Statistica Sinica **25** (2015), 853-861
doi:http://dx.doi.org/10.5705/ss.2013.311

A SYSTEMATIC APPROACH FOR THE CONSTRUCTION OF DEFINITIVE SCREENING DESIGNS

Frederick Kin Hing Phoa and Dennis K. J. Lin

Academia Sinica and Pennsylvaniva State University

Abstract: Definitive screening (DS) designs, first proposed by Jones and Nachtsheim (2011), have received much attention due to their good design properties and run-size economy. This paper investigates the structure of three-level DS designs and suggests a theoretically-driven approach to constructing DS designs for any number of run size. This construction method is computationally efficient and universal to designs with even or odd number of factors. Many constructed DS designs possess high (if not optimal) D-efficiencies, that are comparable to those found in the literature.

Key words and phrases: Definitive screening (DS) designs, conference matrix, circulant matrix, D-optimal.

1. Introduction

There is considerable scope for reducing resources used in research by designing more efficient studies. Experiments are increasingly complex, and are subject to rising experimental cost and competing resources. Careful design considerations, even with only minor variation to traditional designs, can lead to a more efficient study in terms of more precise estimates or the ability to estimate more effects at the same cost.

Resolution III and IV two-level fractional factorial designs have been widely used for early-stage screening experimentation (Box and Hunter (1961)). However, resolution III designs fail to disentangle the confounding between main effects and two-factor interactions, and similar situations exist in resolution IV designs a pair of two-factor interactions (Phoa, Wong, and Xu (2009)). A nonregular design has the ability to disentangle the partial aliased structure, but it is not simple to construct and analyze (Phoa, Xu, and Wong (2009)). Quaternary-code designs (Phoa and Xu (2009); Phoa (2012)) possess similar structure as regular designs, but their structure-based analysis is still under investigation.

In addition to confounding, two-level designs have no capability for capturing curvature due to pure-quadratic effects, and this led to a new class of three-level screening designs called definitive screening (DS) designs (Jones and Nachtsheim (2011)). In essence, DS designs are 3-level designs, for studying m quantitative factors with the following desired properties: the number of required runs is $N = 2m+1$, so they is saturated for estimating the intercept, m main effects and m quadratic effects; all main effects are orthogonal to other main effects, all quadratic effects and all two-factor interactions.

Stylianou (2011) and Xiao, Lin, and Bai (2012) pointed out that the DS designs can be constructed by stacking a positive and a negative C, plus a row of zeros as the center points. C is generally known as a conference matrix of order m: an $m \times m$ matrix whose diagonal entries of C vanish, its off-diagonal entries lie in $\{-1, 1\}$ and $C'C = (m-1)I$, where I is the $m \times m$ identity matrix. A conference matrix is only available for even m. Nguyen and Stylianou (2012) connected conference matrices to incomplete block designs, and they proposed a general algorithm to construct an $m \times m$ $(0, \pm 1)$-matrix with zero diagonal. This method shrinks the searching space and it faster than the computerized enumeration in Jones and Nachtsheim (2011). Still, it is a purely computerized search without structural information.

We propose a systematic approach to constructing an $m \times m$ $(0, \pm 1)$-matrix with zero diagonal, called a C matrix, and its corresponding DS design. Section 2 gives the notations and definitions that we use. The general structure of C is given in this section. Section 3 states the properties of the components in the structure of C: our results describe the criteria on how the C matrix is generated, and the D-efficiency of the corresponding DS designs is discussed. Section 4 compares the DS designs generated in this paper to the best DS designs found in the literature. Section 5 is for the discussion and conclusion. Proofs are given in the Appendix.

2. Notations, Definitions and Structure

Let m be the number of factors, and set $m = 2n+2$ for even m and $2n+1$ for odd m. We need two additional matrices, T and S. T can be obtained cyclically via a generator vector $\vec{t}$ as follows. Given $\vec{t} = (0, t_2, \ldots, t_n)'$, we form the lower-triangular matrix

$$T_i = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ t_2 & 0 & 0 & \cdots & 0 & 0 \\ t_3 & t_2 & 0 & \cdots & 0 & 0 \\ t_4 & t_3 & t_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ t_n & t_{n-1} & t_{n-2} & \cdots & t_2 & 0 \end{pmatrix},$$

and $T = T_l + T_l'\delta$, where T_l' is the transpose of T and $\delta = 1$ when n is even and $\delta = -1$ when n is odd. Similarly, S can be obtained cyclicly via a generator vector $\vec{s}$ as follows.

