Statistica Sinica **21** (2011), 1557-1570
doi:http://dx.doi.org/10.5705/ss.2009.202

OPTIMAL DESIGNS FOR ESTIMATING THE DERIVATIVE IN NONLINEAR REGRESSION

Holger Dette[1], Viatcheslav B. Melas[2] and Petr Shpilev[2]

[1]*Ruhr-Universität Bochum and* [2]*St. Petersburg State University*

Abstract: We consider the problem of estimating the derivative of the expected response in nonlinear regression models. It is demonstrated that in many cases the optimal designs for estimating the derivative have either on m or $m-1$ support points, where m denotes the number of unknown parameters in the model. It is also shown that the support points and weights of the optimal designs are analytic functions, and this result is used to construct a numerical procedure for the calculation of the optimal designs. The results are illustrated in exponential regression and rational regression models.

Key words and phrases: Chebyshev system, c-optimal designs, implicit function theorem, nonlinear regression.

1. Introduction

Nonlinear regression models are widely used to describe the dependencies between a response and an explanatory variable (see e.g., Ratkowsky (1983)). In these models the problem of experimental design has found considerable interest. Many authors have discussed the problem of determining optimal designs for parameter estimation in nonlinear regression models (see for example Chernoff (1953) for an early, and He, Studden and Sun (1996) for a recent reference). Robust design strategies have been proposed by Chaloner and Larntz (1989), Dette (1997) and Müller and Pázman (1998) using a Bayesian or minimax approach. Most of the literature concentrates on optimal designs maximizing a functional of the Fisher information matrix for the parameters in the model, which is related to the problem of estimating the response function most precisely.

The present paper is devoted to the problem of optimal designing of experiments for estimating the derivative of the expected response in a nonlinear regression model. Some motivation for this problem can be found in the pioneering work of Atkinson (1970), and the problem has subsequently been taken up by many other authors (see e.g., Murty and Studden (1972), Mukerjee and Huda (1985), and Melas, Pepelyshev and Cheng (2003), among others). While most of these papers consider linear regression models, we take a closer look at design problems of this type in the context of nonlinearity. In particular we consider the problem of constructing locally optimal designs for a class of nonlinear regression models of the form

$$Y = \eta(t, a, b) + \varepsilon = \sum_{i=1}^{k} a_i \varphi(t, b_i) + \varepsilon \ , \tag{1.1}$$

where φ is a known function, the explanatory variable t varies in an interval $T \subset \mathbb{R}$, ε denotes a random error with mean zero and constant variance, and $\lambda = (a_1, \ldots, a_k, b_1, \ldots, b_k)^T \in \mathbb{R}^{2k}$ denotes the vector of unknown parameters in the model. The problem of designing experiments for models of the form (1.1) has been studied by Melas (2006) and Dette, Melas and Pepelyshev (2006), who considered the case of exponential models. Another popular class is that of rational regression models because of their appealing approximation properties (see Petrushev and Popov (1987) for some theoretical properties, and Ratkowsky (1983) for an application of this model). Optimal design problems for estimation of all parameters have been discussed in Dette, Melas and Pepelyshev (2004a).

The present work considers the problem of designing experiments for the estimation of the derivative of the expected response, at a given point x, in models of the form (1.1). In Section 2 we present some general results for this problem. It is shown that the support points and weights of the locally optimal designs in the regression model (1.1) are analytic functions of the point x where the derivative has to be estimated. This result is used to provide a Taylor expansion for the weights and support points as functions of x, which can easily be used for numerical calculation of the optimal designs. Section 3 considers the case of exponential regression models, while rational functions are discussed in Section 4. We use a general method to determine numerically the optimal design for estimating the derivative, and study their properties. In particular, it is shown that the optimal designs for estimating the derivative of the expected response at the point x have either $2k$ or $2k-1$ support points. Moreover, the locally optimal designs are rather robust with respect to changes in the nonlinear parameters.

2. Optimal Designs for Estimating the Derivative

Consider the regression model at (1.1), where the design space is the interval $T = [T_0, T_1]$, $T_1 \in (0, \infty)$ and $0 \leq T_0 < T_1$. We assume that for each $t \in T$ an observation Y could be made, where different observations are assumed to be independent with the same variance, say $\sigma^2 > 0$. We call any probability measure $\xi = \{t_1, \ldots, t_{n-1}, t_n; \omega_1, \ldots, \omega_{n-1}, \omega_n\}$ with finite support $t_1, \ldots, t_n \in T$, $t_i \neq t_j (i \neq j)$, and masses $\omega_i > 0$, $\sum_{i=1}^{n} \omega_i = 1$ an experimental design. If N experiments can be performed, a rounding procedure is applied to obtain the samples sizes $N_i \approx w_i N$ at the experimental conditions t_i, $i = 1, \ldots, n$ such that

