Statistica Sinica **22** (2012), 1765-1786
doi:http://dx.doi.org/10.5705/ss.2010.261

ON LOCALLY OPTIMAL DESIGNS FOR GENERALIZED LINEAR MODELS WITH GROUP EFFECTS

John Stufken and Min Yang

University of Georgia and University of Illinois at Chicago

Abstract: Generalized linear models with group effects are commonly used in scientific studies. However, there appear to be no results for selecting optimal designs. In this paper, we identify the structure of locally optimal designs, provide a general strategy to determine the design points and the corresponding weights for optimal designs, and present theoretical results for the special case of D-optimality. The results can be applied to many commonly studied models, including the logistic, probit, and loglinear models. The design region can be restricted or unrestricted, and the results can also be applied for a multi-stage approach.

Key words and phrases: A-optimality, binary response, D-optimality, Loewner ordering, logistic model, loglinear model, probit model.

1. Introduction

Categorical response variables are common in such areas of research as public health, medical sciences, social sciences, and marketing. While using generalized linear models (GLMs) for analyzing such data has become common with advances in computational tools, the study of optimal design for experiments with such data is in a very underdeveloped stage. Even though a number of notable contributions have been made in the area (e.g., Ford, Torsney, and Wu (1992); Biedermann, Dette, and Zhu (2006)), Khuri et al. (2006) surveyed design issues for GLMs and noted that "The research on designs for generalized linear models is still very much in its developmental stage. Not much work has been accomplished either in terms of theory or in terms of computational methods to evaluate the optimal design when the dimension of the design space is high. The situations where one has several covariates (control variables) or multiple responses corresponding to each subject demand extensive work to evaluate "optimal" or at least efficient designs." This is especially true for models with multiple parameters. In particular, there appear to be no optimal designs for generalized linear models with group effects.

Yang and Stufken (2009) proposed a new algebraic approach to the study of locally optimal designs for GLMs with two parameters. For a given model, their approach identifies a class of relatively simple designs so that for any design d that does not belong to this class, there is a design in the class that has an information matrix that dominates d in the Loewner ordering. The result can be used for restricted or unrestricted design regions and can also be applied for a multi-stage approach. It makes identifying locally optimal designs a straightforward task for many important models and optimality criteria.

In this paper, we extend that approach to models that include group effects. This is an important extension that allows for heterogeneity among subjects (see, for example, Cook and Thibodeau (1980), and Tighiouart and Rogatko (2006)). Focusing on A- or D-optimality for estimable functions, for various models we provide a strategy to determine design points and corresponding weights for locally optimal designs. The result is significant since it provides a feasible strategy for finding optimal designs under the A- or D-optimality criterion while allowing arbitrary subsets of estimable parameter functions, restricted or unrestricted design regions, and one-stage or multi-stage approaches. We refer to Yang and Stufken (2009) for more detail on the importance of this flexibility.

This paper is organized as follows. In Section 2, we introduce the various models. In Section 3, we identify the structure of optimal designs for GLMs with group effects. This is used to derive explicit forms of D-optimal designs for special cases in Section 4. A strategy for finding optimal weights for given design points is presented in Section 5, followed in Section 6 by two examples to illustrate the computations required for finding optimal designs. A closing discussion is presented in Section 7, while the proof of a technical result can be found in the Appendix.

2. Statistical Models and Information Matrices

We first present the GLMs that we study in later sections. We distinguish between models for binary data and count data. Subsequently we present information matrices for these models. The focus is on models that include parameters for group effects, such as race, ethnicity, gender, or other categorical variables. These models have been studied extensively for data analysis (see for example, Agresti (2002)), but little is known about design selection.

2.1. Generalized linear regression models for binary data

The simplest models are of the form

$$Prob(Y_i = 1) = P(\alpha + \beta x_i). \tag{2.1}$$

Here, Y_i and x_i are the response and covariate for subject i, $i = 1, \ldots, n$, α and β are the intercept and slope parameters, and $P(x)$ is a cumulative distribution function, such as $e^x/(1+e^x)$ for the logistic model or $\Phi(x)$, the cdf for the standard normal, for the probit model. Model (2.1) has been studied extensively in the optimal design literature and we refer to Yang and Stufken (2009) for selected references.

These models, however, do not include parameters that allow for subject heterogeneity. In presenting models that do, we assume that there are L factors, $1, \ldots, L$, with numbers of levels $s_1, \ldots, s_L$, that partition the subjects into $k = s_1 \ldots s_L$ groups. We consider both a model with a common slope for all groups and a model that allows for different slopes for different groups.

The simplest presentation for the model with a common slope for all subject groups is

$$Prob(Y_{ij} = 1) = P(\alpha_0 + \alpha_i + \beta x_{ij}),$$

where Y_{ij} and x_{ij} are the response and covariate value for the jth subject in group i, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, β is the common slope effect, and α_i is the effect for the ith group. For example, with only main-effects, α_i could be parametrized as $\alpha_i = \alpha_1(i) + \ldots + \alpha_L(i)$, where $\alpha_l(i)$ is an effect due to the level of the lth factor in the ith group but, more generally α_i could contain interaction effects of two or more factors.

We write the model as

$$Prob(Y_{ij} = 1) = P((X^{ij})^T \theta), \tag{2.2}$$

where $\theta = (\alpha_0, \alpha^T, \beta)^T$, a $(k+2) \times 1$ vector; $\alpha = (\alpha_1, \ldots, \alpha_k)^T$; $X^{ij} = (1, X_i^T, x_{ij})^T$; and X_i is a $k \times 1$ vector with a 1 in position i and zeros elsewhere. Note that simplifying assumptions about the model (such as the absence of some or all interactions) would allow a reparametrization with fewer parameters.

For the model that facilitates different slopes for the k different groups, using notation as in (2.2), we write the model as

$$\begin{aligned} Prob(Y_{ij} = 1) &= P(\alpha_0 + \alpha_i + \beta_i x_{ij}) = P(\alpha_0 + X_i^T \alpha + X_i^T \beta x_{ij}) \\ &= P((X^{ij})^T \theta). \end{aligned} \tag{2.3}$$

Here $\beta = (\beta_1, \ldots, \beta_k)^T$ is a vector instead of a scalar. Now $\theta = (\alpha_0, \alpha^T, \beta^T)^T$ and $X^{ij} = (1, X_i^T, x_{ij} X_i^T)^T$ are $(2k+1) \times 1$ vectors. It will be clear from the context whether θ, β, and X^{ij} are as in (2.2) or as in (2.3).

2.2. Loglinear regression models for count data

In the medical and social sciences one finds experiments with a response variable based on counts, such as the number of times that a certain event occurs during a given time period or within a territory. Such counts, or the rate of occurrence, is usually modeled by a loglinear regression model (Agresti (2002, Chap. 9)).

In the presence of L factors forming k groups, as in Subsection 2.1, let Y_{ij}, the response of subject j in group i, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, have a Poisson distribution with parameter λ_{ij}. Let x_{ij} be the covariate value (for example the concentrate of a drug) for this subject. Using the notation from Model (2.2), a common slope model can now be written as

$$\begin{aligned} \log(\lambda_{ij}) &= \alpha_0 + \alpha_i + \beta x_{ij} \\ &= (X^{ij})^T \theta. \end{aligned} \tag{2.4}$$

Using the notation from Model (2.3), the model with different slopes for different groups can be written as

$$\begin{aligned} \log(\lambda_{ij}) &= \alpha_0 + \alpha_i + X_i^T \beta x_{ij} \\ &= (X^{ij})^T \theta. \end{aligned} \tag{2.5}$$

2.3. Information matrices

For the problem considered, an exact design can be presented as $\{(x_{ij}, n_{ij}),$ $i = 1, \ldots, k$, $j = 1, \ldots, m_i\}$, where x_{ij} is the j-th distinct covariate value used in group i, m_i is the number of distinct predictor values used in group i, and n_{ij} is the number of subjects assigned to covariate value x_{ij}. With n denoting the total number of subjects, we have that $\sum_i \sum_j n_{ij} = n$. Since finding an optimal exact design is a difficult and often intractable optimization problem, the corresponding approximate design, in which n_{ij}/n is replaced by ω_{ij}, is considered. Thus a design can be denoted by $\xi = \{(x_{ij}, \omega_{ij}), i = 1, \ldots, k, j = 1, \ldots, m_i\}$, where $\omega_{ij} > 0$ and $\sum_i \sum_j \omega_{ij} = 1$. For known parameters, in each group i, there is a one to one mapping between x_{ij} and c_{ij}, where $c_{ij} = (X^{ij})^T \theta$. It turns out to be convenient to denote design ξ as $\xi = \{(c_{ij}, \omega_{ij}), i = 1, \ldots, k, j = 1, \ldots, m_i\}$.

