Definitive Screening Designs with Added Two-Level Categorical Factors*

BRADLEY JONES

SAS Institute, Cary, NC 27513

CHRISTOPHER J. NACHTSHEIM

Carlson School of Management, University of Minnesota, Minneapolis, MN 55455

Recently, Jones and Nachtsheim (2011) proposed a new class of designs called definitive screening designs (DSDs). These designs have three levels, provide estimates of main effects that are unbiased by any second-order effect, require only one more than twice as many runs as there are factors, and avoid confounding of any pair of second-order effects. For designs having six factors or more, these designs project to efficient response surface designs with three or fewer factors. A limitation of these designs is that all factors must be quantitative. In this paper, we develop column-augmented DSDs that can accommodate any number of two-level qualitative factors using two methods. The *DSD-augment* method provides highly efficient designs that are still *definitive* in the sense that the estimates of all main effects continue to be unbiased by any active second-order effects. An alternative procedure, the *ORTH-augment* approach, leads to designs that are orthogonal linear main effects plans; however, some partial aliasing between main effects and interactions involving the categorical factors is present.

Key Words: Alias; Conference Matrix; Confounding; Coordinate Exchange Algorithm; D-Efficiency; Robust Designs; Screening Designs.

1. Introduction

A new class of small three-level designs called definitive screening designs (DSDs) was recently introduced by Jones and Nachtsheim (2011, JN herein). For m factors, these designs require $2m+1$ runs and have the structure described in Table 1, where $x_{i,j}$ denotes the setting of the jth factor for the ith run. From the table, it is easily seen that the $2m+1$ runs are comprised of m fold-over pairs and an overall center point. Each run (excluding the centerpoint) has exactly one factor level at its center point and all others at the extremes.

JN gave the following list of advantages that these designs provide in comparison with the 2^{k-p} system as given by Box and Hunter (1961).

1. The number of required runs is only one more than twice the number of factors.
2. Unlike resolution III designs, main effects are completely independent of two-factor interactions. As a result, estimates of main effects are not biased by the presence of active two-factor interactions, regardless of whether the interactions are included in the model.
3. Unlike resolution IV designs, two-factor interactions are not completely confounded with other two-factor interactions, although they may be correlated.
4. Unlike resolution III, IV, and V designs with added center points, all quadratic effects are

*The editorial review process for this paper was handled by the past editor, Douglas C. Montgomery.

Dr. Jones is Principal Research Fellow for the JMP Division of SAS and Guest Professor at the University of Antwerp. His email address is bradley.jones@sas.com.

Dr. Nachtsheim is the Frank A. Donaldson Chair of Operations Management, Chair of the Supply Chain and Operations Department at the Carlson School of Management and is a member of the Graduate Faculty of the School of Statistics at the University of Minnesota. His email address is nacht001@umn.edu.

TABLE 1. General Definitive Screening Design Structure for *m* Factors

Foldover pair	Run (i)	Factor levels				
		$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$\cdots$	$x_{i,m}$
1	1	0	± 1	± 1	$\cdots$	± 1
	2	0	∓ 1	∓ 1	$\cdots$	∓ 1
2	3	± 1	0	± 1	$\cdots$	± 1
	4	∓ 1	0	∓ 1	$\cdots$	∓ 1
3	5	± 1	± 1	0	$\cdots$	± 1
	6	∓ 1	∓ 1	0	$\cdots$	∓ 1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
m	$2m-1$	± 1	± 1	± 1	$\cdots$	0
	$2m$	∓ 1	∓ 1	∓ 1	$\cdots$	0
Centerpoint	$m+1$	0	0	0	$\cdots$	0

estimable in models comprised of any number of linear and quadratic main effects terms.

5. Quadratic effects are orthogonal to main effects and not completely confounded (though correlated) with interaction effects.
6. With 6 through (at least) 12 factors, the designs are capable of estimating all possible full-quadratic models involving three or fewer factors with very high levels of statistical efficiency.

JN used numerical methods to identify D-optimal DSDs and found that, for $m = 4$, 6, 8, and 10, the designs were orthogonal main effects plans. More recently, Xiao et al. (2012) showed how conference matrices can be used to construct orthogonal DSDs for most values of even $m \geq 4$. For odd m and for values of m that are even and for which no conference-matrix–based construction exists, we simply find the next higher value of m, say m', for which an orthogonal DSD exists, and drop $m' - m$ columns to obtain an orthogonal $(2m' + 1)$-run DSD for m factors. In what follows, we limit attention to orthogonal DSDs obtained in this manner.

