DOCTORAL THESIS

Some New Developments of Experimental Designs and Their Applications

LIN, Yuxuan

Date of Award:
2022

HONG KONG BAPTIST UNIVERSITY

Doctor of Philosophy

THESIS ACCEPTANCE

DATE: July 4, 2022

STUDENT'S NAME: LIN Yuxuan

THESIS TITLE: Some New Developments of Experimental Designs and Their Applications

This is to certify that the above student's thesis has been examined by the following panel members and has received full approval for acceptance in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Chairman:	Prof PAN Jianxin Chair Professor, Division of Science and Technology, BNU-HKBU UIC (Designated by Dean of Faculty of Science)
Internal Members:	Dr Ye Huajun Associate Professor, Division of Science and Technology, BNU-HKBU UIC (Designated by Head of Department of Mathematics)
	Dr Elsawah Ahmed Associate Professor, Division of Science and Technology, BNU-HKBU UIC
External Examiners:	Prof Yue Rongxian Professor College of Mathematics and Physics Shanghai Normal University
	Prof Ou Zujun Professor and Dean School of Mathematics and Statistics Jishou University
In-attendance:	Prof Fang Kaitai Director, Institute of Statistics and Computational Intelligence, BNU-HKBU UIC

Issued by Graduate School, HKBU

Some New Developments of Experimental Designs and Their Applications

LIN Yuxuan

A thesis submitted in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy

Principal Supervisor:
Prof. FANG Kai-Tai (BNU-HKBU UIC)
Dr. PENG Heng (Hong Kong Baptist University)

July 2022

DECLARATION

I hereby declare that this thesis represents my own work which has been done after registration for the degree of PhD at Hong Kong Baptist University, and has not been previously included in a thesis or dissertation submitted to this or any other institution for a degree, diploma or other qualifications.

I have read the University's current research ethics guidelines, and accept responsibility for the conduct of the procedures in accordance with the University's Research Ethics Committee (REC). I have attempted to identify all the risks related to this research that may arise in conducting this research, obtained the relevant ethical approval, and acknowledged my obligations and the rights of the participants.

Signature: 

Date: July 2022

Abstract

Experimental design is an important branch in statistics and has been widely applied to various fields of industry, system engineering, and others. There are diverse design and modeling methods for different purposes. Among them, the orthogonal design (OD), optimal regression design including the D-optimal design (DOD) and uniform design (UD) have been widely employed. However, there are various open problems involving these designs. My research during the doctor study period tries to solve several problems. Most of the new results have been published in several international journals (see the list of my publications). My dissertation is based on some of these publications.

The orthogonal design is the most popular method having a long history. It based on some additive ANOVA models incorporating main effects, interaction effects and random error. A good orthogonal design is able to obtain an efficient estimation for requested effects with a small number of experimental runs. The effects that are less important according to the hierarchical ordering principle can be confounded for saving the number of runs. There are various criteria including the minimum aberration and the uniformity measure for evaluating the orthogonal designs in the literature. To evaluate and avoid the confounding situation of projected saturated symmetric orthogonal designs, we (Lin and Fang (2019)) proposed a criterion "the main effect confounding pattern" (MECP). The new criterion MECP is consistent with other criteria including discrepancies, and the generalized word-length pattern. In the meanwhile, MECP can provide more information about statistical performances in the classification for projection designs than the other criteria, providing an approach to finding the best main effect arrangement for the experimenter. I also participated in the projects "New non-isomorphic detection methods for orthogonal designs" and "Detecting non-isomorphic orthogonal designs" for developing more techniques of the identification and detection of non-isomorphic orthogonal designs. In conservative view, isomorphic orthogonal designs ought to have the same statistical performance. However, MECP indicates that the isomorphism does not imply the equivalence of orthogonal designs.

A good design for experiments should consider both effectiveness and robustness. Each design method is based on a given model. The OD is based on ANOVA models, whereas the D-optimal regression design is based on regression models, and the uniform design is on the overall mean model. These models have many unknown parameters to be estimated. A design is effective if it provides a more accurate (or even the best) estimate for the unknown model parameters. If the underlying model is not completely known, the robustness requests the design to perform at the same level when the model changes. If the underlying regression model is known, the D-optimal design (DOD) is the most effective on parameter estimation, but DOD is not robust against the model change. The uniform design is robust against the model change, but they are less efficient if the underlying model is completely known and more accurate than the overall mean model. The orthogonal design has good performance in both of efficiency and robustness, but it has a large space for improvements. To incorporate both robustness and effectiveness, we (Fang, Lin and Peng (2022)) proposed a new type of composite designs. According to the performances

on prediction mean square error in selected practical models, the recommendation of designs is addressed. Many case studies are investigated, and the application to the chemometrics is mentioned.

With the development of science and technology, computer experiments have gained more and more attention in past decades. Engineers and scientists have implemented computer simulations on physical systems due to the complex relationships between the inputs and outputs. Many space-filling designs including the latin hypercube design and the uniform design have been proposed and widely used in real case studies. Fang, Li, and Sudjianto (2005) gave a comprehensive introduction to the design and modeling of computer experiments. Due to the computational complexity of constructing UDs, the current widely used the uniform designs (UD) are constructed over a discrete (lattice points) space. If the factors of interest are continuous and involved in a computer experiment, we wish the levels of factors are allowed to be continuous in the design since the corresponding experimental cost is not affected. In the literature, most of UDs are constructed by a stochastic heuristic, threshold accepting algorithm. However, this algorithm is not suitable on the continuous domain. Note that constructing UD is minimizing a given discrepancy and is an optimization problem. We (Lai, Fang, Peng, Lin (2021)) proposed an approach to searching UDs on a continuous domain by coordinate descent methods. Several case studies show that the UDs on the continuous domain can improve the experimental achievement compared to the UDs on the discrete domain.

Acknowledgements

I would like to express my deep gratitude to my principal supervisor Prof. Fang Kai-Tai and my co-supervisor Dr. Peng Heng for all of their help and encouragement. It has been a great privilege to work with them. I have been learning from Prof. Fang since my undergraduate study. It was nine years since the first time I met Prof. Fang for inquiring about transferring my undergraduate program into statistics. It was him who firmed my determination to dedicate myself into statistics. My undergraduate final year project was instructed by Prof. Fang, and it was my first time to learn about experimental design. This research experience gave me great courage and confidence to pursue a research degree. With Prof. Fang and Dr. Peng's support and help, I successfully applied for Mphil and transferred to PhD after one year of Mphil study.

I am indebted to Prof. Fang's sophisticated instructions and detailed helpful suggestions that are contributive to both my PhD study and life. It is such a fortune that I have the opportunity to know and learn from Prof. Fang. I have encountered many difficulties in my study and life. Without Prof. Fang's encouragement and support, my study would not reach the current fruition. He is not only the supervisor of my study, but also the instructor of my whole life. The nine years I spent with Prof. Fang in UIC and HKBU are the most precious experience in my life. The benefits I obtained from this memorable experience of research are profound and will influence me in the long term.

A similar gratitude belongs to the teachers of our program. As program director, Dr. Ye Huajun and Dr. He Ping have provided a lot of support and help for my study affairs and life. Every time I have difficulty in life, they both care and help me a lot. I also have enjoyable cooperation with Dr. Peng Xiaoling and Dr. Ahmed Elsawah. I am gratitude to all the teachers of my program who are also my mentors and friends.

I also would like to thank my father, family, and friends. My father and family do not know English and experimental design, but without their constant support, I would never have had a chance to start and finish my PhD study. My best friend Li Haoci has encouraged me a lot every time I felt disappointed in my study and life.

My friends Lai Jianfa, Ke Xiao, Zhang Jiahua, Tang Yihan have paid many efforts and contribution to our publications. The cooperation with them is pleasant and memorable.

The PhD study period is challenging and undoubtedly exciting that I will always remember in my heart. Thank you to every one I have met along the way.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

1. ASC: Average squared correlation
2. CD: Centered L_2-discrepancy
3. CGD: Coordinate gradient descent
4. DE: Distance enumerator
5. GWLP: Generalized word-length pattern
6. HDP: Hamming distance pattern
7. LHS: Latin hypercube samping
8. MD: Mixture L_2-discrepancy
9. MECP: Main effect confounding pattern
10. MLHS: Mid-point Latin hypercube sampling
11. OD: Orthogonal design
12. PGWLP: Projection generalized word-length pattern
13. PHDP: Projection Hamming distance pattern
14. TA: Threshold accepting
15. UD: Uniform design

List of Publications

1. Lin, Y. X., Tang, Y. H., Zhang, J. H., & Fang K. T. (2022). Detecting non-isomorphic orthogonal designs, *Journal of Statistical Planning and Inference*, **221**(2022), 299-312, DOI: 10.1016/j.jspi.2022.05.003.

2. Fang, K. T., Lin, Y., & Peng, H. (2022). A new type of robust designs for chemometrics and computer experiments. *Chemometrics and Intelligent Laboratory Systems*, **221**, 104474, DOI: 10.1016/j.chemolab.2021.104474.

3. Lai, J., Fang, K. T., Peng, X., & Lin, Y. (2021). Construction of uniform designs over continuous domain in computer experiments. *Communications in Statistics-Simulation and Computation*, 1-17, DOI: 10.1080/03610918.2021.2011924.

4. Ke, X., Fang, K. T., Elsawah, A. M., & Lin, Y. (2020). New non-isomorphic detection methods for orthogonal designs. *Communications in Statistics-Simulation and Computation*, 1-16, DOI: 10.1080/03610918.2020.1844895.

5. Lin, Y., & Fang, K. T. (2019). The main effect confounding pattern for saturated orthogonal designs. *Metrika*, **82**(7), 843-861, DOI: 10.1007/s00184-019-00713-w.

Chapter 1

Introduction

1.1 Experimental Designs

Statistical experimental design is prevalent in various fields, including chemistry, clinical medicine and engineering. It refers to scientifically arranging a series of experimental runs and subsequently exploring the relationship between the response and the factors through the experimental outcomes. In some cases, the relationship is demonstrated by a predeterminate statistical model, such as a regression model or an ANOVA model. Under such circumstances where the underlying model is known, plenty of experimental design methods can be utilized, including the orthogonal designs and optimal regression designs. Nevertheless, for computer experiments, the true model is too complicated or have no analytical form. Furthermore, some cases related too many factors such that there is a large experimental region to be explored. Under such situation, those experimental design methods based on a specific model are no longer suitable. Instead, *space filling designs* have been prevalently utilized. The goal of space filling designs is to fill up the experimental region with design points under certain uniform fashion. The *Uniform Design* (UD) is one of space filling designs.

With the advantaged computing technology, computer simulations are prevalently applied for solving the problems that the relationships between responses and factors of interest are complicated nonlinear and even unknown. Computer experiments seek an approximate model, *metamodel*, as a surrogate for the true complex model. Fang

et al. [18] have elaborately addressed a study on design and modeling for computer experiments. The true model can be expressed as

$$y = f(x_1, \ldots, x_s) \equiv f(\boldsymbol{x}),\ \boldsymbol{x} \in \mathcal{T}, \tag{1.1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_s)$ is the input, y the output, $\mathcal{T}$ the experimental domain, and function f is known and has no analytic expression. The metamodel is given by

$$y = g(x_1, \ldots, x_s) \equiv g(\boldsymbol{x}), \tag{1.2}$$

such that the difference of $|f(\boldsymbol{x}) - g(\boldsymbol{x})|$ is small over the domain $\mathcal{T}$ in a certain sense. There are two essential keys for computer experiments, *design and modeling*. Fang et al. [18] provides many popularly used designs and modeling techniques. To obtain an adaptive metamodel, a set of points, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, in $\mathcal{T}$, is selected (*design*), and their corresponding outputs are calculated to form a data set $\{(\boldsymbol{x}_i, y_i), i = 1, \ \ldots, \ n\}$. Then some useful modeling techniques are applied to find a good model fitting the data. An adaptive metamodel should have less computation complexity, and be capable of representing non-linearity and well predicting at untried points, such that some tasks including optimization, prediction or sensitivity analysis are much easier to conduct. There are various *modeling* techniques, for instance, Kriging models, regression models, Bayesian methods, and/or neural networks.

To yield a good metamodel, the choice of design method is critical. The so-called space filling design has been recommended for computer experiments. The Latin hypercube sampling (LHS) proposed by McKay et al. [47] is one of the most popular space filling design. With design points uniformly scattered on the experimental domain, the uniform design proposed by Fang [16], Wang and Fang[54] is another widely known space-filling design.

The Latin hypercube sampling and uniform design are based on the estimation of overall mean of Y:

$$E(Y) = \int_{\mathcal{T}} h(\boldsymbol{x}) d\boldsymbol{x} = E(h(\boldsymbol{x})), \tag{1.3}$$

where $\boldsymbol{x} = (x_1, \ldots, x_s)$ follows the uniform distribution on $\mathcal{T}$. Then $E(Y)$ can be estimated by the mean $\bar{h} = \frac{1}{n} \sum_{i=1}^{n} h(\boldsymbol{x}_i)$. LHS provides a more efficient estimate

of $E(Y)$ than the estimate by the simple random sampling. The Koksma-Hlawka inequality (see [32]) gives the upper error bounds of the estimate as

$$|E(h(\boldsymbol{x})) - \bar{h}| \leq D(\mathcal{P})V(h), \tag{1.4}$$

where $V(h)$ is a measure of the variation of h, and $D(\mathcal{P})$ is the star-discrepancy of $\mathcal{P}$, a measure of the uniformity of $\mathcal{P}$. The definition of $V(h)$ is in the sense of Hardy and Krause ([32]). Note that $V(h)$ is independent of the design points. Thus, given a bounded $V(h)$, the Koksma-Hlawka inequality (1.4) indicates that the more uniform a set $\mathcal{P}$ of points is over the experimental region $\mathcal{T}$, the more accurate $\bar{h}$ is as an estimator of $E(h(\boldsymbol{x}))$. Both LHS and UD are based on this overall mean model.

This thesis focuses on orthogonal designs, orthogonal D-optimal designs and uniform designs, and hence a detailed introduction to the three designs are as following.

Optimal Regression Design

The optimal design, proposed by Kiefer [37], searches design points in order to optimize the estimates of the coefficients under the known underlying regression model. For a given linear regression model, it can be expressed in matrix form, $\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where each ϵ is identically independently distributed in $N(0, \sigma^2)$. The corresponding information matrix is defined by $\boldsymbol{M} = \boldsymbol{G}'\boldsymbol{G}/n$, where n is the number of experimental runs. There are various criteria for optimality such as D-, A-, E-optimalities. D-optimal designs maximize the determinant of information matrix $\boldsymbol{M}$, equivalently minimizing the confidence region of estimation, leading to estimates with high accuracy if the model assumption is correct. Obtaining a D-optimal design is a complicated optimization problem and needs some software.

On the other way, Atkinson and Donev [2] indicated the supporting point sets of D-optimal designs for the one-factor polynomial regression model. As the polynomials with lower order, the supporting point sets can approximate many non-linear regression models well.

Lemma 1. *For the one-factor polynomial regression model*

$$Y = \beta_0 + \sum_{i=1}^{q-1} \beta_i X^i + \epsilon, -1 \leq X \leq 1, \tag{1.5}$$

where X is the factor, Y is the response and ϵ is a random error with zero mean and constant variance, its D-optimal design has q support points that can be found from the Legendre polynomial defined by $L_j(u) = \frac{1}{j!2^j}\frac{d^j}{du^j}(u^2-1)^j$ where j is a positive integer.

Table 1.1 lists the support points of D-optimal design for $q \leq 5$.

Table 1.1: Support points of D-optimal design under model (1.5)

q	Support points
2	$-1 \quad 1$
3	$-1 \quad 0 \quad 1$
4	$-1 \quad -1/\sqrt{5} \quad 1/\sqrt{5} \quad 1$
5	$-1 \quad -\sqrt{3/7} \quad 0 \quad \sqrt{3/7} \quad 1$

Atkinson et al.[1] first introduced D-optimal designs for estimating rate constants in chemical kinetic model of a reversible reaction while in traditional way, chemists adopted the deterministic methods according to the characteristic of a reaction. However, the performance of D-optimal design is dependent on the initial chosen parameters of experiment. Moreover, optimal regression designs are not robust under the model changes.

Orthogonal and Orthogonal D-optimal Designs

Factorial designs and profile analysis are useful in chemometrics for estimating main effects of factors and their interactions. The orthogonal design is a sort of fractional factorial design.

Definition 1. *An $n \times s$ matrix, denoted as $L_n(q^s)$, with levels 1, 2, ..., q at each column, is called a symmetric orthogonal design table if it meets the following conditions:*

1. Each level appears equally often in each column.

2. Each level combination of any 2 columns appears equally often.

Let $\mathcal{L}_n(q^s)$ denote the set of the orthogonal designs, $L_n(q^s)$. For each set $\mathcal{L}_n(q^s)$, the orthogonal designs can be classified into non-isomorphic groups. Isomorphism plays a significant role for classification of orthogonal experimental designs. Two

$L_n(q^s)$ designs are called combinatorially isomorphic if one can be obtained from the other by the joint operations of relabeling factors, reordering the runs and/or switching the levels of the factors. Otherwise, they are combinatorially non-isomorphic, referred to Cheng and Ye [8]. There are $n!s!(q!)^s$ combinatorially isomorphic designs for each $L_n(q^s)$ design. If two factorial designs are isomorphic, then their performances in ANOVA model are at the same level. Hence, in traditional view, the experimenter only needs to compare all non-isomorphic designs and choose one. Chapter 2 introduces some new ways to detect non-isomorphism. However, Chapter 3 provides a novel idea that isomorphic designs perform differently in modeling.

Liang et al.[41] pointed out that the meaning of "orthogonality" implies the design $L_n(q^s)$ results in a decomposition of R^n with s orthogonal subspace to each other, such that the total sum of squares of the responses can be decomposed as s sum of squares, including factor sum of squares and/or interaction sum of squares. The convenience of obtaining tabulated orthogonal designs is one of the reasons for its popularity. Cheng [7] proved that orthogonal designs are also optimal designs under ANOVA model.

If the underlying model is polynomial in each factor, Chan et al. [6] considered the regression model

$$Y = \beta_0 + \sum_{t=1}^{s}[\beta_{t1}X_t + \beta_{t2}X_t^2 + \ldots + \beta_{t,q-1}X_t^{q-1}] + \epsilon, -1 \leq X_t \leq 1, t = 1, \ldots, s, \quad (1.6)$$

and found its D-optimal design, where the design uses the corresponding supporting point set in Lemma 1. Model (1.6) can be used as a metamodel for computer experiments with s quantitative factors. The number of runs for this experiment should be $n = kq^s$ for some positive integer k. For saving the number of runs, we can use an orthogonal design $L_n(q^s)$ where the q levels determined by the support points in Table 1.1. This kind of design is called as *orthogonal D-optimal design*, denoted as $L_n^D(q^s)$. For the convenience of notation, we denote the orthogonal designs taking levels evenly as $L_n^E(q^s)$. For instance, $L_{16}^E(4^5)$ takes levels $\{-1, -\frac{1}{3}, \frac{1}{3}, 1\}$ on $[-1, 1]^5$ while $L_{16}^D(4^5)$ takes levels $\{-1, -\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 1\}$. For increasing the effectiveness of parameter estimation in the given model, Fang et al. [22] showed that the orthogonal D-optimal design (for example, $L_{16}^D(4^5)$) can increases the D-efficiency by the use of $L_{16}^E(4^5)$.

Uniform Design

The uniform design is a kind of space-filling design for computer experiments and is another kind of fractional factorial design for physical and computer experiments. Fang [16], Wang and Fang [54] proposed the uniform design that evenly spread the experimental points throughout the experimental region under a given uniformity measure. Let $U_n(q^s)$ be a uniform design with n runs, s factors and each factor having q levels. When the factors are quantitative and their domain is a rectangle $a_i \leq X_i \leq b_i, i = 1, \ldots, s$, without loss of generality, we can assume the experimental domain is a unit cube $C^s = [0, 1]^s$. To measure the uniformity of a design, Hickernell [30] proposed the centered L_2-discrepancy (CD) and the wrap-around L_2-discrepancy (WD). Moreover, Zhou et al.[62] then proposed the mixture L_2-discrepancy (MD). For a point set $\mathcal{P} = (x_{ik})_{n\times s}$ on a unit hypercube, $\mathcal{C} = [0, 1]^s$, the squared CD has a computational formula

$$CD^2(\mathcal{P}) = (\frac{13}{12})^s - \frac{2}{n}\sum_{i=1}^{n}\prod_{k=1}^{s}(1 + \frac{1}{2}|x_{ik} - 0.5| - \frac{1}{2}|x_{ik} - 0.5|^2)$$
$$+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{k=1}^{s}(1 + \frac{1}{2}|x_{ik} - 0.5| + \frac{1}{2}|x_{jk} - 0.5| - \frac{1}{2}|x_{ik} - x_{jk}|). \tag{1.7}$$

Similarly, the squared MD has a computational formula as

$$MD^2(\mathcal{P}) = (\frac{19}{12})^s - \frac{2}{n}\sum_{i=1}^{n}\prod_{k=1}^{s}(\frac{5}{3} - \frac{1}{4}|x_{ik} - 0.5| - \frac{1}{4}|x_{ik} - 0.5|^2) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\prod_{k=1}^{s}$$
$$(\frac{15}{8} - \frac{1}{4}|x_{ik} - 0.5| - \frac{1}{4}|x_{jk} - 0.5| - \frac{3}{4}|x_{ik} - x_{jk}| + \frac{1}{2}|x_{ik} - x_{jk}|^2). \tag{1.8}$$

The squared WD has a computational formula as

$$\mathrm{WD}^2(\mathcal{P}) = -\left(\frac{4}{3}\right)^s + \frac{1}{n^2}\sum_{i,k=1}^{n}\prod_{j=1}^{s}\left(\frac{3}{2} - |x_{ij} - x_{kj}| + |x_{ij} - x_{kj}|^2\right). \tag{1.9}$$

A uniform design minimizes a given discrepancy of the experimental points over the experimental domain, defined as follows.

Definition 2. *For an experiment consisting of n runs relating to s factors on an experimental region $\mathcal{T}$, denote the experimental domain by $\mathcal{U}$ that is the collection of*

sets $\mathcal{P} = (x_{ik})_{n\times s} \subset \mathcal{T}$. *Then a design,* $\mathcal{P}^* = (x^*_{ik})_{n\times s} \subset \mathcal{T}$, *is called a uniform design under a given measure discrepancy* $\mathcal{D}$ *if*

$$\mathcal{D}(\mathcal{P}^*) = min_{\mathcal{P}\in\mathcal{U}}\mathcal{D}(\mathcal{P}), \tag{1.10}$$

where the discrepancy can choose CD, MD or others.

Constructing a uniform design is an NP hard problem that there is no exact algorithm for global optima within polynomial time. Besides, in practical experiments, the feasible values of factors are limited so that the corresponding levels in designs are required to be discrete. Therefore, for constructing uniform designs, the traditional methods often focused on a lattice domain $\mathcal{U}$ instead of the full domain $\mathcal{T}$. The U-type domain is prevalently accepted in the existing literatures.

