The Annals of Statistics
2015, Vol. 43, No. 1, 30–56
DOI: 10.1214/14-AOS1263

SATURATED LOCALLY OPTIMAL DESIGNS UNDER DIFFERENTIABLE OPTIMALITY CRITERIA

BY LINWEI HU[1], MIN YANG[2] AND JOHN STUFKEN[3]

University of Georgia, University of Illinois at Chicago and Arizona State University

We develop general theory for finding locally optimal designs in a class of single-covariate models under any differentiable optimality criterion. Yang and Stufken [*Ann. Statist.* **40** (2012) 1665–1681] and Dette and Schorning [*Ann. Statist.* **41** (2013) 1260–1267] gave complete class results for optimal designs under such models. Based on their results, saturated optimal designs exist; however, how to find such designs has not been addressed. We develop tools to find saturated optimal designs, and also prove their uniqueness under mild conditions.

1. Introduction. We consider the problem of finding locally optimal designs for a class of single-covariate models under differentiable optimality criteria. In order to avoid intricacies caused by the discreteness of the problem, we will work with *approximate* designs (see Section 2). Because the information matrix usually depends on the unknown parameters, we consider *locally* optimal designs by plugging in values for the parameters in the information matrix. This gives good designs when prior knowledge of the parameters is available, and it also provides a benchmark for evaluating other designs. For the sake of simplicity, we omit the word locally hereafter.

We provide general theoretical results that help to find saturated optimal designs for many of the models for which previous results, such as in Yang and Stufken (2012) and Dette and Schorning (2013), have established so-called *complete class results*. While efficient numerical algorithms, even without using the complete class results, can be developed to approximate optimal designs, theory provides unified results, both with respect to models and optimality criteria, and offers insights that cannot be obtained from algorithms. In some instances the theory enables us to find closed-form optimal designs; moreover, it can be used to develop faster and better algorithms. For example, because of the theory we can avoid having to discretize the design space. We also use the theory to develop

Received March 2014.
[1]Supported in part by NSF Grant DMS-10-07507.
[2]Supported in part by NSF Grants DMS-13-22797 and DMS-14-07518.
[3]Supported in part by NSF Grants DMS-10-07507 and DMS-14-06760.
MSC2010 subject classifications. Primary 62K05; secondary 62J02.
Key words and phrases. Chebyshev system, complete class, generalized linear model, locally optimal design, nonlinear model.

uniqueness results under mild conditions, which cannot be obtained from an algorithm approach.

Our work is based on the complete class results given in a series of papers, including most recently Yang and Stufken (2012) and Dette and Schorning (2013). Based on their results, optimal designs can be found in a small class of designs called the complete class, and in many cases, this complete class only contains designs with at most d design points, where d is the number of parameters. However, theory and tools to identify optimal designs for multiple optimality criteria within the complete class have not been developed. So in Section 2, we will present theorems to find optimal designs in these classes and prove their uniqueness. Section 3 applies the theorems to a variety of different models including polynomial regression models, nonlinear regression models and generalized linear models. The computational benefits will be shown in Section 4. Finally, Section 5 gives a short discussion about limitations of the approach. The technical proofs have been relegated to the Appendix.

2. Locally optimal design. The models under consideration include polynomial regression models, nonlinear regression models and generalized linear models, with a univariate response y and a single covariate x which belongs to the design space $[L, U]$ (L or U could be $-\infty$ or ∞, resp., with $[L, U]$ being half open or open). The unknown parameter is a $d \times 1$ vector denoted as $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$. To be specific, for polynomial regression models and nonlinear models $\boldsymbol{\theta}$ is the unknown parameter in the mean response $\eta(x, \boldsymbol{\theta}) = \mathrm{E}(y)$. We assume the variance to be constant unless otherwise specified, and take its value to be 1 since it does not affect the optimal design. For generalized linear models, $\boldsymbol{\theta}$ is the unknown parameter in the linear predictor $\eta(x, \boldsymbol{\theta}) = h(\mathrm{E}(y))$, where h is the link function.

In approximate design context, a design ξ with at most q design points can be written as $\xi = \{(x_i, \omega_i)\}_{i=1}^q$, where $x_i \in [L, U]$, $\omega_i \geq 0$, $i = 1, \dots, q$, x_i's and ω_i's are the design points and corresponding design weights, and $\sum_{i=1}^q \omega_i = 1$. If the weight of a certain design point is positive, then that design point is a *support point* of the design, and the number of support points is the *support size* of the design.

Under the assumption of independent responses, the Fisher information matrix for $\boldsymbol{\theta}$ under design ξ can be written as $n \sum_{i=1}^q \omega_i \mathbf{M}_{x_i}(\boldsymbol{\theta})$, where n is the total sample size and $\mathbf{M}_{x_i}(\boldsymbol{\theta})$ is the information matrix of a single observation at x_i. Since n is only a multiplicative factor, we prefer using the normalized information matrix, which is $\mathbf{M}_\xi(\boldsymbol{\theta}) = \sum_{i=1}^q \omega_i \mathbf{M}_{x_i}(\boldsymbol{\theta})$.

An optimal design is a design that maximizes the Fisher information matrix $\mathbf{M}_\xi(\boldsymbol{\theta})$ under a certain criterion Φ. In this paper, we focus on a general class of differentiable optimality criteria. Specifically, let NND(d) be the set of all $d \times d$ nonnegative definite matrices, PD(d) be the set of all $d \times d$ positive definite matrices, and Φ be any function defined on NND(d) that satisfies Assumption A below [see Pukelsheim (1993), page 115].

ASSUMPTION A. Suppose the optimality criterion Φ is a nonnegative, nonconstant function defined on NND(d) such that:

(1) it is concave, that is, $\Phi(\alpha\mathbf{M}_1 + (1-\alpha)\mathbf{M}_2) \geq \alpha\Phi(\mathbf{M}_1) + (1-\alpha)\Phi(\mathbf{M}_2)$, where $\alpha \in (0,1)$, $\mathbf{M}_1, \mathbf{M}_2 \in \mathrm{NND}(d)$;
(2) it is isotonic, that is, $\Phi(\mathbf{M}_1) \geq \Phi(\mathbf{M}_2)$ if $\mathbf{M}_1 \geq \mathbf{M}_2$ under the Loewner ordering, $\mathbf{M}_1, \mathbf{M}_2 \in \mathrm{NND}(d)$;
(3) it is smooth on PD(d). By smooth, we mean the function is differentiable and the first-order partial derivatives are continuous [for matrix differentiation, Φ is to be interpreted as a function of the $d(d+1)/2$-dimensional vector of elements in the upper triangle of $\mathbf{M}$].

A design ξ^* is Φ-optimal if it maximizes $\Phi(\mathbf{M}_\xi(\boldsymbol{\theta}))$ with respect to ξ.

This class of optimality criteria is very broad and includes, for example, the well-known Φ_p-optimality criteria with $-\infty < p \leq 1$, which are defined as follows. Suppose we are interested in estimating a smooth function of $\boldsymbol{\theta}$, say $g(\boldsymbol{\theta}) : \mathbb{R}^d \to \mathbb{R}^v$, where $v \leq d$ and $\mathbf{K}(\boldsymbol{\theta}) = (\partial g(\boldsymbol{\theta})/\partial\boldsymbol{\theta})^T$ has full column rank v. It can be estimated as long as the columns of $\mathbf{K}(\boldsymbol{\theta})$ are contained in the range of $\mathbf{M}_\xi(\boldsymbol{\theta})$. The information matrix for $g(\boldsymbol{\theta})$ under design ξ is then defined as $\mathbf{I}_\xi(\boldsymbol{\theta}) = (\mathbf{K}(\boldsymbol{\theta})^T\mathbf{M}_\xi(\boldsymbol{\theta})^-\mathbf{K}(\boldsymbol{\theta}))^{-1}$, where $\mathbf{M}_\xi(\boldsymbol{\theta})^-$ is a generalized inverse if $\mathbf{M}_\xi(\boldsymbol{\theta})$ is singular. Then a Φ_p-optimal design for $g(\boldsymbol{\theta})$ is defined to maximize

$$\Phi\bigl(\mathbf{M}_\xi(\boldsymbol{\theta})\bigr) = \Phi_p\bigl(\mathbf{I}_\xi(\boldsymbol{\theta})\bigr) = \left(\frac{1}{v}\operatorname{trace}\bigl(\mathbf{I}_\xi^p(\boldsymbol{\theta})\bigr)\right)^{1/p}, \qquad p \in (-\infty, 1].$$

However, E-optimality where $g(\boldsymbol{\theta}) = \boldsymbol{\theta}$ and $p = -\infty$, is not included here since generally it does not satisfy the smoothness condition on PD(d); a short discussion about this can be found in Section 5. In addition to the Φ_p-optimality criteria, our general Φ-optimality criteria also include compound optimality criteria, criteria for evaluating a mixture of information matrices obtained from nested models [see Pukelsheim (1993), Chapter 11] and so on.

2.1. *Preliminary results.* While finding optimal designs is an optimization problem, the dimensionality of the optimization problem is unknown since the number of design points, q, is unknown. However, it has been observed in the literature that optimal designs are often saturated designs. This phenomenon was first discovered in de la Garza (1954), and was generalized to a class of models in Yang and Stufken (2009, 2012), Yang (2010) as well as in Dette and Melas (2011) and Dette and Schorning (2013), where the latter two papers provided a different perspective on this phenomenon using *Chebyshev systems*. Based on these results, optimal designs can be found in a small *complete class* of designs, denoted as Ξ, and in many cases Ξ only consists of designs with at most d design points. Here, we briefly introduce a fundamental theorem from Yang and Stufken (2012) for our later use. Using the techniques there, we decompose the Fisher information matrix in the following way (an example is given at the end of Section 2.1):

$$\mathbf{M}_{\xi}(\boldsymbol{\theta}) = \mathbf{P}(\boldsymbol{\theta})\mathbf{C}_{\xi}(\boldsymbol{\theta})\mathbf{P}(\boldsymbol{\theta})^T, \qquad \mathbf{C}_{\xi}(\boldsymbol{\theta}) = \left(\sum_{i=1}^{q} \omega_i \mathbf{C}(\boldsymbol{\theta}, c_i)\right), \tag{2.1}$$

where $\mathbf{C}(\boldsymbol{\theta}, c)$ is a $d \times d$ symmetric matrix,

$$\mathbf{C}(\boldsymbol{\theta}, c) = \begin{pmatrix} \Psi_{11}(\boldsymbol{\theta}, c) & & & \\ \Psi_{21}(\boldsymbol{\theta}, c) & \Psi_{22}(\boldsymbol{\theta}, c) & & \\ \vdots & \vdots & \ddots & \\ \Psi_{d1}(\boldsymbol{\theta}, c) & \Psi_{d2}(\boldsymbol{\theta}, c) & \cdots & \Psi_{dd}(\boldsymbol{\theta}, c) \end{pmatrix},$$

$\mathbf{P}(\boldsymbol{\theta})$ is a $d \times d$ nonsingular matrix that only depends on $\boldsymbol{\theta}$, and $c \in [A, B]$ is a smooth monotonic transformation of x that depends on $\boldsymbol{\theta}$. For the sake of simplicity, we drop $\boldsymbol{\theta}$ from the notation of matrix $\mathbf{C}(\boldsymbol{\theta}, c)$ and its elements hereafter [in fact, in many cases a nice decomposition can be found so that $\mathbf{C}(\boldsymbol{\theta}, c)$ only depends on $\boldsymbol{\theta}$ through c, and $\boldsymbol{\theta}$ becomes redundant in the notation].

