Optimal designs for a class of nonlinear regression models

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 70 94 559

Viatcheslav B. Melas
St. Petersburg State University
Department of Mathematics
St. Petersburg
Russia
email: v.melas@pobox.spbu.ru

Andrey Pepelyshev
St. Petersburg State University
Department of Mathematics
St. Petersburg
Russia
email: andrey@ap7236.spb.edu

December 9, 2002

Abstract

For a broad class of nonlinear regression models we investigate the locally E- and c-optimal design problem. It is demonstrated that in many cases the optimal designs with respect to these optimality criteria are supported at the Chebyshev points, which are the local extrema of the equi-oscillating best approximation of the function $f_0 \equiv 0$ by a normalized linear combination of the regression functions in the corresponding linearized model. The class of models includes rational, logistic and exponential models and for the rational regression models the E- and c-optimal design problem is solved explicitly in many cases. It is also demonstrated that in the models under consideration E-optimal designs are usually more efficient for estimating individual parameters than D-optimal designs.

AMS Subject Classification: 62K05, 41A50
Keywords and phrases: E-optimal design, c-optimal design, rational models, locally optimal designs, Chebyshev systems

1 Introduction

Nonlinear regression models 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)]. An appropriate choice of the experimental conditions can improve the quality of statistical inference substantially and therefore many authors have discussed the problem of designing experiments for nonlinear regression models. We refer to Chernoff (1953), Melas (1978) for early references and Ford, Torsney and Wu (1992), He, Studden and Sun (1996), Dette, Haines and Imhof (1999) for more recent references on locally optimal designs. Because locally optimal designs depend on an initial guess for the unknown parameter several authors propose alternative design strategies. Bayesian or robust optimal designs have been discussed by Pronzato and Walter (1985) and Chaloner and Larntz (1989) among many others [see Chaloner and Verdinelli (1995) and the references in this paper]. Other authors propose sequential methods, which update the information about the unknown parameter sequentially [see e.g. Ford and Silvey (1980), Wu (1985)]. Most of the literature concentrates on D-optimal designs (independent of the particular approach), which maximize the determinant of the Fisher information matrix for the parameters in the model, but much less attention has been paid to E-optimal designs in nonlinear regression models, which maximize the minimum eigenvalue of the Fisher information matrix [see Dette and Haines (1994) or Dette and Wong (1999), who gave some results for models with two parameters].

Because locally optimal designs are the basis for all advanced design strategies, it is the purpose of the present paper to study locally E-optimal designs for a class of nonlinear regression models, which can be represented in the form

$$Y = \sum_{i=1}^{s} a_i h_i(t) + \sum_{i=1}^{k} a_{s+i} \varphi(t, b_i) + \varepsilon \ . \qquad (1.1)$$

Here φ is a given 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 $a_1, \ldots, a_{s+k}, b_1, \ldots, b_k$ denote the unknown parameters of the model. The consideration of this type of model was motivated by the recent work of Imhof and Studden (2001), who considered a class of rational models of the form

$$Y = \sum_{i=1}^{s} a_i t^{i-1} + \sum_{i=1}^{k} \frac{a_{s+i}}{t - b_i} + \varepsilon \qquad (1.2)$$

where $t \in I, b_i \neq b_j (i \neq j)$ and the parameters $b_i \notin I$ are assumed to be <u>known</u> for all $i = 1, \ldots, k$. Note that model (1.2) is in fact linear, because Imhof and Studden (2001) assumed the b_i to be known. These models are very popular because they have appealing approximation properties [see Petrushev and Popov (1987) for some theoretical properties or Dudzinsky and Mykytowycz (1961), Ratkowsky (1983), p. 120 for an application of this model]. In this paper [in contrast to the work of Imhof and Studden (2001)] the nonlinear parameters in the model (1.1) are not assumed to be known, but also have to be estimated from the data. Moreover, the model (1.1) considered here includes numerous other regression functions. For example, in environmental and ecological statistics exponential models of the form

$$a_1 e^{b_1 t} + a_2 e^{b_2 t}$$

are frequently used in toxicokinetic experiments [see e.g. Becka and Urfer (1996) or Becka, Bolt and Urfer (1993)] and this corresponds to the choice $\varphi(t, x) = e^{tx}$ in (1.1). Another popular class of logarithmic models is obtained from the equation (1.1) by the choice $\varphi(t, x) = \log(t - x)$.

Imhof and Studden (2001) studied E-optimal designs for the model (1.2) with $s = 1$ under the assumption that the nonlinear parameters $b_1, \dots, b_k$ are known by the experimenter and do not have to be estimated from the data. In particular they proved that the support of the E-optimal design for estimating a subset of the parameters $a_1, \dots, a_{\ell+1}$ is given by the Chebyshev points corresponding to the regression functions in the model (1.2). These points are the extremal points of the function

$$1 + \sum_{i=1}^{k} \frac{a_i^*}{x - b_i} = p^*(x),$$

in the interval I, which has the smallest deviation from zero, that is

$$\sup_{x \in I} |p^*(x)| = \min_{a_2, \dots, a_{k+1}} \sup_{x \in I} \left| 1 + \sum_{i=1}^{k} \frac{a_i}{x - b_i} \right|. \tag{1.3}$$

The universality of this solution is due to the fact that any subsystem of the regression functions in the model (1.2), which is obtained by deleting one of the basis functions, forms a weak Chebyshev system on the interval I [see Karlin and Studden (1966) or the discussion in Section 2]. However, in the case where the parameters $b_1, \dots, b_k$ are unknown and also have to be estimated from the data, the locally optimal design problem for the model (1.2) is equivalent to an optimal design problem in the linear regression model

$$Y = \sum_{i=1}^{s} \beta_i t^{i-1} + \sum_{i=1}^{2k} \left(\frac{\beta_{s+2i-1}}{t - b_i} + \frac{\beta_{s+2i}}{(t - b_i)^2} \right) + \varepsilon, \tag{1.4}$$

for which the corresponding regression function do not satisfy the weak Chebyshev property mentioned above. Nevertheless, we will prove in this paper that in cases with $k \geq 2$, where the quantity

$$\max_{i \neq j} |b_i - b_j|$$

is sufficiently small, locally E-optimal designs and many locally c-optimal designs for estimating linear combinations of the parameters are still supported on Chebyshev points. This substantially simplifies the construction of locally E-optimal designs. Moreover, we show that this result does not depend on the specific form of the model (1.2) and (1.4) but can be established for the general model (1.1) (or its equivalent linearized model). Additionally it can be shown numerically that in many cases the E-optimal design is in fact supported on the Chebyshev points for all admissible values of the parameters $b_1, \dots, b_k$ ($b_i \neq b_j; i \neq j$). Our approach is based on a study of the limiting behaviour of the information matrix in the model (1.1) in the case, where all nonlinear parameters in the model (1.1) tend to the same limit. We show that in this case the locally E-optimal and many locally optimal designs for estimating linear combinations of the coefficients $a_{s+1}, b_{s+1}, \dots, a_{s+k}, b_{s+k}$ in the model (1.1) have the same limiting design. This indicates that E-optimal designs in models of the type (1.1) yield more precise estimates of the individual coefficients than the popular D-optimal designs and we will illustrate this fact in several examples.

The remaining part of the paper is organized as follows. In Section 2 we introduce the basic concepts, notation and present some preliminary results. Section 3 is devoted to an asymptotic analysis of the model (1.1), which is based on a linear transformation introduced in the Appendix [see Section 5]. Finally, some applications to the rational model (1.2) and its equivalent linear regression model (1.4) are presented in Section 4, which extend the results of Imhof and Studden (2001) to the case, where the nonlinear parameters in the model (1.2) are not known and have to be estimated from the data.

2 Preliminary results

Consider the nonlinear regression model (1.1) and define

$$
\begin{aligned}
f(t,b) &= (f_1(t,b),\dots,f_m(t,b))^T \\
&= (h_1(t),\dots,h_s(t),\varphi(t,b_1),\varphi'(t,b_1),\dots,\varphi(t,b_k),\varphi'(t,b_k))^T
\end{aligned} \tag{2.1}
$$

as a vector of $m = s + 2k$ regression functions, where the derivatives of the function φ are taken with respect to the second argument. It is straightforward to show that the Fisher information for the parameter $(a_1,\dots,a_s,a_{s+1},b_{s+1},\dots,a_{s+k},b_{s+k})^T = (\beta_1,\dots,\beta_m)^T = \beta$ in the equivalent linear regression model

$$
Y = \beta^T f(t,b) + \varepsilon = \sum_{i=1}^{s} \beta_i h_i(t) + \sum_{i=1}^{k} (\beta_{s+2i-1}\varphi(t,b_i) + \beta_{s+2i}\varphi'(t,b_i)) + \varepsilon \tag{2.2}
$$

is given by

$$
I(b,t) = f(t,b) f^T(t,b) \tag{2.3}
$$

The dependence on the parameter b is omitted, whenever it is clear from the context. Following Kiefer (1974) we call any probability measure ξ with finite support on the interval I an (approximate) design. The support points give the locations where observations have to be taken, while the masses correspond to the relative proportions of total observations to be taken at the particular points. For a design ξ the information matrix in the model (2.2) is defined by

$$
M(\xi,b) = \int_I I(b,t)\, d\xi(t), \tag{2.4}
$$

and a locally optimal design maximizes an appropriate function of the information matrix [see Silvey (1980) or Pukelsheim (1993)]. Among the numerous optimality criteria proposed in the literature we consider the D-, E- and c-optimality criteria in this paper. A D-optimal design ξ_D^* for the regression model (2.2) maximizes the determinant

$$
|M(\xi,b)| \tag{2.5}
$$

over the set of all approximate designs on the interval I. Similary, an E-optimal design ξ_E^* maximizes the minimum eigenvalue

$$
\lambda_{\min}(M(\xi,b)), \tag{2.6}
$$

while for a given vector $c \in \mathbb{R}^m$ a c-optimal design minimizes the expression

$$
c^T M^{-}(\xi,b) c, \tag{2.7}
$$

where the minimum is taken over the set of all designs for which the linear combination $c^T\beta$ is estimable, i.e. $c \in \operatorname{range}(M(\xi,b))\ \forall\ b$.

Note that a locally optimal design problem in a nonlinear model (1.1) corresponds to an optimal design problem in the model (2.2) for the transformed vector of parameters $K_a^T b$, where the matrix $K_a \in \mathbb{R}^{m\times m}$ is given by

$$
K_a = \operatorname{diag}\Big(\underbrace{1,\dots,1}_{s},\underbrace{1,\frac{1}{a_1},1,\dots,1,\frac{1}{a_k}}_{2k}\Big). \tag{2.8}
$$

For example, a locally D-optimal design in the model (1.1) maximizes the determinant

$$|K_a^{-1}M(\xi,b)K_a^{-1}| = |K_a^{-1}|^2|M(\xi,b)|,$$

does not depend on the parameters $a_1,\ldots,a_k$ and coincides with the D-optimal design in the model (2.2). Similary, the c-optimal design for the model (1.1) can be obtained from the $\bar{c}$-optimal design in the model (2.2), where the vector $\bar{c}$ is given by $\bar{c} = K_a c$. Finally, the locally E-optimal design in the nonlinear regression model (1.1) maximizes $\lambda_{\min}(K_a^{-1}M(\xi,b)K_a^{-1})$, where $M(\xi,b)$ is the information matrix in the equivalent linear regression model (2.2). For the sake of transparency we will mainly concentrate on the linearized version (2.2). The corresponding results in the nonlinear regression model (1.1) will be briefly mentioned, whenever it is necessary.

