Noname manuscript No.
(will be inserted by the editor)

Orthogonal Array Based Locally D-optimal Designs for Binary Responses in the Presence of Factorial Effects

Zhongshen Wang · John Stufken

Received: date / Accepted: date

Abstract Wang and Stufken [2020] identified locally D-optimal designs for Generalized Linear Models with factorial effects and one continuous covariate. Using an approximate design approach, the design problem consists of selecting values for the covariate and design weights for each group formed by the various factors. For the logistic and probit link, the optimal designs in Wang and Stufken [2020] use two covariate values for each of the groups and equal weights. We establish that smaller D-optimal designs can often be obtained by using orthogonal arrays so that an optimal design uses only some of the groups with at most two covariate values in those groups. The general theory is illustrated through an application.

Keywords Locally optimal design · D-optimality · factorial experiment · orthogonal array

PACS 62K05 · 62K15

1 Introduction

Factors play an important role in many experiments. Understanding their effects on a response variable can be the primary reason for experimentation. If that response variable is categorical, then the model is typically a Generalized

Research partially supported by NSF grant DMS-1935729.

Zhongshen Wang
Incyte Corporation
Wilmington, DE 19803
E-mail: zwang@incyte.com

John Stufken
University of North Carolina at Greensboro
Greensboro, NC 27412
E-mail: jstufken@uncg.edu

Linear Model (GLM). As in Wang and Stufken [2020], we focus on finding optimal designs for a binary response variable using GLMs in which, besides factorial effects, a controllable continuous covariate can also effect the response. We focus on two frequently used link functions: the logistic link and probit link.

For the scenario described in the previous paragraph, finding an optimal design is equivalent to selecting combinations for the values and levels of the covariate and factors, respectively, that are to be used in the experiment. As noted in Wang and Stufken [2020], an optimal design can improve parameter estimation for a fixed number of runs or reduce costs to achieve a desired level of precision. While there are computational approaches to this problem, which are flexible and useful, theoretical considerations can provide more insight in the structure of optimal designs. For a model without interactions, Stufken and Yang [2012] provided an explicit expression for optimal designs under D-optimality. Then, Tan [2015] obtained smaller D-optimal designs using orthogonal arrays (OAs). However, Stufken and Yang [2012] and Tan [2015] are restricted to models that include all factorial effects up to a certain order. Wang and Stufken [2020] generalized the results and obtained optimal designs for models that, for any order, need not include all interactions of that order.

The optimal designs in Wang and Stufken [2020] under an approximate design approach require the selection of two covariate values in each group formed by the factorial structure, with all points having equal weight. With a large number of factors or with some factors that have a large number of levels, these designs can be quite large. We show that smaller optimal designs can often be found by using orthogonal arrays. Recall that a $N \times k$ array is called an orthogonal array with s levels and strength t if, for every $N \times t$ subarray, all possible combinations of t symbols occur equally often as a row (Hedayat et al. [1999]). We denote such an array as $OA(N, k, s, t)$ or $OA(N, s^k, t)$ where "s^k" indicates that there are k factors with s levels each. To allow the factors to have different levels, a slight modification of $OA(N, s^k, t)$ defines mixed orthogonal arrays. A mixed orthogonal array $OA(N, s_1^{k_1} s_2^{k_2} \cdots s_v^{k_v}, t)$ is a $N \times k$ array, where $k = k_1 + k_2 + \cdots + k_v$ is the total number of factors, such that the first k_1 columns have symbols from $\{1, 2, \ldots, s_1\}$, the next k_2 columns have symbols from $\{1, 2, \ldots s_2\}$, and so on, with the property that in any $N \times t$ subarray every possible t-tuple occurs an equal number of times as a row. We will also write $OA(N, s_1 s_2 \cdots s_k, t)$ if it is unclear which s_l's are equal. Tan [2015] already used OAs for finding smaller designs, but was only able to do this for models that included all factorial effects up to a certain order. As a result, the required orthogonal arrays needed to be of higher strength, leading to a smaller reduction in the size of optimal designs. For example, if a second-order model was under consideration, designs in Tan [2015] would require the use of strength 4 OAs; if the model includes only some two-factor interactions, we will see that we may be able to use OAs of strength 2 that have some additional properties. The latter arrays require, generally, far fewer runs than the strength 4 arrays, so that fewer groups are needed to build an optimal designs. Precise results will be presented in Section 3.

Hedayat [1989] (see also Hedayat [1990]) introduced the concept of strength $t+$ OAs. An $OA(N, k, s, t+)$ is an $OA(N, k, s, t)$ that is not of strength $t+1$, but has one or more subarrays that form an $OA(N, k', s, t+1)$. In fractional factorial experiments, strength $t+$ OAs can be desirable because they allow orthogonal estimation of factorial effects for a wider class of models than arbitrary strength t OAs while being more economical than strength $t+1$ OAs. A similar idea appears in recent work on so called strong OAs (He and Tang [2013]) for use in computer experiments. The notion of strong OAs of strength two plus was introduced in He et al. [2018] (see also Zhou and Tang [2019] and Shi and Tang [2019]). Unlike these authors, who seek properties for strong OAs that exceed those of strength two without requiring those of strength three, we do not require the same type of structure that they need for space-filling properties for their designs.

Under effect sparsity, and supported by subject matter knowledge, it is plausible that one would have some inkling which interactions, small in number, might be important. We show that the use of OAs can lead to smaller optimal designs under such models, which were also studied in Wang and Stufken [2020]. These models need not include all interactions of a given order. We establish our results by showing that the information matrices for OA-based designs are equal to those of the larger D-optimal designs obtained in Wang and Stufken [2020]. In Section 2, we formally present the model and the structure of the information matrix for the optimal designs as obtained in Wang and Stufken [2020]. Thereafter, in Section 3, we present the main results for obtaining smaller D-optimal designs. Subsequently, in Section 4 we illustrate the theoretical results by using the Electrostatic Discharge (ESD) failure voltage experiment in Whitman et al. [2006]. We end with a summary and discussion in the final section.

2 The model and information matrix

For clarity, and since we need to introduce the notation, we now describe the model considered in Wang and Stufken [2020]. We also present their main result about optimal designs and the structure of the information matrix for those optimal designs. In doing so, we refer to experimental units as subjects. Considering L factors with $s_1, s_2, \cdots, s_L$ levels respectively, each subject could belong to one of $s = s_1 \cdot s_2 \cdots s_L$ groups. Further, we assume that the slope for the continuous covariate in the linear predictor is the same across all of the groups. The model under consideration can then be written as

$$\begin{aligned} Prob(Y_{i_1 i_2 \ldots i_L u} = 1) = P(&\alpha_0 + \alpha_1^{i_1} + \cdots + \alpha_L^{i_L} \\ &+ \sum_{t=2}^{L} \sum_{(l_1, l_2, \cdots, l_t) \in G_t} \alpha_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}} + \beta x_{i_1 i_2 \cdots i_L u}), \end{aligned} \quad (1)$$

where $Y_{i_1 i_2 \ldots i_L u}$ is the response from the u^{th} subject in group $(i_1, i_2, \cdots, i_L)$, $i_l = 1, \cdots, s_l$; $u = 1, \cdots, M_{i_1 i_2 \cdots i_L}$, and $M_{i_1 i_2 \cdots i_L}$ is the number of subjects in group $(i_1, i_2, \cdots, i_L)$. Further, $P(\cdot)$ is a cumulative distribution function; α_0 is the overall mean, $\alpha_l^{i_l}$ is the effect of the i_l^{th} level of factor l, $\alpha_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}}$ is the effect of the level combination $(i_{l_1}, i_{l_2}, \cdots, i_{l_t})$ for the t-th order effect for the factors $(l_1, l_2, \cdots, l_t)$, $t = 2, \cdots, L$; and G_t is a set of t-tuples representing the t-th order effects included in the model. Since we take all main effects to be in the model, we also write $G_1 = \{1, 2, \cdots, L\}$. Moreover, β is the common slope parameter; and $x_{i_1 i_2 \cdots i_L u}$ is the covariate value for the uth subject in group $(i_1, i_2, \cdots, i_L)$, which must be in the design region denoted by $[L_{i_1 i_2 \cdots i_L}, U_{i_1 i_2 \cdots i_L}]$. The endpoints $L_{i_1 i_2 \cdots i_L}$ and $U_{i_1 i_2 \cdots i_L}$ can be $-\infty$ or ∞, respectively.