By standard methods, the information matrix for θ under Models (2.2), (2.3), (2.4), and (2.5), can be written as

$$I_\xi(\theta) = n \sum_{i=1}^{k} \sum_{j=1}^{m_i} \omega_{ij} X^{ij} \Psi(c_{ij}) (X^{ij})^T \tag{2.6}$$

$$= nXV\Omega VX^T, \tag{2.7}$$

where $\Psi(x) = [P'(x)]^2/[P(x)(1 - P(x))]$ (Models 2.2 and 2.3) or $\Psi(x) = \exp(x)$ (Models 2.4 and 2.5), $X = (X^{11}, X^{12}, \ldots, X^{k,m_k})$, V is a diagonal matrix with diagonal elements $(\sqrt{\Psi(c_{11})}, \sqrt{\Psi(c_{12})}, \ldots, \sqrt{\Psi(c_{k,m_k})})$, and Ω is a diagonal matrix with diagonal elements $(\omega_{11}, \omega_{12}, \ldots, \omega_{k,m_k})$. While for simplification we use the same notation for all models, note that the definitions of X^{ij} and θ are different under different models.

We write

$$X^{ij} = A_i(\alpha,\beta)C^{ij}. \tag{2.8}$$

Here, $C^{ij} = (1, X_i^T, c_{ij})^T$ (Models (2.2) and (2.4)) or $C^{ij} = (1, X_i^T, c_{ij}X_i^T)^T$ (Models (2.3) and (2.5)); and $A_i(\alpha,\beta)$ is of the form $\begin{pmatrix} I_{k+1} & 0 \\ A_{i(1)}(\alpha,\beta) & A_{(2)}(\beta) \end{pmatrix}$, where 0 is the zero matrix of appropriate dimensions. Matrices $A_{i(1)}(\alpha,\beta)$ and $A_{(2)}(\beta)$ depend on the model. Under Models (2.2) and (2.4) (where β is a scalar), $A_{i(1)}(\alpha,\beta) = (-\alpha_0/\beta, -\alpha^T/\beta)$ and $A_{(2)}(\beta) = 1/\beta$. Under Models (2.3) and (2.5) (where β is a vector) $A_{i(1)}(\alpha,\beta)$ is a $k \times (k+1)$ matrix with all elements zero except the ith row; the ith row is $(-\alpha_0/\beta_i, -\alpha^T/\beta_i)$. $A_{(2)}(\beta)$ is the $k\times k$ diagonal matrix with elements $(1/\beta_1, \cdots, 1/\beta_k)$.

Using (2.8), the information matrix $I_\xi(\theta)$ in (2.6) can be rewritten as

$$I_\xi(\theta) = n\sum_{i=1}^{k}\sum_{j=1}^{m_i}\omega_{ij}A_i(\alpha,\beta)C^{ij}\Psi(c_{ij})(C^{ij})^T A_i^T(\alpha,\beta). \tag{2.9}$$

Suppose we are interested in $\eta = B\theta$. Since the models provide information for $X^T\theta$ only, the rows of B must belong to the row space of X^T, i.e., $B = DX^T$ for some matrix D. With $F(\eta)$ as a vector valued function of η, the covariance matrix of $F(\hat\eta)$, where $\hat\eta$ is the MLE of η, can be expressed as

$$\Sigma_\xi(F(\hat\eta)) = \frac{\partial F(\eta)}{\partial\eta^T}BI_\xi^-(\theta)B^T(\frac{\partial F(\eta)}{\partial\eta^T})^T. \tag{2.10}$$

From (2.7), it follows that $X^T I_\xi^-(\theta)X$ is invariant to the choice of the g-inverse $I_\xi^-(\theta)$, which implies that the same is true for (2.10).

3. Structure of Optimal Designs

An optimal design for $F(\eta)$ maximizes the corresponding information matrix in some way, or equivalently minimizes the covariance matrix in (2.10) under a selected optimality criterion. Notice that for any two designs ξ_1 and ξ_2, if $I_{\xi_1}(\theta) \le I_{\xi_2}(\theta)$ (here and elsewhere, matrix inequalities are under the Loewner ordering), then there exist g-inverses $I_{\xi_1}^-(\theta)$ and $I_{\xi_2}^-(\theta)$ such that $I_{\xi_1}^-(\theta) \ge I_{\xi_2}^-(\theta)$ (see Theorem 5(i) of Wu (1980)). By (2.10), this implies that $\Sigma_{\xi_2}(F(\hat\eta)) \le \Sigma_{\xi_1}(F(\hat\eta))$. Thus design ξ_2 is at least as good as design ξ_1 for $F(\eta)$ under commonly used optimality criteria. Hence we can focus our attention on the matrices $I_\xi(\theta)$.

In this section, we show that for any given design $\xi = \{(c_{ij}, \omega_{ij}), i = 1, \ldots, k, j = 1, \ldots, m_i\}$, there exists a design ξ^* with a simple form such that $I_\xi(\theta) \le I_{\xi^*}(\theta)$. To identify optimal designs for $F(\eta)$ under the common optimality criteria based on information matrices, we can then restrict attention to designs with the simple form presented in this section. These optimality criteria include not just A-, D-, E-, L-, and Φ_p-optimality etc., but also standardized versions of optimality criteria proposed by Dette (1997).

Our results extend those of Yang and Stufken (2009), who considered models without group effects. Since we need their results here, we summarize them in two lemmas. Let $c_j = \alpha + \beta x_j$ and $c_j \in [D_1, D_2]$, a bounded or unbounded design region. From Yang and Stufken (2009), the information matrix for (α, β) in Model (2.1) under design $\xi = \{(c_j, w_j), j = 1, \ldots, m\}$, $I_\xi(\alpha, \beta)$, can be written as

$$I_\xi(\alpha, \beta) = A^T C_\xi(\alpha, \beta) A,$$

for a non-singular matrix A that does not depend on ξ, where

$$C_\xi(\alpha, \beta) = \sum_{j=1}^{m} \omega_j \begin{pmatrix} \Psi(c_j) & c_j \Psi(c_j) \\ c_j \Psi(c_j) & c_j^2 \Psi(c_j) \end{pmatrix}$$

and $\Psi(x) = [P'(x)]^2 / [P(x)[1 - P(x)]]$. Therefore, studying dominance in the Loewner ordering of one design over another can be done by studying $C_\xi(\alpha, \beta)$ rather than $I_\xi(\alpha, \beta)$.

Lemma 1. *For the logistic and probit models, as in Model* (2.1)*, for any design* $\xi = \{(c_j, \omega_j), j = 1, \ldots, m\}$*,* $m \geq 2$*, there exists a design* ξ^* *such that*

$$C_\xi(\alpha, \beta) \leq C_{\xi^*}(\alpha, \beta), \tag{3.1}$$

where ξ^* *has two support points. The two support points are* (i) c *and* $-c$ *if* $D_1 = -D_2$*;* (ii) D_1 *and* c *if* $D_1 > 0$*; (iii)* D_2 *and* c *if* $D_2 < 0$*; (iv)* D_1 *and* $c \in (|D_1|, D_2]$ *or* c *and* $-c$ *if* $D_1 < 0$ *and* $|D_1| < D_2$*; or (v)* D_2 *and* $c \in [D_1, -D_2)$ *or* c *and* $-c$ *if* $D_2 > 0$ *and* $|D_1| > D_2$*.*

Yang and Stufken (2009) establish a similar result for the loglinear model $\log(\lambda_j) = \alpha + \beta x_j$, using the same set up and notation as in Lemma 1, but now with $\Psi(x) = \exp(x)$.

