

Locally D-optimal Designs for Generalized Linear Models

by

Zhongshen Wang

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

Approved April 2018 by the
Graduate Supervisory Committee:

John Stufken, Chair
Ioannis Kamarianakis
Ming-Hung Kao
Mark Reiser
Yi Zheng

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Generalized Linear Models (GLMs) are widely used for modeling responses with non-normal error distributions. When the values of the covariates in such models are controllable, finding an optimal (or at least efficient) design could greatly facilitate the work of collecting and analyzing data. In fact, many theoretical results are obtained on a case-by-case basis, while in other situations, researchers also rely heavily on computational tools for design selection.

Three topics are investigated in this dissertation with each one focusing on one type of GLMs. Topic I considers GLMs with factorial effects and one continuous covariate. Factors can have interactions among each other and there is no restriction on the possible values of the continuous covariate. The locally D-optimal design structures for such models are identified and results for obtaining smaller optimal designs using orthogonal arrays (OAs) are presented. Topic II considers GLMs with multiple covariates under the assumptions that all but one covariate are bounded within specified intervals and interaction effects among those bounded covariates may also exist. An explicit formula for D-optimal designs is derived and OA-based smaller D-optimal designs for models with one or two two-factor interactions are also constructed. Topic III considers multiple-covariate logistic models. All covariates are nonnegative and there is no interaction among them. Two types of D-optimal design structures are identified and their global D-optimality is proved using the celebrated equivalence theorem.

Keywords: equivalence theorem, orthogonal arrays, locally optimal designs, D-optimality

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere appreciation to my advisor, Dr. John Stufken, for his tremendous support and constant encouragement throughout my PhD study. I cannot imagine what my experience at ASU would have been without his mentorship. I'm truly grateful for everything he has done for me and it has been such a rewarding time working with him.

I would also like to thank my committee members: Dr. Mark Reiser, Dr. Ming-Hung Kao, Dr. Ioannis Kamarianakis and Dr. Yi Zheng for their valuable suggestions and feedback for my dissertation, as well as their continual help and professional advice for the past five years.

I would like to expend my appreciation to all other faculties, staff and graduate students in the School of Mathematical and Statistical Sciences and the Graduate Statistics Club for their assistance and kindness. They make my life at ASU colorful and enjoyable. I also want to take this opportunity to thank my school for financial support.

Finally, I would like to thank my parents, Yin Wang and Yafen Tian, for their endless love and companion. Without them, this dissertation would not have been possible. Special thanks to my fiancée, Yutong Lu, who never hesitates to help and encourage me to chase my dreams.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vii
CHAPTER	
1 INTRODUCTION	1
1.1 Generalized Linear Models	3
1.2 Optimal Designs	5
1.3 Orthogonal Arrays	7
1.4 Equivalence Theorem	9
2 TOPIC I: D-OPTIMAL DESIGNS FOR GLMS WITH FACTORIAL EFFECTS AND ONE COVARIATE	11
2.1 Model Description	13
2.2 Theoretical Results	14
2.3 Results Based on Orthogonal Arrays	21
2.4 Connections Between Resolution III Designs and Strength 2+ OAs .	32
2.5 Illustrative Examples	34
2.6 Summaries and Discussions	41
3 TOPIC II: D-OPTIMAL DESIGNS FOR GLMS WITH MULTIPLE COVARIATES	45
3.1 Model Description	47
3.2 Results for Main-effect Models	47
3.3 Main Results for Interaction Models	51
3.4 Smaller Optimal Designs Based on Orthogonal Arrays	53
3.5 Summaries and Discussions	58

CHAPTER	Page
4 TOPIC III: D-OPTIMAL DESIGNS FOR LOGISTIC MODLES WITH MULTIPLE COVARIATES	61
4.1 Model Description	63
4.2 D-optimal Designs with $p + 1$ Support Points	66
4.3 D-optimal Designs with $2p$ Support Points	72
4.4 Summaries and Discussions	79
5 CONCLUSION	83
REFERENCES	85
APPENDIX	
A ADDITIONAL PROOFS FOR TOPIC I	88
B ADDITIONAL PROOFS FOR TOPIC III.....	97

LIST OF TABLES

Table	Page
1.1 An $OA(4, 3, 2, 2)$	8
1.2 A Mixed Orthogonal Array.	9
2.1 Support Points for a Locally D-optimal Design.	21
2.2 c^* That Maximizes $c^2(\Psi(c))^r$ for Logistic and Probit Models.	21
2.3 An $OA(8, 2^4, 2+)$ with Columns (1,2,3) and (1,2,4) Being of Strength 3.	25
2.4 Smaller D-optimal Design Based on Orthogonal Arrays	25
2.5 An $OA(8, 2^{42}, 2+)$	27
2.6 Smaller D-optimal Design Based on Orthogonal Arrays.	27
2.7 A Transposed $OA(27, 3^4 3^1, 2+)$	29
2.8 A D-optimal Design Based on $OA(27, 3^4 3^1, 2+)$	29
2.9 A 2_{III}^{5-2} Design with Generators $I = ABD$ and $I = ACE$	33
2.10 Factors and Covariate for ESD Experiment.	35
2.11 D-optimal Design for the ESD Experiment Using a Full Factorial.	36
2.12 Smaller D-optimal Designs for the ESD Experiment Using Orthogonal Arrays.	37
2.13 8-point Threshold Design for the ESD Experiment.	38
2.14 Factors and Covariate for the PCB Experiment.	39
2.15 Original Design for the PCB Experiment.	40
2.16 D-optimal Design for the PCB Experiment Using a Full Factorial.	41
2.17 A Near D-optimal Design for the PCB Experiment.	42
2.18 A Transposed $OA(18, 2^1 3^3, 2+)$	42
2.19 A Smaller D-optimal Design for the PCB Experiment.	43
2.20 Another D-optimal Design for Example 2.2.5.	44
3.1 Support Points and Weights for a Locally D-optimal Design.	51

Table	Page
3.2 Support Points and Weights for a Locally D-optimal Design.	53
3.3 An $OA(8, 2^4, 2+)$ ($OA(8, 2^4, 3)$) with Columns (1,2,4) and (1,3,4) Being of Strength 3.	57
3.4 A Smaller D-optimal Design Based on Orthogonal Arrays	58
3.5 An $OA(8, 2^5, 2+)$ with Columns (1,2,4), (1,2,5), (1,3,4) and (1,3,5) Being of Strength 3.	59
3.6 A Smaller D-optimal Design Based on Orthogonal Arrays	59
4.1 Relationship Between c^* and β_0	72
4.2 D-optimal Design for Example 4.3.2 in Transformed Design Space.	80
4.3 D-optimal Design for Example 4.3.2 in the Original Design Space.	80

LIST OF FIGURES

Figure		Page
4.1	Pattern of a 4-point D-optimal Design for Three-covariate Logistic Model.	73
4.2	Sensitivity Plot for Design ξ_z^*	74
4.3	Pattern of a 6-point D-optimal Design for Three-covariate Logistic Model.	81
4.4	Sensitivity Plot for Design ξ_z^*	82

Chapter 1

INTRODUCTION

Experiments are conducted in a wide range of areas such as agriculture, marketing and pharmaceutical industries. For situations where the experimenter controls the factors and covariates, the evaluation of optimal or efficient designs requires extensive work. The study of optimal experimental designs dates at least back to the early 20th century when Smith (1918) proposed G-optimal designs for polynomial regression models. Since then, a lot of theories have been established but most of the early results focused on linear models. Nelder and Wedderburn (1972) introduced the concept of generalized linear models and proposed a unified likelihood-based procedure for fitting responses with non-normal error distributions. More methods and inferences for such models can be found in Agresti and Kateri (2011). There are a number of recent papers that have made significant contributions in this field both algebraically and computationally. Yang and Stufken (2009) studied locally optimal designs for GLMs with two parameters using a complete class approach. They extended this approach and obtained theoretical results for GLMs with group effects (Stufken and Yang (2012)). Hu and Stufken (2016a) identified locally optimal designs for a class of linear and nonlinear mixed effects models. For GLMs with more than one covariate, Jia and Myers (2001) identified D-optimal designs for two-variable logistic models with interactions. Yang *et al.* (2011) studied optimal designs for multiple covariates without interactions where all but one of the covariates are restricted. Hu and Stufken (2016b) identified complete classes of optimal designs for multiple-covariate GLMs with and without interactions. For computational studies, Yu (2011) proposed a “Cocktail algorithm” for searching D-optimal designs and proved that it outperforms

all existing algorithms in a situation where all parameters are of interest (Yu (2010)). Yang *et al.* (2013) proposed an optimal weight exchange algorithm (OWEA) which not only outperforms the Cocktail algorithm, but can also be applied to a broader class of optimality problems. Particle Swarm Optimization (PSO) is a relatively new meta-heuristic algorithm which has recently received considerable attention for finding optimal designs due to its simplicity and effectiveness (see Qiu *et al.* (2014); Chen *et al.* (2015)).

In spite of all these inspiring breakthroughs, there are still a number of open questions left. When considering GLMs with factorial effects, most of the literature focuses on main effects only whereas interactions among factors are also quite common in applications. For multiple-covariate models, not much work has been done when interaction effects are included in the models and an explicit expression of optimal designs for such models is in great need. The more complicated models where both factorial effects and several controllable covariates are presented are barely explored and require extensive studies.

This dissertation is organized as follows. We will briefly review some background knowledge and necessary notations in chapter 1 following this introduction. In chapter 2, we explore the GLMs with factorial effects and one continuous covariate. Chapter 3 and Chapter 4 both consider GLMs with multiple covariates, but they have different restrictions on the design space as well as the parameter settings. In Chapter 3 we assume that all but one covariates are bounded in intervals and interactions among bounded covariates may exist. In Chapter 4 we assume that all covariates are nonnegative and the model includes main effects only. Summaries and discussions are provided in the end.

1.1 Generalized Linear Models

Ordinary least squares (OLS) method is frequently used when the response variable follows a normal distribution. When this assumption is violated, for example, for binary data or count data, the traditional OLS method is no longer appropriate. Generalized Linear Models (GLM) consider a broader class of distributions which allows the response variable to have an error distribution other than normal. Such distributions include Binomial, Poisson, Inverse Gaussian, Gamma and so on. There are three components for a GLM:

1. random component — a probability distribution from the exponential family for response y with mean μ
2. systematic component — a linear predictor $\nu = \mathbf{X}^T \boldsymbol{\beta}$, where $\mathbf{X}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is an unknown parameter vector
3. a monotone differentiable link function g which connects the mean μ and linear predictor ν , i.e. $\nu = g(\mu)$

In this proposal, we focus on binary responses. Perhaps the simplest representation is of the form

$$Prob(Y_i = 1) = P(\alpha + \beta x_i) \quad (1.1)$$

where Y_i , being the response of subject i , can only take the values of 0 or 1; x_i is the covariate for subject i , α and β are intercept and slope parameters, $P(\cdot)$ is the cumulative distribution function (cdf). For logistic models, $P(x) = \frac{e^x}{e^x + 1}$; for probit models, $P(x) = \Phi(x)$, the cdf of the standard normal distribution.

Note model (1.1) does not take subject heterogeneity into consideration. To do so, assume there are L factors with number of levels $s_1, s_2, \dots, s_L$. This divides the

subjects into $s = s_1 \cdot s_2 \cdots s_L$ groups. We further assume that the model has a common slope for all subjects, then model (1.1) becomes

$$Prob(Y_{ij} = 1) = P(\alpha_0 + \alpha_i + \beta x_{ij}) \quad (1.2)$$

Formula (1.2) is not the only way to write this model, but it is a very simple and convenient presentation. Later on we will provide other presentations as well. Here Y_{ij} and x_{ij} represent the response and covariate for subject j in group i , $i = 1, \dots, s$, $j = 1, \dots, n_i$; α_0 is the grand mean, α_i is the effect for group i , β is the common slope parameter and $P(\cdot)$ as before, is the cumulative distribution function. For main-effects only models, if group i corresponds to level combination $(i_1, i_2, \dots, i_L)$, then α_i could be partitioned as $\alpha_i = \alpha_1^{i_1} + \alpha_2^{i_2} + \cdots + \alpha_L^{i_L}$ where $\alpha_l^{i_l}$ is the effect of the i_l^{th} level of factor l . But in general, α_i can also include interactions of two or more factors.

In vector notation, we can rewrite (1.2) as

$$Prob(Y_{ij} = 1) = P(\mathbf{X}_{ij}^T \boldsymbol{\theta}) \quad (1.3)$$

where $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \dots, \alpha_s, \beta)^T$ is a $(s+2) \times 1$ vector, $\mathbf{X}_{ij} = (1, \mathbf{X}_i^T, x_{ij})^T$ and $\mathbf{X}_i$ is a $s \times 1$ vector with 1 in position i and 0's elsewhere. Other notations are the same as before.

Assuming the independence of Y_{ij} , the likelihood function for $\boldsymbol{\theta}$ can be written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^s \prod_{j=1}^{n_i} P(\mathbf{X}_{ij}^T \boldsymbol{\theta})^{Y_{ij}} (1 - P(\mathbf{X}_{ij}^T \boldsymbol{\theta}))^{1-Y_{ij}} \quad (1.4)$$

To obtain the Fisher information matrix, take the second derivative of the log-likelihood function

$$\begin{aligned} \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial^2 \boldsymbol{\theta}} = \sum_{i=1}^s \sum_{j=1}^{n_i} & \left[Y_{ij} \frac{P''(\mathbf{X}_{ij}^T \boldsymbol{\theta}) P(\mathbf{X}_{ij}^T \boldsymbol{\theta}) - (P'(\mathbf{X}_{ij}^T \boldsymbol{\theta}))^2}{P^2(\mathbf{X}_{ij}^T \boldsymbol{\theta})} \mathbf{X}_{ij} \mathbf{X}_{ij}^T \right. \\ & \left. - (1 - Y_{ij}) \frac{P''(\mathbf{X}_{ij}^T \boldsymbol{\theta}) (1 - P(\mathbf{X}_{ij}^T \boldsymbol{\theta})) + (P'(\mathbf{X}_{ij}^T \boldsymbol{\theta}))^2}{[1 - P(\mathbf{X}_{ij}^T \boldsymbol{\theta})]^2} \mathbf{X}_{ij} \mathbf{X}_{ij}^T \right] \end{aligned} \quad (1.5)$$

So the information matrix can be computed as

$$I(\boldsymbol{\theta}) = E\left(-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial^2 \boldsymbol{\theta}}\right) = \sum_{i=1}^s \sum_{j=1}^{n_i} \frac{[P'(\mathbf{X}_{ij}^T \boldsymbol{\theta})]^2}{P(\mathbf{X}_{ij}^T \boldsymbol{\theta})(1 - P(\mathbf{X}_{ij}^T \boldsymbol{\theta}))} \mathbf{X}_{ij} \mathbf{X}_{ij}^T = \sum_{i=1}^s \sum_{j=1}^{n_i} I_{\mathbf{X}_{ij}} \quad (1.6)$$

where $I_{\mathbf{X}_{ij}}$ is the information matrix of $\boldsymbol{\theta}$ for the j^{th} subject in the i^{th} group.

If we further define $\Psi(x) = \frac{[P'(x)]^2}{P(x)(1-P(x))}$, then (1.6) becomes

$$I(\boldsymbol{\theta}) = \sum_{i=1}^s \sum_{j=1}^{n_i} \Psi(\mathbf{X}_{ij}^T \boldsymbol{\theta}) \mathbf{X}_{ij} \mathbf{X}_{ij}^T \quad (1.7)$$

For two popular link functions used for binary response, the $\Psi(x)$ functions are

$$\Psi(x) = \begin{cases} \frac{e^x}{(1+e^x)^2}, & \text{for logistic link} \\ \frac{[\Phi'(x)]^2}{\Phi(x)(1-\Phi(x))}, & \text{for probit link} \end{cases} \quad (1.8)$$

and they are both even functions.

Note that a generalized inverse of the information matrix can also be used to obtain asymptotic covariance matrices for functions of $\boldsymbol{\theta}$ that are of interest, which will be further explained in the next section.

1.2 Optimal Designs

For a design with n runs, take model (1.2) as an example, the design can be written as $\xi = \{(\mathbf{X}_{ij}, n_{ij}), i = 1, \dots, s, j = 1, \dots, m_i\}$, where the $\mathbf{X}_{ij}$'s, called design points, are distinct vectors in a design space χ , note that previously $\mathbf{X}_{ij}$ denotes the j^{th} subject in the i^{th} group, now it refers to the j^{th} distinct vector in group i ; n_{ij} is the number of runs assigned to $\mathbf{X}_{ij}$, satisfying $\sum_i \sum_j n_{ij} = n$. m_i is the number of distinct design points in group i . Such a design is called an exact design. Therefore, the optimal exact design problem is to select both $\mathbf{X}_{ij}$'s and n_{ij} 's such that the resulting design ξ is the best in terms of a certain optimality criterion. However, due to the discreteness of the n_{ij} , such optimization problems are usually difficult

to solve. Instead, we work with the corresponding approximate designs, where n_{ij}/n is replaced by w_{ij} . The w_{ij} 's are called design weights and $\sum_i \sum_j w_{ij} = 1$. So the approximate design becomes $\xi = \{(\mathbf{X}_{ij}, w_{ij}), i = 1, \dots, s, j = 1, \dots, m_i\}$. There is an expense for doing such a reformulation because an optimal approximate design may not readily correspond to an exact design depending on the value of n .

For model (1.2) with approximate design ξ , the corresponding information matrix for $\boldsymbol{\theta}$ is

$$I_\xi(\boldsymbol{\theta}) = \sum_{i=1}^s \sum_{j=1}^{m_i} w_{ij} I_{\mathbf{X}_{ij}} \quad (1.9)$$

Sometimes the parameter of interest is not $\boldsymbol{\theta}$, but a vector of $\boldsymbol{\theta}$, say $g(\boldsymbol{\theta})$, possibly a vector function. Assuming g to be a differentiable function which includes a set of linearly independent estimable functions of $\boldsymbol{\theta}$, we can obtain the asymptotic covariance matrix of $g(\boldsymbol{\theta})$,

$$\Sigma_\xi = \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right) I_\xi^- \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right)^T, \quad (1.10)$$

where I_ξ^- is a generalized inverse of I_ξ .

It is possible that for some given $g(\cdot)$ function, Σ_ξ is singular, but we will restrict our attention to the situation where Σ_ξ is nonsingular. The information matrix for $g(\boldsymbol{\theta})$ is therefore the inverse of Σ_ξ ,

$$I_\xi(g(\boldsymbol{\theta})) = \Sigma_\xi^{-1} = \left(\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right) I_\xi^- \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right)^T \right)^{-1}. \quad (1.11)$$

Returning to the beginning of this section, to make sure the selected design is the best, we here introduce some of the most prominent optimality criterion.

1. D-optimality. A design is called D-optimal for $g(\boldsymbol{\theta})$ if it minimizes the determinant of the covariance matrix Σ_ξ , or equivalently, maximizes the determinant of the information matrix $I_\xi(g(\boldsymbol{\theta}))$. From a statistical perspective, a D-optimal design minimizes the expected volume of the joint confidence ellipsoid of $g(\boldsymbol{\theta})$

for a given confidence level. One of the properties for D-optimal designs is that D-optimality is invariant under reparametrization.

2. A-optimality. A design is called A-optimal for $g(\boldsymbol{\theta})$ if it minimizes the trace of the covariance matrix Σ_ξ . Therefore, such designs minimize the average of the asymptotic variance of the maximum likelihood estimates of $g(\boldsymbol{\theta})$.
3. E-optimality. A design is called E-optimal for $g(\boldsymbol{\theta})$ if it minimizes the largest eigenvalue of the covariance matrix Σ_ξ . E-optimal designs protect against the worst scenario for inference.

Kiefer (1974) unified all of the above criteria by introducing the class of functions

$$\Phi_p(\Sigma_\xi) = \left[\frac{1}{v} \text{Tr}((\Sigma_\xi)^p) \right]^{1/p}, 0 < p < \infty \quad (1.12)$$

where v is the dimension of Σ_ξ . When $p \rightarrow 0$, it is equivalent to D-optimality; when $p = 1$, it is A-optimality; when $p \rightarrow \infty$, it is E-optimality.

One of the biggest challenges when studying optimal designs for GLMs is the fact that the information matrix for the parameters of interest usually depends on the unknown parameters themselves (as seen in (1.7) and (1.11)). One way to overcome this is to provide a “best guess” for the unknown parameters. This leads to the idea of locally optimal designs. In fact, many experimenters have some level of prior knowledge before the experiment, which makes this a reasonable and typical approach. Alternatives include adopting a Bayesian approach, which also requires some prior information about the parameters, or using a multi-stage approach (see Silvey (1980)). In this study, we focus on locally optimal designs.

1.3 Orthogonal Arrays

Orthogonal arrays, introduced by Rao (1946, 1947, 1949), are widely used in planning experiments. A $N \times k$ array is called an orthogonal array with s levels and

strength t if, for every $N \times t$ subarray, all possible combinations of t symbols occur equally often as a row. We denote such an array as $OA(N, k, s, t)$. For example, Table 1.1 is an $OA(4, 3, 2, 2)$. For more details about orthogonal arrays, we refer to the book written by Hedayat *et al.* (1999).

Table 1.1: An $OA(4, 3, 2, 2)$.

1	1	1
1	2	2
2	1	2
2	2	1

In design applications, the rows of an orthogonal array correspond to the runs or tests to be performed and the columns represent the factors or variables of interest. When dealing with mixed-level factors, we write the orthogonal array as $OA(N, s_1 \cdots s_k, t)$, where $s_1, \dots, s_k$ are the number of levels for the k factors. For instance, the array in Table 1.2 is an $OA(8, 2^4 4, 2)$.

Hedayat (1989, 1990) introduced the concept of strength $t+$ orthogonal arrays. An $OA(N, k, s, t+)$ is an $OA(N, k, s, t)$ that is not of strength $t+1$, but has one or more subarrays that form an $OA(N, k', s, t+1)$. In many fractional factorial experiments, people prefer such OAs compared to those without this property because strength $t+$ OAs also provide orthogonal estimates for the interactions that correspond to the higher strength subarrays.. More details will be discussed in the next chapter.

Table 1.2: A Mixed Orthogonal Array.

1	1	1	1	1
2	2	2	2	1
1	1	2	2	2
2	2	1	1	2
1	2	1	2	3
2	1	2	1	3
1	2	2	1	4
2	1	1	2	4

1.4 Equivalence Theorem

To verify the optimality of a candidate design is challenging. Kiefer and Wolfowitz (1960) presented a powerful tool for checking D-optimality of a given design under linear model setup, known as equivalence theorem. Here, we will formulate an equivalence theorem for binary response GLMs which will be used in the next chapter. With a slight change in notation, for a GLM with binary responses, we have

$$Prob(Y_i = 1) = P(\mathbf{x}_i^T \boldsymbol{\theta}), \quad (1.13)$$

where Y_i is the response for subject i , $\mathbf{x}_i$ is a vector of k covariates, $\boldsymbol{\theta}$ is a $k \times 1$ vector of unknown parameters and $P(\cdot)$ is a cumulative distribution function as before.

For an approximate design with t support points: $\xi = \{(\mathbf{x}_1, w_1), \dots, (\mathbf{x}_t, w_t)\}$, from (1.7) and (1.9), the information matrix for ξ is

$$I_\xi(\boldsymbol{\theta}) = \sum_{i=1}^t w_i \Psi(\mathbf{x}_i^T \boldsymbol{\theta}) \mathbf{x}_i \mathbf{x}_i^T, \quad (1.14)$$

where Ψ is the same as in (1.8) .

Define $\mathbf{z}_i = \sqrt{\Psi(\mathbf{x}_i^T \boldsymbol{\theta})} \mathbf{x}_i$, and consider linear model

$$Z_i = \mathbf{z}_i^T \boldsymbol{\gamma} + \epsilon_i, \quad (1.15)$$

where Z_i is the i^{th} observation, $\boldsymbol{\gamma}$ is a $k \times 1$ vector of unknown parameters and ϵ_i 's are iid $N(0, \sigma^2)$.

Then for the same design ξ , the information matrix for $\boldsymbol{\gamma}$ is

$$I_\xi(\boldsymbol{\gamma}) = \sum_{i=1}^t w_i \mathbf{z}_i \mathbf{z}_i^T = \sum_{i=1}^t w_i \Psi(\mathbf{x}_i^T \boldsymbol{\theta}) \mathbf{x}_i \mathbf{x}_i^T. \quad (1.16)$$

Therefore, the information matrix for $\boldsymbol{\gamma}$ in the linear model is the same as the information matrix for $\boldsymbol{\theta}$ under the GLM. Hence the optimality for $\boldsymbol{\gamma}$ under the linear model corresponds to the optimality for $\boldsymbol{\theta}$ under the GLM.

In linear model (1.15), for a given design ξ and a design point $\mathbf{z}$, define $d(\mathbf{z}, \xi) = \mathbf{z}^T I_\xi^{-1}(\boldsymbol{\gamma}) \mathbf{z}$, then from Kiefer and Wolfowitz (1960), ξ is D-optimal if and only if

$$\max_{\mathbf{z}} d(\mathbf{z}, \xi) = k, \quad (1.17)$$

In a binary response generalized linear model (1.13), (1.17) becomes

$$\max_{\mathbf{x}} \Psi(\mathbf{x}^T \boldsymbol{\theta}) \mathbf{x}^T I_\xi^{-1}(\boldsymbol{\theta}) \mathbf{x} = k. \quad (1.18)$$

Chapter 2

TOPIC I: D-OPTIMAL DESIGNS FOR GLMS WITH FACTORIAL EFFECTS AND ONE COVARIATE

In this chapter, we study generalized linear models with factorial effects and one continuous covariate. As mentioned earlier, when there are no interaction effects among factors, Stufken and Yang (2012) identified the structure of locally optimal designs under logistic, probit and loglinear models. They also provided an explicit expression of optimal designs for binary data under D-optimality criterion. Tan (2016) extended their results and obtained smaller optimal designs using strength 2 orthogonal arrays. They also explored models with interactions where all interactions up to a specified order are included. When there's no continuous covariate in the model, Yang *et al.* (2012) and Yang and Mandal (2015) provided both theoretical results and algorithms about D-optimal designs for factorial experiments with binary response.

