

# Construction of Supersaturated Designs with Small Coherence for Variable Selection

YOURAN QI AND PETER CHIEN\*

## Abstract

The supersaturated design is often used to discover important factors in an experiment with a large number of factors and a small number of runs. We propose a method for constructing supersaturated designs with small coherence. Such designs are useful for variable selection methods such as the Lasso. Examples are provided to illustrate the proposed method.

KEYWORDS AND PHRASES: Design of Experiments, Supersaturated Design, Unbalanced Design, Coherence, Linear Model, Variable Selection.

## 1. INTRODUCTION

Modern experiments can involve a large number of factors. Assuming the effect sparsity principle [3], an experiment with high-dimensional inputs only has a small number of significant factors. Two-level supersaturated designs are often used to identify important factors in screening experiments. Data from such a design can be modeled by a linear model [33] given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where  $\mathbf{X}$  is the  $n \times p$  design matrix at two levels  $-1$  and  $+1$ , denoted by  $-$  and  $+$ ,  $\mathbf{y}$  is the response vector of the  $n$  observations,  $\boldsymbol{\beta}$  is the regression vector of  $p$  coefficients, the error vector  $\boldsymbol{\epsilon}$  follows the multivariate normal distribution  $N_n(\mathbf{0}, \sigma^2\mathbf{I})$  and  $\mathbf{I}$  is the  $n \times n$  identity matrix. In  $\mathbf{X}$ , the first column is the intercept with all  $+$ 's and the remaining columns represent the  $p - 1$  factors. We call  $\mathbf{X}$  balanced if all its factors consist of an equal number of  $-$  and  $+$ . If  $p = n$ ,  $\mathbf{X}$  is a *saturated* design since all the degrees of freedom are used to estimate  $\boldsymbol{\beta}$ . If  $p > n$ ,  $\mathbf{X}$  is a *supersaturated* design since there are not enough degrees of freedom to estimate all components of  $\boldsymbol{\beta}$ . Construction of supersaturated designs that can accommodate a large number of factors relative to the number of runs has been gaining more and more attention [5, 16, 30, 26]. Popular criteria for comparing supersaturated designs include minimax [2],  $E(s^2)$  [18, 27] and mean square correlation [8].

It is increasingly common to use modern variable selection methods to analyze data from a supersaturated design and identify important factors [18, 32, 17]. For example, when the number of variables is greater than the number of observations, Osborne, Presnell and Turlach [22] advised to

use the Lasso [28] for an initial selection, and then conduct a best subset selection (or similar method) on the variables selected by the Lasso. We focus on discussing and demonstrating our method for the Lasso given its popularity. The proposed designs also work with other penalized regression methods. For variable selection methods such as the Lasso, Zhao and Yu [34] showed that their performance depends more critically on the worst case column correlation than the average column correlation of the design matrix. More details about this point can be found in Appendix B.

Motivated by the importance of controlling the worst case column correlation, we propose a new method to construct supersaturated designs according to the *coherence* criterion [7], which measures the worst case column correlation. The coherence of an  $n \times p$  matrix  $\mathbf{X}$  is

$$\mu(\mathbf{X}) = \max_{1 \leq i < j \leq p} \frac{|\langle \mathbf{x}_i, \mathbf{x}_j \rangle|}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|},$$

where the subscript  $i$  denotes the  $i$ th column,  $\langle \cdot, \cdot \rangle$  is the dot product and  $\|\cdot\|$  is the  $L_2$ -norm. Whenever there is no confusion, hereinafter we will drop the symbol  $\mathbf{X}$  in  $\mu(\mathbf{X})$ . Clearly,  $0 \leq \mu \leq 1$ . The smaller  $\mu$  is, the closer the matrix is to orthogonal. If the columns of  $\mathbf{X}$  are centered,  $\mu$  is reduced to the maximum absolute column correlation of  $\mathbf{X}$ , which is equivalent to the correlation criterion used in [19]. Note that

$$\mu(\mathbf{X}) = \max_{1 \leq i < j \leq p} \frac{|\mathbf{x}_i^\top \mathbf{x}_j|}{n} = \frac{s_{\max}}{n},$$

where  $s_{\max} = \max_{1 \leq i < j \leq p} |\mathbf{x}_i^\top \mathbf{x}_j|$  is the value used in the minimax criterion [2]. Different from  $s_{\max}$ ,  $\mu(\mathbf{X})$  reflects the sample size  $n$  in its definition. Because the sample size matters in many data analysis procedures, we will construct designs under  $\mu(\mathbf{X})$  instead of  $s_{\max}$  (we will also compare designs with different sample sizes in Section 3.1).

\*Corresponding author.

Our construction method allows some columns of the design matrix to be unbalanced. Factor balance in a supersaturated design guarantees an accurate estimate of the intercept but unbalance provides other significant benefits in constructing designs. Being unbalanced provides flexibility in controlling the column correlations between every pair of the main effects to achieve a lower average column correlation [29]. Sacrificing the precision of the estimate of the intercept can better estimate the main effects. In addition, allowing unbalance enables us to construct designs with more flexible sizes.

The remainder of the article will unfold as follows. In Section 2, we detail our construction method. In Section 3, we use several examples to demonstrate the advantage of the constructed designs against some benchmark designs for a variable selection problem. We conclude the paper and provide some discussions in Section 4.

## 2. THE CONSTRUCTION METHOD

### 2.1 Notation and Definitions

Throughout, “#Balance” denotes the number of balanced columns of a design. Let  $\mathbf{1}_n$  denote the  $n$ -dimensional unit vector with +1’s. Let  $\mathbf{I}$  denote the identity matrix. For a matrix  $\mathbf{A}$  with entries in  $\{-1, +1\}$  and an even number of rows, let  $\mathbf{A}^*$  denote a matrix obtained by changing the signs of entries in all odd rows and  $\mathbf{A}^{**}$  denote a matrix obtained by changing the signs of entries in all even rows, where  $*$  and  $**$  are two matrix operators. Note that  $(-\mathbf{A})^{**} = \mathbf{A}^*$ . An  $n \times n$  Hadamard matrix  $\mathbf{H}$  is a matrix with entries in  $\{-1, +1\}$  and  $\mathbf{H}^T \mathbf{H} = n\mathbf{I}$ , where  $n$  can only be 1, 2 or a multiple of 4.

### 2.2 Construction Steps

Our construction method expands a  $6m \times p$  supersaturated design  $\mathbf{D}_0$  with  $\mu \leq 1/3$  to a  $12m \times 4p$  design with  $\mu \leq 1/3$ , where  $m, p$  are positive integers with  $6m < p$ . Partition  $\mathbf{D}_0$  as

$$\mathbf{D}_0 = \begin{bmatrix} \mathbf{U} \\ \mathbf{L} \end{bmatrix}, \tag{2.1}$$

where  $\mathbf{U}$  and  $\mathbf{L}$  have  $4m$  and  $2m$  rows, respectively. The proposed method has three steps.

- Step 1: Use two copies of  $\mathbf{D}_0$  to obtain a matrix

$$\mathbf{D}_1 = \begin{bmatrix} \mathbf{U} \\ \mathbf{U} \\ \mathbf{L} \\ \mathbf{L} \end{bmatrix}.$$

- Step 2: Expand  $\mathbf{D}_1$  to obtain a matrix

$$\mathbf{D}_2 = \begin{bmatrix} \mathbf{U} & \mathbf{U} \\ \mathbf{U} & \mathbf{U}^* \\ \mathbf{L} & \mathbf{L} \\ \mathbf{L} & -\mathbf{L} \end{bmatrix}. \tag{2.2}$$

- Step 3: Expand  $\mathbf{D}_2$  to obtain a matrix

$$\mathbf{D}_3 = \begin{bmatrix} \mathbf{U} & \mathbf{U} & \mathbf{U}^{**} & \mathbf{U}^{**} \\ \mathbf{U} & \mathbf{U}^* & \mathbf{U} & \mathbf{U}^* \\ \mathbf{L} & \mathbf{L} & \mathbf{L}^* & \mathbf{L}^* \\ \mathbf{L} & -\mathbf{L} & \mathbf{L}^{**} & \mathbf{L}^* \end{bmatrix}.$$

If  $\mathbf{D}_1$  is viewed as a block matrix with four rows, Step 2 applies the operator  $*$  to the second row and changes the signs of all entries in the fourth row to obtain  $\mathbf{D}_2$ . If  $\mathbf{D}_2$  is viewed as a block matrix with four rows and two columns, Step 3 applies the operator  $**$  to the first and fourth rows and applies the operator  $*$  to the third row to obtain  $\mathbf{D}_3$ . It can be proved that the coherence of  $\mathbf{D}_1$ ,  $\mathbf{D}_2$  and  $\mathbf{D}_3$  is no greater than  $1/3$  by Lemmas 1, 2, 3 and Theorem 1 in Appendix A.

