Metrika
https://doi.org/10.1007/s00184-024-00964-2

Theory and applications of stratification criteria based on space-filling pattern and projection pattern

Xinqi Zhang[1,3] · Yaping Wang[2] · Fasheng Sun[1]

Received: 6 July 2023 / Accepted: 31 March 2024

Abstract
Space-filling designs are crucial for computer experiments. The quality of a space-filling design can be appropriately reflected by its stratification properties. In a recent paper, Tian and Xu (Biometrika 109(2):489–501, 2022) introduced the concept of a space-filling pattern to properly characterize a design's stratification properties on various grids. In this study, we generalize the space-filling pattern using arbitrary orthonormal contrasts. We also propose a new pattern called the two-dimensional projection pattern to capture the stratification properties of balanced designs in two dimensions more comprehensively. We derive some theoretical results for both patterns and show that they are easier to compute and apply to a wider range of designs. We further show the use of the two patterns in constructing space-filling designs based on existing strong orthogonal arrays.

Keywords Computer experiment · Stratification property · Strong orthogonal array · Space-filling design

1 Introduction

Computer experiments are widely used in scientific research and product development. It is widely accepted that space-filling designs, which can fill the experimental

Xinqi Zhang and Yaping Wang have contributed equally to this work.

✉ Fasheng Sun
sunfs359@nenu.edu.cn

1 KLAS and School of Mathematics and Statistics, Northeast Normal University, No.5268, Renmin Street, Changchun 130024, China

2 KLATASDS–MOE, School of Statistics, East China Normal University, No.3663, North Zhongshan Road, Shanghai 200062, China

3 Department of Science and Technology, Hulunbuir University, No.26, Middle Genghis Khan Road, Hailar District, Hulunbuir 021008, China

Published online: 22 April 2024

region evenly, are appropriate choices for computer experiments (Santner et al. 2018). Among the methods available for constructing space-filling designs, the one based on orthogonal arrays is particularly attractive because it produces designs that enjoy some guaranteed space-filling properties in low-dimensional projections (Mckay et al. 1979; Owen 1992; Tang 1993). Strong orthogonal arrays proposed by He and Tang (2012) are a watershed moment in this line of research. The idea behind strong orthogonal arrays inspired researchers to study a design's space-filling properties by its stratifications in low-dimensional projections. Recent years have witnessed a growing amount of interest in this topic. Most of the studies focused on the construction and selection of strong orthogonal arrays; see He and Tang (2014), Liu and Liu (2015), He et al. (2018), Zhou and Tang (2019), Shi and Tang (2019, 2020), Cheng et al. (2020) among others.

It is crucial to have a good criterion for evaluating and ranking various strong orthogonal arrays and other space-filling designs with similar structures. Tian and Xu (2022) introduced the concept of a space-filling pattern, which accurately describes the stratification properties of designs on various grids. Based on the space-filling pattern, they proposed a minimum aberration type criterion for evaluating the space-filling properties of designs. This new criterion allows any two designs to be compared according to their space-filling pattern. However, their space-filling pattern is defined through complex contrasts, and the calculation is cumbersome, especially for designs with large run size or number of factors. Moreover, the space-filling pattern and its related minimum aberration type criterion are not adequate for capturing and ranking strong orthogonal arrays of strength 2+ proposed by He et al. (2018).

In this paper, we propose a new definition of the space-filling pattern that is equivalent to Tian and Xu (2022)'s but avoids computing with complex numbers. The new definition is based on general orthonormal contrasts. It is more flexible and general, applicable to a wider class of designs, and covers the original definition as a special case. We also introduce a variant of the space-filling pattern, which we call the two-dimensional projection pattern, to characterize the two-dimensional projection properties of balanced designs with $\alpha \times s$ levels more elaborately. The two-dimensional projection pattern is easy to compute and can be used to compare large designs effectively. We derive some theoretical properties of the space-filling pattern and the two-dimensional projection pattern. Furthermore, we demonstrate two of their applications. First, we use the two patterns to generate some column-expanded designs based on existing strong orthogonal arrays. These designs can accommodate more factors than the original strong orthogonal arrays and perform well under the minimum aberration type criterion based on the space-filling pattern or the two-dimensional projection pattern. Second, we use the space-filling pattern as a criterion to rank and select subarrays from some existing strong orthogonal arrays.

The remainder of this paper is organized as follows. Section 2 presents some basic notation and background. Section 3 introduces a new definition of the space-filling pattern based on general orthonormal contrasts, as well as the concept of a two-dimensional projection pattern, which better captures the two-dimensional stratification properties of a design. The next two sections are devoted to two applications of the results in Sect. 3. Based on existing strong orthogonal arrays, Sect. 4 constructs optimal column-expanded designs, and Sect. 5 selects optimal subarrays. Section 6 concludes the paper with a discussion.

2 Notation and background

For an integer $s \geq 2$, let $\mathbb{Z}_s = \{0, 1, \ldots, s-1\}$. An $n \times m$ matrix D with entries from $\mathbb{Z}_{s_j} = \{0, 1, \ldots, s_j - 1\}$ in the j-th column is said to be an asymmetrical (or mixed-level) design of n runs, m factors with the j-th factor having s_j levels, $j = 1, 2, \ldots, m$. We denote such a design as $D \in \mathcal{D}(n, s_1 \cdots s_m)$, or $D \in \mathcal{D}(s_1 s_2 \cdots s_m)$ if we consider only the levels of the design. When $s_1 = \cdots = s_m = s$, $D \in \mathcal{D}(n, s_1 \cdots s_m)$ is said to be symmetrical and simply denoted by $D \in \mathcal{D}(n, s^m)$. A design $D \in \mathcal{D}(n, s_1 \cdots s_m)$ is called an orthogonal array (OA) of strength t if all possible level combinations appear with the same frequency in any of its $n \times t$ submatrices. We use $\mathrm{OA}(n, m, s_1 \times \cdots \times s_m, t)$ to denote such an array; for the symmetrical case we simply use the notation $\mathrm{OA}(n, m, s, t)$. In particular, we call $D \in \mathcal{D}(n, s_1 \cdots s_m)$ a full factorial design or full design if it is an $\mathrm{OA}(n, m, s_1 \times \cdots \times s_m, m)$, and we call D a balanced design if it is an $\mathrm{OA}(n, m, s_1 \times \cdots \times s_m, 1)$. We consider only balanced designs in this paper. Two columns of the same length are called (combinatorially) orthogonal if they form a two-factor full design.

Tian and Xu (2022) defined the concept of general strong orthogonal arrays (GSOAs), which covers a broad class of existing space-filling designs. We adopt this concept to simplify our notation. A design $D \in \mathcal{D}(n, (s^p)^m)$ is called a GSOA of strength t and denoted by $\mathrm{GSOA}(n, m, s^p, t)$, if any g-column subarray where $1 \leq g \leq t$ can be collapsed into an $\mathrm{OA}(n, g, s^{u_1} \times \cdots \times s^{u_g}, g)$ for any positive integers $u_1, \ldots, u_g$ satisfying $u_1 + \cdots + u_g = t$ and $u_i \leq p$ for $i = 1, \ldots, g$. Throughout, collapsing s^p levels into s^{u_i} levels is done by $\lfloor x/s^{p-u_i} \rfloor$ for $x \in \mathbb{Z}_{s^p}$. By definition, GSOAs of strength t achieve stratification on grids of volume s^t regardless of how the design space is divided into s^t equal-volume grids from projection (Tian and Xu 2022). Specifically, when $p = t$, i.e., the number of levels equals s^t, $\mathrm{GSOA}(n, m, s^t, t)$ is a strong orthogonal array (SOA) of strength t defined in He and Tang (2012) and is also denoted by $\mathrm{SOA}(n, m, s^t, t)$. When $p = 2$ and $t = 3$, $\mathrm{GSOA}(n, m, s^2, 3)$ is an SOA of strength 3− defined in Zhou and Tang (2019) and is also denoted by $\mathrm{SOA}(n, m, s^2, 3-)$. When $p = 1$, $\mathrm{GSOA}(n, m, s^1, t)$ is an ordinary $\mathrm{OA}(n, m, s, t)$.

We also study SOAs of strength 2+ proposed by He et al. (2018). A design $D \in \mathcal{D}(n, (s^2)^m)$ is called a strong orthogonal array of strength 2+ and denoted by $\mathrm{SOA}(n, m, s^2, 2+)$, if any two columns of D can be collapsed into an $\mathrm{OA}(n, 2, s^2 \times s, 2)$ and an $\mathrm{OA}(n, 2, s \times s^2, 2)$. SOAs of strength 2+ are GSOAs of strength 2, but with better two-dimensional stratification properties. They can achieve stratifications on $s^2 \times s$ and $s \times s^2$ grids in any two dimensions like GSOAs of strength 3, but they do not need to achieve stratifications on $s \times s \times s$ grids in any three dimensions. Such SOAs are desirable in applications because they are more space-filling than comparable OAs in two-dimensional projections and can accommodate more factors than SOAs of the same run sizes and higher strength.

3 Theoretical results for stratification criteria

3.1 Defining the space-filling pattern by using general orthonormal contrasts

Let $D \in \mathcal{D}(n, (s^p)^m)$ be a design with s^p levels. Tian and Xu introduced a space-filling pattern in Tian and Xu (2022) to characterize the stratification properties of D and proposed a minimum aberration type space-filling criterion based on the space-filling pattern. They used complex contrasts to define their space-filling pattern. In this subsection, we redefine the space-filling pattern of D using arbitrary orthonormal contrasts, which avoids complex number calculations. The new definition is more general and flexible than Tian and Xu (2022)'s and covers it as a special case. It also has the potential for generalization (e.g., it can be easily extended to designs with other numbers of levels; see Sect. 3.2).

For a given s, let $C_0^s(x), C_1^s(x), \ldots, C_{s-1}^s(x)$ be a set of functions on $\mathbb{Z}_s$ such that

$$\sum_{x \in \mathbb{Z}_s} C_u^s(x)\overline{C_v^s(x)} = \begin{cases} 0, & \text{if } u \neq v, \\ s, & \text{if } u = v, \end{cases} \tag{1}$$

where $\overline{C_v^s(x)}$ is the complex conjugate of $C_v^s(x)$, and $\overline{C_v^s(x)} = C_v^s(x)$ for the case of real contrasts. We call $C_0^s(x), C_1^s(x), \ldots, C_{s-1}^s(x)$ a set of orthonormal contrasts of order s. We make the convention $C_0^s(x) = 1$ for any $x \in \mathbb{Z}_s$.