Lemma 2. *With the loglinear model* $\log(\lambda_j) = \alpha + \beta x_j = c_j \in [D_1, D_2]$*, with* $D_2 < \infty$*, for any design* ξ*, there exists a design* ξ^* *such that*

$$C_\xi(\alpha, \beta) \leq C_{\xi^*}(\alpha, \beta), \tag{3.2}$$

where ξ^* *has two support points and one of these is* D_2*.*

The next theorems show how these results can be applied to Models (2.2), (2.3), (2.4), and (2.5). Due to possible constraints on the covariate value x_{ij}, we assume that $c_{ij} \in [D_{i1}, D_{i2}]$ for each $i = 1, \ldots, k$. For the loglinear model (Theorem 2), the D_{i2}'s are assumed to be finite.

Theorem 1. *In Models* (2.2) *and* (2.3)*, for any design* $\xi = \{(c_{ij}, \omega_{ij}),\ i = 1, \ldots, k,\ j = 1, \ldots, m_i\}$*, there exists a design* ξ^* *with at most two support points in each of the* k *groups such that* $I_\xi(\theta) \le I_{\xi^*}(\theta)$*. For each group where* ξ *has at least two support points, the two support points of* ξ^* *may be* (i) c_i *and* $-c_i$ *if* $D_{i1} = -D_{i2}$*;* (ii) D_{i1} *and* c_i *if* $D_{i1} > 0$*;* (iii) D_{i2} *and* c_i *if* $D_{i2} < 0$*; (iv)* D_{i1} *and* $c_i \in (|D_{i1}|, D_{i2}]$ *or* c_i *and* $-c_i$ *if* $D_{i1} < 0$ *and* $|D_{i1}| < D_{i2}$*; or (v)* D_{i2} *and* $c_i \in [D_{i1}, -D_{i2})$ *or* c_i *and* $-c_i$ *if* $D_{i2} > 0$ *and* $|D_{i1}| > D_{i2}$*. For groups where* ξ *has less then two support points,* ξ^* *can be taken to coincide with* ξ*.*

Proof. Recall that $C^{ij} = (1, X_i^T, c_{ij})^T$ (Model (2.2)) or $(1, X_i^T, c_{ij}X_i^T)^T$ (Model (2.3)). Thus C^{ij} can be written as $C^{ij} = B_i \begin{pmatrix} 1 \\ c_{ij} \end{pmatrix}$, where

$$B_i^T = \begin{pmatrix} 1 & X_i^T & 0 \\ 0 & 0_{1\times k} & 1 \end{pmatrix}$$

for Model (2.2), and

$$B_i^T = \begin{pmatrix} 1 & X_i^T & 0_{1\times k} \\ 0 & 0_{1\times k} & X_i^T \end{pmatrix}$$

for Model (2.3). Using (2.9), $I_\xi(\theta)$ can now be written as

$$I_\xi(\theta) = n \sum_{i=1}^{k} A_i(\alpha,\beta) B_i \underbrace{\sum_{j=1}^{m_i} \omega_{ij} \begin{pmatrix} \Psi(c_{ij}) & c_{ij}\Psi(c_{ij}) \\ c_{ij}\Psi(c_{ij}) & c_{ij}^2\Psi(c_{ij}) \end{pmatrix}}_{=C_\xi^i,\ \text{say}} B_i^T A_i^T(\alpha,\beta). \tag{3.3}$$

By (3.1), there exists a design ξ^* of the form mentioned in the statement of the theorem, such that for each i where ξ has at least two support points, $C_\xi^i \le C_{\xi^*}^i$. If ξ has less than two support points for some i, then we take ξ^* exactly the same as ξ for that group. This implies that, for each i,

$$A_i(\alpha,\beta) B_i C_\xi^i B_i^T A_i^T(\alpha,\beta) \le A_i(\alpha,\beta) B_i C_{\xi^*}^i B_i^T A_i^T(\alpha,\beta), \tag{3.4}$$

allowing the conclusion that $I_\xi(\theta) \le I_{\xi^*}(\theta)$.

Applying Lemma 2 and the same argument as in the proof of Theorem 1, we have similar results for Models (2.4) and (2.5).

Theorem 2. *In Models* (2.4) *and* (2.5)*, for any design* $\xi = \{(c_{ij}, \omega_{ij}),\ i = 1, \ldots, k, j = 1, \ldots, m_i\}$*, there exists a design* ξ^* *with at most two support points in each of the* k *groups such that* $I_\xi(\theta) \le I_{\xi^*}(\theta)$*. For each group where* ξ *has at least two support points, one of the two support points of* ξ^* *may be taken as* D_{i2}*; for groups where* ξ *has less then two support points,* ξ^* *can be taken to coincide with* ξ*.*

Note that, unlike Cook and Thibodeau (1980) whose study was in the context of linear models, we allow the weight for each group to be decided by optimality considerations. It should however be pointed out that if we would fix the group weights based on practical considerations, so that these are not subject to control by design, then the conclusion of Theorem 2 still holds.

4. D-Optimal Designs

While the characterizations in Theorems 1 and 2 generally require some computation for finding optimal designs, they can be used for deriving explicit expressions for D-optimal designs for certain families. In this section we first do this for Model (2.2) with a single factor at s levels and the design region $(-\infty, \infty)$. The results also apply with L factors provided that the model is the full factorial model; in that case the problem can be reparametrized as a single factor problem with $s = s_1 \cdots s_L$ levels. We consider two cases, one in which the parameters of interest correspond to the group effects and the slope parameter, $\alpha_0 + \alpha_1, \ldots, \alpha_0 + \alpha_s, \beta$, and the other in which the interest is in $s-1$ linearly independent contrasts of the group effects as well as the slope parameter.

Due to invariance of the D-optimality criterion under reparametrization (see, for example, Pukelsheim (2006)), we may take the parameter vectors for the two cases as $\eta_1 = ((\alpha_0+\alpha_1)/\beta, \ldots, (\alpha_0+\alpha_s)/\beta, \beta)^T$ and $\eta_2 = ((\alpha_1-\alpha_s)/\beta, \ldots, (\alpha_{s-1}-\alpha_s)/\beta, \beta)^T$, respectively. While a similar approach as used here can be used to derive explicit expressions for A-optimal designs for η_1 and η_2, we focus on deriving D-optimal designs since A-optimality is not invariant under reparametrizations (cf. Dette (1997)). With ξ as the design, let $I_\xi(\eta_1)$ and $I_\xi(\eta_2)$ denote the information matrices for the two cases.

Theorem 3. *Let design $\xi^* = \{(c_{i1} = c^*, \omega_{i1} = 1/(2s))$, $(c_{i2} = -c^*$, $\omega_{i2} = 1/(2s))$, $i = 1, \ldots, s\}$ for some c^*. For Model* (2.2) *with one factor at s levels and no constraint on the design space,*

(i) *ξ^* is D-optimal for η_1 if c^* maximizes $c^2\Psi^{s+1}(c)$; and*

(ii) *ξ^* is D-optimal for η_2 if c^* maximizes $c^2\Psi^s(c)$.*

Proof. Based on Theorem 1, we can restrict attention to designs $\xi = (c_{ij}, \omega_{ij})$ with (at most) two design points in each group. Let $\omega_{i+} = \omega_{i1} + \omega_{i2}$ and $\omega_{i-} = \omega_{i1} - \omega_{i2}$. The information matrix $I_\xi(\eta_1)$ can be written as

$$\begin{pmatrix} \beta^2\omega_{1+}\Psi(c_1) & 0 & \ldots & 0 & \omega_{1-}c_1\Psi(c_1) \\ 0 & \beta^2\omega_{2+}\Psi(c_2) & \ldots & 0 & \omega_{2-}c_2\Psi(c_2) \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \ldots & \beta^2\omega_{s+}\Psi(c_s) & \omega_{s-}c_s\Psi(c_s) \\ \omega_{1-}c_1\Psi(c_1) & \omega_{2-}c_2\Psi(c_2) & \ldots & \omega_{s-}c_s\Psi(c_s) & \frac{1}{\beta^2}\sum_{i=1}^s \omega_{i+}c_i^2\Psi(c_i) \end{pmatrix}.$$

