

**EXISTENCE AND CONSTRUCTION
OF TWO-LEVEL ORTHOGONAL ARRAYS
FOR ESTIMATING MAIN EFFECTS AND
SOME SPECIFIED TWO-FACTOR INTERACTIONS**

Boxin Tang and Julie Zhou

Simon Fraser University and University of Victoria

Abstract: This paper considers two-level orthogonal arrays that allow joint estimation of all main effects and a set of prespecified two-factor interactions. We obtain some theoretical results that provide a simple characterization of when such designs exist, and how to construct them if they do. General as well as concrete applications of the results are discussed.

Key words and phrases: Clear two-factor interaction, compromise plan, nonregular factorial design, requirement set, robust parameter design.

1. Introduction

We discuss two-level factorial designs that allow joint estimation of all main effects and selected two-factor interactions (2fi's). Consideration of such a problem in the context of regular factorial designs dates back to Addelman (1962). Greenfield (1976) used the concept of requirement set to denote a set of effects the experimenter is interested in estimating. This line of research was further pursued by Franklin and Bailey (1977), Wu and Chen (1992), Dey and Suen (2002), Ke and Tang (2003), and Cheng and Tang (2005). Without restricting to regular factorial designs, Hedayat and Pesotan (1992) investigated the existence and construction of saturated designs for main effects and selected 2fi's. Earlier, Rechtschaffner (1967) provided a simple construction of saturated designs for all main effects and all 2fi's. Cheng (2003) showed that this idea of construction also works for a general requirement set provided that the effects in the requirement set have a nested structure. In general, the saturated designs in Rechtschaffner (1967), Hedayat and Pesotan (1992), and Cheng (2003) are not orthogonal arrays. One situation where certain 2fi's are of importance is the setting of robust parameter design. Designs suitable for such a situation include compound arrays as discussed in Rosenbaum (1996) and Hedayat and Stufken (1999), and combined arrays selected via some aberration criteria as done in Wu and Zhu (2003) and Ke and Tang (2003). The work on clear 2fi's goes further than that on the

requirement set. We refer the reader to Wu and Chen (1992), Chen and Hedayat (1998), Wu and Hamada (2000), Wu and Wu (2002) and Tang (2006) for details.

This paper presents some characterizing results on the existence and construction of orthogonal arrays for jointly estimating all main effects and some specified 2fi's. Our characterization is simple to use, yet powerful enough to allow many useful designs to be found. If a regular factorial is found for a given requirement set, it achieves full efficiency; otherwise, one has to consider designs with at least doubled run sizes. In contrast, saturated designs of Rechtschaffner (1967) and Hedayat and Pesotan (1992) are most economical but suffer low efficiency. Orthogonal arrays provide a compromise. The run sizes of two-level orthogonal arrays are multiples of four, leaving only small gaps between adjacent run sizes. Orthogonal arrays at least guarantee that all main effects are mutually orthogonal, making them more efficient than the designs of Rechtschaffner (1967) and Hedayat and Pesotan (1992), and much more so if there are only a small number of 2fi's in the requirement set.

2. Characterizing Results

We use an $n \times m$ matrix of ± 1 to denote a two-level factorial design of n runs for m factors. Such a design is an orthogonal array of strength t if in each $n \times t$ submatrix, the 2^t level combinations occur with the same frequency. If we speak of an orthogonal array without specifying its strength, we mean that the strength is at least two. When $n = m + 1$, the orthogonal array is called saturated. The existence of a saturated orthogonal array of size n is equivalent to the existence of a Hadamard matrix of order n . A Hadamard matrix is a square orthogonal matrix of ± 1 . The order of a Hadamard matrix is necessarily equal to 1, 2, or a multiple of 4. For a general discussion on orthogonal arrays and Hadamard matrices, we refer to Hedayat, Sloane and Stufken (1999). Our attention in this paper is paid to orthogonal arrays that can be obtained by selecting columns from saturated orthogonal arrays. Designs from saturated orthogonal arrays are very rich, although it is not true that every orthogonal array can be obtained this way. Beder (1998) and Li, Deng and Tang (2004) contain some nontrivial examples of orthogonal arrays that cannot be imbedded into Hadamard matrices. For all practical purposes, assuming the existence of Hadamard matrices is not as severe of a restriction as it seems, as the only values of n in the range $n \leq 1,000$, for which Hadamard matrices have not been found, are 668, 716, 764, and 892.

