Robust T-optimal discriminating designs

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de
FAX: +49 234 32 14559

Viatcheslav B. Melas, Petr Shpilev
St. Petersburg State University
Department of Mathematics
St. Petersburg, Russia
email: v.melas@pobox.spbu.ru
email: pitshp@hotmail.com

Abstract

This paper considers the problem of constructing optimal discriminating experimental designs for competing regression models on the basis of the T-optimality criterion introduced by Atkinson and Fedorov (1975a). T-optimal designs depend on unknown model parameters and it is demonstrated that these designs are sensitive with respect to misspecification. As a solution of this problem we propose a Bayesian and standardized maximin approach to construct robust and efficient discriminating designs on the basis of the T-optimality criterion. It is shown that the corresponding Bayesian and standardized maximin optimality criteria are closely related to linear optimality criteria. For the problem of discriminating between two polynomial regression models which differ in the degree by two the robust T-optimal discriminating designs can be found explicitly. The results are illustrated in several examples.

AMS Subject Classification: 62K05

Keywords and Phrases: optimal design; model discrimination; robust design; linear optimality criteria; Chebyshev polynomial

1 Introduction

An important problem of regression analysis is the identification of an appropriate model to describe the relation between the response and a predictor. Typical examples include dose response studies [see e.g. Bretz et al. (2005)] in medicine or toxicology or problems in pharmacokinetics, where a model has usually to be chosen from a class of competing regression functions [see e.g. Atkinson et al. (1998), Asprey and Macchietto (2000), Ucinski and Bogacka (2005) or Foo and Duffull (2011)]. Because a misspecification of a regression model can result in an inefficient - in the worst case incorrect - data analysis, several authors argue that the design of the experiment should take the problem of model identification into account. Meanwhile a huge amount of literature can be found which addresses the construction of efficient designs for model discrimination. The literature can be roughly decomposed into two groups.

Hunter and Reiner (1965), Stigler (1971), Hill (1978), Studden (1982), Spruill (1990), Dette (1994, 1995), Dette and Haller (1998), Song and Wong (1999) (among many others) considered two nested models, where the extended model reduces to the "smaller" model for a specific choice of a subset of the parameters. The optimal discriminating designs are then constructed such that these parameters are estimated most precisely. This concept relies heavily on the assumption of nested models, and as an alternative Atkinson and Fedorov (1975a) introduced in a fundamental paper the T-optimality criterion for discriminating between two competing regression models. Since its introduction this criterion has been studied by numerous authors [Atkinson and Fedorov (1975b), Ucinski and Bogacka (2005), Waterhouse et al. (2008), Dette and Titoff (2009), Atkinson (2010), Tommasi and López-Fidalgo (2010), Wiens (2009, 2010) or Dette et al. (2012) among others].

The T-optimal design problem is essentially a maximin problem and the criterion can also be applied for non-nested models. Except for very simple models, T-optimal discriminating designs are not easy to find and even their numerical determination is a very challenging task. Moreover, an important drawback of this approach consists in the fact that the criterion and, as a consequence, the corresponding optimal discriminating designs depend sensitively on the parameters of one of the competing regression models. In contrast to other optimality criteria this dependence appears even in the case where only linear models have to be discriminated. Therefore T-optimal designs are locally optimal in the sense of Chernoff (1953) as they can only be implemented if some prior information regarding these parameters is available. Moreover, we will demonstrate in Example 2.1 that the efficiency of a T-optimal design depends sensitively on a precise specification of the unknown parameters in the criterion. This problem has already been recognized by Atkinson and Fedorov (1975a) who proposed Bayesian or minimax versions of the T-optimality criterion. However - to the best knowledge of the authors - there exist no results in the literature investigating optimal design problems of this type more rigorously (we are even not aware of any numerical solutions).

The present paper is devoted to a more detailed discussion of robust T-optimal discriminating designs. We will study a Bayesian and a standardized maximin version of the T-optimal discriminating design problem [see Chaloner and Verdinelli (1995) and Dette (1997)]. It is demonstrated that optimal designs with respect to these criteria are closely related to optimal designs with respect to linear optimality criteria. For the particular case of discriminating between two competing polynomial regression models which differ in the degree by two, robust T-optimal discriminating designs are found explicitly. These results provide - to our best knowledge - the first explicit solution in this context. Interestingly, the structure of these Bayesian and standardized maximin T-optimal discriminating designs is closely related to the structure of designs for a most precise estimation of the two highest coefficients in a polynomial regression model [see Gaffke (1987) or Studden (1989)].

The remaining part of the paper is organized as follows. In Section 2 we revisit the T-optimality criterion introduced by Atkinson and Fedorov (1975a) for two regression models, which will be called locally T-optimality criterion in order to reflect the dependency on the parameters of one of the competing models. In particular it is demonstrated that locally T-optimal designs can be inefficient if the parameters in the optimality criterion have been misspecified. Section 3 is devoted to robust versions of the T-optimality criterion and properties of the corresponding optimal designs, while Section 4 gives explicit results for Bayesian and standardized maximin T-optimal discriminating designs for two competing polynomial regression models. In Section 5 we illustrate the results and construct robust optimal discriminating designs for a constant and quadratic regression. These two models have been proposed in Bretz et al. (2005) to detect dose response signal in phase II clinical trial if there is some evidence that the shape of the dose response might be u-shaped.

2 Locally T-optimal designs

We assume that the relation between a predictor x and response y is described by the regression model

$$y = \eta(x) + \varepsilon,$$

where x varies in a compact designs space $\mathcal{X} \subset \mathbb{R}^k$ and ε denotes a centered random variable with finite variance. We also assume that observations at experimental conditions x_1 and x_2 are independent and that there exist two competing continuous parametric models, say η_1 or η_2, for the regression function η with corresponding parameters $\theta_1 \in \mathbb{R}^{m_1}$; $\theta_2 \in \mathbb{R}^{m_2}$, respectively. In order to find "good" designs for discriminating between the models η_1 and η_2 we consider approximate designs in the sense of Kiefer (1974), which are defined as probability measures on the design space $\mathcal{X}$ with finite support. The support points, say $x_1, \dots, x_s$, of an (approximate) design ξ give 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 $\omega_i > 0$ at the different points x_i ($i = 1, \dots, s$) and n observations can be made by the experimenter, the quantities $\omega_i n$ are rounded to integers, say n_i, satisfying $\sum_{i=1}^s n_i = n$, and the experimenter takes n_i observations at each location x_i ($i = 1, \dots, s$).

