Original Article | Published: 11 October 2021

Connecting U-type Designs Before and After Level Permutations and Expansions

Yaping Wang, Fei Wang, Yabo Yuan & Qian Xiao

Journal of Statistical Theory and Practice **15**, Article number: 81 (2021)

Abstract

In both physical and computer experiments, U-type designs, including Latin hypercube designs, are commonly used. Two major approaches for evaluating U-type designs are orthogonality and space-filling criteria. Level permutations and level expansions are powerful tools for generating good U-type designs under the above criteria in the literature. In this paper, we systematically study the theoretical properties of U-type designs before and after level permutations and expansions. We establish the relationships between the initial designs' generalized word-length patterns (GWLP) and the generated designs' orthogonal and space-filling properties. Our findings generalize the existing results and provide theoretical justifications for the current level permutation and expansion algorithms.

References

1. Bingham D, Sitter RR, Tang B (2009) Orthogonal and nearly orthogonal designs for computer experiments. Biometrika 96:51–65

2. Ba S, Myers WR, Brenneman WA (2015) Optimal sliced Latin hypercube designs. Technometrics

Noname manuscript No.
(will be inserted by the editor)

Connecting U-type designs before and after level permutations and expansions

Yaping Wang · Fei Wang · Yabo Yuan · Qian Xiao

Received: date / Accepted: date

Abstract In both physical and computer experiments, U-type designs, including Latin hypercube designs, are commonly used. Two major approaches to evaluating U-type designs are orthogonality and space-filling criteria. Level permutations and level expansions are powerful tools for generating good U-type designs under the above criteria in the literature. In this paper, we systematically study the theoretical properties of U-type designs before and after level permutations and expansions. We establish the relationships between the initial designs' generalized word-length patterns (GWLP) and the generated designs' orthogonal and space-filling properties. Our findings generalize the existing results and provide theoretical justifications for the current level permutation and expansion algorithms.

Keywords Factorial design · Latin hypercube design · Maximin distance design · Orthogonal array · Uniform design

Mathematics Subject Classification (2010) 62K99 · 62K15

1 Introduction

U-type designs, aka. balanced designs, are commonly used in experiments, which include orthogonal arrays (Hedayat et al., 1999), Latin hypercube designs (McKay et al., 1979) and balanced fractional factorial designs (Tang et al., 2012). They are

Yaping Wang, Yabo Yuan
KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, 200062, China
E-mail: ypwang@fem.ecnu.edu.cn, yaboyuanyabo@163.com

Fei Wang
LMAM, School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing, 100871, China
E-mail: wangfeizz@pku.edu.cn

Qian Xiao (Corresponding Author)
Department of Statistics, University of Georgia, Athens, Georgia, 30602, U.S.A.
E-mail: qx69137@uga.edu

factorial designs where every level appears equally often for each factor. Specifically, Latin hypercube designs (LHDs) have the level sizes equal to the run sizes. LHDs are widely used in modern industrial and computer experiments due to their uniform one-dimensional projection properties (Fang et al., 2006; Bingham et al., 2009; Zhou and Xu, 2014).

Various optimality criteria have been proposed for the optimization and construction of U-type designs, with two major aims: orthogonality and space-filling. Orthogonal designs minimize the associations among factors and space-filling designs minimize the similarities among runs. Specifically, the orthogonality criterion seeks to optimize designs by minimizing the average squared correlations or the maximum absolute correlations between factors (Owen, 1994; Tang, 1998; Ye, 1998). Orthogonal designs whose column-wise correlations are all zero are directly useful in fitting data using main effect linear models, because they allow uncorrelated estimates of linear main effects (Lin et al., 2009; Lin and Tang, 2015). Orthogonality is also theoretically connected with space-filling properties (Wang et al., 2020). Space-filling designs, especially space-filling LHDs, are appealing for computer experiments whose outputs are deterministic (Fang et al., 2006; Santner et al., 2018; Xiao and Xu, 2018). Space-filling U-type designs with multiple levels are also found useful in some pharmaceutical experiments (Xiao and Xu, 2018; Xiao et al., 2019). Space-filling designs are robust to model misspecification and can decrease the bias of fitted models (Gramacy, 2020). To measure designs' space-filling properties, two popular criteria are the maximin distance (Johnson et al., 1990; Morris and Mitchell, 1995) and uniformity (Fang, 1980; Hickernell, 1998). The maximin distance criterion seeks to spread out the design points evenly over the entire design region via maximizing the smallest distance between any pair of points. The uniformity criterion aims to scatter points as uniformly as possible in the design space by minimizing certain discrepancy metric.

Constructing orthogonal or space-filling U-type designs with moderate or large sizes is often challenging (Hedayat et al., 1999; Lin and Tang, 2015; Wang et al., 2018a). Since the candidate spaces of U-type designs grow exponentially fast with the design sizes, searching over the whole spaces can be inefficient for identifying large optimal designs. Thus, theoretical results are often needed to guide the search algorithm only searching over some promising sub-spaces. Motivated by this idea, researchers focused on searching over the candidate U-type designs that can be generated by level permutations and level expansions. Tang et al. (2012), Tang and Xu (2013), and Xu et al. (2014) proposed to use level permutations to construct uniform U-type designs. Zhou and Xu (2014) generalized the level permutation method to construct space-filling U-type designs including uniform and maximin distance designs as special cases. Tang (1993) proposed to generate orthogonal array-based LHDs (OALHDs) by expanding levels in randomized OAs. Jiang and Ai (2017) constructed uniform OALHDs via permuting and expanding the levels of OAs. Xiao and Xu (2018) developed an efficient procedure to generate maximin LHDs as well as maximin multi-level designs from existing orthogonal or nearly orthogonal designs via level permutations and expansions.

Specifically, Xiao and Xu (2018) introduced a general framework of the level permutation and expansion algorithm for constructing good U-type designs. Denote a U-type design with N runs, n factors and s levels by $D(N, s^n)$ where each column takes values from the set $\mathcal{Z}_s = \{0, 1, \ldots, s-1\}$. To generate a good high-level U-type design $\tilde{D}(N, (ms)^n)$ where m is a positive integer satisfying that $m \leq N/s$ and $N/(ms)$ is an integer, Xiao and Xu (2018) proposed to first select a good initial U-type design $D(N, s^n)$, then permute the levels of D to improve its property, and finally expand each of its levels to m distinct levels. This procedure restricts the entire candidate space of $\tilde{D}$ to a much smaller sub-space — the candidate designs that can be generated by the level permutations and expansions of the initial design D. Different initial designs D will lead to different sub-spaces, and the one leading to the best average properties of candidate designs should be chosen. Standard stochastic searching algorithms can be used to find the best design in the selected sub-space. Xiao and Xu (2018) used the threshold accepting algorithm and Jiang and Ai (2017) used the simulated annealing algorithm. Through simulation studies, Jiang and Ai (2017) and Xiao and Xu (2018) showed that the level permutation and expansion method is very efficient for generating good high-level U-type designs. Many new optimal designs were found by these authors.

