Statistica Sinica Preprint No: SS-2024-0145	
Title	Construction of Maximin Distance Latin Hypercube Designs via Good Lattice Point Sets
Manuscript ID	SS-2024-0145
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0145
Complete List of Authors	Xueru Zhang, Dennis K.J. Lin and Wei Zheng
Corresponding Authors	Wei Zheng
E-mails	wzheng9@utk.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

CONSTRUCTION OF MAXIMIN DISTANCE LATIN HYPERCUBE DESIGNS VIA GOOD LATTICE POINT SETS

Xueru Zhang[1], Dennis K.J. Lin[2] and Wei Zheng[3]

[1] *University of Science and Technology Beijing,* [2] *Purdue University and* [3] *University of Tennessee*

Abstract: Space-filling Latin hypercube designs have found widespread applications in computer experiments, yet the construction methods for such designs pose significant challenges. Algebraic methods are only applicable to a very limited number of runs and factors, while algorithmic searches often struggle with computational feasibility for large designs, especially when there is a need to maintain the statistical efficiency above a certain level. To address these limitations, an approach is proposed for producing space-filling Latin hypercube designs that can accommodate flexible numbers of runs and factors. The proposed approach is hybrid in nature, incorporating an algebraic method and its corresponding algorithm. The algebraic method, built on good lattice point sets and level permutation techniques, applies to any run size and flexible numbers of factors. The proposed algorithmic search can further accommodate any number of factors, especially those not covered by the algebraic method. A theoretical analysis of optimality is provided for the algebraic component. Numerical studies demonstrate the superior L_p-distance properties of the proposed designs. Furthermore, it is shown that the proposed designs exhibit good column-orthogonality and projection uniformity as well.

Key words and phrases: column-orthogonality, computer experiment, level permutation, projection uniformity, space-filling design

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

1. Introduction

Space-filling designs are widely utilized in computer experiments, such as Latin hypercube design (LHD) proposed by McKay, Beckman and Conovcr (1979) and uniform design (Fang and Wang, 1981). It aims to select experimental runs that efficiently cover the entire design space. Various criteria have been proposed to assess space-filling properties, such as the uniformity criterion (Weyl, 1916; Zhou, Fang and Ning, 2013), maximin distance criterion (Johnson, Moore and Ylvisaker, 1990), column-orthogonality (Owen, 1994), ϕ_p criterion (Joseph and Hung, 2008), projection uniformity (Moon, Dean and Santner, 2011) and minimum aberration type criterion (Tian and Xu, 2022). Among them, the maximin distance criterion, which maximizes the minimum pairwise distance between design points, is closely connected to other criteria (Joseph and Hung, 2008; Sun, Wang and Xu, 2019; Wang, Sun and Xu, 2022). For example, Wang, Sun and Xu (2022) demonstrated that maximin L_1- or L_2-distance designs often perform well on criteria of column-orthogonality and projection uniformity, especially for supersaturated designs (Lin, 1993).

Constructing space-filling designs poses significant challenges. Numerous algebraic construction methods have been proposed, such as Lin, Mukerjee and Tang (2009), Zhou and Xu (2015), Xiao and Xu (2017) and Wang, Xiao and Xu (2018). These algebraic methods are restricted to very specific numbers of runs and factors. For instance, Xiao and Xu (2017) constructed maximin distance LHDs. Their focus was on saturated and supersaturated designs, accommodating cases where the run size (n) is a prime number and equals the number of factors (m) or $m+1$. Zhou and Xu (2015)

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

and Vazquez and Xu (2024) proposed the maximin distance design with $m = n(n-1)$ or $m = n(n-1)/2$ for a prime n. To overcome this limitation, algorithmic search methods are recognized for their adaptability to flexible numbers of runs and factors (Moon, Dean and Santner, 2011; Ba, 2013; Joseph, Gul and Ba, 2015; Carnell, 2022). However, the search space grows exponentially with each additional factor and run, making it challenging for many existing algorithms to find optimal or near-optimal solutions within a reasonable time frame.

To address the limitations outlined in the aforementioned existing works, an efficient method based upon the maximin L_p-distance criterion is proposed. There are two major components in the proposed method: (a) an algebraic method and (b) a corresponding algorithm. The algebraic method, built on good lattice point (GLP) sets (Korobov, 1959) and the level permutation technique, allows an arbitrary n and a flexible m. Specially $m \approx \tilde{n}\psi(\tilde{n})$ or $m \approx \tilde{n}\psi(\tilde{n})/2$, the L_1- and L_2-distance efficiencies of the resulting designs are above 93.333% in all cases, where $\psi(\cdot)$ is the Euler function, and $\tilde{n} = n$ for odd n and $\tilde{n} = n+1$ for even n. An algorithm based on the algebraic method is developed for accommodating any number of factors (m) not exceeding $\tilde{n}\psi(\tilde{n})$. It is shown both theoretically and numerically that these proposed designs exhibit excellent space-filling properties in terms of the L_p-distance, column-orthogonality and projection uniformity.

The rest of the paper is organized as follows. Section 2 introduces a construction method and discusses the L_1- and L_2-distance of the proposed design. Using those constructed designs and simulated annealing, this section proposes a hybrid approach

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

to accommodate any number of columns not exceeding $\tilde{n}\psi(\tilde{n})$. Section 3 provides a method to improve the L_1- and L_2-distance for the even number of runs. Section 4 shows (theoretically and numerically) that the proposed designs have other excellent space-filling properties: column-orthogonality and projection uniformity. Section 6 provides both quantitative and qualitative comparisons between the proposed method and the existing works. It is shown that the proposed methods can produce more efficient LHDs than the existing works under these space-filling criteria: the L_p-distance, column-orthogonality, projection uniformity and ϕ_p. Section 7 presents concluding remarks.

2. The proposed construction method: additive column expansion

2.1 Preliminaries

Let $\left(n, s^k\right)$ be an $n \times k$ (n-run and k-factor) design where each factor takes levels from the set $Z_s = \{1, \ldots, s\}$. When each level appears equally often in each column, $\left(n, s^k\right)$ is called a U-type design and denoted by $\mathrm{U}(n, s^k)$. In particular, a U-type design is called a LHD if $s = n$. One such example well-studied in the literature is the GLP set (Korobov, 1959; Zhou and Xu, 2015; Fang et al., 2018). Let $n = q_1^{r_1} \cdots q_t^{r_t}$ be the prime decomposition of n, such that $q_1, \ldots, q_t$ are distinct primes in ascending order and $r_1, \ldots, r_t$ are positive integers. Let $\mathcal{H}_n = \{h | \gcd(n, h) = 1, h \in \{1, \ldots, n-1\}\}$, where gcd is the greatest common divisor between two numbers. The cardinality of $\mathcal{H}_n$ is $|\mathcal{H}_n| = \psi(n) = n\prod_{\ell=1}^{t}(1 - 1/q_\ell)$ where $\psi(\cdot)$ is also called the Euler function in the literature. A design D is called an $n \times k$ GLP set if its $(i, j)^{\text{th}}$ element is

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

$x_{ij} = ih_j \pmod n$, where h_j's are distinct elements of $\mathcal{H}_n$ and $k \leq \psi(n)$. When ih_j is divisible by n, the convention of $ih_j \pmod n = 0$ is replaced by n. The $\boldsymbol{h} = (h_1, \ldots, h_k)$ is called the generating vector of this GLP set. Note that each column of the GLP set is a permutation of $\{1, \ldots, n\}$. If $k = \psi(n)$, this GLP set has the largest number of columns and will be denoted by D_0 hereafter.

Define $d_p(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_{l=1}^{k} |x_{il} - x_{jl}|^p$ as the L_p-distance of the i^{th} and j^{th} rows of a design D, denoted by $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, respectively. Let $d_p(D) = \min\{d_p(\boldsymbol{x}_i, \boldsymbol{x}_j) : i \neq j, \boldsymbol{x}_i, \boldsymbol{x}_j \in D\}$ be the L_p-distance of D. When D is a $\mathrm{U}(n, s^k)$, Zhou and Xu (2015) provided an upper bound of the L_p-distance. This bound is defined as the integer part of the average pairwise L_p-distance between rows, denoted as $\lfloor d_{p,\text{ave}}(D) \rfloor$. The formulation of this upper bound for an LHD D is presented in the following lemma.

Lemma 1. (Zhou and Xu, 2015) *For an $n \times k$ LHD as D, $d_p(D) \leq \lfloor d_{p,\text{ave}}(D) \rfloor = \lfloor n^{p-1}(n+1)\ k/(3 \times 2^{p-1}) \rfloor$ for $p = 1, 2$. This bound is achieved when all pairwise distances of D are equal.*

This upper bound is applied to evaluate the L_p-distance efficiency of a design D as $\text{Eff}_p(D) = d_p(D)/\lfloor d_{p,\text{ave}}(D) \rfloor$. Note that a design with 100% efficiency is automatically optimal, while the reverse is not necessarily true.

2.2 Additive column expansion

Let $D = (x_{ij})$ be an $n \times k$ GLP, generated by $\boldsymbol{h} = (h_1, \ldots, h_k)$. Under the context of space-filling designs, Zhou and Xu (2015) showed that the new design $D_{\{u\}} = (x_{ij} \oplus u)_{1 \leq i \leq n, 1 \leq j \leq k}$ does not have a smaller L_p-distance than D where $u \in \mathcal{U}_0 = \{0, \ldots, n-1\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Here, the notation $\oplus$ is a linear level permutation operator as defined by

$$x_{ij} \oplus u = \begin{cases} n & \text{if } \ u = n - x_{ij}, \\ x_{ij} + u \pmod{n} & \text{otherwise.} \end{cases}$$

To accommodate a large number of factors, we propose to juxtapose a collection of such designs as

$$\mathcal{E}(D, \mathcal{U}) = \left(D_{\{u_1\}}, \ldots, D_{\{u_v\}}\right),$$

and call it the additive column expansion (ACE) of D on $\mathcal{U} = \{u_1, \ldots, u_v\} \subseteq \mathcal{U}_0$. Whenever $\mathcal{U}_0$ is used for the construction, the abbreviation $\mathcal{E}(D) = \mathcal{E}(D, \mathcal{U}_0)$ is adopted for any GLP set D. It is shown that $\mathcal{E}(D, \mathcal{U})$ is an $n \times (kv)$ LHD but not a GLP set. The L_p-distance of $\mathcal{E}(D, \mathcal{U})$ is closely related to that of D, as demonstrated below.

Lemma 2. *For arbitrary GLP set D and $\mathcal{U} \subseteq \mathcal{U}_0$, we have $d_p(\mathcal{E}(D, \mathcal{U})) \geq v d_p(D)$ where $v = |\mathcal{U}|$.*

The bound in Lemma 2 is achieved for specific combinations of D and $\mathcal{U}$, such as when D is an $n \times \psi(n)$ GLP set with n being a prime power, and $\mathcal{U} = \{0, 1\}$. As v increases, the difference $d_p(\mathcal{E}(D, \mathcal{U})) - v d_p(D)$ does not decrease and tends to grow larger. It has been demonstrated that GLP sets exhibit favorable space-filling properties, characterized by discrepancy and the L_p-distance (Hua and Wang, 1981). These point sets make the ACE an attractive framework for constructing maximin distance LHDs.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

2.3 Theoretical L_p-distances of ACEs

This section aims to investigate the theoretical lower bounds for the L_p-distance efficiencies of ACEs. We shall focus on cases of $p = 1, 2$, as the L_1- and L_2-distances, are commonly referred to as the rectangular distance and the squared Euclidean distance, respectively. They are widely employed to evaluate space-filling properties in experimental design, more than distances with $p > 2$ (Johnson, Moore and Ylvisaker, 1990; Zhou and Xu, 2015; Wang, Xiao and Xu, 2018; Wang, Sun and Xu, 2022).

A closed-form expression of the L_p-distances for ACEs becomes intractable for other values of m. However, their lower bounds are obtained as follows.

