Optimal designs for estimating the slope in nonlinear regression

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: v.melas@pobox.spbu.ru

Petr Shpilev
St. Petersburg State University
Department of Mathematics
St. Petersburg , Russia
email: pitshp@hotmail.com

August 20, 2009

Abstract

We consider the problem of estimating the slope of the expected response in nonlinear regression models. It is demonstrated that in many cases the optimal designs for estimating the slope are either on k or $k-1$ points, where k denotes a 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.

Keywords and phrases: C-optimal designs, nonlinear regression, implicit function theorem, Chebyshev system

AMS subject classification: 62K05

1 Introduction

A common tool in statistical inference are nonlinear regression models, which are widely used to describe the dependencies between a response and an explanatory variable [see e.g. Seber and Wild (1989), Ratkowsky (1983) or Ratkowsky (1990)]. 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), Melas (1978) for early references and Ford et al. (1992), He et al. (1996), Dette et al. (1999) for more recent references on local optimal designs]. Robust design strategies have been proposed by Pronzato and Walter (1985) and Chaloner and Larntz (1989), Dette (1995), Müller and Pázman (1998) using a Bayesian or minimax approach. Most of the literature concentrates on optimal designs (independent of the particular approach) maximizing a functional of the Fisher information matrix for the parameters in the model. This approach is somehow related to the problem of estimating the response function most precisely. However in many experiments differences in the response will often be of more importance than the absolute response. In such case, in particular, if one is interested in a difference at two points close together, a precise estimation of the slope is of particular interest and often one of the main objectives of the statistical inference in the experiment.

The present paper is devoted to the problem of optimal designing experiments for estimating the slope of the expected response in a nonlinear regression model. Pioneering work in this direction has been done by Atkinson (1970) and the problem has subsequently been taken up by many other authors [see e.g. Ott and Mendenhall (1972), Murty and Studden (1972), Myres and Lahoda (1975), Hader and Park (1978), Mukerjee and Huda (1985), Mandal and Heiligers (1992), Pronzato and Walter (1993) and Melas et al. (2003)]. While most of these papers consider linear regression models, the present paper takes 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 $I \subset \mathbb{R}$, 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 (1978), Dette et al. (2006) and Biedermann et al. (2007), who considered the case of exponential models, that is

$$\varphi(t, b_i) = \exp(b_i t) \ . \tag{1.2}$$

These models have numerous applications in environmental and ecological experiments, toxicology and pharmacokinetics [see for example Landaw and DiStefano (1984), Becka and

Urfer (1996) or Becka et al. (1993), among many others]. For the choice

$$\varphi(t, \lambda) = \frac{1}{b_i + t} \, , \tag{1.3}$$

one obtains a class of rational regression models, which are very popular because they have appealing approximation properties [see Petrushev and Popov (1987) for some theoretical properties and Dudzinski and Mykytowycz (1961), Ratkowsky (1983), p.120 for an application of this model]. Optimal design problems for parameter estimation of the coefficients $a_1, \dots, a_k$ have been discussed in Imhof and Studden (2001), who assumed that the nonlinear parameters $b_1, \dots, b_k$ are known and do not have to be estimated. Optimal design problems for estimating the full vector of parameters $a_1, \dots, a_k, b_1, \dots, b_k$ have been determined in Dette et al. (2004a).

While all papers cited in the previous paragraph have their focus on the estimation of parameters, which is related to the estimation of the expected response, the present work considers the problem of designing experiments for the estimation of the slope of the expected response in models of the form (1.1) at a given point x. In Section 2 we introduce the necessary notation and present some general results about designing experiment for estimating the slope, if the components of the gradient of the expected response (with respect to the unknown parameters) form a Chebyshev system. 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 slope has to be estimated. This result is used to provide a Taylor expansion for the weights and support points as functions of the point x, which can easily be used for the numerical calculation of the optimal designs. Section 3 considers the case, where the function φ is given by (1.2), while the rational functions are discussed in Section 4. We use the general method to determine numerically the optimal design for estimating the slope and study their properties as functions of the unknown parameters and of the point, where the slope has to be estimated. In particular, it is shown that the optimal designs for estimating the slope of the expected response at the point x have either $2m$ or $2m-1$ support points, and this property changes with the value of x. On the other hand, the locally optimal designs are rather robust with respect to changes in the nonlinear parameters.

2 Optimal designs for estimating the slope

Consider the regression model defined by (1.1), where the design space is given by the interval $T = [0, T_1]$, where $T_1 \in (0, \infty)$. We assume that - in principle - for each $t \in T$ an observation Y could be made, where different observations are assumed to be independent with the same variance, say $\sigma^2 > 0$. Following Kiefer (1974) we call any probability measure

$$\xi = \begin{pmatrix} t_1 & \dots & t_{n-1} & t_n \\ \omega_1 & \dots & \omega_{n-1} & \omega_n \end{pmatrix} , \tag{2.1}$$

with finite support $t_1, \dots, t_n \in T$, $t_i \neq t_j (i \neq j)$ 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, $i = 1, 2, \dots, n$ 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 defined by

$$M(\xi, \lambda) = \int_0^\infty f(t, \lambda) f^T(t, \lambda) d\xi(t) \ , \tag{2.2}$$

where

$$f(x, \lambda) = \frac{\eta(t, \lambda)}{\partial \lambda} = (f_1(t, \lambda), \dots, f_{2m}(t, \lambda))^T \tag{2.3}$$

is the vector of partial derivatives of the response function with respect to the parameter $\lambda = (a_1, \dots, a_k, b_1, \dots b_k)^T$. It has been shown by Jennrich (1969) that for uncorrelated observations (obtained from approximate design 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 which can be used to discriminate among competing designs [see Silvey (1980) or Pukelsheim (1993)].
Most of these criteria, reflect the problem of efficient parameter estimation. If the estimation of the slope $\eta'(x, \lambda)$ is of interest, a common estimate is given by

