Optimal Regression Design under Second-Order Least Squares Estimator: Theory, Algorithm and Applications

Chi-Kuang Yeh

2024-05-18

Contents

Chapter 1

Preface

This thesis is written by Chi-Kuang Yeh, under the guidance of Professor Julie Zhou. I create this bookdown project to fix some typos within it.

1.1 Chi-Kuang Yeh

B.Sc. (Hons.), University of Victoria, 2016

1.2 Supervisory Committee

Supervisor: Dr. Julie Zhou (Professor, Department of Mathematics and Statistics)

Departmental Member: Dr. Xuekui Zhang (Assistant Professor, Department of Mathematics and Statistics)

External Examiner: Dr. Xiaodai Dong (Professor, Department of Electrical and Computer Engineering)

1.3 Abstract

In this thesis, we first review the current development of optimal regression designs under the second-order least squares estimator in the literature. The criteria include A- and D-optimality. We then introduce a new formulation of A-optimality criterion so the result can be extended to c-optimality which has not been studied before. Following Kiefer's equivalence results, we derive the optimality conditions for A-, c- and D-optimal designs under the second-order least squares estimator. In addition, we study the number of support points for various regression models including Peleg models, trigonometric models, regular and fractional polynomial models. A generalized scale invariance property for

D-optimal designs is also explored. Furthermore, we discuss one computing algorithm to find optimal designs numerically. Several interesting applications are presented and related MATLAB code are provided in the thesis.

1.4 List of Abbreviations

- **BLUE**: Best Linear Unbiased Estimator
- **CVX**: A MATLAB-based modeling system for convex optimization
- **FRM**: Fractional Polynomial Model
- **GSI**: Generalized Scale Invariance
- **MATLAB**: A script language that uses matrix as default format and a computing environment, which is developed by MathWorks
- **OLSE**: Ordinary Least Squares Estimator
- **SDP**: Semi-Definite Programming
- **SLSE**: Second-Order Least Squares Estimator

Chapter 2

Introduction

Design of experiment is a sub-field in statistics that has a long history in its developments. After Sir Ronald A. Fisher's pioneering work on the analysis of variance and fractional factorial design concepts, while working in Rothamsted Experimental Station in the 1920s and 1930s (**?**, p. 21), many statisticians worked in this research area and made significant contributions. **?** gave many vital examples in the development of optimal design theory over the years, including **?**, **?**, **?**, **?**, **?** and **?**. Their inputs to this field have had huge impacts towards today's design framework. One century later, design techniques have been found to be effective and now are widely used in other disciplines, such as agriculture, engineering, environmental science, biomedical and pharmaceutical studies. See examples in **?**, **?**, **?**, **?**, **?**, **?** and **?**. Its primary objective is to study the cause and effect between variables under some systematic approaches and procedures.

Over the decades, many theoretical results and algorithms have been developed to construct different kinds of optimal designs, which include factorial design, fractional factorial design, response surface design, and regression design. In this thesis, we will focus on optimal regression design under a recently proposed statistical estimator, second-order least squares estimator (SLSE) in **?**.

2.1 Optimal regression design problem

"Regression analysis is a statistical technique for investigating and modeling the relationship between variables" (**?**). It is one of the most widely used statistical methods to explore the relationship between variables based on observed data. Although there are different kinds of regression models, we mainly focus on one response linear and nonlinear regression models.

Suppose we want to study the relationship between $x \in \mathbb{R}^p$ (a vector of explanatory variables) and y (a response variable). Consider a general regression model

for (x_i, y_i),

$$y_i = g(x_i; \theta) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\theta \in \mathbb{R}^q$ is an unknown parameter vector, n is the sample size, $g(\cdot)$ is a known linear or nonlinear expectation function depending on θ, and ϵ_i is a random error of the regression model. The random error $\epsilon_i's$ are assumed to be independent identically distributed with zero mean and variance σ^2. The random error terms are further assumed to be homoscedastic in this thesis. Since x_i and y_i are observed data and $g(\cdot)$ is also known, the only unknown component is θ. A question naturally comes in mind is how to estimate θ efficiently.

Suppose $\hat{\theta}$ is an estimator of θ. The design problem aims to get the most information about θ or $g(x_i; \theta)$ by selecting the best probability distribution on $x_1, x_2, \ldots, x_n$ that maximizes some scalar functions of the Fisher's information matrix of $\hat{\theta}$. The sample points x_i and space are called design points and design space, respectively. It is known that Fisher's information matrix is proportional to the inverse of the variance-covariance matrix of $\hat{\theta}$. Thus, the design problem aims to minimize some scalar functions of the variance-covariance matrix of $\hat{\theta}$, which are called objective functions or loss functions. The resulted probability measure ξ contains two components that are the support points and the corresponding probabilities associated with these points. The choice of the loss functions is determined based on the design interest. Various design criteria have been studied in the literature, such as A- and D-optimality criteria. D-optimality is one of the most widely used design criterion that minimizes the determinant of the variance-covariance matrix. The most desired property of D-optimal design is its scale invariant property. A-optimality minimizes the trace that leads to minimize the sum of the variances of the estimated parameters. See **?**, **?**, **?**, **?**, and **?** for other optimality criteria.

Let us consider Gompertz growth model to illustrate a D-optimal design. The model is given by

$$y_i = \theta_1 e^{-\theta_2 e^{-\theta_3 x_i}} + \epsilon_i, \ i = 1, 2, \cdots, n, \quad \theta = (\theta_1, \theta_2, \theta_3)^\top, \quad x_i \in S,$$

where θ_1 describes the maximum growing capacity, θ_2 explains the initial status of the subject, θ_3 determines the growth rate, y is the overall growth at the current time point and x is the time. Note that x, θ_1 , θ_2 and θ_3 are assumed to be positive in this context. We want to study how one subject's total growth associated with time. The model has broad applications in biological science and cancer studies. See some examples in **?** and **?**. Suppose the design space S is $[0, 10]$ (i.e. $x \in [0, 10]$) and the true parameters of θ is given by $\theta_o = (1, 1, 1)^\top$. For this model, D-optimal design aims to select the probability measure on S that minimizes $\det(\hat{\theta})$, where $\hat{\theta}$ is the SLSE. The details of the SLSE will be discussed in Section **??** and Chapter **??**. The resulted probability measure under D-optimality is

$$\xi_D^* = \begin{bmatrix} 0.000 & 1.350 & 10.000 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}, \tag{2.2}$$

where the top row represents the support points and the second row describes the corresponding probabilities on the points. The resulted design has three support points at $x = 0.000$, 1.350 and 10.000 having equal weight $(1/3)$. The interpretation of (**??**) is that we will distribute resources evenly at three points, 0.000, 1.350 and 10.000. For instance, if the maximum number of runs available is fifteen due to the scarcity of resources, the researcher will make five observations at each of the three points. In Figure **??**, the line represents the behavior of expectation function $g(x; \theta)$ in the design space S, and the $*$ represents the support point.

Many studies are conducted by using ordinary least squares estimator (OLSE) as the estimator $(\widehat{\theta})$ in optimal regression design framework. OLSE is the best linear unbiased estimator (BLUE) in the regression context. However, if the error distribution is asymmetric, the SLSE is more efficient than OLSE from **?**, which is reviewed in next section.

2.2 Second-order least squares estimator

We first discuss the relationship between OLSE and SLSE, as well as the advantages of SLSE over OLSE. OLSE is an estimator to estimate the parameter vector θ in regression model (**??**), which is defined to be

$$\widehat{\theta} := \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} (y_i - g(x_i; \theta))^2.$$

The assumptions for using OLSE are: the error terms are assumed to be homoscedastic and independently identically distributed with zero mean and finite constant variance. It has many desired properties such as consistency and it is the BLUE, which is widely used. In practice, however, other estimators might outperform OLSE in some scenarios. If the error distribution is asymmetric, **?** has shown that SLSE is asymptotically more efficient than OLSE. When the random errors are symmetrically distributed, SLSE and OLSE have the same asymptotic efficiency. SLSE has caught attentions in optimal regression design context due to these reasons.