For $x \in \mathbb{Z}_{s^p}$, let $f_i(x)$ be the i-th digit of x in the base-s numeral system. Then, $f_i(x) = \lfloor x/s^{p-i} \rfloor \pmod s$ and $x = \sum_{i=1}^p f_i(x)s^{p-i}$, where $\lfloor x \rfloor$ denotes the largest integer not exceeding x. Let

$$\rho(x) = \begin{cases} p + 1 - \min\{i : f_i(x) \neq 0, i = 1, \ldots, p\}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases} \tag{2}$$

be the weight defined in Tian and Xu (2022). The weight $\rho(x)$ counts the number of digits needed to express x in the base-s numeral system after ignoring all the leading zeros. For $x, u \in \mathbb{Z}_{s^p}$, define character

$$\chi_u(x) = C_{f_1(u)}^s\left(f_p(x)\right) \times C_{f_2(u)}^s\left(f_{p-1}(x)\right) \times \cdots \times C_{f_p(u)}^s\left(f_1(x)\right). \tag{3}$$

For vectors $\boldsymbol{u} = (u_1, \ldots, u_m), \boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{Z}_{s^p}^m$, define the character $\chi_{\boldsymbol{u}}(\boldsymbol{x}) = \prod_{i=1}^m \chi_{u_i}(x_i)$ and the weight $\rho(\boldsymbol{u}) = \sum_{i=1}^m \rho(u_i)$. For a design $D \in \mathcal{D}(n, (s^p)^m)$, let $\chi_{\boldsymbol{u}}(D) = \sum_{\boldsymbol{x} \in D} \chi_{\boldsymbol{u}}(\boldsymbol{x})$, where $\boldsymbol{x} \in D$ means $\boldsymbol{x}$ is a row of D and the summation $\sum_{\boldsymbol{x} \in D}$ is over all rows of D. Let $\tau = s^{mp}$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_\tau$ denote all possible $\boldsymbol{u} \in \mathbb{Z}_{s^p}^m$ in Yates order. The vector $\chi(D) = (\chi_{\boldsymbol{u}_1}(D), \ldots, \chi_{\boldsymbol{u}_\tau}(D))$ is called the characteristic of design D.

Definition 1 Let $D \in \mathcal{D}(n, (s^p)^m)$. For $r = 1, \ldots, mp$, let us define

$$S_r(D) = \frac{1}{n^2} \sum_{\rho(\boldsymbol{u})=r} |\chi_{\boldsymbol{u}}(D)|^2 = \frac{1}{n^2} \sum_{\rho(\boldsymbol{u})=r} \chi_{\boldsymbol{u}}(D)\overline{\chi_{\boldsymbol{u}}(D)}, \tag{4}$$

where the summation is over all $\boldsymbol{u} \in \mathbb{Z}_{s^p}^m$ with $\rho(\boldsymbol{u}) = r$ and $\overline{\chi_{\boldsymbol{u}}(D)}$ is the complex conjugate of $\chi_{\boldsymbol{u}}(D)$. The vector $\left(S_1(D), \ldots, S_{mp}(D)\right)$ is called the space-filling pattern of the design D.

Theorem 1 *The space-filling pattern in Definition 1 is independent of the choice of orthonormal contrasts.*

Proof of Theorem 1 For any $D \in \mathcal{D}(n, (s^p)^m)$, there exists a unique set of p designs $D_1, \ldots, D_p \in \mathcal{D}(n, s^m)$ such that $D = \sum_{k=1}^p s^{p-k} D_k$. Let $D^* = (d_1^{(1)}, \ldots, d_m^{(1)}, \ldots, d_1^{(p)}, \ldots, d_m^{(p)})$, where $d_i^{(k)}$ is the i-th column of D_k for $k = 1, \ldots, p$. Given any orthonormal contrast, we take the space-filling pattern defined by that contrast in Definition 1. From the definitions of the space-filling pattern in Definition 1 and the generalized wordlength pattern proposed by Xu and Wu (2001), each $S_r(D)$, $r = 1, \ldots, mp$, is a linear combination of the generalized wordlength pattern of D^*'s subarrays. For example, $S_2(D) = A_2(D_1)$ and $S_3(D) = A_3(D_1) + \sum_{1 \le i \ne j \le m} [A_3(d_i^{(1)}, d_i^{(2)}, d_j^{(1)}) + A_2(d_i^{(2)}, d_j^{(1)})] + \sum_{i=1}^m A_2(d_i^{(1)}, d_i^{(2)})$. Because the generalized wordlength pattern of a design is independent of the choice of the orthonormal contrast (see, Xu and Wu 2001; Cheng and Ye 2004), so is the space-filling pattern. □

Among all designs in $\mathcal{D}(n, (s^p)^m)$, the minimum aberration space-filling criterion proposed by Tian and Xu (2022) is to select designs that sequentially minimize the space-filling pattern $S_r(D)$ for $r = 1, \ldots, mp$.

The character defined in (3) follows the inverse inner product manner used in Tian and Xu (2022). In this paper, we also consider an alternative definition of weight (2) for $x \in \mathbb{Z}_{s^p}$:

$$\rho'(x) = \begin{cases} \max\{i : f_i(x) \neq 0, i = 1, \ldots, p\}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases} \tag{5}$$

Here, $\rho'(x)$ counts the number of digits needed to express x in the base-s numeral system after removing all the trailing zeros. For example, for $x = 3 \in \mathbb{Z}_{2^3}$, $f_1(x) = 0$, $f_2(x) = f_3(x) = 1$, $\rho(x) = 2$ and $\rho'(x) = 3$. We show that by using the weight ρ', we can skip the "inverse inner product" process in Tian and Xu (2022) without changing the values of the space-filling pattern. Specifically, for $x, u \in \mathbb{Z}_{s^p}$, when the weight $\rho'(x)$ in (5) is used, we can define the character as

$$\chi'_u(x) = C^s_{f_1(u)}(f_1(x)) \times C^s_{f_2(u)}(f_2(x)) \times \cdots \times C^s_{f_p(u)}\left(f_p(x)\right). \tag{6}$$

Correspondingly, for vectors $\boldsymbol{u} = (u_1, \ldots, u_m), \boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{Z}_{s^p}^m$, we define character $\chi'_{\boldsymbol{u}}(\boldsymbol{x}) = \prod_{i=1}^m \chi'_{u_i}(x_i)$ and weight $\rho'(\boldsymbol{u}) = \sum_{i=1}^m \rho'(u_i)$.

Proposition 2 *Let $D \in \mathcal{D}(n, (s^p)^m)$. For $r = 1, \ldots, mp$, let us define*

$$S'_r(D) = \frac{1}{n^2} \sum_{\rho'(\boldsymbol{u})=r} |\chi'_{\boldsymbol{u}}(D)|^2 = \frac{1}{n^2} \sum_{\rho'(\boldsymbol{u})=r} \chi'_{\boldsymbol{u}}(D)\overline{\chi'_{\boldsymbol{u}}(D)},$$

where the summation is over all $\boldsymbol{u} \in \mathbb{Z}_{s^p}^m$ with $\rho'(\boldsymbol{u}) = r$, $\rho'(\cdot)$ is defined in (5), $\overline{\chi'_{\boldsymbol{u}}(D)}$ is the complex conjugate of $\chi'_{\boldsymbol{u}}(D)$, and $\chi'_u(\cdot) = \sum_{\boldsymbol{x} \in D} \prod_{i=1}^m \chi'_{u_i}(x_i)$. Then, $S'_r(D) = S_r(D)$, $r = 1, \ldots, mp$, where $S_r(D)$ is defined in (4).

Proof of Proposition 2 For $D \in \mathcal{D}(n, (s^p)^m)$, from the definitions of $S_r(D)$, $S'_r(D)$, $\rho(\boldsymbol{u})$, $\rho'(\boldsymbol{u})$ and the fact $C_0^s(x) = 1$, we have

$$
\begin{aligned}
S_r(D) &= \frac{1}{n^2} \sum_{\rho(\boldsymbol{u})=r} \left| \sum_{\boldsymbol{x} \in D} \prod_{i=1}^m C^s_{f_1(u_i)}(f_p(x_i)) \cdots C^s_{f_p(u_i)}(f_1(x_i)) \right|^2 \\
&= \frac{1}{n^2} \sum_{\rho(\boldsymbol{u})=r} \left| \sum_{\boldsymbol{x} \in D} \prod_{i=1}^m C^s_{f_1(u_i)}(f_r(x_i)) \cdots C^s_{f_p(u_i)}(f_1(x_i)) \right|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
S'_r(D) &= \frac{1}{n^2} \sum_{\rho'(\boldsymbol{u})=r} \left| \sum_{\boldsymbol{x} \in D} \prod_{i=1}^m C^s_{f_1(u_i)}(f_1(x_i)) \cdots C^s_{f_p(u_i)}(f_p(x_i)) \right|^2 \\
&= \frac{1}{n^2} \sum_{\rho'(\boldsymbol{u})=r} \left| \sum_{\boldsymbol{x} \in D} \prod_{i=1}^m C^s_{f_1(u_i)}(f_1(x_i)) \cdots C^s_{f_r(u_i)}(f_r(x_i)) \right|^2.
\end{aligned}
$$

The sums $\sum_{\rho(\boldsymbol{u})=r}$ and $\sum_{\rho'(\boldsymbol{u})=r}$ are taken over $\{\boldsymbol{u} = (u_1, \ldots, u_m) \in \mathbb{Z}_{s^p}^m : \sum_{k=1}^m \rho(u_i) = r\}$ and $\{\boldsymbol{u} = (u_1, \ldots, u_m) \in \mathbb{Z}_{s^p}^m : \sum_{k=1}^m \rho'(u_i) = r\}$, respectively. These two sets can be written as $\{\boldsymbol{u} = (u_1, \ldots, u_m) \in \mathbb{Z}_{s^p}^m : \rho(u_1) = r_1, \ldots, \rho(u_m) = r_m, \sum_{k=1}^m r_k = r\}$ and $\{\boldsymbol{u} = (u_1, \ldots, u_m) \in \mathbb{Z}_{s^p}^m : \rho'(u_1) = r_1, \ldots, \rho'(u_m) = r_m, \sum_{k=1}^m r_k = r\}$. For fixed $r_1, \ldots, r_m$ with $\sum_{k=1}^m r_k = r$, we have $\{u_k : \rho(u_k) = r_k\} = \{u_k : f_{p-r_k+1}(u_k) \neq 0, f_1(u_k) = \cdots = f_{p-r_k}(u_k) = 0\}$ and $\{u_k : \rho'(u_k) = r_k\} = \{u_k : f_{r_k}(u_k) \neq 0, f_{r_k+1}(u_k) = \cdots = f_p(u_k) = 0\}$. Thus, for $k = 1, \ldots, m$, $f_{p-r_k+1}(u_k)$ takes $1, \ldots, s-1$, and $f_{p-r_k+2}(u_k), \ldots, f_p(u_k)$ take $0, 1, \ldots, s-1$ when calculating $S_r(D)$. Similarly, $f_{r_k}(u_k)$ takes $1, \ldots, s-1$, and $f_1(u_k), \ldots, f_{r_k-1}(u_k)$ take $0, 1, \ldots, s-1$ when calculating $S'_r(D)$. These facts imply $S_r(D) = S'_r(D)$. □

Definition 1 and Proposition 2 give two equivalent definitions of the space-filling pattern. Henceforth, we do not distinguish between $S_r(D)$ and $S'_r(D)$, $\rho(\cdot)$ and $\rho'(\cdot)$, and $\chi(\cdot)$ and $\chi'(\cdot)$, and we use the common notation $S_r(D)$, $\rho(\cdot)$ and $\chi(\cdot)$.

Remark 1 Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\tau$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_\tau$ denote all possible $\boldsymbol{x}, \boldsymbol{u} \in \mathbb{Z}_{s^p}^m$ in Yates order, where $\tau = s^{mp}$. Let $H = \left(\chi_{\boldsymbol{u}_j}(\boldsymbol{x}_i)\right)$ be the $\tau \times \tau$ matrix of characters evaluated at all possible points in $\mathbb{Z}_{s^p}^m$.

(i) Both the first row and column of H are vectors of ones. The character matrix H is symmetrical because $\chi_{\boldsymbol{u}}(\boldsymbol{x}) = \chi_{\boldsymbol{x}}(\boldsymbol{u})$.

