Biometrika (2022), **109**, 2, *pp.* 489–501
doi: 10.1093/biomet/asab021
Advance Access publication 8 September 2021

A minimum aberration-type criterion for selecting space-filling designs

BY YE TIAN AND HONGQUAN XU

Department of Statistics, University of California, 8125 Math Sciences Building, Box 951554, Los Angeles, California 90095, U.S.A.

yetian@ucla.edu hqxu@stat.ucla.edu

SUMMARY

Space-filling designs are widely used in computer experiments. Inspired by the stratified orthogonality of strong orthogonal arrays, we propose a criterion of minimum aberration-type for assessing the space-filling properties of designs based on design stratification properties on various grids. A space-filling hierarchy principle is proposed as a basic assumption of the criterion. The new criterion provides a systematic way of classifying and ranking space-filling designs, including various types of strong orthogonal arrays and Latin hypercube designs. Theoretical results and examples are presented to show that strong orthogonal arrays of maximum strength are favourable under the proposed criterion. For strong orthogonal arrays of the same strength, the space-filling criterion can further rank them based on their space-filling patterns.

Some key words: Computer experiment; Generalized minimum aberration; Space-filling hierarchy principle; Space-filling pattern; Strong orthogonal array.

1. INTRODUCTION

Computer models are computer codes used to simulate complicated, hard-to-solve systems. Computer experiments aim to build statistical surrogate models efficiently based on the data from computer models (Santner et al., 2003; Fang et al., 2006). Space-filling designs are widely used in computer experiments. There are a number of different types of space-filling designs, such as Latin hypercube designs and variations (Ba et al., 2015; Lin & Tang, 2015; Xiao & Xu, 2017), maximin distance designs (Wang et al., 2018; Xiao & Xu, 2018; Li et al., 2020) and uniform designs (Fang et al., 2018). Joseph et al. (2015) and Sun et al. (2019) pointed out that maximin distance designs and uniform designs may have poor low-dimensional projections which are not space-filling.

We focus on an alternative approach that explicitly takes into account low-dimensional projections. Latin hypercube designs achieve stratifications in all univariate projections. Using orthogonal arrays of strength $t \geqslant 2$, Owen (1992) and Tang (1993) provided constructions of space-filling designs that achieve stratifications in all projections of dimension t and lower. He & Tang (2013) proposed strong orthogonal arrays that are more space-filling than comparable ordinary orthogonal arrays. A strong orthogonal array of strength t achieves stratifications in all t-dimensional margins as comparable ordinary orthogonal arrays do. Further, it achieves stratifications on finer grids in all g-dimensional margins for any $g < t$. He et al. (2018) introduced strong orthogonal arrays of strength 2+ and presented construction results for such designs. Strong orthogonal arrays of strength 2+ achieve the same two-dimensional stratifications as strong orthogonal arrays of strength 3 while keeping run sizes small. Shi & Tang (2019) proposed methods of distinguishing strong orthogonal arrays of strength 2+ and 2 based on three-dimensional and two-dimensional projections, respectively. Zhou & Tang (2019) studied the construction of strong orthogonal arrays of strength 2+ and 3− with column orthogonality. Shi & Tang (2020) further proposed construction methods for strong orthogonal arrays of strength 3 with additional stratification properties. Despite all this progress, the important topic of design selection of strong orthogonal arrays has not been systematically addressed. As we will show, there are many strong orthogonal arrays that are of the same strength yet have quite different space-filling properties.

In this paper, we propose a minimum aberration-type criterion for classifying and selecting space-filling designs in a systematic way. The new criterion is inspired by the popular minimum aberration criterion and its extensions, which are widely used for selecting and ranking fractional factorial designs; see Mukerjee & Wu (2006), Wu & Hamada (2009) and Cheng (2014). The underlying assumption for the minimum aberration criterion is the effect hierarchy principle (Wu & Hamada, 2009): (i) lower-order effects are more likely to be important than higher-order effects; and (ii) effects of the same order are equally likely to be important. However, the minimum aberration criterion and its extensions cannot be used to assess the space-filling properties of Latin hypercube designs and strong orthogonal arrays; for example, all Latin hypercubes of the same size, whether orthogonal array-based or not, have the same generalized wordlength pattern.

Instead of considering factorial effects, we take stratification properties into account when assessing space-filling properties. Our space-filling criterion is based on the following space-filling hierarchy principle: (i) stratifications on larger grids are more likely to be important than stratifications on smaller grids; and (ii) stratifications on grids of the same volume are equally likely to be important. We formalize the principle here, although it was implicitly used in the development of strong orthogonal arrays. We carefully define the space-filling pattern to characterize the stratification properties on various grids according to this principle. The space-filling criterion selects the designs that sequentially minimize the space-filling pattern. The new criterion can be applied to a broader class of designs, including the various strong orthogonal arrays and Latin hypercube designs mentioned earlier. We show that strong orthogonal arrays of maximum strength are favourable under the space-filling criterion. We further present examples to show that the new criterion can classify and rank designs of the same strength.

2. NOTATION AND BACKGROUND

Let $\mathbb{Z}_s = \{0, 1, \ldots, s-1\}$ be the ring of integers modulo s. An orthogonal array of strength t, denoted by $\text{OA}(n, m, s_1 \times \cdots \times s_m, t)$, is an $n \times m$ matrix in which the entries of the jth column are taken from $\mathbb{Z}_{s_j}$ and such that in any t-column subarray, all possible level combinations appear equally often. If $s_1 = \cdots = s_m = s$, the orthogonal array is symmetric and can be written as $\text{OA}(n, m, s, t)$. For a symmetric orthogonal array of strength t, the number of rows must be a multiple of s^t. Define $n = \lambda s^t$ where λ is called the index of the orthogonal array. Latin hypercubes are $n \times m$ matrices in which each column is strictly a permutation of n evenly spread levels, say $\{0, \ldots, n-1\}$; they are special orthogonal arrays of strength 1 with $\lambda = 1$.

