Journal of Statistical Planning and Inference 231 (2024) 106131

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Construction of mixed-level screening designs using Hadamard matrices

Bo Hu [a], Dongying Wang [b], Fasheng Sun [a,*]

[a] *KLAS and School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, China*
[b] *School of Statistics, Jilin University of Finance and Economics, Changchun, Jilin 130117, China*

ARTICLE INFO

MSC:
62K15
62K99

Keywords:
Correlation
Definitive screening design
Factorial design
Fold-over
Optimal design

ABSTRACT

Modern experiments typically involve a very large number of variables. Screening designs allow experimenters to identify active factors in a minimum number of trials. To save costs, only low-level factorial designs are considered for screening experiments, especially two- and three-level designs. In this article, we provide a systematic method to construct screening designs that contain both two- and three-level factors based on Hadamard matrices with the fold-over structure. The proposed designs have good performance in terms of D-optimal and A-optimal criteria, and the estimates of the main effects are unbiased by the second-order effects, making them very suitable for screening experiments. Besides, some theoretical results on D- and A-optimality are obtained as a by-product.

1. Introduction

Screening experiments are frequently used to identify dominant factors out of a large number of potential factors. The active factors will be investigated further in some follow-up experiments, while the inactive factors will be discarded. Screening designs that allow the experimenters to identify the active factors in a minimum number of trials (i.e., with minimum budget and resources) are valuable in practical applications. A good screening design should be able to accommodate a large number of factors with relatively few runs, and orthogonality or lower correlations among factorial effects are desirable properties to identify active factors effectively. Besides, the biases of estimates of main effects caused by the presence of active two-factor interactions are as small as possible. Fold-over technique is common used to deal with the issue.

Generally, to save on costs, only low-level factorial designs, especially two- and three-level factorial designs, are considered for screening experiments. Additionally, the run size is usually not large enough for estimating all the main effects of the factors. Such designs are called supersaturated designs. The analysis relies on the assumption of effect sparsity (Box and Meyer, 1986) that only a few factors are active. Some practical applications and constructions on supersaturated designs can be found in Lin (1993), Wu (1993), Nguyen (1996), Tang and Wu (1997), Ai et al. (2007) and Sun et al. (2011). Georgiou (2014) is a good reference for supersaturated designs.

Supersaturated designs commonly assume the absence of all interaction effects. However, if some of the active factors can potentially interact with each other, then supersaturated designs cannot detect the confounding influences between main effects and second-order effects. To address this problem, Shi and Tang (2019) proposed a new class of fold-over supersaturated designs. Such designs are robust to two-factor interactions but limit each factor to only two levels. To further estimate the pure-quadratic effects, Jones and Nachtsheim (2011) provided a useful class of economic three-level designs called definitive screening designs

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

https://doi.org/10.1016/j.jspi.2023.106131
Received 26 January 2022; Received in revised form 18 November 2023; Accepted 20 November 2023
Available online 23 November 2023

(DSDs), which require $2m+1$ runs consisting of m fold-over pairs and an overall center point for m factors. Numerical constructions on DSDs are studied in Jones and Nachtsheim (2011) and Nguyen and Stylianou (2013). To avoid infeasible computational time, Xiao et al. (2012) provided a systematic construction of DSDs using conference matrices. Moreover, Phoa and Lin (2015) proposed a theoretically driven approach to construct DSDs.

However, much practical evidence indicates that some categorical factors are able to remain at two levels, while the other factors must be at three levels, especially quantitative factors. To construct screening designs containing both two- and three-level factors, referred to as mixed-level screening designs, Jones and Nachtsheim (2013) presented a DSD-augment method (hereafter abbreviated as JNs). This method uses a search algorithm to convert some three-level columns of a conference matrix to two-level columns. To do so, it needs to evaluate the determinants of 2^{2m_2} information matrices, where m_2 is the number of two-level factors. Hence, if m_2 is very large, this procedure will lead to a large amount of calculation. Another limitation of JNs is that as the number of categorical factors increases, the correlations between the pure-quadratic effect columns increase and approach $1/2$. In this paper, to save computational work and reduce the correlations between the pure-quadratic effect columns, we provide a systematic method for constructing mixed-level screening designs.

Throughout, we assume that the response follows the second-order model, which can be written as:

$$y = \beta_0 + \sum_{i=1}^{m_2+m_3} \beta_i x_i + \sum_{i=1}^{m_2+m_3-1} \sum_{j=i+1}^{m_2+m_3} \beta_{ij} x_i x_j + \sum_{i=1}^{m_3} \beta_{ii} x_i^2 + \epsilon, \tag{1}$$

where y is the response variable, $x_1, x_2, \ldots, x_{m_3}$ are three-level factors, $x_{m_3+1}, x_{m_3+2}, \ldots, x_{m_3+m_2}$ are two-level factors, and $x_i x_j$ and x_i^2 are the interactions of factors and the pure-quadratic effects of three-level factors, respectively. β_i, β_{ij} and β_{ii} denote the unknown constant coefficients, and ϵ is the random error with zero mean and a finite variance σ^2. Because our goal is screening, we assume that the experimenters will fit the first-order model to the response at the beginning, which can be written as:

$$y = \beta_0 + \sum_{i=1}^{m_2+m_3} \beta_i x_i + \epsilon. \tag{2}$$

In this paper, the primary goal is to identify the active factors using the first-order model (2), but an important secondary goal is to use the second-order model (1) to capture the essential features of the relationship between these active factors and the response. Therefore, it is desirable that the main effects are all orthogonal to potential two-factor interactions. It turns out that the fold-over structure is well suited to achieve this. Accordingly, we present a construction method for mixed-level screening designs with the fold-over structure. Although the method is simple, the resulting designs enjoy some attractive properties, making them very suitable for screening experiments.

An outline of the remainder of the paper is as follows. Section 2 provides the construction method for mixed-level screening designs utilizing Hadamard matrices and the fold-over structure. Section 3 demonstrates that the proposed designs possess good properties in terms of D-optimal and A-optimal criteria. The correlations between columns are also presented in this section. Section 4 provides the discussion and conclusions.

2. Construction of mixed-level screening designs based on Hadamard matrices

A construction method for mixed-level screening designs with m_2 two-level factors and m_3 three-level factors based on Hadamard matrices and the fold-over structure is introduced in this section. First, we now introduce the definition of Hadamard matrix.

