The Annals of Statistics
2012, Vol. 40, No. 3, 1665–1681
DOI: 10.1214/12-AOS992

IDENTIFYING LOCALLY OPTIMAL DESIGNS FOR NONLINEAR MODELS: A SIMPLE EXTENSION WITH PROFOUND CONSEQUENCES

BY MIN YANG[1] AND JOHN STUFKEN[2]

University of Illinois at Chicago and University of Georgia

We extend the approach in [*Ann. Statist.* **38** (2010) 2499–2524] for identifying locally optimal designs for nonlinear models. Conceptually the extension is relatively simple, but the consequences in terms of applications are profound. As we will demonstrate, we can obtain results for locally optimal designs under many optimality criteria and for a larger class of models than has been done hitherto. In many cases the results lead to optimal designs with the minimal number of support points.

1. Introduction. During the last decades nonlinear models have become a workhorse for data analysis in many applications. While there is now an extensive literature on data analysis for such models, research on design selection has not kept pace, even though there has seen a spike in activity in recent years. Identifying optimal designs for nonlinear models is indeed much more difficult than the much better studied corresponding problem for linear models. For nonlinear models results can typically only be obtained on a case-by-case basis, meaning that each combination of model, optimality criterion and objective of the experiment requires its own proof.

Another challenge is that for a nonlinear model an optimal design typically depends on the unknown parameters. This leads to the concept of locally optimal designs, which are optimal for a priori chosen values of the parameters. The designs may be poor if the choice of values is far from the true values. Where feasible, a multistage approach could help with this. A small initial design is then used to obtain some information about the parameters, and this information is used at the next stage to estimate the true parameter values and to extend the initial design in a locally optimal way to a larger design. The design at this second stage could be the final design, or there could be additional stages at which more design points are selected. The solution presented in this paper is applicable for a one-shot approach for finding a locally optimal design as well as for a multistage approach. The argument that our method can immediately be applied for the multistage approach is exactly as in Yang and Stufken (2009).

Received August 2011; revised February 2012.
[1]Supported by NSF Grants DMS-07-07013 and DMS-07-48409.
[2]Supported by NSF Grants DMS-07-06917 and DMS-10-07507.
MSC2010 subject classifications. Primary 62K05; secondary 62J12.
Key words and phrases. Locally optimal, Loewner ordering, principal submatrix, support points.

For a broader discussion on the challenges to identify optimal designs for generalized linear models, many of which apply also for other nonlinear models, we refer the reader to Khuri et al. (2006).

The work presented here is an extension of Yang and Stufken (2009), Yang (2010) and Dette and Melas (2011). The analytic approach in those papers unified and extended many of the results on locally optimal designs that were available through the so-called geometric approach. The extension in the current paper has major consequences for two reasons. First, it enables the application of the basic approach in the three earlier papers to many models for which it could until now not be used. As a result, this paper opens the door to finding locally optimal designs for models where no feasible approach was known so far. Second, for a number of models for which answers could be obtained by earlier work, the current extension enables the identification of locally optimal designs with a smaller support. This is important because it simplifies the search for optimal designs, whether by computational or analytical methods. Section 4 will illustrate the impact of our results.

The basic approach in Yang and Stufken (2009), Yang (2010) and Dette and Melas (2011), which is also adopted here, is to identify a subclass of designs with a simple format, so that for any given design ξ, there exists a design ξ^* in that subclass with $I_{\xi^*} \geq I_\xi$ under the Loewner ordering. We will refer to this subclass as a complete class for this problem. Here, I_{ξ^*} and I_ξ are information matrices for a parameter vector θ under ξ^* and ξ, respectively. Others, such as Pukelsheim (1989) have called such a class *essentially* complete, which is admittedly indeed more accurate, but also more cumbersome. When searching for a locally optimal design, for the common information-based optimality criteria, including A-, D-, E- and Φ_p-criteria, one can thus restrict consideration to this complete class, both for a one-shot or multistage approach. Also, as shown in Yang and Stufken (2009), this conclusion holds for arbitrary functions of the parameters. Ideally, the same complete class results would apply for all a priori values of the parameter vector θ. However, it turns out, as we will see in Section 4, that there are instances where complete class results hold only for certain a priori values of θ.

Yang and Stufken (2009), Yang (2010) and Dette and Melas (2011) identify small complete classes for certain models. They do so by showing that for any design ξ that is not in their complete class, there is a design ξ^* that is in the complete class such that all elements of I_{ξ^*} are the same as the corresponding elements in I_ξ, except that *one diagonal element* in I_{ξ^*} is at least as large as that in I_ξ. This guarantees of course that $I_{\xi^*} \geq I_\xi$. The contribution of this paper is that we focus on increasing a *principal submatrix* rather than just a single *diagonal element*. This allows us to obtain results for more models than could be addressed by Yang and Stufken (2009), Yang (2010) and Dette and Melas (2011), and also facilitates the identification of smaller complete classes for some models considered in these earlier papers.

In Section 2 we will present the necessary background, while the main results are featured in Section 3. The power of the proposed extension is seen through applications in Section 4. We conclude with a short discussion in Section 5.

2. Information matrix and approximate designs. Consider a nonlinear regression model for which a response variable y depends on a single regression variable x. We assume that the y's are independent and follow some exponential distribution G with mean $\eta(x,\theta)$, where θ is the $p \times 1$ parameter vector, and the values of x can be chosen by the experimenter. Typically, approximate designs are used to study optimality in this context. An approximate design ξ can be written as $\xi = \{(x_i, \omega_i), i = 1, \dots, N\}$, where $\omega_i > 0$ is the weight for design point x_i and $\sum_{i=1}^N \omega_i = 1$. It is often more convenient to present ξ as $\xi = \{(c_i, \omega_i), i = 1, \dots, N\}$, $c_i \in [A, B]$, with the c_i's obtained from the x_i's through a bijection that may depend on θ. Typically, the information matrix for θ under design ξ can be written as

$$I_\xi(\theta) = P(\theta)\left(\sum_{i=1}^N \omega_i C(\theta, c_i)\right)(P(\theta))^T, \tag{2.1}$$

where

$$C(\theta, c) = \begin{pmatrix} \Psi_{11}(c) & & & \\ \Psi_{21}(c) & \Psi_{22}(c) & & \\ \vdots & \vdots & \ddots & \\ \Psi_{p1}(c) & \Psi_{p2}(c) & \cdots & \Psi_{pp}(c) \end{pmatrix}. \tag{2.2}$$

The functions Ψ are allowed to depend on θ not just through c, but in an attempt to simplify notation we write, for example, $\Psi_{11}(c)$ rather than $\Psi_{11}(\theta, c)$. In (2.2), $C(\theta, c)$ is a symmetric matrix, and $P(\theta)$ is a $p \times p$ nonsingular matrix that depends only on θ. Some examples of (2.1) and (2.2) will be seen in Section 4.