For some d_1, $1 \le d_1 < d$, define $\mathbf{C}_{22}(c)$ as the lower $d_1 \times d_1$ principal submatrix of $\mathbf{C}(c)$, that is,

$$\mathbf{C}_{22}(c) = \begin{pmatrix} \Psi_{d-d_1+1,d-d_1+1}(c) & \cdots & \Psi_{d-d_1+1,d}(c) \\ \vdots & \ddots & \vdots \\ \Psi_{d,d-d_1+1}(c) & \cdots & \Psi_{dd}(c) \end{pmatrix}.$$

Choose a maximal set of linearly independent nonconstant functions from the first $d - d_1$ columns of the matrix $\mathbf{C}(c)$, let the number of functions in this set be $k - 1$, and rename them as $\Psi_\ell(c)$, $\ell = 1, \dots, k - 1$. Let $\Psi_k(c) = \mathbf{C}_{22}(c)$, and define the functions $f_{\ell,t}(c)$, $1 \le t \le \ell \le k$, to be

$$\begin{pmatrix} f_{1,1} = \Psi_1' & & & & & \\ f_{2,1} = \Psi_2' & f_{2,2} = \left(\frac{f_{2,1}}{f_{1,1}}\right)' & & & & \\ f_{3,1} = \Psi_3' & f_{3,2} = \left(\frac{f_{3,1}}{f_{1,1}}\right)' & f_{3,3} = \left(\frac{f_{3,2}}{f_{2,2}}\right)' & & & \\ f_{4,1} = \Psi_4' & f_{4,2} = \left(\frac{f_{4,1}}{f_{1,1}}\right)' & f_{4,3} = \left(\frac{f_{4,2}}{f_{2,2}}\right)' & f_{4,4} = \left(\frac{f_{4,3}}{f_{3,3}}\right)' & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ f_{k,1} = \Psi_k' & f_{k,2} = \left(\frac{f_{k,1}}{f_{1,1}}\right)' & f_{k,3} = \left(\frac{f_{k,2}}{f_{2,2}}\right)' & f_{k,4} = \left(\frac{f_{k,3}}{f_{3,3}}\right)' & \cdots & f_{k,k} = \left(\frac{f_{k,k-1}}{f_{k-1,k-1}}\right)' \end{pmatrix},$$

where the entries in the last row are matrices, and the derivatives of matrices are element-wise derivatives (assuming all derivatives exist). Define matrix $\mathbf{F}(c) = \prod_{\ell=1}^{k} f_{\ell,\ell}(c)$. Then the following theorem due to Yang and Stufken (2012) is available [see also Dette and Schorning (2013), Theorem 3.1].

THEOREM 2.1 [Yang and Stufken (2012)]. *For a regression model with a single covariate, suppose that either* $\mathbf{F}(c)$ *or* $-\mathbf{F}(c)$ *is positive definite for all* $c \in [A, B]$. *Then the following results hold*:

(a) *If* $k = 2m - 1$ *is odd and* $\mathbf{F}(c) < 0$, *then designs with at most* m *design points, including point* A, *form a complete class* Ξ.

(b) *If* $k = 2m - 1$ *is odd and* $\mathbf{F}(c) > 0$, *then designs with at most* m *design points, including point* B, *form a complete class* Ξ.

(c) *If* $k = 2m$ *is even and* $\mathbf{F}(c) < 0$, *then designs with at most* m *design points, form a complete class* Ξ.

(d) *If* $k = 2m - 2$ *is even and* $\mathbf{F}(c) > 0$, *then designs with at most* m *design points, including both* A *and* B, *form a complete class* Ξ.

It is helpful to sketch how Theorem 2.1 is proved. For some carefully chosen d_1 (see example below) where one of the conditions in Theorem 2.1 holds, it can be proved that for any design $\xi \notin \Xi$, we can find a design $\tilde{\xi} \in \Xi$ such that $\mathbf{C}_{\tilde{\xi}}(\boldsymbol{\theta}) \geq \mathbf{C}_{\xi}(\boldsymbol{\theta})$ under the Loewner ordering, hence $\mathbf{M}_{\tilde{\xi}}(\boldsymbol{\theta}) \geq \mathbf{M}_{\xi}(\boldsymbol{\theta})$. To be specific, $\mathbf{C}_{\tilde{\xi}}(\boldsymbol{\theta}) - \mathbf{C}_{\xi}(\boldsymbol{\theta})$ has a *positive definite* lower $d_1 \times d_1$ principal submatrix, and is 0 everywhere else. So the search for optimal designs can be restricted within Ξ.

Theorem 2.1 also applies to generalized linear models. Besides, while it is stated in terms of the "transformed design point" c, the result can be easily translated back into x using the relationship between them, and we will state results in x unless otherwise specified.

In Theorem 2.1, there are four different types of complete classes, the difference being whether one or both of the endpoints are fixed design points (note however a fixed design point can have weight 0 so that it need not be a support point). To make it easier to distinguish, let $\text{fix}(\Xi)$ denote the set of fixed design points for the designs in the complete class Ξ. For example, $\text{fix}(\Xi) = \varnothing$ and $\{L, U\}$ refers to the complete classes in Theorem 2.1(c) and (d), respectively.

Applications of Theorem 2.1 can be found in Yang and Stufken (2009, 2012) and Yang (2010). Obviously, $m \geq d$, however, in many applications we actually find $m = d$. Take the LINEXP model from Yang and Stufken (2012) as an example.

The LINEXP model is used to characterize tumor growth delay and regrowth. The natural logarithm of tumor volume is modeled using a nonlinear regression model with mean

$$\eta(x, \boldsymbol{\theta}) = \theta_1 + \theta_2 e^{\theta_3 x} + \theta_4 x, \tag{2.2}$$

where $x \in [L, U]$ is the time, $\theta_1 + \theta_2$ is the logarithm of initial tumor volume, $\theta_3 < 0$ is the rate at which killed cells are eliminated, $\theta_4 > 0$ is the final growth rate.

The information matrix for $\boldsymbol{\theta}$ can be written in the form of (2.1) with

$$P(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \theta_2/\theta_3 \\ 0 & 0 & 1/\theta_3 & 0 \end{pmatrix}, \qquad \mathbf{C}(c) = \begin{pmatrix} 1 & & & \\ e^c & e^{2c} & & \\ c & ce^c & c^2 & \\ ce^c & ce^{2c} & c^2e^c & c^2e^{2c} \end{pmatrix},$$

where $c = \theta_3 x \in [A, B] = [\theta_3 U, \theta_3 L]$. Let $d_1 = 2$, $\mathbf{C}_{22}(c)$ be the lower 2×2 principal submatrix of $\mathbf{C}(c)$, and $\Psi_1(c) = c$, $\Psi_2(c) = e^c$, $\Psi_3(c) = ce^c$, $\Psi_4(c) = e^{2c}$, $\Psi_5(c) = ce^{2c}$ be the set of linearly independent nonconstant functions from the first two columns of $\mathbf{C}(c)$. Then $k = 6$, $f_{1,1} = 1$, $f_{2,2} = e^c$, $f_{3,3} = 1$, $f_{4,4} = 4e^c$, $f_{5,5} = 1$, and

$$f_{6,6}(c) = \begin{pmatrix} 2e^{-2c} & e^{-c}/2 \\ e^{-c}/2 & 2 \end{pmatrix}, \qquad \mathbf{F}(c) = \prod_{\ell=1}^{6} f_{\ell,\ell}(c) = \begin{pmatrix} 8 & 2e^c \\ 2e^c & 8e^{2c} \end{pmatrix}.$$

Because $\mathbf{F}(c) > 0$, Theorem 2.1(d) can be applied with $m = 4 = d$, and Ξ consists of designs with at most four design points including both endpoints, thus $\text{fix}(\Xi) = \{L, U\}$.

2.2. *Identifying the optimal design.* If one of the cases in Theorem 2.1 holds, an optimal design exists of the form $\xi = \{(x_i, \omega_i)\}_{i=1}^m$, where x_i's are strictly increasing, with x_1 or x_m possibly fixed to be L or U, respectively; ω_i's are nonnegative, and $\omega_1 = 1 - \sum_{i=2}^m \omega_i$. Let $\mathbf{Z}$ be the vector of unknown design points (i.e., exclude x_1 or x_m if fixed to be the endpoint) and $m-1$ unknown weights $\omega_2, \dots, \omega_m$. For example, for the LINEXP model in (2.2), $\mathbf{Z} = (x_2, x_3, \omega_2, \omega_3, \omega_4)^T$ since $m = 4$ and $x_1 = L, x_4 = U$. Thus we can use $\mathbf{Z}$ to represent the design ξ. Now the objective function $\Phi(\mathbf{M}_\xi(\boldsymbol{\theta}))$ is a function of $\mathbf{Z}$, denoted as $\tilde{\Phi}(\mathbf{Z})$, and it is smooth by the smoothness of Φ. To find an optimal design, we need to maximize $\tilde{\Phi}(\mathbf{Z})$ with respect to $\mathbf{Z}$. The simplest way is to find the critical points, specifically, the *feasible critical points*, as defined below.

DEFINITION 2.1. A critical point of $\tilde{\Phi}(\mathbf{Z})$, $\mathbf{Z}^c$, is a feasible critical point if all the design points in $\mathbf{Z}^c$ are within $[L, U]$ and all $m-1$ weights are positive with summation less than 1.

With Definition 2.1, a feasible critical point gives a design with m support points. Moreover, Theorem 2.2 states the conditions such that a feasible critical point gives a globally optimal design.

THEOREM 2.2. *Assume one of the cases in Theorem* 2.1 *holds, then for any feasible critical point of* $\tilde{\Phi}(\mathbf{Z})$, *its corresponding design is a* Φ*-optimal design.*

PROOF. See the Appendix. □

Theorem 2.2 gives an implicit solution of an optimal design if there exists a feasible critical point. Such a point can be given explicitly in special situations, but not in general due to the complexity of the objective function. Nevertheless, we have an implicit solution and it can be easily solved using Newton's algorithm. However, we need to guarantee the existence of a feasible critical point in the first place. Theorem 2.3 gives some sufficient conditions that a feasible critical point exists.

THEOREM 2.3. *Suppose one of the cases in Theorem* 2.1 *holds and any* Φ-*optimal design has at least m support points*. *Further assume one of the following four conditions holds*:

(a) $\text{fix}(\Xi) = \{L\}$, *and the information matrix* $\mathbf{M}_U(\boldsymbol{\theta})$ *is* 0;
(b) $\text{fix}(\Xi) = \{U\}$, *and* $\mathbf{M}_L(\boldsymbol{\theta}) = 0$;
(c) $\text{fix}(\Xi) = \varnothing$, *and* $\mathbf{M}_U(\boldsymbol{\theta}) = \mathbf{M}_L(\boldsymbol{\theta}) = 0$;
(d) $\text{fix}(\Xi) = \{L, U\}$.

Then a feasible critical point of $\tilde{\Phi}(\mathbf{Z})$ *must exist*, *and by Theorem* 2.2, *any such point gives a* Φ-*optimal design*.

PROOF. Let $\xi^* \in \Xi$ be a Φ-optimal design, then ξ^* has at least m support points. By Theorem 2.1, designs in the complete class have at most m support points, hence ξ^* has exactly m support points. Let $\mathbf{Z}^*$ be the vector corresponding to ξ^* according to the definition in the beginning of Section 2.2. For each of conditions (a)∼(d), we know the design points in $\mathbf{Z}^*$ do not include any of the endpoints (recall the fixed design points are excluded in $\mathbf{Z}^*$), hence they all belong to the open interval (L, U). The weights in $\mathbf{Z}^*$ are all positive, hence all belong to the open interval (0, 1), so $\mathbf{Z}^*$ is not on the boundary and must be a critical point of $\tilde{\Phi}(\mathbf{Z})$. This proves the existence. □

The condition in Theorem 2.3 that every Φ-optimal design has at least m support points is met with $m = d$ for many models and optimality criteria. For example, when $K(\theta)$ is a nonsingular matrix, any Φ-optimal design has at least d support points for commonly used optimality criteria. On the other hand, as we have stated, for many models, the complete class given by Theorem 2.1 only consists of designs with at most d support points. The condition (d) is found to be satisfied for several models, as we will see in Section 3. For condition (a), usually $\mathbf{M}_U(\boldsymbol{\theta}) = 0$ only when $U = \infty$, so the condition fails if we are interested in a finite design region, and so do conditions (b) and (c). This issue will be addressed later in Theorem 2.6.

Useful results can be obtained by applying Theorem 2.3 to the most commonly used Φ_p-optimality criteria. In particular, we are interested in Φ_p-optimal designs for $\boldsymbol{\theta}$ or $\mathbf{a}^T\boldsymbol{\theta}$, where $\mathbf{a} = (a_1, \ldots, a_d)^T$ is a $d \times 1$ vector such that $\mathbf{a}^T\boldsymbol{\theta}$ is only estimable with at least d support points. Adopting the notation in Kiefer and Wolfowitz (1965), define

$$A^* = \{\mathbf{a} | \mathbf{a}^T \boldsymbol{\theta} \text{ is only estimable with at least } d \text{ support points}\}.$$

Now Corollary 2.4 gives applications of Theorem 2.3 to Φ_p-optimal designs.