$\sum_{i=1}^{n} N_i = N$ (see e.g., Pukelsheim (1993)). The information matrix of a design ξ for the model (1.1) is

$$M(\xi,\lambda) = \int_T f(t,\lambda) f^T(t,\lambda) d\xi(t) \ , \tag{2.1}$$

where

$$f(t,\lambda) = \frac{\partial \eta(t,\lambda)}{\partial \lambda} = (f_1(t,\lambda),\ldots,f_{2k}(t,\lambda))^T \tag{2.2}$$

is the vector of partial derivatives of the response function with respect to the parameter $\lambda = (a_1,\ldots,a_k,b_1,\ldots b_k)^T$. It is well known that for uncorrelated observations (obtained from approximate designs using an appropriate rounding procedure) the covariance matrix of the least squares estimator for the parameter λ is approximately proportional to the inverse of the information matrix. Consequently, an optimal design maximizes (or minimizes) an appropriate concave (or convex) function of the information matrix or its inverse, and there are numerous optimality criteria that can be used to discriminate among competing designs (see Pukelsheim (1993)).

Most optimality criteria reflect the problem of efficient parameter estimation. If the estimation of the derivative $\eta'(x,\lambda)$, i.e., the gradient of $\eta(t,\lambda)$ with respect to t at the point $t = x$, is of interest, a common estimate is $\hat{\eta} = \eta'(x,\hat{\lambda})$, where $\hat{\lambda}$ denotes the nonlinear least squares estimate. A straightforward application of the delta method now shows that the variance of this estimate is approximately

$$\mathrm{Var}(\hat{\eta}) = \frac{\sigma^2}{N} \left(f'(x,\lambda)\right)^T M^{-}(\xi,\lambda) f'(x,\lambda) \cdot (1 + o(1)),$$

where it is assumed that the vector $f'(x,\lambda) = (f_1'(x,\lambda),\ldots,f_{2k}'(x,\lambda))^T$ is estimable by the design ξ, i.e. $f'(x,\lambda) \in$ Range $(M(\xi,\lambda))$ and $f'(x,\lambda) = \frac{\partial}{\partial x} f(x,\lambda)$. Throughout, we take

$$\Phi(x,\xi,\lambda) = \begin{cases} (f'(x,\lambda))^T M^{-}(\xi,\lambda) f'(x,\lambda) & \text{if} \quad f'(x,\lambda) \in \text{Range } (M(\xi,\lambda)), \\ \infty & \text{else.} \end{cases} \tag{2.3}$$

as the term depending on the design ξ in this expression, and call a design ξ^* minimizing $\Phi(x,\xi,\lambda)$ in the class of all (approximate) designs satisfying $f'(x,\lambda) \in$ Range $(M(\xi,\lambda))$ optimal for estimating the derivative of the expected response in model (1.1). Note that the criterion (2.3) corresponds to a c-optimal design problem in the linear regression model $\theta^T f(t,\lambda)$, which has found considerable interest in the literature. In particular, it follows that there always exists an optimal design for estimating the expected derivative with at most $2k$ support points (Fellmann (1974), Pukelsheim (1993, Chap. 8.3)). Moreover, the criterion depends on the parameter λ and, following Chernoff (1953), we assume that a preliminary guess for this parameter vector is available.

Throughout this paper we assume that the functions $f_1, \ldots, f_{2k}$ constitute a Chebyshev system on the interval T (see Karlin and Studden (1966)). Recall that a set of functions $g_1, \ldots, g_m : T \to \mathbb{R}$ is called a weak Chebyshev system (on the interval T) if there exists an $\varepsilon \in \{-1, 1\}$ such that

$$\varepsilon \cdot \begin{vmatrix} g_1(t_1) & \ldots & g_1(t_m) \\ \vdots & \ddots & \vdots \\ g_m(t_1) & \ldots & g_m(t_m) \end{vmatrix} \geq 0 \tag{2.4}$$

for all $t_1, \ldots, t_m \in T$ with $t_1 < \ldots < t_m$. If the inequality in (2.4) is strict, then $\{g_1, \ldots, g_m\}$ is called Chebyshev system. It is well known (see Karlin and Studden (1966, Thm. II 10.2)) that if $\{g_1, \ldots, g_m\}$ is a weak Chebyshev system, then there exists a unique function $\sum_{i=1}^m c_i^* g_i(t) = c^{*T} g(t)$, $g^T(t) = (g_1(t), \ldots, g_m(t))$, satisfying

$$|c^{*T} g(t)| \leq 1 \quad \forall\, t \in T, \tag{2.5}$$

$$\text{there exist } m \text{ points } s_1 < \ldots < s_m \text{ such that } c^{*T} g(s_i) = (-1)^i \; i = 1, \ldots, m. \tag{2.6}$$

The function $c^{*T} g(t)$ is called Chebyshev polynomial, the points $s_1, \ldots, s_m$ are called Chebyshev points and need not be unique. They are unique if $1 \in \text{span}\{g_1, \ldots, g_m\}, m \geq 1$ and T is a bounded and closed interval, where in this case $s_1 = \min_{t \in T} t$, $s_m = \max_{t \in T} t$. It is well-known (see Studden (1968)) that in many cases c-optimal designs are supported at Chebyshev points.

