View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Eldorado - Ressourcen aus und für Lehre, Studium und Forschung

SFB 823

Optimal designs for comparing curves

Holger Dette, Kirsten Schorning

Nr. 40/2015

Discussion Paper

SFB 823

Submitted to the Annals of Statistics
arXiv: 1410.7616

OPTIMAL DESIGNS FOR COMPARING CURVES

By Holger Dette and Kirsten Schorning

Ruhr-Universität Bochum

We consider the optimal design problem for a comparison of two regression curves, which is used to establish the similarity between the dose response relationships of two groups. An optimal pair of designs minimizes the width of the confidence band for the difference between the two regression functions. Optimal design theory (equivalence theorems, efficiency bounds) is developed for this non standard design problem and for some commonly used dose response models optimal designs are found explicitly. The results are illustrated in several examples modeling dose response relationships. It is demonstrated that the optimal pair of designs for the comparison of the regression curves is **not** the pair of the optimal designs for the individual models. In particular it is shown that the use of the optimal designs proposed in this paper instead of commonly used "non-optimal" designs yields a reduction of the width of the confidence band by more than 50%.

1. Introduction. An important problem in many scientific research areas is the comparison of two regression models that describe the relation between a common response and the same covariates for two groups. Such comparisons are typically used to establish the non-superiority of one model to the other or to check whether the difference between two regression models can be neglected. These investigations have important applications in drug development and several methods for assessing non-superiority, non-inferiority or equivalence have been proposed in the recent literature [for a recent reference see for example Gsteiger, Bretz and Liu (2011)]. For example, if the "equivalence" between two regression models describing the dose response relationships in the groups individually has been established subsequent inference in drug development could be based on the combined samples. This results in more precise estimates of the relevant parameters, for example the minimum effective dose. Comparison of curves problems have been investigated in linear and nonlinear models [see Liu et al. (2009), Gsteiger, Bretz and Liu (2011), Liu, Jamshidian and Zhang (2011)] and also in nonparametric regression models [see for example Hall and Hart (1990)

AMS 2000 subject classifications: Primary62K05
Keywords and phrases: similarity of regression curves, Confidence band, Optimal design, Robust design

and Dette and Neumeyer (2001)]. A common approach in all these references is to estimate regression curves in the different samples and to investigate the maximum or an L_2-distance (taken over the possible range of the covariates) of the difference between these estimates (after an appropriate standardization by a variance estimate).

This paper is devoted to the construction of efficient designs for the comparison of two parametric curves. Although the consideration of optimal designs for dose response models has found considerable interest in the recent literature [see for example Dette et al. (2008), Dragalin et al. (2010) and Dette, Bornkamp and Bretz (2013) for recent references], we are not aware of any work on design of experiments for the comparison of two parametric regression curves. However, the effective planning of the experiments in the comparison of curves will yield to a substantially more accurate statistical inference. We demonstrate these advantages in Section 5 showing that the width of the (simultaneous) confidence bands proposed by Gsteiger, Bretz and Liu (2011) for the difference of the curves is about two times smaller if a design constructed in this paper is used instead of a standard design.

The remaining part of this paper is organized as follows. Some terminology (for the comparison of two parametric curves) will be introduced in Section 2, where we also give an introduction to optimal design theory in the present context. The particular difference to the classical setup is that for the comparison of two curves two designs have to be chosen simultaneously (each for one group or regression model). A pair of optimal designs minimizes an integral or the maximum of the variance of the prediction for the difference of the two regression curves calculated in the common region of interest. Section 3 is devoted to some optimal design theory and we derive particular equivalence theorems corresponding to the new optimality criteria and a lower bound for the efficiencies, which can be used without knowing the optimal designs. It turns out that in general the optimal pair of designs is not the pair of the optimal designs in the individual models.

In general, the problem of constructing optimal designs is very difficult and has to be solved numerically in most cases of practical interest. Some analytical results are given in Section 4 for the commonly used Michaelis Menten, Emax and loglinear model. In Section 5 we use the developed theory to investigate specific optimal design problems for the comparison of nonlinear regression models, which are frequently used in drug development. In particular we demonstrate by means of a simulation study that the derived optimal designs yield substantially narrower confidence bands. Some further discussion is given in Section 6. In Section 6.1 we briefly indicate how the results can be generalized if optimization can also be performed with respect to the allocation of patients to the different groups, while some robustness issues are discussed in Section 6.2. Finally, all proofs and technical details are deferred to an Appendix in Section 7.

2. Comparing parametric curves.

Consider the regression models

$$Y_{ijk} = m_i(t_{ij}, \vartheta_i) + \varepsilon_{ijk}\ ; \quad i = 1, 2; \quad j = 1, \dots, \ell_i\ ; \quad k = 1, \dots, n_{ij}, \tag{2.1}$$

where ε_{ijk} are independent random variables, such that $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, 2$. This means that two groups ($i = 1, 2$) are investigated and in each group observations are taken at ℓ_i different experimental conditions $t_{i1}, \dots, t_{i\ell_i}$, which vary in the design space (for example the dose range) $\mathcal{X} \subset \mathbb{R}$, and n_{ij} observations are taken at each t_{ij} ($i = 1, 2;\ j = 1, \dots, \ell_i$). Let $n_i = \sum_{j=1}^{\ell_i} n_{ij}$ denote the total number of observations in group i ($= 1, 2$) and $n = n_1 + n_2$ the total sample size. Two regression models m_1 and m_2 with d_1- and d_2-dimensional parameters ϑ_1 and ϑ_2 are used to describe the dependence between response and predictor in the two groups. For asymptotic arguments we assume that $\lim_{n_i \to \infty} \frac{n_{ij}}{n_i} = \xi_{ij} \in (0, 1)$ and collect this information in the matrix

$$\xi_i = \begin{pmatrix} t_{i1} & \dots & t_{i\ell_i} \\ \xi_{i1} & \dots & \xi_{i\ell_i} \end{pmatrix}, \quad i = 1, 2.$$

Following Kiefer (1974) we call ξ_i an approximate design on the design space $\mathcal{X}$. This means that the support points t_{ij} define the distinct experimental conditions where observations are to be taken and the weights ξ_{ij} represent the relative proportion of observations at the corresponding support point t_{ij} (in each group). If an approximate design is given and n_i observations can be taken, a rounding procedure is applied to obtain integers n_{ij} ($i = 1, 2$, $j = 1, \dots, \ell_i$) from the not necessarily integer valued quantities $\xi_{ij} n_i$ [see Pukelsheim and Rieder (1992)]. We note that d_1 and d_2 are determined by the models m_1 and m_2 under consideration and that in this section the sample sizes n_1 and n_2 for the two groups are also fixed. The optimal allocation of patients to the two different groups (for a fixed total sample size) will be discussed in Section 6.1.

Assume that observations are taken according to an approximate design and that an appropriate rounding procedure has been applied. In order to measure the quality of an experimental design we use an asymptotic argument and assume that $\lim_{n_i \to \infty} \frac{n_{ij}}{n_i} = \xi_{ij} \in (0, 1)$. Then, under the common assumptions of regularity, the maximum likelihood estimates, say $\hat{\vartheta}_1, \hat{\vartheta}_2$ in both samples are asymptotically normally distributed (after appropriate standardization). Moreover, the prediction for the difference of the experimental condition t satisfies

$$\sqrt{n}\big(m_1(t,\hat\vartheta_1) - m_2(t,\hat\vartheta_2) - (m_1(t,\vartheta_1) - m_2(t,\vartheta_2))\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0,\varphi(t,\xi_1,\xi_2)\big),$$

where the symbol $\xrightarrow{\mathcal{D}}$ denotes weak convergence, the function φ is defined by

$$\varphi(t,\xi_1,\xi_2) = \frac{\sigma_1^2}{\gamma_1} f_1^T(t) M_1^{-1}(\xi_1,\vartheta_1) f_1(t) + \frac{\sigma_2^2}{\gamma_2} f_2^T(t) M_2^{-1}(\xi_2,\vartheta_2) f_2(t), \tag{2.2}$$

$$M_i(\xi_i,\vartheta_i) = \int_{\mathcal{X}} f_i(t) f_i^T(t) d\xi_i(t)$$

is the information matrix of the design ξ_i in model m_i and $f_i(t) = \frac{\partial}{\partial\vartheta_i} m_i(t,\vartheta_i) \in \mathbb{R}^{d_i}$ is the gradient of m_i with respect to the parameter $\vartheta_i \in \mathbb{R}^{d_i}$ $(i=1,2)$. For these calculations we assume in particular that the limit

$$\gamma_i = \lim_{n\to\infty} \frac{n_i}{n} \in (0,1), \qquad i = 1,2$$

exists and that m_1, m_2 are continuously differentiable with respect to the parameters ϑ_1, ϑ_2. Note that under different distributional assumptions on the errors ε_{ijk} in model (2.1) similar statements can be derived with different covariance matrices in the asymptotic distribution.
Therefore the asymptotic variance of the prediction $m_1(t,\hat\vartheta_1) - \hat m_2(t,\hat\vartheta_2)$ at an experimental condition t is given by $\varphi(t,\xi_1,\xi_2)$, where $\xi = (\xi_1,\xi_2)$ is the pair of designs under consideration. Gsteiger, Bretz and Liu (2011) used this result to obtain a simultaneous confidence band for the difference of the two curves. More precisely, if $\mathcal{Z}$ is a range where the two curves should be compared (note that in contrast to Gsteiger, Bretz and Liu (2011) here the set $\mathcal{Z}$ does not necessarily coincide with the design space $\mathcal{X}$) the confidence band is defined by

$$\hat T \equiv \sup_{t\in\mathcal{Z}} \frac{|m_1(t,\hat\vartheta_1) - m_2(t,\hat\vartheta_2) - (m_1(t,\vartheta_1) - m_2(t,\vartheta_2))|}{\{\frac{\hat\sigma_1^2}{\gamma_1}\hat f_1(t) M_1^{-1}(\xi_1,\hat\vartheta_1)\hat f_1(t) + \frac{\hat\sigma_2^2}{\gamma_2}\hat f_2(t) M_2^{-1}(\xi_2,\hat\vartheta_2)\hat f_2(t)\}^{1/2}} \le D. \tag{2.3}$$

Here, $\hat\sigma_1^2, \hat\sigma_2^2, \hat f_1, \hat f_2$ denote estimates of the quantities $\sigma_1^2, \sigma_2^2, f_1, f_2$, respectively and the constant D is chosen, such that $\mathbb{P}(\hat T \le D) \approx 1-\alpha$. Note that Gsteiger, Bretz and Liu (2011) proposed the parametric bootstrap for this purpose. Consequently, a "good" design, more precisely, a pair $\xi = (\xi_1,\xi_2)$ of two designs on $\mathcal{X}$, should make the width of this band as small as possible at each $t \in \mathcal{Z}$. This corresponds to a simultaneous minimization of

the asymptotic variance in (2.2) with respect to the choice of the designs ξ_1 and ξ_2. Obviously, this is only possible in rare circumstances and we propose to minimize a norm of the function φ as a design criterion. For a precise definition of the optimality criterion we assume that the set $\mathcal{Z}$ contains at least $d \geq \max\{d_1, d_2\}$ points, say $t_1, \ldots, t_d$, such that the vectors $f_1(t_1), \ldots, f_1(t_{d_1})$ and $f_2(t_1), \ldots, f_2(t_{d_2})$ are linearly independent in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively. It then follows, that a pair of designs $\xi = (\xi_1, \xi_2)$, which allows to predict the regression function m_1 and m_2 at all points $t_1, \ldots, t_{d_1}$ and $t_1, \ldots, t_{d_2}$, respectively, must have nonsingular information matrices $M_1(\xi_1, \vartheta_1)$ and $M_2(\xi_2, \vartheta_2)$. Therefore the optimization will be restricted to the class of all designs ξ_1 and ξ_2 with non-singular information matrices throughout this paper.
A worst case criterion is to minimize

$$\mu_\infty(\xi) \;=\; \mu_\infty(\xi_1, \xi_2) = \sup_{t \in \mathcal{Z}} \{\varphi(t, \xi_1, \xi_2)\} \tag{2.4}$$

with respect to $\xi = (\xi_1, \xi_2)$ over a region of interest $\mathcal{Z}$. Alternatively, one could use an L_p-norm

$$\mu_p(\xi) = \mu_p(\xi_1, \xi_2) = \Big(\int_{\mathcal{Z}} \varphi^p(t, \xi_1, \xi_2) d\lambda(t) \Big)^{1/p} \tag{2.5}$$

of the function φ defined in (2.2) with respect to a given measure λ on the region $\mathcal{Z}$ ($p \in [1, \infty)$), where the measure λ has at least $d \geq \max\{d_1, d_2\}$ support points, say $t_1, \ldots, t_d$, such that the vectors $f_1(t_1), \ldots, f_1(t_{d_1})$ and $f_2(t_2), \ldots, f_2(t_{d_2})$ are linearly independent in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively.

