Title
Space-Filling Designs and Big Data Subsampling

Author
Wang, Lin

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Space-Filling Designs and Big Data Subsampling

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Statistics

by

Lin Wang

2019

ABSTRACT OF THE DISSERTATION

Space-Filling Designs and Big Data Subsampling

by

Lin Wang

Doctor of Philosophy in Statistics

University of California, Los Angeles, 2019

Professor Hongquan Xu, Chair

Space-filling designs are commonly used in computer experiments and other scenarios for investigating complex systems, but the construction of such designs is challenging. In this thesis, we construct a series of maximin-distance Latin hypercube designs via Williams transformations of good lattice point designs. Some constructed designs are optimal under the maximin L_1-distance criterion, while others are asymptotically optimal. Moreover, these designs are also shown to have small pairwise correlations between columns. The procedure is further extended to the construction of multi-level nonregular fractional factorial designs which have better properties than regular designs. Existing research on the construction of nonregular designs focuses on two-level designs. We construct a novel class of multilevel nonregular designs by permuting levels of regular designs via the Williams transformation. The constructed designs can reduce aliasing among effects without increasing the run size. They are more efficient than regular designs for studying quantitative factors. In addition, we explore the application of experimental design strategies to data-driven problems and develop a subsampling framework for big data linear regression. The subsampling procedure inherits optimality from the design matrices and therefore minimizes the mean squared error of coefficient estimations for sufficiently large data. It works especially well for the problem of label-constrained regression where a large covariate dataset is available but only a small set of labels are observable. The subsampling procedure can also be used for big data reduction where computation and storage issues are the primary concern.

The dissertation of Lin Wang is approved.

Arash Ali Amini

Weng Kee Wong

Qing Zhou

Hongquan Xu, Committee Chair

University of California, Los Angeles

2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I owe a great deal of gratitude to my advisor, Professor Hongquan Xu. He provided invaluable strategic advice at every stage of my graduate study at UCLA, from choosing a thesis topic to succeeding in the job market, as well as work-family balance. Working with him was among the most rewarding and pleasant experience in my life. Nobody has had a greater positive impact on my academic career than Professor Xu. From him, I learned how good an advisor can be, and the enormous amounts of time and effort required to lead a student to grow up as a researcher. I hope to become the kind of professor to my students that Professor Xu has been to his.

I am very fortunate to have Professors Arash Ali Amini, Weng Kee Wong, and Qing Zhou in my committee. Their classes provided the backbone for the statistical training that I rely on in my career. Discussions with them during my graduate study are always encouraging and motivating. The collaborative project with Professor Wong provided essential support to me on the job market.

I also want to thank Professor Jingyi Jessica Li for the enormous insights she gave to me on being a female researcher. The collaborative project with her extended my research area, and I look forward to continued collaboration in the coming years. Her great academic passion would always inspire me in my career.

I am very proud to work with my fellow students and junior researchers, especially Qian Xiao, Jianfeng Yang, Fasheng Sun, Yaping Wang, Jake Richard Kramer, Zachary Stokes, Ye Tian, and Yuhao Yin. Thanks for the help and advice they bring to my research and life. Their friendship and academic support made my time at UCLA much more pleasant. I am also very grateful to my parents and my husband. All of this would not have been possible without their undivided love, support and encouragement.

VITA

2008	B.S. in Mathematics, Dalian University of Technology, Liaoning, China
2011	M.S. in Mathematics, Dalian University of Technology, Liaoning, China
2015	Ph.D. in Mathematics, Nankai University, Tianjin, China
2015–2016	Reader, Statistics Department, UCLA.
2016	Research Assistant, Statistics Department, UCLA.
2016 – 2018	Teaching Assistant/Associate, Statistics Department, UCLA.

PUBLICATIONS

- Wang, L., Yang, J.-F., Lin, D. K. J., and Liu, M.-Q. (2015). Nearly Orthogonal Latin Hypercube Designs for Many Design Columns. Statistica Sinica, 25, 1599–1612.

- Wang, L., Xiao, Q., and Xu, H. (2018). Optimal maximin L_1-distance Latin hypercube designs based on good lattice point designs. The Annals of Statistics, 46, 3741–3766.

- Wang, L., Sun, F., Lin, D. K. J., and Liu, M.-Q. (2018). Construction of orthogonal symmetric Latin hypercube designs. Statistica Sinica, 28, 1503–1520.

- Xiao, Q., Wang, L., and Xu, H. (2019). Application of kriging models for a drug combination experiment on lung cancer. Statistics in Medicine, 38, 236–246.

- Wang, L. and Xu, H. (2018). A class of multilevel nonregular fractional factorial designs for studying quantitative factors. arXiv: 1812.05202.

- Wang, L., Liu, M.-Q., and Xu, H. (2019). Fractional factorial designs for fourier-cosine models. Manuscript.

- Xi, N., Wang, L., and Li, J. J. (2019). Prediction of Unobservable Individual Cell RNA Expression under Varied Experimental Conditions. Manuscript.

CHAPTER 1

Introduction

Computer experiments are increasingly being used to investigate complex systems (Sacks et al., 1989; Santner et al., 2003; Fang et al., 2005; Morris and Moore, 2015). A general design approach to planning computer experiments is to seek design points that fill a design region as uniformly as possible (Lin and Tang, 2015). Representative designs include Latin hypercube designs (LHDs) and their modifications, maximin distance designs (Johnson et al., 1990) and uniform designs (Fang and Wang, 1993). LHDs have uniform one-dimensional projections and orthogonal-array based LHDs (Tang, 1993; He and Tang, 2012, 2014; He et al., 2018) have improved two- or three-dimensional projections. Many researchers have constructed orthogonal or nearly orthogonal LHDs; see, among others, Ye (1998), Steinberg and Lin (2006), Cioppa and Lucas (2007), Lin et al. (2009), Sun et al. (2009), Yang and Liu (2012), Georgiou and Efthimiou (2014), Lin and Tang (2015), and Sun and Tang (2017). However, these LHDs are often not space-filling in high dimensions (Joseph and Hung, 2008; Xiao and Xu, 2018).

Computer experiments are often modeled as Gaussian processes. When the correlations between observations rapidly decrease as the distances between design points increase, maximin distance designs are asymptotically D-optimal in the sense that they maximize the determinant of the correlation matrix (Johnson et al., 1990). A maximin distance design spreads design points over the design space in such a way that the separation distance, i.e., the minimal distance between pairs of points, is maximized. Some researchers proposed algorithms such as simulated annealing (Morris and Mitchell, 1995; Joseph and Hung, 2008; Ba et al., 2015) and swarm optimization algorithms (Moon et al., 2011; Chen et al., 2013) to construct maximin distance LHDs. However, such methods are not efficient for constructing large designs due to their computational complexity. Nevertheless, large designs are needed for computer experiments; for example, Morris (1991) considered many simulation models involving hundreds of factors. Therefore, efficient approaches to the generation of large designs are in high demand.

Fractional factorial designs are also widely used in various scientific investigations and industrial applications. These designs are classified into two broad types: regular designs and nonregular designs. Designs that can be constructed through defining relations among factors are called regular designs, while all other designs are nonregular. There are many more nonregular designs than regular designs. Good nonregular designs can either fill the gaps between regular designs in terms of various run sizes or provide better estimation for factorial effects. The construction of good nonregular designs is important and challenging. Constructions for two-level nonregular designs include Plackett and Burman (1946), Deng and Tang (2002), Xu and Deng (2005), Fang et al. (2007), Phoa and Xu (2009), among others. While numerous constructions are available for two-level designs, constructions for designs of three or more levels rarely exist (Xu et al., 2009). This is because the number of multilevel nonregular designs is huge so that providing an efficient algorithm for searching the design space is super challenging.

Because data are now easier to gather, data-driven models, rooted in big data sets, are gaining more ground as one of the best tools in decision-making processes. However, the analysis of big data usually involves critical issues. First, the fast-growing computational powers are still far from sufficient to handle the explosion of modern data sets. Also, while we are taking advantages of big data, in many applications, however, labelling all data points is infeasible due to the limit of time and budget. We are often encountered with the problem where we are given a large data set of n data points but can only observe a small subset of $k < n$ labels. These issues present a new challenge of choosing a representative subdata set so that maximum information can be extracted. The space-filling and fractional factorial design strategies can both be applied to subdata selection so that the selected data achieve some optimality for particular statistical models.

In Chapter 2, we will propose a series of systematic methods to construct maximin L_1-distance LHDs. The L_1-distance provides a lower bound for the L_2-distance so that the constructed designs also perform well regarding the L_2-distance. The proposed method is based on the Williams transformation and its modification. The Williams transformation was first used by Williams (1949) to construct Latin square designs that are balanced for nearest neighbors. Bailey (1982) and Edmondson (1993) used the transformation to construct designs orthogonal to polynomial trends. Butler (2001) used the transformation to construct optimal and orthogonal LHDs under a second-order cosine model. Our purpose is different from theirs. We apply the Williams transformation to good lattice point (GLP) designs and construct a class of asymptotically optimal maximin LHDs. Applying the leave-one-out method we obtain another class of asymptotically optimal maximin LHDs. By modifying the Williams transformation, we obtain a class of exactly optimal maximin LHDs. Moreover, all resulting designs have small pairwise correlations between columns and the average correlations converge to zero as the design sizes increase. This near orthogonality is desirable for estimating potential linear trend efficiently in a Gaussian process.

In Chapter 3, we will provide a class of multilevel nonregular designs via the Williams transformation. We construct a class of nonregular designs by manipulating nonlinear level permutations on regular designs via the Williams transformation. While linear level permutations have been studied by Cheng and Wu (2001), Xu et al. (2004), Ye et al. (2007) for three-level designs, and by Tang and Xu (2014) to improve properties of regular designs, as far as we know, nonlinear level permutations have not been studied. Note that linearly permuted regular designs can be still considered as regular because they are just cosets of regular designs and share the same defining relationship.

In Chapter 4, we will develop a sequential addition-elimination algorithm for subdata selection. The algorithm is inspired by the fact that an orthogonal array of two levels is D-, A-, and G-optimal for linear regression. We define a discrepancy to measure how well a subdata set approximates an orthogonal array. Based on this criterion, we develop an algorithm which sequentially selects data points from the full data as well as eliminating points from the full data to reduce the number of candidate points and speed up the selecting process. Simulations show that the algorithm outperforms existing methods in minimizing mean squared errors of parameter estimations and maximizing D- and A-efficiencies of the design matrices.

CHAPTER 2

Optimal Maximin L_1-Distance Latin Hypercube Designs Based on Good Lattice Point Designs

This chapter proposes a series of systematic methods to construct maximin L_1-distance LHDs. The L_1-distance provides a lower bound for the L_2-distance so that the constructed designs also perform well regarding the L_2-distance. The proposed method is based on the Williams transformation and its modification. The Williams transformation was first used by Williams (1949) to construct Latin square designs that are balanced for nearest neighbors. Bailey (1982) and Edmondson (1993) used the transformation to construct designs orthogonal to polynomial trends. Butler (2001) used the transformation to construct optimal and orthogonal LHDs under a second-order cosine model. Our purpose is different from theirs. We apply the Williams transformation to GLP designs and construct a class of asymptotically optimal maximin LHDs. Applying the leave-one-out method we obtain another class of asymptotically optimal maximin LHDs. By modifying the Williams transformation, we obtain a class of exactly optimal maximin LHDs. Moreover, all resulting designs have small pairwise correlations between columns and the average correlations converge to zero as the design sizes increase. This near orthogonality is desirable for estimating potential linear trend efficiently in a Gaussian process.

2.1 Construction methods

An $N \times n$ LHD is an $N \times n$ matrix where each column is a permutation of N equally spaced levels, denoted by $0, \ldots, N-1$ or $1, \ldots, N$. The L_1-distance between two vectors $x_1 = (x_{11}, \ldots, x_{1n})$ and $x_2 = (x_{21}, \ldots, x_{2n})$ is $d(x_1, x_2) = \sum_{j=1}^{n} |x_{1j} - x_{2j}|$. For an $N \times n$ design matrix D, let x_i be the ith row, $i = 1, \ldots, N$, and $d_{ik}(D)$ be the L_1-distance between the ith and kth rows of D, i.e., $d_{ik}(D) = d(x_i, x_k)$. The L_1-distance of D, denoted by $d(D) = \min\{d_{ik}(D) : i \neq k,\ i, k = 1, \ldots, N\}$, is the minimum L_1-distance between any two distinct rows in D. The maximin distance criterion (Johnson et al., 1990) is to maximize $d(D)$ among all possible designs. For an $N \times n$ LHD, the average pairwise L_1-distance between rows is $(N + 1)n/3$ (Zhou and Xu, 2015). Because the minimum pairwise L_1-distance cannot exceed the integer part of the average, we have the following result.

Lemma 2.1. *For any $N \times n$ LHD D, $d(D) \leq d_{upper} = \lfloor (N + 1)n/3 \rfloor$, where $\lfloor x \rfloor$ is the integer part of x.*

Let $h = (h_1, \ldots, h_n)$ be a set of positive integers smaller than and coprime to N. An $N \times n$ GLP design $D = (x_{ij})$ is defined by $x_{ij} = i \times h_j \bmod N$ for $i = 1, \ldots, N$ and $j = 1, \ldots, n$. The last row of D is a vector of zeros. Each column of D is a permutation of $\{0, \ldots, N - 1\}$. Thus a GLP design is an LHD. We can construct an $N \times n$ GLP design for any $n \leq \phi(N)$, where $\phi(N)$ is the Euler function, i.e., the number of positive integers smaller than and coprime to N. Let $D_b = D + b \bmod N$ for $b = 0, \ldots, N - 1$, that is, D_b is a linearly permuted GLP design. Then D_b is still an LHD. Zhou and Xu (2015) showed that $d(D_b) \geq d(D)$ for any b and proposed to search b that maximizes $d(D_b)$.

2.1.1 Williams transformation

Given an integer N, for $x = 0, \ldots, N - 1$, the Williams transformation is defined by

$$W(x) = \begin{cases} 2x, & \text{for } 0 \leq x < N/2; \\ 2(N - x) - 1, & \text{for } N/2 \leq x < N. \end{cases} \tag{2.1}$$

The Williams transformation is a permutation of $\{0, \ldots, N - 1\}$. Hence, for an LHD $D = (x_{ij})$, $W(D) = (W(x_{ij}))$ is also an LHD. The following example shows that the Williams transformation can further increase the L_1-distance of linearly permuted GLP designs.

Example 2.1. *Consider $N = 11$ and $h = (1, \ldots, 10)$. The GLP design $D = (x_{ij})$ is an 11×10 LHD with $x_{ij} = i \times j$ (mod 11) and $d(D) = 30$. For each $b = 0, \ldots, 10$, we obtain two designs via linear permutation and Williams transformation, namely, $D_b = D + b$ (mod 11) and $E_b = W(D_b)$.*

Table 2.1: The L_1-distances of D_b and E_b in Example 2.1

b	0	1	2	3	4	5	6	7	8	9	10
$d(D_b)$	30	34	30	32	31	30	31	32	30	34	30
$d(E_b)$	10	39	31	31	39	10	28	34	30	34	28

Table 2.1 shows the L_1-distances of D_b and E_b. The linearly permuted designs D_b's have distances ranging from 30 to 34, while the distances for E_b's vary from 10 to 39. The upper bound from Lemma 2.1 is 40. The best design from D_b's is D_1 or D_9 with $d(D_1) = d(D_9) = 34$, while the best design from E_b's is E_1 or E_4 with $d(E_1) = d(E_4) = 39$.

Example 2.1 shows that the Williams transformation can generate designs with larger distances than the linear permutation. Inspired by this, we propose a new construction for maximin LHDs:

Algorithm 2.1 (Williams transformation of linearly permuted GLP designs)**.**

Step 1. *Given a pair of integers N and $n \leq \phi(N)$, generate an $N \times n$ GLP design D.*

Step 2. *For $b = 0, \ldots, N-1$, generate $D_b = D + b \bmod N$ and $E_b = W(D_b)$.*

Step 3. *Find the best D_b and E_b which maximize $d(D_b)$ and $d(E_b)$, respectively.*

As an illustration, we apply Algorithm 2.1 for $N = 7, \ldots, 30$ and $n = \phi(N)$. Table 2.2 compares LHDs generated by the linear permutation, the Williams transformation, R package SLHD provided by Ba et al. (2015), and the Gilbert and Golomb methods proposed by Xiao and Xu (2017). For the SLHD method, the command `maximinSLHD` adopts L_2-distance as the measure. We ran the command with option $t = 1$ and default settings for 100 times, and chose the design with the largest L_1-distance. The Williams transformation always offers better designs than the linear permuation except for $N = 13$, and consistently outperforms the Gilbert and Golomb methods, which only work for prime N. Compared to the SLHD package, the Williams transformation performs better for designs with moderate to large sizes. The Williams transformation performs specially well when N is a prime.

Table 2.2: Comparison of L_1-distances of $N \times n$ LHDs

N	n	LP	WT	SLHD	Gil	Gol	N	n	LP	WT	SLHD	Gil	Gol
7	6	13	16	15	14	14	19	18	106	115	108	102	106
8	4	8	10	11			20	8	32	42	43		
9	6	15	16	18			21	12	66	76	73		
10	4	8	11	11			22	10	60	68	61		
11	10	34	39	36	34	34	23	22	154	168	160	154	158
12	4	8	10	13			24	8	32	36	50		
13	12	54	52	52	46	48	25	20	147	162	153		
14	6	22	24	23			26	12	84	98	87		
15	8	29	36	35			27	18	135	156	145		
16	8	32	36	37			28	12	72	94	92		
17	16	84	94	86	86	80	29	28	250	274	254	250	244
18	6	18	28	28			30	8	40	62	57		

Note: LP, linear permutation; WT, Williams transformation; SLHD, R package SLHD; Gil, Gilbert method; Gol, Golomb method.

Table 2.3: Comparison of L_1-distances of $(N-1)\times n$ LHDs

N	n	LP-1	WT-1	SLHD	Gil	Gol	N	n	LP-1	WT-1	SLHD	Gil	Gol
7	6	12	14	14	14	14	19	18	104	112	103	102	106
8	4	8	9	9			20	8	37	40	41		
9	6	14	14	16			21	12	64	74	71		
10	4	10	10	11			22	10	56	64	60		
11	10	34	36	34	34	34	23	22	152	166	152	154	158
12	4	8	10	12			24	8	32	36	47		
13	12	52	50	47	46	48	25	20	146	156	146		
14	6	19	23	22			26	12	80	93	85		
15	8	28	34	34			27	18	134	152	139		
16	8	32	34	36			28	12	81	91	89		
17	16	82	88	82	86	80	29	28	244	268	247	250	244
18	6	18	27	26			30	8	40	60	56		

Note: LP-1, leave-one-out linear permutation; WT-1, leave-one-out Williams transformation.

2.1.2 Leave-one-out method

Since the last row of a GLP design D is $(0,\ldots,0)$, then the last rows of D_b and E_b are $(b,\ldots,b)$ and $(W(b),\ldots,W(b))$, respectively. The leave-one-out method is to delete the constant row of a design and rearrange the levels so that the resulting design is still an LHD. Specifically, starting from D_b, we delete the last row and reduce the levels $b+1,\ldots,N-1$ by one, which gives us an $(N-1)\times n$ LHD, denoted by D_b^*. Similarly, from E_b, we obtain another $(N-1)\times n$ LHD, denoted by E_b^*. Table 2.3 compares the L_1-distances of D_b^* and E_b^* for $N=7,\ldots,30$, as well as the $(N-1)\times n$ designs generated by R package SLHD and the Gilbert and Golomb methods. From Table 2.3, the leave-one-out Williams transformation generates designs with larger L_1-distance than other methods in most cases. It performs specially well when N is a prime.

Table 2.4: The design matrices of D and $w(D)/2$ in Example 2.2

D										$w(D)/2$									
1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	5	4	3	2	1
2	4	6	8	10	1	3	5	7	9	2	4	5	3	1	1	3	5	4	2
3	6	9	1	4	7	10	2	5	8	3	5	2	1	4	4	1	2	5	3
4	8	1	5	9	2	6	10	3	7	4	3	1	5	2	2	5	1	3	4
5	10	4	9	3	8	2	7	1	6	5	1	4	2	3	3	2	4	1	5
6	1	7	2	8	3	9	4	10	5	5	1	4	2	3	3	2	4	1	5
7	3	10	6	2	9	5	1	8	4	4	3	1	5	2	2	5	1	3	4
8	5	2	10	7	4	1	9	6	3	3	5	2	1	4	4	1	2	5	3
9	7	5	3	1	10	8	6	4	2	2	4	5	3	1	1	3	5	4	2
10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	5	4	3	2	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

2.1.3 Modified Williams transformation

To construct other maximin LHDs, we propose a modified Williams transformation. For $x = 0, \ldots, N-1$, define

$$w(x) = \begin{cases} 2x, & \text{for } 0 \leq x < N/2; \\ 2(N-x), & \text{for } N/2 \leq x < N. \end{cases} \tag{2.2}$$

The following lemma shows an important connection between the original and modified Williams transformations.

Lemma 2.2. *Let N be an odd prime, D be an $N \times (N-1)$ GLP design, and $D_b = D + b \bmod N$ for $b = 0, \ldots, N-1$. Then $d_{ik}(w(D_b)) = d_{ik}(W(D_b))$ for $i + k \neq N$ and $i, k = 1, \ldots, N-1$.*

The $w(x)$ is always an even number, so $w(D_b)$ is not an LHD. We can construct LHDs by selecting some submatrices of $w(D)/2$. Let us see an illustrating example.

Example 2.2. *Consider $N = 11$ and the 11×10 GLP design D. The design matrices of D and $w(D)/2$ are shown in Table 2.4. If we divide the design matrix of $w(D)/2$ into four blocks as shown in Table 2.4, then each block is an LHD. Denote H_1 and H_2 as the top two blocks, and*

Table 2.5: Comparison of L_1-distances of $m \times m$ LHDs

m	MWT	SLHD	Wel	Gil	Gol	m	MWT	SLHD	Wel	Gil	Gol
5	10	10	10	10	8	23	184	167	166	164	
6	14	14	12	14	14	26	234	212			
8	24	22				29	290	263	264	266	270
9	30	28			26	30	310	281	240	276	292
11	44	40	40	40	40	33	374	340			
14	70	64				35	420	383			386
15	80	72			72	36	444	402	342	408	404
18	114	103	90	102	106	39	520	473			482
20	140	126				41	574	523	524	534	520
21	154	141			140	44	660	604			

Note: MWT, modified Williams transformation; Wel, Welch.

