INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeaib Road
Ann Arbor, Michigan 48106
EBELING, Charles Evan, 1942-  
ON THE CONSTRUCTION OF OPTIMAL REGRESSION DESIGNS.  
The Ohio State University, Ph.D., 1973  
Statistics  

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.
ON THE CONSTRUCTION OF
OPTIMAL REGRESSION DESIGNS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Charles Evan Ebeling, B.S., M.S.

* * * * *

The Ohio State University

Reading Committee:

Prof. J. B. Neuhardt
Prof. A. B. Bishop
Prof. V. H. Bond
Prof. R. W. Swain

Approved By

Adviser
Department of Industrial and Systems Engineering
ACKNOWLEDGEMENTS

I wish to thank my adviser, Professor John Neuhardt, for his help and sound advice, my reading committee members for their many excellent comments and suggestions, and my typist, Mrs. Bobbi Elwell, for her infinite patience. A special thanks goes to my wife, Pat, for her encouragement and sacrifices, and to my daughter, Michelle, who was born in the middle of this effort.
VITA

December 18, 1942 .......... Born – Pittsburgh, Pennsylvania

1965 ........................ B.S., Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania

1965-1969 ................... Communications Operations Officer, United States Air Force

1970 ........................ M.S., Logistics Management, Air Force Institute of Technology, Dayton, Ohio

1971-1972 ................... Student, Department of Industrial Engineering, Ohio State University, Columbus, Ohio

1972-1973 ................... Instructor, Air Force Institute of Technology, Dayton, Ohio

FIELDS OF STUDY

Major Field: Operations Research


# TABLE OF CONTENTS

## LIST OF TABLES

vi

## LIST OF ILLUSTRATIONS

vii

## Chapter

### I. INTRODUCTION

1

- Research Statement
- The Linear Statistical Model
- The Optimum Experimental Design Problem
- Design Criteria - Literature Review
- Problem Formulation - Overview

### II. A RELATED PROBLEM

18

- Concavity of P2
- Kuhn-Tucker Theory
- An Equivalent Problem
- An Additional Result

### III. AN ALGORITHM

34

- Some Considerations For An Algorithm
- Derivation of the Algorithm

### IV. THE GRID PROBLEM

44

- Choosing A Grid
- A Measure of Accuracy
- Improving a Solution
- The Procedure for Improving A Solution
- Convergence to the D-Optimum
- Summary of Procedure
Chapter

V. APPLICATIONS ...................................... 58
   The Single Variable Problem
   Improving A Solution
   The Multi-Variable Problem
   Determining A Grid Size

VI. CONCLUSIONS AND EXTENSIONS ............... 80
   Some Extensions
   Areas For Further Research

APPENDICES .......................................... 86
   Appendix A - Weighted Least Square Models
   Appendix B - The Augmentation Problem
   Appendix C - The Model Discrimination
   Problem
   Appendix D - Glossary

BIBLIOGRAPHY ...................................... 105
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Single Variable Regression Models</td>
<td>60</td>
</tr>
<tr>
<td>2.</td>
<td>Optimum Regression Designs</td>
<td>61</td>
</tr>
<tr>
<td>3.</td>
<td>Standard Designs</td>
<td>63</td>
</tr>
<tr>
<td>4.</td>
<td>Solutions to Irregular Experimental Regions</td>
<td>68</td>
</tr>
<tr>
<td>5.</td>
<td>Regression Models for Grid Size Comparisons</td>
<td>76</td>
</tr>
<tr>
<td>6.</td>
<td>Determinantal Values for Four Different Grid Sizes</td>
<td>78</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.</td>
<td>Irregular Experimental Regions</td>
<td>67</td>
</tr>
<tr>
<td>2.</td>
<td>An Experimental Region</td>
<td>70</td>
</tr>
<tr>
<td>3.</td>
<td>The Unit Square</td>
<td>74</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Experimental design is concerned with the planning of an experiment. More specifically, an experimental design describes the manner in which the observations are to be taken.\(^1\) Since early in the 1930's, when R. A. Fisher began the development of the analysis of variance technique, experimental design has been primarily concerned with developing, analyzing, and applying a standard set of designs to various experimental situations. Designs such as random blocks, Latin squares, factorials, fractional factorials, rotatable designs, etc., have found wide acceptance and have proved to be quite useful. These designs are easily understood, intuitively appealing, and relatively simple to analyze. However, in this traditional framework of standard designs, it is difficult and many times impossible to compare two or more designs and to state which is "best" for a particular situation.

Twenty years ago, Kempthorne in his book *The Design and Analysis of Experiments* stated,

\(^1\)In this dissertation, experimental design refers to regression analysis as well as qualitative analysis of variance models. A clear distinction between the two will be made later.
The purpose of the theory of the design of experiments is to insure that the experimenter obtains data relevant to his hypothesis in as economical way as possible. (31)

Using a standard textbook design, an experimenter is not able to explicitly consider the economics of his experiment.

Independent of the development of these standard designs and the analysis of variance technique, there has appeared in the literature a considerable amount of theory concerning optimal experimental designs. This has been an attempt to correct the deficiencies pointed out above of choosing somewhat arbitrarily a standard design. Instead various optimality criteria have been proposed for comparing designs, and whenever possible, using these criteria, optimal designs have been determined. Unfortunately, much of the theory of optimal experimental designs is developed in difficult mathematical language and is not readily available for use by the experimenter whose expertise is in unrelated fields. In addition, the application of this theory is in most cases extremely difficult. This is borne out by the relatively few situations in which it has been possible to completely characterize optimal designs.

From the above discussion, it can be seen that experimental design has followed two somewhat separate paths in its development. On one hand, a set of textbook designs are available which have some advantages but are not
necessarily optimal, nor do they allow for explicit economic considerations. On the other hand, the theory of optimal designs corrects these disadvantages but is difficult to apply. Because of the increasing importance of the economics of experimentation in a world that is becoming more and more aware of its resource limitations, it is essential that techniques be developed to apply this theory in as general a setting as possible. This dissertation is one approach to meeting this need.

Research Statement

The purpose of this dissertation is to develop a methodology along with the supporting theory that will determine an optimum experimental design for a wide variety of experimental design models. In many cases this methodology will determine the true optimum design, while more generally, an approximately optimum design can be found within a desired degree of accuracy. The main advantage of this approach is its generality and ease of application.

The Linear Statistical Model

In order to motivate an optimum experimental design criterion and to introduce the necessary mathematical notation, a brief discussion of the linear statistical model is needed. The linear model is the basic mathematical structure for experimental designs, and it will
provide the general framework for the methodology to be developed.

Consider the following functional relationship between a single dependent variable \( Y \) and one or more independent variables \( (x_1, x_2, \ldots, x_k) \):

\[
(1) \quad Y = B_0 + B_1 f_1(X) + B_2 f_2(X) + \ldots + B_q f_q(X) + e
\]

where \( X^t = (x_1, x_2, \ldots, x_k) \).

In the above model \( B_j, j = 0, 1, \ldots, q \) represent unknown parameters which are to be estimated, and \( e \) is a random variable. The functions \( f_j(X) \) are linearly independent and are known.

If \( n \) experiments are conducted in which observations are made on the dependent variable for various levels or values of the independent variables, the results can be summarized in the following matrix notation:

\[
(2) \quad \begin{bmatrix}
Y_1 \\
\vdots \\
Y_n
\end{bmatrix}
= \begin{bmatrix}
1 & f_1(X_1) & \ldots & f_q(X_1) \\
1 & f_1(X_2) & \ldots & f_q(X_2) \\
\vdots & \vdots & \ddots & \vdots \\
1 & f_1(X_n) & \ldots & f_q(X_n)
\end{bmatrix}
\begin{bmatrix}
B_0 \\
B_1 \\
\vdots \\
B_q
\end{bmatrix}
+ \begin{bmatrix}
e_1 \\
e_2 \\
\vdots \\
e_n
\end{bmatrix}
\]

or in vector notation, \( Y = \bar{X}B + e \).

In this model \( Y_i, i = 1, 2, \ldots, n \), is the \( i \)th observation and \( X_i^t = (x_{i1}, x_{i2}, \ldots, x_{ik}) \), \( i = 1, 2, \ldots, n \), is a vector
of levels for the independent variables in the $i^{th}$ experiment. The vector $X_i^t$ will be called the $i^{th}$ experiment point. The matrix $X$ will be referred to as the design matrix. Minimum assumptions on the random vector $e$ are:

$$E(e) = 0$$

$$\text{Cov} (e) = \sigma^2 V$$

where $|V| \neq 0$

and $\sigma^2 V$ is a known covariance matrix except for the constant $\sigma^2$. In many situations it will be assumed that the $e_i$ are uncorrelated and have equal variances. That is,

$$\text{Cov} (e) = \sigma^2 I.$$

It is well known that the best (unbiased-minimum variance) linear estimate for $B, \hat{B}$, is given by the solution to the following set of normal equations:

$$(3) \quad X^t V^{-1} X \hat{B} = X^t V^{-1} Y$$

If $X^t V^{-1} X$ is of full rank then a unique solution to (3) exists and is given by

$$\hat{B} = (X^t V^{-1} X)^{-1} X^t V^{-1} Y$$

where $E(\hat{B}) = B$

and (5) $\text{Cov} (\hat{B}) = (X^t V^{-1} X)^{-1} \sigma^2$.

When $\text{Cov} (e) = \sigma^2 I$, the above reduces to

$$\hat{B} = (X^t X)^{-1} X^t Y$$

and $\text{Cov} (\hat{B}) = (X^t X)^{-1} \sigma^2$.

$|V|$ refers to the determinant of a matrix throughout this dissertation.
In many experimental situations the independent variables will take on values of 0 or 1 only. In this context, the independent variables commonly act as "dummy" or qualitative variables indicating that a particular factor (or treatment) is present at a given level. A model containing only such indicator variables will be called the qualitative model. A model in which all independent variables range over a continuous spectrum will be called the quantitative model. Both qualitative and quantitative variables may be present in the same model. In the literature, the qualitative model is sometimes referred to as the analysis-of-variance model while the quantitative model is called the regression model. (e.g., see 55)

The Optimum Experimental Design Problem

The optimum experimental design problem is to choose a design matrix $\bar{X}$ for the linear model described above in such a manner as to optimize some measure of effectiveness subject to physical and resource constraints. From both a traditional point of view and for the purpose of this dissertation, it is beneficial to characterize a design matrix in a slightly different manner than has been done previously.
Consider the following design matrix:

\[
\begin{bmatrix}
  f_0(X_1) & f_1(X_1) & \cdots & f_q(X_1) \\
  f_0(X_1) & f_1(X_1) & \cdots & f_q(X_1) \\
  \vdots & \vdots & \ddots & \vdots \\
  f_0(X_n) & f_1(X_n) & \cdots & f_q(X_n) \\
  f_0(X_n) & f_1(X_n) & \cdots & f_q(X_n)
\end{bmatrix}
\]

Then define the matrix $\bar{X}$ such that:

\[
N^t\bar{X} = \begin{bmatrix}
  f_0(X_1) & \cdots & f_0(X_1) & \cdots & f_0(X_n) \\
  f_1(X_1) & \cdots & f_1(X_1) & \cdots & f_1(X_n) \\
  \vdots & \vdots & \vdots & \vdots & \vdots \\
  f_q(X_1) & \cdots & f_q(X_1) & \cdots & f_q(X_n)
\end{bmatrix} \begin{bmatrix}
  f_0(X_1) & \cdots & f_q(X_1) \\
  f_0(X_1) & \cdots & f_q(X_1) \\
  \vdots & \vdots & \vdots \\
  f_0(X_n) & \cdots & f_q(X_n)
\end{bmatrix}
\]

\[
= N \begin{bmatrix}
  \sum_{i=1}^{n} p_i[f_0(X_i)]^2 & \sum_{i=1}^{n} p_i f_0(X_i) f_1(X_i) & \cdots \\
  \sum_{i=1}^{n} p_i f_0(X_i) f_q(X_i) & \sum_{i=1}^{n} p_i[f_q(X_i)]^2 & \cdots \\
  \sum_{i=1}^{n} p_i f_1(X_i) f_q(X_i) & \cdots & \sum_{i=1}^{n} p_i[f_q(X_i)]^2
\end{bmatrix}
\]

where $\sum_{i=1}^{n} p_i = 1$, $p_i > 0$, $i = 1, 2, \ldots, n$, and $f_0(X) = 1$. 
In this context, the $p_i$'s represent the proportion of times the $i^{th}$ experimental point is repeated in the design. $N p_i = n_i$, $i = 1, 2, \ldots, n$, is the number of times the experiment is repeated at the $i^{th}$ design point. The $p_i$'s will be referred to as the proportionality measure of the design and the $X_i^+$'s as the design spectrum or points of support of the design. If the $n_i$'s are restricted to integer values the problem is called exact; if not the problem will be referred to as the non-integer problem.

To determine a design both its spectrum and its proportionality measure must be specified. The matrix $X^T X$ is sometimes referred to as the information matrix.

When the model under consideration is qualitative, then the design spectrum is fixed with its specific values determined by the coding scheme adopted by the experimenter. Because the qualitative model has only the proportionality measure as decision variables and because of the nature of the solution procedure to be developed, the qualitative model can be considered in this dissertation as a special case of the quantitative model. Only the quantitative model will be considered. Because the number of experimental runs, $N$, is often quite small for the qualitative model, it may require solving for the exact problem.

A general formulation of the optimization problem for the non-integer quantitative model can now be expressed as follows:
Max \( f(p_1, \ldots, p_n; x_1^t, \ldots, x_n^t) \)

subject to:

\[
\begin{align*}
(x_1^t, \ldots, x_n^t) \in & \Xi \\
\sum_{i=1}^{n} p_i &= 1 \\
p_i &\geq 0 \quad i = 1, 2, \ldots, n
\end{align*}
\]

\( n \), an integer

\( \Xi \) defines an experimental region which is convex. \( \Xi \) will always be nonempty, closed and bounded (i.e., compact).

