Journal of Statistical Planning and Inference 232 (2024) 106150

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Construction of high-dimensional high-separation distance designs

Xu He [a], Fasheng Sun [b,*]

[a] MADIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
[b] KLAS and School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, China

ARTICLE INFO

MSC:
62K99

Keywords:
Gaussian process model
Hadamard matrix
Kriging
Orthogonal array
Space-filling design

ABSTRACT

Space-filling designs that possess high separation distance are useful for computer experiments. We propose a novel method to construct high-dimensional high-separation distance designs. The construction involves taking the Kronecker product of sub-Hadamard matrices and rotation. In addition to possessing better separation distance than most existing types of space-filling designs, our newly proposed designs enjoy orthogonality and projection uniformity and are more flexible in the numbers of runs and factors than that from most algebraic constructions. From numerical results, such designs are excellent in Gaussian process emulation of high-dimensional computer experiments. An R package on design construction is available online.

1. Introduction

Space-filling designs that distribute points apart from each other are useful to determine input values of time-consuming computer simulations (Santner et al., 2018). The L_2 separation distance of a design $D \subset [0,1]^p$ is the minimum distance between a pair of design points,

$$d(D) = \min_{\mathbf{x},\mathbf{y} \in D} \left\{ \sum_{k=1}^{p} |x_k - y_k|^2 \right\}^{1/2},$$

and designs that achieve the maximal separation distance are called maximin distance designs. Johnson et al. (1990) showed that maximin distance designs are desirable for computer experiments because they are asymptotically D-optimal for Gaussian process models with stationary and isotropic correlation functions in the sense that determinant of correlation matrices are maximized. Haaland and Qian (2011) and He and Chien (2018) showed that designs with high separation distance are appealing in reducing numeric error, which may become the major source of emulation error for experiments with large sample size. A design with n points and p factors can also be expressed as an $n \times p$ matrix whose ith row represents the ith design point. Using this definition, a design is called orthogonal if its column-wise correlations are all zero. As discussed in Owen (1994) and Bingham et al. (2009), orthogonality is a desirable property for designs of computer experiment. Furthermore, low-level orthogonal designs are also useful in constructing higher-level space-filling designs (Steinberg and Lin, 2006; Sun and Tang, 2017a).

There is a recent surge of interests on the construction of maximin distance designs or designs with high separation distance. Algebraic constructions include Fries and Hunter (1980), Zhou and Xu (2014, 2015), Sun and Tang (2017b), Xiao and Xu (2017), Wang et al. (2018a,b), Zhou et al. (2020), Yang et al. (2021), Li et al. (2021), Wang et al. (2022), and Yuan et al. (2025), among others. Due to the complexity of mathematical tools, each of these methods has its own constraint on n, p, and s, the number of

* Corresponding author.
E-mail address: sunfs359@nenu.edu.cn (F. Sun).

https://doi.org/10.1016/j.jspi.2024.106150
Received 15 April 2023; Received in revised form 5 October 2023; Accepted 26 January 2024
Available online 1 February 2024

levels. On the other hand, algorithmic methods such as maximin distance Latin hypercube designs (Ba et al., 2015), maximum projection designs (Joseph et al., 2015), interleaved lattice-based maximin distance designs (He, 2019), and orthogonal-maximin Latin hypercube designs (Joseph and Hung, 2008) become inefficient for high-dimensional problems. Finally, existing types of designs that possess projection uniformity, such as orthogonal array-based Latin hypercube designs (Tang, 1993), orthogonal designs (Sun and Tang, 2017a), and strong orthogonal arrays (Sun and Tang, 2023), allow flexible choices of n and p but have no separation distance property.

In this paper, we propose a method to construct high-dimensional high-separation distance designs in 2, 4, 8, and 16 levels. We use Kronecker product of Hadamard matrices and rotation techniques in the construction. The method is applicable to n that is a multiple of 16 with p such that $p < n$, $p < n-1$, $p < n-3$, and $p < n-3$ when s is 2, 4, 8, and 16, respectively. Compared to orthogonal array-based Latin hypercube designs and orthogonal designs, our proposed designs enjoy additional separation distance properties. Compared to most other constructions of high-separation distance designs, our method yields designs that possess better separation distance and additional orthogonality and projection uniformity properties. From numerical results our newly proposed designs outperform several existing types of space-filling designs in emulating high-dimensional computer experiments.

The rest of the paper is organized as follows. We give preliminary definitions and properties of Hadamard matrices and separation distance efficiency in Section 2. In Section 3, we give our construction method and theoretical results. In Sections 4 and 5, we numerically compare designs generated from several constructions using test functions and a real example, respectively. Some final remarks are given in Section 6. Examples to illustrate our algorithms, proofs of theorems as well as some extra theoretical results are provided in supplementary material. Computer code to construct our proposed designs is included in the R package titled "HDOMDesign", which has been distributed to the Comprehensive R Archive Network.

2. Hadamard matrices and separation distance

An $m \times m$ matrix H_m of $+1$ and -1 is called a Hadamard matrix if $H_m H_m^T = mI_m$, where I_m denotes the $m \times m$ identity matrix. A Hadamard matrix has orthogonal rows and columns. If H_a and H_b are two Hadamard matrices of orders a and b, respectively, then $H_a \otimes H_b$, i.e., the Kronecker product of H_a and H_b, is a Hadamard matrix of order ab. Hadamard matrices whose order is a power of 2 can be constructed by applying

$$H_1 = \left(+1\right), \quad H_2 = \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}, \tag{1}$$

and $H_m = H_2 \otimes H_{m/2}$ for $m = 2^z$ and $z \in \mathbb{N}$. Such constructed matrices are called Walsh Hadamard matrices. There are sometimes multiple ways to construct a Hadamard matrix of given order. Throughout this paper, we only consider the matrix be generated via $H_m = H_{2^z} \otimes H_{m/2^z}$ with the highest possible z, where H_{2^z} is a Walsh Hadamard matrix. We also require the first column of the matrix be consisted of "+1" only. Clearly, each of the other columns of them consists of half "+1" and half "−1". Such Hadamard matrices exist for orders 1, 2, all multiples of 4 that is no greater than 664, and most higher multiples of 4. See Hedayat et al. (1999) for further reference on Hadamard matrices.

A design is called a U-type design if its levels are equally spaced and appear equally often in each of its columns. Zhou and Xu (2015) derived that

$$d(D)^2 \leq \lfloor \{n/(n-1)\}\{(s+1)(s-1)\}p/6 \rfloor \tag{2}$$

for any U-type design with n points, p dimensions, and levels $\{1, \dots, s\}$, where $\lfloor x \rfloor$ denotes the highest integer not exceeding x. To apply such a design to a specific problem, one needs to relabel the levels based on the range of input values. For example, we can use either $(D-1)/(s-1)$ or $(D-1/2)/s$ if the input space is $[0,1]^p$. While the former transformation produces designs with maximized separation distance, designs obtained from the latter transformation is more "uniformly" distributed in the input space in the sense that the s levels locate in the center of the s equally spaced intervals of $[0, 1/s], [1/s, 2/s], \dots, [(s-1)/s, 1]$. Provided that the goal is to sample representative points of the input space, the latter transformation is clearly most suitable, but for the emulation purpose it is not clear which transformation is more desirable. Nevertheless, U-type designs with different scales can be fairly compared using separation distance efficiency (Li et al., 2021) defined by

$$e(D) = d(D) \left[\{n/(n-1)\}\{(s+1)(s-1)\}p/6\right]^{-1/2} / z, \tag{3}$$

where n, p, s, and z denote the number of points, the number of factors, the number of levels, and the gap distance between the levels of the U-type design D, respectively. Furthermore, a little derivation based on (2) reveals that designs with lower s tend to have much higher separation distance than designs with higher s from the $(D-1)/(s-1)$ transformation, while from the $(D-1/2)/s$ transformation designs with higher s tend to have slightly higher separation distance. From using $e(D)$ in (3), U-type designs of different s can as well be fairly compared. Also note that in the definition of $e(D)$ we do not round $\{n/(n-1)\}\{(s+1)(s-1)\}p/6$, as is done in (2). This is to make the $e(D)$ function smoother. From the definition, a design whose efficiency is close to 1 is clearly nearly optimal in separation distance. On the other hand, the efficiency of a nearly optimal design may not necessarily be close to one since the bound in (2) may not be tight.

