Journal of Statistical Theory and Practice (2020) 14:19
https://doi.org/10.1007/s42519-020-0085-0

ORIGINAL ARTICLE

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

Zhongshen Wang[1] · **John Stufken**[2]

Published online: 18 February 2020

Abstract

We consider the problem of finding D-optimal designs for certain generalized linear models (GLMs). In particular, we study GLMs with factorial effects and one continuous covariate. The factorial effects include both main effects and interactions, and the design region for the continuous covariate can be unrestricted. The local D-optimality of the proposed design is established through the equivalence theorem, and the results are illustrated with an electrostatic discharge (ESD) experiment.

Keywords Locally optimal design · D-optimality · Factorial experiment · Equivalence theorem

1 Introduction

Factorial experiments are commonly used to study the effects of multiple factors on a response variable of interest. When the response variable is categorical, generalized linear models (GLMs) form a very useful class of models. In this paper, we focus on a binary response variable. In that case, two frequently used link functions for GLMs are the logistic and probit links.

Research partially supported by NSF grants DMS-1506125 and DMS-1935729.

Part of special issue guest edited by Pritam Ranjan and Min Yang—Algorithms, Analysis and Advanced Methodologies in the Design of Experiments.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s42519-020-0085-0) contains supplementary material, which is available to authorized users.

✉ John Stufken
jstufken@uncg.edu

Zhongshen Wang
zwang@incyte.com

[1] Incyte Corporation, Wilmington, DE 19803, USA

[2] University of North Carolina at Greensboro, Greensboro, NC 27412, USA

When a controllable continuous covariate is included in addition to the factorial effects, optimal design questions for these models center on the combinations of values and levels for the covariate and factors, respectively, that should be selected in the experiment. These choices are extremely important because good designs can improve parameter estimation for a fixed number of runs or reduce costs to achieve a desired level of precision. Previous studies for this problem focus primarily on computational approaches, with rather few theoretical results. For models with main effects only, [4] provided an explicit expression for optimal designs under D-optimality. [5] extended their results and obtained smaller optimal designs using orthogonal arrays. He also explored interaction models where all interactions up to a specified order are included in the model. The results of our paper are more general. For example, with 3 factors, for a second-order model, [5] required all three of the 2-factor interactions to be included. We extend this to the case that only one or two of these 2-factor interactions are in the model. Similar extensions are presented for higher-order models.

Studies with binary responses are extremely common. A drug may or may not relieve a symptom, a rat may or may not develop a tumor, a stimulant may or may not lead to increased performance and so on. When possible, in order to establish causal relationships, such studies are conducted in the form of experiments. Many factors are often studied, such as gender, age, body weight, and so on. Continuous covariates, for example measurements at the beginning of a study, such as body mass index, blood pressure, concentration of a certain chemical in the blood, are also often of interest to account for natural differences between subjects. Whether all of the factors and covariates can be controlled for the experiment, or whether some can only be observed, depends on the specific experiment. As a result, many different situations, objectives and models can occur. The goal of this paper is to develop optimal design theory for one of such situations, namely the situation with only one continuous covariate, where all factors and the covariate can be controlled during the experiment, and where the objective is to study the effect that the factors have on the binary response.

For such situations, we propose designs that are then shown to be locally D-optimal by using the equivalence theorem (see [1, 2]). The main theorem is formulated in Sect. 2. In Sect. 3 we provide an illustrative example about determining factors affecting ESD failure voltage ([7]). An optimal design is obtained using our proposed theorem and is compared in terms of D-efficiency to the design used in the original study. A summary and discussion are presented in Sect. 4, followed by an "Appendix" with several of the proofs.

2 Theory

Referring to experimental units as subjects, for L factors with $s_1, s_2, \dots, s_L$ levels, respectively, each subject can be assigned to one of the $s = s_1 \cdot s_2 \cdots s_L$ groups. We assume that the linear predictor has a common slope for the continuous covariate across all of the groups. Then the model that we will consider can be written as

$$Prob(Y_{i_1 i_2 \ldots i_L j} = 1) = P\left(\alpha_0 + \alpha_1^{i_1} + \cdots + \alpha_L^{i_L} + \sum_{t=2}^{L} \sum_{(l_1, l_2, \ldots, l_t) \in G_t} \alpha_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}} + \beta x_{i_1 i_2 \cdots i_L j}\right), \tag{1}$$

where $Y_{i_1 i_2 \ldots i_L j}$ is the response from the jth subject in group $(i_1, i_2, \ldots, i_L)$, $i_l = 1, \ldots, s_l$; $j = 1, \ldots, m_{i_1 i_2 \cdots i_L}$, and $m_{i_1 i_2 \cdots i_L}$ is the number of subjects in group $(i_1, i_2, \ldots, i_L)$. Further, $P(\cdot)$ is a cumulative distribution function; α_0 is the grand 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 $(i_{l_1}, i_{l_2}, \ldots, i_{l_t})$ for the tth-order effect for the factors $(l_1, l_2, \ldots, l_t)$, $t = 2, \ldots, L$; and G_t is a set of t-tuples representing the tth-order effects included in the model. For simplicity, we also denote $G_1 = \{1, 2, \ldots, L\}$. Moreover, β is the common slope parameter, and $x_{i_1 i_2 \cdots i_L j}$ is the covariate value for the jth subject in group $(i_1, i_2, \ldots, i_L)$, 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. The results in this paper apply only to models as in (1).

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

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

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

While the notation in the previous paragraphs is convenient to present the model, we will now immediately change it to facilitate our discussion about designs. Some subjects in the same group could be assigned to the same covariate value, others to a second or even third value used for that group. To that end, we will now use $m_{i_1 i_2 \cdots i_L}$ to denote the distinct number of covariate values used in group $(i_1, i_2, \ldots, i_L)$ and introduce $n_{i_1 i_2 \cdots i_L j}$ to denote the number of subjects in group $(i_1, i_2, \ldots, i_L)$ who are assigned to the jth covariate value in that group, $j = 1, \ldots, m_{i_1 i_2 \cdots i_L}$.

