Optimal Design Construction with respect to Several Probability Distributions using Clustering Approach

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

Master of Science

submitted by

Bodawatte Gedara Muditha Lakmali

Department of Statistics

University of Manitoba

Winnipeg, Manitoba

ABSTRACT

Optimization has been a basic tool in most of the areas of theoretical and applied sciences. Optimal design theory focuses on identifying an experimental design that makes the variances of a model's parameter estimates as small as possible, thereby allowing the model to make the most accurate predictions.

In this thesis, we have tried to address an important problem in optimal regression design, namely the application of a clustering approach to solve optimization problems with respect to several probability distributions and to further improve the convergence of a class of algorithms using the properties of directional derivatives. When we run a multiplicative algorithm to construct an optimal design, the design turns out to be a distribution defined on a disjoint set of clusters of points. This situation arises when many design weights converge to zero at the optimum. We replace the single distribution by conditional distributions and a marginal distribution across the clusters. Motivated by this, we transform this clustering approach to a general problem of optimization with respect to several distributions. The number of probability distributions depends on the number of parameters in the model.

We focus on constructing designs for two criteria of interest such as D-optimality and D_s-optimality criteria. The D-optimality is the most important and popular design criterion in the literature. The D_s-optimality is also quite important when we are interested in a subset of parameters. This situation arises when we are more interested in some of the terms (for example, the even or odd power terms) in the model. We also constructed some D_s-optimal designs using analytic approach. We explore several mod-

els both in one and two design variables. The graphical interpretation was carried out using the plots of weights versus design points as well as plots of variance functions versus design points. We did a powerful improvement in the convergence of the algorithms by combining the clustering approach and the properties of the directional derivatives. The results are promising. This approach is instrumental in improving the convergence of the algorithm and allowing the model to obtain the optimal design saving cost and time.

ACKNOWLEDGEMENTS

First and foremost I am extremely grateful to my supervisor, Prof. Saumen Mandal for his invaluable advice, continuous support, and patience during my MSc program. His immense knowledge and plentiful experience have helped to conduct my research successfully. It's a privilege to work with such a generous and knowledgeable person like you.

I am also thankful to my committee members Dr. Po Yang and Dr. Srimantoorao Appadoo for monitoring my work and spending time reading this thesis and providing valuable suggestions to deepen my understanding of the research topic.

I gratefully acknowledge the financial support from the Department of Statistics, Faculty of Graduate Studies, Faculty of Science, University of Manitoba and Prof. Saumen Mandal's NSERC grant.

I would like to thank all the professors and instructors at the Department of Statistics at University of Manitoba for their wonderful lessons in all courses. Also, I am grateful to all the support staff of the department for taking care of all our administrative work.

Last but not least, I would like to thank my beloved parents and family for their endless support and encouragement. Finally, I want to express my deepest gratitude to my loving husband for his kindness and encouragement when the times got rough. I greatly appreciate that.

Contents

List of Figures

Chapter 1

Introduction

It might lead us to answer several questions when we are conducting research with an experimental design problem, since it might have specific constraints and objectives. Questions arise such as "(i) Which design is the best? (ii) How to reduce the sampling cost (iii) Can we reduce the sample size further by making the design as efficient as possible at minimal cost? (iv) Can I use this fitted model to make inferences about the people older than 30 years?"

Optimal design theory is an useful tool in such situations. It offers a good foundation to our design by answering those questions. This theory helps to identify the best design using computer algorithms and it can be done analytically for simpler problems. Moreover, the optimal design theory enables a researcher to construct flexible designs in a way that they meet the goals of the studies more realistically.

1.1 Optimal Design Theory in Linear Regression

In regression analysis, we are interested in finding the relationship between two or more variables (X's as independent variables and Y as the dependent variable). Here our primary interest is to obtain the parameter estimators efficiently. In general, the efficiency of an estimator increases as its variance becomes smaller since it expresses less uncertainty. There are 3 approaches to measure the uncertainty of estimators:

- The variance of the estimators.
- Confidence interval for the parameters.
- The variance of the predicted response $\hat{y}$ for an arbitrary value of x.

A confidence ellipse is a figure that contains all the above information about the uncertainty of parameters. In Figure 1.1, the intersection of the dotted axes represents the point estimators. Their lengths are related to the variances of estimators. The covariance of the estimators determines the direction of the axes. When adding more parameters to the model, this ellipse can be extended to form an ellipsoid in more than two dimensions. Generally, the volume or contour of such ellipsoids and the length of their axes can be used to measure the uncertainty. The most significant fact about this ellipsoid is that it can be used to find an optimal design for the simultaneous estimation of the parameters. It has been proved that the variance covariance

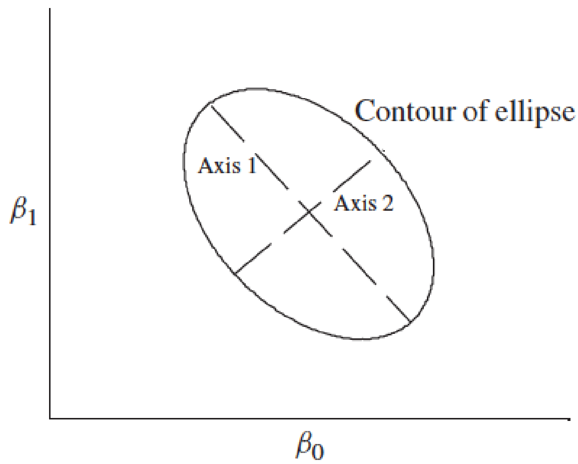


Figure 1.1: Confidence ellipse for two parameters β_0 and β_1 in a simple linear regression

matrix of parameter estimators can be used to determine the shape and the form of the confidence ellipsoid as well as to measure the efficiency of a design. In the literature several optimality criteria have been discussed which can be explained by the properties of confidence ellipsoid and these will be discussed later under Section 1.5.

Next, we will study some methodological concepts of the linear models along with the basic concepts of optimal design theory such as definition of a design, variance function, information matrix including their properties.

Consider a model in the following form:

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

where

- y : response variable.

- $p(.)$: probability model.

- $\underline{x}$: design variables such that $\underline{x} = (x_1, x_2, ..., x_m)^T$ which can be chosen by the experimenter. Values of these variables fall under a m-dimensional space called a design space and it is denoted by $\mathcal{X}$ where $\underline{x} \in \mathcal{X} \in \mathbb{R}^m$. In most cases, this is considered compact.

- $\underline{\theta}$: a k-dimensional vector of unknown parameters such that $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_k)^T$.

- σ : a nuisance parameter which is fixed and unknown. Generally, this is supposed to be independent of $\underline{x}$. (Nuisance parameter is a parameter which is not of our main interest. But it has to be accounted in the analysis of the parameters of interest)

The expected value of a linear model can be expressed in the form:

$$E(y|\underline{x}, \underline{\theta}, \sigma) = \underline{f}^T(\underline{x})\underline{\theta} \tag{1.2}$$

where $\underline{f}(\underline{x}) = (f_1(\underline{x}), \cdots, f_k(\underline{x}))^T$ and $f_1, f_2, ..., f_k$ are the regression functions defined on $\mathcal{X}$.

As mentioned earlier, our primary interest is to obtain 'best' point estimates of $\hat{\theta}$ for some or all the parameters of $\underline{\theta}$. Hence, the selection of $\underline{x}$ has to be made carefully. This allocation of observations, say n to the elements of $\mathcal{X}$ is identified as an optimal design.

For the moment, let us assume that the inferences are made using point estimation and $\hat{\theta}$ is an unbiased estimator for $\underline{\theta}$. So, the dispersion matrix of $\hat{\theta}$ about $\underline{\theta}$ can be expressed as $D(\hat{\underline{\theta}}) = E([\hat{\underline{\theta}} - \underline{\theta}][\hat{\underline{\theta}} - \underline{\theta}]^T)$. For a reliable and accurate value for $\hat{\theta}$, $D(\hat{\underline{\theta}})$ has to be as small as possible.

Suppose that $y_1, y_2, ..., y_n$ are the observations obtained at the same $\underline{x}$ value. If they are independent with equal variance σ^2, then model 1.2 can be expressed in the following form:

$$E(Y) = X\underline{\theta}, \qquad D(Y) = \sigma^2 I_n. \tag{1.3}$$

Here,

- $Y = (y_1, y_2, \cdots, y_n)$
- $X : n \times k$ design matrix
- $I_n : n \times n$ identity matrix
- $D(Y)$: Dispersion/Covariance matrix of Y.

The Best Linear Unbiased Estimators of $\underline{\theta}$ can be obtained using the equation

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

But, if the interest is on all the parameters of $\underline{\theta}$, $(X^T X)$ has to be non-singular and the least squares estimators of $\hat{\underline{\theta}}$ are given by

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

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

Let $\hat{y}(\underline{x})$ be the predicted value of the response at $\underline{x}$ and it can be expressed as

$$\begin{aligned} \hat{y}(\underline{x}) &= f_1(\underline{x})\hat{\theta}_1 + f_2(\underline{x})\hat{\theta}_2 + \cdots + f_k(\underline{x})\hat{\theta}_k \\ &= \underline{f}^T(\underline{x})\underline{\hat{\theta}} \end{aligned} \tag{1.6}$$

where $\underline{f}(\underline{x}) = (f_1(\underline{x}), f_2(\underline{x}), \cdots, f_k(\underline{x}))^T$.

1.2 Exact Designs versus Approximate Designs

Suppose that we are dealing with a sample whose total sample size is N. Let d_j where $j = 1, 2, ..., m$ be the design points from a pre-selected design space and n_j be the number of observations taken from each design point. So that, $\sum_{j=1}^{m} n_j = N$. So, the design can be written as

$$\xi = \left\{ \begin{array}{cccc} d_1 & d_2 & \cdots & d_m \\ n_1 & n_2 & \cdots & n_m \end{array} \right\}.$$

This design tells us how many subjects are assigned to each design point. The subjects are given as integers. Hence it is identified as an "exact or discrete design". Usually, for the implementation, every design has to be discrete since whole units to be assigned for each and every design point.

The term "approximate or continuous designs" comes at this point. It is a mathematically convenient way over exact designs. Generally, an approximate design is defined in terms of proportions and can be denoted in the following way:

$$\xi = \left\{ \begin{array}{cccc} d_1 & d_2 & \cdots & d_m \\ w_1 & w_2 & \cdots & w_m \end{array} \right\}.$$

Here, $w_j = n_j/N$, $j = 1, 2, ..., m$, $\sum_{j=1}^{m} w_j = 1$ and $0 \leq w_j \leq 1$ for all j.

An approximate design can be converted to an exact design before implementation. Typically, an exact design formation is done by multiplying each and every weight of an approximate design by the total sample size N of the particular sample and rounding up to the nearest integer, in a way that the summation of those rounded values will be equal to the total sample size N. Kiefer (1985) mentioned these three significant reasons below to highlight the importance of working with approximate designs instead of exact designs.

- It is very difficult to find optimal exact designs and they mostly depend on the sample size N. This results in different optimal designs for different sample N. As for the case of approximate designs, they are independent of the sample size N.
- Since there is a rounding involve in implementing optimal approximate designs, it will have a similarity and closeness with an optimal exact design. It can be shown that the difference between those two designs will vanish when increasing the sample size N.
- Implementation of exact designs requires complex mathematical theories and most of the times it is even difficult to find an exact design

for simple problems. On the other hand, optimal approximate designs can be found analytically or with the use of iterations in computer algorithms.

So, in summary, an approximate optimal design will be preferred over an exact optimal design.

1.3 Discretization of the Design Space

Consider the alternative form of model 1.2 given below

$$E(y|\underline{v}, \underline{\theta}, \sigma) = \underline{v}^T \underline{\theta} \tag{1.7}$$

where $\underline{v} = (f_1(\underline{x}), \cdots, f_k(\underline{x}))^T$, $\underline{v} \in \mathcal{V}$, $\mathcal{V} = \{\underline{v} \in \mathbf{R}^k : \underline{v} = (f_1(\underline{x}), \cdots, f_k(\underline{x}))^T$, $\underline{x} \in \mathcal{X}\}$.

Generally, the original design space is continuous. Hence, we need to discretize the design space which will be easy to work with. After discretization, the design space is identified as an "induced design space" and it can be assumed to be discrete. Here $\mathcal{V}$ is an induced design space. Suppose that $\mathcal{V}$ is made up of J distinct vectors $\underline{v_1}, \underline{v_2}, ..., \underline{v_J}$. Let us suppose that, n_j observations out of n observations have been taken at $\underline{v_j}$ such that $\sum_{j=1}^{J} n_j = n$. Then, the matrix $(X^T X)$ can be expressed as

$$X^T X = M(\underline{n}),\ \underline{n} = (n_1, n_2, \cdots, n_J)^T \tag{1.8}$$

Here,

$$M(\underline{n}) = \sum_{j=1}^{J} n_j \underline{v}_j \underline{v}_j^T = VNV^T \tag{1.9}$$

where $V = [\underline{v_1}, \underline{v_2}, ..., \underline{v_J}]$ and $N = diag(n_1, n_2, ..., n_J)$.

Our goal in here is to make $M(\underline{n})$ large by choosing $\underline{n}$ efficiently. This design can be identified as an exact design since n_j' s are expressed in integers. As mentioned earlier, it will be flexible and simpler to solve, if we can convert it to an approximate design. Let us look at the following conversion

$$M(\underline{n}) = nM(p). \tag{1.10}$$

Here,

$$M(p) = \sum_{j=1}^{J} p_j \underline{v}_j \underline{v}_j^T = VPV^T$$

where $P = diag(p_1, p_2, ..., p_J)$ and $p_j = n_j/n$, the proportions of observations taken at $\underline{v_j}$ such that $p_j \geq 0$, $\sum_{j=1}^{J} p_j = 1$.

We have to make sure to choose p in a way that $M(p)$ becomes large. Additionally, this can be converted to the nearest exact design using np^* and by doing that it can be preferred to the original exact design.

Furthermore, using the definition of expectation, $M(p) = E_p[\underline{v}\underline{v}^T]$, where p_j is the probability corresponding to the vertex such that $P(\underline{v} = \underline{v_j}) = p_j$.

The matrix $M(p)$ is identified as the information matrix.

1.3.1 Specific Properties of the Information Matrix

As we discussed, the information matrix $M(p)$ can be written as

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

- The first property is that $M(p)$ is symmetric. This follows from its definition.

- $M(p)$ is non-negative definite. The proof is given below.

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

- The inverse of the dispersion matrix $[D(\hat{\theta})]$ is the information matrix. So these two matrices have the reciprocity property. This implies, minimizing the variance is the same as maximizing the information.

1.4 Important Definitions in Optimal Design

1.4.1 Design Measure

Any design measure can be denoted in the following notation

$$\xi = \left\{ \begin{array}{cccc} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_J \\ p_1 & p_2 & \cdots & p_J \end{array} \right\}.$$

Here, the first line denotes the locations of the design points x_j ; $x_j \in \mathcal{X}$ whereas the second line gives the associated design weights p_j such that $\sum_{j=1}^{J} p_j = 1$ and $0 \leq p_j \leq 1$ for all j.

1.4.2 Support of a Design Measure

Suppose we want to denote support of a design measure in the design space $\mathcal{V}$. What we actually denote here are the vertices $\underline{v}_j$ with non-zero weights under p. The support is given by

$$Supp(\xi) \quad = \quad \{\underline{v}_j \in \mathcal{V} : \, p_j > 0, \, j = 1, 2, \cdots, J\}.$$

1.4.3 Standardized Variance of the Predicted Response

Another important term that we will need in optimal designs is, the standardized variance of the predicted response y at $\underline{x}$ and which is defined as follows:

$$d(\underline{x}, p) \quad = \quad \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x}) \qquad (1.11)$$

where $M(p)$ is the information matrix. We will discuss about it in Chapter 3.

1.5 Design Criteria

We intend to obtain the best inference on the unknown parameters $\underline{\theta}$ and to make the information matrix $M(p)$ as large as possible. At this point, the term "Criterion Function" comes in. It can be denoted as, $\phi(p) = \psi\{M(p)\}$ where $\phi(p)$ is the criterion function. Here, ϕ is known as ϕ-optimality whereas the design maximizing the $\phi(p)$ is identified as a ϕ-optimal design.

In this chapter, different types of optimality criteria which use variance-covariance matrix in a unique way (Box, 1982) will be discussed. Atkinson and Donev (1992) have reviewed about the design optimality criteria such as D, A, G and E which will be discussed next.

1.5.1 D-optimality Criterion

Our goal is to obtain more accurate estimators by minimizing the volume of a confidence ellipsoid. This criterion minimizes the volume of the ellipsoid by minimizing the product of the squared lengths of the axes of the ellipsoid. D-optimality criterion is defined as the determinant of variance-covariance matrix $cov(\hat{\theta})$. By reciprocity property of the variance-covariance matrix and the information matrix, this will be equivalent to maximization of the determinant of the information matrix. Hence, it can be written as

$$\phi_D(p) = \psi_D\{M(p)\} = log\,det\{M(p)\} = -log\,det\{M^{-1}(p)\}. \qquad (1.12)$$

Kiefer and Wolfowitz (1960) have shown a relationship between the D-optimal design and the standardized variance of the predicted response

$$inf\ sup\ d(x,p)\ = sup\ d(x,p^*). \qquad (1.13)$$

Here, $sup\ d(x,p^*) = k$, where k is the number of parameters and $d(x,p)$ is the standardized variance of the predicted response.

Furthermore, this criterion can be expressed in terms of the eigenvalues of the information matrix ($M(p)$) as well. The eigenvalues of $M^{-1}(p)$ are proportional to the lengths of the axes of the confidence ellipsoid that we mentioned earlier. So, if the eigenvalues of $M(p)$ are $\lambda_1, \lambda_2, ..., \lambda_k$ the D-optimal design will minimize the product of the eigenvalues of $M^{-1}(p)$: $M^{-1}(p) : \prod_{i=1}^{k} 1/\lambda_i$. The studies related to this theory have been conducted by Mandal, Torsney and Carriere (2005), Atkinson and Donev (1992), Atkinson, Donev and Tobias (2007).

Advantages of D-optimality Criterion

- This is a concave function of the positive definite symmetric matrices and is differentiable when it is finite. Suppose, d_j is the first partial derivative of this criterion and it is given by

$$d_j = \frac{\partial \phi_D}{\partial p_j} = \underline{v}_j^T M^{-1}(p)\underline{v}_j. \tag{1.14}$$

- D-optimal design is invariant under the linear transformation of the scale of the independent variable.

Drawbacks of D-optimality Criterion

- Minimizing determinant of variance-covariance matrix may lead to an elongation in the direction of one axis of the ellipsoid. Therefore only one of the parameters is estimated efficiently while the others are not.

- May be inefficient in estimating certain linear functions of the parameters.

1.5.2 D_A-optimality Criterion

This criterion can be used when our interest is on some linear combinations of the parameters of the linear model. Let us say we are interested in s linear combinations where the elements of the vector $\underline{\alpha} = A\theta$ such that A is a $s \times k$ matrix of rank $s \leq k$. Then the variance-covariance matrix of $A\hat{\underline{\theta}}$ is

$AM^{-1}(p)A^T$. The criterion function is defined as

$$\phi_{D_A}(p) = \psi_{D_A}\{M(p)\} = -log\, det\{AM^{-1}(p)A^T\}. \tag{1.15}$$

The criterion function was named as D_A-optimality by Sibson (1974). Furthermore, the partial derivatives of this criterion are given by

$$\frac{\partial \phi_{D_A}}{\partial p_j} = \underline{v}_j^T M^{-1}(p)A^T[AM^{-1}(p)A^T]^{-1}AM^{-1}(p)\underline{v}_j. \tag{1.16}$$

1.5.3 D_s-optimality Criterion

D_s-optimality criterion can be used when we are interested in estimating a subset of s parameters accurately [Karlin and Studden (1966), Atwood (1969)]. In D_A-optimality, matrix A is composed of $s \times s$ identity matrix and $s \times (k-s)$ zero matrix. In this criterion, interest is on the first s parameters $\theta_1, \theta_2, ..., \theta_s$. Rhode (1965) partitioned $M(p)$ matrix as follows.

Suppose our matrix $A = [I_s : O]$ and $M(p)$ can be partitioned as below

$$M(p) = \begin{bmatrix} M_{11}^{s\times s} & M_{12}^{s\times(k-s)} \\ M_{12}^T & M_{22}^{(k-s)\times(k-s)} \end{bmatrix}.$$

We need to choose the design that maximizes the following criterion

$$\phi_{D_s}(p) = log\, det\{M_{11} - M_{12}M_{22}^{-1}M_{12}^T\}. \tag{1.17}$$

1.5.4 A-optimality Criterion (Average Criterion)

This criterion minimizes the sum of squared lengths of the axes by minimizing the trace of variance-covariance matrix $cov(\hat{\theta})$. It can be expressed in the following notation.

$$\phi_A(p) = \psi_A\{M(p)\} = -Trace\{M^{-1}(p)\} \tag{1.18}$$

One significant feature of this criterion is that it neglects the correlations between the estimates. The partial derivatives of this criterion are given by

$$\frac{\partial \phi_A}{\partial p_j} = \underline{v}_j^T M^{-2}(p) \underline{v}_j. \tag{1.19}$$

Advantages of A-optimality Criterion

- Easy to evaluate since it takes only the diagonal entries of the variance-covariance matrix $M^{-1}(p)$.

Drawbacks of A-optimality Criterion

- This is variant under linear transformation of the scale of the independent variables.

1.5.5 G-optimality Criterion (Global Criterion)

This is useful when the researcher is interested in predicting the response variable Y efficiently over the design space. It has been proved that to obtain an accurate prediction for Y, we need to have a design that gives the smallest possible standardized variance of the predicted response. So, this is defined as a design that minimizes the maximum standardized variance of the predicted response over the design space. It is given by

$$\phi_G(p) = \psi_G\{M(p)\} = -Max_{\underline{v}\in\mathcal{V}}\, \underline{v}^T M^{-1}(p)\underline{v}. \tag{1.20}$$

Kiefer and Wolfowitz (1960) proved a similarity between this criterion and D-optimality criterion. The partial derivative of this G-criterion is given below

$$\frac{\partial\phi_G}{\partial p_j} = [\underline{v}_j^T M^{-1}(p)\underline{v}_j]^2. \tag{1.21}$$

Advantages of G-optimality Criterion

- Efficient in predicting the response variable over the design space by minimizing the maximum standardized variance of the predicted response $d(\underline{x}, p)$. Hence, $d(\underline{x}, p^*) \leq k$ where p^* is a G-optimal design and k is the number of parameters in the model. Moreover, this inequality can be used to check whether a given design is D-optimal or not.

- This is also invariant under linear transformation of the scale of the independent variables as in D-optimal designs.

1.5.6 E-optimality Criterion

This criterion minimizes the squared length of the largest or the extreme axis of the confidence ellipsoid. It can be shown algebraically that, this is same as minimizing the maximum eigenvalues of the variance-covariance matrix $cov(\hat{\theta})$. So, this will maximize the following criterion function

$$\phi_E(p) = \psi_E\{M(p)\} = -\lambda_{max}\{M^{-1}(p)\} \tag{1.22}$$

where $\lambda_{max}\{M^{-1}(p)\}$ is the largest eigenvalue of $M^{-1}(p)$.

Generally, this is more suitable for a design where all the factors are qualitative like a block design.

1.5.7 Linear Optimality Criterion

Suppose we have L, a k x k matrix of coefficients. Consider the criterion

$$\phi_L(p) = \psi_L\{M(p)\} = -tr\{M^{-1}(p)L\}. \tag{1.23}$$

The maximization of the criterion function given above leads to a linear-optimum (L-optimum) design. Here what is meant by linear is, linearity in the elements of the covariance matrix $M^{-1}(p)$. If $L = A^T A$, then the criterion function above can be expressed as

$$\phi_L(p) = -tr\{M^{-1}(p)L\} = -tr\{M^{-1}(p)A^T A\} = -tr\{AM^{-1}(p)A^T\} \tag{1.24}$$

where A is a $s \times k$ matrix of rank $s \leq k$. The partial derivatives of linear optimality can be written as

$$\frac{\partial \phi_L}{\partial p_j} = \underline{v}_j^T M^{-1}(p) A^T A M^{-1}(p) \underline{v}_j. \tag{1.25}$$

Next, we will focus on a special case of linear optimality which is identified as c-optimality.

1.5.8 c-optimality Criterion

This is the case where $A = \underline{c}^T$ in linear optimality where $\underline{c}$ is a $k \times 1$ vector. As mentioned in literature, this is also a standard criterion which is identified as c-optimality criterion. This maximizes the criterion function given below

$$\phi_c(p) = \psi_E\{M(p)\} = -\underline{c}^T M^{-1}(p)\underline{c}. \tag{1.26}$$

Hence our goal is to estimate the linear parametric function $\underline{c}^T\underline{\theta}$ with minimum variance. The partial derivative for this criterion can be expressed as

$$\frac{\partial \phi_c}{\partial p_j} = [\underline{c}^T M^{-1}(p)\underline{v}_j]^2. \tag{1.27}$$

Chapter 2

Optimality Conditions and a Class of Algorithms

So far, we have learnt about some fundamental theory of optimal design and different optimality criteria. In this chapter, we will determine the optimality conditions and a class of algorithms.

Suppose our general problem is to maximize a criterion function $\phi(p)$ subject to the constraints $p_j \geq 0$ and $\sum_{j=1}^{J} p_j = 1$ where $j = 1, 2, ..., J$. Our goal here is to find an optimizing distribution (p^*) by considering this general problem. This means, we are going to find an optimal design according to a particular criterion function.

2.1 Determining Optimality Conditions

When constructing an optimal design, first thing to be done is determining the optimality conditions. We will determine optimality conditions using some directional derivatives. We will consider the following two kinds of directional derivatives.

2.1.1 Directional Derivatives

Suppose that we want to find the directional derivatives of a particular criterion function at p in the direction of q (Whittle, 1973).

$$F_{\phi}\{p,q\} \quad = \quad lim_{\varepsilon \downarrow 0} \frac{\phi\{(1-\varepsilon)p+\varepsilon q\}-\phi(p)}{\varepsilon}. \tag{2.1}$$

2.1.2 G$\hat{a}$teaux Derivative

Kiefer (1974), in one of his design, used this concept. This is another type of directional derivative of a particular criterion function at p in the direction of m which can be denoted by $G_{\phi}\{p,m\}$ and expressed as

$$G_{\phi}\{p,m\} \quad = \quad lim_{\varepsilon \downarrow 0} \frac{\phi\{p+\varepsilon m\}-\phi(p)}{\varepsilon}. \tag{2.2}$$

Some significant properties of the directional derivatives are discussed below.

- Let $p,q \in S$ given that S is a convex set. Then $\{(1-\epsilon)p+\epsilon q\}$ is also in S. This is very useful if we want to find $F_{\phi}\{p,q\}$ where $p,q \in S$.

But $G_\phi\{p,q\}$ does not have this property.

- Given that $\phi(.)$ is concave, the following inequality can be presented. $F_\phi\{p,q\} \geq \phi(q) - \phi(p)$.

If $\phi(.)$ is differentiable, then $F_\phi\{p,\, q\}$ can be expressed as

$$F_\phi(p,q) \;=\; (q-p)^T \frac{\partial \phi}{\partial p} \;=\; \sum_{j=1}^{J} (q_j - p_j)\, d_j, \text{ where } d_j = \partial\phi/\partial p_j,\ j = 1, 2, \ldots, J.$$

If $q = e_j$, where e_j is the j^{th} unit vector in $\mathbb{R}^J$, we can simplify

$$F_j = F_\phi(p, e_j) = \frac{\partial \phi}{\partial p_j} - \sum_{i=1}^{J} p_i \frac{\partial \phi}{\partial p_i}$$

Here F_j's are the vertex directional derivatives of $\phi(.)$.

