Fractional Factorial and Related Designs– Optimality and Construction

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

of the

Indian Institute of Technology Bombay, India

and

Monash University, Australia

by

Rakhi Singh

Id: [redacted] (IITB) and [redacted](Monash Uni.)

Supervisors:

Prof. Ashish Das (IIT Bombay)

Dr. Daniel Horsley (Monash University)

The course of study for this award was developed jointly by
Monash University, Australia and the Indian Institute of Technology, Bombay
and was given academic recognition by each of them.
The programme was administrated by The IITB-Monash Research Academy.

(Year 2018)

Acknowledgments

At the very outset, as I submit this thesis, I would like to express my sincere gratitude to my IITB supervisor Prof. Ashish Das for his continued guidance, inspiration, patience, motivation, and immense knowledge in the design of experiments and life, in general. I am forever indebted to his fatherly affection and love throughout my Ph.D. He has always treated me more like a colleague and less like a student which has immensely helped me in my growth as a human being. He has greatly contributed to my journey of becoming a scholar from being a student. His relentlessly and tirelessly working for several hours of many continuous days before writing a manuscript or sending a revision has been inspiring to me. His compassion for both research and family has left me awestruck and taught me life-lessons to be cherished forever.

Next, I would like to thank my Monash supervisor Dr. Daniel Horsley. He has always been supportive, gentle and kind with me despite the fact that I knew very little in the area of combinatorics that he is an expert of. His perseverance, perfectionism, clear articulation, hard work, work-ethics and enthusiasm has been an epitome of inspiration to me. He has also facilitated my travel to several conferences in Australia without being asked. I have also been a secret admirer of his being a fitness enthusiast so much so that I gathered courage to run my first 21K race while pursuing Ph.D. even when I had never really ran for more than 5 minutes a day. If I could become even 10% of the academician that he is, I would be forever grateful.

The third most important person who has contributed to my thesis is the love of my life, my husband, Prakash. I still remember how when I started my Ph.D. and it all (a Ph.D. in Mathematics) seemed so difficult, he would sit with me (and provide me ideas) for hours on pavements for solving an inequality, for finding a lower bound that I was supposed to find. He was the first person who continuously motivated me to not give up. Even when he will be extremely busy in his own Ph.D., he has always been just one call away to listen to all my woes and bring that inner happiness and relaxation to me. His ever-changing passions ranging from structural engineering, to being a specialized statistician, philosophy, astronomy, being fit and to play squash has always kept us sane and inspired and has made our Ph.D. journey seemed easier. Saying thanks to him is not enough, perhaps I could only pay him by being there for him the same way as he has been for me and more. This thesis truly belongs to him as much as it belongs to me.

I would especially like to thank Prof. Aloke Dey and Prof. Ching-Shui Cheng, every conversation with them was an academic treat to me. I would also like to thank Prof. Feng Shun Chai and Prof. Pi-wen Tsai for valuable discussions during our research collaborations. I would also like to thank Dr. Heiko Großmann for his valuable insights and discussions on Chapter 8 of the thesis. I would like to thank Dr. Praggya Das for the numerous ways in which she has motivated me and for all the love that she has showered on me and Prakash. She and Prof. Das have truly been a family away from home to us. I would also like to thank Prof. Siuli Mukhopadhyay for not only being a co-author in a manuscript but also for being a friend in the department. Thanks are also due to my RPC members Prof. Ian Wanless and Prof. Sanjeev Sabnis for their patient hearing and valuable suggestions during discussions of my Ph.D. program.

I would like to thank my friends Dr. Satya Prakash Singh and (soon-to-be) Dr. Sreelekshmy for the indefinite hours we have spent on statistical discussions, career options and on all the other important or silly things in life. I would also like to thank Vrinda, Devika, Rosie and Shashibhushan for the infinite memories that we have of the time that we have spent cooking and roaming around together while being in Monash University. Among other friends, I would like to specially thank Aakansha and Sheeba for their support. I would also like to take a moment to thank both mine and Prakash's parents and siblings for being always there for us.

I would also like to take this opportunity to thank all the other faculty and staff at IITB, Monash University and IITB-Monash Research Academy for through and smooth sailing of my boat. Last but not the least, I am much thankful to the facilities (especially, Squash courts) in IIT Bombay campus which has kept me going through all these years fearlessly and compassionately.

Thesis Approval

The thesis entitled

Fractional Factorial and Related Designs– Optimality and Construction

by

Rakhi Singh

is approved for the degree of

Doctor of Philosophy

[redacted]

Guide: Prof. Ashish Das

[redacted]

Guide: Dr. Daniel Horsley

[redacted]

External Examiner

[redacted]

Internal Examiner

[redacted]

Chairman

Date: November 26, 2018

Place: IITB-Monash Research Academy

Thesis Declaration

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

[redacted]

Name: Rakhi Singh

IITB Roll Number: [redacted]

Monash Id: [redacted]

List of Publications

- Singh, Rakhi; Chai, Feng-Shun; Das, Ashish. Optimal two-level choice designs for any number of choice sets. *Biometrika 102* (2015), no. 4, 967–973
- Chai, Feng-Shun; Das, Ashish; Singh, Rakhi. Three-level A- and D-optimal paired choice designs. *Statist. Probab. Lett. 122* (2017), 211–217
- Dey, Aloke; Singh, Rakhi; Das, Ashish. Efficient paired choice designs with fewer choice pairs. *Metrika 80* (2017), no. 3, 309–317
- Chai, Feng-Shun; Das, Ashish; Singh, Rakhi. Optimal two-level choice designs for estimating main and specified two-factor interaction effects. *J. Stat. Theory Pract. 12* (2018), no. 1, 82–92
- Cheng, Ching-Shui; Das, Ashish; Singh Rakhi; Tsai, Pi-Wen; $E(s^2)$- and UE(s^2)-Optimal Supersaturated Designs. *J. Statist. Plann. Inference 196* (2018), 105–114
- Horsley, Daniel; Singh, Rakhi. New lower bounds for t-coverings. *J. Combin. Des 26* (2018), no. 8, 369–386
- Singh, Rakhi; Das, Ashish; Chai, Feng-Shun. Optimal Paired Choice Block Designs. *Stat. Sinica* (2018), *accepted*, doi: 10.5705/ss.202016.0084
- Das, Ashish; Horsley, Daniel; Singh, Rakhi. Pseudo Generalized Youden Designs. *J. Combin. Des 26* (2018), no. 9, 439–454
- Singh, Rakhi. On three-level D-optimal paired choice designs. *Statist. Probab. Lett. 145* (2019), 127–132.
- Das, Ashish; Singh, Rakhi. A unified approach to discrete choice experiments. Ongoing.

Abstract

Design of experiments is a useful and practical branch of statistics. One often needs to predict a relationship between any phenomenon (or, response) and the causes (often, called as the predictors) for that phenomenon. Experimental design involves not only the selection of suitable predictors and outcomes, but planning the delivery of the experiment under statistically optimal conditions given the constraints of available resources.

In this thesis, I have worked in four areas of design of experiments namely, (i) choice designs, (ii) supersaturated designs, (iii) covering designs (or, coverings), and (iv) pseudo generalized Youden designs.

(i) Choice designs: Choice experiments mirror real-world situations closely and help manufacturers, service-providers, policy-makers and other researchers in taking business decisions on the characteristics of their products and services based on the perceived utility. In a paired choice experiment, several pairs of options are shown to respondents. The respondents are asked to give their preference among the many options for each of the choice sets shown to them. In order to conduct an experiment, a choice design is customarily used to efficiently estimate the parameters of interest which essentially consist of either the main effects only or the main plus two-factor interaction effects of the attributes. For two-level paired choice experiments, we have obtained a simple form of the information matrix of a choice design for estimating the main effects, and provided D- and MS-optimal paired choice designs with distinct choice sets under the main effects model for any number of choice sets. It is also shown that the optimal designs under the main effects model are also optimal under the broader main effects model. We found that optimal choice designs with a choice set size two often outperform their counterparts with larger choice set sizes.

Traditionally, while using designs for discrete choice experiments, every respondent is shown the same collection of choice pairs (that is, the choice design). Also, as the

attributes and/or the number of levels under each attribute increases, the number of choice pairs in an optimal paired choice design increases rapidly. Moreover, in the literature under the utility-neutral setup, random subsets of the theoretically obtained optimal designs are often allocated to respondents. The question therefore is whether one can do better than a random allocation of subsets. To address these concerns, in the linear paired comparison model (or, equivalently the multinomial logit model), we first incorporate the fixed respondent effects (also referred to as the block effects) and then obtain optimal designs for the parameters of interest. Our approach is simple and theoretically tractable, unlike other approaches which are algorithmic in nature. We present several constructions of optimal block designs for estimating main effects or main plus two-factor interaction effects. Our results show when and how an optimal design for the model without blocks can be split into blocks so as to retain the optimality properties under the block model. For paired choice designs, two new construction methods are also proposed for the estimation of the main effects. These designs require about 30-50% fewer choice pairs than the existing designs and at the same time have reasonable high D-efficiencies for the estimation of the main effects.

Considering all factors at 3 levels each, and for paired choice designs, we have also obtained a sharper lower bound to the A- and D-values for estimating the main effects under the utility-neutral multinomial logit model. New A- and D-optimal (and efficient) designs are also provided. Considering three-level paired choice designs for estimating all the main effects and two-factor interaction effects under the utility-neutral multinomial logit model, we have provided a general technique involving generators to reduce the number of choice pairs in a D-optimal design. Generators are identified allowing significant reduction in the total number of choice pairs for D-optimal designs.

For two-level choice experiments with k factors, we consider a model involving the main plus all two-factor interaction effects with our interest lying in the estimation of the main effects and a specified set of two-factor interaction effects. The two-factor interaction effects of interest are either (i) one factor interacting with each of the remaining $n-1$ factors or (ii) each of the two factors interacting with each of the remaining $n-2$ factors. For the two models, we first characterized the information matrix and then constructed universally optimal choice designs for choice set sizes 3 and 4.

Several author-groups have contributed to the theoretical development of discrete choice experiments and for finding optimal choice designs under the multinomial logit model. The author-groups Street–Burgess and Huber–Zwerina have adopted different approaches and used seemingly different information matrices under the multinomial logit model. The information matrix plays a crucial role for finding optimal designs in both approaches. Since the expressions for the relevant matrices look very different and it is not obvious how the two approaches are related, this has given rise to some confusion in the literature. We resolve this confusion by showing, in general, how the information matrices under the two approaches are related. There have also been some confusion regarding the inference parameters expressed as linear functions of the utility parameter vector τ. We theoretically establish a unified approach to discrete choice experiments and introduce the general inference problem in terms of a simple linear function of τ. This allows us to show that the commonly used effects coding under the A-criterion for the non-singular full-rank inference problem inherently attaches unequal importance to the elementary contrasts of attribute levels. On the contrary, we see that the orthonormal coding leads to attaching equal importance to the elementary contrasts of attribute levels. However, for a singular full-rank inference problem involving the full set of effects-coded parameters, we show that the orthonormal coding provides an equivalent approach to obtain A-optimal designs.

(ii) Supersaturated designs: Supersaturated designs are useful for factor screening experiments under the factor sparsity assumption that only a small number of factors are active. The popular $E(s^2)$-criterion for choosing two-level supersaturated designs minimizes the sum of squares of the entries of the information matrix over the designs in which the two levels of each factor appear equal number of times. Recently Jones and Majumdar (2014) proposed the $UE(s^2)$-criterion which is essentially the same as the $E(s^2)$-criterion except that the requirement of factor-level-balance is dropped. Since this requirement is bypassed, usually there are many $UE(s^2)$-optimal designs with diverse characteristics and performances. It is necessary to choose better designs from them. We proposed additional criteria and provided constructions for superior $UE(s^2)$-optimal designs having good projection properties. Usually $E(s^2)$-optimal designs are difficult to construct, whereas our construction methods of superior $UE(s^2)$-optimal designs are simple and systematic. We also identified several families of designs that are both $E(s^2)$- and $UE(s^2)$-optimal.

(iii) Coverings: A t-(v, k, λ) covering is a collection of k-element subsets, called blocks, of a v-set of points such that each t-subset of points occurs in at least λ blocks. If each t-subset of points occurs in exactly λ blocks the covering is a t-(v, k, λ) design. Fisher's inequality famously states that every 2-(v, k, λ) design has at least v blocks. In 1975 Ray-Chaudhuri and Wilson generalised this result to higher t by showing that every t-(v, k, λ) design has at least $\binom{v}{\lfloor t/2 \rfloor}$ blocks, and Wilson gave a streamlined proof of this result in 1982. Horsley (2017) adapted a well-known proof of Fishers inequality to produce a new lower bound on the number of blocks in some 2-(v, k, λ) coverings. In this thesis, we have shown how ideas from these papers can be combined to obtain improved lower bounds on the number of blocks in t-(v, k, λ) coverings for $t > 2$. We have also identified some infinite families of parameter sets where our bound exists and is an improvement over the best available lower bounds. We also found an infinite family where our bound is tight, that is, there exists a t-(v, k, λ) covering attaining our bound.

(iv) Pseudo generalized Youden designs: Sixty years ago, Kiefer (1958) introduced generalized Youden designs (GYDs) for eliminating heterogeneity in two directions. A GYD is a row-column design whose k rows form a balanced block design (BBD) and whose b columns do likewise. Later Cheng (1981b) introduced pseudo Youden designs (PYDs) in which $k = b$ and where the k rows and the b columns, considered together as blocks, form a BBD. Kiefer (1975b) proved a number of results on the optimality of GYDs. A PYD has the same optimality properties as a GYD. In this thesis, we have introduced and investigated pseudo generalized Youden designs (PGYDs) which generalise both GYDs and PYDs. A PGYD is a row-column design where the k rows and b columns, considered together as blocks, form an equireplicate generalized binary variance balanced design. Every GYD is a PGYD and a PYD is exactly a PGYD with $k = b$. We have shown, however, that there are situations where a PGYD exists but neither a GYD nor a PYD does. We also obtained necessary conditions, in terms of v, k and b, for the existence of a PGYD. Using these conditions, we provided an exhaustive list of parameter sets satisfying $v \leqslant 25, k \leqslant 50, b \leqslant 50$ for which a PGYD exists. We constructed families of PGYDs using patchwork methods based on affine planes.

Contents

List of Tables

Chapter 1

Introduction

1.1 Motivation and Background

Design of experiments is a useful and practical branch of statistics. One often needs to predict a relationship between any phenomenon (or, response) and the causes (often, called the predictors) for that phenomenon. Design of experiments helps in such prediction. Experimental design involves not only the selection of suitable predictors and outcomes, but planning the delivery of the experiment under statistically optimal conditions given the constraints of available resources.

The experiments with only one predictor variable at various discrete levels are commonly known as block designs. Block designs consist of many blocks in which predictor variables are replicated appropriately to achieve the desired precision of the experiment. Typically, blocking is done using a blocking factor which controls the variability that is not of primary interest to the experimenter. For example, in an experiment involving prediction of marks of students of class X, the gender of students can be treated as a blocking factor.

If an experiment, however, involves many predictor variables each with discrete possible values or "levels", then it is termed as a factorial experiment and the design layout is known as a factorial design. Such an experiment allows the investigator to study the effect of each predictor (or, factor) on the response variable, as well as the effects of interactions between factors on the response variable. For the vast majority of factorial experiments, each factor has only two or three levels.

Correctly designed experiments help in gaining better knowledge in many practical fields like nature sciences, social sciences, engineering, marketing, policy making, etc.

In this report, we discuss some open problems in four areas of design of experiments namely, (i) choice designs, (ii) supersaturated designs, (iii) covering designs (or, coverings), and (iv) pseudo generalized Youden designs.

1.1.1 Choice Designs

Choice experiments are widely used in marketing, transport, environmental resource economics and public welfare analysis. They mirror real-world situations closely and help manufacturers, service-providers, policy-makers and other researchers in taking business decisions on the characteristics of their products and services based on the perceived utility.

A choice experiment consists of N choice sets, each containing m options. A respondent is shown each of the choice sets in turn and is asked for the preferred option as per his or her perceived utility. Each option in a choice set is described by a set of k attributes, where each attribute has two or more levels. We assume that there are no repeated options in a choice set. Furthermore, in this thesis, we consider that for a set of k attributes, for $i = 1, \ldots, k$, the ith attribute has v_i levels, $v_i \geq 2$. We represent the v_i levels by $0, \ldots, v_i - 1$, unless stated otherwise. Thus, there are a total of $\prod_i v_i$ options. It is ensured that respondents choose one of the options in each choice set. A choice design is a collection of choice sets that are employed in a choice experiment. Though choice designs may contain repeated choice sets, one may prefer that no two choice sets are repeated. For excellent reviews of designs for choice experiments, see Großmann and Schwabe (2015) and Street and Burgess (2012).

Huber and Zwerina (1996), following the seminal work of McFadden (1974), used a modelling approach to compare choice designs. Subsequently Street and Burgess (2007), using the approach of El-Helbawy and Bradley (1978), presented a comprehensive exposition of designs for choice experiments under the multinomial logit model. The model specifies the probability of an individual choosing one of the m options from a choice set. For a paired choice design, for example, the multinomial logit model supposes that the probability of preferring option 1 over option 2 in the ith choice pair is $\pi_{12i} = e^{u_{1i}}/(e^{u_{1i}} + e^{u_{2i}})$, where u_{1i} and u_{2i} represent the systematic part of the utilities attached to the two options in the ith choice pair. Similarly $\pi_{21i} = 1 - \pi_{12i}$ is the probability that option 2 is preferred over option 1. It follows that for the ith choice pair, the choice probabilities depend only on the utility difference $u_{1i} - u_{2i}$. For a design d with N choice pairs, since options are described by k factors, following Huber and Zwerina (1996), the utilities are modelled using the linear predictor $u_j = P_j\theta$, where θ is a vector representing the main effects, P_j is an effects coded matrix for the jth option, and $u_j = (u_{ji})$ is an $N \times 1$ utility vector for the jth option, $j = 1, 2; i = 1, \ldots, N$. The utility difference $u_1 - u_2 = (P_1 - P_2)\theta = X\theta$ is then a linear function of the parameter vector θ. In what follows, we refer to X as the design matrix of design d. Since multinomial logit choice models are non-linear in the parameters and the information matrix is a function of the parameters, a utility-neutral approach (that is, taking $\theta = 0$) of finding the information matrix has been developed over the last two decades. Under such a utility-neutral multinomial logit model, the Fisher information matrix for a design d is $(1/4)M_d$, where $M_d = X^T X$.

Simultaneously, Graßhoff et al. (2004) studied linear paired comparison designs which are analyzed under the linear paired comparison model. Here, the quantitative response Z is the observed utility difference between the two options and is described by the model, $Z = U_1 - U_2 + \epsilon = (P_1 - P_2)\theta + \epsilon = X\theta + \epsilon$, where ϵ is a random error vector. For a design d, the matrix M_d is the information matrix under the linear paired comparison model. Although the linear paired comparison and the multinomial logit models follow different approaches, the information matrix for a choice design for $m = 2$ under the latter with equal choice probability (Huber and Zwerina, 1996) is proportional to that under the former.

One objective of a choice experiment is to optimally or efficiently estimate the parameters of interest which essentially consist of either only the main effects or the main plus two-factor interaction effects of the k attributes. As noted in Großmann and Schwabe (2015), most optimality results for choice designs are available for the D-criterion. A D-optimal design has the maximum determinant of the information matrix among all competing information matrices. D-criterion is invariant to reparameterizations or in other words, it does not depend on the coding of the attribute levels. Furthermore, Großmann and Schwabe (2015) observed that the paired choice designs that are optimal under the linear paired comparison model are also D-optimal under the multinomial logit model and vice versa.

Optimal designs have been obtained theoretically under the utility-neutral setup, for example, see Graßhoff et al. (2003), Graßhoff et al. (2004), Street and Burgess (2007), Street and Burgess (2012), Demirkale, Donovan and Street (2013), Bush (2014), Großmann and Schwabe (2015) and Singh, Chai and Das (2015). We refer the reader to comprehensive reviews provided by Street and Burgess (2007) and Großmann and Schwabe (2015).

1.1.2 Supersaturated Designs

In an n-run factorial experiment involving m two-level factors, for the general mean and all the main effects to be estimable, we must have $n \geqslant m + 1$. A design is called supersaturated if $n < m + 1$. Under the assumption that only a small number of factors are active (factor sparsity), a supersaturated design can provide considerable cost saving in factor screening. In supersaturated designs, as in factorial experiments, most of the results correspond to the situations where each of the factors has two levels. Each two-level supersaturated design can be represented by an $n \times m$ matrix having entries 1s and -1s, with each column of X_d corresponding to one factor and each row representing a factor-level combination. A factor is said to be level-balanced if the corresponding column of X_d has the same numbers of 1s and -1s. This is possible only if n is even. For an odd n, a factor is said to be nearly level-balanced if in the corresponding column the numbers of times 1 and -1 appear differ by one. Without loss of generality, we require that 1 appears $(n-1)/2$ times and -1 appears $(n+1)/2$ times. A design is said to be level-balanced (respectively, nearly level-balanced) if all the factors are level-balanced (respectively, nearly level-balanced). Usually, under the main-effects model, one is interested in finding the best lower bounds for $E(s^2)$ (which is defined later and is a measure of non-orthogonality of the design) and then one is also interested in finding the designs which attain the lower bounds to $E(s^2)$. More details are given in Chapter 9.

1.1.3 Coverings

An *incidence structure* is a pair $(V, \mathcal{B})$ where V is a set of *points* and $\mathcal{B}$ is a collection of subsets of V called *blocks*. For positive integers t, v, k and λ with $t < k < v$, a t-(v, k, λ) *covering* is an incidence structure $(V, \mathcal{B})$ such that $|V| = v$, $|B| = k$ for all $B \in \mathcal{B}$, and each t-subset of V is contained in at least λ blocks in $\mathcal{B}$. If each t-subset of V is contained in exactly λ blocks in $\mathcal{B}$, then $(V, \mathcal{B})$ is a t-(v, k, λ) *design*. Usually we are interested in finding coverings with as few blocks as possible. The *covering number* $C_\lambda(v, k, t)$ is the minimum number of blocks in any t-(v, k, λ) covering. The *Schönheim bound* for the covering number is given by

$$C_\lambda(v,k,t) \geqslant L_\lambda(v,k,t) \quad \text{where} \quad L_\lambda(v,k,t) = \left\lceil \frac{v}{k} \left\lceil \frac{v-1}{k-1} \cdots \left\lceil \frac{v-t+2}{k-t+2} \left\lceil \frac{\lambda(v-t+1)}{k-t+1} \right\rceil \right\rceil \cdots \right\rceil \right\rceil .$$

We are interested in improving lower bounds on covering numbers, wherever possible, and in constructing coverings attaining the improved lower bounds. More details are given in Chapter 10.

1.1.4 Pseudo Generalized Youden Designs

A GYD is a row-column design whose k rows form a balanced block design (BBD) and whose b columns do likewise. Pseudo Youden designs (PYDs) are designs in which $k = b$ and where the k rows and the b columns, considered together as blocks, form a BBD. We introduce and investigate pseudo generalized Youden designs (PGYDs) which generalize both GYDs and PYDs. A PGYD is a row-column design where the k rows and b columns, considered together as blocks, form an equireplicate generalized binary variance balanced design. Every GYD is a PGYD and a PYD is exactly a PGYD with $k = b$. More details are given in Chapter 11.

1.2 Outline of the Thesis

In this thesis, each chapter is independent in itself and therefore notations are consistent only throughout each chapter. Chapters 2–8 is a collection of work done in the area of choice experiments. All chapters except Chapter 6 and Chapter 8 are published or accepted in peer-reviewed journals. Chapter 6 has been submitted to *Statistics & Probability Letters* and Chapter 8 is an ongoing work. Chapter 9 is in the area of supersaturated designs. Chapter 10 and Chapter 11 are in the area of coverings and pseudo generalized Youden designs, respectively. Chapters 9–11 are also published/accepted in peer-reviewed journals. Chapter 12 provides conclusions and future research in the areas I have covered in this thesis.

Chapter 2

Optimal two-level choice designs for any number of choice sets

This chapter is based on the following work:

Singh et al. (2015): Singh, Rakhi; Chai, Feng-Shun; Das, Ashish. Optimal two-level choice designs for any number of choice sets. *Biometrika 102* (2015), no. 4, 967–973.

2.1 Introduction and preliminaries

In this chapter, we consider each attribute to be at two levels, -1 and 1, leading to a total of 2^k options. Let $T_i = (t_{i1}, \ldots, t_{im})$ denote the ith choice set, where $t_{i\alpha}$ is the αth option in the ith choice set ($i = 1, \ldots, N$; $\alpha = 1, \ldots, m$). The collection of all such choice sets T_i is called a choice design T, with parameters N, k and m. As in Street and Burgess (2007), under the multinomial model and equal choice probabilities, the information matrix for options of a choice design with N choice sets is $\Lambda = (\Lambda_{(r,s)})$, where

$$m^2 N \Lambda_{(r,s)} = \begin{cases} (m-1)n_r, & r = s, \\ -n_{r,s}, & r \neq s \end{cases}$$

with r and s the labels of the corresponding options, n_r the number of times option label r appears in the choice design and $n_{r,s}$ the number of times option labels r and s occur together in choice sets of the design.

We consider choice experiments where our interest is restricted to the main effects of the attributes. For a 2^k choice experiment, let B represent the orthonormal contrast matrix for the k main effects and let $B_{(2)}$ represent the orthonormal contrast matrix for all the $k(k-1)/2$ two-factor interactions. Then, in the main effects model with no interactions, the information matrix, also called the C-matrix, of the main effects is $C_m = B\Lambda B^T$. Similarly, in the broader main effects model, where all interactions that involve three or more factors are absent, the C-matrix of the main effects is

$$C_m^{(2)} = C_m - B\Lambda B_{(2)}^T(B_{(2)}\Lambda B_{(2)}^T)^- B_{(2)}\Lambda B^T. \tag{2.1}$$

A choice design is connected if all the main effects are estimable under the main effects model, and this happens if and only if C_m has rank k. In what follows, the class of all connected choice designs that involve k two-level attributes and N choice sets each of size m is denoted by $\mathcal{D}_{N,k,m}$. For a choice design $T \in \mathcal{D}_{N,k,m}$, let $0 < \gamma_1 \leq \cdots \leq \gamma_k$ be the eigenvalues of C_m. Then, $T^* \in \mathcal{D}_{N,k,m}$ is said to be A-, D-, or E-optimal in $\mathcal{D}_{N,k,m}$ if, respectively, $\sum_{i=1}^k \gamma_i^{-1}$, $\prod_{i=1}^k \gamma_i^{-1}$, or γ_1^{-1} is minimum for the design T^*. Furthermore, Eccleston and Hedayat (1974) introduced the MS-optimality criterion: $T^* \in \mathcal{D}_{N,k,m}$ is said to be MS-optimal in $\mathcal{D}_{N,k,m}$ if $\sum_{i=1}^k \gamma_i^2$ is minimum for the design T^* among all designs $T \in \mathcal{D}_{N,k,m}$ having maximum $\sum_{i=1}^k \gamma_i$.

Graßhoff et al. (2004) and Demirkale et al. (2013) obtained D-optimal paired choice designs, which are also A-, E- and MS-optimal, under the main effects model. Though their results are exhaustive, for two-level choice designs, their C-matrix is necessarily a scalar multiple of identity matrix, with N being a multiple of 4. Graßhoff et al. (2004) also noted that under their setup the model of paired comparisons is equivalent to the weighing of k objects in a chemical balance. Under the main effects model, we derive a simple form of the C-matrix in terms of the design matrix of the paired choice design. We see that even under a broader main effects model, there is a one-one correspondence between optimal paired choice designs and chemical balance weighing designs. Thus, by suitably modifying the constructions on the weighing designs, we construct new D- and MS-optimal paired choice designs in $\mathcal{D}_{N,k,2}$, for all N under the broader main effects model. We also find that the optimal choice designs with $m = 2$ often outperform their counterparts with $m > 2$.

2.2 Information Matrix and D-optimal Designs

Let $m = 2$. For the ith choice set T_i, let $t_{i\alpha} = (t_{i\alpha}^{(1)} \cdots t_{i\alpha}^{(k)})$ where $t_{i\alpha}^{(j)}$ represents the level of the jth attribute in the αth option. For $\alpha = 1, 2$, define an $N \times k$ matrix $P_\alpha = (t_{i\alpha}^{(j)})$ such that $\{P_1, P_2\}$ represent the paired choice design T. Also, let $X = (P_1 - P_2)/2$. Henceforth, we will refer to the matrix X as the paired choice design matrix, or simply, the design matrix. Since X is a matrix with elements ± 1 and 0, it is similar to a chemical balance weighing design matrix. We now present a simple form of the information matrix of a choice design for estimating the main effects.

Theorem 2.1. *For a paired choice design T with parameters N and k, $C_2 = X^T X/(N2^k)$. Also, $rank(C_2) = k$ only if $k \leqslant N$.*

Großmann and Schwabe (2015) established how the information matrix under the equal choice probability approach of Huber and Zwerina (1996) for $m \geq 2$ is related to the information matrices for pairs. Supplementing this, we show that under the equal choice probability multinomial logit model approach of Street and Burgess (2007) the information matrix for a choice design with $m \geq 2$ is proportional to the sum of $m(m-1)/2$ information matrices of paired choice designs. This generalizes Theorem 2.1, details of which are as follows.

As in Street and Burgess (2007), under the multinomial model and equal choice probabilities, the information matrix, for options, of a choice design with N choice sets is,

$$\Lambda = \frac{1}{N} \sum_{i=1}^{N} \Lambda_i = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{m^2} \sum_{\alpha=1}^{m-1} \sum_{\alpha_1=\alpha+1}^{m} \Delta_{i(\alpha\alpha_1)} \right), \tag{2.2}$$

where Λ_i is the information matrix of the ith choice set T_i and for $\alpha \neq \alpha_1 = 1, \ldots, m$, $\Delta_{i(\alpha\alpha_1)}$, with elements $\Delta_{i(\alpha\alpha_1)}(r, s)$, is a matrix of order 2^k corresponding to αth and α_1th options in the ith choice set, with

$$\Delta_{i(\alpha\alpha_1)}(r, s) = \begin{cases} n_r^{(i)}, & r = s, \\ -n_{r,s}^{(i)}, & r \neq s. \end{cases}$$

Here, r and s are the labels of the corresponding options, $n_r^{(i)}$ is the number of times option label r appears in the pair $(t_{i\alpha}, t_{i\alpha_1})$ and $n_{r,s}^{(i)}$ is the number of times option labels r and s occur together in the pair $(t_{i\alpha}, t_{i\alpha_1})$.

Under the main effects model, the following lemma derives the C-matrix for general m in terms of the paired choice design matrices. We then see that Theorem 2.1 is a special case of Lemma 2.2.

Lemma 2.2. *Let T be a choice design with parameters N, k and m. For any $\alpha \neq \alpha_1 = 1, \ldots, m$, define the $N \times k$ matrix $P_\alpha = (t_{i\alpha}^{(j)})$ and $X_{\alpha\alpha_1} = (P_\alpha - P_{\alpha_1})/2$. Then,*

$$C_m = \frac{1}{m^2 N 2^{k-2}} \sum_{\alpha=1}^{m-1} \sum_{\alpha_1=\alpha+1}^{m} X_{\alpha\alpha_1}^T X_{\alpha\alpha_1}.$$

Proof. Let $t_{i\alpha}$ and $t_{i\alpha_1}$ be the ith rows of P_α and P_{α_1}, respectively, and let $x_{i\alpha\alpha_1}$ be the ith row of $X_{\alpha\alpha_1}$. Without loss of generality, let $t_{i\alpha}$ and $t_{i\alpha_1}$ correspond to the rth and the sth lexicographic labels, with $r < s$.

Without any loss of generality, we take the $k \times 2^k$ orthonormal contrast matrix B for main effects, as defined in Street and Burgess (2007). The columns of B are lexicographic arrangement of all 2^k options. Let $2^{(k/2)}B = (B_1 \;\; b_r \;\; B_2 \;\; b_s \;\; B_3)$, where B_1 is of order $k \times (r-1)$, B_2 is of order $k \times (s-r-1)$, and B_3 is of order $k \times (2^k - s)$. Since rth and sth column of B are the rth and sth treatment combinations in lexicographic order, respectively, $b_r = t_{i\alpha}^T$ and $b_s = t_{i\alpha_1}^T$.

Also, by definition, $\Delta_{i(\alpha\alpha_1)} = \begin{pmatrix} 0_{2^k\times(r-1)} & w_{i\alpha\alpha_1}^T & 0_{2^k\times(s-r-1)} & -w_{i\alpha\alpha_1}^T & 0_{2^k\times(2^k-s)} \end{pmatrix}$ where

$w_{i\alpha\alpha_1} = \begin{pmatrix} 0_{1\times(r-1)} & 1 & 0_{1\times(s-r-1)} & -1 & 0_{1\times(2^k-s)} \end{pmatrix}$. Then,

$$\begin{aligned} 2^k B\Delta_{i(\alpha\alpha_1)}B^T &= 2^{(k/2)}\begin{pmatrix} 0_{k\times(r-1)} & (t_{i\alpha}^T - t_{i\alpha_1}^T) & 0_{k\times(s-r-1)} & (t_{i\alpha_1}^T - t_{i\alpha}^T) & 0_{k\times(2^k-s)} \end{pmatrix}B^T \\ &= 2^{(k/2)}\begin{pmatrix} 0_{k\times(r-1)} & 2x_{i\alpha\alpha_1}^T & 0_{k\times(s-r-1)} & -2x_{i\alpha\alpha_1}^T & 0_{k\times(2^k-s)} \end{pmatrix}B^T \\ &= 2(x_{i\alpha\alpha_1}^T t_{i\alpha} - x_{i\alpha\alpha_1}^T t_{i\alpha_1}) = 2x_{i\alpha\alpha_1}^T(t_{i\alpha} - t_{i\alpha_1}) = 4x_{i\alpha\alpha_1}^T x_{i\alpha\alpha_1}. \end{aligned}$$

From (2.2), we get,

$$\begin{aligned} C_m &= B\Lambda B^T = \tfrac{1}{N} B \textstyle\sum_{i=1}^N \left(\frac{1}{m^2} \sum_{\alpha=1}^{m-1} \sum_{\alpha_1=\alpha+1}^{m} \Delta_{i(\alpha,\alpha_1)}\right) B^T \\ &= \tfrac{1}{m^2 N 2^{k-2}} \textstyle\sum_{\alpha=1}^{m-1} \sum_{\alpha_1=\alpha+1}^{m} \left(\sum_{i=1}^N x_{i\alpha\alpha_1}^T x_{i\alpha\alpha_1}\right) \qquad \square \\ &= \tfrac{1}{m^2 N 2^{k-2}} \textstyle\sum_{\alpha=1}^{m-1} \sum_{\alpha_1=\alpha+1}^{m} X_{\alpha\alpha_1}^T X_{\alpha\alpha_1}. \end{aligned}$$

Remark 2.3. *Großmann and Schwabe (2015) show that the information matrix for the D-optimal two-level paired choice design d^* under the equal choice probability multinomial logit model approach of Street and Burgess (2007) is proportional to the information matrix of the approximate uniform D-optimal two-level design ξ^* under the linear paired* *comparison model. In contrast, Theorem 2.1 establishes that the result holds true for any two-level paired choice design d.*

We now provide optimal designs in $\mathcal{D}_{N,k,2}$. From Theorem 2.1, it follows that finding an optimal paired choice design T is equivalent to finding an optimal design matrix X with elements ± 1 and 0. Galil and Kiefer (1980) showed that for D-optimal designs, within the class of all choice design matrices, it suffices to find a D-optimal design within the class of choice design matrices with elements ± 1 only.

From (2.1) we see that a sufficient condition for $C_m^{(2)} = C_m$ is $B\Lambda B_{(2)}^T = 0$. Furthermore, corresponding to any X with elements ± 1 only, the choice design $d \in \mathcal{D}_{N,k,2}$ has $B\Lambda B_{(2)}^T = 0$. This orthogonality condition holds since $\Lambda B_{(2)}^T = 0$ if X has elements ± 1 only. Details are as follows.

The following result establishes a sufficient condition for orthogonality of main effects and two-factor interaction effects in a paired choice design.

Lemma 2.4. *Let T be a paired choice design with parameters N and k. Then $\Lambda B_{(2)}^T = 0$ if X has elements ± 1 only.*

Proof. Let X have elements ± 1 only. Then, for the corresponding paired choice design T, the ith choice set $T_i = (t_{i1},\ -t_{i1})$, $i = 1, \ldots, N$. Without loss of generality, let t_{i1} will correspond to the rth lexicographic label, $r \leq 2^{k-1}$. Then $-t_{i1}$ corresponds to the $(2^k - r + 1)$th lexicographic label. Let $2^k B_{(2)} = (B_1^* \quad b_r^* \quad B_2^* \quad b_{2^k-r+1}^* \quad B_3^*)$, where B_1^* is of order $k^* \times (r-1)$, B_2^* is of order $k^* \times (2^k - 2r)$, and B_3^* is of order $k^* \times (r-1)$ where $k^* = k(k-1)/2$. It is easy to see that $b_r^* = b_{2^k-r+1}^*$. Also since, $\Lambda_i = \begin{pmatrix} 0_{2^k\times(r-1)} & w_i^T & 0_{2^k\times(2^k-2r)} & -w_i^T & 0_{2^k\times(r-1)} \end{pmatrix}$ with $4w_i = \begin{pmatrix} 0_{1\times(r-1)} & 1 & 0_{1\times(2^k-2r)} & -1 & 0_{1\times(r-1)} \end{pmatrix}$, therefore, $N2^k\Lambda B_{(2)}^T = 2^k \sum_{i=1}^N \Lambda_i B_{(2)}^T = \sum_{i=1}^N w_i^T (b_r^{*T} - b_{2^k-r+1}^{*T}) = 0$. □

Thus, if a design matrix X with elements ± 1 is optimal under the main effects model then it is also optimal under the broader main effects model. In this context, we note that Graßhoff et al. (2003) established optimality of a uniform design $\bar{\xi}_d$ for the main effects having a block diagonal information matrix with block matrices corresponding to main effects and two-factor interactions respectively. In contrast, our orthogonality condition $B\Lambda B_{(2)}^T = 0$ establishes that any paired choice design d such that X_d has elements ± 1

only has a block-diagonal information matrix where the two block matrices need not be scalar multiples of the identity matrix.

Let $H_{N,N}$ denote a Hadamard matrix of order $N \equiv 0 \pmod 4$ in its normal form. For every $k \leq N$, by deleting any $N-k$ columns of $H_{N,N}$ we get $H_{N,k}$ such that $H_{N,k}^T H_{N,k} = NI_k$, where I_k is the identity matrix of order k. In such a situation $X = H_{N,k}$ is A-, D-, E- and MS-optimal.

A choice design with no two choice sets repeated has a design matrix with distinct rows. Two rows of a design matrix are distinct if the absolute value of their inner product is less than k. Modifying the constructions of D-optimal matrices given in Payne (1974) and Galil and Kiefer (1980), we now provide new constructions such that the rows of X are all distinct. We take up the cases $N \equiv i \pmod 4$ $(i = 0, 1, 2, 3)$ separately.

Construction-(0) for $N \equiv 0$ *(mod 4),* $k \leq N$: Starting from $H_{N,N}$, it is easy to see that one can randomly delete a maximum of $N/2 - 1$ columns resulting in an optimal design matrix $X_0 = H_{N,k}$ with $k > N/2$, such that no two rows have an inner product equal to $\pm k$, that is, all rows are distinct. However, in order to delete $N/2$ or more columns, one would need to delete columns carefully to ensure that all rows are distinct. As established in Graßhoff et al. (2004) and Demirkale et al. (2013), the choice design d_0 corresponding to X_0 is A-, D- and E-optimal in $\mathcal{D}_{N,k,2}$.

Construction-(I) for $N \equiv 1$ *(mod 4),* $k < N$: Consider $X_0 = H_{N-1,k}$ of Construction-(0). To ensure that no two choice sets are repeated, one may add to X_0 any row of ± 1's not present in X_0 or $-X_0$ to get a design matrix, say X_1. The following theorem shows that the resultant X_1 is D-optimal.

Theorem 2.5. *For* $N \equiv 1$ *(mod 4) and* $k < N$*, one can add any row consisting of entries* ± 1 *to* $H_{N-1,k}$*, and the resultant paired choice design* d_1 *corresponding to* X_1 *is* D*-optimal in* $\mathcal{D}_{N,k,2}$*.*

Proof. Let $H_{N-1,N-1}$ be a Hadamard matrix of order $N-1$ and a be column vector of order k consisting entirely of entries ± 1. Keeping any k columns of $H_{N-1,N-1}$, a $(N-1) \times k$ matrix $H_{N-1,k}$ is obtained. Then, $H_{N-1,k}^T H_{N-1,k} = (N-1) I_k$.

Let X_1 be an $N \times k$ matrix with $k < N$ such that $X_1 = (\ H_{N-1,k}\ a^T\ )^T$. Then, $X_1^T X_1 = H_{N-1,k}^T H_{N-1,k} + aa^T$, or $X_1^T X_1 = (N-1) I_k + aa^T$, and the eigenvalues of

$X_1^T X_1$ are $N-1$ and $N-1+a^Ta$ with respective multiplicities $k-1$ and 1. Thus, $\det(X_1^T X_1) = (N-1+k)(N-1)^{k-1}$, attaining the theoretical bound as obtained in Payne (1974). Therefore, X_1 is D-optimal. □

Unlike the construction of Payne (1974), Theorem 2.5 allows us to broaden the selection of the D-optimal paired choice designs by adding any one of the $2^k - 2(N-1)$ possible options, which are not options in the rows of $\pm X_0$. Furthermore, based on the results of Cheng (1980), it follows that the paired choice design d_1 is also A- and E-optimal in $\mathcal{D}_{N,k,2}$.

Construction-(II) for $N \equiv 2$ (mod 4), $k \leq N$: For $k \leq N-2$, consider $X_0 = H_{N-2,k}$ of Construction-(0). Then to obtain X_2, we add to X_0 two rows as follows. A row of all 1's is added after multiplying any column of X_0 by -1. As second row, one can add any row consisting of entries ± 1 such that number of 1's and -1's differ by at most 1 and is distinct from the other $N-1$ rows. The resultant paired choice design d_2, corresponding to the design matrix X_2, is D-optimal in $\mathcal{D}_{N,k,2}$, as multiplying any column by -1 doesn't change the Hadamard properties of $H_{N-2,k}$. Cheng (1980) and Jacroux et al. (1983) showed that X_2 is also E- and A-optimal within the restricted class of choice design matrices with elements ± 1 only. Furthermore, for certain values of k, Cheng et al. (1985) showed that X_2 with respective choice design d_2 is A-optimal in $\mathcal{D}_{N,k,2}$.

For $k = N$ and $N-1$, consider $X_0 = H_{N+2,k}$ of Construction-(0). Then delete from X_0 the first row of all 1's and a row such that number of 1's and -1's differ by at most 1. This results in X_2 corresponding to a paired choice design d_2. Payne (1974) and Galil and Kiefer (1980) did not provide any constructions for $k = N$ and $N-1$.

Construction-(III) for $N \equiv 3$ (mod 4), $k \leq N$: Consider $X_0 = H_{N+1,k}$ of Construction-(0). Delete any row from X_0 to get a design matrix, say X_3. This would facilitate to get N distinct rows of X_3 if X_0 had at most two repeated rows. The following theorem, proof of which follows on lines similar to Theorem 2.5, establishes D-optimality of X_3.

Theorem 2.6. *For $N \equiv 3$ (mod 4) and $k \leq N$, one can delete any row from $H_{N+1,k}$, and the resultant paired choice design d_3 corresponding to X_3 is D-optimal in $\mathcal{D}_{N,k,2}$ for $k \leq (N+5)/2$.*

Unlike the constructions of Payne (1974) and Galil and Kiefer (1980), this result allows us to broaden the selection of D-optimal paired choice designs by deleting any one of the $N+1$ possible options which are options in the rows of $H_{N+1,k}$. For certain values of k, Cheng et al. (1985) showed that d_3 is A-optimal in $\mathcal{D}_{N,k,2}$

Remark 2.7. *So far, under the broader main effects model, we have generally provided D-optimal, and in some cases, A- and E-optimal, paired choice designs with distinct choice pairs. However, there are several situations where systematic construction of D-optimal design matrices are not available. Below, we summarize the cases in which uncertainties remain:*

(a) when $N \equiv 1 \pmod 4$, no systematic construction is available for $k = N$ except when $2N-1$ is a perfect square;

(b) when $N \equiv 2 \pmod 4$, no systematic construction is available for $k = N$ and $k = N-1$;

(c) when $N \equiv 3 \pmod 4$, neither sharp upper bounds to $\det(X^TX)$, nor systematic constructions are available for $(N+5)/2 < k \leqslant N$.

The link `http://www.indiana.edu/~maxdet/` *and Galil and Kiefer (1982) provide examples of D-optimal matrices for $k = N \leq 119$ and $N \equiv 3 \pmod 4$, $(N+5)/2 < k \leqslant N$, respectively.*

2.3 *MS*-optimal Designs

In order to address situations where D-optimal designs could not be identified, we now find *MS*-optimal designs in $\mathcal{D}_{N,k,2}$. From Theorem 2.1 it follows that, a paired choice design in $\mathcal{D}_{N,k,2}$ with its C-matrix having maximum $\sum_{i=1}^{k} \gamma_i$ has a corresponding design matrix necessarily belonging to the class of choice design matrices with elements ± 1 only. Thus, finding a *MS*-optimal paired choice design is equivalent to finding a design matrix X with elements ± 1 only, such that $\sum_{i=1}^{k} \lambda_i^2$ is minimum where $0 < \lambda_1 \leq \cdots \leq \lambda_k$ are the eigenvalues of X^TX. The eigenvalues of C_2 are $\gamma_i = \lambda_i/(N2^k)$ ($i = 1, \ldots, k$). First we provide a lower bound to $\sum_{i=1}^{k} \lambda_i^2$.

Theorem 2.8. *Let X be a matrix with elements ± 1 only. Then $\sum_{i=1}^{k} \lambda_i^2 \geqslant N^2k + L$, where*

$$L = \begin{cases} 0, & N \equiv 0 \ (mod\ 4), \\ 2k(k-2), & N \equiv 2 \ (mod\ 4),\ k\ even, \\ 2\{k(k-2)+1\}, & N \equiv 2 \ (mod\ 4),\ k\ odd, \\ k(k-1), & N \equiv 1 \ (mod\ 4)\ or\ N \equiv 3 \ (mod\ 4). \end{cases}$$

Proof. Let $X^TX = M = (m_{ij})$. Then $\sum_{i=1}^{k} \lambda_i^2 = \mathrm{tr}(M^2) = N^2k + \sum_{i=1}^{k}\sum_{j(\neq i)=1}^{k} m_{ij}^2$, which is used to get the final bound. For every given row of X, the four possible values for the ith and jth column entries are $(1,1)$, $(1,-1)$, $(-1,1)$ and $(-1,-1)$. For the ith and jth columns of X, let f_1, f_2, f_3 and f_4 be the number of rows of X with entries $(1,1)$, $(1,-1)$, $(-1,1)$ and $(-1,-1)$, respectively. Then, $f_1+f_2+f_3+f_4 = N$ and $(f_1+f_4)-(f_2+f_3) = m_{ij}$, which implies that

$$f_1 + f_4 = (m_{ij} + N)/2, \quad f_2 + f_3 = (N - m_{ij})/2. \tag{2.3}$$

Since $f_1 + f_4$ is an integer, m_{ij} is even when N is even and m_{ij} is odd when N is odd.

Now we define $e_{ij} = (-1)^{2f_1+f_2+f_3}$. For $N \equiv 2$ (mod 4), using (2.3), when $m_{ij} \equiv 0$ (mod 4) then $e_{ij} = -1$, and when $m_{ij} \equiv 2$ (mod 4) then $e_{ij} = 1$. We observe that, for $k \geq 3$, there is no matrix X such that $m_{ij} = 0$ for all i, j. Thus, we minimize the number of m_{ij}'s taking the value 2. The minimum number of cases for $e_{ij} = 1$ and $e_{ji} = 1$ are $k(k-2)/2$ for k even, and are $\{k(k-2)+1\}/2$ for k odd. For each such case $\mid m_{ij} \mid \geq 2$ for all i, j. Hence, $\sum_{i=1}^{k}\sum_{j(\neq i)=1}^{k} m_{ij}^2 \geq 2k(k-2)$ for k even, and $\sum_{i=1}^{k}\sum_{j(\neq i)=1}^{k} m_{ij}^2 \geq 2\{k(k-2)+1\}$ for k odd.

When $N \equiv i$ (mod 4) ($i = 0, 1, 3$), the bounds follow on similar lines. □

Note that X attains the *MS*-optimality lower bound if the off-diagonal elements of X^TX are:

(i) 0, when $N \equiv 0$ (mod 4);

(ii) ± 2 for $k(k-2)/2$ elements, when $N \equiv 2$ (mod 4) and k is even;

(iii) ± 2 for $\{k(k-2)+1\}/2$ elements, when $N \equiv 2$ (mod 4) and k is odd; or

(iv) ± 1, when either $N \equiv 1$ (mod 4) or $N \equiv 3$ (mod 4).

For $i = 0, 1, 2, 3$, based on X_i, as given in Constructions(0)-(III), the off-diagonal elements of $X_i^TX_i$ satisfies the above structure and thus attains the *MS*-optimality lower bound. Therefore,

Theorem 2.9. *For $i = 0, 1, 2, 3$, a paired choice design d_i corresponding to the design matrix X_i is MS-optimal in $\mathcal{D}_{N,k,2}$.*

The D-optimal choice designs obtained by Graßhoff et al. (2004) and Demirkale et al. (2013) for $N \equiv 0 \pmod 4$ are also MS-optimal. To conclude, under the broader main effects model, we have provided MS-optimal paired choice designs for every N and k except $k = N \equiv 1 \pmod 4$, $k = N \geq 9$. For $k = N = 5$, §4 gives a design.

2.4 Comparing Designs with $m = 2$ and $m > 2$

Burgess and Street (2006) and Großmann and Schwabe (2015) have studied optimality aspects of choice designs with respect to choice set size. For two-level choice designs, Burgess and Street (2006) established that D-optimal choice designs are equivalent so long as the choice set sizes are multiples of 2. However, for their paired choice D-optimal designs, it is necessary that $N \equiv 0 \pmod 4$. Similarly, Großmann and Schwabe (2015) observed that if the number of levels is small, then sometimes using choice sets of size $m = 2$ may be better than using $m > 2$. We now broaden their results by considering optimal designs for all N, and show that designs with $m = 2$ are D-better than the best possible designs with $m = 3, 5$. For two designs, that with the bigger $\det(C_m)$ is said to be D-better than the other. Similarly, a design is MS-better than another design when compared with respect to the MS-criteria. Contrary to the results of Burgess and Street (2006), we first show that there could be designs with $m = 4$ which are better than a D-optimal design with $m = 2$ for situations where $N \not\equiv 0 \pmod 4$.

Consider two designs $d^2 \in \mathcal{D}_{5,5,2}$ and $d^4 \in \mathcal{D}_{5,5,4}$. Let $d^2 = \{(P \;\; p),\ (-P \;\; -p)\}$ and $d^4 = \{(P \;\; p),\ (-P \;\; -p),\ (P \;\; -p),\ (-P \;\; p)\}$, where P is a 5×4 matrix with -1 in the (i,i)th position $(i = 1, 2, 3, 4)$, and 1 elsewhere and $p = \begin{pmatrix} 1 & 1 & 1 & 1 & -1 \end{pmatrix}^T$. Here, d^2 is D- and MS-optimal in $\mathcal{D}_{5,5,2}$. Now, $\det(160C_4^{(2)})_{d^4} = \det(160C_4)_{d^4} = 2560 > 2304 = \det(160C_2)_{d^2} = \det(160C_2^{(2)})_{d^2}$, and thus $d^4 \in \mathcal{D}_{5,5,4}$ is D-better than $d^2 \in \mathcal{D}_{5,5,2}$. In fact, d^4 is also MS-better than d^2. Here, $\det(C)_d$ denotes the determinant of C corresponding to a choice design d.

Choice designs with $m > 5$ are expected to be less useful when there exist optimal designs with $m \leqslant 5$. We now concentrate on choice designs X_1, X_2, X_3, constructed in §2, and establish that they are D-better than the best possible designs with $m = 3$ and

$m = 5$. We show this in the next two theorems.

Theorem 2.10. *For $N \equiv i$ (mod 4), $i = 1, 2, 3$, and $k < N$, $d_i \in \mathcal{D}_{N,k,2}$ is D-better than the D-optimal design in $\mathcal{D}_{N,k,3}$, except possibly when $i = 2$ and $k = N - 1$.*

Proof. An upper bound to $\mathrm{tr}(C_m^{(2)})$ for m odd is $2^k \mathrm{tr}(C_m^{(2)}) \leqslant 2^k \mathrm{tr}(C_m) \leqslant k(m^2-1)/m^2$. Therefore, for $m = 3$ we have $\mathrm{tr}(C_3^{(2)}) \leq 2k/\{9(2^{k-2})\}$ and since the sum of the eigenvalues equals the trace of a matrix, $\det(C_3^{(2)}) \leqslant [2/\{9(2^{k-2})\}]^k$ or $\det(N2^k C_3^{(2)}) \leqslant (8N/9)^k$. To prove the result, it suffices to show that $f(k) = \det(N2^k C_2)_{d_i} - (8N/9)^k > 0$. For $N \equiv i$ (mod 4), we treat the three cases separately.

For $i = 1$, from Payne (1974), $\det(N2^k C_2)_{d_1} = (N-1+k)(N-1)^{k-1}$. Thus, it suffices to show that, for fixed N, $f(k) = (N-1+k)(N-1)^{k-1} - (8N/9)^k > 0$. For a proof by induction, first note that $f(1) = N/9 > 0$. Now, assuming that $f(k) > 0$, we prove that $f(k+1) = (N-1+k)(N-1)^k - (8N/9)^{k+1} > 0$. Since $f(k) = (N-1+k)(N-1)^{k-1} - (8N/9)^k$,

$$f(k+1) > \frac{\{(N+k)(N-1)^k\}\{(8N/9)^k\} - \{(8N/9)^{k+1}\}\{(N-1+k)(N-1)^{k-1}\}}{(N-1+k)(N-1)^{k-1}(8N/9)^k} = A.$$

Now,
$$A = \frac{(N+k)(N-1)^k}{(N-1+k)(N-1)^{k-1}} - \frac{(8N/9)^{k+1}}{(8N/9)^k} = \frac{N(N-1)+k(N-9)}{9(N-1+k)}.$$
Thus, $A > 0$, for $N \geq 9$. When $N = 5$, $A > 0$ for $k < 5$.

Case (ii) $i = 2$: From Payne (1974), for $k < N-1$ and the design d_2, $\det(N2^k C_2)_{d_2} = \{(N-2+k)^2 - \mu\}(N-2)^{k-2}$, where $\mu = 0$ if k is even and 1 otherwise. It suffices to show that, for fixed N, $f(k) = \{(N-2+k)^2 - \mu\}(N-2)^{k-2} - (8N/9)^k > 0$. Using induction we get, for k even,

$$f(k+2) > A = \frac{(N+k)^2(N-2)^2}{(N-2+k)^2} - \left(\frac{8N}{9}\right)^2,$$

and for k odd,
$$\frac{(N+k+1)(N-2)^2}{(N-3+k)} - \left(\frac{8N}{9}\right)^2.$$

To show $A > 0$, we may equivalently show that $9(N+k)(N-2) > 8N(N-2+k)$ for k even or $9(N-2) \geq 8N$ for k odd, which is always true for $N \geq 18$. Complete enumeration shows that the result holds for $N = 6, 10, 14; k < N-1$.

Case (iii) $i = 3$: From Payne (1974), $\det(N2^kC_2)_{d_3} = (N+1-k)(N+1)^{k-1}$. It suffices to show that, for fixed N, $f(k) = (N+1-k)(N+1)^{k-1} - (8N/9)^k > 0$. We first prove that the above inequality holds for $k \leq N-8$. Using induction, we get,

$$f(k+1) > A = \frac{N(N+1) - k(N+9)}{9(N+1-k)}.$$

Putting $k = N-8$, we see that $A = 8/9 > 0$ for any N. Thus, for $k \leq N-8$, the above inequality holds.

Now, for other cases, that is, $k = N-7, \ldots, N-1$. Let $k = N-\alpha$, $\alpha = 1, \ldots, 7$. Hence, $f(k) = f(N-\alpha) = 9^{N-\alpha}(\alpha+1)(N+1)^{N-\alpha-1} - 8^{N-\alpha}NN^{N-\alpha-1}$. Since $N+1 > N$, it suffices to show that for each k, $9^{N-\alpha}(\alpha+1) > 8^{N-\alpha}N$. It can be easily seen that $f_1(N,\alpha) = 9^{N-\alpha}(\alpha+1) - 8^{N-\alpha}N$ is an increasing function in N for $N \geq 23$ and $f_1(23,\alpha) > 0$ for $\alpha = 1, \ldots, 7$. Thus, for $k = N-7, \ldots, N-1$ and $N \geq 23$, the inequality holds. Complete enumeration shows that the result holds for $k < N; 7 \leq N \leq 19$. $\square$

Theorem 2.11. *For $N \equiv i$ (mod 4) ($i = 1, 2, 3$), and $k < N$, $d_i \in \mathcal{D}_{N,k,2}$ is D-better than the D-optimal design in $\mathcal{D}_{N,k,5}$, except possibly when (i) $i = 1$ and $k = N-1 = 4$; (ii) $i = 2$ and $k = N$ or $k = N-2 = 4, 8$ or $k = N-3 = 7$; (iii) $i = 3$ and $k = N-1 \leq 58$ or $k = N-2 \leq 41$ or $8 \leq k = N-3 \leq 28$ or $11 \leq k = N-4 \leq 15$.*

Proof. An upper bound to $\mathrm{tr}(C_5^{(2)})$ is obtained as in the proof of Theorem 2.10. For $m = 5$, we have $\mathrm{tr}(C_5^{(2)}) \leq 6k/\{25(2^{k-2})\}$ and thus, $\det(C_5^{(2)}) \leqslant [6/\{25(2^{k-2})\}]^k$ or $\det(N2^kC_5^{(2)}) \leqslant (24N/25)^k$. Thus, to prove the result, it suffices to show that $f(k) = \det(N2^kC_2)_{d_i} - (24N/25)^k > 0$.

For $N \equiv i$ (mod 4), we take up the three cases separately.

Case (i) $i = 1$: Working on lines similar to the proof of Theorem 2.10, we get $25(N-1+k)A = N(N-1) + k(N-25)$. Thus, $A > 0$, for $N \geq 25$. Complete enumeration shows that $A > 0$ for $k < N = 9, 13, 17, 21$ and for $N = 5$ with $k < 4$.

Case (ii) $i = 2$: Let $k < N-1$. Working on lines similar to the proof of Theorem 2.10, we only have to show that for k even, $N(N-2) + k(N-50) > 0$, and for k odd, $25(N-2) \geq 24N$. It is easy to see that the inequalities are always true for $N \geq 50$. For $N = 14$ through 42, since $k < N-1$, complete enumeration shows that the inequality holds. Also, for $N = 10$ with $k < 7$ and for $N = 6$ with $k < 4$, the above inequalities hold.

Case (iii) $i = 3$: We first prove that the inequality holds for $k \leq N - 24$ and then treat the remaining cases $k = N - \alpha$, $\alpha = 1, \ldots, 23$ separately. Working on lines similar to the proof of Theorem 2.10, we get $25(N + 1 - k)A = N(N + 1) - k(N + 25)$. Putting $k = N - 24$, we see that $A = 24/25 > 0$ for any N. Thus, we have shown that for $k \leq N - 24$, the above inequality holds. Now, we take up the other cases, that is, $k = N - 23, \ldots, N - 1$. Let $k = N - \alpha$, $\alpha = 1, \ldots, 23$. Then on lines similar to the proof of Theorem 2.10, we see that the inequality holds for $N \geq 59$ and $k = N - 23, \ldots, N - 2$ and for $N \geq 63$ and $k = N - 1$. Complete enumeration for remaining $N \leq 59$ shows that $(N + 1 - k)(N + 1)^{k-1} > (24N/25)^k$ for all N and $k < N$ except when (i) $N \leq 59$, $k = N - 1$ (ii) $N \leq 43$, $k = N - 2$ (iii) $11 \leq N \leq 31$, $k = N - 3$ and (iv) $15 \leq N \leq 19$, $k = N - 4$. □

2.5 Concluding Remarks

The *D*- and *MS*-optimal two-level paired choice designs found in this chapter provide solutions in situations where, for every $N \not\equiv 0 \pmod 4$, the information matrix of an optimal exact design is different from the information matrix of the optimal approximate design, for which the corresponding exact optimal design would not be available. However, *D*-optimal design constructions for situations as mentioned in Remark 2.7, and *MS*-optimal designs for $k = N \equiv 1 \pmod 4$ can be further explored. This work complements previous work giving optimal exact designs only for $N \equiv 0 \pmod 4$. Thus experimenters can now use optimal designs for any number of choice sets N. Designs in this chapter are optimal for estimating the main effects under a broader model containing all two-factor interactions, which is more realistic in practice. From a statistical perspective we have established that one should prefer optimal paired choice designs to choice designs with $m = 3$ or $m = 5$. This also adds in achieving the desired quality of response through reduced choice set size.

Chapter 3

Optimal paired choice block designs

This chapter is based on the following work:

Singh et al. (2018): Singh, Rakhi; Das, Ashish; Chai, Feng-Shun. Optimal Paired Choice Block Designs. Stat. Sinica (2018), *accepted*, doi: 10.5705/ss.202016.0084.

3.1 Introduction

In a choice experiment, respondents are shown multiple choice sets of options and from each set they choose the preferred option. Considering choice sets of size two and r given respondents, a paired choice experiment is usually perceived as showing the same set of N choice pairs to each of the r respondents. The respondents are asked to give their preference among the two options for each of the N choice pairs shown to them.

D-optimal designs have been obtained theoretically under the utility-neutral setup, for example, see Graßhoff et al. (2003), Graßhoff et al. (2004), Street and Burgess (2007), Street and Burgess (2012), Demirkale, Donovan and Street (2013), Bush (2014), Großmann and Schwabe (2015) and Singh, Chai and Das (2015). In contrast, in the locally-optimal and the Bayesian approach, D-optimal designs have been obtained using computer algorithms (see, Huber and Zwerina (1996), Sándor and Wedel (2001), Sándor and Wedel (2002), Sándor and Wedel (2005), Kessels, Goos and Vandebroek (2006), Kessels, Goos and Vandebroek (2008), Kessels, Jones, Goos and Vandebroek (2008), Kessels et al. (2009), Yu, Goos and Vandebroek (2009)). In this chapter, we follow the utility-neutral approach.

Traditionally, in a choice experiment, respondents are shown the same collection of

N choice pairs under the assumption that the respondents are alike. A choice experiment with the inherent premise that the respondents are alike is not quite practical since respondents, being a random sample from a population, are more likely to be heterogeneous. Kessels, Goos and Vandebroek (2008) also noted that heterogeneity leads to responses from different respondents being different.

In a paired choice experiment, there is always a constraint on the maximum number of choice pairs that can be shown to each respondent so as to maintain overall response quality. A major concern with the traditional optimal paired choice designs is that the number of choice pairs in the design increases rapidly as k and/or v_i's are moderately increased.

Attempts have been made to address the issue of heterogeneity through different models and approaches. Sándor and Wedel (2002) have addressed the heterogeneity in respondents by constructing designs through a computer-intensive algorithmic approach under the so called mixed logit model. In their approach, same set of N choice pairs are shown to every respondent. Subsequently, Sándor and Wedel (2005) demonstrated that the use of different choice designs for different respondents and the random allocation of respondents to these designs yields substantially higher efficiency than the designs obtained in Sándor and Wedel (2002). Later Kessels, Goos and Vandebroek (2008), for catering to heterogeneity in conjoint experiments, introduced a random respondent effects model for estimating the main effects and used algorithmic methods for constructing D-optimal designs. The conjoint designs under their setup consists of identifying as many sets of options as there are respondents. Therefore, the approach, though similar, is not applicable to our setup.

Often in practice, there is a pool of choice sets and respondents are allocated a random subset of choice sets (Street and Burgess, 2007). This process is continued until all choice sets are used once. Thereafter the process is started again. To address the *ad hoc* approach in the random allocation of choice sets, we use an additional fixed-effect term in the model to systematically split the pool of choice sets. In experimental design theory, the concept of blocking, as a tool to eliminate systematic heterogeneity in the experimental material, has been used extensively. Following the same approach, we consider the respondents as blocks. Thus, in contrast to the computer-intensive algorithmic approaches of Sándor and Wedel (2005) and Kessels, Goos and Vandebroek (2008), we treat the respondent heterogeneity as a nuisance factor by including respondent-level block effect terms in the model and then design experiments to optimally estimate the parameters of interest after eliminating the respondent (block) effects. Adopting such an approach also enables the experimenter to get optimal designs with reasonable number of choice pairs $s(< N)$ shown to each of the r respondents. Later in Section 3.2, we discuss the kind of heterogeneity that is being taken care of in our approach and the seemingly similar approaches.

In what follows, a design with b blocks each of size s is generated and that each block is associated to a respondent. Usually t copies of a proposed design is used for larger numbers of respondents $r = tb$, since replicating the design does not affect its optimality. We therefore, restrict ourselves to optimal paired choice block designs with b blocks each of size s with $N = bs$.

In this context, the traditional paired choice designs reduce to $b = 1, s = N$ and $r = t$ where s is necessarily atleast the number of model parameters. However, for $b > 1$, the block size s can be smaller than the number of model parameters, but the paired choice design with b blocks can still estimate all model parameters. In order to estimate the model parameters, we provide optimal designs with block sizes that are flexible and practical under our setup.

In Section 3.2, treating respondent heterogeneity as a nuisance factor and incorporating the fixed respondent (block) effects in the model, we obtain the information matrix for estimating the parameters of interest after eliminating the respondent (block) effects. In Section 3.3, under the main effects block model, we provide optimal paired choice block designs for estimating the main effects for symmetric and asymmetric attributes. We also give a simple solution to the problem of identifying generators in the constructions of optimal paired choice designs. In Section 3.4, under a broader main effects block model, we provide optimal paired choice block designs for symmetric and asymmetric attributes. The broader main effects model constitutes the main effects and the two-factor interaction effects with interest lying only in the estimation of the main effects. Finally, in Section 3.5, we provide optimal paired choice block designs for estimating the main plus two-factor interaction effects. Finally, we provide a Discussion in Section 3.6.

3.2 Preliminaries and the model incorporating respondent effects

Most of the work on optimal choice designs is based on the multinomial logit model approach of either Huber and Zwerina (1996) or that followed in Street and Burgess (2007). Großmann and Schwabe (2015) observed that the two approaches are equivalent for the purpose of finding optimal designs. We work with the multinomial logit model approach of Huber and Zwerina (1996). The multinomial logit model supposes that the probability of preferring option 1 over option 2 in the ith choice pair can be expressed as $\pi_{12i} = e^{u_{1i}}/(e^{u_{1i}} + e^{u_{2i}})$, where u_{1i} and u_{2i} represent the systematic part of the utilities attached to the two options in choice pair i. Similarly $\pi_{21i} = 1 - \pi_{12i}$ is the probability that option 2 is preferred over option 1. It follows that for the ith choice pair, the choice probabilities depend only on the utility difference $u_{1i} - u_{2i}$. For a design d with N choice pairs, since options are described by k attributes, the utilities are modeled using the linear predictor $u_j = P_{pj}\theta$, where θ is a $p \times 1$ vector representing the parameters of interest, P_{pj} is an $N \times p$ effects-coded matrix for the jth option, and $u_j = (u_{ji})$ is an $N \times 1$ utility vector for the jth option, $j = 1, 2$. The utility difference $u_1 - u_2 = (P_{p1} - P_{p2})\theta = P_p\theta$ is then a linear function of the parameter vector θ. For the purpose of deriving optimal designs, it is often assumed that $\theta = 0$. This indifference or the utility-neutral assumption means that the two options in a choice set are equally attractive and leads to a considerable simplification of the information matrix and the design problem. Under the utility-neutral multinomial logit model, the Fisher information matrix is $(1/4)P_p'P_p$ (see, Großmann and Schwabe (2015)).