In this paper, we systematically study the theoretical properties of U-type designs before and after all level permutations and expansions. We focus on the theoretical results on the choices of the initial designs and their connections with the average properties of the generated designs, which provides theoretical supports for the level permutation and expansion algorithms in Jiang and Ai (2017) and Xiao and Xu (2018). The established theorems justify the use of the generalized minimum aberration (GMA) designs as the initial designs in the level permutation and expansion method, where the GMA designs (Xu and Wu, 2001) have the sequentially minimized generalized word-length pattern (GWLP; see Section 2 for details). Specifically, we prove that the orthogonality and space-filling properties of the generated high-level designs on average are determined by the GWLPs of the initial low-level designs. Consequently, starting with a GMA U-type design, the mean squared correlations, the maximin distance metrics and the uniformity discrepancies of all candidate designs generated via all possible level permutations and expansions are minimized on average, i.e. the "mean-best" searching sub-space.

The theoretical results in this paper can be viewed as the generalizations of the results in Zhou and Xu (2014), Jiang and Ai (2017), and Xiao and Xu (2018), where the "mean-best" sub-spaces based on the GMA designs were used. For constructing orthogonal and nearly orthogonal designs, Zhou and Xu (2014) only focused on generating fractional factorial designs via level permutations; while, we prove results for generating LHDs as well as multi-level designs via both level permutations and expansions. For constructing maximin distance designs, Xiao and Xu (2018) only connected the generated designs' expected distance variations with $A_2(D)$, i.e., the second element of the GWLP of the initial design; while, we establish the relationship between the generated designs' expected measures of whole distance structures and the initial design's entire GWLP. Jiang and Ai (2017) only considered uniform LHDs; while, we discuss uniform U-type designs under various discrepancy criteria. Note that our work only contributes to the theory, and it adopts the same algorithms for generating practical designs as those in Jiang and Ai (2017) and Xiao and Xu (2018).

The remainder of this paper is organized as follows. Section 2 introduces some notation and preliminaries. Section 3 shows the theoretical results connecting the U-type designs before and after level permutations and expansions under the orthogonality, maximin distance and uniformity discrepancy criteria. Section 4 summarizes the paper and discusses some possible future work. All proofs are relegated to the Appendix.

2 Notation and preliminaries

A design $D(N, s^n)$ is called an orthogonal array (OA) of strength t, denoted by $OA(N, n, s, t)$, if all possible level-combinations in each t-columns of D appear the same number of times. A U-type design is an $OA(N, n, s, 1)$, where each of the s levels occurs exactly N/s times. In particular, a U-type (N, N^n) design is called an LHD, denoted by $LHD(N, n)$. Throughout the paper, we assume $n \geq 2$.

The generalized minimum aberration (GMA) criterion was proposed by Xu and Wu (2001) for evaluating fractional factorial designs. For a design $D(N, s^n)$, consider the full ANOVA model:

$$Y = X_0\alpha_0 + X_1\alpha_1 + \cdots + X_n\alpha_n + \epsilon,$$

where Y is the vector of N observations, α_0 is the intercept, X_0 is an $N \times 1$ vector of 1's, α_j is the $(s-1)^j\binom{n}{j} \times 1$ vector of all jth-order factorial effects for $j = 1, \ldots, n$, X_j is the $N \times (s-1)^j\binom{n}{j}$ matrix of orthonormal contrast coefficients for α_j and ϵ is the random error term. Xu and Wu (2001) defined

$$A_j(D) = N^{-2}\|X_0^{\mathrm{T}}X_j\|^2 \text{ for } j = 0, \ldots, n,$$

to measure the overall aliasing between all jth-order factorial effects and the intercept, where $\|X\|^2 = tr(X^{\mathrm{T}}X)$. Obviously, $A_0(D) = 1$ for U-type designs by the definition. The generalized word-length pattern (GWLP) of D is defined to be the vector $(A_1(D), A_2(D), \ldots, A_n(D))$. Note that $\|X_0^{\mathrm{T}}X_j\|^2$ is independent of the choice of orthonormal contrasts, thus the value of $A_j(D)$ is independent of the parameterizations of the full ANOVA model (Xu and Wu, 2001). The GMA criterion is to sequentially minimize the GWLP. For two designs D_1 and D_2, D_1 is said to have less aberration than D_2 if there exists a $k \in \{1, 2, \ldots, n\}$, such that $A_k(D_1) < A_k(D_2)$ and $A_i(D_1) = A_i(D_2)$ for $i = 1, \ldots, k-1$. If no other design has less aberration than D_1, it is said to be a GMA design. Xu and Wu (2001) showed that a design $D(N, s^n)$ is an $OA(N, n, s, t)$ if and only if $A_1(D) = \cdots = A_t(D) = 0$.

Next, we present some mathematical background of the level permutation and expansion methods in Jiang and Ai (2017) and Xiao and Xu (2018). For an initial U-type design $D(N, s^n)$, by permuting the s levels of each columns in D, we can obtain a new U-type design of the same size. When all possible level permutations are considered, there are a total of $(s!)^n$ generated designs from D. Denote them by $\mathcal{P}(D)$. Since the level permutation does not change designs' combinatorial structures, all designs in $\mathcal{P}(D)$ have the same GWLP. By expanding the s levels of each column in D to ms levels, we can obtain a new U-type design $\widetilde{D}(N, (ms)^n)$, where

$m \leq N/s$ and both m and $N/(ms)$ are positive integers. In particular, design $\tilde{D}$ is an $LHD(N,n)$ when $m = N/s$. The level expansion procedure is carried out by replacing the N/s entries of level ℓ ($\ell \in \mathcal{Z}_s$) with a random sequence of $N/(ms)$ replicates of $\ell m + i$ ($i = 0, \ldots, m-1$) for each column in the initial design D, which ensures that the expanded (ms)-levels in $\tilde{D}$ take values from $\mathcal{Z}_{ms}$. For example, consider expanding a two-level column $(0,1,0,1,0,1,0,1)^{\mathrm{T}}$ to a random four-level column $(0,2,1,3,0,2,1,3)^{\mathrm{T}}$. Here, the original level 0 is expanded to $N/(ms) = 2$ replicates of levels (0,1) in some random orders, and the original level 1 is expanded to 2 replicates of levels (2,3) in some random orders. From an initial $D(N, s^n)$, when first performing all possible level permutations and then performing all possible level expansions, the total number of generated U-type designs $\tilde{D}(N,(ms)^n)$ (including isomorphic ones) is

$$n_0 = (s!)^n \left(\frac{\frac{N}{s}!}{\left(\frac{N}{ms}!\right)^m} \right)^{sn}. \tag{1}$$

Denote the set of these designs by $\mathcal{EP}(D)$. As a special case, when $m = 1$, $\mathcal{EP}(D) = \mathcal{P}(D)$. To obtain space-filling designs $\tilde{D}$ via level permutations and expansions, Zhou and Xu (2014), Jiang and Ai (2017), and Xiao and Xu (2018) proposed to start from a GMA design D and then adopt stochastic searching algorithms to identify the best generated designs in $\mathcal{EP}(D)$. Next, we will justify this procedure by theories.

3 Theoretical results

In this section, we present some theoretical results connecting the U-type designs before and after level permutations and expansions under the orthogonality, maximin distance and uniformity discrepancy criteria. These results provide justifications for the use of GMA designs as initial designs for the level permutation and expansion method.