2.1.3 General Equivalence Theorem

Suppose $\phi(p)$ is differentiable at an optimizing distribution p^*, and we are going to find the first-order conditions for $\phi(p^*)$ to be a local maximum of $\phi(p)$ in the feasible region in the general problem. The conditions are given as follows:

$$F_j^* = F_\phi\{p^*, e_j\} \begin{cases} = 0 & \text{if } p_j^* > 0 \\ \leq 0 & \text{if } p_j^* = 0 \end{cases} \tag{2.3}$$

When $\phi(p)$ is a concave function, the first-order stationarity condition mentioned above is necessary and sufficient for optimality. This is identified as "The General Equivalence Theorem in Optimal Design" by Kiefer in 1974.

2.2 A Class of Algorithms to find an Optimal Design for a Given Model Defined on a Given Design Space

Research on the optimal design construction has a long history. Theoretical method of constructing an optimal design is with the use of mathematical derivation based on model assumptions. However, this theoretical approach can have some limitations in terms of usability. In theoretical approach, the optimal design is constructed using a particular set of assumptions. Hence the results are only appropriate to a specific setting of conditions. This means, analytical results of a particular setting have a limited use in practice when the conditions or the assumptions are changed.

Algorithms of constructing optimal designs have been introduced to overcome these problems associated with theoretical approach. We provide a review on some such algorithms. Wynn (1970) in one of his article, has mentioned about this sequential algorithm of constructing optimal approximate designs. Generally, the algorithm consists of a starting point along with a stopping rule. When running the algorithm, it will proceed further by adding a point to the current design until the stopping rule conditions are met.

In exchange algorithms, the current design is iteratively modified by deleting existing design points and adding new points from the design space. Since this poses some issues related to local maximizers, the algorithm is run multiple times using different starting designs. Fedorov's exchange algorithm is one of the earliest and significant algorithm that falls in this category. Yang et al (2012) proposed a method to obtain optimal designs for generalized linear models by modifying this algorithm. Furthermore, Wynn's (1972) algorithm, modified Fedorov's algorithm by Cook and Nachtsheim (1980) and k-exchange algorithms by Johnson and Nachtsheim (1983) also can be included in this category. These algorithms differ from each other in the basic step which is the selection of exchanging the design point in current design.

Another algorithm is Fedorov-Wynn type algorithm (Mandal and Wong (2015)) which has become more popular as one of the earliest algorithm converging to an optimal design even with a large number of runs. The convergence of this algorithm has proved only for the linear models.

2.2.1 Multiplicative Algorithm

Another significant class of algorithms of constructing optimal designs is the class of multiplicative algorithm which has become more popular in the last decade. This algorithm was initially proposed by Torsney (1977). Later Mandal and Torsney (2006) modified this algorithm to increase the efficiency of the design construction. Torsney and Martin (2009) used this algorithm to find optimal designs when responses are correlated. Dette (2008) made some modifications in this algorithm in order to take larger steps at each iteration even maintaining monotonic convergence. My current research work is based on the construction of optimal designs using the multiplicative algorithms.

We will now discuss about the multiplicative algorithm and its properties. There are two basic constraints of optimal weights that has to be satisfied by an algorithm. One is the weights $p_1, p_2, ..., p_J$ must be non-negative and the next one is their summation should be equal to 1. The form of multiplicative algorithm is given below:

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(x_j^{(r)})}{\sum_{i=1}^{J} p_i^{(r)} f(x_i^{(r)})}. \tag{2.4}$$

Here, $x_j^{(r)} = d_j^{(r)}$ or $F_j^{(r)}$, the partial derivatives or the directional derivatives of the criterion function. It also has to be noted that the function $f(x)$ may depend on a free positive parameter δ. The function is positive and strictly increasing in x.

The algorithm has the following properties:

- $p^{(r)}$ is always feasible.
- $F_\phi\{p^{(r)}, p^{(r+1)}\} \geq 0$ where equality occurs when the d_j's corresponding to nonzero p_j's are equal.
- $supp(p^{(r+1)}) \subseteq supp(p^{(r)})$ will be satisfied under the above mentioned iteration and some weights will converge to zero.
- If the derivatives corresponding to nonzero $p_j^{(r)}$ are all equal, an iterate

$p^{(r)}$ is identified as a fixed point of the iteration.

As $f(x)$ depends on a parameter δ, let us denote it by $f(x, \delta)$. This function plays a major role in the algorithm of constructing optimal designs. By selecting a symmetric function which is centered around zero, we can improve the convergence rate of the algorithm. The function $f(x)$ for $x = d$ is not symmetric nor centered around zero. But if we replace partial derivatives d of the criterion function by directional derivatives F in the function, our goal can be achieved.

Consider the normal cumulative distribution function $f(x, \delta) = \Phi(\delta x)$ and the logistic function $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$. These two functions satisfy the conditions of the above mentioned algorithm and change quickly at $x = F = 0$. But for the function $f(x, \delta) = e^{\delta x}$, it does not make any difference in the performance if we replace $x = d$ by $x = F$.

Chapter 3

Construction of D-optimal Designs using Traditional Approach

In this chapter we are going to focus on the construction of D-optimal designs using traditional approach. First we discuss some analytic solution for this approach. Next we will learn about the application of a class of algorithms in D-optimal design construction.

3.1 Analytic Solutions for the Construction of D-optimal Designs

Let us consider a polynomial regression model of order $k-1$ for a one variable, which is given by

$$E(y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + ... + \theta_{k-1} x^{k-1} = \underline{v}_x^T \underline{\theta}. \tag{3.1}$$

Here, $\underline{v}_x = (1, x, x^2, ..., x^{k-1})^T$, $x \in [-1, 1]$, $\underline{\theta} = (\theta_0, \theta_1, \cdots, \theta_{k-1})^T$ and the induced design space is given by,

$\underline{v}_x \in \mathcal{V} = \{\underline{v}_x : \underline{v}_x = (1, x, x^2, \cdots, x^{k-1})^T,\ -1 \leq x \leq 1\}.$

This design space is a standardized continuous design space. Fedorov (1972) stated about the discrete D-optimal designs. He mentioned that they are unique and consist of k number of minimal support points which can be found using the k roots of the polynomials, $(1 - x^2)P'_{k-1}(x)$. In here $P_k(x)$ is the k^{th} Legendre polynomial:

$$P_k(x) = \sum_{n=0}^{N} \left[\frac{(-1)^n (2k-2n)! x^{k-2n}}{2^k n!(k-n)!(k-2n)!} \right] \tag{3.2}$$

where $N = k/2$ when k is even and $N = (k-1)/2$ when k is odd. Since there are k support points in a minimal support design, there has to be $(1/k)$ weight assigned to each of these design points.

Let us consider the analytic solutions for cubic regression model where $k = 3$ and quartic regression model where $k = 4$.

3.1.1 Analytic Solution for the Cubic Regression Model

The model is given by

$$E(y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3.$$

The design space $\mathcal{X}$ is $[-1, 1]$. We will find the D-optimal design for the cubic regression model.

- Here, the total number of parameters is $k = 4$. This means we have to consider $(1 - x^2)P'_{k-1}(x) = (1 - x^2)P'_3(x)$.

- First we need to find $P_3(x)$.

- By using the conditions of Legendre polynomial $N = (k - 1)/2 = (3 - 1)/2 = 1$; $P_k(x) = P_3(x)$.

- By substituting $k = 3$ and $N = 1$ in Legendre polynomial equation 3.1, $P_3(x) = \sum_{n=0}^{1} \left[\frac{(-1)^n (6-2n)! x^{3-2n}}{2^3 n!(3-n)!(3-2n)!}\right]$.

- By solving the above,
  $P_3(x) = (5x^3 - 3x)/2$.

- Next we have to calculate the first partial derivative of the above solution to obtain $P'_3(x)$
  $P'_3(x) = (15x^2 - 3)/2$.

- The support points are found using $(1 - x^2)P'_{k-1}(x) = 0$. Therefore,
  $(1 - x^2)P'_3(x) = (1 - x^2)(15x^2 - 3)/2 = 0$

- The solutions for x are $\pm 1, \pm 0.447$
- Therefore, the D-optimal design for the cubic regression model is

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.447 & -0.447 & 1.00 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{array} \right\}.$$

3.1.2 Analytic Solution for the Quartic Regression Model

The model is given by

$$E(Y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4.$$

The design space $\mathcal{X}$ is $[-1, 1]$. We need to find the D-optimal design for the quartic regression model. That means a design space of optimal design points with their corresponding probabilities.

- Here, the total number of parameters is $k = 5$. So, we have to consider $(1 - x^2)P'_{k-1}(x) = (1 - x^2)P'_4(x)$.
- First we need to find $P_4(x)$.
- By using the conditions of Legendre polynomial $N = k/2 = 4/2 = 2$; Here $P_k(x) = P_4(x)$.
- By substituting $k = 4$ and $N = 2$ in Legendre polynomial equation 3.1, $P_4(x) = \sum_{n=0}^{2} \left[\frac{(-1)^n (8-2n)! x^{4-2n}}{2^4 n!(4-n)!(4-2n)!} \right]$.

- By solving the above,

  $P_4(x) = (35x^4 - 30x^2 + 3)/8.$

- Next we have to calculate the first partial derivative of the above solution to obtain $P_4'(x)$

  $P_4'(x) = 5x(7x^2 - 3)/2.$

- The support points are found using $(1 - x^2)P_{k-1}'(x) = 0$. Therefore,

  $(1 - x^2)P_4'(x) = (1 - x^2)5x(7x^2 - 3)/2 = 0.$

- The solutions for x are $0, \pm 1, \pm 0.655$.

- Therefore, the D-optimal design for the quartic regression model is,

$$P^* = \left\{ \begin{array}{ccccc} -1.00 & -0.655 & -0.00 & 0.655 & 1.000 \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \end{array} \right\}.$$

3.2 Construction of D-optimal Designs using a Class of Algorithms

In Chapter 2, we have discussed about the class of algorithms including multiplicative algorithm used to find optimal design for a given model defined on a given design space. In this section we are going to focus specifically on the construction of D-optimal designs using a class of multiplicative algorithms based on a function of the derivatives of D-optimality criterion.

3.2.1 Multiplicative Algorithm for D-optimal Design Construction

A general introduction on the class of algorithms was given in Chapter 2. We have learnt about the multiplicative algorithm and its properties. As for the function used in the algorithm, we know that it depends on a free positive parameter δ which may differ according to the regression model being considered. The algorithm is given here in the following:

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(x_j^{(r)}, \delta)}{\sum_{i=1}^{J} p_i^{(r)} f(x_i^{(r)}, \delta)} \tag{3.3}$$

where, $x_j^{(r)} = d_j^{(r)}$ or $F_j^{(r)}$, the partial derivatives and the directional derivatives of the criterion function. Furthermore,

- $d_j^{(r)} = \left.\frac{\partial \phi}{\partial p_j}\right|_{p=p^{(r)}}$, the partial derivatives at r^{th} iteration.
- $F_j^{(r)} = d_j^{(r)} - \sum_{i=1}^{J} p_i^{(r)} d_i^{(r)}$, the vertex directional derivatives at $p = p^{(r)}$.

All the properties of this algorithm are similar to the ones we discussed in Section 2.2.1. The performance of the algorithm for each model will be tested by taking $f(x, \delta)$ where $x = d$. Later, the results are reported by replacing the partial derivatives in the function with directional derivatives. This will improve the convergence of the algorithms.

3.3 Construction of Optimal Designs for One Variable Regression Models using Different Functions

The optimal designs have been constructed for quadratic regression model, cubic regression model and quartic regression model. The four functions being considered when constructing these models are x^δ, $e^{\delta x}$, $e^{\delta x}/1+e^{\delta x}$ and normal cumulative distribution function. The results are reported in the tables including the best choices for δ with bold font.

Moreover, the results have been interpreted graphically using the plots of weights versus design points and variance function versus design points. The plots have been drawn running the algorithm for small number of iterations and at the optimum. As for the plot of variance function versus design points, the standardized variance of the predicted responses at a given $\underline{x}$ were recorded. This is defined as

$$d(\underline{x}, p) = \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x}).$$

The proof of this equation is as follows.

Consider equation (1.6) defined in Chapter 1. So we have, $\hat{y}(\underline{x}) = \underline{f}^T(\underline{x})\hat{\underline{\theta}}$. We also have $D(\hat{\underline{\theta}}) = \sigma^2(X^TX)^{-1}$. Furthermore we know that $M(\underline{n}) = X^TX = nM(p)$. Then, the standardized variance of the predicted response on y at $\underline{x}$ can be expressed as

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

After we scale the variance σ^2 and the number of trials n, the variance function is

$$d(\underline{x},p) = D(\hat{y})/(\sigma^2/n).$$

3.3.1 Quadratic Regression Model

Let us consider the quadratic regression model which is given by

$$E(Y|x) = \theta_0 + \theta_1 x + \theta_2 x^2.$$

The initial design space consists of 101 points equally spaced at intervals of 0.01 between -1.00 and 1.00. Total number of iterations is 100 000.

We report the results in the following tables, first by taking the argument of $f(x,\delta)$ as $x = d$. We record for $t = 1,\ldots,6$ the number of iterations needed to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$, where F_j's are the vertex directional

Table 3.1: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for quadratic regression model

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.600	34	245	2494	21763	49740	75549
1.000	21	148	1497	13058	29844	45330
1.250	17	118	1197	10447	23875	36264
1.375	16	108	1089	9497	21705	32968
1.400	15	106	1069	9327	21318	32379
1.900	**15**	**79**	**788**	**6873**	**15708**	**23858**
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.200	31	244	2491	21757	49736	75547
0.375	17	131	1329	11604	26526	40292
0.425	15	115	1172	10239	23405	35552
0.475	13	103	1049	9161	20941	31810
0.500	**12**	**98**	**996**	**8703**	**19894**	**30219**
0.575	15	84	866	7567	17299	26277
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.300	77	564	5748	50189	114716	174245
0.375	75	532	5425	47356	108237	164403
0.400	75	529	5385	47007	107439	163190
0.450	**76**	**528**	**5383**	**46983**	**107380**	**163099**
0.500	79	537	5468	47720	109063	165655
0.575	84	563	5737	50059	114404	173764
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.150	81	609	6208	54210	123912	188215
0.250	71	503	5122	44710	102190	155218
0.255	**71**	**501**	**5108**	**44592**	**101920**	**154807**
0.290	73	499	5084	44373	101417	154042
0.300	73	500	5097	44490	101683	154445
0.350	78	519	5289	46155	105484	160217

derivatives.

According to the results obtained in Table 3.1, the best choices of δ for $x = d$ for the four different functions are

- 1.900 for $f(x, \delta) = x^{\delta}$
- 0.500 for $f(x, \delta) = e^{\delta x}$
- 0.450 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$
- 0.255 for $f(x, \delta) = \Phi(\delta x)$.

In addition to that it can be seen, when moving away from this δ value, the number of iterations needed have been increased. Moreover, for the logistic cumulative density function (CDF) $e^{\delta x}/1 + e^{\delta x}$ and normal CDF the total number of iterations were taken as 200 000 unlike for the other two functions.

The D-optimal design obtained for the quadratic regression model is given below. The optimal design corresponding to a particular model in a predetermined design space will be the same, even though the function in the algorithm is different.

$$P^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 0.333 & 0.333 & 0.333 \end{array} \right\}$$

Next we are going to check the convergence of the algorithm by replacing $x = d$ by $x = F$. This means we are replacing partial derivatives of the criterion function $f(x, \delta)$ with the directional derivatives. It is noticeable that, the results and the convergence rate will not be changed for the function $f(x, \delta) = e^{\delta x}$. It can be proved theoretically using functional form of the algorithm. In addition, we cannot use $f(x, \delta) = x^{\delta}$ for $x = F$ because the directional derivatives could take negative values.

The number of iterations used to reach $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ by replacing partial derivatives with directional derivatives for quadratic regression model using different δ values are given in Table 3.2.

Table 3.2: Number of iterations need to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for quadratic regression model

$f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.300	44	326	3323	29013	66317	100731
0.375	36	261	2659	23211	53054	80585
0.400	34	245	2493	21761	49738	75549
0.450	31	218	2216	19343	44212	67155
0.575	25	171	1735	15139	34601	52556
0.575	**25**	**171**	**1735**	**15139**	**34601**	**52556**
$f(x, \delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.150	54	409	4164	36362	83115	126247
0.250	34	246	2499	21818	49870	75749
0.255	34	241	2450	21391	48892	74264
0.290	30	212	2155	18810	42992	65301
0.300	30	205	2083	18183	41559	63124
0.350	**26**	**176**	**1786**	**15586**	**35622**	**54107**

3.3.2 Cubic Regression Model

Consider the application of D-optimality to the cubic regression model given below

$$E(Y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3.$$

A discretized design space of 101 points equally spaced at intervals of 0.01 between -1.00 and 1.00 is considered in here.

For the logistic CDF $e^{\delta x}/1+e^{\delta x}$, the design space being considered is 0.00 and 1.00, which was approximated by a grid of 101 points equally spaced at intervals of 0.01.

The best choices of δ for the four different functions are obtained as

- 1.750 for $f(x,\delta) = x^{\delta}$
- 0.375 for $f(x,\delta) = e^{\delta x}$
- 0.325 for $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$
- 0.200 for $f(x,\delta) = \Phi(\delta x)$

The D-optimal design obtained using the functions x^{δ}, $e^{\delta x}$ and $\Phi(\delta x)$ will be the same since their design space is [-1,1]. It is as follows:

$$P^* = \left\{ \begin{array}{cccccc} -1.00 & -0.45 & -0.44 & 0.44 & 0.45 & 1.00 \\ 0.250 & 0.231 & 0.019 & 0.019 & 0.231 & 0.250 \end{array} \right\}.$$

Table 3.3: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for cubic regression model

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.900	23	217	2028	10867	30890	71850
1.000	21	196	1825	9781	27801	64665
1.250	17	157	1460	7825	22241	51732
1.500	14	131	1217	6521	18534	43110
1.750	**13**	**112**	**1043**	**5590**	**15886**	**36951**
1.900	19	104	961	5148	14632	34034
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.180	29	270	2532	13580	38608	89809
0.200	26	243	2279	12222	34747	80828
0.250	21	195	1823	9778	27798	64662
0.300	17	162	1519	8148	23165	53884
0.325	16	149	1402	7521	21382	49739
0.375	**13**	**128**	**1213**	**6516**	**18529**	**43105**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.250	74	688	4684	15984	32925	51445
0.300	71	666	4536	15476	31879	49811
0.325	**70**	**665**	**4526**	**15441**	**31806**	**49696**
0.350	70	668	4550	15523	31975	49960
0.375	71	676	4605	15711	32363	50564
0.500	79	776	5287	18034	37146	58038
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.190	70	666	6235	33427	95021	221024
0.199	70	663	6206	33271	94576	219988
0.200	**70**	**663**	**6204**	**33259**	**94543**	**219912**
0.230	72	666	6233	33412	94975	220914
0.240	83	724	6778	36325	103249	240154
0.259	76	686	6418	34398	97774	227423

The D-optimal design obtained using the logistic probability density function $e^{\delta x}/1 + e^{\delta x}$ in the design space [0,1] is given below.

$$P^* = \left\{ \begin{array}{cccccc} 0.00 & 0.27 & 0.28 & 0.72 & 0.73 & 1.00 \\ 0.250 & 0.058 & 0.192 & 0.192 & 0.058 & 0.250 \end{array} \right\}$$

It is noticeable that in both of the above situations, the support points consist of 4 clusters of points. These clusters can be a pair of neighboring points or a single point. It should be kept in mind; the design space obtained here is the discretized design space. Therefore, the desired solution for the continuous design space has to be a 4-point design. By taking the convex combinations of the particular cluster members, the following solution can be obtained.

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.447 & -0.447 & 1.00 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{array} \right\}$$

$$P^* = \left\{ \begin{array}{cccc} 0.00 & 0.275 & 0.725 & 1.00 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{array} \right\}$$

Here, the first solution can be obtained under the design space [-1, 1] while the second solution can be obtained under the design space [0, 1].

Next we will see the convergence rate by replacing partial derivatives with directional derivatives in the function (Table 3.4). Here also, by the same argument as before, when using $x = F$ we do not use the functions $f(x, \delta) = x^{\delta}$ and $f(x, \delta) = e^{\delta x}$.

Table 3.4: Number of iterations need to achieve $\max_{1 \leq j \leq J}\{F_j\} \leq 10^{-t}$ with $x = F$ for cubic regression model

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.250	40	370	2519	8597	17709	27670
0.300	34	309	2100	7164	14758	23058
0.325	31	285	1938	6613	13622	21285
0.350	29	265	1800	6141	12649	19764
0.375	27	247	1680	5732	11806	18446
0.500	**21**	**186**	**1261**	**4299**	**8855**	**13835**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.190	34	322	3008	16127	45844	106637
0.199	32	307	2872	15398	43771	101814
0.200	32	306	2858	15321	43552	101305
0.230	28	266	2485	13323	37872	88092
0.240	27	255	2382	12768	36294	84421
0.259	**26**	**236**	**2207**	**11832**	**33632**	**78228**

3.3.3 Quartic Regression Model

We now apply the D-optimality to the quartic regression model:

$$E(Y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4.$$

The design space was considered as a grid of 101 points, i.e., equally spaced at intervals of 0.01 between -1.00 and 1.00.

For the logistic cumulative density function and normal distribution, approximately 150 000 iterations have been used unlike for the other two functions. The results obtained for the total number of iterations when using

Table 3.5: Number of iterations need to achieve $\max_{1\le j\le J}\{F_j\} \le 10^{-t}$ with $x = d$ for quartic regression model

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.300	83	815	8308	38630	66634	93923
0.500	50	490	4985	23178	39981	56354
0.550	46	445	4532	21071	36346	51231
0.650	39	377	3835	17830	30755	43349
0.950	27	258	2624	12200	21043	29660
1.000	**26**	**245**	**2493**	**11590**	**19991**	**28177**
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.190	27	257	2622	12197	21041	29660
0.200	26	244	2491	11587	19989	28177
0.230	22	212	2166	10075	17382	24501
0.250	20	195	1992	9269	15991	22541
0.260	18	187	1915	8912	15375	21673
0.265	**17**	**182**	**1878**	**8743**	**15084**	**21263**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.100	133	1294	13197	61384	105893	149265
0.200	93	909	9265	43089	74327	104769
0.250	**89**	**878**	**8952**	**41630**	**71809**	**101218**
0.275	89	881	8981	41763	72038	101539
0.300	90	893	9108	42352	73053	102970
0.350	94	943	9620	44733	77157	108754
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.100	99	960	9786	45518	78520	110679
0.130	88	864	8809	40967	70668	99611
0.145	86	841	8580	39904	68833	97024
0.150	**85**	**837**	**8533**	**39682**	**68450**	**96483**
0.190	85	839	8559	39803	68656	96772
0.200	86	849	8666	40296	69506	97970

these functions where $x = d$ are displayed in Table 3.5. Since the design space is [-1,1] for all four functions, the D-optimal design obtained for the quartic regression model is as follows.

$$P^* = \left\{ \begin{array}{ccccccc} -1.00 & -0.66 & -0.65 & 0.00 & 0.65 & 0.66 & 1.00 \\ 0.200 & 0.085 & 0.115 & 0.200 & 0.115 & 0.085 & 0.200 \end{array} \right\}$$

Here also, the convex combinations of the particular cluster members yield the solution

$$P^* = \left\{ \begin{array}{ccccc} -1.00 & -0.655 & 0.00 & 0.655 & 1.00 \\ 0.200 & 0.200 & 0.200 & 0.200 & 0.200 \end{array} \right\}$$

Additionally, by the results obtained, the best choices of δ for the four different functions are

- 1.000 for $f(x, \delta) = x^\delta$
- 0.265 for $f(x, \delta) = e^{\delta x}$
- 0.250 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$
- 0.150 for $f(x, \delta) = \Phi(\delta x)$

The results obtained using $x = F$ instead of $x = d$ in the function for the quartic regression model is given in the Table 3.6.

Table 3.6: Number of iterations need to achieve $\max\limits_{1\leq j\leq J}\{F_j\} \leq 10^{-t}$ with $x = F$ for quartic regression model

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.100	101	977	9965	46350	79958	112708
0.200	51	489	4983	23176	39979	56354
0.250	41	391	3987	18541	31984	45083
0.275	37	356	3625	16856	29076	40985
0.300	34	326	3323	15451	26653	37569
0.350	**29**	**280**	**2848**	**13244**	**22846**	**32202**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.100	63	613	6245	29046	50106	70629
0.130	49	471	4804	22344	38544	54330
0.145	44	423	4307	20033	34556	48710
0.150	43	409	4164	19365	33405	47086
0.190	34	323	3288	15289	26372	37173
0.200	**32**	**307**	**3123**	**14524**	**25054**	**35315**

3.3.4 Application of D-optimality Criterion to a Practical Problem in Chemistry

This practical model was considered by Torsney and Alahmadi (1995), and Mandal, Torsney and Chowdhury (2017). The main objective of this problem was to study the relationship between the viscosity and the concentration of a chemical solution using a three parameter regression model. They wanted to construct a design in a way that the parameters of interest are as uncorrelated as possible. The model is

$$E(y|x) \quad = \quad \theta_0 x + \theta_1 x^{1/2} + \theta_2 x^2. \tag{3.4}$$

Here, the viscosity is the response variable y and the concentration of a chemical solution is the explanatory variable x. The design space is (0,0.2] which is so short. Therefore, minimizing the covariance between the parameters of interest is very challenging under any design.

The initial design space consists of 20 points equally spaced at intervals of 0.01 between 0.0 and 0.2. The functions being considered are x^δ, $e^{\delta x}$, logistic CDF $e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x, \delta) = \phi(\delta x)$ where $x = d$, partial derivatives.

The results obtained for the total number of iterations when using the above mentioned functions where $x = d$ are displayed in Table 3.7.

The D-optimal design obtained for the chemistry model is as follows.

$$P^* = \left\{ \begin{array}{cccc} 0.02 & 0.11 & 0.12 & 0.20 \\ 0.333 & 0.255 & 0.079 & 0.333 \end{array} \right\}$$

Moreover, the best choices for the δ for each of the functions are,

- 1.800 for $f(x, \delta) = x^\delta$
- 0.550 for $f(x, \delta) = e^{\delta x}$
- 0.450 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$
- 0.250 for $f(x, \delta) = \Phi(\delta x)$

Next we will observe the convergence rate of obtaining an D-optimal design when using $x = F$, the directional derivatives with the function. The promising results were achieved with logistic CDF, $e^{\delta x}/1 + e^{\delta x}$ and normal

Table 3.7: Number of iterations need to achieve $\max_{1 \le j \le J} \{F_j\} \le 10^{-t}$ with $x = d$ for practical model in chemistry

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.500	32	167	789	1958	3297	4661
0.600	27	139	657	1632	2747	3883
0.750	22	112	526	1305	2197	3106
0.950	18	88	415	1030	1734	2451
1.500	12	56	263	652	1097	1551
1.800	**9**	**47**	**219**	**543**	**914**	**1292**
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.100	51	278	1316	3266	5498	7772
0.250	21	112	526	1306	2198	3107
0.300	18	93	439	1088	1831	2588
0.450	12	62	292	725	1220	1724
0.500	11	56	263	652	1098	1551
0.550	**11**	**51**	**239**	**592**	**997**	**1410**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.350	58	306	1448	3597	6057	8563
0.450	**57**	**300**	**1417**	**3522**	**5931**	**8385**
0.500	58	304	1439	3576	6023	8516
0.650	66	342	1619	4027	6784	9592
0.750	75	386	1831	4558	7679	10858
0.900	95	484	2308	5748	9686	13698
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.150	65	346	1638	4067	6848	9680
0.250	**54**	**285**	**1349**	**3352**	**5645**	**7980**
0.350	57	294	1390	3457	5824	8235
0.400	61	315	1492	3713	6256	8846
0.500	78	396	1883	4692	7909	11185
0.600	110	553	2652	6619	11162	15790

CDF where $x = F$. The iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ are given in Table 3.8.