Simultaneously, Graßhoff et al. (2003) and Graßhoff et al. (2004) studied linear paired comparison designs which are analyzed under the linear paired comparison model. The observed utility difference Z between the two options again depends on the difference matrix $P_p = P_{p1} - P_{p2}$. More precisely, the response is described by the model, $Z = u_1 - u_2 + \epsilon = (P_{p1} - P_{p2})\theta + \epsilon = P_p\theta + \epsilon$, where ϵ is the random error vector. The matrix $C = P_p'P_p$ is the information matrix under the linear paired comparison model. Since C is proportional to the information matrix under the utility neutral multinomial logit model, it follows that the designs optimal under the linear paired comparison model are also optimal under the multinomial logit model and vice versa.

We discuss only D-optimality since, as noted in Großmann and Schwabe (2015), most of the optimality results for choice designs and linear paired comparison designs are available for the D-criterion. A D-optimal design has the maximum determinant of the information matrix among all competing designs.

For paired choice experiments, the multinomial logit model as well as the linear paired comparison model are based on the utility difference $u_1 - u_2$. By incorporating respondent effects, the relevant utility differences under the block model, with blocks being the respondents, becomes

$$u_1 - u_2 = (P_{p1} - P_{p2})\theta + W\beta = P_p\theta + W\beta, \tag{3.1}$$

where $\beta = (\beta_1, \ldots, \beta_b)'$ represents the $b \times 1$ vector of block effects, and $W = (w_{ij})$ is an $N \times b$ incidence matrix with $w_{ij} = 1$ if the ith choice pair belongs to the jth block and 0 otherwise. Without loss of generality, we take $W = I_b \otimes 1_s$, where I_a and 1_a denotes the identity matrix of order a and the $a \times 1$ vector of all ones, respectively. Here, $\otimes$ denotes the Kronecker product. Note that (3.1) corresponds to a paired choice block design with b blocks each of size s and that such b blocks are repeated t times to accommodate for $r = tb$ respondents. Each of the r respondents is associated to a single block of the design.

Unlike Sándor and Wedel (2005) and Kessels, Goos and Vandebroek (2008), where an assumed distribution on the model parameters takes care of the respondent effects, our approach, following the standard block design theory, has been to consider β_j as a fixed-effects term. While the vast literature on theoretically obtained D-optimal designs for choice experiments rests on a multinomial logit model without any respondent effects, our fixed-effects block model attempts to obtain the optimal block designs theoretically under the utility-neutral setup.

In either the multinomial logit model or the linear paired comparison model, including respondent effects β can be regarded as adding b two-level attributes to the set of p predictor variables. Then, the corresponding difference matrix for the pairs, in b blocks, has an additional component and can be written as (P_p, W). Thus, under the utility-neutral multinomial logit block model, it follows that the information matrix for estimating θ and β is

$$M = \frac{1}{4}\begin{bmatrix} C & P_p'W \\ W'P_p & W'W \end{bmatrix} \tag{3.2}$$

where $C = P_p'P_p$, as defined earlier. Moreover, upto a constant factor of $1/4$, M coincides with the information matrix in the linear paired comparison block model. Thus, optimal designs under the linear paired comparison block model are also optimal under the utlity neutral multinomial logit block model. The information matrix for estimating θ under the linear paired comparison block model after eliminating the block effects is

$$\tilde{C} = C - P_p'W(W'W)^{-1}W'P_p = C - (1/s)P_p'WW'P_p. \tag{3.3}$$

This follows from the standard linear model theory where a parameter vector is partitioned into a parameter vector of interest and the nuisance parameters (see, for example, Page 68 of Haines (2015)).

A paired choice block design is connected if all the parameters of interest are estimable, and this happens if and only if $\tilde{C}$ has rank p. In what follows, the class of all connected paired choice block designs with k attributes in b blocks each of size s is denoted by $\mathcal{D}_{k,b,s}$. From (3.3), since $C - \tilde{C}$ is a non-negative definite matrix, if in the class of unblocked designs with $N = bs$, a paired choice design d is D-optimal, then d, considered as a design in $\mathcal{D}_{k,b,s}$, is also D-optimal, provided $\tilde{C} = C$.

It is observed that eliminating respondent effects simultaneously controls the within-pair order effects (see, Goos and Großmann (2011) and Bush, Street and Burgess (2012)).

3.3 Optimal block designs under the main effects model

Under the main effects block model, from (3.1) it follows that $u_1 - u_2 = (P_{M1} - P_{M2})\tau + W\beta = P_M\tau + W\beta$, where τ is a $\sum_{i=1}^{k}(v_i - 1) \times 1$ parameter vector for main effects, P_{Mj} is an $N \times \sum_{i=1}^{k}(v_i - 1)$ effects-coded matrix of the main effects for the jth option, $j = 1, 2$, and $P_M = P_{M1} - P_{M2}$. In a row of P_{Mj} is embedded the effects-coded row vector of length $v_i - 1$ for the ith attribute. The effects coding for level l is represented by a unit vector with 1 in the $(l + 1)$th position for $l = 0, \ldots, v_i - 2$, and for level $v_i - 1$ is represented by -1 in each of the $v_i - 1$ positions, $i = 1, \ldots, k$. For example, for $v = 3$, effects-coded vectors for $l = 0, 1, 2$ are $(1\ 0), (0\ 1)$ and $(-1\ \ -1)$, respectively.

From (3.3), the information matrix for estimating the main effects after eliminating the block effects is

$$\tilde{C}_M = C_M - (1/s)P_M'WW'P_M, \tag{3.1}$$

where $C_M = P_M' P_M$ is the information matrix for estimating the main effects under the unblocked model. From (3.1), it follows that a necessary and sufficient condition for $\tilde{C}_M = C_M$ to hold is $W' P_M = 0$. Therefore, by suitably blocking the choice pairs of an optimal paired choice design into b blocks such that $W' P_M = 0$, one can obtain an optimal paired choice block design. We provide a simple condition to achieve the same, proof of which is provided in the Appendix A.

Theorem 3.1. *$\tilde{C}_M = C_M$ if for each block, the levels of every attribute appear equally often in the first option as well as in the second option.*

This property of every level of an attribute appearing the same number of times in the first and second option of pairs is also known as *position-balance* (see, Großmann and Schwabe (2015)).

An orthogonal array $OA(n, k, v_1 \times \cdots \times v_k, t)$, of strength t, is an $n \times k$ array with elements in the ith column from a set of v_i distinct symbols $\{0, 1, \ldots, v_i - 1\}$ $(i = 1, \ldots, k)$, such that all possible combinations of symbols appear equally often as rows in every $n \times t$ subarray. An orthogonal array is symmetric if $v_i = v$ for all i and the corresponding OA is denoted by $OA(n, k, v^k, t)$, else it is an asymmetric orthogonal array.

Street and Burgess (2007), Demirkale, Donovan and Street (2013) and Bush (2014) provide the $OA + G$ method for constructing optimal paired choice designs using orthogonal arrays and generators G. Let G be a collection of h generators $G_1, \ldots, G_h$ where $G_j = (g_{j_1}, g_{j_2}, \ldots, g_{j_k})$. The $OA + G$ method gives a paired choice design $(A, B_j), j = 1, \ldots, h$ where $A = (A_{li})$ is an $OA(n_1, k, v_1 \times \cdots \times v_k, t)$ and $B_j = (B_{li}^j)$ with $B_{li}^j = A_{li} + g_{ji}$ reduced mod v_i, $l = 1, \ldots, n_1$, $i = 1 \ldots, k$, $j = 1, \ldots, h$. This method depends on the availability of the required orthogonal array, which may not always exist. The SAS link `http://support.sas.com/techsup/technote/ts723.html`, the Sloane link `http://neilsloane.com/oadir/` and Hedayat, Sloane and Stufken (1999) provide a comprehensive summary of orthogonal arrays and their constructions.

In the literature, arriving at the generators G has been usually through a trial-and-error approach, and no general results on the structure of such generators appear to exist. In fact, Bush (2014) highlights the complexities involved in choosing the sets of generators. We present a simple result that systematically provides the h generators, proof of which is provided in the Appendix A. Let $lcm(a_1, \ldots, a_k)$ denotes the least common multiple of $a_1, \ldots, a_k$.

Theorem 3.2. *Number of generators for the optimal paired choice design with k attributes is* $h = lcm(h_1, \ldots, h_k)$ *where* $h_i = v_i - 1$ *for* v_i *even and* $h_i = (v_i - 1)/2$ *for* v_i *odd,* $i = 1, \ldots, k$. *The generators are then given by* $G_j = (g_{j_1}, g_{j_2}, \ldots, g_{j_k})$, *where* g_{j_i} *takes each of the values from the set* $\{1, \ldots, h_i\}$ *with frequency* h/h_i, $j = 1, \ldots, h$, $i = 1, \ldots, k$.

Note that Theorem 3.2 provides generators for the unblocked paired choice designs. As in Street and Burgess (2007), we use several sets of generators to create the final design and that the number of generators given in Theorem 3.2 may not be the smallest possible.

Example 3.3. *Suppose there are three attributes with* $v_1 = 2, v_2 = 3$ *and* $v_3 = 4$. *Then we have* $h_1 = 1$, $g_{j_1} = 1$; $h_2 = 1$, $g_{j_2} = 1$; *and* $h_3 = 3$, $g_{j_3} = 1, 2, 3$. *Thus,* $h = lcm(1, 1, 3) = 3$. *This leads to the generators* $G_1 = (111), G_2 = (112)$ *and* $G_3 = (113)$. *Thus, for a given* $OA(24, 3, 2 \times 3 \times 4, 2)$, *the corresponding optimal paired choice design with parameters k,* $v_1 = 2, v_2 = 3, v_3 = 4$, $b = 1$, $N = s = hn_1 = 3 \times 24 = 72$, *is obtained using the* $OA + G$ *method of construction with three generators. The corresponding design is given in the Appendix A.*

Example 3.4. *As another example, suppose there are two attributes with* $v_1 = 4$ *and* $v_2 = 5$. *Then we have* $h_1 = 3$, $g_{j_1} = 1, 2, 3$ *and* $h_2 = 2$, $g_{j_2} = 1, 2$. *Thus,* $h = lcm(3, 2) = 6$. *This leads to the six generators* $G_1 = (11), G_2 = (12)$, $G_3 = (21)$, $G_4 = (22)$, $G_5 = (31)$ *and* $G_6 = (32)$ *which will give an optimal paired choice design when used in conjunction with* $OA(20, 2, 4 \times 5, 2)$.

In general, for a given $OA(n_1, k, v_1 \times \cdots \times v_k, 2)$, the corresponding optimal paired choice design d_1 with parameters k, $v_1, \ldots, v_k$, $b = 1$, $N = s = hn_1$, is obtained using the $OA+G$ method of construction with generators $G_j, j = 1, \ldots, h$. When $N = s$ is large, we find that practitioners advocate allocation of the choice pairs into more than one blocks either randomly or using a spare attribute (see, Street and Burgess (2007), Bliemer and Rose (2011)). Based on Theorem 3.1, it follows that under our block model, we can retain optimality of the design obtained through the $OA + G$ method if blocking is done using a column corresponding to an attribute. Any other blocking approach may jeopardize the characteristics of the design. We now provide four theorems and their constructions, detailed proofs of which are provided in the Appendix A.

Theorem 3.5. *For $\delta \geq 1$ and an $OA(n_1, k+1, v_1 \times \cdots \times v_k \times \delta, 2)$, there exists an optimal paired choice block design $d_2 \in \mathcal{D}_{k,b,s}$ with parameters k, $v_1, \ldots, v_k$, $b = h\delta$, $s = n_1/\delta$, where $h = lcm(h_1, \ldots, h_k)$.*

Construction. *For a given $OA(n_1, k+1, v_1 \times \cdots \times v_k \times \delta, 2)$, corresponding to the k attributes at levels $v_i, i = 1, \ldots, k$, let d_1 be the design constructed through $OA+G$ method using $h = lcm(h_1, \ldots, h_k)$ generators from Theorem 3.2. Then d_1 with parameters k, $v_i, i = 1, \ldots, k$, $b = 1$, $s = hn_1$ is an optimal paired choice design. From d_1, the choice pairs obtained through each of the h generators constitute a block of size n_1. Finally, we use the δ symbols of the $(k+1)$th column of the orthogonal array for further blocking. This gives us a paired choice block design d_2 with parameters k, $v_1, \ldots, v_k$, $b = h\delta$, $s = n_1/\delta$.*

Example 3.6. *From an $OA(24, 15, 2^{13}\times 3\times 4, 2)$, for estimating the main effects of $k = 14$ attributes of which 13 attributes are at 2 levels and 1 attribute is at 3 levels, an optimal paired choice block design can be constructed for $\delta = 4, h = 1, k = 14, b = 4, s = 6$ are optimal. As an illustration, we give a $2^4 \times 3$ paired choice block design d_2 with parameters $k = 5, b = 4, s = 6$.*

$d_2 =$

B_1	B_2	B_3	B_4
(00000, 11111)	(01102,10010)	(10112,01000)	(10001,01112)
(11010,00101)	(11110,00001)	(00111,11002)	(00012,11100)
(01101,10012)	(11011,00102)	(01002,10110)	(10100,01011)
(11002,00110)	(00100,11011)	(11101,00012)	(01011,10102)
(10111,01002)	(10012,01100)	(01010,10101)	(01110,10001)
(00112,11000)	(00001,11112)	(10000,01111)	(11102,00010)

It is noted that when the attributes have mixed levels greater than 3, the $OA+G$ method leads to choice designs with a large number of choice pairs. However, blocking still helps in reducing the number of choice pairs shown to respondent from $N = s = 96$ to $s = 24$. For example, an $OA(32, 11, 2^3 \times 4^7 \times 8, 2)$ can be used to construct a paired choice block design having three 2-level attributes and seven 4-level attributes in $N = 96$ choice pairs with $b = 24$ and $s = 4$.

For many parameter sets corresponding to k attributes each at v levels, Graßhoff et al. (2004) and Demirkale, Donovan and Street (2013) have provided constructions of optimal paired choice designs with a reduced number of choice pairs in comparison to the $OA+G$ method of construction. We now show how an optimal paired choice block design can be constructed starting from their designs.

Theorem 3.7. *For a Hadamard matrix H_m, an optimal paired choice design d_3 with parameters $k, v, b = 1, s = mv(v-1)/2$, $k \leqslant m$ exists. Furthermore for v odd, a paired choice block design d_4 with parameters $k, v, b = m(v-1)/2, s = v$ exists, which is optimal in $\mathcal{D}_{k,b,s}$.*

Construction. *For a given H_m, an optimal paired choice design d_3 is obtained through Theorem* 3 *of Graßhoff et al. (2004) with parameters $k, v, b = 1, s = mv(v-1)/2$. Moreover, for v odd, the choice pairs corresponding to each of the rows of $\{H_m, -H_m\}$ forms a block and the design so obtained is an optimal paired choice block design. Now, using a result from Dey (2009), $v(v-1)/2$ combinations involving v levels taken two at a time can be grouped into $(v-1)/2$ replicate each comprising v elements. Therefore, the blocks generated by each row of H_m can be further broken into $(v-1)/2$ blocks each of size v, which gives the optimal paired choice block design d_4.*

Example 3.8. *Consider $v = 3$ with combinations $(0,1), (1,2), (2,0)$ and the Hadamard matrix H_4. An optimal paired choice design d_3 with parameters $k = 4, v = 3, b = 1, s = 12$ exists. Furthermore, since v is odd, an optimal paired choice block design d_4 is constructed with parameters $k = 4, v = 3, b = 4, s = 3$ by considering choice pairs generated by each row of $\{H_4, -H_4\}$ as a block.*

$d_4 =$

B_1	B_2	B_3	B_4
(0000,1111)	(0101,1010)	(0011,1100)	(0110,1001)
(1111,2222)	(1212,2121)	(1122,2211)	(1221,2112)
(2222,0000)	(2020,0202)	(2200,0022)	(2002,0220)

Theorem 3.9. *For an $OA(n_2, k+1, v^k \times v_{k+1}, 2)$ with $v_{k+1} = n_2/v$, an optimal paired choice design d_5 with parameters $k, v, b = 1, s = n_2(v-1)/2$ exists. Furthermore for v odd, a paired choice block design d_6 with parameters $k, v, b = n_2(v-1)/2v, s = v$ exists, which is optimal in $\mathcal{D}_{k,b,s}$.*

Construction. *For a given $OA(n_2, k+1, v^k \times v_{k+1}, 2)$ with $v_{k+1} = n_2/v$, an optimal paired choice design d_5 is obtained through Construction 3.2 of Demirkale, Donovan and Street (2013) with parameters $k, v, b = 1, s = v_{k+1}\binom{v}{2}$. Moreover, for v odd, the choice pairs corresponding to each of the parallel sets of the orthogonal array forms a block and the design so obtained is an optimal paired choice block design. Now, following Dey (2009), the blocks generated by each parallel set can be further broken into $(v-1)/2$ blocks each of size v, which gives the optimal paired choice block design d_6.*

Theorem 3.10. *For $\delta \geq 1$ and an $OA(n_3, k+1, m_1 \times \cdots \times m_k \times \delta, 2)$ with $m_i = v_i(v_i-1)/2$ for some odd v_i, an optimal paired choice block design d_8 with parameters $k, v_i, \ldots, v_k, b = \delta, s = n_3/\delta$ exists.*

Construction. *For a given $OA(n_3, k+1, m_1 \times \cdots \times m_k \times \delta, 2)$ with $m_i = v_i(v_i-1)/2$ for some odd v_i, an optimal paired choice design d_7 is obtained through Theorem 4 of Graßhoff et al. (2004) with parameters $k, v_i, \ldots, v_k, b = 1, s = n_3$. Then, similar to construction of Theorem 3.5, we use the δ (≥ 1) symbols of the $(k+1)$th column of the orthogonal array for blocking. This gives us an optimal paired choice block design d_8 with parameters $k, v_i, \ldots, v_k, b = \delta, s = n_3/\delta$. Note that this method of blocking is applicable only for odd v_i.*

Table 3.1 highlights the flexibility in the number of blocks while blocking the traditional optimal symmetric paired choice designs as listed in Table 2 of Demirkale, Donovan and Street (2013). We list the values of s and b corresponding to the optimal designs obtained through Theorem 3.5 and Theorem 3.7. It is observed that in the parameter range of Table 3.1, Theorems 3.9 and 3.10 do not provide any additional designs that are not obtainable from Theorem 3.5 and Theorem 3.7. Some of the traditional optimal paired choice designs, marked $*$, are not optimal under the block setup for blocks of size $s = N$ and $b = 1$ since the design matrices are not orthogonal to the vector of all ones. However, by having $b > 1$, optimal designs having blocks of size $s = N/b$ are feasible using Theorem 3.5.

Note that, from a given optimal paired choice design in $\mathcal{D}_{k,b,s}$, we can randomly group the b blocks into b/x blocks each of size xs to obtain optimal paired choice designs in $\mathcal{D}_{k,b/x,sx}$. In Table 3.1, the designs with $x = 1$ are first obtained using the Theorems as mentioned in the corresponding column headers whereas the designs with $x > 1$ are obtained thereafter through random grouping. One could obtain a table similar to Table 3.1, for optimal asymmetric paired choice designs based on a list of more than 600 orthogonal arrays with $n \leq 100$.

Table 3.1: Optimal designs in $\mathcal{D}_{k,b,s}$

v	k	Traditional (s,1)	Theorem 3.5 (s,b)	Theorem 3.7 (s,b)
2	3	4	(4,1)	
2	4	4*	(4x,2/x), x=1,2	
2	5-6	8	(4x,2/x), x=1,2 (6x,2/x), x=1,2	
2	7	8	(8,1) (6x,2/x), x=1,2 (4x,4/x), x=1,2,4	
2	8	8*	(6x,2/x), x=1,2 (4x,4/x), x=1,2,4	
2	9-10	12	(6x,2/x), x=1,2 (4x,4/x), x=1,2,4 (10x,2/x), x=1,2	
2	11	12	(12,1) (4x,4/x), x=1,2,4 (10x,2/x), x=1,2 (6x,4/x), x=1,2,4	
2	12	12*	(4x,4/x), x=1,2,4 (10x,2/x), x=1,2 (6x,4/x), x=1,2,4	
3	3	9,12	(3x,3/x), x=1,3	(3x,4/x), x=1,2,4
3	4	9,12,18	(9,1) (3x,6/x), x=1,2,3,6	(3x,4/x), x=1,2,4
3	5,6	18,24	(3x,6/x), x=1,2,3,6	(3x,8/x), x=1,2,4,8
3	7	18,24,27	(9x,2/x), x=1,2 (3x,9/x), x=1,3,9	(3x,8/x), x=1,2,4,8
3	8	24,27	(3x,9/x), x=1,3,9	(3x,8/x), x=1,2,4,8
3	9	27,36	(3x,9/x), x=1,3,9	(3x,12/x), x=1,2,3,4,6,12
3	10-12	27,36	(9x,3/x), x=1,3 (3x,12/x), x=1,2,3,4,6,12	(3x,12/x), x=1,2,3,4,6,12
4	3-4	24*,28	(4x,12/x), x=1,2,3,4,6,12	
4	5	48	(16x,3/x), x=1,3 (4x,24/x), x=1,2,3,4,6,8,12,24	
4	6-8	48*,96	(4x,24/x), x=1,2,3,4,6,8,12,24	
4	9	72*,96	(16x,6/x), x=1,2,3,6 (4x,36/x), x=1,2,3,4,6,9,12,18,36	
4	10-12	72*,144	(4x,36/x), x=1,2,3,4,6,9,12,18,36	
5	3-4	40,50	(5x,10/x), x=1,2,5,10	(5x,8/x), x=1,2,4,8
5	5	50,80	(5x,10/x), x=1,2,5,10	(5x,16/x), x=1,2,4,8,16
5	6	50,80,100	(25x,2/x), x=1,2 (5x,20/x), x=1,2,4,5,10,20	(5x,16/x), x=1,2,4,8,16
5	7-8	80,100	(5x,20/x), x=1,2,4,5,10,20	(5x,16/x), x=1,2,4,8,16
5	9-10	100,120	(5x,20/x), x=1,2,4,5,10,20	(5x,24/x), x=1,2,3,4,6,12,24
6	3	60*,180	(12x,15/x), x=1,3,5,15 (18x,10/x), x=1,2,5,10	
6	4	60*,180*,360	(6x,60/x), x=1-6,10,12,15,20,30,60	
6	5-6	120*,180*,360	(6x,60/x), x=1-6,10,12,15,20,30,60	
7	3-4	84,147	(7x,21/x), x=1,3,7,21	(7x,12/x), x=1,2,3,4,6,12
7	5-7	147,168	(7x,21/x), x=1,3,7,21	(21x,8/x), x=1,2,4,8
7	8	147,168,294	(49x,3/x), x=1,3 (7x,42/x), x=1,2,3,6,7,14,21,42	(21x,8/x), x=1,2,4,8

3.4 Optimal block designs under the broader main effects model

In this section, we consider estimation of the main effects under the broader main effects model for an asymmetric paired choice design where the ith attribute is at v_i levels, $i = 1, \ldots, k$. The broader main effects model constitutes the main effects and the two-factor interaction effects with interest lying only in the estimation of the main effects. For the symmetric paired choice designs, Graßhoff et al. (2003) characterized optimal paired choice designs under the broader main effects model. More recently, for $v_i = 2$, Singh, Chai and Das (2015) obtained optimal designs under such a model.

With the introduction of the respondent effects, from (3.1), the relevant utility differences become

$$u_1 - u_2 = (P_{M1} - P_{M2})\tau + (P_{I1} - P_{I2})\gamma + W'\beta = P_M\tau + P_I\gamma + W'\beta, \qquad (3.1)$$

where γ is a $\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(v_i-1)(v_j-1)\times 1$ parameter vector for the two-factor interaction effects, P_{Ij} is an $N \times \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(v_i-1)(v_j-1)$ effects-coded matrix of the two-factor interaction effects for the jth option, $j = 1, 2$, and $P_I = P_{I1} - P_{I2}$. Let $P_{Ij} = (P_{Ij}^{1'}, \ldots, P_{Ij}^{n'})'$ where P_{Ij}^{l} corresponds to the lth choice pair in P_{Ij}. Also, let $P_{Mj(i)}^{l}$ represent the columns of P_{Mj} corresponding to the lth choice pair and ith attribute. Then, $P_{Ij}^{l} = (P_{Mj(1)}^{l} \otimes P_{Mj(2)}^{l}, P_{Mj(1)}^{l} \otimes P_{Mj(3)}^{l}, \ldots, P_{Mj(k-1)}^{l} \otimes P_{Mj(k)}^{l})$.

The information matrix for estimating the main effects after eliminating the two-factor interaction effects and the block effects is

$$\tilde{C}_B = C_M - [P_M'P_I \; P_M'W]\begin{bmatrix} P_I'P_I & P_I'W \\ P_I'W & W'W \end{bmatrix}^{-}\begin{bmatrix} P_I'P_M \\ W'P_M \end{bmatrix}. \qquad (3.2)$$

Therefore, a paired choice design which is optimal under the main effects model is also optimal under the broader main effects block model if $\tilde{C}_B = C_M$, that is, if $P_I'P_M = 0$ and $W'P_M = 0$. The designs in Theorem 3.5 satisfy $W'P_M = 0$ and for symmetric designs with $v = 2$, it follows from Singh, Chai and Das (2015) that the designs additionally satisfy $P_I'P_M = 0$. Therefore, in particular, for symmetric designs with $v = 2$, the paired choice block designs of Theorem 3.5 are also optimal under the broader main effects block model.

We now give the following construction for optimal paired choice block designs under the broader main effects model.

Theorem 3.11. *Under the broader main effects model, for an $OA(n_1, k, v_1 \times \cdots \times v_k, 3)$ and $h = lcm(v_1, \ldots, v_k)$, there exists a paired choice block design d_1^B with parameters $k, v_1, \ldots, v_k, b = 1, s = hn_1$, which is optimal in $\mathcal{D}_{k,b,s}$.*

Construction. *We obtain d_1^B through the $OA + G$ method of construction using h generators as in Theorem 3.2. Detailed proof is provided in the Appendix A.*

Theorem 3.12. *Under the broader main effects model, for $\delta \geq 1$ and an $OA(n_1, k+1, v_1 \times \cdots \times v_k \times \delta, 3)$, there exists a paired choice block design d_2^B with parameters $k, v_1, \ldots, v_k, b = h\delta, s = n_1/\delta$, which is optimal in $\mathcal{D}_{k,b,s}$.*

Construction. *On lines similar to Theorem 3.5, the construction here is based on using sets of generators, from Theorem 3.2, on an orthogonal array of strength* 3.

We now provide another method to obtain symmetric optimal paired choice block designs with $s = v; v \geq 3$.

Theorem 3.13. *For an $OA(n_1, k-1, v^{k-1}, 3)$, there exists a paired choice block design d_3^B with parameters $k, v \geq 3, s = v, b = hn_1$, which is optimal in $\mathcal{D}_{k,b,s}$.*

Construction. *We adopt the following method of construction.*
(i) Following Theorem 3.12, construct d_2^B from an $OA(n_1, k, v^{k-1} \times 1, 3)$ for $k-1$ attributes each at v levels. While constructing d_2^B, the h generators, as in Theorem 3.2, are $(k-1)$-tuples of the form $(1 \ldots 1), \cdots, (v-1 \ldots v-1)$ for v even ($h = (v-1)$), and of the form $(1 \ldots 1), \cdots, ((v-1)/2 \ldots (v-1)/2)$ for v odd ($h = (v-1)/2$). Then, for each choice pair, add the kth attribute at level 0 in the option 1 and similarly, the kth attribute in the second option is generated using the same generator as that used for the other $k-1$ attributes.

(ii) For each of the h generators, generate $v-1$ additional copies of the design obtained in (i) by adding 1 (mod v),$\ldots, (v-1)$ (mod v) in every attribute under both the options. Note that every copy in (ii) is just the recoding of the design obtained in (i), and hence the resultant design with parameters $k, v, s = hn_1v, b = 1$ is also optimal.

(iii) Finally for each of the h generators, the ith block of size v comprises of the ith row from each of the v copies created in (ii), $i = 1, \ldots, n_1$.

The hn_1 blocks so obtained with $s = v$ forms the required optimal design d_3^B. The design so obtained has distinct choice pairs in every block.

3.5 Optimal block designs for estimating the main plus two-factor interaction effects

The literature on optimal paired choice designs for estimating the main plus two-factor interaction effects is very limited since such designs require a large number of choice pairs to be shown to every respondent. Graßhoff et al. (2003), Street and Burgess (2004) and Großmann, Schwabe and Gilmour (2012) have provided optimal and/or efficient paired choice designs under this setup for k attributes each at two levels. In this Section, we consider each of the k attributes to be at two levels. Let $q = \lceil k/2 \rceil$, where $\lceil z \rceil$ represents the smallest integer greater than or equal to z. The construction method of Street and Burgess (2007) entails starting with an orthogonal array $OA(n_1, k, 2^k, 4)$ as a set of n_1 first options, and then taking the foldover of α attributes in the second option, keeping the rest of the $k - \alpha$ attributes same for each of the n_1 choice pairs. Here $\alpha = q$ for k odd and $\alpha = q$ and $q + 1$ for k even. This process is repeated for $\binom{k}{\alpha}$ possible combinations of the attributes. Here, the foldover of an attribute in the second option of a choice pair means that the attribute level in the second option is different from that in the first. Such a paired choice design d_1^I with parameters $k, v, s, b = 1$ is optimal where $s = n_1\binom{k}{q}$ for k odd and $s = n_1\binom{k+1}{q+1}$ for k even.

Incorporating respondent effects, the model is as given in (3.1). However, in contrast to Section 3.4, interest here lies in the estimation of both the main-effects and the two-factor interaction effects. The information matrix for estimating the main plus two-factor interaction effects under the multinomial logit model incorporating respondent effects is

$$\tilde{C}_I = \begin{bmatrix} C_M & P_M' P_I \\ P_I' P_M & P_I' P_I \end{bmatrix} - (1/s)[P_M' W \; P_I' W] \begin{bmatrix} W' P_M \\ W' P_I \end{bmatrix}. \tag{3.1}$$

As earlier, in order to achieve optimal paired choice block designs, we start with an optimal paired choice design d_1^I and enforce blocking such that $W' P_M = 0$ and $W' P_I = 0$.

We provide a simple condition to achieve the same, proof of which is provided in the Appendix A.

Let pair (a_1, b_1) means that a_1 and b_1 are the levels corresponding to an attribute for the first and second options, respectively. Similarly, let pair (a_1a_2, b_1b_2) means that a_1a_2 and b_1b_2 are the levels corresponding to the two attributes for the first and second options, respectively.

Theorem 3.14. *$W'P_M = 0$ and $W'P_I = 0$ if and only if for every block,*

(i) the frequency of the pair $(1, 0)$ is same as the frequency of the pair $(0, 1)$ for every attribute;

(ii) the frequency of the pairs from the set $\{(01, 00),\ (01, 11),\ (10, 00),\ (10, 11)\}$ is same as the frequency of the pairs from the set $\{(00, 01),\ (00, 10),\ (11, 01),\ (11, 10)\}$ for every two attributes.

We now provide a method of construction for optimal paired choice block designs with $s = 4$.

Theorem 3.15. *For $k > 4$, there exists a paired choice block design d_2^I with parameters $k, v = 2, s = 4, b$, which is optimal in $\mathcal{D}_{k,b,s}$. Here $b = 2^{k-3}\binom{k}{q}$ for k odd and $b = 2^{k-3}\binom{k+1}{q+1}$ for k even.*

Construction. *Let F be a set of $\binom{k}{\alpha}$ attribute indices of size $\alpha = q$ obtained from the attribute labels $1, \cdots, k$ taking α labels at a time such that $2 \leq \alpha \leq k-2$. For an element $f = (f_1, \ldots, f_i, \ldots, f_\alpha)$ of F, let $f' = \{1, \ldots, k\} - f = (f_1', \ldots, f_j', \ldots, f_{(k-\alpha)}')$ be the complement of f. Keeping in view the construction of the design d_1^I, we adopt the steps (i)-(v) to construct an optimal paired choice block design d_2^I for k attributes.*

(i) Write the complete factorial involving 2^α combinations. Divide this set into two-halves such that the second half is a foldover of the first half.

(ii) Write the complete factorial involving $2^{k-\alpha}$ combinations. Divide this set into two-halves such that the second half is a foldover of the first half.

(iii) Take one combination from the first half of (i), say a, and two combinations from the first half of (ii), say b and c. Let a', b' and c' be the foldovers of a, b and c, respectively. Corresponding to the element f of F, make a block having choice pairs $(ab, a'b)$, $(ab', a'b')$, $(a'c, ac)$, $(a'c', ac')$. Here, in a choice pair, the option ab implies that if $a = a_1 \cdots a_i \cdots a_\alpha$

and $b = b_1 \cdots b_j \cdots b_{k-\alpha}$*, then* a_i *corresponds to the attribute index* f_i *and* b_j *corresponds to the attribute index* f'_j*.*

(iv) Repeat (iii) for each of the $2^{\alpha-1}$ *combinations in the first half of (i) using the same b and c as in (iii). Then, repeat the entire process for two different combinations from the first half of (ii).*

(v) Repeating (i)-(iv) for every element f *of* F *corresponding to* $\alpha = q$ *for* k *odd and* $\alpha = q$ *and* $q+1$ *for* k *even, an optimal paired choice block design* d_2^I *is obtained with parameters* $k, v = 2, s = 4, b$ *where* $b = 2^{k-3}\binom{k}{q}$ *for* k *odd and* $b = 2^{k-3}\binom{k+1}{q+1}$ *for* k *even.*

Example 3.16. *Let* $k = 4, v = 2, b = 10, s = 8$*. For* $k = 4$*,* α *takes the values* 2 *and* 3*. Since* $\alpha = 3 > 2 = k - 2$*, Theorem 3.15 does not allow to achieve* d_2^I *from* d_1^I*. However, for* $\alpha = 2$*, the proposed construction method still holds, for which we get* 12 *blocks each of size* 4*, as below.*

B_1	B_2	B_3	B_4	B_5	B_6
(0000,1100)	(0100,1000)	(0000,1010)	(0010,1000)	(0000,1001)	(0001,1000)
(0011,1111)	(0111,1011)	(0101,1111)	(0111,1101)	(0110,1111)	(0111,1110)
(1101,0001)	(1001,0101)	(1011,0001)	(1001,0011)	(1011,0010)	(1010,0011)
(1110,0010)	(1010,0110)	(1110,0100)	(1100,0110)	(1101,0100)	(1100,0101)
B_7	B_8	B_9	B_{10}	B_{11}	B_{12}
(0000,0110)	(0010,0100)	(0000,0101)	(0001,0100)	(0000,0011)	(0001,0010)
(1001,1111)	(1011,1101)	(0110,0011)	(0111,0010)	(1100,1111)	(1101,0010)
(0111,0001)	(0101,0011)	(1011,1110)	(1010,1111)	(0111,0100)	(0110,0101)
(1110,1000)	(1100,1010)	(1101,1000)	(1100,1001)	(1011,1000)	(1010,1001)

For $\alpha = 3$*, we provide a design in* 4 *blocks each of size* 8*, as below.*

B_{13}	B_{14}	B_{15}	B_{16}
(0000,1110)	(0000,1011)	(0000,1101)	(0000,0111)
(0110,1000)	(0011,1000)	(0101,1000)	(0011,0100)
(1010,0100)	(1010,0001)	(1100,0001)	(0101,0010)
(1100,0010)	(1001,0010)	(1001,0100)	(0110,0001)
(0001,1111)	(0100,1111)	(0010,1111)	(1000,1111)
(0111,1001)	(0111,1100)	(0111,1010)	(1011,1100)
(1011,0101)	(1110,0101)	(1110,0011)	(1101,1010)
(1101,0011)	(1101,0110)	(1011,0110)	(1110,1001)

We form 6 *blocks each of size* 8 *by combining blocks* B_i *and* B_{i+6}*,* $i = 1, \ldots, 6$*, which in combination with the* 4 *blocks* B_i*,* $i = 13, \ldots, 16$ *gives the optimal design with parameters* $k = 4, b = 10, s = 8$*.*

3.6 Discussion

In situations where an optimal design has more choice pairs than a respondent can complete, the N choice pairs can be split among the respondents (blocks) either randomly or using a spare attribute, if there is one available (see, Street and Burgess (2007)). To this effect, we have instances of respondents being considered as blocks in various choice experiments, although without much theoretical rigor. Bliemer and Rose (2011) reported that 64% of studies used a blocking column to allocate choice sets to respondents, 13% assigned choice sets randomly to respondents, 5% studies provided the full factorial to each respondent and for the remaining 18% of the studies, it could not be determined how choice sets were assigned to respondents.

With an objective to assess the main or interaction effects, wherever practical, the same set of N optimal choice pairs are shown to every respondent. As such there are no theoretical results on optimal designs, under the utility-neutral setup, where different respondent sees smaller and different designs. In contrast, the approach that is adopted here allows the construction of optimal designs with smaller and flexible number of choice pairs, to be shown to every respondent. Even in situations where simple techniques like blocking using a spare attribute can not be used, we provide optimal paired choice block designs.

In contrast to the approaches of Sándor and Wedel (2005) and Kessels, Goos and Vandebroek (2008), following the block design theory, we adopt the fixed-effects block model for obtaining optimal designs. The approach adopted here treats respondent heterogeneity as a nuisance factor by including respondent-level fixed-effect terms in the model and enables the derivation of analytical results. Though there is no guarantee that the optimal block designs obtained under this setup and the heterogeneous designs obtained by Sándor and Wedel (2005) would be same, it would require a separate study to compare optimal designs obtained under the two approaches.

Furthermore, unlike their designs, which are available only for situations when estimation of the main effects is of interest, we have provided optimal paired choice block designs not only under the main effects model but also under the broader main effects model and under the main plus two-factor interaction effects model.

Chapter 4

Efficient paired choice designs with fewer choice pairs

This chapter is based on the following work:

Dey et al. (2017): Dey, Aloke; Singh, Rakhi; Das, Ashish. Efficient paired choice designs with fewer choice pairs. *Metrika 80* (2017), no. 3, 309–317.

4.1 Introduction

In this chapter, we consider paired choice experiments with each option being described by k factors each at v levels. We denote these levels by $0, \ldots, v-1$. For an attribute, the level l is coded and represented by a unit vector with 1 in the $(l+1)$th position for $l = 0, \ldots, v-2$, and for level $v-1$ is represented by -1 in each of the $v-1$ positions. For example, for $v = 3$, effects-coded vectors for $l = 1, 2, 3$ are $(1\ 0), (0\ 1)$ and $(-1\ \ -1)$, respectively.

We now consider designs under the linear paired comparison model for estimating the main effects. The response is described by the model, $Z = U_1 - U_2 + \epsilon = (P_1 - P_2)\theta + \epsilon = X\theta + \epsilon$, where ϵ is the random error vector, θ is the parameter vector for the main effects, P_j is an $N \times k(v-1)$ effects-coded matrix of the main effects for the jth option, $j = 1, 2$, and $X = P_1 - P_2$. For a paired choice design d, the matrix $M_d = X^T X$ is the information matrix for estimating the main effects under the linear paired comparison model. A paired choice design d is connected if all the main effects are estimable, and this happens if and only if M_d has rank $k(v-1)$. We consider D-optimality in this chapter. The D-efficiency

$\mathit{eff}_D(d)$ of a paired choice designs d with N choice pairs is defined as

$$\mathit{eff}_D(d) = \left(\frac{\det(M_d/N)}{det(M_{d^*}/N^*)} \right)^{1/(k(v-1))},$$

where d^* is a D-optimal design in N^* choice pairs. In view of invariance, the D-efficiency of a design d is same under the linear paired comparison approach and under the multinomial logit model approach.

A major concern with the available optimal paired choice designs is that the number of choice pairs in the design increases rapidly as k and/or v are even moderately increased. Such large designs may not be attractive to an experimenter. In this chapter, we propose two construction methods yielding highly efficient paired choice designs with fewer choice pairs.

4.2 Construction of D-efficient designs

In this section we provide two construction methods for D-efficient paired choice designs with fewer choice pairs. The following result is well known (see e.g., Dey (2009)).

Lemma 4.1. *Consider $v(v-1)/2$ combinations involving v levels taken two at a time. Then,*

(i) For v odd, the combinations can be grouped into $g = (v-1)/2$ groups $G_0, \dots, G_{g-1}$ each comprising $s = v$ combinations. Here $G_i = \{(i, v-2-i), (i+1, v-1-i), \dots, (i+v-1, 2v-3-i)\}$ and the levels are reduced modulo v; $i = 0, \dots, g-1$.

(i) For v even, the combinations can be grouped into $g = v-1$ groups $G_0, \dots, G_{g-1}$ each comprising $s = v/2$ combinations. Here $G_i = \{(i, \infty), (i+1, i+v-2), (i+2, i+v-3), \dots, (i+v/2-1, i+v/2)\}$ and the levels are reduced modulo $v-1$; $i = 0, \dots, g-1$. Here, ∞ is the invariant level $v-1$.

For paired choice designs, only constructions 3.2 and 3.4 of Demirkale et al. (2013) are applicable and construction 3.4 of Demirkale et al. (2013) is same as the construction in Theorem 3 of Graßhoff et al. (2004). Modifications to these constructions, using g groups as in Lemma 4.1 gives rise to efficient designs with fewer choice pairs.

For $k \le m$, let $H_{m,k}$ be a $m \times k$ matrix with elements ± 1 such that $H^T_{m,k} H_{m,k} = mI_k$, where I_k is the identity matrix of order k. When $k = m$, $H_{m,k}$ is called a Hadamard matrix.

Construction 1: Graßhoff et al. (2004) used an $H_{m,k}$ to construct an optimal paired choice design d with parameters $N = mgs = mv(v-1)/2, k, v$, by associating the $gs = v(v-1)/2$ combinations of v levels taken two at a time with the rows of $\{H_{m,k}, -H_{m,k}\}$. From every row of $\{H_{m,k}, -H_{m,k}\}$, gs choice pairs are obtained by replacing '1' in the row by the first element of the combinations and '-1' in the row by the second element of the combinations.

An efficient design is obtained by associating only $g's$ combinations (where $g' = g-1$ or $g' = g-2$) with each row of $\{H_{m,k}, -H_{m,k}\}$. Corresponding to the jth row of $\{H_{m,k}, -H_{m,k}\}$, $(j = 1, \ldots, m)$, the combinations in the g' group $G_{(j-1)}(\text{mod } g), G_j(\text{mod } g)$, $\ldots, G_{(j+g'-2)}(\text{mod } g)$ are used for generating the choice pairs. This gives rise to an efficient paired choice design with k attributes each at v levels and $N = mg's$ choice pairs with $g' = g-1$ or $g' = g-2$.

Example 4.2. *For $k = 9, v = 4$, starting from a normal $H_{12,12}$, a $H_{12,9}$ is obtained by retaining the 2nd to 10th column of $H_{12,12}$. Then generating $gs = v(v-1)/2 = 6$ choice pairs from each row of $H_{12,9}$, an optimal paired choice design is obtained. Since $v = 4$, from Lemma 4.1, the $g = 3$ groups $G_0 = \{(0,3),\ (1,2)\}$, $G_1 = \{(1,3),\ (2,0)\}$ and $G_2 = \{(2,3),\ (0,1)\}$ are each of size $s = 2$.*

We choose $g' = v-2 = 2$ and associate 4 combinations to each row of $\{H_{12,9}, -H_{12,9}\}$. For the jth row of $H_{12,9}$, the combinations in the groups $G_{(j-1)(mod\ 3)}$ and $G_{j(mod\ 3)}$ are used, $j = 1, \ldots, 12$. This gives a design in 48 choice pairs with $\mathit{eff}_D(d) = 0.96$. Notice that the optimal design is available in 72 choice pairs. Therefore, we observe that with 4% loss in D-efficiency, a 33% reduction in the number of choice pairs is achieved.

We now recall the definition of an orthogonal array. An orthogonal array $OA(n, k, v_1 \times \cdots \times v_k, t)$, of strength t, is an $n \times k$ array with elements in the ith column from a set of v_i distinct symbols $\{0, \ldots, v_i - 1\}$ $(i = 1, \ldots, k)$, such that all possible combinations of symbols appear equally often as rows in every $n \times t$ subarray. An orthogonal array is symmetric if $v_i = v$ for all i and the corresponding OA is denoted by $OA(n, k, v^k, t)$. The link `http://support.sas.com/techsup/technote/ts723.html` and Hedayat et al. (1999) provide a comprehensive description of orthogonal arrays and their constructions. An orthogonal array is said to be 1-resolvable (or, simply, resolvable) if its rows can be partitioned into sets of rows (also called parallel classes) such that each set is an orthogonal array of strength unity.

Construction 2: Demirkale et al. (2013) used an $OA(nv, k+1, v^k \times n, 2)$ to form a 1-resolvable $OA(nv, k, v^k, 2)$ having n parallel sets of v rows each, and then to construct an optimal paired choice design d with parameters $N = ngs = nv(v-1)/2, k, v$. Let $\{i, j\}$ be a typical combination among the $gs = v(v-1)/2$ combinations of v numbers $\{0, 1, \ldots, v-1\}$ taken two at a time. Then, for each such combination and from each of the n parallel sets, $(i+1)$th row and $(j+1)$th row are chosen to form the choice pairs of the optimal paired choice design.

An efficient design is obtained by associating only $g's$ combinations (where $g' = g-1$ or $g' = g-2$) with each of the n parallel sets of 1-resolvable $OA(vn, k, v^k, 2)$. Then, corresponding to the jth parallel set ($j = 1, \ldots, n$), the combinations in the group $G_{(j-1)(\text{mod } g)}, G_{j(\text{mod } g)}, \ldots, G_{(j+g'-2)(\text{mod } g)}$ are used for generating the choice pairs. This gives rise to an efficient paired choice design with k attributes each at v levels and $N = ng's$ choice pairs with $g' = g-1$ or $g' = g-2$.

Example 4.3. *For $k = 5, v = 7$, from an $OA(49, 8, 7, 2)$ in the website link mentioned and using the first factor as the resolving factor, 7 parallel sets are created and an optimal paired choice design is obtained by constructing $gs = v(v-1)/2 = 21$ choice pairs corresponding to each parallel set. Since $v = 7$, from Lemma 4.1, three groups formed are $G_0 = \{(0,5), (1,6), (2,0), (3,1), (4,2), (5,3), (6,4)\}$, $G_1 = \{(1,4), (2,5), (3,6), (4,0), (5,1), (6,2), (0,3)\}$ and $G_2 = \{(2,3), (3,4), (4,5), (5,6), (6,0), (0,1), (1,2)\}$.*

We choose $g' = (v-3)/2 = 2$ and hence associate 14 combinations to each of the parallel sets. For the jth parallel set, the combinations in the group $G_{(j-1)(mod\ 3)}$ and $G_{j(mod\ 3)}$ are used, $j = 1, \ldots, 7$. This gives a design in 98 choice pairs with $\mathit{eff}_D(d) = 0.98$. Notice that the optimal design is available in 147 choice pairs. Therefore, we observe that with 2% loss in D-efficiency, a 33% reduction in the number of choice pairs is achieved.

4.3 Tables of Designs and Concluding Remarks

We have two methods of construction of efficient designs. Since, for $v = 2, 3$, the number of choice pairs involved are not very large, it may be preferable to use optimal designs for such cases. However, as the number of levels increases, the number of choice pairs in an optimal design increases rapidly, and thus, it is preferable to use efficient designs with fewer choice pairs. The significant gain through the reduced number of choice pairs compensates for the marginal loss in D-efficiency. In most practical situations, very large values of v and/or k are not useful. We thus restrict ourselves to the values of v, k as in Table 2 of Demirkale et al. (2013). We provide a list of efficient designs in Tables 4.1-4.4. From these tables, we see that retaining a D-efficiency of more than 0.9, on an average there is $30 - 50\%$ reduction in the number of choice pairs in the design.

$H_{g'/g}$ in the Table represent the designs constructed following the Graßhoff et al. (2004) approach using g' groups as against the g groups in the optimal designs, and $D_{g'/g}$ represent the designs constructed following the Demirkale et al. (2013) approach using g' groups as against the g groups in the optimal designs. N^* represents the least number of choice pairs for an optimal design as in Table 2 of Demirkale et al. (2013). The last column 'Reduction' depicts the percentage reduction in the number of choice pairs vis-á-vis the designs as in Table 2 of Demirkale et al. (2013).

Table 4.1: Efficient designs for $v = 4$

k	N^*	N	$eff_D(d)$	Design	Reduction	k	N^*	N	$eff_D(d)$	Design	Reduction
3	24	16	0.9596	$H_{2/3}$	33%	8	48	32	0.9449	$H_{2/3}$	33%
4	24	16	0.9449	$H_{2/3}$	33%	9	72	48	0.9590	$H_{2/3}$	33%
5	48	32	0.9691	$D_{2/3}$	33%	10	72	48	0.9542	$H_{2/3}$	33%
6	48	32	0.9596	$D_{2/3}$	33%	11	72	48	0.9491	$H_{2/3}$	33%
7	48	32	0.9515	$H_{2/3}$	33%	12	72	48	0.9449	$H_{2/3}$	33%

Table 4.2: Efficient designs for $v = 5$

k	N^*	N	$eff_D(d)$	Design	Reduction	k	N^*	N	$eff_D(d)$	Design	Reduction
3	40	25	0.9469	$D_{1/2}$	38%	7	80	40	0.9088	$H_{1/2}$	50%
3	40	20	0.9283	$H_{1/2}$	50%	8	80	50	0.9174	$D_{1/2}$	38%
4	40	25	0.9188	$D_{1/2}$	38%	8	80	40	0.8944	$H_{1/2}$	50%
4	40	20	0.8944	$H_{1/2}$	50%	9	100	60	0.9162	$H_{1/2}$	40%
5	50	40	0.9146	$H_{1/2}$	20%	9	100	50	0.9056	$D_{1/2}$	50%
5	50	25	0.8944	$D_{1/2}$	50%	10	100	60	0.9061	$H_{1/2}$	40%
6	50	40	0.9283	$H_{1/2}$	20%	10	100	50	0.8944	$D_{(1/2}$	50%
7	80	50	0.9300	$D_{1/2}$	38%	11	100	60	0.9035	$H_{1/2}$	40%

Table 4.3: Efficient designs for $v = 6$

k	N^*	N	$eff_D(d)$	Design	Reduction	k	N^*	N	$eff_D(d)$	Design	Reduction
3	60	48	0.9860	$H_{4/5}$	20%	5	120	96	0.9886	$H_{4/5}$	20%
3	60	36	0.9606	$H_{3/5}$	40%	5	120	72	0.9681	$H_{3/5}$	40%
4	60	48	0.9801	$H_{4/5}$	20%	6	120	108	0.9736	$D_{3/5}$	10%
4	60	36	0.9426	$H_{3/5}$	40%	6	120	96	0.9850	$H_{4/5}$	20%
5	120	108	0.9781	$D_{3/5}$	10%	6	120	72	0.9585	$H_{3/5}$	40%

Table 4.4: Efficient designs for $v = 7$

k	N^*	N	$eff_D(d)$	Design	Reduction	k	N^*	N	$eff_D(d)$	Design	Reduction
3	84	56	0.9752	$H_{2/3}$	33%	5	147	49	0.8878	$D_{1/3}$	67%
3	84	49	0.9327	$D_{1/3}$	42%	6	147	112	0.9706	$H_{2/3}$	24%
3	84	28	0.8847	$H_{1/3}$	67%	6	147	98	0.9697	$D_{2/3}$	33%
4	84	56	0.9638	$H_{2/3}$	33%	6	147	56	0.8556	$H_{1/3}$	62%
4	84	49	0.9099	$D_{1/3}$	42%	6	147	49	0.8530	$D_{1/3}$	67%
4	84	28	0.8198	$H_{1/3}$	67%	7	147	112	0.9689	$H_{2/3}$	24%
5	147	112	0.9764	$H_{2/3}$	24%	7	147	98	0.9638	$D_{2/3}$	33%
5	147	98	0.9764	$D_{2/3}$	33%	7	147	56	0.8483	$H_{1/3}$	62%
5	147	56	0.8876	$H_{1/3}$	62%	7	147	49	0.8198	$D_{1/3}$	67%

Chapter 5

Three-level A- and D-optimal paired choice designs

This chapter is based on the following work:

Chai et al. (2017): Chai, Feng-Shun; Das, Ashish; Singh, Rakhi. Three-level A- and D-optimal paired choice designs. *Statist. Probab. Lett. 122* (2017), 211–217.

5.1 Introduction

In this chapter, we consider each factor to be at three levels, 0, 1 and 2 (say). For a design d with N choice pairs, since options are described by k factors, following Huber and Zwerina (1996), the utilities are modeled using the linear predictor $u_j = P_j\theta$, where θ is a $2k \times 1$ vector representing the main effects, P_j is an $N \times 2k$ effects coded matrix for the jth option, and $u_j = (u_{ji})$ is an $N \times 1$ utility vector for the jth option, $j = 1, 2; i = 1, \ldots, N$. The utility difference $u_1 - u_2 = (P_1 - P_2)\theta = X\theta$ is then a linear function of the parameter vector θ. In what follows, we refer to X as the design matrix of design d. Since multinomial logit choice models are non-linear in the parameters and the information matrix is a function of the parameters, a utility-neutral approach (that is, taking $\theta = 0$) of finding the information matrix has been developed over the last two decades. Under such a utility-neutral multinomial logit model, the Fisher information matrix for a design d is $(1/4)M_d$, where $M_d = X^T X$.

Simultaneously, Graßhoff et al. (2004) for a design d, the matrix M_d is the information matrix under the linear paired comparison model.

Recently, Sun and Dean (2016) have provided an efficient computer algorithm to obtain two-level A-optimal choice designs under the locally-optimal approach (that is, taking $\theta = \theta_0$, for an a priori θ_0). In this chapter, we consider each factor to be at three levels and theoretically obtain new A- and D-optimal designs under the utility-neutral multinomial logit model setup.

For three-level factors, a choice design d is connected if all the main effects are estimable, and this happens if and only if M_d has rank $2k$. In what follows, the class of all connected paired choice designs with k three-level factors and N choice pairs is denoted by $\mathcal{D}_{k,N}$. As a performance measure, we use the standard A- and D-optimality criteria. The A-value of a design d is $trace(M_d^{-1})$ and the D-value is $det(M_d^{-1})$. A design that minimizes the A-value (the D-value) among all designs in $\mathcal{D}_{k,N}$ is said to be A-optimal (D-optimal).

In this chapter, we provide constructions of A- and D-optimal designs for estimating the main effects under the utility-neutral multinomial logit model using effects coding. We also provide designs having high A- and D-efficiencies. Finally, we investigate optimal designs under the utility-neutral multinomial logit model approach of Street and Burgess (2007) and show that the D-optimal designs obtained under the model using effects coding are also A- and D-optimal under the Street–Burgess approach.

5.2 Lower bounds to the A-value

Considering each factor at 3 levels, the ith row of the $N \times 2k$ effects coded matrix P_j contains the effects coding for the jth option in ith choice pair, $i = 1, \dots, N$, $j = 1, 2$. For each of the k factors, level 0 is effects coded as $(1\ \ 0)$, level 1 as $(0\ \ 1)$ and level 2 as $(-1\ \ -1)$. The design matrix $X = P_1 - P_2$ for the pairs can be partitioned as $X = (X_{(1)}|X_{(2)}|\cdots|X_{(k)})$, where $X_{(p)}$ is a $N \times 2$ matrix corresponding to the pth factor. A row in $X_{(p)}$ determine the corresponding options in a design for the pth factor. In $X_{(p)}$, rows $(+2,\ +1)$, $(-2,\ -1)$, $(+1,\ +2)$, $(-1,\ -2)$, $(+1,\ -1)$ and $(-1,\ +1)$ correspond to choice pairs $(0,\ 2)$, $(2,\ 0)$, $(1,\ 2)$, $(2,\ 1)$, $(0,\ 1)$ and $(1,\ 0)$ respectively. Similarly, row $(0,\ 0)$ corresponds to any of the choice pairs $(0,\ 0)$, $(1,\ 1)$ or $(2,\ 2)$. Let, $M_{pq} = X_{(p)}^T X_{(q)}$; $p = 1, \dots, k; q = 1, \dots, k$. Clearly, for a design d, $M_d = (M_{dpq})$, where M_{dpq} denotes M_{pq} for design d. Let d_1 have $w > 0$ rows of $X_{(p)}$ that are equal to $(0,\ 0)$ and let d_2 be the design obtained from d_1 by replacing the w rows $(0,\ 0)$ by either $(1,\ -1)$ or $(-1,\ 1)$. Then, $M_{d_2pp} - M_{d_1pp} = w \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ is a non-negative definite matrix. Thus, without affecting the generality of the results that follow, we write M_{dpp} having $X_{(p)}$ with no rows $(0,\ 0)$. It is easy to see that

$$M_{dpp} = \begin{bmatrix} 3y+N & 3(y+z)-N \\ 3(y+z)-N & 3z+N \end{bmatrix}, \tag{5.1}$$

where y is the number of rows of $X_{(p)}$ that are equal to either $(2,\ 1)$ or $(-2,-1)$ and z is the number of rows of $X_{(p)}$ that are equal to either $(1,\ 2)$ or $(-1,-2)$. Then the remaining $N-(y+z)$ rows of $X_{(p)}$ are necessarily equal to either $(1,\ -1)$ or $(-1,\ 1)$.

We first obtain a lower bound to $trace(M_{dpp}^{-1})$. From (5.1), $trace(M_{dpp}^{-1}) = (3(y+z)+2N)/h_N(y,z) = g_N(y,z)$ (say) where $h_N(y,z) = det(M_{dpp}) = 9(yz+N(y+z)-(y+z)^2)$. Note that since both $h_N(y,z)$ and $g_N(y,z)$ are symmetric in y and z, it follows that $h_N(y,z) = h_N(z,y)$ and $g_N(y,z) = g_N(z,y)$. We now find the values y and z for which $g_N(y,z)$ is minimized for $1 \le y+z \le N$, $y \ne N$, $z \ne N$. These conditions are required so that the design d is connected, that is, M_d has full rank. Even though it appears that these conditions are only needed for ensuring that $rank(M_{dpp}) = 2$, in fact if these conditions are not satisfied for every p, then $rank(M_d) < 2k$. This can be seen easily since if one of the factors (say the pth factor) does not meet the conditions, all the N pairs would assume only one type of values which would mean that the two columns of $X_{(p)}$ are linearly dependent. Let $min_{d\in\mathcal{D}_{1,N}} trace(M_{dpp}^{-1}) = min_{1\le y+z\le N, y\ne N, z\ne N}\, g_N(y,z) = L_a$. Also, let $\lfloor x \rfloor$ denote the greatest integer contained in x.

Lemma 5.1. *For a single-factor design $d \in \mathcal{D}_{1,N}$ with $N > 4$, $trace(M_{dpp}^{-1}) = g_N(y,z) \ge L_a = g_N(a^*,b^*)$ where $g_N(a^*,b^*)$ is $min\{g_N(a_1,b_1), g_N(a_2,b_2), g_N(a_3,b_3)\}$ with (i) $a_1 = b_1 = t$, (ii) $a_2 = b_2 = t+1$, (iii) $a_3 = t, b_3 = t+1$ and $t = \lfloor N(\sqrt{3}-1)/3 \rfloor$. For $N = 4$, $trace(M_{dpp}^{-1}) = g_N(y,z) \ge L_a = g_N(a^*,b^*) = 14/45$ with $a^* = b^* = 1$.*

Proof. Treating y and z as continuous variables and adopting the usual derivative approach to minimize $g_N(y,z)$, we get $\partial g_N(y,z)/\partial y = 9(3y^2-2N^2+6yz+4Ny+2Nz)/h_N^2(y,z)$. Similarly, $\partial g_N(y,z)/\partial z = 9(3z^2-2N^2+6yz+4Nz+2Ny)/h_N^2(y,z)$.

Now, $\partial g_N(y,z)/\partial y = \partial g_N(y,z)/\partial z = 0$ implies that $(y-z)(3(y+z)+2N) = 0$. In other words, y=z, since $3(y+z)+2N > 0$.

Now, for $y = z$, it follows that $\partial g_N(y,z)/\partial y = 0$ implies that $9y^2 + 6Ny - 2N^2$=0 or $y = N(\pm\sqrt{3}-1)/3$. However, since $y \geq 0$, the only feasible solution of y is $N(\sqrt{3}-1)/3 = t_1$.

Similarly, checking the matrix of second derivatives, we see that the minimum of $g_N(y,z)$ is attained at $y = z = t_1$. Since t_1 is non-integer, $g_N(y,z) = L_a$ at one of the integer points nearest to (t_1, t_1).

For $N = 4$, since $y + z \geq 1$, the only valid integer points nearest to (t_1, t_1) are $(1,1)$ or $(0,1)$. It is then easy to see that $g_N(0,1) \geq g_N(1,1)$ and therefore, $trace(M_{dpp}^{-1}) = g_N(y,z) \geq L_a = g_N(a^*, b^*) = 14/45$ with $a^* = b^* = 1$. □

Using Lemma 5.1, we have computed the values of a^* and b^* for $4 \leq N \leq 64$ and summarize it below.

Remark 5.2. *For a single-factor design $d \in \mathcal{D}_{1,N}$, the values of a^* and b^* are*

1. $a^* = b^* = t$ *when*
(i) $N = 4i+1, i = 1, \ldots, 10$, *(ii)* $N = 4i+2, i = 7, \ldots, 15$.

2. $a^* = t, b^* = t+1$ *when*
(i) $N = 4i+2, i = 1, \ldots, 6$, *(ii)* $N = 4i+3, i = 4, \ldots, 15$, *(iii)* $N = 60, 64$.

3. $a^* = b^* = t+1$ *when*
(i) $N = 4i, i = 1, \ldots, 13$, *(ii)* $N = 4i+1, i = 11, \ldots, 15$, *(iii)* $N = 7, 11$.

4. $a^* = 3, b^* = 4$ *or* $a^* = b^* = 4$ *when* $N = 15$.

5. $a^* = 13, b^* = 14$ *or* $a^* = b^* = 14$ *when* $N = 56$.

Note that since $g_N(y,z)$ is symmetric in y and z, interchanging the values of a^* and b^* in Lemma 5.1 and in Remark 5.2 would yield the same values of $g_N(a^*, b^*)$ and therefore for the sake of simplicity, we have reported only one of the a^* and b^* values. Let $L_A = min_{d \in \mathcal{D}_{k,N}} trace(M_d^{-1})$. We now give a lower bound for the A-value for paired choice designs with k factors in N choice pairs.

Theorem 5.3. *For a paired choice design $d \in \mathcal{D}_{k,N}$, $trace(M_d^{-1}) \geq L_A \geq kL_a = kg_N(a^*, b^*)$ where a^* and b^* are as in Lemma 5.1.*

Proof. First we apply the inequality $trace(M_d^{-1}) \geq \sum_{p=1}^{k} trace(M_{dpp}^{-1})$ which, using Schur complement and the inverse of partitioned matrices, follows easily for $k = 2$. Thereafter, it is easy to see, using the method of induction, that the inequality holds for general k (see, Appendix B for detailed proof). Finally, using Lemma 5.1, the proof follows. □

5.3 Lower bounds to the D-value

In this section, we provide general results for the lower bounds to D-values. We first obtain a lower bound to $det(M_{dpp}^{-1})$. From (5.1), $det(M_{dpp}^{-1}) = 1/h_N(y, z)$. We now find the values y and z for which $h_N(y, z)$ is minimized for $1 \leq y + z \leq N$, $y \neq N$, $z \neq N$. Let $min_{d \in \mathcal{D}_{1,N}} det(M_{dpp}^{-1}) = min_{1 \leq y+z \leq N, y \neq N, z \neq N}(1/h_N(y, z)) = L_d$.

Lemma 5.4. *For a single-factor design $d \in \mathcal{D}_{1,N}$, $det(M_{dpp}^{-1}) = 1/h_N(y, z) \geq L_d = 1/h_N(a^*, b^*)$ where the values of a^* and b^* are (i) $a^* = l, b^* = l$ if $N = 3l$, (ii) $a^* = l, b^* = l$ or $a^* = l, b^* = l + 1$ if $N = 3l + 1$ and (iii) $a^* = l, b^* = l + 1$ or $a^* = l + 1, b^* = l + 1$ if $N = 3l + 2$. Also, if $N = 3l$, then $h_N(a^*, b^*) = 3N^2$, else $h_N(a^*, b^*) = 3(N^2 - 1)$.*

Proof. Assuming y, z to be continuous, we find their values such that $1/h_N(y, z)$ is minimized or equivalently $h_N(y, z)$ is maximized. Now $\frac{\partial h_N(y,z)}{\partial y}$=$9(N - 2y - z)$, while $\frac{\partial h_N(y,z)}{\partial z}$=$9(N - 2z - y)$. Equating these equations to zero, we get a^*=b^*=$\frac{N}{3}$ as a feasible solution when N is divisible by 3. Therefore, for $N = 3l$, $a^* = b^* = l$ and $h_N(l, l) = 3N^2$. For situations when N is not divisible by 3, we have the following two cases.

Case 1. $N = 3l + 1$. For integers i, j, consider the difference $h_N(l, l) - h_N(l + i, l + j)$. On simplification we see that, $h_N(l, l) - h_N(l + i, l + j) = 9(i^2 + j^2 + ij - i - j)$. We have the following four cases. Case (a) $i \geq 0, j \geq 0$, Case (b) $i < 0, j < 0$, Case (c) $i \geq 0, j < 0$ and Case (d) $i < 0, j \geq 0$. Clearly, for Case (a) and Case (b), $h_N(l, l) - h_N(l + i, l + j) \geq 0$. For Case (c), $h_N(l, l) - h_N(l + i, l + j) = 9((i + j)^2 - ij - i - j) = 9((i + j)^2 + i(-j - 1) - j) \geq 0$. Similarly, for Case (d), $h_N(l, l) - h_N(l + i, l + j) \geq 0$.

Now, $h_N(l, l) = h_N(l + i, l + j)$ when $(i^2 + j^2 + ij - i - j) = i(i - 1) + j(j - 1) + ij = 0$, that is, when either (i) $i = j = 0$, or (ii) $i = 0, j = 1$, or (iii) $i = 1, j = 0$. Therefore, $a^* = l, b^* = l$ or $a^* = l, b^* = l + 1$ or $a^* = l + 1, b^* = l$ and $h_N(a^*, b^*) = 3(N^2 - 1)$.

Case 2. $N = 3l+2$. For integers i, j, consider the difference $h_N(l+1, l+1) - h_N(l+1+i, l+1+j)$. On lines similar to that in Case 1, we see that, $h_N(l+1, l+1) - h_N(l+1+i, l+1+j) = 9(i^2 + j^2 + ij + i + j) \geq 0$.

Now, $h_N(l+1, l+1) = h_N(l+1+i, l+1+j)$ when $(i^2 + j^2 + ij + i + j) = i(i+1) + j(j+1) + ij = 0$, that is, when either (i) $i = j = 0$, or (ii) $i = 0, j = -1$, or (iii) $i = -1, j = 0$. Therefore, $a^* = l+1, b^* = l+1$ or $a^* = l+1, b^* = l$ or $a^* = l, b^* = l+1$ and $h_N(a^*, b^*) = 3(N^2 - 1)$. □

Note that since $h_N(y, z)$ is symmetric in y and z, interchanging the values of a^* and b^* in Lemma 5.4 would yield the same values of $h_N(a^*, b^*)$ and therefore for the sake of simplicity, we have reported only one of the a^* and b^* values.