We now review some properties of the SLSE in the regression model ((**??**). SLSE is defined as

$$(\widehat{\theta}^{\top}, \widehat{\sigma}^2)^{\top} := \operatorname*{argmin}_{\theta, \sigma^2} \sum_{i=1}^{n} \begin{pmatrix} y_i - g(x_i; \theta) \\ y_i^2 - g^2(x_i; \theta) - \sigma^2 \end{pmatrix}^{\top} W(x_i) \begin{pmatrix} y_i - g(x_i; \theta) \\ y_i^2 - g^2(x_i; \theta) - \sigma^2 \end{pmatrix}.$$

Note that $W(x_i)$ is a 2×2 non-negative semi-definite matrix which may or may not depend on x_i (**?**). It is clear that SLSE is a natural extension of the OLSE which is defined based on the first-order difference function (i.e. $y_i - \mathbb{E}[y_i] = y_i - g(x_i; \theta)$). On the other hand, SLSE is defined using not only the fist-order difference function, but also second-order difference function (i.e. $y_i^2 -$

$\mathbb{E}[y_i^2] = y_i^2 - (g^2(x_i; \theta) + \sigma^2))$. One might think about the downsides of the SLSE after talking about the advantages of SLSE over OLSE. SLSE does have its disadvantages indeed. It is not a linear estimator and there is no closed-form solution. It requires more computational resources compared to the OLSE due to the nonlinearity. However, numerical results can be easily computed for SLSE nowadays. As the result, SLSE is a powerful alternative estimator to be considered in research studies and real-life applications.

2.3   Research problem

As introduced in the previous section, **?** showed that SLSE is asymptotically more efficient than OLSE when the error distribution is asymmetric. Optimal designs under the SLSE was proposed in **?**. **? ?** and **?** did further investigations, including the convexity analysis, numerical methods, transformation and scale invariance properties for both A- and D-optimality criteria. There are other commonly used design criteria under the SLSE that have not been studied in the literature. Our goal is to fill this gap by extending the results to other design criteria, as well as exploring and deriving more theoretical results for the optimal designs under the SLSE.

The rest of the thesis is organized as follows. Chapter 2 describes the detailed formulation of optimal regression designs under SLSE. We derive several analytical results including equivalence theorem and the number of support points in optimal designs. Chapter 3 explains how to use numerical algorithms to solve the proposed optimal regression design problems via convex programming. We also present several interesting applications of the optimal designs studied in this thesis. Chapter 4 provides concluding remarks and discusses possible future research topics. MATLAB code are given in the Appendix.

2.4   Main Contributions

Here is a summary of the main contributions in this thesis.

1. We have studied the c-optimality design criterion under the SLSE, which has not been studied before.

2. We have applied Kiefer's equivalence theorem (**?**) to obtain the conditions for the A-, c- and D-optimal designs under the SLSE.

3. We have obtained the number of support points in optimal designs under the SLSE analytically for various regression models.

4. We have studied the generalized scale invariance property of D-optimal designs under the SLSE.

5. We have given one efficient and effective computing algorithm based on the program in MATLAB for computing optimal designs under the SLSE

on discrete design spaces.

Chapter 3

Optimal Regression Designs Under SLSE

In this chapter, we first review the results and properties of optimal regression designs under the SLSE. We then derive several new analytical results for the optimal designs under the SLSE. We begin with recalling the formulation of the optimal regression designs under SLSE based on three different criteria, A-, D-, and c-optimality. The formulation of A-, and D-optimality under the SLSE was first proposed in **?** while A-optimality was further investigated in **?**. We formulate optimal design problems under A-optimality differently so that the properties can be extended to c-optimality which has not been studied yet. Equivalence results for verifying optimal designs are also obtained. In addition, analytical results are derived for the number of support points for several regression models.

3.1   Design criteria under SLSE

Let us introduce the notations first. Assume σ_o and θ_o are the true parameter values of σ and θ, respectively. Let $S \subset \mathbb{R}^p$ be the design space for x. Let $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ be the trace and determinant functions of a matrix, respectively. Moreover, let Ξ denote the class of all probability measures on S. Define, for any $\xi \in \Xi$,

$$g_1 = g_1(\xi, \theta_o) = \mathbb{E}_\xi\left[\frac{\partial g(x; \theta)}{\partial \theta}\Big|_{\theta=\theta_o}\right],$$

and

$$G_2 = G_2(\xi, \theta_o) = \mathbb{E}_\xi\left[\frac{\partial g(x; \theta)}{\partial \theta}\frac{\partial g(x; \theta)}{\partial \theta^\top}\Big|_{\theta=\theta_o}\right].$$

For a discrete probability measure $\xi \in \Xi$, we write it as

$$\xi = \begin{bmatrix} x_1 & x_2 & ... & x_m \\ p_1 & p_2 & ... & p_m \end{bmatrix},$$

where $x_1, x_2, ..., x_m$ are the support points in S, and $p_1, ..., p_m$ are the probabilities associated with those points. From **?**, the asymptotic variance-covariance matrix of $\hat{\theta}$ under the SLSE is given by

$$\mathbb{V}(\hat{\theta}) = \sigma_o^2(1 - t)(G_2 - tg_1g_1^\top)^{-1}, \tag{3.1}$$

where $t = \frac{\mu_3^2}{\sigma_o^2(\mu_4 - \sigma_o^4)}$, $\mu_3 = \mathbb{E}[\epsilon_i^3|x]$ and $\mu_4 = \mathbb{E}[\epsilon_i^4|x]$. Note that **?** discussed that $t \in [0, 1)$ for any error distributions, and $t = 0$ for symmetric error distributions. Define a matrix $J = J(\xi, \theta_o, t) = G_2 - tg_1g_1^T$. It is clear that matrix J is proportional to the inverse of the variance-covariance matrix (**??**). For the rest of the thesis, we will be working on the design problems using matrix J.

As discussed in Chapter **??**, we aim to minimize the loss functions in optimal design problems, and the loss functions for D-, A- and c-optimality criteria under SLSE can be expressed as

$$\begin{aligned} \phi_D(\xi, \theta_o, t) &= \det(J^{-1}(\xi, \theta_o, t)), \\ \phi_A(\xi, \theta_o, t) &= \operatorname{tr}(J^{-1}(\xi, \theta_o, t)), \\ \phi_c(\xi, \theta_o, t) &= c_1^\top J^{-1}(\xi, \theta_o, t)c_1, \end{aligned} \tag{3.2}$$

when J is non-singular, and c_1 is a given vector in $\mathbb{R}^q$. If J is singular, all the three loss functions are defined to be $+\infty$. We use ξ_D^*, ξ_A^* and ξ_c^* to denote for D-, A- and c-optimal designs, respectively. For two measures ξ_1 and $\xi_2 \in \Xi$, define $\xi_\alpha = (1 - \alpha)\xi_1 + \alpha\xi_2$ for $\alpha \in [0, 1]$.

Lemma 3.1. $\log(\phi_D(\xi_\alpha, \theta_o, t))$, $\phi_A(\xi_\alpha, \theta_o, t)$ and $\phi_c(\xi_\alpha, \theta_o, t)$ *are convex functions of* α.

The convexity results are discussed in **?** and **?**. Similar convexity results are given in **?**. We will use $\log(\det(J^{-1}(\xi, \theta_o, t)))$ for D-optimal design for the rest of the thesis as $\log(\cdot)$ is a monotonic increasing function which does not change the optimality.