Since the design  $\mathbf{D}_3$  from the construction is a  $12m \times 4p$  design with  $\mu \leq 1/3$ , repeating the above procedure multiple times with  $\mathbf{D}_0$  being  $\mathbf{D}_3$  produces a class of designs with  $(2^k \times 6m)$  rows,  $(4^k \times p)$  columns and  $\mu \leq 1/3$  for every positive integer  $k$ .

The only requirement for  $\mathbf{D}_0$  is to be a  $6m \times p$  two-level supersaturated design with  $\mu \leq 1/3$ . For a given pair of  $m$  and  $p$ , different choices of  $\mathbf{D}_0$  yield different forms of  $\mathbf{D}_3$  but all of them have guaranteed small coherence. As suggested by Chen and Lin [6] and Liu, Ruan and Dean [20], a supersaturated design with coherence of  $1/3$  can identify most of the important factors.

### 2.3 Examples

We now provide several examples for the proposed method with different choices of  $\mathbf{D}_0$ . In the following examples, we show the first several designs constructed by the proposed method and choose  $\mathbf{D}_0$  to have as many columns as possible.

**Example 1.** Let  $\mathbf{D}_0$  be a  $6 \times 16$  design given by

$$\begin{bmatrix} + & + & + & + & + & + & + & + & + & + & + & + & + & + & + & + \\ + & + & + & - & - & - & - & - & - & - & - & - & - & - & - & - \\ + & + & - & - & + & + & - & - & + & - & - & + & - & + & + & - \\ + & - & - & - & + & + & - & - & - & + & - & - & + & - & - & + \\ + & - & + & - & + & - & + & - & + & - & + & - & + & - & + & - \\ + & - & + & - & - & + & - & + & + & - & + & - & - & + & - & + \end{bmatrix},$$

where  $-$  and  $+$  denote  $-1$  and  $+1$ , respectively, and its coherence is  $1/3$ . The coherence of the above design attains the lower bound derived in [31]. The sizes, coherence values and numbers of the balanced columns of the first several designs constructed by the proposed method are given in Table 1.

**Example 2.** Let

$$\mathbf{D}_0 = [ \mathbf{1}_{24} \ \mathbf{D}_{01} \ \mathbf{D}_{02} \ \mathbf{D}_{03} ],$$

where  $\mathbf{D}_{01}$  is the  $24 \times 23$  Plackett and Burman design [23],  $\mathbf{D}_{02}$  is formed by taking the two-order interaction terms of  $\mathbf{D}_{01}$  and  $\mathbf{D}_{03}$  is formed by taking the three-order interaction terms of  $\mathbf{D}_{01}$ . Then  $\mathbf{D}_0$  is a  $24 \times 2048$  design with  $\mu = 1/3$ .

Table 1. Designs Constructed from a  $6 \times 16 \mathbf{D}_0$ .

Size	$\mu$	#Balance
$12 \times 64$	1/3	24
$24 \times 256$	1/3	168
$48 \times 1024$	1/3	840
$96 \times 4096$	1/3	3720
$192 \times 16384$	1/3	15624

Table 2. Designs Constructed from a  $24 \times 2048 \mathbf{D}_0$ .

Size	$\mu$	#Balance
$48 \times 8192$	1/3	2968
$96 \times 32768$	1/3	18616
$192 \times 131072$	1/3	99064

Table 3. Designs Constructed from a  $30 \times 59 \mathbf{D}_0$ .

Size	$\mu$	#Balance
$60 \times 236$	1/3	101
$120 \times 944$	1/3	550
$240 \times 3776$	1/3	2864
$480 \times 15104$	1/3	13156
$960 \times 60416$	1/3	56396

The first 277 columns of this design form the design in [32]. The sizes, coherence values and numbers of the balanced columns of the first several designs constructed by the proposed method are given in Table 2.

**Example 3.** Let  $\mathbf{D}_0$  be the  $30 \times 59$  design from Lin [18], where an additional column  $\mathbf{1}_{30}$  is added as the first column. This design has  $\mu = 0.2$ . The sizes, coherence values and numbers of the balanced columns of the first several designs constructed by the proposed method are given in Table 3. This example illustrates that our method can use a design  $\mathbf{D}_0$  with  $\mu < 1/3$ .

Whenever necessary, one can select some columns of the designs constructed for use. Selecting a subset of columns could remain coherence the same or further decrease it, but will never increase coherence. For example, for a given  $p$ , first, one can select as many balanced columns as possible. If there are no sufficient balanced columns, then one can select some unbalanced columns to achieve  $p$  columns. In addition, one may also select the columns according to an additional criterion such as  $E(s^2)$ .

### 2.4 Generalization

We now generalize the proposed method to obtain a supersaturated design with  $\mu$  smaller than  $1/3$ . Recall that our method partitions  $\mathbf{D}_0$  into two parts. The upper part  $\mathbf{U}$  has  $6mr$  rows and the lower part  $\mathbf{L}$  has  $6m(1-r)$  rows with a partition ratio  $r = 2/3$ . We generalize the original construction by using a different partition ratio and stopping at Step 2. Suppose that  $\mathbf{D}_0$  is an  $n \times p$  two-level supersaturated design

Table 4. Designs Constructed from a  $30 \times 59 \mathbf{D}_0$  by the Generalization.

Size	$\mu$	#Balance
$60 \times 118$	0.2	76
$120 \times 236$	0.2	112
$240 \times 472$	0.2	184
$480 \times 944$	0.2	328
$960 \times 1888$	0.2	616

with  $\mu \leq t/n$  for  $t \in \{2, 4, \dots, n/2\}$  and an even  $n$ . Partition  $\mathbf{D}_0$  as in (2.1), where  $\mathbf{U}$  has  $2t$  rows and  $\mathbf{L}$  has  $n - 2t$  rows with  $r = 2t/n$ . Define  $\mathbf{D}_2$  as in (2.2). Then  $\mathbf{D}_2$  is the supersaturated design constructed by our generalized method. It can be proved that the coherence of  $\mathbf{D}_2$  constructed here is no greater than  $t/n$  by Theorem 2 in Appendix A.

The generalization indicates the possibility of expanding any two-level supersaturated design with  $\mu \leq 1/2$  and an even number of runs while retaining its coherence. For the original construction, coherence must be no greater than  $1/3$  and the initial design must have  $6m$  runs but the initial design with  $p$  columns can be expanded to a design with  $4p$  columns and  $\mu \leq 1/3$ . The generalization only requires that the coherence is no greater than  $1/2$  and the initial design has an even number of runs but the initial design with  $p$  columns can only be expanded to a design with  $2p$  columns and the same coherence. In addition, the number of balanced columns can be fewer than the number of runs for the designs constructed by the generalization. Here is an example.

**Example 4.** Let  $\mathbf{D}_0$  be the  $30 \times 59$  design in Example 3 with  $\mu = 0.2$ . The sizes, coherence values and numbers of the balanced columns of the first several designs constructed by the generalization are given in Table 4.

## 3. SIMULATION STUDY

In this section, we compare the proposed designs with four popular classes of supersaturated designs: Lin’s designs [18], Wu’s designs [32], the Bayesian D-optimal supersaturated designs (Jones, Lin, and Nachtsheim 2008) and the  $UE(s^2)$ -optimal designs [15] for the Lasso problem by simulations. These designs are denoted by “LIN”, “WU”, “BAYES” and “JM” respectively and our proposed designs are denoted by “Proposed”.

Lin’s designs are constructed from the Plackett and Burman design [23]. According to Bulutoglu and Cheng [4], Lin’s designs are  $E(s^2)$ -optimal among all the balanced two-level supersaturated designs. Wu’s designs can be obtained by appending interaction columns to a Hadamard matrix. Bulutoglu and Cheng [4] showed that in certain cases, Wu’s designs are  $E(s^2)$ -optimal among all the balanced two-level supersaturated designs. The Bayesian D-optimal supersaturated design is obtained by the coordinate exchange algorithm [9, 14]. Any  $n$  rows of a  $p \times p$  Hadamard matrix form a type of  $UE(s^2)$ -optimal design [15, 5].

Table 5. Active Coefficients Settings.