The objective function, \( f \), is some measure of effectiveness. It is assumed that \( n \) is finite; otherwise the design would not be useful. It will be shown in Chapter II that an optimum design exists with \( n \) finite.
Design Criteria - Literature Review

The linear statistical model provides the motivation for most of the optimality criteria which have been considered in the literature. As can be seen from the previous discussion, the least-square estimates are minimum variance, unbiased estimates of the unknown parameters in the linear model. However, the covariance matrix of the least-square estimates depends upon the design matrix $\bar{X}$ through the matrix $(X^tX)^{-1}$. Therefore it would appear reasonable in choosing the elements of $\bar{X}$ (i.e., determining the design) to consider the resulting elements of $(X^tX)^{-1}$.

Ideally an experimenter would like the least-square estimates to have as small a variance as possible and at the same time have zero correlation. Unfortunately in most situations it is impossible to simultaneously minimize each variance while maintaining zero correlation. Therefore, various scalar measures of effectiveness have been proposed each having a particular statistical appeal.

One of the first attempts at finding an optimum design was due to K. Smith (56). He was able to find the optimum proportions of the independent variable in a fixed interval for a polynomial up to degree six which minimized the maximum variance of the estimated dependent variable. A design satisfying this minimax criterion is called G-optimum. Elfving did some of the pioneering work in
optimum designs. He proposed a geometrical solution for determining a design which minimizes the variance of a single estimable function$^3$ and a design which minimizes the average variance (or trace of $[X^t X]^{-1}$) of the least-squares estimates (19). Chernoff generalized Elfving's work by proposing an approximation to his optimum design for large numbers of experimental runs (9).

A different criterion was motivated by Ehrenfeld. He defined the efficiency of a design by the ratio $\lambda \min / u$ where $u = \text{maximum value of } \lambda \min$ and $\lambda \min$ is the minimum eigenvalue of $X^t X$. This criterion (i.e., minimizing the maximum eigenvalue of the covariance matrix) is equivalent to minimizing the maximum variance of all linearly independent estimable functions (18).

The most popular criterion encountered in the literature, and the one used throughout this dissertation, is that of maximizing the determinant of the matrix $X^t X$ or equivalently minimizing the determinant of its inverse. This criterion has several appealing properties. If the error terms are normally distributed, it leads to a design having the smallest hypervolume for a given confidence region of the parameter estimates (37). In addition, it is invariant under a non-singular linear transformation.

$^3$See Scheffe (55) for a discussion of estimable functions.
on the experimental space. (37) Referred to as D-optimality this criterion is equivalent to minimizing the product of the eigenvalues of \((\bar{X}^t \bar{X})^{-1}\) or in other words minimizing the generalized variance.

Hoel using the D-optimality criterion showed that the optimum design for an \(n^{th}\) degree polynomial could be determined by finding the zeros of the derivatives of Legendre polynomials (26). Guest showed that the same design was G-optimal (24). This equivalence was shown to be true regardless of the regression model. Kiefer and Wolfowitz proved that D-optimum and G-optimum designs were equivalent for the quantitative non-integer problem (39). This rather interesting result has helped motivate the use of the D-optimum criterion.

More recent work in this area has resulted in solutions to the D-optimum design problem for particular single variable regression models in terms of zeros of other classical polynomials. The reader is referred to Karlin and Studden (30) or Fedorov (20) for specific regression models and their solutions.

Several attempts have been made to apply much of the optimum design theory to more general design models. Most notable is that of Box and Draper (8) and Nalimov, Golikov and Mikeshina (48). Other recent work in this area is by Bloom, Pfaffenberger, and Kochenberger (4).
These have been attempts to approximate D-optimum designs and to conduct direct searches for a local optimum design. They are somewhat successful for models with a small number of parameters and independent variables. However, they are computationally difficult for complex regression models. Few attempts have been made to solve the more general optimization problem with constraints on the experimental region (other than upper and lower bounds on the independent variables). Fedorov has developed an algorithm which will solve this problem but it requires repeatedly solving a non-linear constrained optimization problem (20). Kennard and Stone (32) have provided a sequential method that generates design points uniformly over an arbitrary experimental space. However, their criterion for choosing the next design point is to pick that point from a finite set of points that is farthest from existing design points. This method of choosing a design from a finite set of candidate points is similar to the methodology developed in this dissertation. An essential difference, and an important one, is that the criterion used in this dissertation is that of D-optimality. Furthermore, the procedure described in this dissertation allows for an improved (in terms of the objective function) set of candidate points to be considered.

Finally, there has appeared in the literature discussions of the augmentation problem and the model discrimination problem. The augmentation problem is one
of choosing \( m \) additional experimental points, given that \( n \) experiments have already been conducted. The reader should consult Wynn (64), Dykstra (17), and Covey-Crump and Silvey (11) for theoretical results in this area. The model discrimination problem is concerned with finding optimum designs which allow for the assumed model to be incorrect. A disadvantage of D-optimality (as well as most other criteria) is that it is model dependent. If the incorrect regression model is hypothesized, then the optimum design based upon that model will not be optimum for the true model. For example, if a first degree polynomial is hypothesized, then the optimum design is to take an equal number of observations at both end points of the experimental region. If, however, the true curve is a second degree polynomial, then that design is no longer optimum. In fact, the second degree curve cannot even be fitted using a design with only two distinct points. Stigler has offered one approach to solving this problem by modifying the D-optimum criterion (57). Both of these problems can be addressed from a slightly different view point based upon the methodology developed in this dissertation.

**Problem Formulation - Overview**

The specific problem (P1) to be addressed in this dissertation is the following:
PI: $\text{Max } |N^tX|$
\[ p_i, X_i, n \]

subject to: $X_i^t \in \Xi$

\[ i = 1, 2, \ldots, n \]

\[ \sum_{i=1}^{n} p_i = 1 \]

\[ p_i \geq 0 \quad i = 1, 2, \ldots, n \]

$n$ an integer

It will be assumed that the constraint set $\Xi$ is closed and bounded. That is, the experimental region is compact. This is not an unrealistic assumption since in all meaningful experiments a well defined region of interest can be determined. The problem as defined in PI is to solve for the non-integer solution; that is, the optimum set of $p_i$'s will be such that $n_i = Np_i$ is not necessarily an integer.

The solution procedure to be developed in this dissertation is as follows. First a finite set $(n)$ of support points referred to as candidate or grid points is determined which "adequately" covers the experimental region. Then problem PI is optimized with respect to the proportionality measure $p_i$ only. The constraint set $\Xi$ is implicitly considered by choosing the candidate points such that these constraints are not violated. Only the non-integer problem will be solved. However, the integer solution can be approximated particularly for a large number of experimental
runs. For most regression models, it will not be possible to choose the best set of candidate points. However by establishing a fine enough grid structure over the experimental region, the optimum solution to PI can be accurately approximated. The optimal set of $p_i$'s will then determine the proportion of experiments to be run at each of the pre-established grid points. It will be shown that once an estimate for the optimum design is found based upon the initial set of candidate points, then a new candidate point can be determined, and the optimization carried out again leading to a better solution to problem PI. This procedure can be repeated and will converge to the D-optimum design.

Once the non-integer problem is solved, then an estimate of the integer solution can be determined by searching in the area of the non-integer solution. The non-integer solution will provide an upper bound on the exact problem while any feasible solution to the exact problem will provide a lower bound. Thus the search can terminate when the experimenter determines that a reasonably good solution has been obtained or no more improvement can be found from the search. For large $N$, rounding off may provide an accurate solution. Also exact solutions will be found for certain values of $N$. That is, those values of $N$ for which $Np_i = n_i$ are integers for all $i$.

There are several advantages to imposing a grid structure on the problem. It enables a solution (although
many times approximate) to be obtained which would be difficult or impossible to obtain otherwise. The resulting problem has nice properties such as a concave objective function to be maximized. Computationally the problem is easier to solve. The approach is completely general in that any regression model can be considered and irregular experimental regions are permitted.

In Chapter II, theoretical results will be obtained for the problem. The concavity of \( f(p_1, \ldots, p_n) \) will be established, necessary and sufficient conditions for the maximum will be found, and an equivalence relationship proved. Chapter III will develop an algorithm for finding the best design over a given set of experimental points. Chapter IV examines the problem of determining candidate or grid points. This leads to a procedure for improving a given solution. An estimate of the accuracy of a design relative to the D-optimum is determined and convergence to the D-optimum design is shown. Chapter V examines some examples and applications of the procedure. Extensions to problem P1 are discussed in the appendices.
CHAPTER II

A RELATED PROBLEM

The problem as defined in Chapter I (Problem P1) is to maximize the determinant of the matrix $X^tX$. The decision variables are the points of support of the design $(X^t_i)$, the proportion of experiments to be performed at each point of support $(p_i)$, and the total number of distinct points $(n)$ in the design. This can be a formidable problem if an attempt is made to optimize over all the variables simultaneously. In fact, until $n$ is specified it is impossible to know how many decision variables are in the problem. If there are $k$ independent variables and $n$ distinct experimental points, then the total number of decision variables is given by $n (k+1)$. Since the objective function is non-linear, any attempt at directly solving the constrained maximization problem as a mathematical program will result in a locally optimum solution. Therefore, initially it is advantageous to solve a related but simpler problem. The purpose of this chapter is to define this new problem, investigate its characteristics, and show its equivalence to a third problem. This equivalence relationship will aid in solving both the new problem and the original problem (Problem P1).
Of main concern in this chapter will be an approximation to problem P1 determined by establishing a finite set of support points as candidates for the experimental design. Assume that a set of \( n \) experimental points \( X_i^t = (x_{i1}, x_{i2}, \ldots, x_{ik}) \), \( i = 1, 2, \ldots, n \), have been selected in some manner such that the constraint set

\[
X_i^t \in \mathbb{E} \quad i = 1, 2, \ldots, n
\]

is not violated.

The problem of interest then becomes

\[
P_2: \max \quad |N\bar{X}^t\bar{X}|
\]

Subject to:

\[
\sum_{i=1}^{n} p_i = 1 \quad p_i \geq 0 \quad i = 1, 2, \ldots, n
\]

The objective function can be written in terms of the \( p_i \) and the \( n \) candidate points as follows. Letting \( [f(X_i)]^t = [f_0(X_i) \ldots f_q(X_i)] \), then

\[
|N\bar{X}^t\bar{X}| = |N| \sum_{i=1}^{n} p_i [f(X_i)] [f(X_i)]^t
\]

\[
= N^{q+1} |\sum_{i=1}^{n} p_i [f(X_i)] [f(X_i)]^t|.
\]

Note that the matrix \( \bar{X}^t\bar{X} \) is symmetrical, and if the rank of \( \bar{X} \) is \( q + 1 \) (full rank), then the rank of \( \bar{X}^t\bar{X} \) is also \( q + 1 \) (see e.g., 55, p. 399). In either case the matrix \( \bar{X}^t\bar{X} \) will be positive semi-definite. Since the objective is to maximize the determinant of \( \bar{X}^t\bar{X} \), only the full rank case (\( \bar{X}^t\bar{X} \) positive-definite) will be of interest; otherwise
If we let $X_f$ be that design matrix having $n_i = 1$ for $i=1, 2, \ldots, n$, and if $n \geq q+1$, then $X_f$ will be of full rank if and only if the columns of $X_f$ are linearly independent. If the grid points are chosen such that $X_f$ is of full rank, then any design consisting of one or more repetitions of these points will also be of full rank since the columns of the augmented design matrix will also be linearly independent. It follows from this discussion that a minimum of $q+1$ distinct grid points must be chosen in order that $X_f$ is of full rank. It will be assumed throughout this dissertation that this is the case.

Concavity of P2

In order to establish an important property of the objective function in problem P2, these two lemmas are needed:

Lemma 1.

Let $h(X_1, X_2, \ldots, X_m)$ be a concave function. If $f_1(Y_1, \ldots, Y_n), \ldots, f_m(Y_1, \ldots, Y_n)$ are $m$ linear functions in $Y \in S \subset \mathbb{R}^n$, then $h[f_1(Y_1, \ldots, Y_n), \ldots, f_m(Y_1, \ldots, Y_n)]$ is a concave function in $Y$ over the set $S$ where $S$ is convex.

Proof.

Let $Y^1$ and $Y^2$ be two elements of the set $S$. Then

$$h[f_1(\alpha Y^1 + (1-\alpha)Y^2), \ldots, f_m(\alpha Y^1 + (1-\alpha)Y^2)]$$

$$= h[\alpha f_1(Y^1) + (1-\alpha) f_1(Y^2), \ldots, \alpha f_m(Y^1) + (1-\alpha) f_m(Y^2)]$$

by the linearity of $f_1, \ldots, f_m$. 

by the concavity of $h$. 

**Lemma 2.**

Given the closed convex set $\mathbb{R}^n$, the set of information matrices $X^tX$ (based upon $n$ finite points of support) determined by $p_1, \ldots, p_n \in T$ where $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ is convex. That is, the convex combination of any two feasible information matrices is a feasible information matrix.

**Proof.**

Let $p_i = (p_1^{i}, \ldots, p_n^{i}) \in T$, $i = 1, 2$ be two proportionality measures. It follows immediately that $ap_1^{i} + (1-a)p_2^{i} > 0$, $\sum_{i=1}^{n} (ap_1^{i} + (1-a)p_2^{i}) = 1$ and $ap_1^{i} + (1-a)p_2^{i} \in T$.