With that notation, a design can be presented as

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

where the design points $\boldsymbol{X}^{i_1 \cdots i_L j}$ are distinct vectors (they either represent different groups or correspond to different covariate values in the same group), while $\sum_{i_1} \cdots \sum_{i_L} \sum_j n_{i_1 i_2 \cdots i_L j} = n$, the design size. Such a design is called an exact design. Therefore, the optimal exact design problem for a design of size n is to select both $\boldsymbol{X}^{i_1 \cdots i_L j}$'s and $n_{i_1 i_2 \cdots i_L j}$'s, the latter summing to n, such that the resulting design ξ is the best in terms of a certain optimality criterion. However, due to the discreteness of the $n_{i_1 i_2 \cdots i_L j}$'s, such optimization problems are difficult to solve. Instead, we work with the corresponding approximate designs, where $n_{i_1 i_2 \cdots i_L j}/n$ is replaced by $w_{i_1 i_2 \cdots i_L j}$. The $w_{i_1 i_2 \cdots i_L j}$'s are called design weights and $\sum_{i_1} \cdots \sum_{i_L} \sum_j w_{i_1 i_2 \cdots i_L j} = 1$. So an approximate design becomes

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

and when searching for an optimal design the $w_{i_1 i_2 \cdots i_L j}$'s can take any non-negative values that sum to 1, so that the optimal design problem no longer depends on n. The disadvantage of this approach is that it may be difficult to convert an optimal approximate design to an optimal exact design depending on the value of n.

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

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

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 $\boldsymbol{X}^{i_1 \cdots i_L j}$.

Inference could focus on a vector of $\boldsymbol{\theta}$, say $g(\boldsymbol{\theta})$. Assuming g to be differentiable, from the Delta method we obtain the asymptotic covariance matrix of $g(\hat{\boldsymbol{\theta}})$ as

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

where $I_\xi(\boldsymbol{\theta})^-$ is a generalized inverse of $I_\xi(\boldsymbol{\theta})$.

It is possible that for some function $g(\cdot)$, $\Sigma_\xi(\boldsymbol{\theta})$ is singular, but we will restrict our attention to situations where $\Sigma_\xi(\boldsymbol{\theta})$ is non-singular. The information matrix for $g(\boldsymbol{\theta})$ is therefore the inverse of $\Sigma_\xi(\boldsymbol{\theta})$,

$$I_\xi(g(\boldsymbol{\theta})) = \Sigma_\xi(\boldsymbol{\theta})^{-1} = \left(\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right) I_\xi(\boldsymbol{\theta})^- \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right)^T\right)^{-1}. \tag{5}$$

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. One way to overcome this is to replace $\boldsymbol{\theta}$ in $I_\xi(g(\boldsymbol{\theta}))$ by a "best guess," which could be based on prior experiments. The resulting D-optimal designs are known as locally D-optimal designs (for that guessed value of $\boldsymbol{\theta}$).

The model in (2) is overparameterized, and $\boldsymbol{\theta}$ is not estimable. Instead, we consider a maximal set of linearly independent estimable functions of $\boldsymbol{\theta}$. Since D-optimality is invariant under reparameterization, the optimal design result is invariant to the choice of this maximal set. Let $g(\boldsymbol{\theta}) = B\boldsymbol{\theta} = \boldsymbol{\eta}$ denote one particular maximal set. Note that

$$rank(B) = 1 + \sum_{t=1}^{L} \sum_{(l_1,\ldots,l_t)\in G_t} \left[\prod_{i=1}^{t}(s_{l_i} - 1)\right] + 1. \tag{6}$$

We will denote the rank in (6) by r.

We define $c_{i_1\cdots i_L j} = (\boldsymbol{X}^{i_1\cdots i_L j})^T\boldsymbol{\theta}$, 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 $x_{i_1 i_2\cdots i_L j}$. In this way, rather than specifying covariate values for each group, a design can now also be presented by specifying values for $c_{i_1\cdots i_L j}$ for each group. By doing so, Theorem 1 provides a locally D-optimal design for $\boldsymbol{\eta}$ under models as in (2).

Theorem 1 *For a model of 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* $(i_l, \ldots, i_L)$, *where* $c^* > 0$ *maximizes* $f(c) = c^2(\Psi(c))^r$ *on* $(-\infty, \infty)$, *the design* $\xi^* = \{(c_{i_1\cdots i_L 1} = c^*, w_{i_1\cdots i_L 1} = \frac{1}{2s}), (c_{i_1\cdots i_L 2} = -c^*, w_{i_1\cdots i_L 2} = \frac{1}{2s}), i_l = 1, \ldots, s_l, l = 1, \ldots, L\}$ *is a locally D-optimal design for* $\boldsymbol{\eta}$. *Here* $s = s_1 \times \cdots \times s_L$ *and* $\Psi(x)$ *is given by*

$$\Psi(x) = \begin{cases} \frac{e^x}{(1+e^x)^2}, & \text{for the logistic link} \\ \frac{[\Phi'(x)]^2}{\Phi(x)(1-\Phi(x))}, & \text{for the probit link} \end{cases}. \tag{7}$$

Proof We start with a reparameterization of model (1), the proof of which is relegated to "Appendix." To formulate this reparameterization, define the sets $H_1, \ldots, H_L$ from $G_1, \ldots, G_L$ as follows: For $1 \le t \le L$,

$$\begin{aligned} H_t =&\{(l_1, \ldots, l_t) : \text{ there is an index } t' \ge t \text{ and a } t'\text{-tuple} \\ &(j_1, \ldots, j_{t'}) \in G_{t'} \text{ so that } \{l_1, \ldots, l_t\} \text{ is a subset of } \{j_1, \ldots, j_{t'}\}\}. \end{aligned} \tag{8}$$

