Locally Optimal Experimental Designs for Mixed Responses Models

by

Hazar Khogeer

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved March 2020 by the
Graduate Supervisory Committee:

Ming-Hung Kao, Chair
John Stufken
Mark Reiser
Yi Zheng
Dan Cheng

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

Bivariate responses that comprise mixtures of binary and continuous variables are common in medical, engineering, and other scientific fields. There exist many works concerning the analysis of such mixed data. However, the research on optimal designs for this type of experiments is still scarce. The joint mixed responses model that is considered here involves a mixture of ordinary linear models for the continuous response and a generalized linear model for the binary response. Using the complete class approach, tighter upper bounds on the number of support points required for finding locally optimal designs are derived for the mixed responses models studied in this work.

In the first part of this dissertation, a theoretical result was developed to facilitate the search of locally symmetric optimal designs for mixed responses models with one continuous covariate. Then, the study was extended to mixed responses models that include group effects. Two types of mixed responses models with group effects were investigated. The first type includes models having no common parameters across subject group, and the second type of models allows some common parameters (e.g., a common slope) across groups. In addition to complete class results, an efficient algorithm (PSO-FM) was proposed to search for the A- and D-optimal designs. Finally, the first-order mixed responses model is extended to a type of a quadratic mixed responses model with a quadratic polynomial predictor placed in its linear model.

INDEX WORDS: complete class, generalized linear models, logistic models, locally optimal design, equivalence theorem.

Dedicated

to my beloved husband Dr. Mohammed J. Alawi

whose patience, support, and encouragement have helped me pursue my doctoral degree.

ACKNOWLEDGMENTS

First and foremost, I would like to heartily dedicate my success to God almighty for his protection, grace, and guidance. I would like to express my profound and special thanks and gratitude to my academic adviser, Dr. Ming-Hung (Jason) Kao, for his invaluable guidance throughout my research. I am highly obliged to take this opportunity to thank his true motivation, vision, and patience. I want to include a special note to Dr. Mark Reiser for inspiring me to start this program and life-changing voyage. I could not be more appreciative of my committee members, Dr. John Stufken, Dr. Yi Zheng, and Dr. Dan Cheng, for their generous attitude and support. I would like to extend my extreme gratefulness to my beloved kids for their sacrifices and understanding. Their presence is the sole reason behind the perusal of a doctoral degree. They continuously conveyed bravery and a spirit of adventure into me. My completion of this dissertation would not have been successfully accomplished without my loving husband and partner, Dr. Mohammed J. Alawi. I want to acknowledge his encouragement, enthusiasm, and transparency through this extraordinary milestone. I would also like to express my deep, sincere gratitude to my parents Rabha Shaibi and Dr. Abdulrahman Khogeer and brother Dr. Feras Khogeer, whom their prayers and emotional supports have strengthened me throughout this journey. I also thank my dear friend and classmate, Amani Alrumayh, for her support and positive criticism through our educational expedition. Also, I am greatly indebted to the School of Mathematical and Statistical Sciences at Arizona State University that became my home for more than a decade. I could not complete this acknowledgment without mentioning my home country, The Kingdom of Saudi Arabia, for their financial support. No valuable words could be enough to express the support I received and the persons behind this success.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

In most experiments, statistical design and analysis of data are important tools for experimenters to study the connection between the output (response variables) and the input (explanatory variables) in order to draw conclusions and make recommendations. There is much research on the design and analysis of experiments, where the response(s) collected from each individual is (are) of the same data type (continuous or categorical). However, not much has been done in the design of experiments where the response variables are of the mixed (continuous and categorical) data type, although there is a great demand for this research. Mixed responses models attract more and more researchers as they have been increasingly seen in fields such as the pharmaceutical industry, chemical, and engineering.

The research of developmental toxicity is among the many experiments that have mixed responses (see also Aerts, Molenberghs, Geys, & Ryan, 2002). As an example, the study of the effect of dose (ethylene glycol) on fetal weight and malformation of pregnant mice is observed in laboratory animal experiments. The two outcomes of interest include a continuous response like fetal weight and a binary response such as fetal death or malformation. Mixed responses are also seen in dose-finding trials, where the main interests are efficacy and toxicity of the drug. As observed in the phase II lung cancer study, the two endpoints measured are the efficacy and toxicity, where efficacy is frequently considered as the continuous response (shrinkage in tumor size) while the toxicity is often the binary response (low/high adverse events). The major difficulty in modeling mixed responses arises from the absence of a natural multivariate (bivariate) distribution (Fedorov, Wu, & Zhang, 2012).

A simple technique for modeling mixed responses with the assumption that the responses are independent is to separately model and analyze each response variable. However, this approach is unsatisfactory since the assumption is often violated (Fedorov et al., 2012; Teixeira-Pinto & Normand, 2009). As a result, two common likelihood-based techniques are popular among researchers to deal with the dilemma of analyzing mixed responses. The first technique is inserting a continuous unobserved (latent) variable to model the association within several mixed responses. The initial attempt to use the latent variable in modeling the discrete variables goes back to Pearson (1900) (Rabe-Hesketh & Skrondal, 2004).

Sammel, Ryan, and Legler (1997) proposed a model for multiple responses where the covariates effects x_i are included in the model through the unobservable latent variable b_i by using a linear regression model $b_i = \lambda x_i + \delta_i$. Here, λ represents the association between the covariates and the latent variable, and δ_i is an error term. Then the responses y_i are modeled by conditioning on the unobservable latent variables b_i. Each response is assumed to be independent and can be any member of the one-parameter exponential family model. For $n = 1, \cdots, N$ responses, $f_n(y_{in}|b_i) = \exp[(y_{in}\theta_n - d(\theta_n))/\phi_n + c_n(y_{in}, \phi_n)]$, where $\theta_n = \alpha_n + \beta_n b_i$ and α, β represent the association between the responses and the latent variable.

On the other hand, Arminger and Küsters (1988) proposed an alternative latent variable model in which the effects of the covariates are included directly into the model rather than through the latent variables. However, the sets of latent variables are considered jointly normally distributed. Dunson (2000) generalized the model introduced by Arminger and Küsters (1988) by allowing the latent variables and the responses to follow any member of the one-parameter exponential family model. However, the downside of this model is that it had several non-estimable parameters that are related to the variance of the responses.

The second common likelihood-based technique is by factorizing of the joint distribution of mixed responses in a rather straight-forward way (Catalano & Ryan, 1992; Cox, 1972; Deng & Jin, 2015; Fitzmaurice & Laird, 1995). Cox (1972) was the first to consider modeling the joint distribution for the discrete and continuous responses using the direct factorization technique where the joint distribution of bivariate responses is computed by the product of the normal distribution for the marginal continuous response and the logistic conditional distribution for the binary response given the value of the continuous response.

Catalano and Ryan (1992) used the concept of the factorization approach with the latent variable to model the bivariate mixed outcome. Their research was motivated by the development of toxicity in laboratory animals, in which they studied the association between fetal weight and malformations. In their study, the fetal weight is a continuous variable y, and the malformation is a binary response z. The assumption was that the binary response had some corresponding unobserved continuous latent variable b_i. They modeled the joint distribution as the product of the continuous variable and the latent variable condition on the continuous variable, i.e., $f(y, z) = f(y)f(z_{b_i}|y)$ that follows a bivariate normal distribution. They used the probit link to connect the marginal distribution of the binary response to covariate variables and used the linear link function to connect the marginal distribution of the continuous response to covariate variables. The model accounted for the correlation between the continuous variable and the latent variable and was extended to accommodate for any clustering in the data. According to Fitzmaurice and Laird (1995), the models proposed by Cox (1972) and Catalano and Ryan (1992) have some downsides: the regression parameters in both models have no marginal interpretation, and the efficiency of estimating the parameters depends on the correct specifications of the association between the responses.

By only using direct factorization techniques for modeling bivariate binary and continuous responses, Fitzmaurice and Laird (1995) proposed a model for analyzing data from developmental toxicity studies. The joint distribution of bivariate responses is computed by the product of the marginal distribution of the binary response and the conditional distribution of the continuous response given the value of the binary response. Fitzmaurice and Laird (1997) extended their method to include multivariate responses of mixed types with missing data.

More recently, in manufacturing systems where multiple endpoints of the mixed data type are obtained, Deng and Jin (2015) considered a model termed as the "QQ model". They were interested in studying the wafer lapping process that involves a continuous response, the total thickness variation (TTV), and a binary response, the sit total indicator readings (STIR). When modeling the bivariate responses, Deng and Jin (2015) also considered the previously mentioned factorization idea. Their model formulation tends to be more flexible than that in the previous studies, and their focus was on the association between the responses. This is in contrast to Fitzmaurice and Laird (1995), in which the main concern was on modeling the marginal expectation of the continuous response. As claimed by Deng and Jin (2015), their proposed model allows for accurate predictions of both binary and continuous responses, which are informative in the manufacturing system. Many models on the analysis of mixed responses were proposed by different researchers. The book of De Leon and Chough (2013) contains an aggregation of models on mixed responses analysis.

Another important question that often emerges is how to design the experiment so that the study objectives can be effectively achieved with minimum cost. Our focus here is on finding an optimal design for bivariate mixed responses models. Several researchers have addressed the design problems for mixed responses models of certain types. For example, Fedorov et al. (2012) searched for D-optimal designs for mixed responses with one continuous efficacy response and one binary response for toxicity in clinical trials. The modeling technique that was considered involves a latent variable for determining the value of the binary response, and then the latent variable and the continuous variable assumed to have a joint normal distribution. The first order exchange algorithm was applied for optimal design selection.

Biswas and López-Fidalgo (2013) worked on a similar dose-finding design problem in which the bivariate response variables can possibly be of different types (quantitative/qualitative). To reach their goal of maximizing efficacy with no or low toxicity, they defined a criterion function that additively combined two optimality criteria the D-optimality criterion and a specialized continuous criterion function for active controlled dose-finding experiment. The direct factorization approach was used when the bivariate responses include a continuous variable and a binary variable. The model that they considered contains two sub-models: a logistic model for the binary toxicity response and a linear model for the continuous efficacy response given the value of the binary response. By using the first-order algorithm, they completed their search for the optimal design.

More recently, Kim and Kao (2019) investigated designs for a mixed responses model with a single covariate and identified a complete class of designs having two to four support points with two of them being the endpoints of the design region. The complete class approach will be explained in the next chapter. They then searched for the A- and D-optimality designs for some cases and confirmed the optimality of the obtained designs by the general equivalence theorem. The model that Kim and Kao (2019) had studied involved a first-order polynomial predictor and a form of quadratic polynomial predictor. However, the research in optimal experimental design for mixed responses models still needs additional investigation with the increasingly sophisticated models and data analysis procedures.

In practice, when researchers adopt mixed responses models, the responses may depend on one or more groups. Finding optimal designs by including group effects in the model is important in scientific studies. To our knowledge, there is currently no systematic study on this design issue. One of our study objectives is to build upon existing works to provide additional useful results in this research direction.

In this study, we are concerned with optimal experimental designs for mixed responses models. Our first focus is on a model that has the same form as the model advocated by Deng and Jin (2015). This model has also been studied in Kim and Kao (2019). Partly inspired by their research results, we work in this research direction. We provide some relevant background knowledge about the complete class approach in Chapter 2. In Chapter 3, we derive some additional results by considering the same model as Kim and Kao (2019), which we will refer to as the simple mixed responses model in this work. Some design issues and results for mixed responses models with group effects are then explored in Chapter 4. In Chapter 5, the strategy used to obtain and verify the results for the mixed responses models with group effects are discussed. Then we proposed in Chapter 6 an extension on the mixed responses model with one covariate by inserting a quadratic polynomial predictor into the model. Finally, in Chapter 7, summary and discussion of future work for this study are provided.

Chapter 2

LITERATURE REVIEW

2.1 Optimal Design Approach

Research on optimal experimental designs has been receiving much interest and continues to develop over the past several decades. The optimal designs are a type of designs that are optimal with respect to some criterion, which for example, allows the most precise estimate of the model parameters. The goal of the experimenter is to gain as much information as possible from the experiments with a given budget to allow researchers to obtain high-quality statistical analysis results. In this case, given the model and a particular optimality criterion, the researchers will need an optimal design approach that guides them to the best number of combination levels of the covariates, known as design points and the corresponding number of replicates.

The initial recorded work on the concept of optimal design might go back to Smith (1918) who obtained optimum designs for a series of single-factor polynomial models of order up to six (Atkinson, Donev, & Tobias, 2007). The criterion that was proposed by Smith (1918), which minimizes the maximum variance of any predictor value over the design region, was later referred to as G-optimality by Kiefer and Wolfowitz (1959). Some fundamental optimal design theories were investigated and evolved by Jack Kiefer in the 1950's through the 1970's (Fedorov & Leonov, 2013). Since then, there has been much development in this important line of research. To provide the relevant background on this research, let us consider N independent responses, $z_1, \cdots, z_N$, and a linear model:

$$z_i = \mathbf{f}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\theta} + e_i, \tag{2.1}$$

where $\mathbf{x}_i$ is the vector of q predictor variables of the i^{th} experimental unit, $\mathbf{f}(\mathbf{x}_i)$ is a vector of known functions of $\mathbf{x}_i$ with $\mathbf{f}(\mathbf{x}_i) = (f_1(\mathbf{x}_i), \cdots, f_p(\mathbf{x}_i))^{\mathrm{T}}$, $\boldsymbol{\theta}$ is a vector of p unknown parameters, and e_i is the error. A common assumption for linear model is that the e_i's are iid normally distributed with mean 0 and variance σ^2. With model (2.1), the (Fisher) information matrix for $\boldsymbol{\theta}$ is $M(\boldsymbol{\theta}) = \dfrac{\mathbf{F}^{\mathbf{T}}\mathbf{F}}{\sigma^2} = (1/\sigma^2)\sum\limits_{i=1}^{N} \mathbf{f}(\mathbf{x}_i)\mathbf{f}(\mathbf{x}_i)^{\mathrm{T}}$, where $\mathbf{F}$ is an $N \times p$ design matrix that contains the $1 \times p$ vectors $\mathbf{f}^{\mathbf{T}}(\mathbf{x}_i)$ as its rows. Under certain assumptions, the inverse of the information matrix is the variance-covariance matrix of the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, which is $\Sigma(\hat{\boldsymbol{\theta}})= \sigma^2(\mathbf{F}^{\mathbf{T}}\mathbf{F})^{-1}$. To estimate the model parameters as precise as possible, we select a design that provides the minimum variance of the parameter estimator. Such an optimal design for a given model is often achieved by maximizing the information matrix (or minimizing the variance-covariance matrix). A design ξ for model (2.1) can be expressed as

$$\xi = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ n_1 & n_2 & \cdots & n_m \end{pmatrix},$$

where $\mathbf{x}_i$ is a design point indicating the values of the q predictor variables, m is the total number of distinct design points, n_i is the number of replicates of $\mathbf{x}_i$, and $i = 1, 2, \cdots, m$. Here, the total number of observations is $N = \sum\limits_{i=1}^{m} n_i$. The design space χ, which contains all the possible values of $\mathbf{x}_i$, may be an interval or hyper rectangle and is normally determined by the experimenter. With n_i being an integer, such a design is often referred to as an exact design. Some major challenges of finding an optimal exact design come from the discreteness of n_i and the dependence of the optimal designs on the total number of observations N. Because of these and other challenges, many research works focus heavily on the continuous design, which is also known as the approximate design to be described below.

In order to avoid the mathematical complexity of discrete optimization, Kiefer (1959) introduced a continuous design theory which became a common design approach. A continuous design can be illustrated as a probability measure with finite support points that are defined on a given compact design space χ. In particular, a continuous (approximate) design is normally denoted as

$$\xi = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ w_1 & w_2 & \cdots & w_m \end{pmatrix}, \text{ or } \xi = \{(\mathbf{x}_i, w_i), i = 1, ..., m\}.$$

Here, $w_i = \dfrac{n_i}{N}$ is called the weight and is the proportion of the total number of observations that is assigned to the corresponding distinct design point $\mathbf{x}_i$; $w_i \geq 0$, and $\sum_{i=1}^{m} w_i = 1$. Only when w_i is greater than zero, $\mathbf{x}_i$ is known as a support point, and the size of the design, ξ, is determined by the number of support points in it. When studying continuous designs, we drop the constraint that Nw_i has to be an integer and allow it to be any real value between 0 and N. However in practice, the number of trials of any design point must be an integer. When Nw_i is not an integer, a rounding procedure may be applied to obtain the number of replicates for $\mathbf{x}_i$ with the constraint that Nw_i's sum to N (Pukelsheim & Rieder, 1992). For a large N, the rounded continuous design (exact design) and the continuous design ought to be close (Berger & Wong, 2009). Note that, the total number of observations, N, will not affect the search for an optimal continuous design.

The theory of optimal design is mainly built upon the information matrix that, in some sense, summarizes the amount of information for the parameters of interest. The normalized information matrix for a continuous design with m support points corresponding to the linear model in (2.1), can be defined as $M(\xi) = (1/\sigma^2) \sum_{i=1}^{m} w_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}(\mathbf{x}_i)^{\mathrm{T}}$. We denote the inverse of a nonsingular information matrix for $\boldsymbol{\theta}$ by $\Sigma(\xi, \boldsymbol{\theta})$, which corresponds to the variance-covariance matrix of the MLE of $\boldsymbol{\theta}$ under the normality assumption. We aim at an optimal design that possesses the 'smallest' variance-covariance matrix or the 'largest' information matrix. Since matrix ranking is often not practical, the selection of the optimal design is commonly based on an optimality criterion. The optimal criterion is a real-valued function of the information matrix that, in some sense, summarizes how good a design is. The optimal design may vary with the selected optimality criterion. In other words, a design that is optimal for one criterion may not be optimal for another criterion (Atkinson et al., 2007). Several criteria coincide with particular statistical concepts, so the selection of the criterion is often made depending on the objectives of the experiment. For example, if the objective is to estimate all the parameters of the model, then the D-optimal criterion, Φ_D, may be considered. On the other hand, if the objective is to estimate a specific linear function of the model parameters then the c-optimal criterion is suitable. When the design factors are qualitative and of the same scale as in optimal block designs, the A-optimal criterion, Φ_A, is considered (Atkinson et al., 2007). We now define the most commonly used optimality criteria, known as "alphabet optimal criteria."

D-optimality : this criterion is one of the most popular optimality criteria. It is the determinant of the variance-covariance matrix of the parameter-estimator $|\Sigma(\xi, \boldsymbol{\theta})|$, or $\log|\Sigma(\xi, \boldsymbol{\theta})|$, which is also known as (the log of) the generalized variance of the parameter estimates. Since the volume of the (asymptotic) ellipsoidal confidence regions of the parameter vector $\boldsymbol{\theta}$ is proportional to the square root of the determinant of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, the criterion helps to identify designs that give 'small' joint confidence ellipsoid for $\boldsymbol{\theta}$.

A-optimality : this criterion is the trace of the variance-covariance matrix for the parameter-estimator, i.e. tr $(\Sigma(\xi, \boldsymbol{\theta}))$. The A-optimal criterion aims to minimize the average asymptotic variance of the estimates of the parameters $\boldsymbol{\theta}$.

c-optimality : this criterion focuses on estimating some desirable functions, i.e., $c^{\mathrm{T}}\boldsymbol{\theta}$ of the model parameters. The criterion minimizes $c^{\mathrm{T}}\Sigma(\xi, \boldsymbol{\theta})c$, where c is a $p \times 1$ vector.

G-optimality : this criterion focuses on minimizing the maximum of the standardized variance $d(x, \xi)$ over the design space. The standardized variance related to the prediction model is defined as $d(x, \xi) = \dfrac{N\mathrm{var}(\hat{y}(x))}{\sigma^2}$, where $\mathrm{var}(\hat{y}(x)) = \sigma^2 f^{\mathrm{T}}(x)\Sigma(\xi, \boldsymbol{\theta})f(x)$. Therefore, the criterion seeks to minimize the worst prediction variance over the design space.

E-optimality : this criterion focuses on minimizing the maximum eigenvalue ,λ_i, of the variance-covariance matrix, $\lambda_{max}(\Sigma(\xi, \boldsymbol{\theta}))$. The criterion helps to identify designs that minimize the length of the major axis of the confidence ellipsoid for $\boldsymbol{\theta}$. This criterion is a spacial case of the G-optimality when $c^{\mathrm{T}}c$ is a unit sphere, thus it will minimize the worst variance of linear combinations of the parameter estimates $\mathrm{var}(c^{\mathrm{T}}\hat{\boldsymbol{\theta}}) = c^T\Sigma(\xi, \boldsymbol{\theta})c$, i.e., min max$\dfrac{1}{\lambda} = \dfrac{c^{\mathrm{T}}\Sigma(\xi, \boldsymbol{\theta})c}{c^{\mathrm{T}}c}$ is reached when $c^{\mathrm{T}}c = 1$.

Kiefer (1974) proposed a quite general criterion named as the Φ_p-optimality criterion that includes the popular D-, A-, and E-optimality criteria as special cases, and is defined as

$$\Phi_p\{\Sigma(\xi, \boldsymbol{\theta})\} = \left[\frac{1}{v}tr\{(\Sigma(\xi, \boldsymbol{\theta}))^p\}\right]^{1/p} \qquad p \in (0, \infty)$$

where v is the dimension of $\Sigma(\xi, \boldsymbol{\theta})$. When $p \to 0$ the criterion reduces to the D-optimality criterion, and when $p \to \infty$ we have the E-optimality criterion. We get the A-optimality criterion when $p = 1$. The optimality criteria mentioned previously (or their equivalent forms) are convex functions of the information matrix. With optimality criteria that are convex, differentiable, and a compact design space χ, the general equivalence theorem can be applied to verify the optimality of a given continuous design (Kiefer, 1974; Kiefer & Wolfowitz, 1960). The general equivalence theorem will be explained in Section 2.4.

2.2 Design for Generalized Linear Models

Another type of models that is useful when the linear model assumptions are violated is the Generalized Linear Models (GLM). This type of models includes the classical linear models as special cases. It also contains models that are suitable for cases where the response variable does not have a normal distribution, but some other distributions such as Poisson or binomial distributions that belong to the exponential family for which the mean and variance might be highly related. McCullagh and Nelder (1989) and Dobson and Barnett (2008) provided a historical summary with examples of the progress and evolution of GLMs. The construction of experimental design for a GLM presents a level of complexity. The main reason is that the information matrix normally contains unknown model parameters. As a result, experiments having the same model but with different model parameter values will normally require different optimal designs. To see this, let us look at the following GLM:

$$\mathsf{g}[E(z_i)] = \mathsf{g}(\mu_i) = \eta_i, \tag{2.2}$$

where η_i is the linear predictor defined as $\eta_i = \mathbf{f}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\theta}$, μ_i is the mean of z_i, and $\mathsf{g}(\cdot)$ is the link function which relates the mean of the response variable to the linear predictor. The normalized information matrix for $\boldsymbol{\theta}$ is $M(\xi, \boldsymbol{\theta}) = \sum_{i=1}^{m} w_i \Gamma(\mathbf{f}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\theta})\mathbf{f}(\mathbf{x}_i)\mathbf{f}(\mathbf{x}_i)^{\mathrm{T}}$, where $\Gamma(\cdot)$ is called the GLM weight defined as $\Gamma(\eta_i) = \frac{1}{V(\mu_i)}\left(\frac{d\mu_i}{d\eta_i}\right)^2$, the variance of the response is $\mathrm{var}(z_i) = \phi V(\mu_i)$, and ϕ is a dispersion parameter. Note that $\Gamma(\eta_i)$ depends on the model parameters through μ_i, and so does $M(\xi, \boldsymbol{\theta})$. Table (2.1) lists some examples of commonly used GLMs. Obtaining optimal designs for GLMs normally requires some prior information regarding the possible values of the unknown parameters. Nevertheless, the theory of optimal design for linear models, including the general equivalence theorem, can be extended to GLMs (and other nonlinear models) by fixing the values of the parameters that the information matrix depends on (Pukelsheim, 2006; Stufken & Yang, 2012b).

Table 2.1: Some Commonly Used GLMs

Model	Linear	Logistic	Probit	Poisson
Response	continous	Binary	Binary	Count
Distribution	Normal	Binomial	Binomial	Poisson
Link-Function	Identity : $\eta = \mu$	Logit : $\eta = \log\left(\frac{\mu}{1-\mu}\right)$	Probit : $\eta = \Phi^{-1}(\mu)$	Log : $\eta = \log(\mu)$
Γ-Function	$\Gamma(\eta) = \frac{1}{V(\mu)}$	$\Gamma(\eta) = \frac{\exp(\eta)}{(1+\exp(\eta))^2}$	$\Gamma(\eta) = \frac{\Phi'(\eta)^2}{\Phi(\eta)(1-\Phi(\eta))}$	$\Gamma(\eta) = \exp(\eta)$

One of the widespread approaches to tackle the design issues for GLMs (or nonlinear models) was introduced by Chernoff (1953), who suggested obtaining the “locally optimal” designs based on the best guess of the unknown model parameters. In principle, a locally optimal design is optimal in some sense when the guessed parameters value turns out to be true, but can be sub-optimal when the guessed value is far from the true parameter value. In the absence of any prior knowledge on the unknown parameters, the locally optimal design is still useful since it can be considered as a benchmark to evaluate the performance of other designs. A reasonable initial guess on the unknown parameters can be obtained from previous experiments conducted on the same objective. However, if there is no knowledge of the possible parameter value, one possible solution is to use the multistage approach. For example, an arbitrary design may be considered at the first stage to gain some knowledge about the unknown parameters. In the next steps, a locally optimal design is constructed based on the information gained from the previous step for estimating the unknown parameters. Other common procedures include the Bayesian design approach that places a prior distribution on the unknown parameters (see also Khuri, Mukherjee, Sinha, & Ghosh, 2006).

For obtaining locally optimal designs, a powerful algebraic procedure was introduced by Yang and Stufken (2009). This algebraic procedure allows us to find a "complete class" of designs for a variety of models under many commonly used optimality criteria.

2.3 Complete Class Approach

The complete class approach allows us to restrict our search for optimal designs to a small subclass, called complete class, of candidate designs. This subclass of designs, denoted as Ξ, is constructed so that, for any given design, there is at least one design in the subclass that is at least as efficient.

The idea for using a complete class had started some time ago (e.g., Ehrenfeld et al., 1956; Kiefer & Wolfowitz, 1959), and the research on this direction has been developed by several authors such as Pukelsheim (1989) who identified it as "essentially complete class" under what he called the Kiefer ordering. There are several recent contributions that are related to the development of the complete class approach, such as the work by Mathew and Sinha (2001) on the two parameters logistic regression model. They demonstrate that a locally D-optimal design for estimating the two model parameters of a simple logistic regression model can be found in the class of two-point symmetric designs. A symmetric design is a design whose design points are symmetric around 0 and the paired design points, x and $-x$, have the same weights; note that $x = -x = 0$ can also be included in a symmetric design. However, their numerical results for the A-optimality criterion revealed a different conclusion; the design points of a locally A-optimal design are symmetric but the corresponding weights of the symmetric design points may not be the same. The work of Mathew and Sinha (2001) leads to the seminal work by Yang and Stufken (2009) who proposed an algebraic approach to identify a complete class of locally optimal designs for nonlinear models with two parameters. Their complete class suggests that the search for locally optimal designs for the previously mentioned setting can be done within a subclass of designs with at most two support points. Specifically, for any given design ξ, we can find a design ξ^* within a subclass of designs with at most two support points that is not inferior to the design ξ such that $M(\xi^*) \geq M(\xi)$ under the Loewner ordering. Note that the Loewner ordering is a type of ordering among (information) matrices. We say that an information matrix M^* (or the corresponding designs ξ^*) is no worse than another M (or the corresponding designs ξ) in Loewner ordering when $M^*-\ M$ is nonnegative definite; we denote this as $M^* \geq$ M. Most popular optimality criteria, like Φ_p-optimality criterion, obey the Loewner ordering, so if $M^* \geq M$ then $\Phi_p(\mathrm{M}^*) \leq \Phi_p(M)$. The construction of complete class was extended to include nonlinear models with more than two parameters by Yang (2010b) who derived sufficient conditions for the De la Garza phenomenon to be applied to nonlinear models. The de la Garza phenomenon is due to De la Garza et al. (1954), which suggests that for a polynomial regression model with a degree p, an optimal design can be found within the class of designs having $p + 1$ support points. Dette and Melas (2011) then gave a more general result than Yang (2010b) by including a broader class of nonlinear regression models. They utilized the Chebyshev systems, see also Karlin and Studden (1966), for the study of the De la Garza phenomenon. Yang and Stufken (2012) further extended the work to give a general and powerful approach that allows researchers to easily identify complete classes for an even wider class of models. For many of these models, the complete class achieved by this latter approach can be much smaller than those identified by the previous methods. Since then, this became an important research field that caught major attention.

The complete class approach of Yang and Stufken (2012) can help determine a sharp upper bound on the number of support points for (locally) optimal designs. This upper bound is normally tighter than the classical Caratheodory's bound which restricted the maximum number of support points of the optimal design to $p(p+1)/2+1$, where p is the number of parameters of interest. To explain the approach of Yang and Stufken (2012), a bijection, which will be determined by the model, is used to relate the induced design point $c_i \in [A, B]$ to the design point $x_i \in [L, U]$. This helps to rewrite the information matrix of a given nonlinear model as:

$$M(\xi, \boldsymbol{\theta}) = B(\boldsymbol{\theta})(\sum w_i C(\boldsymbol{\theta}, c_i))B^{\mathrm{T}}(\boldsymbol{\theta}). \tag{2.3}$$

By doing so, we can restrict our attention to the matrix $C(\boldsymbol{\theta}, c_i)$ instead of the entire information matrix $M(\xi, \boldsymbol{\theta})$ since $B(\boldsymbol{\theta})$ is a non-singular matrix that depends only on the parameters, and it is not affected by the selected design. In many cases, $C(\boldsymbol{\theta}, c_i)$ is a matrix whose elements are smooth functions of c_i defined on $[A, B]$. We denote the (i, j)th element of $C(\boldsymbol{\theta}, c_i)$ as $\Gamma_{ij}(c_i)$. With this decomposition, we may focus on finding a design $\xi = \{(c_i, w_i), i = 1, \cdots, m\}$ that 'maximizes' $M(\xi, \boldsymbol{\theta})$ for a given value of $\boldsymbol{\theta}$, and then rewrite the design in term of x_i. The strategy of Yang and Stufken (2012) is performed by first partitioning the $n \times n$ symmetric matrix $C(\boldsymbol{\theta}, c_i)$ by a chosen permutation matrix P as follows:

$$PC(\boldsymbol{\theta}, c_j)P^T = \begin{pmatrix} C_{11}(\boldsymbol{\theta}, c_i) & C_{12}(\boldsymbol{\theta}, c_i) \\ C_{12}^T(\boldsymbol{\theta}, c_i) & C_{22}(\boldsymbol{\theta}, c_i) \end{pmatrix}, \tag{2.4}$$

where $C_{22}(\boldsymbol{\theta}, c_i)$ is a symmetric principal sub-matrix with dimension $n_1 \times n_1$; $1 \leq n_1 < n$. It can be seen that, for two designs $\xi^* = \{(c_i^*, w_i^*), i = 1, \cdots, m^*\}$, and $\xi = \{(c_i, w_i), i = 1, \cdots, m\}$, we have $M(\xi^*, \boldsymbol{\theta}) \geq M(\xi, \boldsymbol{\theta})$, whenever $\sum_{i=1}^{m^*} w_i^* C(\boldsymbol{\theta}, c_i^*) \geq$

$\sum_{i=1}^{m} w_i C(\boldsymbol{\theta}, c_i)$. The latter is satisfied if the following hold:

$$\sum_{i=1}^{m^*} w_i^* C_{11}(\boldsymbol{\theta}, c_i^*) = \sum_{i=1}^{m} w_i C_{11}(\theta, c_i),$$
$$\sum_{i=1}^{m^*} w_i^* C_{12}(\boldsymbol{\theta}, c_i^*) = \sum_{i=1}^{m} w_i C_{12}(\boldsymbol{\theta}, c_i), \text{ and}$$
$$\sum_{i=1}^{m^*} w_i^* C_{22}(\boldsymbol{\theta}, c_i^*) \geq \sum_{i=1}^{m} w_i C_{22}(\boldsymbol{\theta}, c_i).$$

This means that the weighted sum of the two matrices $C(\boldsymbol{\theta}, c_i)$ and $C(\boldsymbol{\theta}, c_i^*)$ are almost the same, except for the elements in the principle sub-matrix C_{22} that creates the information inequality. With this fact, the procedure of Yang and Stufken (2012) suggests to first find a maximal set of linearly independent non-constant Γ_{ij}'s from the first $n - n_1$ rows of the $PC(\boldsymbol{\theta}, c_i)P^T$ matrix. Denote these Γ-functions by $\Psi_1, \cdots, \Psi_{k-1}$. Set $\Psi_0 = 1$, and $\Psi_k^q = q^T C_{22} q$ for a nonzero vector q. The next step is to check if the sets $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}\}$ and $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}, \Psi_k^q\}$ or $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}\}$ and $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}, -\Psi_k^q\}$ form Chebyshev systems for all $q \neq 0$. The Chebyshev systems (Dette & Melas, 2011; Karlin & Studden, 1966) used in this step can be explained as follows. Suppose we have a set of $k+1$ continuous functions $\Psi_0, \cdots, \Psi_k : [A, B] \to \mathbb{R}$. We say that $\{\Psi_0, \cdots, \Psi_k\}$ form a Chebyshev system if for all $A \leq z_0 < z_1 < \cdots < z_k \leq B$ the inequality

$$\begin{vmatrix} \Psi_0(z_0) & \Psi_0(z_1) & \cdots & \Psi_0(z_k) \\ \Psi_1(z_0) & \Psi_1(z_1) & \cdots & \Psi_1(z_k) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_k(z_0) & \Psi_k(z_1) & \cdots & \Psi_k(z_k) \end{vmatrix} > 0 \tag{2.5}$$

is satisfied. A rather simple way to identify a Chebyshev system other than using the definition directly is to follow the strategy found in Proposition 4 of Yang and Stufken (2012). Assume that all the Ψ functions are at least k times differentiable on

(A, B). Let $f_{l,t}(c), 1 \leq t \leq l \leq k$ be defined as in the following matrix:

$$\begin{pmatrix} f_{1,1} = \Psi_1'(c) & & & & & \\ f_{2,1} = \Psi_2'(c) & f_{2,2} = \left(\frac{f_{2,1}}{f_{1,1}}\right)' & & & & \\ f_{3,1} = \Psi_3'(c) & f_{3,2} = \left(\frac{f_{3,1}}{f_{1,1}}\right)' & f_{3,3} = \left(\frac{f_{3,2}}{f_{2,2}}\right)' & & & \\ f_{4,1} = \Psi_4'(c) & f_{4,2} = \left(\frac{f_{4,1}}{f_{1,1}}\right)' & f_{4,3} = \left(\frac{f_{4,2}}{f_{2,2}}\right)' & f_{4,4} = \left(\frac{f_{4,3}}{f_{3,3}}\right)' & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ f_{k,1} = \Psi_k' & f_{k,2} = \left(\frac{f_{k,1}}{f_{1,1}}\right)' & f_{k,3} = \left(\frac{f_{k,2}}{f_{2,2}}\right)' & f_{k,4} = \left(\frac{f_{k,3}}{f_{3,3}}\right)' & \cdots & f_{k,k} = \left(\frac{f_{k,k-1}}{f_{k-1,k-1}}\right)' \end{pmatrix}. \quad (2.6)$$

If $f_{l,l}(c) > 0 \ \forall c \in [A, B]$ and $l = 1, \cdots, k-1$, then the set $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}, \Psi_k^q\}$ forms a Chebyshev system if $f_{k,k}(c) > 0$, whereas the set $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}, -\Psi_k^q\}$ forms a Chebyshev system if $-f_{k,k}(c) > 0$. Following this, we can conclude that the sets $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}\}$ and $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}, \Psi_k^q\}$ form Chebyshev systems if $F(c) > 0$, where as the sets $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}\}$ and $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}, -\Psi_k^q\}$ form Chebyshev systems if $-F(c) > 0$, where $F(c)$ is defined as:

$$F(c) = \prod_{l=1}^{k} f_{l,l}(c) \quad \forall c \in [A, B]. \quad (2.7)$$

If the desired Chebyshev system can be formed, then based on Lemma 2 of Yang and Stufken (2012) which is restated again in Lemma 2.3.1, for any given set $S = \{(c_i, w_i) : w_i > 0, A \leq c_i \leq B, i = 1, \cdots, N\}$ with a sufficiently large N (to be specified in the Lemma), there exists a dominant set $S^* = \{(c_i^*, w_i^*) : w_i^* > 0, A \leq c_i^* \leq B, i = 1, \cdots, n^*\}$ such that

$$\begin{aligned} \sum_{i=1}^{n^*} w_i^* &= \sum_{i=1}^{N} w_i \\ \sum_{i=1}^{n^*} w_i^* \Psi_l(c_i^*) &= \sum_{i=1}^{N} w_i \Psi_l(c_i), \qquad l = 1, \cdots, k-1; \\ \sum_{i=1}^{n^*} w_i^* \Psi_k^q(c_i^*) &> \sum_{i=1}^{N} w_i \Psi_k^q(c_i) \qquad \text{for every nonzero vector q.} \end{aligned} \quad (2.8)$$

From (2.8), it is clear that we may change the sign of some Ψ_l for $l = 1, \cdots, k-1$ without changing the equality. The complete class results is thus invariant to a sign change of these Ψ functions. We now state the Lemma 2 of Yang and Stufken (2012) below.