Definition 2.1. *For $p \in [1, \infty]$, γ_1, γ_2 fixed, a pair of designs $\xi^{\star,p} = (\xi_1^{\star,p}, \xi_2^{\star,p})$ is called locally μ_p-optimal design (for the comparison of the curves m_1 and m_2) if it minimizes the function $\mu_p(\xi_1, \xi_2)$ over the space of all approximate pairs of designs (ξ_1, ξ_2) on $\mathcal{X} \times \mathcal{X}$ with nonsingular information matrices $M_1(\xi_1, \vartheta_1)$, $M_2(\xi_2, \vartheta_2)$.*

Remark 2.2.

1. The space $\mathcal{Z}$ does not necessarily coincide with the design space $\mathcal{X}$. The special case $\mathcal{Z} \cap \mathcal{X} = \emptyset$ corresponds to the problem of extrapolation and will be discussed in more detail in Section 4.
2. If one requires $\xi_1 = \xi_2$ (for example by logistic reasons) and $\mathcal{Z} = \mathcal{X}$ the criterion μ_∞ is given by

$$\max_{t \in \mathcal{X}} \Big\{ \frac{\sigma_1^2}{\gamma_1} f_1^T(t) M_1^{-1}(\xi, \vartheta_1) f_1(t) + \frac{\sigma_2^2}{\gamma_2} f_2^T(t) M_2^{-1}(\xi, \vartheta_2) f_2(t) \Big\}.$$

It then follows from Theorem 1 in Läuter (1974) that this criterion is equivalent to the weighted D-optimality criterion $(\det M_1(\xi, \vartheta_1))^{\omega_1} (\det M_2(\xi, \vartheta_2))^{\omega_2}$, where the weights are given by $\omega_1 = \frac{\sigma_1^2}{\gamma_1}$ and $\omega_2 = \frac{\sigma_2^2}{\gamma_2}$. Criteria of this type have been studied intensively in the literature [see Lau and Studden (1985), Dette (1990), Zen and Tsai (2004) among others]. Similarly, the criterion μ_1 corresponds to a weighted sum of I-optimality criteria in the case $\mathcal{X} = \mathcal{Z}$.

3. It follows from Minkowski's inequality that in general the pair of the optimal designs for the individual models m_i $(i = 1, 2)$, is not necessarily μ_p-optimal in terms of Definition 2.1.

In some applications it might not be possible to conduct the experiments for both groups simultaneously. This situation arises, for example, in the analysis of clinical trials where data from different sources is available and one trial has already been conducted, while the other is planned in order to compare the corresponding two response curves. In this case only one design (for one group), say ξ_1, can be chosen, while the other is fixed, say η. The corresponding criteria are defined as

$$\nu_p(\xi_1) = \mu_p(\xi_1, \eta), \qquad p \in [1, \infty], \tag{2.6}$$

and ν_p is minimized in the class of all designs on the design space $\mathcal{X}$ with non-singular information matrix $M_1(\xi_1, \vartheta_1)$. The corresponding design minimizing ν_p is called ν_p-optimal throughout this paper.

3. Optimal Design Theory. A main tool of optimal design theory are equivalence theorems which, on the one hand, provide a characterization of the optimal design and, on the other hand, are the basis of many procedures for their numerical construction [see for example Dette, Pepelyshev and Zhigljavsky (2008) or Yu (2010), Yang, Biedermann and Tang (2013)]. Moreover, they are frequently used to reduce the infinite dimensional optimization problems arising in optimal design theory to finite dimensional ones by deriving upper bounds on the number of support points of the optimal design. As the criteria under consideration are convex we can derive corresponding characterizations for the μ_p-criteria. The following two results give the equivalence theorems in the cases $p \in [1, \infty)$ (Theorem 3.1) and $p = \infty$ (Theorem 3.2). These statements are used in Section 5 to check optimality of numerically determined designs. Moreover, Theorem 3.2 is used in an efficient algorithm for the determination of μ_∞-optimal designs in Section 5. Proofs can be found in Section 7. Throughout this paper $\text{supp}(\xi)$ denotes the support of the design ξ on $\mathcal{X}$.

Theorem 3.1. *Let $p \in [1,\infty)$. The design $\xi^{\star,p} = (\xi_1^{\star,p}, \xi_2^{\star,p})$ is μ_p-optimal if and only if the inequality*

$$\int_{\mathcal{Z}} \varphi(t, \xi_1^{\star,p}, \xi_2^{\star,p})^{p-1}\Big(\frac{\gamma_1}{\sigma_1^2}\varphi_1^2(t_1, t, \xi_1^{\star,p}) + \frac{\gamma_2}{\sigma_2^2}\varphi_2^2(t_2,t, \xi_2^{\star,p})\Big)d\lambda(t) - \mu_p^p(\xi_1^{\star,p}, \xi_2^{\star,p}) \leq 0 \tag{3.1}$$

holds for all $t_1, t_2 \in \mathcal{X}$, where

$$\varphi_i(d, t, \xi_i^{\star,p}) = \frac{\sigma_i^2}{\gamma_i} f_i^T(d) M_i^{-1}(\xi_i^{\star,p}, \vartheta_i) f_i(t), \;\; i = 1, 2 \tag{3.2}$$

and the function $\varphi(t, \xi_1^{\star,p}, \xi_2^{\star,p})$ is defined in (2.2). Moreover, equality is achieved in (3.1) for any $(t_1, t_2) \in supp(\xi_1^{\star,p}) \times supp(\xi_2^{\star,p})$.

Theorem 3.2. *The design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ is μ_∞-optimal if and only if there exists a measure $\varrho^\star$ on the set of the extremal points*

$$\mathcal{Z}(\xi^{\star,\infty}) = \Big\{ t_0 \in \mathcal{Z} : \varphi(t_0, \xi_1^{\star,\infty}, \xi_2^{\star,\infty}) = \sup_{t \in \mathcal{Z}} \varphi(t, \xi_1^{\star,\infty}, \xi_2^{\star,\infty}) \Big\} \tag{3.3}$$

of the function $\varphi(t, \xi_1^{\star,\infty}, \xi_2^{\star,\infty})$, such that the inequality

$$\int_{\mathcal{Z}(\xi^{\star,\infty})} \Big(\frac{\gamma_1}{\sigma_1^2}\varphi_1^2(t_1, t, \xi_1^{\star,\infty}) + \frac{\gamma_2}{\sigma_2^2}\varphi_2^2(t_2, t, \xi_2^{\star,\infty})\Big) d\varrho^\star(t) - \mu_\infty(\xi^{\star,\infty}) \leq 0 \tag{3.4}$$

holds for all $t_1, t_2 \in \mathcal{X}$, where the functions φ_1 and φ_2 are defined in (3.2). Moreover, equality is achieved in (3.4) for any $(t_1, t_2) \in supp(\xi_1^{\star,\infty}) \times supp(\xi_2^{\star,\infty})$.

Theorem 3.1 and Theorem 3.2 can be used to check the optimality of a given design. However, in general the explicit calculation of locally μ_p-optimal designs is very difficult. In order to investigate the quality of a (non-optimal) design $\xi = (\xi_1, \xi_2)$ for the purpose of comparing curves, we consider its μ_p-efficiency which is defined by

$$\text{eff}_p(\xi) = \frac{\mu_p(\xi^{\star,p})}{\mu_p(\xi)} \in [0, 1]. \tag{3.5}$$

The following theorem provides a lower bound for the efficiency of a design $\xi = (\xi_1, \xi_2)$ in terms of the functions appearing in the equivalence Theorems 3.1 and 3.2. It is remarkable that this bound does not require knowledge of the optimal design.

THEOREM 3.3. *Let $\xi = (\xi_1, \xi_2)$ be a pair of designs with non singular information matrices $M_1(\xi_1, \vartheta_1)$, $M_2(\xi_2, \vartheta_2)$.*

(a) If $p \in [1, \infty)$, then

$$\mathrm{eff}_p(\xi) \geq \frac{\mu_p^p(\xi)}{\max_{t_1,t_2 \in \mathcal{X}} \int_{\mathcal{Z}} \varphi(t, \xi_1, \xi_2)^{p-1} \big(\frac{\gamma_1}{\sigma_1^2} \varphi_1^2(t, t_1, \xi_1) + \frac{\gamma_2}{\sigma_2^2} \varphi_2^2(t, t_2, \xi_2) \big) d\lambda(t)}. \tag{3.6}$$

(b) If $p = \infty$, then

$$\mathrm{eff}_\infty(\xi) \geq \frac{\mu_\infty(\xi)}{\min_{\varrho \in \Xi(\mathcal{Z}(\xi))} \max_{t_1,t_2 \in \mathcal{X}} \int_{\mathcal{Z}(\xi)} \frac{\gamma_1}{\sigma_1^2} \varphi_1^2(t_1, t, \xi_1) + \frac{\gamma_2}{\sigma_2^2} \varphi_2^2(t_2, t, \xi_2) d\varrho(t)}, \tag{3.7}$$

where $\Xi(\mathcal{Z}(\xi))$ is the set of all measures on $\mathcal{Z}(\xi)$ defined in (3.3).

Roughly speaking the lower bound for the efficiency is the ratio of the two terms in the equivalence Theorem 3.1 (in the case $p < \infty$) and Theorem 3.2 (in the case $p = \infty$). Consequently, for an optimal design the bound is 1 and for a nearly optimal design the bound is close to 1.
Now, we consider the case where one design η is already fixed and the criterion can only be optimized by the other design. The proofs of the following two results are omitted since they are similar to the proofs of Theorems 3.1 and 3.2.

THEOREM 3.4. *Let $p \in [1, \infty)$. The design $\xi_1^{\star,p}$ is ν_p-optimal if and only if the inequality*

$$\int_{\mathcal{Z}} \varphi^{p-1}(t, \xi_1^{\star,p}, \eta) \Big(\frac{\gamma_1}{\sigma_1^2} \varphi_1^2(t_1, t, \xi_1^{\star,p}) + \varphi_2(t, t, \eta) \Big) d\lambda(t) - \nu_p^p(\xi_1^{\star,p}) \leq 0 \tag{3.8}$$

holds for all $t_1 \in \mathcal{X}$, where φ_i and φ are defined in (3.2) and (2.2), respectively. Moreover, equality is achieved in (3.8) for any $t_1 \in supp(\xi_1^{\star,p})$.

THEOREM 3.5. *The design $\xi_1^{\star,\infty}$ is ν_∞-optimal if and only if there exists a measure $\varrho^\star$ on the set of the extremal points*

$$\mathcal{Z}(\xi_1^{\star,\infty}) = \Big\{ t_0 \in \mathcal{Z} : \varphi(t_0, \xi_1^{\star,\infty}, \eta) = \sup_{t \in \mathcal{Z}} \varphi(t, \xi_1^{\star,\infty}, \eta) \Big\}$$

of the function $\varphi(t, \xi_1^{\star,\infty}, \eta)$, such that the inequality

$$\int_{\mathcal{Z}(\xi_1^{\star,\infty})} \frac{\gamma_1}{\sigma_1^2} \varphi_1^2(t_1, t, \xi_1^{\star,\infty}) d\varrho^\star(t) - \int_{\mathcal{Z}(\xi_1^{\star,\infty})} \varphi_1(t, t, \xi_1^{\star,\infty}) d\varrho^\star(t) \leq 0 \tag{3.9}$$

holds for all $t_1 \in \mathcal{X}$, where the functions φ_1 is defined in (3.2). Moreover, equality is achieved in (3.9) for any $t_1 \in supp(\xi_1^{\star,\infty})$.