The situations considered in this paper are that the requirement sets contain all main effects and some selected 2fi's. Such a requirement set specifies a model with the grand mean, all main effects, and selected 2fi's. For a given requirement set S , we define its *core*, denoted by $C(S)$, to be the subset of S such that $C(S)$ includes all the 2fi's in S and that the factor of every main effect in $C(S)$ must occur

in at least one 2fi. Clearly, every requirement set has a unique core. For example, the core of $\{F_1, F_2, F_3, F_4, F_5, F_6, F_1F_2, F_3F_4\}$ is $\{F_1, F_2, F_3, F_4, F_1F_2, F_3F_4\}$ and that of $\{F_1, F_2, F_3, F_4, F_5, F_1F_2, F_2F_3\}$ is $\{F_1, F_2, F_3, F_1F_2, F_2F_3\}$, where F_1 denotes the main effect of the first factor and F_1F_2 is the 2fi between factors 1 and 2. A requirement set can be represented by a graph if we associate a main effect with a vertex and a 2fi with an edge in the graph. Then the core of a requirement set is obtained by simply deleting all isolated vertices. We say that a design supports a requirement set if it allows joint estimation of the effects in the requirement set.

Theorem 1. *An orthogonal array that supports a requirement set S exists if and only if an orthogonal array that supports its core $C(S)$ exists.*

The beauty of this result is that the problem of finding an orthogonal array for S reduces to that of finding an orthogonal array for $C(S)$. This result is especially powerful if only a few 2fi's are to be estimated, in which case $C(S)$ has a much smaller size than S . The necessity part of Theorem 1 is obvious. Only the sufficiency part needs some explanation. Let the requirement set S consist of m main effects and e 2fi's. Then its core $C(S)$ consists of m_1 of the m main effects and all e 2fi's in S . For an orthogonal array to support the requirement set S , its run size n must satisfy that $n \geq 1 + m + e$, which we assume throughout the paper. Let H be a saturated orthogonal array of n runs and D_1 be a subarray of H with m_1 columns that supports $C(S)$. Now write $H = (D_1, D_2, D_3)$, where D_2 has $m_2 = m - m_1$ columns and D_3 has $m_3 = n - 1 - m$ columns. Then Theorem 1 says that one can always obtain $D = (D_1, D_2)$ that supports S by deleting certain $m_3 = n - 1 - m$ columns, which form D_3 , from H . Any D_3 can be deleted as long as it satisfies

$$\det(X_2^T D_3 D_3^T X_2) > 0, \tag{2.1}$$

and such a D_3 always exists provided that D_1 supports $C(S)$. In (2.1), \det denotes determinant and X_2 denotes the model matrix for the e 2fi's. There are in all $\binom{m_2+m_3}{m_3}$ possible candidates for D_3 and many of these may satisfy (2.1). Which is the best? Theorem 2 answers this question. To present this result, we need the concept of the D -efficiency, defined as $[\det(n^{-1}X^T X)]^p$, where X is the model matrix and p is the number of parameters in the model. A design is D -optimal if it maximizes $[\det(n^{-1}X^T X)]^p$ or equivalently $\det(X^T X)$.

Theorem 2. *For a given requirement set S , suppose that an orthogonal array D_1 supports $C(S)$. Then, in terms of D -efficiency, design $D = (D_1, D_2)$ is best if D_3 maximizes $\det(X_2^T D_3 D_3^T X_2)$.*

Note that Theorem 2 is a conditional optimality result in the sense that it tells how to choose D_3 for given D_1 . We provide more discussion on this issue in Section 4.