A limitation of DSDs is that all of the factors must be quantitative. In an effort to broaden the applicability of DSDs, we consider the following question: Can DSDs be augmented by two-level categorical factors efficiently and in a way that preserves all or most of the properties discussed above? For the most part, the answer to this question is in the affirmative, with a few caveats that we delineate in the development below.

The designs we produce fall into the general area of mixed-level screening designs. Other approaches to constructing mixed-level designs include Hedayat et al. (1999), Nguyen (1996), Wang and Wu (1992), and Xu (2002).

An outline of the remainder of the paper is as follows. In Section 2, we give a simple construction method that leads to main effects plans whose estimated main effects are completely independent of second-order effects. In Section 3, we provide an alternative procedure that produces orthogonal linear main effects plans when the number of added categorical factors is less than or equal to four, and nearly orthogonal plans when the number of added categorical factors is small but greater than four. While these plans are D-optimal (in a sense to be described), they lead to some partial aliasing between main effects and two-factor interactions. Finally, in Section 4, we summarize results and discuss the relative strengths of the two approaches.

Throughout, we assume that the response y_i follows the normal theory linear model

$$y_i = \beta_0 + \sum_{j=1}^{m+c} \beta_j x_{i,j} + \sum_{j=1}^{m+c-1} \sum_{k=j+1}^{m+c} \beta_{jk} x_{i,j} x_{i,k} + \sum_{j=1}^{m} \beta_{jj} x_{i,j}^2 + \varepsilon_i \qquad i = 1, \ldots, n, \tag{1}$$

where m is the number of quantitative factors, c is the number of qualitative factors, where the parameters $\beta_0, \ldots, \beta_{mm}$ are unknown constants (of which many are zero by the sparsity of effects assumption), and the $\{\varepsilon_i\}$ are *iid* $\mathrm{N}(0, \sigma^2)$. In matrix form, we write $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, where $\mathbf{X}_1$ is the $n \times (m+1)$ design matrix for the intercept term and the linear main effects in $\boldsymbol{\beta}_1$ and $\mathbf{X}_2$ is the design matrix corresponding to the $m(m+1)/2$ second-order effects (including linear-by-linear interactions and quadratic main effects) in $\boldsymbol{\beta}_2$. Because the goal is screening, we assume that, at least initially, the experimenter will fit the first-order model to the response.

2. DSD-Based Column Augmentation

Our goal is to add columns in such a way that the design is highly efficient for estimation of the first-order model while maintaining complete independence between estimates of all main effects and second-order effects. As will be shown, the designs to be described achieve both of these goals with two limitations: (1) partial aliasing between the intercept and interaction terms involving categorical factors remains; and (2) the resulting designs, while highly efficient, are not completely orthogonal.

The design construction method employs the following steps:

1. Let $n_1 + 1$ denote the size of the smallest orthogonal DSD for $m + c$ factors, so that n_1 is the number of runs omitting the center-point run. The design matrix for this design is shown in the upper part of Table 2 for $c = 4$. In Table 2, $\mathbf{1}$ denotes the constant column; $\mathbf{d}_i^+$ denotes the ith column from an n_1-run DSD (with center point omitted), augmented by two center points; and $\mathbf{a}_j^+ = (\mathbf{a}_j', \mathbf{b}_j')'$ denotes the jth augmented column, where $\mathbf{b}_j$ is the jth column of the $(2 \times c)$ $\mathbf{B}$ matrix, which is identified in Step 2. The jth augmented column contains two unknowns, $z_{j,1}$ and $z_{j,2}$, in the two entries that would have zeros in an $m + c$ factor DSD. There may or may not be added fold-over pairs with no center points, depending on the value of $m + c$. These are runs $2m + 9$ through n_1 if present; if not, this block of runs in the table is deleted and we have $n_1 = 2m + 8$.
2. Set $\mathbf{B}$ equal to the $2 \times c$ matrix,
$$\mathbf{B} = \begin{pmatrix} b_1 & b_2 & \cdots & b_c \\ -b_1 & -b_2 & \cdots & -b_c. \end{pmatrix}$$
3. For $j = 1, \ldots, c$, set $z_{j,1} = z_j$, and $z_{j,2} = -z_j$.
3. Steps 2 and 3 guarantee that the resulting design will be made up of $n/2$ fold-over pairs and, hence, there will be no aliasing between main effects and second-order effects. With this structure, there are $2c$ unknowns in the design, $b_1, \ldots, b_c$, and $z_1, \ldots, z_c$. Now evaluate all 2^{2c} possibilities for setting $b_i = \pm 1$ and $z_i = \pm 1$. Choose any design that maximizes $|\mathbf{X}_1'\mathbf{X}_1|$ from among the 2^{2c} alternatives.