3. Construction

In this section, we give our method to construct high-dimensional high-separation distance designs, beginning with the method to construct binary designs. Our constructions make use of three types of sub-Hadamard matrices that possess high separation distance. For a binary matrix M, let M_P denotes the sub-matrix of M consisting of columns that are indexed by P. When B is a $b \times b$ Hadamard matrix, we use B_p where the index set P is given by

$$P = \begin{cases} \{b-p+1,\dots,b\}, & b/2 < p < b, \\ L(b), & p \le b/2 \text{ and } b = 2^{p-1} \ge 16, \\ \{b/2-p+1,\dots,b/2\}, & \text{otherwise}, \end{cases} \tag{4}$$

where

$$L(b) = \begin{cases} \{b-2^{\log_2(b)-1},\dots,b-2^0,b\}, & \log_2(b) \text{ is even}, \\ \{2,b-2^{\log_2(b)-1},\dots,b-2^1,b\}, & \log_2(b) \text{ is odd}. \end{cases} \tag{5}$$

Let $\rho(M)$ denote the minimum number of positions not in common among the row pairs of M, $I()$ denote the identity function, and $\lambda(b,p)$ be 0, $b-p-2$, $p-4$, and $b/2-p$ when $p=b$, $b/2<p<b$, $b=2^{p-1}\ge 16$, and other cases with $b/4<p\le b/2$, respectively. Proposition 1 below summarizes the separation distance property of sub-Hadamard matrices introduced in (4).

Proposition 1. *Suppose B_P is a $p\times b$ sub-Hadamard matrix with either $b/4 < p \le b$ or $b = 2^{p-1} \ge 16$ and P is given by (4). When $b = 2^{p-1} \ge 16$, further suppose that B is the Walsh Hadamard matrix of order b. When $b/4 < p \le b/2$, further suppose that B can be expressed as the Kronecker product of H_2 in (1) and a Hadamard matrix of order $b/2$.*

Then columns of B_P are mutually orthogonal, $\rho(B_P) \ge p/2 - \lambda(b,p)/2$, and $e(B_P)^2 \ge \{1-\lambda(b,p)/p\}(1-1/n)$. Furthermore, each column of B_P consists of half $+1$ and half -1 unless $p=n$.

Indicated from Proposition 1, whenever $\lambda(b,p)/p$ is close to zero, B_P is nearly optimal in separation distance. That is to say, for p/b being either nearly 1 or nearly 1/2, we can generate nearly optimal designs and the choices of b or p are flexible. In particular, from some derivations based on (2), B_P is optimal when $b-3 \le p \le b$ or $b/2-1 \le p \le b/2$. Although not guaranteed to be optimal, designs with $b = 2^{p-1}$ are also appealing as we are not aware of better designs with such low p/b. On the other hand, for many other p/n the generated designs are not excellent.

To construct nearly optimal designs with more flexible p/n, we consider Kronecker product of sub-Hadamard matrices. Theorem 1 below gives the basic properties of product sub-Hadamard matrices.

Theorem 1. *Suppose A and B are two Hadamard matrices of orders a and b, respectively, $H = A\otimes B$, and $Q = \bar{P} \cup \{x+yb : x\in \tilde{P}, 1 \le y \le a-1\}$. Then*

$$\rho(H_Q) \ge q/2 - \max\{(a-1)\lambda(b,\tilde{p})/2 + \lambda(b,\bar{p})/2, \bar{p}/2-\tilde{p}/2, \tilde{p}/2+\lambda(b,\bar{p})/2\}.$$

From Theorem 1, H_Q has favorable separation distance whenever both $B_{\tilde{P}}$ and $B_{\bar{P}}$ are nearly optimal in separation distance. An improved but much more complicated bound on separation distance of H_Q is provided in supplementary material. For a set Z, let $|Z|$ denote its cardinality. Theorem 2 below shows two further ways to make the p flexible:

Theorem 2. *Suppose $U \subset \{2,\dots,b\}\setminus P$ and $V \subset P$. Then $e(H_{Q\cup U})^2 \ge e(H_Q)^2|Q|/(|Q|+|U|)$ and $e(H_{Q\setminus V})^2 \ge e(H_Q)^2|Q|/(|Q|-|V|) - 2(1-1/n)|V|/(|Q|-|V|)$.*

From Theorem 2, provided that $|U|$ and $|V|$ are small compared to $|Q|$, $H_{Q\cup U}$ and $H_{Q\setminus V}$ are also appealing. This allows us to add or remove a few columns from H_Q. Employing the above techniques, we propose to construct 2-level designs using Algorithm 1. An example to illustrate Algorithm 1 is provided in the supplementary material.

In Algorithm 1, we try different combinations of a, b, $\tilde{p}$, and $\bar{p}$, hoping to find one choice that yield a high-separation distance design. From Algorithm 1, a much divide n, b is determined by (n,a), there are at most $(b+1)/(a-1)+1$ choices of $\tilde{p}$, and two choices of $\bar{p}$. Consequently, Algorithm 1 does not need many iterations and thus is fast in computation.

Because the order of most Hadamard matrices is a multiple of four, most choices of a and b are multiples of four, requiring n to be a multiple of 16. Besides, to allow orthogonal columns, p has to be less than n. These are the major constraints on (n,p), which is much looser than most algebraic constructions of high-separation distance designs.