(ii) Let $P = \underbrace{C_s \otimes \cdots \otimes C_s}_{mp}$, where $\otimes$ denotes the Kronecker product and

$$C_s = \begin{pmatrix} C_0^s(0) & C_1^s(0) & \cdots & C_{s-1}^s(0) \\ C_0^s(1) & C_1^s(1) & \cdots & C_{s-1}^s(1) \\ \vdots & \vdots & \ddots & \vdots \\ C_0^s(s-1) & C_1^s(s-1) & \cdots & C_{s-1}^s(s-1) \end{pmatrix}.$$

Then, $H = P$ when the columns and rows of P are properly permutated, and $HH^* = PP^* = \tau I$, where I is the identity matrix of order τ and H^* is the conjugate transpose of H.

(iii) If we take

$$C_s = \begin{pmatrix} C_0^s(0) & C_1^s(0) & \cdots & C_{s-1}^s(0) \\ C_0^s(1) & C_1^s(1) & \cdots & C_{s-1}^s(1) \\ \vdots & \vdots & \ddots & \vdots \\ C_0^s(s-1) & C_1^s(s-1) & \cdots & C_{s-1}^s(s-1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{(s-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{(s-1)} & \omega^{2(s-1)} & \cdots & \omega^{(s-1)(s-1)} \end{pmatrix}, \tag{7}$$

where $\omega = e^{2\pi i/s}$ is the primitive sth root of unity and $i = (-1)^{1/2}$, then Definition 1 reduces to the definition of the space-filling pattern in Tian and Xu (2022).

Different orthonormal contrasts $\{C_0^s(\cdot), C_1^s(\cdot), \dots, C_{s-1}^s(\cdot)\}$ generate different character matrices, but the space-filling pattern of D is always invariant by Theorem 1. Compared to the original definition of the space-filling pattern based on complex contrasts by Tian and Xu (2022), the new definition is based on a class of general orthonormal contrasts and thus is more flexible. In practice, we can choose some real orthonormal contrasts to facilitate the calculation of the space-filling pattern in (4). For example, we can use

$$C_3 = \begin{pmatrix} C_0^3(0) & C_1^3(0) & C_2^3(0) \\ C_0^3(1) & C_1^3(1) & C_2^3(1) \\ C_0^3(2) & C_1^3(2) & C_2^3(2) \end{pmatrix} = \begin{pmatrix} 1 & -\sqrt{6}/2 & \sqrt{2}/2 \\ 1 & 0 & -\sqrt{2} \\ 1 & \sqrt{6}/2 & \sqrt{2}/2 \end{pmatrix} \text{ for } s = 3 \tag{8}$$

instead of (7) with $\omega = e^{2\pi i/3}$ to avoid complex number calculations.

Because the space-filling pattern defined by (4) is the same as Tian and Xu (2022)'s by Theorem 1, the properties shown in their paper all hold. We rephrase three of them as a proposition.

Proposition 3 *Let $D \in \mathcal{D}(n, (s^p)^m)$, then*

(i) the sum of the space-filling pattern has a lower bound

$$\sum_{r=1}^{mp} S_r(D) \geq \frac{s^{mp}}{n} - 1$$

with the equality holds if and only if D has no replicated points;

(ii) $S_r(D) = A_r(D)$ when $p = 1$, where $(A_1(D), \ldots, A_m(D))$ is the generalized wordlength pattern defined in Xu and Wu (2001);

(iii) D is a $\mathrm{GSOA}(n, m, s^p, t)$ if and only if $S_r(D) = 0$ for $1 \leq r \leq t$.

3.2 A new criterion based on two-dimensional projection pattern

The space-filling pattern defined in Sect. 3.1 reflects the stratification properties of a design. If the first r entries of the space-filling pattern are zeros, then a design achieves stratification on any s^r grids from projection. However, the space-filling pattern does not capture some specific stratification properties. For instance, for a design $D \in \mathcal{D}(n, (s^p)^m)$, if $S_3(D) > 0$, we cannot tell by the values of $S_1(D)$, $S_2(D)$, $S_3(D)$ whether D achieves stratification on $s^2 \times s$ or $s \times s^2$ grids in two dimensions. In addition, the calculation of the space-filling pattern is time-consuming, especially when mp is large. Because obtaining the space-filling pattern by (1) requires $O(nmps^{mp})$ operations to go over all $u \in \mathbb{Z}_{s^p}^m$. However, the comparison of two designs can often be done by a few leading entries of the space-filling patterns in practice.

Therefore, in this subsection, we propose a new criterion based on a variant of the space-filling pattern. Both SOAs of strength 3 and strength 2+ have superior two-dimensional projection stratification properties than comparable OAs. We hope that the new criterion will assist in evaluating and ranking designs based on these properties. Furthermore, improving the three or higher dimensional projections of the design often comes at the cost of economy. We prefer designs that exhibit superior two-dimensional projections while accommodating more factors. Thus, the new criterion will focus on the two-dimensional projection stratification properties of the design.

We define the new criterion for designs in $\mathcal{D}(n, (\alpha s)^m)$, where α and s are two positive integers. A design $D \in \mathcal{D}(n, (\alpha s)^m)$ is called an SOA of strength 2+ with αs levels, denoted by $\mathrm{SOA}_\alpha(n, m, \alpha s, 2+)$, if any two columns of D can be collapsed into an $\mathrm{OA}(n, 2, (\alpha s) \times s, 2)$ and an $\mathrm{OA}(n, 2, s \times (\alpha s), 2)$. Here, collapsing αs levels into s levels is done by $\lfloor x/\alpha \rfloor$ for $x \in \mathbb{Z}_{\alpha s}$. In other words, such designs can achieve stratifications on $(\alpha s) \times s$ and $s \times (\alpha s)$ grids in any two dimensions. This class of SOAs was first proposed by He et al. (2018) as a generalization of $\mathrm{SOA}(n, m, s^2, 2+)$. An $\mathrm{SOA}(n, m, s^2, 2+)$ is a special case of an $\mathrm{SOA}_\alpha(n, m, \alpha s, 2+)$ with $\alpha = s$. The following lemma gives a useful characterization for $\mathrm{SOA}_\alpha(n, m, \alpha s, 2+)$.

Lemma 4 *An $\mathrm{SOA}_\alpha(n, m, \alpha s, 2+)$, say D, exists if and only if there exist $A = (a_1, \ldots, a_m)$ in $\mathcal{D}(n, s^m)$ and $B = (b_1, \ldots, b_m)$ in $\mathcal{D}(n, \alpha^m)$ where a_i and b_i denote the i-th column of A and B, such that A is an $\mathrm{OA}(n, m, s, 2)$ and (a_i, b_i, a_j) is an*

OA$(n, 3, s \times \alpha \times s, 3)$ *for all* $1 \leq i \neq j \leq m$. *The three arrays* A, B *and* D *are related through* $D = \alpha A + B$.

For $x \in \mathbb{Z}_{\alpha s}$, let $g_1(x) = \lfloor x/\alpha \rfloor$ and $g_2(x) = x \pmod{\alpha}$. We define weight

$$\rho(x) = \begin{cases} \max\{i : g_i(x) \neq 0, i = 1, 2\}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases} \tag{9}$$

which is a generalization of the weight $\rho'(x)$ for $x \in \mathbb{Z}_{s^p}$ defined in (5). For $x, u \in \mathbb{Z}_{\alpha s}$, let us define the character $\chi_u(x)$ as

$$\chi_u(x) = C^s_{g_1(u)}(g_1(x)) C^\alpha_{g_2(u)}(g_2(x)), \tag{10}$$

where $\{C^s_0(\cdot), \ldots, C^s_{s-1}(\cdot)\}$ and $\{C^\alpha_0(\cdot), \ldots, C^\alpha_{\alpha-1}(\cdot)\}$ are the sets of orthogonal contrasts of order s and α, respectively. Here, (10) is a generalization of the character $\chi'_u(x)$ for $x \in \mathbb{Z}_{s^p}$ defined in (6). Correspondingly, for vectors $\boldsymbol{u} = (u_1, \ldots, u_m)$, $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{Z}^m_{\alpha s}$, let us define character $\chi_{\boldsymbol{u}}(\boldsymbol{x}) = \prod_{i=1}^m \chi_{u_i}(x_i)$, and for design $D \in \mathcal{D}(n, (\alpha s)^m)$, define $\chi_{\boldsymbol{u}}(D) = \sum_{\boldsymbol{x} \in D} \chi_{\boldsymbol{u}}(\boldsymbol{x})$.

Now, we are ready to define the two-dimensional projection pattern for $D \in \mathcal{D}(n, (\alpha s)^m)$.

Definition 2 Let $D \in \mathcal{D}(n, (\alpha s)^m)$. Let $U_{a,b} = \{\boldsymbol{u} : \boldsymbol{u} \in \mathbb{Z}^m_{\alpha s}, \rho(u_i) = a, \rho(u_j) = b \text{ for } 1 \leq i \neq j \leq m, u_k = 0 \text{ for } k \neq i, j\}$, where $a, b \in \{1, 2\}$, $\rho(\cdot)$ is the weight defined in (9). Let

$$S^{(a,b)}(D) = \frac{1}{n^2} \sum_{\boldsymbol{u} \in U_{a,b}} |\chi_{\boldsymbol{u}}(D)|^2 = \frac{1}{n^2} \sum_{\boldsymbol{u} \in U_{a,b}} \chi_{\boldsymbol{u}}(D) \overline{\chi_{\boldsymbol{u}}(D)}, \tag{11}$$

where the summation is over all $\boldsymbol{u} \in U_{a,b}$, $\overline{\chi_{\boldsymbol{u}}(D)}$ is the complex conjugate of $\chi_{\boldsymbol{u}}(D)$ and $\chi(\cdot)$ is defined in (10). The vector $\left(S^{(1,1)}(D), S^{(1,2)}(D), S^{(2,2)}(D)\right)$ is called the *two-dimensional projection pattern* of the design D.

Example 1 Consider the SOA$_2(18, 4, 6, 2+)$ shown in He et al. (2018), the levels 2 and 4 are of weight 1, the levels 1, 3 and 5 are of weight 2. Therefore, $U_{1,1}$ contains $(2, 2, 0, 0)$, $(2, 4, 0, 0)$, $(4, 4, 0, 0)$ and 21 other $\boldsymbol{u}$'s with the same entries as these 3 vectors but in different positions. $U_{1,2}$ contains $(2, 1, 0, 0)$, $(2, 3, 0, 0)$, $(2, 5, 0, 0)$, $(4, 1, 0, 0)$, $(4, 3, 0, 0)$, $(4, 5, 0, 0)$ and 66 other $\boldsymbol{u}$'s with the same entries as these 6 vectors but in different positions. $U_{2,2}$ contains $(1, 1, 0, 0)$, $(1, 3, 0, 0)$, $(1, 5, 0, 0)$, $(3, 3, 0, 0)$, $(3, 5, 0, 0)$, $(5, 5, 0, 0)$ and 48 other $\boldsymbol{u}$'s with the same entries as these 6 vectors but in different positions. Then, by the definitions of given before, we have $S^{(1,1)}(D) = S^{(1,2)}(D) = 0$, $S^{(2,2)}(D) = 6$.

From Definition 2, $S^{(1,1)}(D)$, i.e., the first entry of the two-dimensional projection pattern, evaluates the stratification properties of D on $s \times s$ grids in any two dimensions. The second entry satisfies $S^{(1,2)}(D) = S^{(2,1)}(D)$. The sum of the first two entries $S^{(1,1)}(D) + S^{(1,2)}(D)$ evaluates the stratification properties of D on $s \times (\alpha s)$ and $(\alpha s) \times$

s grids in any two dimensions. The sum of all three entries $S^{(1,1)}(D) + S^{(1,2)}(D) + S^{(2,2)}(D)$ evaluates the stratification properties of D on $(\alpha s) \times (\alpha s)$ grids in any two dimensions.