Proofs of Theorems 1 and 2. Let $X_1 = (I, D_1)$, where I is the column of all plus ones and corresponds to the grand mean. Then (X_1, X_2) is the model matrix for the requirement set $C(S)$. We first have $\det[(X_1, X_2)^T(X_1, X_2)] = \det(X_1^T X_1) \det[X_2^T X_2 - X_2^T X_1 (X_1^T X_1)^{-1} X_1^T X_2]$. Noting that (X_1, D_2, D_3) is a Hadamard matrix of order n , we obtain

$$\det[(X_1, X_2)^T(X_1, X_2)] = n^{m_1+1-e} \det[X_2^T D^* D^{*T} X_2], \quad (2.2)$$

where $D^* = (D_2, D_3)$. As design D_1 supports $C(S)$, we must have $\det[X_2^T D^* D^{*T} X_2] > 0$. The fact that matrix $X_2^T D^* D^{*T} X_2$ has full rank implies that the $e \times (m_2 + m_3)$ matrix $X_2^T D^*$ has rank e . Note that $m_2 + m_3 \geq m_3 \geq e$. Then the matrix $X_2^T D^*$ must contain a submatrix of m_3 columns with rank e . This shows that we can select m_3 columns from D^* to obtain a D_3 such that $\det[X_2^T D_3 D_3^T X_2] > 0$. Now let $X = (I, D_1, D_2)$ so that (X, X_2) is the model matrix for the requirement set S . Similar to (2.2), we have $\det[(X, X_2)^T(X, X_2)] = n^{m+1-e} \det[X_2^T D_3 D_3^T X_2]$, which is strictly positive for the choice we have just made for D_3 . This shows that $D = (D_1, D_2)$ supports S , proving Theorem 1. It is also immediate that $\det[(X, X_2)^T(X, X_2)]$ is maximized if $\det[X_2^T D_3 D_3^T X_2]$ is maximized, which establishes Theorem 2.

Remark 1. In the proofs, the fact that X_2 consists of only 2fi's never gets used. Then Theorems 1 and 2 are also valid for a general requirement set, thus allowing interactions of higher order to be included. The core of a general requirement set is similarly defined, and given by removing all main effects not occurring in any interaction.

To gain some intuitive understanding of the results in Theorems 1 and 2, consider regular fractional factorial designs. In this case, the columns of X_2 form a subset of distinct columns from $D^* = (D_2, D_3)$ if D_1 supports $C(S)$. Then if D_3 consists of all the e columns of X_2 and any extra $m_3 - e$ columns from the remaining columns of D^* , the columns of (D_1, D_2, X_2) are all distinct and belong to H . Including all the columns of X_2 in D_3 is equivalent to excluding all columns of X_2 from D_2 . That is, D_2 can consist of any m_2 columns from D^* that are not in X_2 . In the general case of orthogonal arrays, though X_2 is not necessarily a subset of D^* , it nevertheless occupies an e -dimensional space in the $(m_2 + m_3)$ -dimensional space, making it possible to choose D_2 of m_2 columns from D^* to maintain linear independence among the column vectors of (D_1, D_2, X_2) .

Example 1. Suppose that we want an orthogonal array of 20 runs and 15 factors for the requirement set $S = \{F_1, \dots, F_{15}, F_1 F_2, F_2 F_3, F_3 F_4, F_1 F_5\}$. The core of S

is given by $C(S) = \{F_1, F_2, F_3, F_4, F_5, F_1F_2, F_2F_3, F_3F_4, F_1F_5\}$. Now consider the 20-run saturated orthogonal array H generated by using $(1, 1, -1, -1, 1, 1, 1, 1, -1, 1, -1, 1, -1, -1, -1, 1, 1, -1)$ as the first row, cyclically shifting this row vector one place to the left 18 times and then adding a row of all minus ones. Let $H = (d_1, \dots, d_{19})$. We can easily check that $D_1 = (d_1, d_2, d_3, d_4, d_5)$ supports $C(S)$ with the factors F_1, \dots, F_5 assigned to d_1, \dots, d_5 , respectively. Now consider $X_2 = (d_1d_2, d_2d_3, d_3d_4, d_1d_5)$ and $D^* = (D_2, D_3) = (d_6, \dots, d_{19})$. Then $X_2^T D^*$ is given by

$$\begin{bmatrix} 4 & -4 & 4 & -12 & -4 & -4 & 4 & 4 & -4 & 4 & -4 & -4 & -4 & -4 \\ 4 & 4 & -4 & 4 & -12 & -4 & -4 & 4 & 4 & -4 & 4 & -4 & -4 & -4 \\ 4 & 4 & 4 & -4 & 4 & -12 & -4 & -4 & 4 & 4 & -4 & 4 & -4 & -4 \\ -4 & 4 & -4 & -4 & 4 & 4 & -4 & 4 & -12 & -4 & -4 & 4 & -4 & 4 \end{bmatrix}. \tag{2.3}$$