H_3 and H_4 as the bottom two blocks, respectively. It can be verified that H_1 and H_2 are 5×5 LHDs with $d(H_1) = d(H_2) = 10$, which attains the upper bound of L_1-distance in Lemma 2.1. In fact, H_1 and H_2 are the same design up to column permutations; in addition, H_3 and H_4 can be obtained by adding a row of zeros to H_1 and H_2, respectively.

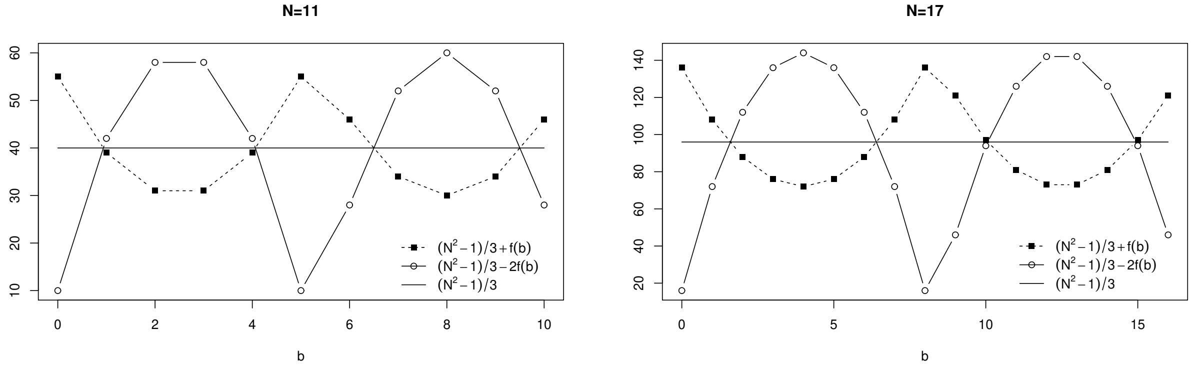
Generally, suppose that N is an odd prime with $N = 2m+1$ and $D = (x_{ij})$ is the $N \times (N-1)$ GLP design. Since $x_{ij} + x_{(N-i)j} = N$ and $x_{ij} + x_{i(N-j)} = N$ for any $i, j = 1, \ldots, N-1$, then

$$D = \begin{pmatrix} A_1 & N - A_2 \\ N - A_3 & A_4 \\ 0_m & 0_m \end{pmatrix} \text{ and } w(D) = \begin{pmatrix} w(A_1) & w(A_2) \\ w(A_3) & w(A_4) \\ 0_m & 0_m \end{pmatrix}, \tag{2.3}$$

where A_1 is the $m \times m$ leading principal submatrix of D, and A_2, A_3, and A_4 can be obtained from A_1 by reversing the order of columns, rows, and both, respectively. In fact, $w(A_1), \ldots, w(A_4)$ are the same design up to row and column permutations, each column of which is a permutation of $\{2, 4, \ldots, 2m\}$. Let

$$H = w(A_1)/2 \tag{2.4}$$

Figure 2.1: The three possible values of pairwise L_1-distance of E_b for $N = 11$ or 17.



be an $m \times m$ LHD from the modified Williams transformation. Table 2.5 compares LHDs generated by the modified Williams transformation, the R package SLHD, and the Welch, Gilbert and Golomb methods from Xiao and Xu (2017). The modified Williams transformation always provides better designs than any other methods. In fact, the L_1-distance of each design generated by the modified Williams transformation in Table 2.5 attains the upper bound given in Lemma 2.1.

2.2 Theoretical results

The Williams transformation leads to a remarkably simple design structure in terms of the L_1-distance when N is an odd prime.

Theorem 2.1. *Let N be an odd prime, D be an $N \times (N-1)$ GLP design, $D_b = D + b$ mod N and $E_b = W(D_b)$ for $b = 0, \ldots, N-1$. Then for $i \neq k$,*

$$d_{ik}(E_b) = \begin{cases} (N^2-1)/3 + f(b), & \textit{for } i = N \textit{ or } k = N, \\ (N^2-1)/3 - 2f(b), & \textit{for } i = N - k, \\ (N^2-1)/3, & \textit{otherwise,} \end{cases}$$

and $d(E_b) = (N^2-1)/3 + \min\{f(b), -2f(b)\}$, where $f(b) = (W(b) - (N-1)/2)^2 - (N^2-1)/12$.

The pairwise L_1-distance between any two distinct rows of E_b takes on only three possible values. One attains $d_{upper} = (N^2-1)/3$ given in Lemma 2.1, and the other two vary around d_{upper}. Figure 2.1 shows the three values for $N = 11$ and $N = 17$ for each $b = 0, \ldots, N-1$.

To maximize $d(E_b)$, we need to maximize $\min\{f(b), -2f(b)\}$. Let $c_0 = \lfloor\sqrt{(N^2-1)/12}\rfloor$,

$$c = \begin{cases} c_0, & \text{if } c_0^2 + 2(c_0+1)^2 \geq (N^2-1)/4; \\ c_0+1, & \text{otherwise,} \end{cases}$$

and

$$b = W^{-1}\left(\frac{N-1}{2} \pm c\right) \tag{2.5}$$

It can be verified that either choice of b defined in (2.5) maximizes $\min\{f(b), -2f(b)\}$ and leads to the best E_b.

Example 2.3. *Consider $N = 11$. Then $c_0 = \lfloor\sqrt{(11^2-1)/12}\rfloor = 3$. Since $c_0^2 + 2(c_0+1)^2 \geq (N^2-1)/4$, set $c = 3$. By* (2.5)*, $b = 1$ or 4. For either $b = 1$ or $b = 4$, by Theorem 2.1, for $i \neq k$,*

$$d_{ik}(E_b) = \begin{cases} 39, & \textit{for } i = 11 \textit{ or } k = 11, \\ 42, & \textit{for } i = 11 - k, \\ 40, & \textit{otherwise.} \end{cases}$$

Hence, $d(E_1) = d(E_4) = 39$.

Based on the upper bound in Lemma 2.1, we define the distance efficiency as

$$d_{\mathrm{e}ff}(D) = d(D)/d_{upper} = d(D)/\lfloor(N+1)n/3\rfloor. \tag{2.6}$$

When N is a prime, $n = \phi(N) = N - 1$ and $(N+1)n/3 = (N^2-1)/3$ is an integer. In this case, $d_{\mathrm{e}ff}(D) = d(D)/((N+1)n/3)$. For example, for the designs E_1 and E_4 in Example 2.3, $d_{\mathrm{e}ff}(E_1) = d_{\mathrm{e}ff}(E_4) = 39/40 = 0.975$. Generally, we have the following result.

Theorem 2.2. *For an odd prime N and b defined in* (2.5)*,*

$$d(E_b) \geq \frac{N^2-1}{3} - \frac{2}{3}\sqrt{\frac{N^2-1}{3}} \textit{ and } d_{eff}(E_b) \geq 1 - \frac{2}{\sqrt{3(N^2-1)}}.$$

As $N \to \infty$, $d_{\mathrm{e}ff}(E_b) \to 1$; so E_b is asymptotically optimal under the maximin distance criterion. For the leave-one-out design E_b^* defined in Section 2.1.2, we have the following result.

Theorem 2.3. *For an odd prime N and b defined in* (2.5)*,*

$$d(E_b^*) \geq \frac{N^2-7}{3} + \frac{1}{3}\sqrt{\frac{N^2-1}{3}} - (N-1).$$

When $N \geq 7$, $d_{eff}(E_b^*) \geq 1 - (3 - 1/\sqrt{3})/N > 1 - 2.43/N$.

For an odd prime $N = 2m+1$ and the $m \times m$ design H constructed in (2.4), we have even better results. By Lemma 2.2 and Theorem 2.1, $d_{ik}(w(D)) = (N^2 - 1)/3$ for $i \neq k$, $i, k = 1, \ldots, m$. By the structure of $w(D)$ shown in (2.3), $d_{ik}(w(A_1)) = d_{ik}(w(D))/2 = (N^2 - 1)/6$; so H is an equidistant LHD and $d(H) = (N^2 - 1)/12 = (m+1)m/3$.

Theorem 2.4. *Let $N = 2m + 1$ be an odd prime, $D = (x_{ij})$ be an $N \times (N - 1)$ GLP design, and A_1 be the $m \times m$ leading principal submatrix of D, that is, $A_1 = (x_{ij})$ with $i, j = 1, \ldots, m$. Then $H = w(A_1)/2$ is a maximin distance LHD with $d(H) = (m + 1)m/3$.*

The modified Williams transformation generates exact maximin LHDs when N is an odd prime. The constructed H is a cyclic Latin square, with each level occurring once in each row and once in each column. We can add a row of zeros to H to obtain an $(m+1) \times m$ LHD, denoted by H^*. It is easy to see that $d(H^*) = d(H) = (m+1)m/3$ and $d_{eff}(H^*) = (m+1)/(m+2) \to 1$ as $m \to \infty$.

The proposed methods are also useful in the construction of maximin L_2-distance designs. An upper bound for the L_2-distance of an $N \times n$ LHD is $d^{(2)}_{upper} = \sqrt{N(N+1)n/6}$ (Zhou and Xu, 2015). Because $\|x\|_2 \geq \|x\|_1/\sqrt{n}$ for any n-vector x, we have $d^{(2)}_{\mathrm{e}ff} > \sqrt{2/3}\, d_{\mathrm{e}ff}$, where $d^{(2)}_{\mathrm{e}ff}$ is the L_2-distance efficiency. Therefore, for an (asymptotically) optimal design under the maximin L_1-distance criterion, its L_2-distance efficiency will tend to be greater than $\sqrt{2/3} > 0.816$. This is a loose lower bound, and yet it illustrates the good performance of our constructed designs regarding the L_2-distance. Numerical calculation shows that our proposed methods are able to produce designs with L_2-distance efficiencies greater than 0.95 for large N.

2.3 Additional results on correlations

We now consider the pairwise correlation between columns for the constructed designs. For any $N \times n$ design $D = (x_{ij})$, define

$$\rho_{ave}(D) = \frac{\sum_{j \neq k} |\rho_{jk}|}{n(n-1)}, \tag{2.7}$$

where ρ_{jk} is the correlation between columns j and k of D. The ρ_{ave} in (2.7) is a performance measure on the overall pairwise column correlations for design D. A good design should have

Table 2.6: Comparison of the ρ_{ave} values for $N \times (N-1)$ LHDs

N	LP	WT	Gil	Gol	N	LP	WT	Gil	Gol
7	.25	.086	.25	.25	47	.09	.015	.09	.11
11	.16	.054	.19	.17	53	.08	.014	.07	.07
13	.07	.065	.16	.18	59	.08	.013	.08	.07
17	.17	.043	.13	.15	61	.07	.012	.07	.07
19	.16	.027	.18	.13	67	.06	.011	.08	.06
23	.14	.022	.12	.09	71	.06	.010	.07	.07
29	.12	.023	.11	.12	73	.06	.011	.06	.08
31	.10	.024	.09	.09	79	.06	.010	.06	.08
37	.11	.017	.10	.10	83	.06	.010	.06	.07
41	.11	.019	.11	.09	89	.06	.009	.07	.06
43	.09	.017	.09	.11	97	.06	.008	.07	.06

a low ρ_{ave} value to reduce correlations between factors and reduce the variance of coefficients estimates.

Consider the ρ_{ave} values for the designs from the Williams transformation. For each prime N, Table 2.6 compares the ρ_{ave} values of designs from the linear permutation, Williams transformation (with b chosen by (2.5)), Gilbert, and Golomb methods. The Williams transformation always generates designs with the smallest ρ_{ave} values. In fact, we have a general result on the average correlation $\rho_{ave}(E_b)$ for any $b = 0, \ldots, N-1$, not restricted to the b defined in (2.5).

Theorem 2.5. *Let N be an odd prime and D be an $N \times (N-1)$ GLP design, $D_b = D + b$ mod N, and $E_b = W(D_b)$ for $b = 0, \ldots, N-1$. Then $\rho_{ave}(E_b) < 2/(N-2)$.*

For a prime N, $\rho_{ave}(E_b) \to 0$ as $N \to \infty$ for any $b = 0, \ldots, N-1$. This property makes it possible to generate large LHDs with tiny pairwise column correlations without any computer search. For the leave-one-out Williams transformation, we have the following result.

Theorem 2.6. *Let N be an odd prime, D be an $N \times (N-1)$ GLP design, $D_b = D + b$ mod N, $E_b = W(D_b)$, and E_b^* be the leave-one-out design obtained from E_b for $b = 0, \ldots, N-1$. Then*

Table 2.7: Comparison of the ρ_{ave} values for $(N-1) \times (N-1)$ LHDs

N	LP-1	WT-1	Gil	Gol	N	LP-1	WT-1	Gil	Gol
7	.35	.211	.21	.20	47	.09	.029	.08	.10
11	.18	.121	.15	.16	53	.07	.027	.06	.06
13	.09	.140	.17	.18	59	.08	.026	.07	.07
17	.14	.095	.11	.14	61	.07	.023	.06	.07
19	.12	.063	.15	.10	67	.06	.022	.08	.06
23	.12	.050	.11	.07	71	.06	.020	.07	.06
29	.11	.046	.09	.13	73	.06	.021	.06	.08
31	.11	.049	.11	.07	79	.07	.020	.06	.08
37	.10	.034	.08	.10	83	.07	.019	.05	.07
41	.09	.038	.09	.09	89	.07	.018	.06	.06
43	.09	.032	.09	.11	97	.06	.016	.07	.06

$\rho_{ave}(E_b^*) < 5(N+1)/(N-2)^2$ *for any* $b = 0, \ldots, N-1$.

Table 2.7 compares designs obtained from the leave-one-out linear permutation, leave-one-out Williams transformation, Gilbert, and Golomb methods. The leave-one-out Williams transformation generates designs with the smallest ρ_{ave} values except for $N = 13$.

For the modified Williams transformation, we have the following result.

Theorem 2.7. *Let* $N = 2m+1$ *be an odd prime,* $D = (x_{ij})$ *be an* $N \times (N-1)$ *GLP design,* A_1 *be the* $m \times m$ *leading principal submatrix of D, that is,* $A_1 = (x_{ij})$ *with* $i, j = 1, \ldots, m$*, and* $H = w(A_1)/2$*. Then* $\rho_{ave}(H) < 2/(m-1)$.

Table 2.8 compares the ρ_{ave} values of designs generated by the modified Williams transformation and some other available methods. The modified Williams transformation always provides designs with the smallest ρ_{ave} values.

Table 2.8: Comparison of the ρ_{ave} values for $m \times m$ LHDs

m	MWT	Wel	Gil	Gol	m	MWT	Wel	Gil	Gol
5	.250	.25	.25	.45	23	.055	.12	.14	
6	.200	.29	.21	.20	26	.049			
8	.143				29	.045	.11	.09	.08
9	.125			.20	30	.044	.11	.11	.07
11	.100	.17	.14	.15	33	.040			
14	.080				35	.038			.09
15	.077			.17	36	.037	.13	.08	.10
18	.067	.17	.15	.10	39	.035			.09
20	.061				41	.033	.11	.11	.11
21	.059			.11	44	.031			

2.4 Extension

We consider extending the results to a general case where $N = kp$ with k and p being prime numbers. Let

$$b = \lfloor N(1 + 1/\sqrt{3})/4 \rfloor, \tag{2.8}$$

and E_b be the $N \times \phi(N)$ design constructed by the Williams transformation. Figure 2.2 (top) shows the values of $d_{eff}(E_b)$ for $N = 2p, 3p, 5p$ and $7p$ and $p \leq 200$. The $d_{eff}(E_b)$ increases quickly as N increases and reaches 0.9 when N is around 30. When $N > 100$, the $d_{eff}(E_b)$ values are typically greater than 0.95 and converge to 1 for $N = 2p$ and $N = 7p$. The $d_{eff}(E_b)$ values do not converge to 1 for $N = 3p$ and $N = 5p$, possibly due to the looseness of the upper bound d_{upper}. In addition, Figure 2.2 (bottom) shows that $\rho_{ave}(E_b)$ goes to 0 quickly as N increases.

We present the asymptotic optimality of E_b for $N = 2p$ based on the theoretical results in Section 2.2. It is possible to establish similar results for other cases with more elaborate arguments, which we do not pursue here.

Theorem 2.8. *Let p be an odd prime, $N = 2p$, D be an $N \times \phi(N)$ GLP design, $D_b = D + b \bmod N$, and $E_b = W(D_b)$. For b defined in* (2.8), *$d_{eff}(E_b) = 1 - O(1/N)$. As $N \to \infty$, $d_{eff}(E_b) \to 1$.*

Now we consider an extension of the leave-one-out procedure. We can generate many asymptotically optimal LHDs by applying the following leave-out-one procedure for rows or columns. When we delete any row from an $N \times n$ LHD D and rearrange the levels as in the leave-one-out method in Section 2.1.2, the distance of the resulting design will reduce at most by n. When we delete any column from an $N \times n$ LHD D, the distance will reduce at most by $N - 1$. Deleting multiple columns and rows together is equivalent to repeating the leave-one-out procedure for multiple times. The following result can be derived.

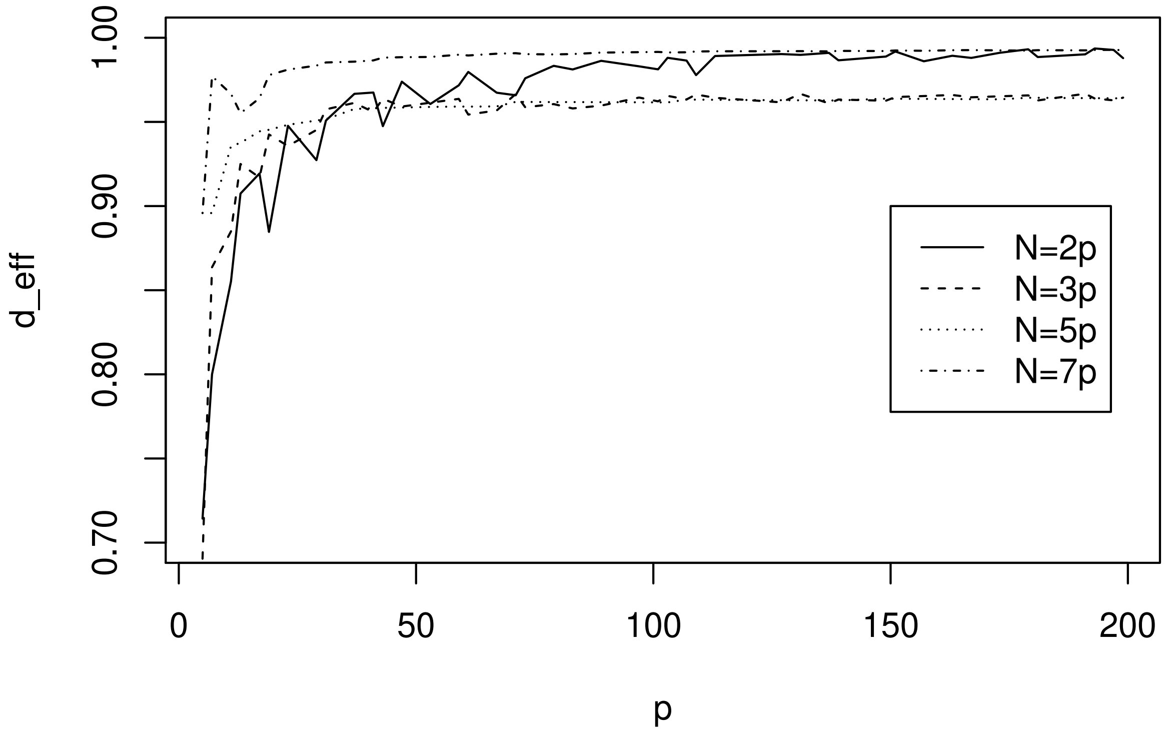
Theorem 2.9. *Let D be an $N \times n$ LHD. Deleting any k_r rows and k_c columns and rearranging the levels yields an $(N - k_r) \times (n - k_c)$ LHD, denoted by D^*. Then $d_{eff}(D^*) \geq d_{eff}(D) - 3k_r/(N - k_r) - 3k_c/(n - k_c)$.*

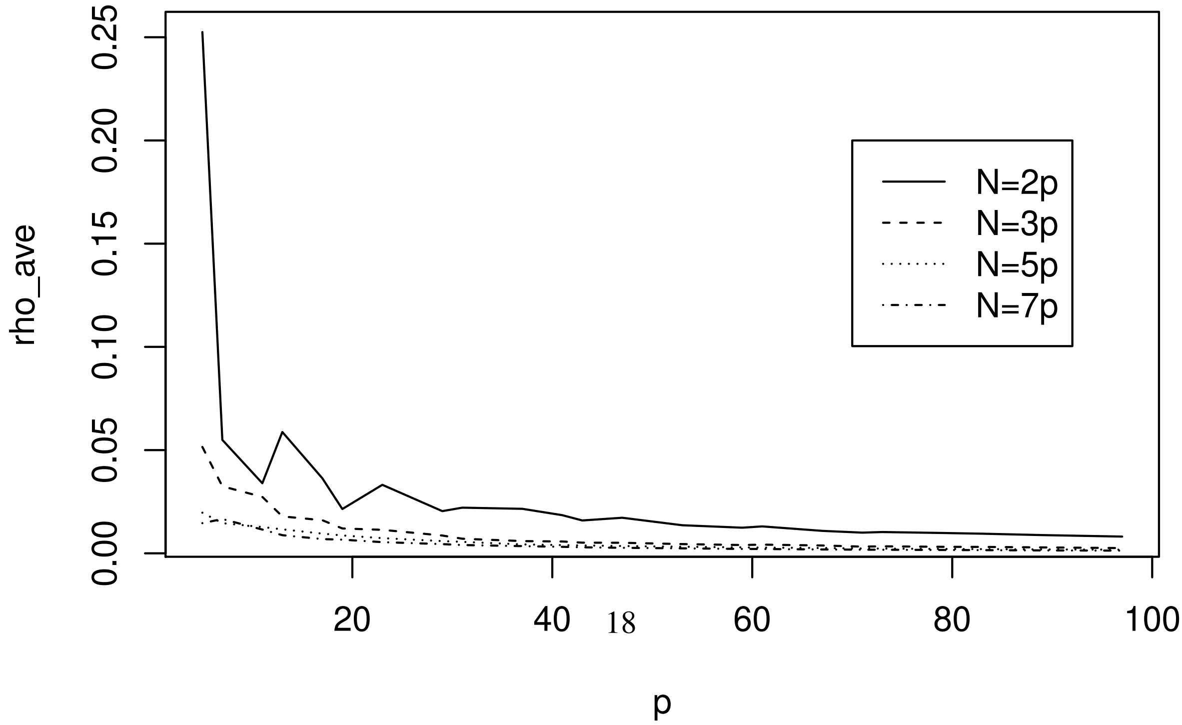
For $N = kp$ and $n = \phi(N)$, $n \to \infty$ as $N \to \infty$. If k_r and k_c are fixed constants not increasing with N, $d_{eff}(D^*) \to 1$ as $N \to \infty$. This multiple leave-one-out procedure yields many asymptotically optimal LHDs with different sizes. For example, let $k = 3$ and $p = 41$, we obtain a 123×80 LHD with $d_{eff} = 0.956$. Delete the last 22 rows and rearrange the levels; we obtain a 101×80 LHD with $d_{eff} = 0.948$. Let $k = 2$ and $p = 61$, we obtain a 122×60 LHD with $d_{eff} = 0.980$. Delete the last 21 rows and rearrange the levels; we obtain a 101×60 LHD with $d_{eff} = 0.961$. Let $k = 5$ and $p = 103$, we obtain a 515×408 LHD with $d_{eff} = 0.962$. Delete the last 3 rows and the last 8 columns, and rearrange the levels, we obtain a 512×400 LHD with $d_{eff} = 0.953$. A distinctive feature of our method is the excellent performance for moderate and large designs. Many other methods slow down quickly as the design size increases and usually give designs with poor distance efficiencies. In contrast, our method generates moderate and large designs with guaranteed high distance efficiencies without search, as long as the ratios of k_r/N and $k_c/\phi(N)$ are small. When the ratios are relatively large, this simple procedure may not work well and further research is needed.