COROLLARY 2.4. *Suppose that one of the cases in Theorem 2.1 holds with $m = d$, and one of the four conditions in Theorem 2.3 is met. Consider Φ_p-optimal design for $g(\boldsymbol{\theta})$ where $g(\boldsymbol{\theta})$ satisfies either case* (i) *or* (ii) *below*:

(i) *$g(\boldsymbol{\theta}) = \boldsymbol{\theta}$ or a reparameterization of $\boldsymbol{\theta}$*;
(ii) *$g(\boldsymbol{\theta}) = \mathbf{a}^T \boldsymbol{\theta}, \mathbf{a} \in A^*$.*

Then a feasible critical point of $\tilde{\Phi}(\mathbf{Z})$ exists, and any such point gives a Φ_p-optimal design for $g(\boldsymbol{\theta})$.

REMARK 2.1. In Corollary 2.4(i), a special case of a reparameterization is $g(\boldsymbol{\theta}) = \mathbf{W}\boldsymbol{\theta}$, where $\mathbf{W}$ is a diagonal matrix with positive diagonal elements. This makes $\text{cov}(g(\hat{\boldsymbol{\theta}}))$ a rescaled version of $\text{cov}(\hat{\boldsymbol{\theta}})$, and it makes sense when $\text{var}(\hat{\theta}_i)$'s are of different orders of magnitude. For example, in Dette (1997), the author proposed "standardized" optimality criteria, where the matrix $\mathbf{W}$ has diagonal elements $\mathbf{W}_{ii} = \sqrt{1/(\mathbf{M}_{\xi_i^*}^{-1})_{ii}}$, ξ_i^* is the c-optimal design for estimating θ_i alone, $i = 1, \ldots, d$. Under the conditions of Corollary 2.4(i), finding such optimal designs is easy after we find ξ_i^*'s.

REMARK 2.2. Corollary 2.4(ii) considers c-optimality. When $\mathbf{a} \in A^*$, the c-optimal design is supported at the full set of Chebyshev points in many cases [see Studden (1968)], but our method gives another way of finding c-optimal designs. When $\mathbf{a} \notin A^*$, sometimes a feasible critical point still exists, and it still gives an optimal design. However, if there is no such critical point, then the c-optimal design must be supported at fewer points, which may not be the Chebyshev points, and this problem becomes harder. Nevertheless, we can approximate such c-optimal designs. Suppose $a_1 \neq 0$, consider $g_\epsilon(\boldsymbol{\theta}) = (\mathbf{a}^T \boldsymbol{\theta}, \epsilon\theta_2, \ldots, \epsilon\theta_d)^T, \epsilon > 0$. A Φ_p-optimal design for $g_\epsilon(\boldsymbol{\theta})$ can be found easily by Corollary 2.4(i). Let $\epsilon \to 0$, it can be shown that these Φ_p-optimal designs will eventually converge to the c-optimal design for $\mathbf{a}^T \boldsymbol{\theta}$ (i.e., the efficiencies of these Φ_p-optimal designs under c-optimality will converge to 1), for any $p \leq -1$. Some examples are provided in Section 3.2.

To verify the condition $\mathbf{a} \in A^*$, let $\mathbf{f}(x, \boldsymbol{\theta}) = (f_1(x, \boldsymbol{\theta}), \ldots, f_d(x, \boldsymbol{\theta})) = \partial\eta(x, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$. The condition $\mathbf{a} \in A^*$ is equivalent to

$$\begin{vmatrix} f_1(x_1, \boldsymbol{\theta}) & \cdots & f_1(x_{d-1}, \boldsymbol{\theta}) & a_1 \\ f_2(x_1, \boldsymbol{\theta}) & \cdots & f_2(x_{d-1}, \boldsymbol{\theta}) & a_2 \\ \vdots & \ddots & \vdots & \vdots \\ f_d(x_1, \boldsymbol{\theta}) & \cdots & f_d(x_{d-1}, \boldsymbol{\theta}) & a_d \end{vmatrix} \neq 0 \tag{2.3}$$

for all $L \leq x_1 < x_2 < \cdots < x_{d-1} \leq U$ (this is also true for generalized linear models). In particular, if we are interested in estimating the individual parameter θ_i, that is, $\mathbf{a} = \mathbf{e}_i$ where $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ denotes the ith unit vector, then $\mathbf{e}_i \in A^*$ is equivalent to $\mathbf{f}_{-i} = \{f_j | j \in \{1, \ldots, d\} \setminus \{i\}\}$ being a Chebyshev system [see Karlin and Studden (1966)], which is easier to verify. Here, the traditional definition of a Chebyshev system is used, which only requires the determinant in (2.3) to be *nonzero* instead of positive.

Next, the uniqueness of optimal designs can also be established under mild conditions. We first introduce some additional terminology. A criterion Φ is called *strictly isotonic* on PD(d) if

$$\Phi(\mathbf{M}_1) > \Phi(\mathbf{M}_2) \qquad \text{for any } \mathbf{M}_1 \geq \mathbf{M}_2 > 0 \text{ and } \mathbf{M}_1 \neq \mathbf{M}_2.$$

It is called *strictly concave* on PD(d) if

$$\Phi\big(\alpha\mathbf{M}_1 + (1-\alpha)\mathbf{M}_2\big) > \alpha\Phi(\mathbf{M}_1) + (1-\alpha)\Phi(\mathbf{M}_2),$$
$$\text{for any } \alpha \in (0, 1), \mathbf{M}_1 > 0, \mathbf{M}_2 \geq 0 \text{ and } \mathbf{M}_2 \not\propto \mathbf{M}_1.$$

For example, Φ_p-optimality criteria are both strictly isotonic and strictly concave on PD(d) when $g(\boldsymbol{\theta})$ is $\boldsymbol{\theta}$ or a reparameterization of $\boldsymbol{\theta}$ and $p \in (-\infty, 1)$ [see Pukelsheim (1993), page 151]. Moreover, a compound optimality criterion which involves a strictly isotonic and strictly concave criterion is also strictly isotonic and strictly concave. For these criteria, we have Theorem 2.5.

THEOREM 2.5. *Assume that one of the cases in Theorem 2.1 holds. If Φ is both strictly isotonic and strictly concave on PD(d) and there exists a Φ-optimal design ξ^* which has at least d support points, then ξ^* is the unique Φ-optimal design. In particular, the Φ_p-optimal design under Corollary 2.4*(i) *is unique for $p \in (-\infty, 1)$.*

PROOF. See the Appendix. □

REMARK 2.3. The c-optimality criterion with $g(\boldsymbol{\theta}) = \mathbf{a}^T\boldsymbol{\theta}$ maybe neither strictly concave nor strictly isotonic on PD(d). However, if $\mathbf{a} \in A^*$ and $\mathbf{f}(x, \boldsymbol{\theta})$ is a Chebyshev system, the uniqueness is proved in Studden (1968).

The uniqueness is not only of interest in itself, but also has implications for finding optimal designs. As we have stated earlier, conditions (a), (b) and (c) in Theorem 2.3 may only hold on a large design region, call it the *full design region*. Let ξ^{**} be a Φ-optimal design on the full design region with smallest support point $x^{**}_{\min}$ and largest support point $x^{**}_{\max}$. Then for a smaller design region $[L, U]$, under the same optimality criterion Φ, we have Theorem 2.6.

THEOREM 2.6. *Assume that one of the cases in Theorem 2.1 holds for the full design region, and both Φ-optimal designs on $[L, U]$ and the full design region are unique with support size m, then we have*:

(a) *under* $\operatorname{fix}(\Xi) = \{L\}$, *if* $U < x^{**}_{\max}$, *then the Φ-optimal design on $[L, U]$ has both L and U as support points*; *otherwise*, *the optimal design is ξ^{**}*;
(b) *under* $\operatorname{fix}(\Xi) = \{U\}$, *if* $x^{**}_{\min} < L$, *then the Φ-optimal design on $[L, U]$ has both L and U as support points*; *otherwise*, *the optimal design is ξ^{**}*;
(c) *under* $\operatorname{fix}(\Xi) = \varnothing$, *if* $x^{**}_{\min} < L$ *or* $U < x^{**}_{\max}$, *then the Φ-optimal design on $[L, U]$ has at least one endpoint as a support point*; *otherwise*, *the optimal design is ξ^{**}*.

PROOF. We only give the proof for case (a), others being similar. When $U \geq x^{**}_{\max}$, the design ξ^{**} is still a feasible design on the region $[L, U]$, and it is optimal because it is optimal on the full design region. When $U < x^{**}_{\max}$, ξ^{**} is no longer a feasible design, let ξ^* be the optimal design on $[L, U]$. A complete class of the same type exists for design region $[L, U]$ because, for example, $\mathbf{F}(c) > 0$ on the full design region implies $\mathbf{F}(c) > 0$ on the smaller design region. So $x^*_1 = L$. If the largest support point $x^*_m < U$, then $\mathbf{Z}^* = (x^*_2, \dots, x^*_m, \omega^*_2, \dots, \omega^*_m)^T$ must be a critical point of $\tilde{\Phi}(\mathbf{Z})$. Now if we consider the optimal design problem on the full design region again, $\mathbf{Z}^*$ is a feasible critical point, and by Theorem 2.2, ξ^* must be an optimal design on the full design region. However, $\xi^* \neq \xi^{**}$, this contradicts the uniqueness assumption. $\square$

3. Application. The theorems we have established can be used to find optimal designs for many models. In Sections 3.1 through 3.3, we consider Φ_p-optimal designs for models with two, three and four or six parameters, respectively. In Section 3.4, we consider polynomial regression models with arbitrary d parameters under more general optimality criteria.

3.1. *Models with two parameters*. Yang and Stufken (2009) considered complete class results for two-parameter models, including logistic/probit regression model, Poisson regression model and Michaelis–Menten model. The theorems we have established can be used to find the optimal designs. Take the Poisson regression model as an example (the applications to other models are similar). It has the following form:

$$\eta(x, \boldsymbol{\theta}) = \log\bigl(\mathrm{E}(y)\bigr) = \theta_1 + \theta_2 x, \qquad x \in [L, U].$$

TABLE 1
A-optimal designs for Poisson regression model on $[0, \infty)$

	A-optimal	
θ_2	(x_1, x_2)	(ω_1, ω_2)
−1	(0, 2.261)	(0.444, 0.556)
−2	(0, 1.193)	(0.320, 0.680)

Theorem 2.1(b) can be applied to this model, and a complete class consists of designs with at most 2 design points including one boundary point [see Yang and Stufken (2009), Theorem 4]. Specifically, when $\theta_2 > 0$, U is a fixed design point, and $\mathbf{M}_{-\infty}(\boldsymbol{\theta}) = 0$ [since $\mathbf{M}_x(\boldsymbol{\theta}) = e^{\theta_1+\theta_2 x}(1, x)^T(1, x)$]; when $\theta_2 < 0$, L is a fixed design point, and $\mathbf{M}_{\infty}(\boldsymbol{\theta}) = 0$. Thus, on any one-sided restricted region $(-\infty, U]$ (when $\theta_2 > 0$) or $[L, \infty)$ (when $\theta_2 < 0$), Φ_p-optimal designs for $\boldsymbol{\theta}$ can be found by solving for the critical points, according to Corollary 2.4(i). For c-optimality, recall $\mathbf{f}(x, \boldsymbol{\theta}) = \partial\eta(x, \boldsymbol{\theta})/\partial\boldsymbol{\theta} = (1, x)$, thus $\mathbf{f}_{-2} = \{1\}$ is a Chebyshev system, which means θ_2 can only be estimated with at least $d = 2$ support points. Therefore, according to Corollary 2.4(ii), an $\mathbf{e}_2$-optimal design (c-optimal design for θ_2) can also be found by solving for the critical points.