For some p_1, $1 \le p_1 < p$, we partition $C(\theta, c)$ as

$$C(\theta, c) = \begin{pmatrix} C_{11}(c) & C_{21}^T(c) \\ C_{21}(c) & C_{22}(c) \end{pmatrix}. \tag{2.3}$$

Here, $C_{22}(c)$ is the lower $p_1 \times p_1$ principal submatrix of $C(\theta, c)$, that is,

$$C_{22}(c) = \begin{pmatrix} \Psi_{p-p_1+1, p-p_1+1}(c) & \cdots & \Psi_{p-p_1+1, p}(c) \\ \vdots & \ddots & \vdots \\ \Psi_{p, p-p_1+1}(c) & \cdots & \Psi_{pp}(c) \end{pmatrix}. \tag{2.4}$$

In the context of local optimality, if designs $\xi = \{(c_i, \omega_i), i = 1, \dots, N\}$ and $\tilde{\xi} = \{(\tilde{c}_j, \tilde{\omega}_j), j = 1, \dots, \tilde{N}\}$ satisfy $\sum_{i=1}^N \omega_i C(\theta, c_i) \le \sum_{i=1}^{\tilde{N}} \tilde{\omega}_i C(\theta, \tilde{c}_i)$, then it follows from (2.1) that $I_\xi(\theta) \le I_{\tilde{\xi}}(\theta)$. Hence, $I_\xi(\theta) \le I_{\tilde{\xi}}(\theta)$ follows if it holds that

$$\sum_{i=1}^N \omega_i C_{11}(c_i) = \sum_{i=1}^{\tilde{N}} \tilde{\omega}_i C_{11}(\tilde{c}_i),$$

$$\sum_{i=1}^{N} \omega_i C_{12}(c_i) = \sum_{i=1}^{\tilde{N}} \tilde{\omega}_i C_{12}(\tilde{c}_i) \quad \text{and} \tag{2.5}$$

$$\sum_{i=1}^{N} \omega_i C_{22}(c_i) \leq \sum_{i=1}^{\tilde{N}} \tilde{\omega}_i C_{22}(\tilde{c}_i).$$

This is what we explore in this paper. Note that this is more general than Yang and Stufken (2009), Yang (2010) and Dette and Melas (2011), where $p_1 = 1$. We develop a theoretical framework for general values of p_1.

3. Main results. Following Karlin and Studden (1966) and Dette and Melas (2011), a set of $k+1$ real-valued continuous functions $u_0, \dots, u_k$ defined on an interval $[A, B]$ is called a Chebyshev system on $[A, B]$ if

$$\begin{vmatrix} u_0(z_0) & u_0(z_1) & \cdots & u_0(z_k) \\ u_1(z_0) & u_1(z_1) & \cdots & u_1(z_k) \\ \vdots & \vdots & \ddots & \vdots \\ u_k(z_0) & u_k(z_1) & \cdots & u_k(z_k) \end{vmatrix} \tag{3.1}$$

is strictly positive whenever $A \leq z_0 < z_1 < \cdots < z_k \leq B$.

Along the lines of Yang (2010), we select a maximal set of linearly independent nonconstant functions from the Ψ functions that appear in the first $p - p_1$ columns of the matrix $C(\theta, c)$ defined in (2.2), and rename the selected functions as $\Psi_1, \dots, \Psi_{k-1}$. For a given nonzero $p_1 \times 1$ vector Q, let

$$\Psi_k^Q = Q^T C_{22}(c) Q, \tag{3.2}$$

where $C_{22}(c)$ is as defined in (2.4).

For $\Psi_0 = 1, \Psi_1, \dots, \Psi_{k-1}$ and $C_{22}(c)$, we will say that a set of n_1 pairs (c_i, ω_i) is dominated by a set of n_2 pairs $(\tilde{c}_i, \tilde{\omega}_i)$ if

$$\sum_i \omega_i \Psi_l(c_i) = \sum_i \tilde{\omega}_i \Psi_l(\tilde{c}_i), \qquad l = 0, 1, \dots, k-1; \tag{3.3}$$

$$\sum_i \omega_i \Psi_k^Q(c_i) < \sum_i \tilde{\omega}_i \Psi_k^Q(\tilde{c}_i) \qquad \text{for every nonzero vector } Q, \tag{3.4}$$

where the summations on the left-hand sides are over the n_1 subscripts for the pairs (c_i, ω_i) and those on the right-hand sides over the n_2 subscripts for the pairs $(\tilde{c}_i, \tilde{\omega}_i)$.

The following two lemmas provide the basic tools for the main results. We point out that the pairs (c_i, ω_i) in these lemmas need not form a design; in particular, the ω_i's need not add to 1.

LEMMA 1. *For the functions* $\Psi_0 = 1, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k^Q$ *defined on an interval* $[A, B]$, *suppose that either*

$$\{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}\} \text{ and } \{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k^Q\} \tag{3.5}$$
$$\text{form Chebyshev systems for every nonzero vector } Q$$

or

$$\{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}\} \text{ and } \{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}, -\Psi_k^Q\} \tag{3.6}$$
$$\text{form Chebyshev systems for every nonzero vector } Q.$$

Then the following conclusions hold:

(a) *For* $k = 2n-1$, *if* (3.5) *holds, then for any set* $S_1 = \{(c_i, \omega_i) : \omega_i > 0, i = 1, \ldots, n\}$ *with* $A \le c_1 < \cdots < c_n < B$, *there exists a set* $S_2 = \{(\tilde{c}_i, \tilde{\omega}_i) : \tilde{\omega}_i > 0, i = 1, \ldots, n\}$ *with* $c_1 < \tilde{c}_1 < c_2 < \cdots < \tilde{c}_{n-1} < c_n < \tilde{c}_n = B$, *such that* S_1 *is dominated by* S_2.

(b) *For* $k = 2n-1$, *if* (3.6) *holds, then for any set* $S_1 = \{(c_i, \omega_i) : \omega_i > 0, i = 1, \ldots, n\}$ *with* $A < c_1 < \cdots < c_n \le B$, *there exists a set* $S_2 = \{(\tilde{c}_i, \tilde{\omega}_i) : \tilde{\omega}_i > 0, i = 0, \ldots, n-1\}$ *with* $A = \tilde{c}_0 < c_1 < \tilde{c}_1 < c_2 < \cdots < \tilde{c}_{n-1} < c_n$, *such that* S_1 *is dominated by* S_2.

(c) *For* $k = 2n$, *if* (3.5) *holds, then for any set* $S_1 = \{(c_i, \omega_i) : \omega_i > 0, i = 1, \ldots, n\}$ *with* $A < c_1 < \cdots < c_n < B$, *there exists a set* $S_2 = \{(\tilde{c}_i, \tilde{\omega}_i) : \tilde{\omega}_i > 0, i = 0, \ldots, n\}$ *with* $A = \tilde{c}_0 < c_1 < \tilde{c}_1 < \cdots < c_n < \tilde{c}_n = B$, *such that* S_1 *is dominated by* S_2.

(d) *For* $k = 2n$, *if* (3.6) *holds, then for any set* $S_1 = \{(c_i, \omega_i), \omega_i > 0, i = 1, \ldots, n+1$ *with* $A \le c_1 < \cdots < c_{n+1} \le B$, *there exists a set* $S_2 = \{(\tilde{c}_i, \tilde{\omega}_i) : \tilde{\omega}_i > 0, i = 1, \ldots, n\}$ *with* $c_1 < \tilde{c}_1 < \cdots < c_n < \tilde{c}_n < c_{n+1}$, *such that* S_1 *is dominated by* S_2.