To formulate one of our basic results we need the concept of an extended Chebyshev system of order 2. For this purpose let $t_1 \leq \ldots \leq t_m$ denote m points in T, where equality occurs at most at two consecutive points. Consider the determinant

$$U^* \begin{pmatrix} 1 \ldots m \\ t_1 \ldots t_m \end{pmatrix} = \det(g(t_1), \ldots, g(t_m)),$$

where the two columns $g(t_i), g(t_{i+1})$ are replaced by $g(t_i), g'(t_{i+1})$ if the points t_i and t_{i+1} coincide. The functions $g_1(t), \ldots, g_m(t)$ generate an extended Chebyshev system of order 2 on the set T if and only if

$$U^* \begin{pmatrix} 1 \ldots m \\ t_1 \ldots t_m \end{pmatrix} > 0$$

for all $t_1 \leq \cdots \leq t_m$ ($t_j \in T$; $j = 1, \ldots, m$), where equality occurs at most at two consecutive points t_j. Note that under this assumption any linear combination

$\sum_{i=1}^{m} \alpha_i g_i(t)$ $(\alpha_1, \dots, \alpha_m \in \mathbb{R}, \sum_{i=1}^{m} \alpha_i^2 \neq 0)$ has at most $m-1$ roots in T, where multiple roots are counted twice regardless of multiplicity (see Karlin and Studden (1966, Chap. 1)).

We begin with the result that the optimal design for estimating the derivative in the nonlinear regression model (1.1) only depends on the "nonlinear" parameters $b_1, \dots, b_k$ of the model. The proof is straightforward and therefore omitted.

Lemma 1. *In the nonlinear regression model* (1.1) *the optimal design for estimating the derivative of the expected response at a point x does not depend on the parameters $a_1, \dots, a_m$.*

Our next result specifies the number of support points of the locally optimal design for estimating the derivative of the expected response in the nonlinear regression model (1.1). A proof of this result is omitted because it can be obtained by exactly the arguments in Dette, Melas and Pepelyshev (2009), who proved a similar result.

Theorem 1. *Assume that the functions $f_1, \dots, f_{2k}$ defined in* (2.2) *form an extended Chebyshev system of order* 2 *on the interval T, then the number of support points of any optimal design for estimating the derivative of the expected response in the nonlinear regression model* (1.1) *is at least $2k-1$. Moreover, if the number of support points is $2k$, then these points must be Chebyshev points with at least one point that coincides with a boundary of the design interval. If the constant function is an element of span$\{f_1, \dots, f_{2k}\}$ then the number of support points is at most $2k$.*

Remark 1. If the design has $2k$ support points it follows by standard arguments of optimal design theory (see for example Pukelsheim and Torsney (1991), Pukelsheim (1993)) that the weights at the support points are

$$\omega_i^* = \frac{|e_i^T F^{-1} f'(x, \lambda)|}{\sum_{i=1}^{2m} |e_i^T F^{-1} f'(x, \lambda)|}, \; i = 1, \dots, 2k, \tag{2.7}$$

where $e_1^T = (1, 0, \dots, 0), \dots, e_{2k}^T = (0, \dots, 0, 1)$ denote the standard basis of $\mathbb{R}^{2k}$, the $2k \times 2k$ matrix F is given by $F = (f(s_1, \lambda), \dots, f(s_{2k}, \lambda))$, and $s_1, \dots, s_{2k}$ denote the support points of the optimal design. Moreover, it follows from Theorem 1 that the support points do not depend on the particular point x where the estimation of the derivative has to be performed.

For the construction of the locally optimal designs for estimating the derivative we use the functional approach, described in Melas (2006), that allows us to calculate support points and weights of the optimal design ξ_x^* for estimating the derivative as a function of the point x. We assume that the number of support points of the design ξ_x^* is constant, say $n \in \mathbb{N}$, for all x contained in some interval, say $[a^*, b^*)$. Then either $n = 2k-1$ or $2k$, and the smallest or largest support point of the design can be a boundary point of the design space T, a total of eight possible scenarios. All cases can be studied in a similar fashion.