Now, to obtain a lower bound to the D-value for paired choice designs with k factors in N choice pairs, we use the inequality $det(M_d^{-1}) \geq \prod_{p=1}^{k} det(M_{dpp}^{-1})$ which is easy to establish by using Schur complement and the method of induction (see, Appendix B for detailed proof). Let $L_D = min_{d \in \mathcal{D}_{k,N}} det(M_d^{-1})$. Thus, using Lemma 5.4, we have

Theorem 5.5. *For a paired choice design $d \in \mathcal{D}_{k,N}$, $det(M_d^{-1}) \geq L_D \geq (L_d)^k = (1/h_N(a^*, b^*))^k$ where a^* and b^* are as in Lemma 5.4. Also, $h_N(a^*, b^*) = 3N^2$, if $N \equiv 0$ (mod 3) and $h_N(a^*, b^*) = 3(N^2 - 1)$ otherwise.*

In situations where N is not a multiple of 3, the lower bound for $det(M_d^{-1})$ obtained above is an improvement over the bounds obtainable from optimal approximate designs of Graßhoff et al. (2004). When N is a multiple of 3, the two bounds are the same.

In the next section, we provide some optimal designs attaining the lower bounds of Theorem 5.3 and Theorem 5.5. In some situations, since we are not able to provide designs attaining the A- and D-lower bounds, A- and D-efficiencies are given.

5.4 Design Constructions

Let $d \in \mathcal{D}_{1,N}$ be an optimal design. Since d is not unique, we have many such designs. Let d_p, $p = 1, \ldots, k$ be k such designs satisfying $X_{(p)}^T X_{(p')} = 0, p \neq p'; p = 1, \ldots, k; p' = 1, \ldots, k$, where the $N \times 2$ matrix $X_{(p)}$ denotes the design matrix of $d_p \in D_{1,N}$. Then, from Theorem 5.3 and Theorem 5.5, $X = (X_{(1)} | X_{(2)} | \cdots | X_{(k)})$ gives rise to an optimal paired choice design in $\mathcal{D}_{k,N}$. It may be noted that single-factor optimal designs $d_p, p = 1, \ldots, k$, satisfying the stated orthogonality condition $X_{(p)}^T X_{(p')} = 0$, may not always exist.

For $N \leq 9$ and certain values of k, through a computer search we are able to identify d_p's and the corresponding optimal designs in $\mathcal{D}_{k,N}$. In particular, we find D-optimal designs for $k = 2$ and $N = 4, 5, 6, 8, 9$ as well as for $k = 3$ and $N = 7$. Similarly, we find A-optimal designs for $k = 1$ and $N = 5$, for $k = 2$ and $N = 4, 6, 9$, for $k = 3$ and $N = 7$, and for $k = 4$ and $N = 8$. We refer to these designs as *base designs*. The base designs are subsequently used for constructing optimal and efficient designs for larger numbers of factors k and $N > 9$.

Since for $k = 2$ and $N = 5$ there is no design in $\mathcal{D}_{2,5}$ satisfying $X_{(1)}^T X_{(2)} = 0$ and attaining the lower bound $2L_a$, we did an exhaustive search in $\mathcal{D}_{2,5}$ to obtain an A-optimal design. The search established that the D-optimal design in $\mathcal{D}_{2,5}$ is also A-optimal. This shows that the lower bound of Theorem 5.3 is not always attainable.

For a paired choice design with k factors and N choice pairs, we denote a D-optimal design by $d_{(k,N)}$, an A-optimal design by $a_{(k,N)}$ and a design which is both A- and D-optimal by $ad_{(k,N)}$. In the Appendix B, we provide the base designs for $k = 2$ and $k = 3$.

A choice design where no two choice pairs are repeated has distinct choice pairs. While obtaining A-optimal design for $k = 2$ and $N = 9$, a complete search indicates that even though there exist $X_{(1)}$ and $X_{(2)}$ satisfying Theorem 5.3 such that $X_{(1)}^T X_{(2)} = 0$, they lead to choice designs with repeated choice pairs. The A-optimal design $a_{(2,9)}$ is one such example provided in the Appendix B. However, such a design is not recommended for experimentation. Accordingly, a complete search was made among designs with distinct choice pairs to arrive at an A-optimal design $a_{(2,9)}^+$, which is also provided in the Appendix B. However, $a_{(2,9)}^+$ does not attain the lower bound $2L_a$ in Theorem 5.3.

All the base designs in the Appendix B satisfy the orthogonality condition $X_{(p)}^T X_{(p')} = 0$ for $p \neq p'$. Moreover, the lower bounds in Theorem 5.3 and Theorem 5.5 are attained by all A- and D-optimal base designs respectively, except for the A-optimal designs $ad_{(2,5)}$ and $a_{(2,9)}^+$.

We now propose a general method of construction to obtain optimal and efficient designs with $k \geq 4$. We use Hadamard matrices for our construction. A Hadamard matrix H_m is a $m \times m$ matrix with elements ± 1 such that $H_m^T H_m = H_m H_m^T = mI_m$.

Consider A- and D-optimal base designs $d' \in \mathcal{D}_{k',N'}$ with $L_A = k' g_{N'}(a', b')$ and $L_D =$

$(1/h_{N'}(a',b'))^{k'}$ respectively. Note that, we are considering the A- and D-optimal base designs in $\mathcal{D}_{k',N'}$ for which the bound values L_A and L_D respectively, are not necessarily equal to the lower bounds in Theorem 5.3 and Theorem 5.5. This also implies that a' and b' are equal to a^* and b^* respectively, only in cases where the base design attains the lower bounds in Theorem 5.3 and Theorem 5.5. Using such a base design d', we construct a paired choice design d_H with parameters $k = mk'$, $N = mN'$ having corresponding design matrix $X_H = H_m \otimes X$, where X is the design matrix of d'. The idea behind the construction is essentially the same as in Theorem 5 of Graßhoff et al. (2004), where optimal designs are obtained as the Kronecker product using an orthogonal array as the base design. This chapter extends this idea in the sense that the base designs are not restricted to orthogonal arrays and that the two- or three-factor base designs are found by means of complete computer searches. This enables one to construct designs with smaller number of choice pairs than other researchers.

For any design $d \in \mathcal{D}_{k,N}$, from Theorem 5.3, the lower bound to the A-value is $kg_N(a^*,b^*)$. In contrast, $d_H \in \mathcal{D}_{k,N}$ has the A-value $trace(M_{d_H}^{-1}) = trace(M_{d'}^{-1}) = k'g_{N'}(a',b')$, since $M_{d_H} = X_H^T X_H = H_m^T H_m \otimes X^T X = mI_m \otimes X^T X = mI_m \otimes M_{d'}$. Therefore the A-efficiency is given by $\phi_A = \frac{kg_N(a^*,b^*)}{k'g_{N'}(a',b')}$. Similarly, $\phi_D = \left(\frac{(m^2 h_{N'}(a',b'))^k}{(h_N(a^*,b^*))^k}\right)^{1/(2k)} = \sqrt{\frac{m^2 h_{N'}(a',b')}{h_N(a^*,b^*)}}$ is the D-efficiency of $d_H \in \mathcal{D}_{k,N}$ where $h_N(a^*,b^*)$ is as in Theorem 5.5.

Also, note that the D-efficiency ϕ_D is based on the lower bounds in Theorem 5.5 for exact designs in $\mathcal{D}_{k,N}$. Therefore, ϕ_D only agrees with the efficiency based on the optimal approximate design when N is a multiple of three.

In Table 5.1, we provide designs with distinct choice pairs that are A-optimal and D-optimal (wherever we get one), and A-efficient and D-efficient. G_{2004} and S_{2007} respectively represents designs constructed as in Graßhoff et al. (2004) and Street and Burgess (2007). In Table 5.1 we denote designs d_H obtained using the Hadamard matrix H_m and a base design $ad_{(k',N')}$ by $H_m \otimes ad_{(k',N')}$. Similar representations are used for base designs $a_{(k',N')}$ and $d_{(k',N')}$. A design with a smaller k retains its optimality property for given N when factors are deleted from a design with larger k.

We see that there are several situations where for given k, a highly D-efficient design with smaller N is available through the method given in this chapter as compared to the D-optimal designs available in the literature. For example, for $k = 8$, we get an A-optimal design in $N = 16$ choice pairs with $\phi_D = 0.976$ as against a D-optimal design with 24 choice pairs. Similarly, for $k = 6$, we get a design with $\phi_D = 0.992$ in $N = 14$ choice pairs as against an A-optimal design in 16 and a D-optimal design in 18 choice pairs. We also have a few situations where for given k and N, we have two different designs of which one is A-optimal or A-efficient and the other is D-optimal or D-efficient. Unlike the new optimal designs constructed here, the D-optimal designs that follow from Graßhoff et al. (2004) and Street and Burgess (2007) force N to be a multiple of 9 or 12, except for $N = 6$.

Table 5.1: A-optimal and D-optimal designs with distinct choice pairs

k	N	ϕ_A	ϕ_D	Method	k	N	ϕ_A	ϕ_D	Method
2	4	Opt	Opt	$ad_{(2,4)}$	3–4	18	0.983	0.981	$H_2 \otimes a^+_{(2,9)}$
2	5	Opt	Opt	$ad_{(2,5)}$	3–7	18	0.936	Opt	G_{2004}, S_{2007}
2	6	Opt	0.971	$a_{(2,6)}$	5–8	20	0.943	0.981	$H_4 \otimes ad_{(2,5)}$
2	6	0.955	Opt	G_{2004}	5–8	24	0.978	0.957	$H_4 \otimes a_{(2,6)}$
2–3	7	Opt	Opt	$ad_{(3,7)}$	5–8	24	0.933	Opt	G_{2004}
2	8	0.948	Opt	$d_{(2,8)}$	7–13	27	0.934	Opt	G_{2004}, S_{2007}
2–4	8	Opt	0.976	$H_2 \otimes ad_{(2,4)}$	5–12	28	0.985	0.99	$H_4 \otimes ad_{(3,7)}$
2	9	0.985	0.981	$a^+_{(2,9)}$	9–16	32	Opt	0.969	$H_8 \otimes ad_{(2,4)}$
2–4	9	0.938	Opt	G_{2004}, S_{2007}	9–12	36	0.933	Opt	G_{2004}
3–4	10	0.951	0.985	$H_2 \otimes ad_{(2,5)}$	9–16	40	0.943	0.98	$H_8 \otimes ad_{(2,5)}$
3–4	12	0.978	0.957	$H_2 \otimes a_{(2,6)}$	9–16	48	0.978	0.957	$H_8 \otimes a_{(2,6)}$
3–4	12	0.933	Opt	G_{2004}	13–16	48	0.933	Opt	G_{2004}
3–6	14	0.989	0.992	$H_2 \otimes ad_{(3,7)}$	17–25	54	0.933	Opt	G_{2004}, S_{2007}
3–4	16	0.948	0.994	$H_2 \otimes d_{(2,8)}$	17–24	56	0.985	0.99	$H_8 \otimes ad_{(3,7)}$
5–8	16	Opt	0.976	$H_4 \otimes ad_{(2,4)}$	17–32	64	0.999	0.968	$H_{16} \otimes ad_{(2,4)}$

5.5 Optimal designs under the Street–Burgess approach

In Section 5.2 and Section 5.3, we have obtained A- and D-optimal designs under the utility-neutral multinomial logit model approach using effects coding (see, for example, Huber and Zwerina (1996), Graßhoff et al. (2004) and Großmann and Schwabe (2015)). However, a different approach has been adopted by the school comprising researchers like

Street, Burgess, Bush and others for obtaining choice designs under the utility-neutral multinomial logit model (see, Street and Burgess (2007), Großmann and Schwabe (2015)). The approach varies in the sense that an information matrix for the option effects is obtained first and then a suitable contrast of the option effects constitutes the main effects. In this section, we investigate optimal designs following the approach of Street and Burgess (2007). Note that since the D-criterion is invariant to re-parameterizations, the designs that are D-optimal under the effects coding setup are also D-optimal under the utility-neutral multinomial logit model approach of Street and Burgess (2007) (see, Großmann and Schwabe (2015)). However, A-optimal designs may be different for the two approaches.

Under the utility-neutral multinomial logit model approach of Street and Burgess (2007), the information matrix for estimating main effects is given by $C_d = B\Lambda_d B^T$ where B is the $2k \times 3^k$ orthonormal contrast matrix for the k main effects and Λ_d is the $3^k \times 3^k$ information matrix for the options of a paired choice design d. Hereafter, for notational simplicity we drop the subscript d in C_d and Λ_d. For a paired choice design d with N choice pairs, $\Lambda = (\lambda_{rs})$, where

$$4N\lambda_{rs} = \begin{cases} n_r & \text{for } r = s, \\ -n_{r,s} & \text{for } r \neq s \end{cases}$$

with r and s being the labels of the corresponding options, n_r being the number of times r appears in the choice design and $n_{r,s} = 1$ or 0 depending on whether r and s forms a choice pair in the design or not.

Similar to the approach followed in Section 5.2 and Section 5.3, we now find the information matrix for $k = 1$. Let for the pth factor in N choice pairs, Λ_p be the corresponding 3×3 information matrix of the levels, and B_o is a 2×3 orthonormal contrast matrix given by,

$$B_o = \begin{pmatrix} -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{pmatrix}.$$

Note that B_o is not unique. As in the earlier sections, let the design d for a single factor have y choice pairs equal to either $(0,\ 2)$ or $(2,\ 0)$ and have z choice pairs equal to either $(1,\ 2)$ or $(2,\ 1)$. Then the remaining $N - (y + z)$ choice pairs are necessarily equal to either $(0,\ 1)$ or $(1,\ 0)$. Then the 2×2 information matrix for estimating the main effect for the pth factor is

$$C_{(p)} = B_o \Lambda_p B_o^T = \frac{1}{N}\sum_{i=1}^{N} B_o \Lambda_p^i B_o^T = \frac{1}{N}\sum_{i=1}^{N} C_{(p)}^i,$$

where Λ_p^i is the Λ_p corresponding to the ith choice pair and $C_{(p)}^i$ is the information matrix of the ith choice pair for estimating the main effect for the pth factor. It is easy to see that $C_{(p)}^i = \frac{1}{2}\begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} = C_{0,2}$ (say) corresponding to the ith choice pair equal to either $(0,\ 2)$ or $(2,\ 0)$. Similarly, $C_{(p)}^i = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{pmatrix} = C_{1,2}$ (say) corresponds to the ith choice pair equal to either $(1,\ 2)$ or $(2,\ 1)$ and $C_{(p)}^i = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & 3 \end{pmatrix} = C_{0,1}$ (say) corresponds to the ith choice pair equal to either $(0,\ 1)$ or $(1,\ 0)$. Therefore,

$$C_{(p)} = \frac{1}{N}(yC_{0,2} + zC_{1,2} + (N-y-z)C_{0,1}) = \frac{1}{2N}\begin{pmatrix} N+3y & \sqrt{3}(y+2z-N) \\ \sqrt{3}(y+2z-N) & 3(N-y) \end{pmatrix},$$

and $det(C_{(p)}) = 3(yz + N(y+z) - (y+z)^2)/N^2 = h_N(y,z)/3N^2$ where $h_N(y,z)$ is as defined in Section 5.2. Thus, $det(C_{(p)}^{-1}) = 3N^2/h_N(y,z)$. Also, $trace(C_{(p)}^{-1}) = 2det(C_{(p)}^{-1}) = 6N^2/h_N(y,z) \geq 6N^2 min_{1\leq y+z\leq N, y\neq N, z\neq N}(1/h_N(y,z)) = 6N^2 L_d$.

Now, let $C = (C_{pq})$ where C_{pq} is the pqth 2×2 sub-matrix of C, $p = 1,\ldots,k; q = 1,\ldots,k$. Therefore, since $3^{k-1}C_{pp} = C_{(p)}$, we have,

$$trace(C^{-1}) \geq \sum_{p=1}^{k} trace(C_{pp}^{-1}) = 3^{k-1}\sum_{p=1}^{k} trace(C_{(p)}^{-1}) = 2\times 3^{k-1}\sum_{p=1}^{k} det(C_{(p)}^{-1}) \geq 2\times 3^k N^2 k L_d.$$

This establishes the following result.

Theorem 5.6. *The D-optimal designs that satisfy the bounds obtained in Theorem 5.5 are also A-optimal and D-optimal under the utility-neutral multinomial logit model approach of Street–Burgess.*

This shows that A-optimal designs could be different under the two approaches as envisaged in Großmann and Schwabe (2015). The D-optimal designs in Table 5.1 are A-optimal and D-optimal under the Street–Burgess approach.

5.6 Concluding Remarks

In the literature, most of the theoretical results on optimal paired choice designs are ones where the information matrix of the optimal design has a certain balanced structure.

This, for a three-level optimal paired choice design, forces N to be a multiple of 3. Unlike the above, the present chapter adopts an approach to theoretically identify three-level A- and D-optimal designs where N is not necessarily a multiple of 3. Table 5.1 provides such new three-level A- and D-optimal designs. The optimal designs provided have distinct choice pairs. However, the base designs constructed in the current chapter can contain repeated pairs (for example, $a_{(2,9)}$). Therefore, one should always check that choice pairs in the base design are distinct. Our general method of construction using Hadamard matrices would always give rise to an optimal or efficient design having distinct choice pairs provided the base design has distinct choice pairs.

The D-optimal designs under effects coding are also A- and D-optimal under orthonormal contrasts. However, under effects coding, A-optimal designs are usually not D-optimal, even if N is a multiple of three; for example the design $a_{(2,6)}$.

While considering the paired choice designs with parameters $k = 2, N = 6$, we observe that for each of the two factors, the A-optimal design $a_{(2,6)}$ has three pairs which are either $(0, 1)$ or $(1, 0)$, two pairs which are either $(1, 2)$ or $(2, 1)$, and one pair which is either $(0, 2)$ or $(2, 0)$. Thus, it appears that under effects coding the A-optimal design $a_{(2,6)}$ attaches more importance to compare the factor-levels 0 and 1 for each of the two factors. On the contrary, for the orthonormal contrasts, the A-optimal design with the same parameters $k = 2, N = 6$ (the design being the same as the D-optimal design in G_{2004}) appears to give equal importance to the pairwise comparison between the three factor-levels since this design has two pairs of each of the 3 pair-types. This example illustrates the need for more work to understand whether one should recommend A-optimal designs for orthonormal contrasts or should one recommend A-optimal designs under effects coding.

Chapter 6

On three-level D-optimal paired choice designs

This chapter is based on the following work:

Singh (2019): Singh, Rakhi. On three-level D-optimal paired choice designs. *Statist. Probab. Lett.* *145* (2019), 127–132.

6.1 Introduction

In this chapter, we consider N paired choice sets with k three-level factors, employed in a choice experiment. We are interested in the estimation of all the main effects and all two-factor interaction effects. For k three-level factors, following Huber and Zwerina (1996), the utilities u_j are modeled as $u_j = P_j\theta$, where θ is a $\left(2k + 4\binom{k}{2}\right) \times 1$ vector representing the main and two-factor interaction effects, P_j is an $N \times \left(2k + 4\binom{k}{2}\right)$ effects-coded matrix for the jth option, and $u_j = (u_{ji})$ is an $N \times 1$ utility vector for the jth option, $j = 1, 2; i = 1, \ldots, N$. We also define $P = P_1 - P_2$ and refer to it as the design matrix of design d. For attaining theoretically optimal designs under the multinomial logit model, a utility-neutral approach (that is, taking $\theta = 0$) is in practice for finding the information matrix. Under such a utility-neutral multinomial logit model, the Fisher information matrix for a design d reduces to $(1/4)M_d$, where $M_d = P^T P$.

In this chapter, we are interested in the estimation of all the main effects and all two-factor interaction effects. As an example, interest on such main-effects and all two-factors interaction effects may arise when say, a fast-food joint wants to assess the effect of four factors (food, drinks, sides, and price) and their interactions on its marketing strategies. These four factors are say at 3 levels each: food (vegetarian, egg, and chicken), drinks (hot coffee, fruit juice, and soft-drinks), sides (fries, onion rings, and popcorn), and price $(3, 5, 7)$. In such situation, when the fast-food joint wants to assess not just the impact of these factors as main-effects but also the impact of interaction effects of each the two factors (interaction of price and food, interaction of sides and drinks, etc.) on their marketing strategies, the designs in this chapter will be useful. Designs for such estimation problems in choice experiments are studied by several authors (see, Street and Burgess (2007), Großmann and Schwabe (2015), etc.).

Graßhoff et al. (2003) provided D-optimal designs for estimating main effects and two-factor interaction effects with total number of choice pairs $N = g3^k$, where $g = \binom{k}{t^*}2^{t^*}$ when 3 does not divide $k-2$ and $g = \binom{k}{t^*}2^{t^*} + \binom{k}{t^*+1}2^{t^*+1}$, otherwise. Here, $t^* = k - 1 - [\frac{k-2}{3}]$. Street and Burgess (2007) reduced the total number of choice pairs to $N = gn$, where g is same as Graßhoff et al. (2003) and n is the size of a strength four orthogonal array on k three-level factors. Thus, Street and Burgess (2007) reduced the number of choice pairs by using an orthogonal array instead of a complete factorial design. In this chapter, we further provide a significant reduction in the number of choice pairs for such an optimal design by reducing the number of generators g. For example, for $k = 4$, currently a D-optimal design would need $N = 32n$ choice pairs, whereas we provide construction of D-optimal design in $N = 4n$ choice pairs, implying a reduction of 88% in the number of choice pairs. We provide construction of such designs for $k = 3, 4, 5, 6$ factors after obtaining generators with much reduced values of g. Using the approach of Singh et al. (2018), we also provide a way to further reduce the number of choice pairs by using orthogonal blocking methodology.

6.2 Preliminaries

In this section, we introduce some notations and discuss the existing work done in details. Let P_j for main effects and two-factor interaction effects be denoted by X_j and Y_j respectively, $j = 1, 2$. Also, let $X = X_1 - X_2$ and $Y = Y_1 - Y_2$. When our interest lies in the estimation of both the main effects and the two-factor interaction effects, the corresponding information matrix M_d under the linear paired comparison model (Graßhoff et al., 2003) is

$$M_d = P^T P = \begin{bmatrix} X^T X & X^T Y \\ Y^T X & Y^T Y \end{bmatrix}. \tag{6.1}$$

For main effects, the effects-coded vectors for levels $0, 1$ and 2 are $(1\ 0), (0\ 1)$ and $(-1\ \ -1)$, respectively. Let $X^i_{j\ell}$ represent X_j corresponding to the ith choice pair and ℓth factor. Then, ith row of X_j is $(X^i_{j1}\ X^i_{j2}\ \ldots\ X^i_{jk})$. Also, ith row of Y_j is defined as $(X^i_{j1} \otimes X^i_{j2}, X^i_{j1} \otimes X^i_{j3}, \ldots, X^i_{j(k-1)} \otimes X^i_{jk})$.

In our context, a choice design d is connected if each of the main effects and the two-factor interaction effects are estimable, and this happens if and only if M_d has rank $2k + 4\binom{k}{2} = 2k^2$. In what follows, the class of all connected paired choice designs with k three-level factors and N choice pairs is denoted by $\mathcal{D}_{k,N}$. We make use of the standard D-optimality criteria. A design that minimizes $det(M_d^{-1})$ among all designs in $\mathcal{D}_{k,N}$ is said to be D-optimal.

A design is said to be a uniform design (Graßhoff et al., 2003) if it assigns equal weight to all choice pairs with meaningful comparisons, that is, for each factor, equal weight is given to each of six choice pairs (s, t) of distinct levels, $s \neq t$. The comparison depth t in a design d is an integer such that exactly t of the k factors have different levels in both the options and in each of the choice pairs. For estimating main effects and two-factor interaction effects, Graßhoff et al. (2003) showed that the information matrix M_d in (6.1) for any uniform design d with comparison depth t can be written as

$$M_{d(t)} = \begin{pmatrix} h_1(t) I_k \otimes M_2 & 0 \\ 0 & h_2(t) I_{k(k+1)/2} \otimes M_2 \otimes M_2 \end{pmatrix} \tag{1}$$

where $M_2 = (I_2 + J_2)$, I_ℓ denotes the identity matrix of order ℓ and J_ℓ denotes the $\ell \times \ell$ matrix of all ones, and $\otimes$ denotes the Kronecker product. Also, $h_1(t) = Nt/k$ and $h_2(t) = N\frac{t}{k}(\frac{2}{3} - \frac{t-1}{2(k-1)})$, where N is the total number of choice pairs in a choice design d.

Let $t^* = k - 1 - [\frac{k-2}{3}]$ and $w^* = (t^* + 1)/(3t^* + 1)$, where $[x]$ denotes the largest integer less than or equal to x. Graßhoff et al. (2003) showed that if 3 does not divide $k - 2$, then a uniform design $d(t^*)$, which gives equal weight to all $N = g3^k = \binom{k}{t^*}2^{t^*}3^k$ choice pairs with comparison depth t^*, is D-optimal in $\mathcal{D}(k, N)$. These N choice pairs are formed by pairing each of the 3^k options to 2^{t^*} options obtained such that t^* positions in second option is different than the corresponding positions in the first option and this needs to be done for each of the $\binom{k}{t^*}$ possibilities. The information matrix for such an optimal design $d(t^*)$ is then given by $M_{d(t^*)}$. Furthermore, if 3 divides $k-2$, then an optimal design d is a combination of two uniform designs d_1 and d_2 with weights w^* and $1-w^*$ respectively. Here, d_1 and d_2 are paired choice designs with all the choice pairs having comparison depths of t^* and t^*+1, respectively. The information matrix for such an optimal design $d(t^*)$ is then given by $w^* M_{d(t^*)} + (1-w^*) M_{d(t^*+1)}$. In this case, total number of choice pairs in an optimal design are $N = g3^k = \{\binom{k}{t^*}2^{t^*} + \binom{k}{t^*+1}2^{t^*+1}\}3^k$.

An orthogonal array $OA(n, 3^k, t)$, of strength t, is an $n \times k$ array with elements from a set of 3 distinct symbols $\{0, 1, 2\}$, such that all possible combinations of symbols appear equally often as rows in every $n \times t$ subarray. Street and Burgess (2007), Demirkale et al. (2013) and Bush (2014) provided the $OA + G$ method for constructing optimal paired choice designs using orthogonal arrays and generators G. Let G be a collection of h generators $G_1, \ldots, G_h$ where $G_u = (g_{u_1}, g_{u_2}, \ldots, g_{u_k})$. The $OA + G$ method gives a paired choice design $(A, B_u), u = 1, \ldots, h$ where $A = (A_{i\ell})$ is an $OA(n_1, 3^k, t)$ and $B_u = (B^u_{i\ell})$ with $B^u_{i\ell} = (A_{i\ell} + g_{u\ell})$ reduced modulo 3, $i = 1, \ldots, n_1$, $\ell = 1 \ldots, k$, $u = 1, \ldots, h$. This method depends on the availability of the required orthogonal array, which may not always exist. For this problem, Street and Burgess (2007) reduced the total number of choice pairs to $N = gn$, where g is same as Graßhoff et al. (2003) and n is the size of a strength four orthogonal array on k three-level factors. They reduced the number of choice pairs by using an orthogonal array instead of a complete factorial design.

It is obvious that the construction method of Graßhoff et al. (2003) can also be framed as an $OA + G$ construction method where, for example, for the case when 3 does not divide $k-2$, the 3^k options act as an orthogonal array and there are $g = \binom{k}{t^*}2^{t^*}$ generators such that exactly t^* of the k factors change their levels between the two options in a pair.

6.3 New generators and construction

In this section, we first provide a general result on the required properties of generators for constructing optimal designs for estimating all main effects and two-factor interaction effects. We now give the following result, proof of which is in the Appendix C.

Theorem 6.1. *In an $OA + G$ method, starting from an $OA(n, 3^k, 4)$, let a paired choice design d_1 be constructed using*

- $\binom{k}{t^*}$ *distinct generators such that each of the generators has non-zeros in all possible t^* positions and zeros in remaining $k - t^*$ positions, when 3 does not divide $k - 2$;*

- $\binom{k}{t^*}$ *distinct generators such that each of the generators has non-zeros in all possible t^* positions and zeros in remaining $k - t^*$ positions and $\binom{k}{t^*+1}$ distinct generators such that each of the generators has non-zeros in all possible $t^* + 1$ positions and zeros in remaining $k - t^* - 1$ positions, when 3 divides $k - 2$.*

Additionally, for any two-factors, all the generators (rows) with both non zero entry should be such that they can be clubbed in several groups of two generators of the type $\{(11, 12), (11, 21), (22, 21), (22, 12)\}$. *Then, the resultant design d_1 with $N = g_S n$ is optimal for estimating all main effects and all two-factor interaction effects. Here, $g_S = \binom{k}{t^*}$ when 3 does not divide $k - 2$ and $g_S = \binom{k}{t^*} + \binom{k}{t^*+1}$ otherwise.*

Theorem 6.1 reduces N by 2^{t^*} times when 3 does not divide $k - 2$. When 3 divides $k - 2$, then the reduction is calculated using the $\frac{g_S - g}{g_S} = 1 - \frac{\binom{k}{t^*}2^{t^*} + \binom{k}{t^*+1}2^{t^*+1}}{\binom{k}{t^*} + \binom{k}{t^*+1}}$. The percentage reduction in N and the generators, using Theorem 6.1 is given in Table 6.1 for a few example of k's. The columns N_S denotes the total number of choice pairs required from Theorem 6.1 and N_{SB} denotes the number of best available choice pairs from Street and Burgess (2007). These examples have mostly been obtained by a hit-and-trial approach. Note that it is not possible to get generators of type as in Theorem 6.1 for $k = 3$ and therefore, the least number of generators needed for $k = 3$ is 6. As a next step, one may obtain the generators for higher k.

In practice, there is a pool of choice sets and respondents are allocated a random subset of choice sets (Street and Burgess, 2007) and this process is continued until all choice sets are used once. To avoid random allocation, Singh et al. (2018) provided a method to systematically split the pool of choice sets by introducing a blocking component in the model. Using their notations, they provided a break-up of N choice pairs into b blocks each of size s such that rather than showing N choice pairs to each respondent, one can show s choice pairs to b respondents. They showed that such a design is optimal under the block model.

Use of Singh et al. (2018) approach would help in further reducing the number of choice pairs to be shown to respondents at one point of time. We now provide a construction result, optimality of which, under the block model, can be easily proved on

Table 6.1: Generators for $k = 3, 4, 5, 6$

k	Generators	N_S	N_{SB}	% Reduction
3	(1,1,0) (1,0,1) (0,1,1) (1,2,0) (2,0,1) (0,1,2)	$6n$	$12n$	50%
4	(2,1,1,0) (1,1,0,2) (1,0,1,1) (0,1,2,1)	$4n$	$32n$	88%
5	(1,1,2,1,0) (1,2,1,0,1) (2,1,0,1,1) (1,0,2,2,1) (0,1,1,2,1) (2,1,1,0,0) (1,1,0,1,0) (2,0,2,2,0) (0,1,2,2,0) (2,2,0,0,1) (1,0,1,0,1) (0,2,2,0,1) (2,0,0,1,1) (0,2,0,1,2) (0,0,1,2,2)	$15n$	$160n$	91%
6	(1,1,2,1,0,0) (1,2,1,0,1,0) (2,1,0,1,1,0) (1,0,2,2,1,0) (0,1,1,2,1,0) (2,1,1,0,0,2) (1,1,0,1,0,1) (2,0,2,2,0,1) (0,1,2,2,0,1) (2,2,0,0,1,1) (1,0,1,0,1,1) (0,2,2,0,1,2) (2,0,0,1,1,1) (0,2,0,1,2,1) (0,0,1,2,2,1)	$15n$	$240n$	94%

the similar lines as in Singh et al. (2018).

Theorem 6.2. *In an $OA+G$ method, starting from an $OA(n, 3^k, 4)$, let a paired choice design d_1 be constructed as in Theorem 6.1. Then, d_2 can be obtained as a paired choice block design with $s = n$, $b = g_S = \binom{k}{t^*}$ when 3 does not divide $k-2$; or $b = g_S = \binom{k}{t^*}+\binom{k}{t^*+1}$ otherwise and d_2 is optimal under the paired choice block model of Singh et al. (2018).*

The paired choice block design d_2 is obtained from d_1 by considering the choice pairs generated using different generators as different blocks.

Example 6.3. We now give an example of a paired choice design with four 3-level factors for estimating all the main effects and all two-factor interaction effects. From Table 6.1, for $k = 4$, the number of generators is $g_S = 4$ and we use an orthogonal array $OA(81, 3^4, 4)$ for four 3-level factors with $n = 81$ runs. Using $g_S = 4$ generators, as provided in Table 6.1, and $n = 81$, the paired choice design exists in $N = 324$ choice pairs and is provided below. Now, invoking Theorem 6.2, the paired choice block design is obtained using each of the generators as a block. Here, $s = 81$, $b = g_S = 4$, and $N = 324$.

Block 1: using generator (2,1,1,0)

(0000, 2110)	(0100, 2210)	(0200, 2010)	(1000, 0110)	(1100, 0210)	(1200, 0010)	(2000, 1110)	(2100, 1210)	(2200, 1010)
(0001, 2111)	(0101, 2211)	(0201, 2011)	(1001, 0111)	(1101, 0211)	(1201, 0011)	(2001, 1111)	(2101, 1211)	(2201, 1011)
(0002, 2112)	(0102, 2212)	(0202, 2012)	(1002, 0112)	(1102, 0212)	(1202, 0012)	(2002, 1112)	(2102, 1212)	(2202, 1012)
(0010, 2120)	(0110, 2220)	(0210, 2020)	(1010, 0120)	(1110, 0220)	(1210, 0020)	(2010, 1120)	(2110, 1220)	(2210, 1020)
(0011, 2121)	(0111, 2221)	(0211, 2021)	(1011, 0121)	(1111, 0221)	(1211, 0021)	(2011, 1121)	(2111, 1221)	(2211, 1021)
(0012, 2122)	(0112, 2222)	(0212, 2022)	(1012, 0122)	(1112, 0222)	(1212, 0022)	(2012, 1122)	(2112, 1222)	(2212, 1022)
(0020, 2100)	(0120, 2200)	(0220, 2000)	(1020, 0100)	(1120, 0200)	(1220, 0000)	(2020, 1100)	(2120, 1200)	(2220, 1000)
(0021, 2101)	(0121, 2201)	(0221, 2001)	(1021, 0101)	(1121, 0201)	(1221, 0001)	(2021, 1101)	(2121, 1201)	(2221, 1001)
(0022, 2102)	(0122, 2202)	(0222, 2002)	(1022, 0102)	(1122, 0202)	(1222, 0002)	(2022, 1102)	(2122, 1202)	(2222, 1002)

Block 2: using generator (1,1,0,2)

(0000, 1102)	(0100, 1202)	(0200, 1002)	(1000, 2102)	(1100, 2202)	(1200, 2002)	(2000, 0102)	(2100, 0202)	(2200, 0002)
(0001, 1100)	(0101, 1200)	(0201, 1000)	(1001, 2100)	(1101, 2200)	(1201, 2000)	(2001, 0100)	(2101, 0200)	(2201, 0000)
(0002, 1101)	(0102, 1201)	(0202, 1001)	(1002, 2101)	(1102, 2201)	(1202, 2001)	(2002, 0101)	(2102, 0201)	(2202, 0001)
(0010, 1112)	(0110, 1212)	(0210, 1012)	(1010, 2112)	(1110, 2212)	(1210, 2012)	(2010, 0112)	(2110, 0212)	(2210, 0012)
(0011, 1110)	(0111, 1210)	(0211, 1010)	(1011, 2110)	(1111, 2210)	(1211, 2010)	(2011, 0110)	(2111, 0210)	(2211, 0010)
(0012, 1111)	(0112, 1211)	(0212, 1011)	(1012, 2111)	(1112, 2211)	(1212, 2011)	(2012, 0111)	(2112, 0211)	(2212, 0011)
(0020, 1122)	(0120, 1222)	(0220, 1022)	(1020, 2122)	(1120, 2222)	(1220, 2022)	(2020, 0122)	(2120, 0222)	(2220, 0022)
(0021, 1120)	(0121, 1220)	(0221, 1020)	(1021, 2120)	(1121, 2220)	(1221, 2020)	(2021, 0120)	(2121, 0220)	(2221, 0020)
(0022, 1121)	(0122, 1221)	(0222, 1021)	(1022, 2121)	(1122, 2221)	(1222, 2021)	(2022, 0121)	(2122, 0221)	(2222, 0021)

Block 3: using generator (1,0,1,1)

(0000, 1011)	(0100, 1111)	(0200, 1211)	(1000, 2011)	(1100, 2111)	(1200, 2211)	(2000, 0011)	(2100, 0111)	(2200, 0211)
(0001, 1012)	(0101, 1112)	(0201, 1212)	(1001, 2012)	(1101, 2112)	(1201, 2212)	(2001, 0012)	(2101, 0112)	(2201, 0212)
(0002, 1010)	(0102, 1110)	(0202, 1210)	(1002, 2010)	(1102, 2110)	(1202, 2210)	(2002, 0010)	(2102, 0110)	(2202, 0210)
(0010, 1021)	(0110, 1121)	(0210, 1221)	(1010, 2021)	(1110, 2121)	(1210, 2221)	(2010, 0021)	(2110, 0121)	(2210, 0221)
(0011, 1022)	(0111, 1122)	(0211, 1222)	(1011, 2022)	(1111, 2122)	(1211, 2222)	(2011, 0022)	(2111, 0122)	(2211, 0222)
(0012, 1020)	(0112, 1120)	(0212, 1220)	(1012, 2020)	(1112, 2120)	(1212, 2220)	(2012, 0020)	(2112, 0120)	(2212, 0220)
(0020, 1001)	(0120, 1101)	(0220, 1201)	(1020, 2001)	(1120, 2101)	(1220, 2201)	(2020, 0001)	(2120, 0101)	(2220, 0201)
(0021, 1002)	(0121, 1102)	(0221, 1202)	(1021, 2002)	(1121, 2102)	(1221, 2202)	(2021, 0002)	(2121, 0102)	(2221, 0202)
(0022, 1000)	(0122, 1100)	(0222, 1200)	(1022, 2000)	(1122, 2100)	(1222, 2200)	(2022, 0000)	(2122, 0100)	(2222, 0200)

Block 4: using generator (0,1,2,1)

(0000, 0121)	(0100, 0221)	(0200, 0021)	(1000, 1121)	(1100, 1221)	(1200, 1021)	(2000, 2121)	(2100, 2221)	(2200, 2021)
(0001, 0122)	(0101, 0222)	(0201, 0022)	(1001, 1122)	(1101, 1222)	(1201, 1022)	(2001, 2122)	(2101, 2222)	(2201, 2022)
(0002, 0120)	(0102, 0220)	(0202, 0020)	(1002, 1120)	(1102, 1220)	(1202, 1020)	(2002, 2120)	(2102, 2220)	(2202, 2020)
(0010, 0101)	(0110, 0201)	(0210, 0001)	(1010, 1101)	(1110, 1201)	(1210, 1001)	(2010, 2101)	(2110, 2201)	(2210, 2001)
(0011, 0102)	(0111, 0202)	(0211, 0002)	(1011, 1102)	(1111, 1202)	(1211, 1002)	(2011, 2102)	(2111, 2202)	(2211, 2002)
(0012, 0100)	(0112, 0200)	(0212, 0000)	(1012, 1100)	(1112, 1200)	(1212, 1000)	(2012, 2100)	(2112, 2200)	(2212, 2000)
(0020, 0111)	(0120, 0211)	(0220, 0011)	(1020, 1111)	(1120, 1211)	(1220, 1011)	(2020, 2111)	(2120, 2211)	(2220, 2011)
(0021, 0112)	(0121, 0212)	(0221, 0012)	(1021, 1112)	(1121, 1212)	(1221, 1012)	(2021, 2112)	(2121, 2212)	(2221, 2012)
(0022, 0110)	(0122, 0210)	(0222, 0010)	(1022, 1110)	(1122, 1210)	(1222, 1010)	(2022, 2110)	(2122, 2210)	(2222, 2010)

Chapter 7

Optimal two-level choice designs for estimating main and specified two-factor interaction effects

This chapter is based on the following work:
Chai et al. (2018): Chai, Feng-Shun; Das, Ashish; Singh, Rakhi. Optimal two-level choice designs for estimating main and specified two-factor interaction effects. *J. Stat. Theory Pract. 12* (2018), no. 1, 82–92.

7.1 Introduction

In this chapter, a choice experiment consists of N choice sets with each set containing m options with no repeated options in a choice set. Each option in a choice set is described by k two-level factors.

We denote αth choice set by $\mathcal{T}_\alpha = (T_{\alpha 1}, T_{\alpha 2}, \ldots, T_{\alpha m})$, where $T_{\alpha i}$ is the ith option in the αth choice set, $\alpha = 1, 2, \ldots, N$ and $i = 1, 2, \ldots, m$. Since an option in the choice set is a representation of k factors, $T_{\alpha i}$ can be written as $(i_1 i_2 \cdots i_n)_\alpha$ where i_q represents the level of the qth factor f_q in the ith option. The collection of all such choice sets $\mathcal{T}_\alpha$, $\alpha = 1, 2, \ldots, N$ is a choice design, say d, with parameters N, k and m.

For the purpose of deriving optimal designs, it is often assumed that $\theta = 0$. This indifference or the utility-neutral assumption means that the m options in a choice set are equally attractive. Großmann and Schwabe (2015) showed that for a choice design with choice set size m, the average information matrix, under the utility-neutral multinomial logit model, using Huber and Zwerina (1996) approach, is

$$C = \frac{4}{Nm^2} \sum_{1 \leq i < j \leq m} C_{ij} \tag{7.1}$$

where $C_{ij} = (P_i - P_j)'(P_i - P_j)$ is the average information matrix corresponding to options i and j and $P_i = (p'_{1i} \mid \cdots \mid p'_{Ni})'$ is the $N \times p$ effects-coded matrix for the ith option, $i = 1, \ldots, m$. This also means that the contribution of the choice set $\mathcal{T}_\alpha$ to the information matrix is equal to $4/(Nm^2)$ times the sum of the individual contributions of the $m(m-1)/2$ different *component pairs* $(T_{\alpha i}, T_{\alpha j}), 1 \leq i < j \leq m$, that $\mathcal{T}_\alpha$ contains. In next chapter (Das and Singh, 2016), we have shown that the approach of Huber and Zwerina (1996) and that of Street and Burgess (2007) are equivalent. Therefore, in this chapter, we restrict ourselves to the Huber and Zwerina (1996) approach for utility-neutral multinomial logit model.

Most of the work on optimal designs for choice experiments under the utility-neutral multinomial logit model is based on an a priori assumption that either only the main effects of the factors or the main effects and all two-factor interaction effects are to be estimated. However, in practice, there are situations where interest lies in the estimation of main plus *some* two-factor interaction effects. For example, interest on such specified two-factor interaction effects arise in situations when one or each of the two factor(s) like price and/or brand of a product interact individually with the other factors of the product. In the traditional factorial design setup, the issue of estimability and optimality in situations of this kind has been addressed by Hedayat and Pesotan (1992), Wu and Chen (1992), Hedayat and Pesotan (1997), Chiu and John (1998), Dey and Mukerjee (1999) and Dey and Suen (2002).

Street and Burgess (2012) and Großmann and Schwabe (2015) observed that there are no general results on the optimal choice designs for estimating main plus *some* two-factor interaction effects, though Street and Burgess (2007) highlighted the problem giving few examples.

For $h = 1, \ldots, k$, let F_h represent the main effects corresponding to the factor f_h, and for $h = k + (q-1)(2k-q)/2 + q' - q = q(2k-q-1)/2 + q', 1 \leqslant q < q' \leqslant k$, F_h represent the two-factor interaction effects corresponding to the factors f_q and $f_{q'}$. Note that for the two-factor interaction effects F_h corresponding to the factors f_q and $f_{q'}$, $1 \leq q < q' \leq k$, the value of h ranges from $k+1$ to $k+k(k-1)/2$. We consider a model that includes main and all two-factor interaction effects with our interest lying in the estimation of the main effects and a specified set of two-factor interaction effects where either one or each of the two factor(s) interact individually with the remaining factors. Without loss of generality, we consider the model class $\mathcal{M}_1$ where the factor f_1 interacts with each of the remaining $k-1$ factors and thus our interest lies in the k main effects $F_h, h = 1, \ldots, k$ and $k-1 = t_1$ (say) specified two-factor interaction effects $F_h, h = k+1, \ldots, 2k-1$, that is, the two-factor interactions between factor f_1 and each of f_2, f_3, $\ldots$, f_k. Similarly, we consider the model class $\mathcal{M}_2$ where each of the factors f_1 and f_2 interact with each of the remaining $k-2$ factors and thus our interest lies in the k main effects $F_h, h = 1, \ldots, k$ and $2k-4 = t_2$ (say) specified two-factor interaction effects $F_h, h = k+2, \ldots, 3k-2$, that is, the two-factor interactions between factor f_1 and each of f_3, f_4, $\ldots$, f_k and that between f_2 and each of f_3, f_4, $\ldots$, f_k. Let S denote the indices of main effects and all two-factor interaction effects, that is, $S = \{1, \ldots, k, k+1, \ldots, k+k(k-1)/2\}$ and let S_s denote the set of indices of the main effects and two-factor interaction effects under model class $\mathcal{M}_s$. Clearly, $S_s \subset S$. Also, cardinality of the model parameters in $\mathcal{M}_1$ is $|S_1| = k + t_1 = 2k-1$ and that in $\mathcal{M}_2$ is $|S_2| = k + t_2 = 3k-4$.

As noted in Großmann and Schwabe (2015), choice sets of size $m = 2, 3$, or 4 are more useful in applications since for $m \geq 5$, the increased information processing requirements affect the quality of the responses. In this chapter, we construct universally optimal choice designs for estimating main effects and the specified set of two-factor interaction effects when $m = 3$ and $m = 4$, under the assumption that all three or higher order interaction effects are absent.

7.2 The information matrix

For ith option, we partition the p parameters into the k main effects and $k(k-1)/2$ two-factor interaction effects such that $P_i = (P_{Mi} \mid P_{Ii})$. The effects coding is used as in Großmann and Schwabe (2015). Furthermore, for any two options $i < j$, we define $X_{ij} = P_{Mi} - P_{Mj}$ and $Y_{ij} = P_{Ii} - P_{Ij}$. Then, from (7.1), the information matrix corresponding to options i and j, for estimating the main plus two-factor interaction effects, is

$$C_{ij} = \begin{bmatrix} X'_{ij}X_{ij} & X'_{ij}Y_{ij} \\ Y'_{ij}X_{ij} & Y'_{ij}Y_{ij} \end{bmatrix}. \tag{7.2}$$

Since, for $s = 1, 2$, under the model class $\mathcal{M}_s$ our interest lies in the k main effects $F_h, h = 1, \ldots, k$ and the t_s specified two-factor interaction effects, let Y_{ij} be further partitioned such that $Y_{ij} = (Y_{s(1)ij}\ Y_{s(2)ij})$, where $Y_{s(1)ij}$ is a $N \times t_s$ matrix corresponding to the selected two-factor interaction effects and $Y_{s(2)ij}$ is a $N \times k(k-1)/2 - t_s$ matrix of the remaining two-factor interaction effects. Denoting summations over $1 \leq i < j \leq m$ by Σ, let $X'X = \Sigma X'_{ij}X_{ij}$, $X'Y_{s(v)} = \Sigma X'_{ij}Y_{s(v)ij}$, and $Y'_{s(u)}Y_{s(v)} = \Sigma Y'_{s(u)ij}Y_{s(v)ij}$; $u = 1, 2$, $v = 1, 2$. Then, using (7.1) and (7.2), the information matrix for estimating the main plus two-factor interaction effects is,

$$C = \frac{4}{Nm^2} \begin{pmatrix} X'X & X'Y_{s(1)} & X'Y_{s(2)} \\ Y'_{s(1)}X & Y'_{s(1)}Y_{s(1)} & Y'_{s(1)}Y_{s(2)} \\ Y'_{s(2)}X & Y'_{s(2)}Y_{s(1)} & Y'_{s(2)}Y_{s(2)} \end{pmatrix}. \tag{7.3}$$

Using (7.3), the information matrix for estimating the main effects and the specified two-factor interaction effects under model class $\mathcal{M}_s$ is $\tilde{C}_s$ where

$$\frac{Nm^2}{4}\tilde{C}_s = \begin{pmatrix} X'X & X'Y_{s(1)} \\ Y'_{s(1)}X & Y'_{s(1)}Y_{s(1)} \end{pmatrix} - \begin{pmatrix} X'Y_{s(2)} \\ Y'_{s(1)}Y_{s(2)} \end{pmatrix} (Y'_{s(2)}Y_{s(2)})^{-}(Y'_{s(2)}X \quad Y'_{s(2)}Y_{s(1)}). \tag{7.4}$$

A choice design for estimating the main effects and the specified two-factor interactions under the model $\mathcal{M}_s$ is said to be *connected* if $rank(\tilde{C}_s) = k + t_s$. In what follows, under the model $\mathcal{M}_s$, the class of all connected choice designs involving k two-level factors and N choice sets each of size m is denoted by $\mathcal{D}^{(s)}_{N,k,m}$.

For F_h, $h = 1, \ldots, k$, we define the hth positional value for the option $T_{\alpha i}$ as i_h. Also, corresponding to the two-factor interactions involving f_q and $f_{q'}$, that is, for $F_h, h = k + 1, \ldots, k + k(k-1)/2$, the hth positional value for the option $T_{\alpha i}$ is defined as $i_q + i_{q'}$ (mod 2) ($= i^*_h$, say).

For the option $T_{\alpha i}$, the hth and ℓth positional value is $(i_h i_\ell)_{h\ell}$ when both the hth and ℓth indices correspond to the main effects, that is, $h \neq \ell, h \in \{1, \ldots, n\}, \ell \in \{1, \ldots, n\}$. For the option $T_{\alpha i}$, the hth and ℓth positional value is $(i_h i^*_\ell)_{h\ell}$ when hth index corresponds to the main effect and ℓth index corresponds to the two-factor interactions, that is, $h \in \{1, \ldots, n\}, \ell \in \{k+1, \ldots, k+k(k-1)/2\}$. Similarly, for the option $T_{\alpha i}$, the hth and ℓth positional value is $(i^*_h i^*_\ell)_{h\ell}$ when hth and ℓth indices both correspond to the two-factor interactions, that is, $h \neq \ell, h \in \{k+1, \ldots, k+k(k-1)/2\}, \ell \in \{k+1, \ldots, k+k(k-1)/2\}$.

Also, for the *component pair* $(T_{\alpha i}, T_{\alpha j})$, the hth and ℓth positional value for the above three cases respectively are $(i_h i_\ell, j_h j_\ell)_{h\ell}$; $(i_h i_\ell^*, j_h j_\ell^*)_{h\ell}$; and $(i_h^* i_\ell^*, j_h^* j_\ell^*)_{h\ell}$.

For F_h and F_ℓ, $h \neq \ell$, let $N_{h\ell}^+$ and $N_{h\ell}^-$ be the total number of *component pairs* of the positional value type $(00, 11)_{h\ell}$ and $(01, 10)_{h\ell}$ respectively, across all $m(m-1)/2$ possible pairs of a choice set of size m and among all such N sets in the choice design.

Theorem 7.1. *The off-diagonal elements of $\tilde{C}$ are zero if $N_{h\ell}^+ = N_{h\ell}^-$, for $h \neq \ell, h \in \{1, \ldots, k + k(k-1)/2\}, \ell \in \{1, \ldots, k + k(k-1)/2\}$.*

Proof. It is easy to see that for $h, \ell \in S$, the exhaustive cases leading to possible values of the (h, ℓ)th entries of $X_{ij}' X_{ij}$, $X_{ij}' Y_{ij}$, $Y_{ij}' Y_{ij}$ for its associated *component pairs* $(T_{\alpha i}, T_{\alpha j})$, are

(i) Case 1: For $h \neq \ell, h \in \{1, \ldots, n\}, \ell \in \{1, \ldots, n\}$, (h, ℓ)th entry corresponding to αth choice set in $X_{ij}' X_{ij}$ is -4 if $(i_h i_\ell, j_h j_\ell)_{h\ell} \equiv (01, 10)_{h\ell}$, is 4 if $(i_h i_\ell, j_h j_\ell)_{h\ell} \equiv (00, 11)_{h\ell}$ and is 0 otherwise.

(ii) Case 2: For $h \in \{1, \ldots, n\}, \ell \in \{k + 1, \ldots, k + k(k-1)/2\}$, (h, ℓ)th entry corresponding to αth choice set in $X_{ij}' Y_{ij}$ is 4 if $(i_h i_\ell^*, j_h j_\ell^*)_{h\ell} \equiv (01, 10)_{h\ell}$, is -4 if $(i_h i_\ell^*, i_h j_\ell^*)_{h\ell} \equiv (00, 11)_{h\ell}$ and is 0 otherwise.

(iii) Case 3: For $h \neq \ell, h \in \{k+1, \ldots, k+k(k-1)/2\}$ and $\ell \in \{k+1, \ldots, k+k(k-1)/2\}$, (h, ℓ)th entry corresponding to αth choice set in $Y_{ij}' Y_{ij}$ is -4 if $(i_h^* i_\ell^*, j_h^* j_\ell^*)_{h\ell} \equiv (01, 10)_{h\ell}$, is 4 if $(i_h^* i_\ell^*, j_h^* j_\ell^*)_{h\ell} \equiv (00, 11)_{h\ell}$ and is 0 otherwise.

Applying the above three cases, proof follows from (7.4) and the definition of $N_{h\ell}^+$ and $N_{h\ell}^-$. $\square$

As a particular case of Theorem 7.1, we have the following.

Corollary 7.2. *The off-diagonal elements of $\tilde{C}_s$, $s = 1, 2$ are zero if,*
(i) $N_{h\ell}^+ = N_{h\ell}^-$, for $h \neq \ell, h \in S_s, \ell \in S_s$, and
(ii) $N_{h\ell}^+ = N_{h\ell}^-$, for $h \in S_s$ and $\ell \in S - S_s$.

In a choice set $\mathcal{T}_\alpha$, let $n_{h_\alpha} \in \{0, 1, 2, \ldots, m\}$ represents the number of options such that the hth positional value is 0. The following Theorem gives upper bound to $trace(\tilde{C}_s)$.

Theorem 7.3. *For a choice design d with N choice sets of size m, under model $\mathcal{M}_s$, an upper bound of $trace(\tilde{C}_s)$ is*

$$trace(\tilde{C}_s) \leqslant \begin{cases} 4(k + t_s) & \text{for } m \text{ even} \\ 4(k + t_s)(m^2 - 1)/m^2 & \text{for } m \text{ odd} \end{cases},$$

with equality attaining when the following two conditions are satisfied:

(i) $n_{h_\alpha} = m/2$ *for* m *even and* $n_{h_\alpha} = (m-1)/2$ *or* $(m+1)/2$ *for* m *odd for every* $h \in S_s$ *and for every choice set* $\mathcal{T}_\alpha$ *and*
(ii) $N^+_{h\ell} = N^-_{h\ell}$*, for* $h \in S_s, \ell \in S - S_s$.

Proof. Note that from (7.4), $trace(\tilde{C}_s) \leqslant (4/Nm^2)\{trace(X'X) + trace(Y'_{s(1)}Y_{s(1)})\}$ since the matrix $(Y'_{s(2)}X \quad Y'_{s(2)}Y_{s(1)})(Y'_{s(2)}Y_{s(2)})^-(Y'_{s(2)}X \quad Y'_{s(2)}Y_{s(1)})'$ is non-negative definite. Now we find the maximum possible trace of $X'X$ and $Y'_{s(1)}Y_{s(1)}$.

For $h \in \{1, \ldots, n\}$, hth diagonal entry in $X'_{ij}X_{ij}$ is 4 if $i_h - j_h = \pm 1$ and is 0 otherwise. For $h \in \{k+1, \ldots, k + k(k-1)/2\}$, hth diagonal entry in $Y'_{ij}Y_{ij}$ is 4 if $i^*_h - j^*_h = \pm 1$ and is 0 otherwise. This implies that the value of hth diagonal entry of $X'X$ and $Y'_{s(1)}Y_{s(1)}$ is non-zero when hth factor differs among two options and this happens $n_{h_\alpha}(m - n_{h_\alpha})$ times. Therefore, every choice set $\mathcal{T}_\alpha$ adds a value $4n_{h_\alpha}(m - n_{h_\alpha})$ to the hth diagonal entry of $X'X$ and $Y'_{s(1)}Y_{s(1)}$. Clearly, $4n_{h_\alpha}(m - n_{h_\alpha})$ is maximum when $n_{h_\alpha} = m/2$ for m even, and $n_{h_\alpha} = (m-1)/2$ or $(m+1)/2$ for m odd. By simple addition of $(1/m^2N)\max(4n_{h_\alpha}(m - n_{h_\alpha}))$ over all choice sets $\alpha = 1, \ldots, N$, we get

$$trace(X'X/m^2N) \leqslant \begin{cases} n & \text{for } m \text{ even} \\ n(m^2-1)/m^2 & \text{for } m \text{ odd} \end{cases}$$

and

$$trace(Y'_{s(1)}Y_{s(1)}/m^2N) \leqslant \begin{cases} t_s & \text{for } m \text{ even} \\ t_s(m^2-1)/m^2 & \text{for } m \text{ odd} \end{cases}.$$

□

Remark 7.4. *For* $m = 2$*, it is noted that the upper bound of* $trace(\tilde{C}_s)$*, as in Theorem 7.3, is not achievable. However, for* $m = 3, 4$ *it is achievable. For given* N *and* k*, with respect to choice designs with maximum* $trace(\tilde{C}_s)$*, (i) all designs with* m *even are equivalent and (ii) a design with* m *odd is always inferior to a design with* m *even.*

7.3 Construction of universally optimal designs

The criteria of *universal optimality* was introduced by Kiefer (1975*b*) and is a strong family of optimality criteria which includes $A-$, $D-$, and $E-$ criteria as particular cases. Kiefer (1975*b*) also obtained the following sufficient condition for universal optimality. *Suppose*

$d^* \in \mathcal{D}$ and $\tilde{C}_{d^*}$ satisfies (i) $\tilde{C}_{d^*}$ is scalar multiple of I_p and, (ii) trace$(\tilde{C}_{d^*}) = max_{d\in\mathcal{D}}$ trace$(\tilde{C}_d)$. Then d^* is universally optimal in $\mathcal{D}$.

We first provide a simple method for constructing universally optimal two-level choice designs with choice set size $m = 3$ and $m = 4$ under the model class $\mathcal{M}_1$.

Theorem 7.5. *Let $k = 4t - j$, where t is a positive integer and $j = 0, 1, 2, 3$. Also, given a Hadamard matrix H of order $4t$ in normal form, let H_1 be the Hadamard matrix derived from H by multiplying the first column of H by -1. Let $Z_1 = H, Z_2 = -H, Z_3 = H_1, Z_4 = -H_1$. For $w = 1, 2, 3, 4$, let A_w be respective matrices obtained by replacing every entry i $(i = 1, -1)$ of Z_w by $(1 + i)/2$, and then deleting rightmost j columns from Z_w, where $j = 4t - k, j \in \{0, 1, 2, 3\}$. Consider rows of A_w as options. Then, $d_1^{(1)} = (A_1,\ A_2,\ A_3,\ A_4)$ and $d_2^{(1)} = \begin{bmatrix} (A_1, & A_2, & A_3) \\ (A_1, & A_2, & A_4) \end{bmatrix}$ are universally optimal two-level choice design in $\mathcal{D}^{(1)}_{4t,k,4}$ and in $\mathcal{D}^{(1)}_{8t,k,3}$, respectively.*

Proof. To prove that $d_1^{(1)}$ and $d_2^{(1)}$ are universally optimal choice designs, we show that the information matrix $\tilde{C}_1$ for the designs $d_1^{(1)}$ and $d_2^{(1)}$ are of the form βI_k for some scalar β and that $d_1^{(1)}$ and $d_2^{(1)}$ maximizes *trace*$(\tilde{C}_1)$ in the respective classes of designs $\mathcal{D}^{(1)}$. First we show that for every $h \neq \ell$, $h \in S$, $\ell \in S$, the (h, ℓ)th element of the $\tilde{C}_1$ is zero. Note that the design $d_1^{(1)}$ consists of the *component pair* designs $\{(A_\delta, A_{\delta'}), 1 \leqslant \delta < \delta' \leqslant 4\}$. Denoting the *component pair* designs of $d_1^{(1)}$ by $d^{(1)}_{1(\delta\delta')}, 1 \leqslant \delta < \delta' \leqslant 4$, we now calculate $N^+_{h\ell}$ and $N^-_{h\ell}$ for the design $d_1^{(1)}$. Since H is a Hadamard matrix of order $4t$, for all $h \neq \ell$, $h \in S$, $\ell \in S$, the combinations from the set $\{(00)_{h\ell}, (11)_{h\ell}\}$ and from the set $\{(10)_{h\ell}, (01)_{h\ell}\}$ occur equally often for each of the *component pair* designs (A_1, A_2), (A_1, A_3), (A_2, A_4), (A_3, A_4), i.e., $N^+_{(\delta\delta')h\ell} = N^-_{(\delta\delta')h\ell} = 0$ or $2t$ for $(\delta, \delta') = (1, 2)$, $(1, 3)$, $(2, 4)$, $(3, 4)$, where $N^+_{(\delta\delta')h\ell}$ is the total number of pairs of the type $(00, 11)_{h\ell}$ corresponding to hth and ℓth positional values in $d^{(1)}_{1(\delta\delta')}$, and $N^-_{(\delta\delta')h\ell}$ is the total number of pairs of the type $(01, 10)_{h\ell}$ corresponding to hth and ℓth positional values in $d^{(1)}_{1(\delta\delta')}$. Furthermore, for $(\delta, \delta') = (1, 4), (2, 3)$, the respective *component pair* designs (A_1, A_4) and (A_2, A_3) have $N^+_{(\delta\delta')h\ell} = N^-_{(\delta\delta')h\ell} = 0$ or $2t$ for all $h \neq \ell$, $h \in S$ and $\ell \in S$, except $(h, k) = (h, k + h - 1)$, $h = 2, 3, \ldots, k$, i.e., (h, ℓ) corresponding to the main effects involving f_h and the two-factor interaction effects involving f_1 and f_h, $h = 2, \ldots, k$. For such (h, ℓ)'s, $N^+_{(14)h\ell} = N^-_{(23)h\ell} = 4t$, and $N^-_{(14)h\ell} = N^+_{(23)h\ell} = 0$. Therefore, using the result of Corollary 7.2 it follows that $\tilde{C}_1$ for the design $d_1^{(1)}$ has off-diagonal elements zero. The design $d_1^{(1)}$ also ensures that

$n_{h_\alpha} = 2$, for $h \in S_1$ and for every choice set. Therefore using Theorem 7.3, it follows that each of the diagonal elements of $\tilde{C}_1$ equals 4 and $trace(\tilde{C}_1)$ is maximum for the design $d_1^{(1)}$. Thus $d_1^{(1)}$ is universally optimal in $\mathcal{D}_{4t,k,4}^{(1)}$.

To establish that the design $d_2^{(1)}$ is universally optimal in $\mathcal{D}_{8t,k,3}^{(1)}$, one can see that the *component pairs* of the design are similar to the ones corresponding to $d_1^{(1)}$ and thus $\tilde{C}_1$ for the design $d_2^{(1)}$ has off-diagonal elements zero. Regarding the diagonal elements of the $\tilde{C}_1$, the design $d_2^{(1)}$ ensures that $n_{h_\alpha} = 1$ or 2, for $h \in S_1$ and for every choice set. Therefore, from Theorem 7.3, each of the diagonal elements of $\tilde{C}_1$ equals 32/9 and $trace(\tilde{C}_1)$ is maximum for $d_2^{(1)}$. Thus, the design $d_2^{(1)}$ is universally optimal in $\mathcal{D}_{8t,k,3}^{(1)}$. □

Remark 7.6. *As an alternative to $d_2^{(1)}$, if situation demands, one may consider a choice design $d_{2'}^{(1)} = \begin{bmatrix} (A_1, & A_2, & A_3) \\ (A_1^*, & A_2^*, & A_4^*) \end{bmatrix}$ with distinct options, which is also universally optimal in $\mathcal{D}_{8t,k,3}^{(1)}$. Here, for $w = 1, 2, 4$, A_w^* is obtained from A_w by adding 1 to the elements of the 2nd column of A_w, reduced mod 2.*

We now provide a simple method for constructing universally optimal two-level choice designs with choice set size $m = 3$ and $m = 4$ under the model class $\mathcal{M}_2$.

