A note on the de la Garza phenomenon for locally optimal designs

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de

Viatcheslav B. Melas
St. Petersburg State University
Department of Mathematics
St. Petersburg, Russia
email: vbmelas@post.ru

July 29, 2010

Abstract

The celebrated de la Garza phenomenon states that for a polynomial regression model of degree $p-1$ any optimal design can be based on at most p design points. In a remarkable paper Yang (2010) showed that this phenomenon exists in many locally optimal design problems for nonlinear models. In the present note we present a different view point on these findings using results about moment theory and Chebyshev systems. In particular, we show that this phenomenon occurs in an even larger class of models than considered so far.

AMS subject classification: Primary 62K05

Keywords and phrases: Locally optimal designs, saturated designs, complete class theorem, moment spaces, Chebyshev systems

1 Introduction

Non linear regression models are widely used for modeling dependencies between response and explanatory variables [see Seber and Wild (1989) or Ratkowsky (1990)]. It is well known that an appropriate choice of an experimental design can improve the quality of statistical analysis substantially, and therefore the problem of constructing optimal designs for nonlinear regression models has found considerable attention in the literature. Most authors concentrate on locally optimal designs which assume that a guess for the unknown parameters of the model is available [see Chernoff (1953), Ford et al. (1992), He et al. (1996), Fang and Hedayat (2008)].

These designs are usually used as benchmarks for commonly used designs. Additionally, they serve as a basis for constructing optimal designs with respect to more sophisticated optimality criteria which address for a less precise knowledge about the unknown parameters [see Pronzato and Walter (1985) or Chaloner and Verdinelli (1995), Dette (1997), Müller and Pázman (1998)]. It is a well known fact that the numerical or analytical calculation of optimal designs simplifies substantially if it is known that the optimal design is saturated, which means that the number of different experimental conditions coincides with the number of parameters in the model [see for example He et al. (1996), Dette and Wong (1996), Imhof and Studden (2001), Imhof (2001), Melas (2006), Fang and Hedayat (2008) among many others].

So, the ideal situation appears if the optimal design is in the sub-class of all saturated designs. In a celebrated paper de la Garza (1954) proved that for a $(p-1)$th-degree polynomial regression model, any optimal design can be based on at most p points. Khuri et al. (2006) considered a nonlinear regression model and introduced the terminology of the de la Garza phenomenon, which means that for any design there exists a saturated design, such that the information matrix of the saturated design is not inferior to that of the given design under the Loewner ordering. In a remarkable paper Yang (2010) derived sufficient conditions on the nonlinear regression model for the occurrence of the de la Garza phenomenon and demonstrated that this situation appears in a broad class of non linear regression models. These results generalize recent findings of Yang and Stufken (2009) for nonlinear models with two parameters.

However, some care is necessary if these results are applied as indicated in the following simple example of homoscedastic linear regression on the interval $[0, 1]$. Here the information matrix of the design which advises the experimenter to take all n observations at the point 0 is given by

$$X_1^T X_1 = \begin{pmatrix} n & 0 \\ 0 & 0 \end{pmatrix}$$

while any other design (using the experimental conditions $x_1, \ldots, x_n$) yields an information matrix

$$X_2^T X_2 = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

It is easy to see that the matrix $X_2^T X_2 - X_1^T X_1$ is indefinite (i.e. it has positive and negative eigenvalues) whenever one of the x_i is positive. Consequently, the design corresponding to $X_1^T X_1$ cannot be improved. On the other hand, it is also easy to see that for any $k \in \{1, \ldots, \lfloor n/2 \rfloor - 1\}$ the information matrix of the design, which takes observations at $x_1 = \ldots = x_{n-2k} = 0$ and at $x_{n-2k+1} = \ldots = x_n = 1/2$ can be improved (with respect to the Loewner ordering) by the information matrix corresponding to the design $x_1 = \ldots = x_{n-k} = 0$ and $x_{n-k+1} = \ldots = x_n = 1$. Thus there exist designs where a "real" improvement is possible, while other designs cannot be improved. Note that the results in Yang (2010) do not provide a classification of the two types of designs.

It is the purpose of the present paper to present a more detailed view point on these problems, which clarifies this – on a first glance – contradiction. In contrast to the method used by Yang (2010), which is mainly algebraic, our approach is analytic and based on the theory of Chebyshev systems and moment spaces [see Karlin and Studden (1966a)]. In particular, we will demonstrate that the de la Garza phenomenon appears in any nonlinear regression model, where the functions in the Fisher information matrix form a Chebyshev system. Additionally, we will solve the problem described in the previous paragraph and we will identify the sufficient conditions stated in Yang (2010) as a special case of an extended Chebyshev system. Therefore, our results generalize the recent findings of Yang (2010) in a non trivial way and, additionally, provide - in our opinion - a more transparent and more complete explanation of the de la Garza phenomenon for optimal designs in nonlinear regression models.

The remaining part of this paper is organized as follows. Section 2 provides a brief introduction in the problem, while Section 3 contains our main results. Finally, the new results are illustrated in a rational regression model, where the currently available methodology cannot be used to establish the de la Garza phenomenon.