2.5 Summary

We have proposed a series of systematic methods for the construction of maximin LHDs via the Williams transformation and its modification. The Williams transformation and leave-one-

Figure 2.2: The values of $d_{eff}(E_b)$ (top) and $\rho_{ave}(E_b)$ (bottom) with b defined in (2.8).





out method produce asymptotically optimal LHDs under the maximin distance criterion, and the modified Williams transformation generates equidistant LHDs under the L_1-distance. Xu (1999) showed that equidistant LHDs are universally optimal for computer experiments. The average correlations between columns of the constructed designs converge to zero as the design sizes increase. Moreover, the constructed designs often have larger L_1-distance and smaller average correlation than existing designs even for designs with small sizes. The proposed methods are also useful in the construction of maximin L_2-distance designs.

The Williams transformation can be applied to other designs as well. We have found that when applied to fractional factorial designs, the Williams transformation can substantially improve their efficiencies for estimating polynomial models. We will show this result in Chapter 3.

2.6 Appendix: Proofs

We need to distinguish two addition operations. For clarify, let $\oplus$ be the addition operation over the Galois field $\{0,\ldots,N-1\}$. Let $D=(x_{ij})$ be the $N\times\phi(N)$ GLP design and $D_b=(x_{ij}\oplus b)$. When N is a prime, $x_i=(x_{i1},\ldots,x_{i(N-1)})$ and $x_i\oplus b=(x_{i1}\oplus b,\ldots,x_{i(N-1)}\oplus b)$ are the ith row of D and D_b, respectively, x_i is a permutation of $\{1,\ldots,N-1\}$ for $i=2,\ldots,N-1$; and $x_1=(1,\ldots,N-1)$. The designs D and D_b have some important properties which are crucial for the proofs of all theoretical results. We first summarize these properties in the following lemma.

Lemma 2.3. *Let N be an odd prime.*

(i) For $i\neq k$ and $i,k=1,\ldots,N-1$, there exists a unique $q\in\{2,\ldots,N-1\}$ such that $k=iq$ mod N. For any given b, the two matrices

$$\begin{pmatrix} x_i\oplus b \\ x_k\oplus b \end{pmatrix} \text{ and } \begin{pmatrix} x_1\oplus b \\ x_q\oplus b \end{pmatrix}$$

are the same up to column permutations. In addition, $q=N-1$ if and only if $i+k=N$.

(ii) For any $b=0,\ldots,N-1$ and $i=2,\ldots,N-2$, denote $a=(1-i)b$ mod N. The two

matrices

$$\begin{pmatrix} x_1 \oplus b & b \\ x_i \oplus b & b \end{pmatrix} \textit{ and } \begin{pmatrix} x_1 & 0 \\ x_i \oplus a & a \end{pmatrix}$$

are the same up to column permutations.

Proof. Part (i) is obvious from the definition of D and D_b. For (ii), denote $\tilde{x}_i = (x_i, 0)$ for $i = 1, \ldots, N$. Then $\tilde{x}_i \oplus b = i(\tilde{x}_1 \oplus b) \oplus a$. The result follows by noting that $\tilde{x}_1 \oplus b$ is a permutation of $\tilde{x}_1$ and $i\tilde{x}_1 \oplus a = \tilde{x}_i \oplus a = (x_i \oplus a, a)$. $\square$

Proof of Lemma 2.2. We divide the proof in four steps.

Step 1. For $i + k \neq N$, $i \neq k$, and $i, k = 1, \ldots, N-1$, by Lemma 2.3(i), there exists a unique $q \in \{2, \ldots, N-2\}$ such that $d_{ik}(W(D_b)) = d_{1q}(W(D_b))$ and $d_{ik}(w(D_b)) = d_{1q}(w(D_b))$. Therefore, it suffices to show that $d_{1i}(W(D_b)) = d_{1i}(w(D_b))$ for any $b = 0, \ldots, N-1$ and $i = 2, \ldots, N-2$.

Step 2. By Lemma 2.3(ii), to prove $d_{1i}(W(D_b)) = d_{1i}(w(D_b))$, we only need to show that $d(W(x_1), W(x_i \oplus a)) + W(a) = d(w(x_1), w(x_i \oplus a)) + w(a)$ for any $a = 0, \ldots, N-1$. Note that $W(a) = w(a)$ if $a < N/2$, and $W(a) = w(a) - 1$ if $a > N/2$. It suffices to show that

$$d(W(x_1), W(x_i \oplus a)) = \begin{cases} d(w(x_1), w(x_i \oplus a)), & \text{if } a < N/2; \\ d(w(x_1), w(x_i \oplus a)) + 1, & \text{if } a > N/2. \end{cases} \tag{2.9}$$

Step 3. Recall that $x_1 = (1, \ldots, N-1)$ and $x_i \oplus a = (x_{i1} \oplus a, \ldots, x_{i(N-1)} \oplus a)$. Then $d(W(x_1), W(x_i \oplus a)) = \sum_{j=1}^{N-1} |W(j) - W(x_{ij} \oplus a)|$ and $d(w(x_1), w(x_i \oplus a)) = \sum_{j=1}^{N-1} |w(j) - w(x_{ij} \oplus a)|$. It can be shown that

$$|W(j) - W(x_{ij} \oplus a)| = \begin{cases} |w(j) - w(x_{ij} \oplus a)|, & \text{for } j \in I \cup J; \\ |w(j) - w(x_{ij} \oplus a)| - 1, & \text{for } j \in U \backslash I; \\ |w(j) - w(x_{ij} \oplus a)| + 1, & \text{for } j \in V \backslash J, \end{cases}$$

where

$$I = \{j : j < N/2, (x_{ij} \oplus a) < N/2\}, \quad J = \{j : j > N/2, (x_{ij} \oplus a) > N/2\},$$
$$U = \{j : j + (x_{ij} \oplus a) < N\}, \text{ and } \quad V = \{j : j + (x_{ij} \oplus a) \geq N\}.$$

Therefore, to prove (2.9), we need to show that if $a < N/2$, $U\backslash I$ and $V\backslash J$ contain the same number of elements; and if $a > N/2$, $U\backslash I$ contains one less element than $V\backslash J$.

Step 4. Denote $\#S$ as the number of elements in a set S. Since $\#(U\backslash I) = \#U - \#I$ and $\#(V\backslash J) = \#V - \#J$, we want to show that

$$\#U = \#V \text{ and } \begin{cases} \#I = \#J, & \text{if } a < N/2; \\ \#I = \#J + 1, & \text{if } a > N/2. \end{cases}$$

Since

$$x_{(i+1)j} \oplus a = \begin{cases} j + (x_{ij} \oplus a), & \text{for } j \in U; \\ j + (x_{ij} \oplus a) - N, & \text{for } j \in V, \end{cases}$$

then $\sum_{j=1}^{N-1}(x_{(i+1)j} \oplus a) = \sum_{j=1}^{N-1}(x_{ij} \oplus a) + \sum_{j=1}^{N-1} j - (\#V)N$. Because both x_i and x_{i+1} are permutations of $\{1, \ldots, N-1\}$, $\sum_{j=1}^{N-1}(x_{(i+1)j} \oplus a) = \sum_{j=1}^{N-1}(x_{ij} \oplus a)$, which leads to $\#V = \sum_{j=1}^{N-1} j/N = (N-1)/2$. Because $\#U + \#V = N - 1$, $\#U = \#V = (N-1)/2$. Denote $I_1 = \{j : j > N/2, (x_{ij} \oplus a) < N/2\}$. If $a < N/2$, $\#I + \#I_1 = \#J + \#I_1 = (N-1)/2$ so $\#I = \#J$. If $a > N/2$, $\#I + \#I_1 = (N+1)/2$ and $\#J + \#I_1 = (N-1)/2$ so $\#I = \#J + 1$. This completes the proof. $\square$

To prove Theorem 2.1, we need the following lemma.

Lemma 2.4. *For all $i = 2, \ldots, N-2$ and $b = 0, \ldots, N-1$, $d(x_1 \oplus b, x_i \oplus b) + d(N - (x_1 \oplus b), x_i \oplus b) = (2N^2+1)/3 - |N - 2b|$.*

Proof. We divide the proofs in three steps.

Step 1. By Lemma 2.3(ii),

$$\begin{aligned} d(x_1 \oplus b, x_i \oplus b) &= d(x_1, x_i \oplus a) + a, \text{ and} \\ d(N - (x_1 \oplus b), x_i \oplus b) + |N - 2b| &= d(N - x_1, x_i \oplus a) + N - a, \end{aligned}$$

where $a = (1-i)b \bmod N$. Then,

$$d(x_1 \oplus b, x_i \oplus b) + d(N - (x_1 \oplus b), x_i \oplus b) = d(x_1, x_i \oplus a) + d(N - x_1, x_i \oplus a) + N - |N - 2b|.$$

Hence, it suffices to show that $d(x_1, x_i \oplus a) + d(N - x_1, x_i \oplus a) = (2N^2+1)/3 - N = (N-1)(2N-1)/3$ for any $a = 0, \ldots, N-1$.

Step 2. Let $g_i(a) = d(x_1, x_i \oplus a) + d(N - x_1, x_i \oplus a)$. If we can prove $g_i(0) = g_i(1) = \cdots = g_i(N-1)$, we will have

$$g_i(a) = \frac{1}{N}\sum_{c=0}^{N-1} g_i(c) = \frac{1}{N}\sum_{c=0}^{N-1}\left(d(x_1, x_i \oplus c) + d(N - x_1, x_i \oplus c)\right).$$

Because $\sum_{c=0}^{N-1} d(N - x_1, x_i \oplus c) = \sum_{c=0}^{N-1} d(x_1, x_i \oplus c)$, then

$$\begin{aligned} g_i(a) &= \frac{2}{N}\sum_{c=0}^{N-1} d(x_1, x_i \oplus c) = \frac{2}{N}\sum_{c=0}^{N-1}\sum_{j=1}^{N-1} |j - (x_{ij} \oplus c)| \\ &= \frac{2}{N}\sum_{j=1}^{N-1}\sum_{k=0}^{N-1} |j - k| = (N-1)(2N-1)/3. \end{aligned}$$

Step 3. Now we prove that $g_i(0) = g_i(1) = \cdots = g_i(N-1)$. It suffices to show that $g_i(a+1) = g_i(a)$ for any $a = 0, \ldots, N-2$. Recall that $g_i(a) = d(x_1, x_i \oplus a) + d(N - x_1, x_i \oplus a) = \sum_{j=1}^{N-1}\left(|j - (x_{ij} \oplus a)| + |N - j - (x_{ij} \oplus a)|\right)$. Since

$$\begin{aligned} & |j - (x_{ij} \oplus (a+1))| + |N - j - (x_{ij} \oplus (a+1))| \\ = & \begin{cases} |j - (x_{ij} \oplus a)| + |N - j - (x_{ij} \oplus a)|, & \text{for } j \in S_1 \cup S_2; \\ |j - (x_{ij} \oplus a)| + |N - j - (x_{ij} \oplus a)| + 2, & \text{for } j \in S_3; \\ |j - (x_{ij} \oplus a)| + |N - j - (x_{ij} \oplus a)| - 2, & \text{for } j \in S_4, \end{cases} \end{aligned}$$

where

$$\begin{aligned} S_1 &= \{j : j \le x_{ij} \oplus a < N - j\}, & S_2 &= \{j : N - j \le x_{ij} \oplus a < j\}, \\ S_3 &= \{j : x_{ij} \oplus a \ge j, x_{ij} \oplus a \ge N - j\}, & S_4 &= \{j : x_{ij} \oplus a < j, x_{ij} \oplus a < N - j\}, \end{aligned}$$

we only need to show that $\#S_3 = \#S_4$. Note that

$$\begin{cases} x_{(i-1)j} \oplus a = x_{ij} \oplus a - j \text{ and } x_{(i+1)j} \oplus a = x_{ij} \oplus a + j, & \text{for } j \in S_1; \\ x_{(i-1)j} \oplus a = x_{ij} \oplus a - j + N \text{ and } x_{(i+1)j} \oplus a = x_{ij} \oplus a + j - N, & \text{for } j \in S_2; \\ x_{(i-1)j} \oplus a = x_{ij} \oplus a - j \text{ and } x_{(i+1)j} \oplus a = x_{ij} \oplus a + j - N, & \text{for } j \in S_3; \\ x_{(i-1)j} \oplus a = x_{ij} \oplus a - j + N \text{ and } x_{(i+1)j} \oplus a = x_{ij} \oplus a + j, & \text{for } j \in S_4. \end{cases}$$

Then

$$\sum_{j=1}^{N-1}\left((x_{(i-1)j} \oplus a) + (x_{(i+1)j} \oplus a)\right) = 2\sum_{j=1}^{N-1}(x_{ij} \oplus a) - N(\#S_3 - \#S_4). \tag{2.10}$$

Because $x_i \oplus a$ is a permutation of $\{0, \ldots, a-1, a+1, \ldots, N-1\}$ for any $i < N$, $\sum_{j=1}^{N-1}(x_{(i-1)j} \oplus a) = \sum_{j=1}^{N-1}(x_{ij} \oplus a) = \sum_{j=1}^{N-1}(x_{(i+1)j} \oplus a)$. By (2.10), $N(\#S_3 - \#S_4) = 0$ so $\#S_3 = \#S_4$. This completes the proof. $\square$

Proof of Theorem 2.1. For the first case, note that $W(x_i \oplus b)$ is a permutation of $\{0, \ldots, W(b) - 1, W(b)+1, \ldots, N-1\}$, and $W(x_N \oplus b)$ is a constant vector with each component equal to $W(b)$, so $d_{iN}(E_b) = d_{Ni}(E_b) = \sum_{j=0}^{N-1} |j - W(b)| = (N^2-1)/3 + f(b)$.

To prove the result for the second case, $i = N-k$, it suffices to prove the result for the third case. This is because the total pairwise L_1-distance between distinct rows of $W(D_b)$ is $t = (N-1)\sum_{j_1=0}^{N-1}\sum_{j_2=0}^{N-1}|j_1 - j_2| = N(N-1)^2(N+1)/6$. Out of all the pairs of distinct rows, $N-1$ pairs belong to the first case with a total distance $t_1 = (N-1)[(N^2-1)/3 + f(b)]$, $(N-1)(N-3)/2$ pairs belong to the third case with a total distance $t_2 = (N^2-1)(N-1)(N-3)/6$, and $(N-1)/2$ pairs belong to the second case. By Lemma 2.3(i), $d_{i(N-i)}(E_b) = d_{1(N-1)}(E_b)$ for any i. Therefore, $d_{i(N-i)}(E_b) = (t - t_1 - t_2)/[(N-1)/2] = (N^2-1)/3 - 2f(b)$.

Now we prove the result for the last case where $i \neq N-k$, $i \neq N$, and $k \neq N$. By Lemmas 2.2 and 2.3(i), it suffices to consider $d_{1i}(E_b) = d(W(x_1 \oplus b), W(x_i \oplus b)) = d(w(x_1 \oplus b), w(x_i \oplus b))$ for $i = 2, \ldots, N-2$. Denote

$$\begin{aligned} B &= \left(B_1 \,\middle|\, B_2 \,\middle|\, B_3 \,\middle|\, B_4 \right) \\ &= \left(\begin{array}{c|c|c|c} w(x_1 \oplus b) & w(x_1 \oplus b) & 2N - w(x_1 \oplus b) & 2N - w(x_1 \oplus b) \\ w(x_i \oplus b) & 2N - w(x_i \oplus b) & w(x_i \oplus b) & 2N - w(x_i \oplus b) \end{array} \right), \end{aligned}$$

then $d_{1i}(E_b) = d(B_1)$. By column permutations, B can be rearranged as

$$C = \left(\begin{array}{c|c|c|c} 2(x_1 \oplus b) & 2(x_1 \oplus b) & 2N - 2(x_1 \oplus b) & 2N - 2(x_1 \oplus b) \\ 2(x_i \oplus b) & 2N - 2(x_i \oplus b) & 2(x_i \oplus b) & 2N - 2(x_i \oplus b) \end{array} \right).$$

By Lemma 2.4, $d(B) = d(C) = 4((2N^2+1)/3 - |N - 2b|)$. Note that $d(B_1) = d(B_4)$ and $d(B_2) = d(B_3)$. For B_2, in both $w(x_1 \oplus b)$ and $w(x_i \oplus b)$, 0 and $w(b)$ appear once and all other even numbers smaller than N appear twice. Then $d(B_2) = \sum_{j=1}^{N-1}(N - w(x_{1j} \oplus b) - w(x_{ij} \oplus b)) = (N^2+1) - 2|N - 2b|$. Therefore, $d_{1i}(E_b) = d(B_1) = (d(B) - 2d(B_2))/2 = (N^2-1)/3$. $\square$

Proof of Theorem 2.2. If $c_0^2 + 2(c_0+1)^2 \geq (N^2-1)/4$, then $c_0 \geq \sqrt{(N^2-1)/12 - 2/9} - 2/3$ and $c_0^2 \geq (N^2-1)/12 - (4/3)\sqrt{(N^2-1)/12}$. Hence, $d(E_b) = (N^2-1)/4 + c_0^2 \geq (N^2-1)/3 - (4/3)\sqrt{(N^2-1)/12}$. Similarly, if $c_0^2+2(c_0+1)^2 < (N^2-1)/4$, $c_0+1 \leq \sqrt{(N^2-1)/12 - 2/9} + 1/3$, and $(c_0+1)^2 \leq (N^2-1)/12 + (2/3)\sqrt{(N^2-1)/12}$. Then $d(E_b) = (N^2-1)/2 - 2(c_0+1)^2 \geq (N^2-1)/3 - (4/3)\sqrt{(N^2-1)/12}$. Therefore,

$$d(E_b) \geq \frac{N^2-1}{3} - \frac{4}{3}\sqrt{\frac{N^2-1}{12}} = \frac{N^2-1}{3} - \frac{2}{3}\sqrt{\frac{N^2-1}{3}}.$$

By the definition in (2.6), $d_{eff}(E_b) = d(E_b)/((N^2-1)/3) \geq 1 - 2/\sqrt{3(N^2-1)}$. □

Proof of Theorem 2.3. Let $e_i = (e_{i1}, \ldots, e_{i(N-1)})$ and $e_k = (e_{k1}, \ldots, e_{k(N-1)})$ be two distinct rows of E_b for $i, k = 1, \ldots, N-1$, and $e_i^* = (e_{i1}^*, \ldots, e_{i(N-1)}^*)$ and $e_k^* = (e_{k1}^*, \ldots, e_{k(N-1)}^*)$ be the corresponding rows of E_b^*. For $j = 1, \ldots, N-1$, if $e_{ij} > W(b) > e_{kj}$ or $e_{kj} > W(b) > e_{ij}$, $|e_{ij}^* - e_{kj}^*| = |e_{ij} - e_{kj}| - 1$; otherwise, $|e_{ij}^* - e_{kj}^*| = |e_{ij} - e_{kj}|$. Since the number of j's such that $e_{ij} > W(b) > e_{kj}$ (or $e_{kj} > W(b) > e_{ij}$) cannot exceed $\min\{W(b), N-1-W(b)\}$, then $d(E_b^*) \geq d(E_b) - 2\min\{W(b), N-1-W(b)\}$. For the b defined in (2.5), $\min\{W(b), N-1-W(b)\} = (N-1)/2 - c$. Then $d(E_b^*) \geq d(E_b) - (N-1) + 2c \geq d(E_b) - (N-1) + 2(\sqrt{(N^2-1)/12} - 1)$. By Theorem 2.2, $d(E_b^*) \geq (N^2-7)/3 + \sqrt{(N^2-1)/3}/3 - (N-1)$. When $N \geq 7$, we have $d_{eff}(E_b^*) = d(E_b^*)/\lfloor N(N-1)/3 \rfloor \geq d(E_b^*)/(N(N-1)/3) \geq 1 + 1/(\sqrt{3}N) - 3/N > 1 - 2.43/N$.
□

Proof of Theorem 2.5. Let ρ_{jk} be the correlation between the jth and kth columns of E_b. Denote the jth column of D_b as $\tilde{z}_j \oplus b$ for $j = 1, \ldots, N-1$, then $\tilde{z}_j \oplus b = (x_j \oplus b, b)^{\mathrm{T}}$. By Lemma 2.3(i), there exists a unique $q \in \{2, \ldots, N-1\}$ such that $\rho_{jk} = \rho_{1q}$. Thus,

$$\rho_{ave}(E_b) = \frac{\sum_{j=2}^{N-1} |\rho_{1j}|}{N-2}, \tag{2.11}$$

where

$$\begin{aligned} \rho_{1j} &= \operatorname{cor}(W(\tilde{z}_1 \oplus b), W(\tilde{z}_j \oplus b)) \\ &= \frac{\sum_{i=1}^{N} \left(W(x_{i1} \oplus b) - \frac{N-1}{2}\right)\left(W(x_{ij} \oplus b) - \frac{N-1}{2}\right)}{(N^3-N)/12}. \end{aligned} \tag{2.12}$$