Corollary 1. *For $n = q_1^{r_1} \dots q_t^{r_t}$, any $\mathcal{U} \subseteq \mathcal{U}_0$ and $D = D_0$, we have*

$$\mathrm{Eff}_1(\mathcal{E}(D,\mathcal{U})) \geq \begin{cases} \dfrac{3n}{4(n+1)} & \text{if } \ t = 1 \text{ and } q_1 = 2, \\ \dfrac{3(n^2+q_1)}{4n(n+1)} & \text{if } \ t = 1 \text{ and } q_1 > 2, \\ \dfrac{3(q_2-1)}{2(n+1)} & \text{if } \ t = 2, q_1 = 2 \text{ and } r_1 = r_2 = 1, \\ \dfrac{3(n-1)}{4(n+1)} & \text{if } \ t = 2, q_1 > 2, \text{ and } r_1 = r_2 = 1. \end{cases}$$

The bound in Corollary 1 is achieved for specific $\mathcal{U}$ and $k(n+1) \pmod 3 = 0$, such as $\mathcal{U} = \{0, 1\}$. As v increases, the value of $\mathrm{Eff}_1(\mathcal{E}(D,\mathcal{U}))$ tends to grow larger. Based on Theorems 2 and 4 of Zhou and Xu (2015), the proof of Corollary 1 follows directly, which is therefore omitted here. Similar results for $p = 2$ can be derived using Theorem 2 of Zhou and Xu (2015) and $d_2(D)$, where $d_2(D)$ can be obtained through a method analogous to that used in the proof of Theorem 4 of Zhou and Xu (2015).

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Table 1: The L_1-distance efficiencies of some ACEs.

n	m	D	$\mathcal{U}$	d_1	Eff_1	n	m	D	$\mathcal{U}$	d_1	Eff_1
3	2	D_0	$\{0\}$	2	1	7	6	D_0	$\{3\}$	13	0.813
	4		$\{0,2\}$	5	1		12		$\{0,2\}$	30	0.938
	6		$\{0,1,2\}$	8	1		18		$\{0,3,4\}$	46	0.958
5	4	D_0	$\{0\}$	6	0.75		24		$\{0,2,3,4\}$	62	0.969
	8		$\{0,2\}$	14	0.875		30		$\{0,2,3,4,6\}$	78	0.975
	12		$\{0,2,3\}$	23	0.958		36		$\{0,1,2,3,4,5\}$	94	0.979
	16		$\{0,1,2,3\}$	30	0.938		42		$\{0,1,2,3,4,5,6\}$	112	1
	20		$\{0,1,2,3,4\}$	40	1		3	D_1	$\{0\}$	6	0.75
	2	D_1	$\{0\}$	3	0.75		6		$\{3,5\}$	13	0.813
	4		$\{0,1\}$	6	0.75		9		$\{2,4,6\}$	22	0.917
	6		$\{0,1,3\}$	10	0.833		12		$\{0,1,2,6\}$	30	0.938
	8		$\{0,1,2,4\}$	14	0.875		15		$\{0,1,3,4,5\}$	38	0.95
	10		$\{0,1,2,3,4\}$	20	1		18		$\{0,1,2,3,4,5\}$	44	0.917
							21		$\{0,1,2,3,4,5,6\}$	56	1

Note: d_1 represents the L_1-distance and Eff_1 represents the L_1-distance efficiency.

Example 1. Given n and D, we seek the $\mathcal{U}$ that optimizes the ACE under the L_1-distance among all possible sets with the same cardinality. The optimal ACEs for $n = 3, 5, 7$ and $D = D_0$ are displayed in Table 1. It is shown that the L_1-distance efficiencies of all ACEs are greater than 0.75 in most cases (some are reaching 100%). The situation on the L_2-distance is shown in Table S8 in the supplementary materials. Overall, ACE performs well on the L_1- and L_2-distance criteria.

First, the ACE with the largest number of factors is considered, i.e., $D = D_0$ and $\mathcal{U} = \mathcal{U}_0$, denoted as $\mathcal{E}(D_0)$. The closed-form expressions of those distances for $\mathcal{E}(D_0)$ are provided as follows.

Theorem 1. *For $n = q_1^{r_1} \dots q_t^{r_t}$, let $k = \psi(n)$ and $p = 1, 2$. The L_p-distance of $\mathcal{E}(D_0)$*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

is

$$d_p(\mathcal{E}(D_0)) = \begin{cases} \dfrac{n^p(n^2+q_1)(q_1-1)}{2^{p-1}3q_1} & \text{if } \ t=1, \\ \dfrac{n^{p+2}(q_1q_2-1)\prod_{\ell=1}^{t}(q_\ell-1)}{2^{p-1}3q_1q_2\prod_{\ell=1}^{t}q_\ell} & \text{if } \ t>1. \end{cases}$$

Theorem 1 provides the closed-form expressions for the L_1- and L_2-distances of $\mathcal{E}(D_0)$ with any run size. This result generalizes the findings from Zhou and Xu (2015) from a prime n to an arbitrary n. To further accommodate a more flexible number of factors, the designs constructed with arbitrary values of $k < \psi(n)$ are considered in the following.

For a given GLP set D generated by $\boldsymbol{h} = (h_1, \ldots, h_k)$, its dual design is the GLP set generated by $\boldsymbol{h}' = (n-h_1, \ldots, n-h_k)$, denoted by D'.

Lemma 3. *For an arbitrary $\mathcal{U} \subset \mathcal{U}_0$ and a pair of dual GLP sets, say D and D', their ACEs $\mathcal{E}(D,\mathcal{U})$ and $\mathcal{E}(D',\mathcal{U})$ have the same L_p-distance for any p.*

When $\boldsymbol{h}$ and $\boldsymbol{h}'$ have no common elements, D and D' are called strictly dual. Note that the column numbers of the strictly dual GLP sets do not exceed $\psi(n)/2$ since $n-h \in \mathcal{H}_n$ if and only if $h \in \mathcal{H}_n$. Let D_1 be a GLP set generated by half of the elements in $\mathcal{H}_n$, such that its dual D_1' is its strict dual. Then, the two proposed designs $\mathcal{E}(D_1)$ and $\mathcal{E}(D_1')$ have the same number of columns as $n\psi(n)/2$. Note that they complement each other in the sense that the column juxtaposition of $\mathcal{E}(D_1)$ and $\mathcal{E}(D_1')$ reassembles $\mathcal{E}(D_0)$. If we change D_0 to D_1, Corollary 1 remains valid. The L_1- and L_2-distance efficiencies of the ACEs with $D = D_1$ are listed in Table 1 and Table S8. The ACEs perform well on these criteria. Based on Lemma 3, Theorem 2 allows

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

us to calculate the L_1- and L_2-distances of $\mathcal{E}(D_1)$ and $\mathcal{E}(D_1')$.

Theorem 2. *For D_0, D_1, D_1' as defined earlier and any p, we have*

(i) $d_p(\mathcal{E}(D_1)) = d_p(\mathcal{E}(D_1')) = d_p(\mathcal{E}(D_0))/2$,

(ii) $d_p(\mathcal{E}(D_1)) \geq d_p(\mathcal{E}(D))$ *for any* $n \times (\psi(n)/2)$ *GLP set* D,

where $d_p(\mathcal{E}(D_0))$ is as shown in Theorem 1.

Combining Theorem 1 and Theorem 2(i), we obtain the closed forms of the L_1- and L_2-distances of $\mathcal{E}(D_1)$ and $\mathcal{E}(D_1')$. Moreover, Theorem 2 proves that both $\mathcal{E}(D_1)$ and $\mathcal{E}(D_1')$ reach the maximal L_p-distance among all $\mathcal{E}(D)$'s, where D is an $n \times (\psi(n)/2)$ GLP set.

Theorem 3. *Given any n and p, we have $d_p(\mathcal{E}(D_1)) = d_p(\mathcal{E}(D_1')) \geq d_p(\mathcal{D})$, for any $\mathcal{D}$ that is an $n \times n\psi(n)/2$ submatrix of $\mathcal{E}(D_0)$*

While Theorem 2 establishes the optimality of $\mathcal{E}(D_1)$ among all $\mathcal{E}(D)$'s with D being an $n \times \psi(n)/2$ GLP set, Theorem 3 strengthens this result by demonstrating the optimality of $\mathcal{E}(D_1)$ over all possible $n\psi(n)/2$-column combinations within $\mathcal{E}(D_0)$. Combining Theorems 1 & 2 and Lemma 1, the lower bounds of the L_p-distance efficiencies for the proposed designs can be derived as follows.

Theorem 4. *Let $p = 1, 2$ and $D = D_0$ or D_1. Given any n, the L_p-distance efficiency of $\mathcal{E}(D)$ follows*

$$\mathrm{Eff}_p(\mathcal{E}(D)) = \frac{d_p(\mathcal{E}(D))}{\lfloor d_{p,\mathrm{ave}}(\mathcal{E}(D)) \rfloor} \geq \frac{d_p(\mathcal{E}(D))}{d_{p,\mathrm{ave}}(\mathcal{E}(D))} = \begin{cases} 1 - O\left(\dfrac{1}{n} - \dfrac{q_1}{n^2}\right) & \text{if } \ t = 1, \\ 1 - O\left(\dfrac{1}{q_1 q_2} + \dfrac{1}{n}\right) & \text{if } \ t > 1. \end{cases}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

The implications of Theorem 4 are as follows. For the prime n, i.e. $t = 1$ and $r_1 = 1$, $\mathcal{E}(D)$ is an equidistant design and its L_p-distance reaches the upper bound $d_{p,\text{ave}}$. This design serves as a maximin distance design with $\text{Eff}_p(\mathcal{E}(D)) = 1$. For the prime power value of n, i.e. $t = 1$, $\mathcal{E}(D)$ is asymptotically optimal under the L_1- and L_2-distance criteria. Otherwise, for $t > 1$, the limiting efficiency is $\text{Eff}_p(\mathcal{E}(D)) \to 1 - 1/(q_1q_2)$ as $n \to \infty$. That is, among the nonprimer power n, the limiting efficiency depends on the two minimum prime numbers in its decomposition. This is because the terms associated with other primes in the numerator are identical to those in the denominator. The limiting efficiency increases as q_1q_2 increases. When $(q_1, q_2) = (2, 3)$, the limiting efficiency reaches its minimum value of $1 - 1/(q_1q_2) = 1 - 1/6 = 83.333\%$. In all other cases, except for $(q_1, q_2) = (2, 3)$, the resulting designs yield the limiting efficiencies of at least 90% occurring at $(q_1, q_2) = (2, 5)$. A remedy for those not-so-ideal cases of $(q_1, q_2) = (2, 3)$ through the leave-one-out method (Fang and Wang, 1981) will be provided in Section 3.

For any $\mathcal{U} = \{u_1, \ldots, u_v\} \subseteq \mathcal{U}_0$, its dual set is defined as $\mathcal{U}' = \{n+1-u_1, \ldots, n+1-u_v\} \pmod n$. The following lemma reveals the relationships between different ACEs in terms of the L_p-distance.

Lemma 4. *Let D and D' be a pair of $n \times k$ dual GLP sets, and $\mathcal{U}$ and $\mathcal{U}'$ be a pair of dual sets contained in $\mathcal{U}_0$. For any p, the two ACEs $\mathcal{E}(D,\mathcal{U})$ and $\mathcal{E}(D',\mathcal{U}')$ have the same L_p-distance.*

Unlike Lemma 3, Lemma 4 allows for both the GLP set D and index set $\mathcal{U}$ to be replaced by their dual simultaneously. Combining Lemmas 3 and 4, it holds that

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

the L_p-distance of $\mathcal{E}(D,\mathcal{U})$ remains unchanged if D is replaced by D' and/or $\mathcal{U}$ is replaced by $\mathcal{U}'$. Note D_0 is the dual of itself, i.e. $D_0 = D_0'$. Next, the ACEs with $D = D_0$ and $\mathcal{U} \in \{\mathcal{U}_1, \mathcal{U}_2\}$ are theoretically investigated for $\mathcal{U}_1 = \{1, \ldots, \lceil (n-1)/2 \rceil\}$ and $\mathcal{U}_2 = \{1, \ldots, \lfloor (n+1)/2 \rfloor\}$. Here, $\mathcal{U}_1$ and $\mathcal{U}_2$ are only one element different when n is odd, and $\mathcal{U}_1 = \mathcal{U}_2$ for an even n. Based on Lemma 4, the L_1- and L_2-distances of $\mathcal{E}(D_0,\mathcal{U}_1)$ and $\mathcal{E}(D_0,\mathcal{U}_2)$ are shown in the following theorem.