A strong orthogonal array of n runs, m factors, s^t levels and strength t is an $n \times m$ matrix with entries from $\mathbb{Z}_{s^t}$ such that any subarray of g columns for any $1 \leqslant g \leqslant t$ can be collapsed into an

$\text{OA}(n, g, s^{u_1} \times \cdots \times s^{u_g}, g)$ for any set of positive integers $\{u_1, \ldots, u_g\}$ satisfying $u_1 + \cdots + u_g = t$. Collapsing s^t levels into s^{u_j} levels is done by calculating $\lfloor a/s^{t-u_j} \rfloor$ for $a = 0, 1, \ldots, s^t - 1$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x. We denote this strong orthogonal array by $\text{SOA}(n, m, s^t, t)$. Similarly, the index λ of a strong orthogonal array is defined as $n = \lambda s^t$. If $\lambda = 1$, the corresponding strong orthogonal array is also a Latin hypercube.

A design matrix with fixed levels can be viewed as a set of design points distributed in a grid space. An $\text{SOA}(n, m, s^t, t)$ can be treated as a set of n design points spread within an s^{tm} hypercube. To assess the space-filling property of a design, especially in a high-dimensional space, we could look at its low-dimensional projections. As a result of stratified orthogonality, strong orthogonal arrays of strength t achieve stratification regardless of how the design space is divided into s^t equal-volume grids from projection. For example, strong orthogonal arrays of strength 3 guarantee stratifications on s^3 grids in any one dimension, $s^2 \times s$ and $s \times s^2$ grids in any two dimensions, and $s \times s \times s$ grids in any three dimensions of the design region. Other examples include strong orthogonal arrays of strength 2+, which achieve stratifications on s^2 grids in any one dimension and on $s^2 \times s$ and $s \times s^2$ grids in any two dimensions, and strong orthogonal arrays of strength 3−, which achieve all the stratifications of those of strength 3 except for s^3 grids in any one dimension, as the total number of levels is only s^2. Stratified orthogonality guarantees good projection properties on finer grids.

To discuss a broader class of space-filling designs, we introduce the concept of general strong orthogonal arrays, which include strong orthogonal arrays as special cases. A general strong orthogonal array of n runs, m factors, s^p levels and strength t, denoted by $\text{GSOA}(n, m, s^p, t)$, is an $n \times m$ matrix with entries from $\mathbb{Z}_{s^p}$ such that any subarray of g columns for any $1 \leqslant g \leqslant t$ can be collapsed into an $\text{OA}(n, g, s^{u_1} \times \cdots \times s^{u_g}, g)$ for any set of positive integers $\{u_1, \ldots, u_g\}$ satisfying $u_1 + \cdots + u_g = t$ and $u_i \leqslant p$ for $i = 1, \ldots, g$. General strong orthogonal arrays of strength t achieve stratification no matter how the design space is divided into s^t equal-volume grids from projection. Strong orthogonal arrays have the constraint $t = p$; that is, a $\text{GSOA}(n, m, s^t, t)$ is an $\text{SOA}(n, m, s^t, t)$. Without this constraint, general strong orthogonal arrays include any design with s^p levels in the framework. Specifically, general strong orthogonal arrays of strength $t = 3$ and $p = 2$ are strong orthogonal arrays of strength 3−; that is, a $\text{GSOA}(n, m, s^2, 3)$ is an $\text{SOA}(n, m, s^2, 3-)$. General strong orthogonal arrays with $p = 1$ are ordinary orthogonal arrays; that is, a $\text{GSOA}(n, m, s^1, t)$ is an $\text{OA}(n, m, s, t)$. There may be situations where s and p are not clear. In such situations we will explicitly state either s or p or both. In this paper we focus on $p > 1$, even though the criterion and results also apply to $p = 1$.

3. A space-filling criterion

3.1. *Characters*

We first define a series of mapping functions f_i from $\mathbb{Z}_{s^p}$ to $\mathbb{Z}_s$. For $i = 1, \ldots, p$ and $x \in \mathbb{Z}_{s^p}$, let $f_i(x) = \lfloor x/s^{p-i} \rfloor \bmod s$. The function $f_i(x)$ outputs the ith digit of x in the base-s numeral system. The set of mapping functions is a bijection that expands x to a vector whose elements are from $\mathbb{Z}_s$. This expansion makes it possible to obtain information on x in each possible division. To transfer back, $x = \sum_{i=1}^{p} f_i(x) s^{p-i}$.

For $x \in \mathbb{Z}_{s^p}$, define its weight by $\rho(x) = p + 1 - \min\{i : f_i(x) \neq 0,\ i = 1, \ldots, p\}$ if $x \neq 0$ and $\rho(0) = 0$. The weight $\rho(x)$ is a generalization of the Hamming weight and represents the number of digits needed to express x in the base-s numeral system after eliminating all the leading zeros. As an example, Table 1 lists the mapping functions and weights for all possible $x \in \mathbb{Z}_{2^3}$.

Table 1. *Mapping functions and weights for* $x \in \mathbb{Z}_{2^3}$

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$\rho(x)$
0	0	0	0	0
1	0	0	1	1
2	0	1	0	2
3	0	1	1	2
4	1	0	0	3
5	1	0	1	3
6	1	1	0	3
7	1	1	1	3

Table 2. *Characters* $\chi_u(x)$ *for* $u, x \in \mathbb{Z}_{2^3}$

x	$\chi_0(x)$	$\chi_1(x)$	$\chi_2(x)$	$\chi_3(x)$	$\chi_4(x)$	$\chi_5(x)$	$\chi_6(x)$	$\chi_7(x)$
0	1	1	1	1	1	1	1	1
1	1	1	1	1	−1	−1	−1	−1
2	1	1	−1	−1	1	1	−1	−1
3	1	1	−1	−1	−1	−1	1	1
4	1	−1	1	−1	1	−1	1	−1
5	1	−1	1	−1	−1	1	−1	1
6	1	−1	−1	1	1	−1	−1	1
7	1	−1	−1	1	−1	1	1	−1

Define an inverse inner product between $u, x \in \mathbb{Z}_{s^p}$ by

$$\langle u, x\rangle = f_p(u)f_1(x) + \cdots + f_1(u)f_p(x) = \sum_{i=1}^{p} f_{p-i+1}(u)f_i(x).$$

As an illustration, suppose that the entries of a design matrix are from $\mathbb{Z}_{2^3}$. The mapping functions for $u = 2$ and $x = 6$ are $\{f_1(2), f_2(2), f_3(2)\} = (0, 1, 0)$ and $\{f_1(6), f_2(6), f_3(6)\} = (1, 1, 0)$. The inverse inner product between u and x is $\langle 2, 6\rangle = f_3(2)f_1(6) + f_2(2)f_2(6) + f_1(2)f_3(6) = 0 \times 1 + 1 \times 1 + 0 \times 0 = 1$.

