MINIMAL DEPENDENT SETS FOR EVALUATING SUPERSATURATED DESIGNS

Arden Miller and Boxin Tang

University of Auckland and Simon Fraser University

Abstract: This paper investigates minimal dependent sets used for evaluating supersaturated designs. Unlike the popular $E(s^2)$ criterion, the criteria based on minimal dependent sets directly capture the properties of designs in terms of estimation and identification of active factors. This paper provides a theoretical investigation into the number and structure of minimal dependent sets in a supersaturated design, and presents some construction results on supersaturated designs with large minimal dependent sets.

Key words and phrases: $E(s^2)$-optimality, estimation capacity, model discrimination, screening design.

1. Introduction

At the outset of a systematic investigation of a process, a screening experiment may be used to sift through a set of candidate factors to find those that have an important impact on the response. As subsequent experiments will focus on these factors, the primary purpose of a screening experiment is model identification rather than model estimation. A good screening design is one that allows the investigator to consider a large number of factors in a relatively small number of runs. The question of how many candidate factors can be screened using a design of a given size arises. If the purpose of an experiment is estimation, then the answer is straightforward: assuming that each factor has two levels and there are no interactions between factors, then main effects for up to $n-1$ factors can be estimated using an n-run experiment. However for screening applications, Booth and Cox (1962) and Lin (1993) raised the possibility that more than $n-1$ factors can be considered in an n-run experiment. The key difference is that for screening applications the investigator may be willing to place a cap on the number of active factors. Thus it is not necessary to simultaneously estimate the effects of all of the candidate factors and the possibility of having the number m of factors exceed $n-1$ exists. Such designs are said to be supersaturated. Supersaturated designs are special fractional factorial designs. For a general discussion on the construction and optimality of fractional factorial designs, we refer the reader to Dey and Mukerjee (1999).

The literature on supersaturated designs is quite rich – for details see Wu (1993), Nguyen (1996), Tang and Wu (1997), Cheng (1997), Butler et al. (2001), Fang, Lin, and Liu (2003), Liu and Dean (2004), Bulutoglu and Cheng (2004) and Xu and Wu (2005). The criterion used most often to evaluate supersaturated designs is $E(s^2)$ which represents the average of the cross products between all possible pairs of columns. It can be thought of as a measure of nonorthogonality – clearly a supersaturated design of n runs cannot be orthogonal because the number m of factors is greater than $n - 1$. This criterion is easy to use and makes intuitive sense, and has played a very important role in the studies of supersaturated designs. However, $E(s^2)$ does not directly measure the ability of a design to screen for active factors.

This paper considers the use of minimal dependent sets (MDS's) to evaluate the screening capability of supersaturated designs. Unlike $E(s^2)$, there is a direct connection between criteria based on MDS's and the ability of a design to identify active factors from a candidate set. Minimal dependent sets were first introduced by Miller and Sitter (2004), and later used by Lin, Miller, and Sitter (2008), to assess nonorthogonal foldover designs. Similar to what is done in Lin, Miller, and Sitter (2008), MDS-resolution and MDS-aberration criteria can be defined for supersaturated designs. The present article provides an investigation into the theoretical properties of MDS's in the context of supersaturated designs. These properties were utilized to design an extensive computer search for MDS-aberration optimal supersaturated designs – the results of this search will be reported in a separate paper (Miller and Tang (2011)).

In Section 2, we define the concepts of MDS's, MDS-resolution and MDS-aberration, and some general theoretical results concerning MDS's are presented. In Section 3, results on two-level balanced supersaturated designs are reported. In particular, a supersaturated design of n runs for n factors that has the maximum MDS-resolution is constructed. Section 4 concludes the paper with a discussion.

2. General Results on Minimal Dependent Sets

2.1 Minimal dependent sets

To evaluate the ability of a design to differentiate between competing models, Miller and Sitter (2004) proposed looking at the MDS's of the column vectors. An MDS is defined as a set of column vectors that are linearly dependent but if any one of them is removed the resulting subset becomes linearly independent. Lin, Miller, and Sitter (2008) further defined the criteria of MDS-resolution and MDS-aberration, and constructed a catalog of nonorthogonal foldover designs based on these criteria. The same concepts can also be used to evaluate supersaturated designs.