Lemma 2.3.1 (Lemma 2 in Yang and Stufken (2012)). *Consider a given set* $S = \{(c_i, w_i) : w_i > 0, A \leq c_i \leq B, i = 1, \cdots, N\}$, *and let the previously described* Ψ_i $i = 0, \cdots, k-1$, *and* Ψ_k^q *be such that either*

$$\{\Psi_0, \Psi_1, \cdots, \Psi_{1-k}\} \text{ and } \{\Psi_0, \Psi_1, \cdots, \Psi_{1-k}, \Psi_k^q\} \tag{2.9}$$

or

$$\{\Psi_0, \Psi_1, \cdots, \Psi_{1-k}\} \text{ and } \{\Psi_0, \Psi_1, \cdots, \Psi_{1-k}, -\Psi_k^q\} \tag{2.10}$$

form Chebyshev systems on the interval $[A, B]$ *for any nonzero vector* q. *Let* $n^* = \lceil k/2 \rceil$ *be the smallest integer* $\geq k/2$. *Then the following results hold:*

a. Assume k *is odd and* $N \geq n^*$. *If* (2.9) *holds, then there exists a set* S^*, *which has* n^* *points including* B, *such that* S *is dominated by* S^*.

b. Assume k *is odd and* $N \geq n^*$. *If* (2.10) *holds, then there exists a set* S^*, *which has* n^* *points including* A, *such that* S *is dominated by* S^*.

c. Assume k *is even and* $N \geq n^*$. *If* (2.9) *holds, then there exists a set* S^*, *which has* $n^* + 1$ *points including both* A *and* B, *such that* S *is dominated by* S^*.

d. Assume k *is even and* $N \geq n^* + 1$. *If* (2.10) *holds, then there exists a set* S^*, *which has* n^* *points, such that* S *is dominated by* S^*.

Note that S and S^* may or may not be designs depending on whether $\sum w_i = 1$ (and $\sum w_i^* = 1$) or not. In addition, to facilitate the construction of the complete class, Yang and Stufken (2012) proposed the use of $F(c)$ defined in (2.7) to verify if the sets of Ψ-functions in (2.9) and (2.10) form Chebyshev systems. Kim and Kao

(2019) made a necessary change to this complete class approach to identify complete classes of locally optimal designs for the mixed responses models that they considered. This main tool used by Kim and Kao (2019) is important, and is restated below with the same notation as previously described. We note that instead of finding a maximal set $\{\Psi_1, \cdots, \Psi_{k-1}\}$ of nonconstant linearly independent elements from $[C_{11}(\theta, c), C_{12}(\theta, c)]$ in (2.4), Kim and Kao (2019) required $\{\Psi_0 = 1, \Psi_1, \cdots, \Psi_{k-1}\}$ to be a maximal set of linearly independent Ψ-functions. Following their work, we developed new complete class results for mixed responses under more complicated settings.

Lemma 2.3.2. *Let $\Psi_1, \cdots,$ and Ψ_{k-1} be nonconstant functions selected from $[C_{11}, C_{12}]$ such that $\{\Psi_0, \Psi_1, \cdots, \Psi_{k-1}\}$ is a maximal set of linearly independent functions and $n^* = \lceil k/2 \rceil$. Then the following results hold:*

(a) If k is odd and $F(c) > 0$ for all $c \in [A, B]$, then the designs containing at most n^* support points including B form a complete class.

(b) If k is odd and $F(c) < 0$ for all $c \in [A, B]$, then the designs containing at most n^* support points including A form a complete class.

(c) If k is even and $F(c) > 0$ for all $c \in [A, B]$, then the designs containing at most $n^* + 1$ support points including both A and B form a complete class.

(d) If k is even and $F(c) < 0$ for all $c \in [A, B]$, then the designs containing at most n^* support points form a complete class.

2.4 General Equivalence Theorem

As mentioned before, the general equivalence theorem (GET) is a powerful tool that can be used to verify the optimality of the selected design. It is one of the traditional approaches that is widely used in the literature to establish optimal designs and to verify their optimality.

Let ξ_x be a design that contains a one design point x with design weight equaling one. The general equivalence theorem for a given criterion function $\Phi\{M(\xi)\}$ states that the following three conditions are equivalent (see also Atkinson et al., 2007):

1. The design ξ^* minimizes $\Phi\{M(\xi)\}$.

2. $\phi(x, \xi^*) \geq 0 \ \forall \ x \in \chi$, and the minimum is achieved at the support points of the design ξ^*.

3. The design ξ^* maximizes $\min\limits_{x\in\chi} \phi(x, \xi)$.

Here, $\phi(x, \xi)$ is the directional derivative of $\Phi(\xi)$ in the direction of ξ_x, and it is defined as:

$$\phi(x, \xi) = \lim_{\alpha \to 0^+} \frac{1}{\alpha}[\Phi\{(1-\alpha)M(\xi) + \alpha M(\xi_x)\} - \Phi\{M(\xi)\}]. \tag{2.11}$$

Interestingly, the general equivalence theorem states that both the D-optimal and G-optimal designs are equivalent under a given model even though the two criteria have different statistical interpretations. While there is a general statement for the equivalence theorem, each optimal criterion has its own form. The condition for optimality criteria are examined as follows:

The D-optimality criterion: the directional derivative of $\Phi_D(M)$ at ξ in the direction of ξ_x is given as:

$$\begin{aligned} \phi(x, \xi) =& - \operatorname{tr}\{M^{-1}(\xi)M(\xi_x) - M^{-1}(\xi)M(\xi)\} \\ =& p - \operatorname{tr}\{M^{-1}(\xi)M(\xi_x)\} = p - d_D(x, \xi), \end{aligned} \tag{2.12}$$

where p is the number of parameters and $d_D(x, \xi)$ is known as the sensitivity function for the corresponding criterion (Fedorov & Leonov, 2013). For the D-optimal designs the function $d_D(x, \xi)$ should achieve its maxima of p at the support points of the design.

The A-optimality criterion: the directional derivative of $\Phi_A(M)$ at ξ in the direction of ξ_x is given as:

$$\begin{aligned}\phi(x,\xi) = -\mathrm{tr}\{M(\xi_x)M^{-2}(\xi) - M^{-1}(\xi)\} =&\mathrm{tr}\{M^{-1}(\xi)\} - \mathrm{tr}\{M(\xi_x)M^{-2}(\xi)\} \\ =& \, q - d_A(x,\xi)\end{aligned} \tag{2.13}$$

where $d_A(x,\xi)$ is the sensitivity function for the A-criterion.

An easy and fast way to validate the optimality of a design is by graphing, whenever possible, the sensitivity function for the obtained ξ for the corresponding criterion. If the maximum values of the graph of the sensitivity function is bounded above by p or q for the D- or A-optimality criterion, respectively, and the tangent points are exactly the obtained support points then the design is optimal.

2.5 The Relative Efficiency and Efficiency Lower Bound

In some cases, the researchers prefer to use other designs rather than the optimal design. For example, when the sample size is small, the rounded optimal continuous design (exact design) applied in practice can not be identical to the continuous optimal design observed. In other words, the optimal design found acts as a benchmark for other designs. In this case, we need a tool to assess how successful the applied design is, in comparison to the optimal design found. The relative efficiency proposed by Yates (1935) is a significant result in which the design of interest ξ is compared to the optimal design ξ^*. If the relative efficiency is less than one, then we can achieve the same estimation precision as the optimal design by multiplying the sample size of the design ξ by the reciprocal of the ratio calculated. If the D- or A-optimality is of interest, then the relative efficiency of the design is as follows:

The D-efficiency:

$$D_{eff} = \left(\frac{|M(\xi)|}{|M(\xi^{*})|} \right)^{1/p}, \text{ where } p \text{ is the number of parameters.} \tag{2.14}$$

The A-efficiency:

$$A_{eff} = \frac{\mathrm{tr}(M(\xi))}{\mathrm{tr}(M(\xi^{*}))}. \tag{2.15}$$

In calculating the relative efficiency of any design, the problem that arises is the absence of the optimal designs. In this case the worth of the design is measured by the efficiency lower bound, ϵ, which can be computed using the equivalence theorem and the relative efficiency.

The D-efficiency of the design ξ can be bounded below by:

$$D_{eff}(\xi) \geq \exp(-\frac{\epsilon}{p}) = D_{lb}(\xi). \tag{2.16}$$

The A-efficiency of the design ξ can be bounded below by:

$$A_{eff}(\xi) \geq (1 - \frac{\epsilon}{\Phi_A(M(\xi))}) = A_{lb}(\xi), \tag{2.17}$$

(see Yang, Biedermann, & Tang, 2013).

Chapter 3

SIMPLE MIXED RESPONSES MODEL

In this chapter, we restrict the search for optimal designs further to the class of symmetric designs. This is especially useful when the D-criterion is considered. We adopted the previously described complete class approach to identify a small class of symmetric designs denoted as Ξ_s, for a given symmetric design space. In other words, for any given continuous symmetric design ξ_s, we can identify a design $\xi_s^* \in \Xi_s$ so that $M(\xi_s, \boldsymbol{\theta}) \leq M(\xi_s^*, \boldsymbol{\theta})$ under Loewner ordering, i.e. $M(\xi_s^*, \boldsymbol{\theta}) - M(\xi_s, \boldsymbol{\theta}) \geq 0$ is non-negative definite. That is, the design ξ_s^* is not inferior to the design ξ_s under the D-optimality criterion. As a result, the search of D-optimal designs will be within Ξ_s, and this can substantially decrease the effort for finding optimal designs.

Many research in optimal experimental designs for nonlinear models resulted in symmetric designs. For example, Mathew and Sinha (2001) considered the logistic model with two parameters. They first confined themselves to the class of two-point symmetric designs when searching for A- and D-optimal designs. They found that the design $\xi_s^* = \{(\pm c^*, 1/2)\}$ with $c^* = 1.5434$ is D-optimal for estimating the two parameters in the logistic model that they considerd. However, their numerical results revealed that, in general, the A-optimal design is not a symmetric design. Similar results were achieved by Yang (2008) who provided a mathematical proof of the numerical findings of Mathew and Sinha (2001). In Chapter 2 of Liski, Mandal, Shah, and Sinha (2002), the optimal designs for a polynomial model of degree one or higher was studied by considering the symmetric design for symmetric domain. They explain that the value of the Φ_p-criterion for symmetric designs is at least as good as any other designs. Thus, in their models the search can be restricted to the class of symmetric designs. Symmetric designs were also studied by Hyun (2013) under a probit model with a quadratic term for dose response relationship in toxicology studies. He identified a four-point symmetric complete class for the probit model. Then he searched for the A- and D-optimality designs and confirmed the optimality of the obtained designs by the general equivalence theorem. Wu and Stufken (2014) also eased the search for optimal designs by focusing on symmetric designs when forming a complete class using Φ_p-optimal design for GLM with a single-variable quadratic polynomial predictor. They showed under some assumptions that the optimal design can be established depending on the value of the three parameters within the class of symmetric designs that have either three or four support points. Kim (2017) considered a quadratic mixed responses model, where she extended the simple mixed responses model by inserting a quadratic term in the logistic sub-model. The focus was on finding a complete class for symmetric designs, where she searched for the locally D-optimal designs. Within the same framework, we consider a symmetric complete class for locally optimal designs that can be applied to the simple mixed responses models.

3.1 Statistical Model

Consider an experiment that has an independent variable $x \in \mathbb{R}$, and bivariate response variables; one response variable is binary, $z \in \{0, 1\}$, and the other is continuous, $y \in \mathbb{R}$. We assume that the relationship between z and x can be described by a generalized linear model (GLM), whereas the conditional distribution of y given z and x follows a normal distribution.

The joint distribution of (y, z) can be obtained by $f(y, z) = f(z)f(y|z)$, where $f(\cdot)$ represents a probability density/mass function. With N observations $i = 1, \cdots, N$, the model that we consider can be represented as follows:

$$y_i|z_i \sim N(\boldsymbol{f}_z^{\mathrm{T}}(x_i)\boldsymbol{\theta}_z, \sigma^2), \quad \text{with} \quad \mu_{y|z}(x_i) = \boldsymbol{f}_z^T(x_i)\boldsymbol{\theta}_z, \quad z = 0, 1$$

and (3.1)

$$\mathrm{Prob}(z_i = 1) = \mathrm{P}(\boldsymbol{f}_2^{\mathrm{T}}(x_i)\boldsymbol{\theta}_2) \equiv p(x_i),$$

where $\boldsymbol{f}_r(x)$'s are vectors of known functions of x for $r = 0, 1, 2$, $\mathrm{P}(\cdot)$ is a cumulative distribution function (cdf), and $\{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2\}$ are unknown parameters. Note, we assume that the bivariate responses (y, z) of the same observational unit might be correlated, but are independent across different units. Moreover, the correlation between the response y and z can be calculated by $\mathrm{cor}(y_i, z_i) = d_i / \Big[\frac{\sigma^2}{\{p(x_i)(1-p(x_i))\}} + d_i^2\Big]^{1/2}$, where $d_i = \mu_{y|z=1}(x_i) - \mu_{y|z=0}(x_i)$ (Olkin, Tate, et al., 1961).

The joint distribution of y and z is given as:

$$\begin{aligned}
f(y_i, z_i) =& f(z_i)f(y_i|z_i) \\
=& [p(x_i)]^{z_i}[1-p(x_i)]^{1-z_i}[f(y_i|z_i=1)]^{z_i}[f(y_i|z_i=0)]^{1-z_i}, \\
=& [p(x_i)]^{z_i}\,[1-p(x_i)]^{1-z_i} \times \\
& \left[\frac{1}{\sigma\sqrt{2\pi}}\exp - \frac{(y_i - \mu_{y|z=1}(x_i))^2}{2\sigma^2}\right]^{z_i} \left[\frac{1}{\sigma\sqrt{2\pi}}\exp - \frac{(y_i - \mu_{y|z=0}(x_i))^2}{2\sigma^2}\right]^{1-z_i}.
\end{aligned}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_0^T, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ be the parameter vector of interest, then the log-likelihood function is:

$$\begin{aligned}
\log L(\boldsymbol{\theta}) &= \log\prod_{i=1}^{N} f(y_i, z_i) = \log\prod_{i=1}^{N} f(z_i)f(y_i|z_i) \\
&= \sum_{i=1}^{N}\Big\{z_i\log\big[\,\mathrm{P}(\boldsymbol{f}_2^{\mathrm{T}}(x_i)\boldsymbol{\theta}_2)\big] + (1-z_i)\log\big[1 - \mathrm{P}(\boldsymbol{f}_2^{\mathrm{T}}(x_i)\boldsymbol{\theta}_2)\big] \\
&+ \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - z_i\frac{(y_i - \boldsymbol{f}_1^{\mathrm{T}}(x_i)\boldsymbol{\theta}_1)^2}{2\sigma^2} - (1-z_i)\frac{(y_i - \boldsymbol{f}_0^{\mathrm{T}}(x_i)\boldsymbol{\theta}_0)^2}{2\sigma^2}\Big\}.
\end{aligned}$$

Based on the log-likelihood function, the information matrix can be calculated by $M(\boldsymbol{\theta}) = -E\left(\frac{\partial^2 logL}{\partial^2\boldsymbol{\theta}}\right)$. Then, the information matrix for $\boldsymbol{\theta}$ under a continuous design $\xi = \{(x_i, w_i), i = 1, \cdots, m\}$ is a symmetric block diagonal matrix:

$$M(\xi, \boldsymbol{\theta}) = \bigoplus_{r=0}^{2} \sum_{i=1}^{m} w_i \Gamma_r(\boldsymbol{f}_2^{\mathrm{T}}(x_i)\boldsymbol{\theta}_2)\boldsymbol{f}_r(x_i)\boldsymbol{f}_r^{\mathrm{T}}(x_i) \tag{3.2}$$

where, $\oplus$ is the direct sum operator,

$$\Gamma_0(x) = \frac{1-\mathrm{P}(x)}{\sigma^2}, \Gamma_1(x) = \frac{\mathrm{P}(x)}{\sigma^2}, \text{ and } \Gamma_2(x) = \frac{[\mathrm{P}'(x)]^2}{\mathrm{P}(x)(1-\mathrm{P}(x))}. \tag{3.3}$$

Note that the model can be extended to the case where $\mathrm{var}(y|z)$ also depends on the value of z (e.g., Cox & Wermuth, 1992).

For simplicity, we first follow Kim and Kao (2019) to consider the case where $\boldsymbol{f}_{\mathrm{r}}(x_i) = (1, x_i)^{\mathrm{T}}$ and $\boldsymbol{\theta}_{\mathrm{r}} = (\alpha_{\mathrm{r}}, \beta_{\mathrm{r}})$ for $r = 0, 1, 2$, and z has a logistic distribution; i.e. $\mathrm{P}(z) = \dfrac{e^z}{1+e^z}$. We also consider an induced design $\xi = \{(c_i, w_i), i = 1, \cdots, m\}$, where $c_i = \alpha_2 + \beta_2 x_i$ is defined through a bijection of x_i. The information matrix of $\boldsymbol{\theta}$ can be re-written, for some nonsingular matrix $B(\boldsymbol{\theta})$ and non-nonegative definite matrix $C(\boldsymbol{\theta}, c_i)$, in the following form:

$$M(\xi, \boldsymbol{\theta}) = B(\boldsymbol{\theta})\tilde{C}(\xi, \boldsymbol{\theta})B^{\mathrm{T}}(\boldsymbol{\theta}), \tag{3.4}$$

where $\tilde{C}(\xi, \boldsymbol{\theta}) = \sum\limits_{i=1}^{m} w_i C(\boldsymbol{\theta}, c_i)$. Specifically, $B(\boldsymbol{\theta}) = \mathrm{diag}(\dfrac{1}{\sigma}B_1, \dfrac{1}{\sigma}B_1, B_1)$, where $B_1 = \begin{pmatrix} 1 & 0 \\ \alpha_2 & \beta_2 \end{pmatrix}^{-1}$. The $C(\boldsymbol{\theta}, c_i)$ matrix is a 6-by-6 symmetric matrix defined as:

$$\begin{pmatrix}
\Gamma_{11} = \frac{1}{1+e^{c_i}} & \Gamma_{12} = c_i \frac{1}{1+e^{c_i}} & 0 & 0 & 0 & 0 \\
\Gamma_{12} = c_i \frac{1}{1+e^{c_i}} & \Gamma_{22} = {c_i}^2 \frac{1}{1+e^{c_i}} & 0 & 0 & 0 & 0 \\
0 & 0 & \Gamma_{33} = \frac{e^{c_i}}{1+e^{c_i}} & \Gamma_{34} = c_i \frac{e^{c_i}}{1+e^{c_i}} & 0 & 0 \\
0 & 0 & \Gamma_{34} = c_i \frac{e^{c_i}}{1+e^{c_i}} & \Gamma_{44} = c_i^2 \frac{e^{c_i}}{1+e^{c_i}} & 0 & 0 \\
0 & 0 & 0 & 0 & \Gamma_{55} = \frac{e^{c_i}}{(1+e^{c_i})^2} & \Gamma_{56} = c_i \frac{e^{c_i}}{(1+e^{c_i})^2} \\
0 & 0 & 0 & 0 & \Gamma_{56} = c_i \frac{e^{c_i}}{(1+e^{c_i})^2} & \Gamma_{66} = c_i^2 \frac{e^{c_i}}{(1+e^{c_i})^2}
\end{pmatrix}$$

$$= \mathrm{diag}(C_0, C_1, C_2), \tag{3.5}$$

where $C_r = \Gamma_r \begin{pmatrix} 1 & c_i \\ c_i & c_i^2 \end{pmatrix}, r = 0, 1, 2$, $\Gamma_0 = \dfrac{1}{1+e^{c_i}}$, $\Gamma_1 = \dfrac{e^{c_i}}{1+e^{c_i}}$, and $\Gamma_2 = \dfrac{e^{c_i}}{(1+e^{c_i})^2}$.

3.2 Symmetric Complete Class Results

Our focus is to identify a small symmetric complete class where we provide an upper bound on the numbers of symmetric design points and weights. By working with the C matrix in (3.5), Kim and Kao (2019) identified a complete class for the simple mixed responses model, which we restate it in the next theorem for later use.

Theorem 3.2.1. *For the simple mixed responses model, for any given design ξ that has at least four support points, there exists a design ξ^* that has at most four support points including both endpoints of the design space which satisfies $\tilde{C}(\xi^*, \boldsymbol{\theta}) \geq \tilde{C}(\xi, \boldsymbol{\theta})$.*

According to their numerical results, it can be conjectured that, when the range of c is symmetric around 0, the locally D-optimal design can be found by searching over the class of symmetric designs of at most 4 support points. Here, the symmetric design is defined as $\xi_s = \{(\pm c_i, w_i/2), c_i \geq 0, w_i > 0, i = 1, \cdots, m\}$, where c_i is a support point, w_i is the corresponding weight for c_i and $-c_i$, and $\sum_{i=1}^{m} \frac{w_i}{2} = 0.5$. Without restricting the number of the support points (m), we have the following simple result whose proof can be found in the Appenedix.

Lemma 3.2.2. *For any design $\xi = \{(c_i, w_i), i = 1, \cdots, m\}$ in the induced design space $[-D, D]$ for some $D > 0$, we define the corresponding symmetric design as $\xi_s = \{(\pm c_i, w_i/2), c_i \geq 0, w_i > 0, i = 1, \cdots, m\}$. Then, ξ_s is at least as good as ξ under D-optimality.*

With this lemma, we may say that the set of all symmetric designs is a complete class under D-optimality. However, this is a large complete class. In what follows, we form a much smaller complete class.

With the same notation as in Lemma 3.2.2, it can be seen that $\tilde{C}(\xi_s, \boldsymbol{\theta}) = \frac{1}{2}[(\tilde{C}(\xi, \boldsymbol{\theta}) + G\tilde{C}(\xi, \boldsymbol{\theta})G^T)] = \frac{1}{2}[(\tilde{C}(\xi, \boldsymbol{\theta}) + \tilde{C}(\xi_r, \boldsymbol{\theta})]$, where ξ_r is a reflected design with $\xi_r = \{(-c_i, w_i), i = 1, \cdots m\}$ generated from $\xi = \{(c_i, w_i), w_i > 0, \sum w_i = 1, i = 1, \cdots m\}$, and G is an orthogonal transformation matrix with $\det(G) = \pm 1$ and GG^T =I. In particular,

$$G = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{3.6}$$

We then have $\tilde{C}(\xi_s, \boldsymbol{\theta}) = \sum_{i=1}^{m} w_i C_s(\boldsymbol{\theta}, c_i)$ where

$$C_s(\boldsymbol{\theta}, c_i) = \begin{pmatrix} \Gamma_{s11}(c_i) & \Gamma_{s12}(c_i) & 0 & 0 & 0 & 0 \\ \Gamma_{s12}(c_i) & \Gamma_{s22}(c_i) & 0 & 0 & 0 & 0 \\ 0 & 0 & \Gamma_{s33}(c_i) & \Gamma_{s34}(c_i) & 0 & 0 \\ 0 & 0 & \Gamma_{s34}(c_i) & \Gamma_{s44}(c_i) & 0 & 0 \\ 0 & 0 & 0 & 0 & \Gamma_{s55}(c_i) & 0 \\ 0 & 0 & 0 & 0 & 0 & \Gamma_{s66}(c_i) \end{pmatrix} \tag{3.7}$$

$$= \begin{pmatrix} \frac{1}{2} & \frac{c_i(1-e^{c_i})}{2(1+e^{c_i})} & 0 & 0 & 0 & 0 \\ \frac{c_i(1-e^{c_i})}{2(1+e^{c_i})} & \frac{c_i^2}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{c_i(e^{c_i}-1)}{2(1+e^{c_i})} & 0 & 0 \\ 0 & 0 & \frac{c_i(e^{c_i}-1)}{2(1+e^{c_i})} & \frac{c_i^2}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{e^{c_i}}{(1+e^{c_i})^2} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_i^2 \frac{e^{c_i}}{(1+e^{c_i})^2} \end{pmatrix}. \tag{3.8}$$

We note that the sum of Γ_{s34} and Γ_{s12} in $C_s(\boldsymbol{\theta}, c_i)$ is: $\Gamma_{s34} + \Gamma_{s12} = \frac{c_i(e^{c_i}-1)}{2(1+e^{c_i})} + \frac{c_i(1-e^{c_i})}{2(1+e^{c_i})} = 0$; these two Γ_{sij} functions are linearly dependent. We thus select either one of them into the set $\{\Psi_1, \cdots, \Psi_{k-1}\}$. To apply the complete class approach described in Chapter 2, we will partition $C_s(\boldsymbol{\theta}, c_i)$, possibly after simultaneously permuting some rows and columns of it, as:

$$\begin{pmatrix} C_{s11}(\boldsymbol{\theta}, c_i) & C_{s12}(\boldsymbol{\theta}, c_i) \\ C_{s12}(\boldsymbol{\theta}, c_i) & C_{s22}(\boldsymbol{\theta}, c_i) \end{pmatrix}. \tag{3.9}$$

The selection of the C_{s22} matrix needed in the procedure is done by permuting the rows and columns of C_s so that Γ_{s44} and Γ_{s66} can be in the lower-right 2-by-2 submatrix. In particular, $C_{s22} = \text{diag}(\Gamma_{s44}, \Gamma_{s66})$. We then have the following result:

Lemma 3.2.3. *Under a simple mixed responses model,* $\{\Psi_0, \Psi_1 = \Gamma_{s34}, \Psi_2 = \Gamma_{s55}\}$ *and* $\{\Psi_0, \Psi_1 = \Gamma_{s34}, \Psi_2 = \Gamma_{s55}, \Psi_3^q\}$ *form Chebyshev systems for any non-zero vector* q *on any closed interval in* $[0, \infty)$*, where* $\Psi_0 = 1$ *and* $\Psi_3^q = q^T \mathit{diag}(\Gamma_{s44}, \Gamma_{s66})q$*.*

Proof. If we take into consideration the set of Ψ functions above and using the definition in (2.6) for $f_{l,l}$'s, then, $f_{1,1} = \dfrac{c + \sinh(\mathrm{c})}{1 + \cosh(\mathrm{c})} > 0\ \forall c > 0$, and by Proposition 1 in the Appendix, we have $f_{2,2} = \dfrac{e^{5c} - 4ce^{3c} - e^c}{(1 + e^c)^2(e^{2c} + 2ce^c - 1)^2} > 0\ \forall c > 0$. We then use the principle minor test to check if the diagonal matrix $f_{3,3}$ is a positive definite matrix. By Proposition 2 in the Appendix, the (1,1) element of $f_{3,3}$ is
$\dfrac{16(\cosh(c) + 1)(\sinh^2(c) - c^2)}{(2c - \sinh(2\mathrm{c}))^2}[2c\cosh(\mathrm{c}) + \cosh(\mathrm{c})\ \sinh(\mathrm{c}) - c - 2\sinh(\mathrm{c})] > 0 \quad \forall c > 0$,
and by Proposition 3 in the Appendix, the (2,2) element of $f_{3,3}$ is
$\dfrac{16\cosh^2(\frac{c}{2})(c + \sinh(c))}{(2c - \sinh(2c))^2}[c^2 - 2\sinh^2(c) + \dfrac{c\sinh(2c)}{2}] > 0 \quad \forall c > 0$. Therefore, $F(c) > 0$ for $c \in (0, \infty)$. Then the conclusion of the Ψ sets forming Chebyshev systems follows directly from the proof of Theorem 2 in Yang and Stufken (2012). $\square$

The following theorem allows us to form a small complete class for symmetric designs.

Theorem 3.2.4. *Under a simple mixed responses model, a complete class for symmetric designs in the induced design space* $[-D, D]$ *can be formed by all the symmetric designs of at most 4 support points including* $-D$ *and* D*.*

Proof. Lemma 3.2.3 and Lemma 2.3.1 assert that for any set $S^+ = \{(c_i, w_i/2) : 0 \le c_i \le D, w_i > 0, i = 1, \cdots, N, \sum_{i=1}^{N} w_i/2 = 0.5\}$, where $N \ge 2$, there exists a dominating set $S^{+*} = \{(c^*, w^*/2), (D, 0.5 - w^*/2), 0 \le c_i \le D, w^* > 0\}$ so that $\sum_{i=1}^{2} \frac{w_i^*}{2} \Psi_t(c_i^*) = \sum_{i=1}^{N} \frac{w_i}{2} \Psi_t(c_i)$, $t = 0, 1, 2$, and $\sum_{i=1}^{2} \frac{w_i^*}{2} \Psi_3^q(c_i^*) > \sum_{i=1}^{N} \frac{w_i}{2} \Psi_3^q(c_i)$ for every non-zero vector q. For $c < 0$, we have $\Psi_t(-c) = \Psi_t(c)$, $t = 0, 1, 2$, and $\Psi_3^q(-c) = \Psi_3^q(c)$ for every non-zero vector q. Thus, $\sum_{i=1}^{2} \frac{w_i^*}{2} \Psi_t(-c_i^*) = \sum_{i=1}^{N} \frac{w_i}{2} \Psi_t(-c_i)$, $t = 0, 1, 2$, and $\sum_{i=1}^{2} \frac{w_i^*}{2} \Psi_3^q(-c_i^*) > \sum_{i=1}^{N} \frac{w_i}{2} \Psi_3^q(-c_i)$. These in turn give that:
$\sum_{i=1}^{2} w_i^* C_{s11}(\boldsymbol{\theta}, c_i^*) = \sum_{i=1}^{N} w_i C_{s11}(\boldsymbol{\theta}, c_i)$,
$\sum_{i=1}^{2} w_i^* C_{s12}(\boldsymbol{\theta}, c_i^*) = \sum_{i=1}^{N} w_i C_{s12}(\boldsymbol{\theta}, c_i)$, and
$\sum_{i=1}^{2} w_i^* C_{s22}(\boldsymbol{\theta}, c_i^*) > \sum_{i=1}^{N} w_i C_{s22}(\boldsymbol{\theta}, c_i)$.
Furthermore, since the matrix $B(\boldsymbol{\theta})$ in (3.4) does not depend on the design, the conclusion $M(\xi_s^*, \boldsymbol{\theta}) \ge M(\xi_s, \boldsymbol{\theta})$ follows, where $\xi_s = \{(\pm c_i, w_i/2), 0 \le c_i \le D, w_i > 0, \sum_{i=1}^{m} w_i/2 = 0.5, i = 1, \cdots, m\}$ with some $m \ge 2$, and $\xi_s^* = \{(\pm D, 0.5 - w^*/2), (\pm c^*, w^*/2), 0 \le c^* \le D, w^* > 0\}$. □

Based on Theorem 3.2.4, we can obtain a locally D-optimal design for the induced design space $[-D, D]$ by searching among the class of designs of the form: $\{(\pm D, 0.5 - w^*/2), (\pm c^*, w^*/2), 0 \le c^* \le D, w^* > 0\}$.