For $u, x \in \mathbb{Z}_{s^p}$, define the character $\chi_u(x) = \xi^{\langle u,x\rangle}$ where $\xi = \exp(2\pi \mathrm{i}/s)$, with $\mathrm{i} = (-1)^{1/2}$, is the primitive sth root of unity. As an example, for $u = 2$ and $x = 6$, we have $\chi_2(6) = \xi^{\langle 2,6\rangle} = (-1)^1 = -1$. Table 2 shows the values of all possible characters $\chi_u(x)$ for $u, x \in \mathbb{Z}_{2^3}$.

We can expand the definitions of weight and character to vectors over $\mathbb{Z}_{s^p}$. For $u = (u_1, \ldots, u_m) \in \mathbb{Z}_{s^p}^m$, the weight $\rho(u) = \sum_{i=1}^m \rho(u_i)$ is defined as the sum of the individual weights, and the character $\chi_u(x) = \prod_{i=1}^m \chi_{u_i}(x_i)$ for any $x = (x_1, \ldots, x_m) \in \mathbb{Z}_{s^p}^m$ is defined as the tensor product of the individual characters. For example, for $u = (2, 3, 6)$ and $x = (6, 5, 4) \in \mathbb{Z}_{2^3}^3$, we have $\rho(u) = \sum_{i=1}^m \rho(u_i) = 2 + 2 + 3 = 7$ and $\chi_u(x) = \chi_2(6)\chi_3(5)\chi_6(4) = \xi^{\langle 2,6\rangle+\langle 3,5\rangle+\langle 6,4\rangle} = (-1)^2 = 1$.

Let $\tau = s^{mp}$. Let $x_1, \ldots, x_\tau$ and $u_1, \ldots, u_\tau$ denote all possible $x, u \in \mathbb{Z}_{s^p}^m$ in Yates order. Let $H = \{\chi_{u_j}(x_i)\}$ be the $\tau \times \tau$ matrix of characters evaluated at all possible points in $\mathbb{Z}_{s^p}^m$. Both the first row and the first column of H are vectors of ones. The character matrix H is symmetric because $\chi_u(x) = \chi_x(u)$. Table 2 shows the H matrix with $s = 2, p = 3$ and $m = 1$. When $s > 2$, H is a matrix of complex numbers. Let H^* be the conjugate transpose of H.

Table 3. *A Latin hypercube (left) and an* SOA$(8, 3, 8, 3)$ *(right)*

0	0	0	0	0	0
1	1	4	2	3	6
2	4	1	3	6	2
3	5	5	1	5	4
4	2	2	6	2	3
5	3	6	4	1	5
6	6	3	5	4	1
7	7	7	7	7	7

Theorem 1. *The character matrix H is symmetric and orthogonal; that is, $H^{\mathrm{T}} = H$ and $H^*H = HH^* = \tau I$, where I is an identity matrix of order τ.*

3.2. *Characteristics and the space-filling pattern*

Let D be a design with n runs, m columns and s^p levels. The design D can be regarded as n points spread out in the design space $\mathbb{Z}_{s^p}^m$. There are a total of τ distinct points in $\mathbb{Z}_{s^p}^m$. For each $x \in \mathbb{Z}_{s^p}^m$, let N_x be the number of times x appears in D. If we ignore row orders, the design matrix can be represented uniquely by the frequency vector $N(D) = (N_{x_1}, \ldots, N_{x_\tau})$, where $x_1, \ldots, x_\tau$ are all the distinct points in $\mathbb{Z}_{s^p}^m$ arranged in Yates order. We call this vector $N(D)$ the frequency representation of D.

For any $u \in \mathbb{Z}_{s^p}^m$, define $\chi_u(D) = \sum_{x \in D} \chi_u(x)$, where x is a row of D and the summation is over all rows of D. The set of characteristics of D is defined as $\chi(D) = \{\chi_{u_1}(D), \ldots, \chi_{u_\tau}(D)\}$, where $u_1, \ldots, u_\tau$ are all the distinct points in $\mathbb{Z}_{s^p}^m$ in Yates order. The set of all $\chi_u(D)$ fully characterizes the properties of D. The following theorem shows that the set of characteristics and the frequency representation are connected through H.

Theorem 2. *The set of characteristics and the frequency representation uniquely determine each other as follows: $\chi(D) = N(D)H$ and $N(D) = \tau^{-1}\chi(D)H^*$.*

We are now ready to define the space-filling pattern. For $j = 0, \ldots, mp$, define

$$S_j(D) = n^{-2} \sum_{\rho(u)=j} |\chi_u(D)|^2 = n^{-2} \sum_{\rho(u)=j} \chi_u(D)\overline{\chi_u(D)}, \tag{1}$$

where the summation is over all $u \in \mathbb{Z}_{s^p}^m$ with $\rho(u) = j$ and $\overline{\chi_u(D)}$ denotes the complex conjugate of $\chi_u(D)$. It is easy to show that $S_0(D) = 1$. The vector $\{S_1(D), \ldots, S_{mp}(D)\}$ is called the space-filling pattern.