2 Locally optimal designs

Consider the common nonlinear regression model

$$Y = \eta(x, \theta) + \varepsilon, \tag{2.1}$$

where $\theta \in \Theta \subset \mathbb{R}^p$ is the vector of unknown parameters, and different observations are assumed to be independent. The errors are normally distributed with mean 0 and variance σ^2. The variable x denotes the explanatory variable, which varies in the design space $[A, B] \subset \mathbb{R}$. We assume that η is a continuous and real valued function of both arguments $(x, \theta) \in [A, B] \times \Theta$ and differentiable with respect to the variable θ. A design is defined as a probability measure ξ on the interval $[A, B]$ with finite support [see Kiefer (1974)]. If the design ξ has masses w_i at the points x_i $(i = 1, \ldots, k)$ and n observations can be made by the experimenter, this means that the quantities $w_i n$ are rounded to integers, say n_i, satisfying $\sum_{i=1}^k n_i = n$, and the experimenter takes n_i observations at each location x_i $(i = 1, \ldots, k)$. The information matrix of an approximate design ξ is defined by

$$M(\xi, \theta) = \int_A^B \Big(\frac{\partial}{\partial \theta}\eta(x, \theta\Big)\Big(\frac{\partial}{\partial \theta}\eta(x, \theta)\Big)^T d\xi(x), \tag{2.2}$$

and it is well known [see Jennrich (1969)] that under appropriate assumptions of regularity the covariance matrix of the least squares estimator is approximately given by $\sigma^2\, M^{-1}(\xi, \theta)/n$, where n denotes the total sample size and we assume that the observations are taken according to the approximate design ξ.

An optimal design maximizes an appropriate functional of the information matrix and numerous criteria have been proposed in the literature to discriminate between competing designs [see Silvey (1980), Pázman (1986) or Pukelsheim (2006) among others]. Note that in nonlinear regression models the information matrix (and as a consequence the corresponding optimal designs) depend on the unknown parameters and are therefore called locally optimal designs [see Chernoff (1953)]. These designs require an initial guess of the unknown parameters in the model and are used as benchmarks for many commonly used designs.

Most of the available optimality criteria satisfy a monotonicity property with respect to the Loewner ordering, that is

$$M(\xi_1,\theta) \leq M(\xi_2,\theta) \Longrightarrow \Phi(M(\xi_1,\theta)) \leq \Phi(M(\xi_2,\theta)), \tag{2.3}$$

where the parameter θ is fixed, ξ_1, ξ_2 are two competing designs and Φ denotes an information function in the sense of Pukelsheim (2006). For this reason it is of interest to derive a complete class theorem in this general context which characterizes the class of designs, which cannot be improved with respect to the Loewner ordering of their information matrices. We call a design ξ_1 admissible if there does not exist a design ξ_2, such that $M(\xi_1,\theta) \neq M(\xi_2,\theta)$ and

$$M(\xi_1,\theta) \leq M(\xi_2,\theta). \tag{2.4}$$

As pointed out in Yang (2010) for many nonlinear regression models the information matrix defined in (2.2) has a representation of the form

$$M(\xi,\theta) = P(\theta)C(\xi,\theta)P^T(\theta), \tag{2.5}$$

where $P(\theta)$ is a nonsingular $p \times p$ matrix, which does not depend on the design ξ, the matrix C is defined by

$$C(\xi,\theta) = \begin{pmatrix} \int_A^B \Psi_{11}(x)d\xi(x) & \cdots & \int_A^B \Psi_{1p}(x)d\xi(x) \\ \vdots & \ddots & \vdots \\ \int_A^B \Psi_{p1}(x)d\xi(x) & \cdots & \int_A^B \Psi_{pp}(x)d\xi(x) \end{pmatrix} \tag{2.6}$$

and $\Psi_{11}, \Psi_{12}, \ldots, \Psi_{pp}$ are functions defined on the interval $[A, B]$. Note that these functions usually depend on the parameter θ, but for the sake of simplicity we do not reflect this dependence in our notation. Obviously the inequality (2.4) is satisfied if and only if the inequality

$$C(\xi_1,\theta) \leq C(\xi_2,\theta). \tag{2.7}$$

is satisfied.

3 Cheyshev systems and complete class theorems

In the following discussion we make extensive use of the property that a system of functions has the Chebyshev property. Following Karlin and Studden (1966a) a set of $k+1$ continuous functions $u_0, \ldots, u_k : [A, B] \to \mathbb{R}$ is called a Chebyshev system (on the interval $[A, B]$) if the inequality

$$(3.1) \qquad \begin{vmatrix} u_0(x_0) & u_0(x_1) & \ldots & u_0(x_k) \\ u_1(x_0) & u_1(x_1) & \ldots & u_1(x_k) \\ \vdots & \vdots & \ddots & \vdots \\ u_k(x_0) & u_k(x_1) & \ldots & u_k(x_k) \end{vmatrix} > 0$$

