The Annals of Statistics
2015, Vol. 43, No. 5, 1959–1985
DOI: 10.1214/15-AOS1333

BAYESIAN T-OPTIMAL DISCRIMINATING DESIGNS

BY HOLGER DETTE[*,1,2], VIATCHESLAV B. MELAS[†,1,3] AND ROMAN GUCHENKO[†]

Ruhr University at Bochum[*] and St. Petersburg State University[†]

The problem of constructing Bayesian optimal discriminating designs for a class of regression models with respect to the T-optimality criterion introduced by Atkinson and Fedorov [*Biometrika* **62** (1975a) 57–70] is considered. It is demonstrated that the discretization of the integral with respect to the prior distribution leads to locally T-optimal discriminating design problems with a large number of model comparisons. Current methodology for the numerical construction of discrimination designs can only deal with a few comparisons, but the discretization of the Bayesian prior easily yields to discrimination design problems for more than 100 competing models. A new efficient method is developed to deal with problems of this type. It combines some features of the classical exchange type algorithm with the gradient methods. Convergence is proved, and it is demonstrated that the new method can find Bayesian optimal discriminating designs in situations where all currently available procedures fail.

1. Introduction. Optimal design theory provides useful tools to improve the accuracy of statistical inference without any additional costs by carefully planning experiments before they are conducted. Numerous authors have worked on the construction of optimal designs in various situations. For many models, optimal designs have been developed explicitly [see the monographs of Atkinson, Donev and Tobias (2007), Pukelsheim (2006)], and several algorithms have been developed for their numerical construction if the optimal designs are not available in explicit form; see Yang, Biedermann and Tang (2013), Yu (2010), among others. On the other hand the construction of such designs depends sensitively on the model assumptions, and an optimal design for a particular model might be inefficient if it is used in a different model. Moreover, in many experiments it is often not obvious which model should be finally fitted to the data, and model building

Received December 2014; revised February 2015.

[1]Supported by the Deutsche Forschungsgemeinschaft (SFB 823: Statistik nichtlinearer dynamischer Prozesse, Teilprojekt C2).

[2]Supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639.

[3]Supported in part by St. Petersburg State University ("Actual problems of design and analysis for regression models," 6.38.435.2015).

MSC2010 subject classifications. 62K05.

Key words and phrases. Design of experiment, Bayesian optimal design, model discrimination, gradient methods, model uncertainty.

is an important part of data analysis. A typical and very important example are Phase II dose-finding studies, where various nonlinear regression models of the form

$$Y = \eta(x, \theta) + \varepsilon, \tag{1.1}$$

have been developed for describing the dose–response relation [see Pinheiro, Bretz and Branson (2006)], but the problem of model uncertainty arises in nearly any other statistical application. As a consequence, the construction of efficient designs for model identification has become an important field in optimal design theory. Early work can be found in Stigler (1971), who determined designs for discriminating between two nested univariate polynomials by minimizing the volume of the confidence ellipsoid for the parameters corresponding to the extension of the smaller model. Several authors have worked on this approach in various other classes of nested models; see, for example, Dette and Haller (1998) or Song and Wong (1999), among others.

A different approach to the problem of constructing optimal designs for model discrimination is given in a pioneering paper by Atkinson and Fedorov (1975a), who proposed the T-optimality criterion to construct designs for discriminating between two competing regression models. Roughly speaking their approach provides a design such that the sum of squares for a lack of fit test is large. Atkinson and Fedorov (1975b) extended this method for discriminating a selected model η_1 from a class of other regression models, say $\{\eta_2, \ldots, \eta_k\}$, $k \geq 2$. In contrast to the work Stigler (1971) and his followers, the T-optimality criterion does not require competing nested models and has found considerable attention in the statistical literature with numerous applications, including such important fields as chemistry and pharmacokinetics; see, for example, Atkinson (2008), Atkinson, Bogacka and Bogacki (1998), López-Fidalgo, Tommasi and Trandafir (2007), Ponce de Leon and Atkinson (1991), Tommasi (2009), Uciński and Bogacka (2005) or Foo and Duffull (2011) for some more recent references. A drawback of the T-optimality criterion consists of the fact that—even in the case of linear models—the criterion depends on the parameters of the model η_1. This means that T-optimality is a local optimality criterion in the sense of Chernoff (1953), and that it requires some preliminary knowledge regarding the parameters. Consequently, most of the cited papers refer to locally T-optimal designs. Although there exist applications where such information is available [e.g., in the analysis of dose–response studies as considered in Pinheiro, Bretz and Branson (2006)], in most situations such knowledge can be rarely provided. Several authors have introduced robust versions of the classical optimality criteria such as Bayesian or minimax D-optimality criteria in order to determine efficient designs for model discrimination, which are less sensitive with respect to the choice of parameters; see Chaloner and Verdinelli (1995), Dette (1997), Pronzato and Walter (1985). The robustness problem of the T-optimality criterion has been already mentioned in Atkinson and Fedorov (1975a), who proposed a Bayesian approach to address the problem of parameter

uncertainty in the T-optimality criterion. Wiens (2009) imposed (linear) neighborhood structures on each regression response and determined least favorable points in these neighborhoods in order to robustify the locally T-optimal design problem. Dette, Melas and Shpilev (2012) considered polynomial regression models and determined explicitly Bayesian T-optimal discriminating designs for the criterion introduced by Atkinson and Fedorov (1975a). Their results indicate the difficulties arising in Bayesian T-optimal design problems.

The scarcity of literature on Bayesian T-optimal discriminating designs can be explained by the fact that in nearly all cases of practical interest, these designs have to be found numerically, and even this is a very hard problem. These numerical difficulties become apparent even in the case of locally T-optimal designs. Atkinson and Fedorov (1975a) proposed an exchange type algorithm, which has a rather slow rate of convergence and has been used by several authors. Braess and Dette (2013) pointed out that besides its slow convergence, this algorithm does not yield an accurate solution of the optimal discriminating design problem if more than 5 model comparisons are under consideration. These authors developed a more efficient algorithm for the determination of locally T-optimal discriminating designs for several competing regression models by exploring relations between optimal design problems and (nonlinear) vector-valued approximation theory. Although the resulting algorithm provides a substantial improvement of the exchange type methods, it cannot deal with Bayesian optimality criteria in general, and the development of an efficient procedure for this purpose is a very challenging and open problem.

The goal of the present paper is to fill this gap. We utilize the fact that in applications, the integral with respect to the prior distribution has to be determined by a discrete approximation, and we show that the discrete Bayesian T-optimal design problem is a special case of the local T-optimality criterion for a very large number of competing models, as considered in Braess and Dette (2013). The competing models arise from the different support points used for the approximation of the prior distribution by a discrete measure, and the number of model comparisons in the resulting criterion easily exceeds the 200. Therefore, the algorithm in Braess and Dette (2013) does not provide an accurate solution of the corresponding optimization problem, and we propose a new method for the numerical construction of Bayesian T-optimal designs with substantial computational advantages. Roughly speaking, the support points of the design in each iteration are determined in a similar manner, as proposed in Atkinson and Fedorov (1975a), but for the calculation of the corresponding weights, we use quadratic programming or a gradient approach. It turns out that the new procedure is extremely efficient and is able to find Bayesian T-optimal designs with a few number of iterations.

The rest of this paper is organized as follows. In Section 2 we give an introduction into the problem of designing experiments for discriminating between competing regression models and also derive some basic properties of locally T-optimal discriminating designs. In particular we show how the Bayesian T-optimal

design problem is related to a local one with a large number of model comparisons; see Section 2.2. Section 3 is devoted to the construction of new numerical procedures (in particular, Algorithm 3.2), for which we prove convergence to a T-optimal discriminating design. Our approach consists of two steps, consecutively optimizing with respect to the support points (step 1) and weights of the design (step 2). For the second step, we also discuss two procedures to speed up the convergence of the algorithm. The results are illustrated in Section 4, calculating several Bayesian T-optimal discriminating designs in examples, where all other available procedures do not provide an accurate numerical solution of the optimal design problem. For example, the new procedure is able to solve locally T-optimal designs with more than 240 model comparisons, as they arise frequently in Bayesian T-optimal design problems. In particular, we illustrate the methodology of calculating Bayesian T-optimal discriminating designs for a dose finding clinical trial, which has recently been discussed in Pinheiro, Bretz and Branson (2006). The corresponding `R`-package is provided in the CRAN library; see http://cran.r-project.org/web/packages/rodd/index.html. Finally all proofs are deferred to the Appendix.

2. T-optimal discriminating designs. Consider the regression model (1.1), where x belongs to some compact set $\mathcal{X}$ and observations at different experimental conditions are independent. For the sake of transparency and a clear representation, we assume that the error ε is normally distributed. The methodology developed in the following discussion can be extended to more general error structures following the line of research in López-Fidalgo, Tommasi and Trandafir (2007), but for the sake of brevity, details are deferred to future research.

Throughout this paper we consider the situation where ν different models, say

$$\eta_i(x, \theta_i), \qquad i = 1, \ldots, \nu, \tag{2.1}$$

are available to describe the dependency of Y on the predictor x. In (2.1) the quantity θ_i denotes a d_i-dimensional parameter, which varies in a compact space, say Θ_i $(i = 1, \ldots, \nu)$. Following Kiefer (1974) we consider approximate designs that are defined as probability measures, say ξ, with finite support. The support points $x_1, \ldots, x_k$ of a design ξ give the locations where observations are taken, while the weights $\omega_1, \ldots, \omega_k$ describe the relative proportions of observations at these points. If an approximate design is given and n observations can be taken, a rounding procedure is applied to obtain integers n_i $(i = 1, \ldots, k)$ from the not necessarily integer valued quantities $\omega_i n$ such that $\sum_{i=1}^k n_i = n$. We are interested in designing an experiment, such that a most appropriate model can be chosen from the given class $\{\eta_1, \ldots, \eta_\nu\}$ of all competing models.