We focus on two-level supersaturated designs in this paper. A supersaturated design with two levels is represented by an $n \times m$ matrix X of ± 1, where the number m of factors is no smaller than the run size n. We consider even n. Each column of a design has the same number of ± 1 to ensure that the main effects are orthogonal to the grand mean. Because $m \geq n$, the m columns of design X cannot be linearly independent. A dependent subset of columns of design X can cause problems in differentiating between certain models. MDS's provide a concise way of capturing all the dependent relationships among the columns of a design as any dependent subset can be obtained by adding columns to an MDS (Lemma 2 of Section 2.2). The size of an MDS is defined to be the number of columns it contains. For design X, let A_j be the number of MDS's of size j. The MDS-resolution is defined as the size of the smallest MDS. The criterion of MDS-aberration selects a design by sequentially minimizing the components $A_1, \ldots, A_m$ in the MDS word length pattern $(A_1, \ldots, A_m)$.

2.2. Some general results

Let $V_1, \ldots, V_m$ be a set of m column vectors whose elements are real numbers. A lemma provides a characterization of when these column vectors form an MDS.

Lemma 1. *The set of vectors $V_1, \ldots, V_m$ is an MDS if and only if V_m can be written as*

$$V_m = a_1 V_1 + \cdots + a_{m-1} V_{m-1},$$

where the coefficients $a_1, \ldots, a_{m-1}$ are non-zero real numbers and form a unique set.

The proof is simple, thus omitted. In Lemma 1, there is nothing special about V_m and the result holds true if V_m is replaced by any other V_j.

Lemma 2. *Every dependent set of vectors $V_1, \ldots, V_m$ contains at least one MDS.*

Let r be the rank of the set of vectors $V_1, \ldots, V_m$ where $r < m$ since the set is dependent. Assume that $V_1, \ldots, V_r$ are independent. Then V_m can be uniquely written as $V_m = a_1 V_1 + \cdots + a_r V_r$. Dropping those V_j's with $a_j = 0$, produces an MDS by Lemma 1.

Two questions arise: how many MDS's exist and what are their structures? The rest of this section addresses them. First, we show that every linear dependency is determined by those that correspond to the MDS's. Again let r denote the rank of the vectors $V_1, \ldots, V_m$ and assume that $V_1, \ldots, V_r$ are independent. To facilitate the discussion, let $U_1, \ldots, U_s$ denote $V_{r+1}, \ldots, V_m$. Therefore the set of dependent vectors is now represented as $\{V_1, \ldots, V_r, U_1, \ldots, U_s\}$ where

$s = m - r \geq 1$. Each U_j can be uniquely written as a linear combination of $V_1, \ldots, V_r$:

$$U_i = a_{i1}V_1 + \cdots + a_{ir}V_r \tag{2.1}$$

for $i = 1, \ldots, s$. It will be seen that the matrix $A = [a_{ij}]_{s \times r}$ contains all the information regarding the MDS's in the set of dependent vectors $V_1, \ldots, V_r, U_1, \ldots, U_s$.

Consider the set of all row vectors $(a_1, \ldots, a_r, a'_1, \ldots, a'_s)$ that satisfy $a_1 V_1 + \cdots + a_r V_r + a'_1 U_1 + \cdots + a'_s U_s = 0$. This set forms an s-dimensional linear subspace with one basis given by the rows of the $s \times (r+s)$ matrix $(A, -I)$ where $A = [a_{ij}]_{s \times r}$ with a_{ij} from (2.1), and I the identity matrix of order s. This leads to the following conclusion.

Lemma 3. *All linear relations among the set of vectors $\{V_1, \ldots, V_r, U_1, \ldots, U_s\}$ can be obtained by linearly combining those in (2.1), and thus are of the form*

$$\sum_{i=1}^{s} c_i U_i = \left(\sum_{i=1}^{s} c_i a_{i1} \right) V_1 + \cdots + \left(\sum_{i=1}^{s} c_i a_{ir} \right) V_r, \tag{2.2}$$

for real coefficients $c_1, \ldots, c_s$.

We are ready to examine the number and structure of MDS's for a given dependent set. First consider the case where $s = 1$. If the terms with zero coefficients are dropped from $U_1 = a_{11}V_1 + \cdots + a_{1r}V_r$ in (2.1), we obtain an MDS by Lemma 1. Lemma 3 confirms that this is the only linear relation (up to a constant) among the vectors $V_1, \ldots, V_r, U_1$. Therefore, there exists exactly one MDS when $s = 1$. For $s = 2$, a complete characterization on the number and structures of the MDS's can also be obtained. In order not to disrupt the flow of the paper, the results and their derivations are presented in Appendix A.