holds for all $A \leq x_0 < x_1 < \ldots < x_k \leq B$. Note that if the determinant in (3.1) does not vanish then either the functions $u_0, u_1, \ldots, u_{k-1}, u_k$ or the functions $u_0, u_1, \ldots, u_{k-1}, -u_k$ form a Chebyshev system. The Chebyshev property has widely been used to determine explicitly c-optimal designs [see He et al. (1996), Dette et al. (2003) or Dette et al. (2008) among many others]. On the other hand, its application to other optimality criteria has not been studied intensively. In the following discussion we will demonstrate that this property will essentially be the reason for the occurrence of the de la Garza phenomenon. In particular, we will show that it is essentially sufficient to obtain a complete class theorem for the design problems associated with the nonlinear regression model (2.1).
For this purpose we define the index $I(\xi)$ of a design ξ on the interval $[A, B]$ as the number of support points, where the boundary points A and B (if they occur as support points) are only counted by 1/2. Recall the definition of the matrix C in (2.6) and denote by $\Psi_1, \ldots, \Psi_k$ the different elements among the functions $\{\Psi_{ij} \mid 1 \leq j, j \leq p\}$, which are not equal to the constant function. Throughout this paper we assume

$$(3.2) \qquad \Psi_k = \Psi_{ll} \text{ for some } l \in \{1, \ldots, p\} \quad \text{and} \quad \Psi_{ij} \neq \Psi_k \text{ for all } (i, j) \neq (l, l)$$

[see Yang (2010)]. Additionally, we put $\Psi_0(x) = 1$ and assume either that

$$(3.3) \quad \{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}\} \quad \text{and} \quad \{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k\} \quad \text{are Chebyshev systems}$$

or that

$$(3.4) \quad \{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}\} \quad \text{and} \quad \{\Psi_0, \Psi_1, \ldots, \Psi_{k-1}, -\Psi_k\} \quad \text{are Chebyshev systems}$$

then the following result characterizes the class of admissible designs.

Theorem 3.1.

(1) If the functions $\Psi_0(x) = 1, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k$ satisfy (3.2) and (3.3), then for any design ξ there exists a design ξ^+ with at most $\frac{k+2}{2}$ support points, such that $M(\xi^+, \theta) \geq M(\xi, \theta)$.

If the index of the design ξ satisfies

$$I(\xi) < \frac{k}{2}$$

then the design ξ^+ is uniquely determined in the class of all designs η satisfying

$$\int_A^B \Psi_i(x) d\eta(x) = \int_A^B \Psi_i(x) d\xi(x), \quad i = 0, \ldots, k-1 \tag{3.5}$$

and coincides with the design ξ. Otherwise (in the case $I(\xi) \geq \frac{k}{2}$) the following two assertions are valid.

(1a) If k is odd, then ξ^+ has at most $\frac{k+1}{2}$ support points and ξ^+ can be chosen such that its support contains the point B.

(1b) If k is even, then ξ^+ has at most $\frac{k}{2}+1$ support points and ξ^+ can be chosen such that the support of ξ^+ contains the points A and B.

(2) If the functions $\Psi_0(x) = 1, \Psi_1, \ldots, \Psi_{k-1}, \Psi_k$ satisfy (3.2) and (3.4), then for any design ξ there exists a design ξ^- with at most $\frac{k+2}{2}$ support points, such that $M(\xi^-, \theta) \geq M(\xi, \theta)$. If the index of the design ξ satisfies

$$I(\xi) < \frac{k}{2}$$

then the design ξ^- is uniquely determined in the class of all designs η satisfying (3.5) and coincides with the design ξ. Otherwise (in the case $I(\xi) \geq \frac{k}{2}$) the following two assertions are valid.

(2a) If k is odd, then ξ^- has at most $\frac{k+1}{2}$ support points and ξ^- can be chosen such that its support contains the point A.

(2b) If k is even, then ξ^- has at most $\frac{k}{2}$ support points.

Proof. We only present a proof of the first part (1) of the theorem, the second part follows by similar arguments. For $i = 0, \ldots, k$ let

$$d_i(\xi) = \int_A^B \Psi_i(x) d\xi(x)$$

denote the i-th "moment" and define

$$\boldsymbol{d_k}(\xi) = (d_0, (\xi), \ldots, d_k(\xi))^T$$

as the vector of all "moments" up to the order k. Consider two designs ξ_1 and ξ_2 with

$$\boldsymbol{d_{k-1}}(\xi_1) = \boldsymbol{d_{k-1}}(\xi_2) \quad \text{and} \quad d_k(\xi_1) \leq d_k(\xi_2),$$

then for any vector $z = (z_1, \dots, z_p)^T \in \mathbb{R}^p$ we have for some $l \in \{1, \dots, p\}$

$$z^T(C(\xi_2, \theta) - C(\xi_1, \theta)z \geq z_l^2(d_k(\xi_2) - d_k(\xi_1)) \geq 0,$$

which means that

$$C(\xi_2, \theta) \geq C(\xi_1, \theta).$$

Now let for a fixed vector of "moments" $\boldsymbol{d_{k-1}}(\xi)$

$$d_k^+ = \sup\Big\{d_k(\eta) \mid \ \eta \text{ design on } [A, B] \text{ with } \boldsymbol{d_{k-1}}(\eta) = \boldsymbol{d_{k-1}}(\xi)\Big\}$$

denote the maximum of the k-th "moment" over the set of all designs with fixed "moments" up to the order $k-1$. Due to the compactness of the design space and the continuity of the functions $\Psi_0, \dots, \Psi_k$ there exists a design ξ^+ such that

$$d_j(\xi^+) = d_j(\xi)\ ; \quad j = 0, \dots, k-1, \tag{3.6}$$

$$d_k(\xi^+) = d^+ \geq d_k(\xi). \tag{3.7}$$

This shows (by the argument at the beginning of the proof and the discussion at the end of the previous section)

$$M(\xi^+, \theta) \geq M(\xi, \theta). \tag{3.8}$$

Moreover, it follows from Chapter II, Section 6 of Karlin and Studden (1966a) that the point $\boldsymbol{d_k}(\xi^+)$ is a boundary point of the "moment space"

$$\mathcal{M}_k = \{\boldsymbol{d_k}(\eta) \mid \eta \quad \text{design on } [A, B]\}.$$

Consequently, we obtain from Theorem 2.1 in Karlin and Studden (1966a) that the design ξ^+ is based on at most $\frac{k+2}{2}$ support points, which proves the first part of the statement.