2.1. *T-optimal designs.* In the case of $\nu = 2$ competing models Atkinson and Fedorov (1975a) proposed to fix one model, say $\eta_1(\cdot, \theta_1)$, with corresponding parameter $\overline{\theta}_1$ and to maximize the function

$$T_{12}(\xi) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \big[\eta_1(x, \overline{\theta}_1) - \eta_2(x, \theta_2)\big]^2 \xi(dx) \tag{2.2}$$

in the class of all (approximate) designs. Roughly speaking, these designs maximize the power of the test of the hypothesis η_1 versus η_2. Note that the resulting optimal design depends on the parameter $\overline{\theta}_1$ for the first model, which has to be fixed by the experimenter. This means that these designs are local in the sense of Chernoff (1953). It was pointed out by Dette, Melas and Shpilev (2013) that locally T-optimal designs may be very sensitive with respect to misspecification of $\overline{\theta}_1$. In a further paper Atkinson and Fedorov (1975b) generalized their approach to construct optimal discriminating designs for more than 2 competing regression models and suggested the criterion

$$T(\xi) = \min_{2\le j\le \nu} T_{1j}(\xi) = \min_{2\le j\le \nu} \inf_{\theta_j\in\Theta_j} \int_{\mathcal{X}} \big[\eta_1(x,\overline{\theta}_1) - \eta_j(x,\theta_j)\big]^2 \xi(dx). \tag{2.3}$$

This criterion determines a "good" design for discriminating the model η_1 against $\eta_2,\dots,\eta_\nu$, where the parameter $\overline{\theta}_1$ has the same meaning as before. As pointed out by Tommasi and López-Fidalgo (2010) and Braess and Dette (2013), there are many situations where it is not clear which model should be considered as fixed, and these authors proposed a symmetrized Bayesian (instead of minimax) version of the T-optimality criterion, that is,

$$\begin{aligned} T_{\mathrm{P}}(\xi) &= \sum_{i,j=1}^{\nu} p_{i,j} T_{i,j}(\xi) \\ &= \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j}\in\Theta_j} \int_{\mathcal{X}} \big[\eta_i(x,\overline{\theta}_i) - \eta_j(x,\theta_{i,j})\big]^2 \xi(dx), \end{aligned} \tag{2.4}$$

where the quantities $p_{i,j}$ denote nonnegative weights reflecting the importance of the comparison between the models η_i and η_j. We note again that this criterion requires the specification of the parameter $\overline{\theta}_i$, whenever the corresponding weight $p_{i,j}$ is positive. Throughout this paper we will call a design maximizing one of the criteria (2.2)–(2.4) locally T-optimal discriminating design, where the specific criterion under consideration is always clear from the context. For some recent references discussing locally T-optimal discriminating designs, we refer to Atkinson (2008), López-Fidalgo, Tommasi and Trandafir (2007), Tommasi (2009), Uciński and Bogacka (2005) or Braess and Dette (2013), among many others. For the formulation of the first result we require the following assumptions, which are sufficient for the convergence results established in this paper.

ASSUMPTION 2.1. For each $i = 1,\dots,\nu$, the functions $\eta_i(\cdot,\theta_i)$ are continuously differentiable with respect to the parameter $\theta_i \in \Theta_i$.

For a design ξ we also introduce the notation

$$\Theta^*_{i,j}(\xi) = \operatorname*{arg\,inf}_{\theta_{i,j}\in\Theta_j} \int_{\mathcal{X}} \big[\eta_i(x,\overline{\theta}_i) - \eta_j(x,\theta_{i,j})\big]^2 \xi(dx), \tag{2.5}$$

which is used in the formulation of the following result.

THEOREM 2.1. *If Assumption 2.1 is satisfied, then the design ξ^* is a locally T_P-optimal discriminating design if and only if there exist distributions μ_{ij}^* on the sets $\Theta_{i,j}^*(\xi^*)$, defined in (2.5), such that the inequality*

$$\sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_{i,j}^*(\xi^*)} \big[\eta_i(x,\overline{\theta}_i) - \eta_j(x,\theta_{i,j})\big]^2 \mu_{ij}^*(d\theta_{i,j}) \le T_P(\xi^*) \tag{2.6}$$

is satisfied for all $x \in \mathcal{X}$. Moreover, there is equality in (2.6) for all support points of the locally T_P-optimal discriminating design ξ^*.

Theorem 2.1 provides an extension of the corresponding theorem in Dette and Titoff (2009), and the proof is similar and therefore omitted. For designs ξ, ζ on $\mathcal{X}$ we next introduce the function

$$Q(\zeta,\xi) = \int_{\mathcal{X}} \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j} \in \Theta_{ij}^*(\xi)} \big[\eta_i(x,\overline{\theta}_i) - \eta_j(x,\theta_{i,j})\big]^2 \zeta(dx), \tag{2.7}$$

which plays an important role in the subsequent discussion. Using Lemma A.1 from the Appendix it is easy to check [see, e.g., Atkinson and Fedorov (1975a)] that

$$\left.\frac{\partial T_P(\xi(\alpha))}{\partial \alpha}\right|_{\alpha=0} = Q(\zeta,\xi) - T_P(\xi),$$

where $\xi(\alpha) = (1-\alpha)\xi + \alpha\zeta$ denotes the convex combination of the designs ξ and ζ. In the following discussion we need an extension of Theorem 2.1.

THEOREM 2.2. *If Assumption 2.1 is satisfied and the design ξ is not T_P-optimal, then there exists a design ζ^*, such that the inequality $Q(\zeta^*,\xi) > T_P(\xi)$ holds.*

In order to obtain a more manageable condition of this result, let $\hat{\mu}_{i,j}(\xi)$ denote a measure on the set $\Theta_{i,j}^*(\xi)$ $(i,j = 1,\dots,\nu)$ for which the function

$$\max_{x\in\mathcal{X}} \sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_{i,j}^*(\xi)} \big[\eta_i(x,\overline{\theta}_i) - \eta_j(x,\theta_{i,j})\big]^2 \mu_{i,j}(d\theta_{i,j})$$

attains its minimal value, and define

$$\Psi(x,\xi) = \sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_{i,j}^*(\xi)} \big[\eta_i(x,\overline{\theta}_i) - \eta_j(x,\theta_{i,j})\big]^2 \hat{\mu}_{ij}(d\theta_{i,j}). \tag{2.8}$$

COROLLARY 2.3. *If Assumption 2.1 is satisfied and the design ξ is not T_P-optimal, then there exists a point $\overline{x} \in \mathcal{X}$ such that*

$$\Psi(\overline{x},\xi) > T_P(\xi).$$

ASSUMPTION 2.2. For any design ξ such that $T_{\mathrm{P}}(\xi) > 0$ and weight $p_{i,j} \neq 0$, the infima in (2.4) are attained at a unique point $\widehat{\theta}_{i,j} = \widehat{\theta}_{i,j}(\xi)$ in the interior of the set Θ_j.

If Assumptions 2.1 and 2.2 are satisfied, the function Q in (2.7) simplifies to

$$Q(\zeta, \xi) = \int_{\mathcal{X}} \sum_{i,j=1}^{\nu} p_{i,j} \big[\eta_i(x, \overline{\theta}_i) - \eta_j(x, \widehat{\theta}_{i,j})\big]^2 \zeta(dx).$$

Moreover, the function Ψ defined in (2.8) simplifies to

$$\Psi(x, \xi) = \sum_{i,j=1}^{\nu} p_{i,j} \big[\eta_i(x, \overline{\theta}_i) - \eta_j(x, \widehat{\theta}_{i,j})\big]^2.$$

2.2. *Bayesian T-optimal designs.* As pointed out by Dette, Melas and Shpilev (2012) locally T-optimal designs are rather sensitive with respect to misspecification of the unknown parameters $\overline{\theta}_i$, and it might be appropriate to construct more robust designs for model discrimination. The problem of robustness was already mentioned in Atkinson and Fedorov (1975a), and these authors proposed a Bayesian version of the T-optimality criterion which reads in the situation of the criterion (2.4) as follows:

$$T_{\mathrm{P}}^{\mathrm{B}}(\xi) = \sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_i} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} \big[\eta_i(x, \lambda_i) - \eta_j(x, \theta_{i,j})\big]^2 \xi(dx) \mathcal{P}_i(d\lambda_i). \tag{2.9}$$

Here for each $i = 1, \ldots, \nu$, the measure $\mathcal{P}_i$ denotes a prior distribution for the parameter θ_i in model η_i, such that all integrals in (2.9) are well defined. Throughout this paper we will call any design maximizing the criterion (2.9) a Bayesian T-optimal discriminating design. For (two) polynomial regression models, Bayesian T-optimal discriminating designs have been explicitly determined by Dette, Melas and Shpilev (2013), and their results indicate the intrinsic difficulties in the construction of optimal designs with respect to this criterion.

In the following discussion we will link the criterion (2.9) to the locally T-optimality criterion (2.4) for a large number of competing models. For this purpose we note that in nearly all situations of practical interest, an explicit evaluation of the integral in (2.9) is not possible, and the criterion has to be evaluated by numerical integration approximating the prior distribution by a measure with finite support. Therefore we assume that the prior distribution $\mathcal{P}_i$ in the criterion is given by a discrete measure with masses $\tau_{i1}, \ldots, \tau_{i\ell_i}$ at the points $\lambda_{i1}, \ldots, \lambda_{i\ell_i}$. The criterion in (2.9) can then be rewritten as

$$T_{\mathrm{P}}^{\mathrm{B}}(\xi) = \sum_{i,j=1}^{\nu} \sum_{k=1}^{\ell_i} p_{i,j} \tau_{ik} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} \big[\eta_i(x, \lambda_{ik}) - \eta_j(x, \theta_{i,j})\big]^2 \xi(dx), \tag{2.10}$$

which is a locally T-optimality criterion of the form (2.4). The choice of the weights $\tau_{i,k}$ and points $\lambda_{i,k}$ depends on the particular models under consideration, where the main goal is to provide a good approximation of the integrals in (2.9) with a small number ℓ_i and a "good" choice of points and weights. Quadrature formulas are recommended for this purpose; see Abramowitz and Stegen (1965).

Note that the only difference between the criterion obtained from the Bayesian approach and (2.4) consists of the fact that the criterion (2.10) involves substantially more comparisons of the functions η_i and η_j (even if a "good" rule for numerical integration has been used). Because for each support point of the prior distribution in the criterion (2.10), the infimum has to be calculated numerically, the optimization of the criterion (2.10) is computationally expensive. Consequently, the computation of Bayesian T-optimal discriminating design problems is particularly challenging. In the following sections we provide an efficient solution to this problem.

3. Calculating locally T-optimal designs. Braess and Dette (2013) proposed an algorithm for the numerical construction of locally T-optimal designs, which is based on vector-valued Chebyshev approximation. This algorithm is quite difficult both in terms of description and implementation. Moreover, it requires substantial computational resources and is therefore only able to deal with a small number of comparisons in the T-optimality criterion. The purpose of this section is to develop a more efficient method which is able to deal with a large number of comparisons in the criterion and avoids the drawbacks of the procedures in Atkinson and Fedorov (1975a) and Braess and Dette (2013).