A set of functions $f_1,\ldots,f_m : I \to \mathbb{R}$ is called a weak Chebyshev system (on the interval I) if there exists an $\varepsilon \in \{-1,1\}$ such that

$$\varepsilon \cdot \begin{vmatrix} f_1(x_1) & \ldots & f_1(x_m) \\ \vdots & \ddots & \vdots \\ f_m(x_1) & \ldots & f_m(x_m) \end{vmatrix} \geq 0 \tag{2.9}$$

for all $x_1,\ldots,x_m \in I$ with $x_1 < x_2 < \ldots < x_m$. If the inequality in (2.9) is strict, then $\{f_1,\ldots,f_m\}$ is called Chebyshev system. It is well known [see Karlin and Studden (1966), Theorem II 10.2] that if $\{f_1,\ldots,f_m\}$ is a weak Chebyshev system, then there exists a unique function

$$\sum_{i=1}^{m} c_i^* f_i(t) = c^{*T} f(t), \tag{2.10}$$

with the following properties

$$\begin{array}{lll} & \text{(i)} & |c^{*T}f(t)| \leq 1 \quad \forall\, t \in I \\ (2.11) & & \\ & \text{(ii)} & \text{there exist } m \text{ points } s_1 < \ldots < s_m \text{ such that } c^{*T}f(s_i) = (-1)^i \quad i = 1,\ldots,m. \end{array}$$

The function $c^{*T}f(t)$ is called Chebyshev polynomial, the points $s_1,\ldots,s_m$ are called Chebyshev points and need not to be unique. They are unique if $1 \in \text{span}\{f_1,\ldots,f_m\}, m \geq 1$ and I is a bounded and closed interval, where in this case

$$s_1 = \min_{x \in I} x, \quad s_m = \max_{x \in I} x.$$

It is well known [see Studden (1968), Pukelsheim and Studden (1993), Heiligers (1994) or Imhof and Studden (2001) among others] that for many linear regression models the E- and c-optimal designs are supported at the Chebyshev points.
For a further discussion assume that the functions $f_1,\ldots,f_m$ generate a Chebyshev system on the interval I with Chebyshev polynomial $c^{*T}f(t)$ and Chebyshev points $s_1,\ldots,s_m$, define the $m \times m$ matrix $F = (f_i(s_j))_{i,j=1}^m$ and consider a vector of weights given by

$$w = (w_1,\ldots,w_m)^T = \frac{JF^{-1}c^*}{\|c^*\|^2}, \tag{2.12}$$

where the matrix J is defined by $J = \operatorname{diag}\{(-1), 1, \ldots, (-1)^m\}$. It is then easy to see that

$$\frac{c^*}{\|c^*\|^2} = FJw = \sum_{j=1}^m f(s_j)(-1)^j w_j \in \partial\mathcal{R}, \tag{2.13}$$

where

$$\mathcal{R} = \operatorname{conv}(f(I) \cup f(-I))$$

denotes the Elfving set [see Elfving (1952)]. Consequently, if all weights in (2.12) are nonnegative, it follows from Elfving's theorem that the design

$$\xi_{c^*}^* = \begin{pmatrix} s_1 & \ldots & s_m \\ w_1 & \ldots & w_m \end{pmatrix} \tag{2.14}$$

is c^*-optimal in the regression model (2.2) [see Elfving (1952)], where $c^* \in \mathbb{R}^m$ denotes the vector of coefficients of the Chebyshev polynomial defined in the previous paragraph. The following result relates this design to the E-optimal design.

Lemma 2.1. *Assume that $f_1, \ldots, f_m$ generate a Chebyshev system on the interval I such that the Chebyshev points are unique. If the minimum eigenvalue of the information matrix of an E-optimal design has multiplicity one, then the design $\xi_{c^*}^*$ defined by (2.12) and (2.14) is E-optimal in the regression model (2.2). Moreover, in this case the E-optimal design is unique.*

Proof. Let ξ_E^* denote an E-optimal design such that the minimum eigenvalue $\lambda = \lambda_{\min}(M(\xi_E^*, b))$ of the information matrix $M(\xi_E^*, b)$ has multiplicity one with corresponding eigenvector $z \in \mathbb{R}^m$. By the equivalence theorem for the E-optimality criterion [see Pukelsheim (1993), p. 181-182] we obtain for the matrix $E = zz^T/\lambda$

$$\left(\frac{1}{\sqrt{\lambda}} z^T f(t)\right)^2 = f^T(t) E f(t) \leq 1$$

for all $t \in I$ with equality at the support points of ξ_E^*. Because the Chebyshev polynomial is unique it follows that (up to the factor ∓ 1)

$$c^* = \frac{1}{\sqrt{\lambda}} z$$

and that $\operatorname{supp}(\xi_E^*) = \{s_1, \ldots, s_m\}$. Now Theorem 3.2 in Dette and Studden (1993) implies that ξ_E^* is also c^*-optimal, where $c^* \in \mathbb{R}^m$ denotes the vector of coefficients of the Chebyshev polynomial. Consequently, by the discussion of the previous paragraph we have $\xi_E^* = \xi_{c^*}^*$, which proves the assertion. $\square$

Lemma 2.2. *Assume that the functions $f_1, \ldots, f_m$ generate a Chebyshev system on the interval I with Chebyshev polynomial $c^{*T} f(t)$ and let $\xi_{c^*}^*$ denote the c^*-optimal design in the regression model (2.2) defined by (2.14). Then c^* is an eigenvector of the information matrix $M(\xi_{c^*}^*, b)$ and if the corresponding eigenvalue $\lambda = \frac{1}{\|c^*\|^2}$ is the minimal eigenvalue, then $\xi_{c^*}^*$ is also E-optimal in the regression model (2.2).*

Proof. From the identity (2.13) and the Chebyshev property (2.11) it follows immediately that c^* is an eigenvector of the matrix

$$M(\xi^*_{c^*}, b) = \sum_{i=1}^{m} f(s_i) f^T(s_i) w_i$$

with corresponding eigenvalue $\lambda = 1/\|c^*\|^2$. Now if $\lambda = \lambda_{\min}(M(\xi^*_{c^*}, b))$ we define the matrix $E = \lambda c^* c^{*T}$ and obtain from the Chebyshev properties (2.11) that

$$f^T(t) E f(t) = \lambda (c^{*T} f(t))^2 \leq \lambda = \lambda_{\min}(M(\xi^*_{c^*}, b))$$

for all $t \in I$. The assertion of the Lemma now follows from the equivalence theorem for E-optimality [see Pukelsheim (1993)].

$\square$

We now discuss the c-optimal design problem in the regression model (2.2) for a general vector $c \in \mathbb{R}^m$ (not necessarily equal to the vector c^* of coefficients of the Chebyshev polynomial). Assume again that $f_1, \ldots, f_m$ generate a Chebyshev system on the interval I. As candidate for the c-optimal design we consider the measure

$$\xi_c = \xi_c(b) = \begin{pmatrix} s_1 & \ldots & s_m \\ w_1 & \ldots & w_m \end{pmatrix}, \tag{2.15}$$

where the support points are the Chebyshev points and the weights are already chosen such that the expression $c^T M^{-1}(\xi_c, b) c$ becomes minimal, that is

$$w_i = \frac{|e_i^T J F^{-1} c|}{\sum_{j=1}^{m} |e_j^T J F^{-1} c|} \qquad i = 1, \ldots, m \tag{2.16}$$

[see Pukelsheim (1993)]. The following result characterizes the optimal designs for estimating the individual coefficients.

Lemma 2.3. *Assume that the functions $f_1, \ldots, f_m$ generate a Chebyshev system on the interval I and let $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^T \in \mathbb{R}^m$ denote the jth unit vector. The design ξ_{e_j} defined by (2.15) and (2.16) for the vector $c = e_j$ is e_j-optimal if the system*

$$\{f_i \mid i \in \{1, \ldots, m\} \backslash \{j\}\}$$

is a weak Chebyshev system on the interval I.

Proof. If $f_1, \ldots, f_m$ generate a weak Chebyshev system on the interval I it follows from Theorem 2.1 in Studden (1968) that the design ξ_{e_j} defined in (2.15) and (2.16) is e_j-optimal if

$$\varepsilon e_i^T J F^{-1} e_j \geq 0 \qquad i = 1, \ldots, m$$

for some $\varepsilon \in \{-1, 1\}$. The assertion of the Lemma now follows by Cramer's rule. $\square$

Remark 2.4. It is worthwhile to mention that in general the sufficient condition of Lemma 2.3 is not satisfied. To see this assume that $k \geq 3$, that the function φ is continuously differentiable with respect to the second argument and that the functions $f_1(\cdot,b),\ldots,f_m(\cdot,b)$ defined by (2.1) generate a Chebyshev system for any b. Define an $(m-1)\times(m-1)$ matrix

$$F_j(x) := \Big(h_1(t_i),\ldots,h_s(t_i),\varphi(t_i,b_1),\varphi'(t_i,b_1),\ldots,\varphi(t_i,b_{j-1}),\varphi'(t_i,b_{j-1}),\\ \varphi(t_i,x),\varphi(t_i,b_{j+1}),\ldots,\varphi(t_i,b_k),\varphi'(t_i,b_k)\Big)_{i=1}^{m-1}$$

where $c < t_1 < \ldots < t_{m-1} < d, b_i \neq b_j$ whenever $i \neq j$ and $x \neq b_i$. We choose $t_1,\ldots,t_{m-1}$ such that

$$g(x) = \det F_j(x) \not\equiv 0$$

(note that the functions $f_1,\ldots,f_m$ form a Chebyshev system and therefore this is always possible) and observe that

$$g(b_i) = 0 \qquad i = 1,\ldots,k; i \neq j.$$

Because $k \geq 3$ and g is continuously differentiable it follows that there exist two points, say x^* and x^{**} such that such that $g'(x^*) < 0$ and $g'(x^{**}) > 0$. Consequently, there exists an $\bar{x}$ such that

$$0 = g'(\bar{x}) = \det\Big(f_\nu(t_i,b_{\bar{x}})\Big)_{i=1,\ldots,m-1}^{\nu=1,\ldots,m,\nu\neq s+2j-1},$$

where the vector $b_{\bar{x}}$ is defined by $b_{\bar{x}} = (b_1,\ldots,b_{j-1},\bar{x},b_{j+1},\ldots,b_k)^T$. Note that the Chebyshev property of the functions $f_1,\ldots,f_{s+2j-2},f_{s+2j},\ldots,f_m$ would imply that all determinants in (2.9) were of the same sign (otherwise there exists a b such that the determinant vanishes for $t_1 < \ldots < t_{m-1}$). Therefore the conditions $g'(x^*) < 0, g'(x^{**}) > 0$ imply that there exists a $\tilde{x} \in (x^*,\bar{x})$ or $\tilde{x} \in (\bar{x},x^{**})$, such that the system of regression functions

$$\Big\{f_1(t,b_{\tilde{x}}),\ldots,f_{s+2j-2}(t,b_{\tilde{x}}),f_{s+2j}(t,b_{\tilde{x}}),\ldots,f_m(t,b_{\tilde{x}})\Big\}\\ = \Big\{h_1(t),\ldots,h_s(t),\varphi(t,b_1),\varphi'(t,b_1),\ldots,\varphi'(t,b_{j-1}),\varphi'(t,\tilde{x}),\varphi(t,b_{j+1}),\varphi'(t,b_{j+1}),\ldots,\varphi'(t,b_k)\Big\}$$

is not a weak Chebyshev system on the interval I. Finally in the case $k = 2$, if

$$\lim_{|b|\to\infty} \varphi(t,b) \to 0$$

it can be shown by a similar argument that there exists a $\tilde{x}$ such that the system

$$\{h_1,(t),\ldots,h_s(t),\varphi(t,b_1)\varphi'(t,b_1)\varphi'(t,\tilde{x})\}$$

is not a Chebyshev system on the interval I.

3 Asymptotic analysis of E- and c-optimal designs

Recall the definition of the information matrix in (2.4) for the model (2.2) with design space given by $I = [c_1,d_1]$ and assume that the nonlinear parameters vary in a compact interval, say

$$b_i \in [c_2,d_2]; \quad i = 1,\ldots,k.$$

We are interested in the asymptotic properties of E- and c-optimal designs if

$$b_i = x + \delta r_i \qquad i = 1, \ldots, k \tag{3.1}$$

for some $x \in [c_2, d_2], \delta > 0,\ r_1 < r_2 < \ldots < r_k$ and $\delta \to 0$. For this purpose we study for fixed $\varepsilon, \Delta > 0$ the set

$$\Omega_{\varepsilon,\Delta} = \Big\{ b \in \mathbb{R}^k \ \Big|\ b_i - b_j = \delta(r_i - r_j);\ i,j = 1, \ldots, k;\ \delta \le \varepsilon;\ b_i \in [c_2, d_2],\ \min_{i \ne j} |r_i - r_j| \ge \Delta \Big\}, \tag{3.2}$$

introduce the functions

$$\begin{aligned} \bar{f}_i(t,x) &= \bar{f}_i(t) = h_i(t) \qquad i = 1, \ldots, s \\ \bar{f}_{s+i}(t,x) = \bar{f}_{s+i}(t) &= \varphi^{(i-1)}(t,x) \qquad i = 1, \ldots, 2k \end{aligned} \tag{3.3}$$

and the corresponding vector of regression functions

$$\bar{f}(t,x) = (\bar{f}_1(t,x), \ldots, \bar{f}_{s+2k}(t,x))^T, \tag{3.4}$$

where the derivatives are taken with respect to the second argument, that is

$$\varphi^{(i)}(t,x) = \frac{\partial^i}{\partial^i u} \varphi(t,u)\Big|_{u=x} \qquad i = 0, \ldots, 2k-1.$$

Again the dependency of the functions $\bar{f}_i$ on the parameter x will be omitted whenever it is clear from the context. The linear model with vector of regression functions given by (3.4) will serve as an approximation for the model (2.2) if the parameters b_i are sufficiently close to each other.