$$\hat{\eta} = \eta'(x, \hat{\lambda}),$$

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 proportional to

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

where it is assumed that the vector $f'(x, \lambda) = (f_1'(x, \lambda), \dots, f_{2m}'(x, \lambda))^T$ is estimable by the design ξ, i.e. $f'(t, \lambda) \in \mathrm{Range}\ (M(\xi, \lambda))$, and

$$f'(x, \lambda) = \frac{\partial}{\partial x} f(x, \lambda) \tag{2.4}$$

denotes the derivative of the vector f with respect to x. Throughout this paper we define

$$\Phi(x, \xi, \lambda) = \begin{cases} (f'(x, \lambda))^T M^-(\xi, \lambda) f'(x, \lambda) & \text{if} \quad f'(x, \lambda) \in \mathrm{Range}\ (M(\xi, \lambda)), \\ \infty & \text{else.} \end{cases} \tag{2.5}$$

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 $f'(x,\lambda) \in \text{Range}\,(M(\xi,\lambda))$ optimal for estimating the slope of expected response in model (1.1). Note that the criterion (2.5) 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 [see e.g. Studden (1968), Ford et al. (1992), Studden (2005) among many others]. Moreover, the criterion depends on the parameter λ and following Chernoff (1953), we assume that a preliminary guess for this parameter is available. This corresponds to the concept of locally optimal designs, which are used as benchmarks of commonly applied designs and form the basis for many optimal designs with respect to more sophisticated optimality criteria.

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

$$\varepsilon \cdot \begin{vmatrix} g_1(t_1) & \dots & g_1(t_m) \\ \vdots & \ddots & \vdots \\ g_m(t_1) & \dots & g_m(t_m) \end{vmatrix} \geq 0 \tag{2.6}$$

for all $t_1,\dots,t_m \in I$ with $t_1 < t_2 < \dots < t_m$. If the inequality in (2.6) is strict, then $\{g_1,\dots,g_m\}$ is called Chebyshev system. It is well known [see Karlin and Studden (1966), Theorem II 10.2] that if $\{g_1,\dots,g_m\}$ is a weak Chebyshev system, then there exists a unique function

$$\sum_{i=1}^{m} c_i^* g_i(t) = c^{*T} g(t), \tag{2.7}$$

with $g^T(t) = (g_1(t),\dots,g_m(t))$ and the following properties

$$\text{(2.8)}\quad \begin{array}{ll} \text{(i)} & |c^{*T} g(t)| \leq 1 \quad \forall\, t \in T \\ \text{(ii)} & \text{there exist } m \text{ points } s_1 < \dots < s_m \text{ such that } c^{*T} g(s_i) = (-1)^i \quad i = 1,\dots,m. \end{array}$$

The function $c^{*T} g(t)$ is called Chebyshev polynomial, the points $s_1,\dots,s_m$ are called Chebyshev points and need not to be unique. They are unique if $1 \in \text{span}\{g_1,\dots,g_m\}, m \geq 1$ and I is a bounded and closed interval, where in this case $s_1 = \min_{x\in I} x$, $s_m = \max_{x\in I} x$. It is well known [see Studden (1968)] that in many cases c-optimal designs are supported at Chebyshev points. Recall that the functions $g_1(x),\dots,g_m(x)$ generate an extended Chebyshev system of order 2 on the set $\mathcal{Z} = [a,b] \cup [a',b']$ if and only if

$$U^* \begin{pmatrix} 1 & \dots & m \\ x_1 & \dots & x_m \end{pmatrix} > 0$$

for all $x_1 \leq \cdots \leq x_m$ ($x_j \in \mathcal{X}$; $j = 1, \ldots, m$) where equality occurs at most at 2 consecutive points x_j, the determinant U^* is defined by

$$U^* \begin{pmatrix} 1 & \ldots & m \\ x_1 & \ldots & x_m \end{pmatrix} = \det(g(x_1), \ldots, g(x_m))$$

and the two columns $g(x_i), g(x_{i+1})$ are replaced by $g(x_i), g'(x_{i+1})$ if the points x_i and x_{i+1} coincide. Note that under this assumption any linear combination

$$\sum_{i=1}^{m} \alpha_i g_i(x)$$

($\alpha_1, \ldots, \alpha_m \in \mathbb{R}, \sum_{i=1}^m \alpha_i^2 \neq 0$) has at most $m-1$ roots, where multiple roots are counted twice [see Karlin and Studden (1966), Ch. 1]. We begin with a result which shows that the optimal design for estimating the slope in the nonlinear regression model (1.1) only depends on the "nonlinear" parameters $b_1, b_2 \ldots, b_m$ of the model.

Lemma 2.1 *In the nonlinear regression model (1.1) the optimal design for estimating the slope of the expected response at a point x does not depend on the parameters $a_1, \ldots, a_m$.*

Proof. With the notation $\lambda_1 = (1, b_1, 1, b_2, \ldots, 1, b_m)^T$ we obtain, observing the definition of the vectors $f(x, \lambda)$ and $f'(x, \lambda)$ in (2.3) and (2.4),

$$f(x, \lambda) = L_a f(x, \lambda_1) \ , \ f'(x, \lambda) = L_a f'(x, \lambda_1),$$

where the matrix L_a is given by

$$L_a = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & a_m \end{pmatrix}.$$

By the definition of the information matrix this yields for $f'(x, \lambda) \in$ Range $(M(\xi, \lambda))$

$$\begin{aligned} \Phi(x, \xi, \lambda) &= (f'(x, \lambda))^T M^-(\xi, \lambda) f'(x, \lambda) = (f'(x, \lambda_1))^T L_a^T (L_a^T)^- M^-(\xi, \lambda_1) L_a^- L_a f'(x, \lambda_1) \\ &= (f'(x, \lambda_1))^T M^-(\xi, \lambda_1) f'(x, \lambda_1) = \Phi(x, \xi, \lambda_1), \end{aligned}$$

which proves the assertion of the Lemma. □

Our next result specifies the number of support points of the locally optimal design for estimating the slope of the expected response in the nonlinear regression model (1.1). A similar result was recently derived in a paper of Dette et al. (2009) and the following proof is obtained by exactly the same arguments.