Next, we propose the method to construct four-level high-separation distance designs from rotating binary designs. Steinberg and Lin (2006) showed that four-level U-type designs can be obtained by rotating an orthogonal binary design in groups of two using the rotation matrix

$$R_2 = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} /5^{1/2}.$$

A set $\mathscr{G}$ is called a k-grouping of $J \subset \mathbb{N}$ if all elements of $\mathscr{G}$ are k-vectors, the entries in these vectors are mutually different, and the entries combined cover J. Let $R_{2,\mathscr{G}}$ denote the $n\times n$ block diagonal matrix such that for each $G\in\mathscr{G}$ the block indexed by G is R_2 and all other diagonal entries are one. A U-type design with even number of equally spaced levels $z-(s-1)l, z-(s-3)l,\dots,z+(s-1)l$

Algorithm 1: Procedure to generate two-level designs

initialize B_P from Proposition 1, $D_{\mathrm{OPT}} = (1 - B_P)/2$, and $d_{\mathrm{MAX}} = d(D_{\mathrm{OPT}})$;
for (a, b) *such that Hadamard matrices of orders* $a \geq 2$ *and* $b \geq 2$ *exist* **do**
  obtain Hadamard matrices A and B and compute $H = A \otimes B$;
  for $\tilde{p}$ *such that* $\lfloor (p - b - 1)/(a - 1) \rfloor \leq \tilde{p} \leq \lceil p/(a - 1) \rceil$ **do**
    if $\lambda(b, \tilde{p}) > 2$ and $b \neq 2^{\tilde{p}-1}$, **continue** to the next $\tilde{p}$;
    generate the indices set $\tilde{P}$ for $\tilde{p} \times b$ matrix given by (4);
    for $\bar{p}$ *being either (the lowest integer no less than* $p - (a - 1)\tilde{p}$ *such that* $\lambda(b, \bar{p}) \leq 2$ *or* $b = 2^{\bar{p}-1}$*) or (the greatest integer no more than* $p - (a - 1)\tilde{p}$ *such that* $\lambda(b, \bar{p}) \leq 2$*, or* $b = 2^{\bar{p}-1}$*, or* $\bar{p} = 0$*)* **do**
      if $\bar{p} = 0$, let $\bar{P} = \emptyset$; **otherwise** generate $\bar{P}$ from (4);
      let $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a - 1\}$ and $q = |Q|$;
      if $q = p$, let $P = Q$;
      if $q > p$, select a V such that $V \subset Q$ and $v = q - p$ and let $P = Q \setminus V$;
      if $q < p$, select a U such that $U \subset \{2, \ldots, n\} \setminus Q$ and $u = p - q$ and let $P = Q \cup U$;
      let $D = (1 - H_P)/2$ and compute $d(D)$;
      if $d(D) > d_{\mathrm{MAX}}$, update $d_{\mathrm{MAX}} = d(D)$ and $D_{\mathrm{OPT}} = D$;
    end
  end
end
return D_{OPT}.

is said to possess 2×2 projection uniformity if all of its column pairs have equal number of rows in the four 2×2 bins of $[z-(s-1)l, z-l] \times [z-(s-1)l, z-l]$, $[z-(s-1)l, z-l] \times [z+l, z+(s-1)l]$, $[z+l, z+(s-1)l] \times [z-(s-1)l, z-l]$, and $[z+l, z+(s-1)l] \times [z+l, z+(s-1)l]$. Theorem 3 below gives the construction method.

Theorem 3. *Suppose* $1 \leq p \leq n - 2$, $H = A \otimes B$ *is an* $n \times n$ *Hadamard matrix,* $Q \subset \{2, \ldots, n\}$, $|Q| = q$, *and* $\lceil q/2 \rceil = \lceil p/2 \rceil$. *Then there exists* $U \subset \{2, \ldots, n\} \setminus Q$, $V \subset Q \cup U$, *and* $\mathscr{G}$ *such that* $|U| \leq 1$, $|V| \leq 1$, $\mathscr{G}$ *is a 2-grouping of* $Q \cup U$, $D = (HR_{2,\mathscr{G}})_P$ *is an* $n \times p$ *four-level orthogonal U-type design that possesses* 2×2 *projection uniformity, and*

$$e(D)^2 \geq e(H_Q)^2 q/p - 2(1 - 1/n)I(q \geq p)\{q - p - (4/5)I(p \text{ is odd})\}/p. \tag{6}$$

Remark that 2-groupings must have even number of elements. Consequently, after obtaining an H_Q that process high separation distance where q is odd, we have to rotate $H_{Q \cup U}$ instead of rotate H_Q. When $U = V = \emptyset$, we have $q = p$ and $e(D)^2 = e(H_Q)^2$, i.e., there is no loss on separation distance efficiency during the rotation step. In other cases, the loss is small for big p. That is to say, D has appealing separation distance if and only if H_P has appealing separation distance. Because the n and p are relatively flexible for binary designs, they are also flexible for 4-level designs. Nevertheless, using the above approach it is not possible to construct a design with $p = n - 1$ columns.

Similar to Algorithm 1, we propose to construct four-level designs for given n and $p < n-1$ using Algorithm 2, which is illustrated by an example provided in the supplementary material.

Finally, we propose our method to construct 8-level and 16-level designs using the rotation matrices

$$R_3 = \begin{pmatrix} 4 & -2 & -1 & 0 \\ 2 & 4 & 0 & 1 \\ 1 & 0 & 4 & -2 \\ 0 & -1 & 2 & 4 \end{pmatrix} / \sqrt{21} \text{ and } R_4 = \begin{pmatrix} 8 & -4 & -2 & 1 \\ 4 & 8 & -1 & -2 \\ 2 & -1 & 8 & -4 \\ 1 & 2 & 4 & 8 \end{pmatrix} / \sqrt{85}.$$

A binary matrix is called k-orthogonal if from all of its k-column submatrices the 2^k level combinations occur equally often from the rows. A binary design H_Q that is a sub-design of H is called j-orthogonal with k-grouping $\mathscr{G}$ if $\mathscr{G}$ is a k-grouping of Q and H_G is j-orthogonal for any $G \in \mathscr{G}$. For a 4-grouping $\mathscr{G}$ and $j \in \{3, 4\}$, let $R_{j,\mathscr{G}}$ denote the $n \times n$ block diagonal matrix such that for each $G \in \mathscr{G}$ the block indexed by G is R_j and all other diagonal entries are one. Sun and Tang (2017a) showed that if a binary balanced design is j-orthogonal with a 4-grouping $\mathscr{G}$, from rotating it by groups using $R_{j,\mathscr{G}}$ we can obtain a 2^j-level orthogonal U-type design with 2×2 projection uniformity. Zhou et al. (2020) generated high-separation distance designs by rotating full factorial designs in groups. However, their construction requires n to be a power of 2 and p to be certain multiples of $n/4$. We summarize their results that are relevant to our construction in Lemma 1 below.

Lemma 1. *Suppose* $k \in \{3, 4\}$, H *is a binary orthogonal U-type design,* $\mathscr{G}$ *is a 4-grouping of* Q, H_Q *is* k*-orthogonal with* $\mathscr{G}$, $W = HR_{k,\mathscr{G}}$, *and* $D = W_Q$. *Then* D *is a* 2^k*-level orthogonal U-type design with* 2×2 *projection uniformity,* $d(D) = d(H_Q)$, *and* $e(D) = e(H_Q)$.

From Lemma 1, D is optimal and nearly optimal in separation distance if and only if H_Q is. we can thus use H_Q to construct 8-level and 16-level designs provided that they are 3-orthogonal or 4-orthogonal, respectively, with a 4-grouping. To uncover whether they can be partitioned into orthogonal groups, we give four types of orthogonal 4-groups in Proposition 2 below.