Lemma 3.1. *Assume that the function*

$$\varphi : [c_1, d_1] \times [c_2, d_2] \to \mathbb{R}$$

in model (1.1) satisfies

$$\varphi \in C^{0,2k-1}([c_1, d_1] \times [c_2, d_2])$$

and that for any fixed $x \in [c_2, d_2]$ the functions $\bar{f}_1, \ldots, \bar{f}_{s+2k}$ defined by (3.3) form a Chebyshev system on the interval $[c_1, d_1]$. For any $\Delta > 0$ and any design on the interval $[c_1, d_1]$ with at least $m = s + 2k$ support points there exists an $\varepsilon > 0$ such that for all $b \in \Omega_{\varepsilon,\Delta}$ the maximum eigenvalue of the inverse information matrix $M^{-1}(\xi, b)$ defined in (2.4) is simple.

Proof. Recall the definition of the functions in (3.3) and let

$$\bar{M}(\xi, x) = \int_c^d \bar{f}(t,x) \bar{f}^T(t,x) d\xi(x) \tag{3.5}$$

denote the information matrix in the corresponding linear regression model. Because of the Chebyshev property of the functions $\bar{f}_1, \ldots, \bar{f}_{s+2k}$ it follows that $|\bar{M}(\xi, x)| \ne 0$ (note that the design ξ has at least $s + 2k$ support points). It will be shown in the Appendix (see Theorem 5.1) that under the condition (3.1) with $\delta \to 0$ the asymptotic expansion

$$\delta^{4k-2}M^{-1}(\xi,b) = h\bar{\gamma}\bar{\gamma}^T + o(1) \tag{3.6}$$

is valid, where the vector $\bar{\gamma} = (\bar{\gamma}_1,\ldots,\bar{\gamma}_{s+2k})^T$ is defined by

$$\begin{aligned} \bar{\gamma}_{s+2i-1} &= -\prod_{j\neq i}(r_i - r_j)^{-2}\cdot\sum_{j\neq i}\frac{2}{r_i - r_j}, \quad i = 1,\ldots,k, \\ &\bar{\gamma}_1 = \ldots = \bar{\gamma}_s = 0; \quad \bar{\gamma}_{s+2i} = 0 \quad i = 1,\ldots,k, \end{aligned} \tag{3.7}$$

and the constant h is given by

$$h = ((2k-1)!)^2(\bar{M}^{-1}(\xi,x))_{m,m}. \tag{3.8}$$

From formula (3.6) it follows that the maximal eigenvalue of the matrix $M^{-1}(\xi,b)$ is simple if δ is sufficiently small.
For a fixed value $r = (r_1,\ldots,r_k)$ and fixed $x \in \mathbb{R}$ in the representation (3.1) denote by $\varepsilon = \varepsilon(x,r)$ the maximal value (possibly ∞) such that the matrix $M^{-1}(\xi,b)$ has a simple maximal eigenvalue for all $\delta \le \varepsilon$. Then the function $\varepsilon : (x,r) \to \varepsilon(x,r)$ is continuous and the infimum

$$\inf\Big\{\varepsilon(x,b)\Big| x \in [c_1,d_1],\ \min_{i\neq j}|r_i - r_j| \ge \Delta,\ \|r\|_2 = 1\Big\}$$

is attained for some $x^* \in [c_1,d_1]$ and r^*, which implies

$$\varepsilon^* = \varepsilon(x^*,r^*) > 0.$$

This means that for any $b \in \Omega_{\varepsilon^*,\Delta}$ the multiplicity of the maximal eigenvalue of the information matrix $M^{-1}(\xi,b)$ is equal one.

□

Theorem 3.2. *Assume that the function $\varphi : [c_1,d_1]\times[c_2,d_2] \to \mathbb{R}$ satisfies*

$$\varphi \in C^{0,2k-1}([c_1,d_1]\times[c_2,d_2])$$

and that the systems of functions

$$\{f_1(t,b),\ldots,f_m(t,b)\}$$
$$\{\bar{f}_1(t,x),\ldots,\bar{f}_m(t,x)\}$$

defined by (2.1) and (3.3), respectively, are Chebyshev systems on the interval $[c_1,d_1]$ (for arbitrary but fixed $x, b_1,\ldots,b_k \in [c_2,d_2]$ with $b_i \neq b_j$ whenever $i \neq j$). If ε is sufficiently small, then for any $b \in \Omega_{\varepsilon,\Delta}$ the design $\xi^*_{c^*}$ defined by (2.12) and (2.14) is the unique E-optimal design in the regression model (2.2).

Proof. The proof is a direct consequence of Lemma 2.2 and Lemma 3.1, which shows that the multiplicity of the maximum eigenvalue of the inverse information matrix of any design has multiplicity one, if $b \in \Omega_{\varepsilon,\Delta}$ and ε is sufficiently small.

□

From Remark 2.4 we may expect that in general c-optimal designs in the regression model (1.1) are not necessarily supported at the Chebyshev points. Nevertheless, an analogue of Lemma 3.1 is available for specific vectors $c \in \mathbb{R}^m$. The proof is similar as the proof of Lemma 3.1 and therefore omitted (see also the proof of Theorem 3.5 below which uses similar arguments).

Lemma 3.3. *Let $e_i = (0,\dots,0,1,0,\dots,0)^T$ denote the ith unit vector in $\mathbb{R}^m$. Under the assumptions of Lemma 3.1 define a vector $\tilde\gamma = (0,\dots,0,\gamma_1,\dots,\gamma_{2k}) \in \mathbb{R}^m$ by*

$$(3.9)\qquad \begin{aligned} \gamma_{2i} &= \prod_{j\neq i}(r_i - r_j)^{-2} \quad i = 1,\dots,k \\ \gamma_{2i-1} &= -\gamma_{2i}\sum_{j\neq i}\frac{2}{r_i - r_j} \quad i = 1,\dots,k. \end{aligned}$$

(i) If $c \in \mathbb{R}^m$ satisfies $c^T\tilde\gamma \neq 0$, then for nay $\Delta > 0$, sufficiently small ε and any $b \in \Omega_{\varepsilon,\Delta}$ the design $\xi_c(b)$ defined in (2.15) and (2.16) is c-optimal in the regression model (2.2).

(ii) The assumption $c^T\tilde\gamma \neq 0$ is in particular satisfied for the vector $c = e_{s+2j-1}$ for any $j = 1,\dots k$ and for the vector $c = e_{s+2j}$ for any $j = 1,\dots,k$, which satisfies condition

$$(3.10)\qquad \sum_{\ell\neq j}\frac{1}{r_j - r_\ell} \neq 0.$$

Remark 3.4. Note that it follows from the proof of Theorem 3.1 that the assumption of compactness of the intervals is only required for the existence of the set $\Omega_{\varepsilon,\Delta}$. In other words if condition (3.1) is satisfied and δ is sufficiently small, the maximum eigenvalue of the matrix $M^{-1}(\xi, b)$ will have multiplicity one (independently of the domain of the function φ). The same remark applies to the statement of Theorem 3.2 and Lemma 3.3.

Our final result of this section shows that under assumption (3.1) with small δ the locally E- and locally c-optimal designs for the vectors c considered in Lemma 3.3 are very close. To be precise we assume that the assumptions of Theorem 3.2 are valid and consider the design

$$(3.11)\qquad \bar\xi_c = \bar\xi_c(x) = \begin{pmatrix} \bar s_1 & \dots & \bar s_m \\ \bar w_1 & \dots & \bar w_m \end{pmatrix}$$

where $\bar s_1,\dots,\bar s_m$ are the Chebyshev points corresponding to the system $\{\bar f_i \mid i = 1,\dots,m\}$ defined in (3.3),

$$(3.12)\qquad \bar w_i = \frac{|e_i^T J\bar F^{-1}c|}{\sum_{j=1}^m |e_j^T J\bar F^{-1}c|} \quad i = 1,\dots,m$$

with $\bar F = (f_i(\bar s_j))_{i,j=1}^m$ and $c \in \mathbb{R}^m$ is a fixed vector.

Theorem 3.5. *Assume that the assumptions of Theorem 3.2 are satisfied and that for the system $\{\bar f_1,\dots,\bar f_m\}$ the Chebyshev points are unique.*

(i) If $\delta \to 0$, the design $\xi^*_{c^*}(b)$ defined by (2.14) and (2.12) converges weakly to the design $\bar\xi_{e_m}(x)$ defined by (3.11) and (3.12) for $c = e_m$.

(ii) If $c \in \mathbb{R}^m$ satisfies $c^T\tilde{\gamma} \neq 0$ for the vector $\tilde{\gamma}$ defined in (3.9) and $\delta \to 0$, then the design $\xi_c^*(b)$ defined by (2.15) and (2.16) converges weakly to the design $\bar{\xi}_{e_m}(x)$.

(iii) The assumption $c^T\tilde{\gamma} \neq 0$ is in particular satisfied for the vector $c = e_{s+2j-1}$ for any $j = 1, \dots k$ and for the vector $c = e_{s+2j}$ for any $j = 1, \dots, k$, which satisfies the condition (3.10).

Proof. It follows from Theorem 3.2 that the design $\xi_{c^*}^* = \xi_{c^*}^*(b)$ is locally E-optimal for sufficiently small $\delta > 0$. In other words, if δ is sufficiently small the design $\xi_{c^*}^*$ minimizes

$$\max_{\|c\|_2=1} c^T M^{-1}(\xi, b)c$$

in the class of all designs. Note that the components of the vector $r = (r_1, \dots, r_k)$ are ordered, which implies

$$e_{s+2i-1}^T\tilde{\gamma} \neq 0 \quad i = 1, k.$$

Multiplying equation (5.4) in the Appendix with δ^{4k-2} it then follows from Theorem 5.1 in the Appendix that for some subsequence $\delta_k \to 0$

$$\xi_{c^*}^* \to \hat{\xi}(x),$$

where the design $\hat{\xi}(x)$ minimizes the function

$$\max_{\|c\|_2=1} (c^T\tilde{\gamma})^2 e_m^T \bar{M}^{-1}(\xi, x)e_m$$

and the vector $\tilde{\gamma}$ is defined by equation (3.7). The maximum is attained for $c = \tilde{\gamma}/\|\tilde{\gamma}\|_2$ (independently of the design ξ) and consequently $\hat{\xi}(x)$ is e_m-optimal in the linear regression model defined by the regression function in (3.4). Now the functions $\bar{f}_1, \dots, \bar{f}_m$ generate a Chebyshev system and the corresponding Chebyshev points are unique, which implies that the e_m-optimal design $\bar{\xi}_{e_m}(x)$ is unique. Consequently, every subsequence of designs $\xi_{c^*}^*(b)$ contains a weakly convergent subsequence with limit $\bar{\xi}_{e_m}(x)$ and this proves the first part of the assertion. For a proof of the second part we note that a c-optimal design minimizes

$$c^T M^{-1}(\xi, b)c$$

in the class of all designs on the interval I. Now if $c^T\tilde{\gamma} \neq 0$ and

$$e_{s+2i-1}^T\tilde{\gamma} = -\prod_{j\neq i}(r_i - r_j)^{-2} \sum_{j\neq i} \frac{2}{r_i - r_j} \neq 0$$

for some $i = 1, \dots, k$, the same argument as in the previous paragraph shows that $\xi_c^*(b)$ converges weakly to the design which maximizes the function

$$(\tilde{\gamma}^T c)^2 e_m^T \bar{M}^{-1}(\xi, x)e_m.$$

If $e_{s+2i-1}^T\tilde{\gamma} = 0$ for all $i = 1, \dots, k$, the condition $c^T\tilde{\gamma} \neq 0$ implies $e_{s+2i}^T\tilde{\gamma} \neq 0$ for some $i = 1, \dots, k$ and the assertion follows by multiplying equation (5.4) in the Appendix with δ^{4k-4} and similar arguments. Finally, the third assertion follows directly from the definition of the vector $\tilde{\gamma}$ in (3.9). $\square$

Remark 3.6. Note that Theorem 3.2, Lemma 3.3 and Theorem 3.5 remain valid for the locally optimal designs in the nonlinear regression model (1.1). This follows by a careful inspection of the proofs of the previous results. For example, Theorem 5.1 in the Appendix shows that

$$\delta^{4k-2} K_a M^{-1}(\xi, b) K_a = h(K_a \tilde{\gamma})(K_a \tilde{\gamma})^T + o(1)$$

where the vector $\tilde{\gamma}$ is defined in Lemma 3.3 and consequently, there exists a set $\Omega_{\varepsilon,\Delta}$ such that for all $b \in \Omega_{\varepsilon,\Delta}$ the maximum eigenvalue of the inverse information matrix in the model (1.1) is simple. Similary, if $\delta \to 0$ and (3.1) is satisfied, c-optimal designs in the nonlinear regression model are given by the design $\xi_{\bar{c}}(b)$ in (2.15) and (2.16) with $\bar{c} = K_a c$ whenever $\tilde{\gamma}^T \bar{c} \neq 0$ and all these designs converge weakly to the e_m-optimal design in the linear regression model defined by the functions (3.4).