Obtaining the space-filling pattern in Tian and Xu (2022) requires $O(nmps^{mp})$ operations for designs in $\mathcal{D}(n, (s^p)^m)$. However, for designs in $\mathcal{D}(n, (\alpha s)^m)$, it is easy to show that the complexity of computing the two-dimensional projection pattern is $O(nm^3(\alpha s)^2)$, which is much smaller than that of the space-filling pattern for large m (when $\alpha = s$ and $p = 2$). It is worth noting that some faster methods for computing the space-filling pattern have been introduced recently by Tian and Xu (2023). Extending these methods to compute the corresponding two-dimensional projection pattern will be studied as a future work.

The space-filling hierarchy principle proposed by Tian and Xu (2022) states that stratifications on larger grids are more likely to be important than stratifications on smaller grids. Following this principle, we hope that the designs can achieve stratifications on $s \times s$ grids first, followed by $s \times (\alpha s)$ and $\alpha s \times s$ grids, and finally $(\alpha s) \times (\alpha s)$ grids in any two dimensions. Therefore, based on the two-dimensional projection pattern, following the space-filling hierarchy principle, we can propose a minimum aberration type space-filling criterion selects designs $D \in \mathcal{D}(n, (\alpha s)^m)$ that sequentially minimize the entries of $\left(S^{(1,1)}(D), S^{(1,2)}(D), S^{(2,2)}(D)\right)$. This criterion can be used for ranking designs in $\mathcal{D}(n, (\alpha s)^m)$.

The next theorem shows that the two-dimensional projection pattern captures the strength of an $\mathrm{SOA}_\alpha(n, m, (\alpha s), 2+)$.

Theorem 5 *A balanced design $D \in \mathcal{D}(n, (\alpha s)^m)$ is an $\mathrm{SOA}_\alpha(n, m, (\alpha s), 2+)$ if and only if $S^{(1,1)}(D) = S^{(1,2)}(D) = 0$.*

Proof of Theorem 5 Suppose $D = (d_1, \ldots, d_m) \in \mathcal{D}(n, (\alpha s)^m)$ is a balanced design, where d_i denotes the i-th column of D. Let A and B be the two matrices obtained by transferring the levels of D through $x \mapsto g_1(x) = \lfloor x/\alpha \rfloor$ and $x \mapsto g_2(x) = x \pmod{\alpha}$, respectively. It is easy to see that A is a balanced design in $\mathcal{D}(n, s^m)$, B is a balanced design in $\mathcal{D}(n, \alpha^m)$ and $D = \alpha A + B$. Denote a_i and b_i as the i-th columns of A and B, respectively. We further have (a_i, b_i) is an $\mathrm{OA}(n, 2, s \times \alpha, 2)$ for $i = 1, \ldots, m$.

By Definition 2, we have

$$
\begin{aligned}
S^{(1,1)}(D) &= \frac{1}{n^2} \sum_{\boldsymbol{u} \in U_{1,1}} \left| \sum_{\boldsymbol{x} \in D} \prod_{i=1}^{m} C^s_{g_1(u_i)}(g_1(x_i)) \right|^2 \\
&= \frac{1}{n^2} \sum_{1 \le i \ne j \le m} \sum_{g_1(u_i)=1}^{s-1} \sum_{g_1(u_j)=1}^{s-1} \left| \sum_{\boldsymbol{x} \in D} C^s_{g_1(u_i)}(g_1(x_i)) C^s_{g_1(u_j)}(g_1(x_j)) \right|^2 .
\end{aligned}
$$

and

$$
S^{(1,2)}(D) = \frac{1}{n^2} \sum_{1 \le i \ne j \le m} \sum_{g_1(u_i)=0}^{s-1}
$$

$$\sum_{g_2(u_i)=1}^{\alpha-1}\sum_{g_1(u_j)=1}^{s-1}\left|\sum_{\boldsymbol{x}\in D} C^s_{g_1(u_i)}(g_1(x_i))C^{\alpha}_{g_2(u_i)}(g_2(x_i))C^s_{g_1(u_j)}(g_1(x_j))\right|^2$$
$$=\frac{1}{n^2}\sum_{1\le i\neq j\le m}\left(\sum_{g_2(u_i)=1}^{\alpha-1}\sum_{g_1(u_j)=1}^{s-1}\left|\sum_{\boldsymbol{x}\in D} C^{\alpha}_{g_2(u_i)}(g_2(x_i))C^s_{g_1(u_j)}(g_1(x_j))\right|^2\right.$$
$$\left.+\sum_{g_1(u_i)=1}^{s-1}\sum_{g_2(u_i)=1}^{\alpha-1}\sum_{g_1(u_j)=1}^{s-1}\left|\sum_{\boldsymbol{x}\in D} C^s_{g_1(u_i)}(g_1(x_i))C^{\alpha}_{g_2(u_i)}(g_2(x_i))C^s_{g_1(u_j)}(g_1(x_j))\right|^2\right).$$

The two equations above reveal that $S^{(1,1)}(D)$ is equivalent to $A_2(A)$ and $S^{(1,1)}(D)+S^{(1,2)}(D)$ is equivalent to $\sum_{1\le i\neq j\le m}(A_3(a_i,b_i,a_j)+A_2(b_i,a_j))$. Therefore, A is an OA$(n,m,s,2)$ if and only if $S^{(1,1)}(D)=0$. For $1\le i\neq j\le m$, (a_i,b_i,a_j) is an OA$(n,m,s\times\alpha\times s,3)$ if and only if $S^{(1,1)}(D)+S^{(1,2)}(D)=0$. The desired conclusion then follows by Lemma 4. □

In particular, when $\alpha=s$, the two-dimensional projection pattern captures the strength of an SOA$(n,m,s^2,2+)$.

Corollary 6 *A balanced design $D\in\mathcal{D}(n,(s^2)^m)$ is an* SOA$(n,m,s^2,2+)$ *if and only if $S^{(1,1)}(D)=S^{(1,2)}(D)=0$.*

Theorem 5 and Corollary 6 show that any SOA$_{\alpha}(n,m,(\alpha s),2+)$ or SOA$(n,m,s^2,2+)$ has $S^{(1,1)}=S^{(1,2)}=0$. Therefore, selecting an optimal SOA of strength 2+ under the minimum aberration type criterion based on the two-dimensional projection pattern is equivalent to minimizing $S^{(2,2)}$.

Example 2 Consider two SOA(81, 11, 9, 2+)'s, denoted by D_1 and D_2, respectively. They are constructed by $D_1=3A_1+B_1$ and $D_2=3A_2+B_2$ with

$$\begin{aligned}&A_1=(e_1e_2^2, e_1e_2^2e_3,\\&e_1e_3^2, e_1e_2^2e_3^2, e_1e_2^2e_3e_4, e_1e_3^2e_4, e_1e_2e_3^2e_4, e_1e_2^2e_3^2e_4, e_2e_4^2, e_1e_2e_4^2, e_2e_3e_4^2)\\&B_1=(e_4, e_2e_3^2e_4, e_2e_3^2e_4^2, e_1, e_1e_3^2e_4^2, e_1e_4^2, e_1e_2e_3e_4, e_1e_2e_3^2, e_1e_2e_3, e_1e_3e_4^2, e_2)\end{aligned}$$

and

$$\begin{aligned}&A_2=(e_1e_2^2, e_1e_3^2, e_1e_4^2, e_2e_3^2, e_2e_4^2, e_3e_4^2, e_1e_2^2e_3, e_1e_2e_3^2, e_1e_2^2e_3^2, e_1e_2^2e_4, e_1e_2e_4^2)\\&B_2=(e_1, e_1, e_1, e_2, e_2, e_3, e_2, e_3, e_1, e_2, e_4),\end{aligned}$$

where e_1,e_2,e_3,e_4 are four independent columns of length 81 whose entries are taken from $\mathbb{Z}_3$ and $e_ie_j^2$ represents $e_i+2e_j \pmod 3$.

Both D_1 and D_2 achieve stratifications on 3×3, 3×9 and 9×3 grids in any two dimensions. The two-dimensional projection patterns of D_1 and D_2 are (0, 0, 8)

Table 1 Comparison of D_1 and D_2 under various criteria

	$S^{(1,1)}$	$S^{(1,2)}$	$S^{(2,2)}$	ϕ_{eff}	$CD_2(\times 100)$	d_{L_1}	d_{L_2}	ρ_{max}	ρ_{ave}
D_1	0	0	8	99.560%	5.483	17	5.745	0	0
D_2	0	0	76	95.638%	6.785	14	4.796	0.1	0.024

Table 2 Design D_3, an $\text{SOA}_2(36, 7, 6, 2+)$

0	0	0	2	2	2	4	4	4	0	0	0	2	2	2	4	4	4
0	0	2	2	2	4	0	4	4	2	4	4	0	0	4	0	2	2
0	2	0	2	4	2	4	0	4	4	2	4	0	4	0	2	0	2
0	2	4	2	4	0	2	0	4	0	4	2	4	0	2	4	2	0
0	4	4	2	0	0	2	2	4	2	2	0	2	4	4	0	0	4
0	4	2	2	0	4	0	2	4	4	0	2	4	2	0	2	4	0
0	2	4	0	2	4	4	2	0	0	2	4	4	2	0	0	2	4
1	1	1	3	3	3	5	5	5	1	1	1	3	3	3	5	5	5
1	1	3	3	3	5	1	5	5	3	5	5	1	1	5	1	3	3
1	3	1	3	5	3	5	1	5	5	3	5	1	5	1	3	1	3
1	3	5	3	5	1	3	1	5	1	5	3	5	1	3	5	3	1
1	5	5	3	1	1	3	3	5	3	3	1	3	5	5	1	1	5
1	5	3	3	1	5	1	3	5	5	1	3	5	3	1	3	5	1
1	3	5	1	3	5	5	3	1	1	3	5	5	3	1	1	3	5

and (0, 0, 76), respectively, which clearly capture the stratification properties of both designs. The values of $S^{(2,2)}(D_1)$ and $S^{(2,2)}(D_2)$ indicate that design D_1 is more space-filling than design D_2 in terms of two-dimensional projections.

We also compare the performance of D_1 and D_2 across other measures, including the ϕ-efficiency (ϕ_{eff}) measure under the uniform projection criterion (Sun et al. 2019; Wang et al. 2022), the squared centered L_2-discrepancy(CD_2) measure under the uniformity criterion (Fang et al. 2000), the minimum row-pairwise L_1 and L_2 distances (d_{L_1} and d_{L_2}) under the maximin distance criterion (Johnson et al. 1990), and the maximum and average column-pairwise sample correlation coefficients (ρ_{max} and ρ_{ave}) under the column-orthogonality criterion (Owen 1994). The results are summarized in Table 1, which clearly shows that D_1 is better than D_2 under all criteria. This confirms the effectiveness of the proposed minimum aberration type criterion in evaluating the space-filling properties of the two SOAs.

Example 3 Consider designs D_3 and $D_4 \in \mathcal{D}(36, 6^7)$ listed in Tables 2 and 3, respectively. Here, D_3 and D_4 are obtained by level expansion based on an $\text{OA}(36, 7, 3, 2)$, with the difference that D_3 is an $\text{SOA}_2(36, 7, 6, 2+)$, and D_4 is obtained by randomly expanding the levels of $\text{OA}(36, 7, 3, 2)$ through $0 \to \{0, 1\}$, $1 \to \{2, 3\}$ and $2 \to \{4, 5\}$. To save space, both designs are presented in transposed forms, with the top half of each table display runs 1–18 and bottom half runs 19–36.

