Journal of Statistical Planning and Inference 215 (2021) 184–192

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Column-orthogonal nearly strong orthogonal arrays

Wenlong Li, Min-Qian Liu, Jian-Feng Yang *

School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin 300071, China

ARTICLE INFO

Article history:
Received 11 August 2019
Received in revised form 23 September 2020
Accepted 28 December 2020
Available online 19 March 2021

MSC:
primary 62K15
secondary 62K99

Keywords:
Computer experiment
Stratification
Space-filling property

ABSTRACT

Strong orthogonal arrays enjoy more attractive space-filling properties than ordinary orthogonal arrays for computer experiments. In this paper, we propose two methods for constructing column-orthogonal nearly strong orthogonal arrays. These designs enjoy column orthogonality, inherit the attractive two-dimensional space-filling property of strong orthogonal arrays, and can accommodate twice or more number of factors than the existing strong orthogonal arrays. In addition, the proposed designs with four levels enjoy an attractive space-filling property under the maximin distance criterion. The construction methods are convenient and flexible, and the resulting designs are useful in computer experiments.

1. Introduction

Computer experiment is a popular and powerful tool to investigate the complex phenomena and systems in engineering and sciences. The implementation of computer experiments needs space-filling designs (Santner et al., 2013; Fang et al., 2006). A commonly used approach for constructing space-filling designs is to adopt orthogonal arrays and similar structures. Randomized orthogonal arrays (Owen, 1992) and orthogonal array-based Latin hypercubes (Tang, 1993) employ orthogonal arrays of strength t to realize t-dimensional space-filling property. Motivated by (t, m, s)-nets from quasi-Monte Carlo (Niederreiter, 1992, Chap. 4), He and Tang (2013) introduced the concept of strong orthogonal arrays (SOAs) and found their attractive space-filling property for computer experiments. These arrays of strength t are more space-filling than ordinary orthogonal arrays in less than t dimensions and have the same space-filling property as the latter ones in t dimensions. However, SOAs, to enjoy more attractive space-filling property than orthogonal arrays, must have strength three or higher. He and Tang (2014) examined the characterizations of SOAs of strength 3. Given the number of runs, the number of factors for an SOA of strength 3 is very small because such arrays are based on orthogonal arrays of strength 3. In order to increase the number of factors and retain the two-dimensional space-filling property of SOAs of strength 3, He et al. (2018) proposed a new class of arrays, called SOAs of strength 2+.

Obviously, both the SOAs of strength 3 in He and Tang (2014) and SOAs of strength 2+ in He et al. (2018) have no column orthogonality. Nevertheless, the column orthogonality is of great importance. Joseph and Hung (2008) argued that minimizing the correlations among the columns will help in estimating the linear trends efficiently when the universal kriging model with linear trends is considered. Furthermore, the column orthogonality, viewed as a stepping stone, helps finding space-filling designs when Gaussian process models are considered (Bingham et al., 2009). Many researchers have discussed the column orthogonality of Latin hypercube designs including Ye (1998), Steinberg and Lin (2006), Sun et al. (2009), Lin et al. (2009), and so on. Liu and Liu (2015) constructed column-orthogonal strong orthogonal arrays (OSOAs)

* Corresponding author.
E-mail address: jfyang@nankai.edu.cn (J.-F. Yang).

https://doi.org/10.1016/j.jspi.2020.12.005

of strength t based on orthogonal arrays of strength t while the numbers of factors are still very small. Zhou and Tang (2019) further examined OSOAs of strength 2+ with relatively more factors.

In this paper, we propose column-orthogonal nearly strong orthogonal arrays (ONSOAs) to accommodate much more factors than the existing ones. The resulting designs have the column orthogonality, attractive stratifications in two dimensions and flexible run sizes with very high factor-to-run ratios, almost equal to that of the orthogonal arrays. Such stratifications, as shown in Definition 1, are a type of space-filling property. In addition, the proposed ONSOAs with four levels enjoy an attractive space-filling property under the maximin distance criterion. The construction methods involve two key ideas. The first is that the specific three-column submatrices form orthogonal arrays of strength three. The second, due to Steinberg and Lin (2006) and Lin et al. (2009), is that rotating the points in an orthogonal array will preserve the column orthogonality of the original orthogonal array.

Compared to stratifications in Definition 1, any two columns of an orthogonal Latin hypercube design from Pang et al. (2009) only achieve a stratification on an $s \times s$ grid, and their designs have a severe restriction of run sizes. Many pairwise columns of orthogonal Latin hypercube designs from Lin et al. (2009) cannot achieve such stratifications. Thus the above two works can be regarded as a motivation for constructing column-orthogonal designs with good stratifications. The proposed ONSOAs with many levels, not requiring the number of levels for each factor to be the same as the number of runs, enjoy good stratifications as shown in Definition 1. In addition, Bingham et al. (2009) discussed the rationale and usefulness for constructing exactly or nearly column-orthogonal designs that are not Latin hypercube designs but still have many levels.

The remainder of this paper is organized as follows. Section 2 provides some preliminaries and defines the ONSOA. Section 3 investigates two construction methods for the ONSOAs. Section 4 is devoted to some comparisons between the ONSOAs and other types of SOAs. Section 5 examines the space-filling property of the constructed ONSOAs under the maximin distance criterion. Section 6 contains some concluding remarks.

2. Definitions and preliminaries

An $n \times m$ matrix with entries from $\{0, 1, \dots, s_j - 1\}$ in the jth column is an orthogonal array of n runs, m factors and strength t if, in any $n \times t$ subarray, all possible level combinations occur equally often. We denote such an array by $\mathrm{OA}(n, m, s_1 \times \cdots \times s_m, t)$. The array is symmetric if $s_1 = \cdots = s_m = s$, simply denoted by $\mathrm{OA}(n, m, s, t)$, and asymmetric otherwise. An $n \times m$ matrix with entries from $\{0, 1, \dots, s^t - 1\}$ is called a strong orthogonal array of n runs, m factors, s^t levels and strength t if any subarray of g columns for any g with $1 \le g \le t$ can be collapsed into an $\mathrm{OA}(n, g, s^{u_1} \times \cdots \times s^{u_g}, g)$ for any positive integers $u_1, \dots, u_g$ with $u_1 + \cdots + u_g = t$, where collapsing s^t levels into s^{u_j} levels is according to $\lfloor x/s^{t-u_j} \rfloor$ for $x = 0, 1, \dots, s^t - 1$, and $\lfloor x \rfloor$ is the largest integer not exceeding x. Denote this array by $\mathrm{SOA}(n, m, s^t, t)$. As a consequence, any $\mathrm{SOA}(n, m, s^3, 3)$ can achieve stratifications on $s^2 \times s$ and $s \times s^2$ grids in two dimensions and on $s \times s \times s$ grids in three dimensions. See He and Tang (2014) for more details of SOAs.