We finally remark that Theorem 3.5 and Remark 3.6 indicate that E-optimal designs are very efficient for estimating the parameters $a_{s+1}, b_1, \ldots, a_{s+k}, b_k$ in the nonlinear regression model (1.1) and the linear model (2.2), because for small differences $|b_i - b_j|$ the E-optimal design and the optimal design for estimating the individual coefficients are close to the optimal design for estimating the coefficient b_k. Therefore we expect E-optimal designs to be more efficient for estimating these parameters than D-optimal designs. We will illustrate this fact in the following section, which discusses the rational model in more detail.

4 Rational models

In this section we discuss the rational model (1.2) in more detail, where the design space is a compact or seminfinite interval I. In contrast to the work of Imhof and Studden (2001) we assume that the nonlinear parameters $b_1, \ldots, b_k \notin I$ are not known by the experimenter but have to be estimated from the data. A typical application of this model can be found in the work of Dudzinski and Mykytowycz (1961), where this model was used to describe the relation between the weight of the dried eye lens of the European rabbit and the age of the animal. In the notation of Section 2 and 3 we have $f(t) = f(t, b) = (f_1(t), \ldots, f_m(t))^T$ with

$$\begin{aligned} f_i(t) &= f_i(t, b) = t^{i-1} \quad i = 1, \ldots, s \\ f_{s+2i-1}(t) &= f_{s+2i-1}(t, b) = \frac{1}{t - b_i} \quad i = 1, \ldots, k \\ f_{s+2i}(t) &= f_{s+2i}(t, b) = \frac{1}{(t - b_i)^2} \quad i = 1, \ldots, k \end{aligned} \tag{4.1}$$

and the equivalent linear regression model is given by (1.4). The corresponding limiting model is determined by the regression functions $\bar{f}(t) = \bar{f}(t, x) = (\bar{f}_1(t, x), \ldots, \bar{f}_m(t, x))^T$ with

$$\begin{aligned} \bar{f}_i(t) &= t^{i-1}, \quad i = 1, \ldots, s \\ \bar{f}_{i+s}(t) &= \bar{f}_{s+i}(t, x) = \frac{1}{(t - x)^i}, \quad i = 1, \ldots, 2k. \end{aligned} \tag{4.2}$$

Some properties of the functions defined by (4.1) and (4.2) are discussed in the following lemma.

Lemma 4.1. *Define*

$$\mathcal{B} = \{b = (b_1, \dots, b_k)^T \in \mathbb{R}^k \mid b_i \notin I; b_i \neq b_j\} \ ,$$

then the following assertions are true.

(i) If I is a finite interval or $I \subset [0, \infty)$ and $b \in \mathcal{B}$, then the system

$$\{f_1(t_1, b), \dots, f_m(t, b)\}$$

defined in (4.1) is a Chebyshev system on the interval I. If $x \notin I$ then the system

$$\{\bar{f}_1(t, x), \dots, \bar{f}_m(t, x)\}$$

defined by (4.2) is a Chebyshev system on the interval I.

(ii) Assume that $b \in \mathcal{B}$ and that one of the following conditions is satisfied

(a) $I \subset [0, \infty)$

(b) $s = 1$ or $s = 0$.

For any $j \in \{1, \dots, k\}$ the system of regression functions

$$\{f_i(t, b) \mid i = 1, \dots, m, i \neq s + 2j\}$$

is a Chebyshev system on the interval I.

(iii) If I is a finite interval or $I \subset [0, \infty)$, $k \geq 2$ and $j \in \{1, \dots, k\}$, then there exists a nonempty set $W_j \subset \mathcal{B}$ such that for all $b \in W_j$ the system of functions

$$\{f_i(t, b) \mid i = 1, \dots, m; i \neq s + 2j - 1\}$$

is not a Chebyshev system on the interval I.

Proof. Part (iii) follows from Remark 2.4. Part (i) and (ii) are proved similary and we restrict ourselves to the first case. For this purpose we introduce the functions $\psi(t, b) = (\psi_1(t, \tilde{b}), \dots, \psi_m(t, \tilde{b}))^T$ with

$$\begin{aligned} \psi_i(t, \tilde{b}) &= t^{i-1} \quad i = 1, \dots, s \\ \psi_{s+i}(t, \tilde{b}) &= \frac{1}{t - \tilde{b}_i} \quad i = 1, \dots, 2k, \end{aligned} \tag{4.3}$$

where $\tilde{b} = (\tilde{b}_1, \dots, \tilde{b}_{2k})^T$ is a fixed vector with pairwise different components. With the notation

$$L(\Delta) = \begin{pmatrix} I_s & 0 \\ 0 & G_k(\Delta) \end{pmatrix} \in \mathbb{R}^{m \times m}$$

$$G_k(\Delta) = \begin{pmatrix} G(\Delta) & & \\ & \ddots & \\ & & G(\Delta) \end{pmatrix} \in \mathbb{R}^{2k \times 2k}; \quad G(\Delta) = \begin{pmatrix} 1 & 0 \\ -\frac{1}{\Delta} & \frac{1}{\Delta} \end{pmatrix} \in \mathbb{R}^{2 \times 2}$$

(here I_s is the $s \times s$ identity matrix) it is easy to verify that

$$(4.4) \qquad f(t,b) = L(\Delta)\psi(t,\tilde{b}_\Delta) + o(1) \ ,$$

where $\tilde{b}_\Delta = (b_1, b_1 + \Delta, \ldots, b_k, b_k + \Delta)^T$. For a fixed vector $T = (t_1, \ldots, t_m)^T \in \mathbb{R}^m$ with ordered components $t_1 < \ldots < t_m$ such that $t_i \in I$ $(i = 1, \ldots, m)$ define the matrices

$$\begin{aligned} F(T,b) &= (f_i(t_j,b))_{i,j=1}^m \ , \\ \psi(T,\tilde{b}) &= (\psi_i(t_j,\tilde{b}))_{i,j=1}^m, \end{aligned}$$

then we obtain from (4.4)

$$(4.5) \qquad \det F(T,b) \ = \ \lim_{\Delta \to 0} \frac{1}{\Delta^k} \psi(T, \tilde{b}_\Delta) \ = \ \frac{\prod_{1 \le i < j \le m}(t_j - t_i) \prod_{1 \le i < j \le k}(b_i - b_j)^4}{\prod_{i=1}^k \prod_{j=1}^m (t_j - b_i)^2},$$

where the last identity follows from the fact that $\psi(T,\tilde{b})$ is a Cauchy-Vandermonde matrix, which implies

$$\det \psi(T,\tilde{b}) = \frac{\prod_{1 \le i < j \le m}(t_j - t_i) \prod_{1 \le i < j \le 2k}(\tilde{b}_i - \tilde{b}_j)}{\prod_{i=1}^{2k} \prod_{j=1}^m (t_j - \tilde{b}_i)}.$$

Now for any $b \in \mathcal{B}$ the right hand side does not vanish and is of one sign independently of T. Consequently $\{f_i(t,b) \mid i = 1, \ldots, m\}$ is a Chebyshev system on the interval I. The assertion regarding the system $\{\bar{f}_i(t,x) \mid i = 1, \ldots, m\}$ is proved similary and therefore left to the reader.

□

The case $k = 1$ will be studied more explicitly in Example 4.5 and 4.6. Note that the third part of Lemma 4.1 shows that for $k \ge 2$ the main condition in Theorem 2.1 in the paper of Imhof and Studden (2001) is <u>not</u> satisfied in general for the linear regression model with the functions given by (4.1). These authors assumed that every subsystem of $\{f_1, \ldots, f_m\}$ which consists of $m - 1$ of these functions is a weak Chebyshev system on the interval I. Because the design problem for this model is equivalent to the design problem for the model (1.2) (where the nonlinear parametes are not known and have to be estimated) it follows that in general we cannot expect locally E-optimal designs for the rational model to be supported at the Chebyshev points. However, the linearized regression model (1.4) is a special case of the general model (2.2) with $\varphi(t,b) = (t-b)^{-1}$ and all results of Section 3 are applicable here. In particular we obtain that the E-optimal designs and the optimal designs for estimating the individual coefficients $a_{s+1}, b_1, \ldots, a_{s+k}, b_k$ are supported at the Chebyshev points if the nonlinear parameters $b_1, \ldots, b_k$ are sufficiently close [see Theorem 3.2, Lemma 3.3 and Remark 3.6].

Theorem 4.2.

(i) If $s = 1$, then the Chebyshev points $s_1 = s_1(b), \ldots, s_m = s_m(b)$ for the system of regression functions in (4.1) on the interval $[-1,1]$ are given the roots of the polynomial

$$(4.6) \qquad (1-t^2) \sum_{i=0}^{4k} d_i U_{-2k+s+i-2}(t),$$

where $U_j(x)$ denotes the jth Chebyshev polynomial of the second kind [see Szegö (1975)], $U_{-1}(x) = 0, U_{-n}(x) = -U_{n-2}(x)$ and the factors $d_0, \dots, d_{4k}$ are defined as the coefficients of the polynomial

$$\sum_{i=0}^{4k} d_i t^i = \prod_{i=1}^{k} (t - \tau_i)^4, \tag{4.7}$$

where

$$2b_i = \tau_i + \frac{1}{\tau_i} \quad i = 1, \dots, k.$$

(ii) Let $\Omega_E \subset \mathcal{B}$ denote the set of all b such that an E-optimal design for the model (1.4) is given by (2.14) and (2.12), then $\Omega_E \neq \emptyset$.

Proof. The second part of the theorem is a direct consequence of Lemma 4.1 and Theorem 3.2, while the first part of the proposition follows by Theorem A.2 in Imhof and Studden (2001).

□

Remark 4.3.

(a) The Chebyshev points for the system (4.1) on an arbitrary finite interval $I \subset \mathbb{R}$ can be obtained by rescaling the points onto the interval $[-1, 1]$. The case $s = 0$ and $I = [0, \infty)$ will be discussed in more detail in Examples 4.5 and 4.7.

(b) It follows from Theorem 3.2 that the set Ω_E defined in the second part of Theorem 4.1 contains the set $\Omega_{\varepsilon,\Delta}$ defined in (3.2) for sufficiently small ε. In other words: if the nonlinear parameters $b_1, \dots, b_k$ are sufficiently close the locally E-optimal design will be supported at the Chebyshev points with weights given by (2.12). Moreover, we will demonstrate in the subsequent examples that in many cases the set Ω_E coincides with the full set $\mathcal{B}$.

(c) In applications the Chebyshev points can be calculated numerically with the Remez algorithm [see Studden and Tsay (1976) or DeVore and Lorentz (1993)]. In some cases these points can be obtained explicitly (see Example 4.5 and 4.6).