Although we have formulated the loss functions, there are some issues associated with the formulation in @ref{eq-loss-J}. The reason is that J is lacking of linearity. From construction of J, $J(\xi_\alpha, \theta_o, t)$ is not a linear combination of $J(\xi_1, \theta_o, t)$ and $J(\xi_2, \theta_o, t)$. Thus, it is difficult to obtain the theoretical results using J. To solve this issue, **?** proposed an alternative expression for characterizing the loss functions. The key is to define a matrix

$$B(\xi) = B(\xi, \theta_o, t) = \begin{pmatrix} 1 & \sqrt{t}g_1^\top \\ \sqrt{t}g_1 & G_2 \end{pmatrix}, \tag{3.3}$$

which plays an important role in the following formulation. Note $B(\xi_\alpha)$ is now an affine function of α, i.e.,

$$B(\xi_\alpha) = (1 - \alpha)B(\xi_1) + \alpha B(\xi_2).$$

This fact ultimately makes B much more useful than J to study optimal designs under SLSE. The inverse of B is given as

$$B^{-1}(\xi) = B^{-1}(\xi, \theta_o, t) = \begin{pmatrix} \frac{1}{r} & \frac{-\sqrt{t}}{r}g_1^\top G_2^{-1} \\ \frac{-\sqrt{t}}{r}G_2^{-1}g_1 & J^{-1} \end{pmatrix}, \qquad (3.4)$$

where $r = 1 - tg_1^\top G_2^{-1}g_1$. Note that if J is invertible, G_2 must also be invertible since $G_2 = J + tg_1g_1^\top$ and $tg_1g_1^\top$ is positive semi-definite. Consequently, B^{-1} exists from (**??**). Now, we are going to present the following lemmas to characterize the loss functions for A-, c- and D-optimal design problems. Lemma **??** is slightly different from a result in **?**, Lemma **??** is a result from **?**, and Lemma **??** is a new result.

Lemma 3.2. *If J is invertible, then*

$$\phi_A(\xi, \theta_o, t) = \operatorname{tr}(J^{-1}) = \operatorname{tr}(C^\top B^{-1}C),$$

where $C = 0 \oplus I_q$, I_q denotes for the $q \times q$ identity matrix, and $\oplus$ denotes for matrix direct sum operator.

Proof. From (**??**) and $C = 0 \oplus I_q$, we get

$$C^\top B^{-1}C = \begin{pmatrix} 0 & 0 \\ 0 & I_q \end{pmatrix}^T \begin{pmatrix} \frac{1}{r} & \frac{-\sqrt{t}}{r}g_1^\top G_2^{-1} \\ \frac{-\sqrt{t}}{r}g_1 G_2^{-1} & J^{-1} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I_q \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 \\ 0 & J^{-1} \end{pmatrix},$$

which implies $\operatorname{tr}(C^\top B^{-1}C) = \operatorname{tr}(J^{-1})$. $\qquad\square$

Lemma 3.3. *If J is invertible, then*

$$\phi_D(\xi, \theta_o, t) = \det(J^{-1}) = \det(B^{-1}).$$

From @ref{eq:B-matrix}, we have

$$\det(B) = \det(1 - tg_1^T G_2^{-1}g_1)\det(G_2)$$
$$= \det(I - tg_1g_1^\top G_2^{-1})\det(G_2)$$
$$= \det(G_2 - tg_1g_1^T)$$
$$= \det(J),$$

which gives $\det(J^{-1}) = \det(B^{-1})$.

Lemma 3.4. *If J is invertible, then*

$$\phi_c(\xi, \theta_o, t) = c_1 \top J^{-1} c_1 = c^\top B^{-1} c,$$

where c_1 is a vector in $\mathbb{R}^q$ and $c^\top = (0, c_1^\top)$.

Thus, by Lemmas **??**, **??** and **??**, the alternative expressions for the loss functions in (**??**) are

$$
\begin{aligned}
\phi_D(\xi, \theta_o, t)) &= \det(B^{-1}(\xi, \theta_o, t)), \\
\phi_A(\xi, \theta_o, t) &= \operatorname{tr}(C^T B^{-1}(\xi, \theta_o, t) C), \\
\phi_c(\xi, \theta_o, t) &= c^T B^{-1}(\xi, \theta_o, t) c,
\end{aligned}
\tag{3.5}
$$

where $C = 0 \oplus I_q$, $c_1 \in \mathbb{R}^q$ and $c^T = (0, c_1{}^T)$. If B is singular, all the three loss functions are defined to be $+\infty$.

3.2   Equivalence theorem for optimal designs under SLSE

In this section we derive the optimality conditions for the optimal designs under the SLSE which follows from the equivalence theorem in **?** and **?**. We also analyze the minimum number of support points in optimal designs for various regression models, and theoretical results are obtained. Note we study approximate designs in this thesis. The advantages of working with approximate designs instead of exact design are well documented in **?**.

Define a vector $f(x, \theta_o) = \left.\frac{\partial g(x;\theta)}{\partial \theta}\right|_{\theta = \theta_o} \in \mathbb{R}^q$ and a matrix

$$
M(x) = M(x, \theta_o, t) = \begin{pmatrix} 1 & \sqrt{t} f^T(x, \theta_o) \\ \sqrt{t} f(x, \theta_o) & f(x, \theta_o) f^T(x, \theta_o) \end{pmatrix}_{(q+1) \times (q+1)}.
\tag{3.6}
$$

Then $B(\xi, \theta_o, t) = \mathbb{E}_\xi[M(x, \theta_o, t)]$. Define dispersion functions

$$
\begin{aligned}
d_D(x, \xi, t) &= \operatorname{tr}(B^{-1} M(x)) - (q + 1), \\
d_A(x, \xi, t) &= \operatorname{tr}(M(x) B^{-1} C^T C B^{-1}) - \operatorname{tr}(C B^{-1} C^T), \\
d_c(x, \xi, t) &= c^T B^{-1} M(x) B^{-1} c - c^T B^{-1} c,
\end{aligned}
\tag{3.7}
$$

where $B = B(\xi, \theta_o, t)$ is invertible.

Theorem 3.1. *We suppose all the dispersion functions are evaluated at θ_o. If ξ_D^*, ξ_A^* and ξ_c^* are the optimal probability measures for D-, A- and c- optimality, respectively, then B is invertible and for any $x \in S$,*

$$d_D(x, \xi_D^*, t) \le 0, \tag{3.8}$$

$$d_A(x, \xi_A^*, t) \le 0, \tag{3.9}$$

$$d_c(x, \xi_c^*, t) \le 0. \tag{3.10}$$

Proof. In this proof, we use $B(\xi)$ for $B(\xi, \theta_o, t)$ and $M(x)$ for $M(x, \theta_o, t)$. Suppose ξ^* is an optimal design to a criterion. Define $\xi_\alpha = (1-\alpha)\xi^* + \alpha\xi$ where ξ is an arbitrary probability measure. This proof is based on Kiefer's general equivalence theorem (**?**), and the optimal condition can be derived from $\frac{\partial \phi(\xi_\alpha)}{\partial \alpha}\big|_{\alpha=0} \geq 0$ for any measure $\xi \in \Xi$, where ϕ is a loss function.

