D-optimal Designs in Linear Regression Models

by
Chang, Hsing-Ming

A thesis submitted to the Faculty of Graduate Studies of
the University of Manitoba at Winnipeg
in partial fulfillment for the degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

D-optimal Designs in Linear Regression Models

BY

Chang, Hsing-Ming

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

Of

MASTER OF SCIENCE

ABSTRACT

CHANG, HSING-MING: D-optimal Designs in Linear Regression Models. (Under the direction of Dr. S., Mandal.)

'Linear regression models' unarguably is the most studied and used topic in applied statistics and other applications. In the aspect of experimental design, once the choice of the model and the optimality criterion are identified, along with specified variable and parameter constraints, optimal experimental design can be used to find the most efficient combination set of design variables and the allocation of the number of trials at the support points.

Tracing back to 1918, K. Smith was the first to consider optimal design problem. Later some of the most profound optimal design theoretical results were explored and published by Kiefer (The General Equivalence Theory, 1959).

In this thesis we first have a broad review of the important optimal design theories and some major optimality criteria. Also the properties of different criteria are discussed. In Chapter 2, a brief introduction is given on types of models describing response and design variables. A simple numerical example of finding optimal design is demonstrated. In Chapter 3, optimality conditions and a class of algorithms are discussed. The optimality conditions are based on directional derivatives of the criterion function of interest. The class of multiplicative algorithms neatly submit to the constraints of the general problem we consider. In Chapter 4, some examples of constructing D-optimal designs for linear regression models are demonstrated. The optimal designs are constructed by using the class of multiplicative algorithms and optimality

conditions discussed in Chapter 3. The iteration results are reported for each example. Some final remarks and further research directions are concluded in the final chapter with references which greatly contributed to the process and completion of this thesis.

ACKNOWLEDGMENTS

Honors and special thanks giving to Him who is always there for me wherever I go. "In nothing be anxious, but in everything, by prayer and petition with thanksgiving, let your requests be made known to God;" (Philippians 4:6). Also to my parents, an ocean of ink is not enough to describe their unlimited support to me in life.

I am deeply grateful to my direct supervisor, Dr. Saumen, Mandal, whose support, teaching, suggestions and encouragement helped me in all the time of learning and writing of this thesis. I am also very thankful to the members of my M.Sc. committee, Dr. S. Mandal, Dr. M. Samanta, and Dr. R. K. Thulasiram for monitoring my work and took effort in reading and providing me with valuable comments. I would like to extend my appreciation to all the professors and instructors of the Department of Statistics at the University of Manitoba for teaching all the wonderful courses. I would also like to offer personal thanks to Margaret for taking care all of our administrative work.

At last, but not the least, I would like to extend my gratitude especially to Dr. Fu and my classmate Yanqing Yi for their constant encouragement and curricular and extracurricular guidance. Many thanks.

Table of Contents

Chapter 1

Introduction

A design of experiment is a structured and organized method widely employed by researchers/scientists to answer questions of many varieties. Typically the goal of an experiment is to determine the relationship between a process and the response variable(outcome) of the process and some factors and interactions between factors quantitatively. The "Design" of an experiment refers to the consideration and structure of the experiment, especially with reference to:

1. the experimental factors(treatment combinations) and their values/levels included in the study
2. the types and number of experimental units included in the study
3. the randomization rules/procedures by which the experimental units are assigned to treatments(or vice versa)
4. the type of measurements made on the experimental units for the purpose of the interests

The optimal design theory was first developed for the linear models. In this thesis we will start with some general descriptions of optimal design theory mainly for linear

models. Then we continue to review some fundamental concepts of optimal design theory and further give some illustration on how to device specifically the optimal design and some practical examples of using optimal designs in experimental design setting.

Optimal designs are useful when certain criteria in the experiment are of importance and/or the standard classical designs are not suitable for two major reasons:

1. standard full or fractional factorial designs require too many runs for the amount of time or resources allocated for the experiment
2. the design space is constrained(or certain factor levels, therefore treatments, are not available)

Some examples are given in Neter, Kutner, Nachtsheim and Wasserman (1996). We may first look at a simple probability model:

$$y \sim g(y|\underline{x}, \underline{\theta}, \sigma) \tag{1.1}$$

where y is the response variable in an experiment, and a random error term, given that:

- $\underline{x} = (x_1, x_2, \cdots, x_m)^T$ are the explanatory variables, input values or design variables in an experiment usually known or selected by the experimenter, $x \in \mathcal{X} \subseteq \mathbb{R}^m$; $\mathcal{X}$ is called the design space;
- $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$;
- σ is a nuisance parameter, which is unknown but not of the main interest when constructing an optimal design;

- $g(.)$ is a distribution model of y and is a function of ($\underline{x}$,θ,σ).

Experimenters are free to vary the experimental conditions by selecting different $\underline{x}$ from the design space($\mathcal{X}$) and observe the changes in the response $y = y(\underline{x})$ which is a random variable with variance $var(y(\underline{x})) = \sigma^2$. σ^2 is generally assumed to be independent of the input variable $\underline{x}$.

Linear regression design models which are linear in the unknown parameter coefficients $\underline{\theta}$, have expected value of $y(\underline{x})$ in an explicit form:

$$\begin{aligned} E[y|\underline{x},\underline{\theta},\sigma] &= \underline{f}^T(\underline{x})\underline{\theta} \\ &= \theta_1 f_1(\underline{x}) + \theta_2 f_2(\underline{x}) + \ldots + \theta_k f_k(\underline{x}) \end{aligned} \tag{1.2}$$

where $\underline{f}^T(\underline{x}) = (f_1(\underline{x}), f_2(\underline{x}), \ldots, f_k(\underline{x}))^T$ is a k dimensional vector of real-valued function on design space $\mathcal{X}$.

From above, one may say that in order to construct an optimal experimental design, the experimenter must first specify the following:

1. the model
2. the feasible treatment combinations region
3. the total number of trials/runs allowed for the experiment, n
4. a specific optimal design criterion for selecting treatment combinations from the feasible region and for allocating experimental units n to the selected treatments

Knowing that only n trials are allowed due to limited time and resources, the problem now would be at what values of $\underline{x}$(levels of factors) should we make observations

on the response y in order to have as accurate and precise as possible of an inference on the parameters $\underline{\theta}$. Such a search for the best selection of treatments or the allocation of n trials to the selected treatments would be the focus of the construction of optimal experimental designs in this thesis.

In linear regression models $y = f^T(\underline{x})\underline{\theta} + \epsilon$, or in expectation:

$$E[Y] = X\underline{\theta} \qquad D[Y] = \sigma^2 I_n \tag{1.3}$$

where $Y = (y_1, y_2, \ldots, y_n)^T$, X is a $n \times k$ matrix with the $(i,j)^{th}$ element being $f_j(\underline{x}_i)$, and it can be easily derived that the best linear unbiased estimate $\hat{\underline{\theta}}$ of $\underline{\theta}$ is the solutions of linear equations:

$$(X^T X)\hat{\underline{\theta}} = X^T Y \tag{1.4}$$

in which the dispersion matrix of $\hat{\underline{\theta}}$, $D(\hat{\underline{\theta}}) = E[(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})^T]$, is minimized.

Since the experimenter selects X, one must ensure that the $k \times k$ matrix $(X^T X)$ is non-singular meaning that the inverse matrix $(X^T X)^{-1}$ exists and therefore there is a unique solution for $\hat{\underline{\theta}}$:

$$\hat{\underline{\theta}} = (X^T X)^{-1} X^T Y \tag{1.5}$$

with $E[\hat{\underline{\theta}}] = \underline{\theta}$ and $D(\hat{\underline{\theta}}) = \sigma^2 (X^T X)^{-1}$

From above we can see that the dispersion matrix $D(\hat{\underline{\theta}})$ does not depend on $\underline{\theta}$ and is proportional to the matrix $(X^T X)^{-1}$. In order to obtain better inference on $\underline{\theta}$, we can select X such that $(X^T X)^{-1}$ is minimized or $X^T X$ is large, so that $D(\hat{\underline{\theta}})$ is as small as possible. However, if the design space is continuous, we first need to discretize the design space.

1.1 Discretizing the Design Space

The linear regression design model, $E[y|\underline{x}, \underline{\theta}, \sigma] = \underline{f}^T(\underline{x})\underline{\theta}$, can be written as

$$E[y|\underline{v}, \underline{\theta}, \sigma] = \underline{v}^T\underline{\theta} \tag{1.6}$$

where $\underline{v} = (f_1(\underline{x}), f_2(\underline{x}), \ldots, f_k(\underline{x}))^T, v \in \mathcal{V}$ and $\mathcal{V} = (\underline{v} \in \mathbb{R}^k : \underline{v} = (f_1(\underline{x}), f_2(\underline{x}), \ldots, f_{\underline{k}}(\underline{x}))^T)$, $x \in \mathcal{X}$).

Intuitively choosing a design condition x in the restricted design space $\mathcal{X}$ is equivalent to choosing a k-element vector $\underline{v}$ in the closed bounded k-dimensional space $\mathcal{V} = \underline{f}(\mathcal{X})$. $\mathcal{V}$ is the image of $\mathcal{X}$ under the vector valued function $\underline{f} = (f_1, f_2, \ldots, f_k)^T$. $\mathcal{V}$ is called the induced design space(or design locus).

In theory the design space is usually assumed to be continuous, but for practical purposes, $\mathcal{V}$ can be assumed to be discrete, which is more close to applications in real problems. A "justification" for this can be seen in Mandal (2000, Theorem 1.3.1: Carathéodory's Theorem). This theorem says that for any experimental design p with more than $[\frac{k(k+1)}{2}+1]$ support points and information matrix $M(p)$, it is always possible to find a design p with a support of at most $[\frac{k(k+1)}{2}+1]$ points. Therefore any continuous optimal design measure can be replaced by at least one finite discrete probability distribution.

Now to obtain an observation y, one must first choose a vector $\underline{v}$ from the discrete design space $\mathcal{V}$ consisting of J distinct vectors, $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_J$. For this we may assume that choosing a $\underline{v}$ can be done without error so that the error-in-variable problem is not a concern.

Two simple but important design problems may be defined in search for an optimal design experiment:

1. At which of the points, or vertices, $\underline{v}_j, j = 1, 2, \ldots, J$, should we obtain a response y on.

2. With limited time and resources, suppose we can only obtain n observations, we must decide how many observations, n_j, to obtain at $\underline{v}_j$, such that $0 \leq n_j \leq n$, $j = 1, 2, \ldots, J$ and $\sum_{j=1}^{J} n_j = n$.

Now the information matrix $X^T X$ can be expressed as:

$$\begin{aligned} X^T X = M(\underline{n}) &= \sum_{j=1}^{J} n_j \underline{v}_j \underline{v}_j^T \\ &= V N V^T \end{aligned} \tag{1.7}$$

where $\underline{n} = (n_1, n_2, \ldots, n_J)^T$, $V = [\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_J])$, and $N = diag(n_1, n_2, \ldots, n_J)$.

Intuitively, now we would like to choose $\underline{n}$ to make the information matrix $X^T X = M(\underline{n})$ as big as possible. Since n_j's must be integers which we are solving for in a design, this type of integer programming problem is described as an exact design problem. In general integer-programming problems are more difficult to solve even without additional constraints in consideration. Calculus theories are more readily applicable to continuous design problem and a solution $\underline{n}$ has to be worked out for different values of n in a design if one chooses so to seek optimal $\underline{n}^*$ directly.

To avoid solving the integer-programming problem, however, we can take a closer look at the matrix $M(\underline{n})$. Suppose we write:

$$\begin{aligned} M(\underline{n}) &= \sum_{j=1}^{J} n_j \underline{v}_j \underline{v}_j^T \\ &= n \sum_{j=1}^{J} \frac{n_j}{n} \underline{v}_j \underline{v}_j^T \\ &= n \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^T \\ &= nVPV^T = nM(p) \end{aligned}$$

where $M(p) = \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^T = VPV^T$, $p_j = \frac{n_j}{n}$, the proportions of n observations taken at $\underline{v}_j$, $P = diag(p_1, p_2, \ldots, p_J)$ such that $p_j \geq 0$ and $\sum_{j=1}^{J} p_j = 1$.

Given n for a design, then to make X^TX large is equivalent to making $M(p)$ large subject to $p_j = \frac{n_j}{n}$, $j = 1, 2, \ldots, J$. The original exact design problem becomes now an approximate design problem with constraints $p_j \geq 0$ and $\sum_{j=1}^{J} p_j = 1$. The approximate design problem is easier and more flexible to work with to find the optimal weights p_j^* and an approximate solution $\underline{n}^*$ of the original exact design problem can be easily obtained by multiplication $n \times p^*$, where $p^* = (p_1^*, p_2^*, \ldots, p_J^*)$ is the optimal design weights. $\underline{n}^*$ is rounded to the nearest "integer" exact design.

It is interesting to note that $p = (p_1, p_2, \ldots, p_J)$ can be thought of as a resultant probability distribution on $\mathcal{V}$ and $M(p)$ can also be expressed as:

$$M(p) = E_p[\underline{v}\,\underline{v}^T]$$

where $p(\underline{v} = \underline{v}_j) = p_j$.

Finding an optimal design is simplified by finding p_j assigned to design points

$\underline{v}_j \in \mathcal{V}$. Such a design may include $p_j = 0$, meaning $\underline{v}_j$ is not selected for a design.

1.2 Design Measure

We have mentioned above that $p = (p_1, p_2, \ldots, p_J)$ can be thought as the resultant probability distribution on $\mathcal{V}$, but since $\mathcal{V}$ is the image of $\mathcal{X}$ under f, an induced design space, we can also think of $p = (p_1, p_2, \ldots, p_J)$ as a probability distribution or a "measure" on the original design space $\mathcal{X}$. By convention the notation used for the design measure is ξ and is represented as:

$$\xi = \left\{ \begin{array}{cccc} x_1, & x_2, & \cdots, & x_J \\ p_1, & p_2, & \cdots, & p_J \end{array} \right\} \qquad \text{over } \mathcal{X}$$

where the first row of ξ gives the possible values of design variables with associated design weights p_j, $j = 1, 2, \ldots, J$, on the second row and that $\int_{\mathcal{X}} \xi \, dx = 1$, $0 \leq p_j \leq 1$, $j = 1, 2, \ldots, J$.