Table 3.8: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for practical model in chemistry

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.350	31	159	751	1865	3140	4438
0.400	27	139	657	1631	2747	3883
0.500	22	112	525	1304	2196	3105
0.650	17	86	403	1003	1689	2387
0.800	15	70	327	814	1371	1939
0.950	**13**	**59**	**275**	**685**	**1154**	**1632**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.200	33	175	824	2045	3444	4868
0.350	20	100	470	1167	1966	2779
0.400	18	87	411	1021	1720	2431
0.500	15	70	328	816	1375	1944
0.650	12	54	252	627	1057	1494
0.800	**8**	**44**	**205**	**509**	**858**	**1213**

When comparing the results of logistic CDF and normal CDF, it can be seen that the convergence rate has increased when using directional derivatives instead of partial derivatives.

3.4 Construction of D-optimal Designs for Regression Models with Two Variables

So far we discussed only about one design variable. When working with statistical models, we also need to focus on models with more than one design variable. In this section, we are going to focus on polynomial regression models with two design variables. The design variables are defined as x_1 and x_2. The full model we are considering is given by

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 = \underline{v}_x^T \underline{\theta}. \tag{3.5}$$

Here, $\underline{v}_x = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)^T$, $x \in [-1, 1]$, $\underline{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5)^T$ and the induced design space is given by $\underline{v}_x \in \mathcal{V} = \{\underline{v}_x : \underline{v}_x = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)^T, -1 \leq x_i \leq 1\}$ where $i = 1, 2$.

We are going to construct D-optimal designs for the two variable full model and two variable reduced models. First model we are going to consider is defined as simple model which does not include squared terms and interaction term. Next we are going to focus on the reduced model without the interaction term. Finally, the full model will be considered. For all three situations, the discretized design space will consist of all the pairs (x_1, x_2) where values of x_i, $i = 1, 2$, will be between -1.0 and 1.0 at intervals of 0.1. In total the design space will consist of $(21)^2 = 441$ pairs of (x_1, x_2). The results will be reported in the same way as in for one variable model using the functions $x^\delta, e^{\delta x}, e^{\delta x}/(1 + e^{\delta x})$ and normal CDF. The convergence rate of the algorithms are compared by replacing partial derivatives of the criterion function with directional derivatives.

3.4.1 Two Variable Model with θ_0, θ_1 and θ_2 as the Parameters of Interest

The D-optimal design for the following two variable model is constructed.

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \tag{3.6}$$

The design space is approximated by 441 pairs of (x_1, x_2). The convergence of the algorithms is assessed using directional derivatives of D-optimality criterion instead of partial derivatives. The number of iterations needed when using the above mentioned functions withe $x = d$ are displayed in Table 3.9.

The optimal design obtained for this model is as follows.

$$P^* = \begin{Bmatrix} -1.00 & 1.00 & -1.00 & 1.00 \\ -1.00 & -1.00 & 1.00 & 1.00 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{Bmatrix}$$

Moreover, the best choices for the δ for each of the functions are,

- 0.400 for $f(x, \delta) = x^\delta$
- 0.900 for $f(x, \delta) = e^{\delta x}$

- 0.400 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$
- 0.250 for $f(x, \delta) = \Phi(\delta x)$

We will observe the convergence rate when using $x = F$, the directional derivatives. The promising results were achieved with logistic CDF, $e^{\delta x}/1 + e^{\delta x}$ and normal CDF. The iteration values need to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ are given in Table 3.10.

Table 3.9: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for two variable regression model with θ_0, θ_1 and θ_2 as the parameters of interest

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.250	93	223	369	518	666	815
0.300	87	208	343	480	617	754
0.400	**85**	**199**	**324**	**451**	**578**	**705**
0.500	89	205	332	459	587	715
0.650	105	236	377	519	660	802
0.750	122	270	429	587	746	905
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.200	35	86	146	206	267	327
0.350	20	50	84	118	153	187
0.500	15	35	59	83	107	131
0.600	12	29	49	69	90	110
0.725	10	24	41	58	74	91
0.900	**8**	**20**	**33**	**46**	**60**	**73**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.250	93	223	369	518	666	815
0.300	87	208	343	480	617	754
0.400	**85**	**199**	**324**	**451**	**578**	**705**
0.500	89	205	332	459	587	715
0.650	105	236	377	519	660	802
0.750	122	270	429	587	746	905
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.175	86	206	341	478	615	753
0.250	**81**	**189**	**308**	**429**	**550**	**672**
0.300	83	191	309	429	548	668
0.400	98	220	350	480	611	742
0.500	132	286	448	609	771	932
0.600	196	415	637	859	1080	1302

Table 3.10: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for two variable regression model with θ_0, θ_1 and θ_2 as the parameters of interest

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.250	59	142	236	332	428	524
0.450	35	82	134	186	239	292
0.550	30	68	110	153	196	239
0.625	27	61	98	135	173	211
0.700	25	55	88	121	155	188
0.800	**22**	**49**	**78**	**107**	**136**	**165**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.350	29	67	109	151	193	235
0.450	24	53	86	118	151	184
0.600	19	42	66	90	114	138
0.700	17	37	57	78	98	119
0.850	15	31	48	65	81	98
0.900	**15**	**30**	**45**	**61**	**77**	**93**

When comparing the results of logistic CDF and normal CDF, it can be seen that the convergence rate has increased when using directional derivatives of the criterion function instead of partial derivatives.

3.4.2 Two Variable Model with $\theta_0, \theta_1, \theta_2, \theta_3$ and θ_4 as the Parameters of Interest

Let us now consider the two variable regression model consisting of the squared terms of the variables.

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 \tag{3.7}$$

Here also, the design space was approximated by 441 pairs of (x_1, x_2) The number of iterations recorded for $x = d$ are displayed in Table 3.11. The optimal design obtained for the model including the best choices for the δ for each of the functions are as follows.

$$P^* = \begin{Bmatrix} -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 \\ -1.00 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 1.00 \\ 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 \end{Bmatrix}$$

- 0.925 for $f(x, \delta) = x^\delta$
- 0.575 for $f(x, \delta) = e^{\delta x}$
- 0.270 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$
- 0.200 for $f(x, \delta) = \Phi(\delta x)$

Next we will find about the convergence rate of obtaining the D-optimal design when using $x = F$, the directional derivatives with the function. The iterations needed to reach $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ with logistic CDF,

$e^{\delta x}/1 + e^{\delta x}$ and normal CDF where $x = F$ are given in Table 3.12.

According to the results, it can be seen that the convergence rate has increased when using the directional derivatives of the criterion function.

3.4.3 Two Variable Full Model

Let us consider the full model (3.5) with all the possible terms

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 = \underline{v}_x^T \underline{\theta}.$$

As in for the earlier situations, the design space was approximated by 441 pairs of (x_1, x_2) where each variable was equally spaced spaced between -1.0 and 1.0 with 0.1 intervals. The results for $x = d$ are displayed in Table 3.13. The D-optimal design and the best choices of δ for each of the functions are as follows.

$$P^* = \begin{Bmatrix} -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 \\ -1.00 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 1.00 \\ 0.1458 & 0.0802 & 0.1458 & 0.0802 & 0.0962 & 0.0802 & 0.1458 & 0.0802 & 0.1458 \end{Bmatrix}$$

- 0.950 for $f(x, \delta) = x^\delta$
- 0.350 for $f(x, \delta) = e^{\delta x}$
- 0.200 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$
- 0.200 for $f(x, \delta) = \Phi(\delta x)$

The results for $x = F$ are displayed in Table 3.14. When comparing the results of these two tables, it can be seen that the convergence rate has improved when using directional derivatives.

Generally, when we run the multiplicative algorithm for a few number of iterations for all these regression models, we will be left with a number of disjoint set of clusters. This means clustering starts happening in early iterations. We will deal with this matter in next chapter.

Table 3.11: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for two variable regression model with $\theta_0, \theta_1, \theta_2, \theta_3$ and θ_4 as the parameters of interest

$f(x,\delta) = x^\delta$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.400	121	763	1435	2081	2725	3368
0.500	97	610	1148	1665	2180	2695
0.600	81	509	957	1388	1817	2246
0.725	67	421	792	1149	1504	1859
0.800	61	382	718	1041	1363	1684
0.925	**53**	**330**	**621**	**901**	**1179**	**1457**
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.200	49	303	573	833	1091	1350
0.275	36	221	417	606	794	982
0.300	33	202	382	555	728	900
0.400	25	152	287	417	546	675
0.500	20	122	229	333	437	540
0.575	**13**	**105**	**199**	**289**	**379**	**469**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.100	257	1608	3036	4409	5776	7143
0.270	**174**	**1098**	**2066**	**2995**	**3921**	**4846**
0.350	185	1181	2217	3213	4204	5195
0.425	210	1352	2535	3671	4802	5932
0.500	249	1619	3032	4387	5737	7085
0.600	327	2161	4043	5845	7639	9433
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.090	201	1262	2380	3456	4527	5598
0.125	173	1089	2053	2979	3901	4824
0.200	**167**	**1062**	**1997**	**2894**	**3788**	**4682**
0.250	187	1200	2251	3260	4265	5269
0.300	226	1475	2762	3996	5225	6452
0.400	404	2793	5207	7516	9813	12110

Table 3.12: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for two variable regression model with $\theta_0, \theta_1, \theta_2, \theta_3$ and θ_4 as the parameters of interest

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.500	39	245	460	667	872	1078
0.600	33	205	384	556	727	898
0.725	28	170	318	461	602	744
0.800	25	155	289	418	546	674
0.900	23	138	257	372	486	599
1.000	**21**	**125**	**232**	**335**	**437**	**540**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.375	33	205	385	557	729	901
0.500	25	155	290	419	547	676
0.575	22	136	252	365	476	588
0.650	20	121	224	323	422	520
0.725	19	109	202	290	378	467
0.800	**18**	**99**	**183**	**263**	**343**	**423**

Table 3.13: Number of iterations need to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for the full model (3.5)

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.300	196	1097	1991	2855	3716	4577
0.500	118	659	1195	1713	2230	2746
0.650	91	507	919	1318	1715	2113
0.700	85	471	854	1224	1593	1262
0.800	74	412	747	1071	1394	1717
0.950	**63**	**347**	**629**	**902**	**1174**	**1446**
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.090	110	607	1104	1586	2067	2547
0.100	99	546	994	1428	1860	2293
0.150	66	364	663	952	1240	1529
0.200	50	273	497	714	930	1146
0.250	40	219	398	571	744	917
0.350	**28**	**156**	**284**	**408**	**531**	**655**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.090	201	1262	2380	3456	4527	5598
0.125	173	1089	2053	2979	3901	4824
0.200	**167**	**1062**	**1997**	**2894**	**3788**	**4682**
0.250	187	1200	2251	3260	4265	5269
0.300	226	1475	2762	3996	5225	6452
0.400	404	2793	5207	7516	9813	12110
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.100	278	1543	2806	4029	5247	6465
0.175	217	1208	2194	3147	4097	5047
0.200	**212**	**1185**	**2150**	**3083**	**4014**	**4944**
0.250	215	1204	2183	3130	4073	5015
0.300	229	1292	2340	3353	4362	5371
0.500	400	2326	4200	6008	7808	9609

Table 3.14: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the full model (3.5)

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.200	99	548	995	1428	1859	2290
0.350	57	314	570	816	1062	1308
0.400	50	275	499	714	930	1145
0.500	42	221	400	572	744	916
0.600	34	185	334	477	620	763
0.725	**29**	**154**	**277**	**396**	**514**	**632**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.150	83	458	832	1193	1553	1913
0.200	62	344	624	895	1165	1435
0.250	50	276	500	716	932	1148
0.300	42	231	417	597	777	956
0.400	32	174	314	448	583	717
0.450	**29**	**155**	**279**	**399**	**518**	**638**

Chapter 4

Construction of D-optimal Designs using Clustering Approach

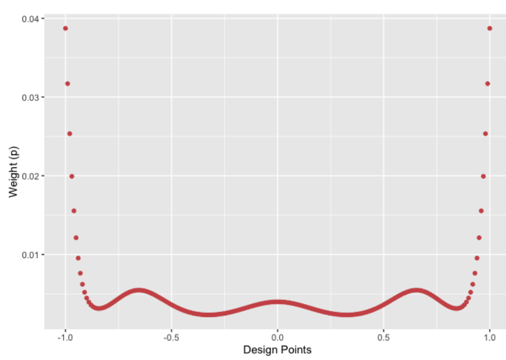
Traditional method of constructing an optimal design will usually result in higher number of iterations which is not easy to implement. Sometimes, with many number of nonsupport points with zero weights, convergence of the algorithm obtaining an optimal design becomes slow. In the previous chapter, we noted that, in some discretized design spaces of the regression models, the optimal support points consist of number of clusters. Therefore, we hope to enhance the convergence rate of the algorithm with the use of clustering approach. With the clustering approach, the number of iterations will be reduced a great deal, and it will save time, cost and resources of the experimenter.

We will consider a general regression problem to explain the idea behind the clustering approach.

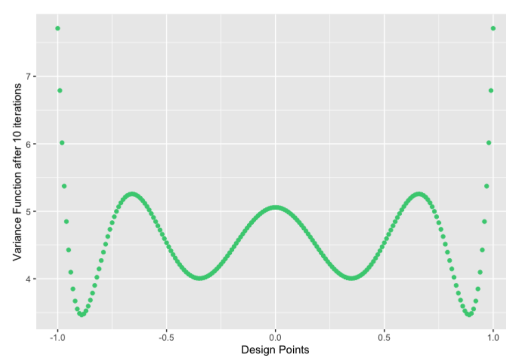
4.1 General Problem

Suppose we run the multiplicative algorithm for a small number of iterations and obtain the plots of weights and the variance function with relative to the design points. These plots will be curves with a number of maximal and minimal turning points.

For an example, consider the following plots of weights and variance function obtained for the quartic regression model after 10 iterations. In Figure



(a) Weights vs design points



(b) Variance vs design points

Figure 4.1: Weights and variance function vs design points for the quartic regression model

4.1, the plots clearly depict curves with three maximal and four minimal turning points. The minimal turning points are related to the regions of zero weights. The weights on these regions of minimal turning points are converging to zero. With the assumption of such zero weights, we will be having a number of disjoint clusters of points that include the optimal support points. If we consider the designs obtained after running the algorithm for small number of iterations using traditional approach, they consist of same number of clusters as the number of support points in the design space of the final solution. This was observed by Mandal and Torsney (2006). We make use of this approach and attempt to further improve the convergence, generalize this approach to optimizing distributions with respect to several distributions as well as to apply other criteria. A similar feature related to this is presented in Wynn(1990) algorithm. Torsney (1983) ran this algorithm for more than 100 iterations in solving a problem related to trigonometric regression.

4.1.1 Transformation of Weights to Within Cluster and Total Cluster Weights

Suppose that for a particular model and a given design space, there are g number of clusters $C_1, C_2, ..., C_g$. Any point within a cluster is denoted by (l, m). This means the m^{th} point in l^{th} cluster (C_l). Let $p_{l,m}$ be the weight such that $1 \leq l \leq g$, $1 \leq m \leq n_l$ and n_l is the l^{th} cluster size where $l = 1, 2, ..., g$.

Next step is to obtain within cluster weights and total cluster weights which are denoted by $s_{l,m}$ and t_l respectively. First, we have to divide the weights $p_{l,m}$ into g number of clusters. The summation of weights $p_{l,m}$ in each of the clusters will be the total cluster weight for a particular cluster. The within cluter weights are formed by dividing the weights $p_{l,m}$ in each of the clusters by its total cluster weight. Consider the following formulas of forming total cluster weights and within cluster weights respectively

$$t_l = \sum_{m=1}^{n_l} p_{l,m}, \qquad 1 \leq l \leq g \tag{4.1}$$

$$s_{l,m} = p_{l,m}/t_l, \qquad 1 \leq m \leq n_l, \quad 1 \leq l \leq g \tag{4.2}$$

Here, t_l denotes the total cluster weight such that $l = 1, 2, ..., g$ and $s_{l,m}$ is the m^{th} within cluster weight for cluster l (C_l) such that $l = 1, 2, ..., g$ and $m = 1, 2, ..., n_l$.

As in for the traditional approach of constructing optimal designs, there are some constraints that has to be considered in clustering approach also. They are,

$$\sum_{l=1}^{g} t_l = 1, \qquad t_l \geq 0, \quad l = 1, 2, ..., g$$

$$\sum_{m=1}^{n_l} s_{l,m} = 1, \qquad s_{l,m} \geq 0, \quad l = 1, 2, ..., g, m = 1, 2, ..., n_l$$

Next step is to formulate the optimization problem using the total cluster weights t_l and the within cluster weights $s_{l,m}$. Then, the criterion function $\phi(p)$ becomes a function of t_l and $s_{l,m}$ and it can be stated as given below

$$\phi(p) = \tilde{\phi}(\underline{t}, \underline{s}_1, \underline{s}_2, ..., \underline{s}_g) \tag{4.3}$$

where,

$\underline{t} = (t_1, t_2, ..., t_g)^T$

$\underline{s}_l = (s_{l,1}, s_{l,2}, ..., s_{l,n_l})^T, \qquad l = 1, 2, ..., g.$

Here, we have to select $\underline{t}, \underline{s}_1, \underline{s}_2, ..., \underline{s}_g$ optimally. Then our optimization problem can be denoted as maximizing $\tilde{\phi}(\underline{t}, \underline{s}_1, \underline{s}_2, ..., \underline{s}_g)$ subject to the constraints

$$\sum_{l=1}^{g} t_l = 1, \qquad t_l \geq 0, \quad l = 1, 2, ..., g$$
$$\sum_{m=1}^{n_l} s_{l,m} = 1, \qquad s_{l,m} \geq 0, \quad l = 1, 2, ..., g, m = 1, 2, ..., n_l$$

Next we will consider the algorithm of constructing optimal designs using clustering approach.

4.1.2 Algorithm for Constructing Optimal Designs using Clustering Approach

The algorithm has to be modified according to different sets of probability distributions. In clustering approach, we have to use two separate algorithms corresponding to total cluster weights t_l and within cluster weights $s_{l,m}$

$$t_l^{(n+1)} = t_l^{(n)} f_t(\tilde{x}_l^{(n)}) / \sum_{k=1}^{g} t_k^{(n)} f_t(\tilde{x}_k^{(n)}), \qquad 1 \leq l \leq g \tag{4.4}$$
$$s_{l,m}^{(n+1)} = s_{l,m}^{(n)} f_l(\tilde{x}_{l,m}^{(n)}) / \sum_{h=1}^{n_l} s_{l,h}^{(n)} f_l(\tilde{x}_{l,h}^{(n)}), \qquad 1 \leq m \leq n_l, 1 \leq l \leq g \tag{4.5}$$

Here,

- $\underline{t}^{(n)}$ = Total cluster weight at nth iteration.

- $\underline{s}_l^{(n)}$ = Within cluster weight for C_l at nth iteration where $l = 1, 2, ..., g$.
- $\tilde{x}_l = \tilde{d}_l = \dfrac{\partial \tilde{\phi}}{\partial t_l} = \sum_{m=1}^{n_l} d_{l,m} s_{l,m}$ where $\tilde{d}_l$ is the partial derivative of $\tilde{\phi}$ with respect to t_l.
- $\tilde{x}_{l,m} = \tilde{d}_{l,m} = \dfrac{\partial \tilde{\phi}}{\partial s_{l,m}} = t_l d_{l,m}$ where $\tilde{d}_{l,m}$ is the partial derivatives of $\tilde{\phi}$ with respect to $s_{l,m}$.
- The above partial derivatives $\tilde{d}_l$ and $\tilde{d}_{l,m}$ are defined using the partial derivatives $d_{l,m}$ of original weights $p_{l,m}$.
- $f_t(.)$ and $f_l(.)$ are strictly increasing positive functions. They might depend on free positive parameters like δ_t and δ_l where $l = 1, 2, ..., g$ respectively.

The properties related to this algorithm is similar to the properties of the algorithm in traditional approach (Mandal and Torsney, 2006; Mandal et al., 2017).

- $\underline{t}^{(n)}, \underline{s}_1^{(n)}, ..., \underline{s}_g^{(n)}$ are always feasible.
- $F_{\tilde{\phi}} \underline{t}^{(n)}, \underline{s}_1^{(n)}, ..., \underline{s}_g^{(n)}; \underline{t}^{(n+1)}, \underline{s}_1^{(n+1)}, ..., \underline{s}_g^{(n+1)} \geq 0$
- When the partial derivatives related to nonzero weights of the distributions t and s are all equal, an iterate $\underline{t}^{(n)}, \underline{s}_1^{(n)}, ..., \underline{s}_g^{(n)}$ will be a fixed point of the iteration. This is applied similarly for the situation where corresponding vertex directional derivatives are zero.

4.2 Quadratic Regression Model

The optimal design for the quadratic regression model was constructed using clustering approach algorithm. The model is given by

$$E(Y|x) = \theta_0 + \theta_1 x + \theta_2 x^2.$$

The functions we considered in here are x^δ, logistic cumulative density function $e^{\delta x}/1 + e^{\delta x}$ and normal cumulative density function $\phi(\delta x)$.

The design space was approximated by a grid of 201 points equally spaced at intervals of 0.01 between -1.00 and 1.00. In order to form the clusters, δ was taken as 0.5 and total number of iterations between 100 and 150 for the function $f(x, \delta) = x^\delta$ when $x = d$ were considered. For the logistic CDF and normal CDF where $x = F$, δ was taken as 0.575.

The results were first obtained for the function $f(x, \delta) = x^\delta$ when $x = d$, partial derivatives. For the logistic CDF and normal CDF, the best results were obtained when we replaced partial derivatives of optimality criterion in the function $f(x, \delta)$ by the directional derivatives.

The number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the function $f(x, \delta) = x^\delta$ where $x = d$ are displayed in Table 4.1. Moreover, the number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the functions logistic CDF $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x, \delta) = \phi(\delta x)$ where $x = F$ are displayed in Table 4.2.

The D-optimal design obtained for the quadratic regression model with

Table 4.1: Number of iterations need to achieve $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-t}$ with $x = d$ for quadratic regression model with the use of clustering approach

$f(x,\delta) = x^{\delta}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
12	1	2	41	80	116	152
15	1	2	33	64	93	122
18	1	2	28	54	78	102
20	1	2	25	49	71	92
23	1	2	22	43	62	80
25	**1**	**2**	**21**	**39**	**57**	**74**

Table 4.2: Number of iterations need to achieve $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-t}$ with $x = F$ for quadratic regression model with the use of clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
50	1	1	4	73	147	217
70	1	1	3	52	105	155
80	1	1	3	46	92	135
95	1	1	3	39	77	114
100	1	1	3	37	73	108
120	**1**	**1**	**2**	**31**	**61**	**90**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
55	1	1	3	41	82	120
65	1	1	3	35	69	101
70	1	1	3	33	64	94
80	1	1	3	29	56	82
90	1	1	2	26	50	73
100	**1**	**1**	**2**	**23**	**45**	**65**

the function $f(x,\delta) = x^{\delta}$ when $x = d$, using clustering approach is as follows. This means, we obtained the same design as traditional approach.

$$P^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 0.333 & 0.333 & 0.333 \end{array} \right\}$$

The D-optimal designs obtained for the quadratic regression model with the functions $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x,\delta) = \phi(\delta x)$ when $x = F$, using clustering approach is as follows.

$$P^* = \left\{ \begin{array}{ccccc} -1.00 & -0.01 & 0.00 & 0.01 & 1.00 \\ 0.333 & 0.015 & 0.303 & 0.015 & 0.333 \end{array} \right\}$$

It is noticeable that, the support points consist of three clusters of points which is either of neighboring points or a single point. Since this is a solution for a discretized design space, the solution for a continuous design space should be a 3-point design. Therefore, by taking the convex combinations of the relevant cluster elements, the solution below can be obtained.

$$P^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ 0.333 & 0.333 & 0.333 \end{array} \right\}$$

The best choices of δ for the above three functions are,

- 25 for $f(x,\delta) = x^\delta$ when $x = d$
- 120 for $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ when $x = F$
- 100 for $f(x,\delta) = \Phi(\delta x)$ when $x = F$

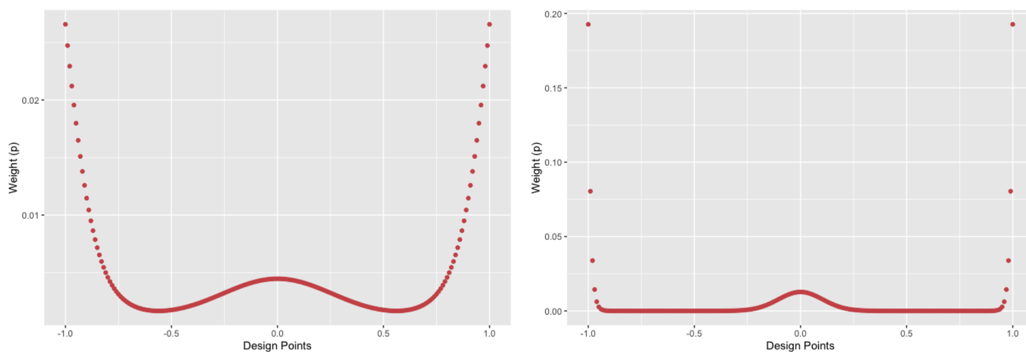
According to the results, it can be seen that when increasing the δ value, the number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ have been decreased. This will increase the convergence rate.

Graphical Interpretation

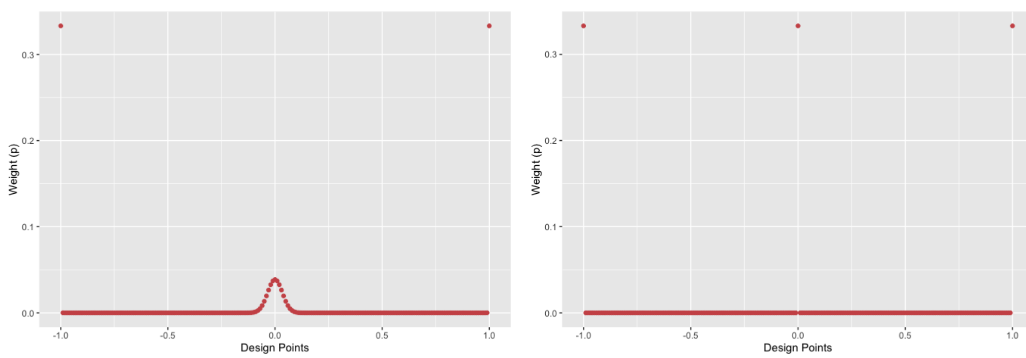
Consider the optimal design construction for the quadratic model using D-optimality criterion. The construction is done using multiplicative algorithm where the function is logistic cumulative density function and δ is 0.45. Figure 4.2 provides the graphs that we obtained after running the algorithm for few iterations(10 iterations (Figure 4.2 (a)) at first and thereafter 100 iterations (Figure 4.2 (b)), 1000 iterations (Figure 4.2 (c))) and at the optimum (Figure 4.2 (d)).

The plot 4.2 (a) clearly depicts curves with two minimal and one maximal turning points. Two peak points are visible at the design point $x = 1.00$ and $x = -1.00$. The weights on the minimal turning points are converging to zero. It is noticeable in plot 4.2 (d), all the minimal turning points have converged to zero when increasing the total number of iterations in the algorithm. It can be seen that at the optimum the weights corresponding to the data points $x = -1.00, 0.00, 1.00$ have converged to 0.333 and the rest have converged to zero.