PROOF. Since the proof is similar for all parts, we only provide a proof for part (a).

Let S_1 be as in part (a). First consider the special case that $Q = (1, 0, \ldots, 0)^T$. By (1a) of Therorem 3.1 in Dette and Melas (2011), there exists a set of at most n pairs $(\tilde{c}_i, \tilde{\omega}_i)$ with one of the points equal to B so that (3.3) and (3.4) hold for this Q. By part (a) of Proposition 1 in the Appendix, the number of distinct points with $\tilde{\omega}_i > 0$ must then be exactly n. Thus we have $\tilde{c}_1 < \cdots < \tilde{c}_n = B$, and the c_i's and $\tilde{c}_i$'s must alternate by part (b) of Proposition 1. The result follows now for an arbitrary nonzero Q by applying Proposition 2 in the Appendix and using (3.5) and (3.4). □

Lemma 2 partially extends Lemma 1 by observing that larger sets S_1 than in Lemma 1 are also dominated by sets S_2 as in that lemma.

LEMMA 2. *With the same notation and assumptions as in Lemma 1, let* $S_1 = \{(c_i, \omega_i) : \omega_i > 0, A \le c_i \le B, i = 1, \dots, N\}$, *where* $N \ge n$ *for cases* (a), (b), *and* (c) *of Lemma 1, and* $N \ge n+1$ *for case* (d). *Then the following conclusions hold*:

(a) *For* $k = 2n-1$, *if* (3.5) *holds, then* S_1 *is dominated by a set* S_2 *of size* n *that includes* B *as one of the points.*

(b) *For* $k = 2n-1$, *if* (3.6) *holds, then* S_1 *is dominated by a set* S_2 *of size* n *that includes* A *as one of the points.*

(c) *For* $k = 2n$, *if* (3.5) *holds, then* S_1 *is dominated by a set* S_2 *of size* $n+1$ *that includes both* A *and* B *as points.*

(d) *For* $k = 2n$, *if* (3.6) *holds, then* S_1 *is dominated by a set* S_2 *of size* n.

PROOF. The results follow by application of Lemma 1. For example, for case (a), if $N = n$, the result follows directly from Lemma 1. If $N > n$, we start with the points $c_1 < c_2 < \cdots < c_N$ in S_1. Using Lemma 1, we obtain points $c_1, \dots, c_{N-n}, \tilde{c}_{N-n+1}, \dots, \tilde{c}_N = B$ in a set $\tilde{S}_1$ that dominates S_1. Using Lemma 1 again on the n largest points other than $\tilde{c}_N$ in $\tilde{S}_1$, we move one more point to B, obtaining a new set with $N-1$ points that dominates $\tilde{S}_1$. Continue until the size of the set is reduced to n; this is the desired set S_2. $\square$

The first main result is an immediate consequence of Lemma 2.

THEOREM 1. *For a regression model with a single regression variable* x, *suppose that the information matrix* $C(\theta, c)$ *can be written as in* (2.1) *for* $c \in [A, B]$. *Partitioning the information matrix as in* (2.3), *let* $\Psi_1, \dots, \Psi_{k-1}$ *be a maximum set of linearly independent nonconstant* Ψ *functions in the first* $p - p_1$ *columns of* $C(\theta, c)$. *Define* Ψ_k^Q *as in* (3.2). *Suppose that either* (3.5) *or* (3.6) *in Lemma 1 holds. Then the following complete class results hold*:

(a) *For* $k = 2n-1$, *if* (3.5) *holds, the designs with at most* n *support points, including* B, *form a complete class.*

(b) *For* $k = 2n-1$, *if* (3.6) *holds, the designs with at most* n *support points, including* A, *form a complete class.*

(c) *For* $k = 2n$, *if* (3.5) *holds, the designs with at most* $n+1$ *support points, including both* A *and* B, *form a complete class.*

(d) *For* $k = 2n$, *if* (3.6) *holds, the designs with at most* n *support points form a complete class.*

Note that if (3.3) holds for $\Psi_l(c)$, $l = 1, \dots, k-1$, then the same is true if we replace one or more of the Ψ_l's by $-\Psi_l$. Therefore, if (3.5) or (3.6) do not hold for the original Ψ_l's, conclusions in Theorem 1 would still be valid if (3.5) and (3.6) hold after multiplying one or more of the Ψ_l's, $l = 1, \dots, k-1$, by -1.

While Theorem 1 is very powerful, applying it directly may not be easy. The next result, which utilizes a generalization of a tool in Yang (2010), will lead to a condition that is easier to verify. Using the notation of Theorem 1, define functions $f_{l,t}$, $1 \le t \le k$; $t \le l \le k$ as follows:

$$f_{l,t}(c) = \begin{cases} \Psi_l'(c), & \text{if } t = 1, l = 1, \dots, k-1, \\ C_{22}'(c), & \text{if } t = 1, l = k, \\ \left(\dfrac{f_{l,t-1}(c)}{f_{t-1,t-1}(c)}\right)', & \text{if } 2 \le t \le k, t \le l \le k. \end{cases} \tag{3.7}$$

The following lower triangular matrix contains all of these functions, and suggest an order in which to compute them:

$$\begin{pmatrix} f_{1,1} = \Psi_1' & & & & \\ f_{2,1} = \Psi_2' & f_{2,2} = (\frac{f_{2,1}}{f_{1,1}})' & & & \\ f_{3,1} = \Psi_3' & f_{3,2} = (\frac{f_{3,1}}{f_{1,1}})' & f_{3,3} = (\frac{f_{3,2}}{f_{2,2}})' & & \\ \vdots & \vdots & \vdots & \ddots & \\ f_{k,1} = C_{22}' & f_{k,2} = (\frac{f_{k,1}}{f_{1,1}})' & f_{k,3} = (\frac{f_{k,2}}{f_{2,2}})' & \vdots & f_{k,k} = (\frac{f_{k,k-1}}{f_{k-1,k-1}})' \end{pmatrix}. \tag{3.8}$$

Note that, for $p_1 \ge 2$, the functions in the last row are matrix functions, which is a key difference with Yang (2010). The derivatives of matrices in (3.7) are element-wise derivatives. For the next result, we will make the following assumptions:

(i) All functions Ψ in the information matrix $C(\theta, c)$ are at least kth order differentiable on (A, B).

(ii) For $1 \le l \le k-1$, the functions $f_{l,l}(c)$ have no roots in $[A, B]$.

For ease of notation, in the remainder we will write $f_{l,l}$ instead of $f_{l,l}(c)$, and $f_{l,l} > 0$ means that $f_{l,l}(c) > 0$ for all $c \in [A, B]$. This also applies for $l = k$, in which case it means that the matrix $f_{k,k}$ is positive definite for all $c \in [A, B]$.