We can also write the model in (1) in vector notation as

$$Prob(Y_{i_1 i_2 \ldots i_L u} = 1) = P((\boldsymbol{X^{i_1 \cdots i_L u}})^T \boldsymbol{\theta}). \tag{2}$$

Here $\boldsymbol{\theta} = (\alpha_0, \boldsymbol{\alpha_1}^T, \cdots, \boldsymbol{\alpha_L}^T, \cdots, \boldsymbol{\alpha_{l_1 l_2}}^T, \cdots, \boldsymbol{\alpha_{l_1 l_2 \cdots l_t}}^T, \cdots, \beta)^T$, and terms in $\boldsymbol{\theta}$ correspond to those in the model where for any $1 \le l_1 < \cdots < l_t \le L$ and $t = 1, \cdots, L$, $\boldsymbol{\alpha_{l_1 \cdots l_t}} = (\alpha_{l_1 \cdots l_t}^{1 \cdots 1}, \cdots, \alpha_{l_1 \cdots l_t}^{1 \cdots s_{l_t}}, \cdots, \alpha_{l_1 \cdots l_t}^{s_{l_1} \cdots s_{l_t}})^T$. Further, $\boldsymbol{X^{i_1 \cdots i_L u}} = (1, (\boldsymbol{X_1^{i_1}})^T, \cdots, (\boldsymbol{X_L^{i_L}})^T, \cdots, (\boldsymbol{X_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T, \cdots, (\boldsymbol{X_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}}})^T, \cdots, x_{i_1 i_2 \cdots i_L u})^T$, where terms in $\boldsymbol{X^{i_1 \cdots i_L u}}$ correspond again to those in the model and $\boldsymbol{X_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}}}$ is a $(s_{l_1} \times \cdots \times s_{l_t}) \times 1$ vector with a 1 in position $(i_{l_1}, \cdots, i_{l_t})$ and 0's elsewhere.

While the model is relatively easy to understand as presented, it is overparameterized. Hence, we follow Wang and Stufken [2020] by introducing a bit more notation and by reparameterizing the model. Some subjects in the same group could be assigned to the same covariate value, and we need a notation for the distinct numbers of covariate values within the groups. We use $m_{i_1 i_2 \cdots i_L}$ to denote the distinct number of covariate values used in group $(i_1, i_2, \cdots, i_L)$. Also, let $n_{i_1 i_2 \cdots i_L j}$ denote the number of subjects in group $(i_1, i_2, \cdots, i_L)$ who are assigned to the jth covariate value in that group, $j = 1, \cdots, m_{i_1 i_2 \cdots i_L}$. With n as the total number of subjects in the experiment, replace $n_{i_1 i_2 \cdots i_L j}/n$ by the design weights $w_{i_1 i_2 \cdots i_L j}$. Then a design can be written as

$$\xi = \{(\boldsymbol{X^{i_1 \cdots i_L j}}, w_{i_1 i_2 \cdots i_L j}), i_l = 1, ..., s_l,\ l = 1, ..., L,\ j = 1, ..., m_{i_1 i_2 \cdots i_L}\}.$$

By allowing the $w_{i_1 i_2 \cdots i_L j}$'s to take any non-negative values that sum to 1, the design becomes an approximate design, and finding an optimal design no longer depends on the value of n.

For model (2) with approximate design ξ, the corresponding information matrix for $\boldsymbol{\theta}$ is

$$I_\xi(\boldsymbol{\theta}) = \sum_{i_1}^{s_1} \cdots \sum_{i_L}^{s_L} \sum_{j=1}^{m_{i_1 i_2 \cdots i_L}} w_{i_1 i_2 \cdots i_L j} I_{\boldsymbol{X^{i_1 \cdots i_L j}}}(\boldsymbol{\theta}), \tag{3}$$

where $I_{\boldsymbol{X^{i_1 \cdots i_L j}}}(\boldsymbol{\theta})$ is the information matrix for the design that places all weight on the single design point $\boldsymbol{X^{i_1 \cdots i_L j}}$. If interest lies in the estimation of a function of $\boldsymbol{\theta}$, say $g(\boldsymbol{\theta})$, then the information matrix for $g(\boldsymbol{\theta})$ is

$$I_\xi(g(\boldsymbol{\theta})) = (\Sigma_\xi(\boldsymbol{\theta}))^{-1} = \left(\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right) I_\xi(\boldsymbol{\theta})^- \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right)^T \right)^{-1}. \tag{4}$$

For design selection, we focus on D-optimality. A design ξ is called D-optimal for $g(\boldsymbol{\theta})$ if it minimizes the determinant of the covariance matrix $\Sigma_\xi(\boldsymbol{\theta})$, or equivalently, maximizes the determinant of the information matrix $I_\xi(g(\boldsymbol{\theta}))$. Note that $I_\xi(g(\boldsymbol{\theta}))$ depends on $\boldsymbol{\theta}$, which is unknown before conducting the experiment. We therefore consider locally D-optimal designs where $\boldsymbol{\theta}$ is replaced by its best guess in $I_\xi(g(\boldsymbol{\theta}))$. The values of these best guesses for $\boldsymbol{\theta}$ are usually based on prior experiments.

Since the model in (2) is overparameterized, we consider a maximal set of linearly independent estimable functions of $\boldsymbol{\theta}$. D-optimality is invariant under reparameterization, so that optimal design results are invariant to the choice of this maximal set. Let $g(\boldsymbol{\theta}) = B\boldsymbol{\theta} = \boldsymbol{\eta}$ denote one particular maximal set. We will only consider models that contain the overall mean, all of the main effects, a slope parameter, and some of the two-factor interactions, so that $G_t = \emptyset$ for $t \geq 3$. In that case,

$$r = rank(B) = 1 + \sum_{l=1}^{L}(s_l - 1) + \sum_{(l_1,l_2)\in G_2}(s_{l_1} - 1)(s_{l_2} - 1) + 1. \tag{5}$$

Additionally, define $c_{i_1\cdots i_L j} = (\boldsymbol{X}^{\boldsymbol{i_1\cdots i_L j}})^T\boldsymbol{\theta}$, which belongs to the design region $[D_{i_1\cdots i_L 1}, D_{i_1\cdots i_L 2}]$ induced by the region $[L_{i_1 i_2\cdots i_L}, U_{i_1 i_2\cdots i_L}]$ for $x_{i_1 i_2\cdots i_L j}$. Then, Lemma 1 provides a locally D-optimal design for $\boldsymbol{\eta}$ under models as in (2).

Lemma 1 *(Theorem 1 of Wang and Stufken [2020]) For a model of the form (2) with the logistic or probit link, if* $\{c^*, -c^*\} \subset [D_{i_1\cdots i_L 1}, D_{i_1\cdots i_L 2}]$ *for all groups* $(i_l, ..., i_L)$*, where* $c^* > 0$ *maximizes* $f(c) = c^2(\Psi(c))^r$ *on* $(-\infty, \infty)$*, the design* $\xi^* = \{(c_{i_1\cdots i_L 1} = c^*, w_{i_1\cdots i_L 1} = \frac{1}{2s}), (c_{i_1\cdots i_L 2} = -c^*, w_{i_1\cdots i_L 2} = \frac{1}{2s}), i_l = 1, ..., s_l, l = 1, ..., L\}$ *is a locally D-optimal design for* $\boldsymbol{\eta}$ *. Here* $s = s_1 \times \cdots \times s_L$ *and* $\Psi(x)$ *is given by*

$$\Psi(x) = \begin{cases} \frac{e^x}{(1+e^x)^2}, & \textit{for the logistic link} \\ \frac{[\Phi'(x)]^2}{\Phi(x)(1-\Phi(x))}, & \textit{for the probit link} \end{cases}. \tag{6}$$

Remark 1 Observe that Lemma 1 requires two support points for each of the s groups formed by the L factors. Thus, the optimal designs in Lemma 1 have $2s$ support points. This number becomes quite large if either L or any of the s_i's are large.

