

Optimal Discrimination Designs for Exponential Regression Models

Stefanie Biedermann

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

email: stefanie.biedermann@ruhr-uni-bochum.de

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

email: holger.dette@ruhr-uni-bochum.de

FAX: +49 2 34 32 14 559

Andrey Pepelyshev

St. Petersburg State University

Department of Mathematics

St. Petersburg, Russia

email: andrey@ap7236.spb.edu

July 6, 2005

Abstract

We investigate optimal designs for discriminating between exponential regression models of different complexity, which are widely used in the biological sciences; see, e.g., Landaw (1995) or Gibaldi and Perrier (1982). We discuss different approaches for the construction of appropriate optimality criteria, and find sharper upper bounds on the number of support points of locally optimal discrimination designs than those given by Caratheodory's Theorem. These results greatly facilitate the numerical construction of optimal designs. Various examples of optimal designs are then presented and compared to different other designs. Moreover, to protect the experiment against misspecifications of the nonlinear model parameters, we adapt the design criteria such that the resulting designs are robust with respect to such misspecifications and, again, provide several examples, which demonstrate the advantages of our approach.

AMS Classification: 62K05, 62J02

Keywords and Phrases: Compartmental Model, Model Discrimination, Discrimination Design, Locally Optimal Design, Robust Optimal Design, Maximin Optimal Design.

1 Introduction

In the biological and chemical sciences, the expected response of an experiment at some experimental condition x is commonly modeled as a function in x depending nonlinearly on some unknown parameters. An important class within the nonlinear regression models are exponential models of the form

$$\eta_k(x, \theta) = \sum_{i=1}^k a_i e^{-\lambda_i x}, \quad k \geq 1, \quad x \in [0, \infty), \quad (1)$$

which have applications in chemical kinetics [see Gibaldi and Perrier (1982)] (in particular toxicokinetic experiments [see Becka, Bolt and Urfer (1992, 1993)]) or microbiology [see Alvarez et al. (2003), who used a model of type (1) to describe *Escherichia coli* inactivation by pulsed electric fields]. Landaw (1985) fitted an exponential model to describe open, noncyclic k -compartmental systems. It is well known that statistical inference can be improved considerably by choosing appropriate experimental conditions, at which the observations are taken and many authors have contributed articles on designing experiments optimally for nonlinear regression models [see, e.g., Melas (1978), Ford, Torsney and Wu (1992), Chaloner (1993), Dette and Biedermann (2003), Biedermann and Dette (2005) among many others]. As optimal designs with respect to classical optimality criteria such as D -, c - or E -optimality depend on the nonlinear model parameters they are termed "locally" optimal. Most authors provide such locally optimal designs in the sense of Chernoff (1953) assuming that a good initial guess for the true parameters will be available. Some more recent articles, however, pursue Bayesian or maximin approaches, which are more robust against misspecifications of the unknown parameters. Locally optimal designs for model (1) have already been discussed in a few articles [see Ermakov and Melas (1995), Dette, Melas and Wong (2005) and Dette, Melas and Pepelyshev (2004b) where the properties of locally D - and E -optimal designs are investigated].

In all the articles cited above the authors assume that the model under consideration is already known except for the values of the unknown parameters. So they deal with optimality criteria that are efficient for parameter estimation within a fixed model. In many applications, however, the form of the regression function will not be known exactly, i.e. the experimenter will have to decide among a set of competing classes of functions, which of them describes the data most adequately. The design problem for discrimination between competing nonlinear models, however, has found much less attention in the literature than problems for parameter estimation [see, e.g., Atkinson and Fedorov (1975) or Dette, Melas and Wong (2004)]. In the case of linear models optimal designs for model discrimination have been discussed by several authors. Stigler (1971) and Studden (1982) considered designs which are on the one hand efficient for estimating the parameters in a given polynomial regression model and on the other hand allow to test this model against a polynomial of higher degree. Läuter (1974) proposed designs which are optimal for a given class of models, and this approach was successfully applied by Dette (1990, 1994a), Zen and Tsai (2002) and Tsai and Zen (2004) to the class of polynomial models. Furthermore, the class of trigonometric regression models has been considered by Zen and Tsai (2004).