In particular, D- and $\mathbf{e}_2$-optimal designs can be found analytically through symbolic computation software (e.g., by using the solve function in Matlab) and are listed in (3.1) and (3.2). Note that they do not depend on θ_1 since e^{θ_1} is merely a multiplicative factor in $\mathbf{M}_x(\boldsymbol{\theta})$:

$$\xi_D^* = \begin{cases} \{(U - 2/\theta_2, 1/2), (U, 1/2)\}, & \theta_2 > 0, \\ \{(L - 2/\theta_2, 1/2), (L, 1/2)\}, & \theta_2 < 0, \end{cases} \tag{3.1}$$

$$\xi_{\mathbf{e}_2}^* = \begin{cases} \{(U - 2.557/\theta_2, 0.782), (U, 0.218)\}, & \theta_2 > 0, \\ \{(L - 2.557/\theta_2, 0.782), (L, 0.218)\}, & \theta_2 < 0. \end{cases} \tag{3.2}$$

However, A-optimal designs do not have explicit forms. Nevertheless, the solutions can be found easily using Newton's algorithm. For the case of $\theta_2 < 0$, some examples are listed in Table 1 (again the optimal designs do not depend on θ_1).

In addition, the Φ_p-optimal design for $\boldsymbol{\theta}$ and $\mathbf{e}_2$-optimal design are unique, due to Theorem 2.5. For finite design regions, Theorem 2.6 can be applied. For example, the A-optimal design for $\boldsymbol{\theta} = (1, -1)^T$ on $[0, U]$ when $U \geq 2.261$ is $\{(0, 0.444), (2.261, 0.556)\}$; when $U < 2.261$, the optimal design is supported at exactly two points 0 and U, and the weights can be determined easily.

3.2. *Models with three parameters*. Dette et al. (2008, 2010) considered optimal designs for the Emax and log-linear models. These models, often used to

TABLE 2
A-optimal designs for the Emax model on [0, 150]

θ_2	θ_3	(x_1, x_2, x_3)	$(\omega_1, \omega_2, \omega_3)$
7/15	15	(0, 12.50, 150)	(0.250, 0.500, 0.250)
7/15	25	(0, 18.75, 150)	(0.250, 0.500, 0.250)
10/15	25	(0, 18.75, 150)	(0.250, 0.500, 0.250)

model dose-response curves, are nonlinear regression models with means

$$\eta(x, \boldsymbol{\theta}) = \begin{cases} \theta_1 + \theta_2 x/(x+\theta_3), & \text{Emax}, \\ \theta_1 + \theta_2 \log(x+\theta_3), & \text{log-linear}. \end{cases}$$

Here, $x \in [L, U] \subseteq (0, \infty)$ is the dose range, $\theta_2 > 0$ and $\theta_3 > 0$. Theorem 2.1(d) can be applied to both models, and a complete class consists of designs with at most 3 design points including *both* endpoints [Yang (2010), Theorem 3]. Hence, Corollary 2.4 is applicable on design space $[L, U]$. In particular, D-optimal designs can be computed explicitly using symbolic computation software, and are listed in (3.3). They are consistent with the results in Dette et al. (2010):

$$\xi_D^* = \begin{cases} \{(L, 1/3), (x_E^*, 1/3), (U, 1/3)\}, & \text{Emax}, \\ \{(L, 1/3), (x_l^*, 1/3), (U, 1/3)\}, & \text{log-linear}, \end{cases} \tag{3.3}$$

where

$$\begin{aligned} x_E^* &= \frac{L(U+\theta_3) + U(L+\theta_3)}{L + U + 2\theta_3}, \\ x_l^* &= \frac{(L+\theta_3)(U+\theta_3)}{U-L} \log\left(\frac{U+\theta_3}{L+\theta_3}\right) - \theta_3. \end{aligned} \tag{3.4}$$

For A-optimality, numerical solutions can be obtained easily by Newton's algorithm. Table 2 gives some examples for the Emax model using parameter settings in Dette et al. (2008) (the optimal designs do not depend on θ_1 since it is not involved in the information matrix; and although it seems that the optimal weights are constant, they do change gradually with θ_2 and θ_3).

For c-optimality, Dette et al. (2010) gave explicit solutions for ED_p-optimal designs, where an ED_p-optimal design is a design that is optimal for estimating the dose that achieves $100p\%$ of the maximum effect in dose range $[L, U]$, $0 < p < 1$. In fact, ED_p-optimality is equivalent to $\mathbf{e}_3$-optimality regardless of p, and we can find the optimal designs using our method. First, we have

$$\mathbf{f}(x, \boldsymbol{\theta}) = \begin{cases} \bigl(1, x/(x+\theta_3), -\theta_2 x/(x+\theta_3)^2\bigr), & \text{Emax}, \\ \bigl(1, \log(x+\theta_3), \theta_2/(x+\theta_3)\bigr), & \text{log-linear}. \end{cases}$$

It is easy to prove for both the Emax and log-linear models that $\mathbf{f}_{-3}$ is a Chebyshev system, which means that θ_3 is only estimable with at least $d = 3$ support points. So $\mathbf{e}_3$-optimal designs can be found by solving for the critical points, by Corollary 2.4(ii). The solutions can be found explicitly using symbolic computation software and are listed in (3.5). They are consistent with the results in Dette et al. (2010):

$$\xi^*_{\mathbf{e}_3} = \xi^*_{ED_p} = \begin{cases} \{(L, 1/4), (x^*_E, 1/2), (U, 1/4)\}, & \text{Emax}, \\ \{(L, \omega^*_l), (x^*_l, 1/2), (U, 1/2 - \omega^*_l)\}, & \text{log-linear}, \end{cases} \tag{3.5}$$

where x^*_E and x^*_l are the same as in (3.4), and

$$\omega^*_l = \frac{\log(x^*_l + \theta_3) - \log(U + \theta_3)}{2(\log(L + \theta_3) - \log(U + \theta_3))}.$$

Regarding $\mathbf{f}_{-2}$, it can be shown that it is always a Chebyshev system for the log-linear model, and it is a Chebyshev system for the Emax model if $\theta_3 \notin (L, U)$. In such cases, $\mathbf{e}_2$-optimal designs can be found according to Corollary 2.4(ii), and the solutions can be derived analytically as shown in (3.6):

$$\xi^*_{\mathbf{e}_2} = \begin{cases} \left\{\left(L, \dfrac{1}{4} - \dfrac{(U-L)\theta_3}{8(\theta_3^2 - LU)}\right), \left(x^*_E, \dfrac{1}{2}\right), \left(U, \dfrac{1}{4} + \dfrac{(U-L)\theta_3}{8(\theta_3^2 - LU)}\right)\right\}, \\ \hfill \text{Emax}, \theta_3 \notin (L, U), \\ \left\{\left(L, \dfrac{(U - x^*_l)(L+\theta_2)}{2(U-L)(x^*_l + \theta_2)}\right), \left(x^*_l, \dfrac{1}{2}\right), \left(U, \dfrac{(x^*_l - L)(U+\theta_2)}{2(U-L)(x^*_l + \theta_2)}\right)\right\}, \\ \hfill \text{log-linear}. \end{cases} \tag{3.6}$$

When $\theta_3 \in (L, U)$, $\mathbf{f}_{-2}$ is no longer a Chebyshev system for the Emax model. However, if $|(U - L)\theta_3| < |2(\theta_3^2 - LU)|$, the weights of $\xi^*_{\mathbf{e}_2}$ in (3.6) are still positive, and the design is still $\mathbf{e}_2$-optimal; otherwise, the optimal design is supported at fewer than 3 points, which may not be the Chebyshev points. Nevertheless, we can approach the optimal design using the method in Remark 2.2. To show this, consider the setting where the dose range is $[0, 150]$, $\theta_2 = 7/15$ and $\theta_3 = 25$. The exact $\mathbf{e}_2$-optimal design can be found to be $\xi^*_{\mathbf{e}_2} = \{(\theta_3^2/U, 0.5), (U, 0.5)\} = \{(25/6, 0.5), (150, 0.5)\}$ using Elfving's method [Elfving (1952)]. Now let $\epsilon = 10^{-5}, 10^{-6}, 10^{-7}$; the Φ_p-optimal designs for estimating $g_\epsilon(\boldsymbol{\theta}) = (\epsilon\theta_1, \theta_2, \epsilon\theta_3)^T$ can be found by Corollary 2.4(i) and are used to approximate the $\mathbf{e}_2$-optimal design. Table 3 shows the errors and $1-$ efficiencies of the approximation for $p = -1$ and -3. As we can see, the error gets sufficiently small after a few iterations, especially when $|p|$ is larger; however, due to singularity issues, the error cannot be made arbitrary small.

3.3. *Models with four or six parameters.* Demidenko (2004) used a double exponential model to characterize the regrowth of tumor after radiation. The natural logarithm of tumor volume can be modeled using a nonlinear regression model with mean

$$\eta(x, \boldsymbol{\theta}) = \theta_1 + \log(\theta_2 e^{\theta_3 x} + (1 - \theta_2) e^{-\theta_4 x}),$$

TABLE 3
Approximating two point $\mathbf{e}_2$*-optimal design using three point designs for the Emax model*

p	ϵ	$\lvert x_2 - \frac{25}{6}\rvert/\frac{25}{6}$	$\lvert\omega_1\rvert$	$\lvert\omega_2 - 0.5\rvert$	$\lvert\omega_3 - 0.5\rvert$	$1 - \text{eff}$
-1	10^{-5}	10^{-2}	$4 \cdot 10^{-3}$	$4 \cdot 10^{-3}$	$2 \cdot 10^{-4}$	$6 \cdot 10^{-4}$
	10^{-6}	10^{-3}	$4 \cdot 10^{-4}$	$4 \cdot 10^{-4}$	$2 \cdot 10^{-5}$	$6 \cdot 10^{-5}$
	10^{-7}	10^{-4}	$4 \cdot 10^{-5}$	$4 \cdot 10^{-5}$	$2 \cdot 10^{-6}$	$6 \cdot 10^{-6}$
-3	10^{-5}	$3 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$9 \cdot 10^{-5}$	$4 \cdot 10^{-6}$	$2 \cdot 10^{-5}$
	10^{-6}	$8 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$5 \cdot 10^{-8}$	$5 \cdot 10^{-7}$
	10^{-7}	$7 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$2 \cdot 10^{-7}$	$1 \cdot 10^{-8}$	$4 \cdot 10^{-8}$

where $0 \le x \in [L, U]$ is the time, θ_1 is the logarithm of the initial tumor volume, $0 < \theta_2 < 1$ is the proportional contribution of the first compartment, and $\theta_3, \theta_4 > 0$ are cell proliferation and death rates.

Demidenko (2006) used the LINEXP model to characterize tumor growth delay and regrowth. The model was described in Section 2.1 and re-presented below:

$$\eta(x, \boldsymbol{\theta}) = \theta_1 + \theta_2 e^{\theta_3 x} + \theta_4 x.$$

Li and Balakrishnan (2011) considered D- and c-optimal designs for these two models, but our approach yields more general results. For both models, Theorem 2.1(d) can be applied, and a complete class consists of designs with at most four design points including both endpoints [see Yang and Stufken (2012)]. Thus, Corollary 2.4 can again be applied on the design space $[L, U]$, and Φ_p-optimal designs for $\boldsymbol{\theta}$ and certain c-optimal designs can be found by solving for the critical points. In particular, $\mathbf{f}_{-3}$ and $\mathbf{f}_{-4}$ are Chebyshev systems under both models [see Li and Balakrishnan (2011)], thus $\mathbf{e}_3$- and $\mathbf{e}_4$-optimal designs for both models can be found by solving for the critical points.

There is no explicit solution for the optimal designs, but numerical solutions can be easily found using Newton's algorithm. Here, we give some A-optimal designs for the LINEXP model in Table 4 (the optimal designs for the LINEXP model do not depend on θ_1 and θ_4 since they are not involved in the information matrix). For D- and c-optimality, our approach gives the same results as in Li and Balakrishnan (2011).

Consider one more example. Dette, Melas and Wong (2006) studied D-optimal designs for exponential regression models, which are nonlinear regression models with mean

$$\eta(x, \boldsymbol{\theta}) = \sum_{s=1}^{S} \theta_{2s-1} e^{-\theta_{2s} x}, \qquad 0 \le x \in [L, U], \tag{3.7}$$

where $\theta_{2s-1} \neq 0, s = 1, \dots, S, 0 < \theta_2 < \cdots < \theta_{2S}$. When $S = 2$ and $\theta_4/\theta_2 < 61.98$ or $S = 3, 2\theta_4 = \theta_2 + \theta_6$ and $\theta_4/\theta_2 < 23.72$, Theorem 2.1(b) can be applied, and a

TABLE 4
A-optimal designs for the LINEXP model on [0, 1]

θ_2	θ_3	(x_1, x_2, x_3, x_4)	$(\omega_1, \omega_2, \omega_3, \omega_4)$
0.5	−1	(0, 0.220, 0.717, 1)	(0.156, 0.324, 0.344, 0.176)
1	−1	(0, 0.220, 0.717, 1)	(0.151, 0.319, 0.349, 0.181)
1	−2	(0, 0.195, 0.681, 1)	(0.146, 0.315, 0.355, 0.184)

complete class consists of designs with at most $2S$ design points including the lower endpoint L [see Yang and Stufken (2012), Theorems 3 and 4]. Moreover, it is easy to see that the information matrix $\mathbf{M}_x$ goes to 0 when x approaches infinity, thus Corollary 2.4 can be applied on any design region $[L, \infty)$. Table 5 gives some A-optimal designs for $S = 2$.