1.3 Support of a Design Measure

In the design space, support of a design measure is defined to be the vertices $\underline{v}_j$, $j = 1, 2, \ldots, J$, associated with positive design weights of p. The notation for support of a design measure is:

$$Supp(\xi) = \{v_j \in \mathcal{V},\, p_j > 0,\, j = 1, 2, \ldots, J\} \tag{1.8}$$

Very often we see in an optimal design ξ^*, such that the support $Supp(\xi^*)$ is strictly a subset of $\mathcal{V}$.

1.4 Standardized Variance of the Predicted Response

The predicted response value at given $\underline{x}$ is:

$$\hat{y}(\underline{x}) = f_1(\underline{x})\hat{\theta}_1 + f_2(\underline{x})\hat{\theta}_2 + \ldots + f_k(\underline{x})\hat{\theta}_k = \underline{f}^T(\underline{x})\hat{\underline{\theta}} \tag{1.9}$$

where $\underline{f}^T(\underline{x}) = (f_1(\underline{x}), f_2(\underline{x}), \ldots, f_k(\underline{x}))$, $\hat{\underline{\theta}} = (X^TX)^{-1}X^TY$.

We also know that $D(\hat{\underline{\theta}}) = \sigma^2(X^TX)^{-1}$, where σ^2 is a constant and $X^TX = nM(p)$.

Then,

$$\begin{aligned} D(\hat{y}(\underline{x})) &= D(\underline{f}^T(\underline{x})\hat{\underline{\theta}}) = \underline{f}^T(\underline{x})D(\hat{\underline{\theta}})\underline{f}(\underline{x}) \\ &= \underline{f}^T(\underline{x})\sigma^2(X^TX)^{-1}\underline{f}(\underline{x}) \\ &= \underline{f}^T(\underline{x})\sigma^2(nM(p))^{-1}\underline{f}(\underline{x}) \\ &= \underline{f}^T(\underline{x})\frac{\sigma^2}{n}M^{-1}(p)\underline{f}(\underline{x}) \qquad (\frac{\sigma^2}{n} \text{ is a constant}) \\ &= \frac{\sigma^2}{n}\underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x}) \\ &= \frac{\sigma^2}{n}d(\underline{x}, p) \end{aligned} \tag{1.10}$$

where $M(p)$ is the information matrix, $d(\underline{x}, p)$ is the standardized variance of the predicted response at given $\underline{x}$ and n is the total number of trials in a design. Thus, in an n trial experimental design with known σ^2, $d(\underline{x}, p)$ can be expressed as

$$d(\underline{x}, p) = \frac{n \times var(\hat{y}(\underline{x}))}{\sigma^2} \tag{1.11}$$

and easily be calculated.

1.5 Properties of the Matrix M(p)

The information matrix $M(p)$ has two major properties:

(1) $M(p)$ is a $J \times J$ symmetric non-negative definite matrix. $M(p)$ is symmetric from it's definition, $M(p) = E_p[\underline{v}\underline{v}^T] = \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^T = VPV^T$. Also,

$$\begin{aligned} \underline{x}^T M(p)\underline{x} &= \underline{x}^T E_p[\underline{v}\underline{v}^T]\underline{x} \\ &= E_p[\underline{x}^T \underline{v}\underline{v}^T \underline{x}] \\ &= E_p[(\underline{x}^T \underline{v})^2] \geq 0 \qquad (Mandal, 2000) \end{aligned}$$

The determinant of the information matrix is zero if the number of support points is less than k, the number of parameters of a design.

(2) Suppose we extend the definition of M to the set G of all probability distributions in $\mathcal{V}$. Let $\mathcal{M} = \{M(p) : p \in G\}$. Each element of $\mathcal{M}$ is a $J \times J$ matrix and $\mathcal{M}$ is the convex hull of the set $\underline{v}\underline{v}^T : v \in \mathcal{V}$. Note that if p is the design measure having probability 1 at $\underline{v}$, then $M(p_v) = \underline{v}\underline{v}^T$(Silvey, 1980).

1.6 Optimal Design Criteria and Their Properties

Recall that our goal is to make the matrix $M(p)$ large in some sense to obtain a best inference on all or some of the unknown parameters $\underline{\theta} \in \Theta$. Different experiments may have different rules on how $M(p)$ should be made large. We may consider making $M(p)$ large as maximizing a real valued function $\phi(p) = \psi(M(p))$. The function ϕ is often called a criterion function. The criterion defined is called ϕ-optimality and such a design which maximizes $\phi(p)$ is called ϕ-optimal design.

In this section we discuss possible design criteria of interest and their properties. For all the optimality we choose a design which maximizes the defined criterion function. Consider the following two cases:

Case I

First we consider the case when the inference on all the parameters $\underline{\theta}$ of a linear model is of interest. As shown in chapter 1, $M(p)$ must be a non-singular and positive definite matrix. In this case, possible design criteria to consider are D-, G-, A- and E- optimal designs.

(1) D-optimality

The most studied optimal design criterion and employed in application is the D-optimality. The criterion function is defined to be

$$\phi_D(p) = \psi_D(M(p)) = \log\ det[M(p)] = -\log\ det(M^{-1}(p)). \tag{1.12}$$

A probability measure p^* is said to be D-optimal when

$$det(M(p^*)) = \sup_p\ det(M(P)) \tag{1.13}$$

or equivalently when p^* minimizes $det(M^{-1}(p))$.

J. Kiefer and J. Wolfowitz (1960) showed that it is equivalent to choosing p^* such that

$$\sup_x\ d(x, p^*) = k \tag{1.14}$$

where $d(x,p) = \underline{f}^T(x)M^{-1}(p)\underline{f}(x)$ is the standardized variance of the predicted response and k is the number of parameters in the linear model.

One motivation for choosing D-optimality is the interpretation of the joint confidence region for the estimate of the vector of unknown parameters $\underline{\theta} \in \Theta$ in linear model like in multivariate analysis. If we assume normality of the errors in the linear model and we know that the variance of the estimator $\hat{\underline{\theta}}$ is proportional to $M^{-1}(p)$, then the joint confidence region of $\underline{\theta}$ can be expressed by an ellipsoid of the form

$$\{\underline{\theta}, (\underline{\theta} - \hat{\underline{\theta}})^T M(p)(\underline{\theta} - \hat{\underline{\theta}}) \leq c\}, \tag{1.15}$$

for some critical value c where $\hat{\underline{\theta}}$ is the least square estimator of $\underline{\theta}$. The volume of the ellipsoid is proportional to $[det(M(p))]^{\frac{-1}{2}}$ and it's desired to have small ellipsoid for better inference. D-optimal criterion chooses $M(p)$ to minimize the ellipsoid and therefore $[det(M(p))]^{\frac{-1}{2}}$ or equivalently to maximize $det(M(p))$ or $\log[det[M(p)]]$. The value of $\log[det(M(p))]$ is finite if and only if $M(p)$ is non-singular, meaning that all the unknown parameters $\underline{\theta}$ are estimable. D-optimality is the most well studied problem which can be seen in published books or articles by Kiefer (1959), Fedorov (1972), Silvey (1980), Atkinson and Donev (1992), Torsney (1981), Mandal (2000) etc., having some properties as the following.

At this point it is interesting to mention I-optimality. An I-optimal design differs from a D-optimal design only that the expected integrated mean squared error of prediction of a model generated by least-squares fitting the responses to a linear statistical model is minimized in an I-optimal design.

Properties of $\phi_D(p) = \psi_D(M(p))$

(i)ψ_D is an increasing function in $M(p) \in \mathcal{M}$ where $\mathcal{M}$ is the set of all positive definite symmetric matrix, ie. for $M_1, M_2 \in \mathcal{M}$,

$$\psi_D(M_1 + M_2) \geq \psi_D(M_1). \tag{1.16}$$

(ii)ψ_D is a concave function in $M(p)$.

(iii)When ϕ_D is finite, it is differentiable with first derivative being

$$\frac{\partial \phi_D}{\partial p_j} = \underline{v}_j^T M^{-1}(p) \underline{v}_j. \tag{1.17}$$

(iv)ϕ_D is invariant under a non-singular linear transformation of $\mathcal{V}$, ie., suppose $V = (v_1, v_2, \ldots, v_J)$ is transformed to $W = (w_1, w_2, \ldots, w_J)$ under the linear transformation $\underline{w}_j = A\underline{v}_j \forall j$, ie. assigning weights p_j to vertex w_j, and A is a $k \times k$ matrix, then

$$\phi_D(M_w(p)) = \phi_D(M_v(p)) + constant. \tag{1.18}$$

To maximize $\phi_D(M_w(p))$ is equivalent to maximizing $\phi_D(M_v(p))$.

Theorem

The weighted sum of the standardized variances of the predicted response over all points of the design equals the number of parameters k in the linear model, see Mandal (2000) for a simple proof.

$$\sum_{j=1}^{J} p_j d(\underline{x}_j, p) = k \tag{1.19}$$

In a continuous design,

$$\int_{\mathcal{X}} d(\underline{x}, p)\, dp(\underline{x}) = k. \tag{1.20}$$

Theorem

The function $\log det[M(p)]$ is a concave function(Mandal, 2000).

(2)G-optimality

The purpose of this criterion is to choose a design which minimizes the maximum value of $\underline{v}^T M^{-1}(p)\underline{v}$ which is proportional to the variance of $\underline{v}^T\underline{\hat{\theta}} = \hat{Y}(\underline{x})$, the predicted response. Kiefer and Wolfowitz (1960) published an article which includes the proof of equivalence of G-optimality and the D-optimality criteria.

Properties

(i)ψ is an increasing function over the set of all positive definite symmetric matrices $M(p) \in \mathcal{M}$.

(ii)ψ is concave on $\mathcal{M}$.

(iii)ϕ_G is invariant under a non-singular linear transformation of $\mathcal{V}$. We can see this simply by considering the same linear transformation in part (iv) of D-optimality. Having $M_w(p) = WPW^T$ and $W = AV$,

$$\begin{aligned}
\phi_G(p) &= \psi_G(M(p)) = -\max_{\underline{w}\in W} \underline{w}^T M_w^{-1}(p)\underline{w} \\
&= -\max_{\underline{w}\in W} \underline{w}^T M_w^{-1}(p)\underline{w} \\
&= -\max_{\underline{w}\in W} \underline{w}^T (WPW^T)^{-1}\underline{w} \\
&= -\max_{A\underline{v}\in W} (A\underline{v})^T (AVP(AV)^T)^{-1} A\underline{v} \\
&= -\max_{\underline{v}\in A^{-1}W} \underline{v}^T A^T (AVPV^T A^T)^{-1} A\underline{v} \\
&= -\max_{\underline{v}\in A^{-1}W} \underline{v}^T A^T (A^T)^{-1}(VPV^T)^{-1} A^{-1} A\underline{v} \\
&= -\max_{\underline{v}\in A^{-1}W} \underline{v}^T M_v^{-1}(p)\underline{v} \\
&= \phi_G(M(p)) = \phi_G(p)
\end{aligned}$$

where $P = diag(p_1, p_2, \ldots, p_J)$ and $W = (w_1, w_2, \ldots, w_J)$

(3)A-optimality

The criterion function of A-optimality is defined by

$$\phi_A(p) = \psi_A(M(p)) = -Trace(M(p)). \tag{1.21}$$

Recall the $k \times k$ dispersion matrix $D(\hat{\theta})$ which is proportional to $M^{-1}(p)$. The criterion function of an A-optimal design desires to minimize the sum of the variances of the parameter estimates $\underline{\hat{\theta}}$, but without information on the correlations among them. Since only the k diagonal elements of $M^{-1}(p)$ need to be calculated, computationally it is simpler to evaluate the $\phi_A(p)$ criterion function than the previous two. The criterion was considered by Elfving (1952) and Chernoff (1953).

Properties

(i)ψ_A is an increasing function over the set of positive definite symmetric matrices $M(p) \in \mathcal{M}$.

(ii)ψ_A is concave on $\mathcal{M}$.

(iii)When $\phi_A(p)$ is finite, it's differentiable with the first derivative being

$$\frac{\partial \phi_A}{\partial p_j} = \underline{v}_j^T M^{-2}(p) \underline{v}_j \tag{1.22}$$

(4)E-optimality

The criterion function of E-optimality is defined by

$$\phi_E(p) = \psi_E(M(p)) = -\lambda_{max}(M^{-1}(p)), \tag{1.23}$$

where $\lambda(M^{-1}(p))$ denotes the largest eigenvalue of the matrix $M^{-1}(p)$, see Kiefer (1974). Since the confidence ellipsoid of the parameter estimates $\underline{\hat{\theta}}$ is proportional to the matrix $M^{-1}(p)$, the length of the principle axes of the ellipsoid is associated with the eigenvalues of $M^{-1}(p)$. Thus the criterion function of E-optimality seeks to minimize the variance of the least well-estimated contrast $\underline{a}^T\underline{\theta}$ such that $\underline{a}^T\underline{a} = 1$.

Properties of $\phi_E(p) = \psi_E(M(p))$

(i)$\psi_E(M(p))$ is an increasing function over the set of positive definite symmetric matrices $M(p)$.

(ii)$\psi_E(M(p))$ is concave on $\mathcal{M}$.

(iii)Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the eigenvalues of $M(p)$ in descending order. ϕ_E has unique first partial derivatives with respect to positive weights p_j when λ_1 is unique, otherwise ϕ_E is not differentiable.

Case II

Now we consider the case when not all k parameter but s parameters or s linear combinations of the k parameters $\theta_1, \theta_2, \ldots, \theta_k$ in the linear model are of interest. Let the s linear combinations be $\underline{\gamma} = A\underline{\theta}$, where A is a $s \times k$ matrix of rank $s \leq k$. For a special case when $A = [I_s \vdots O]$, where I_s is the $s \times s$ identify matrix and O is the $s \times (k-s)$ zero matrix. In this case we are interested in estimating the first s

parameters $\theta_1, \theta_2, \ldots, \theta_s$ of $\theta \in \Theta$.

When $M(p)$ is non-singular, the covariance matrix of the least square estimate of $A\underline{\theta}$ is proportional to the matrix $AM^{-1}(p)A^T$. If $M(p)$ is singular, Graybill (1969) showed that the requirement for estimating the vector $\underline{\gamma} = A\underline{\theta}$ is that the row space of A is in the column space of $M(p)$ which lead to the invariance of the matrix $AM^{-}(p)A^T$ to the choice of the generalized inverse matrix $M^{-}(p)$ of $M(p)$.

A generalized inverse of a matrix M is defined as any matrix M^{-} which satisfy the condition $MM^{-}M = M$. The generalized inverse matrix M^{-} exists but is not unique for a matrix M. It is only unique when M is a square non-singular matrix and $M^{-} = M^{-1}$. Then, in general, to construct a good design would mean to choose p such that $AM^{-}(p)A^T$ is minimized. Specifically we will take a look at D_A-, E_A- and linear optimality.