We extend results from Stufken and Yang (2012) and investigate situations where only some interactions of a certain order are included in the model. Compared with the results in Tan (2016), the interaction effects among factors can be involved in a more flexible way. The chapter is organized as follows. In Section 2.1, we give a detailed description about the models of interest. Then we present our main theorem in Section 2.2, which provides an explicit formula for the locally D-optimal designs. In Section 2.3 we applied the proposed optimal design structure to construct smaller D-optimal designs by using orthogonal arrays with some additional properties. For situations where there are only one or two 2-factor interactions, we presented a series of theorems about using strength 2+ OAs to obtain smaller D-optimal designs. How to construct such orthogonal arrays are discussed in Section 2.4. We were able to

identify the connections between strength 2+ OAs and resolution III fractional factorial designs (FFDs). In Section 2.5, we provide two illustrative examples, one of them is about determining influential factors that affect the failure rate of semiconductor circuit when exposed to electrostatic discharge (ESD) (Whitman *et al.* (2006)). Based on our new theorems, the designs we proposed not only outperform the design used in the original study in terms of D-efficiency, but also require much fewer runs. The other example is about an experiment on the inner layer (IL) manufacturing process of Printed Circuit Boards (PCBs). Summaries and discussions are presented in Section 2.6.

2.1 Model Description

With a change in subscripts, the model in (1.2) can be written as

$$\begin{aligned} \text{Prob}(Y_{i_1 i_2 \dots i_L j} = 1) = P(\alpha_0 + \alpha_1^{i_1} + \dots + \alpha_L^{i_L} + \\ \sum_{t=2}^L \sum_{(l_1, l_2, \dots, l_t) \in G_t} \alpha_{l_1 l_2 \dots l_t}^{i_{l_1} i_{l_2} \dots i_{l_t}} + \beta x_{i_1 i_2 \dots i_L j}), \end{aligned} \quad (2.1)$$

where $Y_{i_1 i_2 \dots i_L j}$ is the j^{th} response in the group $(i_1, i_2, \dots, i_L)$, $i_l = 1, \dots, s_l$, s_l is the number of levels for factor l ; $j = 1, \dots, m_{i_1 i_2 \dots i_L}$, $m_{i_1 i_2 \dots i_L}$ is the number of subjects in group $(i_1, i_2, \dots, i_L)$. Further, $P(\cdot)$ is the cumulative distribution function; α_0 is the grand mean, $\alpha_l^{i_l}$ is the effect of the i_l^{th} level of group l , $\alpha_{l_1 l_2 \dots l_t}^{i_{l_1} i_{l_2} \dots i_{l_t}}$ is the effect of the $(i_{l_1}, i_{l_2}, \dots, i_{l_t})^{\text{th}}$ level of the t -th order effect among groups $(l_1, l_2, \dots, l_t)$, $t = 2, \dots, L$; and G_t is a set of t -tuples representing the t -way interactions included in the model, $t = 2, \dots, L$. For simplicity, we also denote $G_1 = \{1, 2, \dots, L\}$. Moreover, β is the common slope parameter; and $x_{i_1 i_2 \dots i_L j}$ is the covariate in the design region denoted by $[L_{i_1 i_2 \dots i_L}, U_{i_1 i_2 \dots i_L}]$. The endpoints $L_{i_1 i_2 \dots i_L}$ and $U_{i_1 i_2 \dots i_L}$ can be $-\infty$ and ∞ , respectively.

Analogously to the expression of (1.3), we can also write the model in vector notation,

$$\text{Prob}(Y_{i_1 i_2 \dots i_L j} = 1) = P((\mathbf{X}^{i_1 \dots i_L j})^T \boldsymbol{\theta}). \quad (2.2)$$

We now have $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_L^T, \dots, \boldsymbol{\alpha}_{l_1 l_2}^T, \dots, \boldsymbol{\alpha}_{l_1 l_2 \dots l_t}^T, \dots, \beta)^T$, terms in $\boldsymbol{\theta}$ correspond to those in the model where for any $1 \leq l_1 < \dots < l_t \leq L$ and $t = 1, \dots, L$, $\boldsymbol{\alpha}_{l_1 \dots l_t} = (\alpha_{l_1 \dots l_t}^{1 \dots 1}, \dots, \alpha_{l_1 \dots l_t}^{1 \dots s_{l_t}}, \dots, \alpha_{l_1 \dots l_t}^{s_{l_1} \dots s_{l_t}})^T$. Further, $\mathbf{X}^{i_1 \dots i_L j} = (1, (\mathbf{X}_1^{i_1})^T, \dots, (\mathbf{X}_L^{i_L})^T, \dots, (\mathbf{X}_{l_1 l_2}^{i_{l_1} i_{l_2}})^T, \dots, (\mathbf{X}_{l_1 l_2 \dots l_t}^{i_{l_1} i_{l_2} \dots i_{l_t}})^T, \dots, x_{i_1 i_2 \dots i_L j})^T$, where terms in $\mathbf{X}^{i_1 \dots i_L j}$ correspond again to those in the model and $\mathbf{X}_{l_1 l_2 \dots l_t}^{i_{l_1} i_{l_2} \dots i_{l_t}}$ is a $(s_{l_1} \times \dots \times s_{l_t}) \times 1$ vector with a 1 in position $(i_{l_1}, \dots, i_{l_t})$ and 0's elsewhere.

For instance, for a model with two factors at two levels each, if interactions between

the two factors are also included, then the first distinct design point in group (2,2) is $\mathbf{X}^{221} = (1, (0, 1), (0, 1), (0, 0, 0, 1), x_{221})^T, \boldsymbol{\theta} = (\alpha_0, (\alpha_1^1, \alpha_1^2), (\alpha_2^1, \alpha_2^2), (\alpha_{12}^{11}, \alpha_{12}^{12}, \alpha_{12}^{21}, \alpha_{12}^{22}), \beta)^T$. Note that earlier $\mathbf{X}^{221}$ denotes the first subject in group (2,2), now it refers to the first distinct design point in group (2,2).

In this parametrization, the whole vector $\boldsymbol{\theta}$ is not estimable, we consider a maximal set of linearly independent estimable functions of $\boldsymbol{\theta}$. Since D-optimality is invariant under reparametrization, the optimal design result is actually true for all possible maximal sets. Let $g(\boldsymbol{\theta}) = B\boldsymbol{\theta} = \boldsymbol{\eta}$ denote one particular maximal set. Note that

$$\text{rank}(B) = 1 + \sum_{t=1}^L \sum_{(l_1, \dots, l_t) \in G_t} \left[\prod_{m=1}^t (s_{l_m} - 1) \right] + 1 \triangleq r. \quad (2.3)$$

We further define $c_{i_1 \dots i_L j} = (\mathbf{X}^{i_1 \dots i_L j})^T \boldsymbol{\theta}$, which is contained in the design region $[D_{i_1 \dots i_L 1}, D_{i_1 \dots i_L 2}]$ derived from the region $[L_{i_1 i_2 \dots i_L}, U_{i_1 i_2 \dots i_L}]$ for $x_{i_1 i_2 \dots i_L j}$. Then the theorem in Section 2.2 gives a locally D-optimal design for model (2.1).

2.2 Theoretical Results

With the notation introduced in Section 2.1, the following result holds.

Theorem 2.2.1. *For any model of the form (2.2) with logistic or probit link, if $\{c^*, -c^*\} \subset [D_{i_1 \dots i_L 1}, D_{i_1 \dots i_L 2}]$ for all $i_1 = 1, \dots, s_1, \dots, i_L = 1, \dots, s_L$, where $c^* > 0$ maximizes $f(c) = c^2(\Psi(c))^r$ on $(-\infty, \infty)$, then the design $\xi^* = \{(c_{i_1 \dots i_L 1} = c^*, w_{i_1 \dots i_L 1} = \frac{1}{2s}), (c_{i_1 \dots i_L 2} = -c^*, w_{i_1 \dots i_L 2} = \frac{1}{2s}), i_1 = 1, \dots, s_1, \dots, i_L = 1, \dots, s_L\}$ is a locally D-optimal design for $\boldsymbol{\eta}$. Here $s = s_1 \times \dots \times s_L$ and $\Psi(x)$ is given by (1.8).*

Proof. We first rewrite model (2.1) in the following way

$$\text{Prob}(Y_{i_1 i_2 \dots i_L j} = 1) = P(\gamma_0 + \sum_{t=1}^L \sum_{(l_1, \dots, l_t) \in G_t} \left[\sum_{i_1=1}^{s_{l_1}-1} \dots \sum_{i_t=1}^{s_{l_t}-1} \gamma_{l_1 \dots l_t}^{i_1 \dots i_t} z_{l_1}^{i_1} \dots z_{l_t}^{i_t} \right] + \beta x_{i_1 \dots i_L j}), \quad (2.4)$$

where for each factor l , define

$$\begin{aligned} z_l^1 &= \begin{cases} 1, & \text{when factor } l \text{ is at level 1} \\ -\frac{1}{s_l-1}, & \text{otherwise} \end{cases} \\ z_l^2 &= \begin{cases} 1, & \text{when factor } l \text{ is at level 2} \\ -\frac{1}{s_l-1}, & \text{otherwise} \end{cases} \\ &\vdots \\ z_l^{s_l-1} &= \begin{cases} 1, & \text{when factor } l \text{ is at level } s_l - 1 \\ -\frac{1}{s_l-1}, & \text{otherwise.} \end{cases} \end{aligned}$$

Analogously to the expression in (2.2), we define $\boldsymbol{\theta}_1 = (\gamma_0, \boldsymbol{\gamma}_1^T, \dots, \boldsymbol{\gamma}_L^T, \dots, \boldsymbol{\gamma}_{l_1 l_2}^T, \dots, \boldsymbol{\gamma}_{l_1 l_2 \dots l_L}^T, \beta)^T$, terms in $\boldsymbol{\theta}_1$ correspond to those in model (2.4) where for any $1 \leq l_1 < \dots < l_t \leq L$ and $t = 1, \dots, L$, $\boldsymbol{\gamma}_{l_1 \dots l_t} = (\gamma_{l_1 \dots l_t}^{1 \dots 1}, \dots, \gamma_{l_1 \dots l_t}^{1 \dots (s_{l_t}-1)}, \dots, \gamma_{l_1 \dots l_t}^{(s_{l_1}-1) \dots (s_{l_t}-1)})^T$. We also define $\mathbf{Z}^{i_1 \dots i_L j} = (1, (\mathbf{Z}_1^{i_1})^T, \dots, (\mathbf{Z}_L^{i_L})^T, \dots, (\mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}})^T, \dots, (\mathbf{Z}_{l_1 l_2 \dots l_L}^{i_{l_1} i_{l_2} \dots i_{l_L}})^T, x_{i_1 i_2 \dots i_L j})^T$, again terms in $\mathbf{Z}^{i_1 \dots i_L j}$ correspond to those in model (2.4) where for each factor l ,

$$\mathbf{Z}_l^{i_l} = \begin{cases} (-\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1}, 1, -\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1})^T, & \text{when } 1 \leq i_l \leq s_l - 1 \\ & \text{and 1 is at } i_l^{th} \text{ position} \\ (-\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1})^T, & \text{when } i_l = s_l \end{cases}$$

and $\mathbf{Z}_{l_1 l_2 \dots l_t}^{i_{l_1} i_{l_2} \dots i_{l_t}} = \mathbf{Z}_{l_1}^{i_{l_1}} \otimes \dots \otimes \mathbf{Z}_{l_t}^{i_{l_t}}$.

In model (2.4), $\boldsymbol{\theta}_1$ is a reparametrization of $\boldsymbol{\theta}$ (see proof in Appendix). Due to the invariance of D-optimality under such transformation, the D-optimality for

θ_1 under model (2.4) corresponds to the D-optimality for $\eta = g(\theta) = B\theta$ under model (2.1). Besides, $(Z^{i_1 \dots i_L j})^T \theta_1 = (X^{i_1 \dots i_L j})^T \theta = c_{i_1 \dots i_L j}$. Let $D^{i_1 \dots i_L j} = (1, (Z_1^{i_1})^T, \dots, (Z_L^{i_L})^T, \dots, (Z_{l_1 l_2}^{i_1 i_2})^T, \dots, (Z_{l_1 l_2 \dots l_L}^{i_1 i_2 \dots i_L})^T, c_{i_1 i_2 \dots i_L j})^T$, then we have $Z^{i_1 \dots i_L j} = A(\theta_1) D^{i_1 \dots i_L j}$, where

$$A(\theta_1) = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \\ -\gamma_0/\beta & A_{(1)}(\theta_1) & \dots & A_{(L)}(\theta_1) & 1/\beta \end{pmatrix}$$

and $A_{(t)}(\theta_1) = (-(\gamma_1 \dots \gamma_t)^T/\beta, \dots, -(\gamma_{l'_1} \dots \gamma_{l'_t})^T/\beta)$ represents the coefficients of all the t -way interactions included in the model.

Then, for the design ξ^* in Theorem 2.2.1, the information matrix for θ_1 becomes

$$\begin{aligned} I_{\xi^*}(\theta_1) &= A(\theta_1) \left[\sum_{i_1=1}^{s_1} \dots \sum_{i_L=1}^{s_L} \sum_{j=1}^{m_{i_1 \dots i_L}} w_{i_1 \dots i_L j} \Psi(c_{i_1 \dots i_L j}) D^{i_1 \dots i_L j} (D^{i_1 \dots i_L j})^T \right] A^T(\theta_1) \\ &= A(\theta_1) \left[\frac{1}{2^s} \Psi(c^*) \sum_{i_1=1}^{s_1} \dots \sum_{i_L=1}^{s_L} \sum_{j=1}^2 D^{i_1 \dots i_L j} (D^{i_1 \dots i_L j})^T \right] A^T(\theta_1) \\ &\triangleq A(\theta_1) M_{\xi^*}(\theta_1) A^T(\theta_1). \end{aligned} \tag{2.5}$$

To prove that ξ^* is D-optimal for θ_1 , we use the equivalence theorem result presented by Kiefer and Wolfowitz (1960). From what we already derived in (1.18), all we need to show is

$$\Psi(c) (D^{i_1 \dots i_L})^T M_{\xi^*}^{-1}(\theta_1) D^{i_1 \dots i_L} \leq r, \tag{2.6}$$

where the equality is attained at the support points of ξ^* . Here $D^{i_1 \dots i_L}$ is obtained

from $\mathbf{D}^{i_1 \dots i_L j}$ by changing $c_{i_1 \dots i_L j}$ to c , with $(i_1, \dots, i_L, c)$ being any design point in the entire design space. That is, ξ^* is a locally D-optimal design if and only if Equation (2.6) is satisfied for all $(i_1, \dots, i_L, c)$.

We first use a lemma to state that $M_{\xi^*}(\boldsymbol{\theta}_1)$ is a block-diagonal matrix.

Lemma 2.2.2. *$M_{\xi^*}(\boldsymbol{\theta}_1)$ is equal to $\Psi(c^*)$ multiplied by a block-diagonal matrix with top-left element 1 and bottom-right element $(c^*)^2$. For those diagonal blocks, suppose factor l has s_l levels, then the block corresponding to that factor is a $(s_l - 1) \times (s_l - 1)$ matrix*

$$B_l = \begin{pmatrix} \frac{1}{(s_l-1)} & -\frac{1}{(s_l-1)^2} & \cdots & \cdots & -\frac{1}{(s_l-1)^2} \\ & \frac{1}{(s_l-1)} & -\frac{1}{(s_l-1)^2} & \cdots & -\frac{1}{(s_l-1)^2} \\ & & \ddots & \ddots & \vdots \\ & sym. & & \ddots & -\frac{1}{(s_l-1)^2} \\ & & & & \frac{1}{(s_l-1)} \end{pmatrix} = \frac{1}{(s_l - 1)^2} (s_l I - J),$$

where J being a square matrix of 1's. Moreover, for factors $l_1, l_2, \dots, l_k$ with $s_{l_1}, s_{l_2}, \dots, s_{l_k}$ levels respectively, the block corresponding to the interaction of these factors is

$$B_{l_1 l_2 \dots l_k} = B_{l_1} \otimes B_{l_2} \otimes \cdots \otimes B_{l_k}.$$

Proof. See appendix. □

Now we can write down the inverse of $M_{\xi^*}(\boldsymbol{\theta}_1)$. Based on Lemma 2.2.2, $M_{\xi^*}^{-1}(\boldsymbol{\theta}_1)$ is $\frac{1}{\Psi(c^*)}$ multiplied by a block-diagonal matrix, with 1 at the top-left and $\frac{1}{(c^*)^2}$ at the bottom-right. For blocks in the middle, it is easy to verify that

$$B_l^{-1} = \frac{(s_l - 1)^2}{s_l} (I + J).$$

In addition, for the interaction block $B_{l_1 l_2 \dots l_k}$

$$B_{l_1 l_2 \dots l_k}^{-1} = (B_{l_1} \otimes B_{l_2} \otimes \dots \otimes B_{l_k})^{-1} = B_{l_1}^{-1} \otimes \dots \otimes B_{l_k}^{-1}.$$

Before we calculate $(\mathbf{D}^{i_1 \dots i_L})^T M_{\xi^*}^{-1}(\theta_1) \mathbf{D}^{i_1 \dots i_L}$, we introduce another lemma here which will be used later.

Lemma 2.2.3.

$$(\mathbf{Z}_l^{i_l})^T \mathbf{Z}_l^{i_l} + (\mathbf{Z}_l^{i_l})^T J \mathbf{Z}_l^{i_l} = \frac{s_l}{s_l - 1}.$$

Proof. See appendix. □

Now, for $(\mathbf{D}^{i_1 \dots i_L})^T M_{\xi^*}^{-1}(\theta_1) \mathbf{D}^{i_1 \dots i_L}$ in Equation (2.6), since $\mathbf{D}^{i_1 \dots i_L} = (1, (\mathbf{Z}_1^{i_1})^T, \dots, (\mathbf{Z}_L^{i_L})^T, \dots, (\mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}})^T, \dots, (\mathbf{Z}_{l_1 l_2 \dots l_L}^{i_{l_1} i_{l_2} \dots i_{l_L}})^T, c)^T$ and $M_{\xi^*}^{-1}(\theta_1)$ is a block-diagonal matrix, we only need to consider the product of each sub-vector in $\mathbf{D}^{i_1 \dots i_L}$ and its corresponding block in $M_{\xi^*}^{-1}(\theta_1)$. For instance, for main effect $\mathbf{Z}_l^{i_l}$, we have

$$\begin{aligned} (\mathbf{Z}_l^{i_l})^T \cdot \frac{1}{\Psi(c^*)} B_l^{-1} \cdot \mathbf{Z}_l^{i_l} &= \frac{1}{\Psi(c^*)} \frac{(s_l - 1)^2}{s_l} (\mathbf{Z}_l^{i_l})^T (I + J) \mathbf{Z}_l^{i_l} \\ &= \frac{1}{\Psi(c^*)} \frac{(s_l - 1)^2}{s_l} [(\mathbf{Z}_l^{i_l})^T \mathbf{Z}_l^{i_l} + (\mathbf{Z}_l^{i_l})^T J \mathbf{Z}_l^{i_l}] \\ &= \frac{1}{\Psi(c^*)} \frac{(s_l - 1)^2}{s_l} \frac{s_l}{s_l - 1} \quad (\text{from Lemma 2.2.3}) \\ &= \frac{1}{\Psi(c^*)} \cdot (s_l - 1). \end{aligned}$$

For interaction effect $\mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}}$, we have

$$\begin{aligned} (\mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}})^T \frac{1}{\Psi(c^*)} B_{l_1 \dots l_t}^{-1} \mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}} &= \frac{1}{\Psi(c^*)} ((\mathbf{Z}_{l_1}^{i_{l_1}})^T \otimes \dots \otimes (\mathbf{Z}_{l_t}^{i_{l_t}})^T) \cdot (B_{l_1}^{-1} \otimes \dots \otimes B_{l_t}^{-1}) \\ &\quad \cdot (\mathbf{Z}_{l_1}^{i_{l_1}} \otimes \dots \otimes \mathbf{Z}_{l_t}^{i_{l_t}}) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\Psi(c^*)} [(\mathbf{Z}_{l_1}^{i_{l_1}})^T B_{l_1}^{-1} \mathbf{Z}_{l_1}^{i_{l_1}}] \otimes \cdots \otimes [(\mathbf{Z}_{l_t}^{i_{l_t}})^T B_{l_t}^{-1} \mathbf{Z}_{l_t}^{i_{l_t}}] \\
&= \frac{1}{\Psi(c^*)} \prod_{i=1}^t (s_{l_i} - 1).
\end{aligned}$$

Also, for the first element in $\mathbf{D}^{i_1 \cdots i_L}$,

$$(1)^T \cdot \frac{1}{\Psi(c^*)} \cdot 1 = \frac{1}{\Psi(c^*)};$$

for the last element in $\mathbf{D}^{i_1 \cdots i_L}$,

$$(c)^T \cdot \frac{1}{(c^*)^2 \Psi(c^*)} \cdot c = \frac{1}{\Psi(c^*)} \cdot \frac{c^2}{(c^*)^2}.$$

To summarize,

$$(\mathbf{D}^{i_1 \cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta}_1) \mathbf{D}^{i_1 \cdots i_L} = \frac{1}{\Psi(c^*)} \left\{ 1 + \sum_{t=1}^L \sum_{(l_1, \dots, l_t) \in G_t} \left[\prod_{m=1}^t (s_{l_m} - 1) \right] + \frac{c^2}{(c^*)^2} \right\}.$$

So Equation (2.6) becomes

$$\begin{aligned}
\Psi(c) (\mathbf{D}^{i_1 \cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta}_1) \mathbf{D}^{i_1 \cdots i_L} &= \frac{\Psi(c)}{\Psi(c^*)} \left\{ 1 + \sum_{t=1}^L \sum_{(l_1, \dots, l_t) \in G_t} \left[\prod_{m=1}^t (s_{l_m} - 1) \right] + \frac{c^2}{(c^*)^2} \right\} \\
&= \frac{\Psi(c)}{\Psi(c^*)} (r - 1) + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)} \leq r, \quad c \in (-\infty, \infty).
\end{aligned} \tag{2.7}$$

To justify inequality (2.7), we have the following lemma.

Lemma 2.2.4. *For logit and probit link,*

$$\frac{\Psi(c)}{\Psi(c^*)} (r - 1) + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)} \leq r$$

for any $r \geq 2$ and $c \in (-\infty, \infty)$.

Proof. See appendix. □

Therefore, based on Lemma 2.2.4, inequality (2.7) always holds. Then the D-optimality of design ξ^* follows immediately from Equation (2.6) and the equivalence theorem, this concludes the proof.

□

We use an example to illustrate Theorem 2.2.1.

Example 2.2.5. For a logistic model of the form as in (2.1), assume that there are 4 factors with 2 levels each, and that the interaction effect between factor 1 and factor 2 is also in the model. For finding a locally D-optimal design, suppose that $\theta = (1, 0, 1, 0, -1, 0, 0.5, 0, -0.5, 0.1, 0.2, 0.3, 0.4, 1)^T$, consider a maximal set of linear estimable functions for θ , for instance, $g(\theta) = B\theta = (\alpha_0, \alpha_1^2 - \alpha_1^1, \alpha_2^2 - \alpha_2^1, \alpha_3^2 - \alpha_3^1, \alpha_4^2 - \alpha_4^1, \alpha_{12}^{11} - \alpha_{12}^{12} - \alpha_{12}^{21} + \alpha_{12}^{22}, \beta)^T$. Then based on Theorem 2.2.1, we can obtain a D-optimal design with two support points in each group, as shown in Table 2.1. The design structure is easier seen in terms of the c^* 's than in terms of the reported x 's. In fact, the D-optimal design in Table 2.1 has support points c^* and $-c^*$ in each of the 16 groups. Since $\text{rank}(B) = 7$, so c^* maximizes $c^2(\Psi(c))^7$, which is approximately 0.7744. Also, all design points have the same weight $1/32$.

For convenience, Table 2.2 list the c^* values for maximizing $c^2(\Psi(c))^r$ when $3 \leq r \leq 9$. Consider two variations of Example 2.2.5: (1) another interaction term (say the interaction effect between factor 3 and 4) is in the model instead of the one we originally picked. In this case, since $\text{rank}(B)$ is still 7, so based on Theorem 2.2.1, c^* is unchanged, and the corresponding x values may be different; (2) another interaction term (say the interaction effect among factor 1, 2 and 3) is in the model in addition to the one we originally picked. In this case, $\text{rank}(B) = 8$, so based on Theorem 2.2.1, c^* maximizes $c^2(\Psi(c))^8$, which is approximately 0.7222. The corresponding x 's in each group will be changed accordingly.

Table 2.1: Support Points for a Locally D-optimal Design.

Group	Support points	Group	Support points
(1,1,1,1)	(-0.3256, -1.8744)	(2,1,1,1)	(-1.5256, -3.0744)
(1,1,1,2)	(0.1744, -1.3744)	(2,1,1,2)	(-1.0256, -2.5744)
(1,1,2,1)	(-0.8256, -2.3744)	(2,1,2,1)	(-2.0256, -3.5744)
(1,1,2,2)	(-0.3256, -1.8744)	(2,1,2,2)	(-1.5256, -3.0744)
(1,2,1,1)	(0.5744, -0.9744)	(2,2,1,1)	(-0.6256, -2.1744)
(1,2,1,2)	(1.0744, -0.4744)	(2,2,1,2)	(-0.1256, -1.6744)
(1,2,2,1)	(0.0744, -1.4744)	(2,2,2,1)	(-1.1256, -2.6744)
(1,2,2,2)	(0.5744, -0.9744)	(2,2,2,2)	(-0.6256, -2.1744)

Table 2.2: c^* That Maximizes $c^2(\Psi(c))^r$ for Logistic and Probit Models.

r	3	4	5	6	7	8	9	10
logistic	1.2229	1.0436	0.9254	0.8399	0.7744	0.7222	0.6793	0.6432
probit	0.9376	0.8159	0.7320	0.6696	0.6209	0.5815	0.5487	0.5209

2.3 Results Based on Orthogonal Arrays

In the previous theorem, we identified a locally D-optimal design under model (2.1). However, it requires a full factorial so the total number of support points becomes large when there are a large number of factors and levels. In many situations, especially when the number of interactions is small, we can usually find smaller designs that are still optimal.

Theorem 2.3.1. *Let ξ^* be the design in Theorem 2.2.1, for model (2.2) with main effects and only a single interaction effect, α_{12} . Assume that an orthogonal array $OA(N, s_1 \cdots s_L, 2+)$ exists so that columns $(1, 2, j)$, $j \neq 1$ or 2 , form an $OA(N, s_1 s_2 s_j, 3)$. Let H denote the set of rows in the $OA(N, s_1 \cdots s_L, 2+)$ and define $\xi_1 = \{(c_{i_1 \cdots i_{L1}} = c^*, w_{i_1 \cdots i_{L1}} = \frac{1}{2N}), (c_{i_1 \cdots i_{L2}} = -c^*, w_{i_1 \cdots i_{L2}} = \frac{1}{2N}), (i_1, \cdots, i_L) \in H\}$. Then $I_{\xi_1}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1)$, so that ξ_1 is also a D -optimal design for η .*

Proof. To show that design ξ_1 is also optimal, from Equation (2.5), we only need to show that

$$M_{\xi_1}(\boldsymbol{\theta}_1) = M_{\xi^*}(\boldsymbol{\theta}_1),$$

where $M_{\xi^*}(\boldsymbol{\theta}_1)$ is given in Lemma 2.2.2.