Recall the definition of the function Ψ in (2.8), and note that under Assumption 2.1 it follows from Corollary 2.3 that there exists a point $\overline{x} \in \mathcal{X}$, such that the inequality

$$\Psi(\overline{x}, \xi) > T_{\mathrm{P}}(\xi)$$

holds, whenever ξ is *not* a locally T-optimal discriminating design. The algorithm of Atkinson and Fedorov (1975a) uses this property to construct a sequence of designs which converges to the locally T-optimal discriminating design. For further reference, it is stated here.

ALGORITHM 3.1 [Atkinson and Fedorov (1975a)]. Let ξ_0 denote a given (starting) design, and let $(\alpha_s)_{s=0}^{\infty}$ be a sequence of positive numbers, such that $\lim_{s\to\infty} \alpha_s = 0$, $\sum_{s=0}^{\infty} \alpha_s = \infty$, $\sum_{s=0}^{\infty} \alpha_s^2 < \infty$. For $s = 0, 1, \ldots$ define

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi(x_{s+1}),$$

where $\xi(x_{s+1})$ denotes the Dirac measure at the point

$$x_{s+1} = \arg\max_{x \in \mathcal{X}} \Psi(x, \xi_s).$$

It can be shown that this algorithm converges in the sense that

$$\lim_{s\to\infty} T_{\mathrm{P}}(\xi_s) = T_{\mathrm{P}}\big(\xi^*\big),$$

where ξ^* denotes a locally T-optimal discriminating design; see also Aletti, May and Tommasi (2013), Aletti, May and Tommasi (2014). However, a major problem of Algorithm 3.1 is that it yields a sequence of designs with an increasing number of support points. As a consequence, the resulting design (after applying some stopping criterion) is concentrated on a large set of points. Even if this problem can be solved by clustering or by determining the extrema of the final function $\Psi(x, \xi_s)$, it is much more difficult to deal with the accumulation of support points during the iteration. Moreover, Braess and Dette (2013) demonstrated that in many cases the iteration process may take several hundred iterations to obtain a locally T-optimal discriminating design with a required precision, resulting in a high computational complexity for the recalculation of the optimum values

$$\widehat{\theta}_{i,j} \in \operatorname*{arg\,inf}_{\theta_{i,j}\in\Theta^*_{i,j}(\xi)} \int_{\mathcal{X}} \big[\eta_i(x, \overline{\theta}_i) - \eta_j(x, \theta_{i,j})\big]^2 \xi(dx) \tag{3.1}$$

in the optimality criterion (2.4). These authors also showed that Algorithm 3.1 may not find the optimal design if there are too many model comparisons involved in the T-optimality criterion (2.4).

Therefore, we propose the following basic procedure for the calculation of locally T-optimal discriminating designs as an alternative to Algorithm 3.1. Roughly speaking, it consists of two steps treating the maximization with respect to support points (step 1) and weights (step 2) separately, where two methods implementing the second step will be given below; see Sections 3.1 and 3.2 for details.

ALGORITHM 3.2. Let ξ_0 denote a starting design such that $T_{\mathrm{P}}(\xi_0) > 0$, and define recursively a sequence of designs $(\xi_s)_{s=0,1,\ldots}$ as follows:

(1) Let $\mathcal{S}_{[s]}$ denote the support of the design ξ_s. Determine the set $\mathcal{E}_{[s]}$ of all local maxima of the function $\Psi(x, \xi_s)$ on the design space $\mathcal{X}$, and define $\mathcal{S}_{[s+1]} = \mathcal{S}_{[s]} \cup \mathcal{E}_{[s]}$.

(2) We define $\xi = \{\mathcal{S}_{[s+1]}, \omega\}$ as the design supported at $\mathcal{S}_{[s+1]}$ (with a vector w of weights) and determine the locally T_{P}-optimal design in the class of all designs supported at $\mathcal{S}_{[s+1]}$; that is, we determine the vector $\omega_{[s+1]}$ maximizing the function

$$\begin{aligned} g(\omega) &= T_{\mathrm{P}}\big(\{\mathcal{S}_{[s+1]}, \omega\}\big) \\ &= \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j}\in\Theta_j} \sum_{x\in\mathcal{S}_{[s+1]}} \big[\eta_i(x, \overline{\theta}_i) - \eta_j(x, \theta_{i,j})\big]^2 w_x. \end{aligned}$$

(Here w_x denotes the weights at the point $x \in \mathcal{S}_{s+1}$.) All points in $\mathcal{S}_{[s+1]}$ with vanishing components in the vector of weights $\omega_{[s+1]}$ will be removed, and the new set of support points will also be denoted by $\mathcal{S}_{[s+1]}$. Finally the design ξ_{s+1} is

defined as the design with the set of support points $\mathcal{S}_{[s+1]}$ and the corresponding nonzero weights.

THEOREM 3.3. *Let Assumption 2.1 be satisfied, and let* $(\xi_s)_{s=0,1,\ldots}$ *denote the sequence of designs obtained by Algorithm 3.2, then*

$$\lim_{s\to\infty} T_P(\xi_{s+1}) = T_P(\xi^*),$$

where ξ^* *denotes a locally* T*-optimal discriminating design.*

A proof of Theorem 3.3 is deferred to the Appendix. Note that the algorithm adds all local maxima of the function $\Psi(x, \xi_s)$ as possible support points of the design in the next iteration. Consequently, in its current form, Algorithm 3.2 also accumulates too many support points. To avoid this problem, it is suggested to remove at each step those points from the support, whenever their weight is smaller than $m^{0.25}$, where m denote the working precision of the software used in the implementation (which is 2.2×10^{-16} for R). Note also that this refinement does not affect the convergence of the algorithm from a practical point of view. A more important question is how to implement the second step of the procedure, that is, the maximization of function $g(\omega)$. Before we discuss two computationally efficient procedures for this purpose in the following sections, we state an important property of the function $\Psi(x, \xi_{s+1})$ obtained in each iteration.

LEMMA 3.4. *Let Assumptions 2.1 and 2.2 be satisfied. At the end of each iteration of Algorithm 3.2, the function* $\Psi(x, \xi_{s+1})$ *attains one and the same value for all support points of the design* ξ_{s+1}.

3.1. *Quadratic programming.* Let $\mathcal{S}_{[s+1]} = \{x_1, \ldots, x_n\}$ denote the set obtained in the first step of Algorithm 3.2, and define ξ as a design supported at $\mathcal{S}_{[s+1]}$ with corresponding weights $\omega_1, \ldots, \omega_n$, which have to be determined in step 2 of the algorithm by maximizing the function

$$g(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \sum_{k=1}^{n} \omega_k \big[\eta_i(x_k, \overline{\theta}_i) - \eta_j(x_k, \widehat{\theta}_{i,j})\big]^2, \tag{3.2}$$

where $\widehat{\theta}_{i,j} = \widehat{\theta}_{i,j}(\omega) \in \Theta^*_{i,j}(\xi)$ is defined in (3.1). For this purpose we suggest the linearization of the functions $\eta_j(x_k, \theta_{i,j})$ in the neighborhood of point $\widehat{\theta}_{i,j}$. More precisely, we consider the function

$$\begin{aligned}\overline{g}(\omega) &= \sum_{i,j=1}^{\nu} p_{i,j} \min_{\alpha_{i,j}\in\mathbb{R}^{d_j}} \sum_{k=1}^{n} \omega_k \Bigg[\eta_i(x_k, \overline{\theta}_i) - \eta_j(x_k, \widehat{\theta}_{i,j}) \\ &\qquad - \alpha_{i,j}^T \frac{\partial \eta_j(x_k, \theta_{i,j})}{\partial \theta_{i,j}}\bigg|_{\theta_{i,j}=\widehat{\theta}_{i,j}}\Bigg]^2 \\ &= \sum_{i,j=1}^{\nu} p_{i,j} \min_{\alpha_{i,j}\in\mathbb{R}^{d_j}} \big[\alpha_{i,j}^{\mathrm{T}} \mathbf{J}_{i,j}^{\mathrm{T}} \mathbf{\Omega} \mathbf{J}_{i,j} \alpha_{i,j} - 2\omega^{\mathrm{T}} \mathbf{R}_{i,j} \alpha_{i,j} + b_{i,j}^{\mathrm{T}} \omega\big],\end{aligned}$$

where d_j is the dimension of the parameter space Θ_j, $\mathbf{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_n)$ and the matrices $\mathbf{J}_{i,j} \in \mathbb{R}^{n \times d_j}$, $\mathbf{R}_{i,j} \in \mathbb{R}^{n \times d_j}$ and the vectors $b_{i,j} \in \mathbb{R}^n$ are defined by

$$\mathbf{J}_{i,j} = \left(\frac{\partial \eta_j(x_r, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j} = \widehat{\theta}_{i,j}} \right)_{r=1,\ldots,n},$$

$$\mathbf{R}_{i,j} = \left(\left[\eta_i(x_r, \overline{\theta}_i) - \eta_j(x_r, \widehat{\theta}_{i,j}) \right] \frac{\partial \eta_j(x_r, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j} = \widehat{\theta}_{i,j}} \right)_{r=1,\ldots,n},$$

$$b_{i,j} = \left(\left[\eta_i(x_r, \overline{\theta}_i) - \eta_j(x_r, \widehat{\theta}_{i,j}) \right]^2 \right)_{r=1,\ldots,n},$$

respectively. Obviously the minimum with respect to $\alpha_{i,j}$ is achieved by $\alpha_{i,j} = (\mathbf{J}_{i,j}^{\mathrm{T}} \mathbf{\Omega} \mathbf{J}_{i,j})^{-1} \mathbf{R}_{i,j}^{\mathrm{T}} \omega$ which gives

$$\overline{g}(\omega) = -\omega^{\mathrm{T}} \mathbf{Q}(\omega) \omega + b^{\mathrm{T}} \omega,$$

where $b = \sum_{i,j=1}^{\nu} p_{i,j} b_{i,j}$ and

$$\mathbf{Q}(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \mathbf{R}_{i,j} \left(\mathbf{J}_{i,j}^{\mathrm{T}} \mathbf{\Omega} \mathbf{J}_{i,j} \right)^{-1} \mathbf{R}_{i,j}^{\mathrm{T}}.$$

The matrix $\mathbf{Q}(\omega)$ depends on ω, but if we ignore this dependence and take the matrix $\mathbf{\Omega} = \operatorname{diag}(\overline{\omega}_1, \ldots, \overline{\omega}_n)$ as fixed, then we end up with a quadratic programming problem, that is,

$$\begin{aligned} &\phi(\omega, \overline{\omega}) = -\omega^{\mathrm{T}} \mathbf{Q}(\overline{\omega}) \omega + b^{\mathrm{T}} \omega \to \max_{\omega}, \\ &\sum_{k=1}^{n} \omega_k = 1; \qquad \omega_k \geq 0, k = 1, \ldots, n. \end{aligned} \tag{3.3}$$

This problem is solved iteratively until convergence, substituting each time the solution obtained in the previous iteration instead of $\overline{\omega}$. We note that a similar idea has also been proposed by Braess and Dette (2013).