Definition 2.1 (*Hedayat et al., 1999*)**.** An $m \times m$ matrix $H = (h_{ij})$ is called a Hadamard matrix if it satisfies $H'H = mI_m$, with $h_{ij} = \pm 1$ $(i, j = 1, 2, \ldots, m)$, where I_m is the m-order identity matrix.

Hadamard matrices are widely used in the construction of two-level orthogonal designs. A Hadamard matrix is said to be normalized if its first column only contains entry $+1$; any Hadamard matrix can be normalized by sign-switching those rows that have -1 in the first column. Deleting the first column from a normalized Hadamard matrix of order m, we obtain a saturated orthogonal array of strength 2 with m runs and $m-1$ factors. When there are a large scale of two-level factors, the mixed-level screening designs based on Hadamard matrices tend to have high D- and A-efficiencies.

Next, we provide a new construction for mixed-level screening design, denoted as HMD. The construction method employs the following steps:

Step 1. For any $m \times m$ Hadamard matrix H, randomly select $m_2 + m_3$ columns ($m_2 + m_3 \le m$), and denote the resulting matrix as $\widetilde{H}$;

Step 2. Set the first m_3 diagonal elements of $\widetilde{H}$ to 0, namely, $\widetilde{h}(i,i) = 0$, $i = 1, 2, \ldots, m_3$, and denote the new matrix as H^*;

Step 3. Fold over H^* and obtain the following design:

$$D = \begin{pmatrix} H^* \\ -H^* \end{pmatrix}. \tag{3}$$

The above steps provide a convenient procedure for constructing an HMD with $2m$ runs involving m_2 two-level factors and m_3 three-level factors, where $m_2 + m_3 \le m$. We shall illustrate the above procedure with the help of the following example.

Example 2.2. We will construct an HMD with 6 two-level columns and 2 three-level columns. First, we generate a Hadamard matrix H of order 8 by the $Hadamard$ function in R software. Permutate the columns of H randomly, and $\widetilde{H}$ is one member of the matrices after permutation. Then, we set the first 2 diagonal elements of $\widetilde{H}$ to 0; we obtain half fraction of an HMD H^*. Folding over H^* leads to an HMD with 16 runs, 6 two-level columns and 2 three-level columns. H, $\widetilde{H}$ and H^* are given in (4).

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \end{pmatrix}; \widetilde{H} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 \end{pmatrix};$$

$$H^* = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 \end{pmatrix}. \tag{4}$$

3. Design properties

In this section, we discuss the properties of the proposed designs and use the D-optimal criterion, the A-optimal criterion and the column correlations to evaluate the performance of the proposed HMDs.

3.1. The D-optimality

The D-optimal criterion provides a scalar measure of the generalized variance of the least squares estimate of the coefficients which minimizes the volume of the confidence region. In this paper, we use D-efficiency to compare designs with different numbers of design points or runs (Draper and Lin, 1990), that is,

$$D_{\text{eff}} = \frac{|X'X|^{1/p}}{n}, \tag{5}$$

where X is the model matrix of design D, X' is the transpose of X, n is the run size, and p is the number of parameters in the model.

We can now obtain the lower bound of the first-order D-efficiency under the first-order model (2) for the proposed HMD.

Theorem 3.1. *Suppose an HMD given by (3) with $2m$ runs involving m_2 two-level factors and m_3 three-level factors, where $m_2 + m_3 \le m$. If $m_3 \le m/3$, the first-order D-efficiency of the HMD has a lower bound given by:*

$$D_{\text{eff}}(D_{HMD}) \ge LB_{\text{D}_{\text{eff}}(HMD)} = \left(1 - \frac{a}{b}\right)^c, \tag{6}$$

where $a = 4m^3 + m^2\left(16m_3^2 - 24m_3 + 4m_2 - 3\right) + m\left(24m_3 - 12m_2 - 16m_3^2\right) + 24m_2m_3 - 16m_2m_3^2$, $b = m^2(2m-3)^2$ *and* $c = m_3/(m_2 + m_3 + 1)$.

The proof of Theorem 3.1 and later results are deferred to Appendix. Examining (6), $a/b \to 0$ and $D_{\text{eff}}(D_{HMD}) \to 1$, if $m \to \infty$ and $m_3/m \to 0$. Hence, HMDs achieve high estimation efficiency for cases with a small scale of three-level factors. More specifically, Table 1 lists the lower bounds of the first-order D-efficiencies of HMDs for different values of m_2 and m_3. This further illustrates that HMDs have high first-order D-efficiencies, especially when m_3/m is small. The lower bounds of the first-order D-efficiencies listed in Table 1 may not be tight, and the actual first-order D-efficiencies can be much higher than the bounds. Please see Table 1 for details.

Next, we demonstrate a higher first-order D-efficiency characteristic of the HMD compared with the JN through a specific example.

Example 2.2 (*Continued*)**.** For the case of $m_2 = 6, m_3 = 2$, we compare the first-order D-efficiencies of HMD and JN. Let $D_{HMD} = (H^{*\prime}, -H^{*\prime})'$ be the HMD with 16 runs, where H^* is defined in Example 2.2, and let $D_{JN} = (\overline{H}', -\overline{H}')'$ be the JN with 18

Table 1
Comparison of the first-order D-efficiencies between HMDs and JNs.

m	m_3	m_2	HMDs					JNs	
			Runs	$LB_{D_{\rm eff}}$	max $D_{\rm eff}$	min $D_{\rm eff}$	Average $D_{\rm eff}$	Runs	$D_{\rm eff}$
8	1	7	16	0.9701	0.9708	0.9708	0.9708	18	0.9443
8	2	6	16	0.9135	0.9466	0.9381	0.9430	18	0.9123
8	3	5	16	–	0.9270	0.9033	0.9165	18	0.8852
8	4	4	16	–	0.9118	0.8674	0.8911	18	0.8623
12	1	11	24	0.9865	0.9867	0.9867	0.9867	26	0.9667
12	2	10	24	0.9661	0.9748	0.9723	0.9737	26	0.9513
12	3	9	24	0.9202	0.9643	0.9567	0.9609	26	0.9373
12	4	8	24	0.8231	0.9550	0.9405	0.9485	26	0.9246
12	5	7	24	–	0.9449	0.9262	0.9361	26	0.9131
12	6	6	24	–	0.9379	0.9121	0.9247	26	0.9026
16	1	15	32	0.9924	0.9924	0.9924	0.9924	34	–
16	2	14	32	0.9820	0.9854	0.9844	0.9850	34	–
16	3	13	32	0.9617	0.9790	0.9759	0.9776	34	–
16	4	12	32	0.9231	0.9731	0.9667	0.9703	34	0.9512
16	5	11	32	0.8550	0.9677	0.9579	0.9633	34	0.9438
16	6	10	32	–	0.9623	0.9504	0.9562	34	0.9368
16	7	9	32	–	0.9585	0.9412	0.9490	34	0.9303
16	8	8	32	–	0.9508	0.9346	0.9423	34	0.9168