Remark 4.4. We note that a similar result is valid for c-optimal designs in the rational regression model (1.4). For example assume that one of the assertions of Lemma 4.1 is valid and that we are interested in estimating a linear combination $c^T\beta$ of the parameters in the rational model (1.4). We obtain from Lemma 3.3. that if $c \in \mathbb{R}^m$ satisfies $c^T\tilde{\gamma} \neq 0$, then for sufficiently small ε and any $b \in \Omega_{\varepsilon,\Delta}$ the design $\xi_c(b)$ defined in (2.15) and (2.16) is c-optimal. In particular this is true for $c = e_{s+2j-1}$ (for all $j = 1, \dots, k$) and the vector $c = e_{s+2j}$ if the index j satsifies the condition (3.10). Note that due to the third part of Lemma 4.1 in the case $k \geq 2$ there exists $b \in \mathcal{B}$ such that the e_{s+2j}-optimal design is not necessarily supported at the Chebyshev points. However, from Theorem 3.5 it follows that for a vector $b \in \mathcal{B}$ satisfying (3.1) with $\delta \to 0$ and any vector c with $c^T\tilde{\gamma} \neq 0$ we have for the designs $\xi^*_{c^*}(b)$ and $\xi^*_c(b)$ defined by (2.14) and (2.15)

$$\begin{aligned} \xi^*_{c^*}(b) &\to \bar{\xi}_{e_m}(x) \\ \xi^*_c(b) &\to \bar{\xi}_{e_m}(x) , \end{aligned}$$

where the design $\bar{\xi}_{e_m}(x)$ is defined in (3.11) and (3.12), respectively, and e_m-optimal in the limiting model with the regression functions (4.2).

Example 4.5. Consider the rational model

$$Y = \frac{a}{t-b} + \varepsilon; \quad t \in [0,\infty) \tag{4.8}$$

with $b < 0$ (here we have $k = 1, s = 0, I = [0,\infty)$). The corresponding equivalent linear regression model is given by

$$Y = \beta^T f(t,b) = \frac{\beta_1}{t-b} + \frac{\beta_2}{(t-b)^2}. \tag{4.9}$$

In this case it follows from the first part of Lemma 4.1 that the system of regression functions

$$\Big\{\frac{1}{t-b}, \frac{1}{(t-b)^2}\Big\} = \{f_1(t), f_2(t)\}$$

is a Chebyshev system on the interval $[0,\infty)$, whenever $b < 0$. Moreover, any subsystem (consisting of one function) is obviously a Chebyshev system on the interval $[0,\infty)$. The Chebyshev points are the (local) extrema of the function

$$g(t) = \rho\Big(\frac{1}{t-b} + \frac{\kappa}{(t-b)^2}\Big),$$

where ρ and κ are determined by the condition

$$\begin{aligned} g(t) &\le 1 \quad \forall\, t \in [0,\infty) \\ g(s_j) &= (-1)^j \quad j = 1,2. \end{aligned}$$

It is easy to see that $s_1 = 0$ and that s_2 is the positive solution of the equation $g'(t) = 0$, which implies

$$\kappa = \frac{b - s_2}{2}.$$

Observing the relation $g(s_1) = -g(s_2)$ we obtain by a straightforward calculation

$$s_2 = \sqrt{2}|b| = -\sqrt{2}b$$

and the condition $g(s_1) = g(0) = -1$ implies

$$\rho = \frac{-2}{\sqrt{2}-1} b,$$

which determines the Chebyshev polynomial explicitly. Now we consider the design $\xi_c^*(b)$ defined in (2.15) as a candidate for the c-optimal design in the model (4.9). The weights (for any $c \in \mathbb{R}^2$) are obtained from formula (2.16), where the matrix F is given by

$$F = (f_i(s_j))_{i,j=1}^2 = \begin{pmatrix} \frac{1}{|b|} & \frac{1}{(\sqrt{2}+1)|b|} \\ \frac{1}{b^2} & \frac{1}{(\sqrt{2}+1)^2 b^2} \end{pmatrix}.$$

A straightforward calculation shows that

$$F^{-1}c = \frac{1}{2}\begin{pmatrix} |b|(-\sqrt{2}c_1 + (2+\sqrt{2})c_2 b) \\ -|b|(4+3\sqrt{2})(-c_1 + c_2 b) \end{pmatrix},$$

which gives

$$\xi_c^*(b) = \begin{pmatrix} 0 & \sqrt{2}|b| \\ w_1 & w_2 \end{pmatrix}, \tag{4.10}$$

where the weights are given by

$$\omega_1 = 1 - \omega_2 = \frac{|b(-\sqrt{2}c_1 + (2+\sqrt{2})c_2 b)|}{|b|\{-\sqrt{2}c_1 + (2+\sqrt{2})c_2 b\ | + (4+3\sqrt{2})| - c_1 + c_2 b|\}}.$$

It can easily be checked by Elfving's theorem [see Elfving (1952)] or by the equivalence theorem for c-optimality [see Pukelsheim (1993)] that this design is in fact c-optimal in the regression model (4.9) whenever

$$\frac{c_2}{c_1} \notin \Big[\frac{1}{b}, \frac{1}{(1+\sqrt{2})b}\Big].$$

In the remaining cases the c-optimal design is a one point design supported at $t = b - \frac{c_1}{c_2}$. In particular, by Lemma 2.3, the e_1- and e_2-optimal design for estimating the coefficients β_1 and β_2 in the model (4.9) are given by

$$\begin{aligned} \xi_{e_1}^*(b) &= \begin{pmatrix} 0 & \sqrt{2}|b| \\ \frac{1}{4}(2-\sqrt{2}) & \frac{1}{4}(2+\sqrt{2}) \end{pmatrix}, \\ \xi_{e_2}^*(b) &= \begin{pmatrix} 0 & \sqrt{2}|b| \\ 1-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \end{aligned} \tag{4.11}$$

respectively. It follows from the results of Imhof and Studden (2001) that an E-optimal design in the regression model (4.9) is given by the c^*-optimal design for the Chebyshev vector

$$c^* = (1+\sqrt{2})|b|(-2, |b|(1+\sqrt{2}))^T,$$

that is

$$\xi_E^* = \begin{pmatrix} 0 & \sqrt{2}|b| \\ w_1 & w_2 \end{pmatrix}, \tag{4.12}$$

where

$$w_1 = \frac{1}{2}\frac{(2-\sqrt{2})(6-4\sqrt{2}+b^2)}{b^2+12-8\sqrt{2}} = 1 - \frac{1}{2}\frac{\sqrt{2}(2\sqrt{2}-2+b^2)}{b^2+12-8\sqrt{2}} = 1 - w_2.$$

The corresponding information matrix is obtained by a tedious calculation

$$M(\xi_E^*(b), b) = \begin{pmatrix} \dfrac{(\sqrt{2}-1)(b^2+6\sqrt{2}-8)}{b^2(b^2+12-8\sqrt{2})} & \dfrac{2(3-\sqrt{2})(b^2+\sqrt{2}-1)}{b^3(b^2+12-8\sqrt{2})} \\[2ex] \dfrac{2(3-\sqrt{2})(b^2+\sqrt{2}-1)}{b^3(b^2+12-8\sqrt{2})} & \dfrac{(8\sqrt{2}-11)(7b^2+16\sqrt{2}-20)}{7b^4(b^2+12-8\sqrt{2})} \end{pmatrix} \tag{4.13}$$

and has a minimum eigenvalue

$$\lambda_{\min}(M(\xi_E^*(b),b) = \frac{17-2\sqrt{2}}{b^2(b^2+12-8\sqrt{2})} = \frac{1}{\|c^*\|^2}$$

of multiplicity 1 with corresponding eigenvector c^*. Note that for $b \to -\infty$ this design approximates the optimal design $\xi_{e_2}^*(b)$ for estimating the individual coefficient β_2 in the rational model (4.9). It is of some interest to compare these designs with the locally D-optimal design. It follows from the results in He, Studden and Sun (1996) and a straightforward calculation that this design is given by

$$\xi_D^* = \begin{pmatrix} 0 & |b| \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{4.14}$$

The designs are now compared by their efficiencies for estimating the coefficients β_1 and β_2, i.e.

$$\text{eff}_i(\xi) = \left(\frac{e_i^T M^{-1}(\xi,b)e_i}{e_i^T M^{-1}(\xi_{e_i}^*,b)e_i} \right)^{-1} \qquad i=1,2. \tag{4.15}$$

The values $e_i^T M^{-1}(\xi_{e_i}^*,b)e_i$ can be directly obtained from the Chebyshev vector, which gives

$$e_i^T M^{-1}(\xi_{e_i}^*,b)e_i = \begin{cases} 4(1+\sqrt{2})^2 b^2 & \text{if } i=1 \\ (1+\sqrt{2})^4 b^4 & \text{if } i=2. \end{cases}$$

Now a striaghtforward calculation yields for the efficiencies of the D-optimal design defined by (4.14)

$$\text{eff}_i(\xi_D^*) = \begin{cases} \dfrac{4(\sqrt{2}+1)^2}{34} \approx 0.6857 & \text{if } i=1 \\[2ex] \dfrac{(\sqrt{2}+1)^4}{40} \approx 0.8493 & \text{if } i=2. \end{cases}$$

The corresponding efficiencies of the E-optimal design in the regression model (4.9) depend on the parameter b and are obtained by a straightforward but tedious inversion of the matrix $M(\xi_E^*(b),b)$ defined in (4.13), that is

$$\text{eff}_i(\xi_E^*(b)) = \begin{cases} \dfrac{28(b^4(5\sqrt{2}-7)+b^2(34\sqrt{2}-48)+396-280\sqrt{2})}{(9\sqrt{2}-11)(b^2-8\sqrt{2}+12)(7b^2+16\sqrt{2}-20)} & \text{if} \quad i=1 \\[3ex] \dfrac{b^4(\sqrt{2}-1)+(6\sqrt{2}-8)b^2+68-48\sqrt{2}}{(\sqrt{2}-1)(b^2-8\sqrt{2}+12)(b^2-6\sqrt{2}+8)} & \text{if} \quad i=2. \end{cases} \tag{4.16}$$

The corresponding efficiencies are depicted in Figure 4.1 for the range $b \in [-2.5,-1]$. We observe for the e_1-efficiency for all $b \le -1$

$$0.9061 \approx \frac{4(5\sqrt{2}-7)}{(8\sqrt{2}-11)} = \lim_{b\to-\infty} \text{eff}_1(\xi_E^*(b)) \le \text{eff}_1(\xi_E^*(b)) \le \text{eff}_1(\xi_E^*(-1)) \approx 0.9595,$$

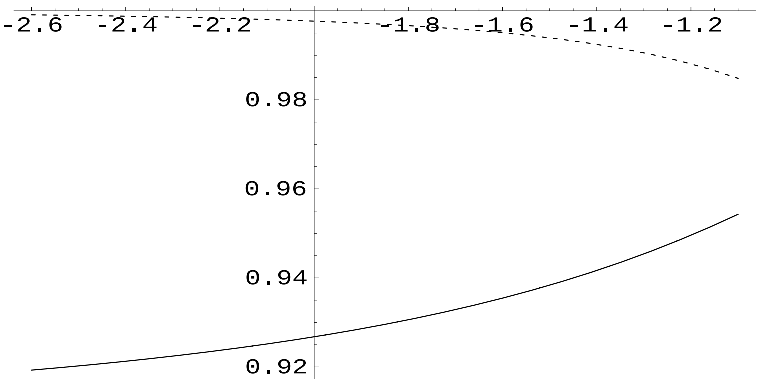

Figure 4.1: *Efficiencies of the E-optimal design $\xi^*(b)$ for estimating the individual coefficients in the regression model (4.9) for various values of $b \in [-2.5, -1]$. Solid line: $\mathit{eff}_1(\xi^*(b))$, dotted line: $\mathit{eff}_2(\xi^*(b))$.*

and similary for the e_2-efficiency

$$0.9805 \approx \text{eff}_2(\xi_E^*(-1)) \leq \text{eff}_2(\xi_E^*(b)) \leq \lim_{b \to -\infty} \text{eff}_2(\xi_E^*(b)) = 1.$$

This demonstrates that the E-optimal designs yield substantially more accurate estimates for the individual parameters in the regression model (4.9) than the D-optimal design.
We finally mention the results for the locally optimal design in the rational model (4.8), which maxi- or minimize the corresponding functional for the matrix $K_a^{-1} M(\xi, b) K_a^{-1}$, where $K_a = \text{diag}(1, -\frac{1}{a})$. Obviously the locally e_1-, e_2- and D-optimal designs are given by (4.11) and (4.14), respectively and coincide with the corresponding designs in the equivalent linear regression model (4.9). On the other hand the c-optimal design for the rational model (4.8) is obtained from the $\bar{c}$-optimal design $\xi_{\bar{c}}^*(b)$ in (4.10) for the model (4.9) with $\bar{c} = K_a c = (c_1, -c_2/a)^T$. Similary, the locally E-optimal design for the rational model (4.8) is given by

$$\xi_E^* = \begin{pmatrix} 0 & \sqrt{2}|b| \\ w_1^* & w_2^* \end{pmatrix},$$

where the weights are given by

$$w_1^* = \frac{2\sqrt{2}a^2 + (4 + 3\sqrt{2})b^2}{2\{4(1 + \sqrt{2})a^2 + (7 + 5\sqrt{2})b^2\}} = 1 - \frac{(4 + 3\sqrt{2})(2a^2 + (1 + \sqrt{2})b^2)}{2\{4(1 + \sqrt{2})a^2 + (7 + 5\sqrt{2})b^2\}} = 1 - w_2^*.$$

A comparison of the efficiencies for the D- and E-optimal design in the rational model (4.8) yields similar results as in the corresponding equivalent linear regression model (4.9). For a broad range of parameter values (a, b) the locally E-optimal designs in the rational model (4.8) are substantially more efficient for estimating the individual parametes than the locally D-optimal designs.