REMARK 3.5. (1) Note that the calculation of the optimal values in (3.1) is computationally easier for Algorithm 3.2 because the number of support points of the designs calculated during the iterations is substantially smaller than for Algorithm 3.1.

(2) The calculation of the optimal weights as proposed in Section 3.1 could also be combined with Algorithm 3.1 to produce an acceleration of the algorithm. We have implemented Algorithm 3.1 with this additional step and found that Algorithm 3.1 is about two times slower than Algorithm 3.2.

(3) In the practical implementation of the procedure, it is recommended to perform only a few iterations of this step such that an improvement in the difference between the value of the criterion of the starting design in step 2 and the design obtained in the iteration of (3.3) is observed. This will speed up the convergence of

the procedure substantially. In this case, equality of the function Ψ at the support points of the calculated design (as stated in Lemma 3.4) is only achieved approximately.

(4) Note that the statement regarding the convergence of the algorithm is only correct if the optimization problem in (3.3) is solved "explicitly." In practice, the maximum has to be found iteratively, and there appears a trade-off between numerical complexity and accuracy in the solution of (3.3). However, in all the examples considered so far, we observed convergence of the procedure, even if only a few iterations of (3.3) are used. In our R program the user can specify the number of iterations used in this part of the algorithm; see http://cran.r-project.org/web/packages/rodd/index.html. Thus if any problem regarding convergence is observed, the number of iterations should be increased (of course, at the cost of the speed of the algorithm).

3.2. *A gradient method.* A further option for the second step in Algorithm 3.2 is a specialized gradient method, which is used for the function

$$g(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \sum_{k=1}^{n} \omega_k \bigl[\eta_i(x_k, \overline{\theta}_i) - \eta_j(x_k, \widehat{\theta}_{i,j})\bigr]^2, \tag{3.4}$$

where $\widehat{\theta}_{i,j} = \widehat{\theta}_{i,j}(\omega)$ is defined in (3.1). For its description we define the functions

$$v_k(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \bigl[\eta_i(x_k, \overline{\theta}_i) - \eta_j\bigl(x_k, \widehat{\theta}_{i,j}(\omega)\bigr)\bigr]^2, \qquad k = 1, \ldots, n,$$

and iteratively calculate a sequence of vectors $(\omega_{(\gamma)})_{\gamma=0,1,\ldots}$. At the beginning we choose $\omega_{(0)} = \overline{\omega}$ (e.g., equal weights). If $\omega_{(\gamma)} = (\omega_{(\gamma),1}, \ldots, \omega_{(\gamma),n})$ is given, we proceed for $\gamma = 0, 1, \ldots$ as follows. We determine indices $\overline{k}$ and $\underline{k}$ corresponding to $\max_{1\le k\le n} v_k(\omega_{(\gamma)})$ and $\min_{1\le k\le n} v_k(\omega_{(\gamma)})$, respectively, and define

$$\alpha^* = \operatorname*{arg\,max}_{0\le\alpha\le\omega_{(\gamma),\underline{k}}} g\bigl(\overline{\omega}_{(\gamma)}(\alpha)\bigr), \tag{3.5}$$

where the vector $\overline{\omega}_{(\gamma)}(\alpha) = (\overline{\omega}_{(\gamma),1}(\alpha), \ldots, \overline{\omega}_{(\gamma),n}(\alpha))$ is given by

$$\overline{\omega}_{(\gamma),i}(\alpha) = \begin{cases} \omega_{(\gamma),i} + \alpha, & \text{if } i = \overline{k}, \\ \omega_{(\gamma),i} - \alpha, & \text{if } i = \underline{k}, \\ \omega_{(\gamma),i}, & \text{else.} \end{cases}$$

The vector $\omega_{(\gamma+1)}$ of the next iteration is then defined by $\omega_{(\gamma+1)} = \overline{\omega}_{(\gamma)}(\alpha^*)$. The following theorem shows that the generated sequence of vectors converges to a maximizer of the function g in (3.4) and is proved in the Appendix.

THEOREM 3.6. *The sequence $(\omega_{(\gamma)})_{\gamma\in\mathbb{N}}$ converges to a vector*

$$\omega^* \in \arg\max g(\omega).$$

REMARK 3.7. It is worthwhile to mention that the one-dimensional optimization problem (3.5) is computationally rather expensive. In the implementation we use a linearization of the optimization problem, which is obtained similarly to the method described in Section 3.1.

4. Implementation and numerical examples. We illustrate the new algorithm in two examples calculating Bayesian T-optimal discriminating designs. For this purpose we have implemented the procedure for the calculation of the locally T-optimal discriminating design in R, where the user has to specify the weights $p_{i,j}$ and the corresponding preliminary information regarding the parameters $\overline{\theta}_i$; see http://cran.r-project.org/web/packages/rodd/index.html. To be precise, we call

$$\mathrm{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,\nu-1} & p_{1,\nu} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{\nu,1} & p_{\nu,2} & \cdots & p_{\nu,\nu-1} & p_{\nu,\nu} \end{bmatrix}$$

the comparison table for the locally T-optimal discriminating design problem under consideration. This table has to be specified by the experimenter. We recall that a Bayesian T-optimal design problem with a discrete prior can be reduced to a locally T-optimal one with a large number of model comparisons. For illustration purposes we consider in a first example the case $\nu = 2$ (see Section 4.1), where the Bayesian T-optimality criterion (2.2) reduces to

$$T_{12}^{\mathrm{B}}(\xi) = \sum_{k=1}^{\ell} \tau_k \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \bigl[\eta_1(x, \lambda_k) - \eta_2(x, \theta_2)\bigr]^2 \xi(dx). \tag{4.1}$$

This is the locally T-optimality criterion (2.4) with $\nu = \ell + 1$, $p_{i,\ell+1} = \tau_i$ $(i = 1, \ldots, \ell)$ and $p_{i,j} = 0$ otherwise, that is,

$$T_{\mathrm{P}}(\xi) = \sum_{i,j=1}^{\ell+1} p_{i,j} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} \bigl[\eta_i(x, \overline{\theta}_i) - \eta_j(x, \theta_{i,j})\bigr]^2 \xi(dx), \tag{4.2}$$

where the comparison table is given by

$$\mathrm{P} = (p_{i,j})_{i,j=1,\ldots,\ell+1} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \tau_1 \\ 0 & 0 & \cdots & 0 & \tau_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \tau_\ell \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{\ell+1 \times \ell+1}, \tag{4.3}$$

$\eta_i(x, \overline{\theta}_i) = \eta_1(x, \lambda_i)$, $i = 1, \ldots, \ell$ and $\eta_{\ell+1}(x, \theta_{i,j}) = \eta_2(x, \theta_{i,\ell+1})$. Thus, instead of making only one comparison, as is required for the locally T-optimality criterion, the Bayesian approach (with a discrete approximation of the prior) yields a criterion with ℓ comparisons, where ℓ denotes the number of support points used for the approximation of the prior distribution. The extension of this approach to

more than two models is easy and left to the reader. This case is considered in our second example; see Section 4.2.

We have implemented both procedures described in Sections 3.1 and 3.2, and the results are similar. For this reason we only represent the Bayesian T-optimal discriminating designs calculated by Algorithm 3.2, where the quadratic programming method was used in step 2; see Section 3.1 for details.

4.1. *Bayesian T-optimal discriminating designs for exponential models.* Consider the problem of discriminating between the two regression models

$$\begin{aligned} \eta_1(x,\theta_1) &= \theta_{1,1} - \theta_{1,2}\exp\bigl(-\theta_{1,3}x^{\theta_{1,4}}\bigr), \\ \eta_2(x,\theta_2) &= \theta_{2,1} - \theta_{2,2}\exp(-\theta_{2,3}x), \end{aligned} \tag{4.4}$$

where the design space is given by the interval $[0, 10]$. Exponential models in the form (4.4) are widely used in applications. For example, the model η_2 is frequently fitted in agricultural sciences, where it is called Mitscherlich's growth law and is used for describing the relation between the yield of a crop and the amount of fertilizer. In fisheries research this model is called Bertalanffy growth curve and is used for the description of the length of a fish in dependence of its age; see Ratkowsky (1990). Optimal designs for exponential regression models have been determined by Han and Chaloner (2003) among others. In the following we will demonstrate the performance of the new algorithm in calculating Bayesian T-optimal discriminating designs for the two exponential models. Note that it only makes sense to consider the Bayesian version of T_{12}, because the model η_2 is obtained as a special case of η_1 for $\theta_{1,4} = 1$. It is easy to see that the locally T-optimal discriminating designs do not depend on the linear parameters of η_1, and we have chosen $\overline{\theta}_{2,1} = 2$ and $\overline{\theta}_{2,2} = 1$ for these parameters. For the parameters $\overline{\theta}_{1,3}$ and $\overline{\theta}_{1,4}$ we considered independent prior distributions supported at the points

$$\mu_j + \frac{\sigma(i-3)}{2}, \qquad i = 1,\dots,5;\ j = 3,4, \tag{4.5}$$

where $\mu_3 = 0.8$, $\mu_4 = 1.5$ and different values of the variance σ^2 are investigated. The corresponding weights at these points are proportional (in both cases) to

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\biggl(-\frac{(i-3)^2}{8}\biggr); \qquad i = 1,\dots,5. \tag{4.6}$$

We note that this yields 25 terms in the Bayesian optimality criterion (4.1). Bayesian T-optimal discriminating designs are depicted in Table 1 for various values of σ^2, where an equidistant design at 11 points $0, 1, \dots, 10$ was used as starting design.