Theorem 5. *Let $p = 1, 2$. For D_0, $\mathcal{U}_1$ and $\mathcal{U}_2$ as defined earlier, we have*

(i) *when n is even, $d_p(\mathcal{E}(D_0,\mathcal{U}_1)) = d_p(\mathcal{E}(D_0,\mathcal{U}_2)) = d_p(\mathcal{E}(D_0))/2$;*

(ii) *when n is odd,*

$$d_p(\mathcal{E}(D_0,\mathcal{U})) = \begin{cases} \dfrac{(n - 2^{2p-3}3^{2-p}) d_p(\mathcal{E}(D_0))}{2n} & \text{if } \mathcal{U} = \mathcal{U}_1, \\ \dfrac{(n + 2^{p-3}3^{2-p}) d_p(\mathcal{E}(D_0))}{2n} & \text{if } \mathcal{U} = \mathcal{U}_2, \end{cases}$$

where $d_p(\mathcal{E}(D_0))$ is as shown in Theorem 1.

The designs considered in Theorem 5 have either $\lceil (n-1)/2 \rceil \psi(n)$ or $\lfloor (n+1)/2 \rfloor \psi(n)$ columns, which is similar to that of $\mathcal{E}(D_1)$. Note that $\mathcal{E}(D_0,\mathcal{U}_1)$ and $\mathcal{E}(D_0,\mathcal{U}_2')$ complement each other in $\mathcal{E}(D_0)$, as well as $\mathcal{E}(D_0,\mathcal{U}_1')$ and $\mathcal{E}(D_0,\mathcal{U}_2)$. Their L_p-distance efficiencies are obtained as follows.

Corollary 2. *Let $p = 1, 2$. For the proposed designs in Theorem 5 and $\mathcal{U} = \mathcal{U}_1$ or $\mathcal{U}_2$,*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

their L_p-distance efficiencies follow

$$\mathrm{Eff}_p(\mathcal{E}(D_0,\mathcal{U})) \to \begin{cases} 1 & \text{if} \quad t=1, \\ 1-\frac{1}{q_1 q_2} & \text{if} \quad t>1, \end{cases} \qquad \text{as} \quad n\to\infty.$$

Furthermore, the L_p-distance efficiency of each design follows the same order as that of the design in Theorem 4.

The L_1- and L_2-distance properties of the proposed designs have been investigated so far for arbitrary values of n. The theoretical results focus on the number of factors being either $m \approx n\psi(n)$ or $m \approx n\psi(n)/2$, depending on whether the GLP set D_1 is used instead of D_0 or the index sets $\mathcal{U}_1$ and $\mathcal{U}_2$ are used (instead of $\mathcal{U}_0$ for the construction).

3. Leave-one-out additive column expansion

The ACE can efficiently deal with any run size n, and it is particularly well-suited for odd values of n, as evident from the theoretical results presented in Theorem 3 and Corollary 2. It has been observed that there could be occasional drops in efficiency for an even n. For instance, when n is a multiple of 6 with $(q_1, q_2) = (2, 3)$, the limiting efficiency lower bound equals 83.333% and $\psi(n)$ does not exceed $n/3$, implying that the maximum number of factors that can be constructed by the ACE is $n^2/3$. In such cases, we shall use the leave-one-out technique (Fang and Wang, 1981) to improve the efficiency and scalability of resulting designs. This method first constructs an $(n+1)$-run ACE and then adjusts all levels to $1,\ldots,n$ by removing the last row of the ACE. As an example when $n = 6$, this method entails removing the last row of the 7-run

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

$\mathcal{E}(D_0)$ as $(7,\ldots,7,1,\ldots,1,\ldots,6,\ldots,6)$ and adjusting all levels to $1,\ldots,6$. It is shown that the resulting design is a 6×42 LHD with an L_p-distance efficiency of 95.918%. We call such a design the leave-one-out additive column expansion (LACE).

The LACE can be systematically constructed by a leave-one-out good lattice point (LGLP) set $\tilde{D}$. An $n \times k$ array $(x_{ij})_{1\le i\le n, 1\le j\le k}$ is called an LGLP set $\tilde{D}$ if $x_{ij} = ih_j \pmod{n+1}$, where h_j's are distinct elements of $\mathcal{H}_{n+1}$ and thus $k \le \psi(n+1)$. Note that each column of the LGLP set is a permutation of $\{1,\ldots,n\}$. The difference between this n-run LGLP set and the $(n+1)$-run GLP set constructed by the same generating vector is that the former excludes the row with all elements being $n+1$. Nevertheless, the two designs have the same L_p-distance. When $k = \psi(n+1)$, $\tilde{D}$ has the largest number of columns and will be denoted by $\tilde{D}_0$ hereafter. In comparison, the n-run GLP set D_0 has $\psi(n)$ columns while the n-run LGLP set $\tilde{D}$ has $\psi(n+1)$ columns. Define the new design $\tilde{D}_{\{u\}} = (x_{ij} \oplus_L u)_{1\le i\le n, 1\le j\le k}$ for $u \in \tilde{\mathcal{U}}_0 = \{0,\ldots,n\}$, where the operator $\oplus_L$ is defined by

$$x_{ij} \oplus_L u = \begin{cases} u & \text{if } \ u = n+1-x_{ij}, \\ x_{ij} + u \pmod{n+1} & \text{otherwise.} \end{cases}$$

Each column of $\tilde{D}_{\{u\}}$ is a permutation of $\{1,\ldots,n\}$. For a set $\mathcal{U} = \{u_1,\ldots,u_v\} \subseteq \tilde{\mathcal{U}}_0$, an LACE can be constructed by the LGLP set $\tilde{D}$ and $\mathcal{U}$, as $\mathcal{L}(\tilde{D},\mathcal{U}) = (\tilde{D}_{\{u_1\}},\ldots,\tilde{D}_{\{u_v\}})$. This expansion is abbreviated as $\mathcal{L}(\tilde{D})$ when $\mathcal{U} = \tilde{\mathcal{U}}_0$. Note that $\mathcal{L}(\tilde{D},\mathcal{U})$ forms an $n \times (kv)$ LHD. For example, the 6×42 LHD discussed in the previous paragraph is an LACE constructed by the 6×6 LGLP set $\tilde{D}_0$ and $\tilde{\mathcal{U}}_0 = \{0,\ldots,6\}$. Let $(n+1) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

$\tilde{q}_1^{\tilde{r}_1} \cdots \tilde{q}_{\tilde{t}}^{\tilde{r}_{\tilde{t}}}$ be the prime decomposition of $n+1$, such that $\tilde{q}_1, \ldots, \tilde{q}_{\tilde{t}}$ are distinct primes in ascending order and $\tilde{r}_1, \ldots, \tilde{r}_{\tilde{t}}$ are positive integers. Without loss of generality, assume that $\tilde{q}_1 < \cdots < \tilde{q}_{\tilde{t}}$. The following theorem provides the closed forms of the L_1- and L_2-distance for the largest design $\mathcal{L}(\tilde{D}_0)$.

Theorem 6. *Let $p = 1, 2$. For an arbitrary even n (i.e. $q_1 = 2$), the L_p-distance of $\mathcal{L}(\tilde{D}_0)$ is*

$$d_p(\mathcal{L}(\tilde{D}_0)) = \begin{cases} \dfrac{(n+1)^{p-1}(\tilde{q}_1 - 1)((n+1)^3 + \tilde{q}_1(n+1) - 2^{p-2}3((n+1)^2 - \tilde{q}_1))}{2^{p-1}3\tilde{q}_1} & \text{if } \tilde{t} = 1, \\ \dfrac{(n+1)^{p+1}\prod_{\ell=1}^{\tilde{t}}(\tilde{q}_\ell - 1)((n+1)(\tilde{q}_1\tilde{q}_2 - 1) - 2^{p-2}3(\tilde{q}_1\tilde{q}_2 + 1))}{2^{p-1}3\tilde{q}_1\tilde{q}_2\prod_{\ell=1}^{\tilde{t}}\tilde{q}_\ell} & \text{if } \tilde{t} > 1. \end{cases}$$

Theorem 6 provides the closed forms of the L_1- and L_2-distances for $\mathcal{L}(\tilde{D}_0)$. For an $n \times m$ LGLP set $\tilde{D}$ generated by $\boldsymbol{h} = (h_1, \ldots, h_k)$, its dual design is the LGLP set generated by $\boldsymbol{h}^+ = (n+1-h_1, \ldots, n+1-h_k)$, denoted by $\tilde{D}^+$. When $\boldsymbol{h}$ and $\boldsymbol{h}^+$ has no common elements, $\tilde{D}$ and $\tilde{D}^+$ are called strictly dual. Let $\tilde{D}_1$ be an LGLP set generated by half of the elements in $\mathcal{H}_{n+1}$, such that its dual $\tilde{D}_1^+$ is its strict dual. For any $\mathcal{U} = \{u_1, \ldots, u_v\} \subseteq \tilde{\mathcal{U}}_0$, its dual set is defined as $\mathcal{U}^+ = \{n+1-u_1, \ldots, n+1-u_v\} \pmod{(n+1)}$. The results in Lemmas 3 and 4, and Theorem 2 regarding ACE also have their LACE versions, as explicitly stated below.

Corollary 3. *For any p and an arbitrary $\mathcal{U} \subseteq \tilde{\mathcal{U}}_0$, we have*

(i) $d_p(\mathcal{L}(\tilde{D}, \mathcal{U})) = d_p(\mathcal{L}(\tilde{D}, \mathcal{U}^+)) = d_p(\mathcal{L}(\tilde{D}^+, \mathcal{U})) = d_p(\mathcal{L}(\tilde{D}^+, \mathcal{U}^+))$;

(ii) $d_p(\mathcal{L}(\tilde{D}_1)) = d_p(\mathcal{L}(\tilde{D}_1^+)) = d_p(\mathcal{L}(\tilde{D}_0))/2$;

(iii) $d_p(\mathcal{L}(\tilde{D}_1)) \geq d_p(\mathcal{L}(\tilde{D}))$ *for any* $n \times (\psi(n+1)/2)$ *LGLP set* $\tilde{D}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

where $d_p(\mathcal{L}(\tilde{D}_0))$ is as shown in Theorem 6.

Corollary 3 shows that the L_p-distance of $\mathcal{L}(\tilde{D},\mathcal{U})$ remains unchanged if $\tilde{D}$ is replaced by $\tilde{D}^+$ and/or $\mathcal{U}$ is replaced by $\mathcal{U}^+$. When $\mathcal{U} = \tilde{\mathcal{U}}_0$, it is shown that $\mathcal{L}(\tilde{D}_1)$ has the largest L_p-distance among all $\mathcal{L}(\tilde{D})$'s, with $\tilde{D}$ being an $n \times (\psi(n+1)/2)$ LGLP set. Moreover, the L_p-distance of $\mathcal{L}(\tilde{D}_1)$ is half as much as $d_p(\mathcal{L}(\tilde{D}_0))$.

Let $\mathcal{U}_3 = \{0, 1, \ldots, \lfloor n/2 \rfloor\}$. Based on Corollary 3, the closed forms of the L_1- and L_2-distances for the two special LACEs of $\tilde{D}_0$ on either $\mathcal{U}_1$ or $\mathcal{U}_3$ are provided as follows.