Because this method guarantees that estimates of linear main effects are completely independent of all second-order effects, we refer to it as the *DSD-augment* method and we denote the run size ($n = n_1 + 2$) by n_{DSD}. The alias matrix for the design having four continuous factors and two categorical factors is shown in Table 3. (To save space, the zero aliasing between the linear main effects terms and the quadratic main effects terms has been omitted.) A disadvantage of DSD-based column augmentation is that the main effects columns for categorical factors exhibit small correlations so that the information matrix is not diagonal. To see this, we observe that $\mathbf{d}_i^{+\prime}\mathbf{a}_j^+ = \pm(z_{i,1} - z_{i,2})$. With $z_{i,1} = \pm 1$ and $z_{i,2} = \mp 1$, $\mathbf{d}_i^{+\prime}\mathbf{a}_j^+ = \pm 2$. Also, for $i \neq j$, it can be shown that $\mathbf{a}_i^{+\prime}\mathbf{a}_j^+ = 2 \pm (z_{i,1} - z_{i,2}) \pm (z_{j,1} - z_{j,2})$, which can potentially take on values -2, 2, and 6. In our experience, the DSD-augment procedure usually eliminates the $\mathbf{a}_i^{+\prime}\mathbf{a}_j^+ = 6$ case, so that $\mathbf{a}_i^{+\prime}\mathbf{a}_j^+ = \pm 2$. For example, the information matrix for the design having four continuous factors and two categorical factors is given by

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 14 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 2 & -2 \\ 0 & 0 & 10 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 10 & 0 & -2 & 2 \\ 0 & 0 & 0 & 0 & 10 & 2 & 2 \\ 0 & 2 & -2 & -2 & 2 & 14 & 2 \\ 0 & -2 & -2 & 2 & 2 & 2 & 14 \end{pmatrix}. \qquad (2)$$

We note that a potential drawback to this procedure is that we must evaluate the determinants of 2^{2c} information matrices and, if c is very large, this might be problematic. We have found that the procedure takes less than a minute for $c \leq 20$ on a PC. For larger values of c, we use an exchange algorithm to quickly optimize the choice of $z_1, \ldots, z_c$ and $b_1, \ldots, b_c$. The latter procedure is not guaranteed to produce the best DSD-augment design, but we have observed that, with multiple random starts, it usually does.

The DSD-augmentation procedure is a heuristic that does not guarantee that the augmentation is D-optimal, subject to the constraint that all estimates of main effects are independent of second-order effects. As a check on the optimality of the heuristic, we computed constrained D-optimal augmentations using a variation of the coordinate-exchange algorithm (Meyer and Nachtsheim (1995)). The results are summarized in Table 4, in the column marked "Efficiency of DSD-Augment vs. Optimal DSD-Augment." We see from the Table that the DSD-augment method was also constrained D-optimal in every case. As a result, we conjecture that the augmentation procedure described will always produce the best such augmentation in terms of

TABLE 2. Design Matrix Structure for Augmented Design with $m + 1$ Through $m + 4$ Factors.
For the DSD-augment method, $n = n_1 + 2$ and **B** is $2 \times c$; for the ORTH-Augment method, $n = n_1 + 4$ and **B** is $4 \times c$