We now consider the cases (1a) and (1b). The vector $\boldsymbol{d_{k-1}}(\xi)$ is either a boundary point or an interior point of the $(k-1)$-th moment space $\mathcal{M}_{k-1}$. The first case is characterized by an index satisfying $I(\xi) < k/2$ and there exists a unique measure $\tilde{\xi}$ with "moments" up to the order k specified by $\boldsymbol{d_{k-1}}(\xi)$. To prove this statement regarding uniqueness suppose that $I(\xi) < \frac{k}{2}$ and that there exists a further design, say $\tilde{\xi}$, with this property. Since

$$\boldsymbol{d_{k-1}}(\xi) - \boldsymbol{d_{k-1}}(\tilde{\xi}) = \boldsymbol{0}$$

it follows that there exist at least k different points

$$A \leq x_0 < x_1 < \cdots < x_{k-1} \leq B$$

such that

$$\begin{vmatrix} \Psi_0(x_0) & \Psi_0(x_1) & \dots & \Psi_0(x_{k-1}) \\ \Psi_1(x_0) & \Psi_1(x_1) & \dots & \Psi_1(x_{k-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{k-1}(x_0) & \Psi_{k-1}(x_1) & \dots & \Psi_{k-1}(x_{k-1}) \end{vmatrix} = 0 \tag{3.9}$$

which is impossible by the definition of Chebyshev systems. Consequently, a design with moments specified by (3.5) is uniquely determined and therefore we take $\xi^+ = \tilde{\xi}$, which has at most $\frac{k+1}{2}$ support points [see Theorem 2.1 in Karlin and Studden (1966a), p. 42].

If the index of the design ξ satisfies $I(\xi) \geq k/2$ it follows from the discussion in Chapter II, Section 6 in Karlin and Studden (1966a) that the design ξ^+ defined by (3.6) and (3.7) is the upper principal representation of the vector $\boldsymbol{d_{k-1}}(\xi)$, which means that its index is precisely $\frac{k}{2}$ and its support includes the point B. Note that for this argument we require condition (3.3). Consequently, if $k = 2m + 1$ is odd, the upper principal representation ξ^+ has index $m + \frac{1}{2}$ and precisely $m + 1$ support points including the point B. On the other hand, if $k = 2m$ is even, ξ^+ has $m + 1$ support points and the boundary points A and B of the design interval are support points because the index of the design ξ^+ is m.

The proof of part (2) of Theorem 3.1 is similar (where the upper principal representation has to be replaced by the lower principal representation using condition (3.4)) and omitted. □

Remark 3.2.

(a) Note that Theorem 2.1 in Karlin and Studden (1966a), Chapter II refers to moment spaces corresponding to not necessarily bounded measures and the inclusion of the constant function in the system under consideration guarantees its application to a moment space corresponding to probability measures as required in the proof of Theorem 3.1. An alternative explanation can be given by the generalized equivalence theorem as stated in Pukelsheim (2006). It follows from this result that for an optimal design (with respect to the commonly used criteria) there exist some constants, say $a_i \in \mathbb{R}$, $i = 1, \ldots, k$ such that for all support points of the optimal design the identity

$$\sum_{i=1}^{k} a_i \Psi_i(x) = c,$$

is satisfied, where c denotes a constant (for example for the D-optimality criterion c is the number of parameters). Since an optimal design is admissible, the inclusion of the constant function guarantees that the index of these designs is at most $k/2$. Note that this is a sufficient but, generally speaking, not necessary condition.

(b) Note that it follows from the proof of Theorem 3.1 that the conditions (3.6) and (3.7) imply (3.8), i.e. the superiority of the information matrix of the design ξ^+ with respect to the Loewner ordering. In many cases (for example polynomial regression models) the converse direction is also true and in these cases it follows from the proof of Theorem 3.1 that a design ξ with index $I(\xi) < \frac{k}{2}$) can only be "improved" (with respect to the Loewner ordering of the corresponding information matrices) by itself. In fact we are not aware of any case where the converse direction does not hold.

(c) Note also that Theorem 3.1 provides a solution to the problem indicated in the example of the introduction. In the linear regression model we have $k = 2$, therefore we can use the given design ξ_1 (concentrating all observations at $x = 0$) as an "improvement" of ξ_1. However, because the index of ξ_1 is $1/2 < 1$ the design ξ_1 can only be improved by itself (see the previous remark). In particular there does not exist a design ξ which takes observations at $x = 1$ and improves ξ_1 in the sense $M(\xi) \geq M(\xi_1)$.