The goal of this article is to find designs that are efficient for discriminating between models of the form (1) where the value of k , i.e. the number of exponential terms contributing to the

model, is not fixed in advance. If for example a small model (small k) proves to be appropriate the number of parameters to be estimated will also be small so that estimation will be more efficient. Thus either experimental costs can be reduced without losing precision in estimation by carrying out less runs, or the precision of the estimation will increase (for a fixed number of runs).

The article at hand is organized as follows. Section 2 proposes several different optimality criteria that are suitable for finding discrimination designs for models of the form (1) with different degrees. In Section 3, we present some analytical results, which facilitate the numerical construction of locally optimal designs substantially. In particular, we derive explicit bounds on the number of support points of the locally optimal designs with respect to the optimality criteria introduced in Section 2. Various examples of optimal designs are displayed in Section 4, where we also investigate their performance compared to different uniform designs and designs, which are optimal for parameter estimation in the largest model. Furthermore, emphasis in this section will be on robust designs with respect to misspecifications of the unknown parameters and their performance. For this purpose we determine maximin optimal designs, and compare their performance with the locally optimal designs. The proofs of our results, finally, are deferred to an appendix.

2 Optimality Criteria

We consider an experiment with N independent observations Y_i , $i = 1, \dots, N$, at experimental conditions x_i , which are modeled by

$$Y_i = \eta_k(x_i, \theta) + \varepsilon_i, \quad i = 1, \dots, N, \quad (2)$$

where the errors ε_i are assumed i.i.d. with zero expectation and finite common variance, and the regression function $\eta_k(x, \theta)$ is of the form (1) for some parameter $k \geq 1$. By θ we will denote the vector of unknown parameters in (1), i.e. $\theta = (a_1, \lambda_1, \dots, a_k, \lambda_k)^T$. Without loss of generality we assume that $a_j \neq 0$, $j = 1, \dots, k$ and $0 < \lambda_1 < \dots < \lambda_k$.

An approximate design

$$\xi = \begin{pmatrix} x_1 & \dots & x_n \\ w_1 & \dots & w_n \end{pmatrix}$$

is a probability measure with finite support, i.e. the observations are taken at the support points x_i of the measure proportional to the corresponding masses w_i . In practice, a rounding procedure is applied to obtain the sample sizes $N_i \approx w_i N$ at the experimental conditions x_i , $i = 1, \dots, n$, subject to $N_1 + \dots + N_n = N$. An optimal rounding procedure is described in Pukelsheim and Rieder (1992). In the following, we will therefore restrict ourselves to the analysis and calculation of approximate designs to avoid the problems of discrete optimization. For the model described in (2), under some assumptions of regularity [see Jennrich (1969)] the least squares estimator is asymptotically unbiased with asymptotic covariance matrix proportional to the inverse of

$$M_k(\xi, \theta) = \sum_{i=1}^n w_i f_k(x_i, \theta) f_k^T(x_i, \theta) = \int f_k(x, \theta) f_k^T(x, \theta) d\xi(x) \in \mathbb{R}^{2k \times 2k}, \quad (3)$$

where the vector $f_k(x, \theta)$ is defined by

$$f_k(x, \theta) = \frac{\partial \eta_k(x, \theta)}{\partial \theta} = (e^{-\lambda_1 x}, -a_1 x e^{-\lambda_1 x}, e^{-\lambda_2 x}, -a_2 x e^{-\lambda_2 x}, \dots, e^{-\lambda_k x}, -a_k x e^{-\lambda_k x})^T.$$

In design literature, the matrix $M_k(\xi, \theta)$ is referred to as the information matrix of the design ξ , and an optimality criterion maximizes some (statistically meaningful) real-valued functional of the information matrix over the class of approximate designs.

The goal of this article is to find optimal designs for discriminating between models of type (2) with different numbers of exponential terms included in the regression function. Our main interest in this problem stems from the fact that often the experimenter will not know in advance, which value for k must be chosen to describe the data adequately. In the following, we will introduce several optimality criteria that are appropriate for problems of this type.