Theorem 7.7. *Let $k = 4t - j$, where t is a positive integer and $j = 0, 1, 2, 3$. Also, given a Hadamard matrix H of order $4t$ in normal form, let H_1 be the Hadamard matrix derived from H by multiplying the first column of H by -1. Let $Z_1 = (H' \;\; H_1')' = (Z_{1a}\, Z_{1b})$, where Z_{1a} is of order $8t \times 2$ and Z_{1b} is of order $8t \times (4t - 2)$. Define $Z_2 = -Z_1$, $Z_3 = (-Z_{1a}\, Z_{1b})$ and $Z_4 = -Z_3$. For $w = 1, 2, 3, 4$, let A_w be respective matrices obtained by replacing every entry i ($i = 1, -1$) of Z_w by $(1 + i)/2$, and then deleting rightmost j columns from Z_w, where $j = 4t - k, j \in \{0, 1, 2, 3\}$. Consider rows of A_w as options. Then, $d_1^{(2)} = (A_1, \;\; A_2, \;\; A_3, \;\; A_4)$ and $d_2^{(2)} = \begin{bmatrix} (A_1, & A_2, & A_3) \\ (A_1, & A_2, & A_4) \end{bmatrix}$ are universally optimal two-level choice design in $\mathcal{D}_{8t,k,4}^{(2)}$ and in $\mathcal{D}_{16t,k,3}^{(2)}$, respectively.*

Proof. On lines similar to the proof of Theorem 7.5, we first note that the design $d_1^{(2)}$ consists of the *component pair* designs $\{(A_\delta, A_{\delta'}), 1 \leqslant \delta < \delta' \leqslant 4\}$. Denoting the *component pair* designs of $d_1^{(2)}$ by $d_{1(\delta\delta')}^{(2)}, 1 \leqslant \delta < \delta' \leqslant 4$, we now calculate $N_{h\ell}^+$ and $N_{h\ell}^-$ for the design $d_1^{(2)}$. It is easy to see that for all $h \neq \ell$, $h \in S$, $\ell \in S$, the combinations from the set $\{(00)_{h\ell}, (11)_{h\ell}\}$ and from the set $\{(10)_{h\ell}, (01)_{h\ell}\}$ occur equally often for each of the *component pair* designs (A_1, A_2), (A_1, A_3), (A_2, A_4), (A_3, A_4), i.e., $N_{(\delta\delta')h\ell}^+ = N_{(\delta\delta')h\ell}^- = 0$ or $4t$

for $(\delta,\delta') = (1,2),\ (1,3),\ (2,4),(3,4)$. Furthermore, as in the proof of Theorem 7.5, for $(\delta,\delta') = (1,4),(2,3)$, the respective *component pair* designs (A_1, A_4) and (A_2, A_3) have $N^{+}_{(14)h\ell} = N^{-}_{(23)h\ell}$ and $N^{-}_{(14)h\ell} = N^{+}_{(23)h\ell}$. Therefore, using the result of Corollary 7.2, it follows that $\tilde{C}_2$ for the design $d_1^{(2)}$ has off-diagonal elements zero. The design $d_1^{(2)}$ also ensures that $n_{h_\alpha} = 2$, for $h \in S_2$ and for every choice set. Therefore using Theorem 7.3, it follows that each of the diagonal elements of $\tilde{C}_2$ equals 4 and $trace(\tilde{C}_2)$ is maximum for the design $d_1^{(2)}$. Thus $d_1^{(2)}$ is universally optimal in $\mathcal{D}^{(2)}_{8t,k,4}$.
To establish that the design $d_2^{(2)}$ is universally optimal in $\mathcal{D}^{(2)}_{16t,k,3}$, one can see that the *component pairs* of the design are similar to the ones corresponding to $d_1^{(2)}$ and thus $\tilde{C}_2$ for the design $d_2^{(2)}$ has off-diagonal elements zero. Regarding the diagonal elements of the $\tilde{C}_2$, the design $d_2^{(2)}$ ensures that $n_{h_\alpha} = 1$ or 2, for $h \in S_2$ and for every choice set. Therefore, from Theorem 7.3, each of the diagonal elements of $\tilde{C}_2$ equals $32/9$ and $trace(\tilde{C}_2)$ is maximum for $d_2^{(2)}$. Thus, the design $d_2^{(2)}$ is universally optimal in $\mathcal{D}^{(2)}_{16t,k,3}$. □

Example 7.8. *Consider a 2^{8-j} choice experiment ($j = 0,1,2,3$) conducted through 8 choice sets of size 4 each. The 2^8 ($j = 0$) choice design $d_1^{(1)}$, is universally optimal in $\mathcal{D}^{(1)}_{8,8,4}$.*

$$d_1^{(1)} = \begin{array}{|llll|}
\hline
(00000000, & 11111111, & 10000000, & 01111111) \\
(01010101, & 10101010, & 11010101, & 00101010) \\
(00110011, & 11001100, & 10110011, & 01001100) \\
(01100110, & 10011001, & 11100110, & 00011001) \\
(00001111, & 11110000, & 10001111, & 01110000) \\
(01011010, & 10100101, & 11011010, & 00100101) \\
(00111100, & 11000011, & 10111100, & 01000011) \\
(01101001, & 10010110, & 11101001, & 00010110) \\
\hline
\end{array}.$$

Deleting the last j factors we get the corresponding universally optimal design in $\mathcal{D}^{(1)}_{8,8-j,4}$, $j = 1,2,3$.
Now consider the design $d_2^{(1)}$.

$$d_2^{(1)} = \begin{array}{|lll|lll|}
\hline
(00000000, & 11111111, & 10000000) & (01101001, & 10010110, & 11101001) \\
(01010101, & 10101010, & 11010101) & (00000000, & 11111111, & 01111111) \\
(00110011, & 11001100, & 10110011) & (01010101, & 10101010, & 00101010) \\
(01100110, & 10011001, & 11100110) & (00110011, & 11001100, & 01001100) \\
(00001111, & 11110000, & 10001111) & (01100110, & 10011001, & 00011001) \\
(01011010, & 10100101, & 11011010) & (00001111, & 11110000, & 01110000) \\
(00111100, & 11000011, & 10111100) & (01011010, & 10100101, & 00100101) \\
(00111100, & 11000011, & 01000011) & (01101001, & 10010110, & 00010110) \\
\hline
\end{array}.$$

Deleting the last j factors ($j = 0, 1, 2, 3$) of the design $d_2^{(1)}$ we get the corresponding universally optimal design in $\mathcal{D}^{(1)}_{16,8-j,3}$, $j = 0, 1, 2, 3$.
Similarly, the design $d_1^{(2)}$ is universally optimal in $\mathcal{D}^{(2)}_{16,8-j,4}$ when deleting the last j factors, $j = 0, 1, 2, 3$.

$$d_1^{(2)} =$$

(00000000,	11111111,	11000000,	00111111)	(10000000,	01111111,	01000000,	10111111)
(01010101,	10101010,	10010101,	01101010)	(11010101,	00101010,	00010101,	11101010)
(00110011,	11001100,	11110011,	00001100)	(10110011,	01001100,	01110011,	10001100)
(01100110,	10011001,	10100110,	00011001)	(11100110,	00011001,	00100110,	10011001)
(00001111,	11110000,	11001111,	00110000)	(10001111,	01110000,	01001111,	10110000)
(01011010,	10100101,	10011010,	01100101)	(11011010,	00100101,	00011010,	11100101)
(00111100,	11000011,	11111100,	00000011)	(10111100,	01000011,	01111100,	10000011)
(01101001,	10010110,	10101001,	01010110)	(11101001,	00010110,	00101001,	11010110)

.

7.4 Concluding Remarks

In this chapter, we have obtained optimal two-level choice designs for estimating main effects and specified two-factor interaction effects in the model class $\mathcal{M}_s$, $s = 1, 2$. As discussed earlier, practical situations arise where factors like f_1 = "price" and/or f_2 = "brand" interact with the other important factors. Apart from all the main effects, these factors interacting with the other factors are of significance, while studying the other interactions are of less consequence in preliminary studies. As indicated earlier, there are no general results on the optimal choice designs for estimating main plus specified two-factor interaction effects in the choice design literature, though Street and Burgess (2007) highlighted the problem giving few examples.

One could argue that the optimal designs available for estimating main effects and all two-factor interactions could be used for our specific problem because of a lack of theoretical results. However, when one increases the parameters of interest (especially 2-factor interactions), theoretically obtained optimal designs usually have large number of choice sets.

As an illustration, an optimal design for 2^5 in 16 choice sets of size $m = 3$ can be obtained for estimating main effects f_q, $q = 1, \ldots, 5$ and two-factor interactions $f_1 f_{q'}$, $q' = 2, \ldots, 5$ (i.e., under the model class $\mathcal{M}_1$). However, if the remaining $10 - 4 = 6$ two-factor interactions are to be additionally estimated optimally, a D-optimal choice design with $m = 2$, suggested by Street and Burgess (2007), is available in 320 choice sets. They also suggest a design with $m = 2$ in 48 choice sets which is 91% efficient. Similarly, a D-optimal design for 2^5 choice experiment with $m = 3$ has been obtained by Burgess and Street (2003) in 1440 choice sets.

As another example, for $k = 4, m = 3$, a D-optimal design available in the literature requires 160 choice sets, while a 96.7% efficient design is available in 32 choice sets (see, Burgess and Street (2003)). However, our design for estimating main and specified two-factor interaction effects can be constructed in 8 choice sets.

Under our model, we have provided theoretical results characterizing optimal designs for any m. However, we provide optimal design constructions for more practical values of m, i.e., $m = 3$ and $m = 4$. Though Remark 7.4 guided us to not consider the case $m = 2$, nevertheless, the case for $m = 2$ still remains a relevant open problem unless one uses large designs that are optimal for estimating main and all two-factor interactions as obtained by Street and Burgess (2007).

As a way forward one can possibly extend this work for factors with asymmetric levels. One could also consider other sets of specified two-factor interaction effects as indicated in Dey and Suen (2002).

Chapter 8

A unified approach to discrete choice experiments

This chapter is based on the following work:

Das and Singh (2016): Das, Ashish; Singh, Rakhi. A unified approach to discrete choice experiments. Ongoing.

8.1 Introduction

As stated before, the primary objective of a discrete choice experiment is to study the impact of k attributes of an option where the ith attribute has v_i (≥ 2) levels labeled $0, \ldots, v_i - 1$; $i = 1, \ldots, k$. With the options being described by the levels of the attributes, each option is a k-tuple and there are a total of $L = \prod_{i=1}^{k} v_i$ possible options. Let the L lexicographically arranged options be denoted by $t_1, t_2, \ldots, t_L$ where, $t_w = (w_1 w_2 \cdots w_k), w_i = 0, 1, \ldots, v_i - 1; i = 1, 2, \ldots, k$. Here, w denotes the lexicographic number of the option t_w and is given by $w = w_1 \prod_{i=2}^{k} v_i + w_2 \prod_{i=3}^{k} v_i + \cdots + w_{k-1} v_k + w_k + 1$. A choice design is a collection of choice sets employed in a choice experiment.

Discrete choice experiments have been discussed primarily under the multinomial logit model (MNL model) setup. Street and Burgess (2007) have derived the information matrix to study choice experiments under the MNL model. Independently, Huber and Zwerina (1996) have also derived the information matrix under the same model. We refer to the two approaches as SB approach and HZ approach. The information matrices under the two approaches look superficially different and moreover, the attributes are also coded differently, even though we have considered the average information matrix in both the SB and the HZ approach.

One of the many objectives of a choice experiment is to optimally or efficiently estimate the parameters of interest which essentially consists of either only the main effects or the main plus two-factor interaction effects of the k attributes. D-optimal designs have been obtained theoretically under the utility-neutral setup, for example, see Graßhoff et al. (2004), Street and Burgess (2007), Demirkale et al. (2013), and Singh et al. (2015). Additionally, Sun and Dean (2016), Sun and Dean (2017), and Chai et al. (2017) have obtained A-optimal choice designs. For such optimal designs, researchers either used the information matrix following the HZ approach under effects coding, or used the information matrix following the SB approach under orthonormal coding.

The author-groups SB and HZ have used seemingly different information matrices under the MNL model. There have also been some confusion regarding the inference parameters expressed as linear functions of the utility parameter vector τ. We theoretically establish a unified approach to discrete choice experiments and introduce the general inference problem in terms of a simple linear function of τ; say $M\tau$.

After introducing the SB and HZ approaches in Section 8.2, we show their equivalence in Section 8.3. In Section 8.4, the inference problem under different codings is expressed in simple terms as a function of τ. In Section 8.5, with respect to the A-criterion, we discuss how different codings may be interpreted. We also propose a related coding which is appropriate for test-control discrete choice experiments wherein some new test levels of an attribute are compared with an existing control level. Finally, in Section 8.6 we summarize the results along with a short discussion.

8.2 The SB and HZ approaches

We now discuss the two approaches, SB and HZ, which have generally been used in the theory of choice experiments.

In a choice design, we denote the nth choice set by $T_n = (t_{(n1)}, t_{(n2)}, \ldots, t_{(nm)})$, where $t_{(nj)}$ is the jth option in the nth choice set, $n = 1, \ldots, N$, $j = 1, \ldots, m$. Each option $t_{(nj)} = t_w$ for some $w \in \{1, \ldots, L\}$, where w, as mentioned before, is the lexicographic number of the option. Corresponding to the jth option in the N choice sets, let $A_j =$

$(t_{(1j)}^T\ t_{(2j)}^T \cdots\ t_{(Nj)}^T)^T$ be a $N \times k$ matrix representing the levels of the k attributes.

8.2.1 SB approach

Let S_n be the set of m lexicographic numbers w corresponding to $T_n = (t_{(nj)}), j = 1, \ldots, m$. Thus, a respondent α assigns some utility $U_{w\alpha}$ to t_w, $w \in S_n, n = 1, \ldots, N$. In a choice experiment, it is assumed that each respondent chooses the option having maximum utility among the other options in a choice set. The respondent chooses t_w in T_n if $U_{w\alpha} > U_{w'\alpha}$, $t_{w'}$ being the other options in T_n. The systematic component of the utility that can be captured, is denoted by the utility parameter $\tau_{w\alpha}$ and that $U_{w\alpha} = \tau_{w\alpha} + \epsilon_{w\alpha}$. If $\epsilon_{w\alpha}$ is independently and identically Gumbel distributed, then the choice model is the MNL model. In this chapter, as is generally the case in the literature, we assume that the respondents are alike and this assumption allows us to drop the subscript α. Following Street and Burgess (2007), the probability of choosing $t_{(nj)}$ (which is say, t_w) from T_n is then given by,

$$P_{nj} = P_w = P(U_w > U_{w'}, \text{for all } w' \neq w \in S_n) = \frac{e^{\tau_w}}{\sum_{w' \in S_n} e^{\tau_{w'}}}.$$

Let $\mathcal{I}(M\tau)$ denote the Fisher information matrix for a linear function $M\tau$ in the MNL model with utility parameter vector $\tau = (\tau_1 \cdots \tau_L)^T$. Street and Burgess (2007), using the approach of El-Helbawy and Bradley (1978), gives the Fisher information matrix for estimating p parameters of interest $B_O\tau$ ($= \beta_O$, say), where B_O is a $p \times L$ orthonormal contrast matrix corresponding to the p parameters of interest. The Fisher information matrix for $B_O\tau$, as obtained by them, is

$$\mathcal{I}(B_O\tau) = B_O \Lambda B_O^T, \tag{8.1}$$

where Λ is the information matrix for τ. For a choice design with N choice sets each containing m options, under the MNL model, the $L \times L$ information matrix for τ is (Street and Burgess, 2007, p. 81) $\Lambda = (\Lambda_{(r,r')})$, where

$$\Lambda_{(r,r')} = \begin{cases} \frac{1}{N} \sum\limits_{n \in T^r} \frac{e^{\tau_r}\left(\sum_{l \neq r \in S_n} e^{\tau_l}\right)}{\left(\sum_{l \in S_n} e^{\tau_l}\right)^2}, & r = r', r = 1, \ldots, L, \\ \\ -\frac{1}{N} \sum\limits_{n \in T^{rr'}} \frac{e^{\tau_r} e^{\tau_{r'}}}{\left(\sum_{l \in S_n} e^{\tau_l}\right)^2}, & r \neq r', r, r' = 1, \ldots, L, \end{cases} \tag{8.2}$$

with T^r being the subset of indices of choice sets containing t_r, and $T^{rr'}$ being the subset of indices of choice sets containing both t_r and $t_{r'}$. Here, the set of indices of choice set is $\{1, \ldots, N\}$ and note that $\Lambda_{(r,r')} = 0$ when $T^{rr'}$ is an empty set and $\Lambda_{(r,r)} = 0$ when T^r is an empty set. Under the utility-neutral MNL model assumption of all options being equally attractive, $\Lambda_{(r,r')}$ reduces to

$$m^2 N \Lambda_{(r,r')} = \begin{cases} -a_{r,r'}, & r \neq r', \\ (m-1)a_r, & r = r', \end{cases} \tag{8.3}$$

where a_r is the number of times t_r appears in the choice design and $a_{r,r'}$ is the number of times options t_r and $t_{r'}$ appear together in the design. In what follows, unless otherwise stated, Λ will refer to Λ as in (8.2).

8.2.1.1 Coding under SB approach

The choice designs are studied by Street and Burgess (2007) under orthonormal coding. The columns of the $p \times L$ matrix B_O correspond to the orthonormal coding for the L options arranged in lexicographic order. We are interested in estimating $B_O \tau = \beta_O$ under this approach.

As an example, for main effects, the $p_M \times L$ matrix B_O with $p_M = \sum_{i=1}^{k}(v_i - 1)$ is,

$$B_O = \begin{pmatrix} B_o^{(1)} \otimes \frac{1}{\sqrt{v_2}} 1_{v_2}^T \otimes \cdots \otimes \frac{1}{\sqrt{v_k}} 1_{v_k}^T \\ \frac{1}{\sqrt{v_1}} 1_{v_1}^T \otimes B_o^{(2)} \otimes \cdots \otimes \frac{1}{\sqrt{v_k}} 1_{v_k}^T \\ \vdots \\ \frac{1}{\sqrt{v_1}} 1_{v_1}^T \otimes \frac{1}{\sqrt{v_2}} 1_{v_2}^T \otimes \cdots \otimes B_o^{(k)} \end{pmatrix} \tag{8.4}$$

where $B_o^{(i)}$ is a $(v_i - 1) \times v_i$ matrix, the v_i columns of which define an orthonormal coding for the v_i levels of the ith attribute, that is, $B_o^{(i)} B_o^{(i)T} = I_{v_i - 1}$ and $B_o^{(i)} 1_{v_i} = 0$. Here, 1_s is a $s \times 1$ vector of all ones, I_s is an identity matrix of order s, and $\otimes$ denotes the Kronecker product.

The subscript O in B_O and β_O implies the orthonormal coding.

8.2.2 HZ approach

In the marketing literature, under the MNL model, usually a different approach of Huber and Zwerina (1996), following the seminal work of McFadden (1974), is followed. In this approach, the utilities corresponding to $t_{(nj)}$ are modelled as $U_{nj} = h_{nj}\beta_H + \epsilon_{nj}$, where h_{nj} is a general coded row vector of order p characterizing $t_{(nj)}$ based on a general coding of the

k attributes and β_H is a column vector of order p representing the estimable parameters of interest. Then, the probability of choosing the jth option from the nth choice set is given by,

$$P_{nj} = \frac{e^{h_{nj}\beta_H}}{\sum_{j'=1}^{m} e^{h_{nj'}\beta_H}}. \tag{8.5}$$

The information matrix for β_H, as given in Huber and Zwerina (1996), is

$$\mathcal{I}(\beta_H) = \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{m}(h_{nj} - \sum_{j'=1}^{m} h_{nj'}P_{nj'})^T P_{nj}(h_{nj} - \sum_{j'=1}^{m} h_{nj'}P_{nj'}). \tag{8.6}$$

Also, under the utility-neutral MNL model (that is, taking $\beta_H = 0$, or equivalently by assuming $P_{nj} = 1/m$ for all n and j), the information matrix reduces to the sum of $m(m-1)/2$ matrices corresponding to all possible pairs of options, that is,

$$\mathcal{I}(\beta_H) = \frac{1}{Nm^2}\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m} X_{H(jj')}^T X_{H(jj')}. \tag{8.7}$$

where $X_{H(jj')} = H_j - H_{j'}$, is the $N \times p$ difference matrix for the jth and j'th options; the $N \times p$ matrix $H_j = (h_{1j}^T\ h_{2j}^T\ \cdots\ h_{Nj}^T)^T$ corresponds to the A_j, $j = 1, \ldots, m$ and is coded with a general coding.

8.2.2.1 Coding under HZ approach: General coding

The derivation of the information matrix in McFadden (1974) and subsequently used by Huber and Zwerina (1996) is based on a general coding. Rows of the matrix H_j consists of general coded vectors h_{nj} having a one-to-one correspondence with the choice sets $t_{(nj)}$. Each of the L options can be coded according to the general coding. We now define the $p \times L$ general coded matrix B_H, columns of which correspond to the general coding corresponding to the L options arranged in lexicographic order. Corresponding to A_j involving $t_{(nj)}$, $n = 1, \ldots, N$, the nth row of H_j is the wth column of B_H, with w being the lexicographic number of the option $t_{(nj)}$.

As an example, for main effects, the $p_M \times L$ general coded matrix B_H is

$$B_H = \begin{pmatrix} B_h^{(1)} \otimes 1_{v_2}^T \otimes \cdots \otimes 1_{v_k}^T \\ 1_{v_1}^T \otimes B_h^{(2)} \otimes \cdots \otimes 1_{v_k}^T \\ \vdots \\ 1_{v_1}^T \otimes 1_{v_2}^T \otimes \cdots \otimes B_h^{(k)} \end{pmatrix}, \tag{8.8}$$

where $B_h^{(i)}$ is a $(v_i - 1) \times v_i$ matrix having full row rank, v_i columns of which define a general coding for the v_i levels of the ith attribute. For an attribute, the v_i columns of $B_h^{(i)}$ represent a general coding for the respective v_i levels labeled $0, \ldots, v_i - 1$.

In what follows, the subscript H in B_H and β_H implies the general coding.

8.2.2.2 Coding under HZ approach: Effects coding

The coding that is generally used in the marketing literature for describing attributes is more formally known as effects coding (see, Großmann and Schwabe (2015)). Corresponding to each option $t_{(nj)}$, we denote the effects coded vector by e_{nj}. For effects coding, we also use B_E and e_{nj} to denote B_H and h_{nj} respectively. The L columns in B_E represent the effects coding for the L lexicographically arranged options. This simply means that, the wth column of B_E is the effects coding for t_w. Under effects coding, the information matrix for the parameter of interest β_E is

$$\mathcal{I}(\beta_E) = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{m} (e_{nj} - \sum_{j'=1}^{m} e_{nj'} P_{nj'})^T P_{nj} (e_{nj} - \sum_{j'=1}^{m} e_{nj'} P_{nj'}), \tag{8.9}$$

where e_{nj} is the effects coded vector corresponding to $t_{(nj)}$.

As an example, for main effects, the effects coding for level labeled l is represented by a unit row vector of length $v_i - 1$ with 1 in the $(l+1)$th position for $l = 0, \ldots, v_i - 2$, and the effects coding for level labeled $v_i - 1$ is represented by $-1_{v_i-1}^T$, $i = 1, \ldots, k$. For example, the effects coded vectors for one factor at three levels $l = 0, 1, 2$ are $(1\ 0), (0\ 1)$ and $(-1\ \ -1)$, respectively. Now for estimation of the main effects, B_E is of the same form B_H given in (8.8) with $B_h^{(i)}$ replaced by $B_e^{(i)}$. For obtaining $B_e^{(i)}$, effects coding corresponding to level l is put as the $(l+1)$th column of $B_e^{(i)}$. For example, for $v_i = 3$, $B_e^{(i)} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$ since the effects coded vectors for $l = 0, 1, 2$ are $(1\ 0), (0\ 1)$ and $(-1\ \ -1)$, respectively.

In what follows, the subscript E in B_E and β_E implies the effect coding.

8.3 Equivalence of SB and HZ approaches

The mathematical derivations used while obtaining the information matrices under the SB approach and under the HZ approach appears to be somewhat different. This difference in the mathematical derivations of the variance-covariance matrix has resulted in significant confusion within the literature (see, for example Rose and Bliemer (2014)). The major differences in the two approaches are: (a) the expressions for the information matrices under the two approaches appear to be different, and (b) coding of levels in the two approaches are different.

Under the utility-neutral effects coding setup, the similarity of the two approaches for a D-optimal paired ($m = 2$) choice design was first addressed by Großmann and Schwabe (2015) (pp. 793). We now show how for any general coded matrix B_H and under a general setup, the two seemingly different structures of the information matrices are related. Below is the equivalence result, a proof of which is in the Appendix D.

Theorem 8.1. *For a general coded matrix B_H, the information matrix $\mathcal{I}(\beta_H)$, under the MNL model, satisfies $\mathcal{I}(\beta_H) = \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{m}(h_{nj} - \sum_{j'=1}^{m} h_{nj'}P_{nj'})^T P_{nj}(h_{nj} - \sum_{j'=1}^{m} h_{nj'}P_{nj'}) = B_H \Lambda B_H^T$, where Λ and P_{nj} are as defined in (8.2) and (8.5), respectively.*

Theorem 8.1 shows that once the coded matrix B_H is decided, the two expressions of the information matrix, which appear different, are in fact the same. This also implies that the two seemingly independent derivations result in the same information matrix.

In what follows, $Var(\hat{\beta}_H)$ represents the asymptotic covariance matrix of the maximum likelihood estimator of β_H and is inverse of the information matrix $\mathcal{I}(\beta_H)$. Similar statement holds for β_O and β_E when $B_H = B_O$ and $B_H = B_E$ respectively. The following are two special cases of Theorem 8.1.

Corollary 8.2. *Under the MNL model,*

(i) For orthonormally coded matrix B_O, the information matrix $\mathcal{I}(\beta_O) = B_O \Lambda B_O^T = \mathcal{I}(B_O \tau)$. In other words, the variance $Var(\hat{\beta}_O) = (B_O \Lambda B_O^T)^{-1} = Var(B_O \hat{\tau})$.

(ii) For effects coded matrix B_E, the information matrix $\mathcal{I}(\beta_E) = B_E \Lambda B_E^T$. In other words, the variance $Var(\hat{\beta}_E) = (B_E \Lambda B_E^T)^{-1}$.

Proof. For orthonormal coding B_O, Street-Burgess derived that $\mathcal{I}(B_O \tau) = B_O \Lambda B_O^T$. Additionally, from Theorem 8.1, $B_O \Lambda B_O^T = \mathcal{I}(\beta_O)$. Therefore, for orthonormal coding, Corollary 8.2(i) holds. It is easy to see that (ii) is just a special case of Theorem 8.1 for effects coding. $\square$

Note that in Corollary 8.2(i), we indicate that $Var(\hat{\beta}_O) = Var(B_O\hat{\tau})$. However, in Corollary 8.2(ii), we mention the variance only of $\hat{\beta}_E$, not mentioning it in terms of $\hat{\tau}$. We elaborate on the same in the next section.

8.4 Inference problem in terms of utility parameters

In the previous section, we have established that for any general coded matrix B_H, the information matrix of β_H is $\mathcal{I}(\beta_H) = B_H \Lambda B_H^T$. This shows the equivalence of the SB and the HZ approaches. For the SB approach, where β_H has been taken as β_O with $B_H = B_O$, the parameter β_O is expressible in terms of the utility parameter vector τ through $B_O\tau$. However, in the effects coded HZ approach, $\beta_H = \beta_E$ is the parameter of interest. Under such an effects coding setup B_E, we need to understand what β_E is in terms of utility parameter vector τ. In this section, we express the general inference problem β_H in terms of utility parameter vector τ.

In comparative experiments, like the choice experiments, the problem of estimation for inferring $M\tau$ may be specified as $\Pi : \beta_H = M\tau$, where M is a $p \times L$ matrix with $M1_L = 0$, τ is a utility parameter vector and 1_L is a column vector of size L having all 1s. Thus β_H contains p parametric contrasts. With reference to Π, we call a design d as *acceptable* if all components of β_H are estimable using d. Let $\mathcal{D}_\Pi$ be the class of all acceptable designs with reference to the problem Π. The problem Π is referred to as (i) *non-singularly estimable* if and only if $rank(M) = p \leq \sum(v_i - 1)$, and, more explicitly as (ii) *non-singularly estimable full-rank problem* if and only if $rank(M) = p = \sum(v_i - 1)$. Furthermore, when $rank(M) = \sum(v_i - 1) < p$, we refer to Π as *singularly estimable full-rank problem.*

We first examine the two different functions of τ that are available in the literature, one given by Großmann and Schwabe (2015) and the other by Street and Burgess (2007). For notational clarity, as and when required, Λ corresponding to a design d is denoted by Λ_d. A generalized inverse of a matrix A is denoted by A^-, while the MoorePenrose inverse is denoted by A^+.

Großmann and Schwabe (2015) have indicated that under the utility-neutral setup, for D-optimal balanced paired choice designs d^*, the information matrix for β_E is,

$$\mathcal{I}(\beta_E) = \mathcal{I}(SG\tau) = (S(G\Lambda_{d^*}G^T)^- S^T)^-. \tag{8.10}$$

Here, G is obtained from (8.4) by replacing every $(v_i - 1) \times v_i$ matrix $B_o^{(i)}$ with the $v_i \times v_i$ centering matrix $K_{v_i} = I_{v_i} - \frac{1}{v_i} J_{v_i}$ and S is a rectangular block diagonal matrix with the ith $(v_i - 1) \times v_i$ diagonal block being $D_{v_i} = (\sqrt{\frac{v_i}{L}} I_{v_i - 1}, 0_{v_i - 1 \times 1})$, $i = 1, \dots, k$.

Remark 8.3. *From the proof of the expression for the information matrix in* (8.10) *(Großmann and Schwabe, 2015, pp. 798), it follows that the Moore-Penrose inverse of* $G\Lambda_{d^*}G^T$ *should be used rather than a generalized inverse. Therefore, the correct version of their expression (as (8.10), above) for the information matrix of the estimable parameter vector* β_E *is*

$$\mathcal{I}(\beta_E) = \mathcal{I}(SG\tau) = (S(G\Lambda_{d^*}G^T)^+ S^T)^{-1}. \tag{8.11}$$

Contrary to Großmann and Schwabe's identification of the matrix SG for the inference problem β_E expressed as a function of τ, Street and Burgess (2007) (pp. 77–78) gives an impression, through an example, that under the utility-neutral setup,

$$\mathcal{I}(\beta_E) = \mathcal{I}(B_E\tau) = B_E \Lambda_{d^*} B_E^T. \tag{8.12}$$

We now study an example of a D-optimal paired choice design.

Example 8.4. *For* $k = 2, v_1 = v_2 = 3$ *and* $N = 9$*, consider a* D*-optimal paired choice design* d^*

$$d^* = \{(00,\ 11), (10,\ 21), (20,\ 01), (01,\ 12), (11,\ 22), (21,\ 02), (02,\ 10), (12,\ 20), (22,\ 00)\}.$$

Then, from (8.11),

$$\mathcal{I}(SG\tau) = (S(G\Lambda_{d^*}G^T)^+ S^T)^{-1} = \begin{pmatrix} 0.50 & 0.25 & 0 & 0 \\ 0.25 & 0.50 & 0 & 0 \\ 0 & 0 & 0.50 & 0.25 \\ 0 & 0 & 0.25 & 0.50 \end{pmatrix} = A, \textit{say}.$$

Again, from (8.12)*, it is easy to see that*

$$B_E \Lambda_{d^*} B_E^T = A.$$

Therefore, here it follows that $\mathcal{I}(SG\tau) = (S(G\Lambda_{d^*}G^T)^+ S^T)^{-1} = B_E \Lambda_{d^*} B_E^T$*. From the impression created by Street and Burgess (2007), it follows that* $B_E \Lambda_{d^*} B_E^T = \mathcal{I}(B_E\tau)$ *and therefore* $\mathcal{I}(SG\tau) = \mathcal{I}(B_E\tau)$*. Also from Theorem 8.1, we have that* $B_E \Lambda_{d^*} B_E^T = \mathcal{I}(\beta_E)$*. Therefore, it seems that the inference problem* β_E *expressed in terms of* τ *is* $SG\tau$ *as well as* $B_E\tau$*. However,* $B_E \neq SG$ *since*

$$B_E = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 \end{pmatrix} \quad \textit{and} \quad SG = \frac{1}{9}\begin{pmatrix} 2 & 2 & 2 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 2 & 2 & 2 & -1 & -1 & -1 \\ 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & -1 \\ -1 & 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 \end{pmatrix}.$$

In the above example, though we have taken a specific D-optimal design, we are able to highlight the lack of clarity on what β_E is in terms of τ. Is it that $\beta_E = B_E\tau$ or $\beta_E = SG\tau$ or is it something else?

In the result that follows, for a general coded matrix B_H, we obtain the corresponding inference problem for inferring on β_H in terms of the utility parameter vector τ. The result has been obtained under a more general unrestricted setup for any design d with $m \geq 2$ and no restriction of utility-neutrality. Proof of the following result is in the Appendix D.

Theorem 8.5. *Under the MNL model, for a general coded matrix B_H,*
(i) $Var(B_H\hat{\tau}) = (B_H B_H^T)(B_H \Lambda B_H^T)^{-1}(B_H B_H^T)$, and
(ii) $Var((B_H B_H^T)^{-1} B_H \hat{\tau}) = (B_H \Lambda B_H^T)^{-1} = Var(\hat{\beta}_H)$.

As a consequence of Theorem 8.5, we have $\mathcal{I}(B_H\tau) = (B_H B_H^T)^{-1}(B_H\Lambda B_H^T)(B_H B_H^T)^{-1}$ and $\mathcal{I}((B_H B_H^T)^{-1} B_H \tau) = B_H \Lambda B_H^T = \mathcal{I}(\beta_H)$. We now have the following Corollary as special cases when (i) $B_H = B_O$, (ii) $B_H = B_E$ and (iii) $B_H = (B_E B_E^T)^{-1} B_E$.

Corollary 8.6. *Under the MNL model, the following holds.*

(i) $Var(B_O\hat{\tau}) = (B_O\Lambda B_O^T)^{-1} = Var(\hat{\beta}_O)$; and $\mathcal{I}(B_O\tau) = B_O\Lambda B_O^T = \mathcal{I}(\beta_O)$.

(ii) $Var((B_E B_E^T)^{-1} B_E\hat{\tau}) = (B_E\Lambda B_E^T)^{-1} = Var(\hat{\beta}_E)$; and $\mathcal{I}((B_E B_E^T)^{-1} B_E\tau) = B_E\Lambda B_E^T = \mathcal{I}(\beta_E)$,

(iii) $Var(B_E\hat{\tau}) = (B_E B_E^T)(B_E\Lambda B_E^T)^{-1}(B_E B_E^T)$; and $\mathcal{I}(B_E\tau) = (B_E B_E^T)^{-1}(B_E\Lambda B_E^T)(B_E B_E^T)^{-1}$.

Under orthonormal coding (Corollary 8.6(i)), the inference problem is $B_O\tau$ with $\mathcal{I}(B_O\tau) = B_O\Lambda B_O^T$. Corollary 8.6(ii) shows that under the usual effects coding setup, the inference problem is not $B_E\tau$ (as illustrated by Street and Burgess (2007), pp. 77–78) but the correct inference problem is $(B_E B_E^T)^{-1} B_E\tau$ with $\mathcal{I}((B_E B_E^T)^{-1} B_E\tau) = B_E\Lambda B_E^T$. Finally, the inference problem $B_E\tau$ (Corollary 8.6(iii)) implies test-control setup with $\mathcal{I}(B_E\tau) = (B_E B_E^T)^{-1}(B_E\Lambda B_E^T)(B_E B_E^T)^{-1}$. We discuss more on the test-control setup in Section 8.5.

It is also easy to establish directly from the definition of the matrices B_E, S and G that $(B_E B_E^T)^{-1} B_E = SG$ (proof in Appendix D), and thus, $B_E = (SGG^T S^T)^{-1} SG$.

Thus, another form of the inference problem $SG\tau$ as in (8.11), when written as a function of effects coded matrix, is $(B_E B_E^T)^{-1} B_E \tau$. However, referring to (8.11), we see that $\mathcal{I}(\beta_E) = \mathcal{I}(SG\tau) = (S(G\Lambda_{d^*}G^T)^+ S^T)^{-1}$ holds only for D-optimal paired choice designs d^* of Graßhoff et al. (2004), while $\mathcal{I}(\beta_E) = \mathcal{I}(SG\tau) = \mathcal{I}((B_E B_E^T)^{-1} B_E \tau) = B_E \Lambda B_E^T$ holds for any arbitrary choice design. Using the identity $B_E = (SGG^T S^T)^{-1} SG$, it thus follows that for any arbitrary choice design, $\mathcal{I}(\beta_E) = \mathcal{I}(SG\tau) = \mathcal{I}((B_E B_E^T)^{-1} B_E \tau) = B_E \Lambda B_E^T = (SGG^T S^T)^{-1} SG\Lambda G^T S^T (SGG^T S^T)^{-1}$.

The most commonly studied optimal designs under the choice experiments are the D-optimal designs. For a general full-rank inference problem $M\tau$, the D-optimality criterion is invariant to the choice of M. However, for some other criterion, different M may lead to different optimal designs (see, Morgan and Stallings (2014)). For example, the A-optimal designs are generally different for the two non-singular full-rank inference problems $\beta_O = B_O \tau$ and $\beta_E = (B_E B_E^T)^{-1} B_E \tau$. When our interest lies in the estimation of main effects, it is important to understand the preferred inference problem. Recently, Sun and Dean (2016) and Sun and Dean (2017) have obtained A-optimal designs under orthonormal coding, while Chai et al. (2017) have obtained three-level A-optimal designs both under effects coding as well as under orthonormal coding. Chai et al. (2017) show that the A-optimal designs under the effects coding are different than those under the orthonormal coding. Their non-singular full-rank inference problem for the A-optimal designs under effects coding is $(B_E B_E^T)^{-1} B_E \tau$, whereas the non-singular full-rank inference problem under the orthonormal coding is $B_O \tau$. We discuss more on this in Section 8.5.

Before we conclude this section, as an application to the equivalence of the SB and the HZ approaches, we generalize the paired choice D-optimality results of Graßhoff et al. (2004) for $m \geq 2$. Street and Burgess (2007) have provided sufficiency conditions to obtain D-optimal choice designs for $m \geq 2$ under the utility-neutral setup. Since D-optimality criterion is invariant with respect to reparameterizations, we now provide the corresponding D-optimality results for estimating $\beta_E = (B_E B_E^T)^{-1} B_E \tau$ and $B_E \tau$. Proof of the following result is in the Appendix D.

Theorem 8.7. *Let d^* be a D-optimal choice design, under the utility-neutral setup, for estimating the main effects $B_O \tau$ and $\mathcal{I}_{d^*}(B_O \tau) = diag(\alpha_1 I_{v_1 - 1}, \ldots, \alpha_k I_{v_k - 1})$, where $\alpha_i = \frac{2 v_i S_i}{m^2 L (v_i - 1)}$ and S_i is as in Street and Burgess (2007)[Theorem 6.3.1]. Then, d^* is also D-optimal under the HZ approach for inferring on β_E and $B_E \tau$. Furthermore, the* *respective information matrices for inferring on β_E and $B_E\tau$ are:*

(i) $\mathcal{I}_{d^*}(\beta_E) = \mathcal{I}_{d^*}((B_E B_E^T)^{-1} B_E\tau) = diag(\alpha_1 V_1, \ldots, \alpha_k V_k)$*, and*

(ii) $\mathcal{I}_{d^*}(B_E\tau) = diag(\alpha_1 V_1^{-1}, \ldots, \alpha_k V_k^{-1})$,

where $V_i = \frac{L}{v_i}(I_{v_i-1} + J_{v_i-1})$ *and* $V_i^{-1} = \frac{1}{L}(v_i I_{v_i-1} - J_{v_i-1})$.

8.5 Inference problem under A-optimality

In this section, we consider different inference problems for optimal estimation of k main effects corresponding to the k attributes with the ith attribute at v_i levels. For the ease of understanding, we first discuss the result for one attribute (the ith attribute) and then generalize it for k attributes.

Let $Z_{(i,u)} = (-I_u \;\; 1_u \;\; 0_{u\times(v_i-u-1)})$ for $u = 1, \ldots, v_i - 1$. Then, the $\binom{v_i}{2} \times v_i$ coefficient matrix of all normalized elementary comparisons between the v_i levels of the ith attribute is

$$Z_i = \frac{1}{\sqrt{2}} \begin{pmatrix} Z_{(i)} \\ \bar{Z}_{(i)} \end{pmatrix},$$

where $Z_{(i)} = (Z_{(i,1)}^T \; Z_{(i,2)}^T \; \cdots \; Z_{(i,v_i-2)}^T)^T$ and $\bar{Z}_{(i)} = Z_{(i,v_i-1)}$. Note that $\bar{Z}_{(i)}$ is a contrast matrix for comparing level labeled $v_i - 1$ of ith attribute to each of the remaining $v_i - 1$ levels labeled $0, 1, \ldots, v_i - 2$. Similarly, $Z_{(i)}$ is a contrast matrix for pairwise comparisons of the levels labeled $0, 1, \ldots, v_i - 2$ of attribute i.

For the inference problem $Z_i\tau$, the matrix Z_i represents all normalized elementary comparisons between levels of the ith attribute. Then, as a consolidated measure of goodness of a choice design, the sum of the variances of the estimates of all normalized elementary comparisons is given by

$$tr[Var(Z_i\hat{\tau})]. \tag{8.13}$$

For measuring goodness of a choice design, (8.13) ensures that equal importance is provided to each of the $\binom{v_i}{2}$ elementary comparisons of which only $v_i - 1$ comparisons are independent. We now show the relationship between the sum of the variances of the estimates of all normalized elementary comparisons and the sum of variances of the estimates under different inference problems. We first have the following theorem whose proof we discuss later.

Theorem 8.8. *Under the main effects model, for attribute i,*

$$tr[Var(B_o^{(i)}\hat{\tau})] = \frac{2}{v_i} tr[Var(Z_i\hat{\tau})].$$

It follows from Theorem 8.8 that under orthonormal coding, for attribute i, the average variance of $B_o^{(i)}\hat{\tau}$ is proportional to the average variance of $\binom{v_i}{2}$ elementary comparisons among the levels of the attribute.

Huber and Zwerina (1996) (pp. 309), for minimizing errors around the estimated parameter $\hat{\beta}_H$, considered the A-criterion under the non-singular full-rank setup. Accordingly, we first focus on the non-singular full-rank inference problem. As in Theorem 8.8, we first consider a single attribute setup, keeping the general result and its proof for a later theorem.

Theorem 8.9. *For attribute i, under a non-singular full-rank inference problem,*

(i) $tr[Var((B_e^{(i)}B_e^{(i)T})^{-1}B_e^{(i)}\hat{\tau})] = \frac{2}{v_i^2}tr[Var(Z_i\hat{\tau})] + \frac{1}{v_i}tr[Var(Z_{(i)}\hat{\tau})]$,

(ii) $tr[Var(B_e^{(i)}\hat{\tau})] = tr[Var(\bar{Z}_{(i)}\hat{\tau})]$.

For the inference problem under effects coding, from Theorem 8.9(i) we find that for A-optimality of the ith attribute, there is a disproportionately higher weight (importance) attached for the estimation of the main effect components representing comparisons among the first $v_i - 1$ levels labeled $0, 1, \dots, v_i - 2$. To compare A-optimal designs under effects coding and under orthonormal coding, we consider the following example.

Example 8.10. *Let $k = 2, N = 6, v_1 = v_2 = 3$. Under the utility-neutral setup, for the estimation of main effects, we consider two designs d_1 and d_2 (Chai et al., 2017), where*

$$d_1 = \begin{matrix} (21,\ 02), & (20,\ 11), & (22,\ 10) \\ (11,\ 00), & (12,\ 01), & (01,\ 10) \end{matrix} \qquad \textit{and} \qquad d_2 = \begin{matrix} (00,\ 11), & (11,\ 22), & (22,\ 00) \\ (01,\ 10), & (12,\ 21), & (20,\ 02) \end{matrix}.$$

The design d_1 is A-optimal under effects coding (the non-singular full-rank inference problem $(B_EB_E^T)^{-1}B_E\tau$) whereas d_2 is A-optimal under orthonormal coding (the non-singular full-rank inference problem $B_O\tau$) .

A closer look at d_1 shows an unequal number of paired comparisons of attribute levels for each attribute. For example, for each of the two attributes in d_1, the unordered pair

(2,0) occurs once, (2,1) occurs twice and (1,0) occurs thrice. However d_2, which is A-optimal under orthonormal coding, gives equal importance to the pairwise comparisons between the three attribute levels, i.e., comparing levels labeled $\{0$ and $1\}$, $\{0$ and $2\}$, and $\{1$ and $2\}$ for each attribute.

Example 8.10 shows that for the non-singular full-rank inference problem $(B_E B_E^T)^{-1} B_E \tau$, A-optimal design is such that it gives more importance to comparsions of the levels (0,1) as against the other comparsions, whereas under the orthonormal coding with non-singular full-rank inference problem $B_O \tau$, equal importance is attached to all the three elementary comparisons (0,1), (0,2) and (1,2).

We now give a generalization of Theorem 8.8 and Theorem 8.9 for k attributes, proof (and a mathematical version of the same) is given in the Appendix D. In order to use effects coding for the purpose of identifying A-optimal designs, parity is achieved among the attributes with different number of levels by considering rows of M, of the inference problem $M\tau$, in its normal form.

Theorem 8.11. *In a choice experiment having k attributes with the ith attribute at v_i levels, under the A-criterion, in the main effects model,*

(i) all elementary comparisons among the levels of each attribute are given equal importance for the inference problem $B_O \tau$,

(ii) $\binom{v_i-1}{2}$ elementary comparisons among the levels $0, \ldots, v_i - 2$ of each attribute are given more importance for the normalized inference problem $(B_E B_E^T)^{-1} B_E \tau$,

(iii) $v_i - 1$ elementary comparisons of levels $0, \ldots, v_i - 2$ with level $v_i - 1$ of each attribute are given more importance for the normalized inference problem $B_E \tau$.

It seems reasonable that one would want to attach equal importance to all elementary comparisons of the attribute levels for finding a good design for estimating each of the main effects. Restricting to a non-singular full-rank inference problem, this suggests that for conducting search of good choice designs under the A-criterion, it is more appropriate to use orthonormal coding than the effects coding.

The inference problem $B_E \tau$, that came up in Theorem 8.9(ii) (and Theorem 8.11 (iii)), addresses situations where the primary interest lies in making test-control comparisons. In test versus control comparisons, some new levels (called test levels) of an attribute are compared with an existing control level.

Optimal designs under the single attribute setup have been amply provided by several researchers; we refer the readers to the review papers by Hedayat et al. (1988) and Majumdar (1996). Under the multi-attribute setup, designs for test-control experiments have been obtained by Gupta (1995) and Gupta (1998). In discrete choice experiments, when manufacturers/service providers or policymakers want to study the effect of few potential attribute-levels on the utility of a product/service, test-control discrete choice experiments would be more practical.

Our primary goal in a test-control discrete choice experiment is to determine which level among the test levels has a significantly more impact on the utility when compared against the control level. Thus, for the ith attribute, we make elementary comparisons between control level labeled $v_i - 1$ and each of the remaining $v_i - 1$ test levels (labeled $0, 1, \ldots, v_i - 2$) with as much precision as possible. There is a need for more work for finding A-efficient or A-optimal test-control choice designs.

As indicated in Großmann and Schwabe (2015), the inference problem $(B_E B_E^T)^{-1} B_E \tau$ has more appeal than the inference problem $B_O \tau$ since the inference problem under effects coding has a clearer interpretation. This is so because the $v_i - 1$ independent comparisons under effects coding are representing the difference between the true unknown latent utility value of a level l_i (where l_i is considered only for $v_i - 1$ levels, i.e., $0, 1, \ldots, v_i - 2$), of an attribute i, and the average (over all levels l_i, where $l_i = 0, 1, \ldots, v_i - 1$) of the true unknown latent utility values, for each $i = 1, \ldots, k$.

Moreover, as also indicated in Graßhoff et al. (2003) (pp. 379) and also in Graßhoff et al. (2004) (pp. 375), to consider A-optimality under effects coding, one needs to resort to the singular full-rank inference problem by additionally considering, for each attribute, the left out comparison of one of the level effects (corresponding to level $v_i - 1$) as deviates from the average of all level effects of attribute i. Accordingly, the revised inference problem would be singular with M of order $(\sum v_i) \times L$ rather than $(\sum (v_i - 1)) \times L$. The corresponding information matrix for such an inference problem $M\tau$ would be of order $(\sum v_i) \times (\sum v_i)$ having rank $\sum (v_i - 1)$. For A-optimality considerations, we now show the equivalence of such a singular full-rank inference problem to that of the non-singular full-rank orthonormal inference problem.

It is easy to see that the normalized form of $(B_E B_E^T)^{-1} B_E$ is $\Upsilon_E (B_E B_E^T)^{-1} B_E$ where $\Upsilon_E = diag(\sqrt{\frac{L}{v_1-1}} I_{v_1-1}, \ldots, \sqrt{\frac{L}{v_k-1}} I_{v_k-1})$. We define a matrix A of order $(\sum v_i) \times p_M$ such that $A = diag(B_e^{(1)T}, \ldots, B_e^{(k)T})$. Also, let $B_n = A\Upsilon_E (B_E B_E^T)^{-1} B_E$. The weighted sum of variances of the $\sum v_i$ normalized comparisons is given by

$$\sum_{i=1}^{k} \frac{v_i - 1}{v_i} tr[Var(B_n^{(i)} \hat{\tau})] = tr[\Gamma_n Var(B_n \hat{\tau})], \tag{8.14}$$

where $B_n^{(i)}$ is B_n corresponding to the ith attribute, and $\Gamma_n = diag(\Gamma_{n1}, \Gamma_{n2}, \ldots, \Gamma_{nk})$ with $\Gamma_{ni} = ((v_i - 1)/v_i) I_{v_i}$.

While obtaining $tr[\Gamma_n Var(B_n \hat{\tau})]$ in (8.14), for each attribute i, we account for the contribution of $Var(B_n^{(i)} \hat{\tau})$ through $\frac{v_i-1}{v_i} tr[Var(B_n^{(i)} \hat{\tau})]$. This ensures providing equal importance to each of the k sets of v_i comparisons of which only $v_i - 1$ comparisons are independent, $i = 1, 2, \ldots, k$. We now have the following result, proof of which is in the Appendix D.

Theorem 8.12. *Under the main effects model,*

$$tr[Var(B_O \hat{\tau})] = tr[\Gamma_n Var(A\Upsilon_E (B_E B_E^T)^{-1} B_E \hat{\tau})] = tr[\Gamma_n Var(B_n \hat{\tau})].$$

It follows from Theorem 8.12 that while considering a full set of $\sum v_i$ normalized contrasts under effects coding, one would get the same A-optimal designs as one would get under orthonormal coding.

8.6 Discussion

In the theory of discrete choice experiments, we show the equivalence of two seemingly different approaches (the SB and the HZ approaches) for deriving the information matrices under the MNL model. Under the utility-neutral effects coding setup, the similarity of the two approaches for a D-optimal paired ($m = 2$) choice design was first addressed by Großmann and Schwabe (2015). We have shown how for any general coded matrix B_H and under a general setup, the two seemingly different structures of the information matrices are related.

We have obtained a simple linear function of τ that is being inferred upon under different inference problems. This allows us to establish that the inference problem being addressed under effects coding is not $B_E\tau$, but $(B_EB_E^T)^{-1}B_E\tau$. This helps us in establishing that the information matrix under the SB approach for the inference problem $(B_EB_E^T)^{-1}B_E\tau$ is $B_E\Lambda B_E^T$, which is same as the information matrix for effects coding under the HZ approach.

Most design criteria are sensitive to the coding of the attributes. As discussed, an exception is the D-optimality criterion because the criterion is not affected by reparameterizations. That is why most optimality results for choice and paired comparison designs have been derived for the D-criterion. For the non-singular full-rank setup, Sun and Dean (2016) and Sun and Dean (2017) obtained A-optimal designs under orthonormal coding, while Chai et al. (2017) have obtained three-level A-optimal paired choice designs, both under effects coding and orthonormal coding. Since A-optimal designs usually differ depending on the inference problem being addressed and the corresponding codings of the attributes, it is pertinent to understand which coding is more appropriate for defining main effects. In this connection, under the non-singular full-rank inference problem, the followers of the HZ approach have been usually adopting the effects coding of the attributes. Although, the orthonormal coding may be technically convenient, the contrast represented by the matrix B_O usually have no natural interpretation for qualitative attributes (see, Großmann and Schwabe (2015)).

To obtain A-optimal designs under a non-singular full-rank main-effects problem, we show that $B_O\tau$ attaches equal importance to each of the $\binom{v_i}{2}$ elementary comparisons of the ith main effect, $i = 1, \ldots, k$. On the contrary, the non-singular full-rank inference problem $(B_EB_E^T)^{-1}B_E\tau$ attaches more importance to $\binom{v_i-1}{2}$ of the overall $\binom{v_i}{2}$ elementary comparisons, for the ith main effect, $i = 1, \ldots, k$. However, to consider A-optimality under effects coding, one needs to resort to the singular full-rank inference problem by additionally considering the left out comparison of one of the level effects for each attribute. We have shown that from the point of obtaining a good A-optimal designs for choice experiments, one can equivalently use either orthonormal coding for the non-singular full-rank inference problem or normalized effects coding for the singular full-rank inference problem.

In situations where the primary interest lies in making test-control comparisons, test-control discrete choice experiments are conducted. The inference problem then, would be to estimate $B_E\tau$. The issue of construction of A-efficient and A-optimal designs for estimating $B_E\tau$ will be discussed in a future work.

Chapter 9

$E(s^2)$- and $UE(s^2)$-Optimal Supersaturated Designs

This chapter is based on the following work:
Cheng et al. (2018): Cheng, Ching-Shui; Das, Ashish; Singh Rakhi; Tsai, Pi-Wen; $E(s^2)$- and UE(s^2)-Optimal Supersaturated Designs. *J. Statist. Plann. Inference 196* (2018), 105–114.

9.1 Introduction

In an n-run factorial experiment involving m two-level factors, for the general mean and all the main effects to be estimable, we must have $n \geqslant m + 1$. A design is called supersaturated if $n < m + 1$. Under the assumption of factor sparsity that only a small number of factors are active, a supersaturated design can provide considerable cost saving in factor screening.

Each two-level supersaturated design d can be represented by an $n \times m$ matrix X_d having entries 1s and -1s, with each column of X_d corresponding to one factor and each row representing a factor-level combination. Let $Z_d = [1\ X_d]$, where 1 is the $n \times 1$ column of 1s, be the model matrix of the main-effects model for d. Two columns u and v of Z_d such that $u = v$ or $u = -v$ are said to be *aliased*. We require that no two columns of Z_d are aliased.

A factor is said to be level-balanced if the corresponding column of X_d has the same numbers of 1s and -1s. This is possible only if n is even. For an odd n, a factor is said to be nearly level-balanced if in the corresponding column the numbers of times 1 and -1 appear differ by one. Without loss of generality, we require that 1 appears $(n-1)/2$ times and -1 appears $(n+1)/2$ times. A design is said to be level-balanced (respectively, nearly level-balanced) if all the factors are level-balanced (respectively, nearly level-balanced). Later in Section 9.3 we provide a motivation behind level-balanced and nearly level-balanced designs.

Ideally one would want the columns of Z_d to be mutually orthogonal, which clearly is not possible for supersaturated designs. Under level-balanced designs, columns of X_d are orthogonal to 1. In this case, a simple measure (Booth and Cox, 1962) of nonorthogonality among the columns of X_d is

$$E_d(s^2) = \frac{1}{\binom{m}{2}} \sum_{1 \leqslant i < j \leqslant m} (x_i^T x_j)^2, \tag{9.1}$$

where x_i is the ith column of X_d. The popular $E(s^2)$-criterion (Lin, 1993*b*) is to minimize (9.1) among the level-balanced designs. For the case when n is odd, Nguyen and Cheng (2008) suggested minimizing (9.1) among nearly level-balanced designs.

Prima facie it appears that there is no need to impose the restriction of level-balance or near-level-balance while identifying a good supersaturated design so long as it minimizes the overall nonorthogonality among the columns of Z_d. Marley and Woods (2010) extended the definition of $E_d(s^2)$ to include the inner products of 1 and the columns of X_d. Jones and Majumdar (2014) also introduced the criterion

$$UE_d(s^2) = \frac{1}{\binom{m+1}{2}} \left[\sum_{i=1}^{m} (1^T x_i)^2 + \sum_{1 \leqslant i < j \leqslant m} (x_i^T x_j)^2 \right]. \tag{9.2}$$

For given m and n, let $\mathcal{D}_U(m,n)$ be the class of all supersaturated designs without the restriction of level-balance or near-level-balance, and let $\mathcal{D}_R(m,n)$ be the subclass of level-balanced or nearly level-balanced supersaturated designs. A supersaturated design $d^* \in \mathcal{D}_U(m,n)$ is said to be $UE(s^2)$-optimal if $UE_{d^*}(s^2) \leqslant UE_d(s^2)$ for all $d \in \mathcal{D}_U(m,n)$. It is clear that an $E(s^2)$-optimal design is $UE(s^2)$-optimal over the subclass $\mathcal{D}_R(m,n)$, but may not be $UE(s^2)$-optimal over the entire class $\mathcal{D}_U(m,n)$.

Removing the constraint of level-balance or near-level-balance makes the construction of $UE(s^2)$-optimal designs very easy, and produces a smaller sum of squares of the entries of the information matrix $Z_d^T Z_d$, which is twice the sum of the two quantities inside the brackets in (9.2). While the former is an advantage because in general $E(s^2)$-optimal designs are difficult to construct, a consequence is that usually there are many $UE(s^2)$-optimal designs with diverse characteristics and performances. Additional criteria are needed to choose among $UE(s^2)$-optimal designs. Among possible secondary criteria, Jones and Majumdar (2014) mentioned the maximization of the number of level-balanced factors among the $UE(s^2)$-optimal designs. Regarding the reduction of the sum of squares of the entries of the information matrix, since $\sum_{i=1}^{m}(1^T x_i)^2$ is minimized by level-balanced or nearly level-balanced designs, the difference in the $E(s^2)$- and $UE(s^2)$-criteria, when both are considered as criteria over the entire $\mathcal{D}_U(m,n)$, is that an $UE(s^2)$-optimal design minimizes the sum of $\sum_{i=1}^{m}(1^T x_i)^2$ and $\sum_{1\leqslant i<j\leqslant m}(x_i^T x_j)^2$, while an $E(s^2)$-optimal design minimizes the former followed by the minimization of the latter. In other words, the $E(s^2)$-criterion places a heavier weight on $\sum_{i=1}^{m}(1^T x_i)^2$. We denote this quantity by SS in the rest of the chapter.

Since only a small number of factors is expected to be active, one way to evaluate the performance of a supersaturated design is to consider its average efficiency over lower dimensional projections. Such an approach based on average D-efficiencies was proposed in Wu (1993). A comparison in Section 9.2 of the $E(s^2)$- and $UE(s^2)$-optimal designs for $n = 12$ and $m = 14$ discussed in Example 2 of Jones and Majumdar (2014) shows that the $UE(s^2)$-optimal design has worse projection properties than the $E(s^2)$-optimal design. Also, this $UE(s^2)$-optimal design, with 11 level-balanced factors, has worse projection properties than some other $UE(s^2)$-optimal designs with fewer level-balanced factors. This indicates that maximizing the number of level-balanced factors is not an appropriate secondary criterion.

In Section 9.3 we show that the traditional $E(s^2)$-criterion is a good surrogate for maximizing the average D-efficiency over f-factor projections for small f (relative to m). A similar argument leads to the minimization of SS as a good secondary criterion for selecting $UE(s^2)$-optimal designs with good lower-dimensional projections. In Section 9.4 we present simple and systematic methods of constructing designs that minimize SS among the $UE(s^2)$-optimal designs constructed by using the method of Jones and Majumdar (2014). Our examples show that, for lower dimensional projections, the best $UE(s^2)$-optimal designs have nearly as good average D-efficiencies as the $E(s^2)$-optimal designs. We also provide $UE(s^2)$-optimal design construction for the case $n = m \equiv 1$ (mod

4), which is missing in Jones and Majumdar (2014). In Section 9.5, we derive conditions under which $E(s^2)$-optimal designs are also $UE(s^2)$-optimal and identify several families of designs that are optimal under both criteria.

9.2 An example

The basis of using a supersaturated design is the inherent assumption that there are very few active factors. The identification of the active factors is usually based on a forward selection method of model building involving projections onto various subsets of factors. Thus it is desirable to use a supersaturated design with good projection properties in the sense that on average the model parameters can be efficiently estimated during the model building process.

Suppose among the m factors, only those in a set F of f factors are active, $f \leqslant n-1$. Let X_{d^F} be the design matrix of d^F consisting of the f corresponding columns of X_d. Here d^F is the projected design of d onto the factors in F. Consider the model matrix $Z_{d^F} = [1\ X_{d^F}]$ and let M_{d^F} be the information matrix of d^F. Then

$$M_{d^F} = Z_{d^F}^T Z_{d^F} = \begin{bmatrix} n & 1^T X_{d^F} \\ X_{d^F}^T 1 & X_{d^F}^T X_{d^F} \end{bmatrix}.$$

Suppose we measure the performance of the projected design by the D-criterion $D(X_{d^F}) = \{(\det[M_{d^F}])^{1/(f+1)}\}/n$. Then the overall D-efficiency of f-factor projections of a supersaturated design d can be measured by the average D-efficiency

$$D_f(d) = \frac{1}{\binom{m}{f}} \sum_{F:|F|=f} D(X_{d^F}),$$

where the sum is over all the subsets F consisting of f factors, and the objective is to maximize $D_f(d)$, $f = 1, 2, \ldots, (n-1)$. Another quantity of interest is the number of non-estimable f-factor projective main-effects models, denoted by NE_f. Note that D_f and NE_f are related to, respectively, information capacity and estimation capacity proposed by Sun (1993).

In Example 9.1 , we compare the performances of several $E(s^2)$- and $UE(s^2)$-optimal designs with $m = 14$ and $n = 12$ with respective to D_f and NE_f for all $f \leqslant 11$. For each design, we also report the values of the following three characteristics which, as demonstrated later, are useful for helping identify $UE(s^2)$-optimal designs with good projection properties:

$LB =$ the number of level-balanced (nearly level-balanced) factors for n even (odd);
$OF =$ the number of orthogonal (nearly orthogonal) pairs of factors among the $\binom{m}{2}$ pairs for n even (odd);
$Q = LB + OF$.
Here two factors are said to be orthogonal (nearly orthogonal) if the inner product of their corresponding columns in the model matrix is zero (± 1).

Example 9.1. For $m = 14$ and $n = 12$, a $UE(s^2)$-optimal design, say d_6, and an $E(s^2)$-optimal design, say d_1, are displayed in Table 9.2 and Table 9.3 respectively, of Jones and Majumdar (2014). Here d_6 has 11 level-balanced factors. We construct four additional $UE(s^2)$-optimal designs d_2, d_3, d_4, and d_5 with 6, 12, 6, and 5 level-balanced factors, respectively. All the six designs can be found in the Appendix E. Table 9.1 shows that $d_i > d_6$ for all $i = 1, \ldots, 5$, where $d_i > d_j$ (d_i dominates d_j) means that d_i is at least as good as d_j under both D_f and NE_f for every $f = 1, \ldots, n-1$, and d_i is better than d_j in some cases. In particular, the $E(s^2)$-optimal design d_1 dominates the $UE(s^2)$-optimal design d_6. We also note that $d_2 > d_j$, $j = 3, 4, 5, 6$. This indicates that, at least in terms of projection properties, maximizing the number of level-balanced factors is not an appropriate secondary criterion. Design d_2 is the best among the five $UE(s^2)$-optimal designs in the table. In fact, based on an exhaustive search, it is the best design among $UE(s^2)$-optimal designs constructed using the method of Jones and Majumdar (2014) with respect to D_f and NE_f. Between d_2 and d_1, d_1 is better than d_2 under D_f for $f \leqslant 5$ (but d_2 is at least 99% efficient as d_1 in all these cases, where the efficiency is measured by the ratio of D_f values), the two designs are tied under NE_f for $f \leqslant 6$, and d_2 is better than d_1 for $6 \leqslant f \leqslant 11$. Thus d_2 can be recommended. However, it takes additional work to identify it among many $UE(s^2)$-optimal designs.

9.3 Projection justification

It is well-known that a good surrogate for maximizing $D(X_{d^F})$ is the (M.S)-criterion proposed by Eccleston and Hedayat (1974): maximizing $\mathrm{tr}[M_{d^F}]$ and minimizing $\mathrm{tr}[M_{d^F}]^2$ among those that maximize $\mathrm{tr}[M_{d^F}]$. This goes back to the result of Kiefer (1958, 1975*b*) that a design is optimal with respect to many criteria if it maximizes the trace of the

Table 9.1: Comparison of six designs with $m = 14$ and $n = 12$

	d_1	d_2	d_3	d_4	d_5	d_6
SS	0	32	32	44	48	48
LB	14	6	12	6	5	11
OF	67	36	72	36	37	73
Q	81	42	84	42	42	84
$(D_1,\ NE_1)$	(1, 0)	(0.9920, 0)	(0.9918, 0)	(0.9889, 0)	(0.9879, 0)	(0.9877, 0)
$(D_2,\ NE_2)$	(0.9898, 0)	(0.9809, 0)	(0.9805, 0)	(0.9765, 0)	(0.9751, 0)	(0.9744, 0)
$(D_3,\ NE_3)$	(0.9759, 0)	(0.9679, 0)	(0.9671, 0)	(0.9626, 0)	(0.9609, 0)	(0.9579, 1)
$(D_4,\ NE_4)$	(0.9590, 0)	(0.9530, 0)	(0.9498, 3)	(0.9470, 0)	(0.9449, 0)	(0.9341, 13)
$(D_5,\ NE_5)$	(0.9391, 0)	(0.9360, 0)	(0.9235, 30)	(0.9291, 0)	(0.9268, 0)	(0.8972, 75)
$(D_6,\ NE_6)$	(0.9157, 0)	(0.9162, 0)	(0.8808, 135)	(0.9084, 0)	(0.9058, 0)	(0.8400, 255)
$(D_7,\ NE_7)$	(0.8869, 3)	(0.8928, 0)	(0.8116, 360)	(0.8832, 3)	(0.8801, 4)	(0.7540, 568)
$(D_8,\ NE_8)$	(0.8476, 25)	(0.8630, 6)	(0.7048, 627)	(0.8484, 24)	(0.8439, 30)	(0.6323, 867)
$(D_9,\ NE_9)$	(0.7831, 82)	(0.8163, 36)	(0.5522, 738)	(0.7876, 84)	(0.7799, 97)	(0.4731, 918)
(D_{10}, NE_{10})	(0.6614, 143)	(0.7189, 93)	(0.3565, 585)	(0.6566, 165)	(0.6455, 177)	(0.2879, 665)
(D_{11}, NE_{11})	(0.4060, 161)	(0.4790, 132)	(0.1477, 300)	(0.3803, 180)	(0.3710, 184)	(0.1108, 316)

information matrix and all the eigenvalues of the information matrix are equal. Since $\mathrm{tr}[M_{d^F}] = n(f+1)$ is a constant, a good surrogate for maximizing $D_f(d)$ is to minimize

$$\frac{1}{\binom{m}{f}} \sum_{F:|F|=f} \mathrm{tr}[M_{d^F}]^2 = \frac{1}{\binom{m}{f}} \sum_{F:|F|=f} \mathrm{tr}[Z_{d^F}^T Z_{d^F}]^2. \tag{9.3}$$

Using the fact that the first- and second-order inclusion probabilities under simple random sampling of size f without replacement from a population of size m are, respectively, f/m and $f(f-1)/[m(m-1)]$, it is easy to see that

$$\frac{1}{\binom{m}{f}} \sum_{F:|F|=f} \mathrm{tr}[Z_{d^F}^T Z_{d^F}]^2 = \text{constant} + \frac{2f}{m}[1^T X_d X_d^T 1] + \frac{f(f-1)}{m(m-1)}\mathrm{tr}[(X_d^T X_d)^2]. \tag{9.4}$$

Due to factor sparsity, f/m is small; thus $f(f-1)/[m(m-1)]$ is much smaller than f/m. It follows that a good surrogate for minimizing (9.4) is the two-step procedure of first minimizing $1^T X_d X_d^T 1 (= SS)$, and then minimizing $\mathrm{tr}[(X_d^T X_d)^2]$. The first step is achieved by level-balanced or nearly level-balanced designs. The second step is equivalent to minimizing $E_d(s^2)$. This justifies the traditional $E(s^2)$-criterion and shows why it is important to restrict to level-balanced or nearly level-balanced designs for achieving good lower-dimensional projection properties. Earlier, Lin (1993*a*) and Tsai and Gilmour (2016) studied two-level saturated main effects screening designs under factor sparsity. They found that those with good lower-dimensional projections are to be found among level-balanced or nearly level-balanced ones.

For each design $d \in D_U(m,n)$, let $UE_f^d(s^2)$ be the average of the $UE_{d^F}(s^2)$ values over all F involving f factors; that is,

$$UE_f^d(s^2) = \frac{1}{\binom{m}{f}} \sum_{F:|F|=f} UE_{d^F}(s^2).$$

Since $Z_{d^F}^T Z_{d^F}$ has constant diagonals n, minimizing $UE_f^d(s^2)$ is equivalent to minimizing (9.3), and thus is also a good surrogate for maximizing $D_f(d)$.

The following can easily be established:

$$UE_f^d(s^2) = \frac{m+1}{(f+1)(m-1)} \left\{ (f-1)\, UE_d(s^2) + \frac{2(m-f)}{m(m+1)} 1^T X_d X_d^T 1 \right\}. \qquad (9.5)$$

A possible secondary criterion for discriminating $UE(s^2)$-optimal designs is to minimize $UE_f^d(s^2)$ among the $UE(s^2)$-optimal designs. By (9.5), this is to minimize $1^T X_d X_d^T 1$ among the $UE(s^2)$-optimal designs. The resulting designs do not depend on f.

The traditional $E(s^2)$-criterion is equivalent to the two-step procedure of minimizing SS and then minimizing $UE_d(s^2)$ among those that minimize SS. Thus using the minimization of SS as a secondary criterion for $UE(s^2)$ amounts to reversing the two steps of the traditional $E(s^2)$-criterion as formulated above: when the minimization of SS is not done before minimizing $UE_d(s^2)$, it should be done afterwards. Let $\mathcal{D}_H(m,n)$ be the class of all $UE(s^2)$-optimal designs constructed from a Hadamard matrix using the method of Jones and Majumdar (2014). A design that minimizes SS over $\mathcal{D}_H(m,n)$ is said to be superior $UE(s^2)$-optimal.