THEOREM 2. *For a regression model with a single regression variable x, let $c \in [A, B]$, $C(\theta, c)$, $\Psi_1, \dots, \Psi_{k-1}$ and Ψ_k^Q be as in Theorem 1. For the functions $f_{l,l}$ in (3.7), define $F(c) = \prod_{l=1}^k f_{l,l}$, $c \in [A, B]$. Suppose that either $F(c)$ or $-F(c)$ is positive definite for all $c \in [A, B]$. Then the following complete class results hold*:

(a) *For $k = 2n-1$, if $F(c) > 0$, the designs with at most n support points, including B, form a complete class.*

(b) *For $k = 2n-1$, if $-F(c) > 0$, the designs with at most n support points, including A, form a complete class.*

(c) *For $k = 2n$, if $F(c) > 0$, the designs with at most $n+1$ support points, including both A and B, form a complete class.*

(d) *For $k = 2n$, if $-F(c) > 0$, the designs with at most n support points form a complete class.*

PROOF. We only present the proof for case (a) since the other cases are similar. For any nonzero vector Q, $Q^T F(c) Q > 0$ for all $c \in [A, B]$. Among all $f_{l,l}$, $l = 1, \dots, k-1$, and $Q^T f_{k,k} Q$, suppose that a of them are negative. Let $1 \le l_1 < \cdots < l_a \le k$ denote the subscripts for these negative terms, and note that a must be even. Note also that the labels $l_1 < \cdots < l_a$ do not depend on the choice of the vector Q since $f_{1,1}, \dots, f_{k-1,k-1}$ do not depend on Q. Finally, note that for any l with $1 \le l \le k-1$, if we replace $\Psi_l(c)$ by $-\Psi_l(c)$, then the signs of $f_{l,l}$ and $f_{l+1,l+1}$ are switched while all others remain unchanged.

We now change some of the Ψ_l's to $-\Psi_l$. This is done for those l that satisfy $l_{2b-1} \le l < l_{2b}$ for some value of $b \in \{1, \dots, a/2\}$. Denote the new Ψ-functions by $\{1, \widehat{\Psi}_1, \dots, \widehat{\Psi}_k^Q\}$. Notice that $\widehat{\Psi}_k^Q = \Psi_k^Q$. From the last observation in the previous paragraph, it is easy to check that $f_{l,l} > 0$, $l = 1, \dots, k$, for the functions $f_{l,l}$ that correspond to this new set of $\widehat{\Psi}$-functions. By Proposition 4 in the Appendix, $\{1, \widehat{\Psi}_1, \dots, \widehat{\Psi}_{k-1}\}$ and $\{1, \widehat{\Psi}_1, \dots, \widehat{\Psi}_{k-1}, \widehat{\Psi}_k^Q\}$ are Chebyshev systems on $[A, B]$, regardless of the choice for $Q \neq 0$. The result follows now from case (a) of Theorem 1 and the observation immediately after Theorem 1. $\square$

For case (a) in Theorem 2, the value of A in the interval $[A, B]$ is allowed to be $-\infty$. In this situation, for any given design ξ, we can choose $A = \min_i c_i$, and the conclusion of the theorem holds. Similarly, B can be ∞ in case (b), and the interval can be unbounded at either side for case (d).

As noted at the end of Section 2, the results in Yang and Stufken (2009), Yang (2010), and Dette and Melas (2011) correspond to $p_1 = 1$. The extension in this paper allows the choice of larger values of p_1 where feasible. Larger values of p_1 lead to designs with smaller support sizes. The reason for this is that the value of k in Theorems 1 and 2 corresponds to the number of equations in (3.3). For a particular model, this number is smaller for larger p_1. Since the support size of the designs is roughly half the value of k, the support size is smaller for larger values of p_1.

We will provide some examples of the application of Theorems 1 and 2 in the next section, and will offer some further thoughts on the ease of their application in Section 5.

4. Applications. Whether the model is for continuous or discrete data, with homogeneous or heterogeneous errors, Theorems 1 and 2 can be applied as long as the information matrix can be written as in (2.1). As the examples in this section will show, in many cases the result of the theorem facilitates the determination of complete classes with the minimal number of support points.

4.1. *Exponential regression models.* Dette, Melas and Wong (2006) studied exponential regression models, which can be written as

$$Y_i = \sum_{l=1}^{L} a_l e^{-\lambda_l x_i} + \varepsilon_i, \tag{4.1}$$

where the ε_i's are i.i.d. with mean 0 and variance σ^2, and $x_i \in [U, V]$ is the value of the regression variable to be selected by the experimenter. Here $\theta = (a_1, \dots, a_L, \lambda_1, \dots, \lambda_L)^T$, with $a_l \neq 0$, $l = 1, \dots, L$, and $0 < \lambda_1 < \cdots < \lambda_L$. For $L = 2$, they showed that there is a D-optimal design for $\theta = (a_1, a_2, \lambda_1, \lambda_2)^T$ based on four points, including the lower limit U. Further, for $L = 3$ and $\lambda_2 = (\lambda_1 + \lambda_3)/2$, they showed that there is a D-optimal design for θ based on six points, again including the lower limit U. By using Theorem 2, we will show that similar conclusions are possible for other optimality criteria, including A- and E-optimality, and other functions of interest for many a priori values of θ.

For $L = 2$, the results in Yang (2010) can be used to obtain a complete class of designs with at most five points. We can do better with Theorem 2. The information matrix for $\theta = (a_1, a_2, \lambda_1, \lambda_2)^T$ under design $\{(x_i, \omega_i), i = 1, \dots, N\}$ can be written in the form of (2.1) with $P(\theta) = \mathrm{diag}(1, 1, \frac{a_1}{\lambda_2 - \lambda_1}, \frac{a_2}{\lambda_2 - \lambda_1})$ and

$$(4.2)\quad C(\theta, c) = \begin{pmatrix} c^\lambda & & & \\ c^{\lambda+1} & c^{\lambda+2} & & \\ \log(c)c^\lambda & \log(c)c^{\lambda+1} & \log^2(c)c^\lambda & \\ \log(c)c^{\lambda+1} & \log(c)c^{\lambda+2} & \log^2(c)c^{\lambda+1} & \log^2(c)c^{\lambda+2} \end{pmatrix},$$

where $c = e^{-(\lambda_2 - \lambda_1)x}$ and $\lambda = \frac{2\lambda_1}{\lambda_2 - \lambda_1}$. Let $\Psi_1(c) = c^\lambda$, $\Psi_2(c) = \log(c)c^\lambda$, $\Psi_3(c) = c^{\lambda+1}$, $\Psi_4(c) = \log(c)c^{\lambda+1}$, $\Psi_5(c) = c^{\lambda+2}$, $\Psi_6(c) = \log(c)c^{\lambda+2}$ and

$$C_{22}(c) = \begin{pmatrix} \log^2(c)c^\lambda & \log^2(c)c^{\lambda+1} \\ \log^2(c)c^{\lambda+1} & \log^2(c)c^{\lambda+2} \end{pmatrix}.$$

Then $f_{1,1} = \lambda c^{\lambda-1}$, $f_{2,2} = \frac{1}{c}$, $f_{3,3} = \frac{\lambda+1}{\lambda}$, $f_{4,4} = \frac{1}{c}$, $f_{5,5} = \frac{4(\lambda+2)}{\lambda+1}$, $f_{6,6} = \frac{1}{c}$ and

$$f_{7,7}(c) = \begin{pmatrix} \dfrac{2\lambda}{(\lambda+2)c^3} & \dfrac{\lambda+1}{2(\lambda+2)c^2} \\ \dfrac{\lambda+1}{2(\lambda+2)c^2} & \dfrac{2}{c} \end{pmatrix}.$$

Note that $c > 0$ and $\lambda > 0$, so that $F(c)$ is positive definite if $|f_{7,7}(c)| > 0$. This is equivalent to $15\lambda^2 + 30\lambda - 1 > 0$, which is satisfied when $\frac{\lambda_2}{\lambda_1} < \frac{\sqrt{960}+30}{\sqrt{960}-30}$. Thus, by (a) of Theorem 2, we have the following result.