We could proceed to provide a complete characterization on the number and structures of the MDS's given a dependent set of $r + s$ vectors for $s = 3$, $4 \ldots$ as is given by Lemmas A1 and A2 for $s = 2$ in Appendix A. We choose not to do so because the presentation of such results would inevitably get very complicated without providing further insight. Instead, we present the following general version of Lemmas A1 and A2.

Theorem 1. *Suppose that a dependent set of $r + s$ vectors has rank r. Let an independent subset of r vectors be $V_1, \ldots, V_r$, and the remaining s vectors $U_1, \ldots, U_s$. Each U_i can be uniquely written as*

$$U_i = a_{i1}V_1 + \cdots + a_{ir}V_r, \ \text{for } i = 1, \ldots, s. \tag{2.3}$$

In each expression drop the terms with zero coefficients to form an MDS and call these s basic MDS's. Let $A = [a_{ij}]$. Then

(i) *the number of MDS's is bounded below by s and is bounded above by $\binom{r+s}{r+1}$;*

(ii) *the lower bound is attained if and only if the s basic MDS's are mutually exclusive;*

(iii) *the upper bound is attained if and only if every $t \times t$ submatrix of A has full rank for each $t = 1, \ldots, \min(r, s)$, in which case every subset of $r + 1$ vectors is an MDS.*

Theorem 1 can be proved in a manner similar to that of establishing Lemmas A1 and A2 in Appendix A, and we omit the details. In addition to providing theoretical insights, the results in this section are practically useful as well. Illustrations are given in Section 3. These general results provide useful tools in the search for supersaturated designs based on the criteria of MDS-resolution and MDS-aberration.

3. Results for Balanced Two-Level Designs

This section presents results for balanced two-level supersaturated designs, which are restricted to those where no column is fully aliased with another column – i.e. no two columns are either identical or mirror images. Thus any two columns are independent, implying that the minimum size of an MDS is at least three. In fact, a stronger result holds.

Lemma 4. *The MDS-resolution of a balanced two-level supersaturated design is at least four.*

To the contrary, suppose three columns V_1, V_2, V_3 constitute an MDS. By Lemma 1, $V_3 = a_1 V_1 + a_2 V_2$ where a_1, a_2 are unique and nonzero. For balanced V_1 and V_2 that are not fully aliased with each other, each of the four possible pairs $(1, 1)$, $(1, -1)$, $(-1, 1)$ and $(-1, -1)$ must occur at least once in the $n \times 2$ matrix (V_1, V_2). The corresponding entries of V_3 for these pairs are $a_1 + a_2$, $a_1 - a_2$, $-a_1 + a_2$, $-a_1 - a_2$, of which at least three must be distinct because a_1 and a_2 are not zero. This contradicts the fact that V_3 has only two possible entries ± 1 and thus Lemma 4 is established.

Lemma 5. *For any even $n \geq 6$, there exists a balanced two-level supersaturated design with MDS-resolution 4.*

The result follows if we can construct four columns that form an MDS. For

$n = 6$, take

$$\begin{bmatrix} 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix},$$

as the sum of these four columns is zero. For $n \geq 8$, simply add $(n-6)/2$ rows of $(1, 1, -1, -1)$ and $(n-6)/2$ rows of $(-1, -1, 1, 1)$.

One attractive way to construct supersaturated designs is to add balanced columns to a saturated two-level orthogonal array – designs created in this manner have been investigated in Deng, Lin, and Wang (1996) and in Yamada and Lin (1997). First, consider adding an *interaction* column to a saturated orthogonal array and note that a saturated orthogonal array of n runs has $n - 1$ factors.