4. Some analytical results - extrapolation. In this section we present some analytical results which illustrate the difficulties in determining designs for the comparison of curves explicitly. These results can also be used to check the accuracy and speed of convergence of the developed algorithms (as the solutions are known). To be precise, we consider the criterion μ_∞ and the case where the design space $\mathcal{X}$ and the space $\mathcal{Z}$ do not intersect, which corresponds to the problem of comparing two curves for extrapolation. In general, extrapolation is not an easy task and has to be addressed very carefully, because it is not clear if the postulated relation between response and predictor holds also in regions, where no data is available. However, in dose response studies (such as Phase II clinical trials or studies in toxicology) experimenters usually have information about the functional form describing this relation. Often models appear as solutions of differential equations which are used to describe chemical reactions. In such cases extrapolation is well justified. Moreover, in toxicology there are many cases where it is in fact necessary to do a reasonable extrapolation, because patients cannot be treated with too high doses.

We are particularly interested in the difference between curves modeled by the Michaelis Menten, Emax and loglinear model. It turns out that the results for these models can be easily obtained from a general result for weighted polynomial regression models, which is of own interest and will be considered first. For this purpose assume that the design space $\mathcal{X}$ and the range $\mathcal{Z}$ are intervals, that is $\mathcal{X} = [L_\mathcal{X}, U_\mathcal{X}]$, $\mathcal{Z} = [L_\mathcal{Z}, U_\mathcal{Z}]$ and that both regression models m_1 and m_2 are given by functions of the type

$$m_i(t) = \omega_i(t) \sum_{j=0}^{p_i} \vartheta_{ij}\, t^j, \quad i = 1, 2, \tag{4.1}$$

where ω_1, ω_2 are known positive weight functions on $\mathcal{X} \cup \mathcal{Z}$. The models m_1, m_2 are called weighted polynomial regression models and in the case of one model several design problems have been discussed in the literature, mainly for the D- and E-optimality criterion [see for example Dette (1993), Heiligers (1994), Antille and Weinberg (2003), Chang (2005a,b) or Dette and Trampisch (2010)]. It is easy to show that the systems $\{\omega_i(t)t^j | j = 0, \ldots, p_i\}$ are Chebyshev systems on the convex hull of $\mathcal{X} \cup \mathcal{Z}$, say $\text{conv}(\mathcal{X} \cup \mathcal{Z})$, which means that for any choice $\vartheta_{i0}, \ldots, \vartheta_{ip_i}$ the equation $\omega_i(t) \sum_{j=0}^{p_i} \vartheta_{ij} t^j = 0$ has at most p_i solutions in $\text{conv}(\mathcal{X} \cup \mathcal{Z})$ [see Karlin and Studden (1966)]. It then follows from this reference that there exist unique polynomials $\underline{v}_i(t) = \omega_i(t) \sum_{j=0}^{p_i} a_{ij} t^j$, $i = 1, 2$ satisfying the properties

1. for all $t \in \mathcal{X}$ the inequality $|\underline{v}_i(t)| \le 1$ holds.

2. there exist p_i+1 points $L_{\mathcal{X}} \le t_{i0} < t_{i1} < \ldots < t_{ip_i} \le U_{\mathcal{X}}$ such that $\underline{v}_i(t_{ij}) = (-1)^j$ for $j = 0, \ldots, p_i$.

The points $t_{i0}, \ldots, t_{ip_i}$ are called Chebyshev points while $\underline{v}_i$ is called Chebyshev or equioscillating polynomial. The following results give an explicit solution of the μ_∞-optimal design problem if the functions m_1 and m_2 are weighted polynomials.

THEOREM 4.1. *Consider the weighted polynomials (4.1) with differentiable, positive weight functions ω_1, ω_2 such that for $\omega_i(t) \ne c \in \mathbb{R}$ $\{1, \omega_i(t), \omega_i(t)t, \ldots, \omega_i(t)t^{2p_i-1}\}$ and $\{1, \omega_i(t), \omega_i(t)t, \ldots, \omega_i(t)t^{2p_i}\}$ are Chebshev systems ($i = 1, 2$). Assume that $\mathcal{X} \cap \mathcal{Z} = [L_{\mathcal{X}}, U_{\mathcal{X}}] \cap [L_{\mathcal{Z}}, U_{\mathcal{Z}}] = \emptyset$.*

1. *If $U_{\mathcal{X}} < L_{\mathcal{Z}}$ and ω_1, ω_2 are strictly increasing on $\mathcal{Z}$, the support points of the μ_∞- optimal design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ are given by the extremal points of the Chebyshev polynomial $\underline{v}_1(t)$ for $\xi_1^{\star,\infty}$ and $\underline{v}_2(t)$ for $\xi_2^{\star,\infty}$ with corresponding weights*

$$\xi_{ij} = \frac{|L_{ij}(U_{\mathcal{Z}})|}{\sum_{k=0}^{p_i} |L_{ik}(U_{\mathcal{Z}})|} \quad j = 0, \ldots, p_i, \quad i = 1, 2. \tag{4.2}$$

   Here $L_{ij}(t) = \omega_i(t) \sum_{j=0}^{p_i} \ell_{ij} t^j$ is the j-th Lagrange interpolation polynomial with knots $t_{i0}, \ldots, t_{ip_i}$, $i = 1, 2$ defined by the properties $L_{ij}(t_{ik}) = \delta_{jk}$, $j, k = 1, \ldots, p_i$ (and δ_{jk} denotes the Kronecker symbol).
2. *If $L_{\mathcal{X}} > U_{\mathcal{Z}}$ and ω_1, ω_2 are strictly decreasing on $\mathcal{Z}$, the support points of the μ_∞-optimal design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ are given by the extremal points of the Chebyshev polynomial $\underline{v}_1(t)$ for $\xi_1^{\star,\infty}$ and $\underline{v}_2(t)$ for $\xi_2^{\star,\infty}$ with corresponding weights*

$$\xi_{ij} = \frac{|L_{ij}(L_{\mathcal{Z}})|}{\sum_{k=0}^{p_i} |L_{ik}(L_{\mathcal{Z}})|}, \quad j = 0, \ldots, p_i, \quad i = 1, 2.$$

REMARK 4.2. It is worthwhile to mention that for general $p \ne \infty$ the μ_p-optimal designs have to be found numerically if the degree of the polynomials is larger than 2. The situation is similar as in the problem of calculating optimal designs with respect to Kiefer's Φ_p-criteria for (unweighted) polynomial regression models. Only in the cases $p = 0$ and $p = \infty$ corresponding to the D- and E-criterion explicit results are available [see Pukelsheim (2006)]. The μ_p-optimal design problems are even harder and only the μ_∞-optimal designs can be found explicitly for weighted polynomial regression models.

EXAMPLE 4.3. If both regression models m_1 and m_2 are given by polynomials of degree p_1 and p_2, we have $\omega_1 \equiv \omega_2 \equiv 1$ and the μ_∞-optimal design can be described even more explicitly. For the sake of brevity we only consider the case $U_{\mathcal{X}} < L_{\mathcal{Z}}$. According to Theorem 4.1 $\xi_1^{\star,\infty}$ and $\xi_2^{\star,\infty}$ are supported at the extremal points of the polynomials $\underline{v}_1(t)$ and $\underline{v}_2(t)$. If $\omega_1 \equiv \omega_2 \equiv 1$ these are given by the Chebyshev polynomials of the first kind on the interval $[L_{\mathcal{X}}, U_{\mathcal{X}}]$, that is

$$\underline{v}_1(t) = T_{p_1}\left(\frac{2t-(U_{\mathcal{X}}+L_{\mathcal{X}})}{U_{\mathcal{X}}-L_{\mathcal{X}}}\right) \quad \text{and} \quad \underline{v}_2(t) = T_{p_2}\left(\frac{2t-(U_{\mathcal{X}}+L_{\mathcal{X}})}{U_{\mathcal{X}}-L_{\mathcal{X}}}\right),$$

where $T_p(x) = \cos(p \arccos x)$, $x \in [-1,1]$. Consequently, the component $\xi_i^{\star,\infty}$ of the optimal design is supported at the $p_i + 1$ Chebyshev points

$$t_{ij} = \frac{(1-\cos(\frac{j}{p_i}\pi))U_{\mathcal{X}} + (1+\cos(\frac{j}{p_i}\pi))L_{\mathcal{X}}}{2}, \quad j = 0, \ldots, p_i$$

with corresponding weights

$$\xi_{ij} = \frac{|L_{ij}(U_{\mathcal{Z}})|}{\sum_{k=0}^{p_i} |L_{ik}(U_{\mathcal{Z}})|}, \quad j = 0, \ldots, p_i \tag{4.3}$$

where

$$L_{ij}(t) = \prod_{k=0, k\neq j}^{p_i} \frac{t - t_{ik}}{t_{ij} - t_{ik}}$$

is the Lagrange interpolation polynomial at the knots $t_{i0}, \ldots, t_{ip_i}$.

While Theorem 4.1 and Example 4.3 are of own interest, they turn out to be particularly useful to find μ_∞-optimal designs for some commonly used dose response models. To be precise we consider the Michaelis Menten model

$$m(t, \vartheta) = \frac{\vartheta_1 t}{\vartheta_2 + t} \tag{4.4}$$

the loglinear model with fixed parameter ϑ_3

$$m(t, \vartheta) = \vartheta_1 + \vartheta_2 \log(t + \vartheta_3) \tag{4.5}$$

and the Emax model

$$m(t, \vartheta) = \vartheta_1 + \frac{\vartheta_2 t}{\vartheta_3 + t}. \tag{4.6}$$

The following result specifies the μ_∞-optimal designs for the comparison of curves if $\mathcal{X} \cap \mathcal{Z} = \emptyset$ and m_1 and m_2 are given by any of these models.

Corollary 4.4. *Assume that the regression models m_1 and m_2 are given by one of the models (4.4) - (4.6), $L_{\mathcal{X}} \geq 0$ and $U_{\mathcal{X}} < L_{\mathcal{Z}}$. The μ_∞-optimal design is given by $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$, where $\xi_i^{\star,\infty}$ is given by*

$$\xi^{\star,\infty}_{MM} = \begin{pmatrix} \frac{\vartheta_2 U_{\mathcal{X}}(\sqrt{2}-1)}{(2-\sqrt{2})U_{\mathcal{X}}+\vartheta_2} & U_{\mathcal{X}} \\ \frac{\vartheta_2(U_{\mathcal{Z}}-U_{\mathcal{X}})}{U_{\mathcal{X}}U_{\mathcal{Z}}(3\sqrt{2}-4)+\vartheta_2(\sqrt{2}U_{\mathcal{Z}}-(4-2\sqrt{2})U_{\mathcal{X}})} & \frac{(\sqrt{2}-1)\left[(2-\sqrt{2})U_{\mathcal{X}}U_{\mathcal{Z}}+\vartheta_2(U_{\mathcal{Z}}-(\sqrt{2}-1)U_{\mathcal{X}})\right]}{U_{\mathcal{X}}U_{\mathcal{Z}}(3\sqrt{2}-4)+\vartheta_2\left[\sqrt{2}U_{\mathcal{Z}}-(4-2\sqrt{2})U_{\mathcal{X}}\right]} \end{pmatrix},$$

if m_i is the Michaelis Menten model and $\frac{\vartheta_2 U_{\mathcal{X}}(\sqrt{2}-1)}{(2-\sqrt{2})U_{\mathcal{X}}+\vartheta_2} \geq L_{\mathcal{X}} > 0$, by

$$\xi^{\star,\infty}_{LogLin} = \begin{pmatrix} L_{\mathcal{X}} & U_{\mathcal{X}} \\ \frac{\log(U_{\mathcal{Z}}+\vartheta_3)-\log(U_{\mathcal{X}}+\vartheta_3)}{2\log(U_{\mathcal{Z}}+\vartheta_3)-(\log(L_{\mathcal{X}}+\vartheta_3)+\log(U_{\mathcal{X}}+\vartheta_3))} & \frac{\log(U_{\mathcal{Z}}+\vartheta_3)-\log(L_{\mathcal{X}}+\vartheta_3)}{2\log(U_{\mathcal{Z}}+\vartheta_3)-(\log(L_{\mathcal{X}}+\vartheta_3)+\log(U_{\mathcal{X}}+\vartheta_3))} \end{pmatrix},$$

if m_i is the loglinear model and by

$$\xi^{\star,\infty}_{Emax} = \begin{pmatrix} L_{\mathcal{X}} & \frac{2U_{\mathcal{X}}L_{\mathcal{X}}+(U_{\mathcal{X}}+L_{\mathcal{X}})\vartheta_3}{2\vartheta_3+U_{\mathcal{X}}+L_{\mathcal{X}}} & U_{\mathcal{X}} \\ \frac{(g(U_{\mathcal{Z}},U_{\mathcal{X}})+g(U_{\mathcal{Z}},L_{\mathcal{X}}))g(U_{\mathcal{Z}},U_{\mathcal{X}})}{L} & \frac{4g(U_{\mathcal{Z}},U_{\mathcal{X}})g(U_{\mathcal{Z}},L_{\mathcal{X}})}{L} & \frac{(g(U_{\mathcal{Z}},U_{\mathcal{X}})+g(U_{\mathcal{Z}},L_{\mathcal{X}}))g(U_{\mathcal{Z}},L_{\mathcal{X}})}{L} \end{pmatrix}$$

if m_i is the Emax model. Here the function g is defined by $g(a,b) = \frac{a}{a+\vartheta_3} - \frac{b}{b+\vartheta_3}$ and L is a normalizing constant, that is $L = g^2(U_{\mathcal{Z}}, U_{\mathcal{X}}) + 6g(U_{\mathcal{Z}}, U_{\mathcal{X}})g(U_{\mathcal{Z}}, L_{\mathcal{X}}) + g^2(U_{\mathcal{Z}}, L_{\mathcal{X}})$.