3.3 Numerical Results

The optimization problem can be attacked in several ways; either analytically by mathematical derivations or numerically by effective computer algorithms. However, in most cases the functionality of analytical approaches can guide the numerical techniques in finding optimal designs in a more efficient way or vice versa. The complete class approach is helpful in narrowing the search of optimal designs to a small class of designs. But, to identify the optimal design under a specific criterion, we still need an efficient optimization algorithm. In literature, there are different types of algorithms that can be considered to find approximate optimal designs reliably and quickly for different study objectives. A brief overview of several algorithms that have been developed to obtain optimal designs is discussed later in Chapter 5. Recently, Kim and Kao (2019) used the fmincon solver in MATLAB to search for D- and A-optimal designs for simple mixed responses model. Thus, we adopt the fmincon solver to search within the complete class for locally optimal designs. The fmincon solver is a constrained nonlinear multivariable function optimizer provided in MATLAB that finds the minimum of a problem specified by linear and nonlinear constraints. There are five different algorithms that fmincon solver can be operated with. These algorithms include the sequential quadratic programming (SQP), SQP-legacy, interior-point algorithm (IPA), trust-region-reflective method (TRRM), and active-set algorithm. Following the recommendation of Kim (2017), both SQP algorithm and IPA can be used due to their comparable speed and accuracy for the simple mixed responses model. Here, the SQP algorithm was primarily used. Our optimization problem for the simple mixed responses model deals with smooth objective function that has some linear inequality constraint in regards to the approximate design setting and design space restriction. In general, constrained nonlinear optimization problems can be formulated as follows:

$$\begin{aligned} &\underset{\xi=\{(c_i,w_i),i=1,\cdots,m\}}{\text{minimize}} && \Phi_p\{\mathbf{M}(\xi)\} \\ &\text{subject to} && \sum w_i = 1, w_i \geq 0, \text{and } D_1 \leq c_i \leq D_2 \qquad (i = 1, ..., m). \end{aligned}$$

We focused on the D-optimality criterion of the form $\Phi_D = \log|M^{-1}(\xi,\boldsymbol{\theta})|$. Since we decomposed the information matrix as in (3.4), then the D-optimality criterion can be written as $\Phi_D = -2\log|B(\boldsymbol{\theta})| - \log|\tilde{C}(\xi,\boldsymbol{\theta})|$ where the search can be done for the induced design points c_i, and the corresponding x_i can be determined through the equation $c_i = \alpha_2+\beta_2 x_i$. Hence, the problem can be reduced to minimize $-\log|\tilde{C}(\xi,\boldsymbol{\theta})|$

since the matrix $B(\boldsymbol{\theta})$ does not depend on c_i. The GET is then used to verify that the obtained design is optimal. Based on the GET for D-optimality in (2.12), the plot of $d_D(x, \xi) = \mathrm{tr}\{\tilde{C}^{-1}(\xi, \boldsymbol{\theta})\tilde{C}(\xi_x, \boldsymbol{\theta})\}$ can be used to verify the optimality of the design over the induced design space.

By using the symmetric complete class results in Theorem 3.2.4, the proposed structure of the symmetric locally D-optimal design for the simple mixed responses model in (3.1) with the induced symmetric design space $[-D, D]$ is given as $\xi_s^*=$ $\{(-D, (1-w^*)/2), (-c^*, w^*/2), (c^*, w^*/2), (D, (1-w^*)/2)\}$. This further simplifies the search, since we only need to find one support point c^* and one weight w^*, which in return decreases the search time for the optimal symmetric design. Thus, the initial design point is set to $\xi^0 = (-c, c, (1-w)/2, w/2, w/2, (1-w)/2)$, where the initial values of c and w are chosen randomly, $0 \leq c < D$, and $0 \leq w \leq 0.5$. We compare the time that the fmincon solver takes to find an optimal design over the 4-point symmetric designs, ξ_s^*, and that over all the 4-point designs.

The optimal design results obtained by fmincon solver were presented in Table 3.1, where the CPU times (in milli-seconds) are displayed. The search was executed using a Dell Desktop that has a 3.4 GMz Intel Core i7 with 32G RAM.

To study the effect of the design space on the number of support points we gradually increased the design space to range from [-1,1] to [-100,100]. As shown in Table 3.1, the number of symmetric support points ranged from two to four support points depending on the size of the symmetric induced design space. For the obtained optimal designs with two support points, the two points are exactly the two end points of the induced design space as for the design space [-0.5,0.5], [-1, 1], and [-1.5,1.5]. For a 3-point design, zero is one of the support points as for the induced design spaces [-2,2], [-2.5,2.5], and [-2.7,2.7]. Here, the weights of the 3-support points are not all equal, and only the weights of the end points are the same. As the induced design

Table 3.1: Locally D-optimal Designs for Symmetric Induced Design Space for c

Design space		Optimized design				Number of points	OF	CPU[a](ms)	CPU[b](ms)
-0.5	0.5	-0.5			0.5	2	9.952	42	22
		0.5			0.5				
-1	1	-1			1	2	6.506	41	21
		0.5			0.5				
-1.5	1.5	-1.5			1.5	2	5.179	41	21
		0.5			0.5				
-2	2	-2	0		2	3	4.779	52	40
		0.431	0.138		0.431				
-2.5	2.5	-2.5	0		2.5	3	4.491	48	35
		0.347	0.306		0.347				
-2.7	2.7	-2.7	0		2.7	3	4.381	52	37
		0.327	0.346		0.327				
-3	3	-3	-0.3016	0.3016	3	4	4.225	50	38
		0.3	0.2	0.2	0.3				
-4	4	-4	-0.923	0.923	4	4	3.699	51	36
		0.218	0.282	0.282	0.218				
-5	5	-5	-1.1067	1.1067	5	4	3.143	51	33
		0.1846	0.3154	0.3154	0.1846				
-10	10	-10	-1.3218	1.3218	10	4	0.6419	54	30
		0.1628	0.3372	0.3372	0.1628				
-20	20	-20	-1.4113	1.4113	20	4	-2.23	74	41
		0.165	0.335	0.335	0.165				
-30	30	-30	-1.45	1.45	30	4	-3.901	86	49
		0.1662	0.3338	0.3338	0.1662				
-40	40	-40	-1.4716	1.4716	40	4	-5.079	100	50
		0.1664	0.3336	0.3336	0.1664				
-50	50	-50	-1.485	1.485	50	4	-5.988	99	51
		0.1665	0.3335	0.3335	0.1665				
-60	60	-60	-1.4943	1.4943	60	4	-6.729	112	52
		0.1666	0.3334	0.3334	0.1666				
-80	80	-80	-1.5063	1.5063	80	4	-7.895	138	58
		0.16663	0.3337	0.3337	0.16663				
-100	100	-100	-1.5133	1.5133	100	4	-8.797	145	71
		0.1667	0.3333	0.3333	0.1667				

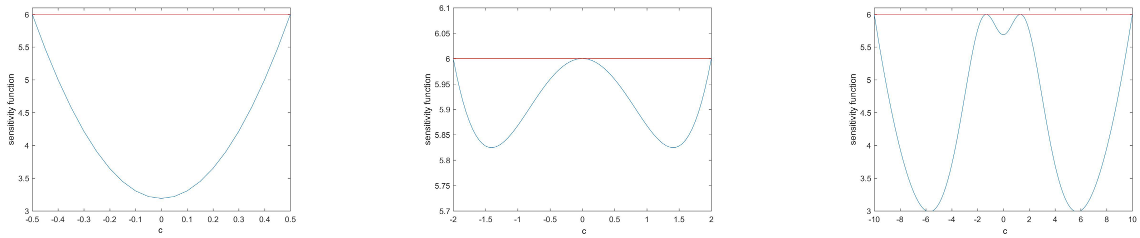
[a] Time consumed by fmincon to search for the two inner points and three weights within a class of four support points

[b] Time consumed by fmincon to search for only one design point and one weight

space gets larger, the obtained optimal designs has 4-support points, where the outer points are the two ends of the induced design space and the inner two points are symmetric about zero. The weights of the paired points are the same, while they might differ across pairs. In Table 3.1, the 4-support points have emerged when the design induced space is [-3,3] and larger. We can also observe from Table 3.1 that the speed gain is more prominent to search over the symmetric complete class of at most 4-point designs in comparison to searching over the space of 4-point designs. For example, when the induced design space is [-3,3], the CPU time it took for finding the optimal designs over the space of 4-point designs is 50 milliseconds compared to 38 milliseconds for searching over the ξ_s^* of Ξ_s. And it took 145 milliseconds to search for ξ^* in comparison to 71 milliseconds search for ξ_s^* for the induced design space [-100,100]. We include the value of the objective function 'OF' , i.e. $-\log|\tilde{C}(\xi^*, \boldsymbol{\theta})|$, for the obtained optimal design. As can be seen from the table, the objective function decreases as the symmetric induced design space gets wider. We observe that as the symmetric induced design space gets wider more weights are placed on the two inner support points until a ratio of about 1:2:2:1 is achieved where the weights become stable.

All the obtained designs found in Table 3.1 have been verified by GET where the plots of the sensitivity function $d_D(c, \xi)$ curve are bounded above by the reference line which is equal to 6 with equality at the support points of the design. As an example, we draw the plots of the functions, $d_D(c, \xi)$, in Figure 3.1 for three different designs: 2-, 3-, and 4-points design. They are the designs that we obtained for the induced design spaces [-0.5,0.5], [-2,2], and [-10,10], respectively. This indicates that the three designs are D-optimal.

Figure 3.1: Locally D-optimal Design Verification for Three Different Symmetric Designs



design for [-0.5,0.5] design for [-2,2] design for [-10,10]

3.4 Discussion

In this chapter, we built on the complete class results established by Kim and Kao (2019) for the simple mixed responses model. For D-optimality, it is easily seen, as in Lemma 3.2.2, that the symmetric design is as good as any other design. If our interest is in finding optimal designs among the symmetric designs for the simple mixed responses model with a symmetric induced design region, Theorem 3.2.4 assures that we can find such an optimal design among those with at most 4 symmetric support points, including the two endpoints of the design region. Further, the D-optimal design found within this complete class will be optimal among all (symmetric and non-symmetric) designs. Based on this, our result facilitates the search for D-optimal designs over symmetric induced design regions. For example, the D-optimal designs that have 2-support points in the induced design region $[-D, D]$ have the form of $\{(-D, 0.5), (D, 0.5)\}$. Meanwhile, D-optimal designs that have 3-support points can take the form $\{(-D, 0.5 - w^*/2), (0, w^*), (D, 0.5 - w^*/2)\}$. Finally, the 4-support points D-optimal designs will have the form $\{(-D, 0.5 - w^*/2), (-c^*, w^*/2), (c^*, w^*/2), (D, 0.5 - w^*/2)\}$.

In regard to the computational search, we showed the advantage in speed for searching for the symmetric locally D-optimal designs using the symmetric complete class and the off-the-shelf computational tool of the fmincon solver in MATLAB.

The simple mixed responses model can be extended by including group effects. The complexity in constructing the optimal designs in this case will be in the selection of the number of support points corresponding to each group. Thus, the focus will be on identifying a complete class of locally optimal designs for the extended model to estimate the model parameters. This optimal problem will be studied in the next chapter.

Chapter 4

MIXED RESPONSES MODELS WITH GROUP EFFECT AND ONE COVARIATE

In many experiments, when using a mixed responses model, the subject's responses may depend on some qualitative explanatory variables such as gender, age group, race, and so on. The search for optimal designs becomes a complicated task for models that include these factor effects. Here, we consider some mixed responses models with group effects and one continuous covariate. The focus is on identifying a complete class of locally optimal designs for estimating the model parameters. The group effects are considered in several experimental studies, which allow for the heterogeneity among units. They are formed by the different level combinations of factor effects included in the model.

There are numerous contributions to studying experimental designs for GLMs that involve factor effects. For example, Yang, Mandal, and Majumdar (2012) investigated the locally D-optimal designs for factorial experiments with only two qualitative factors that are assumed to have two-factor levels. The response variable is binary and is modeled by a GLM with logit, probit, log-log, and complementary log-log links. Yang, Mandal, and Majumdar (2016) generalized their previous results by focusing on k two-level factors with a binary response. They proposed *lift-one algorithm* and *exchange algorithm* to find good designs and compared their performance with commonly used algorithms. Their work was further extended by Yang and Mandal (2015) to include factors with multi-levels and allow the response to follow a single parameter exponential family. They also modified the *lift-one algorithm* and the *exchange algorithm*, so that these algorithms could be applied to their work. The research works mentioned above contain factorial effects with no continuous covariate in the model.

On the other hand, Tan (2015) and Wang (2018) studied optimal designs for some GLMs that have both factorial effects and one covariate. Tan (2015) used the orthogonal array technique to construct locally D-optimal designs that have a small number of support points. Wang (2018) expanded the work of Tan (2015) by allowing the factorial effects to include more flexible interactions of certain orders. He used the concept of strength $t+$ orthogonal array introduced by Hedayat (1989) to find locally D-optimal designs with a reduced number of support points. Both works were built upon previous work on a similar study by Stufken and Yang (2012a). In particular, Stufken and his coauthor considered finding locally optimal designs for GLM with group effects and one covariate.

Introducing factors in the mixed responses model is required in many cases since the responses may be affected by other factors besides the continuous covariate. For example, in dose-finding experiments where both toxicity and efficacy are measured, some factors besides dose such as age, gender, or size of the tumor may have a great impact on patient's responses. The idea of including factor effects in the mixed responses model involving both toxicity and efficacy responses was discussed by Lei, Yuan, and Yin (2011). They included the patient's prognoses as covariates in their Bayesian adaptive randomization procedure, which allows for different response rates for the heterogeneity of patients. Zhang and Hu (2009) explained how the covariance information could influence the responses in clinical trial experiments which lead them to propose an efficient Covariate-Adjusted Response-Adaptive (CARA) designs. Following this line of research, we are concerned with optimal design problems for mixed response models with group effects and a continuous covariate.

4.1 Models With no Common Parameters Across Subject Groups

Now let us describe the mixed responses model with group effects in a more general form. With the same notation as in the simple mixed responses model x is the continuous independent variables, y is the continuous response variable, and $z \in \{0, 1\}$ is the binary response variable. In addition to these terms, we assume that the response variables are influenced by L qualitative factors. Let $s = s_1 s_2 \cdots s_L$ be the total number of groups formed by the L qualitative factors, where s_l is the number of levels of the lth factor, and $(y(l, x), z(l, x))$ denotes the response vector of a subject in the lth group having continuous explanatory variables x. With $(l, x) \in \Omega$, where $\Omega = \bigcup_{l=1}^{s} \{l\} \times \chi_l$ and $\chi_l \subset \mathbb{R}$, we consider the following model:

$$
\begin{aligned}
& y(l,x)|z(l,x) = z \sim N(\mu_{y|z}(l,x), \sigma_z^2), \quad \text{with} \quad \mu_{y|z}(l,x) = \boldsymbol{f}_{zl}^T(x)\boldsymbol{\beta}_{zl}, \quad z = 0, 1, \\
& \text{and} \\
& \text{Prob}(z(l,x) = 1) = \text{P}(\boldsymbol{f}_{2l}^T(x)\boldsymbol{\beta}_{2l}) \equiv p(l,x),
\end{aligned}
\tag{4.1}
$$

where $\boldsymbol{f}_{rl}(x)$ is a vector of some given functions of x and $\boldsymbol{\beta}_{rl} \in \mathbb{R}^{p_{rl}}$ is an unknown parameter vector consisting of p_{rl} unknown coefficients, $r = 0, 1, 2$, $l = 1, \cdots, s$. The terms $\mu_{y|z}(l, x)$ and σ_z^2 denote the mean and the variance of the conditional distribution of $y(l, x)$ given $z(l, x) = z$, respectively. As before, $\text{P}(\cdot)$ is some differentiable cumulative distribution function.

Consequently, the joint probability function for $(y(l, x), z(l, x)) = (y, z)$ can be obtained by direct factorization $f(y, z) = f(z)f(y|z)$ as:

$$
f(y,z) = [p(l,x)]^z [1 - p(l,x)]^{1-z} [f(y|z=1)]^z [f(y|z=0)]^{1-z}. \tag{4.2}
$$

Then, the log-likelihood function is

$$\begin{aligned}\log[f(y,z)] &= z\log[p(l,x)] + (1-z)\log[1-p(l,x)] \\ &\quad + z\left[\frac{[y-\mu_{y|1}(l,x)]^2}{-2\sigma_1^2}\right] + (1-z)\left[\frac{[y-\mu_{y|0}(l,x)]^2}{-2\sigma_0^2}\right] \\ &\quad - \frac{z}{2}\log[\sigma_1^2] - \frac{1-z}{2}\log[\sigma_0^2] - \frac{1}{2}\log[2\pi].\end{aligned} \tag{4.3}$$

As noted previously, we assume that the bivariate responses (y,z) of the same observational unit might be correlated, but are independent between different units. We let the parameter vector of interest be $\boldsymbol{\theta} = (\boldsymbol{\theta}_0^T, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$, where $\boldsymbol{\theta}_r = (\boldsymbol{\beta}_{r1}^{\mathrm{T}}, \cdots, \boldsymbol{\beta}_{rs}^T)^T$ $r = 0,1,2$. We also write $\mu_{y|z}(l,x) = \boldsymbol{g}_z^T(l,x)\boldsymbol{\theta}_z$, $z = 0,1$, and $p(l,x) = P[\boldsymbol{g}_2^T(l,x)\boldsymbol{\theta}_2]$, where $\boldsymbol{g}_r^T(l,x) = (\mathbf{0}_{p_{r,1}}^T, \cdots, \mathbf{0}_{p_{r,l-1}}^T, \boldsymbol{f}_{rl}^T(x), \mathbf{0}_{p_{r,l+1}}^T, \cdots, \mathbf{0}_{p_{r,s}}^T)$. If $p_{rl} = p_r$, then $\boldsymbol{g}_r(l,x) = \boldsymbol{e}_l \otimes \boldsymbol{f}_{rl}(x)$ for all $l = 1, \cdots, s$, where $\otimes$ is the Kronecker product, and $\boldsymbol{e}_l = (0,\cdots,0,1,0,\cdots,0)^T \in \mathbb{R}^s$ with the lth element as 1 and the other elements as 0.

For simplicity, we also write $c_l = \boldsymbol{g}_2^T(l,x)\boldsymbol{\theta}_2$, and for $\boldsymbol{\theta}$, the (individual) information matrix at (l,x) can be calculated from the log-likelihood function as:

$$I(l,x) = \bigoplus_{r=0}^{2} \Gamma_r(c_l)\boldsymbol{g}_r(l,x)\boldsymbol{g}_r^T(l,x), \tag{4.4}$$

where

$$\Gamma_0(c_l) = \frac{1-\mathrm{P}(c_l)}{\sigma_0^2}, \Gamma_1(c_l) = \frac{\mathrm{P}(c_l)}{\sigma_1^2}, \text{and } \Gamma_2(c_l) = \frac{[\mathrm{P}'(c_l)]^2}{\mathrm{P}(c_l)(1-\mathrm{P}(c_l))}. \tag{4.5}$$

Here, $\oplus$ is the direct sum operator, and $\mathrm{P}'(c_l)$ is the first derivative of $\mathrm{P}(c_l)$ with respect to c_l. Note that if $p_{rl} = p_r$, then $\boldsymbol{g}_r(l,x)\boldsymbol{g}_r^T(l,x) = (\boldsymbol{e}_l^T \otimes \boldsymbol{f}_{rl}^T(x))^T(\boldsymbol{e}_l^T \otimes \boldsymbol{f}_{rl}^T(x)) = \boldsymbol{e}_l\boldsymbol{e}_l^T \otimes \boldsymbol{f}_{rl}(x)\boldsymbol{f}_{rl}^T(x)$, where $\boldsymbol{e}_l\boldsymbol{e}_l^T = \mathrm{diag}(\boldsymbol{e}_l) \in \mathbb{R}^{s\times s}$, the diagonal matrix whose diagonal elements are as $\boldsymbol{e}_l$.

In what follows, we focus on the case where $p_{rl} = p_r$, for all $l = 1, \cdots, s$. It should be straightforward to extend our results to cases where not all p_{rl}'s are equal. Our focus is on a continuous design problem for finding an 'optimal' continuous design measure ξ on Ω that allows a precise statistical inference about $\boldsymbol{\theta}$. In particular, ξ can be viewed as a probability measure, and for given explanatory variables $(l, x = x_{lj})$, we write $\xi(l, x_{lj}) = w_{l,j} \in [0,1]$ as the proportion of times (l, x_{lj}) occurs in the experiment. Specifically, $(l, x_{lj}) \in \Omega$, $j =$

$1, \cdots m_l$, $l = 1, \cdots, s$, and m_l is the total number of distinct values of x in group l. Then the continuous design with group effects can be expressed as follows:

$$\xi = \left\{ \begin{array}{ccc|c|ccc} (1, x_{1,1}) & \cdots & (1, x_{1,m_1}) & \cdots & (s, x_{s,1}) & \cdots & (s, x_{s,m_s}) \\ w_{1,1} & \cdots & w_{1,m_1} & \cdots & w_{s,1} & \cdots & w_{s,m_s} \end{array} \right\},$$

where the value of $w_{l,j}$ is the weight for the corresponding design point (l, x_{lj}). Clearly, we may write $w_{l,j} = w_l w_{j|l}$, where w_l is the (marginal) proportion of the total number of observations that are assigned to the lth group, and $w_{j|l}$ is the (conditional) proportion of the number of observations having continuous covariates x_{lj} within the given group l. Specifically, $w_l = \sum_{j=1}^{m_l} w_{l,j}$ and $w_{j|l} = w_{l,j}/w_l$. In the continuous designs framework, we allow $w_{l,j}$ to be any real number between 0 and 1, $\sum_{l=1}^{s} \sum_{j=1}^{m_l} w_{l,j} = 1$. This is mainly for mathematical simplicity, and in cases where we have a total of N subjects, we will need $Nw_{l,j}$ to be an integer, and it is not uncommon to consider the rounding methods after obtaining optimal $w_{l,j}$'s under certain optimality criterion.

We now are ready to write down the information matrix of the parameter vector $\boldsymbol{\theta}$ as:

$$\begin{aligned} M(\xi) &= \sum_{l=1}^{s} \sum_{j=1}^{m_l} w_{l,j} I(l, x_{lj}) \\ &= \sum_{l=1}^{s} w_l \left[\boldsymbol{e}_l \boldsymbol{e}_l^T \otimes \bigoplus_{r=0}^{2} \sum_{j=1}^{m_l} w_{j|l} \Gamma_r(c_{lj}) \boldsymbol{f}_{rl}(x_{lj}) \boldsymbol{f}_{rl}^T(x_{lj}) \right] \\ &= \bigoplus_{l=1}^{s} w_l \left[\bigoplus_{r=0}^{2} \sum_{j=1}^{m_l} w_{j|l} \Gamma_r(c_{lj}) \boldsymbol{f}_{rl}(x_{lj}) \boldsymbol{f}_{rl}^T(x_{lj}) \right] \equiv \bigoplus_{l=1}^{s} [w_l M_l(\tau_l)]. \end{aligned} \tag{4.6}$$

Here, $M_l(\tau_l) = \bigoplus_{r=0}^{2} \sum_{j=1}^{m_l} w_{j|l} \Gamma_r(c_{lj}) \boldsymbol{f}_{rl}(x_{lj}) \boldsymbol{f}_{rl}^T(x_{lj})$ is the information matrix for $(\boldsymbol{\beta}_{0l}^T, \boldsymbol{\beta}_{1l}^T, \boldsymbol{\beta}_{2l}^T)^T$ when we treat the lth group as the only group in the experiment, and $\tau_l(x_{lj})$ is the corresponding conditional design measure for that group with $\tau_l(x_{lj}) = w_{j|l}$; we also set $c_{lj} = \boldsymbol{g}_2^T(l, x_{lj}) \boldsymbol{\theta}_2 = \boldsymbol{f}_{2l}^T(x_{lj}) \boldsymbol{\beta}_{2l}$. Note that the optimal designs are effected by the values of $\boldsymbol{\theta}_2$ and σ_z^2 as shown from the structure of the information matrix, although the whole parameter vector $\boldsymbol{\theta}$ is of interest.

In general, for arbitrary p_{rl}'s, we still have $M(\xi) = \bigoplus_{l=1}^{s} [w_l M_l(\tau_l)]$. Consequently, the information matrix, $M(\xi)$, is nonsingular if and only if (a) $w_l > 0$, and (b) the $p_l \times p_l$ matrix $M_l(\tau_l)$ is nonsingular for all $l = 1, \cdots, s$, where $p_l = \sum_{r=0}^{2} p_{rl}$. The nonsingularity of $M(\xi)$ will ensure the estimability of $\boldsymbol{\theta}$, and when $M^{-1}(\xi)$ exists it is proportional to the asymptotic variance-covariance matrix of the likelihood estimate of $\boldsymbol{\theta}$ (under certain regularity conditions). We would like to find a design ξ that yields the most precise estimate of $\boldsymbol{\theta}$ through optimizing a specific optimality criterion such as the A, D, E, and Φ_p criteria of Kiefer. However, a major challenge for finding such an optimal design is again that the information matrix $M(\xi)$ depends on the unknown parameters (through c_{lj}). In this work, we place our focus on finding locally optimal design ξ^* (Chernoff, 1953).

Since commonly used optimality criteria are functions of the eigenvalues of the information matrix, we would only focus on such functions in this work. The eigenvalues of $M(\xi)$ have the form of $w_l \lambda_i[M_l(\tau_l)]$, where $\lambda_i[M]$ is the ith eigenvalue of M; $i = 1, \cdots, p_l, l = 1, \cdots, s$. For convenience, we write the eigenvalues of $M(\xi)$ as $\lambda_{l,i}[M(\xi)] = w_l \lambda_i[M_l(\tau_l)]$, and when $M(\xi)$ is nonsingular, we have $\lambda_{l,i}[M^{-1}(\xi)] = \lambda_i[M_l^{-1}(\tau_l)]/w_l$. For designs allowing estimable $\boldsymbol{\theta}$, we write the Φ_q-optimality criterion as:

$$\sum_{l=1}^{s} \sum_{i=1}^{p_l} \lambda_{l,i}[M^{-1}(\xi)]^q = \sum_{l=1}^{s} w_l^{-q} \sum_{i=1}^{p_l} \lambda_i[M_l^{-1}(\tau_l)]^q, \text{ for } \quad q \in (0, \infty). \tag{4.7}$$

We note that the set of the Φ_q-criteria is essentially that of the Φ_p-criteria of Kiefer (1974); when $q = 1$, the criterion is reduced to the so-called A-optimality criterion. We also followed Kiefer (1974) to include the two (limiting) cases, namely $\Phi_0[M(\xi)] = \sum_{l=1}^{s} \sum_{i=1}^{p_l} \log \lambda_{l,i}[M^{-1}(\xi)]$, i.e. the D-criterion, and $\Phi_\infty[M(\xi)] = \max \lambda_{l,i}[M^{-1}(\xi)]$, i.e. the E-criterion. For these criteria, it can be easily seen that, among the designs ξ having the same marginal weights w_l's, design ξ^* is Φ_q-optimal if and only if its corresponding τ_l^* minimizes $\Phi_q[M_l(\tau_l)]$ for all $l = 1, \cdots, s$. This result is a direct consequence of the result found in Section 7 of Schwabe (2012) and it is summarized below.

Lemma 4.1.1. *Under a mixed responses model in* (4.1)*, for the designs ξ with given marginal weights w_l, then the design ξ^* is Φ_q-optimal if and only if its corresponding design measure τ_l minimizes $\Phi_q[M_l(\tau_l)]$ for all $l = 1, \cdots, s$.*

Lemma 4.1.1 can be applied to the case where we have no control on the proportion of the marginal weights, but can only select the weights and the values of the continuous explanatory variable in each group. To illustrate Lemma 4.1.1, we consider the following example.

Example 4.1.1. Suppose the mixed responses model in (4.1) has a covariate, such as dose, and the range of the covariate is [0,10], and in addition to the covariate the responses depend on another factor such as gender where we have two groups $s = 2$ (male and female). It is given that, among the experimental subjects, 60% of the patients are male and 40% are female. Suppose that the (4.1) has two parameters, namely scale and location, in each group; i.e., $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^{\mathrm{T}} = (0.4, 0.7), \boldsymbol{\beta}_{22}^{\mathrm{T}} = (0.8, 0.6))^T$, and $\sigma_0^2 = \sigma_1^2 = 1$. Based on Lemma 4.1.1, we can search for the conditional measures, τ_l, for the given marginal weights $w_1 = 0.6$, and $w_2 = 0.4$. With a computer search, the conditional D-optimal measures τ_l are given as:

$$\tau_1 = \left\{ \begin{array}{ccc} (1, x_{1,1} = 0) & (1, x_{1,2} = 3.164) & (1, x_{1,3} = 10) \\ w_{1|1} = 0.4408 & w_{2|1} = 0.4003 & w_{3|1} = 0.1589 \end{array} \right\}, \text{ and}$$

$$\tau_2 = \left\{ \begin{array}{ccc} (2, x_{2,1} = 0) & (2, x_{2,2} = 3.637) & (2, x_{2,3} = 10) \\ w_{1|2} = 0.4541 & w_{2|2} = 0.3818 & w_{3|2} = 0.1641 \end{array} \right\}.$$

The marginal weights $w_{l,j}$ can be obtained based on the conditional measures as follows:

$w_{1,1} = w_1 * \tau_1(0) = 0.6 * 0.4408 = 0.26448$ $\quad$ $w_{2,1} = w_2 * \tau_2(0) = 0.4 * 0.4541 = 0.18164$

$w_{1,2} = w_1 * \tau_1(3.164) = 0.6 * 0.4003 = 0.24018$ $\quad$ $w_{2,2} = w_2 * \tau_2(3.637) = 0.4 * 0.3818 = 0.15272$

$w_{1,3} = w_1 * \tau_1(10) = 0.6 * 0.1589 = 0.09534$ $\quad$ $w_{2,3} = w_2 * \tau_2(10) = 0.4 * 0.1641 = 0.06564$

Thus, the Φ_D-optimal design can be expressed as follows:

$$\xi^* = \left\{ \begin{array}{ccc|ccc} (1, x_{1,1} = 0) & (1, 3.1642) & (1, x_{1,3} = 10) & (2, x_{2,1} = 0) & (2, 3.6372) & (2, x_{2,3} = 10) \\ w_{1,1} = 0.26448 & 0.24018 & w_{1,3} = 0.09534 & w_{2,1} = 0.18164 & 0.15272 & w_{2,3} = 0.06564 \end{array} \right\}.$$

On the other hand, if the conditional measures $\hat{\tau}_l$ are given, then the marginal weights w_l for each group l can be obtained by the following Lemma 4.1.2, where the proof can be found in the Appendix.

Lemma 4.1.2. *Under a mixed responses model in* (4.1)*, for the designs ξ with given conditional measures $\hat{\tau}_l$, then the design ξ^* is Φ_q-optimal if its marginal weights w_l are given as:*

$$w_l = \frac{\left\{\sum_{i=1}^{p_l} \lambda_i [M_l^{-1}(\hat{\tau}_l)]^q\right\}^{1/(q+1)}}{\sum_{l=1}^{s}\left\{\sum_{i=1}^{p_l} \lambda_i [M_l^{-1}(\hat{\tau}_l)]^q\right\}^{1/(q+1)}}, \qquad 0 < q < \infty. \tag{4.8}$$

We also note that for the limiting cases of D- and E-optimality, we have:

$$w_l = p_l / \sum_{l=1}^{s} p_l, \qquad q \longrightarrow 0; \tag{4.9}$$

$$w_l = \frac{\lambda_{max}[M_l^{-1}(\hat{\tau}_l)]}{\sum_{l=1}^{s} \lambda_{max}[M_l^{-1}(\hat{\tau}_l)]}, \qquad q \longrightarrow \infty. \tag{4.10}$$

Now we illustrate the use of Lemma (4.1.2) in the next example.

Example 4.1.2. With the same scenario as Example (4.1.1), assume now that the researchers have conditional measures $\hat{\tau}_l$ that they would like to use. They are A-optimal designs listed below.

$$\hat{\tau}_1 = \left\{\begin{array}{ccc} (1, x_{1,1} = 0) & (1, x_{1,2} = 3.4467) & (1, x_{1,3} = 10) \\ w_{1|1} = 0.6189 & w_{2|1} = 0.3682 & w_{3|1} = 0.0129 \end{array}\right\}, \text{ and}$$

$$\hat{\tau}_2 = \left\{\begin{array}{ccc} (2, x_{2,1} = 0) & (2, x_{2,2} = 3.9888) & (2, x_{2,3} = 10) \\ w_{1|2} = 0.6269 & w_{2|2} = 0.3612 & w_{3|2} = 0.0119 \end{array}\right\}.$$

By equation (4.8) with $q = 1$, the calculated values of $w_1 = 0.4891$ and $w_2 = 0.5109$ are used to compute each $w_{l,j}$, and the resulting design is

$$
\xi^* = \left\{ \begin{array}{ccc|ccc} (1,0) & (1,3.4467) & (1,10) & (2,0) & (2,3.9888) & (2,10) \\ w_{1,1}=0.3027 & 0.1801 & w_{1,3}=0.0063 & w_{2,1}=0.3203 & 0.1845 & w_{2,3}=0.0061 \end{array} \right\}.
$$

It is noteworthy that the given $\hat{\tau}_l$'s are A-optimal, and the resulting ξ^* can be shown to be A-optimal. In fact, the same statement is true for any Φ_q-criterion, and we have the following general result.

Corollary 4.1.3. *Under a mixed responses model in* (4.1)*, the design* ξ *is* Φ_q*-optimal if the conditional measure* τ_l *minimize* $\Phi_q[M_l(\tau_l)]$ *for all* $l = 1, \cdots, s$*, and the marginal weights* w_l *satisfy equation* (4.8) *for* $q \in (0, \infty)$*,* (4.9) *for* $q \longrightarrow 0$*, and* (4.10) *for* $q \longrightarrow \infty$*.*

Recall that the information matrix $M(\xi)$ for model (4.6) is a block diagonal matrix with the information matrices $M_l(\tau_l)$ of the groups as diagonal blocks. Based on Corollary 4.1.3, the determination of the optimal deign ξ^* simplifies to find an optimal conditional measure τ_l^* for each group individually. Then, aggregate these conditional measures together after adjusting the weights using the same method as in Example 4.1.2. We also note that while the results in this section are presented for one continuous covariate, they can be easily extended when there are two or more continuous covariates.