We first consider the discrimination problem between two models $\eta_s(x, \theta)$ and $\eta_k(x, \theta)$, where without loss of generality $s < k$. The benefits of this approach are obvious. If the smaller model already describes the dependencies between response and explanatory variable sufficiently well, the number of parameters to be estimated decreases, yielding more precision in the subsequent estimation. If, however, the smaller model is not adequate for the data, fitting only the regression function $\eta_s(x, \theta)$ will lead to severe problems in the data analysis. Following Stigler (1971) and Studden (1982), who considered polynomial regression models, a reasonable approach to find optimal discrimination designs for this problem, is to use designs that are in some sense optimal for estimating the $2(k-s)$ parameters from the $k-s$ highest terms in (1), to check if fitting these terms is actually necessary. An appropriate choice of optimality criterion is therefore the $D_{2(k-s)}$ -criterion; see, e.g., Silvey (1980), i.e. we maximize the determinant of the matrix

$$C_{A_{s,k}}(\xi, \theta) = (A_{s,k}^T M_k^-(\xi, \theta) A_{s,k})^{-1}, \quad A_{s,k}^T = (0, I_{2(k-s)}) \in \mathbb{R}^{2(k-s) \times 2k},$$

where $M_k^-(\xi, \theta)$ denotes a generalized inverse of $M_k(\xi, \theta)$, since this matrix is not necessarily of full rank. From Pukelsheim (1993), Sec. 3.11, it follows that for $M_k(\xi, \theta)$ with $\text{range}(A_{s,k}) \subset \text{range}(M_k(\xi, \theta))$ the information matrix $C_{A_{s,k}}(\xi, \theta)$ can be expressed as

$$C_{A_{s,k}}(\xi, \theta) = M_{22} - M_{21} M_{11}^- M_{12}, \quad (4)$$

where M_{11} , M_{12} , M_{21} and M_{22} are defined by partitioning the information matrix $M_k(\xi, \theta)$ into the block matrix

$$M_k(\xi, \theta) = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \quad M_{11} \in \mathbb{R}^{2s \times 2s}, \quad M_{22} \in \mathbb{R}^{2(k-s) \times 2(k-s)}.$$

The design maximizing $\det C_{A_{s,k}}(\xi, \theta)$ will therefore be referred to by locally optimal discrimination design η_k vs η_s or, shortly, $D_{2(k-s)}$ -optimal design. Note that if $M_k(\xi, \theta)$ is nonsingular, the determinant of $C_{A_{s,k}}(\xi, \theta)$ can be expressed by

$$\det C_{A_{s,k}}(\xi, \theta) = \frac{\det M_k(\xi, \theta)}{\det M_s(\xi, \theta)}.$$

A more general criterion will be necessary if discrimination between more than two competing models is required. Following Läuter (1974) we call a design ξ optimal discriminating design for a class of models $\{\eta_1, \dots, \eta_k\}$ if ξ maximizes the weighted geometric mean

$$T(\xi, \theta) = \sum_{l=1}^k \beta_l \ln \det C_{A_{l-1,l}}(\xi, \theta) \rightarrow \max_{\xi}$$

where the constants β_l are user-selected weights, chosen as a measure for the interest the experimenter has in discriminating η_l vs η_{l-1} . This criterion simplifies to the $D_{2(k-s)}$ -criterion when the weights β_l , $l = 1, \dots, k$, are given by $\beta_k = \beta_{k-1} = \dots = \beta_{s+1} = 1$, $\beta_s = \beta_{s-1} = \dots = \beta_1 = 0$. For the choice $\beta_k = \beta_{k-1} = \dots = \beta_1 = 1$, moreover, we obtain as a special case the D -criterion for model η_k . Another important special case of the criterion $T(\xi, \theta)$ arises for weights β_l that are inverse proportional to the maximal value of $\det C_{A_{l-1,l}}(\xi, \theta)$, $l = 1, \dots, k$. $T(\xi, \theta)$ can then be interpreted as a weighted sum of $D_{2(l-(l-1))}$ -efficiencies, i.e. a weighted sum of standardized optimality criteria, the use of which is recommended by Dette (1997). Optimal discrimination designs for the class of polynomial regression models have been determined by Dette (1990, 1994a) and Tsai and Zen (2002, 2004), whereas Zen and Tsai (2004) have found results for the class of trigonometric regression models.

Optimal designs with respect to the criterion $T(\xi, \theta)$ for at least two positive weights $\beta_l > 0$ are multiple-objective or compound optimal designs. Another somewhat related approach would be to consider constrained optimal designs, i.e. designs that maximize $\det C_{A_{l-1,l}}(\xi, \theta)$ for one certain value of $l \in \{1, \dots, k\}$ subject to the constraints that the $D_{2(j-(j-1))}$ -efficiencies attain at least some values $\mu_j \geq 0$, $j = 1, \dots, k$, $j \neq l$. It was shown by Dette (1994b) that there is a 1-1 correspondence between the constrained and the compound optimal designs. In the following, we will therefore restrict ourselves to results on compound or T -optimal designs.