Note: JNs with $m_2 > 12$ are not listed in this table because these designs are computational expensive. And – indicates $LB_{D_{\rm eff}}$(HMD) does not exist.

runs, where $\overline{H}$ generated by the construction method in Jones and Nachtsheim (2013) is given by:

$$\overline{H} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 \\ 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}.$$

By calculation, D_{HMD} has a higher first-order D-efficiency of 0.9466 than that of D_{JN} whose the first-order D-efficiency is 0.9123.

Actually, the D-efficiency values resulting from randomly selecting m_2+m_3 columns in $Step1$ are quite close to each other as can be seen in Table 1. This justifies the simple and convenient technique of randomly selecting m_2+m_3 columns in $Step1$. All Hadamard matrices are generated by the $Hadamard$ function in R software, and all JNs in this paper are obtained by the exhaustive search methods proposed by Jones and Nachtsheim (2013). This table also illustrates that HMDs have higher first-order D-efficiencies than JNs and need fewer runs.

3.2. The A-optimality

The A-optimal criterion is also related to the shape of the confidence ellipsoid, which minimizes the average variance of the parameters. Here, we use A-efficiency to discuss the performance of HMDs in terms of the A-optimal criterion. A-efficiency can be stated formally as follows:

$$A_{\rm eff} = \frac{LB}{tr(X'X)^{-1}}, \tag{7}$$

where $tr(\cdot)$ is the trace of the corresponding matrix, $(\cdot)^{-1}$ is the inverse of the corresponding matrix, X is the model matrix of design D. Here $LB = p/n$ because if a matrix X has only 0 s and ± 1 s, we have $tr(X'X)^{-1} \geq LB = p/n$.

Next, we calculate the first-order A-efficiency for the proposed HMD under the first-order model (2).

Theorem 3.2. *Suppose an HMD given by (3) with $2m$ runs involving m_2 two-level factors and m_3 three-level factors, where $m_2+m_3 \leq m$. If $m_3 \leq m/3$, the first-order A-efficiency of the HMD has a lower bound as follows:*

$$A_{\rm eff}(D_{HMD}) \geq LB_{A_{\rm eff}(HMD)} = 1 - \frac{(r-1)m_3}{1+m_2+rm_3}, \tag{8}$$

where $r = \dfrac{(m^2+m_2m_3)(m-3/2)^2}{(m-2m_3)(m+2m_3-3)(m^2-m_2-m)}$.

Table 2
Comparison of the first-order A-efficiencies between HMDs and JNs.

m	m_3	m_2	HMDs					JNs	
			Runs	$LB_{A_{\rm eff}}$	max $A_{\rm eff}$	min $A_{\rm eff}$	Average $A_{\rm eff}$	Runs	$A_{\rm eff}$
8	1	7	16	0.9516	0.9525	0.9525	0.9525	18	0.9211
8	2	6	16	0.8516	0.9184	0.8988	0.9100	18	0.8777
12	1	11	24	0.9786	0.9788	0.9788	0.9788	26	0.9517
12	2	10	24	0.9386	0.9612	0.9556	0.9586	26	0.9308
12	3	9	24	0.8604	0.9467	0.9298	0.9393	26	0.9133
12	4	8	24	0.7141	0.9350	0.9040	0.9211	26	0.8990
16	1	15	32	0.9880	0.9881	0.9881	0.9881	34	–
16	2	14	32	0.9664	0.9776	0.9753	0.9765	34	–
16	3	13	32	0.9282	0.9684	0.9615	0.9653	34	–
16	4	12	32	0.8641	0.9604	0.9463	0.9543	34	0.9329
16	5	11	32	0.7608	0.9534	0.9330	0.9442	34	0.9246

Note: JNs with $m_2 > 12$ are not listed in this table because these designs are computational expensive.

Examining (8), $r \to 1$ and $A_{\rm eff}(D_{HMD}) \to 1$, if $m \to \infty$ and $m_3/m \to 0$; and $A_{\rm eff}(D_{HMD})$ performs well for small m_3/m. For illustrative purposes, Table 2 lists the lower bounds of the first-order A-efficiencies of HMDs for different values of m_2 and m_3. From this table, we know that HMDs have large lower bounds for the first-order A-efficiencies, especially when m_3/m is small. The actual first-order A-efficiencies will be much higher than the lower bounds of the ones. Please see Table 2 for details.

Table 2 exhibits the first-order A-efficiencies for the cases with $m_2 + m_3 = m$, $m_3 \le m/3$ and $m = 8, 12, 16$ for HMDs and JNs. From this table, the maximum, minimum, and average of $A_{\rm eff}$ for HMDs are larger than those for JNs. This implies that HMDs can estimate the main effects more properly than JNs.

3.3. Correlations

The main objectives of screening designs are to select active factors and estimate the effects accurately. Therefore, lower correlations among effects are preferred. Thus, we calculate the correlations among the design columns in the rest of this section. Denote ρ as the correlation between two columns, namely, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ and $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)'$. Then,

$$\rho = \frac{\sum_{i=1}^{n}(\mu_i - \overline{\mu})(\nu_i - \overline{\nu})}{\sqrt{\sum_{i=1}^{n}(\mu_i - \overline{\mu})^2}\sqrt{\sum_{i=1}^{n}(\nu_i - \overline{\nu})^2}},$$

where $\overline{\mu} = \sum_{i=1}^{n} \mu_i/n$, $\overline{\nu} = \sum_{i=1}^{n} \nu_i/n$. In what follows, we call $|\rho|$ the absolute correlation.