Then the reparameterization of model (1) is given by

$$Prob(Y_{i_1 i_2 \ldots i_L j} = 1) = P\left(\gamma_0 + \sum_{t=1}^{L} \sum_{(l_1,\ldots,l_t)\in H_t} \left[\sum_{i_{l_1}=1}^{s_{l_1}-1} \cdots \sum_{i_{l_t}=1}^{s_{l_t}-1} \gamma_{l_1\cdots l_t}^{i_{l_1}\cdots i_{l_t}} z_{l_1}^{i_{l_1}} \cdots z_{l_t}^{i_{l_t}}\right] + \beta x_{i_1\cdots i_L j}\right), \tag{9}$$

where, for each factor l, we define

$$z_l^1 = \begin{cases} 1, & \text{if factor } l \text{ is at level } 1 \\ -\frac{1}{s_l-1}, & \text{otherwise} \end{cases}$$

$$z_l^2 = \begin{cases} 1, & \text{if factor } l \text{ is at level } 2 \\ -\frac{1}{s_l-1}, & \text{otherwise} \end{cases}$$

$$\vdots$$

$$z_l^{s_l-1} = \begin{cases} 1, & \text{if factor } l \text{ is at level } s_l - 1 \\ -\frac{1}{s_l-1}, & \text{otherwise.} \end{cases}$$

Only including terms that appear in the model, the parameter vector $\boldsymbol{\theta_1}$ for model (9) can be written as $\boldsymbol{\theta_1} = (\gamma_0, \boldsymbol{\gamma_1}^T, \dots, \boldsymbol{\gamma_L}^T, \cdots, \boldsymbol{\gamma_{l_1 l_2}}^T, \dots, \boldsymbol{\gamma_{l_1 l_2 \cdots l_L}}^T, \beta)^T$, where for any $t = 1, \dots, L$ and $1 \le l_1 < \cdots < l_t \le L$, we use the notation $\boldsymbol{\gamma_{l_1 \cdots l_t}} = (\gamma_{l_1 \cdots l_t}^{1 \cdots 1}, \dots, \gamma_{l_1 \cdots l_t}^{1 \cdots (s_{l_t}-1)}, \dots, \gamma_{l_1 \cdots l_t}^{(s_{l_1}-1) \cdots (s_{l_t}-1)})^T$. Note that the length of $\boldsymbol{\theta_1}$ is equal to r. Again only focusing on t-tuples that appear in model (9), we define $\boldsymbol{Z}^{i_1 \cdots i_L j} = (1, (\boldsymbol{Z_1^{i_1}})^T, \dots, (\boldsymbol{Z_L^{i_L}})^T, \dots, (\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T, \dots, (\boldsymbol{Z_{l_1 l_2 \cdots l_L}^{i_{l_1} i_{l_2} \cdots i_{l_L}}})^T, x_{i_1 i_2 \cdots i_L j})^T$, where for each factor l,

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

and $\boldsymbol{Z_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}}} = \boldsymbol{Z_{l_1}^{i_{l_1}}} \otimes \cdots \otimes \boldsymbol{Z_{l_t}^{i_{l_t}}}$.

Since model (9) is a reparameterization of the original model, due to the invariance of D-optimality under reparameterizations, we can focus on D-optimality for $\boldsymbol{\theta_1}$ under model (9). Let $\boldsymbol{D}^{i_1 \cdots i_L j} = (1, (\boldsymbol{Z_1^{i_1}})^T, \dots, (\boldsymbol{Z_L^{i_L}})^T, \dots, (\boldsymbol{Z_{l_1 l_2}^{i_{l_1} i_{l_2}}})^T, \dots, (\boldsymbol{Z_{l_1 l_2 \cdots l_L}^{i_{l_1} i_{l_2} \cdots i_{l_L}}})^T, c_{i_1 i_2 \cdots i_L j})^T$, then $\boldsymbol{Z}^{i_1 \cdots i_L j} = A(\boldsymbol{\theta_1}) \boldsymbol{D}^{i_1 \cdots i_L j}$, where

$$A(\boldsymbol{\theta_1}) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \\ -\gamma_0/\beta & A_{(1)}(\boldsymbol{\theta_1}) & \cdots & A_{(L)}(\boldsymbol{\theta_1}) & 1/\beta \end{pmatrix}$$

and $A_{(t)}(\boldsymbol{\theta_1}) = (-(\gamma_{l_1 \cdots l_t})^T/\beta, \dots, -(\gamma_{l_1' \cdots l_t'})^T/\beta)$ represents the coefficients of all the t-th order effects included in the model.

Then, for the design ξ^* in Theorem 1, the information matrix for $\boldsymbol{\theta_1}$ becomes

$$\begin{aligned} I_{\xi^*}(\boldsymbol{\theta_1}) &= A(\boldsymbol{\theta_1}) \left[\sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \right. \\ &\quad \left. \sum_{j=1}^{m_{i_1 \cdots i_L}} w_{i_1 \cdots i_L j} \Psi(c_{i_1 \cdots i_L j}) \boldsymbol{D}^{i_1 \cdots i_L j} (\boldsymbol{D}^{i_1 \cdots i_L j})^T \right] A^T(\boldsymbol{\theta_1}) \\ &= A(\boldsymbol{\theta_1}) \left[\frac{1}{2s} \Psi(c^*) \sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \sum_{j=1}^{2} \boldsymbol{D}^{i_1 \cdots i_L j} (\boldsymbol{D}^{i_1 \cdots i_L j})^T \right] A^T(\boldsymbol{\theta_1}) \\ &= A(\boldsymbol{\theta_1}) M_{\xi^*}(\boldsymbol{\theta_1}) A^T(\boldsymbol{\theta_1}), \end{aligned} \tag{10}$$

where the last expression defines $M_{\xi^*}(\boldsymbol{\theta_1})$.

To prove that ξ^* is D-optimal for $\boldsymbol{\theta_1}$, we use the equivalence theorem in [2]. After simple calculations, ξ^* is a locally D-optimal design for $\boldsymbol{\theta_1}$ if

$$\Psi(c)(\boldsymbol{D}^{i_1\cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta_1})\boldsymbol{D}^{i_1\cdots i_L} \leq r, \tag{11}$$

with equality at the support points of ξ^*, for all design points $(i_1, \dots, i_L, c)$. Here $\boldsymbol{D}^{i_1\cdots i_L}$ is obtained from $\boldsymbol{D}^{i_1\cdots i_L j}$ by writing c instead of $c_{i_1\cdots i_L j}$.