As D_1 supports $C(S)$, the matrix in (2.3) must have rank 4. Let A be a 4×4 submatrix of the matrix in (2.3). We now evaluate $\det(AA^T)$ for all the $\binom{14}{4} = 1,001$ submatrices and present the results for $4^{-8} \det(AA^T)$ in the following table.

$4^{-8} \det(AA^T)$	0	64	256	576	1,024	2,304	4,096
frequency	364	305	242	22	55	6	7

From the table, we see that of 1,001 submatrices, seven take the largest value of $4^{-8} \det(AA^T) = 4,096$, one of which is given by columns 2, 4, 5, 6 of the matrix in (2.3). These columns of the matrix in (2.3) correspond to columns d_7, d_9, d_{10}, d_{11} in D^* . One best design for the requirement set S given by Theorems 1 and 2 is therefore $D = (d_1, \dots, d_6, d_8, d_{12}, \dots, d_{19})$.

In applying Theorem 2, very often more than one D_3 maximizes $\det(X_2^T D_3 D_3^T X_2)$. For instance, we see seven such choices for D_3 in Example 1. Although all these are the same in terms of D -efficiency, the resulting designs may be different in other aspects. One way to take advantage of the situation is to select the design that minimizes the contamination due to the 2fi's outside the requirement set (Ke and Tang (2003)).

3. Applications

This section explores the power of Theorem 1 for the case that S is saturated and has a structure that corresponds to a compromise plan of class one (Addelman (1962)). This special structure for S says that, besides m main effects, S contains all the $\binom{m_1}{2}$ 2fi's among a set of $m_1 \leq m$ factors. We use $S(m_1, m)$ to denote such a requirement set and simply write $S(m_1) = S(m_1, m_1)$, which is the core of $S(m_1, m)$ for any $m \geq m_1$. Any requirement set consisting of main effects and some 2fi's is a subset of an $S(m_1, m)$, provided $n \geq 1 + m + m_1(m_1 - 1)/2$.

That S is saturated means that $m = n - 1 - m_1(m_1 - 1)/2$, leaving no degrees of freedom for estimating the error variance. Consideration of saturated $S(m_1, m)$ provides a quick and simple way for finding a solution to various requirement sets by applying the following result and Theorem 1.

Lemma 1. *If an orthogonal array of n runs for a saturated $S(m_1, m)$ exists, then an orthogonal array of n runs exists for any S as long as $C(S) \subseteq S(m_1, m)$.*

This result, though very simple, allows designs to be constructed for requirement sets with different configurations from those for saturated $S(m_1, m)$. For example, suppose that design $D = (d_1, d_2, d_3, d_4, d_5)$ of 12 runs supports

$$S(m_1 = 4, m = 5) = \{F_1, F_2, F_3, F_4, F_5, F_1F_2, F_1F_3, F_1F_4, F_2F_3, F_2F_4, F_3F_4\}.$$

Then design $D_1 = (d_1, d_2, d_3, d_4)$ supports $\{F_1, F_2, F_3, F_4, F_1F_2, F_1F_3, F_2F_4\}$, from which one can find a design for $\{F_1, F_2, F_3, F_4, F_1F_2, F_1F_3, F_2F_4, F_5, F_6, F_7, F_8\}$ by Theorem 1. The next result allows large designs to be constructed from small designs.

Lemma 2. *For given m_1 , if an orthogonal array of n' runs for saturated $S(m_1, m')$ exists, then an orthogonal array of $n = kn'$ runs for saturated $S(m_1, m)$ exists, where k is such that a Hadamard matrix of order k exists.*

Cheng (1995) showed that any projection design of an orthogonal array onto four factors allows estimation of all main effects and all 2fi's if the run size n (≥ 12) is not a multiple of 8. Cheng (1998) obtained a similar result for orthogonal arrays of strength three. Combining these with Theorem 1, we have the following result.

