Optimal designs for rational regression models

Holger Dette, Christine Kiss
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
holger.dette@ruhr-uni-bochum.de
tina.kiss12@googlemail.com

January 25, 2012

Abstract

In this paper we consider locally optimal designs problems for rational regression models. In the case where the degrees of polynomials in the numerator and denominator differ by at most 1 we identify an invariance property of the optimal designs if the denominator polynomial is palindromic, which reduces the optimization problem by 50%. The results clarify and extend the particular structure of locally c-, D- and E optimal designs for inverse quadratic regression models which have recently been found by Haines (1992) and Dette and Kiss (2009). We also investigate the relation between the D-optimal designs for the Michaelis Menten and EMAX-model from a more general point of view.

AMS Subject Classification: 62K05
Keywords and Phrases: rational regression models, optimal designs, Chebyshev systems, palindromic polynomials

1 Introduction.

Rational functions have appealing approximation properties and are widely used in regression analysis. They define a flexible family of nonlinear regression models which can be used to describe the relationship between a response, say Y, and a univariate predictor, say d

[see Ratkowsky (1983), Ratkowsky (1990) among many others]. In contrast to ordinary polynomials rational regression models can be bounded. As a consequence, they can be used to describe saturation effects, in cases where it is known that the response does not exceed a finite amount. Similarly, a toxic effect can be produced, in situations where the response is decreasing and converges eventually to a constant. Two important examples are given by the Michaelis Menten and the EMAX model which are widely used is such important areas as medicine, economics, environment modeling, toxicology or engineering [see Johansen (1984), Cornish-Browden (1995) or Boroujerdi (2002) among many others]

Despite of their importance optimal designs for rational models have only recently been found. For the Michaelis Menten model optimal designs have been studied by Dunn (1988), Rasch (1990), Song and Wong (1998), López-Fidalgo and Wong (2002), Dette et al. (2003) and Dette and Biedermann (2003) among others. Similarly, optimal designs for the EMAX model have been determined by Merle and Mentre (1995), Wang (2006), Dette et al. (2008) and Dette et al. (2010). Cobby et al. (1986) determined local D-optimal designs numerically and Haines (1992) provided some analytical results for D-optimal designs in the inverse quadratic regression model. Recently Dette and Kiss (2009) extended these results and also derived D_1, E- and optimal extrapolation designs for this class of models. He et al. (1996), Dette et al. (1999), Imhof and Studden (2001) and Dette et al. (2004) investigated D-, E- and c-optimal designs for more general rational models.

In the present paper we will derive further results on the structure of optimal designs for rational regression models. In particular, we consider the case where the degrees of the polynomials in the numerator and denominator differ by one. Several structural results of locally optimal designs in these models are derived, which explain the specific structure found in the case of inverse quadratic regression models by Haines (1992) and Dette and Kiss (2009). More precisely, for a broad class of optimality criteria we prove an invariance property of locally optimal designs if the polynomial in the denominator is palindromic. This reduces the corresponding optimization problems by 50%. In particular we investigate under which circumstances the results found by Haines (1992) and Dette and Kiss (2009) can be transferred to other optimality criteria and more general rational regression models.

The remaining part of this paper is organized as follows. In Section 2 we introduce the class of rational models considered in this paper. Section 3 is devoted to a discussion of the number of support points of locally optimal designs in these models and to the D-optimal design problem. Finally, in Section 4 we consider the special case where the polynomial in the numerator is palindromic (for a precise definition see Section 4). The results of this paper demonstrate that the specific properties of the optimal designs found in Haines (1992) and Dette and Kiss (2009) have a deeper background, namely the palindromic structure of the polynomial in the denominator.

We finally point out that the designs considered in this paper are locally optimal in the sense of Chernoff (1953), because they require the specification of the unknown parameters. These designs are usually used as benchmarks for commonly proposed designs. Moreover, they are the basis for more sophisticated design strategies, which require less precise knowledge about the model parameters, such as sequential, Bayesian or standardized maximin optimality criteria [see Chaloner and Verdinelli (1995) and Dette (1997) among others].

2 Rational regression models.

We consider the common nonlinear regression model

$$\mathbb{E}(Y|d) = \eta(d,\theta) \ , \tag{2.1}$$

where the regression function η is either given by

$$\eta_1(d,\theta) = \frac{P(d,\theta)}{Q(d,\theta)} = \frac{\theta_1 d + \cdots + \theta_p d^p}{1 + \theta_{p+1} d + \cdots + \theta_{p+q} d^q} \tag{2.2}$$

or

$$\eta_2(d,\theta) = \frac{P_0(d,\theta)}{Q(d,\theta)} = \frac{\theta_0 + \theta_1 d + \cdots + \theta_p d^p}{1 + \theta_{p+1} d + \cdots + \theta_{p+q} d^q}. \tag{2.3}$$

We define $\theta = (\theta_0, \theta_1, \ldots, \theta_t)^T$ as the corresponding vector of parameters (where $t = p+q$ or $t = p+q+1$ corresponding to model η_1 or η_2, respectively) and assume that the explanatory variable d varies in the design space $\mathcal{D}$, which is either given by a compact interval $\mathcal{D} = [d_\ell, d_u]$, where $d_\ell \geq 0$, or by interval $\mathcal{D} = [0, \infty)$. If $p \geq q$ we always assume that $\mathcal{D} = [d_\ell, d_u]$ (it is easy to see that otherwise the design problem is not well defined). Additionally, we assume that $Q(d,\theta) \neq 0$ for all $d \in \mathcal{D}$. Note that in the case $p = q = 1$ the models (2.2) and (2.3) give the Michaelis Menten and the EMAX model, respectively, which are widely used is such important areas as medicine, economics, environment modeling, toxicology or engineering [see Johansen (1984), Cornish-Browden (1995) or Boroujerdi (2002) among many others]. On the other hand the choice $p = 1$ and $q = 2$ yields inverse quadratic regression models as discussed in Haines (1992) and Dette and Kiss (2009).