Remark 2 Lemma 1 does not only hold for the special case in this paper, but is valid for general sets G_t. However, as noted in Remark 3, the formula for the rank r in Wang and Stufken [2020] needs some correction in that case.

Remark 3 The expression for the rank r in (5) is in agreement with that in Equation (6) of Wang and Stufken [2020], but for the special case that $G_t = \emptyset$ for $t \geq 3$. However, the general expression in Wang and Stufken [2020] is incorrect. To correct it, G_t should be replace by H_t as defined in their Equation (8). For the special case that we consider here, $G_t = H_t$ for all t, so that this correction makes no difference.

The reparametrization of model (2) in vector notation (cf. Wang and Stufken [2020]) is, for the special case that $G_t = \emptyset$ for $t \geq 3$,

$$Prob(Y_{i_1 i_2 \ldots i_L j} = 1) = P((\boldsymbol{Z}^{\boldsymbol{i_1 \cdots i_L j}})^T \boldsymbol{\theta_1}), \tag{7}$$

with $\boldsymbol{\theta_1} = (\gamma_0, \boldsymbol{\gamma_1}^T, \cdots, \boldsymbol{\gamma_L}^T, \cdots, \boldsymbol{\gamma_{l_1 l_2}}^T, \cdots, \beta)^T$, where $\boldsymbol{\gamma_l} = (\gamma_l^1, \cdots, \gamma_l^{s_l - 1})^T$ and, for $(l_1, l_2) \in G_2$, $\boldsymbol{\gamma_{l_1 l_2}} = (\gamma_{l_1 l_2}^{11}, \cdots, \gamma_{l_1 l_2}^{(s_{l_1}-1)(s_{l_2}-1)})^T$. Note that the length of $\boldsymbol{\theta_1}$ is equal to r in (5). Further, we define $\boldsymbol{Z}^{\boldsymbol{i_1 \cdots i_L j}} = (1, (\boldsymbol{Z_1^{i_1}})^T, \cdots, (\boldsymbol{Z_L^{i_L}})^T, \cdots, (\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T, \cdots, x_{i_1 i_2 \cdots i_L j})^T$, where for each factor l,

$$\boldsymbol{Z_l^{i_l}} = \begin{cases} (-\frac{1}{s_l - 1}, \cdots, -\frac{1}{s_l - 1}, 1, -\frac{1}{s_l - 1}, \cdots, -\frac{1}{s_l - 1})^T, & \text{for } 1 \leq i_l \leq s_l - 1 \text{ and} \\ & \text{the 1 is in position } i_l \\ (-\frac{1}{s_l - 1}, \cdots, -\frac{1}{s_l - 1})^T, & \text{for } i_l = s_l \end{cases}$$

and $\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}} = \boldsymbol{Z_{l_1}^{i_{l_1}}} \otimes \boldsymbol{Z_{l_2}^{i_{l_2}}}$, where the notation $\otimes$ denotes the Kronecker product.

With $\boldsymbol{D}^{\boldsymbol{i_1 \cdots i_L j}} = (1, (\boldsymbol{Z_1^{i_1}})^T, \cdots, (\boldsymbol{Z_L^{i_L}})^T, \cdots, (\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T, \cdots, x_{i_1 i_2 \cdots i_L j})^T$, it follows as in Wang and Stufken [2020] that the information matrix for $\boldsymbol{\theta_1}$ takes the form $A(\boldsymbol{\theta_1}) M_\xi(\boldsymbol{\theta_1}) A^T(\boldsymbol{\theta_1})$ for a matrix $A(\boldsymbol{\theta_1})$ that does not depend on design ξ and with

$$M_\xi(\boldsymbol{\theta_1}) = \sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \sum_{j=1}^{m_{i_1 \cdots i_L}} w_{i_1 \cdots i_L j} \Psi(c_{i_1 \cdots i_L j}) \boldsymbol{D}^{\boldsymbol{i_1 \cdots i_L j}} (\boldsymbol{D}^{\boldsymbol{i_1 \cdots i_L j}})^T. \tag{8}$$

Hence, if $M_{\xi_1}(\boldsymbol{\theta_1}) = M_{\xi_2}(\boldsymbol{\theta_1})$ for designs ξ_1 and ξ_2, then they have identical information matrices for $\boldsymbol{\theta_1}$. This holds for any link function $\Psi(x)$. For the logistic or probit link, Wang and Stufken [2020] established $M_{\xi^*}(\boldsymbol{\theta_1})$ for their optimal design ξ^* from Lemma 1.

Lemma 2 *(Lemma 1 of Wang and Stufken [2020]) For the logistic or probit link, $M_{\xi^*}(\boldsymbol{\theta_1})$ is equal to $\Psi(c^*)$ times a block-diagonal matrix with (1) the top-left element equal to 1; (2) the bottom-right element equal to $(c^*)^2$; (3) the block corresponding to $\boldsymbol{\gamma_l}^T$ equal to $B_l = \frac{1}{(s_l - 1)^2}(s_l I - J)$, where J is a matrix of ones; and (4) the block corresponding to $\boldsymbol{\gamma_{l_1 l_2}}$ equal to*

$$B_{l_1 l_2} = B_{l_1} \otimes B_{l_2}.$$

3 Main results

As observed in Remark 1, the optimal designs ξ^* in Lemma 1 can have a large number of support points. Smaller optimal designs can be found, but their structure depends on the interactions that are present in model (1), that is, on the sets G_t. As already noted in Section 2, we restrict attention to models that include the overall mean, all of the main effects, the slope parameter, and some two-factor interactions. We first present a general result for such a model with an arbitrary number of two-factor interactions. The set G_2 can be any set of pairs for this result. We write $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^g}$ for the corresponding parameter vector (see Theorem 1). After presenting results for a few special choices of G_2 as corollaries, we present additional results to further reduce the number of support points for these special choices. For each case, to show that a design ξ is optimal for a vector $\boldsymbol{\theta_1}$, we show that $M_\xi(\boldsymbol{\theta_1})$ and $M_{\xi^*}(\boldsymbol{\theta_1})$ for design ξ^* in Lemma 2 are identical. We consider the following special cases:

(i) $G_2 = \{(1,2)\}$. So, $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^1}$, say, contains only one two-factor interaction, which we have taken without loss of generality to be γ_{12}, besides the overall mean, main effects and slope parameter (see Corollary 1, Theorems 2 and 3).

(ii) $G_2 = \{(1,2),(1,3)\}$. So, $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^2}$, say, contains two two-factor interactions that have one factor in common (see Corollary 2 and Theorem 4).

(iii) $G_2 = \{(1,2),(3,4)\}$. So, $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^3}$, say, contains also two two-factor interactions, but now involving four different factors (see Corollary 3 and Theorem 5).

We start with the general result, presenting its proof in the Appendix.