For $x \in [0, N]$, the Fourier cosine expansion of $x - N/2$ is given by

$$x - \frac{N}{2} = \sum_{u=1}^{\infty} a_u \cos\left(\frac{u\pi x}{N}\right), \tag{2.13}$$

with

$$a_u = \frac{2}{N}\int_0^N \left(x - \frac{N}{2}\right)\cos\left(\frac{u\pi x}{N}\right)dx = \begin{cases} 0, & \text{if } u \text{ is even;} \\ -4N/(u^2\pi^2), & \text{if } u \text{ is odd.} \end{cases}$$

By (2.13), for any $x + 0.5 \in [0, N]$,

$$x - \frac{N-1}{2} = (x + 0.5) - \frac{N}{2} = \sum_{u=1}^{\infty} a_u \cos\left(\frac{u\pi(x+0.5)}{N}\right).$$

Then the numerator of (2.12) is

$$\begin{aligned} & \sum_{i=1}^{N}\left(W(x_{i1}\oplus b) - \frac{N-1}{2}\right)\left(W(x_{ij}\oplus b) - \frac{N-1}{2}\right) \\ = \; & \sum_{u=1}^{\infty}\sum_{v=1}^{\infty} a_u a_v s(u,v) = \frac{16N^2}{\pi^4}\sum_{\text{odd } u}\sum_{\text{odd } v}\frac{1}{u^2v^2}s(u,v), \end{aligned} \tag{2.14}$$

where

$$s(u,v) = \sum_{i=1}^{N}\cos\left(\frac{u\pi(W(x_{i1}\oplus b)+0.5)}{N}\right)\cos\left(\frac{v\pi(W(x_{ij}\oplus b)+0.5)}{N}\right).$$

By (2.1), for any $x = 0, \ldots, N-1$, $\cos\left(u\pi(W(x)+0.5)/N\right) = \cos(u\pi(2x+0.5)/N)$. Then

$$\begin{aligned} s(u,v) &= \sum_{i=1}^{N}\cos\left(\frac{u\pi(2x_{i1}+2b+0.5)}{N}\right)\cos\left(\frac{v\pi(2x_{ij}+2b+0.5)}{N}\right) \\ &= \frac{1}{2}\sum_{i=1}^{N}\cos\left(\frac{2\pi((jv+u)i+c_1)}{N}\right) + \frac{1}{2}\sum_{i=1}^{N}\cos\left(\frac{2\pi((jv-u)i+c_2)}{N}\right), \end{aligned} \tag{2.15}$$

where $c_1 = (b+0.25)(u+v)$ and $c_2 = (b+0.25)(v-u)$. For positive odd numbers u and v, let $I_1 = \{(u,v) : u = jv \text{ or } -jv, v \neq 0 \bmod N\}$ and $I_2 = \{(u,v) : u = 0 \text{ and } v = 0 \bmod N\}$. For $(u,v) \in I_1$, $|s(u,v)| \leq N/2$ because only one of the two items in (2.15) can be nonzero. For $(u,v) \in I_2$, $|s(u,v)| \leq N$; for $(u,v) \notin I_1 \cup I_2$, $s(u,v) = 0$. Then by (2.11), (2.12), and (2.14),

$$\begin{aligned} \rho_{ave}(E_b) &= \frac{\sum_{j=2}^{N-1}\left|\sum_{i=1}^{N}\left(W(x_{i1}\oplus b) - \frac{N-1}{2}\right)\left(W(x_{ij}\oplus b) - \frac{N-1}{2}\right)\right|}{(N-2)(N^3-N)/12} \\ &\leq \frac{192N^2}{\pi^4(N^3-N)(N-2)}\sum_{j=2}^{N-1}\left(\sum_{I_1}\frac{N}{2}\frac{1}{u^2v^2} + \sum_{I_2}N\frac{1}{u^2v^2}\right) \\ &= \frac{192N^2}{\pi^4(N^2-1)(N-2)}\sum_{j=2}^{N-1}\left(\sum_{I_1}\frac{1}{2u^2v^2} + \sum_{I_2}\frac{1}{u^2v^2}\right). \end{aligned} \tag{2.16}$$

Since

$$\begin{aligned}
&\sum_{j=2}^{N-1}\left(\sum_{I_1}\frac{1}{2u^2v^2}+\sum_{I_2}\frac{1}{u^2v^2}\right)\\
\leq\ & \frac{1}{2}\sum_{\text{odd } v}\frac{1}{v^2}\left(2\sum_{\text{odd } u}\frac{1}{u^2}-\sum_{k=0}^{\infty}\frac{1}{(v+2kN)^2}-2\sum_{\text{odd } k}\frac{1}{k^2N^2}\right)\\
\leq\ & \sum_{\text{odd } v}\frac{1}{v^2}\sum_{\text{odd } u}\frac{1}{u^2}-\frac{1}{2}\sum_{\text{odd } v}\frac{1}{v^4}-\frac{1}{N^2}\sum_{\text{odd } v}\frac{1}{v^2}\sum_{\text{odd } k}\frac{1}{k^2}\\
=\ & \frac{N^2-1}{N^2}\left(\frac{\pi^4}{8^2}\right)-\frac{\pi^4}{192},
\end{aligned}$$

where we used the fact that $\sum_{\text{odd } v} 1/v^2=\pi^2/8$ and $\sum_{\text{odd } v} 1/v^4=\pi^4/96$. Then by (2.16),

$$\begin{aligned}
\rho_{ave}(E_b) &\leq \frac{1}{N-2}\frac{192N^2}{\pi^4(N^2-1)}\left(\frac{N^2-1}{N^2}\left(\frac{\pi^4}{8^2}\right)-\frac{\pi^4}{192}\right)\\
&= \frac{1}{N-2}\left(3-\frac{N^2}{N^2-1}\right)<\frac{2}{N-2}.
\end{aligned}$$

□

Proof of Theorem 2.6. For any $b=0,\ldots,N-1$, let $E_b=(e_{ij})$. Because $\sum_{i=1}^{N}(e_{ij}-(N-1)/2)^2 = N(N^2-1)/12$ for any $j=1,\ldots,N-1$, by Theorem 2.5, we have

$$\sum_{j=2}^{N-1}\left|\sum_{i=1}^{N}\left(e_{i1}-\frac{N-1}{2}\right)\left(e_{ij}-\frac{N-1}{2}\right)\right|<\frac{N(N^2-1)}{6}. \tag{2.17}$$

Let ρ^*_{jk} be the correlation between the jth and kth columns of E^*_b. Similar to (2.11),

$$\rho_{ave}(E^*_b)=\frac{\sum_{j=2}^{N-1}|\rho^*_{1j}|}{N-2}. \tag{2.18}$$

Note that

$$\rho^*_{1j}=\frac{12C_0}{N(N-1)(N-2)} \tag{2.19}$$

with

$$\begin{aligned}
C_0 &= \sum_{\substack{e_{i1}<W(b)\\ e_{ij}<W(b)}}(e_{i1}-\mu)(e_{ij}-\mu)+\sum_{\substack{e_{i1}>W(b)\\ e_{ij}<W(b)}}(e_{i1}-1-\mu)(e_{ij}-\mu)\\
&+\sum_{\substack{e_{i1}<W(b)\\ e_{ij}>W(b)}}(e_{i1}-\mu)(e_{ij}-1-\mu)+\sum_{\substack{e_{i1}>W(b)\\ e_{ij}>W(b)}}(e_{i1}-1-\mu)(e_{ij}-1-\mu)\\
&= \sum_{i=1}^{N}\left(e_{i1}-\frac{N-1}{2}\right)\left(e_{ij}-\frac{N-1}{2}\right)+C_1+C_2,
\end{aligned}$$

where $\mu = (N-2)/2$,

$$\begin{aligned} C_1 &= \frac{1}{2}\left(\sum_{e_{i1}<W(b)} e_{ij} - \sum_{e_{i1}>W(b)} e_{ij} + \sum_{e_{ij}<W(b)} e_{i1} - \sum_{e_{ij}>W(b)} e_{i1}\right) \\ &+ \frac{(N-1)^2}{4} - (W(b))^2 \end{aligned}$$

and

$$C_2 = \frac{1}{4}\left(\sum_{\substack{e_{i1}<W(b)\\ e_{ij}<W(b)}} 1 + \sum_{\substack{e_{i1}>W(b)\\ e_{ij}>W(b)}} 1 - \sum_{\substack{e_{i1}>W(b)\\ e_{ij}<W(b)}} 1 - \sum_{\substack{e_{i1}<W(b)\\ e_{ij}>W(b)}} 1\right).$$

It is easy to see that $|C_1| \leq (N^2-1)/4$ and $|C_2| \leq (N-1)/4$. Hence, by (2.17), (2.18), and (2.19),

$$\begin{aligned} &\rho_{ave}(E_b^*) \\ &< \frac{12}{N(N-1)(N-2)^2}\left(\frac{N(N^2-1)}{6} + \frac{(N-2)(N^2-1)}{4} + \frac{(N-2)(N-1)}{4}\right) \\ &< \frac{5(N+1)}{(N-2)^2}. \end{aligned}$$

□

Proof of Theorem 2.7. The proof is similar to that of Theorem 2.5. By (2.13), for $j = 1, \ldots, (N-1)/2$,

$$\sum_{i=1}^{N}\left(w(x_{i1}) - \frac{N}{2}\right)\left(w(x_{ij}) - \frac{N}{2}\right) = \frac{16N^2}{\pi^4}\sum_{\text{odd } v}\frac{1}{u^2v^2}s(u,v),$$

where

$$s(u,v) = \sum_{i=1}^{N}\cos\left(\frac{u\pi w(x_{i1})}{N}\right)\cos\left(\frac{v\pi w(x_{ij})}{N}\right).$$

Similar to (2.16), we can prove that

$$\sum_{j=2}^{(N-1)/2}\left|\sum_{i=1}^{N}\left(w(x_{i1}) - \frac{N}{2}\right)\left(w(x_{ij}) - \frac{N}{2}\right)\right| \leq \frac{N^3}{24}.$$

Since

$$\begin{aligned} &\sum_{i=1}^{N-1}\left(w(x_{i1}) - \frac{N+1}{2}\right)\left(w(x_{ij}) - \frac{N+1}{2}\right) \\ &= \sum_{i=1}^{N}\left(w(x_{i1}) - \frac{N}{2}\right)\left(w(x_{ij}) - \frac{N}{2}\right) - (N-1) + \frac{(N+1)^2+1}{4}, \end{aligned}$$

then

$$\begin{aligned}
&\sum_{j=2}^{(N-1)/2}\left|\sum_{i=1}^{N-1}\left(w(x_{i1})-\frac{N+1}{2}\right)\left(w(x_{ij})-\frac{N+1}{2}\right)\right| \\
\leq\ & \frac{N^3}{24}+\left(\frac{N-1}{2}-1\right)\left(\frac{(N+1)^2+1}{4}-(N-1)\right) \\
=\ & \frac{N^3}{6}-\frac{5N^2-12N+18}{8} \\
\leq\ & \frac{(N+1)(N-1)(N-3)}{6}.
\end{aligned}$$

Hence,

$$\begin{aligned}
& \rho_{ave}(H)=\rho_{ave}(w(A_1)) \\
=\ & \frac{\sum_{j=2}^{(N-1)/2}\left|\sum_{i=1}^{N-1}\left(w(x_{i1})-\frac{N+1}{2}\right)\left(w(x_{ij})-\frac{N+1}{2}\right)\right|}{(m-1)(N+1)(N-1)(N-3)/12} \\
\leq\ & \frac{2}{m-1}.
\end{aligned}$$

□

Proof of Theorem 2.8. To save space, we sketch only the main steps.

Step 1. For $N=2p$, $\phi(N)=p-1$ and $D=(x_{ij})$ with $x_{ij}=i(2j-1)$ mod N for $i=1,\ldots,2p$ and $j=1,\ldots,p-1$. With proper row and column permutations, D is equivalent to

$$\begin{pmatrix} 2C \\ 2C+p \end{pmatrix} \bmod N, \tag{2.20}$$

where $C=(y_{ij})$ is an $p\times(p-1)$ GLP design with $y_{ij}=i\cdot j$ mod p for $i=1,\ldots,p$ and $j=1,\ldots,p-1$. Then $E_b=W(D_b)$ is equivalent to

$$\tilde{E}_b=\begin{pmatrix} W(2C\oplus b) \\ W(2C\oplus(b+p)) \end{pmatrix}.$$

Step 2. Consider $W(2C\oplus b)$. If b is even, $2C\oplus b=2(C+b/2 \bmod p)$. Then $w(2C\oplus b)=2w_p(C+b/2 \bmod p)$ where w is the modified Williams transformation defined in (2.2) and w_p is the modified Williams with N replaced by p. By Lemma 2.2 and Theorem 2.1, $d_{ik}(w(2C\oplus b))=$

$2[d_{ik}(w_p(C+b/2 \bmod p))] = 2(N^2-1)/3$ for $i \neq k$, $i \neq p, k \neq p$, and $i+k \neq p$. Following the lines of Lemma 2.2 will result $d_{ik}(W(2C \oplus b)) = d_{ik}(w(2C \oplus b))$. Then

$$d_{ik}(W(2C \oplus b)) = (N^2-4)/6 \text{ for } i \neq k, i \neq p, k \neq p, \text{ and } i+k \neq p. \tag{2.21}$$

If b is odd, $W(2C \oplus b) = N-1-W(2C \oplus (b+p))$ and (2.21) also holds.

Step 3. If b is even, the last row of $W(2C \oplus b)$ is $(2b, \ldots, 2b)$ and each other row is a permutation of $\{0, 3, 4, \ldots, 2(p-1)-1, 2(p-1)\} \backslash \{2b\}$. Based on this structure, we get

$$\begin{aligned} d_{ip}\left(W(2C \oplus b)\right) &= \frac{N^2}{6} - \frac{N+2}{4} + \frac{W(b)}{2} + \frac{g(b)}{2}, & (2.22)\\ d_{i(p-i)}\left(W(2C \oplus b)\right) &= \frac{N^2}{6} + \frac{N}{2} - 1 - W(b) - g(b), & (2.23) \end{aligned}$$

where

$$g(b) = \left(W(b) - \frac{1}{2}\left(1+\frac{1}{\sqrt{3}}\right)N\right)\left(W(b) - \frac{1}{2}\left(1-\frac{1}{\sqrt{3}}\right)N\right).$$

Similarly, if b is odd, (2.22) and (2.23) also hold .

Step 4. Because $W(2C \oplus b) = N-1-W(2C \oplus (b+p))$, $W(2C \oplus (b+p))$ has the same distance structure as $W(2C \oplus b)$.

Step 5. By the structure of $W(2C \oplus (b+p))$ and $W(2C \oplus b)$, by computation, we can get

$$d_{i(p+k)}(\tilde{E}(b)) = \begin{cases} N^2/4 - l_1(b), & \text{for } i = k \neq p;\\ (N/2-1)l_1(b), & \text{for } i = k = p;\\ N^2/6 - l_1(b) + 1/3, & \text{for } (i,k) \in I_1;\\ N^2/6 - (N-2)/4 + l_2(b)/2 - l_1(b), & \text{for } (i,k) \in I_2;\\ -N^2/12 + (N/2-1)l_1(b) + N/2 - l_2(b), & \text{for } (i,k) \in I_3. \end{cases} \tag{2.24}$$

where $l_1(b) = |N - 2W(b) - 1|$, $l_2(b) = W(b) + g(b)$, $I_1 = \{(i,k) : i \neq p, k \neq p, i+k \neq p\}$, $I_2 = \{(i,k) : i \neq p, k = p, \text{ or } i = p, k \neq p\}$, and $I_3 = \{(i,k) : i \neq p, k \neq p, i+k = p\}$.

Step 6. For $b = \lfloor N(1+1/\sqrt{3})/4 \rfloor$, $W(b) = 2b = \lfloor N(1+1/\sqrt{3})/2 \rfloor$ or $\lfloor N(1+1/\sqrt{3})/2 \rfloor + 1$, so $-N/\sqrt{3} \leq g(b) \leq 0$. Then $l_1(b) = O(N)$ and $l_2(b) = O(N)$. Since for any $N \times (N/2-1)$ LHD, $d_{\text{upper}} = (N+1)(N-2)/6$, by (2.21)–(2.24), it can be verified that $d_{\text{eff}}(E_b) = d_{\text{eff}}(\tilde{E}_b) = 1 - O(1/N)$.

$\square$

CHAPTER 3

A Class of Multilevel Nonregular Fractional Factorial Designs for Studying Quantitative Factors

This chapter provides a class of multilevel nonregular designs via the Williams transformation. We have applied the Williams transformation to good lattice point sets in Chapter 2 for constructing maximin Latin hypercube designs. In this chapter, we will use the transformation to manipulate nonlinear level permutations and construct a class of nonregular designs. While linear level permutations have been studied by Cheng and Wu (2001), Xu et al. (2004), Ye et al. (2007) for three-level designs, and by Tang and Xu (2014) to improve properties of regular designs, as far as we know, nonlinear level permutations have not been studied. Note that linearly permuted regular designs can be still considered as regular because they are just cosets of regular designs and share the same defining relationship.

Multilevel designs are often used for studying quantitative factors by fitting response surface models such as polynomial models. A commonly accepted principle for polynomial models is that effects of a lower polynomial order are more important than effects of a higher polynomial order, while effects of the same polynomial order are regarded as equally important. Based on this principle, Cheng and Ye (2004) proposed the minimum β-aberration criterion for selecting multilevel designs. For an $N \times n$ design $D = (x_{ij})$, define

$$\beta_k(D) = N^{-2} \sum_{\|u\|_1 = k} \left| \sum_{i=1}^{N} \prod_{j=1}^{n} p_{u_j}(x_{ij}) \right|^2 \quad \text{for } k = 1, \ldots, K, \tag{3.1}$$

where $u = (u_1, \ldots, u_n)$ is a vector in $\{0, \ldots, q-1\}$, $\|u\|_1 = u_1 + \cdots + u_n$, $\{p_0(x), p_1(x), \ldots, p_{q-1}(x)\}$ is a set of orthonormal polynomials, and $K = n(q-1)$. The β_k measures the overall aliasing between jth- and $(k-j)$th-order effects for all j with $0 \le j \le k$. Specifically, β_1 measures the aliasing between the intercept and linear effects, β_2 the aliasing between linear effects, β_3

the aliasing between linear and second-order effects, and β_4 the aliasing between second-order effects. The minimum β-aberration criterion is to find a design D which sequentially minimizes $\beta_k(D)$ for $k = 1, \ldots, K$. Because linear and second-order effects are more important than higher-order effects, the sequential minimization of $\beta_1, \ldots, \beta_4$ would be adequate for choosing designs in practice.

We show that the proposed construction via the Williams transformation can provide better designs than regular designs and linearly permuted regular designs in terms of the minimum β-aberration criterion. We develop a general theory on the construction and apply the theory to construct nonregular designs with five and seven levels.

3.1 Construction via Williams transformation

A design with N runs, n factors and q levels is denoted by an $N \times n$ matrix over $Z_q = \{0, 1, \ldots, q-1\}$, where each row represents a run, and each column represents a factor. For $x \in Z_q$, the Williams transformation is defined by

$$W(x) = \begin{cases} 2x, & \text{for } 0 \leq x < q/2; \\ 2(q-x)-1, & \text{for } q/2 \leq x < q. \end{cases} \tag{3.2}$$

The Williams transformation is a permutation of Z_q. For a design $D = (x_{ij})$, let $W(D) = (W(x_{ij}))$. The following example shows that we can get better designs from the Williams transformation.

Example 3.1. *Consider a 5-level regular design D with three columns x_1, x_2 and $x_3 = x_1 + x_2$. By (3.1), $\beta_1(D) = \beta_2(D) = 0$, $\beta_3(D) = 0.125$, and $\beta_4(D) = 0.525$. For each $b = 0, \ldots, 4$, we obtain two designs via linear permutations and the Williams transformation, namely, D_b with columns x_1, x_2 and $x_3 = x_1 + x_2 + b \bmod 5$ and $E_b = W(D_b)$. It can be verified that all D_b's and E_b's have $\beta_1 = \beta_2 = 0$. Table 3.1 shows their β_3 and β_4. The best design from D_b's is D_3 with $\beta_3 = 0$ and $\beta_4 = 0.686$, while the best design from E_b's is E_4 with $\beta_3 = 0$ and $\beta_4 = 0.027$. Design E_4 performs much better than D_3 under the minimum β-aberration criterion, although they are both better than the original design D.*

Table 3.1: The β-wordlength pattern of D_b and E_b in Example 3.1

b	$\beta_3(D_b)$	$\beta_4(D_b)$	$\beta_3(E_b)$	$\beta_4(E_b)$
0	0.125	0.525	0.442	0.004
1	0.125	0.525	0.168	0.021
2	0.125	0.096	0.168	0.021
3	0.000	0.686	0.442	0.004
4	0.125	0.096	0.000	0.027

Remark 3.1. *In the computation of β_k defined in* (3.1)*, $p_0(x) \equiv 1$ and $p_j(x)$ for $j = 1, \ldots, q-1$ is a polynomial of order j defined on Z_q satisfying*

$$\sum_{x=0}^{q-1} p_i(x)p_j(x) = \begin{cases} 0, & i \neq j; \\ q, & i = j. \end{cases}$$

For example, the orthonormal polynomials for a 5-level factor are $p_0(x) = 1$, $p_1(x) = (x-2)/\sqrt{2}$, $p_2(x) = \sqrt{10/7}\{p_1(x)^2 - 1\}$, $p_3(x) = \{10p_1(x)^3 - 17p_1(x)\}/6$, and $p_4(x) = \{70p_1(x)^4 - 155p_1(x)^2 + 36\}/\sqrt{14}$.