We assume that for each experimental condition d an observation Y is available according to the model (2.1), where different observations are realizations of independent and normally distributed random variables with variance $\sigma^2 > 0$. We consider approximate designs in the sense of Kiefer (1974), which are defined as probability measures on the design space $\mathcal{D}$ with finite support. The support points of an (approximate) design ξ define the locations where observations are taken, while the weights give the corresponding relative proportions of total observations to be taken at these points. If the design ξ has masses $w_i > 0$ at the different points d_i $(i = 1, \ldots, t)$ and n observations can be made, the quantities $w_i n$ are rounded to integers, say n_i, satisfying $\sum_{i=1}^t n_i = n$, and the experimenter takes n_i observations at each location d_i $(i = 1, \ldots, t)$. In this case (under regularity conditions) the asymptotic covariance matrix of the maximum likelihood estimator is given by the matrix $\frac{\sigma^2}{n} M^{-1}(\xi, \theta)$, where

$$M(\xi, \theta) = \int_{\mathcal{D}} f(d, \theta) f^T(d, \theta) d\xi(\theta)$$

denotes the information matrix of the design ξ and

$$f(d, \theta) = \frac{\partial}{\partial \theta} \eta(d, \theta)$$

is the gradient of the regression function in model (2.1) with respect to the parameter θ. For $n \in \mathbb{N}$ we introduce the notation

$$h_n(d) \quad = \quad (1, d, \ldots, d^n)^T,$$

then a similar calculation as in He et al. (1996) shows that for the models (2.2) and (2.3) the information matrix has the representation

$$M_i(\xi, \theta) = B_i(\theta) \bar{M}_i(\xi, \theta) B_i^T(\theta), \quad i = 1, 2, \tag{2.4}$$

where the matrix $\bar{M}_i(\xi, \theta)$ is given by

$$\bar{M}_i(\xi, \theta) = \int_{\mathcal{D}} \frac{d^{2(2-i)}}{Q^4(d, \theta)} h_{p+q-2+i}(d) h^T_{p+q-2+i}(d) d\xi(d) \in \mathbb{R}^{p+q-1+i \times p+q-1+i}, \quad i = 1, 2. \tag{2.5}$$

In the representation (2.4) the symbols $B_i(\theta)$ denote square matrices of appropriate dimension with rows given by

$$b_{1,i} = \begin{cases} (\underbrace{0, 0, \ldots, 0}_{i-1}, 1, \theta_{p+1}, \theta_{p+2} \ldots, \theta_{p+q}, \underbrace{0, 0, \ldots, 0}_{p-i}), & \text{if } 1 \leq i \leq p, \\ -(\underbrace{0, 0, \ldots, 0}_{i-p}, \theta_1, \ldots, \theta_p, \underbrace{0, 0, \ldots, 0}_{p+q-i}), & \text{if } p+1 \leq i \leq p+q. \end{cases}$$

for model (2.2), and by

$$b_{2,i} = \begin{cases} (\underbrace{0, 0, \ldots, 0}_{i}, 1, \theta_{p+1}, \theta_{p+2} \ldots, \theta_{p+q}, \underbrace{0, 0, \ldots, 0}_{p-i}), & \text{if } 0 \leq i \leq p, \\ -(\underbrace{0, 0, \ldots, 0}_{i-p}, \theta_0, \theta_1, \ldots, \theta_p, \underbrace{0, 0, \ldots, 0}_{p+q-i}), & \text{if } p+1 \leq i \leq p+q. \end{cases}$$

for model (2.3), respectively. An optimal (approximate) design maximizes an appropriate concave functional, say Φ, of the information matrix $M_i(\xi, \theta)$ which is proportional to the asymptotic covariance matrix, and there are numerous criteria which can be used for discriminating between competing designs [see Silvey (1980), Pukelsheim (1993) among others].

Note that resulting designs are locally optimal in the sense of Chernoff (1953), because they require the specification of the unknown parameters. There are many situations where such preliminary knowledge is available, such that the application of locally optimal designs is well justified [a typical example are phase II dose finding trials, see Dette et al. (2008)]. However, the most important application of locally optimal designs is their use as benchmarks for commonly proposed designs. Moreover, they are the basis for more sophisticated design strategies, which require less precise knowledge about the model parameters, such as sequential, Bayesian or standardized maximin optimality criteria [see Chaloner and Verdinelli (1995) and Dette (1997) among others].

3 Number of support points and locally D-optimal designs

Throughout this paper we assume that the function Φ is an information function in the sense of Pukelsheim (1993) and are interested in a most precise estimation of $K^T\theta$ where $K \in \mathbb{R}^{s\times t}$ is a given matrix of rank s and $t = p+q$ or $t = p+q+1$ corresponding to models (2.2) and (2.3), respectively. Throughout this section the matrix $M(\xi,\theta)$ is either $M_1(\xi,\theta)$ or $M_2(\xi,\theta)$ corresponding to model (2.2) and (2.3), respectively. A locally Φ-optimal design ξ^* for estimating $K^T\theta$ maximizes the function

$$\Phi(C_K(M(\xi,\theta))) \tag{3.1}$$

in the class of all models for which $K^T\theta$ is estimable, that is range$(K) \in$ range$(M(\xi,\theta))$. In (3.1) the matrix C_K is defined by

$$C_K(M(\xi,\theta)) = (K^T M^-(\xi,\theta)K)^{-1}$$

and A^- denotes a generalized inverse of the matrix A. Our first result refers to the number of support points of optimal designs in the rational regression model (2.3) and requires some concepts of classical approximation theory. Following Karlin and Studden (1966) a set of functions $\{g_0,\dots,g_k\}$ defined on an interval $\mathcal{D}$ is called Chebychev-system, if every linear combination $\sum_{i=0}^k a_i g_i(d)$ with $\sum_{i=0}^k a_i^2 > 0$ has at most k distinct roots on $\mathcal{D}$. This property is equivalent to the fact that

$$\det(g(d_0),\dots,g(d_k)) \neq 0$$

holds for all $d_0, \dots, d_k \in \mathcal{D}$ with $d_i \neq d_j$ $(i \neq j)$, where $g(d) = (g_0(d), \dots, g_k(d))^T$ denotes the vector of all functions [see Karlin and Studden (1966)]. If the functions $g_0, \dots, g_k$ constitute a Chebyshev-system on the set $\mathcal{D}$, then there exists a unique "polynomial"

$$\phi^*(d) := \sum_{i=0}^{k} \alpha_i^* g_i(d) \quad (\alpha_0^*, \dots, \alpha_k^* \in \mathbb{R}) \tag{3.2}$$

with the following properties

(i) $\quad | \phi^*(d) | \leq 1 \quad \forall d \in \mathcal{D}$

(ii) $\quad$ There exist $k+1$ points $s_0 < \cdots < s_k$ such that $\phi^*(s_i) = (-1)^{k-i}$ for $i = 0, \dots, k$.