The last optimality criterion to be investigated in this study is motivated by the following idea. To check if the last $k-s$ exponential terms in model (1) are necessary to fit the data adequately, one could think of performing a statistical test with the null hypothesis

$$H_0 : a_{s+1} = \dots = a_k = 0,$$

or, equivalently, $K_{s,k}^T \theta = 0 \in \mathbb{R}^{k-s}$ for the matrix $K_{s,k} \in \mathbb{R}^{2k \times k-s}$, where the i^{th} column of $K_{s,k}$ is given by the $(2(s+i)-1)^{\text{th}}$ unit vector of dimension $2k$, against the alternative hypothesis

$$H_1 : \exists l \in \{s+1, \dots, k\} \text{ such that } a_l \neq 0$$

or $K_{s,k}^T \theta \neq 0 \in \mathbb{R}^{k-s}$. Under some assumptions of regularity, the noncentrality parameter of the corresponding likelihood ratio test then depends on the underlying design ξ through the matrix

$$(K_{s,k}^T M_k^-(\xi, \theta) K_{s,k})^{-1}.$$

The power of this test can thus be increased by increasing the matrix $(K_{s,k}^T M_k^-(\xi, \theta) K_{s,k})^{-1}$ in the Loewner ordering [see Pukelsheim (1993), chapter 3], which motivates the function

$$\Phi(\xi, \theta) = \det(K_{s,k}^T M_k^-(\xi, \theta) K_{s,k})^{-1} \rightarrow \max_{\xi} \quad (5)$$

as an appropriate optimality criterion for model discrimination in the sense of testing H_0 against H_1 . Note that for information matrices $M_k(\xi, \theta)$ of full rank this optimality criterion simplifies to

$$\Phi(\xi, \theta) = \frac{\det M_k(\xi, \theta)}{\det M_{[s+k]}(\xi, \theta)},$$

where the matrix $M_{[s+k]} \in \mathbb{R}^{(s+k) \times (s+k)}$ is obtained from M_k by deleting the $(2(s+i)-1)^{th}$ rows and columns, $i = 1, \dots, k-s$, in the information matrix M_k . From (5) it follows that the optimality criterion $\Phi(\xi, \theta)$ belongs to the class of D_A -optimality criteria; see, e.g., Silvey (1980), and is thus in some sense related to the $D_{2(k-s)}$ -criterion for discriminating between η_k and η_s defined above. The Φ -optimal design is in this sense the D -optimal design for estimating the subset $\{a_{s+1}, \dots, a_k\}$ of the full parameter vector θ .

All the optimality criteria introduced above are local criteria in the sense of Chernoff (1953), i.e. the optimal designs with respect to these criteria will depend on the (unknown) nonlinear model parameters $\lambda_1, \dots, \lambda_k$. Application of the (locally) optimal designs with respect to these criteria can therefore only be efficient if a good guess for the nonlinear parameters is available. The locally optimal designs, however, provide a basis for the construction of designs, which are more robust with respect to misspecifications of these parameters. Consider, for example, an arbitrary local optimality criterion $\psi(\xi, \theta)$ and assume that the experimenter has at least some knowledge about the parameter values. If sufficient information is available, a prior π on the parameters $\lambda_1, \dots, \lambda_k$ can be specified. This approach leads to so-called Bayesian optimality criteria; see, e.g., Chaloner (1993), where the criterion function $\psi(\xi, \theta)$ is integrated with respect to the prior π , i.e. averaged over the most plausible parameter values. In this article, we will adopt a different approach, which is based on the maximin concept, and only requires the specification of a certain range for the unknown parameters. This seems to be the most realistic scenario in many practical applications, in particular if the number of nonlinear parameters is large. The optimality criteria are constructed as follows. Since the use of standardized criteria is recommended to avoid different scalings; see Dette (1997), we protect the experiment against the worst possible case by maximizing

$$\psi_{-\infty}(\xi) = \min_{\theta \in \Theta} \frac{\psi(\xi, \theta)}{\psi(\xi_{\theta}^*, \theta)}, \quad (6)$$

