University of Alberta

MINIMAX DESIGN FOR APPROXIMATE STRAIGHT LINE REGRESSION

by

Maryam Daemi

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences

©Maryam Daemi
Fall 2012
Edmonton, Alberta

Abstract

This dissertation first reviews the construction of an optimal design for a straight linear regression model with uncorrelated errors when the experimenter seeks protection against the biases which will accrue if her straight line model is slightly erroneous. The optimal design is derived from the minimax method, and is robust against bias caused by a small departure from the fitted model.

The study then points out a gap within the part of the minimax method related to minimizing the maximized loss function based on A- and E-optimality criteria: it is not applicable to finding an optimal design for these criteria when the emphasis is much more on the errors from bias than on those from variation.

Finally, an alternative technique is applied in order to achieve an A- and E-optimal design whether the experimenter places more emphasis on the errors from bias or on the errors from variance.

Acknowledgements

I owe my deepest gratitude to my supervisor, Dr. Wiens, who has encouraged, helped, and supported me throughout my study. His extensive knowledge and experiences were an invaluable resource for me as I undertook and completed my thesis; and without his guidance and persistent help this thesis would not have been possible. In a broader sense, the knowledge and research experience I have gained during my study with him have had a profound influence on my graduate education, and they will be a great asset in my future career.

I am also thankful to Dr. N.G.Narasimha Prasad, Dr. Csaba Szepesvari, and Dr. Linglong Kong for serving on my dissertation examination committee.

I am also grateful to my father, who passed away last spring, and to my mother, for their love and support throughout my life. And last but not least, I am thankful to my husband and best friend, Reza, for his companionship, encouragement and undivided support, which helped me to continue moving forward throughout my program of study.

Contents

List of Figures

Chapter 1: Introduction

Some experiments seek to investigate the effect of an independent variable $\mathbf{x}$, which might be a vector, on a dependent variable y; in them, an experimenter observes the response variable y corresponding to each value of the explanatory variable $\mathbf{x}$, which is under her control. In such cases, the experimenter is confronted with the problem of how to choose the values of the controlled variable $\mathbf{x}$ at which she takes the observations, in order that a limited number of observations provide the maximum information with sufficient efficiency at minimum cost. Therefore, the experimenter first needs to design an experiment by making a structure for the selection of the controlled variable $\mathbf{x}$. This issue - constructing a 'good' design - is challenging, considering some requirements of the experiment such as time, expenditure and the limitation for selected values of the controlled variable $\mathbf{x}$, which are often required in the analysis of medical data, biological or chemical data. From an ethical point of view, it is always advisable to work with a small sample size in these areas of research – for example, by having fewer patients undergoing a controversial treatment or sacrificing fewer animals in a toxicology study.

An example of such research may clarify the importance for the experimenter to

construct a 'good' design. Suppose a radiologist is interested in studying the linear effect of radiation dose-rates x on tumour shrinkage y. A wide range of dose-rates has been used in radiotherapy, expanding from a few rads per day to thousands of rads in a fraction of a second. In clinical practice, the range of dose-rates must be restricted, because an ultra-high dose-rate can harm patients by causing unacceptable damage to the tissue. On the other hand, a dose-rate that is too small might have no effect on tumour reduction. In such a situation, the experimenter should define a domain space for the explanatory variable (radiation dose-rates) x, which is the set of all possible design points – known as the design space – in advance. Suppose as well that in this research, the radiologist can only use a dose rate between 100 rads/minute and 10 rads/hour; Then the region space of the design, which is a set of all possible design points, is symbolized by $S = \left[\frac{1}{6}, 100\right]$. The available funding is only sufficient for $n = 16$ patients. In practice, we linearly transform a design space to the popular scale $[-1, 1]$, which provides easier interpretation for comparing results. The experimenter should create a design carefully to allocate 16 patients to a dose-rate and observe the outcome variable (tumour shrinkage) in order to study the relationship between these two variables with minimum cost and maximum efficiency.

In this introductory chapter, we review a technical definition of 'good' design, namely optimal design, and some other basic concepts in the design of a study such as exact design, and approximate design. We describe some requirements for achieving an optimal design, including such criteria as I-optimality. Next, we review both classical and robust approaches to finding an optimal design. Within the robust approach, we review the minimax method, which is the main subject of this dissertation.

In the second chapter, we apply the minimax method to achieve an optimal design for three different criteria, A-, E-, and I-optimality. Then we address a gap in the existing theory applied within the part of the minimax method using two criteria, A-, and E-optimality. In the final chapter, we apply a technique to remedy this gap.

1.1 Exact versus Approximate Design

In this section, we define exact and approximate design, and clarify the difference between these two types of designs. This distinction is important for understanding the content of this research project, and it suggests the reasons why we work with approximate rather than exact designs.

Let us suppose that an experimenter wishes to select n various values of the specified predictor variables $\mathbf{x}$ and observes the corresponding values of a dependent variable y. An implementable design ξ shows the experimenter which values of the input variable $\mathbf{x}$, with how many repetitions, should be selected from a "design space" S. Let $\mathbf{x}_1, \ldots, \mathbf{x}_m$ be individual points of S, and $m \leq n$. The integer value $n \times \xi(\mathbf{x}_i)$ for $i \in \{1, \ldots, m\}$ is the number of repeated observations at each particular point $\mathbf{x}_i$ in the experiment. This implementable design ξ is called an exact design which determines the frequency n_i at particular point $\mathbf{x}_i$ where the sum of these frequencies

n_i is n. The exact design thus defines a discrete probability measure,

$$\xi_n = \left\{ \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_m \\ \omega_1 = \frac{n_1}{n} & \omega_2 = \frac{n_2}{n} & \ldots & \omega_m = \frac{n_m}{n} \end{array} \right\},$$

$$\xi_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} n_i \, I_{\mathbf{x}_i}(\mathbf{x}),$$

where $I_{\mathbf{x}_i}(\mathbf{x})$ denotes the pointmass 1 at $\mathbf{x}_i$, for all points $\mathbf{x}_i \in S$.

In practice, only discrete (or exact) designs are implementable, since the weights of design points are a ratio of integers. But mathematically an exact design cannot be computed easily, and it might come from a complicated optimization problem such that it is intractable even for a simple situation.

As discussed later, the optimal design depends on the fitted model and optimality criterion. But the exact optimal design also depends on the sample size n. That is, there are different exact optimal designs for different sample sizes using an identified model optimality criterion.

Kiefer (1959) introduced the theory of continuous design in which he eliminated the requirement of the ω_i being a ratio of integers (namely a multiple of $\frac{1}{n}$). The approximate or continuous design is defined with continuous weight

$$\xi = \left\{ \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_m \\ \omega_1 & \omega_2 & \ldots & \omega_m \end{array} \right\},$$

$$0 \le \omega_i \le 1, \ \sum_{i=1}^{m} \omega_i = 1.$$

The essential advantage of an approximate design is that we can compute an implementable optimal design from an optimal approximate design by rounding its $n\omega_i$'s off to the value of the nearest integer. Pukelsheim and Rieder (1992) introduced the "efficient rounding method" to round an approximate design into an exact design.

In fact an implementable exact design can be approximately obtained easily using all calculus techniques in the continuous domain. Of course, the exact design approximated by rounding off a continuous design for implementation is not unique. But if the sample size becomes large, the implemented optimal approximate design is always close to the optimal exact design.

1.2   Optimal Design

As noted above, a 'good' design is desirable for an experimenter in order to achieve the most efficient result at minimal cost. In order to have a technical definition of 'good' design, let us suppose that the following linear regression model completely describes the relationship between the predictor $\mathbf{x}$ and the dependent variable y:

$$y = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \qquad \mathbf{x} \in S, \tag{1.1}$$

where $\mathbf{z}(x) = (1,\ x)^T$, $\boldsymbol{\theta} = (\theta_0,\ \theta_1)^T$, and the error terms ε_i are uncorrelated with zero mean and a common variance σ_ε^2 (that is homoscedasticity). In accordance with the previous example, the predictor variables x is radiation dose-rate which is defined on the design space $S = \left[\frac{1}{6}, 100\right]$, and the dependent variable y is tumour shrinkage.

The radiologist assumes that a simple linear regression model describes the relation between radiation dose-rate and tumour shrinkage.

The structure of choosing n various values of the predictor variable $\mathbf{x}$ from the design space S in an experiment is completely explained by a design ξ, which is a discrete probability measure. Indeed, an implementable design ξ determines which points of the design space S, with how many repetitions, should be selected.

Let the class Ξ be the set of all possible designs of an experiment, which can be performed to obtain n points $\mathbf{x}_1, \ldots, \mathbf{x}_n$, from the design space S. An optimal design is one of the designs belonging to the class Ξ in which a certain criterion is optimized. Hence, we need to define a sensible criterion based on (1.1) to compare all designs in the class Ξ; this is called optimality or design criterion. Indeed, an optimality criterion illuminates how good a design is. It means that the criterion is minimized or maximized by an optimal design.

1.2.1 Classical Approach to Acquiring an Optimal Design

The classical approach was the first method used to compute an optimal design. Suppose an experimenter measures water purity as a function of input variable $x = chlorine$, linearly transformed so that $S = [-1, 1]$. The dependent variable y (*water purity*) is expected to change whenever the independent variable (*chlorine*) is altered, through a vector $\mathbf{z}(x)$ of regressors. Suppose the experimenter fits, by the least squares, a simple straight line regression model and she is also confident that the model fits on

the data. Then

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i = \mathbf{z}^T(x_i)\boldsymbol{\theta} + \varepsilon_i, \tag{1.2}$$

$$E(y_i|x_i) = \theta_0 + \theta_1 x_i = \mathbf{z}^T(x_i)\boldsymbol{\theta},$$

where $\mathbf{z}(x_i) = (1,\ x_i)^T$, $\boldsymbol{\theta} = (\theta_0,\ \theta_1)^T$, and the error terms ε_i are uncorrelated with zero mean and a common variance $\sigma_{\boldsymbol{\varepsilon}}^2$. Define $\mathbf{Z}^T = [\mathbf{z}(x_1), \ldots, \mathbf{z}(x_n)]$ (which is called the design matrix) and $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^T$. Based on the Least Squares method, if $\mathbf{Z}^T\mathbf{Z}$ is nonsingular, the estimate

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\ \mathbf{Y},$$

(based on the Gauss–Markov theorem) is the best estimate of $\boldsymbol{\theta}$ among all linear unbiased estimates. Here "best" means giving the lowest possible variance of the estimate where

$$E(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta},$$

$$\mathbf{COV}\left(\widehat{\boldsymbol{\theta}}\right) = \sigma_{\varepsilon}^2 \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}.$$

Hence, in order to obtain a more efficient estimate, a "smaller" covariance matrix is required. Since the covariance matrix depends only on the data through $\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}$, the efficiency of the estimator $\widehat{\boldsymbol{\theta}}$ can be completely determined by the design ξ. In this example, we focus on minimizing a scalar-valued function of the covariance matrix of

the parameter estimates,

$$\mathbf{COV}\left(\widehat{\theta}\right) = \sigma_\varepsilon^2 \left(Z^T Z\right)^{-1} \tag{1.3}$$

$$= \frac{\sigma_\varepsilon^2}{n \sum x_i^2 - n^2 \overline{x}^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}.$$

In terms of $S_x^2 = \frac{\sum(x_i - \overline{x})^2}{n}$, we have

$$nVAR\left(\widehat{\theta}_1\right) = \frac{\sigma_\varepsilon^2}{S_x^2}, \quad nVAR\left(\widehat{\theta}_0\right) = \sigma_\varepsilon^2(1 + \frac{\overline{x}^2}{S_x^2}).$$

Therefore, both variances are minimized by putting half of the x's at each of ± 1. In other words, the probability distribution of the optimal design is

$$\xi\left(-1\right) = \xi\left(1\right) = \frac{1}{2}.$$

Another scalar function of the covariance matrix is

$$\det\left[\mathbf{COV}\left(\widehat{\theta}\right)\right] = \frac{\sigma_\varepsilon^4}{n \sum x_i^2 - n^2 \overline{x}^2}, \tag{1.4}$$

again this value is minimized by putting half of the x's at each of ± 1. In fact, for a straight line regression, the design

$$\xi\left(-1\right) = \xi\left(1\right) = \frac{1}{2},$$

is optimal with respect to a wide range of criteria.

These methods are called classical approaches to finding the optimal design. They assume that the proposed model (1.2) is a correct model in which the error terms ε_i are uncorrelated. In fact, the optimal design is constructed based on minimizing the covariance matrix or a measure of the size of the covariance matrix. In particular, the determinant is used as a measure of the size of the covariance matrix in (1.4). These optimality criteria (1.3) and (1.4) construct a design that concentrates all mass at extreme points of the design space S when a straight line with intercept is being fitted.

The classical approach has been discussed in great detail, using diverse optimality criteria (which are also known as alphabetical optimality criteria), by several authors. From a historical point of view, Smith (1918) was one of the first to define a specific optimality criterion to obtain optimal designs on a fixed design space for polynomial regression models up to degree six. Later on, Smith's criterion was called global optimality (and was also called G-optimality by Kiefer (1959)). Her optimality criterion was the worst-case prediction error over the design space to be minimized,

$$\min_{x_i,\ i=1,\ldots,n}\ \max_{x\in S} VAR\left(\widehat{y_x}\right).$$

Another optimality criterion, proposed by Wald (1943), is popular and important in life applications. This criterion is defined as maximizing the determinant of the matrix $\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}$, and it is known as the determinant criterion (also called D-optimality by

Kiefer(1959)). It is optimized, as follows:

$$\min \det \left[\left(\mathbf{Z}^T \mathbf{Z} \right)^{-1} \right] = \max \det \left(\mathbf{Z}^T \mathbf{Z} \right).$$

Based on the above definition, D-optimality is essentially a parameter estimation criterion, whereas G-optimality is a response estimation criterion. The equivalence of these two alphabetic criteria, G- and D-optimality, under certain conditions was proved in the Equivalence Theorem by Kiefer and Wolfowitz (1960). The optimality criteria A-optimality and E-optimality which put an emphasis on the quality of the parameter estimate, as does D-optimality, were introduced respectively by Chernoff (1953) and Ehrenfeld (1955). The criterion A-optimality is the average variance of the estimates of the regression coefficients, and E-optimality is the maximum variance of all possible normalized linear combinations of parameter estimates. They are optimized, in order, as follows:

$$\min \operatorname{tr} \left[\left(\mathbf{Z}^T \mathbf{Z} \right)^{-1} \right],$$

$$\min \ ch_{\max} \left[\left(\mathbf{Z}^T \mathbf{Z} \right)^{-1} \right] = \max \ ch_{\min} \left(\mathbf{Z}^T \mathbf{Z} \right),$$

where $ch_{\max} \left[\left(\mathbf{Z}^T \mathbf{Z} \right)^{-1} \right]$ is the maximum characteristic root of the matrix $\left[\left(\mathbf{Z}^T \mathbf{Z} \right)^{-1} \right]$, $ch_{\min} \left(\mathbf{Z}^T \mathbf{Z} \right)$ is the minimum characteristic root of the matrix $\left(\mathbf{Z}^T \mathbf{Z} \right)$.

Other criteria have been introduced and studied by various authors. These include c-optimality (Elfving (1952)), Generalized D-optimality or D_A-optimality (Sibson (1974)), L-optimality (Fedorov (1972)) and so on, in a huge literature. But certain

of the criteria, including G- and D-optimality, have been studied in more depth. Since all of the above mentioned design criteria are related to the matrix $(\mathbf{Z}^T \mathbf{Z})$, which is proportional to the inverse of the covariance matrix for the least-squares estimates of the estimated parameters in the linear regression model (1.1), they are also known as information-based criteria. But for parametric models, these criteria are sometimes called the loss function.

Although some optimality criteria such as G-, D-, A- and E-optimality are quite reasonable measures for obtaining optimum design in the classical approach, the severe practical disadvantages of using these criteria in the classical optimal design problems should be noted. As the results of the *water purity* and *chlorine* example shows, the optimum designs of both criteria, variance of the parameter estimates, and D-optimality for the linear regression model (1.2) concentrate all mass at the extreme points, and the experimenter cannot explore all of the points within the design space. On the other hand, the result of the D-optimum model is extremely model dependent; for if the linear regression model (1.2) is inadequate and a quadratic regression model is more appropriate, it is impossible to detect any curvature (the presence of the term x^2) using just the end points of the design space. In fact, the optimal design based on the classical approach is so dependent on the proposed model that it is impossible to check the adequacy of the model.

As stated above, in a classical optimal design problem, it is assumed that the model suggested by experimenter (1.1) is correctly specified and also it is assumed that the error terms ε_i are uncorrelated. However, such assumptions may not be satisfied in reality, and the violation of either assumption can occur. In this dissertation, our

attention is focused on the case in which the proposed model (1.1) has deviation from the true model. In such a case, as noted above, it is impossible to detect the deviation, if the optimal design is achieved by classical approach. That is why it is necessary to study another method which could be applied in this situation. This alternative approach is called the robust approach.

1.2.2   Acquiring a Robust Optimal Design

In the previous example related to *water purity*, notwithstanding the experimenter's choice that is only approximate:

$$E(Y \mid x) \approx \mathbf{z}^T(x)\boldsymbol{\theta}, \qquad x \in S, \tag{1.5}$$

it is more realistic that we define a precise model

$$E(Y \mid x) = \mathbf{z}^T(x)\boldsymbol{\theta} + f(x), \qquad x \in S, \tag{1.6}$$

for some unknown but small non-linear function f belonging to some class $\mathcal{F}$. The parameter $\boldsymbol{\theta}$ is not identifiable, since we might equally well write the alternate models

$$E(Y \mid x) = \mathbf{z}^T(x)\left(\boldsymbol{\theta} + \phi\right) + \left(f(x) - \mathbf{z}^T(x)\phi\right), \qquad x \in S.$$

To avoid this problem, we define the "true" $\boldsymbol{\theta}$ as that which makes (1.5) most accurate:

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\alpha}} \int_S \left\{E(Y \mid x) - \mathbf{z}^T(x)\boldsymbol{\alpha}\right\}^2 dx. \tag{1.7}$$

12

Taking the derivative of (1.7):

$$2 \int_S \mathbf{z}(x) \left\{ E(Y \mid x) - \mathbf{z}^T(x)\boldsymbol{\theta} \right\} dx = 0,$$

leads us to the orthogonality property for $f(x)$,

$$\int_S f(x)\mathbf{z}(x)dx = 0. \tag{1.8}$$

In such a situation, we have not only a variance due to sampling variation

$$\mathbf{COV}\left(\widehat{\boldsymbol{\theta}}\right) = \sigma_\varepsilon^2 \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1} = \frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1},$$

(where $\mathbf{B} = \sum \frac{n_i}{n}\mathbf{z}(x_i)\mathbf{z}^T(x_i) = E_\xi\left[\mathbf{z}(x)\mathbf{z}^T(x)\right]$ depends only on the design ξ, not on f) but also a bias due to the inadequacy of the linear regression model,

$$\begin{aligned}
\mathbf{bias} &= E\left(\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\left(\mathbf{Z}\boldsymbol{\theta} + f + \varepsilon\right)\right) - \boldsymbol{\theta} \\
&= n\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\sum\frac{n_i}{n}\mathbf{z}(x_i)f(x_i) \\
&= \mathbf{B}^{-1}\mathbf{b}(f,\xi),
\end{aligned}$$

where $\mathbf{b}(f,\xi) = \sum \frac{n_i}{n}\mathbf{z}(x_i) f(x_i) = E_\xi[\mathbf{z}(x) f(x)]$. Therefore the Mean Squared Error (MSE) matrix of $\widehat{\boldsymbol{\theta}}$,

$$\mathbf{MSE} = E\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^T\right]$$

$$= \mathbf{COV}\left(\widehat{\boldsymbol{\theta}}\right) + \mathbf{bias} \cdot \mathbf{bias}^T,$$

might be written as a function of f and ξ, namely:

$$\mathbf{MSE} = \frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{b}(f,\xi)\mathbf{b}^T(f,\xi)\mathbf{B}^{-1}. \tag{1.9}$$

In this thesis we consider a design (ξ) with its design measure, which is a probability measure $\xi(dx)$ on S.