(1)D_A-optimality

To indicate that the design criterion is dependent of the matrix of coefficients A, Sibson (1974) named the criterion D_A-optimality which the criterion function is defined by

$$\phi_{D_A}(p) = \psi_{D_A}(M(p)) = -\log det[AM^{-}(p)A^T)]. \tag{1.24}$$

Properties

(i)ψ_{D_A} is an increasing function over the set of positive definite symmetric matrices $M(p) \in \mathcal{M}$.

(ii)ψ_{D_A} is a concave function on $\mathcal{M}$

(iii)ϕ_{D_A} has the following first partial derivatives with respect to positive weights $p_j \geq 0$.

$$\frac{\partial \phi_{D_A}}{\partial p_j} = \underline{v}_j^T M^-(p) A^T [AM^-(p)A^T]^{-1} AM^-(p)\underline{v}_j \tag{1.25}$$

Since $AM^-(p)A^T$ is invariant with respect to the choice of the generalized inverse M^- of $M(p)$, if $\underline{v}_j$'s and A are in the column space of $M(p)$, the partial derivatives are also invariant for any choice of $M^-(p)$ of $M(p)$, see Graybill (1969) Theorem 6.6.9 and Corollary 6.6.9.2.

In the special case when the first s parameters $\theta_1, \theta_2, \ldots, \theta_s$ are of interest, A can be written as a $[I_s \vdots O]$ matrix and $M(p)$ be expressed as a partitioned matrix as

$$M(p) = \begin{bmatrix} M^{(11)}_{s\times s} & M^{(12)}_{s\times(k-s)} \\ M^{(21)} & M^{(22)}_{(k-s)\times(k-s)} \end{bmatrix}, \tag{1.26}$$

where $M^{(21)}$ is $M^{(12)}_{s\times(k-s)}$ transposed. Rhode (1965) and Torsney (1981) showed that the matrix $[AM^-(p)A^T]^{-1}$ can be expressed as $(M^{(11)} - M^{(12)}M^{(22)} - M^{(21)})$. For this special case, choosing a design p to maximize ϕ_{D_A} is equivalent to maximizing $\phi_{D_A}(p) = \psi_{D_A}(M(p)) = \log det(M^{(11)} - M^{(12)}M^{(22)} - M^{(21)})$, which is called the D_s-optimality. For reference, see Karlin and Studden (1966), Atwood (1969), Silvey and Titterington (1973) and Silvey (1980).

(2)E_A-optimality

The criterion function of E_A-optimality is defined by

$$\phi_{E_A}(p) = \psi_{E_A}(M(p)) = -\lambda_{max}(AM^-(p)A^T) \tag{1.27}$$

where $\lambda_{max}(AM^-(p)A^T)$ denotes the largest eigenvalue of the matrix $AM^-(p)A^T$, see

Pazman (1986). The motivation for E_A-optimality is the same as for E-optimality and the properties of the criterion function ϕ_{E_A} are similar also.

(3)Linear optimality

The criterion function of linear optimality is defined by

$$\phi_L(p) = \psi_L(M(p)) = -tr(M^-(p)L), \tag{1.28}$$

where L is a $k \times k$ matrix of coefficients. $\phi_L(p)$ is linear in the elements of the covariance matrix $M^-(p)$. Note that if L is of rank $s \leq k$, then $\phi_L(p)$ can be written as

$$\phi_L(p) = \psi_L(M(p)) = -tr(M^-(p)L) = -tr(M^-(p)A^TA) = -tr(AM^-(p)A^T), \tag{1.29}$$

where A is a $s \times k$ matrix of rank s. In this case, it's easy to see the contrast between L- and D_A- optimal design where D_A- optimality seeks to minimize the determinant of $AM^-(p)A^T$ and L-optimality seeks to minimize the trace of $AM^-(p)A^T$. Also the problem becomes an A-optimal design where L is an identity matrix I_k.

Properties of $\phi_L(p) = \psi_L(M(p))$

(i)ψ_L is an increasing function over the set of positive definite symmetric matrices $M(p)$.

(ii)ψ_L is concave on $\mathcal{M}$.

(iii)ϕ_L has the following first partial derivatives with respect to positive weights p_j,

$$\frac{\partial \phi_L}{\partial p_j} = \underline{v}_j^T M^-(p)A^TAM^-(p)\underline{v}_j. \tag{1.30}$$

In a special case when $A = \underline{c}^T$ has a rank of 1, where $\underline{c}$ is a $k \times 1$ vector of constants,

Elfving (1952) referred to this as c-optimality with criterion function being

$$\phi_c(p) = -\underline{c}^T M^-(p)\underline{c}. \tag{1.31}$$

Chapter 2

Choices of a Model

Optimal experimental designs for linear models depend upon the model chosen to fit to the collected data. Before the construction of an optimal design of experiment, an appropriate form of the model, either linear or non-linear in the parameters $\underline{\theta}$, must first be known or defined to describe the data. Therefore choosing a correct model is important and the appropriateness of the choice of a model can only be determined by analysis of the experimental results. This chapter intends to give an overview on the choice of an appropriate model. The focus will be on linear models with some discussions of non-linear models in section 2.2.

2.1 Linear Models of One Factor

The underlying relationship between an observed response y and a factor x or a vector of design variables $\underline{x}$ is usually unknown. Therefore we want to choose a model which follows the pattern of the observed response over the region of a factor of interest. We begin by looking at the simplest linear one-factor first-order model of the form $y(x) = \theta_0 + \theta_1 x$.

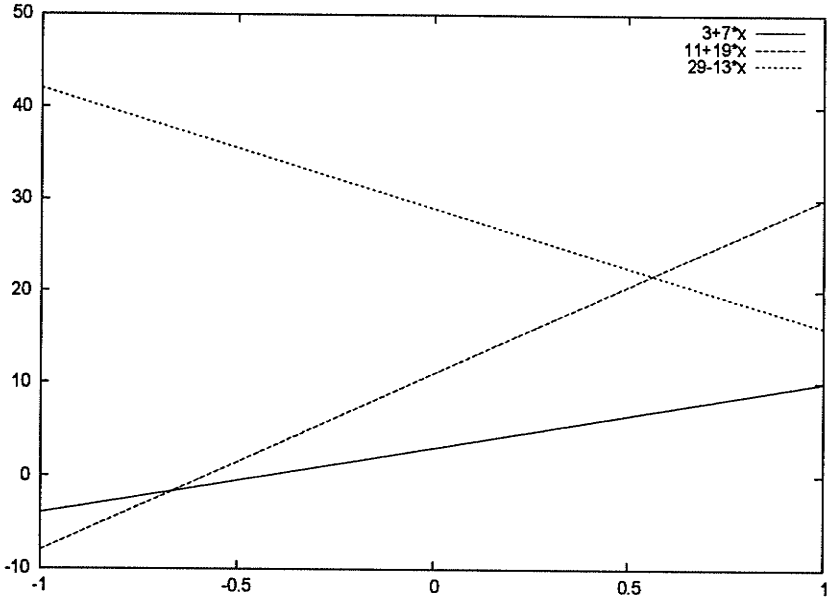

Figure 2.1:

Figure 2.1 gives an illustration of 3 simple first-order functions

$$y(x) = 3 + 7x \tag{2.1}$$

$$y(x) = 11 + 19x \tag{2.2}$$

$$y(x) = 29 - 13x. \tag{2.3}$$

The values of parameters θ_0 and θ_1 are the respective intercept and slope of each line. The rate of increase of $y(x)$ as a function of x only depends on the slope parameter θ_1, not on x. Least square estimation is often the method used to suggest good estimators of θ_0 and θ_1 for this model.

Increase in design variable x can either increase or decrease $y(x)$ without a bound unless a constraint on the region of x where $y(x)$ is fitted is imposed. However, often

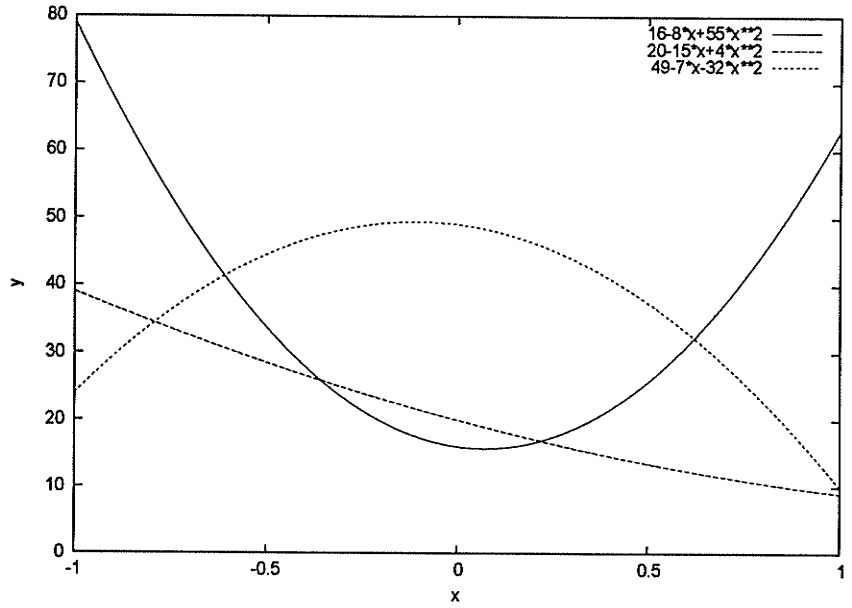

Figure 2.2:

the response does not have a straight-line pattern but a curve with a maximum or minimum or it increases or decreases to an asymptote. Figure 2.2 gives 3 illustrations of second-order curves with symmetry about their maximum or minimum.

$$y(x) = 16 - 8x + 55x^2 \tag{2.4}$$

$$y(x) = 20 - 15x + 4x^2 \tag{2.5}$$

$$y(x) = 49 - 7x - 32x^2 \tag{2.6}$$

The maximum or minimum of this type of functions can easily be found. For function (2.4), the minimum of $y(x)$ is at $x = \frac{8}{110}$. For function (2.5), the minimum of $y(x)$ is at $x = \frac{15}{8}$ which lies outside the region of the plotted region $x = [-1, 1]$. For function (2.6), since the coefficient of the x^2 term is a negative value, the function has a maximum at $x = \frac{-7}{64}$. More complicated curves may be needed to fit the pattern of

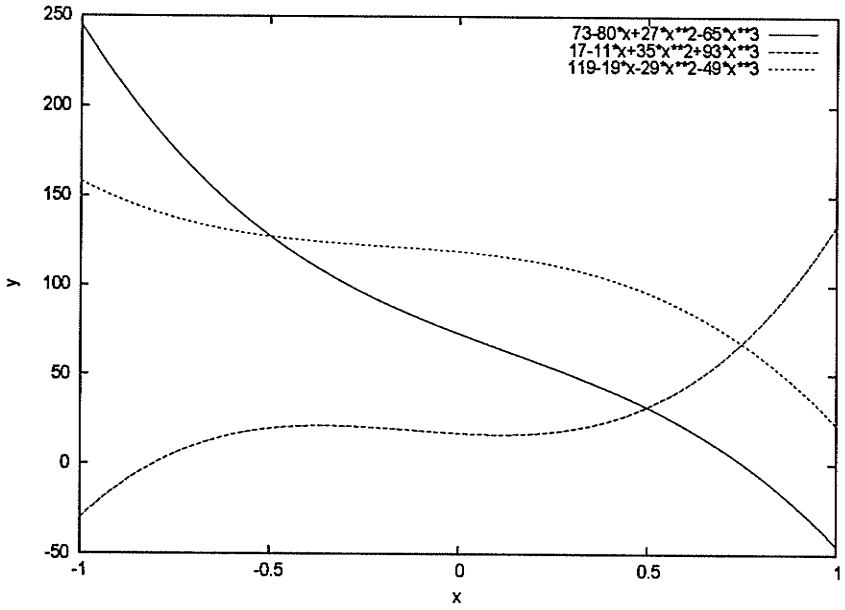

Figure 2.3:

the observed response $y(x)$ over some region of x. These curves can be described by the 3^{rd}-order or higher-order polynomials. Again three examples are illustrated in Figure 2.3 for the functions

$$y(x) = 73 - 80x + 27x^2 - 65x^3 \tag{2.7}$$

$$y(x) = 17 - 11x + 35x^2 + 93x^3 \tag{2.8}$$

$$y(x) = 119 - 19x - 29x^2 - 49x^3 \tag{2.9}$$

Higher-order polynomials are easy to use to fit a trend for data, however adding unnecessary terms will inflate the variance of predicted response from the fitted model. Increasing the number of the parameters in the model may also increase the number of trials in an experiment necessary to estimate the parameters, therefore providing incentives to choose simple models. As mentioned by Atkinson and Donev (1992),

often in experiments, the response can be described well by polynomial models of order no more than two. Other times models modified from those examples in Figure 2.2 with asymmetry around the single maximum or minimum can be used to describe data. For example, the transformation of the factor to $x^{\frac{1}{2}}$ or $\log x$ is often used in models rather than adding additional x^3 or x^4 terms in the models.

2.2 Non-Linear Models

Sometimes models non-linear in parameters may be more appropriate describing certain phenomena. A simple non-linear model example could be

$$y(x) = \theta_0 x^{\theta_1}. \tag{2.10}$$

Assuming multiplicative error term, log transformation can be applied to the model to get

$$\begin{aligned} y'(x) &= \log\theta_0 + \theta_1 \log x_1 + error & (2.11)\\ &= \theta_0' + \theta_1' x_1' + error. & (2.12) \end{aligned}$$

Bates and Watts (1988) mentioned the Hougen-Watson model for chemical reaction kinetics,

$$rate = \frac{\theta_4 x_2 - \frac{1}{\theta_5} x_3}{1 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3} \tag{2.13}$$

where $\theta_1, \ldots, \theta_5$ are unknown parameters with input factors x_1, x_2 and x_3 being hydrogen, n-pentane and isopentane respectively. No simple transformation can be used to make this model linear. There may be cases when no linear transformation for a

non-linear model is possible as above, or cases when the usual error assumptions are not met by the transformed model. In other cases, non-linear models may have good theoretical properties where estimation of parameters is of direct interest. Also, there are cases that the response can only be described by a non-linear model. Atkinson and Donev (1992) gave an example of a simple non-linear model

$$y(x) = \theta_0(1 - exp\{-\theta_1 x\}). \tag{2.14}$$

This model is often used to describe the growth rate of cells or bacteria.