Theorem 1 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^g}$ consisting of an overall mean, all main effects, the slope parameter and two-factor interactions specified by a set G_2 of pairs. Define sets C_3 and C_4 such that*

$$C_3 = \{(j_1,j_2,j_3) | 1 \le j_1 < j_2 < j_3 \le L \text{ such that at least one of } (j_1,j_2), (j_1,j_3) \text{ and } (j_2,j_3) \text{ is in } G_2\} \quad \text{and}$$

$$C_4 = \{(j_1,j_2,j_3,j_4) | 1 \le j_1 < j_2 < j_3 < j_4 \le L \text{ and they can be grouped into 2 pairs that are both in } G_2\}.$$

Let H be the collection of rows for an $OA(N, s_1 \cdots s_L, 2+)$ with the property that for any triplet $(j_1,j_2,j_3) \in C_3$ the corresponding columns in H form an $OA(N, s_{j_1}s_{j_2}s_{j_3}, 3)$ and for any quadruplet $(j_1,j_2,j_3,j_4) \in C_4$ the corresponding columns in H form an $OA(N, s_{j_1}s_{j_2}s_{j_3}s_{j_4}, 4)$. Define

$$\xi_g = \{(c_{i_1\cdots i_L 1} = c^*, w_{i_1\cdots i_L 1} = \frac{1}{2N}), (c_{i_1\cdots i_L 2} = -c^*, w_{i_1\cdots i_L 2} = \frac{1}{2N}), \text{ for all } (i_1,\cdots,i_L) \in H\}.$$

Then, ξ_g is D-optimal for $\boldsymbol{\theta_1^g}$.

We emphasize that the optimality result in Theorem 1 requires that c^* and $-c^*$ are in the design regions $[D_{i_1\cdots i_L 1}, D_{i_1\cdots i_L 2}]$ for all groups $(i_1,\cdots,i_L) \in H$. This is less restrictive than what is needed in Wang and Stufken [2020]. Results in the next theorems provide even more flexibility, but all cases have the constraint that the optimality result only holds if the design points are in the design region. If the design region does not allow the required design structure, then an algorithm like that in Lukemire et al. [2019] can be considered to obtain D-optimal designs.

For special cases with few interactions, we formulate the following corollaries of Theorem 1.

Corollary 1 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^1}$, that is, with γ_{12} as the only interaction effect. Let H be the collection of rows for an $OA(N, s_1 \cdots s_L, 2+)$ with the property that columns $(1, 2, j)$, $j \geq 3$, form an $OA(N, s_1 s_2 s_j, 3)$. Define*

$$\xi_1 = \{(c_{i_1 \cdots i_L 1} = c^*, w_{i_1 \cdots i_L 1} = \frac{1}{2N}), (c_{i_1 \cdots i_L 2} = -c^*, w_{i_1 \cdots i_L 2} = \frac{1}{2N}),$$
$$\textit{for all } (i_1, \cdots, i_L) \in H\}.$$

Then, ξ_1 is D-optimal for $\boldsymbol{\theta_1^1}$.

Corollary 2 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^2}$, that is, with γ_{12} and γ_{13} as the only two interaction effects. Let H be the collection of rows for an $OA(N, s_1 \cdots s_L, 2+)$ with the property that columns $(1, 2, j)$, $j \geq 3$, and $(1, 3, j)$, $j \geq 4$, form an $OA(N, s_1 s_2 s_j, 3)$ and $OA(N, s_1 s_3 s_j, 3)$, respectively. Define*

$$\xi_2 = \{(c_{i_1 \cdots i_L 1} = c^*, w_{i_1 \cdots i_L 1} = \frac{1}{2N}), (c_{i_1 \cdots i_L 2} = -c^*, w_{i_1 \cdots i_L 2} = \frac{1}{2N}),$$
$$(i_1, \cdots, i_L) \in H\}.$$

Then ξ_2 is D-optimal for $\boldsymbol{\theta_1^2}$.

Corollary 3 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^3}$, that is, with γ_{12} and γ_{34} as the only two interaction effects. Let H be the collection of rows for an $OA(N, s_1 \cdots s_L, 2+)$ with the property that columns $(1, 2, j)$ and $(3, 4, j)$, $j \geq 5$, form an $OA(N, s_1 s_2 s_j, 3)$ and $OA(N, s_3 s_4 s_j, 3)$, respectively, and that columns $(1, 2, 3, 4)$ form an $OA(N, s_1 s_2 s_3 s_4, 4)$. Define*

$$\xi_3 = \{(c_{i_1 \cdots i_L 1} = c^*, w_{i_1 \cdots i_L 1} = \frac{1}{2N}), (c_{i_1 \cdots i_L 2} = -c^*, w_{i_1 \cdots i_L 2} = \frac{1}{2N}),$$
$$(i_1, \cdots, i_L) \in H\}.$$

Then ξ_3 is D-optimal for $\boldsymbol{\theta_1^3}$.

Two other cases are worth mentioning: (i) the model has no interactions, and (ii) the model includes all two-factor interactions. For (i), both C_3 and C_4 are empty, and, based on Theorem 1, any OA of strength 2 is sufficient to reduce the support size. For (ii), C_4 contains all quadruplets, so that Theorem 1 requires an OA of strength 4.

While Theorem 1 identifies D-optimal designs with a reduced support size, any group with a support point also includes a second support point. The next theorem shows that this is not always necessary. Its proof can be found in the Appendix.

Theorem 2 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^1}$, that is, with γ_{12} as the only interaction effect. Let H be the collection of rows for an $OA(N, s_1 \cdots s_L 2^1, 2+)$ with the property that columns $(1, 2, j)$, $3 \leq j \leq L+1$, form an $OA(N, s_1 s_2 s_j, 3)$. Let H_1 and H_2 denote the $\frac{N}{2} \times L$ collections obtained by partitioning the rows in H into those with final entry equal to 1 and 2, respectively, and then deleting that final entry from each row. Define*

$$\begin{aligned}\xi_{1a} =&\{(c_{i_1\cdots i_L} = c^*, w_{i_1\cdots i_L} = \frac{1}{N}),\ \textit{for all } (i_1, \cdots, i_L) \in H_1,\ \textit{and}\\ &\{(c_{i_1\cdots i_L} = -c^*, w_{i_1\cdots i_L} = \frac{1}{N}),\ \textit{for all } (i_1, \cdots, i_L) \in H_2\} \quad .\end{aligned}$$

Then ξ_{1a} is D-optimal for $\boldsymbol{\theta_1^1}$.

Sometimes, it might be difficult to have an additional 2-level column, whereas an OA might exist with the extra column having an odd number of levels. If that happens, then an optimal design with less than $2N$ support point can be constructed as in the following result.

Theorem 3 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^1}$, that is, with γ_{12} as the only interaction effect. Let H be the collection of rows for an $OA(N, s_1 \cdots s_L (2u+1)^1, 2+)$ with the property that columns $(1, 2, j)$, $3 \leq j \leq L+1$, form an $OA(N, s_1 s_2 s_j, 3)$. Denoting the levels of the last column as $1, 2, \ldots, (2u+1)$, let H_1, H_2 and H_3 be the $\frac{uN}{2u+1} \times L$, $\frac{uN}{2u+1} \times L$ and $\frac{N}{2u+1} \times L$ collections obtained by partitioning the rows in H into those with final entry in the sets $\{1, 2, \ldots, u\}$, $\{u+1, u+2, \ldots, 2u\}$ and $\{2u+1\}$, respectively, and then deleting that final entry. Define*

$$\begin{aligned}\xi_{1b} = &\{(c_{i_1\cdots i_L} = c^*, w_{i_1\cdots i_L} = \frac{1}{N}), (i_1, \cdots, i_L) \in H_1\} \ \cup\\ &\{(c_{i_1\cdots i_L} = -c^*, w_{i_1\cdots i_L} = \frac{1}{N}), (i_1, \cdots, i_L) \in H_2\} \ \cup\\ &\{(c_{i_1\cdots i_L 1} = c^*, w_{i_1\cdots i_L 1} = \frac{1}{2N}), (c_{i_1\cdots i_L 2} = -c^*, w_{i_1\cdots i_L 2} = \frac{1}{2N}),\\ &(i_1, \cdots, i_L) \in H_3\} \quad .\end{aligned}$$

Then ξ_{1b} is D-optimal for $\boldsymbol{\theta_1^1}$.