A typical determination of the optimal design takes between 0.03 seconds (in the case $\sigma^2 = 0$) and 1.4 seconds (in the case $\sigma^2 = 0.4$) CPU time on a standard

TABLE 1

Bayesian T-optimal discriminating designs for the two exponential models in (4.4). The support points and weights of the independent prior distributions for the parameters $\overline{\theta}_{1,3}$ and $\overline{\theta}_{1,4}$ are given by (4.5) and (4.6), respectively

σ^2		Optimal design			σ^2		Optimal design			
0.0	0.000	0.441	1.952	10.000	0.285	0.000	0.453	1.758	10.000	
	0.209	0.385	0.291	0.115		0.207	0.396	0.292	0.105	
0.1	0.000	0.452	1.877	10.000	0.3	0.000	0.452	1.747	4.951	10.000
	0.209	0.391	0.290	0.110		0.207	0.396	0.292	0.003	0.102
0.2	0.000	0.455	1.811	10.000	0.4	0.000	0.446	1.651	4.699	10.000
	0.208	0.394	0.291	0.107		0.200	0.384	0.290	0.060	0.066

PC (with an Intel core i7-4790K processor). The algorithm using the procedure described in Section 3.2 in step 2 requires between 0.11 seconds (in the case $\sigma^2 = 0$) and 11.6 seconds (in the case $\sigma^2 = 0.4$) CPU time. We observe that for small values of σ^2 the optimal designs are supported at 4 points, while for $\sigma^2 \geq 0.285$ the Bayesian T-optimal discriminating design is supported at 5 points. The corresponding function Ψ from the equivalence Theorem 2.1 is shown in Figure 1.

4.2. *Bayesian T-optimal discrimination designs for dose finding studies.* Nonlinear regression models have also numerous applications in dose response studies, where they are used to describe the dose response relationship. In these and similar situations the first step of the data analysis consists of the identification of an appropriate model, and the design of the experiment should take this task into account. For example, for modeling the dose response relationship of a Phase II clinical trial, Pinheiro, Bretz and Branson (2006) proposed the following plausible models:

$$\begin{aligned} \eta_1(x,\theta_1) &= \theta_{1,1} + \theta_{1,2}x, \\ \eta_2(x,\theta_2) &= \theta_{2,1} + \theta_{2,2}x(\theta_{2,3} - x), \\ \eta_3(x,\theta_3) &= \theta_{3,1} + \theta_{3,2}x/(\theta_{3,3} + x), \\ \eta_4(x,\theta_4) &= \theta_{4,1} + \theta_{4,2}/\big(1 + \exp(\theta_{4,3} - x)/\theta_{4,4}\big), \end{aligned} \tag{4.7}$$

where the designs space (dose range) is given by the interval $\mathcal{X} = [0, 500]$. In this reference some prior information regarding the parameters for the models is also provided, that is,

$$\begin{aligned} &\overline{\theta}_1 = (60, 0.56), \qquad \overline{\theta}_2 = (60, 7/2250, 600), \\ &\overline{\theta}_3 = (60, 294, 25), \qquad \overline{\theta}_4 = (49.62, 290.51, 150, 45.51). \end{aligned}$$

Locally optimal discrimination designs for the models in (4.7) have been determined by Braess and Dette (2013) in the case $p_{i,j} = 1/6$ $(1 \leq j < i \leq 4)$, which

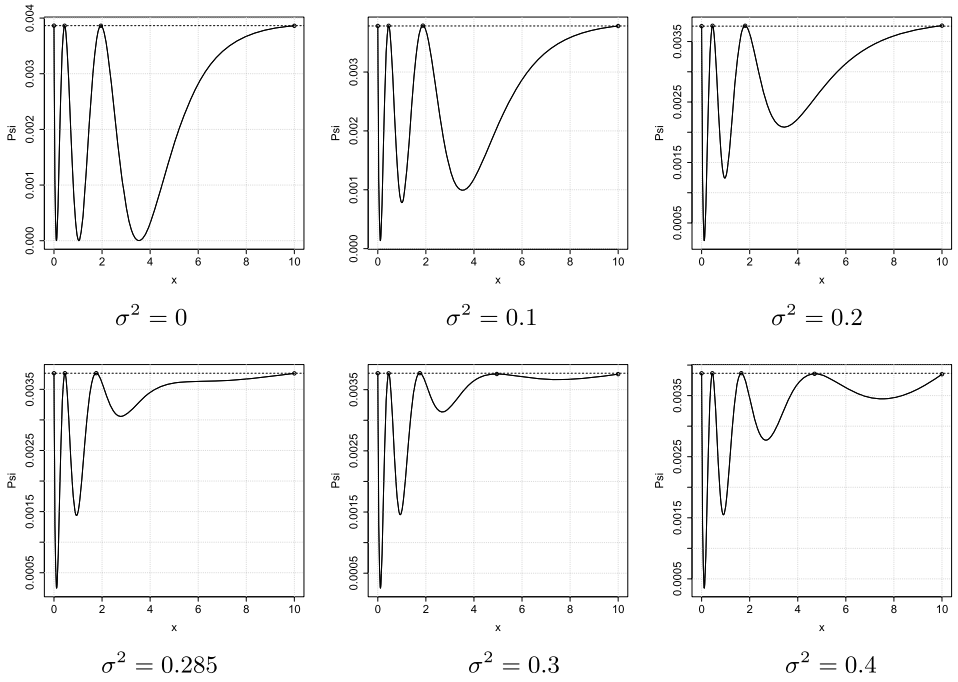

FIG. 1. *The function on the left-hand side of inequality* (2.6) *in the equivalence Theorem* 2.1 *for the numerically calculated Bayesian T-optimal discriminating designs. The competing regression models are given in* (4.4).

means that the resulting local T-optimality criterion (2.4) consists of 6 model comparisons.

We begin with an illustration of the new methodology developed in Section 3, calculating again the locally T-optimal discriminating design for this scenario. The proposed algorithm needs only four iterations for the calculation of a design, say ξ_4, which has at least efficiency

$$\mathrm{Eff}_{T_P}(\xi_4) = \frac{T_P(\tilde{\xi}_4)}{\sup_\zeta T_P(\zeta)} \geq 0.999;$$

for a definition of efficiencies with respect to general optimality criteria, see Pukelsheim (2006). The function $\Psi(\cdot, \xi_1)$ after the first iteration is displayed in Figure 2, where we use the same starting design as in Braess and Dette (2013). The support points of ξ_1 are shown as circles and—as stated in Lemma 3.4—we can see that function $\Psi(x, \xi_1)$ attains one and the same value, which is represented with a dotted line, for all support points. We finally note that the algorithm proposed in Braess and Dette (2013) needs 9 iterations to find a design with the same efficiency.

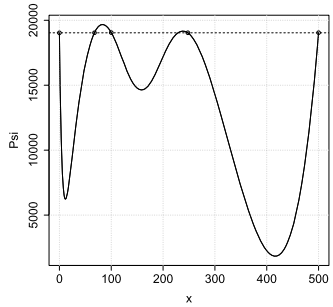

FIG. 2. *The function $\Psi(\cdot, \xi_1)$ after the first iteration of Algorithm 3.2.*

We now investigate Bayesian T-optimal discriminating designs for a similar situation. For the sake of a transparent representation we only specify a prior distribution of the four-dimensional parameter $\overline{\theta}_4$ for the calculation of the discriminating design, while $\overline{\theta}_1$, $\overline{\theta}_2$ and $\overline{\theta}_3$ are considered as fixed. In order to obtain a design which is robust with respect to model misspecification, we chose a discrete prior with 81 points in $\mathbb{R}^4$. More precisely, the support points of the prior distribution are given by the points

$$\{\mu_{e_1,e_2,e_3,e_4} | e_1, e_2, e_3, e_4 \in \{-1, 0, 1\}\}, \tag{4.8}$$

where

$$\mu_{e_1,e_2,e_3,e_4} = (\mu_1 + e_1\sigma, \mu_2 + e_2\sigma, \mu_3 + e_3\sigma, \mu_4 + e_4\sigma),$$
$$\mu = (\mu_1, \mu_2, \mu_3, \mu_4) = (49.62, 290.51, 150, 45.51)$$

and different values for σ^2 are considered. The weights at the corresponding points are proportional (normalized such that their sum is 1) to

$$\frac{1}{(2\pi\sigma^2)^2} \exp\left(\frac{\|\mu_{e_1,e_2,e_3,e_4} - \mu\|_2^2}{2\sigma^2}\right), \qquad e_1, e_2, e_3, e_4 \in \{-1, 0, 1\}, \tag{4.9}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. The resulting Bayesian optimality criterion (2.10) consist of 246 model comparisons. In this case the method of Braess and Dette (2013) fails to find the Bayesian T-optimal discriminating design. Bayesian T-optimal discriminating designs have been calculated by the new Algorithm 3.2 for various values of σ^2, and the results are shown in Table 2. A typical determination of the optimal design takes between 0.09 seconds (in the case $\sigma^2 = 0$) and 7.8 seconds (in the case $\sigma^2 = 37^2$) CPU time on a standard PC. The algorithm using the procedure described in Section 3.2 in step 2 requires between 0.75 seconds (in the case $\sigma^2 = 0$) and 37.1 seconds (in the case $\sigma^2 = 37^2$) CPU time. For small values, the Bayesian T-optimal discriminating designs are supported at 4 points including the boundary of the design space. The smaller (larger) interior support point is increasing (decreasing) if σ^2 is increasing. For larger values of σ^2 even the number of support points of the optimal design increases. For

TABLE 2

Bayesian T-optimal discriminating designs for the models in (4.7). The weights in criterion (2.9) are given by $p_{i,j} = 1/6$; $1 \le i < j \le 4$, and the support and masses of the prior distribution are defined by (4.8) and (4.9), respectively

σ^2	Optimal design					
0	0.000	78.783	241.036	500.0		
	0.255	0.213	0.357	0.175		
20^2	0.000	84.467	234.134	500.0		
	0.257	0.225	0.351	0.167		
30^2	0.000	91.029	225.713	500.0		
	0.259	0.237	0.345	0.159		
33^2	0.000	92.692	222.735	500.0		
	0.260	0.240	0.344	0.156		
35^2	0.000	91.743	129.322	221.118	500.0	
	0.260	0.214	0.036	0.336	0.154	
37^2	0.000	89.881	129.590	170.306	220.191	500.0
	0.260	0.170	0.091	0.019	0.310	0.150

example, if $\sigma^2 = 35^2$ or 37^2, the Bayesian T-optimal discriminating design has 5 or 6 points (including the boundary points of the design space). These observations are in line with the theoretical finding of Braess and Dette (2007) who showed that the number of support points of Bayesian D-optimal designs can become arbitrarily large with an increasing variability in the prior distribution. The corresponding functions from the equivalence Theorem 2.1 are shown in Figure 3.