According to the construction method of HMDs, calculations of the correlations between design columns are straightforward as shown by the following proposition.

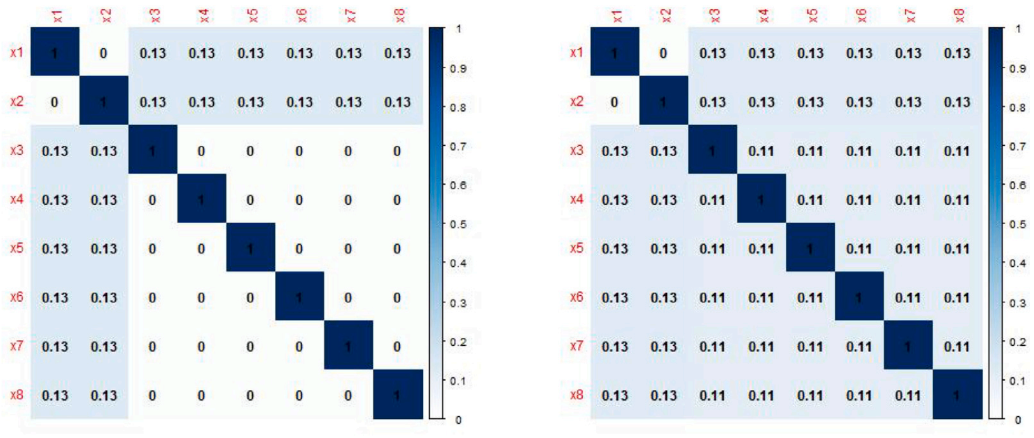
Proposition 3.3. *For an HMD with $n = 2m$ runs involving m_2 two-level factors and m_3 three-level factors ($m_2 + m_3 \le m$), the correlations between design columns are as follows:*
(i) the correlation between two two-level design columns is 0;
(ii) the correlation between two three-level design columns is 0 or $\frac{\pm 2}{m-1}$;
(iii) the correlation between a three-level design column and a two-level design column is $\frac{\pm 1}{\sqrt{m(m-1)}}$.

Proposition 3.3 demonstrates that the correlations between design columns in HMDs are either 0 or decrease to 0 as $m \to +\infty$. This guarantees that the HMDs can have high accuracy when screening active factors for large m.

Compared to JNs, HMDs have orthogonality of two-level columns and low correlations between three-level columns. Therefore, we tend to favor HMDs when two-level columns have a high percentage. We now illustrate the above idea with the help of the specific example.

Example 2.2 (*Continued*). We use an absolute correlation cell plot to demonstrate the correlations between design columns for D_{HMD}, which are also compared with D_{JN}, as shown in Fig. 1. It is obvious that the correlation between any two two-level design columns is 0 for HMD. On the whole, the correlations of HMD are smaller than those of JN.

As a result of the fold-over structure of HMDs, estimates of main effects are robust not only to the presence of two-factor interactions but also to the presence of the pure-quadratic effects. However, when using HMDs to explore the response surface model, the confounding between the potential two-factor interactions and the pure-quadratic effects cannot be ignored. In the second-order model (1), we hope to obtain a global assessment of curvature, that is, to estimate the pure-quadratic effects accurately. As mentioned, a lower correlation permits nearly independent estimates of effects. Thus, we further examine the correlations between the pure-quadratic effect columns and the other second-order effect columns for HMDs. These results are provided in the following proposition.



(a) D_{HMD} (b) D_{JN}

Fig. 1. Absolute correlation cell plots of design columns for D_{HMD} and D_{JN} with 2 three-level factors denoted as x_1, x_2 and 6 two-level factors denoted as $x_3, \dots, x_8$.

Proposition 3.4. *For an HMD with $n = 2m$ runs involving m_2 two-level factors and m_3 three-level factors ($m_2 + m_3 \leq m$), the correlations between the pure-quadratic effect columns and the other second-order effect columns are as follows:*

(i) the correlation between two pure-quadratic effect columns is $\frac{-1}{m-1}$;

(ii) the correlation between a pure-quadratic effect column and a three-level factor interaction effect column involving one common factor is 0 or $\frac{\pm 2}{\sqrt{(m-1)(m^2-2m-4)}}$;

(iii) the correlation between a pure-quadratic effect column and a three-level factor interaction column that involves three different factors is $\frac{\pm(m-2)}{\sqrt{(m-1)(m^2-2m-4)}}$, $\frac{\pm(m+2)}{\sqrt{(m-1)(m^2-2m-4)}}$ or $\frac{\pm\sqrt{m}}{\sqrt{(m-1)(m-2)}}$;

(iv) the correlation between a pure-quadratic effect column and a two-level factor interaction column is $\frac{\pm 1}{\sqrt{m-1}}$;

(v) the correlation between a pure-quadratic effect column and a two-factor interaction column of a two-level design column and a three-level design column involving one common three-level factor is $\frac{\pm 1}{\sqrt{(m-1)(m^2-m-1)}}$;

(vi) the correlation between a pure-quadratic effect column and a two-factor interaction column of a two-level design column and a three-level design column involving two different three-level factors is $\frac{\pm(m-1)}{\sqrt{(m-1)(m^2-m-1)}}$ or $\frac{\pm(m+1)}{\sqrt{(m-1)(m^2-m-1)}}$.

For JNs, the corresponding types of correlations also can be obtained, and the results are presented in the next remark. The proof is similar to that of Proposition 3.4, thus is omitted here.