Box and Draper (1959) considered a similar case in which an experimenter fits a polynomial or a multinomial linear regression model on the data in the circumstance where the true response has higher degree. They adopted the minimization of mean square error of $\widehat{y}$ integrated over a region of interest S as a basic criterion. This criterion is known as $IMSE$ or I-optimality, and it involves both variance and bias. Their unexpected result (at least in the cases considered) was that "......the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if *variance were ignored completely* and the experiment designed so as to *minimize bias alone*." (Box and Draper [1959, p. 622])

For that reason, we need to study optimal designs which are not sensitive to small deviations from the model proposed by an experimenter. In other words, we are in-

terested in designing an experiment that works reasonably well not only in the fitted model, but also in a small neighbourhood around the fitted model. A method of achieving an optimal design considering the contamination term of the true model (1.6) is called robust approach, as distinguished from the classical approach to constructing an optimal design. After the publication of Box and Draper's paper (1959), a series of papers were spawned; in these papers, the authors constructed optimal designs which are robust against bias caused by model misspecification. Based on Box and Draper's results, some authors such as Karson, Manson, and Hader (1969) suggested the choice of the estimator according to minimizing the average squared bias. But Stigler (1971) explained that "minimum bias designs, while they are an important attempt to meet realistically the problem of checking the representational adequacy of the model, may often be inappropriate, inefficient, or both."(Stigler [1971, p. 313]), and he found the optimal designs which permit a check of the adequacy of the model.

The criteria of the robust approach are generally functions of the Mean Squared Error matrix of the estimator (**MSE**), and as shown in (1.9), **MSE** is a function of f and ξ $(\mathbf{M}\,(f,\xi))$.

Minimax is a well-known and well-studied method in the robust approach of achieving optimal design. Considering any function of **MSE** (as a function of f and ξ $(\mathcal{L}\,(f,\xi))$), the minimax method aiming to find ξ^*(optimum design) might be written as follows,

$$\xi^* = \min_{\xi} \max_{f} \; \mathcal{L}\,(f,\xi)\,.$$

Based on the orthogonality condition of the function f and the regressors $\mathbf{z}(x)$

(1.8), the parameter $\boldsymbol{\theta}$ is uniquely defined, provided the matrix $\int_S \mathbf{z}(x)\mathbf{z}^T(x)dx$ is non-singular. In addition, the following boundary condition on the disturbance function $f(x)$ was used by Huber (1975) and Wiens (1990) to prevent the domination of the error due to bias over the error due to variance

$$\int_S f^2(x)dx \leqslant \eta, \tag{1.10}$$

where the radius η is assumed known and small.

Huber (1975) defined a class $\mathcal{F}$ with (1.10) and (1.8) properties in order to construct a specific neighbourhood around a straight line regression as a fitted model, and he found an optimal robust design (from a class of symmetric designs) with respect to the Lebesgue measure on $S = \left[-\frac{1}{2}, \frac{1}{2}\right]$ by minimizing the maximum "Integrated Mean Squared Error" of the predictions (or I-optimality as a loss function),

$$IMSE = \int_S MSE(\widehat{Y}(x))dx,$$

as the true model varying over the full neighborhood where

$$MSE(\widehat{Y}(x)) = E\left(\widehat{Y}(x) - E(Y|x)\right)^2.$$

(It is noteworthy that he used the standard least squares estimates based on the model (1.1).) Huber's contamination class was criticized by Marcus and Sacks (1976) and Li and Notz (1982) as being too large. They pointed out that any discrete design has infinite maximum loss in this class; their claim was proved by Wiens (1992),

who demonstrated that only absolutely continuous design measures with respect to Lebesgue measure on the design space S have finite maximum loss in Huber's contamination class. In fact, Huber restricted himself to absolutely continuous design measures with respect to Lebesgue measure on the design space S. Thus, using Huber's contamination class, we achieve an optimal continuous design (density). This optimal density leads us to another problem; that is, how an implementable design can be extracted from the optimal density. (Computing an implementable design from the optimal density is explained briefly in the next chapter.) However, as Wiens (1992) noted, *"our attitude is that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse."*

Wiens (1990, 1992) extended Huber's work and found a minimax design for multiple linear regression for five different loss functions, I-, A-, E-, G- and D-optimality. He described the neighbourhood of a fitted model in a general form using the contamination class

$$\mathcal{F} = \left\{ f \mid \|f\|_2 = \left(\int_S f^2(x)dx \right)^{\frac{1}{2}} \leqslant \eta, \ \int_S f(x)\mathbf{z}(x)dx = 0, \ x \in S \right\}. \qquad (1.11)$$

To achieve optimal robust designs, he applied Mean Squared Error (**MSE**) rather than variance of the estimate using A-, E-, G- and D-optimality.

Marcus and Sacks (1976) defined a contamination class

$$\acute{\mathcal{F}} = \left\{ f \mid |f(x)| \leqslant \phi(x), \ x \in S \right\}, \qquad (1.12)$$

17

for a straight line regression on $S = [-1, 1]$, where $\phi(x)$ is a given function on S with $\phi(0) = 0$. This class seems to be thinner than Huber's class. They found estimates and designs minimizing the maximum mean squared error,

$$\sup_{f \in \mathcal{F}} E\left(\left(\widehat{\theta}_0 - \theta_0 \right)^2 + \alpha^2 \left(\widehat{\theta}_1 - \theta_1 \right)^2 \right),$$

where $\widehat{\theta}_0$ and $\widehat{\theta}_1$ (not necessarily obtained by the least squares method) denote the estimates of θ_0 and θ_1, and α is a specific constant. Marcus and Sacks found an optimal exact design for some choices of $\phi(x)$. If $\phi(x) \geq mx$ (m is constant) then the (unique) optimal exact design is on the points $\{-1, 0, +1\}$. If ϕ is a convex function there is a wide range of cases for which a design can be found on two points $\{-z, +z\}$ where z depends on ϕ and θ. Since their robust designs are constructed based on the concentration of all mass on a small number of, generally extreme, points over the specified design space, the experimenter has little or no ability to check the validity of the fitted model.

Li and Notz (1982) modified and combined the clever ideas of Marcus and Sacks (1976) and of Huber (1975) to get optimal designs for multivariate straight line regression (where $\mathbf{z}(x_i) = (1, x_{i1}, \ldots, x_{ik})^T$, and $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_k)^T$) on $S = [-1, 1]^k$ using the contamination class

$$\mathcal{F}' = \left\{ f \mid |f(\mathbf{x})| \leqslant \phi(\mathbf{x}), \int_S f(\mathbf{x})\mathbf{z}(\mathbf{x})d\mathbf{x} = \mathbf{0}, \ \mathbf{x} = (x_1, \ldots, x_k) \in S \right\}.$$

The second condition again ensures the uniqueness of the θ_i's in model (1.6). Li and

18

Notz minimized the following weighted mean squared error:

$$\sup_{f \in \mathcal{F}} E\left(\alpha_0^2 \left(\widehat{\theta}_0 - \theta_0 \right)^2 + \sum_{i=1}^{k} \alpha_i^2 \left(\widehat{\theta}_1 - \theta_1 \right)^2 \right),$$

where $\widehat{\theta}_0$ and $\widehat{\theta}_1$ denote the estimates of θ_0 and θ_1 (not necessarily the standard least squares estimate), and α_i's are specific constants. Furthermore, designs were restricted to have finite support. The result for optimal design puts point mass $\frac{1}{2^k}$ on each of the 2^k corners of the cube $S = [-1, 1]^k$. This result has the same shortcoming as the previous result in which the robust design only covers the extreme points of S.

Other authors such as Sacks and Ylvisaker (1978), Li (1984) and Pesotchinsky (1982) used the contamination class (1.12). Pesotchinsky (1982) also extended Marcus and Sacks's work (1976), and he found a minimax design for multiple linear regression using the $\Phi p - family$,

$$\Phi p = \left[\frac{1}{k+1} \, tr \left(\mathbf{M} \left(f, \xi \right)^p \right) \right]^{\frac{1}{p}}$$

$$= \left\{ \frac{1}{k+1} \sum_{j=0}^{k} \lambda_j^p \left(f, \xi \right) \right\}^{\frac{1}{p}},$$

(where $\mathbf{M}\left(f, \xi \right)$ is the mean squared error matrix ($\mathbf{MSE}$) and $\lambda_0 \left(f, \xi \right) \leq \lambda_1 \left(f, \xi \right) \leq \ldots \leq \lambda_k \left(f, \xi \right)$ are the eigenvalues of $\mathbf{M} \left(f, \xi \right)$) as criteria (introduced by Kiefer (1974, 1975)), in particular criteria $\Phi_0, \Phi_1, \Phi_\infty$ corresponding respectively to D- , A- and E-optimality. Like Huber, he confined himself to the use of the standard least squares estimates. He used the contamination class (1.12) for multivariate covariates where $\phi(\mathbf{x})$ is a convex function of $\|\mathbf{x}\|^2 = x_1^2 + \cdots + x_k^2$. He also confined his consideration

19

to the class $\Xi(m)$ of all symmetric designs ξ with fixed second moment $E_\xi(x_i^2) = m$. Then he showed that, given all of the above conditions, any symmetric design ξ_0 belonging to the class $\Xi(m)$ will be D-optimal, if it is supported only by the points of sphere S_R of radius $R = \sqrt{mk}$. He also found unique optimal symmetric designs for A- and E-optimality criteria which correspond to the uniform continuous measures on appropriate spheres.

In general, the robust optimal designs, found in these papers, are quite sensitive to the choice of ϕ, and they tend to concentrate all mass function at a small number of, generally extreme, points in the design space such that we have little or no chance to test the validity of the fitted model.

1.3   Purpose of This Research

In this introductory chapter, we have reviewed some basic concepts of regression design such as exact and approximate designs. We have also described the definition of an optimal design, its requirements, and two different types of approaches, classical and robust, to achieving an optimal design by reviewing some statisticians' contributions. The main concentration in the robust approach was on the minimax method.

In the next chapter, we will review the minimax method, which has been applied by Huber (1975) and Wiens (1992), in order to achieve the robust optimal design (or optimal density) for an approximate straight linear regression (1.2) over the design space $S = [-\frac{1}{2}, \frac{1}{2}]$, using I-, A- and E-optimality as optimality criteria. The model

has been transformed in such a way that the design space is $S = [-\frac{1}{2}, \frac{1}{2}]$. We should note that although this transformation provides easier computation and easier interpretation for comparing results, for A- and E-optimality the optimal robust design obtained from the original design space and the optimal robust design of its transformation $\left(S = [-\frac{1}{2}, \frac{1}{2}]\right)$ are different. This is true even though we apply the same method to achieve these optimal robust designs.

Next, we will briefly show how an exact design can be extracted from optimal density using the methods discussed in Xu and Yuen (2011). Based on the definition of minimax design using Huber's class, the algorithm for acquiring optimal design is divided into two main steps:

1. Maximizing the loss function over the contamination class $\mathcal{F}$, defined at (1.11).

2. Minimizing the maximum loss over the class of all possible symmetric absolutely continuous designs over the design space $S = [-\frac{1}{2}, \frac{1}{2}]$, in order to achieve optimal density.

Wiens (1992) showed that using the existing theory in the second part of this algorithm for A- and E-optimality does not allow us to find a robust design on the design space S when there is much more emphasis on the errors from bias than on those from variation. He pointed out that this problem arises for multiple (even simple) linear regressions without interactions and also for quadratic regressions with an intercept. In this dissertation, we consider only a simple linear regression model. In the last chapter, we apply an alternative method in the second step of the minimax problem to find A- or E-optimal design to satisfy experimenters' preferences to place

more emphasis on the errors from bias or the errors from variation (the experimenter's preference is determined by the variable ν, which is defined in the next chapter). In this method, the form of the density of the minimax design is similar to the optimal density introduced by Shi, Ye and Zhou (2003). In their study, they used the Lagrange Multiplier Rule from nonsmooth optimization theory.

Chapter 2: Minimax Design – a Review of Existing Results

In the previous chapter, we reviewed two strategies – the classical and robust approaches – for achieving the best design (optimal design) for an experiment in order to study the relationship between two variables with minimum cost and maximum efficiency. As noted, this dissertation focuses on the robust approach rather than the classical one.

Minimax, as noted, is a method employed in achieving robust optimal designs which minimize the maximum (over f) loss function, where the loss functions are monotonic functions of the mean squared error matrix. In this chapter, we explain all of the steps of the minimax method in order to determine the optimal design for three different criteria, I-, A-, and E-optimality.

Suppose that an experimenter wants to find an optimal design in order to study the relationship between the specified predictor variables x over the design space $S = [-\frac{1}{2}, \frac{1}{2}]$ and the corresponding observed values of a dependent variable y. The

experimenter fits, by least squares, a linear regression model

$$E(y \mid x) = \mathbf{z}^T(x)\boldsymbol{\theta}, \qquad x \in S.$$

But she suspects that there is some nonlinearity in the regression response, and she thinks that the above model is only approximately true, and that

$$E(y \mid x) = \mathbf{z}^T(x)\boldsymbol{\theta} + f(x), \qquad x \in S,$$

where the small non-linear function f belongs to the class (1.11). The least squares estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}\mathbf{Y},$$

and its mean squared error matrix (1.9) is presented as a function of f and ξ.

As discussed in the previous chapter, Huber's class $\mathcal{F}$ was criticized for this reason: Huber's class $\mathcal{F}$ is very large, and consequently, the loss function $\mathcal{L}(f, \xi)$ (which is a function of $\mathbf{M}(f, \xi)$) of this class has an infinite supremum for any discrete design. In fact, in order to have finite maximum loss, it is necessary that ξ be absolutely continuous with a density $m(x)\,(= \xi'(x))$ when we use Huber's class. This required property of ξ was proven by Wiens (1992) and refined by Heo, Schmuland &Wiens (2001). Wiens assumed the following two properties for the loss function $\mathcal{L}(f, \xi)$ in order to prove the absolutely continuous condition for ξ.

(L1) Monotonicity:

If $\mathbf{M}(f_1, \xi) \geqslant \mathbf{M}(f_2, \xi)$, in the sense of positive semidefiniteness then $\mathcal{L}(f_1, \xi) \geqslant$

$\mathcal{L}(f_2, \xi);$

(L2) Unboundedness:

If the maximum characteristic root of $\mathbf{M}(f, \xi)$, $ch_1(\mathbf{M}(f_n, \xi))$, tends to infinity as $n \to \infty$, then $\mathcal{L}(f_n, \xi) \to \infty$.

He pointed out that in discrete cases, there are sets of Lebesgue measure zero to which ξ assigns positive mass ($\xi(\{x_i\}) > 0$ for $i = 1, \ldots, n$). Then, he assigned a very large value to f at these points such that the bias $\mathbf{b}(f, \xi)$ goes to infinity while having no effect on the integral $\int_S f^2(x)dx$ or $\int_S \mathbf{z}(x) f(x) dx$. Via this restriction on the design ξ, we have the following result for $\mathbf{B}$ and $\mathbf{b}(f, \xi)$ using the density $m(x)$ of the design $\xi(x)$,

$$\mathbf{B} = E_\xi \left[\mathbf{z}(x) \mathbf{z}^T(x) \right] = \int_S \mathbf{z}(x) \mathbf{z}^T(x) m(x) dx,$$

and

$$\mathbf{b}(f, \xi) = E_\xi \left[\mathbf{z}(x) f(x) \right] = \int_S \mathbf{z}(x) f(x) m(x) dx.$$

In order to solve the first step of the minimax method (maximizing the loss function over the class $\mathcal{F}$), Wiens (1992) used the following theorem to reduce the maximization problem over the infinite dimensional class $\mathcal{F}$ to a finite dimensional parametric problem.

Define $p \times p$ matrices:

$$\mathbf{K} = \int_S \mathbf{z}(x) \mathbf{z}^T(x) m^2(x) dx,$$

where p is the number of regressors, and

$$G = K - H,$$

where $\mathbf{H} = \mathbf{BA}^{-1}\mathbf{B}$, and $\mathbf{A} = \int_S \mathbf{z}(x)\mathbf{z}^T(x)\,dx$.

The $p \times p$ matrix $\mathbf{G}$ is positive semi-definite, and it might be singular.

Theorem 1 *Define*

$$\mathbf{r}(x) = \eta\, \mathbf{G}^{\frac{-1}{2}}\left(m(x)\mathbf{I}_p - \mathbf{BA}^{-1}\right)\mathbf{z}(x),$$

$$\mathcal{H} = \left\{h_{\boldsymbol{\beta}}(x) = \mathbf{r}^T(x)\boldsymbol{\beta}; \|\boldsymbol{\beta}\| = 1\right\},$$

where $\mathbf{G}$ *is assumed to be non-singular. Then*

(i) $\mathcal{H} \subset \mathcal{F}$, *with* $\int_S h_{\boldsymbol{\beta}}^2(x)\,dx = \eta^2$.

(ii) $\sup_{f \in \mathcal{F}} \mathcal{L}(f, \xi) = \sup_{\|\boldsymbol{\beta}\|=1} \mathcal{L}(h_{\boldsymbol{\beta}}, \xi)$.

In other words, to find f such that it maximizes the loss function over $\mathcal{F}$, we can focus only on the bias term; and in order to find maximum bias, we take into account those members which belong to $\mathcal{H}$.

As a result of this theorem, we have

$$b(h_{\boldsymbol{\beta}}, \xi) = \eta\mathbf{G}^{\frac{1}{2}}\boldsymbol{\beta}.$$

Based on the definition of a maximum eigenvalue, the "max" is actually attained in $\mathcal{H}$. In addition, based on the proof of this theorem, there is a one-to-one map from

26

the class $\mathcal{F}$ to the class $\mathcal{H}$. Therefore, we can use "max" rather than "sup" over the class of $\mathcal{F}$ and $\mathcal{H}$ in Claim *(ii)* of this theorem.

If we have a design such that its matrix **G** is singular (such as the continuous uniform design), we slightly change the matrix **G** to make it an invertible matrix, and then pass to the limit (more details are available in Heo, Schmuland &Wiens (2001)).

Furthermore, there is another limitation involving the design ξ which provides for easier computation of the optimal density. We assume that all designs ξ's are symmetric. Let the class Ω be a set of all possible symmetric absolutely continuous designs.

In light of the above restrictions, the following sections in this chapter review the minimax method to find the optimal design using I-, A-, and E-optimality criteria.

2.4   Minimax Design Using I-optimality

The loss function corresponding to the I-optimality ($IMSE$) is the Integrated Mean Squared Error of the fitted responses $\widehat{Y}(x)$

$$IMSE = \int_S MSE(\widehat{Y}(x))dx,$$

which is a monotonic function of the Mean Squared Error of the estimate.

$$IMSE = \int_S E\left\{\widehat{Y}(x) - E(Y|x)\right\}^2 dx$$

$$= \int_S \mathbf{z}^T(x) \ \mathbf{MSE} \ \mathbf{z}(x) \, dx$$

$$- 2\left(\int_S f(x)\mathbf{z}^T(x)\, dx\right)\left(E\left[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right]\right) + \int_S f^2(x)\, dx,$$

where $\mathbf{MSE} = E\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^T\right]$.