4.2 Complete Class Results for Mixed Responses Model With Groups and Multiple Slopes Effects

We now consider the following specific joint model for the responses $(y(l,x), z(l,x))$

$$
\begin{aligned}
& y_{lj}|z_{lj} = z \sim N(\boldsymbol{f}_l^T(x_{lj})\boldsymbol{\beta}_{zl} = \alpha_{zl} + \beta_{zl}x_{lj}, \sigma_z^2), \qquad \text{for} \quad z = 0, 1, \\
& \text{and} \\
& \text{Prob}(z_{lj} = 1) = \text{P}(\boldsymbol{f}_l^T(x_{lj})\boldsymbol{\beta}_{2l} = \alpha_{2l} + \beta_{2l}x_{lj}).
\end{aligned} \tag{4.11}
$$

Here, $\boldsymbol{f}_l(x_{lj}) = (1, x_{lj})^T$, and $\boldsymbol{\beta}_{rl} = (\alpha_{rl}, \beta_{rl})^T$; $r = 0, 1, 2$. The unknown parameters in the proceeding models are σ_z^2, and $\boldsymbol{\theta}_r$ where $\boldsymbol{\theta}_r = (\boldsymbol{\beta}_{r1}^{\mathrm{T}}, \cdots, \boldsymbol{\beta}_{rs}^T)^T$ which is a vector of size $2s$. Our goal is to determine a sharp upper bound on the number of support points in each group for (locally) optimal designs. As in the previous section, to ease the employment of the information matrix we work with the induced design points $c_{lj} = \boldsymbol{f}_l(x_{lj})^T \boldsymbol{\beta}_{2l}$. In this case, the design can be expressed as $\xi = \{w_1\tau_1, \cdots, w_s\tau_s\}$ where $\tau_l = \{(c_{lj}, w_{j|l}), l = 1, ..., s, j = 1, ..., m_l\}$ for $c_{lj} \in [D_{l1}, D_{l2}]$. We then have a matrix $B_{1l}^{-1} = \begin{pmatrix} 1 & 0 \\ \alpha_{2l} & \beta_{2l} \end{pmatrix}$, so that $B_{1l}^{-1} \times \boldsymbol{f}_l(x_{lj}) = \boldsymbol{f}_l(c_{lj})$. Thus, with this representation under model (4.11), the information matrix in equation (4.6) can be expressed as:

$$
\begin{aligned}
M(\xi, \boldsymbol{\theta}) &= \bigoplus_{l=1}^{s} w_l B_l(\boldsymbol{\theta}) \Big[\bigoplus_{r=0}^{2} \sum_{j=1}^{m_l} w_{j|l} \Gamma_r(c_{lj}) \boldsymbol{f}_l(c_{lj}) \boldsymbol{f}_l^T(c_{lj}) \Big] B_l^T(\boldsymbol{\theta}) \\
&\equiv \bigoplus_{l=1}^{s} B_l(\boldsymbol{\theta}) \Big[\sum_{j=1}^{m_l} w_l w_{j|l} C_l(\boldsymbol{\theta}, c_{lj}) \Big] B_l^T(\boldsymbol{\theta}),
\end{aligned}
\tag{4.12}
$$

where

$$
C_l(\boldsymbol{\theta}, c_{lj}) = \bigoplus_{r=0}^{2} \Gamma_r(c_{lj}) \boldsymbol{f}_l(c_{lj}) \boldsymbol{f}_l^T(c_{lj}), \tag{4.13}
$$

$B_l(\boldsymbol{\theta}) = \text{diag}({\sigma_0}^{-1} B_{1l}, {\sigma_1}^{-1} B_{1l}, B_{1l})$ is a nonsingular matrix with size 6-by-6, and

$$
\Gamma_0(c_{lj}) = \frac{1}{1 + e^{c_{lj}}}, \Gamma_1(c_{lj}) = \frac{e^{c_{lj}}}{1 + e^{c_{lj}}}, \text{and } \Gamma_2(c_{lj}) = \frac{e^{c_{lj}}}{(1 + e^{c_{lj}})^2}. \tag{4.14}
$$

By decomposing the information matrix to this form, we can direct our attention to the matrix $C_l(\boldsymbol{\theta}, c_{lj})$, since $B_l(\boldsymbol{\theta})$ depends only on the parameters and is not affected by the selected design.

As mentioned before, our interest is to find sharper upper bounds on the number of support points for the locally optimal design for model (4.1) by borrowing the complete class approach. Our model is an extension of the simple mixed responses model used by Kim and Kao (2019), where they made a contribution by finding a complete class for locally optimal designs that are formed by designs of at most four support points, including the end points of the design region. This result, is extended in the next theorem where we include some groups in the model.

Theorem 4.2.1. *For model* (4.11)*, there exists a complete class of optimal designs* $\xi^* = \{w_1\tau_1^*, \cdots, w_s\tau_s^*\}$ *for estimating* $\boldsymbol{\theta}$ *formed by conditional measures* $\tau_l^* = \{(c_{lj}^*, w_{j|l}^*),$
$l = 1, ..., s, j = 1, ..., 4\}$*, including the endpoints* D_{l1} *and* D_{l2}*.*

Proof. The complete class result obtained from Theorem 3.2.1 implies that $\sum_{j=1}^{m^*} w_j^* C(\boldsymbol{\theta}, c_j^*) \geq \sum_{j=1}^{m} w_j C(\boldsymbol{\theta}, c_j)$, where (c_j^*, w_j^*) and (c_j, w_j) are the pairs of the support points and corresponding weights, $m^* \leq 4$, and $m \geq 4$ are the number of support points of ξ^* and ξ, respectively. We note that the above inequality will still hold when $\sum w_j^* = \sum w_j \neq 1$, which can be observed from Lemma 2.3.1. Since all the groups within the total information matrix given in (4.13) have the same form as the C matrix of the simple mixed responses model in (3.5), we can therefore conclude that for any conditional measure $\tau_l = \{(c_{lj}, w_{j|l}), l = 1, ..., s, j = 1, ..., m_l\}$ for group l, there exists a conditional measure $\tau_l^* = \{(c_{lj}^*, w_{j|l}^*), l = 1, ..., s, j = 1, \cdots, 4\}$ with $\sum w_l w_{j|l}^* = \sum w_l w_{j|l}$, $m_l^* \leq 4$, and two of the c_{lj}^* points are the endpoints of the design space of that group so that

$$\sum_{j=1}^{m_l} w_{j|l} C_l(\boldsymbol{\theta}, c_{lj}) \leq \sum_{j=1}^{m_l^*} w_{j|l}^* C_l(\boldsymbol{\theta}, c_{lj}^*). \tag{4.15}$$

This implies that

$$B_l(\boldsymbol{\theta})\Big[\sum_{j=1}^{m_l} w_l w_{j|l} C_l(\boldsymbol{\theta}, c_{lj})\Big] B_l^T(\boldsymbol{\theta}) \leq B_l(\boldsymbol{\theta})\Big[\sum_{j=1}^{m_l^*} w_l w_{j|l}^* C_l(\boldsymbol{\theta}, c_{lj}^*)\Big] B_l^T(\boldsymbol{\theta}), \tag{4.16}$$

which in turn leads to the conclusion. $\square$

In fact, with Corollary 4.1.3, it should be easy to see that a complete class for the optimal designs ξ^* can be formed by combining s complete classes for the conditional design τ_l, $l = 1, \cdots, s$. This holds true even when the s complete classes are not all the same (e.g., in cases with different models for different subject groups). The previous theorem is a special case of this slightly more general result. We also note that Theorem 4.2.1 is helpful in narrowing the search of locally optimal designs for mixed responses models with group effects to a small class of designs. Moreover, the complete class results generalize the results on the simple mixed responses model which can be considered as a special case with $s = 1$. Furthermore, for group l with symmetric induced design range, it is possible to restrict the search to a complete class of symmetric designs under the D-optimality based on Corollary

4.1.3 and Lemma 3.2.2. We then can find the symmetric D-optimal conditional measure $\tau^*_{sl} = \{(\pm c^*_{lj}, w^*_{j|l}/2), l = 1, ..., s, j = 1, \cdots, m^*_l\}$ by searching over the class of symmetric designs of at most 4 support points for that group. We summarize this result in the next Lemma.

Lemma 4.2.2. *Consider the D-optimality criterion for the mixed responses model* (4.11)*, for the groups with symmetric induced design regions $c_{lj} \in [-D_l, D_l]$, the corresponding portions of the complete class in Theorem 4.2.1 can be replaced by the class of symmetric designs with at most four support points including $-D_l$ and D_l.*

Proof. Suppose that group k in model (4.11) has a symmetric induced design region $c_{kj} \in [-D_k, D_k]$, from equation (4.12) the corresponding group information matrix is $M(\tau_k, \boldsymbol{\theta}) = B_k(\boldsymbol{\theta})\Big[\sum\limits_{j=1}^{m_k} w_k w_{j|k} C_k(\boldsymbol{\theta}, c_{kj})\Big] B_k^T(\boldsymbol{\theta})$, which its $C_k(\boldsymbol{\theta}, c_{kj})$ matrix contains the same $\Gamma(\cdot)$ functions as the $C(\boldsymbol{\theta}, c_j)$ matrix for the simple mixed responses model in (3.5). Thus, the symmetric matrix C_{sk} of the C_k matrix has the same form and the number of $\Gamma(\cdot)$ functions as the symmetric matrix C_s for the simple mixed responses model in (3.8). Based on Lemma (3.2.2) and Theorem (3.2.4), then it is implied that there exists a symmetric D-optimal conditional measure in group k such as $\tau^*_{sk} = \{(\pm c^*_{kj}, w^*_{j|k}/2) : 0 \leq, c^*_{kj} \leq D_k, w^*_{k,j} > 0, j = 1, \cdots, m^*_k\}$ where $m^*_k \leq 2$, that dominates the symmetric conditional measure $\tau_{sk} = \{(\pm c_{kj}, w_{j|k}/2) : 0 \leq, c_{kj} \leq D_k, w_{k,j} > 0, j = 1, \cdots, m_k\}$ where $m_k \geq 2$. Furthermore, the matrix $B_k(\boldsymbol{\theta})$ does not depend on the conditional design. Hence, we conclude that $M(\tau^*_{sk}, \boldsymbol{\theta}) \geq M(\tau_{sk}, \boldsymbol{\theta})$. $\square$

Moreover, the computation problem of finding the conditional locally D-optimal designs for groups with symmetric induced design regions under model (4.11) can be obtained by an explicit form presented in the next Lemma.

Lemma 4.2.3. *For the mixed responses model* (4.11)*, the groups with symmetric induced design regions* $c_{lj} \in [-D_l, D_l]$ *have D-optimal conditional measures of the form* $\{(c_{l1} = -D_l, w_{1|l} = \dfrac{1-w_l^*}{2}), (c_{l2} = -c_l^*, w_{2|l} = \dfrac{w_l^*}{2}), (c_{l3} = c_l^*, w_{3|l} = \dfrac{w_l^*}{2}), (c_{l4} = D_l, w_{4|l} = \dfrac{1-w_l^*}{2}), 0 \leq c_l^* \leq D_l, w_l^* \geq 0\}$*. The support points* c_l^* *and the weights* w_l^* *maximize the functions* $\Delta(c_l, w_l)$ *which are defined as* $(E_l - H_l^2)^2\{[(1-w_l)\Gamma_2(D_l) + w_l\Gamma_2(c_l)][(1-w_l)D_l^2\Gamma_2(D_l) + w_l c_l^2\Gamma_2(c_l)]\}$*, where* $E_l = (1-w_l)D_l^2 + w_l c_l^2$*,* $H_l = [(1-w_l)D_l(\Gamma_1(D_l) - \Gamma_0(D_l)) + w_l c_l(\Gamma_1(c_l) - \Gamma_0(c_l))]$*, and* Γ_r *is as defined in* (4.14)*.*

Proof. The proof can be found in the Appendix. □

In most experiments, the used models are flexible when we allow the regression coefficients to have different values across groups as the model (4.11) considered here. However, in other cases, it might be reasonable to allow some coefficients to share the same values across groups. Thus, it is useful to introduce another model in the coming section that assumes common parameters across groups.

4.3 Complete Class Results for Mixed Responses Model With Common Parameters

For simplicity, we focus on the mixed responses model that has the same form as (4.11) but possesses a common slope for l different groups, the joint model for the responses $(y(l,x), z(l,x))$ becomes

$$
\begin{aligned}
& y_{lj}|(z_{lj} = z) \sim N(\boldsymbol{h}_z^T(l, x_{lj})\boldsymbol{\theta}_z, \sigma_z^2), \qquad \text{for} \quad z = 0, 1 \\
& \text{and} \\
& \text{Prob } (z_{lj} = 1) = P(\boldsymbol{h}_2^T(l, x_{lj})\boldsymbol{\theta}_2).
\end{aligned}
\tag{4.17}
$$

Here, $\boldsymbol{\theta}_r = (\boldsymbol{\alpha}_r^T, \beta_r)^T$ are $(s+1) \times 1$ vectors, where $\boldsymbol{\alpha}_r = (\alpha_{r1}, \cdots, \alpha_{rs})^T$ and α_{rl} represents the effect of the lth group in the sub-model r. We focus on the case when $\boldsymbol{h}_r(l, x_{lj}) = \boldsymbol{h}(l, x_{lj}) = (\boldsymbol{e}_l^T, x_{lj})^T$. It should be straightforward to extend our results to cases where not all $\boldsymbol{h}_r(l, x_{lj})$'s are equal. Then, the joint probability function of the responses (y, z) is:

$$f(y_{lj}, z_{lj}) = [\mathrm{P}(\boldsymbol{h}^T(l, x_{lj})\boldsymbol{\theta}_2)]^{z_{lj}}[1 - \mathrm{P}(\boldsymbol{h}^T(l, x_{lj})\boldsymbol{\theta}_2)]^{1-z_{lj}}$$
$$\times \left[\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left\{-\frac{(y_{lj} - \boldsymbol{h}^T(l, x_{lj})\boldsymbol{\theta}_1)^2}{2\sigma_1^2}\right\}\right]^{z_{lj}} \left[\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left\{-\frac{(y_{lj} - \boldsymbol{h}^T(l, x_{lj})\boldsymbol{\theta}_0)^2}{2\sigma_0^2}\right\}\right]^{1-z_{lj}}. \tag{4.18}$$

The information matrix for $\boldsymbol{\theta}$ with a continuous design $\xi = \{(x_{lj}, w_{l,j}), l = 1, ..., s,$ $j = 1, ..., m_l\}$ can be written as:

$$M(\xi, \boldsymbol{\theta}) = \sum_{l=1}^{s}\sum_{j=1}^{m_l} w_{l,j}\Big[\bigoplus_{r=0}^{2}\Gamma_r(c_{lj})\boldsymbol{h}(l, x_{lj})\boldsymbol{h}^T(l, x_{lj})\Big], \tag{4.19}$$

where $c_{lj} = \boldsymbol{h}^T(l, x_{lj})\boldsymbol{\theta}_2$ and Γ_r is defined in (4.5). The information matrix for this model does not possess the nice property of being a block diagonal matrix with the groups information matrices placed on its diagonal blocks. In order to investigate the existence of the complete class in this scenario, we first decompose the information matrix as:

$$M(\xi, \boldsymbol{\theta}) = \sum_{l=1}^{s} B(\boldsymbol{\theta})\sum_{j=1}^{m_l} w_{l,j}\Big[\bigoplus_{r=0}^{2}\Gamma_r(c_{lj})\boldsymbol{h}(l, c_{lj})\boldsymbol{h}^T(l, c_{lj})\Big]B^T(\boldsymbol{\theta}) \tag{4.20}$$

Here, Γ_r is defined in (4.14), $\boldsymbol{h}(l, c_{lj}) = (\boldsymbol{e}_l^T, c_{lj})^T$, and $B(\boldsymbol{\theta}) = \mathrm{diag}({\sigma_0}^{-1}B_1, {\sigma_1}^{-1}B_1, B_1)$ where $B_1^{-1} = \begin{pmatrix} \boldsymbol{I}_s & \boldsymbol{0}_{s\times 1} \\ \boldsymbol{\alpha}_2^T & \beta_2 \end{pmatrix}$ such that $B_1^{-1} \times \boldsymbol{h}(l, x_{lj}) = \boldsymbol{h}(l, c_{lj})$. With the dependency of the matrix $B(\boldsymbol{\theta})$ only on the parameters, we can direct our attention to the matrix $\bigoplus_{r=0}^{2}\Gamma_r(c_{lj})\boldsymbol{h}(l, c_{lj})\boldsymbol{h}^T(l, c_{lj})$ and rewrite it in the form $A_lQ_l(\boldsymbol{\theta}, c_{lj})A_l^T$. Where $A_l =$ $\mathrm{diag}(A_{1l}, A_{1l}, A_{1l})$, with $A_{1l} = \begin{pmatrix} \boldsymbol{e}_l & \boldsymbol{0}_{s\times 1} \\ 0 & 1 \end{pmatrix}$, and the symmetric matrix $Q_l(\boldsymbol{\theta}, c_{lj})$ is given as

$$Q_l(\boldsymbol{\theta}, c_{lj}) = \bigoplus_{r=0}^{2}\Gamma_r(c_{lj})Q_{1l}(\boldsymbol{\theta}, c_{lj}), \text{ with } \quad Q_{1l}(\boldsymbol{\theta}, c_{lj}) = \begin{pmatrix} 1 & c_{lj} \\ c_{lj} & c_{lj}^2 \end{pmatrix}. \tag{4.21}$$

Thus, the information matrix can be expressed in the following form:

$$M(\xi, \boldsymbol{\theta}) = \sum_{l=1}^{s} B(\boldsymbol{\theta})A_l\Big[\sum_{j=1}^{m_l} w_{l,j}Q_l(\boldsymbol{\theta}, c_{lj})\Big]A_l^{\mathrm{T}}B^{\mathrm{T}}(\boldsymbol{\theta}). \tag{4.22}$$

By doing so, we formed a matrix Q_l with a fixed number of Γ functions no matter how many groups we have in the model. With this factorization of the information matrix we are ready to present the complete class result under model (4.17).

Theorem 4.3.1. *For mixed responses model* (4.17)*, a complete class of locally optimal designs can be formed by designs that contain at most 4s support points and* $2s$ *of them are the endpoints of the design space in each group.*

Proof. With the factorization of the group effects in the information matrix in (4.22), the matrix $Q_l(\boldsymbol{\theta}, c_{lj})$ has the same form and number of Γ functions of the $C(c_j)$ matrix in (3.5) no matter how many groups we have in the model. Following the same argument as in the proof for Theorem 4.2.1, then for any conditional measure $\{(c_{lj}, w_{j|l}), l = 1, ..., s, j = 1, ..., m_l\}$ for group l, there exists a conditional optimal measure $\{(c^*_{lj}, w^*_{j|l}), l = 1, ..., s, j = 1, \cdots, 4\}$ with $\sum w^*_{l,j} = \sum w_{l,j}$ and $m^*_l \leq 4$ where two of the c^*_{lj} points are the endpoints of the design space of that group so that

$$\sum_{j=1}^{m_l} w_{l,j} Q_l(\boldsymbol{\theta}, c_{lj}) \leq \sum_{j=1}^{m^*_l} w^*_{l,j} Q_l(\boldsymbol{\theta}, c^*_{lj}). \tag{4.23}$$

This implies that

$$B(\boldsymbol{\theta}) A_l \sum_{j=1}^{m_l} [w_{l,j} Q_l(\boldsymbol{\theta}, c_{lj})] A_l^T B^T(\boldsymbol{\theta}) \leq B(\boldsymbol{\theta}) A_l \sum_{j=1}^{m^*_l} [w^*_{l,j} Q_l(\boldsymbol{\theta}, c^*_{lj})] A_l^T B^T(\boldsymbol{\theta}) \tag{4.24}$$

which in turn leads to the conclusion. □

4.4 Discussion

In this chapter, we extended the simple mixed responses model by introducing the group effects in the model. In the first section, we present a general form of the model, and in the second section, we impose some assumptions on the general form where the model has only one covariate with no common regression coefficients across groups. For simplicity, we will call this model as Model I. With this assumption, the model information matrix has a nice structure with the information matrices for the individual groups on its diagonal blocks. This allows the search for the locally optimal design to be simplified to the (separate) search for the conditional optimal design for each group. Then the overall optimal design is constructed by pooling the conditional optimal designs for the groups after adjusting their weights. This finding was stated in Lemma 4.1.1, Lemma 4.1.2, and Corollary 4.1.3 that facilitate the search for the locally optimal designs. Several examples were used to demonstrate this idea. Based on these findings, we obtain a complete class for a specific model, and the results can be easily extended to some other models.

We then considered a more restricted model than the Model I, namely a model with some common regression coefficients. For convenience, we will call it Model II. For such a model, the structure of the information matrix becomes more complicated, and a complete class for a specific model of this type was stated in Theorem 4.3.1.

Although the theoretical results provide a guide for finding locally optimal designs, but in order to identify the optimal design under a specific criterion, we still need numerical approaches. Dealing with models as complicated as the mixed responses model with group effects will become a challenging task because the number of support points in the optimal designs increases by at most four support points for each additional group. This inspired us in the next chapter to search for algorithms with the best reliable results.

Chapter 5

SEARCH ALGORITHMS

The analytical results of the complete class found in Chapter 4 give a tremendous reduction in the dimension of the problem of finding the optimal design for the mixed responses model with group effects. However, to obtain the optimal design within the complete class Ξ, we need efficient numerical techniques. Finding an algorithm that is suitable for our problem can be a complicated mission since the dimensions of the information matrix can quickly become very large. In this chapter, we will discuss the process of finding some good search algorithms for our needs. These algorithms are used to obtain A- and D-optimal designs whose optimality is then verified by the GET. Finally, a discussion of the obtained results will follow.

5.1 Algorithmic Search

Finding an efficient algorithm for solving optimal design problems is a challenging task, especially for the case with both quantitative covariates and qualitative factors. The problem becomes even more challenging when the mixed responses model is considered. This is mainly due to the complexity of the model structure, and an increased number of parameters that gives an enlarged dimensions of the information matrix. In the process of searching for such an algorithm, the two most important features that attract researchers are the speed and reliability of the algorithm in finding an optimal design. It also helps when the algorithm is easy to code, or there exist user-friendly computer packages, which can be directly applied or easily modified for different situations. A recent review of various algorithms for obtaining optimal designs is made by Mandal et al. (2015).

In general, algorithms can be categorized into two branches depending on their nature, either deterministic algorithms or stochastic algorithms. Deterministic algorithms obey precise steps with repetitive action, while stochastic algorithms follow some random path in solving the problem (Yang, 2010a). Sometimes, a hybrid of the two types of algorithms is utilized to solve some optimization problems.

Some very well-known classical algorithms for finding optimal approximate designs include the Fedorov-Wynn Algorithm (FWA) (Fedorov, 1972; Wynn, 1972) and the Multiplicative Algorithm (MA) developed by Silvey, Titterington, and Torsney (1978). For the Fedorov-Wynn algorithm, the search of optimal designs begins with an initial design which typically is a point in a probability simplex. Specifically, designs considered in this algorithm are normally represented as a (long) vector of weights for all the candidate design points over the (discretized) design space; the weights are allowed to be zero, but they sum to one. In the improvement phase, the algorithm continues to improve the design efficiency by moving the current design towards the vertex, i.e., the design point, of the simplex that gives the greatest improvement (in a small step) at each iteration. This procedure continues until the stopping requirement is achieved. The FWA is known to have some numerical deficiencies (Wu, 1978), and several modifications were later proposed by, e.g., Wu (1978) and Böhning (1986).

The MA also works with updating the weights of the design points. In contrast to FWA, the weights of all points in the design region are updated at each iteration via a multiplicative factor to relatively assign more weights to the design points that give greater information gains. However, MA can be inefficient since it takes many iterations before it reduces the weights of non-support points and eventually deletes them.

Yu (2011) came up with the idea of combining the two classical algorithms; vertex directional method and the multiplicative algorithm in order to optimize the weights and the support points, simultaneously. He also introduced the Nearest Neighbor Exchange (NNE) algorithm that was implemented before applying the MA as a way of speeding it up. The idea of the NNE algorithm is to decide upon the order of the design points sets, then exchange positive weights between the design points and their 'nearest neighbor' adjacent points, where each exchange is supposed to improve the design. The result of the sequential pairwise exchange between the ordered design points discards the design points that are not support points but happen to be close to the support points. The deletion process is done by setting the weights of these points to zero. Yu (2011) called the proposed algorithm, the Cocktail Algorithm (CA) for D-optimal designs. The cocktail algorithm is a combination of three algorithms in succession: vertex directional method, nearest neighbor exchange, and multiplicative algorithm. This combination of algorithms will have a great speed efficiency compared with the performance of each algorithm separately. The main drawback of the CA is its limitation to find D-optimal designs only.

More recently, an exceptional algorithm was proposed by Yang et al. (2013) called the Optimal Weight Exchange Algorithm (OWEA) that aims to update the support points and their corresponding weights in the same fashion as the FWA in conjunction with the Newton optimization method. At each iteration, the support points are updated by adding points to the current support points from a candidate set, and then the weights are optimized using the Newton optimization method and the support points with (nearly) zero weights are removed. The iteration continues until the optimal design is obtained by satisfying the GET. The main reasons for the success of this algorithm are its speed in finding the optimal design, its ability to accommodate almost all members of the Φ_p family, and its capability to handle multi-stage designs.

The strategy of the stochastic algorithms is based on a random search process. The search for the optimal solution is done by exploring different areas of the solution space based on some random mechanism, and then exploiting the identified regions of interest. The advantages of such an algorithm are its speed in identifying solutions that are optimal or near-optimal, it can be assumptions free, and its ability to handle various forms of optimization prblems. However, many algorithms of this type are not supported by strong theoretical evidence of convergence, which means that they may not converge in some situations. But, almost all popularly used stochastic algorithms are shown to provide very good results in practice, although one may need to consider different algorithms for different problems.

Some stochastic algorithms that are used in the statistical literature to search for optimal designs include: Genetic Algorithm (GA) (Holland, 1975), Simulated Annealing (SA) (Kirkpatrick, Gelatt, & Vecchi, 1983), Ant Colony Optimization (ACO) (Dorigo, 1992), Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995), Differential Evolution (DE) (Storn & Price, 1997), and Imperialists Competitive Algorithm (ICA) (Atashpaz-Gargari & Lucas, 2007). All these algorithms fall in the category of nature-inspired algorithms.

One of the most popular nature-inspired metaheuristic optimization techniques is PSO. More recently, PSO was used to search for optimal designs in several fields. For example, Schorning, Dette, Kettelhake, Wong, and Bretz (2017) used PSO to find the approximate optimal design for dose finding studies with bivariate responses. Lukemire, Mandal, and Wong (2018) proposed a modified version of the PSO algorithm where they call it the d-QPSO algorithm to obtain locally D-optimal designs for models with mixed quantitative and qualitative factors and a binary response by assuming that all the discrete factors have two levels. Qiu, Chen, Wang, and Wong (2014) searched for optimal designs in nonlinear models and demonstrated by examples the success of the standard PSO in finding optimal designs in life science.

There also are some previous works that involve algorithmic search of optimal designs. For example, Hu, Yang, and Stufken (2015) identified optimal designs for several nonlinear models under general optimal criteria by deriving theoretical results followed by using both Newton's algorithm and OWEA. Stufken and Yang (2012a) considered a basic algorithm based on a grid search to identify locally optimal design for GLMs with group effects. Tan and Stufken (2016) searched for locally D-optimal design using GLMs with factorial effects and one covariate by a modified OWEA to handle the factor effects in the model. Kim and Kao (2019) employed the fmincon solver in MATLAB to find the desired optimal designs for the simple mixed responses model with one covariate.

For the simple mixed responses model, there are several algorithms that we can consider in searching for locally optimal designs such as the PSO algorithm, the ICA, the OWEA, and the fmincon solver in MATLAB. All of those algorithms that we tried tend to provide satisfactory results for some simple cases. In our case, we observed that the use of the fmincon solver was among the fastest methods, and it is very easy to use without much programming effort. Thus, we select the fmincon solver to search for the optimal designs for the simple mixed responses model adopted in Chapter 3.

Although the fmincon solver is fast and effective in finding optimal designs for simple cases, we start to observe its failing to operate well when the design problem becomes complex. We thus propose a simple idea on borrowing the strengths of different algorithms in finding optimal designs. Similarly to several other approaches, the designs that we found can sometimes include clustered support points. Thus, we also consider to cluster similar support points and remove the ones with very small or zero weights for the obtained designs. We note that the number of support points required for the optimal design problem should be determined in advance when applying the fmincon solver. Therefore, applying the complete class results obtained in Chapter 4 is needed. If such a result is unavailable, one may consider, e.g., the 'complete class' given by the Carathéodory theorem, or by simply trying different number of support points. The optimality of the obtained design can be verified by the GET. In theory, the optimal design ξ is verified when the directional derivative $\phi(x, \xi) = 0$ for all the support points of the optimal design. But in numerical computations, this verification is done with a choice of a small cut-off value, ϵ, which also gives information on the efficiency of the attained design in comparison to the true optimal design. The efficiency lower bound for the D- and A-optimal designs can be found in equations (2.16) and (2.17), respectively. In this work, the relative efficiency to be achieved is 99.99% for pragmatic reasons. A higher relative efficiency can be considered. But, in our experience, the slightly improved design efficiency typically requires much more computational resource without giving a significant difference in the resulting designs.

5.2 Numerical Results for Mixed Responses Model With Groups and Multiple Slopes Effects

For Model I, each subject group is allowed to have its own regression functions for the means with no parameters in common. Based on Corollary 4.1.3, we can first search for optimal conditional measure τ_l^* for each group separately, and we adopt the fmincon solver in MATLAB for this. When needed, a parallel computing technique can be considered at this stage. The fmincon solver requires initial values that can effect the search for the optimal design. In this case we set the initial points for each group to $\tau_l^0 = \{c_{l2}, c_{l3}, 0.25, 0.25, 0.25, 0.25\}$, where $c_{l2} < c_{l3}$ are chosen randomly. After the optimal conditional measure τ_l for each group l is validated by the GET, then the marginal weights w_l are computed by equation (4.8). The weights $w_{j|l}$ of the

τ_l^* are adjusted to aggregate these optimal conditional measures to form the optimal deign ξ^* as shown in Example 4.1.2.

As claimed by Yang et al. (2013), the OWE algorithm was proven to defeat existing algorithms in speed by a large scale and in solving wider optimality problems. Therefore, we also consider the OWE algorithm. Both algorithms are efficient in identifying an optimal design for Model I. Thus, we compare the running time in finding A- and D-optimal designs.

In order to compare the OWE algorithm with the fmincon solver in MATLAB, the OWE algorithm code was transferred from SAS to MATLAB and we used a Desktop computer that has a 3.4 GMz Intel Core i7 with 32G RAM for implementing both algorithms. Similar to the fmincon solver, the OWE algorithm requires initial values. We set the initial points for each group to $\tau_l^0 = \{D_{l1}, c_{l2}, c_{l3}, D_{l2}, 0.25, 0.25, 0.25, 0.25\}$, where c_{l2} and c_{l3} are chosen randomly, and $c_{lj} \in [D_{l1}, D_{l2}]$. The minimum efficiency lower bound is also set to 99.99% to terminate the search.

Locally D-optimal designs

With the D-optimality criterion, we have $\Phi_D = -\log|\bigoplus_{l=1}^{s} B_l(\boldsymbol{\theta})\tilde{C}_l(\xi, \boldsymbol{\theta})B_l^T(\boldsymbol{\theta})| = -2\sum_{l=1}^{s} \log|B_l(\boldsymbol{\theta})| - \sum_{l=1}^{s} \log|\tilde{C}_l(\xi, \boldsymbol{\theta})|$, where $\tilde{C}_l(\xi, \boldsymbol{\theta}) = \sum_{j=1}^{m_l} w_l w_{j|l} C_l(\boldsymbol{\theta}, c_{lj})$. This implies that the matrix $B_l(\boldsymbol{\theta})$ does not have any effect on the optimization procedure as long as they are the same across candidate designs. Thus, for the D-criterion we can only concentrate on the minimization of $-\sum_{l=1}^{s} \log|\tilde{C}_l(\xi, \boldsymbol{\theta})|$.

Table 5.1 reports the CPU time in seconds that the fmincon solver and the OWE algorithm need for finding D-optimal designs for selected design spaces (for x). For the OWE algorithm, we need to discretize the design space, and we consider two different grid sizes, namely $N = 5000$ and $N = 10000$. We note that the fmincon function works directly on the entire design space without discretization. We consider model (4.11), and 4 scenarios for each of the design spaces. Scenario 1 has 4 groups with

$\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (-2, -0.2), \boldsymbol{\beta}_{22}^T = (4, -0.8), \boldsymbol{\beta}_{23}^T = (3, 0.4)$, and $\boldsymbol{\beta}_{24}^T = (-1, 0.6))^T$. Scenario 2 has 6 groups with $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (-1, 0.5), \boldsymbol{\beta}_{22}^T = (4, 0.4), \boldsymbol{\beta}_{23}^T = (2, -0.5), \boldsymbol{\beta}_{24}^T = (3, 0.6), \boldsymbol{\beta}_{25}^T = (4.5, 0.8)$, and $\boldsymbol{\beta}_{26}^T = (3.5, 0.4))^T$. Scenario 3 has 8 groups with $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (1.7, 0.4), \boldsymbol{\beta}_{22}^T = (0.6, 0.6), \boldsymbol{\beta}_{23}^T = (1.3, 0.8), \boldsymbol{\beta}_{24}^T = (0.6, 1), \boldsymbol{\beta}_{25}^T = (0.4, 1.2), \boldsymbol{\beta}_{26}^T = (1.1, 1.4), \boldsymbol{\beta}_{27}^T = (0.8, 1.6)$, and $\boldsymbol{\beta}_{28}^T = (0.7, 1.8))^T$. Scenario 4 contains 9 groups with $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (1.7, 0.2), \boldsymbol{\beta}_{22}^T = (0.6, 0.4), \boldsymbol{\beta}_{23}^T = (1.3, 0.6), \boldsymbol{\beta}_{24}^T = (0.6, 0.8), \boldsymbol{\beta}_{25}^T = (0.4, 1), \boldsymbol{\beta}_{26}^T = (1.1, 1.2), \boldsymbol{\beta}_{27}^T = (0.8, 1.4), \boldsymbol{\beta}_{28}^T = (1.5, 1.6)$, and $\boldsymbol{\beta}_{29}^T = (0.7, 1.8))^T$. We also set $\sigma_0^2 = 1$ and $\sigma_1^2 = 1$ in all the 4 scenarios, although this does not have an effect on the selection of D-optimal designs.