We first prove (**??**). Let ξ_D^* be the optimal measure under D-optimality. We have

$$
\begin{aligned}
\frac{\partial \log(\phi_D(\xi_\alpha))}{\partial \alpha}\bigg|_{\alpha=0} &= -\operatorname{tr}(B^{-1}(\xi_D^*)(-B(\xi_D^*) + B(\xi))) \\
&= -\operatorname{tr}(B^{-1}(\xi_D^*)B(\xi)) + \operatorname{tr}(I_{q+1}) \\
&= -\operatorname{tr}(B^{-1}(\xi_D^*)B(\xi)) + (q+1)) \\
&= -\operatorname{tr}(B^{-1}(\xi_D^*)\mathbb{E}_\xi[M(x)] - (q+1)) \\
&= -\mathbb{E}_\xi[d_D(x, \xi_D^*, t)] \\
&\geq 0, \text{ for any } \xi \text{ on } S,
\end{aligned}
$$

which implies $d_D(x, \xi_D^*, t) \leq 0$, for all $x \in S$.

To prove (**??**), let ξ_A^* be the optimal measure under A-optimality. We have

$$
\begin{aligned}
\frac{\partial \phi_A(\xi_\alpha)}{\partial \alpha}\bigg|_{\alpha=0} &= -\operatorname{tr}(C^T B^{-1}(\xi_A^*)[B(\xi) - B(\xi_A^*)]B^{-1}(\xi_A^*)C) \\
&= -\operatorname{tr}(C^T B^{-1}(\xi_A^*)B(\xi)B^{-1}(\xi_A^*)C) + \operatorname{tr}(C^T B^{-1}(\xi_A^*)C) \\
&= -\operatorname{tr}(C^T B^{-1}(\xi_A^*)\mathbb{E}_\xi[M(x)]B^{-1}(\xi_A^*)C) + \operatorname{tr}(C^T B^{-1}(\xi^*)C) \\
&= -(\operatorname{tr}(\mathbb{E}_\xi[M_\xi(x, \theta_o)]B^{-1}(\xi_A^*)CC^T B^{-1}(\xi^*)) - \operatorname{tr}(C^T B^{-1}(\xi_A^*)C)) \\
&= -\mathbb{E}_\xi[d_A(x, \xi_A^*, t)] \\
&\geq 0, \text{ for any } \xi \text{ on } S,
\end{aligned}
$$

which implies $d_A(x, \xi_A^*, t) \leq 0$, for all $x \in S$.

Lastly, to prove (**??**), let ξ_c^* be the optimal measure under c-optimality. We have

$$
\begin{aligned}
\frac{\partial \phi_c(\xi_\alpha)}{\partial \alpha}\bigg|_{\alpha=0} &= -c^T B^{-1}(\xi_c^*)[-B(\xi) + B(\xi_c^*)]B^{-1}(\xi_c^*)c \\
&= -c^T B^{-1}(\xi_c^*)B(\xi)B^{-1}(\xi_c^*)c + c^T B^{-1}(\xi_c^*)B(\xi_c^*)B^{-1}(\xi_c^*)c \\
&= -c^T B^{-1}(\xi_c^*)\mathbb{E}_\xi[M(x)]B^{-1}(\xi_c^*)c + c^T B(\xi_c^*)^{-1}c \\
&= -\mathbb{E}_\xi[c^T B^{-1}(\xi_c^*)M(x)B^{-1}(\xi_c^*)c + c^T B(\xi_c^*)^{-1}c] \\
&= -\mathbb{E}_\xi[d_c(x, \xi_c^*, t)] \\
&\geq 0, \text{ for any } \xi \text{ on } S,
\end{aligned}
$$

which implies $d_c(x, \xi_c^*, t) \leq 0$, for all $x \in S$. $\qquad\square$

3.3   Results on the number of support points

Using the results in Theorem **??** , we can explore the properties of the optimal designs. In **?** and **?**, there are some discussions about the number of support points based on computational results. However, there is little discussion on the number of support points theoretically in **?**, and there is still a large gap to be filled in. Hence we derive several results about the number of support points for various models, including polynomial models, fractional polynomial models, Michaelis-Menton model, Peleg model and trigonometric models.

A polynomial regression model of degree q ($q \geq 1$) without intercept is given by

$$y_i = \theta_1 x_i + \theta_2 x_i^2 + \cdots + \theta_q x_i^q + \epsilon_i, \ x_i \in S = [-1, +1], \ i = 1, 2, \cdots, n. \quad (3.11)$$

Polynomial regression models are widely used when the response and regressors have curvilinear relationship. Complex nonlinear relationships can be well approximated by polynomials over a small range of the explanatory variables (**?**, p. 223). There are different kinds of polynomial models such as orthogonal polynomial models, multi-variable polynomial models, and one variable polynomial models. Polynomial models are often used in design of experiment for the response surface methodology, and there are many applications in industry. For example, see **?**, **?** and **?**.

A- and D-optimal designs for (**??**) under SLSE are symmetric on S [**?**,**?**]. In (**??**), we have $f(x, \theta) = (x, x^2, \cdots, x^q)^\top$ and

$$M(x, \theta_o, t) = \begin{pmatrix} 1 & \sqrt{t}x & \sqrt{t}x^2 & ... & \sqrt{t}x^q \\ \sqrt{t}x & x^2 & x^3 & \cdots & x^{q+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sqrt{t}x^q & x^{q+1} & ... & ... & x^{2q} \end{pmatrix}_{(q+1) \times (q+1)}. \quad (3.12)$$

Theorem 3.2. *Let* n_A *and* n_D *denote the minimum number of support points in A- and D-optimal designs under SLSE, respectively. For* (**??**)*, we have*

$$n_A = q \text{ or } q+1, \quad (3.13)$$

and

$$n_D = q \text{ or } q+1. \quad (3.14)$$

Proof. The proof includes the following three parts.

(i).   From @ref(thm:dispersion} and (**??**), we can see that $d_A(x, \xi_A^*, t)$ and $d_D(x, \xi_D^*, t)$ are polynomial functions of x with highest degree $2q$. By fundamental theorem of algebra, there are exactly $2q$ roots for x in equations $d_A(x, \xi_A^*, t) = 0$ and $d_D(x, \xi_D^*, t) = 0$. However, we have at most $2q$ real roots.

(ii). By the construction, the determinant of B matrix is not zero if and only if the determinant of G_2 is not zero. Therefore, there are at least q support points in ξ.

(iii). Both boundary points are the support points, so the number of support points are at most $2q - 2$ in the interval $(-1, +1)$. From the equivalence theorem, we know that the dispersion functions are all less or equal to zero (i.e. $d_A(x, \xi_A^*, t) \leq 0$ and $d_D(x, \xi_D^*, t) \leq 0$), so all those support points in $(-1, +1)$ have a multiplicity of two. In total, we have at most $2 + \frac{(2q-2)}{2} = q + 1$ distinct support points.

Thus, the number of support points in ξ_A^* and ξ_D^* is either q or $q+1$. \end{proof}

$\square$

Example 3.1. Consider (**??**) with $q = 2$ and $S = [-1, +1]$. Let $\eta_m = \mathbb{E}[x^m]$ be the m-th moment of distribution $\xi(x)$. **?** showed that the D-optimal design is symmetric in this case (i.e. $\eta_1 = \eta_3 = 0$), and is given by

$$
\xi_D^* = \begin{cases}
\begin{bmatrix} -1 & +1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, & \text{for } t \in [0, \frac{2}{3}), \\[2em]
\begin{bmatrix} -1 & 0 & +1 \\ \frac{1}{3t} & \frac{3t-2}{3t} & \frac{1}{3t} \end{bmatrix}, & \text{for } t \in [\frac{2}{3}, 1).
\end{cases}
$$