Remark 3.5. For a JN with $n = 2m+2$ runs involving m_2 two-level factors and m_3 three-level factors ($m_2 + m_3 \leq m$), the correlations between the pure-quadratic effect columns and the other second-order effect columns are as follows:

(i) the correlation between two pure-quadratic effect columns is $\frac{1}{2} - \frac{1}{m-1}$;

(ii) the correlation between a pure-quadratic effect column and a three-level factor interaction effect column involving one common factor is 0;

(iii) the correlation between a pure-quadratic effect column and a three-level factor interaction column that involves three different factors is $\frac{\pm\sqrt{m+1}}{\sqrt{2(m-1)(m-2)}}$;

(iv) the correlation between a pure-quadratic effect column and a two-level factor interaction column is $\frac{\pm 2}{\sqrt{2(m-1)(m^2+2\ m)}}$, $\frac{\pm 2m}{\sqrt{2(m-1)(m^2+2\ m)}}$, $\frac{\pm(2m+4)}{\sqrt{2(m-1)(m^2+2\ m)}}$, $\frac{\pm(4m+2)}{\sqrt{2(m-1)(m^2+2\ m)}}$, $\frac{\pm(2m-4)}{\sqrt{2(m-1)(m^2+2m-8)}}$, $\frac{\pm(4m-2)}{\sqrt{2(m-1)(m^2+2m-8)}}$, $\frac{\pm 6}{\sqrt{2(m-1)(m^2+2m-8)}}$ or $\frac{\pm 6m}{\sqrt{2(m-1)(m^2+2m-8)}}$;

(v) the correlation between a pure-quadratic effect column and a two-factor interaction column of a two-level design column and a three-level design column involving one common three-level factor is $\frac{\pm 2}{\sqrt{2(m-1)(m^2-2)}}$;

(vi) the correlation between a pure-quadratic effect column and a two-factor interaction column of a two-level design column and a three-level design column involving two different three-level factors is $\frac{\pm(m-1)}{\sqrt{2(m-1)(m^2-2)}}$ or $\frac{\pm(m+3)}{\sqrt{2(m-1)(m^2-2)}}$.

These results show that for HMDs and JNs, most of the above types of correlations approach to 0 as $m \to \infty$, except that the correlation between two pure-quadratic effect columns of JNs is $\frac{1}{2} - \frac{1}{m-1}$ tending to $\frac{1}{2}$ as ($m \to \infty$), while that of HMDs is $\frac{-1}{m-1} \to 0$ as ($m \to \infty$). This also means that HMDs can estimate the pure-quadratic effects more accurately than JNs.

Table 3
The first-order D-efficiencies of HMDs for non-isomorphic Hadamard matrices.

m	m_3	m_2	HMDs$^{(1)}$			HMDs$^{(2)}$		
			max D_{eff}	min D_{eff}	Average D_{eff}	max D_{eff}	min D_{eff}	Average D_{eff}
20	1	18	0.9950	0.9950	0.9950	0.9950	0.9950	0.9950
20	2	17	0.9903	0.9898	0.9903	0.9898	0.9898	0.9898
20	3	16	0.9859	0.9843	0.9858	0.9843	0.9843	0.9843
20	4	15	0.9819	0.9787	0.9814	0.9804	0.9784	0.9788
20	5	14	0.9780	0.9727	0.9772	0.9765	0.9736	0.9749
20	6	13	0.9745	0.9666	0.9730	0.9719	0.9680	0.9705
24	1	22	0.9965	0.9965	0.9965	0.9965	0.9965	0.9965
24	2	21	0.9929	0.9929	0.9929	0.9929	0.9929	0.9929
24	3	20	0.9892	0.9892	0.9892	0.9892	0.9892	0.9892
24	4	19	0.9860	0.9856	0.9859	0.9863	0.9852	0.9854
24	5	18	0.9832	0.9826	0.9831	0.9833	0.9821	0.9823
24	6	17	0.9804	0.9796	0.9802	0.9800	0.9789	0.9791
24	7	16	0.9764	0.9759	0.9763	0.9767	0.9755	0.9757

Note: HMDs$^{(1)}$ and HMDs$^{(2)}$ with $m = 20$ are constructed by had.20.pal and had.20.toncheviv, respectively. HMDs$^{(1)}$ and HMDs$^{(2)}$ with $m = 24$ are constructed by had.24.2 and had.24.23, respectively.

4. Discussion and conclusions

In this paper, we construct mixed-level screening designs (HMDs) utilizing Hadamard matrices and the fold-over structure. HMDs usually require fewer runs than the corresponding JNs and have high D-efficiencies per run for the first-order model. Moreover, HMDs give away the orthogonality between three-level design columns but achieve higher first-order D-efficiencies and first-order A-efficiencies. Therefore, if the experimenters are more concerned about the presence of two-level factors or a low percentage of three-level factors in the experiments, then we suggest using HMDs.

It must be said that the proposed method performs well when there are a low proportion of three-level columns (as stated in Theorems 3.1 and 3.2). When there are a substantial proportion of three-level columns, the JNs have higher efficiencies.

Nonisomorphic Hadamard matrices (Shi and Tang, 2018) generate HMDs with different first-order D-efficiencies. As shown in Table 3, we select two pairs of nonisomorphic Hadamard matrices given in http://neilsloane.com/hadamard/, had.20.pal and had.20.toncheviv, had.24.2 and had.24.23, to construct HMDs. Although distinct Hadamard matrices can result in HMDs with different first-order D-efficiencies, the differences are very small. Thus, in practical applications, we suggest taking any Hadamard matrix to save computational time. Of course, the issues of how to optimally choose the initial Hadamard matrix and how to arrange the columns are worth studying.

Acknowledgments

The authors thank the executive editors and two reviewers for their valuable comments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 11971098, 12371259 and 12101258), the Fundamental Research Funds for the Central Universities, China (2412023YQ003) and the National Key Research and Development Program of China (No. 2020YFA0714102). The first two authors contributed equally to this work.

Appendix

In this appendix, we provide justifications for Theorems 3.1, 3.2, Propositions 3.3 and 3.4.

To prove the theorems, we need the following lemmas:

Lemma 1. *(i) Let A be an $n\times n$ real symmetric matrix, A_m be any principal submatrix of order m of A. There are $\binom{n}{m}$ principal submatrices, then*

$$\lambda_{\min}(A) \le \lambda(A_m) \le \lambda_{\max}(A),$$

where $\lambda(\cdot)$ is an eigenvalue of the corresponding matrix.

(ii) Let A be an $n \times n$ real symmetric matrix, then

$$\min_{x\neq 0} \frac{x'Ax}{x'x} \le \lambda(A) \le \max_{x\neq 0} \frac{x'Ax}{x'x},$$

where x is an n-dimensional column vector.

(iii) Let A and B be $n \times n$ real symmetric matrices, then

$$\lambda_{\min}(B) \le \lambda_i(A + B) - \lambda_i(A) \le \lambda_{\max}(B),\ i = 1, 2, \dots, n,$$

where $\lambda_i(\cdot)$ is the ith largest eigenvalue of the corresponding matrix.