Usually superior $UE(s^2)$-optimal designs are not unique. In fact, the $E(s^2)$-criterion also suffers from this problem. The issue of choosing better $E(s^2)$-optimal designs arises naturally, but it has not received much attention in the literature, perhaps because $E(s^2)$-optimal designs are difficult to construct. The easy construction of superior $UE(s^2)$-optimal designs makes it possible to further choose better designs from them.

An interesting observation in Table 9.1 is that for the two designs with $SS = 32$ (and also those with $SS = 48$), the one with more level-balanced factors has worse projection properties. We offer the following explanation. It is desirable to make the columns of the model matrix as nearly orthogonal as possible, that is, to make the absolute values (or, equivalently, the squares) of the off-diagonal entries of the information matrix as *uniformly* small as possible. For this purpose, we do not want to have too many pairs of orthogonal z_i and z_j once $UE_d(s^2)$ has been minimized, where z_i and z_j are columns of

Z_d. After the minimization of $UE_d(s^2)$, the sum of squares of all pairwise inner products $z_i^T z_j, i \neq j$, is fixed. If there are too many zeros among these inner products, then the overall distribution of the squared pairwise inner products is likely to be more dispersed. Thus, subject to the minimization of $UE_d(s^2)$, it is better not to have too many orthogonal pairs (z_i, z_j), $i \neq j$. Since the number of such pairs is equal to $Q = LB + OF$, it is preferable to have a small Q. In particular, we do not want to have too many level-balanced factors among the designs with the same values of $UE_d(s^2)$ and SS.

The discussion above can be linked to the concept of majorization (Marshall and Olkin, 1979). Given two vectors $x = (x_1, \ldots, x_k)^T$ and $y = (y_1, \ldots, y_k)^T$, we say that x is majorized by y if

$$\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} y_i \tag{9.6}$$

and

$$\sum_{i=1}^{t} x_{[i]} \geqslant \sum_{i=1}^{t} y_{[i]} \text{ for all } 1 \leqslant t \leqslant k-1, \tag{9.7}$$

where $x_{[1]} \leqslant \cdots \leqslant x_{[k]}$ and $y_{[1]} \leqslant \cdots \leqslant y_{[k]}$ are ordered values of $x_1, \ldots, x_k$ and $y_1, \ldots, y_k$, respectively. Under (9.6) and (9.7), the entries in x can be regarded as less dispersed than those in y. For example, the vector with $x_1 = \cdots = x_k$ is majorized by all the y's satisfying (9.6). Majorization is a strong property that requires all the $k-1$ inequalities in (9.7) to hold. Suppose all the components of x and y are nonnegative, (9.6) is satisfied, x has Q_x zero components, and y has Q_y zero components. If $Q_x < Q_y$, then (9.7) is satisfied for all $1 \leqslant t \leqslant Q_y$. In this case (9.7) partially holds for a subset of the inequalities.

Table 9.1 exhibits the patterns that (i) designs with smaller SS and Q tend to have better projection properties and (ii) those with smaller SS but larger Q may have better lower-dimensional and worse higher-dimensional projection properties. For example, d_2 has smaller SS and Q than the other $UE(s^2)$-optimal designs in the table; at the same time it dominates all these designs. At the other extreme, d_6 has larger SS and Q than the other $UE(s^2)$-optimal designs in the table, and it is dominated by all of them. Design d_3 has the same value of SS as d_2, but has a larger Q; we see that it is dominated by d_2. Design d_4 has the same value of Q as d_5, but has a smaller SS; we also have that d_4 dominates d_5. Even though d_3 has a smaller SS than d_4 and d_5, it has a much larger Q. We see that d_3 has better lower-dimensional, but worse higher-dimensional projection properties than these two other designs. A similar observation applies to the comparison of d_1 and d_2.

Example 9.2. Consider $m = 16$ and $n = 10$. Seven designs are listed in Table 9.2. Design d_1 is $E(s^2)$-optimal, while d_2, d_3, d_4, d_5, d_6, and d_7 are $UE(s^2)$-optimal. These designs can be found in the Appendix E. The six $UE(s^2)$-optimal designs in the table have the same values of SS, and are labeled according to increasing values of Q. We have $d_1 > d_2 > d_3 > d_4 > d_6 > d_5 > d_7$. Note that d_5 has a very slightly smaller SS than d_6 (53 vs. 54), but is very slightly dominated by the latter. The $E(s^2)$-optimal design d_1 dominates all the $UE(s^2)$-optimal designs in the table, but a carefully chosen $UE(s^2)$-optimal design such as d_2 is at least 96% efficient as d_1 under the D_f-criterion for $f \leqslant 7$.

Table 9.2: Comparison of seven designs with $m = 16$ and $n = 10$

	d_1	d_2	d_3	d_4	d_5	d_6	d_7
SS	0	60	60	60	60	60	60
LB	16	1	1	7	7	7	9
OF	0	5	8	43	46	47	58
Q	16	6	9	50	53	54	67
(D_1, NE_1)	(1, 0)	(0.9811, 0)	(0.9811, 0)	(0.9807, 0)	(0.9807, 0)	(0.9807, 0)	(0.9799, 0)
(D_2, NE_2)	(0.9792, 0)	(0.9562, 0)	(0.9556, 0)	(0.9552, 0)	(0.9546, 0)	(0.9550, 0)	(0.9521, 0)
(D_3, NE_3)	(0.9504, 0)	(0.9276, 0)	(0.9262, 0)	(0.9247, 1)	(0.9232, 1)	(0.9243, 1)	(0.9165, 3)
(D_4, NE_4)	(0.9157, 0)	(0.8951, 0)	(0.8927, 0)	(0.8866, 16)	(0.8841, 16)	(0.8861, 16)	(0.8663, 48)
(D_5, NE_5)	(0.8748, 0)	(0.8568, 6)	(0.8529, 10)	(0.8367, 114)	(0.8326, 118)	(0.8357, 116)	(0.7914, 336)
(D_6, NE_6)	(0.8261, 4)	(0.8075, 76)	(0.8000, 126)	(0.7693, 486)	(0.7614, 534)	(0.7664, 513)	(0.6817, 1376)
(D_7, NE_7)	(0.7652, 46)	(0.7354, 451)	(0.7181, 725)	(0.6762, 1427)	(0.6597, 1661)	(0.6679, 1568)	(0.5326, 3680)
(D_8, NE_8)	(0.6791, 311)	(0.6141, 1684)	(0.5759, 2526)	(0.5396, 3205)	(0.5111, 3767)	(0.5227, 3562)	(0.3522, 6732)
(D_9, NE_9)	(0.5149, 1679)	(0.3908, 4410)	(0.3316, 5710)	(0.3222, 5835)	(0.2923, 6421)	(0.3009, 6264)	(0.1649, 8648)

The empirical studies we have carried out indicate that a design is likely to dominate those with larger SS and larger (or similar) Q, two designs with about the same SS and Q are expected to perform similarly, and one with smaller SS but a much larger Q may have better lower-dimensional projections and worse higher-dimensional projections. There is no guarantee that a simple surrogate criterion such as what we propose here will produce the best design, but minimizing SS followed by minimizing Q is an effective way of getting $UE(s^2)$-optimal designs with good projection properties. We are interested in using a computationally cheap criterion to identify good designs (such as d_2 in both Example 9.1 and Example 9.2), rather than to rank the $UE(s^2)$-optimal designs.

In certain cases $E(s^2)$-optimal designs are easy to construct; then the strategy of minimizing Q can be used to find better designs. For example, Lin (1993*b*) proposed a simple method of using a Hadamard matrix of order $4t$ to construct a supersaturated design with $m = 4t - 2$ and $n = 2t$. Nguyen (1996) showed that such *half-Hadamard matrices* achieve the lower bound on $E(s^2)$ derived by Nguyen (1996) and Tang and Wu (1997), and hence are $E(s^2)$-optimal. They are also $UE(s^2)$-optimal (see Theorem 9.6 in Section 9.5). Furthermore, with $SS = 0$, they minimize SS among the $UE(s^2)$-optimal designs. Projection properties of several half-Hadamard matrices and $UE(s^2)$-optimal designs for the case $m = 22$ and $n = 12$ are summarized in Table 9.3. All the designs in this table are $UE(s^2)$-optimal and can be found in the Appendix E. Designs d_1 and d_2 are half-Hadamard matrices constructed by using Hadamard matrices of order 24 available in `http://neilsloane.com/hadamard/`. Designs d_3, d_4, and d_5 are obtained by deleting rows and columns of a Hadamard matrix without the restriction of level-balance. We can see that for the two $E(s^2)$-optimal designs, d_1 has smaller Q and dominates d_2. Design d_3 dominates d_4 and d_5. The better $E(s^2)$-optimal design d_1 is better than d_3 with respect to D_f for all $f \leq 10$, but has more nonestimable models when $f \geqslant 7$. Overall, d_3 performs quite well.

Table 9.3: Comparison of five designs with $m = 22$ and $n = 12$

	d_1	d_2	d_3	d_4	d_5
SS	0	0	44	80	140
LB	22	22	11	8	8
OF	132	138	66	67	78
Q	154	160	77	75	86
$(D_1,\ NE_1)$	(1, 0)	(1, 0)	(0.9930, 0)	(0.9872, 0)	(0.9774, 0)
$(D_2,\ NE_2)$	(0.9835, 0)	(0.9833, 0)	(0.9741, 0)	(0.9663, 0)	(0.9529, 0)
$(D_3,\ NE_3)$	(0.9611, 0)	(0.9606, 0)	(0.9505, 0)	(0.9416, 0)	(0.9263, 0)
$(D_4,\ NE_4)$	(0.9347, 0)	(0.9338, 0)	(0.9234, 0)	(0.9138, 0)	(0.8972, 0)
$(D_5,\ NE_5)$	(0.9044, 0)	(0.9032, 0)	(0.8927, 0)	(0.8827, 0)	(0.8651, 0)
$(D_6,\ NE_6)$	(0.8699, 0)	(0.8684, 0)	(0.8580, 0)	(0.8477, 0)	(0.8294, 0)
$(D_7,\ NE_7)$	(0.8302, 10)	(0.8284, 12)	(0.8183, 0)	(0.8077, 0)	(0.7890, 0))
$(D_8,\ NE_8)$	(0.7836, 159)	(0.7816, 212)	(0.7719, 0)	(0.7613, 8)	(0.7424, 0)
$(D_9,\ NE_9)$	(0.7264, 1445)	(0.7238, 1884)	(0.7159, 0)	(0.7053, 151)	(0.6867, 0)
$(D_{10},\ NE_{10})$	(0.6486, 10403)	(0.6448, 12478)	(0.6438, 88)	(0.6328, 1669)	(0.6157, 216)
$(D_{11},\ NE_{11})$	(0.5089, 72727)	(0.5013, 81856)	(0.5337, 5401)	(0.5171, 22181)	(0.5078, 8463)

Remark 9.3. Jones and Majumdar (2014) argued that although insisting on level-balance achieves the highest efficiency for the intercept, "estimation of the intercept... is not a goal for most experiments", and "it comes at the expense of precision of main effect estimation." Suppose the intercept is treated as a nuisance parameter; then for any subset F of $\{1, \ldots, m\}$, the information matrix for the f main effects is $X_{dF}^T(I_n - \frac{1}{n}11^T)X_{dF}$, where I_a denotes the identity matrix of order a. Similar to earlier discussions in this section, a good surrogate for maximizing the average D-efficiency of the main-effect estimates over all f-factor projections is to maximize $\frac{1}{\binom{m}{f}} \sum_{F:|F|=f} \text{tr}[X_{dF}^T(I_n - \frac{1}{n}11^T)X_{dF}]$, and subject to that, minimize $\frac{1}{\binom{m}{f}} \sum_{F:|F|=f} \text{tr}[X_{dF}^T(I_n - \frac{1}{n}11^T)X_{dF}]^2$. A similar derivation as before shows that this is equivalent to minimizing SS, and subject to that, minimizing $\text{tr}[(X_d^T X_d)^2]$, i.e. the traditional $E(s^2)$-optimality. Therefore the $E(s^2)$-optimality is also a good surrogate for maximizing the average D-efficiency even when the mean is treated as a nuisance parameter and only the main effects are of interest. In this case, for the average D-efficiencies of the designs in Table 9.1 and Table 9.2, the pattern remains the same.

9.4 Construction of superior $UE(s^2)$-optimal designs

There is no simple general construction of $E(s^2)$-optimal designs and, except for some limited values of m and n, they are not readily available. This section is devoted to the construction of $UE(s^2)$-optimal designs that minimize SS among the $UE(s^2)$-optimal designs in $\mathcal{D}_H(m, n)$. There is no guarantee that such superior $UE(s^2)$-optimal designs minimize SS among all the $UE(s^2)$-optimal designs since the method given by Jones and Majumdar (2014) does not produce all the $UE(s^2)$-optimal designs. However, our construction method is simple, systematic, and can be applied to all cases. We need the following result:

Lemma 9.4. *A design minimizes SS among the $UE(s^2)$-optimal designs if and only if it maximizes $tr[(X_d X_d^T)^2]$.*

Proof. We have

$$\text{tr}[(Z_d^T Z_d)^2] = n^2 + 2 \cdot 1^T X_d X_d^T 1 + \text{tr}[(X_d^T X_d)^2]. \tag{9.8}$$

Since all the $UE(s^2)$-optimal designs have the same value of $\text{tr}[(Z_d^T Z_d)^2]$, by (9.8), $1^T X_d X_d^T 1$ is minimized if and only if $\text{tr}[(X_d^T X_d)^2] = \text{tr}[(X_d X_d^T)^2]$ is maximized. $\square$

In Jones and Majumdar (2014), the construction of $UE(s^2)$-optimal designs considers four cases depending on whether m is of the form $4t-1$, $4t-2$, $4t$, or $4t+1$, where t is a positive integer.

(a) Construction for $m = 4t-1$.

Let H be a $4t \times 4t$ (normalized) Hadamard matrix with all the entries in the first row and first column equal to 1. To construct a $UE(s^2)$-optimal design, we delete the first column and $4t-n$ rows of H. Then the resulting matrix O is $UE(s^2)$-optimal with $m = 4t-1$, provided that $Y = [1\ O]$ has no aliased columns. Note that Y is an $n \times 4t$ matrix in which any two rows are orthogonal.

We first address the existence of a Y with no aliased columns, an issue not considered in Jones and Majumdar (2014). Without loss of generality, let $H^T = \left[Y^T\ G^T\right]$. Also, let $Y^TY = [u_{ij}]$ and $G^TG = [w_{ij}]$. Then since $H^TH = (4t)I_{4t}$, for all i, j,

$$|u_{ij}| = |w_{ij}|, \quad |u_{ij}| \leqslant n, \text{ and } \quad |w_{ij}| \leqslant 4t-n. \tag{9.9}$$

If Y has two aliased columns, say the i_0th and j_0th columns, then $|u_{i_0j_0}| = n$. Then it follows from (9.9) that $n \leqslant 4t-n$. This shows that as long as $n > 2t$, any matrix Y obtained by deleting $4t-n$ rows from H does not contain aliased columns.

On the other hand, if n is too small, then the aliasing of some columns in Y cannot be avoided. This can be seen as follows. There are at most 2^{n-1} columns of 1s and -1s of size n that are not mutually aliased. Thus in order for Y to have no aliased columns, we must have $2^{n-1} \geqslant 4t$, or $n \geqslant \lceil \log_2 t \rceil + 3$, where $\lceil z \rceil$ is the smallest integer greater than or equal to z.

When $4t$ is a power of 2, say $4t = 2^w$, let H be a Hadamard matrix of order $4t$ that is the Kronecker product of w normalized Hadamard matrices of order 2. In this case, $\lceil \log_2 t \rceil + 3 = w+1$. For any $n \geqslant w+1$, let Y be the $n \times 4t$ submatrix of H consisting of the $w+1$ rows $\otimes_{i=1}^{w}(1,1)$, $(1,-1) \otimes_{i=2}^{w} (1,1)$, $\otimes_{i=1}^{j}(1,1) \otimes (1,-1) \otimes_{i=j+2}^{w} (1,1)$, $j = 1, ..., w-2$, $\otimes_{i=1}^{w-1}(1,1) \otimes (1,-1)$, and any $n-w-1$ additional rows of H if $n > w+1$. Then it can be seen that Y has no aliased columns.

When $4t$ is not a power of 2, for Hadamard matrices of order $4t$, $2 < t < 25$ and $n \geq \lceil \log_2 t \rceil + 4$, we have enumerated and found that there exists at least one set of $4t-n$ rows that can be deleted to get a Y without aliased columns.

Theorem 9.5. *If m is of the form $4t-1$, then all the UE(s^2)-optimal designs have the same value of SS.*

Proof. When m is of the form $4t-1$, each $UE(s^2)$-optimal X_d is such that any two rows of the $n \times (4t)$ matrix $Z_d = [1\ X_d]$ are orthogonal. Then $Z_d Z_d^T = (4t)I_n$ and hence $\mathrm{tr}[(Z_d^T Z_d)^2] = \mathrm{tr}[(Z_d Z_d^T)^2] = (4t)^2 n$. On the other hand, $X_d X_d^T$ is a matrix with all the diagonal entries equal to $4t-1$ and all the off-diagonal entries equal to -1. This implies that $\mathrm{tr}[(X_d^T X_d)^2] = \mathrm{tr}[(X_d X_d^T)^2]$ is a constant. It follows from (9.8) that $1^T X_d X_d^T 1$ is also a constant. □

Note that Theorem 9.5 is not restricted to designs constructed by the method of Jones and Majumdar (2014).

(b) Construction for $m = 4t-2$.

For $m = 4t-2$, according to the construction of Jones and Majumdar (2014), a $UE(s^2)$-optimal X_d can be obtained by deleting 1 and another column $\alpha = (\alpha_1, \ldots, \alpha_n)^T$ from Y, where Y is as described above. To construct a superior $UE(s^2)$-optimal design, we have to ensure that $1^T \alpha \alpha^T 1$ is maximized. This is achieved if α is a least level-balanced column; that is, one such that the difference of the numbers of times 1 and -1 appear is the largest.

(c) Construction for $m = 4t$.

For $m = 4t$, we need to add a column α to O such that $1^T \alpha \alpha^T 1$ is minimized; that is, add a level-balanced or nearly level-balanced column that is not aliased with any column of O.

(d) Construction for $m = 4t+1, n < m$.

For $m = 4t+1$, the construction of Jones and Majumdar (2014) adds to O two columns α and β such that the resulting design has no aliased columns. However, since $n \leqslant 4t$, we have $n < m$.

Suppose in the $n \times 2$ matrix $[\alpha\ \beta]$, $(1,1)$, $(-1,-1)$, $(1,-1)$, and $(-1,1)$ appear a, b, c, and d times as row vectors, respectively. Then in order for X_d to be $UE(s^2)$-optimal, $a+b$ and $c+d$ differ by at most one; see Jones and Majumdar (2014). Since Y has orthogonal rows and X_d is obtained by deleting 1 and adding α and β to Y, by suitably rearranging the rows of X_d, we have

$$X_d X_d^T = \begin{bmatrix} (m-1)I_a + J_{a,a} & -3J_{a,b} & -J_{a,c} & -J_{a,d} \\ -3J_{b,a} & (m-1)I_b + J_{b,b} & -J_{b,c} & -J_{b,d} \\ -J_{c,a} & -J_{c,b} & (m-1)I_c + J_{c,c} & -3J_{c,d} \\ -J_{d,a} & -J_{d,b} & -3J_{d.c} & (m-1)I_d + J_{d,d} \end{bmatrix}, \quad (9.10)$$

where $J_{a,b}$ is the $a \times b$ matrix of 1s. By Lemma 9.4, we need to maximize $\mathrm{tr}[(X_d X_d^T)^2]$, the sum of squares of all the entries of the matrix in (9.10). By direct computation, this is achieved if $ab + cd$ is maximized. It follows that for given $a+b$ and $c+d$, a and b should differ by at most 1, and c and d also should differ by at most 1. Combining this with that $a+b$ and $c+d$ differ by at most 1, we conclude that

(i) for $n = 4s$, $a = b = c = d = s$;

(ii) for $n = 4s+1$, $\{a,b\} = \{s, s+1\}, c = d = s$ or $a = b = s, \{c,d\} = \{s, s+1\}$;

(iii) for $n = 4s+2$, $\{a,b\} = \{c,d\} = \{s, s+1\}$;

(iv) for $n = 4s+3$, $\{a,b\} = \{s, s+1\}, c = d = s+1$ or $a = b = s+1, \{c,d\} = \{s, s+1\}$.

(e) Construction for $m = 4t+1, n = m$.

For the case $m = 4t+1$, as noted earlier, the method of Jones and Majumdar (2014) cannot be applied to the construction of $UE(s^2)$-optimal designs with $n = m$. In this case, instead of adding two columns to O, one can start with a larger Hadamard matrix H' of order $4t+4$ and delete from H' three rows, a column of 1s, and two other columns in which the number of rows that are $(1,1)$ or $(-1,-1)$ and the number of rows that are $(1,-1)$ or $(-1,1)$ differ by at most 1. Two such columns exist because there must be two columns such that the corresponding 1×2 rows in the first two deleted rows are $(1,1)$ and $(-1,1)$. Then no matter what entries in the third row are, the required condition holds. In order to maximize $\mathrm{tr}(X_d X_d^2)$ among the $UE(s^2)$-optimal designs so constructed, we need to maximize the sum of squares of the inner products of the 3×1 vector of 1s and the two 3×1 columns corresponding to the two deleted columns and three deleted rows. Theoretically, such a maximum is achieved when the two columns are $(1, 1, 1)^T$ or $(-1, -1, -1)^T$. This is not possible since otherwise the design would not be $UE(s^2)$-optimal. The next best is to delete two columns with the corresponding column sums in the deleted rows being ± 3 and ± 1. The existence of such columns can be seen as follows. The last $4t + 3$ rows of H form an orthogonal array of size $4t + 4$ and strength two. It follows from Theorem 2.1 of Cheng (1995) that, for any three of the $4t + 3$ rows, either each of the eight possible 3×1 columns of 1s and -1s appears at least once or one of the three rows is aliased with the component-wise product of the other two rows. Using this, one can easily establish the existence of the required columns.

9.5 Cases where traditional $E(s^2)$-optimal designs are also $UE(s^2)$-optimal

Several families of supersaturated designs are known to be $E(s^2)$-optimal:

(a) As mentioned in the paragraph before Remark 9.3, half Hadamad matrices are $E(s^2)$-optimal. Such designs have $n = 2t, m = 4t - 2$.

(b) Cheng (1997) showed that a design obtained by deleting an arbitrary factor from any design in (a) is $E(s^2)$-optimal. Such designs have $n = 2t, m = 4t - 3$.

(c) Extending the result in (a) to the case of odd n, Nguyen and Cheng (2008) constructed $E(s^2)$-optimal designs for $n = 2t - 1$ and $m = 4t - 2$.

(d) A design obtained by deleting an arbitrary factor from any design in (c) is $E(s^2)$-optimal. Such designs have $n = 2t - 1$ and $m = 4t - 3$.

(e) Let H^* be obtained by deleting a column of 1s from a $4t \times 4t$ Hadamard matrix, and let α be a level-balanced column that is not aliased with any column of H^*. Let $X_{d^*} = [\alpha \ H^*]$. Then d^* is $E(s^2)$-optimal. Such designs have $n = m = 4t$.

(f) Let H^* be as in part (e) and let α and β be $4t \times 1$ columns of 1s and -1s such that neither is aliased with any column of H^* and each of $(1, 1), (1, -1), (-1, 1)$, and

$(-1,-1)$ appears t times as a row of $[\alpha\ \beta]$. Let $X_{d^*} = [\alpha\ \beta\ H^*]$. Then d^* is $E(s^2)$-optimal. Such designs have $n = 4t, m = 4t + 1$.

(g) Let H^* be obtained by deleting a column of 1s and a row of 1s from a $4t \times 4t$ Hadamard matrix, and $X_{d^*} = H^*$. Then d^* is $E(s^2)$-optimal. Such designs have $n = m = 4t - 1$.

(h) Let H^* be as in (g). Then the design obtained by adding to H^* a nearly level-balanced column that is not aliased with any column of H^* is $E(s^2)$-optimal. Such designs have $n = 4t - 1, m = 4t$.

(i) Let H^* be as in (g) and let α and β be $(4t-1) \times 1$ nearly level-balanced columns of 1s and -1s such that neither is aliased with any column of H^*, each of $(1,-1), (-1,1)$, and $(-1,-1)$ appears t times and $(1,1)$ appears $t-1$ times as a row of $[\alpha\ \beta]$. Let $X_{d^*} = [\alpha\ \beta\ H^*]$. Then d^* is $E(s^2)$-optimal. Such designs have $n = 4t - 1, m = 4t + 1$.

Theorem 9.6. *All the $E(s^2)$-optimal designs given in (a) – (i) are also $UE(s^2)$-optimal.*

Proof. For each of the cases (a), (c), and (e)-(i), the result follows directly from the lower bound in Theorem 2.1 of Jones and Majumdar (2014). Let d^* be the $E(s^2)$-optimal design specified in (b) and (d). We need to show that it minimizes $\mathrm{tr}\{([1_n\ X_d]^T[1_n\ X_d])^2\} = \mathrm{tr}\{([1_n\ X_d][1_n\ X_d]^T)^2\}$ among all the $n \times m$ matrices X_d with 1 and -1 entries. We note that $[1_n\ X_{d^*}]$ can be obtained by deleting a column of 1's and a level-balanced (or nearly level-balanced) column from an $n \times 4t$ matrix in which any two rows are orthogonal. That d^* minimizes $\mathrm{tr}\{([1\ X_d][1\ X_d]^T)^2\}$ among all the $n \times (4t-3)$ matrices X_d with 1 and -1 entries follows from the fact that $[1_n\ X_{d^*}][1_n\ X_{d^*}]^T$ is of the form

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

where all the diagonal entries of A and B are $4t-2$, all their off-diagonal entries have absolute values 2, and the orders of A and B differ by at most 1. $\square$

The $UE(s^2)$- and $E(s^2)$-optimal designs listed in (a)-(i) are very easy to construct and, since they are level-balanced or nearly level-balanced, SS is automatically minimized. For example, if we want to construct a $UE(s^2)$-optimal design in case (a) by using the method of Jones and Majumdar (2014), then we would have to start with a rather large $4t \times 4t$ Hadamard matrix and choose $2t$ rows from the $4t$ rows. An arbitrary $UE(s^2)$-optimal design so constructed may have a large SS and not so good projection properties.

In contrast, good designs can be obtained very easily by the simple half Hadamard matrix construction. When t is even, an $E(s^2)$- and $UE(s^2)$-optimal design in this case can also be constructed by putting together two $2t \times (2t-1)$ matrices, each of which is obtained by deleting a column of 1's from a $2t \times 2t$ Hadamard matrix. Jones and Majumdar (2014) also observed that such designs are both $E(s^2)$- and $UE(s^2)$-optimal.

Remark 9.7. *It follows from the definition of $UE(s^2)$-optimal designs that level-balanced or nearly level-balanced $UE(s^2)$-optimal designs, i.e., $UE(s^2)$-optimal designs with $SS = 0$ (when n is even) or m (when n is odd) are also $E(s^2)$-optimal. This leaves the possibility of constructing previously unknown $E(s^2)$-optimal designs through our construction of superior $UE(s^2)$-optimal designs.*

For $d \in \mathcal{D}_R(m,n)$, the inner product of each column of X_d and 1 is 0 or -1 when n is even or odd, respectively. Using this, it is easy to see that $(m+1)\,UE_d(s^2) = (m-1)E_d(s^2) + 1 + (-1)^{n+1}$. Thus, since $\min_{d\in\mathcal{D}_R(m,n)} UE_d(s^2) \geq \min_{d\in\mathcal{D}_U(m,n)} UE_d(s^2)$, it follows that

$$(m+1)\min_{d\in\mathcal{D}_U(m,n)} UE_d(s^2) \;\leq\; (m-1)\min_{d\in\mathcal{D}_R(m,n)} E_d(s^2) + 1 + (-1)^{n+1}. \tag{9.11}$$

It follows easily from (9.11) that a necessary and sufficient condition for $E(s^2)$-optimal designs over $\mathcal{D}_R(m,n)$ to be $UE(s^2)$-optimal over $\mathcal{D}_U(m,n)$ is that equality holds in (9.11). In this case, the $E(s^2)$-optimal designs over $\mathcal{D}_R(m,n)$ are also $UE(s^2)$-optimal over $\mathcal{D}_U(m,n)$. In particular, for any m and even n, (9.11) is the same as (2.11) of Jones and Majumdar (2014), and if a lower bound on $E_d(s^2)$ over $\mathcal{D}_R(m,n)$ multiplied by $(m-1)/(m+1)$ is equal to the corresponding lower bound on $UE_d(s^2)$ of Jones and Majumdar (2014), then $E(s^2)$-optimal designs that achieve the lower bound on $E(s^2)$ are also $UE(s^2)$-optimal over $\mathcal{D}_U(m,n)$. The parameter combinations (m,n) for which such equality holds for the sharpest available lower bounds on $E(s^2)$ given by Das et al. (2008) can be determined via some tedious arguments. This leads to the following result, a proof of which is given in the Appendix E.

Theorem 9.8. *For even n, $E(s^2)$-optimal designs achieving the lower bounds on $E_d(s^2)$ given in Das et al. (2008) are also $UE(s^2)$-optimal if and only if one of the following holds:*

(i) $n \equiv 0$ *(mod 2)*, $m = 2(n-1)$,

(ii) $n \equiv 0$ *(mod 4)*, $m = n$,

(iii) $n \equiv 0$ *(mod* $4)$, $m = 4h + 1$, $n/4 \le h \le (n-2)/2$,

(iv) $n \equiv 2$ *(mod* $4)$, $m = 8h + 1$, $(n+2)/8 \le h \le (n-2)/4$.

For example, cases (a), (e), and (f) of Theorem 9.6 correspond to (i), (ii), and (iii) with $h = n/4$, respectively, in Theorem 9.8, and case (b) of the former corresponds to (iii) with $h = (n-2)/2$ and (iv) with $h = (n-2)/4$ in Theorem 9.8.

Let $L(m, n)$ be the lower bound on $E_d(s^2)$ over $d \in \mathcal{D}_R(m, n)$ derived by Das et al. (2008). It is shown in the proof of Theorem 9.8 that $[(m-1)/(m+1)]L(m, n) \geqslant \min_{d \in \mathcal{D}_U(m,n)} UE_d(s^2)$. The four cases in Theorem 9.8 are when the equality holds. In particular, if a design d^* is $E(s^2)$-optimal over $\mathcal{D}_R(m, n)$, then we have $[(m-1)/(m+1)]E_{d^*}(s^2) \geqslant [(m-1)/(m+1)]L(m, n) \geqslant \min_{d \in \mathcal{D}_U(m,n)} UE_d(s^2)$. If d^* is also $UE(s^2)$-optimal, then we must have $[(m-1)/(m+1)]L(m, n) = \min_{d \in \mathcal{D}_U(m,n)} UE_d(s^2)$. It follows that m and n must fall in one of the four cases in Theorem 9.8. This yields a necessary condition for $E(s^2)$-optimal designs over $\mathcal{D}_R(m, n)$ to be also $UE(s^2)$-optimal over $\mathcal{D}_U(m, n)$.

9.6 Concluding Remarks

The $UE(s^2)$-criterion skips the step of minimizing nonorthogonality between the intercept and the main effects. Jones and Majumdar (2014) argued that a consequence of this step is that "the intercept is estimated with the highest efficiency", but "an unintended consequence of the high efficiency of intercept estimation is that it comes at the expense of precision of main effect estimation." It is our opinion that minimizing nonorthogonality between the intercept and the main effects also helps the estimation of main effects and is an important step for achieving good projection properties. Also, minimizing $UE_d(s^2)$ alone produces a large class of $UE(s^2)$-optimal designs that requires secondary criteria to discriminate. An arbitrary $UE(s^2)$-optimal design may have poor projection properties. We have proposed secondary criteria to identify good $UE(s^2)$-optimal designs. A smaller value of *SS* along with minimum Q are common features of many $UE(s^2)$-optimal designs with good projection properties. Although no simple surrogate criterion is expected to always produce the best design, minimizing *SS* followed by minimizing Q is an effective way of getting $UE(s^2)$-optimal designs with good projection properties.

There is no simple general method of constructing $E(s^2)$-optimal designs. We provide easy construction of superior $UE(s^2)$-optimal designs that are almost as efficient as $E(s^2)$-optimal designs (where available) with respect to the D_f-criteria.

Chapter 10

Lower bounds on the sizes of t-(v,k,λ) coverings

This chapter is based on the following work:

Horsley and Singh (2018): Horsley, Daniel; Singh, Rakhi. New lower bounds for t-coverings. *J. Combin. Des 26* (2018), no. 8, 369–386.

10.1 Introduction

For our purposes, an *incidence structure* is a pair $(V, \mathcal{B})$ where V is a set of *points* and $\mathcal{B}$ is a multiset of subsets of V called *blocks*. For positive integers t, v, k and λ with $t \leqslant k \leqslant v$, a t-(v,k,λ) *covering* is an incidence structure $(V, \mathcal{B})$ such that $|V| = v$, $|B| = k$ for all $B \in \mathcal{B}$, and each t-subset of V is contained in at least λ blocks in $\mathcal{B}$. If each t-subset of V is contained in exactly λ blocks in $\mathcal{B}$, then $(V, \mathcal{B})$ is a t-(v,k,λ) *design*. For an incidence structure $(V, \mathcal{B})$ and a subset $X \subseteq V$, define $b(X)$ to be the number of blocks in $\mathcal{B}$ that contain X. Coverings were introduced for $t = 2$ by Erdős and Rényi (1956) and then generalised to arbitrary t by Erdős and Hanani (1963).

Usually we are interested in finding coverings with as few blocks as possible. The *covering number* $C_\lambda(v,k,t)$ is the minimum number of blocks in any t-(v,k,λ) covering. When $\lambda = 1$ we omit the subscript. It is convenient to set $C_\lambda(v,k,0) = \lambda$ for all v, k and λ. In Rödl (1985) introduced the famous *nibble* method to show that $C(v,k,t) \sim \binom{v}{t}/\binom{k}{t}$ as $v \to \infty$.

Observe that if $(V, \mathcal{B})$ is a t-(v,k,λ) covering and X is a subset of V with $|X| \leqslant t$, then $(V \setminus X, \mathcal{B}')$, where $\mathcal{B}' = \{B \setminus X : B \in \mathcal{B}, X \subseteq B\}$, is a $(t-|X|)$-$(v-|X|, k-|X|, \lambda)$ covering and hence

$$b(X) \geqslant C_\lambda(v-|X|, k-|X|, t-|X|). \tag{10.1}$$

Using this fact with $|X| = 1$ and some simple counting gives

$$C_\lambda(v,k,t) \geqslant \left\lceil \frac{v}{k}\, C_\lambda(v-1,k-1,t-1) \right\rceil. \tag{10.2}$$

Iterating this inequality yields the *Schönheim bound* (Schönheim, 1964) which states that $C_\lambda(v,k,t) \geqslant L_\lambda(v,k,t)$ where

$$L_\lambda(v,k,t) = \left\lceil \frac{v}{k} \left\lceil \frac{v-1}{k-1} \cdots \left\lceil \frac{v-t+2}{k-t+2} \left\lceil \frac{\lambda(v-t+1)}{k-t+1} \right\rceil \right\rceil \cdots \right\rceil \right\rceil.$$

Furthermore, Mills and Mullin (1992) have shown that if $vC_\lambda(v-1,k-1,t-1) \not\equiv 0 \pmod{k}$ and $C_\lambda(v-1,k-1,t-1) = (\binom{v-1}{r-1}/\binom{k-1}{r-1})C_\lambda(v-r,k-r,t-r)$ for some $r \in \{2,\dots,t\}$, then

$$C_\lambda(v,k,t) \geqslant \left\lceil \frac{v}{k}(C_\lambda(v-1,k-1,t-1)+r) \right\rceil. \tag{10.3}$$

This result is easiest to apply in the case $r = t = 2$, when it states that if $\lambda(v-1) \equiv 0 \pmod{k-1}$ and $\lambda v(v-1) \equiv 1 \pmod{k}$, then $C_\lambda(v,k,t) \geqslant L_\lambda(v,k,t)+1$. A result (Theorem 6.5) of Keevash (2014) implies that, for a fixed t, k and λ and for all sufficiently large v, $C_\lambda(v,k,t) = h_\lambda(v,k,t)/\binom{k}{t}$ where $h_\lambda(v,k,t)$ is the size of a smallest t-(v,t,λ) covering $(V,\mathcal{B})$ with the property that $\binom{k-|X|}{t-|X|}$ divides $b(X)$ for each subset X of V with $|X| \leqslant t$. In the case $t = 2$, this establishes that the Schönheim bound with the Mills and Mullin improvement is tight for all sufficiently large v. Glock et al. (2016) have recently extended Keevash's main result.

Our interest here is principally in establishing lower bounds for covering numbers $C_\lambda(v,k,t)$ when k is a significant fraction of v. Exact values for $C_\lambda(v,k,t)$ have been determined when $(k,t) \in \{(3,2),(4,2)\}$, when $(t,\lambda) = (2,1)$ and $v \leqslant \frac{13}{4}k$, and for most cases when $(t,\lambda) = (3,1)$ and $v \leqslant \frac{8}{5}k$ (see Gordon and Stinson (2007)). In the case $t = 2$, a number of results have been proved which improve on the Schönheim bound in various cases where k is a significant fraction of v Bluskov et al. (2000); Bose and Connor (1952); Bryant et al. (2011); Füredi (1990); Todorov (1984, 1989). A number of other lower bounds for specific parameter sets, which have been mostly obtained by computer searches, are available in literature (see Gordon (n.d.); Gordon and Stinson (2007)). For surveys on coverings see Gordon and Stinson (2007); Mills and Mullin (1992). Gordon maintains a repository for small coverings Gordon (n.d.).

Fisher's inequality (Fisher, 1940) famously states that every 2-(v, k, λ) design with $v > k$ has at least v blocks. Ray-Chaudhuri and Wilson (1975) generalised this result to higher t by showing that every t-(v, k, λ) design with $v \geqslant k + s$ has at least $\binom{v}{s}$ blocks for any positive integer $s \leqslant \lfloor \frac{t}{2} \rfloor$. Subsequently Wilson (1982) gave an alternate proof of this generalised result using so-called *higher incidence matrices*. In this chapter we demonstrate how an approach based on these matrices can be used to obtain improved lower bounds on covering numbers $C_\lambda(v, k, t)$. Our results generalise both the results of Ray-Chaudhuri and Wilson (1975) and the more recent results of Horsley (2017) which established lower bounds for $C_\lambda(v, k, 2)$.

To avoid triviality, we often consider only t-(v, k, λ) coverings with $2 < k < v$. The bounds we prove in this chapter apply to covering numbers $C_\lambda(v, k, t)$ for arbitrary λ. However in our discussions, as in most of the literature concerning coverings with $t \geqslant 3$, we will concentrate on the case $\lambda = 1$. The methods in this chapter should also be applicable to packings, but we do not pursue this here.

In the next section we discuss our proof strategy and prove some preliminary results. In Sections 10.3, 10.5 and 10.6 we then prove and discuss bounds that generalise Theorems 1, 11 and 14 of Horsley (2017) respectively. The results in Sections 10.5 and 10.6 make use of a result of Caro and Tuza (1991) which guarantees an m-independent set of a certain size in a multigraph with a specified degree sequence. In Section 10.4 we exhibit infinite families of parameter sets t-(v, k, λ) for which our results improve on the best bounds previously known.

10.2 Strategy and preliminary results

To prove our results we will combine ideas from Horsley (2017) with those from a proof by Wilson (1982) of the generalisation of Fisher's inequality to higher t. The methods in Horsley (2017) were, in turn, inspired by a proof by Bose (1949) of Fisher's inequality. Following Wilson (1982), we make use of higher incidence matrices. For a nonnegative integer s, the *s-incidence matrix* of an incidence structure $(V, \mathcal{B})$ is the matrix whose rows are indexed by the s-subsets of V, whose columns are indexed by the blocks in $\mathcal{B}$, and where the entry in row X and column B is 1 if $X \subseteq B$ and 0 otherwise. For a set V and a nonnegative integer i, let $\binom{V}{i}$ denote the set of all i-subsets of V.

We will make use of standard facts about positive definite matrices (see (Hogben, 2013, §9.4)). If A is a square matrix whose rows and columns are indexed by the elements of a set Z, then a *principal submatrix* of A is a square submatrix whose rows and columns are both indexed by the same subset Z' of Z. We say a real matrix is *diagonally dominant* if, in each of its rows, the magnitude of the diagonal entry is strictly greater than the sum of the magnitudes of the other entries in that row. It follows easily from the well-known Gershgorin circle theorem (see (Hogben, 2013, p16-6)) that real diagonally dominant matrices are positive definite. Our bounds rest on the following simple observations.

Lemma 10.1. *Let $(V, \mathcal{B})$ be an incidence structure and let A be the s-incidence matrix of $(V, \mathcal{B})$ for some positive integer s. Then*

(i) AA^T is the symmetric matrix whose row and columns are indexed by $\binom{V}{s}$ and where the entry in row X and column Y is $b(X \cup Y)$; and

(ii) $|\mathcal{B}| \geqslant \operatorname{rank}(AA^T)$.

Proof. Part (i) follows from the definition of matrix multiplication. Because A has only $|\mathcal{B}|$ columns, $\operatorname{rank}(A) \leqslant |\mathcal{B}|$. Thus $|\mathcal{B}| \geqslant \operatorname{rank}(A) \geqslant \operatorname{rank}(AA^T)$, proving part (ii). □

By Lemma 10.1 we can bound the number of blocks in a covering by bounding $\operatorname{rank}(AA^T)$. Our strategy to bound this rank is as follows. We first write $AA^T = P + M$ where P is positive semidefinite. We then find a diagonally dominant, and hence positive definite, principal submatrix M' of M. Because every principal submatrix of P is positive semidefinite, the submatrix of AA^T with row and column indices corresponding to those of M' is positive definite and hence full rank. Thus the rank of AA^T is at least the order of M'.

We choose P so that the entry in row X and column Y for $X \neq Y$ is $b_{|X \cup Y|}$, where $b_{s+1}, \ldots, b_{2s}$ are positive integers chosen so that each i-subset of V is in at least b_i blocks in $\mathcal{B}$ for $i \in \{s+1, \ldots, 2s\}$. The entries on the lead diagonal of P are chosen to be small as possible, given that P must be positive semidefinite. We establish that P is indeed positive semidefinite using an approach from Wilson (1982) in which P is written as a nonnegative linear combination of Gram matrices.

We will require the following simple identity for binomial coefficients.

Lemma 10.2. *Let i and ℓ be nonnegative integers with $i \leqslant \ell$. Then*

$$\sum_{j=i}^{\ell}(-1)^{i+j}\tbinom{\ell}{j}\tbinom{j}{i} = \begin{cases} 0, & \text{if } i < \ell; \\ 1, & \text{if } i = \ell. \end{cases}$$

Proof. The multinomial theorem implies that the coefficient of x^i in the expansion of $(x-1+1)^\ell$ is

$$\sum_{j'=0}^{\ell-i}\tbinom{\ell}{i+j'}\tbinom{i+j'}{i}(-1)^{j'} = \sum_{j=i}^{\ell}(-1)^{i+j}\tbinom{\ell}{j}\tbinom{j}{i},$$

where the equality is obtained by substituting $j = i + j'$. So because $(x-1+1)^\ell = x^\ell$, the result now follows by equating the coefficients of x^i. $\square$

The next lemma establishes that if A is the higher incidence matrix of a t-(v,k,λ) covering, then AA^T has a specific form that we can exploit. Subsequent results in this chapter will often explicitly assume the hypotheses of Lemma 10.3 and use its notation.

Lemma 10.3. *Let t, v, k, λ and s be positive integers such that $t < k < v$ and $s \leqslant \lfloor \frac{t}{2} \rfloor$. Let $b_{2s}, b_{2s-1}, \dots, b_s$ be positive integers such that*

(i) $L_\lambda(v-2s, k-2s, t-2s) \leqslant b_{2s} \leqslant C_\lambda(v-2s, k-2s, t-2s)$;

(ii) $\lceil \frac{v-i}{k-i} b_{i+1} \rceil \leqslant b_i \leqslant C_\lambda(v-i, k-i, t-i)$ for $i = 2s-1, 2s-2, \dots, s$; and

(iii) $a_j \geqslant 0$ for $j \in \{0, \dots, s\}$, where $a_j = \sum_{i=0}^{j}(-1)^{i+j}\binom{j}{i} b_{2s-i}$.

If $(V, \mathcal{B})$ is a t-(v,k,λ) covering and A is the s-incidence matrix of $(V,\mathcal{B})$, then $b(Z) \geqslant b_{|Z|}$ for any $Z \subseteq V$ with $|Z| \in \{s, \dots, 2s\}$ and $AA^T = P + M$ for matrices $P = (p_{XY})$ and $M = (m_{XY})$ such that

$$p_{XY} = \begin{cases} b_{|X\cup Y|} & \text{if } X \neq Y \\ b_s - a_s & \text{if } X = Y \end{cases} \qquad m_{XY} = \begin{cases} b(X \cup Y) - b_{|X\cup Y|} & \text{if } X \neq Y \\ a_s + b(X) - b_s & \text{if } X = Y \end{cases}.$$

Furthermore, the following hold.

(a) $P = \sum_{j=0}^{s-1} a_j Q_j^T Q_j$, where Q_j is the j-incidence matrix of the incidence structure $(V, \binom{V}{s})$. Hence P is positive semidefinite.

(b) For any $X \in \binom{V}{s}$,

$$\sum_{Y \in \binom{V}{s} \setminus \{X\}} m_{XY} = \sum_{Y \in \binom{V}{s} \setminus \{X\}} \left(b(X\cup Y) - b_{|X\cup Y|}\right) = \left(b(X) - b_s\right)\left(\tbinom{k}{s} - 1\right) + d$$

where $d = b_s(\binom{k}{s} - 1) - \sum_{i=0}^{s-1}\binom{s}{i}\binom{v-s}{s-i} b_{2s-i}$ is a nonnegative integer.

Proof. Let $Z \subseteq V$ with $|Z| \in \{s, \dots, 2s\}$. That $b(Z) \geqslant b_{|Z|}$ follows because $b_{|Z|} \leqslant C_\lambda(v - |Z|, k - |Z|, t - |Z|)$ by (i) and (ii) and $C_\lambda(v - |Z|, k - |Z|, t - |Z|) \leqslant b(Z)$ by (10.1). That $AA^T = P + M$ follows immediately from Lemma 10.1 (i) and the definitions of P and M. Let $\mathcal{V} = \binom{V}{s}$ and $\mathcal{V}_0 = \{X \in \mathcal{V} : b(X) = b_s\}$.

We prove (a). Observe that for $j \in \{0, \dots, s\}$, $Q_j^T Q_j$ is the matrix whose rows and columns are indexed by $\binom{V}{s}$ and whose (X, Y) entry is $\binom{|X \cap Y|}{j}$ for all $X, Y \in \binom{V}{s}$. In particular, $Q_s^T Q_s = I$. Let

$$Q' = \sum_{j=0}^{s} a_j Q_j^T Q_j = a_s I + \sum_{j=0}^{s-1} a_j Q_j^T Q_j.$$

It suffices to show that $Q' = a_s I + P$.

Let $X, Y \in \binom{V}{s}$, let $\ell = |X \cap Y|$, and note that $\ell \leqslant s$. For $j \in \{0, \dots, s\}$, the (X, Y) entry of $Q_j^T Q_j$ is $\binom{\ell}{j}$. Thus the (X, Y) entry of Q' is

$$\sum_{j=0}^{s} a_j \tbinom{\ell}{j} = \sum_{j=0}^{s} \sum_{i=0}^{j} (-1)^{i+j} \tbinom{\ell}{j} \tbinom{j}{i} b_{2s-i} = \sum_{i=0}^{s} \sum_{j=i}^{s} (-1)^{i+j} \tbinom{\ell}{j} \tbinom{j}{i} b_{2s-i}.$$

So it follows from Lemma 10.2 that the (X, Y) entry of Q' is $b_{2s-\ell} = b_{|X \cup Y|}$. Thus $Q' = a_s I + P$.

Now we prove (b). For each $X \in \mathcal{V}$,

$$\sum_{Y \in \mathcal{V} \setminus \{X\}} b(X \cup Y) = b(X) \left(\tbinom{k}{s} - 1 \right)$$

because each block that contains X contributes $\binom{k}{s} - 1$ to this sum. Also for each $X \in \mathcal{V}$,

$$\sum_{Y \in \mathcal{V} \setminus \{X\}} b_{|X \cup Y|} = \sum_{i=0}^{s-1} \tbinom{s}{i} \tbinom{v-s}{s-i} b_{2s-i}$$

because, for each $i \in \{0, \dots, s-1\}$, $|\{Y : |X \cap Y| = i\}| = \binom{s}{i}\binom{v-s}{s-i}$. Together, these facts imply that (b) holds provided d is nonnegative. By (ii), $b_{i+1} \leqslant \frac{k-i}{v-i} b_i$ for $i = 2s-1, 2s-2, \dots, s$ and so it can be seen that $b_{2s-i} \leqslant (\binom{k-s}{s-i}/\binom{v-s}{s-i}) b_s$ for $i = s-1, s-2, \dots, 0$. Thus,

$$\sum_{i=0}^{s-1} \tbinom{s}{i} \tbinom{v-s}{s-i} b_{2s-i} \leqslant b_s \sum_{i=0}^{s-1} \tbinom{s}{i} \tbinom{k-s}{s-i} = b_s(\tbinom{k}{s} - 1),$$

and it follows that $d \geqslant 0$. $\square$

Remark 10.4. *In many cases condition (ii) of Lemma 10.3 implies condition (iii). Specifically, we claim that if condition (ii) is satisfied then $a_j \geqslant 0$ for $j \in \{0, \dots, \min(\lfloor \frac{v}{k} \rfloor, s)\}$.*

This means that we can ignore condition (iii) whenever $v \geqslant sk$. *Certainly,* $a_0 = b_{2s} \geqslant 1$. *To see that the rest of our claim is true, fix* $j \in \{1, \ldots, \min(\lfloor \frac{v}{k} \rfloor, s)\}$, *and let* $\delta = 2$ *if* j *is even and* $\delta = 1$ *if* j *is odd. Then, pairing consecutive terms in the definition of* a_j, *we see that*

$$a_j \geqslant \sum_{i \in \{\delta, \delta+2, \ldots, j\}} \left(\tbinom{j}{i} b_{2s-i} - \tbinom{j}{i-1} b_{2s-i+1} \right).$$

For $i \in \{\delta, \delta + 2, \ldots, j\}$, *using condition (ii),*

$$\tbinom{j}{i} = \tfrac{j-i+1}{i} \tbinom{j}{i-1} \geqslant \tfrac{1}{j} \tbinom{j}{i-1} \quad \text{and} \quad b_{2s-i} \geqslant \left\lceil \tfrac{v-2s+i}{k-2s+i} b_{2s-i+1} \right\rceil \geqslant \tfrac{v}{k} b_{2s-i+1} \geqslant j b_{2s-i+1},$$

and hence $\binom{j}{i} b_{2s-i} \geqslant \binom{j}{i-1} b_{2s-i+1}$. *Thus* $a_j \geqslant 0$.

It follows from Lemma 10.3(a) that the diagonal entries $b_s - a_s$ of P are at least $a_0 = b_{2s} > 0$. Hence $b_s > a_s$. This fact will be used several times in later sections. We are now ready to prove Lemma 10.5, which forms the basis of all the lower bounds that we establish in this chapter.

Lemma 10.5. *Suppose the hypotheses of Lemma 10.3 hold. If there is a subset* $\mathcal{S}$ *of* $\binom{V}{s}$ *such that, for each* $X \in \mathcal{S}$,

$$\sum_{Y \in \mathcal{S} \setminus \{X\}} \left(b(X \cup Y) - b_{|X \cup Y|} \right) < a_s + b(X) - b_s,$$

then $|\mathcal{B}| \geqslant |\mathcal{S}|$.

Proof. By Lemma 10.1 (ii), it suffices to show that the principal submatrix of AA^T whose rows and columns are indexed by $\mathcal{S}$ is positive definite and hence full rank.

By Lemma 10.3, AA^T can be written as the sum of a positive semidefinite matrix P and a matrix M whose (X, Y) entry is the nonnegative integer $b(X \cup Y) - b_{|X \cup Y|}$ for all distinct $X, Y \in \binom{V}{s}$ and whose (X, X) entry is the nonnegative integer $a_s + b(X) - b_s$ for all $X \in \binom{V}{s}$. Because every principal submatrix of P is positive semidefinite, it in fact suffices to show that the principal submatrix M' of M whose rows and columns are indexed by $\mathcal{S}$ is positive definite. Given the hypothesis of the lemma that

$$\sum_{Y \in \mathcal{S} \setminus \{X\}} \left(b(X \cup Y) - b_{|X \cup Y|} \right) < a_s + b(X) - b_s,$$

M' is diagonally dominant and hence it is positive definite by the Gershgorin circle theorem (see (Hogben, 2013, p.16-6)). $\square$

10.3 Basic bound

Here we use Lemma 10.5 to prove the simplest and most easily stated of our results, and then discuss when it can be usefully applied.

Theorem 10.6. *Suppose the hypotheses of Lemma 10.3 hold and that $d < a_s$. Then*

$$C_\lambda(v,k,t) \geqslant \left\lceil \frac{\binom{v}{s}(b_s+1)}{\binom{k}{s}+1} \right\rceil.$$

Proof. Let $(V, \mathcal{B})$ be a t-(v,k,λ) covering. Let $\mathcal{V} = \binom{V}{s}$ and $\mathcal{V}_0 = \{X \in \mathcal{V} : b(X) = b_s\}$. Because $d < a_s$, it follows from Lemma 10.3(b) that we can apply Lemma 10.5 with $\mathcal{S} = \mathcal{V}_0$ and hence conclude that $|\mathcal{B}| \geqslant |\mathcal{V}_0|$.

Since each block in $\mathcal{B}$ covers $\binom{k}{s}$ sets in $\mathcal{V}$, we have that $\sum_{X \in \mathcal{V}} b(X) = |\mathcal{B}|\binom{k}{s}$. Thus

$$|\{X \in \mathcal{V} : b(X) > b_s\}| \leqslant |\mathcal{B}|\tbinom{k}{s} - \tbinom{v}{s} b_s$$

because $b(X) \geqslant b_s$ for each $X \in \mathcal{V}$. It follows that $|\mathcal{B}| \geqslant |\mathcal{V}_0| \geqslant \binom{v}{s} - (|\mathcal{B}|\binom{k}{s} - \binom{v}{s}b_s)$. A simple calculation now establishes that

$$|\mathcal{B}| \geqslant \frac{\binom{v}{s}(b_s+1)}{\binom{k}{s}+1}. \qquad \square$$

It is useless to apply Theorem 10.6 with b_s chosen to be less than the best known lower bound for $C_\lambda(v-s, k-s, t-s)$, because the bound of Theorem 10.6 is always inferior to the bound given by s iterated applications of (10.2) to $b_s + 1$ (note this latter bound is at least $\lceil b_s\binom{v}{s}/\binom{k}{s}\rceil$). Furthermore, from the definitions of d and a_s we have that

$$a_s - d = \left(\sum_{i=0}^{s-1} \tbinom{s}{i}\left(\tbinom{v-s}{s-i} + (-1)^{s-i}\right) b_{2s-i} \right) - \left(\tbinom{k}{s} - 2\right) b_s, \tag{10.4}$$

which is increasing in b_{2s-i} for each $i \in \{0, \dots, s-1\}$. Thus, in the absence of condition (iii) of Lemma 10.3, it can be seen that when attempting to apply Theorem 10.6 we only need consider choosing b_i to be the best known lower bound on $C_\lambda(v-i, k-i, t-i)$ for $i \in \{s, \dots, 2s\}$. Throughout the rest of the chapter, we shall refer to this as the natural choice for the b_i. Condition (iii) complicates the picture somewhat, but in view of Remark 10.4 this is only of concern when $v \leqslant (s-1)k$ (note that $a_s > d \geqslant 0$ by our hypotheses and Lemma 10.3). In many cases the best known lower bounds are all given by the Schönheim bound and in these cases the natural choice of the b_i amounts to taking $b_i = L_\lambda(v-i, k-i, t-i)$ for $i \in \{s, \dots, 2s\}$.

For each of the subsequent lower bounds we establish in this chapter (see Theorems 10.15 and 10.18), we will also show that we only need consider the natural choice for b_s. With this choice fixed, the natural choice for the remaining b_i will minimise d and maximise $a_s - d$, by the definition of d and (10.4). Considering this and Remark 10.4, we believe that taking the natural choice for the b_i in our theorems will almost always produce the best results.

For the Theorem 10.6 bound to exceed the bound obtained by s iterated applications of (10.2) to b_s, it must be the case that $b_s < \binom{k}{s}$ (again note the latter bound is at least $\lceil b_s\binom{v}{s}/\binom{k}{s}\rceil$). Furthermore, the other lower bounds we establish in this chapter will explicitly require $b_s < \binom{k}{s}$. We have $b_s < \binom{k}{s}$ only when $v < (\frac{k^t}{s!\lambda})^{1/(t-s)}$ because $\binom{k}{s} \leqslant \frac{k^s}{s!}$ and $b_s \geqslant L_\lambda(v-s, k-s, t-s) \geqslant \lambda\binom{v-s}{t-s}/\binom{k-s}{t-s} > \lambda(\frac{v}{k})^{t-s}$. So none of the lower bounds of this chapter are of use when $v \geqslant (\frac{k^t}{s!\lambda})^{1/(t-s)}$.

Theorem 10.6 implies Ray-Chaudhuri and Wilson (1975) generalisation of Fisher's inequality. If there exists a t-(v, k, λ) design $(V, \mathcal{B})$ with $v \geqslant k+s$ for some positive integer $s \leqslant \lfloor \frac{t}{2} \rfloor$, then applying Theorem 10.6 with $b_i = L_\lambda(v-i, k-i, t-i) = \lambda\binom{v-i}{t-i}/\binom{k-i}{t-i}$ for $i \in \{s, \ldots, 2s\}$ we have $C_\lambda(v, k, t) \geqslant \binom{v}{s}(b_s+1)/(\binom{k}{s}+1)$ (the hypotheses are satisfied because $d = 0$ and $a_j = \lambda\binom{v-2s}{k-2s+j}/\binom{v-t}{k-t}$ for $j \in \{0, \ldots, s\}$). But, because $(V, \mathcal{B})$ is a design, it has exactly $\binom{v}{s}b_s/\binom{k}{s}$ blocks. So we can conclude that $\binom{v}{s}b_s/\binom{k}{s} \geqslant \binom{v}{s}(b_s+1)/(\binom{k}{s}+1)$ which implies $b_s \geqslant \binom{k}{s}$ and hence that $(V, \mathcal{B})$ has at least $\binom{v}{s}$ blocks.

10.4 Infinite families of improvements

In this section we first give, in Lemma 10.7, an infinite family of parameter sets for which applying Theorem 10.6 with $s = 2$ yields an improvement over the Schönheim bound. Then we exhibit, in Theorem 10.10, an infinite family of parameter sets for which applying Theorem 10.6 with $s = 1$ establishes exact covering numbers. In this section we will often use the simple observation that, for given t, k and λ, $C_\lambda(v, k, t) \leqslant C_\lambda(v', k, t)$ when $v \leqslant v'$.

Lemma 10.7. *Let $m \geqslant 6$ be an integer, and let $v = m^2(m-2)+4$ and $k = m(m-1)+2$. An application of Theorem 10.6 with $s = 2$ establishes that $C(v, k, 5) \geqslant L(v, k, 5)+m(m-4)-10$.*

Proof. Let $\ell_i = L(v-i, k-i, t-i)$ for $i = 4, 3, 2$. We can successively calculate

$$\begin{aligned}
\ell_4 &= \lceil \tfrac{v-4}{k-4} \rceil = \lceil m - 1 + \tfrac{m-2}{m(m-1)-2} \rceil = m \\
\ell_3 &= \lceil \tfrac{v-3}{k-3} \ell_4 \rceil = m(m-1) \\
\ell_2 &= \lceil \tfrac{v-2}{k-2} \ell_3 \rceil = m^2(m-2) + 2.
\end{aligned}$$

We will apply Theorem 10.6 with $s = 2$ and $b_i = \ell_i$ for $i = 4, 3, 2$. Routine calculations show that, in the terminology of Lemma 10.3, $a_0 = m$, $a_1 = m(m-2)$, $a_2 = m^3 - 4m^2 + 3m + 2$, and $d = 0$. Using this, and recalling that $m \geqslant 6$, it can be seen that the hypotheses of Theorem 10.6 are satisfied and hence

$$C(v, k, 5) \geqslant \left\lceil \tfrac{v(v-1)}{k(k-1)+2} (\ell_2 + 1) \right\rceil .$$

This implies that $C(v, k, 5) \geqslant m^5 - 4m^4 + 21m^2 - 14m - 55$.

On the other hand,

$$L(v, k, 5) = \lceil \tfrac{v}{k} \lceil \tfrac{v-1}{k-1} \ell_2 \rceil \rceil$$

and for $m \geqslant 14$ we can calculate that this is equal to $m^5 - 4m^4 + 20m^2 - 10m - 45$. Thus it can be seen that the lemma holds for $m \geqslant 14$, and it is routine to check it holds for $6 \leqslant m \leqslant 13$. $\square$

Further routine calculations establish that, for v and k as in Lemma 10.7, neither the result of Mills and Mullin (1992) nor the results of this chapter (including those in Sections 10.5 and 10.6) give improvements over the Schönheim bound for the parameter sets $C(v-1, k-1, 4)$, $C(v-2, k-2, 3)$ or $C(v-3, k-3, 2)$. We believe that, in general, no bound better than the Schönheim bound was previously known for this family of parameter sets. Since $d = 0$ in our application of Theorem 10.6, we could make a slight further improvement to this result by instead applying Theorem 10.18(a) below.

We now move on to show that Theorem 10.6 with $s = 1$ can be applied to establish that certain coverings constructed from affine planes are optimal, and thus obtain a family of exact covering numbers.

Let q be a prime power. It is well known (see Gordon et al. (1995), for example) that if we take V to be the q^t points of the affine geometry $AG(t, q)$ and $\mathcal{B}$ to be the set of its $(t-1)$-flats, then $(V, \mathcal{B})$ is a t-$(q^t, q^{t-1}, 1)$ covering with $q(\frac{q^t-1}{q-1})$ blocks. Further, it is straightforward to calculate that $L(q^t, q^{t-1}, t) = q(\frac{q^t-1}{q-1})$ and hence $C(q^t, q^{t-1}, t) = q(\frac{q^t-1}{q-1})$. The following lemma is based on a well-known "blow up" construction for coverings.

Lemma 10.8. *Let m, t and q be positive integers such that q is a prime power. Then $C(v, mq^{t-1}, t) \leqslant q(\frac{q^t-1}{q-1})$ for each $v \leqslant mq^t$.*

Proof. Let $(U, \mathcal{A})$ be the t-$(q^t, q^{t-1}, 1)$ covering with $q(\frac{q^t-1}{q-1})$ blocks obtained from the $(t-1)$-flats of $AG(t, q)$. Let M be a set of m elements, let $V = U \times M$ and let $\mathcal{B} = \{A \times M : A \in \mathcal{A}\}$. Then $(V, \mathcal{B})$ is an $(mq^t, mq^{t-1}, 1)$-covering with $q(\frac{q^t-1}{q-1})$ blocks. The result now follows because $C(v-1, k, t) \leqslant C(v, k, t)$ for any parameter set (v, k, t). $\square$

Next we determine the value of the Schönheim bound in the cases we are concerned with.

Lemma 10.9. *Let v, m, q and t be positive integers such that q is a prime power, $m \geqslant 2q+2$, $2 \leqslant t < mq^{t-1}$, and $mq^t - 2q + 3 \leqslant v \leqslant mq^t$. Let $\ell_t = 1$ and let $\ell_i = L(v-i, mq^{t-1}-i, t-i)$ for $i = t-1, t-2, \ldots, 0$. Then*

(i) $\ell_i = \frac{q^{t-i+1}-1}{q-1}$ *for* $i = t-1, t-2, \ldots, 0$;

(ii) $\ell_1 = \begin{cases} \frac{q^t-1}{q-1} & \text{if } mq^t - q + 2 \leqslant v \leqslant mq^t \\ q(\frac{q^{t-1}-1}{q-1}) & \text{if } mq^t - 2q + 3 \leqslant v \leqslant mq^t - q + 1; \end{cases}$

(iii) $\ell_0 = \begin{cases} q(\frac{q^t-1}{q-1}) & \text{if } mq^t - q + 2 \leqslant v \leqslant mq^t \\ q^2(\frac{q^{t-1}-1}{q-1}) & \text{if } mq^t - 2q + 3 \leqslant v \leqslant mq^t - q + 1. \end{cases}$

Proof. Let c be the integer such that $v = mq^t - q + 1 + c$. By definition, for $i = t-1, t-2, \ldots, 0$,

$$\ell_i = \left\lceil \frac{(mq^t - q + 1 + c - i)\ell_{i+1}}{mq^{t-1} - i} \right\rceil = q\ell_{i+1} + \left\lceil \frac{((i-1)(q-1)+c)\ell_{i+1}}{mq^{t-1} - i} \right\rceil. \qquad (10.5)$$

Since $c \in \{-q+2, \ldots, q-1\}$, (10.5) implies that $\ell_i = q\ell_{i+1} + 1$ for $i \geqslant 2$, provided $\ell_{i+1} \leqslant \frac{mq^{t-1}-i}{i(q-1)}$. Using this fact, it is easy to prove (i) by induction on i. In particular, we have $\ell_2 = \frac{q^{t-1}-1}{q-1}$, and applying (10.5) once more establishes (ii). Applying (10.5) one final time using (ii) and the hypothesis $m \geqslant 2q+2$ establishes (iii). $\square$

Together, Lemmas 10.8 and 10.9 establish the known result that, under the hypotheses of Lemma 10.9, $C(v, mq^{t-1}, t) = q(\frac{q^t-1}{q-1})$ for $v \in \{mq^t - q + 2, \ldots, mq^t\}$. By applying Theorem 10.6 with $s = 1$ we can strengthen this result to cover some cases where $v \leqslant mq^t - q + 1$.

Theorem 10.10. *Let m, q and t be positive integers such that q is a prime power, $m \geqslant 2q+2$ and $2 \leqslant t < mq^{t-1}$. Then $C(v, mq^{t-1}, t) = q(\frac{q^t-1}{q-1})$ for each integer v such that*

$$mq^t - q + 1 - z \leqslant v \leqslant mq^t \qquad \textit{where} \qquad z = \min\left(q-2, \left\lfloor \frac{m(q-1)q^{t-1}}{q^t-1} \right\rfloor - 2q + 1\right).$$

Proof. Note that $z \geqslant 0$ because $m \geqslant 2q+2$. Let $v' = mq^t - q + 1 - z$. It suffices to show that $C(v', mq^{t-1}, t) \geqslant q(\frac{q^t-1}{q-1})$, because then, for each integer v such that $v' \leqslant v \leqslant mq^t$, we have

$$q\left(\frac{q^t-1}{q-1}\right) \leqslant C(v', mq^{t-1}, t) \leqslant C(v, mq^{t-1}, t) \leqslant C(mq^t, mq^{t-1}, t) \leqslant q\left(\frac{q^t-1}{q-1}\right),$$

where the final inequality follows from Lemma 10.8.