Example 1. Table 3 displays an 8×3 Latin hypercube design D and an SOA$(8, 3, 8, 3)$. The Latin hypercube design is generated from an OA$(8, 3, 2, 3)$ according to Tang (1993), while the SOA$(8, 3, 8, 3)$ is from He & Tang (2014). Here we illustrate how to calculate the space-filling pattern. The entries of the designs are from $\mathbb{Z}_{2^3}$. The mapping functions with weights and the table of characters are given in Tables 1 and 2. The collection of $u \in \mathbb{Z}_{2^3}^3$ with weight 1 is the set $\{(1,\ 0,\ 0), (0,\ 1,\ 0), (0,\ 0,\ 1)\}$, so

$$S_1(D) = n^{-2} \sum_{\rho(u)=1} |\chi_u(D)|^2 = \frac{1}{64}\left(|\chi_{(1,0,0)}(D)|^2 + |\chi_{(0,1,0)}(D)|^2 + |\chi_{(0,0,1)}(D)|^2\right) = 0.$$

The rest of the space-filling pattern can be calculated in a similar way. The sizes of the sets $\{u : u \in \mathbb{Z}_{2^3}^3, \rho(u) = i\}$ for $i = 0, \ldots, 9$ are $1, 3, 9, 25, 42, 72, 104, 96, 96, 64$, respectively. The sum of the sizes is the total number of possible points in $\mathbb{Z}_{2^3}^3$. The space-filling pattern of the Latin hypercube design is $(0, 0, 3, 5, 9, 16, 10, 12, 8)$. Note that $S_i(D) = 0$ for $i = 1, 2$ as all $\chi_u(D) = 0$ for u with $0 < \rho(u) \leqslant 2$. The fact that $S_3(D) = 3$ implies that not all of the $\chi_u(D)$ for u of weight 3 are zeros. Specifically, $\chi_u(D) = 8$ for $u = (1, 2, 0)$, $(1, 0, 2)$ or $(2, 1, 0)$. Furthermore, the space-filling pattern for the SOA$(8, 3, 8, 3)$ is $(0, 0, 0, 12, 6, 13, 12, 12, 8)$. For both designs, $\sum_{j=1}^{9} S_j(D) = 63$.

The following theorem shows that the space-filling pattern captures the strength of a general strong orthogonal array.

THEOREM 3. *A general strong orthogonal array D has strength t if and only if $S_j(D) = 0$ for $1 \leqslant j \leqslant t$.*

Theorem 3 establishes the connection between stratified orthogonality and the space-filling pattern. If the first j elements of the space-filling pattern are zeros, the general strong orthogonal array achieves stratification on any s^j grids from projection. For example, $S_1(D) = 0$ guarantees that there is an equal number of design points when the design region is divided into s equal-volume grids in any one dimension. If $S_1(D) = S_2(D) = 0$, there is an equal number of design points in any s^2 equal-volume grids cut out by projection: either s^2 grids of any one dimension or $s \times s$ grids of any two dimensions.

3.3. *A minimum aberration-type space-filling criterion*

The space-filling pattern describes the space-filling properties sequentially as the design is projected onto margins from coarser to finer. The properties are evaluated in the cluster of projections with respect to the volume of grids rather than the dimensions. The space-filling hierarchy principle suggests that stratifications on larger grids are more likely to be important than stratifications on smaller grids. The minimum aberration-type space-filling criterion selects designs that sequentially minimize the space-filling pattern $S_j(D)$ for $j = 1, \ldots, mp$. Here is a formal definition.

DEFINITION 1. *Suppose that designs D_1 and D_2 have space-filling patterns $\{S_1(D_1), \ldots, S_{mp}(D_1)\}$ and $\{S_1(D_2), \ldots, S_{mp}(D_2)\}$, respectively. If $S_j(D_1) = S_j(D_2)$ for $j = 1, \ldots, l$ and $S_{l+1}(D_1) < S_{l+1}(D_2)$, then D_1 is more space-filling than D_2. Design D_1 is the most space-filling if there is no other design that is more space-filling than D_1.*

Example 2. Recall that the space-filling patterns of the Latin hypercube and SOA$(8, 3, 8, 3)$ in Table 3 are $(0, 0, 3, 4, 9, 16, 10, 12, 8)$ and $(0, 0, 0, 12, 6, 13, 12, 12, 8)$, respectively. According to the space-filling criterion, the SOA$(8, 3, 8, 3)$ is more space-filling as it has $S_3 = 0$, compared with $S_3 = 3$ for the Latin hypercube. Both designs achieve stratifications in each dimension, 2×2 grids in two dimensions and $2 \times 2 \times 2$ grids in three dimensions. Figure 1 presents their 2×4 and 4×2 projection plots. The strong orthogonal array achieves stratifications as each grid has exactly one design point. On the other hand, the Latin hypercube does not have an equal number of points in all 2×4 and 4×2 grids. We highlight the grids that do not have any design points. Our ranking agrees with the conclusion in He & Tang (2013) that Latin hypercubes based on strong orthogonal arrays are more space-filling than comparable Latin hypercubes based on orthogonal arrays.

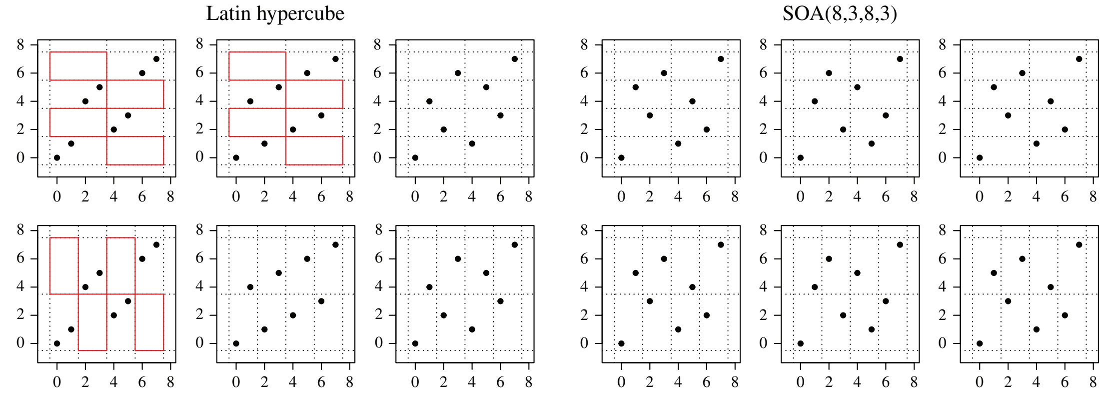

Fig. 1. Projection plots of the Latin hypercube and $\text{SOA}(8, 3, 8, 3)$ in Table 3: 2×4 (upper) and 4×2 (lower).