Example 3.1 shows that from a regular design, we can obtain a series of nonregular designs via linear permutations and the Williams transformation. This series of designs can provide better designs than the original regular design and linearly permuted designs. Generally, for a prime number q, a regular q^{n-m} design has $n-m$ independent columns, denoted as $x_1, \ldots, x_{n-m}$, and m dependent columns, denoted as $x_{n-m+1}, \ldots, x_n$, which can be specified by m generators as

$$x_{n-m+i} = c_{i1}x_1 + \cdots + c_{i(n-m)}x_{n-m} \bmod q, \text{ for } i = 1, \ldots, m, \tag{3.3}$$

where each vector $(c_{i1}, \ldots, c_{i(n-m)})$ is a generator whose entries are constants in Z_q. For each regular q^{n-m} design, denoted by D, let

$$D_b = (x_1, \ldots, x_{n-m}, x_{n-m+1} + b_1, \ldots, x_n + b_m) \bmod q, \tag{3.4}$$

and

$$E_b = W(D_b), \tag{3.5}$$

for $b = (b_1, \ldots, b_m) \in Z_q^m$. Note that we only consider permutations for dependent columns in (3.4) because linearly permuting one or more independent columns is equivalent to linearly permuting some dependent columns, which can be seen from (3.3). From each regular q^{n-m} design D, we can derive q^m D_b's and q^m E_b's. To find the best design, we search over all possible regular q^{n-m} designs defined by different generators and all possible permutations $b \in Z_q^m$ for each design. Tang and Xu (2014) proposed to find the best design among the class of all D_b's whereas we consider searching over the class of E_b's and develop theoretical results to accelerate the search in Section 3.2.

For three-level designs, the class of designs E_b's are geometrically isomorphic to the class of designs D_b's, because any three-level design obtained from any nonlinear level permutations is geometrically isomorphic to a regular design or its coset (Tang and Xu, 2014). Two designs are said to be geometrically isomorphic if one can be obtained from the other by row and column exchanges and possibly reversing the level order of some columns. Geometrically isomorphic designs have the same β_k values for all k (Cheng and Ye, 2004). However, with more than three levels, we will see that the class of E_b's can provide many better designs than the class of D_b's.

3.2 Theoretical results

We study properties of E_b in this section. It is well known that a regular design D is an orthogonal array of strength $t \geq 2$. An orthogonal array is a design in which all q^t level combinations appear equally often in every submatrix formed by t columns. The t is called the strength of the orthogonal array, which is often omitted when $t = 2$. Because the Williams transformation is a permutation of $\{0, \ldots, q-1\}$, if $D = (x_{ij})$ is a q-level orthogonal array, then $W(D) = (W(x_{ij}))$ is still an orthogonal array. The following result is from Tang and Xu (2014).

Lemma 3.1. *For an orthogonal array of strength t, $\beta_k = 0$ for $k = 1, \ldots, t$.*

From the construction in (3.5), E_b is an orthogonal array of the same strength as D and D_b. While we use designs of strength 2 in practice, Lemma 3.1 guarantees $\beta_1(E_b) = \beta_2(E_b) = 0$ so that linear effects are not aliased with the intercept, nor with each other. Then we want to minimize

$\beta_3(E_b)$ in order to minimize the aliasing between linear and second-order effects. The following theorem gives a permutation b theoretically to ensure $\beta_3(E_b) = 0$ so that no aliasing exists between any linear and second-order effects.

Theorem 3.1. *For an odd prime q, let*

$$\gamma = W^{-1}((q-1)/2) = \begin{cases} (q-1)/4, & \textit{if } q = 1 \textit{ mod } 4; \\ (3q-1)/4, & \textit{if } q = 3 \textit{ mod } 4. \end{cases} \tag{3.6}$$

Let D be a regular q^{n-m} design generated by (3.3)*, and E_b be defined by* (3.5)*. Then $\beta_3(E_{b^*}) = 0$ with $b^* = (b_1^*, \ldots, b_m^*)$, where*

$$b_i^* = \left(1 - \sum_{j=1}^{n-m} c_{ij}\right)\gamma \quad (i = 1, \ldots, m). \tag{3.7}$$

Example 3.2. *Consider a 7^{3-1} design D with $x_3 = x_1 + x_2$. Then $\gamma = (3 \times 7 - 1)/4 = 5$, and equation* (3.7) *gives $b_1^* = 2$. It can be verified that $\beta_3(E_2) = 0$ and $\beta_4(E_2) = 0.003$. Consider another 7^{3-1} design D with $x_3 = 2x_1 + 2x_2$. Then $\gamma = 5$, and equation* (3.7) *gives $b_1^* = 6$. It can be verified that $\beta_3(E_6) = 0$ and $\beta_4(E_6) = 0.0196$.*

Theorem 3.1 states that given a regular design D, we can always find an E_{b^*} such that $\beta_3(E_{b^*}) = 0$. In the following, we give a sufficient condition for the E_{b^*} to be the unique design with $\beta_3 = 0$ among all possible q^m E_b's.

Definition 3.1. *Let D be a regular q^{n-m} design. If there exist $n - m$ independent columns of D, $z_1, \ldots, z_{n-m}$, and a series of $s+1$ sets of columns, $T_0 \subset \cdots \subset T_s$, such that $T_0 = \{z_1, \ldots, z_{n-m}\}$,*

$$T_{k+1} = T_k \cup \{w \in D : w = c_1 w_1 + c_2 w_2 \textit{ mod } q, w_1, w_2 \in T_k, c_1, c_2 \in Z_q\} \tag{3.8}$$

for $k = 0, \ldots, s-1$, and $T_s = D$, then D is called recursive. Furthermore, if c_1 or c_2 is 1 or -1 for all k, then D is called ordinary-recursive; if both c_1 and c_2 are either 1 or -1 for all k, then D is called simple-recursive.

Example 3.3. *Consider the 7^{3-1} design D defined by $x_3 = 2x_1 + 2x_2$ in Example 3.2. Clearly, D is recursive. Because $x_3 = 2x_1 + 2x_2$, we have $2x_1 + 2x_2 + 6x_3 = 0$ mod 7, $x_1 + x_2 + 3x_3 = 0$ mod 7 and $x_2 = -x_1 + 4x_3$ mod 7. Then D is also ordinary-recursive, if we take $T_0 = \{x_1, x_3\}$ and $T_1 = \{x_1, x_2, x_3\} = D$. However, D is not simple-recursive.*

Example 3.4. *Consider a 5^{5-2} design D with $x_4 = x_1 + x_2$ and $x_5 = x_1 + x_2 + x_3$. Take $T_0 = \{x_1, x_2, x_3\}$, $T_1 = \{x_1, x_2, x_3, x_4\}$ and $T_2 = \{x_1, x_2, x_3, x_4, x_5\} = D$, then D is simple-recursive. If $x_5 = x_1 + x_2 + 2x_3$ instead, then D is ordinary-recursive but not simple-recursive. Consider another 5^{5-2} design D with $x_4 = x_1 + x_2$ and $x_5 = x_1 + 2x_2 + 2x_3$. This design is not recursive because x_5 is not involved in any word of length three. However, when one more column $x_6 = x_1 + 2x_2$ is added, it is ordinary-recursive.*

Regular designs with q^2 runs are commonly used in practice because they are economical and guarantee that linear effects are uncorrelated. Those designs accommodate two independent columns and up to $q - 1$ dependent columns. By Definition 3.1, they are all recursive by letting T_0 include the two independent columns and $T_1 = D$.

Lemma 3.2. *Let q be an odd prime and D be a regular design of q^2 runs. Then D is recursive.*

Clearly, recursive designs include ordinary-recursive designs, which in turn include simple-recursive designs. For three-level designs, the three types of designs are equivalent, while for designs with more than three levels, they are dramatically different. Table 3.2 compares the numbers of the three types of designs with 25 and 49 runs. The numbers of simple-recursive designs are much smaller than the numbers of the other two types of designs. Although there is a difference between the numbers of ordinary-recursive and recursive designs, the difference is small. As the number of columns increases, all designs tend to be ordinary-recursive.

The next theorem gives a sufficient condition for the E_{b^*} to be the unique design with $\beta_3 = 0$ among all possible q^m E_b's.

Theorem 3.2. *For an odd prime q, let D be a regular q^{n-m} design defined by* (3.3)*, and E_b be defined as* (3.5)*. If D is ordinary-recursive, then E_{b^*} with b^* defined in* (3.7) *is the only design with $\beta_3 = 0$ among all q^m E_b's derived from D.*

Remark 3.2. *We can show that if the number of levels is less than 13, Theorem 3.2 also holds for recursive designs. That is, for a recursive q^{n-m} design D, if $q \leq 13$, the E_{b^*} with b^* defined in* (3.7) *is the only design with $\beta_3 = 0$ among all E_b's. However, this is not the case for $q \geq 17$. A counter example for $q = 17$ comes with a 17^{3-1} design with $x_3 = 2x_1 + 4x_2$. By* (3.7)*, $b^* = 14$. Then*

Table 3.2: The numbers of the three types of recursive designs with 25 and 49 runs

	25-run designs			49-run designs		
n	simple	ordinary	recursive	simple	ordinary	recursive
3	2	6	8	2	10	18
4	6	22	24	6	99	135
5	20	32	32	20	517	540
6	16	16	16	70	1214	1215
7				252	1458	1458
8				267	729	729

E_{14} has $\beta_3 = 0$, while the design E_4 with columns x_1, x_2, and $x_3 + 4$ also has zero β_3. That being said, as the number of columns increases, the number of non-ordinary-recursive regular designs decreases dramatically.

Example 3.5. *Consider a 7^{8-6} design D with $x_3 = x_1 + x_2, x_4 = x_1 + 2x_2, x_5 = x_1 + 4x_2, x_6 = x_1 + 5x_2, x_7 = 2x_1 + 5x_2$, and $x_8 = 2x_1 + 6x_2$. There are $7^6 = 117,649$ E_b's derived from D, which makes it cumbersome, if not impossible, to do an exhaustive search for the best E_b. Note that $x_7 = x_1+x_6$, $x_8 = x_3+x_6$. So D is ordinary-recursive by taking $T_0 = \{x_1, x_2\}$, $T_1 = \{x_1, \dots, x_6\}$ and $T_2 = \{x_1, \dots, x_8\} = D$. Equation (3.7) gives $b_1^* = 2, b_2^* = 4, b_3^* = 1, b_4^* = 3, b_5^* = 5$, and $b_6^* = 0$. It can be verified that $\beta_3(E_{b^*}) = 0$ and $\beta_4(E_{b^*}) = 9.677$. By Theorem 3.2, E_{b^*} is the best design among all E_b's derived from D under the minimum β-aberration criterion.*

By Theorem 3.2 and Remark 3.2, for an ordinary-recursive design or a recursive design with no more than 13 levels, E_{b^*} is the best design among all E_b's, which is obtained without any computer search. To study the property of D_b's defined in (3.4), Tang and Xu (2014) showed that if D is simple-recursive, the design $D_{\tilde{b}}$ given by

$$\tilde{b}_i = \left(1 - \sum_{j=1}^{n-m} c_{ij}\right)(q-1)/2 \quad (i = 1, \dots, m) \tag{3.9}$$

is the unique design with $\beta_3 = 0$ among all D_b's. As we have shown above, only a small amount of regular designs are simple-recursive. Therefore, results on simple-recursive designs are usually not applicable for designs with more than three levels. In contrast, Theorem 3.2 is more general and applies to the broader classes of ordinary-recursive and recursive designs.

Theorem 3.2 does not apply to the class of linearly permuted designs D_b's even if D is ordinary-recursive. Here is a counter example.

Example 3.6. *Consider the design 7^{3-1} design D defined by $x_3 = 2x_1 + 2x_2$ in Example 3.2. Example 3.3 shows that it is ordinary-recursive, so by Theorem 3.2, E_{b^*} is the unique design with $\beta_3 = 0$ among all E_b's. In contrast, there are three D_b's with zero β_3. Equation* (3.9) *gives $\tilde{b} = 5$, which leads to $D_{\tilde{b}}$ with $\beta_3 = 0$ and $\beta_4 = 0.0625$. Other than this, both $b = 0$ and $b = 3$ lead to D_b with $\beta_3 = 0$ and $\beta_4 = 0.0417$. All D_b's are worse than E_{b^*} under the minimum β-aberration criterion.*

Theorem 3.2, together with Lemma 3.2 and Remark 3.2, indicates the following result.

Corollary 3.1. *For an odd prime $q \leq 13$, let D be a regular design of q^2 runs. Then E_{b^*} with b^* defined as* (3.7) *is the unique design with $\beta_3 = 0$ among all E_b's derived from D.*

Now we show another useful property of E_{b^*}. A design D over Z_q is called mirror-symmetric if $(q-1)J - D$ is the same design as D, where J is a matrix of unity. Mirror-symmetric designs include two-level foldover designs as special cases.

Theorem 3.3. *For an odd prime q, let D be a regular q^{n-m} design defined by* (3.3)*, and E_b be defined as* (3.5)*. Then E_{b^*} with b^* defined in* (3.7) *is mirror-symmetric.*

Tang and Xu (2014) showed that a design is mirror-symmetric if and only if it has $\beta_k = 0$ for all odd k. By Theorem 3.3, the E_{b^*} has $\beta_k(E_{b^*}) = 0$ for all odd k. This guarantees that all odd-order effects are not aliased with all even-order effects. Specifically, linear effects are not aliased with second-order or fourth-order effects.

3.3 Comparisons and application

We apply our theoretical results to construct nonregular designs with q^2 runs and compare our designs with regular designs and linearly permuted regular designs. Designs with q^2 runs are

Table 3.3: Comparison of β-wordlength patterns for 25-run designs

	D		$D_{\tilde{b}}$			E_{b^*}		
n	β_3	β_4	Generators	β_3	β_4	Generators	β_3	β_4
3	0.125	0.525	(1,2)	0	0.271	(1,1)	0	0.027
4	0.375	1.361	(1,2) (2,1)	0	1.336	(1,1) (1,2)	0	1.037
5	0.750	3.029	(1,1) (1,3) (2,3)	0	3.793	(1,1) (1,2) (1,3)	0	3.768
6	1.250	6.786	(1,1) (1,2) (1,3) (2,3)	0	8.250	(1,1) (1,2) (1,3) (2,3)	0	8.250

widely used in practice due to their run size economy. A regular design with q^2 runs can study up to $(q+1)$ columns given by

$$x_1, x_2, x_1 + x_2, x_1 + 2x_2, x_1 + 3x_2, \ldots, x_1 + (q-1)x_2. \tag{3.10}$$

The common choice of a design with q^2 runs and n columns is to use the first n columns of (3.10); see Wu and Hamada (2009) and Mukerjee and Wu (2006). Denote such a design as D. We search over all q^{n-m} regular designs with $n - m = 2$ to get the best $D_{\tilde{b}}$ and the best E_{b^*}, where $\tilde{b}$ and b^* are defined in (3.9) and (3.7), respectively. To do this, we search over generators (c_1, c_2) for the $m = n - 2$ dependent columns such that each column can be generated by $c_1x_1 + c_2x_2$. Because $(q - c_1)x_1 + c_2x_2$ is a reflection of $c_1x_1 + (q - c_2)x_2$, which leads to geometrically isomorphic designs, we only consider $c_1 = 1, \ldots, (q-1)/2$ and $c_2 = 1, \ldots, q-1$. This leads to $\binom{q-1}{n-2} \cdot \{(q-1)/2\}^{n-2}$ regular designs with strength $t \geq 2$. Tables 3.3 and 3.4 show the comparisons of the standard regular design D, the best $D_{\tilde{b}}$, and the best E_{b^*} with 25 and 49 runs, respectively, as well as the corresponding generators for the $D_{\tilde{b}}$ and E_{b^*}. We can see that the E_{b^*} always performs the best for any design size. The E_{b^*} given in Tables 3.3 and 3.4 is optimal under the minimum β-aberration criterion within the class of E_b's.

Consider applying the three 25-run designs with 3 columns in Table 3.3 to study three five-level quantitative factors. A traditional method for fitting the data is to use the following second-order polynomial model

$$y_i = \alpha_0 + \sum_{j=1}^{3} p_1(x_{ij})\alpha_j + \sum_{j=1}^{3} p_2(x_{ij})\alpha_{jj} + \sum_{j=1}^{2}\sum_{k=j+1}^{3} p_1(x_{ij})p_1(x_{ik})\alpha_{jk} + \varepsilon,\ i = 1, \ldots, 25, \tag{3.11}$$

Table 3.4: Comparison of β-wordlength patterns for 49-run designs

	D		$D_{\tilde{b}}$			E_{b^*}		
n	β_3	β_4	Generators	β_3	β_4	Generators	β_3	β_4
3	0.063	0.563	(1,3)	0	0.063	(1,1)	0	0.003
4	0.188	1.354	(1,3) (3,1)	0	0.250	(1,1) (2,4)	0	0.055
5	0.375	2.440	(1,2) (3,1) (3,5)	0	1.135	(1,1) (1,3) (2,4)	0	0.836
6	0.625	4.313	(1,2) (1,4) (2,3) (2,5)	0	3.094	(1,1) (1,3) (1,4) (2,4)	0	2.368
7	0.938	7.401	(1,1) (1,3) (1,4) (3,1) (3,4)	0	6.438	(1,1) (1,3) (1,4) (2,3) (2,4)	0	4.928
8	1.312	12.78	(1,1) (1,3) (1,4) (3,1) (3,4) (3,6)	0	11.23	(1,1) (1,2) (1,4) (1,5) (2,5) (2,6)	0	9.677

where $p_1(x) = \sqrt{2}(x-2)/2$, $p_2(x) = \sqrt{5/14}\{(x-2)^2 - 2\}$, $x_{i1}, x_{i2}, x_{i3} \in Z_5$ are levels for the three factors, $\alpha_0, \alpha_j, \alpha_{jj}$, and α_{jk} are the intercept, linear, quadratic and bilinear terms, respectively, and $\varepsilon \sim N(0, \sigma^2)$. Because $\beta_3(D) \neq 0$, linear terms are aliased or correlated with bilinear terms for D. While both $D_{\tilde{b}}$ and E_{b^*} have $\beta_1 = \beta_2 = \beta_3 = 0$, the intercept and all the linear terms are not correlated with the quadratic and bilinear terms and so they can be estimated independently. For any design, let M denote the model matrix. Table 3.5 shows part of the information matrix $M^{\mathrm{T}}M/25$ corresponding to the 3 quadratic and 3 bilinear terms: $\alpha_{11}, \alpha_{22}, \alpha_{33}, \alpha_{12}, \alpha_{13}$ and α_{23} for $D_{\tilde{b}}$ and E_{b^*}. It is easy to see that the terms for E_{b^*} are less correlated than that for $D_{\tilde{b}}$. The variance-covariance matrix of the estimates of parameters for these terms is $\sigma^2(M^{\mathrm{T}}M)^{-1}$. For $D_{\tilde{b}}$, the variances of the estimates for quadratic terms α_{11}, α_{22} and α_{33} are $0.047\sigma^2$, $0.041\sigma^2$, and $0.047\sigma^2$, respectively, and for bilinear terms α_{12}, α_{13} and α_{23} are $0.051\sigma^2$, $0.050\sigma^2$, and $0.051\sigma^2$, respectively. For E_{b^*}, the variance of the estimate for each quadratic term is $0.040\sigma^2$, and for each bilinear term is $0.041\sigma^2$. Furthermore, the correlations between the estimates are smaller for E_{b^*} than $D_{\tilde{b}}$. Therefore, E_{b^*} is better than both D and $D_{\tilde{b}}$ for fitting the model in (3.11). Further, if there are nonnegligible third- or fourth-order effects, the aliasing between linear and third-order effects is smaller for E_b^* than $D_{\tilde{b}}$ because $\beta_4(E_b^*) < \beta_4(D_{\tilde{b}})$, and there is no aliasing between linear and fourth-order effects or between second- and third-order effects for E_b^* because $\beta_5(E_b^*) = 0$.

Table 3.5: Part of information matrices $M^{\mathrm{T}}M/25$ corresponding to quadratic and bilinear terms for designs $D_{\tilde{b}}$ and E_{b^*}

$D_{\tilde{b}}$						E_{b^*}					
1	0	0	0	0	0.36	1	0	0	0	0	0.096
0	1	0	0	−0.12	0	0	1	0	0	0.096	0
0	0	1	−0.36	0	0	0	0	1	−0.096	0	0
0	0	−0.36	1	0.3	−0.1	0	0	−0.096	1	0.08	0.08
0	−0.12	0	0.3	1	−0.3	0	0.096	0	0.08	1	−0.08
0.36	0	0	−0.1	−0.3	1	0.096	0	0	0.08	−0.08	1

3.4 Summary

We provide a new class of nonregular designs via the Williams transformation. While two-level nonregular designs have been catalogued by some researchers, the construction of multilevel nonregular designs was rarely studied. The approach in this chapter is a pioneer work in this field. The constructed designs are easily obtained, and shown to have better properties than regular designs.

The Williams transformation is pairwise linear, which is probably the simplest nonlinear transformation, yet it leads to some remarkable results such as Theorems 3.2 and 3.3. It would be of interest to identify and characterize other nonlinear transformations that have similar properties.

The newly obtained designs can be used to generate orthogonal Latin hypercube designs which are commonly used in computer experiments. Orthogonal Latin hypercube designs have been widely studied; see, e.g., Steinberg and Lin (2006), Pang et al. (2009), Lin et al. (2009), Sun et al. (2009), Sun et al. (2010), Lin et al. (2010), Georgiou and Stylianou (2011), Yang and Liu (2012), Wang et al. (2018b), among others. These designs have $\beta_1 = \beta_2 = 0$ therefore guarantee the orthogonality between linear main effects. A popular construction, proposed by Steinberg and Lin (2006) and Pang et al. (2009), is to rotate a regular design to obtain a Latin hypercube design which inherits the orthogonality from both the rotation matrix and the regular design. Wang et al. (2018b) improved the method by rotating a linearly permuted regular design, that is, the $D_{\tilde{b}}$ with $\tilde{b}$ defined in

(3.9). Such generated Latin hypercube designs have $\beta_3 = 0$ thus can guarantee that nonnegligible quadratic and bilinear effects do not contaminate the estimation of linear main effects. With the results in this chapter, rotating the E_{b^*} will lead to better Latin hypercube designs which have zero β_3 and smaller β_4. When nonnegligible third-degree polynomial effects exist, these designs will provide better estimation for linear terms.

3.5 Appendix: Proofs

We need the following lemmas for the proofs.

Lemma 3.3. *The D_b is the same design as $(D_e + \gamma)$ mod q, where $e = b - b^*$, γ is defined as* (3.6)*, and b^* is defined as* (3.7).