The integral in the cross-product vanishes based on (1.8). Then

$$IMSE = tr\left[\mathbf{A} \cdot \mathbf{M}(f, \xi)\right] + \int_S f^2(x)\, dx.$$

Now based on the result (1.9), we have

$$IMSE = \frac{\sigma_\varepsilon^2}{n} tr\left\{\mathbf{A}\mathbf{B}^{-1}\right\} + \mathbf{b}^T(f, \xi)\mathbf{H}^{-1}\mathbf{b}(f, \xi) + \int_S f^2(x)\, dx.$$

The $IMSE$ is a function of f and ξ. From this point, we use $\mathcal{L}_I(f, \xi)$ instead of $IMSE$.

The loss function based on I-optimality has both conditions (L1) and (L2). The loss function is a monotonic function of $\mathbf{M}(f, \xi)$ due to the monotonic property of the trace. Furthermore, since the trace of a matrix tends to infinity when the maximum characteristic root of the matrix tends to infinity, the limit of $\mathcal{L}_I(f_n, \xi)$ tends to infinity, if $ch_1(\mathbf{M}(f_n, \xi)) \to \infty$. Hence, we have a finite loss function only for absolutely continuous design ξ, and we aim to find optimal design over the class

28

of all symmetric absolutely continuous designs Ω.

In the following subsections (four steps), to achieve an optimal design, we apply the following sets of operations on the loss function as a function of ξ and f, respectively.

2.4.1 Maximizing $IMSE$ over f

In this step, we assume that the design ξ is fixed, and we maximize $IMSE$ based on f over the class $\mathcal{F}$,

$$\max_f \mathcal{L}_I\left(f, \xi\right) = \frac{\sigma_\varepsilon^2}{n} tr\left\{\mathbf{A}\mathbf{B}^{-1}\right\} + \max_f \left(\mathbf{b}^T\left(f, \xi\right)\mathbf{H}^{-1}\mathbf{b}\left(f, \xi\right) + \int_S f^{\,2}\left(x\right) dx\right).$$

In fact, we are seeking to obtain the worst possible case in which the contamination term makes the true model very far from the fitted model, such that we obtain a model with maximum bias for the parameters.

To maximize the loss function, we apply Theorem 1 and its result $\left(\mathbf{b}(h_{\boldsymbol{\beta}}, \xi) = \eta\mathbf{G}^{\frac{1}{2}}\boldsymbol{\beta}\right)$, in order to reduce the size of the search for f. Instead of maximizing f over the infinite dimensional class $\mathcal{F}$, we maximize it over the finite dimensional class $\mathcal{H} = \left\{h_{\boldsymbol{\beta}}\left(x\right) = \mathbf{r}^T(x)\boldsymbol{\beta}; \|\boldsymbol{\beta}\| = 1\right\}$.

Moreover, based on this Theorem, $\mathcal{H}$ is a subset of $\mathcal{F}$, with the properties

$$\int_S h_{\boldsymbol{\beta}}^2\left(x\right) dx = \eta^2, \text{and} \sup_f \mathcal{L}\left(f, \xi\right) = \sup_{\|\boldsymbol{\beta}\|=1} \mathcal{L}\left(h_{\boldsymbol{\beta}}, \xi\right).$$

Then

$$\max_f \mathcal{L}_I\left(f,\xi\right) = \frac{\sigma_\varepsilon^2}{n} tr\left\{\mathbf{AB}^{-1}\right\} + \sup_{\|\boldsymbol{\beta}\|=1}\left(\mathbf{b}^T\left(h_{\boldsymbol{\beta}},\xi\right)\mathbf{H}^{-1}\mathbf{b}\left(h_{\boldsymbol{\beta}},\xi\right)+\eta^2\right)$$

$$= \frac{\sigma_\varepsilon^2}{n} tr\left\{\mathbf{AB}^{-1}\right\} + \eta^2 \sup_{\|\boldsymbol{\beta}\|=1}\left(\boldsymbol{\beta}^T\mathbf{G}^{\frac{1}{2}}\mathbf{H}^{-1}\mathbf{G}^{\frac{1}{2}}\boldsymbol{\beta}+\mathbf{1}\right)$$

$$= \frac{\sigma_\varepsilon^2}{n} tr\left\{\mathbf{AB}^{-1}\right\} + \eta^2\, ch_{\max}\left(\mathbf{GH}^{-1}+\mathbf{I}\right).$$

Since $\mathbf{G} = \mathbf{K} - \mathbf{H}$, the loss function might be written as

$$\max_f \mathcal{L}_I\left(f,\xi\right) = \frac{\sigma_\varepsilon^2}{n} tr\left\{\mathbf{AB}^{-1}\right\} + \eta^2\, ch_{\max}\left(\mathbf{KH}^{-1}\right).$$

Hence, the maximum of the loss function based on f might be shown as a multiple of $\frac{\sigma_\varepsilon^2}{n} + \eta^2$ by the function of ξ,

$$L_I\left(\xi\right) = \left(1-\nu\right)\underbrace{tr\left\{\mathbf{AB}^{-1}\right\}}_{Variance} + \underbrace{\nu ch_{\max}\left(\mathbf{KH}^{-1}\right)}_{Bias^2}, \tag{2.13}$$

where $\nu = \frac{\eta^2}{\frac{\sigma_\varepsilon^2}{n}+\eta^2} \in [0,1]$, and the value of ν is chosen by the experimenter. In fact, the experimenter determines the importance of variance versus bias by varying the value ν over the interval $[0,1]$.

As mentioned in the previous chapter, contamination terms in the class $\mathcal{F}$ (1.11) are bounded by a small value η. The value of η decreases at the rate $\frac{1}{\sqrt{n}}$. Therefore the maximized loss value, even if the loss function is not based on I-optimality, is meaningful asymtotically, since the bias decreases at the same rate as the variance. Otherwise when the sample size goes to infinity, we have either variance alone or bias

alone.

2.4.2   Minimizing the Maximum $IMSE$ over ξ

In this step, we work on minimizing the loss function $L_I(\xi)$ over designs ξ to achieve the optimal design ξ^* (also called minimax design, since we use minimax method to achieve it),

$$\xi^* = \min_{\xi} L_I(\xi) = \min_{\xi} \left[(1 - \nu)\, tr\left\{ \mathbf{AB}^{-1} \right\} + \nu\, ch_{\max}\left(\mathbf{KH}^{-1} \right) \right].$$

This part of the minimax method is complex, and we are not able to solve it in its general form. We should therefore look at an individual model. We consider only an approximate straight line regression model with a symmetric design in this thesis.

Define

$$\gamma_0 = \int_S x^2 dx, \;\; \gamma = \int_S x^2 m(x)dx, \tag{2.14}$$

$$\kappa_0 = \int_S m^2(x)dx, \;\; \kappa_1 = \int_S x^2 m^2(x)dx.$$

Then, based on the definition of $\mathbf{B}$, $\mathbf{A}$, $\mathbf{K}$, $\mathbf{H}$, and $\mathbf{G}$, we have:

$$\mathbf{B} = E_{\xi}\left[\mathbf{z}\left(x\right)\mathbf{z}^{T}\left(x\right)\right] \Rightarrow \mathbf{B}^{-1} = \text{diag}\left(1, \frac{1}{\gamma}\right), \tag{2.15}$$

$$\mathbf{A} = \int_{\mathbf{S}} \mathbf{z}\left(x\right)\mathbf{z}^{T}\left(x\right)dx = \text{diag}\left(1, \gamma_{0}\right),$$

$$\mathbf{K} = \int_{\mathbf{S}} \mathbf{z}\left(x\right)\mathbf{z}^{T}\left(x\right)m^{2}(x)dx = \text{diag}\left(\kappa_{0}, \kappa_{1}\right),$$

$$\mathbf{H} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B} = \text{diag}\left(1, \frac{\gamma^{2}}{\gamma_{0}}\right) \Rightarrow \mathbf{H}^{-1} = \text{diag}\left(1, \frac{\gamma_{0}}{\gamma^{2}}\right),$$

$$\mathbf{G} = \mathbf{K} - \mathbf{H} = \text{diag}\left((\kappa_{0}-1), (\kappa_{1} - \frac{\gamma^{2}}{\gamma_{0}})\right),$$

where γ_{0} is the second moment of the uniform distribution, and it is fixed value $(= \frac{1}{12})$. Hence

$$L_{I}\left(\xi\right) = (1-\nu)\left(1 + \frac{1}{12\gamma}\right) + \nu\left(\max\left(\kappa_{0}, \frac{\kappa_{1}}{12\gamma^{2}}\right)\right),$$

or

$$L_{I}\left(\xi\right) = (1-\nu)\left(1 + \frac{1}{12\gamma}\right) + \nu\left(\max\left(\int_{S} m^{2}(x)dx, \frac{1}{12\gamma^{2}}\int_{S} x^{2}m^{2}(x)dx\right)\right).$$

Based on definition (2.14), γ is a function of a density $(\gamma(m))$. Then, one of the eigenvalues $\int_{S} m^{2}(x)dx$ or $\frac{1}{12\gamma^{2}}\int_{S} x^{2}m^{2}(x)dx$ should be the largest characteristic root of the matrix $\mathbf{K}\mathbf{H}^{-1}$ at (2.13), for different densities $m(x)$. Generally, at this point, we select one of the eigenvalues as a maximum, and then investigate the correctness of our selection. Suppose the first eigenvalue (the simplest one) is the largest one, so

that

$$L_I\left(\xi\right) = (1-\nu)\left(1 + \frac{1}{12\gamma}\right) + \nu\left(\int_S m^2(x)dx\right).$$

To obtain the minimax design, we should find a density $m_0\left(x\right)$ which minimizes the above function of ξ. The value of ν is fixed by the experimenter. If we assume that γ has a fixed value, to minimize $L_I\left(\xi\right)$ based on ξ, we minimize only the bias by minimizing $\int_S m^2(x)dx$ subject to the fixed values $\int_S m(x)dx$ and $\int_S x^2 m(x)dx$. This is a constrained optimization problem that might be solved using the Lagrange Multiplier method. This method is a classical approach to solving a typical constrained optimization problem

$$\min_{\xi}\ f(\xi)\ \text{subject to}\ g_1(\xi) = 1,\ g_2(\xi) = \gamma,$$

where $f(\xi) = \int_S m^2(x)dx$ is the scalar-valued objective function, and $g_1(\xi) = \int_S m(x)dx$ and $g_2(\xi) = \int_S x^2 m(x)dx$ are the fixed constrained functions. This approach converts the constrained optimization problem into a scalar-valued function

$$L(\xi, \lambda, \mu) = f(\xi) + \lambda\ g_1(\xi) + \mu\ g_2(\xi)$$

which is known as the Lagrangian of a constrained optimization problem. Suppose the density $m_1(x)$ of the design ξ_1 is the minimizer of the function $f(\xi)$ subject to the constraints $g_1(\xi) = 1$, and $g_2(\xi) = \gamma$. Then the density $m_1(x)$, which satisfies the constraints, is the minimizer of the Lagrangian of the constrained optimization

problem. This means that for all design

$$\xi_t = (1-t)\xi_1 + t\xi,$$

the function $f(\xi_t)$ is minimized at $t = 0$ if and only if $L(\xi_t, \lambda, \mu)$ is minimum at $t = 0$. If we have an inequality constraint such as $h(\xi) \leqslant 0$, then in the Lagrangian the coefficient of this term is defined as a non-negative variable.

Our Lagrangian is

$$\int_S m^2(x)dx - \lambda \int_S m(x)dx - \mu \int_S x^2 m(x)dx = \int_S (m^2(x) - (\lambda + \mu x^2) m(x))dx,$$

subject to the constraints, determining the multipliers,

$$\int_S m(x)dx = 1, \quad \int_S x^2 m(x)dx = \gamma.$$

The integrand $m^2(x) - (\lambda + \mu x^2) m(x)$ is minimized pointwise over $m(x) (\geq 0)$ by $m_0(x) = (\lambda + \mu\, x^2)^+$. It turns out that $\mu > 0$ (although for $\mu < 0$, we may achieve a density, but in that case $\gamma < \gamma_0$, which is useless for us. This will be clarified later, in the next step). Therefore, m_0 should take the following form,

$$m_0(x) = a\left(x^2 - b\right)^+,$$

or

$$m_0(x) = a\left(x^2 - b\right) I\left(\max(0, b) \leqslant x^2 \leqslant \frac{1}{4}\right), \tag{2.16}$$
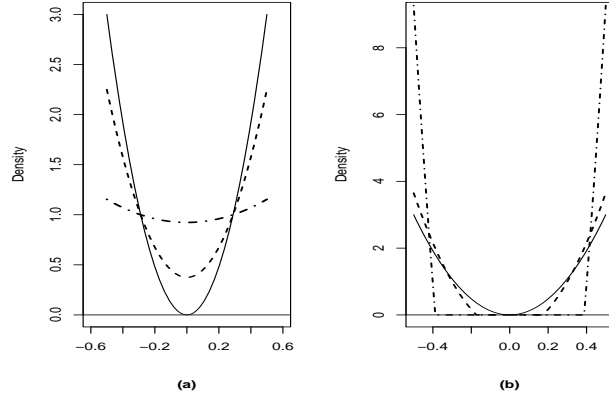
Figure 2.1: Forms of possible I-optimal densities for different values of b. Plot (a) presents three cases of robust densities $m_0(x)$ in which ν tends to one: $b = 0$, $\nu = 0.13$ (solid line), $b = -0.05$, $\nu = 0.26$ (dashed line), and $b = -1$, $\nu = 0.85$ (dotted-dashed Line); and plot (b) presents three cases of robust densities $m_0(x)$ in which ν tends to zero: $b = 0$, $\nu = 0.13$ (solid line), $b = 0.03$, $\nu = 0.083$ (dashed line), and $b = 0.15$, $\nu = 0.0096$ (dotted-dashed Line).

where $a > 0$, $b \leqslant \frac{1}{4}$. Hence, among densities in the class Ω with a fixed second moment $(E(x^2) = \gamma)$, the density m_0 is the minimizer of the maximized loss function $L_I(\xi)$.

Using the constraint $\int_S m_0(x)dx = 1$, we can obtain a and consequently the density $m_0(x)$ as a function of b. Then, by changing the values of b, we attain various densities. We split up these densities in two intervals $b \leqslant 0$ and $0 \leqslant b \leqslant \frac{1}{4}$.

At this point, we drop the assumption about the second moment γ, which was assumed to be fixed. Then, one way to calculate the optimal density for the loss function (based on I-optimality) among all possible densities with different values of second moment γ is to write the loss function $L_I(\nu, \gamma)$ as a function of ν and γ in this manner

$$L_I(\nu, \gamma) = A(\gamma) + \nu B(\gamma),$$

35

where

$$A(\gamma) = (1 + \frac{1}{12\gamma}),$$

$$B(\gamma) = (1 + \frac{1}{12\gamma} + \int_S m^2(x,\gamma)dx),$$

and we minimize it over γ. Hence

$$A'(\gamma) + \nu B'(\gamma) = 0,$$

or

$$\nu = -\frac{A'(\gamma)}{B'(\gamma)}.$$

Then, we vary γ over $\left[\frac{1}{12}, \frac{1}{4}\right]$ to compute the corresponding ν, which is called ν_1. Therefore, the loss $L_I(\nu_1, \gamma)$ is minimum. There is another way to minimize the loss function over γ. That is, we use the first condition of the Lagrange Multiplier method $\left(\int_S m(x)dx = 1\right)$ in order to calculate the value of a based on b. Thus, A and B might be written in terms of b, rather than γ. Then in order to minimize the maximized loss function $L_I(\xi)$, we take the first derivative of $L_I(\xi)$ based on b, and the value of ν becomes

$$\nu = \frac{1}{1 + \frac{12 \; \kappa_0'(b) \; \gamma^2(b)}{\gamma(b)}}.$$

Using this formula, for each particular $b \in (-\infty, \frac{1}{4})$, we calculate the value of ν for which b minimizes the maximized loss function $L_I(\xi)$, rather than finding the value b by minimizing the maximized loss function $L_I(\xi)$ for fixed ν; and we can find the

36

limit of the value of γ and ν on the domain $b \in (-\infty, \frac{1}{4})$, which are summarized as follows:

1. When $b \to \frac{1}{4}$, then $\gamma \to \frac{1}{4}$, $\nu \to 0$.

2. When $b = 0$, then $\gamma = \frac{3}{20}$, $\nu = \frac{25}{187}$.

3. When $b \to -\infty$, then $\gamma \to \frac{1}{12}$, $\nu \to 1$.

Similarly, we can find the corresponding limits of b, using the importance of variance versus bias in the experimenter's mind. The latter can be formulated as follows:

$$L_I(\xi) = (1 - \nu) \, Variance + \nu Bias^2.$$

For example, when $\nu \to 1$ (in which case only the bias is important for the experimenter), the optimum density is that of a design which tends to uniformity. The reason is that for the uniform design we have

$$\mathbf{b}(f, \xi) = \int_S \mathbf{z}(x) f(x) m(x) dx = \mathbf{0},$$

due to (1.8). Hence, based on Part (a) of Figure 2.1, $b \to -\infty$.

Furthermore, when $\nu \to 0$ (in which case only variance is important for the experimenter and the bias term disappears from the loss function), the optimum density is that of a design with all mass at $|x| = \frac{1}{2}$ in which $b \to \frac{1}{4}$. In fact, when $\nu \to 0$, the optimal design tends to the classical optimal design.

The plots in Figure 2.2 show the variation of ν based on the variable b in which the importance of bias, ν, is decreased by increasing the variable b, and also the

37
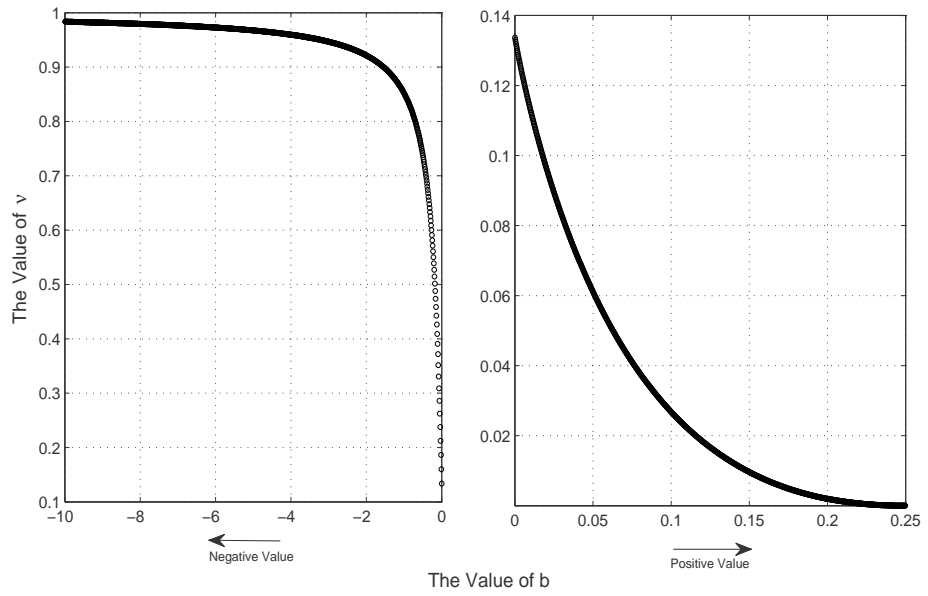
Figure 2.2: The possible values of ν for different values of b (for the loss function corresponding to I-optimality). Left plot presents the values of ν over the values of b less than zero, and right plot presents the values of ν over the values of b in the domain $(0, 0.25)$. When $b = 0$, then the value of ν is $\frac{25}{187}$.
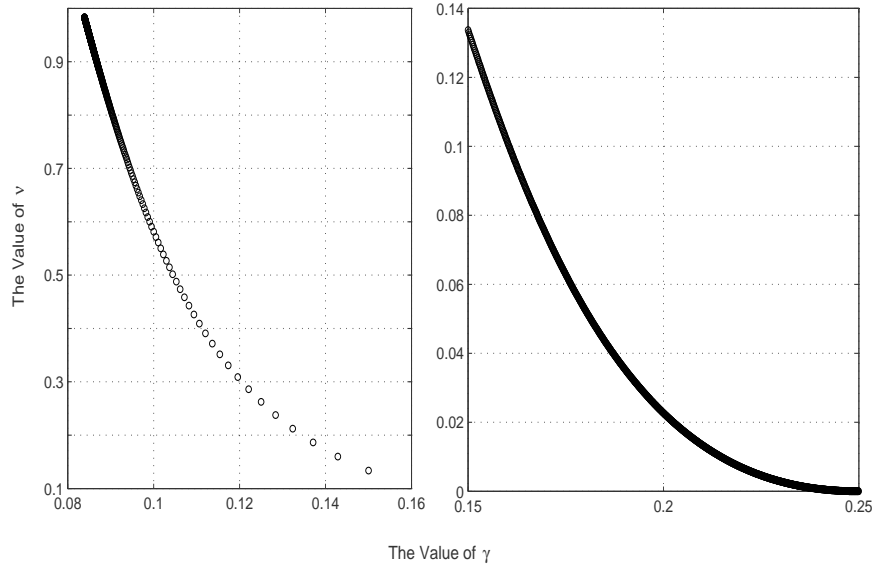
Figure 2.3: The possible values of ν for different values of γ (for the loss function corresponding to I-optimality). Left plot presents the values of ν for different values of γ when the value of b is less than zero, and right plot presents the values of ν for different values of γ when the value of b is in the domain $(0, 0.25)$.

importance of variance, $(1 - \nu)$, is increased by increasing the variable b.