First, it is easy to see that the upper-left element of $M_{\xi_1}(\boldsymbol{\theta}_1)$ is equal to

$$\Psi(c^*) \cdot \frac{1}{2N} \sum_{(i_1, \dots, i_L) \in H} \sum_{j=1}^2 1 = \Psi(c^*).$$

Similarly, the bottom-right element is

$$\Psi(c^*) \cdot \frac{1}{2N} \sum_{(i_1, \dots, i_L) \in H} \sum_{j=1}^2 (c^*)^2 = (c^*)^2 \Psi(c^*).$$

For the last column of $M_{\xi_1}(\boldsymbol{\theta}_1)$, since we are choosing two symmetric points in each group from H , all off-diagonal elements in the last column are 0.

For the off-diagonal blocks, for one particular block, its rows correspond to the effect $\mathbf{Z}_{r_1 \cdots r_m}^{i_{r_1} \cdots i_{r_m}}$; say, and its columns correspond to the effect $\mathbf{Z}_{c_1 \cdots c_n}^{i_{c_1} \cdots i_{c_n}}$. Note that $m, n \in \{1, 2\}$ and they cannot both be 2. Then the entries in the block can be expressed as

$$\Psi(c^*) \cdot \frac{1}{2N} \cdot \sum_{(i_1, \dots, i_L) \in H} \sum_{j=1}^2 \mathbf{Z}_{r_1 \cdots r_m}^{i_{r_1} \cdots i_{r_m}} (\mathbf{Z}_{c_1 \cdots c_n}^{i_{c_1} \cdots i_{c_n}})^T.$$

Let $q_{(1)} < \dots < q_{(D)}$ be the distinct elements in $\{r_1, \dots, r_m, c_1, \dots, c_n\}$. Since either $D = 2$ or $D = 3$, the property of the orthogonal array assures that each possible level combination $(i_{q_{(1)}}, \dots, i_{q_{(D)}})$ appears equally often, more specifically, each appears $\frac{N}{\prod_{d=1}^D s_{q_{(d)}}}$ times. So the block becomes

$$\Psi(c^*) \cdot \frac{1}{2N} \cdot \frac{N}{\prod_{d=1}^D s_{q_{(d)}}} \cdot 2 \cdot \sum_{i_{q_{(1)}}=1}^{s_{q_{(1)}}} \dots \sum_{i_{q_{(D)}}=1}^{s_{q_{(D)}}} \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}} (\mathbf{Z}_{c_1 \dots c_n}^{i_{c_1} \dots i_{c_n}})^T. \quad (2.8)$$

Since the block is off-diagonal, there must exist a group index among $\{q_{(1)}, \dots, q_{(D)}\}$ which only appears once in $\{r_1, \dots, r_m, c_1, \dots, c_n\}$. Without loss of generality, say this holds for $q_{(1)}$ in $\{r_1, \dots, r_m\}$, then (2.8) becomes

$$\begin{aligned} & \Psi(c^*) \cdot \frac{1}{\prod_{d=1}^D s_{q_{(d)}}} \cdot \sum_{i_{q_{(1)}}=1}^{s_{q_{(1)}}} \dots \sum_{i_{q_{(D)}}=1}^{s_{q_{(D)}}} \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}} (\mathbf{Z}_{c_1 \dots c_n}^{i_{c_1} \dots i_{c_n}})^T \\ &= \Psi(c^*) \cdot \frac{1}{\prod_{d=1}^D s_{q_{(d)}}} \cdot \sum_{i_{q_{(2)}}=1}^{s_{q_{(2)}}} \sum_{i_{q_{(D)}}=1}^{s_{q_{(D)}}} (\mathbf{Z}_{r_1}^{i_{r_1}} \otimes \dots \otimes (\sum_{i_{q_{(1)}}=1}^{s_{q_{(1)}}} \mathbf{Z}_{q_{(1)}}^{i_{q_{(1)}}}) \otimes \dots \otimes \mathbf{Z}_{r_m}^{i_{r_m}}) \\ & \quad \cdot (\mathbf{Z}_{c_1 c_2 \dots c_n}^{i_{c_1} i_{c_2} \dots i_{c_n}})^T. \end{aligned}$$

By definition, $\sum_{i_{q_{(u)}}=1}^{s_{q_{(u)}}} \mathbf{Z}_{q_{(u)}}^{i_{q_{(u)}}} = 0$, thus the entire block is 0. Therefore, all off-diagonal blocks are 0.

Last, we look at diagonal blocks. The block corresponding to the effect $\mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}}$, $t \leq 2$ is

$$\Psi(c^*) \cdot \frac{1}{2N} \cdot \sum_{(i_1, \dots, i_L) \in H} \sum_{j=1}^2 \mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}} (\mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}})^T. \quad (2.9)$$

Again, from the property of the orthogonal array, each possible level combination $(i_{l_1}, \dots, i_{l_t})$ appears $\frac{N}{\prod_{m=1}^t s_{l_m}}$ times. So (2.9) becomes

$$\Psi(c^*) \cdot \frac{1}{2N} \cdot \frac{N}{\prod_{m=1}^t s_{l_m}} \cdot 2 \cdot \sum_{i_{l_1}}^{s_{l_1}} \dots \sum_{i_{l_t}}^{s_{l_t}} \mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}} (\mathbf{Z}_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}})^T$$

$$\begin{aligned}
&= \Psi(c^*) \cdot \frac{1}{\prod_{m=1}^t s_{l_m}} \cdot (s_{l_1} \cdot B_{l_1}) \otimes \cdots \otimes (s_{l_t} \cdot B_{l_t}) \text{ (from proof of Lemma 2.2.2)} \\
&= \Psi(c^*) \cdot B_{l_1} \otimes \cdots \otimes B_{l_t}
\end{aligned}$$

which is the same as the expression of $M_{\xi^*}(\boldsymbol{\theta}_1)$ in Lemma 2.2.2.

In summary,

$$M_{\xi_1}(\boldsymbol{\theta}_1) = M_{\xi^*}(\boldsymbol{\theta}_1),$$

and thus,

$$I_{\xi_1}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1).$$

□

Example 2.3.2. Consider again the Example 2.2.5, since an orthogonal array $OA(8, 2^4, 2+)$ exists, with its columns $(1, 2, 3)$ and $(1, 2, 4)$ being strength 3 (see Table 2.3), based on Theorem 2.3.1, we can obtain a smaller design which is still D-optimal as shown in Table 2.4. Now the design only contains 16 points and all of them have the same weight $1/16$.

Table 2.3: An $OA(8, 2^4, 2+)$ with Columns (1,2,3) and (1,2,4) Being of Strength 3.

1	1	1	2
1	1	2	1
1	2	1	2
1	2	2	1
2	1	1	1
2	1	2	2
2	2	1	1
2	2	2	2

Table 2.4: Smaller D-optimal Design Based on Orthogonal Arrays

Group	Support points	Group	Support points
(1,1,1,2)	(0.1744, -1.3744)	(2,1,1,1)	(-1.5256, -3.0744)
(1,1,2,1)	(-0.8256, -2.3744)	(2,1,2,2)	(-1.5256, -3.0744)
(1,2,1,2)	(1.0744, -0.4744)	(2,2,1,1)	(-0.6256, -2.1744)
(1,2,2,1)	(0.0744, -1.4744)	(2,2,2,2)	(-0.6256, -2.1744)

In Theorem 2.3.1, we still need two support points in each group. In Theorem 2.3.3 we will show that when the orthogonal array has an additional column for a 2-level factor, it is possible to use fewer support points to obtain optimal designs.

Theorem 2.3.3. *Under the same assumptions as in Theorem 2.3.1, suppose that there exists an $OA(N, s_1 \cdots s_L 2^1, 2+)$ and columns $(1, 2, j)$, $j \neq 1$ or 2 form an $OA(N, s_1 s_2 s_j, 3)$. Then*

$$I_{\xi_2}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1),$$

where ξ_2 is obtained in the following way: Let H_1 and H_2 be the $\frac{N}{2} \times L$ subarrays obtained by taking the $\frac{N}{2}$ rows of the OA in which the last entry is 1 and 2, respectively, and deleting the last entry. Then $\xi_2 = \{(c_{i_1 \dots i_L} = c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_1\} \cup \{(c_{i_1 \dots i_L} = -c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_2\}$.

Proof. As in the proof of Theorem 2.3.1, we only need to show all the corresponding blocks in $M_{\xi_2}(\boldsymbol{\theta}_1)$ and $M_{\xi^*}(\boldsymbol{\theta}_1)$ are identical. For most of the blocks, the proofs are analogous to those of Theorem 2.3.1 except for the last column and row, which are for the slope parameter β .

For a specific block in the last column, say its rows correspond to the effect $\mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}}$, $m \leq 2$, then based on the properties of the orthogonal array, $(i_{r_1}, \dots, i_{r_m}, c^*)$ and $(i_{r_1}, \dots, i_{r_m}, -c^*)$ appear equally often, namely $\frac{N}{2s_{r_1} \dots s_{r_m}}$ times. Therefore, the block becomes

$$\Psi(c^*) \cdot \frac{1}{N} \cdot \frac{N}{2s_{r_1} \dots s_{r_m}} \sum_{i_{r_1}=1}^{s_{r_1}} \dots \sum_{i_{r_m}=1}^{s_{r_m}} (c^* \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}} - c^* \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}}) = 0$$

Therefore, all blocks except for the diagonal element in the last column are 0. By symmetry, it is also true for the last row. This concludes the proof. □

Example 2.3.4. Consider again the Example 2.2.5, since an orthogonal array $OA(8, 2^4 2, 2+)$ exists, with its columns $(1, 2, 3)$, $(1, 2, 4)$ and $(1, 2, 5)$ being strength 3 (see Table 2.5), based on Theorem 2.3.3, we can obtain an even smaller design which is still D-optimal as shown in Table 2.6. Now the design only contains 8 points and all of them have the same weight $1/8$.

Table 2.5: An $OA(8, 2^4 2, 2+)$.

1	1	1	2	2
1	1	2	1	1
1	2	1	2	1
1	2	2	1	2
2	1	1	1	2
2	1	2	2	1
2	2	1	1	1
2	2	2	2	2

Table 2.6: Smaller D-optimal Design Based on Orthogonal Arrays.

Group	Support point	Group	Support point
(1,1,1,2)	0.1744	(2,1,1,1)	-1.5256
(1,1,2,1)	-2.3744	(2,1,2,2)	-3.0744
(1,2,1,2)	-0.4744	(2,2,1,1)	-2.1744
(1,2,2,1)	0.0744	(2,2,2,2)	-0.6256

Sometimes, an additional 2-level column cannot be added. However, if there exists an N -run orthogonal array with an extra column that has an odd number of levels, then we may still be able to find an optimal design with less than $2N$ points. Some of the groups will have one support point while others will have two.

Theorem 2.3.5. *Under the same assumptions as in Theorem 2.3.1, assume that an $OA(N, s_1 \cdots s_L(2u+1)^1, 2+)$ exists for some $u \geq 1$ and columns $(1, 2, j)$, $j \neq 1$ or*

2, form an $OA(N, s_1 s_2 s_j, 3)$. Without loss of generality, we denote the levels of the last column as 1, 2, ..., $(2u+1)$. Then

$$I_{\xi_3}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1),$$

where ξ_3 is obtained in the following way: Let H_1 , H_2 and H_3 be the $\frac{uN}{2u+1} \times L$, $\frac{uN}{2u+1} \times L$ and $\frac{N}{2u+1} \times L$ subarrays obtained by taking the $\frac{uN}{2u+1}$, $\frac{uN}{2u+1}$ and $\frac{N}{2u+1}$ rows of the OA in which the last entry is $\{1, 2, \dots, u\}$, $\{u+1, u+2, \dots, 2u\}$ and $\{2u+1\}$, respectively, and deleting the last entry. Then $\xi_3 = \{(c_{i_1 \dots i_L} = c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_1\} \cup \{(c_{i_1 \dots i_L} = -c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_2\} \cup \{(c_{i_1 \dots i_{L-1}} = c^*, w_{i_1 \dots i_{L-1}} = \frac{1}{2N}), (c_{i_1 \dots i_{L-1} 2} = -c^*, w_{i_1 \dots i_{L-1} 2} = \frac{1}{2N}), (i_1, \dots, i_{L-1}) \in H_3\}$.

Proof. With equal weight assigned to each group, most of the blocks in $M_{\xi_3}(\boldsymbol{\theta}_1)$ are identical to the corresponding blocks in $M_{\xi^*}(\boldsymbol{\theta}_1)$ except for the last column.

For a specific block in the last column, say its rows correspond to the effect $\mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}}$, $m \leq 2$, then based on the property of the orthogonal array, $(i_{r_1}, \dots, i_{r_m}, c^*)$ appears $\frac{uN}{(2u+1)s_{r_1} \dots s_{r_m}}$ times with weight $1/N$ each and another $\frac{N}{(2u+1)s_{r_1} \dots s_{r_m}}$ times with weight $1/2N$ each, similarly, $(i_{r_1}, \dots, i_{r_m}, -c^*)$ appears $\frac{uN}{(2u+1)s_{r_1} \dots s_{r_m}}$ times with weight $1/N$ each and another $\frac{N}{(2u+1)s_{r_1} \dots s_{r_m}}$ times with weight $1/2N$ each. Therefore, the block becomes

$$\begin{aligned} \Psi(c^*) \frac{1}{N} \sum_{i_{r_1}=1}^{s_{r_1}} \dots \sum_{i_{r_m}=1}^{s_{r_m}} & \left(\frac{uN}{(2u+1)s_{r_1} \dots s_{r_m}} c^* \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}} - \frac{uN}{(2u+1)s_{r_1} \dots s_{r_m}} c^* \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}} \right. \\ & \left. + \frac{N}{(2u+1)s_{r_1} \dots s_{r_m}} c^* \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}} - \frac{N}{(2u+1)s_{r_1} \dots s_{r_m}} c^* \mathbf{Z}_{r_1 \dots r_m}^{i_{r_1} \dots i_{r_m}} \right) = 0 \end{aligned}$$

Therefore, all blocks except for the diagonal element in the last column are 0. By symmetry, it is also true for the last row. This concludes the proof.

□

Example 2.3.6. Following the same assumptions as in Theorem 2.3.1, consider a 4-factor experiment with three levels each, since an orthogonal array $OA(27, 3^4 3^1, 2+)$ exists, with its columns $(1, 2, 3)$, $(1, 2, 4)$ and $(1, 2, 5)$ being strength 3 (see Table 2.7, to save space, we show the transposed array), based on Theorem 2.3.5, we can obtain a smaller design which is D-optimal as shown in Table 2.8. Now the design only contains 36 points and each group has one or two support points.

Table 2.7: A Transposed $OA(27, 3^4 3^1, 2+)$.

1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	
1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	1	1	2	2	2	3	3	3	1	1	1	2	2	2	3	3	3	1	1	1	2	2	2	3	3	3	1	2	3
1	1	1	2	2	2	3	3	3	2	2	2	3	3	3	1	1	1	3	3	3	1	1	1	2	2	2	1	2	3
1	2	3	3	1	2	2	3	1	2	3	1	1	2	3	3	1	2	3	1	2	2	3	1	1	2	3	1	2	3

Table 2.8: A D-optimal Design Based on $OA(27, 3^4 3^1, 2+)$.[illegible]

For models with two two-factor interactions, by the same arguments as in the proof of Theorem 2.3.1 and Theorem 2.3.3 we can obtain the following results.

Theorem 2.3.7. *Consider design ξ^* in Theorem 2.2.1, for model (2.2) with two 2-factor interaction effects. Assume that they have a common factor, say α_{12} and α_{13} . Assume further that an orthogonal array $OA(N, s_1 \cdots s_L, 2+)$ exists so that columns $(1, 2, j)$ and columns $(1, 3, j)$, $j = 4, \dots, L$ are of strength 3. Let H denote the set of rows in this $OA(N, s_1 \cdots s_L, 2+)$. Then*

$$I_{\xi_4}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1),$$

where $\xi_4 = \{(c_{i_1 \dots i_L 1} = c^*, w_{i_1 \dots i_L 1} = \frac{1}{2N}), (c_{i_1 \dots i_L 2} = -c^*, w_{i_1 \dots i_L 2} = \frac{1}{2N}), (i_1, \dots, i_L) \in H\}$.

Moreover, if an $OA(N, s_1 \cdots s_L 2^1, 2+)$ exists with columns $(1, 2, j)$ and columns $(1, 3, j)$, $j = 4, \dots, L + 1$, being of strength 3. Without loss of generality, we denote the levels of the last column as 1 and 2. Then

$$I_{\xi_5}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1)$$

where ξ_5 is obtained in the following way: Let H_1 and H_2 be the $\frac{N}{2} \times L$ subarrays obtained by taking the $\frac{N}{2}$ rows of the OA in which the last entry is 1 and 2, respectively, and deleting the last entry. Then $\xi_5 = \{(c_{i_1 \dots i_L} = c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_1\} \cup \{(c_{i_1 \dots i_L} = -c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_2\}$.

Theorem 2.3.8. *Consider design ξ^* in Theorem 2.2.1, for model (2.2) with two 2-factor interaction effects,. Assume that they don't have a common factor, say α_{12} and α_{34} , and that an orthogonal array $OA(N, s_1 \cdots s_L, 2)$ exists in which columns $(1, 2, j)$ and columns $(3, 4, j)$ $j = 5, \dots, L$, are of strength 3 and columns $(1, 2, 3, 4)$ are of strength 4. Then, with H as the set of rows in this $OA(N, s_1 \cdots s_L, 2)$,*

$$I_{\xi_6}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1),$$

where $\xi_6 = \{(c_{i_1 \dots i_L 1} = c^*, w_{i_1 \dots i_L 1} = \frac{1}{2N}), (c_{i_1 \dots i_L 2} = -c^*, w_{i_1 \dots i_L 2} = \frac{1}{2N}), (i_1, \dots, i_L) \in H\}$.

Moreover, if an $OA(N, s_1 \dots s_L 2^1, 2)$ exists with columns $(1, 2, j)$ and columns $(3, 4, j)$ $j = 5, \dots, L+1$, being of strength 3 and columns $(1, 2, 3, 4)$ being of strength 4. Without loss of generality, we denote the levels of the last column as 1 and 2. Then

$$I_{\xi_7}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1),$$

where ξ_7 is obtained in the following way: Let H_1 and H_2 be the $\frac{N}{2} \times L$ subarrays obtained by taking the $\frac{N}{2}$ rows of the OA in which the last entry is 1 and 2, respectively, and deleting the last entry. Then $\xi_7 = \{(c_{i_1 \dots i_L} = c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_1\} \cup \{(c_{i_1 \dots i_L} = -c^*, w_{i_1 \dots i_L} = \frac{1}{N}), (i_1, \dots, i_L) \in H_2\}$.

Similarly, for models with any number of two-factor interactions, we have the following result.

Theorem 2.3.9. *Let ξ^* be the design in Theorem 2.2.1, for model (2.2) with main effects and any number of two-factor interactions. Let G_2 be a set of 2-tuples representing the two-factor interactions included in the model. We further define*

$$C_3 = \{(j_1, j_2, j_3) | 1 \leq j_1 < j_2 < j_3 \leq L \text{ and at least one of } (j_1, j_2), (j_1, j_3) \text{ and } (j_2, j_3) \text{ is in } G_2\} \quad \text{and}$$

$$C_4 = \{(j_1, j_2, j_3, j_4) | 1 \leq j_1 < j_2 < j_3 < j_4 \leq L \text{ and they can be grouped into 2 pairs so that both belong to a 2-tuple in } G_2\}$$

Assume that an orthogonal array $OA(N, s_1 \cdots s_L, 2)$ exists so that columns corresponding to C_3 and C_4 form strength 3 and 4 OAs, respectively. Let H denote the set of rows in this $OA(N, s_1 \cdots s_L, 2+)$ and define $\xi_8 = \{(c_{i_1 \dots i_L 1} = c^*, w_{i_1 \dots i_L 1} = \frac{1}{2N}), (c_{i_1 \dots i_L 2} = -c^*, w_{i_1 \dots i_L 2} = \frac{1}{2N}), (i_1, \dots, i_L) \in H\}$. Then $I_{\xi_8}(\boldsymbol{\theta}_1) = I_{\xi^*}(\boldsymbol{\theta}_1)$, so that ξ_8 is also a D-optimal design for η .

2.4 Connections Between Resolution III Designs and Strength 2+ OAs

The results in the previous sections have demonstrated the powerfulness of orthogonal arrays for finding D-optimal designs under model (2.1). The OA-based designs not only reduce the number of support points needed in an experiment, but also relax the assumptions made on the design region. For models where only a small number of interactions are present, using strength 2+ orthogonal arrays is a practical way to reduce the cost of experiments. However, the existence of such OAs is rarely studied and how to construct such OAs remains a big problem.

In this section, we focus on the situation where all factors have the same number of levels. We want to point out that our previous results also hold for mixed-level OAs. In fact, the strength of an orthogonal array is closely related to the concept of resolution in a regular fractional factorial design (FFD). We know that a resolution $t + 1$ regular FFD is also an OA with strength t . For instance, a 2_{III}^{5-2} design with generators $D = AB$ and $E = AC$ is shown in Table 2.9 and it is also an orthogonal array of strength 2. Notice that columns (1, 2, 3) form a strength 3 OA, so do (2, 3, 4) and (2, 3, 5). Thus the resulting OA is of strength 2+, and to identify which columns form a higher strength OA, it is important to study the defining relations of such regular FFDs.

Theorem 2.4.1. *Assuming there are L factors with s levels each, for a N run resolution III regular fractional factorial design, if there exist two letters (say F_1 and F_2)*

Table 2.9: A 2_{III}^{5-2} Design with Generators $I = ABD$ and $I = ACE$.

A	B	C	D	E
−	−	−	+	+
−	−	+	+	−
−	+	−	−	+
−	+	+	−	−
+	−	−	−	−
+	−	+	−	+
+	+	−	+	−
+	+	+	+	+

that are not in the same length 3 word, then this design can be used to construct a strength 2+ orthogonal array, where columns (F_1, F_2, F_j) , $j \neq 1$ or 2 form a strength 3 OA.

Proof. First, let $GF(s)$ denote the s levels for each factor. If the two letters, F_1 and F_2 , are not from the same length 3 word, then for any factor F_j , $j \neq 1$ or 2, the equation

$$x_1 + x_2 + x_j = n,$$

where x_i is the value of F_i , has N/s solutions for every $n \in \{0, 1, \dots, s-1\}$. Further, for each n , there are $\frac{N/s}{s^2}$ solutions for each possible combination of (x_1, x_2, x_j) . By definition, the columns (F_1, F_2, F_j) form a strength 3 orthogonal array.

□

Considering the design in Table 2.9 again, notice that the full defining relation is $I = ABD = ACE = BCDE$ and factors B and C do not simultaneously appear in

any of the length 3 words. Therefore, according to Theorem 2.4.1, this design can be used to construct an $OA(8, 2^5, 2+)$ with its columns (2,3,1), (2,3,4) and (2,3,5) being of strength 3. The resulting OA can be incorporated with Theorem 2.3.1 or Theorem 2.3.3 to construct smaller D-optimal designs.

Another example involves a three-level resolution III design 3_{III}^{5-2} with $D = AB$ and $E = AB^2C$. The defining relation is: $I = ABD^2 = AB^2CE^2 = AC^2DE = BCDE^2$. Notice that factors A and C are not in the same length 3 word ABD^2 . Thus again, according to Theorem 2.4.1, this design can be used to construct an $OA(27, 3^5, 2+)$ with its columns (1,3,2), (1,3,4) and (1,3,5) being of strength 3. Then the resulting OA can be applied for Example 2.3.6 using Theorem 2.3.5 and construct a smaller D-optimal design using one or two support points in each selected group. In this case, a full factorial experiment for four 3-level factors requires $3^4 \times 2 = 162$ distinct support points while the smaller optimal design only needs 36.

2.5 Illustrative Examples

In this section, we provide two examples to illustrate our proposed theorems.

Example 1: Electrostatic Discharge (ESD) Experiment

The ESD experiment was originally reported by Whitman *et al.* (2006). In this study, the experimenters considered a logistic model to determine influential factors that affect the failure rate of semiconductor circuit wafers when exposed to electrostatic discharge. Four factors and one continuous covariate are under consideration as shown in Table 2.10.

The first two factors, Lot A and Lot B, indicate the type of wafer used. ESD handling is a factor to indicate whether or not the standard procedure was applied. No ESD handling means that no ESD-safe lab coat/shoes and no wrist strap were

Table 2.10: Factors and Covariate for ESD Experiment.

	Description	levels	
Factors	Lot A	Location 1	Location 2
	Lot B	Location 1	Location 2
	ESD Handling	No	Yes
	Pulse Polarity	Negative	Positive
Covariate	Voltage	continuous	

used. The ESD testing involves "zapping" a part first with a pulse polarity (positive or negative) and then followed by a second pulse of the opposite polarity. Since there is no industry standard specifying the order of pulse polarity, that makes it the fourth factor. The continuous covariate is the voltage each wafer was tested at and the response variable is binary: a wafer either passes or fails the test. Let p be the probability a wafer passes the test and use 1 and -1 to denote the two levels for each of the four factors shown in Table 2.10, then the regression model can be written as:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{LotA} + \beta_2 \text{LotB} + \beta_3 \text{ESD} + \beta_4 \text{Pulse} + \beta_{34} \text{ESD} \times \text{Pulse} + \beta_5 \text{Voltage}. \quad (2.10)$$

Notice that compared with model (2.2), model (2.10) is no longer overparametrized. We have demonstrated the equivalence of these two types of models in Theorem 2.2.1. Also, note that the interaction effect between ESD handling and pulse polarity is also included in the model.

The experimenters decided to conduct a full factorial design with the continuous covariate Voltage being discretized into 5 levels: 25, 30, 35, 40 and 45 Volt. In total there are $2^4 \times 5 = 80$ runs. We should point out that the experimenters didn't provide

the reasons for selecting $[25, 45]$ as the voltage range or why they chose exactly 5 levels to design the experiment. Instead, we will treat Voltage as a continuous covariate in the following analysis.

To find locally optimal designs, we use $\beta_0 = (-7.50, 1.50, -0.20, -0.15, 0.25, 0.40, 0.35)^T$ as reported in Lukemire *et al.* (2018), which were obtained based on parameter estimations in the original study. Then, according to Theorem 2.2.1, a D-optimal design using full factorials and two support points in each group can be obtained and is reported in Table 2.11.

Table 2.11: D-optimal Design for the ESD Experiment Using a Full Factorial.