We now study the A-optimal design. By taking the advantage of the symmetric result in **?**, we have

$$
B(\xi) = \begin{pmatrix} 1 & 0 & \sqrt{t}\eta_2 \\ 0 & \eta_2 & 0 \\ \sqrt{t}\eta_2 & 0 & \eta_4 \end{pmatrix}, \text{ and } B^{-1}(\xi) = \begin{pmatrix} \frac{\eta_4}{\eta_4 - t\eta_2} & 0 & \frac{\sqrt{t}\eta_2}{t\eta_2^2 - \eta_4} \\ 0 & \frac{1}{\eta_2} & 0 \\ \frac{\sqrt{t}\eta_2}{t\eta_2^2 - \eta_4} & 0 & \frac{1}{\eta_4 - t\eta_2^2} \end{pmatrix}.
$$

From **?**, on $S = [-1, +1]$, the even moments of any distributions must satisfy $0 \leq \eta_2^2 \leq \eta_4 \leq \eta_2 \leq 1$. Then our loss function can be expressed as

$$
\phi_A(\xi) = \text{tr}(C^\top B^{-1} C) = \frac{1}{\eta_2} + \frac{1}{\eta_4 - t\eta_2^2}.
$$

The optimal design problem can be written as

$$
\begin{aligned}
\min_{\eta_2, \eta_4} \quad & \frac{1}{\eta_2} + \frac{1}{\eta_4 - t\eta_2^2} \\
\text{s.t.} \quad & 0 \leq \eta_2^2 \leq \eta_4 \leq \eta_2 \leq 1.
\end{aligned}
$$

In order to minimize the loss function, we first fix η_2. Then it is trivial to see we should make η_4 as big as possible, which leads to $\eta_4 = \eta_2$, its ceiling. Now the question becomes

$$
\begin{aligned}
\min_{\eta_2} \quad & \frac{1}{\eta_2} + \frac{1}{\eta_2 - t\eta_2^2} \\
\text{s.t.} \quad & 0 \leq \eta_2 \leq 1.
\end{aligned}
$$

It yields to $\eta_2 = \frac{2-\sqrt{2}}{t}$ or $\eta_2 = \frac{2+\sqrt{2}}{t}$ where $t \neq 0$. Note $t \in [0, +1)$ and $\eta_2 \in [0, +1]$, so the latter solution should be excluded as it is not within the feasible region. Consequently,

$$\eta_2 = \begin{cases} \min\left(\frac{2-\sqrt{2}}{t}, 1\right), & t \in (0, 1), \\ 1, & t = 0. \end{cases}$$

By the symmetric property, the A-optimal design is

$$\xi_A^* = \begin{bmatrix} -1 & 0 & +1 \\ \frac{\eta_2}{2} & 1-\eta_2 & \frac{\eta_2}{2} \end{bmatrix}.$$

Here we have three cases:\ **Case 1** When $t = 0$,

$$\xi_A^* = \begin{bmatrix} -1 & +1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Case 2 When $t \in (0, 2-\sqrt{2})$,

$$\xi_A^* = \begin{bmatrix} -1 & +1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Case 3 When $t \in [2-\sqrt{2}, 1)$,

$$\xi_A^* = \begin{bmatrix} -1 & 0 & +1 \\ \frac{2-\sqrt{2}}{2t} & 1-\frac{2-\sqrt{2}}{t} & \frac{2-\sqrt{2}}{2t} \end{bmatrix}.$$

In summary, the A-optimal design for (**??**) when $q = 2$ is

$$\xi_A^* = \begin{cases} \begin{bmatrix} -1 & +1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, & \text{for } t \in [0, \leq 2-\sqrt{2}), \\ \begin{bmatrix} -1 & 0 & +1 \\ \frac{2-\sqrt{2}}{2t} & \frac{2t+\sqrt{2}-2}{t} & \frac{2-\sqrt{2}}{2t} \end{bmatrix}, & \text{for } t \in [2-\sqrt{2}, 1). \end{cases}$$

From the results above, it is clear to observe that A- and D-optimal designs have either 2 or 3 support points which is consistent with Theorem **??**. $\square$

To show the result in Theorem **??**, we plot $d_A(x, \xi^*, t)$ and $d_D(x, \xi^*, t)$ in Figure [TOFILL] using the optimal designs in Example **??**. It is clear that $d_A(x, \xi^*, t) \leq 0$ and $d_D(x, \xi^*, t) \leq 0$ for all $x \in S$, which is consistent with Theorem **??**. Also $d_A(x, \xi_A^*, t) = 0$ at the support points of ξ_A^* and $d_D(x, \xi_D^*, t) = 0$ at the support points of ξ_D^*. As t increases, the origin becomes a support point which changes the number of support points from 2 to 3. This is again consistent with Theorem **??**.

When the design space S is asymmetric, say $S = [-1, b]$, $b \in (-1, 1)$, the number of support points under A- and D-optimality are either q or $q+1$ for all $q \in \mathbb{Z}^+$.

When $t = 0$, $n_D = q$ for all $q \in \mathbb{Z}^+$. The derivation is similar to that of Theorem **??** and is omitted. We will show some computational results in Chapter **??**.

Fractional polynomial model (FPM) is given by

$$y = \theta_1 x^{q_1} + \theta_2 x^{q_2} + \cdots + \theta_p x^{q_p} + \epsilon, \ q_i \in \mathbb{Q}, \ \forall i, \tag{3.15}$$

which provides more flexible parameterization with wide range of applications in many disciplines. For instance, **?** used FPM for longitudinal epidemiological studies, and **?** applied this model to analyze medical data. This model has also been studied for optimal designs. For example, **?** used FPM to study the effects of concentration x to viscosity y and this model is given by

$$y = \theta_1 x + \theta_2 x^{\frac{1}{2}} + \theta_3 x^2 + \epsilon, \ x \in S = (0.0, 0.2]. \tag{3.16}$$

Let $z = x^{1/2}$, then @ref(exm:frac_poly) becomes a polynomial model of order 4 with the new design space $S' = (0.0, \sqrt{0.2}]$. Now, we can apply the result for the polynomial model (**??**) and conclude that n_A and n_D are at most 4 or 5. Moreover, notice that there are only three parameters in model (**??**). In order to have an invertible matrix $G_2(\xi, \theta_o)$ which is a 3×3 matrix, we need at least 3 support points. Thus, the number of support point for both A- and D-optimal design is at least 3. In summary, the number of support points in A- and D-optimal designs are either 3, 4, or 5. The purpose of this example is for illustrating that FPM can be transformed into a polynomial model so we can use the result for the number of support points in polynomial models.

Michaelis-Menton model is one of the best-known models proposed by Leonor Michaelis and Maud Menten (**?**) that studies the enzyme reactions between the enzyme and the substrate concentration. This model is given by

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \epsilon_i. \ i = 1, 2, \ldots, n, \ x_i \in (0, k_o], \ \theta_1, \theta_2 \geq 0. \tag{3.17}$$

Here, $\theta = (\theta_1, \theta_2)^\top$. Define $z = \frac{1}{\theta_2 + x}$. Then $f(x, \theta) = (\frac{x}{\theta_2 + x}, \frac{-\theta_1 x}{(\theta_2 + x)^2})^T = (1 - z\theta_2, -z\theta_1 + \theta_1 \theta_2 z^2)^\top$, and

$$M(x, \theta_o, t) = \begin{pmatrix} 1 & \sqrt{t}(1 - z\theta_2) & \sqrt{t}(-\theta_1 z(1 - z\theta_2)) \\ \sqrt{t}(1 - z\theta_2) & (1 - z\theta_2)^2 & -z\theta_1(1 - z\theta)^2 \\ \sqrt{t}(-\theta_1 z(1 - z\theta_2)) & -z\theta_1(1 - z\theta)^2 & (z\theta_1(1 - z\theta_2))^2 \end{pmatrix}. \tag{3.18}$$

We can clearly see that, after the transformation, $z \in S' = [\frac{1}{\theta_1 + k_o}, \frac{1}{\theta_1}]$, and since $f(x, \theta_o)$ and $M(x, \theta, t)$ are polynomial functions of z, ϕ_D and ϕ_A can be described as polynomial functions with highest degrees 4. Now, similar to the discussion for model (**??**), we can find the results for n_A and n_D for model @ref(eq:model-mm} and they are presented in the following Lemma.