For $i \in \{0,1,2\}$, let $\ell_i = L(v'-i, mq^{t-1}-i, t-i)$. By Lemma 10.9, $\ell_1 = q(\frac{q^{t-1}-1}{q-1})$ and $\ell_2 = \frac{q^{t-1}-1}{q-1}$. To bound $C(v', mq^{t-1}, t)$ below, we will apply Theorem 10.6 with $s = 1$, $b_1 = \ell_1$ and $b_2 = \ell_2$. Obviously this choice satisfies hypotheses (i) and (ii) of Lemma 10.3. Because $v' \geqslant mq^t - 2q + 3$, a simple calculation establishes that $\ell_1(mq^{t-1}-2) < \ell_2(v'-2)$ and thus $d < a_1$ (because $d \geqslant 0$, this also implies that $a_1 \geqslant 0$ and that hypothesis (iii) of Lemma 10.3 holds). So, by Theorem 10.6, we have

$$C(v,k,t) \geqslant \left\lceil \frac{v'(\ell_1+1)}{mq^{t-1}+1} \right\rceil = q(\ell_1+1) - \left\lfloor \frac{(2q+z-1)(\ell_1+1)}{mq^{t-1}+1} \right\rfloor.$$

A routine calculation shows that the second upper bound on z in our hypotheses is equivalent to $(2q+z-1)(\ell_1+1) \leqslant mq^{t-1}$ and hence $C(v,k,t) \geqslant q(\ell_1+1)$. Observing that $q(\ell_1+1) = q(\frac{q^t-1}{q-1})$ completes the proof. □

Corollary 10.11. *Let m, q and t be positive integers such that q is a prime power, $m \geqslant 3q$ and $2 \leqslant t < mq^{t-1}$. Then $C(v, mq^{t-1}, t) = q(\frac{q^t-1}{q-1})$ for each integer v such that $mq^t - 2q + 3 \leqslant v \leqslant mq^t$.*

Proof. This follows by observing that, in Theorem 10.10, $z = q-2$ if $m \geqslant 3q$. □

10.5 Bounds for the case $d \geqslant a_s$

Using the terminology of Lemma 10.3, Theorem 10.6 applies only when $d < a_s$. In this section we will establish a bound that can be applied when $d \geqslant a_s$. For a multigraph G and a subset S of $V(G)$, let $G[S]$ denote the sub-multigraph of G induced by S. In this section and the next, we will make use of the notion of an n-independent set in a multigraph G, which is defined as a subset S of $V(G)$ such that $G[S]$ has maximum degree strictly less than n. Setting $n = 1$ recovers the usual notion of an independent set. Let $\mu_G(xy)$ denote the number of edges between vertices x and y in a multigraph G.

If M is the matrix defined in Lemma 10.3 and G is the multigraph whose adjacency matrix agrees with M in its off-diagonal entries, then an n-independent set in G corresponds to a principal submatrix of M in which the off-diagonal entries in each row sum to less than n. This allows us to use results that guarantee an n-independent set in a multigraph to find the diagonally dominant principal submatrix of M that we require. In particular we will use the following result of Caro and Tuza (1991).

Theorem 10.12 (Caro and Tuza (1991)). *Let n be a positive integer and let G be a multigraph. There is an n-independent set in G of size at least $\lceil \sum_{u \in V(G)} f_n(\deg_G(u)) \rceil$ where*

$$f_n(x) = \begin{cases} 1 - \frac{x}{2n}, & \text{if } x \leqslant n; \\ \frac{n+1}{2(x+1)}, & \text{if } x \geqslant n. \end{cases}$$

We next prove a technical lemma that enables us to deduce bounds of a specific form that we denote by $CB_{(v,k,\lambda;s)}(\alpha, \beta)$. We will state the bounds in this section and the next in terms of this notation. Observe that the bound of Theorem 10.6 is $CB_{(v,k,\lambda;s)}(1, 0)$.

Lemma 10.13. *Let s and b_s be positive integers and let α and β be nonnegative real numbers such that $\alpha \geqslant 2\beta$. Suppose that any t-(v, k, λ) covering $(V, \mathcal{B})$ has $b(X) \geqslant b_s$ for each $X \in \binom{V}{s}$, and $|\mathcal{B}| \geqslant \alpha|\mathcal{V}_0| + \beta|\mathcal{V}_1|$ where $\mathcal{V}_i = \{X \in \binom{V}{s} : b(X) = b_s + i\}$ for $i \in \{0, 1\}$. Then*

$$C_\lambda(v, k, t) \geqslant \lceil CB_{(v,k,\lambda;s)}(\alpha, \beta) \rceil \quad \text{where} \quad CB_{(v,k,\lambda;s)}(\alpha, \beta) = \frac{b_s(\alpha - \beta)\binom{v}{s} + \alpha\binom{v}{s}}{(\alpha - \beta)\binom{k}{s} + 1}.$$

Proof. Let $(V, \mathcal{B})$ be a t-(v, k, λ) covering. Let $\mathcal{V} = \binom{V}{s}$, $x = |\mathcal{B}|\binom{k}{s} - b_s\binom{v}{s}$ and $v_i = |\mathcal{V}_i|$ for $i \in \{0, 1\}$. Note that $v_1 + 2\left(\binom{v}{s} - v_0 - v_1\right) \leqslant x$ because $b(X) = b_s + i$ for each $X \in \mathcal{V}_i$ for $i \in \{0, 1\}$, $b(X) \geqslant b_s + 2$ for each $X \in \mathcal{V} \setminus (\mathcal{V}_0 \cup \mathcal{V}_1)$, and $\sum_{X \in \mathcal{V}} b(X) = |\mathcal{B}|\binom{k}{s}$. It follows that $v_0 \geqslant \frac{1}{2}(2\binom{v}{s} - v_1 - x)$ and so from our hypotheses we have

$$|\mathcal{B}| \geqslant \tfrac{1}{2}\alpha\left(2\tbinom{v}{s} - v_1 - x\right) + \beta v_1 = \alpha\tbinom{v}{s} - \tfrac{1}{2}\alpha x - \tfrac{1}{2}(\alpha - 2\beta)v_1.$$

Thus, because $\alpha \geqslant 2\beta$, it follows from $v_1 \leqslant |\mathcal{V} \setminus \mathcal{V}_0| \leqslant x$ that

$$|\mathcal{B}| \geqslant \alpha\tbinom{v}{s} - \tfrac{1}{2}\alpha x - \tfrac{1}{2}(\alpha - 2\beta)x = \alpha\tbinom{v}{s} - (\alpha - \beta)x.$$

Since $x = |\mathcal{B}|\binom{k}{s} - b_s\binom{v}{s}$, we can deduce $|\mathcal{B}| \geqslant CB_{(v,k,\lambda;s)}(\alpha,\beta)$. $\square$

Remark 10.14. *A routine calculation shows that if $b_s + 1 > \beta\binom{k}{s}$, then the bound $\lceil CB_{(v,k,\lambda;s)}(\alpha,\beta)\rceil$ is inferior to the bound given by s iterated applications of* (10.2) *to $b_s + 1$.*

Theorem 10.15. *Suppose the hypotheses of Lemma 10.3 hold, that $b_s < \binom{k}{s}$, and that $d \geqslant a_s \geqslant 1$. Then*

$$C_\lambda(v,k,t) \geqslant \left\lceil CB_{(v,k,\lambda;s)}\left(\frac{a_s+1}{2(d+1)}, \frac{a_s+1}{2\left(d+\binom{k}{s}\right)}\right)\right\rceil.$$

Proof. Let $(V,\mathcal{B})$ be a t-(v,k,λ) covering. Let $\mathcal{V}_i = \{X \in \binom{V}{s} : b(X) = b_s + i\}$ for $i \in \{0,1\}$. Let G be the multigraph with vertex set $\binom{V}{s}$ such that $\mu_G(XY) = b(X \cup Y) - b_{|X\cup Y|}$ for each pair of distinct vertices X and Y.

By the definition of G, for a positive integer n, an n-independent set $\mathcal{S}$ in the multigraph G is a subset of $\binom{V}{s}$ with the property that, for all $X \in \mathcal{S}$,

$$\sum_{Y \in \mathcal{S}\setminus\{X\}} \left(b(X\cup Y) - b_{|X\cup Y|}\right) < n.$$

Consequently, if $n \leqslant a_s$, then $\mathcal{S}$ satisfies the hypotheses of Lemma 10.5 and $|\mathcal{B}| \geqslant |\mathcal{S}|$. So, by Lemma 10.13, it suffices to show that G has an a_s-independent set of size at least

$$\frac{a_s+1}{2d+2}|\mathcal{V}_0| + \frac{a_s+1}{2\left(d+\binom{k}{s}\right)}|\mathcal{V}_1|.$$

By Lemma 10.3(b), $\deg_G(X) = d$ for all $X \in \mathcal{V}_0$ and $\deg_G(X) = d + \binom{k}{s} - 1$ for all $X \in \mathcal{V}_1$. Thus, because $d \geqslant a_s$, G has an a_s-independent set of the required size by Theorem 10.12. $\square$

We only need consider the natural choice of b_s in Theorem 10.15. This follows by Remark 10.14 because

$$\frac{(a_s+1)\binom{k}{s}}{2\left(d+\binom{k}{s}\right)} < \frac{a_s+1}{2} < a_s + 1 < b_s + 1.$$

10.6 Improved bounds for the case $d < a_s$

In this section we will show that, by using techniques similar to those of the last section in the case $d < a_s$, we can sometimes improve on Theorem 10.6. We require a slight variant of Lemma 10.5.

Lemma 10.16. *Suppose the hypotheses of Lemma 10.3 hold and there exists a subset $\mathcal{S}$ of $\binom{V}{s}$ and positive real numbers $(c_X)_{X\in\mathcal{S}}$ such that, for each $X \in \mathcal{S}$,*

$$\sum_{Y\in\mathcal{S}\setminus\{X\}} c_Y \left(b(X\cup Y) - b_{|X\cup Y|}\right) < c_X \left(a_s + b(X) - b_s\right),$$

then $|\mathcal{B}| \geqslant |\mathcal{S}|$.

Proof. The proof of Lemma 10.5 applies, except that our hypotheses here imply via the Gershgorin circle theorem (see (Hogben, 2013, p.16-6)) that the matrix M'' rather than M' is positive definite, where M'' is obtained from M' by multiplying the entries in column X by c_X for each $X \in \mathcal{S}$. However, it is easy to see (using Sylvester's criterion (Hogben, 2013, p.9-7), for example) that M' is positive definite if and only if M'' is. $\square$

In Section 10.5 we employed multigraphs, but in this section we will work in a more general setting of edge-weighted graphs. An edge-weighted graph G is a complete (simple) graph in which each edge has been assigned a nonnegative real weight. We denote the weight of an edge uw in such a graph G by $\mathrm{wt}_G(uw)$ and we define the weight of a vertex u of G as $\mathrm{wt}_G(u) = \sum_{w\in V(G)\setminus\{u\}} \mathrm{wt}_G(uw)$. For $S \subseteq V(G)$, let $G[S]$ denote the edge-weighted subgraph of G induced by S. We generalise our notion of an n-independent set by saying, for a positive integer n, that a subset S of the vertices of an edge-weighted graph G is *n-independent* in G if $\mathrm{wt}_{G[S]}(u) < n$ for each $u \in S$.

We will require a technical result which guarantees the existence of an n-independent set of a certain size in an edge-weighted graph of a specific form. This result was effectively proved in Horsley (2017).

Lemma 10.17. *Let n, d and d' be nonnegative integers such that $d < n < d' - d$, and let G be a multigraph on some vertex set $\mathcal{V}_0 \cup \mathcal{V}_1$ such that $\deg_G(X) = d$ for $X \in \mathcal{V}_0$ and $\deg_G(X) = d'$ for $X \in \mathcal{V}_1$. Let c be a real number such that $c > \frac{d}{n}$ and let G^* be the edge-weighted graph on vertex set $\mathcal{V}_0 \cup \mathcal{V}_1$ such that, for all distinct $X, Y \in \mathcal{V}_0 \cup \mathcal{V}_1$,*

$$\mathrm{wt}_{G^*}(XY) = \begin{cases} 0, & \text{if } X, Y \in \mathcal{V}_0; \\ \mu_G(XY), & \text{if } X, Y \in \mathcal{V}_1; \\ c\mu_G(XY), & \text{otherwise.} \end{cases}$$

Let α and β be real numbers such that one of the following holds.

(a) $(\alpha, \beta) = \left(1 - \frac{d^2}{2n(n+1)}, \frac{n+2}{2(d'+1)}\right)$.

(b) $(\alpha,\beta) = \left(1, 1 - \frac{dd'}{n(n+1)}\right)$, $d \geqslant \frac{n}{2}$ *and* $dd' < n(n+1)$.

(c) $(\alpha,\beta) = \left(1, \sqrt{\frac{d(n+2)}{(n+1)(n-d)}} - \frac{d(d'+1)}{2(n+1)(n-d)}\right)$, $d < \frac{n}{2}$, *and* $d(d'+1)^2 < 4(n+1)(n+2)(n-d)$.

Then $\alpha \geqslant 2\beta > 0$ *and, if* c *is sufficiently close to* $\frac{d}{n}$, G^* *has an* $(n+1)$*-independent set* $\mathcal{S}$ *such that* $\mathcal{V}_0 \subseteq \mathcal{S}$ *and* $|\mathcal{S}| \geqslant \alpha|\mathcal{V}_0| + \beta|\mathcal{V}_1|$.

Proof. When (a) holds we obviously have $\beta > 0$ and

$$\alpha - 2\beta = \frac{(d' - n - d - 1)(2n(n+1) - d^2) + (n-d)(2d(n+1) + d^2)}{2n(n+1)(d'+1)}$$

is nonnegative because $d' > n + d$ and $n > d$. When (b) holds we have $\beta > 0$ because $dd' < n(n+1)$ and

$$\alpha - 2\beta = \frac{2dd' - n(n+1)}{n(n+1)}$$

is nonnegative because $d' > n + d$ and $d \geqslant \frac{n}{2}$. When (c) holds we have $\beta > 0$ because $d(d'+1)^2 < 4(n+1)(n+2)(n-d)$ and $\frac{d(d'+1)}{(n+1)(n-d)} \geqslant \frac{d(n+2)}{(n+1)(n-d)}$ because $d' > n$. Thus, since $2\sqrt{x} - x \leqslant 1$ for each nonnegative real number x, we have $\alpha \geqslant 2\beta$.

In the course of the proof of (Horsley, 2017, Theorem 14), the remainder of this result is proved for the case $d' = d + k - 1$. It is a routine exercise to show that the proof given there applies here for any $d' > n + d$. $\square$

We can now establish our improvements on Theorem 10.6.

Theorem 10.18. *Suppose the hypotheses of Lemma 10.3 hold, that* $b_s < \binom{k}{s}$*, and that* $d < a_s$*. Let* $d' = d + \binom{k}{s} - 1$*. Then* $C_\lambda(v,k,t) \geqslant \lceil CB_{(v,k,\lambda;s)}(\alpha,\beta)\rceil$ *when one of the following holds.*

(a) $(\alpha,\beta) = \left(1 - \frac{d^2}{2a_s(a_s+1)}, \frac{a_s+2}{2(d'+1)}\right)$.

(b) $(\alpha,\beta) = \left(1, 1 - \frac{dd'}{a_s(a_s+1)}\right)$, $d \geqslant \frac{a_s}{2}$ *and* $dd' < a_s(a_s+1)$.

(c) $(\alpha,\beta) = \left(1, \sqrt{\frac{d(a_s+2)}{(a_s+1)(a_s-d)}} - \frac{d(d'+1)}{2(a_s+1)(a_s-d)}\right)$, $d < \frac{a_s}{2}$ *and* $d(d'+1)^2 < 4(a_s+1)(a_s+2)(a_s-d)$.

Proof. Let $(V,\mathcal{B})$ be a t-(v,k,λ) covering. Let $\mathcal{V}_i = \{X \in \binom{V}{s} : b(X) = b_s + i\}$ for $i \in \{0,1\}$. Let G be the multigraph with vertex set $\binom{V}{s}$ such that $\mu_G(XY) = b(X \cup Y) - b_{|X\cup Y|}$ for each pair of distinct vertices X and Y. Note that, by Lemma 10.3, $\deg_G(X) = d$ for each $X \in \mathcal{V}_0$ and $\deg_G(X) = d'$ for each $X \in \mathcal{V}_1$. Also, $d < a_s < d' - d$ because

$d' - d = \binom{k}{s} - 1$ and $a_s < b_s < \binom{k}{s}$. Thus, by Lemma 10.17, there is a real number $c > \frac{d}{a_s}$ such that the edge-weighted graph G^* obtained from $G[\mathcal{V}_0 \cup \mathcal{V}_1]$ as in Lemma 10.17 has an (a_s+1)-independent set $\mathcal{S}$ such that $\mathcal{V}_0 \subseteq \mathcal{S}$ and $|\mathcal{S}| \geqslant \alpha|\mathcal{V}_0| + \beta|\mathcal{V}_1|$. We show that we can apply Lemma 10.16 to $\mathcal{S}$ choosing $c_X = c$ for $X \in \mathcal{S} \cap \mathcal{V}_0$ and $c_X = 1$ for $X \in \mathcal{S} \cap \mathcal{V}_1$. By Lemma 10.13 this will suffice to complete the proof.

If $X \in \mathcal{S} \cap \mathcal{V}_0$, then $c_X = c$, $b(X) = b_s$, and

$$\sum_{Y \in \mathcal{S}\setminus\{X\}} c_Y \left(b(X \cup Y) - b_{|X\cup Y|}\right) \leqslant d < ca_s = c_X\left(a_s + b(X) - b_s\right)$$

where the first inequality follows from Lemma 10.3(b). If $X \in \mathcal{S} \cap \mathcal{V}_1$, then $c_X = 1$, $b(X) = b_s + 1$, and

$$\sum_{Y \in \mathcal{S}\setminus\{X\}} c_Y \left(b(X \cup Y) - b_{|X\cup Y|}\right) = \mathrm{wt}_{G^*[\mathcal{S}]}(X) < a_s + 1 = c_X\left(a_s + b(X) - b_s\right)$$

where the first equality follows from the definition of G^* and our choice of c_Y for $Y \in \mathcal{S}$ and the inequality follows from the fact that $\mathcal{S}$ is an $(a_s + 1)$-independent set in G^*. $\square$

Again, we only need consider the natural choice of b_s in Theorem 10.18. To establish this it suffices, by Remark 10.14 and the fact that $b_s > a_s$, to show that $a_s + 2 - \beta\binom{k}{s}$ is positive. When (a) holds this is the case because

$$\frac{(a_s+2)\binom{k}{s}}{2(d'+1)} \leqslant \frac{a_s+2}{2} < a_s + 2.$$

When (b) or (c) holds, $a_s + 2 - \beta\binom{k}{s}$ is a quadratic in $\binom{k}{s}$ (note that $d' = d + \binom{k}{s} - 1$) and we can compute its global minimum in terms of a_s and d. When (b) holds this minimum is equal to

$$\frac{1}{4da_s(a_s+1)}\Big((2d-a_s)(a_s^3+2a_s^2+ad+a_s) + d(2a_s^3-d^3+7a_s^2+4a_s) + d^2(2a_s^2+2d-1)\Big)$$

which is positive since $\frac{a_s}{2} \leqslant d < a$. When (c) holds this minimum is equal to

$$\frac{1}{8(a_s+1)(a_s-d)}\Big(4d\sqrt{d(a_s+1)(a_s+2)(a_s-d)} + (a_s-2d)(2a_s^2+6a_s+12) \\ + (2a_s^3-d^3+16d) + 2a_s(3a_s-2)\Big)$$

which is positive since $0 \leqslant d < \frac{a_s}{2}$.

There are situations in which each of the Theorem 10.18 bounds is superior to both of the others. In the special case when $d = 0$, Theorem 10.18(a) is the best of our bounds.

10.7 Improvements for small parameter sets

We conclude with some tables which detail small parameter sets for which the results in this chapter produce an improvement over the previously best known lower bound on $C(v,k,t)$. For $t=2$ similar tables appear in Horsley (2017), so we concentrate here on the case $t \geqslant 3$. Our methodology in producing these tables is as follows.

To determine whether we see an improvement for $C(v',k',t')$ we successively evaluate a "best known" bound $b_{(v,k,t)}$ for $C(v,k,t)$ for $(v,k,t)=(v'-t'+1,k'-t'+1,1),(v'-t'+2,k'-t'+2,2),\dots,(v',k',t')$. This "best known" bound incorporates the following.

- $C(v,k,1)=\lceil \frac{v}{k} \rceil$.
- $C(v,k,t) \geqslant \lceil \frac{v}{k} b_{(v-1,k-1,t-1)} \rceil$ by (10.2).
- The Mills and Mullin result stated in (10.3).
- Results for a fixed number of blocks from Mills (1979); Greig et al. (2006); Todorov (1985); Todorov and Tonchev (1982). These include results for $t=2$, for $t=3$, and for general t. (The $t \in \{2,3\}$ results are summarised in Gordon and Stinson (2007).)
- Theorems 2.1, 3.1 and 4.4 of Todorov (1989).
- The lower bound of de Caen (1983).
- The lower bounds listed for $t \leqslant 8$, $v \leqslant 99$, $k \leqslant 25$ at the La Jolla Covering Repository Gordon (n.d.).
- Theorems 10.6, 10.15 and 10.18 of this chapter, applied with $s \in \{1,\dots,\lfloor \frac{t}{2} \rfloor\}$ and with b_i chosen as $b_{(v-i,k-i,t-i)}$ for $i \in \{s,\dots,2s\}$ (note that these theorems with $s=1$ specialise to the results in Horsley (2017)).

If the bound provided for $C(v',k',t')$ by one of the theorems of this chapter (using a particular choice of s) strictly exceeds the bound provided by any of the other results, then we include v' in the appropriate location in the tables. If, moreover, the bound provided for $C(v',k',t')$ by Theorem 10.15 or Theorem 10.18 strictly exceeds the bound provided by Theorem 10.6, then the table entry is set in italic or bold font, respectively. All improvements for $k \leqslant 40$ when $t=3$, when $t \in \{4,5\}$ and when $t \in \{6,7,8\}$ are given in Tables 10.1, 10.2, and 10.3 respectively (recall from the discussion after Theorem 10.6 that we obtain no improvements for sufficiently large v). Of course the listed improvements will, via (10.2), imply many further improvements for higher values of t, but we do not include these subsequent improvements in our tables.

Table 10.1: v's with an improved lower bound on $C(v, k, t)$ when $t = 3$

k	$s = 1$
9	19
10	21,**22**
12	**26**
13	29
15	33,42,45
16	35,**36**,45,46,48,49
17	**33**,48,49,51,52,53
18	**35**,40,51,59
19	**37**,42,43,54,**55**,58,62
20	**39**,*44*,57,61,62,66
21	41,47,60,61,64,65,**66**,69
22	43,49,50,63,64,73,88,89
23	45,*51*,66,71,76,87,**88**,**89**,92,93,95,**96**,**97**
24	47,*53*,54,69,74,**75**,80,91,92,93,96,97,99,**101**
25	49,56,57,**72**,73,77,**78**,79,83,95,96,97,100,101
26	51,*58*,75,87,**100**,101,104,105,106
27	53,*60*,61,78,84,90,103,**104**,**105**,108,109,110,114,115
28	55,*62*,63,64,81,**82**,87,88,94,107,108,**109**,112,113,114,117,118,119
29	57,*64*,*65*,84,**85**,90,91,**92**,97,111,112,113,116,117,118,121,122,123,124
30	59,*67*,68,87,101,115,116,117,120,121,122,126,127,**128**
31	61,*69*,70,71,90,**91**,97,104,119,**120**,121,124,125,126,127,130,131,**132**,**133**
32	63,*71*,*72*,93,100,**101**,*107*,108,123,**124**,**125**,129,130,131,135,136,137,160,161
33	65,*73*,*74*,75,96,97,103,**104**,105,111,127,128,**129**,133,134,135,139,140,141, 158,**159**,**160**,**161**,165,166,168,169,**170**,**171**
34	67,*76*,77,78,99,100,*106*,*114*,115,131,132,**133**,137,138,139,143,144,145,**146**, 163,164,165,166,170,171,173,174,**175**,**176**,**177**
35	69,*78*,*79*,102,*109*,110,*117*,118,135,136,**137**,141,142,143,148,149,**150**,168,169, 170,171,175,176,179,**180**,**181**,**182**
36	71,*80*,*81*,82,**105**,113,114,122,139,140,141,145,146,147,**148**,152,153,**154**,155, 174,175,176,180,181,184,**186**,**187**
37	73,*82*,*83*,84,**85**,**108**,109,116,117,118,*124*,125,143,144,145,150,151,**152**,157, 158,159,178,179,**180**,**181**,183,185,186,187,189,**192**,**193**,195,196
38	75,*85*,*86*,**111**,*119*,*128*,129,147,**148**,149,154,155,**156**,161,162,163,183,184,185, **186**,189,190,**191**,192,**197**,198,201
39	77,*87*,*88*,89,114,*122*,123,132,151,**152**,153,158,159,**160**,165,166,167,**168**,188, 189,190,191,194,195,**196**,**197**,201,202,203,206
40	79,*89*,*90*,91,**92**,117,118,*125*,126,127,*134*,*135*,136,155,**156**,157,162,163,**164**, 170,**171**,**172**,193,194,195,196,199,200,**201**,**202**,205,206,207,208,209,212

Table 10.2: v's with an improved lower bound on $C(v,k,t)$ when $t=4,5$

	$t=4$	$t=5$	
k	$s=1$	$s=1$	$s=2$
9			17
11			**29**
14			**47**
15			42
16	33		55
17	**30**,35		**59**
18	32,37		**66**
19	34,39		70
20	**39**,41		
21	37,**41**,43		75,93
22	39,43,45,**46**	36	**79**,**98**
23	41,45,**48**,52		**87**,123
24	43,47,**50**		
25	**37**,45,49,**52**,59	41	113,**135**,141
26	51,**54**		**118**
27	*47*,48,53	44	**127**,147
28	50,55,64,66,**68**,70	46,52	132,**153**
29	43,52,57,61,**69**	54	
30	**54**,59,63,73,75,76	49,54,56	*138*,*147*,161,**192**
31	*54*,**56**,61,65,71,73,74,**80**	51,56	143,171,199,**206**
32	*56*,*57*,63,67,74,78,81	65	**148**,**177**,**206**,213
33	49,59,65,69,76,78,79,80,**81**,**85**,88	54,67	**158**
34	61,67,71,81,86	56,61,69	216
35	*61*,**63**,69,73,81,83,84,85,86,**90**,93	63,**66**,71	**227**,231,235,**259**
36	*63*,65,71,75,**76**,83,86,91,92,93,96,97	59,65,68,73	201
37	55,*65*,*66*,67,73,**78**,88,**89**,90,91	61,67,70,75	**207**,275
38	*68*,**75**,80,88,91,**93**,96,97,101,105	77	**218**,248,**283**
39	*68*,70,**77**,82,90,93,95,96,99,104	64,70,79	224,**255**,264,**287**,**299**
40	*70*,72,**79**,84,95,**96**,98,101,102,103,106,**107**,113,114	66,72,81	*230*,299,307

Table 10.3: v's with an improved lower bound on $C(v,k,t)$ when $t = 6, 7, 8$

	$t=6$			$t=7$			$t=8$		
k	$s=1$	$s=2$	$s=3$	$s=1$	$s=2$	$s=3$	$s=1$	$s=2$	$s=3$
9			**25**						
12		23							
16		**29**							
17		33,38			31				
18		43							
19						**75**			
20					**39**				
21	30	57							
22	33								
23		58			**51**,53				
24	36			33					
25		*58*				**125**			
26	**36**,39								
27		**63**,75							**97**
28		68,78			68	166			
29		**83**,94		40	68				
30		86			*57*				
31	43,49	*89*					41		
32		*82*			61,77	221			
33	52,55	**117**,127		45	65,82,85			63	
34	47				86,92		45		
35	55,58	91,143							
36	50,57,60	124			**102**,**105**,107			71	
37		**122**,139,156					49		170,181
38	60,63	*108*,**122**,136,**143**		52	102		**52**		
39	54,65	*111*		**58**					
40	61,63	114,179					53	96	

Chapter 11

Pseudo Generalized Youden Designs

This chapter is based on the following work:

Das et al. (2018): Das, Ashish; Horsley, Daniel; Singh, Rakhi. Pseudo Generalized Youden Designs. *J. Combin. Des 26* (2018), no. 9, 439–454

11.1 Introduction

This chapter deals with designs on some set of *treatments*. Unless we specify otherwise we will always take this set of treatments to be the set $\{1, \ldots, v\}$ where v is a positive integer called the *order* of the design.

A *balanced block design* (BBD) with order v, block size k and index λ consists of a multiset $\mathcal{B}$ of *blocks* such that

- each block is a multiset of k treatments from $\{1, \ldots, v\}$;
- each treatment appears in blocks an equal number of times in total, and each treatment appears in each block $\lfloor k/v \rfloor$ times or $\lfloor k/v \rfloor + 1$ times;
- each pair of distinct treatments is covered a total of exactly λ times by blocks, where the number of times a block B covers a pair $\{i, j\}$ is given by the product of the number of times i appears in B and the number of times j appears in B.

A balanced incomplete block design (BIBD) is simply a BBD whose block size is less than its order. The *incidence matrix* of a BBD with b blocks is a $v \times b$ matrix whose (i, ℓ) entry is the number of times that treatment i appears in the ℓth block of the design. A BBD can be equivalently defined by demanding that the matrix NN^T be completely symmetric, where N is the incidence matrix of the design and *completely symmetric* means that all of the diagonal entries are equal and all of the off-diagonal entries are equal. It is easily seen that for any BBD with block size $k > v$ and incidence matrix N there is a BIBD with incidence matrix $N - \lfloor k/v \rfloor J$, where J is an all-ones matrix, and that the converse also holds.

In this chapter we are principally concerned with *row-column designs* which are rectangular arrays, each cell of which contains a treatment from $\{1, \ldots, v\}$. Three varieties of row-column designs are of particular significance here.

Youden square designs. Also known as Youden rectangles, these are classical objects in design theory. A Youden rectangle can be defined as a $k \times v$ row-column design such that each treatment appears once in each row of the design and the columns of the design form the blocks of a BIBD.

Generalized Youden designs (GYDs). These were introduced by Kiefer (1958) (although he originally called them generalized Youden squares). A GYD is a $k \times b$ row-column design such that the rows of the design form the blocks of a BBD and, separately, the columns of the design do likewise. Results on the existence and construction of GYDs can be found in Kiefer (1975*b*), Ruiz and Seiden (1974), Seiden and Wu (1978), Ash (1981) and Kunert and Sailer (2007). We give an example of a GYD with $v = 6, k = 10, b = 15$ obtained in Ash (1981).

GYD for $v = 6, k = 10, b = 15$

1	2	3	4	5	6	1	2	3	4	5	6	3	4	2
2	3	4	5	6	1	2	3	4	5	6	1	2	3	1
3	4	5	6	1	2	3	4	5	6	1	2	4	1	5
4	5	6	1	2	3	4	5	6	1	2	3	1	6	4
5	6	1	2	3	4	5	6	1	2	3	4	6	5	3
6	1	2	3	4	5	6	1	2	3	4	5	5	2	6
3	5	1	6	5	4	3	2	1	6	4	2	1	5	3
6	3	2	1	2	5	1	4	5	4	6	3	5	4	2
2	1	5	4	6	1	4	3	6	3	2	5	6	2	1
1	2	4	2	1	3	6	5	4	5	3	6	3	6	4

Pseudo Youden designs (PYDs). These were introduced by Cheng (1981*b*). A PYD is a $k \times k$ row-column design such that the rows and columns of the design, taken together as blocks, form a BBD. The existence and properties of PYDs have been investigated in Cheng (1981*b*), Cheng (1981*a*), Ash (1981), McSorley and Phillips

(2007) and Nilson (2011). We give an example of a PYD with $v = 9, k = 6$ provided in Cheng (1981*b*). This PYD can also be obtained through Theorem 11.9 of the current chapter.

PYD for $v = 9, k = 6$

4	7	8	6	9	5
3	1	2	8	7	9
2	5	1	3	6	4
9	3	6	2	5	8
7	6	9	4	1	3
5	8	4	7	2	1

These three kinds of row-column designs share the properties that they are optimal in various statistically-desirable senses under the most common set of assumptions for evaluating experimental designs. More formally, under the usual additive and homoscedastic fixed effects two-way heterogeneity model, they are A- and E-optimal and, when $v \neq 4$, they are also D-optimal. (The optimality is among all $k \times b$ row-column designs on v treatments that allow estimation of all treatment contrasts.) We refer the reader to Shah and Sinha (1989) for an introduction to these concepts along with the appropriate definitions. Kiefer established this optimality for GYDs in Kiefer (1975*b*), and Cheng (1981*b*) observed that his proof can be generalized to the case of PYDs. It is crucial to this proof that a particular matrix associated with the relevant design, called the *information matrix* or *C-matrix* of the design, is completely symmetric. The information matrix for a row-column design is given by

$$C = R - b^{-1}MM^T - k^{-1}NN^T + (kb)^{-1}rr', \tag{11.1}$$

where

- $R = \mathrm{diag}(r_1, r_2, \ldots, r_v)$ and $r = (r_1, r_2, \ldots, r_v)'$ where r_i is the number of times that treatment i occurs in the design;
- M is the $v \times k$ *treatment-row incidence matrix* of the design whose (i, ℓ) entry is the number of times treatment i appears in ℓth row of the design;
- N is the $v \times b$ *treatment-column incidence matrix* of the design whose (i, ℓ) entry is the number of times treatment i appears in ℓth column of the design.

Note that the different coefficients of MM^T and NN^T in (11.1) mean that if we allowed a $k \times b$ design for $b \neq k$ in the definition of PYD, then the information matrix would no longer necessarily be completely symmetric and the design would not be guaranteed to be optimal.

Here we introduce pseudo generalized Youden designs, which generalize both GYDs and PYDs. A *pseudo generalized Youden design* (PGYD) is a $k \times b$ row-column design such that

(A1) each treatment appears exactly kb/v times in total;

(A2) each treatment appears $\lfloor b/v \rfloor$ or $\lfloor b/v \rfloor + 1$ times in each row, and $\lfloor k/v \rfloor$ or $\lfloor k/v \rfloor + 1$ times in each column;

(A3) $kMM^T + bNN^T$ is completely symmetric, where M is the treatment-row incidence matrix and N is the treatment-column incidence matrix.

In statistical terminology, a PGYD is a row-column design satisfying (A1) and (A2) and where the k rows and b columns, considered together as blocks, form a variance balanced design. For an introduction to variance-balanced block designs with different block sizes, one can refer to Hedayat and Stufken (1989) and references therein.

In view of our comments on the incidence matrix of BBDs it can be seen that every GYD is a PGYD and every PYD is a PGYD. We will show, however, that there are parameter sets (v, k, b) for which a PGYD exists, but neither a GYD nor a PYD does. For some examples of PGYDs, which are non-GYD and non-PYD, one can refer to appendix F. Using the techniques of Kiefer (1975b), it can be seen that PGYDs share the optimality properties of GYDs and PYDs that we discussed. This fact means that, from the perspective of experimental design, our definition of PGYDs is more natural and useful than simply allowing $k \neq b$ in the PYD definition.

In Section 11.2, we obtain necessary conditions, in terms of v, k and b, for the existence of a PGYD. In Section 11.3, we construct families of PGYDs using patchwork methods based on affine planes. Using our necessary conditions, we also provide an exhaustive list of admissible parameter sets satisfying $v \leqslant 25, k \leqslant 50, b \leqslant 50$. For each, we establish that a PGYD exists, except for one where we demonstrate non-existence.

11.2 Necessary conditions for existence of PGYDs

We first show that condition (A3) of the PGYD definition can be rephrased more combinatorially, in terms of the blocks consisting of the treatments that appear more often in a particular row or column.

Definition 11.1. *Let $\mathcal{D}$ be a $k \times b$ row-column design on $v \geqslant 2$ treatments that obeys (A1) and (A2). In what follows, let $k = k'v + k''$ and $b = b'v + b''$ where $k' = \lfloor k/v \rfloor$, $b' = \lfloor b/v \rfloor$ and k'' and b'' are non-negative integers. For $u \in \{1, \dots, k\}$ let $\mathcal{R}_u$ be the set of treatments that occur $b' + 1$ times in row u and for $w \in \{1, \dots, b\}$ let $\mathcal{C}_w$ be the set of treatments that occur $k' + 1$ times in column w. For any two treatments i, j in $\{1, \dots, v\}$, we define*

$$\delta_{ij} = |\{u : \{i, j\} \subseteq \mathcal{R}_u\}|; \text{ and}$$
$$\lambda_{ij} = |\{w : \{i, j\} \subseteq \mathcal{C}_w\}|.$$

Let the collection $\{\mathcal{R}_1, \dots, \mathcal{R}_k\}$ be denoted by $\mathcal{D}_R$ and the collection $\{\mathcal{C}_1, \dots, \mathcal{C}_b\}$ be denoted by $\mathcal{D}_C$.

Theorem 11.2. *Let $\mathcal{D}$ be a $k \times b$ row-column design on $v \geqslant 2$ treatments that obeys (A1) and (A2). Then $\mathcal{D}$ is a PGYD if and only if $k\delta_{ij} + b\lambda_{ij}$ is identical for any two distinct treatments i and j.*

Proof. Let M and N be the treatment-row and treatment-column incidence matrices of $\mathcal{D}$, respectively. We first consider the diagonal entries of $kMM^T + bNN^T$. Because each treatment occurs $b' + 1$ times in exactly $r - kb'$ rows and exactly b' times in the rest, it can be seen that the diagonal elements of MM^T are all equal. Similarly, each treatment occurs $k' + 1$ times in exactly $r - bk'$ columns and exactly k' times in the rest, and the diagonal elements of NN^T are all equal. Thus the diagonal elements of $kMM^T + bNN^T$ are all equal.

We now consider the off-diagonal entries of $kMM^T + bNN^T$. Let i and j be distinct treatments and let $\nu_z = |\{u : |\{i, j\} \cap \mathcal{R}_u| = z\}|$ for $z \in \{0, 1, 2\}$. The (i, j) entry in MM^T is

$$(b')^2\nu_0 + b'(b' + 1)\nu_1 + (b' + 1)^2\nu_2 = k(b')^2 + b'\nu_1 + (2b' + 1)\nu_2$$

where the equality follows because $\nu_0 = k - \nu_1 - \nu_2$. Because there are exactly $r - kb'$ rows in which i occurs $b' + 1$ times and $r - kb'$ rows in which j occurs $b' + 1$ times,

$\nu_1 = 2(r - kb' - \nu_2)$. Also, $\nu_2 = \delta_{ij}$. Thus, the (i, j) entry in MM^T is

$$2b'r - k(b')^2 + \delta_{ij}.$$

Similarly, it can be established that the (i, j) entry in NN^T is

$$2k'r - b(k')^2 + \lambda_{ij}.$$

Thus it can be seen that the off-diagonal elements of $kMM^T + bNN^T$ are all equal if and only if $k\delta_{ij} + b\lambda_{ij}$ is identical for any two distinct treatments i and j. $\square$

We can view Theorem 11.2 in terms of edge decompositions of complete multigraphs. The condition of Theorem 11.2 is equivalent to requiring that the collection consisting of k-fold complete multigraphs on vertex sets $\mathcal{D}_R$ and of b-fold complete multigraphs on vertex sets $\mathcal{D}_C$ forms a decomposition of an x-fold complete multigraph on vertex set $\{1, \ldots, v\}$ for some positive integer x.

Our next result provides necessary conditions for the existence of a PGYD.

Theorem 11.3. *If there exists a $k \times b$ PGYD on $v \geqslant 2$ treatments, then the following hold.*

(1) $k + b \geqslant v$.

(2) $\frac{k(r-kb')(b''-1)+b(r-bk')(k''-1)}{v-1} = t$ *is an integer.*

(3) *There exist $p \geqslant 1$ pairs of non-negative integers $(m_1, n_1), \ldots, (m_p, n_p)$ such that, for $\ell = 1, \ldots, p$,*

   (i) $km_\ell + bn_\ell = t$

   (ii) $2r - 2kb' - k \leqslant m_\ell \leqslant r - kb'$ *and* $2r - 2bk' - b \leqslant n_\ell \leqslant r - bk'$.

(4) *There exist non-negative integers $z_1, \ldots, z_p$ such that*

   (i) $\sum_{\ell=1}^{p} z_\ell = \binom{v}{2}$,

   (ii) $\sum_{\ell=1}^{p} z_\ell m_\ell = k\binom{b''}{2}$,

   (iii) $\sum_{\ell=1}^{p} z_\ell n_\ell = b\binom{k''}{2}$.

Proof. Suppose there exists $k \times b$ PGYD $\mathcal{D}$ on $v \geqslant 2$ treatments. We provide the proofs for each of the conditions (1) – (4) below.

Condition (1): Let M and N be the treatment-row and treatment-column incidence matrices of $\mathcal{D}$, respectively. Elementary linear algebra establishes that rank$([M \vdots N]) \geqslant$ rank$(kMM^T + bNN^T)$. Since $[M \vdots N]$ is $v \times (k+b)$, it has rank at most $k+b$. Following Dey (1975), since $kMM^T + bNN^T$ is $v \times v$ and completely symmetric, it has rank v. Condition (1) follows. This condition is in fact a necessary condition for the existence of the corresponding variance balanced design with block sizes k and b.

Condition (2): This follows from the requirement that $k\delta_{ij} + b\lambda_{ij}$ in Theorem 11.2 is an integer. Let $t = k\delta_{ij} + b\lambda_{ij}$. To find its value, we note that the total number of pairs of treatments in blocks of $\mathcal{D}_R$ and $\mathcal{D}_C$ are respectively,

$$\sum_{i<j} \delta_{ij} = k\binom{b''}{2} \text{ and } \sum_{i<j} \lambda_{ij} = b\binom{k''}{2}. \tag{11.2}$$

This is so because in $\mathcal{D}_R$ there are k blocks and each block is of size b'', and in $\mathcal{D}_C$ there are b blocks and each block is of size k''. Therefore, summing over all $\binom{v}{2}$ treatment pairs, we get $\sum_{i<j}(k\delta_{ij} + b\lambda_{ij}) = \sum_{i<j} t$, which using (11.2) gives

$$t = \frac{k^2\binom{b''}{2} + b^2\binom{k''}{2}}{\binom{v}{2}} = \frac{k(r-kb')(b''-1) + b(r-bk')(k''-1)}{v-1}.$$

Condition (3): Index the distinct pairs in $\{(\delta_{ij}, \lambda_{ij}) : 1 \leqslant i < j \leqslant v\}$ as $(m_1, n_1), \ldots, (m_p, n_p)$. Using Condition (2) above, $km_\ell + bn_\ell = t$ for $\ell = 1, \ldots, p$. From the proof of Theorem 11.2, since $\nu_2 = \delta_{ij}$, $\nu_1 = 2(r - kb' - \nu_2) = 2(r - kb' - \delta_{ij})$ and $\nu_0 = k - \nu_1 - \nu_2 = k - 2r + 2kb' + \delta_{ij}$ must be non-negative, we have $2r - 2kb' - k \leqslant m_\ell \leqslant r - kb'$ for $\ell = 1, \ldots, p$. Similarly, we have $2r - 2bk' - b \leqslant n_\ell \leqslant r - bk'$ for $\ell = 1, \ldots, p$.

Condition (4): Let $z_\ell = |\{(i,j) : (\delta_{ij}, \lambda_{ij}) = (m_\ell, n_\ell), 1 \leqslant i < j \leqslant v\}|$, $\ell = 1, \ldots, p$. It is clear that $\sum_{\ell=1}^p z_\ell = \binom{v}{2}$. Also, from (11.2),

$$\sum_{\ell=1}^{p} z_\ell m_\ell = k\binom{b''}{2} \text{ and } \sum_{\ell=1}^{p} z_\ell n_\ell = b\binom{k''}{2}.$$

$\square$

A $k \times b$ row-column design with v treatments is called *regular* if $k \equiv 0 \pmod v$ or $b \equiv 0 \pmod v$; otherwise it is said to be non-regular. Accordingly, a PGYD with parameters $v, k = k'v + k'', b = b'v + b''$ is regular exactly when $k'' = 0$ or $b'' = 0$. A regular PGYD reduces to a regular GYD, the existence of which depends solely on the existence of a corresponding BIBD Agrawal (1966). Thus, we restrict ourselves to non-regular PGYDs for which v divides neither k nor b (that is, $k'' \neq 0$ and $b'' \neq 0$). Also, without loss of generality, henceforth assume $k \leqslant b$. In view of the proof of Theorem 11.3 we can give additional necessary conditions for the existence of a PGYD that is not a GYD.

Corollary 11.4. *Necessary conditions for the existence of a $k \times b$ non-GYD PGYD with v treatments, in addition to necessary conditions* (1) *and* (2) *in Theorem 11.3 are,*

(3′) $p \geq 2$*, in the condition* (3)*, and*

(4′) *at least two of the z_ℓ's are non-zero, in the condition* (4)*.*

Theorem 11.3 also specialises to give well-known necessary conditions for the existence of a GYD.

Corollary 11.5. *Necessary conditions for the existence of a $k \times b$ non-regular GYD with v treatments are,*

(1′) $k \geqslant v$ *and* $b \geqslant v$*, and*

(2′) $k\binom{b''}{2}/\binom{v}{2}$ *and* $b\binom{k''}{2}/\binom{v}{2}$ *are integers.*

The rows of a GYD form a BBD with k blocks and the columns form a BBD with b blocks. Thus, (1′) follows from Dey's generalization of Fisher's inequality Dey (1975). Also, (2′) follows directly from the condition (4) of Theorem 11.3 since for a GYD exactly one of the z_ℓ's should be non-zero.

Remark 11.6. *In addition to the necessary conditions for a non-regular GYD as given in Corollary 11.5, additional parametric conditions for the existence of the corresponding BIBDs, as given in Theorem 10.3.1 and Thoerem 16.1.3 of Hall (1998), must also hold.*

11.3 Construction of PGYDs

The constructions presented in this section are patchwork methods which go back to Kiefer (1975*a*). These constructions rely heavily on affine planes. For our purposes an affine plane of order q is a BIBD with q^2 treatments and $q(q+1)$ blocks of size q, where any two treatments appear together in exactly one block. The blocks of such a design can be partitioned into $q+1$ parallel classes each containing q blocks such that any two blocks from the same parallel class are disjoint and any two blocks from different parallel classes intersect in exactly one point. We will use this property frequently. An affine plane of order q is known to exist whenever q is a prime power. We will also sometimes consider complements of affine planes. For a block B of an affine plane on treatment set V, let $B^c = V \setminus B$ and for a parallel class $\mathcal{P}$ of such a plane, let $\mathcal{P}^c = \{B^c : B \in \mathcal{P}\}$.

Lemma 11.7. *Let m, n and v be positive integers with $n \equiv 0 \pmod{v}$, let $\{1, \ldots, v\}$ be a set of v treatments, and let $S_1, \ldots, S_n$ be m-subsets of $\{1, \ldots, v\}$. If every treatment occurs exactly mn/v times in the collection $\{S_1, \ldots, S_n\}$, then there is an $m \times n$ matrix A such that the set of treatments in the wth column of A is S_w and each treatment occurs n/v times in each row of A.*

Proof. Let G be the bipartite graph with parts $\{c_1, \ldots, c_n\}$ and $\{1, \ldots, v\}$ such that the set of vertices adjacent to c_w is S_w for $w \in \{1, \ldots, n\}$. Then $\deg_G(c_w) = m$ for $w \in \{1, \ldots, n\}$ and, by our hypothesis, $\deg_G(i) = mn/v$ for each $i \in \{1, \ldots, v\}$. By a result of de Werra (1971) the edges of G can be colored with m colours, say $1, \ldots, m$, such that each vertex in $\{c_1, \ldots, c_n\}$ is incident with exactly one edge of each color, and each vertex in $\{1, \ldots, v\}$ is incident with exactly n/v edges of each color.

Form A by placing in the (u, w) position the unique element i of $\{1, \ldots, v\}$ such that the edge $c_w i$ of G is assigned color u. That the set of treatments in the wth column of A is S_w follows from the definition of G. That each treatment occurs n/v times in each row of A follows from the fact that each vertex in $\{1, \ldots, v\}$ is incident with exactly n/v edges of each color.

This lemma can also be proved using the literature on systems of distinct representatives (see (Ford and Fulkerson, 1958, Theorem 1), for example). □

Lemma 11.8. *Let $\mathcal{P}_1, \ldots, \mathcal{P}_{q-1}$ and $\mathcal{Q}_1, \ldots, \mathcal{Q}_{q-1}$ be parallel classes (not necessarily distinct) of an affine plane of order q such that $\mathcal{P}_x \neq \mathcal{Q}_y$ for $x, y \in \{1, \ldots, q-1\}$.*

(i) For any $x, y \in \{1, \ldots, q-1\}$ there is a $q \times q$ matrix A such that the sets of treatments in the rows of A are the elements of $\mathcal{P}_x$ and the sets of treatments in the columns of A are the elements of $\mathcal{Q}_y$.

(ii) *For any* $x \in \{1, \ldots, q-1\}$ *there is a* $q \times (q^2 - q)$ *matrix* A *such that the sets of treatments in the rows of* A *are the elements of* $\mathcal{P}_x^c$ *and the sets of treatments in the columns of* A *are the elements of* $\mathcal{Q}_1, \ldots, \mathcal{Q}_{q-1}$.

(iii) *There is a* $(q^2 - q) \times (q^2 - q)$ *matrix* A *such that the sets of treatments in the rows of* A *are the elements of* $\mathcal{P}_1^c, \ldots, \mathcal{P}_{q-1}^c$ *and the sets of treatments in the columns of* A *are the elements of* $\mathcal{Q}_1^c, \ldots, \mathcal{Q}_{q-1}^c$.

Proof. For $x, y \in \{1, \ldots, q-1\}$, let $\mathcal{P}_x = \{P_{x,1}, \ldots, P_{x,q}\}$ and let $\mathcal{Q}_y = \{Q_{y,1}, \ldots, Q_{y,q}\}$.

Case (i): We will show that there exists a $q \times q$ matrix A such that the set of treatments in the uth row of A is $P_{x,u}$ and the set of treatments in the wth column of A is $Q_{y,w}$. Because $x \neq y$, $|P_{x,u} \cap Q_{y,w}| = 1$ for all $u, w \in \{1, \ldots, q\}$. So A can be obtained by placing the unique element of $P_{x,u} \cap Q_{y,w}$ in the (u, w) position.

Case (ii): As in the proof of (i) there is, for each $y \in \{1, \ldots, q-1\}$, a $q \times q$ matrix A_y such that the set of treatments in the uth row of A_y is $P_{x,u+y}$ and the set of treatments in the wth column of A_y is $Q_{y,w}$ (where the subscripts are considered modulo q). We take

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{q-1} \end{bmatrix}.$$

The set of treatments in the uth row of A is $P_{x,u}^c$.

Case (iii): As in the proof of (i) there is, for each $x, y \in \{1, \ldots, q-1\}$, a $q \times q$ matrix $A_{x,y}$ such that the set of treatments in the uth row of $A_{x,y}$ is $P_{x,u+y}$ and the set of treatments in the wth column of $A_{x,y}$ is $Q_{y,w+x}$ (where the subscripts are considered modulo q). We take

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,q-1} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,q-1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{q-1,1} & A_{q-1,2} & \cdots & A_{q-1.q-1} \end{bmatrix}.$$

The set of treatments in the uth row of A is $P_{x,u'}^c$ where $u = (x-1)q + u'$ and $u' \in \{1, \ldots, q\}$. Similarly the set of treatments in the wth column of A is $Q_{y,w'}^c$ where $w = (y-1)q + w'$ and $w' \in \{1, \ldots, q\}$. □

The following theorem gives us four families of PGYDs based on the residues of k and b modulo q^2.

Theorem 11.9. *Let q be a prime power. There exists a $k \times b$ PGYD with $v = q^2$ treatments if*

(i) $k \equiv \pm q \pmod{q^2}$*;*

(ii) $b \equiv \pm q \pmod{q^2}$*;*

(iii) $k = k^* q(q+1) + \frac{b}{\gcd(b,k)} q(q+1-n)$ *and* $b = b^* q(q+1) + \frac{k}{\gcd(b,k)} nq$ *for some* $n \in \{0, \ldots, q\}$ *and non-negative integers* k^* *and* b^**.*

Proof. From (i) and (ii) we have $k = k'q^2 + k''$ and $b = b'q^2 + b''$ where $k'', b'' \in \{q, q^2 - q\}$ and k' and b' are non-negative integers. Let $\mathcal{P}_1, \ldots, \mathcal{P}_{q+1}$ be the parallel classes of an affine plane of order q. Let $g = \gcd(b, k)$.

Let $x_1, \ldots, x_{k/q}$ be the unique non-decreasing sequence of indices from $\{1, \ldots, q+1\}$ such that each index in $\{1, 2, \ldots, n\}$ occurs k^* times in the sequence and each index in $\{n+1, \ldots, q+1\}$ occurs $k^* + b/g$ times in the sequence. For $u \in \{1, \ldots, k/q\}$, let $\mathcal{R}_u = \mathcal{P}_{x_u}$ if $b'' = q$ and $\mathcal{R}_u = \mathcal{P}^c_{x_u}$ if $b'' = q^2 - q$. Let $y_1, \ldots, y_{b/q}$ be the unique non-increasing sequence of indices from $\{1, \ldots, q+1\}$ such that each index in $\{1, 2, \ldots, n\}$ occurs $b^* + k/g$ times in the sequence and each index in $\{n+1, \ldots, q+1\}$ occurs b^* times in the sequence. For $w \in \{1, \ldots, b/q\}$, let $\mathcal{C}_w = \mathcal{P}_{y_w}$ if $k'' = q$ and $\mathcal{C}_w = \mathcal{P}^c_{y_w}$ if $k'' = q^2 - q$.

We will form the required design as

$$\begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$$

where

- W is a $(k - k'') \times (b - b'')$ matrix such that each treatment occurs the same number of times in each row of W and the same number of times in each column of W.
- X is a $(k - k'') \times b''$ matrix such that the sets of treatments in the rows of X are the elements of $\mathcal{R}_1, \ldots, \mathcal{R}_{(k-k'')/q}$ and each treatment occurs k' times in each column of X.
- Y is a $k'' \times (b - b'')$ matrix such that the sets of treatments in the columns of Y are the elements of $\mathcal{C}_1, \ldots, \mathcal{C}_{(b-b'')/q}$ and each treatment occurs b' times in each row of Y.

- Z is a $k'' \times b''$ matrix such that the sets of treatments in the rows of Z are the elements of $\mathcal{R}_{(k-k'')/q+1}, \dots, \mathcal{R}_{k/q}$ and the sets of treatments in the columns of Z are the elements of $\mathcal{C}_{(b-b'')/q+1}, \dots, \mathcal{C}_{b/q}$.

We will first show that such a design is a PGYD and then show that we can construct matrices W, X, Y and Z with the required properties.

It is clear that such a design obeys (A1) and (A2). So to show the design is a PGYD it suffices, by Theorem 11.2, to show that $k\delta_{ij} + b\lambda_{ij}$ is identical for each pair of distinct treatments (i, j). Let (i, j) be a pair of distinct treatments. Define

$$\gamma_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ occur together in a block in } \mathcal{P}_1 \cup \dots \cup \mathcal{P}_n; \\ 0, & \text{if } i \text{ and } j \text{ occur together in a block in } \mathcal{P}_{n+1} \cup \dots \cup \mathcal{P}_{q+1}. \end{cases}$$

Note that i and j occur together in $1 - \gamma_{ij}$ blocks in $\mathcal{P}_{n+1} \cup \dots \cup \mathcal{P}_{q+1}$. Note also that i and j occur together in $q-1$ blocks of $\mathcal{P}_\alpha^c$ if i and j occur together in a block of $\mathcal{P}_\alpha$ and i and j occur together in $q-2$ blocks of $\mathcal{P}_\alpha^c$ otherwise. Then from our construction we can calculate that λ_{ij} and δ_{ij} are as given below.

$$\lambda_{ij} = \begin{cases} b^* + \frac{k}{g}\gamma_{ij}, & \text{if } k'' = q; \\ b^*(q^2 - q - 1) + \frac{k}{g}(n(q-2) + \gamma_{ij}), & \text{if } k'' = q^2 - q. \end{cases}$$

$$\delta_{ij} = \begin{cases} k^* + \frac{b}{g}(1 - \gamma_{ij}), & \text{if } b'' = q; \\ k^*(q^2 - q - 1) + \frac{b}{g}((q+1-n)(q-2) + (1 - \gamma_{ij})), & \text{if } b'' = q^2 - q. \end{cases}$$

Considering four cases according to the values of k'' and b'', it is easy to check that the value of $k\delta_{ij} + b\lambda_{ij}$ is independent of γ_{ij}. Hence $k\delta_{ij} + b\lambda_{ij}$ is identical for each pair of distinct treatments (i, j) and the design is a PGYD.

We now show that we can construct matrices W, X, Y and Z with the required properties. It is easy to form W by tiling $q^2 \times q^2$ latin squares. Because each treatment appears once in each parallel class and appears $q-2$ times in the complement of each parallel class, and because $b - b'' \equiv 0 \pmod{q^2}$, Lemma 11.7 can be used to construct a matrix Y with the required properties. Similarly, by applying Lemma 11.7 and taking a transpose, a matrix X with the required properties can be constructed. Finally, Lemma 11.8 yields a matrix Z with the required properties provided that the sets $\{x_{(k-k'')/q+1}, \dots, x_{k/q}\}$ and $\{y_{(b-b'')/q+1}, \dots, y_{b/q}\}$ are disjoint. We complete the proof by establishing this claim.

When $k'' = b'' = q$, $\{y_{b/q}\} = \{1\}$ and $\{x_{k/q}\} = \{q+1\}$. When $b'' = q$ and $k'' = q^2 - q$, $\{y_{b/q}\} = \{1\}$ and $\{x_{(k-k'')/q+1}, \dots, x_{k/q}\} \subseteq \{3, \dots, q+1\}$. When $b'' = q^2 - q$ and $k'' = q$,

$\{x_{k/q}\} = \{q+1\}$ and $\{y_{(b-b'')/q+1}, \ldots, y_{b/q}\} \subseteq \{1, \ldots, q-1\}$. In each of these cases the claim is true, so we may assume that $b'' = k'' = q^2 - q$ and $q \neq 2$. We consider two cases according to whether $b = k$.

Suppose first that $k \neq b$. We are assuming $k \leqslant b$ without loss of generality, so $k < b$. Then

$$g = q\gcd(b'q + q - 1, k'q + q - 1) = q\gcd(b'q + q - 1, (b' - k')q) \leqslant q(b' - k') \leqslant b'q,$$

where the last equality follows because $(b'-k')q = (b'q+q-1)-(k'q+q-1)$ and the first inequality follows because $\gcd(b'q+q-1, q) = 1$. So we have $b/g \geqslant (b'q^2+q^2-q)/(b'q) > q$. Thus $\{x_{(k-k'')/q+1}, \ldots, x_{k/q}\} = \{q+1\}$. Obviously $\{y_{(b-b'')/q+1}, \ldots, y_{b/q}\} \subseteq \{1, \ldots, q-1\}$, and the claim follows.

Now suppose that $k = b$. Then it follows from (iii) that $k = b = b^*q(q+1)+nq$ where $n \in \{0, (q+1)/2\}$. So, because $b \equiv -q \pmod{q^2}$, $b^* \equiv q - n - 1 \pmod{q}$. Thus, it must be the case that $n = 0$ and $b^* \geqslant 2$ or that $n = (q+1)/2$ and $b^* \geqslant 1$ or that $n = (q+1)/2$, $b^* = 0$ and $q = 3$. In each of these cases it can be verified that $\{y_{(b-b'')/q+1}, \ldots, y_{b/q}\} \subseteq \{1, \ldots, \lfloor (q+1)/2 \rfloor\}$ and $\{x_{(k-k'')/q+1}, \ldots, x_{k/q}\} \subseteq \{\lceil (q+3)/2 \rceil, \ldots, q+1\}$. □

Theorem 11.9 produces a PYD when $k = b$. In this case it must be that $n = 0$ or $n = (q+1)/2$. Cheng's construction in Theorem 2.2 of Cheng (1981*b*) necessarily requires that $b \equiv q \pmod{q^2}$ and produces designs for parameter sets covered by Theorem 11.9. However, Theorem 11.9 also produces PYDs for parameter sets not covered by Theorem 2.2 of Cheng (1981*b*). In particular, it does so when $b \equiv -q \pmod{q^2}$, as stated in the following corollary.

Corollary 11.10. *Let q be a prime power and a be a positive integer. Then an $(aq^2 - q) \times (aq^2 - q)$ PYD with q^2 treatments exists*

(i) when q is odd and $a \equiv -1 \pmod{\frac{q+1}{2}}$; and

(ii) when q is even and $a \equiv -1 \pmod{q+1}$.

Remark 11.11. *Theorem 11.9 gives a GYD when $n = 0$ and Corollary 11.10 gives a GYD when $a \equiv -1 \pmod{q+1}$.*

It appears harder to analyze when Theorem 11.9 can be applied with $n \neq 0$ and $b \neq k$ so as to produce a PGYD which is not a GYD or PYD. Instead we present Table 11.1

which, for $q \leqslant 8$, lists the parameter sets of such designs obeying $k \leqslant b \leqslant v^2$ (there are no such parameter sets for $q \in \{2,3,8\}$). For $q = 9$ such a listing would be too lengthy and we instead list those obeying $k \leqslant b \leqslant v^2/3$. It appears we obtain a wider variety of parameter sets when q is a perfect square.

Table 11.1: Parameter sets of non-GYD non-PYD PGYDs given by Theorem 11.9 for $q \leqslant 8$ and $k \leqslant b \leqslant v^2$ and for $q = 9$ and $k \leqslant b \leqslant v^2/3$

$q=4$, $v=16$			$q=5$, $v=25$			$q=7$, $v=49$			$q=9$, $v=81$		
k	b	n	k	b	n	k	b	n	k	b	n
12	36	4	405	495	3	924	1428	4	234	1872	8
28	196	4				924	2100	4	396	2178	6
36	108	2				1428	2100	4	819	1287	9
44	132	3				1820	2100	4	819	1953	9
52	156	4							882	1386	2
68	204	1							1224	1368	6
76	228	2							1368	2016	6
84	132	4							1386	2178	6
84	252	3							1449	1953	9
108	204	4							1872	2016	8
156	228	4									
196	252	4									

We now present a simple method of obtaining a non-GYD PYD from a PYD that is a GYD and has a particular additional property.

Theorem 11.12. *Let $\mathcal{D}$ be a $k \times k$ GYD with v treatments, of the form*

$$\begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$$

where W is $vk' \times vk'$ and is formed by tiling latin squares of order v, Z is $k'' \times k''$, and each treatment occurs k' times in each column of X and k' times in each row of Y. If there is a pair of treatments that occur together in the columns of Z a different number of times from in the rows of Z, then the design $\mathcal{D}^*$ formed from $\mathcal{D}$ by replacing Z with Z^T is a PYD that is not a GYD.

Proof. Using the logic of Theorem 11.2, it follows that in the GYD $\mathcal{D}$ each pair of distinct treatments (i,j) appears μ times in the rows of $[X^T \vdots Z^T]^T$ and μ times in the columns of $[Y \vdots Z]$ for some positive integer μ.

Let (i,j) be a pair of distinct treatments. Say (i,j) appears δ_{ij}^Z times in the rows of Z and λ_{ij}^Z times in the columns of Z. So (i,j) appears $(\mu-\delta_{ij}^Z)$ times in the rows of X and $(\mu-\lambda_{ij}^Z)$ times in the columns of Y. Then $\mathcal{D}^*$ has the form

$$\begin{bmatrix} W & X \\ Y & Z^T \end{bmatrix}$$

and (i,j) appears $\mu-\delta_{ij}^Z+\lambda_{ij}^Z$ times in the rows of $[X^T \vdots Z]^T$ and $\mu-\lambda_{ij}^Z+\delta_{ij}^Z$ times in the columns of $[Y \vdots Z^T]$. So, in $\mathcal{D}^*$, $k\lambda_{ij}+k\delta_{ij}=k(\mu-\lambda_{ij}^Z+\delta_{ij}^Z+\mu-\delta_{ij}^Z+\lambda_{ij}^Z)=2k\mu$.

Thus, in $\mathcal{D}^*$, $k\lambda_{ij}+k\delta_{ij}$ is identical for any pair (i,j) of distinct treatments, and $\mathcal{D}^*$ is a PYD by Theorem 11.2. However, by our hypotheses, there is some pair (i,j) of distinct treatments that appears δ_{ij}^Z times in the rows of Z and λ_{ij}^Z times in the columns of Z where $\delta_{ij}^Z \neq \lambda_{ij}^Z$. So, using our arguments above, $\lambda_{ij} \neq \delta_{ij}$ in $\mathcal{D}^*$ because $\mu-\delta_{ij}^Z+\lambda_{ij}^Z \neq \mu-\lambda_{ij}^Z+\delta_{ij}^Z$. Therefore, $\mathcal{D}^*$ is not a GYD. □

Remark 11.13. *Any GYD with $k=b$ constructed according to the proof of Theorem 11.9 will satisfy the conditions of Theorem 11.12. To see this, note that in the proof of Theorem 11.9,* $1 \in \{y_{(k-k'')/q+1},\ldots,y_{k/q}\}$ *but* $1 \notin \{x_{(b-b'')/q+1},\ldots,x_{b/q}\}$. *It follows that any pair of treatments that appears in a block in $\mathcal{P}_1$ will appear more often in the columns of Z than in the rows of Z.*

Ash (1981) gave constructions of GYDs for all parameter sets satisfying $v \leqslant 25, k \leqslant b \leqslant 50$ and the conditions of Corollary 11.5, with two exceptions. For $(v,k,b)=(15,21,35)$ a GYD is known not to exist by Remark 11.6. For $(v,k,b)=(25,40,40)$ it is not known whether a GYD exists, but Ash provides a PYD. Consequently a PGYD exists trivially for all of Ash's parameter sets except (15,21,35), and in these cases we turn our attention to whether there exists a non-GYD PGYD. Table 11.2 lists these parameter sets together with whether a non-GYD PGYD is known to exist or not exist. The parameters in Table 11.2 where a non-GYD PGYD exists can be obtained from Theorem 11.9 and Theorem 11.12 (see Remark 11.13), except for $(8,14,28)$, $(8,28,28)$, $(9,24,48)$ and $(10,36,45)$. However, a non-GYD PGYD for $(8,28,28)$ can be obtained by applying Theorem 11.12 to the GYD for $(8,28,28)$ provided in Ash (1981) and a non-GYD PGYD for $(9,24,48)$ is provided in the appendix F. The special statuses of $(15,21,35)$ and $(25,40,40)$ are marked in Table 11.2 by an asterisk (*) and an exclamation (!), respectively.

There are seven parameter sets in the range $v \leqslant 25, k \leqslant b \leqslant 50$ that satisfy the conditions of Theorem 11.2 but not those of Corollary 11.5. For all of these a PGYD exists. These parameter sets are listed in Table 11.3 together with the condition of Corollary 11.5 they violate and references to constructions of corresponding PGYDs. Other than two non-GYD PGYDs for $(8, 20, 50)$ and $(18, 12, 48)$ provided in the appendix F, the constructions for the non-GYD PGYDs follow from Theorem 11.9.