LotA	LotB	ESD	Pulse	Volt1	Volt2	LotA	LotB	ESD	Pulse	Volt1	Volt2
-1	-1	-1	-1	22.07	26.50	1	-1	-1	-1	13.50	17.93
-1	-1	-1	1	22.93	27.36	1	-1	-1	1	14.36	18.78
-1	-1	1	-1	25.22	29.64	1	-1	1	-1	16.64	21.07
-1	-1	1	1	21.50	25.93	1	-1	1	1	12.93	17.36
-1	1	-1	-1	23.22	27.64	1	1	-1	-1	14.64	19.07
-1	1	-1	1	24.07	28.50	1	1	-1	1	15.50	19.93
-1	1	1	-1	26.36	30.78	1	1	1	-1	17.79	22.21
-1	1	1	1	22.64	27.07	1	1	1	1	14.07	18.50

In terms of approximate designs, the relative D-efficiency of the original 80-point design ξ_0 compared to the optimal design ξ^* we proposed can be computed using the following formula:

$$E_{\xi_0} = \left[\frac{\det(I_{\xi_0})}{\det(I_{\xi^*})} \right]^{1/p}, \quad (2.11)$$

where p is the number of parameters in the model, which is 7. The idea behind is

that a more efficient design needs fewer runs to achieve the same criterion than a less efficient design. In this case, $E_{\xi_0} = 24.22\%$, suggesting that our optimal design is over four times as D-efficient as the original design.

When applying Theorem 2.3.1 and Theorem 2.3.3, smaller optimal designs can also be obtained. For instance, the left panel in Table 2.12 only requires 16 distinct runs and the right panel only contains 8 different support points and can be executed in as few as 8 runs. To produce an 80-run design as reported in the original study, all we need to do is repeat repeat runs from these optimal designs. There is, however, a trade-off for doing that, the smaller designs cannot distinguish confounding effects under model-misspecification.

Table 2.12: Smaller D-optimal Designs for the ESD Experiment Using Orthogonal Arrays.

LotA	LotB	ESD	Pulse	Volt1	Volt2
-1	-1	1	-1	25.22	29.64
-1	-1	1	1	21.50	25.93
-1	1	-1	-1	23.22	27.64
-1	1	-1	1	24.07	28.50
1	-1	-1	-1	13.50	17.93
1	-1	-1	1	14.36	18.78
1	1	1	-1	17.79	22.21
1	1	1	1	14.07	18.50

LotA	LotB	ESD	Pulse	Volt
-1	-1	1	-1	29.64
-1	-1	1	1	21.50
-1	1	-1	-1	27.64
-1	1	-1	1	24.07
1	-1	-1	-1	13.50
1	-1	-1	1	18.78
1	1	1	-1	17.79
1	1	1	1	18.50

Another useful comparison would be if we threshold the voltage values in our proposed designs based on the $[25, 45]$ voltage range adopted in the original study. It turns out that even after truncation, our proposed designs still performs gener-

ally better than the original design. For instance, Table 2.13 is the second optimal design repored in Table 2.12 after threshold. The relative efficiency of the original 80-point design comparing with this threshold design is about 56.36%, meaning that this threshold design is still almost twice as efficient as the original design.

Table 2.13: 8-point Threshold Design for the ESD Experiment.

LotA	LotB	ESD	Pulse	Volt
-1	-1	1	-1	29.64
-1	-1	1	1	25.00
-1	1	-1	-1	27.64
-1	1	-1	1	25.00
1	-1	-1	-1	25.00
1	-1	-1	1	25.00
1	1	1	-1	25.00
1	1	1	1	25.00

Example 2: Printed Circuit Board (PCB) Experiment

Jeng *et al.* (2008) reported an experiment on the inner layer (IL) manufacturing process of printed circuit boards (PCBs). Many defects may occur during this process, of which shorts and opens in the circuits are the major ones. In this example, we modify the original experiment to suit our purpose and consider the response variable as whether or not there is a short in the circuit. Three factors and one covariate are under consideration as shown in Table 2.14.

The first factor Preheat indicates whether preheating was done or not. Surface

Table 2.14: Factors and Covariate for the PCB Experiment.

	Description	levels		
Factors	Preheat (x_1)	No	Yes	–
	Surface preparation (x_2)	Scrub	Pumice	Chemical
	Lamination pressure (x_3)	20 psi	40 psi	60 psi
Covariate	Exposure energy (x_4)	continuous		

preparation has 3 different methods (Scrub, Pumice and Chemical). Lamination pressure has 3 levels (20psi, 40psi and 60psi) and Exposure energy indicates to how much ultraviolet radiation the laminated panel is exposed. In the original study, the experimenters selected three levels for the exposure energy: 14, 17 and 20. See Maruthi and Roshan Joseph (1999) and Yang and Mandal (2015) for more details of this experiment. Table 2.15 is obtained from Table 2 of Jeng *et al.* (2008).

Let p be the probability that there is a short in the circuit, the following logistic model is under consideration:

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_{21} x_{21} + \beta_{22} x_{22} + \beta_{121} x_1 x_{21} + \beta_{122} x_1 x_{22} + \beta_{31} x_{31} + \beta_{32} x_{32} + \beta_4 x_4, \quad (2.12)$$

where for the three-level factor x_2 (or x_3), two degrees of freedoms are split into x_{21} and x_{22} (or x_{31} and x_{32}). For convenience, we use (1,-0.5), (-0.5,1) and (-0.5,-0.5) to denote the three levels, which is consistent with our proofs for the theorems. For the two-level factor x_1 , we use 1 and -1 to denote the two levels as before. Also, compared with model (2.2), model (2.12) is again no longer overparametrized. We have demonstrated the equivalence of these two types of models in Theorem 2.2.1. Furthermore, notice that the interaction effect between Preheat (x_1) and Surface

Table 2.15: Original Design for the PCB Experiment.

x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
1	1	1	14	2	1	1	20
1	1	2	17	2	1	2	14
1	1	3	20	2	1	3	17
1	2	1	17	2	2	1	20
1	2	2	20	2	2	2	14
1	2	3	14	2	2	3	17
1	3	1	17	2	3	1	14
1	3	2	20	2	3	2	17
1	3	3	14	2	3	3	20

preparation (x_2) is also included in the model.

To find locally optimal designs, we use $\beta_0 = (-8.095, -0.491, 0.045, 0.387, -0.495, -0.115, -0.496, -0.399, 0.373)^T$, which is obtained from the fitting of the logistic model. Then, according to Theorem 2.2.1, a D-optimal design using a full factorial and two support points in each group can be obtained and is reported in Table 2.16.

After applying the relative D-efficiency formula (2.11), we observe that the relative D-efficiency for the original design is $E_{\xi_0} = 58.68\%$, suggesting that our optimal design is almost twice as efficient as the original design. However, we should point out that in reality, the variable Exposure energy (x_4) is usually measured in whole numbers. If we round the optimal x_4 values in Table 2.16, as shown in Table 2.17, the resulting design is still of high D-efficiency (approximately 99.8%).

Smaller designs can also be obtained using orthogonal array results. Table 2.18

Table 2.16: D-optimal Design for the PCB Experiment Using a Full Factorial.

x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
1	1	1	(23.563,27.205)	2	1	1	(18.586,22.229)
1	1	2	(23.175,26.818)	2	1	2	(18.199,21.841)
1	1	3	(21.569,25.212)	2	1	3	(16.592,20.235)
1	2	1	(20.660,24.303)	2	2	1	(18.740,22.382)
1	2	2	(20.272,23.915)	2	2	2	(18.352,21.994)
1	2	3	(18.666,22.309)	2	2	3	(16.746,20.388)
1	3	1	(21.753,25.396)	2	3	1	(20.755,24.398)
1	3	2	(21.366,25.008)	2	3	2	(20.368,24.010)
1	3	3	(19.760,23.402)	2	3	3	(18.761,22.404)

shows a transposed $OA(18, 2^1 3^3, 2+)$, where columns (1,2,3) and (1,2,4) form strength 3 orthogonal arrays. Then, according to Theorem 2.3.5, a smaller D-optimal design can be obtained by deleting the last 3-level column in the OA and using one or two points in each group. The resulting design is shown in Table 2.19. Note that all groups share the same weight $1/18$. For groups with 2 support points, we assign weight $1/36$ to each point. Therefore, from Table 2.17 to Table 2.19, we successfully reduced the number of support points from 36 to 24.

2.6 Summaries and Discussions

In this chapter, we proposed several locally D-optimal designs for GLMs with factorial effects under a very general setup. The factors may or may not have interactions and the interactions can be involved in a more flexible way compared to some

Table 2.17: A Near D-optimal Design for the PCB Experiment.

x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
1	1	1	(24,27)	2	1	1	(19,22)
1	1	2	(23,27)	2	1	2	(18,22)
1	1	3	(22,25)	2	1	3	(17,20)
1	2	1	(21,24)	2	2	1	(19,22)
1	2	2	(20,24)	2	2	2	(18,22)
1	2	3	(19,22)	2	2	3	(17,20)
1	3	1	(22,25)	2	3	1	(21,24)
1	3	2	(21,25)	2	3	2	(20,24)
1	3	3	(20,23)	2	3	3	(19,22)

Table 2.18: A Transposed $OA(18, 2^1 3^3, 2+)$.

1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
1	1	1	2	2	2	3	3	3	1	1	1	2	2	2	3	3	3
1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	2	3	2	3	1	3	1	2	2	3	1	1	2	3	3	1	2

previous research. When the number of factors or the number of levels in each factor is large, a D-optimal design using a full factorial with two points in each cell requires too many distinct support points, which is typically unnecessary since the number of parameters being estimated is usually relatively small. Our OA-based results overcome such problems and greatly reduce the number of support points required in the

Table 2.19: A Smaller D-optimal Design for the PCB Experiment.

x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
1	1	1	27	2	1	1	19
1	1	2	23	2	1	2	(18,22)
1	1	3	(22,25)	2	1	3	20
1	2	1	21	2	2	1	22
1	2	2	(20,24)	2	2	2	18
1	2	3	22	2	2	3	(17,20)
1	3	1	(22,25)	2	3	1	(21,24)
1	3	2	25	2	3	2	24
1	3	3	20	2	3	3	19

optimal designs. Methods for obtaining such orthogonal arrays are briefly discussed. The power of our theorems is demonstrated through two illustrative examples. The results indicate that designs based on our theorems are not only more efficient than the original designs, but also require for fewer distinct runs.

One important observation from our computational studies is that D-optimal designs are often not unique, especially when the number of interaction effects is small. In fact, some of the optimal designs we obtained can not be explained using our proposed theorems. For instance, the design shown in Table 2.20 is another D-optimal design for example 2.2.5, yet it doesn't fit any of our theorems. How to summarize such results into theorems remains an open question. Besides, the models we consider in this chapter only involve the first-order component of the continuous covariate. It would be interesting to see what the optimal design structures look like if we add an

Table 2.20: Another D-optimal Design for Example 2.2.5

Group	Support point	Weight	Group	Support point	Weight
(1,1,1,1)	-0.3256	1/16	(1,2,2,1)	0.0744	1/16
(1,1,1,2)	-1.3744	1/16	(1,2,2,2)	-0.9744	1/16
(1,1,2,1)	-0.8256	1/16	(2,1,1,1)	-3.0744	1/8
(1,1,2,2)	-1.8744	1/16	(2,1,2,2)	-1.5256	1/8
(1,2,1,1)	0.5744	1/16	(2,2,1,2)	-0.1256	1/8
(1,2,1,2)	-0.4744	1/16	(2,2,2,1)	-2.6744	1/8

additional second-order component to the model.

TOPIC II: D-OPTIMAL DESIGNS FOR GLMS WITH MULTIPLE COVARIATES

In this chapter and the following chapter, we study GLMs with multiple covariates. Optimal design results for such models are relatively rare. Sitter and Torsney (1995a) studied optimal designs for binary response experiments with two covariates. They also extended the results to models with more than two design variables (Sitter and Torsney (1995b)). Russell *et al.* (2009) investigated D-optimal design structures for Poisson regression models. Kabera and Haines (2012) considered multiple-covariate logistic models without interaction for D and D_s optimality. Hu and Stufken (2016b) provided complete class results of optimal designs for multiple-covariate GLMs. As discussed in Sitter and Torsney (1995b), for models with two or more design variables, the information matrices can be made arbitrarily large unless proper assumptions are made on the design space, for instance, a bounded interval for each variable. Assumptions like this are actually very common and frequently used in many applications. For example, in a clinical study, the experimenter would not like to use a too high dose level which may cause serious side effects, or a too low dose level which could make the response infeasible to observe.

The reason we use two chapters to cover such models is that even though Chapters 3 and 4 both consider GLMs with multiple covariates, they have different restrictions on the design space as well as the parameter settings. As a result, the D-optimal design structures are completely different and there is no obvious way to summarize these results in a unified statement. In this chapter, we assume in the following sections that for all the covariates, only one of them is unbounded, while the others are restricted to intervals. A main reason for leaving one covariate unbounded is

mathematical convenience. Also, interaction effects may exist among the bounded covariates.

This chapter is organized as follows. In Section 3.1, we introduce the models of interest. Section 3.2 provides the D-optimal design structures when no interactions exist among bounded covariates. The results are first obtained by Yang *et al.* (2011) and we here provide an alternative proof. In Section 3.3, we propose our main theorem for the interaction models. Additional results for obtaining smaller D-optimal designs using orthogonal arrays are presented in Section 3.4, followed by a brief discussion in Section 3.5.

3.1 Model Description

As discussed earlier, the model we consider in this topic can be written as

$$Prob(Y_i = 1) = P(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \sum_{t=2}^{p-1} \sum_{(l_1, l_2, \dots, l_t) \in H_t} \beta_{l_1 l_2 \dots l_t} x_{il_1} x_{il_2} \cdots x_{il_t}) \quad (3.1)$$

where Y_i and $(x_{i1}, x_{i2}, \dots, x_{ip})$ are the response and p covariates for subject i . As usual, $P(\cdot)$ is a cumulative distribution function. The first $p-1$ covariates are assumed to be bounded, with x_{ij} belonging to the (bounded) interval $[L_j, U_j]$, $1 \leq j \leq p-1$. The last covariate is unbounded. In other words, the design space is $\chi = [L_1, U_1] \times \cdots \times [L_{p-1}, U_{p-1}] \times R$. Interactions are among the bounded covariates only, so that H_t is a set of t -tuple pairs representing the t -way interactions included in the model among the first $p-1$ variables.

In vector form, we have the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1}, \dots, \beta_{l_1 \dots l_t}, \dots, \beta_p)^T$ and model vector $\tilde{\mathbf{x}}_i = (1, x_{i1}, \dots, x_{i,p-1}, \dots, x_{il_1} \cdots x_{il_t}, \dots, x_{ip})^T$. Notice that the design point corresponding to $\tilde{\mathbf{x}}_i$ should be $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$, that is, the model vector $\tilde{\mathbf{x}}_i$ excluding the intercept and all interaction terms. Furthermore, as explained in Chapter 2, in model setup, the subscript i in $\mathbf{x}_i$ means “subject i ”, while in design setup, it represents the “ i -th distinct support point” in an arbitrary design. After these clarifications, with $c_i = \tilde{\mathbf{x}}_i^T \boldsymbol{\beta}$, $\tilde{\mathbf{C}}_i = (1, x_{i1}, \dots, x_{i,p-1}, \dots, x_{il_1} \cdots x_{il_t}, \dots, c_i)^T$ and the assumption that $\beta_p \neq 0$, there is a one-to-one relationship between $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{C}}_i$. And an approximate design $\xi = \{(\mathbf{x}_i, w_i), i = 1, \dots, k\}$ can also be written as $\xi = \{(\mathbf{C}_i, w_i), i = 1, \dots, k\}$, where $\mathbf{C}_i = (x_{i1}, \dots, x_{i,p-1}, c_i)$.

3.2 Results for Main-effect Models

For models like (3.1), if we ignore all the interaction effects, then as demonstrated in Yang *et al.* (2011), an explicit expression for locally D-optimal designs for parameter

β can be obtained. Here we restate their result in the following theorem and provide an alternative proof for it.

Theorem 3.2.1. *For the logistic and probit models as in (3.1) with no interaction effects, a D-optimal design for parameter β is given by $\xi^* = \{(\mathbf{C}_{i1}^*, \frac{1}{2^p}) \& (\mathbf{C}_{i2}^*, \frac{1}{2^p}), l = 1, \dots, 2^{p-1}\}$ where $\mathbf{C}_{i1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$ and $\mathbf{C}_{i2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$. Here h_{lj} is either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, 2^{p-1}$ cover all possible combinations; c^* maximizes $c^2(\Psi(c))^{p+1}$ where Ψ is defined in (1.8).*

Proof. First of all, instead of considering the design problem directly, we introduce an equivalent canonical form (see Ford *et al.* (1992), Atkinson and Haines (1996) and Torsney and Gunduz (2001)) of the original design. For each of the first $p - 1$ covariates, consider the following transformation:

$$v_i = \frac{x_i - (U_i + L_i)/2}{(U_i - L_i)/2} \quad (3.2)$$

Then we have $v_i \in [-1, 1], i = 1, \dots, p - 1$. For notational convenience, let $a_i = (U_i + L_i)/2$ and $b_i = (U_i - L_i)/2$, then $v_i = \frac{x_i - a_i}{b_i}$. Also, we let $v_p = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ and write $\mathbf{v} = (v_1, v_2, \dots, v_p)$.

Now, for an arbitrary design point $\mathbf{x} \in \chi$ and the corresponding design vector $\tilde{\mathbf{x}} = (1, \mathbf{x}^T)^T$, we have

$$B\tilde{\mathbf{x}} = \tilde{\mathbf{v}}, \quad (3.3)$$

where $\tilde{\mathbf{v}} = (1, \mathbf{v}^T)^T$ and

$$B = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ -a_1/b_1 & 1/b_1 & 0 & \cdots & \cdots & 0 \\ -a_2/b_2 & 0 & 1/b_2 & 0 & \cdots & 0 \\ & & \vdots & & & \\ -a_{p-1}/b_{p-1} & 0 & \cdots & 0 & 1/b_{p-1} & 0 \\ \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_{p-1} & \beta_p \end{pmatrix}$$

is a $(p+1) \times (p+1)$ nonsingular matrix.

Therefore, the mapping from $\mathbf{x}$ to $\mathbf{v}$ transforms a design in the original variables

$$\xi = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \\ w_1 & w_2 & \cdots & w_k \end{pmatrix}$$

into an induced design

$$\xi_v = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \\ w_1 & w_2 & \cdots & w_k \end{pmatrix}.$$

The induced design space $\chi_v = \{\mathbf{v} : B\tilde{\mathbf{x}} = \tilde{\mathbf{v}}, \mathbf{x} \in \chi\} = [-1, 1] \times \cdots \times [-1, 1] \times R$.

Also, for each $\mathbf{x}_i$, the corresponding $\mathbf{v}_i$ has its last element being $v_{ip} = \tilde{\mathbf{x}}_i^T \boldsymbol{\beta} = c_i$.

Then, let $M_{\xi_v} = \sum_{i=1}^k w_i \Psi(c_i) \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^T$, the information matrix for design ξ is

$$\begin{aligned} I_\xi &= \sum_{i=1}^k w_i \Psi(\tilde{\mathbf{x}}_i^T \boldsymbol{\beta}) \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T \\ &= \sum_{i=1}^k w_i \Psi(c_i) B^{-1} \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^T (B^{-1})^T \\ &= B^{-1} \left[\sum_{i=1}^k w_i \Psi(c_i) \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^T \right] (B^{-1})^T \\ &= B^{-1} M_{\xi_v} (B^{-1})^T. \end{aligned}$$

As a result, $\det(I_\xi) = \det(B^{-1})^2 \cdot \det(M_{\xi_v})$, meaning that maximizing $\det(I_\xi)$ on χ is equivalent to maximizing $\det(M_{\xi_v})$ on χ_v . Once we obtain the optimal design ξ_v^* , we may transform it back to ξ^* by mapping $\mathbf{v}$ back to $\mathbf{x}$. So from now on, we will focus only on designs ξ_v on the induced design space χ_v .

It is easy to show that after the transformation $B\tilde{\mathbf{x}} = \tilde{\mathbf{v}}$, the proposed design ξ^* in Theorem 3.2.1 has become $\xi_v^* = \{(\tilde{\mathbf{C}}_{l1}^*, \frac{1}{2^p}) \& (\tilde{\mathbf{C}}_{l2}^*, \frac{1}{2^p}), l = 1, \dots, 2^{p-1}\}$, where $\tilde{\mathbf{C}}_{l1}^* = (\tilde{h}_{l1}, \dots, \tilde{h}_{l,p-1}, c^*)^T$ and $\tilde{\mathbf{C}}_{l2}^* = (\tilde{h}_{l1}, \dots, \tilde{h}_{l,p-1}, -c^*)^T$. Here $\tilde{h}_{lj}$ is either -1 or 1 and $(\tilde{h}_{l1}, \dots, \tilde{h}_{l,p-1}), l = 1, \dots, 2^{p-1}$ cover all possible combinations.

Therefore, to show the D-optimality of ξ_v^* on χ_v , we apply again the equivalence theorem and all we have to show is that

$$\Psi(c)\tilde{\mathbf{C}}^T I_{\xi_v^*}^{-1} \tilde{\mathbf{C}} \leq p + 1, \quad (3.4)$$

where the equality is attained at the support points of ξ_v^* . Here $\tilde{\mathbf{C}} = (1, v_1, \dots, v_{p-1}, v_p)^T$, where $(v_1, \dots, v_p)$ represents an arbitrary design point in χ_v with $v_i, 1 \leq i \leq p-1$ taking values between -1 and 1 and $v_p = \tilde{\mathbf{x}}^T \boldsymbol{\beta} = c \in (-\infty, \infty)$.

Following a similar procedure as in the proof of Theorem 2.2.1, we have

$$\begin{aligned} \Psi(c)\tilde{\mathbf{C}}^T I_{\xi_v^*}^{-1} \tilde{\mathbf{C}} &= \frac{\Psi(c)}{\Psi(c^*)} \left\{ 1 + v_1^2 + \dots + v_{p-1}^2 + \frac{c^2}{(c^*)^2} \right\} \\ &\leq \frac{\Psi(c)}{\Psi(c^*)} p + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)}, \quad c \in (-\infty, \infty). \end{aligned} \quad (3.5)$$

Then, from Lemma 2.2.4,

$$\Psi(c)\tilde{\mathbf{C}}^T I_{\xi_v^*}^{-1} \tilde{\mathbf{C}} \leq p + 1, \quad c \in (-\infty, \infty), \quad (3.6)$$

and the equality holds for each design point in ξ_v^* . This proves that design ξ_v^* is indeed D-optimal on χ_v and as a result, the proposed design ξ^* is also D-optimal on the original design space χ .

This concludes the proof. □

We use a simple example to illustrate the theorem.

Example 3.2.2. Consider model (3.1) with logit link and $\tilde{\mathbf{x}}_i^T \boldsymbol{\beta} = c_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$. Assume that the first two covariates are restricted to $[0, 2]$ and $[-1, 1]$, respectively, and that there is no restriction on the third covariate. Then, based on Theorem 3.2.1, a locally D-optimal design for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T = (1, -1, 0.5, 1)^T$ is shown in Table 3.1. The value c^* maximizes $c^2(\Psi(c))^{3+1}$, which is approximately 1.0436.

Table 3.1: Support Points and Weights for a Locally D-optimal Design.

Support points	Weights	Support points	Weights
$(0, -1, 0.5436)$	$1/8$	$(2, -1, 0.4564)$	$1/8$
$(0, -1, -1.5436)$	$1/8$	$(2, -1, 2.5436)$	$1/8$
$(0, 1, -2.5436)$	$1/8$	$(2, 1, -0.5436)$	$1/8$
$(0, 1, -0.4564)$	$1/8$	$(2, 1, 1.5436)$	$1/8$

3.3 Main Results for Interaction Models

In Section 3.2, we reviewed the D-optimal design structures for the main-effects only models. A natural question is, for a more general case as described in model (3.1) where interaction effects may exist among those bounded covariates, can we find an explicit formula for D-optimal designs? Unfortunately, the method Yang *et al.* (2011) used for proving Theorem 3.2.1 seems unlikely to be generalized to the interaction models. However, the alternative proof we provided in Section 3.2 can be modified and successfully applied to this problem. It turns out that for hierarchical interaction models, D-optimal designs for model (3.1) have the same structure as in Theorem 3.2.1, only the value of c^* is changed. We summarize our main result in the next theorem.

Theorem 3.3.1. *For the logistic and probit models as in (3.1) with hierarchical interaction effects, that is, when a t -th order effect is included in the model, then all t' -th order ($t' < t$) must also exist in the model, a D-optimal design for parameter β is given by $\xi^* = \{(\mathbf{C}_{l1}^*, \frac{1}{2^p}) \& (\mathbf{C}_{l2}^*, \frac{1}{2^p}), l = 1, \dots, 2^{p-1}\}$ where $\mathbf{C}_{l1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$ and $\mathbf{C}_{l2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$. Here h_{lj} is either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l =$*

$1, \dots, 2^{p-1}$ cover all possible combinations; c^* maximizes $c^2(\Psi(c))^r$ where Ψ is defined in (1.8) and r is the length of $\boldsymbol{\beta}$.

The proof of Theorem 3.3.1 follows the same arguments as in the proof of Theorem 3.2.1, except that the canonical form transformation (3.3) now contains more terms to incorporate the additional interaction effects. For instance, if x_1x_2 is the only interaction in the model, then $\tilde{\mathbf{x}}$ becomes $(1, x_1, x_2, x_1x_2, x_3, \dots, x_p)$, $\tilde{\mathbf{v}} = (1, v_1, v_2, v_1v_2, v_3, \dots, v_p)$ and matrix B will have an additional row and column like this

$$B = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ -a_1/b_1 & 1/b_1 & 0 & \cdots & \cdots & \cdots & 0 \\ -a_2/b_2 & 0 & 1/b_2 & 0 & \cdots & \cdots & 0 \\ a_1a_2/b_1b_2 & -a_2/b_1b_2 & -a_1/b_1b_2 & 1/b_1b_2 & 0 & \cdots & 0 \\ & & \vdots & & & & \\ -a_{p-1}/b_{p-1} & 0 & \cdots & \cdots & 0 & 1/b_{p-1} & 0 \\ \beta_0 & \beta_1 & \beta_2 & \beta_{12} & \cdots & \beta_{p-1} & \beta_p \end{pmatrix}.$$

In general, Equation (3.5) becomes

$$\begin{aligned} \Psi(c)\tilde{\mathbf{C}}^T I_{\xi_v}^{-1} \tilde{\mathbf{C}} &= \frac{\Psi(c)}{\Psi(c^*)} \left\{ 1 + v_1^2 + \cdots + v_{p-1}^2 + \sum_{t=2}^L \sum_{(l_1, \dots, l_t) \in H_t} \left(\prod_{m=1}^t v_{l_m}^2 \right) + \frac{c^2}{(c^*)^2} \right\} \\ &\leq \frac{\Psi(c)}{\Psi(c^*)} (r-1) + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)}, \quad c \in (-\infty, \infty) \end{aligned}$$

and the rest follows immediately from the previous proof.