To determine a good design for discriminating between the two rival regression models η_1 and η_2 Atkinson and Fedorov (1975a) proposed in a fundamental paper to fix one model, say η_2 (more precisely its corresponding parameter θ_2), and to determine the design which maximizes the minimal deviation

$$T(\xi, \theta_2) = \min_{\theta_1 \in \Theta_1} \int_{\chi} \left(\eta_1(x, \theta_1) - \eta_2(x, \theta_2)\right)^2 \xi(dx) \tag{2.1}$$

between the model η_2 and the class of models $\{\eta_1(x, \theta_1) \mid \theta_1 \in \Theta_1\}$ defined by η_1, that is

$$\xi^* = \arg\max_{\xi} T(\xi, \theta_2). \tag{2.2}$$

$\theta_{2,0}$	$\theta_{2,2}$	x^*	w_1	w_2	w_3	$\theta_{2,0}$	$\theta_{2,2}$	x^*	w_1	w_2	w_3
-2	2	1.368	0.206	0.499	0.295	-2	1	1.352	0.211	0.499	0.29
-1	2	1.347	0.176	0.495	0.329	-1	1	1.321	0.165	0.491	0.344
$-1/2$	2	1.211	0.040	0.584	0.376	$-1/2$	1	1.590	0.619	0.336	0.045
$1/2$	2	1.400	0.260	0.498	0.242	$1/2$	1	1.384	0.261	0.498	0.239
1	2	1.390	0.247	0.499	0.254	1	1	1.378	0.253	0.499	0.248
2	2	1.387	0.238	0.499	0.263	2	1	1.337	0.244	0.500	0.256

Table 1: *The support points and the weights of T-optimal discriminating designs for a Michaelis Menten and an EMAX model and various specifications of the parameters $\theta_{2,0}$ and $\theta_{2,2}$ of the EMAX model. The locally T-optimal design puts weights w_1, w_2 and w_3 at the points 1, x^* and 2, respectively.*

Throughout this paper we call the maximizing design and optimality criterion in (2.2) locally T-optimal discriminating design and local T-optimality criterion, respectively, because they will depend on the specification of the parameter θ_2 used for the model η_2. The local T-optimal design problem is a maximin problem and except for very simple models the corresponding optimal designs are extremely hard to find. Even their numerical construction is a difficult and challenging task. Nevertheless, since its introduction the optimal designs with respect to the criterion (2.1) have found considerable interest in the literature and we refer the interested reader to the work of Ucinski and Bogacka (2005) or Dette and Titoff (2009) among others. The latter authors showed that the optimization problem (2.2) is closely related to a problem in nonlinear approximation theory, that is

$$R(\theta_2) := \max_{\xi} \; T(\xi, \theta_2) = \inf_{\theta_1 \in \Theta_1} \sup_{x \in \mathcal{X}} \mid \eta_1(x, \theta_1) - \eta_2(x, \theta_2) \mid^2, \tag{2.3}$$

where $T(\xi, \theta_2)$ is defined in (2.1). Because of its local character locally T-optimal designs are rather sensitive with respect to the misspecification of the unknown parameter and the following example illustrates this fact.

Example 2.1 We consider the problem of constructing a T-optimal discrimination design for the Michaelis Menten model

$$\eta_1(x, \theta_1) \;\; = \;\; \frac{\theta_{1,1} x}{\theta_{1,2} + x}$$

[see for example Cornish-Bowden (1965)] and the EMAX model

$$\eta_2(x, \theta_2) \;\; = \;\; \theta_{2,0} + \frac{\theta_{2,1} x}{\theta_{2,2} + x},$$

[see for example Danesi et al. (2002)]. It is easy to see that the T-optimal discriminating design does not depend on the parameter $\theta_{2,1}$ and therefore we assume without loss of generality $\theta_{2,1} \equiv 1$.

In Table 1 we display some locally T-optimal discriminating designs on the interval $[1,2]$ for various values of parameters $\theta_{2,i},\ i=0,2$. We observe that the resulting designs are rather sensitive with respect to the specification of the values $\theta_{2,0}$ and $\theta_{2,2}$. Note that in contrast to the T-optimal discriminating design the T-efficiency

$$\mathrm{Eff}_T(\xi,\theta_2)=\frac{T(\xi,\theta_2)}{\sup_\eta T(\eta,\theta_2)} \tag{2.4}$$

depends also on the parameter $\theta_{2,1}$ of the EMAX model and some efficiencies are depicted in Figure 1 if the true values are given by $\theta_{2,0}=-1$, $\theta_{2,1}=1$, $\theta_{2,2}\in(2,6)$ and one uses the T-optimal discriminating design calculated under the assumption $\theta_{2,0}=-1/4$, $\theta_{2,1}=1$ and $\theta_{2,2}\in(2,6)$. We observe a substantial loss of T-efficiency in some regions for $\theta_{2,2}$. If $\theta_{2,2}\in(0,2)$ the efficiency is larger than 50%, if $\theta_{2,2}\in(2,3)\cup(5.5,6)$ it varies between 15% and 40%, if $\theta_{2,2}\in(3,5.5)$ the efficiency is smaller than 15% and the locally T-optimal design cannot be recommended. On the basis of these observations it might be desirable to use designs which are less sensitive with respect to misspecification of the parameter $\theta_{2,0}$ and the corresponding methodology will be developed in the following section. Robust T-optimal designs for discriminating between the Michaelis and EMAX model will be discussed at the end of this paper where we construct a uniformly better design [see Section 5.3].

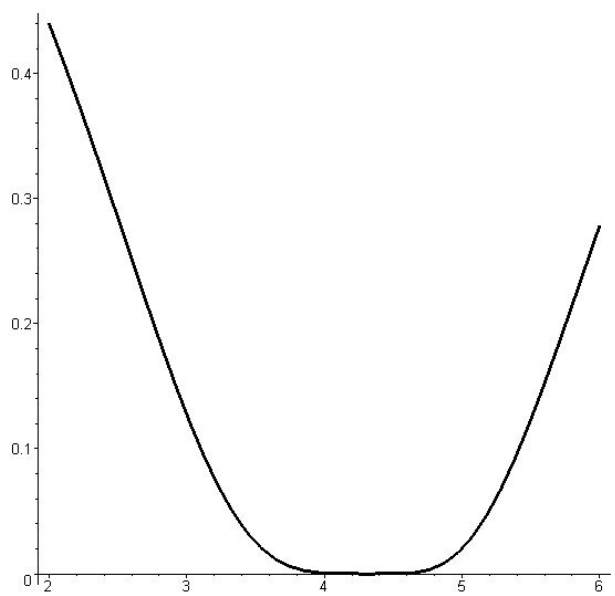

Figure 1: *T-efficiency (2.4) of the locally T-optimal discriminating design for Michaelis Menten and Emax model calculated under the assumption $\theta_{2,0}=-1/4$, $\theta_{2,1}=1$ while the "true" values are given by $\theta_{2,0}=-1$, $\theta_{2,1}=1$. The efficiencies depend on the parameter $\theta_{2,2}\in(2,6)$.*