That is any convex combination of $p_1^{i}$ and $p_2^{i}$ is also a proportionality measure defined on the set $T$. Since

\[
X^tX = \sum_{i=1}^{n} p_if(X_i)f(X_i)^t; \text{ then}
\]

\[
\sum_{i=1}^{n} \left[ ap_i^{1} + (1-a)p_i^{2} \right] f(X_i)f(X_i)^t
\]

\[
= \sum_{i=1}^{n} ap_i^{1}f(X_i)f(X_i)^t + \sum_{i=1}^{n} (1-a)p_i^{2}f(X_i)f(X_i)^t
\]

\[
= \alpha X_1^tX_1 + (1-\alpha) X_2^tX_2
\]
It is noted that the logarithm of the determinant of a symmetric positive definite matrix is a strictly concave function in the elements of the matrix (e.g., see 3, p. 128). That is,

**Fact 1.**

\[ \log |\alpha A + (1-\alpha) B| > \alpha \log |A| + (1-\alpha) \log |B| \]

where \( 0 < \alpha < 1 \), \( A \neq B \), and \( A \) and \( B \) are positive-definite.

The following important result can be obtained.

**Theorem 1.**

If a design matrix \( \bar{X} \) is full rank, then \( h(p_1, \ldots, p_n) = \log_e |\bar{X}^t \bar{X}| \) is a concave function where \( (p_1, \ldots, p_n) \in T \subseteq \mathbb{R}^n \) and \( T \) is a convex set.

**Proof.**

By construction \( \bar{X}^t \bar{X} \) is symmetric and positive-definite. Therefore, \( \log_e |\bar{X}^t \bar{X}| \) is concave in the scalar elements of the matrix \( (\bar{X}^t \bar{X}) \). Since each element of \( \bar{X}^t \bar{X} \) is a linear function of \( p_1, \ldots, p_n \), then the theorem follows from Lemma 1 and the convexity of the set of feasible matrices established by Lemma 2.

From this theorem and the fact that the logarithm is a monotonic function, problem P2 can be transformed into an equivalent concave program.
P2': \[ \text{Max } \log_e |X^t X| = \text{Max } \log_e \left[ \sum_{i=1}^{n} p_i f(X_i) f(X_i)^t \right] \]

Subject to:
\[ \sum_{i=1}^{n} p_i = 1 \quad p_i \geq 0 \]

Observe that \( N \), the total number of observations is not explicitly considered in solving problem P2. Since
\[ \log_e |N X^t| = \log_e N^{q+1} |X^t X| \]
\[ = \log_e N^{q+1} + \log_e |X^t X| , \]

it is sufficient to maximize \( \log_e |\bar{X}^t \bar{X}| \) with respect to the proportionality measure.

**Kuhn-Tucker Theory**

It is helpful to consider the Kuhn-Tucker conditions for problem P2'. The Lagrangian function is:
\[ L = -\log_e \left[ \sum_{i=1}^{n} p_i f(X_i) f(X_i)^t \right] + v \left[ \sum_{i=1}^{n} p_i - 1 \right] - \sum_{i=1}^{n} p_i w_i \]

where \( v \) and \( w_i \) are Lagrangian multipliers. The following fact concerning matrix derivatives is required:
Fact 2.

$$\frac{\partial \log_e |X|}{\partial X} = X^{-1}$$ (see e.g., 49).

The Kuhn-Tucker conditions can be written as:

1. \[ f(X_i)^T S^{-1} f(X_i) + w_i = v \quad i = 1, 2, \ldots, n \]

2. \[ \sum_{i=1}^{n} p_i = 1 \]

3. \[ p_i \geq 0 \]

4. \[ w_i \geq 0 \]

5. \[ p_i w_i = 0 \]

where \( S^{-1} = \{s_{ij}\} = (X^T X)^{-1} \) (i.e., the optimum covariance matrix except for the constant \( \frac{\sigma^2}{N} \)). From (1), (2), and (5)

\[ \sum_{i=1}^{n} p_i f(X_i)^T S^{-1} f(X_i) + \sum_{i=1}^{n} p_i w_i = v \sum_{i=1}^{n} p_i \]

or

\[ \sum_{i=1}^{n} p_i f(X_i)^T S^{-1} f(X_i) = v. \]

Letting \( S^{-1} = \{s_{jk}\} \), it follows that

\[ \sum_{i=1}^{n} q \sum_{j,k=0}^{q} p_i f_j(X_i) \]

\[ \times f_k(X_i) s_{jk} \]

TR refers to the trace of a matrix throughout this dissertation. Using this symbol, the above equals:
Then $v = q + 1$ and $f(X_i)^t S^{-1} f(X_i) + w_i = q + 1$. From the complementary slackness condition it follows that at points having positive proportionality measure, $w_i = 0$ or $f(X_i)^t S^{-1} f(X_i) = q + 1$. However, $f(X_i)^t S^{-1} f(X_i)$ is nothing more than the variance (except for the constant $\frac{c2}{N}$) of the prediction equation $Y_{pred.} = x_i^t \hat{B}$, evaluated at the $i^{th}$ candidate point. This implies that at all points in the optimum design ($p_i > 0$), the variance of the prediction equation is $\frac{(q+1)c^2}{N}$. At any other candidate point, the $\text{var}(Y_{pred.})$ will be less than or equal to $\frac{(q+1)c^2}{N}$ (since $w_i > 0$). As pointed out previously, a minimum of $q + 1$ points must appear in the optimum solution if the design is to be of full rank. Therefore the covariance matrix $S^{-1}$ must be chosen such that the maximum variance over the entire spectrum of grid points is minimized (equal to $q + 1$) and that this maximum variance occurs at no fewer than $q + 1$ grid points. This motivates the consideration of the following problem.
An Equivalent Problem

Before introducing an equivalent problem, the following fact is stated and theorem is proved.

Fact 3.

If $A$ and $B$ are positive-definite matrices and $A \neq B$, then

$$aA^{-1} + (1-a)B^{-1} \geq [aA + (1-a)B]^{-1}, \quad 0 < a < 1$$

where the inequality $A > B$ implies that $A - B$ is a positive-definite matrix (e.g., see 34).

Theorem 2.

Let $h(s_{11}, \ldots, s_{qq}) = f(X)^tS^{-1}f(X)$

where $S = \{s_{ij}\}$ and $S^{-1}$ is positive definite. Then $h$ is a strictly convex function.

Proof.

Let $S_1$ and $S_2$ be two distinct positive definite matrices. This implies $f(X)^tS_1^{-1}f(X) > 0$ and $f(X)^tS_2^{-1}f(X) > 0$ for all $f(X) \neq 0$. Using the previously stated fact that

$$S_1^{-1} + (1-a)S_2^{-1} - [aS_1 + (1-a)S_2]^{-1}$$

is positive definite, then
\[ f(X)^t[aS_1^{-1} + (1-a) S_2^{-1}] - [aS_1 + (1-a)S_2]^{-1} f(X) > 0 \]

or equivalently

\[ af(X)^t[S_1^{-1}f(X) + (1-a)f(X)^tS_2^{-1}f(X)] > f(X)^t[aS_1 + (1-a)S_2]^{-1}f(X) \]

for all \( f(X) \neq 0 \). Which in turn implies

\[ h(aS_1 + (1-a)S_2) < ah(S_1) + (1-a) h(S_2). \]

Lemma 2 insures that the set of feasible \( s_{jk} \) is convex. Hence by definition \( h(s_{11}, \ldots, s_{qq}) \) is strictly convex.

Problem P3 is now defined.

P3:

\[
\begin{align*}
\text{Min} & \quad \text{Max} \quad \{f(X_i)^tS^{-1}f(X_i)\} \\
p_i, s_{jk} & \quad i=1,2,\ldots,n
\end{align*}
\]

Subject to:

\[ s_{jk} = \sum_{i=1}^{n} p_i f_j(X_i) f_k(X_i)^t \quad j<k = 0,1,\ldots,q \]

\[ \sum_{i=1}^{n} p_i = 1 \quad p_i \geq 0 \]

Problem P3 is one of choosing a proportionality measure over the set of candidate points such that the maximum variance of the prediction equation is minimized. Note
that for each i, \( f(X_i)^t S^{-1} f(X_i) \) is a strictly convex function by Theorem 2 since \( S^{-1} \) will always be chosen such that it is positive definite. Because the maximum over a set of strictly convex functions is itself strictly convex (see 43), then \( h(s_{ll}, \ldots, s_{qq}) = \max \{ f(X_i)^t S^{-1} f(X_i) \} \) is a strictly convex function and has a unique global minimum. In order to develop an equivalence relationship, the following preliminary results are needed.

**Lemma 3.**

The weighted sum of the variances \( f(X_i)^t S^{-1} f(X_i) \) taken over all points in the design is equal to the number of unknown parameters, \( q + 1 \).

**Proof.**

\[
\begin{align*}
\sum_{i=1}^{n} \sum_{j=1}^{q} \sum_{k=0}^{q} p_i f_j(X_i) f_k(X_i) s_{jk} &= \sum_{j=0}^{q} s_{jk} \sum_{i=1}^{n} p_i f_j(X_i) f_k(X_i) \\
&= \sum_{j=0}^{q} s_{jk} \sum_{i=1}^{n} p_i f_j(X_i) f_k(X_i) = \sum_{j=0}^{q} s_{jk} s_{jk} \\
&= TR I = q + 1.
\end{align*}
\]
Lemma 4.

The minimum value of $\text{Max} \{f(X_i)^t S^{-1} f(X_i)\}$

$$i = 1, 2, \ldots, n$$

is $q + 1$.

Proof.

Assume $\text{Max} \{f(X_i)^t S^{-1} f(X_i)\} = K < q + 1$.

Then

$$\sum_{i=1}^{n} p_i f(X_i)^t S^{-1} f(X_i) \leq \sum_{i=1}^{n} p_i K = K < q + 1,$$

since $p_i \geq 0$ and $f(X_i)^t S^{-1} f(X_i) \leq \text{Max} f(X_i)^t S^{-1} f(X_i) = K$.

However this result contradicts Lemma 3. Therefore, the minimum value of $\text{Max} \{f(X_i)^t S^{-1} f(X_i)\}$ cannot be less than $q + 1$.

This leads to the following important result.

Theorem 3.

Problems P2 and P3 are equivalent. That is the same covariance matrix (and hence the same design) solves both problems.

Proof.

From the Kuhn-Tucker conditions to P2', it can be
seen that the optimum solution to P2 determines an $S^{-1}$ such that $f(X_i)^t S^{-1} f(X_i) = q + 1$ for all $p_i > 0$. However, by Lemma 4, this is a lower bound on $\max \{ f(X_i)^t S^{-1} f(X_i) \}$. Hence P2 determines a design which minimizes the maximum variance of $f(X_i)^t B$; or in other words solves problem P3. If problem P3 is solved then by Lemma 4 a covariance matrix $S^{-1}$ is found such that

$$\max_{i} \{ f(X_i)^t S^{-1} f(X_i) \} = q + 1$$

Assume for at least one $p_i > 0$

$$f(X_i)^t S^{-1} f(X_i) < q + 1,$$

and then

$$\sum_{i=1}^{n} p_i f(X_i)^t S^{-1} f(X_i) < q + 1$$

contradicting Lemma 3. Therefore if $p_i > 0$, then

$$f(X_i)^t S^{-1} f(X_i) = q + 1$$

and the Kuhn–Tucker conditions of P2' are satisfied and P3 also solves P2.

The proof of Theorem 3 requires that the covariance matrix (and hence the design matrix) be of full rank. It should also be observed that the optimum set of proportionality measures $p_i$ need not be unique. That is, the optimum design is not necessarily unique, but the optimum covariance matrix is.

Theorem 3 is similar to the result proved by Kiefer
and Wolfowitz (39) for the optimum design over a continuous range of the independent variables. Theorem 3 shows the equivalence of the two problems is still true when a grid structure is imposed upon the problem. The value of this theorem lies in the insight it can provide in solving either P2 or P3 and in determining a grid structure. This will be discussed later. Generally P2 will be the easier problem to solve. The equivalence of the two problems provides additional motivation for choosing the determinantal criterion as a measure of effectiveness.

The minimax problem can be rewritten in an equivalent form (P3') for solution.

\[ \text{P3'}: \quad \text{Min} \quad Y \]

\[ \quad p_i, s_{jk} \]

Subject to:

\[ f(X_i)^t S^{-1} f(X_i) \leq Y \quad i = 1, 2, \ldots, n \]

\[ s_{jk} = \sum_{i=1}^{n} p_i f_j(X_i) f_k(X_i) \quad j, k = 0, 1, \ldots, q \]

\[ \sum_{i=1}^{n} p_i = 1, \quad p_i > 0. \]

It is not helpful at this point to investigate the Kuhn-Tucker conditions for this problem. As it turns out, they are very similar to those of P2 but of a more complicated form.
An Additional Result

It is of interest and value to note that a given information matrix can be represented by at most
\[
\frac{(q + 1)(q + 2)}{2} + 1
\]
strictly positive \( p_i \). This can be seen by noting first that, because of the symmetry of the matrix \( \tilde{X}^t \tilde{X} \), there are \( \frac{(q+1)(q+2)}{2} \) elements to be described. Then letting
\[
\tilde{X}^t \tilde{X} = \{s_{jk}\}, \text{ the following relationships must hold:}
\]

\[
s_{jk} = \sum_{i=1}^{n} f_j(X_i)f_k(X_i)p_i \quad j \leq k = 0, 1, \ldots, q
\]

\[
1 = \sum_{i=1}^{n} p_i
\]

and \( p_i \geq 0 \).

If the elements \( \{s_{jk}\} \) are specified, then one solution to the above set of equalities with \( p_i \geq 0 \) can be found by solving a linear program. If a feasible basic solution exists it can have no more than \( \frac{(q + 1)(q + 2)}{2} + 1 \) variables positive. Such a solution can be found by solving Phase I of the two-phase simplex method. Thus while the number of candidate points considered, \( n \), may be quite large, a solution will exist having relatively few nonzero variables.
Therefore in solving problem P2 (or P3), the number of strictly positive $p_i$ denoted by $n'$ will lie in the interval $q + 1 \leq n' \leq \frac{(q + 1)(q + 2)}{2} + 1$. 
CHAPTER III

AN ALGORITHM

The purpose of this chapter is to develop a procedure for solving problem P2. Since the problem has been formulated as a mathematical program, a number of existing optimization techniques can be directly applied. Most of these techniques, however, will not take advantage of the structure of the problem and the known characteristics of its solution. Therefore a modification of an existing algorithm is proposed as one method of obtaining a solution.