Values of the space-filling pattern quantify the stratified orthogonality. Leading zeros indicate stratifications on a certain number of equal-volume grids. The first nonzero element reveals how space-filling the design is when projected onto the particular number of grids. Specifically, the $S_j(D)$ value reveals how uniformly the points are distributed when the design is projected onto s^j grids. When two general strong orthogonal arrays have the same strength, the space-filling criterion selects the design that is more space-filling in the next-finer projections.

THEOREM 4. *For a design D with n runs, m columns and s^p levels, the sum of the values of its space-filling pattern has a lower bound:*

$$\sum_{j=1}^{mp} S_j(D) \geqslant \frac{s^{mp}}{n} - 1. \tag{2}$$

The equality holds if and only if D has no replicated points.

Theorem 4 shows that the space-filling pattern characterizes whether a design has replicated points. Designs without replicated points are more space-filling and have smaller $\sum_{j=1}^{mp} S_j(D)$ than designs with replicated points. Latin hypercube designs do not have replicated points, so the equality in (2) always holds for Latin hypercube designs.

Example 3. Consider the $\text{SOA}(32, 9, 8, 3)$ in Shi & Tang (2020). We search over all 36 two-column subarrays and find a total of four distinct space-filling patterns, which differ in (S_4, S_5, S_6). The four distinct patterns of (S_4, S_5, S_6) are $(0, 0, 1)$ for columns $(1, 2)$, denoted by D_1; $(0, 2, 1)$ for columns $(2, 6)$, denoted by D_2; $(1, 1, 1)$ for columns $(1, 8)$, denoted by D_3; and $(2, 0, 1)$ for columns $(1, 7)$, denoted by D_4. For D_1, there are no replicated points and the equality in (2) holds as $\sum_{j=1}^{6} S_j(D_1) = 1$. For the other three designs, there are only 16 distinct points and $\sum_{j=1}^{6} S_j(D) = 3$. The space-filling criterion ranks D_1 as the best, followed by D_2, D_3 and finally D_4. Figure 2 shows the 2×8 and 8×2 projection plots of these four designs. For D_1 and D_2, $S_4 = 0$ guarantees stratification on any 2^4 grids. Design D_1 with $S_5 = 0$ achieves stratifications on 4×8 and 8×4 grids, whereas D_2 does not. Of D_3 and D_4, D_3 is more space-filling and achieves stratification on 2×8 grids, whereas D_4 does not achieve stratifications on 2×8 and 8×2 grids.

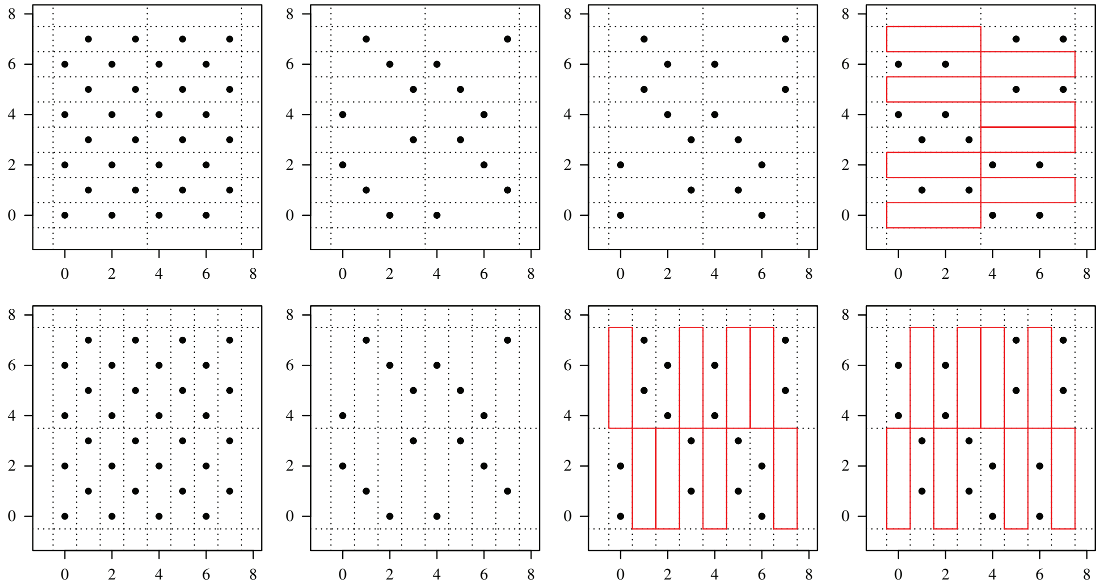

Fig. 2. Projection plots of designs D_1, D_2, D_3 and D_4 (from left to right): 2×8 (upper) and 8×2 (lower).

Example 4. Table 1 in Sun et al. (2019) lists four 25×3 Latin hypercubes: a uniform design, a maximin distance design, a maximum projection design and a uniform projection design. We can compare and rank them using the space-filling criterion. Their space-filling patterns are $(S_1, S_2, S_3, \ldots) = (0, 0.64, 26.08, \ldots), (0, 1.84, 22.48, \ldots), (0, 0.96, 25.12, \ldots)$ and $(0, 0, 28, \ldots)$, respectively. The uniform projection design has strength 2, whereas the other three designs have strength 1. This agrees with the scatterplots in Fig. 1 of Sun et al. (2019). Of these four designs, the uniform projection design is the most space-filling, and the maximin distance design is the least space-filling. The uniform design is more space-filling than the maximum projection design. The ranking of these four designs under the space-filling criterion is consistent with the ranking under the uniform projection criterion used by Sun et al. (2019).

3.4. *Connections with other criteria*

In this subsection we explore the connections between the space-filling criterion and other criteria. When $s = 2$ and $p = 1$, the characteristics $\chi_u(D)$ are referred to as J-characteristics in Tang & Deng (1999) and Tang (2001). For two-level nonregular designs, the space-filling pattern defined in (1) coincides with the generalized wordlength pattern, and the space-filling criterion is equivalent to the minimum G_2-aberration criterion proposed by Tang & Deng (1999).