3 Robust T-optimal discriminating designs

Because the previous example indicates that locally T-optimal discriminating designs are sensitive with respect to misspecification of the parameters θ_2 of the model η_2 in the T-optimality criterion (2.1), the consideration of robust optimality criteria for model discrimination is of great interest. In the context of constructing efficient robust designs for parameter estimation in nonlinear regression models Bayesian and standardized maximin optimality criteria have been discussed intensively in the literature [see Chaloner and Verdinelli (1995), Dette (1997) or Müller and Pázman (1998) among many others]. However – to our best knowledge – these methods have not been investigated rigorously in the context of model discrimination so far and in this section we will define a robust version of the local T-optimality criterion. Recall the definition of this criterion in (2.1) and its optimal value $R(\theta_2)$ in (2.3), then a design ξ_M^* is called standardized maximin T-optimal discriminating (with respect to the set Θ_2) if it maximizes the criterion

$$V_M(\xi) = \inf_{\theta_2 \in \Theta_2} \frac{T(\xi, \theta_2)}{R(\theta_2)}, \tag{3.1}$$

where Θ_2 is a pre-specified set, reflecting the experimenter's belief about the unknown parameter θ_2. Similarly, if π denotes a prior distribution on the set Θ_2, then a design ξ_B^* is called Bayesian T-optimal (with respect to the prior π) if it maximizes the criterion

$$V_B(\xi) = \int_{\Theta_2} T(\xi, \theta_2)\pi(d\theta_2). \tag{3.2}$$

In the following discussion we investigate the problem of constructing robust discriminating designs for two linear regression models

$$\eta_1(x, \theta_1) = \sum_{i=0}^{m_1-1} \theta_{1,i} f_i(x), \quad \eta_2(x, \theta_2) = \sum_{i=0}^{m_2-1} \theta_{2,i} f_i(x), \tag{3.3}$$

where $m_2 > m_1$, $f_0, \dots, f_{m_2-1}$ are given linearly independent regression functions and $\theta_i = (\theta_{i,0}, \dots, \theta_{i,m_i-1})^T$ denotes the parameter in the model η_i $(i = 1, 2)$. We introduce the notation $b_1 = \theta_{2,m_1}/\theta_{2,m_2-1}, \dots, b_{m_2-m_1-1} = \theta_{2,m_2-2}/\theta_{2,m_2-1}$, $m = m_2 - 1$, $s = m_2 - m_1$, $q_i = \theta_{1,i} - \theta_{2,i}$ $(i = 0, 1, \dots, m - s)$ and obtain for the difference $\eta_1(x, \theta_1) - \eta_2(x, \theta_2)$ the representation

$$\bar{\eta}(x, q, \theta_{2,m}^c, \theta_{2,m}) = \sum_{i=0}^{m-s} q_i f_i(x) - (b_1 f_{m-s+1}(x) + \dots + b_{s-1} f_{m-1}(x) + f_m(x))\theta_{2,m}, \tag{3.4}$$

where $\theta_{2,m}^c = (\theta_{2,m-s+1}, \dots, \theta_{2,m-1})^T$. Thus the locally T-optimality criterion in (2.1) can be rewritten as

$$T(\xi, \theta_2) \;=\; \inf_{\theta_1 \in \mathbb{R}^{m-s+1}} \int_{\mathcal{X}} (\eta_1(x, \theta_1) - \eta_2(x, \theta_2))^2 d\xi(x) = \theta_{2,m}^2 \inf_{q \in \mathbb{R}^{m-s+1}} \int_{\mathcal{X}} \bar{\eta}^2(x, q, b, 1) d\xi(x),$$

where $b = (b_1, \dots, b_{s-1})^T$. Consequently, locally T-optimal designs depend only on the ratios $b_i = \theta_{2,m-s+i}/\theta_{2,m}$ $(i = 1, \dots, s-1)$. Similarly, if π is a prior distribution for the vector θ_2, then it follows from these discussions that the Bayesian T-optimality criterion depends only on the induced prior distribution, say $\bar{\pi}$, for the parameter $b = (b_1, \ldots, b_{s-1})$. Therefore we assume that the vector b varies in a subset $\mathcal{B} \subset \mathbb{R}^{s-1}$ and define $\bar{\pi}$ as a prior distribution on $\mathcal{B}$. With these notations the Bayesian T-optimality criterion in (3.2) simplifies to

$$V_B(\xi) = \int_{\mathcal{B}} \inf_{q \in \mathbb{R}^{m-s+1}} \int_{\mathcal{X}} \bar{\eta}^2(x, q, b, 1)\xi(dx)\bar{\pi}(db) \tag{3.5}$$

Similarly, we have with the notation $\bar{\theta}_2 = \theta_2/\theta_{2,m}$

$$R(\theta_2) = \max_{\xi} T(\xi, \theta_2) = \theta_{2,m}^2 \max_{\xi} R(\bar{\theta}_2)$$

and defining $\mathcal{B} = \{(\theta_{2,m-s+1}/\theta_{2,m}, \ldots, \theta_{2,m-1}/\theta_{2,m})^T \mid \theta_2 \in \Theta_2\} \subset \mathbb{R}^{s-1}$ and for $b \in \mathcal{B}$

$$\bar{R}(b) = R((b_1, \ldots, b_{s-1}, 1)^T) \tag{3.6}$$

the factor $\theta_{2,m}^2$ in (3.1) cancels and the standardized maximin T-optimality criterion reduces to

$$V_M(\xi) = \inf_{b \in \mathcal{B}} \frac{\inf_{q \in \mathbb{R}^{m-s+1}} \int_{\mathcal{X}} \bar{\eta}^2(x, q, b, 1)\xi(dx)}{\bar{R}(b)} = \inf_{b \in \mathcal{B}} \mathrm{eff}_T(\xi, b), \tag{3.7}$$

where the efficiency is defined in an obvious manner, that is

$$\mathrm{eff}_T(\xi, b) = \frac{T(\xi, (b_1, \ldots, b_{s-1}, 1)^T)}{\bar{R}(b)}.$$

Throughout this paper we denote by $f(x) = (f_0(x), f_1(x), \ldots, f_m(x))^T$ the vector of regression functions with corresponding decomposition

$$\begin{aligned} f_{(1)}(x) &= (f_0(x), f_1(x), \ldots, f_{m-s}(x))^T \in \mathbb{R}^{m-s+1}, \\ f_{(2)}(x) &= (f_{m-s+1}(x), \ldots, f_m(x))^T \in \mathbb{R}^s. \end{aligned}$$

We assume that the functions $f_0, \ldots, f_m$ are linearly independent and continuous on $\mathcal{X}$ and define

$$M(\xi) = \int_{\mathcal{X}} f(x) f^T(x) \xi(dx)$$

as the information matrix of a design with corresponding blocks

$$M_{ij}(\xi) = \int_{\mathcal{X}} f_{(i)}(x) f_{(j)}^T(x) \xi(dx), \ i, j = 1, 2.$$

and Schur complement

$$M_{(s)}(\xi) = M_{22}(\xi) - X^T M_{11}(\xi) X,$$

where $X \in \mathbb{R}^{m-s+1 \times s}$ is an arbitrary solution of the equation $M_{11}(\xi)X = M_{12}(\xi)$ (if this equation has no solutions, then the matrix $M_{(s)}(\xi)$ remains undefined). Our first main result relates the Bayesian and standardized maximin T-optimality criteria to linear optimality criteria.

Theorem 3.1 *Let $\bar{\pi}$ denote a prior distribution for the vector $b \in \mathcal{B}$, such that the matrix*

$$L = \int_{\mathcal{B}} \begin{pmatrix} bb^T & b \\ b^T & 1 \end{pmatrix} \bar{\pi}(db).$$

exists, then the two following statements are equivalent.