Theorem 7. *Let $p = 1, 2$ and n be an even integer. For D_0, $\tilde{D}_0$, $\mathcal{U}_1$ and $\mathcal{U}_3$ as defined earlier, we have*

(i) $d_1(\mathcal{L}(\tilde{D}_0,\mathcal{U}_1)) = d_1(\mathcal{L}(\tilde{D}_0)) - d_1(\mathcal{E}(D_0))/2$;

(ii) $d_2(\mathcal{L}(\tilde{D}_0,\mathcal{U}_1)) = (d_2(\mathcal{L}(\tilde{D}_0)) - \Lambda(n))/2$;

(iii) $d_1(\mathcal{L}(\tilde{D}_0,\mathcal{U}_3)) = d_1(\mathcal{E}(D_0))/2$;

(iv) $d_2(\mathcal{L}(\tilde{D}_0,\mathcal{U}_3)) = d_2(\mathcal{E}(D_0))/2 - d_1(\mathcal{E}(D_0))/4$ *when* $n+1$ *is a prime,*

where $d_p(\mathcal{L}(\tilde{D}_0))$ is as shown in Theorem 6, $d_p(\mathcal{E}(D_0))$ is as shown in Theorem 1 (replacing n by $n+1$ here) and

$$\Lambda(n) = \begin{cases} \dfrac{(n+1)(\tilde{q}_1 n(n-1) - (n+1-\tilde{q}_1)(n+1-2\tilde{q}_1))}{3\tilde{q}_1} & \text{if } \tilde{t} = 1, \\ \dfrac{(n+1)^3 \left(\prod_{\ell=0}^{2}(\tilde{q}_1\tilde{q}_2 - \ell) - \tilde{q}_1^2 \prod_{\ell=0}^{2}(\tilde{q}_2 - \ell) - \tilde{q}_2^2 \prod_{\ell=0}^{2}(\tilde{q}_1 - \ell)\right) \prod_{\ell=3}^{\tilde{t}}(\tilde{q}_\ell - 1)}{3\tilde{q}_1^2\tilde{q}_2^2 \prod_{\ell=1}^{\tilde{t}} \tilde{q}_\ell} & \text{if } \tilde{t} > 1. \end{cases}$$

In Theorem 7, the LACEs have either $n\psi(n+1)/2$ columns or $(n+2)\psi(n+1)/2$ columns for an arbitrary even n, which has a similar design size as $\mathcal{L}(\tilde{D}_1)$. When $n+1$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

is a prime, the proposed design $\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)$ is an equidistant LHD achieving the upper bound $d_{p,\text{ave}}$ under the L_1- and L_2-distance criteria. Analogous to Theorem 4, we have the following result.

Corollary 4. *Let $p = 1, 2$ and n be an even integer. For each $\mathcal{D}$ in Corollary 3 (ii) and Theorem 7, its L_p-distance efficiency follows*

$$\text{Eff}_p(\mathcal{D}) = \frac{d_p(\mathcal{D})}{\lfloor d_{p,\text{ave}}(\mathcal{D}) \rfloor} \geq \frac{d_p(\mathcal{D})}{d_{p,\text{ave}}(\mathcal{D})} = \begin{cases} 1 - O\left(\dfrac{1}{n} - \dfrac{\tilde{q}_1}{n^2}\right) & \text{if } \tilde{t} = 1, \\ 1 - O\left(\dfrac{1}{\tilde{q}_1\tilde{q}_2} + \dfrac{1}{n}\right) & \text{if } \tilde{t} > 1. \end{cases}$$

When $n + 1$ is a prime power, i.e. $\tilde{t} = 1$, the LACE $\mathcal{D}$ is asymptotically optimal under the L_1- and L_2-distance criteria. When $\tilde{t} \geq 2$, we have the limiting efficiency of $1 - 1/(\tilde{q}_1\tilde{q}_2)$ as $n \to \infty$. The efficiency is at least 93.333%, with the lowest value obtained when $(\tilde{q}_1, \tilde{q}_2) = (3, 5)$.

For an arbitrary even n, i.e., $q_1 = 2$, it holds that $\psi(n) < \psi(n + 1)$, such as $\psi(14) = 6 < \psi(15) = 8$. This implies that for an even n, the number of columns in the LACE with the greatest number of factors exceeds that of the ACE with the greatest number of factors. Moreover, the LACE improves the limiting L_1- and L_2-distance efficiencies from 83.333% (ACE in Section 2.3) to 93.333% for an arbitrary even n. In the context of efficiency and the number of factors, it is advisable to use the ACE for odd values of n, and use the LACE for even values of n.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

4. Design properties: column-orthogonality and projection uniformity

4.1 Column-orthogonality

In variable selection, column-orthogonality is a desirable feature of a two-level supersaturated design since estimates of effects will be uncorrelated. Owen (1994) proposed the mean squared correlation metric to measure the column-orthogonality of a k-factor design D, which is defined as $\rho^2(D) = 2\sum_{j=1}^{k-1}\sum_{l=j+1}^{k}\rho_{jl}^2/(k(k-1))$, where ρ_{jl} is the sample correlation between the jth and lth columns of D. A design D is column-orthogonal if and only if $\rho^2(D) = 0$ $(0 \le \rho^2(D) \le 1)$.

Wang, Sun and Xu (2022) established a connection between ρ^2 and the L_2-distance of a design in their Theorem 2 and Corollary 1. They demonstrated that a larger L_2-distance value corresponds to a smaller ρ^2 value. The ρ^2 reaching its lower bound if and only if the design is equidistant, which corresponds to a maximin distance design. Denote the largest value among all pairwise L_p-distances of a design D by $d_p^{max}(D)$. Based on Theorem 1, we derive the closed forms of $d_p^{max}(\mathcal{E}(D_0))$ and $d_p^{max}(\mathcal{E}(D_1))$ as follows.

Lemma 5. *For $p = 1, 2$, it holds that*

$$d_p^{max}(\mathcal{E}(D_0)) = \begin{cases} \dfrac{n^{p+2}(q_1^2-1)}{2^{p-1}3q_1^2} & \text{if} \quad t = 1, \\ \dfrac{n^{p+2}(q_1+1)\prod_{\ell=1}^{t}(q_\ell-1)}{2^{p-1}3q_1\prod_{\ell=1}^{t}q_\ell} & \text{if} \quad t > 1. \end{cases}$$

Moreover, we have $d_p^{max}(\mathcal{E}(D_1)) = d_p^{max}(\mathcal{E}(D_0))/2$.

Combining Theorems 1–2 with Lemma 5, we obtain the upper bounds of Ξ_2 for

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

$\mathcal{E}(D_0)$ and $\mathcal{E}(D_1)$. Then, we derive the following upper bounds of the ρ^2 for these designs.

Theorem 8. *For D_0 and D_1 as defined earlier, we have*

$$\rho^2(\mathcal{E}(D)) \leq \mathrm{LB}_{\rho^2}(\mathcal{E}(D)) + \frac{n^2\psi(n)(d_2^{max}(\mathcal{E}(D)) - d_{2,\mathrm{ave}}(\mathcal{E}(D)))^2}{(n-1)(n\psi(n)-\Delta)d_{2,\mathrm{ave}}^2(\mathcal{E}(D))},$$

where $\mathrm{LB}_{\rho^2}(D) = \max\{(k+1-n)/(n-1)/(k-1), 0\}$, $\Delta = 1$ *if* $D = D_0$, *and* $\Delta = 2$ *if* $D = D_1$.

When n is a prime, the equal sign holds. The resulting proposed designs $\mathcal{E}(D_0)$ and $\mathcal{E}(D_1)$ are equidistant, ensuring their optimality under the ρ^2 criterion. Moreover, $\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)$ is the optimal LHD under the ρ^2 criterion when $n+1$ is a prime since Theorem 7 proves that $\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)$ is an equidistant design. It is intractable to provide tight upper bounds of the ρ^2 of an arbitrary proposed design except for the designs in Theorem 8. Wang, Sun and Xu (2022) stated that the maximin distance design under the L_2-distance criterion tends to have a small ρ^2. The proposed LHDs often have smaller ρ^2's.

4.2 Projection uniformity

It is well known that good projections to all subspace-fillings of factors are important in computer experiments (Moon, Dean and Santner, 2011; Joseph, Gul and Ba, 2015). Based on the centered L_2-discrepancy, Sun, Wang and Xu (2019) proposed the two-dimensional uniform projection criterion of $D = \left(n, s^k\right)$ as $\phi(D) = 2\sum_{\{\Omega\}} \mathrm{CD}(D_\Omega)/(k(k-1))$ where the (squared) centered L_2-discrepancy is defined as $\mathrm{CD}(D) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

$1/n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \prod_{l=1}^{k} (1 + 1/2\,|z_{il}| + 1/2\,|z_{jl}| - 1/2\,|z_{il} - z_{jl}|) - 2/n \sum_{i=1}^{n} \prod_{l=1}^{k} (1 + 1/2\,|z_{il}| - 1/2\,|z_{il}|^2) + (13/12)^k$, $z_{il} = (2x_{il} - s + 1)/(2s)$, Ω is a two-element subset of $\{1, 2, \ldots, k\}$ and D_Ω is the projected design of D onto dimensions indexed by the elements of Ω. A design is called a uniform projection design if it has the minimum ϕ value. Sun, Wang and Xu (2019) and Wang, Sun and Xu (2022) connected the two-dimensional projection uniformity with the L_1-distance. Wang, Sun and Xu (2022) also established a connection between ϕ and the L_1-distance of a design in their Lemma 4 and Corollary 2.

Combining Theorems 1, 2 and 7, we can obtain the upper bounds of Ξ_1 for $\mathcal{E}(D_0)$, $\mathcal{E}(D_1)$ and $\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)$. Then, we derive the following upper bounds of the ϕ of these designs. The following theorem indicates that $\mathcal{E}(D_0)$ and $\mathcal{E}(D_1)$ are also the optimal LHDs under the ϕ criterion when n is a prime, as well as $\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)$ when $n + 1$ is a prime.

Theorem 9. *For D_0, D_1, $\tilde{D}_0$ and $\mathcal{U}_3$ as defined earlier, we have*

$$\begin{cases} \phi(\mathcal{E}(D))) \le \mathrm{LB}_\phi(\mathcal{E}(D)) + \dfrac{(n+1)^2(n-1)\psi(n)(d_1^{max}(\mathcal{E}(D)) - d_{1,\mathrm{ave}}(\mathcal{E}(D)))^2}{36n^2(n\psi(n) - \Delta)d_{1,\mathrm{ave}}^2(\mathcal{E}(D))}, \\ \phi(\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)) \le \mathrm{LB}_\phi(\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)) + \dfrac{n^2(n+1)(n-2)\psi(n)(d_1^{max}(\mathcal{E}(D_1)) - d_{1,\mathrm{ave}}(\mathcal{E}(D_1)))^2}{36(n-1)^3((n+1)\psi(n) - 2)d_{1,\mathrm{ave}}^2(\mathcal{E}(D_1))}, \end{cases}$$

where $\mathrm{LB}_\phi(D) = (5k(4n^4 + 2(13n - 17)n^2 - n + 5) - (n - 1)(8n^4 + 150n^2 - 33))/(720\, n^4(n - 1)(k - 1)) + (1 + (-1)^n)/(64n^4)$, *$D$ and Δ are defined in Theorem 8.*

The equal sign holds for $\mathcal{E}(D_0)$ and $\mathcal{E}(D_1)$ when n is a prime, and for $\mathcal{L}(\tilde{D}_0, \mathcal{U}_3)$ when $n + 1$ is a prime. These designs exhibit equidistant properties under the L_1-distance, ensuring their optimality under the ϕ criterion. The above uniform projection

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

criterion primarily emphasizes two-dimensional projections. To provide a thorough evaluation of projections beyond two dimensions, Joseph, Gul and Ba (2015) proposed the maximum projection criterion, which is defined as $\tilde{\phi}(D) = \left\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 1/\prod_{l=1}^{k} (x_{il}-x_{jl})^2 / \binom{n}{2}\right\}^{1/k}$. A design D is called a maximum projection design if it has the minimum $\tilde{\phi}(D)$ value. The existing literature can only search for optimal designs under the maximum projection criterion using algorithmic methods, due to the lack of theoretical support to construct them systematically.

In addition to the L_p-distance, a commonly used distance-based criterion for measuring space-filling LHDs is the ϕ_p criterion (Joseph and Hung, 2008). The lower and upper bounds of ϕ_p criterion can be straightforwardly derived from Lemma 1, Theorems 1, 5, 6 and 7. We will utilize the ϕ_1 criterion to conduct comparisons between the proposed algorithms and the existing algorithms in Section 6.

5. The algorithms based on (leave-one-out) column expansions

To further accommodate an arbitrary number of factors (m), a straightforward approach is to apply a stochastic search algorithm to find the best collection of m columns among the $n\psi(n)$ columns within the largest ACE as $\mathcal{E}(D_0)$. This idea can be implemented with the assistance of simulated annealing (SA) proposed by Kirkpatrick, Gelatt and Vecchi (1983). Its pseudo-code is provided in Algorithm 0. Corollary 2 demonstrates that $\mathcal{E}(D_0)$ is a good candidate set of design columns. Not only does it has the maximal number of columns compared to other ACEs, but it also exhibits a high L_p-distance efficiency. The tuning parameters of Algorithm 0 are primarily

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

derived from those used in the Simulated Annealing (SA) algorithm, including the initial temperature T, the decreasing rate: r and the threshold value: δ_1. These parameters have been extensively studied in the literature. Specifically, the initial temperature T is set to approximately $10\Delta E_{\text{avg}}$, where ΔE_{avg} represents the average change in the objective function value when moving between neighboring solutions (Kirkpatrick, Gelatt and Vecchi, 1983; Ben-Ameur, 2004). To balance convergence speed and efficiency, many studies recommend $r = 0.95$ as a standard choice for the SA algorithm (Singh and Baghel, 2021; Blanzeisky and Cunningham, 2022). A common stopping criterion involves halting the algorithm when the temperature falls below a certain threshold (δ_1), typically set to no larger than $10^{-2}, 10^{-3}$ or 10^{-4}. Algorithm 0 is able to accommodate any m not exceeding $n\psi(n)$. An example with $m \leq n$ is provided in Example 2 below.