where Θ is the region of uncertainty for the unknown parameters, and by ξ_{θ}^* the locally ψ -optimal design for the parameter value θ is denoted. The maximization of (6) is thus equivalent to maximizing the minimal ψ -efficiency, so that the resulting designs will either be referred to by maximin ψ -efficient designs or by standardized maximin optimal designs. In the following, the general notation ψ will be replaced by the respective criteria defined above. As an alternative a Bayesian criterion with a non-informative prior could be used. However, it was indicated by Braess and Dette (2005) that optimal designs with respect to maximin criteria usually advise the experimenter to take observations at more different points than optimal designs with respect to Bayesian criteria based on a non-informative prior. As we are interested in optimal designs for model discrimination, optimal designs with a large number of support points are usually preferable, which is our main reason for using the maximin approach.

3 Analytical results

Our first result yields a considerable simplification of the locally optimal design problems at hand by stating an explicit relationship between locally optimal discrimination designs with different parameters. This result holds true for all three local optimality criteria introduced above. Since the locally optimal designs do only depend on the nonlinear parameters λ_i , $i = 1, \dots, k$, but not on the linear parameters a_i , $i = 1, \dots, k$, they will be denoted by ξ_λ^* with $\lambda = (\lambda_1, \dots, \lambda_k)^T$ instead of ξ_θ^* in what follows.

Lemma 3.1 *Let $\xi_\lambda^* = \{x_i^*(\lambda); w_i^*(\lambda); i = 1, \dots, n^*(\lambda)\}$ be a locally T - or Φ -optimal discrimination design with respect to the parameter λ , where $n^*(\lambda)$ denotes the number of support points of $\xi^*(\lambda)$, and $x_i^*(\lambda)$ are the support points with corresponding weights $w_i^*(\lambda)$. Then for any positive constant $\gamma > 0$ the locally optimal discrimination design with respect to the parameter $\gamma\lambda$ is given by*

$$x_i^*(\gamma\lambda) = \frac{1}{\gamma}x_i^*(\lambda), \quad w_i^*(\gamma\lambda) = w_i^*(\lambda).$$

As a consequence of Lemma 3.1, we can without loss of generality restrict ourselves to the derivation of locally optimal designs where one component of the vector of unknown parameters λ is set equal to one.

A further simplification can be achieved by the assertion of the following lemma, which gives the value of one support point of an optimal design with respect to the criteria under consideration.

Lemma 3.2 *Let ξ_λ^* be a locally T - or Φ -optimal discrimination design with respect to the parameter λ . Then the point zero is contained in the support of ξ_λ^* .*

From Caratheodory's Theorem [see, e.g., Karlin and Studden (1966)], we obtain that the number of support points $n^*(\lambda)$ of ξ_λ^* with respect to any local discrimination criterion is bounded by

$$2k - 2s \leq n^*(\lambda) \leq 2k(2k + 1)/2.$$

In the following theorem we will derive considerably sharper upper bounds for $n^*(\lambda)$ to facilitate the numerical calculation of optimal designs.

Theorem 3.1 *Let $\xi_\lambda^* = \{x_i^*(\lambda); w_i^*(\lambda); i = 1, \dots, n^*(\lambda)\}$ be a locally T -optimal discrimination design with respect to the parameter λ , where $n^*(\lambda)$ denotes the number of support points of $\xi^*(\lambda)$. Then we obtain the following results.*

- 1) *For any value of λ the inequality $n^*(\lambda) \leq k(k + 1)/2 + 1$ holds.*
- 2) *For parameters λ of a special form we even find a sharper upper bound. Define a class Λ of vectors λ with components of the form*

$$\lambda_i = (\lambda_{i-1} + \lambda_{i+1})/2, \quad i = 2, \dots, k - 1. \tag{7}$$

There exists a neighborhood $U(\Lambda)$ of Λ such that for all $\lambda \in U(\Lambda)$ the number of support points $n^(\lambda)$ of the T -optimal design is at most $2k$.*