Some Considerations for an Algorithm

It is beneficial to consider briefly some of the characteristics of the problem. In order to obtain a good approximation to the true optimum, a relatively large number of candidate or grid points is desired. As indicated at the end of Chapter II, an optimum solution will exist with at most \((g+1)(g+2) + 1\) of the \(p_i\)'s strictly positive. Hence, it is possible to have a large number of decision variables in the formulation of the problem with relatively few of them non-zero at the optimum. It would increase the speed
of convergence of the algorithm if the zero variables could quickly be identified. A second characteristic is that the constraints on the decision variables, $p_i$, take on a rather simple form. It is therefore possible to solve $P2$ as an unconstrained problem. Third, the objective function ($P2'$) is concave; hence a global maximum can be found by a steepest ascent procedure. As indicated by the Kuhn-Tucker conditions in Chapter II, an expression for the gradient is available. Finally, the equivalence of $P2$ and $P3$ provides an excellent method for determining how far a given solution is from the optimum. These considerations lead to the following choice of an algorithm.

The basic algorithm used is a conjugate gradient method developed by Fletcher and Reeves (22). Modifications of this algorithm are made to take advantage of the structure of this particular problem. The Fletcher-Reeves algorithm was chosen for several reasons. First, it utilizes gradient information. It is not possible to obtain a closed form expression for the gradient in this case without first explicitly writing the determinant of the $X^tX$ matrix in terms of the $p_i$. However, it is possible to numerically calculate the gradient at a given point in the experimental space. Use of gradient information will generally insure faster convergence than will an algorithm that optimizes without using derivatives. Second, because conjugate directions are utilized, some of the disadvantages of a
steepest ascent procedure, particularly the tendency to zigzag when the contours of the objective function are eccentric, are overcome. Initial experience with a steepest ascent algorithm indicates that this is the case with this particular function. Third, the amount of additional storage space required is minimum. Only the actual direction of search and current vector of decision variables need be stored at each iteration. Compared with a more sophisticated algorithm developed by Fletcher and Powell (21) requiring in addition to several vectors of size \( n \), the storage of a positive-definite matrix of order \( n \), the savings in storage can be considerable. This is particularly true when the number of decision variables, \( n \), is quite large. Finally, the application of the Fletcher-Reeves algorithm as well as the necessary modifications are relatively simple. The importance of this fact should not be overlooked since one of the primary objectives is to develop a methodology which can be easily applied to a large variety of design models.

**Derivation of the Algorithm**

The specific problem to be addressed is given by the following:

\[
P2': \quad \max \log |\bar{X}^t\bar{X}|
\]

\( P_1', \ldots, P_n \)
Subject to:
\[
\begin{align*}
\sum_{i=1}^{n} p_i &= 1 \\
p_i &\geq 0 \quad i = 1, 2, \ldots, n
\end{align*}
\]

The normalizing constraint is eliminated by reducing the number of decision variables by one. That is,
\[
(3-1) \quad p_n = 1 - \sum_{i=1}^{n-1} p_i
\]

The non-negativity restrictions are handled implicitly by not allowing the variables to increase in a negative direction beyond the value of zero.

In order to determine the value of the gradient, it is first necessary to rewrite the objective function in terms of the n-1 decision variables. That is, the \( j,k \) element of the \( X^t X \) matrix can be written as
\[
(3-2) \quad s_{jk} = \sum_{i=1}^{n-1} p_i [f_j(x_i)f_k(x_i) - f_j(x_n)f_k(x_n)]
\]

by substituting \( l - \sum_{i=1}^{n-1} p_i \) for \( p_n \).
That is
\[ f(p_1, \ldots, p_{n-1}) = \sum_{i=1}^{n-1} p_i [f(x_i)f(x_i)^t + f(x_n)f(x_n)^t] \]

Using Fact 2 stated in Chapter II concerning matrix derivatives (i.e., \( \frac{\partial \log|C|}{\partial c_{ij}} = c_{ij} \)) and the chain rule,

the gradient of the objective function is given by:

(3-3)
\[ \frac{\partial f(p_1, \ldots, p_{n-1})}{\partial p_i} = \sum_{j,k=0}^{q} [f_j(x_1)f_k(x_1) - f_j(x_n)f_k(x_n)] s_{jk} \]

for \( i = 1, 2, \ldots, n-1 \)

In order to find the gradient at a given point, the inverse of the matrix \( X^tX \) must be found \( \{s_{jk}\} \) for given values of \( p_1, \ldots, p_n \). It is worth noting that the order of the matrix \( X^tX \) is \( q + 1 \) which does not depend upon the number of decision variables, \( n \), but does depend upon the number of parameters in the linear model. Computationally finding inverses can be very time consuming; however, the size of the matrix will be relatively small compared with the number of decision variables. It is this fact that enables the gradient to be used efficiently. That is all \( n-1 \) derivatives can be found once a matrix of order \( q + 1 \) is inverted.
It is now possible to list the steps in the algorithm. Assuming a grid structure has been established, then:

I. Choose an initial starting point. If no information is known concerning possible candidate points that have an exceptionally high possibility of appearing in the optimum solution (e.g., extreme points of the experimental region), then choose \( p_i^0 = \frac{1}{n} \) for the initial vector.

II. Calculate \( X^tX \) using (3-2).

III. Find \( (X^tX)^{-1} \).

IV. Find the gradient using (3-3).

V. Determine the next direction of search, \( d_i^{m+1} \) at the \( m \)th iteration by

\[
(3-4)^1 \\
\frac{\partial f(p_1^0, \ldots, p_{n-1}^0)}{\partial p_i} \\
\frac{\partial f(p_1^{m+1}, \ldots, p_{n-1}^{m+1})}{\partial p_i} + B^{m} d_i^m \hspace{1cm} m = 0, 1, \ldots
\]

\(^1\)This step is in essence the Fletcher-Reeves algorithm.
A recommended modification to this step in the algorithm is to restart the direction of search with the current gradient after every \( n+1 \) iterations. That is let

\[
\begin{align*}
\mathbf{B}^m &= \frac{n-1}{\sum_{i=1}^{n-1} \left[ \frac{\partial f(\mathbf{p}_1, \ldots, \mathbf{p}_{n-1})}{\partial \mathbf{p}_1} \right]^2} \\
&= \frac{n-1}{\sum_{i=1}^{n-1} \left[ \frac{\partial f(\mathbf{p}_1, \ldots, \mathbf{p}_{n-1})}{\partial \mathbf{p}_1} \right]^2}
\end{align*}
\]

This will generally result in a faster rate of convergence since after \( n \) iterations the gradient direction is completely lost (see 22).

VI. Modify the direction of search to insure non-negativity. That is, the new directions of search are given by

\[
d_i^{m+1} = \frac{\partial f(\mathbf{p}_1, \ldots, \mathbf{p}_{n-1})}{\partial \mathbf{p}_i} \\
\begin{cases} 
  d_i^{m+1} & \text{if } p_i^m > 0 \\
  \text{Max} (0, d_i^{m+1}) & \text{if } p_i^m = 0.
\end{cases}
\]

\[ (3-5)^2 \]

\[ s_i^{m} = \begin{cases} 
  d_i^{m+1} & \text{if } p_i^m > 0 \\
  \text{Max} (0, s_i^{m+1}) & \text{if } p_i^m = 0.
\end{cases} \]

\[ 2\]This modification can create a problem in convergence. See Zangwill (65) for a discussion concerning this problem.
VII. Determine the optimum step size by performing a one dimensional search such that the $\max_{0 \leq \alpha \leq U} f(p^m + \alpha s^{m+1})$ is found.

where

$$\begin{aligned}
\min \left\{ \min_{i \in S_{m+1}^{m+1}, \exists S_{m+1}^{m+1} < 0} \left\{ \min \left( -\frac{p_i^{m+1}}{s_i^{m+1}}, \left(1-\sum_{i=1}^{m} p_i^m\right)/\sum_{i=1}^{m} s_i \right) \right\} \right.
\end{aligned}$$

This upper bound insures $p_i^{m+1} > 0$ for all $i$ and $\sum_{i=1}^{m} p_i^{m+1} \leq 1$.

In this dissertation the method of golden sections is used for the one dimensional search (e.g. see 61).

VIII. Determine a new vector $(p_{i}^{m+1}, \ldots, p_{n}^{m+1})$ using the optimum step size $\alpha^*$:

$$\begin{aligned}
p_{i}^{m+1} &= p_i^m + \alpha^* s_i^{m+1} \quad i = 1, 2, \ldots, n-1 \\
p_n^{m+1} &= 1 - \sum_{i=1}^{n-1} p_i^m + 1
\end{aligned}$$
At this point it has been found that a slight modification to the algorithm will speed convergence. If $0 < p_i^m - \epsilon$ and $s_{i}^{m+1} < 0$, then set $p_i^{m+1} = 0$. Since most of the decision variables will be zero at the optimum, this forces those variables to zero much more quickly. Once a given variable reaches zero, it remains at zero unless a positive direction is encountered. The choice of a value for $\epsilon$ is somewhat arbitrary. However, too large a value may result in non-convergence to the optimum. A "good" value for $\epsilon$ will depend upon the relative values of the non-zero variables at the optimum. Good results have been obtained for values of $\epsilon$ between .001 and .01 with the models discussed in Chapter V.

IX. Check some stopping criterion. If the criterion is not met, return to step III.

The particular stopping criterion used in this algorithm utilizes the equivalence relationship of problems P2 and P3. Since the maximum variance of the regression function over the grid points is minimized with a design which maximizes the determinant of $\bar{x}^T\bar{x}$, then the algorithm can be terminated when the maximum variance reaches its lower bound. From the theoretical discussion in Chapter II a lower bound on the maximum variance is $q+1$. Therefore after a specified number of iterations calculate
Max \{ [f(x_i)]^t S^{-1} [f(x_i)] \}_{1 \leq i \leq n} \\
= \max_{i, j, k=0}^q f_j(x_i)f_k(x_i)s_{jk} = (\text{Var } y)_{\text{Max}}

If \((\text{Var } y)_{\text{Max}} \leq q + 1 + \delta, \delta > 0\), then the algorithm is terminated. This stopping criterion should be checked immediately after calculating a new covariance matrix, 
\((X^tX)^{-1}\) (step III).

The efficiency of this procedure will depend to a great extent on the speed at which the \(X^tX\) matrix is inverted and the degree of accuracy desired in conducting the one-dimensional search. Very fast convergence has been obtained in some cases for \(n\) as large as 100 and \(q+1=4\). However, as will be illustrated in Chapter V, \(n\) does not necessarily have to be large in order to obtain a good design.
CHAPTER IV

THE GRID PROBLEM

The original problem (P1) as defined in Chapter I was to determine an optimum design over some convex experimental region. In an attempt to solve this problem an approximation to P1 was formulated (Problem P2) by specifying a finite set of points from which an optimum design over this finite set of points could be found. Chapter II developed the necessary theory and Chapter III developed an algorithm to solve Problem P2. The purpose of this chapter is to determine how close P2 comes to solving P1 and to determine how a better solution can be obtained. That is, attention is now focused on Problem P1.

Choosing A Grid

It should be apparent that if the correct set of candidate points is chosen, then a solution to P1 can be obtained by solving P2. However, generally it is impossible to know in advance which points in the experimental region are the optimum points of support. An exception is the additive model given by
The optimum set of grid points for the additive model consists of extreme points of the experimental region if the region is convex. This can be seen by noting that the variance function,

\[
\text{VAR}(Y) = g(x_1, \ldots, x_q) = (1, x_1, \ldots, x_q)S^{-1} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_q \end{pmatrix}
\]

is convex since \( S^{-1} \) is positive-definite. Therefore, the Max Var (Y) must occur at an extreme point. Therefore, if the experimental region has a finite number of extreme points (if it is defined by a convex polyhedron) and all extreme points are included as grid points, then P2 solves P1.

For most non-additive models, extreme points as well as interior points will be in the optimum solution. In choosing a grid, extreme points, axial points, boundary points, and center points should be included. An equal spacing of grid points over the entire experimental region has led to designs near the D-optimum in most of the examples in Chapter V.

**A Measure of Accuracy**

Once a grid structure has been established, and the optimum design over the grid has been determined,
one question to be answered is how close this solution is to the D-optimum solution. One approach to answering this question is based upon the equivalence of D-optimum and G-optimum designs. While the objective is to maximize the determinant of the matrix $X^tX$, a criterion for determining how good a given design is relative to the D-optimum design is to determine how close the maximum variance of the prediction equation is to the theoretical lower bound of $q+1$. Keifer and Wolfowitz (39) showed that $q + 1$ is also the lower bound of the maximum variance for a D-optimum design over a continuous experimental region. When solving $P_2$ the maximum variance over the finite set of grid points will be $q + 1$. However, the maximum variance over the entire experimental region will be greater than $q + 1$ unless the D-optimum design has been found. Solving the problem

$$P_4: \max_{X \in \mathcal{E}} f(X)^t S^{-1} f(X)$$

where $S^{-1}$ is the covariance matrix of an optimum design over $n$ grid points and comparing this maximum with $q + 1$ will provide a measure of the nearness of the design to the D-optimum design. It should be observed, however, that $P_4$ does not in general have a convex or concave objective function. Solving $P_4$ will guarantee a local maximum point only. Therefore the problem must be solved several times in searching for the global maximum. The objective function is however convex for the additive model.
Once a solution to P4 is obtained it is possible to determine how close $|S|$ is to $|S_D|$ where $S_D$ is the D-optimum information matrix. The following fact will be used.