Algorithm 0: The best collection of m columns from $\mathcal{E}(D_0)$ by simulated annealing

Input: the run size: n, number of factors: m, GLP set: D_0, positive integer: p, initial tuning parameter: T, tuning parameter decreasing rate: $r \in (0, 1)$, threshold value: δ_1;

Output: the design $\mathcal{D}$.

set $D = D_0$ and $\mathcal{D}$ as random m columns among the $n\psi(n)$ columns in $\mathcal{E}(D_0)$;
while $T \geq \delta_1$ **do**
  randomly interchange two columns drawn from $\mathcal{D}$ and $\mathcal{E}(D)\backslash\mathcal{D}$, and denote the updated $\mathcal{D}$ as $\mathcal{D}_{\text{try}}$;
  if $\text{Eff}_p(\mathcal{D}) < \text{Eff}_p(\mathcal{D}_{\text{try}})$ **then**
    set $\pi = 1$;
  else
    set $\pi = \exp\left((\text{Eff}_p(\mathcal{D}_{\text{try}}) - \text{Eff}_p(\mathcal{D}))/T\right)$;
  end
  update design: $\mathcal{D} \leftarrow \mathcal{D}_{\text{try}}$ with probability π;
  update tuning parameter: $T \leftarrow rT$;
end

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Table 2: The L_1-distance efficiencies of the LHDs generated by Algorithms 0 and 0_L.

Algorithm 0								Algorithm 0_L							
n	m	d_1	Eff_1	n	m	d_1	Eff_1	n	m	d_1	Eff_1	n	m	d_1	Eff_1
3	2	2	1	7	4	8	0.8	4	2	3	1	6	3	6	0.857
	3	4	1		5	10	0.769		3	4	0.8		4	8	0.889
5	4	6	0.75		6	13	0.813		4	6	1		5	10	0.909
	5	9	0.9		7	16	0.889						6	13	0.929

Note: d_1 represents the L_1-distance and Eff_1 represents the L_1-distance efficiency.

Example 2. For each n in Example 1, Algorithm 0 is employed to construct some LHDs with $m \leq n$. Their L_1-distances and L_1-distance efficiencies are displayed in the left side of Table 2. The L_1-distance efficiencies of the LHDs generated by Algorithm 0 are higher than that of ACE in Table 1. Furthermore, Algorithm 0 is able to accommodate a wider range of design sizes, including $(n, m) = (3, 3), (5, 5)$, and $(7, 4)$. Additional results for $n < m$ will be provided in Section 6.2.

When the design space is large, Algorithm 0 may not be feasible. To address this issue, we start with a well-chosen initial design. In view of Lemma 2 and Corollary 1, the ACE of either D_0 or D_1 is a suitable choice of the initial design, due to its good space-filling property. Specifically, we add a step in the initial phase of Algorithm 0 to search for a high L_p-distance ACE. This additional step unfolds as follows. An elementary $n \times k$ GLP set D is first selected. Then the SA method is employed to search for the index set $\mathcal{U}$ (and the remaining $m - kv$ columns if $m - kv > 0$), with the constraint $kv \leq m$. In each iteration of SA, one element is randomly chosen from the current $\mathcal{U}$ to exchange with one randomly selected element from $\mathcal{U}_0 \backslash \mathcal{U}$. The new $\mathcal{U}$ is denoted as $\mathcal{U}_{\text{try}}$. If $m - kv > 0$, the remaining $m - kv$ columns are filled with randomly selected columns $\check{D}_{\text{try}}$ from $\mathcal{E}(D, \mathcal{U}_0 \backslash \mathcal{U}_{\text{try}})$. The resulting new design is denoted as

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

$\mathcal{D}_{\text{try}} = (\mathcal{E}(D, \mathcal{U}_{\text{try}}), \check{D}_{\text{try}})$, which is adopted with probability π as specified above. The iteration stops when there is no further improvement of $\mathcal{D}_{\text{try}}$ in the L_p-distance efficiency over several consecutive iterations. Upon iteration completion, the resulting design serves as the initial design for Algorithm 0 if the current tuning parameter remains above a pre-specified threshold. This tuning parameter is transmitted to Algorithm 0 as the initial tuning parameter. This algorithm stops when the tuning parameter falls below a pre-specified threshold.

Corollary 1 indicates that choosing the ACE of D_0 and D_1 is advantageous for generating designs with large L_p-distances. The aforementioned proposed algorithm utilizes D_0, D_1 and D_1' to yield Algorithms 1–3. Specifically, these algorithms individually select ACEs of D_0, ACEs of D_1, and combinations of ACEs of D_1 and D_1' to be the candidate set. The stopping criteria are controlled by the tuning parameters δ_1, δ_2, and κ. A vector $\boldsymbol{v}$ is introduced to record historical L_p-distance efficiencies to continuously check the stopping criterion of the first iteration loop. The details for three specific algorithms based on SA are provided as follows.

In Algorithm 1, D_0 serves as the elementary GLP set to produce the skeleton ACE. The loop in lines 4–16 of the algorithm is designed to identify the optimal $\mathcal{U}$ that works best with D_0, aiming to generate an LHD with high efficiency. The iteration will stop when the variance of the most recent κ L_p-distance efficiencies is smaller than δ_2 or the tuning parameter is below δ_1. The resulting design serves as the initial design for Algorithm 0 if the tuning parameter remains above δ_1. Notably, the recommendation for δ_1 is also applicable to δ_2, as their purposes are similar. Based on empirical evidence,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Algorithm 1: The D_0-based algorithm

Input: the run size: n, number of factors: m, GLP set: D_0, positive integer: p, initial temperature: T, decreasing rate: $r \in (0,1)$, number of consecutive iterations: κ, threshold values: δ_1 and δ_2;

Output: the design $\mathcal{D}$.

set $\mathcal{U}_0 = \{0, \ldots, n-1\}$ and $\boldsymbol{v} = \emptyset$;
set $v = \lfloor m/\psi(n) \rfloor$, $K = m \pmod{\psi(n)}$ and $D = D_0$;
randomly draw a $\mathcal{U} = (u_1, \ldots, u_v) \subseteq \mathcal{U}_0$ and K columns from $\mathcal{E}(D, \mathcal{U}_0 \backslash \mathcal{U})$ to be $\check{D}$, and set $\mathcal{D} = (\mathcal{E}(D, \mathcal{U}), \check{D})$;
while $(T \geq \delta_1)$ *and* $(|\boldsymbol{v}| \leq \kappa$ *or* $\sigma^2 \geq \delta_2)$ **do**
    update historical L_p-distance efficiencies: $\boldsymbol{v} = (\boldsymbol{v}, \mathrm{Eff}_p(\mathcal{D}))$, calculate the variance for the most recent κ L_p-distance efficiencies in $\boldsymbol{v}$ and denote this variance by σ^2;
    randomly interchange two elements drawn from $\mathcal{U}$ and $\mathcal{U}_0 \backslash \mathcal{U}$, and denote the updated $\mathcal{U}$ as $\mathcal{U}_{\mathrm{try}}$;
    randomly draw K columns from $\mathcal{E}(D, \mathcal{U}_0 \backslash \mathcal{U}_{\mathrm{try}})$ to be $\check{D}_{\mathrm{try}}$;
    set $\mathcal{D}_{\mathrm{try}} = (\mathcal{E}(D, \mathcal{U}_{\mathrm{try}}), \check{D}_{\mathrm{try}})$;
    if $\mathrm{Eff}_p(\mathcal{D}) < \mathrm{Eff}_p(\mathcal{D}_{\mathrm{try}})$ **then**
        set $\pi = 1$;
    else
        set $\pi = \exp\left((\mathrm{Eff}_p(\mathcal{D}_{\mathrm{try}}) - \mathrm{Eff}_p(\mathcal{D}))/T\right)$;
    end
    update design and index set: $\mathcal{U} \leftarrow \mathcal{U}_{\mathrm{try}}$ and $\mathcal{D} \leftarrow \mathcal{D}_{\mathrm{try}}$ with probability π;
    update tuning parameter: $T \leftarrow rT$;
end
run lines 2-11 of Algorithm 0;

we recommend setting $\kappa = 10$ in simulations. Algorithm 0 stops when the tuning parameter falls below a pre-specified threshold δ_1. The final output design $\mathcal{D}$ is a (nearly) maximin distance LHD. In Algorithm 2, D_1 serves as the elementary GLP set instead of D_0. For the cases of $m > n\psi(n)/2$, the first $n\psi(n)/2$ columns of the output design are $\mathcal{E}(D_1')$ and the remaining $m - n\psi(n)/2$ columns are selected among columns of $\mathcal{E}(D_1)$ with the same algorithm except for some small adjustments. Algorithm 3 is proposed to study a more flexible framework where ACEs of D_1 and D_1' work together to form an efficient LHD. More discussions and comparisons of Algorithms 1-3 are provided in the supplementary materials.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Algorithm 2: The D_1-based algorithm

Input: the run size: n, number of columns: m, GLP set and its strictly dual: D_1 and D_1', positive integer: p, initial tuning parameter: T, tuning parameter decreasing rate: $r \in (0,1)$, number of consecutive iterations: κ, threshold values: δ_1 and δ_2;

line 2: set $K = m \pmod{(\psi(n)/2)}$, $D = D_1$ and $D' = D_1'$; if $m = n\psi(n)/2$, set $\mathcal{D} = \mathcal{E}(D)$ and stop the algorithm; if $m < n\psi(n)/2$, set $\mathcal{D}_{\text{fix}} = \emptyset$ and $v = \lfloor 2m/\psi(n) \rfloor$; if $m > n\psi(n)/2$, set $\mathcal{D}_{\text{fix}} = \mathcal{E}(D')$ and $v = \lfloor 2m/\psi(n) \rfloor - n$;

line 3: randomly draw a $\mathcal{U} = (u_1, \ldots, u_v) \subseteq \mathcal{U}_0$ and K columns from $\mathcal{E}(D, \mathcal{U}_0 \backslash \mathcal{U})$ to be $\check{D}$, and set $\mathcal{D} = (\mathcal{D}_{\text{fix}}, \mathcal{E}(D, \mathcal{U}), \check{D})$;

line 8: set $\mathcal{D}_{\text{try}} = (\mathcal{D}_{\text{fix}}, \mathcal{E}(D, \mathcal{U}_{\text{try}}), \check{D}_{\text{try}})$;

run lines 1, 4-7, 9-17 as those in Algorithm 1;

Output: the design $\mathcal{D}$.