For c-optimality, first we have

$$\mathbf{f}(x, \boldsymbol{\theta}) = \begin{cases} \left(e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x}, e^{-\theta_4 x}, -\theta_3 x e^{-\theta_4 x}\right), & S = 2, \\ \left(e^{-\theta_2 x}, -\theta_1 x e^{-\theta_2 x}, e^{-\theta_4 x}, -\theta_3 x e^{-\theta_4 x}, e^{-\theta_6 x}, -\theta_5 x e^{-\theta_6 x}\right), & S = 3. \end{cases}$$

Both are Chebyshev systems. In addition, we can show that $\mathbf{f}_{-2s}, s = 1, \dots, S$ are Chebyshev systems for $S = 2$ and $S = 3$, so the c-optimal designs for $\theta_{2s}, s = 1, \dots, S$ on $[L, \infty)$ can be found by solving for the critical points. Table 5 gives some $\mathbf{e}_2$-optimal designs for $S = 2$.

Moreover, the Φ_p-optimal designs for $\boldsymbol{\theta}$ and c-optimal design for θ_{2s}'s are unique by Theorem 2.5. For a finite design region, Theorem 2.6 can be applied. For example, the A-optimal design for $\boldsymbol{\theta} = (1, 1, 1, 2)^T$ on $[0, U]$ when $U \geq 3.416$ is the same as in Table 5; when $U < 3.416$, the optimal design is supported at 4 design points including both 0 and U.

3.4. *Polynomial regression model with d parameters*. Yang (2010) considered the general $(d-1)$th degree polynomial regression model P_{d-1} with variance

TABLE 5
A- and $\mathbf{e}_2$*-optimal designs for exponential regression model on* $[0, \infty)$ *when* $S = 2, \theta_1 = \theta_2 = 1$

Criterion	θ_3	θ_4	(x_1, x_2, x_3, x_4)	$(\omega_1, \omega_2, \omega_3, \omega_4)$
A-optimality	1	2	(0, 0.275, 1.196, 3.416)	(0.078, 0.178, 0.251, 0.493)
	1	4	(0, 0.170, 0.768, 2.472)	(0.118, 0.261, 0.287, 0.334)
	3	4	(0, 0.172, 0.760, 2.450)	(0.083, 0.199, 0.296, 0.422)
$\mathbf{e}_2$-optimality	1	2	(0, 0.273, 1.197, 3.425)	(0.054, 0.124, 0.200, 0.623)
	1	4	(0, 0.168, 0.769, 2.492)	(0.033, 0.082, 0.201, 0.683)
	3	4	(0, 0.168, 0.769, 2.492)	(0.033, 0.082, 0.201, 0.683)

$\sigma^2/\lambda(x)$ and mean

$$\eta(x, \boldsymbol{\theta}) = \theta_1 + \sum_{i=2}^{d} \theta_i x^{i-1}. \tag{3.8}$$

For different choices of the efficiency function $\lambda(x)$, Theorem 2.1 gives the following complete class results [see Yang (2010), Theorem 9]:

(a) When (i) $\lambda(x) = 1 - x, x \in [-1, 1]$ or (ii) $\lambda(x) = e^{-x}, x \in [0, \infty)$, a complete class consists of designs with at most d design points including the left endpoint. Moreover, the information matrix $\mathbf{M}_U(\boldsymbol{\theta}) = 0$.

(b) When $\lambda(x) = 1 + x$, $x \in [-1, 1]$, a complete class consists of designs with at most d design points including the right endpoint. Moreover, the information matrix $\mathbf{M}_L(\boldsymbol{\theta}) = 0$.

(c) When (i) $\lambda(x) = (1 - x)^{u+1}(1 + x)^{v+1}, x \in [-1, 1], u + 1 > 0, v + 1 > 0$ or (ii) $\lambda(x) = x^{u+1}e^{-x}, x \in [0, \infty), u + 1 > 0$ or (iii) $\lambda(x) = e^{-x^2}, x \in (-\infty, \infty)$ or (iv) $\lambda(x) = (1 + x^2)^{-t}, x \in (-\infty, \infty), d \le t$, a complete class consists of designs with at most d design points. Moreover, the information matrices $\mathbf{M}_L(\boldsymbol{\theta}) = \mathbf{M}_U(\boldsymbol{\theta}) = 0$.

(d) When $\lambda(x) \equiv 1, x \in [L, U]$, a complete class consists of designs with at most d design points including both endpoints.

Corollary 2.4 can be applied to the above models on the respective (full) design regions, thus Φ_p-optimal designs for $\boldsymbol{\theta}$ and c-optimal designs for θ_d can be found by solving for the critical points. Furthermore, those designs are unique, so Theorem 2.6 can be used when the design regions are small.

Finally, we apply our theorems to more general optimality criteria. Dette and Studden (1995) considered optimal designs under nested polynomial regression models. To be specific, suppose the degree of the polynomial regression model is an unknown integer between 1 and $d - 1$. The D-optimal design ξ_D^ℓ under a given model P_ℓ, $1 \le \ell \le d - 1$, may not be efficient under another model with a different degree. To take this uncertainty into consideration, the authors proposed the following weighted optimality criteria $\Phi_{p',\boldsymbol{\beta}}$:

$$\Phi_{p',\boldsymbol{\beta}}(\mathbf{M}_\xi) = \left[\sum_{\ell=1}^{d-1} \beta_\ell \big(\mathrm{eff}_D^\ell(\xi)\big)^{p'}\right]^{1/p'}, \tag{3.9}$$

where $p' \in [-\infty, 1]$, $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_{d-1}\}$ is a prior on the set $\{1, \dots, d - 1\}$ with $\beta_{d-1} > 0$,

$$\mathrm{eff}_D^\ell(\xi) = \left(\frac{\det \mathbf{M}_\xi^\ell}{\det \mathbf{M}_{\xi_D^\ell}^\ell}\right)^{1/(\ell+1)}, \qquad \ell = 1, \dots, d - 1,$$

$\mathbf{M}_\xi^\ell$ is the information matrix of ξ under model P_ℓ, and $\mathrm{eff}_D^\ell(\xi)$ is the D-efficiency of ξ under model P_ℓ.

TABLE 6

$\Phi_{p,p',\boldsymbol{\beta}}$-optimal designs for polynomial regression models

p'	(x_1, x_2, x_3, x_4)	$(\omega_1, \omega_2, \omega_3, \omega_4)$	$(\mathrm{eff}_A^1(\xi), \mathrm{eff}_A^2(\xi), \mathrm{eff}_A^3(\xi))$
0	(−0.860, −0.346, 0.346, 0.860)	(0.263, 0.237, 0.237, 0.263)	(0.692, 0.745, 0.902)
−1	(−0.854, −0.343, 0.343, 0.854)	(0.268, 0.232, 0.232, 0.268)	(0.701, 0.753, 0.879)
−3	(−0.846, −0.339, 0.339, 0.846)	(0.273, 0.227, 0.227, 0.273)	(0.714, 0.759, 0.846)

Dette and Studden (1995) gave the solution of $\Phi_{p',\boldsymbol{\beta}}$-optimal design for $\lambda(x) \equiv 1, x \in [-1, 1]$. The solution is rather complicated, and it requires knowledge of canonical moments. An alternative way is to use Theorem 2.3, and it can be applied to more general settings.

First, the D-efficiency in the definition of $\Phi_{p',\boldsymbol{\beta}}$ can be generalized to any Φ_p-efficiency, $p \in (-\infty, 1]$ (e.g., A-efficiency when $p = -1$), and we denote the resulting optimality criteria as $\Phi_{p,p',\boldsymbol{\beta}}$. Second, the efficiency function $\lambda(x)$ can be generalized to any function in cases (a)∼(d) in this subsection, where x belongs to the respective (full) design regions.

Under this general setting, $\Phi_{p,p',\boldsymbol{\beta}}$ always satisfies Assumption A about optimality criteria in Section 2 [see Pukelsheim (1993), page 285]. Moreover, while this optimality criterion is defined on a mixture of different models, these models are nested within the largest model P_{d-1}, thus our complete class result for P_{d-1} can be applied to $\Phi_{p,p',\boldsymbol{\beta}}$. Finally, to use Theorem 2.3, any $\Phi_{p,p',\boldsymbol{\beta}}$-optimal design must have at least d support points. This requirement is reasonable since otherwise the optimal design will not be able to estimate the model P_{d-1}, which may be the true model. To meet the requirement, it is sufficient to restrict ourselves to $p, p' \in (-\infty, 0]$, since any singular matrix will result in $\Phi_{p,p',\boldsymbol{\beta}}$ to be 0. So by Theorem 2.3, $\Phi_{p,p',\boldsymbol{\beta}}$-optimal designs for models in cases (a)∼(d) of this subsection can be found by solving for the critical points. Some examples are given in Table 6 for the case $\lambda(x) = 1 - x^2, x \in [-1, 1]$, $p = -1$ [i.e., for A-efficiency in (3.9)], $d = 4$ and $\boldsymbol{\beta}$ a uniform prior.

In addition, $\Phi_{p,p',\boldsymbol{\beta}}$-optimality is strictly isotonic and strictly concave on PD(d) since $\beta_{d-1} > 0$ and the Φ_p-efficiency under model P_{d-1} is strictly isotonic and strictly concave on PD(d) for $p \in (-\infty, 0]$. Hence by Theorem 2.5, the optimal designs are unique. However, for smaller design regions, the optimality criterion $\Phi_{p,p',\boldsymbol{\beta}}$ changes as the design region changes. For example, when $p = 0$, the design ξ_D^ℓ changes when the design region changes, which causes $\Phi_{p,p',\boldsymbol{\beta}}$ to change. So the optimal design on the full design region cannot be used to obtain the optimal design on a smaller region as we did in Theorem 2.6.

4. Computational advantages. Although it is not the main motivation, our method does provide computational advantages over other algorithms, as Newton's

TABLE 7
Computation time (seconds) for A- and D-optimal designs for the LINEXP model

	A-optimal			D-optimal		
	$\kappa = 100$	$\kappa = 1000$	$\kappa = 10{,}000$	$\kappa = 100$	$\kappa = 1000$	$\kappa = 10{,}000$
Newton's	0.08	0.08	0.08	0.08	0.08	0.08
OWEA I	0.19	0.22	0.63	0.24	0.37	1.28
OWEA II	0.17	0.18	0.21	0.20	0.23	0.29

algorithm is well studied, easy to program and fast. For comparison, we choose the optimal weight exchange algorithm (OWEA) proposed in Yang, Biedermann and Tang (2013), which is among the most general and fastest algorithms.

OWEA algorithm starts with an initial design on a grid of the design space, then iterates between optimizing the weights for the current set of support points and adding a new grid point to the current support points, until the condition for optimality in general equivalence theorem is satisfied. The computing time increases as the grid size κ becomes larger. So to reduce the computing time, the authors proposed a modified algorithm. The modified algorithm starts with a coarse grid and finds the optimal design on the coarse grid. Based on that, the grid near the support points of the optimal design is refined and a more accurate optimal design is found on the finer grid. We refer to their original and modified algorithm as OWEA I and OWEA II, respectively. All algorithms are coded using SAS IML and run on a Dell Desktop (2.5 GHz and 4 Gb RAM). Comparisons are made for different grid sizes, different models and under both A- and D-optimality criterion.

First, we consider the LINEXP model given in (2.2). The parameters are set to be $\boldsymbol{\theta} = (1, 0.5, -1, 1)^T$, and the design space is $[0, 1]$. Three different grid sizes, $\kappa = 100, 1000$ and 10,000, are used for OWEA I and II; and for OWEA II, the initial coarse grid sizes are chosen to be 10, 100 and 100, respectively. The computing times are shown in Table 7. Note the grid size κ is irrelevant for the speed of Newton's algorithm.