Figure 4.3 shows the plots of variance function versus design points after running algorithm for 10 iterations (Figure 4.3 (a)) and at the optimum (Figure 4.3 (b)). In here also plot 4.3 (a) depicts curves with two minimal and one maximal turning points including two peaks at the design points -1.00 and 1.00 as in the previous weights versus design points plot (a). At the optimum, the variance function values corresponding to design points 1.00 and -1.00 have converged to 3. The variance function corresponding to



(a) After 10 iterations (b) After 100 iterations

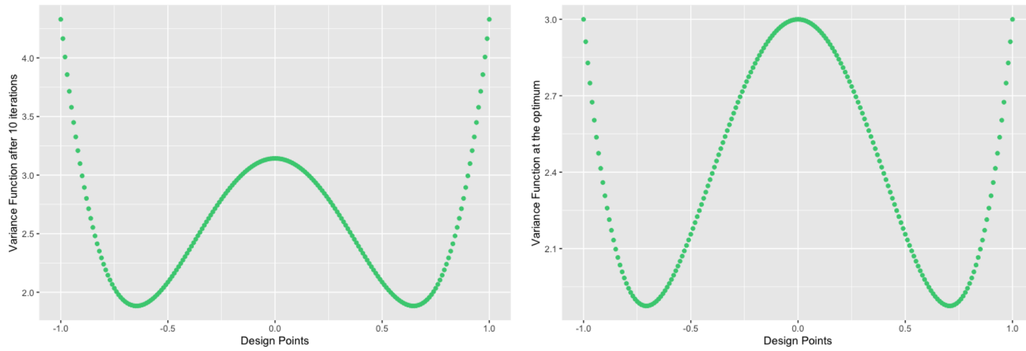


(c) After 1000 iterations (d) At the optimum

Figure 4.2: Weights vs design points for the quadratic regression model

this quadratic model is,

$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) \\ &= 3 - 4.5x^2 + 4.5x^4 \end{aligned}$$



(a) After 10 iterations (b) At the optimum

Figure 4.3: Variance function vs design points for the quadratic regression model

4.3 Cubic Regression Model

The optimal design construction was done for the cubic regression model given below using clustering approach.

$$E(Y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

Here also, the design space was equally spaced at intervals of 0.01 between -1.00 and 1.00. To form the clusters at the beginning, δ was taken as 1.5 and total number of iterations as 85 for the function $f(x, \delta) = x^\delta$ where $x = d$. Then $\delta = 0.5$ and total number of iterations as 250 were considered for logistic CDF and normal CDF where $x = F$. In Table 4.3, the number of iterations are recorded for the function $f(x, \delta) = x^\delta$ where $x = d$ are used.

The D-optimal design obtained for the cubic regression model using the function $f(x, \delta) = x^\delta$ where $x = d$ is given below.

Table 4.3: Number of iterations need to achieve $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-t}$ with $x = d$ for cubic regression model with the use of clustering approach

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
1.5	1	1	822	1740	2574	3392
1.7	1	1	725	1535	2271	2993
1.9	1	1	649	1374	2032	2678
2.05	1	1	602	1274	1884	2483
3	1	1	412	871	1288	1697
3.5	**1**	**1**	**353**	**747**	**1104**	**1455**

$$P^* = \left\{ \begin{array}{cccccc} -1.00 & -0.45 & -0.44 & 0.44 & 0.45 & 1.00 \\ 0.250 & 0.230 & 0.020 & 0.020 & 0.230 & 0.250 \end{array} \right\}$$

Here also, it is noticed that the design space is composed of four clusters with a pair of neighboring points or a single point. Therefore, by taking the convex combinations of the relevant cluster members, the following D-optimal design for the cubic regression model can be obtained.

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.447 & -0.447 & 1.00 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{array} \right\}$$

The iterations needed to achieve $max\{F_j\} \le 10^{-t}$ for $1 \le j \le J$ when using the functions logistic CDF $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x,\delta) = \phi(\delta x)$ where $x = F$ are displayed in Table 4.4.

Here also, the D-optimal designs were constructed using the logistic CDF and normal CDF. We obtained the same number of clusters consist of pair of neighboring points or a single point. The only difference can be seen in the weights of the middle design points. The results are shown below.

Table 4.4: Number of iterations need to achieve $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-t}$ with $x = F$ for cubic regression model with the use of clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
55	1	1	9	50	77	103
65	1	1	7	42	65	87
70	1	1	7	39	60	81
85	1	1	6	32	50	66
95	1	1	6	29	44	59
100	**1**	**1**	**5**	**28**	**42**	**56**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
60	1	1	9	25	37	50
75	1	1	8	20	30	39
80	1	1	8	19	28	37
85	1	1	7	18	26	34
90	1	1	7	17	25	32
100	**1**	**1**	**7**	**15**	**22**	**29**

$$P^* = \left\{ \begin{array}{cccccc} -1.00 & -0.45 & -0.44 & 0.44 & 0.45 & 1.00 \\ 0.250 & 0.145 & 0.105 & 0.105 & 0.145 & 0.250 \end{array} \right\}$$

If we take the convex combinations of the cluster points, these two yield the same solution. Next we report the the best choices of δ values for the above three situations.

- 3.5 for $f(x,\delta) = x^\delta$ when $x = d$
- 100 for $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ when $x = F$
- 100 for $f(x,\delta) = \Phi(\delta x)$ when $x = F$.

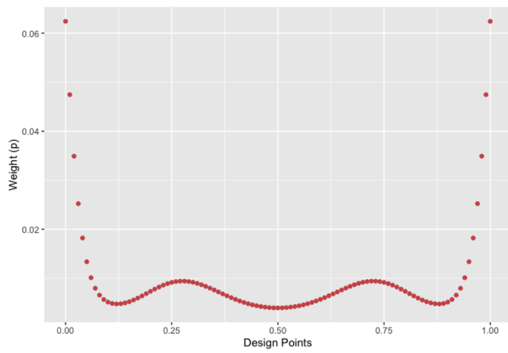
Graphical Interpretation

Figure 4.4 provides the plots of weights versus design points corresponding to several iterations and at the optimum. According to the figure 4.4 (a), it can be seen that, there are three minimal and two maximal turning points in the curve. When increasing the number of iterations, the weights corresponding to these minimal turning points are converging to zero. The weights corresponding to the design points 0.00 and 1.00 have converged to 0.25 at the optimum. It is noticeable that, in the plot 4.4 (d), the weights are arranged as clusters, specifically in the middle between the design points 0.25 and 0.75. This explains the design space for the cubic regression model obtained with the four clusters of points.

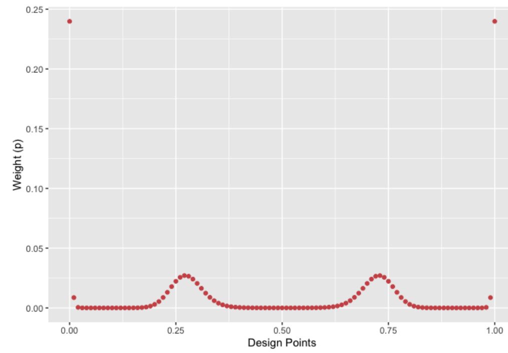
Figure 4.5 provides the corresponding plots of variance function versus design points of the above model. Here in the plot 4.5 (a), the after running the algorithm for 10 iterations, a curve with three minimal and two maximal turning points has been obtained. Two peaks can be seen at the design points 0.00 and 1.00. Then at the optimum, the variance function values corresponding to the design points 0.00 and 1.00 have converged to 4. The variance function corresponding to this cubic regression model is,

$$d(\underline{x}, p) = \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x})$$

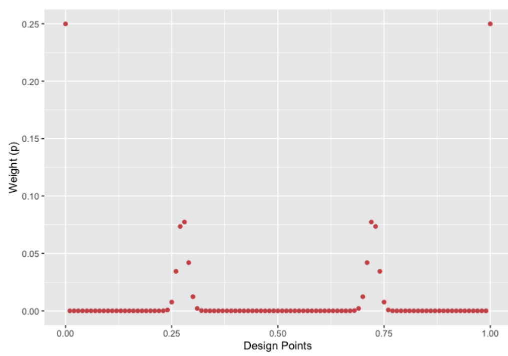
$$= 4 - 47.8924x + 528.4343x^2 - 2166.3436x^3 + 4096.3219x^4 - 3615.78x^5 + 1205.26x^6.$$



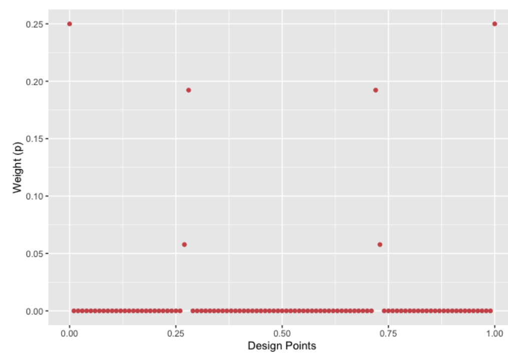
(a) After 10 iterations



(b) After 100 iterations

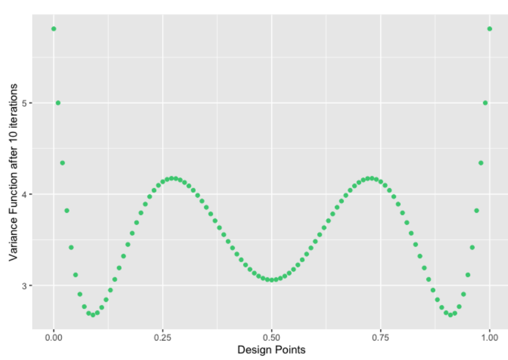


(c) After 1000 iterations

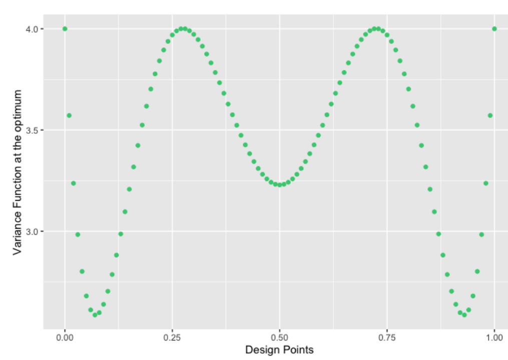


(d) At the optimum

Figure 4.4: Weights vs design points for the cubic regression model



(a) After 10 iterations



(b) At the optimum

Figure 4.5: Variance function vs design points for the cubic regression model

4.4 Quartic Regression Model

The D-optimal design was constructed using clustering approach for the quartic regression model given below.

$$E(Y|x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

When using the clustering approach for the quartic regression model, the best results for the function $f(x, \delta) = x^\delta$ were obtained with $x = d$, the partial derivatives and for the logistic CDF and normal CDF with $x = F$, the directional derivatives. The results were the same even if we use directional derivatives instead of partial derivatives with the function $f(x, \delta) = x^\delta$. The iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the function $f(x, \delta) = x^\delta$ where $x = d$ are displayed in the Table 4.5.

Table 4.5: Number of iterations need to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for quartic regression model with the use of clustering approach

$f(x,\delta) = x^\delta$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
15	1	1	97	180	262	345
20	1	1	73	136	197	259
22	1	1	67	123	179	236
25	1	1	59	109	158	208
26	1	1	57	105	152	200
27	**1**	**1**	**55**	**101**	**147**	**193**

Here also, the design space consists of 5 clusters of points which are either with few neighboring points or a single point.

$$P^*=\left\{\begin{array}{ccccccccc} -1 & -0.66 & -0.65 & -0.01 & 0.00 & 0.01 & 0.65 & 0.66 & 1 \\ 0.2 & 0.085 & 0.115 & 0.003 & 0.194 & 0.003 & 0115 & 0.085 & 0.2 \end{array}\right\}$$

Here also, taking the convex combinations of the relevant cluster members yields the following solution,

$$P^* = \left\{\begin{array}{ccccc} -1.00 & -0.655 & 0.00 & 0.655 & 1.00 \\ 0.200 & 0.200 & 0.200 & 0.200 & 0.200 \end{array}\right\}$$

Now we will focus on the results obtained for quartic regression model using the functions logistic CDF $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x, \delta) = \phi(\delta x)$ where $x = F$.

The number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the functions logistic CDF and normal CDF where $x = F$ are displayed in Table 4.6.

According to the results obtained in Table 4.5 and Table 4.6, it is noticed that the convergence rate of obtaining the optimal design increases when increasing the δ value. The D-optimal design space obtained for these two functions is as follows.

$$P^* = \left\{\begin{array}{ccccccc} -1.00 & -0.66 & -0.65 & 0.00 & 0.65 & 0.66 & 1.00 \\ 0.2 & 0.1 & 0.1 & 0.2 & 0.1 & 0.1 & 0.2 \end{array}\right\}$$

We can obtain the following optimal design by taking the convex combinations of the cluster members as earlier.

$$P^* = \left\{\begin{array}{ccccc} -1.00 & -0.655 & 0.00 & 0.655 & 1.00 \\ 0.200 & 0.200 & 0.200 & 0.200 & 0.200 \end{array}\right\}.$$

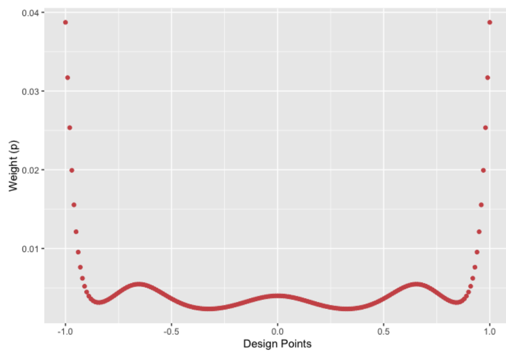
Table 4.6: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for quartic regression model with the use of clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
55	1	1	20	38	54	71
65	1	1	17	32	46	59
80	1	1	14	26	37	48
90	1	1	13	24	33	43
95	1	1	12	22	31	40
100	**1**	**1**	**12**	**21**	**30**	**38**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
50	1	1	13	25	34	44
55	1	1	12	22	31	40
60	1	1	12	21	29	37
75	1	1	10	17	23	29
85	1	1	9	15	20	25
90	**1**	**1**	**8**	**13**	**17**	**22**

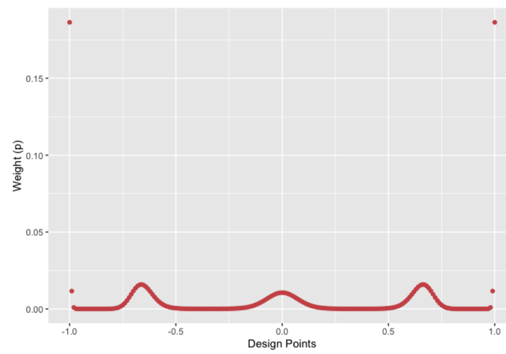
Graphical Interpretation

Next we will consider the graphical interpretation related to the construction of optimal design for the quartic regression model using D-optimality criterion. The function being considered in here is logistic cumulative density function and the δ value is 1.15. Figure 4.6 denotes the plots of weights versus design points for several number of iterations and at the optimum.

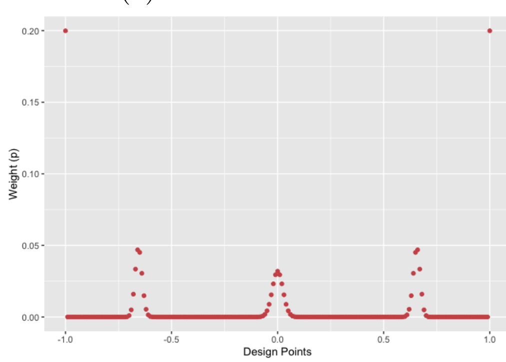
The plot 4.6 (a) shows a curve with four minimal and three maximal turning points. In addition to that, it can be seen there are two peak at $x = -1.00$ and $x = 1.00$ design points. As for the previous cubic and quadratic regression models, here also when increasing the number of iterations, the weights corresponding to minimal turning points are converging to zero. At the optimum, the weights corresponding to the design points -1.00, 0.00 and 1.00 have converged to 0.2. But, some of the weights are arranged as clusters between some design points. In the plot 4.6 (d), it can be seen clearly. That means, the design space for the quartic regression model has been obtained with 5 clusters of points.



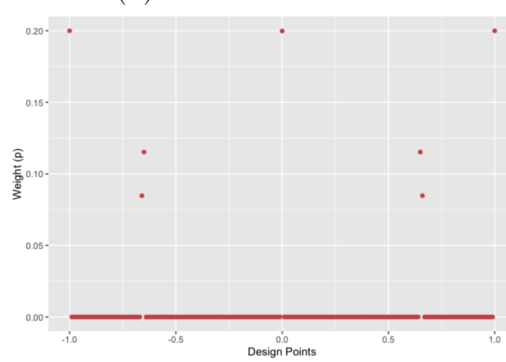
(a) After 10 iterations



(b) After 100 iterations



(c) After 1000 iterations

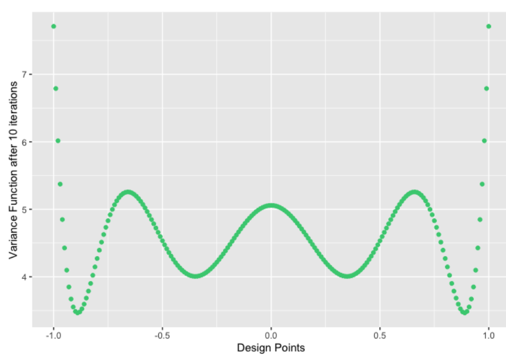


(d) At the optimum

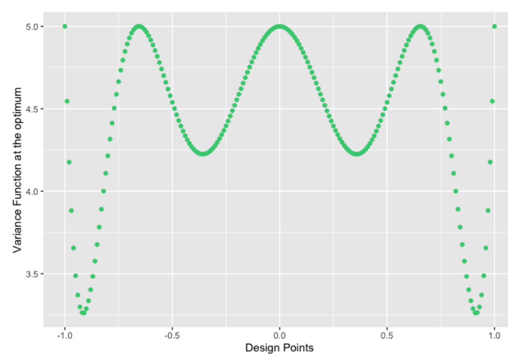
Figure 4.6: Weights vs design points for the quartic regression model

Figure 4.7 displays the plots of variance function versus design points for this optimal design construction. The plot 4.7 (a) depicts the curve after running the algorithm for 10 iterations. This curve includes four minimal and three maximal turning points. In addition to that there are two peak points at $x = 1.00$ and $x = -1.00$ design points. It can be seen that, at the optimum the variance function values corresponding to these design points have converged to 5.0. The variance function related to this quartic regression model is,

$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) \\ &= 5 - 3.8351x^2 + 21.7167x^4 - 38.72x^6 + 20.8385x^8 \end{aligned}$$



(a) After 10 iterations



(b) At the optimum

Figure 4.7: Variance function vs design points for the quartic regression model

4.5 Application of Clustering Approach to the D-optimal Design Construction for the Practical Model in Chemistry

In this section we are going to discuss about the application of clustering approach to the practical problem in chemistry that we mentioned in the previous chapter. The model is denoted as

$$E(y|x) \quad = \quad \theta_0 x + \theta_1 x^{1/2} + \theta_2 x^2.$$

As we have done earlier for the previous models, here also we considered three functions x^δ, logistic cumulative density function $e^{\delta x}/1+e^{\delta x}$ and normal cumulative density function $\phi(\delta x)$.

The same initial design space used in the traditional approach was considered. That means, a grid of 20 points equally spaced at the intervals of 0.01 between 0.0 and 0.2. The best result for the function x^δ was obtained when using $x = d$, the partial derivatives. But for the other two functions, the best results were obtained with the use of $x = F$, the directional derivatives.

The iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ are displayed in the Table 4.7. The D-optimal design was obtained as given below

$$P^* = \left\{ \begin{array}{cccc} 0.02 & 0.11 & 0.12 & 0.20 \\ 0.333 & 0.255 & 0.079 & 0.333 \end{array} \right\}$$

Table 4.7: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for the chemistry model with the use of clustering approach

$f(x,\delta) = x^{\delta}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
7.750	1	1	89	732	1381	2014
8.000	1	1	87	709	1338	1952
8.500	1	1	82	667	1260	1837
8.750	1	1	79	648	1224	1784
9.000	1	1	77	630	1190	1735
9.500	**1**	**1**	**73**	**597**	**1127**	**1648**

The total number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the functions logistic CDF and normal CDF where $x = F$ are displayed in Table 4.8.

D-optimal designs obtained for the chemistry model with the functions $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x,\delta) = \phi(\delta x)$ when $x = F$, using clustering approach are shown respectively below.

$$P^* = \left\{ \begin{array}{cccc} 0.02 & 0.11 & 0.12 & 0.20 \\ 0.333 & 0.240 & 0.094 & 0.333 \end{array} \right\}$$

$$P^* = \left\{ \begin{array}{cccc} 0.02 & 0.11 & 0.12 & 0.20 \\ 0.333 & 0.248 & 0.086 & 0.333 \end{array} \right\}$$

Graphical Interpretation

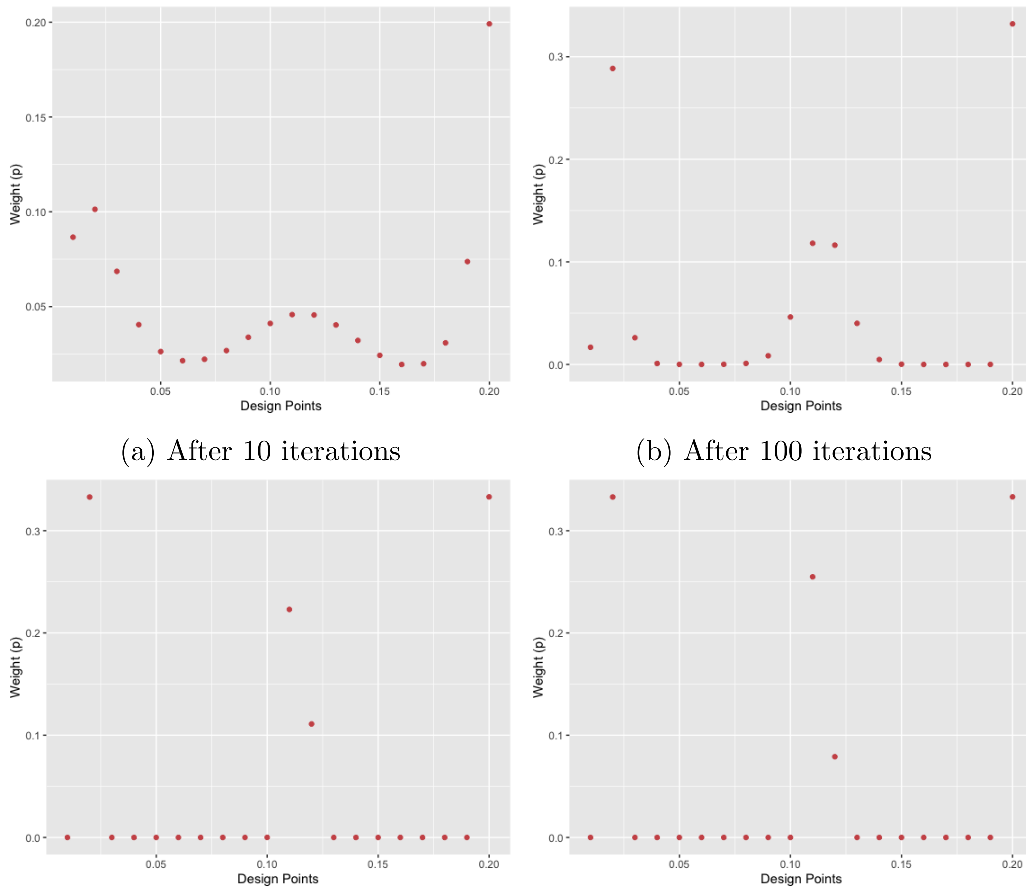
Let us consider the graphical display of weights and variance function related to the construction of D-optimal design for the chemistry model. The design

Table 4.8: Number of iterations need to achieve $\max_{1\leq j\leq J}\{F_j\} \leq 10^{-t}$ with $x = F$ for the chemistry model with the use of clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
50	1	1	16	90	178	264
60	1	1	13	75	148	219
70	1	1	12	65	127	187
75	1	1	11	60	119	175
80	1	1	11	57	111	163
90	**1**	**1**	**10**	**51**	**99**	**145**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
55	1	1	3	62	134	203
60	1	1	3	52	113	171
70	1	1	3	49	105	158
75	1	1	2	46	98	147
80	1	1	2	43	92	138
85	**1**	**1**	**2**	**40**	**86**	**130**

construction was done using the logistic CDF $e^{\delta x}/1 + e^{\delta x}$ where $x = d$. The design space being considered is 0.0 to 0.2. The δ value was taken as 0.45.

The plots of weights versus design points after several number of iterations and at the optimum are displayed in Figure 4.8. According to the plot 4.8 (a), it is noticeable that there are two maximal and two minimal turning points in the curve. Here also, the weights lie in the minimal turning points converged to zero when increasing the number of iterations. These weights at the minimal turning points converged to zero at the optimum as shown in the plot 4.8 (d). In this plot, the weights corresponding to the design points 0.02 and 0.20 have converged to 0.333 as in D-optimal design obtained earlier. The weights and the design points corresponding the D-optimal design for the chemistry model can be seen clearly in plot 4.8 (d).



(a) After 10 iterations

(b) After 100 iterations

(c) After 1000 iterations

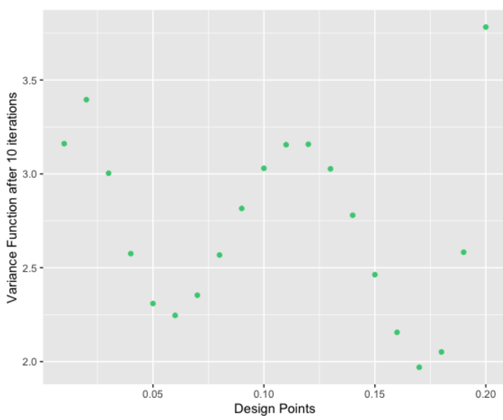
(d) At the optimum

Figure 4.8: Weights vs design points for the practical model in chemistry

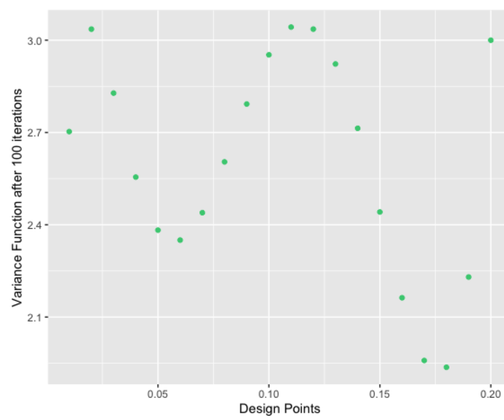
The plots of variance function versus design points after several number of iterations and at the optimum are denoted in the Figure 4.9. Here also in plot 4.9 (a), two minimal and two maximal turning points can be seen like in our weights versus design points plot for chemistry model. When increasing the number of iterations the variance function values around maximal turning points are converging to 3.00. In addition to that there is a peak point at the design point $x = 0.20$ which has converged to 3.00 at the optimum as in the plot 4.9 (c). The variance function related to the chemistry model is,

$$d(\underline{x}, p) = \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x})$$

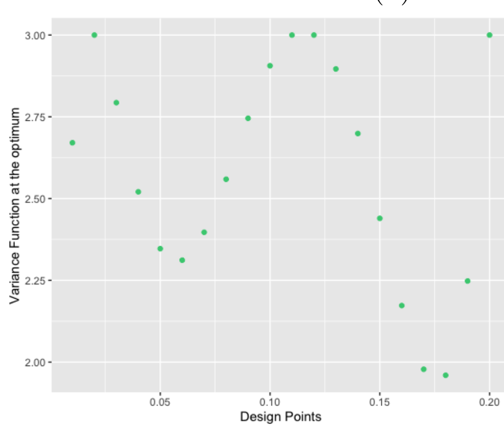
$$= 789.4313x + 16552.652x^2 + 1742.154x^{3/2} - 89578.72x^3 + 8706.886x^{5/4} + 131921.815x^4.$$



(a) After 10 iterations



(b) After 100 iterations



(c) At the optimum

Figure 4.9: Variance function vs design points for the practical model in chemistry

4.6 Application of Clustering Approach to Two Variable Regression Models

In this section we are going to discuss about the D-optimal design construction for the two variable models with the use of clustering approach. We will consider two models (model 3.6 and model 3.7) mentioned in Chapter 3. The functions being considered in the algorithm are $e^{\delta x}$, $e^{\delta x}/1 + e^{\delta x}$ and $\phi(\delta x)$. The total number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ will be displayed. Additionally, the convergence of the algorithm was improved by using directional derivatives of D-optimality criterion.