An $n \times m$ matrix with entries from $\{0, 1, \dots, s^2 - 1\}$ is called a strong orthogonal array of strength 2+ with n runs and m factors of s^2 levels, denoted by $\mathrm{SOA}(n, m, s^2, 2+)$, if any subarray of two columns can be collapsed into an $\mathrm{OA}(n, 2, s^2 \times s, 2)$ and an $\mathrm{OA}(n, 2, s \times s^2, 2)$. An $\mathrm{SOA}(n, m, s^2, 2+)$ enjoys the same attractive two-dimensional space-filling property as that of an $\mathrm{SOA}(n, m, s^3, 3)$, while the former can accommodate more factors.

A design D is called column-orthogonal if the inner product of any two columns of the centered design is zero. Centering a design means that the s levels are converted into $x - (s - 1)/2$ for $x = 0, 1, \dots, s - 1$, and then labeled as in the set $\Omega(s) = \{-(s - 1)/2, -(s - 3)/2, \dots, (s - 3)/2, (s - 1)/2\}$. For example, the levels are $-1/2, 1/2$ if $s = 2$ and $-1, 0, 1$ if $s = 3$. We denote a column-orthogonal $\mathrm{SOA}(n, m, s^2, 2+)$ by $\mathrm{OSOA}(n, m, s^2, 2+)$.

Definition 1. A column-orthogonal nearly strong orthogonal array $\mathrm{ONSOA}(R, \sum_{j=1}^m c_j, s^2, 2+)$ is an $R \times C$ array with s^2 levels in $\Omega(s^2)$, whose $C = \sum_{j=1}^m c_j$ columns can be partitioned into m disjoint groups of $c_1, \dots, c_m$ columns respectively such that

(a) the whole array is column-orthogonal;
(b) any two columns from different groups achieve stratifications on $s^2 \times s$ and $s \times s^2$ grids;
(c) any two columns from the same group achieve a stratification on an $s \times s$ grid.

By Definition 1(a) and (c), an $\mathrm{ONSOA}(R, \sum_{j=1}^m c_j, s^2, 2+)$ is column-orthogonal and can be collapsed into an $\mathrm{OA}(R, \sum_{j=1}^m c_j, s, 2)$. In particular, if $c_j = c$ for each j, then an ONSOA in Definition 1 will be denoted by $\mathrm{ONSOA}(R, c \times m, s^2, 2+)$. Obviously, this array becomes an OSOA when $c = 1$.

From Definition 1(b), in an $\mathrm{ONSOA}(R, \sum_{j=1}^m c_j, s^2, 2+)$, each of the c_j columns in the jth group together with any column in other groups as a whole achieves the stratifications on $s^2 \times s$ and $s \times s^2$ grids. This implies the following measure of the degree of such stratifications among columns

$$\pi = \left(C^2 - \sum_{j=1}^m c_j^2 \right) \Big/ (C(C - 1)) .$$

Table 1
An ONSOA$(16, 2 \times 7, 4, 2+)$ multiplied by 2.

D_1		D_2		D_3		D_4		D_5		D_6		D_7	
3	1	3	1	3	1	3	1	3	1	3	1	3	1
3	1	3	1	−3	−1	3	1	−3	−1	−3	−1	−3	−1
3	1	−3	−1	3	1	−3	−1	3	1	−3	−1	−3	−1
3	1	−3	−1	−3	−1	−3	−1	−3	−1	3	1	3	1
−3	−1	3	1	3	1	−3	−1	−3	−1	3	1	−3	−1
−3	−1	3	1	−3	−1	−3	−1	3	1	−3	−1	3	1
−3	−1	−3	−1	3	1	3	1	−3	−1	−3	−1	3	1
−3	−1	−3	−1	−3	−1	3	1	3	1	3	1	−3	−1
1	−3	1	−3	1	−3	1	−3	1	−3	1	−3	1	−3
1	−3	1	−3	−1	3	1	−3	−1	3	−1	3	−1	3
1	−3	−1	3	1	−3	−1	3	1	−3	−1	3	−1	3
1	−3	−1	3	−1	3	−1	3	−1	3	1	−3	1	−3
−1	3	1	−3	1	−3	−1	3	−1	3	1	−3	−1	3
−1	3	1	−3	−1	3	−1	3	1	−3	−1	3	1	−3
−1	3	−1	3	1	−3	1	−3	−1	3	−1	3	1	−3
−1	3	−1	3	−1	3	1	−3	1	−3	1	−3	−1	3

When $c_j = c$ for any j, we have that

$$\pi = (m-1)c/(\mathrm{cm}-1), \tag{1}$$

which is very high and close to 1 as m gets larger. If $c = 2$, we have $\pi = 92.31\%, 96.55\%, 98.36\%, 99.20\%$ for $m = 7, 15, 31, 63$. This implies that the property of stratifications in two dimensions is almost the same as the SOA of strength 2+. An illustrative example is given below.

Example 1. Table 1 shows an ONSOA$(16, 2 \times 7, 4, 2+)$. The whole design is column-orthogonal. All 14 columns are partitioned into 7 disjoint groups (denoted by $D_1, \ldots, D_7$) of two columns each such that any two columns from distinct groups achieve stratifications on 4×2 and 2×4 grids and any two columns from the same group achieve a stratification on a 2×2 grid. With $m = 7$ in (1), we have the degree of stratifications on 4×2 and 2×4 grids of $\pi = 92.31\%$. This implies that the ONSOA$(16, 2 \times 7, 4, 2+)$ enjoys almost the same two-dimensional space-filling property as those of an SOA$(16, 7, 8, 3)$ in He and Tang (2014), an OSOA$(16, 4, 8, 3)$ in Liu and Liu (2015) and an OSOA$(16, 7, 4, 2+)$ in Zhou and Tang (2019), while the former has much more factors.

3. Construction methods

3.1. Column-orthogonal nearly strong orthogonal arrays with sn runs and 2m factors

Throughout this paper, suppose A is an OA$(n, m, s, 2)$. The following algorithm gives the construction of an ONSOA$(sn, 2 \times m, s^2, 2+)$.

Construction 1.