Lemma 3.5. *For Michaelis-Menton model in* (**??**), *the number of support points for D-optimal design and A-optimal design are either* 2 *or* 3 *for* $t \in [0, 1)$. *Moreover, there are 2 support points under D-optimality when* $t = 0$.

The proof of Lemma **??** is similar to that of Theorem **??** and is omitted. It is easy to observe that $d_D(0, \xi_D^*, 0) < 0$, so the boundary point, 0, is not a support point under D-optimality which leads to the conclusion $n_D = 2$. Moreover, $n_c \leq 3$ for $t \in [0, 1)$.

Peleg model is a statistical model used to investigate the relationship of water absorption for various kinds of food. This model is given by

$$y_i = y_o + \frac{x_i}{\theta_1 + \theta_2 x_i} + \epsilon_i, \quad i = 1, 2, \cdots, n, \tag{3.19}$$

where y_o represents the initial moisture of the food, y_i is the current level of moisture at current time x_i, θ_1 is the Peleg's moisture rate constant, and θ_2 is the asymptotic moisture as time increases. Note that parameters θ_1 and θ_2 must be positive. Optimal experimental designs have been studied for this model using OLSE, for example, **?**. For (**??**), $f(x, \theta) = (\frac{-x}{(\theta_1 + \theta_2 x_i)^2}, \frac{-x^2}{(\theta_1 + \theta_2 x)^2})^T$. We let $z = \frac{1}{\theta_1 + \theta_2 x}$. Then for $S = [0, d]$, the new design space after the transformation is $S' = [\frac{1}{\theta_1 + \theta_2 d}, \frac{1}{\theta_1}]$. Now, $f(x, \theta) = (\frac{-z(1 - \theta_1 z)}{\theta_2}, \frac{-(1 - \theta_1 z)^2}{\theta_2^2})^T$. Since θ_1 and θ_2 are parameters, f depends on z only. Hence, $M(x, \theta, t)$ becomes a polynomial function of z of degree 4. Therefore, the number of support points, n_A and n_D are also either 2 or 3 for the Peleg model. In addition, n_c is either 1, 2 or 3 for $t \in [0, 1)$.

Next we consider trigonometric models without intercept. If the intercept term is included, it has been proven that optimal designs under SLSE and OLSE are the same. The kth order trigonometric model is given by

$$y_i = \sum_{j=1}^{k} [\cos(jx_i)\theta_{1j} + \sin(jx_i)\theta_{2j}] + \epsilon_i, \ i = 1, 2, \cdots, n, \tag{3.20}$$

where $x_i \in S_b = [-b\pi, b\pi]$, $0 < b \leq 1$. Let $\theta = (\theta_{11}, \theta_{12}, \cdots, \theta_{1k}, \theta_{21}, \cdots, \theta_{2k})^T$. From (**??**), we get

$$M(x, \theta_o, t) = \begin{pmatrix} 1 & \sqrt{t}\cos(x) & \dots & \sqrt{t}\sin(kx) \\ \sqrt{t}\cos(x) & \cos^2(x) & \dots & \cos(x)\sin(kx) \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t}\sin(kx) & \cos(x)\sin(kx) & \dots & \sin^2(kx) \end{pmatrix}_{(2k+1) \times (2k+1)}. \tag{3.21}$$

We consider two cases of design space, (i) $b = 1$, full circle and (ii) $0 < b < 1$, the partial circle.

For case (i), the following Lemma is helpful for finding the number of support points for (**??**).

Lemma 3.6. *For trigonometric functions, $j, u = 1, 2, \cdots, k$, $u \neq j$, we have*

1. $\int_{-\pi}^{\pi} \cos(jx)dx = \int_{-\pi}^{\pi} \sin(jx)dx = 0$,

2. $\int_{-\pi}^{\pi} cos^2(jx)dx = \int_{-\pi}^{\pi} sin^2(jx)dx = \pi$,

3. $\int_{-\pi}^{\pi} cos(jx)cos(ux)dx = \int_{-\pi}^{\pi} sin(jx)sin(ux)dx = 0$,

4. $\int_{-\pi}^{\pi} sin(jx)cos(ux)dx = \int_{-\pi}^{\pi} cos(jx)sin(ux)dx = 0$. *This also holds for* $u = j$.

Theorem 3.3. *For model* (**??**) *with design space* $S = [-\pi, \pi]$, *the uniform distribution on* S *is both A- and D-optimal designs.*

Proof. For the uniform distribution ξ^*, we have $B(\xi^*, \theta_o, t) = 1 \oplus \frac{1}{2} I_{2k}$ by Lemma **??**. From (**??**), we obtain

$$tr(M(x, \theta_o, t)B^{-1}(\xi^*, \theta_o, t)) - (q+1)$$
$$= 1 + 2cos^2(x) + 2cos^2(2x) + \cdots + 2sin^2(kx) - (2k+1)$$
$$= 1 + 2[cos^2(x) + sin^2(x)] + \cdots + 2[cos^2(kx) + sin^2(kx)] - (2k+1)$$
$$= 1 + 2k - (2k+1)$$
$$= 0, \quad \text{for all } x \in S.$$

This implies that $d_D(x, \xi^*, t) = 0$ for all $x \in S$. By Theorem **??**, ξ^* is a D-optimal design.

For A-optimality, it is easy to get

$$tr(M(x, \theta_o, t)B^{-1}(\xi^*, \theta_o, t)C^T CB^{-1}(\xi^*, \theta_o, t)) - tr(CB^{-1}(\xi^*, \theta_o, t))$$
$$= 4cos^2(x) + 4cos^2(2x) + \cdots + 4sin^2(kx) - 4k$$
$$= 4[cos^2(x) + sin^2(x)] + \cdots + 4[cos^2(kx) + sin^2(kx)] - 4k$$
$$= 4k - 4k$$
$$= 0, \quad \text{for all } x \in S,$$

which gives $d_A(x, \xi^*, t) = 0$ for all $x \in S$. Thus, by Theorem **??**, ξ^* is an A-optimal design. $\qquad \square$

For case (ii), $0 < b < 1$, the partial circle $S = [-b\pi, +b\pi]$, let $z = cos(x)$. Now, instead of using x directly, we study the number of support point of x through z in $S' = [cos(b\pi), 1]$. Note that cosine is an even function, so each point of $z \in S'$ corresponds to two symmetric points around 0, $\pm x \in S$. **?** discussed that all the elements in $M(x, \theta_o, t)$ in (**??**) can be written as polynomial functions of z. Hence, functions $d_A(x, \xi, t)$, $d_D(x, \xi, t)$ and $d_c(x, \xi, t)$ are all polynomial functions of z with degree $2k$. Using the similar arguments for the polynomial models, A-, D- and c-optimal designs, in terms of $z \in S'$, cannot exceed $k + 1$ support points.

3.4   Scale invariance property of D-optimal design

D-optimality is one of the most used design criteria due to its many advantages. One good property of this criterion is that it is invariant under some scaling of the independent variables using OLSE (**?**). Recently there are some discussions about the invariance properties including scale invariance and shift invariance for the optimal designs under the SLSE. Examples can be found in **?** and **?**. In this section, we focus on finding a new property of scale invariance of D-optimal designs under the SLSE for nonlinear regression models.