Lemma 6. *Consider a supersaturated design of n runs for n factors obtained by adding an interaction column to a saturated orthogonal array. Then we have*

(i) *the MDS-resolution of this design is $n - 2$ or lower;*

(ii) *the MDS-resolution is exactly $n - 2$ if n is not a multiple of 8.*

Let $V_1, \ldots, V_{n-1}$ be the columns of the saturated orthogonal array, and let $U_1 = V_1 V_2$ denote the added interaction column. Because $V_1, \ldots, V_{n-1}$ form an orthogonal basis in the $n - 1$ dimensional space that is orthogonal to the column of all plus ones, we have

$$U_1 = a_1 V_1 + \cdots + a_{n-1} V_{n-1}, \text{ where } a_j = U_1^T \frac{V_j}{n}. \tag{3.1}$$

Since $U_1 = V_1 V_2$ is the interaction column of V_1 and V_2, we have $a_1 = a_2 = 0$. Dropping the terms with $a_j = 0$ in (3.1) gives an MDS, the size of which is at most $n - 2$. If n is not a multiple of 8, we have that $a_j \neq 0$ for $j = 3, \ldots, n - 1$ based on a result in Deng and Tang (2002, Prop. 1). Therefore the MDS obtained from (3.1) is of size $n - 2$.

For a supersaturated design of n runs for n factors, the best possible scenario is to have a single MDS of size n that gives an MDS-resolution of n. Lemma 6 indicates that such a design cannot be constructed by adding an interaction column to a saturated orthogonal array. We have explored the possibility of obtaining MDS-resolution n designs by adding a balanced column that is not an interaction column to a saturated orthogonal array. The results are mixed. We can prove that this is not possible for $n \leq 16$, a result confirmed by an exhaustive computer search. We have also conducted an exhaustive search for $n = 20$, and very extensive searches for $n = 24, 28$ and found no such designs. Somewhat

surprisingly, a computer search for $n = 32$ did find such an MDS-resolution n design. We postulate that for sufficiently large n it is possible to create an MDS-resolution n design of n runs for n factors by adding a balanced column to a saturated orthogonal array.

Better results are obtained if the restriction of starting with a saturated orthogonal array is lifted. In this case, a balanced supersaturated design of n runs for n factors with MDS-resolution n can be constructed for any even $n \geq 6$. The rest of this section presents a construction of such designs. To illustrate the structure of such designs, we present the design for $n = 10$:

$$X = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 \end{bmatrix}.$$

The construction can be described in three steps, and works for any even $n \geq 6$:

Step 1. Let the $(n-1) \times (n-1)$ matrix

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

have A as an $(n/2 - 1) \times n/2$ matrix of all 1's, B as an $(n/2 - 1) \times (n/2 - 1)$ matrix with diagonal elements -1 and off-diagonal elements 1, C as an $n/2 \times n/2$ matrix with diagonal elements 1 and off-diagonal elements -1, and D as an $n/2 \times (n/2 - 1)$ matrix with the entries in the first two rows all 1 and all other entries -1.

Step 2. Add a row of -1's as the top row to obtain an $n \times (n-1)$ matrix $X_0 = (V_1, \ldots, V_{n-1})$, where V_j is the jth column of this matrix.

Step 3. Final design X is $(V_1, \ldots, V_{n-1}, U)$, obtained by adding a column vector U to X_0, where U has its first $n/2$ entries -1 and the other $n/2$ entries 1.

Theorem 2. *The n-run design X for n factors constructed above has MDS-resolution n.*

Proof. It is obvious that all columns of X are balanced. Theorem 2 follows from Lemma 1 if we can prove that (i) the first $n-1$ columns $V_1, \ldots, V_{n-1}$ of

1280 ARDEN MILLER AND BOXIN TANG

design X are linearly independent, and (ii) none of the coefficients in the linear combination $U = \sum_{j=1}^{n-1} a_j V_j$ is zero.

Columns $V_1, \ldots, V_{n-1}$ are linearly independent if and only if matrix $X_0 = (V_1, \ldots, V_{n-1})$ has rank $n-1$. Equivalently, we show that the matrix X_1 obtained by adding a top row of all -1's to

$$\begin{bmatrix} C & D \\ A & B \end{bmatrix},$$

has full rank $n-1$. For $i = 2, \ldots, (n/2+1)$, subtract the first row of X_1 from the ith row and, for $i = (n/2+2), \ldots, n$, add the first row to the ith row. Omitting the first row, the resulting matrix is an upper triangle matrix with diagonal elements ± 2. This shows that matrix X_1 has rank $n-1$. For $n = 10$, the upper triangle matrix is

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 \\ 0 & 2 & 0 & 0 & 0 & 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \end{bmatrix}.$$