Let $A_\xi(\eta_1) = (1/s!)\sum_Q Q^T I_\xi(\eta_1)Q$, where the sum is over all permutation matrices corresponding to permutations of $(1,\ldots,s)$. By Proposition A.2 in the Appendix, we have

$$Det(I_\xi(\eta_1)) \le Det(A_\xi(\eta_1)). \tag{4.1}$$

Since

$$A_\xi(\eta_1) = \begin{pmatrix} aI_{s\times s} & bJ_{s\times 1} \\ bJ_{1\times s} & \frac{1}{\beta^2}\sum_{i=1}^s \omega_{i+}c_i^2\Psi(c_i) \end{pmatrix},$$

where

$$a = \frac{\beta^2}{s}\sum_{i=1}^s \omega_{i+}\Psi(c_i) \text{ and } b = \frac{1}{s}\sum_{i=1}^s \omega_{i-}c_i\Psi(c_i), \tag{4.2}$$

it can be shown that

$$\begin{aligned} Det(A_\xi(\eta_1)) &\le \left(\sum_{i=1}^s \omega_{i+}\Psi(c_i)\right)^s \left(\sum_{i=1}^s \omega_{i+}c_i^2\Psi(c_i)\right)\frac{\beta^{2s-2}}{s^s} \\ &\le (\Psi(c))^s\left(c^2\Psi(c)\right)\frac{\beta^{2s-2}}{s^s}, \end{aligned} \tag{4.3}$$

for an appropriately chosen point c. The last inequality follows from Proposition A.2 of Yang and Stufken (2009), which asserts the existence of a point c, such that

$$\sum_{i=1}^s \omega_{i+}\Psi(c_i) = \Psi(c) \text{ and } \sum_{i=1}^s \omega_{i+}c_i^2\Psi(c_i) \le c^2\Psi(c).$$

Consider a design $\tilde{\xi} = \{(c_{i1} = c, \omega_{i1} = 1/(2s)),\ (c_{i2} = -c,\ \omega_{i2} = 1/(2s)),\ i = 1,\ldots,s\}$. It is easy to see that

$$Det(I_{\tilde{\xi}}(\eta_1)) = (\Psi(c))^s\left(c^2\Psi(c)\right)\frac{\beta^{2s-2}}{s^s}. \tag{4.4}$$

By (4.1), (4.3), and (4.4), the conclusion for case (i) follows.

For case (ii), write

$$I_\xi(\eta_2) = I_{\xi_{22}} - I_{\xi_{21}}I_{\xi_{11}}^{-1}I_{\xi_{21}}^T. \tag{4.5}$$

Here, $I_{\xi_{11}} = \beta^2\sum_{i=1}^s \omega_{i+}\Psi(c_i)$, $I_{\xi_{21}} = (\beta^2\omega_{1+}\Psi(c_1),\ldots,\beta^2\omega_{(s-1)+}\Psi(c_{s-1}), \sum_{i=1}^s \omega_{i-}c_i\Psi(c_i))^T$, and $I_{\xi_{22}}$ can be obtained from $I_\xi(\eta_1)$ by deleting the s'th row and column. Defining $B_\xi(\eta_2) = (1/(s-1)!)\sum_Q Q^T I_\xi(\eta_2)Q$, where the sum is now over all permutation matrices corresponding to permutations of the first

$s-1$ rows and columns of $I_\xi(\eta_2)$, and again applying Proposition A.2 in the Appendix, we have

$$Det(I_\xi(\eta_2)) \le Det(B_\xi(\eta_2)). \tag{4.6}$$

Writing $a_0 = (\beta^2/(s-1))\sum_{i=1}^{s-1}\omega_{i+}\Psi(c_i)$ and letting a be as defined in the first part of the proof, it can be shown that

$$\begin{aligned} Det(B_\xi(\eta_2)) &\le a_0^{s-1}(as - a_0(s-1))\frac{1}{as\beta^2}\sum_{i=1}^{s}\omega_{i+}c_i^2\Psi(c_i) \\ &\le a^s\frac{1}{as\beta^2}\sum_{i=1}^{s}\omega_{i+}c_i^2\Psi(c_i) \le (\Psi(c))^{s-1}\left(c^2\Psi(c)\right)\frac{\beta^{2s-4}}{s^s} \end{aligned} \tag{4.7}$$

for an appropriately chosen point c. The next to last inequality in (4.7) uses that the geometric mean is bounded by the arithmetic mean, while the last inequality follows again from Proposition A.2 of Yang and Stufken (2009).

It can be shown that the design $\tilde{\xi} = \{(c_{i1} = c, \omega_{i1} = 1/(2s)), (c_{i2} = -c, \omega_{i2} = 1/(2s)), i = 1, \ldots, s\}$ satisfies

$$Det(I_{\tilde{\xi}}(\eta_2)) = (\Psi(c))^{s-1}\left(c^2\Psi(c)\right)\frac{\beta^{2s-4}}{s^s}. \tag{4.8}$$

The conclusion for case (ii) follows now from (4.6), (4.7), and (4.8).

The same arguments used in the proof of Theorem 3 can also be used to derive D-optimal designs for the case of Model (2.2) and a main effects model for multiple factors. Writing α_ℓ^i for an effect corresponding to the ith level of factor ℓ, the probability in Model (2.2) is then a function of $\alpha_0 + \sum_{\ell=1}^{L}\alpha_\ell^{i_\ell} + \beta x_{ij}$ if the ith group corresponds to level combination $(i_1, \ldots, i_L)$. A maximal set of linearly independent estimable functions is $\eta_3 = (\alpha_0 + \alpha_1^1 + \alpha_2^1 + \cdots + \alpha_L^1, \alpha_1^2 - \alpha_1^1, \ldots \alpha_1^{s_1} - \alpha_1^1, \ldots, \alpha_L^2 - \alpha_L^1, \ldots, \alpha_L^{s_L} - \alpha_L^1, \beta)^T$. If we are only interested in level comparisons, we form η_4 from η_3 by deleting the first term. By the same arguments as in the proof of Theorem 3 we can obtain the following result.

Theorem 4. *Under Model (2.2) with L factors, s_ℓ levels for factor ℓ, and no constraint on the design space, designs of the form $\xi^* = \{(c_{i1} = c^*, \omega_{i1} = 1/(2k)), (c_{i2} = -c^*, \omega_{i2} = 1/(2k)), i = 1, \ldots, k\}$ are D-optimal for both η_3 and η_4. The point c^* is chosen to maximize $c^2\Psi^{k_1+2}(c)$ for η_3 and $c^2\Psi^{k_1+1}(c)$ for η_4, with $k = \prod_{\ell=1}^{L} s_\ell$ and $k_1 = \sum_{\ell=1}^{L}(s_\ell - 1)$.*

In the next sections we propose computational approaches for finding optimal designs based on Theorems 1 and 2 when explicit optimal designs are not available. Considerations in those sections are not restricted to Model (2.2) or D-optimal designs.

5. Determination of Optimal Weights for a Given Support

Based on Theorems 1 and 2, we can restrict the search for optimal designs to designs of a simple form. However, for a given model, function $F(\eta)$, and optimality criterion, we need to determine the exact support points and the corresponding optimal weights. The number of support points is (at most) $m = 2k$, which are completely determined by k variables. This implies that we also have to determine $m-1$ weights and, in total, $m+k-1$ variables in our optimization problem. When the value of k is small, this can be handled with a relatively simple computer search. However, even for moderate k, this becomes challenging. For example, suppose we have two factors, each with two levels, so $k = 4$. This already results in an optimization problem with 11 variables. Existing approaches for identifying optimal points and weights, such as Pukelsheim and Torsney (1991), do not seem to work here, and relationships between optimal design points and weights, such as in Pukelsheim (2006, p.199), do not provide much help.

