4.5	5.5
X_3 (ml of Biuret)	3.5	4.5
X_4 (Reaction time minutes)	10.0	20.0

TABLE 17

First Approximation to Response Surface

Experiment number	Design Matrix[1]																Observed for Standard	Observed for Unknown	Total Serum Protein gm/100 ml
	X_0	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	X_3X_4	$X_1X_2X_3$	$X_1X_2X_4$	$X_1X_3X_4$	$X_2X_3X_4$	$X_1X_2X_3X_4$			
1	+	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	1.03	1.19	8.0874
2	+	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-	1.04	1.17	7.8750
3	+	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-	0.90	1.10	8.5555
4	+	-	+	-	-	-	+	+	-	-	+	+	+	+	-	-	0.85	1.05	8.6471
5	+	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+	0.95	1.10	8.1053
6	+	-	+	+	-	-	-	+	+	-	-	-	+	-	+	+	0.80	0.95	8.3125
7	+	-	+	+	+	-	-	-	+	+	+	-	-	+	-	-	0.75	0.95	8.8667
8	+	-	+	-	+	-	+	-	-	+	-	+	-	-	+	+	0.83	1.07	9.0241
9	+	+	-	-	-	-	-	-	+	+	+	+	+	-	+	-	1.03	1.15	7.8155
10	+	+	-	-	+	-	-	+	+	-	-	+	-	+	-	+	1.06	1.17	7.7264
11	+	+	-	+	-	-	+	-	-	+	-	-	+	+	-	+	0.93	1.10	8.5555
12	+	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+	0.87	1.06	8.7294
13	+	+	-	+	+	-	+	+	-	-	+	-	-	-	+	-	0.95	1.11	8.1789
14	+	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-	0.78	1.00	8.7500
15	+	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-	0.86	1.07	8.7093
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0.76	0.96	8.8421

[1] + represents the +1 level of the independent variable, - represents the -1 level of the independent variable.

TABLE 18

First Approximation of the Response Surface
(Total Serum Protein)

(X'Y)	(B)		(B)
134.7807	b_0		8.4238
-.1665	b_1		-.0104
4.9817	b_2		.3114
1.5523	b_3		.0970
-.1251	b_4		-.0078
.5273	b_5		.0030
1.1395	b_6		.0712
-.6623	b_7		-.0414
-2.1895	b_8		-.1368
2.1315	b_9	=	.1132
-.2359	b_{10}		-.0412
.1513	b_{11}		.0095
-1.0561	b_{12}		-.0660
-.1147	b_{13}		-.0072
.8147	b_{14}		.0509
-.0153	b_{15}		-.0096

TABLE 19

Analysis of Variance for First Linear Approximation

Source	d.f.	s.s.	m.s.
Total	16	1137.9067	
Mean	1	1135.3648	
Main Effects	4	1.7046	.4261
Lack-of-fit	11	.8373	.0761
x_2x_3	(1)	.3107	.3107
x_2x_4	(1)	.2840	.2840
Remainder	(9)	.2426	.0270
Error[1]	50	------	.0225

[1] Estimated from previous laboratory findings.

$$V(b_i) = C_{ii}\ S^2 = \frac{.0225}{16} = .0014,\ 2S_{b_i} = 0.75,\ F_{(1,50)} = 4.03$$

TABLE 20

Initial Path of Steepest Ascent

	(biuret) X_3	(temp) X_1	(Na_2SO_4) X_2	(time) X_4	(Predicted) $\hat{y}$	(Observed) y
Relative change in design units	0.0907	-0.014	0.3114	-0.0078		
No. of orig. units=1 design unit	0.5	1.5	0.5	5.0		
Relative change in orig. units	0.0485	-0.0156	0.1557	-0.0390		
Change per .5 cc change in X_3	0.5	-0.1608	1.6052	-0.4020		
Rounded	0.5	-0.16	1.60	-0.40		
Initial level	4.0	28.5	5.0	15.0	8.4328	8.4651
Observations along path (1)	4.5	28.34[2]	6.6	14.6	9.5144	8.4858
(2)	5.0	28.18[2]	8.2	14.2	10.6277	8.6471
(3)	5.5	28.02[2]	9.8	13.8	11.7026	8.9250
(4)	6.0	27.86[2]	11.4	13.4	12.7967	8.9688
Coordinates of observations Along path of S.A.					Standard	Observed
Initial value	0	0	0	0	.86	1.04
Obs. 1	1	0.3333	3.2	-.08	.66	.80
2	2	0.3333	6.4	- .16	.51	.63
3	3	0.3333	9.6	-.24	.41	.51
4	4	0.3333	12.8	-.32	.32	.41

[1] $\hat{y} = 8.42379 - .0104X_1 + .3114X_2 + .0970X_3 - .0078X_4$

[2] Since a water bath was being used to control reaction temperature and these increments were so small, all four observations along the path of steepest ascent were made at a constant temperature: that is, 28°.

TABLE 20 (Cont.)

Initial Path of Steepest Ascent

Code

$X_1 = (X-28.5)\ /1.5$ $\qquad$ $X_3 = (X-4.0)\ /1.5$

$X_2 = (X-5.0)\ /.5$ $\qquad$ $X_4 = (X-15)\ /5$

It was immediately evident that the observed responses y's and predicted responses $\hat{y}$'s were substantially different and a new set of factor combinations should be determined. However, in view of the progressive increase in the observed y's, it was decided to continue making observations along the path for a way and then to fit another factorial to the surface.

It is interesting to note in the calculation of the path of steepest ascent (Table 20) that if relative changes are based upon x_2, then there are unreasonably small steps in x_1, x_3, and x_4. In fact, the first several points along the path would lie within the previously examined ranges of the variables x_1, x_3, and x_4. On the other hand, if x_1 were used as the standard and increments were calculated for the other variables for, say, a 1 or 2 degree change in x_1, then the increment in x_2 would be exorbitantly large.

In determining the factor combinations and levels for this second approximation, two things seemed to be indicated. First, the small size of the regression coefficients suggested that a more informative and accurate determination of the regression coefficients might be made if the intervals between the factor levels were increased. Second, the relatively small size of the ''lack-of-fit'' term in the analysis of variance (see Table 19) indicated that the interaction terms were rather small and that a fractional factorial design might be applicable. A breakdown of the analysis of variance shows that even though the ''lack-of-fit'' term is non-significant, the (x_2x_3) and (x_2x_4) interactions are significant. In the one-half replicate of the 2^4 design, these interactions were confounded with the (x_1x_4) and

(x_1x_3) terms respectively. Since the latter two are non-significant this confounding is justifiable, and it was decided that the 1/2 replicate of such a design should be run for the second approximation to the surface and subsequent determination of the path of steepest ascent. The two way tables for x_2 and x_3 and for x_2 and x_4 appear in Table 21.

Adopting the following model:

$\hat{y} = b_0x_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_3x_4 + b_6x_2x_4 + b_7x_2x_3$,

the second approximation of the surface was made. The terms x_3x_4, x_2x_4, and x_2x_3 are confounded with x_1x_2, x_1x_3 and x_1x_4, respectively.

The levels of the various factors and their coded form, the design matrix, and results appear in Table 22. The solution of the normal equations yielded the following regression equation:

$\hat{y} = 8.7310 - .0768x_1 - .0140x_2 - .0475x_3 - .0066x_4 - .0163x_3x_4 + .0160x_2x_4 - .0361x_2x_3$.