From Table 7, we can see all three algorithms are very efficient in finding optimal designs. Newton's algorithm is at least twice as fast as the other two algorithms. The speed gain is more prominent when comparing to OWEA I, especially when the grid size κ is large.

Second, we consider a polynomial regression model given in (3.8) with $d = 6$ and $\lambda(x) = 1 - x^2$, $x \in [-1, 1]$. It has more parameters than the previous example so finding optimal designs takes longer. The results are shown in Table 8, with a similar conclusion as in the previous example.

5. Discussion. In this paper, we present a general theory for finding saturated optimal designs based on the complete class results in Yang and Stufken (2012) as well as Dette and Schorning (2013). While we focus on locally optimal designs,

TABLE 8

Computation time (seconds) for A- and D-optimal designs for polynomial regression model

	A-optimal			D-optimal		
	$\kappa = 100$	$\kappa = 1000$	$\kappa = 10{,}000$	$\kappa = 100$	$\kappa = 1000$	$\kappa = 10{,}000$
Newton's	0.17	0.17	0.17	0.14	0.14	0.14
OWEA I	0.33	0.61	3.49	0.48	1.09	4.83
OWEA II	0.34	0.48	0.89	0.44	0.72	1.35

Theorem 2.2 also applies in a multistage design setting, and we have constructed optimal two-stage designs for the Michaelis–Menten model using this approach. However, unlike in the locally optimal design case, we cannot guarantee the existence of a feasible critical point in the multistage design context, so there is no guarantee this approach always works in that case.

For E-optimality, as long as the smallest eigenvalue of the information matrix $\mathbf{M}_{\xi_E^*}$ has multiplicity 1, where ξ_E^* is the E-optimal design, we have that $\Phi_{-\infty}$ is smooth in a neighborhood of $\mathbf{M}_{\xi_E^*}$, and E-optimal designs can still be found by solving for the critical points. For nonlinear models, we find the smallest eigenvalue of $\mathbf{M}_{\xi_E^*}$ often does have multiplicity 1, but we usually do not know this ahead of time. On the other hand, we can approach E-optimal designs using Φ_p-optimal designs as $|p| \to \infty$. We can show whenever $|p| \ge -\log d / \log 0.95$, the Φ_p-optimal design has at least 95% E-efficiency. This is not a tight bound; in practice, we find a much smaller $|p|$ is enough.

We now point out models that cannot be accommodated. First, this occurs when the complete class given by Theorem 2.1 is not small enough. For example, in Dette et al. (2010), D-optimal designs for a nonlinear model with mean $\eta(x, \boldsymbol{\theta}) = \theta_1 + \theta_2 \exp(x/\theta_3)$, $x \in [L, U]$ are found to be 3-point designs with both endpoints, whereas a complete class consists of designs with at most 3 design points including only the upper endpoint as a fixed design point [Yang (2010), Theorem 3]. So the D-optimal designs are actually on the boundary of the $\mathbf{Z}$-space, hence no feasible critical points can be found, and the approach fails.

Second, the method fails when the model contains multiple covariates. In general, theoretical results are very hard to obtain for multi-covariate models, and only a couple of papers have provided some theoretical guidance. Specific to our approach, complete class results similar to Theorem 2.1 are not available. The reason is that complete class results are built upon Chebyshev systems. However, there is no satisfactory multidimensional generalization of the Chebyshev system yet. While Yang, Zhang and Huang (2011) gave complete class results for logistic and probit models with multiple covariates, the complete classes are not derived using multidimensional Chebyshev systems, and they are not small enough for our method to be applied.

APPENDIX: PROOFS

We will prove Theorems 2.2 and 2.5. Before proving Theorem 2.2, we first provide a lemma. This lemma is easier stated in terms of c, but it can be translated into x. Recall that Theorem 2.1 gives the form of a complete class. For any design ξ, we can find a design $\tilde{\xi} = \{(\tilde{c}_j, \tilde{\omega}_j)\}_{j=1}^m$ in the complete class that is noninferior ($\mathbf{M}_{\tilde{\xi}} \geq \mathbf{M}_\xi$).

In particular, for ξ specified in Lemma A.1, let $\Psi_0(c) \equiv 1$, a design $\tilde{\xi}$ can be found by solving the following nonlinear equation system [see Yang and Stufken (2012) and Dette and Schorning (2013)]:

$$\sum_i \omega_i \Psi_\ell(c_i) = \sum_j \tilde{\omega}_j \Psi_\ell(\tilde{c}_j), \qquad \ell = 0, 1, \dots, k-1, \tag{A.1}$$

where $\tilde{c}_1$ and $\tilde{c}_m$ may be fixed to be boundary points (see Lemma A.1). Multiply both sides of (A.1) by a positive constant, the equation system still holds, so we can remove the constraint of $\sum_i w_i = 1$ for ξ and allow $\sum_i w_i$ to be any positive number in the following Lemma A.1; similarly for $\tilde{\xi}$ (but we still refer to them as designs for convenience). Let $\mathbf{X} = (\mathbf{c}^T, \boldsymbol{\omega}^T)^T$ be the vector of all c_i's and ω_i's in ξ. Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be the sets of all possible vectors $\mathbf{X}$ corresponding to designs in cases (1a)$\sim$(1d) and (2) of Lemma A.1, respectively. Further, let $\mathbf{Y}$ be the vector of all $\tilde{c}_j$'s except those fixed as boundary points (if any) and all $\tilde{\omega}_j$'s in design $\tilde{\xi}$ given in the following Lemma A.1. We will define function H, $H(\mathbf{X}) = Y$, where $\mathbf{X} \in \mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2$, and show this function is smooth on $\mathbf{S}$ under certain conditions.

LEMMA A.1. *Suppose one of the conditions in Theorem 2.1 holds.*

(1a) *If* $k = 2m - 1$ *and* $\mathbf{F}(c) < 0$, *then for any design* $\xi = \{(c_i, \omega_i)\}_{i=1}^m$, $A < c_1 < \cdots < c_m \leq B, \omega_i > 0$ *for* $i \geq 1$, *there exists a noninferior design* $\tilde{\xi} = \{(\tilde{c}_j, \tilde{\omega}_j)\}_{j=1}^m$, *where* $\tilde{c}_1 = A, \tilde{\omega}_j > 0$ *for* $j \geq 1$, *that solves* (A.1).

(1b) *If* $k = 2m - 1$ *and* $\mathbf{F}(c) > 0$, *then for any design* $\xi = \{(c_i, \omega_i)\}_{i=1}^m$, $A \leq c_1 < \cdots < c_m < B, \omega_i > 0$ *for* $i \geq 1$, *there exists a noninferior design* $\tilde{\xi} = \{(\tilde{c}_j, \tilde{\omega}_j)\}_{j=1}^m$, *where* $\tilde{c}_m = B, \tilde{\omega}_j > 0$ *for* $j \geq 1$, *that solves* (A.1).

(1c) *If* $k = 2m$ *and* $\mathbf{F}(c) < 0$, *then for any design* $\xi = \{(c_i, \omega_i)\}_{i=1}^{m+1}$, $A \leq c_1 < \cdots < c_{m+1} \leq B, \omega_i > 0$ *for* $i \geq 1$, *there exists a noninferior design* $\tilde{\xi} = \{(\tilde{c}_j, \tilde{\omega}_j)\}_{j=1}^m$, *where* $\tilde{\omega}_j > 0$ *for* $j \geq 1$, *that solves* (A.1).

(1d) *If* $k = 2m - 2$ *and* $\mathbf{F}(c) > 0$, *then for any design* $\xi = \{(c_i, \omega_i)\}_{i=1}^{m-1}$, $A < c_1 < \cdots < c_{m-1} < B, \omega_i > 0$ *for* $i \geq 1$, *there exists a noninferior design* $\tilde{\xi} = \{(\tilde{c}_j, \tilde{\omega}_j)\}_{j=1}^m$, *where* $\tilde{c}_1 = A, \tilde{c}_m = B, \tilde{\omega}_j > 0$ *for* $j \geq 1$, *that solves* (A.1).

Such solution is unique under each case, hence H *is well defined on* $\mathbf{S}_1$.

(2) *For each case of* (1a)$\sim$(1d), *let* ξ *be similarly defined as above except that there is exactly one* 0 *weight and all other weights are positive. Then rewriting* ξ

in the form of $\tilde{\xi}$ *in each corresponding case solves* (A.1) *and defines* H *on* $\mathbf{S}_2$. *Moreover,* H *is smooth on* $\mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2$.

PROOF. We only prove for case (a), others being similar. First, let us consider (1a). From Lemma 1 in Yang (2010) [see also Dette and Schorning (2013), Theorem 3.1], we know that a solution to (A.1) exists with $\tilde{c}_1 = A$, $\tilde{\omega}_j > 0$, $j \geq 1$. Moreover, $\mathbf{F}(c) < 0$ implies that $\{\Psi_0, \Psi_1, \dots, \Psi_{2m-2}\}$ is a Chebyshev system [see Yang and Stufken (2012), Proposition 4], thus such solution is unique. So H is well defined on $\mathbf{S}_1$. Now we show the smoothness on $\mathbf{S}_1$.

We have $\mathbf{X} = (c_1, \dots, c_m, \omega_1, \dots, \omega_m)^T$, $\mathbf{Y} = (\tilde{c}_2, \dots, \tilde{c}_m, \tilde{\omega}_1, \dots, \tilde{\omega}_m)^T$ by definition ($\tilde{c}_1$ is excluded in $\mathbf{Y}$ since it is fixed to be A). Subtract the left-hand side from the right-hand side in (A.1), we get an equation system $G(\mathbf{X}, \mathbf{Y}) = 0$, where G is smooth. So $\mathbf{Y} = H(\mathbf{X})$ is the implicit function defined by $G(\mathbf{X}, \mathbf{Y}) = 0$. By implicit function theorem, to ensure H to be smooth, we only need the Jacobian matrix $G_{\mathbf{Y}}(\mathbf{X}, \mathbf{Y}) = \partial G(\mathbf{X}, \mathbf{Y})/\partial \mathbf{Y}$ to be nonsingular, that is,

$$
\begin{aligned}
&\det G_{\mathbf{Y}}(\mathbf{X}, \mathbf{Y}) \\
&\quad = \begin{vmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \\ \tilde{\omega}_2 \Psi_1'(\tilde{c}_2) & \cdots & \tilde{\omega}_m \Psi_1'(\tilde{c}_m) & \Psi_1(A) & \cdots & \Psi_1(\tilde{c}_m) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ \tilde{\omega}_2 \Psi_{2m-2}'(\tilde{c}_2) & \cdots & \tilde{\omega}_m \Psi_{2m-2}'(\tilde{c}_m) & \Psi_{2m-2}(A) & \cdots & \Psi_{2m-2}(\tilde{c}_m) \end{vmatrix} \\
&\quad = \left(\prod_{j=2}^{m} \tilde{w}_j \right) d(\tilde{\mathbf{c}}) \neq 0,
\end{aligned}
$$

where

$$
d(\tilde{\mathbf{c}}) = \begin{vmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ \Psi_1(A) & \cdots & \Psi_1(\tilde{c}_m) & \Psi_1'(\tilde{c}_2) & \cdots & \Psi_1'(\tilde{c}_m) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Psi_{2m-2}(A) & \cdots & \Psi_{2m-2}(\tilde{c}_m) & \Psi_{2m-2}'(\tilde{c}_2) & \cdots & \Psi_{2m-2}'(\tilde{c}_m) \end{vmatrix}. \tag{A.2}
$$