Table 11.2: Existence of non-GYD PGYDs for parameter sets satisfying $v \leqslant 25, k \leqslant b \leqslant 50$ and the conditions of Corollary 11.5

v	k	b	Non-GYD PGYD?	v	k	b	Non-GYD PGYD?
4	6	6	Yes: Theorem 11.12	8	28	28	Yes: Theorem 11.12
4	6	18	No: Corollary 11.4 ($3'$)	8	28	42	No: Corollary 11.4 ($4'$)
4	6	30	No: Corollary 11.4 ($3'$)	9	12	12	Yes: Theorem 11.12
4	6	42	No: Corollary 11.4 ($3'$)	9	12	24	No: Corollary 11.4 ($4'$)
4	18	18	Yes: Theorem 11.12	9	12	48	No: Corollary 11.4 ($3'$)
4	18	30	No: Corollary 11.4 ($4'$)	9	24	24	Yes: Theorem 11.12
4	18	42	No: Corollary 11.4 ($3'$)	9	24	48	Yes: appendix F
4	30	30	Yes: Theorem 11.12	9	48	48	Yes: Theorem 11.12
4	30	42	No: Corollary 11.4 ($4'$)	10	15	36	No: Corollary 11.4 ($3'$)
4	42	42	Yes: Theorem 11.12	10	18	45	No: Corollary 11.4 ($3'$)
6	10	15	No: Corollary 11.4 ($3'$)	10	36	45	Unknown
6	10	45	No: Corollary 11.4 ($3'$)	12	33	44	No: Corollary 11.4 ($3'$)
6	15	20	No: Corollary 11.4 ($4'$)	15^*	21	35	No: Corollary 11.4 ($4'$)
6	15	40	No: Corollary 11.4 ($3'$)	15	35	42	No: Corollary 11.4 ($3'$)
6	15	50	No: Corollary 11.4 ($3'$)	16	20	20	Yes: Theorem 11.12
6	20	45	No: Corollary 11.4 ($3'$)	21	30	35	No: Corollary 11.4 ($4'$)
6	40	45	No: Corollary 11.4 ($3'$)	25	30	30	Yes: Theorem 11.12
6	45	50	No: Corollary 11.4 ($4'$)	$25^!$	40	40	Yes: Ash
8	14	28	Unknown				

Table 11.3: Existence of non-GYD PGYDs for parameter sets satisfying $v \leqslant 25, k \leqslant b \leqslant 50$ but failing the conditions of Corollary 11.5

v	k	b	GYD?	Non-GYD PGYD?
8	20	50	No: Corollary 11.5 ($2'$)	Yes: appendix F
9	6	6	No: Corollary 11.5 ($1'$)	Yes: Cheng, Theorem 11.9
9	30	30	No: Corollary 11.5 ($2'$)	Yes: Cheng, Theorem 11.9
9	42	42	No: Corollary 11.5 ($2'$)	Yes: Theorem 11.9
16	12	36	No: Corollary 11.5 ($1'$)	Yes: Theoerm 11.9
18	12	48	No: Corollary 11.5 ($1'$)	Yes: appendix F
25	45	45	No: Corollary 11.5 ($2'$)	Yes: Theorem 11.9

Chapter 12

Summary and future work

In this thesis, we have worked on solving important problems in the areas of optimal design theory, discrete choice experiments, supersaturated designs, coverings and Youden designs. In what follows, we provide concluding remarks and discuss possible future work in these areas.

12.1 Summary

12.1.1 Discrete Choice Experiments

Discrete choice experiments have gained importance over the last few years because of their use in studying people's preferences in a wide range of industries such as marketing, transportation economics, health economics, environmental economics and public economics. The responses in these studies are usually discrete or qualitative choices. Being a relatively new area of study from a statistical perspective, there are a lot of open problems in the area. Several authors have worked in this area and fantastic surveys are available by Street-Burgess and Großmann-Schwabe (Street and Burgess, 2007; Großmann and Schwabe, 2015). One of the biggest challenges is to keep the number of choice sets as small as possible while still being able to achieve the best possible results. For a major part of this thesis, we have worked on choice experiments. Throughout the thesis, we have worked on solving this problem of reducing the number of choice sets under various choice experiment setups including the estimation of either main-effects or main-effects plus two-factor interaction effects. We now give a chapter-wise summary for the work done in this thesis.

12.1.1.1 Chapter 2

For two-level paired choice experiments, we have obtained a simple form of the information matrix of a choice design for estimating the main effects, and provided D- and MS-optimal paired choice designs with distinct choice sets under the main effects model for any number of choice sets. The D- and MS-optimal two-level paired choice designs found in this chapter provide solutions in situations where, for every $N \not\equiv 0 \pmod 4$, the information matrix of an optimal exact design is different from the information matrix of the optimal approximate design, for which the corresponding exact optimal design was not previously available. This work complements previous work giving optimal exact designs only for $N \equiv 0 \pmod 4$. Thus experimenters can now use optimal designs for any number of choice sets N. It is also shown that the optimal designs under the main effects model are also optimal under the broader main effects model. From a statistical perspective we have established that one should prefer optimal paired choice designs to choice designs with $m = 3$ or $m = 5$. This also assists in achieving the desired quality of response through reduced choice set size.

12.1.1.2 Chapter 3

Traditionally, while using designs for discrete choice experiments, every respondent is shown the same collection of choice pairs (that is, the choice design). Also, as the attributes and/or the number of levels under each attribute increases, the number of choice pairs in an optimal paired choice design increases rapidly. Moreover, in the literature under the utility-neutral setup, random subsets of the theoretically obtained optimal designs are often allocated to respondents. The question therefore is whether one can do better than a random allocation of subsets. To address these concerns, in the linear paired comparison model (or, equivalently the multinomial logit model), we first incorporate the fixed respondent effects (also referred to as the block effects) and then obtain optimal designs for the parameters of interest. Our approach is simple and theoretically tractable, unlike other approaches which are algorithmic in nature. We present several constructions of optimal block designs for estimating main effects or main plus two-factor interaction effects. Our results show when and how an optimal design for the model without blocks can be split into blocks so as to retain the optimality properties under the block model.

12.1.1.3 Chapter 4

In this chapter, for paired choice designs, two new construction methods are also proposed for the estimation of the main effects. These designs require about 30-50% fewer choice pairs than the existing designs and at the same time have reasonably high D-efficiencies for the estimation of the main effects. Since, for $v = 2, 3$, the number of choice pairs involved is not very large, it may be preferable to use optimal designs for such cases. However, as the number of levels increases, the number of choice pairs in an optimal design increases rapidly, and thus, it is preferable to use efficient designs with fewer choice pairs. The significant gain through the reduced number of choice pairs compensates for the marginal loss in D-efficiency.

12.1.1.4 Chapter 5

In this chapter, for paired choice designs with all factors having 3 levels, we have obtained a sharper lower bound to the A- and D-values for estimating the main effects under the utility-neutral multinomial logit model in the cases where number of choice pairs N is not necessarily a multiple of 3. New A- and D-optimal (and efficient) designs are also provided. The D-optimal designs under effects coding are also A- and D-optimal under orthonormal contrasts. However, under effects coding, A-optimal designs are usually not D-optimal, even if N is a multiple of three; for example the design $a_{(2,6)}$. The example $a_{(2,6)}$ illustrates the need for more work to understand whether one should recommend A-optimal designs for orthonormal contrasts or one should recommend A-optimal designs under effects coding.

12.1.1.5 Chapter 6

Considering three-level paired choice designs for estimating all the main effects and two-factor interaction effects under the utility-neutral multinomial logit model, we have provided a general technique involving generators to reduce the number of choice pairs in a D-optimal design. Generators are identified allowing significant reduction in the total number of choice pairs for D-optimal designs. We have also given several examples of generators for the practical k's.

12.1.1.6 Chapter 7

For two-level choice experiments with k factors, we consider a model involving the main plus all two-factor interaction effects with our interest lying in the estimation of the main effects and a specified set of two-factor interaction effects. The two-factor interaction effects of interest are either (i) one factor interacting with each of the remaining $n-1$ factors or (ii) each of the two factors interacting with each of the remaining $n-2$ factors. There are no general results on the optimal choice designs for estimating main plus specified two-factor interaction effects in the choice design literature, though Street and Burgess (2007) highlighted the problem by giving a few examples. One could argue that the optimal designs available for estimating main effects and all two-factor interactions could be used for this specific problem because of a lack of theoretical results. However, when one increases the number of parameters of interest (especially 2-factor interactions), theoretically obtained optimal designs usually have a large number of choice sets. Under our model, we have provided theoretical results characterizing optimal designs for any m. However, we provide optimal design constructions for more practical values of m, i.e., $m = 3$ and $m = 4$. The case for $m = 2$ still remains a relevant open problem unless one uses large designs that are optimal for estimating main and all two-factor interactions as obtained by Street and Burgess (2007). As a way forward, one can possibly extend this work to factors with asymmetric levels. One could also consider other sets of specified two-factor interaction effects as indicated in Dey and Suen (2002).

12.1.1.7 Chapter 8

The author-groups Street–Burgess and Huber–Zwerina have adopted different approaches and used seemingly different information matrices under the multinomial logit model. The information matrix plays a crucial role for finding optimal designs in both approaches. Since the expressions for the relevant matrices look very different and it is not obvious how the two approaches are related, this has given rise to some confusion in the literature. We resolved this confusion by showing, in general, how the information matrices under the two approaches are related. There had also been some confusion regarding the inference parameters expressed as linear functions of the utility parameter vector τ. We theoretically established a unified approach to discrete choice experiments and introduce the general inference problem in terms of a simple linear function of τ. This allowed us to show that the commonly used effects coding under the A-criterion for the non-singular full-rank inference problem inherently attaches unequal importance to the elementary contrasts of attribute levels. On the contrary, we see that the orthonormal coding leads to attaching equal importance to the elementary contrasts of attribute levels. However, for a singular full-rank inference problem involving the full set of effects-coded parameters, we showed that the orthonormal coding provides an equivalent approach to obtain A-optimal designs.

12.1.2 Supersaturated Designs

In supersaturated designs, the $E(s^2)$ optimality criterion was proposed by Booth and Cox (1962). Jones and Majumdar (2014) proposed the $UE(s^2)$ optimality criterion. In this thesis (Chapter 9), we have compared the advantages of the two criteria and proposed methods for compromising between the two. Minimizing $UE_d(s^2)$ alone produces a large class of $UE(s^2)$-optimal designs that require secondary criteria to discriminate. An arbitrary $UE(s^2)$-optimal design may have poor projection properties. We have proposed secondary criteria to identify good $UE(s^2)$-optimal designs. A smaller value of SS along with minimum Q are common features of many $UE(s^2)$-optimal designs with good projection properties. Although no simple surrogate criterion is expected to always produce the best design, we have seen that minimizing SS followed by minimizing Q is an effective way of getting $UE(s^2)$-optimal designs with good projection properties. We have also provided easy constructions of superior $UE(s^2)$-optimal designs that are almost as efficient as $E(s^2)$-optimal designs. Alongside, we have also identified several families of designs that are both $E(s^2)$- and $UE(s^2)$-optimal.

12.1.3 Coverings

We have seen the definition of a coverings in Chapter 10. To recall, for positive integers t, v, k and λ with $t < k < v$, a t-(v, k, λ) *covering* is an incidence structure $(V, \mathcal{B})$ such that $|V| = v$, $|B| = k$ for all $B \in \mathcal{B}$, and each t-subset of V is contained in at least λ blocks in $\mathcal{B}$. The *covering number* $C_\lambda(v, k, t)$ is the minimum number of blocks in any t-(v, k, λ) covering. In the area of coverings, it is of crucial importance to find a lower bound on the covering number. We have obtained an improved lower bound for t-coverings and we have also obtained infinite families of coverings attaining our lower bounds (Horsley and Singh, 2018). For these families, our lower bound is an improvement over the best available lower bounds. We also found an infinite family where our bound is tight, that is, there exists a t-(v, k, λ) covering attaining our bound.

12.1.4 Pseudo generalized Youden designs

In Chapter 11, we have seen that in the area of row column designs, Kiefer (1958) introduced generalized Youden designs (GYDs) for eliminating heterogeneity in two directions. A GYD is a row-column design whose k rows form a balanced block design (BBD) and whose b columns do likewise. Later Cheng (1981b) introduced pseudo Youden designs (PYDs) in which $k = b$ and where the k rows and the b columns, considered together as blocks, form a BBD. In our work (Das et al., 2018) (Chapter 11), we have introduced and investigated pseudo generalized Youden designs (PGYDs) which generalise both GYDs and PYDs. We have obtained necessary conditions for the existence of a PGYD and have constructed families of PGYDs based on affine planes. We have also provided an exhaustive list of parameter sets satisfying $v \leqslant 25, k \leqslant 50, b \leqslant 50$ for which a PGYD exists.

12.2 Future work

Discrete choice experiments are studied under a generalized linear model where the responses are non-normal and therefore, the designs are dependent on the unknown parameters of the fitted model. In this thesis, we have mostly obtained theoretically D-optimal designs under the indifference assumption wherein designs are not dependent on the unknown parameters. This has been done to obtain theoretical results in the area. However, another important and interesting sub-area of research in choice experiments forgoes this indifference assumption. This is usually done with the help of Bayesian experiments. My future research will not only involve generalizing the results in this thesis to other well known optimality criteria but will also include obtaining Bayesian optimal designs for many of the practical situations inspired by the work carried out by Peter Goos, Roselinde Kessels, etc. (Kessels et al., 2006; Kessels, Goos and Vandebroek, 2008). A few open problems are also mentioned in Section 12.1.1.4, 12.1.1.6, and 12.1.1.7.

In the area of supersaturated designs, we will work on finding a lower bound for the SS criterion we defined for $UE(s^2)$- optimal designs (Chapter 9), and on constructing designs achieving this lower bound. We have already been working on this front and we will soon be able to present it in a manuscript form. Further works in this area will also be discussed in the said manuscript.

In the area of coverings, we have worked on improving the lower bounds for t-coverings using lessons learned from the lower bounds for 2-coverings. Whether these methods can be extended to generalized covering designs is an area for future work. There are several other problems posed in Bailey et al. (2011) on generalized covering designs. Therefore, in future work, we would like to generalize the results in this thesis to generalized covering designs and solve some of the problems posed in the paper.

In future research on PGYDs, we may try to obtain PGYDs for the parameter sets (given in Chapter 11) for which a non-GYD PGYD is unknown, for example for parameters $v = 8, k = 14, b = 28$. This will help in closing the gaps of the unavailable designs for a larger range of parameter sets. We can also try to construct tables as in Chapter 11 for a larger range of v so as to understand the missing designs and to obtain a general construction for the same, if possible.

Appendix A

Additional Material for Chapter 3

This appendix provides some additional details for the Chapter 3.

Design for Example 3.3 $k = 3, v_1 = 2, v_2 = 3, v_3 = 4, b = 1, N = s = 72.$

(000,111)	(000,222)	(000,333)	(100,211)	(100,322)	(100,033)	(020,131)	(020,313)	(120,302)
(001,112)	(001,223)	(001,330)	(101,212)	(101,323)	(101,030)	(021,132)	(021,310)	(121,303)
(002,113)	(002,220)	(002,331)	(102,213)	(102,320)	(102,031)	(022,133)	(022,311)	(122,300)
(003,110)	(003,221)	(003,332)	(103,210)	(103,321)	(103,032)	(023,130)	(023,312)	(123,301)
(010,121)	(010,232)	(010,303)	(110,221)	(110,332)	(110,003)	(020,202)	(120,231)	(120,013)
(011,122)	(011,233)	(011,300)	(111,222)	(111,333)	(111,000)	(021,203)	(121,232)	(121,010)
(012,123)	(012,230)	(012,301)	(112,223)	(112,330)	(112,001)	(022,200)	(122,233)	(122,011)
(013,120)	(013,231)	(013,302)	(113,220)	(113,331)	(113,002)	(023,201)	(123,230)	(123,012)

Lemma A.1. *A necessary and sufficient condition for $\tilde{C}_M = C_M$ to hold is that for each block and each attribute, the frequency distribution of the levels of the attribute are same for the two options.*

Proof. Let $P_{Mj} = ((P_j)'_1 \cdots (P_j)'_t \cdots (P_j)'_b)'$ where $(P_j)_t$ represents P_{Mj} for the tth block. Then the condition $W'P_M = 0$ is equivalent to the condition $1'(P_1)_t = 1'(P_2)_t$, $t = 1, \ldots, b$. Let $(P_j)_t = ((P_j)^1_t \cdots (P_j)^w_t \cdots (P_j)^k_t)$ where $(P_j)^w_t$ is of order $s \times (v_w - 1)$ and represents $(P_j)_t$ for the wth attribute. Therefore for $t = 1, \ldots, b$, if $1'(P_1)_t = 1'(P_2)_t$, then $1'(P_1)^w_t = 1'(P_2)^w_t$ for every w and t. Now, since the ith column of $(P_j)^w_t$ provides frequency of level i and level v_w in the wth attribute of the jth option in the tth block, therefore, $1'(P_1)^w_t = 1'(P_2)^w_t$ implies that the frequency of each of the levels of attribute w is same in the two options among the s choice pairs in block t.

The converse follows by noting that if for each block and each attribute, the frequency distribution of the levels of the attribute are same for the two options, then $1'(P_1)_t =$

$1'(P_2)_t$ for every t. $\square$

Proof of Theorem 3.1. The proof follows as a special case of Lemma A.1. $\square$

Proof of Theorem 3.2. Under the linear paired comparison model, a design d optimally estimates the main effects if $C_M = diag(C_{(1)}, \dots, C_{(k)})$ (see Großmann and Schwabe (2015)) where $C_{(i)} = z_i(I_{v_i-1} + J_{v_i-1})$ with $z_i = 2N/(v_i - 1)$, $i = 1, \dots, k$. This implies that C_M normalized by number of pairs would attain an optimal structure if $C_{(i)} = z_i(I_{v_i-1} + J_{v_i-1})$ with $z_i = 2/(v_i - 1)$, $i = 1, \dots, k$.

Since the $OA{+}G$ method of construction entails adding generators to the orthogonal array of strength $t, (t \geq 2)$, the off-diagonal elements of $P_M'P_M$ corresponding to two different attributes is zero since under each level of the first attribute, all the levels of the second attribute occur equally often. Also, since in an orthogonal array, under each column (attribute) the levels are equally replicated, to establish that each $C_{(i)}$ attains an optimal structure of the form $z_i(I_{v_i-1} + J_{v_i-1})$, it is enough to show that normalized $P_M'P_M$ corresponding to a paired choice design with one attribute, say at v levels, attains the structure $z(I_{v-1} + J_{v-1})$, where $z = 2/(v - 1)$.

Without loss of generality, we consider only v choice pairs for a typical attribute since under each column, the n rows of the orthogonal array involves v symbols each replicated n/v times. While using the generator g_j, let P_1^0, P_2^j be the $v \times (v-1)$ effects-coded matrix for the main effects for the first and second options, respectively, corresponding to any one attribute at v levels. When $h > 1$, note that P_M is the collection of different matrices generated out of the corresponding $\{P_1^0, P_2^j\}, j = 1 \dots, h$ of choice pairs. For notational simplicity, we denote P_1^0 by P_0 and P_2^j by P_j, $j = 1, \dots, v - 1$. Also, note that $1'P_j = 0$ and $\sum_{j=0}^{v-1} P_j = 0$.

Consider the information matrix $P_M'P_M$ normalized for v even. $v(v-1)P_M'P_M = \sum_{j=1}^{v-1}(P_0 - P_j)'(P_0 - P_j) = \sum_{j=1}^{v-1}(P_0'P_0 + P_j'P_j - P_0'P_j - P_j'P_0) = \sum_{j=1}^{v-1}\{2(I_{v-1} + J_{v-1})\} - P_0'(\sum_{j=1}^{v-1} P_j) - (\sum_{j=1}^{v-1} P_j')P_0 = \{2(v-1)(I_{v-1} + J_{v-1})\} - P_0'(-P_0) - (-P_0')P_0 = 2\{(v-1)(I_{v-1} + J_{v-1})\} + 2P_0'P_0 = 2v(I_{v-1} + J_{v-1})$. Thus, for v even, $h = v - 1$ generators of the type $g_j = 1, \dots, v - 1$ leads to the optimal structure of normalized $P_M'P_M$.

For v odd, we note that, if say, mth row of P_0 corresponds to the level i, then the mth row of P_{v-j} corresponds to the level $i - j$ (mod v). Similarly, if say, lth row of P_j corresponds to the level i, then the lth row of P_0 corresponds to the level $i - j$

(mod v). This makes the lth row of P_j and P_0 same as the mth row of P_0 and P_{v-j} for every two rows $l \neq m = 1, \ldots, v$. Therefore, for v odd, $P_j' P_0 = P_0' P_{v-j}$. Now, $v(v-1)/2 P_M' P_M = \sum_{j=1}^{(v-1)/2} (P_0 - P_j)'(P_0 - P_j) = \sum_{j=1}^{(v-1)/2} (P_0' P_0 + P_j' P_j - P_0' P_j - P_j' P_0) = \sum_{j=1}^{(v-1)/2} \{2(I_{v-1} + J_{v-1})\} - \sum_{j=1}^{(v-1)/2} (P_0' P_j + P_j' P_0) = (v-1)(I_{v-1} + J_{v-1}) - \sum_{j=1}^{(v-1)/2} (P_0' P_j + P_0' P_{v-j}) = (v-1)(I_{v-1} + J_{v-1}) - P_0' \sum_{j=1}^{(v-1)/2} (P_j + P_{v-j}) = (v-1)(I_{v-1} + J_{v-1}) - P_0' \sum_{j=1}^{v-1} P_j = (v-1)(I_{v-1} + J_{v-1}) - P_0'(-P_0) = v(I_{v-1} + J_{v-1})$. Thus, for v odd, $h = (v-1)/2$ generators of the type $g_j = 1, \ldots, (v-1)/2$ leads to the optimal structure of normalized $P_M' P_M$. $\square$

Proof of Theorem 3.5. For a given $OA(n_1, k+1, v_1 \times \cdots \times v_k \times \delta, 2)$, corresponding to the k attributes at levels $v_i, i = 1, \ldots, k$, let d_1 be the design constructed through $OA+G$ method using $h = lcm(v_1, \ldots, v_k)$ generators. Then d_1 with parameters k, $v_1, \ldots, v_k$, $b = 1$, $s = hn_1$ is an optimal paired choice design. From d_1, the choice pairs obtained through each of the h generators constitute a block of size n_1. This is true since n_1 rows of a block form the orthogonal array in the first option and, with labels re-coded through the generator, in the second option and hence the conditions in Theorem 3.1 are satisfied.

Finally, we use the δ symbols of the $(k+1)$th column of the orthogonal array for further blocking giving a paired choice block design d_2 with parameters k, $v_1, \ldots, v_k$, $b = h\delta$, $s = n_1/\delta$. This is true since for every attribute in each of the blocks so formed, each of the v_i levels occurs equally often under ith attribute and hence by Theorem 3.1, d_2 is optimal in $\mathcal{D}_{k,b,s}$. $\square$

Proofs for Theorem 3.7 and Theorem 3.9 require a result from Dey (2009) that is given below.

Lemma A.2 (Dey (2009))**.** *Consider $v(v-1)/2$ combinations involving v levels taken two at a time. Then, for v odd, the combinations can be grouped into $(v-1)/2$ replicates each comprising v combinations. The groups are $\{(i, v-2-i), (i+1, v-1-i), \ldots, (i+v-1, v-2-(i-(v-1)))\}$ and the levels are reduced modulo v; $i = 0, \ldots, (v-3)/2$.*

Proof of Theorem 3.7. Theorem 3 of Graßhoff et al. (2004) states that from $m(\geq k)$ rows of a Hadamard matrix H_m of order m, an optimal paired choice design d_3 with parameters $k, v, b = 1, s = mv(v-1)/2$ is constructed using the $v(v-1)/2$ combinations of v levels taken two at a time. From every row of $\{H_m, -H_m\}$, $v(v-1)/2$ choice pairs are obtained by replacing '1' in the row by the first column of the combinations and '-1' in the row by the second column of the combinations. If v is odd, then $(v-1)/2$ is an integer and the $v(v-1)/2$ combinations can be arranged in rows such that each of the two columns have every level appearing equally often. Such an arrangement is always possible and follows from systems of distinct representatives. Therefore, corresponding to each of the rows of $\{H_m, -H_m\}$, using $v(v-1)/2$ choice pairs as a block, a paired choice block design with parameters $k, v, b = m, s = v(v-1)/2$ is obtained which, following Theorem 3.1 is optimal. Now for v odd, from Dey (2009), $v(v-1)/2$ combinations involving v levels taken two at a time can be grouped into $(v-1)/2$ replicates each comprising v combinations. Therefore, the blocks generated by each row of H_m can be further broken into $(v-1)/2$ blocks each of size v, which gives us d_4. $\square$

Proof of Theorem 3.9. Construction 3.2 of Demirkale, Donovan and Street (2013) uses an $OA(n_2, k+1, v^k \times v_{k+1}, 2)$ with $v_{k+1} = n_2/v$ and forms v_{k+1} parallel sets each having v rows. Then, an optimal paired choice design with parameters $k, v, b = 1, s = v_{k+1}\binom{v}{2}$ is constructed using the $v(v-1)/2$ combinations of v numbers $\{1, \ldots, v\}$ taken two at a time. Let $\{i, j\}$ be a typical row. Then, for each such row of size two, corresponding rows i and j from each of the v_{k+1} parallel sets are chosen to form the choice pairs of the optimal paired choice design d_6. Again as earlier, for v odd, the $v(v-1)/2$ combinations can be arranged in rows such that each of the two columns have every number appearing equally often. Considering the $v(v-1)/2$ choice pairs, obtained from a parallel set, as a block, we get the paired choice block design with parameters $k, v, b = v_{k+1}, s = v(v-1)/2$ which is optimal in $\mathcal{D}_{k,b,s}$. Further proof follows on the same lines as the proof of Theorem 3.7 by treating the pairs generated by each parallel set as blocks. $\square$

Proof of Theorem 3.10. Theorem 4 of Graßhoff et al. (2004) uses an $OA(n_3, k+1, m_1 \times \cdots \times m_k \times \delta, 2)$ with $m_i = v_i(v_i-1)/2$ for some odd v_i to construct an optimal paired choice design d_7 with parameters $k, v_i, \ldots, v_k, b = 1, s = n_3$. This method involves a one-one mapping between m_i levels of orthogonal array to the $v_i(v_i-1)/2$ combinations on v_i symbols. For a combination $\{i, j\}$ corresponding to a symbol of an orthogonal array, the first option in a pair is obtained by replacing i in place of that symbol and the second option has j in the corresponding position. Then, similar to construction of Theorem 3.9, using the δ (≥ 1) symbols of the $(k+1)$th column of the orthogonal array for blocking gives us an optimal paired choice block design d_8 with parameters $k, v_i, \ldots, v_k, b = \delta, s = n_3/\delta$. Note that this method is applicable only for odd v_i since for even v_i, it is not possible to arrange $v_i(v_i-1)/2$ combinations in a position-balanced manner. $\square$

Proof of Theorem 3.11. From Theorem 3.1, for each of the h generators, a paired choice design using the $OA+G$ method of construction is optimal under the broader main effects block model if $P_M'P_I=0$.

For a given generator, to show that $P_M'P_I=0$, it suffices to show that the inner product of the columns of P_M corresponding to the mth main effect and the columns of P_I corresponding to the two-factor interaction effect of ith and jth attribute is zero. Using an $OA(n_1,k,v_1\times\cdots\times v_k,3)$ in the $OA+G$ method of construction, we establish the result through the following two cases.

Case (i) $m=i$: In an orthogonal array of strength 2, each of the v_iv_j combinations occur equally often $n_1/(v_iv_j)$ times as rows. Therefore, since the paired choice design is based on the orthogonal array, for showing that $P_M'P_I=0$, it suffices to show that $P_M'P_I=0$ for one of the $n_1/(v_iv_j)$ sets of v_iv_j rows of the type $(i,j); i=0,\ldots,v_i-1; j=0,\ldots,v_j-1$. For such v_iv_j rows, note that P_{My}, $(y=1,2)$, corresponding to the jth attribute, can be partitioned into v_i sets $P_{My(j)}$ each of v_j distinct rows. Then, $1'P_{My(j)}=0$. Let P_{Iy} corresponding to the ith attribute fixed at level i_l $(i_l=0,\ldots,v_i-1)$ and the jth attribute taking v_j distinct levels be represented by $P_{Iy(i_lj)}$. Then, the columns of $P_{Iy(i_lj)}$ are multiples of either $P_{My(j)}$ or 0_v. Therefore, $1'P_{Iy(i_lj)}=0$ for $y=1,2$.

Let P_M corresponding to the ith attribute at level i_l be represented by X_{i_l}. Then, $X_{i_l}=1x_{i_l}'$ where x_{i_l}' is a row vector of size v_i-1. Therefore, $P_M'P_I=\sum_{i_l=0}^{v_i-1}X_{i_l}'(P_{I1(i_lj)}-P_{I2(i_lj)})=\sum_{i_l=0}^{v_i-1}x_{i_l}(1'P_{I1(i_lj)}-1'P_{I2(i_lj)})=0$.

Case (ii) $m\neq i$: In an orthogonal array of strength 3, each of the $v_iv_jv_m$ combinations occur equally often $n_1/(v_mv_iv_j)$ times as rows. Therefore, as in Case (i), for showing that $P_M'P_I=0$, it suffices to show that $P_M'P_I=0$ for one of the $n_1/(v_mv_iv_j)$ sets of $v_mv_iv_j$ rows of the type $(m,i,j); m=0,\ldots,v_m-1; i=0,\ldots,v_i-1; j=0,\ldots,v_j-1$.

For such $v_mv_iv_j$ rows, note that P_{Iy}, $(y=1,2)$, corresponding to the ith and jth attribute, can be partitioned into v_m sets $P_{Iy(ij)}$ each of v_iv_j distinct rows. Therefore, $1'P_{Iy(ij)}=0$ for $y=1,2$, since from Case (i), $1'P_{Iy(i_lj)}=0$ for the ith attribute at level i_l.

Finally, since for the mth attribute at level m_l $(m_l=0,\ldots,v_m-1)$, the v_iv_j combinations under attributes i and j occur equally often, therefore $P_M'P_I=\sum_{m_l=0}^{v_m-1}X_{m_l}'(P_{I1(ij)}-P_{I2(ij)})=\sum_{m_l=0}^{v_m-1}x_{m_l}(1'P_{I1(ij)}-1'P_{I2(ij)})=0$. $\square$

Proof of Theorem 3.14. From Lemma A.1, $W'P_M = 0$ if and only if for each attribute under the choice pairs having foldover in the second option of a choice pair, the level l ($l = 0, 1$) appears equally often in both the options in every block and thus, the frequency of the pair $(1, 0)$ is same as the frequency of the pair $(0, 1)$ under every attribute in each block.

Let $P_I = (Y_1' \cdots Y_t' \cdots Y_b')'$ where Y_t is the $s \times k(k-1)/2$ matrix corresponding to the tth block. With $(P_{Ij})_t$ representing P_{Ij} for the tth block, $Y_t = (P_{I1})_t - (P_{I2})_t$. Then, the condition $W'P_I = 0$ is equivalent to the condition $1'(P_{I1})_t = 1'(P_{I2})_t$ for every $t = 1, \ldots, b$. Consider $(P_{Ij})_t = ((P_{Ij})_t^{12} \cdots (P_{Ij})_t^{lm} \cdots (P_{Ij})_t^{(k-1)k})$ where $(P_{Ij})_t^{lm}$ is of order $s \times 1$ and represents $(P_{Ij})_t$ for the two-factor interaction between the lth and the mth attribute. Therefore, the necessary and sufficient condition for $1'(P_{I1})_t = 1'(P_{I2})_t$ is that $1'(P_{I1})_t^{lm} = 1'(P_{I2})_t^{lm}$ for every l and m.

In the tth block, for the choice pairs where either both the attributes have a foldover in the second option or both do not have a foldover in the second option, the corresponding rows in $(P_{I2})_t^{lm}$ are same as the corresponding rows in $(P_{I1})_t^{lm}$.

However, for the pairs in which one attribute has a foldover in the second option and another does not have foldover in the second option, the corresponding rows in $(P_{I2})_t^{lm}$ are negative of the corresponding rows in $(P_{I1})_t^{lm}$. In such a case, $1'(P_{I1})_t^{lm} = 1'(P_{I2})_t^{lm}$ if and only if $1'(P_{I1})_t^{lm} = -1'(P_{I2})_t^{lm} = 0$. Now, $1'(P_{I1})_t^{lm} = 0$ if and only if the frequency of the pairs from the set $\{(01, 00), (01, 11), (10, 00), (10, 11)\}$ is same as the frequency of the pairs from the set $\{(00, 01), (00, 10), (11, 01), (11, 10)\}$ under the lth and the mth attribute. $\square$

Proof of Theorem 3.15. In steps (iii)-(iv), corresponding to an element f of F, make the first set of $2^{\alpha-1}2^{k-\alpha-2} = 2^{k-3}$ blocks having choice pairs $(ab, a'b)$, $(ab', a'b')$, $(a'c, ac)$, $(a'c', ac')$. Similarly, following the steps (iii)-(iv), we make an additional set of 2^{k-3} blocks having choice pairs $(ac, a'c)$, $(ac', a'c')$, $(a'b, ab)$, $(a'b', ab')$. Note that each of the constructed blocks satisfy conditions (i) and (ii) of Theorem 3.14. This gives rise to a total of 2^{k-2} sets of blocks each of size 4. The way we have constructed the choice pairs in steps (iii)-(iv), it follows that the collection of first option in the 2^k choice pairs forms a complete factorial having 2^k combinations. Furthermore, the additional set of 2^{k-3} blocks, in the construction, is identical to the first set of 2^{k-3} blocks. Accordingly, we retain only the first set of 2^{k-3} blocks. This gives rise to a total of 2^{k-1} choice pairs divided into 2^{k-3}

blocks each of size 4. Therefore, step (v) gives an optimal paired choice block design d_2^I with parameters $k, v = 2, s = 4, b$ where $b = 2^{k-3}\binom{k}{q}$ for k odd and $b = 2^{k-3}\binom{k+1}{q+1}$ for k even.

$\square$

Appendix B

Additional Material for Chapter 5

This appendix contains some extra details about Chapter 5.

Proof of Theorem 5.3. First we give a proof of $trace(M_d^{-1}) \geq \sum_{p=1}^{k} trace(M_{dpp}^{-1})$.

Let for $k = 2$, the 2×2 partitioned matrix M is $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$. Then $M^{-1} = \begin{bmatrix} (M_{11} - M_{12}M_{22}^{-1}M_{21})^{-1} & -M_{11}^{-1}M_{12}(M_{22} - M_{21}M_{11}^{-1}M_{12})^{-1} \\ -M_{22}^{-1}M_{21}(M_{11} - M_{12}M_{22}^{-1}M_{21})^{-1} & (M_{22} - M_{21}M_{11}^{-1}M_{12})^{-1} \end{bmatrix}$. Since $M_{12}M_{22}^{-1}M_{21}$ is non-negative definite, $(M_{11} - M_{12}M_{22}^{-1}M_{21}) \leq M_{11}$ and therefore $(M_{11} - M_{12}M_{22}^{-1}M_{21})^{-1} \geq M_{11}^{-1}$. Similarly, $(M_{22} - M_{21}M_{11}^{-1}M_{12})^{-1} \geq M_{22}^{-1}$. Therefore, $trace(M^{-1}) = trace(M_{11} - M_{12}M_{22}^{-1}M_{21})^{-1} + trace(M_{22} - M_{21}M_{11}^{-1}M_{12})^{-1} \geq trace(M_{11}^{-1}) + trace(M_{22}^{-1})$.

Using the principle of induction, let the relationship holds for $t = k - 1$ or that $trace(M^{-1}) \geq \sum_{p=1}^{t} trace(M_{pp}^{-1})$.

Now, the relation for $t = k$ can be proved considering one block matrix (say, M_{11}) consisting of $k - 1$ blocks and M_{22} being the last block and proceeding in the similar way as for 2×2 partitioned matrices.

Finally, using Lemma 5.1, the proof follows. $\square$

Proof of Theorem 5.5. First, we give a proof of $det(M_d^{-1}) \geq \prod_{p=1}^{k} det(M_{dpp}^{-1})$ by induction. Let for $k = 2$, the 2×2 partitioned matrix M is $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$. Then $det(M) = det(M_{11})det(M_{22} - M_{21}M_{11}^{-1}M_{12}) \leq det(M_{11})det(M_{22})$ since $M_{21}M_{11}^{-1}M_{12}$ is non-negative definite. Therefore, $det(M^{-1}) \geq det(M_{11}^{-1})det(M_{22}^{-1})$. Let the relationship hold for $t = k - 1$ or that $det(M^{-1}) \geq \prod_{p=1}^{t} det(M_{pp}^{-1})$.

Now, the relation for $t = k$ can be proved considering one block matrix (say, M_{11}) consisting of $k-1$ blocks and M_{22} being the last block and proceeding in the similar way as for 2×2 partitioned matrices.

Finally, using Lemma 5.4, the proof follows. $\square$

Designs as discussed in Chapter 5

$$ad_{(2,4)} = \begin{pmatrix} 20, & 01 \\ 21, & 10 \\ 12, & 00 \\ 02, & 11 \end{pmatrix} \qquad ad_{(2,5)} = \begin{pmatrix} 00, & 22 \\ 01, & 10 \\ 02, & 11 \\ 11, & 20 \\ 12, & 21 \end{pmatrix} \qquad a_{(2,6)} = \begin{pmatrix} 21, & 02 \\ 20, & 11 \\ 22, & 10 \\ 11, & 00 \\ 12, & 01 \\ 01, & 10 \end{pmatrix}$$

$$ad_{(3,7)} = \begin{pmatrix} 211, & 022 \\ 222, & 000 \\ 202, & 121 \\ 220, & 112 \\ 101, & 010 \\ 110, & 001 \\ 011, & 100 \end{pmatrix} \qquad d_{(2,8)} = \begin{pmatrix} 21, & 00 \\ 01, & 22 \\ 10, & 21 \\ 12, & 20 \\ 20, & 11 \\ 01, & 12 \\ 02, & 11 \\ 10, & 02 \end{pmatrix}$$

$$a_{(2,9)} = \begin{pmatrix} 22, & 01 \\ 20, & 02 \\ 21, & 12 \\ 22, & 10 \\ 11, & 00 \\ 11, & 00 \\ 10, & 01 \\ 11, & 00 \\ 01, & 10 \end{pmatrix} \qquad a^{+}_{(2,9)} = \begin{pmatrix} 00, & 11 \\ 00, & 12 \\ 00, & 21 \\ 01, & 12 \\ 01, & 20 \\ 02, & 10 \\ 02, & 20 \\ 10, & 21 \\ 11, & 22 \end{pmatrix}$$

Appendix C

Additional Material for Chapter 6

This appendix contains some extra details about Chapter 6.

Proof of Theorem 6.1. The proof of Theorem 6.1 is immediate from the following Lemma C.1, C.2 and C.3. Lemma C.1 talks about the off-diagonal entries of the information matrix. Lemma C.2 and C.3 talks about the diagonal blocks for main effects matrix and the two-factor interaction effects matrix, respectively. □

Lemma C.1. *If a paired choice design is constructed by adding generators to orthogonal array $OA(n, 3^k, 4)$, then*

(i) $X^T X$ *and* $Y^T Y$ *are block diagonal matrices with blocks of size two and four, respectively;*

(i) $X^T Y = 0$.

Proof. Note that for an factor with three levels, the only possible generators are 0, 1 or 2. Now, for a design with one factor sum of the rows of the main effects matrix X is zero as long as 0,1 and 2 appear equally often in the first part of a paired choice design and that the generators are either $G_1 = 0$, $G_1 = 1$ or $G_1 = 2$. Similarly, for a design with two factors, sum of the rows of the two-factor interaction effects matrix Y is zero as long as each of the 9 pairs 00,01, 01, 10, 11, 12, 20, 21 and 22 appear equally often in the first part of a paired choice design and that the generators are either of the following 9 generators, $G_1 = 00$, $G_1 = 01$, $G_1 = 02$, $G_1 = 10$, $G_1 = 11$, $G_1 = 12$, $G_1 = 20$, $G_1 = 21$ or $G_1 = 22$.

Since a strength four orthogonal array is a strength 2 orthogonal array as well, our construction method implies that for any two factors, each of the 9 combinations appear equally often. Therefore, for each 0 in one factor of the first part of the pair, each of 0, 1 and 2 appear in another factor of the first part of the pair. Then, whatever the generator maybe, the sum of three rows of X for second factor is 0. Similar thing hold for 1 and 2 in that factor of the first part of the pair. Therefore, the off-diagonal blocks corresponding to X^TX are zero matrices. Similarly, for Y^TY, since for each of the 9 combinations for two factors, the other two factors would have all 9 combinations appearing equally often making the sum of the rows of corresponding Y as 0. Since these will happen for any set of 4 factors of Y, we get off-diagonal blocks of Y^TY as 0.

Case (ii) can also be proved on the similar lines by using strength 3 properties of an the staring orthogonal array. □

The construction methods given in Chapter 6 are $OA+G$ methods and the starting orthogonal array is $OA(n, 3^k, 4)$. Therefore, from Lemma C.1(i), one can write X^TX as a block diagonal matrix with k blocks corresponding to each of the k factors, such that $X^TX = diag((X^TX)_1, \ldots, (X^TX)_\ell, \ldots, (X^TX)_k)$ where ℓth block $(X^TX)_\ell$ corresponds to ℓth factor. Similarly, Y^TY is a block diagonal matrix with $\binom{k}{2}$ blocks such that $Y^TY = diag((Y^TY)_{12}, \ldots, (Y^TY)_{\ell m}, \ldots, (Y^TY)_{(k-1)k})$ where ℓmth block $(X^TX)_{\ell m}$ corresponds to a two-factor interaction between ℓth and mth factors.

Lemma C.2. *In an $OA+G$ method, starting from an $OA(n, 3^k, 4)$, let a paired choice design be constructed using $\binom{k}{t}$ distinct generators such that each of the generators has non 0s in all possible t positions and 0s in remaining $k-t$ positions. Then each block in X^TX has the structure $h_1(t)M_2$.*

Proof. Firstly, note that a paired choice design using $\binom{k}{t}$ distinct generators along with an orthogonal array of size n has a total of $N = n\binom{k}{t}$ choice pairs. It is easy to see that for each of the factor, $\binom{k-1}{t}$ generators are 0s and remaining generators $\binom{k}{t}$- $\binom{k-1}{t} = \frac{t}{k}\binom{k}{t}$ are non 0s. Therefore, the number of contributing generators is $\frac{t}{k}\binom{k}{t}$.

It is easy to see that for ℓth factor, paired choice design generated using any non 0 generator gives X^TX corresponding to the ℓth factor as $(X^TX)_\ell = n(I_2 + J_2)$. To see this, for ℓth factor, let $X_1 = (X_{11}^T \quad \ldots X_{1u}^T)^T$ where $n = 3u$ and $X_{1w}, w = 1, \ldots, u$ is a 3×2 matrix with rows being some permutation of the three rows: $(1\ 0), (0\ 1)$ and

$(-1\ -1)$. Without loss of generality we assume that $X_{1w} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$. Using $g_{1\ell} = 1$, we get $X_{2w} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \\ 1 & 0 \end{bmatrix}$ and using $g_{1\ell} = 2$, we get $X_{2w} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$. It is then easy to see that for ith factor $X_w^T X_w = 3(I_2 + J_2)$ where $X_w = X_{1w} - X_{2w}$. Therefore, $(X^T X)_\ell = n(I_2 + J_2)$.

Hence, by counting the total number of generators and their contribution, we get that $(X^T X)_\ell = n\frac{t}{k}\binom{k}{t}(I_2 + J_2) = N\frac{t}{k}(I_2 + J_2)$. $\square$

Lemma C.3. *In an $OA + G$ method, starting from an $OA(n, 3^k, 4)$, let a paired choice design be constructed using $\binom{k}{t}$ distinct generators such that each of the generators has non 0s in all possible t positions and 0s in remaining $k-t$ positions. Additionally, for any two factors, all the generators (rows) with both non 0 entry should be such that they can be clubbed in several groups of two generators of the type $\{(11, 12), (11, 21), (22, 21), (22, 12)\}$. Then, each block in $Y^T Y$ has the structure $h_2(t) M_2 \otimes M_2$.*

Proof. Firstly, note that a paired choice design using $\binom{k}{t}$ distinct generators along with an orthogonal array of size n has a total of $N = n\binom{k}{t}$ choice pairs. It is easy to see that for any two factors ℓ and m, there are

(a) $\binom{k-2}{k-t}$ generators such that generators for both ℓth and mth factors are non-zero;

(b) $\binom{k-2}{t}$ generators such that generators for both ℓth and mth factors are zero;

(c) $\binom{k}{t} - \binom{k-2}{t} - \binom{k-2}{k-t}$ generators such that generators for ℓth factor is zero and mth factor is non-zero or vice-versa.

We now see the contribution from each of these types of generators to $(Y^T Y)_{\ell m}$. For ℓth and mth factor, note that since Y_1 will have effects-coding corresponding to each of the 3^2 factor-level combinations occurring equally often, we can write $Y_1 = (Y_{11}^T \ \ldots Y_{1u}^T)^T$ where $n = 9u$ and $Y_{1w}, w = 1, \ldots, u$ is a 9×4 matrix with rows being some permutation of the nine rows: $(1\ 0\ 0\ 0), (0\ 1\ 0\ 0), (-1\ -1\ 0\ 0), (0\ 0\ 1\ 0), (0\ 0\ 0\ 1), (0\ 0\ -1\ -1), (-1\ 0\ -1\ 0), (0\ -1\ 0\ -1)$, and $(1\ 1\ 1\ 1)$. Then, without loss of generality if we use the generators such that generators for ℓth factor is zero and mth factor is non-zero or vice-versa, we get $Y_{1w}^T Y_{1w} = 3(I_2 + J_2) \otimes (I_2 + J_2)$ for each w. Then, since $u = n/9$, we get $(Y^T Y)_{\ell m} = \sum_w Y_{1w}^T Y_{1w} = \frac{n}{3}(I_2 + J_2) \otimes (I_2 + J_2)$. Similarly, when we use two sets of generators such that it is one among $\{(11, 12), (11, 21), (22, 21), (22, 12)\}$, the total number of pairs are then $2n$ and since we get $Y_{1w}^T Y_{1w} = 3(I_2 + J_2) \otimes (I_2 + J_2)$ and $u = n/9$, we get $(Y^T Y)_{\ell m} = \sum_w Y_{1w}^T Y_{1w} = \frac{n}{3}(I_2 + J_2) \otimes (I_2 + J_2)$.

Now generators of type (a) contribute $\frac{n}{3}(I_2 + J_2) \otimes (I_2 + J_2)$ to $Y^T Y$, one generator of type (c) contribute $\frac{n}{3}(I_2 + J_2) \otimes (I_2 + J_2)$ to $Y^T Y$ and one generator of type (b) contributes nothing to $Y^T Y$.

Since there are a total of $\binom{k}{t}$ generators of type (a)-(c), we get the total contribution to $Y^T Y$ as $\frac{1}{2}\binom{k-2}{k-t}\frac{n}{3}(I_2 + J_2) \otimes (I_2 + J_2) + (\binom{k}{t} - \binom{k-2}{t} - \binom{k-2}{k-t})\frac{n}{3}(I_2 + J_2) \otimes (I_2 + J_2)$. Therefore, we get $(Y^T Y)_{\ell m} = h_2(t) M_2 \otimes M_2$. $\square$

Appendix D

Additional Material for Chapter 8

This appendix provides some additional details for the Chapter 8.

Proof of Theorem 8.1. It is easy to see that

$$\Lambda = \frac{1}{N}\sum_{n=1}^{N}\Lambda_n = \frac{1}{N}\sum_{n=1}^{N}\sum_{1=j<j'=m}\Delta_{n(jj')}(r,r'),$$

where

$$\left(\sum_{l\in S_n} e^{\tau_l}\right)^2 \Delta_{n(jj')}(r,r') = \begin{cases} -e^{\tau_r}e^{\tau_{r'}}, & r\neq r', r, r' = 1,\dots,L, \\ +e^{\tau_r}e^{\tau_{r'}}, & r = r', r = 1,\dots,L, \end{cases}$$

As per the definition, options are lexicographically arranged in Λ as well as B_H. Also, the row in H_j corresponding to an option τ_w, is given by wth column of B_H. Let B_H be as defined in the Chapter 8.

Let $B_H = [B_1 \quad b_r \quad B_2 \quad b_{r'} \quad B_3]$, where B_1 is of order $p\times(r-1)$, B_2 is of order $p\times(r'-r-1)$, and B_3 is of order $p\times(L-r')$. Without loss of generality, let h_{nj} and $h_{nj'}$ correspond to the rth and the r'th lexicographic labels of rth and r'th options respectively, with $r<r'$. Then, $b_r = h_{nj}^T$ and $b_{r'} = h_{nj'}^T$. Here ^{T} denotes the transposition. Therefore,

$$\begin{aligned} B_H\Lambda B_H^T &= \frac{1}{N}B_H\sum_{n=1}^{N}\left(\frac{1}{m^2}\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m}\Delta_{n(jj')}(r,r')\right)B_H^T \\ &= \frac{1}{m^2N}\sum_{n=1}^{N}\sum_{j=1}^{m-1}\sum_{j'=j+1}^{m}B_H\Delta_{n(jj')}(r,r')B_H^T. \end{aligned}$$

Now, by definition,

$$\left(\sum_{l\in S_n} e^{\tau_l}\right)^2 \Delta_{n(jj')} = e^{\tau_r} e^{\tau_{r'}} \begin{bmatrix} 0_{L\times(r-1)} & w_{njj'}^T & 0_{L\times(r'-r-1)} & -w_{njj'}^T & 0_{L\times(L-r')} \end{bmatrix},$$

where $w_{njj'} = \begin{bmatrix} 0_{1\times(r-1)} & 1 & 0_{1\times(r'-r-1)} & -1 & 0_{1\times(L-r')} \end{bmatrix}$. Then,

$$\begin{aligned} B_H \Delta_{n(jj')} B_H^T &= \frac{e^{\tau_r} e^{\tau_{r'}}}{\left(\sum_{l\in S_n} e^{\tau_l}\right)^2} \begin{bmatrix} 0_{p\times(r-1)} & (h_{nj}^T - h_{nj'}^T) & 0_{p\times(r'-r-1)} & (h_{nj'}^T - h_{nj}^T) & 0_{p\times(L-r')} \end{bmatrix} B_H^T \\ &= \frac{e^{\tau_r} e^{\tau_{r'}}}{\left(\sum_{l\in S_n} e^{\tau_l}\right)^2} ((h_{nj}^T - h_{nj'}^T) h_{nj} - (h_{nj}^T - h_{nj'}^T) h_{nj'}) = \frac{e^{\tau_r} e^{\tau_{r'}}}{\left(\sum_{l\in S_n} e^{\tau_l}\right)^2} (h_{nj}^T - h_{nj'}^T)(h_{nj} - h_{nj'}). \end{aligned}$$

From the definition, we get,

$$\begin{aligned} B_H \Lambda B_H^T &= \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\left(\sum_{l\in S_n} e^{\tau_l}\right)^2} \sum_{j=1}^{m-1} \sum_{j'=j+1}^{m} e^{\tau_r} e^{\tau_{r'}} (h_{nj} - h_{nj'})^T (h_{nj} - h_{nj'}) \\ &= \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{m-1} \sum_{j'=j+1}^{m} P_{nj} P_{nj'} (h_{nj} - h_{nj'})^T (h_{nj} - h_{nj'}). \end{aligned}$$

Upon simple rearrangement and using the fact that $\sum_{j=1}^m P_{nj} = 1$ for each $n = 1, \ldots, N$, we get,

$$\begin{aligned} B_H \Lambda B_H^T &= \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{j=1}^{m} h_{nj}^T h_{nj} P_{nj}(1 - P_{nj}) - \sum_{j\neq j'=1}^{m} h_{nj}^T h_{nj'} P_{nj} P_{nj'} \right) \\ &= \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{j=1}^{m} h_{nj}^T h_{nj} (P_{nj} - 2P_{nj}^2) + \sum_{j=1}^{m} h_{nj}^T h_{nj} P_{nj}^2 (\sum_{j'=1}^{m} P_{nj'}) - 2 \sum_{j\neq j'=1}^{m} h_{nj}^T h_{nj'} P_{nj'} P_{nj'} \right. \\ &\quad \left. + \sum_{j_1\neq j_2=1}^{m} h_{nj_1}^T h_{nj_2} P_{nj_1} P_{nj_2} (\sum_{j'=1}^{m} P_{nj'}) \right) \\ &= \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{m} P_{nj} \left(h_{nj}^T h_{nj} - \sum_{j'=1}^{m} h_{nj'}^T h_{nj} P_{nj'} - h_{nj}^T \sum_{j'=1}^{m} h_{nj'} P_{nj'} + \sum_{j'=1}^{m} h_{nj'}^T h_{nj'} P_{nj'}^2 \right. \\ &\quad \left. + \sum_{j_1\neq j_2} h_{nj_1}^T h_{nj_2} P_{nj_1} P_{nj_2} \right) \\ &= \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{m} P_{nj} \left(h_{nj}^T h_{nj} - \sum_{j'=1}^{m} h_{nj'}^T h_{nj} P_{nj'} - h_{nj}^T \sum_{j'=1}^{m} h_{nj'} P_{nj'} + (\sum_{j'=1}^{m} h_{nj'} P_{nj'})^T (\sum_{j'=1}^{m} h_{nj'} P_{nj'}) \right) \\ &= \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{m} (h_{nj}\sqrt{P_{nj}} - \sqrt{P_{nj}} \sum_{j'=1}^{m} h_{nj'} P_{nj'})^T (h_{nj}\sqrt{P_{nj}} - \sqrt{P_{nj}} \sum_{j'=1}^{m} h_{nj'} P_{nj'}) \end{aligned}$$

Hence,

$$B_H \Lambda B_H^T = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{m} (h_{nj} - \sum_{j'=1}^{m} h_{nj'} P_{nj'})^T P_{nj} (h_{nj} - \sum_{j'=1}^{m} h_{nj'} P_{nj'}).$$

□

Proof of $(\mathbf{B_E B_E^T})^{-1}\mathbf{B_E} = \mathbf{SG}$. From (8.4), G can be written as,

$$G = \begin{pmatrix} K_{v_1} & \otimes & \frac{1}{\sqrt{v_2}}1_{v_2}^T & \otimes & \cdots & \otimes & \frac{1}{\sqrt{v_k}}1_{v_k}^T \\ \frac{1}{\sqrt{v_2}}1_{v_1}^T & \otimes & K_{v_2} & \otimes & \cdots & \otimes & \frac{1}{\sqrt{v_k}}1_{v_k}^T \\ \vdots & \otimes & \vdots & \otimes & \cdots & \otimes & \vdots \\ \frac{1}{\sqrt{v_1}}1_{v_1}^T & \otimes & \frac{1}{\sqrt{v_2}}1_{v_2}^T & \otimes & \cdots & \otimes & K_{v_k} \end{pmatrix}$$

Also, from definition, S is a rectangular block diagonal matrix given by,

$$S = \frac{1}{\sqrt{L}} \begin{pmatrix} D_{v_1} & 0_{(v_1-1)\times v_2} & \cdots & 0_{(v_1-1)\times v_k} \\ 0_{(v_2-1)\times v_1} & D_{v_2} & \cdots & 0_{(v_2-1)\times v_k} \\ \vdots & \vdots & \vdots & \vdots \\ 0_{(v_k-1)\times v_1} & 0_{(v_k-1)\times v_2} & \cdots & D_{v_k} \end{pmatrix}.$$

Therefore, on multiplication,

$$SG = \frac{1}{L} \begin{pmatrix} T_{v_1} & \otimes & 1_{v_2}^T & \otimes & \cdots & \otimes & 1_{v_k}^T \\ 1_{v_1}^T & \otimes & T_{v_2} & \otimes & \cdots & \otimes & 1_{v_k}^T \\ \vdots & \otimes & \vdots & \otimes & \cdots & \otimes & \vdots \\ 1_{v_1}^T & \otimes & 1_{v_2}^T & \otimes & \cdots & \otimes & T_{v_k} \end{pmatrix}, \tag{AD.1}$$

where T_{v_i} is the $(v_i - 1) \times v_i$ matrix of the first $v_i - 1$ rows of $v_i K_{v_i}$. Now, it is easy to see that

$$(B_E B_E^T)^{-1} = \frac{1}{L} \begin{pmatrix} v_1 I_{v_1-1} - J_{v_1-1} & 0 & \cdots & 0 \\ 0 & v_2 I_{v_2-1} - J_{v_2-1} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & v_k I_{v_k-1} - J_{v_k-1} \end{pmatrix}.$$

and that $(v_i I_{v_i-1} - J_{v_i-1})B_e^{(i)} = T_{v_i}$. Therefore,

$$(B_E B_E^T)^{-1} B_E = \frac{1}{L} \begin{pmatrix} T_{v_1} & \otimes & 1_{v_2}^T & \otimes & \cdots & \otimes & 1_{v_k}^T \\ 1_{v_1}^T & \otimes & T_{v_2} & \otimes & \cdots & \otimes & 1_{v_k}^T \\ \vdots & \otimes & \vdots & \otimes & \cdots & \otimes & \vdots \\ 1_{v_1}^T & \otimes & 1_{v_2}^T & \otimes & \cdots & \otimes & T_{v_k} \end{pmatrix}. \tag{AD.2}$$

Then the result follows from (AD.1) and (AD.2). □

Proof of Theorem 8.5. For every $p \times L$ matrix B_H whose rows are not necessarily orthogonal but that span the same vector space as the rows of $p \times L$ matrix B_O, there exists a non-singular matrix Q of order p such that

$$B_H = QB_O. \tag{AD.3}$$

Now, $B_H = QB_O$ implies $B_O = Q^{-1}B_H$. Also, since $B_O B_O^T = I_p$, it follows that $B_H B_H^T = QB_O B_O^T Q^T = QQ^T$. Then, from Corollary 3.1 (i),

$$\begin{aligned}
Var(B_H\hat{\tau}) &= Var(QB_O\hat{\tau}) \\
&= QVar(B_O\hat{\tau})Q^T \\
&= Q(B_O\Lambda B_O^T)^{-1}Q^T \\
&= \{(Q^T)^{-1}(B_O\Lambda B_O^T)Q^{-1}\}^{-1} \\
&= \{(Q^T)^{-1}(Q^{-1}B_H\Lambda(Q^{-1}B_H)^T)Q^{-1}\}^{-1} \\
&= \{(QQ^T)^{-1}(B_H\Lambda B_H^T)(QQ^T)^{-1}\}^{-1} \\
&= \{(B_HB_H^T)^{-1}(B_H\Lambda B_H^T)(B_HB_H^T)^{-1}\}^{-1} \\
&= (B_HB_H^T)(B_H\Lambda B_H^T)^{-1}(B_HB_H^T). \qquad \text{(AD.4)}
\end{aligned}$$

Also, from (AD.4) and Theorem 3.1, it follows that,

$$\begin{aligned}
Var((B_HB_H^T)^{-1}B_H\hat{\tau}) &= (B_HB_H^T)^{-1}Var(B_H\hat{\tau})(B_HB_H^T)^{-1} \\
&= (B_H\Lambda B_H^T)^{-1} \\
&= Var(\hat{\beta}_H).
\end{aligned}$$

□

Proof of Theorem 8.7. Let $B_E = (B_{E(1)}^T \; B_{E(2)}^T \cdots B_{E(k)}^T)^T$, where $B_{E(i)} = \left(\otimes_{i'=1}^{i-1} 1_{v_{i'}}^T\right) \otimes B_e^{(i)} \otimes_{i'=i+1}^{k} 1_{v_{i'}}^T$ and $B_e^{(i)}$ is a $(v_i - 1) \times v_i$ effects coded matrix for ith attribute at v_i levels. Now, taking $B_H = B_E$ in (AD.3), we have, $B_{E(i)} = Q_i B_O$, where $Q = (Q_1^T \; Q_2^T \; \cdots \; Q_k^T)^T$ is a non-singular matrix and Q_i is of order $(v_i - 1) \times \sum_{i=1}^{k}(v_i - 1)$. Also note that $B_O B_O^T = I_{\sum_{i=1}^{k}(v_i-1)}$, $B_E B_E^T = diag(B_{E(1)}B_{E(1)}^T, \ldots, B_{E(k)}B_{E(k)}^T)$ and $B_{E(i)}B_{E(i)}^T = V_i$,

$i = 1, \ldots, k$. Therefore, for $d^* \in \mathcal{D}_{k,m,N}$,

$$\begin{aligned}
\mathcal{I}_{d^*}(\beta_E) &= \mathcal{I}_{d^*}((B_E B_E^T)^{-1} B_E \tau) \\
&= B_E \Lambda_{d^*} B_E^T \\
&= Q B_O \Lambda_{d^*} B_O^T Q^T \\
&= Q(diag(\alpha_1 I_{v_1-1}, \ldots, \alpha_k I_{v_k-1}))Q^T \\
&= diag(\alpha_1 Q_1 Q_1^T, \ldots, \alpha_k Q_k Q_k^T) \\
&= diag(\alpha_1 B_{E(1)} B_{E(1)}^T, \ldots, \alpha_k B_{E(k)} B_{E(k)}^T) \\
&= diag(\alpha_1 V_1, \ldots, \alpha_k V_k).
\end{aligned}$$

Similarly, since $(B_E B_E^T)^{-1} = diag((B_{E(1)} B_{E(1)}^T)^{-1}, \ldots, (B_{E(k)} B_{E(k)}^T)^{-1}) = diag(V_1^{-1}, \ldots, V_k^{-1})$, it follows that $\mathcal{I}_{d^*}(B_E \tau) = (B_E B_E^T)^{-1}(B_E \Lambda_{d^*} B_E^T)(B_E B_E^T)^{-1} = diag(\alpha_1 V_1^{-1}, \ldots, \alpha_k V_k^{-1})$.

$\square$

Mathematical version and proof of Theorem 8.11 as Theorem D.1

Let $Z_{(i,u)} = (-I_u \;\; 1_u \;\; 0_{u \times (v_i - u - 1)})$ for $u = 1, \ldots, v_i - 1$. Then, the $\binom{v_i}{2} \times v_i$ coefficient matrix of all normalized elementary comparisons between the v_i levels of the ith attribute is

$$Z_i = \frac{1}{\sqrt{2}} \begin{pmatrix} Z_{(i)} \\ \bar{Z}_{(i)} \end{pmatrix},$$

where $Z_{(i)} = (Z_{(i,1)}^T \; Z_{(i,2)}^T \; \cdots \; Z_{(i,v_i-2)}^T)^T$ and $\bar{Z}_{(i)} = Z_{(i,v_i-1)}$.

Also, define

$$\begin{aligned}
W_i &= \left(\otimes_{i'=1}^{i-1} \frac{1}{\sqrt{v_{i'}}} 1_{v_{i'}}^T\right) \otimes \frac{1}{\sqrt{2}} Z_{(i)} \otimes_{i'=i+1}^{k} \frac{1}{\sqrt{v_{i'}}} 1_{v_{i'}}^T, \\
\overline{W}_i &= \left(\otimes_{i'=1}^{i-1} \frac{1}{\sqrt{v_{i'}}} 1_{v_{i'}}^T\right) \otimes \frac{1}{\sqrt{2}} \bar{Z}_{(i)} \otimes_{i'=i+1}^{k} \frac{1}{\sqrt{v_{i'}}} 1_{v_{i'}}^T.
\end{aligned}$$

Note that $\overline{W}_i$ is the contrast matrix for comparing level labeled $v_i - 1$ of attribute i to each of the remaining $v_i - 1$ levels labeled $0, 1, \ldots, v_i - 2$. Similarly, W_i is the contrast matrix for pairwise comparisons of the levels labeled $0, 1, \ldots, v_i - 2$ of attribute i. Let $Y_i = (W_i^T \;\; \overline{W}_i^T)^T$. Then, the $t \times L$ matrix $Z = (Y_1^T \; Y_2^T \; \cdots \; Y_k^T)^T$ represents the matrix of normalized elementary comparisons between levels of each and every attribute, with $t = (\sum_{i=1}^k \binom{v_i}{2})$.

The weighted sum of the variances of all normalized elementary comparisons is given by

$$\sum_{i=1}^{k} \frac{v_i - 1}{\binom{v_i}{2}} tr[Var(Y_i\hat{\tau})] = \sum_{i=1}^{k} \frac{2}{v_i} tr[Var(Y_i\hat{\tau})] = tr[\Gamma_O Var(Z\hat{\tau})], \qquad \text{(AD.5)}$$

where $\Gamma_O = diag(\Gamma_{O1}, \Gamma_{O2}, \ldots, \Gamma_{Ok})$ and $\Gamma_{Oi} = (2/v_i)I_{\binom{v_i}{2}}$.

While obtaining $tr[\Gamma_O Var(Z\hat{\tau})]$ in (AD.5), for each attribute i, we account for the contribution of $Var(Y_i\hat{\tau})$ through $\frac{v_i-1}{\binom{v_i}{2}} tr[Var(Y_i\hat{\tau})]$. This ensures providing equal importance to each of the k sets of $\binom{v_i}{2}$ elementary comparisons of which only $v_i - 1$ comparisons are independent, $i = 1, 2, \ldots, k$.

For A-optimality considerations, to bring in parity between different attributes with different number of levels, we normalize each of $(B_E B_E^T)^{-1} B_E$ and B_E. It is easy to see that the normalized form of $(B_E B_E^T)^{-1} B_E$ is $\Upsilon_E (B_E B_E^T)^{-1} B_E$ where $\Upsilon_E = diag(\sqrt{\frac{L}{v_1 - 1}} I_{v_1 - 1},$ $\ldots, \sqrt{\frac{L}{v_k - 1}} I_{v_k - 1})$. Also, $\Upsilon_T B_E$ is the normalized form of B_E where $\Upsilon_T = diag(\sqrt{\frac{v_1}{2L}} I_{v_1 - 1}, \ldots,$ $\sqrt{\frac{v_k}{2L}} I_{v_k - 1})$.

Finally, we introduce the two notations $\Gamma_E = diag(\Gamma_{E1}, \Gamma_{E2}, \ldots, \Gamma_{Ek})$ and $\Gamma_W = diag(\Gamma_{W1}, \Gamma_{W2}, \ldots, \Gamma_{Wk})$, where $\Gamma_{Ei} = \frac{2}{v_i(v_i - 1)} I_{\binom{v_i}{2}}$ and $\Gamma_{Wi} = \frac{1}{(v_i - 1)} I_{\binom{v_i - 1}{2}}$.

Theorem D.1. *Under the main effects model, (i)* $tr[Var(B_O\hat{\tau})] = tr[\Gamma_O Var(Z\hat{\tau})]$,

(ii) $tr[Var(\Upsilon_E (B_E B_E^T)^{-1} B_E \hat{\tau})] = \left(tr[\Gamma_E Var(Z\hat{\tau})] + 2\sum_{i=1}^{k} tr[\Gamma_{Wi} Var(W_i\hat{\tau})] \right)$,

(iii) $tr[Var(\Upsilon_T B_E \hat{\tau})] = \sum_{i=1}^{k} tr[Var(\overline{W}_i \hat{\tau})]$.