Here is an example to illustrate Theorem 3.3.1.

Example 3.3.2. Similar to Example 3.2.2, but now with the interaction effect between x_1 and x_2 in the model; so $\mathbf{X}_i^T \boldsymbol{\beta} = c_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \beta_3 x_{i3}$. Then, according to Theorem 3.3.1, a locally D-optimal design for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_{12}, \beta_3)^T = (1, -1, 0.5, 1, 1)^T$ can be obtained and is shown in Table 3.2. The value c^* maximizes $c^2(\Psi(c))^{4+1}$, which is approximately 0.9254.

Table 3.2: Support Points and Weights for a Locally D-optimal Design.

Support points	Weights	Support points	Weights
$(0, -1, -1.4254)$	$1/8$	$(2, -1, 2.5746)$	$1/8$
$(0, -1, 0.4254)$	$1/8$	$(2, -1, 4.4254)$	$1/8$
$(0, 1, -2.4254)$	$1/8$	$(2, 1, -2.4254)$	$1/8$
$(0, 1, -0.5746)$	$1/8$	$(2, 1, -0.5746)$	$1/8$

3.4 Smaller Optimal Designs Based on Orthogonal Arrays

Theorem 3.3.1 provides a nice and relatively simple structure for locally D-optimal designs. However, as the number of covariates p becomes larger, the number of distinct support points increases rapidly. For instance, when p increases from 4 to 8, the support size increases from 16 to 256. The support size being that large is usually unnecessary since the number of parameters being estimated is generally much smaller. In fact, as observed in Sitter and Torsney (1995b), for main-effect models, smaller D-optimal designs can usually be constructed based on a subset of those 2^p support points. They implemented a method using Hadamard matrices and successfully found a 16-point D-optimal design when $p = 8$.

Recall in Chapter 2, we use strength $t+$ orthogonal arrays to find smaller D-optimal designs. The idea can also be applied here for multiple-covariate GLMs problems. In fact, when no interaction effects exists, we can relax the strength of OAs from $t+$ to 2. But since our main focus in this chapter is on interaction models, we will skip those results. As for interaction models, we present the following theorems without proofs. The reason for omitting the proofs is that they can be directly generated following the same arguments used for the proofs in Section 2.3.

Theorem 3.4.1. *Consider design ξ^* defined in Theorem 3.3.1, for model (3.1) with only one interaction effect, namely the two-way interaction x_1x_2 . Assume that an orthogonal array $OA(N, 2^{p-1}, 2+)$ exists so that columns $(1, 2, j)$, $j \neq 1$ or 2 form an $OA(N, 2^3, 3)$. Let H denote the set of rows in an $OA(N, 2^{p-1}, 2+)$ and define $\xi_1 = \{(\mathbf{C}_{i1}^*, \frac{1}{2N}) \& (\mathbf{C}_{i2}^*, \frac{1}{2N}), l = 1, \dots, N\}$ where $\mathbf{C}_{i1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$ and $\mathbf{C}_{i2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$. Here h_{lj} is either L_j or U_j , c^* is the same as in Theorem 3.3.1 and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N$ cover all possible combinations that correspond to the rows in H ; Then $I_{\xi_1}(\boldsymbol{\beta}) = I_{\xi^*}(\boldsymbol{\beta})$, so that ξ_1 is also a D -optimal design for $\boldsymbol{\beta}$.*

In Theorem 3.4.1, we need two support points in each combination of the first $p-1$ covariates, the next theorem shows that when the orthogonal array has an additional 2-level column, it is possible to use fewer support points to obtain optimal designs.

Theorem 3.4.2. *Under the same assumptions as in Theorem 3.4.1, suppose that there exists an $OA(N, 2^p, 2+)$ and columns $(1, 2, j)$, $j \neq 1$ or 2 form an $OA(N, 2^3, 3)$. Then*

$$I_{\xi_2}(\boldsymbol{\beta}) = I_{\xi^*}(\boldsymbol{\beta}),$$

where ξ_2 is obtained in the following way: Let H_1 and H_2 be the $\frac{N}{2} \times (p-1)$ subarrays obtained by taking the $\frac{N}{2}$ rows of the OA in which the last entry is 1 and 2, respectively, and deleting the last entry. Then $\xi_2 = \{(\mathbf{C}_{i1}^*, \frac{1}{N}), l = 1, \dots, \frac{N}{2}\} \cup \{(\mathbf{C}_{i2}^*, \frac{1}{N}), l = 1, \dots, \frac{N}{2}\}$, where $\mathbf{C}_{i1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$, with h_{lj} being either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N/2$ cover all possible combinations that correspond to the rows in H_1 ; $\mathbf{C}_{i2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$, with h_{lj} being either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N/2$ cover all possible combinations that correspond to the rows in H_2 .

For two 2-way interactions, we can obtain similar results.

Theorem 3.4.3. Consider design ξ^* in Theorem 3.3.1, for model (3.1) with two 2-way interaction effects. If they have a common factor, say x_1x_2 and x_1x_3 , and an orthogonal array $OA(N, 2^{p-1}, 2+)$ exists with columns $(1, 2, j)$ and columns $(1, 3, j)$, $j = 4, \dots, p-1$ being of strength 3 orthogonal arrays. Let H denote the set of groups in this $OA(N, 2^{p-1}, 2+)$, then

$$I_{\xi_3}(\beta) = I_{\xi^*}(\beta),$$

where $\xi_3 = \{(\mathbf{C}_{i1}^*, \frac{1}{2N}) \& (\mathbf{C}_{i2}^*, \frac{1}{2N}), l = 1, \dots, N\}$, $\mathbf{C}_{i1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$ and $\mathbf{C}_{i2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$. Here h_{lj} is either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N$ cover all possible combinations that correspond to the rows in H .

Moreover, if an $OA(N, 2^p, 2+)$ exists with columns $(1, 2, j)$ and columns $(1, 3, j)$, $j = 4, \dots, p$ being of strength 3 OAs, then

$$I_{\xi_4}(\beta) = I_{\xi^*}(\beta)$$

where ξ_4 is obtained in the following way: Let H_1 and H_2 be the $\frac{N}{2} \times (p-1)$ subarrays obtained by taking the $\frac{N}{2}$ rows of the OA in which the last entry is 1 and 2, respectively, and deleting the last entry. Then $\xi_4 = \{(\mathbf{C}_{i1}^*, \frac{1}{N}), l = 1, \dots, \frac{N}{2}\} \cup \{(\mathbf{C}_{i2}^*, \frac{1}{N}), l = 1, \dots, \frac{N}{2}\}$, where $\mathbf{C}_{i1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$, with h_{lj} being either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N/2$ cover all possible combinations that correspond to the rows in H_1 ; $\mathbf{C}_{i2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$, with h_{lj} being either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N/2$ cover all possible combinations that correspond to the rows in H_2 .

Theorem 3.4.4. Consider design ξ^* in Theorem 3.3.1, for model (3.1) with two 2-way interaction effects, if they don't have a common factor, say x_1x_2 and x_3x_4 , then if an orthogonal array $OA(N, 2^{p-1}, 2)$ exists in which columns $(1, 2, j)$ and columns

$(3, 4, j)$ $j = 5, \dots, p-1$ form strength 3 OAs and columns $(1, 2, 3, 4)$ form a strength 4 OA, then with H being the set of groups in this $OA(N, 2^{p-1}, 2)$, we have

$$I_{\xi_5}(\beta) = I_{\xi^*}(\beta),$$

where $\xi_5 = \{(\mathbf{C}_{i1}^*, \frac{1}{2N}) \& (\mathbf{C}_{i2}^*, \frac{1}{2N}), l = 1, \dots, N\}$ where $\mathbf{C}_{i1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$ and $\mathbf{C}_{i2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$. Here h_{lj} is either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N$ cover all possible combinations that correspond to the rows in H .

Moreover, if an $OA(N, 2^p, 2)$ exists with columns $(1, 2, j)$ and columns $(3, 4, j)$ $j = 5, \dots, p$ being of strength 3 and columns $(1, 2, 3, 4)$ being of strength 4, Then

$$I_{\xi_6}(\beta) = I_{\xi^*}(\beta),$$

where ξ_6 is obtained in the following way: Let H_1 and H_2 be the $\frac{N}{2} \times (p-1)$ subarrays obtained by taking the $\frac{N}{2}$ rows of the OA in which the last entry is 1 and 2, respectively, and deleting the last entry. Then $\xi_6 = \{(\mathbf{C}_{i1}^*, \frac{1}{N}), l = 1, \dots, \frac{N}{2}\} \cup \{(\mathbf{C}_{i2}^*, \frac{1}{N}), l = 1, \dots, \frac{N}{2}\}$, where $\mathbf{C}_{i1}^* = (h_{l1}, \dots, h_{l,p-1}, c^*)^T$, with h_{lj} being either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N/2$ cover all possible combinations that correspond to the rows in H_1 ; $\mathbf{C}_{i2}^* = (h_{l1}, \dots, h_{l,p-1}, -c^*)^T$, with h_{lj} being either L_j or U_j and $(h_{l1}, \dots, h_{l,p-1}), l = 1, \dots, N/2$ cover all possible combinations that correspond to the rows in H_2 .

The following example is an illustration of Theorem 3.4.3.

Example 3.4.5. Consider model (3.1) with logit link and

$$\tilde{\mathbf{x}}_i^T \beta = c_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_5 x_{i5}.$$

Assume that the first four covariates are restricted to $[-1, 1]$, $[-2, 2]$, $[-1, 1]$ and $[-0.5, 0.5]$, respectively, and that there is no restriction on the last (fifth) covariate. To

find locally optimal designs, we further assume that $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_{12}, \beta_{13}, \beta_5)^T = (1, -0.5, 0.5, -1, 1, -0.5, 0.5, 1)^T$. Notice that there are two 2-way interactions in the model and they have a common covariate x_1 . Then, according to Theorem 3.4.3, we want to find an $OA(N, 2^4, 2+)$ so that the columns (1,2,4) and (1,3,4) both form strength 3 orthogonal arrays. Table 3.3 presents such an OA for $N = 8$. In fact, this is an $OA(8, 2^4, 3)$.

Table 3.3: An $OA(8, 2^4, 2+)$ ($OA(8, 2^4, 3)$) with Columns (1,2,4) and (1,3,4) Being of Strength 3.

1	1	1	2
1	1	2	1
1	2	1	1
1	2	2	2
2	1	1	1
2	1	2	2
2	2	1	2
2	2	2	1

Then, based on Theorem 3.4.3, a locally D-optimal design can be obtained as shown in Table 3.4. For illustration purpose, we use lower bound L_i and upper bound U_i instead of their real values. Also we use c^* and $-c^*$ to replace the real values for the unbounded covariate x_5 . In this case, the value c^* maximizes $c^2(\Psi(c))^8$, which is approximately 0.7222. Notice that the design shown in Table 3.4 only has 16 distinct support points compared with the 32-point design obtained from Theorem 3.3.1.

In fact, the second part of Theorem 3.4.3 can help us obtain an even smaller

Table 3.4: A Smaller D-optimal Design Based on Orthogonal Arrays

(x_1, x_2, x_3, x_4)	x_5	(x_1, x_2, x_3, x_4)	x_5
(L_1, L_2, L_3, U_4)	$c^*; -c^*$	(U_1, L_2, L_3, L_4)	$c^*; -c^*$
(L_1, L_2, U_3, L_4)	$c^*; -c^*$	(U_1, L_2, U_3, U_4)	$c^*; -c^*$
(L_1, U_2, L_3, L_4)	$c^*; -c^*$	(U_1, U_2, L_3, U_4)	$c^*; -c^*$
(L_1, U_2, U_3, U_4)	$c^*; -c^*$	(U_1, U_2, U_3, L_4)	$c^*; -c^*$

optimal design. In which case, we need an $OA(N, 2^5, 2+)$ so that columns (1,2,4), (1,2,5), (1,3,4) and (1,3,5) all form strength 3 orthogonal arrays. The OA in Table 3.5 satisfies all these requirements. Then based on the same theorem, a smaller D-optimal design can be obtained and is shown in Table 3.6. Notice that there are only 8 distinct support points in this design and there are exactly 8 parameters in β that need to be estimated, so this design is actually a saturated D-optimal design (see Hu *et al.* (2015) for more details).

3.5 Summaries and Discussions

In this chapter, we proposed several locally D-optimal designs for multiple-covariate GLMs under a very general setup. There is only one covariate that is unbounded, while the others are bounded in intervals. Hierarchical-structure interactions among those bounded covariates may or may not exist. It turns out that D-optimal designs for such models share a very similar structure as the ones in the previous chapter. Meanwhile, when the number of covariates is large, smaller D-optimal designs are in great need. The OA-based results in Chapter 2 can also be extended here and successfully reduce the number of support points required for D-optimal designs.

Table 3.5: An $OA(8, 2^5, 2+)$ with Columns (1,2,4), (1,2,5), (1,3,4) and (1,3,5) Being of Strength 3.

1	1	1	2	1
1	1	2	1	2
1	2	1	1	2
1	2	2	2	1
2	1	1	1	1
2	1	2	2	2
2	2	1	2	2
2	2	2	1	1

Table 3.6: A Smaller D-optimal Design Based on Orthogonal Arrays

(x_1, x_2, x_3, x_4)	x_5	(x_1, x_2, x_3, x_4)	x_5
(L_1, L_2, L_3, U_4)	$-c^*$	(U_1, L_2, L_3, L_4)	$-c^*$
(L_1, L_2, U_3, L_4)	c^*	(U_1, L_2, U_3, U_4)	c^*
(L_1, U_2, L_3, L_4)	c^*	(U_1, U_2, L_3, U_4)	c^*
(L_1, U_2, U_3, U_4)	$-c^*$	(U_1, U_2, U_3, L_4)	$-c^*$

One main difference between models in Chapter 2 and 3 is that in Chapter 3, the assumption we make on the interaction effects is stronger. In Chapter 2, we assume the interactions can be involved in a more flexible way while in Chapter 3, we require a hierarchical structure for the interactions. The main reason for making such an assumption is due to the proof method we use. How to relax this assumption and

obtain more general results remain open questions.

Another interesting observation is, even with cutting-edge search algorithms, optimal designs that we obtained are not of high accuracy. Most of the traditional search algorithms are based on grid-search, as a result, it takes much more time if we use finer grids during the search. We mentioned earlier that meta-heuristic algorithms like PSO are getting more and more attention. Such algorithms do not rely on grid-search and because of this, the computation time can be greatly reduced. Some of this has been done (see Qiu *et al.* (2014) and Chen *et al.* (2015)). It is worthwhile to apply such algorithms to our models and hopefully propose a faster and more efficient algorithm.

TOPIC III: D-OPTIMAL DESIGNS FOR LOGISTIC MODELS WITH MULTIPLE COVARIATES

As introduced in Chapter 3, we consider again multiple-covariate GLM problems in this chapter. But now the restrictions on parameters and design space are different. We assume in the following sections that all covariates are non-negative and there is no interaction effect among them. We further assume that the coefficients for all covariates are of the same sign, for instance, all positive, and we allow the intercept to vary in a wide range. Models with such assumptions are often seen in clinical studies, where the covariates represent the dose levels of active drugs. There are however limited literatures focusing on the optimal design problems for such models. Sitter and Torsney (1995a) used a canonical transformation to study GLMs with two design variables without interaction. They transformed the original design space into logit-value space and imposed parallel restrictions on the logits. Jia and Myers (2001) proposed hyperbola-based designs for two-covariate logistic models with interaction. Haines *et al.* (2007) investigated D-optimal design structures for main-effect logistic models with two covariates. They observed that under certain restrictions, there were two patterns of D-optimal designs, one with 4 support points and the other with 3. Kabera (2009) provided theoretical justifications of Haines *et al.* (2007)'s findings and proved that D-optimal design structures depend on a cutoff value for the intercept parameter.

In this chapter, we extend results from Haines *et al.* (2007) and Kabera (2009) to logistic models with n covariates ($n \geq 2$) and no interaction effects. We observe that there are still two D-optimal design patterns, only now one of them includes $2n$

support points and the other one has $n + 1$, depending on the same cutoff value of the intercept parameter. This chapter is organized as follows. In Section 4.1, we introduce the models of interest. Section 4.2 provides a theoretical result for D-optimal designs when the intercept is greater than the cutoff value. The global D-optimality is proved using the equivalence theorem. When the intercept is less than the cutoff value, we observed the $2n$ -point optimal design pattern, and that is proved in Section 4.3 for selected values of the intercept. Summaries and discussions are provided in the end.

4.1 Model Description

As introduced earlier, we consider the following model:

$$u = \text{logit}(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p, \quad (4.1)$$

where P is the probability that an event occurs, and $x_i \geq 0$, $i = 1, \dots, p$, are p independent covariates. We also assume that the coefficients corresponding to these x_i 's have the same sign. For simplicity, say $\beta_i \geq 0$, $i = 1, \dots, p$. In fact, if some of the β_i 's have opposite signs, then we are able to show that for certain designs, the determinant of information matrices can go to infinity when the covariates go to infinity, which means there will be no D-optimal designs (See Proposition 4.1.1 below). The intercept β_0 has no restrictions at this stage and the entire parameter vector is denoted by $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$.

Let $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ be an arbitrary design point under model (4.1). Then the information matrix evaluated at this point is given by

$$I_{\mathbf{x}_i} = \frac{e^{u_i}}{(1 + e^{u_i})^2} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T,$$

where $u_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ and $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i^T)^T$.

Then for an approximate design

$$\xi = \left\{ \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \\ w_1 & w_2 & \cdots & w_k \end{array} \right\},$$

the information matrix becomes

$$I_\xi = \sum_{i=1}^k w_i I_{\mathbf{x}_i} = \sum_{i=1}^k w_i \frac{e^{u_i}}{(1 + e^{u_i})^2} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T.$$

We can also formulate the design problem in canonical form, i.e.

$$u = \text{logit}(p) = \beta_0 + z_1 + z_2 + \cdots + z_p, \quad (4.2)$$

where $z_i = \beta_i x_i \geq 0$, $i = 1, \dots, p$. Similarly, the information matrix evaluated at $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})^T = (\beta_1 x_{i1}, \dots, \beta_p x_{ip})^T$ is given by

$$M_{\mathbf{z}_i} = \frac{e^{u_i}}{(1 + e^{u_i})^2} \tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^T = B I_{\mathbf{x}_i} B^T,$$

where $\tilde{\mathbf{z}}_i = (1, \mathbf{z}_i^T)^T$ and $B = \text{diag}(1, \beta_1, \dots, \beta_p)$.

Meanwhile, the transformation of model (4.1) to model (4.2) transforms the design ξ into

$$\xi_z = \begin{Bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_k \\ w_1 & w_2 & \cdots & w_k \end{Bmatrix}.$$

Note that design ξ and design ξ_z are defined on the same design space $[0, \infty)^p$. Also, the information matrix for ξ_z is

$$M_{\xi_z} = \sum_{i=1}^k w_i M_{\mathbf{z}_i} = \sum_{i=1}^k w_i B I_{\mathbf{x}_i} B^T = B I_{\xi} B^T.$$

Since $\det(I_{\xi_z}) = \det(B I_{\xi} B^T) = \det(B)^2 \cdot \det(I_{\xi})$, it implies that maximizing $\det(I_{\xi})$ is equivalent to maximizing $\det(I_{\xi_z})$. Therefore, in the following discussions we will focus on model (4.2) only.

The following proposition explains why all β_i 's, $i = 1, \dots, p$, must have the same sign.

Proposition 4.1.1. *For model (4.1), if there exist some β_i 's that have opposite signs, then a D-optimal design does not exist.*

Proof. Without loss of generality, assume the first m x_i 's have positive coefficients and the rest have negative coefficients. Then, following the same transformation as shown in (4.2), it is sufficient to consider

$$u = \text{logit}(p) = \beta_0 + z_1 + \cdots + z_m - z_{m+1} - \cdots - z_p, \quad (4.3)$$

on design space $[0, \infty)^p$.

If $\beta_0 \geq 0$, then finding a D-optimal design for (4.3) is equivalent to finding a D-optimal design for

$$u = \text{logit}(p) = z_1 + \cdots + z_m - z_{m+1} - \cdots - z_p \quad (4.4)$$

on the design space of $z_i \geq 0$, $i = 1, \dots, p-1$ and $z_p \geq -\beta_0$.

Or if $\beta_0 < 0$, then finding a D-optimal design for (4.3) is equivalent to finding a D-optimal design for (4.4) on the design space of $z_1 \geq \beta_0$ and $z_i \geq 0$, $i = 2, \dots, p$.

In either case, design

$$\xi_t = \left\{ \begin{array}{ccccc} (1, 0, \dots, 0) & (0, 1, 0, \dots, 0) & \cdots & (0, \dots, 0, 1) & (t, 0, \dots, 0, t) \\ \frac{1}{p+1} & \frac{1}{p+1} & \frac{1}{p+1} & \cdots & \frac{1}{p+1} \end{array} \right\},$$

where $t \geq 0$, is always a valid design candidate for model (4.4).

Then for model (4.4), the information matrix for design ξ_t can be written as

$$M_{\xi_t} = \frac{1}{p+1} \begin{pmatrix} p\Psi(1) + \Psi(0) & \Psi(1) + t\Psi(0) & \Psi(1) & \cdots & \Psi(1) + t\Psi(0) \\ \Psi(1) + t\Psi(0) & \Psi(1) + t^2\Psi(0) & 0 & \cdots & t^2\Psi(0) \\ \Psi(1) & 0 & \Psi(1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Psi(1) + t\Psi(0) & t^2\Psi(0) & 0 & \cdots & \Psi(1) + t^2\Psi(0) \end{pmatrix}$$

and the determinant of the information matrix M_{ξ_t} is

$$\det(M_{\xi_t}) = \frac{1}{(p+1)^{p+1}} \Psi(0)\Psi(1)^p (2t-1)^2.$$

With p being fixed, we know that $\det(\xi_t) \rightarrow \infty$ as $t \rightarrow \infty$. Therefore, the determinant of the information matrix can be arbitrarily large, i.e. no D-optimal design exists.

□

4.2 D-optimal Designs with $p + 1$ Support Points

Before stating the main theorem, we first establish two lemmas.

Lemma 4.2.1. *Equation*

$$1 + c + e^c - ce^c = 0 \quad (4.5)$$

has a unique solution for $c \geq 0$, say $\tilde{c}$.

Proof. See Appendix B. □

Lemma 4.2.2. *Equation*

$$2 - \beta_0 + c + 2e^c + \beta_0 e^c - ce^c = 0, \quad (4.6)$$

has a unique solution for $c \geq 0$, namely c^* . When $\beta_0 \in [-\tilde{c}, 0]$, $c^* \geq -\beta_0$; when $\beta_0 \in (0, \infty)$, $c^* > \beta_0$.

Proof. See Appendix B. □

We now state the main theorem.

Theorem 4.2.3. *For Model (4.2) with parameter $\beta_0 \in [-\tilde{c}, \infty)$, where $\tilde{c}$ being the unique positive solution of Equation (4.5) with an approximate value of 1.5434, the design*

$$\xi^* = \left\{ \begin{array}{cccccc} (0, \dots, 0) & (c^* - \beta_0, 0, \dots, 0) & (0, c^* - \beta_0, 0, \dots, 0) & \dots & (0, \dots, 0, c^* - \beta_0) \\ \frac{1}{p+1} & \frac{1}{p+1} & \frac{1}{p+1} & \dots & \frac{1}{p+1} \end{array} \right\},$$

where $c^* > 0$ satisfies Equation (4.6), is D-optimal.

Proof. Let M_{ξ^*} be the information matrix of the proposed design ξ^* , then we have

$$M_{\xi^*} = \frac{1}{p+1} \begin{pmatrix} \Psi(\beta_0) + p\Psi(c^*) & (c^* - \beta_0)\Psi(c^*) & (c^* - \beta_0)\Psi(c^*) & \cdots & (c^* - \beta_0)\Psi(c^*) \\ (c^* - \beta_0)\Psi(c^*) & (c^* - \beta_0)^2\Psi(c^*) & 0 & \cdots & 0 \\ (c^* - \beta_0)\Psi(c^*) & 0 & (c^* - \beta_0)^2\Psi(c^*) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ (c^* - \beta_0)\Psi(c^*) & 0 & \cdots & 0 & (c^* - \beta_0)^2\Psi(c^*) \end{pmatrix},$$

where $\Psi(x) = e^x/(1 + e^x)^2$ as defined in (1.8). Then

$$M_{\xi^*}^{-1} = (p+1) \times \begin{pmatrix} \frac{1}{\Psi(\beta_0)} & \frac{-1}{(c^* - \beta_0)\Psi(\beta_0)} & \frac{-1}{(c^* - \beta_0)\Psi(\beta_0)} & \cdots & \frac{-1}{(c^* - \beta_0)\Psi(\beta_0)} \\ \frac{-1}{(c^* - \beta_0)\Psi(\beta_0)} & \frac{\Psi(\beta_0) + \Psi(c^*)}{(c^* - \beta_0)^2\Psi(\beta_0)\Psi(c^*)} & \frac{1}{(c^* - \beta_0)^2\Psi(\beta_0)} & \cdots & \frac{1}{(c^* - \beta_0)^2\Psi(\beta_0)} \\ \frac{-1}{(c^* - \beta_0)\Psi(\beta_0)} & \frac{1}{(c^* - \beta_0)^2\Psi(\beta_0)} & \frac{\Psi(\beta_0) + \Psi(c^*)}{(c^* - \beta_0)^2\Psi(\beta_0)\Psi(c^*)} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{1}{(c^* - \beta_0)^2\Psi(\beta_0)} \\ \frac{-1}{(c^* - \beta_0)\Psi(\beta_0)} & \frac{1}{(c^* - \beta_0)^2\Psi(\beta_0)} & \cdots & \frac{1}{(c^* - \beta_0)^2\Psi(\beta_0)} & \frac{\Psi(\beta_0) + \Psi(c^*)}{(c^* - \beta_0)^2\Psi(\beta_0)\Psi(c^*)} \end{pmatrix}.$$

Let $\mathbf{z} = (z_1, z_2, \dots, z_p)^T = (z_1, z_2, \dots, z_{p-1}, c - \sum_{i=1}^{p-1} z_i)^T$ be an arbitrary support point in the design space $[0, \infty)^p$, where $c = \text{logit}(P) = \beta_0 + z_1 + z_2 + \cdots + z_p$. Clearly, $c \geq \beta_0$. Then according to the equivalence theorem presented in (1.18), ξ^* is D-optimal if and only if

$$\max d(\mathbf{z}, \xi^*) = \max \Psi(c)(1, \mathbf{z}^T)M_{\xi^*}^{-1}(1, \mathbf{z}^T)^T = p+1, \text{ for any } \mathbf{z} \in [0, \infty)^p. \quad (4.7)$$

Then we have

$$\begin{aligned} d(\mathbf{z}, \xi^*) &= \Psi(c)(1, \mathbf{z}^T)M_{\xi^*}^{-1}(1, \mathbf{z}^T)^T \\ &= (p+1)\Psi(c) \left\{ \frac{1}{\Psi(\beta_0)}(c - c^*)^2 + \frac{1}{\Psi(c^*)} \{ (c - \beta_0)^2 + 2[z_1^2 - (c - \beta_0)z_1] + \cdots + \right. \\ &\quad \left. 2[z_{p-1}^2 - (c - \beta_0)z_{p-1}] + \sum_{i \neq j} z_i z_j \} \right\} / (c^* - \beta_0)^2 \end{aligned}$$

$$\begin{aligned}
&\leq \frac{(p+1)\Psi(c)\left\{\frac{1}{\Psi(\beta_0)}(c-c^*)^2 + \frac{1}{\Psi(c^*)}\{(c-\beta_0)^2\}\right\}}{(c^*-\beta_0)^2} \\
&= \frac{(p+1)e^c[e^{-\beta_0}(1+e^{\beta_0})^2(c-c^*)^2 + e^{-c^*}(1+e^{c^*})^2(c-\beta_0)^2]}{(1+e^c)^2(c^*-\beta_0)^2}.
\end{aligned}$$

The inequality follows from the following Lemma.