Table 5.1: Computation Time (in Seconds) for Locally D-optimal Designs for Model I

	Scenario 1			Scenario 2			Scenario 3			Scenario 4		
Design Space	fmin	OWE 5000	OWE 10000	fmin	OWE 5000	OWE 10000	fmin	OWE 5000	OWE 10000	fmin	OWE 5000	OWE 10000
$[-25, -5]$	0.42	0.81	1.55	0.45	1.48	2.40	0.62	1.49	2.71	0.82	1.78	2.88
$[-20, 0]$	0.33	1.22	2.48	0.42	2.17	3.34	0.48	2.19	3.44	0.51	2.25	3.66
$[-15, 5]$	0.35	1.43	2.41	0.42	2.37	3.87	0.49	4.09	7.32	0.57	4.44	7.75
$[-10, 10]$	0.31	1.51	2.82	0.47	2.64	4.36	0.49	4.13	8.28	0.59	4.45	9.09
$[-5, 15]$	0.25	1.22	2.63	0.46	1.59	2.94	0.47	3.79	7.08	0.55	4.24	7.51
$[0, 20]$	0.31	1.07	1.81	0.40	1.35	2.66	0.41	1.39	2.71	0.55	1.49	2.76
$[5, 25]$	0.37	0.71	1.27	0.45	1.09	2.04	0.74	1.59	2.49	0.87	1.62	2.75

In general, it can be seen from Table 5.1 that as the number of groups increases, the computing time for both algorithms increases as well. In addition, the fmincon solver is clearly faster than the OWE algorithm. We now discuss some of our obtained designs using the following examples.

Example 5.2.1. Let us consider Scenario 1 with a design region $[-5, 15]$ for the continuous x. For this case, the corresponding induced design regions $[D_{l1}, D_{l2}]$ for the four groups are $[-5, -1], [-8, 8], [1, 9]$, and $[-4, 8]$, respectively. As shown in

Table 5.2, the locally executed D-optimal design has 14 support points. When the induced design region is all negative $[-5, -1]$, as in group 1, or all positive $[1, 9]$, as in group 3, then we have 3 support points including the two endpoints of the induced design region. For the other groups whose induced design range contains 0, we have 4 support points including the two endpoints of the induced design region. Furthermore, the conditional measure τ_2 for the symmetric induced design region of Group 2 is a 4-point symmetric design. We also note that the D-criterion value for the design in Table 5.2 is $\Phi_D(\xi^*) = \log|M^{-1}(\xi^*, \boldsymbol{\theta})| = 34.62$. The graph in Figure 5.1 verifies the optimality of the conditional measures obtained for each group. All the sensitivity functions $d_D(x, \xi)$ are bounded above by the straight reference line of $y = 6$, the number of parameters in the group, with equality obtained at the support points.

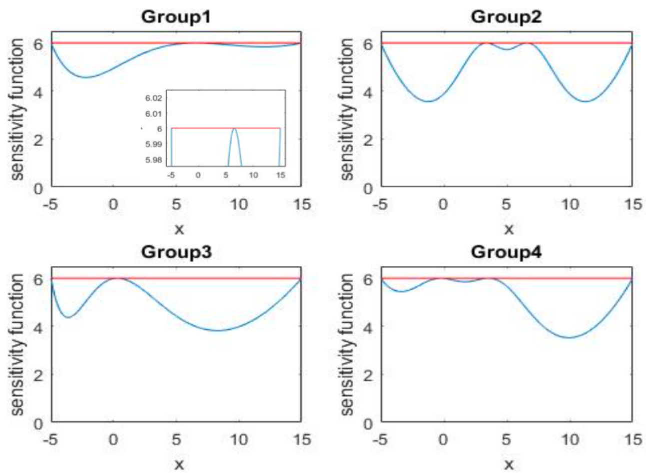
Table 5.2: Locally D-optimal Designs for Model I Using Scenario 1, for $x_{lj} \in [-5, 15]$

Group	Induced design space	Support points(x_{lj})	Support points(c_{lj})	weights(w_{lj})
1	[-5,-1]	(-5,6.557,15)	(-1,-3.3114,-5)	(0.1208,0.072,0.0572)
2	[-8,8]	(-5,3.398,6.602,15)	(8,1.2816,-1.2816,-8)	(0.0408,0.0842,0.0842,0.0408)
3	[1,9]	(-5,0.293,15)	(1,3.1172,9)	(0.11,0.1008,0.0392)
4	[-4,8]	(-5,-0.222,3.546,15)	(-4,-1.1332,1.1276,8)	(0.0539,0.0575,0.0976,0.041)

Scenario1: $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (-2, -0.2), \boldsymbol{\beta}_{22}^T = (4, -0.8), \boldsymbol{\beta}_{23}^T = (3, 0.4)$, and $\boldsymbol{\beta}_{24}^T = (-1, 0.6))^T$

Example 5.2.2. For Scenario 2 with the design region $[-20, 0]$ for x, the induced design regions for the six groups are $[-11, -1]$, $[-4, 4]$,[12, 2], $[-9, 3]$, $[-11.5, 4.5]$, and $[-4.5, 3.5]$, respectively. As shown in Table 5.3, the locally D-optimal design has 22 support points. When the induced design region is all negative $[-11, -1]$, as in Group 1, or all positive $[2, 12]$, as in Group 3, we have 3 support points in that group including the endpoints of the induced design region. For the other groups whose induced design range contains 0 as in Groups 2, 4-6, we have 4 support points including the

Figure 5.1: Locally D-optimal Design Verification for 4-Group Design for $x_{lj} \in [-5, 15]$



two endpoints of the corresponding induced design region. The D-criterion value for this design is $\Phi_D(\xi^*) = \log|M^{-1}(\xi^*, \boldsymbol{\theta})| = 70.7804$. The verification of the optimality of the obtained design is shown in Figure 5.2. The plots show that the conditional measure for each group is D-optimal as claimed.

Table 5.3: Locally D-optimal Designs for Model I Using Scenario 2 for $x_{lj} \in [-20, 0]$

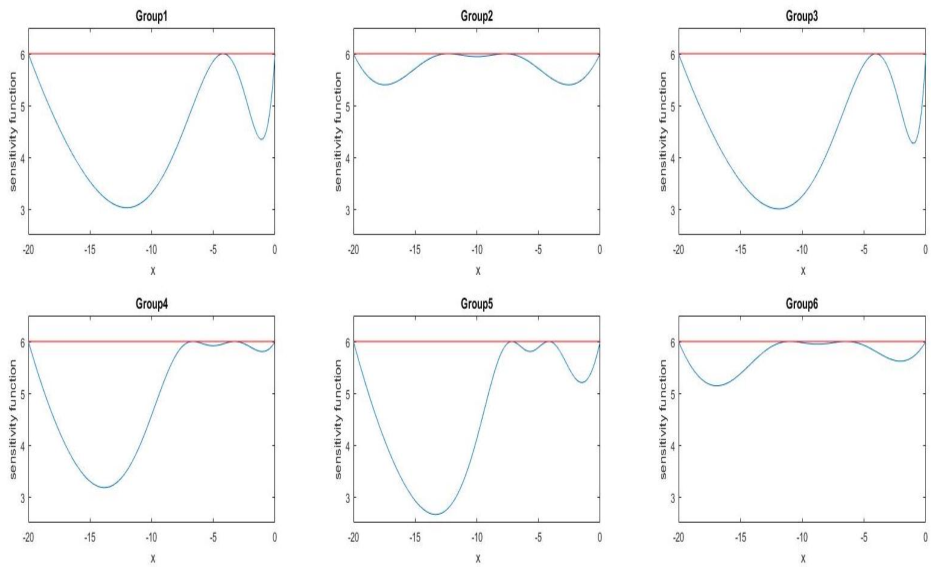
Group	Induced design space	Support points(x_{lj})	Support points(c_{lj})	Weights(w_{lj})
1	[-11,-1]	(-20,-4.208,0)	(-11,-3.104,-1)	(0.264,0.0684,0.0719)
2	[-4,4]	(-20,-12.308,-7.692,0)	(-4,-0.9232,0.9232,4)	(0.0364,0.047,0.047,0.0364)
3	[12,2]	(-20,-4.048,0)	(12,4.024,2)	(0.0264,0.0673,0.073)
4	[-9,3]	(-20,-6.658,-3.273,0)	(-9,-0.9948,1.0362,3)	(0.0273,0.0748,0.0139,0.0506)
5	[-11.5,4.5]	(-20,-7.163,-4.158,0)	(-11.5,-1.2304,1.1736,4.5)	(0.0273,0.0647,0.0423,0.0324)
6	[-4.5,3.5]	(-20,-10.976,-6.439,0)	(-4.5,-0.8904,0.9244,3.5)	(0.0331,0.057,0.0348,0.0417)

Scenario2: $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (-1, 0.5), \boldsymbol{\beta}_{22}^T = (4, 0.4), \boldsymbol{\beta}_{23}^T = (2, -0.5), \boldsymbol{\beta}_{24}^T = (3, 0.6), \boldsymbol{\beta}_{25}^T = (4.5, 0.8)$, and $\boldsymbol{\beta}_{26}^T = (3.5, 0.4))^T$

Lemma 4.2.3 is helpful in calculating the D-optimal symmetric designs. We use two simple cases to illustrate the use of the lemma in the next example.

Example 5.2.3. Two cases of D-optimal symmetric designs are shown in Table 5.4, where the first case is for a one group symmetric design $s = 1$, and the second case is

Figure 5.2: Locally D-optimal Design Verification for 6-Group Design for $x_{lj} \in [-20, 0]$



for a two group symmetric design $s = 2$. Based on Lemma 4.2.3, the values of c_l^* and w_l^* that maximizes the function $\Delta(c_l, w_l)$ can be computed for each symmetric induced design. For the first case with $\boldsymbol{\theta}_2^T = (\boldsymbol{\beta}_{21}^T = (5, 2.5))^T$, the calculated values are $c_1^* = 1.3218$ and $w_1^* = 0.67439$. In the second case with $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (2.5, -0.5), \boldsymbol{\beta}_{22}^T = (-5, 1))^T$ the values for c_l^* and w_l^* for the first group are: $c_1^* = 0.92305$ and $w_1^* = 0.56363$, and for second group are: $c_2^* = 1.2812$ and $w_2^* = 0.67393$.

Table 5.4: Locally D-optimal Designs for c_l^*, w_l^* that Maximizes the Function $\Delta(c_l, w_l)$ in Lemma 4.2.3

Group	Design region (x)	Induced design region (c_l)	Support points(x_{lj})	Support points(c_{lj})	Weights(w_{lj})
1	[-6,2]	[-10,10]	(-6,-2.528,-1.471,2)	(-10,-1.3218,1.3218,10)	(0.1628,0.3372,0.3372,0.1628)
1	[-3,13]	[-4,4]	(-3,3.154,6.846,13)	(4,0.923,-0.923,-4)	(0.1091,0.1409,0.1409,0.1091)
2		[-8,8]	(-3,3.719,6.281,13)	(-8,-1.281,1.281,8)	(0.0815,0.1685,0.1685,0.0815)

Locally A-optimal designs

If the interest is on determining A-optimal designs, then the value of the model parameters is going to affect the solutions due to the form of A-optimal criterion.

To show this, recall that the A-optimal criterion is $\Phi_A = \text{tr}(M^{-1})$, then with the decomposed information matrix in (4.12) the A-optimal criterion becomes $\Phi_A = \sum_{l=1}^{s} \text{tr}(B_l^{-1}(\boldsymbol{\theta}) B_l^{T^{-1}}(\boldsymbol{\theta}) \tilde{C}_l^{-1}(\xi, \boldsymbol{\theta}))$. The matrix $B_l(\boldsymbol{\theta})$ is inseparable in the A-optimality criterion. It also is noteworthy that both σ_0^2 and σ_1^2 play a role in the selection of the locally A-optimal design.

Again, we compare the CPU time needed for obtaining optimal designs between the fmincon solver and the OWE algorithm. As shown in Table 5.5 we used the same previously considered scenarios with $\sigma_0^2 = \sigma_1^2 = 1$.

Table 5.5: Computation Time (in Second) for Locally A-optimal Designs for Model I

	Scenario1			Scenario2			Scenario3			Scenario4		
Design Space	fmin	OWE 5000	OWE 10000	fmin	OWE 5000	OWE 10000	fmin	OWE 5000	OWE 10000	fmin	OWE 5000	OWE 10000
$[-25, -5]$	0.35	0.98	1.51	0.46	1.97	2.78	1.11	8.01	16.51	1.39	8.53	16.85
$[-20, 0]$	0.34	1.42	2.33	0.44	2.97	5.44	0.69	3.02	5.94	0.95	3.26	6.18
$[-15, 5]$	0.56	3.60	4.97	0.69	3.29	6.39	0.83	12.40	13.86	1.37	16.47	19.74
$[-10, 10]$	0.36	15.07	16.16	0.61	10.35	18.02	0.62	14.88	26.31	0.72	19.09	39.51
$[-5, 15]$	0.45	2.64	6.93	0.49	10.10	15.44	1.13	10.54	17.15	1.23	15.50	26.69
$[0, 20]$	0.31	1.13	2.25	0.38	1.44	3.18	0.77	2.85	5.91	0.93	3.38	5.98
$[5, 25]$	0.28	0.86	1.29	0.46	1.22	1.57	0.63	1.38	3.18	0.69	1.44	3.46

It can be seen from Table 5.5 that the fmincon solver is at least two times faster than the OWE algorithms when the grid size is $N = 5000$, and at least three times faster when the grid size is $N = 10000$ in all the scenarios.

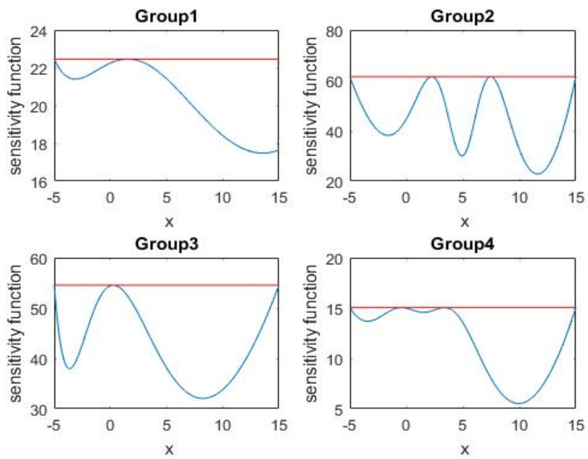
Example 5.2.4. By considering model (4.11) and using Scenario 1 with design region $[-5, 15]$ for x. The corresponding induced design regions for c for the four groups are $[-5, -1], [-8, 8], [1, 9]$, and $[-4, 8]$, respectively. As shown in Table 5.6, the locally executed A-optimal design has 13 support points. The A-criterion value for this design is $\Phi_A = \text{tr}(M^{-1}) = 445.4688$. The optimality of the design is verified by the

Table 5.6: Locally A-optimal Designs for Model I Using Scenario 1, for $x_{lj} \in [-5, 15]$

Group	Induced design space	Support points(x_{lj})	Support points(c_{lj})	weights(w_{lj})
1	[-5,-1]	(-5,1.547)	(-1,-2.3094)	(0.0376,0.1611)
2	[-8,8]	(-5,2.2486,7.4573,15)	(8,2.2011,-1.9658,-8)	(0.0118,0.2305,0.0673,0.0192)
3	[1,9]	(-5,0.2624,15)	(1,3.105,9)	(0.0256,0.2835,0.0008)
4	[-4,8]	(-5,-0.4270,3.3684,15)	(-4,-1.2562,1.021,8)	(0.0055,0.1163,0.0371,0.0039)

Scenario1: $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (-2, -0.2), \boldsymbol{\beta}_{22}^T = (4, -0.8), \boldsymbol{\beta}_{23}^T = (3, 0.4)$, and $\boldsymbol{\beta}_{24}^T = (-1, 0.6))^T$

Figure 5.3: Locally A-optimal Design Verification for 4-Group Design for $x_{lj} \in [-5, 15]$



GET as shown in Figure 5.3.

The effects of the variance on the A-optimal designs

Now let us examine the effects of the values of σ_0^2 and σ_1^2 on the optimal designs. For demonstration purposes, we will consider Model I, and the design region for x is $[-6, 5]$. We assume that the study contains $s = 2$ groups and the guessed values of the parameters are set to $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_{21}^T = (4, 3), \boldsymbol{\beta}_{22}^T = (1, 2))^T$. The corresponding induced design regions for c are $[-14, 19]$ and $[-11, 11]$ for groups 1 and 2, respectively. In Tables 5.7 and Table 5.8 we consider all the combinations of the values of σ_0^2 and σ_1^2 among 0.1,0.5,1,10, and 100. Here, we divide the results into two tables for the ease of presentation. Table 5.7 has the results for the group with the asymmetric induced

design region $[-14, 19]$ while Table 5.8 is for the group with the symmetric induced design region $[-11, 11]$. We report the support points and corresponding weights of the obtained A-optimal designs in these two tables. We also include the ratio,R, of the total weight of the outer-points to that of the two inner support points. We discuss our finding based on the following three scenarios:
(i) σ_1^2 is fixed but σ_0^2 is increased. (ii) σ_0^2 is fixed but σ_1^2 is increased. (iii) $\sigma^2 = \sigma_0^2 = \sigma_1^2$ and σ^2 increases.

For the group with the asymmetric induced design region we have the following observations for each scenario:
Under scenario (i): support points c_{11} and c_{14} have the same values of the endpoints of the induced design region, and the inner support points c_{12} and c_{13} tend to move towards the center of the design region. Regarding the corresponding weights, we get w_{11} increases, w_{12} decreases and then increases, w_{13} and w_{14} decrease, and the weight ratio R increases in value.
Under scenario (ii): support points c_{11} and c_{14} have the same values of the endpoints of the induced design region, c_{12} gets closer to zero, and c_{13} moves away from zero. Regarding the values of the corresponding weights, we find w_{11} and w_{12} decrease, w_{13} decreases and then increases, w_{14} increases, and the weight ratio R also increases in value as well.
Under scenario (iii): support points c_{11} and c_{14} have the same values of the induced design region, and c_{12} and c_{13} get closer to zero. Regarding the values of the corresponding weights, we observe that w_{11} and w_{14} increase. However, w_{12} decreases and then increases, w_{13} decreases, and the weight ratio R increases in value.

For the group with the symmetric induced design region, the obtained A-optimal designs might not be symmetric designs; this is in contrast to the D-optimal designs results. In addition, we have the following observations for each scenario:

Table 5.7: Locally A-optimal Designs for Model I for $c_{1j} \in [-14, 19]$

σ_1^2	σ_0^2	*Group*	c_{11}	c_{12}	c_{13}	c_{14}	w_{11}	w_{12}	w_{13}	w_{14}	R
0.1	0.1	1	-14	-2.228	2.251	19	0.0076	0.2068	0.418	0.004	0.01857
0.1	0.5	1	-14	-2.189	2.212	19	0.0209	0.02042	0.4044	0.0038	0.04058
0.1	1	1	-14	-2.165	2.17	19	0.0298	0.2051	0.3913	0.0036	0.056
0.1	10	1	-14	-2.024	1.759	19	0.0712	0.2338	0.2943	0.0025	0.13956
0.1	100	1	-14	-1.658	1.087	19	0.1021	0.2885	0.1792	0.0011	0.22065
0.5	0.1	1	-14	-2.222	2.26	19	0.0074	0.2015	0.4124	0.0103	0.02883
0.5	0.5	1	-14	-2.18	2.221	19	0.0204	0.1992	0.3994	0.0098	0.05045
0.5	1	1	-14	-2.159	2.179	19	0.0291	0.2004	0.3868	0.0095	0.06574
0.5	10	1	-14	-2.018	1.765	19	0.0701	0.2304	0.2927	0.0068	0.14701
0.5	100	1	-14	-1.655	1.09	19	0.1015	0.2868	0.179	0.0033	0.22499
1	0.1	1	-14	-2.213	2.266	19	0.0072	0.1967	0.408	0.0147	0.03622
1	0.5	1	-14	-2.174	2.23	19	0.02	0.1948	0.3955	0.0142	0.05794
1	1	1	-14	-2.15	2.185	19	0.0285	0.1961	0.3833	0.0137	0.07283
1	10	1	-14	-2.009	1.771	19	0.0692	0.2274	0.2917	0.01	0.15257
1	100	1	-14	-1.649	1.093	19	0.101	0.2855	0.179	0.0049	0.22799
10	0.1	1	-14	-2.09	2.353	19	0.0055	0.1539	0.3788	0.0396	0.08466
10	0.5	1	-14	-2.051	2.314	19	0.0156	0.1543	0.3703	0.0385	0.101313
10	1	1	-14	-2.027	2.266	19	0.0226	0.1572	0.3618	0.0376	0.11599
10	10	1	-14	-1.898	1.834	19	0.0592	0.1978	0.2906	0.0299	0.18243
10	100	1	-14	-1.586	1.126	19	0.0951	0.2722	0.1829	0.0163	0.24478
100	0.1	1	-14	-1.568	2.605	19	0.0021	0.075	0.3537	0.0651	0.15675
100	0.5	1	-14	-1.532	2.563	19	0.0067	0.0768	0.3504	00645	0.16667
100	1	1	-14	-1.511	2.515	19	0.01	0.0796	0.347	0.064	0.17346
100	10	1	-14	-1.403	2.053	19	0.0315	0.117	0.3116	0.0591	0.21139
100	100	1	-14	-1.241	1.261	19	0.0704	0.2169	0.2162	0.0422	0.25999

Table 5.8: Locally A-optimal Designs for Model I for $c_{2j} \in [-11, 11]$

σ_1^2	σ_0^2	*Group*	c_{21}	c_{22}	c_{23}	c_{24}	w_{21}	w_{22}	w_{23}	w_{24}	R
0.1	0.1	2	-11	-1.744	1.762	11	0.0035	0.1373	0.22	0.0028	0.01763
0.1	0.5	2	-11	-1.704	1.73	11	0.0109	0.1371	0.2162	0.0026	0.03821
0.1	1	2	-11	-1.686	1.688	11	0.0159	0.1392	0.2128	0.0024	0.05199
0.1	10	2	-11	-1.548	1.3	11	0.0391	0.1716	0.1864	0.0012	0.11257
0.1	100	2	-11	-1.206	0.786	11	0.0574	0.2384	0.1331	0.0003	0.15532
0.5	0.1	2	-11	-1.722	1.754	11	0.0033	0.1359	0.2203	0.0089	0.03425
0.5	0.5	2	-11	-1.686	1.722	11	0.0106	0.1357	0.2165	0.0084	0.05395
0.5	1	2	-11	-1.666	1.682	11	0.0154	0.1377	0.2131	0.008	0.0667
0.5	10	2	-11	-1.534	1.302	11	0.0383	0.1697	0.1869	0.005	0.12142
0.5	100	2	-11	-1.2	0.79	11	0.057	0.2371	0.1332	0.0021	0.1596
1	0.1	2	-11	-1.696	1.75	11	0.0032	0.1346	0.222	0.0133	0.04624
1	0.5	2	-11	-1.66	1.72	11	0.0102	0.1345	0.2183	0.0126	0.06463
1	1	2	-11	-1.642	1.68	11	0.0149	0.1364	0.215	0.012	0.07655
1	10	2	-11	-1.516	1.304	11	0.0376	0.1681	0.1882	0.0077	0.12714
1	100	2	-11	-1.194	0.794	11	0.0566	0.2361	0.1336	0.0034	0.16229
10	0.1	2	-11	-1.386	1.728	11	0.0019	0.125	0.257	0.0382	0.1049
10	0.5	2	-11	-1.362	1.704	11	0.0071	0.125	0.2524	0.0368	0.11632
10	1	2	-11	-1.35	1.672	11	0.0107	0.1265	0.2482	0.0354	0.12303
10	10	2	-11	-1.292	1.338	11	0.0302	0.1536	0.2135	0.0252	0.15091
10	100	2	-11	-1.104	0.822	11	0.0522	0.2263	0.1425	0.0125	0.17543
100	0.1	2	-11	-0.822	1.534	11	0.0005	0.0988	0.3397	0.0651	0.1496
100	0.5	2	-11	-0.814	1.526	11	0.0028	0.0987	0.3361	0.0641	0.15386
100	1	2	-11	-0.808	1.516	11	0.0044	0.099	0.3328	0.0632	0.15655
100	10	2	-11	-0.784	1.356	11	0.0152	0.111	0.3007	0.0539	0.16784
100	100	2	-11	-0.746	0.896	11	0.0363	0.1808	0.2036	0.0335	0.18158

Under scenario (i): the outer support points c_{21} and c_{24} remain to be the endpoints of the induced design region, and the inner two support points c_{22} and c_{23} get closer to zero. We also observe that w_{21} increases, whereas w_{24} decreases, and R increases. Under scenario (ii): the outer support points c_{21} and c_{24} are still the endpoints of the induced design region, the smaller inner support points c_{22} gets closer to zero, but the larger inner points c_{23} gets closer to zero for small σ_1^2 and moves away from zero for large σ_1^2. The weights w_{21} decreases, but w_{24} increases, and R also increases. Under scenario (iii): c_{21} and c_{24} remain the same, and c_{22} and c_{23} get closer to zero. Regarding the values of the corresponding weights, we observe that both w_{21} and w_{24} increase, and the weight ratio R increases in value.

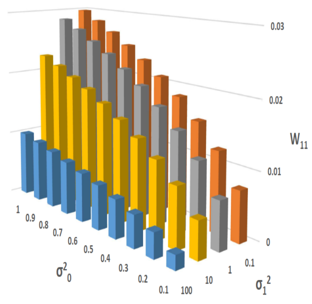
In summary, c_{l1} and c_{l4} in all the cases have the same values of the endpoints of the induced design region. w_{l1} is directly proportional to σ_0^2 and w_{l4} is inversely proportional to σ_0^2. Meanwhile, w_{l4} is directly proportional to σ_1^2 and w_{l1} is inversely proportional to σ_1^2. Notice that $\sigma_0^2 = var(y|z=0)$ goes with $\mathrm{P}(z=0) = 1-\mathrm{P}(z=1)$, where $1-\mathrm{P}(z=1)$ is maximum within the induced design region at its left endpoint. This left endpoint is the support point c_{l1} with corresponding weight w_{l1}. On the other hand, $1-\mathrm{P}(z=1)$ is minimum within the induced design region at its right endpoint, which is the support point c_{l4} with corresponding weight w_{l4}. As previously mentioned the value of the weight w_{l1} is directly proportional to σ_0^2, while the value of the weight w_{l4} is inversely proportional to σ_0^2. This implies that more observations are needed for the situation where the variance is larger.

Considering the case where $\sigma_0^2 = \sigma_1^2 = \sigma^2$ and allowing σ^2 to increase. We see that more weights are assigned to the boundary points to account for the inflated variance of the continuous response. These results coincide with the finding of Kim (2017) when she tested the effects of the inflated variance of the linear models σ^2 under the simple mixed responses model on the locally A-optimal designs.

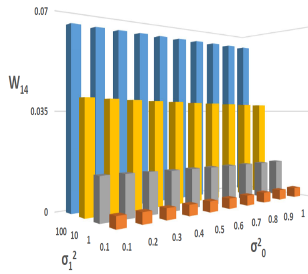
Figure 5.4 summarizes the trend of the obtained optimal designs according to the values of σ_0^2 and σ_1^2 for the group with the asymmetric induced design region. We set three intervals of σ_0^2 and develop ten equally spaced values for each interval of σ_0^2 while σ_1^2 is fixed to one of the values among (0.1, 1, 10, 100). For each (σ_0^2, σ_1^2), we find the A-optimal design under Model I and verify it with the GET. The figures in the first column (a), (d), and (g) show the variation of w_{11} with respect to σ_0^2 at each fixed σ_1^2. The figures in the second column (b), (e), and (h) show the variation of w_{14} with respect to σ_0^2 at each fixed σ_1^2. The figures in the third column (c), (f), and (i) show the variation of the ratio R with respect to σ_0^2 at each fixed σ_1^2.

Next, we will consider searching for the optimal designs under Model II as in model (4.17). This model has a more complex information matrix than Model I. Consequently, this requires an efficient strategy to search for the optimal designs. In the next subsection, we will introduce a combined algorithm that is effective for tackling this challenge.

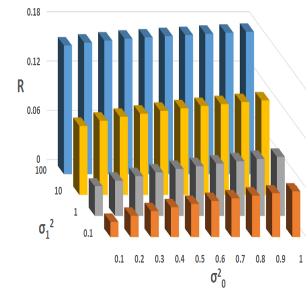
Figure 5.4: Locally A-optimal Designs by Varying σ_0^2 for Fixed σ_1^2



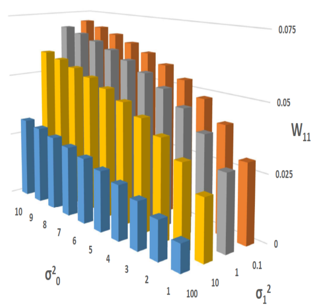
(a) w_{11} vs. $\sigma_0^2 \in [0.1, 1]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



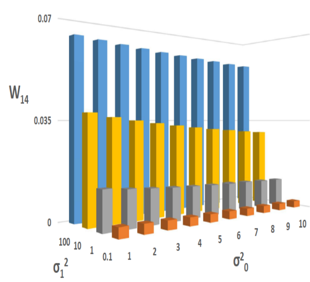
(b) w_{14} vs. $\sigma_0^2 \in [0.1, 1]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



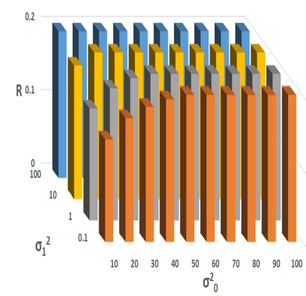
(c) R vs. $\sigma_0^2 \in [0.1, 1]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



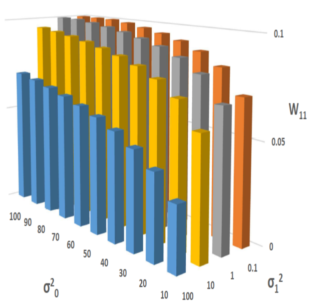
(d) w_{11} vs. $\sigma_0^2 \in [1, 10]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



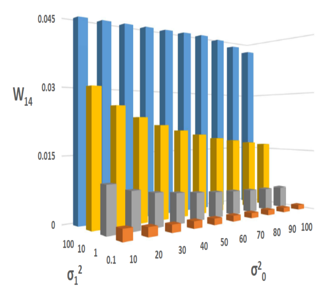
(e) w_{14} vs. $\sigma_0^2 \in [1, 10]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



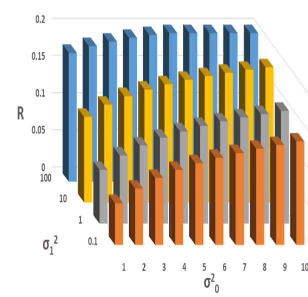
(f) R vs. $\sigma_0^2 \in [1, 10]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



(g) w_{11} vs. $\sigma_0^2 \in [10, 100]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



(h) w_{14} vs. $\sigma_0^2 \in [10, 100]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$



(i) R vs. $\sigma_0^2 \in [10, 100]$ for fixed $\sigma_1^2 = [0.1, 1, 10, 100]$

5.3 Hybrid PSO-FM Algorithm

Although the fmincon solver gave promising optimal designs, it needs to be fed with initial designs before it can search for the optimal design. The fmincon solver has a deficiency of repition by using different initial starting points every time. We observe that, some imprudently selected initials may cause the divergence of the algorithm away from the optimal design or may require more time to converge. This problem frequently emerged in searching for designs for model II with a large number of groups for which we need to search for the optimal design by considering all the groups simultaneously.

A possible remedy that we propose is to combine the fmincon solver with PSO to unite the computational advantages of the two methods. This combination tends to improve the search for optimal designs further. We call this combined computational method the "Hybrid PSO-FM". The PSO algorithm is a good candidate, since it is easy to implement, has few tuning parameters, is fast in exploring the solution space, and can handle complex problems. However, in our case of the mixed responses model, the PSO alone tends to be slow in pinpointing the design that is optimal, especially when the number of groups increases.

The idea of the PSO was inspired by animal behavior such as a flock of birds for food hunting. Each member of the flock is considered as a particle that spreads randomly with its own velocity in the search space in order to determine the food location. At the same time, the flying birds exchange information constantly on food tracking, and there is a tendency for the birds to go after the location that is closer to food. The search for food is affected by both the individual experience of the particle (bird) and the shared knowledge of the swarm (flock). At the beginning of the search, the particles do not know the position they are aiming to reach, which is known as the global best position ($Gpos$). A number of particles, n, is initially created, and these particles are candidate designs for the optimal design problem. The particle is represented by a vector consisting of design points and weights. For our problem, the complete class results obtained in Theorem 4.3.1 allow us to set the number of design points in each group to four with two of them being the endpoints; for generating an initial design for the PSO, the free design points and weights are randomly selected. Then the fitness value of each particle is calculated by the objective function and the algorithm also keeps track of the personal best position ($Bpos$) for each particle. At each iteration, the location of the particle, its velocity, and its fitness value are updated based on its present $Bpos$ and $Gpos$ determined by the whole flock. The two equations that guide the movement of the ith particle in the PSO at each iteration is given by:

$$z_i^{t+1} = z_i^t + v_i^{t+1}, \tag{5.1}$$

where the z_i^t is the position of the particle i at iteration t, and the v_i^{t+1} is the velocity of the particle i at iteration $t+1$ that can be found by:

$$v_i^{t+1} = \delta_t v_i^t + \varphi_1 U_1(0,1) \otimes (Bpos_i^t - z_i^t) + \varphi_2 U_2(0,1) \otimes (Gpos^t - z_i^t). \tag{5.2}$$

Here, $Bpos_i^t$ is the personal best position that the ith particle possesses at iteration t, $Gpos^t$ is the global best position at the t iteration, $\mathrm{U}_j(0,1)$ is a random vector generated from uniform distribution $j = 1, 2$, and $\otimes$ is the Hadamard product. Any particle that moves out of the design space will be dragged back to the specified boundary. The constants φ_1 and φ_2 are the cognitive learning factor and the social learning factor, respectively, to specify how each particle proceeds in the direction of its own personal best position and on the way to the global best position. We allow φ_1 and φ_2 to differ at each iteration following the recommendation of Ratnaweera, Halgamuge, and Watson (2004) where φ_1 decreases from 2.5 to 0.5 and φ_2 increases from 0.5 to 2.5. δ_t is the inertia factor that symbolizes the impact of the prior velocity and can be set to a constant value between $[0, 1]$. In our case, we follow Eberhart and Shi (2000) who set it to be a decreasing linear function from 0.9 to 0.4.