Both D_3 and D_4 achieve stratifications on 3×3 grids in any two dimensions. The design D_3 can further achieve stratifications on 3×6 and 6×3 grids in any

Table 3 Design $D_4 \in \mathcal{D}(36, 6^7)$, a design based on OA(36, 7, 3, 2)

1	0	0	3	2	2	5	5	4	0	1	1	3	3	2	5	4	4
1	0	3	3	2	5	0	4	5	3	5	4	0	1	5	1	2	2
1	3	0	2	4	3	4	1	5	4	2	5	0	5	1	2	1	3
1	3	4	2	4	0	3	0	5	1	4	2	4	0	3	5	2	0
0	5	4	3	0	1	3	2	5	2	3	0	3	4	5	1	0	5
0	5	2	2	0	4	1	3	5	5	1	3	5	2	1	3	4	0
1	3	4	0	3	4	5	3	0	1	2	5	4	3	0	1	2	4
3	2	3	4	5	4	1	1	0	2	3	2	5	4	5	0	1	0
3	2	4	5	4	0	3	1	0	5	1	0	2	2	1	3	4	4
2	4	3	5	0	4	0	3	1	0	5	0	3	1	2	5	2	4
2	4	1	5	0	2	5	2	1	3	0	5	1	3	5	1	4	3
2	1	1	4	2	2	4	4	0	5	4	3	5	1	0	2	3	1
2	0	4	4	3	0	2	4	0	1	2	5	1	4	3	5	1	3
3	4	1	2	5	0	1	5	2	2	5	0	0	5	2	3	4	1

Table 4 Comparison of D_3 and D_4 under various criteria

	$S^{(1,1)}$	$S^{(1,2)}$	$S^{(2,2)}$	ϕ_{eff}	$CD_2(\times 100)$	d_{L_1}	d_{L_2}	ρ_{max}	ρ_{ave}
D_3	0	0	21	96.310%	4.279	7	2.646	0.086	0.086
D_4	0	10.333	6.833	91.368%	5.205	2	1.414	0.238	0.068

two dimensions, whereas D_4 cannot. The two-dimensional projection patterns of D_3 and D_4 are $(0, 0, 21)$ and $(0, 10.333, 6.833)$, respectively, which clearly capture the stratification properties of both designs. By the minimum aberration type criterion, the design D_3 is more space-filling than D_4 in two-dimensional projections.

We also compare the performance of D_3 and D_4 under the criteria in Example 2. Table 4 shows that D_3 is better than D_4 under most criteria.

At the end of this subsection, we remark that $S^{(1,1)}(D) + S^{(1,2)}(D) + S^{(2,2)}(D)$ is equivalent to $A_2(D)$ for a balanced design $D \in \mathcal{D}(n, (\alpha s)^m)$, where $A_2(D)$ is the 2nd entry of the generalized wordlength pattern (Xu and Wu 2001), as stated in the next theorem. Despite this relationship, it is not feasible to simply evaluate the design's space-filling property through its A_2 value, as shown in Example 3.

Theorem 7 *Let $D \in \mathcal{D}(n, (\alpha s)^m)$ be a balanced design, then $S^{(1,1)}(D) + S^{(1,2)}(D) + S^{(2,2)}(D) = A_2(D)$.*

Proof of Theorem 7 Suppose $D = (d_1, \ldots, d_m) \in \mathcal{D}(n, (\alpha s)^m)$ is a balanced design, and the arrays $A = (a_1, \ldots, a_m)$ and $B = (b_1, \ldots, b_m)$ are defined as in the proof of Theorem 5 such that $D = \alpha A + B$. Each column of D, say d_i, can be uniquely determined by the corresponding columns in A and B, that is, (a_i, b_i). For $1 \leq i \neq j \leq m$, Proposition 2 of Chen and Tang (2022a) shows that

$$A_2(d_i, d_j) = A_2(a_i, a_j) + A_2(a_i, b_j) + A_2(b_i, a_j) + A_2(a_i, b_i) \\ + A_2(a_j, b_j) + A_2(b_i, b_j) + A_3(a_i, b_i, a_j) + A_3(a_i, a_j, b_j) \\ + A_3(a_i, b_i, b_j) + A_3(b_i, a_j, b_j) + A_4(a_i, b_i, a_j, b_j),$$

where $A_2(\cdot)$, $A_3(\cdot)$ and $A_4(\cdot)$ are the 2nd, 3rd and 4th entries of the generalized wordlength pattern proposed by Xu and Wu (2001), respectively. By using a proof similar to those for Theorems 1 and 5, we obtain

$$S^{(1,1)}(D) = \sum_{1 \le i < j \le m} A_2(a_i, a_j),$$

$$S^{(1,2)}(D) = \sum_{1 \le i < j \le m} \{A_3(a_i, b_i, a_j) + A_3(a_i, a_j, b_j) + A_2(a_i, b_j) + A_2(b_i, a_j)\},$$

$$S^{(2,2)}(D) \\ = \sum_{1 \le i < j \le m} \{A_4(a_i, b_i, a_j, b_j) + A_3(a_i, b_i, b_j) + A_3(b_i, a_j, b_j) + A_2(b_i, b_j)\}.$$

For $i = 1, \ldots, m$, (a_i, b_i) is an $\mathrm{OA}(n, 2, s \times \alpha, 2)$; thus, $A_2(a_i, b_i) = 0$. Therefore, $S^{(1,1)}(D) + S^{(1,2)}(D) + S^{(2,2)}(D) = \sum_{1 \le i < j \le m} A_2(d_i, d_j) = A_2(D)$. $\square$

4 Application I: constructing optimal column-expanded designs based on GSOAs

In this section, we apply the results in Sect. 3 to construct designs by adding columns to some existing GSOAs of strength 3 and SOAs of strength 2+. We call the generated designs column-expanded designs. These designs can accommodate more factors than the original GSOAs or SOAs for the same run size. They also perform well under the space-filling pattern or two-dimensional projection pattern based minimum aberration type criteria.

We add columns to GSOAs of strength 3 and SOAs of strength 2+. Calculating the design's space-filling pattern is often time-consuming, especially when the design size is large. For the sake of effectiveness, we optimize the first few entries of the space-filling pattern and the two-dimensional projection pattern for column-expanded designs. Specifically, we focus on S_1, S_2, S_3 of the space-filling pattern when expanding GSOAs of strength 3 and $S^{(1,1)}$, $S^{(1,2)}$ of the two-dimensional projection pattern when expanding SOAs of strength 2+.

The following lemma is crucial in this section. The proof is straightforward by using the definitions of S_3 and $S^{(1,2)}$.

Lemma 8 *Let $A = (a_1, \ldots, a_m)$ and $B = (b_1, \ldots, b_m)$ be two balanced designs in $\mathcal{D}(n, s^m)$ satisfying (a_j, b_j) is an* $\mathrm{OA}(n, 2, s, 2)$ *for $j = 1, \ldots, m$, and let a_{+1} and b_{+1} be two balanced columns in $\mathcal{D}(n, s^1)$ satisfying (a_{+1}, b_{+1}) is an* $\mathrm{OA}(n, 2, s, 2)$. *Let $D = (d_1, \ldots, d_m) = sA + B$ and the column-expanded design $D_{+1} = (D, d_{+1}) = sA_{+1} + B_{+1}$, where $d_j = sa_j + b_j$ for $j = 1, \ldots, m$, $d_{+1} = sa_{+1} + b_{+1}$, $A_{+1} = (A, a_{+1})$ and $B_{+1} = (B, b_{+1})$. Then*

$$
\begin{aligned}
S_3(D_{+1}) - S_3(D) &= \alpha(A, a_{+1}) + \beta(A, B, a_{+1}) + \gamma(A, a_{+1}, b_{+1}), \\
S^{(1,2)}(D_{+1}) - S^{(1,2)}(D) &= \beta(A, B, a_{+1}) + \gamma(A, a_{+1}, b_{+1}),
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha(A, a_{+1}) &= \sum_{1 \le j < j' \le m} A_3(a_j, a_{j'}, a_{+1}), \\
\beta(A, B, a_{+1}) &= \sum_{1 \le j \le m} \{A_3(a_j, b_j, a_{+1}) + A_2(b_j, a_{+1})\}, \\
\gamma(A, a_{+1}, b_{+1}) &= \sum_{1 \le j \le m} \{A_3(a_j, a_{+1}, b_{+1}) + A_2(a_j, b_{+1})\}, \qquad (12)
\end{aligned}
$$

$S_3(\cdot)$ is defined in (4), $S^{(1,2)}(\cdot)$ is defined in (11), and $A_2(\cdot)$, $A_3(\cdot)$ are the 2nd and the 3rd entries of the generalized wordlength pattern proposed by Xu and Wu (2001), respectively.

4.1 Column-expanded designs based on GSOAs of strength 3

GSOA$(n, m, s^p, 3)$s have good stratification properties in both two and three-dimensional projections. The column-expanded designs based on them are expected to keep these projection properties. We expand a class of GSOAs with $s = 2$ which achieve the maximum number of columns, whose existence were shown by Zhou and Tang (2019) and He and Tang (2012).

Lemma 9 *For any $k \ge 3$, GSOA$(2^k, 2^{k-1}-1, 2^2, 3)$ and SOA$(2^k, 2^{k-1}-1, 2^3, 3)$ can be constructed by selecting columns in the saturated OA$(2^k, 2^k - 1, 2, 2)$. Moreover, their numbers of factors reach the maximum value.*

The construction method of GSOA$(2^k, 2^{k-1} - 1, 2^2, 3)$ given by Zhou and Tang (2019) follows. For $k \ge 3$, we use $e_1, \ldots, e_k$ to denote the k two-level independent columns and S to denote the saturated regular OA$(2^k, 2^k - 1, 2, 2)$ consisting of $e_1, \ldots, e_k$ and all of their possible interaction columns. We select columns from S to obtain

$$
\begin{aligned}
A &= \left\{ e_1^{u_1} \cdots e_{k-1}^{u_{k-1}} e_k, u_i \in \{0, 1\}, \sum_{i=1}^{k-1} u_i > 0 \right\}, \\
B &= \left\{ e_1^{u_1} \cdots e_{k-1}^{u_{k-1}}, u_i \in \{0, 1\}, \sum_{i=1}^{k-1} u_i > 0 \right\}, \qquad (13)
\end{aligned}
$$

where $e_1^{u_1} \cdots e_{k-1}^{u_{k-1}} e_k$ represents $u_1 e_1 + \cdots + u_{k-1} e_{k-1} + e_k$ (mod 2). Then, A is an OA$(2^k, 2^{k-1} - 1, 2, 3)$ and B is an OA$(2^k, 2^{k-1} - 1, 2, 2)$. Let a_j and b_j denote the j-th column of A and B, respectively, for $1 \le j \le 2^{k-1} - 1$, such that $a_j b_j = e_k$ for $1 \le j \le 2^{k-1} - 1$. Then, $D = 2A + B \in \mathcal{D}(2^k, 4^{2^{k-1}-1})$ is a GSOA$(2^k, 2^{k-1} - 1, 2^2, 3)$.

Given A and B, we can always find $C = (c_1, \ldots, c_{2^{k-1}-1}) \in \mathcal{D}(2^k, 2^{2^{k-1}-1})$, such that (a_j, b_j, c_j) is an OA$(2^k, 3, 2, 3)$ for $1 \le j \le 2^{k-1}-1$. Then, $D = 4A+2B+C \in \mathcal{D}(2^k, 8^{2^{k-1}-1})$ is a SOA$(2^k, 2^{k-1}-1, 2^3, 3)$; see He and Tang (2012) for more details. Because C does not affect the stratification properties of concern (i.e., stratifications on 2×2, 2×4 and 4×2 grids), below we focus on expanding the columns of GSOA$(2^k, 2^{k-1}-1, 2^2, 3)$ only.