**Fact 4.**  $\text{TR} \ AB \geq m \ \frac{1}{|A|^m} \frac{1}{|B|^m}$ where $A$ and $B$ are positive-definite matrices of order $m$ (e.g. see 20).

Note that the $\text{TR} \ S^{-1}S_D$

$$ = \text{TR} \left\{ \sum_{i \in R_D} p_i^D f(X_i) f(X_i)^t \right\} S^{-1}$$

where $R_D = \{i \mid p_i^D > 0\}$

$$ = \sum_{j,k=0}^{q} \sum_{i \in R_D} p_i^D f_j(X_i) f_k(X_i) s_{jk}$$

$$ = \sum_{i \in R_D} p_i^D \sum_{j,k=0}^{q} f_j(X_i) f_k(X_i) s_{jk}$$

$$ = \sum_{i \in R_D} p_i^D f(X_i)^t S^{-1} f(X_i)$$

$$ = \sum_{i \in R_D} p_i^D \text{Max}_{X} \left[ f(X)^t S^{-1} f(X) \right]$$

$$ = \text{Max}_{X} f(X)^t S^{-1} f(X)$$
since
\[ \prod_{i \in R_D} p_i^D = 1. \]

Because
\[ \text{TR } S^{-1} S_D \geq (q+1) |S^{-1}| \frac{1}{q+1} |S_D| \frac{1}{q+1} \]
then
\[ |S_D| \leq \left[ \frac{\text{TR } S^{-1} S_D}{q+1} \right]^{q+1} \]
or
\[ |S| \geq \left[ \max_{X} \frac{q + 1}{f(X) S^{-1} f(X)} |S_D| \right]^{q+1} \]

A similar inequality has been derived by Kiefer (35). It is the author's opinion that the variance criterion will have more meaning to an experimenter than will the determinantal criterion. If \(|S| > t |S_D|, 0 < t < 1\), it is not clear how good a design \(S\) represents in an absolute sense. However, if
\[ \max_{X} f(X) S^{-1} f(X) - (q + 1) < \epsilon \]
then more insight is possible since the variance is a familiar and meaningful concept to an experimenter.

**Improving a Solution**

It should be clear that any new design determined by augmenting the initial grid with one or more new points
from the experimental region and applying the algorithm
to this new grid will have a determinantal value at least
as good as the old design. This follows from the fact
that the solution set to the old design is a subset of
the solution set to the new design.

Specifically, let \( S_r = \sum_{i \in R_r} p_i f(X_i) f(X_i)^t \)
represent an optimum design based upon \( n \) grid points
where \( R_r = \{ i \mid p_i > 0 \} \).

Then let
\[
S_{r+1} = \sum_{i \in R_r} p_i^{r+1} f(X_i) f(X_i)^t + p_{n+1}^{r+1} f(X_{n+1})^t
\]
represent an optimum design based upon the points of positive
proportionality for \( S_r \) and one additional grid point in the
experimental region. Then \(|S_{r+1}| \geq |S_r|\) since the design \( S_r \)
is a feasible solution to the new problem. The problem to
be addressed is how to choose the new grid point \( x_{n+1} \) so
that the best improvement can be obtained.

Fedorov (20) has developed a sequential procedure
for determining a continuous normalized design which con­
verges to a D-optimum design. Starting with an initial
design, the point of maximum variance in the experimental
region is found. A new design is constructed by allocating
a small proportionality to this new point and decreasing all
other positive proportions a corresponding amount. That is,
if

\[
\begin{pmatrix}
x_1, x_2, \ldots, x_n \\
p_1, p_2, \ldots, p_n
\end{pmatrix}
\]

represents the design at a given iteration, then

\[
\text{Max } f(X)^t S^{-1} f(X) = f(x_{n+1})^t S^{-1} f(x_{n+1})
\]

is found and the new design

\[
\begin{pmatrix}
x_1', x_2', \ldots, x_n, x_{n+1}' \\
(1-\alpha) p_1, (1-\alpha) p_2, \ldots, (1-\alpha) p_n, \alpha
\end{pmatrix}
\]

is determined where \(\alpha\) is chosen such that the new determinant is maximized. Fedorov proves convergence to the D-optimum design. Wynn (64) has developed an equivalent procedure in which successive runs are added to an initial design. Each new run is conducted at the maximum variance point.

While both approaches consist of sequentially augmenting existing designs neither is directly applicable to the methodology presented here. What is desired is a method of choosing a new point or points for inclusion in an augmented grid. Once a new grid is established, then an optimum design over this new grid is determined. This new design should then be closer to solving problem \(P_1\) than the previous design. Wynn's approach is concerned with adding sequentially one more experimental run in which there is no change made to the initial \(n\) runs. Fedorov determines a new design which is a convex
combination of the old design and the new experimental point.

The Procedure For Improving A Solution

The following theorem will be proved.

Theorem 4.

A necessary and sufficient condition for

\[ |S_{r+1}| > |S_r| \text{ is for } f(x_{n+1})S_r^{-1} f(x_{n+1}) > (q+1) \]

Proof.

The following inequality can be formed. First

\[
TR S_{r+1}S_r^{-1} = \sum_{j,k=0}^{q} s_{jk}^r s_{jk}^{r+1}
\]

\[
= \sum_{j,k=0}^{q} s_{jk}^r \sum_{i \in R_{r+1}}^{p_i} r+1
\]

\[
\times f_j(X_i)f_k(X_i)
\]

where

\[ R_{r+1} = \{ i \mid i \in R_r \text{ or } i = n+1 \} \]
\[
= \sum_{i \in R_{r+1}} p_i^{r+1} \sum_{j,k=0}^{q} f_j(X_i) f_k(X_i) s_{jk}^{r+1}
\]
\[
= \sum_{i \in R_{r+1}} p_i^{r+1} f(X_i)^t s_{r}^{-1} f(X_i)
\]
\[
= \sum_{i \in R_{r}} p_i^{r+1} f(X_i)^t s_{r}^{-1} f(X_i) + p_{n+1}^{r+1} f(X_{n+1})^t s_{r}^{-1} f(X_{n+1})
\]
\[
\leq \sum_{i \in R_{r}} p_i^{r+1} (q+1) + p_{n+1}^{r+1} f(X_{n+1})^t s_{r}^{-1} f(X_{n+1})
\]
\[
= (1 - p_{n+1}^{r+1}) (q+1) + p_{n+1}^{r+1} f(X_{n+1})^t s_{r}^{-1} f(X_{n+1})
\]

since
\[f(X_i)^t s_{r}^{-1} f(X_i) \leq q+1 \text{ for all } i \in R_{r}.
\]

Then using Fact 4,
\[
(q+1) \frac{1}{|s_{r+1}|} \frac{1}{|s_{r}^{-1}|} \frac{1}{q+1} \leq \text{TR } s_{r+1} s_{r}^{-1}
\]
or
\[
\frac{|s_{r+1}|}{|s_{r}|} \leq \left[ \frac{\text{TR } s_{r+1} s_{r}^{-1}}{(q+1)} \right]^{q+1}
\]
\begin{align*}
\log |a S^r (1-a) S^r_+| &> a \log |S_r| + (1-a) \log |S^r_+| \\
&= \log |S^r_+| \quad 0 < a < 1
\end{align*}

using Fact 1 in Chapter II. But

\begin{align*}
\log |a S^r (1-a) S^r_+| &= \log [a p^r_i + (1-a) p^{r+1}_i] f(X_i) f(X_i)^t \\
&\quad + (1-a) p^{r+1}_n f(X_n+1) f(X_n+1)^t
\end{align*}
= \sum_{i \in R_{r+1}} p_i f(X_i) f(X_i)^t = S^*.

However $S_{r+1}$ is the optimum matrix based upon the points in $R_{r+1}$. Thus $|S_{r+1}|^2 \leq |S^*|$; hence a contradiction.

Theorem 4 indicates that an improved design can always be obtained if a point in the experimental region can be found such that $f(X_{n+1})^t S_r^{-1} f(X_{n+1}) > q^+1$. If no such point exists then $S_r$ is the D-optimum matrix. Notice that it is not necessary for the point $X_{n+1}$ to be the point where $f(X_{n+1})^t S_r^{-1} f(X_{n+1})$ is maximum. Also the grid points, $X_i$, corresponding to $p_i^r = 0$ need not be included in finding $S_{r+1}$. This can reduce the number of decision variables considerably when solving P2.

There appears to be some justification for choosing the point of maximum variance as the next point to include in the grid. Fedorov and Wynn's approaches both lead to the greatest increase in the determinantal value if the point of maximum variance is chosen. For any given grid, the design found by optimizing over the grid will have a determinantal value as large as or larger than Fedorov's design based upon the same set of points. Thus choosing the point of maximum variance will maximize a lower bound on $|S_{r+1}|$. Finding the point of maximum variance provides a measure of how good the design actually is. As will be shown next, it will guarantee convergence to the D-optimum.
Convergence to the D-Optimum

Theorem 4 insures that a sequence is obtained such that

\[ |S_1| < |S_2| < \ldots < |S_r| < |S_{r+1}|. \]

This sequence is bounded above by the determinant of the D-optimum information matrix. Being a monotonic and bounded sequence it converges.

**Theorem 5.** \( \lim_{r \to \infty} |S_r| = |S_D| \)

**Proof.** (Based upon Fedorov's work)

First consider the matrix

\[ S_{r+1} = (1-\alpha) S_r + \alpha f(X^*) f(X^*)^t \]

where \( \max_{x} x^t S_r^{-1} f(x) = f(x^*)^t S_r^{-1} f(x^*) \)

and \( \alpha \) is chosen such that \( |S_{r+1}| \) is maximized. Fedorov (20) shows that

\[ |S_{r+1}| = \left[ \frac{f(x^*)^t S_r^{-1} f(x^*)}{q+1} \right]^{q+1} \frac{q}{f(x^*)^t S_r^{-1} f(x^*)} \]

However, \( |S_{r+1}| \geq |S_{r+1}^1| \) since the design leading to the matrix \( S_{r+1}^1 \) is in the solution set to the problem

\[ |S_{r+1}| = \max_{x} \sum_{i \in R_r} p_i^{r+1} f(X_i) f(X_i)^t + p^{r+1} f(x^*) f(x^*)^t \]
Subject to 
\[ \sum_{i \in R_r} p_i^{r+1} + p^{r+1} = 1 \]
\[ p_i^{r+1} \geq 0, \ p^{r+1} \geq 0 \]

Because the sequence \( \{ |S_r| \} \) converges, there exists an integer \( W \) such that for any \( \varepsilon > 0 \), \( |S_{r+1}| - |S_r| < \varepsilon \) for \( r > W \).

This implies
\[ |S_{r+1}^1| - |S_r| < \varepsilon \]
or
\[ \left[ \frac{f(X*)^t S_r^{-1} f(X*)}{q+1} \right] q+1 \left[ \frac{q}{f(X*)^t S_r^{-1} f(X*) - 1} \right]^q \frac{\varepsilon}{|S_r|} + 1. \]

Let \( F_r = f(X*)^t S_r^{-1} f(X*) \). Then the above inequality implies
\[ \frac{F_r^{q+1}}{(F_r-1)^{q+1}} \frac{(q+1)^{q+1}}{q^q} \]
or equivalently
\[ F_r \to q+1. \]

Therefore for any \( \gamma > 0 \), there exists an \( \varepsilon > 0 \), such that
\[ F_r - (q+1) < \gamma \]
Assume that the sequence does not converge to \( |S_D| \). Then there is a \( \sigma > 0 \) such that \( F_r - (q+1) > \sigma \). Choose \( \gamma < \sigma \) and the contradiction proves the theorem.
Summary of Procedure

I. Choose an initial grid of \( n \geq q + 1 \) distinct points within the experimental region.

II. Find the optimum design over the grid points.

III. Find \( \max_{X \in \mathbb{E}} f(X)S^{-1}f(X) = f(X^*)S^{-1}f(X^*) \) where \( S^{-1} \) is the covariance matrix of the design found in Step II.

IV. If \( f(X^*)S^{-1}f(X^*) = q + 1 + \delta, \delta > 0 \), stop. If not, define a new grid consisting of \( X^* \) and those \( X_i \) that are in the design found in Step II.

V. Return to Step II.

While the determinantal value is strictly increasing at each iteration, this does not necessarily imply that the value of the maximum variance is decreasing. Wynn (64) has derived an inequality, however, which shows that an upper bound on the maximum variance must decrease as the determinantal value increases. This inequality can be formed from Fedorov's relationship given above by observing that \( |S_D| > |S_{r+1}| \). Therefore

\[
\frac{P^q \Gamma_q}{(P_{r-1})^q (q+1)^{q+1}} \geq \frac{|S_D|}{|S_r|}
\]
CHAPTER V
APPLICATIONS

The methodology developed in Chapters II, III, and IV can be applied in a wide variety of experimental situations. The purpose of this chapter is to discuss and illustrate some of these applications. Several types of regression models in one independent variable are solved for their optimum design in a specified closed interval. A two-dimensional linear model defined over four different irregularly shaped experimental regions is considered. A second order polynomial is treated in an irregularly shaped experimental region, and a two-dimensional exponential function is optimized over the unit square.

The Single Variable Problem

A regression model in one independent variable is generally the easiest to solve. Any number of equally spaced grid points can be generated between any two end points on a real line segment. Using the algorithm, the optimum design over these grid points can be determined. Once an initial design is determined, then the point of maximum variance can be found by conducting a one-dimensional
search over the given interval. Further improvement may be obtained by including this new point in an augmented grid and applying the algorithm once more.