The steps for the PSO algorithm can be summarized as follows:

i) Initialize the particle positions $(z_1^0, \cdots, z_n^0)$ and the velocities $(v_1^0, \cdots, v_n^0)$.

ii) Compute the fitness values $\Phi(z_1^0, \cdots, z_n^0)$.

iii) Determine the global best position $Gpos^0$.

iv) Calculate the velocities $(v_1^{t+1}, \cdots, v_n^{t+1})$ using equation (5.2).

v) Update particle positions $(z_1^{t+1}, \cdots, z_n^{t+1})$ using equation (5.1).

vi) Compute the fitness values $\Phi(z_1^{t+1}, \cdots, z_n^{t+1})$.

vii) Determine the personal best position $Bpos_i^{t+1}$ and the global best position $Gpos^{t+1}$.

Steps (iv)-(vii) are repeated until the stopping rule, the maximum number of iterations set by the user, is reached. The particle whose fitness value equals min $\Phi(Gpos^{t+1})$ is chosen as the output of the algorithm. The algorithms can itself be used to search for optimal designs. However, we tend to observe that it can quickly reach a good design, and can then take a long time to converge to the optimal solution. We thus only implement this algorithm for a small number of iterations, use its output as the initial designs for the fmincon solver, and let the fmincon solver finish the job for identifying the optimal design. Finally, the obtained designs by the hybrid PSO-FM algorithm are verified by GET. For Model II, we use the interior-point method (IPM) in the fmincon solver instead of the SQP that is used in the simple mixed responses model. Due to the complicated information matrix in model II, the IPM has an advantage over the SQP in which that the Hessian matrix and the gradient are computed once per optimization iteration rather than updated more than once at each time the active set is modified during the inner iteration in the SQP. Therefore, we can conclude that as the information matrix becomes more complicated with the presence of groups in the model the IPM can be more competitive than the SQP algorithm.

The proposed hybrid PSO-FM algorithm can be described in steps as follows:

i) Start with the PSO algorithm to get the initial design.

ii) Feed the fmincon solver with the design obtained from PSO.

iii) Optimize the inner design points and weights for all the groups using fmincon.

iv) Check the optimality of the design reached using the GET.

v) If the selected design did not pass the GET repeat (i)-(iv).

vi) Cluster similar support points and remove design points with very small or zero weights from the optimal design.

viii) Verify the optimal design by the GET plots.

In the case of the fmincon solver, steps (i) and (ii) above are replaced by selecting random initial two points and weights for each group.

5.4 Numerical Results for Mixed Responses Model With Common Parameters

Now we study the performance of the proposed hybrid PSO-FM algorithm in searching for locally A- and D-optimal designs for Model II in (4.17). When using the hybrid PSO-FM algorithm, the two tuning parameters that need to be determined are the number of particles (flock size) and the number of iterations for the PSO when finding initial designs for the fmincon solver. Based on the recommendation of Qiu et al. (2014) on minimal flock size and the number of iterations, in what they called the "90% rule", we set the flock size to 20 and the number of iterations to 100. In our overall experience, a small flock size and number of iterations, around 20 and 100, gives reasonable initial design points for the fmincon solver to employ. Increasing the maximum number of iterations over 100 and flock size over 20 costs time which is not worthwhile in our case since we only need reasonable initial design points.

Again, we tried to use the OWE algorithm for the same tasks. Unfortunately, the OWE algorithm did not work for all our cases. First, Newton's method is used in the OWE algorithm to update the weights, which requires the calculation of the second derivatives of the $\Phi_p\{M^{-1}(\xi)\}$ with respect to $\mathbf{w}$. This can be computationally cumbersome for some models, e.g. Model II case. Second, the information matrix may be singular which causes Newton's method to produce negative weights during the iteration (Wong, Yin, & Zhou, 2018). Third, the algorithm has a high probability of being stuck in an infinite loop. For example, when running the algorithm, the selected design might not satisfy the efficiency lower bound. In this case, the algorithm adds one new point x^0 to the existing design that maximizes the sensitivity function. This is followed by the calculation for all the weights for the new design with the new point included. But Newton's algorithm assigns a very small weight to the new point which leads the algorithm to delete this point and, as a result, the same previous design is obtained. This process continues by adding and deleting the same point x^0 again and thus the algorithm will be stuck between two designs. To overcome this obstacle, Newton's method for updating the weights in the OWE algorithm can be replaced by other methods such as the multiplicative algorithm (see also Tan, 2015). The modification that needs to be done for the OWE algorithm to work for our case is beyond the scope of this work.

In addition to the OWE algorithm, we also tried some other algorithms, such as the d-QPSO of Lukemire et al. (2018), and the PSO alone (without the fmincon solver). However, it does not seem straightforward to make these algorithms work efficiently for our problems, and for several scenarios, no satisfactory results are achieved by these algorithms after a long computing time. Thus, we choose to demonstrate the performance of the hybrid PSO-FM algorithm by comparing it with the fmincon solver alone (without the PSO) in order to examine the effect of the PSO-generated initials in speeding up the fmincon solver search.

Locally D-optimal designs

We consider four scenarios using Model II for each of the design spaces. Scenario 1 has 6 groups with $\boldsymbol{\theta}_2 = (\alpha_{21} = 1.7, \alpha_{22} = 0.6, \alpha_{23} = 0.4, \alpha_{24} = 1.1, \alpha_{25} = 0.8, \alpha_{26} = 1.3, \beta_2 = 1)^T$, Scenario 2 has 8 groups with $\boldsymbol{\theta}_2 = (\alpha_{21} = 1.7, \alpha_{22} = 0.6, \alpha_{23} = 1.3, \alpha_{24} = 0.2, \alpha_{25} = 0.8, \alpha_{26} = 1.7, \alpha_{27} = 0.4, \alpha_{28} = 0.7, \beta_2 = 1)^T$, Scenario 3 has 9 groups with $\boldsymbol{\theta}_2 = (\alpha_{21} = 1.7, \alpha_{22} = 0.6, \alpha_{23} = 1.3, \alpha_{24} = 0.6, \alpha_{25} = 0.4, \alpha_{26} = 1.1, \alpha_{27} = 0.8, \alpha_{28} = 1.5, \alpha_{29} = 0.7, \beta_2 = 1)^T$, and Scenario 4 has 12 groups with $\boldsymbol{\theta}_2 = (\alpha_{21} = 1.7, \alpha_{22} = 0.6, \alpha_{23} = 1.3, \alpha_{24} = 0.2, \alpha_{25} = 0.8, \alpha_{26} = 1.6, \alpha_{27} = 0.4, \alpha_{28} = 0.7, \alpha_{29} = 0.3, \alpha_{210} = -0.4, \alpha_{211} = -0.3, \alpha_{212} = -0.1, \beta_2 = 1)^T$. In all the 4 scenarios, we set $\sigma_0^2 = \sigma_1^2 = 1$.

Due to the randomness nature of both the fmincon solver and the PSO algorithm, the computation time is checked five times for each design space under all scenarios for both the hybrid PSO-FM algorithm and the fmincon solver. The minimum and maximum computation times are recorded in Table 5.9. Note that we excluded the recorded times for the fmincon solver when it gets stuck in a local minimum. This was observed in cases where the model has a large number of groups. We can observe that in all the cases the maximum computation time for the hybrid PSO-FM algorithm is faster than the fmincon solver under all scenarios. The minimum computation time using the fmincon solver might be faster than that of the hybrid PSO-FM algorithm. This is more likely to occur for models with a small number of groups. It is possible that the random initial points used by the fmincon solver were very close to the optimal design. As the number of groups increases, the variation between the minimum and maximum computation times increases using the fmincon solver. We can conclude that as the number of groups increases, the model becomes more dependable on

Table 5.9: Computation Time (in Second) for Locally D-optimal Designs for Model II

Algorithm	Design Space(x)	Scenario1		Scenario2		Scenario3		Scenario4	
		min	max	min	max	min	max	min	max
PSO-FM	$[-25,-5]$	12.12	15.84	13.61	17.21	17.05	25.77	26.6	34.99
fmincon		12.53	17.93	15.81	83.32	20.71	56.55	35.81	296.11
PSO-FM	$[-20,0]$	9.85	11.27	15.22	18.01	18.3	23.47	31.97	51.68
fmincon		8.02	37.38	21.49	82.81	22.42	60.36	58.12	261.62
PSO-FM	$[-15,5]$	10.34	11.91	16.59	20.9	20.73	24.97	50.15	68.17
fmincon		17.59	40.91	15.95	66.31	25.01	68.45	56.98	248.41
PSO-FM	$[-10,10]$	10.15	10.76	15.01	19.02	19.99	24.55	41.88	51.74
fmincon		12.64	22.99	17.12	70.64	28.61	82.15	46.61	135.27
PSO-FM	$[-5,15]$	12.47	18.07	18.94	22.21	24.18	27.41	58.84	86.68
fmincon		9.91	61.81	22.45	111.47	28.92	123.49	65.9	323.77
PSO-FM	$[0,20]$	13.31	17.57	18.32	23.63	23.19	35.54	47.77	91.66
fmincon		10.72	38.82	21.15	80.38	28.78	127.71	89.31	598.7
PSO-FM	$[5,25]$	9.74	14.74	16.13	18.4	19.91	25.95	42.69	58.05
fmincon		15.15	40.48	19.48	89.15	28.71	148.36	71.82	519.73

the initial points, and thus the hybrid PSO-FM algorithm becomes faster and more stable than the fmincon solver.

The complete class found in theorem 4.3.1 on the maximum number of support points for this model can be illustrated in the following examples.

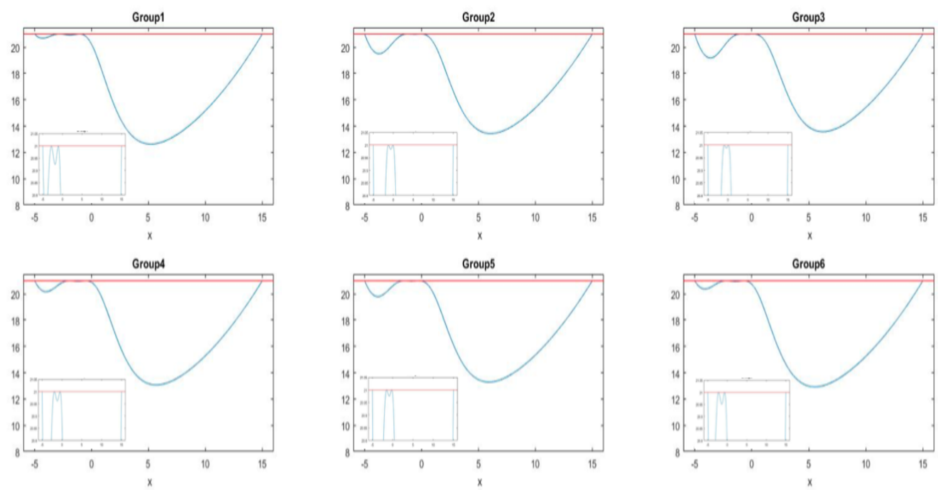
Example 5.4.1. Let us consider Scenario 1 with design region $[-5, 15]$ for the continuous variable x. The corresponding induced design regions $[D_{l1}, D_{l2}]$ for the six groups are $[-3.3, 16.7], [-4.4, 15.6], [-4.6, 15.4], [-3.9, 16.1], [-4.2, 15.8]$, and $[-3.7, 16.3]$, respectively. As it can be seen from Table 5.10, the locally D-optimal design has four support points in each group including the endpoints of the design region. The D-criterion value for this design is $\Phi_D(\xi^*) = \log|M^{-1}(\xi^*, \boldsymbol{\theta})| = 49.7098$. The GET plot conforms the D-optimality of the selected design shown in Figure 5.5.

Table 5.10: Locally D-optimal Designs for Model II Using Scenario 1, for $x_{lj} \in [-5, 15]$

Group	Induced design space	Support points(x_{lj})	Support points(c_{lj})	weights(w_{lj})
1	[-3.3,16.7]	(-5,-2.814,-1.050,15)	(-3.3,-1.114,0.650,16.7)	(0.012,0.053,0.086,0.017)
2	[-4.4,15.6]	(-5,-1.2,-0.060,15)	(-4.4,-0.6,0.54,15.6)	(0.023,0.045,0.084,0.014)
3	[-4.6,15.4]	(-5,-0.912,0.113,15)	(-4.6,-0.512,0.513,15.4)	(0.025,0.045,0.084,0.013)
4	[-3.9,16.1]	(-5,-1.927,-0.502,15)	(-3.9,-0.827,0.598,16.1)	(0.019,0.048,0.085,0.015)
5	[-4.2,15.8]	(-5,-1.49,-0.235,15)	(-4.2,-0.690,0.565,15.8)	(0.021,0.046,0.085,0.014)
6	[-3.7,16.3]	(-5,-2.221,-0.683,15)	(-3.7,-0.921,0.617,16.3)	(0.017,0.049,0.085,0.016)

Scenario1: $\boldsymbol{\theta}_2 = (\alpha_{21} = 1.7, \alpha_{22} = 0.6, \alpha_{23} = 0.4, \alpha_{24} = 1.1, \alpha_{25} = 0.8, \alpha_{26} = 1.3, \beta_2 = 1)^T$

Figure 5.5: Locally D-optimal Design Verification for 6-Group Design for $x_{lj} \in [-5, 15]$



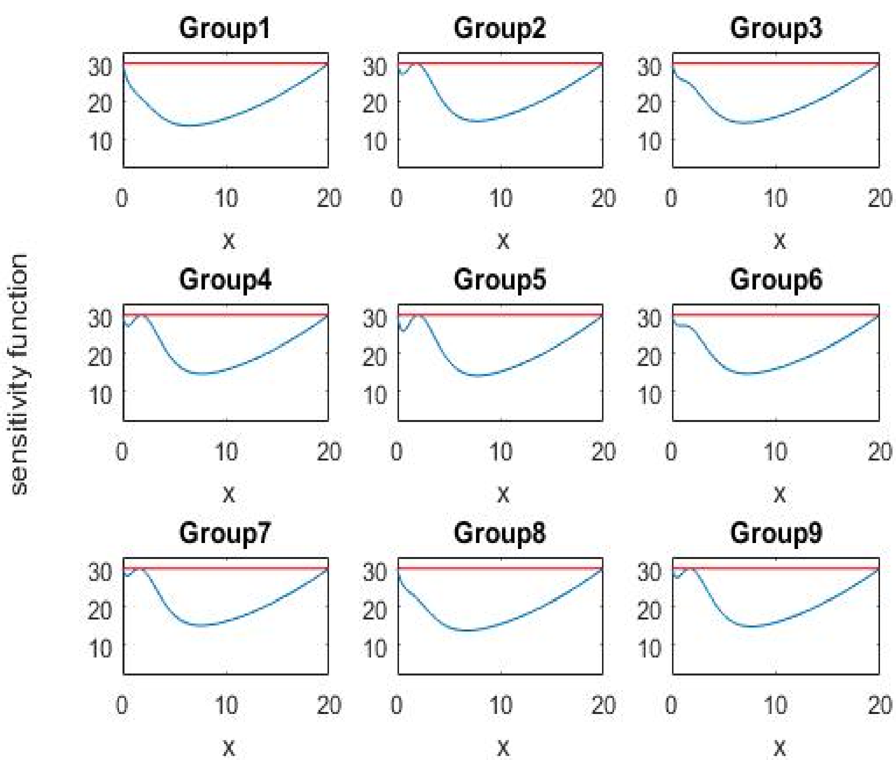
Example 5.4.2. Consider Scenario 3, and the design region for the continuous variable x is $[0, 20]$. The locally D-optimal design found contains 23 support points, and from Table 5.11 we observe that the number of support points ranged from 2 to 3 points in each group. When there are 2 support points in the group, the two endpoints of the design region are included. The 3 support points appeared in groups 2,4,5,7, and 9. The D-criterion value for this design is $\Phi_D(\xi^*) = \log|M^{-1}(\xi^*, \boldsymbol{\theta})| = 95.144$, and the confirmation of the executed D-optimal design is in Figure 5.6.

Table 5.11: Locally D-optimal Designs for Model II Using Scenario 3 for $x_{lj} \in [0, 20]$

Group	Induced design space	Support points(x_{lj})	Support points(c_{lj})	weights(w_{lj})
1	[1.7,21.7]	(0,20)	(1.7,21.7)	(0.098,0.006)
2	[0.6,20.6]	(0,1.805,20)	(0.6,2.405,20.6)	(0.0833,0.0291,0.0044)
3	[1.3,21.3]	(0,20)	(1.3,21.3)	(0.0976,0.0071)
4	[0.6,20.6]	(0,1.805,20)	(0.6,2.405,20.6)	(0.0833,0.0291,0.0044)
5	[0.4,20.4]	(0,2.006,20)	(0.4,2.406,20.4)	(0.0781,0.0492,0.0007)
6	[1.1,21.1]	(0,20)	(1.1,21.1)	(0.0973,0.0077)
7	[0.8,20.8]	(0,1.604,20)	(0.8,2.404,20.8)	(0.0925,0.008,0.0076)
8	[1.5,21.5]	(0,20)	(1.5,21.5)	(0.0978,0.0065)
9	[0.7,20.7]	(0,1.705,20)	(0.7,2.405,20.7)	(0.0873,0.0188,0.0061)

Scenario 4: $\boldsymbol{\theta}_2 = (\alpha_{21} = 1.7, \alpha_{22} = 0.6, \alpha_{23} = 1.3, \alpha_{24} = 0.6, \alpha_{25} = 0.4, \alpha_{26} = 1.1, \alpha_{27} = 0.8, \alpha_{28} = 1.5, \alpha_{29} = 0.7, \beta_2 = 1)^T$

Figure 5.6: Locally D-optimal Design Verification for 9-Group Design for $x_{lj} \in [0, 20]$



Locally A-optimal designs

As we discussed before, the A-optimal designs depend on the values of the unknown parameters. For this case, we again study the effect of the variance of the continuous response σ_z^2, for $z = 0, 1$, on the selection of the A-optimal designs. At first we will evaluate the performance of the hybrid PSO-FM algorithm by comparing the CPU time in seconds between the fmincon solver and the hybrid PSO-FM algorithm in searching for the locally A-optimal designs. The same four scenarios that were used in the first search for locally D-optimal designs for Model II are used here. For comparison purposes, we set $\sigma_0^2 = \sigma_1^2 = 1$. We then will vary the values of both σ_z^2's. As performed before in the D-optimality case, the minimum and maximum

Table 5.12: Computation Time (in Second) for Locally A-optimal Designs for Model II

Algorithm	Design Space(x)	Scenario1		Scenario2		Scenario3		Scenario4	
		min	max	min	max	min	max	min	max
PSO-FM	$[-25, -5]$	11.14	16.78	18.08	32.73	32.75	40.83	53.57	66.44
fmincon		11.20	72.46	24.15	77.56	44.79	236.05	132.87	207.14
PSO-FM	$[-20, 0]$	11.55	14.64	15.84	25.39	32.01	36.57	38.26	61.41
fmincon		7.39	84.98	32.33	122.89	38.17	195.09	42.34	224.12
PSO-FM	$[-10, 10]$	10.91	13.95	24.81	66.08	33.02	40.68	50.63	66.22
fmincon		14.67	64.77	30.63	146.76	45.35	217.61	93.28	511.15
PSO-FM	$[0, 20]$	9.71	14.78	28.18	31.88	21.97	36.17	28.36	54.48
fmincon		9.34	43.06	60.61	200.19	88.62	241.42	109.22	521.1
PSO-FM	$[5, 25]$	8.57	11.28	23.34	25.35	19.49	35.41	24.07	46.31
fmincon		8.11	41.51	20.65	72.01	22.75	130.11	48.27	179.11

computation times are selected for each design space under all scenarios from five runs for the hybrid PSO-FM algorithm and from five runs for the fmincon solver as well excluding the times that the fmincon solver gets stuck in a local minimum as the number of groups increases in the model. These times are recorded in Table 5.12, where we observe the same results as in the case of D-optimality. We can conclude that as the number of groups increases, the model becomes more dependable on the initial points, and thus the hybrid PSO-FM algorithm becomes faster and more stable than the fmincon solver.

For Model II we can find a locally A-optimal designs where some of its groups requires only one support point, which is less than the minimal number of support points needed for Model I. This can be seen in the next example.

Example 5.4.3. Consider model (4.17) and use Scenario 1 with a design region $[5, 25]$ for the continuous variable x. As shown in Table 5.13, the locally A-optimal design found has eight support points where four of the groups (group 1, 4, 5, and 6) requires only one support point in their conditional measures. The A-criterion value for this design is $\Phi_A = \mathrm{tr}(M^{-1}) = 392580$. The design is verified by the GET as shown in Figure 5.7.

Table 5.13: Locally A-optimal Designs for Model II Using Scenario 1, for $x_{lj} \in [5, 25]$

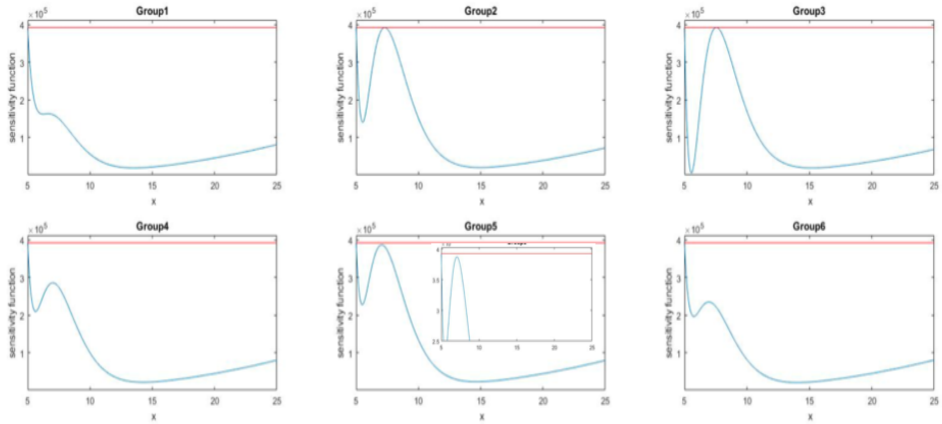
Group	Induced design space	Support points(x_{lj})	Support points(c_{lj})	weights(w_{lj})
1	[6.7,26.7]	(5)	(6.7)	(0.0644)
2	[5.6,25.6]	(5,7.29)	(5.6,7.89)	(0.0413,0.0384)
3	[5.4,25.4]	(5,7.55)	(5.4,7.95)	(0.1691,0.5451)
4	[6.1,26.1]	(5)	(6.1)	(0.0478)
5	[5.8,25.8]	(5)	(5.8)	(0.0411)
6	[6.3,26.3]	(5)	(6.3)	(0.0528)

Scenario1: $\boldsymbol{\theta}_2 = (\alpha_{21} = 1.7, \alpha_{22} = 0.6, \alpha_{23} = 0.4, \alpha_{24} = 1.1, \alpha_{25} = 0.8, \alpha_{26} = 1.3, \beta_2 = 1)^T$

The effects of the variance on the A-optimal designs

Now we will look at the effects of the variance σ_0^2 and σ_1^2 on the selection of the A-optimal designs. To study the variance effects on the optimal design, we use Model II with two groups. The guessed values of the parameters are set to $\boldsymbol{\theta}_2 = (\alpha_{21} = 4, \alpha_{22} = 2, \beta_2 = 2)^T$ and the design region for x is $[-6, 4]$. Then, the corresponding induced

Figure 5.7: Locally A-optimal Design Verification for 6-Group Design for $x_{lj} \in [-25, -5]$



design region for c are $[-8, 12]$ and $[-10, 10]$ for group l, $l = 1, 2$, respectively. As in Model I, we set the values of σ_0^2 and σ_1^2 in Table 5.14 and Table 5.15 to 0.1,0.5,1,10, and 100 by fixing one of the variances and allowing the other one to vary. Table 5.14 has the results for the group with the asymmetric induced design region $[-8, 12]$ while Table 5.15 is for the group with the symmetric induced design region $[-10, 10]$. There, we report the support points and corresponding weights. We also include the ratio, R, of the total weight of the outer-points to that of the two inner points.

We observe that the influence of the variance on the behavior of the support points and weights under both models (Model I/Model II) is almost the same. The optimal designs will tend to place a higher weight (and thus more observations) to the situations where the variance is larger.

5.5 Discussion

In this chapter, we searched for computational methods (algorithms) that are fast and efficient in providing A- and D-optimal designs. With the assistance of the complete class results found for these models, we managed to adjust the fmincon solver to work with models that have mixed responses. We considered the fmincon solver

Table 5.14: Locally A-optimal Designs for Model II for $c_{1j} \in [-8, 12]$

σ_0^2	σ_1^2	*Group*	c_{11}	c_{12}	c_{13}	c_{14}	w_{11}	w_{12}	w_{13}	w_{14}	R
0.1	0.1	1	-8	-1.94	1.92	12	0.0002	0.0711	0.3583	0.0056	0.01351
0.1	0.5	1	-8	-1.92	1.94	12	0.0002	0.0674	0.3557	0.0177	0.04231
0.1	1	1	-8	-1.88	1.96	12	0.0002	0.0637	0.3535	0.0261	0.06304
0.1	10	1	-8	-1.6	2.08	12	0.0001	0.0399	0.3532	0.0698	0.17782
0.1	100	1	-8	-0.94	2.34	12	0	0.0156	0.3759	0.1043	0.26641
0.5	0.1	1	-8	-1.98	1.88	12	0.0003	0.0886	0.3533	0.0052	0.01245
0.5	0.5	1	-8	-1.96	1.9	12	0.0003	0.0846	0.3505	0.0167	0.03907
0.5	1	1	-8	-1.94	1.92	12	0.0002	0.0807	0.3485	0.0248	0.05825
0.5	10	1	-8	-1.68	2	12	0.0002	0.0533	0.347	0.0674	0.16887
0.5	100	1	-8	-1	2.28	12	0	0.0204	0.372	0.1033	0.26325
1	0.1	1	-8	-2.02	1.84	12	0.0004	0.1014	0.3427	0.0049	0.01193
1	0.5	1	-8	-2	1.84	12	0.0004	0.0973	0.3402	0.0157	0.0368
1	1	1	-8	-1.98	1.86	12	0.0002	0.0937	03385	0.0233	0.05437
1	10	1	-8	-1.74	1.94	12	0.0003	0.0648	0.3386	0.0654	0.16287
1	100	1	-8	-1.06	2.2	12	0	0.0257	0.3675	0.1023	0.26017
10	0.1	1	-8	-1.9	1.4	12	0.0155	0.1536	0.2526	0.0028	0.04505
10	0.5	1	-8	-1.88	1.4	12	0.0152	0.1507	0.2521	0.0107	0.0643
10	1	1	-8	-1.88	1.42	12	0.0151	0.1481	0.2521	0.0164	0.07871
10	10	1	-8	-1.68	1.46	12	0.0128	0.1232	0.2635	0.0501	0.16266
10	100	1	-8	-1.14	1.74	12	0.0053	0.0674	0.3188	0.0903	0.24754
100	0.1	1	-8	-1.56	0.88	12	0.0287	0.2103	0.1549	0.001	0.08133
100	0.5	1	-8	-1.54	0.9	12	0.0286	0.209	0.155	0.0049	0.09203
100	1	1	-8	-1.54	0.9	12	0.0284	0.2079	0.1552	0.0076	0.09915
100	10	1	-8	-1.44	0.9	12	0.027	0.1972	0.1609	0.0276	0.15247
100	100	1	-8	-1.06	1.08	12	0.0203	0.1574	0.2078	0.0607	0.2218

Table 5.15: Locally A-optimal Designs for Model II for $c_{2j} \in [-10, 10]$

σ_0^2	σ_1^2	Group	c_{21}	c_{22}	c_{23}	c_{24}	w_{21}	w_{22}	w_{23}	w_{24}	R
0.1	0.1	2	-10	-1.88	1.9	10	0.00088	0.195	0.3574	0.0002	0.01629
0.1	0.5	2	-10	-1.86	1.92	10	0.0085	0.1927	0.3576	0.0002	0.01581
0.1	1	2	-10	-1.84	1.96	10	0.0081	0.188	0.3601	0.0002	0.01514
0.1	10	2	-10	-1.56	2.2	10	0.0051	0.1437	0.3874	0.0008	0.01111
0.1	100	2	-10	-0.92	2.54	10	0.0012	0.0746	0.4108	0.0175	0.03852
0.5	0.1	2	-10	-1.8	1.84	10	0.0314	0.1875	0.3335	0.0002	0.06065
0.5	0.5	2	-10	-1.78	1.86	10	0.0303	0.1827	0.3347	0.0002	0.05895
0.5	1	2	-10	-1.76	1.88	10	0.0293	0.1787	0.3376	0.0002	0.05714
0.5	10	2	-10	-1.48	2.14	10	0.0199	0.1405	0.3709	0.0008	0.04048
0.5	100	2	-10	-0.88	2.5	10	0.0063	0.0752	0.4053	0.0174	0.04932
1	0.1	2	-10	-1.74	1.76	10	0.0464	0.1866	0.3174	0.0002	0.09246
1	0.5	2	-10	-1.72	1.78	10	0.045	0.182	0.3191	0.0002	0.0902
1	1	2	-10	-1.7	1.82	10	0.0436	0.1783	0.3221	0.0002	0.08753
1	10	2	-10	-1.44	2.08	10	0.0306	0.1423	0.3577	0.0003	0.0618
1	100	2	-10	-0.84	2.44	10	0.0103	0.0769	0.3996	0.0178	0.05897
10	0.1	2	-10	-1.52	1.3	10	0.107	0.2266	0.2419	0.0002	0.22882
10	0.5	2	-10	-1.5	1.32	10	0.1051	0.2227	0.2433	0.0002	0.22597
10	1	2	-10	-1.48	1.34	10	0.103	0.2195	0.2455	0.0002	0.22194
10	10	2	-10	-1.28	1.56	10	0.0828	0.1892	0.2781	0.0003	0.17783
10	100	2	-10	-0.72	1.92	10	0.0395	0.1123	0.3457	0.0207	0.13144
100	0.1	2	-10	-1.1	0.74	10	0.1492	0.3023	0.1537	0.0002	0.32763
100	0.5	2	-10	-1.1	0.74	10	0.1481	0.3003	0.154	0.0002	0.32644
100	1	2	-10	-1.1	0.76	10	0.147	0.2988	0.1549	0.0002	0.32444
100	10	2	-10	-1	0.86	10	0.1352	0.2831	0.1687	0.0003	0.29991
100	100	2	-10	-0.62	1.06	10	0.0933	0..2204	0.2193	0.0209	0.25972

to obtain optimal designs for Model I for which we can obtain an optimal design by combining the optimal (conditional) designs for the individual groups. The fmincon solver can easily handle the search of the optimal designs for this model. For cases with a large number of groups, we may consider the parallel computing technique to speed up the search of optimal designs. We have compared the performance of the fmincon solver with the outstanding OWE algorithm. Our study shows that the fmincon solver and the OWE algorithm can both identify optimal designs, although the fmincon has some advantages in speed over the OWE algorithm. Consequently, we conclude that even though the OWE algorithm was proven to outperform other algorithms in design search problems for some types of models, a better alternative method for our model, for which a complete class result is available, is by the use of the off-the-shelf fmincon solver for its speed and its ease of use.

For a more complicated model as Model II, where we need to search for the entire design at once, the hybrid PSO-FM algorithm was proposed. The hybrid PSO-FM algorithm attempts to avoid a poorly selected initial for the fmincon solver. From our numerical results, we observe that the proposed algorithm is especially useful in the case where we have more than four groups in the model. For all the optimal designs generated by the hybrid PSO-FM algorithm, the efficiency lower bound is satisfied if $A_{lb}(\xi_A) > 0.9999$ for A-optimality and $D_{lb}(\xi_D) > 0.9999$ for D-optimality. For the D-optimal design, we can use a higher efficiency lower bound as 0.999999, and the algorithm will still be able to find the optimal design that satisfies this bound easily without much time compromises. But this does not apply for A-optimality and we tend to need a much greater computational effort to find a design with an efficiency higher than 99.99% in some cases. With the model that we consider, the hybrid PSO-FM algorithm can handle a large number of groups (~ 30), but as the number of groups increases in the model, the cost in time increases as well. We also observed that the algorithm works faster using the D-optimality criterion than the A-optimality criterion. Although the hybrid PSO-FM algorithm preformed well in the cases we considered, but it still has some limitations such as there is a slight chance for the algorithm to get stuck in a sub-optimal solution. In addition, there is a variation in the running time of the same model due to the randomness (in PSO) of the algorithm.

Chapter 6

AN EXTINSION: QUADRATIC MIXED RESPONSES MODEL

The mixed responses models that we considered throughout this work include two types of sub-models for different types of outcomes: the linear model for the continuous response and a generalized linear model for the binary response. In the case of the linear model, when there is a nonlinear relation between the predictor and the response variables, higher order terms may be considered to capture the curvature trend that is present in the nature of the experimental data. The same concept applies in the case of the generalized linear model where higher order terms may be required when modeling binary data.

Many optimality results under the GLMs focused on models with a single quadratic covariate. Wu and Stufken (2014) obtained complete class results of optimal designs for GLMs with a 2nd-order polynomial as the linear predictor. In the process of deriving an upper bound for the number of support points, they found that if the focus is on the Φ_p-optimality criteria and the design region is unrestricted, then the optimal designs can be restricted to the class of symmetric designs. With three parameters and a single independent covariate in the model, they obtained a symmetric complete class of 3- and 4-support points, depending on the parameter values. As discussed in Hyun (2013), the probit model with a quadratic term provides a good fit in modeling the dose response functions with a downturn in toxicology studies. He applied the complete class approach to search for the A- and D-optimal designs for a quadratic model with three parameters and revealed a complete class of 4-support points.

In the scope of mixed responses models, Biswas and López-Fidalgo (2013) worked on a mixed responses model for dose-finding design problems in the treatment of breast cancer, where the toxicity is binary, and the efficacy is continuous. They considered a type of quadratic mixed responses model with a single independent covariate. The mixed quadratic model that they considered with a single independent covariate includes two sub-models; a logistic model for the binary toxicity response and a linear model for the continuous efficacy response where they placed a quadratic term.