(d) It is also worthwhile to mention that a design improving the given design ξ is not necessarily unique. Consider for example again the linear regression model on the interval $[0, 1]$ and the design ξ which has equal masses at the points 0 and 3/4. The information matrix of ξ is given by

$$M(\xi) = \begin{pmatrix} 1 & \frac{3}{8} \\ \frac{3}{8} & \frac{9}{32} \end{pmatrix}.$$

Now define for any $p \in [\frac{1}{2}, \frac{5}{8}]$ a design ξ_p^+ with masses p and $1-p$ at the points 0 and $\frac{3}{8(1-p)}$, respectively. Then it follows that

$$M(\xi_p^+) = \begin{pmatrix} 1 & \frac{3}{8} \\ \frac{3}{8} & \frac{9}{64(1-p)} \end{pmatrix}.$$

and $M(\xi_p^+) \geq M(\xi)$ for any $p \in [\frac{1}{2}, \frac{5}{8}]$. Note that the choice $p = \frac{5}{8}$ gives the upper principal representation $\xi^+ = \xi_{5/8}^+$ with index 1 and support points 0 and 1, while for $p \in [\frac{1}{2}, \frac{5}{8})$ we have index $I(\xi_p^+) = 3/2$.

In the remaining part of this section we will relate the result of Theorem 3.1 to the recent findings of Yang (2010). Note that - in contrast to Theorem 1 and 2 of Yang (2010) our Theorem 3.1 does not require the differentiability of the functions Ψ_j. Moreover, in some case it provides a better description of the admissible designs. For a more detailed explanation we note that a Chebyshev system of functions $\{u_0, \ldots, u_k\}$ is called an extended Chebyshev system, if and only if for any $a_0, \ldots, a_k \in \mathbb{R}$ with $\sum_{i=0}^k a_i^2 \neq 0$ the function

$$\sum_{i=0}^k a_i u_i(x)$$

has at most k zeros counted with multiplicities in the interval $[A, B]$ (this definition is equivalent to the definition given in Karlin and Studden (1966a)). A simple way of constructing an extended Chebyshev system is the following [see Karlin and Studden (1966a), p. 19]. Let $w_0, \ldots, w_k$ be functions on the interval $[A, B]$ which are either positive or negative. We now consider the new functions

$$\begin{aligned} u_0(x) &= w_0(x) \\ u_1(x) &= w_0(x) \int_A^x w_1(t_1) dt_1 \\ \vdots \quad & \vdots \quad \vdots \\ u_k(x) &= w_0(x) \int_A^x w_1(t_1) \int_A^{t_2} w_2(t_2) \ldots \int_A^{t_{k-1}} w_k(t_k) dt_k \ldots dt_1. \end{aligned} \tag{3.10}$$

A direct calculation shows that the Wronskian determinant of the functions $u_0, \ldots, u_k$ is given by

$$W_x(u_0,\ldots,u_k) = \begin{vmatrix} u_0(x) & u_0'(x) & \cdots & u_0^{(k)}(x) \\ u_1(x) & u_1'(x) & \cdots & u_1^{(k)}(x) \\ \vdots & \vdots & \ddots & \vdots \\ u_k(x) & u_k'(x) & \cdots & u_k^{(k)}(x) \end{vmatrix}$$

$$= (w_0(x))^{k+1}(w_1(x))^k \ldots (w_{k-1}(x))^2 w_k(x) \tag{3.11}$$

and it is shown in Chapter XI in Karlin and Studden (1966a) that the set $\{u_0, \ldots, u_k\}$ of k times differentiable function is an extended Chebyshev system if and only if

$$W_x(u_0,\ldots,u_k) > 0$$

for all $x \in [A, B]$. On the other hand, this representation provides a constructive method for checking if a given system of k times differentiable functions $\{u_0, \ldots, u_k\}$ is a Chebyshev system on the interval $[A, B]$. To be precise, define $w_0(x) = u_0(x)$ and recursively differential operators

$$D_j f = \frac{d}{dx}\Big(\frac{f}{w_j}\Big)\ ; \qquad j = 0, \ldots, k \tag{3.12}$$

$$w_{j+1} = (D_j D_{j-1} \ldots D_0) u_{j+1}\ ; \qquad j = 0, 1, \ldots, k-1. \tag{3.13}$$

Consequently, the set $\{u_0, \ldots, u_k\}$ is a Chebyshev system if the functions $w_0, \ldots, w_k$ calculated by (3.12) and (3.13) are all positive on the interval $[A, B]$.

Remark 3.3. Yang (2010) constructed a triangle array of functions $\{f_{l,t} \mid t = 1, \ldots, k;\ t \leq l \leq k\}$ from the functions $\Psi_1, \ldots, \Psi_k$ induced by the nonlinear regression model (2.1) using the recursion

$$f_{l,t}(x) = \begin{cases} \Psi_l'(x) & t = 1, \ldots, k \\ \Big(\frac{f_{l,t-1}(x)}{f_{t-1,t-1}(x)}\Big)' & 2 \leq t \leq k\ ;\ t \leq l \leq k \end{cases}.$$

It is now easy to see that the functions $w_1, \ldots, w_k$ obtained from (3.12) and (3.13) with $w_0 = 1$ $u_j = \Psi_j$ $(j = 1, \ldots, k)$ are precisely the functions f_{ll} defined by Yang (2010). As a consequence, we will obtain the main result of Yang (2010) as a special case of our Theorem 3.1 (note that our assumptions regarding the differentiability are slightly weaker than in this reference).