We first use a lemma to state that $M_{\xi^*}(\boldsymbol{\theta_1})$ is a block-diagonal matrix.

Lemma 1 *$M_{\xi^*}(\boldsymbol{\theta_1})$ 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 $\boldsymbol{\gamma}_{l_1 l_2 \dots l_t}$ equal to*

$$B_{l_1 l_2 \cdots l_k} = B_{l_1} \otimes B_{l_2} \otimes \cdots \otimes B_{l_t}.$$

Proof See "Appendix." □

From Lemma 1, $M_{\xi^*}^{-1}(\boldsymbol{\theta_1})$ is equal to $\frac{1}{\Psi(c^*)}$ times a block-diagonal matrix with (1) the top-left element equal to 1; (2) the bottom-right element equal to $\frac{1}{(c^*)^2}$; (3) the block corresponding to $\boldsymbol{\gamma_l}^T$ equal to $B_l^{-1} = \frac{(s_l-1)^2}{s_l}(I + J)$; and (4) the block corresponding to $\boldsymbol{\gamma}_{l_1 l_2 \dots l_t}$ equal to $B_{l_1 l_2 \cdots l_t}^{-1} = B_{l_1}^{-1} \otimes \cdots \otimes B_{l_t}^{-1}$.

Before evaluating the expression on the left-hand side of Eq. (11), we need one more lemma.

$$(\boldsymbol{Z}_l^{i_l})^T \boldsymbol{Z}_l^{i_l} + (\boldsymbol{Z}_l^{i_l})^T J \boldsymbol{Z}_l^{i_l} = \frac{s_l}{s_l - 1}.$$

Lemma 2

Proof See "Appendix." □

From the block-diagonal form of $M_{\xi^*}^{-1}(\boldsymbol{\theta_1})$, it is immediately clear that $(\boldsymbol{D}^{i_1\cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta_1})\boldsymbol{D}^{i_1\cdots i_L}$ in Eq. (11) is a relatively simple sum of various terms. First, corresponding to γ_0, we have

$$1 \cdot \frac{1}{\Psi(c^*)} \cdot 1 = \frac{1}{\Psi(c^*)}.$$

Next, corresponding to β, we have

$$c \cdot \frac{1}{(c^*)^2 \Psi(c^*)} \cdot c = \frac{1}{\Psi(c^*)} \cdot \frac{c^2}{(c^*)^2}.$$

Further, corresponding to $\boldsymbol{\gamma_l}^T$ we obtain

$$
\begin{aligned}
(\boldsymbol{Z}_l^{i_l})^T \cdot \frac{1}{\Psi(c^*)} B_l^{-1} \cdot \boldsymbol{Z}_l^{i_l} &= \frac{1}{\Psi(c^*)} \frac{(s_l-1)^2}{s_l} (\boldsymbol{Z}_l^{i_l})^T (I+J) \boldsymbol{Z}_l^{i_l} \\
&= \frac{1}{\Psi(c^*)} \frac{(s_l-1)^2}{s_l} \left[(\boldsymbol{Z}_l^{i_l})^T \boldsymbol{Z}_l^{i_l} + (\boldsymbol{Z}_l^{i_l})^T J \boldsymbol{Z}_l^{i_l} \right] \\
&= \frac{1}{\Psi(c^*)} \frac{(s_l-1)^2}{s_l} \frac{s_l}{s_l-1} \qquad \text{(from Lemma 2)} \\
&= \frac{1}{\Psi(c^*)} \cdot (s_l-1).
\end{aligned}
$$

And finally, for ${\boldsymbol{\gamma}_{l_1 \cdots l_t}}^T$ the contributing term is given by

$$
\begin{aligned}
&(\boldsymbol{Z}_{l_1 \cdots l_t}^{i_{l_1} \cdots i_{l_t}})^T \frac{1}{\Psi(c^*)} B_{l_1 \cdots l_t}^{-1} \boldsymbol{Z}_{l_1 \cdots l_t}^{i_{l_1} \cdots i_{l_t}} \\
&= \frac{1}{\Psi(c^*)} ((\boldsymbol{Z}_{l_1}^{i_{l_1}})^T \otimes \cdots \otimes (\boldsymbol{Z}_{l_t}^{i_{l_t}})^T)(B_{l_1}^{-1} \otimes \cdots \otimes B_{l_t}^{-1})(\boldsymbol{Z}_{l_1}^{i_{l_1}} \otimes \cdots \otimes \boldsymbol{Z}_{l_t}^{i_{l_t}}) \\
&= \frac{1}{\Psi(c^*)} \left[(\boldsymbol{Z}_{l_1}^{i_{l_1}})^T B_{l_1}^{-1} \boldsymbol{Z}_{l_1}^{i_{l_1}} \right] \otimes \cdots \otimes \left[(\boldsymbol{Z}_{l_t}^{i_{l_t}})^T B_{l_t}^{-1} \boldsymbol{Z}_{l_t}^{i_{l_t}} \right] \\
&= \frac{1}{\Psi(c^*)} \prod_{i=1}^{t} (s_{l_i} - 1).
\end{aligned}
$$

Combining all of the above, we find that

$$
\begin{aligned}
(\boldsymbol{D}^{i_1 \cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta}_1) \boldsymbol{D}^{i_1 \cdots i_L} &= \frac{1}{\Psi(c^*)} \left\{ 1 + \sum_{t=1}^{L} \sum_{(l_1,\ldots,l_t) \in G_t} \left[\prod_{i=1}^{t} (s_{l_i} - 1) \right] + \frac{c^2}{(c^*)^2} \right\} \\
&= \frac{1}{\Psi(c^*)} \left\{ r - 1 + \frac{c^2}{(c^*)^2} \right\},
\end{aligned}
$$

where the last equality follows from the definition of r in Eq. 6. So the requirement in Eq. (11) becomes

$$
\Psi(c)(\boldsymbol{D}^{i_1 \cdots i_L})^T M_{\xi^*}^{-1}(\boldsymbol{\theta}_1) \boldsymbol{D}^{i_1 \cdots i_L} = \frac{\Psi(c)}{\Psi(c^*)}(r-1) + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)} \leq r, \quad c \in (-\infty, \infty). \tag{12}
$$

The validity of this inequality is stated in the following lemma.