The plots in Figure 2.3 show the variation of ν based on the variable γ in which the importance of bias, ν, is decreased by increasing the variable γ from $\frac{1}{12}$ to $\frac{1}{4}$, and the importance of variance, $(1 - \nu)$, is increased by increasing the variable γ from $\frac{1}{12}$ to $\frac{1}{4}$. Hence, the range of γ for all possible values of ν is more than or equal to γ_0 (the second moment of the uniform design); and investigation of the case that γ is less than γ_0 is unnecessary. It is thus more clear why we consider densities with the form $(\lambda + \mu \, x^2)^+$ where μ is positive.

2.4.3 Verifying the guess

In the previous step we found the density

$$m_0(x) = a \left(x^2 - b \right) I \left(\max(0, b) \leqslant x^2 \leqslant \frac{1}{4} \right),$$

where $a > 0$, $b \leqslant \frac{1}{4}$, which minimizes the function

$$L_I \left(\xi \right) = (1 - \nu) \left(1 + \frac{1}{12\gamma} \right) + \nu \left(\max \left(\kappa_0, \ \frac{\kappa_1}{12\gamma^2} \right) \right),$$

assuming κ_0 as a largest eigenvalue. But since (as we discussed in the previous section) κ_0, κ_1, and γ are functions of b, the truth of the inequality

$$\kappa_0 \geqslant \frac{\kappa_1}{12\gamma^2}, \tag{2.17}$$

should be investigated for all values of $b \in (-\infty, 0.25)$. When $b \in (-\infty, 0)$, we analytically found that $\kappa_0 - \frac{\kappa_1}{12\gamma^2}$, as a function of b, is always positive,

$$\kappa_0 - \frac{\kappa_1}{12\gamma^2} = \frac{128}{7} b(4b - 3) + \frac{192}{35} \geqslant 0 \text{ for } \forall b \in (-\infty, 0).$$

But when $b \in (0, 0.25)$, it is difficult to prove the truth of (2.17). Thus, we have to apply numerical methods in R (statistics software). The plots in Figure 2.4 display the difference between κ_0 and $\frac{\kappa_1}{12 \ \gamma^2}$.

These plots show that κ_0 is the largest eigenvalue for all values over the domain
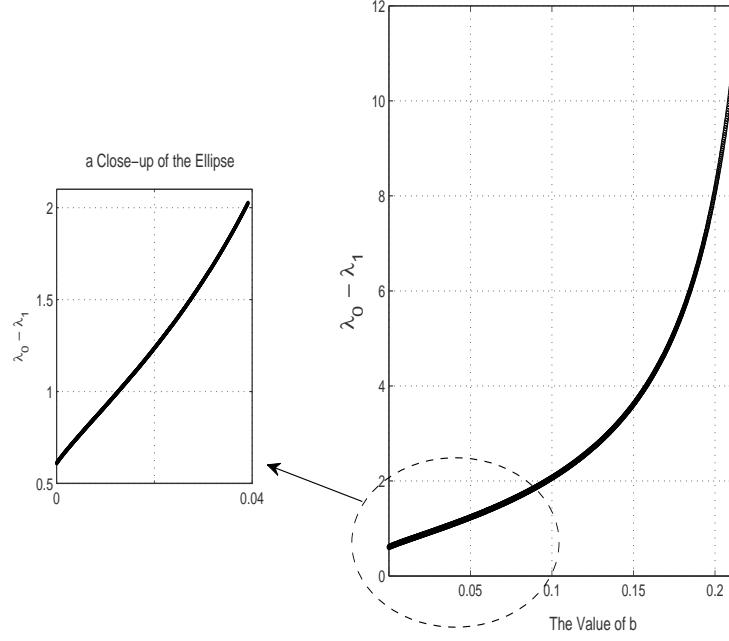
Figure 2.4: The difference between two eigenvalues, $\left(\kappa_0 - \frac{\kappa_1}{12\,\gamma^2}\right)$, in the loss function corresponding to I-optimality over the values of b in the domain $(0, 0.25)$.

of b. Hence, we conclude that the density

$$m_0(x) = a\left(x^2 - b\right) I\left(\max\left(0, b\right) \leqslant x^2 \leqslant \frac{1}{4}\right),$$

where $a > 0$, $b \leqslant \frac{1}{4}$, is the minimizer of the function $L_I\left(\xi\right)$. In other words, the optimal density $m_0(x)$ is the density of the minimax design ξ^*. But as experimenters, we are looking for an implementable design to conduct an experiment. An implementable design should be extracted (in some way) from the optimal density in order to have the best observations from the experiment.

41

2.4.4 Implementing an Optimal Design

Several authors have presented various implementation methods. Wiens (1992) suggested a randomized approach in which we first generate values $u_1, \ldots, u_n$ from the uniform random variable on $(0, 1)$, and then design points x_i are randomly chosen from the optimal density $m_0(x)$, given by

$$x_i = F_{\xi_0}^{-1}(u_i).$$

A more systematic approach was introduced by Wiens and Zhou (1996). They choose design points at quantiles of the optimal density

$$x_i = F_{\xi_0}^{-1}\left(\frac{i - 0.5}{n}\right).$$

This is called Quantile Matching Design (QMD), and in fact the design points are picked by uniformly partitioning the y-axis.

Instead of uniformly partitioning the y-axis, we may uniformly partition the x-axis, a process which is known as Equally Spaced Design (ESD). Suppose a design space $S = \left[-\frac{1}{2}, \frac{1}{2}\right] \subset \mathbb{R}$. In order to select m design points, we divide the design space to the m subspace, select the midpoint of each interval, defined as

$$x_i = -\frac{1}{2} + \frac{(i - 1)}{m} + \frac{1}{2m}, \text{ for } i = 1, \ldots, m,$$

and compute the number of repetition of each point using the integral of the optimal

density $m_0(x)$ over its corresponding interval

$$p_i = \int_{-\frac{1}{2}+\frac{(i-1)}{m}}^{-\frac{1}{2}+\frac{i}{m}} m_0(x) \ dx, \text{for } i = 1, \ldots, m.$$

The number of repetition of each point n_i is np_i – the product of the sample size n and the corresponding probability of the interval related to the design point x_i. In fact, we assign all mass at the midpoint of each interval. For some subintervals, the value of the product np_i is not an integer and we need to round the frequencies of the corresponding design point x_i. We may solve this problem using the efficient rounding method suggested by Pukelsheim and Rieder (1992). In addition, the selection of subintervals and design points are somewhat subjective.

Heo, Schmuland and Wiens (2001) suggested a matching moment approach, in which we may choose design points by matching up the moments of the design with its empirical moments. The matching is done by numerical minimization of

$$\sum_k \left\{ \frac{1}{m} \sum_{i=1}^{m} x_i^{2k} - E_{\xi_0}\left(X^{2k}\right) \right\}^2.$$

Xu and Yuen (2011) proposed an Optimal Approximation Design (OAD) under a criterion based on the Kolmogorov-Smirnov distance which was proposed by Fang and Wang (1994) in order to define the concept of discrepancy to measure the uniformity of a set of points on $[0,1]^s$. Xu and Yuen reasoned as follows: Suppose we aim to have a sample with size n, from m distinct design point $x_1 < \ldots < x_m$ with repetition

$n_1, \ldots, n_m$, respectively where $\sum_{i=1}^{m} n_i = n$. Define $l_0 = 0$ and

$$l_i = n_1 + \ldots + n_i,$$

for $i = 1, \ldots, m$. They assumed $x_0 = -\infty$, and $x_{k+1} = +\infty$, and defined the following e.d.f of the design as an approximating discrete distribution

$$\widehat{F}(x) = \frac{l_i}{n}$$

if $x_i \le x < x_{i+1}$, $i = 0, \ldots, m$. They minimize the Kolmogorov-Smirnov – distance between the e.d.f $\left(\widehat{F}\right)$ and the optimal distribution function F_{ξ_0} –

$$\mathcal{K}_{k,n} = \mathcal{K}_{k,n}(F_{\xi_0}, \widehat{F}) = \max_i \left\{ \max \left[F_{\xi_0} - \frac{l_i - 1}{n}, \frac{l_i}{n} - F_{\xi_0}(x_i) \right] \right\},$$

and they chose m distinct design points $x_1 < \ldots < x_m$, with $n_1, \ldots, n_m$ observations at each point by minimizing $\mathcal{K}_{k,n}$. Figures 2.5 and 2.6 present design points which are selected using different methods for the I-optimal density corresponding to $\nu = 0.25$.

At this point, the two main advantages of a robust design appear. The first one is revealed by the fitted model in which due to uncertainty of the suggested model, we consider a contamination term that is bounded by some conditions. Hence, the optimal design is not sensitive to what we suggest as a fitted model to the data. Secondly, unlike the classical method, the robust approach allows us to explore all the design points over the design space, depending on how much the experimenter is

44
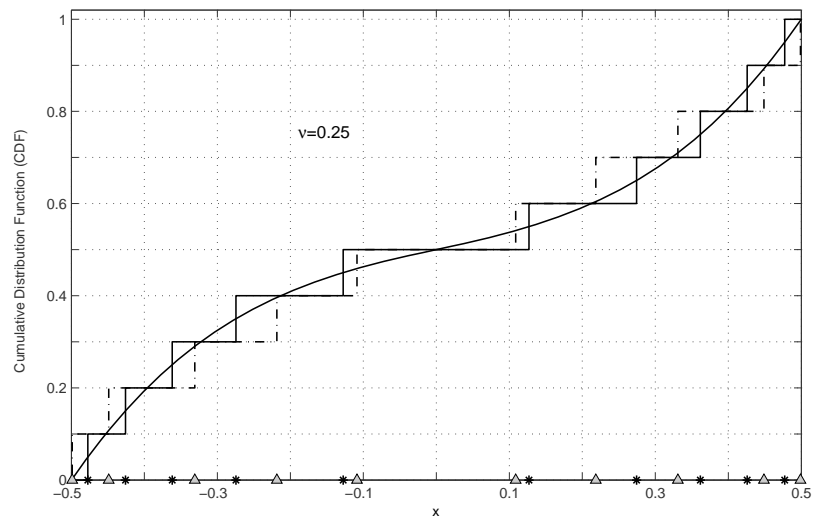
Figure 2.5: The solid line curve depicts the distribution function of the I-optimal density corresponding to $\nu = 0.25$. The solid line and dotted-dashed line stairs show the distribution function of the implemented I-optimal design using the matching moment and quantile matching methods respectively. Triangles and stars show the design points computed by these methods.
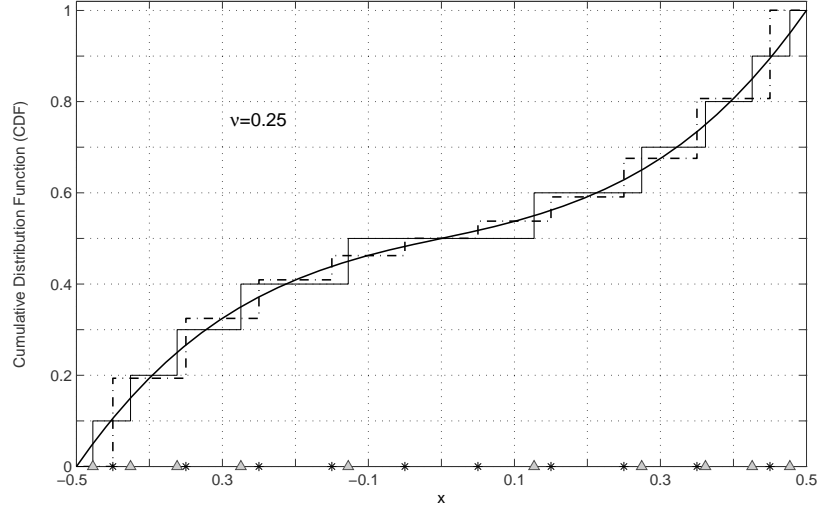
Figure 2.6: The solid line curve depicts the distribution function of the I-optimal density corresponding to $\nu = 0.25$. The solid line and dotted-dashed line stairs show the distribution function of the implemented I-optimal design using the OAD and ESD methods respectively. Triangles and stars show the design points computed by OAD and ESD.

willing to pay in terms of a loss of efficiency in order to have robustness.

2.5   Minimax Design Using A-optimality, for $\nu < 0.692$

The loss function corresponding to A-optimality is the trace of the Mean Squared Error matrix of $\widehat{\boldsymbol{\theta}}$,

$$\mathcal{L}_A\left(f,\xi\right) = tr\left(\mathbf{M}\left(f,\xi\right)\right),$$

where $\mathbf{M}\left(f,\xi\right) = \mathbf{MSE}$. Using the result (1.9), this loss function might be written as a function of ξ and f,

$$\mathcal{L}_A\left(f,\xi\right) = \frac{\sigma_\varepsilon^2}{n} tr\mathbf{B}^{-1} + tr\left(\mathbf{B}^{-1}\mathbf{b}\left(f,\xi\right)\mathbf{b}^T\left(f,\xi\right)\mathbf{B}^{-1}\right).$$

46

The loss function based on A-optimality has both conditions (L1) and (L2). The trace is a monotonic function of $\mathbf{M}(f,\xi)$, and the limit of $tr(\mathbf{M}(f_n,\xi))$ tends to infinity, if $ch_1(\mathbf{M}(f_n,\xi)) \to \infty$. Hence we have a finite loss function only for absolutely continuous design ξ. After presenting this loss function as a function of ξ and f , the following strategies are applied to achieve an optimal design.

2.5.1 Maximizing $\mathcal{L}_A(f,\xi)$ over f

In this step, assuming a fixed design ξ, we would like to maximize $\mathcal{L}_A(f,\xi)$ based on f over the class $\mathcal{F}$. Hence

$$\max_f \mathcal{L}_A(f,\xi) = \frac{\sigma_\varepsilon^2}{n} tr\mathbf{B}^{-1} + \max_f tr\left(\mathbf{B}^{-1}\mathbf{b}(f,\xi)\mathbf{b}^T(f,\xi)\mathbf{B}^{-1}\right).$$

Using Theorem 1 and its result $\left(\mathbf{b}(h_{\boldsymbol{\beta}},\xi) = \eta\mathbf{G}^{\frac{1}{2}}\boldsymbol{\beta}\right)$, the maximization problem over the infinite dimensional class $\mathcal{F}$ is reduced to a finite dimensional parametric problem over the class $\mathcal{H} = \left\{h_{\boldsymbol{\beta}}(x) = \mathbf{r}^T(x)\boldsymbol{\beta}; \|\boldsymbol{\beta}\| = 1\right\}$.

Moreover, based on this Theorem, $\mathcal{H}$ is a subset of $\mathcal{F}$, with the properties

$$\int_S h_{\boldsymbol{\beta}}^2(x)\,dx = \eta^2, \text{and} \sup_f \mathcal{L}(f,\xi) = \sup_{\|\boldsymbol{\beta}\|=1} \mathcal{L}(h_{\boldsymbol{\beta}},\xi).$$

47

Then

$$\max_f \mathcal{L}_A(f,\xi) = \frac{\sigma_\varepsilon^2}{n} tr\mathbf{B}^{-1} + \sup_{\|\boldsymbol{\beta}\|=1} tr\left(\mathbf{B}^{-1}\mathbf{b}(h_{\boldsymbol{\beta}},\xi)\mathbf{b}^T(h_{\boldsymbol{\beta}},\xi)\mathbf{B}^{-1}\right)$$

$$= \frac{\sigma_\varepsilon^2}{n} tr\mathbf{B}^{-1} + \eta^2 \sup_{\|\boldsymbol{\beta}\|=1}\left(\boldsymbol{\beta}^T\mathbf{G}^{\frac{1}{2}}\mathbf{B}^{-1}\,\mathbf{B}^{-1}\mathbf{G}^{\frac{1}{2}}\boldsymbol{\beta}\right)$$

$$= \frac{\sigma_\varepsilon^2}{n} tr\mathbf{B}^{-1} + \eta^2\, ch_{\max}\left(\mathbf{B}^{-1}\,\mathbf{G}\mathbf{B}^{-1}\right).$$

Hence, the maximum loss function (corresponding to A-optimality) over the class $\mathcal{F}$ is a multiple of $\frac{\sigma_\varepsilon^2}{n} + \eta^2$ by the function of ξ

$$L_A(\xi) = (1-\nu)\underbrace{tr\mathbf{B}^{-1}}_{Variance} + \nu\underbrace{ch_{\max}\left(\mathbf{B}^{-1}\,\mathbf{G}\mathbf{B}^{-1}\right)}_{Bias^2}, \qquad (2.18)$$

where $\nu = \frac{\eta^2}{\frac{\sigma_\varepsilon^2}{n}+\eta^2} \in [0,1]$. The variation for ν in its domain is under the control of the experimenter. In practice, the experimenter determines the importance of variance versus bias by taking different values for ν from zero to one.

2.5.2   Minimizing the Maximum of $\mathcal{L}_A(f,\xi)$ over ξ

In the previous step, the function $L_A(\xi)$ was obtained after maximizing the loss function based on A-optimality over all possible contamination terms f in the class $\mathcal{F}$. Now, we must minimize the function $L_A(\xi)$ over designs ξ to achieve the optimal design (or minimax design)

$$\xi^* = \arg\min_\xi L_A(\xi) = \arg\min_\xi \left[(1-\nu)tr\mathbf{B}^{-1} + \nu\, ch_{\max}\left(\mathbf{B}^{-1}\,\mathbf{G}\mathbf{B}^{-1}\right)\right], \qquad (2.19)$$

where $\nu \in [0,1]$. Then based on (2.15) the maximized loss function corresponding to A-optimality is

$$L_A(\xi) = (1-\nu)\left(1+\frac{1}{\gamma}\right) + \nu\left(\max\left(\kappa_0 - 1, \frac{\kappa_1}{\gamma^2} - 12\right)\right),$$

where κ_0, κ_1 and γ are defined at (2.14). Hence

$$L_A(\xi) = (1-\nu)\left(1+\frac{1}{\gamma}\right) + \nu\left(\max\left(\int_S m^2(x)dx - 1, \frac{\int_S x^2 m^2(x)dx}{\gamma^2} - 12\right)\right)$$

is a function of a density $m(x)$. In order to minimize this function over all possible densities, we first select one of the eigenvalues of the matrix $\mathbf{B}^{-1}\,\mathbf{G}\mathbf{B}^{-1}$ at (2.19) as a maximum, and solve the minimization problem. Then we investigate the correctness of our choice using the method applied in the previous section for I-optimality.