Theorem 3.4. *Let $\Psi_1, \ldots, \Psi_k$ denote the k different functions in the information matrix (3.1) corresponding to the nonlinear regression model which are not equal to the constant* *function. Assume that* Ψ_j *is* $(j+1)$ *times continuously differentiable, define* $w_0 = 1$ *and for* $j = 0, \dots, k-1$

$$w_{j+1} = D_j D_{j-1} \dots D_0 \Psi_{j+1}$$

and assume that condition (3.2) *is satisfied. If*

$$F(x) = w_1(x) \dots w_k(x) \neq 0$$

for all $x \in [A, B]$, *then for any given design* ξ *there exists a design* $\tilde{\xi}$, *such that* $I(\tilde{\xi}) \leq \frac{k}{2}$

$$M(\tilde{\xi}, \theta) \geq M(\xi, \theta).$$

If the index of the design ξ *satisfies* $I(\xi) < \frac{k}{2}$ *then* $\tilde{\xi}$ *is uniquely determined in the class of all designs* η *with moments specified by* (3.5) *and coincides with the design* ξ. *Otherwise (in the case* $I(\xi) \geq \frac{k}{2}$*) the following assertions are valid.*

(1a) If k *is odd and* $F(x) < 0$ *on the interval* $[A, B]$, *then the design* $\tilde{\xi}$ *has at most* $(k+1)/2$ *support points and* $\tilde{\xi}$ *can be chosen such that the point* A *is a support point.*

(1b) If k *is odd and* $F(x) > 0$ *on the interval* $[A, B]$, *then the design* $\tilde{\xi}$ *has at most* $(k+1)/2$ *support points and* $\tilde{\xi}$ *can be chosen such that the point* B *is a support point.*

(2a) If k *is even and* $F(x) < 0$ *on the interval* $[A, B]$, *then the design* $\tilde{\xi}$ *has at most* $k/2$ *support points.*

(2b) If k *is even and* $F(x) > 0$ *on the interval* $[A, B]$, *then the design* $\tilde{\xi}$ *has at most* $k/2 + 1$ *support points and* $\tilde{\xi}$ *can be chosen such that the points* A *and* B *are support points.*

Proof. Let us define $\Psi_0(x) = 1$ and note that

$$F(x) = \frac{W_x(\Psi_0, \dots, \Psi_k)}{W_x(\Psi_0, \dots, \Psi_{k-1})}.$$

Thus if $F(x) > 0$ then condition (3.3) is fulfilled and if $F(x) < 0$ then condition (3.4) is fulfilled. Now Theorem 3.4 is an immediate corollary of Theorem 3.1. □

Remark 3.5. Note that if the constant function appears among the different functions $\{\Psi_{ij} \mid 1 \leq i \leq j \leq p\}$ in the information matrix (3.1) it is not counted in Theorem 3.4 or Theorem 2 of Yang (2010) [see the proof of Theorem 3, Theorem 5, Theorem 6 and Theorem 7 in this reference].

A number of interesting applications of Theorem 3.4 are given in Yang (2010). Note that in all examples considered there, as well as in the paper of Yang and Stufken (2009), the functions under consideration generate a special type of Chebyshev systems, namely extended Chebyshev systems that can be generated by formulas (3.7). This follows from Remark 3.3 and the discussion before Theorem 3.4. Note that the main advantage of Theorem 3.1 consists in the fact that the de la Garza phenomenon can be established by proving that the system under consideration is a Chebyshev system. For this purpose, several methods are available which differ from the approach presented in Yang (2010) and in the next section we will consider an example illustrating the usefulness of Theorem 3.1.

4 An application to rational regression models

In this section we present a class of nonlinear regression models where Theorem 3.4 [or Theorem 2 in Yang (2010)] is not directly applicable, but the de la Garza phenomenon can be established by an application of Theorem 3.1. For this purpose we consider rational regression models of the form

$$\eta(x,\theta) = \frac{P(x,\theta_{(1)})}{Q(x,\theta_{(2)})}, \tag{4.1}$$

where

$$\begin{aligned} P(x,\theta_{(1)}) &= \theta_1 + \theta_2 x + \cdots + \theta_l x^{(l-1)}, \\ Q(x,\theta_{(2)}) &= 1 + \theta_{l+1} x + \cdots + \theta_{s+l} x^s. \end{aligned}$$

are polynomials of degree $l-1$ and s, respectively, with corresponding parameters

$$\theta_{(1)} = (\theta_1, \ldots, \theta_l)^T, \theta_{(2)} = (\theta_{1+1}, \ldots, \theta_{l+s})^T.$$

It is shown in He et al. (1996) that the information matrix for this model can be written in the form

$$M(\xi,\theta) = B(\theta)C(\xi,\theta)B(\theta),$$

where $\theta = (\theta_1, \ldots, \theta_{l+s})^T$, B denotes an appropriate matrix [see He et al. (1996)], the matrix C is given by

$$C(\xi,\theta) = \int_A^B [1/Q^4(x)] h(x) h(x)^T d\xi(x),$$

$h(x) = (1, x, \ldots, x^{p-1})^T$ denotes the vector of monomials with $p = l + s$ and $Q(x)$ is a polynomial of degree s. Therefore it follows that the different functions in the information matrix are given by

$$\Psi_1(x) = 1/Q^4(x), \ldots, \Psi_k(x) = x^{k-1}/Q^4(x),$$

where $k = 2p - 1$. Define $\Psi_0(x) = 1$, then it is well known [see Karlin and Studden (1966b)] that under the conditions

(a) $Q(x)$ does not vanish in the interval $[A, B]$;

(b) $[Q^4(x)]^{(2p-1)}$ does not vanish in the interval $[A, B]$

the functions $\Psi_0, \Psi_1, \ldots, \Psi_{2p-1}$ generate a Chebyshev system on the interval $[A, B]$ and Theorem 3.1 is applicable here.