Example 4.6. We now consider the rational model

$$Y = a_1 + \frac{a_2}{t - b} + \varepsilon; \quad t \in [-1, 1], \tag{4.17}$$

where $|b| > 1$. The corresponding equivalent linear regression model is given by

$$Y = \beta_1 + \frac{\beta_2}{t-b} + \frac{\beta_3}{(t-b)^2} + \varepsilon; \quad t \in [-1,1], \tag{4.18}$$

and the first part of Lemma 4.1 shows that this system is a Chebyshev system on the interval $[-1,1]$. Moreover, the three subsystems obtained by deleting one of the regression functions form also weak Chebyshev systems (this follows partially from Lemma 4.1 (ii), while the remaining case has to be checked directly). Therefore the optimal designs for estimating the individual coefficients and the E-optimal design are supported at the Chebyshev points, which are given by $s_1 = -1, s_2 = 1/b, s_3 = 1$. A similar calculation as in Example 4.5 shows that the E-optimal design in the equivalent linear regression model (4.18) is given by

$$\xi_E^* = \begin{pmatrix} -1 & \frac{1}{b} & 1 \\ w_1 & w_2 & w_3 \end{pmatrix},$$

where

$$\begin{aligned}
w_1 &= \frac{b+1}{2} \cdot \frac{2b^7 - 2b^6 + 2b^5 + 2b^4 - 4b^3 - 2b^2 + b + 2}{4b^8 - 4b^4 - 4b^2 + 5}, \\
w_2 &= \frac{(b^2-1)(2b^6 + 2b^4 - 3)}{4b^8 - 4b^4 - 4b^2 + 5}, \\
w_3 &= \frac{b-1}{2} \cdot \frac{2b^7 + 2b^6 + 2b^5 - 2b^4 - 4b^3 + 2b^2 + b - 2}{4b^8 - 4b^4 - 4b^2 + 5},
\end{aligned}$$

Here we have used Lemma 2.2 and the fact that the vector of the coefficients of the Chebyshev polynomial is given by

$$c^* = (2b^2 - 1, 4b(b^2-1), 2(b^2-1)^2)^T.$$

The optimal designs for estimating the individual coefficients $\beta_1, \beta_2, \beta_3$ are given by

$$\begin{aligned}
\xi_{e_1}^* &= \begin{pmatrix} -1 & \frac{1}{b} & 1 \\ \frac{b(1+b)}{2(2b^2-1)} & \frac{b^2-1}{2b^2-1} & \frac{b(b-1)}{2(2b^2-1)} \end{pmatrix}, \\
\xi_{e_2}^* &= \begin{pmatrix} -1 & \frac{1}{b} & 1 \\ \frac{1}{8}(2+\frac{1}{b}) & \frac{1}{2} & \frac{1}{8}(2-\frac{1}{b}) \end{pmatrix}, \\
\xi_{e_3}^* &= \begin{pmatrix} -1 & \frac{1}{b} & 1 \\ -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix},
\end{aligned}$$

respectively. We note again that for $|b| \to \infty$ all designs are approximated by the optimal design $\xi_{e_3}^*$ for estimating the individual coefficient β_3. The corresponding efficiencies $\text{eff}_i(\xi_E^*(b))$ $i = 1,2,3$ are depicted in Figure 4.2 for the interval $[2,4]$ and demonstrate again that the locally E-optimal design is highly efficient for estimating the coefficients $\beta_1, \beta_2, \beta_3$ in the model (4.18). The locally D-optimal design can be obtained by similar arguments as given in Example 4.5, that is

$$\xi_D^*(b) = \begin{pmatrix} -1 & \frac{1}{b} & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix},$$

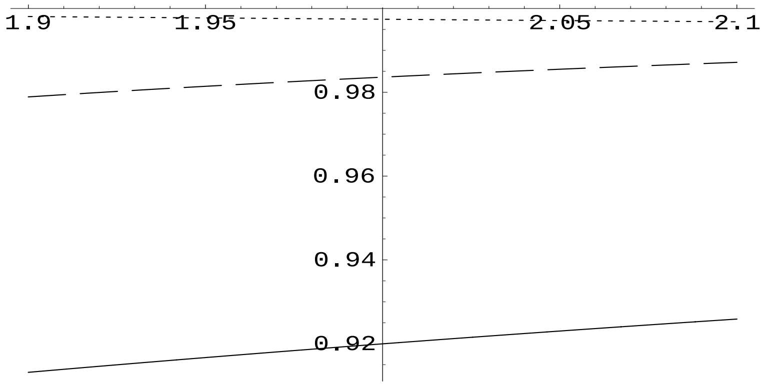

Figure 4.2: *Efficiencies of the E-optimal design $\xi^*(b)$ for estimating the individual coefficients in the regression model (4.18) for various values of $b \in [2,4]$. Solid line: $\mathrm{eff}_1(\xi^*(b))$, dotted line: $\mathrm{eff}_2(\xi^*(b))$, dashed line: $\mathrm{eff}_3(\xi^*(b))$.*

while the corresponding efficiencies can be calculated explicitly and are given by

$$\mathrm{eff}_i(\xi^*_D(b)) = \begin{cases} \dfrac{2(2b^2-1)^2}{3(3b^4-3b^2+2)} & \text{if} \quad i=1 \\[2ex] \dfrac{32b^2}{3+36b^2} & \text{if} \quad i=2 \\[2ex] \dfrac{2}{9} & \text{if} \quad i=3. \end{cases}$$

Again we observe that locally E-optimal design yield substantially more accurate estimates of the individual parameters than D-optimal designs. Finally, the locally optimal designs for the rational model (4.17) are obtained as follows. The optimal designs for estimating the individual coefficients and the locally D-optimal design coincide with the corresponding designs in the linear regression model (4.18) while the locally E-optimal design puts masses

$$\begin{aligned} w_1^* &= \frac{2(b^2-1)^4 + a_2^2 b(8b^5+4b^4-14b^3-6b^2+7b+3)}{2\{4(b^2-1)^4 + a_2^2(16b^6-28b^4+12b^2+1\}}, \\ w_2^* &= \frac{(b^2-1)\{2(b^2-1)^3 + a_2^2(8b^4-6b^2-1)\}}{4(b^2-1) + a_2^2(16b^6-28b^4+12b^2+1)}, \\ w_3^* &= \frac{2(b^2-1)^4 + a_2^2 b_2(8b^5-4b^4-14b^2+6b^2+7b-3)}{2\{4(b^2-1)^4 + a_2^2(16b^6-28b^4+12b^2+1)\}} \end{aligned}$$

at the points $-1, 1/b$ and 1, respectively.

Example 4.7. We now discuss optimal designs for the rational model

$$Y = \frac{a_1}{t-b_1} + \frac{a_2}{t-b_2} + \varepsilon; \quad t \in [0,\infty) \tag{4.19}$$

where $b_1, b_2 < 0; |b_2 - b_2| > 0$ $(k=2, s=0)$. The corresponding equivalent linear regression model is given by

$$Y = \frac{\beta_1}{t-b_1} + \frac{\beta_2}{(t-b_1)^2} + \frac{\beta_3}{t-b_2} + \frac{\beta_4}{(t-b_2)^2} + \varepsilon. \tag{4.20}$$

Locally D-optimal designs for the model (4.19) [or equivalently (4.20)] have been determined by Melas (2001), while the optimal designs for estimating the individual coefficients can be obtained numerically from the results of this paper. We now compare these designs by looking at D-, E- and e_i-efficiencies. For the sake of brevity we restrict ourselves to the model (4.20), which corresponds to the locally optimal design problem for the model (4.19) with $(a_1, a_2) = (1, 1)$. In our comparison we will also include the E-optimal design in the limiting model under assumption (3.1), i.e.

$$Y = \frac{\beta_1}{t-x} + \frac{\beta_2}{(t-x)^2} + \frac{\beta_3}{(t-x)^3} + \frac{\beta_4}{(t-x)^4} + \varepsilon \tag{4.21}$$

where the parameter x is chosen as $x = (b_1 + b_2)/2$. Without loss of generality we assume that $x = -1$, because in the general case the optimal designs can be obtained by a simple scaling argument. The limiting optimal design was obtained numerically and is given by

$$\bar{\xi}_E(-1) = \begin{pmatrix} 0 & 0.18 & 1.08 & 7.9 \\ 0.13 & 0.26 & 0.27 & 0.34 \end{pmatrix}. \tag{4.22}$$

Table 4.1: *D- and E-optimal designs for linear regression model (4.20) on the interval $[0, \infty)$, where $b_1 = -1 - z$, $b_2 = -1 + z$. These designs are locally D- and E-optimal in the rational model (4.19) for the initial parameter $a_1 = a_2 = 1$. Note that the smallest support point of the D-optimal design (t^*_{1D}) and E-optimal design (t^*_{1E}) are equal to 0 and that the masses of the D-optimal design are equal.*

z	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
t^*_{2D}	0.21	0.20	0.20	0.19	0.17	0.15	0.13	0.10	0.06	0.04
t^*_{3D}	1.00	0.98	0.95	0.92	0.87	0.80	0.71	0.60	0.44	0.31
t^*_{4D}	4.78	4.73	4.65	4.54	4.39	4.19	3.94	3.60	3.13	2.78
t^*_{2E}	0.18	0.17	0.17	0.16	0.15	0.13	0.11	0.09	0.05	0.03
t^*_{3E}	1.08	1.06	1.03	0.99	0.94	0.87	0.77	0.65	0.47	0.34
t^*_{4E}	7.85	7.77	7.65	7.46	7.21	6.88	6.45	5.88	5.05	4.43
w^*_{1E}	0.13	0.13	0.13	0.13	0.12	0.10	0.08	0.07	0.05	0.03
w^*_{2E}	0.26	0.26	0.27	0.26	0.25	0.22	0.20	0.17	0.13	0.10
w^*_{3E}	0.27	0.27	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
w^*_{4E}	0.34	0.33	0.33	0.33	0.36	0.39	0.44	0.49	0.54	0.59

From Theorem 3.2 we obtain that for sufficiently small

$$\Delta = \left|\frac{b_1 - b_2}{2}\right|$$

the E-optimal designs for the model (4.20) is given the design $\xi^*_{c^*}(b)$ defined in (2.12) and (2.14). From Lemma 2.2 it follows that the design $\xi^*_{c^*}(b)$ is E-optimal, whenever

$$\lambda_{c^*} := \frac{c^{*T} M(\xi^*_E(b), b) c^*}{c^{*T} c^*} \le \lambda_{(2)}(M(\xi^*_E(b), b)) = \lambda_{(2)},$$