Unfortunately, an error was made in the original calculation of the regression coefficient for x_1. The initial levels for x_1 were 27° C and 30° C. The calculated regression coefficient b_1 should have been -.0104; however, a positive coefficient was recorded. Consequently, in the second approximation to the surface the levels of x_1 were selected at 30° C and 35° C. These levels of the x_1 variable gave a regression coefficient of -.0768, a negative value that was originally interpreted as meaning that a maximum had been passed for the regression coefficient had changed from positive to negative. The subsequent calculation of the path of steepest ascent corrected for this error, since the increased size of the regression coefficient in the negative

TABLE 21

Biuret-Sulfate Interaction

	$X_2(SO_4)^-$		
X_3(B)	-1	1	
-1	31.5043	35.1099	66.6142
1	33.3952	34.7713	68.1665
	64.8995	69.8812	

Time-Sulfate Interaction

	$X_2(SO_4)^-$		
X_4(time)	-1	1	
-1	33.0139	34.4390	67.4529
1	31.8856	35.4422	67.3278
	64.8995	69.8812	

TABLE 22

Second Approximation of the Surface

Factor	Code -1	Code +1
x_1 (temp.)	30	35
x_2 (SO_4)	9	11
x_3 (B)	6	8
x_4 (time)	5	15

Design Matrix[1]

x_o	x_1	x_2	x_3	x_4	x_{34}	x_{24}	x_{23}	obs.	std.	y (observed)
1	-1	-1	-1	-1	1	1	1	.495	.392	8.8393
1	-1	1	-1	1	-1	1	-1	.407	.320	8.9031
1	-1	-1	1	1	1	-1	-1	.407	.324	8.7716
1	-1	1	1	-1	-1	-1	1	.330	.265	8.7170
1	1	-1	-1	1	-1	-1	1	.508	.410	8.6732
1	1	1	-1	-1	1	-1	-1	.420	.338	8.6982
1	1	-1	1	-1	-1	1	-1	.400	.322	8.6956
1	1	1	1	1	1	1	1	.331	.271	8.5498

[1] Each comparison meausres a pair of effects, the pairs are; $(x_1, x_2x_3x_4)$; $(x_2, x_1x_3x_4)$; $(x_3, x_1x_2x_4)$; $(x_4, x_1x_2x_3)$; (x_2x_3, x_1x_4); (x_1x_2, x_3x_4); (x_2x_4, x_1x_3).

For an analysis of this design see O. L. Davies (1956).

direction increased the size of the increments in the calculation of the path, thus compensating by use of a large increment what would have been brought about originally. This demonstrated, unintentionally, one of the advantages of the method, namely the robustness of the technique. It should be noted that when the error was discovered and corrected, the regression coefficient had indeed increased negatively by the change in coordinates of x_1, as could and should have been expected.

The calculation for the path of steepest ascent and observed values are given in Table 23.

Notice that the signs of b_2 and b_3 have changed from positive to negative, thus indicating that a maximum has been passed. In the case of b_4, the sign has not changed, but the minute size of b_4 (b_4 = -.0066) indicates that we are near a maximum. It seemed reasonable, therefore, to conclude that we were in the near vicinity of a maximum and hence a complete factorial should be planned in order that the experiment, if found to be satisfactory, could be augmented by central and axial observations for the estimation of the regression coefficients for a second-order model.

To take into account the above-mentioned change in sign of the regression coefficients and the error made in b_1, the levels of the various factors were adjusted in an attempt to insure that the optimum would be included in the design space. The code[1], design matrix, X'Y, and B matrices follow in Tables 24, 25 and 26. The analysis of variance is also given in Table 26.

---

[1] Notice that the table includes the coded values of the center and axial points for the augmented design. Here, however, we are only interested in the Columns 1 and -1.

TABLE 23

Second Path of Steepest Ascent

		Temp X_1	SO_4 X_2	B X_3	Time X_4	Predicted $\hat{y}$	Observed y
Relative change in design units		-.0768	-.0140	-.0475	-.0066		
No. of orig. units=1 design unit		2.5	1.0	1.0	5.0		
Relative change in orig. units		-.192	-.014	-.0475	-.0330		
Change per 1°C change in X		1.0	.0729	.2473	.1718		
Initial level		32.5	10.0	7.0	10		
Observations along path	(1)	31.5	9.9271	6.7527	9.8282	+8.77708	9.2920
	(2)	30.5	9.8542	6.5054	9.6564		
	(3)	29.5	9.7813	6.2581	9.4846	8.88602	8.9570
	(4)	28.5	9.7084	6.0108	9.3128		
	(5)	27.5	9.6355	5.7635	9.1410	8.90878	8.8480
	(6)	26.5	9.5626	5.5162	8.9692	8.9469	
	(7)	25.5	9.4897	5.2689	8.7974	8.9898	7.8369
	(8)	24.5	9.4168	5.0216	8.6256	9.0205	
	(9)	23.5	9.3439	4.7743	8.4538	9.0428	
Obs. (Rounded)		X_1	X_2	X_3	X_4	Std.	Obs.
	(1)+	31	10	7	10	.339	.45
	(3)+	29	10	6	10	.372	.476
	(5)+	27	10	6	9	.375	.474
	(7)+	25	9	5	9	.460	.515
	(9)+	23	9	5	8		

Code $X_1 = \frac{(\text{temp}-32.5}{2.5}$ $X_2 = \frac{(NaSO_4-10)}{1}$ $X_3 = \frac{(\text{Biuret}-7)}{1}$ $X_4 = \frac{(\text{time}-10)}{5}$

+ These observations are those indicated by arrows above after rounding.

TABLE 24

Third Approximation of Response Surface

	Coded Values				
Factor	-2	-1	0	1	2
X_1 (temp.)	25	27	29	31	33
X_2 (Na_2SO_4)	7	8	9	10	11
X_3 (Biuret)	4	5	6	7	8
X_4 (time)	1	4	7	10	13

CODE:

$$X_1 = \frac{\text{temp. } -29}{2}$$

$$X_2 = \frac{Na_2SO_4 \ -9}{1}$$

$$X_3 = \frac{\text{Biuret } -6}{1}$$

$$X_4 = \frac{\text{time } -7}{3}$$

TABLE 25

Third Linear Approximation of Response Surface

Experiment numbers	X_1	X_2	X_3	X_4	Observed	Standard	y Observed
1	-1	-1	-1	-1	0.670	0.527	8.8994
2	-1	-1	-1	1	0.660	0.532	8.6842
3	-1	-1	1	-1	0.530	0.410	9.0487
4	-1	1	-1	-1	0.530	0.427	8.6885
5	-1	-1	1	1	0.520	0.415	8.7711
6	-1	1	1	-1	0.419	0.338	8.6775
7	-1	1	1	1	0.423	0.348	8.5086
8	-1	1	-1	1	0.541	0.445	8.5101
9	1	-1	-1	-1	0.720	0.564	8.9362
10	1	-1	-1	1	0.740	0.568	9.1197
11	1	-1	1	-1	0.565	0.446	8.8677
12	1	1	-1	-1	0.575	0.453	8.8852
13	1	-1	1	1	0.550	0.440	8.7500
14	1	1	1	-1	0.450	0.358	8.7989
15	1	1	1	1	0.450	0.351	8.9744
16	1	1	-1	1	0.584	0.470	8.6979

TABLE 26

Third Linear Approximation of Response Surface (Cont.)

$$B = \begin{bmatrix} b_0' \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 8.8011 \\ .0776 \\ -.0834 \\ -.0015 \\ -.0491 \end{bmatrix} \qquad X'Y = \begin{bmatrix} 140.8181 \\ 1.2419 \\ -1.3359 \\ -.0243 \\ -.7861 \end{bmatrix}$$

$$\hat{y} = 8.8011 + 0.0776x_1 - 0.0834x_2 - 0.0075x_3 - 0.0491x_4$$

Analysis of Variance

Source	d.f.	s.s.	m.s.
Total	16	1239.8169	
Mean	1	1239.3585	
Single factors	4	.2464	.0616
Lack-of-fit	11	.2120	.01927
Error[1]	50	----	.0225

[1] Estimated from pervious laboratory findings.