Step 1. Define two $sn \times m$ matrices $F_1 = (A^T, \ldots, A^T)^T$, $F_2 = (A^T, A^T+1, \ldots, A^T+s-1)^T$ (mod s) and write $F_i = (f_{i1}, \ldots, f_{im})$ for $i = 1, 2$, where ^{T} denotes the transpose of a matrix.

Step 2. Center the s levels of two matrices F_1 and F_2 by $x-(s-1)/2$ for $x \in \{0, 1, \ldots, s-1\}$ such that they are from $\Omega(s)$. Then arrange the $2m$ centered columns f_{ij}'s as $(F^{(1)}, \ldots, F^{(m)})$, where $F^{(j)} = (f_{1j}, f_{2j})$ has two columns for $j = 1, \ldots, m$.

Step 3. Define

$$D = (D_1, \ldots, D_m), \tag{2}$$

where $D_j = F^{(j)}V$ for $j = 1, \ldots, m$, and

$$V = \begin{pmatrix} s & -1 \\ 1 & s \end{pmatrix}. \tag{3}$$

A theoretical property of the proposed design D in (2) can be stated as follows.

Theorem 1. *If there exists an* OA$(n, m, s, 2)$*, then the design D in (2) is an* ONSOA$(sn, 2 \times m, s^2, 2+)$.

Example 2. Let A be an OA$(8, 7, 2, 2)$, $F_1 = (A^T, A^T)^T$ and $F_2 = (A^T, A^T+1)^T$ (mod 2). Write $F_i = (f_{i1}, \ldots, f_{i7})$ for $i = 1, 2$. Center the levels of F_1 and F_2 to be $\{-1/2, 1/2\}$. Next, arrange the 14 centered columns f_{ij}'s as $(F^{(1)}, \ldots, F^{(7)})$, where $F^{(j)} = (f_{1j}, f_{2j})$ for $j = 1, \ldots, 7$. Then $D = (F^{(1)}V, \ldots, F^{(7)}V)$ is an ONSOA$(16, 2 \times 7, 4, 2+)$ with $\pi = 92.31\%$, as shown in Table 1.

Based on Construction 1 and Theorem 1, various ONSOAs can be constructed from orthogonal arrays of strength 2. Here we summarize three families from distinct existence and construction results of the orthogonal arrays (Hedayat et al., 1999, Chaps. 3 & 7). (i) From a Hadamard matrix of order m, we can obtain an OA$(m, m-1, 2, 2)$. Then Theorem 1 can produce an ONSOA$(2m, 2 \times (m-1), 4, 2+)$. The number of columns $(2m-2)$ for the resulting design is closer to the maximum number of columns $(2m-1)$ for a column-orthogonal design. (ii) For a prime power s, the Rao–Hamming construction gives a saturated regular design OA$(s^{k-1}, m, s, 2)$ with $m = (s^{k-1}-1)/(s-1)$. Based on this, we can obtain an ONSOA$(s^k, 2 \times m, s^2, 2+)$. (iii) The Addelman–Kempthorne construction provides an OA$(2s^{k-1}, m, s, 2)$ with $m = 2(s^{k-1}-1)/(s-1)-1$ for any odd prime power s. So, we can obtain an ONSOA$(2s^k, 2 \times m, s^2, 2+)$ correspondingly.

Note that Zhou and Tang (2019) obtained the OSOA$(sn, m, s^2, 2+)$ by $sF_1 + F_2$ in our symbols. This is a special case of our method if we take $V = (s, 1)^T$ in Step 3, while our designs can accommodate double number of factors. Note that the proposed method works for any $s \geq 2$, which is not limited to a prime power. According to the flexible choices of orthogonal arrays and simple construction, Theorem 1 provides a very powerful method to construct ONSOA$(sn, 2 \times m, s^2, 2+)$ designs.

3.2. Column-orthogonal nearly strong orthogonal arrays with sn runs and cm factors

This subsection provides a construction of ONSOAs with more columns using difference schemes. For a prime power s, suppose the s levels are taken from the Galois field GF$(s) = \{\alpha_0 = 0, \alpha_1 = 1, \ldots, \alpha_{s-1}\}$, which is simplified as $\{0, 1, \ldots, s-1\}$ if s is a prime. Let $\omega = (\alpha_0, \alpha_1, \ldots, \alpha_{s-1})^T$ and $g_i = \alpha_{i-1}\omega$ for $i = 1, \ldots, s$. Consequently, $(g_1, \ldots, g_s)$ is a difference scheme, see Hedayat et al. (1999, Chap. 6.2). Let A be an OA$(n, m, s, 2)$ and $c = 2\lfloor s/2 \rfloor$. The construction method is as follows.

Construction 2.

Step 1. For $i = 1, \ldots, c$, define an $sn \times m$ matrix $F_i = g_i \oplus A$ and write $F_i = (f_{i1}, \ldots, f_{im})$, where $g_i \oplus A$ is the Kronecker sum of g_i and A.

Step 2. For $i = 1, \ldots, c$, replace the s levels $\{\alpha_0, \alpha_1, \ldots, \alpha_{s-1}\}$ of F_i by $\{0, 1, \ldots, s-1\}$ and then center them by $x-(s-1)/2$ for $x \in \{0, 1, \ldots, s-1\}$ such that they are from $\Omega(s)$. Then arrange the cm centered columns f_{ij}'s as $(F^{(1)}, \ldots, F^{(m)})$, where $F^{(j)} = (f_{1j}, \ldots, f_{cj})$ has c columns for $j = 1, \ldots, m$.

Step 3. Define

$$D = (D_1, \ldots, D_m), \tag{4}$$

where $D_j = F^{(j)}R$ for $j = 1, \ldots, m$, and $R = \text{diag}\{V, \ldots, V\}$ with V repeating $c/2$ times in the diagonal.

Similar to Theorem 1, a theoretical property of the design D in (4) is stated as follows.

Theorem 2. *If there exists an* OA$(n, m, s, 2)$*, then the design D in (4) is an* ONSOA$(sn, c \times m, s^2, 2+)$*, where s is a prime power and $c = 2\lfloor s/2 \rfloor$.*

Compared with the condition in Theorem 1, Theorem 2 requires that s is a prime power. In fact, this is not a strong condition for the existence of orthogonal arrays. For $s = 2, 3$, Construction 2 is equivalent to Construction 1. For $s \geq 4$, however, Construction 2 will generate ONSOAs with much more factors than Construction 1.