The function $\phi^*(d)$ is called the Chebychev-polynomial, and the points $s_0, \dots, s_k$ are called Chebychev-points, which are not necessarily unique. They are unique if the constant function is an element of $\text{span}\{g_0, \dots, g_k\}$.

Theorem 3.1 *Assume that the polynomial Q in model* (2.2) *and* (2.3) *satisfies one of the following conditions*

(a) The function $\frac{\partial^{2(p+q)+1}}{\partial d^{2(p+q)+1}} Q^4(d,\theta)$ has no roots in the interior of the interval $\mathcal{D}$.

(b) $p \geq q$.

(c) In model (2.2) *the functions $\{d^2, \dots, d^{2(p+q)}, Q^4(d,\theta)\}$ form a Chebyshev system on the interval $\mathcal{D} \setminus \{0\}$.*

In model (2.3) *the functions $\{1, d, \dots, d^{2(p+q)}, Q^4(d,\theta)\}$ form a Chebyshev system on the interval $\mathcal{D}$.*

In model (2.2) *any locally Φ-optimal design for estimating $K^T\theta$ is supported at at most $p+q$ points. In model* (2.3) *any locally Φ-optimal design for estimating $K^T\theta$ is supported at at most $p+q+1$ points. Moreover, if all roots of the polynomial $Q(d,\theta)$ are smaller than d_ℓ and a locally Φ-optimal design is supported at $p+q+1$ points then its support contains the boundary point d_ℓ of the design space.*

Proof. It follows from Pukelsheim (1993) that a design ξ^* is locally Φ-optimal for estimating $K^T\theta$ if an only if there exists a generalized inverse G of the matrix $M(\xi^*, \theta)$ and a matrix D satisfying

$$\Phi(C_K M((\xi^*, \theta)))\Phi^\infty(D) = \text{trace } (C_K(M(\xi^*, \theta))) = 1,$$

such that the inequality

$$f_i^T(d,\theta) G K C_K(M_i(\xi^*, \theta)) D C_K(M_i(\xi^*, \theta)) K^T G^T f_i(d,\theta) \leq 1 \tag{3.3}$$

is satisfied for all $d \in \mathcal{D}$, where Φ^∞ denotes the polar function of Φ and the vector f_i is defined by

$$f_i(d,\theta) = B_i(\theta) d^{2-i} \frac{h_{p+q-2+i}(d)}{Q^2(d,\theta)}; \quad i = 1,2,$$

corresponding to model (2.2) and (2.3), respectively. For model (2.3) corresponding to the case $i = 2$ it now follows by the same arguments as given in the proof of Theorem 3.6 in Chapter X of Karlin and Studden (1966) that any Φ-optimal design has at most $p+q+1$ support points [note that these authors consider the D-optimality criterion, but the proof does not change if the checking condition is of the form (3.3)]. For the second part assume that the locally Φ-optimal design ξ^* is supported at $p+q+1$ points but does not have the boundary point d_ℓ as support point. From Theorem 4.2 in Dette and Melas (2011) it follows that there exists a design $\tilde{\xi}$ with $p+q+1$ support points including the point d_ℓ, such that

$$M_2(\tilde{\xi},\theta) \geq M_2(\xi^*,\theta).$$

By the concavity of the information function Φ we have for the design $\bar{\xi} = \frac{1}{2}(\xi^* + \tilde{\xi})$

$$\Phi(C_K(M_2(\bar{\xi},\theta))) \geq \Phi(C_K(M_2(\xi^*,\theta))).$$

This means that $\bar{\xi}$ is also Φ-optimal but has at least $p+q+2$ support contradicting the first part of the proof. The results for model (2.2) are derived similarly and the arguments are omitted for the sake of brevity. □

For the D-optimality criterion $\Phi(C_K(M(\xi,\theta))) = |C_K(M(\xi,\theta))|^{1/s}$ there exists an interesting connection between locally D-optimal designs which minimize the criterion in the class of all $(p+q)$-point designs or $(p+q+1)$-point designs for model (2.2) and (2.3), respectively. Note that a standard result in optimal design theory [see Silvey (1980) for example] shows that these designs are equally weighted.

Theorem 3.2 *If the assumptions of Theorem 3.1 are satisfied and*

$$\xi_D^* = \begin{pmatrix} d_1 & \dots & d_{p+q} \\ \frac{1}{p+q} & \dots & \frac{1}{p+q} \end{pmatrix}$$

is the locally D-optimal design for model (2.2)*, then the design*

$$\xi_D^* = \begin{pmatrix} 0 & d_1 & \dots & d_{p+q} \\ \frac{1}{p+q+1} & \frac{1}{p+q+1} & \dots & \frac{1}{p+q+1} \end{pmatrix}$$

is a locally D-optimal for model (2.3)*.*

Proof. We obtain from Theorem 3.1 that locally D-optimal designs for the rational regression models (2.2) and (2.3) have $p+q$ and $p+q+1$ support points, respectively, and a standard argument shows that the optimal designs have equal masses at their support points. Observing the representation (2.4) it follows that a locally D-optimal saturated design for model (2.3) maximizes the determinant of the matrix

$$\bar{M}_2(\xi,\theta) = X_2^T(\xi,\theta)\mathrm{diag}\big(\tfrac{1}{p+q+1},\dots,\tfrac{1}{p+q+1}\big)X_2(\xi,\theta),$$

where the matrix $X_2(\xi,\theta)$ is defined by

$$X_2(\xi,\theta) = \big(\tfrac{1}{Q^2(d_i,\theta)}d_i{}^j\big)_{i,j=0}^{p+q}$$

and $\mathrm{diag}(a_1,\dots,a_k)$ denotes a diagonal matrix with diagonal elements $a_1,\dots,a_k$. A straightforward calculation yields for this determinant

$$\begin{aligned}\det(\bar{M}_2(\xi,\theta)) &= \frac{1}{(p+q+1)^{p+q+1}}\det(X_2(\xi,\theta))^2 \\ &= \frac{1}{(p+q+1)^{p+q+1}}\prod_{i=0}^{p+q}\frac{1}{Q^4(d_i,\theta)}\prod_{0\le i<j\le p+q}(d_i-d_j)^2.\end{aligned}$$