THEOREM 3. *For Model* (4.1) *with* $L = 2$, *if*

$$\frac{\lambda_2}{\lambda_1} < \frac{\sqrt{960} + 30}{\sqrt{960} - 30} \approx 61.98,$$

then the designs with at most four points, *including the lower limit* U, *form a complete class*.

For $L = 3$ and $2\lambda_2 = \lambda_1 + \lambda_3$, the information matrix for $\theta = (a_1, a_2, a_3, \lambda_1, \lambda_2, \lambda_3)^T$ under design $\{(x_i, \omega_i), i = 1, \dots, N\}$ can be written in the form of (2.1) with

$P(\theta) = \text{diag}(1, 1, 1, \frac{a_1}{\lambda_2-\lambda_1}, \frac{a_2}{\lambda_2-\lambda_1}, \frac{a_3}{\lambda_2-\lambda_1})$ and

$$C(\theta, c) = \begin{pmatrix} c^{\lambda} & & & & & \\ c^{\lambda+1} & c^{\lambda+2} & & & & \\ c^{\lambda+2} & c^{\lambda+3} & c^{\lambda+4} & & & \\ \log(c)c^{\lambda} & \log(c)c^{\lambda+1} & \log(c)c^{\lambda+2} & \log^2(c)c^{\lambda} & & \\ \log(c)c^{\lambda+1} & \log(c)c^{\lambda+2} & \log(c)c^{\lambda+3} & \log^2(c)c^{\lambda+1} & \log^2(c)c^{\lambda+2} & \\ \log(c)c^{\lambda+2} & \log(c)c^{\lambda+3} & \log(c)c^{\lambda+4} & \log^2(c)c^{\lambda+2} & \log^2(c)c^{\lambda+3} & \log^2(c)c^{\lambda+4} \end{pmatrix}, \tag{4.3}$$

where $c = e^{-(\lambda_2-\lambda_1)x}$ and $\lambda = \frac{2\lambda_1}{\lambda_2-\lambda_1}$. Let $\Psi_{2l-1}(c) = c^{\lambda+l-1}$ and $\Psi_{2l}(c) = \log(c)c^{\lambda+l-1}$, $l = 1, \dots, 5$, and let

$$C_{22}(c) = \begin{pmatrix} \log^2(c)c^{\lambda} & \log^2(c)c^{\lambda+1} & \log^2(c)c^{\lambda+2} \\ \log^2(c)c^{\lambda+1} & \log^2(c)c^{\lambda+2} & \log^2(c)c^{\lambda+3} \\ \log^2(c)c^{\lambda+2} & \log^2(c)c^{\lambda+3} & \log^2(c)c^{\lambda+4} \end{pmatrix}.$$

Then $f_{1,1} = \lambda c^{\lambda-1}$, $f_{2l,2l} = \frac{1}{c}$, $l = 1, 2, 3, 4, 5$, $f_{2l+1,2l+1} = \frac{l^2(\lambda+l)}{\lambda+l-1}$, $l = 1, 2, 3, 4$, and

$$f_{11,11}(c) = \begin{pmatrix} \dfrac{2\lambda}{(\lambda+4)c^5} & \dfrac{\lambda+1}{8(\lambda+4)c^4} & \dfrac{\lambda+2}{18(\lambda+4)c^3} \\ \dfrac{\lambda+1}{8(\lambda+4)c^4} & \dfrac{\lambda+2}{18(\lambda+4)c^3} & \dfrac{\lambda+3}{8(\lambda+4)c^2} \\ \dfrac{\lambda+2}{18(\lambda+4)c^3} & \dfrac{\lambda+3}{8(\lambda+4)c^2} & \dfrac{2}{c} \end{pmatrix}.$$

Again, $c > 0$ and $\lambda > 0$, so that $F(c)$ is positive definite if $|(f_{11,11}(c))|$ and its leading principal minors are positive. This is equivalent to

$$\begin{aligned} 1505\lambda^3 + 9030\lambda^2 + 11499\lambda - 1082 &> 0, \\ 55\lambda^2 + 110\lambda - 9 &> 0, \\ 1295\lambda^2 + 5180\lambda - 4 &> 0 \\ \text{and} \quad 55\lambda^2 + 330\lambda + 431 &> 0. \end{aligned} \tag{4.4}$$

Simple computation shows that this holds for $\frac{\lambda_2}{\lambda_1} < 23.72$ (or, equivalently, $\frac{\lambda_3}{\lambda_1} < 46.45$). By Theorem 2, we have the following result.

THEOREM 4. *For model* (4.1) *with* $L = 3$ *and* $2\lambda_2 = \lambda_1 + \lambda_3$, *if* $\frac{\lambda_2}{\lambda_1} < 23.72$, *then the designs with at most six points, including the lower limit* U, *form a complete class.*

4.2. *LINEXP model.* Demidenko (2006) proposed a model referred to as the LINEXP model to describe tumor growth delay and regrowth. The natural logarithm of the tumor volume is modeled as

$$Y_i = \alpha + \gamma x_i + \beta(e^{-\delta x_i} - 1) + \varepsilon_i, \tag{4.5}$$

with independent $\varepsilon_i \sim N(0, \sigma^2)$ and $x_i \in [U, V]$ as the value of the single regression variable, which in this case refers to time. Here $\theta = (\alpha, \gamma, \beta, \delta)^T$ is the parameter vector, where α is the baseline logarithm of the tumor volume, γ is the final growth rate and δ is the rate at which killed cells get washed out. The size of the parameter β relative to γ/δ determines whether regrowth is monotonic ($\beta < \gamma/\delta$) or not. Li and Balakrishnan (2011) recently studied this model and showed that a D-optimal design for θ can be based on four points, including U and V. We will now show that Theorem 2 extends this conclusion to other optimality criteria and functions of interest.

The information matrix for θ under design $\{(x_i, \omega_i), i = 1, \dots, N\}$ can be written in the form of (2.1) with

$$P(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -\delta & 0 & 0 \\ 0 & 0 & 0 & \delta/\beta \end{pmatrix}^{-1} \quad \text{and}$$
$$C(\theta, c) = \begin{pmatrix} 1 & & & \\ e^c & e^{2c} & & \\ c & ce^c & c^2 & \\ ce^c & ce^{2c} & c^2e^c & c^2e^{2c} \end{pmatrix}, \tag{4.6}$$

where $c = -\delta x$. With a proper choice of Ψ functions, it can be shown that the result in Yang (2010) yields a complete class of designs with at most five points, including U and V. We can again do better with Theorem 2.