3.1 Orthogonal and nearly orthogonal designs

Orthogonality is a widely used design criterion. Specifically, we adopt the average squared correlations to measure designs' orthogonality (Owen, 1994; Joseph and Hung, 2008; Wang et al., 2020). Let $D = (x_{ik})_{N\times n}$ be an (N, s^n) U-type design. The correlation between the jth and kth columns of D ($j \neq k$) is

$$\rho_{jk}(D) = \frac{\sum_{i=1}^{N}\left(x_{ij} - \frac{s-1}{2}\right)\left(x_{ik} - \frac{s-1}{2}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_{ij} - \frac{s-1}{2}\right)^2 \sum_{i=1}^{N}\left(x_{ik} - \frac{s-1}{2}\right)^2}}.$$

The average squared correlation metric for orthogonality is defined as

$$\rho^2(D) = \frac{1}{\binom{n}{2}} \sum_{j<k} \rho_{jk}^2(D).$$

A design D is orthogonal if and only if $\rho^2(D) = 0$. When orthogonal designs do not exist, the nearly orthogonal designs have the minimized $\rho^2(D)$ values.

From a U-type design $D(N, s^n)$, we consider generating the U-type designs $\tilde{D}(N, (ms)^n)$ via level permutations and expansions. When $m = N/s$, designs $\tilde{D}$ are LHDs. Denote $\overline{\rho^2}(\tilde{D})$ as the average $\rho^2(\tilde{D})$ value over $\mathcal{EP}(D)$. Theorem 1 establishes a connection between $\overline{\rho^2}(\tilde{D})$ and the $A_2(D)$ value of its initial design D.

Theorem 1 *Given a U-type initial design $D(N; s^n)$, consider all possible level permutations and expansions. Then we have*

$$\begin{aligned}\overline{\rho^2}(\tilde{D}) &= \frac{1}{n_0}\sum_{\tilde{D}\in\mathcal{EP}(D)} \rho^2(\tilde{D}) \\ &= \alpha(N,s,m) + \frac{144 n_2^4 c_1^2(s,m)(c_2(s,m)-1)^2}{\binom{n}{2} n_1^4 s^4 (s-1)^2 \left(m^2 s^2 - 1\right)^2} A_2(D),\end{aligned}$$

where

$$\begin{aligned}\alpha\left(N,s,m\right) = &\frac{4n_2^2 m^2 (2ms-1)^2}{N {n_1}^2 (ms+1)^2} + \frac{144 n_2^4 c_1^2(s,m)}{N n_1^4 s^2 (s-1)^2 \left(m^2 s^2-1\right)^2} \times \\ &\left[N\left(\frac{c_2(s,m)+s-1}{s}\right)^2 - c_2^2(s,m)\right],\end{aligned}$$

$$c_1(s,m) = \sum_{\substack{l_1,l_2\in\mathcal{Z}_s\\ l_1\neq l_2}} \sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s}\\ x_{jk}\in\mathcal{T}_{l_2,s}}} \tilde{x}_{ik}\tilde{x}_{jk},$$

$$c_2(s,m) = \frac{n_1 n_3}{n_2^2}\frac{s-1}{c_1(s,m)}\sum_{l\in\mathcal{Z}_s}\left(\sum_{\substack{x_{ik}\in\mathcal{T}_{l,s}\\ x_{jk}\in\mathcal{T}_{l,s}}} \tilde{x}_{ik}\tilde{x}_{jk} - \sum_{x_{ik}\in\mathcal{T}_{l,s}} \tilde{x}_{ik}^2\right),$$

$$\tilde{x}_{ik} = x_{ik} - (ms-1)/2 \textit{ for } x_{ik}\in\mathcal{Z}_{ms},$$

$$n_1 = \frac{(N/s)!}{\left([N/(ms)]!\right)^m},$$

$$n_2 = \frac{(N/s-1)!}{[N/(ms)-1]!\left([N/(ms)]!\right)^{m-1}},$$

$$n_3 = \begin{cases}\frac{(N/s-2)!}{[N/(ms)-2]!([N/(ms)]!)^{m-1}} + \frac{(N/s-2)!}{([N/(ms)-1]!)^2([N/(ms)]!)^{m-2}} & \textit{if } N\neq ms,\\ \frac{(N/s-2)!}{([N/(ms)-1]!)^2([N/(ms)]!)^{m-2}} & \textit{if } N = ms,\end{cases}$$

$\mathcal{T}_{l,s} = \{lm, lm+1, \ldots, (l+1)m-1\}$ *and n_0 is defined in Equation (1).*

Theorem 1 shows that the average $\rho^2(\tilde{D})$ value is a linear function of $A_2(D)$. Therefore, starting from an $OA(N, n, s, 2)$ as the initial design in the level permutation and expansion method tends to generate better final design under the orthogonality criterion.

It is worth mentioning that level permutations and expansions for U-type designs have not been studied under the ρ^2 criterion previously. The special case of $m = 1$ of Theorem 1 reduces to a level-permutation only procedure which is considered in Zhou and Xu (2014).

3.2 Maximin distance designs

To construct space-filling designs, the maximin distance criterion proposed by Johnson et al. (1990) aims to maximize the minimum distance among design points. For an (N, s^n) design $D = (x_{ij})_{N\times n}$, define the L_p-distance (p is a positive integer) between the ith and jth rows of D as

$$d_p(x_i, x_j) = \sum_{k=1}^{n} |x_{ik} - x_{jk}|^p.$$

The maximin L_p-distance designs have the maximized values of $\min\{d_p(x_i, x_j) \mid 1 \le i < j \le N\}$. Here, $d_p(x_i, x_j)$ is an additive function and we can write it as $d_p(x_i, x_j) = \sum_{k=1}^{n} d_p(x_{ik}, x_{jk})$. In practice, $p = 1$ (Manhattan distance) and $p = 2$ (Euclidean distance) are the most commonly used.

Morris and Mitchell (1995) and Zhou and Xu (2014) developed some scalar metrics to quantify the designs' maximin distance properties as extensions of the maximin distance criterion. The metric defined by Morris and Mitchell (1995) is

$$\varphi(D) = \left(\sum_{1\le i<j\le N} \frac{1}{(d_p(x_i, x_j))^{q/p}} \right)^{1/q},$$

where q is a positive integer. Zhou and Xu (2014) further proposed a generalized metric. For any design D and $y \in (0, 1)$, consider the function

$$y^{d_p(x_i,x_j)} = \prod_{k=1}^{n} y^{d_p(x_{ik},x_{jk})},$$

which is a decreasing function of $d_p(x_i, x_j)$. Zhou and Xu (2014) proposed to minimize the scalar value

$$\phi(D) = \frac{1}{N^2} \sum_{i,j=1}^{N} y^{d_p(x_i,x_j)}. \tag{2}$$

Note that minimizing $\varphi(D)$ when $q \to \infty$ or minimizing $\phi(D)$ when $y \to 0$ are both asymptotically equivalent to maximizing the minimum L_p-distance of D. In this paper we consider the ϕ metric.