5. Numerical results.

In most cases of practical interest the μ_p-optimal designs have to be found numerically. In the case $p < \infty$ the optimality criteria are in fact differentiable. In this case - as the criteria under consideration are convex - several procedures from convex optimization theory can be used for this purpose, which have been adapted to the specific optimization problems (such as no upper bound on the dimension) occurring in the determination of optimal experimental designs [see Dette, Pepelyshev and Zhigljavsky (2008), Yang (2010) or Yang, Biedermann and Tang (2013)]. In particular the optimality of the numerically constructed designs can be easily checked using the equivalence Theorem 3.1. For this reason we concentrate on the case $p = \infty$ which is also probably of most practical interest, because it directly refers to the maximum width of the confidence band. The μ_∞-optimality criterion is not necessarily differentiable. As a consequence there appears the unknown measure $\varrho^\star$ in Theorem 3.2, which has also to be calculated in order to check the μ_∞-optimality of a given design (or to obtain a tight lower bound for its efficiency by an application of 3.3). Former algorithms for minimax optimal design problems are based on analogues of Theorem 3.2 such that the measure $\varrho^\star$ has to be calculated simultaneously with the optimal design [see for example Wong and Cook (1993)]. We now derive an alternative procedure using the Particle Swarm Optimization (PSO), which calculates the μ_∞-optimal design and the corresponding measure $\varrho^\star$ consecutively. For this purpose recall the definition of $\tilde{\varphi}_i$ in (3.2), and consider an arbitrary design $\xi = (\xi_1, \xi_2)$ and an arbitrary measure ϱ defined on the set of the extremal points $\mathcal{Z}(\xi)$, then the following inequality holds

$$
\begin{aligned}
& \max_{t_1,t_2\in\mathcal{X}} \int_{\mathcal{Z}(\xi)} \big(\tfrac{\gamma_1}{\sigma_1^2}\varphi_1^2(t_1,t,\xi_1) + \tfrac{\gamma_2}{\sigma_2^2}\varphi_2^2(t_2,t,\xi_2)\big)d\varrho(t) \\
\geq\ & \int_{\mathcal{X}}\int_{\mathcal{Z}(\xi)} \tfrac{\gamma_1}{\sigma_1^2}\varphi_1^2(t_1,t,\xi_1)d\varrho(t)d\xi_1(t_1) + \int_{\mathcal{X}}\int_{\mathcal{Z}(\xi)} \tfrac{\gamma_2}{\sigma_2^2}\varphi_2^2(t_2,t,\xi_2)d\varrho(t)d\xi_2(t_2) \\
=\ & \int_{\mathcal{Z}(\xi)} \varphi(t,\xi_1,\xi_2)d\varrho(t) = \mu_\infty(\xi).
\end{aligned}
$$

On the other hand it follows from the equivalence Theorem 3.2 that the opposite inequality also holds for the μ_∞-optimal design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ and the corresponding measure $\varrho^\star$ on $\mathcal{Z}(\xi^{\star,\infty})$ [see inequality (3.4)]. Consequently, the measure $\varrho^\star$ is the measure on $\mathcal{Z}(\xi^{\star,\infty})$ which minimizes the function

$$
\begin{aligned}
N_\infty(\varrho,\xi^{\star,\infty}) &= \max_{t_1,t_2\in\mathcal{X}} \int_{\mathcal{Z}(\xi^{\star,\infty})} \big(\tfrac{\gamma_1}{\sigma_1^2}\varphi_1^2(t_1,t,\xi_1^{\star,\infty}) + \tfrac{\gamma_2}{\sigma_2^2}\varphi_2^2(t_2,t,\xi_2^{\star,\infty})\big)d\varrho(t) \\
&= \max_{t_1\in\mathcal{X}} \tfrac{\sigma_1^2}{\gamma_1} f_1^T(t_1) M_1^{-1}(\xi_1^{\star,\infty}) M_1(\varrho) M_1^{-1}(\xi_1^{\star,\infty}) f_1(t_1) \\
&+ \max_{t_2\in\mathcal{X}} \tfrac{\sigma_2^2}{\gamma_2} f_2^T(t_2) M_2^{-1}(\xi_2^{\star,\infty}) M_2(\varrho) M_2^{-1}(\xi_2^{\star,\infty}) f_2(t_2).
\end{aligned} \tag{5.1}
$$

The μ_∞-optimal design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ and the corresponding measure $\varrho^\star$ for the equivalence theorems are now calculated numerically in four consecutive steps using Particle Swarm Optimization (PSO) [see for example Clerc (2006)]:

1. We calculate the μ_∞-optimal design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ using PSO.
2. We calculate numerically the set of extremal points $\mathcal{Z}(\xi^{\star,\infty}) = \{z_1, \ldots, z_k\}$ of the function $\varphi(t, \xi_1^{\star,\infty}, \xi_2^{\star,\infty})$.
3. We calculate numerically the measure $\varrho^\star$ on $\mathcal{Z}(\xi^{\star,\infty}) = \{z_1, \ldots, z_k\}$ which minimizes the function $N_\infty(\varrho, \xi^{\star,\infty})$ defined in (5.1) using PSO.
4. We check the optimality of the design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ calculated in step 1 by an application of Theorem 3.2 using the measure $\varrho^\star$ from Step 3.

TABLE 1
Commonly used dose response models with their parameter specifications [from Bretz, Pinheiro and Branson (2005)].

model	$m(t, \vartheta)$	parameters
Emax	$\vartheta_1 + \frac{\vartheta_2 t}{t+\vartheta_3}$	$(0.2, 0.7, 0.2)$
exponential	$\vartheta_1 + \vartheta_2 \exp(t/\vartheta_3)$	$(0.183, 0.017, 0.28)$
loglinear	$\vartheta_1 + \vartheta_2 \log(t + \vartheta_3)$	$(0.74, 0.33, 0.2)$

In its original form the PSO is a metaheuristic algorithm whose convergence can not be proved. However, there exist several modifications of the method, such that convergence can established mathematically [see for example van den Bergh and Engelbrecht (2010) or Bonyadi and Michalewicz (2014) among others]. In our implementation we did not use any modification of this type, but we added Step 4 in the procedure to check the derived designs for optimality by an application of the equivalence Theorem 3.2. Due to the convexity of the corresponding optimization problems this procedure is very reliable and the PSO algorithm with 400 particles and 250 iterations was able to find the μ_∞-optimal design with the required accuracy in all considered examples. This usually requires about 8 minutes cpu time on a standard PC.
In the following discussion we consider the exponential, loglinear and Emax model with their corresponding parameter specifications depicted in Table 1. These models have been proposed by Bretz, Pinheiro and Branson (2005) as a selection of commonly used models to represent dose response relationships on the dose range $[0, 1]$. These authors also proposed a design which allocates 20% of the patients to the dose levels 0, 0.05, 0.2, 0.6 and 1, and which will be called standard design in the following discussion. We consider μ_∞-optimal designs for the three combinations of these models, where the design space and the region of interest are given by $\mathcal{X} = \mathcal{Z} = [0, 1]$. The variances σ_1^2 and σ_2^2 are equal and given by $\sigma^2 = 1.478^2$ as proposed in Bretz, Pinheiro and Branson (2005) and we assume $\gamma_1 = \gamma_2 = 0.5$. The resulting μ_∞-optimal designs are displayed in Table 2. In the diagonal blocks we have two identical designs reflecting the fact that in this case $m_1 = m_2$. These designs are actually the D-optimal designs for the corresponding common model, which follows by a straightforward application of the famous equivalence theorem for D- and G-optimal designs [see Kiefer and Wolfowitz (1960)].
In the other cases the optimal designs are obtained from Table 2 as follows. For example, the μ_∞-optimal design for the combination of the Emax (m_1) and the exponential model (m_2) can be obtained from the right upper block.

Table 2
μ_∞-optimal designs for different model combinations. Upper rows: support points. Lower rows: weights given in percent (%).

m_1/ m_2	Emax			loglinear			exponential		
Emax	0.00	0.14	1.00	0.00	0.22	1.00	0.00	0.74	1.00
	$33.\bar{3}$	$33.\bar{3}$	$33.\bar{3}$	34.0	32.5	33.5	40.3	27.4	32.3
	0.00	0.14	1.00	0.00	0.15	1.00	0.00	0.15	1.00
	$33.\bar{3}$	$33.\bar{3}$	$33.\bar{3}$	33.4	32.7	33.9	32.0	28.2	39.8
loglinear				0.00	0.23	1.00	0.00	0.74	1.00
				$33.\bar{3}$	$33.\bar{3}$	$33.\bar{3}$	39.2	26.8	34.0
				0.00	0.23	1.00	0.00	0.24	1.00
				$33.\bar{3}$	$33.\bar{3}$	$33.\bar{3}$	33.5	27.8	38.7
exponential							0.00	0.75	1.00
							$33.\bar{3}$	$33.\bar{3}$	$33.\bar{3}$
							0.00	0.75	1.00
							$33.\bar{3}$	$33.\bar{3}$	$33.\bar{3}$

The first component is the design for the exponential model, which allocates 40.3% , 27.4% , 32.3% of the patients to the dose levels 0.00, 0.74, 1.00. The second component is the design for the Emax model which allocates 32.0% , 28.2%, 39.8% of the patients to the dose levels 0.00, 0.15, 1.00.

In Figure 1 we demonstrate the application of the equivalence Theorem 3.2 for the combinations Emax and exponential model and exponential and loglinear model.
The advantages of the new designs are illustrated in Figure 2, where we present the improvement of the confidence bands proposed by Gsteiger, Bretz and Liu (2011) for the difference between the two regression functions if the μ_∞-optimal design is used instead of a pair of the standard designs. The sample sizes in both groups are $n_1 = 100$ and $n_2 = 100$, respectively. The presented confidence bands are the averages of uniform confidence bands calculated by 100 simulation runs. We observe that inference on the basis of an μ_∞-optimal design yields a substantial reduction in the (maximal) width of the confidence band.