Table 1 is a list of six regression models and their respective experimental regions. For each model, \( n \) equally spaced grid points were generated in the specified interval. A starting vector giving \( p_i = 1/n, i = 1, 2, \ldots, n \) was used. The stopping criterion was

\[
\max_{i=1,2,\ldots,n} f(X_i)S^{-1}f(X_i) \leq (q+1) + \delta, \delta > 0.
\]

The maximum variance was found based upon the optimum covariance matrix over the \( n \) points by conducting a Fibonacci search. In order to determine the global maximum, the search was conducted over each subregion whose end points coincided with the points of support of the design. Based upon the global maximum and using the inequality derived in Chapter IV, page 48, \( t \) was found where \(|S| > t|S_D|\) and

\[
t = \frac{(q+1)^q+1}{\max_X f(X)S^{-1}f(X)}q+1.
\]

Table 2 summarizes the results.

In each model, based on a relatively small number of grid points, designs within 90 percent of the D-optimum design were found. As can be seen from Table 2, the models which were optimized over the larger number of grid points (21 vs 11) required as expected almost twice the number of iterations. In three of the models, a minimum point design
<table>
<thead>
<tr>
<th>MODEL</th>
<th>EXPERIMENTAL REGION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $E(Y) = B_0 + B_1 X + B_2 X^2$</td>
<td>$-1 \leq X \leq 1$</td>
</tr>
<tr>
<td>2. $E(Y) = B_0 + B_1 X + B_2 X^2 + B_3 X^3$</td>
<td>$0 \leq X \leq 2$</td>
</tr>
<tr>
<td>3. $E(Y) = B_0 + B_1 X + B_2 \ln X$</td>
<td>$1 \leq X \leq 2$</td>
</tr>
<tr>
<td>4. $E(Y) = B_0 + B_1 + \frac{B_2 + B_3}{X}$</td>
<td>$1 \leq X \leq 2$</td>
</tr>
<tr>
<td>5. $E(Y) = B_0 + B_1 \sqrt{X} + \frac{B_2}{X}$</td>
<td>$1 \leq X \leq 2$</td>
</tr>
<tr>
<td>6. $E(Y) = B_0 + B_1 X + B_2 \sin X$</td>
<td>$0 \leq X \leq 1$</td>
</tr>
</tbody>
</table>
## TABLE 2

**OPTIMUM REGRESSION DESIGNS**

| Model n | δ | Optimum Solution | $|S|$ | Max Var | t | Number of Iterations |
|---------|---|------------------|------|---------|---|---------------------|
| 1. 11 .05 | P(-1) = .328 | .148 | 3.05 | .952 | 25 |
|         | P(0) = .335 |      |      |      |    |
|         | P(1) = .337 |      |      |      |    |
| 2. 21 .05 | P(0) = .247 | 5.04x10^{-3} | 4.047 | .942 | 49 |
|         | P(.5) = .130 |      |      |      |    |
|         | P(.6) = .123 |      |      |      |    |
|         | P(1.4) = .117 |      |      |      |    |
|         | P(1.5) = .113 |      |      |      |    |
|         | P(2) = .250 |      |      |      |    |
| 3. 21 .05 | P(1) = .335 | 1.3x10^{-4} | 3.025 | .974 | 45 |
|         | P(1.4) = .113 |      |      |      |    |
|         | P(1.45) = .138 |      |      |      |    |
|         | P(1.5) = .083 |      |      |      |    |
|         | P(2.0) = .331 |      |      |      |    |
| 4. 21 .05 | P(1.0) = .25 | 3.045x10^{-10} | 4.07 | .933 | 49 |
|         | P(1.15) = .25 |      |      |      |    |
|         | P(1.55) = .22 |      |      |      |    |
|         | P(1.60) = .03 |      |      |      |    |
|         | P(2) = .25 |      |      |      |    |
| 5. 11 .05 | P(1) = .333 | 2.648x10^{-5} | 3.02 | .98 | 26 |
|         | P(1.4) = .333 |      |      |      |    |
|         | P(2) = .333 |      |      |      |    |
| 6. 11 .05 | P(0) = .333 | 1.323x10^{-4} | 3.026 | .975 | 25 |
|         | P(.6) = .333 |      |      |      |    |
|         | P(1) = .333 |      |      |      |    |
(q+1 points) was found to be the optimum. While it is known that single variable polynomials have a minimum point design as D-optimum, the cubic polynomial in Table 2 (Model 2 design) has six points of support. This is a result of the finite grid structure imposed upon the problem. In all likelihood, the D-optimum design will have points of support somewhere between .5 and .6 and 1.4 and 1.5. It is interesting to note that the maximum variance occurred at the point 1.44. In other words if one were to improve upon the design given by applying the procedure outlined in Chapter IV, then 1.44 would be included in the new design. However, it can at times be an advantage to have a design which is not a minimum point design. This could be the case when a polynomial of degree $k$ was hypothesized but a polynomial of degree $k+m$ better fit the data.

For comparison purposes, an arbitrary design was determined for each model. This design, called the standard design, is assumed to be a typical regression design which an experimenter might use. In each case the points of support are equally spaced. Table 3 outlines these standard designs. The closeness of the determinant (and maximum variance) of these designs with the designs given in Table 2 varies considerably. Choosing blindly such a design may result in a reasonably good design as in models three and six or in an exceptionally poor design as indicated by model 4. If the functions in the regression model are
| Model | Standard Design | $|S|$ | Max Var | $t$ |
|-------|-----------------|------|--------|-----|
| 1     | $P(\pm 1)=P(\pm .5)=P(0)=1/5$ | .0875 | 4.41   | .317 |
| 2     | $P(0)=P(.7)=P(1.4)=P(2)=1/4$ | $4.47 \times 10^{-3}$ | 4.85   | .463 |
| 3     | $P(1)=P(1.5)=P(2)=1/3$ | $1.28 \times 10^{-4}$ | 3.04   | .961 |
| 4     | $P(1)=P(1.25)=P(1.5)$ $=P(1.75)=P(2)=1/5$ | $1.72 \times 10^{-10}$ | 6.85   | .199 |
| 5     | $P(1)=P(1.5)=P(2)=1/3$ | $2.446 \times 10^{-5}$ | 3.267  | .781 |
| 6     | $P(0)=P(.5)=P(1)=1/3$ | $1.28 \times 10^{-4}$ | 3.11   | .90  |
highly non-linear, it would appear that the equally spaced designs are poor designs.

**Improving A Solution**

The procedure developed in Chapter IV will be illustrated using the model

\[ E(Y) = B_0 + B_1 \frac{\sqrt{X}}{X} + B_2 X^2 \exp(X), \quad 1 \leq X \leq 2. \]

Based upon 11 equally spaced grid points the optimum design was found in 11 iterations to be

- \( P(1) = .332 \)
- \( P(1.5) = .200 \)
- \( P(1.6) = .138 \)
- \( P(2) = .330 \)

with \( |S| = .2795 \). A one-dimensional search determined that the maximum variance was 3.032 occurring at the point 1.514. A new grid containing the points, 1, 1.5, 1.514, 1.6, and 2 was formed. The optimum design over these points is

- \( P(1) = .333 \)
- \( P(1.5) = .106 \)
- \( P(1.514) = .228 \)
- \( P(2) = .333 \)

with \( |S| = .2842 \). With a starting vector of \( P(1) = P(2) = .33, P(1.5) = P(1.6) = P(1.514) = .11 \), the algorithm converged in 5 iterations. A second Fibonacci search found the maximum variance now to be 3.002 (\( t = .997 \)).
A second example is given by the model \( E(Y) = B_0 + B_1 X + B_2 \exp(X) - 15X^2 - 1 \). With \( n = 11 \) and \( \delta = .01 \), the optimum design is

\[
\begin{align*}
P(1) &= .333 \\
P(2) &= .335 \\
P(1) &= .332
\end{align*}
\]

with \(|S| = .0459\). The maximum variance occurred at .15 and equaled 3.0015 (an almost D-optimum design). A new optimum design based upon the new grid consisted of \( P(-1) = P(.15) = P(1) = 1/3 \) with \(|S| = .04605\). The maximum variance of this new design is 3.001 occurring at .165. One more iteration yielded the design

\[
\begin{align*}
P(-1) &= P(.165) = P(1) = 1/3
\end{align*}
\]

with \(|S| = .04606\). The maximum variance is given by 3.00038. This final design is sufficiently close to the D-optimum for any practical situation.

The Multi-Variable Problem

Perhaps one of the main advantages of the methodology developed in this dissertation, is the ability to treat "messy" experimental regions. The standard design procedure is to consider the experimental region as a hypercube or hypersphere. For the few multi-variable regression models for which the optimum design has been found analytically, the experimental region must conform to such a hypercube or hypersphere. However, there are experimental situations
in which this is not the case (for example see 25). As an
illustration of the multi-variable problem, several examples
of regression models having two independent variables de­
fined over irregularly shaped regions are discussed. The
application to three or more variables is straightforward.

Consider first the additive model

\[ E(Y) = B_0 + B_1 X_1 + B_2 X_2 \]

The optimum design must occur at the extreme points
of the experimental region. Figure 1 illustrates four
experimental regions of interest. In order to initiate the
algorithm, a starting vector giving equal weight to each
extreme point was used. The stopping criteria was \( \text{Max Var} \)
(Y) \( \leq 3.01 \). Because the variance function is convex for the
additive model, the maximum variance throughout the experi­
mental region will not exceed 3.01. Table 4 lists the
optimum designs. Included are the determinantal values and
maximum variances for the design giving equal weight to each
extreme point (i.e., the initial design).

Because all the extreme points were included in the
set of candidate points, the solutions given in Table 4 are
solutions to problem P1. In the case of irregular experi­
mental regions, equal-weight designs are not necessarily
optimal. There may appear to be a tendency for minimum
point designs to be optimum. For example in regions (2)
and (4) most of the weight is assigned to three of the
Figure 1. Irregular Experimental Regions.
| Region | Optimum Design | $|S|$ | Max Var | Equal-Weight Design | $|S|$ | Max Var |
|--------|----------------|------|---------|---------------------|------|---------|
| 1      | $P(0,1) = .312$| .1685| 3.001   | $P(0,1) = .1563$    | .1563| 3.60    |
|        | $P(1,1) = .188$|      |         |                     |      |         |
|        | $P(1,0) = .188$|      |         |                     |      |         |
|        | $P(0,-1) = .312$|     |         |                     |      |         |
| 2      | $P(0,0) = .327$| .1336| 3.01    | $P(0,0) = .1016$    | .1016| 3.583   |
|        | $P(1,2) = .321$|      |         |                     |      |         |
|        | $P(3,1) = .051$|      |         |                     |      |         |
|        | $P(3,0) = .301$|      |         |                     |      |         |
| 3      | $P(1,0) = .333$| .1875| 3.005   | $P(1,0) = .162$     | .162 | 3.61    |
|        | $P(2,1) = .333$|      |         |                     |      |         |
|        | $P(.75,2) = .333$|   |         |                     |      |         |
| 4      | $P(2,.5) = .330$| .334 | 3.008   | $P(2,.5) = .254$    | .254 | 3.58    |
|        | $P(2,2) = .308$|      |         |                     |      |         |
|        | $P(1,2) = .036$|      |         |                     |      |         |
|        | $P(0,1) = .326$|      |         |                     |      |         |
extreme points. Region (2) is a minimum point design. Region (1) is not, however. Therefore, it is possible to conclude that the shape of a convex experimental region may require a non-minimum point design for an optimum. See Figure 1.

Figure 2 illustrates another experimental region. For the additive model, a design giving equal weight to each extreme point in Figure 2 has a determinantal value of .164 and a maximum variance of 3.97. A near optimum design was found to be

\[
P(1, -.2) = P(0,1) = P(-1,-.2) = 1/3
\]

having a determinantal value of .213 and a maximum variance of 3.003. Consider next, the additive model with an interaction term. That is

\[
E(Y) = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_1 X_2
\]

For this model equal weight on 12 candidate points denoted in Figure 2 provided the starting vector. The optimum design based upon these 12 points is

\[
\begin{align*}
P(1, -.2) &= .245 \\
P(1, .5) &= .236 \\
P(0, 1) &= .200 \\
P(-1,-.2) &= .244 \\
P(-.5,.5) &= .075
\end{align*}
\]
Figure 2. An Experimental Region.
The determinantal value is .0113. Note that most of the weight is given to four of the five extreme points. The variance is 4.009 (t = .99) occurring at the extreme point \((-1, -.2)\). This value was found using a non-linear programming algorithm based upon Rosen's Gradient Projection Method (53). This type of algorithm is necessary since the search for the maximum variance point in two or more dimensions requires solving for the maximum of a non-linear function in a constrained region. Rosen's Gradient Projection Method is considered to be efficient when all the constraints are linear. The algorithm was initiated at several different points within the experimental region in searching for the global optimum. For comparison, a minimum point design giving equal weight to the points \((1, -.2), (1, .5), (0,1),\) and \((-1, -.2)\) has a determinantal value of .0110. Finally, for the 2nd order model

\[
E(Y) = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_1 X_2 + B_4 X_1^2 + B_5 X_2^2,
\]

the optimum design based upon the same candidate points is

\[
\begin{align*}
P (1, -.2) &= .167 \\
P (1, .5) &= .166 \\
P (0, 1) &= .166 \\
P (-1, 0) &= .018
\end{align*}
\]
P (0, -.2) = .155
P (0, 0) = .011
P (0, -.2) = .157
P (-.5, .5) = .160

The determinantal value is \(1.67 \times 10^{-5}\). Based upon the covariance matrix of the above design, the maximum variance of the regression function within the experimental region was found to be 6.99 \((t = .397)\) at the point (.11, .25). A new grid was formed using the eight optimum points of support and the point (.11, .25). The new optimum design over these nine points is

\[
\begin{align*}
P (1, -.2) &= .165 \\
P (1, .5) &= .162 \\
P (0, 1) &= .164 \\
P (-1, 0) &= .024 \\
P (-1, -.2) &= .150 \\
P (0, -.2) &= .147 \\
P (-.5, .5) &= .147 \\
P (.11, .25) &= .041
\end{align*}
\]

The determinantal value increased to \(1.709 \times 10^{-5}\). The maximum variance decreased to 6.456 \((t = .591)\) occurring at the point (-.64, .35). Including this point in a new grid yielded the design

\[
\begin{align*}
P (1, -.2) &= .165 \\
P (1, .5) &= .162
\end{align*}
\]
\[ P(0, 1) = 0.166 \]
\[ P(-1,-0.2) = 0.162 \]
\[ P(0, -0.2) = 0.145 \]
\[ P(0.11, 0.25) = 0.046 \]
\[ P(-0.64, 0.36) = 0.154 \]

with \(|S| = 1.802 \times 10^{-5}\) and a maximum variance of 6.056 (\(t = 0.946\)).