We propose a new approach for the determination of optimal weights for given supports points (not necessarily the support points of an optimal design). It can be applied to any parameter function of interest. In Section 6 we show how this method, combined with the results in Section 3, can be used to identify optimal designs. Optimal weights are the solution to $m-1$ nonlinear equations and can be found numerically, where a convexity property helps with the speed of convergence. We focus on A-optimality and D-optimality but the conclusions also hold under Φ_p-optimality, where p is any positive integer

As in Section 2, let $F(\eta)$ be a vector of parameter functions of interest, where $\eta = B\theta$ and $B = DX^T$ for some matrix D. The covariance matrix of $F(\hat{\eta})$, $\Sigma_\xi(F(\hat{\eta}))$ (or Σ_ξ for simplicity), is given by (2.10). Under the A-criterion we want to minimize $Tr(\Sigma_\xi)$ for fixed design points. Let $\widetilde{B} = \frac{\partial F(\eta)}{\partial \eta^T} B$, so that $\Sigma_\xi = \widetilde{B} I_\xi^-(\theta) \widetilde{B}^T$. For given parameters, $\widetilde{B}$ is a constant matrix. Also observe that, from $B = DX^T$ and (2.7), we have

$$X^T = X^T I_\xi^-(\theta) I_\xi(\theta), \tag{5.1}$$

$$\widetilde{B} = \widetilde{B} I_\xi^-(\theta) I_\xi(\theta). \tag{5.2}$$

Theorem 5. *Let $\omega = (\omega_{11}, \omega_{12}, \ldots, \omega_{k,m_k-1})^T$, where $\omega_{ij} \geq 0$, $i = 1, \ldots, k$ and $j = 1, \ldots, m_i$, $\sum_{i=1}^{k} \sum_{j=1}^{m_i} \omega_{ij} = 1$. For a given θ and design points x_{ij}, $i = 1, \ldots, k$, and $j = 1, \ldots, m_i$, consider $\Sigma_\xi = \widetilde{B} I_\xi^-(\theta) \widetilde{B}^T$ as a function of ω. The minimum value of $Tr(\Sigma_\xi)$ is achieved at any of its critical points or at a point on the boundary.*

Proof. For simplification, we rename the sequence $(\omega_{11}, \ldots, \omega_{k,m_k})$ as $(\omega_1, \ldots, \omega_m)$ and drop θ from our notation. Let $I_\xi^i = nXV\Omega^i VX^T$, where Ω^i is a diagonal matrix with the last diagonal element equal to -1, the ith element 1, and all others 0. Other notation is as in (2.7). By Lemma 15.10.5 of Harville (1997) and (5.2), for $i = 1, \ldots, m-1$, we have

$$\frac{\partial \Sigma_\xi}{\partial \omega_i} = \widetilde{B}\frac{\partial I_\xi^-}{\partial \omega_i}\widetilde{B}^T = -\widetilde{B}I_\xi^- \frac{\partial I_\xi}{\partial \omega_i} I_\xi^- \widetilde{B}^T = -\widetilde{B}I_\xi^- I_\xi^i I_\xi^- \widetilde{B}^T. \tag{5.3}$$

Similarly, using (5.1) and (5.3), for $i, j = 1, \ldots, m-1$ we have

$$\frac{\partial^2 \Sigma_\xi}{\partial \omega_i \partial \omega_j} = \widetilde{B}I_\xi^- \left(I_\xi^j I_\xi^- I_\xi^i + I_\xi^i I_\xi^- I_\xi^j\right) I_\xi^- \widetilde{B}^T. \tag{5.4}$$

Using that

$$\frac{\partial Tr(\Sigma_\xi)}{\partial \omega_i} = Tr\left(\frac{\partial \Sigma_\xi}{\partial \omega_i}\right), \tag{5.5}$$

for $i, j = 1, \ldots, m-1$, we have

$$\frac{\partial^2 Tr(\Sigma_\xi)}{\partial \omega_i \partial \omega_j} = Tr\left(\frac{\partial^2 \Sigma_\xi}{\partial \omega_i \partial \omega_j}\right). \tag{5.6}$$

Let $H(\omega)$ be the Hessian matrix of $Tr(\Sigma_\xi)$. We show that $H(\omega)$ is a nonnegative definite matrix. Since I_ξ is nonnegative definite, there exists a g-inverse I_ξ^-, which is also nonnegative definite. This g-inverse I_ξ^- can be written as $I_\xi^- = (I_\xi^-)^{1/2}(I_\xi^-)^{1/2}$, where $(I_\xi^-)^{1/2}$ is also nonnegative definite.

By (5.4), the (i, j)th element of $H(\omega)$ is

$$\begin{aligned} H(\omega)[i,j] &= Tr\left(\widetilde{B}I_\xi^- \left(I_\xi^j I_\xi^- I_\xi^i + I_\xi^i I_\xi^- I_\xi^j\right) I_\xi^- \widetilde{B}^T\right) \\ &= 2Tr\left(\widetilde{B}I_\xi^- I_\xi^i I_\xi^- I_\xi^j I_\xi^- \widetilde{B}^T\right) \\ &= 2Tr\left(\widetilde{B}I_\xi^- I_\xi^i (I_\xi^-)^{1/2}(I_\xi^-)^{1/2} I_\xi^j I_\xi^- \widetilde{B}^T\right). \end{aligned} \tag{5.7}$$

By defining $A_i = \widetilde{B}I_\xi^- I_\xi^i (I_\xi^-)^{1/2}$ and applying Proposition A.1 in the Appendix, it follows that $H(\omega)$ is nonnegative definite. From this it follows that $Tr(\Sigma_\xi)$ attains its minimum at any of the critical points (cf. Kaplan (1999, Sec. 9)). It is however possible that none of the critical points satisfy the restriction that each of the ω_{ij} must be nonnegative. In this case, the minimum value is attained on the boundary, i.e., at least one of $\omega_{ij} = 0$.

Consideration of D-optimality only makes sense when Σ_ξ is nonsingular, so we assume that. A D-optimal design maximizes $|\Sigma_\xi^{-1}|$. We have the following counterpart of Theorem 5.

Theorem 6. *Let* $\omega = (\omega_{11}, \omega_{12}, \ldots, \omega_{k,m_k-1})^T$*, where* $\omega_{ij} \geq 0$*,* $i = 1, \ldots, k$ *and* $j = 1, \ldots, m_i$*,* $\sum_{i=1}^{k} \sum_{j=1}^{m_i} \omega_{ij} = 1$*. For a given* θ *and design points* $x_{ij}, i = 1, \ldots, k$*, and* $j = 1, \ldots, m_i$*, consider* $\Sigma_\xi = \widetilde{B} I_\xi^-(\theta) \widetilde{B}^T$ *as a function of* ω*. The maximum value of* $|\Sigma_\xi^{-1}|$ *is achieved at any of its critical points or at a point on the boundary.*

Proof. Maximizing $|\Sigma_\xi^{-1}|$ is equivalent to minimizing $-\log|\Sigma_\xi^{-1}|$. Using Harville (1997, p.309), we have, for $i, j = 1, \ldots, m-1$,

$$\frac{\partial^2 \left(-\log|\Sigma_\xi^{-1}|\right)}{\partial\omega_i \partial\omega_j} = \frac{\partial^2 \left(\log|\Sigma_\xi|\right)}{\partial\omega_i \partial\omega_j} = Tr\left(\Sigma_\xi^{-1} \frac{\partial^2 \Sigma_\xi}{\partial\omega_i \partial\omega_j} - \Sigma_\xi^{-1} \frac{\partial \Sigma_\xi}{\partial\omega_i} \Sigma_\xi^{-1} \frac{\partial \Sigma_\xi}{\partial\omega_j}\right). \tag{5.8}$$

With $H(\omega)$ denoting the Hessian matrix of $-\log|\Sigma_\xi^{-1}|$, the (i,j)th element of $H(\omega)$ is given by (5.8). By the same argument as in the proof of Theorem 5, it is sufficient to show that $H(\omega)$ is a nonnegative definite matrix. As there, let $I_\xi^- = (I_\xi^-)^{1/2}(I_\xi^-)^{1/2}$, where $(I_\xi^-)^{1/2}$ is nonnegative definite. By (5.4) and a similar argument as for (5.7), we have

$$Tr\left(\Sigma_\xi^{-1} \frac{\partial^2 \Sigma_\xi}{\partial\omega_i \partial\omega_j}\right) = 2Tr\left(\Sigma_\xi^{-1/2} \widetilde{B} I_\xi^- I_\xi^i (I_\xi^-)^{1/2} (I_\xi^-)^{1/2} I_\xi^j I_\xi^- \widetilde{B}^T \Sigma_\xi^{-1/2}\right). \tag{5.9}$$

Using (5.3), for the second term in (5.8) we also have

$$Tr\left(\Sigma_\xi^{-1} \frac{\partial \Sigma_\xi}{\partial\omega_i} \Sigma_\xi^{-1} \frac{\partial \Sigma_\xi}{\partial\omega_j}\right) = Tr\left(\Sigma_\xi^{-1/2} \widetilde{B} I_\xi^- I_\xi^i I_\xi^- \widetilde{B}^T \Sigma_\xi^{-1} \widetilde{B} I_\xi^- I_\xi^j I_\xi^- \widetilde{B}^T \Sigma_\xi^{-1/2}\right). \tag{5.10}$$