Assuming the first eigenvalue $\kappa_0 - 1$ as a maximum characteristic root, we proceed using the same Lagrange Multiplier method as was applied for I-optimality. The result is similar to I-optimality, and the density takes the form

$$m_0(x) = a\left(x^2 - b\right) I\left(\max(0,b) \leqslant x^2 \leqslant \frac{1}{4}\right), \tag{2.20}$$

where $a > 0$, $b \leqslant \frac{1}{4}$. But the conjecture that $\kappa_0 - 1$ is the largest root, when evaluated at the minimizing design, fails, as shown in Figure 2.7.

Therefore, we take $\frac{\kappa_1}{\gamma^2} - 12$ as a maximum characteristic root. Hence

$$L_A(\xi) = (1-\nu)\left(1+\frac{1}{\gamma}\right) + \nu\left(\frac{\int_S x^2 m^2(x)dx}{\gamma^2} - 12\right).$$
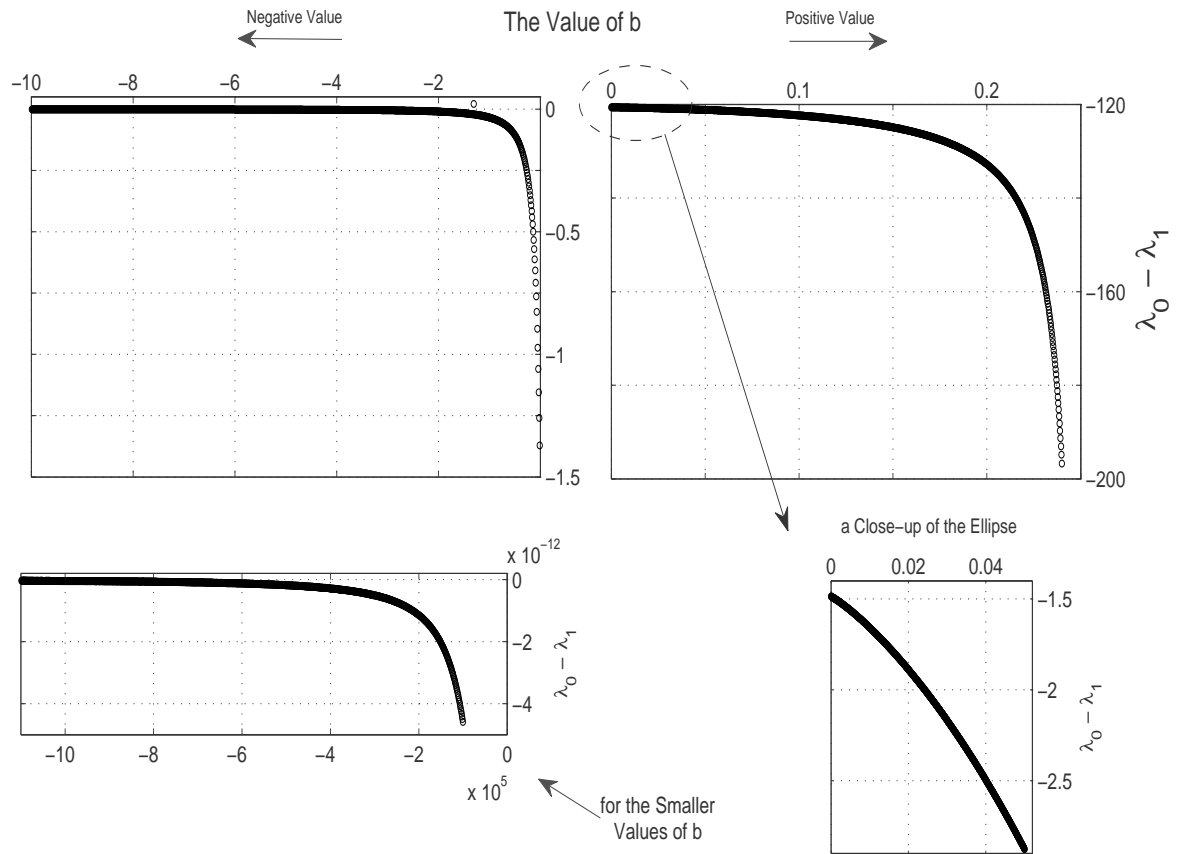
49

Figure 2.7: The difference between two eigenvalues, $\left((\kappa_0 - 1) - \left(\frac{\kappa_1}{\gamma^2} - 12\right)\right)$, in the loss function corresponding to A-optimality for different values of b in the domain $(-\infty, 0.25)$.

Then again we apply the same Lagrange Multiplier method as was used in determining I-optimality in order to minimize this maximized loss function $L_A(\xi)$, as a function of a density $(m(x))$, over all possible absolutely continuous designs with density $m(x)$; and we minimize

$$\int_S x^2 m^2(x)dx - \lambda \int_S m(x)dx - \mu \int_S x^2 m(x)dx.$$

The integrand $x^2 m^2(x) - (\lambda + \mu x^2) m(x)$ is minimized pointwise over $m(x)(\geq 0)$ by $m_0(x) = (\lambda + \frac{\mu}{x^2})^+$. It turns out that $\mu > 0$. Thus m_0 is the form

$$m_0(x) = a \left(1 - \frac{b}{x^2}\right)^+$$

or

$$m_0(x) = a \left(1 - \frac{b}{x^2}\right) I \left(b \leqslant x^2 \leqslant \frac{1}{4}\right), \tag{2.21}$$

where $a > 0$, $0 < b < \frac{1}{4}$.

Next we apply the same trick as we did for I-optimality. We calculate a as a function of b, using the first condition of the Lagrange Multiplier $\left(\int_S m(x)dx = 1\right)$. Substituting the value of a based on b, we write γ and κ_1 as functions of b. Then taking the first derivative of $L_A(\xi)$ based on b, the value of ν as a function of b is

$$\nu = \frac{1}{1 + \left(\frac{\kappa_1'(b)}{\gamma(b)} - \frac{2 \; \kappa_1(b)}{\gamma(b)}\right)}.$$

The above formula determines the value of ν for any variable of b in its domain.

51
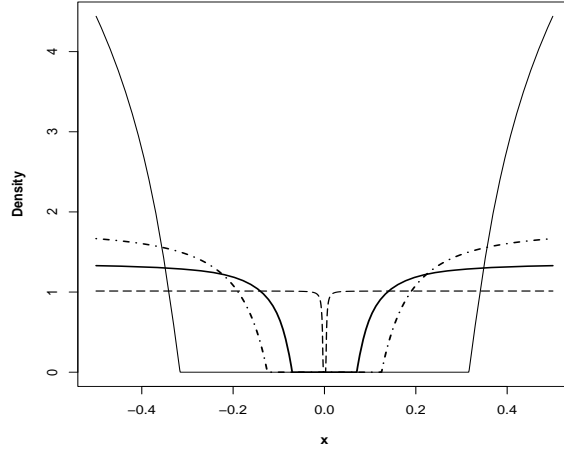
Figure 2.8: The plot presents four different densities $m_0(x)$ that minimize the second eigenvalue using A-optimality. From the center, $b = 0.00001$, $\nu = 0.9998$ (longdash line), $b = 0.005$, $\nu = 0.887$ (bold line), $b = 0.0156$, $\nu = 0.692$ (dotted-dashed line), and $b = 0.1$, $\nu = 0.113$ (solid line).

For this value of ν, the loss function is minimized by its corresponding value of b. In

Figure 2.9, the plots show the limits of the value of γ and ν on the domain $b \in (0, \frac{1}{4})$:

1. when $b \to 0$, then $\gamma \to \frac{1}{12}, \nu \to 1$.

2. when $b \to \frac{1}{4}$, then $\gamma \to \frac{1}{4}, \nu \to 0$.

Therefore, when the experimenter is not too concerned about errors arising from

bias (and so takes a small value of ν, $(\nu \to 0)$) the optimum design is close to the

classical optimal design – a design with all mass at $\|x\| = \frac{1}{2}$ – in which based on

Figure 2.8, $b \to \frac{1}{4}$; and when the experimenter is too concerned about errors arising

from bias (and so takes a large value of ν, $(\nu \to 1)$) due to her uncertainty of the

fitted model, the optimum density is that of a design which tends to uniformity in

which based on Figure 2.8, $b \to 0$.
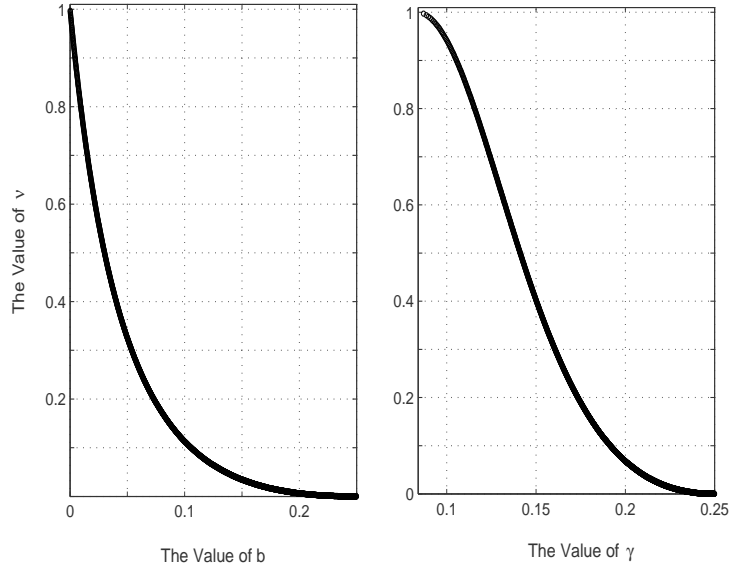
Figure 2.9: The variation of ν based on the variables b and γ (for the loss function corresponding to A-optimality) on the domain $b \in (0, \frac{1}{4})$.

2.5.3 Verifying the Guess

Based on the previous step, the result of the minimax problem for the loss function corresponding to A-optimality

$$L_A\left(\xi\right) = (1 - \nu)\left(1 + \frac{1}{\gamma}\right) + \nu\left(\frac{\kappa_1}{\gamma^2} - 12\right),$$

based on choosing $\frac{\kappa_1}{\gamma^2} - 12$ as a largest eigenvalue, is the density

$$m_0(x) = a\left(1 - \frac{b}{x^2}\right) I\left(b \leqslant x^2 \leqslant \frac{1}{4}\right),$$
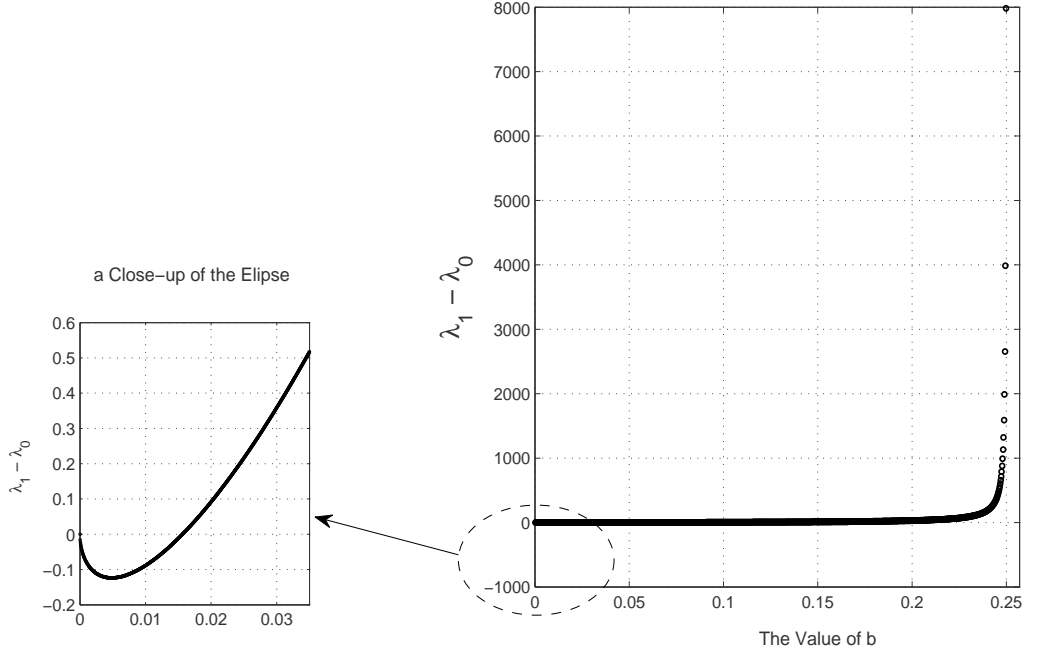
Figure 2.10: The difference between two eigenvalues, $\left(\frac{\kappa_1}{\gamma^2} - 12\right) - (\kappa_0 - 1)$, in the loss function corresponding to A-optimality for different values of b in the domain $(0, 0.25)$.

where $a > 0$, $0 < b < \frac{1}{4}$. Since κ_0, κ_1, and γ are functions of b, in order to accept $m_0(x)$ as an optimal density, we must verify the truth of our guess

$$\frac{\kappa_1}{\gamma^2} - 12 \geqslant \kappa_0 - 1,$$

for all values of the domain of b $(b \in (0, 0.25))$. As Figure 2.10 shows the result of the numerical method applied in R, the difference between $\frac{\kappa_1}{\gamma^2} - 12$ and $\kappa_0 - 1$ is positive for all values of b more than 0.0156 in the domain $(0, 0.25)$.

Since $\frac{\kappa_1}{\gamma^2} - 12$ is the largest eigenvalue for all values b more than 0.0156 in the

54

domain $(0, 0.25)$, the optimal density is

$$m_0(x) = a \left(1 - \frac{b}{x^2} \right) I \left(b \leqslant x^2 \leqslant \frac{1}{4} \right),$$

where $a > 0$, $0 < b < \frac{1}{4}$, for the value of b more than 0.0156 corresponding with ν less than $\nu_A = 0.692$, where

$$\nu = \frac{\eta^2}{\frac{\sigma_\varepsilon^2}{n} + \eta^2} \in [0, 1].$$

In the final stage, an implementable design could be extracted from the optimal densities by various methods as reviewed earlier in the section dealing with I-optimality, when ν is not too large.

For those values of b which are smaller than 0.0156, Figure 2.8 implies that the robust densities are more similar to the classical design than the uniform design.

As this section reveals, then, the technique used in minimizing the maximized loss function $L_A (\xi)$ over all possible absolutely continuous design is not applicable for $\nu \geqslant \nu_A$. In the next chapter, we apply an alternative to this technique in order to find an optimal density for the minimax design problem based on A-optimality, for all values of ν.

2.6 Minimax Design Using E-optimality, for $\nu < 0.997$

The loss function corresponding to E-optimality is the largest characteristic root of the mean squared matrix of $\widehat{\boldsymbol{\theta}}$ ($ch_{\max}\mathbf{MSE}$). Using the result (1.9), we have

$$\mathcal{L}_E\left(f,\xi\right) = ch_{\max}\left\{\frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{b}\left(f,\xi\right)\mathbf{b}^T\left(f,\xi\right)\mathbf{B}^{-1}\right\}. \qquad (2.22)$$

The loss function based on E-optimality has both conditions (L1) and (L2), since the loss function is a monotonic function of $\mathbf{M}\left(f,\xi\right)$ and given that the limit of $\mathcal{L}_E\left(f_n,\xi\right)$ tends to infinity, if $ch_1\left(\mathbf{M}\left(f_n,\xi\right)\right) \to \infty$. Hence we have a finite loss function only for absolutely continuous design ξ. The loss function based on E-optimality is shown as a function of ξ and f in (2.22). Thus we apply the following strategies, respectively, in order to achieve an optimal design.

2.6.1 Maximizing $\mathcal{L}_E\left(f,\xi\right)$ over f

In this step, assuming a fixed design ξ , we aim to maximize $\mathcal{L}_E\left(f,\xi\right)$ based on f over the class $\mathcal{F}$:

$$\max_f \mathcal{L}_E\left(f,\xi\right) = \max_f\ ch_{\max}\left\{\frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{b}\left(f,\xi\right)\mathbf{b}^T\left(f,\xi\right)\mathbf{B}^{-1}\right\}. \qquad (2.23)$$

From Theorem 1 applied to (2.23),

$$\max_f \mathcal{L}_E\left(f, \xi\right) = \sup_{\|\beta\|=1} ch_{\max}\left\{\frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{b}(h_\beta, \xi)\mathbf{b}^T(h_\beta, \xi)\mathbf{B}^{-1}\right\}$$

$$= \sup_{\|\beta\|=1}\sup_{\|\alpha\|=1} \boldsymbol{\alpha}^T\left[\frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1} + \eta^2\mathbf{B}^{-1}\mathbf{G}^{\frac{1}{2}}\boldsymbol{\beta}\,\boldsymbol{\beta}^T\mathbf{G}^{\frac{1}{2}}\mathbf{B}^{-1}\right]\boldsymbol{\alpha}$$

$$= \sup_{\|\alpha\|=1}\left[\frac{\sigma_\varepsilon^2}{n}\boldsymbol{\alpha}^T\mathbf{B}^{-1}\boldsymbol{\alpha} + \eta^2\,ch_{\max}\left(\mathbf{B}^{-1}\mathbf{G}^{\frac{1}{2}}\boldsymbol{\alpha}\,\boldsymbol{\alpha}^{\mathbf{T}}\mathbf{G}^{\frac{1}{2}}\mathbf{B}^{-1}\right)\right]$$

$$= ch_{\max}\left[\frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1} + \eta^2\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}\right].$$

Hence the maximum of loss function E-optimality over the class of $\mathcal{F}$ is a multiple of $\frac{\sigma_\varepsilon^2}{n} + \eta^2$ by the function

$$L_E\left(\xi\right) = ch_{\max}\left[(1-\nu)\underbrace{\mathbf{B}^{-1}}_{Variance} + \nu\underbrace{\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}}_{Bias^2}\right]. \tag{2.24}$$

2.6.2 Minimizing the Maximum $\mathcal{L}_E\left(f, \xi\right)$ over $\boldsymbol{\xi}$

At this point, all of the processes used to minimize the maximum loss function (2.24) are the same as those used to minimize the maximum loss function for A-optimality (2.18). Then based on (2.14) and (2.15), the maximum of loss function over the class $\mathcal{F}$ might be written as a function of the density $m(x)$ as

$$L_E\left(\xi\right) = \max\left[(1-\nu) + \nu\left(\kappa_0 - 1\right),\ \frac{1-\nu}{\gamma} + \nu\left(\frac{\kappa_1}{\gamma^2} - \frac{1}{\gamma_0}\right)\right], \tag{2.25}$$

or

$$L_E\left(\xi\right) = \max\left[\left(1-\nu\right) + \nu\left(\int_S m^2(x)dx - 1\right), \frac{1-\nu}{\gamma} + \nu\left(\frac{\int_S x^2 m^2(x)dx}{\gamma^2} - 12\right)\right].$$

In order to minimize the function $L_E\left(\xi\right)$ over all possible absolutely continu-ous design measures, we proceed by using the same Lagrange Multiplier method as was applied for I- and A-optimality, and we obtain a similar result as we did for A-optimality. That is, the conjecture that $(1-\nu) + \nu\left(\kappa_0 - 1\right)$ is the largest charac-teristic root of the matrix $\left[(1-\nu)\mathbf{B}^{-1} + \nu\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}\right]$ at (2.24), when evaluated at the minimizing design (2.20), fails. Moreover, when $\frac{1-\nu}{\gamma} + \nu\left(\frac{\kappa_1}{\gamma^2} - 12\right)$ is taken as a maximum characteristic root, the minimizer of the function $L_E\left(\xi\right)$ takes the form (2.21).