For a U-type initial design $D(N, s^n)$, let $\overline{\phi}(\tilde{D})$ be the average maximin distance metric values defined in Equation (2) over $\mathcal{EP}(D)$ (i.e. all the generated U-type designs $\tilde{D}(N, (ms)^n)$). The following Theorem 2 shows that $\overline{\phi}(\tilde{D})$ is a linear combination of the GWLP of the design D.

Theorem 2 *Given a U-type initial design $D(N; s^n)$, consider all possible level permutations and expansions. Suppose the maximin distance metric in Equation (2) is* *adopted such that the scalar value* $c_2(s;m)$ *defined in Equation* (3) *is larger than* 1. *Then we have*

$$\begin{aligned}\overline{\phi}(\tilde{D}) &= \frac{1}{n_0}\sum_{\tilde{D}\in\mathcal{EP}(D)}\phi(\tilde{D}) \\ &= \alpha\left(N,n,s,m\right) + \left[\frac{n_2^2}{n_1^2}\frac{c_1(s,m)(c_2(s,m)+s-1)}{s^2(s-1)}\right]^n\sum_{i=2}^{n}\left(\frac{c_2(s,m)-1}{c_2(s,m)+s-1}\right)^i A_i(D),\end{aligned}$$

where

$$\alpha\left(N,n,s,m\right) = \frac{1}{N} - \frac{1}{N}\left[\frac{n_2^2}{n_1^2}\frac{c_1(s,m)c_2(s,m)}{s(s-1)}\right]^n + \left[\frac{n_2^2}{n_1^2}\frac{c_1(s,m)(c_2(s,m)+s-1)}{s^2(s-1)}\right]^n,$$

$$c_1(s,m) = \sum_{\substack{l_1,l_2\in\mathcal{Z}_s\\ l_1\neq l_2}}\sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s}\\ x_{jk}\in\mathcal{T}_{l_2,s}}} y^{|x_{ik}-x_{jk}|^p},$$

$$c_2(s,m) = \frac{n_1 n_3}{n_2^2}\frac{s-1}{c_1(s,m)}\sum_{l\in\mathcal{Z}_s}\left(\sum_{\substack{x_{ik}\in\mathcal{T}_{l,s}\\ x_{jk}\in\mathcal{T}_{l,s}}} y^{|x_{ik}-x_{jk}|^p} - m\right), \tag{3}$$

$\mathcal{T}_{l,s} = \{lm, lm+1, \ldots, (l+1)m-1\}$, (n_1, n_2, n_3) *are given in Theorem 1 and* n_0 *is defined in Equation* (1).

For the commonly used $p = 1$ (Manhattan distance) and $p = 2$ (Euclidean distance), one can easily verify that the condition $c_2(s,m) > 1$ in Theorem 2 holds and thus $(c_2(s,m)-1)/(c_2(s,m)+s-1) \in (0,1)$. This means that the coefficient of $A_i(D)$ in the expression of $\overline{\phi}(\tilde{D})$ decreases exponentially as i increases. Therefore, Theorem 2 shows that starting from the GMA U-type designs in the level permutation and expansion method tends to generate better high-level U-type designs under the maximin distance criterion on average.

In the existing work for generating maximin distance designs, Zhou and Xu (2014) considered level permutations only and their Theorem 2 was a special case of our Theorem 2 with $m = 1$. Xiao and Xu (2018) considered both level permutations and expansions, but their Theorem 2 only connected the generated designs' expected distance variations with the $A_2(D)$ values; whereas, we connect $\overline{\phi}(\tilde{D})$ with the entire GWLP of the design D. Thus, Theorem 2 generalizes the related results in the above two papers.

3.3 Uniform designs

Uniformity is another widely used space-filling criterion. The basic idea behind uniform design is to scatter the design points as uniformly as possible in the design space by minimizing certain discrepancy metric. Hickernell (1998) developed several discrepancies defined on the reproducing kernel Hilbert spaces to measure designs' uniformity. Among them, the centered L_2-discrepancy (CD) and the wrap-around

L_2-discrepancy (WD) are the most commonly used. In addition, Zhou et al. (2013) proposed another popular mixture discrepancy (MD).

Let $\mathcal{K}(x,y) = \prod_{j=1}^{n} f(x_j, y_j)$ be a reproducing kernel defined on $[0,1]^n \times [0,1]^n$, where $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in [0,1]^n$. Here $f(x,y)$ satisfies the condition

$$f(x,y) \geq 0 \text{ and } f(x,x) + f(y,y) > f(x,y) + f(y,x) \quad \forall x \neq y, x, y \in [0,1]. \quad (4)$$

Then for an (N, s^n) design $D = (x_{ij})_{N\times n} = \{x_1, \ldots, x_N\}$, we can obtain the corresponding discrepancy according to the choice of $\mathcal{K}(\cdot,\cdot)$,

$$\mathrm{Disc}(D) = \mathcal{K}_2 - \frac{2}{N}\sum_{i=1}^{N}\prod_{k=1}^{n} f_1(x_{ik}) + \frac{1}{N^2}\sum_{i,j=1}^{N}\prod_{k=1}^{n} f(x_{ik}, x_{jk}), \quad (5)$$

where $\mathcal{K}_2 = \int_{[0,1]^n}\int_{[0,1]^n} \mathcal{K}(x,y)dxdy$ is a constant and $f_1(x) = \int_0^1 f(x,y)dy$. Below, we list the choices of $\mathcal{K}(x,y)$ and their corresponding discrepancies:

(i) for CD, $f(x,y) = 1 + (|x - 1/2| + |y - 1/2| - |x - y|)/2$;
(ii) for WD, $f(x,y) = 3/2 - |x-y| + |x-y|^2$;
(iii) for MD, $f(x,y) = 15/8 - (|x - 1/2| + |y - 1/2| + 3|x-y| - 2|x-y|^2)/4$.

Next, we show the explicit expressions of the squared CD, WD and MD for an (N, s^n) design $D = (x_{ik})_{N\times n}$:

$$\begin{aligned}
\mathrm{CD}(D) &= \left(\frac{13}{12}\right)^n - \frac{2}{N}\sum_{i=1}^{N}\prod_{k=1}^{n}\left(1 + \frac{1}{2}\left|u_{ik} - \frac{1}{2}\right| - \frac{1}{2}\left|u_{ik} - \frac{1}{2}\right|^2\right) \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\prod_{k=1}^{n}\left(1 + \frac{1}{2}\left|u_{ik} - \frac{1}{2}\right| + \frac{1}{2}\left|u_{jk} - \frac{1}{2}\right| - \frac{1}{2}\left|u_{ik} - u_{jk}\right|\right), \\
\mathrm{WD}(D) &= -\left(\frac{4}{3}\right)^n + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\prod_{k=1}^{n}\left(\frac{3}{2} - |u_{ik} - u_{jk}| + |u_{ik} - u_{jk}|^2\right), \\
\mathrm{MD}(D) &= \left(\frac{19}{12}\right)^n - \frac{2}{N}\sum_{i=1}^{N}\prod_{k=1}^{n}\left(\frac{5}{3} - \frac{1}{4}|u_{ik}| - \frac{1}{4}|u_{ik}|^2\right) \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i=1}^{N}\prod_{k=1}^{n}\left(\frac{15}{8} - \frac{1}{4}|u_{ik}| - \frac{1}{4}|u_{jk}| - \frac{3}{4}|u_{ik} - u_{jk}| + \frac{1}{2}|u_{ik} - u_{jk}|^2\right),
\end{aligned}$$

where $u_{ik} = (x_{ik} + 1/2)/s$ which re-scales x_{ik} to the $[0,1]$ range. Please refer to Hickernell (1998) and Zhou et al. (2013) for more details on the discrepancy criteria.