To be specific, take the smallest support point to be the left endpoint, $t_1 = T_0$, and the largest support point to be an interior point of $[a^*, b^*)$. We write the design ξ_x^* as the vector

$$\Theta = \Theta(\xi_x^*) = (t_2, \ldots, t_n, \omega_1, \ldots, \omega_{n-1})^T.$$

Note that since $\omega_n = 1 - \sum_{i=1}^{n-1} \omega_i$ there is one-to-one correspondence between designs with $t_1 = T_0$ and vectors Θ. Fix λ, let $\Psi(x, \Theta) = \Phi^{-1}(x, \xi_x^*, \lambda)$, and consider the system of equations

$$\frac{\partial \Psi(x, \Theta)}{\partial \Theta} = 0. \tag{2.8}$$

Clearly the optimal design ξ_x^* is a solution of the system (2.8). The Jacobi matrix of the system is

$$J(x, \Theta) = \left(\frac{\partial^2}{\partial \Theta_i \partial \Theta_j} \Psi(x, \Theta) \right)_{i,j=1}^{2n-2} \in \mathbb{R}^{(2n-2)\times(2n-2)}. \tag{2.9}$$

If the Jacobi matrix is nonsingular, at some point x_0, then, by a straightforward application of the Implicit Function Theorem (see e.g., Gunning and Rossi (1965)), in a neighbourhood of this point there exists an analytic function $\Theta^*(x)$ that is a solution at (2.8) and the locally optimal design for estimating the derivative in the nonlinear regression model. Moreover, if one is able to find a solution $\Theta^*(x_0)$ of this system at a particular point $x = x_0$, then one can construct a Taylor expansion for the support points and weights of $\Theta^*(x)$ of the optimal design for all x in a neighbourhood of x_0. The coefficients of this expansion can be determined recursively, as proved by Dette, Melas and Pepelyshev (2004b).

Theorem 2. *If the Jacobi matrix at* (2.9) *is nonsingular at* $x_0 \in (-\infty, \infty)$ *with* $\Theta = \Theta^*(x_0)$, *then the coefficients* $\Theta^*(j, x_0)$ *of the Taylor expansion*

$$\Theta^*(x) = \Theta^*(x_0) + \sum_{j=1}^{\infty} \frac{1}{j!} \cdot \Theta^*(j, x_0)(x - x_0)^j$$

in the neighbourhood of x_0 *can be obtained recursively as*

$$\Theta^*(s+1, x_0) = -J^{-1}(x_0, \Theta^*(x_0)) \left(\frac{d}{dx} \right)^{s+1} h(\widetilde{\Theta}^*_{(s)}(x), x) \,|_{x=x_0}, \quad s = 0, 1, \ldots,$$

where

$$\widetilde{\Theta}^*_{(s)}(x) = \Theta^*(x_0) + \sum_{j=1}^{s} \frac{1}{j!} \cdot \Theta^*(j, x_0)(x - x_0)^j,$$

$$h(\widetilde{\Theta}, x) = \frac{\partial}{\partial \Theta} \Psi(x, \Theta) \mid_{\Theta = \widetilde{\Theta}} .$$

From the next result, we get that the coefficients in the Taylor expansion of the function $\Theta^*(x)$, which represents the support points and weights of the locally optimal design for estimating the derivative of the expected response at the point x, can be obtained by the recursive formulas of Theorem 2. The result is proved in the Appendix.

Theorem 3. *The Jacobi matrix at* (2.8) *is nonsingular, whenever the optimal design for estimating the derivative in the nonlinear regression model* (1.1) *has* $n = 2k$ *or* $2k - 1$ *support points.*

If the assumptions of Theorem 2 and Theorem 3 are satisfied, the functional approach can be easily used for constructing any optimal design for estimating the derivative in the nonlinear regression model (1.1). In the following sections we illustrate this with two examples.

3. Optimal Designs for Estimating the Derivative in Exponential Regression Models

For the special choice $\varphi(t, b_i) = \exp(b_i t)$ the nonlinear regression model reduces to the exponential regression model

$$Y = \eta_1(t, \lambda) + \varepsilon = \sum_{i=1}^{k} a_i \exp(b_i\ t) + \varepsilon, \tag{3.1}$$

where $\lambda = (a_1, b_1, a_2, b_2, \ldots, a_k, b_k)^T$ denotes the vector of unknown parameters, and we assume that the explanatory variable varies in the interval $T = [0, T_1]$. It is easy to see that this model satisfies the assumptions of Theorem 3.

To illustrate our general procedure, we considered (3.1) for $k = 2$, and constructed locally optimal designs for estimating the derivative in this model by the functional approach. The vector of parameters is given by $\lambda = (1, 0.5, 1, 1)^T$ and the design interval is $T = [0, 1]$. There are two types of optimal designs: a design with four support points including the boundary points of the design space; a design with three support points $0 \le t_1^*(x) < t_2^*(x)) < t_3^*(x) \le 1$. To begin with, we take $x_0 = 0$. The optimal design $\xi^*(0)$ has masses $0.3509, 0.4438, 0.1491$, and 0.0562 at the points $0, 0.3011, 0.7926$, and 1, respectively.