Lemma 4.2.4. *Let $f(x_1, \dots, x_n) = x_1^2 - ax_1 + \dots + x_n^2 - ax_n + \sum_{i \neq j} x_i x_j$, $a > 0$, $n \geq 2$, with $x_i \geq 0$, $i = 1, \dots, n$, and $\sum_{i=1}^n x_i \leq a$, then*

$$f(x_1, \dots, x_n) \leq 0.$$

Proof. See Appendix B. □

When $\mathbf{z}$ is one of the support points in ξ^* , it is easy to verify that the equality sign in Equation (4.7) holds. Therefore, to complete the proof, it is sufficient to show that

$$g(c) \triangleq \frac{e^c[e^{-\beta_0}(1+e^{\beta_0})^2(c-c^*)^2 + e^{-c^*}(1+e^{c^*})^2(c-\beta_0)^2]}{(1+e^c)^2(c^*-\beta_0)^2} \leq 1, \text{ for all } c \geq \beta_0. \quad (4.8)$$

Next, we consider two situations for β_0 : (i) $\beta_0 = -\tilde{c}$ and (ii) $\beta_0 > -\tilde{c}$.

For situation (i), when $\beta_0 = -\tilde{c}$, from Equation (4.5), we have

$$1 - \beta_0 + e^{-\beta_0} + \beta_0 e^{-\beta_0} = 0.$$

Then, according to (4.6), we have

$$2 - \beta_0 + (-\beta_0) + 2e^{-\beta_0} + \beta_0 e^{-\beta_0} - (-\beta_0)e^{-\beta_0} = 2(1 - \beta_0 + e^{-\beta_0} + \beta_0 e^{-\beta_0} = 0) = 0.$$

Therefore, $c = -\beta_0$ is the solution of Equation (4.6), i.e. $c^* = -\beta_0$. Then $g(c)$ in (4.8) becomes

$$g(c) = \frac{e^c}{2(1+e^c)^2} \left(\frac{e^{\tilde{c}}}{(1+e^{\tilde{c}})^2} \right)^{-1} \left[1 + \frac{c^2}{(\tilde{c})^2} \right].$$

To show (4.4), it is equivalent to show

$$\frac{e^c}{(1+e^c)^2} \left(\frac{e^{\tilde{c}}}{(1+e^{\tilde{c}})^2} \right)^{-1} \left[1 + \frac{c^2}{(\tilde{c})^2} \right] = \frac{\Psi(c)}{\Psi(\tilde{c})} \left[1 + \frac{c^2}{(\tilde{c})^2} \right] \leq 2,$$

which has been shown in Lemma 2.2.4 from Theorem 2.2.1.

For situation (ii), when $\beta_0 > -\tilde{c}$. Since $c \geq \beta_0$, we can write c as $c = \beta_0 + \alpha(c^* - \beta_0)$ with $\alpha \in [0, \infty)$. Specifically, $\alpha = 0$ and $\alpha = 1$ correspond to $c = \beta_0$ and $c = c^*$ respectively. Therefore, $g(c)$ in (4.8) can be written in terms of α as follows:

$$g(\alpha) = \frac{e^{\beta_0 + \alpha(c^* - \beta_0)} [\alpha^2 e^{-c^*} (1 + e^{c^*})^2 + (\alpha - 1)^2 e^{-\beta_0} (1 + e^{\beta_0})^2]}{[1 + e^{\beta_0 + \alpha(c^* - \beta_0)}]^2}.$$

So now we only need to show that $g(\alpha) \leq 1$ for any $\alpha \in [0, \infty)$.

In fact, it is easy to verify that $g(0) = g(1) = 1$. Next we will prove that the function $g(\alpha)$ reaches its maximum at $\alpha = 0$ and $\alpha = 1$, attains a local minimum at some $\alpha \in (0, 1)$, and goes to 0 as α goes to infinity.

To investigate the property of $g(\alpha)$, consider its first derivative

$$g'(\alpha) = \frac{e^{\beta_0 + \alpha(c^* - \beta_0)} [h_1(\alpha) - h_2(\alpha)]}{[1 + e^{\beta_0 + \alpha(c^* - \beta_0)}]^3},$$

where

$$h_1(\alpha) = [\alpha^2(c^* - \beta_0) + 2\alpha]e^{-c^*}(1 + e^{c^*})^2 + [(\alpha - 1)^2(c^* - \beta_0) + 2(\alpha - 1)]e^{-\beta_0}(1 + e^{\beta_0})^2$$

is a convex quadratic function and

$$h_2(\alpha) = e^c \{ [\alpha^2(c^* - \beta_0) - 2\alpha]e^{-c^*}(1 + e^{c^*})^2 + [(\alpha - 1)^2(c^* - \beta_0) - 2(\alpha - 1)]e^{-\beta_0}(1 + e^{\beta_0})^2 \},$$

where $c = \beta_0 + \alpha(c^* - \beta_0)$. Notice that the number of stationary points of $g(\alpha)$ is the number of solutions of $g'(\alpha) = 0$, which is also the number of solutions for $h_1(\alpha) = h_2(\alpha)$. In addition, since c^* is the solution of Equation (4.6), we have

$$g'(1) = \frac{2 - \beta_0 + c^* + 2e^{c^*} + \beta_0 e^{c^*} - c^* e^{c^*}}{1 + e^{c^*}} = 0,$$

which means that $h_1(\alpha)$ and $h_2(\alpha)$ meet at $\alpha = 1$. The next lemma will show that they cross at $\alpha = 1$.

Lemma 4.2.5. $h_1(\alpha)$ and $h_2(\alpha)$ cross at $\alpha = 1$.

Proof. See Appendix B. □

Next we will show that curve $h_2(\alpha)$ has only one inflection point on $[0, \infty)$ and intersects $h_1(\alpha)$ only once other than at $\alpha = 1$.

Take the second derivative of $h_2(\alpha)$,

$$\begin{aligned} h_2''(\alpha) = (c^* - \beta_0)e^{\beta_0 + \alpha(c^* - \beta_0)} & \left\{ (c^* - \beta_0)^2 [e^{-c^*}(1 + e^{c^*})^2 + e^{-\beta_0}(1 + e^{\beta_0})^2] \alpha^2 \right. \\ & + 2(c^* - \beta_0)[e^{-c^*}(1 + e^{c^*})^2 + (1 - c^* + \beta_0)e^{-\beta_0}(1 + e^{\beta_0})^2] \alpha \\ & \left. - 2e^{-c^*}(1 + e^{c^*})^2 - 2(c^* - \beta_0 + 1)e^{-\beta_0}(1 + e^{\beta_0})^2 + (c^* - \beta_0)^2 e^{-\beta_0}(1 + e^{\beta_0})^2 \right\}. \end{aligned}$$

So the inflection points are those that satisfy $h_2''(\alpha) = 0$. Since $(c^* - \beta_0)e^{\beta_0 + \alpha(c^* - \beta_0)} > 0$, thus to find the solutions of $h_2''(\alpha) = 0$ is equivalent to find the zeroes of

$$\begin{aligned} h_3(\alpha) = (c^* - \beta_0)^2 & [e^{-c^*}(1 + e^{c^*})^2 + e^{-\beta_0}(1 + e^{\beta_0})^2] \alpha^2 \\ & + 2(c^* - \beta_0)[e^{-c^*}(1 + e^{c^*})^2 + (1 - c^* + \beta_0)e^{-\beta_0}(1 + e^{\beta_0})^2] \alpha \\ & - 2e^{-c^*}(1 + e^{c^*})^2 - 2(c^* - \beta_0 + 1)e^{-\beta_0}(1 + e^{\beta_0})^2 + (c^* - \beta_0)^2 e^{-\beta_0}(1 + e^{\beta_0})^2. \end{aligned}$$

Notice that $h_3(\alpha)$ is a convex quadratic function, with

$$h_3(0) = -2e^{-c^*}(1 + e^{c^*})^2 - 2(c^* - \beta_0 + 1)e^{-\beta_0}(1 + e^{\beta_0})^2 + (c^* - \beta_0)^2 e^{-\beta_0}(1 + e^{\beta_0})^2.$$

The next lemma states that $h_3(0) < 0$.

Lemma 4.2.6. $h_3(0) < 0$.

Proof. See Appendix B. □

Now we know that $h_3(\alpha)$ is a convex parabola with $h_3(0) < 0$, so there exists one and only one positive root for $h_3(\alpha) = 0$ on $[0, \infty)$, denoted by α_0 . Consider

$$h_3(1) = [(c^* - \beta_0)^2 + 2(c^* - \beta_0 - 1)]e^{-c^*}(1 + e^{c^*})^2 - 2e^{-\beta_0}(1 + e^{\beta_0})^2$$

$$\begin{aligned}
&\geq [(c^* - \beta_0)^2 + 2(c^* - \beta_0 - 1)]e^{-\beta_0}(1 + e^{\beta_0})^2 - 2e^{-\beta_0}(1 + e^{\beta_0})^2 \\
&= [(c^* - \beta_0 + 1)^2 - 5]e^{-\beta_0}(1 + e^{\beta_0})^2.
\end{aligned}$$

The inequality comes from the fact that $e^{-c^*}(1 + e^{c^*})^2 \geq e^{-\beta_0}(1 + e^{\beta_0})^2$ which has been shown in the proof of Lemma 4.2.5. Also, from the proof of Lemma 4.2.6 we know $(c^* - \beta_0 + 1)^2 - 5 > 0$. Therefore, $h_3(1) > 0$. That means $\alpha = 1$ belongs to the subset of $[0, \infty)$ where $h_2(\alpha)$ is convex. Hence $h_2(\alpha)$ is concave on $[0, \alpha_0]$ and convex on $[\alpha_0, \infty]$.

Recall that

$$\begin{aligned}
h_2(0) - h_1(0) &= (1 + e^{\beta_0})^2 \{(c^* - \beta_0 + 2) - (c^* - \beta_0 - 2)e^{-\beta_0}\} \\
&= (1 + e^{\beta_0})^2 (c^* - \beta_0 + 2)(1 - e^{-c^* - \beta_0}) \\
&> 0.
\end{aligned}$$

As a result, $h_2(\alpha)$ only intersects $h_1(\alpha)$ once other than at $\alpha = 1$. We call this point α_m . Clearly, $\alpha_m < 1$.

To summarize,

- for $\alpha \in [0, \alpha_m)$, then $h_1(\alpha) - h_2(\alpha) \leq 0$, i.e. $g(\alpha)$ decreases on $[0, \alpha_m)$;
- for $\alpha \in [\alpha_m, 1)$, then $h_1(\alpha) - h_2(\alpha) \geq 0$, i.e. $g(\alpha)$ increases on $[\alpha_m, 1)$;
- for $\alpha \in [1, \infty)$, then $h_1(\alpha) - h_2(\alpha) \leq 0$, i.e. $g(\alpha)$ decreases on $[1, \infty)$.

Therefore, $g(\alpha)$ attains its maximum at $\alpha = 0$ and $\alpha = 1$ with maximum value $g(0) = g(1) = 1$. Hence $g(\alpha) \leq 1$ for all $\alpha \geq 0$. This concludes the proof.

□

For reader's convenience, Table 4.1 lists some values of c^* for selected values of β_0 . Notice that this relationship between c^* and β_0 does not depend on the number of covariates p .

We use a simple example to illustrate Theorem 4.2.3.

Table 4.1: Relationship Between c^* and β_0 .

β_0	-1.5	-1	-0.5	0	0.5	1	1.5
c^*	1.562	1.796	2.075	2.399	2.768	3.175	3.611

Example 4.2.7. Consider the following three-covariate logistic model:

$$u = \text{logit}(p) = -1 + 2x_1 + 3x_2 + x_3,$$

where $x_i \geq 0$, $i = 1, 2$ and 3 . Since $\beta_0 = -1 > -1.5434$, Theorem 4.2.3 applies. Then according to Table 4.1, $\beta_0 = -1$, $c^* = 1.796$ and $c^* - \beta_0 = 2.796$. Therefore, a D-optimal design in the (z_1, z_2, z_3) -space is

$$\xi_z^* = \begin{pmatrix} (0, 0, 0) & (2.796, 0, 0) & (0, 2.796, 0) & (0, 0, 2.796) \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.$$

With $z_1 = 2x_1$, $z_2 = 3x_2$ and $z_3 = x_3$, the D-optimal design in the original design space is

$$\xi^* = \begin{pmatrix} (0, 0, 0) & (1.398, 0, 0) & (0, 0.932, 0) & (0, 0, 2.796) \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}.$$

Figure 4.1 shows the D-optimal design pattern for this example and Figure 4.2 is an illustration of equivalence theorem check for the sensitivity function.

4.3 D-optimal Designs with $2p$ Support Points

In Section 4.2, we proved the $(p + 1)$ -point D-optimal design pattern under the assumption that β_0 is greater than a cutoff value (approximately -1.5434). So a natural question is, does this pattern still hold when β_0 is less than -1.5434? In

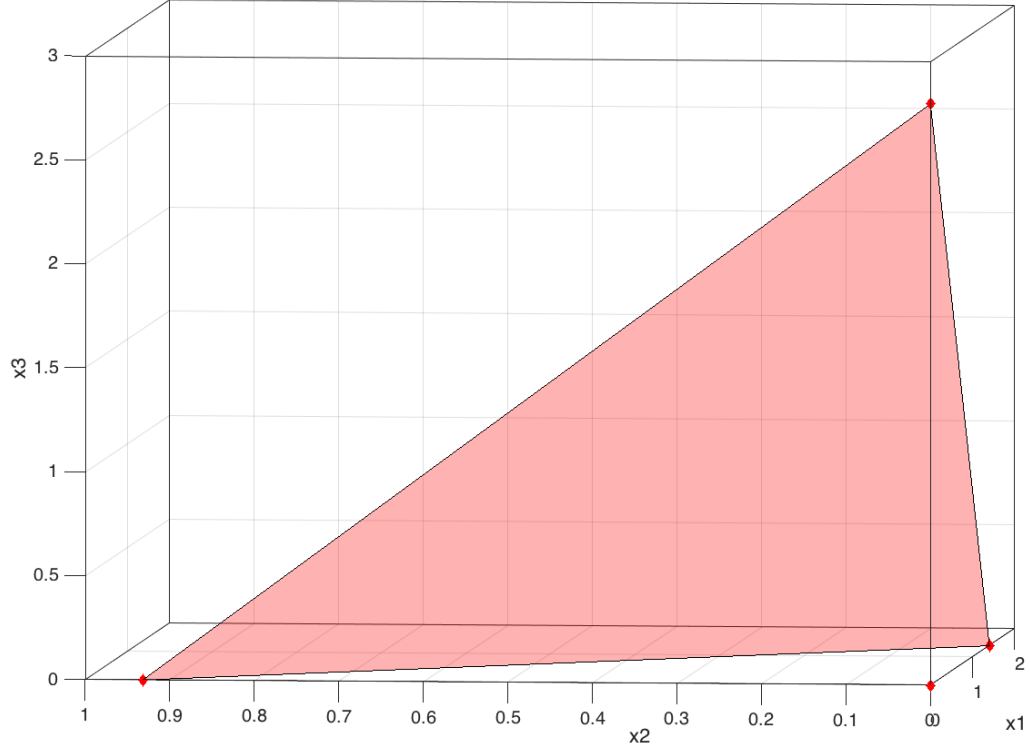


Figure 4.1: Pattern of a 4-point D-optimal Design for Three-covariate Logistic Model.

fact, we observed a different structure for D-optimal designs which takes two support points on each axis and two different weights. The global D-optimality of such designs is proved theoretically for a special range of β_0 .

Theorem 4.3.1. *Consider Model (4.2) with parameter $\beta_0 \in [-2, -\tilde{c})$, where $\tilde{c}$ is the unique positive solution of (4.5) with an approximate value of 1.5434. Assume that $0 < c^* \leq -\beta_0$ and $0 \leq w^* \leq 1$ satisfy the equations*

$$(1 - 2w^*)(c^* + \beta_0)^2 - 4w^*[p - (p + 1)w^*]\beta_0 c^* = 0 \quad (4.9)$$

and

$$(p+1)(1+c^*+e^{c^*}-c^*e^{c^*})[(c^*+\beta_0)^2-4w^*\beta_0c^*]-(p-1)(\beta_0^2-c^{*2})(1+e^{c^*})=0, \quad (4.10)$$

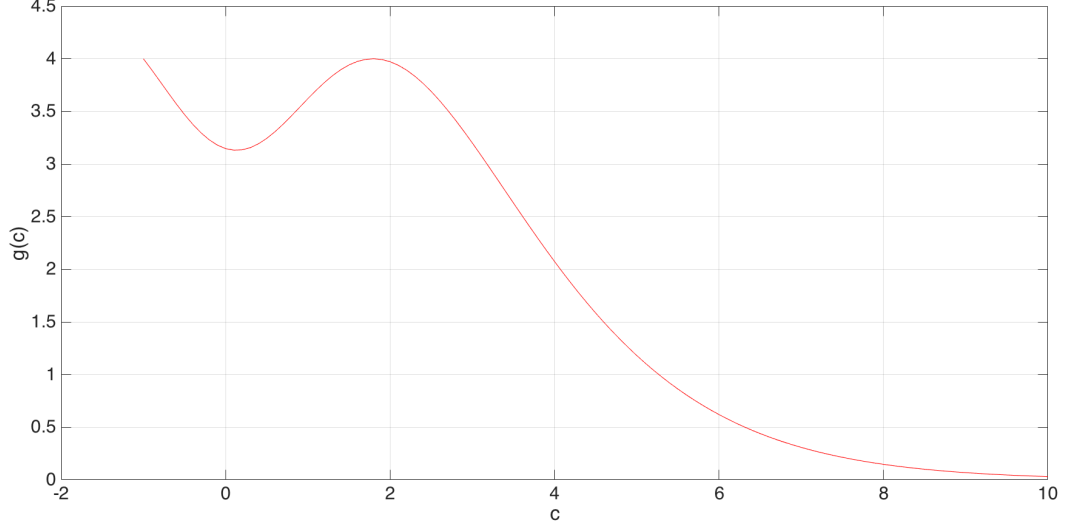


Figure 4.2: Sensitivity Plot for Design ξ_z^*

then the design $\xi^* =$

$$\left\{ \begin{array}{cccccc} (-c^* - \beta_0, 0, \dots, 0) \cdots (0, \dots, 0, -c^* - \beta_0) & (c^* - \beta_0, 0, \dots, 0) \cdots (0, \dots, 0, c^* - \beta_0) \\ \frac{1-w^*}{p} & \dots & \frac{1-w^*}{p} & \frac{w^*}{p} & \dots & \frac{w^*}{p} \end{array} \right\},$$

is D -optimal.

Proof. First, let I_{ξ^*} be the information matrix of the proposed design ξ^* , to simplify notations, let $R = -c^* - \beta_0 + 2w^*c^*$ and $Q = (c^* + \beta_0)^2 - 4w^*\beta_0c^*$, then we have

$$I_{\xi^*} = \frac{1}{p} \Psi(c^*) \begin{pmatrix} p & R & R & \cdots & R \\ R & Q & 0 & \cdots & 0 \\ R & 0 & Q & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ R & 0 & \cdots & 0 & Q \end{pmatrix},$$

and

$$I_{\xi^*}^{-1} = \frac{1}{\Psi(c^*)(Q - P^2)} \begin{pmatrix} Q & -R & -R & \cdots & -R \\ -R & \frac{pQ-(p-1)R^2}{Q} & \frac{R^2}{Q} & \cdots & \frac{R^2}{Q} \\ -R & \frac{R^2}{Q} & \frac{pQ-(p-1)R^2}{Q} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{R^2}{Q} \\ -R & \frac{R^2}{Q} & \cdots & \frac{R^2}{Q} & \frac{pQ-(p-1)R^2}{Q} \end{pmatrix}.$$

Let $\mathbf{z} = (z_1, z_2, \dots, z_p)^T = (z_1, z_2, \dots, z_{p-1}, c - \beta_0 - \sum_{i=1}^{p-1} z_i)^T$ be an arbitrary support point in the design space $[0, \infty)^p$, where $c = \text{logit}(P) = \beta_0 + z_1 + z_2 + \cdots + z_p$. Then according to the equivalence theorem presented in (1.18), after some calculations we have

$$\begin{aligned} d(\mathbf{z}, \xi^*) &= \Psi(c)(1, \mathbf{z}^T)I_{\xi^*}^{-1}(1, \mathbf{z}^T)^T \\ &= \frac{\Psi(c)}{\Psi(c^*)} \left\{ \frac{(c + c^*)^2 - 4w^*cc^*}{4w^*(1 - w^*)c^{*2}} + \right. \\ &\quad \left. \frac{(p-1)(c - \beta_0)^2 + 2p[z_1^2 - (c - \beta_0)z_1 + \cdots + z_{p-1}^2 - (c - \beta_0)z_{p-1} + \sum_{i \neq j} z_i z_j]}{(c^* + \beta_0)^2 - 4w^*\beta_0c^*} \right\} \\ &\leq \frac{\Psi(c)}{\Psi(c^*)} \left\{ \frac{(c + c^*)^2 - 4w^*cc^*}{4w^*(1 - w^*)c^{*2}} + \frac{(p-1)(c - \beta_0)^2}{(c^* + \beta_0)^2 - 4w^*\beta_0c^*} \right\} \triangleq g(c). \end{aligned} \quad (4.11)$$

The inequality follows again from Lemma 4.2.4. To simplify the right hand side of (4.11), we first look at Equation (4.9), which is a quadratic function in design weight w^* . There are at most two solutions:

$$w_1^* = \frac{\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 - \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4}}{4(p+1)\beta_0c^*} \quad (4.12)$$

and

$$w_2^* = \frac{\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 + \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4}}{4(p+1)\beta_0c^*}. \quad (4.13)$$

Consider w_2^* first. Since $\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4 = (\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2)^2 - 4(p+1)\beta_0c^*(c^* + \beta_0)^2 = (\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2)^2 + 4(p+1)|\beta_0|c^*(c^* + \beta_0)^2$,

it implies that

$$\begin{aligned}\sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4} &\geq \sqrt{(\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2)^2} \\ &= |\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2| \geq \beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2,\end{aligned}\tag{4.14}$$

which means the numerator of w_2^* , $\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 +$

$$\sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4} \geq \beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 + |\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2| \geq 0.$$

As a result, $w_2^* \leq 0$, which means that w_2^* is not the desired solution.

For w_1^* , following the same procedure as in (4.14), we have $\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 - \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4} \leq 0$, thus $w_1^* \geq 0$. Also, $[\beta_0^2 - 2(p+1)\beta_0c^* + (c^*)^2]^2 = \beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4 - 4(p+1)\beta_0c^*(c^* + \beta_0)^2 \geq \beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4$, which implies $\sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4} \leq |\beta_0^2 - 2(p+1)\beta_0c^* + (c^*)^2| = \beta_0^2 - 2(p+1)\beta_0c^* + (c^*)^2$. Meanwhile, $4(p+1)\beta_0c^* = \beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 - (\beta_0^2 - 2(p+1)\beta_0c^* + (c^*)^2) \leq \beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 - \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4} \leq 0$. It means that $w_1^* \leq 1$.

To summarize, (4.12) is the only valid solution for Equation (4.9), and it is between 0 and 1. Without causing any confusion, we shall from now on write w_1^* as w^* .

Now, from the expression of (4.12), we have

$$\begin{aligned}\frac{w^*}{(p+1)w^* - 1} &= \frac{\beta_0^2 + 2(p+1)\beta_0c^* + (c^*)^2 - \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4}}{(p+1)(\beta_0^2 + 2(p-1)\beta_0c^* + (c^*)^2 - \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4})} \\ &= \frac{p((c^*)^2 + \beta_0^2) + \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4}}{(p^2 - 1)(c^* - \beta_0)^2},\end{aligned}$$

i.e.,

$$\frac{w^*(c^* - \beta_0)^2}{(p+1)w^* - 1} = \frac{p((c^*)^2 + \beta_0^2) + \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4}}{(p^2 - 1)}.\tag{4.15}$$

Also from (4.12), we have

$$\frac{1}{p-1}[(c^* + \beta_0)^2 - 4w^*\beta_0c^*] = \frac{p((c^*)^2 + \beta_0^2) + \sqrt{\beta_0^4 + (4p^2 - 2)\beta_0^2(c^*)^2 + (c^*)^4}}{(p^2 - 1)}.\tag{4.16}$$

Comparing Equations (4.15) and (4.16) gives the following identity:

$$(c^* + \beta_0)^2 - 4w^*\beta_0c^* = \frac{(p-1)w^*(c^* - \beta_0)^2}{(p+1)w^* - 1}. \quad (4.17)$$

Now we can simplify $g(c)$ in (4.11) using (4.17) as

$$\begin{aligned} g(c) &= \frac{\Psi(c)}{\Psi(c^*)} \left\{ \frac{(c + c^*)^2 - 4w^*cc^*}{4w^*(1 - w^*)c^{*2}} + \frac{(p-1)(c - \beta_0)^2}{(c^* + \beta_0)^2 - 4w^*\beta_0c^*} \right\} \\ &= \frac{\Psi(c)}{\Psi(c^*)} \left\{ \frac{(c + c^*)^2 - 4w^*cc^*}{4w^*(1 - w^*)c^{*2}} + \frac{[(p+1)w^* - 1](c - \beta_0)^2}{w^*(c^* - \beta_0)^2} \right\} \text{ for } c \geq \beta_0 \end{aligned} \quad (4.18)$$

Therefore, based on Equivalence Theorem (1.18), all we have to show is that $g(c) \leq p+1$, for all $c \geq \beta_0$. It is easy to see that equality holds for support points in ξ^* . Next we will show the inequality holds for all $c \geq \beta_0$ with $\beta_0 \in [-2, -\tilde{c})$.