Thus, from (5.8), (5.9), and (5.10), $H(\omega)$ can be written as

$$H(\omega) = H_0(\omega) + H_1(\omega), \tag{5.11}$$

where the (i,j)th element of $H_0(\omega)$ is

$$H_0(\omega)[i,j] = Tr\left(\Sigma_\xi^{-1/2} \widetilde{B} I_\xi^- I_\xi^i (I_\xi^-)^{1/2} (I_\xi^-)^{1/2} I_\xi^j I_\xi^- \widetilde{B}^T \Sigma_\xi^{-1/2}\right) \tag{5.12}$$

and the (i,j) element of $H_1(\omega)$ is

$$H_1(\omega)[i,j] = Tr\left(\Sigma_\xi^{-1/2} \widetilde{B} I_\xi^- I_\xi^i (I_\xi^-)^{1/2} P^\perp\left[(I_\xi^-)^{1/2} \widetilde{B}^T\right] (I_\xi^-)^{1/2} I_\xi^j I_\xi^- \widetilde{B}^T \Sigma_\xi^{-1/2}\right). \tag{5.13}$$

Here $P^\perp\left[(I_\xi^-)^{1/2} \widetilde{B}^T\right]$ denotes the orthogonal projection matrix onto the orthogonal complement of the column space of $(I_\xi^-)^{1/2} \widetilde{B}^T$. That $H_0(\omega)$ is nonnegative definite follows now from Proposition A.1 in the Appendix by taking $A_i = \Sigma_\xi^{-1/2} \widetilde{B} I_\xi^- I_\xi^i (I_\xi^-)^{1/2}$. That $H_1(\omega)$ is nonnegative definite follows by taking

$A_i = \Sigma_\xi^{-1/2}\widetilde{B}I_\xi^- I_\xi^i (I_\xi^-)^{1/2}P^\perp\left[(I_\xi^-)^{1/2}\widetilde{B}^T\right]$ in that same proposition. This implies that $H(\omega)$ is also nonnegative definite.

By Theorems 5 and 6, the optimal weights can be derived by solving $m-1$ equations to find the critical points. These $m-1$ equations are nonlinear in ω. Since there is in general no closed form solution, we have to rely on a numerical approach such as Newton's method. Since the Hessian matrices are nonnegative definite, the convergence can be very fast (see Deuflhard (2004)). However, it can also result in negative weights. If, during the implementation of the algorithm, a step takes us into a region where some weights are negative, the algorithm tries to reduce the size of this step to assure all weights are positive. If this makes the step too small, then a weight of zero is assigned to the weight variable that first attains the value of zero along the path selected by the algorithm. This reduces the number of support points, and we can now apply Theorems 5 or 6 again for this reduced set of support points and search for the optimal weights on that set. This process is repeated until we find weights that satisfy the constraints. We have developed an efficient algorithm that accomplishes this. While there is no guarantee that this algorithm always finds the optimal weights, we illustrate in the next section that, when searching for an optimal design, there is a numerical check as to whether the answer obtained from the algorithm corresponds indeed to an optimal design. In our experience, it almost always does.

6. Examples

Based on the structure for optimal designs in Theorems 1 and 2, and the suggested approach for finding optimal weights based on Theorems 5 and 6, we can contemplate a complete grid search to identify optimal designs. Conceptually, for a design of the form in Theorems 1 or 2, we can use a sufficiently fine grid to consider possible values for the unknown design points and, for each of these, find the optimal weights based on the results in Section 5. However, with multiple groups, this is practically unfeasible except for the simplest problems. For example, even with four groups, besides the optimal weights we would need to determine four unknown design points. A complete grid search for this problem that aims to find optimal support points that are accurate to two decimal places seems excessive.

Instead we use a multi-stage grid search that starts with a coarse grid that is made increasingly finer in later stages. At each stage we identify the best design based on the grid at that stage; for the next stage, a finer grid is, in each group, restricted to neighborhoods of the best support points found at the current stage, still using the structure of the optimal designs in Theorems 1 and 2. The search continues until a specified accuracy for the design points is reached.

While this strategy can greatly reduce computing time, there is no guarantee that the resulting design is indeed optimal. Fortunately, there is a powerful tool to verify whether a design is optimal or not, namely the General Equivalence Theorem (Pukelsheim (2006)). To formulate this tool, we focus on the common slope models, Models (2.2) and (2.4). Straightforward changes can be made for Models (2.3) and (2.5). With $X^i = (1, X_i^T, x)^T$, a design ξ is locally optimal for $\theta = \theta_0$ if there exists a generalized inverse of I_ξ, say G, such that for each i and all x,

$$(X^i)^T G^T \widetilde{B}^T (\widetilde{B} I_\xi^-(\theta_0) \widetilde{B}^T)^{-(p+1)} \widetilde{B} G X^i \leq Tr\left((\widetilde{B} I_\xi^-(\theta_0) \widetilde{B}^T)^{-p} \right),$$

with equality when x is one of the support points of ξ. Here $p = 0$ corresponds to D-optimality and $p = -1$ to A-optimality. Note that this property does not need to hold for every generalized inverse of I_ξ. However, in our experience the Moore-Penrose inverse always works. In our numerical studies, we were able to verify every optimality result in this way. The algorithm can be summarized as follows:

(i) Start with a coarse grid within each of the groups.

(ii) Find optimal weights for each possible design of the form in Theorem 1 or 2 that chooses its support from the grid points, and identify the best design of this type.

(iii) Build a finer grid around the support points of this best design, find optimal weights for each possible design of the form in Theorem 1 or 2, now choosing the support points from the finer grid, and identify the best design of this type.

(iv) Repeat (iii) until no further improvement can be made (up to a specified accuracy).

(v) Verify that the final design is optimal using the General Equivalence Theorem.

Example 1. Consider the logistic model for two factors, each with two levels,

$$\text{logit}\,(Prob(Y_{ij} = 1)) = \alpha_0 + \alpha_i + \beta x_{ij},\ \ i = 1, \ldots, 4. \tag{6.1}$$

Here, the α_i's, $i = 1, \ldots, 4$, represent the group effects of the groups $(1,1)$, $(1,2)$, $(2,1)$, and $(2,2)$, respectively. We assume that there is no restriction on the x_{ij}'s, and consider two cases: (1) The full model with no further assumptions about the α_i's; and (2) the main-effects model with $\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 0$. For the interaction model we use the parameter vector $\eta = ((\alpha_1 + \alpha_2 - \alpha_3 - \alpha_4)/2, (\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4)/2, (\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4)/2, \beta)^T$, while the main-effects model corresponds to

Table 1. Support Points and Weights for Locally Optimal Designs

	Main-effects model		Interaction model	
	A-optimal	D-optimal*	A-optimal	D-optimal*
group	(1.8284, 0.1253)	2.2229	(1.7539, 0.1253)	2.0436
(1,1)	(0.1716, 0.1253)	-0.2229	(0.2461, 0.1253)	-0.0436
group	(1.5784, 0.1521)	1.9729	(1.5039, 0.1532)	1.7936
(1,2)	(-0.0784, 0.0974)	-0.4729	(-0.0039, 0.0963)	-0.2936
group	(2.0784, 0.0974)	2.4729	(2.0039, 0.0963)	2.2936
(2,1)	(0.4216, 0.1521)	0.0271	(0.4961, 0.1532)	0.2064
group	(1.8284, 0.1253)	2.2229	(1.7539, 0.1253)	2.0436
(2,2)	(0.1716, 0.1253)	-0.2229	(0.2461, 0.1253)	-0.0436

* For the D-optimal designs, all support points have weight 1/8.

Table 2. The value of c^* that maximizes $c^2\Psi^p(c)$.

	c^*			c^*	
p	Logistic	Probit	p	Logistic	Probit
1	2.3994	1.5750	6	0.8399	0.6696
2	1.5434	1.1381	7	0.7744	0.6209
3	1.2229	0.9376	8	0.7222	0.5815
4	1.0436	0.8159	9	0.6793	0.5487
5	0.9254	0.7320	10	0.6432	0.5209

$\eta = ((\alpha_1+\alpha_2-\alpha_3-\alpha_4)/2, (\alpha_1-\alpha_2+\alpha_3-\alpha_4)/2, \beta)^T$. For both cases we run the algorithm to search for locally A- and D-optimal designs with local conditions given by $\alpha_0 = -1$, $(\alpha_1, \ldots, \alpha_4) = (0, 0.25, -0.25, 0)$, and $\beta = 1$. The A- and D-optimal designs found by the algorithm are shown in Table 1. For A-optimal designs the support points are followed by the corresponding weights, while only the support points are presented for D-optimal designs since all the weights are 1/8. These designs are not unique, and optimal designs with fewer support points may be found. The designs in the table do have a lot of structure. This is not surprising in view of Theorem 1, but is easier seen in terms of the c_{ij}'s than in terms of the reported x_{ij}'s. For example, the A-optimal design in Table 1 for the main-effects model has support points $(c_{i1}, c_{i2}) = (0.8284, -0.8284)$ for each of the four groups.