Since $\tilde{w}_j > 0$ for all $1 \leq j \leq m$, we only need to show $d(\tilde{\mathbf{c}}) \neq 0$. We first do some column manipulations to the matrix in (A.2). Subtract the first column from the second to the mth column, then for the resulting matrix, subtract the second column from the third to the mth column, continue doing this until finally subtract the $(m-1)$th column from the mth column. Because the determinant does not change during this process,

$$
d(\tilde{\mathbf{c}}) = \begin{vmatrix} \Psi_1(\tilde{c}_2) - \Psi_1(A) & \cdots & \Psi_1(\tilde{c}_m) - \Psi_1(\tilde{c}_{m-1}) & \\ \vdots & \ddots & \vdots & \mathbf{D} \\ \Psi_{2m-2}(\tilde{c}_2) - \Psi_{2m-2}(A) & \cdots & \Psi_{2m-2}(\tilde{c}_m) - \Psi_{2m-2}(\tilde{c}_{m-1}) & \end{vmatrix}, \tag{A.3}
$$

where $\mathbf{D}$ is the $(2m-2)\times(m-1)$ matrix,

$$\mathbf{D} = \begin{pmatrix} \Psi_1'(\tilde{c}_2) & \cdots & \Psi_1'(\tilde{c}_m) \\ \vdots & \ddots & \vdots \\ \Psi_{2m-2}'(\tilde{c}_2) & \cdots & \Psi_{2m-2}'(\tilde{c}_m) \end{pmatrix}.$$

Treat A in the first column of the matrix in (A.3) as a variable and fix everything else, then the determinant becomes a real-valued function of A. Using the mean value theorem, we get

$$\begin{aligned} d(\tilde{\mathbf{c}}) &= (\tilde{c}_2 - A) \\ &\times \begin{vmatrix} \Psi_1'(\hat{c}_1) & \Psi_1(\tilde{c}_3) - \Psi_1(\tilde{c}_2) & \cdots & \Psi_1(\tilde{c}_m) - \Psi_1(\tilde{c}_{m-1}) & \\ \vdots & \vdots & \ddots & \vdots & \mathbf{D} \\ \Psi_{2m-2}'(\hat{c}_1) & \Psi_{2m-2}(\tilde{c}_3) - \Psi_{2m-2}(\tilde{c}_2) & \cdots & \Psi_{2m-2}(\tilde{c}_m) - \Psi_{2m-2}(\tilde{c}_{m-1}) & \end{vmatrix}, \end{aligned} \tag{A.4}$$

where $A < \hat{c}_1 < \tilde{c}_2$. Let $\varepsilon = \operatorname{sign} d(\tilde{\mathbf{c}})$ be the sign of $d(\tilde{\mathbf{c}})$, treat $\tilde{c}_2$ in the second column of the matrix in (A.4) as a variable, and use the mean value theorem again to obtain

$$\varepsilon = \operatorname{sign} \begin{vmatrix} \Psi_1'(\hat{c}_1) & \Psi_1'(\hat{c}_2) & \cdots & \Psi_1(\tilde{c}_m) - \Psi_1(\tilde{c}_{m-1}) & \\ \vdots & \vdots & \ddots & \vdots & \mathbf{D} \\ \Psi_{2m-2}'(\hat{c}_1) & \Psi_{2m-2}'(\hat{c}_2) & \cdots & \Psi_{2m-2}(\tilde{c}_m) - \Psi_{2m-2}(\tilde{c}_{m-1}) & \end{vmatrix},$$

where $\tilde{c}_2 < \hat{c}_2 < \tilde{c}_3$. Keep on doing this, and finally get

$$\varepsilon = \operatorname{sign} \begin{vmatrix} \Psi_1'(\hat{c}_1) & \cdots & \Psi_1'(\hat{c}_{m-1}) & \Psi_1'(\tilde{c}_2) & \cdots & \Psi_1'(\tilde{c}_m) \\ \vdots & \ddots & \cdots & \cdots & \ddots & \cdots \\ \Psi_{2m-2}'(\hat{c}_1) & \cdots & \Psi_{2m-2}'(\hat{c}_{m-1}) & \Psi_{2m-2}'(\tilde{c}_2) & \cdots & \Psi_{2m-2}'(\tilde{c}_m) \end{vmatrix},$$

and $A = \tilde{c}_1 < \hat{c}_1 < \tilde{c}_2 < \hat{c}_2 < \cdots < \hat{c}_{m-1} < \tilde{c}_m$. Since $\{\Psi_1', \ldots, \Psi_{2m-2}'\}$ is a Chebyshev system, $\varepsilon \neq 0$. Hence, the Jacobian matrix is invertible, and the function H is smooth on $\mathbf{S}_1$.

Turning to case (2), without loss of generality, assume $\omega_1 = 0$, $\omega_i > 0$ for $i \geq 2$. If we can show the function $H(\mathbf{X})$ is continuous on $\mathbf{S}_2$ and its partial derivatives can be extended continuously to $\mathbf{S}_2$, then it can be proved that $H(\mathbf{X})$ is also differentiable on $\mathbf{S}_2$. So first, we prove its continuity.

To show this, for any sequence $\mathbf{X}^n = (c_1^n, \ldots, c_m^n, \omega_1^n, \ldots, \omega_m^n)^T$, $n \geq 1$, $\boldsymbol{\omega}^n > 0$ and $\mathbf{X}^n$ approaching $\mathbf{X}^0 = (c_1, \ldots, c_m, 0, \omega_2, \ldots, \omega_m)^T$, we need to show $\mathbf{Y}^n = (\tilde{c}_2^n, \ldots, \tilde{c}_m^n, \tilde{\omega}_1^n, \ldots, \tilde{\omega}_m^n)^T$ approaches $\mathbf{Y}^0 = (c_2, \ldots, c_m, 0, \omega_2, \ldots, \omega_m)^T$.

By definition, we have

$$\sum_{i=1}^{m} \omega_i^n \Psi_\ell(c_i^n) = \sum_{j=1}^{m} \tilde{\omega}_j^n \Psi_\ell(\tilde{c}_j^n), \qquad \ell = 0, \dots, 2m-2. \tag{A.5}$$

Suppose we have $\mathbf{Y}_{j_1}^n$ does not converge to $\mathbf{Y}_{j_1}^0$ for some j_1, then because $\mathbf{Y}^n$ is a bounded sequence, there exists a subsequence $\{n_t | t = 1, 2, \dots\}$ such that $\mathbf{Y}^{n_t}$ converges to some $\bar{\mathbf{Y}}^0 = (\bar{c}_2, \dots, \bar{c}_m, \bar{\omega}_1, \dots, \bar{\omega}_m)^T$ and $\bar{\mathbf{Y}}_{j_1}^0 \neq \mathbf{Y}_{j_1}^0$.

Now let $n_t \to \infty$, take the limit of (A.5) on both sides, we get

$$\sum_{i=2}^{m} \omega_i \Psi_\ell(c_i) = \sum_{j=1}^{m} \bar{\omega}_j \Psi_\ell(\bar{c}_j), \qquad \ell = 0, \dots, 2m-2. \tag{A.6}$$

Since $\{\Psi_0, \dots, \Psi_{2m-2}\}$ is a Chebyshev system and the maximum number of different support points in (A.6) is $2m-1$, (A.6) only holds if $\bar{\omega}_1 = 0$, $\bar{\omega}_i = \omega_i$, $\bar{c}_i = c_i$ for $i \geq 2$, which means $\bar{\mathbf{Y}}^0 = \mathbf{Y}^0$, leading to a contradiction.

Next, we show the partial derivatives can be extended continuously to $\mathbf{S}_2$. Using the implicit function theorem, we know

$$\frac{\partial H(\mathbf{X})}{\partial \mathbf{X}} = -G_{\mathbf{Y}}^{-1}\big(\mathbf{X}, H(\mathbf{X})\big) G_{\mathbf{X}}\big(\mathbf{X}, H(\mathbf{X})\big),$$
$$G_{\mathbf{X}}(\mathbf{X}, \mathbf{Y}) = \frac{\partial G(\mathbf{X}, \mathbf{Y})}{\partial \mathbf{X}},$$

for $\mathbf{X} \in \mathbf{S}_1$. When $\mathbf{X} \to \mathbf{X}^0$, $H(\mathbf{X}) \to H(\mathbf{X}^0)$ by continuity, hence $G_{\mathbf{Y}}(\mathbf{X}, H(\mathbf{X})) \to G_{\mathbf{Y}}(\mathbf{X}^0, H(\mathbf{X}^0))$ since $G_{\mathbf{Y}}(\mathbf{X}, \mathbf{Y})$ is continuous. Furthermore, $G_{\mathbf{Y}}(\mathbf{X}^0, H(\mathbf{X}^0))$ is nonsingular by the similar argument as previously, therefore, $G_{\mathbf{Y}}^{-1}(\mathbf{X}, H(\mathbf{X})) \to G_{\mathbf{Y}}^{-1}(\mathbf{X}^0, H(\mathbf{X}^0))$. It is easy to see $G_{\mathbf{X}}(\mathbf{X}, H(\mathbf{X})) \to G_{\mathbf{X}}(\mathbf{X}^0, H(\mathbf{X}^0))$, therefore, the derivative $\partial H(\mathbf{X})/\partial \mathbf{X} \to -G_{\mathbf{Y}}^{-1}(\mathbf{X}^0, H(\mathbf{X}^0)) \times G_{\mathbf{X}}(\mathbf{X}^0, H(\mathbf{X}^0))$, that is, the derivative can be extended continuously to $\mathbf{S}_2$. So $H(\mathbf{X})$ is differentiable on $\mathbf{S}_2$ and the partial derivatives are continuous. □

Now we are ready to prove Theorem 2.2; the proof is stated in terms of x to be consistent with the theorem.

PROOF OF THEOREM 2.2. We only prove the case where the complete class consists of designs with at most m points including L, other cases being similar. Assume the design ξ^c given by a feasible critical point is not an optimal design, and an optimal design exists as $\xi^* = \{(L, 1 - \sum_{i=2}^m \omega_i^*), \{(x_i^*, \omega_i^*)\}_{i=2}^m\}$, where $L < x_2^* < \cdots < x_m^*$ is a strictly increasing sequence (some of the weights ω_i^* may be 0 if the support size of ξ^* is less than m). We have $\Phi(\mathbf{M}_{\xi^*}) > \Phi(\mathbf{M}_{\xi^c})$. Consider the linear combination of the two designs, $\xi_\epsilon = \epsilon\xi^* + (1-\epsilon)\xi^c$, $0 \leq \epsilon \leq 1$, so

$$\xi_\epsilon = \left\{\left(L, 1-(1-\epsilon)\sum_{i=2}^{m}\omega_i^c - \epsilon\sum_{i=2}^{m}\omega_i^*\right), \{(x_i^c, (1-\epsilon)\omega_i^c)\}_{i=2}^m, \{(x_i^*, \epsilon\omega_i^*)\}_{i=2}^m\right\}.$$

By the concavity of the optimality criterion Φ, we have

$$\Phi(\mathbf{M}_{\xi_\epsilon}) \geq (1-\epsilon)\Phi(\mathbf{M}_{\xi^c}) + \epsilon\Phi(\mathbf{M}_{\xi^*}). \tag{A.7}$$

Utilizing (A.7), we can get

$$\frac{\Phi(\mathbf{M}_{\xi_\epsilon}) - \Phi(\mathbf{M}_{\xi^c})}{\epsilon} \geq \Phi(\mathbf{M}_{\xi^*}) - \Phi(\mathbf{M}_{\xi^c}) > 0. \tag{A.8}$$

Now, if we can find a series of designs with m support points, $\tilde{\xi}_\epsilon = \{(L, 1 - \sum_{i=2}^m \omega_{i,\epsilon}), \{(x_{i,\epsilon}, \omega_{i,\epsilon})\}_{i=2}^m\}$, $\epsilon \geq 0$ belongs to a neighborhood of 0, such that:

1. $\Phi(\mathbf{M}_{\tilde{\xi}_\epsilon}) \geq \Phi(\mathbf{M}_{\xi_\epsilon})$;
2. $\mathbf{Z}_\epsilon = (\mathbf{x}_\epsilon, \boldsymbol{\omega}_\epsilon)$ depends smoothly on ϵ, where $\mathbf{x}_\epsilon = (x_{2,\epsilon}, \ldots, x_{m,\epsilon})$, $\boldsymbol{\omega}_\epsilon = (\omega_{2,\epsilon}, \ldots, \omega_{m,\epsilon})$;
3. $\mathbf{Z}_0 = \mathbf{Z}^c = (\mathbf{x}^c, \boldsymbol{\omega}^c)$, thus $\tilde{\xi}_0 = \xi^c$.