Run	Continuous factors						Categorical factors			
(i)	$\mathbf{1}$	$\mathbf{d}_1^+$	$\mathbf{d}_2^+$	$\mathbf{d}_2^+$	$\cdots$	$\mathbf{d}_m^+$	$\mathbf{a}_1^+$	$\mathbf{a}_2^+$	$\mathbf{a}_3^+$	$\mathbf{a}_4^+$
1	1	0	±1	±1	$\cdots$	±1	±1	±1	±1	±1
2	1	0	∓1	∓1	$\cdots$	∓1	∓1	∓1	∓1	∓1
3	1	±1	0	±1	$\cdots$	±1	±1	±1	±1	±1
4	1	∓1	0	∓1	$\cdots$	∓1	∓1	∓1	∓1	∓1
5	1	±1	±1	0	$\cdots$	±1	±1	±1	±1	±1
6	1	∓1	∓1	0	$\cdots$	∓1	∓1	∓1	∓1	∓1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2m-1$	1	±1	±1	±1	$\cdots$	0	±1	±1	±1	±1
$2m$	1	∓1	∓1	∓1	$\cdots$	0	∓1	∓1	∓1	∓1
$2m+1$	1	±1	±1	±1	$\cdots$	±1	$z_{1,1}$	±1	±1	±1
$2m+2$	1	∓1	∓1	∓1	$\cdots$	∓1	$z_{1,2}$	∓1	∓1	∓1
$2m+3$	1	±1	±1	±1	$\cdots$	±1	±1	$z_{2,1}$	±1	±1
$2m+4$	1	∓1	∓1	∓1	$\cdots$	∓1	∓1	$z_{2,2}$	∓1	∓1
$2m+5$	1	±1	±1	±1	$\cdots$	±1	±1	±1	$z_{3,1}$	±1
$2m+6$	1	∓1	∓1	∓1	$\cdots$	∓1	∓1	∓1	$z_{3,2}$	∓1
$2m+7$	1	±1	±1	±1	$\cdots$	±1	±1	±1	±1	$z_{4,1}$
$2m+8$	1	∓1	∓1	∓1	$\cdots$	∓1	∓1	∓1	∓1	$z_{4,1}$
$2m+9$	1	±1	±1	±1	$\cdots$	±1	±1	±1	±1	±1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_1	1	∓1	∓1	∓1	$\cdots$	∓1	∓1	∓1	∓1	∓1
n_1+1	1	0	0	0	0	0				
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$			**B**	
n	1	0	0	0	0	0				

D-efficiency. We also provide the maximum absolute element of the constant-term row (row 1) of the alias matrix in the rightmost column. The largest value occurred for $m = 4$ and $c = 1$, and for $m = 5$ and $c = 1$, in which case the maximum absolute entry in the first row of the alias matrix was $1/7$.

TABLE 3. Alias Matrix for DSD-Augmented Design Having 4 Continuous Factors and 2 Categorical Factors

Main effects	β_{12}	β_{13}	β_{14}	β_{23}	β_{24}	β_{34}	β_{15}	β_{16}	β_{25}	β_{26}	β_{35}	β_{36}	β_{45}	β_{46}	β_{56}
β_0	0	0	0	0	0	0	0.1429	−0.1429	−0.1429	−0.1429	−0.1429	0.1429	0.1429	0.1429	0.1429
β_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
β_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
β_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
β_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
β_5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
β_6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TABLE 4. Design Diagnostics for ORTH-Augmented and DSD-Augmented Designs for 4 through 12 Continuous Factors and 1 Through 4 Categorical Factors

n_{DSD}	n_{ORTH}	m	c	p	Efficiency of DSD-augment vs. optimal DSD-augment	Normalized (per run) efficiency of DSD-augment vs. ORTH-augment	Maximum Value of DSD-augment alias matrix
14	14	4	1	5	1	0.98	0.1429
14	16	4	2	6	1	1.04	0.1429
18	20	4	3	7	1	1.03	0.1111
18	20	4	4	8	1	1.01	0.1111
14	14	5	1	6	1	0.98	0.1429
18	20	5	2	7	1	1.04	0.1111
18	20	5	3	8	1	1.03	0.1111
22	24	5	4	9	1	1.02	0.0909
18	18	6	1	7	1	0.99	0.1111
18	20	6	2	8	1	1.05	0.1111
22	24	6	3	9	1	1.03	0.0909
22	24	6	4	10	1	1.02	0.0909
18	18	7	1	8	1	0.99	0.1111
22	24	7	2	9	1	1.05	0.0909
22	24	7	3	10	1	1.03	0.0909
26	28	7	4	11	1	1.02	0.0769
22	22	8	1	9	1	0.99	0.0909
22	24	8	2	10	1	1.05	0.0909
26	28	8	3	11	1	1.03	0.0769
26	28	8	4	12	1	1.03	0.0769
22	22	9	1	10	1	0.99	0.0909
26	28	9	2	11	1	1.05	0.0769
26	28	9	3	12	1	1.04	0.0769
30	32	9	4	13	1	1.03	0.0667
26	26	10	1	11	1	0.99	0.0769
26	28	10	2	12	1	1.05	0.0769
30	32	10	3	13	1	1.03	0.0667
30	32	10	4	14	1	1.03	0.0667
26	26	11	1	12	1	0.99	0.0769
30	32	11	2	13	1	1.04	0.0667
30	32	11	3	14	1	1.04	0.0667
34	36	11	4	15	1	1.03	0.0588
30	30	12	1	13	1	1.00	0.0667
30	32	12	2	14	1	1.04	0.0667
34	36	12	3	15	1	1.03	0.0588
34	36	12	4	16	1	1.03	0.0588