Let A and B be defined in (13), and $D = 2A + B$. Let $\bar{A}$ denote the complement of A in S. Since $S = A \cup B \cup \{e_k\}$, we have $\bar{A} = B \cup \{e_k\}$. Let us suppose a column $d_{+1} = 2a_{+1} + b_{+1}$ is added to D to obtain D_{+1}, where $a_{+1}, b_{+1} \in S$ are the corresponding columns added to A and B, respectively. By Lemma 9, D is a GSOA$(2^k, 2^{k-1}-1, 2^2, 3)$ achieving the maximum number of factors and thus can no longer accommodate more factors to keep $S_1(D_{+1}) = S_2(D_{+1}) = S_3(D_{+1}) = 0$. Our aim is to sequentially minimize $S_1(D_{+1})$, $S_2(D_{+1})$ and $S_3(D_{+1})$. First, it is possible and easy to have $S_1(D_{+1}) = S_2(D_{+1}) = 0$. As $\bar{A}$ is the set of all columns in S that are orthogonal to each column of A, it is only possible to select a_{+1} from $\bar{A}$ to keep $S_1(D_{+1}) = S_2(D_{+1}) = 0$. Next, we need to minimize $S_3(D_{+1})$ among $a_{+1} \in \bar{A}$ and $b_{+1} \in S$. By Lemma 8, $S_3(D_{+1}) = S_3(D) + \alpha(A, a_{+1}) + \beta(A, B, a_{+1}) + \gamma(A, a_{+1}, b_{+1})$, where α, β and γ are defined in (12).

Theorem 10 *Let $D_{+1} = (D, d_{+1})$ be a column-expanded design by adding $d_{+1} = 2a_{+1} + b_{+1}$ to D, where $D = 2A + B$ is the* GSOA$(2^k, 2^{k-1}-1, 2^2, 3)$ *with A and B defined in (13), $a_{+1} \in \bar{A} = B \cup \{e_k\}$ and $b_{+1} \in S$. We have:*

(i) $\alpha(A, a_{+1}) = 0, \beta(A, B, a_{+1}) = 2^{k-1} - 1$ *if* $a_{+1} = e_k$;
(ii) $\alpha(A, a_{+1}) = 2^{k-2} - 1, \beta(A, B, a_{+1}) = 1$ *if* $a_{+1} \in B$;
(iii) $\gamma(A, a_{+1}, b_{+1}) = 0$ *if* $b_{+1} \in \bar{A} \setminus \{a_{+1}, e_k\}$ *such that* $a_{+1}b_{+1} \in \bar{A}$.

Proof of Theorem 10 (i) For $a_{+1} = e_k$, we have $A_3(a_j, a_{j'}, a_{+1}) = 0$ for any $1 \le j < j' \le m$ because e_k cannot be linearly represented by any two columns of A. Hence, $\alpha(A, a_{+1}) = 0$. Additionally, $A_3(a_j, b_j, a_{+1}) = 1$ and $A_2(b_j, a_{+1}) = 0$ for $1 \le j \le 2^{k-1} - 1$, which implies $\beta(A, B, a_{+1}) = 2^{k-1} - 1$.

(ii) For $a_{+1} \in B = \bar{A} \setminus \{e_k\}$ and any two columns $a_j, a_{j'} \in A$, if $a_{+1}a_j = a_{j'}$ then $A_3(a_j, a'_j, a_{+1}) = 1$; otherwise, $A_3(a_j, a'_j, a_{+1}) = 0$. Without loss of generality, we suppose $a_{+1} = b_{j^*}$, $1 \le j^* \le 2^{k-1} - 1$. We have $a_{j^*} \in A$ such that $a_{+1}a_{j^*} = e_k$. Each of the remaining $2^{k-1} - 2$ columns of A, say a_j, $j \ne j^*$, there exists exactly a column $a_{j'} \in A$, $j \ne j, j^*$ such that $A_3(a_j, a'_j, a_{+1}) = 1$. Thus, there are $(2^{k-1} - 2)/2$ pairs of $a_j, a_{j'}$ such that $A_3(a_j, a'_j, a_{+1}) = 1$ and $\alpha(A, a_{+1}) = 2^{k-2} - 1$. Additionally, we have $A_2(b_{j^*}, a_{+1}) = 1$, $A_3(a_{j^*}, b_{j^*}, a_{+1}) = 0$, and $A_3(a_j, b_j, a_{+1}) = A_2(b_j, a_{+1}) = 0$ for all $j \ne j^*$, leading to $\beta(A, B, a_{+1}) = 1$.

(iii) The column $b_{+1} \in \bar{A} \setminus \{a_{+1}\}$ is orthogonal to each column of A, which ensures $A_2(a_j, b_{+1}) = 0$ for $1 \le j \le 2^{k-1} - 1$. That $a_{+1}b_{+1} \in \bar{A}$ ensures $A_3(a_j, a_{+1}, b_{+1}) = 0$ for $1 \le j \le 2^{k-1} - 1$. Thus, $\gamma(A, a_{+1}, b_{+1}) = 0$.

□

Theorem 10 shows that the optimal column-expanded design can be obtained by choosing any $a_{+1} \in B$ and $b_{+1} \in \bar{A}\backslash\{a_{+1}\}$ such that $a_{+1}b_{+1} \in \bar{A}$. We denote the corresponding designs as $D_{+1} = (D, d_{+1})$, $A_{+1} = (A, a_{+1})$ and $B_{+1} = (B, b_{+1})$. The design D_{+1} has the smallest $S_3(D_{+1})$ value 2^{k-2}.

More columns can be added to D sequentially by using a similar strategy as Theorem 10. We illustrate how to optimally add a column to the above D_{+1}. The following proposition shows that the best choice of the added columns a_{+2}, b_{+2} and $d_{+2} = sa_{+2} + b_{+2}$.

Proposition 11 *Suppose a column $a_{+2} \in \bar{A}\backslash\{a_{+1}, b_{+1}, a_{+1}b_{+1}, e_k\}$ is added to A_{+1}; then, we have $\alpha(A_{+1}, a_{+2}) = 2^{k-2} - 1$, $\beta(A_{+1}, B_{+1}, a_{+2}) = 1$. Given a_{+2}, we suppose a column $b_{+2} \in \bar{A} \setminus \{a_{+1}, e_k\}$ is added to B_{+1} such that $a_{+2}b_{+2} \in \bar{A}\backslash\{a_{+1}\}$; then, we have $\gamma(A_{+1}, a_{+2}, b_{+2}) = 0$.*

Proof of Proposition 11 If a_{+2} is added to A_{+1}, the increment α is formed by $\sum_j A_3(a_j, a_{+1}, a_{+2})$ and $\sum_{j \neq j'} A_3(a_j, a_{j'}, a_{+2})$. The former must be zero, because the columns of A cannot be linearly represented by a_{+1} and a_{+2}. The latter has been shown in Theorem 10. The increment β is formed by $\sum_j [A_3(a_j, b_j, a_{+2}) + A_2(b_j, a_{+2})]$, $A_3(a_{+1}, b_{+1}, a_{+2}) + A_2(b_{+1}, a_{+2})$. The former has also been shown in theorem 10. The latter must be zero, because $a_{+2} \in \bar{A} \setminus \{a_{+1}, b_{+1}, a_{+1}b_{+1}, e_k\}$.

Now, we consider b_{+2}. The column $b_{+2} \in \bar{A}\backslash\{a_{+1}, e_k\}$ is orthogonal to each column of A_{+1}, which ensures $A_2(a_j, b_{+1}) = 0$ for $1 \leq j \leq 2^{k-1} - 1$ and $A_2(a_{+1}, b_{+1}) = 0$. That $a_{+2}b_{+2} \in \bar{A}$ ensures $A_3(a_j, a_{+2}, b_{+2}) = 0$ for $1 \leq j \leq 2^{k-1} - 1$ and $A_3(a_{+1}, a_{+2}, b_{+2}) = 0$. Thus, $\gamma(A_{+1}, a_{+2}, b_{+2}) = 0$. □

Example 4 Let $s = 2$, $k = 4$ and $e_1, \ldots, e_4$ be 4 independent columns of length 16. The 16×7 arrays,

$$A = (e_1e_4, e_2e_4, e_3e_4, e_1e_2e_4, e_1e_3e_4, e_2e_3e_4, e_1e_2e_3e_4),$$
$$B = (e_1, e_2, e_3, e_1e_2, e_1e_3, e_2e_3, e_1e_2e_3),$$

can be utilized to construct an GSOA(16, 7, 4, 3) by $D = 2A + B$. We have $\alpha(A, e_4) = 0$ and $\beta(A, B, e_4) = 7$ if $a_{+1} = e_4$. According to Theorem 10, a better choice is $a_{+1} = e_1$; then, we have $\alpha(A, e_1) = 3$ and $\beta(A, B, e_1) = 1$, and the same values of α and β can be obtained by taking $a_{+1} = e_2, e_3, e_1e_2, e_1e_3, e_2e_3$ or $e_1e_2e_3$. With $a_{+1} = e_1$, if $b_{+1} = e_2$, then we have $\gamma(A, e_1, e_2) = 0$. The b_{+1} could also take the other columns in $\bar{A} \setminus \{e_1, e_4\}$. The minimum $S_3(D_{+1})$, which equals 4, is obtained when $a_{+1} = e_1$ and $b_{+1} = e_2$. Furthermore, by Proposition 11, if we take $a_{+2} = e_3$ and $b_{+2} = e_1e_2$, then $S_3(D_{+2}) = 8$, and this is the minimum value of $S_3(D_{+2})$.

If we want to add 3 or more columns to the GSOA$(2^k, 2^{k-1} - 1, 2^2, 3)$ above, there is a simple way for $k \geq 4$. First, construct a GSOA$(2^k, 2^{k-2} - 1, 2^2, 3)$ $D' = 2A' + B'$ with

$$A' = \left\{ e_1^{u_1} \cdots e_{k-2}^{u_{k-2}} e_{k-1}, u_i \in \{0, 1\}, \sum_{i=1}^{k-2} u_i > 0 \right\},$$

$$B' = \left\{ e_1^{u_1} \cdots e_{k-2}^{u_{k-2}}, u_i \in \{0, 1\}, \sum_{i=1}^{k-2} u_i > 0 \right\},$$

where $e_1, \ldots, e_{k-1}$ denote the $k-1$ two-level independent columns of length 2^k. Then take any m'-column subdesign from D' with $3 \le m' \le 2^{k-2} - 1$ and add it to D to obtain $D_{+m'}$. According to Theorem 10 and Proposition 11, the increment of S_3 is at least 2^{k-2} when adding any column. This minimum increment can be achieved by adding any column of D' to D, and D' itself will not cause an increase in S_3, so the $D_{+m'}$ will have the smallest S_3 value of all the column-expanded designs based on D. Specifically, we have $S_1(D_{+m'}) = S_2(D_{+m'}) = 0$ and $S_3(D_{+m'}) = m' 2^{k-2}$.

4.2 Column-expanded designs based on SOAs of strength 2+

SOAs of strength 2+ have good two-dimensional stratification properties and economical run sizes. It was shown in He et al. (2018) that an SOA$(n, m, s^2, 2+)$ D exists if and only if there exist an OA$(n, m, s, 2)$ A and an OA$(n, m, s, 1)$ B such that (a_i, a_j, b_i) is an OA$(n, m, s, 3)$ for any $i \ne j$. In particular, the three designs are linked through $D = sA + B$.