For general $s \geqslant 2$ and $p = 1$, the characters $\{\chi_u : u \in \mathbb{Z}_s\}$ form the normalized orthogonal contrasts (Xu & Wu, 2001). For $u = (u_1, \ldots, u_m) \in \mathbb{Z}_s^m$, the weight $\rho(u)$ is the Hamming weight of u, i.e., the number of nonzero elements of u. For a design D with n runs, m factors and s levels, Xu & Wu (2001) defined the generalized wordlength pattern $\{A_1(D), \ldots, A_m(D)\}$, where

$$A_j(D) = n^{-2} \sum_{\rho(u)=j} |\chi_u(D)|^2. \tag{3}$$

The generalized wordlength pattern reveals the aliasing structure of the design. Each $A_j(D)$ measures the overall aliasing between all j-factor interactions and the intercept. It also measures the overall aliasing between all $(j-1)$-factor interactions and all main effects. The generalized

Table 4. *Number of distinct space-filling patterns*

Design	Number of columns 2	3	4	5	6	7	8	9
SOA(32, 7, 8, 3)	2	4	4	2	1	1	—	—
SOA(32, 8, 8, 3)	3	5	8	5	3	1	1	—
SOA(32, 9, 8, 3)	4	23	92	121	83	35	7	1

minimum aberration criterion proposed by Xu & Wu (2001) sequentially minimizes the elements in the generalized wordlength pattern.

The definition of $A_j(D)$ in (3) is a special case of the definition of $S_j(D)$ in (1) with $p = 1$, so the generalized wordlength pattern is also a special case of the space-filling pattern. As a result, the space-filling criterion is more general than the generalized minimum aberration criterion.

Despite the similarity of the definitions, it is important to mark the differences between these two concepts. The generalized wordlength pattern and the generalized minimum aberration criterion are defined for selecting factorial designs under an analysis-of-variance model. The generalized wordlength pattern considers the aliasing among factorial effects, whereas the space-filling pattern considers the stratification properties of the designs. The generalized minimum aberration criterion treats the s levels as nominal symbols so that permuting levels for any column does not alter the generalized wordlength pattern; in contrast, the space-filling criterion treats the s^p levels as numerical values so that permuting levels for any column may alter the design stratification structure and the space-filling pattern. For example, consider the two designs in Table 3. When we view them as ordinary eight-level factorial designs with $s = 8$ and $p = 1$, both have the same generalized wordlength pattern $(A_1, A_2, A_3) = (0, 21, 42)$. In contrast, when we view them as general strong orthogonal arrays with $s = 2$ and $p = 3$, they have different space-filling patterns and different strengths; see Examples 1 and 2 and Fig. 1.

Shi & Tang (2019) considered design selection for strong orthogonal arrays of strength 2+. The criterion they used is equivalent to the minimization of $S_3(D)$. Their criterion works only for strong orthogonal arrays constructed from regular designs, whereas our criterion is more general and works for any type of general strong orthogonal array.

4. Applications and comparison of simulations

We first apply the space-filling criterion for selecting and ranking designs. Shi & Tang (2020) considered constructions of strong orthogonal arrays of strength 3. They gave an SOA$(32, m, 8, 3)$ for each of $m = 7, 8, 9$. We conduct an exhaustive search over all subarrays of these three designs. Table 4 presents the numbers of distinct space-filling patterns for all m-column subarrays of these three designs, for $m = 2, \ldots, 9$. The numbers of distinct space-filling patterns vary substantially between different designs. For either SOA$(32, 7, 8, 3)$ or SOA$(32, 8, 8, 3)$, there are only a few distinct space-filling patterns. Subarrays of these designs are clustered with similar space-filling properties. However, for SOA$(32, 9, 8, 3)$ there is a large number of distinct space-filling patterns. For example, there are 121 distinct space-filling patterns among a total of 126 five-column subarrays of SOA$(32, 9, 8, 3)$. For $m = 6, 7$ and 8, the numbers of distinct space-filling patterns are 83, 35 and 7, just 1 or 2 less than the total numbers of all subarrays, which are 84, 36 and 9, respectively. Almost all subarrays have different space-filling properties.

Tables 5–7 list the best space-filling patterns and sets of representative column indices from these three designs. For $m = 2$, the best designs have strength 5, so they achieve stratifications

Table 5. *Best space-filling designs from* SOA$(32, 7, 8, 3)$

m	S_4, S_5, S_6, S_7	Columns	Strength
2	0, 0, 1, —	1, 4	4
3	0, 1, 7, 3	1, 4, 7	4
4	0, 9, 19, 11	1, 2, 4, 7	4
5	1, 22, 40, 40	1, 2, 3, 4, 6	3
6	3, 42, 83, 104	1, 2, 3, 4, 5, 6	3
7	7, 70, 161, 224	1, 2, 3, 4, 5, 6, 7	3

Table 6. *Best space-filling designs from* SOA$(32, 8, 8, 3)$

m	S_4, S_5, S_6, S_7	Columns	Strength
2	0, 0, 1, —	1, 4	5
3	0, 1, 7, 3	1, 4, 6	4
4	0, 8, 20, 12	1, 2, 3, 5	4
5	3, 17, 44, 38	1, 2, 3, 5, 8	3
6	7, 36, 79, 108	1, 2, 3, 4, 5, 8	3
7	13, 62, 143, 248	1, 2, 3, 4, 5, 6, 7	3
8	22, 96, 252, 496	1, 2, 3, 4, 5, 6, 7, 8	3