Lemma 3 *For the logit and probit links,*

$$
\frac{\Psi(c)}{\Psi(c^*)}(r-1) + \frac{c^2 \Psi(c)}{(c^*)^2 \Psi(c^*)} \leq r
$$

for any $r \geq 2$ *and* $c \in (-\infty, \infty)$.

Proof See "Appendix." □

This concludes the proof of Theorem 1. □

Note that the value of c^* in Theorem 1 does not depend on the values of the unknown parameters $\boldsymbol{\theta}$. It merely depends on the value of the rank r and the link function. However, to compute the covariate values $x_{i_1 i_2 \cdots i_L j}$ from the points $-c^*$ and c^* in each cell, we do have to use the best guess for $\boldsymbol{\theta}$. This observation emphasizes that the designs in Theorem 1 are locally optimal designs.

3 An Illustrative Example

To illustrate the main result, we consider an electrostatic discharge (ESD) study originally reported by [7]. It will highlight strengths and weaknesses of our result. The study used a logistic model to assess the effect of four factors on the failure rate of semiconductors when exposed to electrostatic discharge. The four factors and the continuous covariate are displayed in Table 1.

The original study used the factor wafer lot with four levels (lot 1, 2, 3 and 4), but the experimenters decided to code it using two two-level factors, Lot A and Lot B, and ignore a possible interaction between them. We will use Lot A and Lot B in the same way. ESD handling is a factor to indicate whether or not the standard procedure was applied. No ESD handling means that no ESD-safe laboratory coat/shoes and no wrist strap were used. The ESD testing involves "zapping" a part first with a pulse polarity (positive or negative) and then followed by a second pulse of the opposite polarity. Since there is no industry standard specifying the order of pulse polarity, this makes it the fourth factor. The continuous covariate is the voltage each wafer was tested at, and the response variable is binary: A wafer either passes or fails the test. Taking the factor levels as 1 and -1 and denoting them by x_1 through x_4 (using the same order for the factors as in Table 1), with p and x denoting the probability of a wafer passing the test and the voltage, respectively, the model that was used is

$$logit(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{34} x_3 x_4 + \beta_5 x. \tag{13}$$

Note that the linear predictor includes an interaction term. While Model (13) is a reparameterization of Model (2), Theorem 1 remains applicable.

The experimenters 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. There is no discussion about reasons for selecting the voltage

Table 1 Factors and covariate for ESD experiment

	Description	Levels	
Factors	Lot A	Location 1	Location 2
	Lot B	Location 1	Location 2
	ESD handling	No	Yes
	Pulse polarity	Negative	Positive
Covariate	Voltage	Continuous	

range as [25, 45] or for selecting 5 levels. We will 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 $\boldsymbol{\beta_0} = (-7.50, 1.50, -0.20, -0.15, 0.25, 0.40, 0.35)^T$ as in [3], which is based on estimates from the original study. Then, according to Theorem 1, a D-optimal design using a full factorial and two support points in each group is given by the design in Table 2.

Using approximate design representations, the D-efficiency of the design used in the study, ξ_0, relative to the optimal design in Table 2, ξ^*, can be computed as

$$RE(\xi_0) = \left[\frac{det(I_{\xi_0})}{det(I_{\xi^*})}\right]^{1/p}, \tag{14}$$

where $p = 7$ is the number of parameters in the model. We find that $RE_{(}\xi_0) = 24.22\%$, suggesting that the optimal design is more than four times as efficient as the design that was used. While [7] may not have had good guesses for the parameters in the model, this example shows how to apply our theoretical results if such information is available and demonstrates convincingly that the theoretical results can then lead to the use of better designs.

4 Summary and Discussion

Even though GLMs with factorial effects have been widely used in multiple disciplines, optimal design theory for such models is limited. In this paper, we proposed a locally D-optimal design for logistic and probit links for a general model as in Model (2). This model can include any number of interaction terms between the factors. The power of the main result has been illustrated through an example, which

Table 2 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

shows that the same amount of information could have been obtained with far fewer runs. Such theoretical results are also very useful to provide benchmarks for other designs. What is difficult to incorporate in the theory is possible restrictions on the design space. For example, if the voltage had to be between 25 and 45 Volt, then the design in Table 2 would not have been feasible. Computational approaches as in [3] are important to solve specific problems under such restrictions.

With a large number of factors or levels, the number of support points, $2s$, of the optimal design in Theorem 1 can become large. Rather than using a full factorial design, one could then consider using a fractional factorial design or a design that has less then 2 observations per cell. Results for this will be reported in [6].

5 Appendix

Proof of Lemma 1 By the definition given through Eq. (10),

$$M_{\xi^*}(\boldsymbol{\theta_1}) = \frac{1}{2s}\Psi(c^*)\sum_{i_1=1}^{s_1}\cdots\sum_{i_L=1}^{s_L}\sum_{j=1}^{2}\boldsymbol{D}^{i_1\cdots i_L j}(\boldsymbol{D}^{i_1\cdots i_L j})^T.$$