Algorithm 2: Procedure to generate four-level designs

Initialize H_Q be obtained from Proposition 1, choose U, V, and $\mathscr{G}$ so that (6) is valid, compute $P = (Q \cup U) \setminus V$, R, $W = HR$, $D = W_P$, and $d(D)$;
for (a, b) *such that Hadamard matrices of orders* $a \geq 2$ *and* $b \geq 2$ *exist* **do**
  obtain Hadamard matrices A and B and compute $H = A \otimes B$;
  for $\tilde{p}$ *such that* $\lfloor (p - b - 1)/(a - 1) \rfloor \leq \tilde{p} \leq \lceil p/(a - 1) \rceil$ **do**
    if $\lambda(b, \tilde{p}) > 2$ and $b \neq 2^{\tilde{p}-1}$, **continue** to the next $\tilde{p}$;
    generate $\tilde{P}$ from (4);
    for $\bar{p}$ *being either (the lowest integer no less than* $p - (a - 1)\tilde{p}$ *such that* $\lambda(b, \bar{p}) \leq 2$ *or* $b = 2^{\bar{p}-1}$*) or (the greatest integer no more than* $p - (a - 1)\tilde{p}$ *such that* $\lambda(b, \bar{p}) \leq 2$*, or* $b = 2^{\bar{p}-1}$*, or* $\bar{p} = 0$*)* **do**
      if $\bar{p} = 0$, let $\bar{P} = \emptyset$; **otherwise** generate $\bar{P}$ from (4);
      let $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a - 1\}$ and $q = |Q|$;
      choose U, V, and $\mathscr{G}$ so that (6) is valid, compute $P = (Q \cup U) \setminus V$, R, $W = HR$, $D = W_P$, and $d(D)$;
      if $d(D) > d_{\mathrm{MAX}}$, update $d_{\mathrm{MAX}} = d(D)$ and $D_{\mathrm{OPT}} = D$;
    end
  end
end
return D_{OPT}.

Proposition 2. *Suppose $H = A \otimes B$, A and B are two Hadamard matrices of orders a and b, respectively, $G = \{(i-1)b + x, (j-1)b + y, (k-1)b+z, (l-1)b+w\} \subset \{2, \ldots, ab\}$, $i, j, k, l \in \{1, \ldots, a\}$, and $x, y, z, w \in \{1, \ldots, b\}$. Then H_G is 3-orthogonal and 4-orthogonal provided that (a) $1 \neq i = j = k \neq l \neq 1$; or (b) $1 \neq i = j \neq k = l \neq 1$, $x = z$, and $y \neq w$; or (c) $i = j$, the $i, k, l, 1$ are mutually different, $z = w$, $x \neq 1$, and $y \neq 1$. Besides, H_G is 3-orthogonal provided that $i = j$, $k = l$, $x = z$, and $y = w$.*

Exploiting Propositions 2, we show in Theorem 4 below that most H_Q obtained from Theorem 1 with $a \geq 2$ can be supplemented to a binary orthogonal U-type design that is 3-orthogonal with a 4-grouping and we can thus construct 8-level high-separation distance designs from rotating them.

Theorem 4. *Suppose A and B are two Hadamard matrices of orders $a \geq 2$ and $b \geq 4$, respectively, $H = A \otimes B$, $p \leq ab - 4$, $\tilde{P} \subset \{1, \ldots, b\}$, $\bar{P} \subset \{2, \ldots, b\}$, $1 \notin \tilde{P}$ as long as $\tilde{p} < b$, $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a - 1\}$, and $|Q| = q$. When $a \geq 4$, further suppose $\tilde{p} \geq 2$, $\bar{p} \leq (a-3)\tilde{p} + 2$, $q \leq ab - 4$, and $\hat{p} = \max(4\lceil p/4 \rceil - q, 4\lceil q/4 \rceil - q) \leq b - 1$. When $a = 2$, further suppose $\tilde{p} \leq b - 2$, $\bar{p} \leq b - 2$, either $\tilde{P} \subset \bar{P}$ or $\bar{P} \subset \tilde{P}$, and $\hat{p} = 4\max(\lceil p/4 \rceil, \lceil \tilde{p}/2 \rceil, \lceil \bar{p}/2 \rceil) - q$.*

Then there exist sets $U, V \subset \{2, \ldots, ab\}$ and a 4-grouping $\mathscr{G}$ of $Q \cup U$ such that $U \cap Q = \emptyset$, $|U| = \hat{p}$, $V \subset (U \cup Q)$, $|V| = q + |V| - p$, and $H_{U \cup Q}$ is 3-orthogonal with $\mathscr{G}$.

Let $P = (U \cup Q) \setminus V$, $W = HR_{3,\mathscr{G}}$, and $D = W_P$. Then D is an 8-level orthogonal U-type design with $n = ab$ points and $p = |P| = (a-1)\tilde{p} + \bar{p} + u - v$ factors, D possesses 2×2 projection uniformity, and

$$e(D)^2 \geq e(H_Q)^2 q/p - (1 - 1/n)(14/3)(v/p).$$

From Theorem 4, the difference on separation distance efficiency between binary designs H_Q and 8-level designs D is small for large p. Further suppose that either (a) $a \geq 4$ and $\tilde{p} \leq b - 3$; or (b) $a = 2$ and $\max(\lceil \tilde{p}/2 \rceil, \lceil \bar{p}/2 \rceil) - \min(\lfloor \tilde{p}/2 \rfloor, \lfloor \bar{p}/2 \rfloor) \leq 2$, the gap in $e(D)^2$ and $e(H_Q)^2 q/p$ can be even smaller than $(1 - 1/n)(14/3)(v/p)$. However, because the formula is quite complicated, we provide the results in the supplementary material. In light to Theorem 4, we recommend to generate 8-level designs using Algorithm 3. An example to illustrate Algorithm 3 is provided in the supplementary material.

Similar to Theorem 4, we show in Theorem 5 below that most H_Q obtained from Theorem 1 with $a \geq 4$ can be supplemented to a binary orthogonal U-type design that is 4-orthogonal with a 4-grouping and we can thus construct 16-level high-separation distance designs from rotating them.

Theorem 5. *Suppose A and B are two Hadamard matrices of orders $a \geq 4$ and $b \geq 4$, respectively, $H = A \otimes B$, $p \leq ab - 4$, $\tilde{P} \subset \{1, \ldots, b\}$, $\bar{P} \subset \{2, \ldots, b\}$, $\tilde{p} \geq 2$, $1 \notin \tilde{P}$ as long as $\tilde{p} < b$, $\bar{p} \leq (a-3)\tilde{p} + 2$, $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a - 1\}$, $q = |Q| \leq ab - 4$, and $\hat{p} = \max(4\lceil p/4 \rceil - q, 4\lceil q/4 \rceil - q) \leq b - 1$.*

Then there exist sets $U, V \subset \{2, \ldots, ab\}$ and a 4-grouping $\mathscr{G}$ of $U \cup Q$ such that $U \cap Q = \emptyset$, $|U| = \hat{p}$, $V \subset (U \cup Q)$, $|V| = q + |U| - p$, and $H_{U \cup Q}$ is 4-orthogonal with the grouping $\mathscr{G}$.

Let $P = (U \cup Q) \setminus V$, $W = HR_{4,\mathscr{G}}$, and $D = W_P$. Then D is a 16-level orthogonal U-type design with $n = ab$ points and $p = |P| = (a-1)\tilde{p} + \bar{p} + u - v$ factors, D possesses 2×2 projection uniformity, and

$$e(D)^2 \geq e(H_Q)^2 q/p - (1 - 1/n)(90/17)(v/p).$$

From Theorem 5, we can construct nearly optimal 16-level designs from rotating nearly optimal binary designs. Provided that $\tilde{p} \leq b - 3$, there exist U, V, and $\mathscr{G}$ such that the difference between $e(D)^2$ and $e(H_Q)^2 q/p$ is even smaller. This result is provided in the supplementary material. The algorithm to construct 16-level designs is the same to Algorithm 3 excepts that Steps 14–17 are removed and the $(U, V, \mathscr{G})$ that fulfills Theorem 4 in Steps 7 and 10 are replaced by $(U, V, \mathscr{G})$ that fulfills Theorem 5, respectively.