Given $\vec{s} = (s_1, \ldots, s_n)'$, take

$$S = \begin{pmatrix} s_1 & s_2 & s_3 & \cdots & s_{n-1} & s_n \\ s_2 & s_3 & s_4 & \cdots & s_n & s_1 \\ s_3 & s_4 & s_5 & \cdots & s_1 & s_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s_n & s_1 & s_2 & \cdots & s_{n-2} & s_{n-1} \end{pmatrix}.$$

The C matrix is then

$$C = \begin{pmatrix} 0 & \delta & \vec{\delta}' & \vec{\delta}' \\ 1 & 0 & \vec{\delta}' & -\vec{\delta}' \\ \vec{1} & \vec{1} & T & S\delta \\ \vec{1} & -\vec{1} & S & -T\delta \end{pmatrix}; \text{ when } m \text{ is even;}$$

$$C = \begin{pmatrix} 0 & -\vec{\delta} & -\vec{\delta} \\ \vec{1} & T & S\delta \\ -\vec{1} & S & -T\delta \end{pmatrix}; \text{ when } m \text{ is odd,}$$

where $\vec{1}$ and $\vec{\delta}$ are column vectors of length n with all entries 1 and δ, respectively. The design matrix D for the DS designs can be constructed as

$$D = \begin{pmatrix} C \\ -C \\ \vec{0}' \end{pmatrix},$$

where $\vec{0}$ is a column vector of 0's of length m and C is an $m \times m$ $(0, \pm 1)$-matrix with zero diagonal. Ideally here, C is a conference matrix.

Let D_o be a (hypothetical) D-optimal design of order m. For a fair comparison with DS designs in the literature, we adapt the D-efficiency criterion of Jones and Nachtsheim (2011):

$$d_e(D, D_o) = \left(\frac{|X(D)'X(D)|}{|X(D_o)'X(D_o)|} \right)^{1/p},$$

where $X(D)$ and $X(D_o)$ are design matrices of designs D and D_o respectively, $|M|$ is the determinant of a matrix M, p is the number of terms in the model that consists of the intercept term and all linear effects.

3. Main Theorems

Some lemmas about the properties of T and S are stated for convenience.

Lemma 1. *S is back-circulant and symmetric.*

The back-circulant property of S is obvious from its structure, which implies that the sums of columns and rows of S are all equal, say $s = \sum_{i=1}^{n} s_i$. It is well-known that a back-circulant matrix is always symmetric.

Lemma 2. *For $t_i = t_{n+2-i}\delta$ in $\vec{t}$, T is back-circulant and symmetric when n is even, T is back-circulant and anti-symmetric when n is odd.*

When $t_i = t_{n+2-i}$, $\vec{t} = (0, t_2, t_3, \ldots, t_{(n-1)/2}, t_{n/2}, t_{(n-1)/2}, \ldots, t_3, t_2)$ if n is odd, and $\vec{t} = (0, t_2, t_3, \ldots, t_{n/2-1}, -t_{n/2-1}, \ldots, -t_3, -t_2)$ if n is even. T is circulant no matter which $\vec{t}$ is used as the generator. The circulant property implies that the sums of columns and rows of T are all equal, say $t = \sum_{i=1}^{n} t_i$.

Theorem 1. *Let $\vec{s}$ and $\vec{t}$ be the generators of S and T, m even. If $t_i = t_{n+2-i}$ in $\vec{t}$, $(s, t) = (0, -1)$ for even n or $(1, 0)$ for odd n, and $\sum_i (s_i s_{(i \ mod \ n)+k} + t_i t_{(i \ mod \ n)+k}) = -2$ for all integers $k < (n+1)/2$, then C is a $(2n+2) \times (2n+2)$ conference matrix.*

Theorem 2. *Let $D = (C, -C, \vec{0})'$, where C is the $(2n+2) \times (2n+2)$ conference matrix of Theorem 1. Then the D-efficiency of D is optimal and $d_e(D, D_0) = ((2n+1)/(2n+2))^{(2n+2)/(2n+3)}$.*

Notice that a D-optimal matrix does not exist when a conference matrix structure is assumed. This is mainly because of the existence of zero entries in each column of C: if D-optimal matrix exists, all entries have to be $+1$ or -1. For D-efficiency, adding more zero entries in each column of C reduces the value of the determinant of $C'C$, so the best case is to add only one zero in each column of C. The use of Theorems 1 and 2 in constructing DS designs with even m is illustrated below.