It was also pointed out by a referee that it might be of interest to investigate the sensitivity of this improvement with respect to misspecification of the parameters in the locally μ_∞-optimality criterion. Exemplarily we consider the locally μ_∞-optimal design for the combination for the Emax and the exponential model. The μ_∞-optimal design has been constructed for the parameter constellation given in Table 1, whereas the actual "true" parameters for the Emax and exponential model are given by $\vartheta_{Emax} = (0.1, 0.35, 0.1)^T$

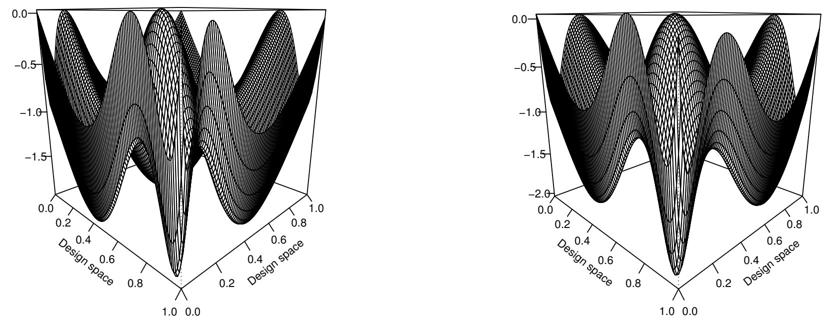

FIG 1. *Illustration of Theorem 3.2. The figures show the function on the left hand side of inequality (3.4). Left figure: The combination of exponential and Emax model. Right figure: The combination of the loglinear and the exponential model.*

TABLE 3
The μ_∞-efficiencies (in %) of the standard design, of the pairs of D-optimal designs (displayed in the diagonal blocks of Table 2) and of the robust optimal design (cf. Section 6.2).

model 1 / model 2	loglin / exp	loglin / Emax	exp / Emax
standard design	58.85	72.83	59.00
D-optimal designs for Emax	2.21	93.81	2.24
D-optimal designs for loglinear	7.31	92.44	7.40
D-optimal designs for exponential	15.08	3.72	4.29
robust optimal design	67.30	89.65	69.57

and $\vartheta_{exp} = (0.1, 0.05, 0.167)^T$ (scenario A) and by $\vartheta_{Emax} = (0.4, 0.7, 0.4)^T$ and $\vartheta_{exp} = (0.4, 0.2, 0.66)^T$ (scenario B), respectively. In Figure 3 we compare the resulting confidence bands obtained from the μ_∞-optimal design (with misspecified parameters) with those obtained from the standard design. We observe that - despite the misspecification of the parameters - the μ_∞-optimal design yields substantially narrower confidence bands in both scenarios. Further investigations, which are not reported for the sake of brevity, showed a similar picture and we conclude that the μ_∞-optimal design is robust against (moderate) misspecification of the parameters. The construction of robust designs with respect to extreme misspecification of the parameters will be discussed in Section 6.2.
Besides the comparison of the different confidence bands produced by the μ_∞-optimal design and the standard design proposed in Bretz, Pinheiro and

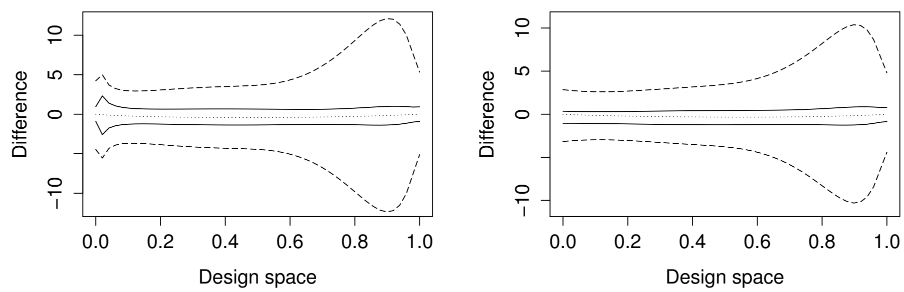

FIG 2. *Confidence bands obtained from the μ_∞-optimal design (solid lines) and a standard design (dashed lines). The dotted line shows the true difference of the curves. Left figure: The combination of exponential and Emax model. Right figure: The combination of the loglinear and the exponential model.*

Branson (2005) we can also compare different designs using the efficiency defined by (3.5). For example the efficiencies of the μ_∞-optimal design for the combination for the Emax and the exponential model with misspecified parameters are given by 79.93% (for scenario A) and by 80.09% (for scenario B), while the standard design has efficiencies 27.31% and 63.32% in these cases.

A more detailed comparison of the designs for different models (with correctly specified parameters) is given in Table 3, where we also investigate the problem of model misspecification. If the model is correctly specified we observe a substantial loss of efficiency if the standard design is used instead of a μ_∞-optimal design. In row 2-4 of Table 3 we show the corresponding efficiencies of the μ_∞-optimal design, if these designs are used for the comparison of different curves. For example, the μ_∞-optimal design for two Emax models has μ_∞-efficiencies 2.21%, 93.81% and 2.24%, if it is used for the comparison of the loglinear and exponential, the loglinear and Emax and the exponential and Emax model, respectively. The results indicate that the optimal designs are sensitive with respect to misspecification of the parametric form of the regression functions, and that they do not necessarily improve the standard design in such cases. In the last row of Table 3 we display the efficiencies of a robust design which will be constructed by the methodology developed in Section 6.2. We observe an improvement of the standard design in all considered scenarios.

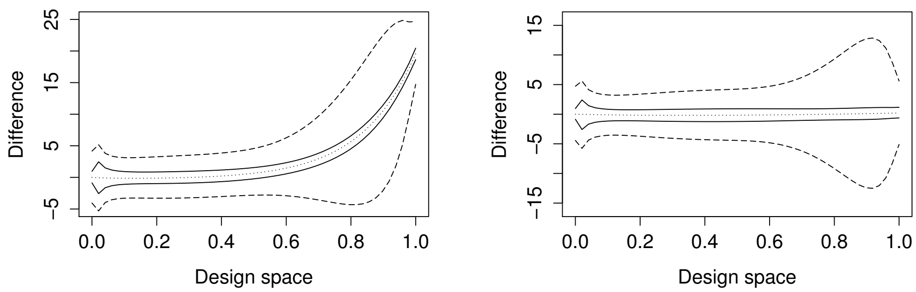

FIG 3. *Confidence bands obtained from the standard design (dashed lines) and the μ_∞-optimal design for the Emax and the exponential model under misspecification of the model parameters (solid lines). The optimal designs have been calculated for the parameters given in Table 1, where the actual "true" parameters are given by $\vartheta_{Emax} = (0.1, 0.35, 0.1)^T$ and $\vartheta_{exp} = (0.1, 0.05, 0.167)^T$ (left panel) and by $\vartheta_{Emax} = (0.4, 0.7, 0.4)^T$ and $\vartheta_{exp} = (0.4, 0.2, 0.66)^T$ (right panel).*

6. Further discussion.

6.1. *Optimal allocation to the two groups.* So far we have assumed that the sample sizes n_1 and n_2 in the two groups are fixed and cannot be chosen by the experimenter. In this section we will briefly indicate some results, if optimization can also be performed with respect to the relative proportions $\gamma_1 = n_1/(n_1 + n_2)$ and $\gamma_2 = n_2/(n_1 + n_2)$ for the two groups. Following the approximate design approach we define γ as a probability measure with masses γ_1 and γ_2 at the points 0 and 1, respectively, and a μ_∞-optimal design as a triple $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star)$, which minimizes the functional

$$\mu_\infty(\xi_1, \xi_2, \gamma) = \sup_{t \in \mathcal{Z}} \varphi(t, \xi_1, \xi_2, \gamma),$$

where

$$\varphi(t, \xi_1, \xi_2, \gamma) = \frac{\sigma_1^2}{\gamma_1} f_1^T(t) M_1^{-1}(\xi_1, \vartheta_1) f_1(t) + \frac{\sigma_1^2}{\gamma_2} f_2^T(t) M_2^{-1}(\xi_2, \vartheta_2) f_2(t).$$

Similar arguments as given in the proof of Theorem 3.1 give a characterization of the optimal designs. The details are omitted for the sake of brevity.

THEOREM 6.1. *A design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star)$ is μ_∞-optimal if and*

Table 4
The μ_∞-optimal design $(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star)$ for the comparison of the Emax and the exponential model, where optimization is also performed with respect to the relative sample sizes $\gamma = (\gamma_1, \gamma_2)$ for the two groups. The weights are given in %.

γ^*	$\xi_1^{\star,\infty}$			$\xi_2^{\star,\infty}$		
(30.2, 69.8)	0.00	0.15	1.00	0.00	0.75	1.00
	32.4	24.9	42.7	36.9	30.4	32.7

only if there exists a measure $\varrho^\star$ on the set

$$\mathcal{Z}(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star) = \{t \in \mathcal{Z} : \mu_\infty(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star) = \varphi(t, \xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star)\}$$

such that the inequality

$$\int_{\mathcal{Z}(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star)} \frac{I\{\omega=0\}}{\sigma_1^2}\varphi_1^2(t, t_1, \xi_1^{\star,\infty}) + \frac{I\{\omega=1\}}{\sigma_2^2}\varphi_2^2(t, t_2, \xi_2^{\star,\infty})\, d\varrho^\star(t) - \mu_\infty(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star) \leq 0 \tag{6.1}$$

is satisfied for all $t_1, t_2 \in \mathcal{X}$ and $\omega \in \{0,1\}$, where φ_i is defined in (3.2) with $\gamma_i = \gamma_i^\star$. Moreover, equality is achieved in (3.4) for any $(t_1, t_2, \omega) \in$ supp$(\xi_1^{\star,\infty}) \times$ supp$(\xi_2^{\star,\infty}) \times \{0,1\}$.

Example 6.2. The μ_∞-optimal design $(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \gamma^\star)$ can be determined numerically in a similar way as described in Section 5, and we briefly illustrate some results for the comparison of the Emax model with the exponential model, where the parameters are given in Table 1. The variances are $\sigma_1^2 = 1.478^2$ in the first group and $\sigma_2^2 = 5 \cdot 1.478^2$ in the second group and the optimal designs (calculated by the PSO) are presented in Table 4. Note that the optimal design allocates only 30.2% of the observations to the first group. A comparison of the optimal designs from Table 4 with the corresponding optimal designs from Table 2 (calculated under the assumptions $\sigma_1^2 = \sigma_2^2$ and $\gamma_1 = \gamma_2 = 0.5$) shows that the support points are very similar, but there appear differences in the weights.

6.2. *Robustness.* For the sake of transparency the discussion presented so far considers locally optimal designs [see Chernoff (1953)] for "known" models in the two samples. Besides the specification of the models these designs require a-priori information about the corresponding parameters. In several situations preliminary knowledge regarding the model and/or unknown parameters is available. A typical example are phase II clinical dose finding trials, where some useful knowledge regarding model and corresponding parameters is already available from phase I [see Dette et al. (2008)].

Moreover, these designs can be applied as benchmarks for commonly used designs, and locally optimal designs serve as basis for constructing optimal designs with respect to more sophisticated optimality criteria, which are efficient and robust against model assumptions [see Läuter (1974), Dette (1990), Chaloner and Verdinelli (1995), Dette (1997) among others].
In this section we briefly indicate how the methodology introduced in the previous sections can be further developed to address uncertainty with respect to the model assumptions. For the sake of brevity we restrict ourselves to the μ_∞-criterion and note that similar results as presented in this section can be obtained for the criterion (2.5). In order to reflect the dependence of the criterion (2.4) on the regression functions m_1, m_2, the parameters ϑ_1, ϑ_2 and the variances σ_1^2, σ_2^2 in our notation we will use the notation $\mu_\infty^{1,2}(\xi_1, \xi_2, \vartheta_1, \vartheta_2, \sigma_1^2, \sigma_2^2)$ for the criterion introduced in equation (2.4). Similarly, we denote the efficiency introduced in (3.5) by

$$\text{eff}_\infty^{1,2}(\xi, \vartheta_1, \vartheta_2, \sigma_1^2, \sigma_2^2) = \frac{\mu_\infty^{1,2}(\xi^{1,2,\star}_{\vartheta_1,\vartheta_2,\sigma_1^2,\sigma_2^2}, \vartheta_1, \vartheta_2, \sigma_1^2, \sigma_2^2)}{\mu_\infty^{1,2}(\xi, \vartheta_1, \vartheta_2, \sigma_1^2, \sigma_2^2)} \tag{6.2}$$

where $\xi^{1,2,\star}_{\vartheta_1,\vartheta_2,\sigma_1^2,\sigma_2^2}$ is the locally μ_∞-optimal design minimizing the functional $\mu_\infty^{1,2}(\cdot, \vartheta_1, \vartheta_2, \sigma_1^2, \sigma_2^2)$ (for fixed models m_1, m_2, fixed parameters ϑ_1, ϑ_2 and fixed variances σ_1^2, σ_2^2). We assume that p models, say $m_1, \ldots, m_p$, are available to describe the relation between predictor and response in both groups. We address uncertainty with respect to the parameters in model m_k by a prior distribution, say π_k, for the corresponding parameter $\vartheta_k \in \Theta_k \subset \mathbb{R}^{d_k}$ and with respect to the variance σ_i^2 by a prior distribution on $\mathbb{R}^+$, say τ_i $(i = 1, 2)$. A design is called robust optimal design for the comparison of the two curves if it maximizes the functional

$$\Phi(\xi) = \sum_{k,l=1}^{p} \alpha_{k,l} \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \int_{\Theta_k} \int_{\Theta_l} \text{eff}_\infty^{k,l}(\xi, \vartheta_k, \vartheta_l, \sigma_1^2, \sigma_2^2) \pi_k(d\vartheta_k) \pi_l(d\vartheta_l) \tau_1(d\sigma_1^2) \tau_2(d\sigma_2^2), \tag{6.3}$$

where the quantities $\alpha_{k,l}$ denote non-negative weights reflecting the experimenters belief about the pair (m_k, m_l) for group 1 and 2 with $\sum_{k,l=1}^p \alpha_{k,l} = 1$ (here and throughout this section we assume the existence of all integrals). Exemplarily, we mention a generalization of Theorem 3.2. The proof is omitted for the sake of brevity.