Algorithm 3: Procedure to generate eight-level designs

initialize $d_{\text{MAX}} = -1$;
for (a, b) *such that Hadamard matrices of orders* $a \geq 4$ *and* $b \geq 4$ *exist* **do**
  obtain Hadamard matrices A and B and compute $H = A \otimes B$;
  for $\tilde{p}$ *with* $\tilde{p} \geq 2$ *and* $\lfloor (p-b-1)/(a-1) \rfloor \leq \tilde{p} \leq \lceil p/(a-1) \rceil$ **do**
    if b is a power of 2 and $\log_2(b) + 1 \leq \tilde{p} < b/4 + 2$, let $\tilde{P}$ be the union of $L(b)$ in (5) and the $\tilde{p} - |L(b)|$ largest elements of $\{1, \ldots, b\} \setminus L(b)$, **otherwise** let $\tilde{P} = \{b - \tilde{p} + 1, \ldots, b\}$;
    let $\bar{p}$ be the integer closest to $p - (a-1)\tilde{p}$ provided that $\bar{p} \geq 0$, $\bar{p} \leq b-1$, and $\bar{p} \leq (a-3)\tilde{p} + 2$ and let $\bar{P} = \{b - \bar{p} + 1, \ldots, b\}$;
    let $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a-1\}$, $\hat{p} = \max(4\lceil p/4 \rceil - q,\ 4\lceil q/4 \rceil - q)$, find $(U, V, \mathscr{G})$ such that the conditions of Theorem 4 are fulfilled; compute $P = (U \cup Q) \setminus V$, R, $W = HR$, $D = W_P$, and $d(D)$; if $d(D) > d_{\text{MAX}}$, update $d_{\text{MAX}} = d(D)$ and $D_{\text{OPT}} = D$;
    if *b is a power of 2 and* $\log_2(b) + 1 \leq \bar{p} < b/4 + 2$ **then**
      let $\bar{P}$ be the union of $L(b)$ in (5) and the $\bar{p} - |L(b)|$ largest elements of $\{1, \ldots, b\} \setminus L(b)$;
      let $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a-1\}$, $\hat{p} = \max(4\lceil p/4 \rceil - q,\ 4\lceil q/4 \rceil - q)$, find $(U, V, \mathscr{G})$ such that conditions of Theorem 4 are fulfilled; compute $P = (U \cup Q) \setminus V$, R, $W = HR$, $D = W_P$, and $d(D)$; if $d(D) > d_{\text{MAX}}$, update $d_{\text{MAX}} = d(D)$ and $D_{\text{OPT}} = D$;
    end
  end
end
let $a = 2$, $b = n/2$, $\tilde{p} = \bar{p} = 2\lceil p/4 \rceil$, $\tilde{P} = \bar{P} = \{b - \bar{p} + 1, \ldots, b\}$, $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a-1\}$, $q = |Q|$, and $\hat{p} = \max(4\lceil p/4 \rceil - q, 4\lceil q/4 \rceil - q)$; obtain A and B, Hadamard matrices of orders a and b, respectively; compute $H = A \otimes B$; find $(U, V, \mathscr{G})$ such that the conditions of Theorem 4 are fulfilled; compute $P = (U \cup Q) \setminus V$, R, $W = HR$, $D = W_P$, and $d(D)$; if $d(D) > d_{\text{MAX}}$, update $d_{\text{MAX}} = d(D)$ and $D_{\text{OPT}} = D$;
if *b is a power of 2 and* $\log_2(b) + 1 \leq \tilde{p} < b/4 + 2$ **then**
  let $\tilde{P} = \bar{P}$ being the union of $L(b)$ in (5) and the $\tilde{p} - |L(b)|$ largest elements of $\{1, \ldots, b\} \setminus L(b)$, $Q = \bar{P} \cup \{x + yb : x \in \tilde{P}, 1 \leq y \leq a-1\}$, $\hat{p} = \max(4\lceil p/4 \rceil - q, 4\lceil q/4 \rceil - q)$, find $(U, V, \mathscr{G})$ such that conditions of Theorem 4 are fulfilled; compute $P = (U \cup Q) \setminus V$, R, $W = HR$, $D = W_P$, and $d(D)$; if $d(D) > d_{\text{MAX}}$, update $d_{\text{MAX}} = d(D)$ and $D_{\text{OPT}} = D$;
end
return D_{OPT}.

4. Numerical comparison

To corroborate the usefulness of our proposed designs, we compare them to four types of space-filling designs below:

HDD Our newly proposed high-dimensional high-separation distance designs.

MmLLT Maximin distance designs proposed by Li et al. (2021).

MmLH Maximin distance Latin hypercube designs generated from the R package "SLHD" (Ba, 2015).

MaxPro Maximum projection Latin hypercube designs generated from the R package "MaxPro" (Ba and Joseph, 2015)

OA Two-level orthogonal arrays. When n is a power of two, we use minimum aberration fractional factorial arrays generated from the R package "FrF2" (Groemping, 2022); otherwise we use randomly selected columns of two-level saturated orthogonal arrays that are generated from Hadamard matrices.

Firstly, Figs. 1 and 2 present separation distance efficiency in (3) of designs with roughly 80 and 256 points, respectively. Observed from the figures, HDD is the best method for $p \geq 30$ and $p \geq 120$ when $n = 80$ and $n = 256$, respectively. Under these cases, HDD designs possess 85% or higher efficiency. In particular, the $n = 80$ two-level HDD designs with p being 39, 40, 41, 57, 59, 60, 61, 71, 72, 73, 75, 76, 77, 78, 79, and the $n = 256$ two-level HDD designs with p being 127, 128, 129, 189, 191, 192, 193, 217, 221, 223, 224, 225, 233, 237, 239, 240, 241, 245, 247, 248, 249, 251, 252, 253, 254, 255 are optimal in separation distance. Also taking into account of the advantages on orthogonality, 2×2 projection uniformity, flexibility on n and p, and applicability to very high n, HDD is arguable the best method for constructing high-dimensional designs. However, HDD designs possess lower efficiency and are inferior to some other designs for lower p, indicating that HDD is less desirable for when p is low. The fact that no design has near-to-one efficiency for small p cues that the bound in (2) may not be tight when p/n is small. Although HDD in $s = 2$ is superior than HDD with higher s in separation distance efficiency, the difference is negligible.

Not producing any design in $n = 80$, we use MmLLT designs of 81 points in the comparison. The performance of MmLLT designs is varying, this is, while some of them are very good, more than half of them are poor. We also find that some of the MmLLT designs have fully confounded column pairs, which are not desirable. While being excellent for cases with low p, MmLH becomes less favorable as p grows, both in the separation distance and the time consumption in generating designs. Owing to the slow computation, MmLH is not suitable to problems with huge n or huge p. MaxPro designs are poor in the efficiency because they optimizes separation distances of projections. OA designs that are generated from randomly selected columns of Hadamard matrices are poor. Finally, two-level minimum aberration fractional factorial arrays are in general excellent in separation distance efficiency. However, minimum aberration fractional factorial arrays can only be constructed when n is a power of s.