However, we will give an alternative proof of this property which yields – as a by-product – a constructive condition under which the condition (b) is fulfilled. Assume that $Q^4(x) > 0$ for all $x \in [A, B]$ and note that a Chebyshev system remains a Chebyshev system after multiplication of all functions by a positive function. Thus in order to apply Theorem 3.1 it is sufficient to prove that the functions

$$1, x, x^2, \ldots, x^{2p-2}, Q^4(x)$$

generate a Chebyshev system on the interval $[A, B]$. The following Lemma provides a sufficient condition for this property.

Lemma 4.1. *Assume that the polynomial $Q(x)$ has only real roots which are either all smaller than A or larger than B. If $s > l - 1$, then the functions*

$$1, x, x^2, \ldots, x^{2p-2}, \epsilon Q^4(x),$$

generate a Chebyshev system on the interval $[A, B]$, where $\epsilon = +1$ if the roots are smaller than A and $\epsilon = -1$ if the roots larger than B.

Proof. We restrict ourselves to the case where all roots of the polynomial $Q(x)$ are real and smaller than A and its leading coefficient is positive. All other cases are treated similarly. Consider a polynomial $R(x)$ of a degree $2n$ with positive leading coefficient, where all its roots are real, simple and smaller than A. Define $x_{\min}$ and $x_{\max}$ as the smallest and largest root of $R(x)$, then all derivatives of $R(x)$ of even order less or equal than $2n$ are positive outside of the interval $[x_{\min}, x_{\max}]$. To show this property note that by Rolle's theorem the first derivative of the polynomial $R(x)$ vanishes between two roots of $R(x)$. This means that the first derivative does not vanish for all x in the complement of the interval $[x_{\min}, x_{\max}]$. Moreover (because the degree of R is even) it is positive for $x > x_{\max}$ and negative for $x < x_{\min}$. By a multiple application of this argument we obtain that all derivatives of even order of the polynomial $R(x)$ are positive, whenever $x \notin [x_{\min}, x_{\max}]$. Finally, if there exist roots of the polynomial $R(x)$, which are not simple, we approximate $R(x)$ by a polynomial with simple roots and obtain the assertion by a limit argument.

Now consider the polynomial $Q(x)$, then $Q^4(x)$ is a polynomial of even degree ($4s = 2n$) and it follows from the discussion of the previous paragraph that $(Q^4(x))^{(2j)} \geq 0, j = 1, \ldots, 2s$ for all $x \notin [x_{\min}, x_{\max}]$. Similarly, it can be shown that $(Q^4(x))^{(2j-1)} \leq 0$ $(j = 1, \ldots, 2s)$

for all $x < x_{\min}$ and $(Q^4(x))^{(2j-1)} \geq 0$ $(j = 1, \dots, 2s)$ for all $x > x_{\max}$. Define $u_0(x) = 1, u_1(x) = x, \dots, u_{2p-2}(x) = x^{2p-2}, u_{2p-1}(x) = Q^4(x)$. By formulas (3.12) and (3.13) we can easily calculate that $w_0(x) = w_1(x) = \dots = w_{2p-2}(x) = 1, w_{2p-1}(x) = [Q^4(x)]^{(2p-1)}$. Thus if $s > l-1$ it follows that $w_{2p-1}(x)$ is negative for $x < x_{\min}$ and positive for $A > x > x_{\max}$. Therefore (note that $[Q^4(x)]^{(2p-1)}$ has no roots in the interval $[A, B]$) we have $w_{2p-1}(x) > 0$ for all $x \in [A, B]$. Now the assertion of Lemma 4.1 follows from the formula for the Wronskian determinant in (3.11) and the fact that a positive Wronskian determinant is sufficient for the Chebyshev property of the functions $u_0, \dots, u_{2p-1}$. □

The following result is now an immediate consequence of Lemma 4.1 and Theorem 3.1 (note that we do not repeat the statement of uniqueness of the latter result).

Theorem 4.2. *Consider the rational regression model (4.1). Assume that $s > l - 1$ and that the polynomial $Q(x)$ has only real roots, which are either all smaller than A or larger than B. Then for any design ξ there exists a design $\tilde{\xi}$ with at most p support points, such that $M(\xi, \theta) \leq M(\tilde{\xi}, \theta)$. Moreover,*

(1) if the index of ξ satisfies $I(\xi) \geq p - \frac{1}{2}$ and all roots of the polynomial Q are smaller than A then $\tilde{\xi}$ can be chosen such that the support of $\tilde{\xi}$ contains the point B,

(2) if the index of ξ satisfies $I(\xi) \geq p - \frac{1}{2}$ and all roots of the polynomial Q are larger than B then $\tilde{\xi}$ can be chosen such that the support of $\tilde{\xi}$ contains the point A.