For linear regression models, D-optimal designs are often scale invariant. On the other hand, if the model is nonlinear, the scale invariance property is no longer available. The optimal designs for nonlinear models are called locally optimal, since they depend on the true parameter vector θ_o. Thus, D-optimal designs have to be constructed for each θ_o and for each S. **?** proposed a generalized scale invariance (GSI) concept for studying scale invariance property of D-optimal designs for nonlinear models and generalized linear models. The GSI property is also useful for studying D-optimal designs under the SLSE such that the designs may be constructed on a scaled design space S^V instead of the original design space S.

Denote the D-optimal design for a given model, with true parameter vector θ_o on a design space S by $\xi_D^*(S, \theta_o)$. Also, denote the scaled design space from the original design space S by $S^V = \{Vx | x \in S\}$.

Definition 3.1. The matrix V is called scale matrix which is a diagonal matrix defined as $V = diag(v_1, v_2, \cdots, v_p)$, where all v_i are positive.

Definition 3.2. Transforming x to Vx is called scale transformation.

Definition 3.3. $\xi_D^*(S, \theta_o)$ is said to be scale invariant for a model if there exists a parameter vector $\tilde{\theta}_o$ such that the D-optimal design $\xi_D^*(S^V, \tilde{\theta}_o)$ can be obtained from $\xi_D^*(S, \theta_o)$ using some scale transformations.

This property of the design, $\xi_D^*(S, \theta_o)$, is defined as a generalized scale invariance. Note that $\tilde{\theta}_o$ and θ_o are closed related through the elements $v_1, \cdots, v_p$ in the scale matrix V. When the model is linear, GSI of $\xi_D^*(S, \theta_o)$ becomes the traditional scale invariance since it does not depend on θ_o. For many nonlinear regression models, the GSI property holds for D-optimal designs under both OLSE and SLSE. The following lemma provides a condition to check for this property.

Lemma 3.7. *For a given model and a scale matrix V, if there exists a parameter vector $\tilde{\theta}_o$ and an invertible diagonal matrix K which does not depend on x, such that $f(Vx, \tilde{\theta}_o) = Kf(x, \theta_o)$ for all $x \in S$, then the design $\xi_D^*(S, \theta_o)$ has the generalized scale invariance property. \end{lemma}*

Proof. For design space S and parameter vector θ_o, $\xi_D^*(S, \theta_o)$ minimizes $\phi_D^*(\xi, \theta_o) = \det(B^{-1}(\xi, \theta_o))$ where $B(\xi, \theta_o) = \mathbb{E}[M(x, \theta_o)]$ and $M(x, \theta_o)$ is given by (**??**), and the expectation is taken with respect to $\xi(x)$ on S.

For design space $S^V = \{Vx | x \in S\}$ and parameter vector $\tilde{\theta}_o$, we minimize the following function to find $\xi_D^*(S^V, \tilde{\theta}_o)$, $\phi_D^*(\xi, \tilde{\theta}_o) = \det(B^{-1}(\xi, \tilde{\theta}_o))$, where $B(\xi, \tilde{\theta}_o) = \mathbb{E}[M(z, \tilde{\theta}_o)]$, and the expectation is taken with respect to $\xi(z)$ on S^V. Each $z \in S^V$ can be written as Vx followed by definition, where $x \in S$. Thus, we have $M(z, \tilde{\theta}_o) = M(Vx, \tilde{\theta}_o)$.

By the assumption of Lemma **??** and (**??**), we get

$$M(Vx, \tilde{\theta}_o) = (1 \oplus K)M(x, \theta_o)(1 \oplus K), \ \ for \ all \ x \in S.$$

Therefore, $B(\xi, \tilde{\theta}_o) = (1 \oplus K)B(\xi, \theta_o)(1 \oplus K)$ and $\phi_D(\xi, \tilde{\theta}_o) = \frac{\phi_D(\xi, \theta_o)}{\det^2(K)}$. This implies that we can minimize $\phi_D(\xi, \theta_o)$ to obtain $\xi_D^*(S^V, \tilde{\theta}_o)$, where $\xi_D^*(S^V, \tilde{\theta}_o)$ is the scale transformation from $\xi_D^*(S, \theta_o)$. $\qquad\square$

Example 3.2. Consider Peleg regression model in (**??**). In this model, since $p = 1$, the scale matrix $V = v_1 > 0$ is a scalar and $\theta_o = (a, b)^\top$. Now, let $\tilde{\theta}_o = (a, \frac{b}{v_1})^\top$ and we can obtain

$$
\begin{aligned}
f(Vx, \tilde{\theta}_o) &= f(v_1 x, \tilde{\theta}_o) \\
&= \left(\frac{-v_1 x}{(a + \frac{b}{v_1} v_1 x)^2}, \quad \frac{-(v_1 x)^2}{(a + \frac{b}{v_1} v_1 x)^2} \right)^\top \\
&= \left(v_1 \frac{-x}{(a + bx)^2}, \quad v_1^2 \frac{-x^2}{(a + bx)^2} \right)^\top \\
&= \begin{pmatrix} v_1 & 0 \\ 0 & v_1^2 \end{pmatrix} \left(\frac{-x}{(a + bx)^2}, \quad \frac{-(v_1 x)^2}{(a + bx)^2} \right)^\top \\
&= K f(x, \theta_o).
\end{aligned}
$$

Hence, Peleg model has GSI property based on Lemma **??** by choosing $K = diag(v_1, v_1^2)$ and $\tilde{\theta}_o = (a, \frac{b}{v_1})^T$. $\qquad\square$

It is worth noting that when matrix $B(\xi, \theta_o, t)$ is ill-conditioned (i.e. the condition number of the matrix is very large), numerical algorithm computing $B^{-1}(\xi, \theta_o, t)$ may fail as the numerical inverse is inaccurate and imprecise. In these situations, the GSI property may be helpful. Here we provide one example to demonstrate the usefulness of the GSI property.

Example 3.3. Piecewise polynomial regression using knots is frequently used and has various applications. See **?** and the references therein. They investigated optimal designs under OLSE for piecewise polynomial regression model with unknown knots, and obtained results for the number of support points

under D-optimality and various other properties about designs. Here we consider one model they used, a cubic spline regression model with unknown knots, which is given by

$$y = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3 + \theta_5 (x - \lambda)_+^3 + \epsilon, \ x \in [0, b] \qquad (3.22)$$

where $(x - \lambda)_+ = \max(0, (x - \lambda))$. Model (**??**) is nonlinear with parameter vector $\theta_o = (\theta_1, \cdots, \theta_5, \lambda)^T$, $f(x, \theta_o) = (1, x, x^2, x^3, (x - \lambda)_+^3, -3\theta_5 (x - \lambda)_+^2)^T$. We now start to illustrate GSI property for this model. Consider a scale matrix $V = v_1 > 0$ and let $\tilde{\theta}_o = (\theta_1, \cdots, \theta_5, v_1 \lambda)^T$. Then we have, for all $x \in [0, b]$,

$$
\begin{aligned}
f(Vx, \tilde{\theta}_o) &= f(v_1 x, \tilde{\theta}_o) \\
&= (1, v_1 x, (v_1 x)^2, (v_1 x)^3, (v_1 x - v_1 \lambda)_+^3, -3\theta_5 (v_1 x - v_1 \lambda)_+^2)^T \\
&= (1, v_1, x, (v_1 x)^2, (v_1 x)^3, v_1^3 (x - \lambda)_+^3, -3\theta_5 v_1^2 (x - \lambda)_+^2)^T \\
&= diag(1, v_1, v_1^2, v_1^3, v_1^3, v_1^2)(1, x, x^2, x^3, (x - \lambda)_+^3, -3\theta_5 (x - \lambda)_+^2)^T \\
&= K f(x, \theta_o), \quad \text{with } K = diag(1, v_1, v_1^2, v_1^3, v_1^3, v_1^2).
\end{aligned}
$$

Hence, by Lemma **??**, the D-optimal design under SLSE has the GSI property. Moreover, the model is linear in $\theta'_o = (\theta_1, \theta_2, \cdots, \theta_5)^\top$, so the D-optimal designs under both OLSE and SLSE do not depend on θ'_o **?**. Thus, for $S = [0, b]$ and $S^V = [0, v_1 b]$, the D-optimal designs $\xi^*_D(S, \theta_o)$ and $\xi^*_D(S^V, \tilde{\theta}_o)$ can be obtained from each other by using the corresponding θ_o and $\tilde{\theta}$. This property can help us dramatically for finding the D-optimal designs when matrix $B(\xi, \theta_o, t)$ is ill-conditioned. Numerical results of D-optimal designs are presented in Example **??** after we discuss about numerical algorithms in Chapter **??**. $\square$

We also want to note that in model (**??**), the intercept is included so the D-optimal designs under both OLSE and SLSE are the same (**?**). Moreover, the result may easily be extended for piecewise regression models with multiple unknown knots.