Case	#Active	Active Coefficients and Their Values								
1	2	$\beta_1 = 1$	$\beta_{21} = 5$	$\beta_{23} = 10$						
2	2	$\beta_1 = 0.1$	$\beta_{21} = 1$	$\beta_{23} = 1.5$						
3	4	$\beta_1 = 1$	$\beta_6 = 5$	$\beta_8 = 10$	$\beta_9 = 10$	$\beta_{26} = 15$				
4	4	$\beta_1 = 0.1$	$\beta_6 = 0.6$	$\beta_8 = 1.7$	$\beta_9 = 0.9$	$\beta_{26} = 1$				
5	4	$\beta_1 = 1$	$\beta_{19} = 5$	$\beta_{59} = 14$	$\beta_{143} = 10$	$\beta_{214} = 13$				
6	4	$\beta_1 = 0.1$	$\beta_{19} = 1.3$	$\beta_{59} = 0.9$	$\beta_{143} = 1$	$\beta_{214} = 1.2$				
7	6	$\beta_1 = 1$	$\beta_{38} = 5$	$\beta_{111} = 5$	$\beta_{122} = 9$	$\beta_{123} = 9$	$\beta_{147} = 9$	$\beta_{151} = 11$		
8	6	$\beta_1 = 0.1$	$\beta_{38} = 1.5$	$\beta_{111} = 1.5$	$\beta_{122} = 1.9$	$\beta_{123} = 1.9$	$\beta_{147} = 1.9$	$\beta_{151} = 2.8$		

Table 6. Comparison with Lin’s Design.

Case	#Active	Design	Size	$\mu$	$E(s^2)$	#Balance	AFDR	AMR	MSE	EME
1	2	Proposed	$12 \times 27$	0.33	9.70	24	<b>0.49</b>	<b>0.00</b>	<b>0.64</b>	<b>7.22</b>
		LIN	$14 \times 27$	0.43	7.84	26	0.59	0.00	0.93	9.18
2	2	Proposed	$12 \times 27$	0.33	9.70	24	<b>0.51</b>	<b>0.02</b>	<b>0.74</b>	<b>9.15</b>
		LIN	$14 \times 27$	0.43	7.84	26	0.59	0.08	1.13	12.51
3	4	Proposed	$12 \times 27$	0.33	9.70	24	<b>0.25</b>	<b>0.18</b>	<b>73.58</b>	<b>537.69</b>
		LIN	$14 \times 27$	0.43	7.84	26	0.40	0.28	230.66	794.73
4	4	Proposed	$12 \times 27$	0.33	9.70	24	<b>0.49</b>	<b>0.29</b>	<b>2.15</b>	<b>18.52</b>
		LIN	$14 \times 27$	0.43	7.84	26	0.64	0.53	4.19	19.79

We simulate data from the linear model in (1.1) with different designs and active coefficients, where the error vector  $\epsilon$  follows the standard multivariate normal distribution  $N_n(\mathbf{0}, \mathbf{I})$ . We use the Lasso to select the active factors and fix the intercept term  $\beta_1$  as active. We repeat the above data generation and variable selection procedure  $N = 300$  times. We use the “cv.glmnet” function [12] in the R software [24] and calculate  $\hat{\beta}$  with “s=lambda.min”.

We conduct simulations using the eight active coefficients settings in Table 5 and denote them as Case 1, . . . , Case 8, respectively. We use “#Active” to denote the number of active factors. For each group of active coefficients, we consider both the case where they take large values, and the case where they take small values. For example, for  $\beta_1, \beta_{21}$  and  $\beta_{23}$ , we consider both the case where they take values of 1, 5 and 10, and the case where they take values of 0.1, 1 and 1.5. We let the regression coefficients of all inactive factors be zero and do not write them in Table 5. We neither use a single setting of active coefficients for all the comparisons, nor use all settings in Table 5 for each comparison, because of the following reasons. First, the designs under consideration have different numbers of columns and some designs do not have enough columns for a given setting, which makes the setting not applicable. Second, in practice, the more factors we are considering, the more active factors there tends to be, so we use settings with more active coefficients for designs with large numbers of columns, and settings with fewer active coefficients for designs with small numbers of columns.

For each design in our simulations, we will show its size, coherence, and number of balanced columns. Although  $E(s^2)$  is not our focused criterion, we will also show the

$E(s^2)$  of each design for reference. We use the following four criteria to compare variable selection and model fitting accuracy of each design:

1. Average False Discovery Rate (AFDR)  $\sum_{i=1}^N \text{FDR}_i / N$ , where

$$\text{FDR}_i = \frac{\text{the number of falsely discovered coefficients}}{\text{the number of discovered coefficients}}$$

if the discovered model is not a null model, and  $\text{FDR}_i = 1$  if the discovered model is a null model.

2. Average Miss Rate (AMR)  $\sum_{i=1}^N \text{MR}_i / N$ , where

$$\text{MR}_i = \frac{\text{the number of undiscovered active coefficients}}{\text{the number of active coefficients}}.$$

3. Mean Squared Error (MSE)  $\sum_{i=1}^N \|\beta - \hat{\beta}^{(i)}\|^2 / N$  to estimate  $E(\|\beta - \hat{\beta}\|^2)$ .
4. Expected Model Error (EME)  $\sum_{i=1}^N \|\mathbf{X}\beta - \mathbf{X}\hat{\beta}^{(i)}\|^2 / N$  to estimate  $E(\|\mathbf{X}\beta - \mathbf{X}\hat{\beta}\|^2)$ .

Smaller values of AFDR, AMR, MSE and EME are desirable. Note that AFDR and AMR are computed based on the number of discovered or active coefficients, not the number of discovered or active factors.

### 3.1 Comparison with Lin’s Design

We obtain the  $12 \times 27$  proposed design by selecting the intercept, the 24 balanced columns and the first two unbalanced columns of the  $12 \times 64$  design in Example 1. We use Lin’s  $14 \times 27$  design directly. The result is shown in Table 6, where and in other tables below, the smaller values

Table 7. Comparison with Lin's Design with Randomly Selected Active Factors.

Case	#Active	Design	Size	$\mu$	$E(s^2)$	#Balance	AFDR	AMR	MSE	EME
1	2	Proposed	$12 \times 27$	0.33	9.70	24	<b>0.61</b>	<b>0.00</b>	5.01	<b>9.02</b>
		LIN	$14 \times 27$	0.43	7.84	26	0.61	0.00	<b>1.59</b>	11.58
2	2	Proposed	$12 \times 27$	0.33	9.70	24	<b>0.58</b>	<b>0.09</b>	<b>1.34</b>	<b>10.83</b>
		LIN	$14 \times 27$	0.43	7.84	26	0.59	0.14	1.36	12.93
3	4	Proposed	$12 \times 27$	0.33	9.70	24	0.47	<b>0.00</b>	51.62	<b>27.23</b>
		LIN	$14 \times 27$	0.43	7.84	26	<b>0.41</b>	0.04	<b>33.89</b>	135.42
4	4	Proposed	$12 \times 27$	0.33	9.70	24	0.55	<b>0.15</b>	<b>2.11</b>	<b>13.23</b>
		LIN	$14 \times 27$	0.43	7.84	26	<b>0.45</b>	0.27	2.13	18.56

Table 8. Comparison with Wu's Designs.

Case	#Active	Design	Size	$\mu$	$E(s^2)$	#Balance	AFDR	AMR	MSE	EME
1	2	Proposed	$12 \times 64$	0.33	9.90	24	0.65	<b>0.00</b>	0.95	<b>8.41</b>
		WU	$12 \times 64$	0.33	11.06	63	<b>0.54</b>	0.00	<b>0.89</b>	9.03
2	2	Proposed	$12 \times 64$	0.33	9.90	24	0.66	0.05	1.20	12.46
		WU	$12 \times 64$	0.33	11.06	63	<b>0.56</b>	<b>0.04</b>	<b>0.98</b>	<b>11.22</b>
3	4	Proposed	$24 \times 256$	0.33	21.84	168	<b>0.41</b>	<b>0.00</b>	<b>0.96</b>	<b>19.96</b>
		WU	$24 \times 256$	0.33	23.00	255	0.65	0.00	2.08	23.25
4	4	Proposed	$24 \times 256$	0.33	21.84	168	<b>0.67</b>	<b>0.04</b>	<b>0.94</b>	<b>18.20</b>
		WU	$24 \times 256$	0.33	23.00	255	0.68	0.14	1.76	24.90
7	6	Proposed	$48 \times 300$	0.33	93.05	299	<b>0.62</b>	<b>0.00</b>	<b>1.76</b>	<b>27.15</b>
		WU	$48 \times 300$	0.33	41.99	299	0.75	0.00	2.52	37.94
8	6	Proposed	$48 \times 300$	0.33	93.05	299	<b>0.71</b>	<b>0.00</b>	<b>1.74</b>	<b>21.65</b>
		WU	$48 \times 300$	0.33	41.99	299	0.77	0.01	2.58	44.09

of the four criteria are in boldface. Table 6 indicates that the proposed design outperforms Lin's design in terms of variable selection and parameter estimation with the Lasso. In addition, the proposed design has fewer runs than Lin's design.