Both the design D in (4) and a mappable nearly orthogonal array (Mukerjee et al., 2014) can be collapsed into an orthogonal array of strength 2, while the latter does not enjoy column orthogonality. In a word, the former possesses both the attractive two-dimensional stratifications and column orthogonality.

Example 3. Let A be an OA$(16, 5, 4, 2)$ and GF$(4) = \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$. For $i = 1, \ldots, 4$, let $F_i = g_i \oplus A$, where $g_i = \alpha_{i-1}(\alpha_0, \alpha_1, \alpha_2, \alpha_3)^T$, and write $F_i = (f_{i1}, \ldots, f_{i5})$. Replace the levels of each F_i by $\{0, 1, 2, 3\}$ and then center them to be $\{-3/2, -1/2, 1/2, 3/2\}$. Next, arrange the 20 centered columns f_{ij}'s as $(F^{(1)}, \ldots, F^{(5)})$, where $F^{(j)} = (f_{1j}, \ldots, f_{4j})$ for $j = 1, \ldots, 5$. Then we have that $D = (D_1, \ldots, D_5) = (F^{(1)}R, \ldots, F^{(5)}R)$ is an ONSOA$(64, 4 \times 5, 16, 2+)$ with $\pi = 84.21\%$, where $R = \text{diag}\{V, V\}$.

For a prime power s, the Bush construction (Hedayat et al., 1999, Chap. 3.2) and an ovoid provide an OA$(s^3, s+1, s, 3)$ and an OA$(s^4, s^2+1, s, 3)$, respectively, where a set of s^2+1 points $\{(w, x, y, z)\}$ in projective geometry PG$(3, s)$ is called an ovoid if any three points of the set are not collinear; see Hedayat et al. (1999, Chap. 5.9) and Calderbank and Kantor (1986) for further details. Based on the two orthogonal arrays, He and Tang (2014) obtained an SOA$(s^3, s+1, s^3, 3)$ and an SOA$(s^4, s^2+1, s^3, 3)$. For comparison, based on the OA$(s^{k-1}, (s^{k-1}-1)/(s-1), s, 2)$ from the Rao–Hamming construction, our method can generate an ONSOA$(s^3, c(s+1), s^2, 2+)$ and an ONSOA$(s^4, c(s^2+s+1), s^2, 2+)$ with $c = 2\lfloor s/2 \rfloor$. Take $s = 4$ as an example. Our design ONSOA$(64, 4 \times 5, 16, 2+)$ has almost the same two-dimensional stratifications (with $\pi = 84.21\%$) as that of the SOA$(64, 5, 64, 3)$, while the former enjoys much more columns and additional column orthogonality. Similar observations can also be obtained from the comparison between the ONSOA$(256, 4 \times 21, 16, 2+)$ and SOA$(256, 17, 64, 3)$.

Table 2
Comparisons of the number of factors for SOA($n, m_1, s^3, 3$), OSOA($n, m_2, s^3, 3$), OSOA($n, m_3, s^2, 2+$) and ONSOA($n, m_4, s^2, 2+$), and the degree of stratifications π.

Families	s	k	n	m_1	m_2	m_3	m_4	π(%)
HM	2	–	8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96	$n/2-1$	$n/4$	$n/2-1$	$n-2$	$1-1/(n-3)$
RH	2	4	16	7	4	7	14	92.31
RH	2	5	32	15	8	15	30	96.55
RH	2	6	64	31	16	31	62	98.36
RH	2	7	128	63	32	63	126	99.20
RH	2	8	256	127	64	127	254	99.60
RH	3	3	27	4	2	4	8	85.71
RH	3	4	81	10	5	13	26	96.00
RH	3	5	243	19	10	40	80	98.73
RH	4	3	64	5	2	5	20	84.21
RH	4	4	256	17	8	21	84	96.39
RH	4	5	1024	–	–	85	340	99.12
RH	5	3	125	6	3	6	24	86.96
RH	5	4	625	26	13	31	124	97.56
RH	7	3	343	8	4	8	48	89.36
RH	8	3	512	9	4	9	72	90.14
RH	9	3	729	10	5	10	80	91.14
RH	10	3	1000	11	5	11	110	91.74
AK	3	3	54	5	–	7	14	92.31
AK	3	4	162	–	–	25	50	97.96
AK	5	3	250	–	–	11	44	93.02
AK	7	3	686	–	–	15	90	94.38

Note: HM, Hadamard matrix; RH, the Rao–Hamming construction; AK, the Addelman–Kempthorne construction.

4. Comparisons with existing SOAs

This section employs Constructions 1 and 2 to obtain ONSOAs based on the three families of orthogonal arrays mentioned in Section 3.1. For the run sizes $n \le 1000$, some of the resulting designs are shown in Table 2 and many others, which can be obtained similarly, are omitted.

Table 2 shows some comparisons of the numbers of factors, denoted by m_1, m_2, m_3 and m_4, respectively for the four kinds of designs, including the SOA($n, m_1, s^3, 3$) from He and Tang (2014), OSOA($n, m_2, s^3, 3$) from Liu and Liu (2015), OSOA($n, m_3, s^2, 2+$) from Zhou and Tang (2019) and ONSOA($n, m_4, s^2, 2+$) from our methods, where $n = s^k$ for the Rao–Hamming construction and $n = 2s^k$ for the Addelman–Kempthorne construction. The numbers of columns for the first three kinds of designs can be calculated accordingly, and $m_4 = c \times m_3$ with $c = 2\lfloor s/2 \rfloor$. The second row of Table 2 shows the cases of $n \le 100$ being a multiple of eight ($n = 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96$) with $s = 2$, where the designs are constructed from Hadamard matrices. Many other cases of $n > 100$ can be similarly obtained accordingly, and a library of Hadamard matrices can be found from Dr. N. J. A. Sloane's website, http://neilsloane.com/hadamard/. The last column of Table 2 lists the degree of stratifications π in (1) for our designs. For simplicity, the four kinds of designs are denoted by SOA(3), OSOA(3), OSOA(2+) and ONSOA(2+) respectively.

We summarize the following observations from Table 2. Compared with SOA(3), our design ONSOA(2+) possesses column orthogonality. Compared with SOA(3) and OSOA(3), if they exist, our design ONSOA(2+) can accommodate much more columns with column orthogonality and comparable two-dimensional stratifications, although the three-dimensional stratifications cannot be guaranteed. Compared with OSOA(2+), our design ONSOA(2+) can accommodate much more columns, with column orthogonality retained and two-dimensional stratifications not sacrificed too much. Compared with SOA(3) and OSOA(3), OSOA(2+) and ONSOA(2+) gain column orthogonality and/or more columns with almost the same two-dimensional stratifications, although the number of levels decreases from s^3 to s^2.