Algorithm 3: The (D_1, D_1')-based algorithm

Input: the run size: n, number of columns: m, GLP set: D_0, GLP set and its strictly dual: D_1 and D_1', positive integer: p, initial tuning parameter: T, tuning parameter decreasing rate: $r \in (0,1)$, number of consecutive iterations: κ, threshold values: δ_1 and δ_2;

line 2: randomly draw an integer partition as $m = m_1 + m_2$ with non-negative m_1 and m_2, set $v_i = \lfloor 2m_i/\psi(n) \rfloor$, $K_i = m_i \pmod{(\psi(n)/2)}$ for $i = 1, 2$, $D = D_1$ and $D' = D_1'$;

line 3: randomly draw the v_i-element subset of $\mathcal{U}_0$ to be $\mathcal{U}^i$ for $i = 1, 2$, K_1 columns from $\mathcal{E}(D, \mathcal{U}_0 \backslash \mathcal{U}^1)$ to be $\check{D}^1$ and K_2 columns from $\mathcal{E}(D', \mathcal{U}_0 \backslash \mathcal{U}^2)$ to be $\check{D}^2$, and set $\mathcal{D} = (\mathcal{E}(D, \mathcal{U}^1), \mathcal{E}(D', \mathcal{U}^2), \check{D}^1, \check{D}^2)$;

line 6: randomly draw an $i \in \{1, 2\}$, randomly select one of the two operations: (i) randomly interchange two elements drawn from $\mathcal{U}^i$ and $\mathcal{U}_0 \backslash \mathcal{U}^i$, (ii) randomly delete an element of $\mathcal{U}^i$ and add an element from $\mathcal{U}_0 \backslash \mathcal{U}^j$ to $\mathcal{U}^j$ for $j \in \{1, 2\}$ and $j \neq i$, and denote the updated $\mathcal{U}^1$ and $\mathcal{U}^2$ as $\mathcal{U}^1_{\text{try}}$ and $\mathcal{U}^2_{\text{try}}$ respectively;

line 7: randomly draw K_1 columns from $\mathcal{E}(D, \mathcal{U}_0 \backslash \mathcal{U}^1_{\text{try}})$ to be $\check{D}^1_{\text{try}}$, and K_2 columns from $\mathcal{E}(D', \mathcal{U}_0 \backslash \mathcal{U}^2_{\text{try}})$ to be $\check{D}^2_{\text{try}}$;

line 8: set $\mathcal{D}_{\text{try}} = (\mathcal{E}(D, \mathcal{U}^1_{\text{try}}), \mathcal{E}(D', \mathcal{U}^2_{\text{try}}), \check{D}^1_{\text{try}}, \check{D}^2_{\text{try}})$;

line 13: update design and index set: $\mathcal{U}^1 \leftarrow \mathcal{U}^1_{\text{try}}$, $\mathcal{U}^2 \leftarrow \mathcal{U}^2_{\text{try}}$ and $\mathcal{D} \leftarrow \mathcal{D}_{\text{try}}$ with probability π;

run lines 1, 4-5, 9-13, 15-17 as those in Algorithm 1.

Output: the design $\mathcal{D}$.

Besides the SA algorithm, the proposed hybrid approaches can accommodate various stochastic algorithms, such as the genetic algorithm (Mitchell, 1998) and the particle swarm algorithm (Kennedy and Eberhart, 1995). In contrast to directly applying

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

these algorithms for the search of maximin distance designs, Algorithms 0, 1, 2 and 3, leveraging the ACE structure with high L_p-distance efficiency, can yield more efficient designs. Algorithms 1–3 employ a well-chosen initial design via the geometric structure of ACE. Compared to Algorithm 0, Algorithms 1–3 will accelerate the search process. This is particularly important when n is large. These three algorithms are inclined to construct highly efficient designs. Numerical comparisons will be presented in Section 6.2.

For an even n and an arbitrary m, we shall modify Algorithms 0, 1, 2 and 3 by replacing the ACE component therein with the corresponding LACE, and we shall refer to these new algorithms as Algorithms 1_L, 2_L, 3_L and $\mathrm{S1}_L$. Further details on them are deferred to the supplemental materials. They are expected to yield improvements over Algorithms 0, 1, 2 and 3 when n is even.

Example 3. For $n = 4$ and 6, Algorithm 0_L is used to construct LHDs under the L_1-distance with $m \leq n$. Their L_1-distances and L_1-distance efficiencies are displayed in the right side of Table 2. The L_1-distance efficiencies of the LHDs generated by Algorithm 0_L are notably high, ranging from 80% to 100%. More results are presented in Section 6.2.

6. Comparisons with the existing works

6.1 Qualitative comparisons

Existing algebraic methods are applicable to specific configurations of run size (n) and number of factors (m), limiting their flexibility in accommodating arbitrary design

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

sizes. For instance, the construction methods introduced in Xiao and Xu (2017) and Wang, Xiao and Xu (2018) are only applicable when either n or $2n+1$ is a prime and $m \in \{n-1, n\}$. Zhou and Xu (2015) and Vazquez and Xu (2024) proposed maximin distance designs with $m = n(n-1)$ or $m = n(n-1)/2$ where n is a prime. Besides, there is a category of innovative construction methods that expand a space-filling design (B) through level and/or column expansions based on a specific algebraic structure (A) to achieve larger space-filling designs (D). The space-filling property of the input design B ensures that of the resulting design D. However, achieving this goal is challenging since constructing the input design B is only possible for limited run sizes. In contrast, the algebraic component of our method allows an arbitrary n and $m = kv$. The cases of $m \approx \tilde{n}\psi(\tilde{n})$ or $m \approx \tilde{n}\psi(\tilde{n})/2$ have been theoretically studied and their asymptotic optimality has been established.

Algorithmic search methods are only available for small numbers of runs and factors, such as Ba (2013) and Carnell (2022). In contrast, our algorithmic component handles an arbitrary $m \leq \tilde{n}\psi(\tilde{n})$ for all values of n. Numerical studies demonstrate that the proposed algorithms can efficiently yield (nearly) optimal LHDs under the L_1- and L_2-distance criteria.

To accommodate arbitrary values of n and m, one can employ a stochastic algorithmic search to augment $n - n_1$ rows and $m - n_1$ columns for the $n_1 \times n_1$ ($n \geq n_1$, $m \geq n_1$) optimal L_1-distance LHD proposed by Wang, Xiao and Xu (2018). This LHD is limited to the case where $2n_1 + 1$ is a prime. Such a method generally underperform the proposed algorithms. More details are presented in the supplementary materials

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

(Section S3.2).

6.2 Quantitative comparisons

In this section, the proposed method is compared with two of the most commonly used algorithms for constructing arbitrary space-filling LHDs: "SLHD" (Ba, 2013) and "lhs" (Carnell, 2022) in the R-packages. Set $T = 1$, $r = 0.95$, $\kappa = 10$ and $\delta_2 = 10^{-4}$. The quantitative comparisons demonstrate that the proposed algorithms outperform "SLHD" and "lhs" under the space-filling criteria of the L_p-distance, ρ^2, ϕ, ϕ_1 and $\tilde{\phi}$.

For $n = 7$ and $m \in \{2, \ldots, n\psi(n)\} = \{2, \ldots, 42\}$, Algorithms 0, 1, 2 and 3 are compared with "SLHD" and "lhs" under the L_1- and L_2-distance criteria. Three choices of (δ_1, R) as $(3.5 \times 10^{-1}, 50)$, $(7.5 \times 10^{-2}, 200)$ and $(3.5 \times 10^{-5}, 1000)$ are provided where δ_1 is the tuning parameter and R is the number of iterations corresponding to δ_1. For each (m, R), each algorithm is repeated 500 times and the averages of the L_1- and L_2-distance efficiencies are drawn in Figure 1. Specifically, Figures 1 (a)–(c) display the average L_1-distance efficiencies of all algorithms for $m \in \{2, \ldots, 42\}$ with R set to 50, 200, and 1000 iterations, respectively. Figures 1 (d)–(f) depict the L_2-distance efficiency versions of Figures 1 (a)–(c), respectively. As shown in Figure 1 (a)–(f), the L_p-distance efficiencies of the proposed designs via Algorithms 0, 1, 2 and 3 consistently outperform those of the designs generated by "SLHD" and "lhs" across all combinations of R, m, and p. As R increases, the growth of efficiency is slower for "SLHD" and "lhs" than for Algorithms 0, 1, 2 and 3. As m approaches $n\psi(n)/2$, Algorithm 2 significantly enhances its efficiency due to the established optimality around $m = n\psi(n)/2$. This ensures a highly efficient design. From $m = 21$ to 42, the optimality near the boundary

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

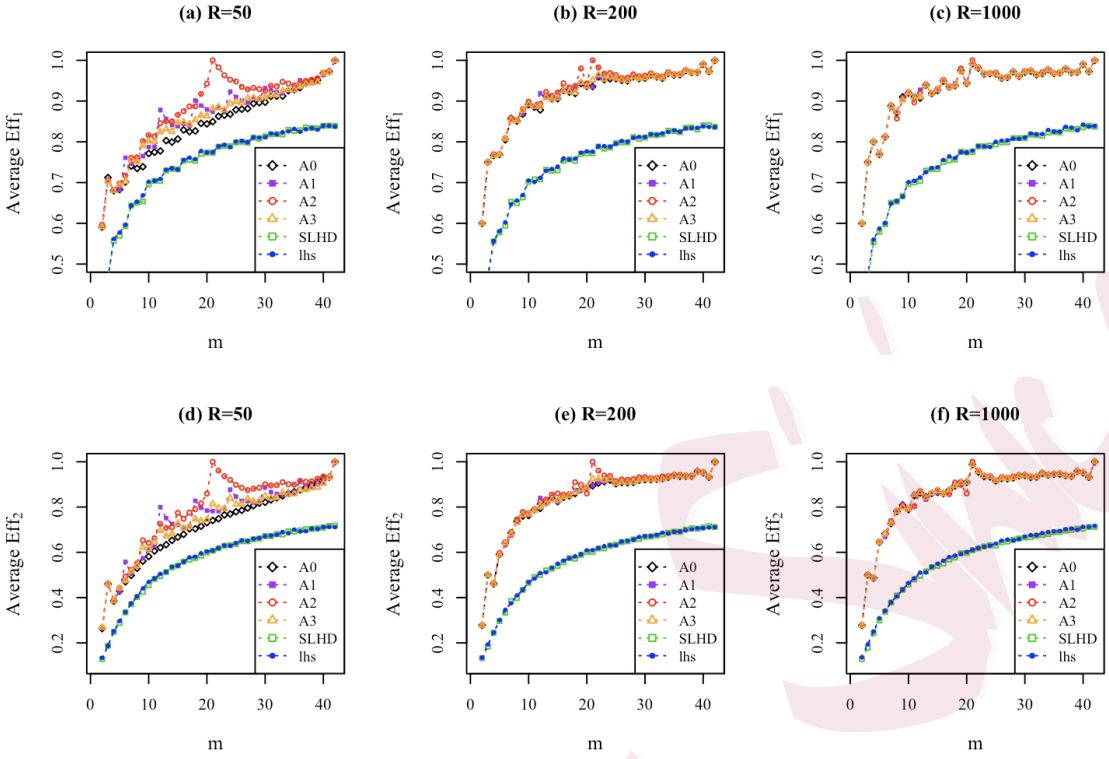

Figure 1: The L_1- and L_2-distance efficiencies under various scenarios: Algorithms 0, 1, 2, 3, "SLHD" and "lhs" in the R-packages (marked as A0, A1, A2, A3, SLHD and lhs in sequence).

values of m ensures high efficiencies. Even when m deviates from these boundaries, the L_1-distance efficiency of the proposed design remains above 90% in Figure 1(a) for $m \geq 35$, and in Figures 1(b)–(c) for $m \geq 19$.