By Theorem 3.1 the smallest support point of the locally D-optimal design is given by $d_0=0$ and the corresponding determinant reduces to

$$\det(\bar{M}_2(\xi,\theta)) = \frac{1}{(p+q+1)^{p+q+1}}\frac{1}{Q^4(0,\theta)}\prod_{i=1}^{p+q}\frac{d_i^2}{Q^4(d_i,\theta)}\prod_{1\le i<j\le p+q}(d_i-d_j)^2$$

which has to be maximized with respect to the choice of $d_1\dots d_{p+q}$. Now a similar calculation shows that the same optimization problem arises in the maximization of the determinant of the information matrix in model (2.2) in the class of all $(p+q)$-point designs and the assertion of Theorem 3.2 follows. □

Example 3.1 Consider the case $p=q=1$ and $\mathcal{D}=[0,1]$, where (2.2) and (2.3) correspond to the Michaelis-Menten and EMAX model, respectively. The locally D-optimal design has been determined in Rasch (1990) and puts equal masses at the two points $\frac{1}{2+\theta_2}$ and 1. The corresponding EMAX model is given by

$$\eta_2(d,\eta) = \frac{\theta_0+\theta_1 d}{1+\theta_2 d}$$

for which the locally D-optimal design has not been stated explicitly in the literature. By Theorem 3.1 and 3.2 this design has 3 support points and puts equal masses at the points $0, \frac{1}{2+\theta_2}$ and 1.

4 Palindromic polynomials

In this section we investigate the case $q = p + 1$ in model (2.2) and $q = p$ in model (2.3) in more detail in the case where the polynomial in the denominator is palindromic, which means that the coefficients of the polynomial in the denominator

$$Q_\ell(d, \theta) = 1 + \theta_{p+1} d + \cdots + \theta_{2p+\ell} d^{p+\ell} \tag{4.1}$$

satisfy $\theta_{2p+\ell} = 1$ and $\theta_{2p-1+\ell-i} = \theta_{p+1+i}$ $(i = 0, \ldots, \lfloor \frac{p-1+\ell}{2} \rfloor)$, where the choice $\ell = 1$ and $\ell = 0$ correspond to model (2.2) and (2.3), respectively. It is easy that this condition is equivalent to the equation

$$d^{2(p+\ell)} Q_\ell^2(\frac{1}{d}, \theta) = Q_\ell^2(d, \theta) \qquad (\ell = 0, 1). \tag{4.2}$$

4.1 Locally c-optimal designs for model (2.2)

We begin with an investigation of c-optimal designs, which maximize the function

$$(c^T M_1^-(\xi, \theta) c)^{-1}$$

for a given vector $c \in \mathbb{R}^{2p+1}$ in the class of all designs satisfying range$(c) \in$ range$(M(\xi, \theta))$ (note that we have $s = 1, K = c$ in the general optimality criterion). By the discussion in Section 2 the gradient of the regression function η_1 in model (2.3) is given by

$$f_1(d, \theta) = B_1(\theta) \frac{d}{Q^2(d, \theta)} \big(1, d, \ldots, d^{2p}\big)^T = B_1(\theta) \frac{d}{Q^2(d, \theta)} h_{2p}(d), \tag{4.3}$$

where $Q(d, \theta)$ is defined in (4.1) and satisfies (4.2). Therefore a straightforward calculation shows that

$$\begin{aligned} f_1(\frac{1}{d}, \theta) &= B_1(\theta) \frac{d}{Q^2(d, \theta)} \big(d^{2p}, \ldots, d, 1\big)^T \\ &= B_1(\theta) D \frac{d}{Q^2(d, \theta)} h_{2p}(d) = B_1(\theta) D B_1^{-1}(\theta) f_1(d, \theta) = \tilde{D} f_1(d, \theta), \end{aligned}$$

where the matrices D and $\tilde{D}$ are given by

$$D = \begin{pmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{pmatrix}$$

(all other entries in the matrix D are 0) and

$$\tilde{D}_1 = B_1(\theta) D B_1^{-1}(\theta), \tag{4.4}$$

respectively. Obviously the functions

$$g_1(d) = \frac{d}{Q^2(d,\theta)}, \ \ldots, \ g_{2p+1}(d) = \frac{d^{2p+1}}{Q^2(d,\theta)} \tag{4.5}$$

form a Chebyshev system on the interval $\mathcal{D}\setminus\{0\}$. This means that for all $\alpha_1,\ldots,\alpha_{2p+1}$ with $\sum_{j=1}^{2p+1}\alpha_j^2 > 0$ the function

$$\sum_{j=1}^{2p+1} \alpha_j g_j(d)$$

has at most $2p$ roots, where roots in the interior of $\mathcal{D}\setminus\{0\}$, where no sign changes are counted twice [see Karlin and Studden (1966), p. 23].

Define the set $\mathcal{A}^* \subset \mathbb{R}^{2p+1}$ as the set of all vectors $\hat{c} = (\hat{c}_1,\ldots,\hat{c}_{2p+1})^T \in \mathbb{R}^{2p+1}$ such that the condition

$$\begin{vmatrix} g_1(d_1) & \ldots & g_1(d_{2p}) & \hat{c}_1 \\ g_2(d_1) & \ldots & g_2(d_{2p}) & \hat{c}_2 \\ \vdots & & \vdots & \vdots \\ g_{2p+1}(d_1) & \ldots & g_{2p+1}(d_{2p}) & \hat{c}_{2p+1} \end{vmatrix} \neq 0 \tag{4.6}$$

is satisfied for all $d_1,\ldots,d_{2p} \in \mathcal{D}\setminus\{0\}$ (with $d_i \neq d_j$). If $d_1^*,\ldots,d_k^*$ are Chebyshev points (see the discussion in Section 3) and ϕ^* is the corresponding Chebyshev polynomial then the inequality $(\phi^*)^2(d) \leq 1$ for all $d \in \mathcal{D}$ is equivalent to the inequality

$$\Delta(d) = d^2 S_{4p}(d) - \tilde{S}_{4p+4}(d) \leq 0$$

for all $d \in \mathcal{D}$, where S_{4p} and $\tilde{S}_{4p+4}$ are polynomials of degree $4p$ and $4p+4$ with positive leading coefficients, respectively. Because $\Delta(d_i^*) = 0 \quad (i = 1,\ldots,k)$ a careful counting argument shows that $k = 2p+1$ and consequently the Chebyshev points $d_1^*,\ldots,d_{2p+1}^*$ are uniquely determined. If $\hat{c} \in \mathcal{A}^*$, define the weights

$$w_i^* = \frac{|v_i|}{\sum_{j=0}^{2p+1}|v_j|} \qquad i = 1,\ldots,2p+1, \tag{4.7}$$

where the vector v is given by

$$v = (X^TX)^{-1}X^T\hat{c},$$

and $X = \big(g_i(d_j^*)\big)_{i,j=1}^{2p+1}$. It now follows from Theorem 7.7 in Chapter X of Karlin and Studden (1966) that the design, which puts masses $w_1^*,\ldots w_{2p+1}^*$ at the points $d_1^*,\ldots d_{2p+1}^*$ minimizes $\hat{c}^T\bar{M}_1^-(\xi,\theta)\hat{c}$.