5. Space-filling property

This section is devoted to studying the space-filling property of the resulting ONSOAs under the maximin distance criterion. A theoretical justification for the space-filling property is provided.

Let U(n, s^m) denote a balanced design with n runs, m factors, and s levels from $\{0, 1, \ldots, s-1\}$ where the s levels appear equally often for each factor. Let D be a U(n, s^m). The L_2-distance between two distinct rows $x_i = (x_{i1}, \ldots, x_{im})$ and $x_j = (x_{j1}, \ldots, x_{jm})$ in D is defined to be $d(x_i, x_j) = \sum_{k=1}^{m} |x_{ik} - x_{jk}|^2$. Define the L_2-distance of D to be $d(D) = \min_{i \ne j} d(x_i, x_j)$. A maximin distance design is to maximize the $d(D)$ value (Johnson et al., 1990). For a $D \in$ U(n, s^m), its average pairwise L_2-distance between rows is $d_{\text{ave}}(D) = n(s^2-1)m/(6n-6)$ (Zhou and Xu, 2015). Since the minimum pairwise L_2-distance cannot exceed the integer part of the average, we have $d(D) \le \lfloor d_{\text{ave}}(D) \rfloor$, where $\lfloor d_{\text{ave}}(D) \rfloor$ is called the upper bound for the L_2-distance of D. Based on this, we define the distance efficiency as

$$d_{\text{eff}}(D) = d(D)/\lfloor d_{\text{ave}}(D) \rfloor = d(D)/\lfloor n(s^2-1)m/(6n-6) \rfloor. \quad (5)$$

Table 3
The distance efficiencies of our designs and L_2-distances of $32 \times m$ designs with $m = 2k$ columns, $k = 4, 5, \ldots, 15$.

m	8	10	12	14	16	18	20	22	24	26	28	30
d_{eff}	0.50	0.48	0.67	0.61	0.73	0.72	0.80	0.82	0.85	0.90	0.97	0.97
Ours	1.11	1.33	2.22	2.44	3.33	3.67	4.56	5.11	5.78	6.67	7.78	8.33
SLHD	0.91	1.26	1.62	2.00	2.37	2.75	3.20	3.47	3.86	4.23	4.61	5.04

Note: Ours, our designs; SLHD, R package "SLHD".

The larger $d_{\text{eff}}(D)$, the better. Because of strict restrictions on the numbers of runs, levels and factors, the upper bound for an ONSOA may be not attainable. In many situations, even though the $d_{\text{ave}}(D)$ in (5) is an integer, this upper bound may not be achieved and there is still a need for further research. Before we introduce the space-filling property of the proposed ONSOAs, we need the following.

In Step 2 of Construction 1, write $F = (F^{(1)}, \ldots, F^{(m)})$, where each $F^{(j)}$ has two columns. We know that F must be an OA$(sn, 2m, s, 2)$ with entries from $\Omega(s)$. Since the D in (2) can be rewritten as $D = FR_1$, where $R_1 = \text{diag}\{V, \ldots, V\}$ with V in (3) repeating m times in the diagonal, we have the following result.

Lemma 1. *Let $D = FR_1$, where F is an* OA$(sn, 2m, s, 2)$ *and $R_1 = \text{diag}\{V, \ldots, V\}$ with V repeating m times in the diagonal. Then we have that*

$$d(D) = (s^2 + 1)d(F) \quad \text{and} \quad d_{\text{eff}}(D) \geq d(F)/d_{\text{ave}}(F).$$

Lemma 1 shows that the L_2-distance of D is determined by that of F, and the distance efficiency of D also relies on that of F since a large $d_{\text{eff}}(F)$ must lead to a large $d(F)/d_{\text{ave}}(F)$.

We now investigate the L_2-distance and distance efficiency of an ONSOA$(2n, 2\times(n-1), 4, 2+)$, which is constructed by taking A to be an OA$(n, n-1, 2, 2)$ via Construction 1. Note that $F = (F^{(1)}, \ldots, F^{(n-1)})$ is an OA$(2n, 2(n-1), 2, 2)$. According to the structure of F, a saturated OA$(2n, 2n-1, 2, 2)$ can be obtained if we add to F a $2n \times 1$ two-level column with the first n entries being $1/2$ and the next n entries being $-1/2$. Since such a saturated OA$(2n, 2n-1, 2, 2)$ has an L_2-distance of n with any two distinct rows being equidistant, we have that F must have an L_2-distance of $n-1$, i.e., $d(F) = n-1$. Taking $s = 2$ and $m = n-1$ in Lemma 1, we have the following result.

Theorem 3. *Let D be an* ONSOA$(2n, 2 \times (n-1), 4, 2+)$ *constructed in (2) via Construction 1. Then we have that*

$$d(D) = 5(n-1) \quad \text{and} \quad d_{\text{eff}}(D) \geq (2n-1)/(2n).$$

From Theorem 3, we can see that $d_{\text{eff}}(D)$ converges to one as n goes to infinity. This shows that the proposed ONSOAs with four levels perform well in terms of the maximin distance criterion. Thus an ONSOA$(2n, 2\times(n-1), 4, 2+)$ achieves the attractive space-filling properties not only in two-dimensional projections, but also in the full space.