For purposes of illustration, we display the DSD-augmented designs for $m = 4$ continuous factors and $c = 1, \ldots, 4$ categorical factors in Figure 1a.

3. Orthogonal Column Augmentation: ORTH-Augment Procedure

We now consider the problem of adding two-level categorical columns to a DSD having m factors in a way that preserves orthogonality of the main effects design. An optimal column-augmentation scheme for $c \leq 4$ is accomplished via the following steps. (The $c > 4$ case is examined at the end of this section.)

1. Augment the rows of the design as indicated in Table 2. The base DSD is augmented with four center points, while the jth added column $\mathbf{a}_j$ is augmented by $\mathbf{b}_j$, a 4×1 vector having a one in

(a) DSD-augmented designs

Run	m = 4, c = 1	m = 4, c = 2	m = 4, c = 3	m = 4, c = 4
1	0++++	0+++++	0++++++	0+++++++
2	0----	0-----	0------	0-------
3	+0+--	+0+--+	-0++-+-	-0++-+--
4	-0-++	-0-++-	+0--+-+	+0--+-++
5	++0+-	++0+--	--0++-+	--0++-+-
6	--0-+	--0-++	++0--+-	++0--+-+
7	+-+0+	+-+0+-	---0++-	---0++-+
8	-+-0-	-+-0-+	+++0--+	+++0--+-
9	+--++	+--+++	-+--+++	-+--+++-
10	-++--	-++---	+-++---	+-++---+
11	++--+	++--++	--+--++	--+--+++
12	--++-	--++--	++-++--	++-++---
13	0000-	0000--	-+-+--+	-+-+--++
14	0000+	0000++	+-+-++-	+-+-++--
15			-++-+--	-++-+--+
16			+--+-++	+--+-++-
17			0000---	0000----
18			0000+++	0000++++

(b) ORTH-augmented designs

Run	m = 4, c = 1	m = 4, c = 2	m = 4, c = 3	m = 4, c = 4
1	0++++	0+++++	0++++++	0+++++++
2	0----	0-----	0------	0-------
3	+0+--	+0+--+	-0++-+-	-0++-+--
4	-0-++	-0-++-	+0--+-+	+0--+-++
5	++0+-	++0+--	--0++-+	--0++-+-
6	--0-+	--0-++	++0--+-	++0--+-+
7	+-+0+	+-+0+-	---0++-	---0++-+
8	-+-0-	-+-0-+	+++0--+	+++0--+-
9	+--++	+--+++	-+--+++	-+--+++-
10	-++-+	-++-+-	+-+++--	+-+++--+
11	++--+	++--++	--+--++	--+--+++
12	--++-	--++-+	++-+++-	++-+++--
13	0000-	0000--	-+-+--+	-+-+--++
14	0000-	0000--	+-+-+++	+-+-+++-
15		0000-+	-++-+--	-++-+--+
16		0000+-	+--+-++	+--+-+++
17			0000---	0000---+
18			0000--+	0000--+-
19			0000-+-	0000-+--
20			0000+--	0000+---

FIGURE 1. Designs Constructed by the DSD-Augment Procedure (a) and by the ORTH-Augment procedure (b) for $m = 4$ Continuous Factors and for $c = 1$ Through $c = 4$ Categorical Factors.

the $(5-j)$th row and negative ones elsewhere, for $j = 1, \ldots, c \leq 4$. For $c = 4$, $\mathbf{B}$ the 4×4 matrix $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4)$.)

2. If $c < 4$, the design structure obtained in Step 2 will be similar to that of Table 2, except that the rightmost $4-c$ column(s) of Table 2 would be deleted along with the $2(4-c)$ rows that contained the $z_{i,j}$ values in the deleted column(s).
3. In each of the remaining augmented columns (i.e., for $j = 1, \ldots, c$), set $z_{j,1} = z_{j,2} = 1$.