To verify that this result is not due to some better property of the selected active factors in one design versus another, we also conducted a follow-up comparison with randomly selected active factors. More specifically, for a given setting in Table 5, we randomly select an #Active number of active factors with the same values given in the setting, and repeat 10 times to get 10 sets of active factors. Then, for each set of randomly selected active factors, we repeat the above data generation and variable selection procedure  $N = 30$  times for the proposed design and Lin's design. The final criteria are averaged over the criteria from the 10 sets of active factors, and they can be found in Table 7. According to Table 7, the proposed design still generally outperforms Lin's design, even with randomly selected active factors. To save computational cost and keep our paper concise, we will omit such a verification for the following comparisons.

### 3.2 Comparison with Wu's Designs

We use the  $12 \times 64$ ,  $24 \times 256$  proposed designs in Example 1 directly. We obtain the  $48 \times 300$  proposed design by selecting the intercept and the first 299 balanced columns of the  $48 \times 8192$  design in Example 2. We obtain Wu's  $12 \times 64$ ,

$24 \times 256$  and  $48 \times 300$  designs by selecting the first 64 columns of Wu's  $12 \times 67$  design, the first 256 columns of Wu's  $24 \times 277$  design and the first 300 columns of Wu's  $48 \times 1129$  design, respectively. The result shown in Table 8 indicates that the proposed designs are generally better than Wu's designs in terms of variable selection and parameter estimation with the Lasso. In almost all cases, the proposed designs have better performance while in the first two cases, Wu's designs perform relatively better.

### 3.3 Comparison with the Bayesian D-Optimal Supersaturated Designs

We use the  $12 \times 64$  proposed design in Example 1 directly. We obtain the  $24 \times 253$  proposed design by selecting the intercept, the 168 balanced columns and the first 84 unbalanced columns of the  $24 \times 256$  design in Example 1. We first search a  $12 \times 64$  and a  $24 \times 256$  Bayesian D-optimal supersaturated design by the JMP<sup>®</sup> software [13] with 500 random starting designs. We use the  $12 \times 64$  design directly. We remove the 22th, 134th and 209th columns of the  $24 \times 256$  design to obtain the  $24 \times 253$  design used in the simulations since the original searched  $24 \times 256$  Bayesian D-optimal supersaturated design has  $\mu = 0.83$ . The result shown in Table 9 indicates that the proposed designs generally outperform the Bayesian D-optimal supersaturated designs in terms of variable selection and parameter estimation with the Lasso.

Table 9. Comparison with the Bayesian D-Optimal Supersaturated Designs.

Case	#Active	Design	Size	$\mu$	$E(s^2)$	#Balance	AFDR	AMR	MSE	EME
1	2	Proposed	$12 \times 64$	0.33	9.90	24	<b>0.65</b>	<b>0.00</b>	<b>0.95</b>	<b>8.41</b>
		BAYES	$12 \times 64$	0.67	11.04	63	0.76	0.06	45.02	76.31
2	2	Proposed	$12 \times 64$	0.33	9.90	24	<b>0.66</b>	<b>0.05</b>	<b>1.20</b>	<b>12.46</b>
		BAYES	$12 \times 64$	0.67	11.04	63	0.78	0.45	3.20	16.31
3	4	Proposed	$12 \times 64$	0.33	9.90	24	<b>0.30</b>	<b>0.00</b>	<b>1.83</b>	<b>18.04</b>
		BAYES	$12 \times 64$	0.67	11.04	63	0.66	0.16	240.96	535.21
4	4	Proposed	$12 \times 64$	0.33	9.90	24	0.65	<b>0.06</b>	<b>1.48</b>	<b>12.77</b>
		BAYES	$12 \times 64$	0.67	11.04	63	<b>0.63</b>	0.46	4.61	30.40
5	4	Proposed	$24 \times 253$	0.33	21.84	168	<b>0.60</b>	<b>0.00</b>	<b>1.61</b>	<b>21.59</b>
		BAYES	$24 \times 253$	0.67	22.91	252	0.67	0.00	3.62	27.99
6	4	Proposed	$24 \times 253$	0.33	21.84	168	0.70	<b>0.05</b>	<b>1.85</b>	<b>29.70</b>
		BAYES	$24 \times 253$	0.67	22.91	252	<b>0.61</b>	0.28	3.36	50.61
7	6	Proposed	$24 \times 253$	0.33	21.84	168	<b>0.54</b>	<b>0.00</b>	<b>4.24</b>	<b>38.37</b>
		BAYES	$24 \times 253$	0.67	22.91	252	0.67	0.28	162.70	934.85
8	6	Proposed	$24 \times 253$	0.33	21.84	168	0.72	<b>0.01</b>	<b>3.49</b>	<b>25.34</b>
		BAYES	$24 \times 253$	0.67	22.91	252	<b>0.70</b>	0.44	16.57	158.04

Table 10. Comparison with the  $UE(s^2)$ -Optimal Designs.

Case	#Active	Design	Size	$\mu$	$E(s^2)$	#Balance	MAFDR	MAMR	MMSE	MEME
1	2	Proposed	$12 \times 64$	0.33	9.90	24	0.67	<b>0.00</b>	<b>0.87</b>	<b>7.99</b>
		JM	$12 \times 64$	0.83	9.90	16	<b>0.61</b>	0.00	2.53	11.02
2	2	Proposed	$12 \times 64$	0.33	9.90	24	0.67	<b>0.03</b>	<b>1.04</b>	<b>10.38</b>
		JM	$12 \times 64$	0.83	9.90	16	<b>0.66</b>	0.26	2.35	15.26
3	4	Proposed	$12 \times 64$	0.33	9.90	24	<b>0.30</b>	<b>0.00</b>	<b>2.03</b>	<b>21.28</b>
		JM	$12 \times 64$	0.83	9.90	16	0.62	0.45	351.16	956.38
4	4	Proposed	$12 \times 64$	0.33	9.90	24	<b>0.59</b>	<b>0.09</b>	<b>1.44</b>	<b>13.18</b>
		JM	$12 \times 64$	0.83	9.90	16	0.66	0.55	4.47	24.95
5	4	Proposed	$24 \times 256$	0.33	21.84	168	<b>0.58</b>	<b>0.00</b>	<b>1.68</b>	<b>26.33</b>
		JM	$24 \times 256$	0.58	21.84	44	0.69	0.00	66.86	61.37
6	4	Proposed	$24 \times 256$	0.33	21.84	168	<b>0.72</b>	<b>0.02</b>	<b>1.42</b>	<b>20.00</b>
		JM	$24 \times 256$	0.58	21.84	44	0.73	0.31	3.73	41.41
7	6	Proposed	$24 \times 256$	0.33	21.84	168	<b>0.71</b>	<b>0.30</b>	<b>260.42</b>	<b>398.54</b>
		JM	$24 \times 256$	0.58	21.84	44	0.72	0.32	294.96	1336.26
8	6	Proposed	$24 \times 256$	0.33	21.84	168	<b>0.62</b>	<b>0.35</b>	<b>13.62</b>	<b>110.92</b>
		JM	$24 \times 256$	0.58	21.84	44	0.72	0.43	18.54	147.44

### 3.4 Comparison with the $UE(s^2)$ -Optimal Designs

According to Jones and Majumdar [15], for given  $n$  and  $p$ , there are a large number of  $UE(s^2)$ -optimal designs. In practical scenarios, people may randomly generate only one  $UE(s^2)$ -optimal design and use it to screen important factors. In our comparison, we want to mimic such a scenario, i.e., we want to know whether people will have a higher chance ( $> 0.5$ ) of getting a worse design (compared with our design) if they generate a  $UE(s^2)$ -optimal design in this way. Therefore, we randomly generate 500  $UE(s^2)$ -optimal designs and for each of them, repeat the simulation procedure  $N = 30$  times (rather than  $N = 300$  in comparisons with other competing designs). We take the median values for the four criteria introduced in the beginning of the section over the 500 randomly generated designs and compare these

median values with the corresponding criteria of the proposed design. If the AFDR of a proposed design is less than the median AFDR (MAFDR) of all the 500  $UE(s^2)$ -optimal designs, it means the AFDR of the proposed design is less than that of the  $UE(s^2)$ -optimal design with more than half probability. The coherence,  $E(s^2)$  and number of balanced columns in the table are also the corresponding medians. In addition, it is possible that a generated  $UE(s^2)$ -optimal design has  $\mu = 1$  and we will avoid using these designs in the simulations, because in practice if people get a bad design with  $\mu = 1$ , they will probably regenerate another one.