THEOREM 6.3. *The design $\xi^\star = (\xi_1^\star, \xi_2^\star)$ is robust optimal for the comparison of the two curves if and only if for all $k, l = 1, \ldots, p$, for all $\vartheta_k \in supp(\pi_k)$, $\vartheta_l \in supp(\pi_l)$ and for all $\sigma_i^2 \in supp(\tau_i)$ $(i = 1, 2)$ there exist* *measures* $\varrho^{\star,k,l}_{\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2}$ *on the sets of the extremal points*

$$\mathcal{Z}^{k,l}_{\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2}(\xi^\star) = \Big\{ t_0 \in \mathcal{Z} \;\Big|\; \frac{\sigma_1^2}{\gamma_1} f_k^T(t_0,\vartheta_k) M_k^{-1}(\xi_1^\star,\vartheta_k) f_k(t_0,\vartheta_k) + \frac{\sigma_2^2}{\gamma_2} f_l^T(t_0,\vartheta_l) M_l^{-1}(\xi_2^\star,\vartheta_l) f_l(t_0,\vartheta_l) = \mu_\infty^{k,l}(\xi_1^\star,\xi_2^\star,\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2) \Big\},$$

such that the inequality

$$\int_{\mathbb{R}^{>0}} \int_{\mathbb{R}^{>0}} \sum_{k,l=1}^{p} \alpha_{k,l} \int_{\Theta_k} \int_{\Theta_l} \Big\{ \frac{\mathrm{eff}_\infty^{k,l}(\xi^\star,\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2)}{\mu_\infty^{k,l}(\xi^\star,\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2)} L^{k,l}(\zeta_1,\zeta_2,\xi^\star,\varrho^{\star,k,l}_{\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2}) - \mathrm{eff}_\infty^{k,l}(\xi^\star,\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2) \Big\} \pi_k(d\vartheta_k)\pi_l(d\vartheta_l)\tau_1(d\sigma_1^2)\tau_2(d\sigma_2^2) \quad \le 0$$

with $L^{k,l} := L^{k,l}(\zeta_1,\zeta_2,\xi^\star,\varrho^{\star,k,l}_{\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2})$ *defined by*

$$L^{k,l} = \Big(\frac{\sigma_1^2}{\gamma_1} \mathrm{tr}\big(M_k(\varrho^{\star,k,l}_{\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2},\vartheta_k) M_k^{-1}(\xi_1^\star,\vartheta_k) M_k(\zeta_1,\vartheta_k) M_k^{-1}(\xi_1^\star,\vartheta_k)\big) + \frac{\sigma_2^2}{\gamma_2} \mathrm{tr}\big(M_l(\varrho^{\star,k,l}_{\vartheta_k,\vartheta_l,\sigma_1^2,\sigma_2^2},\vartheta_l) M_l^{-1}(\xi_2^\star,\vartheta_l) M_l(\zeta_2,\vartheta_l) M_l^{-1}(\xi_2^\star,\vartheta_l)\big) \Big)$$

holds for all approximate pairs of designs $\zeta = (\zeta_1,\zeta_2)$ *on* $\mathcal{X} \times \mathcal{X}$.

Efficient robust designs can be calculated by a generalization of the algorithm developed in Section 5 and we illustrate the application of the algorithm in the examples considered in Section 5. As indicated in this section, the μ_∞-optimal designs are robust with respect to moderate misspecification of the model parameters [see the discussion at the end of Section 5], however, they are less robust with respect to misspecification of the parametric regression models [see Table 3]. For this reason we use one-point priors π_k, π_l, τ_1, τ_2, in equation (6.3) (supported at the parameters specified in Table 1 for the three models). For the weights $\alpha_{k,l}$ in (6.3) we choose $\alpha_{1,1} = \alpha_{2,2} = \alpha_{3,3} = 1/3$ and $\alpha_{k,l} = 0$, for $k \neq l$. This means that we construct a robust design for the comparison of two Emax, two exponential and two log-linear models.

The robust optimal design is given by $\xi_{\mathrm{robust}} = (\xi_{1,\mathrm{robust}}, \xi_{2,\mathrm{robust}})$, where $\xi_{1,\mathrm{robust}}$ has masses 30.27%, 30.72%, 19.56%, 19.44% at the points 0.00 , 0.17, 0.77, 1.00 and $\xi_{2,\mathrm{robust}}$ has masses 29.96%, 30.18%, 19.38%, 20.48% at the points 0.00, 0.17, 0.76 1.00. Its efficiencies are given in the last row of Table 3. We observe a substantial improvement of the standard design. It might also be of interest to compare the two designs for equal models for the two groups. The efficiencies of the standard design are 67.98%, 42.87% and 74.29% for the Emax/Emax, exp/exp and loglin/loglin case, respectively, while the corresponding efficiencies of the robust optimal design are 90.40%, 59.96% and 90.79%. In all scenarios the robust optimal design provides a substantial improvement of the standard design.

Acknowledgements. The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. We are also grateful to Kathrin Möllenhoff for interesting discussions on the subject of comparing curves and for computational assistance. This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under the IDEAL Grant Agreement no 602552. The work has also been supported in part by the Deutsche Forschungsgemeinschaft (SFB 823, Teilprojekt C2) and the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. 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 would like to thank three referees and the associate editor for their constructive comments on an earlier version of this paper.

7. Proofs.

Let Ξ denote the space of all approximate designs on $\mathcal{X}$ and define for $\xi_1, \xi_2 \in \Xi$

$$M(\xi_1, \xi_2, \vartheta_1, \vartheta_2) = \begin{pmatrix} \frac{\gamma_1}{\sigma_1^2} M_1(\xi_1, \vartheta_1) & 0_{d_1 \times d_2} \\ 0_{d_2 \times d_1} & \frac{\gamma_2}{\sigma_2^2} M_2(\xi_2, \vartheta_2) \end{pmatrix} \tag{7.1}$$

as the block diagonal matrix with information matrices $\frac{\gamma_1}{\sigma_1^2} M_1(\xi_1, \vartheta_1)$ and $\frac{\gamma_2}{\sigma_2^2} M_2(\xi_2, \vartheta_2)$ in the diagonal. The set

$$\mathcal{M}^{(2)} = \{M(\xi_1, \xi_2, \vartheta_1, \vartheta_2) : \xi_1, \xi_2 \in \Xi\}$$

is obviously a convex subset of the set $\mathrm{NND}(d_1 + d_2)$ of all non-negative definite $(d_1 + d_2) \times (d_1 + d_2)$ matrices. Moreover, if δ_t denotes the Dirac measure at the point $t \in \mathcal{X}$ it is easy to see that $\mathcal{M}^{(2)}$ is the convex hull of the set

$$\mathcal{D}^{(2)} = \{M(\delta_{t_1}, \delta_{t_2}, \vartheta_1, \vartheta_2) : t_1, t_2 \in \mathcal{X}\},$$

and that for any $p \in [1, \infty]$ the function $\mu_p(\xi) = \mu_p((\xi_1, \xi_2))$ defined in (2.5) and (2.4) is convex on the set $\Xi \times \Xi$.

Proof of Theorem 3.1 Note that the function φ in (2.2) can be written as $\varphi(t,\xi_1,\xi_2) = f^T(t)M^{-1}(\xi_1,\xi_2,\vartheta_1,\vartheta_2)f(t)$, where $f^T(t) = (f_1^T(t), f_2^T(t))$ and $M(\xi_1,\xi_2) \in \mathcal{M}^{(2)}$ is defined in (7.1). Similarly, we introduce the notation $\Phi(M,t) = f^T(t)M^{-1}f(t)$ for a matrix $M \in \mathcal{M}^{(2)}$ and we rewrite the function $\mu_p(\xi_1,\xi_2)$ as

(7.2)
$$\tilde{\mu}_p(M) = \Big(\int_{\mathcal{Z}} (\Phi(M,t))^p \, d\lambda(t)\Big)^{1/p} = \Big(\int_{\mathcal{Z}} \big(f^T(t)M^{-1}f(t)\big)^p \, d\lambda(t)\Big)^{1/p}.$$

Because of the convexity of μ_p the design $\xi^{\star,p} = (\xi_1^{\star,p}, \xi_2^{\star,p})$ is μ_p-optimal if and only if the derivative of $\tilde{\mu}_p(M)$ evaluated in $M_0 = M(\xi_1^{\star,p}, \xi_2^{\star,p}, \vartheta_1, \vartheta_2)$ is non-negative for all directions $E_0 = E - M_0$, where $E \in \mathcal{M}^{(2)}$, i.e. $\partial\tilde{\mu}_p(M_0,E_0) \geq 0$. Since $\mathcal{M}^{(2)} = \text{conv}(\mathcal{D}^{(2)})$ it is sufficient to verify this inequality for all $E \in \mathcal{D}^{(2)}$.
Assuming that integration and differentiation are interchangeable, it follows by standard calculations that the derivative at $M_0 = M(\xi_1,\xi_2,\vartheta_1,\vartheta_2)$ in direction $E_0 = M(\delta_{t_1}\delta_{t_2},\vartheta_1,\vartheta_2) - M_0$ is given by

$$\partial\tilde{\mu}_p(M_0,E_0) = \mu_p(\xi_1,\xi_2)\Big[1 - \mu_p(\xi_1,\xi_2)^{-p}\int_{\mathcal{Z}} \beta(t,t_1,t_2) d\lambda(t)\Big] \tag{7.3}$$

where the function β is given by

$$\beta(t,t_1,t_2) = \varphi(t,\xi_1,\xi_2)^{p-1}\big(\tfrac{\gamma_1}{\sigma_1^2}(\varphi_1(t,t_1,\xi_1))^2 + \tfrac{\gamma_2}{\sigma_2^2}(\varphi_2(t,t_2,\xi_2))^2\big). \tag{7.4}$$

Consequently, the design $\xi^{\star,p} = (\xi_1^{\star,p}, \xi_2^{\star,p})$ is μ_p-optimal if and only if the inequality

$$\int_{\mathcal{Z}} \beta(t,t_1,t_2) d\lambda(t) - \big(\mu_p(\xi_1^{\star,p}, \xi_2^{\star,p})\big)^p \leq 0 \tag{7.5}$$

is satisfied for all $t_1, t_2 \in \mathcal{X}$, which proves the first part of the assertion.
It remains to prove that equality holds for any point $(t_1,t_2) \in \text{supp}(\xi_1^{\star,p}) \times \text{supp}(\xi_2^{\star,p})$. For this purpose we assume the opposite, i.e. there exists a point $(t_1,t_2) \in \text{supp}(\xi_1^{\star,p}) \times \text{supp}(\xi_2^{\star,p})$, such that there is strict inequality in (7.5). This gives

$$\int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Z}} \beta(t,t_1,t_2) d\lambda(t) d\xi_1^{\star,p}(t_1) d\xi_2^{\star,p}(t_2) < \big(\mu_p(\xi_1^{\star,p}, \xi_2^{\star,p})\big)^p .$$

On the other hand, we have

$$\int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{Z}} \beta(t,t_1,t_2) d\lambda(t) d\xi_1^{\star,p}(t_1) d\xi_2^{\star,p}(t_2)$$
$$= \int_{\mathcal{Z}} \varphi(t,\xi_1^{\star,p}, \xi_2^{\star,p})^p d\lambda(t) = \big(\mu_p(\xi_1^{\star,p}, \xi_2^{\star,p})\big)^p.$$

This contradiction shows that equality in (7.5) must hold whenever $(t_1, t_2) \in \text{supp}(\xi_1^{\star,p}) \times \text{supp}(\xi_2^{\star,p})$.