4.3. *A brief comparison with alternative procedures*. In this section we will provide a brief comparison of the method proposed in this paper with the currently available algorithms. First, note that the performance of the algorithms of Yu (2010) and Yang, Biedermann and Tang (2013) depends on the specification of a grid, which approximates the design space $\mathcal{X}$. In contrast to these methods, the approach proposed in the present paper does not use such a discretization.

For locally T-optimal discriminating design problems, we have already provided a comparison with the method proposed in Braess and Dette (2013) in the previous sections. For a large number of model comparisons, as required by the Bayesian T-optimality criterion considered in this paper, it is very difficult to implement this method, and for these reasons the procedure of Braess and Dette (2013) is not included in the comparison. Similarly, the method of Yu (2010) is particularly designed for the construction of optimal designs with respect to Kiefer's Φ_p-optimality criteria. It is not clear how this approach can be modified for the Bayesian T-optimality criterion. On the other hand, we were able to adapt the method proposed by Yang, Biedermann and Tang (2013) to the current problem

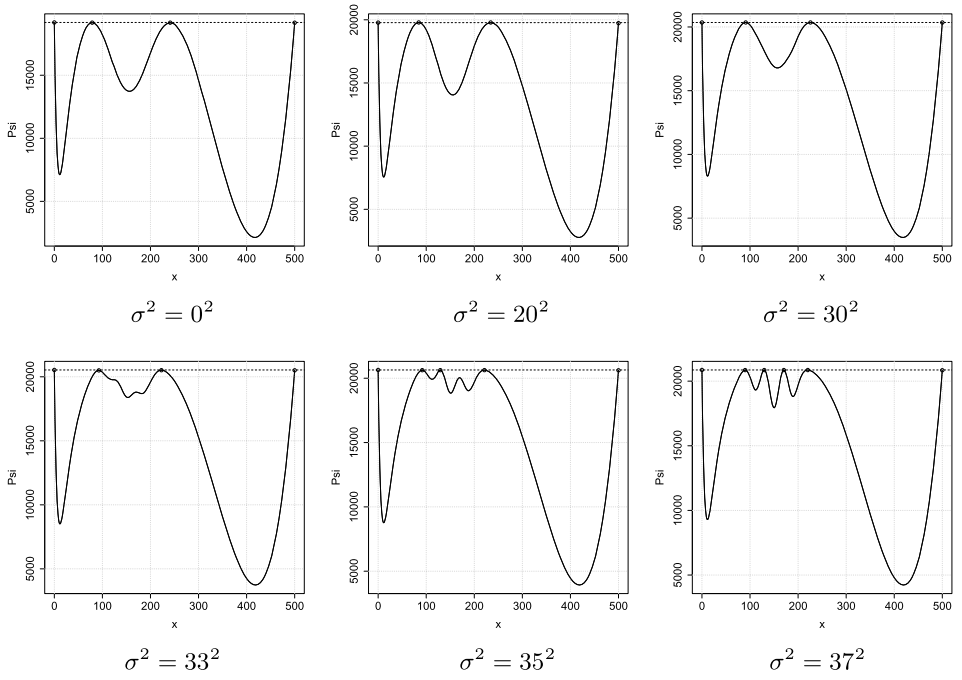

FIG. 3. *The function on the left-hand side of inequality* (2.6) *in the equivalence Theorem* 2.1 *for the numerically calculated Bayesian T-optimal discriminating designs. The competing regression models are given in* (4.7).

of constructing Bayesian T-optimal discriminating designs. We also consider a modification of the algorithm proposed by Atkinson and Fedorov (1975a) in our comparison.

To be precise the implementation of the algorithm of Atkinson and Fedorov (1975a) follows the description in Algorithm 3.1. In every iteration the point x_{s+1} corresponding to the global maximum of the function Ψ defined in (2.8) is added to the support of the design. For the choice of the weights, several suggestions have been made in the literature, and we have investigated some of these proposals for the construction of the Bayesian T-optimal discriminating designs. The performance of Algorithm 3.1 depends very sensitively on this choice, and for some rules the procedure is not even able to find the Bayesian T-optimal discriminating design with the required accuracy. For example, recall that in the $(s+1)$st iteration of Algorithm 3.1 the weights at the points $x_{s,1}, \ldots, x_{s,n}, x_{s+1}$ are given by $(1-\alpha_s)w_{s,1}, \ldots, (1-\alpha_s)w_{s,n}, \alpha_s$, respectively, where $\{x_{s,i}\}_{i=1}^n$ $\{w_{s,i}\}_{i=1}^n$ are the support points and weights of the design ξ_s from the previous iteration. If the scaling factor α_s in each iteration is determined such that the new vector of weights maximizes the function g defined in (3.2), Algorithm 3.1 is not able to find a design with an accuracy of (at least) 0.999. In our numerical study of Algorithm 3.1 we

found that the most efficient method in Algorithm 3.1 from the previous iteration is to choose the weights according to the rule $\alpha_s = 1/(s+1)$ in the sth iteration, and this "best" version of Algorithm 3.1 was considered in our comparison. We would also like to emphasize once again that Algorithm 3.1 produces a sequence of designs with a large number of support points and several modifications to reduce the number of support points are numerically not stable.

The algorithm proposed in Yang, Biedermann and Tang (2013) combines features of Algorithm 3.1 for the choice of the new support with an optimization procedure for the corresponding weights. This method has a particular focus on Kiefer's Φ_p-optimal designs. The general idea of this algorithm (adding in a first step a new support point and optimizing in a second step with respect to the weights in order to reduce the number of support points) is similar to the methods proposed in this paper. However, there are important differences between the two procedures. On one hand, the algorithm proposed in Yang, Biedermann and Tang (2013) adds only one support point in each step of the iteration, whereas the methods proposed here include all local extreme points of the sensitive function Ψ defined in (2.8). On the other hand, a general Newton method is used by Yang, Biedermann and Tang (2013) for the optimization with respect to the weights in step 2, while Algorithm 3.2 (with its two versions for step 2) incorporates special techniques for this purpose, which address the specific features of the T-optimality criterion under consideration.

The algorithm proposed in Yang, Biedermann and Tang (2013) can be adapted for the construction of Bayesian T-optimal discriminating designs (by replacing the Newton step for the calculation of the weights for the Φ_p-optimality criteria by a similar nonlinear programming procedure), and this modification will also be considered in our comparison.

In Table 3 we display the computing time (in seconds) for the different algorithms in the examples considered in Sections 4.1 and 4.2. We consider two different cases for the variance σ^2, where the case $\sigma^2 = 0$ corresponds to the situation of locally T-optimal discriminating designs. We observe that Algorithm 3.2, which uses quadratic programming in step 2 of the procedure, has the best performance. Algorithm 3.2, which uses the gradient method in step 2 of the procedure, is about 3–8 times slower. On the one hand, the most efficient procedure based on Algorithm 3.1 of Atkinson and Fedorov (1975a) needs a substantially larger computing time for the calculation of the Bayesian T-optimal discriminating design. In most cases under consideration the computing time is about 100 times larger than the time required by Algorithm 3.2 with the quadratic programming method. The adaptation of the algorithm proposed by Yang, Biedermann and Tang (2013) shows a better performance than Algorithm 3.1 in Example 4.4 in the case $\sigma^2 = 0$. In all other cases Algorithm 3.1 provides better results with respect to computing time. In particular, the adaptation of the method proposed by Yang, Biedermann and Tang (2013) was not able to find the Bayesian T-optimal discriminating design in Example 4.7 in the case $\sigma^2 = 37^2$. In all cases under consideration the new

TABLE 3

Computing times of various algorithms for the determination of Bayesian T-optimal discriminating designs in Examples 4.4 and 4.7. Alg. 3.2(1): *Algorithm 3.2 using quadratic programming in step* 2; *Alg.* 3.2(2): *Algorithm 3.2 using the gradient method in step* 2; *AF*: *Algorithm 3.1 of Atkinson and Fedorov* (1975a) *with an optimized choice of weights*; *YTB*: *adaptation of the method proposed by Yang, Biedermann and Tang* (2013)

Example	σ^2	Alg. 3.2(1)	Alg. 3.2(2)	AF	YTB
(4.4)	0.0	0.03	0.11	12.4	3.0
	0.4	1.4	11.6	218.3	369.1
(4.7)	0.0	0.09	0.75	5.7	12.1
	37^2	7.8	37.1	762.3	–

algorithms proposed in this paper were able to calculate the Bayesian T-optimal discriminating design with substantially shorter computing times than the currently available procedures.

5. Conclusions. In this paper we have investigated the problem of constructing optimal designs for model discrimination with a particular focus on Bayesian T-optimality criteria. It is demonstrated that—from a practical point of view—the Bayesian T-optimality criterion is closely related to a local T-optimality criterion for a very large class of competing models because the integrals in the criterion have to be approximated by quadrature formulas.

As a consequence, the numerical determination of Bayesian T-optimal designs is a particular challenging task, and currently available methods are computationally too expensive to solve the corresponding optimization problem in a reasonable time with the required accuracy. In this paper a new algorithm is developed, which combines the features of classical exchange type algorithms with two commonly used optimization approaches, namely quadratic programming and gradient methods. The convergence of the sequence of designs generated by this method is proved, and its application is illustrated in several examples, where the currently available methods for the determination of Bayesian T-optimal designs do not yield a satisfactory solution of the optimal design problems. The new methodology is available in `R` and therefore is easily accessible for practitioners; see http://cran.r-project.org/web/packages/rodd/index.html.

In this paper the methodology is developed for the situation regression models with Gaussian errors as considered in the classical work of Atkinson and Fedorov (1975a). An interesting direction for future research consists of an extension of the results for more general distribution families, as discussed, for example, in Uciński and Bogacka (2005) and López-Fidalgo, Tommasi and Trandafir (2007). We expect that similar methods for the determination of Bayesian optimal discriminating designs can also be developed for these models.