A note on the advantage of non-linear models is that if based firmly on theory, extrapolation over the values of x often does not produce seriously erroneous predictions in response where the same does not apply to linear models of polynomials. However, a disadvantage of non-linear models in optimal design is that estimation of parameters depends on the values of the parameters.

2.3 Response Surface Models

When two or more factors are included in a model, we call it a response surface model. Figure 2.4 gives an example of contours of a function of a response surface with two factors without interaction.

$$y(x) = 3 + 5x_1 + 7x_2 \tag{2.15}$$

The surface of this function is a plane in a 3-D space. The contour lines are equi-spaced and the rate of increase of $y(x)$ by increasing one unit of x_1 is independent of x_2 and is the same for x_2 of x_1.

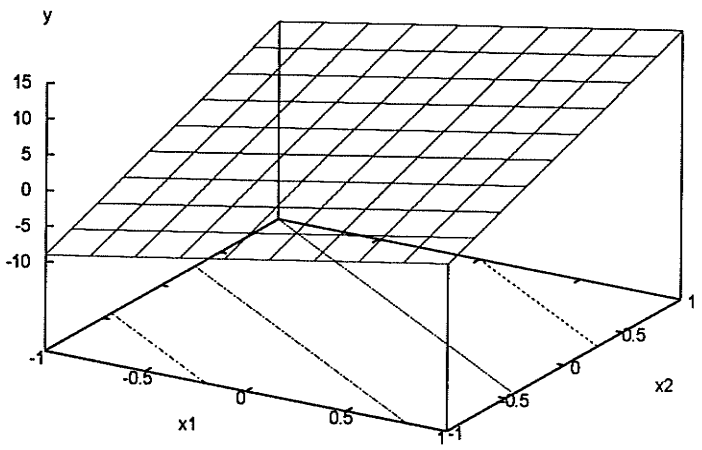

Figure 2.4:

If an interaction term is present in the model as in equation (2.16), the contours now becomes hyperbolae instead of parallel lines as illustrated in Figure 2.5. The rate of increase of $y(x)$ with the increment of x_1 by one unit is now dependent of x_2 and vice versa. The same applies to models with more than two factors.

$$y(x) = 3 + 5x_1 + 7x_2 + 9.8x_1x_2 \tag{2.16}$$

Furthermore, to describe curvature, if present in the response surface, square terms of factors need to be included in the model. Figure 2.6 gives an example of contours of

$$y(x) = (x_1 - 4)^2 + (2x_2 - 3)^2 + 1.5x_1x_2 \tag{2.17}$$

$$= x_1^2 + 4x_2^2 - 8x_1 - 12x_2 + 1.5x_1x_2 + 25 \tag{2.18}$$

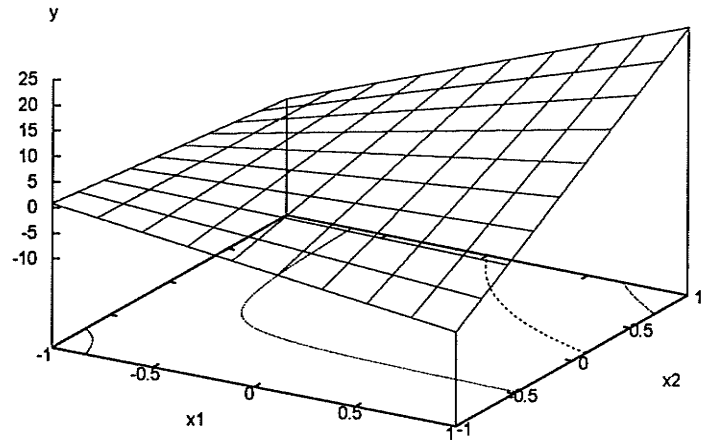

Figure 2.5:

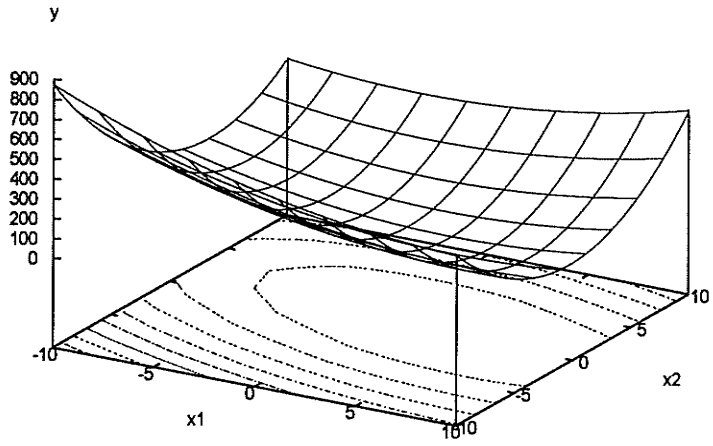

Figure 2.6:

with a single minimum value at $x_1 = \frac{184}{55}$ and $x_2 = \frac{48}{55}$. Often not all of the 2^{nd}-order terms are needed in order to model the curvature in the response when there are many factors included in the model.

2.4 Examples

Examples 5.1 and 5.2 in Atkinson and Donev (1992) are good examples of models of a single quantitative factor and choice of design. Similarly suppose we have a simple linear regression model,

$$y(x) = \theta_0 + \theta_1 x. \tag{2.19}$$

For a design having $n = 3$ trials, we can write

$$E\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \tag{2.20}$$

$$= X\theta \tag{2.21}$$

where $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix}$, and $\underline{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$.

Clearly the matrix X depends on $\underline{x}$ and the model of choice. Suppose the design variable $x \in \mathcal{X} = [-1, 1]$. As mentioned in Chapter 1, the design problem can be formulated as to choose n points in the design space $\mathcal{X}$ to give best estimate of θ. Three possible designs are to choose $\underline{x} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$.

For model 2.19, it is easy to see that the information matrix

$$X^TX = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \tag{2.22}$$

and

$$\begin{aligned} det(X^TX) &= det \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \\ &= n\sum_{i=1}^{n}(x_i - \bar{x})^2. \end{aligned} \tag{2.23}$$

The covariance matrix for the least square estimate $\underline{\hat{\theta}} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{bmatrix}$ is

$$\sigma^2(X^TX)^{-1} = \frac{\sigma^2}{det(X^TX)} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} \tag{2.24}$$

For the design $\underline{x} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $X = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $X^TX = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$. $M(p) = \frac{1}{n}X^TX = \begin{bmatrix} 1 & 0 \\ 0 & \frac{2}{3} \end{bmatrix}$, $M^{-1}(p) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{3}{2} \end{bmatrix}$, $\log det[M(p)] = \log(\frac{2}{3}) = -0.4055$ and $-trace[M(p)] = \frac{-5}{3}$.

For the design $\underline{x} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$, $X = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $X^TX = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$. Similarly we can find $M(p)$ to be $\begin{bmatrix} 1 & \frac{-1}{3} \\ \frac{-1}{3} & 1 \end{bmatrix}$, $M^{-1}(p) = \begin{bmatrix} \frac{9}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{9}{8} \end{bmatrix}$, $\log det[M(p)] = -0.1178$

and $-trace[M(p)] = -2$.

A third possible choice of design is $\underline{x} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$, with $M^{-}1(p) = \begin{bmatrix} \frac{9}{8} & \frac{-3}{8} \\ \frac{-3}{8} & \frac{9}{8} \end{bmatrix}$, $\log det[M(p)]$ and $-trace[M(p)]$ are also found to be -0.1178 and -2 respectively.

We can see from this example that for the 1^{st}-order model $y(x) = \theta_0 + \theta_1 x$ given only three trials available for the experiment, with the given three possible designs, according to D-optimality where the criterion function $\log det[M(p)]$ is to be maximized, it is preferable to choose the design at $\underline{x} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ over the design $x = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

If the criterion is to minimize the maximum of the standardized variance $d(x,p)$, for $\underline{x} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, the maximum of $d(x,p)$ is

$$\begin{aligned} \max_x d(x,p) &= \max_x \underline{f}^T(x) M^{-1}(p) \underline{f}(x) \\ &= \max_x \begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} \\ &= \max_x (1 + \frac{3x^2}{2}) \\ &= 2.5 \end{aligned}$$

at x = -1 or 1. For $x = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$, $\max_x d(x,p)$ is 3 at $x = 1$ and for $x = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$, $\max_x d(x,p)$ is 3 at x = -1. $x = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ would be preferable.

In this example we see that the same design satisfying one optimal criterion may not satisfy another.

Chapter 3

Optimality Conditions and Algorithms

Before determining optimality conditions, it is important to discuss the optimization problems we want to consider. In this thesis we mainly discuss two classes of optimization problems.

Problem (P1)

To maximize a criterion $\phi(p)$ over the measure $\mathcal{P} \equiv \{p = (p_1, p_2, \ldots, p_J), \sum_{j=1}^{J} p_j = 1, p_j \geq 0\}$. The condition $\sum_{j=1}^{J} p_j = 1$ makes this a non-degenerate constrained optimization problem over a closed bounded support region.

Problem (P2)

To maximize $\psi(x)$ over the convex hull(with vertices $G(\underline{v}_1), G(\underline{v}_2), \ldots, G(\underline{v}_J)$),

$$\mathcal{CH}\{\mathcal{G}(\mathcal{V})\} = \{x = x(p) = \sum_{j=1}^{J} p_j G(\underline{v}_j), p = (p_1, p_2, \ldots, p_J \in \mathcal{P}\} \tag{3.1}$$

where G is a one-to-one transformation and $\mathcal{V} = (\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_J)$ is a known set of vector

vertices of fixed dimension. Statistically we may also render that $x(p) = E_p[G(\underline{v})]$ assuming $G(\underline{v})$ is a random variable with probability p_j at vertex $G(\underline{v}_j)$.

Clearly (P2) is a more generalized and practical optimal linear regression problem. Carathéodory's Theorem guarantees that over the continuous unbounded space $\mathcal{V}$ there is at least one discrete optimizing distribution p^*. Problem (P2) may differ from problem (P1) in several ways, such as

- The interest may solely be the optimizing support x^*.
- A unique optimizing x^* does not imply a unique optimizing p^*.
- An optimizing p^* may have zero weights, ie., the optimum lies on the boundary of $\mathcal{P}$ and $p_j = 0$ for some j.

The approach we will adopt is to seek out p^* rather than directly seeking out x^* maximizing $\psi(x)$ over $\mathcal{CH}\{\mathcal{G}(\mathcal{V})\}$ and finding a distribution p^* such that $x(p^*) = x^*$.

3.1 Optimality Theory

We can analytically solve for solutions to construction of optimal designs in simple cases. In practical problems, it's often not possible to find an explicit solution p^*, or to derive such a design explicitly as the size of the design space increases. Iterative techniques are required and special algorithms are devised to solve constrained optimal design problems with aid of computer software. From Section 1.1 we know that by discretizing a continuous design space, there always exists an optimal measure with finite support points. We wish to find the optimal measure p^* which typically will be on the boundary of convex hull $\mathcal{CH}\{\mathcal{G}(\mathcal{V})\}$, especially when $\mathcal{V}$ is a discretization

of a continuous space. This implies that at the optimum there will be zero weights corresponding to some support points. We define optimality conditions in forms of directional derivatives of a criterion function $\phi(.)$.

3.1.1 Directional Derivatives

Let

$$g(p,q,\varepsilon) = \phi\{(1-\varepsilon)p + \varepsilon q\} \tag{3.2}$$

where $\phi(.)$ is a criterion function. The directional derivative $F_\phi\{p,q\}$ of the criterion function $\phi(.)$ at p in the direction of q is defined as (Whittle, 1973)

$$F_\phi\{p,q\} = \lim_{\varepsilon \to 0^+} \frac{g(p,q,\varepsilon) - \phi(p)}{\varepsilon} = \frac{dg(p,q,\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=0^+}. \tag{3.3}$$

The derivative $F_\phi\{p,q\}$ exists even if $\phi(.)$ is not differentiable. If $\phi(.)$ is differentiable, then (3.3) becomes

$$F_\phi\{p,q\} = (q-p)^T \frac{\partial \phi}{\partial p} = (q-p)^T d \tag{3.4}$$

$$= \sum_{i=1}^{J} (q_i - p_i) d_i, \tag{3.5}$$

where $d_i = \frac{\partial \phi}{\partial p_i}$, $i = 1, 2, \ldots, J$ and $d = \frac{\partial \phi}{\partial p}$.

In problem (P1) when $p \in \mathcal{P}$, F_j can be expressed as

$$F_j = F_\phi\{p, e_j\} = d_j - \sum_{i=1}^{J} p_i d_i, \tag{3.6}$$

where $d_j = \frac{\partial \phi}{\partial p_j}$ is the partial derivative of ϕ with respect to the jth component of p and e_j is the jth unit vector. We call F_j the vertex directional derivatives of $\phi(.)$ at p.

The importance of $F_\phi\{p,q\}$ is its utilization in our algorithms to check for convergence and therefore finding optimal measure p^*. To determine the optimality conditions, we use an important theorem(Kiefer, 1959) as stated below.

Vertex Direction Optimality Theorem *Let $\mathcal{S} = \mathcal{CH}\{\mathcal{G}(\mathcal{V})\}$ and assume that $\psi(x)$ is concave on $\mathcal{S}$ and $\psi(.)$ is differentiable at the point $x(p^*)$, then $x(p^*)$ maximizes $\psi(.)$ on $\mathcal{S}$ if and only if*

$$F_\psi\{x(p^*), G(v_j)\} = 0 \qquad \text{when } p_j^* > 0 \tag{3.7}$$

$$F_\psi\{x(p^*), G(v_j)\} \leq 0 \qquad \text{when } p_j^* = 0. \tag{3.8}$$

From above theorem and equation (3.6), a corollary of the Vertex Direction Optimality Theorem can be made.