Definition 3. *Suppose a design* $\boldsymbol{U}$ *is an* $n \times s$ *matrix. Then it is said to be a* q*-level U-type design denoted as* $\boldsymbol{U}(n, q^s)$ *if in each column of* $\boldsymbol{U}$*, the entries,* $\frac{l-0.5}{q}$ *for* $l = 1, \ldots, q$*, appear equally often.*

Most authors restrict a U-type design in an experimental region, $[0, 1]^s$, defined in Definition 3. The uniform design obtained from a U-type domain is called a U-type uniform design for distinguishing purpose in this paper, denoted as $\boldsymbol{U}_n(q^s)$. However, Definition 3 is not suitable for continuous domain addressed in Chapter 5. Hence, we adjust the definition of U-type requirement for a continuous design as follows.

Definition 4. *Suppose an* $n \times s$ *design* $\boldsymbol{U}$ *is on a continuous domain,* $\mathcal{C}^s = [0, 1]^s$. *In each dimension, separate the interval* $[0, 1]$ *into* n *subintervals, denoted as*

$$[0, u_1],\ (u_1, u_2], \ldots, (u_{n-1}, 1],$$

where $u_i = \frac{i}{n},\ i = 1, 2, \ldots, n-1$. *Then the continuous design* $\boldsymbol{U}$ *satisfies U-type requirement, if in each column of* $\boldsymbol{U}$*, each entry appears equally in each of the* n *subintervals respectively.*

Fang et al. [18] and Fang et al. [19] gave a comprehensive introduction to the uniform design. Liang et al. [41] and Xu et al. [61] gave a systematic introduction to the applications of the uniform design in chemistry, chemical engineering and chemometrics. Liang et al. [41] pointed out that the uniform designs offer the following advantages:

1. more choices of number of levels and/or factors,
2. tabulated designs,
3. the application of both factorial and computer experiments,
4. less requirement of prior information about the underlying model,
5. robust against model changes,
6. nearly orthogonal.

Similarly as orthogonal tables, uniform designs are tabulated in many literature, such as the following website, `https://dst.uic.edu.cn/en/isci/unifor/m-design/uniform-design-tables`. Note that all UD tables in the webpage have levels labeled as $\{1, 2, \ldots, q\}$ for q levels that can be shifted into $\{\frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{2q-1}{2q}\}$ over the domain (0,1).

Motivated by Atkinson et al.'s [1] work, Xu et al. [60] put uniform designs, orthogonal designs and D-optimal designs into the same framework, and presented the effects of different design methods on parameter estimation in the kinetics of a reversible chemical reaction, where average MSEs of predicting 100 samples and variance of the estimated parameters are adopted as criteria. Xu et al. [60] pointed out their conclusion:

1. If the experimenter has a proper initial guess of parameters that are close to the true ones, and the random error is not too large, OD, UD and D-optimal designs are all suitable but D-optimal designs are strongly recommended.
2. If the experimenter has no prior information about the parameters, the parameter estimation of D-optimal designs may fail if there is no appropriate initial parameters.
3. OD is robust against to the position of initial parameters, but the estimation may also fail when the random error increases to a certain level.
4. The performance of UD is stable in all cases. Although under some circumstances, UD can not give the best performance, it often performs better than OD in estimating parameters.

Each design method has its advantages and weaknesses. The orthogonal design is convenient for factorial design with small number of levels; the uniform design is robust against the model change, but less effective on model parameter estimation when the underlying model is known; the D-optimal design guarantees the optimal parameter estimation when the underlying regression model is known, but lack of robustness against the model change. This motivates us to construct a composition of them elaborated in Chapter 4 such that the resulting designs are both effective and robust.

1.2 Criteria of Experimental Designs

There exist various criteria including minimum aberration and the uniformity measure that are used for detecting or comparing non-isomorphic designs. Box et al. [4] proposed resolution, and Fries and Hunter [29] proposed minimum aberration (word-length pattern) for regular fractional factorial designs (FFD). For extending the concept to non-regular FFD, Ma and Fang [45] and Xu and Wu [59] proposed generalized word-length pattern (GWLP). Moreover, the Hamming distance pattern (HDP) has been proposed by Ma et al. [46] and by Clark and Dean [10]. From the overall mean model, various discrepancies as uniformity measure, are used to evaluate the geometrical structures of design points and can be used detecting non-isomorphic ODs. Ma et al. [46] also proposed the distance enumerator (DE) to detect non-isomorphism and provided the relationship among DE, discrepancies, HDP and GWLP. The DE shows its advantages in non-isomorphism detection. All the above criteria have analytic expressions related to HDP when the number of levels is two or three. Booth and Cox [3] proposed an efficient criterion for supersaturated 2-level designs, that is the variance of the sum of products by each pair of columns, denoted as Es^2. Evangelaras et al. [14] extended the concept to three-level designs, that is the so-called Average Squared Correlation (ASC). A lower ASC value among main effects indicates less confounding in modeling.

The following are some basic definitions for the convenience of elaboration.

Definition 5. *Hamming distance pattern and distance enumerator*

Let $d_H(i,j)$ be the Hamming distance between the i^{th} and j^{th} runs of a design $\boldsymbol{D}$, indicating the number of positions where the two runs are different. Let $E_k(\boldsymbol{D}) = \frac{1}{n}\#\{(i,j)|(i,j) \in \boldsymbol{D}, d_H(i,j) = k\}$. The vector $\boldsymbol{H}(\boldsymbol{D})$ is called the Hamming distance pattern of $\boldsymbol{D}$, where

$$\boldsymbol{H}(\boldsymbol{D}) = (E_0(\boldsymbol{D}), E_1(\boldsymbol{D}), E_2(\boldsymbol{D}) \cdots E_s(\boldsymbol{D})).$$

The distance enumerator of $\boldsymbol{D}$ is defined by Roman [48] as

$$B_a(\boldsymbol{D}) = \sum_{k=1}^{s} E_k(\boldsymbol{D}) a^k, \tag{1.11}$$

where a is a pre-determined positive constant, often recommended to set as an irrational number, such as $\sqrt{2}$.

It is interesting to note that many useful criteria in theory of experimental design have close relationships with HDP. The following are some of these criteria. The word-length patten has been used for assessing regular fractional factorial designs. Ma and Fang [45] and Xu and Wu [59] extended this concept to generalized word-length patten, $\boldsymbol{W}(\boldsymbol{D})$, for assessing regular and non-regular designs. Ma and Fang [45] used the MacWilliams identities to express the $\boldsymbol{W}(\boldsymbol{D})$ by the $\boldsymbol{H}(\boldsymbol{D})$.

Definition 6. *General word-length pattern*
For any design $\boldsymbol{D}$, the GWLP is defined as

$$\boldsymbol{W}(\boldsymbol{D}) = (A_1(\boldsymbol{D}), A_2(\boldsymbol{D}), \cdots, A_s(\boldsymbol{D})),$$

where

$$A_i(\boldsymbol{D}) = \frac{1}{n(q-1)} \sum_{j=0}^{s} P_i(j;s) E_j(D), i = 1, \cdots, s, \tag{1.12}$$

and $P_i(j;s)$ are the Krawtchouk Polynomials. For two designs $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, $\boldsymbol{D}_1$ is said to have less aberration than $\boldsymbol{D}_2$ if there exists a positive integer $\theta \leq s$, such that $A_\theta(\boldsymbol{D}_1) < A_\theta(\boldsymbol{D}_2)$ and $A_i(\boldsymbol{D}_1) = A_i(\boldsymbol{D}_2)$ for $i = 1, \cdots, \theta - 1$.

There are several measures of uniformity that also have a close relationship with HDP and GWLP. Fang and Mukerjee [23] found an analytic connection between uniformity and aberration in two-level regular factorial designs. Ma and Fang [45]

extended this relationship into any two-level factorial designs with some connections between WD and GWLP. Zhou et al. [62] referred to the relations between MD and GWLP. Fang et al. [19] gave a comprehensive discussion in this direction. Ma et al. [46] proved the analytic formula between discrepancies and HDP as

$$CD^2(\boldsymbol{D}) = \left(\frac{13}{12}\right)^s - 2\left(\frac{35}{32}\right)^s + \frac{1}{n}\left(\frac{5}{4}\right)^s \sum_{r=0}^{s} E_r(\boldsymbol{D}) \left(\frac{4}{5}\right)^r . \tag{1.13}$$

This also leads to the relationship between discrepancies and the distance enumerator when replacing the last term of (1.13) with $B_{4/5}(\boldsymbol{D})$ in (5) as follows

$$B_{4/5}(\boldsymbol{D}) = \sum_{r=1}^{s} E_r(\boldsymbol{D}) \left(\frac{4}{5}\right)^r .$$

HDP evaluates designs from geometric perspective as uniformity measures. However, it has statistical meaning as in Definition 6. Moreover, Elsawah et al. [13] indicated that HDP and GWLP have similar performance on non-isomorphism detection. Based on the above review, it is reasonable that we use HDP to define another criterion for evaluating designs. From the joint work with Ms. Tang Yihan, Mr. Zhang Jiahua and Prof. Fang Kai-Tai in Lin et al.[43], we propose a criterion, the entropy based on projection HDP, denoted by $S(\mathcal{D}_k)$.

The entropy based on projection HDP

The concept of entropy in thermodynamics was first used for quantifying the amount of information by Shannon [50] in 1948.

Definition 7. *Suppose a signal can take values $X_1, \ldots, X_n$ with the corresponding positive probabilities $p_1, \ldots, p_n$. The information entropy is defined by*

$$S(X) = -\sum_{i=1}^{n} p_i log\left(p_i\right). \tag{1.14}$$

Considering the signal as a discrete random variable, Shannon used the information entropy to measure the uncertainty. It inspires us that for any given discrete random variable, we can construct its information entropy. For any saturated or unsaturated, regular or irregular factorial design, HDP is suitable for evaluating designs as the generalized minimum aberration criterion. In practical applications, the experimenter may focus on a certain number of factors less than the number of columns

of a given design table. Hence, evaluating the projection property of factorial designs plays a significant role as well.

Given any design $\boldsymbol{D} \in \mathcal{L}_n(q^s)$, there are $\binom{s}{k}$ k-dimensional projection designs, $1 < k < s$. Denote the collection of all k-dimensional projection designs of $\boldsymbol{D}$ as $\mathcal{D}^k$. After calculating their HDPs, we suppose that the collection $\mathcal{D}^k$ have m unique projection Hamming distance patterns (PHDP), each denoted as $\boldsymbol{H}^j = (H_i^j) = (H_0^j,\ H_1^j,\ ...,\ H_k^j)$, $j = 1, \ldots, m$, and the corresponding probabilities as $\boldsymbol{p}(\boldsymbol{H}^j) = (\ p(H_i^j)\ )$, where

$$p(H_i^j) = \frac{H_i^j + 1}{\sum_{i=0}^{k}(H_i^j + 1)}.$$

Here we use $H_i^j + 1$ to avoid 0 element in $\boldsymbol{H}^j$. Then we have the entropy based on projection HDP of $\mathcal{D}^k$ defined as follows.

Definition 8. *Entropy based on projection Hamming distance pattern, $S(\mathcal{D}^k)$*
Denote the entropy of each $\boldsymbol{H}^j$ in $\mathcal{D}^k$ as

$$S_j(\mathcal{D}^k) = -\sum_{i=0}^{k} p(H_i^j) log\left(p(H_i^j)\right). \tag{1.15}$$

Define the probabilities of each $S_j(\mathcal{D}^k)$ as $p(S_j(\mathcal{D}^k)) = \frac{S_j(\mathcal{D}^k)}{\sum_{j=1}^{m} S_j(\mathcal{D}^k)}$. Hence, the entropy based on projection Hamming distance pattern of each $\mathcal{D}^k$, $1 < k < s$, is defined by

$$S(\mathcal{D}^k) = -\sum_{j=1}^{m} p(S_j(\mathcal{D}^k)) log\left(p(S_j(\mathcal{D}^k))\right). \tag{1.16}$$

Two isomorphic designs have the same projection HDP distribution and subsequently have the same $S(\mathcal{D}^k)$. Hence, the criterion, $S(\mathcal{D}^k)$ can be used for detecting non-isomorphism for two orthogonal designs, $\boldsymbol{D}_1, \boldsymbol{D}_2 \in \mathcal{L}_n(q^s)$. If two designs have different $S(\mathcal{D}^k)$ values for any k, where $1 < k < s$, then we conclude that they are combinatorially non-isomorphic, otherwise we need to do more investigation. For instance, Evangelaras et al. [15] proved that there exist only three classes in $\mathcal{L}_{18}(3^7)$ listed in Table 1.4, which is downloaded from `http://pietereendebak.nl/oapage/index.html`. Denote these three designs by $L_{18}^{\zeta}(3^7), \zeta = 1, 2, 3$. For each class, we explore the k-factor projection designs for $k = 3, 4, 5, 6$. Their $S(\mathcal{D}^k)$ values for each k-factor projection are listed in Table 1.2.

To interpret the values of $S(\mathcal{D}^k)$, we list the values of each PHDP and the corresponding $S(\mathcal{D}^k)$ values with frequencies of $L_{18}^2(3^7)$ and $L_{18}^3(3^7)$ in Table 1.3. The

Table 1.2: The $S(\mathcal{D}^k)$ of k-factor projection for $L_{18}^{\zeta}(3^7)$ designs

k	3	4	5	6	Sum
$L_{18}^1(3^7)$	5.128136	5.129087	4.392251	2.806419	17.455893
$L_{18}^2(3^7)$	5.128132	5.128994	4.390699	2.778008	17.425833
$L_{18}^3(3^7)$	5.128141	5.129082	4.392291	2.807287	17.456801

comparison procedure of HDP is similar to GWLP defined in Definition 6. It is recommended that the leading entries of HDP who have higher priority are zeros. The smaller HDP implies that the design have less aberration. In view of projection HDP, for five- or six-factor projections in Table 1.3, the best projection HDP of $L_{18}^2(3^7)$ is better than the best one of $L_{18}^3(3^7)$, and hence $L_{18}^2(3^7)$ is recommended. As for three- or four-factor projection in Table 1.3, both classes are able to derive the same best projection HDPs, $[0, 27, 81, 45]$ and $[0, 0, 54, 72, 27]$, but $L_{18}^2(3^7)$ has greater frequencies of the two best projection HDPs. That implies in this consideration, $L_{18}^2(3^7)$ is also recommended. To interpret $S(\mathcal{D}^k)$, that is defined through entropy, it is known that the less uncertainty corresponds to smaller entropy. In our case, the greater frequency of the best projection HDP corresponds the smaller entropy, which is preferred when comparing two designs. The $S(\mathcal{D}^k)$ values of $L_{18}^2(3^7)$ for all k-factor projection are smaller than the ones of $L_{18}^3(3^7)$, which is consistent with the conclusion of projection HDP.

The entropy based on projection HDP provides a new perspective of evaluating experimental designs. Moreover, its detection process saves a great amount of time. However, in some cases, $S(\mathcal{D}^k)$ may fail to distinguish non-isomorphism due to the limit of HDP on detection task. Hence, other criteria can be adopted as an amendment in the detection task. In Lin et al.[43], we adopt projection ASC distribution.

Es^2 and Average Squared Correlation

Booth and Cox [3] proposed the Es^2 criterion to evaluate the efficiency of two-level supersaturated designs, that are fractional factorial designs with too small number of runs to estimate all the main effects. Let $\boldsymbol{X}$ denote an $n \times k$ design, where $n < k+1$,

Table 1.3: The k-factor PHDP and $S_j(\mathcal{D}^k)$ with Frequencies of $L_{18}^2(3^7)$ and $L_{18}^3(3^7)$

	$L_{18}^2(3^7)$				$L_{18}^3(3^7)$			
$k=3$	PHDP	$S_j(\mathcal{D}^3)$	Frequency	$S(\mathcal{D}^3)$	PHDP	$S_j(\mathcal{D}^3)$	Frequency	$S(\mathcal{D}^3)$
	[0,27,81,45]	1.446648	28	5.128132	[0,27,81,45]	1.446648	16	5.128141
	[3,18,90,42]	1.43675	6		[1,24,84,44]	1.458638	18	
	[9,0,108,36]	1.086313	1		[9,0,108,36]	1.086313	1	
$k=4$	PHDP	$S_j(\mathcal{D}^4)$	Frequency	$S(\mathcal{D}^4)$	PHDP	$S_j(\mathcal{D}^4)$	Frequency	$S(\mathcal{D}^4)$
	[0,0,54,72,27]	1.483659	15	5.128994	[0,0,54,72,27]	1.483659	1	5.129082
	[0,3,45,81,24]	1.535458	12		[0,2,48,78,25]	1.529048	18	
	[0,9,27,99,18]	1.451661	8		[0,3,45,81,24]	1.535458	12	
					[0,9,27,99,18]	1.451661	4	
$k=5$	PHDP	$S_j(\mathcal{D}^5)$	Frequency	$S(\mathcal{D}^5)$	PHDP	$S_j(\mathcal{D}^5)$	Frequency	$S(\mathcal{D}^5)$
	[0,0,0,90,45,18]	1.33282	6	4.390699	[0,0,6,72,63,12]	1.510116	15	4.392291
	[0,0,9,63,72,9]	1.519728	8		[0,0,12,54,81,6]	1.487315	6	
	[0,0,12,54,81,6]	1.487315	6					
	[0,0,18,36,99,0]	1.260772	1					
$k=6$	PHDP	$S_j(\mathcal{D}^6)$	Frequency	$S(\mathcal{D}^6)$	PHDP	$S_j(\mathcal{D}^6)$	Frequency	$S(\mathcal{D}^6)$
	[0,0,0,0,135,0,18]	0.522559	1	2.778008	[0,0,0,12,99,36,6]	1.368807	3	2.807287
	[0,0,0,18,81,54,0]	1.37928	6		[0,0,0,18,81,54,0]	1.37928	4	

and the levels are denoted by 1 and -1. Each column of $\boldsymbol{X}$ has the same frequency of 1 and -1 (U-type design), and all columns are different. Let s_{ij} denote the $(ij)^{th}$ element of $\boldsymbol{X'X}$. Clearly, $s_{ij} = 0$ if the i^{th} and j^{th} columns of $\boldsymbol{X}$ are orthogonal. Booth and Cox [3] indicated that a design is called Es^2-optimal if it has the minimum average s_{ij}^2, that implies it has the minimum average squared correlation.

Evangelaras et al. [14] extended the concept of Es^2 to three-level designs, that is the average squared correlation (ASC). For each three-level factor of the model matrix $\boldsymbol{X}$, there are two orthogonal contrasts allocated to linear and quadratic main effects respectively. For each pair of two factors of the model matrix $\boldsymbol{X}$, there are four orthogonal contrasts allocated to two-factor interactions. According to Johnson and Wichern [35], if $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ are column vectors of the model $\boldsymbol{X}$, representing either the main effect contrasts of two factors or interaction contrasts, then the correlation between the contrasts is calculated as the angle between the two vectors, denoted by $\rho_{i;j}$. Remark that s_{ij}^2 is the square of the $(ij)^{th}$ element of $\boldsymbol{X'X}$ for two-level designs, whereas $\rho_{i;j}^2$ is the square of the angle between the i^{th} and j^{th} vectors. A design is Es^2-optimal if it has the minimum average $\rho_{i;j}^2$.

Table 1.4: Three Representative Non-isomorphic Designs of $L_{18}(3^7)$

No	$L_{18}^1(3^7)$							$L_{18}^2(3^7)$							$L_{18}^3(3^7)$						
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2	1	1	1	2	2	2	2	1	1	1	2	2	2	2
3	1	2	2	1	1	3	3	1	2	2	1	1	3	3	1	2	2	1	1	3	3
4	1	2	2	3	3	2	2	1	2	2	3	3	2	2	1	2	2	3	3	2	2
5	1	3	3	2	2	3	3	1	3	3	2	2	3	3	1	3	3	2	3	1	3
6	1	3	3	3	3	1	1	1	3	3	3	3	1	1	1	3	3	3	2	3	1
7	2	1	2	2	3	1	3	2	1	2	2	3	1	3	2	1	2	2	2	3	3
8	2	1	2	3	2	3	1	2	1	2	3	2	3	1	2	1	2	3	3	1	1
9	2	2	3	1	2	1	2	2	2	3	1	2	2	1	2	2	3	1	2	1	2
10	2	2	3	2	1	2	1	2	2	3	2	1	1	2	2	2	3	2	1	2	1
11	2	3	1	1	3	2	3	2	3	1	1	3	2	3	2	3	1	1	3	3	2
12	2	3	1	3	1	3	2	2	3	1	3	1	3	2	2	3	1	3	1	2	3
13	3	1	3	1	3	3	2	3	1	3	1	3	3	2	3	1	3	1	3	2	3
14	3	1	3	3	1	2	3	3	1	3	3	1	2	3	3	1	3	3	1	3	2
15	3	2	1	2	3	3	1	3	2	1	2	3	3	1	3	2	1	2	3	3	1
16	3	2	1	3	2	1	3	3	2	1	3	2	1	3	3	2	1	3	2	1	3
17	3	3	2	1	2	2	1	3	3	2	1	2	1	2	3	3	2	1	2	2	1
18	3	3	2	2	1	1	2	3	3	2	2	1	2	1	3	3	2	2	1	1	2

Evangelaras et al. [14] also proved that the average squared correlation between complete set of orthogonal contrasts for any two factorial effects is invariant to the particular sets of orthogonal contrasts selected. For any orthogonal design, each column represents a factor of the model X. Fang and Ma [21] suggested that a main effect can be decomposed into linear main effect and quadratic main effect by the orthogonal polynomial technique. Fang et al. [24] indicated the decomposition was derived by replacing levels $\{1\}$, $\{2\}$, $\{3\}$ with $\{-1, 1\}$, $\{0, -2\}$, $\{1, 1\}$ respectively. Therefore, for each factor, say A, we obtain two orthogonal contrasts of main effects representing linear trend and quadratic trend respectively, say A_l and A_q. Moreover, the orthogonal trend contrasts for each two-factor interaction are calculated by dot

products. For instance, factors A and B are decomposed into $A_l,\ A_q$ and $B_l,\ B_q$ respectively, and the four orthogonal trend contrasts for AB interaction are denoted by $A_lB_l,\ A_lB_q,\ A_qB_l,\ A_qB_q$, derived by dot products of the corresponding decomposed main effects, (refer to Lin and Fang [42]).

Evangelaras et al.[14] considered that the correlation between main effect and two-factor interaction is of different order from the correlation among the four orthogonal trend contrasts of two-factor interaction. For instance, let C, D denote another two factors with linear main effects, C_l and D_l. The correlation between C_l and A_lB_l is of order 3, whereas the correlation between C_lD_l and A_lB_l is of order 4. Based on Evangelaras et al.[14], let $Ave(\rho^2_{t;uv})$ denote the average squared correlation of order 3, that is the average of the squared correlations between each of the two orthogonal trend contrasts for the t^{th} main effect and each of the four orthogonal trend contrasts for the two-factor interaction between factors u and v, for all $t \neq u < v$. Similarly, the average squared correlation of order 4, denoted by $Ave(\rho^2_{tw;uv})$, is calculated between each of the four orthogonal trend contrasts for the $(tw)^{th}$ two-factor interaction (between factors t and w) and those for the $(uv)^{th}$ two-factor interaction, for all $t < w \neq u < v$ (Evangelaras et al. [14]).