Consider generating U-type designs $\tilde{D}(N, (ms)^n)$ from a U-type design $D(N, s^n)$ via level permutations and expansions. Let $\overline{\mathrm{Disc}}(\tilde{D})$ be the average value of any kind of discrepancy defined by Equation (5) over $\mathcal{EP}(D)$. The following Theorem 3 establishes a connection between $\overline{\mathrm{Disc}}(\tilde{D})$ and the GWLP of the design D.

Theorem 3 *Given a U-type initial design $D(N; s^n)$, consider all possible level permutations and expansions. Suppose any kind of discrepancy defined by Equation (5) is adopted such that the scalar value $c_2(s;m)$ defined in Equation (6) is larger than 1. Then we have*

$$\overline{Disc}(\tilde{D}) = \frac{1}{n_0} \sum_{\tilde{D}\in\mathcal{EP}(D)} Disc(\tilde{D})$$
$$= \alpha\left(N,n,s,m\right) + \left[\frac{n_2^2}{n_1^2}\frac{c_1(s,m)(c_2(s,m)+s-1)}{s^2(s-1)}\right]^n \sum_{i=2}^{n}\left(\frac{c_2(s,m)-1}{c_2(s,m)+s-1}\right)^i A_i(D),$$

where

$$\alpha\left(N,n,s,m\right) = \mathcal{K}_2 - 2\left(\frac{n_2}{n_1 s}\sum_{l=0}^{ms-1} f_1\left(\frac{l+1/2}{ms}\right)\right)^n + \frac{1}{N}\left(\frac{n_2}{n_1 s}\sum_{l=0}^{ms-1} f\left(\frac{l+1/2}{ms},\frac{l+1/2}{ms}\right)\right)^n$$
$$-\frac{1}{N}\left[\frac{n_2^2}{n_1^2}\frac{c_1(s,m)c_2(s,m)}{s(s-1)}\right]^n + \left[\frac{n_2^2}{n_1^2}\frac{c_1(s,m)(c_2(s,m)+s-1)}{s^2(s-1)}\right]^n,$$

$$c_1(s,m) = \sum_{\substack{l_1,l_2\in\mathcal{Z}_s\\ l_1\neq l_2}}\ \sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s}\\ x_{jk}\in\mathcal{T}_{l_2,s}}} f(u_{ik},u_{jk}),$$

$$c_2(s,m) = \frac{n_1 n_3}{n_2^2}\frac{s-1}{c_1(s,m)}\sum_{l\in\mathcal{Z}_s}\left(\sum_{\substack{x_{ik}\in\mathcal{T}_{l,s}\\ x_{jk}\in\mathcal{T}_{l,s}}} f(u_{ik},u_{jk}) - \sum_{x_{ik}\in\mathcal{T}_{l,s}} f(u_{ik},u_{ik})\right), \tag{6}$$

$u_{ik} = (x_{ik}+1/2)/ms$ for $x_{ik}\in\mathcal{Z}_{ms}$, (n_1,n_2,n_3) are given in Theorem 1, $\mathcal{K}_2 = \int_{[0,1]^n}\int_{[0,1]^n}\mathcal{K}(x,y)dxdy$ and $\mathcal{T}_{l,s} = \{lm, lm+1,\ldots,(l+1)m-1\}$.

Theorem 3 shows that the average discrepancies of the designs generated by level permutations and expansions can be linearly expressed by the initial designs' GWLPs. It is straightforward to verify that for the commonly used discrepancies CD, WD and MD, the condition $c_2(s,m) > 1$ in Theorem 3 holds and thus $(c_2(s,m)-1)/(c_2(s,m)+s-1)\in(0,1)$. This means that the coefficient of $A_i(D)$ in the expression of $\overline{\text{Disc}}(\tilde{D})$ decreases exponentially as i increases. Therefore, Theorem 3 shows that starting from the GMA U-type designs in the level permutation and expansion method tends to give the best generated high-level U-type designs on average under the CD, WD and MD criteria.

Generating uniform designs has been previously studied by Zhou and Xu (2014) and Jiang and Ai (2017). Theorem 1 in Zhou and Xu (2014) is a special case of our Theorem 3 when $m = 1$, which considers level permutations only. Corollary 2 in Jiang and Ai (2017) is a special case of our Theorem 3 for $m = N/s$, which corresponds to LHDs.

4 Concluding remarks

The method of level permutation and expansion is a powerful tool to construct good high-level designs from low-level designs. In this paper, we establish theoretical connections between the U-type designs before and after level permutations and expansions under the orthogonality, maximin distance and uniformity criteria. These results generalize the existing results and provide theoretical guidance on the selections of good initial designs for the level permutation and expansion algorithms in Zhou and Xu (2014), Jiang and Ai (2017), and Xiao and Xu (2018). When following these algorithms to generate optimal U-type designs, we justify that OAs and GMA designs are good choices for the initial designs, which will lead to the best candidate sub-spaces to search from.

In this paper, we focus on pure-level designs that are not supersaturated, where each factor has the same number of levels and the run size is larger than the factor size. In practice, good mixed-level designs (i.e. factors can have different numbers of levels) and supersaturated designs (i.e. the run size is too small to estimate all the main effects) are also needed (Yamada and Matsui, 2002; Georgiou, 2014). It is promising to generalize the established results in this paper for mixed level designs and supersaturated designs. New algorithms based on the level permutation and expansion can be developed to construct such designs, which will be an interesting future work.

Most space-filling designs only consider the uniformity in full-dimensional projections. To guarantee the space-filling properties on all possible dimensions, Joseph et al. (2015) proposed the maximum projection (MaxPro) designs. How to construct MaxPro designs with flexible sizes is also a challenging question. Another possible future work is to generalize the established results for constructing the MaxPro designs via the level permutations and expansions.

Appendix: Proofs

To prove the theorems in this paper, the following lemma by Tang et al. (2012) and Zhou and Xu (2014) is needed.