Proof. For D_b, permuting all columns x_j to $x_j - \gamma$ for $j = 1, \ldots, n$ is equivalent to keeping the independent columns unchanged while permuting the dependent columns $x_{n-m+i} + b_i$ to $x_{n-m+i} + b_i - b_i^*$ for $i = 1, \ldots, m$. Hence, $D_b - \gamma$ is the same design as D_e with $e = b - b^*$. Equivalently, D_b is the same design as $D_e + \gamma$ mod q. $\square$

Lemma 3.4. *If x is a real number which is not an integer, then*

$$\sum_{n=-\infty}^{\infty} \frac{(-1)^{n-1}}{(n+x)^2} = \frac{\pi^2 \cos \pi x}{(\sin \pi x)^2}.$$

Proof. It is known that $\sum_{n=-\infty}^{\infty} 1/(n+x)^2 = \pi^2/(\sin \pi x)^2$. Then

$$\sum_{n=-\infty}^{\infty} \frac{(-1)^{n-1}}{(n+x)^2} = \sum_{n=-\infty}^{\infty} \frac{1}{(n+x)^2} - 2 \sum_{\text{even } n} \frac{1}{(n+x)^2} = \frac{\pi^2}{(\sin \pi x)^2} - \frac{1}{2} \frac{\pi^2}{(\sin(\pi x/2))^2} = \frac{\pi^2 \cos \pi x}{(\sin \pi x)^2}.$$

$\square$

Lemma 3.5. *Let $p_1(x) = \rho[x - (q-1)/2]$ be the linear orthogonal polynomial, where $\rho = \sqrt{12/[(q+1)(q-1)]}$. Then for $x = 0, \ldots, q-1$,*

$$p_1(x) = -\frac{\rho}{2q} \sum_{v=0}^{q-1} g(v) \cos\left\{\frac{(2v+1)\pi(x+0.5)}{q}\right\}.$$

where

$$g(v) = \frac{\cos(\pi(v+0.5)/q)}{\{\sin(\pi(v+0.5)/q)\}^2}. \tag{3.12}$$

Proof. For $x \in (0, q)$, the Fourier-cosine expansion of $x - q/2$ is given by

$$x - \frac{q}{2} = \sum_{v=1}^{\infty} a_v \cos\left(\frac{v\pi x}{q}\right),$$

with

$$a_v = \frac{2}{q}\int_0^q \left(x - \frac{q}{2}\right)\cos\left(\frac{v\pi x}{q}\right) dx = \begin{cases} 0, & \text{if } v \text{ is even;} \\ -4q/(v^2\pi^2), & \text{if } v \text{ is odd.} \end{cases}$$

Then

$$\begin{aligned} p_1(x) &= -\frac{4\rho q}{\pi^2} \sum_{\text{odd } v>0} \frac{1}{v^2} \cos\left(\frac{v\pi(x+0.5)}{q}\right) \\ &= -\frac{2\rho q}{\pi^2} \sum_{v=-\infty}^{\infty} \frac{1}{(2v+1)^2} \cos\left\{\frac{(2v+1)\pi(x+0.5)}{q}\right\} \\ &= -\frac{2\rho q}{\pi^2} \sum_{k=-\infty}^{\infty} \sum_{v=0}^{q-1} \frac{1}{(2kq+2v+1)^2} \cos\left\{\frac{(2kq+2v+1)\pi(x+0.5)}{q}\right\}. \end{aligned}$$

Since for any integers k and x,

$$\cos\left\{\frac{(2kq+2v+1)\pi(x+0.5)}{q}\right\} = (-1)^k \cos\left\{\frac{(2v+1)\pi(x+0.5)}{q}\right\},$$

we have

$$p_1(x) = -\frac{2\rho q}{\pi^2} \sum_{v=0}^{q-1} \sum_{k=-\infty}^{\infty} \frac{(-1)^k}{(2kq+2v+1)^2} \cos\left\{\frac{(2v+1)\pi(x+0.5)}{q}\right\}.$$

By Lemma 3.4 and (3.12), we have

$$p_1(x) = -\frac{\rho}{2q} \sum_{v=0}^{q-1} g(v) \cos\left\{\frac{(2v+1)\pi(x+0.5)}{q}\right\}.$$

$\square$

Proof of Theorem 3.1. Denote $e = b - b^*$ and $D_e = (y_{ij})$. By Lemma 3.3, D_b is the same design as $(D_e + \gamma)$ mod q, so $E_b = W(D_b) = W(D_e + \gamma)$. By Lemma 3.5,

$$\begin{aligned} p_1(W(x)) &= -\frac{\rho}{2q} \sum_{v=0}^{q-1} g(v) \cos\left\{\frac{(2v+1)\pi(W(x)+0.5)}{q}\right\} \\ &= -\frac{\rho}{2q} \sum_{v=0}^{q-1} g(v) \cos\left\{\frac{(2v+1)\pi(2x+0.5)}{q}\right\} \end{aligned}$$

because $\cos\{(2v+1)\pi(W(x)+0.5)/q\} = \cos\{(2v+1)\pi(2x+0.5)/q\}$ for any integer v. Then we have

$$\begin{aligned}
\beta_3(E_b) &= \beta_3(W(D_e+\gamma)) \\
&= N^{-2}\sum_{y_1,y_2,y_3}\left|\sum_{i=1}^{N} p_1(W(y_{i1}+\gamma))p_1(W(y_{i2}+\gamma))p_1(W(y_{i3}+\gamma))\right|^2 \\
&= N^{-2}\left(\frac{\rho}{2q}\right)^6 \sum_{y_1,y_2,y_3}\left|\sum_{v_1=0}^{q-1}\sum_{v_2=0}^{q-1}\sum_{v_3=0}^{q-1} g(v_1)g(v_2)g(v_3)S(y,v)\right|^2, \qquad (3.13)
\end{aligned}$$

where $\sum_{y_1,y_2,y_3}$ sums over all three different columns y_1, y_2, y_3 in D_e, $y_j = (y_{1j},\ldots,y_{Nj})$ for $j = 1, 2, 3$, and

$$\begin{aligned}
S(y,v) &= \sum_{i=1}^{N}\prod_{j=1}^{3}\cos\left\{\frac{(2v_j+1)\pi(2y_{ij}+2\gamma+0.5)}{q}\right\} \\
&= \sum_{i=1}^{N}\prod_{j=1}^{3}(-1)^{(q+1)/2+v_j}\sin\left\{\frac{2(2v_j+1)\pi y_{ij}}{q}\right\} \\
&= (-1)^{(q+1)/2+v_1+v_2+v_3}\sum_{i=1}^{N}\prod_{j=1}^{3}\sin\left\{\frac{2(2v_j+1)\pi y_{ij}}{q}\right\}.
\end{aligned}$$

If $b = b^*$, $e = 0$ and $D_e = D$. Because D is a regular design, it is a linear space over Z_q. Thus, $(q-y_{i1},\ldots,q-y_{in}) \in D$ whenever $(y_{i1},\ldots,y_{in}) \in D$. Then $S(y,v) = 0$ for any $y = (y_1, y_2, y_3)$ and $v = (v_1, v_2, v_3)$. By (3.13), $\beta_3(E_{b^*}) = 0$. □

Proof of Theorem 3.2. Following the proof of Theorem 1, if $b \neq b^*$, then $e = b - b^*$ has nonzero components. Since D is ordinary-recursive, there exist three columns, say z_1, z_2, z_3, in D such that $z_3 = c_1 z_1 + c_2 z_2$, $c_1 = 1$ or -1, $c_2 \in Z_q$, and z_1, z_2 and $z_3 + e_0$ are three columns in D_e, where e_0 is a nonzero component of e. We only consider $c_1 = 1$ below as the proof for $c_1 = -1$ is similar. Let d be the design formed by z_1, z_2, and $z_3 + e_0$. By (3.13), we only need to show that $\beta_3(W(d)) \neq 0$. Note that

$$\beta_3(W(d)) = N^{-2}\left(\frac{\rho}{2q}\right)^6 \left|\sum_{v_1=0}^{q-1}\sum_{v_2=0}^{q-1}\sum_{v_3=0}^{q-1}(-1)^{v_1+v_2+v_3} g(v_1)g(v_2)g(v_3)S(z,v)\right|^2, \qquad (3.14)$$

where $g(v)$ is defined in (3.12), and

$$S(z,v) = \sum_{i=1}^{N}\sin\left(\frac{2(2v_1+1)\pi z_{i1}}{q}\right)\sin\left(\frac{2(2v_2+1)\pi z_{i2}}{q}\right)\sin\left(\frac{2(2v_3+1)\pi(z_{i3}+e_0)}{q}\right).$$

By applying the product-to-sum identities twice, we have

$$\begin{aligned}S(z,v) \quad &= \frac{1}{4}\Bigg\{\sum_{i=1}^{N}\sin\left(\frac{2\pi(t_1 z_{i1} - t_4 z_{i2} + (2v_3+1)e_0)}{q}\right)\\ &+\sum_{i=1}^{N}\sin\left(\frac{2\pi(t_2 z_{i1} + t_4 z_{i2} - (2v_3+1)e_0)}{q}\right)\\ &-\sum_{i=1}^{N}\sin\left(\frac{2\pi(t_1 z_{i1} + t_3 z_{i2} + (2v_3+1)e_0)}{q}\right)\\ &-\sum_{i=1}^{N}\sin\left(\frac{2\pi(t_2 z_{i1} - t_3 z_{i2} - (2v_3+1)e_0)}{q}\right)\Bigg\},\end{aligned} \tag{3.15}$$

where $t_1 = 2(v_1+v_3)+2$, $t_2 = 2(v_1-v_3)$, $t_3 = 2(v_2+v_3c_2)+c_2+1$, and $t_4 = 2(v_2-v_3c_2)-c_2+1$. Let

$$v_{10} = q-1-v_3 \text{ and } v_{20} = v_3c_2 + (c_2-1)(q+1)/2 \bmod q. \tag{3.16}$$

When $v_1 = v_{10}$ and $v_2 = v_{20}$, $t_1 = t_4 = 0 \bmod q$ and the first item in the right hand side of (3.15), $\sum_{i=1}^{N}\sin\left(2\pi(t_1 z_{i1} - t_4 z_{i2} + (2v_3+1)e_0)/q\right)$, equals $N\sin(2\pi(2v_3+1)e_0/q)$. When $v_1 \neq v_{10}$ or $v_2 \neq v_{20}$, the item is zero. By similar analysis to other items in (3.15), we have

$$S(z,v) = \begin{cases} \frac{N}{4}\sin\left\{\frac{2\pi(2v_3+1)e_0}{q}\right\}, & \text{if } (v_1,v_2) = (v_{10},v_{20}) \text{ or } (q-1-v_{10}, q-1-v_{20});\\ -\frac{N}{4}\sin\left\{\frac{2\pi(2v_3+1)e_0}{q}\right\}, & \text{if } (v_1,v_2) = (v_{10},q-1-v_{20}) \text{ or } (q-1-v_{10}, v_{20});\\ 0, & \text{otherwise.}\end{cases}$$

Note that $g(q-1-v) = -g(v)$ for any v. Then by (3.14),

$$\beta_3(W(d)) = \left(\frac{\rho}{2q}\right)^6 \left|\sum_{v_3=0}^{q-1}(-1)^{v_3c_2} g(v_{20})(g(v_3))^2 \sin\left\{\frac{2\pi(2v_3+1)e_0}{q}\right\}\right|^2, \tag{3.17}$$

where v_{20} is defined in (3.16). Applying $g(q-1-v) = -g(v)$ again, we can simply (3.17) as

$$\beta_3(W(d)) = \frac{\rho^6}{16q^6}\left|\sum_{v_3=0}^{(q-1)/2}(-1)^{v_3c_2} g(v_{20})(g(v_3))^2 \sin\left\{\frac{2\pi(2v_3+1)e_0}{q}\right\}\right|^2. \tag{3.18}$$

By considering the Taylor expansion of $g(v)$, we can see that the sum in (3.18) is dominated by the first two items with $v_3 = 0$ and $v_3 = 1$. It can be verified that (3.18) is nonzero for $e_0 = 1, \ldots, q-1$. This completes the proof. □

Proof of Theorem 3.3. We need to show that for any run $W(x_1, \dots, x_n)$ in E_{b^*}, $(q-1)-W(x_1, \dots, x_n)$ also belongs to E_{b^*}. This is equivalent to show that for each run $(x_1, \dots, x_n)$ in D_{b^*}, $W^{-1}(q-1-W(x_1, \dots, x_n))$ also belongs to D_{b^*}. Since the design D contains the zero point $(0, \dots, 0)$, by Lemma 3.3, D_{b^*} contains the point $(\gamma, \dots, \gamma)$. Because all design points of D form a linear space and D_b is a coset of D, then $\gamma - (x_1, \dots, x_n)$ belongs to the null space of D_{b^*}. Hence, $\gamma - (x_1, \dots, x_n) + \gamma = 2\gamma - (x_1, \dots, x_n)$ belongs to D_{b^*}. For $x = 0, \dots, q-1$,

$$W^{-1}(x) = \begin{cases} x/2, & \text{for even } x; \\ q-(x+1)/2, & \text{for odd } x, \end{cases}$$

and

$$\begin{aligned} W^{-1}(q-1-x) &= \begin{cases} (q-1)/2 - W^{-1}(x), & \text{for even } x; \\ (3q-1)/2 - W^{-1}(x), & \text{for odd } x, \end{cases} \\ &= 2\gamma - W^{-1}(x). \end{aligned}$$

Then $W^{-1}(q-1-W(x_1, \dots, x_n)) = 2\gamma - (x_1, \dots, x_n)$. Hence, $W^{-1}(q-1-W(x_1, \dots, x_n))$ belongs to D_{b^*}. This completes the proof. $\square$

CHAPTER 4

Orthogonal Array-Based Subdata Selection for Big Data Regression

The dramatic growth of large datasets has enabled the study of many scientific problems. While we are taking advantages of big data, in many applications, however, labelling all data points is infeasible due to the limit of time and budget. We are often encountered with the problem where we are given a large data set of n data points but can only observe a small subset of $k < n$ labels. Wang et al. (2017) considered three application examples which cover material synthesis, CPU benchmarking, and wind speed prediction. In all examples, collecting labels for data points is either time-consuming or costly, so only a subset of data points can be labeled. An intuitive solution is to randomly select k points to label, while this may end up with a big loss of information of the full big data. The selection of an informative subdata set is crucial.

In this chapter, we develop an orthogonal array (OA)-based method for subdata selection. The method is inspired by the fact that an OA of two levels is D-, A-, and G-optimal for linear regression. We define a discrepancy to measure how well a subdata set approximates an OA. Based on the discrepancy, we develop an algorithm which sequentially selects data points as well as eliminating points from the full data to reduce the number of candidate points and speed up the selecting process. Simulation results show that the algorithm outperforms existing methods in minimizing mean squared errors of parameter estimations and maximizing D- and A-efficiencies of the design matrices.

4.1 The framework

We consider the linear regression problem

$$y = \tilde{X}\beta + \varepsilon, \tag{4.1}$$

where $y = (y_1, \ldots, y_n)^T$ is a vector of all observations, $\tilde{X} = (1, X)$ is the design matrix, and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a vector of parameters. When using the full data (X, y), the least-squares (LS) estimator of β is

$$\hat{\beta} = (\tilde{X}^T\tilde{X})^{-1}(\tilde{X}^T y).$$

Now consider taking a subdata set of size k from the full data. Denote the subdata as (X_s, y_s). Then the LS estimator based on the subdata is given by

$$\hat{\beta}_s = (\tilde{X}_s^T\tilde{X}_s)^{-1}(\tilde{X}_s^T y_s),$$

where $\tilde{X}_s = (1, X_s)$. The covariance matrix of $\hat{\beta}_s$ is $\sigma^2 M^{-1}$ with

$$M_s = \tilde{X}_s^T\tilde{X}_s.$$

To minimize the variance of $\hat{\beta}_s$, we seek the subdata X_s which, in some sense, maximizes M_s. This is typically done, in optimal experimental design strategy, by minimizing an optimality function of the matrix M_s^{-1}. Denote ψ as the optimality function, then we want to find the X_s that minimizes $\psi(M_s^{-1})$. Denote $\xi = (\xi_1, \ldots, \xi_n)$ as the indicator vector that signifies whether the data points in X are included in X_s or not, that is, $\xi_i = 1$ if the ith data point in X is included in X_s and $\xi_i = 0$ otherwise, then $\sum_{i=1}^n \xi_i = k$ where k is the number of data points in X_s. With the help of ξ, M_s can be rewritten as

$$M_s = M_s(\xi) = \tilde{X}^T\text{diag}(\xi)\tilde{X}.$$

Then the problem can be presented as the following optimization problem:

$$\begin{aligned} \xi^* &= \arg\min_{\xi} \psi\left\{M_s^{-1}\right\} = \arg\min_{\xi} \psi\left\{\left(\tilde{X}^T\text{diag}(\xi)\tilde{X}\right)^{-1}\right\}, \\ s.t. & \quad \sum_{i=1}^{n} \xi_i = k. \end{aligned} \tag{4.2}$$

Popular optimality criteria include the D-optimality criterion that minimizes the determinant of M_s^{-1}, A-optimality criterion that minimizes the trace of M_s^{-1}, and G-optimality criterion that minimizes the maximum entry in the diagonal of the hat matrix $\tilde{X}_s M_s^{-1} \tilde{X}_s^T$. Wang et al. (2017) considers the A-optimality and proposes a computationally tractable stochastic subsampling algorithm. They consider a continuous relaxation of the combinatorial optimization problem in (4.2) to get an optimal probability following which a stochastic sample is then drawn. Wang et al. (2018a) considers the D-optimality and proposes an information-based optimal subdata selection (IBOSS) method. They approximate the optimality by only including data points with extreme (largest and smallest) covariate values into X_s to maximize the diagonal entries of M_s without any consideration of the off-diagonal entries (that is, correlation between variables). Both methods try to approximate the combinatorial optimality in some sense.

4.2 Orthogonal arrays

Recall that a two-level orthogonal array (OA) with strength t is an $n \times p$ matrix in which all 2^t level combinations appear equally often in every $n \times t$ submatrix. In this chapter, the two levels are denoted by -1 and 1. The following theorem shows that ideally, a subdata set X_s is D-optimal if and only if X_s forms a two-level orthogonal array.

Theorem 4.1. *Suppose all covariates are scaled to* $[-1, 1]$. *For a subdata set* X_s *of size* k*,*

$$\det(M_s^{-1}) \geq \frac{1}{k^{p+1}},$$

and the equality holds if and only if X_s *forms a two-level OA with levels from* $\{-1, 1\}$ *and strength* $t \geq 2$.

The following theorem shows that ideally, a subdata set X_s is A-optimal if and only if X_s forms a two-level orthogonal array.

Theorem 4.2. *Suppose all covariates are scalled to* $[-1, 1]$. *For a subdata set* X_s *of size* k*, denote the eigenvalues of* M_s^{-1} *as* $\lambda_0(M_s^{-1}), \lambda_1(M_s^{-1}), \ldots, \lambda_p(M_s^{-1})$*, then*

$$\sum_{j=0}^{p} \lambda_j(M_s^{-1}) \geq \frac{p+1}{k},$$

and the equality holds if and only if X_s forms a two-level OA with levels from $\{-1, 1\}$ and strength $t \geq 2$.

Kiefer and Wolfowitz (1959, 1960) showed the equivalence theorem between D- and G-optimality. Thus, we can have the following result regarding the G-optimality.

Theorem 4.3. *Suppose all covariates are scalled to $[-1, 1]$. For a subdata set X_s of size k and any point $x \in [-1, 1]^p$, denote $\tilde{x} = (1, x^T)^T$ and*

$$d(x, \xi) = \tilde{x}^T M_s^{-1} \tilde{x},$$

then

$$\max_x d(\xi) \geq p + 1,$$

and the equality holds if and only if X_s forms a two-level OA with levels from $\{-1, 1\}$ and strength $t \geq 2$.

Theorems 4.1–4.3 show that subdata forming OAs are universally optimal in all of the criteria. Often the full data do not contain any subset of k points forming an OA, so that the lower bounds in Theorems 4.1–4.3 would not be attained. However, we can always find a subset approximating an OA. We will introduce an algorithm in the next section.

4.3 A sequential addition-elimination algorithm

To select subdata X_s following an OA, we need to define a discrepancy function that measures the similarity between X_s and an OA. Considering that data points selected following an OA should have two features: (i) they are at the corners of the data region, and (ii) their signs are as dissimilar as possible. Therefore, the discrepancy function should contain two parts corresponding to the two features. For Feature (i), it is intuitive to maximize $\|x_i\|$ for any point $x_i = (x_{i1}, \ldots, x_{ip})$ included in X_s, where $\| \cdot \|$ is the Euclidean norm. For Feature (ii), denote $s(x)$ as the sign of x and $s(x_i) = (s(x_{i1}), \ldots, s(x_{ip}))$. Define

$$\delta(x, y) = \begin{cases} 1, & \text{if } x = y; \\ 0, & \text{otherwise,} \end{cases} \tag{4.3}$$

and let $\delta(s(x_i), s(x_j)) = \sum_{l=1}^{p} \delta(s(x_{il}), s(x_{jl}))$ for any two points x_i and x_j. Then $\delta(s(x_i), s(x_j))$ is the number of components in x_i and x_j that have the same signs. We want to minimize a function of $\delta(s(x_i), s(x_j))$. Based on these considerations, we define a D_2-discrepancy criterion as

$$D_2(X_s) = \sum_{1 \le i < j \le k} [\delta(s(x_i), s(x_j)) + p - \|x_i\|^2/2 - \|x_j\|^2/2]^2. \tag{4.4}$$

The following result shows an important lower bound of $D_2(X_s)$.

Theorem 4.4. *For a subdata set X_s,*

$$D_2(X_s) \ge \frac{k^2 p(p+1) - 4kp^2}{8},$$

with equality if and only if X_s forms a two-level OA with levels from $\{-1, 1\}$ and strength $t \ge 2$.

Now the subdata selection based on OAs can be presented as the following optimality problem:

$$\begin{aligned} X_s^* &= \arg\min_{X_s} D_2(X_s), \\ & s.t. \quad X_s \text{ contains } k \text{ points}. \end{aligned}$$

This optimality problem is combinatorial in nature and the optimal subset X_s^* is difficult to get. An exhaustive search over all possible X_s of size k requires $O(n^k k^2 p)$ operations, which is infeasible for even moderate X and X_s. We propose a sequential addition-elimination algorithm that approximately achieves the optimality. The algorithm selects data points iteratively as well as eliminating candidate points from X to speed up the search.