Theorem 2.1 *Assume that the functions $f_1, \ldots, f_{2m}$ defined in (2.3) form an extended Chebyshev system of second order on the interval T, then the number of support points of any optimal design for estimating the slope of the expected response in the nonlinear regression model (1.1) is at least $2m-1$. Moreover, if the number of support points is $2m$, then these points must be Chebyshev points defined by (2.8). In this case at least one of these points coincides with a boundary of the design interval. If the constant function is an element of span$\{f_1, \ldots, f_{2m}\}$ then the number of support points is at most $2m$.*

Remark 2.1 If the design has $2m$ support points it follows by standard arguments of optimal design theory [see for example Pukelsheim and Torsney (1991)] that the weights at the support points are given by

$$\omega_i^* = \frac{|e_i^T F^{-1} f'(x, \lambda)|}{\sum_{i=1}^{2m} |e_i^T F^{-1} f'(x, \lambda)|}, \quad i = 1, \ldots, 2m$$

where the matrix F is given by $F = (f(s_1, \lambda), \ldots, f(s_{2m}, \lambda))$ and $s_1, \ldots, s_{2m}$ denote the support points of the optimal design (i.e. the Chebyhev points of the system $f_1, \ldots, f_{2m}$). Moreover, it is also worthwhile to mention that in this case if follows from Theorem 2.1 that the support points do not depend on the particular point x, where the estimation of the slope has to be performed.

For the construction of the locally optimal designs for estimating the slope we use the functional approach, which is described in Melas (2006) and allows us to calculate support points and weights of the optimal design ξ_x^* for estimating the slope as a function of the point x, where the estimate of the slope of the regression is required. To be precise, we assume that the number of support points of the design ξ_x^* is constant, say $n \in \mathbb{N}$, the smallest support point, say t_1, is equal to the left boundary of the design interval T and the largest support point, say t_n, is an interior point of the design space, if x is contained in some interval, say $[a^*, b^*) \subseteq [0, \infty)$. All other cases can be considered in a similar way. We collect the information of the design ξ_x^* given by its support points $t_i^*(x)$ and its weights $\omega_i^*(x)$ $(i = 1, \ldots, n)$ in a vector valued function

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

and consider a system of equations

$$\frac{\partial \Phi(x, \xi, \lambda)}{\partial \Theta} = 0. \tag{2.9}$$

From the necessary conditions for an extremum it follows that the vector valued function $\Theta^*(x)$ corresponding to the optimal design ξ_x^* for estimating the slope in the nonlinear regression (1.1) is a solution of the system (2.9). The Jacobi matrix of this system is given by

$$J(x,\xi) = \left(\frac{\partial^2}{\partial \Theta_i \partial \Theta_j} \Phi(x,\xi,\lambda) \right)_{i,j=1}^{2n-2} \in \mathbb{R}^{(2n-2)\times(2n-2)}. \tag{2.10}$$

If the Jacobi matrix is nonsingular, for some point x_0, then we obtain by a straightforward application of the implicit function theorem [see e.g. Gunning and Rossi (1965)] that in a neighbourhood of this point there exists an analytic function $\Theta^*(x)$, which is a solution of system (2.9) and corresponds to the locally optimal design for estimating the slope 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 neighborhood of point x_0. The coefficients of this expansion can be determined recursively as stated in the following theorem, which has been proved by Dette et al. (2004b).

Theorem 2.2 *If the Jacobi matrix defined in (2.10) is a nonsingular matrix at some point $x_0 \in (-\infty,\infty)$, then the coefficients $\Theta^*(j,x_0)$ of the Taylor expansion of the vector valued function*

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

can be obtained recursively in the neighborhood of point x_0, that is

$$\Theta^*(s+1,x_0) = -J^{-1}(x_0,\xi_{x_0}^x) \left(\frac{d}{dx} \right)^{s+1} h(\widetilde{\Theta}^*_{(s)}(x),x) \,|_{x=x_0}, \quad s = 0,1,2,\dots$$

where the polynomial $\widetilde{\Theta}^*_{(s)}(x)$ of s-th degree is defined by

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

and the function h is given by

$$h(\widetilde{\Theta},x) = \frac{\partial}{\partial \Theta} \Phi(x,\xi,\lambda) \,|_{\Theta=\widetilde{\Theta}} \,.$$

In the following result we will prove that the Jacobi matrix of the system (2.9) is nonsingular if the number of support points of the locally optimal design for estimating the slope in the regression model (1.1) is $2m-1$ or $2m$. Note that there always exists an optimal design for estimating the slope of the expected response with at most $2m$ support points [see for example Pukelsheim (1993), p. 190], and consequently by Theorem 2.1 there exist optimal designs with $2m-1$ or $2m$ support points. As a consequence, 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 slope of the expected response at the point x, can be obtained by the recursive formulas stated in the previous theorem.