Because $V_1, \ldots, V_{n-1}$ are linearly independent, column U can be uniquely written as

$$U = a_1 V_1 + \cdots + a_{n-1} V_{n-1}. \tag{3.2}$$

We show that $a_j = -(n/2 - 2)$ for $j = 1, 2$ and $a_j = 1$ for $j = 3, \ldots, n-1$. Let $(v_1, \ldots, v_{n-1}, u)$ denote a row of design X. We then need to show that

$$u = -(\frac{n}{2} - 2)v_1 - (\frac{n}{2} - 2)v_2 + v_3 + \cdots + v_{n-1}, \tag{3.3}$$

is true for every row $(v_1, \ldots, v_{n-1}, u)$ of design X. This can be verified separately for (a) row 1, (b) rows 2 to $n/2$, (c) rows $n/2+1$ and $n/2+2$ and (d) rows $n/2+3$ to n. All verifications are straightforward and we only give some details here. For row 1, (3.3) easily follows from the fact that $v_1 = \cdots = v_{n-1} = u = -1$. For row 2, we have that $u = -1$, $v_j = -1$ for $j = n/2 + 1$ and $v_j = 1$ for all other j's. For row $n/2 + 1$, we have that $u = 1$, $v_1 = 1$, $v_j = -1$ for $j = 2, \ldots, n/2$ and $v_j = 1$ for $j = n/2 + 1, \ldots, n - 1$. Since $n \geq 6$, the coefficient a_j in (3.2) is nonzero for every j. Theorem 2 is established.

Consider a supersaturated design of n runs for n factors, obtained by adding one balanced column to a saturated orthogonal array. According to Cheng (1997), such a design is $E(s^2)$ optimal with $E(s^2) = 2.29, 2.18, 2.13$ for $n = 8, 12, 16$, respectively. The MDS-resolution of such a design is 5, 10, and 11 for $n = 8, 12$, and 16, respectively. The design in Theorem 2 has the maximum MDS-resolution of n but its performance in terms of $E(s^2)$ is not ideal, with $E(s^2) = 10.86, 45.82, 111.33$ for $n = 8, 12, 16$, respectively. In view of these sharp contrasts, it is of great interest to construct designs that, though not optimal, perform well under both the MDS-resolution and $E(s^2)$ criteria.

4. Discussion and Concluding Remarks

This paper focuses on the relationship between the ability of a supersaturated design to differentiate between alternative models and the MDS's of the columns of its design matrix. If the combined set of columns for two competing models form a linearly dependent set, problems can arise in differentiating between the models. Thus the ability of a design to differentiate between models is directly related to the linear dependencies of its design matrix. Each dependent set (DS) of columns must either be an MDS or contain one. As a result, the set of MDS's for a design determine the set of DS's and it is sufficient to study the (much smaller) set of MDS's. For a supersaturated design, there is a direct connection between the structures and number of its MDS's and its ability to differentiate between models. It follows that the MDS-resolution and MDS-aberration criteria are directly related to the capability of a design to identify the active factors. Note that such a connection does not exist for the $E(s^2)$ criterion.

There are clear connections between the MDS's and two other criteria that have been used to evaluate supersaturated designs: estimation capacity (Cheng, Steinberg, and Sun (1999)) and resolution rank (Deng, Lin, and Wang (1999)). These connections are obvious if we consider the DS's in a supersaturated design and note that a dependent set (DS) of columns corresponds to a non-estimable model. Define DS-resolution and DS-aberration as follows: If B_j represents the number of DS's of size j in a supersaturated design, then the DS-resolution is the smallest j such that $B_j \neq 0$, and the DS-aberration criterion selects designs by sequentially minimizing $B_1, B_2, \ldots$. Clearly, sequentially minimizing $B_1, B_2, \ldots$ is equivalent to sequentially maximizing $E_1, E_2, \ldots$, where E_j is the proportion of estimable models of size j – this is the estimation capacity criterion used by Cheng, Steinberg, and Sun (1999). Also note that if a design has DS-resolution j then all subsets of $j - 1$ columns are linearly independent. Thus maximizing DS-resolution is equivalent to maximizing c, where c is the largest value such that all sets of columns of size c are linearly independent – this is the resolution rank criterion used by Deng, Lin, and Wang (1999).