Corollary

If $\mathcal{S} = \mathcal{P}$, $\phi(p)$ is (weakly) concave on $\mathcal{P}$ and p^* is a differentiable point of ϕ on $\mathcal{P}$, then p^* maximizes ϕ on $\mathcal{P}$ if and only if

$$\frac{\partial \phi}{\partial p_j^*} = \sum_{i=1}^{J} p_i^* \frac{\partial \phi}{\partial p_i^*} \qquad \text{when } p_j^* > 0 \tag{3.9}$$

$$\frac{\partial \phi}{\partial p_j^*} \leq \sum_{i=1}^{J} p_i^* \frac{\partial \phi}{\partial p_i^*} \qquad \text{when } p_j^* = 0, \tag{3.10}$$

or equivalently,

$$F_j = 0 \qquad \text{where } p_j^* > 0 \tag{3.11}$$

$$F_j \leq 0 \qquad \text{where } p_j^* = 0. \tag{3.12}$$

3.2 Algorithm

Having the usual constraint on the discrete probability measure $p_1, p_2, \ldots, p_J$ being non-negative and summing to 1, a suitable multiplicative algorithm (Torsney, 1977) which induce nice properties is

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(x_j^{(r)})}{\sum_{i=1}^{J} p_i^{(r)} f(x_i^{(r)})}, \tag{3.13}$$

where $f(x)$ is strictly a positive increasing function in x and may depend on one or more free parameters, and $x_j^{(r)} = d_j^{(r)}$ or $F_j^{(r)}$ defined as

$$d_j^{(r)} = \frac{\partial \phi}{\partial p_j}\Big|_{p=p(r)} \tag{3.14}$$

$$F_j^{(r)} = d_j^{(r)} - \sum_{i}^{J} p_i^{(r)} d_i^{(r)} \text{ directional derivatives at } p = p^{(r)}. \tag{3.15}$$

This type of iteration was first proposed by Torsney (1977), taking $x = d$, $f(d) = d^{\delta}$ with $\delta > 0$. This requires the derivatives d_j to be positive. Subsequently Silvey, Titterington and Torsney (1978) studied choices of δ for the same $f(d)$; Torsney (1988) considered $f(d) = e^{\delta d}$; Torsney and Alahmadi (1992), Mandal and Torsney (2000) and Torsney and Mandal (2001) considered and explored other choices of $f(.)$. See Mandal, Torsney and Carriere (2004) for further developments of the algorithm for optimal designs with two constraints.

Having such conditions imposed on $f(.)$, above authors found some properties as the following:

- $p^{(r)}$ is always feasible, with respect to their constraints.

- $F_\phi\{p^{(r)}, p^{(r+1)}\} \geq 0$. The equality is simple to see when the d_j's corresponding to non-zero weights p_j's have a common value, d, ie. $x_j = d_j = d$ or when $x_j = F_j = 0$.

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(x_j)}{\sum_{i=1}^{J} p_i^{(r)} f(x_i)} = \frac{p_j^{(r)} f(d)}{f(d) \sum_{i=1}^{J} p_i^{(r)}} = p_j^{(r)}$$

  By (3.2) and (3.3), it's easy to see $F_\phi\{p, p\} = 0$. An iterate $p^{(r)}$ is a fixed point of the iteration if the partial derivatives($d_j^{(r)}$) corresponding to non-zero $p_j^{(r)}$ are all equal. This is a necessary but not a sufficient condition for $p^{(r)}$ to solve problem (P1). So we may say that at optimality conditions, the solution to the design is a fixed point of the iteration for any subset of $\mathcal{V}$.

Some modification of the algorithm would be needed if there are many zero optimal weights(p_j). Mandal and Torsney (2004) explore one such modification based on a clustering approach. This is related to the fact that the support points of a discretized design space can be viewed as consisting of some clusters of design points.

The choice of $f(.)$ plays an important role in the convergence of the algorithm. Convergence rates also vary according to the choice of the free parameter(s) in the algorithm.

Next in this chapter, we will consider different choices of $f(.)$ and its argument on their rate of convergence in finding optimal measures for different models linear in parameters. Then we will work out examples of constructing D-optimal designs.

Chapter 4

Construction of Optimal Designs

Now we consider some examples in constructing D-optimal designs having the usual constraint on the probability measure $\sum_{j=1}^{J} p_j = 1$, $p_j \geq 0$ $\forall j$ and the support from a region of a closed bounded convex set. If extra constraints were imposed on p, then the Lagrangian approach can be used to find the optimal design (Mandal et al., 2004). As mentioned in Chapter 1, the criterion function and the information matrix for D-optimality respectively are

1. $\psi_D(M(p)) = \log det[M(p)]$ and
2. $M(p) = \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^T$.

We first consider some choices of $f(x)$ for the five examples which were studied by Silvey et al (1978), Wu (1978) and Mandal (2000). Example 1 was originated by Wynn (1970). The examples were defined by their design spaces as the following.

Example 1: $\mathcal{V} = \{(1,-1,-1)^T, (1,-1,1)^T, (1,1,-1)^T, (1,2,2)^T\}$

Example 2: $\mathcal{V} = \{(1,-1,-1)^T, (1,-1,1)^T, (1,1,-1)^T, (1,2,3)^T\}$

Example 3: $\mathcal{V} = \{(1,-1,-2)^T, (1,-1,1)^T, (1,1,-1)^T, (1,2,2)^T\}$

Example 4: $\mathcal{V} = \{(1,1,-1,-1)^T, (1,-1,1,-1)^T, (1,-1,-1,-1)^T,$
$(1,2,2,-1)^T, (1,1,-1,1)^T, (1,-1.5,1,1)^T,$
$(1,-1,-1,2)^T\}$

Example 5: $\mathcal{V} = \{(1,1,-1,-1)^T, (1,-1,1,-1)^T, (1,-1,-1,-1)^T,$
$(1,2,2,-1)^T, (1,1,-1,1)^T, (1,-1.5,1,1)^T,$
$(1,-1,-1,2)^T, (1,1,1.5,1)^T\}$

Finding an optimizing p^* is an example of problem (P1) with criterion function and information matrix being

- $\phi(p) = log\{det(M(p))\}$
- $M(p) = \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^T$.

Solving for an optimizing p^*(for Examples 1 to 5) is an example of problem (p2) where

- $G(\underline{v}) = \underline{v}\underline{v}^T$;
- $v \in \mathcal{V} \subseteq \mathbb{R}^k$, $\mathcal{V}$ is the induced design space;
- $x = x(p) = M(p)$ is a symmetric $k \times k$ matrix and
- $\psi(x) = log\{det(x)\} = \phi(p)$.

Tables 4.1 - 4.3 summarize the performance of algorithm (3.13) in finding D-optimal designs of Examples 1 - 5 and some linear models with different choices of $f(x)$ satisfying the conditions imposed.

We first considered choices of $f(x)$ with $x = d$, in particular, $f(d) = d^{\delta}$, then $f(d) = exp(\delta d)$, $f(d) = ln(e+\delta d)$, $f(d) = [exp(\delta d)]/[1+exp(\delta d)]$ and $f(d) = 1.0001+exp(-\delta d)$. In Tables 4.1 - 4.3, the number of iterations needed to achieve $\max_{1\leq j\leq J}\{F_j\} \leq 10^{-n}$, for n = 1, 2, 3 and 4, were recorded. F_j are the vertex directional derivatives at $p_j^{(r)}$ towards $p_j^{(r+1)}$. In all cases the initial weights $p_j^{(0)}$ were assigned equal weights $1/J$, $j = 1, 2, \ldots, J$. The programming and results were obtained using Minitab version 14. In all tables of iteration counts, the relative best choices of δ are in bold font. For each example or regression model, different choices of $f(x)$ yield similar optimal design solution and only one approximated design space at intervals of 0.1 with corresponding weights are reported.

Examples 1 - 5 have fixed vertices and their optimal design measures are reported in Table 4.1. The solutions are similar to those of Mandal (2000) but the number of iterations required for convergence are different due to the fact that the criterion function we considered is $\phi_D(p)$ in (1.12), not the standardized criterion.

Note that in all tables, a symbol "x" is used to replace an iteration number when a convergence of the algorithm can not be reached with the corresponding choice of $f(d)$ or $f(F)$ and δ value(s).

Table 4.1

Example 1: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}

δ	n=1	n=2	n=3	n=4	weights(p)	v
0.5	4	16	31	47	0.12507	(1,-1,-1)
1	1	7	14	22	0.28122	(1,-1,1)
1.25	2	6	11	16	0.28122	(1,1,-1)
1.4	2	5	9	14	0.31249	(1,2,2)
1.5	2	4	9	13		
1.6	**2**	**5**	**8**	**12**		
1.65	3	5	9	13		
1.7	3	7	11	15		
2	5	17	29	41		

f(d)=exp(δd)

δ	n=1	n=2	n=3	n=4
0.49	2	4	9	13
0.5	2	5	9	13
0.51	2	5	8	13
0.52	**2**	**5**	**8**	**13**
0.53	3	5	9	12
0.54	3	6	9	12
0.55	3	6	9	13
0.56	3	7	10	13
0.6	3	9	13	19

f(d)=ln(e+δd)

δ	n=1	n=2	n=3	n=4
0.5	8	33	65	98
1	6	27	53	80
1.5	6	25	50	76
2	**6**	**25**	**50**	**75**
2.5	6	25	50	76
3	6	25	51	77

$f(d)=1.0001 + \exp(-\delta d)$

δ	n=1	n=2	n=3	n=4
0.0001	2	10	20	30
0.001	1	8	15	22
0.005	**1**	**7**	**14**	**22**
0.01	1	7	15	22
0.05	2	8	16	24
0.1	2	9	17	26
0.25	3	11	22	34
0.5	4	18	35	54
0.6	5	21	43	66

$f(d)=\Phi(\delta d)$

δ	n=1	n=2	n=3	n=4
0.1	10	45	88	133
0.2	7	29	58	88
0.25	6	28	54	82
0.3	**6**	**27**	**54**	**82**
0.4	7	30	60	91
0.5	9	37	75	115

$f(d)=[\exp(\delta d)]/[1+\exp(\delta d)]$

δ	n=1	n=2	n=3	n=4
0.25	8	34	67	101
0.35	7	30	59	89
0.4	**7**	**29**	**57**	**87**
0.5	7	29	58	88
0.6	7	31	62	94
0.75	9	36	73	112

Example 2: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}

δ	n=1	n=2	n=3	n=4	weights(p)	v
0.5	4	26	56	88	0.073455	(1,-1,-1)
1	3	12	27	42	0.291395	(1,-1,1)
1.5	2	8	17	27	0.311248	(1,1,-1)
1.6	3	8	16	25	0.323901	(1,2,3)
1.65	**3**	**8**	**15**	**24**		
1.7	4	9	14	24		
1.75	4	9	16	24		
2	11	41	71	101		

f(d)=exp(δd)

δ	n=1	n=2	n=3	n=4
0.45	2	9	20	31
0.475	2	9	18	29
0.5	3	8	17	28
0.55	4	8	16	25
0.575	4	9	16	24
0.58	**4**	**10**	**16**	**22**
0.59	5	11	17	25
0.6	5	11	19	27

f(d)=ln(e+δd)

δ	n=1	n=2	n=3	n=4
0.5	8	54	115	183
0.6	7	51	108	171
0.75	7	47	100	159
1	6	44	94	149
1.5	6	42	89	142
2	**6**	**41**	**89**	**141**
2.5	6	42	89	141
3	6	42	90	143

$f(d)=1.0001+\exp(-\delta d)$

δ	n=1	n=2	n=3	n=4
0.001	4	13	28	44
0.01	**3**	**13**	**27**	**43**
0.1	4	15	31	50
0.25	3	19	41	65
0.5	4	29	64	102
0.75	7	49	104	166

$f(d)=\Phi(\delta d)$

δ	n=1	n=2	n=3	n=4
0.1	11	74	156	246
0.25	7	46	97	154
0.3	**6**	**45**	**96**	**152**
0.31	7	45	96	153
0.32	7	45	97	154
0.35	7	46	99	158
0.4	7	49	106	169
0.5	9	61	133	213
0.75	26	164	369	601

$f(d)=[\exp(\delta d)]/[1+\exp(\delta d)]$

δ	n=1	n=2	n=3	n=4
0.25	8	56	119	188
0.3	7	52	109	173
0.38	7	48	103	163
0.4	**7**	**48**	**102**	**162**
0.5	7	48	103	163
0.75	9	60	130	207

Example 3: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}

δ	n=1	n=2	n=3	n=4	weights(p)	v
0.5	3	15	28	41	0.243194	(1,-1,-2)
1	2	7	13	19	0.305285	(1,-1,1)
1.5	2	4	8	11	0.160567	(1,1,-1)
1.6	**2**	**4**	**7**	**10**	0.290954	(1,2,2)
1.65	2	5	7	11		
1.7	2	5	9	12		
1.75	2	5	9	13		
2	5	13	23	31		

f(d)=exp(δd)

δ	n=1	n=2	n=3	n=4
0.4	1	6	10	15
0.45	2	5	9	13
0.475	2	5	8	12
0.5	2	4	8	11
0.51	2	4	8	11
0.525	2	4	7	10
0.53	**2**	**4**	**7**	**10**
0.54	2	5	7	10
0.55	2	5	8	11
0.6	3	7	11	16

f(d)=ln(e+δd)

δ	n=1	n=2	n=3	n=4
0.5	7	32	59	86
1	6	26	48	70
1.5	6	24	46	67
2	6	24	45	66
2.25	**6**	**24**	**45**	**66**
2.5	6	24	45	66
3	6	25	46	67
3.5	6	25	47	68

$f(d)=1.0001+\exp(-\delta d)$

δ	n=1	n=2	n=3	n=4
0.001	**2**	**7**	**13**	**19**
0.01	2	7	13	19
0.1	2	8	15	22
0.25	2	11	20	29
0.5	4	17	32	47
0.75	7	28	53	79

$f(d)=\Phi(\delta d)$

δ	n=1	n=2	n=3	n=4
0.1	10	43	80	117
0.2	7	28	53	77
0.25	6	27	49	72
0.3	**6**	**26**	**49**	**72**
0.32	6	26	50	72
0.33	6	27	50	73
0.34	6	27	50	74
0.5	9	36	69	101

$f(d)=[\exp(\delta d)]/[1+\exp(\delta d)]$

δ	n=1	n=2	n=3	n=4
0.25	8	33	61	89
0.4	**6**	**28**	**52**	**76**
0.5	7	28	53	77
0.6	7	30	56	83
0.75	8	35	67	99

Example 4: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}

δ	n=1	n=2	n=3	n=4	weights(p)	v
0.5	11	77	215	453	0.029954	(1,1,-1,-1)
1	6	38	107	225	0.01202	(1,-1,1,-1)
1.5	4	25	71	149	0.230944	(1,-1,-1,-1)
2	**7**	**20**	**53**	**112**	0.233506	(1,2,2,-1)
2.05	9	28	56	110	0.183396	(1,1,-1,1)
2.1	13	56	204	392	0.208186	(1,-1.5,1,1)
					0.101994	(1,-1,-1,2)

f(d)=exp(δd)