For the symmetrical summand we have

$$\boldsymbol{D}^{i_1\cdots i_L j}(\boldsymbol{D}^{i_1\cdots i_L j})^T = \begin{pmatrix} 1 & (\boldsymbol{Z}_1^{i_1})^T & \cdots & (\boldsymbol{Z}_L^{i_L})^T & \cdots & (\boldsymbol{Z}_{l_1l_2}^{i_{l_1}i_{l_2}})^T & \cdots & c_{i_1\cdots i_L j} \\ & \boldsymbol{Z}_1^{i_1}(\boldsymbol{Z}_1^{i_1})^T & \cdots & \boldsymbol{Z}_1^{i_1}(\boldsymbol{Z}_L^{i_L})^T & \cdots & \boldsymbol{Z}_1^{i_1}(\boldsymbol{Z}_{l_1l_2}^{i_{l_1}i_{l_2}})^T & \cdots & c_{i_1\cdots i_L j}\boldsymbol{Z}_1^{i_1} \\ & & \ddots & \vdots & & \vdots & & \vdots \\ & & & \boldsymbol{Z}_L^{i_L}(\boldsymbol{Z}_L^{i_L})^T & \cdots & \boldsymbol{Z}_L^{i_L}(\boldsymbol{Z}_{l_1l_2}^{i_{l_1}i_{l_2}})^T & \cdots & c_{i_1\cdots i_L j}\boldsymbol{Z}_L^{i_L} \\ & & & & \ddots & \vdots & & \vdots \\ & & & & & \boldsymbol{Z}_{l_1l_2}^{i_{l_1}i_{l_2}}(\boldsymbol{Z}_{l_1l_2}^{i_{l_1}i_{l_2}})^T & \cdots & c_{i_1\cdots i_L j}\boldsymbol{Z}_{l_1l_2}^{i_{l_1}i_{l_2}} \\ & & & & & & \ddots & \vdots \\ & & & & & & & c_{i_1\cdots i_L j}^2 \end{pmatrix}$$

The top-left element of $M_{\xi^*}(\boldsymbol{\theta_1})$ is thus

$$\frac{1}{2s}\Psi(c^*)\sum_{i_1=1}^{s_1}\cdots\sum_{i_L=1}^{s_L}\sum_{j=1}^{2}1 = \Psi(c^*),$$

while the bottom-right element is

$$\frac{1}{2s}\Psi(c^*)\sum_{i_1=1}^{s_1}\cdots\sum_{i_L=1}^{s_L}\sum_{j=1}^{2}(c^*)^2 = (c^*)^2\Psi(c^*).$$

All other elements in the last column of $M_{\xi^*}(\boldsymbol{\theta_1})$ are 0 because each cell has two opposite values for the c values.

An off-diagonal block of $M_{\xi^*}(\boldsymbol{\theta_1})$ is of the form

$$\boldsymbol{Z}_{r_1r_2\cdots r_m}^{i_{r_1}i_{r_2}\cdots i_{r_m}} \cdot (\boldsymbol{Z}_{c_1c_2\cdots c_n}^{i_{c_1}i_{c_2}\cdots i_{c_n}})^T = (\boldsymbol{Z}_{r_1}^{i_{r_1}} \otimes \cdots \otimes \boldsymbol{Z}_{r_m}^{i_{r_m}}) \cdot ((\boldsymbol{Z}_{c_1}^{i_{c_1}})^T \otimes \cdots \otimes (\boldsymbol{Z}_{c_n}^{i_{c_n}})^T).$$

Since it is an off-diagonal block, at least one factor appears only once in the two sets of factors. Without loss of generality, say r_1 does not appear in $(c_1, \dots, c_n)$. By summing over the levels of factor r_1, we obtain

$$\begin{aligned}
&\sum_{i_{r_1}=1}^{s_{r_1}} \boldsymbol{Z}_{r_1r_2\cdots r_m}^{i_{r_1}i_{r_2}\cdots i_{r_m}} \cdot (\boldsymbol{Z}_{c_1c_2\cdots c_n}^{i_{c_1}i_{c_2}\cdots i_{c_n}})^T \\
&\quad = \left(\left(\sum_{i_{r_1}=1}^{s_{r_1}} \boldsymbol{Z}_{r_1}^{i_{r_1}} \right) \otimes \boldsymbol{Z}_{r_2}^{i_{r_2}} \otimes \cdots \otimes \boldsymbol{Z}_{r_m}^{i_{r_m}} \right) \cdot (\boldsymbol{Z}_{c_1c_2\cdots c_n}^{i_{c_1}i_{c_2}\cdots i_{c_n}})^T = 0.
\end{aligned}$$

The last equality follows immediately from the definition of the $\boldsymbol{Z}_l^{i_l}$'s.

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

$$\begin{aligned}
&\frac{1}{2s}\Psi(c^*) \sum_{i_1=1}^{s_1} \cdots \sum_{i_L=1}^{s_L} \sum_{j=1}^{2} \boldsymbol{Z}_l^{i_l} (\boldsymbol{Z}_l^{i_l})^T \\
&\quad = \frac{\Psi(c^*)}{s_l} \cdot \left[\begin{pmatrix} 1 \\ -\frac{1}{s_l-1} \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \left(1, -\frac{1}{s_l - 1}, \dots, -\frac{1}{s_l - 1}\right) \right. \\
&\qquad \left. + \cdots + \begin{pmatrix} -\frac{1}{s_l-1} \\ -\frac{1}{s_l-1} \\ \vdots \\ -\frac{1}{s_l-1} \end{pmatrix} \left(-\frac{1}{s_l - 1}, -\frac{1}{s_l - 1}, \dots, -\frac{1}{s_l - 1} \right) \right] \\
&\quad = \frac{\Psi(c^*)}{s_l} \cdot s_l \cdot \begin{pmatrix} \frac{1}{(s_l-1)} & -\frac{1}{(s_l-1)^2} & \cdots & \cdots & -\frac{1}{(s_l-1)^2} \\ & \frac{1}{(s_l-1)} & -\frac{1}{(s_l-1)^2} & \cdots & -\frac{1}{(s_l-1)^2} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & -\frac{1}{(s_l-1)^2} \\ & & & & \frac{1}{(s_l-1)} \end{pmatrix} = \Psi(c^*) \cdot B_l.
\end{aligned}$$