We refer to this as the *ORTH-augment* method.

TABLE 5. Alias Matrix for Orth-Augmented Design Having 4 Continuous Factors and 2 Categorical Factors

Main effects	β_{12}	β_{13}	β_{14}	β_{23}	β_{24}	β_{34}	β_{15}	β_{16}	β_{25}	β_{26}	β_{35}	β_{36}	β_{45}	β_{46}	β_{56}
β_0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
β_1	0	0	0	0	0	0	0.2	0.2	−0.2	0.2	−0.2	−0.2	0.2	−0.2	0.4
β_2	0	0	0	0	0	0	−0.2	0.2	0.2	0.2	0.2	−0.2	−0.2	−0.2	0
β_3	0	0	0	0	0	0	−0.2	−0.2	0.2	−0.2	0.2	0.2	−0.2	0.2	−0.4
β_4	0	0	0	0	0	0	0.2	−0.2	−0.2	−0.2	−0.2	0.2	0.2	0.2	0
β_5	−0.125	−0.125	0.125	0.125	−0.125	−0.125	0	0.25	0	0	0	−0.25	0	0	0
β_6	0.125	−0.125	−0.125	−0.125	−0.125	0.125	0.25	0	0	0	−0.25	0	0	0	0

For $c \leq 4$, it is straightforward to show that the information matrix, $\mathbf{X}'\mathbf{X}$, for the first-order model of any design produced by the ORTH-augment method is diagonal, with diagonal elements

$$\mathbf{X}'\mathbf{X}_{i,i} = \begin{cases} n & \text{for } i = 1 \text{ and } i = m+2, \\ & \quad \ldots, m+c+1 \\ n_1 = n-4 & \text{for } i = 2, \ldots, m+1. \end{cases}$$

To verify orthogonality, do as follows:

1. We know that $\mathbf{1}'\mathbf{d}_i^+ = 0$ for $i = 2, \ldots, m+1$, and we know that $\mathbf{d}_i^{+\prime}\mathbf{d}_j^+ = 0$, for $i \neq j$ because $\mathbf{d}_i^+$ and $\mathbf{d}_j^+$ are columns from the original DSD (without the center point run), augmented by four center points.
2. Consider any augmented column $\mathbf{a}_j^+ = (\mathbf{a}_j', \mathbf{b}_j')'$. We know that, with $z_{j,1} = z_{j,2} = 0$, $\mathbf{a}_j'$ is balanced, so changing these two values to $+1$ gives $\mathbf{1}_{n_1\times 1}'\mathbf{a}_j' = +2$. But $\mathbf{1}_{4\times 1}'\mathbf{b}_j = -2$ and we therefore have $\mathbf{1}'\mathbf{a}_j^+ = 0$.
3. Consider now the $\mathbf{d}_i^{+\prime}\mathbf{a}_j^+$ cross product for $i = 2, \ldots, m+1$ and $j = m+2, \ldots, m+c+1$. $\mathbf{d}_i^{+\prime}\mathbf{a}_j^{+\prime} = (\mathbf{d}_i', \mathbf{0}_{4\times 1}')(\mathbf{a}_j', \mathbf{b}_j')' = \mathbf{d}_i'\mathbf{a}_j'$. Because $\mathbf{d}_i'\mathbf{a}_j = 0$ in the original DSD with $z_{i,1} = z_{i,2} = 0$ and because the rows containing $z_{i,1}$ and $z_{i,2}$ are a fold-over pair (ignoring column $\mathbf{a}_j^+$), it follows that $\mathbf{d}_i'\mathbf{a}_j = \pm(z_{i,1} - z_{i,2})$. With $z_{i,1} = z_{i,2} = 1$, we have $\mathbf{d}_i^{+\prime}\mathbf{a}_j^+ = 0$.
4. Now consider the $\mathbf{a}_i^{+\prime}\mathbf{a}_j^+$ crossproduct for $i \neq j$. $\mathbf{a}_i^{+\prime}\mathbf{a}_j^+ = (\mathbf{a}_i', \mathbf{b}_i')(\mathbf{a}_j', \mathbf{b}_j')' = \mathbf{a}_i'\mathbf{a}_j + \mathbf{b}_i'\mathbf{b}_j = \mathbf{a}_i'\mathbf{a}_j$, since $\mathbf{b}_i'\mathbf{b}_j = 0$ by construction. Now $\mathbf{a}_i'\mathbf{a}_j = \pm(z_{i,1} - z_{i,2}) + \pm(z_{j,1} - z_{j,2})$ because the rows containing $z_{i,1}$ and $z_{i,2}$ are a fold-over pair (ignoring column $\mathbf{a}_i^+$) and the rows containing $z_{j,1}$ and $z_{j,2}$ are another fold-over pair (ignoring column $\mathbf{a}_j^+$). Setting $z_{i,1} = z_{i,2} = z_{j,1} = z_{j,2} = 1$, we have $\mathbf{a}_i'\mathbf{a}_j = 0$. Thus, $\mathbf{a}_i^{+\prime}\mathbf{a}_j^+ = 0$.