(1) The design ξ^* is a Bayesian T-optimal discriminating design with respect to the prior $\bar{\pi}$ for the linear regression models defined in (3.3)*.*

(2) The design ξ^* maximizes the linear criterion

$$\operatorname{tr}\, LM_{(s)}(\xi),$$

in the class of all approximate designs ξ, for which there exists a solution $X \in \mathbb{R}^{m-s+1\times s}$ of the equation

$$M_{11}(\xi)X = M_{12}(\xi). \tag{3.8}$$

Proof. If the matrix $M_{(s)}(\xi)$ is non-singular, then it follows from Karlin and Studden (1966), Section 10.8, that

$$\big(M_{(s)}(\xi)\big)^{-1} = (\mathbf{O}^T \,:\, I_s)M^-(\xi) \begin{pmatrix} \mathbf{O} \\ I_s \end{pmatrix},$$

where $I_s \in \mathbb{R}^{s\times s}$ is the identity matrix, $\mathbf{O} \in \mathbb{R}^{m-s+1\times s}$ is the matrix with all entries equal to 0 and $M^-(\xi)$ is an arbitrary generalized inverse of the matrix $M(\xi)$. For any $(m-s+1)\times s$ matrix K we have the inequality

$$(-K^T \,:\, I_s) \begin{pmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) \end{pmatrix} \begin{pmatrix} -K \\ I_s \end{pmatrix} \geq M_{(s)}(\xi),$$

where there is equality if and only if the matrix K is a solution of the equation (3.8) [see Karlin and Studden (1966), Section 10.8]. From (3.4) and the discussion in the subsequent paragraph we obtain the representation

$$T(\xi,\theta_2) = \theta_{2,m}^2 \min_{q\in\mathbb{R}^{m-s+1}} (q^T, b^T, 1)M(\xi)(q^T, b^T, 1)^T = \theta_{2,m}^2 (b^T, 1)M_{(s)}(\xi)(b^T, 1)^T, \tag{3.9}$$

where the last equality follows from the fact that each vector $(q^T, b^T, 1)^T$ can be represented in the form

$$(q^T, b^T, 1)^T = (-K^T : I_s)^T (b^T, 1)^T$$

for some appropriate matrix $K \in \mathbb{R}^{m-s+1\times s}$ (just use the matrix $K = -q(b^T, 1)/(b^Tb+1)$). The assertion of Theorem 3.1 is now obvious. $\Box$

A similar result for standardized maximin T-optimal discriminating designs is formulated in the following Theorem. Throughout this paper we will use the notation $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ with the usual compactification.

Theorem 3.2 *If $\mathcal{B} \subset \bar{\mathbb{R}}^{s-1}$ be a given compact set, then the following two statements are equivalent.*

(1) The design ξ^* is a standardized maximin T-optimal discriminating design for the regression models defined in (3.3) *with respect to the set $\mathcal{B}$.*

(2) For the design ξ^* there exists a solution of the equation (3.8) and a matrix $L^* \in \mathbb{R}^{s \times s}$ such that the pair (L^*, ξ^*) satisfies

$$\operatorname{tr} L^* M_{(s)}(\xi^*) = \sup_{\xi} \operatorname{tr} L^* M_{(s)}(\xi), \tag{3.10}$$

$$\operatorname{tr} L^* M_{(s)}(\xi^*) = \inf_{\mathcal{L}} \operatorname{tr} L M_{(s)}(\xi^*), \tag{3.11}$$

where the supremum in (3.10) *is taken with respect to all approximate designs and the set $\mathcal{L}$ in* (3.11) *is defined by*

$$\Big\{ \sum_{i=1}^{k} (b_i^T, 1)^T (b_i^T, 1) \frac{\omega_i}{\bar{R}(b_i)} \;\Big|\; b_i \in \mathcal{B},\ \omega_i > 0 \quad \sum_{i=1}^{k} \omega_i = 1 \Big\}.$$

Proof. By a similar argument as used in the proof of Theorem 3.1 the standardized T-optimality criterion in (3.7) can be represented as

$$\begin{aligned} V_M(\xi) &= \inf_{b \in \mathcal{B}} \inf_{q \in \mathbb{R}^{m-s+1}} \frac{(q^T, b^T, 1) M(\xi) (q^T, b^T, 1)^T}{\bar{R}(b)} \\ &= \inf_{b \in \mathcal{B}} \frac{(b^T, 1) M_{(s)}(\xi) (b^T, 1)^T}{\bar{R}(b)} = \inf_{L \in \mathcal{L}} \operatorname{tr} L M_{(s)}(\xi). \end{aligned}$$

The assertion now follows from the von Neumann theorem on minimax problems [see Osborne and Rubinstein (1994)]. $\Box$

A lower bound for the efficiencies of a standardized maximin T-optimal discriminating design is given in the following theorem.

Theorem 3.3 *Let ξ^* denote a standardized maximin T-optimal discriminating design for the linear regression models defined in* (3.3) *with respect to set $\mathcal{B}$. Then for all $b \in \mathcal{B}$*

$$\mathit{eff}_T(\xi^*, b) \geq \frac{1}{s}.$$

Proof. Recall the definition of the standardized maximin optimality criterion in (3.7). Because for any $b \in \mathcal{B}$

$$\operatorname{eff}_T(\xi^*, b) \geq \inf_{b \in \mathcal{B}} \operatorname{eff}_T(\xi^*, b) = V_M(\xi^*)$$

the assertion follows, if the inequality

$$\sup_{\xi} V_M(\xi) \geq \frac{1}{s}$$

can be established. For this purpose we define the function

$$\psi(x) = f_{(2)}(x) - X^T f_{(1)}(x), \tag{3.12}$$

where X is an $(m - s + 1) \times s$-matrix (the dependence of the function ψ on this matrix is not reflected in the notation). Let ξ be an arbitrary design such that the matrix $M_{(s)}(\xi)$ is non-singular, then it follows from the Cauchy Schwartz inequality that

$$\inf_{l \in \mathbb{R}^s \setminus 0} \frac{l^T M_{(s)}(\xi) l}{\sup_{x \in \mathcal{X}} (l^T \psi(x))^2} \geq \frac{1}{\sup_{x \in \mathcal{X}} \psi^T(x) M_{(s)}^{-1}(\xi) \psi(x)}. \tag{3.13}$$

By the equivalence theorem for D_s-optimal designs [see (Karlin and Studden (1966), Section 10.8)] there exists a design $\tilde{\xi}$ and a matrix $\tilde{X}$ satisfying $M_{11}(\tilde{\xi})\tilde{X} = M_{12}(\tilde{\xi})$, such that the corresponding matrix $M_{(s)}(\tilde{\xi})$ and the vector $\tilde{\psi}(x) = f_{(2)}(x) - \tilde{X}^T f_{(1)}(x)$ satisfy

$$\max_{x \in \mathcal{X}} \tilde{\psi}^T(x) M_{(s)}^{-1}(\tilde{\xi}) \tilde{\psi}(x) = s.$$