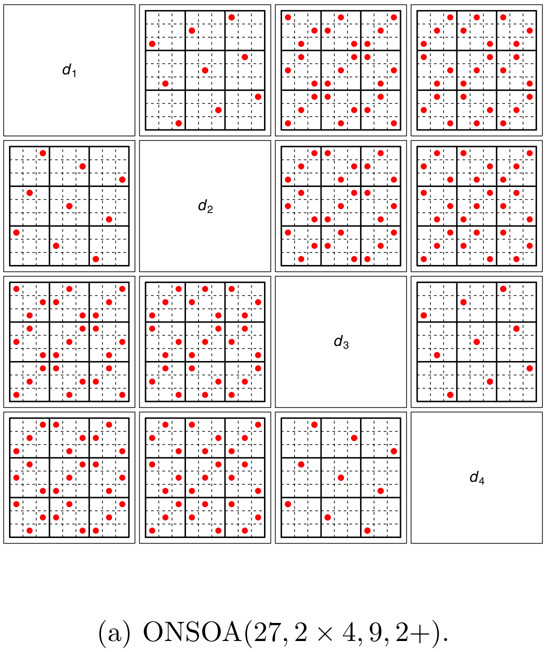
As suggested by one referee, we next consider the L_2-distances of our designs with $m \leq 2(n-1)$ columns. Some comparisons are also made with the designs generated by the R package "SLHD" (Ba et al., 2015). For a fair comparison, the L_2-distance here is calculated after the levels are scaled into the interval [0, 1]. For our designs with $m \leq 2(n-1)$ columns, we randomly generate m-dimensional projections 100 times based on the ONSOA$(2n, 2 \times (n-1), 4, 2+)$ and choose the best one. For the designs by the R package "SLHD", we run the R command *maximinLHS* 100 times with default settings and choose the best design for each m. Table 3 shows the distance efficiencies of our designs and L_2-distances of both $32\times m$ designs with $m = 2k$ columns, $k = 4, 5, \ldots, 15$. It can be seen that our designs have high distance efficiencies for large values of m. For small values of m, although the distance efficiencies get small, our designs still have slightly larger L_2-distances than SLHDs. Similar phenomenon also occurs for designs with other run sizes.

To study the space-filling property of the ONSOA D constructed via Construction 2, we first introduce a strategy to improve the distance efficiency. Let F be $(F^{(1)}, \ldots, F^{(m)})$ in Step 2 of Construction 2. We use the linear level permutation on F to obtain a new F, where the levels of F are adaptively chosen to be $\{0, 1, \ldots, s-1\}$ for the linear level permutation and $\Omega(s)$ for Construction 2. For example, write $F = (f_1, \ldots, f_{cm})$ and $u = (u_1, \ldots, u_{cm})$ with $u_j \in \{0, 1, \ldots, s-1\}$, and define $F + u = (f_1 + u_1, \ldots, f_{cm} + u_{cm}) \pmod{s}$ as the new F. Based on this new F, we obtain a new ONSOA via Construction 2. Repeat the linear level permutation 100 times by randomly generating a u each time, and then select the best one (say D^*) among the 100 generated ONSOAs in terms of the distance efficiency. Since a systematic study on this point is out of the scope of the paper, we just give an illustrative example below.

Example 4. According to the above strategy, let D be an ONSOA$(256, 4\times 21, 16, 2+)$ constructed via Construction 2 and D^* be the improved one, which have the distance efficiencies of 0.30 and 0.65, respectively. Therefore, D^* is preferred. Similar to the discussion in Table 3, we also consider the L_2-distances of m-dimensional projections of D^*, in comparison with those generated by the R package "SLHD" (Ba et al., 2015), where the setting of 10^4 iterations is utilized in the R command *maximinLHS*. The results in Table 4 indicate that the distance efficiency decreases as the number of columns gets smaller, while our designs are relatively comparable to those from the R package "SLHD" in terms of the L_2-distance.

Table 4
The distance efficiencies of our designs and L_2-distances of $256 \times m$ designs with $m = 6k$ columns, $k = 2, 3, \ldots, 14$.

m	12	18	24	30	36	42	48	54	60	66	72	78	84
d_{eff}	0.19	0.28	0.34	0.42	0.45	0.50	0.52	0.55	0.58	0.60	0.63	0.64	0.65
Ours	0.44	0.95	1.55	2.37	3.07	4.00	4.78	5.64	6.62	7.52	8.54	9.46	10.28
SLHD	0.54	1.19	1.92	2.67	3.44	4.23	5.18	5.92	6.67	7.50	8.35	9.30	10.17



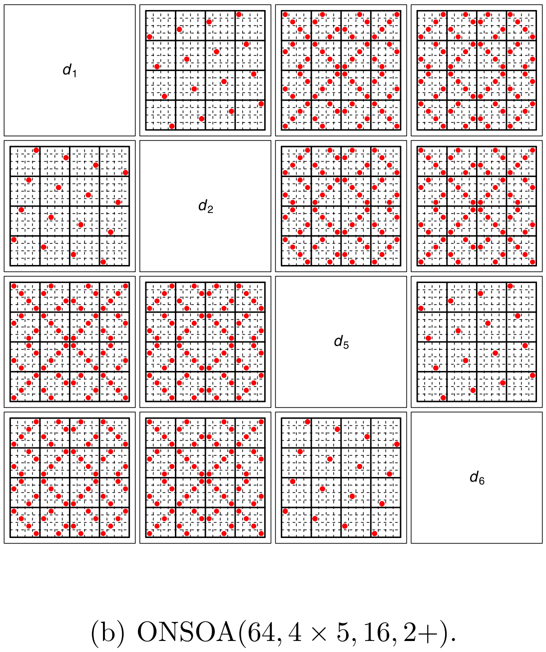


Fig. 1. Bivariate projections of (a) the first four columns of ONSOA$(27, 2 \times 4, 9, 2+)$ and (b) the columns (d_1, d_2, d_5, d_6) of ONSOA$(64, 4 \times 5, 16, 2+)$.

As the number of levels s^2 increases, the upper bound in (5) is hard to achieve. In such cases, the space-filling property of an ONSOA can be partly explained by the stratification properties given in Definition 1. Such stratification properties have motivated the development of strong orthogonal arrays (He and Tang, 2013, 2014; He et al., 2018; Zhou and Tang, 2019). We now use an example to illustrate the stratification property of ONSOAs with more than four levels.

Example 5. For $s = 3$ and $s = 4$, let D_1 and D_2 be an ONSOA$(27, 2 \times 4, 9, 2+)$ and an ONSOA$(64, 4 \times 5, 16, 2+)$, respectively. Their stratification properties can be seen intuitively in Fig. 1(a) and (b), respectively, where d_j stands for the jth column of each design. More specifically, the 8 columns of D_1 can be partitioned into 4 groups of 2 successive columns each such that any two columns achieve stratifications on 9×3 and 3×9 grids if they are from distinct groups (e.g., (d_1, d_3)), and on a 3×3 grid if they are from the same group (e.g., (d_1, d_2)). Similarly, the 20 columns of D_2 can be partitioned into 5 groups of 4 successive columns each such that any two columns achieve stratifications on 16×4 and 4×16 grids if they are from distinct groups (e.g., (d_1, d_5)), and on a 4×4 grid if they are from the same group (e.g., (d_1, d_2)). From the bivariate projections of D_1 and D_2, they enjoy appealing space-filling properties in terms of the stratifications.