The results for the limit of the value of γ and ν on the domain $b \in (0, \frac{1}{4})$ are similar to those found for A-optimality. In other words, when $\nu \to 1$ (in which case the variance term vanishes), the optimum design is a design which tends to uniformity, and hence $b \to 0$. In addition, when $\nu \to 0$ (in which case the variance is the only dominant term in the model), the optimum design tends to the classical optimal design. The changes in the values of γ and ν on the domain $b \in (0, \frac{1}{4})$ are exactly the same as in the plots of Figure 2.9 related to the A-optimality.

2.6.3   Verifying the Guess

In a manner similar to that discussed in the previous section, the density (2.21) is the E-optimal density, assuming $\frac{1-\nu}{\gamma} + \nu\left(\frac{\kappa_1}{\gamma^2} - \frac{1}{\gamma_0}\right)$ as the largest eigenvalue. In order to
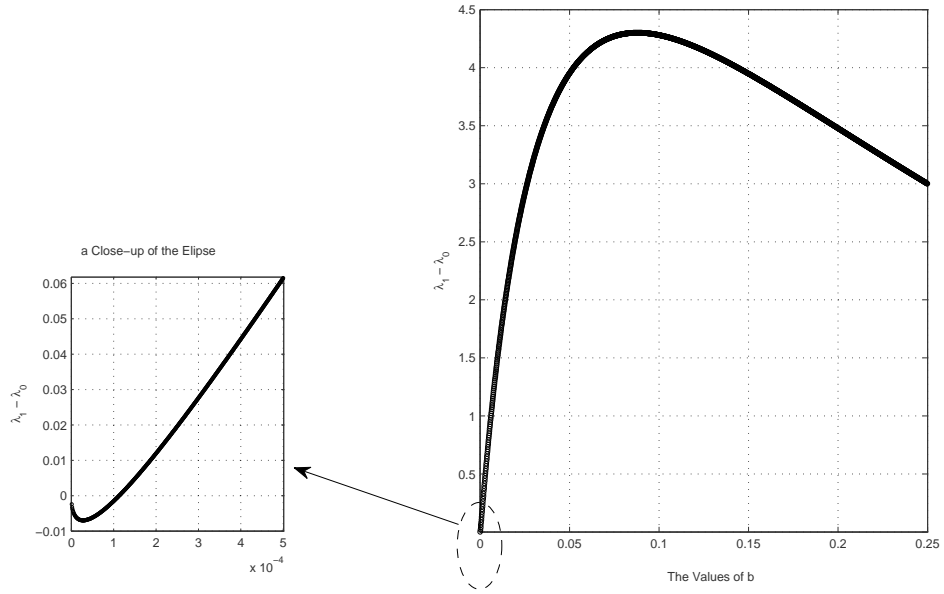
Figure 2.11: The difference between two eigenvalues, $\left(\frac{\kappa_1}{\gamma^2} - 12\right) - (\kappa_0 - 1)$, in the loss function corresponding to E-optimality for different values of b in the domain $(0, 0.25)$.

examine the correctness of the inequality

$$(1 - \nu) + \nu(\kappa_0 - 1) \geqslant \frac{1 - \nu}{\gamma} + \nu\left(\frac{\kappa_1}{\gamma^2} - \frac{1}{\gamma_0}\right),$$

we apply the same numerical method in R as we did for A-optimality. The result shows the correctness of this inequality for some values of b in its domain $(0, 0.25)$ which correspond to a value of ν of less than $\nu_E = 0.997$.

Then in the final stage, we compute an implementable design from the optimal density using various methods as they are reviewed in the I-optimality section.

To summarize this chapter: the density $m_0(x)$ at (2.16) is an I-optimal density for all values of ν, and the density $m_0(x)$ at (2.21) is an optimal density using A-
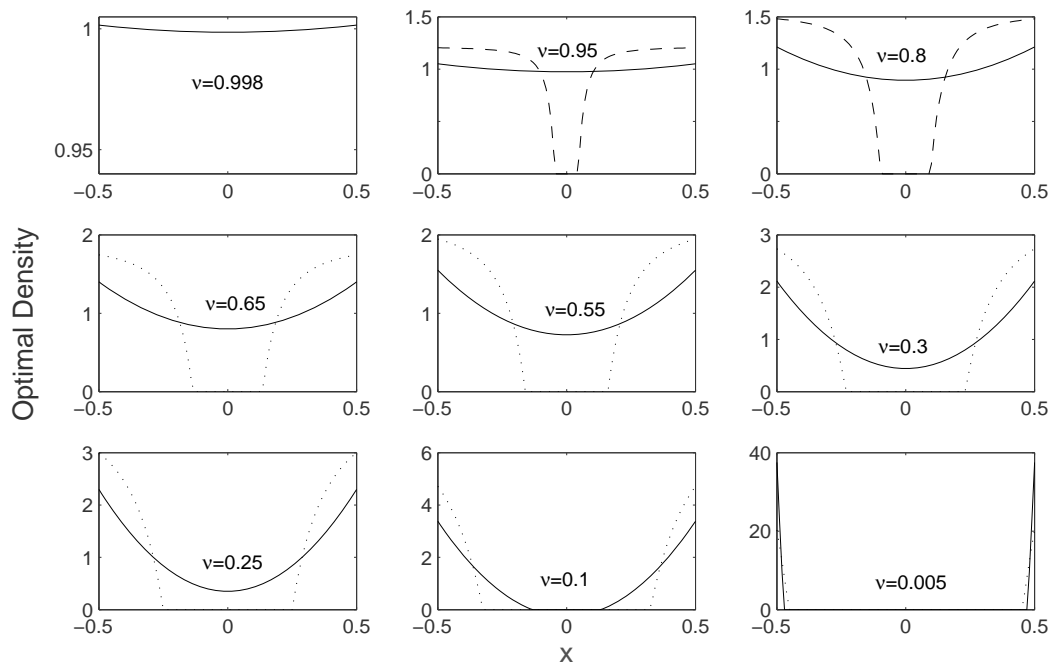
Figure 2.12: These plots show optimal densities for different values of υ, using three optimality criteria. The solid line and the dashed line curves present the I-optimal and E-optimal densities respectively. The dotted line curve shows the optimal density corresponding to both the A- and E-optimality criteria.

optimality or E-optimality when ν is not too large. In other words, a part of the method used to minimize the maximized loss function $L_A(\xi)$ or $L_E(\xi)$ over all possible absolutely continuous design is not applicable to finding a minimax design for an approximate straight regression line, when we select a sufficiently large value for ν. In the next chapter, we apply a different technique to address this gap within the existing method. The new technique is applicable for experiments with all ν values when we wish to achieve an optimal design using the loss function corresponding to A-optimality or E-optimality. Figure 2.12 clearly illustrates the gap within the method applied in this chapter for A- and E- optimality.

As Figure 2.12 shows, the A- and E-optimal densities are identical for $\nu < \nu_A$. The maximized loss functions for A- and E-optimality criteria are multiples of $\frac{\sigma_\varepsilon^2}{n} + \eta^2$ by

$$L_A(\xi) = (1 - \nu)\left(1 + \frac{1}{\gamma}\right) + \nu\left(\max\left(\kappa_0 - 1, \frac{\kappa_1}{\gamma^2} - 12\right)\right),$$

and

$$L_E(\xi) = \max\left[(1 - \nu) + \nu(\kappa_0 - 1), \frac{1 - \nu}{\gamma} + \nu\left(\frac{\kappa_1}{\gamma^2} - 12\right)\right]. \qquad (2.26)$$

Let's $L_{A,i}(\xi)$ and $L_{E,i}(\xi)$, $i = 0, 1$ be the loss function using the first and second eigenvalues respectively. As noted in the second section of this chapter related to A-optimality, for all values of ν less than ν_A, the conjecture that the second eigenvalue $\left(\frac{\kappa_1}{\gamma^2} - 12\right)$, evaluated at A-optimal design is the largest characteristic root, succeeded. On the other hand, since the second moment of the optimal distribution γ varies over the interval $\left[\frac{1}{12}, \frac{1}{4}\right]$, we always have

$$\frac{1 - \nu}{\gamma} > 1 - \nu. \qquad (2.27)$$

Suppose ξ^* is the A-optimal density corresponding to a value of ν (ν^*), less than ν_A. In order to prove that the A- and E-optimal designs corresponding to ν^* are equal, we aim to show that ξ^* is the E-optimal design. In other words, for any design ξ in Ω, the loss value based on E-optimality criteria correspond to ν^* is

$$L_E(\xi) \geqslant L_E(\xi^*).$$

Based on (2.26)

$$L_E\left(\xi\right) = \max(L_{E,1}\left(\xi\right), L_{E,2}\left(\xi\right)) \geqslant L_{E,2}\left(\xi\right),$$

$$L_{E,2}\left(\xi\right) = L_{A,2}\left(\xi\right) - \left(1 - \nu^*\right) \geqslant L_{A,2}\left(\xi^*\right) - \left(1 - \nu^*\right),$$

where $L_{A,2}\left(\xi^*\right) - \left(1 - \nu^*\right) = L_{E,2}\left(\xi^*\right).$ Therefore

$$L_E\left(\xi\right) \geqslant L_{E,2}\left(\xi^*\right).$$

On the other hand, since ξ^* is the A-optimal density corresponding to ν^*,

$$L_{A,2}\left(\xi^*\right) - \left(1 - \nu^*\right) \geqslant L_{A,1}\left(\xi^*\right) - \left(1 - \nu^*\right),$$

and based on (2.27), we have

$$L_{A,1}\left(\xi^*\right) - \left(1 - \nu^*\right) \geqslant L_{A,1}\left(\xi^*\right) - \frac{1 - \nu^*}{\gamma} = L_{E,1}\left(\xi^*\right).$$

This means that $L_{E,2}\left(\xi^*\right) \geqslant L_{E,1}\left(\xi^*\right).$ Based on all of the above results

$$L_E\left(\xi\right) \geqslant \max(L_{E,1}\left(\xi^*\right), L_{E,2}\left(\xi^*\right)) = L_E\left(\xi^*\right).$$

Hence for all fixed values of ν less than ν_A, there exists an E-optimal design which is exactly the same as the A-optimal design. The existence of an E-optimal design for $\nu < \nu_A$, implies that $\nu_E \geqslant \nu_A$.

Chapter 3: Robust A- and E-optimal Designs

As demonstrated in the previous chapter, one part of the method used there for finding optimal design is not applicable using A- and E-optimality criteria for a straight regression model when the value of ν exceeds ν_A and ν_E respectively. In this chapter, we apply a method for minimizing the maximized loss function corresponding to A- and E-optimality criteria based on the following theorem, which does not require one to know in advance which eigenvalue is largest. Let $l_i(\xi)$, for $i = 0, 1$ be the loss functions using the first and second eigenvalues (of the matrix $\mathbf{B}^{-1}\,\mathbf{G}\mathbf{B}^{-1}$ at (2.18) for A-optimality criteria or of the matrix $\left[(1-\nu)\,\mathbf{B}^{-1} + \nu\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}\right]$ at (2.24) for E-optimality criteria) respectively such that we aim to minimize $l(\xi) = \max(l_0(\xi), l_1(\xi))$.

Theorem 2 *Suppose an experimenter intends to find an optimal design based on a loss function on the design space S. The experimenter has found the designs ξ_i that*

minimize $l_i(\xi)$ subject to the constraint $l_i(\xi) \geqslant l_{1-i}(\xi)$. Define

$$
\xi^* = \begin{cases} \xi_0 & if \quad l_0(\xi_0) \leqslant l_1(\xi_1), \\[2ex] \xi_1 & if \quad l_1(\xi_1) \leqslant l_0(\xi_0), \end{cases}
\tag{3.28}
$$

so that

$$
l(\xi^*) = \min(l_0(\xi_0), l_1(\xi_1)).
$$

Then ξ^* is the minimax or optimal design.

Proof: Let ξ be any design on the design space S . Then ξ belongs to one of the following two classes

$$
A = \{\xi \mid l_0(\xi) \geqslant l_1(\xi)\},
$$

$$
B = \{\xi \mid l_1(\xi) \geqslant l_0(\xi)\}.
$$

Considering two possible situations, based on the definition of the ξ_i we have

$$
l(\xi) = \begin{cases} l_0(\xi) \geqslant l_0(\xi_0) \geqslant \min(l_0(\xi_0), l_1(\xi_1)) & if \quad \xi \in A, \\[2ex] l_1(\xi) \geqslant l_1(\xi_1) \geqslant \min(l_0(\xi_0), l_1(\xi_1)) & if \quad \xi \in B. \end{cases}
$$

On the other hand, based on the definition of the ξ_i, we have

$$
l_0(\xi_0) \geqslant l_1(\xi_0) \text{ and } l_1(\xi_1) \geqslant l_0(\xi_1).
\tag{3.29}
$$

Suppose that $\xi^* = \xi_0$. Then

64

$$l_1\left(\xi^*\right) = l_1\left(\xi_0\right) \overset{3.28}{\leqslant} l_0\left(\xi_0\right) = l_0\left(\xi^*\right)$$

$$\Rightarrow l\left(\xi^*\right) = \max\left(l_0\left(\xi^*\right), l_1\left(\xi^*\right)\right) = l_0\left(\xi^*\right) = l_0\left(\xi_0\right) \overset{3.29}{\leqslant} l_1\left(\xi_1\right)$$

$$\Rightarrow l\left(\xi^*\right) \leqslant \min\left(l_0\left(\xi_0\right), l_1\left(\xi_1\right)\right).$$

Similarly if $\xi^* = \xi_1$, we have

$$l\left(\xi^*\right) \leqslant \min\left(l_0\left(\xi_0\right), l_1\left(\xi_1\right)\right).$$

Thus $l\left(\xi\right) \geqslant l\left(\xi^*\right)$ for any design ξ from the design space S. Hence ξ^* is the optimal design.

3.7 Minimax Design Using A-optimality

Given that, as noted in the previous chapter, the part of the minimax method for A-optimality (related to minimizing the loss function over all possible absolutely continuous designs) was not applicable to $\nu \geqslant \nu_A$, we must return to the result of the first part of this method (2.18) – maximizing the loss function corresponding to A-optimality over all contamination terms in Huber's class.

We aim to minimize the function

$$L_A\left(\xi\right) = \left(1 - \nu\right)\left(1 + \frac{1}{\gamma}\right) + \nu\left(\max\left(\int_S m^2(x)dx - 1, \frac{1}{\gamma^2}\int_S x^2 m^2(x)dx - 12\right)\right),$$

over the class of all possible symmetric absolutely continuous designs (Ω) on the design space $S = [-\frac{1}{2}, \frac{1}{2}]$.

In order to minimize this function over ξ, we use the Lagrange Multiplier method with different set of constraints than those which we applied in the last approach. Following Theorem 2, we minimize

$$L_{A_0}(\xi) = (1-\nu)\left(1 + \frac{1}{\gamma}\right) + \nu\left(\int_S m^2(x)dx - 1\right),$$

subject to the following constraints

$$\begin{cases} \int_S m(x)dx = 1, \\ \int_S x^2 m(x)dx = \gamma, \\ \int_S m^2(x)dx - 1 \geq \frac{1}{\gamma^2}\int_S x^2 m^2(x)dx - 12. \end{cases}$$

The Lagrangian is

$$\int_S m^2(x)dx + \lambda_1 \int_S m(x)dx + \lambda_2 \int_S x^2 m(x)dx + \lambda_3\left(-\int_S m^2(x)dx + \frac{1}{\gamma^2}\int_S x^2 m^2(x)dx - 11\right),$$

where $\lambda_3 \geqslant 0$. The integrand

$$\left(1 - \lambda_3 + \frac{\lambda_3 x^2}{\gamma^2}\right) m^2(x) + \left(\lambda_1 + \lambda_2 x^2 - 11\lambda_3\right) m(x),$$

is minimized pointwise over $m(x)(\geq 0)$ by

$$m_0(x) = \left(\frac{-\lambda_1 + 11\lambda_3 - \lambda_2 x^2}{2 \left(1 - \lambda_3 + \frac{\lambda_3 x^2}{\gamma^2} \right)} \right)^+,$$

or it might be presented as a rational function

$$m_0(x) = \left(\frac{a + bx^2}{c + dx^2} \right)^+, \tag{3.30}$$

where a, b, c and d are some values in the real line. The form of the optimal density is the same as the optimal density found by Shi, Ye and Zhou (2003). They used the Lagrange Multiplier Rule from nonsmooth optimization theory and showed that the density of the optimal design takes the form

$$m(x) = \left(\frac{\mathbf{z}^T(x)\mathbf{B}\mathbf{z}(x) + d}{\mathbf{z}^T(x)\mathbf{D}\mathbf{z}(x)} \right)^+,$$

where $\mathbf{B}$ and $\mathbf{D}$ are constant symmetric matrices and d is constant. Then they used a numerical method, the SPlus function "nlmin", in order to determine $\mathbf{B}$, $\mathbf{D}$ and d.

We also obtain similar form for the optimal density at (3.30) if we minimize

$$L_{A_1}(\xi) = (1 - \nu)\left(1 + \frac{1}{\gamma} \right) + \nu \left(\frac{1}{\gamma^2} \int_S x^2 m^2(x)dx - 12 \right),$$

67

subject to the following constraints

$$
\begin{cases}
\int_S m(x)dx = 1, \\[2mm]
\int_S x^2 m(x)dx = \gamma, \\[2mm]
\frac{1}{\gamma^2}\int_S x^2 m^2(x)dx - 12 \geq \int_S m^2(x)dx - 1.
\end{cases}
$$

Hence, in each case the optimal density is in the form of (3.30). Then we use a numerical method to find optimal density at (3.30) for solving the minimization problem

$$
L_A(\xi) = (1-\nu)\left(1 + \frac{1}{\gamma}\right) + \nu\left(\max\left(\int_S m^2(x)dx - 1, \frac{1}{\gamma^2}\int_S x^2 m^2(x)dx - 12\right)\right),
$$
(3.31)

for any specific value of ν between zero and one. To enable easier computation of the optimal density, we simplify it to the following form

$$
m_0(x) = \left(\frac{\alpha + x^2}{\beta + \theta x^2}\right)^+.
$$
(3.32)

In the next step, as a check, we compute optimal densities for numerous values of ν less than ν_A using the following numerical minimization algorithm in Matlab. For these values of ν, based on the result of the previous chapter, we have correctly guessed that the second eigenvalue is the largest characteristic root to be minimized, and we have computed analytically optimal densities using the regular Lagrange Multiplier method outlined in the previous chapter. Using the results of the following numerical minimization algorithm, we are able to evaluate the accuracy of our new technique

by comparing them with the results of the technique used in the last chapter.