By orthogonality, this augmentation is easily seen to be D-optimal among all such column augmentations.

We note that, if $c = 1$, then an orthogonal augmentation exists that employs only two center points for the continuous factors and therefore just $n_1 + 2$ runs. This design structure is obtained by deleting the rightmost three columns in Table 2, the rows that contained the $z_{i,j}$ entries in the three deleted columns, and the last two (center point) rows.

For illustration, we have listed the ORTH-augmented DSDs for $m = 4$ continuous factors and $c = 1, \ldots, 4$ categorical factors in Figure 1b.

While orthogonality is a valuable property of the proposed designs, partial aliasing between main effects and interactions involving the categorical factors does exist. The alias matrix for the orthogonal design having four continuous factors and two categorical factors is shown in Table 5. From the table, it is clear that the aliasing structure for the continuous factors is retained. However, all of the main effects may be biased by potential two-factor interactions involving categorical factors. The alias matrix entries range in magnitude from 0 to 0.4, so that, while these numbers are not zero, they are typically small.

It is interesting to compute the D-efficiencies of the DSD-augmented designs relative to those of the corresponding ORTH-augmented designs for the linear main effects model. Because the run sizes are not the same for $c > 1$, the comparisons must be carried out on a normalized (information per run) basis. The normalized D-efficiencies of the DSD-augmented designs relative to the D-optimal (orthogonal) designs for 4 through 12 continuous factors and 1 through 4 categorical factors are listed in Table 4 in the column headed "Normalized (per run) Efficiency of DSD-Augment vs. ORTH-Augment." From the Table, we see that the worst-case D-efficiency of a DSD-based design is 98%. Perhaps surprisingly, the normalized relative efficiencies of the DSD-augmented designs for $c > 1$ are always larger than one. This is due to the fact that the DSD-augmented design uses fewer center points. Thus, on a per-run basis, more information is obtained with regard to linear main effects.

Our results described thus far in this section pertain to the addition of up to four categorical factors. The methods can be extended to more than four factors heuristically by column concatenating the lower-right corner $\mathbf{B}$ matrix as often as required. For example, if an additional six categorical factors are required, a matrix structure similar to that of Table 2 having an additional eight categorical factors, with the lower right section of the table replaced by $\mathbf{B}^{(2)} = (\mathbf{B},\quad \mathbf{B})$. As before, design structure for six categorical factors would be obtained by deleting the two rightmost columns and the rows that contain the $z_{i,j}$ entries in the two deleted columns.

It is straightforward to show that, for relatively small c, the resulting designs will still be nearly orthogonal. To be specific,

- For $c = 5$ through $c = 8$, each categorical factor added beyond four adds one new nonzero correlation to the upper triangular portion of the information matrix.
- For $c = 9$ through $c = 12$, each categorical factor added beyond eight adds two new nonzero correlations to the upper triangular portion of the information matrix.
- For $c = 13$ through $c = 16$, each categorical factor added beyond eight adds three new nonzero correlations to the upper triangular portion of the information matrix, and so on.

Let $n_B = [c/4]^-$ denote the number of complete $\mathbf{B}$ matrices used to construct the design. Then the number of nonzero entries in the information matrix is $n_{nz} = 2n_B(c - 2n_B - 2)$. The fraction of nonzero entries among the off-diagonal elements of the information matrix is

$$f_{nz} = \frac{2n_B(c - 2n_B - 2)}{(m + c + 1)(m + c).}$$

This fraction approaches 25% as $c \to \infty$, but it is very small for the numbers of categorical factors typically encountered in practice. Table 6 gives the number of nonzero entries and the fraction of nonzero off-diagonal elements in the information matrix as m varies from four to 12 in steps of four and for c ranging from four through 10 in steps of two. From the table, we observe that the worst-case fraction occurs for $m = 4$ and $c = 10$, in which case $f_{nz} = 0.0762$. Again, for $c < 5$, this fraction will always be zero.