Consider any design ξ for which a solution X of (3.8) exists, then we have for the corresponding function ψ in (3.12)

$$M_{(s)}(\xi) = \int_{\mathcal{X}} \psi(x) \psi^T(x) \xi(dx).$$

Therefore we obtain from formula (3.9)

$$\bar{R}(b) = \theta_{2,m}^2 \max_{\xi} \int_{\mathcal{X}} ((b^T, 1)\psi(x))^2 \xi(dx) = \theta_{2,m}^2 \max_{x \in \mathcal{X}} ((b^T, 1)\psi(x))^2,$$

which gives for the vector $l = (\theta_{2,m-s+1}, \ldots, \theta_{2,m})^T$

$$\sup_{x \in \mathcal{X}} \left(l^T \psi(x) \right)^2 = \bar{R}(b),$$

where $b = (\theta_{2,m-s+1}/\theta_{2,m}, \ldots, \theta_{2,m-1}/\theta_{2,m})^T$. Thus the left hand side in (3.13) equals $V_M(\xi)$ and

$$\sup_{\xi} V_M(\xi) \geq V_M(\tilde{\xi}) \geq \frac{1}{s},$$

which proves the assertion of Theorem 3.3. $\square$

4 Robust T-optimal designs for polynomial regression

In general locally T-optimal discriminating designs have to be found numerically and this statement also applies to the construction of robust T-optimal discriminating designs with respect to the Bayesian or standardized maximin criterion. In order to get more insight in the corresponding optimal design problems we consider in this section the case of two competing polynomial regression models which differ in the degree by two. Remarkably, for this situation the robust T-optimal discriminating designs can be found explicitly. To be precise, let $s = 2$, consider the vectors of monomials

$$f_{(1)}(x) = (1, x, \dots, x^{m-2})^T,\ f_{(2)} = (1, x, \dots, x^m)^T,$$

and define

$$U_n(x) = \frac{\sin((n+1)\text{arcos } x)}{\sin(\text{arcos } x)}$$

as the Chebyshev polynomial of the second kind [see Szegö (1959)]. We assume that the design space is given by the symmetric interval $[-a, a]$ and consider for $\beta > 0$ designs $\xi_{m,\beta}$ defined as follows. If $\beta = 1$ then the design $\xi_{m,1}$ puts masses $1/(2(m-1))$ at the points $-a$, a and masses $1/(m-1)$ at the $m-2$ roots of the polynomial $U_{m-2}(x/a)$. If $\beta \neq 1$ the design $\xi_{m,\beta}$ is supported at the $m+1$ roots $-a = x_0 < x_1 < \ldots < x_{m-1} < x_m = a$ of the polynomial

$$(x^2 - a^2)\left\{U_{m-1}\left(\frac{x}{a}\right) + \beta U_{m-3}\left(\frac{x}{a}\right)\right\},$$

where the corresponding weights are given by

$$\begin{aligned}
\xi_{m,\beta}(\mp a) &= \frac{1+\beta}{2[m + \beta(m-2)]}, \\
\xi_{m,\beta}(x_j) &= \left[m - 1 - \frac{(1+\beta)U_{m-2}\left(\frac{x_j}{a}\right)}{U_m\left(\frac{x_j}{a}\right) + \beta U_{m-2}\left(\frac{x_j}{a}\right)}\right]^{-1},\ j = 1, \dots, m-1,
\end{aligned}$$

Theorem 4.1

(1) Let $\bar{\pi}$ denote a symmetric prior distribution on $\mathcal{B} \subseteq (-\infty, \infty)$ with existing second moment, and define

$$\beta_B = \min\left\{1, \frac{\int_{\mathcal{B}} b^2 \bar{\pi}(db)}{a^2}\right\}. \tag{4.1}$$

The design ξ_{m,β_B} is a Bayesian T-optimal discriminating on the interval $[-a, a]$ for the polynomial regression models of degree $m-2$ and m.

(2) Define $\beta_M = 1 - 2h^*$, where h^* is the unique maximizer of the function

$$\inf_{b \in \mathcal{B}} \frac{b^2 + a^2 h}{a^2 \bar{R}(b, a)}(1-h), \tag{4.2}$$

where

$$\bar{R}(b,a) = \inf_{q_0,\ldots,q_{m-2}\in\mathbb{R}} \sup_{x\in[-1,1]} a^{2m} \mid x^m + \frac{b}{a}x^{m-1} + q_{m-2}x^{m-2} + \cdots + q_1 x + q_0 \mid^2$$

in the interval $[0, \frac{1}{2}]$. *Then the design* ξ_{m,β_M} *is a standardized maximin* T*-optimal discriminating design on the interval* $[-a, a]$ *for the polynomial regression models of degree* $m-2$ *and* m.

Proof of Theorem 4.1. We will prove the statement using some basic facts of the theory of canonical moments [see Dette and Studden (1997) for details]. To be precise, let $\mathcal{P}([-a,a])$ denote the set of all probability measures on the interval $[-a,a]$, and denote for a design $\xi \in \mathcal{P}([-a,a])$ its moments by

$$c_i = c_i(\xi) = \int_{-a}^{a} x^i \xi(dx), \;\; i = 1, 2 \ldots .$$

Define $\mathcal{M}_k = \{(c_1,\ldots,c_k)^T \mid \xi \in \mathcal{P}([-a,a])\}$ as the kth moment space and $\Phi_k(x) = (x,\ldots,x^k)$ as the vector of monomials of order k. Consider for a fixed vector $c = (c_1,\ldots,c_k)^T \in \mathcal{M}_k$ the set

$$\mathcal{S}_k(c) := \left\{ \mu \in \mathcal{P}([-a,a]) : \int_{-a}^{a} \Phi_k(x)\mu(dx) = c \right\}$$

of all probability measures on the interval $[0,1]$ whose moments up to the order k coincide with $c = (c_1,\ldots,c_k)^T$. For $k = 2,3,\ldots$ and for a given point $(c_1,\ldots,c_{k-1})^T \in \mathcal{M}_{k-1}$ we define $c_k^+ = c_k^+(c_1,\ldots,c_{k-1})$ and $c_k^- = c_k^-(c_1,\ldots,c_{k-1})$ as the largest and smallest value of c_k such that $(c_1,\ldots,c_k)^T \in \partial\mathcal{M}_k$, that is

$$\begin{aligned} c_k^- &= \min\Big\{ \int_{-a}^{a} x^k \mu(dx) \mid \mu \in S_{k-1}(c_1,\ldots,c_{k-1}) \Big\}, \\ c_k^+ &= \max\Big\{ \int_{-a}^{a} x^k \mu(dx) \mid \mu \in S_{k-1}(c_1,\ldots,c_{k-1}) \Big\}. \end{aligned}$$

Note that $c_k^- \le c_k \le c_k^+$ and that both inequalities are strict if and only if $(c_1,\ldots,c_{k-1})^T \in \mathcal{M}_{k-1}^0$ where $\mathcal{M}_{k-1}^0$ denotes the interior of the set $\mathcal{M}_{k-1}$ [see Dette and Studden (1997)]. For a moment point $c = (c_1,\ldots,c_n)^T$, such that $c = (c_1,\ldots,c_{n-1})^T$ is in the interior of the moment space $\mathcal{M}_{n-1}$, the canonical moments or canonical coordinates of the vector c are defined by $p_1 = c_1$ and

$$p_k = \frac{c_k - c_k^-}{c_k^+ - c_k^-}, \quad k = 2,\ldots,n \; . \tag{4.3}$$

Note that $p_k \in (0,1)$, $k = 1,\ldots,n-1$ and $p_n \in \{0,1\}$ if and only if $(c_1,\ldots,c_{n-1}) \in \mathcal{M}_{n-1}^0$ and $(c_1,\ldots,c_n)^T \in \partial\mathcal{M}_n$. In this case the canonical moments p_i or order $i > n$ remain undefined.