Lemma 1 *For two rows x_i and x_j of an (N, s^n) design D, denote $\delta_{ij}(D)$ be the number of places where they take the same value. Then for any real number $z > 1$, we have*

$$\sum_{i,j=1}^{N} z^{\delta_{ij}(D)} = N^2 \left(\frac{z+s-1}{s}\right)^n \sum_{i=0}^{n} \left(\frac{z-1}{z+s-1}\right)^i A_i(D). \tag{7}$$

Proof of Theorem 1. Let $\tilde{D} = (x_{ik})$ be a U-type $(N, (ms)^n)$ design obtained by permutating and expanding levels of D. Because all the ms levels in $\mathcal{Z}_{ms}$ appear $N/(ms)$ times in each column of $\tilde{D}$, for the kth and k'th columns of $\tilde{D}$, their correlation is

$$\rho_{kk'} = \frac{12}{N\left(m^2 s^2 - 1\right)} \sum_{i=1}^{N} \tilde{x}_{ik} \tilde{x}_{ik'},$$

where $\tilde{x}_{ik} = x_{ik} - (ms-1)/2$. Let $D_{(k,k')}$ and $\tilde{D}_{(k,k')}$ respectively represent the sub-design of D and $\tilde{D}$ projected onto the kth and k'th columns. When all level permutations and expansions are considered, we can get a set of U-type $(N,(ms)^2)$ designs $\tilde{D}_{(k,k')}$, say $\mathcal{EP}(D_{(k,k')})$. Then the average ρ^2 value over $\mathcal{EP}(D_{(k,k')})$, denoted by $\overline{\rho^2}(\tilde{D}_{(k,k')})$, equals

$$\frac{144}{(s!)^2 n_1^{2s} N^2 \left(m^2s^2-1\right)^2} \sum_{\tilde{D}_{(j,k)} \in \mathcal{EP}(D_{(k,k')})} \left(\sum_{i=1}^{N} \tilde{x}_{ik}\tilde{x}_{ik'}\right)^2 .$$

Now we calculate the term $\sum_{\tilde{D}_{(k,k')} \in \mathcal{EP}(D_{(k,k')})} \left(\sum_{i=1}^{N} \tilde{x}_{ik}\tilde{x}_{ik'}\right)^2$. Expand the squares we further get the quadratic terms and interaction terms. First, the quadratic terms are

$$\begin{aligned} \sum_{\tilde{D}_{(k,k')} \in \mathcal{EP}(D_{(k,k')})} \sum_{i=1}^{N} \tilde{x}_{ik}^2\tilde{x}_{ik'}^2 &= \sum_{i=1}^{N} \sum_{\tilde{D}_{(k,k')} \in \mathcal{EP}(D_{(k,k')})} \tilde{x}_{ik}^2\tilde{x}_{ik'}^2 \\ &= \sum_{i=1}^{N} \left((s-1)! n_1^{s-1} n_2 \sum_{l=0}^{ms-1} l^2\right)^2 \\ &= N\left((s-1)! n_1^{s-1} n_2 \frac{(2ms-1)(ms-1)ms}{6}\right)^2 . \end{aligned} \tag{8}$$

Next, the interactions are

$$\begin{aligned} &\sum_{\tilde{D}_{(k,k')} \in \mathcal{EP}(D_{(k,k')})} \sum_{i,j=1, i\neq j}^{N} \tilde{x}_{ik}\tilde{x}_{ik'}\tilde{x}_{jk}\tilde{x}_{jk'} = \sum_{i,j=1,i\neq j}^{N} \sum_{\tilde{D}_{(k,k')} \in \mathcal{EP}(D_{(k,k')})} \tilde{x}_{ik}\tilde{x}_{jk}\tilde{x}_{ik'}\tilde{x}_{jk'} \\ &= \sum_{i,j=1,i\neq j}^{N} \left((s-2)! n_1^{s-2} n_2^2 \sum_{\substack{l_1,l_2\in\mathcal{Z}_s \\ l_1\neq l_2}} \sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s} \\ x_{jk}\in\mathcal{T}_{l_2,s}}} \tilde{x}_{ik}\tilde{x}_{jk}\right)^{2-\delta_{ij}(D_{(k,k')})} \times \\ &\sum_{i,j=1,i\neq j}^{N} \left((s-1)! n_1^{s-1} n_3 \sum_{\substack{l_1,l_2\in\mathcal{Z}_s \\ l_1 = l_2}} \left(\sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s} \\ x_{jk}\in\mathcal{T}_{l_2,s}}} \tilde{x}_{ik}\tilde{x}_{jk} - \sum_{x_{ik}\in\mathcal{T}_{l_1,s}} \tilde{x}_{ik}^2\right)\right)^{\delta_{ij}(D_{(k,k')})} \\ &= \left[(s-2)! n_1^{s-2} n_2^2 c_1(s,m)\right]^2 \sum_{i,j=1,i\neq j}^{N} c_2(s,m)^{\delta_{ij}(D_{(k,k')})} \\ &= \left[(s-2)! n_1^{s-2} n_2^2 c_1(s,m)\right]^2 \sum_{i,j=1}^{N} c_2(s,m)^{\delta_{ij}(D_{(k,k')})} - N\left[(s-2)! n_1^{s-2} n_2^2 c_1(s,m) c_2(s,m)\right]^2 , \end{aligned} \tag{9}$$

where $c_1(s, m)$ and $c_2(s, m)$ are defined in Theorem 1, and $\delta_{ij}(D_{(k,k')})$ is the number of places where the ith and jth rows of $D_{(k,k')}$ take the same value. Combining (8) with (9) and simplifying gives

$$\overline{\rho^2}(\tilde{D}_{(k,k')}) = \alpha_0(N, s, m) + \beta_0(N, s, m) \sum_{i,j=1}^{N} c_2(s, m)^{\delta_{ij}(D_{(k,k')})},$$

where

$$\alpha_0(N, s, m) = \frac{4{n_2}^2 m^2 (2ms-1)^2}{N{n_1}^2(ms+1)^2} - \frac{144 n_2^4 c_1^2(s, m) c_2^2(s, m)}{N n_1^4 s^2 (s-1)^2 \left(m^2 s^2 - 1\right)^2},$$

$$\beta_0(N, s, m) = \frac{144 n_2^4 c_1^2(s, m)}{N^2 n_1^4 s^2 (s-1)^2 \left(m^2 s^2 - 1\right)^2}.$$

Applying Lemma 1 for $n = 2$ and using the fact that $A_1(D_{(k,k')}) = 0$, we have

$$\begin{aligned}\overline{\rho^2}(\tilde{D}_{(k,k')}) = {} & \alpha_0(N, s, m) + \beta_0(N, s, m) N^2 \times \\ & \left[\left(\frac{c_2(s, m) + s - 1}{s}\right)^2 + \left(\frac{c_2(s, m) - 1}{s}\right)^2 A_2(D_{(k,k')})\right].\end{aligned} \tag{10}$$