In this subsection, we study column-expanded designs based on SOAs of strength 2+ constructed from regular designs. Suppose $s \ge 2$ is a prime power. For $k \ge 3$, let $e_1, \ldots, e_k$ denote the k independent columns of s levels, with entries from the Galois field GF$(s) = \{w_0 = 0, w_1 = 1, w_2, \ldots, w_{s-1}\}$. Let S denote the saturated regular OA$(s^k, (s^k - 1)/(s-1), s, 2)$ consisting of $e_1, \ldots, e_k$ and all of their possible interaction columns. A method of constructing SOA$(s^k, m, s^2, 2+)$ is as follows: first select columns of an OA$(s^k, m, s, 2)$ A and an OA$(s^k, m, s, 1)$ B from S, such that the condition (a_i, a_j, b_i) is an OA$(s^k, m, s, 3)$ for any $i \ne j$ is satisfied, then we obtain $D = sA + B$. Before applying $sA + B$, one should convert the symbols of GF$(s) = \{w_0 = 0, w_1 = 1, w_2, \ldots, w_{s-1}\}$ to $\{0, 1, \ldots, s-1\}$ since the columns of A and B are selected from S and have entries from GF(s). No conversion is necessary if s is a prime. Several such types of construction methods are given in He et al. (2018). For example, Theorem 4 of He et al. (2018) shows that for any $k \ge 3$ and any prime power $s \ge 3$, an SOA$(s^k, m, s^2, 2+)$ where $m = (s^k - 1)/(s-1) - ((s-1)^k - 1)/(s-2)$ can be constructed.

As SOAs of strength 2+ focus on two-dimensional stratification properties, the two-dimensional projection pattern is suitable for evaluating the corresponding column-expanded designs. Given an SOA$(s^k, m, s^2, 2+)$ $D = sA + B$ constructed via saturated regular design S, we have $S^{(1,1)}(D) = S^{(1,2)}(D) = 0$ by Corollary 6, where $S^{(1,1)}$ and $S^{(1,2)}$ are defined in (11). Let $\bar{A}$ denote the complement of A in S. Let us suppose a column $d_{+1} = sa_{+1} + b_{+1}$ is added to D to obtain D_{+1}, where $a_{+1}, b_{+1} \in S$ are the corresponding columns added to A and B, respectively. The criterion we use is to sequentially minimize $S^{(1,1)}(D_{+1})$ and $S^{(1,2)}(D_{+1})$. It is easy to have $S^{(1,1)}(D_{+1}) = 0$ by selecting any column in $\bar{A}$ as a_{+1}. It remains to minimize $S^{(1,2)}(D_{+1})$ among $a_{+1} \in \bar{A}$ and $b_{+1} \in S$. By Lemma 8, $S^{(1,2)}(D_{+1}) - S^{(1,2)}(D) = \beta(A, B, a_{+1}) + \gamma(A, a_{+1}, b_{+1})$, where β and γ are defined in (12). Theorem 10 shows how to choose a_{+1} and b_{+1}.

Theorem 12 *Let $D_{+1} = (D, d_{+1})$ be a column-expanded design by adding $d_{+1} = 2a_{+1} + b_{+1}$ to D, where $D = sA + B$ is an* SOA$(s^k, m, s^2, 2+)$ *with A and B's columns selected from S, $a_{+1} \in \bar{A}$ and $b_{+1} \in S$. Then, $S^{(1,2)}(D_{+1})$ is minimized if a_{+1} has the minimum frequency in*

$$\begin{cases} (b_1, a_1b_1, b_2, a_2b_2, \ldots, b_m, a_mb_m), & \text{when } s = 2, \\ (b_1, a_1b_1, \ldots, a_1b_1^{s-1}, b_2, a_2b_2, \ldots, a_2b_2^{s-1}, \ldots, b_m, a_mb_m, \ldots, a_mb_m^{s-1}), & \text{when } s \geq 3, \end{cases}$$

and $b_{+1} \in \bar{A} \setminus \{a_{+1}\}$ such that $a_{+1}b_{+1} \in \bar{A}$. Here, $a_ib_i^j$ represents $a_i + w_jb_i$ with calculations in GF(s) *for $s \geq 3$.*

Proof of Theorem 12 For $j = 1, \ldots, m$, we have $A_2(b_j, a_{+1}) = s - 1$ if $a_{+1} = b_i$ or $A_2(a_j, b_j, a_{+1}) = s - 1$ if a_{+1} is a linear combination of a_j and b_j. Thus, the values of $\beta(A, B, a_{+1})$ are equal to $1 (= s - 1)$ times the frequency of a_{+1} in $(b_1, a_1b_1, \ldots, b_m, a_mb_m)$ for $s = 2$ and $s - 1$ times the frequency of a_{+1} in $(b_1, a_1b_1, \ldots, a_1b_1^{s-1}, \ldots, b_m, a_mb_m, \ldots, a_mb_m^{s-1})$ for $s \geq 3$. Since b_{+1} is selected from $\bar{A} \backslash \{a_{+1}\}$, we have $A_2(A, b_{+1}) = 0$. Furthermore, the restriction $a_{+1}b_{+1} \in \bar{A}$ ensures $A_3(a_j, a_{+1}, b_{+1}) = 0$ for $j = 1, \ldots, m$. □

In addition, if the frequency of a_{+1} in $(b_1, a_1b_1, \ldots, b_m, a_mb_m)$ is exactly 1 for $s = 2$, then the number of (d_j, d_u) of D_{+1} that achieves stratifications on 2×4 grids is the maximum value M given in Proposition 2 of Shi and Tang (2019). Here is an example illustrating Theorem 12.

Example 5 Let $s = 2$, $k = 5$ and $e_1, \ldots, e_5$ be 5 independent columns of length 32. Let us take

$$P = \{e_1, e_2, e_3, e_1e_2, e_1e_3, e_2e_3, e_1e_2e_3\},$$
$$Q = \{e_4, e_5, e_4e_5\}$$

and $C = P \cup Q$. We obtain $A = (a_1, \ldots, a_{21})$ and $B = (b_1, \ldots, b_{21})$ as

$$\begin{aligned} A = S \setminus C = (&e_1e_4, e_1e_5, e_2e_4, e_2e_5, e_3e_4, e_3e_5, e_1e_2e_4, e_1e_2e_5, e_1e_3e_4, \\ &e_1e_3e_5, e_1e_4e_5, e_2e_3e_4, e_2e_3e_5, e_2e_4e_5, e_3e_4e_5, e_1e_2e_3e_4, e_1e_2e_3e_5, \\ &e_1e_2e_4e_5, e_1e_3e_4e_5, e_2e_3e_4e_5, e_1e_2e_3e_4e_5), \end{aligned}$$

and

$$\begin{aligned} B = (&e_1, e_1, e_2, e_2, e_3, e_3, e_1e_2, e_1e_2, e_1e_3, e_1e_3, e_1, e_2e_3, e_2e_3, \\ &e_2, e_3, e_1e_2e_3, e_1e_2e_3, e_1e_2, e_1e_3, e_2e_3, e_1e_2e_3). \end{aligned}$$

Then, by He et al. (2018), $D = 2A + B$ is an SOA$(32, 21, 4, 2+)$.

In $(b_1, a_1b_1, \ldots, b_{21}, a_{21}b_{21})$, each column in P appears 3 times, and each column in Q appears 7 times. If we choose $a_{+1} \in Q$, e.g., $a_{+1} = e_4$, then we have $\beta(A, B, a_{+1}) = 7$. A better choice according to Theorem 12 is $a_{+1} \in P$, e.g.,

$a_{+1} = e_1$, leading to $\beta(A, B, a_{+1}) = 3$. Given $a_{+1} = e_1$, take $b_{+1} = e_2$, which satisfies $a_{+1}b_{+1} = e_1e_2 \in \bar{A}$. Then, by Theorem 10, $\gamma(A, a_{+1}, b_{+1}) = 0$, and $S^{(1,2)}(D_{+1}) = 3$ is minimized.

More columns can be added to D sequentially. Let us suppose $i - 1$ columns have been added sequentially to D to obtain $D_{+(i-1)} = sA_{+(i-1)} + B_{+(i-1)}, i \geq 2$. Simply denote the columns of $A_{+(i-1)}$ and $B_{+(i-1)}$ by $a_1, \ldots, a_{m+i-1}$ and $b_1, \ldots, b_{m+i-1}$, respectively. We can choose a_{+i} according to the frequencies of the columns of $\bar{A}_{+(i-1)} = S \backslash A_{+(i-1)}$ in

$$\begin{cases} (b_1, a_1b_1, \ldots, b_{m+i-1}, a_{m+i-1}b_{m+i-1}), & \text{when } s = 2, \\ (b_1, a_1b_1, \ldots, a_1b_1^{s-1}, \ldots, b_{m+i-1}, a_{m+i-1}b_{m+i-1}^{s-1}), & \text{when } s \geq 3, \end{cases}$$

and choose $b_{+i} \in \bar{A}_{+(i-1)} \setminus \{a_{+i}\}$ such that $a_{+i}b_{+i} \in \bar{A}_{+(i-1)}$. As an illustration, for D_{+1} in Example 5, suppose we want to add another column d_{+2}. We have $\bar{A}_{+1} = P \cup Q \setminus \{e_1\}$. The frequencies of e_2 and e_1e_2 in $(b_1, a_1b_1, \ldots, b_{22}, a_{22}b_{22})$ become 4, while the frequencies of the other columns in $\bar{A}_{+1}$ remain the same as in Example 5. Therefore, we can choose a_{+2} to be any column in $P \setminus \{e_1, e_2\}$, e.g., $a_{+2} = e_3$. Then, let $b_{+2} = e_1$ to generate d_{+2}. This will lead to $\beta(A_{+1}, B_{+1}, a_{+2}) = 3$, $\gamma(A_{+1}, a_{+2}, b_{+2}) = 1$, and $S^{(1,2)}(D_{+2}) = 7$.

5 Application II: selecting optimal subarrays of SOAs

Space-filling patterns and two-dimensional projection patterns are a powerful tool for evaluating designs according to their stratification properties. In this section, we apply the minimum aberration type criterion based on the space-filling pattern to rank and to select subarrays of some existing SOAs of strength 3 and SOAs of strength 2+.

First, we select subarrays of two SOAs of strength 3 according to the space-filling pattern. He and Tang constructed two nonisomorphic SOA$(54, 5, 3^3, 3)$s in He and Tang (2014) (see their Tables 1 and 2). We denote the two designs by $E^{(1)}$ and $E^{(2)}$, respectively. By calculating their space-filling pattern, we have $(S_3(E^{(1)}), S_4(E^{(1)}), S_5(E^{(1)})) = (0, 55.61, 128.22)$ and $(S_3(E^{(2)}), S_4(E^{(2)}), S_5(E^{(2)})) = (0, 53.61, 131.28)$, which implies that $E^{(2)}$ is more space-filling than $E^{(1)}$.

We select the best 4-column and 3-column subarrays from $E^{(1)}$ and $E^{(2)}$. To do this, we calculate the (S_3, S_4, S_5) values of all possible 4-column and 3-column subarrays of the two SOA$(54, 5, 3^3, 3)$s, as shown in Table 5 . The optimal space-filling patterns of the subarrays are shown in bold. The optimal 4-column subarray consists of the 1, 3, 4, 5th columns of $E^{(2)}$, and the optimal 3-column subarray consists of the 1, 3, 4th columns of $E^{(2)}$. In addition, we see that the subarrays of $E^{(2)}$ are generally better than those of $E^{(1)}$, in terms of S_4. This implies that designs with better space-filling patterns tend to have more space-filling projections.