Checking the analysis of variance and the regression coefficient for x_1 (Table 25) and finding it now to be positive rather than negative as it appeared in Table 23, and taking into consideration the above discussion of the variables x_2, x_3 and x_4, it seemed reasonable that the design space did indeed include the desired optimum and that the second-order model

$$y = b_o x_o + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + b_{44} x_4^2 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{14} x_1 x_4 + b_{23} x_2 x_3 + b_{24} x_2 x_4 + b_{34} x_3 x_4 + e.$$

should be adopted.

The code for the augmented design[1] is given in Table 24 and the design matrix for the entire central composite design is given in Table 27 followed by the X'Y and B matrices in Table 28.

Due to the number of observations necessary for the composite design, it was recognized that blocking over days would be desirable. Therefore, following the recommended procedure discussed in the first section of this dissertation, only six observations were made at the origin, rather than the required seven for uniform information, thus allowing for orthogonal blocking. The analysis of variance is given in Table 29 where the blocking has been done over time.

Contour Lines for Total Serum Protein

Since the response is dependent upon four variables, to make a graphical representation of the response surface, it is necessary to hold two of the independent variables fixed and show response levels as the remaining two variables are allowed to change. To this end, x_4

---

[1] This design has only ''nearly'' uniform information due to the number of observations that were made at the origin.

TABLE 27

Central Composite Design Matrix[1] and Results

Observation number	X_0	X_1	X_2	X_3	X_4	X_{11}	X_{22}	X_{33}	X_{44}	X_{12}	X_{13}	X_{14}	X_{23}	X_{24}	X_{34}	Y
1	+	-	-	-	-	+	+	+	+	+	+	+	+	+	+	8.8994
2	+	-	-	-	+	+	+	+	+	+	+	-	+	-	-	8.6842
3	+	-	-	+	-	+	+	+	+	+	-	+	-	+	-	9.0487
4	+	-	+	-	-	+	+	+	+	-	+	+	-	-	+	8.6885
5	+	-	-	+	+	+	+	+	+	+	-	-	-	-	+	8.7711
6	+	-	+	+	-	+	+	+	+	-	-	+	+	-	-	8.6775
7	+	-	+	+	+	+	+	+	+	-	-	-	+	+	+	8.5086
8	+	-	+	-	+	+	+	+	+	-	+	-	-	+	-	8.5101
9	+	+	-	-	-	+	+	+	+	-	-	-	+	+	+	8.9362
10	+	+	-	-	+	+	+	+	+	-	-	+	+	-	-	9.1197
11	+	+	-	+	-	+	+	+	+	-	+	-	-	+	-	8.8677
12	+	+	+	-	-	+	+	+	+	+	-	-	-	-	+	8.8852
13	+	+	-	+	+	+	+	+	+	-	+	+	-	-	+	8.7500
14	+	+	+	+	-	+	+	+	+	+	+	-	+	-	-	8.7989
15	+	+	+	-	+	+	+	+	+	+	-	+	-	+	-	8.6979
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	8.9744
17	+	-2				4										7.8369
18	+	2				4										8.7906
19	+		-2				4									8.5090
20	+		2				4									8.8000
21	+			-2				4								7.7777
22	+			2				4								7.7538
23	+				-2				4							8.9655
24	+				2				4							8.5790
25	+															8.5938
26	+															8.5573
27	+															8.5610
28	+															8.7500
29	+															8.6333
30	+															8.6666

[1] + represents the +1 level of the independent variable, - represents the -1 level of the independent variable, and no entry represents the 0 level.

TABLE 28

Regression Coefficients and X'Y Matrix for Composite Design

X'Y	B		B	$S_{b_i}^2$
2595926	b_0		8.6280461	-
31493	b_1		.13122083	.00382404
-7539	b_2		-.03141250	.00382404
-721	b_3		-.00300416	.00382404
-15591	b_4	=	-.06496250	.00382404
2073281	b_{11}		-.00787813	.00329824
2100541	b_{22}		.07730937	.00329824
2029441	b_{33}		-.14487813	.00329824
2109961	b_{44}		.10674687	.00329824
7015	b_{12}		.04384375	.00573606
-4717	b_{13}		-.02948125	.00573606
8941	b_{14}		.05588125	.00573606
3797	b_{23}		.02373125	.00573606
679	b_{24}		.00424375	.00573606
87	b_{34}		.00054375	.00573606

TABLE 29

Analysis of Variance for Central Composite Rotatable Design with Orthogonal Blocking

Source	d.f.	s.s.	m.s.
Mean	1	2239.7769	
Regression	14	8.3692	
Residual	15	1.3768	0.0918
SS blocks	(2)	0.8588	0.4294
SS_e	(3)	0.0192	0.0064
$R\text{-}SS_b - SS_e$	(10)	0.4988	0.0499
Total	30	2249.5229	

was held at (+1, 0, -1) and then nine levels of x_3 were selected, that is, x_3 = (-2.0, -1.5, -1.0, -.5, 0, +.5, +1.0, +1.5, +2.0). The IBM 650 program was limited to a maximum of nine levels or values for the dependent variable. The values selected and the increment are given along with the contour lines in Figures 10, 11, 12, 13, 14 and 15.

In order to study further the fitted surface it is often convenient to reduce it to the canonical form. This reduction is accomplished by the shifting of the origin to the optimum, and rotating the coordinate axes so that they correspond to the axes of the original fitted regression equation. The general regression equation is then reduced to the following form:

$$\hat{y} - y_o = \lambda_1 x_1^2 + \lambda_2 x_2^2 \ldots + \lambda_k x_k^2 .$$

Examination of the above equation shows that large values of are indicative of rapid changes in the response, and that small values of λ are indicative of slow changes in the response. If the λ's are all negative, the fitted contour surfaces are ellipsoids and the response surface has a true maximum.

If one or more of the λ's are positive the contour surfaces are elliptic hyperboloids and there is a minimum. If one or more of the λ's approach zero the surface is attenuated along these axes and the surface is an elliptic or hyperbolic cylinder. In such cases the response surface forms a ridge.

To obtain the coordinates of the center or maximum, take the partial derivative of the regression equation given in Table 29 with respect to x_1, x_2, x_3 and x_4 respectively. Setting these partial derivatives equal to zero one obtains the following system of equations:

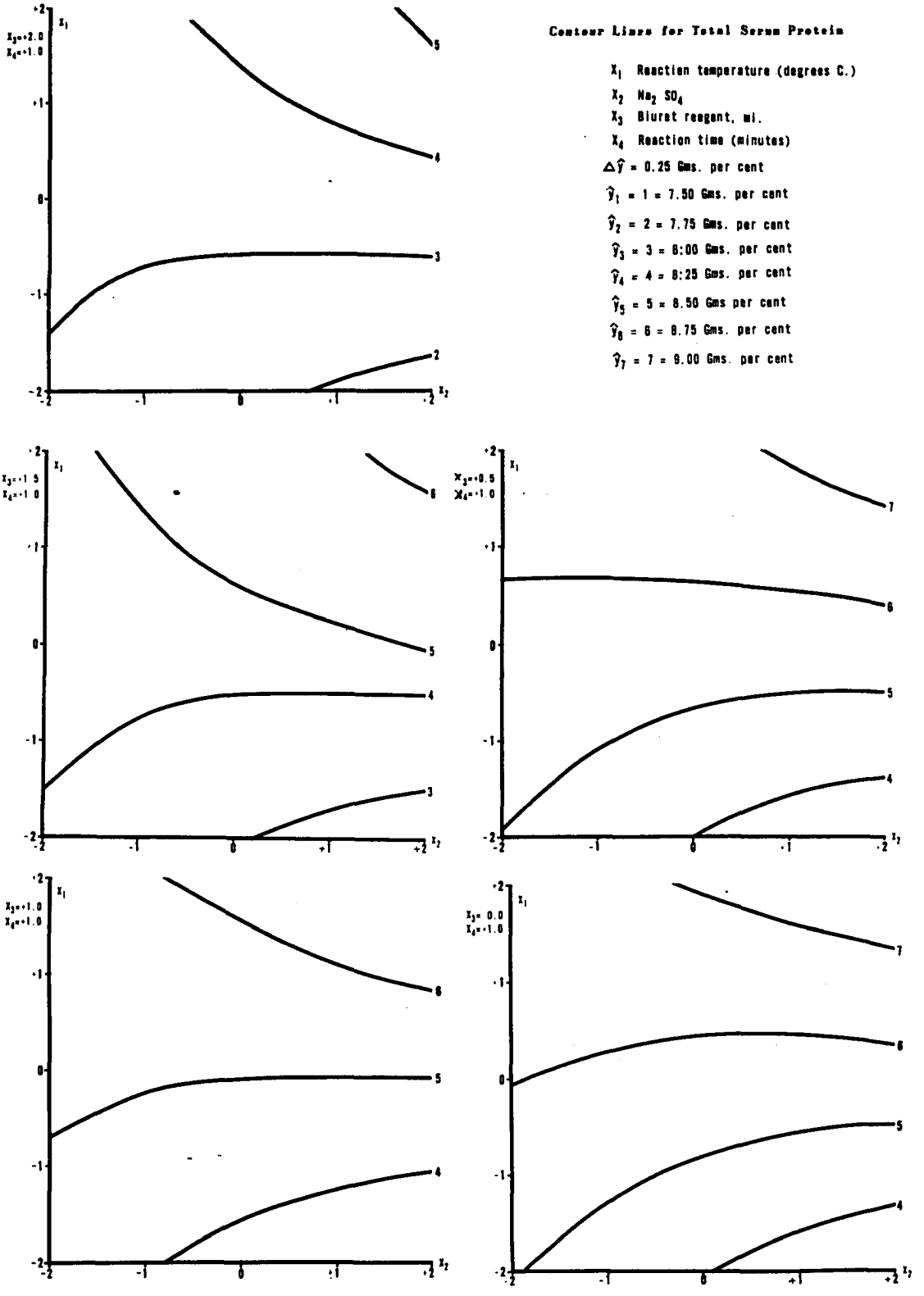

Contour Lines for Total Serum Protein
X_1 Reaction temperature (degrees C.)
X_2 Na_2 SO_4
X_3 Biuret reagent, ml.
X_4 Reaction time (minutes)
$\Delta\hat{y}$ = 0.25 Gms. per cent
$\hat{y}_1$ = 1 = 7.50 Gms. per cent
$\hat{y}_2$ = 2 = 7.75 Gms. per cent
$\hat{y}_3$ = 3 = 8:00 Gms. per cent
$\hat{y}_4$ = 4 = 8:25 Gms. per cent
$\hat{y}_5$ = 5 = 8.50 Gms per cent
$\hat{y}_6$ = 6 = 8.75 Gms. per cent
$\hat{y}_7$ = 7 = 9.00 Gms. per cent
X_3=+2.0
X_4=+1.0
X_3=+1.5
X_4=+1.0
X_3=+0.5
X_4=+1.0
X_3=+1.0
X_4=+1.0
X_3= 0.0
X_4=+1.0

Figure 10.

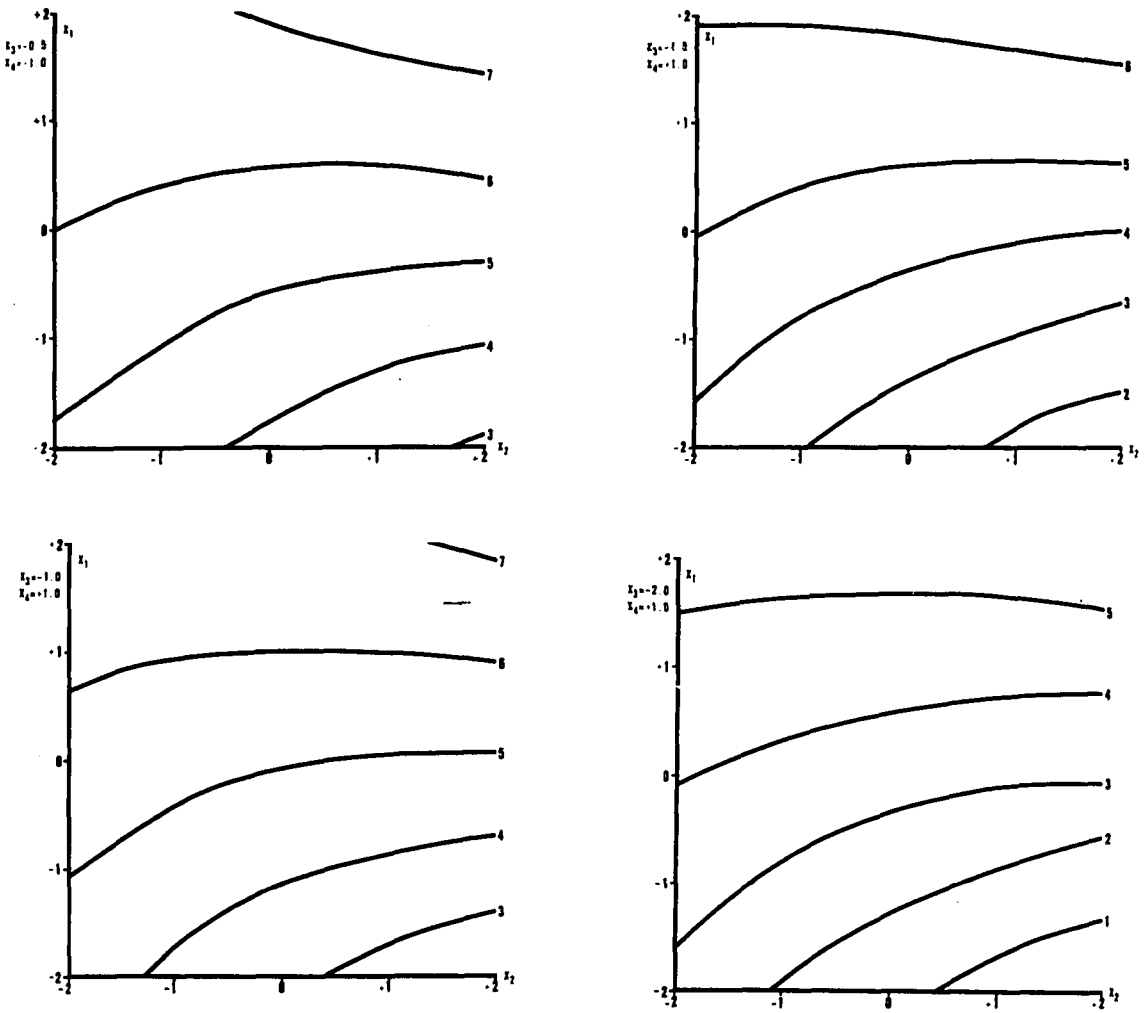


Figure 11.