δ	n=1	n=2	n=3	n=4
0.4	4	24	67	140
0.45	6	21	59	125
0.49	7	18	54	114
0.5	**9**	**22**	**53**	**112**
0.51	11	28	56	110
0.52	13	44	94	164

f(d)=ln(e+δd)

δ	n=1	n=2	n=3	n=4
0.5	21	142	396	832
1	18	124	345	726
1.5	**18**	**121**	**339**	**714**
2	18	123	343	721

$f(d)=1.0001+\exp(-\delta d)$

δ	n=1	n=2	n=3	n=4
0.001	6	39	110	232
0.005	**6**	**39**	**109**	**229**
0.01	6	39	109	231
0.1	7	47	132	277
0.25	10	66	184	388
0.5	18	122	342	722

$f(d)=\Phi(\delta d)$

δ	n=1	n=2	n=3	n=4
0.1	25	173	483	1013
0.19	19	132	369	776
0.195	19	132	368	774
0.2	19	131	367	772
0.21	**19**	**131**	**367**	**771**
0.25	20	134	375	789
0.5	47	340	960	2036

$f(d)=[\exp(\delta d)]/[1+\exp(\delta d)]$

δ	n=1	n=2	n=3	n=4
0.2	23	157	437	917
0.29	21	140	390	820
0.3	20	139	389	818
0.31	20	139	388	816
0.32	**20**	**139**	**388**	**815**
0.33	20	139	388	816
0.4	21	143	401	844
0.5	23	161	451	950

Example 5: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}

δ	n=1	n=2	n=3	n=4	weights(p)	v
0.5	8	121	310	559	0.029698	(1,1,-1,-1)
0.75	6	80	206	372	0.011689	(1,-1,1,-1)
1	5	60	155	279	0.231184	(1,-1,-1,-1)
1.25	4	48	123	222	0.233576	(1,2,2,-1)
2	4	30	77	138	0.183612	(1,1,-1,1)
2.1	**5**	**29**	**73**	**132**	0.208382	(1,-1.5,1,1)
2.11	5	28	73	134	0.101859	(1,-1,-1,2)
2.15	5	28	76	x	0	(1,1,1.5,1)
2.2	7	30	x	x		
2.25	7	x	x	x		

f(d)=exp(δd)

δ	n=1	n=2	n=3	n=4
0.4	3	38	97	174
0.5	5	30	77	139
0.51	5	30	76	136
0.52	**5**	**29**	**74**	**134**
0.525	6	29	73	134
0.53	6	28	74	x
0.55	7	x	x	x

f(d)=ln(e+δd)

δ	n=1	n=2	n=3	n=4
0.5	16	221	569	1025
1	14	193	496	895
1.5	**13**	**190**	**488**	**880**
1.6	13	190	488	880
1.7	13	190	489	882
2	14	192	492	888
2.5	14	195	501	904

$f(d)=1.0001+\exp(-\delta d)$

δ	n=1	n=2	n=3	n=4
0.001	5	62	159	286
0.005	**5**	**61**	**157**	**283**
0.01	5	62	158	285
0.1	6	74	190	343
0.25	7	103	265	478
0.5	14	191	491	888

$f(d)=\Phi(\delta d)$

δ	n=1	n=2	n=3	n=4
0.1	19	270	693	1248
0.175	15	210	538	971
0.19	15	206	530	956
0.2	15	205	527	951
0.21	**15**	**205**	**526**	**949**
0.3	17	228	586	1059
0.4	25	319	817	1478
0.5	44	538	1378	2497

$f(d)=[\exp(\delta d)]/[1+\exp(\delta d)]$

δ	n=1	n=2	n=3	n=4
0.2	17	244	627	1130
0.25	16	225	578	1041
0.29	16	218	560	1011
0.3	15	217	558	1007
0.31	15	217	557	1005
0.32	**15**	**217**	**557**	**1004**
0.33	16	217	557	1005
0.4	16	224	576	1039
0.5	19	252	647	1169

The results for different linear models we considered below are reported in Table 4.2, with the same choices of $f(x)$ considered in Examples 1 - 5, where $x = d$.

Linear Models:

$$1.\ E[y|x] = \theta_0 + \theta_1 x \tag{4.1}$$

$$2.\ E[y|x] = \theta_0 + \theta_1 x + \theta_2 x^2 \tag{4.2}$$

$$3.\ E[y|x] = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \tag{4.3}$$

$$4.\ E[y|x] = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4 \tag{4.4}$$

Let's take a closer look at the optimal design results for the simple linear model $E[y|x] = \theta_0 + \theta_1 x$. The optimal design solution converges to

$$p^* = \left\{ \begin{array}{ccccccccc} \text{weights(p)} & 0.499992 & 0.000008 & 0 & \ldots & 0 & 0.000008 & 0.499992 \\ \text{x} & -1 & -0.9 & -0.8 & \ldots & 0.8 & 0.9 & 1 \end{array} \right\} \tag{4.5}$$

where "..." are the supports with corresponding zero weights.

Clearly we see above for a simple linear model with the support x discretized into 0.1 intervals, it has four support points corresponding to non-zero weights. But, note that the support points can be clearly grouped into two clusters of points, either neighboring points or single points corresponding to non-zero weights. The two clusters suggest that for the continuous design space it is a two-point design with one point within each cluster. Each point has a corresponding weight of the total design weight of its cluster. It was suggested by Atwood (1976) that approximations to these design weights are the convex combinations of relevant points within clusters with convex weights proportional to the design weights.

Therefore, for the simple linear model, on one end we have a support point

$$x = \frac{-1 \times 0.499992 + -0.9 \times 0.000008}{0.499992 + 0.000008} \cong -1 \qquad \text{with weight 0.5,}$$

and other the other hand we have a support point

$$x = \frac{1 \times 0.499992 + 0.9 \times 0.000008}{0.499992 + 0.000008} \cong 1 \qquad \text{with weight 0.5.}$$

$$p^* = \left\{ \begin{array}{cccc} \text{weights(p)} & 0.5 & 0.5 \\ \text{x} & -1 & 1 \end{array} \right\}. \tag{4.6}$$

Similarly, the solutions for models 4.2 - 4.4 are

$$p^* = \left\{ \begin{array}{cccc} \text{weights(p)} & 0.33 & 0.33 & 0.33 \\ \text{x} & -1 & 0 & 1 \end{array} \right\}, \tag{4.7}$$

$$p^* = \left\{ \begin{array}{ccccc} \text{weights(p)} & 0.25 & 0.25 & 0.25 & 0.25 \\ \text{x} & -1 & -0.445 & 0.445 & 1 \end{array} \right\}, \text{ and} \tag{4.8}$$

$$p^* = \left\{ \begin{array}{cccccc} \text{weights(p)} & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ \text{x} & -1 & -0.66 & 0 & 0.66 & 1 \end{array} \right\}. \tag{4.9}$$

respectively.

Table 4.2

Simple Linear Regression Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)= d^{δ}						
δ	n=1	n=2	n=3	n=4	weights(p)	x
1	10	28	51	75	0.499992	-1
2	5	14	26	37	0.000008	-0.9
3	3	10	17	25	0	-0.8
4	3	7	13	19	0	-0.7
5	2	6	10	15	0	-0.6
10	1	3	5	8	0	-0.5
15	1	2	3	5	0	-0.4
16.8	1	2	3	4	0	-0.3
21.6	**1**	**1**	**2**	**3**	0	-0.2
f(d)=exp(δd)					0	-0.1
δ	n=1	n=2	n=3	n=4	0	0
1	4	14	25	37	0	0.1
2	2	6	12	18	0	0.2
3	1	4	8	12	0	0.3
4	1	3	6	9	0	0.4
5	1	2	4	7	0	0.5
6	1	2	3	5	0	0.6
7	**1**	**1**	**3**	**4**	0	0.7
					0	0.8
f(d)=ln(e+δd)					0.000008	0.9
δ	n=1	n=2	n=3	n=4	0.499992	1
0.5	46	136	248	363		
1	35	103	186	272		
1.5	32	94	169	246		
2	30	90	163	237		
2.5	30	89	160	233		
3	**30**	**89**	**160**	**232**		
3.5	30	89	160	233		

$f(d)=1.0001+\exp(-\delta d)$

δ	n=1	n=2	n=3	n=4
0.0001	14	42	76	111
0.001	10	30	54	78
0.025	10	29	52	76
0.05	10	30	54	78
0.01	**10**	**29**	**52**	**75**
0.1	11	32	56	82
0.25	13	37	66	96
0.5	17	50	87	126

$f(d)=\Phi(\delta d)$

δ	n=1	n=2	n=3	n=4
0.1	68	204	376	553
0.25	37	110	200	293
0.3	35	102	185	270
0.4	**33**	**97**	**174**	**253**
0.45	34	98	175	253
0.5	35	100	178	257
0.6	39	111	194	279
1	95	265	446	627

$f(d)=[\exp(\delta d)]/[1+\exp(\delta d)]$

δ	n=1	n=2	n=3	n=4
0.1	102	306	564	830
0.4	38	113	206	300
0.5	36	105	190	277
0.6	**35**	**103**	**184**	**267**
0.7	35	104	185	267
0.8	37	107	190	275
1	42	122	214	307

Quadratic Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}

δ	n=1	n=2	n=3	n=4	weights(p)	x
0.5	29	257	592	903	0.333331	-1
1	15	128	296	451	0	-0.9
1.5	10	86	197	301	0	-0.8
1.6	9	80	185	282	0	-0.7
1.7	8	75	174	265	0	-0.6
1.71	**8**	**75**	**173**	**264**	0	-0.5
					0	-0.4

f(d)=exp(δd)

δ	n=1	n=2	n=3	n=4	0	-0.3
					0	-0.2
0.45	11	94	219	335	0.000373	-0.1
0.5	10	84	197	301	0.332591	0
0.525	6	80	187	287	0.000373	0.1
0.55	8	76	197	273	0	0.2
0.56	8	75	175	268	0	0.3
0.57	10	73	172	263	0	0.4
0.58	**12**	**72**	**169**	**259**	0	0.5
					0	0.6

f(d)=ln(e+δd)

δ	n=1	n=2	n=3	n=4	0	0.7
					0	0.8
0.5	60	518	1199	1829	0	0.9
1	29	426	984	1501	0.333331	1
2	**46**	**404**	**932**	**1420**		
3	46	412	949	1446		

f(d)=1.0001+exp(-δd)

δ	n=1	n=2	n=3	n=4
0.0001	20	171	395	602
0.001	15	133	306	467
0.005	**15**	**130**	**300**	**458**
0.01	15	131	302	460
0.05	16	139	320	487
0.1	17	150	345	526
0.5	32	299	687	1045

f(d)=Φ(δd)

δ	n=1	n=2	n=3	n=4
0.25	50	439	1014	1546
0.28	**50**	**436**	**1006**	**1533**
0.3	49	438	1009	1537
0.35	51	455	1046	1594
0.4	53	489	1124	1710
0.5	63	620	1420	2157
0.6	78	876	2000	3034

f(d)=[exp(δd)]/[1+exp(δd)]

δ	n=1	n=2	n=3	n=4
0.35	54	471	1088	1659
0.39	53	463	1069	1629
0.4	53	462	1066	1625
0.45	**52**	**462**	**1066**	**1624**
0.46	52	463	1068	1627
0.475	52	465	1072	1633
0.5	53	470	1082	1648

Cubic Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}

δ	n=1	n=2	n=3	n=4	weights(p)	x
1.5	11	43	132	227	0.249528	-1
1.6	10	40	124	213	0	-0.9
1.7	8	38	117	200	0	-0.8
1.85	**12**	**24**	**107**	**184**	0	-0.7
1.925	22	48	100	176	0	-0.6
1.9455	30	66	102	172	0.113076	-0.5
1.95	32	72	110	164	0.137396	-0.4
1.9555	36	78	122	166	0	-0.3
1.96	40	86	134	182	0	-0.2
1.975	58	128	198	268	0	-0.1
					0	0

f(d)=exp(δd)

					0	0.1
δ	n=1	n=2	n=3	n=4	0	0.2
0.2	20	83	252	431	0	0.3
0.3	13	55	168	286	0.137396	0.4
0.35	11	47	143	245	0.113076	0.5
0.375	8	44	133	228	0	0.6
0.3825	8	42	130	223	0	0.7
0.39	10	41	127	218	0	0.8
0.3925	12	40	126	216	0	0.9
0.393	**12**	**40**	**125**	**216**	0.249528	1
0.4	x	x	x	x		

f(d)=ln(e+δd)

δ	n=1	n=2	n=3	n=4
1	50	208	638	1097
1.7	**49**	**204**	**628**	**1081**
1.8	49	204	629	1083
1.9	49	205	630	1086
2	49	205	632	1089
2.5	50	209	643	1108
3	53	221	681	1173

$f(d)=1.0001+\exp(-\delta d)$

δ	n=1	n=2	n=3	n=4
0.0001	20	81	249	428
0.001	16	67	204	351
0.005	**16**	**66**	**202**	**347**
0.01	16	66	203	349
0.1	19	79	244	420
0.5	48	197	623	1082

$f(d)=\Phi(\delta d)$

δ	n=1	n=2	n=3	n=4
0.1	70	293	893	1533
0.18	54	224	688	1183
0.2	54	220	677	1166
0.22	**53**	**219**	**676**	**1164**
0.25	54	223	689	1188
0.3	59	240	748	1293
0.4	79	324	1032	1798
0.5	124	521	1718	3019

$f(d)=[\exp(\delta d)]/[1+\exp(\delta d)]$

δ	n=1	n=2	n=3	n=4
0.2	64	265	808	1387
0.28	57	236	724	1245
0.3	57	233	717	1235
0.31	56	233	716	1232
0.32	**56**	**232**	**715**	**1231**
0.4	58	238	737	1272
0.5	64	264	826	1429

Quartic Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(d)=d^{δ}						
δ	n=1	n=2	n=3	n=4	weights(p)	x
0.5	48	228	410	606	0.199235	-1
1.5	16	76	136	201	0	-0.9
1.75	10	65	117	172	0	-0.8
1.8	12	63	114	167	0.113795	-0.7
1.9	24	58	108	159	0.0885	-0.6
1.925	32	56	104	156	0	-0.5
1.95	**43**	**80**	**116**	**152**	0	-0.4
1.9555	47	88	128	166	0	-0.3
					0	-0.2
f(d)=exp(δd)					0.00006	-0.1
δ	n=1	n=2	n=3	n=4	0.196819	0
0.1	49	227	412	609	0.00006	0.1
0.2	24	113	205	304	0	0.2
0.3	12	73	135	200	0	0.3
0.301	**12**	**72**	**134**	**199**	0	0.4
0.302	16	70	131	196	0	0.5
0.5	x	x	x	x	0.0885	0.6
					0.113795	0.7
f(d)=ln(e+δd)					0	0.8
δ	n=1	n=2	n=3	n=4	0	0.9
0.8	77	364	656	970	0.199235	1
1	**76**	**359**	**646**	**956**		
1.5	76	361	648	959		
2	78	368	662	979		

f(d)=1.0001+exp(-δd)