For instance, Table 1.5 displayed the frequency of each ASC value of each class in $\mathcal{L}_{18}(3^7)$. Similar as the entropy of projection HDP, if two designs have the same ASC, we can further explore the distribution of ASC in projection designs. The detection algorithm of projection ASC distribution will be illustrated in next chapter.

Table 1.5: The ASC for $L^{\zeta}_{18}(3^7)$, $\zeta = 1, 2, 3$

	ASC of order 3					ASC of order 4							
	0	0.0625	0.0833	0.125	0.25	0	0.0313	0.0417	0.0625	0.0729	0.0938	0.125	0.25
$L^1_{18}(3^7)$	42	59	33	10	3	11	62	41	36	49	3	2	6
$L^2_{18}(3^7)$	42	75	0	27	3	23	71	0	87	1	20	2	6
$L^3_{18}(3^7)$	42	48	54	0	3	11	53	54	24	60	0	2	6

Chapter 2

New Non-isomorphic Detection Methods for Orthogonal Designs

2.1 A Review to NIU Algorithm

As introduced in the previous chapter, when two designs perform the same, their projection properties are explored, and hence the projection distributions are compared. Ma et al. [46] first introduced NIU algorithm based on projection CD for two-level designs and based on projection distance enumerator for higher levels. Ke et al. [36] adopted NIU algorithm based on projection HDP (Algorithm 1) and based on projection MD respectively. Elsawah et al. [13] adopted NIU algorithm based on projection HDP (Algorithm 1) and projection GWLP respectively. It was found that projection HDP and projection GWLP have the same classification results. Elsawah et al. [13] also indicated that the number of non-isomorphic subclasses detected by projection HDP (or GWLP) for any k-factor projection is equal to the number of non-isomorphic subclasses for its complement, i.e. k^c(=(s-k))-factor projection. Hence, we do not need to exploit PHDP for $k = [\frac{s}{2}] + 1 : s - 1$ in **Step 3** of Algorithm 1.

Algorithm 1. *NIU algorithm based on projection HDP distribution*

Step 1*: If the given designs, $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ in $\mathcal{L}_n(q^s)$, are saturated with the same size, then they have the same HDP. Therefore, directly go to step 3. Otherwise, go to step 2.*

Step 2*: Comparing $\boldsymbol{H}(\boldsymbol{D}_1)$ and $\boldsymbol{H}(\boldsymbol{D}_2)$, we conclude $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are not isomor-*

phic, if $\boldsymbol{H}(\boldsymbol{D}_1) \neq \boldsymbol{H}(\boldsymbol{D}_2)$. *Otherwise go to step 3.*

Step 3: **For** $k = 3 : [\frac{s}{2}]$, **DO** *find all projection HDP of k-factor projection designs of* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$.

Step 4: **IF** *the numbers of projection HDPs are different, then conclude that* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are non-isomorphic and* **Break**.

Step 5: **Else IF** *the values of projection HDPs are different, then conclude that* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are non-isomorphic and* **Break**.

Step 6: **Else IF** *the frequencies of projection HDPs are different, then conclude that* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are non-isomorphic and* **Break**.

Step 7: **End For**.

Step 8: *If the projection HDP distribution of* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are the same for all k, we have no conclusion about the detection and reserve the two designs for other detection methods.*

2.2 New NIU Algorithms

Similar to Algorithm 1, we propose NIU algorithm with the entropy based on projection HDP ($S(\mathcal{D}^k)$) as following Algorithm 2.

Algorithm 2. *NIU Algorithm with* $S(\mathcal{D}^k)$

Step 1: *If the given designs,* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$, *are saturated with the same size, then they have the same HDP. Therefore, directly go to step 3. Otherwise, go to step 2.*

Step 2: *Calculate* $\boldsymbol{H}(\boldsymbol{D}_1)$ *and* $\boldsymbol{H}(\boldsymbol{D}_2)$. *We conclude* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are not isomorphic, if* $\boldsymbol{H}(\boldsymbol{D}_1) \neq \boldsymbol{H}(\boldsymbol{D}_2)$. *Otherwise go to step 3.*

Step 3: **For** $k = 3 : [\frac{s}{2}]$, **DO** *compare* $S(\mathcal{D}_1^k)$ *and* $S(\mathcal{D}_2^k)$ *by* (1.16). **IF** $S(\mathcal{D}_1^k) \neq S(\mathcal{D}_2^k)$, *conclude that* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are non-isomorphic and* **Break**.

Step 4: *If* $S(\mathcal{D}_1^k) = S(\mathcal{D}_2^k)$ *for all k, then we have no conclusion about the detection between* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *and reserve the two designs for other detection methods.*

We do not consider one or two-factor projection designs in our study. Hence, we take k starting with 3 in **Step 3**. Since the numbers of non-isomorphic classes detected by $S(\mathcal{D}^k)$ and projection HDP distribution are consistent, we also omit projections for $k = [\frac{s}{2}] + 1 : s - 1$ in Algorithm 2.

For projection ASC, we also propose an NIU algorithm as following Algorithm 3.

Algorithm 3. *NIU algorithm with projection ASC distribution*

Step 1*: If the given two designs have the same ASC value, go to step 2. Otherwise, conclude these two designs are non-isomorphic and* **Break**.

Step 2*: k=s-1.*

Step 3*: Find all projection ASC of k-factor projection designs of* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$.

Step 4*:* **IF** *the numbers of projection ASC are different, then conclude that* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are non-isomorphic and* **Break**.

Step 5*:* **Else IF** *the values of projection ASC are different, then conclude that* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are non-isomorphic and* **Break**.

Step 6*:* **Else IF** *the frequencies of projection ASC are different, then conclude that* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are non-isomorphic and* **Break**.

Step 7*: k=s-2.*

Step 8*: Run* ***Step 3 to 6***.

Step 9*:* **For** $k = 3 : [\frac{s}{2}]$, **DO** *run* ***Step 3 to 6****, and let* $k = s - k$, *run* ***Step 3 to 6***.

Step 10*: If the projection ASC distribution of* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$ *are the same for all k, we have no conclusion about the detection between* $\boldsymbol{D}_1$ *and* $\boldsymbol{D}_2$.

In NIU algorithm, once a difference is detected in any k-factor projection, the algorithm directly stops and leads to a non-isomorphism conclusion. To compare the elapsed time of NIU algorithms based on different criteria, we applied them to detect non-isomorphic groups of $\mathcal{L}_{18}(3^7)$, $\mathcal{L}_{27}(3^{13})$, and $\mathcal{L}_{32}(4^9)$ respectively, as shown in Tables 2.1, 2.2, and 2.3. Lam and Tonchev [40] has proved that there are 68 non-isomorphic classes of $\mathcal{L}_{27}(3^{13})$. However, these 68 classes of designs have the same GWLP and HDP values. Evangelaras et al. [15] proved that there exist only three classes for $\mathcal{L}_{18}(3^7)$, but these three classes have the same HDP and GWLP values as well. From the website `http://pietereendebak.nl/oapage/index.html`, there are 20 non-isomorphic classes of $\mathcal{L}_{32}(4^9)$.

In Table 2.1, NIU based on distance enumerator (DE) is the fastest while the entropy based on projection HDP ($S(\mathcal{D}^k)$) is still the second best. However, in larger designs, $S(\mathcal{D}^k)$ is faster than DE. $S(\mathcal{D}^k)$ is always faster than projection HDP. From Table 2.2, $S(\mathcal{D}^k)$ is slower than projection ASC since all 68 classes have the

Table 2.1: The Elapsed Time of Detecting $\mathcal{L}_{18}(3^7)$ by NIU with DE, projection HDP, $S(\mathcal{D}^k)$, and projection ASC

Number of classes	Criterion	Elapsed Time (s)
3	DE	0.014507
3	projection HDP	0.028911
3	$S(\mathcal{D}^k)$	0.025121
3	projection ASC	0.135572

Table 2.2: The Elapsed Time of Detecting $\mathcal{L}_{27}(3^{13})$ by NIU with DE, projection HDP, $S(\mathcal{D}^k)$, and projection ASC

Number of classes	Criterion	Elapsed Time (s)
67	DE	59.70944
67	projection HDP	20.294655
67	$S(\mathcal{D}^k)$	9.682262
68	projection ASC	2.1909

Table 2.3: The Elapsed Time of Detecting $\mathcal{L}_{32}(4^9)$ by NIU with DE, projection HDP, $S(\mathcal{D}^k)$, and projection ASC

Number of classes	Criterion	Elapsed Time (s)
18	DE	1.361057
18	projection HDP	0.212654
18	$S(\mathcal{D}^k)$	0.130290
20	projection ASC	37.173814

same HDP, and hence the detection by $S(\mathcal{D}^k)$ needs to conduct projections 68 times. However, 67 classes have different ASC, and hence NIU based on projection ASC can distinguish 67 classes in Step 1. $S(\mathcal{D}^k)$ is the fastest in Table 2.3, but two classes can not be detected. In three applications, projection ASC can distinguish all classes. Hence, $S(\mathcal{D}^k)$ is the most efficient detection criterion, while projection ASC is the most effective one. Therefore, we propose the detection flow chart as following Figure 2.1, in which NIU algorithm based on $S(\mathcal{D}^k)$ is conducted at the first stage, and NIU algorithm based on projection ASC is conducted at the second stage.

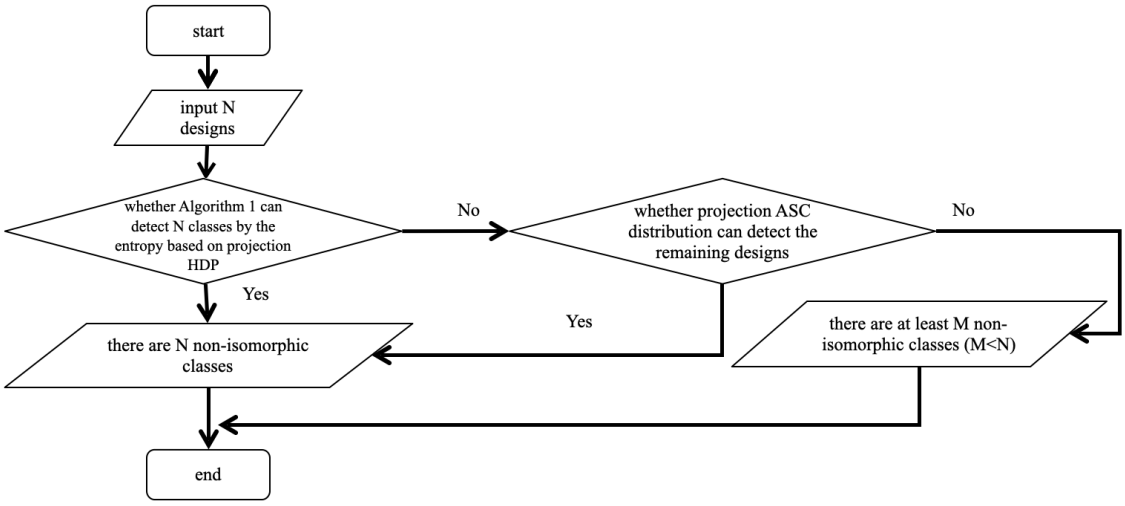


Figure 2.1: Suggested Flowchart of Detecting Combinatorially Non-isomorphic Designs

In this chapter, the non-isomorphism detected is called combinatorial. In traditional view, combinatorially non-isomorphic orthogonal designs are assumed to perform at the same level in statistical modeling. Starting with an example, the next chapter indicates that isomorphic orthogonal designs performs differently in the view of confounding pattern. Isomorphic operations can be separated into two sets so that the equivalence of designs is defined in Section 3.2. Based on this phenomenon, a new criterion evaluating the confounding pattern of saturated orthogonal designs is proposed in Section 3.3.

Chapter 3

A New Criterion Evaluating the Confounding Effects of Saturated Orthogonal Designs

3.1 An Example of Different Confounding Effects

During a very long period in the past, traditional opinions directly abandoned all isomorphic designs and considered that any isomorphic designs had the same performance in statistical modeling. Two isomorphic designs are considered to be equivalent because they share the same statistical properties in some classical ANOVA model where the main effects and some interactions of the factors are involved. However, Fang and Ma [20] indicated that two isomorphic designs may have different statistical performance in view of projection confounding effect after linear-quadratic decomposition. More details can be found in Fang and Ma [21] and some are given as follows.

It is known that there exists only one non-isomorphic class of $\mathcal{L}_9(3^4)$. In Table 3.1, two $L_9(3^4)$ designs, denoted by $L_9(3^4)$ and $UL_9(3^4)$, are isomorphic to each other. In this chapter, we denote a U-type design with n runs, s factors each having f levels as $U(n, f^s)$, and denote the set of U-type designs as $\mathcal{U}(n, f^s)$. $L_9(3^4)$ can be easily found in the literature while the second one $UL_9(3^4)$ was obtained by minimizing CD-value over the U-type class of $\mathcal{U}(9, 3^4)$, and is a uniform orthogonal $L_9(3^4)$ design called

by Fang and Ma [21]. A design is called uniform orthogonal, denoted by $UL_n(f^s)$, if it is orthogonal with the minimum discrepancy (CD for example) over the domain $\mathcal{U}(n, f^s)$. Fang and Ma [21] pointed out that $L_9(3^4)$ and $UL_9(3^4)$ have different CD-values 0.050059 and 0.0493645 respectively. However, $L_9(3^4)$ and $UL_9(3^4)$ have the same GWLP as $[1, 0, 0, 8, 0]$.

Table 3.1: Two $L_9(3^4)$ designs

No	$L_9(3^4)$				$UL_9(3^4)$			
1	1	1	1	1	1	1	1	2
2	1	2	2	2	1	2	3	1
3	1	3	3	3	1	3	2	3
4	2	1	2	3	2	1	3	3
5	2	2	3	1	2	2	2	2
6	2	3	1	2	2	3	1	1
7	3	1	3	2	3	1	2	1
8	3	2	1	3	3	2	1	3
9	3	3	2	1	3	3	3	2

Suppose that there are three factors A, B, and C with each having 3 levels in an experiment. We can choose any 3 columns from $L_9(3^4)$ or $UL_9(3^4)$ for the factors, as suggested by many authors. Suppose two projection designs are formed by the first three columns of $L_9(3^4)$ or $UL_9(3^4)$ respectively, the two projection designs have the same GWLP as $[1, 0, 0, 2]$. Recall that we have introduced an orthogonal polynomial decomposition technique in Chapter 1. Consequently, each column of $L_9(3^4)$ or $UL_9(3^4)$ can be split into two columns. The first six columns in Table 3.2 are the split result of the first three columns in $L_9(3^4)$. Table 3.2 also lists all two-factor interactions based on $L_9(3^4)$, where columns $7 - 18$ are produced by the dot product from the first six columns. We can also produce a similar table for $UL_9(3^4)$.

Through the regression analysis technique, the confounding situation by the use

Table 3.2: Main effects and interactions in $L_9(3^4)$ design

A_l	A_q	B_l	B_q	C_l	C_q	A_lB_l	A_lC_l	B_lC_l	A_qB_q	A_qC_q	B_qC_q	A_lB_q	A_lC_q	A_qB_l	A_qC_l	B_lC_q	B_qC_l
-1	1	-1	1	-1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
-1	1	0	-2	0	-2	0	0	0	-2	-2	4	2	2	0	0	0	0
-1	1	1	1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	1	1
0	-2	-1	1	0	-2	0	0	0	-2	4	-2	0	0	2	0	2	0
0	-2	0	-2	1	1	0	0	0	4	-2	-2	0	0	0	-2	0	-2
0	-2	1	1	-1	1	0	0	-1	-2	-2	1	0	0	-2	2	1	-1
1	1	-1	1	1	1	-1	1	-1	1	1	1	1	1	-1	1	-1	1
1	1	0	-2	-1	1	0	-1	0	-2	1	-2	-2	1	0	-1	0	2
1	1	1	1	0	-2	1	0	0	1	-2	-2	1	-2	1	0	-2	0

of $UL_9(3^4)$ is given by

$$\begin{cases} A_l = 0.5B_lC_q + 0.5B_qC_l, \ A_q = 1.5B_lC_l - 0.5B_qC_q, \\ B_l = 0.5A_lC_q + 0.5A_qC_l, \ B_q = 1.5A_lC_l - 0.5A_qC_q, \\ C_l = 0.5A_lB_q + 0.5A_qB_l, \ C_q = 1.5A_lB_l - 0.5A_qB_q. \end{cases} \tag{3.1}$$

On the other hand, by the use of $L_9(3^4)$, the aliasing statements are

$$\begin{cases} A_l = -0.75B_lC_l - 0.25B_lC_q + 0.25B_qC_l - 0.25B_qC_q, \\ A_q = 0.75B_lC_l - 0.75B_lC_q + 0.75B_qC_l + 0.25B_qC_q, \\ B_l = -0.75A_lC_l - 0.25A_lC_q + 0.25A_qC_l - 0.25A_qC_q, \\ B_q = 0.75A_lC_l - 0.75A_lC_q + 0.75A_qC_l + 0.25A_qC_q, \\ C_l = -0.75A_lB_l + 0.25A_lB_q + 0.25A_qB_l + 0.25A_qB_q, \\ C_q = -0.75A_lB_l - 0.75A_lB_q - 0.75A_qB_l + 0.25A_qB_q. \end{cases} \tag{3.2}$$

If the higher-order interactions $A_lB_q, A_qB_l, A_qB_q, \ldots, B_qC_q$ can be ignored, the alias statements for $UL_9(3^4)$ become

$$\begin{cases} A_q = 1.5B_lC_l, \\ B_q = 1.5A_lC_l, \\ C_q = 1.5A_lB_l. \end{cases} \tag{3.3}$$

In this case we can estimate all the linear effects A_l, B_l and C_l without any confound-

ing. Moreover, the alias statements for $L_9(3^4)$ become

$$\begin{cases} A_l = -0.75B_lC_l, & A_q = 0.75B_lC_l, \\ B_l = -0.75A_lC_l, & B_q = 0.75A_lC_l, \\ C_l = -0.75A_lB_l, & C_q = -0.75A_lB_l. \end{cases} \tag{3.4}$$

Therefore, the projection design by $UL_9(3^4)$ has better statistical performance in the sense of confounding among the main effects and two-factor interactions.

This example gives us several considerations:

1) the *isomorphism* concept is not suitable for quantitative factors in experimental design;

2) the confounding situation can be as a criterion for comparing different orthogonal designs;

3) the uniformity is useful in the theory of orthogonal designs for detecting non-equivalent factorial designs and we can define the so-called *uniform orthogonal design* and *uniform minimum aberration orthogonal design*.

This chapter is to develop these considerations. The level permutation operations of a given design are separated into two equivalent groups in which geometrically equivalent designs have the same discrepancy-value. Hence, when the experimenter considers to compare the level permuted designs of a given design, only $(\frac{f!}{2})^s$ permutations need to be explored instead of $(f!)^s$. A new criterion, the main effect confounding pattern (MECP), is proposed for any saturated orthogonal design (OD) in order to evaluate the confounding situation of its projection designs. MECP presents more details about the design in statistical performance, and hence it can be used to further distinguish equivalent ODs.

3.2 Equivalence of Designs

Section 3.1 shows that the concept of isomorphism is not suitable for quantitative factors in experimental design. Recall that two designs are (combinatorially) isomorphic if one can be obtained from the other by the joint operation of row permutation, column permutation and level permutation. Clack and Dean [10] used 'equivalence' instead of 'isomorphism'. Cheng and Ye [8] split the concept *isomorphism* into two:

"combinatorial isomorphism" and "geometric isomorphism" depending on whether the factors are qualitative or quantitative. This may cause some confusion between the original combinatorial isomorphism and new combinatorial isomorphism; hence, we prefer to use "combinatorial equivalence" and "geometric equivalence" instead of "combinatorial isomorphism" and "geometric isomorphism" for distinguishing between the concepts of isomorphism and equivalence.

Definition 9. *Two designs for a fractional factorial experiment with qualitative factors are combinatorially equivalent if one can be obtained from the other by (i) reordering the treatment combinations; (ii) relabeling the factors; (iii) relabeling the levels of one or more factors.*

Two designs for a fractional factorial experiment with quantitative factors are geometrically equivalent if one can be obtained from the other by (i) reordering the treatment combinations; (ii) relabeling the factors; (iii) reversing the label order of the levels of one or more factors.

From definition 9, we can see that the combinatorial equivalence is just the isomorphism for qualitative factors and the geometric equivalence for the quantitative factors is related to the measure of uniformity. In fact, we can find more general results.

Theorem 1. *All the geometrically equivalent $L_n(f^s)$ designs have the same WD/CD/MD discrepancy-values.*

Proof. It is easy to verify that the terms in CD (1.7), MD (1.8), and WD (1.9) are invariant under the three operations (i), (ii) and (iii) for the geometric equivalence in definition 9. For instance, we may consider $\mathrm{MD}^2(\mathcal{P})$ as a summation of m_{ij}, where

$$m_{ij} = \begin{cases} \frac{1}{n^2}\prod_{k=1}^{s}\frac{1}{2}(\frac{15}{4} - \frac{1}{2}|z_{ik}| - \frac{1}{2}|z_{jk}| - \frac{3}{2}|z_{ik}-z_{jk}| + (z_{ik}-z_{jk})^2), & \text{if } i \neq j \\ \frac{1}{n^2}\prod_{k=1}^{s}\frac{1}{2}(\frac{15}{4} - |z_{ik}|) - \frac{2}{n}\prod_{k=1}^{s}(\frac{5}{3} - \frac{1}{4}|z_{ik}| - \frac{1}{4}z_{ik}^2), & \text{otherwise,} \end{cases} \tag{3.5}$$

where $z_{ik} = x_{ik} - 0.5$. It can be verified that

$$\mathrm{MD}^2(\mathcal{P}) = (\frac{19}{12})^s + \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}. \tag{3.6}$$

Exchanging rows or columns, *i.e.* operations (i), (ii) of the geometric equivalence, will not change the values of m_{ij}. Furthermore, if we reverse the label ordering of

the levels of any column of $\boldsymbol{X}$, the values of m_{ij} are still invariant. Hence, operation (iii) of the geometric equivalence will not change the MD-values, either. In a similar way, we can prove that WD and CD-values are invariant under the geometrically equivalent operations as above. □

Subsequently, for any $L_n(f^s)$ design, the $(f!)^s$ level permutations in combinatorial equivalence can be reduced to $(\frac{f!}{2})^s$ level permutations. For instance, considering each column of a given $L_n(3^s)$ design, let $\{0\ 1\ 2\}$ denote the original column where $0, 1, 2$ represents the three levels respectively. Then level permutation plan $\{1\ 0\ 2\}$ represents exchanging the levels 0 and 1 of the corresponding column. As shown in Table 3.3, there are 6 level permutations in total that can be separated into two geometrically equivalent groups. Similarly, for 4-level $L_n(4^s)$ designs, 24 level permutations of each column can be considered as 12 pairs of geometrically equivalent operations (see Table 3.4). As mentioned in Cheng and Ye [8] and Tang and Xu [52], the geometric structure of a design is invariant by reversing the label ordering of levels. Hence, as suggested in Tang and Xu (2014), when considering level permutations of a given design, the experimenter can reduce a great amount of computation time based on Theorem 1.