Table 4.2: *The efficiency of the E-optimal designs ξ_E^* in the linear regression model (4.20) on the interval $[0,\infty)$ with $b_1=-1-z$, $b_2=-1+z$ and the efficiency of the E-optimal design $\xi_E^*(-1)$ given in (4.22) in the corresponding limiting model (4.21). The effiencies $\mathrm{eff}_D(\xi)$, $d_i(\xi)$ and $C_E(\xi)$are defined in (4.23), (4.24) and (4.25), respectively.*

z	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$d_1(\xi_E^*)$	0.81	0.81	0.81	0.83	0.87	1.04	1.28	0.72	0.52	0.48
$d_2(\xi_E^*)$	0.80	0.79	0.79	0.78	0.76	0.74	0.71	0.68	0.63	0.59
$d_3(\xi_E^*)$	0.81	0.81	0.81	0.81	0.83	0.86	0.94	1.08	1.38	1.79
$d_4(\xi_E^*)$	0.82	0.82	0.84	0.85	0.89	0.97	1.12	1.36	1.89	2.53
$d_1(\bar{\xi}_E^*(-1))$	0.81	0.81	0.82	0.83	0.87	0.93	0.95	0.92	1.14	1.37
$d_2(\bar{\xi}_E^*(-1))$	0.80	0.79	0.80	0.82	0.86	0.94	1.09	1.38	2.04	2.81
$d_3(\bar{\xi}_E^*(-1))$	0.81	0.81	0.81	0.83	0.86	0.93	1.09	1.51	3.42	10.00
$d_4(\bar{\xi}_E^*(-1))$	0.82	0.82	0.84	0.85	0.88	0.94	1.08	1.49	3.48	10.59
$\mathrm{eff}_D(\xi_E^*)$	0.89	0.89	0.89	0.89	0.88	0.85	0.81	0.75	0.67	0.60
$\mathrm{eff}_D(\bar{\xi}_E^*(-1))$	0.89	0.89	0.89	0.88	0.88	0.87	0.84	0.78	0.63	0.48
$C_E(\xi_E^*)$	1.23	1.23	1.23	1.25	1.27	1.32	1.39	1.47	1.61	1.75
$C_E(\bar{\xi}_E^*(-1))$	1.23	1.23	1.23	1.22	1.16	1.08	0.92	0.72	0.50	0.38

where $\lambda_{\min}(M(\xi_E^*(b),b)) \le \lambda_{(2)} \le \ldots \le \lambda_{(m)}$ denote the ordered eigenvalues of the matrix $M(\xi_E^*(b),b)$. The ratio $\lambda_{(2)}/\lambda_{c^*}$ is exemplarily depicted in Figure 4.3 for $b_1=1$ and a broad range of b_2 values, which shows that it is always bigger than 1. Other cases yield a similar picture and practically the locally E-optimal design for the rational model (4.19) and the equivalent linear regression model (4.20) is always supported at the Chebyshev points and given by (2.12) and (2.14). In Table 4.1 and 4.2 we give the main characteristics and efficiencies for the locally E- and D-optimal design $\xi_E^*(b), \xi_D^*(b)$ and for the E-optimal design $\bar{\xi}_E^*(\frac{b_1+b_2}{2})$ in the limiting regression model (4.21). The efficiencies are calculated with respect to the D-optimal design for various values of the nonlinear parameters b_1, b_2 and are defined by

$$\mathrm{eff}_D(\xi) = \left(\frac{\det M(\xi,b)}{\det M(\xi_D^*,b)}\right)^{1/m} \tag{4.23}$$

$$d_i(\xi) = \frac{e_i^T M^{-1}(\xi,b)e_i}{e_i^T M^{-1}(\xi_D^*,b)e_i} \tag{4.24}$$

(in other words: we compare the performance of the design ξ for estimating individual coefficients with respect to the D-optimal design) and

$$C_E(\xi) = \frac{\lambda_{\min}(M(\xi,b))}{\lambda_{\min}(M(\xi_D^*,b)}. \tag{4.25}$$

Again we observe a very good performance of the E-optimal designs. These designs produce a reasonable D-efficiency for a moderate size of the difference $|b_1-b_2|$, but are in many cases substantially more efficient than the D-optimal designs for estimating the individual coefficients.

The behaviour of the design $\bar{\xi}_E$ in the limiting regression model (4.20) is interesting from a practical point of view because it is very similar to the performance of the E-optimal design for a broad range of b_1 and b_2 values. Consequently, this design might be appropriate if rather unprecise prior information for the nonlinear parameters is available. For example, if it is known (by scientific background) that $b_1 \in [\underline{b}_1, \bar{b}_1]$; $b_2 \in [\underline{b}_2, \bar{b}_2]$ the design

$$\bar{\xi}_E\Big(\frac{\underline{b}_1 + \bar{b}_2}{2}\Big)$$

might be a robust choice for practical experiments.

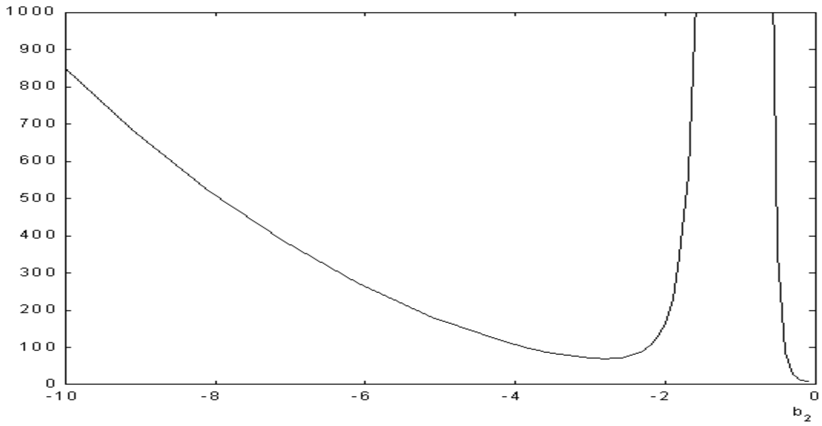

Figure 4.3: *The ratio $\lambda_{(2)}/\lambda_{c^*}$ for the design $\xi_E^*(b)$, where $b = (-1, b_2)$. The designs are E-optimal if this ratio is larger or equal than* 1.

Example 4.8. Our final example discusses the rational model (4.19) with an additional term for the intercept

$$Y = a_1 + \frac{a_2}{t - b_1} + \frac{a_3}{t - b_2} + \varepsilon; \quad t \in [-1, 1] \tag{4.26}$$

where $|b_i| > 1$ $(i = 1, 2)$ and $|b_2 - b_2| > 0$ (this corresponds to the case $k = 2, s = 1$ in the general model (1.4). The limiting model is given by

$$Y = \beta_1 + \frac{\beta_2}{t - x} + \frac{\beta_3}{(t - x)^2} + \frac{\beta_4}{(t - x)^3} + \frac{\beta_5}{(t - x)^4} + \varepsilon. \tag{4.27}$$

The notation is essentially the same as in the previous example. Our numerical study showed that the locally E-optimal design for the model (4.26) is supported at the Chebyshev points for all choices of the parameters (b_1, b_2) ($|b_i| > 1, b_1 \neq b_2$). In Table 4.3 and 4.4 we display the main features of the locally E- and D-optimal designs ξ_E^*, ξ_D^* and the E-optimal design $\bar{\xi}_E\Big(\frac{b_1+b_2}{2}\Big)$ in the limiting regression model (4.27), which is given by

$$\bar{\xi}_E(-3) = \begin{pmatrix} -1 & -0.84 & -0.33 & 0.49 & 1 \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \end{pmatrix}. \tag{4.28}$$

The conclusions are very similar as in the previous Example 4.7. This indicates that the observations from this example are in some sense representative.

Table 4.3: *Locally D- and E-optimal designs for the rational regression model (4.26) on the interval* $[-1, 1]$, *where* $b_1 = -3 - z$, $b_2 = -3 + z$, $a_3 = a_2 = 1$. *Note that the largest and smallest support point of the locally E- and D-optimal design satisfy* $t^*_{5E} = t^*_{5D} = 1$ *and* $t^*_{1E} = t^*_{1D} = -1$, *respectively, while the masses of the locally D-optimal design are all equal.*

z	0.1	0.2	0.3	0.5	1	1.5	1.9
t^*_{2D}	-0.81	-0.81	-0.81	-0.82	-0.83	-0.87	-0.95
t^*_{3D}	-0.32	-0.34	-0.34	-0.34	-0.38	-0.47	-0.70
t^*_{4D}	0.41	0.41	0.41	0.40	0.37	0.29	0.08
t^*_{2E}	-0.84	-0.84	-0.84	-0.85	-0.86	-0.89	-0.96
t^*_{3E}	-0.33	-0.33	-0.34	-0.34	-0.38	-0.47	-0.70
t^*_{4E}	0.49	0.49	0.49	0.48	0.45	0.38	0.17
w^*_{1E}	0.13	0.13	0.13	0.12	0.11	0.09	0.05
w^*_{2E}	0.25	0.25	0.25	0.25	0.22	0.20	0.14
w^*_{3E}	0.25	0.25	0.25	0.25	0.25	0.25	0.25
w^*_{4E}	0.25	0.25	0.25	0.25	0.28	0.30	0.36
w^*_{5E}	0.12	0.12	0.12	0.13	0.14	0.16	0.20

Table 4.4: *The efficiency of the E-optimal designs* ξ^*_E *in the rational regression model (4.26) on the interval* $[-1, 1]$ *with* $b_1 = -3 - z$, $b_2 = -3 + z$, $a_3 = a_2 = 1$ *and the efficiency of the E-optimal design* $\bar{\xi}_E(-1)$ *given in (4.28) in the corresponding limiting model (4.27). The effiencies* $\text{eff}_D(\xi)$, $d_i(\xi)$ *and* $C_E(\xi)$ *are defined in (4.23), (4.24) and (4.25), respectively.*

z	0.1	0.2	0.3	0.5	1	1.5	1.9
$d_1(\xi^*_E)$	0.86	0.87	0.87	0.87	0.84	0.82	0.75
$d_2(\xi^*_E)$	0.83	0.84	0.84	0.84	0.85	0.90	1.21
$d_3(\xi^*_E)$	0.83	0.84	0.84	0.84	0.87	0.97	1.53
$d_4(\xi^*_E)$	0.83	0.84	0.84	0.83	0.88	0.81	0.74
$d_5(\xi^*_E)$	0.83	0.84	0.84	0.84	0.83	0.82	0.76
$d_1(\bar{\xi}^*_E(-3))$	0.86	0.88	0.88	0.89	0.96	1.31	3.62
$d_2(\bar{\xi}^*_E(-3))$	0.83	0.84	0.84	0.84	0.85	1.05	5.74
$d_3(\bar{\xi}^*_E(-3))$	0.83	0.84	0.84	0.84	0.84	1.01	5.72
$d_4(\bar{\xi}^*_E(-3))$	0.83	0.84	0.84	0.83	1.08	1.28	3.74
$d_5(\bar{\xi}^*_E(-3))$	0.83	0.84	0.84	0.84	0.88	1.21	3.94
$\text{eff}_D(\xi^*_E)$	0.93	0.93	0.93	0.93	0.93	0.91	0.83
$\text{eff}_D(\bar{\xi}^*_E(-3))$	0.93	0.93	0.93	0.93	0.93	0.91	0.66
$C_E(\xi^*_E)$	1.20	1.19	1.19	1.19	1.20	1.22	1.33
$C_E(\bar{\xi}^*_E(-3))$	1.20	1.19	1.19	1.19	1.14	0.82	0.26

5 Appendix: Some auxiliary results

Recall the notation in Section 2 and 3

$$
\begin{aligned}
f_i(t) &= h_i(t) \quad t = 1, \dots, s \\
f_{s+2i-1}(t) &= f_{s+2i-1}(t, b) = \varphi(t, b_i) \quad i = 1, \dots, k \\
f_{s+2i}(t) &= f_{s+2i}(t, b) = \varphi'(t, b_i) \quad i = 1, \dots, k
\end{aligned} \tag{5.1}
$$

$$
\begin{aligned}
\bar{f}_i(t) &= h_i(t) \quad i = 1, \dots, s \\
\bar{f}_{s+i}(t) &= \bar{f}_{s+i}(t, x) = \varphi^{(i)}(t, x) \quad i = 1, \dots, 2k.
\end{aligned} \tag{5.2}
$$

Let $f(t, b) = (f_1(t), \dots, f_m(t))^T$ and $\bar{f}(t, x) = (\bar{f}_1(t), \dots, \bar{f}_m(t))^T$ denote the corresponding vectors of regression functions ($m = s + 2k$) and consider a design ξ on the interval I with at least m support points. In this appendix we investigate the relation between the information matrices

$$M(\xi, b) = \int_I f(t, b) f^T(t, b) d\xi(t)$$

and

$$\bar{M}(\xi, b) = \int_I \bar{f}(t, x) \bar{f}^T(t, x) d\xi(t)$$

defined by (2.4) and (3.5), respectively, if

$$\delta_i = r_i \delta = b_i - x \to 0 \quad i = 1, \dots, k \tag{5.3}$$

[see condition (3.1)], where the components of the vector $r = (r_1, \dots, r_k)$ are different and ordered.