Since HDD is excellent in separation distance, we are curious to see if HDD is appealing in emulating high-dimensional computer experiments. In the rest of this section, we compare designs on interpolating seven batches of randomly generated test functions. We generate the first batch of test functions via

$$f(\mathbf{x}) = \sum_{k=1}^{p} \beta_k x_k + Z(\mathbf{x}), \tag{7}$$

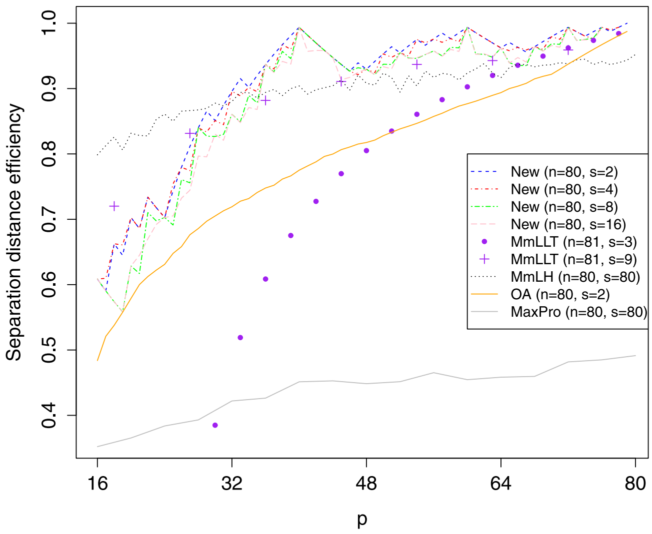

Fig. 1. Separation distance efficiencies for designs with roughly 80 points.

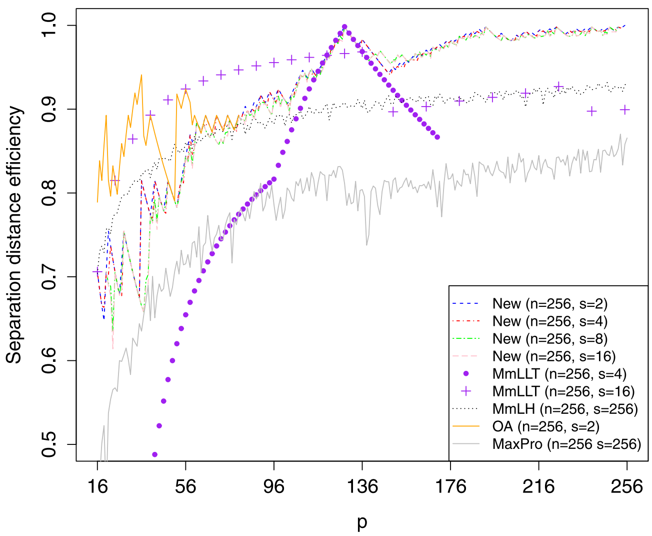

Fig. 2. Separation distance efficiencies for designs with 256 points.

where $f(\mathbf{x})$ gives the output value, the β_k are independently sampled from the uniform distribution on $[-\gamma, \gamma]$, γ is chosen to be 0 or 5, and $Z(\mathbf{x})$ is the realization of a mean zero Gaussian process whose correlation function is Matérn-type with smoothness parameter being $5/2$ and scale parameter being 5. That is to say, when $\gamma = 0$, $f(\mathbf{x})$ is a randomly generated Gaussian process; when $\gamma = 5$, strong linear trends are added to the Gaussian process. We generate the second batch of test functions in a similar way excepts that the Matérn-type correlation function is replaced by the Gaussian correlation function.

The third batch of test functions is given by

$$f(\mathbf{x}) = \exp\left\{-\sum_{i=1}^{p}(x_i - a_i)^2 - \sum_{i=1}^{p}(x_i - b_i)^2\right\}, \tag{8}$$

where the a_i and b_i follow uniform distribution on $[0, 1]$. This function is modified from the function

$$f(\mathbf{x}) = \exp\left\{-50\sum_{i=1}^{p}(x_i - 1/3)^2 - 50\sum_{i=1}^{p}(x_i - 2/3)^2\right\}(100/\pi)^{p/2}/2,$$

which was used in An and Owen (2001) as a test function that cannot be approximated by a low order polynomial to compare emulation methods. Here we let the a_i and b_i to vary so that we can generate 100 distinct functions. We drop the parameter 50 because it does not fit the $p = 40$ or $p = 60$ scenarios. The fourth batch of test functions is the so-called G function,

Table 1
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) when data are generated from the random Gaussian processes by (7) with Matérn correlation function.

n	γ	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	0	SK	0.499	0.481	0.489	0.492	0.410	0.402	0.464	0.445
80	0	UK	0.509	2.4×10^5	0.512	0.496	0.396	0.378	0.408	0.382
80	5	SK	0.497	4.015	0.505	0.469	0.405	0.389	0.403	0.400
80	5	UK	0.509	3.4×10^5	0.511	0.494	0.398	0.380	0.398	0.385
256	0	SK	0.330	0.365	0.369	0.369	0.360	0.382	0.452	0.423
256	0	UK	0.327	0.392	0.400	0.394	0.359	0.378	0.423	0.427
256	5	SK	0.321	0.356	0.372	0.360	0.425	0.370	0.435	0.423
256	5	UK	0.330	0.365	0.369	0.369	0.360	0.382	0.452	0.423

Table 2
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) when data are generated from the random Gaussian processes by (7) with Gaussian correlation function.

n	γ	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	0	SK	0.356	0.409	0.387	0.358	0.290	0.286	0.340	0.327
80	0	UK	0.365	6.7×10^4	0.385	0.375	0.292	0.282	0.290	0.290
80	5	SK	0.349	4.014	0.372	0.351	0.288	0.283	0.296	0.292
80	5	UK	0.365	2.3×10^5	0.381	0.372	0.276	0.285	0.297	0.293
256	0	SK	0.221	0.254	0.254	0.259	0.254	0.264	0.307	0.287
256	0	UK	0.220	0.260	0.278	0.268	0.255	0.256	0.289	0.292
256	5	SK	0.221	0.255	0.254	0.258	0.254	0.264	0.307	0.287
256	5	UK	0.217	0.235	0.257	0.255	0.254	0.258	0.287	0.275

$$f(\mathbf{x}) = \prod_{i=1}^{p} \left\{ 2(4|x_i - a_i| + i/2 - 1)/i \right\}, \tag{9}$$

where the a_i follow uniform distribution on $[0, 1]$. Marrel et al. (2008) have used this function with all a_i being fixed to $1/2$ as a test function to compare emulation methods because it has strongly nonlinear and non-monotonic relationship.

The fifth batch of test functions is the 3-degree polynomials with randomly generated coefficients given by

$$f(\mathbf{x}) = \sum_{i=1}^{p} \beta_i (x_i - c_i) + \sum_{i,j=1}^{p} \beta_{i,j} (x_i - c_i)(x_j - c_j) + \sum_{i,j,k=1}^{p} \beta_{i,j,k} (x_i - c_i)(x_j - c_j)(x_k - c_k), \tag{10}$$

where the β_i, $\beta_{i,j}$, and $\beta_{i,j,k}$ follow the uniform distribution on $[-3, 3]$ and the c_i follow the uniform distribution on $[0, 1]$. The sixth and seventh batch of test functions is the same to the fifth batch of functions excepts that only 50% and 25% selected variables are active, respectively. That is, the summation in (10) is on i from 1 to $p/2$ and from 1 to $p/4$, respectively. Remark that the columns of designs are shuffled before used and thus the variables selected are random.