Example 1. Consider a construction of DS designs with $m = 12$, a 25×12 matrix. An example of this kind is given as D, the formula (1) in Xiao, Lin, and Bai (2012). Here $m = 12$, $n = 5$. To fulfill conditions (1) - (3) of Theorem 1, only two combinations are possible: $\vec{t} = (0, 1, 1, -1, -1)$ and $\vec{s} = c(1, 1, -1, 1, -1)$ or $\vec{t} = (0, 1, -1, 1, -1)$ and $\vec{s} = c(1, 1, 1, -1, -1)$. If we choose the first, then

$$C = \begin{pmatrix}
0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
+1 & 0 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 \\
+1 & +1 & 0 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 & +1 \\
+1 & +1 & +1 & 0 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\
+1 & +1 & +1 & +1 & 0 & -1 & -1 & +1 & -1 & +1 & -1 & -1 \\
+1 & +1 & -1 & +1 & +1 & 0 & -1 & -1 & +1 & -1 & -1 & +1 \\
+1 & +1 & -1 & -1 & +1 & +1 & 0 & +1 & -1 & -1 & +1 & -1 \\
+1 & -1 & +1 & +1 & -1 & +1 & -1 & 0 & -1 & -1 & +1 & +1 \\
+1 & -1 & +1 & -1 & +1 & -1 & +1 & +1 & 0 & -1 & -1 & +1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & +1 & 0 & -1 & -1 \\
+1 & -1 & +1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 & 0 & -1 \\
+1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & +1 & 0
\end{pmatrix}$$

and $D = (C, -C, \vec{0}_{12})'$, where $\vec{0}_{12}$ is a vector of length 12 with all entries zeros, and C is a 12×12 conference matrix. According to Theorem 2, D is a DS design with D-efficiency 92.2823% which is optimal, and equivalent to those in Jones and Nachtsheim (2011), Xiao, Lin, and Bai (2012), and Nguyen and Stylianou (2012).

If the second coombination is chosen, a different 12×12 matrix C results due to different T and S matrices. However, C is still a conference matrix and the resulting DS design has optimal D-efficiency.

Theorem 3. *Given $\vec{s}$ and $\vec{t}$ as the generators of S and T, m odd. If the conditions of Theorem 1 hold, then*

$$C = \begin{pmatrix} 2n & -\vec{1}' & -\vec{1}' \\ -\vec{1} & A & 1_{n\times n} \\ -\vec{1} & 1_{n\times n} & A \end{pmatrix},$$

where $\vec{1}$ is a vector of length n that all entries are 1, $1_{n\times n}$ is a $n \times n$ matrix that all entries are 1, and A is a $n \times n$ matrix such that the diagonal entries of A are $2n$ and the off-diagonal entries of A are -1.

Let $\{a_i\}$, $\{o_i\}$, and $\{b_i\}$ be sequences of length $n+1$, n, and $2n+1$, repectively, with $a_0 = b_0 = 2n$, $a_1 = b_1 = 2n - (1/2n)$, and $o_1 = 1 - (1/2n)$. For $i = 2, \ldots, n+1$, take the

$$\begin{aligned} a_i &= (2n+1)(2 - \frac{2n+1}{a_i - 1}), \\ o_i &= o_{i-1}\frac{2n+1}{a_i - 1}, \\ b_i &= b_{i-1} - \frac{o_i^2 a_{i-1}}{(2n+1)^2}, \end{aligned}$$

and $i = n+2, \ldots, 2n$, take

$$b_i = (2n+1)(2 - \frac{2n+1}{b_i - 1}).$$

Theorem 4. *If $D = (C, -C, \vec{0})'$, where C is the $(2n+1) \times (2n+1)$ matrix in Theorem 3, it has D efficiency $(|C'C|/(2n)^{2n+1})^{1/(2n+2)}$, where $|C'C|$ is equal to $\prod_{i=0}^{2n} a_i I(i \le n) + b_i I(i > n)$.*

An example of the construction of a DS design with odd m is given below.