Kim (2017) considered a different form of a quadratic mixed responses model, where she extended the simple mixed responses model by inserting a quadratic term in the logistic sub-models only, leaving the linear sub-models unchanged. A sharp bound on the number of support points was identified by forming the symmetric complete class of 6-support points, including the boundaries of the design region under the D-optimality. The mathematical programming (fmincon solver) was used by Kim (2017) to search for the locally D-optimal designs.

Another interesting extension of our current results is thus to examine the case where the linear sub-model involves a quadratic term while we set the GLM unchanged with only the simple linear predictor. This is to consider the model studies in Biswas and López-Fidalgo (2013). According to Biswas and López-Fidalgo (2013), "The whole model considered is quite natural and widely used in many areas." The goal is to investigate the existence of the complete class in this scenario, which to our knowledge has never been studied in the literature. We then compare our results with those obtained by Biswas and López-Fidalgo (2013).

6.1 Statistical Model

As in the earlier chapters, the observable data is (x_j, y_j, z_j), $j = 1, \ldots, N$, where $x_j, y_j \in \mathbb{R}$ and $z_j \in \{0, 1\}$. The binary variable z is modeled by a logistic regression model, and $y_j|z_j$ is described by a normal regression model with a quadratic term. Specifically, the model that we consider in this chapter is:

$$y_j|z_j \sim N((\boldsymbol{f}_1^T(x_j)\boldsymbol{\theta}_1 = \alpha_1 + \beta_{11}x_j + \beta_{12}x_j^2 + \gamma z_j, \sigma^2), \quad \text{for} \quad z_j = 0, 1,$$

and

$$P(z_j = 1|x) = \mathrm{P}(\boldsymbol{f}_2^T(x_j)\boldsymbol{\theta}_2 = \alpha_2 + \beta_{21}x_j) \equiv p(x_j), \tag{6.1}$$

where $\boldsymbol{f}_1(x_j) = (1, x_j, x_j^2, z_j)^T$ and $\boldsymbol{f}_2(x_j) = (1, x_j)^T$. Here, γ represents the association parameter of the conditional distribution for y given z, σ^2 denotes the variance of the conditional distribution for y given z, and $\mathrm{P}(\cdot)$ is a cumulative distribution function that follows a logistic model; following Biswas and López-Fidalgo (2013), we set $\mathrm{P}(\alpha_2 + \beta_{21}x_j) = \dfrac{1}{1 + e^{\alpha_2 + \beta_{21}x_j}}$. The unknown parameters are $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2\}$, where $\boldsymbol{\theta}_1 = (\alpha_1, \beta_{11}, \beta_{12}, \gamma)^T$ and $\boldsymbol{\theta}_2 = (\alpha_2, \beta_{21})^T$. For easy referral, we will call this model the quadratic mixed responses model.

The joint probability function for (y, z) is computed by direct factorization as:

$$\begin{aligned} f(y_j, z_j) &= f(z_j)f(y_j|z_j) \\ &= [p(x_j)]^{z_j}[1 - p(x_j)]^{1-z_j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[- \frac{(y_j - \boldsymbol{f}_1^T(x_j)\boldsymbol{\theta}_1)^2}{2\sigma^2} \Big]. \end{aligned}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ be the parameter vector of interest, then the log-likelihood function is:

$$\begin{aligned} \log L(\boldsymbol{\theta}) =& \log \prod_{j=1}^{N} (f(y_j, z_j)) \\ =& \sum_{j=1}^{N} \Big\{ z_j \log[p(x_j)] + (1 - z_j)\log[1 - p(x_j)] - \frac{(y_i - \boldsymbol{f}_1(x_j)^{\mathrm{T}}\boldsymbol{\theta}_1)^2}{2\sigma^2} - \frac{1}{2}\log[2\pi\sigma^2] \Big\}. \end{aligned} \tag{6.2}$$

From the likelihood function we can derive the information matrix as:

$$M(\boldsymbol{\theta}, x_j) = -E\begin{pmatrix} \frac{\partial^2 logL}{\partial^2\alpha_1} & \frac{\partial^2 logL}{\partial\alpha_1\partial\beta_{11}} & \frac{\partial^2 logL}{\partial\alpha_1\partial\beta_{12}} & \frac{\partial^2 logL}{\partial\alpha_1\partial\gamma} & \frac{\partial^2 logL}{\partial\alpha_1\partial\alpha_2} & \frac{\partial^2 logL}{\partial\alpha_1\partial\beta_{21}} \\ \frac{\partial^2 logL}{\partial\beta_{11}\partial\alpha_1} & \frac{\partial^2 logL}{\partial^2\beta_{11}} & \frac{\partial^2 logL}{\partial\beta_{11}\partial\beta_{12}} & \frac{\partial^2 logL}{\partial\beta_{11}\partial\gamma} & \frac{\partial^2 logL}{\partial\beta_{11}\partial\alpha_2} & \frac{\partial^2 logL}{\partial\beta_{11}\partial\beta_{21}} \\ \frac{\partial^2 logL}{\partial\beta_{12}\partial\alpha_1} & \frac{\partial^2 logL}{\partial\beta_{12}\partial\beta_{11}} & \frac{\partial^2 logL}{\partial^2\beta_{12}} & \frac{\partial^2 logL}{\partial\beta_{12}\partial\gamma} & \frac{\partial^2 logL}{\partial\beta_{12}\partial\alpha_2} & \frac{\partial^2 logL}{\partial\beta_{12}\partial\beta_{21}} \\ \frac{\partial^2 logL}{\partial\gamma\partial\alpha_1} & \frac{\partial^2 logL}{\partial\gamma\partial\beta_{11}} & \frac{\partial^2 logL}{\partial\gamma\partial\beta_{12}} & \frac{\partial^2 logL}{\partial^2\gamma} & \frac{\partial^2 logL}{\partial\gamma\partial\alpha_2} & \frac{\partial^2 logL}{\partial\gamma\partial\beta_{21}} \\ \frac{\partial^2 logL}{\partial\alpha_2\partial\alpha_1} & \frac{\partial^2 logL}{\partial\alpha_2\partial\beta_{11}} & \frac{\partial^2 logL}{\partial\alpha_2\partial\beta_{12}} & \frac{\partial^2 logL}{\partial\alpha_2\partial\gamma} & \frac{\partial^2 logL}{\partial^2\alpha_2} & \frac{\partial^2 logL}{\partial\alpha_2\partial\beta_{21}} \\ \frac{\partial^2 logL}{\partial\beta_{21}\partial\alpha_1} & \frac{\partial^2 logL}{\partial\beta_{21}\partial\beta_{11}} & \frac{\partial^2 logL}{\partial\beta_{21}\partial\beta_{12}} & \frac{\partial^2 logL}{\partial\beta_{21}\partial\gamma} & \frac{\partial^2 logL}{\partial\beta_{21}\partial\alpha_2} & \frac{\partial^2 logL}{\partial^2\beta_{21}} \end{pmatrix}.$$

Let us denote $c_j = \alpha_2 + \beta_{21}x_j$, then the (individual) information matrix at x_j can be calculated as:

$$M(\boldsymbol{\theta}, x_j) = \begin{pmatrix} \frac{1}{\sigma^2} & \frac{x_j}{\sigma^2} & \frac{x_j^2}{\sigma^2} & \frac{1}{\sigma^2(1+e^{c_j})} & 0 & 0 \\ \frac{x_j}{\sigma^2} & \frac{x_j^2}{\sigma^2} & \frac{x_j^3}{\sigma^2} & \frac{x_j}{\sigma^2(1+e^{c_j})} & 0 & 0 \\ \frac{x_j^2}{\sigma^2} & \frac{x_j^3}{\sigma^2} & \frac{x_j^4}{\sigma^2} & \frac{x_j^2}{\sigma^2(1+e^{c_j})} & 0 & 0 \\ \frac{1}{\sigma^2(1+e^{c_j})} & \frac{x_j}{\sigma^2(1+e^{c_j})} & \frac{x_j^2}{\sigma^2(1+e^{c_j})} & \frac{1}{\sigma^2(1+e^{c_j})} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{e^{c_j}}{(1+e^{c_j})^2} & x_j\frac{e^{c_j}}{(1+e^{c_j})^2} \\ 0 & 0 & 0 & 0 & x_j\frac{e^{c_j}}{(1+e^{c_j})^2} & x_j^2\frac{e^{c_j}}{(1+e^{c_j})^2} \end{pmatrix}. \quad (6.3)$$

The computed information matrix for $\boldsymbol{\theta}$ under a continuous design $\xi = \{(x_j, w_j), j = 1, \cdots, m\}$ can be represented as:

$$M(\xi, \boldsymbol{\theta}) = \sum_{j=1}^{m} w_j M(\boldsymbol{\theta}, x_j). \quad (6.4)$$

Notice that the information matrix depends on the value of $\boldsymbol{\theta}_2$ and σ^2. The representation for c_j is helpful to decompose the information matrix as $\mathrm{M}(\xi, \boldsymbol{\theta}) = B(\boldsymbol{\theta}) \sum_{j=1}^{m} w_j C(\boldsymbol{\theta}, c_j) B^{\mathrm{T}}(\boldsymbol{\theta})$, where $C(\boldsymbol{\theta}, c_j) =$

$$\left(\begin{array}{cccccc} \Gamma_{11}=1 & \Gamma_{12}=c_j & \Gamma_{13}=c_j^2 & \Gamma_{14}=\frac{1}{1+e^{c_j}} & 0 & 0 \\ \Gamma_{21}=c_j & \Gamma_{22}=c_j^2 & \Gamma_{23}=c_j^3 & \Gamma_{24}=\frac{c_j}{1+e^{c_j}} & 0 & 0 \\ \Gamma_{31}=c_j^2 & \Gamma_{32}=c_j^3 & \Gamma_{33}=c_j^4 & \Gamma_{34}=\frac{c_j^2}{1+e^{c_j}} & 0 & 0 \\ \Gamma_{41}=\frac{1}{1+e^{c_j}} & \Gamma_{42}=\frac{c_j}{1+e^{c_j}} & \Gamma_{43}=\frac{c_j^2}{1+e^{c_j}} & \Gamma_{44}=\frac{1}{1+e^{c_j}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \Gamma_{55}=\frac{e^{c_j}}{(1+e^{c_j})^2} & \Gamma_{56}=c_j\frac{e^{c_j}}{(1+e^{c_j})^2} \\ 0 & 0 & 0 & 0 & \Gamma_{65}=c_j\frac{e^{c_j}}{(1+e^{c_j})^2} & \Gamma_{66}=c_j^2\frac{e^{c_j}}{(1+e^{c_j})^2} \end{array}\right), \quad (6.5)$$

and $B(\boldsymbol{\theta}) = \mathrm{diag}(\frac{1}{\sigma}B_1, B_2)$. Here, $B_1 = \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \frac{-\alpha_2}{\beta_{21}} & \frac{1}{\beta_{21}} & 0 & 0 \\ \frac{\alpha_2^2}{\beta_{21}^2} & \frac{-2\alpha_2}{\beta_{21}^2} & \frac{1}{\beta_{21}^2} & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$ and $B_2 = \left(\begin{array}{cc} 1 & 0 \\ \frac{-\alpha_2}{\beta_{21}} & \frac{1}{\beta_{21}} \end{array}\right)$.

6.2 Complete Class Results

In this section, we search for a complete class of locally optimal designs for the quadratic mixed responses model. To do so, we apply Lemma 2.3.2 which requires identifying the Chebyshev systems first. We note that if a Chebyshev system is found, then it can be used to form other Chebyshev systems based on the next Lemma.

Lemma 6.2.1. *Suppose that the set of functions* $\{\Psi_0^*, \cdots, \Psi_k^*\}$ *forms a Chebyshev system and there exists a matrix* A *such that* $A(\Psi_0^*, \cdots, \Psi_k^*)^T = (\Psi_0, \cdots, \Psi_k)^T$ *with* $det(A) > 0$. *Then, the set of functions* $\{\Psi_0, \cdots, \Psi_k\}$ *forms a Chebyshev system.*

Proof. Let the matrix A contains the elements $a_{i,j}$ for $i, j = 0, \cdots, k$, then we have $\Psi_i(c) = \sum_j a_{i,j}\Psi_j^*(c)$ for $D_1 \le c_0 < \cdots < c_k \le D_2$, and $c_j \in [D_1, D_2]$. Since the set of functions $\{\Psi_0^*, \cdots, \Psi_k^*\}$ forms a Chebyshev system, it follows that $det[(\Psi_j^*(c))_{i,j=0,\cdots,k}] > 0$. Thus, $det[A(\Psi_j^*(c))_{i,j=0,\cdots,k}] = det[A]det[(\Psi_j^*(c))_{i,j=0,\cdots,k}] > 0$. With $det(A) > 0$, then the conclusion of the set $\{\Psi_0, \cdots, \Psi_k\}$ forming a Chebyshev system follows. $\square$

Lemma 6.2.2. *The set of functions* $\{\Psi_0^* = 1, \Psi_1^* = \dfrac{e^{2c_j}}{(1+e^{c_j})^2}, \Psi_2^* = \dfrac{ce^{2c_j}}{(1+e^{c_j})^2}, \Psi_3^* =$

$\dfrac{c^2e^{2c_j}}{(1+e^{c_j})^2}, \Psi_4^* = \dfrac{-e^{c_j}}{1+e^{c_j}}, \Psi_5^* = \dfrac{-ce^{c_j}}{1+e^{c_j}}, \Psi_6^* = \dfrac{-c^2e^{c_j}}{1+e^{c_j}}, \Psi_7^* = c, \Psi_8^* = c^2, \Psi_9^* = c^3,$ $\Psi_{10}^* = c^4\}$ *forms a Chebyshev system.*

Proof. Using the definition for $f_{l,l}$'s in (2.6), $f_{1,1} = \dfrac{2e^{2c}}{(1+e^c)^3}$, $f_{2,2} = 1 + \dfrac{e^c}{2}$, $f_{3,3} = 2 + \dfrac{2e^c}{(2+e^c)^2}$, $f_{4,4} = 2\dfrac{2e^{-c}+1}{(4+e^c)^2}$, $f_{5,5} = 1 + \dfrac{e^c}{4}$, $f_{6,6} = 2$, $f_{7,7} = 4e^{-c}$, $f_{8,8} = 2$, $f_{9,9} = \dfrac{21}{4}e^c + 3(e^c-1)^2$. It is clear that $f_{l,l} > 0 \; \forall c \in \mathbb{R}$ where $l = 1, \cdots, 9$. $f_{10,10} = 4\dfrac{321e^{2c} - 26e^{3c} + 16e^{4c} - 26e^c + 16}{(4e^{2c} - e^c + 4)^2} > 0 \; \forall c \in \mathbb{R}$ by Proposition 4 found in the Appendix. Thus, $F(c) = \prod_{l=1}^{10} f_{l,l} = 24[8 + 27\dfrac{11 - 4\cosh(c)}{(1+\cosh(c))(8\cosh(c)-1)}] > 0$. According to Proposition 4 found in Yang and Stufken (2012), the conclusion of the Ψ sets forming Chebyshev systems follows. □

From the C matrix in (6.5), the number of independent Γ_{ij} functions that is used to form a Chebyshev systems are $k = 10$. We then have the following complete class results:

Lemma 6.2.3. *For a quadratic mixed responses model, where* $k = 10$, $\{\Psi_0, \Psi_1 = \Gamma_{12}, \Psi_2 = \Gamma_{13}, \Psi_3 = \Gamma_{14}, \Psi_4 = \Gamma_{24}, \Psi_5 = \Gamma_{34}, \Psi_6 = \Gamma_{55}, \Psi_7 = \Gamma_{56}, \Psi_8 = \Gamma_{66}, \Psi_9 = \Gamma_{23}\}$ *and* $\{\Psi_0, \Psi_1 = \Gamma_{12}, \Psi_2 = \Gamma_{13}, \Psi_3 = \Gamma_{14}, \Psi_4 = \Gamma_{24}, \Psi_5 = \Gamma_{34}, \Psi_6 = \Gamma_{55}, \Psi_7 = \Gamma_{56}, \Psi_8 = \Gamma_{66}, \Psi_9 = \Gamma_{23}, \Psi_{10}\}$ *form a Chebyshev system, where* $\Psi_0 = \Gamma_{11} = 1$ *and* $\Psi_{10} = \Gamma_{33}$.

Proof. Note that: $\Psi_0 = \Psi_0^*, \Psi_1 = \Psi_7^*, \Psi_2 = \Psi_8^*, \Psi_3 = \Psi_0^* + \Psi_4^*, \Psi_4 = \Psi_5^* + \Psi_7^*, \Psi_5 = \Psi_6^* + \Psi_8^*, \Psi_6 = -\Psi_1^* - \Psi_4^*, \Psi_7 = -\Psi_2^* - \Psi_5^*, \Psi_8 = -\Psi_3^* - \Psi_6^*, \Psi_9 = \Psi_9^*$, and $\Psi_{10} = \Psi_{10}^*$. Based on Lemma 6.2.1 and Lemma 6.2.2, the set of functions $\{\Psi_0, \Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5, \Psi_6, \Psi_7, \Psi_8, \Psi_9, \Psi_{10}\}$ forms a Chebyshev system.

□

Theorem 6.2.4. *Under a quadratic mixed responses model the designs with at most six support points, including the two endpoints of the design region, form a complete class.*

Proof. Based on Lemma 6.2.2 and Lemma 2.3.2 part(c), where $k = 10$ and $n^* = \dfrac{k}{2} = 5$ implies that the designs with at most 6, $n^* + 1$, support points, including both end-points of the design region. Which implies that for any design $\xi = \{(c_j, w_j), j = 1, \cdots, m\}$ with $m \geq 6$, we can find a design $\xi^* = \{(c_j^*, w_j^*), j = 1, \cdots, 6\}$ so that $\sum_{j=1}^{6} w_j^* \Psi_k(c_j^*) = \sum_{j=1}^{N} w_j \Psi_k(c_j)$, $k = 0, \cdots, 9$ and $\sum_{j=1}^{6} w_j^* \Psi_{10}(c_j^*) \geq \sum_{j=1}^{N} w_j \Psi_{10}(c_j)$ is satisfied. Then we conclude that $M(\xi^*, \theta) \geq M(\xi, \theta)$ and identify the complete class. □

6.3 Model With Group Effects

To make the quadratic mixed responses model in (6.1) more practical, we can include additional qualitative factors for group effects. As in Chapter 4, the total number of groups formed by the L qualitative factors is $s = s_1 s_2 \cdots s_L$, where s_l is the number of levels of the lth factor and (y_{lj}, z_{lj}) denotes the response vector of a subject in the lth group having continuous explanatory variables x. Then the quadratic mixed responses model can be expressed as:

$$
\begin{aligned}
& y_{lj}|z_{lj} \sim N(\boldsymbol{f}_1^T(l, x_{lj})\boldsymbol{\theta}_1 = \alpha_{1l} + \beta_{11}x_{lj} + \beta_{12}x_{lj}^2 + \gamma z_{lj}, \sigma^2), \quad \text{for}, z_{lj} = 0, 1, \\
& \text{and} \\
& P(z_{lj} = 1|x) = \mathrm{P}(\boldsymbol{f}_2^T(l, x_{lj})\boldsymbol{\theta}_2 = \alpha_{2l} + \beta_{21}x_{lj}).
\end{aligned} \tag{6.6}
$$

Here, $\boldsymbol{f}_1^T(l, x_{lj}) = (\boldsymbol{e}_l^T, x_{lj}, x_{lj}^2, z_{lj})^T$, $\boldsymbol{f}_2^T(l, x_{lj}) = (\boldsymbol{e}_l^T, x_{lj})$, $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}_{1l}, \beta_{11}, \beta_{12}, \gamma)^T$, and $\boldsymbol{\theta}_2 = (\boldsymbol{\alpha}_{2l}, \beta_{21})^T$, where $\boldsymbol{\alpha}_{rl} = (\alpha_{r1}, \cdots, \alpha_{rs})^T$ represents the effect of the lth group in the sub-model $r = 1, 2$, and $\boldsymbol{e}_l = (0, \cdots, 0, 1, 0, \cdots, 0)^T \in \mathbb{R}^s$ with the lth group as 1 and the other elements as 0.

The information matrix for $\boldsymbol{\theta}$ under a continuous design $\xi = \{(x_{lj}, w_{lj}), j = 1, \cdots, m_l\}$ can be represented as:

$$M(\xi,\boldsymbol{\theta})=\sum_{l=1}^{s}\sum_{j=1}^{m_l}w_{lj}\begin{pmatrix}\boldsymbol{e}_l\boldsymbol{e}_l^T & \frac{x_{lj}}{\sigma^2}\boldsymbol{e}_l & \frac{x_{lj}^2}{\sigma^2}\boldsymbol{e}_l & \frac{\Gamma_1(c_{lj})}{\sigma^2}\boldsymbol{e}_l & \mathbf{0}\\ \frac{x_{lj}}{\sigma^2}\boldsymbol{e}_l^T & \frac{x_{lj}^2}{\sigma^2} & \frac{x_{lj}^3}{\sigma^2} & \frac{x_{lj}}{\sigma^2}\Gamma_1(c_{lj}) & \mathbf{0}\\ \frac{x_{lj}^2}{\sigma^2}\boldsymbol{e}_l^T & \frac{x_{lj}^3}{\sigma^2} & \frac{x_{lj}^4}{\sigma^2} & \frac{x_{lj}^2}{\sigma^2}\Gamma_1(c_{lj}) & \mathbf{0}\\ \frac{\Gamma_1(c_{lj})}{\sigma^2}\boldsymbol{e}_l^T & x_{lj}\frac{\Gamma_1(c_{lj})}{\sigma^2} & \frac{x_{lj}^2}{\sigma^2}\Gamma_1(c_{lj}) & \frac{\Gamma_1(c_{lj})}{\sigma^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \Gamma_2(c_{lj})\boldsymbol{f}_2(x_{lj})\boldsymbol{f}_2^{\mathrm{T}}(x_{lj})\end{pmatrix}. \quad (6.7)$$

Here, $c_{lj}=\alpha_{2l}+\beta_{21}x_{lj}$, $\Gamma_1(c_{lj})=\dfrac{1}{(1+e^{c_{lj}})}$, and $\Gamma_2(c_{lj})=\dfrac{e^{c_{lj}}}{(1+e^{c_{lj}})^2}$.

The previously obtained complete class for the quadratic mixed responses model can be extended to the current model group effects. Theorem 6.3.1 gives the result for the complete class for the quadratic mixed responses model that includes group effects.

Theorem 6.3.1. *For mixed responses model* (6.6)*, a complete class of locally optimal designs can be formed by designs that contain at most six support points, including the two endpoint of the design space in each group.*

Proof. The same arguments used to proof Theorem 4.3.1 can be adopted in proofing this result. □

6.4 Numerical Results

The results of the complete class in the previous section has provided a great reduction in the search for the optimal design under the quadratic mixed responses model. This finding helps to limit the number of support points required for searching for an optimal design. With the current quadratic mixed responses model, we did face the same difficulties as in the case of the previously studied Model II. Thus, the hybrid PSO-FM algorithm proposed and described in Chapter 5 is again used for the search of the locally optimal design. For the fmincon solver the decision vector is set to $\xi^0 = \{c_2, c_3, c_4, c_5, w_1, w_2, w_3, w_4, w_5, w_6\}$. The initial points c_i, $i = 2, \cdots, 5$, are imported from the preceding PSO search. The fmincon solver may report an optimal design that has design points with very small weights or design points that are essentially the same. Thus, the design points with very small weights will be discarded, and the design points that are similar will be combined. Lastly, the locally optimal design obtained is verified by the general equivalence theorem.

We now search for the D-optimal designs by setting $\boldsymbol{\theta}_2 = (\alpha_2 = 2, \beta_{21} = 1)$ and $\sigma^2 = 1$. Our results are represented in Table 6.1, where we consider different design spaces. As shown from the table, we obtain 4 and 5 support points designs for different design spaces. When the induced design region is all negative $[-24, -4]$, or all positive $[4, 24]$, we have 4 support points including the endpoints of the induced design region. When the induced design region contains zero as $[-17, 3]$, and $[-3, 17]$, we have 5 support points including the two endpoints of the corresponding induced design region. We note when the induced design region is symmetric $[-10, 10]$, the optimal design is a 4 support pints symmetric design. All the designs obtained in the table are verified by the GET.

Table 6.1: Locally D-optimal Designs Under Model (6.1) With Different Design Spaces

Design space (x)	Design space (c)	Support points(x_j)	Support points(c_j)	weights(w_j)
[-26,-6]	[-24,-4]	(-26,-16.232,-8.017,-6)	(-24,-14.232,-6.017,-4)	(0.1655,0.1437,0.2313,0.4595)
[-19,1]	[-17,3]	(-19,-10.613,-3.329,-0.841,1)	(-17,-8.613,-1.329,1.159,3)	(0.1636,0.066,0.3607,0.2003,0.2094)
[-12,8]	[-10,10]	(-12,-3.385,-0.615,8)	(-10,-1.385,1.385,10)	(0.1649,0.3351,0.3351,0.1649)
[-5,15]	[-3,17]	(-5,-3.159,-0.671,6.613,15)	(-3,-1.159,1.329,8.613,17)	(0.2094,0.2003,0.3607,0.066,0.1636)
[2,22]	[4,24]	(2,4.017,12.232,22)	(4,6.017,14.232,24)	(0.4595,0.2313,0.1437,0.1655)

6.5 Case Study: Dose-Finding Experiment for Breast Cancer

In this section, we will compare our results with that of Biswas and López-Fidalgo (2013) findings. The quadratic mixed responses model in (6.1) was considered in their study to determine the dose, x, for a new drug called Ridaforolimus. Ridaforolimus is a potent inhibitor of mammalian target of rapamycin known as mTOR, used to treat patients with cancer. The efficacy which is the continuous response y in model (6.1) is evaluated by the decrease in protein Ki67, and the toxicity is the binary response z that determines the appearance of adverse effects in the model or not.

With milligram (mg) being the unit for measuring the amount of dose, the dose under the experiment ranges from $[0, 50]$. Thus, a linear transformation was done on the range of the dose to transform it to $[0, 1]$. The criterion used to search for the optimal design is the D-optimality criterion with $\boldsymbol{\theta}_2 = (\alpha_2 = 7, \beta_{21} = -10)$ and $\sigma^2 = 0.05$.

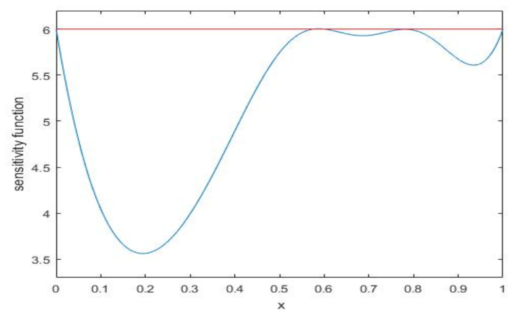
Based on Theorem 6.2.4, we managed to identify the locally D-optimal design within the complete class of designs which is reported in Table 6.2. The design obtained by Biswas and López-Fidalgo (2013), is also presented there, and the two designs appear to be the same. With this design, about 17.2% of patients will be given the usual dose of dalotuzumab without any dose of the new drug. On the other hand, around 39.7% of patients will be given 29.3934 ($\sim 0.587868 \times 50$) mg of Ridaforolimus. About 14.4% of patients will be given 39.075 ($\sim 0.781515 \times 50$) mg of Ridaforolimus, and about 28.7% of patients will be given the maximum dose of 50 mg of Ridaforolimus. The verification for our selected design can be found in Figure 6.1.

As mentioned in Biswas and López-Fidalgo (2013), when searching for the most 'successful' dose for patients with cancer, the amount of doses subscribed for the patient usually depends on other characteristics such as gender, age groups, or type of tumor and so on, that account for the heterogeneity in the patient population. The successful dose is based

Table 6.2: Locally D-optimal Designs Under Model (6.1) for $x_j \in [0,1]$

case	Design for x				$\Phi_D = \log\lvert M^{-1}(\xi)\rvert$
Biswas and López-Fidalgo (2013)	0	0.587868	0.781515	1	
	0.17174	0.396904	0.14462	0.286736	
	0	0.587868	0.781515	1	4.8752
Optimal design obtained	0.1717404	0.3969041	0.1446199	0.2867355	

Figure 6.1: The Verification of Locally D-optimal Design for $x_j \in [0,1]$



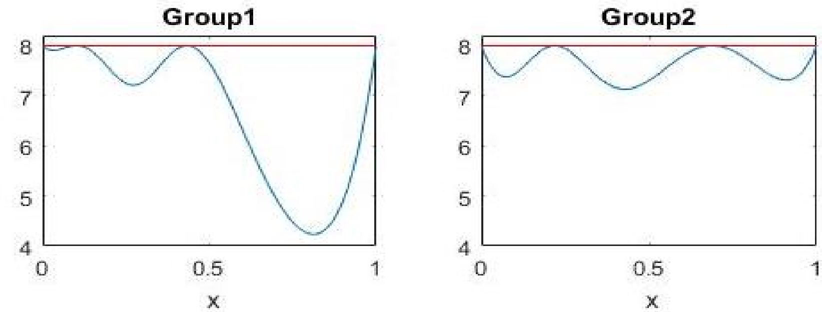
on choosing the dose levels that provide the most effective combination of both efficacy and toxicity. For illustration purposes, let us include the factor of gender (male or female) in model (6.6). If we set $\boldsymbol{\theta}_2 = (\alpha_{21} = 3, \alpha_{22} = 1, \beta_{21} = -10,)$, $\sigma^2 = 0.05$, and $x \in [0,1]$, then the locally D-optimal design selected is in Table 6.3, with $\Phi_D(\xi^*) = \log\lvert M^{-1}(\xi)\rvert = 15.8356$. The result is justified by the GET and a plot of the verification is found in Figure 6.2.

Table 6.3: Locally D-optimal Designs Under Model (6.6) for $x_{lj} \in [0,1]$

Group	Induced design space(c_l)	Support points(x_{lj})	Support points(c_{lj})	weights(w_{lj})
1	[3,-7]	(0,0.102,0.433,1)	(3,1.98,-1.33,-7)	(0.0347,0.1838,0.2039,0.1053)
2	[1,-9]	(0,0.219,0.689,1)	(1,-1.19,-5.89,-9)	(0.1868,0.1307,0.0920,0.0629)

Here, we assumed that the gender factor is controllable by the experimenter, and the optimal design obtained by the algorithm has marginal weights of around 47% for males

Figure 6.2: The Verification of Locally D-optimal Design for $x_{lj} \in [0, 1]$



and around 53% for female patients. However, if the experimenters have no control over the marginal weights of the gender factor then Lemma 4.1.1 can be applied.

6.6 Discussion

In this chapter, we examine a practical quadratic mixed responses model that is useful in the clinical trial experiments. The required number of support points based on the complete class approach for this model was determined by at most 6-support points, including the two boundary points of the design region. We then showed how we could extend the quadratic mixed responses model to include the group effects. The proposed hybrid PSO-FM algorithm was used to search for optimal designs.

As an application on our approach for finding the optimal designs, we provide a case study about the dose-finding problem for a new drug Ridaforolimus that treats breast cancer (Biswas & López-Fidalgo, 2013). Based on our complete class for this model, we obtain the same D-optimal design as Biswas and López-Fidalgo (2013).

Chapter 7

CONCLUDING REMARKS

In this dissertation, the focus was on the bivariate mixed responses data. While there has been many models for the experiments that measure mixtures of discrete and continuous responses, there is very little research on designing such experiments. To address this scarcity, we provide some of our research results on identifying locally optimal designs for some mixed responses models. We used the complete class approach that can restrict the search for the optimal design to a small class which contains designs with a small number of support points. The optimal design is selected based on minimizing the variance-covariance matrix of the parameter estimates, and its optimality is verified by the GET.

In Chapter 3, we considered the simple mixed responses model, where we identified a symmetric complete class. The complete class contained candidate symmetric designs with at most four support points including the two endpoints. We showed how to determine D-optimal designs within the class of symmetric designs where only one support point and one weight are required to be determined. For implementation of our results, we considered the constrained nonlinear algorithm, the fmincon solver in MATLAB, to search for the D-optimal designs.

The mixed responses model with group effects was considered next and was covered in Chapter 4 and Chapter 5. We studied the mixed responses model with group effects for two different cases; the first case is the model that has no common parameters across subject groups, and the second case is the model with a common parameter (i.e. the slope for x) for all the groups. For the former Model I, we showed how the block diagonal structure of its information matrix allows the search for the optimal design to be divided into the search within each individual group. With this finding, the fmincon solver, and many other algorithms were able to handle the search for the optimal designs.

Unlike Model I, the information matrix for the latter Model II is rather complex, and the search for the optimal design is rather involved. Thus, the hybrid optimization algorithm, the PSO algorithm followed by the fmincon solver, was proposed to deal with these kinds of models. The performance of the hybrid PSO-FM algorithm was shown to work well in all the cases that we considered. The idea of combining two algorithms in this fashion was also seen in complex situations such as in Gueorguieva et al. (2006) where a hybrid algorithm was used to search for the D-optimal design under a multivariate response Pharmacokinetic models.

Finally, we extended our results to a quadratic mixed responses model. The quadratic model introduced here was used in breast cancer dose-finding experiments. We found a complete class for the quadratic mixed responses model which was formed by designs with at most six support points including the endpoints of the design region. Our approach also allowed some factors such as age and sex to be included in the model.

The results we obtained in this research can serve as a guide for other similar mixed responses models and their extensions. For future work, it would be interesting to investigate the case where the mixed responses models with group effects contain only the main effects and some interactions of certain order instead of the full factorial setting. Also, there is still a lack of efficient, general, and consistent optimization algorithm that performs well for high dimensional cases. Therefore, an important further research is to develop an algorithm that is easy to adjust, efficient for different models, and can work for mixed responses and mixed covariates design problems. An extension of our results to cases with two or more continuous covariates is another challenging problem for future research.

REFERENCES

Aerts, M., Molenberghs, G., Geys, H., & Ryan, L. M. (2002). *Topics in modelling of clustered data*. Chapman and Hall/CRC.

Arminger, G., & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In *Latent trait and latent class models* (pp. 51–73). Springer.