Note that DS-resolution is identical to MDS-resolution since the smallest DS must be an MDS. However, it is difficult (and may be impossible) to establish an exact equivalence between DS-aberration and MDS-aberration. As in Section 2, let A_j be the number of MDS's of size j. Then the MDS-aberration criterion sequentially minimizes $A_1, A_2, \ldots$. If a supersaturated design has MDS-resolution r, then we must have $A_j = B_j$ for $j = 1, \ldots, r$, and $A_j < B_j$ for $j \geq r + 1$. Here $A_j < B_j$ occurs because B_j is A_j plus the number DS's of size j that can be created by adding columns to an MDS of size $< j$, and for $j \geq r + 1$ there must be some such DS's. The difficulty in establishing an exact equivalence between DS-aberration and MDS-aberration arises because a DS may contain more than one MDS. Thus the possibility of finding two designs that have identical A_j's for all j but non-identical B_j's appears to exist. However, we believe that such examples are rare (if they exist at all), and thus DS-aberration and MDS-aberration are (at least) near equivalent criteria.

MDS's in a supersaturated design capture the dependent relations among its columns. One may wish to supplement this assessment of a supersaturated design with an evaluation of the extent of these dependent relations. Some of the ways in which this can be done can be found in Jones et al. (2007) and Cheng, Deng, and Tang (2002).

Acknowledgement

Appendix A. MDS's in $r + 2$ Vectors with Rank r

For $s = 2$, consider the U_1 and U_2 as in (2.1). Dropping the terms with zero coefficients from $U_1 = a_{11}V_1 + \cdots + a_{1r}V_r$ gives one MDS containing U_1 but not U_2; it is the only MDS for the set $\{U_1, V_1, \ldots V_r\}$. Doing the same for $U_2 = a_{21}V_1 + \cdots + a_{2r}V_r$ gives an MDS that contains U_2 but not U_1; it is the only MDS for the set $\{U_2, V_1, \ldots V_r\}$. For convenience, these two MDS's are referred to as the basic MDS's. Now consider the question: are there any MDS's that contain both U_1 and U_2? A complete answer can be given. Without loss of generality, the V_j's can be labeled such that

$$U_1 = a_1 V_1 + \cdots + a_p V_p, \qquad U_2 = b_q V_q + \cdots + b_k V_k, \qquad (A.1)$$

where the a_j's and the b_j's are all non-zero, $1 \leq p$ and $q \leq k \leq r$.

If $p < q$, then the two basic MDS's $\{U_1, V_1, \ldots, V_p\}$ and $\{U_2, V_q, \ldots, V_k\}$ are disjoint and no MDS containing both U_1 and U_2 exists. To see this, note that

applying Lemmas 1 and 3 when $p < q$ dictates that any MDS that contains U_1 and U_2 must contain all of $V_1, \ldots, V_p, V_q, \ldots, V_k$ as well. Thus it would contain both basic MDS's as proper subsets, violating the definition of an MDS.

If $p \geq q$, $\{U_1, V_1, \ldots, V_p\}$ and $\{U_2, V_q, \ldots, V_k\}$ have an overlap of the vectors $V_q, \ldots, V_p$, and it can be shown that an MDS that contains both U_1 and U_2 exists. By Lemmas 1 and 3, to obtain an MDS containing both U_1 and U_2, we only need to consider the linear relation $U_1 = -cU_2 + a_1V_1 + \cdots + a_pV_p + c(b_qV_q + \cdots + b_kV_k)$ for nonzero c. Rewrite this relation as

$$U_1 = -cU_2 + \sum_{j=1}^{q-1} a_jV_j + c\left(\sum_{j=p+1}^{k} b_jV_j\right) + \sum_{j=q}^{p}(a_j + cb_j)V_j. \qquad \text{(A.2)}$$

Setting $c = -a_q/b_q$ eliminates V_q in (A.2). Note that this choice results in non-zero coefficients for U_2 and for all V_j's where $j = 1, \ldots, q-1$ and $j = p+1, \ldots, k$. For the V_j's where $j = q+1, \ldots, p$, the coefficients are zero if and only if $a_j/b_j = a_q/b_q$. Since $c = -a_q/b_q$ is the only choice that will eliminate V_q in (A.2), the set of non-zero coefficients must be unique and by Lemma 1, we have identified an MDS that contains both U_1 and U_2; it also contains all V_j's except V_q and, for $j = q+1, \ldots, p$, those V_j's with $a_j/b_j = a_q/b_q$.

It is straightforward to expand this result to identify the number of MDS's that contain both U_1 and U_2. Let g be the number of distinct values in the sequence $\beta_q, \ldots, \beta_p$ where $\beta_j = a_j/b_j$. Corresponding to these g distinct values, the set of vectors $V_q, \ldots, V_p$ is partitioned into $g \geq 1$ subsets of vectors. Let $S_1, \ldots, S_g$ denote these subsets.