(iv) Let A and B be $n \times n$ real symmetric matrices. If A and B are positive semidefinite matrices, then

$$0 \le tr(AB) \le tr(A) \cdot tr(B).$$

Lemma 2. *Kantorovich-type inequality (Marshall et al., 2009, P102). If $0 < m \le a_i \le M$, $i = 1, 2, \ldots, n$, then*

$$\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_i}\right) \le \frac{(M+m)^2}{4mM}.$$

Appendix A. The proof of Theorem 3.1

The design matrix is D, which has $2m$ rows and $m_2 + m_3$ columns ($m_2 + m_3 \le m$). Denote X_H as the model matrix for the model (2). For the HMD, without loss of generality, X_H can be expressed as

$$X_H = \begin{pmatrix} \mathbf{1} & X_3 & X_2 \\ \mathbf{1} & -X_3 & -X_2 \end{pmatrix},$$

where the first column corresponds to the intercept term, the columns 2 to $m_3 + 1$, X_3, correspond to the three-level factors, and the last m_2 columns, X_2, correspond to the two-level factors. Then, by the construction of the HMD, we have

$$|X_H' X_H| = 2^{m_2+m_3+1} m^{m_2+1} \left| X_3'X_3 - \frac{1}{m} X_3'X_2X_2'X_3 \right|, \tag{9}$$

where $X_2'X_2 = mI_{m_2}$, the diagonal elements of $X_3'X_3$ are $(m-1)$ s, the off-diagonal elements of $X_3'X_3$ are 0 s or ± 2 s, and all the elements of $X_3'X_2$ are ± 1 s.

Let $T = X_3'X_3 - \frac{1}{m}X_3'X_2X_2'X_3$ and $X_3 = Y_1 + Y_2$, where Y_1 is an $m \times m_3$ matrix consisting of the first m_3 columns of $\widetilde{H}$ in *Step 1*. Since $X_2'X_2 = mI_{m_2}$, we have $\frac{1}{m}X_3'X_2X_2'X_3 = Y_2'X_2(X_2'X_2)^{-1}X_2'Y_2$. Since $0 \le \lambda(Y_2'X_2(X_2'X_2)^{-1}X_2'Y_2) \le 1$ by Lemma 1(i) and $m - 2m_3 + 1 \le \lambda(X_3'X_3) \le m + 2m_3 - 3$ by Lemma 1(ii). Therefore, from Lemma 1(iii), we obtain $m - 2m_3 \le \lambda(T) \le m + 2m_3 - 3$. Thus, combining Lemma 2 and $tr(T) = (m - m_2/m - 1)m_3$, we have

$$|T| \ge \left(\frac{4\left(m - 2m_3\right)\left(m + 2m_3 - 3\right)\left(m^2 - m_2 - m\right)}{m\,(2m-3)^2}\right)^{m_3}, \quad \text{for } m_3 \le m/3. \tag{10}$$

Combining (5), (9) and (10), we can obtain (6). Hence, Theorem 3.1 is proved.

Appendix B. The proof of Theorem 3.2

Continuing with the notation in the proof of Theorem 3.1, we have

$$X_H'X_H = \begin{pmatrix} 2m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2X_3'X_3 & 2X_3'X_2 \\ \mathbf{0} & 2X_2'X_3 & 2X_2'X_2 \end{pmatrix} = 2\begin{pmatrix} m & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix},$$

where $A = \begin{pmatrix} X_3'X_3 & X_3'X_2 \\ X_2'X_3 & X_2'X_2 \end{pmatrix}$. Thus, the $tr(X_H'X_H)^{-1}$ of an HMD can be expressed as

$$tr(X_H'X_H)^{-1} = 1/(2m) + 1/2 \cdot tr(A^{-1}), \tag{11}$$

where

$$A^{-1} = \begin{pmatrix} \left(X_3'X_3 - \frac{1}{m}X_3'X_2X_2'X_3\right)^{-1} & * \\ * & \left(X_2'X_2\right)^{-1} + \left(X_2'X_2\right)^{-1} X_2'X_3 \left(X_3'X_3 - \frac{1}{m}X_3'X_2X_2'X_3\right)^{-1} X_3'X_2 \left(X_2'X_2\right)^{-1} \end{pmatrix}.$$

From Lemma 1(iv), we obtain

$$\begin{aligned} tr(A^{-1}) &= m_2/m + tr(T^{-1}) + 1/m^2 \cdot tr(X_2'X_3T^{-1}X_3'X_2) \\ &\le m_2/m + (1 + m_2m_3/m^2)\textstyle\sum_{i=1}^{m_3} 1/\lambda_i(T), \ \text{for } m_3 \le m/3, \end{aligned} \tag{12}$$

where $T = X_3'X_3 - \frac{1}{m}X_3'X_2X_2'X_3$. Employing Lemma 2 and $tr(T) = (m - m_2/m - 1)m_3$, we have

$$\sum_{i=1}^{m_3} 1/\lambda_i(T) \le \frac{mm_3(2m-3)^2}{4(m-2m_3)(m+2m_3-3)(m^2-m_2-m)}, \ \text{for } m_3 \le m/3. \tag{13}$$

Combining (7), (11), (12) and (13), we can obtain (8).

Hence, Theorem 3.2 is proved.

Appendix C. The proof of Proposition 3.3

Let $\boldsymbol{x}_{f_1},\boldsymbol{x}_{f_2}$ denote any two distinct three-level design columns in the design matrix of the HMD and $\boldsymbol{x}_{g_1},\boldsymbol{x}_{g_2}$ denote any two distinct two-level design columns in the design matrix of the HMD. Let $\overline{x}_*$ denote the average of corresponding column $\boldsymbol{x}_*$, where $*\in\{f_1,f_2,g_1,g_2\}$. By the construction of the HMDs, we have $\overline{x}_*=0$. Then, the correlation between any two design columns $(x_*,x_\diamond)$ is

$$\rho_{*,\diamond}=\frac{\sum_{i=1}^{2m}x_{i,*}x_{i,\diamond}}{\sqrt{\sum_{i=1}^{2m}x_{i,*}^2}\sqrt{\sum_{i=1}^{2m}x_{i,\diamond}^2}},$$

where $*,\diamond\in\{f_1,f_2,g_1,g_2\}$. By the construction of the HMD, we have $\sum_{i=1}^{2m}x_{i,g_1}x_{i,g_2}=0$, $\sum_{i=1}^{2m}x_{i,f_1}^2=\sum_{i=1}^{2m}x_{i,f_2}^2=2(m-1)$, $\sum_{i=1}^{2m}x_{i,g_1}^2=2m$, $\sum_{i=1}^{2m}x_{i,f_1}x_{i,f_2}=0$ or ±4 and $\sum_{i=1}^{2m}x_{i,f_1}x_{i,g_1}=\pm2$. Therefore, by calculation, Proposition 3.3 is straightforwardly established.