We begin with a proof of the first part of Theorem 4.1. By Theorem 3.1 the determination of Bayesian T-optimal discriminating designs can be obtained by minimizing the linear optimality criterion

$$\mathrm{tr} L M_{(2)}(\xi)$$

for some appropriate matrix L, which is diagonal by the symmetry of the prior distribution. A standard argument of optimal design theory shows that there exists a symmetric Bayesian T-optimal discriminating design, say ξ, for which the corresponding 2×2 matrix $M_{(2)}(\xi)$ is also diagonal, that is

$$M_{(2)}(\xi) = \begin{pmatrix} a_{m-1} & 0 \\ 0 & a_m \end{pmatrix}.$$

It now follows from Dette and Studden (1997), Section 5.7, that for such a design the elements in this matrix are given by

$$(4.4) \qquad a_k(\xi) = (2a)^{2k} \prod_{i=1}^{k} q_{2i-2} p_{2i-1} q_{2i-1} p_{2i} \ , \quad k = m-1, m,$$

where $q_0 = 1$, $q_i = 1 - p_i$ $(i \geq 1)$. Consequently, by Theorem 3.1 the Bayesian T-optimal discriminating design problem is reduced to maximization of the function

$$(4.5) \qquad \mathrm{tr} L M_{(2)}(\xi) = a_m(\xi) + \beta a_{m-1}(\xi),$$

where the quantities $a_m(\xi)$ are defined in (4.4) and $\beta = \int b^2 \bar{\pi}(db)$ denotes the second moment of the prior distribution. This expression can now be directly maximized in terms of the canonical moments, which gives $p_{2m} = 1$, $p_i = \frac{1}{2}$, $i = 1, 2, \ldots, 2m-1$, $i \neq 2m-2$ and

$$p_{2m-2} = \min \left\{ \frac{a^2 + \beta}{2a^2}, 1 \right\} = \frac{1 + \beta_B}{2},$$

where β_B is defined in (4.1). The corresponding design is uniquely determined and can be obtained from Theorem 4.4.4 and 1.3.2 in Dette and Studden (1997), which proves the first part of the Theorem.

For a proof of the second part we note that it follows from the proof of Theorem 3.2 that the standardized maximin T-optimal criterion reduces to

$$(4.6) \qquad \inf_{b \in \mathcal{B}} \frac{a_m(\xi) + b^2 a_{m-1}(\xi)}{\bar{R}(b)} \to \sup_{\xi} .$$

where $\bar{R}(b)$ is defined in (3.6), that is

$$\bar{R}(b) = \inf_{q_0, \ldots, q_{m-2} \in \mathbb{R}} \sup_{x \in [-a,a]} | x^m + b x^{m-1} + q_{m-2} x^{m-2} + \cdots + q_1 x + q_0 |^2 = \bar{R}(b, a).$$

From (4.4) it is obvious that the canonical moments of a (symmetric) standardized maximin T-optimal discriminating design satisfy $p_{2m} = 1$,

$$p_i = \frac{1}{2}, \ i = 1, 2, \ldots, 2m-3, 2m-1,$$

and it remains to maximize (4.6) with respect to the quantity p_{2m-2}. A straightforward calculation shows that the optimal value of p_{2m-2} is determined by the condition $p_{2m-2} = 1 - h^*$, where h^* is a solution of the problem

$$\inf_{b \in \mathcal{B}} \frac{a^2 h + b^2}{\bar{R}(b,a) a^2} (1-h) \to \max_{0 \leq h \leq 1/2} .$$

The corresponding design is uniquely determined and can again be obtained from Theorem 4.4.4 and 1.3.2 in Dette and Studden (1997), which completes the proof of Theorem 4.1. □

Remark 4.1 The structure of the Bayesian and standardized maximin T-optimal designs determined in Theorem 4.1 is the same as the structure of the ϕ_p-optimal design for estimating the two coefficients corresponding to the powers x^m and x^{m-1} in a polynomial regression model of degree m on the interval $[-a, a]$. More precisely, it was shown in Gaffke (1987), Studden (1989) (for the interval $[-1, 1]$) and in Dette and Studden (1997) (for arbitrary symmetric intervals) that the designs minimizing

$$\phi_p(\xi) = \left(\mathrm{tr} M_{(2)}^{-p}(\xi)\right)^{1/p}, \ -1 < p \leq \infty,$$

is given by the design $\xi_{m,\beta(p)}$ where $\beta(p)$ is the unique solution of the equation

$$\left(\frac{1-\beta}{2}\right)^{p+1} - a^{-2p}\beta = 0$$

in the interval $[0, 1]$.

5 Some illustrative examples

In this section we illustrate the results in a few examples. We restrict ourselves to the problem of discriminating between a constant and the quadratic regression model on the interval $[-1, 1]$. Additionally, we construct robust designs for the situation considered in Example 2.1. Further results for other models are available from the authors.

5.1 Standardized maximin T-optimal discriminating designs for quadratic regression

Consider the problem of discriminating between a constant and a quadratic regression on the interval $\mathcal{X} = [-1, 1]$. As pointed out in Bretz et al. (2005), these models are of importance for detecting dose response signals in phase II clinical trials. If $\mathcal{B} = [-d, d]$, then it it follows from Theorem 4.1 ($m = 2$) that a standardized maximin T-optimal design is given by

$$\xi_M^* = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1-h^*}{2} & h^* & \frac{1-h^*}{2} \end{pmatrix},$$

where h^* is a solution of the problem (4.2). Due to formula (3.9) in Dette et al. (2012) we have

$$\bar{R}(b) = \bar{R}(b,1) = \begin{cases} \frac{1}{4}\left(1+\frac{|b|}{2}\right)^4, & |b| \leq 2, \\ b^2, & |b| \geq 2. \end{cases} \tag{5.1}$$

We define

$$K(h,b) = \frac{h+b^2}{\overline{R}(b)}(1-h)$$

then the solution of the problem

$$\max_{h\in[0,0.5]} \inf_{b\in[-d,d]} K(h,b)$$

can be obtained by straightforward but tedious calculaions, which are omitted for the sake of brevity. For the solution one has to distinguish 3 cases

(1) If $0 < d \leq \frac{1}{2}$ the minimum of the function $K(h,b)$ with respect to the variable b is attained at the boundary of the interval $\mathcal{B} = [-d,d]$ and the optimal value is given by $h^* = (1-d^2)/2$. A typical situation is depicted in the left part of Figure 2. A standardized maximin T-optimal discriminating design has masses $(1+d^2)/4$, $(1-d^2)/2$ and $(1+d^2)/4$ at the points -1, 0 and 1, respectively.