Table 3.3: Equivalent permutations in $L_n(3^s)$

#	Group 1	Geo. Equivalent	Group 2
1	0 1 2	$\Longleftrightarrow$	2 1 0
2	1 2 0	$\Longleftrightarrow$	1 0 2
3	2 0 1	$\Longleftrightarrow$	0 2 1

Table 3.4: Equivalent permutations in 4-level case

#	Group 1	Geo. Equivalent	Group 2
1	1 2 3 4	$\Longleftrightarrow$	4 3 2 1
2	2 3 4 1	$\Longleftrightarrow$	3 2 1 4
3	3 4 1 2	$\Longleftrightarrow$	2 1 4 3
4	4 1 2 3	$\Longleftrightarrow$	1 4 3 2
5	1 2 4 3	$\Longleftrightarrow$	4 3 1 2
6	2 3 1 4	$\Longleftrightarrow$	3 2 4 1
7	3 4 2 1	$\Longleftrightarrow$	2 1 3 4
8	4 1 3 2	$\Longleftrightarrow$	1 4 2 3
9	1 3 2 4	$\Longleftrightarrow$	4 2 3 1
10	2 4 3 1	$\Longleftrightarrow$	3 1 2 4
11	3 1 4 2	$\Longleftrightarrow$	2 4 1 3
12	4 2 1 3	$\Longleftrightarrow$	1 3 4 2

3.3 The Main Effect Confounding Pattern

Let us continue the discussion on the example of Section 3.1. Following the effect hierarchy principle indicated by Cheng and Ye [8], we indicate that linear main effects, such as A_l, is of order one; quadratic main effects and interactions between linear main effects, such as A_q, A_lB_l, is of order 2; interactions between linear and quadratic main effects, such as A_lB_q, is of order 3, and interactions between quadratic main effects, such as A_qB_q, is of order 4. From (3.1), we can see that there is no confounding between A_l and any item with order 1, 2 and 4, and there are 2 confounding situations between A_l and B_lC_q plus A_l and B_qC_l. We express this situation as $A_l : (0, 0, 2, 0)$. In a similar way we obtain:

$$\begin{aligned}
&A_l : (0, 0, 2, 0), \quad A_q : (0, 1, 0, 1),\\
&B_l : (0, 0, 2, 0), \quad B_q : (0, 1, 0, 1),\\
&C_l : (0, 0, 2, 0), \quad C_q : (0, 1, 0, 1),\\
&\boldsymbol{l}(UL_9(3^4)) = (0, 0, 6, 0), \quad \boldsymbol{q}(UL_9(3^4)) = (0, 3, 0, 3),
\end{aligned}$$

where $\boldsymbol{l}(UL_9(3^4)) = (0, 0, 6, 0)$ is the sum of vectors for the linear main effects and $\boldsymbol{q}(UL_9(3^4)) = (0, 3, 0, 3)$ is the sum for the quadratic main effects. The main effect confounding pattern (MECP) is defined by

$$\boldsymbol{m}(UL_9(3^4)) = (\boldsymbol{l}(UL_9(3^4)), \boldsymbol{q}(UL_9(3^4))) = (0, 0, 6, 0, 0, 3, 0, 3). \tag{3.7}$$

Alternatively, we may further merge $\boldsymbol{l}(UL_9(3^4))$ and $\boldsymbol{q}(UL_9(3^4))$ according to the confounding order as follows.

$$\boldsymbol{m}^*(UL_9(3^4)) = (0, 0, 6, 0, 0) + (0, 0, 3, 0, 3) = (0, 0, 9, 0, 3). \tag{3.8}$$

For the detailed definition of $\boldsymbol{m}^*(D)$, please refer to (3.10). Similarly, we can calculate the corresponding confounding vectors for $L_9(3^4)$

$$\begin{aligned}
&\boldsymbol{l}(L_9(3^4)) = (0, 3, 6, 3), \quad \boldsymbol{q}(L_9(3^4)) = (0, 3, 6, 3), \text{and}\\
&\boldsymbol{m}(L_9(3^4)) = (0, 3, 6, 3, 0, 3, 6, 3), \ \boldsymbol{m}^*(UL_9(3^4)) = (0, 3, 9, 9, 3).
\end{aligned}$$

If we take the dictionary ordering rule (see Fang et al [25]), we can conclude that $UL_9(3^4)$ has less confounding than $L_9(3^4)$. Based on the above discussion we can

define a new criterion for 3-level designs. It depends on the practical consideration about higher order main effects to adopt either the MECP or the merged MECP. For example, if quadratic main effects in this case can be ignored in the practical experiments, then MECP should be adopted. Otherwise, the merged MECP may be more fair for quadratic main effects than the MECP. For physical experiments, experimenters can calculate these two MECPs and decide which design should be chosen.

Definition 10. *Let D be an orthogonal design matrix with n runs and s factors (denoted by $F_1, \ldots, F_s$) each having 3 levels. Decompose each column of D into two orthogonal columns (denoted by $F_{il}, F_{iq}, i = 1, \ldots, s$) by the orthogonal polynomial technique given in Section 3.1. Let m_{il}^j denote the confounding number between F_{il} and the items of the main effect/interactions with order j, and let m_{iq}^j denote the confounding number between F_{iq} and the items of the main effect/interactions with order j. The main effect confounding pattern (MECP)is defined by*

$$\boldsymbol{m}(D) = (l_1(D), \ldots, l_4(D), q_1(D), \ldots, q_4(D)) = (\boldsymbol{l}(D), \boldsymbol{q}(D)), \tag{3.9}$$

where

$$l_j(D) = \sum_{i=1}^{s} m_{il}^j,\ q_j(D) = \sum_{i=1}^{s} m_{iq}^j,\ j = 1, \ldots, 4,$$
$$\boldsymbol{l}(D) = (\sum_{i=1}^{s} m_{il}^1, \ldots, \sum_{i=1}^{s} m_{il}^4),\ \boldsymbol{q}(D) = (\sum_{i=1}^{s} m_{iq}^1, \ldots, \sum_{i=1}^{s} m_{iq}^4).$$

We may further merge $\boldsymbol{l}(D)$ and $\boldsymbol{q}(D)$ by the order of confounding relation and derived merged MECP $\boldsymbol{m}^*(D)$

$$\boldsymbol{m}^*(D) = (l_1(D),\ l_2(D) + q_1(D),\ l_3(D) + q_2(D),\ l_4(D) + q_3(D),\ q_4(D)). \tag{3.10}$$

The MECP can be extended to comparing two $L_n(f^s), f > 2$ designs by decomposing one column of the design matrix into $(f-1)$ orthogonal columns with the orthogonal polynomial method. Similar to comparing GWLP, by the use of the dictionary ordering we can compare two design D_1 and D_2 in $\mathcal{L}_n(f^s)$ as in definition 11.

Definition 11. *For two $L_n(f^s), f > 2$ designs D_1 and D_2, D_1 is said to have less (better) confounding than D_2 if there exists an $\theta \in \{1, 2, \ldots, 2(f-1)^2\}$, such that $\boldsymbol{m}_\theta(D_1) < \boldsymbol{m}_\theta(D_2)$ and $\boldsymbol{m}_i(D_1) = \boldsymbol{m}_i(D_2)$ for $i = 1, \ldots, \theta - 1$.*

If the experimenter adopts the merged MECP, $\boldsymbol{m}^*(D)$, the comparison follows the similar manner.

3.4 Applications to Three- and Four-level Orthogonal Designs

Application to $L_{27}(3^{13})$

The MECP can help to select the best design for estimating the main effects and interested interactions. Recall in Section 3.1, the projection designs formed by the first three columns of $L_9(3^4)$ and $UL_9(3^4)$ have the same GWLP value but differ in the confounding situation. We explore 4-factor and 5-factor projection designs of $L_{27}(3^{13})$ and find the consistency between MECP and GWLP. However, MECP can distinguish two designs having the same GWLP. We also compare the best projection designs in the view of MECP and other criteria including GWLP and discrepancies.

For 4-factor projection, there are $\binom{13}{4} \times 68 = 48620$ subdesigns in total. Elsawah et al. [13] used MD, WD, CD, and GWLP to classify these projection designs. For each set of k-factor projection designs of each non-isomorphic group, where $3 \leq k \leq 13$, Elsawah et al. [13] classified the subdesigns through either the distribution of discrepancy values under permutations or the distribution of projection GWLP (PGWLP) values without permutation. The result shows that the number of k-factor subgroups classified by these different criteria is the same. Let n_ζ^4 denote the number of non-isomorphic groups for 4-factor projection designs of $L_{27}^\zeta(3^{13})$, $\zeta = 1, ..., 68$, classified by [13]. If we use MECP to replace PGWLP, then we will have the results shown in Table 3.7, where m_ζ^4 represents the number of non-equivalent 4-factor projection designs of $L_{27}^\zeta(3^{13})$ by MECP. Since level permutation may alter MECP values, we use "non-equivalence" for two designs having different MECP. In Table 3.7, m_ζ^4 is all greater than n_ζ^4 since the MECP provides more detailed information of the design than the GWLP.

If we sort the MECP of all 48620 4-factor projection designs, then we will find 2086 MECP values partially displayed in Table 3.5. The best MECP is the vector of all zeros that can be obtained by all 68 groups of $L_{27}(3^{13})$ with 2440 occurrences

in total. This implies that the 4-factor main effects can be estimated without any two-factor interaction for each out of 68 non-isomorphic groups in $L_{27}(3^{13})$.

If we compare the best 4-factor projection designs in MECP with the ones in GWLP and uniformity, we find that the 2440 best designs in MECP indicated in Table 3.5 are the same as in GWLP. Moreover, some of these best designs in MECP can obtain the best MD and CD-values respectively. The first column in Table 3.6 is the best values over the whole 4-factor space provided by Elsawah et al. [13].

Similarly, let n_ζ^5 denote the number of non-isomorphic groups classified by [13] for 5-factor projection designs of $L_{27}^\zeta(3^{13})$, $\zeta = 1, \ldots, 68$, and let m_ζ^5 denote the number of MECP values. Table 3.8 presents the results we obtained. If we sort all $\binom{13}{5} \times 68$=87516 5-factor projection designs, then we can get 30476 different MECP values. When comparing the best MD, CD and GWLP values of the 18 5-factor projection designs that have the best MECP indicated in Table 3.9 with the record of [13], we found that only $L_{27}^1(3^{13})$ can obtain the best 5-factor projection MECP, [0,0,6,0,0,0,12,3], with 18 occurrences. However, only $L_{27}^7(3^{13})$ can obtain the best MD and CD values. Even though $L_{27}^1(3^{13})$ performs not so good as $L_{27}^7(3^{13})$ in the view of discrepency, it performs the best in the view of MECP. Moreover, $L_{27}^1(3^{13})$ is a regular orthogonal design. As for the GWLP, all the 18 projection designs having the best MECP have the best GWLP as well. There are $L_{27}^\zeta(3^{13})$, ζ =1, ..., 66, that are able to obtain the best projection GWLP, but only $L_{27}^1(3^{13})$ has the best projection MECP. Hence, $L_{27}^1(3^{13})$ may be the recommended design among the 68 non-isomorphic groups.

Table 3.5: 2086 MECP values of 4-factor projection designs of $L_{27}^{\zeta}(3^{13})$ with frequencies

No	MECP	Frequency
1	[0,0,0,0,0,0,0,0]	2440
2	[0,0,4,0,0,2,0,0]	516
3	[0,0,4,2,0,2,4,0]	152
4	[0,0,6,0,0,3,0,3]	2238
5	[0,0,6,2,0,3,4,0]	36
6	[0,0,8,0,0,3,0,3]	152
7	[0,0,8,0,0,3,2,3]	76
8	[0,0,8,0,0,4,0,0]	279
9	[0,0,8,2,0,4,4,0]	71
10	[0,0,8,4,0,4,8,0]	144
...	...	...
2080	[0,12,24,12,0,12,22,11]	4
2081	[0,12,24,12,0,12,22,12]	31
2082	[0,12,24,12,0,12,23,11]	8
2083	[0,12,24,12,0,12,23,12]	112
2084	[0,12,24,12,0,12,24,10]	3
2085	[0,12,24,12,0,12,24,11]	28
2086	[0,12,24,12,0,12,24,12]	636
Total		48620

Table 3.6: The best MD, CD and GWLP values of the 2440 4-factor projection designs having the best MECP of $L_{27}^{\zeta}(3^{13})$, ζ =1, ..., 68

	Best Values in the record	Best Values out of 2440 projection designs	Frequency out of 2440	The Corresponding Non-isomorphic groups, $\zeta = 1, \ldots, 68$
MD	0.2340031	0.2340031	308	[1,2,3,8,9,10,11,12,59,60,61,62,63,64,65,66,67,68]
CD	0.0465474	0.0465474	308	[1,2,3,8,9,10,11,12,59,60,61,62,63,64,65,66,67,68]
GWLP	[1,0,0,0,2]	[1,0,0,0,2]	2440	All 68 Non-isomorphic groups.

Table 3.7: Group number of 4-factor projection designs of $L_{27}^{\zeta}(3^{13})$ by PGWLP and MECP

	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	n_{16}	n_{17}
n_ζ^4	3	9	6	9	11	12	6	25	18	15	16	10	11	26	16	27	20
m_ζ^4	7	48	54	61	48	83	30	147	153	146	124	139	52	159	97	157	103
	n_{18}	n_{19}	n_{20}	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	n_{26}	n_{27}	n_{28}	n_{29}	n_{30}	n_{31}	n_{32}	n_{33}	n_{34}
n_ζ^4	30	29	29	30	11	30	21	12	19	29	16	18	28	27	9	16	29
m_ζ^4	159	156	163	177	44	154	93	48	74	137	77	67	130	140	25	66	181
	n_{35}	n_{36}	n_{37}	n_{38}	n_{39}	n_{40}	n_{41}	n_{42}	n_{43}	n_{44}	n_{45}	n_{46}	n_{47}	n_{48}	n_{49}	n_{50}	n_{51}
n_ζ^4	33	35	33	29	24	30	23	14	12	14	24	33	20	22	31	33	27
m_ζ^4	178	194	185	170	136	165	156	148	119	160	132	184	198	196	184	194	234
	n_{52}	n_{53}	n_{54}	n_{55}	n_{56}	n_{57}	n_{58}	n_{59}	n_{60}	n_{61}	n_{62}	n_{63}	n_{64}	n_{65}	n_{66}	n_{67}	n_{68}
n_ζ^4	27	31	31	29	17	9	10	36	40	41	35	36	16	27	9	8	13
m_ζ^4	238	192	181	178	142	48	75	256	205	237	168	236	127	287	290	169	248

Table 3.8: Group number of 5-factor projection designs of $L_{27}^{\zeta}(3^{13})$ by PGWLP and MECP

	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	n_{16}	n_{17}
n_ζ^5	3	13	8	12	17	19	7	57	40	34	32	19	15	55	32	63	47
m_ζ^5	24	415	493	610	514	547	420	777	817	830	809	894	372	786	705	750	599
	n_{18}	n_{19}	n_{20}	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	n_{26}	n_{27}	n_{28}	n_{29}	n_{30}	n_{31}	n_{32}	n_{33}	n_{34}
n_ζ^5	77	70	77	76	15	76	53	18	41	75	31	40	68	64	13	36	62
m_ζ^5	726	775	844	888	327	699	551	381	490	713	606	531	710	881	236	495	831
	n_{35}	n_{36}	n_{37}	n_{38}	n_{39}	n_{40}	n_{41}	n_{42}	n_{43}	n_{44}	n_{45}	n_{46}	n_{47}	n_{48}	n_{49}	n_{50}	n_{51}
n_ζ^5	74	76	71	74	57	74	53	26	24	26	57	76	39	44	63	68	66
m_ζ^5	795	830	819	807	768	778	816	895	855	885	749	849	895	1022	796	785	970
	n_{52}	n_{53}	n_{54}	n_{55}	n_{56}	n_{57}	n_{58}	n_{59}	n_{60}	n_{61}	n_{62}	n_{63}	n_{64}	n_{65}	n_{66}	n_{67}	n_{68}
n_ζ^5	60	66	74	64	29	11	14	78	119	128	101	94	37	73	16	19	22
m_ζ^5	963	818	819	810	989	339	518	1062	778	1005	695	944	638	1072	1113	694	910

Table 3.9: 30476 MECP values of 5-factor projection designs of $L_{27}^{\zeta}(3^{13})$ with frequencies

#	MECP	Frequency
1	[0,0,6,0,0,0,12,3]	18
2	[0,0,6,0,0,3,0,3]	50
3	[0,0,6,0,0,3,0,4]	70
4	[0,0,6,0,0,3,0,5]	138
5	[0,0,6,0,0,3,0,6]	238
6	[0,0,7,0,0,3,0,3]	5
7	[0,0,7,0,0,3,0,4]	133
8	[0,0,7,0,0,3,0,5]	221
9	[0,0,7,0,0,3,0,6]	83
10	[0,0,8,0,0,3,0,3]	3
...	...	...
30470	[0,30,60,30,0,29,60,30]	9
30471	[0,30,60,30,0,30,57,30]	2
30472	[0,30,60,30,0,30,58,30]	2
30473	[0,30,60,30,0,30,59,29]	6
30474	[0,30,60,30,0,30,59,30]	7
30475	[0,30,60,30,0,30,60,29]	5
30476	[0,30,60,30,0,30,60,30]	37
Total		87516

Table 3.10: The best MD, CD and GWLP values of the 18 5-factor projection designs having the best MECP of $L_{27}^{\zeta}(3^{13})$, ζ =1, ..., 68

	Best Values in the record	Best Values out of 18 projection designs	Frequency out of 18	The Corresponding Non-isomorphic groups, ζ =1∼68
MD	0.4742819	0.4745399	18	$L_{27}^{1}(3^{13})$
CD	0.0633356	0.0636887	18	$L_{27}^{1}(3^{13})$
GWLP	[1,0,0,2,6,0]	[1,0,0,2,6,0]	18	$L_{27}^{1}(3^{13})$.

Application to 4-level saturated orthogonal designs

MECP is defined for 3-level saturated orthogonal designs in definition 10; however, it can be easily extended to more level saturated designs.

Definition 12. *Let D be an orthogonal design matrix with n runs and s factors (denoted by $F_1, \ldots, F_s$) each having 4 levels. Decompose each column of D into three orthogonal columns for linear, quadratic and cubic main effects (denoted by $F_{il}, F_{iq}, F_{ic},\ i = 1, \ldots, s$) by the orthogonal polynomial technique. Fang et al. [24] indicated the decomposition was derived by replacing levels $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$ with $\{-3, 1, -1\}$, $\{-1, -1, 3\}$, $\{1, -1, -3\}$, $\{3, 1, 1\}$ respectively. Let m_{il}^j denote the confounding number between F_{il} and the items of the interactions with type j, and let m_{iq}^j denote the confounding number between F_{iq} and the items of the interactions with type j, and let m_{ic}^j denote the confounding number between F_{ic} and the items of the interactions with type j The main effect confounding patter (MECP) is defined by*

$$\boldsymbol{m}(D) = (l_1(D), \ldots, l_6(D), q_1(D), \ldots, q_6(D), c_1(D), \ldots, c_6(D)) = (\boldsymbol{l}(D), \boldsymbol{q}(D), \boldsymbol{c}(D)),$$

where

$$\begin{aligned}
&l_j(D) = \sum_{i=1}^{s} m_{il}^j,\ q_j(D) = \sum_{i=1}^{s} m_{iq}^j,\quad c_j(D) = \sum_{i=1}^{s} m_{ic}^j,\ j = 1, \ldots, 6,\\
&\boldsymbol{l}(D) = (\sum_{i=1}^{s} m_{il}^1, \ldots, \sum_{i=1}^{s} m_{il}^6),\ \boldsymbol{q}(D) = (\sum_{i=1}^{s} m_{iq}^1, \ldots, \sum_{i=1}^{s} m_{iq}^6),\\
&\boldsymbol{c}(D) = (\sum_{i=1}^{s} m_{ic}^1, \ldots, \sum_{i=1}^{s} m_{ic}^6).
\end{aligned}$$

We may further merge $\boldsymbol{l}(D)$, $\boldsymbol{q}(D)$ and $\boldsymbol{c}(D)$ by the order of relation and derived merged MECP $\boldsymbol{m}^*(D)$

$$\begin{aligned}
\boldsymbol{m}^*(D) \quad &= (l_1(D),\ l_2(D) + q_1(D),\ l_3(D) + q_2(D) + c_1(D),\ l_4(D) + q_3(D) + c_2(D),\\
&\quad l_5(D) + q_4(D) + c_3(D),\ l_6(D) + q_5(D) + c_4(D),\ q_6(D) + c_5(D),\ c_6(D)).
\end{aligned}$$

We explore projection designs of $L_{16}(4^5)$ in a similar way as we did for the case of $L_{27}(3^{13})$. The design $L_{16}(4^5)$ in Table 3.11 is downloaded from the website established by Eendebak and Schoen, `http://pietereendebak.nl/oapage/index.html`. There is only one non-isomorphic group for $L_{16}(4^5)$.

When exploring its 3-factor and 4-factor projection designs, we calculate the GWLP, HDP, MD, CD and MECP values of all $\binom{5}{k}, k = 3, 4$ projection designs and use these values to classify projection designs. For either 3 or 4-factor projection designs, there is only one GWLP or HDP value even with level permutations. For 3-factor projection designs, the GWLP value is $[1, 0, 0, 3]$, and the HDP value is $[0, 0, 1, 72]$. For 4-factor, the GWLP value is $[1, 0, 0, 12, 3]$, and the HDP value is $[0, 0, 0, 1, 96]$. However, there are 3 MECP values for 3-factor projection designs and 4 MECP values for 4-factor as in Table 3.12. For 3-factor projection designs, the best MECP value can be obtained by the 4 designs formed by the columns {2, 3, 4}, {2, 3, 5}, {2, 4, 5}, {3, 4, 5} of $L_{16}(4^5)$ in Table 3.11. As for 4-factor, there exists only one best projection design formed by column {2, 3, 4, 5} of $L_{16}(4^5)$ in Table 3.11.

There are 4 MD values for 3-factor projection designs and 2 MD values for 4-factor as in Table 3.13. There are 3 CD values for 3-factor projection designs and 2 CD values for 4-factor as in Table 3.14.