We use the  $12 \times 64$  and  $24 \times 256$  proposed designs in Example 1 directly. We obtain the  $UE(s^2)$ -optimal designs by randomly sampling 12 rows of a  $64 \times 64$  Hadamard matrix 500 times and 24 rows of a  $256 \times 256$  Hadamard matrix 500 times, respectively. The result shown in Table 10 indicates

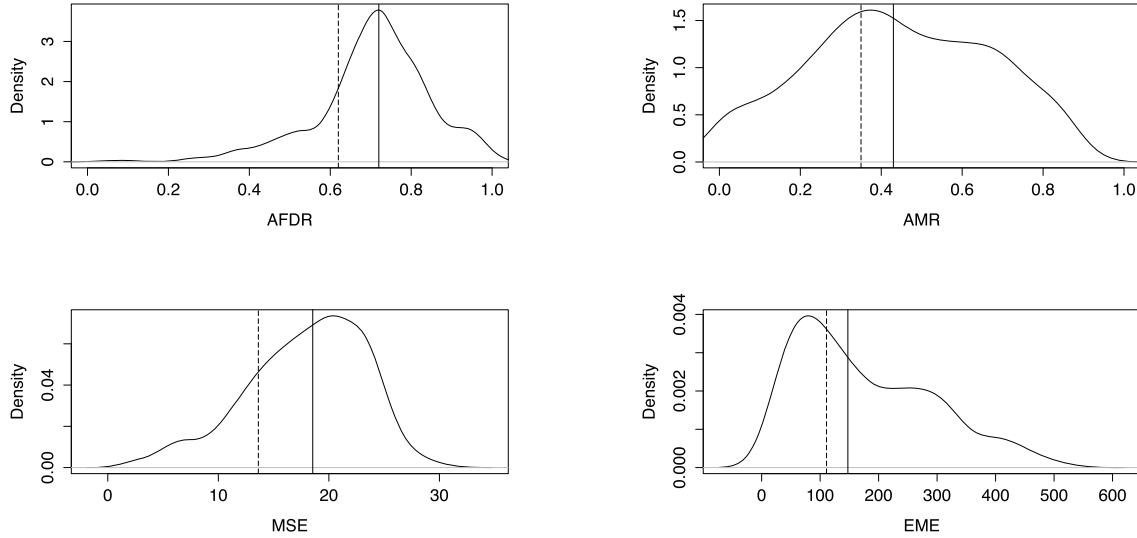


Figure 1: Comparison between the  $UE(s^2)$ -optimal design and the proposed design in Case 8. The solid curves are the kernel density estimates of the probability density functions of  $UE(s^2)$ -optimal design’s criteria and the solid lines are the corresponding median values. The dashed lines are the values of the proposed design’s criteria.

Table 11. Comparison Between Designs with the Same Size and Active Coefficients.

Case	#Active	Design	Size	$\mu$	$E(s^2)$	#Balance	AFDR	AMR	MSE	EME
1	2	Proposed	$12 \times 64$	0.33	9.90	24	0.65	<b>0.00</b>	0.95	<b>8.41</b>
		WU	$12 \times 64$	0.33	11.06	63	<b>0.54</b>	0.00	<b>0.89</b>	9.03
		BAYES	$12 \times 64$	0.67	11.04	63	0.76	0.06	45.02	76.31
		JM	$12 \times 64$	0.83	9.90	16	0.61	0.00	2.53	11.02
2	2	Proposed	$12 \times 64$	0.33	9.90	24	0.66	0.05	1.20	12.46
		WU	$12 \times 64$	0.33	11.06	63	<b>0.56</b>	<b>0.04</b>	<b>0.98</b>	<b>11.22</b>
		BAYES	$12 \times 64$	0.67	11.04	63	0.78	0.45	3.20	16.31
		JM	$12 \times 64$	0.83	9.90	16	0.66	0.26	2.35	15.26

that with more than half probability, the proposed designs outperform the  $UE(s^2)$ -optimal designs in terms of variable selection and parameter estimation with the Lasso. Furthermore, one may want to see the distributions of the criteria of 500  $UE(s^2)$ -optimal designs. We take Case 8 for an example and show the distributions of the criteria in Figure 1.

### 3.5 Summary of Comparisons

To help people better understand how the proposed design performs compared to competing designs with the same size and under the same active coefficient setting, we collect the comparisons between designs with exactly the same size and active coefficient setting all together in Table 11. Lin’s design is not of size  $12 \times 64$ , so it is not included in the table (it is actually impossible to construct a design of this size with Lin’s method).

According to Table 11, the proposed design significantly outperforms the Bayesian D-optimal design and the  $UE(s^2)$ -optimal design. In this scenario, Wu’s design performs slightly better than the proposed design, but there are also

scenarios where the proposed design performs better than Wu’s design (see Table 8).

In summary, for Wu’s design, the proposed design is generally comparable to it, because of the same coherence they have. For Lin’s design, the Bayesian D-optimal design and the  $UE(s^2)$ -optimal design, the proposed design generally outperforms them, except that it sometimes tends to select more factors and gives a larger AFDR. However, the proposed design always gives a smaller AMR when compared to these three designs. In practice, a small AMR is more important than a small AFDR, because missing active factors is a more serious problem than falsely discovering inactive factors. This is because we can conduct follow-up experiments to screen out spurious factors, but we can never detect active factors that were already removed in the screening phase [21].

## 4. CONCLUSION AND DISCUSSION

We have proposed a method for constructing supersaturated designs with small coherence. The constructed designs

are allowed to be unbalanced to achieve more flexible sample sizes. The proposed method uses direct constructions and it entails no intensive computing even for large  $p$ . Since the proposed method can efficiently construct a supersaturated design with a large number of columns, it can be applied to high-dimensional variable selection problems in marketing, biology, engineering and other areas.

Here are possible directions for future work. First, the proposed method can expand a  $6m \times p$  design  $\mathbf{D}_0$  with  $\mu \leq 1/3$  to a larger design with  $\mu \leq 1/3$ . One may be interested in finding a special  $\mathbf{D}_0$  such that the expanded design has coherence strictly smaller than  $1/3$  or using a different construction method to reduce the coherence upper bound. Second, beyond the Lasso, the proposed design can be applicable to other variable selection and high-dimensional problems where controlling coherence of the design matrix is important. Moreover, many new powerful variable selection techniques have emerged. For example, Fan and Lv [10] and Fan, Feng and Song [11] proposed the sure independence screening and nonparametric independence screening for ultrahigh dimensional variable selection. Another example is the best subset regression. Shen, Pan, Zhu and Zhou [25] provided theoretical support for the use of best subset regression. Bertsimas, King and Mazumder [1] proposed a mixed integer optimization approach for the best subset selection problem, where they showed that coherence plays a role in parameter specification of their approach. In future work, our proposed designs will be applied to other penalized variable selection methods with results reported elsewhere.

### APPENDIX A. PROOF FOR THE UPPER BOUND OF COHERENCE

**Lemma 1.** *For any matrix with levels  $-$  and  $+$ , if the signs of all the entries in a row are changed simultaneously, then*

$$\mathbf{A} = \begin{bmatrix} d_1^{(i)} & d_2^{(i)} & \cdots & d_{4m-1}^{(i)} & d_{4m}^{(i)} & \vdots & d_1^{(i)} & d_2^{(i)} & \cdots & d_{4m-1}^{(i)} & d_{4m}^{(i)} \\ d_1^{(j)} & d_2^{(j)} & \cdots & d_{4m-1}^{(j)} & d_{4m}^{(j)} & \vdots & -d_1^{(j)} & d_2^{(j)} & \cdots & -d_{4m-1}^{(j)} & d_{4m}^{(j)} \end{bmatrix}^\top,$$

$$\mathbf{B} = \begin{bmatrix} d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \cdots & d_{10m-1}^{(i)} & d_{10m}^{(i)} & \vdots & d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \cdots & d_{10m-1}^{(i)} & d_{10m}^{(i)} \\ d_{8m+1}^{(j)} & d_{8m+2}^{(j)} & \cdots & d_{10m-1}^{(j)} & d_{10m}^{(j)} & \vdots & -d_{8m+1}^{(j)} & -d_{8m+2}^{(j)} & \cdots & -d_{10m-1}^{(j)} & -d_{10m}^{(j)} \end{bmatrix}^\top.$$

Note that  $\mathbf{A}$  consists of  $2m$  blocks with 4 rows and 2 columns

$$\begin{bmatrix} d_{2l-1}^{(i)} & d_{2l-1}^{(j)} \\ d_{2l}^{(i)} & d_{2l}^{(j)} \\ d_{2l-1}^{(i)} & -d_{2l-1}^{(j)} \\ d_{2l}^{(i)} & d_{2l}^{(j)} \end{bmatrix}, \quad l = 1, \dots, 2m.$$

the dot product of every two columns remains the same. Thus, the coherence of the matrix remains the same.