Shi and Tang (2020) proposed and constructed SOA$(n, m, s^3, 3)$s with additional stratification properties of strength-four SOAs. The idea behind their criteria is to make most part of S_4 of the design zero. Here, our strategy is to select subarrays of

Table 5 Space-filling patterns of subdesigns of $E^{(1)}$ and $E^{(2)}$

Columns	(S_3, S_4, S_5) of	
	$E^{(1)}$'s subarrays	$E^{(2)}$'s subarrays
2, 3, 4, 5	0, 27.00, 56.39	0, 26.17, 55.56
1, 3, 4, 5	0, 26.67, 57.72	**0, 25.17, 58.78**
1, 2, 4, 5	0, 29.17, 56.72	0, 28.89, 56.44
1, 2, 3, 5	0, 29.67, 55.39	0, 28.94, 56.22
1, 2, 3, 4	0, 27.11, 54.33	0, 25.50, 57.67
1, 2, 3	0, 12.17, 18.06	0, 11.44, 18.83
1, 2, 4	0, 10.67, 19.89	0, 9.94, 19.94
1, 2, 5	0, 15.67, 19.83	0, 15.39, 19.94
1, 3, 4	0, 10.39, 19.06	**0, 9.61, 20.50**
1, 3, 5	0, 10.22, 20.78	0, 9.83, 20.94
1, 4, 5	0, 11.22, 19.44	0, 11.28, 19.72
2, 3, 4	0, 11.22, 18.56	0, 11.11, 18.56
2, 3, 5	0, 10.89, 19.11	0, 10.83, 18.61
2, 4, 5	0, 10.39, 19.61	0, 10.83, 19.06
3, 4, 5	0, 12.56, 19.17	0, 11.22, 18.56

SOA$(n, m, s^3, 3)$s by the minimum aberration type criterion based on the space-filling pattern, which shares a similar idea with Shi and Tang (2020).

Next, we consider subarrays of SOAs of strength 2+. Shi and Tang (2019) addressed the problem of design selection for some existing SOAs of strength 2+ by examining their three-dimensional projections, but they focused only on the two-level SOAs. This task can be better completed by comparing the space-filling pattern of the subarrays, especially for $s \neq 2$. With a fixed A, as long as B conforms to the conditions required for construction, all SOAs of strength 2+, $D = sA + B$, have the same $S^{(1,1)}$, $S^{(1,2)}$ and S_1, S_2, S_3.

Now, we illustrate some optimal subarrays of three existing SOAs of strength 2+ with $s \geq 3$. We still focus on SOAs of strength 2+ constructed from regular designs and follow the notation in Sect. 4.2. The three SOAs $D^{(1)}$, $D^{(2)}$ and $D^{(3)}$ are listed below, constructed by the methods in He et al. (2018).

- For $s = 3$, an SOA(27, 6, 9, 2+) can be constructed by $D^{(1)} = 3A^{(1)} + B^{(1)}$ with

$$A^{(1)} = \{e_1e_2^2, e_1e_3^2, e_2e_3^2, e_1e_2^2e_3, e_1e_2e_3^2, e_1e_2^2e_3^2\},$$
$$B^{(1)} = \{e_1, e_1, e_2, e_2, e_3, e_1\}.$$

- For $s = 4$, an SOA(64, 8, 16, 2+) can be constructed by $D^{(2)} = 4A^{(2)} + B^{(2)}$ with

$$A^{(2)} = \{e_1e_3^{1+x}, e_2e_3^{1+x},$$
$$e_1e_2e_3^{1+x}, e_1e_2^xe_3^{1+x}, e_1e_2^{1+x}, e_1e_2^{1+x}e_3, e_1e_2^{1+x}e_3^x, e_1e_2^{1+x}e_3^{1+x}\},$$

Table 6 Optimal subarrays of strength 2+ SOAs $D^{(1)}$, $D^{(2)}$, $D^{(3)}$ and $D^{(4)}$

Original	Subarray	Columns	$S^{(2,2)}$, S_3, S_4, S_5
$D^{(1)}$	5-column	1,2,3,4,5	20*, 4*, 70, 98
$D^{(2)}$	7-column	1,2,3,4,5,6,7	63*, 15*, 378, 1365
	6-column	1,3,4,5,6,7	45*, 6*, 234, 630
$D^{(3)}$	9-column	1,2,3,4,5,6,7,8,9	144*, 48*, 1264, 8928
	8-column	1,2,3,4,5,6,7,8	112*, 32*, 828, 5216
	7-column	1,2,3,4,5,7,9	84*, 16*, 532, 2800
$D^{(4)}$	11-column	1,2,4,5,6,7,8,9,10,11,12	63.5*, 27.5, 440.5, 2120
	11-column	2,3,4,5,6,7,8,9,10,11,12	74, 23*, 460, 2123
	10-column	1,2,4,5,6,8,9,10,11,12	50*, 21.5, 310.5, 1382
	10-column	2,4,5,6,7,8,9,10,11,12	54, 15*, 333, 1398
	9-column	1,2,4,5,8,9,10,11,12	38.5*, 17.5, 206.5, 860
	9-column	4,5,6,7,8,9,10,11,12	42, 7*, 246, 864
	8-column	1,2,4,5,8,9,10,11	28.5*, 11, 142.5, 500
	8-column	4,5,6,7,8,9,10,11	32, 4*, 165.5, 505.5
	7-column	1,2,4,8,9,10,11	20*, 8.5, 84.5, 275
	7-column	4,5,7,8,9,10,12	22.5, 2.5*, 100.5, 285
	6-column	1,4,5,8,9,10	13*, 3, 53, 144.5
	6-column	2,5,7,8,9,11	17, 0.5*, 63, 134

We mark an $S^{(2,2)}$ entry or an S_3 entry with an asterisk if it is optimal

$$B^{(2)} = \{e_3, e_3, e_3, e_3, e_2, e_2, e_2, e_2e_3\}.$$

- For $s = 5$, an SOA(125, 10, 25, 2+) can be constructed by $D^{(3)} = 5A^{(3)} + B^{(3)}$ with

$$A^{(3)} = \{e_1e_2^4, e_1e_3^4, e_2e_3^4, e_1e_2e_3^4, e_1e_2^2e_3^4, e_1e_2^3e_3^4, e_1e_2^4e_3, e_1e_2^4e_3^2, e_1e_2^4e_3^3, e_1e_2^4e_3^4\},$$
$$B^{(3)} = \{e_2, e_3, e_3, e_3, e_3, e_3, e_2, e_2, e_2, e_2e_3\}.$$

We select the 5-column subarray of $D^{(1)}$, the 6 and 7-column subarrays of $D^{(2)}$, and the 7, 8 and 9-column subarrays of $D^{(3)}$. By comparing the space-filling patterns and two-projection patterns of all possible subarrays, we obtain the optimal subarrays, as shown in Table 6.

Recently, Chen and Tang (2022b) found an SOA(54, 12, 9, 2+). Compared with the existing 54-run SOAs of strength 3, the number of factors it can accommodate is more than double. We denote SOA(54, 12, 9, 2+) in Chen and Tang (2022b) by $D^{(4)}$. We obtain the optimal subarrays in the same way as above and list them in Table 6. Unlike the subarrays of $D^{(1)}$, $D^{(2)}$ and $D^{(3)}$, the optimal subarrays under the two criteria are usually different.

6 Discussion

In this study, we propose a new definition of the space-filling pattern introduced by Tian and Xu (2022). Our definition is more general and simplifies the computation to some extent. To account for the importance of the two-dimensional projections of designs and the computational efficiency, we also introduce the two-dimensional projection pattern as a variant of the space-filling pattern. This variant has a simpler form and reveals the stratification properties of the designs on two-dimensional projections for a broad class of designs. We develop a strategy to augment existing GSOAs of strength 3 and SOAs of strength 2+ with additional columns by using the space-filling pattern and the two-dimensional projection pattern, and we obtain some column-augmented designs with desirable space-filling properties. We also use the space-filling pattern to select optimal subarrays from some SOAs of strength 3 and 2+. Exploring the relationship between the two-dimensional projection pattern and other criteria based on the design's two-dimensional projections, such as the uniform projection criterion (Sun et al. 2019), is our follow-up goal.

Acknowledgements We are grateful to the referee editor and two reviewers for their insightful comments and constructive suggestions. This study was supported by the National Natural Science Foundation of China (Grant Nos. 12371259, 11971098, 12271166, 11901199 and 12061033) and the National Key Research and Development Program of China (Nos. 2020YFA0714102 and 2022YFA1003701).

Declarations

Conflict of interest The authors state that there is no conflict of interest.

References

Chen G, Tang B (2022a) A study of orthogonal array-based designs under a broad class of space-filling criteria. Ann Stat 50(5):2925–2949

Chen G, Tang B (2022b) Using nonregular designs to generate space-filling designs. J Stat Plan Inference 221:201–211

Cheng S-W, Ye KQ (2004) Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. Ann Stat 32(5):2168–2185

Cheng C-S, He Y, Tang B (2020) Minimal second order saturated designs and their applications to space-filling designs. Stat Sin 31(2):867–890

Fang K-T, Lin DKJ, Winker P, Zhang Y (2000) Uniform design: theory and application. Technometrics 42(3):237–248

He Y, Tang B (2012) Strong orthogonal arrays and associated Latin hypercubes for computer experiments. Biometrika 100(1):254–260

He Y, Tang B (2014) A characterization of strong orthogonal arrays of strength three. Ann Stat 42(4):1347–1360

He Y, Cheng C-S, Tang B (2018) Strong orthogonal arrays of strength two plus. Ann Stat 46(2):457–468

Johnson ME, Moore LM, Ylvisaker D (1990) Minimax and maximin distance designs. J Stat Plan Inference 26(2):131–148

Liu H, Liu M-Q (2015) Column-orthogonal strong orthogonal arrays and sliced strong orthogonal arrays. Stat Sin 25(4):1713–1734

Mckay MD, Beckman RJ, Conover WJ (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

Owen AB (1992) Orthogonal arrays for computer experiments, integration and visualization. Stat Sin 2(2):439–452

Owen AB (1994) Controlling correlations in Latin hypercube samples. J Am Stat Assoc 89(428):1517–1522

Santner TJ, Williams BJ, Notz WI (2018) The design and analysis of computer experiments, 2nd edn. Springer, New York

Shi C, Tang B (2019) Design selection for strong orthogonal arrays. Can J Stat 47(2):302–314

Shi C, Tang B (2020) Construction results for strong orthogonal arrays of strength three. Bernoulli 26(1):418–431

Sun F, Wang Y, Xu H (2019) Uniform projection designs. Ann Stat 47(1):641–661

Tang B (1993) Orthogonal array-based Latin hypercubes. J Am Stat Assoc 88(424):1392–1397

Tian Y, Xu H (2022) A minimum aberration-type criterion for selecting space-filling designs. Biometrika 109(2):489–501

Tian Y, Xu H (2023) Stratification pattern enumerator and its applications. J R Stat Soc Ser B Stat Methodol 86(2):364–385

Wang Y, Sun F, Xu H (2022) On design orthogonality, maximin distance, and projection uniformity for computer experiments. J Am Stat Assoc 117(537):375–385

Xu H, Wu CFJ (2001) Generalized minimum aberration for asymmetrical fractional factorial designs. Ann Stat 29(4):1066–1077

Zhou Y, Tang B (2019) Column-orthogonal strong orthogonal arrays of strength two plus and three minus. Biometrika 106(4):997–1004