Theorem 2 provides optimal designs with N support points for the special case of $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^1}$. Similarly, the following two theorems provide smaller optimal designs with N support points for the special cases of $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^2}$ and $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^3}$. Proofs of these are omitted since they follow along the same lines as the proof of Theorem 2.

Theorem 4 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^2}$. Let H be the collection of rows for an $OA(N, s_1 \cdots s_L 2^1, 2+)$ with the property that columns $(1, 2, j)$, $3 \leq j \leq L+1$, and $(1, 3, j)$, $4 \leq j \leq$*

$l+1$, form an $OA(N, s_1 s_2 s_j, 3)$ and an $OA(N, s_1 s_3 s_j, 3)$, respectively. Let H_1 and H_2 denote the $\frac{N}{2} \times L$ collections obtained by partitioning the rows in H into those with final entry equal to 1 and 2, respectively, and then deleting that final entry from each row. Define

$$\xi_{2a} = \{(c_{i_1 \cdots i_L} = c^*, w_{i_1 \cdots i_L} = \frac{1}{N}),\ \text{for all } (i_1, \cdots, i_L) \in H_1,\ \text{and}$$
$$(c_{i_1 \cdots i_L} = -c^*, w_{i_1 \cdots i_L} = \frac{1}{N}),\ \text{for all } (i_1, \cdots, i_L) \in H_2\} \qquad .$$

Then ξ_{2a} is D-optimal for $\boldsymbol{\theta_1^2}$.

Theorem 5 *Let c^* be as in Lemma 1 for model (7) with the logistic or probit link and with $\boldsymbol{\theta_1} = \boldsymbol{\theta_1^3}$. Let H be the collection of rows for an $OA(N, s_1 \cdots s_L 2^1, 2+)$ with the property that columns $(1, 2, j)$ and $(3, 4, j)$, $5 \le j \le L+1$, form an $OA(N, s_1 s_2 s_j, 3)$ and $OA(N, s_3 s_4 s_j, 3)$, respectively, and the columns $(1, 2, 3, 4)$ form an $OA(N, s_1 s_2 s_3 s_4, 4)$. Let H_1 and H_2 denote the $\frac{N}{2} \times L$ collections obtained by partitioning the rows in H into those with final entry equal to 1 and 2, respectively, and then deleting that final entry from each row. Define*

$$\xi_{3a} = \{(c_{i_1 \cdots i_L} = c^*, w_{i_1 \cdots i_L} = \frac{1}{N}),\ \text{for all } (i_1, \cdots, i_L) \in H_1,\ \text{and}$$
$$(c_{i_1 \cdots i_L} = -c^*, w_{i_1 \cdots i_L} = \frac{1}{N}),\ \text{for all } (i_1, \cdots, i_L) \in H_2\}.$$

Then ξ_{3a} is D-optimal for $\boldsymbol{\theta_1^3}$.

When an OA with the additional two-level factor as needed in Theorems 4 and 5 does not exist, one may consider whether adding a factor with an odd number of levels is possible. Results along the line of Theorem 3 can then be obtained as expansions of Theorems 4 and 5. Details are omitted here.

4 An Illustrative Example

For the example studied by Wang and Stufken [2020], using results obtained in Section 3, we now obtain optimal designs that are smaller than the ones in Wang and Stufken [2020]. The study assesses the effect of four factors and a covariate on the failure rate of semiconductors when exposed to electrostatic discharge (ESD). Since the response variable is binary, the study, originally reported in Whitman et al. [2006], uses a logistic model. The four two-level factors are:

(i) Lot A (Location 1 or Location 2),
(ii) Lot B (Location 1 or Location 2),
(iii) ESD Handling (No or Yes), and
(iv) Pulse Polarity (Negative or Positive).

The continuous covariate is the voltage used to test a wafer and the binary response variable is pass or fail for a wafer. Taking the factor levels as 1 and -1 and denoting them by x_1 through x_4 (in the order listed above), with p and x denoting the probability of a wafer passing the test and the voltage, respectively, the model that was used is

$$logit(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{34} x_3 x_4 + \beta_5 x. \qquad (9)$$

Thus, the linear predictor includes an interaction term, and $G_2 = \{(3,4)\}$. We take $\boldsymbol{\theta_1} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_{34}, \beta_5)$. While this model is a reparameterization of a model used for the theoretical results, D-optimality is invariant to this. Since G_2 contains only one interaction, Theorems 1 through 3 are applicable provided that the required orthogonal arrays exist.

The original experimenter used a full factorial design in the four factors, and ran each combination at 5 voltage levels: 25, 30, 35, 40 and 45 Volt. This resulted in $2^4 \times 5 = 80$ runs. No discussion about reasons for selecting the voltage range as $[25, 45]$ or for selecting 5 levels was provided. Therefore, following Wang and Stufken [2020], we continue to treat voltage as a continuous covariate that is not necessarily restricted to the range used in the experiment.

To find locally D-optimal designs, we use a guess for the parameter vector $\boldsymbol{\theta_1}$ given by $(-7.50, 1.50, -0.20, -0.15, 0.25, 0.40, 0.35)^T$ as in Lukemire et al. [2019] and Wang and Stufken [2020]. This is based on estimates from the original study. The D-optimal design reported in Wang and Stufken [2020] with 32 runs uses two support points in each group corresponding to a full factorial in four factors. For convenience, this design is given in Table 1.

Table 1 D-optimal design for the ESD experiment using a full factorial.

x_1	x_2	x_3	x_4	Volt1	Volt2	x_1	x_2	x_3	x_4	Volt1	Volt2
-1	-1	-1	-1	22.07	26.50	1	-1	-1	-1	13.50	17.93
-1	-1	-1	1	22.93	27.36	1	-1	-1	1	14.36	18.78
-1	-1	1	-1	25.22	29.64	1	-1	1	-1	16.64	21.07
-1	-1	1	1	21.50	25.93	1	-1	1	1	12.93	17.36
-1	1	-1	-1	23.22	27.64	1	1	-1	-1	14.64	19.07
-1	1	-1	1	24.07	28.50	1	1	-1	1	15.50	19.93
-1	1	1	-1	26.36	30.78	1	1	1	-1	17.79	22.21
-1	1	1	1	22.64	27.07	1	1	1	1	14.07	18.50

Corollary 1 gives a smaller D-optimal design that requires only 16 support points, whereas Theorem 2 gives a D-optimal design with only 8 support points. These two designs are given in the left and right panel of Table 2, respectively. For each design, all support points have the same weight. To produce an 80-run design as reported in the original study, we can simply use each of the support points multiple times. In terms of structure, these designs are much simpler. They also allow pure error estimates. But there is a trade-off because the smaller designs are not robust to model-misspecification.

Table 2 Smaller D-optimal designs for the ESD experiment using orthogonal arrays.

LotA	LotB	ESD	Pulse	Volt1	Volt2	LotA	LotB	ESD	Pulse	Volt
-1	-1	1	-1	25.22	29.64	-1	-1	1	-1	29.64
-1	-1	1	1	21.50	25.93	-1	-1	1	1	21.50
-1	1	-1	-1	23.22	27.64	-1	1	-1	-1	27.64
-1	1	-1	1	24.07	28.50	-1	1	-1	1	24.07
1	-1	-1	-1	13.50	17.93	1	-1	-1	-1	13.50
1	-1	-1	1	14.36	18.78	1	-1	-1	1	18.78
1	1	1	-1	17.79	22.21	1	1	1	-1	17.79
1	1	1	1	14.07	18.50	1	1	1	1	18.50

Using approximate designs again, the D-efficiency of the design used in the study, ξ_0, relative to any of the optimal design, such as ξ^* in Table 1, can be computed as

$$RE(\xi_0) = [\frac{det(I_{\xi_0})}{det(I_{\xi^*})}]^{1/p}, \tag{10}$$

where $p = 7$ is the number of parameters in the model. We find that $RE(\xi_0) = 24.22\%$, suggesting that an optimal design is more than four times as efficient as the design that was used.