APPENDIX: PROOFS

A.1. An auxiliary result.

LEMMA A.1. *Let $\varphi(v, y)$ be a twice continuously differentiable function of two variables $v \in \mathcal{V} \subset \mathbb{R}^k$ and $y \in \mathcal{Y}$, where $\mathcal{Y}$ is a compact set. Denote by $\mathcal{Y}^*$ the set of all points where the minimum $\min_{y\in\mathcal{Y}} \varphi(v, y)$ is attained, and let $q \in \mathbb{R}^k$ be an arbitrary direction. Then*

$$\frac{\partial}{\partial q} \min_{y\in\mathcal{Y}} \varphi(v, y) = \min_{y\in\mathcal{Y}^*} \frac{\partial \varphi(v, y)}{\partial q}, \tag{A.1}$$

where $\partial/\partial q$ denotes the directional derivative in the direction of q.

PROOF. See Pshenichnyi (1971), page 75. □

A.2. Proofs.

PROOF OF THEOREM 2.2. Assume without loss of generality that $p_{i,j} > 0$ for all $i, j = 1, \dots, \nu$. Let ξ^* denote any locally T-optimal discriminating design, and let $\theta = (\theta_{i,j})_{i,j=1,\dots,\nu}$ denote the vector consisting of all $\theta_{i,j} \in \Theta^*_{i,j}(\xi^*)$. We introduce the function

$$\varphi(x, \theta, \xi^*) = \sum_{i,j=1}^{\nu} p_{i,j}\big[\eta_i(x, \overline{\theta}_i) - \eta_j(x, \theta_{i,j})\big]^2 \tag{A.2}$$

and consider the product measure

$$\mu(d\theta) = \prod_{i,j=1,\dots,\nu} \mu_{i,j}(d\theta_{i,j}), \tag{A.3}$$

where μ_{ij} are measures on the sets $\Theta^*_{i,j}(\xi^*)$ defined by (2.5). Similarly, we define $\mu^*(d\theta) = \prod_{i,j=1,\dots,\nu} \mu^*_{i,j}(d\theta_{i,j})$ as the product measure of the measures $\mu^*_{i,j}$ in Theorem 2.1. From this result we have

$$\begin{aligned} T_{\mathrm{P}}(\xi^*) &\geq \sup_{\zeta} \int_{\mathcal{X}} \int_{\Theta^*(\xi^*)} \varphi(x, \theta, \xi^*) \mu^*(d\theta) \zeta(dx) \\ &\geq \inf_{\mu} \sup_{\zeta} \int_{\mathcal{X}} \int_{\Theta^*(\xi^*)} \varphi(x, \theta, \xi^*) \mu(d\theta) \zeta(dx) \\ &= \sup_{\zeta} \inf_{\mu} \int_{\mathcal{X}} \int_{\Theta^*(\xi^*)} \varphi(x, \theta, \xi^*) \mu(d\theta) \zeta(dx), \end{aligned}$$

where the sup and inf are calculated in the class of designs ζ on $\mathcal{X}$ and product measures μ on $\Theta^*(\xi^*) = \bigotimes_{i,j=1}^{\nu} \Theta^*_{i,j(\xi^*)}$, respectively. It now follows that the characterizing inequality (2.6) in Theorem 2.1 is equivalent to the inequality

$$\sup_{\zeta} Q(\zeta, \xi^*) \leq T_{\mathrm{P}}(\xi^*).$$

Consequently, any nonoptimal design must satisfy the opposite inequality. □

PROOF OF COROLLARY 2.3. Let ξ denote a design such that $T_{\mathrm{P}}(\xi) > 0$, and recall the definition of the set $\Theta^*_{ij}(\xi)$ in (2.5). We consider for a vector $\theta = (\theta_{i,j})_{i,j=1,\ldots,v} \in \Theta^*(\xi) = \bigotimes_{i,j=1,\ldots,v} \Theta^*_{i,j}(\xi)$, the function $\varphi(\cdot,\cdot,\xi)$ is defined in (A.2) and product measures $\mu(d\theta)$ of the form (A.3) on $\Theta^*(\xi)$. Now the well-known minimax theorem [see, e.g., Sion (1958)] and the definition of the function Q in (2.7) yield

$$\begin{aligned}
\max_{x\in\mathcal{X}} \Psi(x,\xi) &= \inf_{\mu} \max_{x\in\mathcal{X}} \int_{\Theta^*(\xi)} \varphi(x,\theta,\xi)\mu(d\theta) \\
&= \inf_{\mu} \sup_{\zeta} \int_{\mathcal{X}} \int_{\Theta^*(\xi)} \varphi(x,\theta,\xi)\mu(d\theta)\zeta(dx) \\
&= \sup_{\zeta} \inf_{\mu} \int_{\mathcal{X}} \int_{\Theta^*(\xi)} \varphi(x,\theta,\xi)\mu(d\theta)\zeta(dx) \\
&= \sup_{\zeta} \inf_{\theta\in\Theta^*(\xi)} \int \varphi(x,\theta,\xi)\zeta(dx) = \sup_{\zeta} Q(\zeta,\xi),
\end{aligned}$$

where the infimum is calculated with respect to all measures μ in the form of (A.3), and the supremum is calculated with respect to all experimental designs ζ on $\mathcal{X}$. Note that $\mathcal{X}$ is compact by assumption, and it can be checked that the set $\Theta^*(\xi)$ is also compact as a closed subset of a compact set. Consequently all suprema and infima are achieved, and there exists a design ζ^* supported at the set of local maxima of the function $\Psi(x,\xi)$, such that

$$Q(\zeta^*,\xi) = \sup_{\zeta} Q(\zeta,\xi) = \max_{x\in\mathcal{X}} \Psi(x,\xi).$$

The assertion of Corollary 2.3 now follows from Theorem 2.2. □

PROOF OF THEOREM 3.3. Obviously, the inequality

$$T_{\mathrm{P}}\bigl(\{\mathcal{S}_{[s]},\omega_{[s]}\}\bigr) \le T_{\mathrm{P}}\bigl(\{\mathcal{S}_{[s+1]},\omega_{[s+1]}\}\bigr)$$

holds for all s as optimization with respect to ω occurs on a larger set. Moreover, the sequence $T_{\mathrm{P}}(\xi_s)$ is bounded from above by $T_{\mathrm{P}}(\xi^*)$ and has a limit, which is denoted by T^{**}_{P}. Consequently, there exists a subsequence of designs, say ξ_{s_j}, $j = 1, 2, \ldots$ converging to a design, say ξ^{**}. Note that T_{P} is upper semi-continuous as the infimum of continuous functions, which implies $T_{\mathrm{P}}(\xi^{**}) = T^{**}_{\mathrm{P}}$. Now, assume that $T_{\mathrm{P}}(\xi^{**}) < T_{\mathrm{P}}(\xi^*)$. Then ξ^{**} is not locally T-optimal, and by Theorem 2.2 there exists a constant $\delta > 0$ such that

$$\sup_{\zeta} Q(\zeta,\xi^{**}) - T_{\mathrm{P}}(\xi^{**}) = 2\delta,$$

where the function Q is defined in (2.7). Therefore for sufficiently large j, say, $j \ge N$ we obtain [using again the lower semi-continuity of $\sup_\zeta Q(\zeta,\xi)$] that

$$\sup_{\zeta} Q(\zeta,\xi_{s_j}) - T_{\mathrm{P}}(\xi_{s_j}) > \delta,$$

whenever $j \geq N$. Note that by construction the sequence $(T_P(\xi_s))_{s\in\mathbb{N}}$ is increasing and therefore

$$T_P(\xi_{s_{j+1}}) - T_P(\xi_{s_j}) \geq T_P(\xi_{s_j+1}) - T_P(\xi_{s_j}). \tag{A.4}$$

In order to estimate the right-hand side, we consider for $j \geq N$ and $\alpha \in [0,1]$, the design

$$\tilde{\xi}_{s_{j+1}}(\alpha) = (1-\alpha)\xi_{s_j} + \alpha\zeta_j,$$

where ζ_j is the measure for which the function $Q(\zeta, \xi_{s_j})$ attains its maximal value in the class of all experimental designs supported at the local maxima of the function $\Psi(x, \xi_{s_j})$, and define

$$\alpha_{s_{j+1}} = \arg\max_{0\leq\alpha\leq 1} T_P\big(\tilde{\xi}_{s_{j+1}}(\alpha)\big).$$

By construction of ξ_{s_j+1} is the best design supported at $\operatorname{supp}(\xi_{s_j}) \cup \operatorname{supp}(\zeta_j)$, and (A.4) yields

$$T_P(\xi_{s_{j+1}}) \geq T_P(\xi_{s_j+1}) \geq T_P\big(\tilde{\xi}_{s_{j+1}}(\alpha_{s_{j+1}})\big). \tag{A.5}$$

We introduce the notation $h(j,\alpha) = T_P(\tilde{\xi}_{s_j}(\alpha))$ and note that

$$\frac{\partial T_P(\tilde{\xi}_{s_{j+1}}(\alpha))}{\partial\alpha}\bigg|_{\alpha=0} = Q(\zeta_j, \xi_{s_j}) - T_P(\xi_{s_j}) = \sup_{\zeta} Q(\zeta, \xi_{s_j}) - T_P(\xi_{s_j}) > \delta.$$

A Taylor expansion gives

$$\begin{aligned} h(j+1, \alpha_{s_{j+1}}) - h(j+1, 0) &= \max_{\alpha\in[0,1]} \big[T_P\big(\tilde{\xi}_{s_{j+1}}(\alpha)\big) - T_P\big(\tilde{\xi}_{s_{j+1}}(0)\big)\big] \\ &\geq \max_{\alpha\in[0,1]} \left[\alpha \frac{\partial T_P(\tilde{\xi}_{s_{j+1}}(\alpha))}{\partial\alpha}\bigg|_{\alpha=0} - \frac{1}{2}\alpha^2 K\right] \\ &> \max_{\alpha\in[0,1]} \left[\alpha\delta - \frac{1}{2}\alpha^2 K\right] = \frac{\delta^2}{2K}, \end{aligned}$$

where K is an absolute upper bound of the second derivative. Therefore it follows from (A.5) that

$$\begin{aligned} T_P(\xi_{s_{j+1}}) - T_P(\xi_{s_j}) &\geq T_P(\xi_{s_j+1}) - T_P(\xi_{s_j}) \\ &\geq T_P\big(\tilde{\xi}_{s_{j+1}}(\alpha_{s_{j+1}})\big) - T_P(\xi_{s_j}) \\ &= h(j+1, \alpha_{s_{j+1}}) - h(j+1, 0) \geq \frac{\delta^2}{2K}, \end{aligned}$$

which gives for $L > N+1$,

$$T_P(\xi_{s_L}) - T_P(\xi_{s_N}) = \sum_{j=N}^{L-1} \big[T_P(\xi_{s_{j+1}}) - T_P(\xi_{s_j})\big] \geq [L-N]\frac{\delta^2}{2K}.$$