The function, fmincon, is one of the numerical minimization codes in Matlab that can be applied to problem (3.31). This function numerically computes a local minimum for our loss function (3.31) subject to some constraints. These constraints are defined such that the non-integrable and negative rational functions on the design space are excluded from all possible rational functions that can be searched by fmincon. For example, those rational functions that have two asymptotes in our design space are excluded by the condition

$$\beta(\beta + \frac{\theta}{4}) > 0.$$

Numerical methods always need a starting point at which to initialize the optimization. This starting point is very crucial in order to have a global minimum. We made a large sequence values of ν between zero and ν_A, and for the first non-zero value of ν we use the right starting point (by trial and error) in order to get an optimal density that has a minimum value for loss function (3.31) (the minimum value of (3.31) is computed analytically by the regular Lagrange Multiplier method). Then for the next value of ν in the sequence, we use value of the parameters of the last computed optimal density as a starting point for fmincon function, and we obtain the optimal density for this value of ν. This process continues for all values of ν in the sequence. The results of this process (the parameters of optimal densities, the loss value, and the second moment (γ)) for each value of ν in the sequence are compared with the corresponding values that are computed using the regular Lagrange Multi-

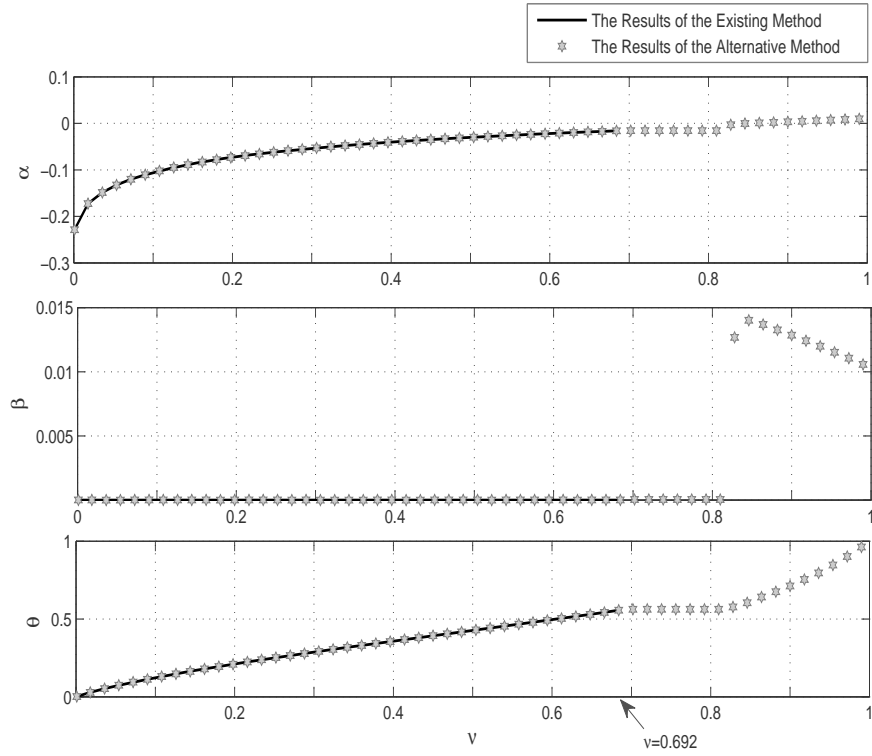Figure 3.13: The uppermost plot depicts the values of the first parameter (α) of the A-optimal densities (3.32) with various ν values (between zero and one) which are computed using two different techniques, and the second and third plots from the top show the values of the second and third parameters (β, θ) of the the A-optimal denisities with different ν values.

plier method (as described in the previous chapter). Figures 3.13 and 3.14 for ν less than ν_A show the difference between the results of these two techniques.

Figure 3.13 shows that the values of all parameters of the optimal densities for a $\nu \geqslant \nu_A$ behave as we expect, considering their behaviour for $\nu < \nu_A$, except the second parameter, which is still close to zero, but not as close as we would expect. Figure 3.14 reveals that the loss values, obtained by two methods, continuously decrease after one peak at around 0.115, and the second moment of the density continuously decreases when the value of ν increases to ν_A. These plots clearly reveal that these two techniques have very similar results, such that we can fairly confidently apply
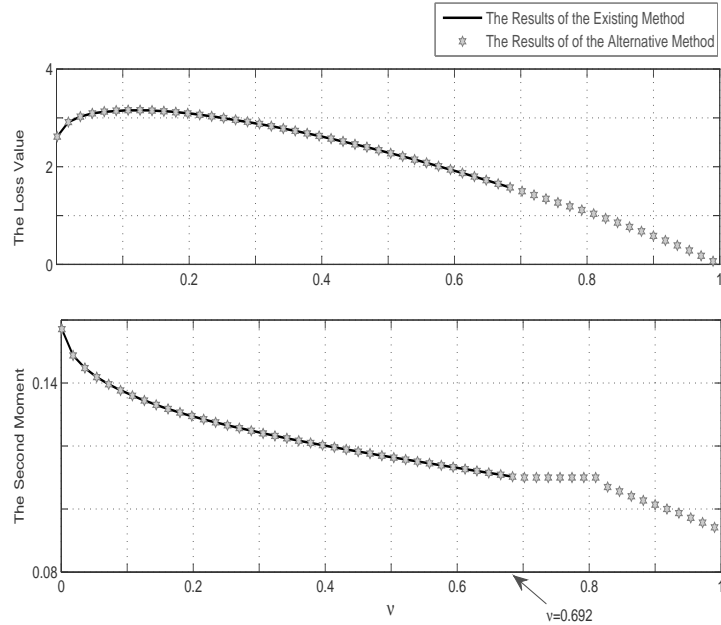
Figure 3.14: The upper and lower plots from the top depict the loss values and the second moment of the A-optimal densities for various ν values between zero and one, which are computed using two different techniques.

the second technique for the remaining values of ν in order to obtain the optimal density. Furthermore, the loss value in the first plot in Figure 3.14 also behaves as we expect, namely, it continuously decreases for values of ν greater than ν_A. However, the second moment is fixed for several values of ν (more than ν_A) and then decreases. Figure 3.15 shows that for values of ν more than ν_A, these two eigenvalues are equal.

On the other hand, when we look at the optimal densities for values of ν greater than ν_A, these densities take the form we expect. That is, we expect to have densities closer to uniform design for values of ν close to one, and these optimal densities behave as we expect (Figure 3.16).

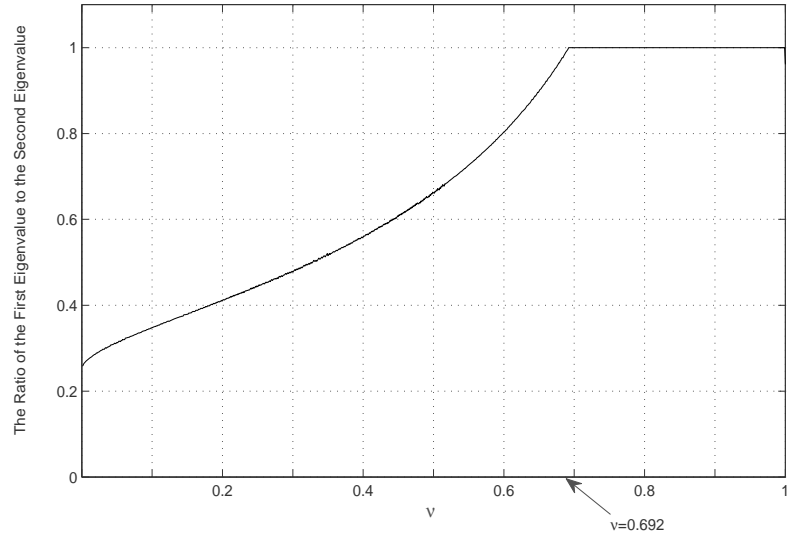Going through the above procedure in the neighbourhood of the parameters of

Figure 3.15: This plot presents the ratio of the first eigenvalue to the second eigenvalue for various ν values between zero and one. These eigenvalues are computed using the technique applied for the A-optimality criteria in this chapter.
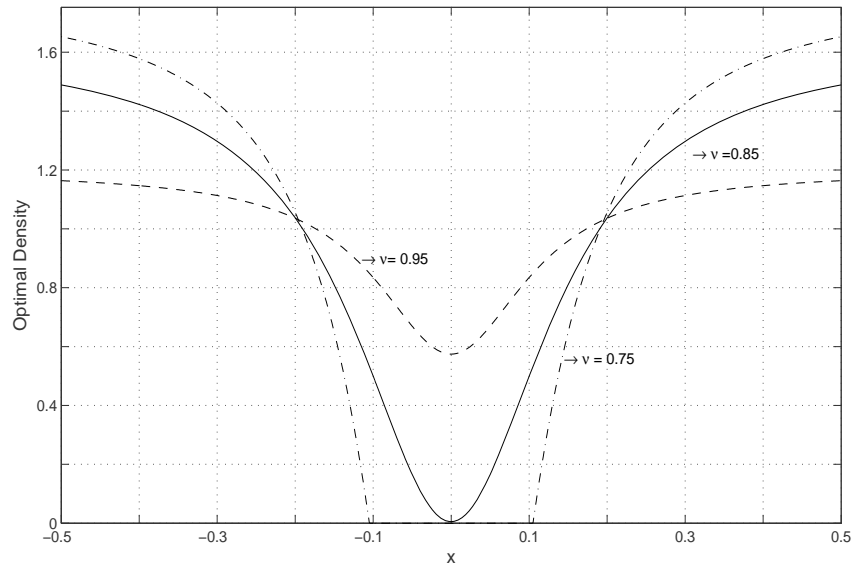

Figure 3.16: Three different A-optimal densities corresponding to three various ν values (0.75, 0.85, 0.95).
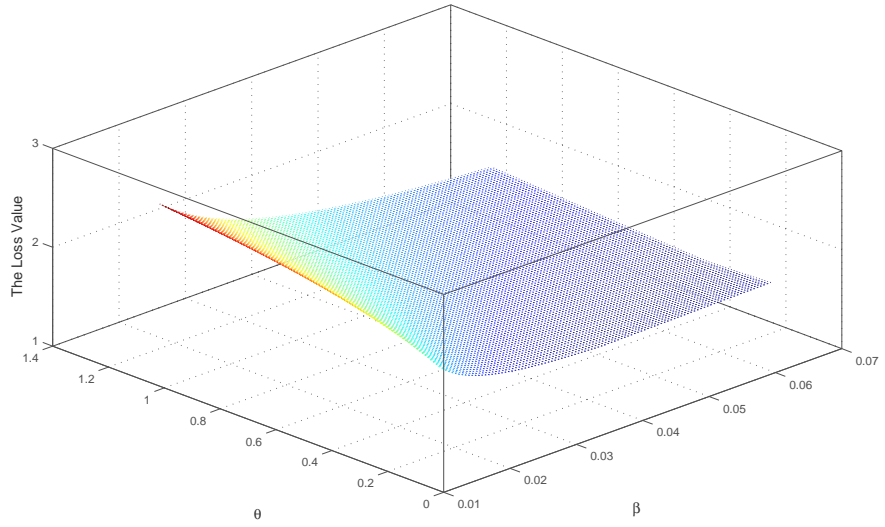
Figure 3.17: The loss value for different densities in the neighbourhood of the A-optimal density (3.32) by fixing the first parameter (α) of the A-optimal density and varying the other two parameters (β, θ) in a small neighbourhood around the values of these parameters at the A-optimal density.
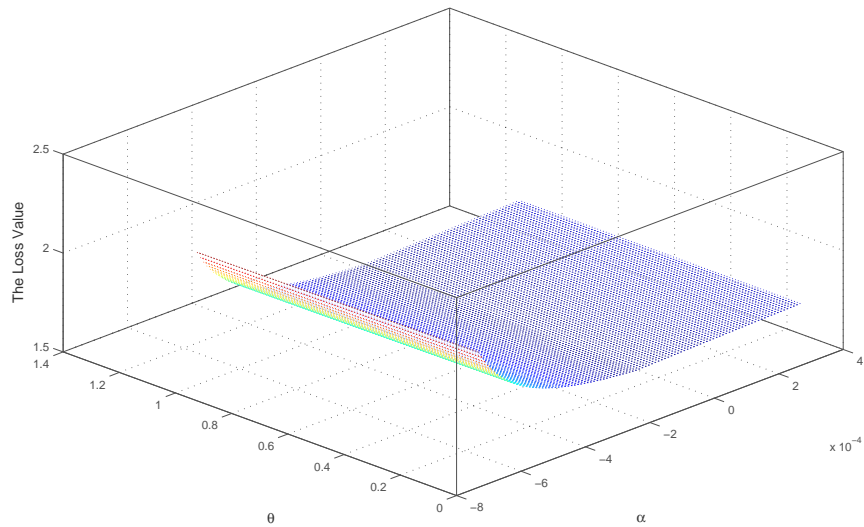

Figure 3.18: The loss value for different densities in the neighbourhood of the A-optimal density (3.32) by fixing the second parameter (β) of the A-optimal density and varying the other two parameters (α, θ) in a small neighbourhood around the values of these parameters at the A-optimal density.
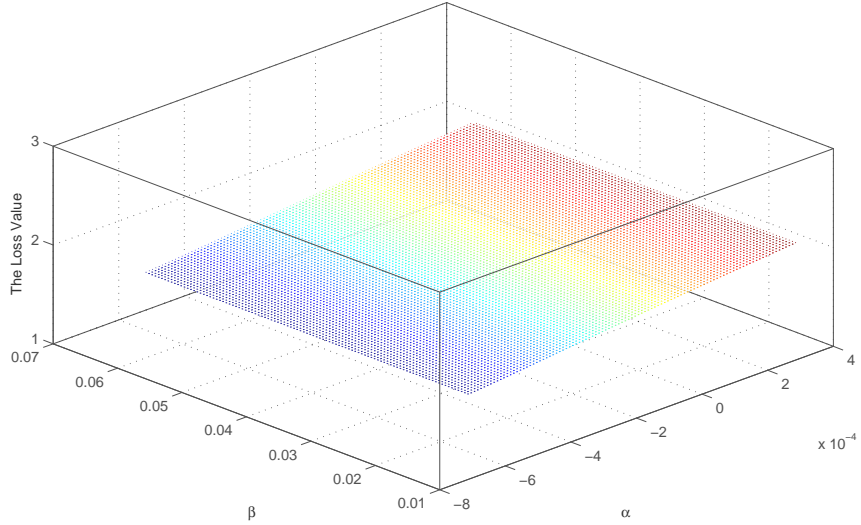
Figure 3.19: The loss value for different densities in the neighbourhood of the A-optimal density (3.32) by fixing the third parameter (θ) of the A-optimal density and varying the other two parameters (α, β) in a small neighbourhood of the values of these parameters at the A-optimal density.

the optimal density we compute that the loss values are the same as the minimum loss value corresponding to the optimal densities for the specific ν ($\nu = 0.85$). When we plot the loss values by fixing one parameter and varying two other parameters, the plot shows a flat curvature in space around the minimum loss value, and the plots of the rational functions at (3.32) corresponding to these parameters are the same as the optimal density analytically computed in the previous chapter.

3.8   Minimax Design Using E-optimality

Proceeding as in the previous section, we minimize

$$L_E\left(\xi\right) = \max\left[\left(1 - \nu\right) + \nu\left(\kappa_0 - 1\right), \ \frac{1 - \nu}{\gamma} + \nu\left(\frac{\kappa_1}{\gamma^2} - 12\right)\right]. \qquad (3.33)$$

74

For this, we first minimize

$$L_{E_0}(\xi) = (1-\nu) + \nu \left(\int_S m^2(x)dx - 1 \right),$$

subject to the constraints

$$\begin{cases} \int_S m(x)dx = 1, \\ \int_S x^2 m(x)dx = \gamma, \\ (1-\nu) + \nu \left(\int_S m^2(x)dx - 1 \right) \geq \frac{1-\nu}{\gamma} + \nu \left(\frac{1}{\gamma^2} \int_S x^2 m^2(x)dx - 12 \right). \end{cases}$$

Since the values of ν and γ are presumed to be fixed, we might summarize the minimization problem into the Lagrangian

$$\int_S m^2(x)dx + \lambda_1 \int_S m(x)dx + \lambda_2 \int_S x^2 m(x)dx$$
$$+ \lambda_3 \left(\frac{1-\nu}{\gamma} + \nu \left(\frac{1}{\gamma^2} \int_S x^2 m^2(x)dx - 12 \right) - (1-\nu) - \nu \left(\int_S m^2(x)dx - 1 \right) \right),$$

where $\lambda_3 \geqslant 0$. The integrand

$$\left(1 - \lambda_3\nu + \frac{\lambda_3\nu x^2}{\gamma^2} \right) m^2(x) + \left(\lambda_1 + \lambda_2 x^2 \right) m(x) + \left(\frac{1-\nu}{\gamma} - 1 - 10\nu \right) \lambda_3,$$

is minimized pointwise over $m(x)(\geq 0)$ using the following optimal density

$$m_0(x) = \left(\frac{-\lambda_1 - \lambda_2 x^2}{2 \left(1 - \lambda_3\nu + \frac{\lambda_3\nu x^2}{\gamma^2} \right)} \right)^+,$$

75

which might be presented as a rational function

$$m_0(x) = \left(\frac{a + bx^2}{c + dx^2} \right)^+,$$

where a, b, c and d are some values in the real line. As in the case of A-optimality, we obtain a similar optimal density by changing the assumption of largest eigenvalue,

$$L_{E_1}(\xi) = \frac{1 - \nu}{\gamma} + \nu \left(\frac{\kappa_1}{\gamma^2} - 12 \right),$$

subject to the constraints

$$\begin{cases} \int_S m(x)dx = 1, \\ \int_S x^2 m(x)dx = \gamma, \\ \frac{1-\nu}{\gamma} + \nu \left(\frac{1}{\gamma^2} \int_S x^2 m^2(x)dx - 12 \right) \geq (1 - \nu) + \nu \left(\int_S m^2(x)dx - 1 \right). \end{cases}$$

Therefore, the minimization problem (3.33) might be solved by this form of density

$$m_0(x) = \left(\frac{\alpha + x^2}{\beta + \theta x^2} \right)^+, \tag{3.34}$$

regardless of which eigenvalue is the largest.

In the final step, we proceed as before, in order to compute the optimal densities for numerous values of ν less than ν_E. For these values of ν, based on the result of the previous chapter, we correctly guessed the second eigenvalue to be the largest characteristic root to be minimized, and we analytically computed optimal densities.
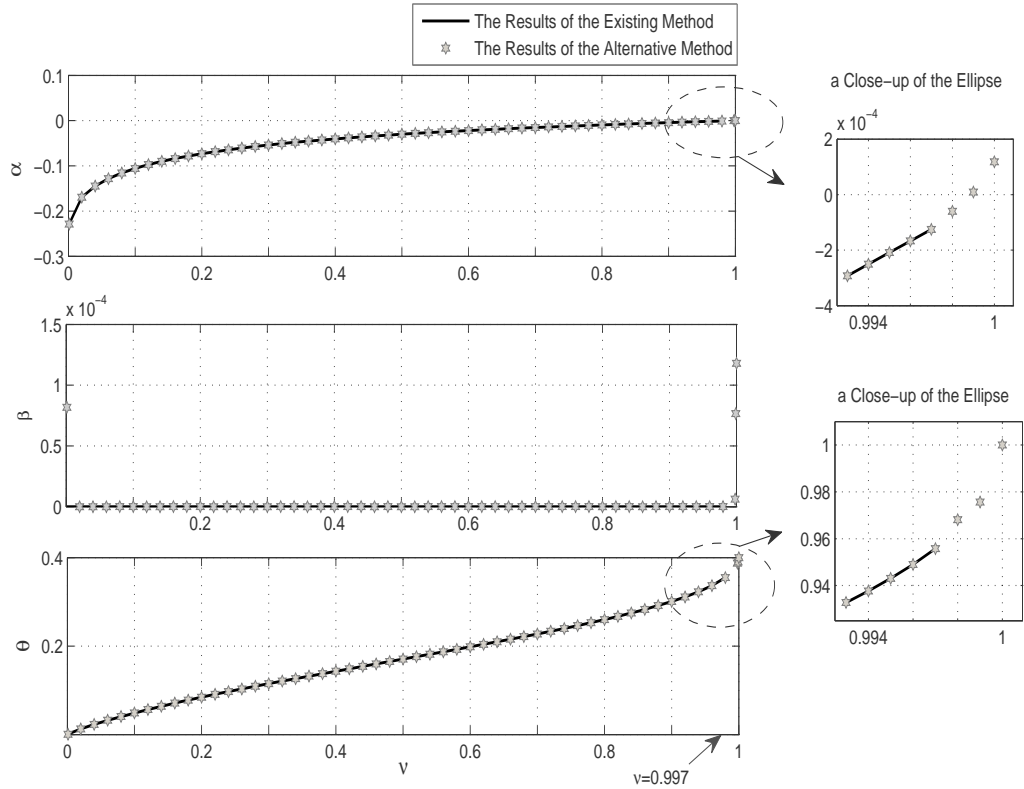
76

Figure 3.20: The uppermost plot depicts the values of the first parameter (α) of E-optimal densities (3.34) with various ν values (between zero and one) which are computed using two different techniques, and the second and third plots from the top depict the values of the second and third parameters (β, θ) of the E-optimal denisities with different ν values.