Theorem 2.3 *Consider the nonlinear regression model (1.1) with corresponding system (2.9). The Jacobi matrix of this system is nonsingular, whenever the optimal design for estimating the slope in the nonlinear regression model* (1.1) *on the design space* $T = [a, b]$, *where* $a < b$ *are arbitrary numbers, has* $n = 2m$ *or* $2m - 1$ *support points.*

Proof. We only consider the case $n = 2m - 1$, $t_1^* = a$, the other cases are treated similarly. An application of Cauchys 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 we put without loss of generality $q_{2m}^* = 1$.
Let us introduce the notation

$$\begin{aligned} \Psi(x, q, \xi, \lambda) &= \frac{q^T M(\xi, \lambda) q}{(q^T f'(x, \lambda))^2}, \ q \in \mathbb{R}^{2m} \\ \widehat{\Theta} &= (q_1, \ldots, q_{2m-1}, t_2(x), \ldots, t_{2m-1}(x), \omega_2(x), \ldots, \omega_{2m-1}(x))^T, \\ \Theta &= (t_2(x), \ldots, t_{2m-1}(x), \omega_2(x), \ldots, \omega_{2m-1}(x))^T. \end{aligned}$$

(note that we consider the case where the point a is a support point of the optimal design). Note that $\Phi(x, \xi, \lambda) = \Psi(x, q, \xi, \lambda)$, where $q = q^*(\xi)$. Denote by J the Jacobi matrix of the system of equations

$$\frac{\partial \Phi(x, \xi, \lambda)}{\partial \Theta} = 0. \tag{2.11}$$

and by $\hat{J}$ the Jacobi matrix of the system

$$\frac{\partial \Psi(x, q, \xi, \lambda)}{\partial \widehat{\Theta}} = 0. \tag{2.12}$$

Note that the non-singularity of J follows from the non-singularity of $\hat{J}$. More precisely, we have by the formula for the derivative that

$$J = G^T \hat{J} G,$$

where $G^T = (I \vdots R) \in \mathbb{R}^{4m-4 \times 6m-5}$, I is the identity matrix of size $(4m - 4) \times (2m - 4)$ and R is a $(4m - 4) \times (2m - 1)$ matrix. Note that both matrices J and $\hat{J}$ are nonnegative definite (since they correspond to a local maximum of the determinant of the information matrix).
Suppose that J is a singular matrix. Then there exists a vector $c \neq 0$ such that $c^T J c = 0$, and due to the above formula, there exists vector $b = Gc \neq 0$ and such $b^T \hat{J} b = 0$. Therefore it follows that if the matrix $\hat{J}$ is nonsingular then the matrix J is also nonsingular.

In order to prove that the matrix $\hat{J}$ 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 $U(q,\Theta) = q^T M(\xi,\lambda)q$, $V(q) = (q^T f'(x,\lambda))^2$, $c_1 = (V(q^*(\xi^*)))^{-1}$, $c_2 = U(q^*(\xi^*),\Theta^*)c_1$ we obtain observing the condition

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

the identity

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

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, that is

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

We now prove that this matrix is nonnegative definite. For this purpose note that we have 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 $\widehat{D} = c_1(M(\xi^*, \lambda) - c_2 f'(x) f'(x)^T) \geq 0$ deleting the last column and row and matrices B_1, B_2 and E are the same as in Dette et al. (2004b), p.208, formula (3.18). 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 obtain that the matrix $\hat{J}$ is a nonsingular matrix. Consequently the assertion of the Theorem follows. □

If the assumptions of Theorem 2.2 and Theorem 2.3 are satisfied, the functional approach can be easily used for constructing any optimal design for estimating the slope in the nonlinear regression model (1.1). In the following sections we will illustrate this concept in two concrete examples.

3 Optimal designs for estimating the slope in exponential regression models

For the special choice (1.2) the nonlinear regression model reduced to the exponential regression model

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

where $\lambda = (a_1, b_1, a_2, b_2, \ldots, a_m, b_m)^T$ denotes the vector of unknown parameters and the explanatory variable varies in the interval $T = [0, T_1]$. It is easy to see that this model satisfies the assumptions of Theorem 2.3. In order to illustrate the general procedure we have considered the model (3.1) for $m = 2$ and have constructed locally optimal designs for estimating the slope in this model by the functional approach, which was described in previous section. The vector of parameters is given by $\lambda = (1, 0.5, 1, 1)^T$ and the design interval is $T = [0, 1]$. As it was pointed out in the previous section, there exist two types of optimal designs, namely a design with four support points including the boundary points of the design space, i.e.

$$\xi^*(x) = \begin{pmatrix} 0 & t_2^*(x) & t_3^*(x) & 1 \\ \omega_1^*(x) & \omega_2^*(x) & \omega_3^*(x) & \omega_4^*(x) \end{pmatrix},$$

where $0 < t_2^*(x) < t_3^*(x) < 1$ and the weights are described in Remark 2.1. The design of the second type has only three support points and is of the form

$$\xi^*(x) = \begin{pmatrix} t_1^*(x) & t_2^*(x) & t_3^*(x) \\ \omega_1^*(x) & \omega_2^*(x) & \omega_3^*(x) \end{pmatrix},$$

where $0 \leq t_1^*(x) < t_2^*(x)) < t_3^*(x) \leq 1$. In a first step we have to find the optimal design for the estimation of the slope in the exponential regression model (3.1) at a particular point, and we chose $x_0 = 0$ for this purpose. The optimal design is of the first type and given by

$$\xi^*(0) = \begin{pmatrix} 0 & 0.3011 & 0.7926 & 1 \\ 0.3509 & 0.4438 & 0.1491 & 0.0562 \end{pmatrix}.$$

By Theorem 2.2 the design is of this form in a neighbourhood of the point x_0, where the support remains unchanged. Therefore, we can use the representation of the weights in Remark 2.1 to determine the point, where the type of the design changes. To be precise we determine the minimal point $x_1 > x_0 = 0$ such that one of the equations

$$\omega_i^*(x) = \frac{|e_i^T F^{-1} f'(x, \lambda)|}{\sum_{i=1}^m |e_i^T F^{-1} f'(x, \lambda)|} \;=\; 0,$$