For a diagonal block that corresponds to an interaction, say for factors $(l_1, \dots, l_t)$, we get

$$
\begin{aligned}
&\frac{1}{2s}\Psi(c^*)\sum_{i_1=1}^{s_1}\cdots\sum_{i_L=1}^{s_L}\sum_{j=1}^{2}\boldsymbol{Z}_{l_1\cdots l_t}^{i_{l_1}\cdots i_{l_t}}(\boldsymbol{Z}_{l_1\cdots l_t}^{i_{l_1}\cdots i_{l_t}})^T \\
&= \frac{1}{s}\Psi(c^*)\underbrace{\sum\cdots\sum}_{not\,(l_1,\ldots,l_t)}\underbrace{\sum\cdots\sum}_{(l_1,\ldots,l_t)}\boldsymbol{Z}_{l_1\cdots l_t}^{i_{l_1}\cdots i_{l_t}}(\boldsymbol{Z}_{l_1\cdots l_t}^{i_{l_1}\cdots i_{l_t}})^T \\
&= \frac{1}{s}\Psi(c^*)\underbrace{\sum\cdots\sum}_{not\,(l_1,\ldots,l_t)}\underbrace{\sum\cdots\sum}_{(l_1,\ldots,l_t)}(\boldsymbol{Z}_{l_1}^{i_{l_1}}\otimes\cdots\otimes\boldsymbol{Z}_{l_t}^{i_{l_t}})\cdot((\boldsymbol{Z}_{l_1}^{i_{l_1}})^T\otimes\cdots\otimes(\boldsymbol{Z}_{l_t}^{i_{l_t}})^T) \\
&= \frac{1}{s}\Psi(c^*)\underbrace{\sum\cdots\sum}_{not\,(l_1,\ldots,l_t)}\left(\sum_{i_{l_1}=1}^{s_{l_1}}\boldsymbol{Z}_{l_1}^{i_{l_1}}(\boldsymbol{Z}_{l_1}^{i_{l_1}})^T\right)\otimes\cdots\otimes\left(\sum_{i_{l_t}=1}^{s_{l_t}}\boldsymbol{Z}_{l_t}^{i_{l_t}}(\boldsymbol{Z}_{l_t}^{i_{l_t}})^T\right) \\
&= \frac{1}{s}\Psi(c^*)\underbrace{\sum\cdots\sum}_{not\,(l_1,\ldots,l_t)}(s_{l_1}\cdot B_{l_1})\otimes\cdots\otimes(s_{l_t}\cdot B_{l_t}) \\
&= \frac{1}{s}\Psi(c^*)\cdot s\cdot B_{l_1}\otimes\cdots\otimes B_{l_t} = \Psi(c^*)\cdot B_{l_1}\otimes\cdots\otimes B_{l_t}
\end{aligned}
$$

This concludes the proof. □

Proof of Lemma 2 If $i_l < s_l$, $\boldsymbol{Z}_l^{i_l}$ is of the form $(-\frac{1}{s_l-1},\ldots,1,\ldots,-\frac{1}{s_l-1})^T$. So

$$
\begin{aligned}
&(\boldsymbol{Z}_l^{i_l})^T\boldsymbol{Z}_l^{i_l} + (\boldsymbol{Z}_l^{i_l})^T J\boldsymbol{Z}_l^{i_l} \\
&= \left(-\frac{1}{s_l-1},\ldots,1,\ldots,-\frac{1}{s_l-1}\right)\left(-\frac{1}{s_l-1},\ldots,1,\ldots,-\frac{1}{s_l-1}\right)^T \\
&\quad + \left(-\frac{1}{s_l-1},\ldots,1,\ldots,-\frac{1}{s_l-1}\right)\cdot J\cdot\left(-\frac{1}{s_l-1},\ldots,1,\ldots,-\frac{1}{s_l-1}\right)^T \\
&= \left(\frac{1}{(s_l-1)^2}\cdot(s_l-2)+1\right)+\frac{1}{s_l-1}\cdot 1'\left(-\frac{1}{s_l-1},\ldots,1,\ldots,-\frac{1}{s_l-1}\right)^T \\
&= \frac{s_l-2}{(s_l-1)^2}+1+\frac{1}{(s_l-1)^2} = \frac{s_l}{s_l-1}
\end{aligned}
$$

If $i_l = s_l$, then $\boldsymbol{Z}_l^{i_l}$ is in the form of $(-\frac{1}{s_l-1},\ldots,-\frac{1}{s_l-1})^T$. So

$$
\begin{aligned}
(\boldsymbol{Z}_l^{i_l})^T\boldsymbol{Z}_l^{i_l} + (\boldsymbol{Z}_l^{i_l})^T J\boldsymbol{Z}_l^{i_l} &= \left(-\frac{1}{s_l-1},\ldots,-\frac{1}{s_l-1}\right)\left(-\frac{1}{s_l-1},\ldots,-\frac{1}{s_l-1}\right)^T \\
&\quad + \left(-\frac{1}{s_l-1},\ldots,-\frac{1}{s_l-1}\right)\cdot J\cdot\left(-\frac{1}{s_l-1},\ldots,-\frac{1}{s_l-1}\right)^T \\
&= \frac{1}{(s_l-1)^2}\cdot(s_l-1)+(-1)^2 = \frac{s_l}{s_l-1}
\end{aligned}
$$

Together, this completes the proof. □

Proof of Lemma 3 For models of form (1) with logistic or probit link, if there is no interaction effect among factors, then according to Theorem 4 in [4], a locally D-optimal design for a maximal set of estimable functions of the parameter vector still has the same design structure as described in Theorem 1, and only the power term r in $f(c) = c^2(\Psi(c))^r$ refers to the main-effects model, with an explicit expression as $r = \sum_1^L (s_l - 1) + 2$.

Then, following the same reparameterization method as shown in the proof of Theorem 1, since the equivalence theorem can be used to verify the locally D-optimal design for the main-effects model, we have

$$\frac{\Psi(c)}{\Psi(c^*)}(r-1) + \frac{c^2\Psi(c)}{(c^*)^2\Psi(c^*)} \le r.$$

This concludes the lemma. □

Proof of Reparameterization To show that model (9) is a reparameterization of model (1), we need to show that the column spaces spanned by the two design matrices are the same. Since the vector of covariate values is the same in both, define $X = [\boldsymbol{x_0}, \boldsymbol{x_1^1}, \dots, \boldsymbol{x_1^{s_1}}, \dots, \boldsymbol{x_L^{s_L}}, \dots, \boldsymbol{x_{l_1 l_2}^{i_{l_1} i_{l_2}}}, \dots]$ and $Z = [\boldsymbol{z_0}, \boldsymbol{z_1^1}, \dots, \boldsymbol{z_1^{s_1-1}}, \dots, \boldsymbol{z_L^{s_L-1}}, \dots, \boldsymbol{z_{l_1 l_2}^{i_{l_1} i_{l_2}}}, \dots]$ for models (1) and (9), respectively. In both cases, the only interaction terms included are those that appear in the model, but the columns of X correspond to the sets G_t, while those in Z correspond to the H_t's, $t = 1, \dots, L$. We need to show that $\mathcal{C}(X) = \mathcal{C}(Z)$.