5 Summary and Discussion

Even though GLMs with factorial effects have been widely used in many research areas, limited results exist on optimal designs. In this paper, we proposed theorems to obtain smaller D-optimal designs for models with the overall mean, all the main effects, a slope parameter, and some or all two-factor interactions using strength 2+ orthogonal arrays. The theoretical results provide great insight into the structure of families of D-optimal designs, irrespective of the number of factors and their levels. The usefulness of our theorems has been demonstrated through a real-life example. The results indicate that the designs based on the proposed theorems are not only more efficient than the original design but require far fewer support points (the three designs we obtained required 32, 16, and 8 support points, respectively, compared to 80 support points in the original design).

The results in this paper would require knowledge about the existence of OAs of strength 2+ and their construction. There are special cases where such results are readily available, such as when the required OAs are regular OAs that have a defining relation. For example, for the 2^{5-2} fraction $I = ABCD = ABE = CDE$ of strength 2, with respect to factors A, C and any third factor, the array has strength 3. This is precisely the type of condition that we need in some of the results in Section 3. If C_4 in Theorem 1 is not empty, then we need a requirement that certain two-factor interactions are not aliased with each other. This will be satisfied for clear two-factor interactions (cf. Chen and Hedayat [1998] and Wu et al. [2012]), but that is a stronger requirement than needed here.

Finally, while the theoretical results provide insight into the desired structure of D-optimal designs, when the values of c^* or $-c^*$ are not within the design region for groups where they are needed, it may be necessary to search for D-optimal designs by using an algorithm such as in Lukemire et al. [2019].

Acknowledgment

Research was partially supported through NSF grant DMS-1935729.

Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

H. Chen and A. Hedayat. 2^{n-m} designs with resolution iii or iv containing clear two-factor interactions. *Journal of Statistical Planning and Inference*, 75(1):147–158, 1998.

Y. He and B. Tang. Strong orthogonal arrays and associated latin hypercubes for computer experiments. *Biometrika*, 100(1):254–260, 2013.

Y. He, C.-S. Cheng, and B. Tang. Strong orthogonal arrays of strength two plus. *The Annals of Statistics*, 46(2):457–468, 2018.

A. Hedayat. Orthogonal arrays of strength $t+$ and their statistical applications. Technical report, University of Illinois, Chicago, 1989.

A. Hedayat. New properties of orthogonal arrays and their statistical applications. In S. Ghosh, editor, *Statistical Design and Analysis of Industrial Experiments*, chapter 13, pages 407–422. Dekker, New York, 1990.

A. Hedayat, N. Sloane, and J. Stufken. *Orthogonal arrays: theory and applications*. Springer Verlag, New York, 1999.

J. Lukemire, A. Mandal, and W. K. Wong. d-QPSO: A quantum-behaved particle swarm technique for finding D-optimal designs with discrete and continuous factors and a binary response. *Technometrics*, 61:77–87, 2019.

C. Shi and B. Tang. Design selection for strong orthogonal arrays. *Canadian Journal of Statistics*, 47(2):302–314, 2019.

J. Stufken and M. Yang. On locally optimal designs for generalized linear models with group effects. *Statistica Sinica*, 22:1765–1786, 2012.

X. Tan. Optimal designs for generalized linear models. *Unpublished Ph.D. dissertation, University of Georgia*, 2015.

Z. Wang and J. Stufken. Locally D-optimal designs for Binary Responses in the Presence of Factorial Effects. *Journal of Statistical Theory and Practice*, 14(2):1–15, 2020.

C. Whitman, T. M. Gilbert, A. M. Rahn, and J. A. Antonell. Determining factors affecting ESD failure voltage using DOE. *Microelectronics Reliability*, 46(8):1228–1237, 2006.

H. Wu, R. Mee, and B. Tang. Fractional factorial designs with admissible sets of clear two-factor interactions. *Technometrics*, 54(2):191–197, 2012.

Y. Zhou and B. Tang. Column-orthogonal strong orthogonal arrays of strength two plus and three minus. *Biometrika*, 106(4):997–1004, 2019.

Appendix

Proof of Theorem 1. From Equation (8), we have that

$$M_{\xi_g}(\boldsymbol{\theta_1^g}) = \frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}\boldsymbol{D^{i_1\cdots i_L j}}(\boldsymbol{D^{i_1\cdots i_L j}})^T,$$

where, for any $(i_1,\cdots,i_L)\in H$, the matrix $\boldsymbol{D^{i_1\cdots i_L j}}(\boldsymbol{D^{i_1\cdots i_L j}})^T$ is

$$\begin{pmatrix}
1 & (\boldsymbol{Z_1^{i_1}})^T & \cdots & (\boldsymbol{Z_L^{i_L}})^T & \cdots & (\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T & \cdots & c_{i_1\cdots i_L j} \\
 & \boldsymbol{Z_1^{i_1}}(\boldsymbol{Z_1^{i_1}})^T & \cdots & \boldsymbol{Z_1^{i_1}}(\boldsymbol{Z_L^{i_L}})^T & \cdots & \boldsymbol{Z_1^{i_1}}(\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T & \cdots & c_{i_1\cdots i_L j}\boldsymbol{Z_1^{i_1}} \\
 & & \ddots & \vdots & & \vdots & & \vdots \\
 & & & \boldsymbol{Z_L^{i_L}}(\boldsymbol{Z_L^{i_L}})^T & \cdots & \boldsymbol{Z_L^{i_L}}(\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T & \cdots & c_{i_1\cdots i_L j}\boldsymbol{Z_L^{i_L}} \\
 & & & & \ddots & \vdots & & \vdots \\
 & & & & & \boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}}(\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T & \cdots & c_{i_1\cdots i_L j}\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}} \\
 & & & & & & \ddots & \vdots \\
 & & & & & & & c_{i_1\cdots i_L j}^2
\end{pmatrix}$$

Since H has N rows, the top-left element of $M_{\xi_g}(\boldsymbol{\theta_1^g})$ is

$$\frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}1 = \Psi(c^*),$$

while the bottom-right element is

$$\frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}(c^*)^2 = (c^*)^2\Psi(c^*).$$

All other elements in the last column of $M_{\xi_g}(\boldsymbol{\theta_1^g})$ are 0 because each cell has c-values c^* and $-c^*$ with equal weights.

Other off-diagonal blocks in $\boldsymbol{D^{i_1\cdots i_L j}}(\boldsymbol{D^{i_1\cdots i_L j}})^T$ are of the form $(\boldsymbol{Z_l^{i_l}})^T$, $(\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T$, $\boldsymbol{Z_{l_1}^{i_{l_1}}}(\boldsymbol{Z_{l_2}^{i_{l_2}}})^T$ for $l_1 \neq l_2$, $\boldsymbol{Z_l^{i_l}}(\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T$, $(\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})(\boldsymbol{Z_{l_3 l_4}^{i_{l_3} i_{l_4}}})^T$ for $(l_1,l_2)\neq(l_3,l_4)$, or their transposes. Considering what this means for $M_{\xi_g}(\boldsymbol{\theta_1^g})$, first, from the definition of $\boldsymbol{Z_l^{i_l}}$ and since each level of factor l appears equally often, we see that

$$\frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}(\boldsymbol{Z_l^{i_l}})^T = \frac{1}{2N}\Psi(c^*)\frac{2N}{s_l}\sum_{i_l=1}^{s_l}(\boldsymbol{Z_l^{i_l}})^T = 0^T.$$