THEOREM 4.1 *Consider the model (2.2) with $\mathcal{D} = (0, \infty)$ with $q = p + 1$ and polynomial Q satisfying (4.2). If $\hat{c} = B_1^{-1}(\theta)c \in \mathcal{A}^*$ then the c-optimal design ξ_c^* in model (2.2) is uniquely determined and has $2p + 1$ points $d_1^* < d_2^* < \cdots < d_{2p+1}^*$ satisfying*

$$d_{2p+2-j}^* = \frac{1}{d_j^*} \quad j = 1, \ldots, p + 1. \tag{4.8}$$

In particular $d_{p+1}^* = 1$ and the weights are given by (4.7).

Proof. A c-optimal design minimizes the criterion

$$c^T M_1^{-1}(\xi, \theta)c = c^T B_1^T(\theta)^{-1} \bar{M}_1^{-1}(\xi, \theta) B_1(\theta)^{-1} c = \hat{c}^T \bar{M}_1^{-1}(\xi, \theta)\hat{c},$$

where the matrix $\bar{M}_1$ and the vector $\hat{c}$ are defined by (2.5) and $\hat{c} = B^{-1}(\theta)c$, respectively. Consider the Chebyshev polynomial

$$\phi^*(d) = \sum_{i=1}^{2p+1} \alpha_i^* g_i(d, \theta)$$

defined in (3.2). By the discussion in Section 3 and in the previous paragraph there exist exactly $2p+1$ points $0 < d_1^* < \ldots < d_{2p+1}^*$ in $\mathcal{D}$ such that the values $\phi^*(d_i)$ alternate in sign and $|\phi^*(d_i)| = \sup_{d \in \mathcal{D}} |\phi^*(d)|$; $\quad i = 1, \ldots, 2p + 1$.

Because $\hat{c} = B_1^{-1}(\theta)c \in \mathcal{A}^*$ the discussion in the previous paragraph also shows that the design ξ_c^* with masses w_i^* defined in (4.7) at the points d_i^* $(i = 1, \ldots, 2p + 1)$ is the unique design which minimizes $\hat{c}^T \bar{M}_1^{-1}(\xi, \theta)\hat{c}$, that is the design ξ_c^* is the unique c-optimal design for the model (2.2). We show at the end of the proof that the Chebyshev polynomial satisfies

$$\phi^*(d) = \phi^*(1/d), \tag{4.9}$$

then it follows that for each Chebyshev point d_i^* the point $1/d_i^*$ is also a Chebyshev point.

Now consider the $p + 1$ smallest Chebyshev points $0 < d_1^* < \cdots < d_p^* < d_{p+1}^*$ and note that $\phi^*(d_i^*) = (-1)^i \varepsilon \sup\{|\phi^*(d)| \big| d \in \mathcal{D}\}$ $\quad (i = 1, \ldots, p + 1)$ for some $\varepsilon \in \{-1, 1\}$ and $d_p^* < 1$. Define $\tilde{d}_j = d_j^*$ $(j = 1, \ldots, p + 1)$ and $\tilde{d}_{j+p+1} = 1/d_{2p+2-j}^*$ $\quad (j = 1, \ldots, p)$. Then the points $\tilde{d}_1, \ldots, \tilde{d}_{2p+1}$ obviously satisfy (4.8) and are also the Chebyshev points, by (4.9). Consequently, they must coincide with $d_1^*, \ldots, d_{2p+1}^*$. By the discussion in the previous paragraph the c-optimal design is supported at these points. Moreover, $d_{p+1}^* = 1$, because otherwise there would exist $2p + 2$ Chebyshev points.

The proof will now be completed by showing the property (4.9). For this purpose we consider the problem of approximating the function g_{2p+1} by a linear combination of the functions $g_1, \ldots, g_{2p}$ in (4.5) with respect to the sup-norm, that is

$$m^* = \min_{\alpha_1, \ldots, \alpha_{2p}} \sup_{d \in \mathcal{D}} \Big| g_{2p+1}(d) - \sum_{j=1}^{2p} \alpha_j g_j(d) \Big| \tag{4.10}$$

by Theorem 1.1 in Karlin and Studden (1966), Chapter IX, it follows that there exist $2p+1$ points $d_1^* < \cdots < d_{2p+1}^*$ in $\mathcal{D}$, such that the solution $\psi^*(d) = g_{2p+1}(d) - \sum_{j=1}^{2p} \alpha_j^* g_j(d)$ of (4.10) satisfies $\sup_{d\in\mathcal{D}} |\psi^*(d)| = m^*$ and

$$m^*(-1)^i \varepsilon = \psi^*(d_i) \quad i = 1, \ldots, 2p+1$$

for some $\varepsilon > 0$. Therefore ψ^* must be proportional to the Chebyshev polynomial, that is $\phi^* = \varepsilon\psi^*/m^*$ for some $\varepsilon \in \{-1, 1\}$, where $1/m^*$ is the coefficient of g_{2p+1} in ϕ^*. On the other hand, $m^*\phi^*(1/d)$ is the unique solution of the minimax problem

$$m^* \;=\; \min_{\beta_1,\ldots,\beta_{2p}} \sup_{d\in\mathcal{D}} \Big| g_{2p+1}(\frac{1}{d}) - \sum_{j=1}^{2p} \beta_j g_j(\frac{1}{d}) \Big| = \min_{\beta_2,\ldots,\beta_{2p+1}} \sup_{d\in\mathcal{D}} \Big| g_1(d) - \sum_{j=2}^{2p+1} \beta_j g_j(d) \Big|,$$

where we have used the fact that

$$g_j(\frac{1}{d}) = \frac{d^{-j}}{Q^2(\frac{1}{d}, \theta)} = \frac{d^{2p+2-j}}{Q^2(d, \theta)} = g_{2p+2-j}(d) \quad j = 1, \ldots, 2p+1,$$

which follows from (4.2). By the same argument as in the previous paragraph it follows that the coefficient of g_1 in ϕ^* is $1/m^*$, which means that the coefficients of g_1 and g_{2p+1} in the representation of ϕ^* coincide. Repeating these arguments for the other coefficients of ϕ^* we obtain

$$\phi^*(\frac{1}{d}) = \phi^*(d),$$

which completes the proof of Theorem 4.1. □

Example 4.1 Dette et al. (2010) considered locally optimal design problems for the inverse quadratic regression model

$$\frac{u}{\kappa_0 + \kappa_1 u + \kappa_2 u^2}; \quad u \in (0, \infty). \tag{4.11}$$