Atashpaz-Gargari, E., & Lucas, C. (2007). Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In *Evolutionary computation, 2007. cec 2007. ieee congress on* (pp. 4661–4667).

Atkinson, A., Donev, A., & Tobias, R. (2007). *Optimum experimental designs, with sas* (Vol. 34). Oxford University Press.

Berger, M. P., & Wong, W.-K. (2009). *An introduction to optimal designs for social and biomedical research* (Vol. 83). John Wiley & Sons.

Biswas, A., & López-Fidalgo, J. (2013). Compound designs for dose-finding in the presence of nondesignable covariates. *Pharmaceutical statistics*, *12*(2), 92–101.

Böhning, D. (1986). A vertex-exchange-method in d-optimal design theory. *Metrika*, *33*(1), 337–347.

Catalano, P. J., & Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, *87*(419), 651–658.

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

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied statistics*, 113–120.

Cox, D. R., & Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika*, *79*(3), 441–461.

De la Garza, A., et al. (1954). Spacing of information in polynomial regression. *The Annals of Mathematical Statistics*, *25*(1), 123–130.

De Leon, A. R., & Chough, K. C. (2013). *Analysis of mixed data: methods & applications*. CRC Press.

Deng, X., & Jin, R. (2015). Qq models: joint modeling for quantitative and qualitative quality responses in manufacturing systems. *Technometrics*, *57*(3), 320–331.

Dette, H., & Melas, V. B. (2011). A note on the de la garza phenomenon for locally optimal designs. *The Annals of Statistics*, *39*(2), 1266–1281.

Dobson, A. J., & Barnett, A. (2008). *An introduction to generalized linear models*. CRC press.

Dorigo, M. (1992). "optimization, learning and natural algorithms", unpublished ph.d. dissertation,university of politecnico di milano, italy,(1992).

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(2), 355–366.

Eberhart, R. C., & Shi, Y. (2000). Comparing inertia weights and constriction factors in particle swarm optimization. In *Evolutionary computation, 2000. proceedings of the 2000 congress on* (Vol. 1, pp. 84–88).

Ehrenfeld, S., et al. (1956). Complete class theorems in experimental design. In *Proceedings of the third berkeley symposium on mathematical statistics and probability, volume 1: Contributions to the theory of statistics.*

Fedorov. (1972). *Theory of optimal experiments*. New York: Academic.

Fedorov, & Leonov. (2013). *Optimal design for nonlinear response models*. CRC Press.

Fedorov, Wu, Y., & Zhang, R. (2012). Optimal dose-finding designs with correlated continuous and discrete responses. *Statistics in medicine*, *31*(3), 217–234.

Fitzmaurice, G. M., & Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American statistical Association*, *90*(431), 845–852.

Fitzmaurice, G. M., & Laird, N. M. (1997). Regression models for mixed discrete and continuous responses with potentially missing values. *Biometrics*, 110–122.

Gueorguieva, I., Aarons, L., Ogungbenro, K., Jorga, K. M., Rodgers, T., & Rowland, M. (2006). Optimal design for multivariate response pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics*, *33*(2), 97.

Holland, J. H. (1975). Adaptation in natural and artificial systems. an introductory analysis with application to biology, control, and artificial intelligence. *Ann Arbor, MI: University of Michigan Press*, 439–444.

Hu, L., Yang, M., & Stufken, J. (2015). Saturated locally optimal designs under differentiable optimality criteria. *The Annals of Statistics*, *43*(1), 30–56.

Hyun, S. W. (2013). Optimal designs for a probit model with a quadratic term. *Statistics in Biopharmaceutical Research*, *5*(1), 18–26.

Karlin, S., & Studden, W. J. (1966). *Tchebycheff systems: With applications in analysis and statistics*. Interscience New York.

Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings of the ieee international conference on neural networks* (Vol. 4, pp. 1942–1948).

Khuri, A. I., Mukherjee, B., Sinha, B. K., & Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science*, 376–399.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 272–319.

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

Kiefer, J., & Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 271–294.

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

Kim, S. (2017). "optimal experimental design for mixed categorical and continuous responses", unpublished ph.d. dissertation,university of arizona(2017).

Kim, S., & Kao, M.-H. (2019). Locally optimal designs for mixed binary and continuous responses. *Statistics & Probability Letters*.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, *220*(4598), 671–680.

Lei, X., Yuan, Y., & Yin, G. (2011). Bayesian phase ii adaptive randomization by jointly modeling time-to-event efficacy and binary toxicity. *Lifetime data analysis*, *17*(1), 156–174.

Liski, E. P., Mandal, N. K., Shah, K. R., & Sinha, B. K. (2002). *Topics in optimal design* (Vol. 163). Springer Science & Business Media.

Lukemire, J., Mandal, A., & Wong, W. K. (2018). d-qpso: A quantum-behaved particle swarm technique for finding d-optimal designs with discrete and continuous factors and a binary response. *Technometrics*(just-accepted), 1–27.

Mandal, W. W. Y., A., et al. (2015). *Handbook of design and analysis of experiments* (Vol. 7). CRC Press.

Mathew, T., & Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference*, *93*(1), 295–307.

McCullagh, P., & Nelder, J. (1989). Binary data. In *Generalized linear models* (pp. 98–148). Springer.

Olkin, I., Tate, R. F., et al. (1961). Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics*, *32*(2), 448–465.

Pearson, K. (1900). Mathematical contributions to the theory of evolution. viii. on the correlation of characters not quantitatively measurable. *Proceedings of the Royal Society of London*, *66*(424-433), 241–244.

Pukelsheim, F. (1989). Complete class results for linear regression designs over the multi-dimensional cube. In *Contributions to probability and statistics* (pp. 349–356). Springer.

Pukelsheim, F. (2006). *Optimal design of experiments.* SIAM.

Pukelsheim, F., & Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika*, *79*(4), 763–770.

Qiu, J., Chen, R.-B., Wang, W., & Wong, W. K. (2014). Using animal instincts to design efficient biomedical studies via particle swarm optimization. *Swarm and evolutionary computation*, *18*, 1–10.

Rabe-Hesketh, S., & Skrondal, A. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models.* Chapman and Hall/CRC.

Ratnaweera, A., Halgamuge, S. K., & Watson, H. C. (2004). Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *IEEE Transactions on evolutionary computation*, *8*(3), 240–255.

Sammel, M. D., Ryan, L. M., & Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(3), 667–678.

Schorning, K., Dette, H., Kettelhake, K., Wong, W., & Bretz, F. (2017). Optimal designs for active controlled dose-finding trials with efficacy-toxicity outcomes. *Biometrika*, *104*(4), 1003–1010.

Schwabe, R. (2012). *Optimum designs for multi-factor models* (Vol. 113). Springer Science & Business Media.

Silvey, S., Titterington, D., & Torsney, B. (1978). An algorithm for optimal designs on a design space. *Communications in Statistics-Theory and Methods*, *7*(14), 1379–1389.

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 observations. *Biometrika*, *12*(1/2), 1–85.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341–359.

Stufken, J., & Yang, M. (2012a). On locally optimal designs for generalized linear models with group effects. *Statistica Sinica*, 1765–1786.

Stufken, J., & Yang, M. (2012b). Optimal designs for generalized linear models. *Design and Analysis of Experiments, Special Designs and Applications*, *3*, 137.

Tan, X. (2015). "optimal designs for generalized linear models", unpublished ph.d. dissertation,university of georgia(2016).

Tan, X., & Stufken, J. (2016). "optimal designs for generalized linear models", unpublished ph.d. dissertation,university of georgia(2016).

Teixeira-Pinto, A., & Normand, S.-L. T. (2009). Correlated bivariate continuous and binary outcomes: issues and applications. *Statistics in medicine*, *28*(13), 1753–1773.

Wang, Z. (2018). "locally d-optimal designs for generalized linear models", unpublished ph.d. dissertation,university of arizona(2016).

Wong, W. K., Yin, Y., & Zhou, J. (2018). Optimal designs for multi-response nonlinear regression models with several factors via semidefinite programming. *Journal of Computational and Graphical Statistics*, 1–13.

Wu. (1978). Some algorithmic aspects of the theory of optimal designs. *The Annals of Statistics*, 1286–1301.

Wu, & Stufken. (2014). Locally ϕ p-optimal designs for generalized linear models with a single-variable quadratic polynomial predictor. *Biometrika*, *101*(2), 365–375.

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

Yang. (2008). A-optimal designs for generalized linear models with two parameters. *Journal of Statistical Planning and Inference*, *138*(3), 624–641.

Yang. (2010a). *Nature-inspired metaheuristic algorithms*. Luniver press.

Yang. (2010b). On the de la garza phenomenon. *The Annals of Statistics*, *38*(4), 2499–2524.

Yang, Biedermann, S., & Tang, E. (2013). On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association*, *108*(504), 1411–1420.

Yang, & Mandal, A. (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics-Simulation and Computation*, *44*(9), 2264–2277.

Yang, Mandal, A., & Majumdar, D. (2012). Optimal designs for two-level factorial experiments with binary response. *Statistica Sinica*, 885–907.

Yang, Mandal, A., & Majumdar, D. (2016). Optimal designs for 2ˆ k factorial experiments with binary response. *Statistica Sinica*, 385–411.

Yang, & Stufken. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *The Annals of Statistics*, 518–541.

Yang, & Stufken. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *The Annals of Statistics*, *40*(3), 1665–1681.

Yates, F. (1935). Complex experiments. *Supplement to the Journal of the Royal Statistical Society*, *2*(2), 181–247.

Yu, Y. (2011). D-optimal designs via a cocktail algorithm. *Statistics and Computing*, *21*(4), 475–481.

Zhang, L.-X., & Hu, F.-f. (2009). A new family of covariate-adjusted response adaptive designs and their properties. *Applied Mathematics-A Journal of Chinese Universities*, *24*(1), 1–13.

APPENDIX A

ADDITIONAL PROOFS

Proof of Lemma 3.2.2. Under a simple mixed response model in (3.1) with information matrix defined as in (3.4), for any arbitrary design $\xi = \{(c_i, w_i), \sum_{i=1}^{m} w_i = 1, i = 1, \cdots m\}$ we can find a reflected design $\xi_r = \{(-c_i, w_i), i = 1, \cdots m\}$, and the symmetrized design is $\xi_s = \frac{1}{2}(\xi + \xi_r)$, which implies that $\mathrm{M}(\xi_s, \theta) = \frac{1}{2}\mathrm{M}(\xi, \theta) + \frac{1}{2}\mathrm{M}(\xi_r, \theta)$ and $\tilde{C}(\xi_s, \theta) = \frac{1}{2}\tilde{C}(\xi, \theta) + \frac{1}{2}\tilde{C}(\xi_r, \theta)$. Since Φ_D is a convex function and both B and $\tilde{C}(\xi, \theta)$ are nonsingular matrices, then

$$
\begin{aligned}
-\log|\mathrm{M}(\xi_s, \theta)| &\leq -\frac{1}{2}\log|\mathrm{M}(\xi, \theta)| - \frac{1}{2}\log|\mathrm{M}(\xi_r, \theta)| \\
&= -\frac{1}{2}\log|\mathrm{M}(\xi, \theta)| - \frac{1}{2}\log|BG\tilde{C}(\xi, \theta)G^{\mathrm{T}}B^{\mathrm{T}}| \\
&= -\frac{1}{2}\log|\mathrm{M}(\xi, \theta)| - \frac{1}{2}(\log|B| + \log|G\tilde{C}(\xi, \theta)G^{\mathrm{T}}| + \log|B^{\mathrm{T}}|) \\
&= -\frac{1}{2}\log|\mathrm{M}(\xi, \theta)| - \frac{1}{2}(\log|B| + \log|\tilde{C}(\xi, \theta)| + \log|B^{\mathrm{T}}|) \\
&= -\frac{1}{2}\log|\mathrm{M}(\xi, \theta)| - \frac{1}{2}\log|B\tilde{C}(\xi, \theta)B^{\mathrm{T}}| \\
&= -\frac{1}{2}\log|\mathrm{M}(\xi, \theta)| - \frac{1}{2}\log|\mathrm{M}(\xi, \theta)| = -\log|\mathrm{M}(\xi, \theta)|,
\end{aligned}
$$

where G is an orthogonal matrix defined in(3.6). Thus, we can conclude that $\Phi_D(\mathrm{M}(\xi_s, \theta)) \leq \Phi_D(\mathrm{M}(\xi, \theta))$. □

Proposition1: The function $f(c) = \dfrac{e^{5c} - 4ce^{3c} - e^c}{(e^c + 1)^2(e^{2c} + 2ce^c - 1)^2} > 0$ for c $\in (0, \infty)$.

Proof. It is clear that $\dfrac{e^c}{(e^c + 1)^2(e^{2c} + 2ce^c - 1)^2}$ is positive for $\forall c > 0$. Let $v(c) = e^{4c} - 4ce^{2c} - 1$, than the first derivative of $v(c)$ is $v^{'}(c) = 4e^{2c}(e^{2c} - 2c - 1) = 4e^{2c}u(c)$ where, $u(c) = e^{2c} - 2c - 1$. Clearly, $4e^{2c} > 0\ \forall c$. $u(c)$ has a root at $c = 0$, since $u(0) = 0$. To show that $u(c) > 0$, suppose $u(c)$ has two roots c_1 and c_2, thus $u(c_1) = u(c_2) = 0$. From Rolle's theorem, there exists a $c_3 \in (c_1, c_2)$ such that $u^{'}(c_3) = \dfrac{u(c_2) - u(c_1)}{c_2 - c_1} = 0$. But, $u^{'}(c) = 2e^{2c} - 2 = 2(e^{2c} - 1) > 0\ \forall c > 0$. By contradiction, $u(c)$ has only a unique root at $c = 0$. Thus, $v^{'}(c)$ has only one root at $c = 0$. Also, for $\forall c > 0$ $v^{'}(c)$ is positive since $v^{'}(1) = 129.72 > 0$. Hence, $v(c)$ is an increasing function for $c > 0$. Therefore, we can conclude that $f(c) > 0\ \forall c > 0$. □

Proposition2: The function $f(c) = \dfrac{16(\cosh(c)+1)(\sinh^2(c)-c^2)}{(2c-\sinh(2c))^2}[2c\cosh(c)+\cosh(c)\sinh(c)-c-2\sinh(c)] > 0$ for $c \in (0,\infty)$.

Proof. Clearly, $\dfrac{16(\cosh(c)+1)}{(2c-\sinh(2c))^2} > 0\ \forall c > 0$. Also, $\sinh^2(c)-c^2 > 0$, since $\sinh(c)+c > 0\ \forall c > 0$ and $\sinh(c)-c > 0\ \forall c > 0$. Let $u(c) = 2c\cosh(c)+\cosh(c)\sinh(c)-c-2\sinh(c)$, then $u^{'}(c) = 2\sinh(c)[c+\sinh(c)] = 2\sinh(c)v(c)$. Clearly, $2\sinh(c) > 0$ $\forall c > 0$. $v(c)$ has a root at $c = 0$, since $v(0) = 0$. Suppose that $v(c)$ has two roots c_1 and c_2, thus $v(c_1) = v(c_2) = 0$. From Rolle's theorem, there exists a $c_3 \in (c_1, c_2)$ such that $v^{'}(c_3) = \dfrac{v(c_2)-v(c_1)}{c_2-c_1} = 0$. But $v^{'}(c) = \cosh(c)+1 > 0\ \forall c$. By contradiction, $v(c)$ has only a unique root at $c = 0$. Thus $u^{'}(c)$ has only one root at $c = 0$. Also, for $\forall c > 0, u^{'}(c)$ is positive since $u^{'}(1) = 5.1126 > 0$. Hence, $v(c)$ is an increasing function $\forall c > 0$. Therefore, the conclusion $f(c) > 0\ \forall c > 0$ follows. □

Proposition3: The function $f(c) = \dfrac{16\cosh^2(c/2)(c+\sinh(c))}{(2c-\sinh(2c))^2}\left[c^2-2\sinh^2(c)+\dfrac{c\sinh(2c)}{2}\right] > 0$ for c $\in (0,\infty)$.

Proof. Clearly, $\dfrac{16\cosh^2(c/2)(c+\sinh(c))}{(2c-\sinh(2c))^2}\ \forall\ c > 0$. Let $u(c) = c^2-2\sinh^2(c)+\dfrac{c\sinh(2c)}{2}$. Taking first, second, third, and fourth derivatives of $u(c)$ yields: $u^{'}(c) = c-\dfrac{3}{2}\sinh(2c)+2c\cosh^2(c)$, $u^{''}(c) = 2c\sinh(2c)-4\sinh^2(c)$, $u^{'''}(c) = 4c\cosh(2c)-2\sinh(2c)$, and $u^{iv}(c) = 8c\sinh(2c)$, respectively. It is clear that $u^{iv}(c) > 0\ \forall c > 0$. $u^{'''}(c)$ has a root at $c = 0$ since $u^{'''}(0) = 0$. Suppose that $u^{'''}(c)$ has two roots c_1 and c_2, thus $u^{'''}(c_1) = u^{'''}(c_2) = 0$. From Rolle's theorem, there exists a $c_3 \in (c_1, c_2)$ such that $u^{iv}(c_3) = \dfrac{u^{'''}(c_2)-u^{'''}(c_1)}{c_2-c_1} = 0$. But, as shown above, $u^{iv}(c) > 0\ \forall c > 0$.Thus $u^{'''}(c)$ has only a unique root at $c = 0$. Also, for $c > 0$ $u^{'''}(c)$ is positive since $u^{'''}(1) = 7.795$. Hence $u^{'''}(c)$ is increasing function $\forall c > 0$. Similarly, $u^{''}(c)$ has a root at $c = 0$ since $u^{''}(0) = 0$. Suppose $u^{''}(c)$ has two roots c_1 and c_2, thus $u^{''}(c_1) = u^{''}(c_2) = 0$. From Rolle's theorem, there exists a $c_3 \in (c_1, c_2)$ such that $u^{'''}(c_3) = \dfrac{u^{''}(c_2)-u^{''}(c_1)}{c_2-c_1} = 0$. However, $u^{'''}(c)$ is proved to be positive and increasing $\forall c > 0$. Thus $u^{''}(c)$ has only a unique root at $c = 0$. Also, for $c > 0$, $u^{''}(c)$ is positive since $u^{''}(1) = 1.729$. Hence, $u^{''}(c)$ is an increasing function $\forall c > 0$. Following the same procedure, $u^{'}(c)$ has a root at $c = 0$ since $u^{'}(0) = 0$. Suppose $u^{'}(c)$ has two roots c_1 and c_2, thus $u^{'}(c_1) = u^{'}(c_2) = 0$. From Rolle's theorem, there exists a $c_3 \in (c_1, c_2)$ such that $u^{''}(c_3) = \dfrac{u^{'}(c_2)-u^{'}(c_1)}{c_2-c_1} = 0$. But $u^{''}(c)$ is proved positive and increasing $\forall c > 0$.

Thus $u^{'}(c)$ has only a unique root at $c = 0$. Also, $\forall c > 0$, $u^{'}(c)$ is positive since $u^{'}(1) = 0.322$. Hence, $u^{'}(c)$ is an increasing function $\forall c > 0$. Finally, $u(c)$ has a root at $c = 0$ since and $u(0) = 0$. Suppose $u(c)$ has two roots c_1 and c_2, thus $u(c_1)$ and $u(c_2) = 0$. From Rolle's theorem, there exists a $c_3 \in (c_1, c_2)$ such that $u^{'}(c_3) = \dfrac{u(c_2) - u(c_1)}{c_2 - c_1} = 0$. But $u^{'}(c)$ is proved to be positive and increasing $\forall c > 0$. Thus $u(c)$ has only a unique root at $c = 0$. Also, $u(c)$ is an increasing positive function since $u(2) = 4.982$. Hence, $u(c)$ is an increasing function $\forall c > 0$.
Therefore, we can conclude that $f(c) > 0 \ \forall c > 0$. $\square$

Proof of Lemma 4.1.2. By Equation (4.7),
let $f(w_l) = \sum_{l=1}^{s} w_l^{-q} \sum_{i=1}^{p_l} \lambda_i [M_l^{-1}(\tau_l)]^q, \quad \text{and } g(w_l) = \sum_{l=1}^{s} w_l - 1.$
Taking the partial derivatives with respect to w_l for both $f(w_l)$ and $g(w_l)$, then

$$\begin{aligned}\frac{\partial f}{\partial w_l} &= -q \sum_{l=1}^{s} w_l^{-q-1} \sum_{i=1}^{p_l} \lambda_i [M_l^{-1}(\tau_l)]^q \\ &= -q\{w_1^{-q-1} \sum_i \lambda_i [M_1^{-1}(\tau_1)]^q + \cdots + w_s^{-q-1} \sum_i \lambda_i [M_s^{-1}(\tau_s)]^q\},\end{aligned}$$

$\dfrac{\partial(\mu * g)}{\partial w_l} = \mu$, where μ is the Lagrange multiplier. Then set
$\dfrac{\partial f}{\partial w_l} = \dfrac{\partial(\mu * g)}{\partial w_l}$, therefore $w_1^{-q-1} \sum_i \lambda_i [M_1^{-1}(\tau_1)]^q = \cdots = w_s^{-q-1} \sum_i \lambda_i [M_s^{-1}(\tau_s)]^q$.

For any w_k, $\quad \left(\dfrac{w_1}{w_k}\right)^{-q-1} = \dfrac{\sum_i \lambda_i [M_k^{-1}(\tau_k)]^q}{\sum_i \lambda_i [M_1^{-1}(\tau_1)]^q}$,

then

$$w_k = w_1 \frac{\{\sum_i \lambda_i [M_k^{-1}(\tau_k)]^q\}^{1/q+1}}{\{\sum_i \lambda_i [M_1^{-1}(\tau_1)]^q\}^{1/q+1}}. \tag{A.1}$$

Since $\sum_{k=1}^{s} w_k = 1$, then $\sum_{k=1}^{s} \left\{ w_1 \dfrac{\{\sum_i \lambda_i [M_k^{-1}(\tau_k)]^q\}^{1/q+1}}{\{\sum_i \lambda_i [M_1^{-1}(\tau_1)]^q\}^{1/q+1}} \right\} = 1$,

by solving for w_1 we get; $w_1 = \dfrac{\{\sum_i \lambda_i [M_1^{-1}(\tau_1)]^q\}^{1/q+1}}{\sum_{k=1}^{s} \{\sum_i \lambda_i [M_k^{-1}(\tau_k)]^q\}^{1/q+1}}$.

Substituting w_1 into equation (A.1) we get

$$w_l = \frac{\{\sum_{i=1}^{p_l} \lambda_i[M_l^{-1}(\tau_l)]^q\}^{1/(q+1)}}{\sum_{l=1}^{s}\{\sum_{i=1}^{p_l} \lambda_i[M_l^{-1}(\tau_l)]^q\}^{1/(q+1)}}.$$

$\square$

Proof of Lemma 4.2.3. Suppose that there exists an induced symmetric design region for any number of groups in model (4.11). The information matrix in (4.12) can be written as

$$\begin{aligned} M(\xi, \boldsymbol{\theta}) &= \bigoplus_{l=1}^{s} B_l(\boldsymbol{\theta})\tilde{C}_l(\xi, \boldsymbol{\theta})B_l^T(\boldsymbol{\theta}), \\ \text{where } \tilde{C}_l(\xi, \boldsymbol{\theta}) &= \sum_{j=1}^{m_l} w_{l,j} C_l(\boldsymbol{\theta}, c_{lj}) = \bigoplus_{r=0}^{2} \tilde{C}_{lr}, \qquad \tilde{C}_{lr} = \begin{pmatrix} \eta_{lr} & \eta c_{lr} \\ \eta c_{lr} & \eta c_{lr}^2 \end{pmatrix}, \\ \text{and } \eta_{lr} &= \sum_{j=1}^{m_l} w_{l,j}\Gamma_r(c_{lj}), \eta c_{lr} = \sum_{j=1}^{m_l} w_{l,j} c_{lj}\Gamma_r(c_{lj}), \eta c_{lr}^2 = \sum_{j=1}^{m_l} w_{l,j} c_{lj}^2 \Gamma_r(c_{lj}). \end{aligned} \tag{A.2}$$

Based on Corollary 4.1.3 we can determine the optimal conditional measures separately, by working with the block information matrix that corresponded to the group with the symmetric induced design region. Hence, The information matrix for a conditional measure for group l is $M(\tau_l, \boldsymbol{\theta}) = B_l(\boldsymbol{\theta})\tilde{C}(\tau_l, \boldsymbol{\theta})B_l^T(\boldsymbol{\theta})$. Therefore, $\Phi_D = \det(M(\tau_l, \boldsymbol{\theta})) = |B_l(\boldsymbol{\theta})||\tilde{C}(\tau_l, \boldsymbol{\theta})||B_l^T(\boldsymbol{\theta})|$, note that the selection of the design is not affected by the B_l matrix. Then,
$|\tilde{C}(\tau_l, \boldsymbol{\theta})| = |\tilde{C}_{l0}||\tilde{C}_{l1}||\tilde{C}_{l2}| = \prod_{r=0}^{2} |\tilde{C}_{lr}|$, where $|\tilde{C}_{lr}| = \eta_{lr} * \eta c_{lr}^2 - (\eta c_{lr})^2$.

Using Hadamard Inequality, which states that the determinant of a matrix is less than or equal to the products of its diagonal entries. Thus,

$$\begin{aligned} |\tilde{C}(\tau_l, \boldsymbol{\theta})| = |\tilde{C}_{l0}||\tilde{C}_{l1}||\tilde{C}_{l2}| &\le |\tilde{C}_{l0}||\tilde{C}_{l1}| * \eta_{l2} * \eta c_{l2}^2 \\ &= [\eta_{l0} * \eta c_{l0}^2 - (\eta c_{l0})^2] * [\eta_{l1} * \eta c_{l1}^2 - (\eta c_{l1})^2] * \eta_{l2} * \eta c_{l2}^2. \end{aligned}$$

Based on Theorem 3.2.4 where we showed that under a simple mixed response model, the equality holds between any symmetric conditional measure τ_l and the optimal symmetric conditional measure τ_l^* based on equations given in (2.8). With the assumption that the lth group has a symmetric induced design range, thus the symmetric optimal conditional measure has the form $\{(-D_l, w_{1|l} = w_{lD}), (-c_l, w_{2|l} = w_{lc}), (c_l, w_{3|l} = w_{lc}), (D_l, w_{4|l} = w_{lD})\}$, with the weights w_{lD} for the points $\pm D_l$, and w_{lc} for the points $\pm c_l$. Using the fact that $\Gamma_0(c_{lj}) + \Gamma_1(c_{lj}) = 1, \Gamma_0(-c_{lj}) = \Gamma_1(c_{lj}), \Gamma_1(-c_{lj}) = \Gamma_0(c_{lj}),$ and $\Gamma_2(c_{lj})$ is an even function, then the following can be computed as:

$$
\begin{aligned}
|\tilde{C}_{l0}| =&[\eta_{l0} * \eta c_{l0}^2 - (\eta c_{l0})^2] \\
=&\{w_{lD}[\Gamma_0(-D_l) + \Gamma_0(D_l)] + w_{lc}[\Gamma_0(-c_l) + \Gamma_0(c_l)]\}\times \\
&\{w_{lD}D_l^2[\Gamma_0(-D_l) + \Gamma_0(D_l)] + w_{lc}c_l^2[\Gamma_0(-c_l) + \Gamma_0(c_l)]\}- \\
&\{w_{lD}D_l[\Gamma_0(D_l) - \Gamma_0(-D_l)] + w_{lc}c_l[\Gamma_0(c_l) - \Gamma_0(-c_l)]\}^2, \\
=&\{w_{lD}[\Gamma_1(D_l) + \Gamma_0(D_l)] + w_{lc}[\Gamma_1(c_l) + \Gamma_0(c_l)]\}\times \\
&\{w_{lD}D_l^2[\Gamma_1(D_l) + \Gamma_0(D_l)] + w_{lc}c_l^2[\Gamma_1(c_l) + \Gamma_0(c_l)]\}- \\
&\{w_{lD}D_l[\Gamma_0(D_l) - \Gamma_1(D_l)] + w_{lc}c_l[\Gamma_0(c_l) - \Gamma_1(c_l)]\}^2, \\
=&[w_{lD} + w_{lc}][w_{lD}D_l^2 + w_{lc}c_l^2] - \{w_{lD}D_l[\Gamma_0(D_l) - \Gamma_1(D_l)] + w_{lc}c_l[\Gamma_0(c_l) - \Gamma_1(c_l)]\}^2, \\
=&\frac{1}{2}[w_{lD}D_l^2 + w_{lc}c_l^2] - \{w_{lD}D_l[\Gamma_1(D_l) - \Gamma_0(D_l)] + w_{lc}c_l[\Gamma_1(c_l) - \Gamma_0(c_l)]\}^2.
\end{aligned}
$$

Similarly, for $|\tilde{C}_{l1}| = \frac{1}{2}[w_{lD}D_l^2 + w_{lc}c_l^2] - \{w_{lD}D_l[\Gamma_1(D_l) - \Gamma_0(D_l)] + w_{lc}c_l[\Gamma_1(c_l) - \Gamma_0(c_l)]\}^2$.

$$
\begin{aligned}
|\tilde{C}_{l2}| \leq&\eta_{l2} * \eta c_{l2}^2 \\
=&\{w_{lD}[\Gamma_2(-D_l) + \Gamma_2(D_l)] + w_{lc}[\Gamma_2(-c_l) + \Gamma_2(c_l)]\}\times \\
&\{w_{lD}D_l^2[\Gamma_2(-D_l) + \Gamma_2(D_l)] + w_{lc}c_l^2[\Gamma_2(-c_l) + \Gamma_2(c_l)]\}, \\
=&4[w_{lD}\Gamma_2(D_l) + w_{lc}\Gamma_2(c_l)][w_{lD}D_l^2\Gamma_2(D_l) + w_{lc}c_l^2\Gamma_2(c_l)].
\end{aligned}
$$

Hence, the upper bound for $|\tilde{C}(\tau_l, \boldsymbol{\theta})|$ is,

$$
\begin{aligned}
|\tilde{C}(\tau_l, \boldsymbol{\theta})| \leq& 4\{\frac{1}{2}[w_{lD}D_l^2 + w_{lc}c_l^2] - \{w_{lD}D_l[\Gamma_1(D_l) - \Gamma_0(D_l)] + w_{lc}c_l[\Gamma_1(c_l) - \Gamma_0(c_l)]\}^2\}^2\times \\
&\{[w_{lD}\Gamma_2(D_l) + w_{lc}\Gamma_2(c_l)][w_{lD}D_l^2\Gamma_2(D_l) + w_{lc}c_l^2\Gamma_2(c_l)]\}.
\end{aligned}
\tag{A.3}
$$

For an appropriate chosen points of c_l and D_l, and for a symmetric conditional measure of the form $\{(c_{l1} = -D_l, w_{1|l} = \frac{1-w_l}{2}), (c_{l2} = -c_l, w_{2|l} = \frac{w_l}{2}), (c_{l3} = c_l, w_{3|l} = \frac{w_l}{2}), (c_{l4} = D_l, w_{4|l} = \frac{1-w_l}{2})\}$, the determent of $|\tilde{C}(\tau_l, \boldsymbol{\theta})| = |\tilde{C}_{l0}||\tilde{C}_{l1}||\tilde{C}_{l2}|$ can be found by substituting the symmetric conditional measure for group l directly as follows:

$$
\begin{aligned}
|\tilde{C}(\tau_l, \boldsymbol{\theta})| =&4\{\frac{1}{4}[(1-w_l)D_l^2 + w_lc_l^2] - \frac{1}{4}\{(1-w_l)D_l[\Gamma_1(D_l) - \Gamma_0(D_l)] + w_lc_l[\Gamma_1(c_l) - \Gamma_0(c_l)]\}^2\}\times \\
&\{\frac{1}{4}[(1-w_l)D_l^2 + w_lc_l^2] - \frac{1}{4}\{(1-w_l)D_l[\Gamma_1(D_l) - \Gamma_0(D_l)] + w_lc_l[\Gamma_1(c_l) - \Gamma_0(c_l)]\}^2\}\times \\
&\{\frac{1}{4}[(1-w_l)\Gamma_2(D_l) + w_l\Gamma_2(c_l)][(1-w_l)D_l^2\Gamma_2(D_l) + w_lc_l^2\Gamma_2(c_l)]\}, \\
=& \frac{1}{2^4}\{[(1-w_l)\Gamma_2(D_l) + w_l\Gamma_2(c_l)][(1-w_l)D_l^2\Gamma_2(D_l) + w_lc_l^2\Gamma_2(c_l)]\}\times \\
&\{[(1-w_l)D_l^2 + w_lc_l^2] - [(1-w_l)D_l(\Gamma_1(c_l) - \Gamma_0(c_l)) + w_lc_l(\Gamma_1(c_l) - \Gamma_0(c_l))]^2\}^2.
\end{aligned}
$$

Thus,

$$
|\tilde{C}(\tau_l, \boldsymbol{\theta})| = \frac{(E_l - H_l^2)^2}{2^4}\Big\{[(1-w_l)\Gamma_2(D_l) + w_l\Gamma_2(c_l)][(1-w_l)D_l^2\Gamma_2(D_l) + w_lc_l^2\Gamma_2(c_l)]\Big\}. \tag{A.4}
$$

Where $E_l = (1-w_l)D_l^2 + w_lc_l^2$, and $H_l = (1-w_l)D_l(\Gamma_1(D_l) - \Gamma_0(D_l)) + w_lc_l(\Gamma_1(c_l) - \Gamma_0(c_l))$. Therefore, the proposed symmetric conditional measure for group l reached the upper bound in equation (A.3).

□

Proposition4: The function $f(c) = 4\dfrac{321e^{2c} - 26e^{3c} + 16e^{4c} - 26e^{c} + 16}{(4e^{2c} - e^{c} + 4)^2} > 0\ \forall c \in \mathbb{R}$.

Proof. $f^{'}(c) = \dfrac{144\ \sinh(c)[16\ \sinh^2(c/2) - 119]}{[16\ \sinh^2(c/2) + 7]^3}$. The critical points of $f(c)$, zeros of $f^{'}(c)$, are -∞, 0, and ∞. $f(-\infty) = f(\infty) = 4$, which is a minimum points for $f(c)$, and $f(0) = 24.5714$, which is a maximum points for $f(c)$, so, $4 < f(c) \leq 24.5714$. Therefore, $f(c) > 0\ \forall c \in \mathbb{R}$. $\square$