4.6.1 Model with Parameters θ_0, θ_1 and θ_2

We have constructed the optimal design for the reduced model below with parameters θ_0, θ_1 and θ_2 using clustering approach.

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

The design space was approximated by a grid of 441 (x_1, x_2) pairs where each variable takes values between -1.0 to 1.0 with 0.1 intervals. The best results were obtained for the function $f(x, \delta) = e^{\delta x}$ when $x = d$, partial derivatives. But for the logistic CDF and normal CDF, the best results were obtained when we replaced partial derivatives in the function $f(x, \delta)$ by the directional derivatives. The number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the function $f(x, \delta) = e^{\delta x}$ where

$x = d$ are displayed in Table 4.9. Moreover, the number of iterations for logistic CDF $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x, \delta) = \phi(\delta x)$ where $x = F$ are displayed in Table 4.10.

The δ values and the total number of iterations which were used to form clusters are

- $e^{\delta x}$ where $x = d$: $\delta = 0.8$, $n = 30$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.5$, $n = 80$
- Normal CDF where $x = F$: $\delta = 0.5$, $n = 30$

Table 4.9: Number of iterations need to achieve $\underset{1 \le j \le J}{max}\{F_j\} \le 10^{-t}$ with $x = d$ for the reduced model with parameters θ_0, θ_1 and θ_2 using clustering approach

$f(x, \delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
10	1	1	18	28	38	48
15	1	1	12	19	26	33
20	1	1	10	15	20	25
25	1	1	8	12	16	20
30	1	1	7	11	14	17
35	**1**	**1**	**7**	**9**	**12**	**15**

The D-optimal design obtained for this reduced regression model with the function $f(x, \delta) = e^{\delta x}$ when $x = d$, using clustering approach is as follows.

Table 4.10: Number of iterations need to achieve $\max_{1\leq j\leq J}\{F_j\} \leq 10^{-t}$ with $x = F$ for the reduced model with parameters θ_0, θ_1 and θ_2 using clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
8	1	1	23	33	42	52
10	1	1	18	26	33	41
15	1	1	12	17	22	26
18	1	1	10	14	18	21
20	1	1	9	13	16	19
25	**1**	**1**	**8**	**10**	**12**	**15**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
1.50	1	7	26	45	64	83
1.60	1	7	24	42	60	78
1.80	1	6	22	37	53	69
2.00	1	5	20	34	48	62
2.25	1	5	17	30	42	54
2.40	**1**	**5**	**15**	**27**	**39**	**51**

$$P^* = \begin{Bmatrix} -1.00 & 1.00 & -1.00 & 1.00 \\ -1.00 & -1.00 & 1.00 & 1.00 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{Bmatrix}$$

The D-optimal designs obtained for this model with the functions $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x,\delta) = \phi(\delta x)$ when $x = F$, using clustering approach is as follows.

$$P^* = \begin{Bmatrix} -1.00 & 1.00 & 1.00 & -1.00 & -1.00 & 1.00 \\ -1.00 & -1.00 & -0.90 & 0.90 & 1.00 & 1.00 \\ 0.25 & 0.23 & 0.02 & 0.02 & 0.23 & 0.25 \end{Bmatrix}$$

It is noticeable that, the support points consist of four clusters of points which is either of neighboring points or a single point. Since this is a solution for a discretized design space, the solution for a continuous design space should be a 4-pair design. Therefore, by taking the convex combinations of the relevant cluster elements, the solution below can be obtained.

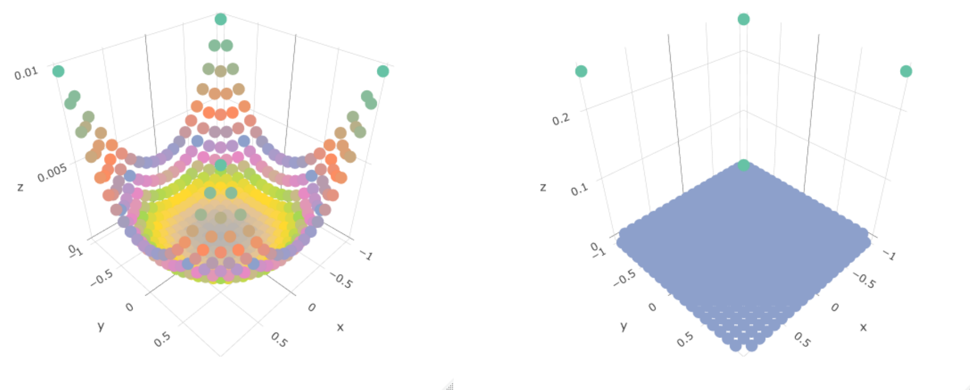
$$P^* = \begin{Bmatrix} -1.00 & 1.00 & -1.00 & 1.00 \\ -1.00 & -1.00 & 1.00 & 1.00 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{Bmatrix}$$

The best choices of δ for the above three functions are,

- 35 for $f(x, \delta) = e^{\delta x}$ when $x = d$
- 25 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ when $x = F$
- 2.40 for $f(x, \delta) = \Phi(\delta x)$ when $x = F$

Graphical Interpretation

Next we will consider the graphical display of weights and variance function versus design points for the above model. The function being considered was logistic CDF $e^{\delta x}/(1+e^{\delta x})$ and δ was taken as 0.50. Figure 4.10 denotes the plots of weights versus design points after running the algorithm for 10 iterations and at the optimum respectively.



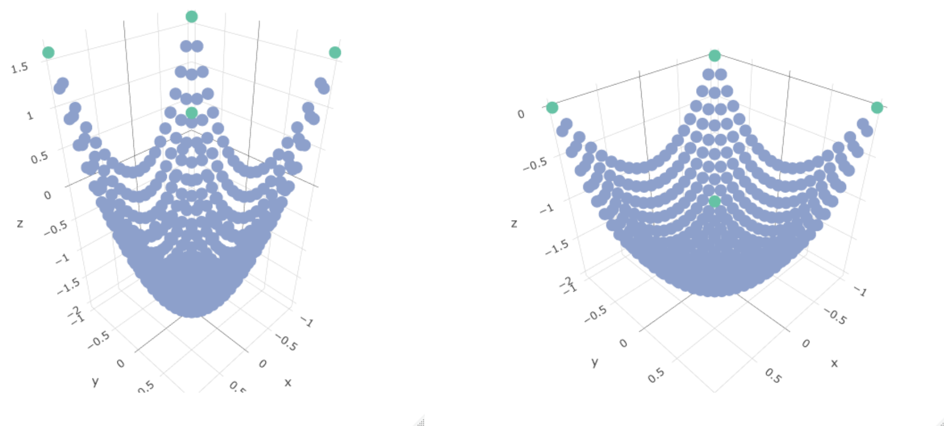
(a) After 10 iterations (b) At the optimum

Figure 4.10: Weights vs design points for the two variable model with θ_0, θ_1 and θ_2 as the parameters of interest

According to the plot 4.10 (b), the D-optimal design have weights only at the corners $(-1,-1)$, $(-1,1)$, $(1,1)$ and $(1,-1)$. Each of them take a weight of 0.25. All the other weights have converged to zero at the optimum.

Figure 4.11 depicts the plots of variance function versus design points for the same model after 10 iterations and at the optimum. It is noticeable that at the optimum (plot 4.11 (b)) the variance function values corresponding to the design pairs $(-1,-1)$, $(-1,1)$, $(1,1)$ and $(1,-1)$ have converged to 0. The variance function related to this quartic regression model is,

$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) \\ &= 1 + x_1^2 + x_1^2. \end{aligned}$$



(a) After 10 iterations (b) At the optimum

Figure 4.11: Variance function vs design points for the two variable model with θ_0, θ_1 and θ_2 as the parameters of interest

4.6.2 Model with Parameters θ_0, θ_1, θ_2, θ_3 and θ_4

Here we are going to focus on the construction of optimal design for the model with parameters θ_0, θ_1, θ_2, θ_3 and θ_4 using clustering approach.

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2$$

The iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the function $f(x, \delta) = e^{\delta x}$ where $x = d$ are displayed in Table 4.11. Furthermore, the number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the functions logistic CDF $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x, \delta) = \phi(\delta x)$ where $x = F$ are displayed in Table 4.12.

The δ values and total number of iterations used in forming clusters are

- $e^{\delta x}$ where $x = d$: $\delta = 0.27$, $n = 70$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.50$, $n = 70$
- Normal CDF where $x = F$: $\delta = 0.60$, $n = 55$

Table 4.11: Number of iterations need to achieve $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-t}$ with $x = d$ for the reduced model with parameters θ_0, θ_1, θ_2, θ_3 and θ_4 using clustering approach

$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
55	1	3	8	12	17	24
60	1	3	7	11	16	23
70	1	3	6	10	15	21
75	1	3	6	10	14	20
80	1	3	6	9	14	20
90	**1**	**2**	**5**	**8**	**13**	**19**

The D-optimal design obtained for this regression model is as follows.

$$P^* = \begin{Bmatrix} -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 \\ -1.00 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 1.00 \\ 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 & 0.111 \end{Bmatrix}$$

The best choices for the δ for each of the functions are as follows.

- 90 for $f(x,\delta) = e^{\delta x}$ when $x = d$
- 90 for $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ when $x = F$

Table 4.12: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the reduced model with parameters θ_0, θ_1, θ_2, θ_3 and θ_4 using clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
50	1	6	17	28	40	54
65	1	5	15	25	40	56
70	1	5	14	25	41	58
74	1	5	13	25	43	61
80	1	4	13	25	46	67
90	**1**	**4**	**12**	**27**	**55**	**83**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
45	1	4	24	50	72	93
50	1	4	21	41	59	76
68	1	3	15	24	33	42
75	1	3	13	21	28	35
80	1	3	12	19	25	31
90	**1**	**3**	**10**	**15**	**20**	**25**

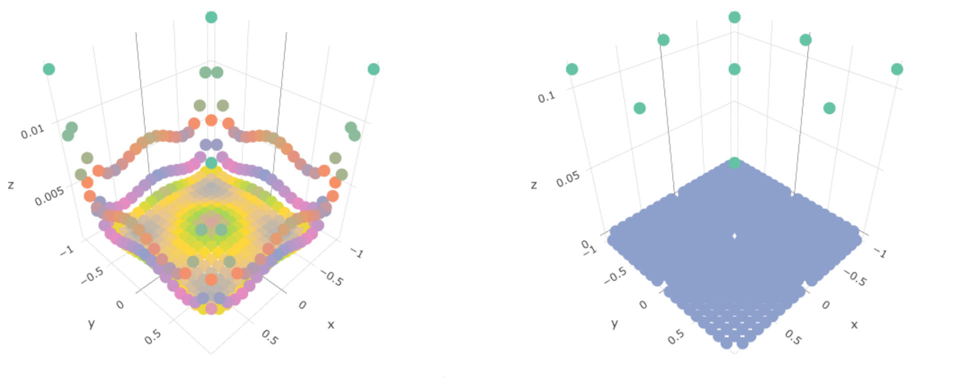
- 90 for $f(x, \delta) = \Phi(\delta x)$ when $x = F$

According to the results, it can be seen that when increasing the δ value, the number of iterations needed to achieve $max\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ have been decreased. This will increase the convergence rate of the optimal design.

Graphical Interpretation

Let us consider the graphical display of weights and variance function versus design points for this model. The function being considered was logistic CDF

$e^{\delta x}/(1+e^{\delta x})$ and δ was taken as 0.27. Figure 4.12 denotes the plots of weights versus design points after running the algorithm for 10 iterations and at the optimum.



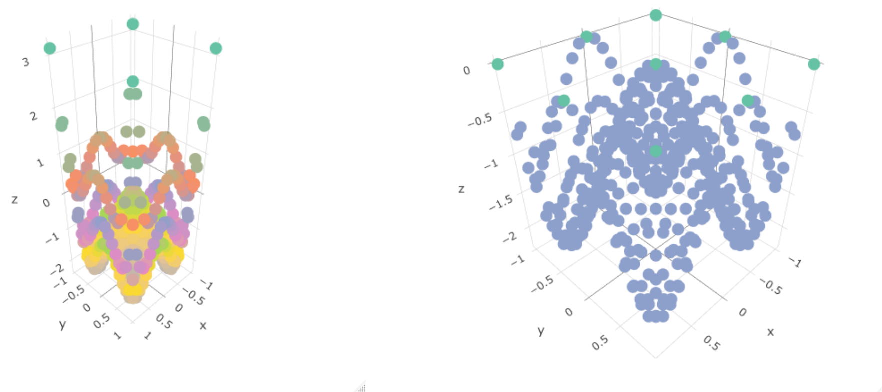
(a) After 10 iterations (b) At the optimum

Figure 4.12: Weights vs design points for the two variable model with $\theta_0, \theta_1, \theta_2, \theta_3$ and θ_4 as the parameters of interest

It is noticeable that in the D-optimal design the weights corresponding to the pairs $(-1,-1)$, $(0,-1)$, $(1,-1)$, $(-1,0)$, $(0,0)$, $(1,0)$, $(-1,1)$, $(0,1)$, $(1,1)$ have converged to 0.111. All the other weights have converged to zero at the optimum.

Moreover, the Figure 4.13 depicts the plots of variance function versus design points for this model after 10 iterations and at the optimum. It is noticeable that at the optimum (plot 4.13 (b)) the variance function values corresponding to the nine design pairs we mentioned earlier have converged to 0. The variance function related to this quartic regression model is,

$$
\begin{aligned}
d(\underline{x}, p) &= \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x}) \\
&= 5 - 0.0008x_1^2x_2^2 + 4.5x_2^4 + 4.5x_1^4 - 4.5x_2^2 - 4.5x_1^2
\end{aligned}
$$



(a) After 10 iterations

(b) At the optimum

Figure 4.13: Variance function vs design points for the two variable model with $\theta_0, \theta_1, \theta_2, \theta_3$ and θ_4 as the parameters of interest

4.6.3 Full Model

Next we are going to focus on the construction of D-optimal design for the two variable full model using clustering approach. The model is given by

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 = \underline{v}_x^T \underline{\theta}.$$

The best results were obtained for the function $f(x, \delta) = e^{\delta x}$ when $x = d$, partial derivatives. But for the logistic CDF and normal CDF, the best results were obtained when we replaced partial derivatives in the function $f(x, \delta)$ by the directional derivatives. Results are given in Tables 4.13 and 4.14.

The δ values and the total number of iterations which were used to form clusters are,

- $e^{\delta x}$ where $x = d$: $\delta = 0.300$, $n = 150$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.725$, $n = 120$
- $\Phi(\delta x)$ where $x = F$: $\delta = 0.450$, $n = 120$

The optimal design is obtained as

$$P^* = \left\{ \begin{matrix} -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 \\ -1.00 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 1.00 \\ 0.146 & 0.080 & 0.146 & 0.086 & 0.082 & 0.086 & 0.146 & 0.080 & 0.146 \end{matrix} \right\}$$

The D-optimal designs obtained for this model with the functions $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ and normal CDF $f(x, \delta) = \phi(\delta x)$ when $x = F$, using clustering

Table 4.13: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for the two variable full model using clustering approach

$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
60	1	1	12	38	60	83
65	1	1	12	35	55	75
74	1	1	11	30	48	65
80	1	1	11	28	44	59
85	1	1	10	26	41	55
90	**1**	**1**	**10**	**25**	**38**	**51**

Table 4.14: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the two variable full model using clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
64	1	1	17	67	113	158
70	1	1	16	60	101	141
75	1	1	16	56	92	129
80	1	1	15	52	85	118
86	1	1	15	47	78	108
90	**1**	**1**	**14**	**45**	**73**	**101**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
55	1	1	15	46	76	105
65	1	1	13	38	61	84
70	1	1	13	34	55	76
78	1	1	12	30	48	65
86	1	1	11	27	42	57
92	**1**	**1**	**10**	**24**	**38**	**51**

approach is as follows.

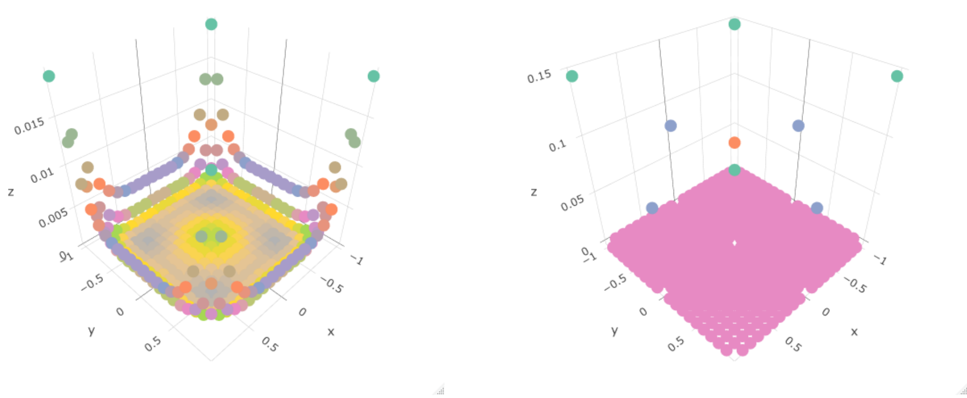
$$P^* = \left\{\begin{array}{ccccccccc} -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 \\ -1.00 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 1.00 \\ 0.146 & 0.080 & 0.146 & 0.087 & 0.080 & 0.087 & 0.146 & 0.080 & 0.146 \end{array}\right\}$$

The best choices of δ for the above three functions are,

- 90 for $f(x,\delta) = e^{\delta x}$ when $x = d$
- 90 for $f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$ when $x = F$
- 92 for $f(x,\delta) = \Phi(\delta x)$ when $x = F$.

Graphical Interpretation

Consider the graphical display of weights and variance function versus design points for the full model (3.5). The function being considered was logistic CDF $e^{\delta x}/(1+e^{\delta x})$ and δ was taken as 0.20. Figure 4.14 gives the plots of weights versus design points after running the algorithm for 10 iterations and at the optimum. At the optimum (plot 4.14 (b)), D-optimal design have same weights on the pairs $(-1,-1)$, $(0,-1)$, $(1,-1)$, $(-1,0)$, $(0,0)$, $(1,0)$, $(-1,1)$, $(0,1)$, $(1,1)$ which are similar to the weights we obtained before.



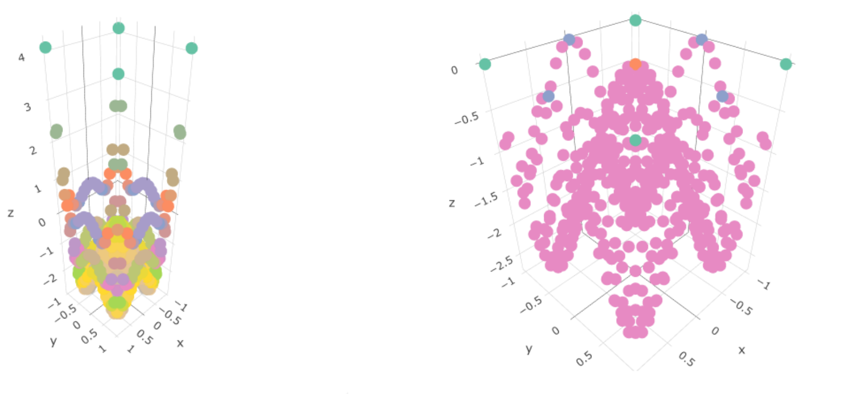
(a) After 10 iterations (b) At the optimum

Figure 4.14: Weights vs design points for the two variable full model

Figure 4.15 depicts the plots of variance function versus design points for the same model after 10 iterations and at the optimum. The variance function related to this quartic regression model is,

$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x}) \\ &= 6 - 4.5936x_1^2x_2^2 + 5.3801x_1^4 + 5.3801x_2^4 - 5.38x_2^2 - 5.38x_1^2 \end{aligned}$$

In all of the results we see that after combining the clustering approach with the properties of the directional derivatives of the criterion, the convergence is improved a lot.



(a) After 10 iterations (b) At the optimum

Figure 4.15: Variance function vs design points for the two variable full model

Chapter 5

Construction of D_S-optimal Designs using both Analytic and Algorithmic Approaches

We are going to focus on D_s optimality criterion which is a special case of D_A-optimality from this chapter onward. D_s optimality is quite important when we are interested in a subset of parameters.

Suppose we are interested in s parameters $\theta_1, \theta_2, ..., \theta_s$ of $\underline{\theta} \in \Theta$. In D_A-optimality, suppose that the A matrix is expressed in terms of a $s \times s$ identity matrix (I_s) and $s \times (k-s)$ zero matrix(O) such that $A = [I_s : O]$. This converts to D_s optimality. Thus, with the above choice of A, $M(p)$

matrix can be partitioned in the following manner

$$M(p) \;=\; \begin{bmatrix} M_{11}^{s\times s} & M_{12}^{s\times(k-s)} \\ M_{12}^{T} & M_{22}^{(k-s)\times(k-s)} \end{bmatrix}.$$

Then in D_s optimality the matrix $(AM^{-1}(p)A^T)^{-1}$ is expressed as $M_{11} - M_{12}M_{22}^{-1}M_{12}^T$ (Rhode (1965)). So the criterion can be written as

$$\phi_{D_s}(p) \;=\; log\,det\{M_{11} - M_{12}M_{22}^{-1}M_{12}^T\}. \tag{5.1}$$

This is known as D_s-optimality criterion as in Karlin and Studden (1966), Atwood (1969) and Silvey (1980). This criterion has similar properties as D-optimality criterion since this is a special case of D_A-optimality.

5.1 Analytic Solutions for the Construction of D_s-optimal Designs

In this section we are going to discuss some analytical solutions for polynomial regression models. If we have the support points for a given model, we can derive the optimal solution for D_s-optimality criterion.

First we consider the analytical solution to quadratic regression model taking θ_2 as the parameter of interest. So our model can be rearranged to the following format.

$$E(y|x) = \theta_2 x^2 + \theta_1 x + \theta_0$$

The D_s-optimal design is,

$$P^* = \left\{ \begin{array}{ccc} -1.00 & 0.00 & 1.00 \\ 0.25 & 0.50 & 0.25 \end{array} \right\}$$

Proof:

Let the weights corresponding to support points -1.00, 0.00 and 1.00 be p_1, p_2 and p_3 respectively. V_j^T and A matrices for the model are expressed as

$V_j^T = [x_j^2 \quad x_j \quad 1]; \quad j = 1, 2, 3$

$$A = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}.$$

- First we find the information matrix $M(p)$ for the model.

$$\begin{aligned} M(p) &= \sum_{j=1}^{4} p_j v_j v_j^T \\ &= \sum_{j=1}^{3} p_j \begin{bmatrix} x_j^4 & x_j^3 & x_j^2 \\ x_j^3 & x_j^2 & x_j \\ x_j^2 & x_j & 1 \end{bmatrix} \end{aligned}$$

- Then by substituting $x_1 = -1.00$, $x_2 = 0.00$ and $x_3 = 1.00$ in $M(p)$, a 3×3 $M(p)$ matrix is obtained. The elements of this matrix include expressions in terms of p_1 and p_3. As $p_1 + p_2 + p_3 = 1$, $p_2 = 1 - p_1 - p_3$.

$$M(p) = \begin{bmatrix} p_3 + p_1 & p_3 - p_1 & p_3 + p_1 \\ p_3 - p_1 & p_3 + p_1 & p_3 - p_1 \\ p_3 + p_1 & p_3 - p_1 & 1 \end{bmatrix} \tag{5.2}$$

- Next, we partition the $M(p)$ matrix in the way we discussed earlier.

$$M_{11} = \begin{bmatrix} p_3 + p_1 \end{bmatrix}$$

$$M_{12} = \begin{bmatrix} p_3 - p_1 & p_3 + p_1 \end{bmatrix}$$

$$M_{21} = \begin{bmatrix} p_3 - p_1 \\ p_3 + p_1 \end{bmatrix}$$

$$M_{22} = \begin{bmatrix} p_3 + p_1 & p_3 - p_1 \\ p_3 - p_1 & 1 \end{bmatrix}$$

- Then, we obtain D_s-optimality criterion using the equation (5.1).

$$\phi_{D_s}(p) = log\, det\{M_{11} - M_{12} M_{22}^{-1} M_{12}^T\}$$

- We calculate partial derivatives of the criterion function obtained, with respect to p_1 and p_3.

$$\frac{\partial \phi_{D_s}(p)}{\partial p_1} = -\frac{p_3\,(3p_1 + p_3 - 1)\,(p_1 - p_3 + 1)}{(p_1 + p_3 - 1)\,p_1\,(p_1^2 + (-2p_3 - 1)\,p_1 + p_3^2 - p_3)}$$

$$\frac{\partial \phi_{D_s}(p)}{\partial p_1} = \frac{p_3\,(p_1+3p_3-1)\,(p_1-p_3-1)}{p_3\,(p_1+p_3-1)\,(p_1^2+(-2p_3-1)\,p_1+p_3^2-p_3)}$$

- Then, we solve following three algebraic equations and calculate the values of p_1, p_2, and p_3.

$$\frac{\partial \phi_{D_s}(p)}{\partial p_1} = 0$$

$$\frac{\partial \phi_{D_s}(p)}{\partial p_3} = 0$$

$$p_1+p_2+p_3 = 1$$

- We obtain the answers: $p_1 = 0.25$, $p_2 = 0.50$ and $p_3 = 0.25$.

- Hence, the D_s-optimal design is,

$$P^* = \left\{ \begin{array}{ccc} -1.00 & 0.00 & 1.00 \\ 0.25 & 0.50 & 0.25 \end{array} \right\}$$

Next we will consider the application of D_s-optimality criterion to the practical model in chemistry we discussed before. Suppose we are taking θ_1 and θ_2 as the parameters of interest. So, the model can be written in the following format.

$$E(y|x) = \theta_1 x^{1/2} + \theta_2 x^2 + \theta_0 x$$

Suppose that we know the support points of this D_s-optimal design are 0.02, 0.10 and 0.20. Then the optimal design is

$$P^* = \left\{ \begin{array}{ccc} 0.02 & 0.10 & 0.20 \\ 0.49 & 0.34 & 0.17 \end{array} \right\}$$

Proof:

Let the weights corresponding to support points 0.02, 0.10 and 0.20 be p_1, p_2 and p_3. The matrices A and V_j^T can be expressed as

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$V_j^T = [x_j^{1/2} \quad x_j^2 \quad x_j]; \quad j = 1, 2, 3.$$

- First we obtain information matrix $M(p)$ for this model.

$$\begin{aligned} M(p) &= \sum_{j=1}^{3} p_j v_j v_j^T \\ &= \sum_{j=1}^{3} p_j \begin{bmatrix} x_j & x_j^{5/2} & x_j^{3/2} \\ x_j^{5/2} & x_j^4 & x_j^3 \\ x_j^{3/2} & x_j^3 & x_j^2 \end{bmatrix} \end{aligned}$$

- Then by substituting $x_1 = 0.02$, $x_2 = 0.10$ and $x_3 = 0.20$ in $M(p)$ the following is obtained as the $M(p)$ matrix.

$$\begin{bmatrix} 0.02p_1 + 0.1p_2 + 0.2p_3 & 0.0032p_2 + 0.0179p_3 & 0.0028p_1 + 0.0316p_2 + 0.0894p_3 \\ 0.0032p_2 + 0.0179p_3 & 0.0001p_2 + 0.0016p_3 & 0.001p_2 + 0.008p_3 \\ 0.0028p_1 + 0.0316p_2 + 0.0894p_3 & 0.001p_2 + 0.008p_3 & 0.0004p_1 + 0.01p_2 + 0.04p_3 \end{bmatrix}.$$

- Next, we partition the $M(p)$ matrix as we discussed in earlier section. So the results are,

$$M_{11} = \begin{bmatrix} 0.02p_1 + 0.1p_2 + 0.2p_3 & 0.0032p_2 + 0.0179p_3 \\ 0.0032p_2 + 0.0179p_3 & 0.0001p_2 + 0.0016p_3 \end{bmatrix}$$

$$M_{12} = \begin{bmatrix} 0.0028p_1 + 0.0316p_2 + 0.0894p_3 \\ 0.001p_2 + 0.008p_3 \end{bmatrix}$$

$$M_{21} = \begin{bmatrix} 0.0028p_1 + 0.0316p_2 + 0.0894p_3 & 0.001p_2 + 0.008p_3 \end{bmatrix}$$

$$M_{22} = \begin{bmatrix} 0.0004p_1 + 0.01p_2 + 0.04p_3 \end{bmatrix}$$

- Next we find the solution of $M_{11} - M_{12}M_{22}^{-1}M_{12}^T$. It is a 2×2 matrix.
- Then, we obtain D_s-optimality criterion:

$$\phi_{D_s}(p) = log\, det\{M_{11} - M_{12}M_{22}^{-1}M_{12}^T\}$$

- We obtain partial derivatives of the criterion function obtained, with respect to p_1, p_2 and p_3.