Proof of Theorem 3.2 By the discussion at the beginning of the proof of Theroem 3.1 the minimization of the function $\mu_\infty(\xi_1, \xi_2)$ is equivalent to the minimization of

$$\tilde{\mu}_\infty(M) = \sup_{t\in\mathcal{Z}} \Phi(M,t) = \sup_{t\in\mathcal{Z}} f^T(t) M^{-1} f(t) \tag{7.6}$$

for $M \in \mathcal{M}^{(2)}$. From Theorem 3.5 in Pshenichnyi (1971) the subgradient of $\tilde{\mu}_\infty(M)$ evaluated at a matrix M_0 in direction E is given by

$$\mathrm{D}\tilde{\mu}_\infty(M_0, E) = \Big\{ \int_{\mathcal{Z}(M_0)} \partial\Phi(M_0, E, t) d\varrho(t) : \varrho \text{ measure on } \mathcal{Z}(M_0) \Big\},$$

where the set $\mathcal{Z}(M_0)$ is defined by $\mathcal{Z}(M_0) = \{t \in \mathcal{Z} : \tilde{\mu}_\infty(M_0) = \Phi(M_0, t)\}$, and the derivative of $\Phi(M_0, t)$ in direction E is given by $\partial\Phi(M_0, E, t) = -f^T(t) M_0^{-1} E M_0^{-1} f(t)$. Applying the results from page 59 in Pshenichnyi (1971) it therefore follows that the design $\xi^{\star,\infty} = (\xi_1^{\star,\infty}, \xi_2^{\star,\infty})$ is μ_∞-optimal if and only if there exists a measure $\varrho^\star$ on $\mathcal{Z}(M(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \vartheta_1, \vartheta_2))$ such that the inequality

$$\begin{aligned}\int_{\mathcal{Z}(M_0)} & \partial\Phi(M_0, E_0, t) d\varrho^\star(t) \\ &= \int_{\mathcal{Z}(M_0)} \partial\Phi(M_0, E, t) d\varrho^\star(t) + \int_{\mathcal{Z}(M_0)} f^T(t) M_0^{-1} f(t) d\varrho^\star(t) \geq 0\end{aligned}$$

holds for all $E_0 = E - M_0$, $E \in \mathcal{M}^{(2)}$. Since $\mathcal{M}^{(2)} = \text{conv}(\mathcal{D}^{(2)})$ it is sufficient to consider the directions $E_0 = E - M_0$, where $E \in \mathcal{D}^{(2)}$. Thus, this inequality is fulfilled if and only if there exists a measure $\varrho^\star$ on $\mathcal{Z}(M_0) = \mathcal{Z}(\xi^{\star,\infty}) = \mathcal{Z}^\star$, such that the inequality

$$\begin{aligned}&\int_{\mathcal{Z}^\star} f^T(t) M^{-1}(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \vartheta_1, \vartheta_2) M(\delta_{t_1}, \delta_{t_2}, \vartheta_1, \vartheta_2) M^{-1}(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \vartheta_1, \vartheta_2) f(t) d\varrho^\star(t) \\ &\leq \int_{\mathcal{Z}^\star} f^T(t) M^{-1}(\xi_1^{\star,\infty}, \xi_2^{\star,\infty}, \vartheta_1, \vartheta_2) f(t)\, d\varrho^\star(t) = \mu_\infty(\xi_1^{\star,\infty}, \xi_2^{\star,\infty})\end{aligned} \tag{7.7}$$

is satisfied for all $M(\delta_{t_1}, \delta_{t_2}, \vartheta_1, \vartheta_2) \in \mathcal{D}^{(2)}$. Observing the definition of φ_i in (3.2), the left-hand part of (7.7) can be rewritten as $\int_{\mathcal{Z}(\xi^{\star,\infty})} \frac{\gamma_1}{\sigma_1^2} \varphi_1^2(t_1, t, \xi_1^{\star,\infty}) +$

$\frac{\gamma_2}{\sigma_2^2}\varphi_2^2(t_2, t, \xi_2^{\star,\infty})\, d\varrho^\star(t)$, and the inequality (7.7) reduces to (3.4). The remaining statement regarding the equality at the support points follows by the same arguments as in the proof of Theorem 3.1 and the details are omitted for the sake of brevity.

Proof of Theorem 3.3 For both cases consider the function $(\tilde{\mu}_p(M))^{-1}$ where $\tilde{\mu}_p$ has been defined in (7.2) and (7.6). Note that for each $t \in \mathcal{Z}$ the function $M \to (f(t)^T M^{-1} f(t))^{-1}$ is concave [see Pukelsheim (2006), p. 77], and consequently the function

$$(\tilde{\mu}_\infty(M))^{-1} = \frac{1}{\max_{t\in\mathcal{Z}} f(t)^T M^{-1} f(t)} = \min_{t\in\mathcal{Z}} (f(t)^T M^{-1} f(t))^{-1}$$

is also concave. The concavity of $(\tilde{\mu}_p(M))^{-1}$ in the case $1 \le p < \infty$ follows by similar arguments. For $p \in [1, \infty]$ the directional derivative of $(\tilde{\mu}_p(M))^{-1}$ at the point M_0 in direction $E_0 = M - M_0$ is given by

$$\partial(\tilde{\mu}_p(M_0, E_0))^{-1} = -(\tilde{\mu}_p(M_0))^{-2}\partial\tilde{\mu}_p(M_0, E_0).$$

We now consider the case $p \in [1, \infty)$, the remaining case $p = \infty$ is briefly indicated at the end of this proof. Observing (7.3) a lower bound of the directional derivative of $\tilde{\mu}_p$ at $M_0 = M(\xi_1, \xi_2, \vartheta_1, \vartheta_2)$ in direction $E_0 = M(\delta_{t_1}\delta_{t_2}, \vartheta_1, \vartheta_2) - M_0$ is given by

$$\partial\tilde{\mu}_p(M_0, E_0) \quad \ge \quad \tilde{\mu}_p(M_0)\Big[1 - \frac{\max_{t_1,t_2}\int_{\mathcal{Z}} \beta(t, t_1, t_2) d\lambda(t)}{\tilde{\mu}_p^p(M_0)}\Big]$$

where $\beta(t, t_1, t_2)$ is defined in (7.4). Consequently, we have

$$\partial(\tilde{\mu}_p(M_0, E_0))^{-1} \le \frac{1}{\tilde{\mu}_p(M_0)}\Big[\frac{\max_{t_1,t_2}\int_{\mathcal{Z}} \beta(t, t_1, t_2) d\lambda(t)}{\tilde{\mu}_p^p(M_0)} - 1\Big]. \tag{7.8}$$

Now, we consider the matrices $M = M(\xi_1^{\star,p}, \xi_2^{\star,p}, \vartheta_1, \vartheta_2)$ of the μ_p-optimal design and $M_0 = M(\xi_1, \xi_2, \vartheta_1, \vartheta_2)$ of any fixed design $\xi = (\xi_1, \xi_2)$ with nonsingular information matrices $M_1(\xi_1, \vartheta_1)$ and $M_2(\xi_2, \vartheta_2)$ and define the function $g_p(\alpha) = \tilde{\mu}_p((1-\alpha)M_0 + \alpha M))^{-1}$, which is concave because of the concavity of $(\tilde{\mu}_p(M))^{-1}$. This yields

$$\frac{1}{\tilde{\mu}_p(M)} - \frac{1}{\tilde{\mu}_p(M_0)} \quad = \quad g_p(1) - g_p(0) \le \frac{\partial g_p(\alpha)}{\partial \alpha}\Big|_{\alpha=0} = \partial(\tilde{\mu}_p(M_0, E_0))^{-1}$$

Consequently, we obtain from (7.8) the inequality

$$\text{eff}_p(\xi) = \frac{\tilde{\mu}_p(M)}{\tilde{\mu}_p(M_0)} \ge \frac{\tilde{\mu}_p^p(M_0)}{\max_{t_1,t_2}\int_{\mathcal{Z}} \beta(t, t_1, t_2) d\lambda(t)},$$

which proves the assertion of Theorem 3.3 in the case $1 \leq p < \infty$. In the case $p = \infty$ we use similar arguments which provides the upper bound

$$\partial(\tilde{\mu}_\infty(M_0, E_0))^{-1} \leq \frac{1}{\tilde{\mu}_\infty(M_0)}\Big\{\frac{\min_{\varrho\in\Xi(\mathcal{Z}_0}\max_{t_1,t_2\in\mathcal{X}}\int_{\mathcal{Z}_0}(f^T(t)M_0^{-1}f(t_1,t_2))^2 d\varrho(t)}{\tilde{\mu}_\infty(M_0)} - 1\Big\}, \tag{7.9}$$

where $f(t_1, t_2)$ is defined by $f^T(t_1, t_2) = (f_1^T(t_1), f_2^T(t_2))^T$ $\mathcal{Z}_0 = \mathcal{Z}(M_0)$ and $\Xi(\mathcal{Z}_0)$ is set of all measures ϱ supported on $\mathcal{Z}_0$. The details are omitted for the sake of brevity.

Proof of Theorem 4.1 For the sake of brevity we now restrict ourselves to the proof of the first part of Theorem 4.1. The second part can be proved analogously. Let $U_{\mathcal{X}} < L_{\mathcal{Z}}$ and recall the definition of the function $\varphi(t, \xi_1, \xi_2)$ defined in (2.2). The function $\varphi(t, \xi_1, \xi_2)$ is obviously increasing on $\mathcal{Z}$, if the functions

$$\varphi_i(t, t, \xi_i) = \frac{\sigma_i^2}{\gamma_i} f_i^T(t)M_i^{-1}(\xi_i)f_i(t) = \frac{\sigma_i^2}{\gamma_i}\omega_i^2(t)(1, t, \dots, t^{p_i})M_i^{-1}(\xi_i)(1, t, \dots, t^{p_i})^T$$

are increasing on $\mathcal{Z}$ for $i = 1, 2$. In this case we have

$$\max_{t\in\mathcal{Z}} \varphi(t, \xi_1, \xi_2) = \varphi(U_{\mathcal{Z}}, \xi_1, \xi_2) = \varphi_1(U_{\mathcal{Z}}, \xi_1) + \varphi_2(U_{\mathcal{Z}}, \xi_2). \tag{7.10}$$

Because of this structure the components of the optimal design can be calculated separately for φ_1 and φ_2. Since both $\{\omega_1(t), \omega_1(t)t, \dots, \omega_1(t)t^{p_1}\}$ and $\{\omega_2(t), \omega_2(t)t, \dots, \omega_2(t)t^{p_2}\}$ are Chebyshev systems on $\mathcal{X} \cup \mathcal{Z}$ it follows from Theorem X.7.7 in Karlin and Studden (1966) that the support points of the design ξ_i minimizing $\varphi_i(U_{\mathcal{Z}}, \xi_i)$ are given by the extremal points of the equioscillating polynomials $\underline{v}_i(t)$, while the corresponding weights are given by (4.2).

In order to prove the monotonicity of φ_i, $(i = 1, 2)$ let ξ_i denote a design with k_i support points $t_{i0}, \dots, t_{ik_i-1} \in \mathcal{X}$ and corresponding weights $\xi_{i0}, \dots, \xi_{ik_i-1}$.

Since $\{1, \omega_i(t), \omega_i(t)t, \dots, \omega_i(t)t^{2p_i-1}\}$ and $\{1, \omega_i(t), \omega_i(t)t, \dots, \omega_i(t)t^{2p_i}\}$ are Chebshev systems for $\omega_i(t) \neq c \in \mathbb{R}$, the complete class theorem of Dette and Melas (2011) can be applied and it is sufficient to consider minimal supported designs ξ_i. Consequently, we set $k_i = p_i + 1$.