Lemma 3. *Suppose that there exists a Hadamard matrix of order n , where $n \geq 12$ is not a multiple of 8. Then we can construct an orthogonal array of n runs for saturated $S(m_1 = 4, m)$, and an orthogonal array of $2n$ runs for saturated $S(m_1 = 5, m)$.*

Cheng's results (1995; 1998) are concerned with all projection designs onto four or five factors. To apply Theorem 1, we only need one projection design that supports $S(m_1)$. We, therefore, expect that stronger results than those in Lemma 3 are available for many values of run size n . To this end, consider the problem of finding an orthogonal array with the largest m_1 factors such that it supports $S(m_1)$ from a given saturated orthogonal array. Table 1 provides the largest m_1 values for which a design for $S(m_1)$ can be found from a saturated orthogonal array, $12 \leq n \leq 60$. In all the cases in Table 1, the results are the best in the sense that any larger value for m_1 will break the degree of freedom constraint $n \geq 1 + m_1 + m_1(m_1 - 1)/2$. The entries for $n = 12, 16$ are obvious, and those for $n = 20, 24, 28$ are available from Loeppky, Sitter, and Tang (2007).

Table 1. The largest m_1 values for which a design for $S(m_1)$ can be found.

n	m_1	Hadamard matrix
12	4	Cheng (1995)
16	5	resolution V design
20	5	Loeppky et al. (2007)
24	6	Loeppky et al. (2007)
28	6	Loeppky et al. (2007)
32	7	had.32.pal
36	7	had.36.pal2
40	8	had.40.tpal
44	8	had.44.pal
48	9	had.48.pal
52	9	had.52.will
56	10	had.56.tpal2
60	10	had.60.pal

For $32 \leq n \leq 60$, Table 1 provides the Hadamard matrices from which we obtain our results. These Hadamard matrices and their labels are from Neil Sloane’s webpage (<http://www.research.att.com/~njas/>).

The existence of orthogonal arrays for various requirement sets can be established by combining Table 1 with Theorem 1 and Lemmas 1 and 2. For example, the existence of an orthogonal array of 44 runs for $S(m_1 = 8)$ in Table 1 implies the existence of a design of 44 runs for $S = \{F_1, \dots, F_8, F_1F_2, F_3F_4, F_5F_6, F_7F_8\}$ by Lemma 1. Applying Theorem 1, we establish the existence of a design of 44 runs for $S = \{F_1, \dots, F_8, F_1F_2, F_3F_4, F_5F_6, F_7F_8, F_9, \dots, F_{39}\}$. The same idea in Lemma 2 further implies the existence of a design of 176 runs for $S = \{F_1, \dots, F_8, F_1F_2, F_3F_4, F_5F_6, F_7F_8, F_9, \dots, F_{171}\}$.

Our focus in this section has been to establish the existence of orthogonal arrays for various requirement sets. We conclude the section with a brief comment on the performance of these designs under the D -criterion. Unless an orthogonal array provides an orthogonal design for the given requirement set, full efficiency is not achieved. The efficiency loss is due to the possible nonorthogonality between main effects and the 2fi’s and between the 2fi’s themselves. When the number of 2fi’s is small relative to that of main effects, one would expect that the designs for such requirement sets achieve high efficiency.

4. Discussion

A more ambitious research problem than what has been done in this paper is to find a D -optimal orthogonal array for a given requirement set from all