For each test function, we generate two designs in $p = 40$ for each method, the first with 80 points and the second with 256 points. However, for MmLLT the first design has 81 points. We assume that the input space is $[0, 1]^p$ and the levels of s-level designs are $1/(2s), 3/(2s), \dots, (2s-1)/(2s)$. Their results are similar to those from designs in levels $0, 1/(s-1), \dots, 1$, which we omit. We fit each data set with either the simple Kriging model (SK),

$$f(\mathbf{x}) = \beta_0 + Z(\mathbf{x}), \tag{11}$$

or the universal Kriging model (UK),

$$f(\mathbf{x}) = \beta_0 + \sum_{k=1}^{p} \beta_k x_k + Z(\mathbf{x}), \tag{12}$$

where $Z(\mathbf{x})$ is a mean zero Gaussian process. We use the Matérn-type correlation function with the smoothness parameter being $5/2$ for the first batch of test functions and the Gaussian correlation function for the rest functions. The scale parameters of the Gaussian processes and the coefficients of the linear trends from the UK model are estimated using the R package "RobustGaSP" (Gu et al., 2020) with default settings. For each test function and each design, we estimate the integrated prediction error using at least 1000 testing samples that are uniformly and independently generated from $[0, 1]^p$.

Tables 1–7 provide averaged the integrated squared prediction errors for the seven batches of test functions. From the results, when $n = 80$, the HDD designs are excellent for all the seven batches of test functions excepts that HDD with $s = 2$ is poor for the third batch. OA designs in $n = 80$, which consist of randomly selected columns of 2-level saturated orthogonal arrays, are substantially worse than 2-level HDD designs.

On the other hand, when $n = 256$, the OA designs, which are 2-level minimum aberration fractional factorial arrays, performs robustly better than the 2-level HDD designs. The HDD designs are always among the best designs excepts OA, while HDD4 is the most robust method among the HDD methods. Remark that although the HDD designs are slightly inferior to minimum aberration fractional factorial arrays, the latter designs are not as flexible as HDD on sample size.

Table 3
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) when data are generated by (8).

n	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	SK	1.038	1.818	0.578	0.533	1.073	0.563	0.514	0.520
80	UK	1.065	1.3×10^5	0.629	0.539	1.053	0.562	0.516	0.513
256	SK	0.736	0.431	0.460	0.434	0.759	0.441	0.448	0.455
256	UK	0.736	0.419	0.433	0.421	0.786	0.455	0.465	0.465

Table 4
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) when data are generated from the G function in (9).

n	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	SK	18.87	16.8	15.12	15.89	12.17	13.06	13.53	12.92
80	UK	19.23	4.1×10^5	16.82	16.03	13.73	14.05	14.01	13.74
256	SK	9.57	10.42	11.30	10.17	9.79	10.26	10.65	10.60
256	UK	9.71	10.63	11.11	10.08	10.60	10.78	10.94	11.38

Table 5
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) when data are generated from the 3-degree polynomials in (10) with all variables being active.

n	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	SK	0.518	0.605	0.565	0.505	0.473	0.485	0.468	0.465
80	UK	0.546	2.2×10^6	0.581	0.526	0.462	0.472	0.449	0.457
256	SK	0.448	0.522	0.528	0.498	0.450	0.466	0.518	0.520
256	UK	0.446	0.501	0.505	0.472	0.458	0.479	0.544	0.554

Table 6
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) when data are generated from the 3-degree polynomials in (10) with 20 active variables.

n	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	SK	0.200	0.208	0.207	0.196	0.177	0.179	0.181	0.176
80	UK	0.209	3.6×10^5	0.219	0.193	0.171	0.171	0.175	0.173
256	SK	0.172	0.200	0.205	0.192	0.175	0.180	0.200	0.198
256	UK	0.171	0.193	0.195	0.183	0.177	0.184	0.212	0.211

From comparing the $n = 80$ and $n = 256$ cases, it seems that the HDD designs are more desirable when p/n is relatively high. Here both HDD2 and OA are binary unsaturated orthogonal arrays of strength two. Their only difference lies in the separation distance. The fact that HDD2 is considerably better than OA when $n = 80$ and OA is better than HDD2 when $n = 256$ implies that designs with high separation distance are indeed desirable in emulating high-dimensional computer experiments.

While MmLLT, MaxPro, and MmLH are excellent for some test functions, none of these methods works robustly well in all the seven batches. We are astonished to see that MmLLT performs poorly when $n = 80$ and there are strong linear trends, i.e., when $\gamma = 5$ or UK is used. In fact, the estimation on linear trends are very poor when the data set comes from MmLLT. The poor fits are presumably due to the fact that each column of the MmLLT design in $n = 80$ is fully confounded with another column with correlation coefficient -1. This demonstrates the advantage of using a design with low, or ideally zero, absolute correlations.

While binary designs are poor for the third batch of test functions, for other test functions smaller s in general leads to slightly better performance. Overall, for the test functions we have tried, $s = 4$ seems to be the most robust choice for HDD. Recall that it is commonly believed that space-filling designs with higher number of levels are more desirable because from using them it is possible to capture stronger nonlinearity of the response surface. However, implied from the results this may not be valid for high-dimensional cases, in which it may not be feasible to model strong nonlinearity even if the design has many levels. Having better separation distance efficiency, our constructed binary and four-level designs thus become superior to eight-level and sixteen-level designs. Having said that, two levels may be too few and therefore four-levels designs are better. However, because the gap in performance is not substantial, the best choice of s remains unclear.

5. A real example

The Leslie model (Leslie, 1945) is one kind of widely used deterministic simulation model for estimating the population of species. Its original motivating example is on estimating the age distribution of one class of brown rat called rattus norvegicus. The inputs of the simulation model include the rate of fertility and mortality and initial population that are divided by successive intervals of time. The outputs of the model are the number of survivors of each age group after certain time.

In years past, the Leslie model has been developed into much more complex models. However, here we use the original model with the original example to compare space-filling designs because we believe its input–output relationship is representative of

Table 7
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) when data are generated from the 3-degree polynomials in (10) with 10 active variables.

n	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	SK	0.082	0.067	0.085	0.081	0.071	0.072	0.072	0.073
80	UK	0.086	2.7×10^5	0.089	0.081	0.071	0.070	0.070	0.070
256	SK	0.069	0.080	0.081	0.077	0.071	0.072	0.079	0.079
256	UK	0.069	0.077	0.077	0.074	0.072	0.074	0.084	0.085

Table 8
Averaged integrated prediction error from the SK model in (11) and the UK model in (12) for the real example.

n	Model	OA	MmLLT	MaxPro	MmLH	HDD2	HDD4	HDD8	HDD16
80	SK	63.2	107.8	65.5	61.8	50.2	53.7	54.1	55.7
80	UK	75.7	6.6×10^8	74.4	68.0	67.2	63.5	62.5	62.9
256	SK	33.9	35.1	32.0	30.0	33.9	31.6	32.2	32.5
256	UK	38.5	42.6	41.5	37.9	41.6	34.4	40.3	37.9

many of its extensions. Here we choose the response to be the total number of survivors in millions after 30 months. We consider 40 covariates, $x_1, \ldots, x_{40}$. For $i = 1, \ldots, 20$, $x_i = u_i s_i$ and s_i gives the survival rate from month $i-1$ to month i that are given in the original data. Because the survival rate must not exceed one, we enforce that $x_i \leq 1$. For $i = 21, \ldots, 39$, $x_i = u_i f_{i-19}$ and f_i gives the fertility rate of individuals at month i that are given in the original data. Remark that in the data the rats start to breed from month 2 and thus the fertility rates of earlier months are zero and not considered to be covariates. The $x_{40} = 1000u_{40}$ gives the initial number of rats. All these rats are assumed to be at month 0. Finally, we let the range of u_i to be $[.85, 1.15]$ for each i. In this setting we essentially model the population size when the key parameters of the model deviate from that provided in the original data for at most 15%. Here the high-dimensionality of the input space roots from the fact that the key rates are given in 21 time units.