Define $X_i = \big(\omega_i(t_{ik})t_{ik}^l\big)_{k,l=0,\dots,p_i}$, then it is easy to see that the jth Langrange interpolation polynomial is given by $L_{ij}(t) = e_j^T X_i^{-1} \left(\omega_i(t), \omega_i(t)t, \dots, \omega_i(t)t^{p_i}\right)^T$, where e_j denotes the jth unit vector (just check the defining condition $L_{ij}(t_{i\ell}) = \delta_{j\ell}$). With these notations the function $\varphi_i(t,\xi_i)$ can be rewritten as

$$
\begin{aligned}
\varphi_i(t,t,\xi_i) &= \frac{\sigma_i^2}{\gamma_i}\omega_i^2(t)\big(1,t,\dots,t^{p_i}\big)X_i^{-T}W_i^{-1}X_i^{-1}\big(1,t,\dots,t^{p_i}\big)^T \\
&:= \frac{\sigma_i^2}{\gamma_i}\sum_{j=0}^{p_i}\frac{1}{\xi_{ij}}(L_{ij}(t))^2,
\end{aligned}
\tag{7.11}
$$

where $W_i = \mathrm{diag}(\xi_{i0},\dots,\xi_{ip_i})$. Now Cramer's rule and a straightforward calculation yields the following representation for the Lagrange interpolation polynomial

$$
\begin{aligned}
L_{ij}(t) &= (-1)^{p_i-j}\omega_i(t)\frac{\prod_{k=0,k\neq j}^{p_i}\omega_i(t_{ik})}{\prod_{k=0}^{p_i}\omega_i(t_{ik})}\frac{\det\begin{pmatrix}1 & \dots & 1 & 1 & \dots & 1 & 1\\ t_{i0} & \dots & t_{ij-1} & t_{ij+1} & \dots & t_{ip_i} & t\\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots\\ t_{i0}^{p_i} & \dots & t_{ij-1}^{p_i} & t_{ij+1}^{p_i} & \dots & t_{ip_i}^{p_i} & t^{p_i}\end{pmatrix}}{\det\begin{pmatrix}1 & \dots & 1\\ t_{i0} & \dots & t_{ip_i}\\ \vdots & \dots & \vdots\\ t_{i0}^{p_i} & \dots & t_{ip_i}^{p_i}\end{pmatrix}} \\
&= \frac{\omega_i(t)}{\omega_i(t_j)}\prod_{k=0,k\neq j}^{p_i}\frac{t-t_{ik}}{t_{ij}-t_{ik}}.
\end{aligned}
$$

Therefore the partial derivative of $\varphi_i(t,\xi_i)$ with respect to t is given by

$$
\frac{\partial}{\partial t}\varphi_i(t,t,\xi_i) = \frac{\sigma_i^2}{\gamma_i}\sum_{j=0}^{p_i}\frac{2}{\xi_{ij}}(L_{ij}(t))^2\Big(\frac{\omega_i'(t)}{\omega_i(t)}+\sum_{l=0}^{p_i}\frac{1}{t-t_{il}}\Big).
$$

Note that $t_{il} < t$ for all $t_{il}\in\mathcal{X}$ and $t\in\mathcal{Z}$ and that both $\omega_i(t)$ and $\omega_i'(t)$ are positive. Consequently, the partial derivative is positive and the function $\varphi_i(t,\xi_i)$ is increasing in $t\in\mathcal{Z}$. Thus, the maximum value of $\varphi_i(t,\xi_i)$ is attained in $U_{\mathcal{Z}}\in\mathcal{Z}$ and (7.10) follows.

Proof of Corollary 4.4 For the sake of brevity we only prove the result for the Emax model (4.6), where it essentially follows by an application of Theorem 4.1 with $\omega(t)\equiv 1$. The proofs for the Michaelis Menten model and for the loglinear model are similar. In the Emax model the gradient is given by $f(t,\vartheta) = (1, \frac{t}{t+\vartheta_3}, \frac{-\vartheta_2 t}{(t+\vartheta_3)^2})$. Using the strictly increasing transformation $z = v(t) = \frac{t}{\vartheta_3+t}$ the function f can be rewritten by

$$
f(t,\vartheta) = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -\frac{\vartheta_2}{\vartheta_3} & \frac{\vartheta_2}{\vartheta_3}\end{pmatrix}\begin{pmatrix}1\\ z\\ z^2\end{pmatrix} := P_\vartheta\begin{pmatrix}1\\ z\\ z^2\end{pmatrix}.
$$

Thus, for an arbitrary design ξ the function $f^T(t)M^{-1}(\xi)f(t)$ reduces to

$$\begin{aligned}\varphi(t,\xi) = f^T(t)M^{-1}(\xi)f(t) &= (1,z,z^2)P_\vartheta^T\left(P_\vartheta\tilde{M}(\tilde{\xi})P_\vartheta^T\right)^{-1}P_\vartheta\,(1,z,z^2)^T \\ &= (1,z,z^2)\tilde{M}^{-1}(\tilde{\xi})(1,z,z^2)^T = \tilde{\varphi}(z,\tilde{\xi})\end{aligned}$$

where $\tilde{M}(\tilde{\xi}) = (\int_{\mathcal{X}} z^{i+j}d\tilde{\xi}(z))_{i,j=0,1,2}$ and $\tilde{\xi}$ is the design on the design space $\tilde{\mathcal{X}} = [\frac{L_\mathcal{X}}{\vartheta_3+L_\mathcal{X}}, \frac{U_\mathcal{X}}{\vartheta_3+U_\mathcal{X}}]$ induced from the actual design ξ by the transformation $z = \frac{t}{\vartheta_3+t}$. The function $\tilde{\varphi}(z,\tilde{\xi})$ coincides with the variance function of a polynomial regression model with degree 2 and constant weight function $\omega(t) \equiv 1$. The corresponding design and extrapolation space are given by $\tilde{\mathcal{X}} = [\frac{L_\mathcal{X}}{\vartheta_3+L_\mathcal{X}}, \frac{U_\mathcal{X}}{\vartheta_3+U_\mathcal{X}}]$ and $\tilde{\mathcal{Z}} = [\frac{L_\mathcal{Z}}{\vartheta_3+L_\mathcal{Z}}, \frac{U_\mathcal{Z}}{\vartheta_3+U_\mathcal{Z}}]$, respectively. According to Example 4.3 ($p_1 = 2$) the component $\tilde{\xi}_i$ of the μ_∞-optimal design is supported at the extremal points of the Chebyshev polynomial of the first kind on the interval $\mathcal{X}$, which are given by

$$\frac{L_\mathcal{X}}{\vartheta_3+L_\mathcal{X}},\ \frac{1}{2}\Big(\frac{L_\mathcal{X}}{\vartheta_3+L_\mathcal{X}} + \frac{U_\mathcal{X}}{\vartheta_3+U_\mathcal{X}}\Big),\ \frac{U_\mathcal{X}}{\vartheta_3+U_\mathcal{X}}$$

For the weights we obtain by the same result $\xi_0 = \frac{|L_0|}{L}, \xi_1 = \frac{|L_1|}{L}, \xi_2 = \frac{|L_2|}{L}$ where

$$\begin{aligned}|L_0| &= \left(2\tfrac{U_\mathcal{Z}}{U_\mathcal{Z}+\vartheta_3} - \left(\tfrac{U_\mathcal{X}}{U_\mathcal{X}+\vartheta_3} + \tfrac{L_\mathcal{X}}{L_\mathcal{X}+\vartheta_3}\right)\right)\left(\tfrac{U_\mathcal{Z}}{U_\mathcal{Z}+\vartheta_3} - \tfrac{U_\mathcal{X}}{U_\mathcal{X}+\vartheta_3}\right) \\ |L_1| &= 4\left(\tfrac{U_\mathcal{Z}}{U_\mathcal{Z}+\vartheta_3} - \tfrac{L_\mathcal{X}}{L_\mathcal{X}+\vartheta_3}\right)\left(\tfrac{U_\mathcal{Z}}{U_\mathcal{Z}+\vartheta_3} - \tfrac{U_\mathcal{X}}{U_\mathcal{X}+\vartheta_3}\right) \\ |L_2| &= \left(\tfrac{U_\mathcal{Z}}{U_\mathcal{Z}+\vartheta_3} - \tfrac{L_\mathcal{X}}{L_\mathcal{X}+\vartheta_3}\right)\left(2\tfrac{U_\mathcal{Z}}{U_\mathcal{Z}+\vartheta_3} - \left(\tfrac{U_\mathcal{X}}{U_\mathcal{X}+\vartheta_3} + \tfrac{L_\mathcal{X}}{L_\mathcal{X}+\vartheta_3}\right)\right) \\ L &= |L_0| + |L_1| + |L_2|.\end{aligned}$$

The support points of the μ_∞-optimal design ξ are now obtained by the inverse of the transformation and the assertion for the Emax model follows from the definition of the function g and a straightforward calculation.

References.

ANTILLE, G. and WEINBERG, A. (2003). A note on optimal designs in weighted polynomial regression for the classical efficiency functions. *Journal of Statistical Planning and Inference* **113** 285-292.

BONYADI, M. R. and MICHALEWICZ, Z. (2014). A locally convergent rotationally invariant particle swarm optimization algorithm. *Swarm Intelligence* **8** 159-198.

BRETZ, F., PINHEIRO, J. C. and BRANSON, M. (2005). Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies. *Biometrics* **61** 738-748.

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

Chang, F. C. (2005a). D-Optimal designs for weighted polynomial regression - A functional-algebraic approach. *Statistica Sinica* **15** 153-163.

Chang, F. C. (2005b). D-Optimal designs for weighted polynomial regression - A functional approach. *Annals of the Institute of Statistical Mathematics* **57** 833-844.

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

Clerc, M. (2006). *Particle Swarm Optimization*. Iste Publishing Company, London.

Dette, H. (1990). A generalization of D- and D_1-optimal designs in polynomial regression. *The Annals of Statistics* **18** 1784-1805.

Dette, H. (1993). A note on E-optimal designs for weighted polynomial regression. *The Annals of Statistics* **21** 767-771.

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

Dette, H., Bornkamp, B. and Bretz, F. (2013). On the efficiency of two-stage response-adaptive designs. *Statistics in Medicine* **32** 1646-1660.

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

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *The Annals of Statistics* **29** 1361-1400.

Dette, H., Pepelyshev, A. and Zhigljavsky, A. (2008). Improving updating rules in multiplicative algorithms for computing D-optimal designs. *Computational Statistics and Data Analysis* **53** 312-320.

Dette, H. and Trampisch, M. (2010). A general approach to D-optimal designs for weighted univariate polynomial regression models. *Journal of the Korean Statistical Society* **39** 1-26.

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

Dragalin, V., Bornkamp, B., Bretz, F., Miller, F., Padmanabhan, S. K., Patel, N., Perevozskaya, I., Pinheiro, J. and Smith, J. R. (2010). A simulation study to compare new adaptive dose-ranging designs. *Statistics in Biopharmaceutical Research* **2**.

Gsteiger, S., Bretz, F. and Liu, W. (2011). Simultaneous confidence bands for nonlinear regression models with application to population pharmacokinetic analyses. *Journal of Biopharmaceutical Statistics* **21** 708-725.

Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association* **85** 1039-1049.

Heiligers, B. (1994). E-optimal designs in weighted polynomial regression. *The Annals of Statistics* **22** 917-929.

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). *The Annals of Statistics* **2** 849-879.

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

Lau, T. S. and Studden, W. J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. *The Annals of Statistics* **13** 383-394.

Läuter, E. (1974). Experimental design in a class of models. *Math. Operationsforsch. Statist.* **5** 379-398.

Liu, W., Jamshidian, M. and Zhang, Y. (2011). Multiple comparison of several linear regression models. *Journal of the American Statistical Association* **99** 395-403.

Liu, W., Bretz, F., Hayter, A. J. and Wynn, H. P. (2009). Assessing non-superiority, non-inferiority of equivalence when comparing two regression models over a restricted covariate region. *Biometrics* **65** 1279-1287.

Pshenichnyi, B. N. (1971). *Necessary Conditions for an Extremum.* Marcel Dekker, Inc.

Pukelsheim, F. (2006). *Optimal Design of Experiments.* SIAM, Philadelphia.

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79** 763-770.

van den Bergh, F. and Engelbrecht, A. P. (2010). A convergence proof for the particle swarm optimiser. *Fundamenta Informaticae - FUIN* **105** 341-374.

Wong, W. K. and Cook, R. D. (1993). Heteroscedastic G-Optimal Designs. *Journal of the Royal Statistical Society. Series B (Methodological)* **55** 871-880.

Yang, M. (2010). On the de la Garza phenomenon. *Annals of Statistics* **38** 2499-2524.

Yang, M., Biedermann, S. and Tang, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *Journal of the American Statistical Association* **108** 1411-1420.

Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics* **38** 1593-1606.

Zen, M. M. and Tsai, M. H. (2004). Criterion-robust optimal designs for model discrimination and parameter estimation in Fourier regression models. *Journal of Statistical Planning and Inference* **124** 475-487.

Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
E-mail: holger.dette@rub.de
kirsten.schorning@rub.de