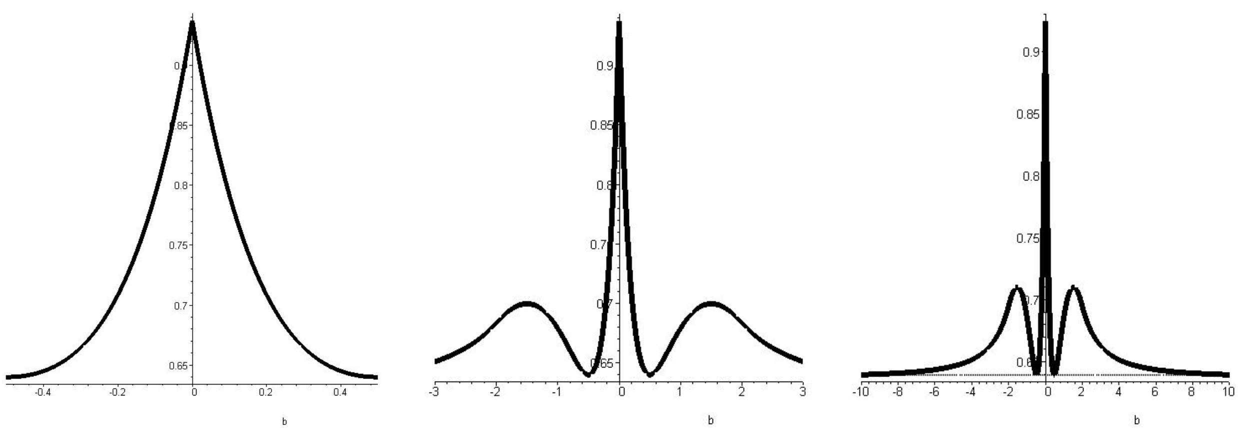

Figure 2: *The behavior of the function $K(h^*,b)$, for different values of d. Left panel $d = 1/2$, middle panel $d = 2$, right panel $d = 10$.*

(2) In the case $1/2 < d \leq 5\sqrt{10}/4$ the solution is given by $h^* = 3/8$, $b^* = 1/2$. Therefore the design with masses 5/16, 3/8 and 5/16 at the points -1, 0 and 1 is a standardized maximin T-optimal discriminating design. The behavior of the function $K(h^*,b)$ in this case is depicted in the middle panel of Figure 2.

(3) In the case $d \in [\frac{5\sqrt{10}}{4}, \infty]$ the structure of the solution changes again. For this interval the optimal pair h^*, b^* is obtained as a solution of the system

$$K(h,b) = K(h,d)\ ,\ \frac{\partial}{\partial b}K(h,b) = 0,$$

and we find by a direct calculation that b^* is the unique root of the equation

$$x^4 + 6x^3 + (-2d^2 + 12)x^2 + (-16d^2 + 8)x + 8d^2 = 0$$

in the interval $[-4 + 2\sqrt{5}, 1/2]$. We have $h^* = b^* - \frac{(b^*)^2}{2}$ and a standardized maximin T-optimal discriminating design has masses $1/2 - b^*/2 + (b^*)^2/4$, $b^* - (b^*)^2/2$, and $1/2 - b^*/2 + (b^*)^2/4$ at the points -1, 0 and 1, respectively. In the limiting case $d = \infty$, that is $\mathcal{B} = \mathbb{R}$, we have $b^* = -4 + 2\sqrt{5}$, $h^* = -22 + 10\sqrt{5}$ and a standardized maximin T-optimal discriminating design has masses $23/2 - 5\sqrt{5}$, $-22 + 10\sqrt{5}$, and $23/2 - 5\sqrt{5}$ at the points -1, 0 and 1, respectively. A typical case for the function $K(h^*, b)$ in this case is depicted in right panel the Figure 2 for $d = 10$.

5.2 Bayesian T-optimal discriminating designs for quadratic regression

For the Bayesian T-optimality criterion a prior has to be chosen and we propose to maximize an average of the efficiencies

$$\int_{-a}^{a} \text{eff}_T(\xi, b) db$$

with respect to the uniform distribution on the interval $[-a, a]$. In the criterion (3.2) this correspond to an absolute continuous prior with density proportional to

$$f(b) = \begin{cases} \frac{3(2+a)^3}{16a(12+6a+a^2)} \frac{1}{\bar{R}(b)}, & a \leq 2 \\ \frac{3a}{17a-6} \frac{1}{\bar{R}(b)}, & a \geq 2. \end{cases},$$

where $\bar{R}(b)$ is defined in (5.1) and the term depending on a is the corresponding normalizing constant. By direct calculations we obtain

$$\int_{-a}^{a} b^2 f(b) db = 2 \int_{0}^{a} b^2 f(b) db = \begin{cases} \frac{4a^2}{12+6a+a^2}, & a \leq 2, \\ \frac{6a^2-4a}{17a-6}, & a \geq 2. \end{cases}$$

In order to apply Theorem 4.1 we consider $\beta_B = \min\{1, \int_{-a}^{a} b^2 f(b) db\}$ and again 3 cases have to be considered.

(1) If $0 < a \leq 2$ we have $\beta_B = \frac{4a^2}{12+6a+a^2}$ and a Bayesian T-optimal discriminating design has masses $\frac{5a^2+6a+12}{4(12+6a+a^2)}$, $\frac{-3a^2+6a+12}{2(12+6a+a^2)}$, and $\frac{5a^2+6a+12}{4(12+6a+a^2)}$ at the points -1, 0 and 1.

(2) If $2 \leq a \leq \frac{7+\sqrt{33}}{4}$ we have $\beta_B = \frac{3a^2-4a}{17a-6}$, and a Bayesian T-optimal discriminating design has masses $\frac{6a^2+13a-6}{4(17a-6)}$, $\frac{-6a^2+21a-6}{2(17a-6)}$, and $\frac{6a^2+13a-6}{4(17a-6)}$ at the points -1, 0 and 1.

(3) If $\frac{7+\sqrt{33}}{4} \leq a$ we have $\beta = 1$ and a Bayesian T-optimal discriminating design has masses 1/2 and 1/2 at the points -1 and 1.