Besides the L_p-distance criterion, four other space-filling criteria are used: ρ^2 (Owen, 1994), ϕ_1 (Joseph and Hung, 2008), $\tilde{\phi}$ (Joseph, Gul and Ba, 2015) and ϕ (Sun, Wang and Xu, 2019) to compare the proposed algorithms with "SLHD" and "lhs". Denote the ρ^2 efficiency of design D by $\text{Eff}_{\rho^2}(D) = \text{LB}_{\rho^2}(D)/\ \rho^2(D)$, the ϕ efficiency of D by $\text{Eff}_{\phi}(D) = \text{LB}_{\phi}(D)/\phi(D)$ and the ϕ_1 efficiency of design D by $\text{Eff}_{\phi_1}(D) = n(n-1)/(2d_{1,ave}(D)\phi_1(D))$. For lack of the theoretical lower bound of $\tilde{\phi}$,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

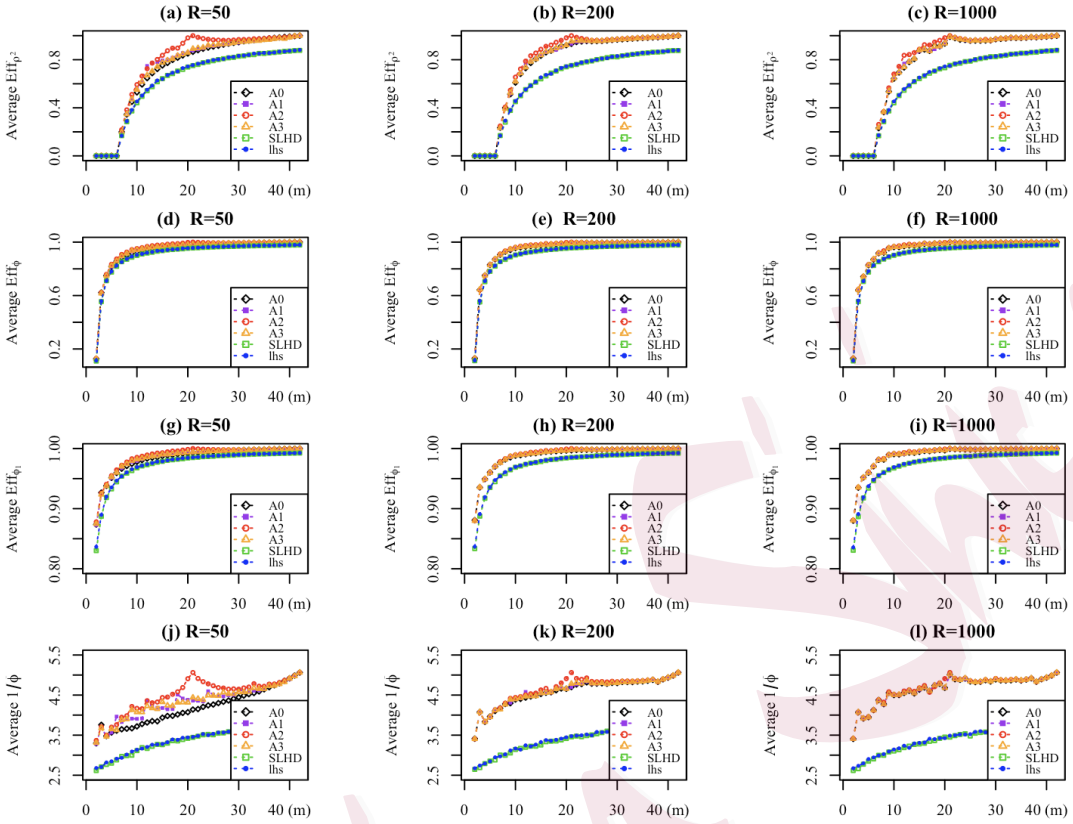

Figure 2: The ρ^2, ϕ, ϕ_1 and $\tilde{\phi}$ efficiencies under various scenarios: Algorithms 0, 1, 2, 3, "SLHD" and "lhs" in the R-packages (marked as A0, A1, A2, A3, SLHD and lhs in sequence).

we utilize its reciprocal $(1/\tilde{\phi})$ to measure the space-filling property. A higher value of $1/\tilde{\phi}$ indicates a better space-filling property. For each configuration of m and R, the variants of Figure 1 based on these criteria are plotted in Figure 2. It turns out that the designs by Algorithms 1–3 often have smaller ρ^2, ϕ, ϕ_1 and $\tilde{\phi}$ compared to those from "SLHD" and "lhs".

Furthermore, we calculate the number of iterations R required for achieving an efficiency not less than $\delta = 0.8$ and 0.9. For each algorithm, we perform 100 repetitions and report $min(R, 200)$, with 200 serving as the cutoff point if the needed R exceeds

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

it. For each algorithm, the average numbers of needed iterations against m are plotted in Figure S3 in the supplementary materials under the L_1- and L_2-distance criteria. The results show that Algorithms 0–3 generally require fewer iterations than "SLHD" and "lhs" for achieving high efficiency (greater than δ) across arbitrary configurations of m, p and δ, particularly Algorithms 1–3.

For $n = 6$, we use LACE as the skeleton for our algorithms since n is even. The tuning parameters keep the same settings as those used for $n = 7$. For $m \in \{2, \ldots, (n+1)\psi(n+1)\} = \{2, \ldots, 42\}$, we plot the averages of the L_1- and L_2-distance, ρ^2, ϕ, ϕ_1 efficiencies and $1/\tilde{\phi}$ against m in Figures S1–S2 in the supplementary materials. Moreover, we also plot the LACE version of Figure S3 in Figure S4 in the supplementary materials. We have a similar observation of the comparison as that for $n = 7$. The computational time for each proposed design shown in Figures 1–2 ranges between 0.2 and 0.3 seconds.

Other values of n categorized based on their prime decomposition $n = q_1^{r_1} \ldots q_t^{r_t}$ are considered. When $t = 1$, four distinct values of n: 5, 8, 9, and 11 are examined. Similarly, for $t = 2$, four different n values: 10, 12, 18, and 22 are examined. Set $\delta_1 = 3.5 \times 10^{-5}$. For each n and any $m \leq n\psi(n)$, each algorithm is repeated 100 times among Algorithms 1–3, "SLHD" and "lhs" to calculate the average L_1-distance of the output design $\mathcal{D}$ as $\bar{d}_1(\mathcal{D})$. For each algorithm and each n, the median along with minimum and maximum (as specified in parentheses) of $\bar{d}_1(\mathcal{D})$'s with $m \in \{2, 3, \ldots, n\psi(n)\}$ are listed in Table 3. It is shown that Algorithms 1–3 have larger L_1-distance efficiencies than "SLHD" and "lhs" when $n \in \{5, 8, 9, 11\}$. When $n \in \{10, 12, 18, 22\}$, Algorithms

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

1–3 have the L_1-distance efficiencies similar to those of "SLHD" and "lhs". For those selected n's that are even, we also repeat each algorithm 100 times among Algorithms 1_L–3_L for $m \leq (n+1)\psi(n+1)$. These results are presented in Table 3. It is shown that Algorithms 1_L–3_L outperform Algorithms 1–3, "SLHD" and "lhs" for each selected even n. This indicates that the LACE indeed improves the non-ideal efficiency based on the ACE for the even n. Moreover, the proposed designs have high L_1-distance efficiencies (exceeding 90%) for at least half of all m values. The computational time for each proposed design presented in Table 3 is under 1 second. For all selected n's, similar comparisons under other criteria are carried out: the L_2-distance, ρ^2, ϕ, ϕ_1 and $\tilde{\phi}$ criteria, which are listed in Tables S1–S5 in the supplementary materials. Under each criterion, we have a similar observation of the comparison as that under L_1-distance criterion in Table 3.

The comparison for large design types under all criteria is extended, specifically for those with $n \in \{60 = 2^2 \times 3 \times 5, 100 = 2^2 \times 5^2, 210 = 2 \times 3 \times 5 \times 7, 1200 = 2^4 \times 3 \times 5^2\}$. Due to the limitations of "lhs" in handling these large design sizes, our comparison is limited to Algorithms 1_L–3_L and "SLHD". The results are displayed in Table 4. It is shown that the designs by Algorithms 1_L–3_L often outperform these by "SLHD" under these criteria. In Table 4, the computational time for each proposed design is under 5 minutes when $n \leq 210$ and is under 2 hours when $n = 1200$.

6.3 Discussions

The proposed designs are intended for scenarios where space-filling properties are beneficial, particularly in tasks such as global surrogate modeling in computer exper-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Table 3: Quantitative comparisons under the L_1-distance criterion.

Method	n							
	$t=1$				$t=2$			
	$r_1=1$		$r_1=2$	$r_1=3$	$r_1=r_2=1$		$r_1=2$ $r_2=1$	$r_1=1$ $r_2=2$
	5	11	9	8	10	22	12	18
A1	**0.933** **(0.667, 1)**	0.940 **(0.5, 1)**	0.875 **(0.5, 0.933)**	0.859 (**0.556**, 0.917)	0.759 (0.455, 0.822)	0.858 (**0.400**, 0.913)	0.707 (0.462, 0.769)	0.741 (**0.474**, 0.789)
$A1_L$	–	–	–	0.877 (0.5, 0.900)	0.945 (**0.571**, 0.974)	0.953 (0.388, 0.982)	0.945 (0.587, 0.976)	0.949 (0.417, 0.980)
A2	**0.933** **(0.667, 1)**	**0.964** **(0.5, 1)**	**0.893** **(0.5, 0.933)**	0.860 (**0.556**, 0.917)	0.761 (0.455, 0.822)	0.883 (**0.400**, 0.913)	0.728 (0.462, 0.769)	0.758 (**0.474**, 0.789)
$A2_L$	–	–	–	**0.886** (0.5, **0.933**)	**0.968** **(0.571, 1)**	**0.976** (0.389, **1**)	**0.970** (**0.598, 1**)	**0.974** (0.417, **1**)
A3	0.930 **(0.667, 1)**	0.940 **(0.5, 1)**	0.877 **(0.5, 0.933)**	0.859 (**0.556**, 0.917)	0.751 (0.455, 0.822)	0.856 (**0.400**, 0.913)	0.712 (0.462, 0.769)	0.740 (**0.474**, 0.789)
$A3_L$	–	–	–	0.878 (0.5, 0.900)	0.944 (**0.571**, 0.975)	0.951 (0.385, 0.982)	0.943 (0.585, 0.977)	0.948 (0.417, 0.980)
SLHD $m \le n\psi(n)$	0.772 (0.512, 0.830)	0.816 (0.261, 0.871)	0.773 (0.357, 0.836)	0.717 (0.350, 0.801)	0.718 (0.303, 0.800)	0.828 (0.139, 0.879)	0.720 (0.275, 0.798)	0.777 (0.178, 0.842)
SLHD $m \le (n+1)\psi(n+1)$	–	–	–	0.78 (0.357, 0.845)	0.827 (0.310, 0.878)	0.886 (0.143, 0.920)	0.840 (0.270, 0.886)	0.871 (0.178, 0.909)
lhs $m \le n\psi(n)$	0.770 (0.540, 0.835)	0.819 (0.260, 0.868)	0.769 (0.352, 0.834)	0.720 (0.358, 0.798)	0.716 (0.306, 0.799)	0.830 (0.144, 0.878)	0.720 (0.268, 0.800)	0.777 (0.178, 0.838)
lhs $m \le (n+1)\psi(n+1)$	–	–	–	0.78 (0.347, 0.846)	0.825 (0.307, 0.878)	0.886 (0.143, 0.921)	0.837 (0.259, 0.888)	0.872 (0.178, 0.912)

Note: Algorithms 1, 1_L, ..., 3, 3_L, "SLHD" and "lhs" in the R-packages (marked as A1, $A1_L$, ...,A3, $A3_L$, SLHD and lhs in sequence). For each case, the listed data is the median along with minimum and maximum (as specified in parentheses) of $\bar{d}_1(\mathcal{D})$'s with the numbers of columns of $\mathcal{D}$'s being either $2, \ldots, n\psi(n)$ or $2, \ldots, (n+1)\psi(n+1)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Table 4: Quantitative comparisons with large values of n under all criteria.