Remark 4.3.

(a) Theorem 4.2 is an extension of Theorem 5 in He et al. (1996) who investigated only locally D-optimal designs.

(b) Note that Yang (2010) considered the classical weighted polynomial regression model where the different functions in the information matrix are given by $\Psi_j(x) = \lambda(x)x^{j-1}, j = 1, \dots, 2p-1$, where λ is a positive function on the interior of the design space, which is called efficiency function [see Dette and Trampisch (2010)]. His findings can be generalized in the following way. If there exists a function $g(x)$ such that

$$\frac{\partial}{\partial x}\Big((\frac{\partial}{\partial x}\lambda(x))g(x) \Big) = c \tag{4.2}$$

for some constant $c \in \mathbb{R}$, then one can use

$$\Psi_0(x) = \int_0^x g(t)dt$$

and obtains a system of functions satisfying the assumptions of Theorem 3.4. In particular, in Theorem 9 of Yang (2010) for the case $\lambda(x) = \exp(x^2)$ the function $g(x) = 1/\lambda(x) =$

$\exp(-x^2)$ is appropriate, while the case $\lambda(x) = (1-x)^{\alpha+1}(1+x)^{\beta+1}$ requires the choice $g(x) = (1-x)^{\alpha}(1+x)^{\beta}$. Moreover, the differential equation (4.2) shows that there are many other efficiency functions for which the de la Garza phenomenon in the weighted polynomial regression model occurs. For example, if $\lambda(x) = \exp(x^{2n})$ ($n \in \mathbb{N}$) one could use

$$g(x) = \frac{\exp(x^{-2n})}{2n\ x^{2n-2}}$$

and it follows that for the weighted polynomial regression model with this efficiency function any optimal design can be based on at most p points. However, for the rational model of the form (4.1) such a technique seemingly does not work. The alternative way is to prove that the functions $1, x, \ldots, x^k, \lambda(x)^{-1}$ generate a Chebyshev system and to use the new Theorem 3.1 to establish the de la Garza phenomenon. Such a method has been realized for the rational model (4.1) in the proof of Theorem 4.2.

Acknowledgements The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. We are also grateful to M. Yang and W.J. Studden for helpful discussions on an earlier version of this paper. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG). The work of Viatcheslav Melas was also partly supported by Russian Foundation for Basic Research (RFBR), project 09-01-00508.

References

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602.

de la Garza, A. (1954). Spacing of information in polynomial regression. *Annals of Mathematical Statistics*, 25:123–130.

Dette, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. C. (2008). Optimal designs for dose finding studies. *Journal of the American Statistical Association*, 103(483):1225–1237.

Dette, H., Melas, V. B., Pepelyshev, A., and Strigul, N. (2003). Efficient design of experiments in the Monod model. *Journal of the Royal Statistical Society, Ser. B*, 65:725–742.

Dette, H. and Trampisch, M. (2010). A general approach to D-optimal designs for weighted univariate polynomial regression models (with discussion). *Journal of the Korean Statistical Society*, 39(1):1–26.

Dette, H. and Wong, W. K. (1996). Optimal bayesian designs for models with partially specified heteroscedastic structure. *Annals of Statistics*, 24:2108–2127.

Fang, X. and Hedayat, A. S. (2008). Locally D-optimal designs based on a class of composed models resulted from blending Emax and one-compartment models. *Annals of Statistics*, 36:428–444.

Ford, I., Torsney, B., and Wu, C. F. J. (1992). The use of canonical form in the construction of locally optimum designs for nonlinear problems. *Journal of the Royal Statistical Society, Ser. B*, 54:569–583.

He, Z., Studden, W. J., and Sun, D. (1996). Optimal designs for rational models. *Annals of Statistics*, 24:2128–2142.

Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics*, 29:561–576.

Imhof, L. A. and Studden, W. J. (2001). E-optimal designs for rational models. *Annals of Statistics*, 29:763–783.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40(2):633–643.

Karlin, S. and Studden, W. (1966a). *Tchebysheff Systems: with Application in Analysis and Statistics*. Wiley, New York.

Karlin, S. and Studden, W. J. (1966b). Optimal experimental designs. *Annals of Mathematical Statistics*, 37(4):783–815.

Khuri, A., Mukherjee, B., Sinha, B., and Ghosh, M. (2006). Design issues for generalized linear models. *Statistical Science*, 21(3):376–399.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2:849–879.

Melas, V. B. (2006). *Functional Approach to Optimal Experimental Design*. Springer Verlag, Berlin.

Müller, C. H. and Pázman, A. (1998). Applications of necessary and sufficient conditions for maximum efficient designs. *Metrika*, 48:1–19.

Pázman, A. (1986). *Foundations of Optimum Experimental Design.* D. Reidel Publishing Company, Dordrecht.

Pronzato, L. and Walter, E. (1985). Robust experimental design via stochastic approximation. *Math. Biosciences*, 75:103–120.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models*. Dekker, New York.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. John Wiley and Sons Inc., New York.

Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

Yang, M. (2010). On the de la Garza Phenomenon. *Annals of Statisticsr*, 38:2499–2524.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Annals of Statistics*, 37:518–541.