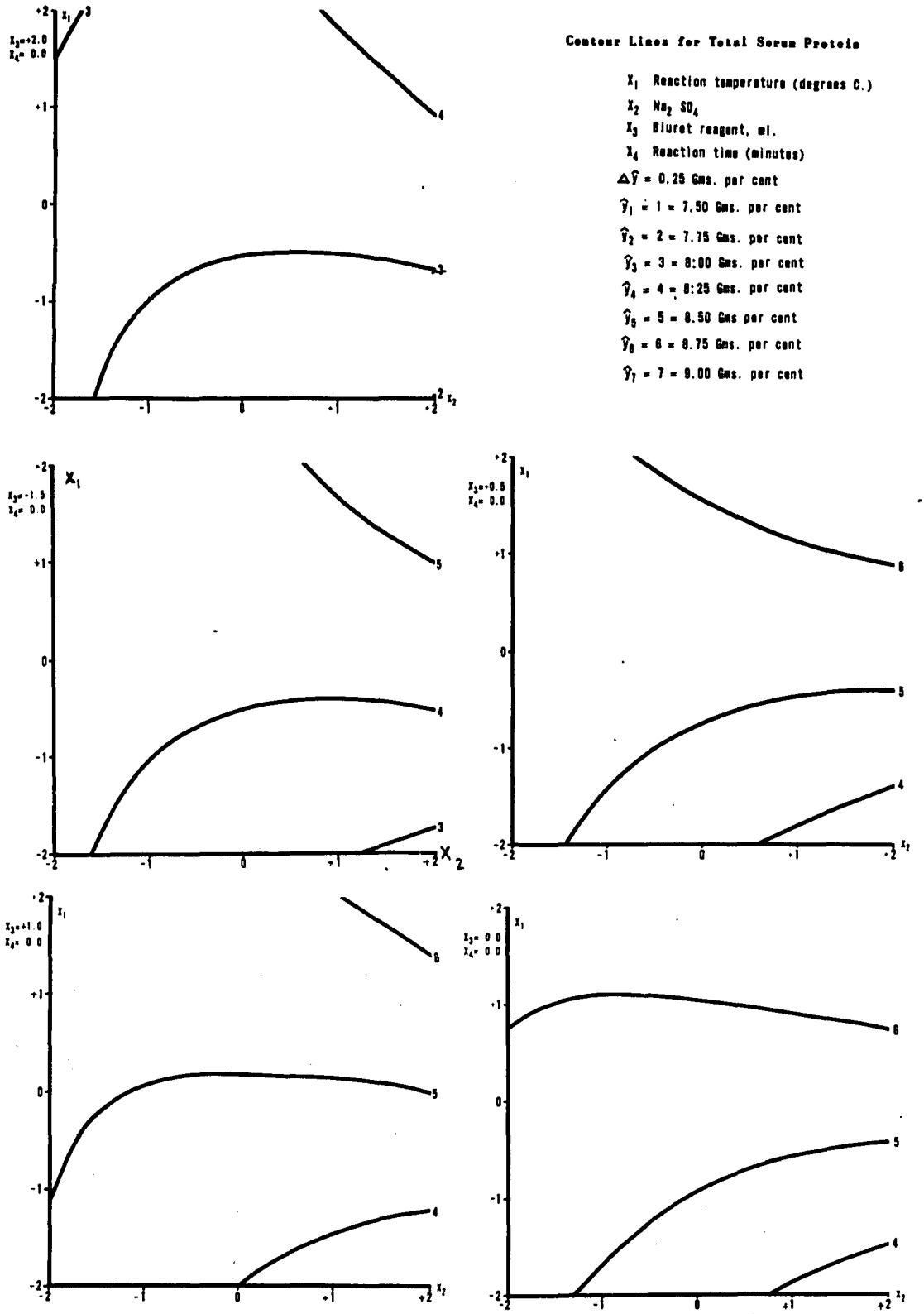

Contour Lines for Total Serum Protein
X_1 Reaction temperature (degrees C.)
X_2 Na_2SO_4
X_3 Biuret reagent, ml.
X_4 Reaction time (minutes)
$\Delta\hat{Y}$ = 0.25 Gms. per cent
$\hat{Y}_1$ = 1 = 7.50 Gms. per cent
$\hat{Y}_2$ = 2 = 7.75 Gms. per cent
$\hat{Y}_3$ = 3 = 8.00 Gms. per cent
$\hat{Y}_4$ = 4 = 8.25 Gms. per cent
$\hat{Y}_5$ = 5 = 8.50 Gms per cent
$\hat{Y}_6$ = 6 = 8.75 Gms. per cent
$\hat{Y}_7$ = 7 = 9.00 Gms. per cent
X_3=+2.0 X_4= 0.0
X_3=+1.5 X_4= 0.0
X_3=+0.5 X_4= 0.0
X_3=+1.0 X_4= 0.0
X_3= 0.0 X_4= 0.0

Figure 12.

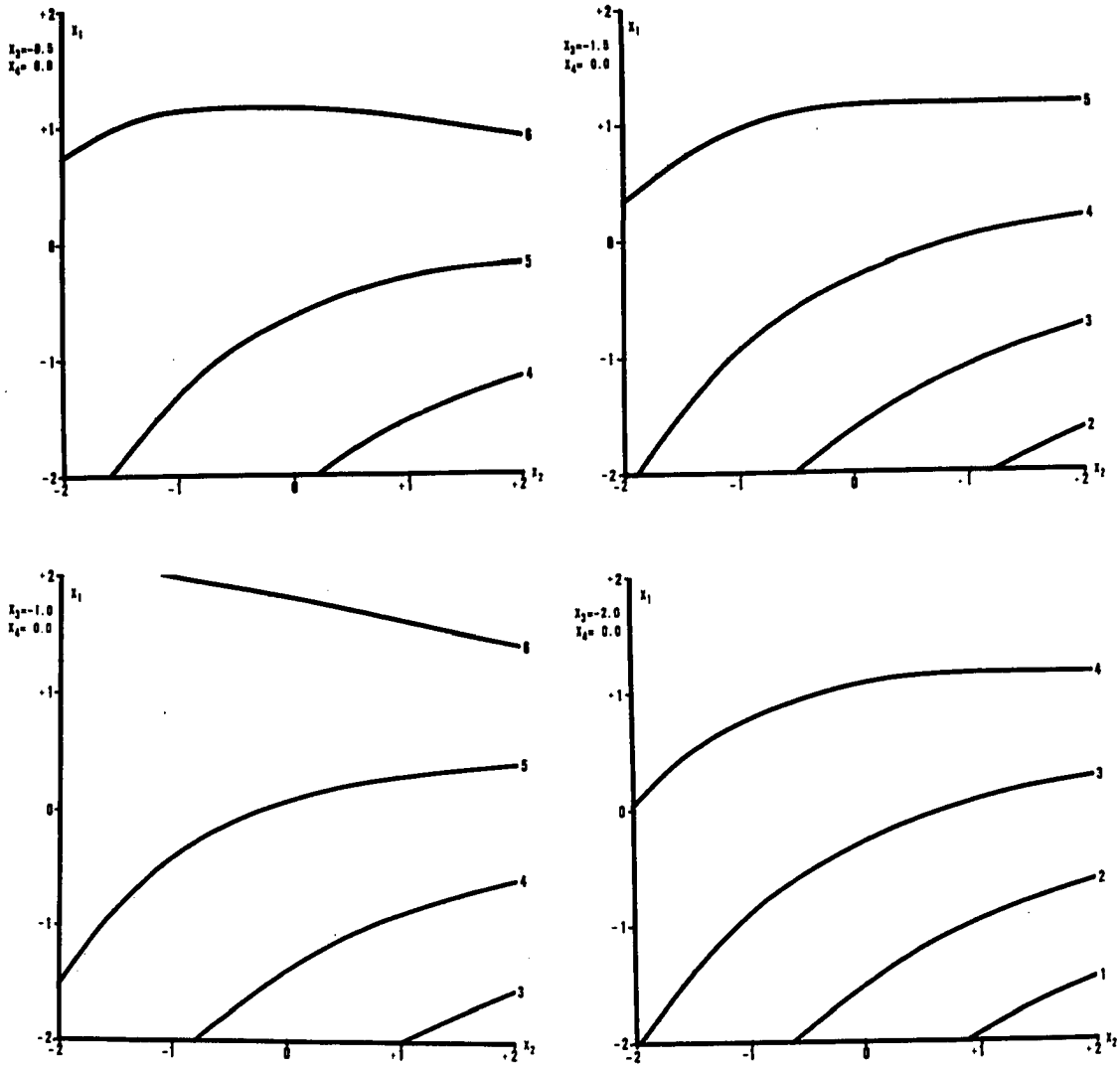
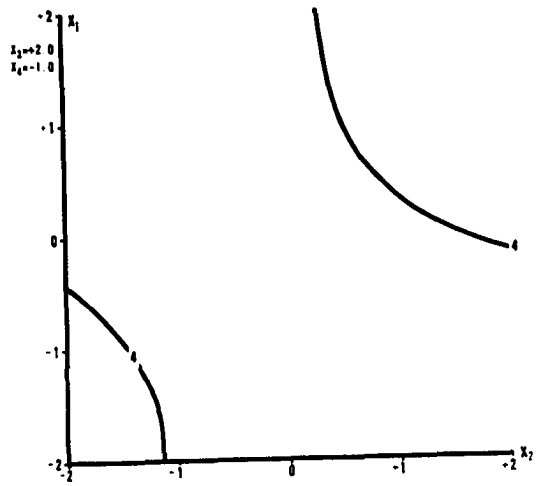
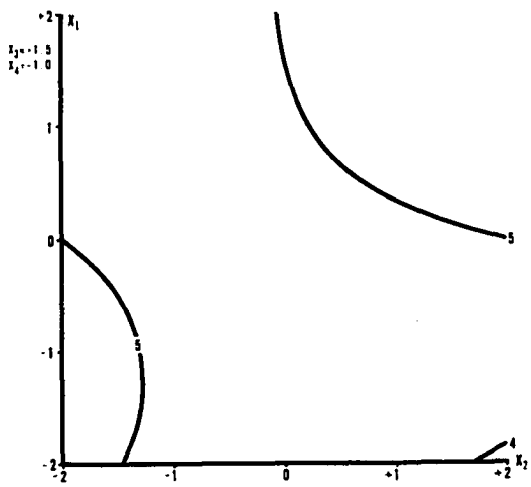