$(i = 1, \ldots, 4)$ is satisfied, which yields $x_1 = 0.1457165222$. In the interval $I_0 = [x_0, x_1)$ the Jacobi matrix of the system (2.11) is non-singular and therefore we can use the formulas from the Theorem 2.2 to determine the coefficients in the Taylor expansion of the function $\Theta^*(x)$. Note that there exists an interval $I_1 = (x_1, x_2)$ such that for $x \in I_1$, the optimal design for estimating the slope in the exponential regression model (3.1) at the point x is of type 2 and 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 3 as a function of the point x, where the slope has to be estimated. We observe that the type of design changes several times, when x varies in the interval $[0, 2.7]$. In particular it is of type one if and only if $x \in [0, 0.1457165222] \cup [0.500137, 0.587461] \cup [0.9092241459, 2.7]$

In the previous example the vector of parameters required for the calculation of the locally optimal design was fixed and we have varied the point x, where the estimation of the slope should be performed. In order to study the sensitivity of the locally optimal design with respect to the choice of the initial parameters we next construct optimal designs on the interval $[0, 1]$ for estimating the slope of the expected response in the nonlinear regression model (3.1) at the point $x = 1$, where the parameter b_1 varies in the interval $[0.1, 4]$ and the parameter $b_2 = 1$ is fixed such that $(b_1 \neq b_2)$. The weights and points of the locally

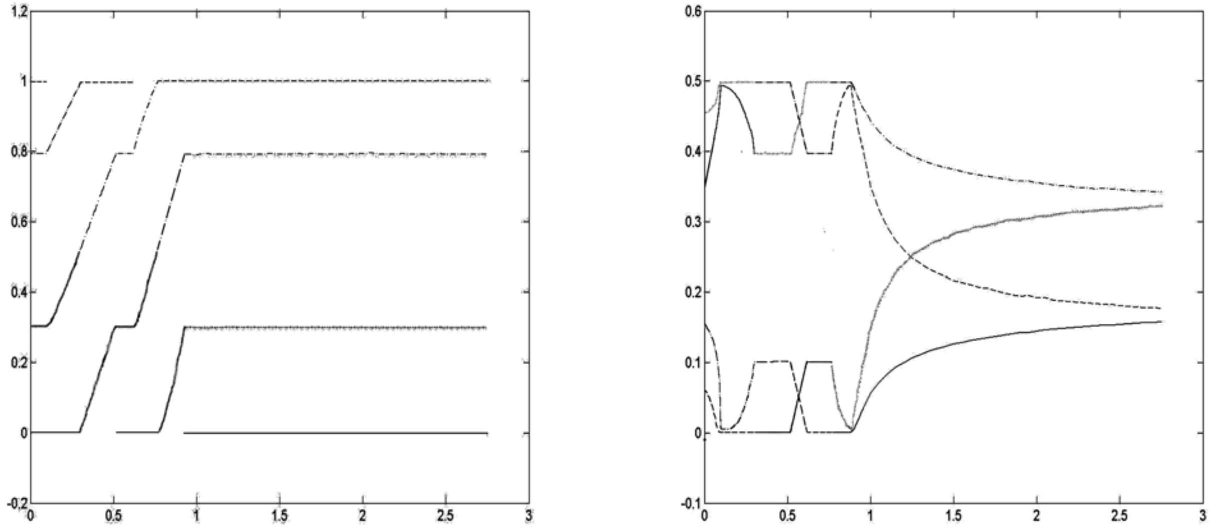

Figure 1: *The points (left) and weights (right) of the optimal design for estimating the slope of the expected response in the nonlinear regression model* (3.1) *at the point* $x \in [0, 2.7]$. *The design interval is given by* $[0, 1]$, $m = 2$, *and vector of parameters is* $\lambda = (1, 0.5, 1,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

optimal design for estimating the slope of the expected response in the nonlinear regression model (3.1) at the point $x = 1$ are depicted in Figure 2. We observe that the locally optimal designs are rather robust with respect to changes in the initial parameter b_1. In particular the weights are nearly not changing, while there appear small changes in the interior support points.

The D-optimal design is efficient for estimating the parameters. By the famous equivalence theorem of Kiefer and Wolfowitz (1960) it is also (minimax-) efficient for estimating the expected response. Therefore, it is of particular interest to investigate the efficiency of the this design for estimating the slope of the expected response. To be precise, we define the function

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

which is called the efficiency of the design ξ_1 relative to the design ξ_2 for estimating the slope of the expected response in the nonlinear regression model (1.1). Note that these efficiencies will depend on the particular point x, where the estimation of the slope is performed, and on the nonlinear parameters in the model. We first fix the vector of parameters, say $\lambda = (1, 0.5, 1, 1)^T$ and vary the point x. The corresponding efficiencies of the D-optimal design