Table 3.11: $L_{16}(4^5)$

0	0	0	0	0
0	1	1	1	1
0	2	2	2	2
0	3	3	3	3
1	0	1	2	3
1	1	0	3	2
1	2	3	0	1
1	3	2	1	0
2	0	2	3	1
2	1	3	2	0
2	2	0	1	3
2	3	1	0	2
3	0	3	1	2
3	1	2	0	3
3	2	1	3	0
3	3	0	2	1

Table 3.12: The MECP classification of all $\binom{5}{k}$ projection designs for $L_{16}(4^5)$

	3-factor of $L_{16}(4^5)$		4-factor of $L_{16}(4^5)$	
	MECP Values	Frequency	MECP Values	Frequency
1	[0,0,6,0,6,0,0,3,0,6,0,3,0,0,6,0,6,0]	4	[0,0,8,0,8,0,0,4,0,8,0,4,0,0,8,0,8,0]	1
2	[0,3,0,6,0,3,0,0,0,3,0,0,0,3,0,6,0,3]	2	[0,5,0,10,0,5,0,4,0,9,0,5,0,5,0,10,0,5]	1
3	[0,3,2,7,2,3,0,1,2,2,2,1,0,3,2,7,2,3]	4	[0,5,0,10,0,5,0,5,0,9,0,4,0,5,0,10,0,5]	1
4			[0,6,0,12,0,6,0,5,0,9,0,4,0,6,0,12,0,6]	2
Total		10		5

Table 3.13: The MD classification of all $\binom{5}{k}$ projection designs for $L_{16}(4^5)$

	3-factor of $L_{16}(4^5)$		4-factor of $L_{16}(4^5)$	
	MD Values	Frequency	MD Values	Frequency
1	0.0596003	4	0.1278681	1
2	0.0596651	2	0.1285225	4
3	0.0596880	2		
4	0.0598749	2		
Total		10		5

Table 3.14: The CD classification of all $\binom{5}{k}$ projection designs for $L_{16}(4^5)$

	3-factor of $L_{16}(4^5)$		4-factor of $L_{16}(4^5)$	
	CD Values	Frequency	CD Values	Frequency
1	0.0189220	4	0.0284552	1
2	0.0190441	4	0.0293979	4
3	0.0195934	2		
Total		10		5

In this chapter, MECP provides a way for an experimenter to obtain the best arrangement of main effects for saturated orthogonal designs. In many applications of the existing literature reviewed in the next chapter, the modeling performance of orthogonal designs has space for improvement. Especially for many problems in chemometrics, the performance of orthogonal designs is not the best, compared to D-optimal designs and uniform designs. The D-optimal design is the most effective whereas the uniform design is the most robust. In next chapter, we propose a way to

combine these three experimental design methods in order to improve the modeling performance and guarantee both effectiveness and robustness.

Chapter 4

New Robust Composite Designs for Chemometrics and Computer Experiments

If the factors are qualitative, we have introduced various traditional experimental design methods including orthogonal designs, orthogonal D-optimal designs, and uniform designs in Chapter 1. The traditional design methods have been successfully implemented to solve problems in chemometrics. For instance, traditionally, chemists used the deterministic methods according to the characteristic of a reaction to obtain the kinetic rate constants. Once Atkinson et al. [1] first introduced D-optimal design (DOD for short) to estimate rate constants in chemical kinetic model of a reversible reaction by computer experiments, their method showed evidently improvement compared to the traditional ones. Therefore, for chemists, improving experimental design methods has been an issue of common concern.

Xu et al. [60] considered the uniform design (UD) and orthogonal design (OD) as alternatives for the reversible reaction, and showed that the uniform design is the most robust against model change (also see Hickernell and Liu [31]). Xu et al. [61] provided a brief introduction to the recent theoretical developments on uniform design and its applications in chemometrics. They reviewed that UD has been applied in various chemical problems. For instance, there are some optimization problems of the experimental conditions without modeling, in which the optimal results are

derived from the best one of experiments arranged through uniformly distributed design points. In Fang et al [28], UD has been used for the optimization of experimental factors in micelle-mediated extraction and preconcentration of ginsenosides from Chinese herbal medicine. Ji et al. [34] has used UD to find the best condition for a fingerprint of Ginkgo biloba extracts by high-performance liquid chromatography. Another kind of optimization problem is based on a regression model, and UD is a good way to estimate the parameters of the regression model. As for parameter estimation, UD sequential optimization has been applied in Lu et al. [44] for optimizing the extraction conditions of emodin and physcion with supercritical CO_2 and an ethanol modifier. Wang et al. [53] used UD and 3-dimensional response surfaces to investigate the effects of storage condition factors on fungal invasion of Radix Ophiopogonis. Space-filling designs are also of great importance for multivariate nonlinear chemical system. The aforementioned instance of rate constants in chemical kinetic model of a reversible reaction is one of such problem.

Liang et al. [41] gave a comprehensive study on the performance of three kinds of designs, orthogonal design, D-optimal design and uniform design, in chemistry and chemical engineering. They concluded that the D-optimal design is the most efficient in the sense of statistical estimation when the underlying model is known, but is not robust against model change. They also mentioned that several experimental design specialists have claimed that when D-optimality is not feasible for application, a design is recommended if having an evenly spaced distribution of points. The uniform design perfectly satisfying the requirement is indeed robust against model change in their study, but it may be lack of effectiveness under some circumstances. The orthogonal design has a good performance in many situations, but there is a space for improvement.

For improving the D-efficiency, Chan et al. [5, 6] combined the orthogonal design (OD) and D-optimal design under the polynomial regression model and gave a theoretic approach for this kind of design. Fang et al. [22] gave some applications according to Chan et al. [6]'s suggestion, related to the so-called *orthogonal D-optimal designs*. However, the effectiveness of D-optimal designs can be achieved to the best only when the underlying model is known.

Chemometrics especially quantitative structure activity relationship (QSAR) and quantitative structure-property relationship (QSPR) attempt to establish relationships between molecule descriptors and activities or properties. Most models in QSAR research, including ordinary least squares regression, principal components regression, partial least squares regression, and multivariate adaptive regression splines, assume the random errors to be independently identical distributed. This assumption can not be met in many practical cases. Fang et al. [27] indicated that Kriging models which allow dependent random errors in the model improve the performances of many existing methods on problems in QSAR.

There is another perspective from that we propose a new robust composite design for chemometrics and computer experiments guaranteeing both robustness and effectiveness.

4.1 Construction of Composite Designs

We propose to use two designs (called as initial design) to form a composite design. Each initial design should have some advantages in efficiency and/or robustness. According to the review in Chapter 1, we adopt five candidate initial designs from orthogonal, uniform and orthogonal D-optimal designs respectively. In this thesis, we explore on 16-run experiments as example. For a factor with the experimental domain $[a, b]$, we can easily get four levels from the five candidate initial designs in $[-1, 1]$, denoted as $q_{[-1,1]}$, into the practical four levels in $[a, b]$, denoted as $q_{[a,b]}$, by a linear transformation, $q_{[a,b]} : a + \frac{b-a}{2}(q_{[-1,1]} + 1)$. The detailed construction of the five initial designs over $[-1, 1]^s$ is as follows.

Chan et al. [6] indicated that for one-factor q-level experiments, the set of equidistant points, $\{\frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{2q-1}{2q}\}$, is the unique uniform design on $[0, 1]$ under many discrepancies including star discrepancy and centered L_2-discrepancy. According to this manner, we further consider an orthogonal design taking evenly spread levels over a shrunk experimental region, denoted as $L_n^S(q^s)$. For instance, $L_{16}^S(4^5)$ takes levels $\{\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}\}$ on $[0, 1]^5$, that is $\{-\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4}\}$ on $[-1, 1]^5$, where the experimental region has actually been shrunk by $\frac{1}{2q}$. Similarly, we also consider a uniform de-

sign $U_n(q^s)$ over an experimental region shrunk by $\frac{1}{2q}$, denoted as $U_n^R(q^s)$. We first pick up an orthogonal table, $L_{16}(4^5)$, from a website established by Eendebak and Schoen, `http://www.pietereendebak.nl/oapackage/series.html` and a uniform table, $U_{16}(16^5)$, from Fang and Ma [21] or `https://dst.uic.edu.cn/isci/Uniform/Uniform_Design_Tables.htm` respectively. Subsequently, we list the aforementioned five candidate initial designs over $[-1,1]^5$ design region, where Table 4.1 gives three initial orthogonal designs $L_{16}^E(4^5), L_{16}^S(4^5), L_{16}^D(4^5)$ and Table 4.2 presents other two initial uniform designs $U_{16}(16^5), U_{16}^R(16^5)$.

Table 4.1: $L_{16}^E(4^5)$, $L_{16}^S(4^5)$, and $L_{16}^D(4^5)$ over $[-1,1]^5$ design region

$\boldsymbol{L}^E = L_{16}^E(4^5)$					$\boldsymbol{L}^S = L_{16}^S(4^5)$					$\boldsymbol{L}^D = L_{16}^D(4^5)$				
−1.00	−1.00	−1.00	−1.00	−1.00	−0.75	−0.75	−0.75	−0.75	−0.75	−1.00	−1.00	−1.00	−1.00	−1.00
−1.00	−0.33	−0.33	−0.33	−0.33	−0.75	−0.25	−0.25	−0.25	−0.25	−1.00	−0.45	−0.45	−0.45	−0.45
−1.00	0.33	0.33	0.33	0.33	−0.75	0.25	0.25	0.25	0.25	−1.00	0.45	0.45	0.45	0.45
−1.00	1.00	1.00	1.00	1.00	−0.75	0.75	0.75	0.75	0.75	−1.00	1.00	1.00	1.00	1.00
−0.33	−1.00	−0.33	0.33	1.00	−0.25	−0.75	−0.25	0.25	0.75	−0.45	−1.00	−0.45	0.45	1.00
−0.33	−0.33	−1.00	1.00	0.33	−0.25	−0.25	−0.75	0.75	0.25	−0.45	−0.45	−1.00	1.00	0.45
−0.33	0.33	1.00	−1.00	−0.33	−0.25	0.25	0.75	−0.75	−0.25	−0.45	0.45	1.00	−1.00	−0.45
−0.33	1.00	0.33	−0.33	−1.00	−0.25	0.75	0.25	−0.25	−0.75	−0.45	1.00	0.45	−0.45	−1.00
0.33	−1.00	0.33	1.00	−0.33	0.25	−0.75	0.25	0.75	−0.25	0.45	−1.00	0.45	1.00	−0.45
0.33	−0.33	1.00	0.33	−1.00	0.25	−0.25	0.75	0.25	−0.75	0.45	−0.45	1.00	0.45	−1.00
0.33	0.33	−1.00	−0.33	1.00	0.25	0.25	−0.75	−0.25	0.75	0.45	0.45	−1.00	−0.45	1.00
0.33	1.00	−0.33	−1.00	0.33	0.25	0.75	−0.25	−0.75	0.25	0.45	1.00	−0.45	−1.00	0.45
1.00	−1.00	1.00	−0.33	0.33	0.75	−0.75	0.75	−0.25	0.25	1.00	−1.00	1.00	−0.45	0.45
1.00	−0.33	0.33	−1.00	1.00	0.75	−0.25	0.25	−0.75	0.75	1.00	−0.45	0.45	−1.00	1.00
1.00	0.33	−0.33	1.00	−1.00	0.75	0.25	−0.25	0.75	−0.75	1.00	0.45	−0.45	1.00	−1.00
1.00	1.00	−1.00	0.33	−0.33	0.75	0.75	−0.75	0.25	−0.25	1.00	1.00	−1.00	0.45	−0.45

Denote the corresponding design matrices of the five initial designs above by $\boldsymbol{L}^E$, $\boldsymbol{L}^S$, $\boldsymbol{L}^D$, $\boldsymbol{U}$, $\boldsymbol{U}^R$, respectively. We in this thesis consider seven composite designs in 32 runs, denoted as $\begin{bmatrix}\boldsymbol{L}^S\\ \boldsymbol{L}^D\end{bmatrix}$, $\begin{bmatrix}\boldsymbol{U}\\ \boldsymbol{L}^D\end{bmatrix}$, $\begin{bmatrix}\boldsymbol{U}^R\\ \boldsymbol{L}^D\end{bmatrix}$, $\begin{bmatrix}\boldsymbol{L}^E\\ \boldsymbol{U}^R\end{bmatrix}$, $\begin{bmatrix}\boldsymbol{L}^S\\ \boldsymbol{U}^R\end{bmatrix}$, $\begin{bmatrix}\boldsymbol{U}\\ \boldsymbol{L}^S\end{bmatrix}$ and $\begin{bmatrix}\boldsymbol{L}^E\\ \boldsymbol{U}\end{bmatrix}$. Note that we ignore the composition, $\begin{bmatrix}\boldsymbol{L}^E\\ \boldsymbol{L}^D\end{bmatrix}$, since these two components have identical experimental runs that is pointless in Kriging models for computer experiments without

Table 4.2: $U_{16}(16^5)$, $U_{16}^R(16^5)$ over $[-1,1]^5$ design region

$\boldsymbol{U} = U_{16}(16^5)$					$\boldsymbol{U}^R = U_{16}^R(16^5)$				
−1.0000	0.2000	0.6000	−0.3333	0.6000	−0.9375	0.1875	0.5625	−0.3125	0.5625
−0.8667	−0.4667	−0.7333	0.7333	−0.0667	−0.8125	−0.4375	−0.6875	0.6875	−0.0625
−0.7333	0.8667	0.0667	0.2000	−0.8667	−0.6875	0.8125	0.0625	0.1875	−0.8125
−0.6000	−1.0000	−0.3333	−0.2000	0.2000	−0.5625	−0.9375	−0.3125	−0.1875	0.1875
−0.4667	−0.0667	0.8667	−0.7333	−0.4667	−0.4375	−0.0625	0.8125	−0.6875	−0.4375
−0.3333	−0.2000	0.2000	1.0000	0.8667	−0.3125	−0.1875	0.1875	0.9375	0.8125
−0.2000	0.4667	−0.8667	−1.0000	−0.3333	−0.1875	0.4375	−0.8125	−0.9375	−0.3125
−0.0667	−0.8667	0.4667	0.4667	−0.7333	−0.0625	−0.8125	0.4375	0.4375	−0.6875
0.0667	0.7333	−0.4667	−0.0667	1.0000	0.0625	0.6875	−0.4375	−0.0625	0.9375
0.2000	0.6000	1.0000	0.6000	0.3333	0.1875	0.5625	0.9375	0.5625	0.3125
0.3333	−0.3333	−0.6000	−0.4667	−1.0000	0.3125	−0.3125	−0.5625	−0.4375	−0.9375
0.4667	−0.7333	−0.0667	−0.8667	0.7333	0.4375	−0.6875	−0.0625	−0.8125	0.6875
0.6000	0.3333	−0.2000	0.8667	−0.6000	0.5625	0.3125	−0.1875	0.8125	−0.5625
0.7333	1.0000	0.3333	−0.6000	0.0667	0.6875	0.9375	0.3125	−0.5625	0.0625
0.8667	0.0667	−1.0000	0.3333	0.4667	0.8125	0.0625	−0.9375	0.3125	0.4375
1.0000	−0.6000	0.7333	0.0667	−0.2000	0.9375	−0.5625	0.6875	0.0625	−0.1875

error term in our study. Moreover, we do not consider two possible composite designs, $\begin{bmatrix}\boldsymbol{U}\\ \boldsymbol{U}^R\end{bmatrix}$ and $\begin{bmatrix}\boldsymbol{L}^E\\ \boldsymbol{L}^S\end{bmatrix}$ as these two pairs are composed of the same kind of design method, where one is the uniform design and the other is the orthogonal design. In this thesis, we aim to obtain a composite design guaranteeing both effectiveness and robustness. Therefore, the possible composition proposed is supposed to be consist of two different design methods. For presentation convenience, we use $\{\boldsymbol{L}_{16}^S, \boldsymbol{U}_{16}\}$ for $\begin{bmatrix}\boldsymbol{L}^S\\ \boldsymbol{U}\end{bmatrix}$ and similarly for other composite designs.

4.2 Case Studies

In this section, under various cases, we compare the performance of predicting $M = 1000$ untried points for the seven composite designs through the mean squared error

(MSE), defined by

$$MSE = \frac{1}{M}\sum_{m=1}^{M}(y_m - \hat{y}_m)^2.$$

The modeling technique we adopt is Kriging modeling. Fang et al. [18] gave an elaborate introduction to Kriging models.

Definition 13. *Suppose that* $\boldsymbol{x}_i,\ i = 1, \ldots, n$ *are design points over an* s*-dimensional experimental domain, and* $y_i = y(\boldsymbol{x}_i)$ *is the corresponding output. The universal Gaussian Kriging model is defined as*

$$y(\boldsymbol{x}) = \sum_{j=0}^{L}\beta_j B_j(\boldsymbol{x}) + z(\boldsymbol{x}),$$

where the set of B_j *is a chosen polynomial basis over the design region and* $z(\boldsymbol{x})$ *is a Gaussian process with zero mean, variance* σ^2*, and a pre-specified correlation function. The ordinary Kriging model is the most commonly used in practice, defined as*

$$y(\boldsymbol{x}) = \mu + z(\boldsymbol{x}),$$

where μ *is the overall mean of* y*.*

The DACE packakge in MATLAB is employed for Kriging modeling. There are several candidate bases in this package, such as *Poly0* representing the ordinary Kriging models, *Poly1* representing the first-order polynomial function as basis, and *Poly2* representing the second-order polynomial function as basis.

As for the selected models in this paper, we refer to five models mentioned in Fang et al. [18], and Fang et al. [26], that are often used to evaluate performance of optimization algorithm, such as six-hump camelback function, Rosenbrock function, and wood function. Besides, we also consider a laser cutting process model refer to Hung et al. [33]. We present the prediction MSE when only adopting the five initial designs respectively as references. Moreover, in some cases, Kriging model with second-order polynomials, denoted as $Poly2$, can not be applied since the number of runs is not enough leading to an underdetermined least squares problem.

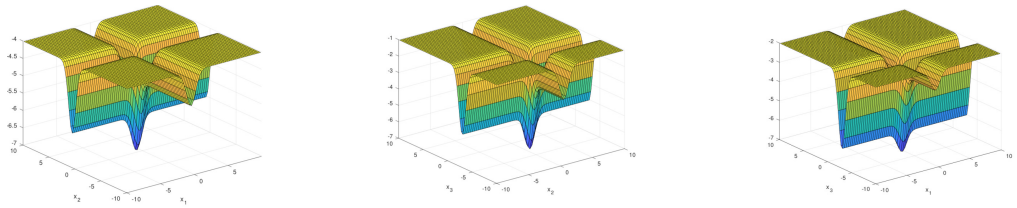
There is no best combination among these seven possible composite designs through theoretic justification. However, from the results of our case studies, $\begin{bmatrix}\boldsymbol{U}^R\\ \boldsymbol{L}^D\end{bmatrix}$ and $\begin{bmatrix}\boldsymbol{U}\\ \boldsymbol{L}^S\end{bmatrix}$ are recommended based on our experience.

4.2.1 f_1 Model

Consider a computer experiment where there are three factors x_1, x_2, x_3 under the model of f_1 as

$$Y = -[e^{-(x_1+0.5)^2} + 2e^{-(x_2-0.5)^2} + 4e^{-(x_3+3)^2}],\ (x_1, x_2, x_3) \in [-2, 2]^2 \times [-1, 4],$$

where a global minimum of y is -7 at $(-0.5, 0.5, -3)$. We first set $x_3 = -3$, and present the 3-dimensional function surface over $x_1, x_2 \in [-10, 10]$. Similarly, we do the same way when fixing $x_1 = -0.5$ or $x_2 = 0.5$ respectively as shown in Figure 4.1. Obviously, if the initial optimization point is at the flat region of the surface shown in Figure 4.1, the minimum is difficult to achieve.



(a) Set $x_3 = -3$ (b) Set $x_1 = -0.5$ (c) Set $x_2 = 0.5$

Figure 4.1: 3D-plot of f_1

Table 4.3: The MSEs of 12 Kriging models in f_1

	$\boldsymbol{L}_{16}^{E}(4^3)$	$\boldsymbol{L}_{16}^{S}(4^3)$	$\boldsymbol{U}_{16}(16^3)$	$\boldsymbol{U}_{16}^{R}(16^3)$	$\boldsymbol{L}_{16}^{D}(4^3)$
Poly0	0.52632410	0.16913826	0.16213913	0.18208480	0.66606326
Poly1	0.50383495	0.42006052	0.11146057	0.13817748	0.62484610
Poly2	0.30706285	0.22536301	0.07277907	0.07623050	0.35369135

		$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$	
Poly0	-	0.04643003	0.03128684	0.03184959	-
Poly1	-	0.04488037	0.03104374	0.03076403	-
Poly2	-	0.04097664	0.02613034	0.02519156	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$		
Poly0	0.03863799	0.01881534	0.04200471	-	-
Poly1	0.03597246	0.04041984	0.03699726	-	-
Poly2	0.02578189	0.04280454	0.03563434	-	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}\}$				
Poly0	0.03827427	-	-	-	-
Poly1	0.03620194	-	-	-	-
Poly2	0.06929245	-	-	-	-

Table 4.3 shows the MSE performance of the corresponding 12 designs. We observe some conclusions:

(1) composite design has a significant advantage in Kriging models;

(2) the optimal composite designs are $\{\boldsymbol{L}^S, \boldsymbol{U}^R\}$, $\{\boldsymbol{U}^R, \boldsymbol{L}^D\}$ and $\{\boldsymbol{U}, \boldsymbol{L}^D\}$, but they are under different basis in Kriging models, where the best one is ordinary Kriging model and the latter two are in second-order Kriging basis. Shrunk design region is preferred in this case since there is little information near the boundary of the region, refer to Figure 4.1;

(3) composition of (shrunk) uniform designs and orthogonal D-optimal designs perform better than other composite designs in Kriging models even though the optimal one is the composition of shrunk orthogonal and uniform designs.

4.2.2 Six-hump Camelback Model

The function with six-hump camelback surface and six local minima is defined by

$$Y = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4, \ (x_1, x_2) \in [-3, 3] \times [-2, 2].$$

The global minimum is $Y^* = -1.03$, when $\boldsymbol{x}^* = [0.09, -0.71]$ or $[-0.09, 0.71]$. Figure 4.2 presents the surface of the six-hump Camelback function on $[-2, 2]\times[-1, 1]$. Table 4.4 gives the MSE performance of the 12 designs under f_2. The optimal design is derived by $\{\boldsymbol{U}^R, \boldsymbol{L}^D\}$, and $\{\boldsymbol{U}, \boldsymbol{L}^D\}$ is the second best.

Table 4.4: The MSEs of 12 Kriging models in f_2

	$\boldsymbol{L}_{16}^E(4^2)$	$\boldsymbol{L}_{16}^S(4^2)$	$\boldsymbol{U}_{16}(16^2)$	$\boldsymbol{U}_{16}^R(16^2)$	$\boldsymbol{L}_{16}^D(4^2)$
Poly0	1308.54470745	469.68964896	346.67121439	324.37707334	781.10670842
Poly1	1308.54470745	469.68964896	346.67121439	324.37707334	781.10670842
Poly2	873.80331034	442.26173123	273.42307887	204.60434167	706.02124138

		$\{\boldsymbol{L}_{16}^S, \boldsymbol{L}_{16}^D\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^D\}$	$\{\boldsymbol{U}_{16}^R, \boldsymbol{L}_{16}^D\}$	
Poly0	-	152.34308825	85.32753948	59.02353737	-
Poly1	-	152.34308825	85.32753948	59.02353737	-
Poly2	-	219.52160123	374.04492573	498.77585792	-
	$\{\boldsymbol{L}_{16}^E, \boldsymbol{U}_{16}^R\}$	$\{\boldsymbol{L}_{16}^S, \boldsymbol{U}_{16}^R\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^S\}$		
Poly0	274.09543833	242.57021945	341.62918231	-	-
Poly1	311.16780784	242.57021945	341.62918231	-	-
Poly2	311.16780784	114.55463698	93.86530241	-	-
	$\{\boldsymbol{L}_{16}^E, \boldsymbol{U}_{16}\}$				
Poly0	102.92632727	-	-	-	-
Poly1	102.92632727	-	-	-	-
Poly2	367.28811737	-	-	-	-

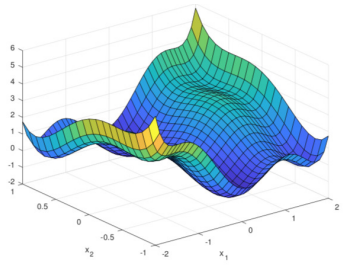


Figure 4.2: The surface of six-hump camelback function

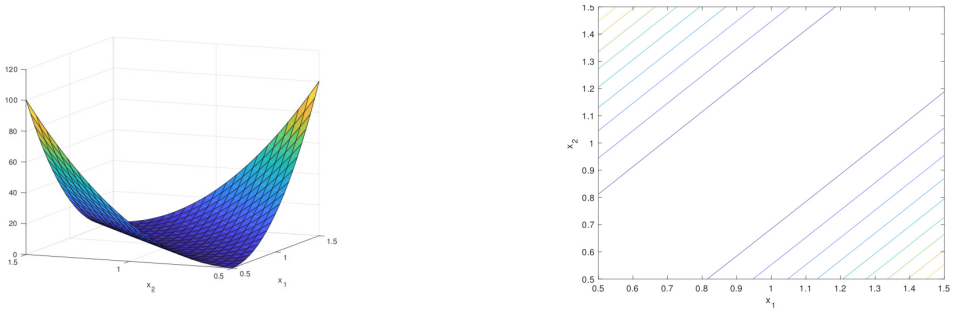
4.2.3 Rosenbrock Model

Minimizing Rosenbrock's function f_3 is a notoriously difficult optimization problem for many algorithms, see Figure 4.3. As a non-convex function, Rosenbrock function is usually used as a performance test problem for optimization algorithms. The function is defined by

$$Y = 100(x_2 - x_1^2)^2 + (x_1 - 1)^2, \ (x_1, x_2) \in [-2, 2]^2.$$

The function is minimized at the point $\boldsymbol{x} = [1, 1]$, with minimum value 0.