Example 2. Take $m = 11$ to find a 23×11 matrix. As in Example 1, the first combination of $\vec{t}$ and $\vec{s}$ is used for design construction and, according to Theorem 3,

$$C'C = \begin{pmatrix} 10 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 10 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 10 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 10 & -1 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 10 & -1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 10 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 & 1 & 10 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 & 1 & -1 & 10 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & 10 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & 10 & -1 \\ -1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 10 \end{pmatrix}.$$

Although C is not a conference matrix, its generated D possesses good D-efficiency. We obtain sequences $\{a\} = \{10.0000, 9.9000, 9.7778, 9.6250, 9.4286, 9.1667\}$ and $\{b\} = \{10.0000, 9.9000, 9.8182, 9.7159, 9.5844, 9.4091, 9.1636, 8.7956, 8.2432, 7.3212, 5.4726\}$, which leads to D-efficiency of the DS design D of 87.9553%, via Theoream 4.

4. A Table of Suggested DS Designs

Table 1 details some DS designs constructed via the proposed method.

For each n, or each pair of $\vec{t}$ and $\vec{s}$, one has in two DS designs. The first is generated via the method when C is odd, see Example 2, the second is generated via the method when C is even, see Example 1.

Except for $n = 10$, all $\vec{t}$ and $\vec{s}$ listed in Table 1 fulfill the conditions stated in Theorem 1. Thus the D-efficiencies of all listed DS designs with even m are optimal, while those with odd m are high but not optimal. For $n = 10$, a pair $\vec{t}$ and $\vec{s}$ that fulfills all conditions of Theorem 1 does not exist. The $\vec{t}$ and $\vec{s}$ in Table 1 fulfill the first two conditions, and the DS design with $m = 22$ may not be D-optimal.

5. Discussions and Concluding Remarks

When C has an odd number of factors, one can search for some DS designs with better D-efficiencies than ours, see the results in Jones and Nachtsheim (2011) and Nguyen and Stylianou (2012). However, the proposed method suggests itself because of its simple construction, its guaranteed high D-efficiency, and its universality on all matrix orders.

As the dimension of DS designs increases, computer search is generally inefficient due to the increasingly large search space. Our construction method reduces the search space to a manageable size via the conditions in Theorem 1. Our construction method is universal to all matrix orders, and no matter if the

Table 1. Some suggested DS designs for $n = 3, \ldots, 15$.

n	m	$d_e(D, D_o)$	Generators
3	7	86.339	$\vec{t} = (0 + -)$
	8	88.808	$\vec{s} = (+ + -)$
4	9	87.173	$\vec{t} = (0 - + -)$
	10	90.866	$\vec{s} = (+ + - -)$
5	11	87.955	$\vec{t} = (0 + + - -)$
	12	92.282	$\vec{s} = (+ + - + -)$
6	13	88.664	$\vec{t} = (0 + - - - +)$
	14	93.317	$\vec{s} = (+ + - + - -)$
7	15	89.298	$\vec{t} = (0 + + - + - -)$
	16	94.107	$\vec{s} = (+ + + - + - -)$
8	17	89.863	$\vec{t} = (0 + - - + - - +)$
	18	94.729	$\vec{s} = (+ + + - + - - -)$
9	19	90.369	$\vec{t} = (0 + + + - + - - -)$
	20	95.232	$\vec{s} = (+ + - + - + + - -)$
10	21	85.386	$\vec{t} = (0 + - - + - + - - +)$
	22	90.163	$\vec{s} = (+ + + - + - - - + -)$
11	23	91.233	$\vec{t} = (0 + + + - + - + - - -)$
	24	95.997	$\vec{s} = (+ + - + + - + + - - -)$
12	25	91.604	$\vec{t} = (0 - + + - - + - - + + -)$
	26	96.293	$\vec{s} = (+ + + + - + - + - - - -)$
13	27	91.942	$\vec{t} = (0 + + + - + + - - + - - -)$
	28	96.550	$\vec{s} = (+ + + + - - + - + - + - -)$
14	29	92.251	$\vec{t} = (0 + + - - - + - + - - - + +)$
	30	96.772	$\vec{s} = (+ + + - + + - + - - + - - -)$
15	31	92.534	$\vec{t} = (0 + + + - + + - + - - + - - -)$
	32	96.968	$\vec{s} = (+ + + - + - - - + - + + + - -)$

conference matrix exists in the given order. In contrast, the method in Xiao, Lin, and Bai (2012) is workable only when the conference matrix exists, and the method in Georgiou, Stylianou, and Aggarwal (2013) allows matrix of all even orders to be constructed using weighing matrix $W(m, m-1)$ with $s = 1$. The method in Nguyen and Stylianou (2012) allows matrix of even or odd orders to be constructed, but see m =10,12, 20, 22, 24, 26, 28, 34, 36, 44, 48, 50 in their Table 1 for even order; it is a computationally intensive search for odd order.