Define $\Psi_1(c) = c$, $\Psi_2(c) = e^c$, $\Psi_3(c) = ce^c$, $\Psi_4(c) = e^{2c}$, $\Psi_5(c) = ce^{2c}$ and

$$C_{22}(c) = \begin{pmatrix} c^2 & c^2e^c \\ c^2e^c & c^2e^{2c} \end{pmatrix}.$$

This yields $f_{1,1} = 1$, $f_{2,2} = e^c$, $f_{3,3} = 1$, $f_{4,4} = 4e^c$, $f_{5,5} = 1$ and

$$f_{6,6}(c) = \begin{pmatrix} 2e^{-2c} & e^{-c}/2 \\ e^{-c}/2 & 2 \end{pmatrix}.$$

Clearly $F(c)$ is a positive definite matrix. Therefore, by part (c) of Theorem 2, we reach the following conclusion.

THEOREM 5. *For the LINEXP model* (4.5), *the designs with at most four points, including U and V, form a complete class.*

4.3. *Double-exponential regrowth model.* Demidenko (2004), using a two-compartment model, developed a double-exponential regrowth model to describe the dynamics of post-irradiated tumors. The model can be written as

$$Y_i = \alpha + \ln[\beta e^{\nu x_i} + (1 - \beta)e^{-\phi x_i}] + \varepsilon_{ij}, \tag{4.7}$$

with independent $\varepsilon_i \sim N(0, \sigma^2)$ and $x_i \in [U, V]$ again as the value for the variable time. Here $\theta = (\alpha, \beta, \nu, \phi)^T$ is the parameter vector, where α is the logarithm of the initial tumor volume, $0 < \beta < 1$ is the proportional contribution of the first compartment and ν and ϕ are cell proliferation and death rates.

Using Chebyshev systems and an equivalence theorem, Li and Balakrishnan (2011) showed that a D-optimal design for θ can be based on four points including U and V. Theorem 1 allows us to extend this result to a complete class result, thereby covering many other optimality criteria and any functions of interest.

The information matrix for θ under design $\{(x_i, \omega_i), i = 1, \dots, N\}$ is of the form (2.1) with

$$P(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1-\beta & 0 & 0 \\ 0 & 0 & 1/\beta & 0 \\ 0 & 0 & 0 & -1/(1-\beta) \end{pmatrix}^{-1}$$

and with $C(\theta, x)$ a 4×4 matrix as in (2.2), where $\Psi_{11} = 1$, $\Psi_{21} = e^{\nu x}/g(x)$, $\Psi_{22} = e^{2\nu x}/g^2(x)$, $\Psi_{31} = xe^{\nu x}/g(x)$, $\Psi_{32} = xe^{2\nu x}/g^2(x)$, $\Psi_{33} = x^2 e^{2\nu x}/g^2(x)$, $\Psi_{41} = xe^{-\phi x}/g(x)$, $\Psi_{42} = xe^{(\nu-\phi)x}/g^2(x)$, $\Psi_{43} = x^2 e^{(\nu-\phi)x}/g^2(x)$ and $\Psi_{44} = x^2 e^{-2\phi x}/g^2(x)$. Here, $g(x) = \beta e^{\nu x} + (1-\beta)e^{-\phi x}$. Note that Ψ_{42} can be written as a linear combination of Ψ_{31} and Ψ_{32}. We can apply Theorem 1 if we can show that both $\{1, \Psi_{21}, \Psi_{22}, \Psi_{41}, -\Psi_{31}, \Psi_{32}\}$ and $\{1, \Psi_{21}, \Psi_{22}, \Psi_{41}, -\Psi_{31}, \Psi_{32}, Q^T C_{22}(x) Q\}$ are Chebyshev systems for any nonzero vector Q, where $C_{22}(x) = \begin{pmatrix} \Psi_{33} & \Psi_{43} \\ \Psi_{43} & \Psi_{44} \end{pmatrix}$.

Rather than do this directly, we first simplify the problem. We multiply each of the Ψ's by the positive function $e^{2\phi x} g(x)^2$, which preserves the Chebyshev system property. After further simplifications by replacing some of the resulting functions by independent linear combinations of these functions, which also preserves the Chebyshev system property, we arrive at the systems $\{1, e^{(\nu+\phi)x}, e^{2(\nu+\phi)x}, x, -xe^{(\nu+\phi)x}, xe^{2(\nu+\phi)x}\}$ and $\{1, e^{(\nu+\phi)x}, e^{2(\nu+\phi)x}, x, -xe^{(\nu+\phi)x}, xe^{2(\nu+\phi)x}, g^2(x) e^{2\phi x} Q^T C_{22}(x) Q\}$. It suffices to show that these are Chebyshev systems for any nonzero vector Q, which follows from Proposition 4 if we show that $f_{l,l} > 0$, $l = 1, \dots, 6$, for the latter system. It can be shown that $f_{1,1} = f_{2,2}/2 = 2f_{4,4} = f_{5,5}/4 = ae^{ax}$, $f_{3,3} = e^{-2ax}$ and $f_{6,6} = \begin{pmatrix} 2 & e^{-ax}/2 \\ e^{-ax}/2 & 2e^{-2ax} \end{pmatrix}$, where $a = \nu + \phi$. Thus both systems are Chebyshev systems, and by part (c) of Theorem 1, we reach the following conclusion.

THEOREM 6. *For the double-exponential regrowth model* (4.7), *the designs with at most four points, including U and V, form a complete class.*

5. Discussion. We have given a powerful extension of the result in Yang (2010) that has potential for providing a small complete class of designs whenever the information matrix can be written as in (2.1). Irrespective of the optimality criterion (provided that it does not violate the Loewner ordering) and of the function of θ that is of interest, the search for an optimal design can be restricted to the small complete class. As the examples in Section 4 show, the results lead us to conclusions that were not possible using the results in Yang (2010) and Dette and Melas (2011).

As already pointed out, direct application of Theorem 1 may not be easy. Section 4.3 shows some tricks that can be useful when using Theorem 1. Direct application of Theorem 2 is easier because the condition for the function $F(c)$ can be verified with the help of software for symbolic computations. Sometimes it is more convenient to do this after multiplying each of the Ψ functions by the same positive function (see Section 4.3).

There remain, however, some basic questions related the application of either Theorem 1 or Theorem 2 that do not have simple general answers. For example, what is a good choice for p_1 in forming the matrix $C_{22}(c)$ in (2.4)? In Section 4, the choice $p_1 = p/2$ worked well, and selecting p_1 approximately equal to $p/2$ may be a good general starting point. Moreover, there is the question of how to order the rows and columns of the information matrix. By reordering the elements in the parameter vector θ, we could wind up with different matrices $C_{22}(c)$, even after fixing p_1. So what ordering is best? In all of the examples in Section 4, we have used an ordering that makes "higher-order terms" appear in $C_{22}(c)$, and this may offer the best general strategy. There is still another issue related to ordering: In renaming the independent Ψ-functions in the first $p - p_1$ columns of $C(\theta, c)$, different orders will result in different $f_{l,l}$-functions. In some cases, but not for all, these functions will result in a function $F(c)$ that satisfies the condition in Theorem 2. In the examples, we have tended to associate "lower-order terms" with the earlier Ψ-functions, but what order is best may require some trial and error.