6. Concluding remarks

This paper presents two methods for constructing column-orthogonal nearly strong orthogonal arrays (ONSOAs). The resulting designs enjoy column orthogonality, large number of factors, and attractive two-dimensional space-filling properties. These designs are based on orthogonal arrays (regular or nonregular) and have flexible run sizes. In addition, the proposed ONSOAs with four levels enjoy an attractive space-filling property under the maximin distance criterion.

One interesting research issue is to construct higher-strength versions of ONSOAs based on orthogonal arrays of strength three or more. It would be interesting to explore a new type of arrays with two-dimensional stratification on finer grids and/or even higher-dimensional space-filling property. If second-order effects exist, a 3-orthogonal ONSOA is more suitable. Such a design can be constructed by $(D^T, -D^T)^T$, where D is an ONSOA. A column-orthogonal design with centered levels is called 3-orthogonal if the sum of elementwise products of any three columns, whether they are distinct or not, is zero (Bingham et al., 2009).

CRediT authorship contribution statement

Wenlong Li: Conceptualization, Methodology, Software, Writing - original draft. **Min-Qian Liu:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition. **Jian-Feng Yang:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

The authors are grateful to the executive editors, an associate editor and a reviewer for their insightful comments and constructive suggestions. We thank Professor Yong-Dao Zhou for his valuable comments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 11771220, 11871033 and 11971204), National Ten Thousand Talents Program of China, Tianjin Development Program for Innovation and Entrepreneurship, China, Natural Science Foundation of Tianjin (20JCYBJC01050), Fundamental Research Funds for the Central Universities and the Ph.D. Candidate Research Innovation Fund of Nankai University, China. The authorship is listed in alphabetic order.

Appendix. Proofs

In the following proofs, according to the specific cases, the levels of OA(sn, m, s, t) can be adaptively chosen to be $\{0, 1, \ldots, s-1\}$ or $\Omega(s)$. To prove Theorems 1 and 2, Lemma 3 in Sun and Tang (2017) is rewritten in our symbols as follows.

Lemma 2. *A three-column submatrix, given by (f_{1i}, f_{2i}, f_{1j}) in Constructions 1 and 2, is an* OA$(sn, 3, s, 3)$ *for any $i \neq j$.*

A.1. Proof of Theorem 1

From the construction of D in (2), we have $D = FR_1$, where $F = (F^{(1)}, \ldots, F^{(m)})$ and $R_1 = \text{diag}\{V, \ldots, V\}$ with V repeating m times in the diagonal. By Lemma 2, F is an orthogonal array of strength 2. Therefore, $D^TD = (FR_1)^TFR_1 = R_1^T(F^TF)R_1 = e_1R_1^TR_1 = e_2I_{2m}$, where e_1 and e_2 are two constants, and I_{2m} is the identity matrix of order $2m$. This shows that D is column-orthogonal.

Next, we will show the stratifications of D in Definition 1(c). Note that any column d of D has the following form: $d = bs \pm b'$, where (b, b') up to a column permutation is (f_{1j}, f_{2j}) for $j = 1, \ldots, m$. Consider the mapping

$$h(x) = \lfloor \{x + (s^2-1)/2\}/s \rfloor - (s-1)/2 \quad \text{with} \quad x \in \Omega(s^2),$$

which collapses the s^2 levels in $\Omega(s^2)$ into the s levels in $\Omega(s)$. According to the structures of V and F, we only need to show that $h(d) = b$, which means that the column d becomes the column b after the mapping h is applied to each component of d. By letting $r = b + (s-1)/2$ and $r' = \pm b' + (s-1)/2$, we have

$$\begin{aligned} h(d) &= \lfloor \{bs \pm b' + (s^2-1)/2\}/s \rfloor - (s-1)/2 \\ &= \lfloor (rs + r')/s \rfloor - (s-1)/2. \end{aligned}$$

Since all entries of (b, b') are in $\Omega(s)$, all entries of (r, r') must take values from $\{0, 1, \ldots, s-1\}$. We thus have $h(d) = r - (s-1)/2 = b$. This shows that D satisfies Definition 1(c).

Finally, we need to prove the stratifications of D in Definition 1(b). Write $D_i = (d_{i1}, d_{i2})$ and $D_j = (d_{j1}, d_{j2})$. Without loss of generality, we only prove that (d_{i1}, d_{j1}) can be collapsed into an OA$(sn, 2, s^2 \times s, 2)$ and an OA$(sn, 2, s \times s^2, 2)$, that is to say, $(d_{i1}, h(d_{j1}))$ and $(h(d_{i1}), d_{j1})$ are an OA$(sn, 2, s^2 \times s, 2)$ and an OA$(sn, 2, s \times s^2, 2)$, respectively. In fact, this is true by noting the following two facts: (i) from Lemma 2, (f_{1i}, f_{2i}, f_{1j}) is an OA$(sn, 3, s, 3)$ for any $i \neq j$; (ii) $sx_1 + x_2$ establishes a one-to-one correspondence between the s^2 pairs (x_1, x_2) with $x_1, x_2 \in \Omega(s)$ and the s^2 levels in $\Omega(s^2)$. □

A.2. Proof of Theorem 2

Note that $(g_1, \ldots, g_c)$ is a difference scheme and A is an orthogonal array. From Hedayat et al. (1999), we know that $(F_1, \ldots, F_c) = (g_1, \ldots, g_c) \oplus A$ is an orthogonal array. This also means that $(F^{(1)}, \ldots, F^{(m)})$ is an orthogonal array. According to the proof of Theorem 1, by noting that $h(D) = (F^{(1)}, \ldots, F^{(m)})$, Definition 1(a) and (c) can be verified easily for the D in (4).

Next, we will show the stratifications of D in Definition 1(b). Write two different groups as $D_i = (d_{i1}, \ldots, d_{ic})$ and $D_j = (d_{j1}, \ldots, d_{jc})$ for any $i \neq j$. We only need to prove that (d_{i1}, d_{jk}) $(k = 1, \ldots, c)$ satisfies the property in Definition 1(b). Theorem 1 guarantees the cases of $k = 1, 2$. We now verify that (d_{i1}, d_{jk}) $(k = 3, \ldots, c)$ satisfies Definition 1(b). For simplicity, we only consider the case of (d_{i1}, d_{j3}) since other cases follow similarly. That is, we only need to prove that $(d_{i1}, h(d_{j3}))$ and $(h(d_{i1}), d_{j3})$ are an OA$(sn, 2, s^2 \times s, 2)$ and an OA$(sn, 2, s \times s^2, 2)$, respectively. Note that $d_{i1} = sf_{1i} + f_{2i}$, $d_{j3} = sf_{3j} + f_{4j}$, $h(d_{i1}) = f_{1i}$ and $h(d_{j3}) = f_{3j}$. So, the results follow if we can show that both (f_{1i}, f_{2i}, f_{3j}) and (f_{1i}, f_{3j}, f_{4j}) are orthogonal arrays of strength three. To do so, write $A = (a_1, \ldots, a_m)$ to be the OA$(n, m, s, 2)$.