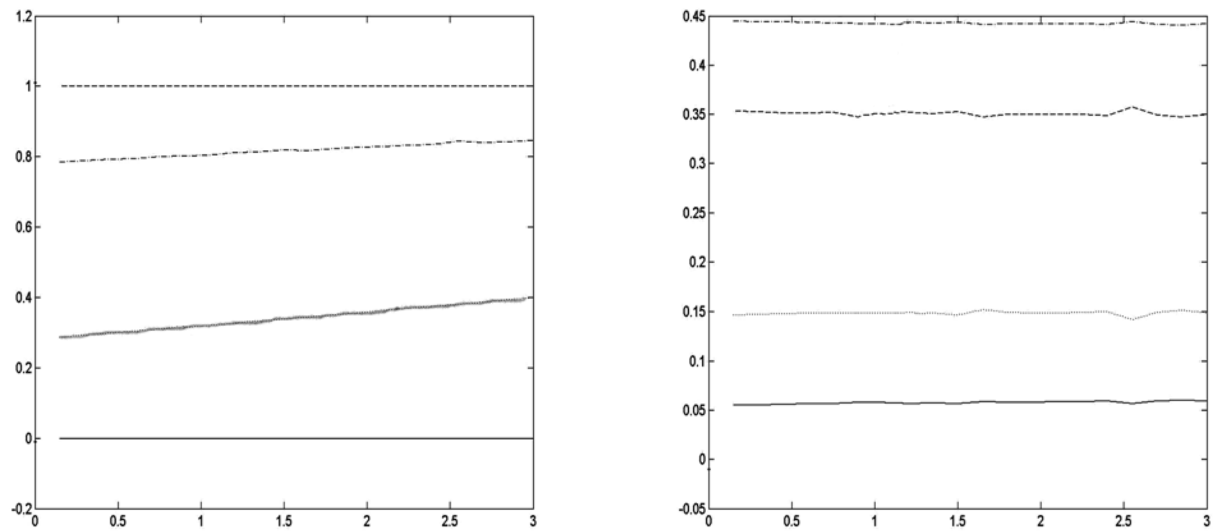

Figure 2: *The points (left) and weights (right) of the optimal design for estimating the slope of the expected response in the nonlinear regression model* (3.1) *at the point* $x = 1$. *The design interval is given by* $[0, 1]$, $m = 2$, *and vector of parameters is* $\lambda = (1, b_1, 1, 1)^T$, *where* b_1 *varies in the interval* $[0.1, 4]$.

for estimating the slope of the expected response in the regression model (3.1) are depicted in Figure 3. We observe that the efficiency is first decreasing to values smaller than 55%, but for larger x the D-optimal design is rather efficient for estimating the slope of the expected response in the regression model (3.1). It is interesting to note that the lowest efficiencies are obtained for those values of x, where the design moves (as a function of x) from a type one design to a type two design. 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 given by 72% and do not change substantially with (b_1, b_2).

4 Optimal designs for estimating the slope in rational regression models

For the special choice (1.3) the nonlinear regression model (1.1) reduces to the rational regression model, that is

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

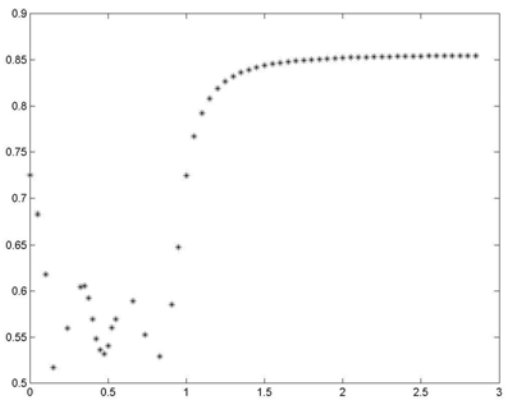

Figure 3: *The efficiency of the D-optimal design relative to the optimal design for estimating the slope of the expected response in the regression model* (3.1) *at the point* $x \in [0, 2.7]$. *The design interval is given by* $[0, 1]$, $m = 2$, *and the vector of parameters is given by* $\lambda = (1, 0.5, 1, 1)$.

where $\lambda = (a_1, b_1, a_2, b_2 \ldots, a_m, b_m)^T$ are the unknown parameters and the explanatory variable varies in the interval $T = [0, T_1]$. This model satisfies the assumptions of Theorem 2.3. Again we consider the model of second order, i.e. $m = 2$, and construct locally optimal designs for estimating the expected response using the functional approach. The design interval is given by $[0, 1]$. The locally optimal designs are either three or four point designs, where in the latter case observations have to be taken at the boundary of the design interval. For the vector $\lambda = (1, 0.5, 1, 1)^T$ and the point $x = 0$ the locally optimal design for estimating the slope of the expected response in the model (4.1) is given by

$$\xi^*(0) = \begin{pmatrix} 0 & 0.09519 & 0.47065 & 1 \\ 0.35088 & 0.44128 & 0.14785 & 0.05999 \end{pmatrix}$$

If $x < 0.0574321$ the optimal design is of the same structure, but for $x > 0.0574321$ a three point design is optimal as long as $x < 0.1973301$. The weights and points of the optimal design for estimating the slope in the rational regression model (4.1) are depicted in Figure 4. We observe that the type of design (3 or 4 support points) is changing several times. In particular the optimal design for estimating the slope in the expected response of the rational regression model (4.1) is supported at 4 points, whenever $x \in [0, 0.0574321] \cup [0.1973301, 0.2801163] \cup [0.69737, 3.01762] \cup [4.478661, \infty)$.

Next we study the sensitivity of the locally optimal design for estimating the slope on the initial parameters b_1 and b_2. Similarly as in the exponential case we fix the point where the slope has to be estimated, i.e. $x = 0$, and vary the parameter b_1 in the interval $[0.1, 4]$. The weights and support points of the optimal design for estimating the slope of the expected

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	-

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

response in the rational regression model (4.1) are depicted in Figure 4. We observe again that the design is rather stable with respect to the changes in the parameter b_1.

Finally we consider the efficiency of the D-optimal design for estimating the slope of the expected response in the regression model (4.1). First we fix the vector of parameters $\lambda = (1, 0.5, 1, 1)^T$ and consider the efficiency of the D-optimal design for estimating the slope in the rational regression at the point $x \in [0, 6.2]$. These efficiencies are depicted in Figure 6. For values of x, where the design changes from type one to type two, the efficiencies are smaller than 50%, while the largest efficiencies are approximately 80%. The efficiencies of the D-optimal design for estimating the slope of the expected response at the point $x = 0$ for various values of the parameters b_1 and b_2 are shown in Table 2. We observe again that there are no substantial changes in the efficiencies for different parameters (b_1, b_2). All efficiencies vary between 70% and 75%.