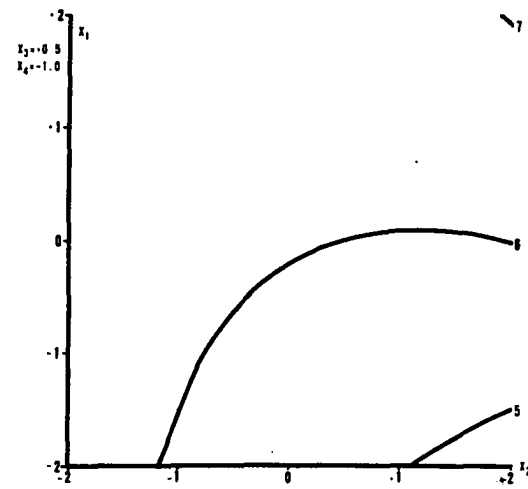
Figure 13.

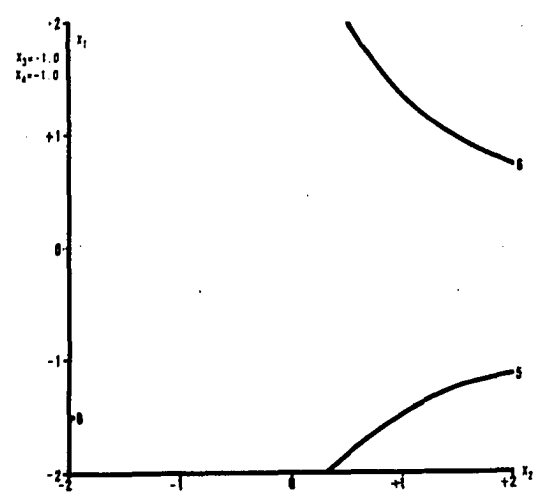


Contour Lines for Total Serum Protein

X_1 Reaction temperature (degrees C.)
X_2 Na_2SO_4
X_3 Biuret reagent, ml.
X_4 Reaction time (minutes)
$\Delta\hat{Y}$ = 0.25 Gms. per cent
$\hat{Y}_1$ = 1 = 7.50 Gms. per cent
$\hat{Y}_2$ = 2 = 7.75 Gms. per cent
$\hat{Y}_3$ = 3 = 8:00 Gms. per cent
$\hat{Y}_4$ = 4 = 8:25 Gms. per cent
$\hat{Y}_5$ = 5 = 8.50 Gms per cent
$\hat{Y}_6$ = 6 = 8.75 Gms. per cent
$\hat{Y}_7$ = 7 = 9.00 Gms per cent







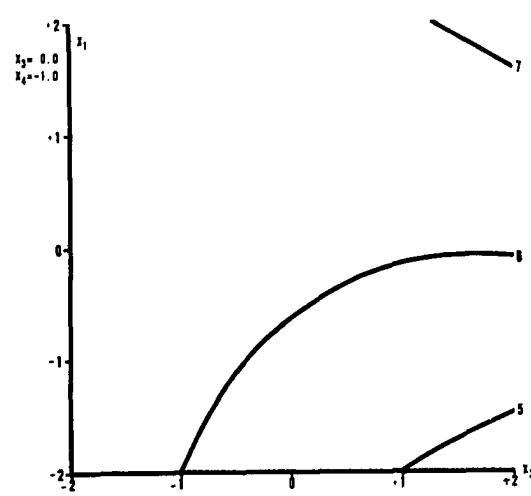


Figure 14.