Appendix D. The proof of Proposition 3.4

Let $\boldsymbol{x}_{f_1f_1},\boldsymbol{x}_{f_2f_2}$ denote any two distinct three-level pure-quadratic effect columns in the model matrix of the HMD, $\boldsymbol{x}_{f_1f_2},\boldsymbol{x}_{f_2f_3}$ denote any two distinct three-level interaction effect columns in the model matrix of the HMD and $\boldsymbol{x}_{g_1g_2}$, $\boldsymbol{x}_{f_1g_1}$ (or $\boldsymbol{x}_{f_2g_1}$) denote a two-level factor interaction column and a two-factor interaction column of a three-level design column and a two-level design column in the model matrix of the HMD, respectively. Let $\overline{x}_{**}$ denote the average of corresponding column $\boldsymbol{x}_{**}$, where $**\in\{f_1f_1,f_2f_2,f_1f_2,f_2f_3,g_1g_2,f_1g_1,f_2g_1\}$.

(i) We calculate the correlation between two three-level pure-quadratic effect columns, denoted as $\rho_{f_1f_1,f_2f_2}$. By the construction of the HMD, we have $\overline{x}_{f_1f_1}=\overline{x}_{f_2f_2}=(m-1)/m$, $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_2f_2}=2(m-2)$, $\sum_{i=1}^{2m}x_{i,f_1f_1}=\sum_{i=1}^{2m}x_{i,f_2f_2}=\sum_{i=1}^{2m}x_{i,f_1f_1}^2=2(m-1)$, then by some simple calculation, we have

$$\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})(x_{i,f_2f_2}-\overline{x}_{f_2f_2})=\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_2f_2}-2m(\overline{x}_{f_1f_1})^2=-2/m, \tag{14}$$

$$\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})^2=\sum_{i=1}^{2m}x_{i,f_1f_1}^2-2m(\overline{x}_{f_1f_1})^2=2(m-1)/m. \tag{15}$$

Substituting them into the expression of correlation, we obtain Proposition 3.4 (i).

(ii) We calculate the correlation between a three-level pure-quadratic effect column and a three-level interaction column involving one common three-level factor, denoted as $\rho_{f_1f_1,f_1f_2}$. By the construction of the HMD, we have $\overline{x}_{f_1f_2}=0$ or $\pm2/m$. For the case with $\overline{x}_{f_1f_2}=0$, the construction of the HMD leads to $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_1f_2}=0$, then $\rho_{f_1f_1,f_1f_2}=0$. For the case with $\overline{x}_{f_1f_2}=\pm2/m$, we know that $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_1f_2}=\pm4$, $\sum_{i=1}^{2m}x_{i,f_1f_1}=\sum_{i=1}^{2m}x_{i,f_1f_1}^2=2(m-1)$, $\overline{x}_{f_1f_1}=(m-1)/m$, $\sum_{i=1}^{2m}x_{i,f_1f_2}=\pm4$, $\sum_{i=1}^{2m}x_{i,f_1f_2}^2=2(m-2)$, then $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})(x_{i,f_1f_2}-\overline{x}_{f_1f_2})=\pm4/m$, $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})^2=2(m-1)/m$, $\sum_{i=1}^{2m}(x_{i,f_1f_2}-\overline{x}_{f_1f_2})^2=2(m^2-2m-4)/m$, thus $\rho_{f_1f_1,f_1f_2}=\frac{\pm2}{\sqrt{(m-1)(m^2-2m-4)}}$. Thus, Proposition 3.4 (ii) is proved.

(iii) We calculate the correlation between a three-level pure-quadratic effect column and a three-level factor interaction column, which involves three different three-level factors, denoted as $\rho_{f_1f_1,f_2f_3}$. By the construction of the HMD, we know that $\overline{x}_{f_2f_3}=0$ or $\pm2/m$. The detailed proof for each case is as follows:

For the case with $\overline{x}_{f_2f_3}=0$, it is easy to obtain the following: $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_2f_3}=\pm2$, $\overline{x}_{f_1f_1}=(m-1)/m$, $\sum_{i=1}^{2m}x_{i,f_2f_3}=0$, $\sum_{i=1}^{2m}x_{i,f_2f_3}^2=2(m-2)$, and $\sum_{i=1}^{2m}x_{i,f_1f_1}=2(m-1)$, then $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})(x_{i,f_2f_3}-\overline{x}_{f_2f_3})=\pm2$, $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})^2=2(m-1)/m$, and $\sum_{i=1}^{2m}(x_{i,f_2f_3}-\overline{x}_{f_2f_3})^2=2(m-2)$. Therefore, we obtain $\rho_{f_1f_1,f_2f_3}=\frac{\pm\sqrt{m}}{\sqrt{(m-1)(m-2)}}$.

For the case with $\overline{x}_{f_2f_3}=\frac{\pm2}{m}$, we have $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_2f_3}=\pm6$ or ±2. If $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_2f_3}=\pm6$, by the construction of the HMD, we know $\overline{x}_{f_1f_1}=(m-1)/m$, $\sum_{i=1}^{2m}x_{i,f_1f_1}=2(m-1)$, $\sum_{i=1}^{2m}x_{i,f_2f_3}=\pm4$, and $\sum_{i=1}^{2m}x_{i,f_2f_3}^2=2(m-2)$, then $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})(x_{i,f_2f_3}-\overline{x}_{f_2f_3})=\pm2(m+2)/m$, $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})^2=2(m-1)/m$, and $\sum_{i=1}^{2m}(x_{i,f_2f_3}-\overline{x}_{f_2f_3})^2=2(m^2-2m-4)/m$. Therefore, we obtain $\rho_{f_1f_1,f_2f_3}=\frac{\pm(m+2)}{\sqrt{(m-1)(m^2-2m-4)}}$. Similarly, if $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_2f_3}=\pm2$, we can obtain $\rho_{f_1f_1,f_2f_3}=\frac{\pm(m-2)}{\sqrt{(m-1)(m^2-2m-4)}}$. Thus, Proposition 3.4 (iii) is proved.