Whereas we have demonstrated that the main results of the paper are powerful, regrettably we cannot offer any guarantees that they will always give results as desired, even when the information matrix can be written in the form (2.1).

APPENDIX

PROPOSITION 1. *Assume that* $\{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}\}$ *is a Chebyshev system defined on an interval* $[A, B]$. *Let* $A \le z_1 < z_2 < \cdots < z_t \le B$, *and let* $r_1, \ldots, r_t$ *be coefficients that satisfy the following* k *equations*:

$$\sum_{i=1}^{t} r_i \Psi_l(z_i) = 0, \qquad l = 0, 1, \ldots, k-1. \tag{A.1}$$

Then we have:

(a) *If* $t \le k$, *then* $r_i = 0, i = 1, \ldots, t$.

(b) *If* $t = k + 1$ *and one* r_i *is not zero*, *then all are nonzero*; *moreover all* r_i*'s for odd* i *must then have the same sign*, *which is opposite to that of the* r_i*'s for even* i.

PROOF. For part (a), if $t < k$, we can expand $z_1, \dots, z_t$ to a set of k distinct points, taking $r_i = 0$ for the added points. Thus without loss of generality, take $t = k$. Consider the matrix

$$\Psi(z_1, z_2, \dots, z_k) = \begin{pmatrix} \Psi_0(z_1) & \Psi_0(z_2) & \cdots & \Psi_0(z_k) \\ \Psi_1(z_1) & \Psi_1(z_2) & \cdots & \Psi_1(z_k) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{k-1}(z_1) & \Psi_{k-1}(z_2) & \cdots & \Psi_{k-1}(z_k) \end{pmatrix}. \tag{A.2}$$

Then (A.1) can be written as

$$\Psi(z_1, z_2, \dots, z_k)R = 0,$$

where $R = (r_1, \dots, r_k)^T$. Since $\{\Psi_0, \Psi_1, \dots, \Psi_{k-1}\}$ is a Chebyshev system, $\Psi(z_1, z_2, \dots, z_k)$ is nonsingular, so that $R = 0$.

For part (b), if one r_i is 0, then it follows from part (a) that all r_i's are 0. Therefore, if at least one r_i is nonzero, then all of them must be nonzero. With the notation from the previous paragraph, we can write (A.1) as

$$\Psi(z_1, z_2, \dots, z_k)R = -r_{k+1}\psi(z_{k+1}),$$

where $\psi(z_{k+1}) = (\Psi_0(z_{k+1}), \Psi_1(z_{k+1}), \dots, \Psi_{k-1}(z_{k+1}))^T$. It follows that

$$r_i = -r_{k+1}\frac{|\Psi(z_1, \dots, z_{i-1}, z_{k+1}, z_{i+1}, \dots, z_k)|}{|\Psi(z_1, z_2, \dots, z_k)|}, \qquad i = 1, \dots, k. \tag{A.3}$$

By the Chebyshev system assumption, the denominator $|\Psi(z_1, z_2, \dots, z_k)|$ in (A.3) is positive, while the numerator $|\Psi(z_1, \dots, z_{i-1}, z_{k+1}, z_{i+1}, \dots, z_k)|$ is positive for $i = k, k-2, \dots$ and negative otherwise. The result in (b) follows. □

PROPOSITION 2. *Let $\{\Psi_0 = 1, \Psi_1, \dots, \Psi_{k-1}\}$ be a Chebyshev system on an interval $[A, B]$, and suppose that $k = 2n - 1$. Consider n pairs (c_i, ω_i), $i = 1, \dots, n$, and n pairs $(\tilde{c}_i, \tilde{\omega}_i)$, $i = 1, \dots, n$, with $\omega_i > 0$, $\tilde{\omega}_i > 0$ and $A \le c_1 < \tilde{c}_1 < \cdots < c_n < \tilde{c}_n = B$. Suppose further that the following k equations hold*:

$$\sum_i \omega_i \Psi_l(c_i) = \sum_i \tilde{\omega}_i \Psi_l(\tilde{c}_i), \qquad l = 0, 1, \dots, k-1. \tag{A.4}$$

Then, for any function Ψ_k on $[A, B]$, we can conclude that

$$\sum_i \omega_i \Psi_k(c_i) < \sum_i \tilde{\omega}_i \Psi_k(\tilde{c}_i) \tag{A.5}$$

if $\{\Psi_0 = 1, \Psi_1, \dots, \Psi_{k-1}, \Psi_k\}$ is also a Chebyshev system.

PROOF. With

$$R = (\omega_1, -\tilde{\omega}_1, \omega_2, -\tilde{\omega}_2, \dots, \omega_n)^T,$$

the k equations in (A.4) can be written as

$$\Psi(c_1, \tilde{c}_1, \dots, c_n)R = \tilde{\omega}_n \psi(\tilde{c}_n), \tag{A.6}$$

where Ψ and ψ are as defined in the proof of Proposition 1. Further, (A.5) is equivalent to

$$(\Psi_k(c_1), \Psi_k(\tilde{c}_1), \dots, \Psi_k(c_n))R < \tilde{\omega}_n \Psi_k(\tilde{c}_n). \tag{A.7}$$

Using (A.6) to solve for R, and using that $\tilde{\omega}_n > 0$, we see that (A.7) is equivalent to

$$(\Psi_k(c_1), \Psi_k(\tilde{c}_1), \dots, \Psi_k(c_n))\Psi^{-1}(c_1, \tilde{c}_1, \dots, c_n)\psi(\tilde{c}_n) - \Psi_k(\tilde{c}_n) < 0. \tag{A.8}$$

From an elementary matrix result [see, e.g., Theorem 13.3.8 of Harville (1997)], the left-hand side of (A.8) can be written as

$$-\frac{|\Psi^*(c_1, \tilde{c}_1, \dots, c_n, \tilde{c}_n)|}{|\Psi(c_1, \tilde{c}_1, \dots, c_n)|}, \tag{A.9}$$

where

$$\begin{aligned} &\Psi^*(c_1, \tilde{c}_1, \dots, c_n, \tilde{c}_n) \\ &= \begin{pmatrix} \Psi_0(c_1) & \Psi_0(\tilde{c}_1) & \cdots & \Psi_0(c_n) & \Psi_0(\tilde{c}_n) \\ \Psi_1(c_1) & \Psi_1(\tilde{c}_1) & \cdots & \Psi_1(c_n) & \Psi_1(\tilde{c}_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Psi_{k-1}(c_1) & \Psi_{k-1}(\tilde{c}_1) & \cdots & \Psi_{k-1}(c_n) & \Psi_{k-1}(\tilde{c}_n) \\ \Psi_k(c_1) & \Psi_k(\tilde{c}_1) & \cdots & \Psi_k(c_n) & \Psi_k(\tilde{c}_n) \end{pmatrix}. \end{aligned} \tag{A.10}$$

Since both $\{\Psi_0, \Psi_1, \dots, \Psi_{k-1}\}$ and $\{\Psi_0, \Psi_1, \dots, \Psi_{k-1}, \Psi_k\}$ are Chebyshev systems and $c_1 < \tilde{c}_1 < \cdots < c_n < \tilde{c}_n$, it follows that (A.9) is negative, which is what had to be shown. □

A similar argument as for Proposition 2 can be used for the next result.