Optimum designs in two or more dimensions for functions other than polynomials have not been in general characterized. However, they present no additional problems when the method outlined in this dissertation is applied. For example, the regression model

\[ E(Y) = B_0 + B_1 X_1 + B_2 X_1 \text{EXP}(X_2) + B_3 X_2^2 \]

is solved for its optimum design over the unit square based upon the 15 points shown in Figure 3. Beginning with a vector giving equal weight to each point, the algorithm converged in 52 iterations, to the solution

\[ P(1, -1) = P(-1, -1) = 0.115 \]
\[ P(1, 0) = P(-1, 0) = 0.185 \]
\[ P(1, 1) = P(-1, 1) = 0.200 \]

with \(|S| = 0.228\). Using Rosen's gradient projection method the problem,

\[
\text{Max} \quad (1, X_1, X_1e^{X_2}, X_2^2)S^{-1}
\]

\[-1 \leq X_1 \leq 1 \]
\[-1 \leq X_2 \leq 1 \]

becomes

\[
\begin{pmatrix}
1 \\
X_1 \\
X_1e^{X_2} \\
X_2^2
\end{pmatrix}
\]
$E(Y) = B_0 + B_1 X_1 + B_2 X_1 \exp(X_2) + B_3 X_2^2$

Figure 3. The Unit Square.
where $S^{-1}$ is the covariance matrix for the above design, was solved. The maximum variance was found to be 4.04 occurring at the points $(1, -.11)$ and $(-1, -.11)$.

In an attempt to improve upon this design, the algorithm was applied again with the following eight points as candidate points: $(± 1, ± 1), (± 1, -.11), (+1,0)$. The algorithm converged to the solution

\[
\begin{align*}
P (± 1, -1) &= .0999 \\
P (± 1,-.11) &= .1935 \\
P (± 1, 1) &= .2065
\end{align*}
\]

with $|S| = .233$.

Again using the gradient projection method, the maximum variance over the entire experimental region was found to be 4.001 ($t = .999$) occurring at the points $(1,1)$ and $(-1,-1)$. This design was considered to be sufficiently close to the D-optimum, and the procedure was terminated.

**Determining A Grid Size**

In order to obtain some insight as to the optimum number of grid points needed for initially approximating a D-optimum design, five different models were solved over four grid spacings. Table 5 lists the five regression models and Table 6 lists the maximum determinantal values for each model over each of the four spacings. Grids consisting of 6, 11, 21, and 41 points were generated.
TABLE 5

REGRESSION MODELS FOR GRID SIZE COMPARISONS

<table>
<thead>
<tr>
<th>MODEL</th>
<th>EXPERIMENTAL REGION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $E(Y) = B_0 + B_1 \text{Exp}(X) + \frac{B_2}{X \text{Exp}(2X)}$</td>
<td>$1 \leq X \leq 2$</td>
</tr>
<tr>
<td>2. $E(Y) = B_0 + B_1 + \frac{B_2}{X} + \frac{B_3}{X^2}$</td>
<td>$1 \leq X \leq 2$</td>
</tr>
<tr>
<td>3. $E(Y) = B_0 + B_1 X + B_2 X \text{Exp}(X) + B_3 X^2 \text{Exp}(2X)$</td>
<td>$1 \leq X \leq 1$</td>
</tr>
<tr>
<td>4. $E(Y) = B_0 + B_1 \sqrt{X} + B_2 X + B_3 X^2$</td>
<td>$1 \leq X \leq 3$</td>
</tr>
<tr>
<td>5. $E(Y) = B_0 + B_1 X + B_2 \text{Exp}(X) + B_3 \text{Exp}(2X) + B_4 \text{Exp}(3X)$</td>
<td>$1 \leq X \leq 1$</td>
</tr>
</tbody>
</table>
with equal spacing in each experimental region. This provided a spacing of .2, .1, .05, and .025 respectively between points defined in an interval of unit length (models 1 and 2) and a spacing of .4, .2, .1, and .05 respectively in an interval of length 2 (models 3, 4 and 5). This spacing insured that each previous set of grid points were included in the augmented set. If this were not the case, then it would be possible for a small number of grid points to yield a better design than a larger number because of the location of the points in the interval. The starting vector was \( p_1 = \frac{1}{N} \), and the stopping criterion was given by \( \delta = .05 \).

Table 6 shows that for all models except model 1, the determinantal value increased as the number of grid points increased. Model 1 however actually decreased slightly for the larger grid sizes. There were 3 points of support for the optimum design for model 1 over 6 and 11 points in the grid. By changing the stopping criterion to \( \delta = .01 \) and searching for maximum variance points, it was found that those three points were very close to the D-optimum points of support. As the grid size increased to 21 and 41 points, there was a tendency for other points in the grid near one of the original three to have small positive proportionalities. This had the effect of slightly reducing the determinantal value.
### TABLE 6

**DETERMINANTAL VALUES FOR FOUR DIFFERENT GRID SIZES**

<table>
<thead>
<tr>
<th>MODEL</th>
<th>6</th>
<th>11</th>
<th>21</th>
<th>41</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.022</td>
<td>2.022</td>
<td>2.003</td>
<td>2.015</td>
</tr>
<tr>
<td></td>
<td>(16)*</td>
<td>(28)</td>
<td>(24)</td>
<td>(30)</td>
</tr>
<tr>
<td>2</td>
<td>3.470 x 10^{-5}</td>
<td>3.551 x 10^{-5}</td>
<td>3.576 x 10^{-5}</td>
<td>3.599 x 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>(18)</td>
<td>(12)</td>
<td>(25)</td>
<td>(38)</td>
</tr>
<tr>
<td>3</td>
<td>5.366 x 10^{-2}</td>
<td>5.368 x 10^{-2}</td>
<td>5.751 x 10^{-2}</td>
<td>5.773 x 10^{-2}</td>
</tr>
<tr>
<td></td>
<td>(8)</td>
<td>(37)</td>
<td>(25)</td>
<td>(46)</td>
</tr>
<tr>
<td>4</td>
<td>8.085 x 10^{-7}</td>
<td>8.616 x 10^{-7}</td>
<td>8.760 x 10^{-7}</td>
<td>8.767 x 10^{-7}</td>
</tr>
<tr>
<td></td>
<td>(24)</td>
<td>(66)</td>
<td>(22)</td>
<td>(44)</td>
</tr>
<tr>
<td>5</td>
<td>6.361 x 10^{-6}</td>
<td>1.187 x 10^{-5}</td>
<td>1.241 x 10^{-5}</td>
<td>1.244 x 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>(22)</td>
<td>(37)</td>
<td>(68)</td>
<td>(90)</td>
</tr>
</tbody>
</table>

*Numbers in parenthesis indicate the number of iterations.*
While the determinantal value increased for each larger set of grid points for the other four models, the amount of change from 21 to 41 points was not large. It appears that once grid points are included that are "close" to the D-optimum points of support, then there is not much benefit to be gained by making the grid structure finer. In fact, as model 1 illustrates, it can have a negative effect by forcing points near the D-optimum points of support to share small positive proportions and thus decrease the determinantal value. Reducing the stopping criterion may eliminate this effect to some extent but at the expense of increasing the number of iterations.

The best procedure would appear to be that of choosing a rather coarse grid structure, optimizing over that grid, and then use the procedure outlined in Chapter IV to improve the solution. As Table 2 shows, for the single-variable regression model, grids consisting of 11 and 21 points can lead to designs having a determinantal value greater than 90 percent of the D-optimum value.
CHAPTER VI

CONCLUSIONS AND EXTENSIONS

In the preceding chapters a procedure has been developed for finding optimum or near-optimum experimental designs based upon the determinantal criterion. A disadvantage of optimum design has been the difficulty in many cases of solving for the optimum regardless of the criterion used. It is hoped that this procedure at least in part overcomes this difficulty. A second disadvantage has been the mathematical sophistication required to understand and apply much of the theory that supports the concept of optimal experimental design. The approach taken in this dissertation has been to formulate and solve an optimum design problem as a mathematical program or as a series of mathematical programs. It is felt that the theory of mathematical programming is rather well developed and widely understood. Therefore, the hope is that an experimenter who might have some difficulty in directly applying optimal design criteria to his specific problem can, nevertheless, take advantage of the approach outlined in the previous chapters and find a good if not best design that suits his needs.
One suspects that the approach outlined here will be of the most benefit in those experimental situations involving costly or limited resources. It is these situations in which it is desirable to obtain the greatest amount of information (minimize uncertainty) per unit of resource. Uncertainty expressed in terms of variances are very dependent upon the spacing of the observations as indicated by some of the examples in Chapter V.

Perhaps one criticism of the method outlined here is the requirement to construct some sort of grid from which an optimum design is chosen. It should be pointed out, however, that in any regression analysis, the experimenter must determine a finite set of points at which the experiment is to be conducted. In addition, the grid size does not need to be large in order to obtain a reasonably good design.

Some Extensions

One of the objectives in developing the methodology presented here was to enable a large range of design problems to be treated using essentially the same basic concepts. By making some relatively simple modifications to the algorithm, it is possible to consider several important but more difficult problems.

A generalization of the regression model discussed in the previous chapters is the weighted least squares problem. It is not uncommon in some situations for the
variances of the observations to be unequal. A more
general class of problems can be defined in which the co-
variance matrix of the error terms is diagonal with unequal
elements. It is assumed that these variances are known
except for the constant $\sigma^2$ or are functions of the inde-
pendent variable(s). The changes required in the algorithm
to handle this problem are discussed in Appendix A.

If an experiment has already been conducted in a
non-optimum manner with a total of $M$ observations at $m$
distinct points, the problem may arise as to how to best
augment this initial design. That is $N$ additional experi-
mental runs is desired. The augmentation problem is to
choose these $N$ additional runs in such a manner that the
total design of $N+M$ observations is the best possible.
Appendix B outlines one approach to this problem based
upon the procedure developed in this dissertation.

**Areas For Further Research**

As indicated in Chapter I, a disadvantage of most
design criteria is that they are model dependent. Many
times an experimenter must hypothesize a mathematical
relationship between a dependent and one or more independent
variables. However, until an experiment is carried out and
the resulting data analyzed, the experimenter cannot statisti-
cally confirm that his model was a good one. In planning
for an experiment, it may be desirable to account for the
fact that the hypothesized model may not adequately fit the
data. A method is presented in Appendix C that enables a design to be found by weighing the determinant of the $X^T X$ matrix of two or more hypothesized models and maximizing the sum of these weighted determinants. A disadvantage of this approach is the fact that the resulting design may not be optimum for any of the hypothesized models. However, if uncertainty exists as to the correct model, it would appear that this method would provide a reasonably good design for any of the models considered.

As indicated in Chapter IV, Fedorov (and Wynn) has developed a procedure for constructing D-optimum designs. Because Fedorov's approach is of a general nature, as is the one described in the previous chapters, it would be of value for a comparison to be made between the two. While both methods require searches for maximum variance points, it is felt that the approach in this dissertation will determine the optimum points of support much more quickly. This is based on the fact that if both procedures begin with the same initial design, the procedure outlined here will by optimizing over the initial design points begin with a covariance matrix that is much closer to the D-optimum covariance matrix. That is, it is felt that this procedure will converge much more quickly; although, in all fairness, more work is required at each iteration to find the optimum design over the set of grid points. In addition, Fedorov's method does not allow any
points of support to leave the design once it becomes part of it. There is a tendency for points of support to appear that are clustered in groups with small proportions assigned to these points. Therefore, some heuristics are introduced to round-off these designs and reduce the number of support points.

While no attempt has been made in Chapter III to find the most efficient algorithm (and computer program) to solve $P_{2'}$, it is certainly likely that a better algorithm is possible. Advances in unconstrained optimization are continuously being made. The objective in developing the algorithm presented here was to indicate how the problem can be solved by taking advantage of its structure. Although considerable success has been achieved with this algorithm, perhaps another method of solving $P_{2'}$ will enable even more efficient solution of quite large problems.

Finally, a generalization of the problem formulated in Chapter I (Problem $P_1$) is to introduce an additional set of constraints given by

$$g_t(M_1, \ldots, M_n) \leq 0 \quad t = 1, 2, \ldots, T$$

where

$$M_i = N p_i, \quad i = 1, 2, \ldots, n.$$ 

These constraints would allow resource limitations on the experimental runs to be explicitly considered. The
approach taken in this dissertation should provide some insight into solving this more general problem.
APPENDIX A

WEIGHTED LEAST SQUARES MODELS
APPENDIX A
WEIGHTED LEAST SQUARES MODELS

Assume that for the linear model \( Y = X \beta + e \),
\[
\text{COV}(e) = \sigma^2 \begin{bmatrix} V_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & V_n \end{bmatrix} = \sigma^2 V
\]
where \( V_i \) are known constants or are functions of the independent variables. Then \( \text{COV}(\beta) = (\bar{X}^t V^{-1} \bar{X})^{-1} \frac{\sigma^2}{N} \) and
the determinantal criterion is to
\[
\text{MAX } |\bar{X}^t V^{-1} \bar{X}|.
\]
But
\[
\bar{X}^t V^{-1} \bar{X} = \sum_{i=1}^{n} \frac{p_i}{V_i} f(X_i) f(X_i)^t
\]
and
\[
\text{LOG } |\sum_{i=1}^{n} \frac{p_i}{V_i} f(X_i) f(X_i)^t|
\]
is concave based on Theorem 1. The Kuhn-Tucker conditions for this problem are
(1) \( \frac{1}{V_i} f(X_i)^t S^{-1} f(X_i) + w_i = z \quad i = 1, 2, \ldots, n \)

(2) \( \sum_{i=1}^{n} p_i = 1 \)

(3) \( p_i \geq 0 \)

(4) \( p_i \geq 0 \)

(5) \( p_i w_i = 0 \)

where \( \{ s_{jk} \} = X^t V^{-1} X \).