Now suppose we are at the ith iteration where X_s^i is the new matrix obtained by adding x_i^* to X_s^{i-1}, $i = 1, \ldots, k-1$. Then by (4.4),

$$D_2(X_s^i) = \sum_{j=1}^{i-1} D_2(x_i^*, x_j^*) + D_2(X_s^{i-1})$$

where

$$D_2(x_i^*, x_j^*) = [\delta(s(x_i^*), s(x_j^*)) + p - \|x_i^*\|^2/2 - \|x_j^*\|^2/2]^2$$

is the D_2-score of x_i^* relative to x_j^*. To minimize $D_2(X_s^i)$, select x_i^* which minimizes the sum of the scores, that is,

$$x_i^* = \arg\min_{x \in X} \sum_{j=1}^{i-1} D_2(x, x_j^*). \tag{4.5}$$

The computational complexity for choosing the x_i^* following (4.5) is $O(nip)$. However, note that $D_2(x, x_j^*)$ for $j = 1, \dots, i-2$ was already calculated in the $(i-1)$th iteration when searching for x_{i-1}^*. Thus, for the current iteration, only the computation of $D_2(x, x_{i-1}^*)$ is required so the computational complexity is reduced to $O(np)$ in each iteration. To further reduce the computation, we can delete some data points in X with large values of $\sum_{j=1}^{i-1} D_2(x, x_j^*)$ so that these points will not be considered in the $(i+1)$th iteration. The algorithm proceeds as follows. Suppose each variable of X is scaled to $[-1, 1]$.

Algorithm 4.1. *[Sequential addition-elimination]*

Step 1. [Initiation] Let $i = 1$. Find the point in X with the largest Euclidean norm, denoted as x_1^*. Include x_1^* in X_s and remove it from X. Let $\mathscr{D} = (0, \dots, 0)$ being an $(n-1)$-vector with each component corresponding to each data point in X.

Step 2. [Addition] Increase i by 1. For each $x \in X$, add the D_2-score

$$D_2(x, x_{i-1}^*) = [\delta(s(x), s(x_{i-1}^*)) + p - \|x\|^2/2 - \|x_{i-1}^*\|^2/2]^2 \tag{4.6}$$

to the corresponding component in $\mathscr{D}$. Find x_i^* with the smallest component in $\mathscr{D}$ and add it to X_s.

Step 3. [Elimination] Keep $t = \lfloor n/i \rfloor$ points in X with t smallest components in $\mathscr{D}$. Remove x_i^* and other points from X as well as their corresponding components from $\mathscr{D}$.

Step 4. [End] Iterate Steps 2 and 3 until X_s contains k points.

By Theorem 4.2, X_s minimizes the average eigenvalue of M_s^{-1}, that is, the average variance of coefficient estimations, if it forms an OA. Therefore, it is easy to see that Algorithm 4.1 tends to generate subdata X_s which minimize the sum of the variances of coefficient estimations when n is large. Note that in the Addition step, X consists $t = \lfloor n/i \rfloor$ points so the computational complexity for finding x_i^* is $O(np/i)$. Therefore, the complexity for selecting k data points is $O(np/1) + \cdots + O(np/k) = O(np \log k)$.

To examine the performance of the proposed algorithm, simulations are conducted and empirical mean squared errors (MSE) for the slope parameters are calculated using

$$\text{MSE} = S^{-1} \sum_{s=1}^{S} \|\hat{\beta}^s - \beta\|^2, \tag{4.7}$$

where S is the number of times a simulation is repeated and $\hat{\beta}^s$ is the estimate of slope parameters in the sth repetition. Other than that, we also calculate the D- and A-efficiencies using

$$D_{eff} = \{(1/k^{p+1})/[\det(M_s)]^{-1}\}^{1/(p+1)} = \det(M_s)^{1/(p+1)}/k \tag{4.8}$$

and

$$A_{eff} = [(p+1)/k] / \sum_{j=0}^{p} \lambda_j(M_s^{-1}) = (p+1)/[k \sum_{j=0}^{p} \lambda_j(M_s^{-1})] \tag{4.9}$$

by noting the lower bounds shown in Theorems 4.1 and 4.2.

The following toy example illustrates the subdata selected by the algorithm.

Example 4.1. *Consider selecting $k = 100$ data points from a full data set with $n = 1000$ points. Let $p = 2$ and each point $x_i \sim$ Unif$[-1, 1]^2$ where Unif$[-1, 1]^2$ is a uniform distribution on $[-1, 1]^2$. The response y is generated through the model*

$$y_i = 1 + x_{i1} + x_{i2} + \varepsilon$$

where $\varepsilon \sim N(0, 1)$. Figure 4.1 shows the subdata selected by the IBOSS (Wang et al., 2018a) and the Sequential addition-elimination algorithm (OA-based). The IBOSS chooses boundary points while the proposed algorithm chooses data points at the corners. Figure 4.2 shows the MSE, D-efficiency, and A-efficiency for the subdata selected by Uniform subsampling, IBOSS, and the proposed algorithm (OA-based). The proposed algorithm outperforms the other two methods in each of the criteria. This is because corner points selected by the proposed algorithm are more informative for linear models thus form more efficient subdata and provide better estimation for parameters.

Example 4.1 is a toy example with $p = 2$ from which we can tell some outperformance of the proposed method than available methods. We will see from more numerical results that the

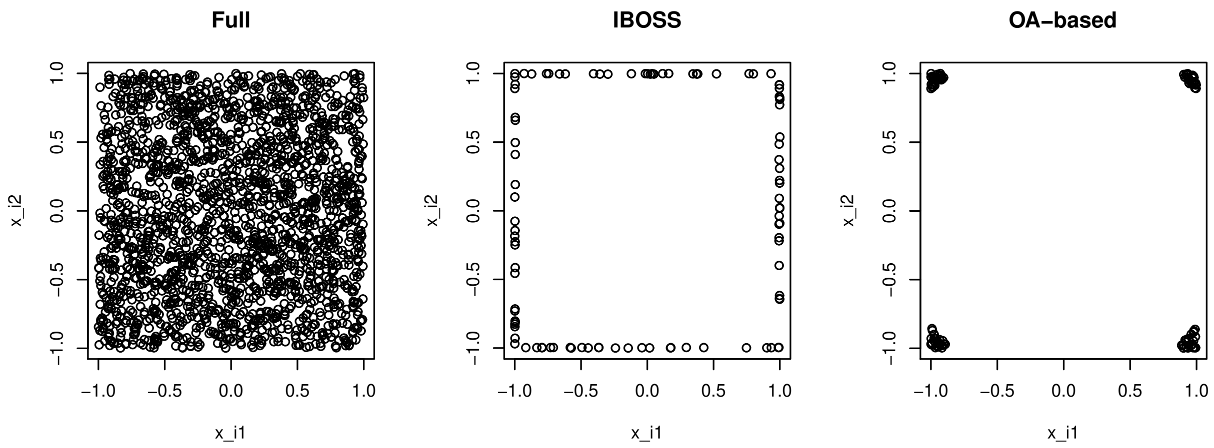


Figure 4.1: The subdata selected by IBOSS and OA-based methods.

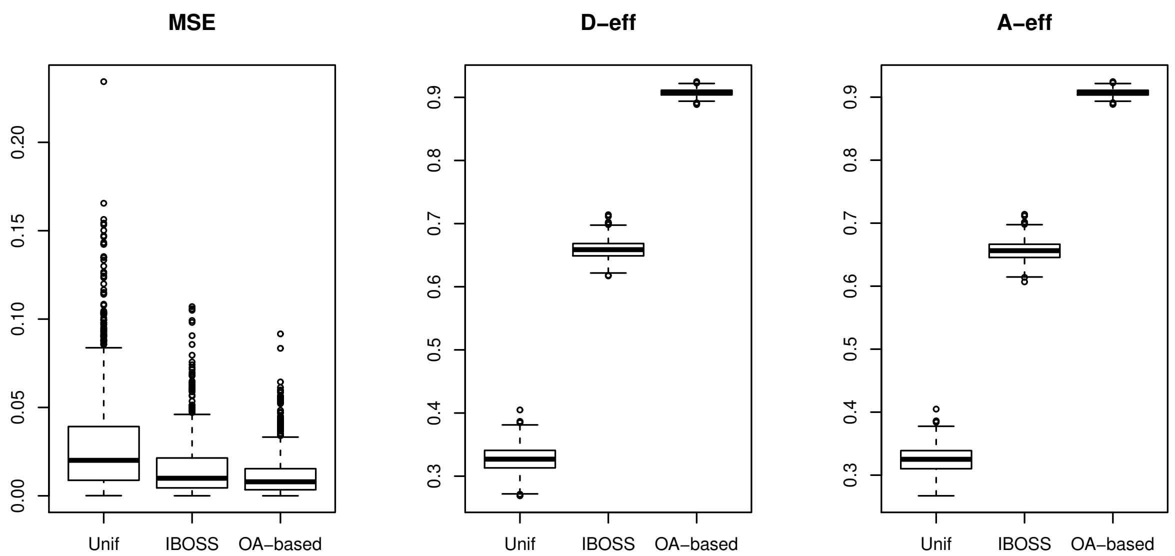


Figure 4.2: The MSE, D- and A-efficiencies for the subdata selected by different methods

proposed algorithm performs much better for larger p. Note that Figure 4.1 may raise questions about potential outliers because it seems that the selected subdata by the proposed algorithm only capture extreme covariate values. However, as we will see in more numerical results, this is not the case for moderate and large p. As p increases, the selected subdata are spreading out over the data region, so the issue of potential outliers does not really exist with the proposed algorithm.

4.4 Model with interactions

Consider the linear regression

$$y = \tilde{X}\tilde{\beta}_1 + X^{inter}\tilde{\beta}_2 + \varepsilon, \tag{4.10}$$

where $y = (y_1, \ldots, y_n)^T$ is a vector of all observations, $\tilde{X} = (1, X)$, X^{inter} contains all interaction terms, that is, each column of X^{inter} is an element-wise product of two columns in X, $\tilde{\beta}_1 = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a vector of linear effects, and $\tilde{\beta}_2 = (\beta_{12}, \ldots, \beta_{(p-1)p})^T$ is a vector of interactions effects. The information matrix for the model in (4.10) is given by

$$M = (\tilde{X}, X^{inter})^T(\tilde{X}, X^{inter}),$$

and the information matrix with a subdata set X_s is given by

$$M_s = (\tilde{X}_s, X_s^{inter})^T(\tilde{X}_s, X_s^{inter}), \tag{4.11}$$

where $\tilde{X}_s = (1, X_s)$ and X_s^{inter} contains all columns which are element-wise product between columns of X_s. A similar result to Theorem 4.1 applies here.

Theorem 4.5. *Suppose all covariates are scaled to* $[-1, 1]$ *and* M_s *is defined in* (4.11)*. For a subdata set* X_s *of size k,*

$$\det(M_s^{-1}) \geq \frac{1}{k^{p(p+1)/2+1}},$$

and the equality holds if and only if X_s *forms a two-level OA with levels from* $\{-1, 1\}$ *and strength* $t \geq 4$.

Theorem 4.5 shows that a two-level OA with strength $t \geq 4$ is D-optimal for models with interactions. We can establish similar results for A- and G-optimality as in Theorems 4.2 and

4.3, which is tedious thus omitted here. Theorem 4.5 indicates that we should follow an OA with strength $t \geq 4$ to select subdata for models with interactions. To do this, define

$$D_4(X_s) = \sum_{1 \leq i < j \leq k} [\delta(s(x_i), s(x_j)) + p - \|x_i\|^2/2 - \|x_j\|^2/2]^4. \tag{4.12}$$

The difference between D_2 in (4.4) and D_4 in (4.12) is only at the power taken for the discrepancy, while the following result shows that D_4 is able to measure the similarity between a subdata set X_s and an OA with strength $t \geq 4$.

Theorem 4.6. *For a subdata set X_s,*

$$D_4(X_s) \geq \frac{k^2 p(p^3 + 6p^2 + 3p - 2) - 16kp^4}{16},$$

with equality if and only if X_s forms a two-level OA with levels from $\{-1, 1\}$ and strength $t \geq 4$.

Theorem 4.6 shows that D_4 is powerful as a criterion for selecting subdata for models with interactions. Therefore, Algorithm 4.1 can be applied to the subdata selection by replacing the D_2 with D_4, that is, replacing the criterion in (4.6) with

$$D_4(x, x_{i-1}^*) = [\delta(s(x), s(x_{i-1}^*)) + p - \|x\|^2/2 - \|x_{i-1}^*\|^2/2]^4.$$

4.5 Numerical results

We show simulation results in this section. We consider 7 scenarios. For Case 1–6, data are generated from the linear model in (4.1) with true value of β being a vector of unity and $\sigma^2 = 9$. An intercept is included so β is a $(p+1)$-dimensional vector. For Case 7, data are generated from the model with interactions, that is, the model in (4.10), with true values of $(\tilde{\beta}_1^T, \tilde{\beta}_2^T)^T$ being a $[1 + p(p+1)/2]$-dimensional vector of unity and $\sigma^2 = 9$. Covariates are generated according to the following scenarios.

Case 1. $n = 10000$, $p = 10$, $k = 100$, and x_i's have a multivariate uniform distribution with all covariates independent.

Case 2. $n = 100000$, $p = 50$, $k = 1000$, and x_i's have a multivariate uniform distribution with all covariates independent.

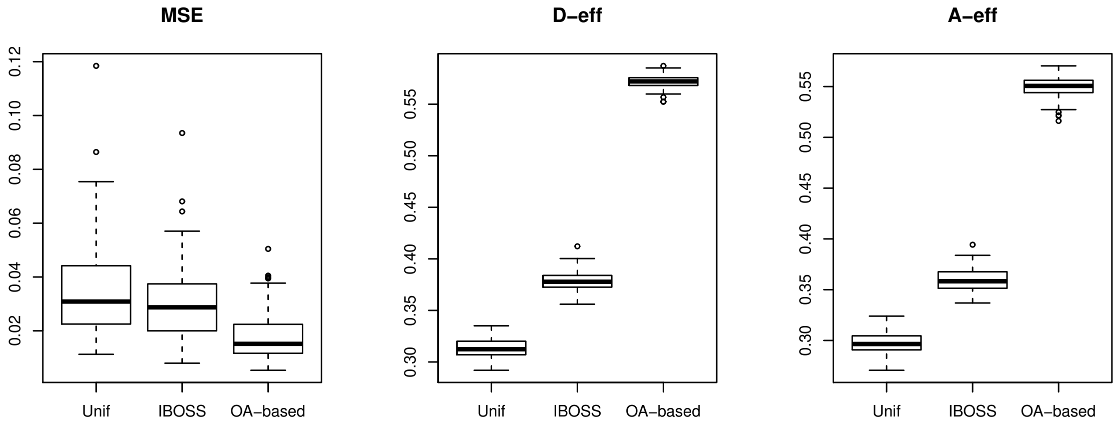


Figure 4.3: MSE, D- and A-efficiencies of X_s selected from different methods for Case 1.

Case 3. $n = 10000$, $p = 10$, $k = 100$, and x_i's have a multivariate normal distribution with all covariates independent.

Case 4. $n = 100000$, $p = 50$, $k = 1000$, and x_i's have a multivariate normal distribution with all covariates independent.

Case 5. $n = 10000$, $p = 10$, $k = 100$, and x_i's have a multivariate normal distribution with covariance matrix $\Sigma = 0.5^{I(i \neq j)}$, where $I()$ is the indicator function.

Case 6. $n = 100000$, $p = 50$, $k = 1000$, and x_i's have a multivariate normal distribution with covariance matrix $\Sigma = 0.5^{I(i \neq j)}$.

Case 7. $n = 10000$, $p = 10$, $k = 100$, and x_i's have a multivariate uniform distribution with all covariates independent.

The simulation is repeated $S = 100$ times. We consider three approaches: Uniform subsampling (Unif), IBOSS algorithm, and the proposed algorithm (OA-based). Empirical mean squared errors (MSE) for the slope parameters, D- and A-efficiencies of X_s are calculated using (4.7), (4.8), and (4.9). Figures 4.3–4.9 show the comparison of the subdata X_s selected from the three different approaches. We can see that the OA-based method outperforms the other two methods for all scenarios. Specifically, the OA-based method performs especially well when the covariates follow a multivariate uniform distribution, which is a common case in many applications. The D- and A-efficiencies of the proposed method are always much larger than the other two methods,

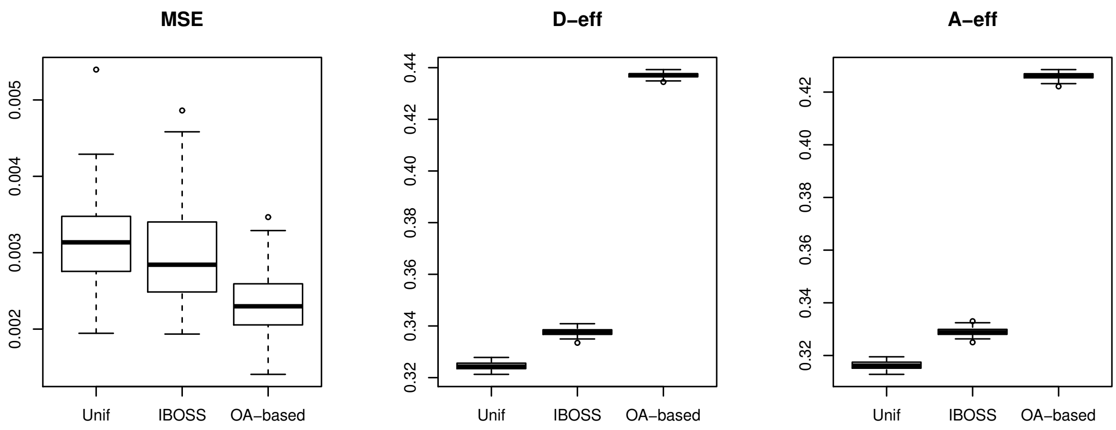


Figure 4.4: MSE, D- and A-efficiencies of X_s selected from different methods for Case 2.

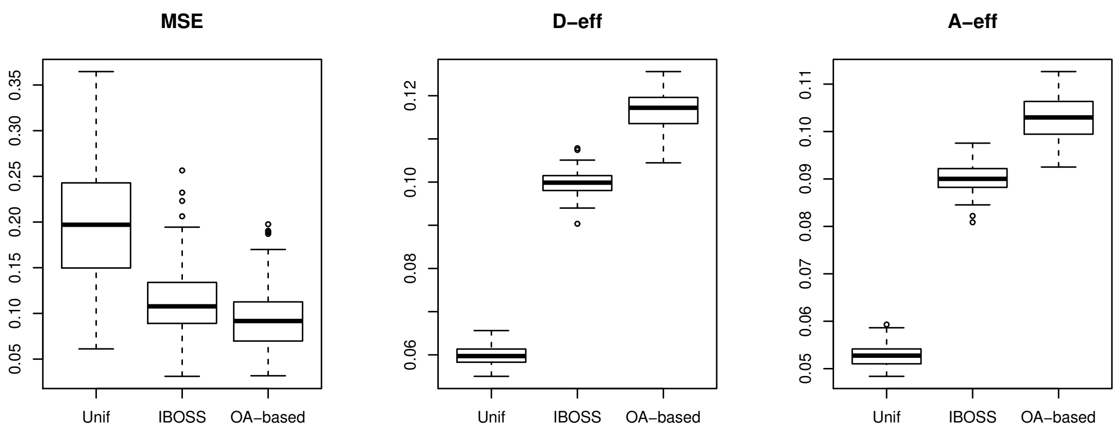


Figure 4.5: MSE, D- and A-efficiencies of X_s selected from different methods for Case 3.

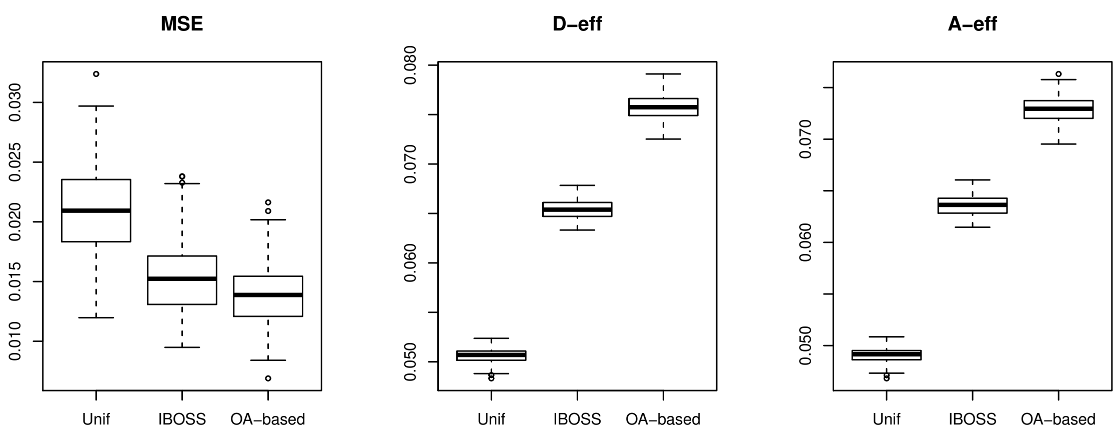


Figure 4.6: MSE, D- and A-efficiencies of X_s selected from different methods for Case 4.

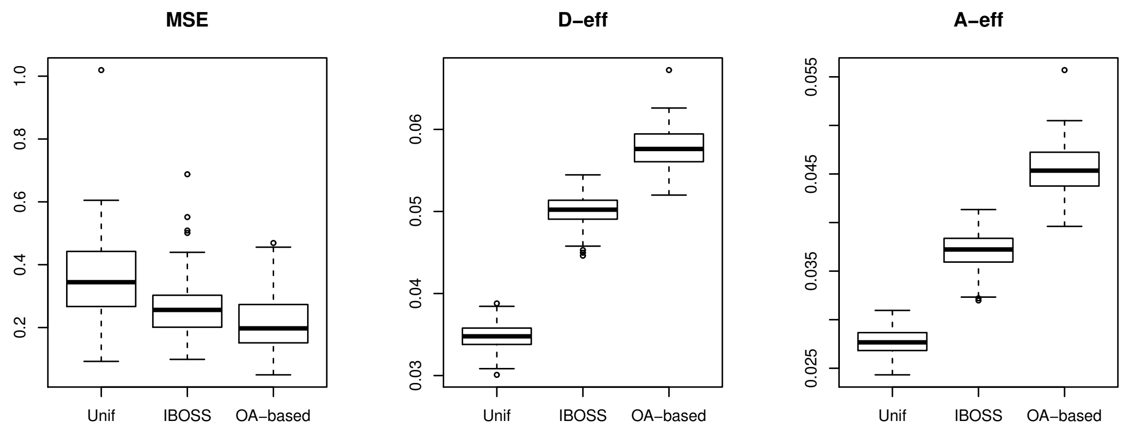


Figure 4.7: MSE, D- and A-efficiencies of X_s selected from different methods for Case 5.