We fit Gaussian process models using 40-dimensional space-filling designs in the same way as in Section 4. Table 8 provides averaged the integrated squared prediction errors for this example. From the results, HDD4 is certainly the best method when $n = 80$ and one of the best methods when $n = 256$. HDD designs of other numbers of levels are sightly inferior to HDD4. These results verify that HDD is a useful method at least for some real applications.

6. Discussion

In this paper, we propose a novel method to construct high-dimensional high-separation distance designs. Many of our generated designs are remarkably better than existing types of designs in separation distance efficiency. Moreover, from our construction the balanceness, orthogonality, projection uniformity, flexibility in n and p, and near-optimality in separation distance are achieved simultaneously. In our simulations our newly proposed designs perform well in Kriging interpolation. While $s = 4$ appears to be a robust choice, it remains an unsolved problem on what is the optimal s to choose.

In the paper we only consider designs with $p < n$. It is not difficult to construct designs with $p \geq n$ by stacking two or more of our proposed designs together. Since the bound on the separation distance is proportional to p, the combined design will be nearly optimal in the separation distance if all ingredient designs are nearly optimal. However, the combined design will not be orthogonal, as no design with $p > n$ can. Nevertheless, it might be possible to construct high separation distance design in $p > n$ that do not have fully aliased column pairs. It is an interesting problem for future work.

Acknowledgments

The authors thank the editors and two referees for their valuable comments and suggestions. This research is supported by National Key R&D Program of China 2021YFA 1000300, 2021YFA 1000301, National Natural Science Foundation of China (Nos. 12022115, 11971465, 71988101, 1232500250, 12371259), the Fundamental Research Funds for the Central Universities, China (2412023YQ003), China Institute of Marine Technology and Economy (Contact Number 2019A128), and National Center for Mathematics and Interdisciplinary Sciences, CAS.

Appendix A. Supplementary data

Examples for our proposed algorithms, proofs of theorems, and additional theoretical results are provided in the supplementary material.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jspi.2024.106150.

References

An, J., Owen, A., 2001. Quasi-regression. J. Complexity 17 (4), 588–607.

Ba, S., 2015. SLHD: Maximin-distance (sliced) latin hypercube designs. URL http://CRAN.R-project.org/package=SLHD. R package version 2.1-1.

Ba, S., Brenneman, W.A., Myers, W.R., 2015. Optimal sliced latin hypercube designs. Technometrics 57 (4), 479–487.

Ba, S., Joseph, V.R., 2015. MaxPro: Maximum projection designs. URL http://CRAN.R-project.org/package=MaxPro. R package version 3.1-2.

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

Fries, A., Hunter, W.G., 1980. Minimum aberration 2^{k-p} designs. Technometrics 22, 601–608.

Groemping, U., 2022. FrF2: Fractional factorial designs with 2-level factors. URL http://CRAN.R-project.org/package=FrF2. R package version 2.2-3.

Gu, M., Palomo, J., Berger, J., 2020. RobustGaSP: Robust Gaussian stochastic process emulation. URL http://CRAN.R-project.org/package=RobustGaSP. R package version 0.6.1.

Haaland, B., Qian, P.Z.G., 2011. Accurate emulators for large-scale computer experiments. Ann. Statist. 39 (6), 2974–3002.

He, X., 2019. Interleaved lattice-based maximin distance designs. Biometrika 106 (2), 453–464.

He, X., Chien, P., 2018. On the instability issue of gradient-enhanced Gaussian process emulators for computer experiments. SIAM/ASA J. Uncertain. Quantif. 6 (2), 627–644.

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

Johnson, M.E., Moore, L.M., Ylvisaker, D., 1990. Minimax and maximin distance designs.. J. Statist. Plann. Inference 26, 131–148.

Joseph, V.R., Gul, E., Ba, S., 2015. Maximum projection designs for computer experiments. Biometrika 102 (2), 371–380.

Joseph, V.R., Hung, Y., 2008. Orthogonal-maximin latin hypercube designs. Statist. Sinica 18, 171–186.

Leslie, P.H., 1945. On the use of matrices in certain population mathematics. Biometrika 33, 183–212.

Li, W., Liu, M., Tang, B., 2021. A method of constructing maximin distance designs. Biometrika 108 (4), 845–855.

Marrel, A., Iooss, B., Van Dorpe, F., Volkova, E., 2008. An efficient methodology for modeling complex computer codes with Gaussian processes. Comput. Statist. Data Anal. 52 (10), 4731–4744.

Owen, A.B., 1994. Controlling correlations in latin hypercube samples. J. Amer. Statist. Assoc. 89, 1517–1522.

Santner, T.J., Williams, B.J., Notz, W.I., 2018. The Design and Analysis of Computer Experiments, second ed. Springer, New York.

Steinberg, D.M., Lin, D.K., 2006. A construction method for orthogonal latin hypercube designs. Biometrika 93, 279–288.

Sun, F., Tang, B., 2017a. A general rotation method for orthogonal latin hypercubes. Biometrika 104 (2), 465–472.

Sun, F., Tang, B., 2017b. A method of constructing space-filling orthogonal designs. J. Amer. Statist. Assoc. 112 (518), 683–689.

Sun, C.Y., Tang, B., 2023. Uniform projection designs and strong orthogonal arrays. J. Amer. Statist. Assoc. 118 (541), 417–423.

Tang, B., 1993. Orthogonal array-based Latin hypercubes. J. Amer. Statist. Assoc. 88, 1392–1397.

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

Wang, L., Xiao, Q., Xu, H., 2018a. Optimal maximin L_1-distance latin hypercube designs based on good lattice point designs. Ann. Statist. 46 (6B), 3741–3766.

Wang, Y., Yang, J., Xu, H., 2018b. On the connection between maximin distance designs and orthogonal designs. Biometrika 105 (2), 471–477.

Xiao, Q., Xu, H., 2017. Construction of maximin distance Latin squares and related Latin hypercube designs. Biometrika 104 (2), 455–464.

Yang, L., Zhou, Y., Liu, M.Q., 2021. Maximin distance designs based on densest packings. Metrika 84, 615–634.

Yuan, R., Yin, Y., Xu, H., Liu, M.Q., 2025. Construction of maximin distance latin hypercube designs. Statist. Sinica 35 (2).

Zhou, Y., Xu, H., 2014. Space-filling fractional factorial designs. J. Amer. Statist. Assoc. 109 (507), 1134–1144.

Zhou, Y., Xu, H., 2015. Space-filling properties of good lattice point sets. Biometrika 102 (4), 959–966.

Zhou, W., Yang, J., Liu, M., 2020. Optimal maximin L_2-distance latin hypercube designs. J. Statist. Plann. Inference 207, 113–122.