It follows that

\[
\sum_{i=1}^{n} \frac{p_i}{V_i} f(X_i)^t S^{-1} f(X_i) = \text{TR I}_{q+1} = q+1 = z
\]

then for \( p_i > 0 \),

\[
\frac{1}{V_i} f(X_i)^t S^{-1} f(X_i) = q + 1
\]

To apply the algorithm to this model requires that equation (3-2) be changed to

\[
s_{jk} = \sum_{i=1}^{n} \frac{p_i}{V_i} [f_j(X_i)f_k(X_i) - f_j(X_n)f_k(X_n)]
\]

\[
+ f_j(X_n)f_k(X_n)
\]

and equation (3-3) be changed to
The stopping criterion is given by

$$\forall f(p_i, \ldots, p_{n-1}) = \sum_{j=0}^{q} [f_j(X_i)f_k(X_i) - f_j(X_n)f_k(X_n)] \times \frac{s_{jk}}{V_i}$$

for $i = 1, 2, \ldots, n-1$.

The stopping criterion is also given by

$$\max_{i} \left\{ -\frac{1}{V_i} f(X_i)^t S^{-1} f(X_i) \right\} \leq q + \delta, \; \delta > 0.$$
APPENDIX B

THE AUGMENTATION PROBLEM
APPENDIX B
THE AUGMENTATION PROBLEM

Assume that an experiment has been conducted in a non-optimal fashion at \( m \) distinct points in an experimental region. A total of \( M \) observations are taken over the \( m \) points. It is desired to augment this design with \( N \) additional experiments such that the determinant of the resulting information matrix is maximized.

If \( \bar{X}_o \) represents the initial design matrix, then the information matrix of the augmented design can be written as

\[
\bar{X}_o^t \bar{X}_o + N \bar{X}^t \bar{X}.
\]

If a grid structure consisting of \( n \) distinct points is established as a basis for determining the augmented design, then the \( j, k \)th element of the above matrix can be written as

\[
s_{jk} = a_{jk} + \sum_{i=1}^{n} p_i f_j(X_i)f_k(X_i)^t
\]

where

\[
a_{jk} = M \sum_{i=1}^{m} p_{io} f_j(X_{io})f_k(X_{io})^t \text{ and } X_{io}, i=1, 2, \ldots, m \text{ are } m \text{ points of support in the original design.}
\]

The optimization problem is

\[
\text{MAX} \sum_{i=1}^{n} p_i \log|A + N \sum_{i=1}^{n} p_i f(X_i)f(X_i)^t|
\]

subject to

\[
\sum_{i=1}^{n} p_i = 1
\]
where $A = \{a_{jk}\}$.

This function is concave based upon Theorem 1 in Chapter II. The only change is that a constant is being added to each element of the matrix $N \bar{X}^t \bar{X}$. That is each element of the resulting matrix is still a linear combination of $p_i$. The gradient of this function is

$$N f(X_i)^t S^{-1} f(X_i) = N \sum_{j,k=0}^{q} f_j(X_i) f_k(X_i) s^{jk}$$

where $S = A + N \sum_{i=1}^{n} p_i f(X_i)f(X_i)^t$

In applying the algorithm in Chapter III, equation (3-2) is replaced by

$$s_{jk} = a_{jk} + N \sum_{i=1}^{n-1} p_i [f_j(X_i)f_k(X_i) - f_j(X_n)f_k(X_n)]$$

$$+ N f_j(X_n)f_k(X_n),$$

and $s^{jk}$ in equation (3-3) is replaced by $N s^{jk}$ where

$\{s^{jk}\} = [A + N \sum_{i=1}^{n} p_i f(X_i)f(X_i)^t]^{-1}$

The stopping criterion must change since the lower bound on $\max_{i=1,2,\ldots,n} f(X_i)^t S^{-1} f(X_i)$ is no longer $q+1$.

One disadvantage of this problem formulation is that $N$ must be specified (unless $N = M$). The optimum
augmented design depends upon the number of additional observations that are to be taken. The dependence of \( N \) can be shown by the following inequality which follows from the concavity of the logarithm of the determinant.

\[
\log|A+N \bar{x}^t \bar{x}| = \log N^{q+1} |\frac{1}{N} A + \bar{x}^t \bar{x}| \\
= \log N^{q+1} + \log |\frac{1}{N} (A+ \bar{x}^t \bar{x}) + (1-\frac{l}{N}) \bar{x}^t \bar{x}| \\
> \log N^{q+1} + \log |A + \bar{x}^t \bar{x}| + (1-\frac{l}{N}) \log |\bar{x}^t \bar{x}|
\]

when \( \bar{x}^t \bar{x} \) is positive - definite.

Therefore for large \( N \), a reasonably good design can be obtained by finding

\[
\max \quad |\bar{x}^t \bar{x}|
\]

\( p_i > 0 \)

\[
\sum_{i=1}^{n} p_i = 1
\]
APPENDIX C

THE MODEL DISCRIMINATION PROBLEM
APPENDIX C

THE MODEL DISCRIMINATION PROBLEM

Assume that an experimenter is entertaining two different models for a regression analysis. He desires to choose a design that will provide good estimation regardless of which model is the correct one. Based upon the determinantal criterion, a reasonable objective might be to choose a design such that it maximizes

\[ k|\bar{X}_1^t \bar{X}_1| + (1-k) |\bar{X}_2^t \bar{X}_2| \quad 0 < k < 1 \]

where \( \bar{X}_1 \) is the design matrix corresponding to Model 1, and \( \bar{X}_2 \) is the design matrix corresponding to Model 2. The constant \( k \) reflects the weight (probably subjective) given to each model.

The objective function can be made concave by taking the logarithm of the function. That is

\[ \text{LOG}_e [k|\bar{X}_1^t \bar{X}_1| + (k-1) |\bar{X}_2^t \bar{X}_2|] \]

\[ \geq k \text{LOG}_e |\bar{X}_1^t \bar{X}_1| + (1-k) \text{LOG}_e |\bar{X}_2^t \bar{X}_2| \]

since the logarithm is itself a concave function. The concavity of the \( \text{LOG}_e |\bar{X}^t \bar{X}| \) was shown in Chapter II, and the sum of two concave functions is itself concave.
While it is possible to maximize this function directly, it is much simpler to maximize the right hand side of the above inequality. If a set of grid points is established, then the problem can be formulated as follows:

$$\text{MAX} \quad k \log_e | \sum_{i=1}^{n} p_i f_1(X_i) f_1(X_i)^t |$$

$$p_i \geq 0$$

$$+ (1-k) \log_e | \sum_{i=1}^{n} p_i f_2(X_i) f_2(X_i)^t |$$

$$\sum_{i=1}^{n} p_i = 1$$

$$0 < k < 1$$

where $f_1(X)$ and $f_2(X)$ are $q_1+1$ and $q_2+1$ vectors of the functionals of Model 1 and Model 2 respectively.

After the experiment has been performed, then one of the two models hypothesized will be used. It would seem reasonably therefore to include in the grid the optimum points of support for both models if known.

From the Kuhn-Tucker conditions for this problem, it can be shown that for $p_i > 0$,

$$k f_1(X_i)^t S_1^{-1} f_1(X_i) + (1-k) f_2(X_i)^t S_2^{-1} f_2(X_i)$$

$$= k q_1 + (1-k) q_2 + 1$$

where $S_1^{-1}$ and $S_2^{-1}$ are the covariance matrices for Models 1 and 2 respectively.
Applying the algorithm in Chapter III to this problem, equation (3-2) is replaced by two equations:

\[ s_{jk}^1 = \sum_{i=1}^{n-1} p_i [f_{jl}(x_i)f_{kl}(x_i)-f_{jl}(x_n)f_{kl}(x_n)] \]

\[ + f_{jl}(x_n)f_{kl}(x_n) \]

and

\[ s_{jk}^2 = \sum_{i=1}^{n-1} p_i [f_{j2}(x_i)f_{k2}(x_i)-f_{j2}(x_n)f_{k2}(x_n)] \]

\[ + f_{j2}(x_n)f_{k2}(x_n). \]

Equation (3-3) is replaced by

\[ \beta f(p_1, \ldots, p_{n-1}) = k \sum_{j,k=0}^{q} \beta p_i \]

\[ -f_{jl}(x_n)f_{kl}(x_n)s_{jk}^1 \]

\[ + (1-k) \sum_{j,k=0}^{q} [f_{j2}(x_i)f_{k2}(x_i)-f_{j2}(x_n)f_{k2}(x_n)] s_{jk}^2 \]

the algorithm may be terminated when the weighted sum of the variance functions is equal to or less than \( k q_1 + (1-k) q_2 + 1 + \delta, \delta > 0. \) This procedure can be generalized to include more than two hypothesized models.
EXAMPLE

In order to illustrate the behavior of the objective function defined above, an analytical example will be given. Consider the two hypothesized models given below:

**MODEL 1**

\[
Y = B_0 + B_1 X \\
-1 \leq X \leq 1
\]

It is known that the optimum design for Model 1 is \( P(-1) = P(1) = \frac{1}{2} \) and for Model 2 is \( P(-1) = P(0) = P(1) = \frac{1}{3} \).

Therefore the points \(-1, 0, 1\) will form the grid points.

The \( X^t X \) matrix for each model is

\[
\begin{bmatrix}
1 & p_1 - p_3 \\
p_1 - p_3 & p_1 + p_3
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & p_1 - p_3 & p_1 + p_3 \\
p_1 - p_3 & p_1 + p_3 & p_1 - p_3 \\
p_1 + p_3 & p_1 - p_3 & p_1 + p_3
\end{bmatrix}
\]

where \( p_1 \) is proportion of runs conducted at \(+1\), \( p_2 \) at \( 0 \), and \( p_3 \) at \(-1\). Then
The objective function is given by

\[ k \log_e |\bar{x}_1^t\bar{x}_1| + (1-k) \log_e |\bar{x}_2^t\bar{x}_2| ; \quad 0<k<1 \]

Taking derivatives,

\[ \frac{\partial L}{\partial p_1} = -2k(p_1-p_3) + (1-k) \frac{[2(1-p_2)(p_1-p_3) - 2(p_1-p_3)]}{|\bar{x}_1^t\bar{x}_1|} \]

\[ \frac{\partial L}{\partial p_2} = -k + (1-k) \frac{[-2(1-p_2) - (p_1-p_3)^2 + 3(1-p_2)^2]}{|\bar{x}_2^t\bar{x}_2|} \]

\[ \frac{\partial L}{\partial p_2} = 2k(p_1-p_3) + (1-k) \frac{[-2(1-p_2)(p_1-p_3) + 2(p_1-p_3)]}{|\bar{x}_1^t\bar{x}_1|} \]

If \( p_1 = p_3 \), then \( \frac{\partial L}{\partial p_1} = \frac{\partial L}{\partial p_2} = 0 \)

and \( \frac{\partial L}{\partial p_2} = -k \frac{1-p_2}{1-p_2} + (1-k) \frac{[-2(1-p_2) + 3(1-p_2)^2]}{(1-p_2)^2} - (1-p_2)^3 \)

setting \( \frac{\partial L}{\partial p_2} = 0 \) and solving for \( p_2 \):

\[ p_2 = \frac{1-k}{3-2k} \]
Thus if \( k = 1 \), \( p_2 = 0 \) and if \( k = 0 \), \( p_2 = \frac{1}{3} \) as expected. For \( 0 < k < 1 \) then \( 0 < p_2 < \frac{1}{3} \). As \( k \) decreases, more weight is placed on the origin while the balance of the runs is equally shared between \(+1\) and \(-1\).
APPENDIX D
GLOSSARY

CANDIDATE POINTS (OR GRID POINTS)

A finite, discrete set of points in an experimental region over which an approximation to the D-optimum design is found.

D-OPTIMUM DESIGN

The design matrix $\bar{X}$ which maximizes $|\bar{X}^t\bar{X}|$ over a given continuous experimental region.

EXACT PROBLEM

Solving for the D-optimum design where the number of repeated runs at each experimental point is integer valued for a given number of total experiments.

EXPERIMENTAL REGION

A subset of euclidean N-space which defines the set of values the independent variables in the regression model can assume.

INFORMATION MATRIX

The matrix $\bar{X}^t\bar{X}$ where $\bar{X}$ is the design matrix.
MINIMUM POINT DESIGN

A design having the same number of distinct points of support as there are unknown parameters in the regression model. That is, the minimum number of points necessary for a full rank design matrix.

NON-INTEGER PROBLEM

Solving for the D-optimum design where the number of repeated runs at each experimental point is not necessarily integer valued.

OPTIMUM DESIGN

The design matrix $\bar{X}$ which maximizes $|\bar{X}^t\bar{X}|$ over a set of candidate points.

POINTS OF SUPPORT

A distinct set of points in the experimental region which determine where one or more experimental runs are to be conducted for a given design.

PROPORTIONALITY MEASURE

A set of values between zero and one and summing to one which indicates the percent of experimental runs to be conducted at each point of support of a given design.
VARIANCE FUNCTION

A function defined over the experimental region which provides the variance of the estimated dependent variable of the regression model at each point in the region.
BIBLIOGRAPHY
BIBLIOGRAPHY


<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title and Details</th>
</tr>
</thead>
</table>