The MSE of the 12 designs in f_3 are listed in Table 4.5. In Kriging models, a second-order basis reduces the MSE in prediction to zero. This should be overfitting, and we ignore Poly2 modeling. Besides that, the composition of shrunk uniform design and orthogonal D-optimal design, $\{\boldsymbol{U}^R, \boldsymbol{L}^D\}$, achieves the optimal MSE, and composite designs have great improvement than the initial designs through Kriging models.



(a) (b)

Figure 4.3: Surface and Contour of f_3

Table 4.5: The MSEs of 12 Kriging models in f_3

	$\boldsymbol{L}_{16}^{E}(4^2)$	$\boldsymbol{L}_{16}^{S}(4^2)$	$\boldsymbol{U}_{16}(16^2)$	$\boldsymbol{U}_{16}^{R}(16^2)$	$\boldsymbol{L}_{16}^{D}(4^2)$
Poly0	247.43129957	789.47499329	79.54610450	114.20747101	125.43030890
Poly1	247.43322373	788.23008543	79.51763760	114.09981641	125.40336835

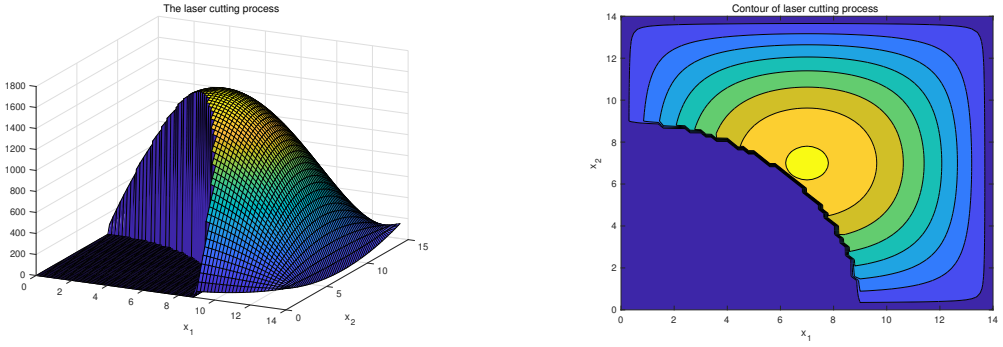
		$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$	
Poly0	-	1.32969713	0.48989703	0.21944395	-
Poly1	-	1.32969713	0.48986503	0.21946112	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$		
Poly0	4.59718677	38.91501096	44.45932527	-	-
Poly1	4.59613907	38.91501096	44.45932527	-	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}\}$				
Poly0	3.01584817	-	-	-	-
Poly1	3.01509954	-	-	-	-

4.2.4 The Laser Cutting Process

We denote a laser cutting functional relationship as f_4. The laser cutting process is expensive, hence simulation with a surrogate model is commonly adopted. According to Hung et al. [33], while other factors are constant, the cutting width (Y) has the following emulated relationship with the pulse duration (x_1) and the laser power (x_2).

$$Y = \begin{cases} 0.01\{[(x_1-7)^2+(x_2-7)^2-62]^2 + \\ [(x_1-7)^2-0.5(x_1-7)-0.5(x_2-7)-1.5]^2\}+10, & if\ x_1^2+x_2^2>80 \\ 0, & elsewhere \end{cases}$$

for $(x_1, x_2) \in [0, 14]^2$.



(a) f_4 surface (b) f_4 contour

Figure 4.4: Surface and contour of f_4

The surface and contour plots are presented as in Figure 4.4. The area where the cutting width (Y) is zero implies the cutting through does not occur. The composite design $\{\boldsymbol{L}^S, \boldsymbol{L}^D\}$ is the best while $\{\boldsymbol{U}^R, \boldsymbol{L}^D\}$ is good as well, see Table 4.6. $\boldsymbol{L}^D$ plays an important role in this case.

Table 4.6: The $\sqrt{MSE}$s of 12 Kriging models in f_4

	$\boldsymbol{L}_{16}^{E}(4^2)$	$\boldsymbol{L}_{16}^{S}(4^2)$	$\boldsymbol{U}_{16}(16^2)$	$\boldsymbol{U}_{16}^{R}(16^2)$	$\boldsymbol{L}_{16}^{D}(4^2)$
Poly0	4.64347707	10.59300108	6.37217202	7.93754519	11.53195776
Poly1	3.89101944	6.20167077	5.31049521	5.78195298	3.56806826
Poly2	7.12546621	43.05636375	23.28175276	30.03484548	5.06876139

		$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$	
Poly0	-	5.34290789	5.15102626	5.29337299	-
Poly1	-	0.45458354	1.60631438	1.03052923	-
Poly2	-	3.37798293	4.85382755	4.43833994	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$		
Poly0	6.06349404	8.79985224	8.26915358	-	-
Poly1	1.64606602	6.31114698	5.01408445	-	-
Poly2	7.91217652	32.07818472	24.10698829	-	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}\}$				
Poly0	11.00490204	-	-	-	-
Poly1	1.66807352	-	-	-	-
Poly2	7.95011362	-	-	-	-

4.2.5 Wood Model

A wood function is notorious due to its difficulty on optimization, defined by

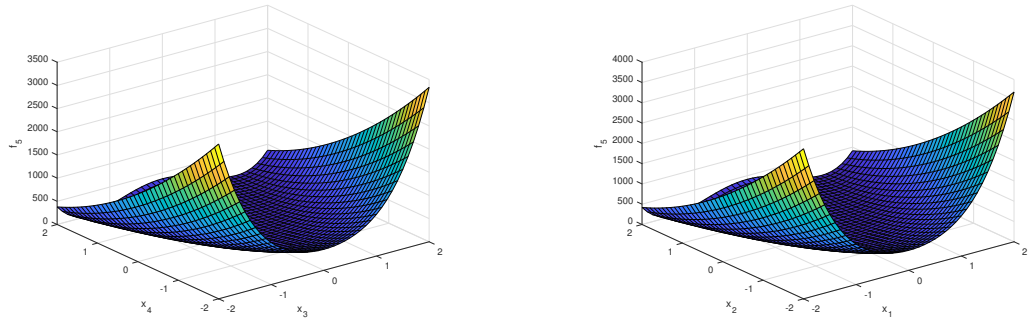
$$\begin{aligned} Y =& 100(x_1^2 - x_2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 \\ &+ 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1), \end{aligned}$$

for $(x_1, x_2, x_3, x_4) \in [-2, 2]^4$. The global minimum is at $x^* = [1, 1, 1, 1]$ with $Y^* = 0$. The surface of the wood model with some fixed input values is displayed as Figure 4.5. Table 4.7 gives the MSE performance of the 12 designs under model f_5. Similarly, due to the large scale of MSE values, we present $\sqrt{MSE}$ values instead. The optimal design is $\{\boldsymbol{L}^E, \boldsymbol{U}^R\}$ in this case. Orthogonal designs are preferred in composition, which is much different from other cases. Uniform designs are always a part of optimal composition.

Table 4.7: The $\sqrt{MSE}$s of 12 Kriging models in f_5

	$\boldsymbol{L}_{16}^{E}(4^4)$	$\boldsymbol{L}_{16}^{S}(4^4)$	$\boldsymbol{U}_{16}(16^4)$	$\boldsymbol{U}_{16}^{R}(16^4)$	$\boldsymbol{L}_{16}^{D}(4^4)$
Poly0	1220.74355273	854.85500018	893.36428973	891.25423550	1344.06395026
Poly1	1220.74355273	1250.16558142	816.82162592	801.13144541	1560.30810885

		$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$	
Poly0	-	585.80050706	687.98078044	568.20911516	-
Poly1	-	565.68943134	1062.06884542	1010.60389630	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$		
Poly0	577.05086085	659.19461620	603.48296368	-	-
Poly1	469.80246602	603.49835078	545.60205753	-	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}\}$				
Poly0	621.09686158	-	-	-	-
Poly1	1022.76859957	-	-	-	-



(a) Fix $x_1 = x_2 = 1$ (b) Fix $x_3 = x_4 = 1$

Figure 4.5: Surface of f_5

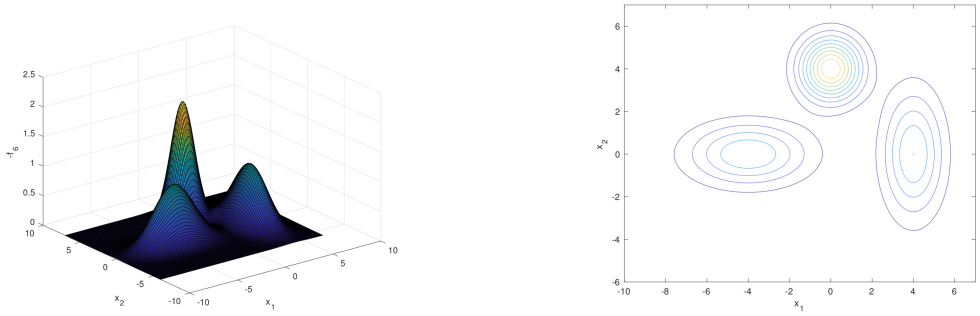
4.2.6 Multi-local Minima Model

The underlying model of f_6 is defined as

$$Y = -(2e^{-0.5(x_1^2+(x_2-4)^2)} + e^{-0.5((x_1-4)^2+\frac{x_2^2}{4})} + e^{-0.5(\frac{(x_1+4)^2}{4}+x_2^2)})$$

for $(x_1, x_2) \in [-10, 7]\times[-6, 7]$ with three local minima. The global minimum achieved from MATLAB built-in functions is at $x^* = [0, 4]$ with $f_6^* = -2$. Figure 4.6 shows

the surface and contour plots of $-Y$ for the convenience of observation. There are three local minima in Y such that the optimization is easy to get stuck.



(a) Surface of $-f_6$ (b) Contour of Y under $-f_6$

Figure 4.6: Surface and Contour of $-f_6$

Table 4.8 gives the MSE performance of the 12 designs in f_6. The optimal design is $\{\boldsymbol{U}, \boldsymbol{L}^S\}$. In this case, the shrunk orthogonal design is preferred and uniform designs are robust. Thus, orthogonal D-optimal designs have no advantage in composition candidates. We can tell from Table 4.8 that in composite designs, $\{\boldsymbol{U}, \boldsymbol{L}^S\}$ and $\{\boldsymbol{L}^S, \boldsymbol{U}^R\}$ perform better than others. Since the information on boundary is too little, orthogonal D-optimal composition are not recommended.

In our study, orthogonal D-optimal design $\boldsymbol{L}^D$ usually plays a significant role in optimal composite designs. There are certain patterns through the above case studies. The following recommendation may help the experimenter quickly select an appropriate composite design.

In most cases, such as f_2, f_3 and f_4, the composition of uniform designs and orthogonal D-optimal designs is recommended. However, for the cases where information on the boundary is little, such as f_1 and f_6, shrunk designs are advantaged in composition, but the composition of uniform designs and orthogonal D-optimal designs still keep competitiveness among others, sometimes with second-best performance. Liang et al. [41] has indicated that there is no best design method for all kinds of models. For instance, if the underlying model is known, D-optimal design performs the best in modeling, while it becomes worse when the underlying model assumption is not correct. Thus, to recommend composite designs based on our study, we choose two having the greatest possibility to perform well. From our study,

Table 4.8: The MSEs of 12 Kriging models in f_6

	$\boldsymbol{L}_{16}^{E}(4^2)$	$\boldsymbol{L}_{16}^{S}(4^2)$	$\boldsymbol{U}_{16}(16^2)$	$\boldsymbol{U}_{16}^{R}(16^2)$	$\boldsymbol{L}_{16}^{D}(4^2)$
Poly0	0.08020388	0.04636583	0.06535556	0.06630720	0.08894426
Poly1	0.07978386	0.06088272	0.06579181	0.06671173	0.08844147
Poly2	0.06748180	0.05010082	0.05085463	0.05433258	0.07836891

		$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{D}\}$	$\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$	
Poly0	-	0.05579098	0.05685204	0.05622184	-
Poly1	-	0.05502589	0.05692332	0.05627756	-
Poly2	-	0.04630765	0.05656794	0.05151787	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{L}_{16}^{S}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$		
Poly0	0.04930164	0.03121931	0.02844730	-	-
Poly1	0.04934869	0.03135780	0.03176334	-	-
Poly2	0.04971465	0.03101581	0.02943805	-	-
	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}\}$				
Poly0	0.05178223	-	-	-	-
Poly1	0.05189215	-	-	-	-
Poly2	0.04871994	-	-	-	-

$\{\boldsymbol{U}^{R}, \boldsymbol{L}^{D}\}$ and $\{\boldsymbol{U}, \boldsymbol{L}^{S}\}$ are recommended for general cases. On the other hand, not all composite designs are valuable, such as $\{\boldsymbol{L}^{E}, \boldsymbol{U}\}$.

To draw fair conclusions, Table 4.9 further provides the prediction MSE in f_5 model of 32-run orthogonal design, $\boldsymbol{L}_{32}^{E}(4^4)$, and 32-run orthogonal D-optimal design, $\boldsymbol{L}_{32}^{D}(4^4)$, and 32-run uniform design, $\boldsymbol{U}_{32}(32^4)$, respectively.

We select three composite designs from Table 4.7, among which $\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$ and $\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$ are the recommended ones according to the most case studies in our research, while $\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$ has the best performance among all Kriging models in f_5. The best composite design in Table 4.9 is $\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$ with linear polynomial basis, that is much better than the best one of pure 32-run initial designs. The best $\sqrt{MSE}$ of pure 32-run initial designs is 711.9625 derived by $\boldsymbol{U}_{32}(32^4)$ with linear polynomial basis. With either basis in Kriging model, $\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$ and $\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$ perform better

than the best record of pure initial designs. Even though with linear basis, $\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$ performs worse, it still has better performance in the ordinary Kriging model, compared to $\boldsymbol{U}_{32}(32^4)$ with linear polynomial basis. Hence, in generally speaking, it shows that the composite designs perform better in prediction when having the same number of experimental runs, compared to the three pure initial design methods, orthogonal designs, orthogonal $D-$optimal designs, and uniform designs.

Table 4.9: The $\sqrt{MSE}$s of 6 Kriging models in f_5 through 32-run initial designs and composite designs

	$\boldsymbol{L}_{32}^{E}(4^4)$	$\boldsymbol{L}_{32}^{D}(4^4)$	$\boldsymbol{U}_{32}(32^4)$	$\{\boldsymbol{L}_{16}^{E}, \boldsymbol{U}_{16}^{R}\}$	$\{\boldsymbol{U}_{16}, \boldsymbol{L}_{16}^{S}\}$	$\{\boldsymbol{U}_{16}^{R}, \boldsymbol{L}_{16}^{D}\}$
Poly0	1344.0143	759.4650	855.8775	577.0509	603.4830	568.2091
Poly1	1334.3019	1191.6372	711.9652	469.8025	545.6021	1010.6039

From the perspective of modeling technique, Kriging model with first-order polynomial basis performs the best, and it is also the second-best in other cases. Ordinary Kriging model may give a surprised performance on prediction in some cases. Kriging model with second-order polynomial basis is limited in our cases since the number of runs is small and usually leading to an underdetermined least squares problem. Moreover, overfitting is also a challenge for second-order basis.

In this chapter, three traditional experimental design methods are combined in order to guarantee both effectiveness and robustness with improved modeling performance. In most existing literature, the design methods require that the levels of factors are discrete. The resulting experimental domain is restricted within a lattice domain so that the resulting designs may lose information when the practical factors of interest are continuous. The construction of uniform designs are requested to be extended to continuous experimental domain. The next chapter propose a construction method for deriving a continuous uniform design.

Chapter 5

Construction of Uniform Designs over Continuous Domain

5.1 Construction of U-type Uniform Designs

There are many construction approaches for $\boldsymbol{U}_n(q^s)$, such as the good lattice point method, the combinatorial construction method, cutting method, and the foldover technique. Stochastic optimization algorithm is also feasible for this problem. More details can refer to Fang et al. [19]. Proposed by Kirkpatrick [38] through the emulation of the physical annealing process in solids, simulated annealing (SA) algorithm is the most widely used stochastic optimization problem. However, since SA algorithm involves too many parameters for the user choosing, its process convergence rate is slow. As a variation of SA, threshold accepting (TA) algorithm, proposed by Dueck and Scheuer [11] has better performance than SA since it provides quicker convergence rate and better results during the same computational time. Many multivariate or combinatorial optimization problems, including some NP-complete or NP-hard problems, have been tackled through TA, such as portfolio optimization problem (Dueck and Winker [12]), NP-complete problem of optimal aggregation (Chipman and Winker [9]) and NP-hard traveling salesman problem (Winker [55]).

Referred to the definition of a uniform design in (1.10), generating a UD is a large scale integer programming problem with multimodal objective function (*i.e.* discrepancies) over a discrete, large and complicated domain (*i.e.* experimental re-

gion). Winker and Fang [56] and Winker and Fang [57] firstly applied the TA to the calculation of the star discrepancy and construction of uniform designs on the lattice domain. Fang et al. [17] suggested some way to improve TA performance. Most uniform designs recorded on the website `https://dst.uic.edu.cn/isci/Uniform/Uniform_Design_Tables.htm` have been obtained by TA algorithm.

A Review to Threshold Accepting Algorithm

TA algorithm starts with an arbitrarily generated U-type design, say $\boldsymbol{U}^0(n, q^s)$, and then it performs a large number of iterations. During each iteration, a new solution is generated and used for substituting the current solution, say $\boldsymbol{U}^c(n, q^s)$, while in the first iteration $\boldsymbol{U}^c(n, q^s) = \boldsymbol{U}^0(n, q^s)$. The new solution $\boldsymbol{U}^{new}(n, q^s)$ is derived from a so-called *neighborhood* of the current solution $\boldsymbol{U}^c(n, q^s)$, which is often defined as a small permutation of the current solution. A decision rule in each iteration is to determine whether to accept the new solution and update the current objective function value or not. We present the flow chart of TA algorithm for generating a U-type UD in Figure 5.1. In our study, we always choose $q = n$.

In uniform design construction, the purpose is to minimize the discrepancy. Hence, in each iteration, the objective discrepancy values of the new candidate and the current solution are compared through $\Delta D = D(\boldsymbol{U}^c(n, q^s)) - D(\boldsymbol{U}^{new}(n, q^s))$. A trivial local search algorithm accepts $\boldsymbol{U}^{new}(n, q^s)$ if and only if $\Delta D \geq 0$. In this algorithm, the resulting solution has a large probability that gets stuck in a bad local optimum especially as the objective function is multimodal. Winker and Fang [57] indicated that in the application to the traveling salesman problem, the result of the trivial local search algorithm got stuck in a bad level and could not be improved by further increasing the number of iterations. Therefore, TA adopts another acceptance criterion that allows certain temporary worsening with respect to a predetermined decreasing threshold sequence, say $T \geq 0$. That implies in each iteration, a new candidate is accepted if and only if $\Delta D \leq T$. Thus TA is also regarded as the refined local search algorithm.

The threshold value is not a constant but a finite sequence that is positive and decreasing to zero. As the threshold value is nonzero in most of the iterations, even

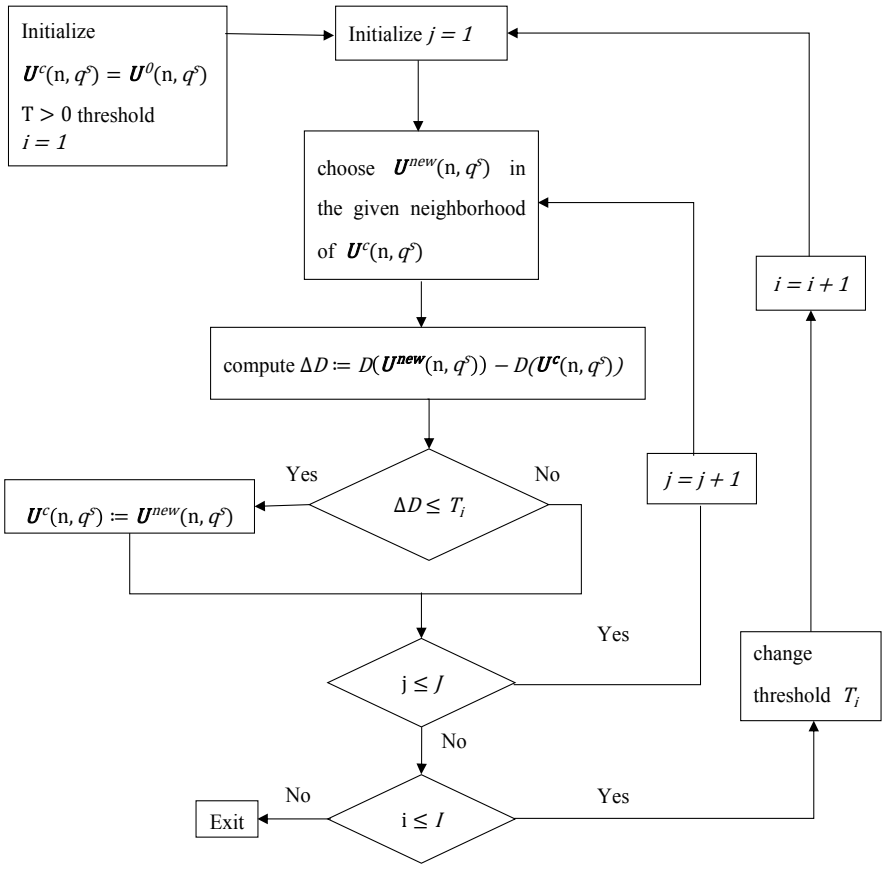


Figure 5.1: Threshold Accepting Algorithm for Generating Uniform Designs

if the current solution is stuck in a bad local optimum, it has chance to leave. At the very end, the threshold value ends up at zero, that ensures the solution of TA algorithm is still a local optimum. However, the local optimum of TA has high quality and at least is approaching to the global optimum.