Proof. Since the rows of Z spans the same vector space as the rows of B_O, there exists a matrix R_o of order $t \times p_M$ such that $Z = R_o B_O$. Now, $Z = R_o B_O$ implies $Z B_O^T = R_o B_O B_O^T = R_o$. Therefore, $Z = Z B_O^T B_O$. Also, it is easy to see that $B_O Z^T \Gamma_O Z B_O^T = I_{p_M}$, where $p_M = \sum_{i=1}^{k} (v_i - 1)$. Then,

$$\begin{aligned}
tr[\Gamma_O \, Var(Z\hat{\tau})] &= tr[\Gamma_O \, Var(Z B_O^T B_O \hat{\tau})] \\
&= tr[\Gamma_O \, Z B_O^T Var(B_O\hat{\tau}) B_O Z^T] \\
&= tr[B_O Z^T \Gamma_O \, Z B_O^T Var(B_O\hat{\tau})] \\
&= tr[Var(B_O\hat{\tau})].
\end{aligned}$$

There exists a matrix R_e of order $t \times p_M$ such that $Z = R_e \Upsilon_E (B_E B_E^T)^{-1} B_E$. Now, $Z = R_e \Upsilon_E (B_E B_E^T)^{-1} B_E$ implies $Z B_E^T = R_e \Upsilon_E$. Also multiplying by Υ_E^{-1} on both sides, we get $ZB_E^T\Upsilon_E^{-1} = R_e$. Therefore, $Z = ZB_E^T\Upsilon_E^{-1}\Upsilon_E(B_EB_E^T)^{-1}B_E$. Also, it is easy to see that $\Upsilon_E^{-1}B_EZ^T\Gamma_EZB_E^T\Upsilon_E^{-1} = diag(V_1, \ldots, V_k)$ where $V_i = \frac{1}{v_i}(I_{v_i-1} + J_{v_i-1})$. Then,

$$\begin{aligned}
tr[\Gamma_E\,Var(Z\hat{\tau})] &= tr[\Gamma_E\,Var(ZB_E^T\Upsilon_E^{-1}\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&= tr[\Gamma_E\,ZB_E^T\Upsilon_E^{-1}Var(\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})\Upsilon_E^{-1}B_EZ^T] \\
&= tr[\Upsilon_E^{-1}B_EZ^T\Gamma_E\,ZB_E^T\Upsilon_E^{-1}Var(\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&= tr[diag(V_1, \ldots, V_k)Var(\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&= \sum_{i=1}^{k} tr[V_iVar_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&= \sum_{i=1}^{k} \frac{1}{v_i}\Big\{tr[Var_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&\quad + tr[J_{v_i-1}Var_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})]\Big\}, \qquad \text{(AD.6)}
\end{aligned}$$

where $Var_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})$, of order $(v_i-1)\times(v_i-1)$, is the ith diagonal sub-matrix of $Var(\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})$. Let t_i be the ith column of I_k. Then, it is easy to see that $\Upsilon_E^{-1}B_EW_i^T\Gamma_{Wi}W_iB_E^T\Upsilon_E^{-1} = t_it_i^T \otimes G_i$, where $G_i = \frac{1}{2v_i}((v_i-1)I_{v_i-1} - J_{v_i-1})$. Now,

$$\begin{aligned}
tr[\Gamma_{Wi}Var(W_i\hat{\tau})] &= tr[\Gamma_{Wi}Var(W_iB_E^T\Upsilon_E^{-1}\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&= tr[\Gamma_{Wi}W_iB_E^T\Upsilon_E^{-1}Var(\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})\Upsilon_E^{-1}B_EW_i^T] \\
&= tr[\Upsilon_E^{-1}B_EW_i^T\Gamma_{Wi}W_iB_E^T\Upsilon_E^{-1}Var(\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&= tr[\Upsilon_E^{-1}(t_it_i^T \otimes G_i)\Upsilon_E^{-1}Var(\Upsilon_E(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&= \frac{1}{2v_i}\Big\{(v_i-1)tr[Var_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&\quad - tr[J_{v_i-1}Var_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})]\Big\}. \qquad \text{(AD.7)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
2\sum_{i=1}^{k} tr[\Gamma_{Wi}Var(W_i\hat{\tau})] = \sum_{i=1}^{k}\frac{1}{v_i}\Big\{&(v_i-1)tr[Var_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})] \\
&- tr[J_{v_i-1}Var_i(\sqrt{L/(v_i-1)}(B_EB_E^T)^{-1}B_E\hat{\tau})]\Big\} \qquad \text{(AD.8)}
\end{aligned}$$

From (AD.6) and (AD.8), we get,

$$tr[\Gamma_EVar(Z\hat{\tau})] + 2\sum_{i=1}^{k} tr[\Gamma_{Wi}Var(W_i\hat{\tau})] = \sum_{i=1}^{k} tr[Var_i(\sqrt{(v_i-1)/L}(B_EB_E^T)^{-1}B_E\hat{\tau})],$$

or

$$tr[\Gamma_EVar(Z\hat{\tau})] + 2\sum_{i=1}^{k} tr[\Gamma_{Wi}Var(W_i\hat{\tau})] = tr[Var(\Upsilon_E((B_EB_E^T)^{-1}B_E\hat{\tau})].$$

For establishing $tr[Var(\Upsilon_T B_E \hat{\tau})] = \sum_{i=1}^{k} tr[Var(\overline{W}_i \hat{\tau})]$, it is easy to see that $\Upsilon_T B_E \hat{\tau} = (\overline{W}_1^T, \overline{W}_2^T, \ldots, \overline{W}_k^T)^T \hat{\tau}$ and therefore the result follows by taking trace of variance on either sides. $\square$

The relationship between $tr[Var(B_O \hat{\tau})]$ and $tr[\Gamma_n Var(A \Upsilon_E (B_E B_E^T)^{-1} B_E \hat{\tau})]$ is now established where $A = diag(B_e^{(1)T}, \ldots, B_e^{(k)T})$ is a $(\sum v_i) \times p_M$ matrix. We also define $\Gamma_n = diag(\Gamma_{n1}, \Gamma_{n2}, \ldots, \Gamma_{nk})$, where $\Gamma_{ni} = ((v_i - 1)/v_i) I_{v_i}$. Furthermore, for representational ease, let $B_n = A \Upsilon_E (B_E B_E^T)^{-1} B_E$.

Proof of Theorem 8.12. Since the rows of B_n spans the same vector space as the rows of B_O, there exists a matrix R_n of order $(\sum v_i) \times p_M$ such that $B_n = R_n B_O$. Now, $B_n = R_n B_O$ implies $B_n B_O^T = R_n B_O B_O^T = R_n$. Therefore, $B_n = B_n B_O^T B_O$. Also, it is easy to see that $B_O B_n^T \Gamma_n B_n B_O^T = I_{p_M}$, where $p_M = \sum_{i=1}^{k} (v_i - 1)$. Then,

$$\begin{aligned} tr[\Gamma_n \, Var(B_n \hat{\tau})] &= tr[\Gamma_n \, Var(B_n B_O^T B_O \hat{\tau})] \\ &= tr[\Gamma_n \, B_n B_O^T Var(B_O \hat{\tau}) B_O B_n^T] \\ &= tr[B_O B_n^T \Gamma_n \, B_n B_O^T Var(B_O \hat{\tau})] \\ &= tr[Var(B_O \hat{\tau})]. \end{aligned}$$

$\square$

Appendix E

Additional Material for Chapter 9

This appendix provides some additional details for the Chapter 9. It includes: designs discussed in Tables 9.1, 9.2, and Table 9.3 to demonstrate that $UE(s^2)$-optimal designs with better projection properties tend to have smaller values of SS and Q, and a proof of Theorem 9.8.

Designs for Tables 9.1, 9.2, and 9.3

(a). Designs for the case $m = 14$ and $n = 12$ (Table 9.1):

d_1	d_2	d_3
1 1 1 1 1-1 1-1-1 1 1-1 1-1	1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1-1-1-1 1 1 1-1 1-1-1 1	1 1-1-1-1-1 1 1 1 1-1-1-1-1	1 1 1 1-1-1-1-1-1-1-1-1 1 1
-1 1-1-1 1 1 1-1 1-1-1-1 1 1	-1-1 1 1-1-1 1 1-1-1 1 1-1-1	-1-1-1-1 1 1 1 1-1-1-1-1 1 1
1-1 1 1 1-1-1-1-1-1-1 1-1 1	-1-1-1-1 1 1 1 1-1-1-1-1 1 1	-1-1-1-1-1-1-1-1 1 1 1 1 1 1
1-1-1-1-1-1 1 1-1-1-1 1 1-1	-1-1-1-1 1 1-1-1 1 1 1 1-1-1	1 1-1-1 1 1-1-1 1 1-1-1-1-1
-1-1 1 1-1 1 1 1-1 1-1-1 1 1	-1-1 1 1-1-1-1-1 1 1-1-1 1 1	1 1-1-1-1-1 1 1-1-1 1 1-1-1
-1 1-1 1 1 1-1 1-1-1 1-1-1-1	1 1-1-1-1-1-1-1-1-1 1 1 1 1	-1-1 1 1 1 1-1-1-1-1 1 1-1-1
-1-1-1 1 1-1-1 1 1 1 1 1 1 1	1 1 1 1 1 1-1-1-1-1-1-1-1-1	-1-1 1 1-1-1 1 1 1 1-1-1-1-1
-1 1-1 1-1-1 1-1 1 1-1 1-1-1	1-1 1-1 1-1 1-1 1-1 1-1 1-1	1-1 1-1 1-1 1-1 1-1 1-1 1-1
-1-1 1-1-1 1-1-1 1-1 1 1 1-1	1-1-1 1-1 1 1-1 1-1-1 1-1 1	1-1 1-1-1 1-1 1-1 1-1 1 1-1
1-1 1-1 1 1-1 1 1 1-1-1-1-1	-1 1 1-1-1 1 1-1-1 1 1-1-1 1	-1 1-1 1 1-1 1-1-1 1-1 1 1-1
1 1-1-1-1 1-1-1-1 1 1 1-1 1	-1 1-1 1 1-1-1 1 1-1 1-1-1 1	-1 1-1 1-1 1-1 1 1-1 1-1 1-1

d_4	d_5	d_6
1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1-1-1-1-1 1 1 1-1-1-1-1 1	1-1-1-1-1 1 1 1 1-1-1-1-1 1	1 1 1-1-1-1-1-1-1-1-1 1 1 1
-1-1 1 1-1-1 1-1-1 1 1-1-1 1	-1 1 1-1-1 1 1-1-1 1 1-1-1 1	-1-1-1 1 1 1 1-1-1-1-1 1 1 1
-1-1-1-1 1 1 1-1-1-1-1 1 1 1	-1-1-1 1 1 1 1-1-1-1-1 1 1 1	-1-1-1-1-1-1-1 1 1 1 1 1 1 1
-1-1-1-1 1 1-1 1 1 1 1-1-1 1	-1-1-1 1 1-1-1 1 1 1 1-1-1 1	1-1-1 1 1-1-1 1 1-1-1 1-1-1
-1-1 1 1-1-1-1 1 1-1-1 1 1 1	-1 1 1-1-1-1-1 1 1-1-1 1 1 1	1-1-1-1-1 1 1-1-1 1 1 1-1-1
1 1-1-1-1-1-1-1-1 1 1 1 1 1	1-1-1-1-1-1-1-1-1 1 1 1 1 1	-1 1 1 1 1-1-1-1-1 1 1 1-1-1
1 1 1 1 1 1-1-1-1-1-1-1-1 1	1 1 1 1 1-1-1-1-1-1-1-1-1 1	-1 1 1-1-1 1 1 1 1-1-1 1-1-1
1-1 1-1 1-1-1 1-1 1-1 1-1-1	-1 1-1 1-1 1-1 1-1 1-1 1-1-1	-1 1-1 1-1 1-1 1-1 1-1-1 1-1
1-1-1 1-1 1-1 1-1-1 1-1 1-1	-1-1 1-1 1 1-1 1-1-1 1-1 1-1	-1 1-1-1 1-1 1-1 1-1 1-1 1-1
-1 1 1-1-1 1-1-1 1 1-1-1 1-1	1 1-1-1 1 1-1-1 1 1-1-1 1-1	1-1 1 1-1 1-1-1 1-1 1-1 1-1
-1 1-1 1 1-1 1 1-1 1-1-1 1-1	1-1 1 1-1-1 1 1-1 1-1-1 1-1	1-1 1-1 1-1 1 1-1 1-1-1 1-1

(b). Designs for the case $m = 16$ and $n = 10$ (Table 9.2):

d_1

```
 1-1-1-1 1-1 1 1-1-1 1 1-1-1-1 1
-1 1-1 1 1 1-1 1 1-1-1-1-1-1-1-1
 1-1 1 1 1 1-1-1 1 1 1 1-1-1 1 1
-1 1 1-1 1 1-1-1-1-1-1 1 1 1-1 1
 1-1 1-1 1-1-1 1-1 1-1-1 1 1 1-1
-1 1 1-1-1-1 1-1 1-1-1 1-1-1 1-1
 1 1 1 1-1-1 1-1-1 1 1-1-1 1-1-1
-1-1-1 1-1-1-1 1 1 1 1 1 1 1-1-1
-1-1-1 1-1 1 1-1-1 1-1-1 1-1 1 1
 1 1-1-1-1 1 1 1 1-1 1-1 1 1 1 1
```

d_2

```
-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1
-1 1-1 1-1 1-1 1-1 1-1 1-1 1-1 1
 1-1 1 1-1-1 1 1-1-1 1 1-1-1 1 1
-1-1-1-1 1 1 1 1-1-1-1-1 1 1 1 1
 1 1-1 1 1-1 1-1-1 1-1 1 1-1 1-1
 1-1 1 1 1 1-1-1-1-1 1 1 1 1-1-1
 1-1-1-1-1-1-1-1 1 1 1 1 1 1 1 1
 1-1 1 1-1 1-1 1 1 1-1-1 1-1 1-1
-1 1-1 1-1 1 1-1 1-1 1-1 1-1-1 1
-1 1 1-1 1 1-1-1 1-1-1 1-1-1 1 1
```

d_3

```
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1-1-1-1-1-1-1-1-1
-1 1 1 1-1-1-1-1 1 1 1 1-1-1-1-1
 1 1 1 1-1-1-1-1-1-1-1-1 1 1 1 1
 1 1-1-1 1 1-1-1 1 1-1-1 1 1-1-1
 1 1-1-1 1 1-1-1-1-1 1 1-1-1 1 1
-1-1 1-1 1-1 1-1 1-1 1-1 1-1 1-1
-1-1 1-1-1 1-1 1 1-1 1-1-1 1-1 1
-1-1-1 1 1-1-1 1 1-1-1 1 1-1-1 1
-1-1-1 1-1 1 1-1-1 1 1-1 1-1-1 1
```

d_4

```
 1 1 1 1 1 1-1 1 1 1 1 1 1 1 1 1
 1 1-1-1-1-1-1 1 1 1-1-1-1-1 1 1
-1-1 1 1-1-1 1 1-1-1 1 1-1-1 1 1
-1-1-1-1 1 1 1 1-1-1-1-1 1 1 1 1
 1-1 1-1 1-1-1-1 1-1 1-1 1-1 1-1
 1-1-1 1-1 1 1-1 1-1-1 1-1 1 1-1
-1 1 1-1-1 1-1-1-1 1 1-1-1 1 1-1
-1 1-1 1-1 1-1-1 1-1 1-1 1-1-1 1
-1 1 1-1 1-1 1-1 1-1-1 1-1 1-1 1
 1-1-1 1 1-1 1-1-1 1 1-1-1 1-1 1
```

d_5

```
 1 1 1 1 1 1-1 1 1 1 1 1 1 1 1 1
 1 1-1-1-1-1-1 1 1 1-1-1-1-1 1 1
-1-1 1 1-1-1 1 1-1-1 1 1-1-1 1 1
-1-1-1-1 1 1 1 1-1-1-1-1 1 1 1 1
 1-1 1-1 1-1 1-1 1-1 1-1 1-1 1-1
 1-1-1 1-1 1-1-1 1-1-1 1-1 1 1-1
-1 1 1-1-1 1-1-1-1 1 1-1-1 1 1-1
-1 1-1 1-1 1 1-1 1-1 1-1 1-1-1 1
-1 1 1-1 1-1 1-1 1-1-1 1-1 1-1 1
 1-1-1 1 1-1-1-1-1 1 1-1-1 1-1 1
```

d_6

```
 1 1 1 1 1 1-1 1 1 1 1 1 1 1 1 1
 1 1-1-1-1-1-1 1 1 1-1-1-1-1 1 1
-1-1 1 1-1-1 1 1-1-1 1 1-1-1 1 1
-1-1-1-1 1 1 1 1-1-1-1-1 1 1 1 1
 1-1 1-1 1-1 1-1 1-1 1-1 1-1 1-1
 1-1-1 1-1 1-1-1 1-1-1 1-1 1 1-1
-1 1 1-1-1 1 1-1-1 1 1-1-1 1 1-1
-1 1-1 1-1 1-1-1 1-1 1-1 1-1-1 1
-1 1 1-1 1-1 1-1 1-1-1 1-1 1-1 1
 1-1-1 1 1-1-1-1-1 1 1-1-1 1-1 1
```

d_7

```
 1 1 1 1 1 1 1 1-1 1 1 1 1 1 1 1
-1-1-1-1-1-1-1-1-1 1 1 1 1 1 1 1
 1 1 1 1-1-1-1-1 1 1 1-1-1-1-1 1
-1-1-1-1 1 1 1 1 1 1 1-1-1-1-1 1
 1 1-1-1 1 1-1-1-1-1-1 1 1-1-1 1
-1-1 1 1-1-1 1 1 1-1-1 1 1-1-1 1
 1 1-1-1-1-1 1 1 1-1-1-1-1 1 1 1
-1-1 1 1 1 1-1-1 1-1-1-1-1 1 1 1
 1-1 1-1 1-1 1-1-1 1-1 1-1 1-1-1
-1 1-1 1-1 1-1 1-1 1-1 1-1 1-1-1
```

(c). Designs for the case $m = 22$ and $n = 12$ (Table 9.3):

d_1

```
-1 -1 -1 -1  1 -1  1 -1 -1 -1  1  1  1  1 -1 -1 -1  1  1  1  1  1
 1 -1  1  1 -1 -1 -1 -1 -1  1 -1  1  1  1  1  1 -1 -1 -1  1  1 -1
-1  1 -1 -1  1 -1  1  1  1 -1  1 -1 -1 -1  1  1 -1  1 -1 -1 -1 -1
-1 -1  1  1 -1  1  1  1  1 -1 -1 -1  1 -1  1 -1  1  1  1  1 -1 -1
 1  1 -1 -1 -1  1  1  1  1  1  1  1 -1  1  1 -1  1 -1 -1 -1 -1  1
 1  1  1  1 -1 -1 -1  1  1 -1 -1 -1  1 -1 -1  1 -1  1  1  1  1  1
-1 -1  1  1 -1  1 -1 -1 -1 -1 -1 -1  1 -1 -1 -1  1 -1 -1  1  1  1
 1 -1  1 -1  1  1  1  1 -1 -1 -1  1  1  1 -1  1  1 -1 -1 -1 -1  1
-1  1  1 -1  1 -1 -1 -1 -1  1  1 -1 -1 -1 -1  1 -1  1 -1 -1 -1  1
 1 -1 -1  1 -1  1  1  1  1  1  1  1 -1  1  1 -1 -1 -1 -1 -1  1 -1
 1 -1 -1 -1  1 -1 -1  1  1  1 -1 -1  1 -1 -1  1 -1  1  1  1 -1  1
 1  1 -1  1  1 -1 -1 -1 -1  1  1 -1 -1  1  1 -1  1  1  1  1 -1 -1
```

d_2

```
 1 -1 -1 -1 -1  1 -1  1  1  1  1  1  1 -1 -1  1 -1 -1  1 -1  1  1
 1 -1  1  1 -1 -1  1  1  1  1 -1 -1 -1  1  1 -1  1  1  1 -1  1 -1
-1  1  1  1  1 -1 -1 -1 -1  1  1 -1  1 -1  1 -1  1  1 -1  1  1  1
-1 -1 -1 -1 -1  1 -1 -1 -1 -1 -1 -1 -1  1  1  1 -1 -1  1  1  1 -1
 1 -1 -1  1 -1 -1  1 -1  1  1 -1 -1  1  1 -1  1 -1  1  1  1 -1  1
-1  1  1 -1  1  1  1 -1  1 -1  1  1 -1  1 -1  1 -1 -1  1  1 -1 -1
 1 -1  1 -1  1  1 -1  1  1  1 -1  1  1 -1 -1 -1  1  1  1 -1 -1 -1
-1  1  1  1 -1 -1  1  1  1 -1 -1 -1  1 -1  1  1  1 -1  1 -1 -1 -1
 1  1 -1  1 -1  1  1  1 -1  1 -1  1 -1 -1 -1 -1  1 -1  1  1  1  1
-1  1 -1  1 -1 -1  1  1 -1 -1  1  1 -1  1  1 -1 -1  1  1  1  1 -1
 1 -1 -1 -1  1  1  1 -1 -1 -1  1 -1  1  1  1  1 -1 -1 -1 -1  1  1
 1 -1  1  1  1 -1  1 -1 -1 -1 -1 -1 -1 -1 -1 -1  1 -1 -1 -1 -1 -1
```

d_3

1	-1	-1	1	1	1	-1	-1	-1	1	1	-1	1	-1	1	1	1	1	1	1	1	1
1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	-1	-1	1	-1	-1
-1	1	1	-1	1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	1	1	-1
-1	1	1	1	-1	1	1	-1	-1	-1	1	1	-1	1	1	-1	1	-1	-1	1	-1	1
1	-1	1	1	1	1	1	1	1	1	1	-1	1	1	-1	1	-1	-1	1	1	1	1
1	-1	-1	-1	-1	-1	-1	1	-1	-1	-1	1	-1	1	-1	-1	1	1	1	1	-1	-1
-1	-1	1	1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	1	-1
-1	1	1	-1	1	-1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	-1	-1	-1
1	1	-1	1	-1	-1	1	1	1	1	-1	1	-1	-1	1	1	1	-1	-1	-1	1	1
-1	1	-1	-1	1	1	1	1	-1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1
1	-1	-1	-1	-1	1	1	-1	1	-1	1	-1	1	1	-1	1	-1	-1	-1	-1	1	-1
1	-1	-1	1	1	-1	1	-1	-1	-1	-1	-1	1	1	1	-1	1	1	-1	-1	-1	1

d_4

1	-1	-1	-1	-1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	1	-1	-1	1	-1
1	-1	-1	-1	1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	-1	-1	-1	-1	-1	1
1	1	1	1	1	1	1	1	1	1	-1	-1	-1	1	1	-1	-1	1	-1	1	-1	1
1	1	1	1	1	-1	-1	-1	1	-1	1	1	1	-1	1	-1	-1	1	1	-1	1	-1
-1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	-1
-1	-1	-1	1	-1	-1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	-1	-1	-1	-1
-1	1	1	-1	-1	1	-1	1	-1	1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1
1	-1	1	-1	-1	1	1	-1	1	-1	-1	-1	1	-1	-1	1	1	-1	-1	1	1	-1
-1	1	1	1	1	1	1	1	1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	1
-1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	1	-1	-1	1	1	1	-1
1	-1	-1	-1	-1	1	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1
1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	1	1	1	-1	-1	-1	1	-1	1

d_5

1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	-1	1	1	-1	1	-1
1	-1	-1	-1	-1	-1	1	1	-1	1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	1	-1
1	1	1	-1	1	-1	-1	-1	1	1	1	1	-1	1	1	1	1	-1	1	1	-1	-1
1	1	1	1	-1	1	1	1	1	1	-1	-1	1	-1	-1	-1	-1	-1	-1	1	-1	-1
-1	-1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	1	-1	1	1	1	1	1	1
-1	-1	-1	-1	1	-1	-1	1	1	-1	-1	-1	-1	1	-1	1	1	1	-1	1	1	-1
-1	1	1	1	1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	-1	-1	-1	-1
1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	-1	1	-1	1	1
-1	1	1	-1	1	1	1	1	1	1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1
-1	1	-1	1	1	1	-1	1	1	-1	1	1	-1	-1	-1	-1	-1	1	1	-1	1	-1
1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	1	1	1	-1	1	-1	-1	-1	1	1	1
1	-1	1	1	-1	-1	1	-1	1	1	-1	1	1	1	1	-1	1	1	1	1	1	-1

Proof of Theorem 9.8. We first mention the sharpest available lower bounds for $E_d(s^2)$ as per Das et al. (2008).

When n is even, let $m = q(n-1) + r$ (q positive, $-\frac{n}{2} < r < \frac{n}{2}$). Then,

(1) When $n \equiv 0 (\text{mod } 4)$,

$$E_d(s^2) \geq L(m,n) = \frac{n^2(m-n+1)}{(n-1)(m-1)} + \frac{n}{m(m-1)} \left\{ D - \frac{r^2}{n-1} \right\}, \tag{AE.1}$$

where

$$D = \begin{cases} n + 2|r| - 3 & \text{for } |r| \equiv 1 \ (\text{mod } 4) \\ 2n - 4 & \text{for } |r| \equiv 2 \ (\text{mod } 4) \\ n + 2|r| + 1 & \text{for } |r| \equiv 3 \ (\text{mod } 4) \\ 4|r| & \text{for } |r| \equiv 0 \ (\text{mod } 4). \end{cases}$$

(2) When $n \equiv 2$ (mod 4),

$$E_d(s^2) \geq L(m,n) = \max \left\{ \frac{n^2(m-n+1)}{(n-1)(m-1)} + \frac{n}{m(m-1)} \left\{ D - \frac{r^2}{n-1} \right\}, 4 \right\}, \tag{AE.2}$$

where

$$D = \begin{cases} n+2|r|-3+x/n & \text{for } |r| \equiv 1 \pmod 4 \text{ and } q \text{ even} \\ 2|r|-8|r|/n+n-16/n+9 & \text{for } |r| \equiv 1 \pmod 4 \text{ and } q \text{ odd} \\ 2n-4+8/n & \text{for } |r| \equiv 2 \pmod 4 \text{ and } q \text{ even} \\ 4|r|-8|r|/n-8/n+8 & \text{for } |r| \equiv 2 \pmod 4 \text{ and } q \text{ odd} \\ n+2|r|+1 & \text{for } |r| \equiv 3 \pmod 4 \text{ and } q \text{ even} \\ 2|r|+n+8/n-3 & \text{for } |r| \equiv 3 \pmod 4 \text{ and } q \text{ odd} \\ 4|r| & \text{for } |r| \equiv 0 \pmod 4 \text{ and } q \text{ even} \\ 2n-4+x/n & \text{for } |r| \equiv 0 \pmod 4 \text{ and } q \text{ odd.} \end{cases}$$

and $x = 32$ if $\left\{\frac{m-1-2i}{4} + \lfloor \frac{m+(1+2i)(n-1)}{4(n-1)} \rfloor \right\} \equiv (1-i) \pmod 2$, for $i = 0$ or 1; else $x = 0$. Here $\lfloor z \rfloor$ denotes the largest integer less than or equal to z.

Jones and Majumdar (2014) obtained the attainable lower bounds for $UE_d(s^2)$ as below.

$$\min_{d\in\mathcal{D}_U(m,n)} UE_d(s^2) = \frac{n(m+1-n)}{m} + \frac{nB}{m(m+1)}, \tag{AE.3}$$

where

$$B = \begin{cases} 0 & \text{for } m+1 \equiv 0 \pmod 4, \\ 2(n-2) & \text{for } m+1 \equiv 2 \pmod 4 \text{ and even } n, \\ 2\{(n-2)+1/n\} & \text{for } m+1 \equiv 2 \pmod 4 \text{ and odd } n, \\ n-1 & \text{for } m+1 \equiv 1 \pmod 4 \text{ or } m+1 \equiv 3 \pmod 4. \end{cases}$$

We now show that

$$\frac{m-1}{m+1}L(m,n) - \min_{d\in\mathcal{D}_U(m,n)} UE_d(s^2) = \frac{nI}{m(n-1)(m+1)} \geq 0,$$

where using (AE.1), (AE.2) and (AE.3),

$$I = A + (n-1)(D-B) \tag{AE.4}$$

with $A = (m+1-n)^2 - r^2$. In other words, we shall identify the cases where $I = 0$ and show that $I > 0$ for the remaining cases .

Additionally, we define $W = m - (n-1)\lfloor \frac{m}{n-1} \rfloor$, $0 \leq W \leq n-2$. Then, since $m = q(n-1) + r$, $-n/2 < r < n/2$,

$$r = \begin{cases} W & \text{for } 0 \leq W < \frac{n}{2}, \quad \text{or equivalently } 0 \leq r < \frac{n}{2}, \\ W-n+1 & \text{for } \frac{n}{2} \leq W \leq n-2, \quad \text{or equivalently } -\frac{n}{2} < r < 0 \end{cases} \tag{AE.5}$$

Using (AE.5), A can also be written in terms of n, q, r and W as below,

$$A = \begin{cases} (n-1)^2(q-1)^2 + 2r(n-1)(q-1), & \text{for } 0 \leq r < \frac{n}{2}, \\ ((q-2)^2-1)(n-1)^2 + 2W(n-1)(q-1), & \text{for } -\frac{n}{2} < r < 0. \end{cases} \tag{AE.6}$$

From (AE.6), it is clear that $A \geq 0$ since

$$\begin{array}{lll} \text{a) } A > 0, & \text{for } \ q \geq 2, & \\ \text{b) } A = 0, & \text{for } q = 1, 0 \leq r < \frac{n}{2}, & \\ \text{c) } A = -(n-1)^2 & \text{for } q = 1, -\frac{n}{2} < r < 0, & \text{which is not possible since then } m < n. \end{array} \tag{AE.7}$$

We take up the following exhaustive cases to arrive at the required conditions.

For $n \equiv 0 \pmod 4$

Case A. $|r| \equiv 0 \pmod 4$

Case A1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

First note that in this case the situation $q = 1$ does not arise. Now from (AE.1) and (AE.3), we get $D - B = 4|r| - n + 1$. Therefore, from (AE.4), $I = A + (n-1)(4|r| - n + 1)$ which, by (AE.6), reduces to

$$I = \begin{cases} (n-1)^2((q-1)^2 - 1) + 2r(n-1)(q+1) & \text{if } 0 \leq r < \frac{n}{2}, \\ (n-1)^2((q-2)^2 + 2) + 2W(n-1)(q-3) & \text{if } -\frac{n}{2} < r < 0. \end{cases}$$

This implies that for $0 \leq r < \frac{n}{2}$, $I > 0$ when $q \geq 3$ and when $q = 2, r \neq 0$. Furthermore, $I = 0$ (equality holds) for $r = 0, q = 2$, i.e., for $m = 2(n-1)$. Similarly for $-\frac{n}{2} < r < 0$, $I > 0$ for $q \geq 2$.

Case A2. $m \equiv 1 \pmod 4$

First note that in this case the situations $q = 1, 2$ do not arise. Now from (AE.1) and (AE.3), we get $D - B = 2(2|r| - n + 2)$. Therefore, from (AE.4), $I = A + 2(n-1)(2|r| - n + 2)$ which, by (AE.6), reduces to

$$I = \begin{cases} (n-1)^2((q-1)^2 - 2) + 2(n-1)(r(q+1) + 1) & \text{if } 0 \leq r < \frac{n}{2}, \\ (n-1)^2((q-2)^2 - 1) + 2(n-1)(W(q-3) + n) & \text{if } -\frac{n}{2} < r < 0. \end{cases}$$

This implies that $I > 0$ when $q \geq 3$.

Case A3. $m \equiv 3 \pmod 4$

First note that in this case the situation $q = 1, r = 0$ does not arise. Now from (AE.1) and (AE.3), we get $D - B = 4|r|$. Therefore, from (AE.4), $I = A + 4(n-1)|r| > 0$ since from (AE.7) $A = 0$ only when $q = 1, 0 \leq r < \frac{n}{2}$.

Case B. $|r| \equiv 1 \pmod 4$

Case B1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

First note that in this case the situation $q = 1, r = -1$ does not arise. Now from (AE.1) and (AE.3), we get $D - B = 2(|r| - 1)$. Therefore, from (AE.4), $I = A + 2(n - 1)(|r| - 1) > 0$ except when $A = 0, r = 1$. Thus, from (AE.7), $I = 0$ (equality holds) for $r = 1, q = 1$, i.e., for $m = n$.

Case B2. $m \equiv 1 \pmod 4$

First note that in this case the situation $q = 1$ does not arise. Now from (AE.1) and (AE.3), we get $D - B = 2|r| - (n-1)$. Therefore, from (AE.4), $I = A + 2(n-1)|r| - (n-1)^2$ which, by (AE.6), reduces to

$$I = \begin{cases} (n-1)^2((q-1)^2 - 1) + 2qr(n-1) & \text{if } 0 \leq r < \frac{n}{2}, \\ (q-2)^2(n-1)^2 + 2W(n-1)(q-2) & \text{if } -\frac{n}{2} < r < 0. \end{cases}$$

This implies that $I > 0$ when $q \geq 2, 0 \leq r < \frac{n}{2}$ and when $q \geq 3, -\frac{n}{2} < r < 0$. Furthermore, $I = 0$ for $-\frac{n}{2} < r < 0, q = 2$, i.e., for $m = 2(n-1) + r, -n/2 < r < 0$. Let $m = 4s + 1$, where s is an integer. Then $-n/2 < m - 2(n-1) \leq -1$ and substituting $m = 4s + 1$ yields $3(n-2)/8 < s \leq (n-2)/2$. Therefore, $I = 0$ (equality holds) when $m = 4s + 1, 3(n-2)/8 < s \leq (n-2)/2, n > 2$.

Case B3. $m \equiv 3 \pmod 4$

From (AE.1) and (AE.3), we get $D - B = (n + 2|r| - 3)$. Therefore, from (AE.4), $I = A + (n + 2|r| - 3)(n - 1)$, which is greater than 0 since the second term is greater than 0.

Case C. $|r| \equiv 2 \pmod 4$

Case C1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

From (AE.1) and (AE.3), we get $D - B = n - 3$. Therefore, from (AE.4), $I = A + (n-1)(n-3)$, which is greater than 0 since the second term is greater than 0.

Case C2. $m \equiv 1 \pmod 4$

From (AE.1) and (AE.3), we get $D - B = 0$. Therefore, from (AE.4) and (AE.7), $I = A > 0$ except when $q = 1, 0 \leq r < \frac{n}{2}$. Thus, $I = 0$ for $0 \leq r < \frac{n}{2}, q = 1$, i.e., for

$m = (n-1)+r$. Let $m = 4s+1$, where s is an integer. Then $2 \le m-n+1 < n/2$ and substituting $m = 4s+1$ yields $n/4 \le s < (3n-4)/8$. Therefore, $I = 0$ (equality holds) when $n/4 \le s < (3n-4)/8, n > 4$.

Case C3. $m \equiv 3 \pmod 4$

From (AE.1) and (AE.3), we get $D - B = (2n-4)$. Therefore, from (AE.4), $I = A + 2(n-2)(n-1)$, which is greater than 0 since the second term is greater than 0.

Case D. $|r| \equiv 3 \pmod 4$

Case D1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

From (AE.1) and (AE.3), we get $D - B = 2(|r|+1)$. Therefore, from (AE.4), $I = A + 2(n-1)(|r|+1)$, which is greater than 0 since the second term is greater than 0.

Case D2. $m \equiv 1 \pmod 4$

First note that in this case the situation $q = 1$ does not arise. Now from (AE.1) and (AE.3), we get $D - B = 2|r| - (n-5)$. Therefore, from (AE.4), $I = A + 2(n-1)|r| - (n-1)(n-5)$ which, by (AE.6), reduces to

$$I = \begin{cases} (n-1)^2 q(q-2) + 2(n-1)(rq+2) & \text{if } 0 \le r < \frac{n}{2}, \\ (n-1)^2 (q-2)^2 + 2W(n-1)(q-2) + 4(n-1) & \text{if } -\frac{n}{2} < r < 0. \end{cases}$$

This implies that $I > 0$ when $q \ge 2$.

Case D3. $m \equiv 3 \pmod 4$

From (AE.1) and (AE.3), we get $D - B = (n + 2|r| + 1)$. Therefore, from (AE.4), $I = A + 2(n-1)(n+2|r|+1)$, which is greater than 0 since the second term is greater than 0.

For $n \equiv 2 \pmod 4$

Case A. $|r| \equiv 0 \pmod 4$

Case A1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

First note that in this case the situation $q = 1$ does not arise. Now from (AE.2) and (AE.3), we get $D - B = 4|r| - n + 1$. Therefore, from (AE.4), $I = A + (n-1)(4|r| - n + 1)$, which is the same as the corresponding expression for the case $n \equiv 0 \pmod 4$. Therefore, $I = 0$ (equality holds) for $r = 0, q = 2$, i.e., for $m = 2(n-1)$.

Case A2. $m \equiv 1 \pmod 4$

Now from (AE.2) and (AE.3), we get $D - B = x/n$. Therefore, from (AE.4), $I = A + (n-1)(x/n)$. Thus, $I = 0$ when $A = 0$ and $x = 0$ or from (AE.7), $I = 0$ when $q = 1, r > 0, x = 0$, i.e., $m = (n-1) + r, m \equiv 1 \pmod 8$. Let $m = 8s + 1$, where s is an integer. Then $4 \leq m - n + 1 < n/2$ and substituting $m = 8s + 1$ yields $(n+2)/8 \leq s < (3n-4)/16$. Therefore, $I = 0$ (equality holds) when $m = 8s+1, (n+2)/8 \leq s < (3n-4)/16, n > 8$.

Case A3. $m \equiv 3 \pmod 4$

From (AE.2) and (AE.3), we get $D - B = (2n - 4 + x/n)$. Therefore, from (AE.4), $I = A + (n-1)(2n - 4 + x/n)$, which is greater than 0 since the second term is greater than 0.

Case B. $|r| \equiv 1 \pmod 4$

Case B1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

First note that in this case the situation $q = 1, -\frac{n}{2} < r < 0$ does not arise. Now from (AE.2) and (AE.3), we get $D - B = 2(|r| - 4|r|/n - 8/n + 5)$. Therefore, from (AE.4), $I = A + 2(n-1)(|r| - 4|r|/n - 8/n + 5)$, which is greater than 0 since the second term is greater than 0.

Case B2. $m \equiv 1 \pmod 4$

First note that in this case the situation $q = 1$ does not arise. Now from (AE.2) and (AE.3), we get $D - B = 2|r|(n-1) - (n-1)^2 + (x/n)(n-1)$. Therefore, from (AE.4), $I = A + 2|r|(n-1) - (n-1)^2 + (x/n)(n-1)$, which is same as the corresponding expression for the case $n \equiv 0 \pmod 4$ except the addition of the last term. Therefore, $I = 0$ for $-\frac{n}{2} < r < 0, q = 2, x = 0$, i.e., for $m = 2(n-1) + r, -n/2 < r < 0, m \equiv 1 \pmod 8$. Let $m = 8s + 1$, where s is an integer. Then $-n/2 < m - 2(n-1) \leq -1$ and substituting $m = 8s + 1$ yields $3(n-2)/16 < s \leq (n-2)/4$. Therefore, $I = 0$ (equality holds) for $m = 8s + 1, 3(n-2)/16 < s \leq (n-2)/4, n > 2$.

Case B3. $m \equiv 3 \pmod 4$

From (AE.2) and (AE.3), we get $D - B = (n + 2|r| - 3 + x/n)$. Therefore, from (AE.4), $I = A + (n-1)(n+2|r|-3+x/n)$, which is greater than 0 since the second term is greater than 0.

Case C. $|r| \equiv 2 \pmod 4$

Case C1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

From (AE.2) and (AE.3), we get $D - B = n - 3 + 8/n$. Therefore, from (AE.4), $I = A + (n-1)(n-3) + (8/n)(n-1)$, which is greater than 0 since the last two terms are greater than 0.

Case C2. $m \equiv 1 \pmod 4$

First note that in this case the situations $q = 1, 2$ do not arise. Now from (AE.2) and (AE.3), we get $D - B = 2(2|r| + 4 - 4|r|/n - 4/n - n + 2)$. Therefore, from (AE.4), $I = A + (n-1)(4|r| + 8 - 8|r|/n - 8/n) - 2(n-1)(n-2)$ which, by (AE.6), reduces to

$$I = \begin{cases} (n-1)((n-1)(q-1)^2 - 2n) + (2/n)(n-1)(r(n(q+1)-4) + 2(3n-2) & \text{if } 0 \le r < \frac{n}{2}, \\ ((q-2)^2 - 1)(n-1)^2 + 2(n-1)(W((q-3) + 4/n) + n) & \text{if } -\frac{n}{2} < r < 0. \end{cases}$$

This implies that $I > 0$ when $q \ge 3$.

Case C3. $m \equiv 3 \pmod 4$

From (AE.2) and (AE.3), we get $D - B = (4|r| + 8 - 8|r|/n - 8/n)$. Therefore, from (AE.4), $I = A + (4|r| + 8 - 8|r|/n - 8/n)$, which is greater than 0 since the second term is greater than 0.

Case D. $|r| \equiv 3 \pmod 4$

Case D1. $m \equiv 0 \pmod 4$ or $m \equiv 2 \pmod 4$

From (AE.2) and (AE.3), we get $D - B = 2(|r| - 1) + 8/n$. Therefore, from (AE.4), $I = A + 2(|r| - 1) + 8/n$, which is greater than 0 since the last term is greater than 0.

Case D2. $m \equiv 1 \pmod 4$

First note that in this case the situation $q = 1$ does not arise. Now from (AE.2) and (AE.3), we get $D - B = 2|r| - (n-5)$. Therefore, from (AE.4), $I = A + 2|r|(n-1) -$

$(n-1)(n+5)$, which is same as the corresponding expression for the case $n \equiv 0 \pmod 4$. Therefore, I is greater than 0.

Case D3. $m \equiv 3 \pmod 4$

From (AE.2) and (AE.3), we get $D - B = (n + 2|r| + 1)$. Therefore, from (AE.4), $I = A + 2(n-1)(n + 2|r| + 1)$, which is greater than 0 since the second term is greater than 0. $\square$

Appendix F

Additional Material for Chapter 11

This appendix provides extra details for Chapter 11. Non-GYD PGYD for $(v, k, b) = (8, 20, 50)$ is the transpose of the following matrix. The twelve 8×8 grids are latin squares of order 8. The blocks in $\mathcal{D}_C$ cover each pair of treatments eleven times, except for the pairs $\{1,4\}, \{2,3\}, \{5,8\}, \{6,7\}$ which are covered nine times. The blocks in $\mathcal{D}_R$ cover each of the pairs $\{1,4\}, \{2,3\}, \{5,8\}, \{6,7\}$ five times.

1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	7	8
2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	7	8	3	4
3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6
4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	5	6	1	2
5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	1	3	6	8
6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	8	2	4
7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	2	4	5	7
8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	5	7	1	3
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	4	6	7
2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	6	7	2	3
3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	2	3	5	8
4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	5	8	1	4
5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	2	1	8	7
6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	8	7	4	3
7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	4	3	6	5
8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	6	5	2	1
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	3	1	8	6
2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	8	6	4	2
3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	4	2	7	5
4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	7	5	3	1
5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	4	1	7	6
6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	7	6	3	2
7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	3	2	8	5
8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	5	4	1
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4
2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	1
3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	6	7	5	8
4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	7	6	8	5
5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	3	8	1	6
6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	4	5	2	7
7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	5	1	7	3
8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	4	6	2
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4
2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	7	3	2	6
3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	6	5
4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	6	7	1
5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	2	8	5	3
6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	5	1	4	8
7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	8	7	1	2
8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	6	5	8	7
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	5	6
2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	4	3	8	7
3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	8	7	1	2
4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	5	6	4	3
5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	3	1	7	5
6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	2	4	6	8
7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	6	8	3	1
8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	7	5	2	4
5	8	6	7	1	4	2	3	5	8	6	7	1	4	2	3	5	1	6	2
8	5	7	6	4	1	3	2	8	5	7	6	4	1	3	2	8	4	7	3

Non-GYD PGYD for $(v, k, b) = (9, 24, 48)$ is the transpose of the following matrix. The ten 9×9 grids are latin squares of order 9. The blocks in $\mathcal{D}_R$ form four copies of one parallel class of an affine plane of order 9 and two copies of two other parallel classes. The blocks in $\mathcal{D}_C$ form the complements of five copies of the parallel class not occurring in $\mathcal{D}_R$, four copies of the parallel classes occurring twice in $\mathcal{D}_R$, and three copies of the parallel class occurring four times in $\mathcal{D}_R$.

1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	4	6	5	7	8	9
2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	7	8	9	1	2	3
3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	1	3	2	4	5	6
4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	6	9	8	2	3	5
5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	9	1	4	3	6	7
6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	8	7	1	5	4	2
7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	3	2	6	8	7	4
8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	5	4	3	9	1	8
9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	2	5	7	6	9	1
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	4	6	5	7	8	9
2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	7	8	9	1	2	3
3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	1	3	2	4	5	6
4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	6	9	8	2	3	5
5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	9	1	4	3	6	7
6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	8	7	1	5	4	2
7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	3	2	6	8	7	4
8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	5	4	3	9	1	8
9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	2	5	7	6	9	1
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	4	6	5	7	8	9
2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	7	8	9	1	2	3
3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	1	3	2	4	5	6
4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	6	9	8	2	3	5
5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	9	1	4	3	6	7
6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	8	7	1	5	4	2
7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	3	2	6	8	7	4
8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	5	4	3	9	1	8
9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	2	5	7	6	9	1
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	2	7	9	4	3	5
2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	9	2	7	5	4	3
3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	7	9	2	3	5	4
4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	3	6	5	7	1	8
5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	3	6	8	7	1
6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	5	3	1	8	7
7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	4	1	8	6	9	2
8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	4	1	2	6	9
9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	1	8	4	9	2	6
1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	4	6	5	7	8	9
2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	7	8	9	1	2	3
3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	1	3	2	4	6	5
4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	3	2	6	8	7	4
5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	4	3	9	1	8
6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	2	5	7	6	9	1
7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	9	7	4	5	3	2
8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	6	1	8	3	5	7
9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	8	9	1	2	4	6
1	2	3	9	7	8	6	4	5	1	2	3	9	7	8	6	4	5	1	2	3	4	5	6
4	5	6	1	2	3	8	9	7	4	5	6	1	2	3	8	9	7	4	5	6	7	8	9
7	8	9	5	6	4	1	2	3	7	8	9	5	6	4	1	2	3	7	8	9	1	2	3

Non-GYD PGYD for $(v, k, b) = (18, 12, 48)$ is the transpose of the following matrix. The blocks in $\mathcal{D}_C$ form the complements of the blocks of a BIBD with order 18, block size 6 and index 5, and with three blocks removed. The blocks in $\mathcal{D}_R$ form the complements of four copies of each of these three removed blocks.

4	5	6	10	11	12	13	14	15	16	17	18
1	2	3	7	8	9	16	17	18	10	11	12
7	8	9	13	14	15	1	2	3	4	5	6
13	14	15	16	17	18	4	5	6	7	8	9
10	11	12	1	2	3	7	8	9	13	14	15
16	17	18	4	5	6	10	11	12	1	2	3
5	6	7	8	9	10	11	12	16	17	18	4
17	18	13	14	15	7	8	9	1	2	3	16
14	15	1	2	3	4	5	6	10	11	12	13
11	12	4	5	6	13	14	15	7	8	9	10
2	3	10	11	12	16	17	18	13	14	15	1
8	9	16	17	18	1	2	3	4	5	6	7
3	4	2	15	7	5	18	10	8	12	13	17
6	10	5	3	1	8	9	13	11	18	16	14
15	1	14	6	4	2	12	16	17	9	7	11
9	13	17	18	10	11	3	7	14	6	4	2
12	16	8	9	13	17	6	1	2	15	10	5
18	7	11	12	16	14	15	4	5	3	1	8
2	3	5	6	7	8	10	11	13	15	16	18
1	17	3	8	6	9	11	12	4	13	14	16
4	1	2	5	15	7	12	10	14	18	9	17
3	2	9	7	5	6	14	16	10	12	17	13
8	5	1	2	4	11	9	13	12	16	18	15
7	8	4	1	3	17	6	18	15	14	10	11
18	16	14	13	12	10	8	9	5	4	3	2
14	9	17	11	10	16	15	7	6	5	1	3
15	13	11	17	2	12	18	8	7	1	6	4
9	11	10	16	17	15	13	4	3	2	8	6
10	18	12	15	16	14	7	5	8	6	2	1
11	12	7	9	14	18	4	3	1	17	13	5
12	7	13	4	11	5	16	17	2	3	15	9
17	10	6	14	13	3	5	1	18	8	12	7
6	14	16	10	18	1	2	15	11	9	4	8
13	4	18	12	8	2	3	6	16	7	11	14
5	6	15	18	9	13	1	2	17	11	7	10
16	15	8	3	1	4	17	14	9	10	5	12
2	3	4	5	7	9	10	11	15	16	18	14
1	6	5	3	11	7	12	8	13	15	16	17
4	1	2	6	12	10	8	9	14	13	17	18
3	2	6	4	10	8	7	12	17	14	15	16
6	5	1	2	9	12	11	7	16	18	14	13
5	4	3	1	8	11	9	10	18	17	13	15
8	12	11	9	14	17	15	18	2	5	3	6
12	8	9	11	17	14	18	15	5	2	6	3
7	9	10	12	15	18	13	16	3	6	1	4
9	7	12	10	18	15	16	13	6	3	4	1
11	10	7	8	13	16	14	17	1	4	2	5
10	11	8	7	16	13	17	14	4	1	5	2

Bibliography

Agrawal, H. (1966), 'Some generalizations of distinct representatives with applications to statistical designs', *Ann. Math. Statist.* **37**, 525–528.

Ash, A. (1981), 'Generalized Youden designs: construction and tables', *J. Statist. Plann. Inference* **5**(1), 1–25.

Bailey, R. F., Burgess, A. C., Cavers, M. S. and Meagher, K. (2011), 'Generalized covering designs and clique coverings.', *J. Combin. Des* **19**(5), 378–406.

Bliemer, M. C. and Rose, J. M. (2011), 'Experimental design influences on stated choice outputs: an empirical study in air travel choice', *Transportation Research Part A: Policy and Practice* **45**(1), 63–79.

Bluskov, I., Greig, M. and Heinrich, K. (2000), 'Infinite classes of covering numbers', *Canadian Mathematical Bulletin* **43**(4), 385–396.

Booth, K. H. V. and Cox, D. R. (1962), 'Some systematic supersaturated designs', *Technometrics* **4**, 489–495.

Bose, R. C. (1949), 'A note on fisher's inequality for balanced incomplete block designs', *The Annals of Mathematical Statistics* **20**(4), 619–620.

Bose, R. C. and Connor, W. S. (1952), 'Combinatorial properties of group divisible incomplete block designs', *The Annals of Mathematical Statistics* pp. 367–383.

Bryant, D., Buchanan, M., Horsley, D., Maenhaut, B. and Scharaschkin, V. (2011), 'On the non-existence of pair covering designs with at least as many points as blocks', *Combinatorica* **31**(5), 507–528.

Burgess, L. and Street, D. J. (2003), 'Optimal designs for 2^k choice experiments', *Comm. Statist. Theory Methods* **32**(11), 2185–2206.

Burgess, L. and Street, D. J. (2006), 'The optimal size of choice sets in choice experiments', *Statistics* **40**(6), 507–515.

Bush, S. (2014), 'Optimal designs for stated choice experiments generated from fractional factorial designs', *J. Stat. Theory Pract.* **8**(2), 367–381.

Bush, S., Street, D. J. and Burgess, L. (2012), 'Optimal designs for stated choice experiments that incorporate position effects', *Comm. Statist. Theory Methods* **41**(10), 1771–1795.

Caro, Y. and Tuza, Z. (1991), 'Improved lower bounds on k-independence', *Journal of Graph Theory* **15**(1), 99–107.

Chai, F.-S., Das, A. and Singh, R. (2017), 'Three-level A-and D-optimal paired choice designs', *Statistics & Probability Letters* **122**, 211–217.

Chai, F.-S., Das, A. and Singh, R. (2018), 'Optimal two-level choice designs for estimating main and specified two-factor interaction effects', *Journal of Statistical Theory and Practice* **12**(1), 82–92.

Cheng, C.-S. (1980), 'Optimality of some weighing and 2^n fractional factorial designs', *Ann. Statist.* **8**, 436–446.

Cheng, C.-S. (1981*a*), 'A family of pseudo-Youden designs with row size less than the number of symbols', *J. Combin. Theory Ser. A* **31**(2), 219–221.

Cheng, C.-S. (1981*b*), 'Optimality and construction of pseudo-Youden designs', *Ann. Statist.* **9**(1), 201–205.

Cheng, C.-S. (1995), 'Some projection properties of orthogonal arrays', *Ann. Statist.* **23**(4), 1223–1233.

Cheng, C.-S. (1997), '$E(s^2)$-optimal supersaturated designs', *Statist. Sinica* **7**(4), 929–939.

Cheng, C.-S., Das, A., Singh, R. and Tsai, P.-W. (2018), 'E(s^2)-and UE(s^2)-optimal supersaturated designs', *Journal of Statistical Planning and Inference* **196**, 105–114.

Cheng, C.-S., Masaro, J. C. and Wong, C. S. (1985), 'Optimal weighing designs', *SIAM Journal on Algebraic Discrete Methods* **6**(2), 259–267.

Chiu, W.-Y. and John, P. W. M. (1998), 'D-optimal fractional factorial designs', *Statist. Probab. Lett.* **37**(4), 367–373.

Das, A., Dey, A., Chan, L.-Y. and Chatterjee, K. (2008), 'On $E(s^2)$-optimal supersaturated designs', *J. Statist. Plann. Inference* **138**(12), 3749–3757.

Das, A., Horsley, D. and Singh, R. (2018), 'Pseudo generalized Youden designs', *J. Combin. Des* **26**(9), 439–454.

Das, A. and Singh, R. (2016), 'A unified approach to choice experiments', *Technical Report, IIT Bombay, http://dspace.library.iitb.ac.in/jspui/handle/100/18429* .

de Caen, D. (1983), 'Extension of a theorem of moon and moser on complete subgraphs', *Ars Combinatoria* **16**, 5–10.

de Werra, D. (1971), 'Equitable colorations of graphs', *Rev. Française Informat. Recherche Opérationnelle* **5**(Sér. R-3), 3–8.

Demirkale, F., Donovan, D. and Street, D. J. (2013), 'Constructing D-optimal symmetric stated preference discrete choice experiments', *J. Statist. Plann. Inference* **143**(8), 1380–1391.

Dey, A. (1975), 'A note on balanced designs', *Sankhyā Ser. B* **37**(4), 461–462.

Dey, A. (2009), 'Orthogonally blocked three-level second order designs', *J. Statist. Plann. Inference* **139**(10), 3698–3705.

Dey, A. and Mukerjee, R. (1999), 'Inter-effect orthogonality and optimality in hierarchical models', *Sankhyā Ser. B* **61**(3), 460–468.

Dey, A., Singh, R. and Das, A. (2017), 'Efficient paired choice designs with fewer choice pairs', *Metrika* **80**(3), 309–317.

Dey, A. and Suen, C.-Y. (2002), 'Optimal fractional factorial plans for main effects and specified two-factor interactions: a projective geometric approach', *Ann. Statist.* **30**(5), 1512–1523.

Eccleston, J. A. and Hedayat, A. (1974), 'On the theory of connected designs: characterization and optimality', *Ann. Statist.* **2**, 1238–1255.

El-Helbawy, A. T. and Bradley, R. A. (1978), 'Treatment contrasts in paired comparisons: Large-sample results, applications, and some optimal designs', *J. Amer. Statist. Assoc.* **73**(364), 831–839.

Erdős, P. and Hanani, H. (1963), 'On a limit theorem in combinatorial analysis', *Publ. Math. Debrecen* **10**, 10–13.

Erdős, P. and Rényi, A. (1956), 'On some combinatorical problems', *Publ. Math. Debrecen* **4**, 398–405.

Fisher, R. A. (1940), 'An examination of the different possible solutions of a problem in incomplete blocks', *Annals of Human Genetics* **10**(1), 52–75.

Ford, Jr., L. R. and Fulkerson, D. R. (1958), 'Network flow and systems of representatives', *Canad. J. Math.* **10**, 78–84.

Füredi, Z. (1990), 'Covering pairs by q2+ q+ 1 sets', *Journal of Combinatorial Theory, Series A* **54**(2), 248–271.

Galil, Z. and Kiefer, J. (1980), 'D-optimum weighing designs', *Ann. Statist.* **8**, 1293–1306.

Galil, Z. and Kiefer, J. (1982), 'Construction methods for D-optimum weighing designs when n $\equiv$ 3 (mod 4)', *Ann. Statist.* **10**(2), 502–510.

Glock, S., Kühn, D., Lo, A. and Osthus, D. (2016), 'The existence of designs via iterative absorption', *arXiv:1611.06827* .

Goos, P. and Großmann, H. (2011), 'Optimal design of factorial paired comparison experiments in the presence of within-pair order effects', *Food quality and preference* **22**(2), 198–204.

Gordon, D. M. (n.d.), 'La jolla covering repository', *http://www.ccrwest.org/cover.html* .

Gordon, D. M., Patashnik, O. and Kuperberg, G. (1995), 'New constructions for covering designs', *Journal of Combinatorial Designs* **3**(4), 269–284.

Gordon, D. and Stinson, D. (2007), Coverings, *in* C. Colbourn and J. Dinitz, eds, 'The CRC Handbook of Combinatorial Designs, 2nd edition', CRC Press, pp. 365–373.

Graßhoff, U., Großmann, H., Holling, H. and Schwabe, R. (2003), 'Optimal paired comparison designs for first-order interactions', *Statistics* **37**(5), 373–386.

Graßhoff, U., Großmann, H., Holling, H. and Schwabe, R. (2004), 'Optimal designs for main effects in linear paired comparison models', *J. Statist. Plann. Inference* **126**(1), 361–376.

Greig, M., Li, P. and Van Rees, G. (2006), 'Covering designs on 13 blocks revisited', *Utilitas Mathematica* **70**, 221.

Großmann, H. and Schwabe, R. (2015), Design for discrete choice experiments, *in* A. Dean, M. Morris, J. Stufken and D. Bingham, eds, 'Handbook of Design and Analysis of Experiments', Chapman and Hall, Boca Raton, FL, pp. 791–835.

Großmann, H., Schwabe, R. and Gilmour, S. G. (2012), 'Designs for first-order interactions in paired comparison experiments with two-level factors', *J. Statist. Plann. Inference* **142**(8), 2395–2401.

Gupta, S. (1995), 'Multifactor designs for test versus control comparisons', *Utilitas Math.* **47**, 199–210.

Gupta, S. (1998), 'A class of multi-factor designs for test versus control comparisons', *J. Statist. Plann. Inference* **72**(1), 291–302.

Haines, L. M. (2015), Introduction to linear models, *in* A. Dean, M. Morris, J. Stufken and D. Bingham, eds, 'Handbook of Design and Analysis of Experiments', Chapman and Hall, Boca Raton, FL, pp. 63–95.

Hall, Jr., M. (1998), *Combinatorial theory*, second edn, John Wiley & Sons, Inc., New York.

Hedayat, A. S., Jacroux, M. and Majumdar, D. (1988), 'Optimal designs for comparing test treatments with controls', *Statist. Sci.* **4**, 462–491.

Hedayat, A. S. and Pesotan, H. (1992), 'Two-level factorial designs for main effects and selected two-factor interactions', *Statist. Sinica* **2**(2), 453–464.

Hedayat, A. S. and Pesotan, H. (1997), 'Designs for two-level factorial experiments with linear models containing main effects and selected two-factor interactions', *J. Statist. Plann. Inference* **64**(1), 109–124.

Hedayat, A. S., Sloane, N. J. A. and Stufken, J. (1999), *Orthogonal arrays: Theory and applications*, Springer Series in Statistics, Springer-Verlag, New York.

Hedayat, A. and Stufken, J. (1989), 'A relation between pairwise balanced and variance balanced block designs', *J. Amer. Statist. Assoc.* **84**(407), 753–755.

Hogben, L. (2013), *Handbook of linear algebra*, Chapman and Hall/CRC.

Horsley, D. (2017), 'Generalising fishers inequality to coverings and packings', *Combinatorica* **37**(4), 673–696.

Horsley, D. and Singh, R. (2018), 'New lower bounds for t-coverings', *J. Combin. Des* **26**(8), 369–386.

Huber, J. and Zwerina, K. (1996), 'The importance of utility balance in efficient choice designs', *J. Marketing Res.* **33**, 307–317.

Jacroux, M., Wong, C. and Masaro, J. (1983), 'On the optimality of chemical balance weighing designs', *J. Statist. Plann. Inference* **8**(2), 231–240.

Jones, B. and Majumdar, D. (2014), 'Optimal supersaturated designs', *J. Amer. Statist. Assoc.* **109**(508), 1592–1600.

Keevash, P. (2014), 'The existence of designs', *arXiv:1401.3665* .

Kessels, R., Goos, P. and Vandebroek, M. (2006), 'A comparison of criteria to design efficient choice experiments', *Journal of Marketing Research* **43**(3), 409–419.

Kessels, R., Goos, P. and Vandebroek, M. (2008), 'Optimal designs for conjoint experiments', *Computational statistics & data analysis* **52**(5), 2369–2387.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2008), 'Recommendations on the use of bayesian optimal designs for choice experiments', *Quality and Reliability Engineering International* **24**(6), 737–744.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2009), 'An efficient algorithm for constructing bayesian optimal choice designs', *Journal of Business & Economic Statistics* **27**(2), 279–291.

Kiefer, J. (1958), 'On the nonrandomized optimality and randomized nonoptimality of symmetrical designs', *Ann. Math. Statist.* **29**, 675–699.

Kiefer, J. (1975*a*), 'Balanced block designs and generalized Youden designs. I. Construction (patchwork)', *Ann. Statist.* **3**, 109–118.

Kiefer, J. (1975*b*), Construction and optimality of generalized Youden designs, *in* S. N. Srivastava, ed., 'A survey of statistical design and linear models', North-Holland, Amsterdam, pp. 333–353.

Kunert, J. and Sailer, O. (2007), 'Randomization of neighbour balanced generalized Youden designs', *J. Statist. Plann. Inference* **137**(6), 2045–2055.

Lin, D. K. J. (1993*a*), 'Another look at first-order saturated design: The p-efficient designs.', **35**, 284–292.

Lin, D. K. J. (1993*b*), 'A new class of supersaturated designs.', **35**, 28–31.

Majumdar, D. (1996), Optimal and efficient treatment-control designs, *in* S. Ghosh and C. Rao, eds, 'Handbook of Statistics 13: Design and Analysis of Experiments', Elsevier, North-Holland, Amsterdam, pp. 1007–1053.

Marley, C. J. and Woods, D. C. (2010), 'A comparison of design and model selection methods for supersaturated experiments', *Comput. Statist. Data Anal.* **54**(12), 3158–3167.

Marshall, A. W. and Olkin, I. (1979), *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, New York, pp. 105–142.

McSorley, J. and Phillips, N. (2007), 'Complete enumeration and properties of binary pseudo-Youden designs PYD$(9, 6, 6)$', *J. Statist. Plann. Inference* **137**(4), 1464–1473.

Mills, W. (1979), 'Covering designs i: coverings by a small number of subsets', *Ars Combin* **8**, 199–315.

Mills, W. and Mullin, R. (1992), Coverings and packings, *in* J. Dinitz and D. Stinson, eds, 'Contemporary Design Theory', Wiley, pp. 371–399.

Morgan, J. and Stallings, J. (2014), 'On the A- criterion of experimental design', *Journal of Statistical Theory and Practice* **8**(3), 418–422.

Nguyen, N.-K. (1996), 'An algorithmic approach to constructing supersaturated designs', *Technometrics* **38**(1), 69–73.

Nguyen, N.-K. and Cheng, C.-S. (2008), 'New $E(s^2)$-optimal supersaturated designs constructed from incomplete block designs', *Technometrics* **50**(1), 26–31.

Nilson, T. (2011), 'Pseudo-Youden designs balanced for intersection', *J. Statist. Plann. Inference* **141**(6), 2030–2034.

Payne, S. E. (1974), 'On maximizing det $(A^T A)$', *Discrete Math.* **10**(1), 145–158.

Ray-Chaudhuri, D. and Wilson, R. (1975), 'On t-designs', *Osaka J. Math.* **12**, 737–744.

Rödl, V. (1985), 'On a packing and covering problem', *European Journal of Combinatorics* **6**(1), 69–78.

Rose, J. and Bliemer, M. C. (2014), Stated choice experimental design theory: the who, the what and the why, *in* S. Hess and A. Daly, eds, 'Handbook of Choice Modelling', Edward Elgar Publishing, Padstow, United Kingdom, pp. 152–177.

Ruiz, F. and Seiden, E. (1974), 'On construction of some families of generalized Youden designs', *Ann. Statist.* **2**, 503–519.

Sándor, Z. and Wedel, M. (2001), 'Designing conjoint choice experiments using managers prior beliefs', *Journal of Marketing Research* **38**(4), 430–444.

Sándor, Z. and Wedel, M. (2002), 'Profile construction in experimental choice designs for mixed logit models', *Marketing Science* **21**(4), 455–475.

Sándor, Z. and Wedel, M. (2005), 'Heterogeneous conjoint choice designs', *Journal of Marketing Research* **42**(2), 210–218.

Schönheim, J. (1964), 'On coverings', *Pacific J. Math.* **14**, 1405–1411.

Seiden, E. and Wu, C. J. (1978), 'A geometric construction of generalized Youden designs for v a power of a prime', *Ann. Statist.* **6**(2), 452–460.

Shah, K. and Sinha, B. (1989), *Theory of optimal designs*, Vol. 54 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

Singh, R. (2019), 'On three-level d-optimal paired choice designs', *Statistics & Probability Letters* **145**, 127–132.

Singh, R., Chai, F.-S. and Das, A. (2015), 'Optimal two-level choice designs for any number of choice sets', *Biometrika* **102**(4), 967–973.

Singh, R., Das, A. and Chai, F.-S. (2018), 'Optimal paired choice block designs', *Stat. Sinica* p. accepted.

Street, D. J. and Burgess, L. (2004), 'Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments', *J. Statist. Plann. Inference* **118**(1-2), 185–199.

Street, D. J. and Burgess, L. (2007), *The Construction of Optimal Stated Choice Experiments: Theory and Methods*, Wiley, Hoboken, New Jersey.

Street, D. J. and Burgess, L. (2012), Designs for choice experiments for the multinomial logit model, *in* K. Hinkelmann, ed., 'Design and Analysis of Experiments, Special Designs and Applications', Wiley, Hoboken, New Jersey, pp. 331–378.

Sun, D. X. (1993), 'Estimation capacity and related topics in experimental designs. ph.d. dissertation, university of waterloo.'.

Sun, F. and Dean, A. (2016), 'A-optimal and A-efficient designs for discrete choice experiments', *J. Statist. Plann. Inference* **170**, 144–157.

Sun, F. and Dean, A. (2017), 'A-efficient discrete choice designs for attributes with unequal numbers of levels', *J. Stat. Theory Pract.* **11**(2), 322–338.

Tang, B. and Wu, C. F. J. (1997), 'A method for constructing super-saturated designs and its $E(s^2)$ optimality', *Canad. J. Statist.* **25**, 191–201.

Todorov, D. T. (1984), 'Some coverings derived from finite planes', *Colloq. Math. Soc. János Bolyai* **37**, 697–710.

Todorov, D. T. (1985), 'On some covering designs', *J. Combin. Theory Ser. A* **39**, 83–101.

Todorov, D. T. (1989), 'Lower bounds for coverings of pairs by large blocks', *Combinatorica* **9**, 217–225.

Todorov, D. T. and Tonchev, V. D. (1982), 'On some coverings of triples', *C. R. Acad. Bulgare Sci.* **35**, 1209–1211.

Tsai, P.-W. and Gilmour, S. G. (2016), 'New families of q_b-optimal saturated two-level main effects screening designs.', *Statist. Sinica* **26**, 605–617.

Wilson, R. (1982), 'Incidence matrices of t-designs', *Linear Algebra Appl.* **46**, 73–82.

Wu, C. and Chen, Y. (1992), 'A graph-aided method for planning two-level experiments when certain interactions are important', *Technometrics* **34**(2), 162–175.

Wu, C.-F. J. (1993), 'Construction of supersaturated designs through partially aliased interactions', *Biometrika* **80**(3), 661–669.

Yu, J., Goos, P. and Vandebroek, M. (2009), 'Efficient conjoint choice designs in the presence of respondent heterogeneity', *Marketing Science* **28**(1), 122–135.