Acknowledgements 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.

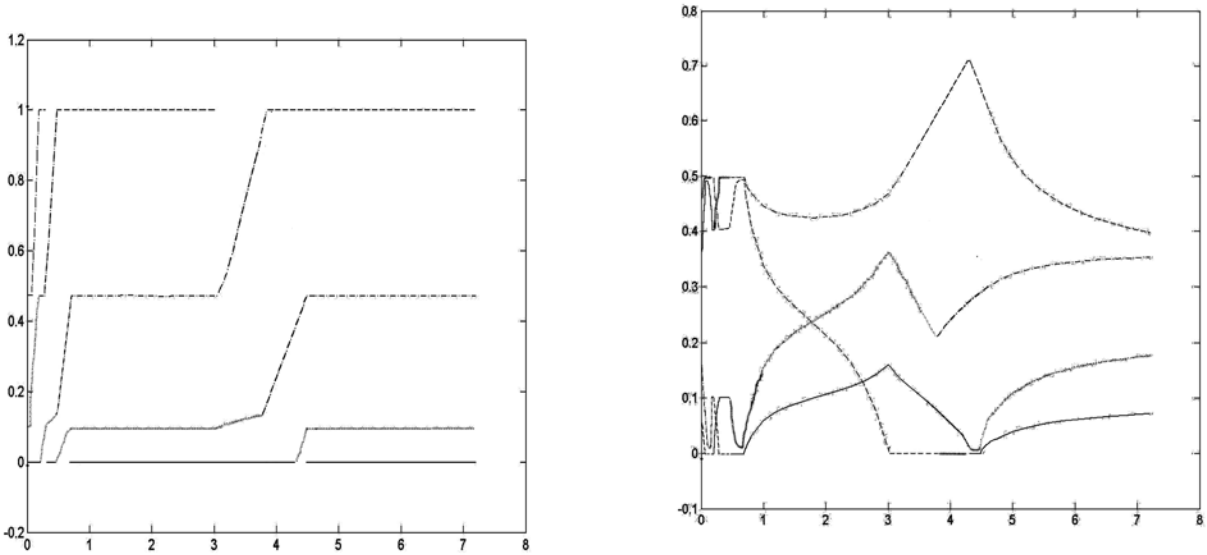

Figure 4: *The points (left) and weights (right) of the optimal design for estimating the slope of the expected response in the nonlinear regression model* (4.1) *at the point* $x \in [0, 7]$. *The design interval is given by* $[0, 1]$, $m = 2$, *and vector of parameters is* $\lambda = (1, 0.5, 1, 1)^T$.

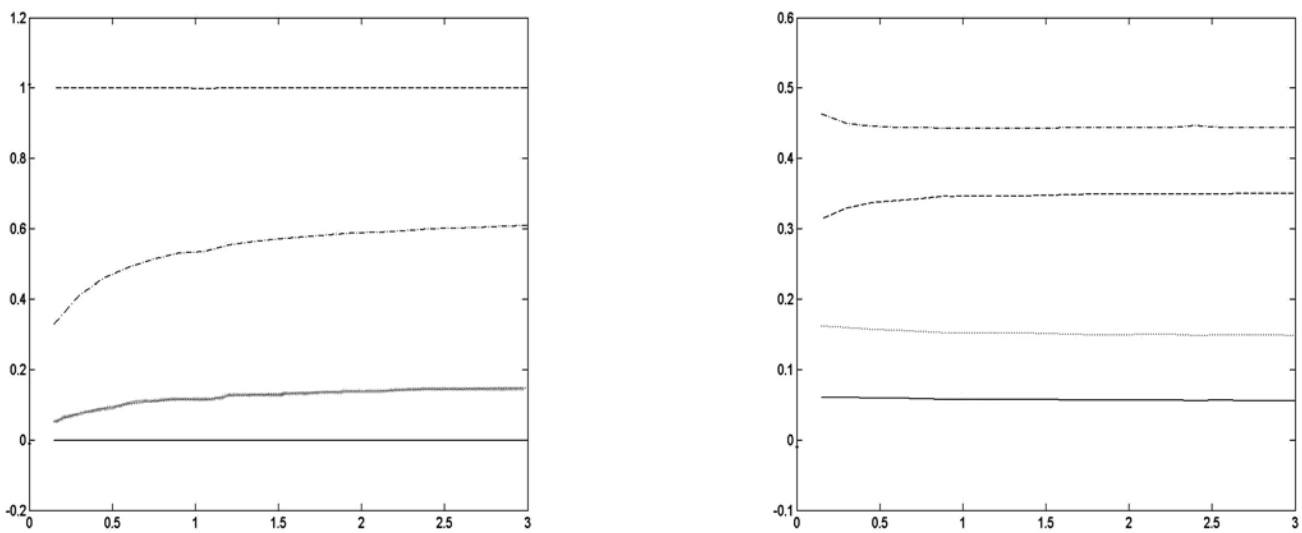

Figure 5: *The points (left) and weights (right) of the optimal design for estimating the slope of the expected response in the nonlinear regression model* (4.1) *at the point* $x = 0$. *The design interval is given by* $[0, 1]$, $m = 2$, *and vector of parameters is* $\lambda = (1, b_1, 1, 1)^T$, *where* b_1 *varies in the interval* $[0.1, 4]$.