Theorem 5.1. *Assume that $\varphi \in C^{0,2k-1}$ and ξ is an arbitrary design, such that the matrix $\bar{M}(\xi, b)$ is nonsingular. If assumption (5.3) is satisfied, it follows that for sufficiently small δ the matrix $M(\xi, b)$ is invertible and if $\delta \to 0$*

$$M^{-1}(\xi, b) = \delta^{-4k+4} T(\delta) \begin{pmatrix} \bar{M}^{(1)}(\xi) & \bar{M}^{(2)}(\xi) F \\ F^T \bar{M}^{(2)^T}(\xi) & \gamma\gamma^T h + o(1) \end{pmatrix} T(\delta) + o(1), \tag{5.4}$$

where the matrices $T(\delta) \in \mathbb{R}^{m \times m}$ and $\bar{M}^{(1)}(\xi) \in \mathbb{R}^{s \times s}, \bar{M}^{(2)}(\xi) \in \mathbb{R}^{s \times 2k}$ and $\bar{M}^{(3)}(\xi) \in \mathbb{R}^{2k \times 2k}$ are defined by

$$T(\delta) = \; diag\Big(\underbrace{\delta^{2k-2}, \dots, \delta^{2k-2}}_{s}, \underbrace{\frac{1}{\delta}, 1, \frac{1}{\delta}, 1, \dots, \frac{1}{\delta}, 1}_{2k}\Big),$$

$$\begin{pmatrix} \bar{M}^{(1)} & \bar{M}^{(2)}(\xi) \\ \bar{M}^{(2)^T}(\xi) & \bar{M}^{(3)}(\xi) \end{pmatrix} = \bar{M}^{-1}(\xi, x),$$

the vector $\gamma = (\gamma_1, \dots, \gamma_{2k})^T$ and $h \in \mathbb{R}$ are given by $h = [(2k-1)!]^2 e_m^T \bar{M}^{-1}(\xi, x) e_m$,

$$
\begin{aligned}
\gamma_{2i} &= \prod_{j \neq i} (r_i - r_j)^{-2} \quad i = 1, \dots, k \\
\gamma_{2i-1} &= -\gamma_{2i} \sum_{j \neq i} \frac{2}{r_i - r_j} \quad i = 1, \dots, k,
\end{aligned} \tag{5.5}
$$

and the matrix $F \in \mathbb{R}^{2k\times 2k}$ is defined by

$$F = \begin{pmatrix} 0 & \dots & 0 & \frac{\gamma_1}{0!} \\ \vdots & & & \\ 0 & \dots & 0 & \frac{\gamma_{2k}}{(2k-1)!} \end{pmatrix}.$$

Proof. Define $\psi(\delta) = (1, \delta, \dots, \delta^{2k-1})^T$ and introduce the matrices

$$L = (\ell_1, \dots, \ell_{2k})^T \in \mathbb{R}^{2k\times 2k} \tag{5.6}$$

$$U = \operatorname{diag}\Big(1, \frac{1}{1!}, \frac{1}{2!}, \dots, \frac{1}{(2k-1)!}\Big) \in \mathbb{R}^{2k\times 2k} \tag{5.7}$$

where $\ell_{2i-1} = \psi(\delta_i); \ell_{2i} = \psi'(\delta_i)\quad i = 1, \dots, k)$. For fixed $t \in I$ we use the Taylor expansions

$$\begin{aligned} \varphi(t, x+\delta) &= \sum_{j=0}^{2k-1} \frac{\varphi^{(i)}(t,x)}{j!}\delta^j + o(\delta^{2k-1}) \\ \varphi'(t, x+\delta) &= \sum_{j=1}^{2k-1} \frac{\varphi^{(i)}(t,x)}{(j-1)!}\delta^{j-1} + o(\delta^{2k-2}) \end{aligned}$$

to obtain the representation

$$f(t, b+\delta r) = \begin{pmatrix} I_s & 0 \\ 0 & LU \end{pmatrix} \bar{f}(t,x) + \begin{pmatrix} 0 \\ \tilde{f}(t) \end{pmatrix}, \tag{5.8}$$

where $I_s \in \mathbb{R}^{s\times s}$ denotes the identity matrix and the vector $\tilde{f}$ is of order

$$\tilde{f}(t) = (o(\delta^{2k-1}), o(\delta^{2k-2}), o(\delta^{2k-1}), \dots, o(\delta^{2k-2}))^T. \tag{5.9}$$

It follows from p. 127-129 in Karlin and Studden (1966) that

$$\det L = \prod_{1\le i<j\le k} (\delta_i - \delta_j)^4$$

and consequently $V = (v_1, \dots, v_{2k}) := L^{-1}$ exists. The equality $LV = I_m$ implies the equations

$$\begin{aligned} v_{2i}^T\psi(\delta_j) &= 0, \qquad v_{2i}^T\psi'(\delta_j) = 0, \quad j \neq i, \\ v_{2i}^T\psi(\delta_i) &= 0, \qquad v_{2i}^T\psi'(\delta_i) = 1, \end{aligned}$$

which shows that $\delta_1, \dots, \delta_{i-1}, \delta_{i+1}, \dots, \delta_k$ are roots of multiplicity two of the polynomial $v_{2i}^T\psi(\delta)$ and δ_i is a root of multiplicity one. Because this polynomial has degree $2k-1$ it follows that

$$v_{2i}^T\psi(\delta) = (\delta - \delta_i)\prod_{j\neq i}\Big(\frac{\delta - \delta_j}{\delta_j - \delta_i}\Big)^2, \tag{5.10}$$

and a similar argument shows

$$v_{2i-1}^T\psi(\delta) = \frac{\delta - \alpha_i}{\delta_i - \alpha_i}\prod_{j\neq i}\Big(\frac{\delta - \delta_j}{\delta_i - \delta_j}\Big)^2, \tag{5.11}$$

where the constants $\alpha_1, \ldots, \alpha_k$ are given by

$$\alpha_i = \delta_i + \Big(\sum_{j \neq i} \frac{2}{\delta_i - \delta_j} \Big)^{-1} \quad i = 1, \ldots, k. \tag{5.12}$$

From (5.8) and (5.9) we therefore obtain

$$f(t, b+\delta r) f^T(t, b+\delta r) = \begin{pmatrix} I_s & 0 \\ 0 & LU \end{pmatrix} \bar{f}(t,x) \bar{f}^T(t,x) \begin{pmatrix} I_s & 0 \\ 0 & LU \end{pmatrix}^T + o(\delta^{2k-2}),$$

and integrating the right hand side with respect to the design ξ shows that

$$M(\xi, b+\delta r) = \begin{pmatrix} I_s & 0 \\ 0 & LU \end{pmatrix} \bar{M}(\xi, x) \begin{pmatrix} I_s & 0 \\ 0 & LU \end{pmatrix}^T + o(\delta^{2k-2}). \tag{5.13}$$

Now define $H_1(\delta) = \operatorname{diag}(\delta^{2k-1}, \delta^{2k-2}, \delta^{2k-1}, \ldots, \delta^{2k-1}, \delta^{2k-2}) \in \mathbb{R}^{2k \times 2k}$ and

$$H(\delta) = \begin{pmatrix} I_s & 0 \\ 0 & H_1(\delta) \end{pmatrix} \in \mathbb{R}^{m \times m},$$

then we obtain from (5.10) and (5.11) that

$$H_1(\delta)(L^{-1})^T = (0 \mid \gamma) + o(1) \ ,$$

where $\gamma = (\gamma_1, \ldots, \gamma_{2k})^T$ is defined by formula (5.5) and $0 \in \mathbb{R}^{2k \times 2k-1}$ denotes the matrix with all entries equal to zero. By (5.2) this implies for the inverse of the matrix $M(\xi, b+\delta r)$

$$\begin{aligned} M^{-1}(\xi, b+\delta r) &= H^{-1}(\delta) \left\{ \begin{pmatrix} I & 0 \\ 0 & F \end{pmatrix} \bar{M}^{-1}(\xi, x) \begin{pmatrix} I & 0 \\ 0 & F^T \end{pmatrix} + o(1) \right\} H^{-1}(\delta) \\ &= \delta^{-4k+4} T(\delta) \left\{ \begin{pmatrix} \bar{M}^{(1)}(\xi) & \bar{M}^{(2)}(\xi) F^T \\ F \bar{M}^{(2)^T}(\xi) & F \bar{M}^{(3)}(\xi) F^T \end{pmatrix} + o(1) \right\} T(\delta), \end{aligned}$$

where the matrix F is given by $F = (0|\gamma)U^{-1} \in \mathbb{R}^{2k \times 2k}$. The assertion now follows by a straightforward calculation which shows that

$$F \bar{M}^{(3)}(\xi) F^T = h \gamma \gamma^T.$$

□

Acknowledgements. The work of V.B. Melas was partially supported by the Russian Foundation of Basic Research (grant No. 00-01-00495). The work of H. Dette and V.B. Melas was also supported by the Deutsche Forschungsgemeinschaft (SFB 475: Komplexitätsreduktion in multivariaten Datenstrukturen). The authors would also like to thank Isolde Gottschlich, who typed numerous versions of this paper with considerable technical expertise.

References

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

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

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

K. Chaloner, I. Verdinelli (1995). Bayesian experimental design: a review. Statistical Science 10, 273-304.

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

H. Dette, W.J. Studden (1993). Geometry of E-optimality. Ann. Statist. 21, 416-433.

H. Dette, L. Haines (1994). E-optimal designs for linear and nonlinear models with two parameters. Biometrika 81, 739–754.

H. Dette, L. Haines and L.A. Imhof (1999). Optimal designs for rational models and weighted polynomial regression. Ann. Statist. 27, 1272-1293.

H. Dette, W.K. Wong (1999). E-optimal designs for the Michaelis Menten model. Statistics & Probability Letters 44, 405-408.

R.A. De Vore, G.G. Lorentz (1993). Constructive Approximation. Springer, N.Y.

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

G. Elfving (1952). Optimum allocation in linear regression. Ann. Math. Statist. 23, 255-262.

I. Ford, B. Torsney, C.F.J. Wu (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. J.R. Stat. Soc., Ser. B 54, 569-583.

I. Ford, S.D. Silvey (1980). A sequentially constructed design for estimating a nonlinear parametric function. Biometrika 67, 381-388.

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

B. Heiligers (1994). E-optimal designs in weighted polynomial regression. Ann. Statist. 22, 917-929.

L.A. Imhof, W.J. Studden (2001). E-optimal designs for rational models. Ann. Statist. 29, 763-783.

S. Karlin, W.J. Studden (1966). Tchebycheff Systems: with Applications in Analysis and Statistics. Interscience, New York.

V.B. Melas (1978). Optimal designs for exponential regression. Statistics 9, 45-59.

V.B. Melas (2001). Analytical properties of locally D-optimal designs for rational models. In: MODA 6 – Advances in Model-Oriented Design and Analysis (eds. A.C. Atkinson, P. Hackel, W.G. Müller). Physica Verlag, Heidelberg, 201-210.

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

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

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

F. Pukelsheim, W.J. Studden (1993). E-optimal designs for polynomial regression. Ann. Statist. 21, 402-415.

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

D.A. Ratkowski (1983). Nonlinear Regression. Dekker, New York.

G.A.J. Seber, C.J. Wild (1989). Nonlinear Regression. Wiley, New York.

S.D. Silvey (1980). Optimum design. Chapman and Hall, London.

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

W.J. Studden, J.Y. Tsay (1976). Remez's procedure for finding optimal designs. Ann. Statist. 4, 1271-1279.

G. Szegö (1975). Orthogonal Polynomials. Amer. Math. Soc. Providence, RI.

C.F.J. Wu (1985). Efficient sequential designs with binary data. J. Amer. Stat. Assoc. 80, 974-984.