Fletcher, Gysin, and Seberry (2001) provided a list of GL(q)-pair from n = 3 to 55 via exhaustive computer searches, some of which are similar to our reported vectors $\vec{t}$ and $\vec{s}$ with the modification that the first entry in the first sequence of GL(q)-pair changes from 1 to 0 in our $\vec{t}$ in order to cooperate with the zero-diagonal property of conference matrix. We note some other differences.

The GL(q)-pair sequences reported in Fletcher, Gysin, and Seberry (2001) have odd length, these pairs do not exist for even lengths. The condition in Theorem 1 greatly reduces the search space for our vectors, while their GL(q)-pairs result from exhaustive computer searches. Hadamard matrices are constructed via GL(q)-pairs, while conference matrix can only be obtained from a skew symmetric Hadamard matrix via $C = H - I_n$. This is only possible for an even number of factors; conference matrices with an odd and even number of factors can be constructed using the proposed method.

Acknowledgement

The authors thank the associate editor and two reviewers for their consructive comments and suggestions. This work was supported by National Science Council of Taiwan ROC grant numbers 100-2118-M-001-002-MY2 and 102-2628-M-001-002-MY3 and Thematic Research Program of Academia Sinica (Taiwan) grant number AS-103-TP-C03.

References

Box, G. E .P and Hunter, J. S. (1961). The 2^{k-p} fractional factorial designs. *Technometrics* **3**, 311-351.

Fletcher, R. J., Gysin, M. and Seberry, J. (2001). Application of the discrete Fourier transform to the search for generalised Legendre pairs and Hadamard matrices. *Austral J. Combinatorics* **23**, 75-86.

Georgiou, S. D., Stylianou, S. and Aggarwal, M. (2013). Efficient three-level screening designs using weighing matrices. *Statistics: A Journal of Theoretical and Applied Statistics.* In press.

Jones, B. and Nachtsheim, C. J. (2011). A class of three levels designs for definitive screening in the presence of second order effects. *J. Quality Tech.* **43**, 1-15.

Nguyen, N. K. and Stylianou, S. (2012). Constructing definitive screening designs using cyclic generators. *J. Statist. Theory and Practice* **7**, 713-724.

Phoa, F. K. H. (2012). A code arithmetic approach for Quaternary code designs and its applications to $(1/64)^{th}$-fractions. *Ann. Statist.* **40**, 3161-3175.

Phoa, F. K. H., Wong, W. K. and Xu, H. (2009). The need of considering the interactions in the analysis of screening designs. *J. Chemometrics* **23**, 545-553.

Phoa, F. K. H. and Xu, H. (2009). Quarter-fraction factorial design constructed via Quaternary codes. *Ann. Statist.* **37**, 2561-2581.

Phoa, F. K. H., Xu, H., and Wong, W. K. (2009). The use of nonregular fractional factorial designs in combination toxicity studies. *Food and Chemical Toxicology* **47**, 2183-2188.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

Stylianou, S. (2011). Three-level screening designs applicable to models with second order terms. Paper presented at the International Conference on Design of Experiments (ICODOE 2011). May 10-13, 2011, Department of Mathematical Sciences, University of Memphis, Memphis, USA.

Wallis, W. D., Street, A. P. and Wallis, J. S. (1972). *Combinatorics: Room Squares, Sum-free sets, Hadamard Matrices.* Lecture Notes in Mathematics, Springer Verlag, Berlin-Heidelberg-New York.

Xiao, L., Lin, D. K. J. and Bai, F. (2012). Constructing definitive screening designs using conference matrices. *J. Quality Tech.* **44**, 2-8.

Institute of Statistical Science, Academia Sinica, Taipei City 115, Taiwan R.O.C.

E-mail: fredphoa@stat.sinica.edu.tw

Department of Statistics, The Pennsylvania State University, 317 Thomas Building University Park, PA 16802, USA.

E-mail: dkl5@psu.edu

(Received October 2013; accepted May 2014)