If the explanatory variable u is scaled by the transformation $d = d(u) = \sqrt{\kappa_2/\kappa_0}\,u$ it is easy to see that locally $\hat{c}$-optimal designs in model (4.11) can be obtained from the locally c-optimal designs for the model (2.3) with $p = 1$ and $q = 2$ with

$$c = B_1(\theta) E_1^{-1}(\theta) \hat{c}$$

by transforming the support points via $d \to \sqrt{\kappa_0/\kappa_2}\; d$ and leaving the weights unchanged. Here the matrices $B_1(\theta)$ and $E_1(\theta)$ are defined by

$$B_1(\theta) = \begin{pmatrix} 1 & \theta_2 & \theta_3 \\ 0 & -\theta_1 & 0 \\ 0 & 0 & -\theta_1 \end{pmatrix}; \quad E_1^{-1}(\theta) = \mathrm{diag}(\kappa_0^{1/2}\kappa_2^{1/2}, \kappa_2, \kappa_0^{-1/2}\kappa_2^{3/2})$$

and $\theta_3 = 1,\ \theta_2 = \kappa_1(\kappa_0\kappa_2)^{-1/2}$ (note that it is not necessary to specify θ_1). Consequently, the support of a $\hat{c}$-optimal design is of the form $\{1/\rho, 1, \rho\}$ if $E^{-1}(\theta)\hat{c} \in \mathcal{A}^*$, which implies that in the original parametrization (4.11) the support of the $\hat{c}$-optimal design is of the form

$$\left\{\frac{1}{\rho}\sqrt{\frac{\kappa_0}{\kappa_2}}, \sqrt{\frac{\kappa_0}{\kappa_2}}, \rho\sqrt{\frac{\kappa_0}{\kappa_2}}\right\} \tag{4.12}$$

if $E^{-1}(\theta)\hat{c} \in \mathcal{A}^*$. Dette and Kiss (2009) considered the case $\hat{c} = (0,0,1)^T$ for which $E^{-1}(\theta)\hat{c}$ is of the form $(0, 0, \kappa_0^{-1/2}\kappa_2^{3/2})^T$ and obviously an element of set $\mathcal{A}^*$. Similarly, if extrapolation at a point u_0 in model (4.11) with design space $\mathcal{U} = (0, d_u]$ is of interest, then

$$E^{-1}(\theta)\hat{c} = \kappa_0^{1/2}\kappa_2^{1/2}\frac{u_0}{(\kappa_0 + \kappa_1 u_0 + \kappa_2 u_0^2)^2}\begin{pmatrix} 1 \\ \kappa_0^{-1/2}\kappa_2^{1/2}u_0 \\ \kappa_0^{-1}\kappa_2 u_0^2 \end{pmatrix},$$

and it is easy to see that this vector satisfies $E^{-1}(\theta)\hat{c} \in \mathcal{A}^*$. Consequently, the support of the optimal extrapolation design in model (4.11) on the interval $(0, d_u]$ is of the form (4.12) if $t \geq \rho\sqrt{\kappa_0/\kappa_2}$. Therefore the first parts of Theorem 3.1 and 3.3 in Dette and Kiss (2009) are consequences of Theorem 4.1 of this paper.

4.2 Locally Φ_ℓ-optimal designs for model (2.2)

We now consider general Φ_ℓ-optimal designs for estimating $K^T\theta$ in model (2.2) with $q = p+1$ and polynomial Q satisfying (4.2). The following result identifies a condition, which implies that the support of the locally Φ_ℓ-optimal design satisfies a similar invariance property as specified in (4.8).

THEOREM 4.2 *For $\ell \in [-\infty, 1]$ let $\Phi_\ell(C_K(M(\xi,\theta))) = (\frac{1}{s}\mathrm{tr}(\mathrm{C}_\mathrm{K}^\ell(\mathrm{M}(\xi,\theta))))^{1/\ell}$ denote the Φ_ℓ-optimality criterion and assume that*

$$\tilde{D}_1 K A^{-1} = K \tag{4.13}$$

for some orthogonal matrix $A \in \mathbb{R}^{s\times s}$ where the matrix $\tilde{D}_1$ is defined in (4.4). Then there exists a number $t^* \in \{1, \ldots, 2p+1\}$ and a locally Φ_ℓ-optimal design ξ^* for estimating $K^T\theta$ in model (2.2) with $\mathcal{D} = (0, \infty)$ and $q = p+1$ has masses $w_1^*, \ldots, w_{t^*}^*$ at points $d_1^* < \ldots, d_{t^*}^*$, which satisfy

$$d^*_{t^*+1-j} = \frac{1}{d^*_j} \quad j = 1, \ldots, t^* \tag{4.14}$$

$$w^*_{t^*+1-j} = w^*_j \quad j = 1, \ldots, t^* \tag{4.15}$$

In particular, if $s = 2p+1$ and the matrix K is non-singular, then condition (4.13) *can be rewritten as*

$$(KK^T)^{-1}\tilde{D}_1^{-1}KK^T = \tilde{D}_1, \tag{4.16}$$

$t^* = 2p+1$*, the locally Φ_ℓ-optimal design is uniquely determined and* $d^*_{p+1} = 1$.