We will now study another important property of the locally optimal designs with respect to the T - and Φ -criterion. Consider that λ is in some neighborhood of the vector $\gamma(1, \dots, 1)^T$ where $\gamma > 0$. As λ converges to $\gamma(1, \dots, 1)^T$ the information matrix M_k defined in (3) becomes singular. The corresponding locally optimal designs, however, are still weakly convergent to some limiting design. At the same time, the following theorem points out some connection between the discrimination design problems for the model at hand and the corresponding design problems in a heteroscedastic polynomial model. Define a regression model with iid errors by

$$E[Y(x)] = \tilde{\eta}_{2k}(x) = \sum_{i=1}^{2k} c_i x^{i-1}, \quad Var(Y(x)) = e^{2\gamma x}, \quad \gamma > 0, \quad (8)$$

where the value of γ is assumed to be known. The corresponding information matrix for the estimation of the polynomial coefficients c_i , $i = 1, \dots, 2k$, is then given by

$$\widetilde{M}_{2k}(\xi) = \int e^{-2\gamma x} \tilde{f}_k(x) \tilde{f}_k^T(x) d\xi(x), \quad \tilde{f}_k^T(x) = (1, x, \dots, x^{2k-1}).$$

Theorem 3.2 *Let $\lambda_i = \gamma - \delta_i$, $\delta_i = r_i \delta$, $i = 1, \dots, k$, and $r_i \in \mathbb{R} \setminus \{0\}$ are fixed numbers with $r_i \neq r_j$ for $i \neq j$ and $r_1 > r_2 > \dots > r_k$, i.e. all components of the vector of nonlinear parameters λ tend to γ for $\delta \rightarrow 0$. If an optimal design ξ with respect to one of the respective criteria is supported on at least $2k$ points so that the information matrix $M_k(\xi, \theta)$ is non-singular the following assertions hold.*

- 1) *If $\delta \rightarrow 0$, and λ is of the form described above, the locally optimal discrimination design η_k vs η_s converges weakly to the corresponding optimal discrimination design $\tilde{\eta}_{2k}$ vs $\tilde{\eta}_{2s}$ for the heteroscedastic polynomial regression model (8) with unknown parameters c_i .*
- 2) *More general, under the same conditions, the T -optimal designs with respect to the criterion $T(\cdot, \theta)$ converge weakly to the corresponding T -optimal design for model (8) if $\delta \rightarrow 0$.*
- 3) *For $\delta \rightarrow 0$ the Φ -optimal discrimination designs converge weakly to the discrimination design for testing the hypothesis $H_0 : c_{2k-1} = c_{2k-3} = \dots = c_{2k-2s+1} = 0$ in model (8) against the alternative that at least one of this parameters is non-zero.*

The restriction to non-singular information matrices in Theorem 3.2 does not mean a great limitation for this result since for all examples in our numerical study given in the following section – as well as for numerous other examples not displayed here for brevity – we found that the optimal designs with respect to all criteria under consideration feature at least $2k$ distinct points of support. Furthermore, it is reasonable to consider only designs with at least $2k$ different support points, because such designs can also be used for the subsequent data analysis if the largest model η_k turns out to be the “true model”.

4 Numerical results

In this section we will apply the results obtained in the previous section to construct optimal designs for all criteria under consideration numerically. All designs presented in this section have

been carefully checked for optimality by the respective corresponding equivalence theorems. For the sake of brevity, we restrict ourselves to the presentation of results for exponential models with not more than six parameters, since these models are most widely used within the whole class of models of type (1).

4.1 Locally optimal designs

As a first example we choose the discrimination problem between the models $\eta_3(x, \theta)$ and $\eta_2(x, \theta)$, i.e. the question if the third term in model (1) is actually necessary to fit the data. Table 1 displays several examples of optimal discrimination designs η_3 vs η_2 . As a consequence of Lemma 3.2 the point $x_1 = 0$ is contained in the support of any locally optimal design ξ_λ^* . We will thus not present it explicitly in what follows. The weight corresponding to the largest optimal support point will also be omitted for brevity.

Table 1: *Selected locally $D_{2(3-2)}$ -optimal discrimination designs for the problem of discriminating between the exponential regression models η_3 vs η_2 .*

λ_1	λ_2	λ_3	x_2	x_3	x_4	x_5	x_6	w_1	w_2	w_3	w_4	w_5	eff $_D$
1	0.2	1.5	0.329	1.276	2.74	5.60	13.75	0.100	0.196	0.174	0.167	0.183	0.96
1	0.2	5	0.156	0.547	1.48	4.06	12.14	0.173	0.270	0.200	0.175	0.108	0.89
1	0.5	1.5	0.288	1.135	2.47	4.57	9.11	0.088	0.172	0.158	0.143	0.166	0.92
1	0.5	5	0.146	0.518	1.30	3.14	7.58	0.159	0.258	0.186	0.164	0.117	0.93
1	0.8	1.5	0.257	1.025	2.26	4.06	7.58	0.078	0.154	0.148	0.137	0.158	0.89
1	0.8	5	0.138	0.494	1.18	2.68	6.11	0.147	0.248	0.178	0.160	0.124	0.94