PROPOSITION 3. *Let $\{\Psi_0 = 1, \Psi_1, \dots, \Psi_{k-1}\}$ be a Chebyshev system on an interval $[A, B]$ and suppose that $k = 2n$. Consider n pairs (c_i, ω_i), $i = 1, \dots, n$, and $n + 1$ pairs $(\tilde{c}_i, \tilde{\omega}_i)$, $i = 0, 1, \dots, n$, with $\omega_i > 0$, $\tilde{\omega}_i > 0$ and $A = \tilde{c}_0 < c_1 < \tilde{c}_1 < \cdots < c_n < \tilde{c}_n = B$. Suppose further that the following k equations hold*:

$$\sum_i \omega_i \Psi_l(c_i) = \sum_i \tilde{\omega}_i \Psi_l(\tilde{c}_i), \qquad l = 0, 1, \dots, k-1. \tag{A.11}$$

Then, for any function Ψ_k on $[A, B]$, we can conclude that

$$\sum_i \omega_i \Psi_k(c_i) < \sum_i \tilde{\omega}_i \Psi_k(\tilde{c}_i) \tag{A.12}$$

if $\{\Psi_0 = 1, \Psi_1, \dots, \Psi_{k-1}, \Psi_k\}$ is also a Chebyshev system.

PROPOSITION 4. *Consider functions $\Psi_0 = 1, \Psi_1, \dots, \Psi_k$ on an interval $[A, B]$. Compute the corresponding functions $f_{l,l}$ as in (3.7), but with $C_{22}(c)$ replaced by Ψ_k, and suppose that $f_{l,l} > 0$, $l = 1, \dots, k-1$. Then $\{1, \Psi_1, \dots, \Psi_k\}$ is a Chebyshev system if $f_{k,k} > 0$, while $\{1, \Psi_1, \dots, -\Psi_k\}$ is a Chebyshev system if $f_{k,k} < 0$.*

PROOF. The conclusion for the case $f_{k,k} < 0$ follows immediately from that for $f_{k,k} > 0$, so that we will only focus on the latter. We need to show that

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \Psi_1(z_0) & \Psi_1(z_1) & \cdots & \Psi_1(z_k) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_k(z_0) & \Psi_k(z_1) & \cdots & \Psi_k(z_k) \end{vmatrix} > 0 \tag{A.13}$$

for any given $A \le z_0 < z_1 < \cdots < z_k \le B$. Consider (A.13) as a function of z_k. The determinant is 0 if $z_k = z_{k-1}$, so that it suffices to show that the derivative of (A.13) with respect to z_k is positive on (z_{k-1}, B), that is,

$$\begin{vmatrix} 1 & 1 & \cdots & 1 & 0 \\ \Psi_1(z_0) & \Psi_1(z_1) & \cdots & \Psi_1(z_{k-1}) & f_{1,1}(z_k) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Psi_k(z_0) & \Psi_k(z_1) & \cdots & \Psi_k(z_{k-1}) & f_{k,1}(z_k) \end{vmatrix} > 0 \tag{A.14}$$

for any $z_k \in (z_{k-1}, B)$. Now consider (A.14) as a function of z_{k-1}, and use a similar argument. It suffices to show that for $z_{k-1} \in (z_{k-2}, z_k)$,

$$\begin{vmatrix} 1 & 1 & \cdots & 0 & 0 \\ \Psi_1(z_0) & \Psi_1(z_1) & \cdots & f_{1,1}(z_{k-1}) & f_{1,1}(z_k) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Psi_k(z_0) & \Psi_k(z_1) & \cdots & f_{k,1}(z_{k-1}) & f_{k,1}(z_k) \end{vmatrix} > 0. \tag{A.15}$$

Continuing like this, it suffices to show that

$$\begin{vmatrix} f_{1,1}(z_1) & f_{1,1}(z_2) & \cdots & f_{1,1}(z_k) \\ \vdots & \vdots & \ddots & \vdots \\ f_{k,1}(z_1) & f_{k,1}(z_2) & \cdots & f_{k,1}(z_k) \end{vmatrix} > 0 \tag{A.16}$$

for any $A \le z_1 < z_2 < \cdots < z_k \le B$. Since $f_{1,1}(c) > 0$ for $c \in [A, B]$, (A.16) is equivalent to

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \dfrac{f_{2,1}(z_1)}{f_{1,1}(z_1)} & \dfrac{f_{2,1}(z_2)}{f_{1,1}(z_2)} & \cdots & \dfrac{f_{2,1}(z_k)}{f_{1,1}(z_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{f_{k,1}(z_1)}{f_{1,1}(z_1)} & \dfrac{f_{k,1}(z_2)}{f_{1,1}(z_2)} & \cdots & \dfrac{f_{k,1}(z_k)}{f_{1,1}(z_k)} \end{vmatrix} > 0. \tag{A.17}$$

Recall that the entries in the last $k-1$ rows of this matrix are by definition simply values of $f_{l,2}$, $l = 2, \ldots, k$. Hence, applying the same arguments used for (A.13) to (A.17) and using that $f_{2,2}(c) > 0$ for $c \in [A, B]$, it is sufficient to show that

$$\text{(A.18)} \qquad \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \dfrac{f_{3,2}(z_2)}{f_{2,2}(z_2)} & \dfrac{f_{3,2}(z_3)}{f_{2,2}(z_3)} & \cdots & \dfrac{f_{3,2}(z_k)}{f_{2,2}(z_k)} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{f_{k,2}(z_2)}{f_{2,2}(z_2)} & \dfrac{f_{k,2}(z_3)}{f_{2,2}(z_3)} & \cdots & \dfrac{f_{k,2}(z_k)}{f_{2,2}(z_k)} \end{vmatrix} > 0.$$

Continuing like this, the ultimate sufficient condition is that $f_{k,k}(c) > 0$ for $c \in [A, B]$, which is precisely our assumption. Thus the conclusion follows. □

REFERENCES

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

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

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

HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York. MR1467237

KARLIN, S. and STUDDEN, W. J. (1966). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815. MR0196871

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. MR2339137

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

PUKELSHEIM, F. (1989). Complete class results for linear regression designs over the multidimensional cube. In *Contributions to Probability and Statistics* (L. J. Gleser, M. D. Perlman, S. J. Press and A. R. Sampson, eds.) 349–356. Springer, New York. MR1024342

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

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

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

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602-7952
USA
E-MAIL: jstufken@uga.edu