Proof. Let

$$\xi = \begin{pmatrix} d_1, \dots, d_t \\ w_1, \dots, w_t \end{pmatrix}$$

denote a locally Φ_ℓ-optimal design for estimating $K^T\theta$ in model (2.2) and define

$$\bar{\xi} = \begin{pmatrix} \frac{1}{d_1}, \dots, \frac{1}{d_t} \\ w_1, \dots, w_t \end{pmatrix}.$$

with $t \in \{1, \dots, 2p+1\}$. Observing (4.3) and the definition of the matrix $\tilde{D}_1$ in (4.4) it follows by a straightforward calculation that

$$\begin{aligned} M_1(\bar{\xi}, \theta) &= \sum_{i=1}^t f_1(\frac{1}{d_i}, \theta) f_1^T(\frac{1}{d_i}, \theta) w_i \\ &= \tilde{D}_1 \sum_{i=1}^t f_1(d_i, \theta) f_1^T(d_i, \theta) w_i \tilde{D}_1^T = \tilde{D}_1 M_1(\xi, \theta) \tilde{D}_1^T. \end{aligned}$$

¿¿From the assumption $\tilde{D}KA^{-1} = K$ we have (Pukelsheim (1993), Chapter 3.2 and 3.21)

$$\begin{aligned} C_K(M(\bar{\xi}_1, \theta)) &= \min\left\{ LM_1(\bar{\xi}, \theta)L^T \middle| LK = I_s,\ L \in \mathbb{R}^{s\times(2p+1)} \right\} \\ &= A^{-1} \min\left\{ ALM_1(\bar{\xi}, \theta)L^T A^T \middle| ALKA^{-1} = I_s,\ L \in \mathbb{R}^{s\times(2p+1)} \right\} A \\ &= A^{-1} \min\left\{ AL\tilde{D}_1 M_1(\xi, \theta)\tilde{D}_1^T L^T A^T \middle| AL\tilde{D}_1\tilde{D}_1 KA^{-1} = I_s,\ L \in \mathbb{R}^{s\times(2p+1)} \right\} A \\ &= A^{-1} \min\left\{ \tilde{L}M_1(\xi, \theta)\tilde{L}^T \middle| \tilde{L}K = I_s,\ L \in \mathbb{R}^{s\times(2p+1)} \right\} A \\ &= A^{-1} C_K(M_1(\xi, \theta))A, \end{aligned}$$

and the orthogonality of the matrix A shows that the matrices $C_K(M_1(\xi, \theta))$ and $C_K(M_1(\bar{\xi}, \theta))$ have the same eigenvalues, which implies $\Phi_\ell(C_K(M_1(\xi, \theta))) = \Phi_\ell(C_K(M_1(\bar{\xi}, \theta)))$. If $1 \in \text{supp}(\xi)$, define $t^* = 2t-1$, otherwise define $t^* = 2t$ and consider the design

$$\xi^* = \frac{\xi + \bar{\xi}}{2} = \begin{pmatrix} d_1^*, \dots, d_{t^*}^* \\ w_1^*, \dots, w_{t^*}^* \end{pmatrix}$$

with support points $d_1^* < \dots < d_{t^*}^*$ and corresponding weights $w_1^*, \dots, w_{t^*}^*$. Then it is easy to see that the support points and weights of ξ^* satisfy (4.14) and (4.15), respectively.

Moreover, by the concavity of the Φ_ℓ-optimality criterion and the mapping $M \to C_K(M)$ [see Pukelsheim (1993), Chapter 3.13] we have

$$\Phi_\ell(C_K(M_1(\xi^*,\theta))) \geq \frac{1}{2}\Big(\Phi_\ell(C_K(M_1(\xi,\theta))) + \Phi_\ell(C_K(M_1(\bar{\xi},\theta)))\Big) = \Phi_\ell\big(C_K(M_1(\xi,\theta))\big)$$

which shows that the design ξ^* is also Φ_ℓ-optimal for estimating $K^T\theta$ in model (2.2) and proves the first part of Theorem 4.2.

For a proof of the second part note that it follows from Theorem 3.1 that for a nonsingular matrix $K \in \mathbb{R}^{2p+1\times 2p+1}$ the Φ_ℓ-optimal design is supported at exactly $2p+1$ points and that the condition (4.16) is a direct consequence of (4.13) in this case. □

For the D-optimality criterion a stronger version of Theorem 4.2 is available, which follows directly from its proof.

Corollary 4.1 *If*

$$\tilde{D}_1 K A^{-1} = K$$

for some nonsingular matrix $A \in \mathbb{R}^{s\times s}$, *then there exists a locally* D*-optimal design for estimating* $K^T\theta$ *in model* (2.2) *with* $q = p+1$ *and* $\mathcal{D} = (0,\infty)$ *with masses* $w_1^*,\ldots,w_{t^*}^*$ *at the points* $d_1^* < \ldots < d_{t^*}^*$ *with* $t^* \in \{1,\ldots,2p+1\}$, *which satisfies* (4.14) *and* (4.15).

4.3 Locally optimal designs for model (2.3)

In this section we briefly discuss similar results for the model (2.3) with $p = q$ and palindromic polynomial in the denominator. In this case the design space is bounded (otherwise the design problems are not well defined) and it follows from the discussion in Section 2 that the gradient of the expected response with respect to the parameter θ is given by

$$f_2(d,\theta) = B_2(\theta)\frac{h_{2p}(d)}{Q^2(d,\theta)}$$

where Q is a polynomial of a degree p. It follows from Theorem 3.1 that any Φ-optimal design is supported at at most $p+q+1$ points. Moreover, a similar argument as used in the proof of this result (using the equivalence theorem in (3.3)) shows that if the locally Φ-optimal design is supported at exactly $p+q+1$ points, then the support includes both boundary points of the design space (note that $p = q$).

Now assume that Q is palindromic, then the coefficients in the polynomial $Q(d,\theta) = \sum_{j=0}^p \theta_j d^j$ satisfy $\theta_0 = 1$ and

$$\theta_{p-j} = \theta_j \ ; \qquad j = 0,\ldots,p.$$

It is now easy to see that $Q(\frac{1}{d},\theta) = d^{-p}Q(d,\theta)$, which implies

$$f_2(\frac{1}{d},\theta) = B_2(\theta)D\frac{h_2(d)}{Q^2(d,\theta)} = \tilde{D}_2 f_2(d,\theta)$$

where $\tilde{D}_2 = B_2(\theta)DB_2^{-1}(\theta)$ and D is defined in Section 4.1. It is now easy to see that the statement of the previous remain valid. For a precise statement, define the set $\mathcal{B}^* \subset \mathbb{R}^{2p+1}$ as the set of all vectors $\hat{c} = (\hat{c}_1,\ldots,\hat{c}_{2p+1})^T$ satisfying (4.6), where the functions $g_1,\ldots,g_{2p+1}$ in the determinant are given by $g_j(d) = d^{j-1}/Q^2(d,\theta)$ $(j = 1,\ldots,2p+1)$.

THEOREM 4.3 *Consider the model* (2.3) *with* $p = q$*, design space* $\mathcal{D} = [1/d_u, d_u]$ *and polynomial* $Q(d,\theta)$ *satisfying* (4.2).