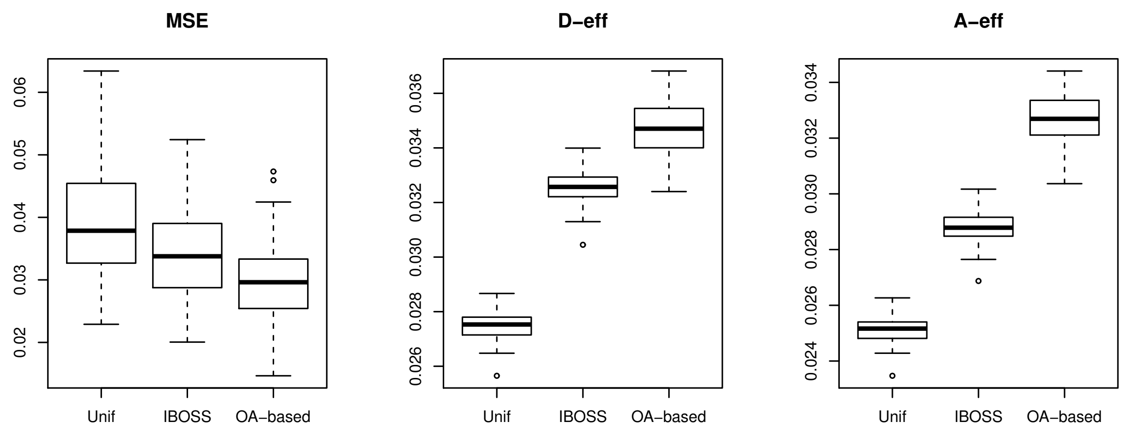


Figure 4.8: MSE, D- and A-efficiencies of X_s selected from different methods for Case 6.

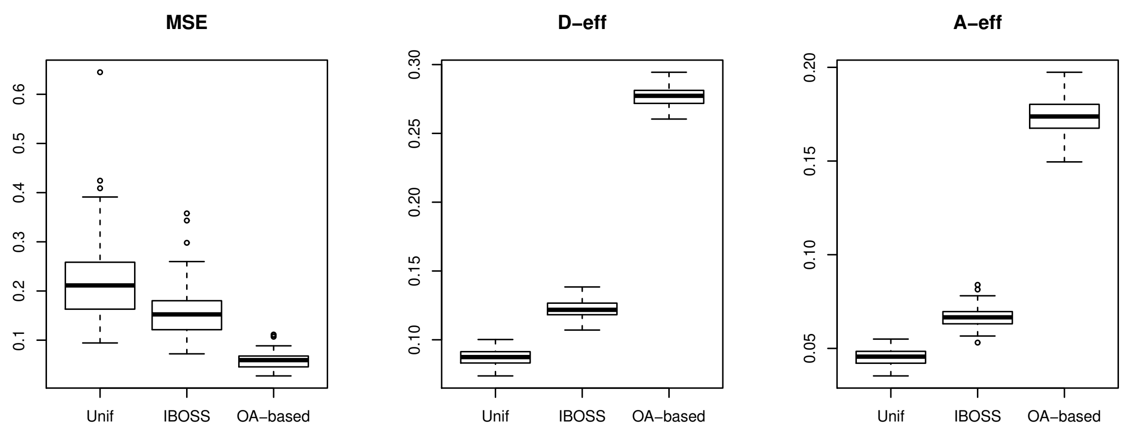


Figure 4.9: MSE, D- and A-efficiencies of X_s selected from different methods for Case 7.

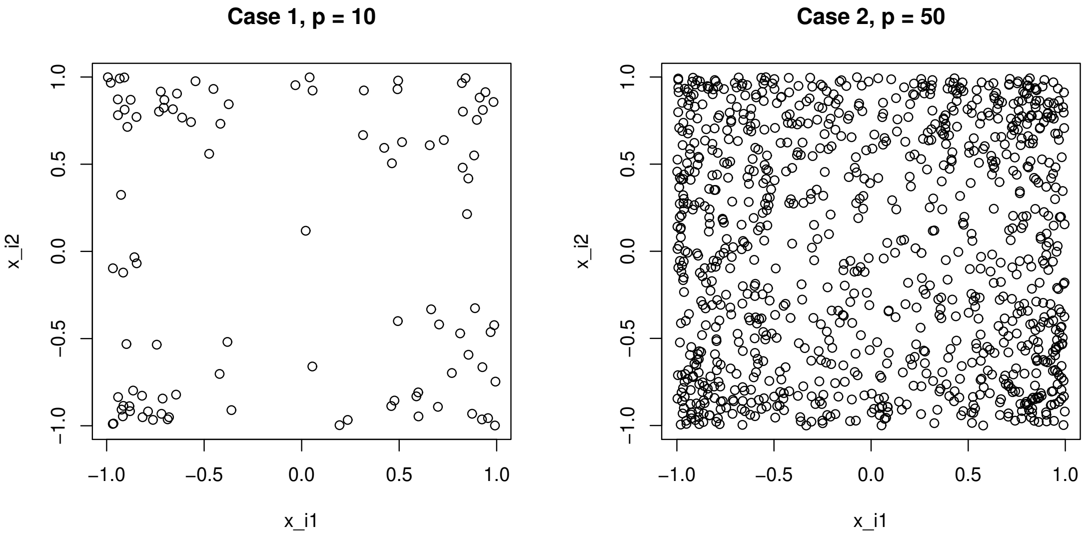


Figure 4.10: Two dimensional projection plot of the subdata selected by the proposed algorithm for Cases 1 and 2.

making the subsdata obtained from the proposed method robust to different settings of the error σ^2.

Figure 4.10 shows the two dimensional projection plot of the subdata selected by the proposed algorithm for Cases 1 and 2. Without loss of generality, we only show the projection onto the first two covariates. As we can see, the selected points are not concentrating on the corners of the region, which is different from the case shown in Example 4.1 for $p = 2$. As p increases, the selected subdata are spreading out over the region, so the issue of potential outliers does not really exist with the proposed algorithm.

4.6 Discussion

We develop a sequential addition-elimination algorithm for subdata selection. The algorithm inherits optimality from OAs and gives approximately optimal subdata for linear regression with or without interactions.

For linear regression with both interactions and quadratic terms, the proposed algorithm also performs better than available methods. To further increase the efficiency of the subdata, following the central composite design strategy (Box and Wilson, 1951), we can add some center and axial points into the subdata. To make the subdata still containing k points, we only select $k_0 < k$ data points from the algorithm and select $k - k_0$ data points close to the center or axes. In our simulation, this modification increases the efficiency of the subdata for small p, say, for $p \leq 4$, while for moderate and large p, the addition of those points usually ends up with a big reduction on the efficiency. This is because, as was shown in the Figure 4.10, the subdata already cover the center and axes for larger p. So the addition of those points would bring a waste of data points instead of new information.

4.7 Appendix: Proofs

Proof of Theorem 4.1. Denote $X_s = (x_{ij}^*)$, then we have

$$M_s = \begin{pmatrix} k & \sum_{i=1}^k x_{i1}^* & \cdots & \sum_{i=1}^k x_{ip}^* \\ \sum_{i=1}^k x_{i1}^* & \sum_{i=1}^k (x_{i1}^*)^2 & \cdots & \sum_{i=1}^k x_{i1}^* x_{ip}^* \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^k x_{ip}^* & \sum_{i=1}^k x_{i1}^* x_{ip}^* & \cdots & \sum_{i=1}^k (x_{ip}^*)^2 \end{pmatrix} \tag{4.13}$$

Because $-1 \leq x_{ij}^* \leq 1$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, p$, $\sum_{i=1}^k (x_{ij}^*)^2 \leq k$ for all j. Thus,

$$\begin{aligned} \det(M_s) &= \prod_{j=0}^{p} \lambda_j(M_s) \\ &\leq \left(\frac{\sum_{j=0}^p \lambda_j(M_s)}{p+1} \right)^{p+1} \qquad (4.14) \\ &= \left(\frac{k + \sum_{j=1}^p \sum_{i=1}^k (x_{ij}^*)^2}{p+1} \right)^{p+1} \\ &\leq k^{p+1}, \qquad (4.15) \end{aligned}$$

where $\lambda_j(M_s)$'s for $j = 0, 1, \ldots, p$ are eigenvalues of M_s, the equality in (4.14) holds if and only if $\lambda_0(M_s) = \lambda_1(M_s) = \cdots = \lambda_p(M_s)$, and the equality in (4.15) holds if and only if x_{ij}^* is either 1 or -1 for all i and j. Therefore, $\det(M_s) = k^{p+1}$ if and only if X_s forms an OA. This completes the proof. □

Proof of Theorem 4.2. Because $\lambda_j(M_s^{-1}) = 1/\lambda_j(M_s)$,

$$\sum_{j=0}^{p} \lambda_j(M_s^{-1}) \geq \frac{(p+1)^2}{\sum_{j=0}^{p} \lambda_j(M_s)}. \tag{4.16}$$

From (4.14) and (4.15), we have $\sum_{j=0}^{p} \lambda_j(M_s) \leq k(p+1)$. Then by (4.16),

$$\sum_{j=0}^{p} \lambda_j(M_s^{-1}) \geq \frac{(p+1)^2}{k(p+1)} = \frac{p+1}{k}.$$

□

Proof of Theorem 4.4. For a two-level $k \times p$ design matrix X with entries from $\{-1, 1\}$, Xu (2003) defined the tth power moment as

$$K_t(X) = \left(\frac{k(k-1)}{2}\right)^{-1} \sum_{1 \leq i < j \leq k} [\delta(d_i, d_j)]^t, \tag{4.17}$$

where d_i for $i = 1, \ldots, k$ is the kth row of X and δ is defined in (4.3). For the second power moment K_2, Xu (2003) showed that

$$K_2(X) \geq \left(\frac{k(k-1)}{2}\right)^{-1} \frac{k^2p(p+1) - 4kp^2}{8},$$

and the equality holds if and only if X is an OA of two levels with strength $t \geq 2$. Note that $\delta(s(x_i), s(x_j)) + p - \|x_i\|^2/2 - \|x_j\|^2/2 \geq \delta(s(x_i), s(x_j))$, then

$$D_2(X_s) \geq \frac{k(k-1)}{2} K_2(s(X_s)) \geq \frac{k^2p(p+1) - 4kp^2}{8},$$

where $s(X_s)$ is the sign matrix of X_s, and the equality holds if and only if X_s is an OA of two levels with strength $t \geq 2$. □

Proof of Theorem 4.6. For the fourth power moment defined in (4.17), Xu (2003) showed that

$$K_4(X) \geq \left(\frac{k(k-1)}{2}\right)^{-1} \frac{k^2p(p^3 + 6p^2 + 3p - 2) - 16kp^4}{16},$$

and the equality holds if and only if X is an OA of two levels with strength $t \geq 4$. Note that $\delta(s(x_i), s(x_j)) + p - \|x_i\|^2/2 - \|x_j\|^2/2 \geq \delta(s(x_i), s(x_j))$, then

$$D_4(X_s) \geq \frac{k(k-1)}{2} K_4(s(X_s)) \geq \frac{k^2p(p^3 + 6p^2 + 3p - 2) - 16kp^4}{16},$$

where $s(X_s)$ is the sign matrix of X_s, and the equality holds if and only if X_s is an OA of two levels with strength $t \geq 4$. □

CHAPTER 5

Conclusion

Space-filling designs and fractional factorial designs are two crucial tools in planning experiments. Space-filling designs spread design points evenly and uniformly in the design domain, so they are suitable for multiple modeling techniques and are model robust. Fractional factorial designs aim at linear models with or without interactions to screen important factorial effects. Both types of designs are commonly used in practical applications depending on different aims of experimenters.

Chapter 2 proposes a series of systematic methods for the construction of space-filling designs via the Williams transformation and its modification. The methods efficiently generate large and high-dimensional designs without any computer search. The generated designs are shown to be optimal under the maximin distance criterion and have small pairwise correlations between variables. Chapter 3 further explores the application of Williams transformation to the construction of nonregular fractional factorial designs. We provide a class of multilevel nonregular designs by manipulating nonlinear level permutations on regular designs via the Williams transformation. While two-level nonregular designs have been catalogued by some researchers, the construction of multilevel nonregular designs was rarely studied. The approach in Chapter 3 is a pioneer work in this field. The constructed designs are easily obtained, and shown to have better properties than regular designs.

In viewing that data-driven modeling is gaining more ground as one of the best tools in decision-making processes, we explore the extension of experimental design strategies into data-driven problems. The analysis of big data usually involves critical issues in computation and storage, and an intuitive way to solve the issues is to only store and analyse an informative subsample instead of the full data. There are a couple of pioneer works in this field, while further exploration is still in high demand. Chapter 4 studies the subdata selection problem in big-data scenarios. We develop a sequential addition-elimination algorithm for subdata selection. The algorithm is inspired by the fact that an orthogonal array of two levels is D-, A-, and G-optimal for linear regression. We define a discrepancy to measure how well a subdata set approximates an orthogonal array. Based on this discrepancy, we develop an algorithm which sequentially selects data points as well as eliminating data points from the full data to reduce the number of candidate points and speed up the selecting process. Compared with available methods, the proposed algorithm works much better in minimizing the sum of variances of coefficient estimations.

Bibliography

Ba, S., Myers, W. R., and Brenneman, W. A. (2015), "Optimal sliced Latin hypercube designs," *Technometrics*, 57, 479–487.

Bailey, R. (1982), "The decomposition of treatment degrees of freedom in quantitative factorial experiments," *Journal of the Royal Statistical Society. Series B (Methodological)*, 44, 63–70.

Box, G. E. and Wilson, K. B. (1951), "On the experimental attainment of optimum conditions," *Journal of the Royal Statistical Society: Series B (Methodological)*, 13, 1–38.

Butler, N. A. (2001), "Optimal and orthogonal Latin hypercube designs for computer experiments," *Biometrika*, 88, 847–857.

Chen, R.-B., Hsieh, D.-N., Hung, Y., and Wang, W. (2013), "Optimizing Latin hypercube designs by particle swarm," *Statistics and computing*, 23, 663–676.

Cheng, S.-W. and Wu, C. F. J. (2001), "Factor screening and response surface exploration," *Statistica Sinica*, 11, 553–580.

Cheng, S.-W. and Ye, K. Q. (2004), "Geometric isomorphism and minimum aberration for factorial designs with quantitative factors," *Annals of Statistics*, 32, 2168–2185.

Cioppa, T. M. and Lucas, T. W. (2007), "Efficient nearly orthogonal and space-filling Latin hypercubes," *Technometrics*, 49, 45–55.

Deng, L. Y. and Tang, B. (2002), "Design selection and classification for Hadamard matrices using generalized minimum aberration criteria," *Technometrics*, 44, 173–184.

Edmondson, R. (1993), "Systematic row-and-column designs balanced for low order polynomial interactions between rows and columns," *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, 707–723.

Fang, K.-T., Li, R., and Sudjianto, A. (2005), *Design and modeling for computer experiments*, Chapman and Hall/CRC.

Fang, K.-T. and Wang, Y. (1993), *Number-theoretic methods in statistics*, vol. 51, CRC Press.

Fang, K.-T., Zhang, A., and Li, R. (2007), “An effective algorithm for generation of factorial designs with generalized minimum aberration,” *Journal of Complexity*, 23, 740.

Georgiou, S. and Stylianou, S. (2011), “Block-circulant matrices for constructing optimal Latin hypercube designs,” *Journal of Statistical Planning and Inference*, 141, 1933–1943.

Georgiou, S. D. and Efthimiou, I. (2014), “Some classes of orthogonal Latin hypercube designs,” *Statistica Sinica*, 24, 101–120.

He, Y., Cheng, C.-S., and Tang, B. (2018), “Strong orthogonal arrays of strength two plus,” *The Annals of Statistics*, 46, 457–468.

He, Y. and Tang, B. (2012), “Strong orthogonal arrays and associated Latin hypercubes for computer experiments,” *Biometrika*, 100, 254–260.

— (2014), “A characterization of strong orthogonal arrays of strength three,” *The Annals of Statistics*, 42, 1347–1360.

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), “Minimax and maximin distance designs,” *Journal of Statistical Planning and Inference*, 26, 131–148.

Joseph, V. R. and Hung, Y. (2008), “Orthogonal-maximin Latin hypercube designs,” *Statistica Sinica*, 171–186.

Kiefer, J. and Wolfowitz, J. (1959), “Optimum designs in regression problems,” *The Annals of Mathematical Statistics*, 30, 271–294.

— (1960), “The equivalence of two extremum problems,” *Canadian Journal of Mathematics*, 12, 363–366.

Lin, C. D., Bingham, D., Sitter, R. R., and Tang, B. (2010), “A new and flexible method for constructing designs for computer experiments,” *Annals of Statistics*, 38, 1460–1477.

Lin, C. D., Mukerjee, R., and Tang, B. (2009), "Construction of orthogonal and nearly orthogonal Latin hypercubes," *Biometrika*, 96, 243–247.

Lin, C. D. and Tang, B. (2015), "Latin hypercubes and space-filling designs," *Handbook of Design and Analysis of Experiments*, 593–625.

Moon, H., Dean, A., and Santner, T. (2011), "Algorithms for generating maximin Latin hypercube and orthogonal designs," *Journal of Statistical Theory and Practice*, 5, 81–98.

Morris, M. D. (1991), "Factorial sampling plans for preliminary computational experiments," *Technometrics*, 33, 161–174.

Morris, M. D. and Mitchell, T. J. (1995), "Exploratory designs for computational experiments," *Journal of Statistical Planning and Inference*, 43, 381–402.

Morris, M. D. and Moore, L. M. (2015), "Design of computer experiments: Introduction and background," in *Handbook of Design and Analysis of Experiments*, Chapman and Hall/CRC, pp. 597–612.

Mukerjee, R. and Wu, C. F. J. (2006), *A modern theory of factorial design*, Springer.

Pang, F., Liu, M.-Q., and Lin, D. K. J. (2009), "A construction method for orthogonal Latin hypercube designs with prime power levels," *Statistica Sinica*, 19, 1721–1728.

Phoa, F. K. H. and Xu, H. (2009), "Quarter-fraction factorial designs constructed via quaternary codes," *Annals of Statistics*, 37, 2561–2581.

Plackett, R. L. and Burman, J. P. (1946), "The design of optimum multifactorial experiments," *Biometrika*, 33, 305–325.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and analysis of computer experiments," *Statistical science*, 409–423.

Santner, T. J., Williams, B. J., and Notz, W. (2003), *The design and analysis of computer experiments*, Springer.

Steinberg, D. M. and Lin, D. K. J. (2006), "A construction method for orthogonal Latin hypercube designs," *Biometrika*, 93, 279–288.

Sun, F., Liu, M.-Q., and Lin, D. K. J. (2009), "Construction of orthogonal Latin hypercube designs," *Biometrika*, 96, 971–974.

— (2010), "Construction of orthogonal Latin hypercube designs with flexible run sizes," *Journal of Statistical Planning and Inference*, 140, 3236–3242.

Sun, F. and Tang, B. (2017), "A general rotation method for orthogonal Latin hypercubes," *Biometrika*, 104, 465–472.

Tang, B. (1993), "Orthogonal array-based Latin hypercubes," *Journal of the American statistical association*, 88, 1392–1397.

Tang, Y. and Xu, H. (2014), "Permuting regular fractional factorial designs for screening quantitative factors," *Biometrika*, 101, 333–350.

Wang, H., Yang, M., and Stufken, J. (2018a), "Information-based optimal subdata selection for big data linear regression," *Journal of the American Statistical Association*, 1–13.

Wang, L., Sun, F., Lin, D. K. J., and Liu, M.-Q. (2018b), "Construction of orthogonal symmetric Latin hypercube designs," *Statistica Sinica*, 28, 1503–1520.

Wang, Y., Yu, A. W., and Singh, A. (2017), "On computationally tractable selection of experiments in measurement-constrained regression models," *The Journal of Machine Learning Research*, 18, 5238–5278.

Williams, E. J. (1949), "Experimental designs balanced for the estimation of residual effects of treatments," *Australian Journal of Scientific Research*, 2, 149–168.

Wu, C. F. J. and Hamada, M. S. (2009), *Experiments: Planning, Analysis, and Optimization*, John Wiley & Sons, 2nd ed.

Xiao, Q. and Xu, H. (2017), "Construction of maximin distance Latin squares and related Latin hypercube designs," *Biometrika*, 104, 455–464.

— (2018), "Construction of maximin distance designs via level permutation and expansion," *Statist. Sinica*, 28, 1395–1414.

Xu, H. (1999), "Universally optimal designs for computer experiments," *Statistica Sinica*, 1083–1088.

— (2003), "Minimum moment aberration for nonregular designs and supersaturated designs," *Statistica Sinica*, 691–708.

Xu, H., Cheng, S.-W., and Wu, C. F. J. (2004), "Optimal projective three-level designs for factor screening and interaction detection," *Technometrics*, 46, 280–292.

Xu, H. and Deng, L. Y. (2005), "Moment aberration projection for nonregular fractional factorial designs," *Technometrics*, 47, 121–131.

Xu, H., Phoa, F. K. H., and Wong, W. K. (2009), "Recent developments in nonregular fractional factorial designs," *Statistics Surveys*, 3, 18–46.

Yang, J. and Liu, M.-Q. (2012), "Construction of orthogonal and nearly orthogonal Latin hypercube designs from orthogonal designs," *Statistica Sinica*, 22, 433–442.

Ye, K. Q. (1998), "Orthogonal column Latin hypercubes and their application in computer experiments," *Journal of the American Statistical Association*, 93, 1430–1439.

Ye, K. Q., Tsai, K.-J., and Li, W. (2007), "Optimal orthogonal three-level factorial designs for factor screening and response surface exploration," in *mODa 8-Advances in Model-Oriented Design and Analysis*, Springer, pp. 221–228.

Zhou, Y. and Xu, H. (2015), "Space-filling properties of good lattice point sets," *Biometrika*, 102, 959–966.