orthogonal arrays. The results in this paper only provide a partial solution to this problem, as Theorem 2 is an optimality result conditional on a given orthogonal array that supports $C(S)$. Completely solving this problem is nontrivial if not impossible. Two complications arise. The first is that we need to consider all nonisomorphic saturated orthogonal arrays for a given run size. The complete set of nonisomorphic saturated orthogonal arrays is available for run size $n \leq 24$. Although the complete set of nonisomorphic Hadamard matrices of order 28 is available, the complete set of nonisomorphic saturated orthogonal arrays of run size 28 has not been identified. There is no simple way of resolving this complication, which also occurs in almost all other studies of design selection. A realistic approach would be simply considering some saturated orthogonal arrays one can easily obtain when the complete set is unavailable. The second complication is that in terms of D -efficiency, the best orthogonal array for S does not necessarily come from the best orthogonal array for $C(S)$. Despite this, we can still establish that $\det[(X, X_2)^T(X, X_2)] \leq n^{m-m_1} \det[(X_1, X_2)^T(X_1, X_2)]$, which is useful for developing computational algorithms in the search of the best design for S . This is a topic for future research.

Acknowledgement

The research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

- Addelman, S. (1962). Symmetrical and asymmetrical fractional factorial plans. *Technometrics* **1**, 47-58.
- Beder, J. H. (1998). Conjectures about Hadamard matrices. *J. Statist. Plann. Inference* **72**, 7-14.
- Chen, H. and Hedayat, A. S. (1998). 2^{n-m} designs with resolution III or IV containing clear two-factor interactions. *J. Statist. Plann. Inference* **75**, 147-158.
- Cheng, C. S. (1995). Some projection properties of orthogonal arrays. *Ann. Statist.* **23**, 1223-1233.
- Cheng, C. S. (1998). Some hidden projection properties of orthogonal arrays with strength three. *Biometrika* **85**, 491-495.
- Cheng, C. S. (2003). Personal communication.
- Cheng, C. S. and Tang, B. (2005). A general theory of minimum aberration and its applications. *Ann. Statist.* **33**, 944-958.
- Dey, A. and Suen, C.-Y. (2002). Optimal fractional factorial plans for main effects and specified two-factor interactions: a projective geometric approach. *Ann. Statist.* **30**, 1512-1523.
- Franklin, M. F. and Bailey, R. A. (1977). Selection of defining contrasts and confounded effects in two-level experiments. *Appl. Statist.* **26**, 321-326.
- Greenfield, A. A. (1976). Selection of defining contrasts in two-level experiments. *Appl. Statist.* **25**, 64-67.

- Hedayat, A. S. and Pesotan, H. (1992). Two-level factorial designs for main effects and selected two-factor interactions. *Statist. Sinica* **2**, 453-464.
- Hedayat, A. S., Sloane, N. J. A. and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York.
- Hedayat, A. S. and Stufken, J. (1999). Compound orthogonal arrays. *Technometrics* **41**, 57-61.
- Ke, W. and Tang, B. (2003). Selecting 2^{m-p} designs using a minimum aberration criterion when some two-factor interactions are important. *Technometrics* **45**, 352-360.
- Li, Y., Deng, L. Y. and Tang, B. (2004). Design catalog based on minimum G -aberration. *J. Statist. Plann. Inference* **124**, 219-230.
- Loeppky, J. L., Sitter, R. R. and Tang, B. (2007). Non-regular designs with desirable projection properties. *Technometrics* **49**, 454-467.
- Rechtschaffner, R. L. (1967). Saturated fractions of 2^n and 3^n factorial design. *Technometrics* **9**, 569-575.
- Rosenbaum, P. R. (1996). Some useful compound dispersion experiments in quality design. *Technometrics* **38**, 354-364.
- Tang, B. (2006). Orthogonal arrays robust to nonnegligible two-factor interactions. *Biometrika* **93**, 137-146.
- Wu, C. F. J. and Chen, Y. (1992). A graph-aided method for planning two-level experiments when certain interactions are important. *Technometrics* **34**, 162-175.
- Wu, C. F. J. and Hamada, M. (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization*. Wiley, New York.
- Wu, C. F. J. and Zhu, Y. (2003). Optimal selection of single arrays for parameter design experiments. *Statist. Sinica* **13**, 1179-1199.
- Wu, H. and Wu, C. F. J. (2002). Clear two-factor interactions and minimum aberration. *Ann. Statist.* **30**, 1496-1511.

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

E-mail: boxint@stat.sfu.ca

Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia V8W 3R4, Canada.

E-mail: jzhou@math.uvic.ca

(Received June 2007; accepted June 2008)