δ	n=1	n=2	n=3	n=4
0.0001	29	137	246	363
0.0005	25	118	213	315
0.001	**25**	**116**	**209**	**309**
0.01	25	117	210	311
0.1	31	148	265	392
0.5	103	508	906	1340

f(d)=Φ(δd)

δ	n=1	n=2	n=3	n=4
0.1	96	447	807	1194
0.16	82	387	697	1031
0.17	**82**	**387**	**696**	**1029**
0.18	82	388	698	1032
0.19	83	391	711	1052
0.2	84	396	711	1052
0.22	86	411	737	1090
0.3	112	546	975	1442

f(d)=[exp(δd)]/[1+exp(δd)]

δ	n=1	n=2	n=3	n=4
0.2	90	423	763	1129
0.24	87	410	738	1092
0.25	87	409	736	1089
0.26	**87**	**409**	**736**	**1088**
0.3	88	416	748	1106
0.4	100	478	856	1265
0.5	123	600	1072	1585

4.1 Replacing d_j by F_j in algorithm (3.13)

The idea of attempting to improve the iteration performance is always under consideration. We can improve the performance in D-optimal designs replacing d_j by F_j for some choices of $f(d)$. The motivation is that for D-optimal criterion, the partial derivatives of $\phi_D(p)$ are positive and centered at k(the number of parameters $\underline{\theta}$ in the linear model). This may be an disadvantage of some choices of $f(x)$, where $x = d$, ie. $f(d) = \Phi(\delta d)$. However, we know that the vertex directional derivatives are $F_j = d_j - \sum_{i=1}^{J} p_i d_i$. This implies that $\sum_{j=1}^{J} p_j F_j = 0$. So, if we choose a function $f(.)$ which is centred at zero, we may get better convergence rates of the algorithm. Thus, we replace d_j by F_j. Algorithm (3.13) now becomes

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(F_j^{(r)})}{\sum_{i=1}^{J} p_i^{(r)} f(F_i^{(r)})} \tag{4.10}$$

where $F_j^{(r)}$ are the vertex directional derivatives at $p_j^{(r)}$ and $f(F)$ satisfying the conditions as in (3.13). We consider only choices of $f(x) = \Phi(x)$ and $f(x) = \frac{exp(x)}{1+exp(x)}$, where $x = F$. The function $f(x) = exp(\delta x)$ is also a possible choice but the performance of the algorithm has no difference, for the reason that p_j at the next iteration becomes

$$
\begin{aligned}
p_j^{(r+1)} &= \frac{p_j^{(r)} f(F_j^{(r)})}{\sum_{i=1}^{J} p_i^{(r)} f(F_i^{(r)})} = \frac{p_j^{(r)} e^{\delta F_j^{(r)}}}{\sum_{i=1}^{J} p_i^{(r)} e^{\delta F_i^{(r)}}} \\
&= \frac{p_j^{(r)} e^{\delta d_j} \cdot e^{-\delta \sum_{i=1}^{J} p_i d_i}}{\sum_{i=1}^{J} [p_i(r) e^{\delta d_i} \cdot e^{-\delta \sum_{k=1}^{J} p_k d_k}]} \\
&= \frac{p_j^{(r)} e^{\delta d_j}}{\sum_{i=1}^{J} p_i^{(r)} e^{\delta d_i}} = \frac{p_j^{(r)} f(d_j^{(r)})}{\sum_{i=1}^{J} p_i^{(r)} f(d_i^{(r)})}.
\end{aligned}
$$

In Table 4.3, we report the performance of the algorithm for models (4.1) - (4.4) the same way as in Tables 4.1 and 4.2, but replacing d_j by F_j. Mandal (2000) reported that for Examples 1-5, both choices of $f(x) = \Phi(x)$ and $f(x) = \frac{exp(x)}{1+exp(x)}$ with $x = F$ improve the convergence in all cases over all other choices of $f(x)$ with $x = d$. This observation almost reflects the results here that $f(x) = \Phi(x)$ **OR** $f(x) = \frac{exp(x)}{1+exp(x)}$ improved the convergence in all cases over all other choices of $f(.)$ except for some models. Specifically, for the Simple Linear Regression Model (4.1), the choice of $f(d) = d^{\delta}$ seems to perform equally well, if not better, comparing to both $f(F) = \Phi(\delta F)$ and $f(F) = \frac{exp(\delta F)}{1+exp(\delta F)}$ in the algorithm. For the Quartic Model (4.4), relatively $f(F) = \frac{exp(\delta F)}{1+exp(\delta F)}$ performs better than all other choices of $f(.)$ in the algorithm.

Note that for the Simple Linear Regression Model (4.1), further exploration of the choices of δ values for the three functions $f(d) = d^{\delta}$, $f(F) = \Phi(\delta F)$ and $f(F) = \frac{exp(\delta F)}{1+exp(\delta F)}$ is possible and better convergence rates may be obtained. We stop our investigation for the time being.

Table 4.3

Linear Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(F)=$\Phi(\delta F)$					weights(p)	x
δ	n=1	n=2	n=3	n=4	0.49986	-1
1	7	19	33	47	0.00014	-0.9
2	5	11	17	24	0	-0.8
3	4	8	12	16	0	-0.7
4	3	6	9	12	0	-0.6
5	3	6	8	10	0	-0.5
6	3	5	7	8	0	-0.4
8.4	3	4	5	5	0	-0.3
15.1	3	4	4	4	0	-0.2
31.3	**3**	**3**	**3**	**3**	0	-0.1
f(F)=[exp(δF)]/[1+exp(δF)]					0	0
δ	n=1	n=2	n=3	n=4	0	0.1
0.5	19	56	102	149	0	0.2
0.8	12	36	64	93	0	0.3
1	10	29	52	75	0	0.4
1.5	7	20	35	50	0	0.5
2	6	16	27	38	0	0.6
3	5	11	18	25	0	0.7
5	4	8	12	16	0	0.8
7.5	3	6	8	11	0.00014	0.9
14.75	3	4	5	6	0.49986	1
29	3	4	4	4		
72	**3**	**3**	**3**	**3**		

Quadratic Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(F)=$\Phi(\delta F)$						
δ	n=1	n=2	n=3	n=4	weights(p)	x
0.5	12	107	247	377	0.333331	-1
0.75	8	72	165	251	0	-0.9
0.8	8	68	155	235	0	-0.8
0.81	8	66	153	232	0	-0.7
0.82	**8**	**64**	**150**	**230**	0	-0.6
0.83	8	48	114	276	0	-0.5
					0	-0.4
f(F)=[exp(δF)]/[1+exp(δF)]					0	-0.3
δ	n=1	n=2	n=3	n=4	0	-0.2
0.5	20	171	395	602	0.000363	-0.1
0.8	12	107	247	376	0.33261	0
0.9	11	95	219	334	0.000363	0.1
1	10	86	198	301	0	0.2
1.3	8	66	152	231	0	0.3
1.31	8	64	150	230	0	0.4
1.32	**8**	**60**	**142**	**212**	0	0.5
1.33	8	50	180	638	0	0.6
					0	0.7
					0	0.8
					0	0.9
					0.333331	1

Cubic Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(F)=$\Phi(\delta F)$

δ	n=1	n=2	n=3	n=4	weights(p)	x
0.5	10	40	123	212	0.249528	-1
0.6	9	31	102	176	0	-0.9
0.61	9	29	99	173	0	-0.8
0.62	**9**	**25**	**79**	**157**	0	-0.7
					0	-0.6
					0.113067	-0.5

f(F)=[exp(δF)]/[1+exp(δF)]

δ	n=1	n=2	n=3	n=4	0.137404	-0.4
0.5	16	65	198	341	0	-0.3
0.7	11	46	141	242	0	-0.2
0.8	10	40	123	212	0	-0.1
0.9	9	35	109	188	0	0
0.95	9	33	103	178	0	0.1
0.97	9	31	101	174	0	0.2
0.98	9	27	97	171	0	0.3
0.99	**9**	**25**	**83**	**165**	0.137404	0.4
1	9	23	282	586	0.113067	0.5
					0	0.6
					0	0.7
					0	0.8
					0	0.9
					0.249528	1

Quartic Model: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(F)=$\Phi(\delta F)$						
δ	n=1	n=2	n=3	n=4	weights(p)	x
0.3	20	95	171	252	0.199235	-1
0.4	15	71	128	188	0	-0.9
0.44	13	65	116	171	0	-0.8
0.45	13	63	114	167	0.113794	-0.7
0.46	11	62	111	164	0.088502	-0.6
0.47	9	61	109	160	0	-0.5
0.48	11	59	107	157	0	-0.4
0.49	**15**	**51**	**95**	**151**	0	-0.3
0.5	37	160	294	426	0	-0.2
					0.000057	-0.1
f(F)=[exp(δF)]/[1+exp(δF)]					0.196825	0
δ	n=1	n=2	n=3	n=4	0.000057	0.1
0.5	19	91	164	241	0	0.2
0.7	14	65	117	172	0	0.3
0.72	13	63	115	169	0	0.4
0.74	11	62	110	162	0	0.5
0.76	9	60	108	158	0.088502	0.6
0.77	11	59	106	156	0.113794	0.7
0.775	**11**	**55**	**105**	**155**	0	0.8
0.78	13	110	253	386	0	0.9
0.8	42	222	414	608	0.199235	1

4.2 D-optimal Design in Second-Order Model with Two Variables

It is also of interest to work on the models with more than one design variable. In Table 4.4, we worked out the design as an example for a linear model in two variables with interaction terms, $E[y|x_1, x_2] = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2$. We again used the $f(F) = \frac{exp(\delta F)}{1+exp(\delta F)}$ in the algorithm with $\delta = 0.5$ chosen without concern of the rate of convergence. Taking the convex combinations of the approximated solution, the optimal support points with corresponding design weights for this model are

$$p^* = \left\{ \begin{array}{ccccccccccc} \text{weights(p)} & 0.146 & 0.08 & 0.146 & 0.08 & 0.096 & .08 & 0.146 & 0.08 & 0.146 \\ \text{x1} & -1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 \\ \text{x2} & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \end{array} \right\}. \tag{4.11}$$

Table 4.4 Quadratic Model of 2 Variables with Interaction: Number of iterations needed to achieve $max\{F_j\} \leq 10^{-n}$.

f(F)=[exp(δF)]/[1+exp(δF)]

δ	n=1	n=2	n=3	n=4
0.5	39	220	399	571

weights(p)	x1	x2	weights(p)	x1	x2
0.145791	-1	-1	0	-0.9	-1
0	-1	-0.9	0	-0.9	-0.9
0	-1	-0.8	0	-0.9	-0.8
0	-1	-0.7	0	-0.9	-0.7
0	-1	-0.6	0	-0.9	-0.6
0	-1	-0.5	0	-0.9	-0.5
0	-1	-0.4	0	-0.9	-0.4
0	-1	-0.3	0	-0.9	-0.3
0	-1	-0.2	0	-0.9	-0.2
0.000024	-1	-0.1	0	-0.9	-0.1
0.080113	-1	0	0	-0.9	0
0.000024	-1	0.1	0	-0.9	0.1
0	-1	0.2	0	-0.9	0.2
0	-1	0.3	0	-0.9	0.3
0	-1	0.4	0	-0.9	0.4
0	-1	0.5	0	-0.9	0.5
0	-1	0.6	0	-0.9	0.6
0	-1	0.7	0	-0.9	0.7
0	-1	0.8	0	-0.9	0.8
0	-1	0.9	0	-0.9	0.9
0.145791	-1	1	0	-0.9	1

weights(p)	x1	x2	weights(p)	x1	x2
0	-0.8	-1	0	-0.7	-1
0	-0.8	-0.9	0	-0.7	-0.9
0	-0.8	-0.8	0	-0.7	-0.8
0	-0.8	-0.7	0	-0.7	-0.7
0	-0.8	-0.6	0	-0.7	-0.6
0	-0.8	-0.5	0	-0.7	-0.5
0	-0.8	-0.4	0	-0.7	-0.4
0	-0.8	-0.3	0	-0.7	-0.3
0	-0.8	-0.2	0	-0.7	-0.2
0	-0.8	-0.1	0	-0.7	-0.1
0	-0.8	0	0	-0.7	0
0	-0.8	0.1	0	-0.7	0.1
0	-0.8	0.2	0	-0.7	0.2
0	-0.8	0.3	0	-0.7	0.3
0	-0.8	0.4	0	-0.7	0.4
0	-0.8	0.5	0	-0.7	0.5
0	-0.8	0.6	0	-0.7	0.6
0	-0.8	0.7	0	-0.7	0.7
0	-0.8	0.8	0	-0.7	0.8
0	-0.8	0.9	0	-0.7	0.9
0	-0.8	1	0	-0.7	1

weights(p)	x1	x2	weights(p)	x1	x2
0	-0.6	-1	0	-0.5	-1
0	-0.6	-0.9	0	-0.5	-0.9
0	-0.6	-0.8	0	-0.5	-0.8
0	-0.6	-0.7	0	-0.5	-0.7
0	-0.6	-0.6	0	-0.5	-0.6
0	-0.6	-0.5	0	-0.5	-0.5
0	-0.6	-0.4	0	-0.5	-0.4
0	-0.6	-0.3	0	-0.5	-0.3
0	-0.6	-0.2	0	-0.5	-0.2
0	-0.6	-0.1	0	-0.5	-0.1
0	-0.6	0	0	-0.5	0
0	-0.6	0.1	0	-0.5	0.1
0	-0.6	0.2	0	-0.5	0.2
0	-0.6	0.3	0	-0.5	0.3
0	-0.6	0.4	0	-0.5	0.4
0	-0.6	0.5	0	-0.5	0.5
0	-0.6	0.6	0	-0.5	0.6
0	-0.6	0.7	0	-0.5	0.7
0	-0.6	0.8	0	-0.5	0.8
0	-0.6	0.9	0	-0.5	0.9
0	-0.6	1	0	-0.5	1