Now, corresponding to $(\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})^T$, since all level combinations for any two factors (l_1, l_2) appear equally often, we have

$$\begin{aligned}
&\frac{1}{2N}\Psi(c^*) \sum_{(i_1,\cdots,i_L)\in H} \sum_{j=1}^{2} (\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})^T \\
=&\frac{1}{2N}\Psi(c^*) \frac{2N}{s_{l_1} s_{l_2}} \sum_{i_{l_1}=1}^{s_{l_1}} \sum_{i_{l_2}=1}^{s_{l_2}} (\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}} \otimes \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})^T \\
=&\frac{1}{2N}\Psi(c^*) \frac{2N}{s_{l_1} s_{l_2}} \sum_{i_{l_1}=1}^{s_{l_1}} (\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}} \otimes \sum_{i_{l_2}=1}^{s_{l_2}} \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})^T = 0.
\end{aligned}$$

Next, using that the elements of H form an OA of strength 2 and that $l_1 \neq l_2$,

$$\begin{aligned}
&\frac{1}{2N}\Psi(c^*) \sum_{(i_1,\cdots,i_L)\in H} \sum_{j=1}^{2} \boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}} (\boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})^T \\
=&\frac{1}{2N}\Psi(c^*) \frac{2N}{s_{l_1} s_{l_2}} \left(\sum_{i_{l_1}=1}^{s_{l_1}} \boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}}\right) \left(\sum_{i_{l_2}=1}^{s_{l_2}} (\boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})^T\right) = 0.
\end{aligned}$$

Further, if $l = l_1$, again using that the elements of H form an OA of strength 2,

$$\begin{aligned}
&\frac{1}{2N}\Psi(c^*) \sum_{(i_1,\cdots,i_L)\in H} \sum_{j=1}^{2} \boldsymbol{Z}_{\boldsymbol{l}}^{\boldsymbol{i_l}} (\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})^T \\
=&\frac{1}{2N}\Psi(c^*) \frac{2N}{s_{l_1} s_{l_2}} \sum_{i_{l_1}=1}^{s_{l_1}} \sum_{i_{l_2}=1}^{s_{l_2}} \boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}} (\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}} \otimes \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})^T \\
=&\frac{1}{2N}\Psi(c^*) \frac{2N}{s_{l_1} s_{l_2}} \sum_{i_{l_1}=1}^{s_{l_1}} \boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}} (\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}} \otimes (\sum_{i_{l_2}=1}^{s_{l_2}} \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}}))^T = 0.
\end{aligned}$$

A similar argument applies for $l = l_2$. If $l \neq l_1$ and $l \neq l_2$, then

$$\begin{aligned}
&\frac{1}{2N}\Psi(c^*) \sum_{(i_1,\cdots,i_L)\in H} \sum_{j=1}^{2} \boldsymbol{Z}_{\boldsymbol{l}}^{\boldsymbol{i_l}} (\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})^T \\
=&\frac{1}{2N}\Psi(c^*) \frac{2N}{s_{l_1} s_{l_2} s_l} \sum_{i_{l_1}=1}^{s_{l_1}} \sum_{i_{l_2}=1}^{s_{l_2}} \left(\sum_{i_l=1}^{s_l} \boldsymbol{Z}_{\boldsymbol{l}}^{\boldsymbol{i_l}}\right) (\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})^T = 0,
\end{aligned}$$

where we have used that $(l, l_1, l_2) \in C_3$ and H is an OA of strength 3 for such a set of 3 columns.

Finally, for the off-diagonal blocks corresponding to $(\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})(\boldsymbol{Z}_{\boldsymbol{l_3 l_4}}^{\boldsymbol{i_{l_3} i_{l_4}}})^T$ for $(l_1, l_2) \neq (l_3, l_4)$, we could have two situations: the two interactions have one factor in common, say $l_1 = l_3$ (and $l_2 \neq l_4$), or they represent four different factors. Looking at the first case, we have

$$\begin{aligned}
&\frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}(\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})(\boldsymbol{Z}_{\boldsymbol{l_3 l_4}}^{\boldsymbol{i_{l_3} i_{l_4}}})^T \\
=&\frac{1}{2N}\Psi(c^*)\frac{2N}{s_{l_1}s_{l_2}s_{l_4}}\sum_{i_{l_1}=1}^{s_{l_1}}\sum_{i_{l_2}=1}^{s_{l_2}}\sum_{i_{l_4}=1}^{s_{l_4}}(\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}}\otimes \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})(\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}}\otimes \boldsymbol{Z}_{\boldsymbol{l_4}}^{\boldsymbol{i_{l_4}}})^T \\
=&\frac{1}{2N}\Psi(c^*)\frac{2N}{s_{l_1}s_{l_2}s_{l_4}}\sum_{i_{l_1}=1}^{s_{l_1}}\sum_{i_{l_2}=1}^{s_{l_2}}(\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}}\otimes \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})(\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}}\otimes \sum_{i_{l_4}=1}^{s_{l_4}}\boldsymbol{Z}_{\boldsymbol{l_4}}^{\boldsymbol{i_{l_4}}})^T = 0,
\end{aligned}$$

where we have used that $(l_1, l_2, l_4) \in C_3$ and H is an OA of strength 3 for such a set of 3 columns.

For the second case, with no common factors, $(l_1, l_2, l_3, l_4) \in C_4$ and H is an OA of strength 4 for such a set of factors. Hence,

$$\begin{aligned}
&\frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}(\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})(\boldsymbol{Z}_{\boldsymbol{l_3 l_4}}^{\boldsymbol{i_{l_3} i_{l_4}}})^T \\
=&\frac{1}{2N}\Psi(c^*)\frac{2N}{s_{l_1}s_{l_2}s_{l_3}s_{l_4}}\sum_{i_{l_1}=1}^{s_{l_1}}\sum_{i_{l_2}=1}^{s_{l_2}}\sum_{i_{l_3}=1}^{s_{l_3}}\sum_{i_{l_4}=1}^{s_{l_4}}(\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}}\otimes \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})(\boldsymbol{Z}_{\boldsymbol{l_3}}^{\boldsymbol{i_{l_3}}}\otimes \boldsymbol{Z}_{\boldsymbol{l_4}}^{\boldsymbol{i_{l_4}}})^T \\
=&\frac{1}{2N}\Psi(c^*)\frac{2N}{s_{l_1}s_{l_2}s_{l_3}s_{l_4}}\sum_{i_{l_1}=1}^{s_{l_1}}\sum_{i_{l_2}=1}^{s_{l_2}}\sum_{i_{l_3}=1}^{s_{l_3}}(\boldsymbol{Z}_{\boldsymbol{l_1}}^{\boldsymbol{i_{l_1}}}\otimes \boldsymbol{Z}_{\boldsymbol{l_2}}^{\boldsymbol{i_{l_2}}})(\boldsymbol{Z}_{\boldsymbol{l_3}}^{\boldsymbol{i_{l_3}}}\otimes \sum_{i_{l_4}=1}^{s_{l_4}}\boldsymbol{Z}_{\boldsymbol{l_4}}^{\boldsymbol{i_{l_4}}})^T = 0.
\end{aligned}$$

For a diagonal block of $M_{\xi_g}(\boldsymbol{\theta_1^g})$ that corresponds to a main effect, say for factor l, we obtain

$$\frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}\boldsymbol{Z}_{\boldsymbol{l}}^{\boldsymbol{i_l}}(\boldsymbol{Z}_{\boldsymbol{l}}^{\boldsymbol{i_l}})^T = \frac{\Psi(c^*)}{s_l}\sum_{i_l=1}^{s_l}\boldsymbol{Z}_{\boldsymbol{l}}^{\boldsymbol{i_l}}(\boldsymbol{Z}_{\boldsymbol{l}}^{\boldsymbol{i_l}})^T = \Psi(c^*)B_l,$$

where $B_l = \frac{1}{(s_l-1)^2}(s_l I - J)$. The first equality follows since every level of factor l comes N/s_l times. The second equality follows as in the proof of Lemma 1 in Wang and Stufken [2020].