- Then, we have following four algebraic equations which can be solved in order to get the answers for p_1, p_2 and p_3.

$\frac{\partial \phi_{D_s}(p)}{\partial p_1} = 0$

$(-(2.0(10^74) * ((1.0(10^{-42}) * (-2.583(10^{74}) * p_1^3 + 1.13(10^{89}) * p_1^2 * p_2 + 2.047(10^{90}) * p_1^2 * p_3 + 4.549(10^{89}) * p_1 * p_2^2 - 1.606(10^{102}) * p_1 * p_2 * p_3 + 1.227(10^{92}) * p_1 * p_3^2 + 1.676(10^{80}) * p_2^3 - 1.912(10^{92}) * p_2^2 * p_3 - 3.894(10^{91}) * p_2 * p_3^2 + 3.965(10^{81}) * p_3^3))/(2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3)^2 - (5.0(10^{-75}) * (-7.748(10^{74}) * p_1^2 + 2.259(10^{89}) * p_1 * p_2 + 4.094(10^{90}) * p_1 * p_3 + 4.549(10^{89}) * p_2^2 - 1.606(10^{102}) * p_2 * p_3 + 1.227(10^{92}) * p_3^2))/(2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3)) * (2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3))/(-2.583(10^{74}) * p_1^3 + 1.13(10^{89}) * p_1^2 * p_2 + 2.047(10^{90}) * p_1^2 * p_3 + 4.549(10^{89}) * p_1 * p_2^2 - 1.606(10^{102}) * p_1 * p_2 * p_3 + 1.227(10^{92}) * p_1 * p_3^2 + 1.676(10^{80}) * p_2^3 - 1.912(10^{92}) * p_2^2 * p_3 - 3.894(10^{91}) * p_2 * p_3^2 + 3.965(10^{81}) * p_3^3)) = 0.0$

$\frac{\partial \phi_{D_s}(p)}{\partial p_2} = 0$

$(2.0(10^{74}) * ((5.0(10^{-75}) * (1.13(10^{89}) * p_1^2 + 9.098(10^{89}) * p_1 * p_2 - 1.606(10^{102}) * p_1 * p_3 + 5.029(10^{80}) * p_2^2 - 3.825(10^{92}) * p_2 * p_3 - 3.894(10^{91}) * p_3^2))/(2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3) - (2.5(10^{-41}) * (-2.583(10^{74}) * p_1^3 + 1.13(10^{89}) * p_1^2 * p_2 + 2.047(10^{90}) * p_1^2 * p_3 + 4.549(10^{89}) * p_1 * p_2^2 - 1.606(10^{102}) * p_1 * p_2 * p_3 + 1.227(e+92) * p_1 * p_3^2 + 1.676(10^{80}) * p_2^3 - 1.912(10^{92}) * p_2^2 * p_3 - 3.894(10^{91}) * p_2 * p_3^2 + 3.965(10^{81}) * p_3^3))/(2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3)^2) * (2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 +$

$2.0(10^{34}) * p_3))/(-2.583(10^{74}) * p_1^3 + 1.13(10^{89}) * p_1^2 * p_2 + 2.047(10^{90}) * p_1^2 * p_3 + 4.549(10^{89}) * p_1 * p_2^2 - 1.606(10^{102}) * p_1 * p_2 * p_3 + 1.227(10^{92}) * p_1 * p_3^2 + 1.676(10^e + 80) * p_2^3 - 1.912(10^{92}) * p_2^2 * p_3 - 3.894(10^{91}) * p_2 * p_3^2 + 3.965(10^{81}) * p_3^3) = 0.0$

$\frac{\partial \phi_{D_s}(p)}{\partial p_3} = 0$

$-(2.0(10^{74}) * ((5.0(10^{-75}) * (-2.047(10^{90}) * p_1^2 + 1.606(10^{102}) * p_1 * p_2 - 2.454(10^{92}) * p_1 * p_3 + 1.912(10^{92}) * p_2^2 + 7.789(10^{91}) * p_2 * p_3 - 1.19(10^{82}) * p_3^2))/(2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3) + (1.0(10^{-40}) * (-2.583(10^{74}) * p_1^3 + 1.13(10^{89}) * p_1^2 * p_2 + 2.047(10^{90}) * p_1^2 * p_3 + 4.549(10^{89}) * p_1 * p_2^2 - 1.606(10^{102}) * p_1 * p_2 * p_3 + 1.227(10^{92}) * p_1 * p_3^2 + 1.676(10^{80}) * p_2^3 - 1.912(10^{92}) * p_2^2 * p_3 - 3.894(10^{91}) * p_2 * p_3^2 + 3.965(10^{81}) * p_3^3))/(2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3)^2) * (2.0(10^{32}) * p_1 + 5.0(10^{33}) * p_2 + 2.0(10^{34}) * p_3))/(-2.583(10^{74}) * p_1^3 + 1.13(10^{89}) * p_1^2 * p_2 + 2.047(10^{90}) * p_1^2 * p_3 + 4.549(10^{89}) * p_1 * p_2^2 - 1.606(10^{102}) * p_1 * p_2 * p_3 + 1.227(10^{92}) * p_1 * p_3^2 + 1.676(10^{80}) * p_2^3 - 1.912(10^{92}) * p_2^2 * p_3 - 3.894(10^{91}) * p_2 * p_3^2 + 3.965(10^{81}) * p_3^3) = 0.0$

$p_1 + p_2 + p_3 = 1$

- These equations are too complex to solve by hand. Therefore, we used MATLAB software for the expressions. The answers are $p_1 = 0.49$, $p_2 = 0.34$ and $p_3 = 0.17$.

- The optimal design is,

$$P^* = \left\{ \begin{array}{ccc} 0.02 & 0.10 & 0.20 \\ 0.49 & 0.34 & 0.17 \end{array} \right\}.$$

5.2 Algorithmic Approach for the Construction of D_s-optimal Designs

We have constructed D_s-optimal designs for some regression models analytically. Now we will use the multiplicative algorithm. Here also we use the functions such as $x^\delta, e^{\delta x}, e^{\delta x}/(1+e^{\delta x})$ and $\Phi(\delta x)$.

Moreover, to interpret the results graphically the plots of weights and variance function versus design points are used. The standardized variance for a D_s-optimal design can be expressed in the following manner.

$$d(\underline{x}, p) = \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x}) - \underline{f}_2^T(\underline{x})M_{22}^{-1}(p)\underline{f}_2(\underline{x})$$

Consider the partitioned information matrix $M(p)$ we mentioned earlier. The matrix M_{22} is the lower right corner part of the information matrix which does not include the information about the parameters of interest. Furthermore, $\underline{f}_2^T(\underline{x})$ denotes the vector including the design variable terms corresponding to the parameters which are not of our interest.

5.2.1 Cubic Regression Model with θ_1 and θ_3 as the Parameters of Interest

For the cubic regression model, consider the polynomial regression model (3.1) in chapter 3 where $k = 4$. Then $\underline{v}_x = (1, x, x^2, x^3)^T, x \in [-1, 1]$, $\underline{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)^T$. First, the D_s-optimal designs have been constructed for the parameters θ_1 and θ_3 using $x = d$, the partial derivatives of D_s-optimality criterion in the function. Then, the model can be rearranged as below.

$$E(Y|x) = \theta_1 x + \theta_3 x^3 + \theta_0 + \theta_2 x^2$$

The functions being considered are $e^{\delta x}, e^{\delta x}/(1 + e^{\delta x})$ and normal CDF. The number of iterations needed to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ for each of these functions are recorded in Table 5.1. The design space was approximated by a grid of 201 points equally spaced between -1.00 and 1.00 with 0.01. Then, we will see the convergence rate by replacing partial derivatives with directional derivatives (Table 5.2). Regardless of the function, the support points and the corresponding weights converged to the same optimal design shown below

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.58 & 0.58 & 1.00 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{array} \right\}.$$

Table 5.1: Number of iterations need to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for cubic regression model with θ_1 and θ_3 as the parameters of interest using D_s-optimality criterion

$f(x, \delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.250	22	196	1795	10141	24058	38286
0.300	18	164	1496	8451	20049	31905
0.350	16	140	1282	7244	17185	27348
0.400	14	123	1122	6339	15037	23930
0.450	12	109	997	5635	13367	21271
0.500	**11**	**98**	**897**	**5071**	**12030**	**19144**
$f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.350	49	425	3868	21859	51821	82446
0.400	47	399	3624	20475	48535	77215
0.500	45	369	3343	18891	44769	71217
0.550	45	361	3274	18498	43833	69725
0.600	**45**	**358**	**3239**	**18299**	**43355**	**68962**
0.700	47	359	3250	18361	43492	69175
$f(x, \delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.250	46	388	3530	19947	47284	75225
0.300	44	360	3263	18441	43704	69524
0.350	43	344	3119	17626	41763	66431
0.400	**44**	**338**	**3060**	**17291**	**40961**	**65150**
0.450	45	339	3067	17330	41046	65279
0.500	48	347	3131	17693	41895	66625

5.2.2 Quartic Regression Model with θ_3 as the Parameter of Interest

For the quartic regression model consider the model 3.1 in chapter 3, where $k = 5$. So, we have $\underline{v}_x = (1, x, x^2, x^3, x^4)^T, x \in [-1, 1]$, $\underline{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)^T$.

Table 5.2: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for cubic regression model with θ_1 and θ_3 as the parameters of interest using D_s-optimality criterion

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.350	32	282	2567	14505	34389	54713
0.400	29	247	2246	12695	30094	47878
0.500	24	198	1798	10161	24081	38303
0.550	23	181	1635	9240	21895	34828
0.600	21	166	1499	8472	20073	31928
0.700	**19**	**143**	**1286**	**7266**	**17209**	**27371**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.250	29	248	2252	12728	30173	48004
0.300	25	207	1878	10611	25149	40008
0.350	22	178	1610	9099	21561	34296
0.400	20	156	1409	7965	18869	30013
0.450	18	139	1253	7083	16776	266681
0.500	**17**	**125**	**1128**	**6377**	**15101**	**24016**

We have constructed D_s-optimal design for this model considering θ_3 as the only parameter of interest. So, the model will be,

$$E(Y|x) = \theta_3 x^3 + \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_4 x^4$$

Here also we considered the three functions $e^{\delta x}, e^{\delta x}/(1+e^{\delta x})$ and normal CDF where $x = d$, the partial derivatives of D_s-optimality criterion. The results are recorded in the Table 5.3 and Table 5.4. Eventually, the support points and the corresponding weights converged to the same optimal design shown below with different functions and derivatives.

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.50 & 0.50 & 1.00 \\ 0.1667 & 0.3333 & 0.3333 & 0.1667 \end{array} \right\}$$

Table 5.3: Number of iterations need to achieve $\underset{1\le j\le J}{max}\{F_j\} \le 10^{-t}$ with $x = d$ for quartic regression model with θ_3 as the parameter of interest using D_s-optimality criterion

$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.200	25	243	2488	12865	22788	32418
0.300	17	163	1659	8577	15192	21612
0.400	13	122	1245	6433	11394	16209
0.500	11	98	996	5146	9115	12967
0.600	9	81	830	4289	7596	10806
0.700	**8**	**69**	**711**	**3676**	**6510**	**9262**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.500	30	260	2640	13635	24147	34348
0.650	27	221	2237	11547	20447	29084
0.800	26	199	2012	10381	18380	26144
0.900	25	190	1919	9898	17524	24925
1.000	25	184	1857	9575	16952	24110
1.200	**26**	**179**	**1799**	**9272**	**16414**	**23344**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.200	39	362	3689	19061	33759	48022
0.500	25	194	1960	10113	17907	25470
0.650	25	175	1765	9103	16117	22923
0.700	25	172	1732	8931	15811	22488
0.800	**26**	**169**	**1703**	**8777**	**15535**	**22093**
0.950	27	171	1717	8848	15660	22271

Table 5.4: Number of iterations need to achieve $\max_{1\le j\le J}\{F_j\} \le 10^{-t}$ with $x = F$ for quartic regression model with θ_3 as the parameter of interest using D_s-optimality criterion

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.500	23	196	1993	10295	18233	25936
0.650	19	151	1534	7920	14026	19951
0.800	16	123	1247	6436	11397	16211
0.900	15	110	1109	5721	10131	14410
1.000	14	99	998	5150	9118	12969
1.200	**12**	**83**	**832**	**4292**	**7599**	**10808**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.200	33	306	3121	16127	28562	40631
0.500	16	124	1250	6453	11427	16254
0.650	13	96	962	4965	8791	12503
0.700	13	89	894	4610	8163	11610
0.800	12	78	783	4035	7143	10159
0.950	**10**	**66**	**660**	**3398**	**6015**	**8555**

5.2.3 Quartic Regression Model with θ_2 and θ_4 as the Parameters of Interest

For the quartic regression model, consider the model 3.1 in chapter 3, where $k = 5$. So, we have $\underline{v}_x = (1, x, x^2, x^3, x^4)^T, x \in [-1, 1]$, $\underline{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)^T$. The D_s-optimal design was constructed for this model considering θ_2 and θ_4 as the parameters of interest. The rearranged model will be,

$$E(Y|x) = \theta_2 x^2 + \theta_4 x^4 + \theta_0 + \theta_1 x + \theta_3 x^3$$

The results are reported in Tables 5.5 and 5.6. The D_s-optimal design obtained for this model for the parameters θ_2 and θ_4 is as follows.

$$P^* = \left\{ \begin{array}{ccccc} -1.00 & -0.71 & 0.00 & 0.71 & 1.00 \\ 0.1667 & 0.1667 & 0.3333 & 0.1667 & 0.1667 \end{array} \right\}$$

5.3 Application of D_s-optimality Criterion to the Practical Model in Chemistry

In this section, we are going to consider the application of D_s-optimality criterion to the practical model (3.4) in chemistry. We have $\underline{v}_x = (x, x^{1/2}, x^2)^T$, $x \in (0, 0.2]$, $\underline{\theta} = (\theta_0, \theta_1, \theta_2)^T$.We considered θ_1 and θ_2, the parameters corresponding to the design variable terms $x^{1/2}$ and x^2 as the parameters of interest. The model can be denoted as follows.

$$E(y|x) \quad = \quad \theta_1 x^{1/2} + \theta_2 x^2 + \theta_0 x$$

The design space was approximated by a grid of 20 points equally spaced between 0.01 and 0.20 with 0.01 interval. Results are reported in Tables 5.7 and 5.8. As we see using the directional derivatives improves the convergence a lot. The D_s-optimal design obtained for the chemistry model is as follows.

$$P^* = \left\{ \begin{array}{ccc} 0.02 & 0.10 & 0.20 \\ 0.49 & 0.34 & 0.17 \end{array} \right\}$$

5.4 D_s-optimal Designs for Two Variable Regression Models

In this chapter, we are going to focus on the construction of D_s-optimal designs for polynomial regression models with two design variables. The design variables are x_1 and x_2. Consider the following two variable full model which we discussed earlier.

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 = \underline{v}_x^T \underline{\theta}$$

First we are going to consider the two variable reduced model obtained by excluding the interaction term. Then we are considering the full model. The discretized design space will consist of all the pairs (x_1, x_2) where values of x_i will be between -1.0 and 1.0 at intervals of 0.1. The design space will consist of $(21)^2 = 441$ pairs of (x_1, x_2). The results will be reported in the same way as in for one variable model using the functions $x^\delta, e^{\delta x}, e^{\delta x}/(1 + e^{\delta x})$ and normal CDF. The convergence of the algorithms were compared by replacing partial derivatives of the criterion function with directional derivatives.

5.4.1 Two Variable Reduced Model taking θ_3 and θ_4 as the Parameters of Interest

Consider the following reduced model.

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2$$

The D_s-optimal design for this model was constructed taking θ_3 and θ_4 as the parameters of interest. Results are reported in Tables 5.9 and 5.10 and the optimal design is obtained as

$$P^* = \left\{ \begin{array}{ccccccccc} -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 \\ -1.00 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 1.00 \\ 0.0625 & 0.1250 & 0.0625 & 0.1250 & 0.2500 & 0.1250 & 0.0625 & 0.1250 & 0.0625 \end{array} \right\}$$

Best δ values are

- 0.800 for $f(x, \delta) = e^{\delta x}$ where $x = d$
- 0.600 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ where $x = d$
- 0.440 for $f(x, \delta) = \Phi(\delta x)$ where $x = d$
- 0.900 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ where $x = F$
- 0.800 for $f(x, \delta) = \Phi(\delta x)$ where $x = F$

When comparing the results of logistic CDF and normal CDF, it can be seen that the convergence rate has increased when using directional derivatives of the criterion function instead of partial derivatives.

5.4.2 Two Variable Full Model taking θ_1 and θ_2 as the Parameters of Interest

Next we will consider the application of D_s-optimality criterion to the two variable full model. The model is given by

$$E(y|x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 = \underline{v}_x^T \underline{\theta}.$$

We consider θ_1 and θ_2 as the parameters of interest. Then, the convergence of the algorithms was assessed using directional derivatives of D_s-optimality criterion instead of partial derivatives. Results are reported in Tables 5.11 and 5.12 and the optimal design is obtained as

$$P^* = \left\{ \begin{matrix} -1.00 & 1.00 & -1.00 & 1.00 \\ -1.00 & -1.00 & 1.00 & 1.00 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{matrix} \right\}$$

Moreover, the best choices for the δ for each of the functions are

- 0.500 for $f(x, \delta) = e^{\delta x}$ where $x = d$
- 0.500 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ where $x = d$
- 0.400 for $f(x, \delta) = \Phi(\delta x)$ where $x = d$
- 1.000 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ where $x = F$
- 0.650 for $f(x, \delta) = \Phi(\delta x)$ where $x = F$

Table 5.5: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for quartic regression model with θ_2 and θ_4 as the parameters of interest using D_s-optimality criterion

$f(x,\delta) = x^{\delta}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.500	20	198	1995	7837	16402	27602
0.600	17	165	1663	6531	13669	23002
0.700	15	142	1425	5598	11716	19716
0.800	13	124	1247	4899	10252	17252
0.900	12	111	1109	4354	9113	15335
1.000	**11**	**100**	**998**	**3919**	**8202**	**13802**
$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.180	29	271	2764	10877	22762	38320
0.200	26	244	2487	9789	20486	34488
0.250	21	195	1990	7831	16389	27590
0.300	17	163	1658	6526	13657	22992
0.350	15	139	1421	5594	11706	19707
0.400	**12**	**122**	**1243**	**4894**	**10242**	**17243**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.400	41	397	4017	15794	33062	55642
0.450	39	378	3830	15058	31524	53051
0.500	38	366	3706	14567	30495	51318
0.550	37	359	3628	14261	29857	50242
0.650	**36**	**355**	**3582**	**14074**	**29465**	**49579**
0.700	37	357	3601	14149	29624	49844
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.250	40	386	3913	15386	32209	54208
0.300	37	357	3617	14220	29769	50098
0.350	35	342	3457	13587	28446	47868
0.400	**34**	**336**	**3391**	**13324**	**27899**	**46943**
0.450	35	337	3398	13351	27955	47035
0.500	38	344	3469	13626	28533	48003

Table 5.6: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for quartic regression model with θ_2 and θ_4 as the parameters of interest using D_s-optimality criterion

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.400	26	246	2490	9792	20499	34500
0.450	23	219	2214	8704	18223	30668
0.500	21	197	1993	7834	16402	27602
0.550	19	179	1812	7122	14912	25094
0.650	17	152	1534	6027	12620	21235
0.700	**16**	**141**	**1424**	**5597**	**11719**	**19719**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.250	26	246	2496	9818	20553	34591
0.300	22	206	2081	8182	17130	28828
0.350	19	177	1784	7014	14684	24712
0.400	17	155	1562	6138	12851	21624
0.450	15	138	1388	5456	11424	19223
0.500	**14**	**124**	**1250**	**4911**	**10283**	**17302**

Table 5.7: Number of iterations need to achieve $\max_{1\leq j\leq J}\{F_j\} \leq 10^{-t}$ with $x = d$ for practical model in chemistry using D_s-optimality criterion

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.400	26	201	529	951	1395	1841
0.550	19	146	385	692	1015	1339
0.600	18	134	353	634	930	1228
0.750	14	108	283	507	744	982
0.800	**14**	**101**	**265**	**476**	**698**	**921**
1.000	22	79	211	381	558	737
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.250	22	159	422	762	1119	1479
0.300	18	133	352	635	933	1232
0.350	16	114	302	544	800	1056
0.400	14	100	264	476	700	924
0.450	**13**	**89**	**235**	**424**	**622**	**822**
0.500	28	91	209	381	560	740
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.350	45	344	910	1638	2406	3177
0.550	38	292	770	1384	2032	2682
0.700	**37**	**290**	**764**	**1373**	**2014**	**2658**
0.800	38	299	787	1414	2073	2736
0.900	39	315	828	1487	2181	2877
1.000	42	337	887	1592	2334	3079
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.250	41	314	830	1495	2195	2898
0.300	38	290	767	1381	2028	2677
0.400	**36**	**273**	**719**	**1293**	**1898**	**2505**
0.500	36	279	735	1321	1938	2557
0.600	38	306	803	1442	2113	2788
0.800	49	429	1125	2015	2951	3891

Table 5.8: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for practical model in chemistry using D_s-optimality criterion

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.350	31	228	604	1088	1597	2109
0.550	20	146	385	692	1016	1342
0.700	16	115	302	544	798	1054
0.800	14	101	265	476	698	922
0.900	12	90	235	423	621	819
1.000	**12**	**80**	**212**	**381**	**558**	**737**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.250	27	200	530	954	1401	1850
0.300	23	167	442	795	1168	1542
0.400	17	126	332	596	875	1156
0.450	15	112	295	530	778	1027
0.500	14	101	265	477	700	924
0.600	**12**	**85**	**221**	**398**	**583**	**770**

Table 5.9: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for two variable reduced regression model with θ_3 and θ_4 as the parameters of interest

$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.200	48	323	628	920	1211	1502
0.300	32	216	419	614	802	1002
0.400	24	162	314	461	606	751
0.600	16	108	210	307	404	501
0.720	14	90	175	256	337	418
0.800	**11**	**81**	**157**	**230**	**303**	**376**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.350	84	563	1086	1587	2086	2584
0.450	75	504	972	1418	1863	2307
0.600	**71**	**475**	**912**	**1329**	**1745**	**2159**
0.750	73	486	928	1351	1771	2191
0.820	75	500	954	1387	1818	2249
0.900	78	523	997	1448	1897	2345
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.300	71	476	918	1339	1759	2178
0.440	**67**	**451**	**863**	**1256**	**1648**	**2039**
0.500	69	463	884	1285	1685	2084
0.620	77	522	992	1440	1885	2329
0.700	86	592	1120	1623	2122	2621
0.800	103	724	1364	1970	2574	3176

Table 5.10: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for two variable reduced regression model with θ_3 and θ_4 as the parameters of interest

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.350	56	373	721	1053	1385	1716
0.450	44	291	561	820	1077	1334
0.600	33	219	422	616	808	1001
0.750	27	176	338	493	647	801
0.820	25	162	310	451	592	733
0.900	**23**	**148**	**283**	**411**	**539**	**668**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
0.300	41	274	528	771	1013	1254
0.440	29	188	361	526	691	855
0.500	25	166	318	463	608	753
0.620	21	135	257	374	491	607
0.700	19	120	228	332	435	538
0.800	**17**	**105**	**200**	**291**	**381**	**471**

Table 5.11: Number of iterations need to achieve $\max_{1 \leq j \leq J}\{F_j\} \leq 10^{-t}$ with $x = d$ for two variable full model with θ_1 and θ_2 as the parameters of interest

$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.050	137	343	580	822	1064	1306
0.100	69	172	291	411	532	654
0.200	35	86	146	206	267	327
0.300	24	58	97	138	178	219
0.400	18	44	73	104	134	164
0.500	**15**	**35**	**59**	**83**	**107**	**131**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.400	63	148	242	337	432	527
0.500	**60**	**139**	**225**	**312**	**399**	**486**
0.800	66	145	229	314	399	487
1.000	77	167	260	354	447	541
1.200	95	202	312	421	531	640
1.500	138	288	438	587	735	883
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.200	67	160	265	370	476	582
0.300	58	135	219	304	390	475
0.400	**58**	**130**	**208**	**287**	**366**	**445**
0.500	62	137	216	295	374	453
0.650	78	165	255	345	434	524
0.700	86	180	277	373	468	564

Table 5.12: Number of iterations need to achieve $\max_{1 \le j \le J}\{F_j\} \le 10^{-t}$ with $x = F$ for two variable full model with θ_1 and θ_2 as the parameters of interest

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.400	39	91	150	209	269	328
0.500	32	74	121	168	216	263
0.700	25	55	88	121	155	188
0.800	22	49	78	107	136	165
0.900	20	44	70	95	121	147
1.000	**19**	**40**	**63**	**86**	**110**	**133**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.150	61	148	247	346	446	547
0.200	47	113	186	261	336	410
0.300	33	77	126	175	225	274
0.400	26	59	96	132	169	206
0.500	22	49	78	107	136	165
0.650	**18**	**39**	**61**	**83**	**105**	**128**

Chapter 6

Construction of D_S-optimal Designs using Clustering Approach

As mentioned earlier, D_s-optimality is quite important when we are interested in a subset of parameters. This happens sometimes when we are more interested in some terms (for example, even power terms or odd power terms) in the model. In this situation we can simply construct D_s-optimal designs for the corresponding parameters.

Also, it is good to apply the clustering approach for D_s-optimality because clusters can be formed according to the design points as well as the convergence will be faster.

Promising results are obtained using the clustering approach along with using the logistic CDF and normal CDF and using directional derivatives in the algorithmic function.

6.1 Cubic Regression Model with θ_1 and θ_3 as the Parameters of Interest

The D_s-optimal designs are constructed using clustering approach for the cubic regression model by taking θ_1 and θ_3 as the parameters of interest. The rearranged model is given by

$$E(Y|x) = \theta_1 x + \theta_3 x^3 + \theta_0 + \theta_2 x^2.$$

We considered the three functions $x^\delta, e^{\delta x}/(1+e^{\delta x})$ and normal CDF. The design space was equally spaced between -1.00 and 1.00 with 0.01 intervals. Therefore, at the beginning the total number of 201 design points were there.

Results were obtained for the function x^δ when using $x = d$, the partial derivatives of D_s-optimality criterion. But for the logistic CDF and normal CDF the best results were obtained when using directional derivatives in the function. The δ values and the total number of iterations which were used to form the clusters at the beginning are given below.

- x^δ where $x = d$: $\delta = 0.9$, $n = 100$
- $e^{\delta x}/(1+e^{\delta x})$ where $x = F$: $\delta = 0.6$, $n = 100$
- Normal CDF where $x = F$: $\delta = 0.35$, $n = 100$

In Table 6.1, the number of iterations needed to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ for $1 \leq j \leq J$ when using the function $f(x, \delta) = x^{\delta}$ where $x = d$ are displayed. Table 6.2 gives results for $x = F$.

Table 6.1: Number of iterations need to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for the D_s-optimal design of cubic regression model with θ_1 and θ_3 as the parameters of interest using clustering approach

$f(x,\delta) = x^{\delta}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
12	1	1	14	225	372	515
13	1	1	13	208	343	475
14	1	1	12	193	319	441
15	1	1	11	180	298	412
18	1	1	10	151	248	344
20	**1**	**1**	**9**	**136**	**224**	**310**

The D_s-optimal design obtained for the cubic regression model is given below.