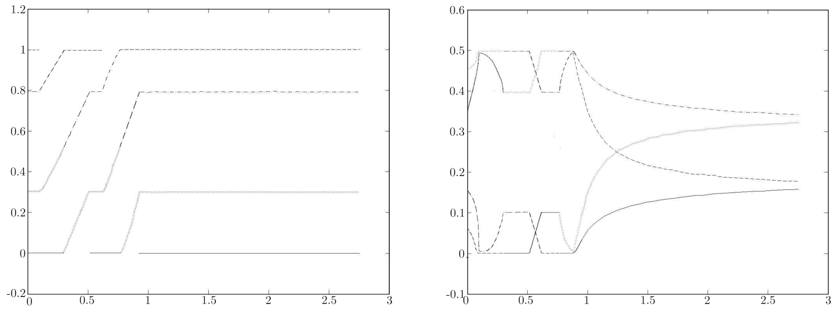

Figure 1. The points (left) and weights (right) of the optimal design for estimating the derivative of the expected response in the nonlinear regression model (3.1) at the point $x \in [0, 2.7]$. The design interval is $[0, 1]$, $k = 2$, and the vector of parameters is $\lambda = (1, 0.5, 1, 1)^T$.

By Theorem 2 the design is of this form in a neighbourhood of the point x_0, where the support remains unchanged. We can then use the representation of the weights. See Remark 2.1 to determine the point where the type of the design changes. Thus the minimal point $x_1 > x_0 = 0$ such that one of the equations $\omega_i^* = 0$, $(i = 1, \dots, 4)$ is satisfied, ω_i^* as at (2.7) is $x_1 = 0.1457$; in the interval $I_0 = [x_0, x_1)$ the Jacobi matrix is non-singular and we can use Theorem 2 to determine the coefficients in the Taylor expansion of the function $\Theta^*(x)$. Note that there is an interval $I_1 = (x_1, x_2)$ such that, for $x \in I_1$, the optimal design for estimating the derivative in the exponential regression model (3.1) at the point x has only three support points. The points and weights are now obtained by a further Taylor expansion and the procedure is continued for the other intervals. The weights and points are depicted in Figure 1 as a function of x, where the derivative is to be estimated. We observe that the type of design changes several times, when x varies in the interval $[0, 2.7]$. In particular, there are four support points if $x \in [0, 0.1457] \cup [0.5001, 0.5875] \cup [0.9092, 2.7]$

In this example the vector of parameters required for the calculation of the locally optimal design was fixed and we have varied x. We also studied the sensitivity of the locally optimal design with respect to the choice of the initial parameters and found the locally optimal designs to be rather robust with respect to changes in the initial parameter b_1 and b_2.

The D-optimal design is efficient for estimating the parameters. By the Equivalence Theorem of Kiefer and Wolfowitz (1960), it is also (minimax-) efficient for estimating the expected response. We look at its efficiency for estimating the derivative of the expected response. Let

$$\text{eff}(x, \xi_1, \xi_2, \lambda) = \frac{\Phi(x, \xi_2, \lambda)}{\Phi(x, \xi_1, \lambda)}, \tag{3.2}$$

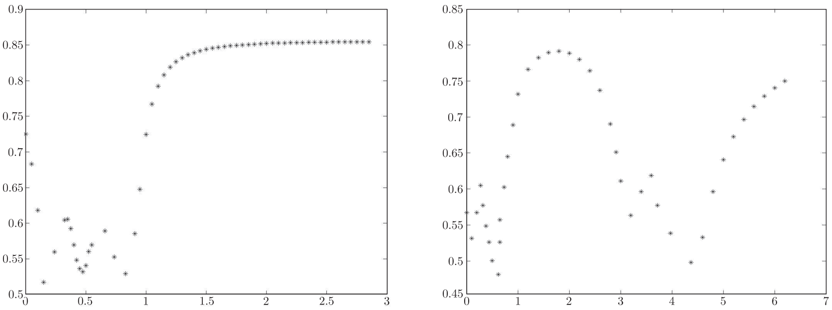

Figure 2. The efficiency of the D-optimal design relative to the optimal design for estimating the derivative of the expected response. The design interval is $[0, 1]$, $k = 2$, and the vector of parameters is $\lambda = (1, 0.5, 1, 1)$. Left panel: the exponential model (3.1) for $x \in [0, 2.7]$; right panel: the rational model (4.1) for $x \in [0, 6.2]$.

be the efficiency of the design ξ_1 relative to the design ξ_2. These efficiencies depend on the particular point x where the estimation of the derivative is performed, and on the nonlinear parameters in the model. We first fix the vector of parameters at $\lambda = (1, 0.5, 1, 1)^T$ and vary the point x. The corresponding efficiencies of the D-optimal design are depicted in the left part of Figure 2. We observe that the efficiency is first decreasing to values smaller than 55%, but for larger x the D-optimal design is rather efficient. It is interesting to note that the lowest efficiencies are obtained for those values of x where the type of the design changes. Corresponding results for a fixed $x = 0$ and various combinations of the nonlinear parameters (b_1, b_2) are shown in Table 1. We observe that the efficiencies are approximately 72% and do not change substantially with (b_1, b_2).