Finding one of these optimal designs takes about 2 seconds of CPU time on a desktop PC with a 3.2GHz Intel Pentium processor.

The D-optimal designs agree with those derived in Theorem 3 (for the interaction model) and Theorem 4 (for the main-effects model). To see this, in Table 2 we present the point c^* that maximizes $c^2\Psi^p(c)$ for the logistic model (and also for the probit model) and small values of p.

For the main-effects model here, Theorem 4 gives $c^* = 1.2229$ (corresponding to $p = 3$). These c_{ij} values correspond, for the given $\theta = \theta_0$, exactly to the x_{ij} values for the D-optimal design for the main-effects model in Table 1. Similarly, using the equivalence of the full interaction model with the model for a single factor at 3 levels, Theorem 3 asserts that the optimal c^* for the interaction model corresponds to $p = 4$, $c^* = 1.0436$. This also corresponds exactly to the x_{ij}'s for the D-optimal design presented in Table 1 for that case.

While the designs in Table 1 are locally optimal if our specification of θ_0 is correct, it is useful to know how efficient they are if θ_0 were misspecified. Following Dror and Steinberg (2006) and Woods et al. (2006), we study the robustness of the optimal designs by randomly drawing other possible true values for θ. If our choice of θ_0 was based on previous information, then it may not be unreasonable to draw θ from a distribution with mean θ_0. Here we took $\theta \sim N(\theta_0, \sigma^2 I_{6\times 6})$. If θ_1 is the value drawn from this distribution for a selected value of σ^2, then we first derived a locally optimal design for θ_1. If ξ_0 and ξ_1 denote locally optimal designs for θ_0 and θ_1, respectively, then we computed the efficiency of ξ_0 for the case that θ_1 is the true value, $eff_{\xi_0}(\theta_1)$, as

$$
\begin{aligned}
&\frac{|(\widetilde{B} I^-_{\xi_0}(\theta_1)\widetilde{B}^T)^{-1}|^{1/r}}{|(\widetilde{B} I^-_{\xi_1}(\theta_1)\widetilde{B}^T)^{-1}|^{1/r}} \text{ under } D\text{-optimality, and} \\
&\frac{Tr(\widetilde{B} I^-_{\xi_1}(\theta_1)\widetilde{B}^T)}{Tr(\widetilde{B} I^-_{\xi_0}(\theta_1)\widetilde{B}^T)} \text{ under } A\text{-optimality.}
\end{aligned}
\tag{6.2}
$$

Here, r is the rank of $\widetilde{B}$. The values that we took for σ were 0.4, 0.2, and 0.1. For each scenario, we drew 1,000 random θ_1 values, leading to 1,000 measurements of the efficiency of ξ_0. Summary statistics for the efficiencies are reported in Table 3. In this table, ξ_{0A1} and ξ_{0A2} denote the locally optimal designs for θ_0 under the A-optimality criterion for the main-effects model and the interaction model, respectively. Similarly, ξ_{0D1} and ξ_{0D2} denote the D-optimal designs.

The results show how the performance of the locally optimal designs can degrade with increased uncertainty about the value of θ_0. It is also worth observing that the A-optimal designs are not D-optimal, and vice versa. More precisely, with $\theta_0 = (-1, 0, 0.25, -0.25, 0, 1)^T$, under the D-optimality criterion, the efficiency of ξ_{0A1} for the main-effects model is 0.921 and that of ξ_{0A2} for the interaction model is 0.954. Conversely, for the A-optimality criterion, ξ_{0D1} has an efficiency of 0.902 for the main-effects model and ξ_{0D2} has an efficiency of 0.939 for the interaction model.

Example 2. The proposed method also works for a larger number of groups. To demonstrate this, consider again an example with two factors, but this time both

Table 3. Efficiencies of the locally optimal designs.

Design	σ	Mean	Std Dev	Minimum	Maximum
ξ_{0A1}	0.4	0.9196	0.0903	0.2703	0.9986
	0.2	0.9765	0.0284	0.7559	0.9998
	0.1	0.9942	0.0066	0.9436	0.9999
ξ_{0D1}	0.4	0.8579	0.1459	0.0244	0.9988
	0.2	0.9628	0.0413	0.6747	0.9995
	0.1	0.9902	0.0116	0.8921	0.9999
ξ_{0A2}	0.4	0.9240	0.0786	0.4672	0.9994
	0.2	0.9783	0.0236	0.8268	0.9997
	0.1	0.9945	0.0006	0.9505	0.9998
ξ_{0D2}	0.4	0.8779	0.1291	0.0742	0.9975
	0.2	0.9668	0.0344	0.7122	0.9988
	0.1	0.9914	0.0088	0.9286	0.9999

with four levels. The form of the model is as in (6.1), but we assume that the two-factor interaction is negligible. We took $\theta_0 = (-1, 0.05, 0, -0.05, 0.1, 0.15, 0.1, 0.05, 0.2, 0.25, 0.2, 0.15, 0.3, -0.05, -0.1, -0.15, 0, 1)^T$, where the first entry is for α_0, the next four for the groups (1,1), (1,2), (1,3) and (1,4), and so on. There was no restriction on the covariate value x. For η we used three main-effect contrasts for the first factor and three for the second, and the slope parameter β. For the contrasts in both cases we took the orthonormal contrasts with coefficients $(-3, -1, 1, 3)/\sqrt{20}$, $(1, -1, -1, 1)/2$, and $(-1, 3, -3, 1)/\sqrt{20}$. The 32 support points and weights for A- and D-optimal designs are given in Table 4.

The D-optimal design corresponds exactly to the design given by Theorem 4 and Table 2 using $p = k_1 + 1 = 7$.

Under the D-optimality criterion, the A-optimal design has an efficiency of .988, while the D-optimal design has an efficiency of .982 under the A-optimality criterion. To find these optimal designs took about 70 seconds of CPU time on a desktop PC with a 3.2GHz Intel Pentium processor.

7. Discussion

This paper provides theoretical results for optimal designs under various models that include group effects. Theorems 1 and 2 extend the results in Yang and Stufken (2009).

The results also hold for multi-stage approaches, by which we mean that if we add design points to an existing initial design, we can do so in an optimal way by restricting attention to designs described in the two theorems, no matter

Table 4. Support points and weights for locally optimal designs.

Group	A-optimal	D-optimal*
(1,1)	(1.5843, 0.0184); (0.3157, 0.0442)	1.7244; 0.1756
(1,2)	(1.6343, 0.0191); (0.3657, 0.0434)	1.7744; 0.2256
(1,3)	(1.6843, 0.0363); (0.4157, 0.0262)	1.8244; 0.2756
(1,4)	(1.5343, 0.0453); (0.2657, 0.0173)	1.6744; 0.1256
(2,1)	(1.4843, 0.0529); (0.2157, 0.0097)	1.6244; 0.0756
(2,2)	(1.5343, 0.0263); (0.2657, 0.0363)	1.6744; 0.1256
(2,3)	(1.5843, 0.0222); (0.3157, 0.0405)	1.7244; 0.1756
(2,4)	(1.4343, 0.0298); (0.1657, 0.0327)	1.5744; 0.0256
(3,1)	(1.3843, 0.0417); (0.1157, 0.0207)	1.5244; -0.0244
(3,2)	(1.4343, 0.0472); (0.1657, 0.0153)	1.5744; 0.0256
(3,3)	(1.4843, 0.0281); (0.2157, 0.0345)	1.6244; 0.0756
(3,4)	(1.3343, 0.0259); (0.0657, 0.0363)	1.4744; -0.0744
(4,1)	(1.6843, 0.0150); (0.4157, 0.0475)	1.8244; 0.2756
(4,2)	(1.7343, 0.0294); (0.4657, 0.0329)	1.8744; 0.3256
(4,3)	(1.7843, 0.0294); (0.5157, 0.0328)	1.9244; 0.3756
(4,4)	(1.6343, 0.0329); (0.3657, 0.0297)	1.7744; 0.2256

* For the D-optimal design, all support points have weight 1/32.

what the initial design is. The reason that this holds is the one in Yang and Stufken (2009). This is important, because in a multi-stage approach the first stage may give us information about the unknown parameters that can then be used in the local optimality approach for adding additional design points at the second stage.