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.58 & 0.58 & 1.00 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{array} \right\}$$

In here, we noticed that the support points consist of four clusters of points which is either of neighboring points or a single point. As we have mentioned earlier, since this solution is for a discretized design space, the solution for a continuous design space should be a 4-point design. Therefore, by taking convex combinations of particular clusters of points, we obtained the same optimal design.

Table 6.2: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the D_s-optimal design of cubic regression model with θ_1 and θ_3 as the parameters of interest using clustering approach

$f(x,\delta) = e^{\delta x}/1+e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
50	1	1	14	49	77	104
60	1	1	12	41	64	86
75	1	1	10	35	51	69
85	1	1	9	29	45	61
90	1	1	9	28	43	57
100	**1**	**1**	**8**	**25**	**38**	**51**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
30	1	1	16	49	76	102
40	1	1	12	37	57	76
50	1	1	10	30	45	61
65	1	1	8	23	35	46
70	1	1	8	22	32	43
80	**1**	**1**	**7**	**19**	**28**	**37**

Graphical Interpretation

Let us consider the graphical interpretation for the construction of D_s-optimal design for the cubic regression model taking θ_1 and θ_3 as the parameters of interest. For the construction the function being considered is $e^{\delta x}$ where $x = d$, the partial derivatives of D_s-optimality criterion and δ was taken as 0.2. Figure 6.1 provides the plots of weights versus design points obtained after running the algorithm for few number of iterations and at the optimum. The curve in plot 6.1 (a), depicts three minimal and two maximal turning points. Two peak points can be seen at the design points -1.00 and 1.00. The weights in the minimal turning points converge to zero when increasing the number of iteration. The converging pattern is noticeable when observing graphs from plot 6.1 (a) to plot 6.1 (d). By plot 6.1 (d), it can be seen that, the weights corresponding to the design points -1.00, -0.58, 0.58 and 1.00 converged to 0.25 at the optimum.

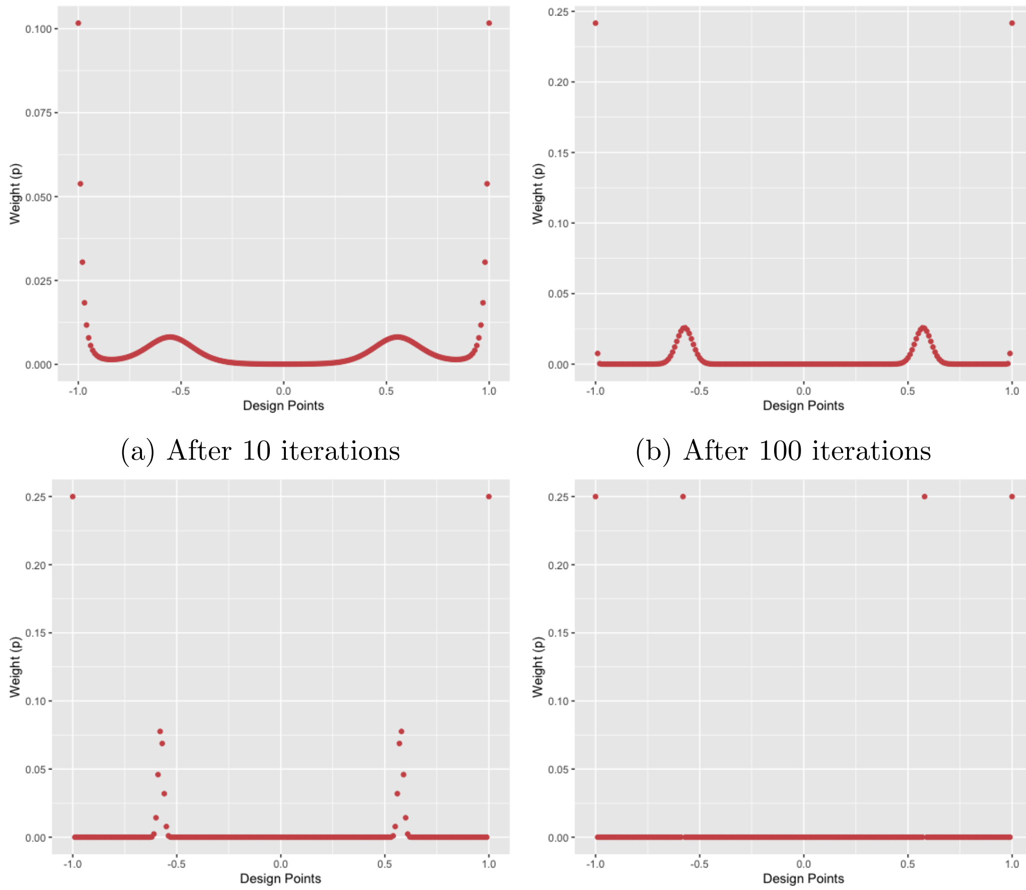
Figure 6.2 gives the plots of variance function versus design points after running the algorithm for 10 iterations and at the optimum. It can be seen that, the curve in plot 6.2 (a) have three minimal and two maximal turning points as in our weights versus design point plots in figure 6.1. At the optimum, the variance function values corresponding to design points -1.00, 1.00 and two maximal turning points converged to 2.0 (plot 6.2 (b)).

The variance function corresponding to this cubic D_s-optimal design is,

$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) - \underline{f}_2^T(\underline{x}) M_{22}^{-1}(p) \underline{f}_2(\underline{x}) \\ &= -4.9045x^6 - 8.1689x^4 + 3.6337x^2 + 0.1095 \end{aligned}$$

6.2 Quartic Regression Model with θ_3 as the Parameter of Interest

In this section we are going to discuss about the application of clustering approach to quartic regression model by taking θ_3 as the parameter of interest.



(a) After 10 iterations (b) After 100 iterations

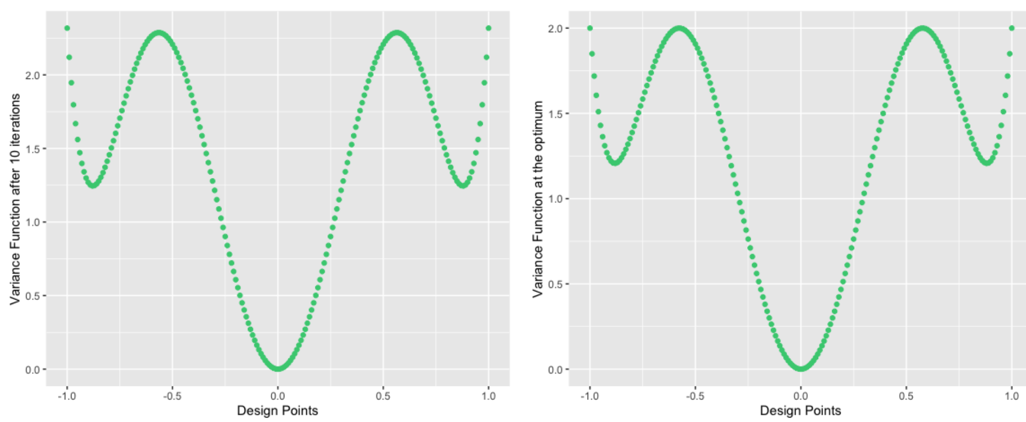
(c) After 1000 iterations (d) At the optimum

Figure 6.1: Weights vs design points for the cubic regression model with θ_1 and θ_3 as the parameters of interest

The model can be denoted as follows.

$$E(Y|x) = \theta_3 x^3 + \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_4 x^4$$

We observe the convergence of the algorithm using $x = d$, and then compared these results using $x = F$. The results are given in Tables 6.3 and 6.4. The δ values and the number of iterations which were used to form the clusters



(a) After 10 iterations (b) At the optimum

Figure 6.2: Variance function vs design points for the cubic regression model with θ_1 and θ_3 as the parameters of interest

at the beginning are given below.

- x^{δ} where $x = d$: $\delta = 0.9$, $n = 50$
- $e^{\delta x}$ where $x = d$: $\delta = 0.7$, $n = 150$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = d$: $\delta = 0.75$, $n = 100$
- Normal CDF where $x = d$: $\delta = 0.85$, $n = 200$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.85$, $n = 150$
- Normal CDF where $x = F$: $\delta = 1.2$, $n = 80$

The optimal design obtained for the functions x^{δ} and $e^{\delta x}$ is given below.

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.50 & 0.50 & 1.00 \\ 0.1667 & 0.3333 & 0.3333 & 0.1667 \end{array} \right\}$$

As for the logistic CDF and normal CDF, the support points of the final designs consist of four clusters of points with neighbouring points or a single point. The design spaces for the two functions are given below respectively

$$P^* = \left\{ \begin{array}{cccccc} -1.00 & -0.50 & -0.49 & 0.49 & 0.50 & 1.00 \\ 0.1667 & 0.2159 & 0.1179 & 0.2159 & 0.1179 & 0.1667 \end{array} \right\}$$

$$P^* = \left\{ \begin{array}{cccccccc} -1.00 & -0.51 & -0.50 & -0.49 & 0.49 & 0.50 & 0.51 & 1.00 \\ 0.1667 & 0.0547 & 0.2166 & 0.0625 & 0.0625 & 0.2166 & 0.0547 & 0.1667 \end{array} \right\}$$

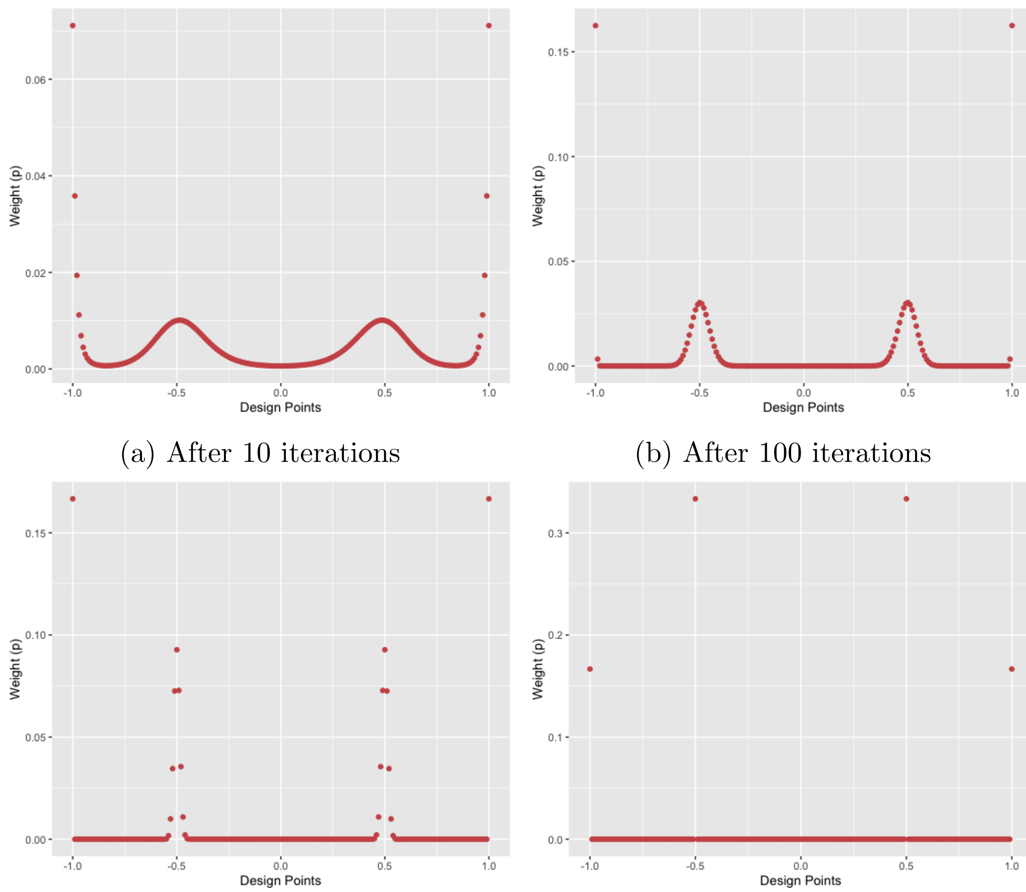
By taking convex combinations of the relevant clusters, these functions also yield the same solution as earlier.

$$P^* = \left\{ \begin{array}{cccc} -1.00 & -0.50 & 0.50 & 1.00 \\ 0.1667 & 0.3333 & 0.3333 & 0.1667 \end{array} \right\}$$

Graphical Interpretation

In this section, we are going to discuss about the graphical interpretation for the D_s-optimal design construction for the quartic regression model taking θ_3 as the parameter of interest. The construction was done using a class of algorithms. The function $e^{\delta x}$, where $x = d$, the partial derivatives of D_s-optimality criterion and δ value 0.2 were considered. Figure 6.3 displays the plots of weights versus design points for several number of iterations and at the optimum. There are three minimal and two maximal turning points in the plot 6.3 (a). When increasing the number of iterations, the weights corresponding to the minimal turning points converged to zero as in the plot

6.3 (d). In addition to that, it is noticeable the weights corresponding to the design points -1.00 and 1.00 have approached to 0.1667 while the weights corresponding to maximal turning points have converged to 0.3333.



(a) After 10 iterations (b) After 100 iterations

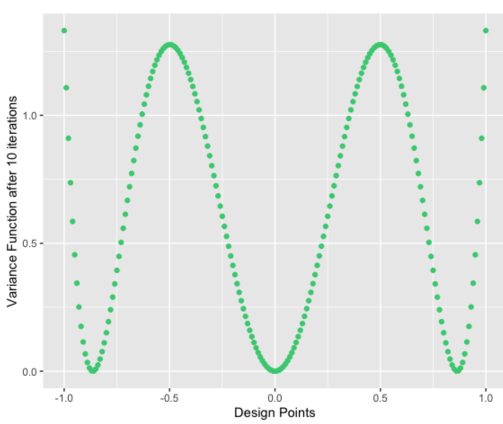
(c) After 1000 iterations (d) At the optimum

Figure 6.3: Weights vs design points for the quartic regression model with θ_3 as the parameter of interest

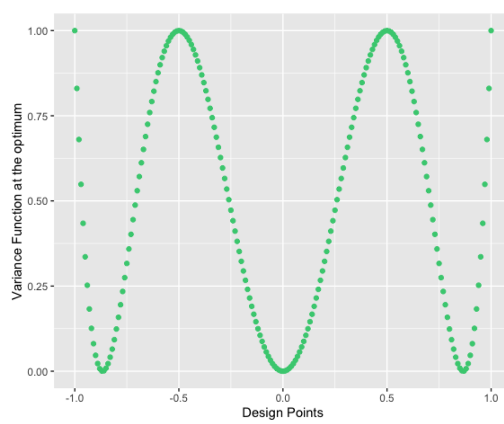
The plots of variance function versus design points for this quartic D_s-optimal design is shown in Figure 6.4. After running the algorithm for 10 iterations, a curve with two maximal and three minimal turning points was obtained. Two peak points can be seen in plot 6.4 (a) which are corresponding to the design points -1.00 and 1.00. The variance function values of these two peak points and maximal turning points converged to 1.00 at the optimum as in plot 6.4 (b). The variance function corresponding to this quartic D_s-optimal design is given below

$$d(\underline{x}, p) = \underline{f}^T(\underline{x})M^{-1}(p)\underline{f}(\underline{x}) - \underline{f}_2^T(\underline{x})M_{22}^{-1}(p)\underline{f}_2(\underline{x})$$

$$= -0.005x^7 + 4.3493x^6 - 0.2904x^5 - 6.5205x^4 - 0.0094x^3 + 2.4423x^2 - 0.1378x + 0.0030.$$



(a) After 10 iterations



(b) At the optimum

Figure 6.4: Variance function vs design points for the quartic regression model with θ_3 as the parameter of interest

6.3 Quartic Regression Model with θ_2 and θ_4 as the Parameters of Interest

The D_s-optimal design was constructed for the quartic regression model with the use of clustering approach by taking θ_2 and θ_4 as the parameters of interest. The considered model is given below.

$$E(Y|x) = \theta_2 x^2 + \theta_4 x^4 + \theta_0 + \theta_1 x + \theta_3 x^3$$

The results are given in Tables 6.5 and 6.6. The δ values and the total number of iterations used in forming clusters for each of the scenarios are given below.

- $e^{\delta x}$ where $x = d$: $\delta = 0.4$, $n = 100$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.5$, $n = 100$
- Normal CDF where $x = F$: $\delta = 0.5$, $n = 100$

The D_s-optimal design obtained with the clustering approach is as follows.

$$P^* = \left\{ \begin{array}{ccccc} -1.00 & -0.71 & 0.00 & 0.71 & 1.00 \\ 0.1667 & 0.1667 & 0.3333 & 0.1667 & 0.1667 \end{array} \right\}$$

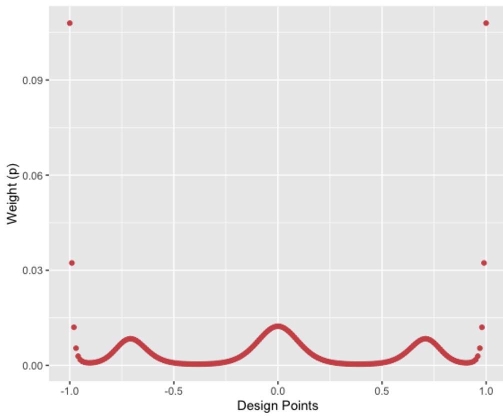
Here also we noticed that the D_s-quartic optimal design obtained consisted of five clusters of support points with neighbouring or a single points. But after taking the convex combinations of the relevant clusters we obtained the same design for all functions.

Graphical Interpretation

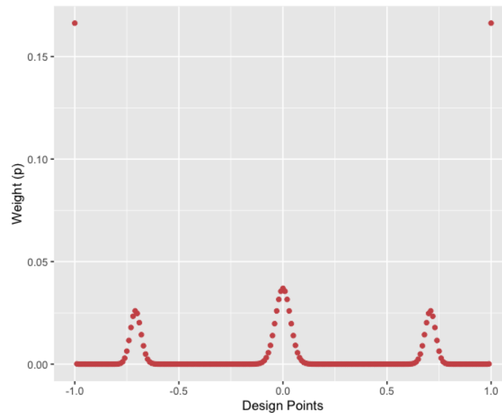
We now focus on the graphical interpretation for the D_s-optimal design for the above model. The function being considered in here is $e^{\delta x}$, where $x = d$, the partial derivatives of D_s-optimality criterion and δ value 0.2. The construction was done using a class of algorithms. Figure 6.5 provides the plots of weights versus design points after running algorithm for a several number of iterations and at the optimum. The plot 6.5 (a) depicts a curve with four minimal and three maximal turning points. The weights in minimal regions converged to zero at the optimum as in plot 6.5 (d). The weights corresponding to the design points -1.00, -0.71, 0.71 and 1.00 have converged to 0.1667 whereas the weight corresponding to the design point 0.00 has converged to 0.3333.

Figure 6.6 provides you with two plots of variance function versus design points for this design. Here also, a curve with three maximal and four minimal turning points was obtained after running the algorithm for 10 iterations. The variance function values of the two peak points and three maximal turning points converged to 2.00 at the optimum as in plot 6.6 (b). The variance function related to this quartic D_s-optimal design is

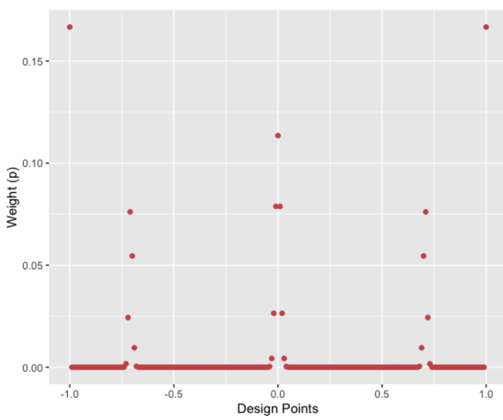
$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) - \underline{f}_2^T(\underline{x}) M_{22}^{-1}(p) \underline{f}_2(\underline{x}) \\ &= 19.5747x^8 - 39.1280x^6 + 33.0549x^4 - 17.9026x^2 + 7.4501 \end{aligned}$$



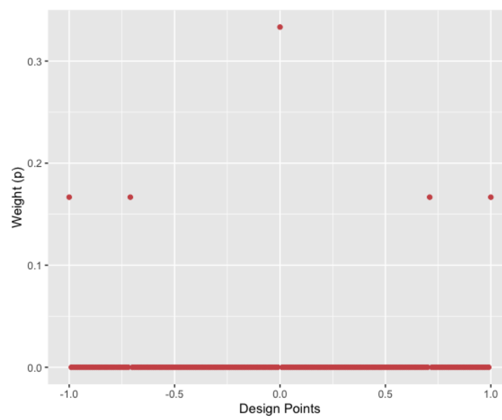
(a) After 10 iterations



(b) After 100 iterations



(c) After 1000 iterations

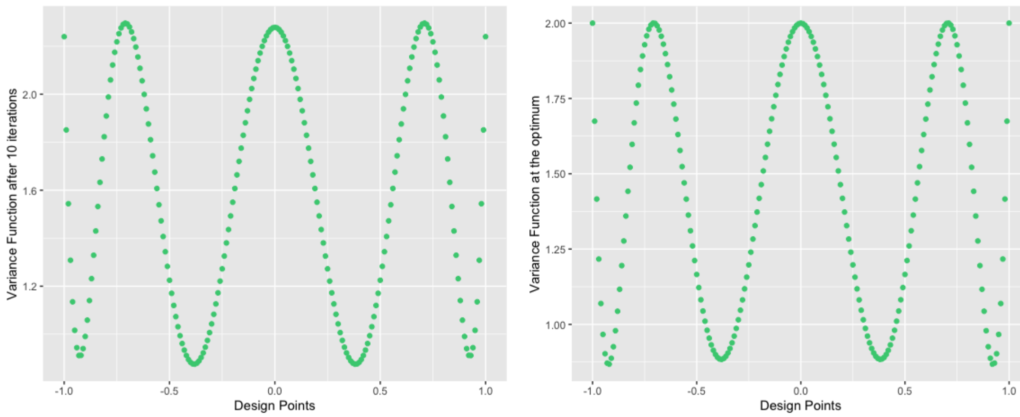


(d) At the optimum

Figure 6.5: Weights vs design points for the quartic regression model with θ_2 and θ_4 as the parameters of interest

6.4 Application of D_s-optimality Criterion to the Practical Model in Chemistry with the use of Clustering Approach

In Chapter 5, we did an application of D_s-optimality to the practical model in chemistry using θ_1 and θ_2 as the parameters of interest. We now construct



(a) After 10 iterations (b) At the optimum

Figure 6.6: Variance function vs design points for the quartic regression model with θ_2 and θ_4 as the parameters of interest

this design using clustering approach. The rearranged model is

$$E(y|x) \quad = \quad \theta_1 x^{1/2} + \theta_2 x^2 + \theta_0 x$$

As in for the previous quartic model, the results were obtained when using $x = d$ with the functions x^δ and $e^{\delta x}$ and $x = F$ with logistic and normal CDF. The results are given in Tables 6.7 and 6.8. The δ values and the total number of iterations used in forming clusters for each of the scenarios are given below.

- x^δ where $x = d$: $\delta = 0.8$, $n = 80$
- $e^{\delta x}$ where $x = d$: $\delta = 0.4$, $n = 80$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.7$, $n = 100$
- Normal CDF where $x = F$: $\delta = 0.5$, $n = 100$

The D_s-optimal design obtained for this model is as follows.

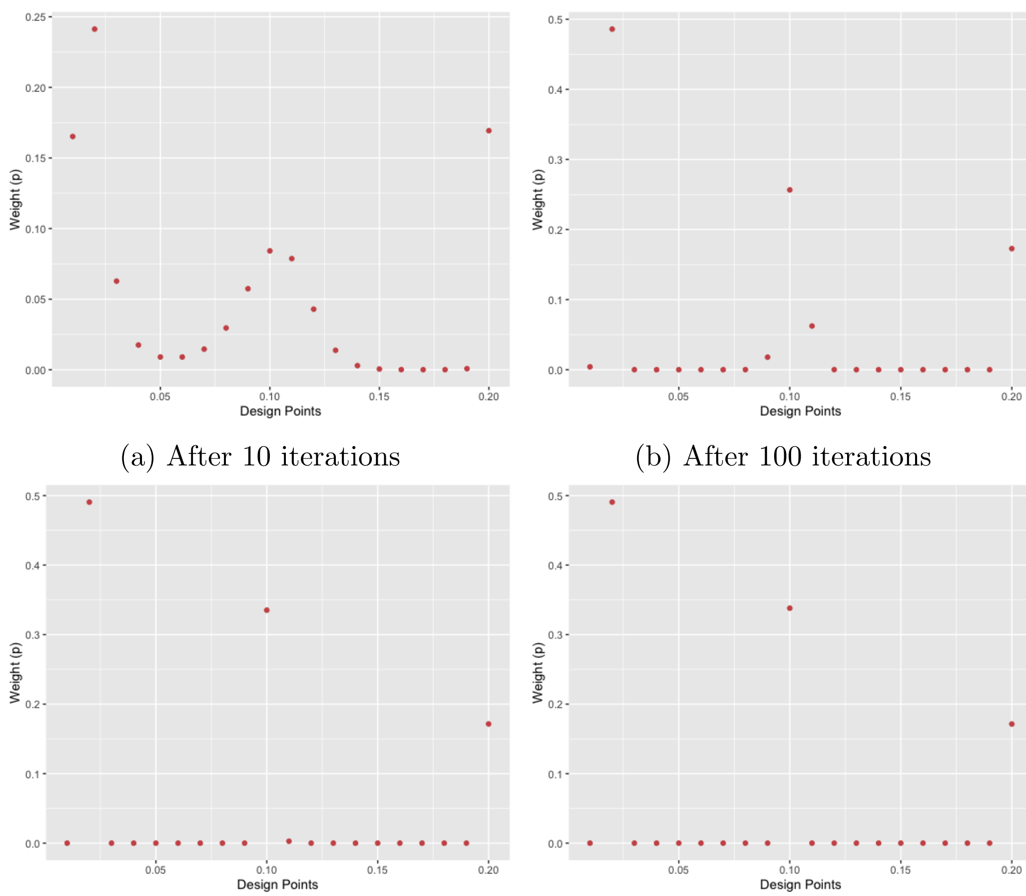
$$P^* = \left\{ \begin{array}{ccc} 0.02 & 0.10 & 0.20 \\ 0.49 & 0.34 & 0.17 \end{array} \right\}$$

Graphical Interpretation

The function $e^{\delta x}$ where δ equals to 0.4 and $x = d$, the partial derivatives of D_s-optimality criterion were used in the design construction. The construction was carried out using class of algorithms. The plots of weights versus design points after several number of iterations (10, 100 and 400) and at the optimum are displayed in Figure 6.7. There are two maximal and two minimal turning points in the curve. When increasing the number of iterations, the weights lie in the minimal turning points converge to zero as in plot 6.7 (d). According to plot 6.7 (d), we see that the weights related to the design points 0.02, 0.10 and 0.20 have converged to 0.49, 0.34 and 0.17 respectively. The convergence pattern can be visualized when going through the plots in ascending order of the iterations.