4. Optimal Designs for Estimating the Derivative in Rational Regression Models

For $\varphi(t, b_i) = 1/(b_i + t)$, the nonlinear regression model (1.1) reduces to the rational regression model

$$Y = \eta_2(t, \lambda) + \varepsilon = \sum_{i=1}^{k} \frac{a_i}{t + b_i} + \varepsilon, \tag{4.1}$$

where $\lambda = (a_1, b_1, a_2, b_2 \ldots, a_k, b_k)^T$ are the unknown parameters, and the explanatory variable varies in the interval $T = [0, T_1]$. This model satisfies the assumptions of Theorem 3. Again we consider $k = 2$, and construct locally optimal designs for estimating the derivative of the expected response using the

Table 1. The efficiency of the D-optimal design relative to the optimal design for estimating the derivative of the expected response in the exponential model (3.1) at the point $x = 0$. The design interval is $[0, 1]$, $k = 2$, and various combinations of the nonlinear parameters (b_1, b_2) are considered.

b_1/b_2	0.1	0.2	0.3	0.4	0.5	1	1.5	2	5
0.1	-	0.7234	0.7235	0.7233	0.7233	0.7230	0.7223	0.7212	0.7015
0.2	0.7228	-	0.7226	0.7233	0.7232	0.7235	0.7230	0.7222	0.7049
0.3	0.7233	0.7241	-	0.7238	0.7239	0.7239	0.7238	0.7232	0.7071
0.4	0.7232	0.7233	0.7234	-	0.7220	0.7243	0.7245	0.7241	0.7102
0.5	0.7231	0.7235	0.7237	0.7261	-	0.7248	0.7251	0.7251	0.7126
1	0.7230	0.7235	0.7239	0.7244	0.7248	-	0.7282	0.7295	0.7240
1.5	0.7224	0.7230	0.7238	0.7244	0.7252	0.7281	-	0.7328	0.7287
2	0.7212	0.7222	0.7232	0.7241	0.7251	0.7295	0.7333	-	0.7163
5	0.7015	0.7049	0.7072	0.7102	0.7126	0.7240	0.7287	0.7163	-

functional approach. The design interval is $[0, 1]$. The locally optimal designs are either three or four point designs where, in the latter case, observations have to be taken at 0 and 1. For $\lambda = (1, 0.5, 1, 1)^T$ and $x = 0$, the locally optimal design $\xi^*(0)$ for estimating the derivative of the expected response in the model (4.1) has weights $0.3509, 0.4419, 0.1479$, and 0.0597 at the points $0, 0.0952, 0.4707$, and 1, respectively. For $x < 0.0574$ the optimal design is of the same structure, but for $x > 0.0574$ a three point design is optimal as long as $x < 0.1973$. The weights and points of the optimal design for estimating the derivative in the rational regression model (4.1) are depicted in Figure 3. We observe that the type of design changes several times. In particular the optimal design for the rational regression model (4.1) is supported at four points whenever $x \in [0, 0.0574] \cup [0.1973, 0.2801] \cup [0.6973, 3.0176] \cup [4.4786, \infty)$.

A study of the sensitivity of the locally optimal design to the initial parameters b_1 and b_2 shows similar results as in the exponential case. We observe again that the design is rather stable with respect to the changes in the parameters.

Finally we consider the efficiency of the D-optimal design for estimating the derivative of the expected response in the regression model (4.1). First we fix $\lambda = (1, 0.5, 1, 1)^T$ and take $x \in [0, 6.2]$. The efficiencies are depicted in the right part of Figure 2. For values of x where the type of design changes, efficiencies are smaller than 50%, while the largest efficiencies are approximately 80%. The efficiencies of the D-optimal design when $x = 0$ for various values of the parameters b_1 and b_2 show a similar behaviour as in the case of exponential regression. All efficiencies vary between 70% and 75%.

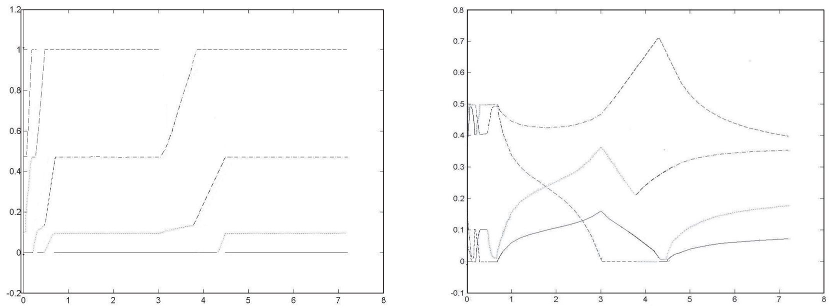

Figure 3. The points (left) and weights (right) of the optimal design for estimating the derivative of the expected response in the rational model (4.1) at a point $x \in [0, 7]$. The design interval is $[0, 1]$, $k = 2$, and $\lambda = (1, 0.5, 1, 1)^T$.

Acknowledgement

The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. 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 BMBF Project SKAVOE and the NIH grant award IR01GM072876:01A1. The work of Viatcheslav Melas and Petr Shpilev was also partly supported by Russian Foundation for Basic Research (RFBR), Project 09-01-00508. The authors would also like to thank two anonymous referees for very constructive comments on an earlier version of this paper.