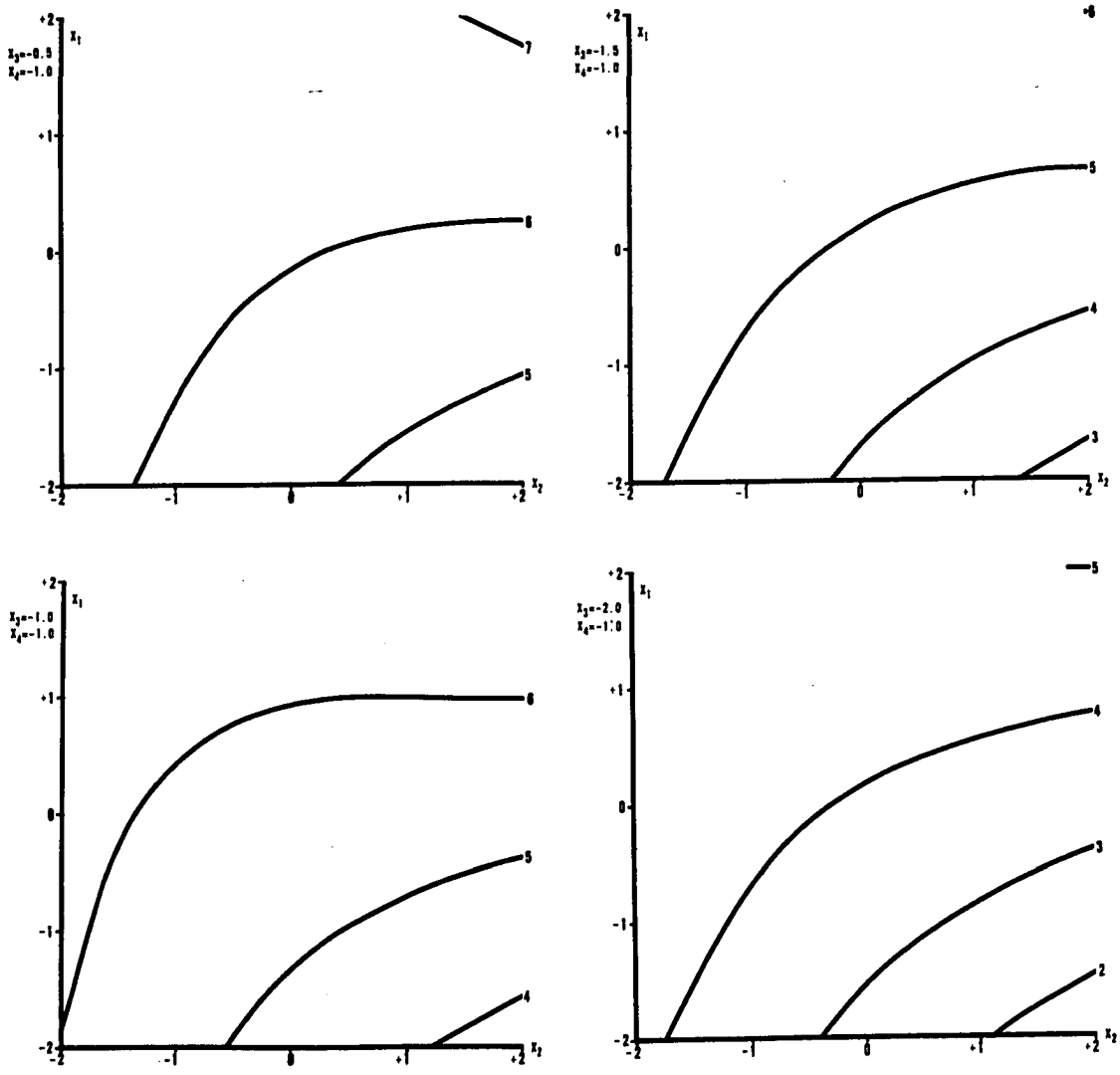


Figure 15.

$$0 = .13122083 - .01575626x_1 + .04384375x_2 - .02948125x_3 + .05588125x_4$$

$$0 = .03141250 + .04384375x_1 + .15461874x_2 + .02373125x_3 + .00424375x_4$$

$$0 = -.00300416 - .02948125x_1 + .02373125x_2 - .28975626x_3 + .00054375x_4$$

$$0 = -.06496250 + .05588125x_1 + .00424375x_2 + .00054375x_3 + .21349374x_4$$

The solution of these equations give:

$$x_1 = 4.1426$$

$$x_2 = -0.8732$$

$$x_3 = -0.5048$$

$$x_4 = -0.7614$$

Substituting these values into the fitted second-order regression equation the predicted value for the maximum is found to be $\hat{y}_0 = 8.8914$.

For a theoretical discussion of the reduction of the above regression equation to this canonical form one is referred to Birkhoff and MacLane (1953). A program now available for the IBM 650 digital computer for solving secular equations was used here, however. The discriminating quartic is:

$$H = \begin{vmatrix} b_{11} - \lambda & \frac{b_{12}}{2} & \frac{b_{13}}{2} & \frac{b_{14}}{2} \\ b_{12} & b_{22} - \lambda & \frac{b_{23}}{2} & \frac{b_{24}}{2} \\ \frac{b_{13}}{2} & \frac{b_{23}}{2} & b_{33} - \lambda & \frac{b_{34}}{2} \\ \frac{b_{14}}{2} & \frac{b_{24}}{2} & \frac{b_{34}}{2} & b_{44} - \lambda \end{vmatrix} = 0$$

which has the roots

$$\lambda_1 = -.016718$$

$$\lambda_2 = .080768$$

$$\lambda_3 = -.147382$$

$$\lambda_4 = .114635.$$

By substituting these values and the predicted value at the center into the canonical form above, one obtains

$$\hat{y} - 8.8914 = -.016718x_1^2 + .080768x_2^2 - .147382x_3^2 + .114635x_4^2 .$$

Following the recommended analysis outlined by Box and Wilson (1951), except using the IBM program mentioned above, the following orthogonal transformation for the old variable to the new was found to be

$$Q' = \begin{bmatrix} .0948192519 & -.0200151948 & -.0128069775 & -.0210875010 \\ .0153260296 & .0956904206 & .0039994488 & -.0243405137 \\ .0117416065 & -.0063647127 & .0990949699 & -.0013460907 \\ .0252300792 & .0200555674 & -.0004154501 & .0946627943 \end{bmatrix}$$

where, H^D [1] $= Q' HQ$.

From the smallness of λ_1 compared to the other coefficients, it can be shown that the surface is attenuated along the x_1 axis, that is, there is a ridge running in this direction. The fitted response surfaces are hyperboloids of one sheet. Thus the sections by the plane $x_2 = 0$ and $x_4 = 0$ are hyperboloids and those by $x_1 = 0$ and $x_3 = 0$, ellipsoids.

This ridge is indicative of a continuing change of the response due to variations in the temperature. However, careful examination of the response surface reveals that the surface is quite flat and that

---

[1] H^D is the diagonalized form of the regression equation; that is, the diagonal elements represent the coefficients of the canonical form of the second-order regression equation.

these variations are not large. The remaining independent variables (amount of Na_2SO_4, amount of Biuret and time) are all more influential on the response.

CHAPTER V

SUMMARY

The purpose of this dissertation has been to demonstrate the applicability of the method of steepest ascent to medical data and to illustrate the subsequent characterization of the response surface in the optimization problem as it appears in certain fields of medical research. To this end, two specific examples were introduced and modifications in the method of steepest ascent adopted to illustrate empirically the feasibility of the method. These modifications have been discussed and the optimal conditions found for the two processes. Specifically, a study was undertaken to determine the optimum conditions for, and interrelations of, temperature, pH and concentration of electron donor on the rate of nitrate reduction by the nitrate-nitrite reductase systems in a pooled sample of salivary sediment. By varying the conditions of pH, temperature and electron donor concentration (yeast extract) the minimum reduction time was established in four sequential sets of experiments. The method employed was that of steepest ascent which consisted of two phases; first, the establishment of the response surface and the conditions for the optimum, and, second, the calculation of the associated contour lines. Three linear approximations of the response surface and their subsequent paths of

steepest ascent were utilized in locating the near-optimal region. A fourth, and final, linear approximation was augmented by additional experimental points such that the property of rotatability was preserved and a second-order approximation of the response surface made possible. This method permitted the calculation of the response surface from a much smaller number of experiments than would have been necessary with a complete factorial experiment (or by the single-factor method). It also provided the opportunity for the recognition of possible interactions among the variables. While the three variables examined were found to influence the enzyme activity, temperature contributed the predominating influence above pH 6.0 and electron donor concentration of 0.7 g/ml. Through the utilization of the determined response surface, the estimated optimal conditions were found to be ; pH 6.98, temperature, 45.9° C, and electron donor concentration, yeast extract, 0.97 g/ml. The basis for the construction of the design for the study was discussed and the details of the statistical analysis presented.

The second application was concerned with the optimization and interrelations of temperature, amount of Na_2SO_4, amount of Biuret reagent and reaction time on the determination of total serum protein. Following the procedure described above, the estimated optimal conditions were found to be

temp. (x_1) = 37.2852° C

Na_2SO_4 (x_2) = 8.1268 c.c.

Biuret (x_3) = 5.4952 c.c.

time (x_4) = 0.2842 min.

The second-order approximation of the response surface was determined, the subsequent contour lines calculated, and the canonical form of the response surface found.

It appears to the writer that the researcher in biological phenomena will frequently be unable to control the independent variables as precisely as might be desired. To further compound this problem one often observes large variation in the dependent variable, thus making it doubly easy for the ''masking'' effect to occur in the application of the technique described. To compensate for these negative effects, one may replicate experimental observations and in some cases increase the independent variable spacing intervals. If the latter measure is adopted, however, the consequent lack-of-fit term in the linear approximation of the response surface might be significant even though it might not have been if the independent variable intervals had been smaller. This writer's experience shows, however, that the path can be calculated and used as a guide for progression toward the maximum, even though the expected values along the path of steepest ascent are unrealistic.

The method of steepest ascent and the subsequent characterization of the response surface seem to offer an excellent opportunity for solving optimization problems in several areas, in particular, bacteriology, biochemistry, enzymology, physiology, pharmacology, radiology, and virology. The restrictions, of course, revolve around the identification of, the number of, the ability to control, and the continuity of the independent variables.