Now we calculate $\overline{\rho^2}(\tilde{D})$. We have

$$\begin{aligned}\overline{\rho^2}(\tilde{D}) &= \frac{1}{n_0} \sum_{\tilde{D} \in \mathcal{EP}(D)} \rho^2(\tilde{D}) \\ &= \frac{1}{n_0} \sum_{\tilde{D} \in \mathcal{EP}(D)} \frac{1}{n(n-1)} \sum_{k,k'=1,k\neq k'}^{n} \rho_{kk'}^2 \\ &= \frac{1}{n(n-1)} \sum_{k,k'=1,k\neq k'}^{n} \frac{1}{n_0} \sum_{\tilde{D} \in \mathcal{EP}(D)} \rho_{kk'}^2 \\ &= \frac{1}{n(n-1)} \sum_{k,k'=1,k\neq k'}^{n} \overline{\rho^2}(\tilde{D}_{(k,k')}),\end{aligned}$$

where the last equality uses the fact that the average of $\rho_{kk'}^2(\tilde{D})$ over $\mathcal{EP}(D)$ is equal to the average of $\rho_{kk'}^2(\tilde{D}_{(k,k')})$ over $\mathcal{EP}(D_{(k,k')})$. Finally, by (10), the above equation and the fact that $A_2(D) = \sum_{k,k'=1,k\neq k'}^{n} A_2(D_{(k,k')})$, we have

$$\overline{\rho^2}(\tilde{D}) = \alpha(N, s, m) + \frac{144 n_2^4 c_1^2(s, m)(c_2(s, m) - 1)^2}{\binom{n}{2} n_1^4 s^4 (s-1)^2 \left(m^2 s^2 - 1\right)^2} A_2(D),$$

where $\alpha(N, s, m)$ is defined in Theorem 1. □

Proof of Theorem 2. Let $\tilde{D} = (x_{ik})$ be a U-type $(N, (ms)^n)$ design obtained by permutating and expanding levels of D. Then the entries $x_{ik} \in \mathcal{Z}_{ms}$. Because of the additivity of $d_p(\cdot,\cdot)$, we have

$$\begin{aligned}\overline{\phi}(\tilde{D}) &= \frac{1}{n_0 N^2}\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1}^{N}\prod_{k=1}^{n} y^{d_p(x_{ik},x_{jk})}\\ &= \frac{1}{n_0 N^2}\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i=1}^{N}\prod_{k=1}^{n} y^{0} + \frac{1}{n_0 N^2}\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1,i\neq j}^{N}\prod_{k=1}^{n} y^{d_p(x_{ik},x_{jk})}\\ &= \frac{1}{N} + \frac{1}{n_0 N^2}\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1,i\neq j}^{N}\prod_{k=1}^{n} y^{d_p(x_{ik},x_{jk})}\end{aligned} \tag{11}$$

In the last equation, the term $\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1,i\neq j}^{N}\prod_{k=1}^{n} y^{d_p(x_{ik},x_{jk})}$ can be similarly simplified as follows:

$$\begin{aligned}&\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1,i\neq j}^{N}\prod_{k=1}^{n} y^{d_p(x_{ik},x_{jk})} = \sum_{i,j=1,i\neq j}^{N}\sum_{\tilde{D}\in\mathcal{EP}(D)}\prod_{k=1}^{n} y^{d_p(x_{ik},x_{jk})}\\ &= \sum_{i,j=1,i\neq j}^{N}\sum_{\tilde{D}\in\mathcal{EP}(D)}\left(\prod_{k:\lfloor\frac{x_{ik}}{m}\rfloor=\lfloor\frac{x_{jk}}{m}\rfloor} y^{d_p(x_{ik},x_{jk})}\right)\left(\prod_{k:\lfloor\frac{x_{ik}}{m}\rfloor\neq\lfloor\frac{x_{jk}}{m}\rfloor} y^{d_p(x_{ik},x_{jk})}\right)\\ &= \sum_{i,j=1,i\neq j}^{N}\left((s-2)!n_1^{s-2}n_2^2\sum_{\substack{l_1,l_2\in\mathcal{Z}_s\\ l_1\neq l_2}}\sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s}\\ x_{jk}\in\mathcal{T}_{l_2,s}}} y^{d_p(x_{ik},x_{jk})}\right)^{n-\delta_{ij}(D)} \times\\ &\sum_{i,j=1,i\neq j}^{N}\left((s-1)!n_1^{s-1}n_3\sum_{\substack{l_1,l_2\in\mathcal{Z}_s\\ l_1=l_2}}\left(\sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s}\\ x_{jk}\in\mathcal{T}_{l_2,s}}} y^{d_p(x_{ik},x_{jk})} - \sum_{x_{ik}\in\mathcal{T}_{l_1,s}} y^{d_p(x_{ik},x_{ik})}\right)\right)^{\delta_{ij}(D)}\\ &= \left[(s-2)!n_1^{s-2}n_2^2c_1(s,m)\right]^n\sum_{i,j=1}^{N}c_2(s,m)^{\delta_{ij}(D)} - N\left[(s-2)!n_1^{s-2}n_2^2c_1(s,m)c_2(s,m)\right]^n,\end{aligned}$$

where $c_1(s,m)$ and $c_2(s,m)$ are defined in Theorem 2. The desired result then follows by substituting the above equation into (11) and applying Lemma 1 and the fact that $A_1(D) = 0$. $\square$

Proof of Theorem 3. For an $(N,(ms)^n)$ design $\tilde{D}=(x_{ij})$, the entries $x_{ik}\in\mathcal{Z}_{ms}$, denote $u_{ik}=(x_{ik}+1/2)/ms$. By (5), we have

$$\begin{aligned}
\overline{\mathrm{Disc}}(\tilde{D}) &= \frac{1}{n_0}\sum_{\tilde{D}\in\mathcal{EP}(D)}\mathrm{Disc}(\tilde{D}) \\
&= \frac{1}{n_0}\sum_{\tilde{D}\in\mathcal{EP}(D)}\left(\mathcal{K}_2-\frac{2}{N}\sum_{i=1}^{N}\prod_{k=1}^{n}f_1(u_{ik})+\frac{1}{N^2}\sum_{i,j=1}^{N}\prod_{k=1}^{n}f(u_{ik},u_{jk})\right) \\
&= \mathcal{K}_2-\frac{2}{n_0N}\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i=1}^{N}\prod_{k=1}^{n}f_1(u_{ik})+\frac{2}{n_0N^2}\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1}^{N}\prod_{k=1}^{n}f(u_{ik},u_{jk}).
\end{aligned} \tag{12}$$

Now we calculate the two terms in (12), $\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i=1}^{N}\prod_{k=1}^{n}f_1(u_{ik})$ and $\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1}^{N}\prod_{k=1}^{n}f(u_{ik},u_{jk})$, respectively. The first term is

$$\begin{aligned}
\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i=1}^{N}\prod_{k=1}^{n}f_1(u_{ik}) &= \sum_{i=1}^{N}\sum_{\tilde{D}\in\mathcal{EP}(D)}\prod_{k=1}^{n}f_1(u_{ik}) \\
&= N\left((s-1)!n_1^{s-1}n_2\sum_{l=0}^{ms-1}f_1\left(\frac{l+1/2}{ms}\right)\right)^n.
\end{aligned}$$

The second term can be further splitted into two terms as

$$\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i=1}^{N}\prod_{k=1}^{n}f(u_{ik},u_{ik})+\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1,i\neq j}^{N}\prod_{k=1}^{n}f(u_{ik},u_{jk}).$$