First notice that $g(c)$ is an even function:

$$\begin{aligned} g(c) - g(-c) &= \frac{c\Psi(c)[(1 - 2w^*)(c^* + \beta_0)^2 - 4w^*[p - (p+1)w^*]\beta_0c^*]}{\Psi(c^*)w^*(1 - w^*)c^*(c^* - \beta_0)^2} \\ &= 0 \quad (\text{from Equation (4.9)}) \end{aligned}$$

Therefore, we only need to show that $g(c) \leq p+1$ on $[0, \infty)$. It is easy to verify that $g(c^*) = p+1$, so we want to show that $g'(c) \geq 0$ on $[0, c^*]$ and $g'(c) \leq 0$ on (c^*, ∞) . Consider the first derivative of $g(c)$, after some simplifications, we have

$$g'(c) = \frac{e^{c-c^*}(1 + e^{c^*})^2 h(c)}{4w^*(1 - w^*)(c^*)^2(c^* - \beta_0)^2(1 + e^c)^3}, \quad (4.19)$$

where

$$\begin{aligned} h(c) &= (c^* - \beta_0)^2[(c^*)^2 + c^2 + 2c] + 4(c^*)^2[(p+1)w^* - 1](1 - w^*)(\beta_0^2 + c^2 + 2c) \\ &\quad - e^c\{(c^* - \beta_0)^2[(c^*)^2 + c^2 + 2c] + 4(c^*)^2[(p+1)w^* - 1](1 - w^*)(\beta_0^2 + c^2 - 2c)\} \end{aligned} \quad (4.20)$$

Therefore, it is sufficient to show that $h(c) \geq 0$ on $[0, c^*]$ and $h(c) \leq 0$ on (c^*, ∞) .

Notice that Equation (4.10) can be written as

$$e^{c^*} = \frac{(c^*)^2[(p+1)c^* + 2p] + \beta_0^2[(p+1)c^* + 2] - 2(p+1)\beta_0c^*(2w^* - 1)(c^* + 1)}{(c^*)^2[(p+1)c^* - 2p] + \beta_0^2[(p+1)c^* - 2] - 2(p+1)\beta_0c^*(2w^* - 1)(c^* - 1)}. \quad (4.21)$$

Replacing e^c in (4.20) with the right hand side of (4.21), we can obtain

$$\begin{aligned}
h(c^*) &= \\
& \frac{4[(p+3) - 2(p+1)w^*]c^{*2}(c^{*2} - \beta_0^2)[(1 - 2w^*)(c^* + \beta_0)^2 - 4w^*[p - (p+1)w^*]\beta_0c^*]}{\beta_0^2[2 - (p+1)c^*] + 2(p+1)c^*(2w^* - 1)(c^* - 1) + [2p - (p+1)c^*](c^*)^2} \\
&= 0 \quad (\text{from Equation (4.9)}) \tag{4.22}
\end{aligned}$$

In addition, it is easy to verify that $h(0) = 0$. With $h(0) = h(c^*) = 0$, we know that there is at least one stationary point on $(0, c^*)$. Also, since $h(c) \rightarrow -\infty$ as $c \rightarrow \infty$, then if there is only one stationary point for $h(c)$ on $[0, \infty)$, then that point must be a maximum. To obtain stationary points for $h(c)$, consider $s(c) = h'(c) = s_1(c) - s_2(c)$, where

$$s_1(c) = 2(c+1)\{(c^* - \beta_0)^2 + 4(c^*)^2[(p+1)w^* - 1](1 - w^*)\}$$

is a linear function with intercept and slope both being $2\{(c^* - \beta_0)^2 + 4(c^*)^2[(p+1)w^* - 1](1 - w^*)\}$ and

$$s_2(c) = e^c\{(c^* - \beta_0)^2(c^{*2} + c^2 - 2) + 4(c^*)^2[(p+1)w^* - 1](1 - w^*)(\beta_0^2 + c^2 - 2)\}.$$

Notice that

$$s_2''(c) = e^c\{(c^* - \beta_0)^2(c^{*2} + c^2 + 4c) + 4(c^*)^2[(p+1)w^* - 1](1 - w^*)(\beta_0^2 + c^2 + 4c)\}.$$

From Equation (4.9), we have

$$\frac{1 - 2w^*}{p - (p+1)w^*} = \frac{4w^*\beta_0c^*}{(c^* + \beta_0)^2} < 0,$$

then we can obtain the range for w^* , i.e. $\frac{1}{2} < w^* < \frac{p}{p+1}$, which results in $s_2''(c) > 0$ on $[0, \infty)$, therefore, $s_2(c)$ is a convex function. Meanwhile, evaluating $s_1(c) - s_2(c)$ at $c = 0$ gives

$$s_1(0) - s_2(0) = (c^* - \beta_0)^2(4 - c^{*2}) + 4(c^*)^2[(p+1)w^* - 1](1 - w^*)(4 - \beta_0^2).$$

From Equation (4.10), we have

$$1 + c^* + e^{c^*} - c^* e^{c^*} = \frac{(p-1)(\beta_0^2 - c^{*2})(1 + e^{c^*})}{(p+1)[(c^* + \beta_0)^2 - 4w^*\beta_0 c^*]}.$$

Under the assumption that $0 < c^* < -\beta_0$, we know $1 + c^* + e^{c^*} - c^* e^{c^*} > 0$. Recall Lemma 4.2.1, it gives the range for c^* as $0 < c^* < \tilde{c}$, where $\tilde{c}$ has an approximate value of 1.5434. Since we also assume that $\beta_0 \in [-2, -\tilde{c})$, therefore $s_1(0) - s_2(0) > 0$, which means that $s_1(c)$ is above $s_2(c)$ at $c = 0$. That implies $s_1(c)$ and $s_c(c)$ only intersect once on $[0, \infty)$. As a result, there is only one stationary point for $h(c)$ on $[0, \infty)$ and that point is indeed a maximum. Therefore, $g(c)$ in (4.18) has a unique maximum at $c = c^*$ with $g(c^*) = p + 1$. This completes the proof. $\square$

We use an example to illustrate Theorem 4.3.1.

Example 4.3.2. Consider the following three-covariate logistic model:

$$u = \text{logit}(p) = -2 + 2x_1 + 3x_2 + x_3,$$

where $x_i \geq 0$, $i = 1, 2$ and 3 . Since $\beta_0 = -2 \in [-2, -\tilde{c})$, Theorem 4.3.1 applies. Then solving Equation (4.9) and (4.10) gives $c^* = 1.4441$ and $w^* = 0.7456$. Then we have $-c^* - \beta_0 = 0.5559$ and $c^* - \beta_0 = 3.4441$. Therefore, a D-optimal design in (z_1, z_2, z_3) -space is shown in Table 4.2.

With $z_1 = 2x_1$, $z_2 = 3x_2$ and $z_3 = x_3$, the D-optimal design in the original design space is given in Table 4.3.

Figure 4.3 shows the D-optimal design pattern for this example and Figure 4.4 is an illustration of the equivalence theorem check for the sensitivity function.

4.4 Summaries and Discussions

In this chapter, we considered multiple-covariate logistic models under a different setup compared with the assumptions we make in Chapter 3. Here we consider main-effects models and all covariates are restricted to be nonnegative. Also, the coefficients

Table 4.2: D-optimal Design for Example 4.3.2 in Transformed Design Space.

Support Points	Weights
(0.5559, 0, 0)	0.0855
(0, 0.5559, 0)	0.0855
(0, 0, 0.5559)	0.0855
(3.4441, 0, 0)	0.2485
(0, 3.4441, 0)	0.2485
(0, 0, 3.4441)	0.2485

Table 4.3: D-optimal Design for Example 4.3.2 in the Original Design Space

Support Points	Weights
(0.2780, 0, 0)	0.0855
(0, 0.1853, 0)	0.0855
(0, 0, 0.5559)	0.0855
(1.7221, 0, 0)	0.2485
(0, 1.1480, 0)	0.2485
(0, 0, 3.4441)	0.2485

corresponding to the covariates are assumed to have the same sign and the value of intercept β_0 can vary from -2 to ∞ .

Two D-optimal design patterns are observed. One of them is a $(p + 1)$ -point saturated design with one point on each axis plus the origin, all design points have the same weight and for those on the axes, they all have the same logit value. We

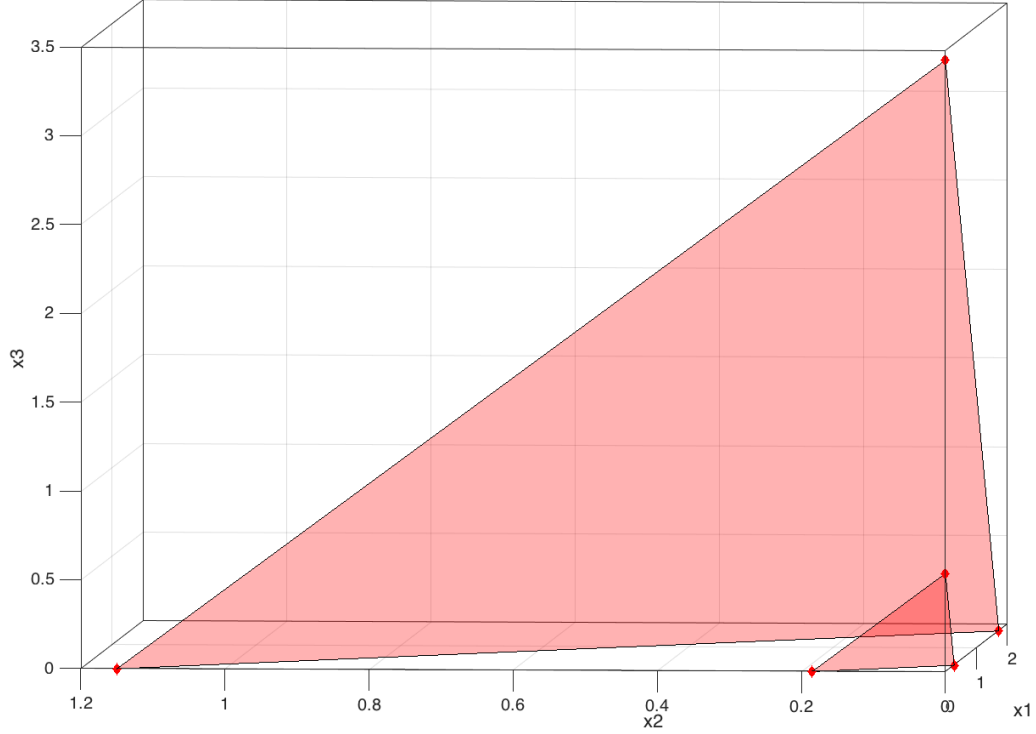


Figure 4.3: Pattern of a 6-point D-optimal Design for Three-covariate Logistic Model.

should note that the optimal logit value only depends on the intercept β_0 irrespective of the model dimension. The second pattern is a $2p$ -point design with two points on each axis. We can envision that there are two equal-logit hyperplanes intersecting the axes and intersections with the same hyperplane share the same weight. Unlike the first pattern, the optimal weight as well as the optimal logit value depends on both the value of the intercept and the number of covariates in the model. For the first design pattern, we analytically proved its global D-optimality using the equivalence theorem when β_0 is greater than a cutoff value of approximately -1.5434. For the second design pattern, we were able to prove a theorem for a limited range of β_0 values, i.e. $-2 \leq \beta_0 \leq -\tilde{c}$, where $-\tilde{c}$ is the same cutoff value.

There are however a number of open questions. The first question is for the second design pattern, we were only able to prove it for $-2 \leq \beta_0 \leq -\tilde{c}$. Computational

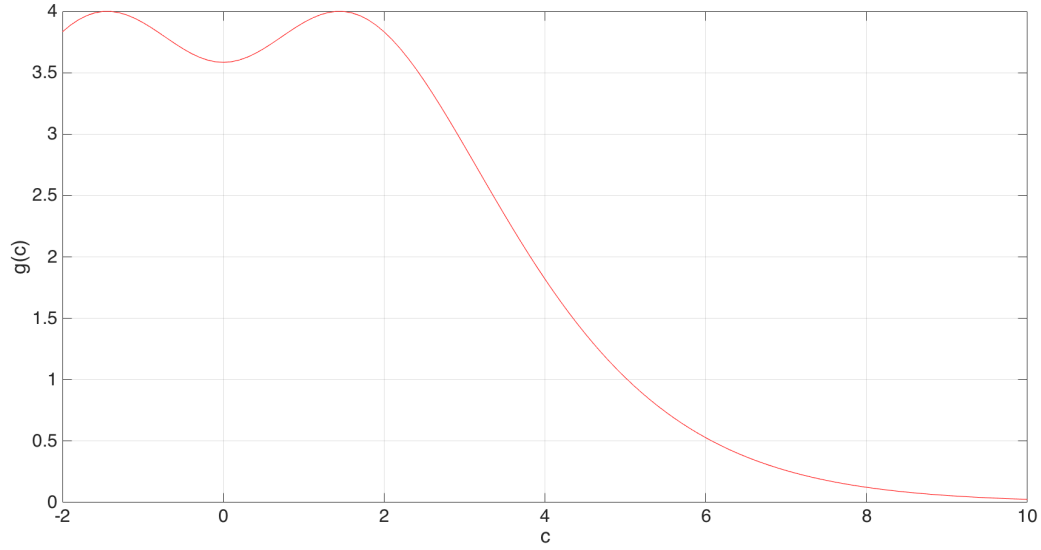


Figure 4.4: Sensitivity Plot for Design ξ_z^*

studies suggest that when β_0 is less than -2 , this optimal design pattern remains valid. How to generalize our proof to a larger range for β_0 remains unknown. Another question is, how do optimal designs look like if we add interactions to the model? We observed that with the existence of interactions, even the most basic two-covariate GLMs become much more complicated, and it appears infeasible to apply our previous methods to the interaction models. More research is needed in these directions.

CONCLUSION

In this dissertation, we investigate locally D-optimal designs for different types of generalized linear models.

In Chapter 2, GLMs with factorial effects and one covariate are under consideration. We provide an explicit expression of D-optimal designs for such models when interactions among factors also exist. In addition, smaller D-optimal designs are obtained by using strength $t+$ orthogonal arrays. Methods on how to obtain such OAs are also discussed. Two illustrative examples about electrostatic discharge (ESD) and Printed Circuit Board (PCB) experiments are provided in the end. By applying our proposed theorems, the designs we obtained are not only more efficient (D-optimal, in fact), but also require much fewer distinct runs compared to the design implemented in the original studies.

In Chapter 3, we consider multiple-covariate GLMs where all but one covariate are bounded in intervals and interactions among those bounded covariates may exist. We first provide a proof for the main-effects model and then extend the proof to derive an explicit formula of locally D-optimal designs for interaction models. Similar to the results in Chapter 2, smaller D-optimal designs can also be obtained using orthogonal arrays. Chapter 4 considers multiple-covariate logistic models where all covariates are nonnegative and there is no interaction among them. We observe two optimal design patterns, one with $p + 1$ support points and the other with $2p$, depending on the value of the intercept parameter β_0 . When the intercept is greater than a cutoff value of approximately -1.5434, we prove theoretically that a saturated design with one point on each axis plus the origin is D-optimal. When the intercept is less than

the cutoff value, we prove that on a limited range of β_0 , a D-optimal design takes two support points on each axis. We can envision that there are two equal-logit hyperplanes intersecting the axes and intersections with the same hyperplane share the same weight. Unlike the first design pattern, now the optimal weights and logit values depend not only on β_0 , but also on the dimension of the parameter space.

While many theoretical results are presented in this dissertation, there remain open questions for each of the three topics. As noted in Martin *et al.* (2016), "No design is completely immune to things going wrong, but we should be prepared for certain problems." There is still much work to be done in this area.

REFERENCES

- Agresti, A. and M. Kateri, *Categorical data analysis* (Springer, 2011).
- Atkinson, A. C. and L. M. Haines, “Designs for nonlinear and generalized linear models”, in “Handbook of Statistics (Vol. 13)”, edited by S. Ghosh and C. R. Rao, pp. 437–475 (Elsevier, 1996).
- Chen, R.-B., S.-P. Chang, W. Wang, H.-C. Tung and W. K. Wong, “Minimax optimal designs via particle swarm optimization methods”, *Statistics and Computing* **25**, 975–988 (2015).
- Ford, I., B. Torsney and C. J. Wu, “The use of a canonical form in the construction of locally optimal designs for non-linear problems”, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 569–583 (1992).
- Haines, L. M., G. Kabera and T. E. O’Brien, “D-optimal designs for logistic regression in two variables”, *mODa 8-Advances in Model-Oriented Design and Analysis* pp. 91–98 (2007).
- Hedayat, A., “Orthogonal arrays of strength $t+$ and their statistical applications”, Technical Report, Statistical Laboratory, University of Illinois, Chicago (1989).
- Hedayat, A., “New properties of orthogonal arrays and their statistical applications”, *Statistical Design and Analysis of Industrial Experiments* pp. 407–422 (1990).
- Hedayat, A., N. Sloane and J. Stufken, *Orthogonal Arrays: Theory and Applications* (Springer Science & Business Media, 1999).
- Hu, L. and J. Stufken, “Optimal designs for mixed effects models”, Submitted (2016a).
- Hu, L. and J. Stufken, “Optimal designs for multiple-covariate generalized linear models”, unpublished (2016b).
- Hu, L., M. Yang and J. Stufken, “Saturated locally optimal designs under differentiable optimality criteria”, *The Annals of Statistics* **43**, 1, 30–56 (2015).
- Jeng, S.-L., V. R. Joseph and C. J. Wu, “Modeling and analysis strategies for failure amplification method”, *Journal of Quality Technology* **40**, 2, 128–139 (2008).
- Jia, Y. and R. H. Myers, “D-optimal designs for two-variable logistic regression models”, in “Proceedings of the Annual Meeting of the American Statistical Association”, (2001).
- Kabera, G., “D-optimal designs for drug synergy”, unpublished Ph.D. dissertation, University of KwaZulu-Natal (2009).
- Kabera, M. G. and L. M. Haines, “A note on the construction of locally d-and ds-optimal designs for the binary logistic model with several explanatory variables”, *Statistics & Probability Letters* **82**, 5, 865–870 (2012).

- Kiefer, J., “General equivalence theory for optimum designs (approximate theory)”, The Annals of Statistics **2**, 849–879 (1974).
- Kiefer, J. and J. Wolfowitz, “The equivalence of two extremum problems”, Canadian Journal of Mathematics **12**, 363–365 (1960).
- Lukemire, J., A. Mandal and W. K. Wong, “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 (2018).
- Martin, R., J. Stufken and M. Yang, “A conversation with Samad Hedayat”, Statistical Science **31**, 4, 637–647 (2016).
- Maruthi, B. and V. Roshan Joseph, “Improving the yield of printed circuit boards using design of experiments”, Quality Engineering **12**, 2, 259–265 (1999).
- Nelder, J. A. and R. W. Wedderburn, “Generalized linear models”, Journal of the Royal Statistical Society Series A (General) **13**, 370–384 (1972).
- Qiu, J., R.-B. Chen, W. Wang and W. K. Wong, “Using animal instincts to design efficient biomedical studies via particle swarm optimization”, Swarm and evolutionary computation **18**, 1–10 (2014).
- Rao, C. R., “Hypercubes of strength d leading to confounded designs in factorial experiments”, Bull. Calcutta Math. Soc **38**, 67–78 (1946).
- Rao, C. R., “Factorial experiments derivable from combinatorial arrangements of arrays”, Supplement to the Journal of the Royal Statistical Society **9**, 128–139 (1947).
- Rao, C. R., “On a class of arrangements”, Proceedings of the Edinburgh Mathematical Society (Series 2) **8**, 119–125 (1949).
- Russell, K. G., D. C. Woods, S. Lewis and J. Eccleston, “D-optimal designs for poisson regression models”, Statistica Sinica **19**, 721–730 (2009).
- Silvey, S. D., *Optimal design* (Springer, 1980).
- Sitter, R. R. and B. Torsney, “Optimal designs for binary response experiments with two design variables”, Statistica Sinica **5**, 405–419 (1995a).
- Sitter, R. R. and B. Torsney, “Optimal designs for generalised linear models”, in “MODA4 - Advances in Model-Oriented Data Analysis”, edited by C. Kitsos and W. Miller, pp. 87–102 (Physica-Verlag, 1995b).
- Smith, K., “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**, 1–85 (1918).
- Stufken, J. and M. Yang, “On locally optimal designs for generalized linear models with group effects”, Statistica Sinica **22**, 1765–1786 (2012).

- Tan, X., “Optimal designs for generalized linear models”, unpublished Ph.D. dissertation, University of Georgia (2016).
- Torsney, B. and N. Gunduz, “On optimal designs for high dimensional binary regression models”, in “Optimum Design 2000”, edited by A. Atkinson, B. Bogacka and A. Zhiglavskii, pp. 275–285 (Springer, 2001).
- Whitman, C., T. M. Gilbert, A. M. Rahn and J. A. Antonell, “Determining factors affecting ESD failure voltage using DOE”, *Microelectronics Reliability* **46**, 1228–1237 (2006).
- Yang, J. and A. Mandal, “D-optimal factorial designs under generalized linear models”, *Communications in Statistics-Simulation and Computation* **44**, 9, 2264–2277 (2015).
- Yang, J., A. Mandal and D. Majumdar, “Optimal designs for two-level factorial experiments with binary response”, *Statistica Sinica* **22**, 885–907 (2012).
- Yang, M., S. Biedermann and E. Tang, “On optimal designs for nonlinear models: a general and efficient algorithm”, *Journal of the American Statistical Association* **108**, 1411–1420 (2013).
- Yang, M. and J. Stufken, “Support points of locally optimal designs for nonlinear models with two parameters”, *The Annals of Statistics* **37**, 518–541 (2009).
- Yang, M., B. Zhang and S. Huang, “Optimal designs for generalized linear models with multiple design variables”, *Statistica Sinica* **21**, 1415–1430 (2011).
- Yu, Y., “Monotonic convergence of a general algorithm for computing optimal designs”, *The Annals of Statistics* **21**, 1593–1606 (2010).
- Yu, Y., “D-optimal designs via a cocktail algorithm”, *Statistics and Computing* **21**, 475–481 (2011).

APPENDIX A

ADDITIONAL PROOFS FOR TOPIC I

Proof of Lemma 2.2.2. By Equation (2.5),

$$M_{\xi^*}(\boldsymbol{\theta}_1) = \frac{1}{2s} \Psi(c^*) \sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \sum_{j=1}^2 D^{i_1 \cdots i_L j} (D^{i_1 \cdots i_L j})^T.$$

Inside the summation, we have $D^{i_1 \cdots i_L j} (D^{i_1 \cdots i_L j})^T$

$$= \begin{pmatrix} 1 & (\mathbf{Z}_1^{i_1})^T & \cdots & (\mathbf{Z}_L^{i_L})^T & \cdots & (\mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}})^T & \cdots & c_{i_1 \cdots i_L j} \\ \mathbf{Z}_1^{i_1} (\mathbf{Z}_1^{i_1})^T & \cdots & \mathbf{Z}_1^{i_1} (\mathbf{Z}_L^{i_L})^T & \cdots & \mathbf{Z}_1^{i_1} (\mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}})^T & \cdots & c_{i_1 \cdots i_L j} \mathbf{Z}_1^{i_1} & \\ & \ddots & \vdots & & \vdots & & \vdots & \\ & & \mathbf{Z}_L^{i_L} (\mathbf{Z}_L^{i_L})^T & \cdots & \mathbf{Z}_L^{i_L} (\mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}})^T & \cdots & c_{i_1 \cdots i_L j} \mathbf{Z}_L^{i_L} & \\ & & & \ddots & \vdots & & \vdots & \\ & & & & \mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}} (\mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}})^T & \cdots & c_{i_1 \cdots i_L j} \mathbf{Z}_{l_1 l_2}^{i_{l_1} i_{l_2}} & \\ & & & & & \ddots & \vdots & \\ & & & & & & \text{sym.} & c_{i_1 \cdots i_L j}^2 \end{pmatrix}$$

Clearly, the top-left element of $M_{\xi^*}(\boldsymbol{\theta}_1)$ is

$$\frac{1}{2s} \Psi(c^*) \sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \sum_{j=1}^2 1 = \Psi(c^*)$$

and the bottom-right element of $M_{\xi^*}(\boldsymbol{\theta}_1)$ is

$$\frac{1}{2s} \Psi(c^*) \sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \sum_{j=1}^2 (c^*)^2 = (c^*)^2 \Psi(c^*).$$

For the last column of $M_{\xi^*}(\boldsymbol{\theta}_1)$, the summation of each element is 0 due to the fact that we are using two symmetric points in each group.

For any other off-diagonal block, say its rows correspond to the effect $\mathbf{Z}_{r_1 r_2 \cdots r_m}^{i_{r_1} i_{r_2} \cdots i_{r_m}}$ and its columns correspond to the effect $\mathbf{Z}_{c_1 c_2 \cdots c_n}^{i_{c_1} i_{c_2} \cdots i_{c_n}}$, then the block becomes

$$\mathbf{Z}_{r_1 r_2 \cdots r_m}^{i_{r_1} i_{r_2} \cdots i_{r_m}} \cdot (\mathbf{Z}_{c_1 c_2 \cdots c_n}^{i_{c_1} i_{c_2} \cdots i_{c_n}})^T = (\mathbf{Z}_{r_1}^{i_{r_1}} \otimes \cdots \otimes \mathbf{Z}_{r_m}^{i_{r_m}}) \cdot ((\mathbf{Z}_{c_1}^{i_{c_1}})^T \otimes \cdots \otimes (\mathbf{Z}_{c_n}^{i_{c_n}})^T).$$

Since it is off-diagonal, there must exist a group index which only appears once.

Without loss of generality, say it is r_1 in $(r_1, \dots, r_m)$. Then taking the sum of

$\mathbf{Z}_{r_1 r_2 \dots r_m}^{i_{r_1} i_{r_2} \dots i_{r_m}} \cdot (\mathbf{Z}_{c_1 c_2 \dots c_n}^{i_{c_1} i_{c_2} \dots i_{c_n}})^T$ we have

$$\begin{aligned} \sum \dots \sum_{\text{not } r_1} \sum_{i_{r_1}=1}^{s_{r_1}} \mathbf{Z}_{r_1 r_2 \dots r_m}^{i_{r_1} i_{r_2} \dots i_{r_m}} \cdot (\mathbf{Z}_{c_1 c_2 \dots c_n}^{i_{c_1} i_{c_2} \dots i_{c_n}})^T \\ = \sum \dots \sum_{\text{not } r_1} \left(\left(\sum_{i_{r_1}=1}^{s_{r_1}} \mathbf{Z}_{r_1}^{i_{r_1}} \right) \otimes \mathbf{Z}_{r_2}^{i_{r_2}} \otimes \dots \otimes \mathbf{Z}_{r_m}^{i_{r_m}} \right) \cdot (\mathbf{Z}_{c_1 c_2 \dots c_n}^{i_{c_1} i_{c_2} \dots i_{c_n}})^T. \end{aligned}$$

By definition, we have $\sum_{i_{r_1}=1}^{s_{r_1}} \mathbf{Z}_{r_1}^{i_{r_1}} = \mathbf{0}$, thus the above equation is 0.