For any $\boldsymbol{v} \in \mathcal{C}(Z)$, there exist $a_0, a_1^1, \dots, a_1^{s_1-1}, \dots$ such that

$$\begin{aligned}\boldsymbol{v} = {} & a_0 \boldsymbol{z_0} + a_1^1 \boldsymbol{z_1^1} \cdots + a_L^{s_L-1} \boldsymbol{z_L^{s_L-1}} \\ & + \sum_{t=2}^{L} \sum_{(l_1,\cdots,l_t)\in H_t} \left[\sum_{i_{l_1}=1}^{s_{l_1}-1} \cdots \sum_{i_{l_t}=1}^{s_{l_t}-1} a_{l_1\cdots l_t}^{i_{l_1}\cdots i_{l_t}} \boldsymbol{z_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}}} \right].\end{aligned}$$

Since $\boldsymbol{z_0} = \boldsymbol{x_0}$ and, for each factor l,

$$\boldsymbol{z_l^{i_l}} = \boldsymbol{x_l^{i_l}} - \frac{1}{s_l - 1} \sum_{i \ne i_l} \boldsymbol{x_l^i} = \frac{1}{s_l - 1}\left(s_l \boldsymbol{x_l^{i_l}} - \boldsymbol{x_0}\right), \tag{15}$$

we see that $\boldsymbol{z_l^{i_l}}$ can be written as a linear combination of the columns in X.

For interaction terms, we have that

$$\boldsymbol{z_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}}} = \boldsymbol{z_{l_1}^{i_{l_1}}} \circ \boldsymbol{z_{l_2}^{i_{l_2}}} \circ \cdots \circ \boldsymbol{z_{l_t}^{i_{l_t}}} = \frac{\left(s_{l_1} \boldsymbol{x_{l_1}^{i_{l_1}}} - \boldsymbol{x_0}\right) \circ \cdots \circ \left(s_{l_t} \boldsymbol{x_{l_t}^{i_{l_t}}} - \boldsymbol{x_0}\right)}{\prod_{i=1}^{t} (s_{l_i} - 1)} \tag{16}$$

where "∘" represents the Hadamard product. The expression in (16) is a linear combination of $\boldsymbol{x_0}$ and Hadamard products of t or fewer $\boldsymbol{x_l^i}$'s. The subscripts in these Hadamard product terms form subsets of $\{l_1, \dots, l_t\}$ and correspond to interaction terms. For example, $\boldsymbol{x_1^1} \circ \boldsymbol{x_2^1} = \boldsymbol{x_{12}^{11}}$. Not all of these interactions appear in the model though. But by definition of H_t, there is a $t' \ge t$ and a t'-tuple $(j_1, \dots, j_{t'}) \in G_{t'}$ that contains all of $(i_1, \ldots, i_t)$. Thus the columns $\boldsymbol{x}_{j_1 \cdots j_{t'}}^{i_{j_1} \cdots i_{j_{t'}}}$ do appear in X, and each column corresponding to a Hadamard product term in (16) can be written as a linear combination of these columns. Hence, all columns in Z can be written as a linear combination of columns in X. Hence $\mathcal{C}(Z) \subset \mathcal{C}(X)$

For any $\boldsymbol{v} \in \mathcal{C}(X)$, there exist $b_0, b_1^1, \ldots, b_1^{s_1}, \ldots$ such that

$$\begin{aligned}\boldsymbol{v} = {} & b_0\boldsymbol{x_0} + b_1^1\boldsymbol{x_1^1} + \cdots + b_L^{s_L}\boldsymbol{x_L^{s_L}} \\ & + \sum_{t=2}^{L} \sum_{(l_1,\cdots,l_t)\in G_t} \left[\sum_{i_{l_1}=1}^{s_{l_1}} \cdots \sum_{i_{l_t}=1}^{s_{l_t}} b_{l_1\cdots l_t}^{i_{l_1}\cdots i_{l_t}} \boldsymbol{x}_{l_1 l_2 \cdots l_t}^{i_{l_1} i_{l_2} \cdots i_{l_t}} \right].\end{aligned}$$

It is easy to verify that each of these terms is a linear combination of $\boldsymbol{z_0}$, the $\boldsymbol{z}_l^{i_l}$'s and Hadamard products of these vectors. Along the same lines as in the first part of this proof, it follows that $\mathcal{C}(X) \subset \mathcal{C}(Z)$

Combined this gives that $\mathcal{C}(X) = \mathcal{C}(Z)$. □

Compliance with Ethical Standards

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

References

1. Kiefer J (1974) General equivalence theory for optimum designs (approximate theory). Ann Stat 2:849–879
2. Kiefer J, Wolfowitz J (1960) The equivalence of two extremum problems. Can J Math 12:363–366
3. Lukemire J, Mandal A, Wong WK (2019) 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
4. Stufken J, Yang M (2012) On locally optimal designs for generalized linear models with group effects. Stat Sinica 22:1765–1786
5. Tan X (2015) Optimal designs for generalized linear models. Unpublished Ph.D. dissertation, University of Georgia
6. Wang Z, Stufken J (2019) Smaller D-optimal designs for GLMs with factorial effects (working title). In preparation
7. Whitman C, Gilbert TM, Rahn AM, Antonell JA (2006) Determining factors affecting ESD failure voltage using DOE. Microelectron Reliab 46(8):1228–1237