Then, applying (A.8), we obtain

$$\frac{\Phi(\mathbf{M}_{\tilde{\xi}_\epsilon}) - \Phi(\mathbf{M}_{\tilde{\xi}_0})}{\epsilon} \geq \frac{\Phi(\mathbf{M}_{\xi_\epsilon}) - \Phi(\mathbf{M}_{\xi^c})}{\epsilon} \geq \Phi(\mathbf{M}_{\xi^*}) - \Phi(\mathbf{M}_{\xi^c}) > 0.$$

Because $\tilde{\xi}_\epsilon$ has $m \geq d$ support points, $\mathbf{M}_{\tilde{\xi}_\epsilon}$ must belong to PD(d). By our smoothness assumption of Φ, $\Phi(\mathbf{M}_{\tilde{\xi}_\epsilon})$ is a smooth function of ϵ. Take the limit as $\epsilon \to 0$, it gives

$$\left.\frac{\partial\Phi(\mathbf{M}_{\tilde{\xi}_\epsilon})}{\partial\epsilon}\right|_{\epsilon=0} > 0. \tag{A.9}$$

On the other hand, by our definition, $\Phi(\mathbf{M}_{\tilde{\xi}_\epsilon}) = \tilde{\Phi}(\mathbf{Z}_\epsilon)$. Applying the chain rule and using the fact that $\mathbf{Z}_0 = \mathbf{Z}^c$ is a critical point of $\tilde{\Phi}(\mathbf{Z})$, we can get

$$\left.\frac{\partial\Phi(\mathbf{M}_{\tilde{\xi}_\epsilon})}{\partial\epsilon}\right|_{\epsilon=0} = \left.\frac{\partial\tilde{\Phi}(\mathbf{Z}_\epsilon)}{\partial\epsilon}\right|_{\epsilon=0} = \left.\frac{\partial\tilde{\Phi}(\mathbf{Z})}{\partial\mathbf{Z}}\right|_{\mathbf{Z}=\mathbf{Z}_0} \left.\frac{\partial\mathbf{Z}_\epsilon}{\partial\epsilon}\right|_{\epsilon=0} = 0.$$

This contradicts with (A.9). Hence, ξ^c must be an optimal design.

To find such designs $\tilde{\xi}_\epsilon$, first, if the design ξ^* does not have new design points other than those in ξ^c, that is, $\forall 2 \leq i \leq m$, we have either $\omega_i^* = 0$ or $x_i^* \in \mathbf{x}^c$, then the design ξ_ϵ is itself a design with m support points, we can simply let $\tilde{\xi}_\epsilon = \xi_\epsilon$, and conditions $1 \sim 3$ are satisfied.

Otherwise, suppose we have $r > 0$ new design points $x^*_{i_1}, \dots, x^*_{i_r}$ introduced by ξ^*, with $\omega^*_{i_k} > 0, k = 1, \dots, r$. Let $\delta_{ii'} = 1$ if $x^c_i = x^*_{i'}$ and 0 otherwise. Rewrite the design ξ_ϵ as

$$\begin{aligned}\xi_\epsilon &= \left\{\left(L, 1-(1-\epsilon)\sum_{i=2}^m \omega^c_i - \epsilon\sum_{i=2}^m \omega^*_i\right), \left\{\left(x^c_i, (1-\epsilon)\omega^c_i + \epsilon\sum_{i'=1}^m \omega^*_{i'}\delta_{ii'}\right)\right\}_{i=2}^m\right\} \\ &\quad \cup \left\{\left(x^*_{i_k}, \epsilon\omega^*_{i_k}\right)\right\}_{k=1}^r \\ &= \left\{\left(L, \omega^{(0)}_{1,\epsilon}\right), \left\{\left(x^{(0)}_{i,\epsilon}, \omega^{(0)}_{i,\epsilon}\right)\right\}_{i=2}^m\right\} \cup \left\{\left(x^*_{i_k}, \epsilon\omega^*_{i_k}\right)\right\}_{k=1}^r,\end{aligned}$$

where the second equation simply renames the design points and design weights. It is easy to verify that conditions $2 \sim 3$ are satisfied for $\mathbf{Z}^{(0)}_\epsilon = (\mathbf{x}^{(0)}_\epsilon, \boldsymbol{\omega}^{(0)}_\epsilon) = (x^{(0)}_{2,\epsilon}, \dots, x^{(0)}_{m,\epsilon}, \omega^{(0)}_{2,\epsilon}, \dots, \omega^{(0)}_{m,\epsilon})$.

To find the desired m-point design $\tilde{\xi}_\epsilon$, we need to reduce the number of design points in a "smooth" way. We reduce one point at a time. First, consider the design $\{(x^{(0)}_{i,\epsilon}, \omega^{(0)}_{i,\epsilon})\}_{i=2}^m \cup \{(x^*_{i_1}, \epsilon\omega^*_{i_1})\}$, all the weights are positive when $0 < \epsilon < 1$, and when $\epsilon = 0$, only one weight is 0. So applying Lemma A.1 to this design we can get a new design $\{(L, \omega^{(1)}_{1,\epsilon}), \{(x^{(1)}_{i,\epsilon}, \omega^{(1)}_{i,\epsilon})\}_{i=2}^m\}$ that is noninferior, and conditions $2 \sim 3$ are satisfied for $\mathbf{Z}^{(1)}_\epsilon = (\mathbf{x}^{(1)}_\epsilon, \boldsymbol{\omega}^{(1)}_\epsilon)$, where $\boldsymbol{\omega}^{(1)}_\epsilon > 0$ for $0 \le \epsilon < 1$.

Next, we add point $x^*_{i_2}$ to $\{(x^{(1)}_{i,\epsilon}, \omega^{(1)}_{i,\epsilon})\}_{i=2}^m$ (we can always assume $x^*_{i_2}$ is a new point to $\mathbf{x}^{(1)}_\epsilon$ by taking ϵ small enough). Again, all the weights are positive when $\epsilon > 0$, and when $\epsilon = 0$, only one weight is 0. Use the same method to reduce one design point again. Keep on doing this until all r new points have been added and reduced, and we finally get $\tilde{\xi}_\epsilon = \{(L, 1 - \sum_{i=2}^m \omega^{(r)}_{i,\epsilon}), \{(x^{(r)}_{i,\epsilon}, \omega^{(r)}_{i,\epsilon})\}_{i=2}^m\}$, that is not inferior to ξ_ϵ, with the conditions $1 \sim 3$ satisfied. □

Finally, we prove Theorem 2.5, the proof is stated in terms of c for convenience.

PROOF OF THEOREM 2.5. We only consider the case of Theorem 2.1(a). First, ξ^* must belong to the complete class. Otherwise, we can find a design $\tilde{\xi}^*$ with $\mathbf{M}_{\tilde{\xi}^*} \ge \mathbf{M}_{\xi^*}$ and $\mathbf{M}_{\tilde{\xi}^*} \ne \mathbf{M}_{\xi^*}$. Because ξ^* has at least d support points, $\mathbf{M}_{\xi^*}$ is positive definite. Since Φ is strictly isotonic on PD(d), we have $\Phi(\mathbf{M}_{\tilde{\xi}^*}) > \Phi(\mathbf{M}_{\xi^*})$, which is a contradiction.

Now suppose there is another optimal design $\tilde{\xi}^*$.

(i) If $\tilde{\xi}^*$ also has at least d support points, then it also belongs to the complete class by previous arguments, and we can write $\xi^* = \{(c^*_i, \omega^*_i)\}_{i=1}^m$, $\tilde{\xi}^* = \{(\tilde{c}^*_i, \tilde{\omega}^*_i)\}_{i=1}^m$, $c^*_1 = \tilde{c}^*_1 = A$. By strict concavity, we must have $\mathbf{M}_{\xi^*} \propto \mathbf{M}_{\tilde{\xi}^*}$ since otherwise $\Phi(\alpha\mathbf{M}_{\xi^*} + (1-\alpha)\mathbf{M}_{\tilde{\xi}^*}) > \alpha\Phi(\mathbf{M}_{\xi^*}) + (1-\alpha)\Phi(\mathbf{M}_{\tilde{\xi}^*}) = \Phi(\mathbf{M}_{\xi^*})$ for all $\alpha \in (0, 1)$. Let $\mathbf{M}_{\xi^*} = \delta\mathbf{M}_{\tilde{\xi}^*}$, then $\Phi(\delta\mathbf{M}_{\tilde{\xi}^*}) = \Phi(\mathbf{M}_{\tilde{\xi}^*})$. The strict isotonicity of Φ implies $\delta = 1$, hence $\mathbf{M}_{\xi^*} = \mathbf{M}_{\tilde{\xi}^*}$ and $\mathbf{C}_{\xi^*} = \mathbf{C}_{\tilde{\xi}^*}$. Then we have (A.1) holds. Because $\mathbf{F}(c) < 0$, $\{\Psi_0, \ldots, \Psi_{2m-2}\}$ is a Chebyshev system. The maximum number of different support points in (A.1) is $2m - 1$, so (A.1) only holds if the design points and weights on two sides of the equations are equal, which means $\xi^* = \tilde{\xi}^*$.

(ii) If $\tilde{\xi}^*$ has less than d support points, let $\xi_\alpha = \alpha\xi^* + (1 - \alpha)\tilde{\xi}^*, 0 < \alpha < 1$. By concavity, ξ_α is also an optimal design, moreover, it has at least d support points. Thus following the arguments in case (i), we have $\xi_\alpha = \xi^*$, which means $\xi^* = \tilde{\xi}^*$. This contradicts with the fact that $\tilde{\xi}^*$ has less than d support points. $\square$

REFERENCES

DE LA GARZA, A. (1954). Spacing of information in polynomial regression. *Ann. Math. Statist.* **25** 123–130. MR0060777

DEMIDENKO, E. (2004). *Mixed Models: Theory and Applications*. Wiley, Hoboken, NJ. MR2077875

DEMIDENKO, E. (2006). The assessment of tumour response to treatment. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **55** 365–377. MR2224231

DETTE, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **59** 97–110. MR1436556

DETTE, H., BRETZ, F., PEPELYSHEV, A. and PINHEIRO, J. (2008). Optimal designs for dose-finding studies. *J. Amer. Statist. Assoc.* **103** 1225–1237. MR2462895

DETTE, H., KISS, C., BEVANDA, M. and BRETZ, F. (2010). Optimal designs for the emax, log-linear and exponential models. *Biometrika* **97** 513–518. MR2650755

DETTE, H., MELAS, V. B. and WONG, W. K. (2006). Locally D-optimal designs for exponential regression models. *Statist. Sinica* **16** 789–803. MR2281302

DETTE, H. and MELAS, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Ann. Statist.* **39** 1266–1281. MR2816354

DETTE, H. and SCHORNING, K. (2013). Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *Ann. Statist.* **41** 1260–1267. MR3113810

DETTE, H. and STUDDEN, W. J. (1995). Optimal designs for polynomial regression when the degree is not known. *Statist. Sinica* **5** 459–473. MR1347600

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255–262. MR0047998

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics. Pure and Applied Mathematics* **XV**. Interscience, New York. MR0204922

KIEFER, J. and WOLFOWITZ, J. (1965). On a theorem of Hoel and Levine on extrapolation designs. *Ann. Math. Statist.* **36** 1627–1655. MR0185769

LI, G. and BALAKRISHNAN, N. (2011). Optimal designs for tumor regrowth models. *J. Statist. Plann. Inference* **141** 644–654. MR2732935

PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. Wiley, New York. MR1211416

STUDDEN, W. J. (1968). Optimal designs on Tchebycheff points. *Ann. Math. Statist.* **39** 1435–1447. MR0231497

YANG, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38** 2499–2524. MR2676896

YANG, M., BIEDERMANN, S. and TANG, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *J. Amer. Statist. Assoc.* **108** 1411–1420. MR3174717

YANG, M. and STUFKEN, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37** 518–541. MR2488361

YANG, M. and STUFKEN, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *Ann. Statist.* **40** 1665–1681. MR3015039

YANG, M., ZHANG, B. and HUANG, S. (2011). Optimal designs for generalized linear models with multiple design variables. *Statist*. *Sinica* **21** 1415–1430. MR2827529

L. HU
DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602
USA
E-MAIL: pangaea@uga.edu

M. YANG
DEPARTMENT OF MATHEMATICS, STATISTICS
AND COMPUTER SCIENCES
UNIVERSITY OF ILLINOIS AT CHICAGO
CHICAGO, ILLINOIS 60607
USA
E-MAIL: myang2@uic.edu

J. STUFKEN
SCHOOL OF MATHEMATICAL
AND STATISTICAL SCIENCES
ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA 85287
USA
E-MAIL: jstufken@asu.edu