(a) If $\hat{c} = B_2^{-1}(\theta)c \in \mathcal{B}^*$*, then the c-optimal design has* $2p+1$ *support points* $1/d_u = d_1 < d_2 < \cdots < d_{2p+1} = d_u$ *satisfying* (4.8) *and the weights are given by* (4.7).

(b) If there exists an orthogonal matrix $A \in \mathbb{R}^{s\times s}$ *such that* $\tilde{D}_2 K A^{-1} = K$*, then there exists a number* $t^* \in \{1,\ldots,2p+1\}$ *such that the* Φ_ℓ*-optimal design has masses* $w_1^*,\ldots,w_{t^*}^*$ *at the points* $d_1^* < d_2^* < \cdots < d_{t^*}^*$*, which satisfy* (4.14) *and* (4.15)*, respectively. Moreover, if* $s = 2p+1$ *and* K *is non-singular, then* $t^* = 2p+1$ *and the support of the optimal design contains the boundary points* $1/d_u$ *and* d_u *of the design space. For the D-optimality criterion* ($\ell = 0$) *these statements remain valid if the matrix* A *is non-singular (but not necessarily orthogonal).*

Example 4.2 Consider the problem of constructing D-optimal designs for the model

$$\eta_2(x,\theta) = \frac{\theta_0 + \theta_1 x + \theta_2 x^2}{1 + \theta_3 x + \theta_4 x^2} \tag{4.17}$$

where $x = [0.2, 5]$. In this case we have $K = I_s$ and the condition of part (b) of Theorem 4.3 is obviously satisfied. The D-optimal design (with $\theta_4 = 1$) puts equal masses at the points $0.2, 1/x, 1, x, 5$. The remaining point x can now easily be found numerically. For example, if $\theta_3 = 2,\ \theta_4 = 1$, the support points are given by 0.2, 0.3923, 1, 2.54884, 5 while for $\theta_3 = 8,\ \theta_4 = 1$ the support is given by 0.2, 0.4031, 1, 2.8408, 5. The corresponding plots of the equivalence theorem are depicted in Figure 1.

Acknowledgements The authors 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, Teilprojekt C1) of the German Research Foundation (DFG).

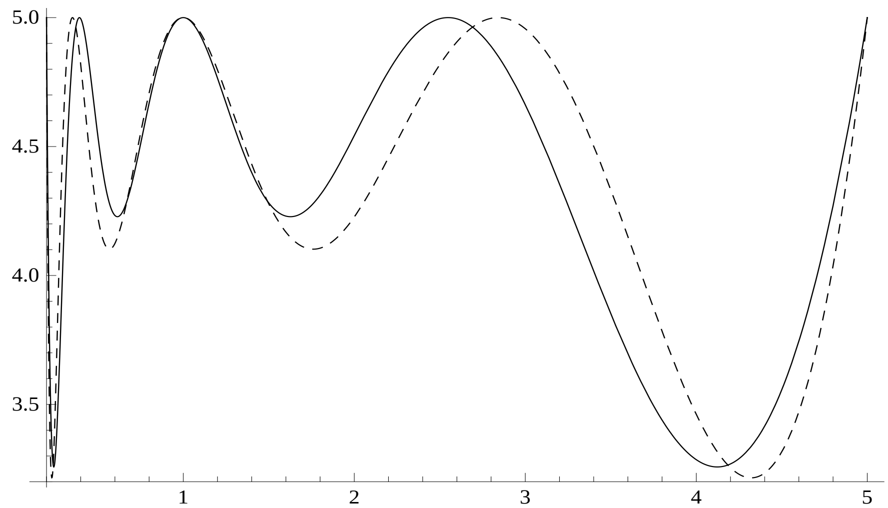

Figure 1: *The checking condition (3.3) for the D-optimality criterion for the model (4.17). Solid curve* $\theta_3 = 2$*; dashed curve* $\theta_3 = 8$*.*

References

Boroujerdi, M. (2002). *Pharmacokinetics: Principles and Applications*. New York: Mc-Graw Hill.

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

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

Cobby, J. M., Chapman, P. F., and Pike, D. J. (1986). Design of experiments for estimating inverse quadratic polynomial responses. *Biometrics*, 42:659–664.

Cornish-Browden, A. (1995). *Fundamentals of Enzyme Kinetics (Rev. Ed.)*. Portland Press, London.

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

Dette, H. and Biedermann, S. (2003). Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association*, 98:679–686.

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

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

Dette, H. and Kiss, C. (2009). Optimal experimental designs for inverse quadratic regression models. *Statistica Sinica*, 19(4):1567–1586.

Dette, H., Kiss, C., Bevanda, M., and Bretz, F. (2010). Optimal designs for the EMAX, log-linear and exponential models. *Biometrika*, 97(2):513–518.

Dette, H. and Melas, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Annals of Statistics*, 39(2):1266–1281.

Dette, H., Melas, V. B., and Pepelyshev, A. (2003). Standardized maximin E-optimal designs for the Michaelis-Menten model. *Statistica Sinica*, 13:1147–1163.

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

Dunn, G. (1988). "Optimal" designs for drug, neurotransmitter and hormone receptor assays. *Statistics in Medicine*, 7:805–815.

Haines, L. M. (1992). Optimal design for inverse quadratic polynomials. *South African Statistical Journal*, 26:25–41.

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(3):763–783.

Johansen, S. (1984). Functional relations, random coefficients and nonlinear regression, with application to kinetic data. *Lecture Notes in Statistics*, 22.

Karlin, S. and Studden, W. J. (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.

López-Fidalgo, J. and Wong, W. K. (2002). Design issues for the Michaelis−Menten model. *Journal of Theoretical Biology*, 215:1–11.

Merle, Y. and Mentre, F. (1995). Bayesian design criteria - computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. *Journal of Pharmacokinetics And Biopharmaceutics*, 23:101–125.

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

Rasch, D. (1990). Optimum experimental design in nonlinear regression. *Comm. Statist. Theory Methods*, 19(12):4789–4806.

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.

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

Song, D. and Wong, W. K. (1998). Optimal two-point designs for the Michaelis-Menten model with heteroscedastic errors. *Communications in Statistics: Theory and Methods*, 27(6):1503–1516.

Wang, J. (2006). Optimal parametric design with applications to pharmacokinetic and pharmacodynamic trials. *Journal of Applied Statistics*, 33, 8:837–852.