(iv) We calculate the correlation between a three-level pure-quadratic effect column and a two-level factor interaction column, denoted as $\rho_{f_1f_1,g_1g_2}$. By the construction of the HMD, we have $\overline{x}_{g_1g_2}=0$. Further, the construction of the HMD leads to $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,g_1g_2}=\pm2$, $\overline{x}_{f_1f_1}=(m-1)/m$, $\sum_{i=1}^{2m}x_{i,f_1f_1}=2(m-1)$, $\sum_{i=1}^{2m}x_{i,g_1g_2}^2=2m$, $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})(x_{i,g_1g_2}-\overline{x}_{g_1g_2})=\pm2$, $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})^2=2(m-1)/m$, and $\sum_{i=1}^{2m}(x_{i,g_1g_2}-\overline{x}_{g_1g_2})^2=2m$. Therefore, we obtain $\rho_{f_1f_1,g_1g_2}=\frac{\pm1}{\sqrt{m-1}}$. Thus, Proposition 3.4 (iv) is proved.

(v) We calculate the correlation between a three-level pure-quadratic effect column and a two-factor interaction column of a three-level factor and a two-level factor involving a common three-level factor, denoted as $\rho_{f_1f_1,f_1g_1}$. By the construction of the HMD, we have $\overline{x}_{f_1g_1}=\pm1/m$. Further, the construction of the HMD leads to $\sum_{i=1}^{2m}x_{i,f_1f_1}x_{i,f_1g_1}=\pm2$, $\sum_{i=1}^{2m}x_{i,f_1g_1}=\pm2$, $\sum_{i=1}^{2m}x_{i,f_1g_1}^2=2(m-1)$, then $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})(x_{i,f_1g_1}-\overline{x}_{f_1g_1}) = \pm 2/m$, $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})^2 = 2(m-1)/m$, and $\sum_{i=1}^{2m}(x_{i,f_1g_1}-\overline{x}_{f_1g_1})^2 = 2(m^2-m-1)/m$. Therefore, we obtain $\rho_{f_1f_1,f_1g_1} = \frac{\pm 1}{\sqrt{(m-1)(m^2-m-1)}}$. Thus, Proposition 3.4 (v) is proved.

(vi) We calculate the correlation between a pure-quadratic effect column and a two-factor interaction column of a three-level design column and a two-level design column involving different three-level factors, denoted as $\rho_{f_1f_1,f_2g_1}$. By the construction of the HMD, we have $\overline{x}_{f_2g_1} = \pm 1/m$. Further, the construction of the HMD leads to $\sum_{i=1}^{2m} x_{i,f_1f_1}x_{i,f_2g_1} = 0$ or ± 4, then $\sum_{i=1}^{2m}(x_{i,f_1f_1} - \overline{x}_{f_1f_1})(x_{i,f_2g_1}-\overline{x}_{f_2g_1}) = \pm 2(m-1)/m$ or $\pm 2(m+1)/m$, $\sum_{i=1}^{2m}(x_{i,f_1f_1}-\overline{x}_{f_1f_1})^2 = 2(m-1)/m$, and $\sum_{i=1}^{2m}(x_{i,f_2g_1}-\overline{x}_{f_2g_1})^2 = 2(m^2-m-1)/m$. Therefore, we obtain $\rho_{f_1f_1,f_2g_1} = \frac{\pm(m-1)}{\sqrt{(m-1)(m^2-m-1)}}$ or $\frac{\pm(m+1)}{\sqrt{(m-1)(m^2-m-1)}}$. Thus, Proposition 3.4 (vi) is proved.

References

Ai, M., Fang, K.T., He, S., 2007. E(χ^2)-optimal mixed-level supersaturated designs. J. Statist. Plann. Inference 137, 306–316.
Box, G.E.P., Meyer, R.D., 1986. An analysis for unreplicated fractional factorials. Technometrics 28, 11–18.
Draper, N.R., Lin, D.K.J., 1990. Small response surface designs. Technometrics 32, 187–194.
Georgiou, S.D., 2014. Supersaturated designs: A review of their construction and analysis. J. Statist. Plann. Inference 144, 92–109.
Hedayat, A.S., Sloane, N.J.A., Stufken, John, 1999. Orthogonal Arrays: Theory and Applications. Springer, New York.
Jones, B., Nachtsheim, C.J., 2011. A class of three-level designs for definitive screening in the presence of second-order effects. J. Qual. Technol. 43, 1–15.
Jones, B., Nachtsheim, C.J., 2013. Definitive screening designs with added two-level categorical factors. J. Qual. Technol. 45, 121–129.
Lin, D.K.J., 1993. A new class of supersaturated designs. Technometrics 35, 28–31.
Marshall, A.W., Olkin, I., Arnold, B.C., 2009. Inequalities: Theory of Majorization and Its Applications, second ed. Springer, New York.
Nguyen, N.K., 1996. An algorithmic approach to constructing supersaturated designs. Technometrics 38, 69–73.
Nguyen, N.K., Stylianou, S., 2013. Constructing definitive screening designs using cyclic generators. J. Stat. Theory Pract. 7, 713–724.
Phoa, F.K.H., Lin, D.K.J., 2015. A systematic approach for the construction of definitive screening designs. Statist. Sinica 25, 853–861.
Shi, C., Tang, B., 2018. Designs from good Hadamard matrices. Bernoulli 24, 661–671.
Shi, C., Tang, B., 2019. Supersaturated designs robust to two-factor interactions. J. Statist. Plann. Inference 200, 119–128.
Sun, F.S., Lin, D.K.J., Liu, M.Q., 2011. On construction of optimal mixed-level supersaturated designs. Ann. Statist. 39, 1310–1333.
Tang, B., Wu, C.F.J., 1997. A method for constructing supersaturated designs and its $E(s^2)$-optimality. Canad. J. Statist. 25, 191–201.
Wu, C.F.J., 1993. Construction of supersaturated designs through partially aliased interactions. Biometrika 80, 661–669.
Xiao, L., Lin, D.K.J., Bai, F., 2012. Constructing definitive screening designs using conference matrices. J. Qual. Technol. 44, 2–8.