The plots of variance function versus design points are given in Figure 6.8. In plot 6.8 (a), we see two minimal and two maximal turning points. When increasing the iterations the variance function values around maximal points are converging to 2.00. In addition to that there is a peak point at the design point $x = 0.20$ in which the variance function has converged to 2.00 at the optimum as in the plot 6.8 (b). The variance function related to the chemistry model using D_s-optimality criterion is

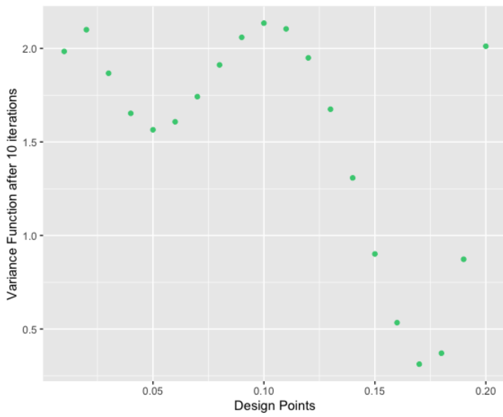
$$
\begin{aligned}
d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) - \underline{f}_2^T(\underline{x}) M_{22}^{-1}(p) \underline{f}_2(\underline{x}) \\
&= 13690x^4 - 84820x^2 + 14619.9896x^2 + 15448x^{5/2} - 5756x^{3/2} + 612.6x.
\end{aligned}
$$



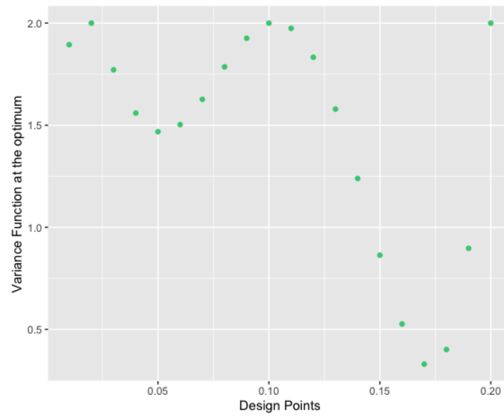
(a) After 10 iterations (b) After 100 iterations

(c) After 400 iterations (d) At the optimum

Figure 6.7: Weights vs design points for the practical model in chemistry with θ_1 and θ_2 as the parameters of interest



(a) After 10 iterations



(b) At the optimum

Figure 6.8: Variance function vs design points for the practical model in chemistry with θ_1 and θ_2 as the parameters of interest

6.5 Application of D_s-optimality Criterion to Two Variable Regression Models using Clustering Approach

We are going to discuss about the D_s-optimal design construction for the two variable models with the use of clustering approach. We will consider two models in sections 5.4.1 and 5.4.2.

6.5.1 Two Variable Reduced Model taking θ_3 and θ_4 as the Parameters of Interest

The model is given by

$$E(y|x) = \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

The design space was approximated by a grid of 441 (x_1, x_2) pairs where each variable takes values between -1.0 to 1.0 with 0.1 intervals. The best results were obtained for the function $f(x, \delta) = e^{\delta x}$ when $x = d$, partial derivatives. But for the logistic CDF and normal CDF, the best results were obtained when we replaced partial derivatives in the function $f(x, \delta)$ by the directional derivatives. The results are given in Tables 6.9 and 6.10. The δ values and the total number of iterations which were used to form clusters are,

- $e^{\delta x}$ where $x = d$: $\delta = 0.4$, $n = 160$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.75$, $n = 160$
- $\Phi(\delta x)$ where $x = F$: $\delta = 0.6$, $n = 130$

The D_s-optimal design obtained for this reduced regression model for every above situations using clustering approach is as follows.

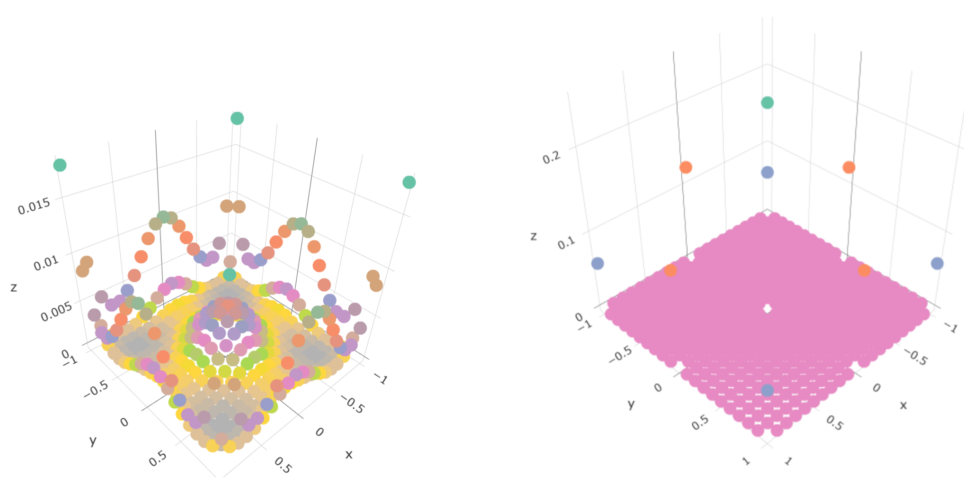
$$P^* = \begin{Bmatrix} -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 & -1.00 & 0.00 & 1.00 \\ -1.00 & -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 1.00 & 1.00 & 1.00 \\ 0.0625 & 0.1250 & 0.0625 & 0.1410 & 0.2180 & 0.1410 & 0.0625 & 0.1250 & 0.0625 \end{Bmatrix}$$

The best choices of δ for the above three functions are

- 64 for $f(x, \delta) = e^{\delta x}$ when $x = d$
- 92 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ when $x = F$
- 94 for $f(x, \delta) = \Phi(\delta x)$ when $x = F$

Graphical Interpretation

The function being considered was logistic CDF $e^{\delta x}/(1+e^{\delta x})$ and δ was taken as 0.6. Figure 6.9 gives the plots of weights versus design points after running the algorithm for 10 iterations and at the optimum.

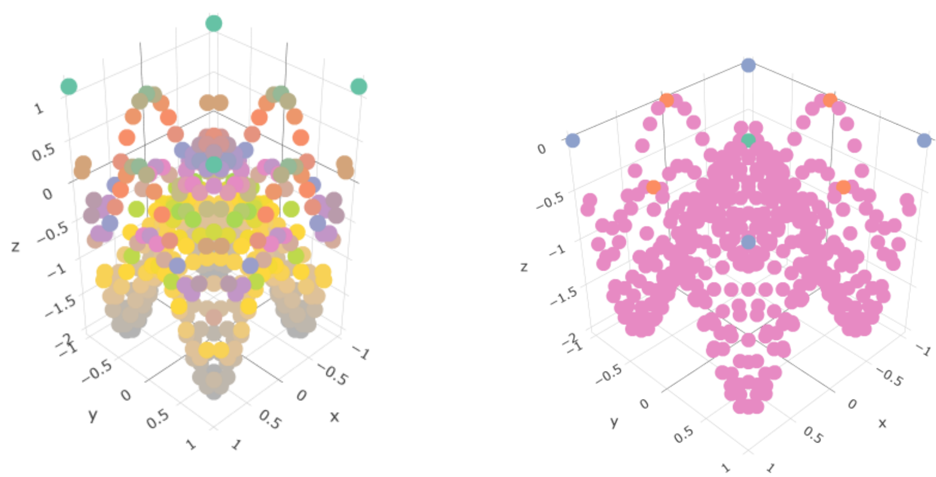


(a) After 10 iterations (b) At the optimum

Figure 6.9: Weights vs design points for the two variable reduced model taking θ_3 and θ_4 as the parameters of interest

At the optimum (plot 6.9 (b)), D-optimal design have some weights on the pairs $(-1,-1)$, $(0,-1)$, $(1,-1)$, $(-1,0)$, $(0,0)$, $(1,0)$, $(-1,1)$, $(0,1)$, (1, 1) which are similar to the weights in the design space we obtained. All the other weights have converged to zero at the optimum. Figure 6.10 depicts the plots of variance function versus design points for the same model after 10 iterations and at the optimum. The variance function related to this this model is

$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) \\ &= 4x_1^4 + 4x_2^4 - 4x_1^2 - 4x_2^2 + 2. \end{aligned}$$



(a) After 10 iterations

(b) At the optimum

Figure 6.10: Variance function vs design points for the two variable reduced model taking θ_3 and θ_4 as the parameters of interest

6.5.2 Two Variable Full Model taking θ_1 and θ_2 as the Parameters of Interest

The model is given by

$$E(y|x) = \theta_1 x_1 + \theta_2 x_2 + \theta_0 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2$$

Results were obtained for the function $f(x, \delta) = e^{\delta x}$ when $x = d$, partial derivatives. But for the logistic CDF and normal CDF, the best results were obtained when we replaced partial derivatives in the function $f(x, \delta)$ by the directional derivatives. The results are reported in Tables 6.11 and 6.12. The δ values and total number of iterations which were used to form clusters are

- $e^{\delta x}$ where $x = d$: $\delta = 0.300$, $n = 60$
- $e^{\delta x}/(1 + e^{\delta x})$ where $x = F$: $\delta = 0.500$, $n = 70$
- Normal CDF where $x = F$: $\delta = 0.500$, $n = 50$

The D_s-optimal design obtained for the two variable full model using clustering approach is given below:

$$P^* = \begin{Bmatrix} -1.00 & 1.00 & -1.00 & 1.00 \\ -1.00 & -1.00 & 1.00 & 1.00 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{Bmatrix}$$

The best choices of δ for the above three functions are

- 8.4 for $f(x, \delta) = e^{\delta x}$ when $x = d$

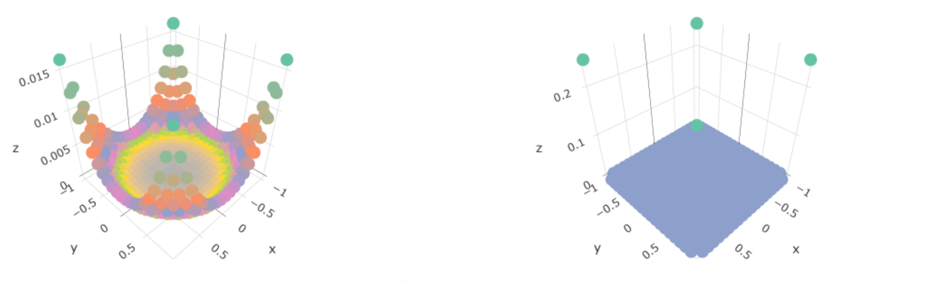
- 12.5 for $f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$ when $x = F$

- 12.0 for $f(x, \delta) = \Phi(\delta x)$ when $x = F$

According to the results, we see that when increasing the δ value and using clustering approach, the number of iterations need to reach optimality have been decreased. Thus the convergence rate is improved by using the clustering aspproach.

Graphical Interpretation

The function being considered was logistic CDF $e^{\delta x}/(1+e^{\delta x})$ and δ was taken as 0.5. Figure 6.11 provides the plots of weights versus design points after running the algorithm for 10 iterations and at the optimum.



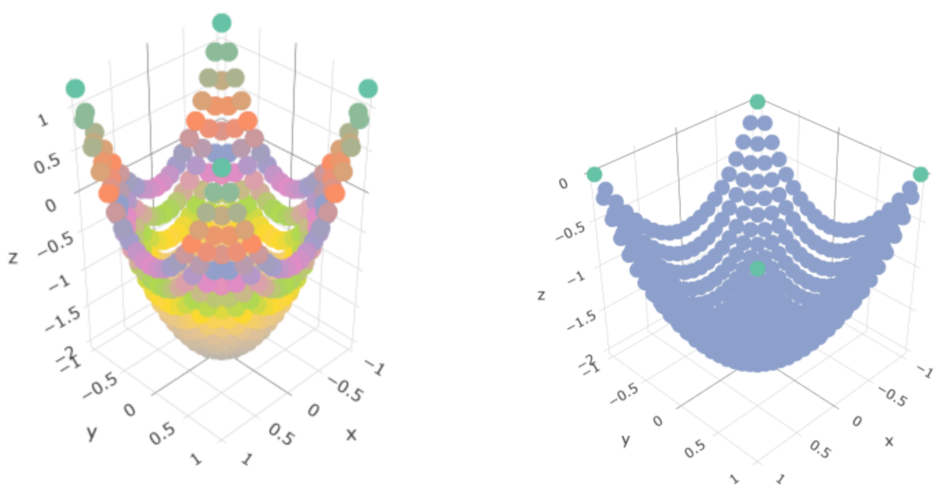
(a) After 10 iterations (b) At the optimum

Figure 6.11: Weights vs design points for the two variable full model taking θ_1 and θ_2 as the parameters of interest

At the optimum (plot 6.11 (b)), D_s-optimal design have same weights as we obtained earlier. All the other weights have converged to zero at the optimum. Figure 6.12 depicts the plots of variance function versus design points for the same model after 10 iterations and at the optimum. The variance function related to this this model is

$$\begin{aligned} d(\underline{x}, p) &= \underline{f}^T(\underline{x}) M^{-1}(p) \underline{f}(\underline{x}) \\ &= x_1^2 + x_2^2. \end{aligned}$$

In all of the results it is clear that after combining the clustering approach with the properties of the directional derivatives of the criterion, the convergence is improved a lot.



(a) After 10 iterations (b) At the optimum

Figure 6.12: Variance function vs design points for the two variable full model taking θ_1 and θ_2 as the parameters of interest

Table 6.3: Number of iterations need to achieve $\underset{1 \leq j \leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for the D_s-optimal design of quartic regression model with θ_3 as the parameter of interest using clustering approach

$f(x,\delta) = x^{\delta}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
0.700	1	1	563	1004	1414	1820
0.800	1	1	493	878	1238	1592
0.950	1	1	415	740	1043	1341
1.000	**1**	**1**	**395**	**703**	**991**	**1274**
1.200	1	1	572	779	965	1218
1.500	1	1	646	827	1008	1193
$f(x,\delta) = e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
1.500	1	1	119	2951	5970	8569
1.600	1	1	112	2766	5597	8033
1.700	1	1	105	2604	5268	7561
1.800	1	1	99	2459	4975	7141
1.900	1	1	94	2330	4713	6765
2.000	**1**	**1**	**90**	**2214**	**4478**	**6427**
$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
1.000	1	1	1734	3213	4581	5940
1.200	1	1	1649	3057	4363	5660
1.400	**1**	**1**	**1625**	**3014**	**4305**	**5588**
1.500	1	1	1629	3024	4322	5612
1.650	1	1	1655	3074	4397	5712
2.000	1	1	1790	3335	4780	6216
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
1.200	**1**	**1**	**1676**	**3084**	**4397**	**5701**
1.300	1	1	1760	3243	4628	6005
1.400	1	1	1872	3456	4937	6410
1.500	1	1	2017	3730	5336	6933
1.650	1	1	2307	4283	6139	7985
2.100	1	1	4069	7659	11053	14428

Table 6.4: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the D_s-optimal design of quartic regression model with θ_3 as the parameter of interest using clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
8.450	1	1	269	365	455	544
8.500	1	1	268	363	453	541
8.750	1	1	258	350	437	523
9.000	1	1	254	344	428	512
9.250	1	1	247	335	416	498
9.500	**1**	**1**	**243**	**328**	**407**	**487**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
3.000	1	1	241	464	662	858
3.500	1	1	207	397	567	735
4.000	1	1	181	348	496	643
5.000	1	1	145	278	397	514
5.300	1	1	254	333	408	483
5.400	**1**	**1**	**136**	**258**	**368**	**475**

Table 6.5: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the D_s-optimal design of quartic regression model with θ_2 and θ_4 as the parameters of interest using clustering approach

$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
10	1	1	47	319	2117	4324
30	1	1	17	108	707	1442
35	1	1	15	92	606	1236
40	1	1	13	81	530	1082
70	1	1	8	47	304	619
80	**1**	**1**	**7**	**41**	**266**	**541**

Table 6.6: Number of iterations need to achieve $\underset{1 \le j \le J}{max}\{F_j\} \le 10^{-t}$ with $x = F$ for the D_s-optimal design of quartic regression model with θ_2 and θ_4 as the parameters of interest using clustering approach

$f(x, \delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
50	1	1	22	128	988	2236
62	1	1	18	103	798	1805
70	1	1	16	92	707	1600
76	1	1	15	85	652	1475
80	1	1	14	81	619	1402
95	**1**	**1**	**12**	**68**	**522**	**1182**
$f(x, \delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
60	1	1	11	68	491	2334
70	1	1	10	58	421	2003
76	1	1	9	54	387	1846
85	1	1	8	48	346	1652
95	1	1	8	44	310	1480
100	**1**	**1**	**7**	**41**	**294**	**1407**

Table 6.7: Number of iterations need to achieve $\max\limits_{1\leq j\leq J}\{F_j\} \leq 10^{-t}$ with $x = d$ for the D_s-optimal design of chemistry practical model with θ_1 and θ_2 as the parameters of interest using clustering approach

$f(x,\delta) = x^{\delta}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
8	1	1	77	130	182	233
10	1	1	62	105	146	187
12	1	1	52	87	122	156
16	1	1	39	66	92	117
18	1	1	35	59	82	105
20	**1**	**1**	**32**	**53**	**74**	**94**
$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
8	1	1	49	76	102	127
10	1	1	40	61	82	102
15	1	1	27	42	55	69
18	1	1	23	35	46	58
20	1	1	21	32	42	52
25	**1**	**1**	**17**	**26**	**34**	**42**

Table 6.8: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the D_s-optimal design of chemistry practical model with θ_1 and θ_2 as the parameters of interest

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
10	1	1	77	120	161	201
12	1	1	65	100	134	168
15	1	1	52	80	107	134
18	1	1	44	67	89	111
22	1	1	36	55	73	91
25	**1**	**1**	**32**	**48**	**64**	**80**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
10	1	1	49	76	102	128
14	1	1	35	54	73	91
18	1	1	28	42	57	71
20	1	1	25	38	51	63
25	1	1	20	31	41	50
28	**1**	**1**	**18**	**28**	**36**	**45**

Table 6.9: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for two variable reduced model taking θ_3 and θ_4 as the parameters of interest using clustering approach

$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
10	1	1	48	146	269	391
20	1	1	25	74	135	196
35	1	1	15	43	78	113
40	1	1	13	38	68	99
52	1	1	10	29	53	76
64	**1**	**1**	**9**	**24**	**43**	**62**

Table 6.10: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for two variable reduced model taking θ_3 and θ_4 as the parameters of interest using clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
55	1	1	16	85	199	307
64	1	1	14	74	171	264
70	1	1	13	68	157	242
75	1	1	12	63	147	226
84	1	1	11	57	131	202
92	**1**	**1**	**10**	**52**	**120**	**185**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
40	1	1	15	43	76	109
50	1	1	13	35	61	87
62	1	1	11	29	50	70
75	1	1	9	24	41	58
80	1	1	9	23	39	55
94	**1**	**1**	**8**	**20**	**33**	**47**

Table 6.11: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = d$ for the two variable full model taking θ_1 and θ_2 as the parameters of interest using clustering approach

$f(x,\delta) = e^{\delta x}$						
δ	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
4.0	1	1	23	33	42	52
5.0	1	1	19	26	34	41
6.5	1	1	15	21	26	32
7.0	1	1	14	19	24	30
8.0	1	1	12	17	22	26
8.4	**1**	**1**	**12**	**16**	**21**	**25**

Table 6.12: Number of iterations need to achieve $\underset{1\leq j\leq J}{max}\{F_j\} \leq 10^{-t}$ with $x = F$ for the two variable full model taking θ_1 and θ_2 as the parameters of interest using clustering approach

$f(x,\delta) = e^{\delta x}/1 + e^{\delta x}$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
3.0	1	1	48	68	87	107
4.0	1	1	36	50	65	79
5.0	1	1	29	40	51	62
8.0	1	1	18	25	31	38
10.0	1	1	14	19	24	29
12.5	**1**	**1**	**12**	**15**	**19**	**23**
$f(x,\delta) = \Phi(\delta x)$						
δ	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
4.0	1	1	27	38	49	60
5.0	1	1	21	30	39	47
6.6	1	1	16	23	29	35
8.0	1	1	13	18	23	28
10.0	1	1	11	15	18	22
12.0	**1**	**1**	**9**	**12**	**15**	**18**

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis, we have tried to address an important problem in optimal regression design, namely the application of optimal design theory to solve optimization problems with respect to several probability distributions and to improve the convergence of algorithms using the properties of directional derivatives. The focus was on constructing optimal designs using clustering approach, using the properties of the directional derivatives and on modifying the algorithms according to some criteria of interest such as D-optimality and D_s-optimality criteria. In particular, D_s-optimality is quite important when we are interested in a subset of parameters. This situation arises when we are more interested in some of the terms (for example, even or odd power terms) in the model.

We first gave a detailed introduction on optimal design theory including the definitions and different terms in optimal design. We discussed about two main types of designs, namely exact designs and approximate designs. In our study, we used approximate designs since it is mathematically convenient over exact designs. We did not need to worry about integer programming problem. Moreover we can use calculus in order to solve optimization problems. Then we discussed about the way of discretizing a design space, a way of finding an optimal design space and thereby learnt some specific properties of the information matrix. Next we focused on standardized variance of the predicted response and subsequently applied this knowledge to obtain variance function plots for several optimal designs. We then discussed about several optimality criteria. Our main interest is on D-optimality and D_s-optimality criteria, under which the full set or a sub-set of parameters are of interest. D-optimality is the widely using criterion in optimal design theory because of its invariant nature under linear transformation of the scale of independent variables. D_s-optimality is a special case of D_A-optimality since it possess the same properties as D-optimality.

We formulated optimal design construction using three approaches, namely analytic approach, traditional approach and clustering approach. Analytic approach is quite challenging. We tried to solve some optimal designs for some models including one practical model in chemistry. We discussed the solutions using a class of algorithms (multiplicative algorithm). When working with class of algorithms we have to define a function which has to be positive and strictly increasing. This depends on a positive parameter which is defined as δ in our thesis. All the other significant properties related to multiplicative algorithm have been mentioned prior to the results. As for the functions, we considered $x^\delta, e^{\delta x}$, logistic CDF $e^{\delta x}/(1+e^{\delta x})$ and normal CDF $\Phi(\delta, x)$. The optimal design construction was done for the models such as quadratic, cubic, quartic regression models including a practical model in chemistry. Later, we also used some two variable models (reduced and full models). Moreover, we investigated the convergence of the algorithm using the properties of directional derivatives of the criterion function. We compared the results for each regression model using different functions.

The graphical interpretation was carried out using the plots of weights versus design points and variance function versus design points. The purpose of producing the different graphics was that we made sure each of the constructed design was correct by satisfying the fact that maximum of the variance functions cannot exceed the number of parameters of interest and the maximum occurs at the optimal support points.

The traditional approach of constructing an optimal design uses multiple iterations to assess the criterion. The other approach called the clustering approach in which the main idea is that when we run an algorithm to construct optimal design, the design turns out to be a distribution defined on disjoint sets of clusters of the design points. So we replaced the single distribution by conditional distributions and a marginal distribution across the clusters. We transformed this clustering approach to a general problem of optimization with respect to several distributions.

Finally, we did a powerful improvement in the convergence of the algorithms by combining the clustering approach and the properties of the directional derivatives. The results were promising. This approach was instrumental in improving the convergence of the algorithm and thereby allowed the model to obtain the optimal design saving more time.

7.2 Future Work

As we have seen that the clustering approach is a very powerful approach for constructing optimal design as well as for improving convergence of algorithms. We focussed on using a class of multiplicative algorithms. Note that in order to use this approach we do not have to stick with a particular algorithm. So we plan to explore the clustering approach for other types of algorithms such as Wynn's algorithm, steepest ascent, EM algorithm and Newton type iterations. In particular, in our context, Wynn's algorithm will be more appropriate as this algorithm is useful when many weights are zero at the optimum as happens in our design problems. At other contexts, when all weights are positive at the optimum (for example, if we start with a small set of design points with a prior knowledge that they are more likely to be the support points, as we considered some designs in Chapter 5), constrained steepest ascent or Newton type iterations will be appropriate.

Another possible future work could be to explore other types of optimization problems (e.g., Mandal et al. 2017) when the criterion is neither convex nor concave. This type of optimization problems are quite challenging in the sense that we cannot directly use the general equivalence theorem (optimality conditions) as the criterion is not concave. However, we could easily explore the Hessian matrix to determine the optimality conditions.

Bibliography

Al Labadi, L. and Z. Wang (2010). Modified wynn's sequential algorithm for constructing d-optimal designs: Adding two points at a time. *Communications in Statistics—Theory and Methods 39*(15), 2818–2828.

Atkinson, A., A. Donev, and R. Tobias (2007). *Optimum experimental designs, with SAS*, Volume 34. Oxford: Oxford University Press.

Atkinson, A. C. and A. N. Donev (1992). *Optimum experimental designs*. Oxford: Oxford University Press.

Atwood, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist. 40*, 1570–1602.

Box, G. E. (1982). Choice of response surface design and alphabetic optimality. In *Proceedings of the... Conference on the Design of Experiments in Army Research, Development and Testing*, Volume 28, pp. 237. US Army Research Office.

Cook, R. D. and C. J. Nachtrheim (1980). A comparison of algorithms for constructing exact d-optimal designs. *Technometrics 22*(3), 315–324.

Dette, H., A. Pepelyshev, and A. Zhigljavsky (2008). Improving updating rules in multiplicative algorithms for computing d-optimal designs. *Computational Statistics & Data Analysis 53*(2), 312–320.

Fedorov, V. V. (1972). *Theory of optimal experiments.* New York and London: Academic Press.

Johnson, M. E. and C. J. Nachtsheim (1983). Some guidelines for constructing exact d-optimal designs on convex design spaces. *Technometrics 25*(3), 271–277.

Karlin, S. and W. J. Studden (1966). Optimal experimental designs. *Ann. Math. Statist. 37*(4), 783–815.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological) 21*(2), 272–304.

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

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics 12*(5), 363–365.

Mandal, A., W. K. Wong, and Y. Yu (2015). Algorithmic searches for optimal designs. *Handbook of design and analysis of experiments*, 755–783.

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

Mandal, S. and B. Torsney (2006). Construction of optimal designs using a clustering approach. *Journal of statistical planning and inference 136*(3), 1120–1134.

Mandal, S., B. Torsney, and K. Carriere (2005). Constructing optimal designs with constraints. *Journal of statistical planning and inference 128*(2), 609–621.

Mandal, S., B. Torsney, and M. Chowdhury (2017). Optimal designs for minimising covariances among parameter estimators in a linear model. *Australian & New Zealand Journal of Statistics 59*(3), 255–273.

Rohde, C. A. (1965). Generalized inverses of partitioned matrices. *Journal of the Society for Industrial and Applied Mathematics 13*(4), 1033–1035.

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

Silvey, S. and D. Titterington (1974). A lagrangian approach to optimal design. *Biometrika 61*(2), 299–302.

Titterington, D. M. (1976). Algorithms for computing D-optimal designs on a finite design space. In *Proc. 1976 Conf. on Information Sciences and Systems*, pp. 213–216. Dept. of Elect. Eng., John Hopkins Univ., Baltimore, MD.

Torsney, B. (1977). Contribution to discussion of "maximum likelihood from incomplete data via the EM algorithm" by dempster et al. *Journal of the Royal Statistical Society B 39*, 26–27.

Torsney, B. (1983). A moment inequality and monotonicity of an algorithm. In *Semi-infinite programming and applications*, pp. 249–260. Springer.

Torsney, B. (1988). Computing optimising distributions with applications in design, estimation and image processing. optimal design and analysis of experiments (edited by y. dodge, v. v. fedorov and h. p. wynn). *Elsevier Science Publishers B. V., North Holland*, 361–370.

Torsney, B. and A. Alahmadi (1992). Further development of algorithms for constructing optimizing distributions. *Model oriented data analysis 215*, 121–129.

Torsney, B. and R. Martín-Martín (2009). Multiplicative algorithms for computing optimum designs. *Journal of Statistical Planning and Inference 139*(12), 3947–3961.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society: Series B (Methodological) 35*(1), 123–130.

Wu, C.-F. and H. P. Wynn (1978). The convergence of general step-length algorithms for regular optimum design criteria. *The Annals of Statistics*, 1273–1285.

Wynn, H. P. (1970). The sequential generation of d-optimum experimental designs. *The Annals of Mathematical Statistics*, 1655–1664.

Wynn, H. P. (1972). Results in the theory and construction of d-optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 133–147.

Yang, J., A. Mandal, and D. Majumdar (2012). Optimal designs for 2ˆ k factorial experiments with binary response. *arXiv preprint arXiv:1109.5320*.