Appendix: Proof of Theorem 3

We only consider the case $n = 2k-1$ and $t_1^* = T_0$, the other cases are treated similarly. An application of Cauchy's inequality yields

$$\Phi(x,\xi,\lambda) = f'(x,\lambda)^T M^{-}(\xi,\lambda) f'(x,\lambda) = \sup_{q \in \mathbb{R}^{2m}} \frac{(q^T f'(x,\lambda))^2}{q^T M(\xi,\lambda) q} = \frac{(q^*(\xi)^T f'(x,\lambda))^2}{q^*(\xi)^T M(\xi,\lambda) q^*(\xi)},$$

where the last identity defines the vector q^* in an obvious manner and, without loss of generality, $q_{2k}^* = 1$.

Let

$$\bar{\Phi}^{-1}(x,q,\xi,\lambda) = \frac{q^T M(\xi,\lambda) q}{(q^T f'(x,\lambda))^2}, \ q \in \mathbb{R}^{2k},$$

$$\widehat{\Theta} = (q_1, \ldots, q_{2k-1}, t_2, \ldots, t_{2k-1}, \omega_2, \ldots, \omega_{2k-1})^T,$$

$$\Theta = (t_2, \ldots, t_{2k-1}, \omega_2, \ldots, \omega_{2k-1})^T.$$

Note that we consider the case where the point a is a support point of the optimal design. Note also that $\Phi(x,\xi,\lambda) = \Psi(x,q,\xi,\lambda)$, where $q = q^*(\xi)$. Fix λ and write $\bar{\Psi}(x,\bar{\Theta}) = \bar{\Phi}^{-1}(x,q,\xi,\lambda)$. Denote by J the Jacobi matrix of the system of equations in (2.8) and by $\hat{J}$ the Jacobi matrix of the system

$$\frac{\partial \bar{\Psi}(x,\bar{\Theta})}{\partial \widehat{\Theta}} = 0. \tag{A.1}$$

Note that the non-singularity of J follows from the non-singularity of $\hat{J}$. More precisely, we have that $J = G^T \hat{J} G$, where $G^T = (I \vdots R) \in \mathbb{R}^{4k-4 \times 6k-5}$, I is the identity matrix of size $(4k-4) \times (2k-4)$, and R is a $(4k-4) \times (2k-1)$ matrix.

Note that G has full rank. Therefore if $\hat{J}$ is nonsingular, J is also nonsingular.

To show that $\hat{J}$ is nonsingular we use the formulas

$$\frac{\partial^2}{\partial x^2}\left(\frac{U(x,y)}{V(x)}\right) = \frac{\frac{\partial^2 U(x,y)}{\partial x^2}}{V(x)} - 2\frac{\frac{\partial U(x,y)}{\partial x}\frac{\partial V(x)}{\partial x}}{V(x)^2} + 2\frac{U(x,y)\left(\frac{\partial V(x)}{\partial x}\right)^2}{V(x)^3} - \frac{U(x,y)\frac{\partial^2 V(x,y)}{\partial x^2}}{V(x)^2},$$

$$\frac{\partial}{\partial x}\left(\frac{U(x,y)}{V(x)}\right) = \frac{\frac{\partial U(x,y)}{\partial x}}{V(x)} - \frac{U(x,y)\frac{\partial V(x)}{\partial x}}{V(x)^2},$$

$$\frac{\partial^2}{\partial x \partial y}\left(\frac{U(x,y)}{V(x)}\right) = \frac{\frac{\partial^2 U(x,y)}{\partial x \partial y}}{V(x)} - \frac{\frac{\partial U(x,y)}{\partial y}\frac{\partial V(x)}{\partial x}}{V(x)^2}.$$

With the notation $U(q,\Theta) = q^T M(\xi,\lambda) q$, $V(q) = (q^T f'(x,\lambda))^2$, $c_1 = (V(q^*(\xi^*)))^{-1}$, and $c_2 = U(q^*(\xi^*),\Theta^*)c_1$, and observing the condition

$$\frac{\partial}{\partial q}\left(\frac{U(q,\Theta)}{V(q)}\right)\bigg|_{q=q^*} = 0$$

we obtain the identity

$$\frac{\partial^2}{\partial q^2}\left(\frac{U(q,\Theta)}{V(q)}\right)|_{q=q^*} = \frac{\frac{\partial^2 U(q,\Theta)}{\partial q^2}}{V(q)} - \frac{U(q,\Theta)\frac{\partial^2 V(q,\Theta)}{\partial q^2}}{V(q)^2}\bigg|_{q=q^*}.$$