For the diagonal blocks, we first investigate the blocks corresponding to a main effect, say group l with s_l levels. Then

$$\begin{aligned} \frac{1}{2s} \Psi(c^*) \sum_{i_1=1}^{s_1} \dots \sum_{i_L=1}^{s_L} \sum_{j=1}^2 \mathbf{Z}_l^{i_l} (\mathbf{Z}_l^{i_l})^T \\ = \frac{\Psi(c^*)}{s_l} \cdot \left[\begin{pmatrix} 1 \\ -\frac{1}{s_l-1} \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \left(1, -\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1} \right) + \right. \\ \left. \dots + \begin{pmatrix} -\frac{1}{s_l-1} \\ -\frac{1}{s_l-1} \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \left(-\frac{1}{s_l-1}, -\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1} \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\Psi(c^*)}{s_l} \cdot s_l \cdot \begin{pmatrix} \frac{1}{(s_l-1)} & -\frac{1}{(s_l-1)^2} & \cdots & \cdots & -\frac{1}{(s_l-1)^2} \\ & \frac{1}{(s_l-1)} & -\frac{1}{(s_l-1)^2} & \cdots & -\frac{1}{(s_l-1)^2} \\ & & \ddots & \ddots & \vdots \\ & \text{sym.} & & \ddots & -\frac{1}{(s_l-1)^2} \\ & & & & \frac{1}{(s_l-1)} \end{pmatrix} \\
&= \Psi(c^*) \cdot B_l
\end{aligned}$$

For diagonal blocks which correspond to interactions, assume the interaction is among group $(l_1, \dots, l_k)$. Then

$$\begin{aligned}
&\frac{1}{2s} \Psi(c^*) \sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \sum_{j=1}^2 \mathbf{Z}_{l_1 \dots l_k}^{i_{l_1} \dots i_{l_k}} (\mathbf{Z}_{l_1 \dots l_k}^{i_{l_1} \dots i_{l_k}})^T \\
&= \frac{1}{s} \Psi(c^*) \sum_{\text{not } (l_1, \dots, l_k)} \cdots \sum_{(l_1, \dots, l_k)} \cdots \sum \mathbf{Z}_{l_1 \dots l_k}^{i_{l_1} \dots i_{l_k}} (\mathbf{Z}_{l_1 \dots l_k}^{i_{l_1} \dots i_{l_k}})^T \\
&= \frac{1}{s} \Psi(c^*) \sum_{\text{not } (l_1, \dots, l_k)} \cdots \sum_{(l_1, \dots, l_k)} \cdots \sum (\mathbf{Z}_{l_1}^{i_{l_1}} \otimes \cdots \otimes \mathbf{Z}_{l_k}^{i_{l_k}}) \cdot ((\mathbf{Z}_{l_1}^{i_{l_1}})^T \otimes \cdots \otimes (\mathbf{Z}_{l_k}^{i_{l_k}})^T) \\
&= \frac{1}{s} \Psi(c^*) \sum_{\text{not } (l_1, \dots, l_k)} \cdots \sum_{i_{l_1}=1}^{s_{l_1}} (\mathbf{Z}_{l_1}^{i_{l_1}} (\mathbf{Z}_{l_1}^{i_{l_1}})^T) \otimes \cdots \otimes (\sum_{i_{l_k}=1}^{s_{l_k}} \mathbf{Z}_{l_k}^{i_{l_k}} (\mathbf{Z}_{l_k}^{i_{l_k}})^T) \\
&= \frac{1}{s} \Psi(c^*) \sum_{\text{not } (l_1, \dots, l_k)} \cdots \sum (s_{l_1} \cdot B_{l_1}) \otimes \cdots \otimes (s_{l_k} \cdot B_{l_k}) \\
&= \frac{1}{s} \Psi(c^*) \cdot s \cdot B_{l_1} \otimes \cdots \otimes B_{l_k} = \Psi(c^*) \cdot B_{l_1} \otimes \cdots \otimes B_{l_k}
\end{aligned}$$

This concludes the proof. □

Proof of Lemma 2.2.3. When $\mathbf{Z}_l^i$ is in the form of $(-\frac{1}{s_l-1}, \dots, 1, \dots, -\frac{1}{s_l-1})^T$, then

$$\begin{aligned}
(\mathbf{Z}_l^{i_l})^T \mathbf{Z}_l^{i_l} + (\mathbf{Z}_l^{i_l})^T J \mathbf{Z}_l^{i_l} &= \left(-\frac{1}{s_l-1}, \dots, 1, \dots, -\frac{1}{s_l-1}\right) \begin{pmatrix} -\frac{1}{s_l-1} \\ \vdots \\ 1 \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \\
&+ \left(-\frac{1}{s_l-1}, \dots, 1, \dots, -\frac{1}{s_l-1}\right) \cdot J \cdot \begin{pmatrix} -\frac{1}{s_l-1} \\ \vdots \\ 1 \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \\
&= \left(\frac{1}{(s_l-1)^2} \cdot (s_l-2) + 1\right) + \frac{1}{s_l-1} \cdot 1' \begin{pmatrix} -\frac{1}{s_l-1} \\ \vdots \\ 1 \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \\
&= \frac{s_l-2}{(s_l-1)^2} + 1 + \frac{1}{(s_l-1)^2} = \frac{s_l}{s_l-1}
\end{aligned}$$

When $\mathbf{Z}_l^{i_l}$ is in the form of $(-\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1})^T$, then

$$(\mathbf{Z}_l^{i_l})^T \mathbf{Z}_l^{i_l} + (\mathbf{Z}_l^{i_l})^T J \mathbf{Z}_l^{i_l} = \left(-\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1}\right) \begin{pmatrix} -\frac{1}{s_l-1} \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix}$$

$$\begin{aligned}
& + \left(-\frac{1}{s_l-1}, \dots, -\frac{1}{s_l-1}\right) \cdot J \cdot \begin{pmatrix} -\frac{1}{s_l-1} \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \\
& = \frac{1}{(s_l-1)^2} \cdot (s_l-1) + (-1) \cdot 1' \begin{pmatrix} -\frac{1}{s_l-1} \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \\
& = \frac{1}{s_l-1} + \frac{1}{s_l-1} \cdot (s_l-1) = \frac{s_l}{s_l-1}
\end{aligned}$$

Together, it concludes the lemma. □

Proof of Lemma 2.2.4. First, notice that the functional inequality in Lemma 2.2.4 does not depend on models, but here we provide a proof for it using model (2.2) with main-effects only.

Analogously to the expression in (2.4), we rewrite model (2.2) with main-effects only in the following way:

$$Prob(Y_{i_1 i_2 \dots i_L j} = 1) = P(\gamma_0 + \sum_{l=1}^L \sum_{i_l=1}^{s_l-1} \gamma_l^{i_l} z_l^{i_l} + \beta x_{i_1 \dots i_L j}), \quad (\text{A.1})$$

where for each factor l , $z_l^{i_l}$ is defined in the same way as shown in (2.4). Then, without interaction terms, the $\boldsymbol{\theta}_1$ in (2.4) reduces to $\boldsymbol{\theta}_1 = (\gamma_0, \boldsymbol{\gamma}_1^T, \dots, \boldsymbol{\gamma}_L^T, \beta)^T$ where for any $1 \leq l \leq L$, $\boldsymbol{\gamma}_l = (\gamma_l^1, \dots, \gamma_l^{(s_l-1)})^T$. Similarly, $\mathbf{Z}^{i_1 \dots i_L j}$ can be simplified as $\mathbf{Z}^{i_1 \dots i_L j} = (1, (\mathbf{Z}_1^{i_1})^T, \dots, (\mathbf{Z}_L^{i_L})^T, x_{i_1 i_2 \dots i_L j})^T$ where for each factor l , $\mathbf{Z}_l^{i_l}$ is defined in the same way as shown in (2.4).

According to Theorem 4 in Stufken and Yang (2012), a locally D-optimal design for a maximal set of estimable functions of the parameter vector of model (2.2) with main-effects only can be written as $\xi^* = \{(c_{i_1 \dots i_L 1} = c^*, w_{i_1 \dots i_L 1} = \frac{1}{2s}), (c_{i_1 \dots i_L 2} = -c^*, w_{i_1 \dots i_L 2} = \frac{1}{2s}), i_1 = 1, \dots, s_1, \dots, i_L = 1, \dots, s_L\}$, where $c^* > 0$ maximizes $f(c) =$

$c^2(\Psi(c))^r$ on $(-\infty, \infty)$ and $r = \sum_1^L (s_l - 1) + 2$. Then, following the similar steps from (2.4) to (2.6), the following inequality

$$\Psi(c)(\mathbf{D}^{i_1 \cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta}_1) \mathbf{D}^{i_1 \cdots i_L} \leq r, \quad (\text{A.2})$$

holds for the main-effects model, where terms in (A.2) can be obtained analogously to the expressions in (2.4) $\sim$ (2.6) but with main-effects only.

The left-hand-side of (A.2) can be further simplified using Lemmas 2.2.2 and 2.2.3.

As a result,

$$\begin{aligned} \Psi(c)(\mathbf{D}^{i_1 \cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta}_1) \mathbf{D}^{i_1 \cdots i_L} &= \frac{\Psi(c)}{\Psi(c^*)} \left\{ 1 + \sum_{l=1}^L (s_l - 1) + \frac{c^2}{(c^*)^2} \right\} \\ &= \frac{\Psi(c)}{\Psi(c^*)} (r - 1) + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)}. \end{aligned} \quad (\text{A.3})$$

Comparing (A.2) and (A.3), we have

$$\frac{\Psi(c)}{\Psi(c^*)} (r - 1) + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)} \leq r,$$

and this concludes the proof. □

Proof of reparametrization. To show that model (2.4) is a reparametrization of model (2.1), all we need is that the column spaces spanned by the two design matrices are the same. Since they have the same slope parameter β , define $X = [\mathbf{x}_0, \mathbf{x}_1^1, \dots, \mathbf{x}_1^{s_1-1}, \dots, \mathbf{x}_L^{s_L-1}, \dots, \mathbf{x}_{l_1 l_2}^{i_{l_1} i_{l_2}}, \dots]$, where the interaction terms correspond to those in the model, as the reduced design matrix for model (2.1) and similarly, define $Z = [\mathbf{z}_0, \mathbf{z}_1^1, \dots, \mathbf{z}_1^{s_1-1}, \dots, \mathbf{z}_L^{s_L-1}, \dots, \mathbf{z}_{l_1 l_2}^{i_{l_1} i_{l_2}}, \dots]$ as the reduced design matrix for model (2.4), so we want to show $\mathcal{C}(X) = \mathcal{C}(Z)$.

For any $\mathbf{v} \in \mathcal{C}(Z)$, there exist $a_0, a_1^1, \dots, a_1^{s_1-1}, \dots$ such that

$$\mathbf{v} = a_0 \mathbf{z}_0 + a_1^1 \mathbf{z}_1^1 + \dots + a_L^{s_L-1} \mathbf{z}_L^{s_L-1} + \sum_{t=2}^L \sum_{(l_1, \dots, l_t) \in G_t} \left[\sum_{i_{l_1}=1}^{s_{l_1}-1} \dots \sum_{i_{l_t}=1}^{s_{l_t}-1} a_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}} \mathbf{z}_{l_1 l_2 \dots l_t}^{i_{l_1} i_{l_2} \dots i_{l_t}} \right].$$

We first notice that $\mathbf{z}_0 = \mathbf{x}_0$ and it is easy to verify that for each factor l ,

$$\mathbf{z}_l^{i_l} = \mathbf{x}_l^{i_l} - \frac{1}{s_l - 1} \sum_{t \neq i_l} \mathbf{x}_l^t, \quad (\text{A.4})$$

which means that each $\mathbf{z}_l^{i_l}$ can be written as a linear combination of the columns in $\mathbf{X}$.

For any interaction term, for instance, $\mathbf{z}_{12}^{11} = \mathbf{z}_1^1 \circ \mathbf{z}_2^1$, where “ $\circ$ ” represents the Hadamard (pairwise) product, from (A.4) we have

$$\begin{aligned} \mathbf{z}_{12}^{11} &= (\mathbf{x}_1^1 - \frac{1}{s_1 - 1} \sum_{t_1=2}^{s_1} \mathbf{x}_1^{t_1}) \circ (\mathbf{x}_2^1 - \frac{1}{s_2 - 1} \sum_{t_2=2}^{s_2} \mathbf{x}_2^{t_2}) \\ &= \mathbf{x}_1^1 \circ \mathbf{x}_2^1 - \frac{1}{s_1 - 1} \sum_{t_1=2}^{s_1} \mathbf{x}_1^{t_1} \circ \mathbf{x}_2^1 - \frac{1}{s_2 - 1} \sum_{t_2=2}^{s_2} \mathbf{x}_1^1 \circ \mathbf{x}_2^{t_2} + \\ &\quad \frac{1}{(s_1 - 1)(s_2 - 1)} \sum_{t_1=2}^{s_1} \sum_{t_2=2}^{s_2} \mathbf{x}_1^{t_1} \circ \mathbf{x}_2^{t_2}. \end{aligned}$$

Notice for each Hadamard product $\mathbf{x}_{l_1}^{i_{l_1}} \circ \mathbf{x}_{l_2}^{i_{l_2}}$, we have

$$\mathbf{x}_{l_1}^{i_{l_1}} \circ \mathbf{x}_{l_2}^{i_{l_2}} = \mathbf{x}_{l_1 l_2}^{i_{l_1} i_{l_2}}. \quad (\text{A.5})$$

Therefore, each interaction term in $\mathbf{Z}$ can also be written as a linear combination of the columns in $\mathbf{X}$. To summarize, $\mathbf{v} \in \mathcal{C}(\mathbf{X})$.

For any $\mathbf{v} \in \mathcal{C}(\mathbf{X})$, there exist $b_0, b_1^1, \dots, b_1^{s_1}, \dots$ such that

$$\mathbf{v} = b_0 \mathbf{x}_0 + b_1^1 \mathbf{x}_1^1 + \dots + b_L^{s_L} \mathbf{x}_L^{s_L} + \sum_{t=2}^L \sum_{(l_1, \dots, l_t) \in G_t} \left[\sum_{i_{l_1}=1}^{s_{l_1}} \dots \sum_{i_{l_t}=1}^{s_{l_t}} b_{l_1 \dots l_t}^{i_{l_1} \dots i_{l_t}} \mathbf{x}_{l_1 l_2 \dots l_t}^{i_{l_1} i_{l_2} \dots i_{l_t}} \right].$$

It is easy to verify that for each factor l ,

$$\mathbf{x}_l^{i_l} = \begin{cases} \frac{\mathbf{z}_0 + (s_l - 1)\mathbf{z}_l^{i_l}}{s_l}, & \text{when } 1 \leq i_l \leq s_l - 1 \\ \frac{\mathbf{z}_0 - (s_l - 1)\sum_{t=1}^{s_l-1} \mathbf{z}_l^t}{s_l}, & \text{when } i_l = s_l \end{cases}$$

which means that each $\mathbf{x}_l^{i_l}$ can be written as a linear combination of the columns in Z .

For any interaction term, for instance, $\mathbf{x}_{12}^{11} = \mathbf{x}_1^1 \circ \mathbf{x}_2^1$, following the same procedure as before, we are able to show that each interaction term in X can also be written as a linear combination of the columns in Z . To summarize, $\mathbf{v} \in \mathcal{C}(Z)$.

Therefore, $\mathcal{C}(X) = \mathcal{C}(Z)$.

□

APPENDIX B

ADDITIONAL PROOFS FOR TOPIC III

Proof of Lemma 4.2.2. Let $f(c) = 1 + c + e^c - ce^c$, then $f(0) = 2$ and $\lim_{c \rightarrow \infty} f(c) = -\infty$. Thus $f(c) = 0$ has at least one solution on $[0, \infty)$. Also, since $f''(c) = -e^c(c + 1) < 0$, indicating that $f(c)$ is concave on $[0, \infty)$, so $f(c)$ intersects the horizontal axis only once. Therefore, Equation (4.5) has a unique solution for $c \geq 0$. The numerical solution is approximately 1.5434. $\square$

Proof of Lemma 4.2.3. When $\beta_0 \in [-\tilde{c}, 0]$, consider $g(c) = 2 - \beta_0 + c + 2e^c + \beta_0 e^c - ce^c$ on $[0, \infty)$, we have $g(0) = 4 > 0$ and $\lim_{c \rightarrow \infty} g(c) = -\infty$, thus $g(c)$ has at least one solution on $[0, \infty)$. Also, $g''(c) = (\beta_0 - c)e^c < 0$, suggesting $g(c)$ is concave on $[0, \infty)$, therefore $g(c)$ intersects the horizontal axis only once, or equivalently, Equation (4.6) has a unique solution on $[0, \infty)$, denoted by c^* .

To show that $c^* \geq -\beta_0$, it is sufficient to show that

$$g(-\beta_0) = 2 - 2\beta_0 + 2(1 + \beta_0)e^{-\beta_0} \geq 0, \text{ for } \forall \beta_0 \in [-\tilde{c}, 0]$$

For convenience, define $h(\beta_0) = g(\beta_0)$, then $h(0) = 4$, $h(-\tilde{c}) = 0$ and $h''(\beta_0) = -2(1 - \beta_0)e^{-\beta_0} < 0$ on $[-\tilde{c}, 0]$. Therefore, $h(\beta_0)$ is strictly concave on $[-\tilde{c}, 0]$, which implies that $h(\beta_0) \geq 0$ on $[-\tilde{c}, 0]$, hence $c^* \geq -\beta_0$.

When $\beta_0 \in (0, \infty)$, consider again $g(c)$, we still have $g(0) = 4 > 0$ and $\lim_{c \rightarrow \infty} g(c) = -\infty$, thus $g(c)$ has at least one solution on $[0, \infty)$. Now, $g''(c) = (\beta_0 - c)e^c$, suggesting that $g(c)$ is convex on $[0, \beta_0]$ and concave on $[\beta_0, \infty)$. Meanwhile,

$$g'(c) = 1 + 2e^c + \beta_0 e^c - e^c - ce^c = 1 + (1 + \beta_0 - c)e^c \geq 0, \text{ when } c \leq \beta_0.$$

So $g(c)$ is increasing on $[0, \beta_0]$, and therefore there's only one solution c^* and $c^* > \beta_0$. $\square$

Proof of Lemma 4.2.4. We use induction. First, when $n = 2$,

$$f(x_1, x_2) = x_1^2 - ax_1 + x_2^2 - ax_2 + x_1x_2.$$

For fixed x_2 with $0 \leq x_2 \leq a$, we have $0 \leq x_1 \leq a - x_2$. Setting $\partial f(x_1, x_2)/\partial x_1 = 2x_1 - a + x_2 = 0$ gives $x_1 = (a - x_2)/2$, i.e. $f(x_1, x_2)$ reaches its minimum at $x_1 = (a - x_2)/2$. Also, the maximum of $f(x_1, x_2)$ is attained at $x_1 = 0$ and $x_1 = a - x_2$, with the maximum value being $x_2^2 - ax_2 = (x_2 - \frac{a}{2})^2 - \frac{a^2}{4} \leq 0$.

Therefore, for any $0 \leq x_2 \leq a$, $f(x_1, x_2) \leq 0$. Hence, $f(x_1, x_2) \leq 0$, for all $x_1, x_2 \geq 0$ with $x_1 + x_2 \leq a$.

Next, assume $f(x_1, \dots, x_n) \leq 0$ holds for $n = k$, then when $n = k + 1$,

$$f(x_1, \dots, x_{k+1}) = x_1^2 - ax_1 + \dots + x_{k+1}^2 - ax_{k+1} + \sum_{i \neq j} x_i x_j.$$

Fix $x_2, \dots, x_{k+1}$, we have $0 \leq x_1 \leq a - \sum_{i=2}^{k+1} x_i$. Again, setting $\partial f(x_1, \dots, x_{k+1})/\partial x_1 = 2x_1 - a + \sum_{i=2}^{k+1} x_i = 0$ gives $x_1 = (a - \sum_{i=2}^{k+1} x_i)/2$, i.e. $f(x_1, \dots, x_{k+1})$ reaches its minimum at $x_1 = (a - \sum_{i=2}^{k+1} x_i)/2$. Also, the maximum of $f(x_1, \dots, x_{k+1})$ is attained at $x_1 = 0$ and $x_1 = a - \sum_{i=2}^{k+1} x_i$, with the maximum value being

$$f(x_2, \dots, x_{k+1}) = x_2^2 - ax_2 + \dots + x_{k+1}^2 - ax_{k+1} + \sum_{i \neq j} x_i x_j \leq 0.$$

The last equality results from the induction assumption.

Therefore, $f(x_1, \dots, x_n) \leq 0$ also holds for $n = k + 1$. This concludes the proof. $\square$

Proof of Lemma 4.2.5. To show that $h_1(\alpha)$ and $h_2(\alpha)$ are crossing, instead of touching each other at $\alpha = 1$, it is sufficient to show that $h_2(1) - h_1(1) \neq 0$. In fact, we will show that $h'_2(1) - h'_1(1) > 0$.

The derivatives of the two functions are

$$h'_1(\alpha) = e^{-c^*} (1 + e^{c^*})^2 [2\alpha(c^* - \beta_0) + 2] + e^{-\beta_0} (1 + e^{\beta_0})^2 [2(\alpha - 1)(c^* - \beta_0) + 2]$$

and

$$h'_2(\alpha) = e^c \left\{ e^{-c^*} (1 + e^{c^*})^2 [2\alpha(c^* - \beta_0) - 2] + e^{-\beta_0} (1 + e^{\beta_0})^2 [2(\alpha - 1)(c^* - \beta_0) - 2] \right\}$$

$$+ e^c(c^* - \beta_0) \left\{ e^{-c^*}(1 + e^{c^*})^2 [\alpha^2(c^* - \beta_0) - 2\alpha] + e^{-\beta_0}(1 + e^{\beta_0})^2 \right. \\ \left. [(\alpha - 1)^2(c^* - \beta_0) - 2(\alpha - 1)] \right\}.$$

When evaluated at $\alpha = 1$,

$$h'_1(1) = e^{-c^*}(1 + e^{c^*})^2[2(c^* - \beta_0) + 2] + 2e^{-\beta_0}(1 + e^{\beta_0})^2$$

and

$$h'_2(1) = e^{c^*} \left\{ e^{-c^*}(1 + e^{c^*})^2[2(c^* - \beta_0) - 2] - 2e^{-\beta_0}(1 + e^{\beta_0})^2 \right\} + \\ e^{c^*}(c^* - \beta_0) \left\{ e^{-c^*}(1 + e^{c^*})^2[c^* - \beta_0 - 2] \right\}.$$

Therefore,

$$h'_2(1) - h'_1(1) = 2e^{-c^*}(1 + e^{c^*})^2[(e^{c^*} - 1)(c^* - \beta_0) - e^{c^*} - 1] - 2e^{-\beta_0}(1 + e^{\beta_0})^2(e^{c^*} + 1) \\ + e^{c^*}(c^* - \beta_0)[e^{-c^*}(1 + e^{c^*})^2(c^* - \beta_0 - 2)].$$

First of all, from Equation (4.6), $c^* - \beta_0 - 2 = (c^* - \beta_0 + 2)e^{-c^*} > 0$, so the last term is positive. Also from Equation (4.6), inside the brackets of the first term, $(e^{c^*} - 1)(c^* - \beta_0) - e^{c^*} - 1 = e^{c^*} + 1$. Thus the summation of the first two terms becomes

$$2(e^{c^*} + 1)[e^{-c^*}(1 + e^{c^*})^2 - e^{-\beta_0}(1 + e^{\beta_0})^2].$$

Note the function $r(x) = e^{-x}(1 + e^x)^2$ is symmetric around $x = 0$ and it is decreasing on $(-\infty, 0]$ and increasing on $[0, \infty)$. So when $\beta \in (-\tilde{c}, 0]$, we have

$$e^{-c^*}(1 + e^{c^*})^2 - e^{-\beta_0}(1 + e^{\beta_0})^2 = r(c^*) - r(\beta_0) = r(c^*) - r(-\beta_0) \geq 0;$$

and when $\beta \in (0, \infty)$, we have

$$e^{-c^*}(1 + e^{c^*})^2 - e^{-\beta_0}(1 + e^{\beta_0})^2 = r(c^*) - r(\beta_0) > 0;$$

Therefore, $h'_2(1) - h'_1(1) > 0$ for any $\beta_0 \in (-\tilde{c}, \infty)$, i.e. $h_1(\alpha)$ and $h_2(\alpha)$ are crossing at $\alpha = 1$. □

Proof of Lemma 4.2.6. First of all, from Lemma 4.2.1 and Equation (4.6), we can write β_0 as a function of c^* ,

$$\beta_0(c^*) = c^* - 2 - \frac{4}{e^{c^*} - 1}.$$

Take the derivative,

$$\beta'_0(c^*) = 1 + \frac{4e^{c^*}}{(e^{c^*} - 1)^2} > 0.$$

Thus $\beta_0(c^*)$ is strictly increasing. As a result, $c^*(\beta_0)$ being a function of β_0 , is also strictly increasing. Thus, with $\beta_0 > -\tilde{c}$, we have $c^* > 1.5$.

Also, from Equation (4.6),

$$c^* - \beta_0 = -2 + \frac{4}{1 - e^{-c^*}}.$$

With $c^* > 1.5$, we know $2 < c^* - \beta_0 < 3.1$.

Now, back to $h_3(0)$. In Lemma 4.2.5 we have shown that $e^{-c^*}(1 + e^{c^*})^2 \geq e^{-\beta_0}(1 + e^{\beta_0})^2$, therefore,

$$\begin{aligned} h_3(0) &\leq -2e^{-\beta_0}(1 + e^{\beta_0})^2 - 2(c^* - \beta_0 + 1)e^{-\beta_0}(1 + e^{\beta_0})^2 + (c^* - \beta_0)^2 e^{-\beta_0}(1 + e^{\beta_0})^2 \\ &= e^{-\beta_0}(1 + e^{\beta_0})^2 [(c^* - \beta_0 - 1)^2 - 5] \\ &< 0. \end{aligned}$$

□