From Step 1 of Construction 2, we have $f_{1i} = g_1 \oplus a_i, f_{2i} = g_2 \oplus a_i$ and $f_{3j} = g_3 \oplus a_j$. Let $\gamma_1 = (\alpha_0, \ldots, \alpha_{s-1})^T \oplus 0_s$, $\gamma_2 = 0_s \oplus (\alpha_0, \ldots, \alpha_{s-1})^T$, $\beta_1 = 0_{n/s} \oplus \gamma_1$ and $\beta_2 = 0_{n/s} \oplus \gamma_2$, where 0_s means an $s \times 1$ column vector with all entries zero. Note that $f_{1i} = (g_1 - g_2) \oplus 0_n + f_{2i}$ and (f_{2i}, f_{3j}) is an OA$(sn, 2, s, 2)$. Hence, the rows of (f_{1i}, f_{2i}, f_{3j}) can be rearranged to be $((g_1 - g_2) \oplus 0_n + \beta_1, \beta_1, \beta_2)$ based on the last two columns. This form implies that (f_{1i}, f_{2i}, f_{3j}) has strength three since $g_1 - g_2$ contains each entry of GF(s) exactly once. A similar discussion can also show that (f_{1i}, f_{3j}, f_{4j}) has strength three. This completes the proof. □

A.3. Proof of Lemma 1

Let $m' = 2m$ and $R_1 = (r_{ij})$. Define any two rows of F to be $y_1 = (y_{11}, \ldots, y_{1m'})$ and $y_2 = (y_{21}, \ldots, y_{2m'})$. Since $D = FR_1$, the resulting two rows of D are $x_1 = y_1R_1$ and $x_2 = y_2R_1$. Due to $x_1 - x_2 = (y_1 - y_1)R_1$, we have $d(x_1, x_2) = (x_1 - x_2)(x_1 - x_2)^T = (y_1 - y_2)R_1R_1^T(y_1 - y_2)^T = (s^2 + 1)(y_1 - y_2)(y_1 - y_2)^T = (s^2 + 1)d(y_1, y_2)$. By the definition of the L_2-distance, we have $d(D) = (s^2 + 1)d(F)$. From (5), we have

$$\begin{aligned} d_{\text{eff}}(D) &\geq d(D)/d_{\text{ave}}(D) \\ &= (s^2 + 1)d(F)/[sn(s^4 - 1)m'/(6sn - 6)] \\ &= d(F)/[sn(s^2 - 1)m'/(6sn - 6)] \\ &= d(F)/d_{\text{ave}}(F). \end{aligned}$$

This completes the proof. □

References

Ba, S., Myers, W.R., Brenneman, W.A., 2015. Optimal sliced latin hypercube designs. Technometrics 57, 479–487.
Bingham, D., Sitter, R.R., Tang, B., 2009. Orthogonal and nearly orthogonal designs for computer experiments. Biometrika 96, 51–65.
Calderbank, R., Kantor, W.M., 1986. The geometry of two-weight codes. Bull. Lond. Math. Soc. 18, 97–122.
Fang, K.T., Li, R., Sudjianto, A., 2006. Design and Modeling for Computer Experiments. Chapman and Hall/CRC, Boca Raton.
He, Y., Cheng, C.S., Tang, B., 2018. Strong orthogonal arrays of strength two plus. Ann. Statist. 46, 457–468.
He, Y., Tang, B., 2013. Strong orthogonal arrays and associated latin hypercubes for computer experiments. Biometrika 100, 254–260.
He, Y., Tang, B., 2014. A characterization of strong orthogonal arrays of strength three. Ann. Statist 42, 1347–1360.
Hedayat, A.S., Sloane, N.J.A., Stufken, J., 1999. Orthogonal Arrays: Theory and Applications. Springer, New York.
Johnson, M.E., Moore, L.M., Ylvisaker, D., 1990. Minimax and maximin distance designs. J. Statist. Plann. Inference 26, 31–148.
Joseph, V.R., Hung, Y., 2008. Orthogonal-maximin latin hypercube designs. Statist. Sinica 18, 171–186.
Lin, C.D., Mukerjee, R., Tang, B., 2009. Construction of orthogonal and nearly orthogonal latin hypercubes. Biometrika 96, 243–247.
Liu, H., Liu, M.Q., 2015. Column-orthogonal strong orthogonal arrays and sliced strong orthogonal arrays. Statist. Sinica 25, 1713–1734.
Mukerjee, R., Sun, F., Tang, B., 2014. Nearly orthogonal arrays mappable into fully orthogonal arrays. Biometrika 101, 957–963.
Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia.
Owen, A.B., 1992. Orthogonal arrays for computer experiments, integration and visualization. Statist. Sinica 2, 439–452.
Pang, F., Liu, M.Q., Lin, D.K.J., 2009. A construction method for orthogonal latin hypercube designs with prime power levels. Statist. Sinica 19, 1721–1728.
Santner, T.J., Williams, B.J., Notz, W.I., 2013. The Design and Analysis of Computer Experiments. Springer, New York.
Steinberg, D.M., Lin, D.K.J., 2006. A construction method for orthogonal latin hypercube designs. Biometrika 93, 279–288.
Sun, F., Liu, M.Q., Lin, D.K.J., 2009. Construction of orthogonal latin hypercube designs. Biometrika 96, 971–974.
Sun, F., Tang, B., 2017. A general rotation method for orthogonal latin hypercubes. Biometrika 104, 465–472.
Tang, B., 1993. Orthogonal array-based latin hypercubes. J. Amer. Statist. Assoc. 88, 1392–1397.
Ye, K.Q., 1998. Orthogonal column latin hypercubes and their application in computer experiments. J. Amer. Statist. Assoc. 93, 1430–1439.
Zhou, Y.D., Tang, B., 2019. Column–orthogonal strong orthogonal arrays of strength two plus and three minus. Biometrika 106, 997–1004.
Zhou, Y.D., Xu, H., 2015. Space-filling properties of good lattice point sets. Biometrika 102, 959–966.