Similarly, the condition

$$\frac{\partial}{\partial \Theta}\left(\frac{U(q,\Theta)}{V(q)}\right)\bigg|_{\Theta=\Theta^*} = \frac{\frac{\partial U(q,\Theta)}{\partial \Theta}}{V(q)}\bigg|_{\Theta=\Theta^*} = 0$$

yields

$$\frac{\partial^2}{\partial q \partial \Theta}\left(\frac{U(q,\Theta)}{V(q)}\right)\bigg|_{\widehat{\Theta}=\widehat{\Theta}^*} = c_1 \frac{\partial^2 U(q,\Theta)}{\partial q \partial \Theta}\bigg|_{\widehat{\Theta}=\widehat{\Theta}^*},$$

The derivatives can now be easily calculated, namely

$$\frac{\partial^2}{\partial q^2}\left(\frac{U(q,\Theta)}{V(q)}\right)\bigg|_{\widehat{\Theta}=\widehat{\Theta}^*} = c_1 M(\xi^*,\lambda) - c_2 c_1 f'(x) f'(x)^T.$$

We now prove that this matrix is nonnegative definite. Note that for any vector p such that $p \neq q^*(\xi^*)$ and $p^T f'(x) f'(x)^T p \neq 0$,

$$p^T(c_1 M(\xi^*,\lambda) - c_2 c_1 f'(x) f'(x)^T)p = c_1 (p^T f'(x))^2 \left(\frac{p^T M(\xi^*,\lambda)p}{(p^T f'(x))^2} - c_2\right) > 0.$$

In particular, if $p = q^*(\xi^*)$ then it follows that

$$c_1 q^*(\xi^*)^T M(\xi^*,\lambda) q^*(\xi^*) - c_2 c_1 (q^*(\xi^*)^T f'(x))^2 = c_2 - c_2 = 0.$$

Consequently, the Jacobi matrix is given by

$$\hat{J} = \begin{pmatrix} D & c_1 B_1^T & c_1 B_2^T \\ c_1 B_1 & c_1 E & 0 \\ c_1 B_2 & 0 & 0 \end{pmatrix},$$

where D is obtained from the matrix $\widehat{D} = c_1(M(\xi^*,\lambda) - c_2 f'(x) f'(x)^T) \geq 0$ by deleting the last column and row, and B_1, B_2, and E are the same as in Dette, Melas and Pepelyshev (2004b, p.208). In that paper a polynomial regression is considered, but all arguments require only the Chebyshev properties of polynomials.

Repeating the arguments from that paper we find that $\hat{J}$ is a nonsingular matrix. The assertion of the theorem follows.

References

Atkinson, A. C. (1970). The design of experiments to estimate the slope of a response surface. *Biometrika* **57**, 319-328.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. *J. Statist. Plann. Inference* **21**, 191-208.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.

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

Dette, H., Melas, V. B. and Pepelyshev, A. (2004a). Optimal designs for a class of nonlinear regression models. *Ann. Statist.* **32**, 2142-2167.

Dette, H., Melas, V. B. and Pepelyshev, A. (2004b). Optimal designs for estimating individual coefficient in polynomial regression - a functional approach. *J. Statist. Plann. Inference* **118**, 201-219.

Dette, H., Melas, V. B. and Pepelyshev, A. (2006). Locally E-optimal designs for exponential regression models. *Ann. Inst. Statist. Math.* **58**, 407-426.

Dette, H., Melas, V. B. and Pepelyshev, A. (2009). Optimal designs for estimating the slope of a regression. To appear in *Statistics*.

Fellmann, J. (1974). On the allocation of linear observations. *Societas Scientiarum Fennica Commentationes Physico-Mathematicae* **44**, 27-78.

Gunning, R. C. and Rossi, H. (1965). *Analytic Functions of Several Complex Variables.* Prentice-Hall, New Jersey.

He, Z., Studden, W. J. and Sun, D. (1996). Optimal designs for rational models. *Ann. Statist.* **24**, 2128-2142.

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

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12**, 363-366.

Melas, V., Pepelyshev, A. and Cheng, R. (2003). Designs for estimating an extremal point of quadratic regression models in a hyperball. *Metrika* **58**, 193-208.

Melas, V. B. (2006). *Functional Approach to Optimal Experimental Design.* (Lecture Notes in Statistics 184). Springer, New York.

Mukerjee, R. and Huda, S. (1985). Minimax second- and third-order designs to estimate the slope of a response surface. *Biometrika* **72**, 173-178.

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

Murty, V. and Studden, W. J. (1972). Optimal designs for estimating the slope of a polynomial regression. *J. Amer. Statist. Assoc.* **67**, 869-873.

Petrushev, P. P. and Popov, V. A. (1987). *Rational Approximation of Real Functions.* Cambridge University Press, Cambridge.

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

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

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach.* Marcel Dekker, New York.

Studden, W. J. (1968). Optimal designs on Tchebycheff points. *Ann. Math. Statist.* **39**, 1435-1447.

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.

E-mail: holger.dette@rub.de

Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia.

E-mail: vbmelas@post.ru

Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia.

E-mail: pitshp@hotmail.com

(Received August 2009; accepted August 2010)