For a diagonal block of $M_{\xi_g}(\boldsymbol{\theta_1^g})$ that corresponds to a two-factor interaction, say for factors (l_1, l_2), we get

$$\begin{aligned}
&= \frac{1}{2N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H}\sum_{j=1}^{2}\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}}(\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})^T \\
&= \frac{1}{2N}\Psi(c^*)\frac{2N}{s_{l_1}s_{l_2}}\sum_{i_{l_1}=1}^{s_{l_1}}\sum_{i_{l_2}=1}^{s_{l_2}}\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}}(\boldsymbol{Z}_{\boldsymbol{l_1 l_2}}^{\boldsymbol{i_{l_1} i_{l_2}}})^T \\
&= \frac{1}{2N}\Psi(c^*)\frac{2N}{s_{l_1}s_{l_2}}(s_{l_1}\cdot B_{l_1})\otimes(s_{l_2}\cdot B_{l_2}) = \Psi(c^*)(B_{l_1}\otimes B_{l_2}),
\end{aligned}$$

where the penultimate equality follows as in Wang and Stufken [2020].

Combined, the previous steps show that $M_{\xi_g}(\boldsymbol{\theta_1^g})$ is identical to $M_{\xi^*}(\boldsymbol{\theta_1^g})$ in Lemma 2, so that ξ_g is also D-optimal for $\boldsymbol{\theta_1^g}$. □

Proof of Theorem 2. As in the proof of Theorem 1, we now need to show that $M_{\xi_{1a}}(\boldsymbol{\theta_1^1})$ is the same as $M_{\xi^*}(\boldsymbol{\theta_1^1})$ in Lemma 2. Since H is an OA as in Corollary 1, our work is simplified. Only the entries in $M_{\xi_{1a}}(\boldsymbol{\theta_1^1})$ that depend on the c-values need verification. In other words, we only need to consider the last column in $M_{\xi_{1a}}(\boldsymbol{\theta_1^1})$.

First, the bottom-right diagonal element in $M_{\xi_{1a}}(\boldsymbol{\theta_1^1})$ is

$$\frac{1}{N}\Psi(c^*)\sum_{(i_1,\cdots,i_L)\in H_1\cup H_2}(c^*)^2=(c^*)^2\Psi(c^*).$$

Further, for the entry in the final column of $M_{\xi_{1a}}(\boldsymbol{\theta_1^1})$ that corresponds to the main effect of factor l, using Equation (8), we get

$$\begin{aligned}&\sum_{(i_1,\cdots,i_L)\in H_1} w_{i_1\cdots i_L}\Psi(c_{i_1\cdots i_L})c_{i_1\cdots i_L}\boldsymbol{Z_l^{i_l}}+\sum_{(i_1,\cdots,i_L)\in H_2} w_{i_1\cdots i_L}\Psi(c_{i_1\cdots i_L})c_{i_1\cdots i_L}\boldsymbol{Z_l^{i_l}}\\&=\frac{1}{N}\Psi(c^*)\frac{N}{2s_l}\left(\sum_{i_l=1}^{s_l}c^*\boldsymbol{Z_l^{i_l}}-\sum_{i_l=1}^{s_l}c^*\boldsymbol{Z_l^{i_l}}\right)=0.\end{aligned}$$

The first equality follows because every level of factor l appears equally often in both H_1 and H_2.

The only other entry in the final column of $M_{\xi_{1a}}(\boldsymbol{\theta_1^1})$ corresponds to the interaction of factors 1 and 2. Because H is an orthogonal array of strength 3 for factors 1, 2 and the additional 2-level column (the $(L+1)$st column), it follows that every level combination (i_1,i_2) for the first two factors appears equally often in H_1 and H_2. Hence,

$$\begin{aligned}&\sum_{(i_1,\cdots,i_L)\in H_1} w_{i_1\cdots i_L}\Psi(c_{i_1\cdots i_L})c_{i_1\cdots i_L}\boldsymbol{Z_{12}^{i_1i_2}}+\sum_{(i_1,\cdots,i_L)\in H_2} w_{i_1\cdots i_L}\Psi(c_{i_1\cdots i_L})c_{i_1\cdots i_L}\boldsymbol{Z_{12}^{i_1i_2}}\\&=\frac{1}{N}\Psi(c^*)\left(\frac{N}{2s_1s_2}\right)\sum_{i_1=1}^{s_1}\sum_{i_2=1}^{s_2}(c^*\boldsymbol{Z_{12}^{i_1i_2}}-c^*\boldsymbol{Z_{12}^{i_1i_2}})=0.\end{aligned}$$

Therefore, all entries in the last column of $M_{\xi_{1a}}(\boldsymbol{\theta_1^1})$ are also equal to those of $M_{\xi^*}(\boldsymbol{\theta_1^1})$. This concludes the proof. □

Proof of Theorem 3. As in the proof of Theorem 2, it suffices to verify that entries in the final column of $M_{\xi_{1b}}(\boldsymbol{\theta_1^1}))$ are equal to those in $M_{\xi^*}(\boldsymbol{\theta_1^1})$ in Lemma 2. For the bottom-right diagonal element in $M_{\xi_{1b}}$, using Equation (8), we get

$$\begin{aligned}&\Psi(c^*)\left(\frac{1}{N}\sum_{(i_1,\cdots,i_L)\in H_1}(c^*)^2+\frac{1}{N}\sum_{(i_1,\cdots,i_L)\in H_2}(c^*)^2+\frac{1}{2N}\sum_{(i_1,\cdots,i_L)\in H_3}\sum_{j=1}^{2}(c^*)^2\right)\\&=(c^*)^2\Psi(c^*),\end{aligned}$$

by counting the sizes of the H_i's.

Further, for the entry in the final column of $M_{\xi_{1b}}(\boldsymbol{\theta_1^1})$ that corresponds to the main effect of factor l, we get

$$\Psi(c^*)\frac{1}{N}\left(\frac{uN}{(2u+1)s_l}\right)\sum_{i_l=1}^{s_l}(c^*\boldsymbol{Z_l^{i_l}} - c^*\boldsymbol{Z_l^{i_l}})$$
$$+\Psi(c^*)\frac{1}{2N}\left(\frac{N}{(2u+1)s_l}\right)\sum_{i_l=1}^{s_l}(c^*\boldsymbol{Z_l^{i_l}} - c^*\boldsymbol{Z_l^{i_l}}) = 0,$$

where we have used that each level of factor l appears $\frac{uN}{(2u+1)s_l}$ times in each of H_1 and H_2 and $\frac{N}{(2u+1)s_l}$ times in H_3.

The only other entry in the final column of $M_{\xi_{1b}}(\boldsymbol{\theta_1^1})$ corresponds to the interaction of factors 1 and 2. Because H is an orthogonal array of strength 3 for factors 1, 2 and the additional $(2u+1)$-level column (the $(L+1)$st column), it follows that every level combination (i_1, i_2) for the first two factors appears $\frac{Nu}{(2u+1)s_1s_2}$ times in each of H_1 and H_2 and $\frac{N}{(2u+1)s_1s_2}$ times in H_3. Using Equation 8, we obtain that the final entry in the final column is equal to

$$\Psi(c^*)\frac{1}{N}\left(\frac{uN}{(2u+1)s_1s_2}\right)\sum_{i_1=1}^{s_1}\sum_{i_2=1}^{s_2}\left(c^*\boldsymbol{Z_{12}^{i_1i_2}} - c^*\boldsymbol{Z_{12}^{i_1i_2}}\right)$$
$$+\Psi(c^*)\frac{1}{2N}\left(\frac{N}{(2u+1)s_1s_2}\right)\sum_{i_1=1}^{s_1}\sum_{i_2=1}^{s_2}\left(c^*\boldsymbol{Z_{12}^{i_1i_2}} - c^*\boldsymbol{Z_{12}^{i_1i_2}}\right) = 0.$$

Therefore, all entries in the last column of $M_{\xi_{1b}}(\boldsymbol{\theta_1^1})$ are also equal to those of $M_{\xi^*}(\boldsymbol{\theta_1^1})$. This concludes the proof. □