Chapter 4

Numerical Algorithm and Applications

We have studied the optimality conditions and the number of support points for optimal designs under SLSE in Chapter **??**. However, it is still challenging to find optimal designs analytically, except for simple regression models. Often numerical algorithms are used to compute optimal designs, and various algorithms have been studied including **?**, **?**, **?**, and **?**. For optimal designs under the SLSE, **?** used multiplicative algorithms for finding the optimal designs whereas **?** used convex programming and semi-definite programming (SDP) in MATLAB. Here we want to further extend the computational strategies for finding optimal designs under the SLSE. In this chapter, we discuss an effective algorithm to compute optimal designs based on CVX programming in MATLAB.

4.1 Optimization problem

We first discretize the design space $S \subset \mathbb{R}^p$, and the discretized space is denoted as $S_N \subset S$, where N is the number of points in S_N. The magnitude of N can be very large, but often the designs are highly efficient already with moderate N. More discussions about the choice of N is given in the examples in this chapter The discretization can be done in many different ways, but we discretize the design space using equally spaced gird points for the sake of simplicity. For an one dimensional design space, say $S = [a, b]$, the design points are $a + (i - 1)\frac{(b-a)}{N}, i = 1, 2, ..., N$. For higher dimensional design spaces, we use equally spaced grid points for each variable, and S_N is formed by Cartesian product.

Suppose ξ^* is the optimal probability measure in Ξ_N, where Ξ_N includes all the distributions on S_N. We denote $S_N = \{u_1, u_2, ..., u_N\} \subset S$. Any probability

measure $\xi \in \Xi_N$ can be express as

$$\xi = \begin{bmatrix} u_1 & u_2 & ... & u_N \\ w_1 & w_2 & ... & w_N \end{bmatrix},$$

where w_i is the weight at u_i, which is non-negative and $\sum_{i=1}^{N} w_i = 1$. The optimal regression design problem can be expressed as

$$\min_{w_1,...,w_N,\delta} \quad \delta$$

$$\text{s.t.} \qquad \sum_{i=1}^{N} w_i = 1, \qquad (4.1)$$

$$- w_i \leq 0, \text{ for } i = 1, 2, ..., N,$$

$$\phi \leq \delta,$$

where ϕ is a loss function, which can be ϕ_A, $\log(\phi_D)$ or ϕ_c defined in (**??**). Notice that for any $\xi \in \Xi_N$, $B(\xi, \theta_o, t) = \mathbb{E}_\xi[M(x, \theta_o, t)] = \sum_{i=1}^{N} M(u_i, \theta_o, t)w_i$, where $M(u_i, \theta_o, t)$ is defined in (**??**). We also note that a design point is selected if its weight is greater than a very small number δ_1 (a small positive number, say 10^{-5}) in practice. Lastly, we report the $d_D^{\max} = \max\{d_D | x \in S\}$ ($d_A^{\max}$, $d_c^{\max}$) for D-optimality (A-, c-optimality), where d_D, d_A and d_c are defined in (**??**). If $d_D^{\max} < \delta_2$ (a small positive number, say 10^{-4}), then the numerical solution to the above optimization problem is a D-optimal design by Theorem **??**. Similarly we can check for A- and c-optimal designs. Problem (**??**) is a convex optimization problem. We will discuss an algorithm to solve (**??**) in the next section.

4.2   Computational Algorithm

There are some numerical algorithms which have been studied under the SLSE. **?** applied multiplicative algorithms for finding the optimal designs to binary design points. **?** computed D-optimal designs using the moments of distribution $\xi \in \Xi$. **?** formulated the design problems as convex optimization problem and applied convex optimization programs CVX and SeDuMi in MATLAB to find D- and A-optimal designs, respectively. In their work, the A-optimal design problems are solved by SeDuMi which is hard to write the code in MATLAB. In our newly proposed formulation in (**??**), we are able to use CVX to find A-optimal designs, which is much easier to code. Moreover, the CVX programming can be also applied to find c-optimal designs, which the previous algorithms were not applied before.

Figure **??** illustrates a simple algorithm with six steps to apply CVX for finding optimal designs. Users input the variables, partition the design space into N grid points, and compute the M matrix in the first three steps. Steps (i), (ii) and (iii) depend on the regression model and the design space. Step (iv) solves the optimal design problem by CVX program, which depends on the

optimality criterion. The details of using CVX and MATLAB code can be found in Appendix **??** and Appendix **??**. We present several examples using this algorithm to show optimal designs in Section **??**.

4.3 Applications

We compute A-, D- and c-optimal designs for various regression models. Notice that when the intercept term presents, the resulting optimal designs under the OLSE and the SLSE are the same **?**. Since there are many studies in optimal regression designs under the OLSE in the literature already, we only consider the models without the intercept. All the examples are computed via CVX package in MATLAB. We provide the results with different values of t, where t is related to a measure of skewness of the error distribution. The related MATLAB code can be found in Appendix **?**.

All the examples in this thesis are computed on a PC with Intel Core I5 6500 4 cores CPU, 3.20 GHz. The MATLAB version is R2018a academic student version with CVX version 2.1 and build number 1123. The RAM and the platform for this particular machine are 16 Gigabyte and Windows 10 Professional version, respectively.

Example 4.1. This example is for Gompertz growth model which was briefly described in Chapter **??** to showcase the output. Here, we use this model to show the effects of N on the D-optimal designs under SLSE with $\theta_o = (1, 1, 1)^\top$, $S = [0.0, 10.0]$ and $t = 0$.

Table **??** gives D-optimal designs for various values of N. Note that the D-optimal design in Figure **??** in Chapter **??** is for $N = 2001$. We can see that the optimal design changes when N increases. However, when $N \geq 1001$, the optimal designs do not change much and converge to a distribution having three support points with equal weights. We also show the change of the loss function as N changes in Figure **??**. As $N \to +\infty$, the loss function converges too. In this example, moderate N already gives a highly efficient design as shown in Figure **??**. Similar results are obtained for A- and c-optimal designs, and representative results are given in Table **??**, where $c_1 = (2, 0.5, 1)^T$ is used for c-optimal designs, and $t = 0.0, 0.3, 0.5, 0.7$ and 0.9.

It is clear that $d_A^{\max} = \max\limits_{x}(d_A(x, \xi_A^*, t)$ (or $d_c(x, \xi_c^*, t))$ is small, which satisfies the optimality conditions in Theorem **??**. We should also point out that for some situations when two numerical support points are very close to each other, they are probably representing only one theoretical support point which is between the two numerical support points. For instance, when $t = 0.0$, the two support points 1.315 and 1.320 are very close to each other in A-optimal design. If we discretize the design space to more points (i.e. choose larger N), we can find this particular point. For instance, the A-optimal design with $N = 9001$ is