**Lemma 2.** *The coherence of  $\mathbf{D}_1$  constructed in Section 2.2 is no greater than  $1/3$ .*

*Proof.* Throughout, for two columns  $\mathbf{d}_i$  and  $\mathbf{d}_j$  of a two-level design  $\mathbf{D}$  with  $n$  rows, we call  $|\mathbf{d}_i^\top \mathbf{d}_j|/n$  their absolute column correlation. Since  $\mathbf{D}_1$  is obtained by two copies of  $\mathbf{D}_0$ , the absolute column correlations of any two columns of  $\mathbf{D}_1$  are the same as those of  $\mathbf{D}_0$ . Thus, the coherence of  $\mathbf{D}_1$  equals that of  $\mathbf{D}_0$ , which is no greater than  $1/3$ .  $\square$

**Lemma 3.** *The coherence of  $\mathbf{D}_2$  constructed in Section 2.2 is no greater than  $1/3$ .*

*Proof.* Let  $\mathbf{d}^{(i)}$  denote the  $i$ th column of  $\mathbf{D}_2$  and  $d_j^{(i)}$  denote the  $j$ th entry of  $\mathbf{d}^{(i)}$ . Put

$$\mathbf{D}_2 = [ \mathbf{C}_1 \mid \mathbf{C}_2 ] = \begin{bmatrix} \mathbf{U} & \mathbf{U} \\ \mathbf{U} & \mathbf{U}^* \\ \mathbf{L} & \mathbf{L} \\ \mathbf{L} & -\mathbf{L} \end{bmatrix},$$

where  $\mathbf{C}_1$  and  $\mathbf{C}_2$  are  $12m \times p$  matrices.

Pick two arbitrary columns of  $\mathbf{D}_2$ .

- (i) Suppose that both of them are from  $\mathbf{C}_1$  or  $\mathbf{C}_2$ . By Lemma 2,  $\mu(\mathbf{C}_1)$  is  $\leq 1/3$ . By Lemmas 1 and 2,  $\mu(\mathbf{C}_2)$  is  $\leq 1/3$ . Thus, the absolute column correlation of the two columns is  $\leq 1/3$ .
- (ii) Suppose that one of the two columns is the  $i$ th column  $\mathbf{d}^{(i)}$  of  $\mathbf{C}_1$  and the other column is the  $j$ th column  $\mathbf{d}^{(p+j)}$  of  $\mathbf{C}_2$  with  $1 \leq i, j \leq p$ . Partition the two columns as

$$[ \mathbf{d}^{(i)} \quad \mathbf{d}^{(p+j)} ] = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix},$$

where  $\mathbf{A}$  and  $\mathbf{B}$  have  $8m$  and  $4m$  rows, respectively. By Step 2 in Section 2.2,

Since the absolute dot products of the two columns in these blocks are  $\leq 2$ , the absolute dot product of the two columns in  $\mathbf{A}$  is  $\leq 4m$ . Similarly,  $\mathbf{B}$  consists of  $m$  blocks with 4 rows and 2 columns

$$\begin{bmatrix} d_{8m+2l-1}^{(i)} & d_{8m+2l-1}^{(j)} \\ d_{8m+2l}^{(i)} & d_{8m+2l}^{(j)} \\ d_{8m+2l-1}^{(i)} & -d_{8m+2l-1}^{(j)} \\ d_{8m+2l}^{(i)} & -d_{8m+2l}^{(j)} \end{bmatrix}, \quad l = 1, \dots, m.$$



Since the absolute dot products of the two columns in these blocks are 0, the absolute dot product of the two columns in  $\mathbf{B}$  is 0. Thus, the absolute dot product of  $\mathbf{d}^{(i)}$  and  $\mathbf{d}^{(p+j)}$  is  $\leq 4m$  and their absolute column correlation is  $\leq 4m/12m = 1/3$ .

Combining (i) and (ii) proves that  $\mu(\mathbf{D}_2)$  is  $\leq 1/3$ .  $\square$

**Theorem 1.** *The coherence of  $\mathbf{D}_3$  constructed in Section 2.2 is no greater than  $1/3$ .*

*Proof.* Let  $\mathbf{d}^{(i)}$  denote the  $i$ th column of  $\mathbf{D}_3$  and  $d_j^{(i)}$  denote the  $j$ th entry of  $\mathbf{d}^{(i)}$ . Put

$$\mathbf{D}_3 = [ \mathbf{C}_1 \mid \mathbf{C}_2 \mid \mathbf{C}_3 \mid \mathbf{C}_4 ] = \left[ \begin{array}{c|c|c|c} \mathbf{U} & \mathbf{U} & \mathbf{U}^{**} & \mathbf{U}^{**} \\ \mathbf{U} & \mathbf{U}^* & \mathbf{U} & \mathbf{U}^* \\ \mathbf{L} & \mathbf{L} & \mathbf{L}^* & \mathbf{L}^* \\ \mathbf{L} & -\mathbf{L} & \mathbf{L}^{**} & \mathbf{L}^* \end{array} \right],$$

where  $\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3$  and  $\mathbf{C}_4$  are  $12m \times p$  matrices.

Pick two arbitrary columns of  $\mathbf{D}_3$ .

- (i) Suppose that both of them are from  $[\mathbf{C}_1, \mathbf{C}_2]$  or  $[\mathbf{C}_3, \mathbf{C}_4]$ . By Lemma 3, the coherence of  $[\mathbf{C}_1, \mathbf{C}_2]$  is  $\leq 1/3$ . By Lemmas 1 and 3, the coherence of  $[\mathbf{C}_3, \mathbf{C}_4]$  is  $\leq 1/3$ . Thus, the absolute column correlation of these two columns is  $\leq 1/3$ .
- (ii) Suppose that one of the two columns is the  $i$ th column  $\mathbf{d}^{(i)}$  of  $\mathbf{C}_1$  and the other column is the  $j$ th column  $\mathbf{d}^{(2p+j)}$  of  $\mathbf{C}_3$  with  $1 \leq i, j \leq p$ . Partition these columns as

$$[ \mathbf{d}^{(i)} \quad \mathbf{d}^{(2p+j)} ] = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix},$$

where  $\mathbf{A}$  and  $\mathbf{B}$  have  $8m$  and  $4m$  rows, respectively. By Step 3 in Section 2.2,

$$\mathbf{A} = \left[ \begin{array}{c|c|c|c|c|c} d_1^{(i)} & d_2^{(i)} & \dots & d_{4m-1}^{(i)} & d_{4m}^{(i)} & d_1^{(i)} \\ d_1^{(j)} & -d_2^{(j)} & \dots & d_{4m-1}^{(j)} & -d_{4m}^{(j)} & d_1^{(j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \dots & d_{10m-1}^{(i)} & d_{10m}^{(i)} & d_{8m+1}^{(i)} \\ -d_{8m+1}^{(j)} & d_{8m+2}^{(j)} & \dots & -d_{10m-1}^{(j)} & d_{10m}^{(j)} & -d_{8m+1}^{(j)} \end{array} \right]^T,$$

$$\mathbf{B} = \left[ \begin{array}{c|c|c|c|c|c} d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \dots & d_{10m-1}^{(i)} & d_{10m}^{(i)} & d_{8m+1}^{(i)} \\ -d_{8m+1}^{(j)} & d_{8m+2}^{(j)} & \dots & -d_{10m-1}^{(j)} & d_{10m}^{(j)} & -d_{8m+1}^{(j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \dots & d_{10m-1}^{(i)} & d_{10m}^{(i)} & d_{8m+1}^{(i)} \\ -d_{8m+1}^{(j)} & d_{8m+2}^{(j)} & \dots & -d_{10m-1}^{(j)} & d_{10m}^{(j)} & -d_{8m+1}^{(j)} \end{array} \right]^T.$$

Note that  $\mathbf{A}$  consists of  $2m$  blocks with 4 rows and 2 columns

$$\left[ \begin{array}{c|c} d_{2l-1}^{(i)} & d_{2l-1}^{(j)} \\ d_{2l}^{(i)} & -d_{2l}^{(j)} \\ \vdots & \vdots \\ d_{2l-1}^{(i)} & d_{2l-1}^{(j)} \\ d_{2l}^{(i)} & d_{2l}^{(j)} \end{array} \right], \quad l = 1, \dots, 2m.$$

Since the absolute dot products of the two columns in these blocks are  $\leq 2$ , the absolute dot product of the two columns in  $\mathbf{A}$  is  $\leq 4m$ . Similarly,  $\mathbf{B}$  consists of  $m$  blocks with 4 rows and 2 columns

$$\left[ \begin{array}{c|c} d_{8m+2l-1}^{(i)} & -d_{8m+2l-1}^{(j)} \\ d_{8m+2l}^{(i)} & d_{8m+2l}^{(j)} \\ \vdots & \vdots \\ d_{8m+2l-1}^{(i)} & d_{8m+2l-1}^{(j)} \\ d_{8m+2l}^{(i)} & -d_{8m+2l}^{(j)} \end{array} \right], \quad l = 1, \dots, m.$$

Since the absolute dot products of the two columns in these blocks are 0, the absolute dot product of the two columns in  $\mathbf{B}$  is 0. Thus, the absolute dot product of  $\mathbf{d}^{(i)}$  and  $\mathbf{d}^{(2p+j)}$  is  $\leq 4m$  and their absolute column correlation is  $\leq 4m/12m = 1/3$ .