The results of this process (namely, the parameters of optimal densities, the loss value, and the second moment(γ)) for each value of ν in the sequence are compared with the corresponding values that have been computed using the regular Lagrange Multiplier method. The comparison shows that these results were nearly the same for this sequence. Figures 3.20 and 3.21 show the results of these two techniques for this sequence, and also the results of the new technique for some values of ν more than ν_E.

Figure 3.20 shows that the values of all parameters of the optimal densities for ν more than ν_E behave as we expect, considering their behaviours for ν less than ν_E.
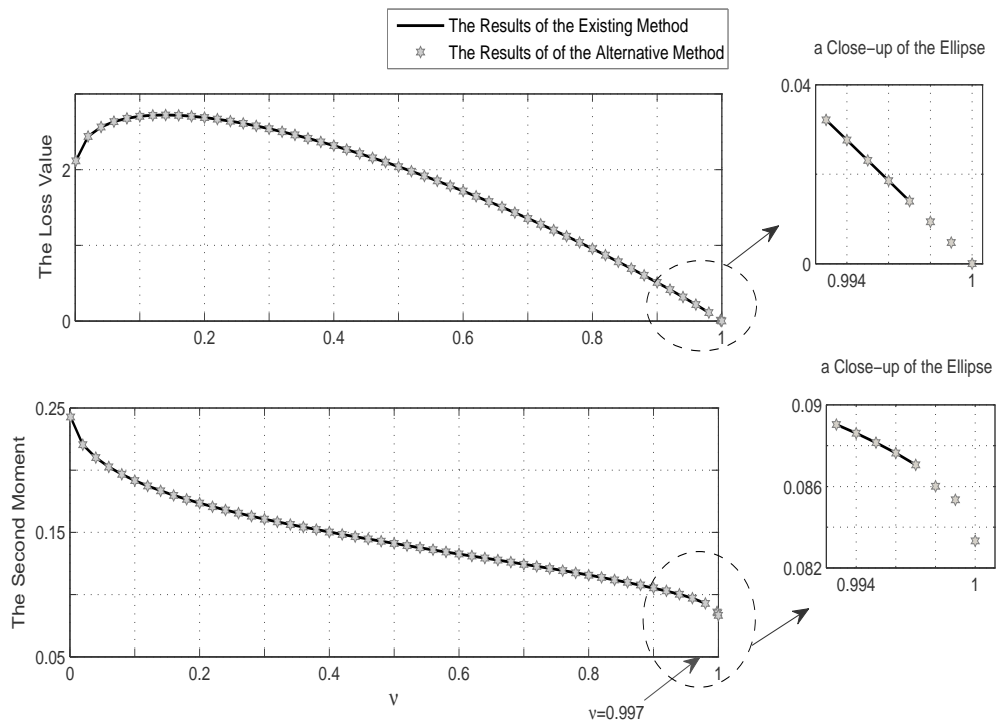
77

Figure 3.21: The upper and lower plots from the top show the loss values and the second moments of the E-optimal densities for various ν values between zero and one which are computed using two different techniques.
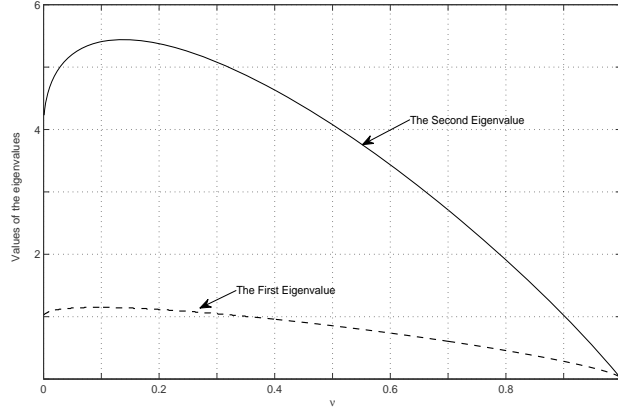
Figure 3.22: Values of the eigenvalues, $(1 - \nu) + \nu \left(\kappa_0 - 1\right)$ and $\frac{1-\nu}{\gamma} + \nu \left(\frac{\kappa_1}{\gamma^2} - 12\right)$, for various ν values between zero and one. These eigenvalues are computed using the technique applied for the E-optimality criteria in this chapter.

Furthermore, the loss value in the first plot, and the second moment in the second plot in Figure 3.21, also behave as we expect. They both decrease when ν increases. Figure 3.22 shows that for values of ν more than ν_E, these two eigenvalues are equal.

On the other hand, when we look at the optimal densities for values of ν greater than ν_E, these densities take the form that we expect. That is, we expect to have densities closer to uniform design for the values of ν close to one, and these optimal densities behave as we expect.

Figures 3.25, 3.24, and 3.26 reveal that for $\nu = 0.998$, the optimal density computed using above algorithm is not unique. In fact for some values of ν, when we change the value of the parameters of the optimal density in a small neighbourhood by fixing one parameter and varying two other parameters, a flat curvature appears around the minimum loss value.

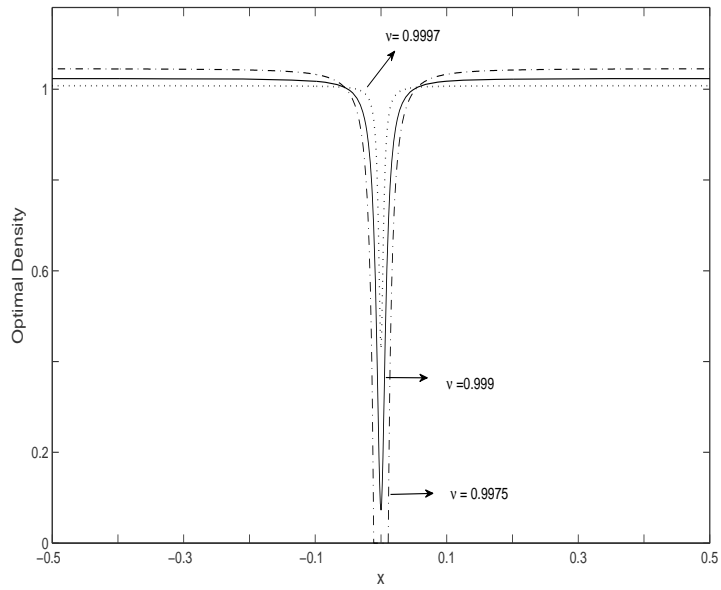Figure 3.27 shows the remedy for the gap in the method applied in the second

Figure 3.23: Three different E-optimal densities corresponding to three various ν values $(0.9975, 0.999, 0.9997)$.
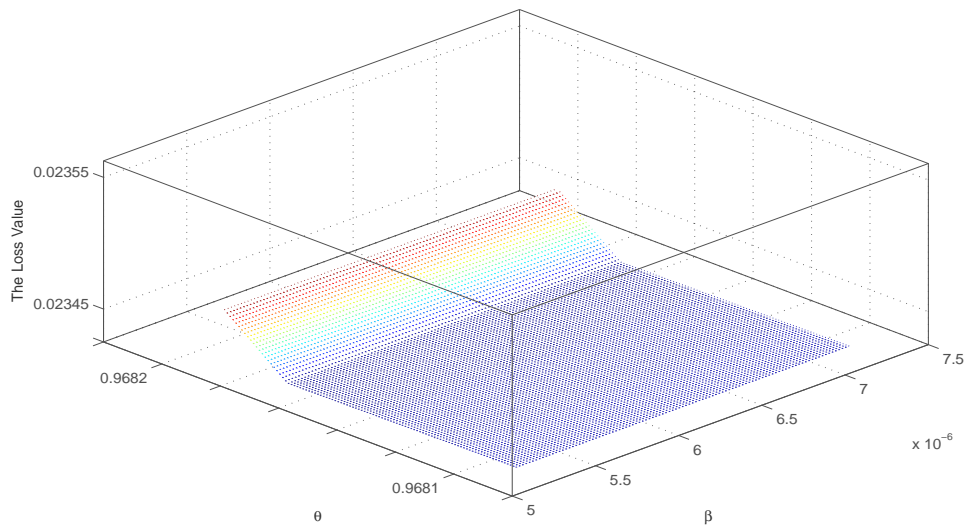

Figure 3.24: The loss value for different densities in the neighbourhood of the E-optimal density (3.34) by fixing the first parameter (α) of the E-optimal density and varying the other two parameters (β, θ) in a small neighbourhood around the values of these parameters at the E-optimal density.
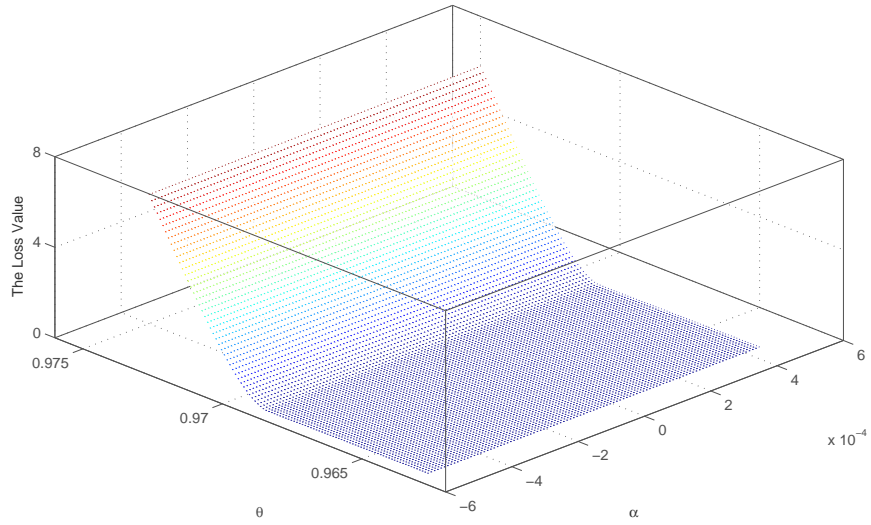
Figure 3.25: The loss value for different densities in the neighbourhood of the E-optimal density (3.34) by fixing the second parameter (β) of the E-optimal density and varying the other two parameters (α, θ) in a small neighbourhood around the values of these parameters at the E-optimal density.
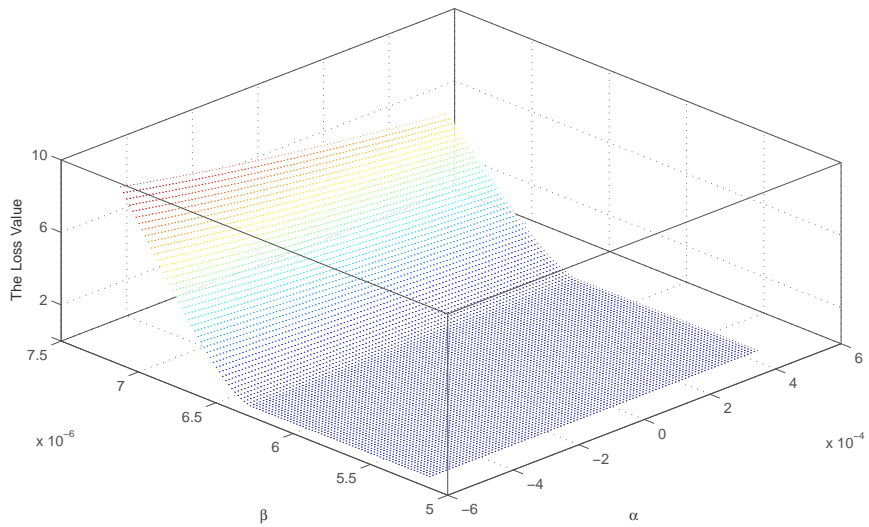

Figure 3.26: The loss value for different densities in the neighbourhood of the E-optimal density (3.34) by fixing the third parameter (θ) of the E-optimal density and varying the other two parameters (α, β) in a small neighbourhood around the values of these parameters at the E-optimal density.
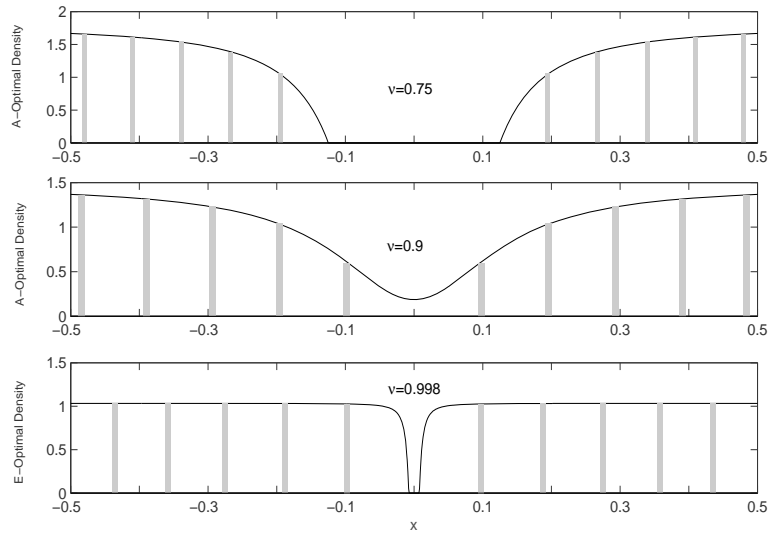
Figure 3.27: The solid line curves depict the A- and E-optimal densities for specific values of ν for the approximate straight line model. The bars show the implemented designs based on the matching moment method for $n = 10$ specific points.

chapter for A- and E- optimality, using the new technique. In this figure the number of distinct points in the implemented methods (Matching Moment), and also the replication in each point have been arbitrarily chosen. In the future it would be interesting to consider how to overcome this arbitrariness in order to make an appropriate implemented optimal design.

In this study, we first reviewed the existing method of acquiring the optimal design for different optimality criteria applied to the straight line regression model. Then we addressed a gap in the existing method of acquiring A- and E-optimal designs when the experimenter places a large enough emphasis on the errors from bias $(\nu_A \geqslant 0.692, \ \nu_E \geqslant 0.997)$. Finally, we showed that the form of the optimal robust density for A- and E-optimality criteria is a rational function (3.32) no matter which eigenvalue (of the matrix $\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}$ at (2.18) for A-optimality criteria or of the

82

matrix $\left[(1-\nu)\,\mathbf{B}^{-1}+\nu\mathbf{B}^{-1}\mathbf{G}\mathbf{B}^{-1}\right]$ at (2.24) for E-optimality criteria) is the largest.

Based on this important result, we were able to find A- and E-optimal densities for all situations - whether experimenters place more emphasis on errors from bias or on errors from variance.

Bibliography

Berger, M. P. F. and Wong, W. K. (2009), *An Introduction to Optimal Designs for Social and Biomedical Research*, Chichester, West Sussex, U.K: Wiley.

Bertsekas, D. P. (1982), *Constrained Optimization and Lagrange Multiplier Methods*, New York: Academic Press.

Box, G.E.P., and Draper, N.R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-654.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 29, 586–602.

Ehrenfeld, E. (1955), "On the Efficiency of Experimental Design," *Annals of Mathematical Statistics*, 26, 247–255.

Elfving, G. (1952), "Optimum Allocation in Linear Regression Theory," *Annals of Mathematical Statistics*, 23, 255–262.

Fang, K. T., and Wang, Y. (1994), *Number-theoretic Methods in Statistics* (1st ed.), London; New York: Chapman & Hall.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, New York: Academic Press.

Heo, G., Schmuland, B., and Wiens, D. P. (2001), "Restricted Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 29, 117-128.

Huber, P. J. (1975), "Robustness and Designs," in: *A Survey of Statistical Design and Linear Models*, ed. J.N. Srivastava, Amsterdam: North Holland, pp. 287-303.

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.

Karson, M. J., Manson, A. R., and Hader, R. J. (1969), "Minimum Bias Estimation and Experimental Design for Response Surfaces," *Technometrics*, 11, 461-475.

Kiefer, J. (1959), "Optimal Experimental Designs," *Journal of the Royal Statistical Society*, B 21, 273–319.

Kiefer, J., and Wolfowitz, J. (1960), "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, 12, 363-366.

Kiefer, J. (1974), "General Equivalence Theory for Optimum Designs (Approximate Theory)," *Ann. Statist.*, 2, 849-879.

Kiefer, J. (1975), "Optimal Design: Variation in Structure and Performance Under Change of Criterion," *Biometrika*, 62, 277-288.

Li, K. C. (1984), "Robust Regression Designs when the Design Space Consists of Finitely Many Points," *Annals of Statistics*, 12, 269–282.

Li, K.C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," *Journal of Statistical Planning and Inference*, 6, 135-151.

Marcus, M.B., and Sacks, J. (1976), "Robust Designs for Regression Problems," in: *Statistical Theory and Related Topics II*, ed. S.S. Gupta and D.S. Moore, New York: Academic Press, pp. 245-268.

Pukelsheim, F., and Rieder, S. (1992), "Efficient Rounding of Approximate Designs," *Biometrika*, 79, 763-770.

Pesotchinsky, L. (1982), "Optimal Robust Designs: Linear Regression in R^k," *The Annals of Statistics*, 10, 511-525.

Sacks, J., and Ylvisaker, D. (1978), "Linear Estimation for Approximately Linear Models," *The Annals of Statistics*, 6, 1122-1137.

Shi, P., Ye, J., and Zhou, J. (2003), "Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 31, 397-414.

Sibson, R. (1974), "D_A–Optimality and Duality," *Progress in Statistics. Proceeding of the 9th European Meeting of Statisticians, Budapest*, 677–692.

Smith, K. (1918), "On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and its Constants and the Guidance They Give Towards a Proper Choice of the Distribution of the Observations," *Biometrika,* 12, 1–85.

Stigler, S. M. (1971), "Optimal Experimental Design for Polynomial Regression," *Journal of the American Statistical Association*, 66, 311-318.

Xu, X., and Yuen, W. K. (2011), "Applications and Implementations of Continuous Robust Designs," *Communications in Statistics - Theory and Methods*, 40:6, 969–988.

Wald, A. (1943), "On the Efficient Design of Statistical Investigations," *Annals of Mathematical Statistics*, 14, 134–140.

Wiens, D. P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*; 127, 327 - 340.

Wiens, D. P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference,* 31, 353-371.

Wiens, D. P., and Zhou, J. (1996), "Minimax Regression Designs for Approximately Linear Models With Autocorrelated Errors," *Journal of Statistical Planning and Inference*, 55, 95-106.