Whether in a multi-stage or single-stage approach, the designs that are obtained are often large if there are many groups; with k groups, as many as $2k$ support points. This may be unavoidable, especially in a single-stage approach. For example, for Model (2.3) there are potentially $2k$ independent estimable functions, so that $2k$ is the minimum number of support points needed to enable unbiased estimation of all of these functions. For the single-slope models or for models with additional assumptions (for example, a main-effects model or a model with main-effects and two-factor interactions) we may hope to get by with fewer support points.

For the special case of Model (2.2) and D-optimality, we used Theorems 1 and 2 to derive explicit solutions for optimal designs in Section 4. For other cases, while Theorems 1 and 2 make finding optimal designs much easier, this can be a formidable problem for larger k. While there is no theoretical guarantee that our algorithm works, empirical evidence for it is very good. Moreover, as described in Section 6, the General Equivalence Theorem allows one to check whether a design found by the algorithm is optimal. Generally, optimal designs of the forms described in Theorems 1 and 2 are not unique. Depending on the model and on the vector η of interest, our algorithm may find optimal designs that are supported on less than $2k$ points, but in general does not find designs with the smallest possible support size. While the algorithm can handle fairly large cases, there is a need for an algorithms that handles even larger cases.

Another feature that we observed is that, in terms of the c_{ij}'s, the design points for an optimal design are often (but not always) the same in each of the groups. It is an interesting open question to identify conditions that allow an optimal design of the forms described in Theorems 1 and 2 but with the same c_{ij}'s in each of the groups.

Acknowledgement

Research was partially supported by NSF grants DMS-0706917 and DMS-1007507 (for JS); and by NSF grants DMS-0707013 and DMS-0748409 (for MY).

Appendix

Proposition A.1. Let A_i, $i = 1, \ldots, n$, be $p \times q$ matrices. The $n \times n$ matrix M, with element $M[i, j] = Tr(A_i A_j^T)$ in position (i, j), is nonnegative definite.

Proof. Consider the matrix $A = (A_1^T, A_2^T, \ldots, A_n^T)^T$. It is clear that the $np \times np$ matrix AA^T is nonnegative definite. AA^T can be written as

$$AA^T = \begin{pmatrix} A_1A_1^T & A_1A_2^T & \ldots & A_1A_n^T \\ A_2A_1^T & A_2A_2^T & \ldots & A_2A_n^T \\ \vdots & \vdots & \ddots & \vdots \\ A_nA_1^T & A_nA_2^T & \ldots & A_nA_n^T \end{pmatrix}. \tag{A.1}$$

Let J_i, $i = 1, \ldots, p$, be a $1 \times p$ vector with the the ith element 1 and all others 0. Define the $n \times np$ matrix B_i as

$$B_i = \begin{pmatrix} J_i & 0_{1\times p} & \ldots & 0_{1\times p} \\ 0_{1\times p} & J_i & \ldots & 0_{1\times p} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{1\times p} & 0_{1\times p} & \ldots & J_i \end{pmatrix}. \tag{A.2}$$

It is obvious that the $n \times n$ matrix $B_iAA^TB_i^T$ is nonnegative definite. Its (k, l)th element is given by $J_iA_kA_l^TJ_i^T$, which is the ith diagonal element of $A_kA_l^T$. Thus, we have

$$M = \sum_{i=1}^{p} B_iAA^TB_i^T. \tag{A.3}$$

By the fact that $B_iAA^TB_i^T$, $i = 1, \ldots, p$, is nonnegative definite, the conclusion follows.

Proposition A.2. For $t \times t$ positive definite matrices Q_i and positive numbers λ_i, $i = 1, \ldots, n$, with $\sum_{i=1}^{n} \lambda_i = 1$, we have

$$\log Det \left(\sum_{i=1}^{n} \lambda_i Q_i \right) \geq \sum_{i=1}^{n} \lambda_i \log Det(Q_i). \tag{A.4}$$

Equality holds only when all Q_i's are the same.

Proof. It suffices to prove (A.4) for $n = 2$. Since Q_1 is positive definite, it is enough to prove that, for $0 < \lambda < 1$,

$$\begin{aligned} &\log Det \left(\lambda I + (1-\lambda) Q_1^{-1/2} Q_2 Q_1^{-1/2} \right) \\ &\quad \geq \lambda \log Det(I) + (1-\lambda) \log Det(Q_1^{-1/2} Q_2 Q_1^{-1/2}). \end{aligned} \tag{A.5}$$

Since $Q_1^{-1/2} Q_2 Q_1^{-1/2}$ is a positive definite matrix, there exists an orthonormal matrix P, such that

$$P Q_1^{-1/2} Q_2 Q_1^{-1/2} P^T = diag(\mu_1, \ldots, \mu_t), \tag{A.6}$$

where $\mu_i > 0, i = 1, \ldots, t$ are the eigenvalues of $Q_1^{-1/2} Q_2 Q_1^{-1/2}$. By (A.6), a basic property of orthonormal matrices, and the fact that $-\log(x)$ is strictly convex, we have

$$\begin{aligned} &\log Det \left(\lambda I + (1-\lambda) Q_1^{-1/2} Q_2 Q_1^{-1/2} \right) \\ &\quad = \sum_{i=1}^{t} \log \left(\lambda + (1-\lambda) \mu_i \right) \\ &\quad \geq (1-\lambda) \sum_{i=1}^{t} \log \mu_i = (1-\lambda) \log Det(Q_1^{-1/2} Q_2 Q_1^{-1/2}). \end{aligned} \tag{A.7}$$

Moreover, equality in (A.7) holds only when $\mu_i = 1$, $i = 1, \ldots, t$, which implies that $Q_1 = Q_2$ by (A.6). This completes the proof.

References

Agresti, A. (2002). *Categorical Data Analysis*. 2nd edition. Wiley, New York.

Biedermann, S., Dette, H. and Zhu, W. (2006). Geometric construction of optimal designs for dose-reponse models with two parameters. *J. Amer. Statist. Assoc.* **101**, 747-759.

Cook, R. D. and Thibodeau, L. A. (1980). Marginally restricted D-optimal designs. *J. Amer. Statist. Assoc.* **75**, 366-371.

Dette, H. (1997). Designing experiments with respect to 'Standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B*, **59**, 97-110.

Deuflhard, P. (2004). *Newton Methods for Nonlinear Problems*, Springer-Verlag, Berlin.

Dror, H. and Steinberg, D. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520-529.

Ford, I., Torsney, B. and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Statist. Soc. Ser. B* **54**, 569-583.

Harville, D. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag, New York.

Kaplan, W. (1999). *Maxima and Minima with Applications*, Wiley, New York.

Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statist. Sci.* **21**, 376-399.

Pukelsheim, F. (2006). *Optimal Design of Experiments.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Pukelsheim, F., and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Ann. Statist.* **19**, 1614-1625.

Tighiouart, M. and Rogatko, A. (2006). Dose Escalation with Overdose Control. In *Statistical Methods for Dose-Finding Experiments* (Edited by Chevret Sylvie), 173-188, Wiley, New York.

Woods, D., Lewis, S., Eccleston, J. and Russell, K. (2006). Designs for Generalized Linear Models with Several Variables and Model Uncertainty. *Technometrics* **48**, 284-292.

Wu, C. F. (1980). On some ordering properties of the generalized inverse of nonnegative definite matrices. *Linear Algebra Appl.* **32**, 49-60.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37**, 518-541.

Department of Statistics, University of Georgia, Athens, GA 30602-7952, USA.

E-mail: jstufken@uga.edu

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 322 Science and Engineering Offices (M/C 249), 851 S. Morgan Street, Chicago, IL 60607-7045, USA.

E-mail: minyang_stat@yahoo.com

(Received November 2010; accepted August 2011)