Table 7. *Best space-filling designs from* SOA$(32, 9, 8, 3)$

m	S_4, S_5, S_6, S_7	Columns	Strength
2	0, 0, 1, —	1, 2	5
3	0, 1, 7, 3	1, 5, 9	4
4	0, 9, 19, 11	3, 7, 8, 9	4
5	3, 16, 42, 46	1, 5, 6, 8, 9	3
6	8, 26, 89, 121	1, 4, 5, 6, 8, 9	3
7	15, 52, 145, 278	1, 2, 4, 5, 6, 8, 9	3
8	27, 80, 248, 546	1, 2, 3, 4, 5, 6, 8, 9	3
9	42, 124, 400, 976	1, 2, 3, 4, 5, 6, 7, 8, 9	3

on 8×4 and 4×8 grids. For $m = 3$ and 4, the best designs have strength 4, so they are strong orthogonal arrays of strength 4−, analogous to the definition of strong orthogonal arrays of strength 3−, and achieve stratifications on 4×4, 8×2, 2×8, $4 \times 2 \times 2$, $2 \times 4 \times 2$ and $2 \times 2 \times 4$ grids, as well as $2 \times 2 \times 2 \times 2$ grids for $m = 4$. For $m = 5$–9, the best designs have strength 3. When $m = 4$–8, the most space-filling design is from either SOA$(32, 7, 8, 3)$ or SOA$(32, 8, 8, 3)$. This indicates that neither SOA$(32, 7, 8, 2)$ nor SOA$(32, 8, 8, 3)$ is a subarray of SOA$(32, 9, 8, 3)$. In the Supplementary Material, we list all possible space-filling patterns of m-column subarrays from SOA$(32, 7, 8, 3)$, SOA$(32, 8, 8, 3)$ and SOA$(32, 9, 8, 3)$ for $m = 2, \ldots, 9$.

We next evaluate the performance of general strong orthogonal arrays and compare them with other types of space-filling designs in building statistical surrogate models. We conduct simulations and generate data from the eight-dimensional borehole function, which has been used by Fang et al. (2006), Chen et al. (2016) and many others. We apply a log-transformation to the response as suggested by Fang et al. (2006). We fit a Gaussian process model with a constant mean and the Gaussian correlation function to approximate the borehole function. To measure

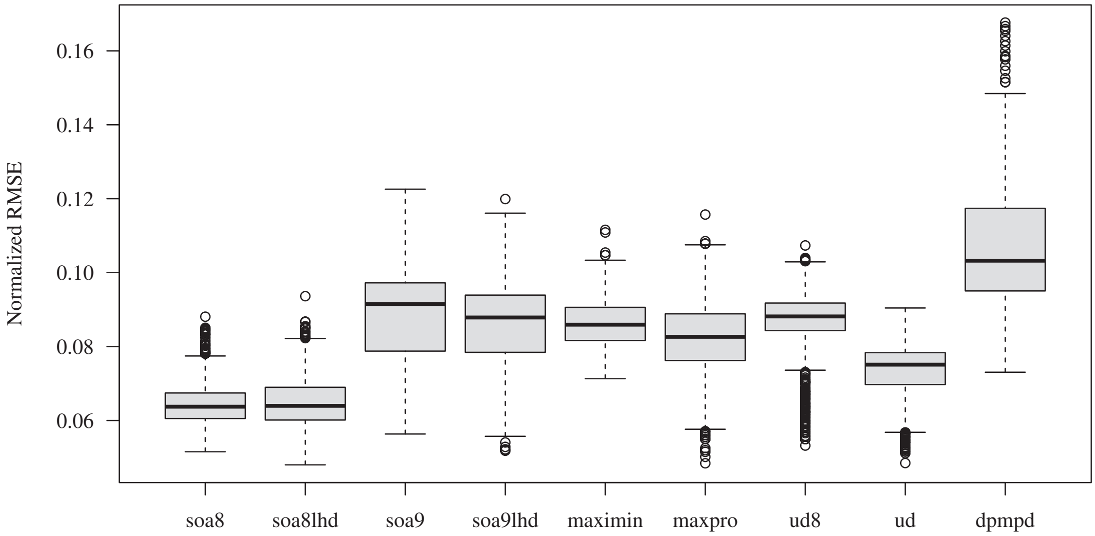

Fig. 3. Normalized root mean square errors for SOA$(32, 8, 8, 3)$ with $S_4 = 22$ (soa8) and its associated Latin hypercube designs (soa8lhd), SOA$(32, 8, 8, 3)$ with $S_4 = 27$ (soa9) and its associated Latin hypercube designs (soa9lhd), maximin Latin hypercube designs (maximin), maximum projection Latin hypercube designs (maxpro), 8-level uniform designs (ud8), 32-level uniform designs (ud), and densest packing-based maximum projection designs (dpmpd).

the prediction error, we use the normalized root mean square error,

$$\text{normalized RMSE} = \left[\frac{N^{-1}\sum_{i=1}^{N}\{\hat{y}(x_i) - y(x_i)\}^2}{N^{-1}\sum_{i=1}^{N}\{\bar{y} - y(x_i)\}^2}\right]^{1/2},$$

where $\{x_1, \ldots, x_N\}$ is a set of N test data points, $y(x_i)$ is the true response at x_i, $\hat{y}(x_i)$ is the predicted response from the Gaussian process model, and $\bar{y}$ is the mean response of the data used to build the model. We generate a test dataset using a random Latin hypercube design with $N = 10\,000$ runs.

We consider two SOA$(32, 8, 8, 3)$s according to Tables 6 and 7, which have $S_4 = 22$ and 27. We also generate random Latin hypercube designs from these two 8-level designs by expanding the 8 levels to 32 levels following Tang (1993). These Latin hypercube designs are GSOA$(32, 8, 32, 3)$s and have the same S_4 values as the corresponding original SOA$(32, 8, 8, 3)$s, while their S_5 values may vary. We also consider four other types of space-filling designs: maximin Latin hypercube designs, maximum projection Latin hypercube designs, uniform designs, and densest packing-based maximum projection designs. The maximin Latin hypercube design and the maximum projection Latin hypercube design are generated using the `R` (R Development Core Team, 2022) packages `SLHD` (Ba et al., 2015) and `MaxPro` (Joseph et al., 2015), respectively. We generate an 8-level and a 32-level uniform design using the `R` package `UniDOE` (Zhang et al., 2018). The densest packing-based maximum projection design is generated using the `R` package `LatticeDesign` (He, 2020). All designs have 32 runs and 8 columns, and each variable is scaled to $[0, 1]$. Given any design, we consider permuting the column labels and reflecting within columns for a random subset of inputs. These operations do not change the design's geometrical structure and space-filling pattern. Figure 3 shows the normalized root mean square errors from 1000 random permutations and reflections for each design. The SOA$(32, 8, 8, 3)$ with $S_4 = 22$ and its associated Latin hypercube designs clearly outperform the other designs, including the SOA$(32, 8, 8, 3)$ with $S_4 = 27$ and its associated Latin hypercube designs. Our new space-filling criterion is capable of selecting efficient space-filling designs for building statistical surrogate models. The simulation also demonstrates that designs with good asymptotic properties, such as the densest packing-based designs and maximin designs, may not work well when the run sizes are not large.