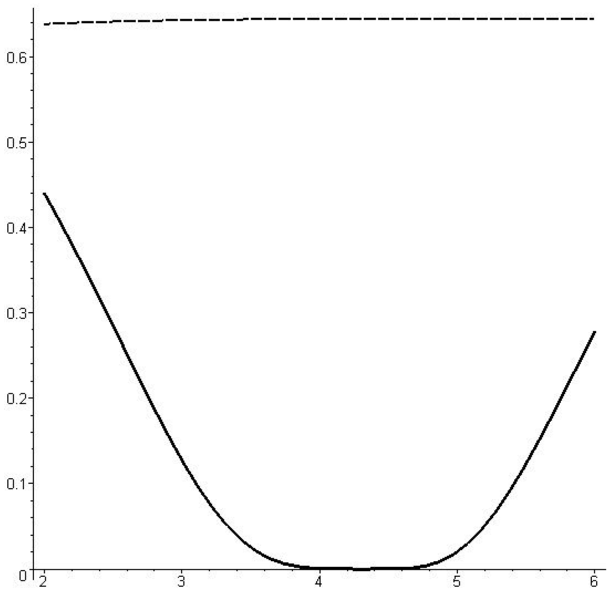

Figure 3: *T-efficiency (2.4) of the standardized maximin optimal discriminating design (dotted line) and the locally T-optimal discriminating design for the Michaelis Menten and Emax model (calculated under the assumption $\theta_{2,0} = -1/4$, $\theta_{2,2} = 1$, solid line). The "true" values are given by $\theta_{2,0} = -1$, $\theta_{2,1} = 1$ and the efficiencies depend on the parameter $\theta_{2,2} \in (2, 6)$.*

5.3 Robust T-optimal discriminating designs for the Michaelis Menten and EMAX model

In this section we briefly illustrate the application of the methodology in the situation described in Example 2.1, where the interest is in designs with good properties for discriminating between the Michaelis Menten and EMAX model. We have calculated the standardized maximin T-optimal discriminating design for the Michaelis Menten and EMAX model, where the region for the parameter $(\theta_{2,0}, \theta_{2,1}, \theta_{2,2})$ is given by $[-1.1, -0.2] \times \{1\} \times [2, 6]$. The corresponding robust design is given by

$$\xi = \begin{pmatrix} 1 & 1.36 & 2 \\ 0.410 & 0.205 & 0.385 \end{pmatrix}$$

As pointed out in Example 2.1, the efficiency of locally T-optimal discriminating designs can be low if some of the parameters of the regression models have been misspecified, and in Figure 3 we compare the performance of the locally and robust optimal discriminating design if the true values are $\theta_{2,0} = -1$, $\theta_{2,1} = 1$ and $\theta_{2,2} \in (2, 6)$. We observe a substantial improvement by the standardized maximin T-optimal discriminating design. Other scenarios showed a similar picture and are not displayed for the sake of brevity.

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 C2) of the German Research Foundation (DFG). The work of V. Melas and P. Shpilev was also additionally supported by Russian Foundation for Basic Research (project number 12-01-00747). This paper was initiated at the Isaac Newton Institute for Mathematical Sciences in Cambridge, England, during the 2011 programme on the Design and Analysis of Experiments.

References

Asprey, S. P. and Macchietto, S. (2000). Statistical tools for optimal dynamic model building. *Computers and Chemical Engineering*, 24:1261–1267.

Atkinson, A. C. (2010). The non-uniqueness of some designs for discriminating between two polynomial models in one variablel. *MODA 9, Advances in Model-Oriented Design and Analysis*, pages 9–16.

Atkinson, A. C., Bogacka, B., and Bogacki, M. B. (1998). D- and T-optimum designs for the kinetics of a reversible chemical reaction. *Chemometrics and Intelligent Laboratory Systems*, 43:185–198.

Atkinson, A. C. and Fedorov, V. V. (1975a). The designs of experiments for discriminating between two rival models. *Biometrika*, 62:57–70.

Atkinson, A. C. and Fedorov, V. V. (1975b). Optimal design: experiments for discriminating between several models. *Biometrika*, 62:289–303.

Bretz, F., Pinheiro, J. C., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*, 61(3):738–748.

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.

Cornish-Bowden, A. (1965). *Fundamentals of Enzyme Kinetics (Rev. Ed.)*. Portland Press, London.

Danesi, R., Innocenti, F., Fogli, S., Gennari, A., Baldini, E., Di Paolo, A., Salvadori, B., Bocci, G., Conte, P. F., and Del Tacca, M. (2002). Pharmacokinetics and pharmacodynamics of combination chemotherapy with paclitaxel and epirubicin in breast cancer patients. *Br J Clin Pharmacol*, 53(5):508–518.

Dette, H. (1994). Discrimination designs for polynomial regression on a compact interval. *Annals of Statistics*, 22:890–904.

Dette, H. (1995). Optimal designs for identifying the degree of a polynomial regression. *Annals of Statistics*, 23:1248–1267.

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 Haller, G. (1998). Optimal designs for the identification of the order of a Fourier regression. *Annals of Statistics*, 26:1496–1521.

Dette, H., Melas, V. B., and Shpilev, P. (2012). T-optimal designs for discrimination between two polynomial models. *Annals of Statistics*, 40(1):188–205.

Dette, H. and Studden, W. J. (1997). *Canonical Moments with Applications in Statistics, Probability and Analysis*. Wiley and Sons, New York.

Dette, H. and Titoff, S. (2009). Optimal discrimination designs. *Annals of Statistics*, 37(4):2056–2082.

Foo, L. K. and Duffull, S. (2011). Optimal design of pharmacokinetic-pharmacodynamic studies. In *Pharmacokinetics in Drug Development, Advances and Applications*. Springer.

Gaffke, N. (1987). Further characterizations of design optimality and admissibility for partial parameter estimation on linear regression. *Annals of Statistics*, 15(3):942–957.

Hill, P. D. (1978). A review of experimental design procedures for regression model discrimination. *Technometrics*, 20(1):15–21.

Hunter, W. G. and Reiner, A. M. (1965). Designs for discriminating between two rival models. *Technometrics*, 7(3):307–323.

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.

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

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT Press, Massachusetts.

Song, D. and Wong, W. K. (1999). On the construction of g_{rm}-optimal designs. *Statistica Sinica*, 9:263–272.

Spruill, M. C. (1990). Good designs for testing the degree of a polynomial mean. *Sankhya, Ser. B*, 52(1):67–74.

Stigler, S. (1971). Optimal experimental design for polynomial regression. *Journal of the American Statistical Association*, 66:311–318.

Studden, W. J. (1982). Some robust-type D-optimal designs in polynomial regression. *Journal of the American Statistical Association*, 77(380):916–921.

Studden, W. J. (1989). Note of some ϕ_p-optimal designs for polynomial regressions. *Annals of Statistics*, 17(2):618–623.

Szegö, G. (1959). *Orthogonal Polynomials*. American Mathematical Society Colloquium Publications, Vol. 23. Revised ed. American Mathematical Society, Providence, R.I.

Tommasi, C. and López-Fidalgo, J. (2010). Bayesian optimum designs for discriminating between models with any distribution. *Computational Statistics & Data Analysis*, 54(1):143–150.

Ucinski, D. and Bogacka, B. (2005). T-optimum designs for discrimination between two multiresponse dynamic models. *Journal of the Royal Statistical Society, Ser. B*, 67:3–18.

Waterhouse, T. H., Woods, D. C., Eccleston, J. A., and Lewis, S. M. (2008). Design selection criteria for discrimination/estimation for nested models and a binomial response. *Journal of Statistical Planning and Inference*, 138:132–144.

Wiens, D. P. (2009). Robust discrimination designs, with Matlab code. *Journal of the Royal Statistical Society, Ser. B*, 71:805–829.

Wiens, D. P. (2010). Robustness of design for the testing of lack of fit and for estimation in binary response models. *Computational Statistics and Data Analysis*, 54:3371–3378.