Lemma A.1. If the two basic MDS's in (A.1) overlap with each other, then there are precisely g MDS's containing both U_1 and U_2. The ith of these can be obtained by deleting the vectors in S_i from the set $U_1, U_2, V_1, \ldots, V_k$ for $i = 1, \ldots, g$.

The maximum value that g can take is r, which occurs when $q = 1$ and $p = k = r$, and the β_j's are all distinct. In this case, every subset of $r+1$ vectors in the set of vectors $\{V_1, \ldots, V_r, U_1, U_2\}$ is an MDS.

Lemma A.2. Suppose that a dependent set of $r+2$ vectors has rank r. Then the number of MDS's can be any integer between 2 and $r+2$. This number is 2 if and only if the two basic MDS's are disjoint, and is $r+2$ if and only if every subset of $r+1$ vectors is an MDS.

References

Booth, E. H. and Cox, D. R. (1962). Some systematic supersaturated designs. *Technometrics* **4**, 489-495.

Bulutoglu, D. A. and Cheng, C. S. (2004). Construction of $E(s^2)$-optimal supersaturated designs. *Ann. Statist.* **32**, 1662-1678.

Butler, N. A., Mead, R., Eskridge, K. M. and Gilmour, S. G. (2001). A general method of constructing $E(s^2)$-optimal supersaturated designs. *J. Roy. Statist. Soc. Ser. B* **63**, 621-632.

Cheng, C. S. (1997). $E(s^2)$-optimal supersaturated designs. *Statist. Sinica* **7**, 929-939.

Cheng, C. S., Deng, L. Y. and Tang, B. (2002). Generalized minimum aberration and design efficiency for nonregular fractional factorial designs. *Statist. Sinica* **12**, 991-1000.

Cheng, C. S., Steinberg, D. M. and Sun, D. X. (1999). Minimum aberration and model robustness for two level fractional factorial designs. *J. Roy. Statist. Soc. Ser. B* **61**, 85-93.

Deng, L. Y., Lin, D. K. J. and Wang, J. (1996). Marginally oversaturated designs. *Comm. Statist. Theory Methods* **25**, 2557-2573.

Deng, L. Y., Lin, D. K. J. and Wang, J. (1999). A resolution rank criterion for supersaturated designs. *Statist. Sinica* **9**, 605-610.

Deng, L. Y. and Tang, B. (2002). Design selection and classification for Hadamard matrices using generalized minimum aberration criteria. *Technometrics* **44**, 173-184.

Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans.* Wiley, New York.

Fang, K. T., Lin, D. K. J. and Liu, M. Q. (2003). Optimal mixed-level supersaturated design. *Metrika* **58**, 279-291.

Jones, B. A., Li, W., Nachtsheim, C. J. and Ye, K. Q. (2007). Model discrimination - another perspective on model-robust designs. *J. Statist. Plann. Inference* **137**, 1576-1583.

Lin, C. D., Miller, A. E. and Sitter, R. R. (2008). Folded over non-orthogonal designs. *J. Statist. Plann. Inference* **138**, 3107-3124.

Lin, D. K. J. (1993). A new class of supersaturated designs. *Technometrics* **35**, 28-31.

Liu, Y. and Dean, A. M. (2004). *k*-circulant supersaturated designs. *Technometrics* **46**, 32-43.

Miller, A. E. and Sitter, R. R. (2004). Choosing columns from the 12-run Plackett-Burman design. *Statistics and Probability Letters* **67**, 193-201.

Miller, A. E. and Tang, B. (2011). Finding MDS-optimal supersaturated designs using computer searches. In preparation.

Nguyen, N. K. (1996). An algorithmic approach to constructing supersaturated designs. *Technometrics* **38**, 69-73.

Tang, B. and 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.

Yamada, S. and Lin, D. K. J. (1997). Supersaturated design including an orthogonal base. *Canadian Journal of Statistics* **25**, 203-213.

Xu, H. and Wu, C.F.J. (2005). Construction of optimal multi-level supersaturated designs. *Ann. Statist.* **33**, 2811-2836.

Department of Statistics, University of Auckland, Auckland, New Zealand.

E-mail: miller@stat.auckland.ac.nz

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.

E-mail: boxint@sfu.ca