The key of TA algorithm is to define the neighborhood and threshold sequence. A U-type design is regarded as a neighborhood of the current design if they differ by elements in at most $k \leq s$ columns. Many authors set $k = 1$, and define each neighborhood by randomly exchanging two elements in a randomly chosen column of the current design (so-called TA neighborhood). The definition of the neighborhoods provides an endogenous data-driven approach to generating a threshold sequence. In our case, we adopt a recursive formula

$$T_i = \frac{I - i}{I} \times T_{i-1}$$

to derive the threshold sequence in an exponentially decreasing manner, refer to Fang et al. [17]. With the given parameters $\alpha \in (0,1)$, I and J, and a discrepancy $\mathcal{D}(\cdot)$, the procedure of generating a threshold sequence is as follows:

1. generate the initial design of the TA algorithm, $\boldsymbol{U}^0(n, q^s)$;
2. find J neighborhoods of the initial design, $\boldsymbol{U}^j(n, q^s)$, $j = 1, \ldots, J$;
3. calculate the discrepancy values for each neighborhood, $\mathcal{D}(\boldsymbol{U}^j(n, q^s))$;
4. compute the range of the discrepancy values of the J neighborhoods, $range(\mathcal{D}(\boldsymbol{U}^j(n, q^s)))$;
5. set threshold basis $T_0 = range(\mathcal{D}(\boldsymbol{U}^j(n, q^s))) \times \alpha$ and $T_1 = \frac{I-1}{I} \times T_0$;
6. generate the rest threshold sequence through $T_i = \frac{I-i}{I} \times T_{i-1}$, $i = 2, \ldots, I$.

However, the definition of the TA neighborhood is for the convenience of computation. It is different from the mathematical neighborhood with respect to the discrepancy function. Hence, this definition of neighborhood is not feasible for continuous uniform designs. In the next section, we introduce coordinate descent methods for further searching a UD over continuous domain starting from a U-type UD derived by TA algorithm.

5.2 Coordinate Descent Algorithms

Computer simulations play a dominant role in modern applications. Traditional computer experimental designs are in a lattice domain, in which the factors of interest are limited with discrete levels in experiments. However, with the improvement of computational ability, factors in practical may refer to a continuous domain, that leads to a great loss of information if the experimenter still choose a design from a lattice domain. The uniform designs over continuous domain are suitable for this case, and easily to be carried out. They should have better uniformity than U-type UDs as well, and as consequence, perform better in modeling. Hence, constructing uniform designs over continuous domain is in badly need. The joint work with Dr. Lai Jianfa, Prof. Fang Kai-Tai and Dr. Peng Xiaoling, Lai et al. [39], is to propose

an approach to obtaining the UDs over continuous domain, especially in a rectangle. Finding a uniform design in a continuous $\mathcal{C}^s = [0,1]^s$ is a high dimensional non-convex optimization problem. Here we only consider the case in which $n = q$ in Definition 3. For a point set $\mathcal{P} = \{\boldsymbol{x}_1 \ldots, \boldsymbol{x}_n\} = (x_{ik})_{n\times s}$ from $\mathcal{C}^s = [0,1]^s$, the objective is to minimize any given measure of uniformity, say $\mathcal{D}$, such as CD in (1.7) or MD in (1.8):

$$X = \arg\min_{\mathcal{P}} \mathcal{D}(\mathcal{P}). \tag{5.1}$$

Gradient descent (Ruder [49]) is a good way to solve such problem with a suitable initialization. However, it costs too much for large sized designs (i.e. n and s are large), which are prevalent in computer experiments. Coordinate descent (Wright [58]) is another feasible alternative approach. Coordinate descent algorithms are a class of large-scale optimization algorithms that are feasible for various objective functions including non-smooth and non-convex ones. They have been successfully implemented on the problems in various modern applications such as machine learning and large-scale computational statistics. For some structured problems, coordinate descent algorithms have shown advantage on computation speed and convergence proof over traditional algorithms, such as gradient descent (GD), refer to Shi et al. [51]. A comprehensive introduction with applications to coordinate descent variants has been elaborated by Wright [58] and Shi et al. [51]. In our study, three coordinate descent algorithms are introduced and discussed on their performance in constructing UDs.

In traditional gradient optimization algorithm, there are two possible actions in each iteration. Coordinate descent is a simple but efficient non-gradient optimization algorithm. The gradient optimization algorithm seeks for the minimum value of the function along the direction of the most rapid descent. However, coordinate descent minimizes the value of the objective function along the coordinate gradient in each step.

Since coordinate descent methods are local optimization algorithms, a good enough initialization is required to be a guarantee of the uniformity. Otherwise, the local search process is easily stuck at a bad local optima. Hence, we adopt the U-type UD derived through TA algorithm as the initialization of the following coordinate descent process. Moreover, the existing U-type UD yields from a U-type lattice domain, in

which the one-dimensional uniformity is guaranteed. Hence, to avoid breaking the uniformity in one-dimension of the new resulting UD over continuous domain, in each update of the current solution, each column of the new design matrix will be checked to satisfy the U-type requirement defined in Definition 4. The new UD over the continuous experimental region derived by such procedure, is close to the initial U-type UD, referred to Figures 5.2 and 5.3.

Coordinate Gradient Descent (CGD)

To construct a UD over continuous domain, denoted as $U_n((0,1)^s)$ in order to distinguish from U-type UD, $U_n(q^s)$, the optimization problem can be regarded as an s-dimensional problem with n points on $[0,1]^s$. An $n \times s$ uniform design matrix can be viewed as n design points each having s dimension. Coordinate gradient descent is to adjust the position of each design point in each dimension so that the whole design point set can be more uniformly distributed in the domain. Figure 5.2 visualizes the design points during the optimization process. Figure 5.3 presents the design points of $U_9((0,1)^4)$ from any two-dimensional projection points. The visualization figures may help us comprehend the objective uniformity measure CD adopted in our study. Figures 5.2 and 5.3 indicate that the whole design points become centrally aggregated after coordinate descent. The following algorithm is the basic coordinate gradient descent method.

Algorithm 1 Coordinate Gradient Descent(CGD)

Given n, s, and a discrepancy $\mathcal{D}(\cdot)$
Initialization X^0 and choose ϵ, the step size δ and the maximum epoch M.
Create discrepancy list and matrix list
for each $m \in [0, M]$ **do**
    for each $i \in [0, n]$ **do**
        for each $j \in [0, s]$ **do**
            Calculate the coordinate gradient, $g = \frac{d\mathcal{D}(X^m)}{dx_{ij}^m}$
            $\hat{x_{ij}} = x_{ij}^m + g * \delta$
            Calculate $d = \mathcal{D}(\hat{X})$ and put into discrepancy list
            Put $\hat{X}$ into matrix list
        end for
    end for
    Find the matrix with minimum discrepancy in the matrix list, say X^{m+1}
    if the corresponding discrepancy value is smaller than the one of X^m, and X^{m+1} satisfies U-type requirement **then** X^{m+1} is returned
    else break the for-loop and return X^m
end for

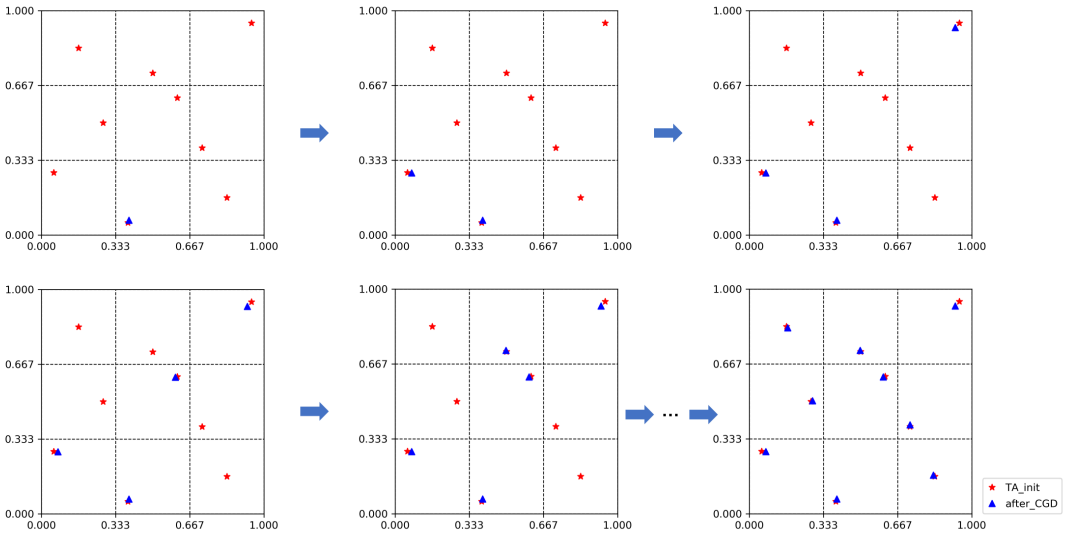


Figure 5.2: Visualization of design points during the optimization process

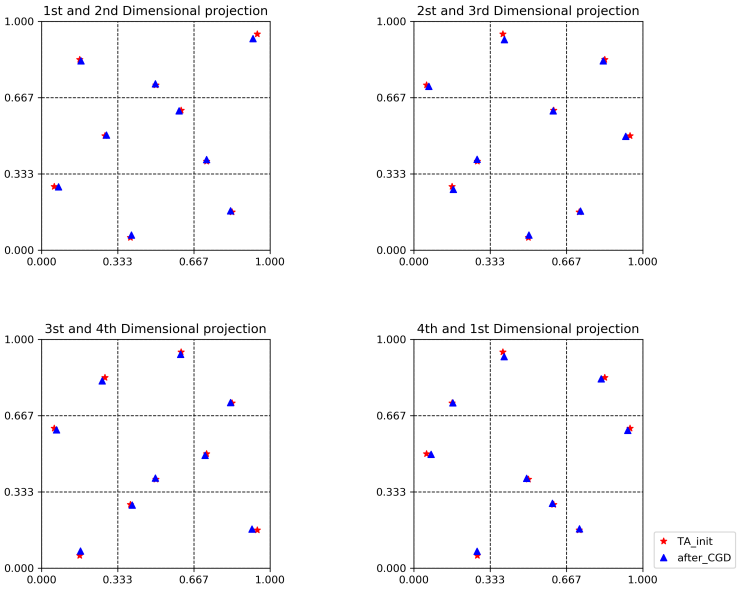


Figure 5.3: Design points from any two-dimensional projection points

Coordinate Descent with Fixed Step Size (CDFSS)

Considering the limit of minimum unit of the variables in practical experiments, coordinate descent with fixed step size are suitable under such circumstances. The idea of this method is to replace the coordinate gradient with $\{-1, 1\}$ and the step size can be set as the minimum unit of the variable of interest. In each step, the coordinate with the rapidest descent will be chosen. The algorithm will be stopped until there is no further descent in any coordinate.

This method can be easily applied in the experiments with minimum unit due to the equipment or technology limitation. For example, one of the variable, temperature, in the industry experiment is from 0 to T. Due to the limitation of temperature detection and control technology, the minimum unit is 0.1 degree. With CDFSS, the step size of the variable can be set as $\frac{0.1}{T-0}$. Similarly for other variables, the step size can be set different values according to the corresponding limitation. Thus, the resulting uniform design is feasible under practical experimental requirements.

Algorithm 2 Coordinate Descent with Fixed Step Size

Given n, s, and discrepancy $\mathcal{D}(\cdot)$
Initialization X^0 and choose the step size in each column $\{\delta_j, j \leq s\}$ and the maximum epoch M.
Create discrepancy list and matrix list
for each $m \in [0, M)$ **do**
    for each $i \in [0, n)$ **do**
        for each $j \in [0, s)]$ **do**
            for each $a \in \{-1, 1\}$ **do**
                $\hat{X}$ be the copy of X^m
                $\hat{x}_{ij} = \hat{x}_{ij} + a * \delta_j$
                Calculate $d = \mathcal{D}(\hat{X})$ and put into discrepancy list
                Put $\hat{X}$ into matrix list
            end for
        end for
    end for
    Find the matrix with minimum discrepancy in the matrix list, say X^{m+1}
    if the corresponding discrepancy value is smaller than the one of X^m, and X^{m+1} satisfies U-type requirement **then** X^{m+1} is returned
    else break the for-loop and return X^m
end for

Coordinate Zero-gradient (CZG)

Under CD, the coordinate gradient in Algorithm 1, $g = \frac{dCD^2(X^m)}{dx_{ij}^m}$, is known, given as follows, and we can set $g = 0$ to find the current coordinate value while fixing others.

$$\begin{aligned}
\frac{dCD^2(\mathcal{P})}{dx_{ij}} &= -\frac{2}{n}\left\{\prod_{q\neq j}^{s}\left[1+\frac{1}{2}|x_{iq}-\frac{1}{2}|-\frac{1}{2}(x_{iq}-\frac{1}{2})^2\right]\right\} \\
&\times\left[\frac{1}{2}sgn(x_{ij}-\frac{1}{2})-(x_{ij}-\frac{1}{2})\right] \\
&+\frac{1}{n^2}\prod_{q\neq j}^{s}\frac{1}{2}(1+|x_{iq}-\frac{1}{2}|)sgn(x_{ij}-\frac{1}{2}) \\
&+\frac{2}{n^2}\sum_{k\neq i}^{n}\left\{\prod_{q\neq j}^{s}\left[1+\frac{1}{2}|x_{iq}-\frac{1}{2}|+\frac{1}{2}|x_{kq}-\frac{1}{2}|-\frac{1}{2}|x_{iq}-x_{kq}|\right]\right. \\
&\left.\times\left[\frac{1}{2}sgn(x_{ij}-\frac{1}{2})-\frac{1}{2}sgn(x_{ij}-x_{kj})\right]\right\},
\end{aligned} \tag{5.2}$$

where the sgn function is:

$$sgn(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases} \tag{5.3}$$

The coordinate value, x_{ij}, can be derived through taking $\frac{d\mathcal{D}(\mathcal{P})}{dx_{ij}} = 0$. However, some of x_{ij} in sgn function can not be solved through explicit expression. Here we adopt recursion with an initialization, denoted as x_{ij}^0, and a maximum number of iterations, T, while other coordinate values $\{x_{kq}, k \neq i, q \neq j\}$, are fixed. The recursive formula is given as

$$x_{ij}^t = \frac{A}{B} + \frac{1}{2}, \tag{5.4}$$

where

$$A = -\frac{2}{n}\left\{\prod_{q\neq j}^{s}\left[1+\frac{1}{2}|x_{iq}-\frac{1}{2}|-\frac{1}{2}(x_{iq}-\frac{1}{2})^2\right]\right\}\times\frac{1}{2}sgn(x_{ij}^{t-1}-\frac{1}{2})$$
$$+\frac{1}{n^2}\prod_{q\neq j}^{s}\frac{1}{2}(1+|x_{iq}-\frac{1}{2}|)sgn(x_{ij}^{t-1}-\frac{1}{2})$$
$$+\frac{2}{n^2}\sum_{k\neq i}^{n}\left\{\prod_{q\neq j}^{s}\left[1+\frac{1}{2}|x_{iq}-\frac{1}{2}|+\frac{1}{2}|x_{kq}-\frac{1}{2}|-\frac{1}{2}|x_{iq}-x_{kq}|\right]\right.$$
$$\left.\times\left[\frac{1}{2}sgn(x_{ij}^{t-1}-\frac{1}{2})-\frac{1}{2}sgn(x_{ij}^{t-1}-x_{kj})\right]\right\} \tag{5.5}$$

and

$$B = -\frac{2}{n}\prod_{q\neq j}^{s}\left[1+\frac{1}{2}|x_{iq}-\frac{1}{2}|-\frac{1}{2}(x_{iq}-\frac{1}{2})^2\right] \tag{5.6}$$

for $t \in [0, T]$. Through plenty of experiments, $T = 1$ is enough for getting good results.

Algorithm 3 Coordinate zero-gradient

Given n, s, and a discrepancy $\mathcal{D}(\cdot)$
Initialization X^0 and choose ϵ and the maximum epoch M.
Create discrepancy list and matrix list
for each $m \in [0, M]$ **do**
    for each $i \in [0, n]$ **do**
        for each $j \in [0, s]$ **do**
            Let $\{x_{kl}^m,\ k \neq i,\ l \neq j\}$ be fixed.
            Calculate $\hat{x}_{ij}^m$ such that the coordinate gradient, $\frac{d\mathcal{D}(\mathcal{P})}{d\hat{x}_{ij}^m} = 0$.
            if $\hat{X}^m$ satisfies U-type requirement **then** $X^{m+1} = \hat{X}^m$
            else break all the for-loop and return X^m
        end for
    end for
    if $|\mathcal{D}(X^{m+1}) - \mathcal{D}(X^m)| < \epsilon$, then break the for-loop
end for

In our experiments, the coordinate zero-gradient method is advantaged in computation speed. The comparisons of the convergence rate among the three algorithms mentioned in our case are given in Figure 5.4. For the same optimization task, it

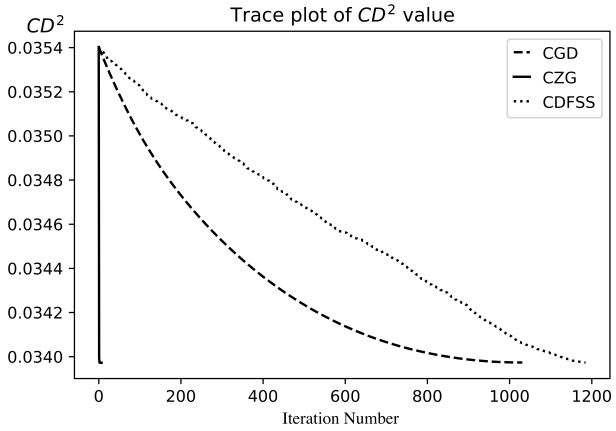


Figure 5.4: CD^2 trace plot of three algorithms

takes several minutes through coordinate gradient descent method but several seconds through coordinate zero-gradient method. However, even though the final optimization result is good enough, the CD values during the optimization process is unstable. For instance, sometimes there may be violent fluctuations of the CD values. Furthermore, since MD overcome the limitation of CD due to dimensionality, when the number of columns is large, MD is recommended as the measure of uniformity. We also provide the coordinate gradient function with respect to MD as follows.

$$\begin{aligned}
\frac{dMD^2(\mathcal{P})}{dx_{ij}} &= -\frac{2}{n}\left\{\prod_{q\neq j}^{s}\left[\frac{5}{3}-\frac{1}{4}|x_{iq}-\frac{1}{2}|-\frac{1}{4}(x_{iq}-\frac{1}{2})^2\right]\right\} \\
&\times\left[-\frac{1}{4}sgn(x_{ij}-\frac{1}{2})-\frac{1}{2}(x_{ij}-\frac{1}{2})\right]-\frac{1}{n^2}\prod_{q\neq j}^{s}(\frac{15}{8}-\frac{1}{2}|x_{iq}-\frac{1}{2}|)\times\frac{1}{2}sgn(x_{ij}-\frac{1}{2}) \\
&+\frac{2}{n^2}\sum_{k\neq i}^{n}\left\{\prod_{q\neq j}^{s}\left[\frac{15}{8}-\frac{1}{4}|x_{iq}-\frac{1}{2}|-\frac{1}{4}|x_{kq}-\frac{1}{2}|-\frac{3}{4}|x_{iq}-x_{kq}|+\frac{1}{2}(x_{iq}-x_{kq})^2\right]\right. \\
&\left.\times\left[-\frac{1}{4}sgn(x_{ij}-\frac{1}{2})-\frac{3}{4}sgn(x_{ij}-x_{kj})+(x_{ij}-x_{kj})\right]\right\}. \qquad (5.7)
\end{aligned}$$

For the three coordinate descent variants, they are all capable of obtaining continuous UD. Coordinate zero-gradient is the fastest algorithm among them. As for the coordinate descent with fixed step size, it does not need the coordinate gradient. Similar to gradient descent, coordinate gradient descent is stable in each step, and in addition, adjusts the position of each design point in each dimension. Compared with gradient descent algorithm, coordinate descent algorithms are simpler and cheaper.

Moreover, each step in coordinate descent is meaningful, which can be visualized during the optimization process. The visualization is helpful for comprehending uniformity and the optimization process. Furthermore, it is convenient for the new UDs to be generalized into practical implementation on the experiments with required minimum unit of factors.

5.3 Modeling Performance of Uniform Designs over Continuous Domain

Table 5.1 presents the CD values of two existing U-type UDs derived by TA and the corresponding new UDs over a continuous domain through TA+CGD. Compared to the U-type UD, the new UD over continuous domain is capable of improving the uniformity with a great amount, and as a result, performing better in modeling. The new UD under CD criterion derived by TA+CGD, $U_{18}((0,1)^7)$, is listed in Table 5.2.

Table 5.1: CD values of uniform designs through TA and TA+CGD

Algorithms	$U_{18}(18^7)$ / $U_{18}((0,1)^7)$	$U_{27}(27^{13})$ / $U_{27}((0,1)^{13})$
TA	0.035403	0.228455
TA + CGD	0.033972	0.198073

To further explore the performance of the new uniform designs, we implement them on several computer experiments. We use Kriging modeling technique to obtain an approximate model for the responses and factors, and evaluate the prediction mean squared errors (MSE) when predicting 1000 untried points through the approximate models trained by the new uniform designs, compared with the U-type UD recorded on the website, and the Latin hypercube sampling.

Since Latin hypercube sampling (LHS) has been popular in computer experiments, we also provide the prediction MSE of the Kriging models trained by LHS and midpoint LHS (MLHS), defined in Fang et al. [18]. Since each LHS or MLHS is a random sample, we generate 100 LHS or MLHS each time and conduct Kriging modeling respectively. Subsequently, we present the average, minimum and maximum

Table 5.2: $U_{18}((0,1)^7)$ under CD obtained by TA+CGD

0.9080	0.5238	0.6906	0.8001	0.7509	0.3578	0.9071
0.3050	0.5838	0.6351	0.0531	0.6986	0.7447	0.1393
0.7475	0.2567	0.4738	0.2078	0.9052	0.8544	0.6415
0.9457	0.1462	0.3609	0.4796	0.3579	0.6288	0.0542
0.4704	0.6925	0.5757	0.9079	0.1977	0.2049	0.0972
0.1438	0.4809	0.0946	0.2525	0.0549	0.5210	0.2979
0.5264	0.0986	0.8501	0.1441	0.1533	0.4260	0.7990
0.3570	0.0529	0.1464	0.5804	0.8045	0.2527	0.4264
0.0957	0.7482	0.9468	0.5135	0.8528	0.5862	0.5889
0.6366	0.4197	0.0528	0.9471	0.5825	0.6931	0.7531
0.0546	0.3076	0.4149	0.6973	0.3062	0.0959	0.6922
0.2009	0.1964	0.7458	0.8505	0.4748	0.9032	0.3564
0.7971	0.8605	0.8011	0.6355	0.0960	0.7947	0.4694
0.4159	0.6462	0.2524	0.4225	0.2570	0.9456	0.9456
0.5711	0.9058	0.2994	0.7430	0.9456	0.4756	0.2040
0.2435	0.9512	0.5211	0.3061	0.5330	0.3026	0.8508
0.6876	0.3689	0.9043	0.3633	0.6321	0.0539	0.2502
0.8534	0.7923	0.1928	0.0974	0.4286	0.1500	0.5260

prediction MSE of each Kriging model trained by these 100 LHS or MLHS, denoted as Avg., Min., and Max., in Tables 5.4 and 5.5.