- (iii) Suppose that one of the two columns is the  $i$ th column  $\mathbf{d}^{(p+i)}$  of  $\mathbf{C}_2$  and the other column is the  $j$ th column  $\mathbf{d}^{(3p+j)}$  of  $\mathbf{C}_4$  with  $1 \leq i, j \leq p$ . This case can be proved similarly to (ii).
- (iv) Suppose that one of the two columns is the  $i$ th column  $\mathbf{d}^{(i)}$  of  $\mathbf{C}_1$  and the other column is the  $j$ th column  $\mathbf{d}^{(3p+j)}$  of  $\mathbf{C}_4$  with  $1 \leq i, j \leq p$ . Partition these columns as

$$[ \mathbf{d}^{(i)} \quad \mathbf{d}^{(3p+j)} ] = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix},$$

where  $\mathbf{A}$  and  $\mathbf{B}$  have  $8m$  and  $4m$  rows, respectively. By Step 3 in Section 2.2,

$$\mathbf{A} = \left[ \begin{array}{c|c|c|c|c|c} d_1^{(i)} & d_2^{(i)} & \dots & d_{4m-1}^{(i)} & d_{4m}^{(i)} & d_1^{(i)} \\ d_1^{(j)} & -d_2^{(j)} & \dots & d_{4m-1}^{(j)} & -d_{4m}^{(j)} & d_1^{(j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \dots & d_{10m-1}^{(i)} & d_{10m}^{(i)} & d_{8m+1}^{(i)} \\ -d_{8m+1}^{(j)} & d_{8m+2}^{(j)} & \dots & -d_{10m-1}^{(j)} & d_{10m}^{(j)} & -d_{8m+1}^{(j)} \end{array} \right]^T,$$

$$\mathbf{B} = \left[ \begin{array}{c|c|c|c|c|c} d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \dots & d_{10m-1}^{(i)} & d_{10m}^{(i)} & d_{8m+1}^{(i)} \\ -d_{8m+1}^{(j)} & d_{8m+2}^{(j)} & \dots & -d_{10m-1}^{(j)} & d_{10m}^{(j)} & -d_{8m+1}^{(j)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{8m+1}^{(i)} & d_{8m+2}^{(i)} & \dots & d_{10m-1}^{(i)} & d_{10m}^{(i)} & d_{8m+1}^{(i)} \\ -d_{8m+1}^{(j)} & d_{8m+2}^{(j)} & \dots & -d_{10m-1}^{(j)} & d_{10m}^{(j)} & -d_{8m+1}^{(j)} \end{array} \right]^T.$$

Note that  $\mathbf{A}$  consists of  $2m$  blocks with 4 rows and 2 columns

$$\begin{bmatrix} d_{2l-1}^{(i)} & d_{2l-1}^{(j)} \\ d_{2l}^{(i)} & -d_{2l}^{(j)} \\ d_{2l-1}^{(i)} & -d_{2l-1}^{(j)} \\ d_{2l}^{(i)} & d_{2l}^{(j)} \end{bmatrix}, \quad l = 1, \dots, 2m.$$

Since the absolute dot products of the two columns in these blocks are 0, the absolute dot product of the two columns in  $\mathbf{A}$  is 0. Similarly,  $\mathbf{B}$  consists of  $m$  blocks with 4 rows and 2 columns

$$\begin{bmatrix} d_{8m+2l-1}^{(i)} & -d_{8m+2l-1}^{(j)} \\ d_{8m+2l}^{(i)} & d_{8m+2l}^{(j)} \\ d_{8m+2l-1}^{(i)} & -d_{8m+2l-1}^{(j)} \\ d_{8m+2l}^{(i)} & d_{8m+2l}^{(j)} \end{bmatrix}, \quad l = 1, \dots, m.$$

The absolute dot product of the two columns in the  $l$ th block is

$$2| -d_{8m+2l-1}^{(i)}d_{8m+2l-1}^{(j)} + d_{8m+2l}^{(i)}d_{8m+2l}^{(j)} |,$$

which is either 0 or 4. This implies that the absolute dot product of the two columns in  $\mathbf{B}$  is  $\leq 4m$ . Thus, the absolute dot product of  $\mathbf{d}^{(i)}$  and  $\mathbf{d}^{(3p+j)}$  is  $\leq 4m$  and their absolute column correlation is  $\leq 4m/12m = 1/3$ .

- (v) Suppose that one of the two columns is the  $i$ th column  $\mathbf{d}^{(p+i)}$  of  $\mathbf{C}_2$  and the other column is the  $j$ th column  $\mathbf{d}^{(2p+j)}$  of  $\mathbf{C}_3$  with  $1 \leq i, j \leq p$ . This case can be proved similarly to (iv).

Combining (i), (ii), (iii), (iv) and (v) proves that  $\mu(\mathbf{D}_3)$  is  $\leq 1/3$ . □

**Theorem 2.** *The coherence of  $\mathbf{D}_2$  constructed in Section 2.4 is no greater than  $t/n$ .*

*Proof.* To prove this theorem, we use the same notation from the proof of Lemma 3, and pick two arbitrary columns of  $\mathbf{D}_2$ .

- (i) Suppose that both of the two columns are from  $\mathbf{C}_1$  or  $\mathbf{C}_2$ . By a similar argument to Lemma 2, the coherence of  $\mathbf{C}_1$  is  $\leq t/n$ . Further, by Lemma 1, the coherence of  $\mathbf{C}_2$  is  $\leq t/n$ . Thus, the absolute column correlation of these two columns is  $\leq t/n$ .
- (ii) Suppose that one of the two columns is from  $\mathbf{C}_1$  and the other column is from  $\mathbf{C}_2$ . Since the matrices  $\mathbf{A}$  and  $\mathbf{B}$  have  $4t$  and  $2n-4t$  rows respectively, both of which are multiples of 4, a similar argument to Lemma 3 (ii) still works. Hence, the absolute column correlation of these two columns is  $\leq (4t/4 \times 2)/2n = t/n$ .

Combining (i) and (ii) proves that the coherence of  $\mathbf{D}_2$  is  $\leq t/n$ . □

## APPENDIX B. COHERENCE AND MODEL SELECTION CONSISTENCY OF THE LASSO

Zhao and Yu [34] studied the model selection consistency of the Lasso, i.e., the ability of the Lasso to exactly identify all the active factors from a large number of factors as the sample size  $n$  gets large. Assuming the model in (1.1), they defined the *Strong Irrepresentable Condition*, which requires the existence of a positive constant vector  $\boldsymbol{\eta}$  such that

$$|\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\text{sign}(\boldsymbol{\beta}_{(1)})| \leq \mathbf{1} - \boldsymbol{\eta},$$

where  $\mathbf{C}_{21} = \mathbf{X}(2)^\top \mathbf{X}(1)/n$ ,  $\mathbf{C}_{11} = \mathbf{X}(1)^\top \mathbf{X}(1)/n$  is invertible,  $\mathbf{X}(1)$  and  $\mathbf{X}(2)$  are the submatrices consisting of the columns of  $\mathbf{X}$  corresponding to the active and inactive factors respectively, the element-wise function  $\text{sign}(\cdot)$  maps zero to zero, positive numbers to +1 and negative numbers to -1,  $\boldsymbol{\beta}_{(1)}$  is the vector of coefficients of active factors,  $\mathbf{1}$  is a vector of +1's, and the inequality holds element-wise. They showed that for fixed  $p$  and  $\boldsymbol{\beta}$ , under regularity conditions

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \mathbf{X}^\top \mathbf{X} &= \mathbf{C}, \quad \mathbf{C} \text{ is a positive definite matrix,} \\ \lim_{n \rightarrow +\infty} \frac{1}{n} \max_{1 \leq i \leq n} \mathbf{x}_{(i)}^\top \mathbf{x}_{(i)} &= 0, \end{aligned}$$

where  $\mathbf{x}_{(i)}$  is the  $i$ th row of  $\mathbf{X}$ , if the Strong Irrepresentable Condition holds, then there exists a  $\lambda_n = f(n)$  such that

$$\lim_{n \rightarrow +\infty} P(\text{sign}(\hat{\boldsymbol{\beta}}(\lambda_n)) = \text{sign}(\boldsymbol{\beta})) = 1,$$

where  $\hat{\boldsymbol{\beta}}(\lambda_n)$  is the Lasso estimate. Further, in Corollary 2 of their paper, they showed that if the worst case column correlation is no greater than  $c/(2\|\boldsymbol{\beta}\|_0 - 1)$  for a constant  $0 \leq c < 1$ , then the Strong Irrepresentable Condition holds. In sum, a small worst case column correlation implies the Strong Irrepresentable Condition, which in turn guarantees the model selection consistency of the Lasso.