TABLE 6. Number and Fraction of Nonzero Entries in the Information Matrix of Orth-Augmented Designs for Selected Values of $m \leq 12$ and $c \leq 10$

m	c	Number of nonzero entries	Fraction of nonzero entries
4	4	0	0.0000
4	6	4	0.0364
4	8	8	0.0513
4	10	16	0.0762
8	4	0	0.0000
8	6	4	0.0190
8	8	8	0.0294
8	10	16	0.0468
12	4	0	0.0000
12	6	4	0.0117
12	8	8	0.0190
12	10	16	0.0316

We note that the inner product between the linear main effects columns and the quadratic main effects columns is equal to +2, so that a small correlation between these columns is present. For the DSD-augment procedure, this inner product is zero.

4. Discussion

In this paper, we have given two conference matrix-based methods for adding two-level categorical factors to DSDs. The first method, termed the DSD-augment method, produces designs that are highly D-efficient and for which all estimates of main effects and second-order effects are independent. Some partial aliasing between two-factor interactions and the constant term remains. The second method, called the ORTH-augment method, leads to D-optimal, orthogonal augmentation for $c \leq 4$ and near-orthogonal augmentation for small $c > 4$. The disadvantage of this approach is that partial aliasing between main effects and two-factor interactions involving the categorical factors exists.

One limitation of these designs is that, as the number of categorical factors increases, the correlations between quadratic effects increase. For the DSD-augment procedure, it can be shown that the correlation between quadratic main effects columns is $1/2-2/(n-4)$. This correlation is increasing in n (or c) and it approaches $1/2$ as $n \to \infty$. For the ORTH-augment designs, this correlation is $2/3 - 2/(n-6)$, which is also increasing—approaching 2/3 in the limit. For a standard DSD design, JN showed that the correlation is $1/3 - 2/(n-3)$, which is bounded above by 1/3. Because this upper bound is lower than those for the DSDs with categorical effects, we conclude that the ability to discriminate among quadratic effects will tend to decrease as we add categorical factors. The development of alternative designs that reduce these correlations is the subject of a future paper.

The choice between the DSD-augment approach and the ORTH-augment approach is clearly up to the experimenter. In our view, the use of a DSD layout for the continuous factors suggests that there is concern about the presence of two-factor interactions. Also, the DSD-augmented designs, which usually require two fewer center point runs than the corresponding ORTH-augmented designs, frequently have higher D-efficiencies per run for the first-order main effects model. For these reasons, we tend to favor the use of the DSD-augmentation when adding categorical factors to DSD designs.

A MATLAB code and a JMP script for constructing DSDs with or without added categorical factors can be obtained from the authors.

References

Box, G. E. P. and Hunter, J. S. (1961). "The 2^{k-p} Fractional Factorial Designs". *Technometrics* 3, pp.449–458.

Hedayat, A. S.; Sloane, N. J. A.; and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications*. New York, NY: Springer.

Jones, B. and Nachtsheim, C. J. (2011). "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects". *Journal of Quality Technology* 43, pp. 1–14.

Meyer, R. K. and Nachtsheim, C. J. (1995). "The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs". *Technometrics* 37, pp. 60–69.

Nguyen, N.-K. (1996). "A Note on the Construction of Near-Orthogonal Arrays with Mixed Levels and Economic Run Size". *Technometrics* 38, pp. 279–283.

Wang, J. C. and Wu, C. F. J. (1992)."Nearly Orthogonal Arrays With Mixed Levels and Small Runs." *Technometrics* 34, pp. 409–422.

Xiao, L.; Lin, D. K. J.; and Fengshan, B. (2012) "Constructing Definitive Screening Designs Using Conference Matrices". *Journal of Quality Technology* 44, pp. 1–7.

Xu, H. (2002). "An Algorithm for Constructing Orthogonal and Nearly-Orthogonal Arrays with Mixed levels and Small Runs". *Technometrics* 44, pp. 356–368.