weights(p)	x1	x2	weights(p)	x1	x2
0	-0.4	-1	0	-0.3	-1
0	-0.4	-0.9	0	-0.3	-0.9
0	-0.4	-0.8	0	-0.3	-0.8
0	-0.4	-0.7	0	-0.3	-0.7
0	-0.4	-0.6	0	-0.3	-0.6
0	-0.4	-0.5	0	-0.3	-0.5
0	-0.4	-0.4	0	-0.3	-0.4
0	-0.4	-0.3	0	-0.3	-0.3
0	-0.4	-0.2	0	-0.3	-0.2
0	-0.4	-0.1	0	-0.3	-0.1
0	-0.4	0	0	-0.3	0
0	-0.4	0.1	0	-0.3	0.1
0	-0.4	0.2	0	-0.3	0.2
0	-0.4	0.3	0	-0.3	0.3
0	-0.4	0.4	0	-0.3	0.4
0	-0.4	0.5	0	-0.3	0.5
0	-0.4	0.6	0	-0.3	0.6
0	-0.4	0.7	0	-0.3	0.7
0	-0.4	0.8	0	-0.3	0.8
0	-0.4	0.9	0	-0.3	0.9
0	-0.4	1	0	-0.3	1

weights(p)	x1	x2	weights(p)	x1	x2
0	-0.2	-1	0.000024	-0.1	-1
0	-0.2	-0.9	0	-0.1	-0.9
0	-0.2	-0.8	0	-0.1	-0.8
0	-0.2	-0.7	0	-0.1	-0.7
0	-0.2	-0.6	0	-0.1	-0.6
0	-0.2	-0.5	0	-0.1	-0.5
0	-0.2	-0.4	0	-0.1	-0.4
0	-0.2	-0.3	0	-0.1	-0.3
0	-0.2	-0.2	0	-0.1	-0.2
0	-0.2	-0.1	0	-0.1	-0.1
0	-0.2	0	0.000028	-0.1	0
0	-0.2	0.1	0	-0.1	0.1
0	-0.2	0.2	0	-0.1	0.2
0	-0.2	0.3	0	-0.1	0.3
0	-0.2	0.4	0	-0.1	0.4
0	-0.2	0.5	0	-0.1	0.5
0	-0.2	0.6	0	-0.1	0.6
0	-0.2	0.7	0	-0.1	0.7
0	-0.2	0.8	0	-0.1	0.8
0	-0.2	0.9	0	-0.1	0.9
0	-0.2	1	0.000024	-0.1	1

weights(p)	x1	x2	weights(p)	x1	x2
0.080113	0	-1	0.000024	0.1	-1
0	0	-0.9	0	0.1	-0.9
0	0	-0.8	0	0.1	-0.8
0	0	-0.7	0	0.1	-0.7
0	0	-0.6	0	0.1	-0.6
0	0	-0.5	0	0.1	-0.5
0	0	-0.4	0	0.1	-0.4
0	0	-0.3	0	0.1	-0.3
0	0	-0.2	0	0.1	-0.2
0.000028	0	-0.1	0	0.1	-0.1
0.096081	0	0	0.000028	0.1	0
0.000028	0	0.1	0	0.1	0.1
0	0	0.2	0	0.1	0.2
0	0	0.3	0	0.1	0.3
0	0	0.4	0	0.1	0.4
0	0	0.5	0	0.1	0.5
0	0	0.6	0	0.1	0.6
0	0	0.7	0	0.1	0.7
0	0	0.8	0	0.1	0.8
0	0	0.9	0	0.1	0.9
0.080113	0	1	0.000024	0.1	1

weights(p)	x1	x2	weights(p)	x1	x2
0	0.2	-1	0	0.3	-1
0	0.2	-0.9	0	0.3	-0.9
0	0.2	-0.8	0	0.3	-0.8
0	0.2	-0.7	0	0.3	-0.7
0	0.2	-0.6	0	0.3	-0.6
0	0.2	-0.5	0	0.3	-0.5
0	0.2	-0.4	0	0.3	-0.4
0	0.2	-0.3	0	0.3	-0.3
0	0.2	-0.2	0	0.3	-0.2
0	0.2	-0.1	0	0.3	-0.1
0	0.2	0	0	0.3	0
0	0.2	0.1	0	0.3	0.1
0	0.2	0.2	0	0.3	0.2
0	0.2	0.3	0	0.3	0.3
0	0.2	0.4	0	0.3	0.4
0	0.2	0.5	0	0.3	0.5
0	0.2	0.6	0	0.3	0.6
0	0.2	0.7	0	0.3	0.7
0	0.2	0.8	0	0.3	0.8
0	0.2	0.9	0	0.3	0.9
0	0.2	1	0	0.3	1

weights(p)	x1	x2	weights(p)	x1	x2
0	0.4	-1	0	0.5	-1
0	0.4	-0.9	0	0.5	-0.9
0	0.4	-0.8	0	0.5	-0.8
0	0.4	-0.7	0	0.5	-0.7
0	0.4	-0.6	0	0.5	-0.6
0	0.4	-0.5	0	0.5	-0.5
0	0.4	-0.4	0	0.5	-0.4
0	0.4	-0.3	0	0.5	-0.3
0	0.4	-0.2	0	0.5	-0.2
0	0.4	-0.1	0	0.5	-0.1
0	0.4	0	0	0.5	0
0	0.4	0.1	0	0.5	0.1
0	0.4	0.2	0	0.5	0.2
0	0.4	0.3	0	0.5	0.3
0	0.4	0.4	0	0.5	0.4
0	0.4	0.5	0	0.5	0.5
0	0.4	0.6	0	0.5	0.6
0	0.4	0.7	0	0.5	0.7
0	0.4	0.8	0	0.5	0.8
0	0.4	0.9	0	0.5	0.9
0	0.4	1	0	0.5	1

weights(p)	x1	x2	weights(p)	x1	x2
0	0.6	-1	0	0.7	-1
0	0.6	-0.9	0	0.7	-0.9
0	0.6	-0.8	0	0.7	-0.8
0	0.6	-0.7	0	0.7	-0.7
0	0.6	-0.6	0	0.7	-0.6
0	0.6	-0.5	0	0.7	-0.5
0	0.6	-0.4	0	0.7	-0.4
0	0.6	-0.3	0	0.7	-0.3
0	0.6	-0.2	0	0.7	-0.2
0	0.6	-0.1	0	0.7	-0.1
0	0.6	0	0	0.7	0
0	0.6	0.1	0	0.7	0.1
0	0.6	0.2	0	0.7	0.2
0	0.6	0.3	0	0.7	0.3
0	0.6	0.4	0	0.7	0.4
0	0.6	0.5	0	0.7	0.5
0	0.6	0.6	0	0.7	0.6
0	0.6	0.7	0	0.7	0.7
0	0.6	0.8	0	0.7	0.8
0	0.6	0.9	0	0.7	0.9
0	0.6	1	0	0.7	1

weights(p)	x1	x2	weights(p)	x1	x2
0	0.8	-1	0	0.9	-1
0	0.8	-0.9	0	0.9	-0.9
0	0.8	-0.8	0	0.9	-0.8
0	0.8	-0.7	0	0.9	-0.7
0	0.8	-0.6	0	0.9	-0.6
0	0.8	-0.5	0	0.9	-0.5
0	0.8	-0.4	0	0.9	-0.4
0	0.8	-0.3	0	0.9	-0.3
0	0.8	-0.2	0	0.9	-0.2
0	0.8	-0.1	0	0.9	-0.1
0	0.8	0	0	0.9	0
0	0.8	0.1	0	0.9	0.1
0	0.8	0.2	0	0.9	0.2
0	0.8	0.3	0	0.9	0.3
0	0.8	0.4	0	0.9	0.4
0	0.8	0.5	0	0.9	0.5
0	0.8	0.6	0	0.9	0.6
0	0.8	0.7	0	0.9	0.7
0	0.8	0.8	0	0.9	0.8
0	0.8	0.9	0	0.9	0.9
0	0.8	1	0	0.9	1

weights(p)	x1	x2
0.145791	1	-1
0	1	-0.9
0	1	-0.8
0	1	-0.7
0	1	-0.6
0	1	-0.5
0	1	-0.4
0	1	-0.3
0	1	-0.2
0.000024	1	-0.1
0.080113	1	0
0.000024	1	0.1
0	1	0.2
0	1	0.3
0	1	0.4
0	1	0.5
0	1	0.6
0	1	0.7
0	1	0.8
0	1	0.9
0.145791	1	1

Chapter 5

Conclusion

In conclusion to this thesis, we learn that by the Carathéodory's Theorem it guarantees that any continuous measure can be replaced by at least one finite discrete probability distribution. Discretization of design space is very helpful in finding an optimal design measure by programmed numerical technique. We also learned to construct an optimal design using a class of multiplicative algorithm which will obtain a convergence when the criterion function is concave, for example, the concavity of the D-optimal design criterion function $\psi(x)$. By the Vertex Direction Optimality Theorem, we can determine whether a solution is at optimum by checking the conditions of vertex directional derivatives. An optimal design depends on the type of criterion we are interested in and the choice of model selected to fit the data. A relative fast algorithm was suggested. In particular, some examples were provided in the thesis on how to construct a D-optimal design. For linear regression models (4.1) - (4.4), we determined at which points in the design space $[-1, 1]$, discretized in an equal interval of 0.1, should observations be taken and with what proportions. This application can be made to other optimal design criteria.

For linear regression models (4.1) - (4.4), we noticed that optimal designs have k

optimal design points evenly spread out over the design space $[-1, 1]$ and are symmetric about zero. Optimal design points have corresponding optimal weights of $1/k$. Here k is the number of parameters in the models. This means that we achieved the correct solution for the D-optimal design.

One way to obtain better approximate designs is to discretize the interval $[-1, 1]$ or $[0, 1]$ into finer equal space of 0.01 or 0.001, sacrificing some program running time of the algorithm since larger vectors and matrices need to be processed by computer software. In the future we would like to work on this finer discretization of the design space with a faster convergence rate in finding optimal designs.

We also would like to work on constructing optimal designs with A- and E- and other optimality criteria. Some further explorations of the construction of optimal designs are such as, other choices of $f(.)$ or variations of $f(d)$ or $f(F)$ with two or more free parameters for improving the performance of algorithm (3.13), and construction of optimal designs for models non-linear in parameters.

Bibliography

Atkinson, A. C. and Donev, A. N. (1992). Optimum Experimental Designs. Oxford Statistical Science Series-8. Oxford University Press, Oxford.

Atwood, C. L. (1969). Optimal and efficient designs of experiments. Annals of Statistics, 40:1570–1602.

Atwood, C. L. (1976). Convergent design sequences for sufficiently regular optimality criteria, ii: Singular case. Annals of Statistics, 4:1124–1138.

Bates, D. M. and Watts, D. G. (1988). Nonlinear Regression Analysis and its Applications. J. Wiley & Sons, NY.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. Annals of Mathematical Statistics, 24:586–602.

Elfving, G. (1952). Optimum allocation in linear regression theory. Annals of Mathematical Statistics, 23:255–262.

Fedorov, V. V. (1972). Theory of Optimal Experiments. Academic Press, New York and London.

Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth International Group, Belmont, California.

Karlin, S. and Studden, W. J. (1966). Optimal experimental designs. Annals Mathematical Statistics, 37:783–815.

Kiefer, J. (1959). Optimum experimental designs(with duscussion). J. Roy. Statist. Soc. Series B, 21:272–319.

Kiefer, J. (1974). General equivalence theory for optimum designs(approximate theory). Annals of Statistics, 2:849–879.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canadian J. Math., 12:363–366.

Mandal, S. (2000). Construction of Optimizing Distributions with Applications in Estimation and Optimal Design. PhD thesis, University of Glasgow.

Mandal, S. and Torsney, B. (2000). Algorithms for the construction of optimizing distributions. Communications in Statistics - Theory and Methods, 29:1219–1231.

Mandal, S. and Torsney, B. (2004). Constructon of optimal designs using a clustering approach. Journal of Statistical Planning and Inference, (To Appear).

Mandal, S., Torsney, B., and Carriere, K. (2004). Constructing optimal designs with constraints. Journal of Statistical Planning and Inference, (To Appear).

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996). Applied Linear Statistical Models, 4th Edition. WCB McGraw-Hill, 4th edition.

Pazman, A. (1986). Foundations of Optimum Experimental Design. Reidel, Dordrecht.

Rhode, C. A. (1965). Generalised inverse of partitioned matrices. J. Soc. Indust. Appl. Math, 13:1033–1053.

Sibson, R. (1974). D_A-optimality and duality. progress in statistics. Colloq. Math. Soc. Janos. Bolyai, 9:677–692.

Silvey, S. (1980). Optimal Design. Chapman and Hall, London.

Silvey, S. D. and Titterington, D. M. (1973). A geometric approach to optimal design theory. Biometrika, 60:21–32.

Silvey, S. D., Titterington, D. M., and Torsney, B. (1978). An algorithm for optimal designs on a finite design space. Communications in Statistics A, 7:1379–1389.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constraints and the guidance they give towards a proper choice of the distribution of observations. Biometrika, 12:1–85.

Torsney, B. (1977). Contribution to discussion of "maximum likelihood from imcomplete data via the em algorithm" by dempster et al. J. Roy. Statist. Soc. Series B, 39:26–27.

Torsney, B. (1981). Algorithms for a Constrained Optimisation Problem with Applications in Statistics and Optimum Design. PhD thesis, University of Glasgow.

Torsney, B. (1988). Optimal design and analyais of experiments(edited by dodge, y., fedorov, v. v. and wynn, h. p.). In Computing Optimising Distributions with Applications in Design, Estimation and Image Processing, pages 361–370. Elsevier Science Publishers B. V., North Holland.

Torsney, B. and Alahmadi, A. M. (1992). Model oriented data analysis. proc. 2nd iiasa workshop in st. kyrik, bulgaria(edited by fedorov, v. v., müller, w. g. and vuchkov, i. n.). In Further Development of Algorithms for Constructing Optimizing Distributions., pages 361–370. Physica-Verlag.

Torsney, B. and Mandal, S. (2001). Construction of constrained optimal designs. In Optimum Design 2000, pages 141–152. Kluwer Academic Publishers.

Whittle, P. (1973). Some general points in the theory of optimal experimental designs. J. Roy. Statist. Soc. Series B, 35:123–130.

Wu, C. F. J. (1978). Some iterative procedures for generating nonsngular optimal designs. Communications in Statistics A, 7:1399–1412.

Wynn, H. P. (1970). The sequential generation of d-optimal experiment designs. Annals of Mathematical Statistics, 41:1655–1664.