*Accepted 19 April 2023*

## REFERENCES

- [1] BERTSIMAS, D., KING, A. and MAZUMDER, R. Best subset selection via a modern optimization lens. *The Annals of Statistics* **44**(2) 813–852 (2016). <https://doi.org/10.1214/15-AOS1388>. MR3476618
- [2] BOOTH, K. H. and COX, D. R. Some systematic supersaturated designs. *Technometrics* **4**(4) 489–495 (1962). <https://doi.org/10.2307/1266285>. MR0184369
- [3] BOX, G. E. and MEYER, R. D. An analysis for unreplicated fractional factorials. *Technometrics* **28**(1) 11–18 (1986). <https://doi.org/10.2307/1269599>. MR0824728
- [4] BULUTOGLU, D. A. and CHENG, C. S. Hidden projection properties of some nonregular fractional factorial designs and their applications. *The Annals of Statistics* **31**(3) 1012–1026 (2003). <https://doi.org/10.1214/aos/1056562472>. MR1994740

- [5] CHAI, F. S., DAS, A., SINGH, R. and STUFKEN, J. Discriminating between superior  $UE(s^2)$ -optimal supersaturated designs. *Statistics and Applications* **18**(2) (2020).
- [6] CHEN, J. and LIN, D. K. On the identifiability of a supersaturated design. *Journal of Statistical Planning and Inference* **72**(1–2) 99–107 (1998). [https://doi.org/10.1016/S0378-3758\(98\)00025-1](https://doi.org/10.1016/S0378-3758(98)00025-1). MR1655185
- [7] DAVENPORT, M. A., DUARTE, M. F., ELДАР, Y. C. and KUTYNIOK, G. Introduction to Compressed Sensing. Citeseer (2012). <https://doi.org/10.1017/CBO9780511794308.002>. MR2963166
- [8] DENG, L. Y. and LIN, D. K. Criteria for supersaturated designs. In *Proceedings of the Section on Physical and Engineering Sciences* 124–128. American Statistical Association (1994).
- [9] DUMOUCHEL, W. and JONES, B. A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics* **36**(1) 37–47 (1994).
- [10] FAN, J. and LV, J. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society, Series B* **70**(5) 849–911 (2008). <https://doi.org/10.1111/j.1467-9868.2008.00674.x>. MR2530322
- [11] FAN, J., FENG, Y. and SONG, R. Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association* **106**(494) 544–557 (2011). <https://doi.org/10.1198/jasa.2011.tm09779>. MR2847969
- [12] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**(1) 1–22 (2010).
- [13] JMP (version 11). SAS Institute Inc., Cary, NC (1989–2007).
- [14] JONES, B., LIN, D. K. and NACHTSHEIM, C. J. Bayesian D-optimal supersaturated designs. *Journal of Statistical Planning and Inference* **138**(1) 86–92 (2008). <https://doi.org/10.1016/j.jspi.2007.05.021>. MR2369616
- [15] JONES, B. and MAJUMDAR, D. Optimal supersaturated designs. *Journal of the American Statistical Association* **109**(508) 1592–1600 (2014). <https://doi.org/10.1080/01621459.2014.938810>. MR3293612
- [16] JONES, B., LEKIVETZ, R., MAJUMDAR, D., NACHTSHEIM, C. J. and STALLRICH, J. W. Construction, properties, and analysis of group-orthogonal supersaturated designs. *Technometrics* **62**(3) 403–414 (2020). <https://doi.org/10.1080/00401706.2019.1654926>. MR4125505
- [17] LI, R. and LIN, D. K. Data analysis in supersaturated designs. *Statistics & Probability Letters* **59**(2) 135–144 (2002). [https://doi.org/10.1016/S0167-7152\(02\)00140-2](https://doi.org/10.1016/S0167-7152(02)00140-2). MR1927525
- [18] LIN, D. K. A new class of supersaturated designs. *Technometrics* **35**(1) 28–31 (1993).
- [19] LIN, D. K. Generating systematic supersaturated designs. *Technometrics* **37**(2) 213–225 (1995). <https://doi.org/10.2307/1269765>. MR1365722
- [20] LIU, Y., RUAN, S. and DEAN, A. M. Construction and analysis of  $E(s^2)$  efficient supersaturated designs. *Journal of Statistical Planning and Inference* **137**(5) 1516–1529 (2007). <https://doi.org/10.1016/j.jspi.2006.09.001>. MR2303773
- [21] MARLEY, C. J. and WOODS, D. C. A comparison of design and model selection methods for supersaturated experiments. *Computational Statistics & Data Analysis* **54**(12) 3158–3167 (2010). <https://doi.org/10.1016/j.csda.2010.02.017>. MR2727742
- [22] OSBORNE, M. R., PRESNELL, B. and TURLACH, B. A. Knot selection for regression splines via the lasso. *Computing Science and Statistics*. 44–49 (1998).
- [23] PLACKETT, R. L. and BURMAN, J. P. The design of optimum multifactorial experiments. *Biometrika* **33**(4) 305–325 (1946). <https://doi.org/10.1093/biomet/33.4.305>. MR0016624
- [24] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). <https://www.R-project.org/>.
- [25] SHEN, X., PAN, W., ZHU, Y. and ZHOU, H. On constrained and regularized high-dimensional regression. *Annals of the Institute of Statistical Mathematics* **65**(5) 807–832 (2013). <https://doi.org/10.1007/s10463-012-0396-3>. MR3105798
- [26] SINGH, R. and STUFKEN, J. Selection of two-level supersaturated designs for main effects models. *Technometrics* **65** 1–9 (2022). <https://doi.org/10.1080/00401706.2022.2102080>. MR4543063
- [27] TANG, B. and WU, C. F. J. A method for constructing supersaturated designs and its  $E(s^2)$  optimality. *Canadian Journal of Statistics* **25**(2) 191–201 (1997). <https://doi.org/10.2307/3315731>. MR1463319
- [28] TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* **58** 267–288 (1996). MR1379242
- [29] WEESE, M. L., SMUCKER, B. J. and EDWARDS, D. J. Searching for powerful supersaturated designs. *Journal of Quality Technology* **47**(1) 66–84 (2015).
- [30] WEESE, M. L., STALLRICH, J. W., SMUCKER, B. J. and EDWARDS, D. J. Strategies for supersaturated screening: Group orthogonal and constrained var (s) designs. *Technometrics* **63**(4) 443–455 (2021). <https://doi.org/10.1080/00401706.2020.1850529>. MR4331445
- [31] WELCH, L. R. Lower bounds on the maximum cross correlation of signals (corresp.). *IEEE Transactions on Information Theory* **20**(3) 397–399 (1974).
- [32] WU, C. F. J. Construction of supersaturated designs through partially aliased interactions. *Biometrika* **80**(3) 661–669 (1993). <https://doi.org/10.1093/biomet/80.3.661>. MR1248029
- [33] WU, C. F. J. and HAMADA, M. S. *Experiments: planning, analysis, and optimization* 2nd ed. John Wiley & Sons, Hoboken, NJ (2009). MR2583259
- [34] ZHAO, P. and YU, B. On model selection consistency of Lasso. *Journal of Machine Learning Research* **7** 2541–2563 (2006). MR2274449

Youran Qi. Department of Statistics, University of Wisconsin-Madison, 1300 University Ave, Madison, WI 53706, USA. E-mail address: [yqi28@wisc.edu](mailto:yqi28@wisc.edu)

Peter Chien. Department of Statistics, University of Wisconsin-Madison, 1300 University Ave, Madison, WI 53706, USA. E-mail address: [peter.chien@wisc.edu](mailto:peter.chien@wisc.edu)