5. Concluding remarks

This work suggests many new research directions. The proposed space-filling criterion has a clear geometrical meaning, but it would be helpful to have additional statistical justifications for the performance and properties of the criterion. The generalized minimum aberration criterion is closely related to various uniformity measures and the maximin criterion when all possible level permutations are considered (Tang et al., 2012; Zhou & Xu, 2014; Fang et al., 2018). It would be interesting to investigate whether these connections could be extended to the new space-filling criterion. One difficulty is that we have to restrict level permutations in order to keep the space-filling pattern invariant. The calculation of the space-filling pattern by the definition (1) is tedious for large designs. A future research problem is to find an efficient calculation method to support the use of our criterion. For designs with the same space-filling pattern, it is worth considering the definition of isomorphism in terms of space-filling properties. Construction of optimal general strong orthogonal arrays might be an important topic to pursue. We hope to develop more theoretical results on general strong orthogonal arrays in the future.

Acknowledgement

The authors thank the editor and two referees for their helpful comments.

Supplementary material

Supplementary Material available at *Biometrika* online includes proofs of the theorems and the search results for all possible space-filling patterns of m-column subarrays from $\text{SOA}(32, 7, 8, 3)$, $\text{SOA}(32, 8, 8, 3)$ and $\text{SOA}(32, 9, 8, 3)$ for $m = 2, \dots, 9$ in Shi & Tang (2020).

References

Ba, S., Myers, W. R. & Brenneman, W. A. (2015). Optimal sliced Latin hypercube designs. *Technometrics* **57**, 479–87.

Chen, H., Loeppky, J. L., Sacks, J. & Welch, W. J. (2016). Analysis methods for computer experiments: How to assess and what counts? *Statist. Sci.* **31**, 40–60.

Cheng, C.-S. (2014). *Theory of Factorial Design: Single- and Multi-Stratum Experiments*. Boca Raton, Florida: CRC Press.

Fang, K.-T., Li, R. & Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Boca Raton, Florida: Chapman and Hall/CRC.

Fang, K.-T., Liu, M.-Q., Qin, H. & Zhou, Y.-D. (2018). *Theory and Application of Uniform Experimental Designs*. Singapore: Springer.

He, X. (2020). Lattice-based designs with quasi-optimal separation distance on all projections. *Biometrika* **108**, 443–54.

He, Y., Cheng, C.-S. & Tang, B. (2018). Strong orthogonal arrays of strength two plus. *Ann. Statist.* **46**, 457–68.

He, Y. & Tang, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100**, 254–60.

He, Y. & Tang, B. (2014). A characterization of strong orthogonal arrays of strength three. *Ann. Statist.* **42**, 1347–60.

Joseph, V. R., Gul, E. & Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika* **102**, 371–80.

LI, W., LIU, M.-Q. & TANG, B. (2020). A method of constructing maximin distance designs. *Biometrika* **108**, 845–55.

LIN, C. D. & TANG, B. (2015). Latin hypercubes and space-filling designs. In *Handbook of Design and Analysis of Experiments*, A. Dean, M. Morris, J. Stufken & D. Bingham, eds. Boca Raton, Florida: Chapman and Hall/CRC, pp. 593–625.

MUKERJEE, R. & WU, C. F. J. (2006). *A Modern Theory of Factorial Design*. New York: Springer.

OWEN, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2**, 439–52.

R DEVELOPMENT CORE TEAM (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

SANTNER, T. J., WILLIAMS, B. J. & NOTZ, W. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer.

SHI, C. & TANG, B. (2019). Design selection for strong orthogonal arrays. *Can. J. Statist.* **47**, 302–14.

SHI, C. & TANG, B. (2020). Construction results for strong orthogonal arrays of strength three. *Bernoulli* **26**, 418–31.

SUN, F., WANG, Y. & XU, H. (2019). Uniform projection designs. *Ann. Statist.* **47**, 641–61.

TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Am. Statist. Assoc.* **88**, 1392–7.

TANG, B. (2001). Theory of J-characteristics for fractional factorial designs and projection justification of minimum G_2-aberration. *Biometrika* **88**, 401–7.

TANG, B. & DENG, L.-Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Ann. Statist.* **27**, 1914–26.

TANG, Y., XU, H. & LIN, D. K. J. (2012). Uniform fractional factorial designs. *Ann. Statist.* **40**, 891–907.

WANG, L., XIAO, Q. & XU, H. (2018). Optimal maximin L_1-distance Latin hypercube designs based on good lattice point designs. *Ann. Statist.* **46**, 3741–66.

WU, C. F. J. & HAMADA, M. S. (2009). *Experiments: Planning, Analysis, and Optimization*. Hoboken, New Jersey: John Wiley & Sons, 2nd ed.

XIAO, Q. & XU, H. (2017). Construction of maximin distance Latin squares and related Latin hypercube designs. *Biometrika* **104**, 455–64.

XIAO, Q. & XU, H. (2018). Construction of maximin distance designs via level permutation and expansion. *Statist. Sinica* **28**, 1395–414.

XU, H. & WU, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Ann. Statist.* **29**, 1066–77.

ZHANG, A., LI, H., QUAN, S. & YANG, Z. (2018). *UniDOE: Uniform Design of Experiments.* R package version 1.0.2, available at https://CRAN.R-project.org/package=UniDOE.

ZHOU, Y. & TANG, B. (2019). Column-orthogonal strong orthogonal arrays of strength two plus and three minus. *Biometrika* **106**, 997–1004.

ZHOU, Y.-D. & XU, H. (2014). Space-filling fractional factorial designs. *J. Am. Statist. Assoc.* **109**, 1134–44.

[*Received on* 17 *October* 2020. *Editorial decision on* 25 *March* 2021]