The left-hand side of this inequality converges to the finite value $T(\xi^{**}) - T(\xi_{s_N})$ as $L \to \infty$, while the right-hand side converges to infinity. Therefore we obtain a contradiction to our assumption $T_P(\xi^{**}) < T_P(\xi^*)$, which proves the assertion of Theorem 3.3. □

PROOF OF LEMMA 3.4. Fix $t \in \{1, \dots, n\}$, and note that $w_t = 1 - \sum_{\ell=1, \ell \neq t}^{n} w_\ell$. Under Assumptions 2.1 and 2.2 we obtain by formula (A.1)

$$\frac{\partial g(\omega)}{\partial \omega_k} = \sum_{i,j=1}^{\nu} p_{i,j}\big[\eta_i(x_k, \overline{\theta}_i) - \eta_j\big(x_k, \widehat{\theta}_{i,j}(\omega)\big)\big]^2 - \sum_{i,j=1}^{\nu} p_{i,j}\big[\eta_i(x_t, \overline{\theta}_i) - \eta_j\big(x_t, \widehat{\theta}_{i,j}(\omega)\big)\big]^2.$$

The condition $\frac{\partial g(\omega)}{\partial \omega_k} = 0, k = 1, \dots, n, k \neq t$ is the necessary condition for weight optimality, and consequently it follows from the definition of the function $\Psi(x, \xi_{s+1})$ that this function attains one and the same value for all support points of the design ξ_{s+1}. □

PROOF OF THEOREM 3.6. The proof is similar to the proof of Theorem 3.3. Denote

$$\kappa(\gamma, \alpha) = g\big(\overline{\omega}_{(\gamma)}(\alpha)\big),$$

where the vector $\overline{\omega}_{(\gamma)}(\alpha^*)$ is calculated at the γth iteration using the definition (3.4). Since the sequence $g(\omega_{(\gamma)})$ is bounded and increasing (by construction), it converges to some limit, say g^{**}. Consequently there exists a subsequence of vector of weights, say $\overline{\omega}_{(\gamma_j)}$, $j = 1, 2, \dots$ converging to a vector, say $\overline{\omega}^{**}$. Note that g is upper semi-continuous as the infimum of continuous functions, which implies $g(\overline{\omega}^{**}) = g^{**}$. Now, assume that $g(\overline{\omega}^{**}) < g(\omega^*)$. Then it follows by an application of Theorem 2.1 with $\mathcal{X} = \{x_1, \dots, x_n\}$ that there exists a constant $\delta > 0$ such that

$$\left.\frac{\partial g(\overline{\omega}(\alpha))}{\partial \alpha}\right|_{\alpha=0} = 2\delta > 0.$$

Here the vector $\overline{\omega}(\alpha)$ is defined in the same way as $\overline{\omega}_{(\gamma)}(\alpha)$, where $\omega_{(\gamma)}$ is replaced by $\omega = \omega^{**}$. Therefore for sufficiently large j, say, $j \geq N$, we obtain (using the lower semi-continuity of g) that $\kappa(\gamma_j, 0) > \delta$, and a Taylor expansion yields

$$\kappa\big(\gamma_{j+1}, \alpha^*_{(\gamma_{j+1})}\big) - \kappa\big(\gamma_j, \alpha^*_{(\gamma_j)}\big) \geq \max_{\alpha}\left(\alpha \frac{\partial g(\overline{\omega}(\alpha))}{\partial \alpha} - \frac{1}{2}\alpha^2 K\right) = \frac{\delta^2}{2K},$$

where $\alpha^*_{(\gamma_j)}$ is the value α^* defined in (3.4) from the γ_jth iteration, and K is an absolute upper bound of the second derivative. Using the same arguments as in the proof of Theorem 3.3, we obtain a contradiction, which proves the assertion of the theorem. □

Acknowledgments. Parts of this work were completed during a visit of the second author to the Department of Mathematics, Ruhr-Universität Bochum, Germany. The authors would like to thank M. Stein who typed this manuscript with considerable technical expertise. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors are also grateful to two unknown referees for their constructive comments on an earlier version of this paper, which led to a substantial improvement of our work.

REFERENCES

ABRAMOWITZ, M. and STEGEN, I. A. (1965). *Handbook of Mathematical Functions with Formulas, Graph, and Mathematical Tables*. Dover, New York.

ALETTI, G., MAY, C. and TOMMASI, C. (2013). A convergent algorithm for finding KL-optimum designs and related properties. In *MODa* 10—*Advances in Model-Oriented Design and Analysis* (D. Ucinski, A. C. Atkinson and M. Patan, eds.) 1–9. Springer, Berlin.

ALETTI, G., MAY, C. and TOMMASI, C. (2014). KL-optimum designs: Theoretical properties and practical computation. *Statistics and Computing*. DOI:10.1007/s11222-014-9515-8.

ATKINSON, A. C. (2008). Examples of the use of an equivalence theorem in constructing optimum experimental designs for random-effects nonlinear regression models. *J. Statist. Plann. Inference* **138** 2595–2606. MR2439971

ATKINSON, A. C., BOGACKA, B. and BOGACKI, M. B. (1998). D- and T-optimum designs for the kinetics of a reversible chemical reaction. *Chemom. Intell. Lab. Syst.* **43** 185–198.

ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs, with SAS. Oxford Statistical Science Series* **34**. Oxford Univ. Press, Oxford. MR2323647

ATKINSON, A. C. and FEDOROV, V. V. (1975a). The design of experiments for discriminating between two rival models. *Biometrika* **62** 57–70. MR0370955

ATKINSON, A. C. and FEDOROV, V. V. (1975b). Optimal design: Experiments for discriminating between several models. *Biometrika* **62** 289–303. MR0381163

BRAESS, D. and DETTE, H. (2007). On the number of support points of maximin and Bayesian optimal designs. *Ann. Statist.* **35** 772–792. MR2336868

BRAESS, D. and DETTE, H. (2013). Optimal discriminating designs for several competing regression models. *Ann. Statist.* **41** 897–922. MR3099125

CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: A review. *Statist. Sci.* **10** 273–304. MR1390519

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Stat.* **24** 586–602. MR0058932

DETTE, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59** 97–110. MR1436556

DETTE, H. and HALLER, G. (1998). Optimal designs for the identification of the order of a Fourier regression. *Ann. Statist.* **26** 1496–1521. MR1647689

DETTE, H., MELAS, V. B. and SHPILEV, P. (2012). T-optimal designs for discrimination between two polynomial models. *Ann. Statist.* **40** 188–205. MR3013184

DETTE, H., MELAS, V. B. and SHPILEV, P. (2013). Robust T-optimal discriminating designs. *Ann. Statist.* **41** 1693–1715. MR3127846

DETTE, H. and TITOFF, S. (2009). Optimal discrimination designs. *Ann. Statist.* **37** 2056–2082. MR2533479

FOO, L. K. and DUFFULL, S. (2011). Optimal design of pharmacokinetic–pharmacodynamic studies. In *Pharmacokinetics in Drug Development, Advances and Applications* (P. L. Bonate and D. R. Howard, eds.). Springer, Berlin.

HAN, C. and CHALONER, K. (2003). D- and c-optimal designs for exponential regression models used in viral dynamics and other applications. *J. Statist. Plann. Inference* **115** 585–601. MR1985885

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879. MR0356386

LÓPEZ-FIDALGO, J., TOMMASI, C. and TRANDAFIR, P. C. (2007). An optimal experimental design criterion for discriminating between non-normal models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 231–242. MR2325274

PINHEIRO, J., BRETZ, F. and BRANSON, M. (2006). Analysis of dose–response studies: Modeling approaches. In *Dose Finding in Drug Development* (N. Ting, ed.) 146–171. Springer, New York.

PONCE DE LEON, A. C. and ATKINSON, A. C. (1991). Optimum experimental design for discriminating between two rival models in the presence of prior information. *Biometrika* **78** 601–608. MR1130928

PRONZATO, L. and WALTER, E. (1985). Robust experiment design via stochastic approximation. *Math. Biosci.* **75** 103–120. MR0800967

PSHENICHNYI, B. N. (1971). *Necessary Conditions for an Extremum*. Dekker, New York. MR0276845

PUKELSHEIM, F. (2006). *Optimal Design of Experiments*. *Classics in Applied Mathematics* **50**. SIAM, Philadelphia. MR2224698

RATKOWSKY, D. (1990). *Handbook of Nonlinear Regression Models*. Dekker, New York.

SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. MR0097026

SONG, D. and WONG, W. K. (1999). On the construction of G_{rm}-optimal designs. *Statist. Sinica* **9** 263–272. MR1678893

STIGLER, S. (1971). Optimal experimental design for polynomial regression. *J. Amer. Statist. Assoc.* **66** 311–318.

TOMMASI, C. (2009). Optimal designs for both model discrimination and parameter estimation. *J. Statist. Plann. Inference* **139** 4123–4132. MR2558355

TOMMASI, C. and LÓPEZ-FIDALGO, J. (2010). Bayesian optimum designs for discriminating between models with any distribution. *Comput. Statist. Data Anal.* **54** 143–150. MR2558465

UCIŃSKI, D. and BOGACKA, B. (2005). T-optimum designs for discrimination between two multiresponse dynamic models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 3–18. MR2136636

WIENS, D. P. (2009). Robust discrimination designs. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 805–829. MR2750096

YANG, M., BIEDERMANN, S. and TANG, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *J. Amer. Statist. Assoc.* **108** 1411–1420. MR3174717

YU, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *Ann. Statist.* **38** 1593–1606. MR2662353

H. DETTE
FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: holger.dette@rub.de

V. B. MELAS
R. GUCHENKO
FACULTY OF MATHEMATICS AND MECHANICS
ST. PETERSBURG STATE UNIVERSITY
ST. PETERSBURG
RUSSIA
E-MAIL: vbmelas@post.ru
romanguchenko@ya.ru