The former is equal to

$$N\left((s-1)!n_1^{s-1}n_2\sum_{l=0}^{ms-1}f\left(\frac{l+1/2}{ms},\frac{l+1/2}{ms}\right)\right)^n$$

and the latter is

$$\sum_{\tilde{D}\in\mathcal{EP}(D)}\sum_{i,j=1,i\neq j}^{N}\prod_{k=1}^{n} f(u_{ik},u_{jk}) = \sum_{i,j=1,i\neq j}^{N}\sum_{\tilde{D}\in\mathcal{EP}(D)}\prod_{k=1}^{n} f(u_{ik},u_{jk})$$

$$= \sum_{i,j=1,i\neq j}^{N}\sum_{\tilde{D}\in\mathcal{EP}(D)}\left(\prod_{k:\lfloor\frac{x_{ik}}{m}\rfloor=\lfloor\frac{x_{jk}}{m}\rfloor} f(u_{ik},u_{jk})\right)\left(\prod_{k:\lfloor\frac{x_{ik}}{m}\rfloor\neq\lfloor\frac{x_{jk}}{m}\rfloor} f(u_{ik},u_{jk})\right)$$

$$= \sum_{i,j=1,i\neq j}^{N}\left((s-2)!n_1^{s-2}n_2^2\sum_{\substack{l_1,l_2\in\mathcal{Z}_s\\ l_1\neq l_2}}\sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s}\\ x_{jk}\in\mathcal{T}_{l_2,s}}} f(u_{ik},u_{jk})\right)^{n-\delta_{ij}(D)} \times$$

$$\sum_{i,j=1,i\neq j}^{N}\left((s-1)!n_1^{s-1}n_3\sum_{\substack{l_1,l_2\in\mathcal{Z}_s\\ l_1=l_2}}\left(\sum_{\substack{x_{ik}\in\mathcal{T}_{l_1,s}\\ x_{jk}\in\mathcal{T}_{l_2,s}}} f(u_{ik},u_{jk}) - \sum_{x_{ik}\in\mathcal{T}_{l_1,s}} f(u_{ik},u_{ik})\right)\right)^{\delta_{ij}(D)}$$

$$= \left[(s-2)!n_1^{s-2}n_2^2c_1(s,m)\right]^n\sum_{i,j=1,i\neq j}^{N} c_2(s,m)^{\delta_{ij}(D)}$$

$$= \left[(s-2)!n_1^{s-2}n_2^2c_1(s,m)\right]^n\sum_{i,j=1}^{N} c_2(s,m)^{\delta_{ij}(D)} - N\left[(s-2)!n_1^{s-2}n_2^2c_1(s,m)c_2(s,m)\right]^n,$$

where $c_1(s,m)$ and $c_2(s,m)$ are defined in Theorem 3, and $\mathcal{T}_{l,s} = \{lm, lm+1,\dots,(l+1)m-1\}$.

Substituting the above equations into (12) and simplifying, we obtain

$$\overline{\text{Disc}}(\tilde{D}) = \alpha(N,n,s,m) + \frac{1}{N^2}\left(\frac{n_2^2}{n_1^2}\frac{c_1(s,m)}{s(s-1)}\right)^n\sum_{i,j=1}^{N} c_2(s,m)^{\delta_{ij}(D)},$$

where $\alpha(N,n,s,m)$ is defined in Theorem 3. Then the result follows by the above equation, Lemma 1 and the fact that $A_1(D)=0$. □

Acknowledgements

The research of Y. Wang was partly supported by NSFC (11901199 and 71931004) and Shanghai Sailing Program (19YF1412800). The research of F. Wang was partly supported by NSFC (12071014) and LMEQF. The research of Xiao was partly supported by the Junior Faculty Seed Grant in STEM Program (JFSGFY20) of the University of Georgia.

The authors thank the editor, the associate editor, and two referees for their valuable comments, which have led to the improvement of the paper.

Conflicts of interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Bingham, D., Sitter, R. R., and Tang, B., Orthogonal and nearly orthogonal designs for computer experiments, *Biometrika*, **96**, 51–65 (2009)

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

Fang, K. T., The uniform design: Application of number-theoretic methods in experimental design, *Acta Mathematicae Applicatae Sinica-English Series*, **3**, 363–372 (1980)

Fang, K. T., Li, R., and Sudjianto, A., *Design and Modeling for Computer Experiments*, CRC Press, New York (2006)

Georgiou, S. D., Supersaturated designs: A review of their construction and analysis, *Journal of Statistical Planning and Inference*, **144**, 92–109 (2014)

Gramacy, R. B., *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*, CRC Press, New York (2020)

Hickernell, F. J., A generalized discrepancy and quadrature error bound, *Mathematics of Computation*, **67**, 299–322 (1998)

Hedayat, A. S., Sloane, N. J. A., and Stufken, J., *Orthogonal Arrays: Theory and Applications*, Springer, New York (1999)

Jiang, B. and Ai, M., Construction of uniform U-designs, *Journal of Statistical Planning and Inference*, **181**, 1–10 (2017)

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

Joseph, V. R. and Hung, Y., Orthogonal-maximin Latin hypercube designs, *Statistica Sinica*, **18**, 171–186 (2008)

Joseph, V. R., Gul, E. and Ba, S., Maximum projection designs for computer experiments, *Biometrika*, **102**, 371–380 (2015)

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

Lin, C. D. and Tang, B., Latin hypercubes and space-filling designs, In *Handbook of Design and Analysis of Experiments*, A. Dean, M. Morris, J. Stufken and D. Bingham, eds., Chapman and Hall/CRC, New York, 593–625 (2015)

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

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

Owen, A. B., Controlling correlations in Latin hypercube samples, *Journal of the American Statistical Association*, **89**, 1517–1522 (1994)

Pronzato, L. and Müller, W. G., Design of computer experiments: space filling and beyond, *Statistics and Computing*, **22**, 681–701 (2012)

Santner, T. J., Williams, B. J. and Notz, W. I., *The Design and Analysis of Computer Experiments (2nd eds)*, Springer, New York (2018)

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

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

Tang, B., Selecting Latin hypercubes using correlation criteria, *Statistica Sinica*, **8**, 965–1397 (1998)

Tang, Y., Xu, H. and Lin, D. K. J., Uniform fractional factorial designs, *Annals of Statistics*, **40**, 891–907 (2012)

Tang, Y. and Xu, H., An effective construction method for multi-level uniform designs, *Journal of Statistical Planning and Inference*, **143**, 1583–1589 (2013)

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

Wang, Y., Yang, J. and Xu, H., On the connection between maximin distance designs and orthogonal designs, *Biometrika*, **105**, 471–477 (2018b)

Wang, Y., Sun, F. and Xu, H., On design orthogonality, maximin distance and projection uniformity for computer experiments, *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2020.1782221 (2020)

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

Xiao, Q. and Xu, H., Construction of maximin distance designs via level permutation and expansion, *Statistica Sinica*, **28**, 1395–1414 (2018)

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

Xu, G., Zhang, J. and Tang, Y., Level permutation method for constructing uniform designs under the wrap-around L_2-discrepancy, *Journal of Complexity*, **30**, 46–53 (2014)

Xu, H., Minimum moment aberration for nonregular designs and supersaturated designs, *Statistica Sinica*, **13**, 691–708 (2003)

Xu, H. and Wu, C. F. J., Generalized minimum aberration for asymmetrical fractional factorial designs, *Annals of Statistics*, **29**, 549–560 (2001)

Yamada, S., and Matsui, T., Optimality of mixed-level supersaturated designs, *Journal of Statistical Planning and Inference*, **104**, 459–468 (2002)

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

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

Zhou, Y. D. and Xu, H., Space-filling fractional factorial designs, *Journal of the American Statistical Association*, **109**, 1134–1144 (2014)