Since the only time an experiment can be properly designed is after the experiment has been performed, one might suspect that the mathematical sophistication for the application of the method of steepest ascent, indeed, any experimental technique, depends upon the experience and knowledge of the researcher regarding the statistical tool itself and the functional phenomena being investigated. If the design is unsatisfactory, it may be frequently attributed to one of the following:

1. One or more variables have been neglected.
2. A poor (or at least a less favorable) choice of code for the variables has been made.
3. Improper ranges for the variables have been selected.
4. A basic error in selection of the design has been made.

The above-mentioned illustrations of the application of the technique have not followed Box's outlined procedure unremittingly but have captured the intent and the general concepts which Box and his co-workers have proposed. Certainly the method of steepest ascent is not a purely mathematical concept; but, just as certainly, it is a methodology for solving a particular type of problem. The method has been shown to be reasonably flexible and successful.

REFERENCES CITED

Anderson, R. L. 1953 Recent Advances in Finding Best Operating Conditions. Journal American Statistical Association, 48: 789-798.

Birkoff, G. and MacLane, S. 1953 A Survey of Modern Algebra. Macmillan Company, N. Y. pp. 248-251.

Bose, R. C. and Draper, N. 1959 Rotatable Designs of Second and Third Order in Three or More Dimensions. Institute of Statistics Mimeo Series, No. 197.

Box, G. E. P. 1952a Multi-factor Designs of First Order. Biometrika, 39: 49-57.

_____ 1952b The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples Biometrics, 10: 16-60.

_____ and Behnken, D. W. 1959 Derivation of Second Order Rotatable Designs from those of First Order. Statistical Techniques Research Group (Princeton University), Technical Report No. 17.

_____ and Draper, N. L 1959 A Basis for the Selection of a Response Surface Design. Statistical Techniques Research Group (Princeton University), Technical Report No. 23. Journal of the American Statistical Association, 54: 622-654.

_____ and Hunter, J. S. 1954 A confidence Region for the Solution of a Set of Stimultaneous Equations With an Application to Experimental Design. Biometrika, 41: 190-199.

_____ and _____ 1957 Multi-factor Experimental Designs for Exploring Response Surfaces, Annals of Mathematical Statistics. 28: 195-241.

_____ and Lucas, H. L 1959 Design of Experiments in Non-Linear Situations. Statistical Techniques Research Group (Princeton University), Technical Report No. 15 Biometrika, 46: Parts 1 and 2.

_____ and Wilson, K. B. 1951 On the Experimental Attainment of Optimum Conditions. Journal Royal Statistical Society, Series B, 13: 1-45.

_____ and Youle, P. W. 1955 The Exploration and Exploitation of Response Surfaces. An Example of the Link Between the Fitted Surface and the Basic Mechanism of the System. Biometrics, 11: 287-323.

Brooks, S. H. 1959 A Comparison of Maximum-Seeking Methods. Operations Research, 7: 430-457.

Budne, T. A. 1959 The Application of Random Balance Designs. Technometrics, 1: No. 2, 139-155.

Chew, Victor (ed) 1958 Experimental Designs in Industry. A Symposium Held at North Carolina State College, Nov. 1956. Experimental Designs for Exploring Response Surfaces, by Box, G. E. P. and Hunter, J. S. John Wiley and Sons, Inc., New York.

Cochran, W. G. and Cox, G. M. 1957 Experimental Designs Second Edition. Some Methods for the Study of Response Surfaces. John Wiley and Sons, Inc., New York. Chapter 8A.

Davies, O. L. (ed) 1956 Design and Analysis of Industrial Experiments. The Determination of Optimum Conditions. Hafner Publishing Co. New York. Chapter 11.

_____ and Hay, W. A. 1950 The Construction and Use of Fractional Factorial Designs in Industrial Research Biometrics, 6: 233-249.

David, H. A. and Arens, B. E. 1949 Optimal Spacing in Regression Analysis. Annals of Mathematical Statistics., 30: 1072-1081.

DeBaun, R. M. 1956 Block Effects in the Determination of Optimum Conditions. Biometrics, 12: 20-22.

_____ 1959 Response Surface Designs for Three Factors at Three Levels. Technometircs, 1: No. 1 1-8.

de la Garza, A. 1954 Spacing of Information in Ploynomial Regression. Annals of Mathematical Statistics, 25: 123-130.

Dykstra, O. 1960 Partial Duplication of Response Surface Designs. Technometrics, 2: No. 2, 185-195.

Ferro, Patrick, Ham and Ana Bell, 1961 Commerical Report for Improved Biuret Method for Total Protein and AG Ratio, Dade Reagent Inc. Miami, Florida.

Finney, D. J. 1945 The Fractional Replication of Factorial Experiments. Annals of Eugenics, 12: 291-301.

_____ 1946 Recent Developments in the Design of Field Experiments. Fractional Replication. Jounal of Agricultural Science, 36: 184-191.

Flosdorf, E. W. 1949 Freeze-Drying (Drying by Sublimation) Reinhold Publishing Corporation, New York.

Friedman, M. and Savage, L. J. 1947 Planning Experiments Seeking Maxima. In Eisenhar, Hastay and Wallis. Selected Techniques of Statistical Analysis. McGraw-Hill Book Co., New York. pp. 363-372.

Fisher, R. A. 1949 The Design of Experiments. Oliver and Boyd London.

Gardiner, D. A., Grandage, A. H. E. and Hader, R. J. 1959 Some Third Order Rotatable Designs. Mimeo Series No. 14 (April, 1956). Institute of Statistics (Raleigh, N. C.) Annals of Mathematical Statistics., 30: 1082-1096.

Goaz, P. W. and Biswell, H. A. 1961 Nitrate Reduction in Whole Saliva. Journal of Dental Research, 40: 355-365.

Graybill, F. A. 1961 An Introduction to Linear Statistical Models. McGraw-Hill Book Co., New York. Chapter 6: 149-164.

Hader, R. J., Hayward, M. E., Mason, D. D. and Moore, D. P. 1957 An Investigation of Some of the Relationships Between Copper, Iron, and Molybdenum in the Growth and Nutrition of Lettuce: Part I. Experimental Design and Statistical Methods for Characterizing the Response Surface. Soil Science Society of America. Proceedings. 21: No. 1, 65-74.

Hawk, P. B., Oser, B. L., and Summerson, W. H. 1951 The Practical Physiological Chemistry. Twlfth Edition. The Blakiston Company, New York.

Kaplan, W. 1957 a Advanced Calculus. Addison-Wesley. Reading, Mass. pp. 79-82.

_____ 1957 b Advanced Calculus. Addison-Wesley. Reading, Mass. pp. 120-132.

Kempthorne, O. 1947 A Simple Approach to Confounding and Fractional Replication in Factorial Experiments. Biometrika, 34: 255-72.

Kiefer, J. and Wolfowitz, J. 1952 Stochastic Estimation of A Maximum of a Regression Function. Annals of Mathematical Statistics., 23: 462-466.

Nielsen, K. L. 1957 Methods in Numerical Analysis. Macmillan Company. New York. Chapter VIII.

Piper, W. W. 1959 A Program for the Solution of Secular Equations. General Electric, Schenectady, New York. P. O. 1088. (Personal Communication).

Plackett, R. L. and Burnam, J. P. 1946 The Design of Optimum Multifactor Experiments. Biometrika, 33: 305-25.

Rao, C. R. 1947 Factorial Experiments Derivable from Combinatorial Arrangements of Arrays. Journal of the Royal Statistical Society Supplement, 9: 288-

Satterthwaite, J. E. 1959 Random Balance Experimentation Technometrics, 1: No. 2, 111-137.

Snedecor, G. W. 1956 Statistical Methods. Fifth Edition. The Iowa State College Press, Ames, Iowa, Chapter 13.

Yates, F. 1937 Design and Analysis of Factorial Experiments. Imperial Bureau of Soil Science Technical Communication, No. 35.