Wood Model

We have introduced the wood model in Chapter 4. We apply several new uniform designs over continuous domain, denoted as $U_n((0,1)^s)$, and the corresponding recorded ones, $U_n(n^s)$ on the website respectively into the wood model, in which 9-run and 16-run designs are shown in Table 5.3. Besides that, we also implement 18-run and 27-run designs, LHS, and MLHS. We use the above mentioned designs with *Poly0* and *Poly1* bases to yield Kriging models. For convenience of reading, $\sqrt{MSE}$ are calculated and listed in Table 5.4. It is obvious that new uniform designs over continuous domain perform better than the recorded ones, and the average of

Table 5.3: $U_9(9^4)$, $U_9((0,1)^4)$, $U_{16}(16^4)$, and $U_{16}((0,1)^4)$ in the wood model

$U_9(9^4)$				$U_9(9^4) \in [-2,2]^4$				$U_9((0,1)^4) \in [0,1]^4$				$U_9((0,1)^4) \in [-2,2]^4$			
4	1	7	5	-0.5	-2	1	0	0.3942	0.0690	0.7114	0.5978	-0.4232	-1.7242	0.8457	0.3912
1	3	4	3	-2	-1	-0.5	-1	0.0743	0.2762	0.3966	0.2761	-1.7030	-0.8953	-0.4136	-0.8957
9	9	5	4	2	2	0	-0.5	0.9303	0.7215	0.6060	0.3940	1.7214	0.8861	0.4240	-0.4240
6	6	6	9	0.5	0.5	0.5	2	0.6086	0.5114	0.4910	0.7172	0.4344	0.0456	-0.0360	0.8689
5	7	2	1	0	1	-1.5	-2	0.5000	0.9207	0.2751	0.0793	0.0000	1.6830	-0.8997	-1.6830
2	8	8	7	-1.5	1.5	1.5	1	0.1758	0.8242	0.8242	0.8242	-1.2969	1.2969	1.2969	1.2969
3	5	1	6	-1	0	-2	0.5	0.2837	0.6065	0.0658	0.5000	-0.8653	0.4260	-1.7370	0.0000
8	2	3	8	1.5	-1.5	-1	1.5	0.8208	0.1875	0.1792	0.9065	1.2833	-1.2501	-1.2833	1.6262
7	4	9	2	1	-0.5	2	-1.5	0.7127	0.3945	0.9214	0.1705	0.8509	-0.4220	1.6858	-1.3181
$U_{16}(16^4)$				$U_{16}(16^4) \in [-2,2]^4$				$U_{16}((0,1)^4) \in [0,1]^4$				$U_{16}((0,1)^4) \in [-2,2]^4$			
1	10	4	6	-2.00	0.40	-1.20	-0.67	0.4678	0.6593	0.4717	0.7195	-0.1290	0.6370	-0.1134	0.8778
2	4	13	15	-1.73	-1.20	1.20	1.73	0.2873	0.3469	0.0440	0.7769	-0.8510	-0.6126	-1.8242	1.1074
3	13	10	10	-1.47	1.20	0.40	0.40	0.1599	0.7847	0.5913	0.5975	-1.3606	1.1386	0.3650	0.3898
4	8	7	1	-1.20	-0.13	-0.40	-2.00	0.0415	0.5965	0.2213	0.3425	-1.8342	0.3858	-1.1150	-0.6302
5	6	1	2	-0.93	-0.67	-2.00	-1.73	0.7808	0.0927	0.3474	0.6514	1.1230	-1.6294	-0.6106	0.6054
6	15	15	4	-0.67	1.73	1.73	-1.20	0.2165	0.4656	0.4057	0.0389	-1.1342	-0.1378	-0.3774	-1.8446
7	1	11	7	-0.40	-2.00	0.67	-0.40	0.7252	0.7091	0.9508	0.0981	0.9006	0.8362	1.8030	-1.6078
8	16	8	14	-0.13	2.00	-0.13	1.47	0.3397	0.9060	0.7747	0.2254	-0.6414	1.6238	1.0986	-1.0986
9	3	3	3	0.13	-1.47	-1.47	-1.47	0.8519	0.5300	0.7172	0.8348	1.4074	0.1198	0.8686	1.3390
10	7	16	9	0.40	-0.40	2.00	0.13	0.6437	0.9514	0.2860	0.9514	0.5746	1.8054	-0.8562	1.8054
11	11	5	16	0.67	0.67	-0.93	2.00	0.4103	0.0373	0.6566	0.4060	-0.3590	-1.8510	0.6262	-0.3762
12	12	12	2	0.93	0.93	0.93	-1.73	0.5944	0.4045	0.8392	0.5308	0.3774	-0.3822	1.3566	0.1230
13	2	6	11	1.20	-1.73	-0.67	0.67	0.9013	0.8316	0.0971	0.4668	1.6050	1.3262	-1.6118	-0.1330
14	14	2	8	1.47	1.47	-1.73	-0.13	0.9600	0.2836	0.5329	0.2779	1.8398	-0.8658	0.1314	-0.8886
15	9	14	13	1.73	0.13	1.47	1.20	0.5341	0.2096	0.1563	0.1535	0.1362	-1.1618	-1.3750	-1.3862
16	5	9	5	2.00	-0.93	0.13	-0.93	0.1032	0.1681	0.8969	0.8969	-1.5874	-1.3278	1.5874	1.5874

LHS/MLHS designs in the wood model.

Six-hump Camelback Model

The definition of the six-hump camelback model is introduced in Chapter 4. Table 5.5 presents the prediction MSE of the Kriging models yielding from the 16-run and 27-run designs, LHS and MLHS with *Poly0*, *Poly1* and *Poly2* bases respectively. The new uniform design, $U_{16}((0,1)^2)$, performs better in the six-hump camelback model as well. However, $U_{27}((0,1)^2)$ performs worse than the U-type UD. The practical performance depends on the underlying model. In the comparison of 27-run designs, the new design with better uniformity performs worse. Nevertheless, the demand of designs having continuous factors is also important.

Table 5.4: The $\sqrt{MSE}$s of Kriging models in the wood model

	$U_9(9^4)$	$U_9((0,1)^4)$	Avg. LHS	Min. LHS	Max. LHS	Avg. $MLHS$	Min. $MLHS$	Max. $MLHS$
Poly0	957.9618	826.0637	801.4158	644.5038	958.0141	794.9072	612.7375	910.9978
Poly1	878.1982	695.0498	943.1197	632.9920	1993.1489	914.5818	629.1848	1911.4974
	$U_{16}(16^4)$	$U_{16}((0,1)^4)$	Avg. LHS	Min. LHS	Max. LHS	Avg. $MLHS$	Min. $MLHS$	Max. $MLHS$
Poly0	829.4668	743.3496	723.1160	504.7169	870.5672	719.2164	542.2015	903.0956
Poly1	779.9872	722.6574	733.4497	556.4965	1005.5512	740.2478	551.4009	999.5673
	$U_{18}(18^4)$	$U_{18}((0,1)^4)$	Avg. LHS	Min. LHS	Max. LHS	Avg. $MLHS$	Min. $MLHS$	Max. $MLHS$
Poly0	691.7101	668.6842	716.2950	540.4776	871.4179	711.2033	498.6695	871.2249
Poly1	618.7809	587.3990	725.4645	474.1841	1242.4868	713.8811	516.4029	1032.2576
	$U_{27}(27^4)$	$U_{27}((0,1)^4)$	Avg. LHS	Min. LHS	Max. LHS	Avg. $MLHS$	Min. $MLHS$	Max. $MLHS$
Poly0	2768.7	543.0675	608.0075	431.3616	878.0648	600.6795	423.5607	861.2940
Poly1	1743.9	534.7528	591.2014	401.7304	819.2450	579.7672	400.7702	810.9931

Table 5.5: The MSEs of Kriging models in the six-hump camelback model

	$U_{16}(16^2)$	$U_{16}((0,1)^2)$	Avg. LHS	Min. LHS	Max. LHS	Avg. $MLHS$	Min. $MLHS$	Max. $MLHS$
Poly0	316.0265	210.8945	418.4559	192.2220	745.3818	420.7049	228.3404	918.7642
Poly1	316.0265	210.8945	473.7306	213.0605	1005.8727	461.3111	240.3622	953.4489
Poly2	253.8087	182.3029	252.4274	74.8913	654.5391	242.9823	111.0450	481.7922
	$U_{27}(27^2)$	$U_{27}((0,1)^2)$	Avg. LHS	Min. LHS	Max. LHS	Avg. $MLHS$	Min. $MLHS$	Max. $MLHS$
Poly0	159.2230	178.3153	221.2963	84.5307	541.6582	210.1022	95.7539	767.8275
Poly1	159.7014	178.9219	238.4213	78.5837	574.5531	224.9911	96.9844	905.1754
Poly2	78.9778	93.0508	145.4119	51.6424	381.6309	142.6485	63.7796	288.7547

Chapter 6

Concluding Remarks

My thesis incorporate some of my new developments of experimental designs during the doctor study period. It focuses on three experimental design methods, orthogonal designs, D-optimal regression designs, and uniform designs. Chapter 1 introduced a new criterion, the entropy based on projection HDP, $S(\mathcal{D}^k)$, in order to fast detect combinatorially non-isomorphic orthogonal designs. $S(\mathcal{D}^k)$ proposes a new perspective for evaluating experimental designs. Various criteria are based on HDP as reviewed. $S(\mathcal{D}^k)$ is derived from HDP and hence its validity is ensured. It also incorporates the evaluation of projection designs, that are valuable for the experimenter in the case of allocating main effects to the columns of a given design. Furthermore, $S(\mathcal{D}^k)$ is not limited for saturated or regular designs. The related definition or application of $S(\mathcal{D}^k)$ can be easily extended to other cases. However, there may exist some non-isomorphic classes that cannot be successfully distinguished. Only adopting one criterion may not be able to derive a perfect result. Therefore, Chapter 2 proposes a two-stage detection algorithm, in which projection ASC distribution is adopted as an amendment.

Uniformity measure evaluates the uniform geometric structure of design points, while minimum aberration (GWLP) presents the ability of designs in statistical inference. In traditional view, combinatorially isomorphic orthogonal designs are considered to perform the same in statistical modeling. However, the MECP is able to provide much more information than GWLP about confounding among the main effects and interaction effects. Hence, combinatorially isomorphic designs having the

same GWLP may differ in MECP in many cases. Chapter 3 further distinguishes the combinatorial isomorphism into combinatorial and geometric equivalence. The MECP can be easily extended to any regular or non-regular saturated orthogonal designs with multiple levels. However, there may be some difficulty for non-saturated ODs since some aliasing terms cannot be estimated. Nevertheless, MECP may give the experimenter more specific information about the design, such as the number of aliasing interactions for certain factors. With projection MECP, the best k-factor projection designs can be obtained in order to provide the best arrangement of main effects for the experimenter. Therefore, when the experimenter has more background information, the effectiveness of the design may rapidly increase if it has the optimal MECP.

The MECP is defined for only saturated orthogonal designs. The three widely known traditional design methods, orthogonal design, orthogonal D-optimal design and uniform design, have advantages and limitations. Chapter 4 proposes a natural idea that a composite design combining the three design methods ought to guarantee both effectiveness and robustness. In our study, orthogonal D-optimal design usually plays a significant role in optimal composite designs. With the same number of runs, compared to adopting only one design method, the composite design is more effective and robust in prediction. We believe that the new composite design can be applied to other multi-level designs and more modeling techniques as well as other physical experiments.

Another research gap is that most existing design methods take discrete levels. If the experimenter is interested in continuous factors but takes designs on lattice domain, the experiments may lose a lot of information. Uniform designs are requested to be extended to continuous domain. Hence, Chapter 5 adopts three variants of coordinate descent method with a U-type UD derived by TA algorithm as the initialization, in order to construct UDs over continuous domain. It turns out that the discrepancy values of UDs over continuous domain have been reduced considerably. Subsequently, to evaluate the new UDs over continuous domain, we compare the modeling performance of the new designs with the recorded UDs on the website, as well as Latin hypercube sampling in two computer experiments as illustrative instances.

The prediction MSE is adopted as evaluation criteria. In our study, we take Kriging modeling technique to test the chosen metamodels, that may affect the prediction MSE as well. Nevertheless, it indicates that the new uniform designs are advantaged on modeling in computer experiments. Uniform designs over continuous domain are promising since designs with continuous factors become feasible and in demand with the improvement of computational ability.

Bibliography

[1] Atkinson, A. C., Bogacka, B., and Bogacki, M. B. (1998), D- and T-optimum designs for the kinetics of a reversible chemical reaction, *Chemometrics and Intelligent Laboratory Systems*, **43**, 185-198.

[2] Atkinson, A. C. and Donev, A. N. (1992), *Optimum Experimental Designs*, Oxford Science Publications, Oxford.

[3] Booth, K. H., and Cox, D. R. (1962). Some systematic supersaturated designs. *Technometrics*, **4**(4), 489-495.

[4] Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters*. Wiley, New York.

[5] Chan, L.Y., Fang, K.T. and Mukerjee, R. (2001). A characterization for orthogonal arrays of strength two via a regression model, *Statistics & Probability Letters*, **54**, 189-192.

[6] Chan, L.Y., Fang, K.T. and Winker, P. (1998), An equivalence theorem for orthogonality and D-optimality, Technical Report MATH-186, Hong Kong Baptist University.

[7] Cheng, C. S. (1980), Orthogonal arrays with variable numbers of symbols. *The Annals of Statistics*, **8**(2), 447-453.

[8] Cheng, S. W., Ye, K. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *The Annals of Statistics*, **32**(5), 2168-2185.

[9] Chipman, J. S. and Winker, P. (1995). Optimal industrial classification by threshold accepting. *Control and Cybernetics*, **24**:477–494.

[10] Clark, J. B., and Dean, A. M. (2001). Equivalence of fractional factorial designs. *Statistica Sinica*, **11**(2), 537-547.

[11] Dueck, G. and Scheuer, T. (1990). Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing. *Journal of Computational Physics*, **90**(1):161–175.

[12] Dueck, G. and Winker, P. (1992). New concepts and algorithms for portfolio choice. *Applied Stochastic Models and Data Analysis*, **8**(3):159–178.

[13] Elsawah, A. M., Fang, K. T., and Ke, X. (2021). New recommended designs for screening either qualitative or quantitative factors. *Statistical Papers*, **62**(1), 267-307.

[14] Evangelaras, H., Koukouvinos, C., Dean, A. M., & Dingus, C. A. (2005). Projection properties of certain three level orthogonal arrays. *Metrika*, **62**(2), 241-257.

[15] Evangelaras, H., Koukouvinos, C., and Lappas, E. (2007). 18-run nonisomorphic three level orthogonal arrays. *Metrika*, **66**(1), 31-37.

[16] Fang, K. T. (1980), The uniform design: application of number-theoretic methods in experimental design. *Acta Mathematicae Applicatae Sinica*, **3**(4), 363-372.

[17] Fang, K. T., Ke, X., and Elsawah, A. M. (2017). Construction of uniform designs via an adjusted threshold accepting algorithm. *Journal of Complexity*, **43**:28–37.

[18] Fang, K. T., Li, R. and Sudjianto, A. (2005), *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC Press, London.

[19] Fang, K. T., Liu, M. Q., Qin, H. and Zhou, Y. D. (2018), *Theory and Application of Uniform Experimental Designs*, Science Press and Springer, Singapore.

[20] Fang K. T., Ma C. X. (2000), The uniformity – a useful criterion in experimental design, First Midwest Conference for New Directions in Experimental Design, May 2000, Columbus, Ohio.

[21] Fang K. T., Ma C. X. (2001), Orthogonal and Uniform Experimental Designs (in Chinese), *Science Press*, 175-183.

[22] Fang, K. T., Ma, C. X. and Li, J. K. (1999), Recent development of orthogonal factorial designs and their applications (III) - D-optimality of orthogonal designs, *Journal of Applied Statistics and Management*, **19**(4), 43-52 (in Chinese).

[23] Fang, K. T., and Mukerjee, R. (2000). A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika*, **87**(1), 193-198.

[24] Fang, K. T., Quan, H., and Chen, Q. Y. (1988). *Practical Regression Analysis* (in Chinese). Science Press, 348-349.

[25] Fang K. T., Tang Y., and Yin J. X. (2008), Lower bounds of various criteria in experimental designs, *Journal of Statistical Planning and Inference*, **138**:184-195.

[26] Fang, K. T., Wang, Y. and Bentler, P. M. (1994). Some applications of number-theoretic methods in statistics, *Statistical Science*, **9**, 416-428.

[27] Fang, K. T., Yin, H. and Liang, Y. Z. (2004). New approach by Kriging methods to problems in QSAR. *Journal of Chemical Information and Computer Sciences*, **44**(6), 2106-2113.

[28] Fang, Q., Yeung, H. W., Leung, H. W., and Huie, C. W. (2000), Micelle-mediated extraction and preconcentration of ginsenosides from Chinese herbal medicine. *Journal of Chromatography A*, **904**(1), 47-55.

[29] Fries, A., and Hunter, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics*, **22**(4), 601-608.

[30] Hickernell, F. J. (1998). A generalized discrepancy and quadrature error bound. *Mathematics of Computation of the American Mathematical Society, 67*(221), 299-322.

[31] Hickernell, F. J. and Liu, M. Q. (2002), Uniform designs limit aliasing, *Biometrika*, **89**, 893-904.

[32] Hua, L. K., and Wang, Y. (1981). *Applications of Number Theory to Numerical Analysis*. Springer and Science Press.

[33] Hung, Y. C. , Michailidis, G. , and Lok, P. H. (2020). Locating infinite discontinuities in computer experiments. *Journal on Uncertainty Quantification*, **8**(2), 717-747.

[34] Ji, Y. B., Xu, Q-S, Hu, Y. Z., and Heyden, Y. V. (2005). Development, optimization and validation of a fingerprint of Ginkgo biloba extracts by high-performance liquid chromatography, *Journal of Chromatography A*, **1066**(1-2), 97-104.

[35] Johnson, R. A., and Wichern, D. W. (2014). *Applied Multivariate Statistical Analysis*. Pearson, London, UK.

[36] Ke, X., Fang, K. T., Elsawah, A. M., and Lin, Y. (2020). New non-isomorphic detection methods for orthogonal designs. *Communications in Statistics-Simulation and Computation*, 1-16, DOI: 10.1080/03610918.2020.1844895.

[37] Kiefer, J. (1959), Optimal experimental designs (with discussion). *Journal of the Royal Statistical Society: Series B*, **21**, 272-319.

[38] Kirkpatrick, S. (1984). Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, **34**(5-6):975–986.

[39] Lai, J., Fang, K. T., Peng, X., and Lin, Y. (2021). Construction of uniform designs over continuous domain in computer experiments. *Communications in Statistics-Simulation and Computation*, 1-17, DOI: 10.1080/03610918.2021.2011924.

[40] Lam, C., and Tonchev, V. D. (1996). Classification of affine resolvable $2-(27,9,4)$ designs. *Journal of Statistical Planning and Inference*, **56**(2), 187-202.

[41] Liang, Y. Z., Fang, K. T., and Xu, Q. S. (2001), Uniform design and its applications in chemistry and chemical engineering, *Chemometrics and Intelligent Laboratory Systems*, **58**, 43-57.

[42] Lin, Y., and Fang, K. T. (2019). The main effect confounding pattern for saturated orthogonal designs. *Metrika*, **82**(7), 843-861, DOI: 10.1007/s00184-019-00713-w.

[43] Lin, Y., Tang, Y., Zhang, J., and Fang K. T. (2022). Detecting non-isomorphic orthogonal designs, *Journal of Statistical Planning and Inference*, **221**(2022), 299-312, DOI: 10.1016/j.jspi.2022.05.003.

[44] Lu, H. M., Ni, W. D., Liang, Y. Z., and Man, R. L. (2006). Supercritical CO2 extraction of emodin and physcion from Polygonum cuspidatum and subsequent isolation by semipreparative chromatography, *Journal of Separation Science*, **29**(14), 2136-2142.

[45] Ma, C. X., and Fang, K. T. (2001). A note on generalized aberration in fractional designs. *Metrika*, **53**(1), 85-93.

[46] Ma, C. X, Fang, K. T., and Lin, D. K. J. (2001). On the isomorphism of fractional factorial designs. *Journal of Complexity*, **17**(1), 86-97.

[47] McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245.

[48] Roman, S. (1992). *Coding and Information Theory* (Vol. 134). Springer Science & Business Media.

[49] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

[50] Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, **5**(1), 3-55.

[51] Shi, H. J. M., Tu, S., Xu, Y., and Yin, W. (2016). A primer on coordinate descent algorithms. *arXiv preprint arXiv:1610.00040*.

[52] Tang Y, Xu H. Q. (2014), Permuting regular fractional factorial designs for screening quantitative factors, *Biometrika*, **101**(2):333-350.

[53] Wang, H., Qi, J., Xu, T., Liu, J. H., Qin, M. J., Zhu, D. N., and Yu, B. Y. (2010). Effects of storage condition factors on fungal invasion of Radix Ophiopogonis. *Journal of Agricultural and Food Chemistry*, **58**(9), 5432-5437.

[54] Wang, Y., and Fang, K. T. (1981). A note on uniform distribution and experimental design. *A Monthly Journal of Science, 26*(6), 485-485.

[55] Winker, P. (1994). The tuning of the threshold accepting heuristic for travelling salesman problems. *Sonderforschungsbereich*, **178**. Universität Konstanz.

[56] Winker, P. and Fang, K. T. (1997). Application of threshold-accepting to the evaluation of the discrepancy of a set of points. *SIAM Journal on Numerical Analysis*, **34**(5):2028–2042.

[57] Winker, P. and Fang, K. T. (1998). Optimal u-type designs. *Monte Carlo and Quasi-Monte Carlo Methods 1996*, 436–448. Springer.

[58] Wright, S. J. (2015). Coordinate descent algorithms. *Mathematical Programming*, **151**(1):3–34.

[59] Xu, H., and Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *The Annals of Statistics*, **29**(2), 549-560.

[60] Xu, Q. S., Liang, Y. Z. and Fang, K. T. (2000), The effects of different experimental designs on parameter estimation in the kinetics of a reversible chemical reaction, *Chemometrics and Intelligent Laboratory Systems*, **52**, 155-166.

[61] Xu, Q. S., Xu, Y. D., Li, L., and Fang, K. T. (2018), Uniform experimental design in Chemometrics, *Journal of Chemometrics*, **32**(11), e3020, DOI: 10.1002/cem.3020.

[62] Zhou, Y. D., Fang, K. T., and Ning, J. H. (2013). Mixture discrepancy for quasi-random point sets. *Journal of Complexity, 29*(3), 283-301.

CURRICULUM VITAE

Academic qualification of the thesis author, Ms. LIN Yuxuan:

- Received the degree of Bachelor of Science (Honours) from Beijing Normal University-Hong Kong Baptist University United International College, November 2017.

Date: July 2022