n	m	Method	Eff_1	Eff_2	Eff_{ρ^2}	Eff_{ϕ}	Eff_{ϕ_1}	$1/\phi$	n	m	Method	Eff_1	Eff_2	Eff_{ρ^2}	Eff_{ϕ}	Eff_{ϕ_1}	$1/\phi$
60	50	$A1_L$	0.741	**0.551**	0	0.588	0.991	117.333	60	600	$A1_L$	0.949	0.901	0.931	0.985	**1.000**	**185.948**
		$A2_L$	0.682	0.499	0	0.562	0.989	104.977			$A2_L$	**0.950**	**0.920**	0.931	**0.987**	**1.000**	185.944
		$A3_L$	**0.776**	0.545	0	**0.683**	**0.994**	**138.227**			$A3_L$	0.925	0.867	**0.950**	0.981	**1.000**	179.421
		SLHD	0.686	0.489	0	0.580	0.990	75.221			SLHD	0.911	0.827	0.898	0.964	0.999	165.707
	60	$A1_L$	**0.915**	0.823	0.037	**0.861**	**0.998**	**186.114**		1800	$A1_L$	0.975	0.925	0.983	0.996	**1.000**	197.452
		$A2_L$	0.818	**0.832**	**0.041**	0.793	0.996	150.908			$A2_L$	**0.990**	**0.985**	**0.999**	**1.000**	**1.000**	**203.463**
		$A3_L$	0.795	0.696	0.025	0.783	0.995	147.765			$A3_L$	0.964	0.951	0.990	0.995	**1.000**	193.770
		SLHD	0.721	0.551	0.017	0.656	0.993	100.452			SLHD	0.951	0.916	0.969	0.989	**1.000**	181.130
100	100	$A1_L$	0.857	0.748	**0.017**	0.768	0.998	417.122	210	100	$A1_L$	0.804	**0.660**	0	0.203	0.995	**1495.290**
		$A2_L$	0.830	0.727	0.015	0.748	0.997	384.150			$A2_L$	0.794	0.650	0	0.200	0.995	1331.878
		$A3_L$	**0.885**	**0.749**	0.015	**0.840**	**0.999**	**471.617**			$A3_L$	**0.815**	0.649	0	**0.216**	**0.996**	1482.491
		SLHD	0.719	0.563	0.010	0.628	0.995	276.106			SLHD	0.738	0.578	0	0.187	0.995	1030.483
	1000	$A1_L$	**0.962**	0.908	0.943	**0.986**	**1.000**	**504.381**		200	$A1_L$	0.857	**0.760**	0	0.618	**0.998**	1597.577
		$A2_L$	0.915	**0.935**	**0.953**	0.975	**1.000**	477.036			$A2_L$	0.851	0.746	0	**0.635**	**0.998**	1537.790
		$A3_L$	0.939	0.911	0.937	0.984	**1.000**	484.772			$A3_L$	**0.861**	0.741	0	**0.635**	**0.998**	**1695.839**
		SLHD	0.920	0.864	0.901	0.964	**1.000**	430.906			SLHD	0.769	0.689	0	0.587	0.997	1432.261
210	500	$A1_L$	0.912	**0.877**	0.674	0.903	0.999	1842.205	1200	600	$A1_L$	0.855	**0.793**	0	0.204	**0.999**	51353.750
		$A2_L$	0.916	0.872	0.688	0.896	0.999	**1956.835**			$A2_L$	**0.873**	0.787	0	0.202	**0.999**	**53147.620**
		$A3_L$	**0.927**	0.856	**0.705**	**0.909**	**1.000**	1929.198			$A3_L$	0.823	0.618	0	**0.259**	**0.999**	50840.490
		SLHD	0.878	0.808	0.583	0.840	0.999	1644.199			SLHD	0.849	0.757	0	0.206	**0.999**	46063.140
	1000	$A1_L$	0.930	0.916	**0.900**	0.940	**1.000**	2009.328		2000	$A1_L$	0.927	**0.883**	**0.479**	0.830	**1.000**	61247.193
		$A2_L$	**0.946**	**0.922**	0.875	**0.960**	**1.000**	**2018.350**			$A2_L$	**0.936**	0.877	0.469	0.842	**1.000**	**62895.417**
		$A3_L$	0.929	0.902	0.816	0.948	**1.000**	1975.750			$A3_L$	0.912	0.843	0.475	**0.850**	**1.000**	60017.242
		SLHD	0.911	0.846	0.791	0.919	**1.000**	1769.360			SLHD	0.910	0.860	0.401	0.762	**1.000**	57457.881

Note: Algorithms 1_L, 2_L, 3_L and R-package "SLHD" (marked as $A1_L$, $A2_L$, $A3_L$ and SLHD in sequence). For the case of $(n, m) = (60, 50), (210, 100), (210, 200)$ or $(1200, 600)$, Eff_{ρ^2} is equal to 0 since the lower bounds is equal to 0.

iments (Johnson, Moore and Ylvisaker, 1990; Zhou and Xu, 2015), as well as initial sampling in active learning frameworks (Crombecq et al., 2009; Zhang et al., 2021) and multi-start methods for global optimization (Regis and Shoemaker, 2013; Yu et al., 2019). When $n \geq m$, space-filling LHDs are widely utilized in computer experiments. Johnson, Moore and Ylvisaker (1990) demonstrated that maximin distance designs are asymptotically optimal for fitting Gaussian process models within a Bayesian framework. The proposed hybrid method enables the efficient construction of these designs. Regardless of whether $n \geq m$ or $n < m$, the proposed method is a beneficial tool for constructing larger space-filling designs. As discussed in Section 6.1, these construction techniques require flexible space-filling designs (B) to achieve larger and flexible space-filling designs (D). The proposed designs precisely meet the critical requirements for the input design (B). For instance, Li, Liu and Tang (2021) constructed a maximin distance U-type design (D) by replacing the u^{th} level of a U-type design (A) with a

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

large Hamming distance by the u^{th} row of a U-type design (B) for $u = 1, \ldots, s$. Let A be a 144×7 orthogonal array with 12 levels as shown in the R-package "DoE.base", with a Hamming distance efficiency of 92.875%. The 12×12 LACE of $\tilde{D}_1$ on $\mathcal{U}$ is the maximal L_1-distance design, where $\mathcal{U} = \{1, 12\}$. This proposed design can serve as B, enabling the resulting design D as a 144×84 U-type design. It is shown that D exhibits the L_1-distance and L_2-distance efficiencies of 92.875% and 89.881%, respectively. If $\mathcal{U} = \{1, 12\}$ is replaced by $\mathcal{U} = \{0, 1, 12\}$, the proposed design B becomes a 12×18 LACE with the L_1-distance efficiency of 93.590%. Consequently, the resulting design D is a 144×126 U-type design with the L_1-distance efficiency of 86.904%.

A space-filling LHD stands out as an excellent choice for initial points in multi-start methods for global optimization (Regis and Shoemaker, 2013; Yu et al., 2019), especially beneficial for multimodal optimization. To avoid local optima, diverse starting points are needed to explore the landscape and identify what appears to be the global optimum. Compared with random starting points, a space-filling LHD often offers multi-start methods greater chances of reaching global optima, as it distributes more uniformly in space and exhibits superior projection uniformity. When the dimension (m) is high, the proposed designs with $n < m$ are more economical in practice. More results are provided in the supplementary materials (Section S3.3).

Furthermore, space-filling LHDs have been recommended for applications in variable selection, especially when $n < m$. Butler (2005) mentioned supersaturated LHDs in variable selection. Chien, Deng and Lin (2022) demonstrated that space-filling LHDs significantly enhance the variable selection accuracy of Lasso regression. The proposed

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

hybrid method is able to efficiently construct these supersaturated space-filling LHDs. For predictive accuracy, space-filling designs are advantageous in computer experiments or surrogate modeling, as they uniformly explore the input parameter space and minimize bias in response surface estimation. More results are provided in the supplementary materials (Section S3.7).

7. Concluding Remarks

The proposed hybrid method efficiently generates space-filling LHDs to accommodate any configuration (n, m) with $m \leq \tilde{n}\psi(\tilde{n})$ where $\psi(\cdot)$ is the Euler function, and $\tilde{n} = n$ for an odd n and $\tilde{n} = n + 1$ for an even n. The algebraic component, ACE (LACE), is identified as a high-quality candidate set for constructing space-filling LHDs. This paper provides a solid theoretical foundation for this choice, expanding the scope of run sizes outlined in Vazquez and Xu (2024) from prime numbers to all integers. We recommend ACE for all odd values of n and LACE for all even values of n, as they provide high efficiencies in terms of L_1- or L_2-distance and provide greater flexibility in the number of columns. This candidate set ensures that the L_1- and L_2-distance efficiencies are both greater than 93.333%. These theoretical findings validate the feasibility of the proposed method in accommodating any run size. To accommodate an arbitrary number of factors, the proposed algorithm, leveraging the ACE or LACE structure with high L_p-distance efficiency, enables the acceleration of the search process and the generation of highly efficient designs. Furthermore, theoretical and simulation-based evidence shows that the proposed LHDs also perform well on criteria

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

based on column-orthogonality, projection uniformity, and ϕ_p.

In addition to the maximin distance criterion, uniformity serves as another key criterion for space-filling designs. It would be interesting to study the space-filling properties of the proposed designs under discrepancy criteria in future studies.

Supplementary Material

The online Supplementary Material includes detailed proofs, algorithms based on the leave-one-out additive column expansion, quantitative comparisons between the proposed method and existing methods based on criteria such as the L_p-distance, column-orthogonality, projection uniformity and computational costs.

Acknowledgments

Zhang was supported by Fundamental Research Funds for the Central Universities Grant FRF-TP-25-044. Lin was supported by National Science Foundation via Grant DMS-18102925.

References

Ba, S. (2013). *SLHD: Maximin-distance (sliced) Latin hypercube designs*. R package version 2.1.1.

Ben-Ameur, W. (2004). Computing the initial temperature of simulated annealing. *Computational Optimization and Applications* **29**, 369–385.

Blanzeisky, W. and P. Cunningham (2022). Using Pareto simulated annealing to address algorithmic bias in machine learning. *The Knowledge Engineering Review* **37**, e5. Cambridge University Press.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Butler, N. A. (2005). Supersaturated Latin hypercube designs. *Communications in Statistics–Theory and Methods* **34**, 417–428.

Carnell, R. (2022). *lhs: Latin hypercube samples*. R package version 1.1.4.

Chien, P., X. W. Deng and C. D. Lin (2022). Efficient experimental design for Lasso regression. In *Advances and Innovations in Statistics and Data Science* (Edited by W. Q. He, L. Q. Wang, J. H. Chen and C. D. Lin), 295–309. Cham Springer International Publishing.

Crombecq, K., I. Couckuyt, D. Gorissen and T. Dhaene (2009). Space-filling sequential design strategies for adaptive surrogate modeling. In: *Proceedings of the First International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering* (Edited by B.H.V. Topping and Y. Tsompanakis), 1–20. Civil-Comp Press.

Fang, K. T., M.-Q. Liu, H. Qin and Y. D. Zhou (2018). *Theory and application of uniform experimental designs*. Springer.

Fang, K. T. and Y. Wang (1981). A note on uniform distribution and experiment design. *Chinese Science Bulletin* **26**, 485–489.

Hua, L. K. and Y. Wang (1981). *Applications of number theory to numerical analysis*. Springer.

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

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

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

Kennedy, J. and R. Eberhart (1995). Particle swarm optimization. In *Proceedings of ICNN'95-international*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

conference on neural networks **4**, 1942–1948. IEEE.

Kirkpatrick, S., C. D. Gelatt Jr and M. P. Vecchi (1983). Optimization by simulated annealing. *Science* **220**, 671–680.

Korobov, A. (1959). The approximate computation of multiple integrals. *In Doklady Akademii Nauk SSSR* **124**, 1207–1210.

Li, W. L., M.-Q. Liu and B. Tang (2021). A method of constructing maximin distance designs. *Biometrika* **108**, 845–855.

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

Lin, D. K. J. (1993). A new class of supersaturated designs. *Technometrics* **35**, 28–31.

McKay, D., R. Beckman and W. Conovcr (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.

Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT Press.

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

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

Regis, R. G. and C. A. Shoemaker (2013). A quasi-multistart framework for global optimization of expensive functions using response surface models. *Journal of Global Optimization* **56**, 1719–1753.

Singh, R. B. and A. S. Baghel (2021). IC floorplanning optimization using simulated annealing with order-based representation. *International Journal of Intelligent Systems & Applications* **13**(2), 62–70.

Sun, F. S., Y. P. Wang and H. Xu (2019). Uniform projection designs. *The Annals of Statistics* **47**, 641–661.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Tian, Y. and H. Xu (2022). A minimum aberration-type criterion for selecting space-filling designs. *Biometrika* **109**, 489–501.

Vazquez, A. R. and H. Xu (2024). An integer programming algorithm for constructing maximin distance designs from good lattice point sets. *Statistica Sinica* **4**, 1–40.

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

Wang, Y. P., F. S. Sun and H. Xu (2022). On design orthogonality, maximin distance, and projection uniformity for computer experiments. *Journal of the American Statistical Association* **117**, 375–385.

Weyl, H. (1916). Über die gleichverteilung von zahlen mod. eins. *Mathematische Annalen* **77**, 313–352.

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

Yu, H., Y. Tan, C. Sun and J. Zeng (2019). A generation-based optimal restart strategy for surrogate-assisted social learning particle swarm optimization. *Knowledge-Based Systems* **163**, 14–25.

Zhang, Y., S. Wang, C. Zhou, L. Lv and X. Song (2021). A fast active learning method in design of experiments: multipeak parallel adaptive infilling strategy based on expected improvement. *Structural and Multidisciplinary Optimization* **64**, 1259–1284.

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

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

School of Mathematics and Physics, University of Science and Technology Beijing, Beijing, BJ 100083, China

Email: xueruzhang@ustb.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0145

Department of Statistics, Purdue University, West Lafayette, IN 47907, U.S.A

E-mail: dkjlin@purdue.edu

Department of Business Analytics and Statistics, University of Tennessee, Knoxville, TN 37996, U.S.A

E-mail: wzheng9@utk.edu