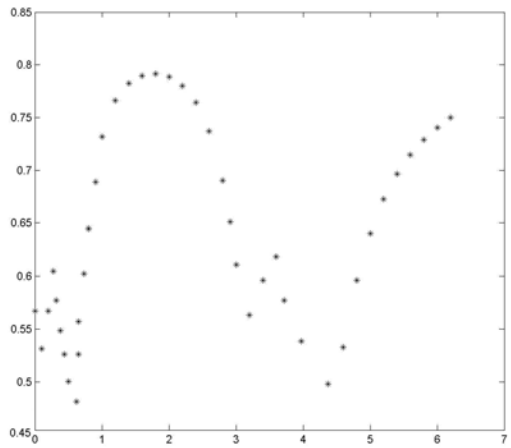

Figure 6: *The efficiency of the D-optimal design relative to the optimal design for estimating the slope of the expected response in the regression model* (4.1) *at the point* $x \in [0, 6.2]$. *The design interval is given by* $[0, 1]$, $m = 2$, *and the vector of parameters is given by* $\lambda = (1, 0.5, 1, 1)$.

b_1/b_2	0.1	0.2	0.3	0.4	0.5	1	1.5	2	5
0.1	-	0.7422	0.7491	0.7531	0.7451	0.7243	0.7155	0.7107	0.7008
0.2	0.7422	-	0.7573	0.7502	0.7450	0.7301	0.7234	0.7196	0.7116
0.3	0.7491	0.7573	-	0.7462	0.7424	0.7313	0.7261	0.7230	0.7162
0.4	0.7531	0.7502	0.7462	-	0.7399	0.7314	0.7272	0.7246	0.7188
0.5	0.7451	0.7448	0.7425	0.7402	-	0.7311	0.7275	0.7253	0.7203
1.0	0.7244	0.7301	0.7314	0.7314	0.7311	-	0.7270	0.7260	0.7231
1.5	0.7155	0.7234	0.7261	0.7272	0.7276	0.7273	-	0.7256	0.7239
2.0	0.7107	0.7196	0.7230	0.7245	0.7253	0.7259	0.7259	-	0.7239
5.0	0.7008	0.7115	0.7162	0.7188	0.7203	0.7231	0.7239	0.7239	-

Table 2: *The efficiency of the D-optimal design relative to the optimal design for estimating the slope of the expected response in the regression model* (4.1) *at the point* $x = 0$. *The design interval is given by* $[0, 1]$, $m = 2$, *and the vector of parameters is given by* $\lambda = (1, 0.5, 1, 1)$.

References

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

Becka, M., Bolt, H. M., and Urfer, W. (1993). Statistical evalutation of toxicokinetic data. *Environmetrics*, 4:311–322.

Becka, M. and Urfer, W. (1996). Statistical aspects of inhalation toxicokinetics. *Environmental and Ecological Statistics*, 3:51–64.

Biedermann, S., Dette, H., and Pepelyshev, A. (2007). Optimal discrimination designs for exponential regression models. *Journal of Statistical Planning and Inference*, 137:2579–2592.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208.

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

Dette, H. (1995). Designing of experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59:97–110.

Dette, H., Haines, L., and Imhof, L. A. (1999). Optimal designs for rational models and weighted polynomial regression. *Annals of Statistics*, 27:1272–1293.

Dette, H., Melas, V., and Pepelyshev, A. (2004a). Optimal designs for a class of nonlinear regression models. *Annals of Statistics*, 32:2142–2167.

Dette, H., Melas, V. B., and Pepelyshev, A. (2004b). Optimal designs for estimating individual coefficient in polynomial regression - a functional approach. *Journal of Statistical Planning and Inference*, 118:201–219.

Dette, H., Melas, V. B., and Pepelyshev, A. (2006). Locally E-optimal designs for exponential regression models. *Annals of the Institute of Statistical Mathematics*, 58:407–426.

Dette, H., Melas, V. B., and Pepelyshev, A. (2009). Optimal designs for estimating the slope of a regression. *Submitted, http://www.ruhr-uni-bochum.de/imperia/md/content/mathematik3/publications/detmelpep08version2.pdf.*

Dudzinski, M. L. and Mykytowycz, R. (1961). The eye lens as an indicator of age in the wild rabbit in australia. *CSIRO Wildl. Res.*, 6:156–159.

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.

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

Hader, R. and Park, S. (1978). Slope-rotatable central composite designs. *Technometrics*, 20:413–417.

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

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*, 40:633–643.

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

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

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremem problems. *Canadian Journal of Mathematics*, 12:363–366.

Landaw, E. M. and DiStefano, J. J. (1984). Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 246:R665–R677.

Mandal, N. and Heiligers, B. (1992). Minimax designs for estimating the optimum point in a quadratic response surface. *Journal of Statistical Planning and Inference*, 31:235–244.

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. (1978). Optimal designs for exponential regression. *Mathematische Operationsforschung und Statistik*, 9(1):45–59.

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

Murty, V. and Studden, W. (1972). Optimal designs for estimating the slope of a polynomial regression. *Journal of the American Statistical Association*, 67:869–873.

Myres, R. and Lahoda, S. (1975). A generalization of the response surface mean square error criterion with a specific application to the slope. *Technometrics*, 17:481–486.

Ott, L. and Mendenhall, W. (1972). Designs for estimating the slope of a second order linear model. *Technometrics*, 14:341–353.

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

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

Pronzato, L. and Walter, E. (1993). Experimental design for estimating the optimum point in a response surface. *Acta Applicandae Mathematicae*, 33:45–68.

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

Pukelsheim, F. and Torsney, B. (1991). Optimal designs for experimental designs on linearly independent support points. *Annals of Statistics*, 19:1614–1625.

Ratkowsky, D. A. (1983). *Nonlinear regression modeling: a unified practical approach*. Marcel Dekker, New York.

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

Studden, W. J. (1968). Optimal designs on Tchebycheff points. *Annals of Mathematical Statistics*, 39:1435–1447.

Studden, W. J. (2005). Elfving's theorem revisited. *Journal of Statistical Planning and Inference*, 130:85–94.