For comparison, the following table (Table 2) shows the locally optimal designs for testing the hypothesis $H_0 : a_3 = 0$ with respect to the same parameters λ_1 , λ_2 and λ_3 . Tables 1 and 2

Table 2: *Selected locally optimal discrimination designs for testing the hypothesis $H_0 : a_3 = 0$ in the exponential regression model η_3 .*

λ_1	λ_2	λ_3	x_2	x_3	x_4	x_5	x_6	w_1	w_2	w_3	w_4	w_5	eff $_D$
1	0.2	1.5	0.276	1.157	2.853	6.044	14.367	0.067	0.143	0.167	0.206	0.212	0.92
1	0.2	5	0.128	0.571	1.621	4.250	12.342	0.011	0.135	0.282	0.270	0.177	0.87
1	0.5	1.5	0.246	1.020	2.448	4.880	9.696	0.062	0.129	0.143	0.163	0.193	0.87
1	0.5	5	0.120	0.525	1.409	3.315	7.779	0.028	0.135	0.241	0.236	0.176	0.90
1	0.8	1.5	0.223	0.918	2.182	4.266	8.145	0.058	0.120	0.129	0.147	0.182	0.84
1	0.8	5	0.113	0.489	1.273	2.851	6.311	0.040	0.135	0.216	0.218	0.177	0.91

also display the D -efficiency

$$\text{eff}_D(\xi) = \left(\frac{\det(M_k(\xi, \theta))}{\det(M_k(\xi_D, \theta))} \right)^{1/(2k)}$$

of the respective designs with respect to the D -optimal design ξ_D in the model η_k to evaluate how the optimal discrimination design performs for estimating the model parameters if η_k has been found to be the true model. This gives us the opportunity to compare the different optimality criteria under consideration in this article. We find that the $D_{2(3-2)}$ -optimal designs have somewhat higher D -efficiencies than the optimal designs for testing $a_3 = 0$. The $D_{2(k-s)}$ -optimal designs therefore have a slight advantage over their competitors in the multiple-objective sense, i.e. they are more efficient in the parameter estimation subsequent to the model discrimination procedure. There is a heuristic explanation for this property. Note that the $D_{2(k-s)}$ -optimal design is optimal for estimating the $2(k-s)$ parameters $\alpha_k, \dots, \alpha_{s+1}$ and $\lambda_k, \dots, \lambda_{s+1}$ in model η_k , whereas the optimal design for testing $\alpha_k = \dots = \alpha_{s+1} = 0$ in model η_k is in fact optimal for estimating the $k-s$ parameters $\alpha_k, \dots, \alpha_{s+1}$. A design, which is by the construction of its optimality criterion optimal for estimating $2(k-s)$ parameters will thus be more efficient for estimating $2k$ parameters compared with a design that is optimal for the estimation of $k-s$ parameters in the same model. It is therefore recommended to use the $D_{2(k-s)}$ -optimal design if the data collected for discrimination between models η_k and η_s are also to be employed for parameter estimation.

In the example $\lambda = (1, 0.5, 1.5)^T$, part 2) of Theorem 3.1 applies so that the upper bound for $n^*(\lambda)$ in the design problem η_3 vs η_2 is given by $2k = 6$. From Table 1 we obtain that the optimal design attains this bound, i.e. the bound is sharp. It turns out that the other optimal designs in the examples given in Table 1 are also supported at 6 points, i.e. the bound $k(k+1)/2 + 1 = 7$ from part 1) of Theorem 3.1 is not necessarily attained by the optimal design. Since the two criteria under consideration are relatively similar the optimal designs with respect to these criteria are also close. We further observe from Tables 1 and 2 that if λ_1 and λ_2 are close the optimal support points are less spread on the real axis than for more distant values λ_1 and λ_2 for both criteria.

From a heuristic point of view, it seems reasonable to use designs, which allocate equal weights to some equidistant points on the real axis. In order to investigate the performance of such uniform designs, we calculated several of their $D_{2(3-2)}$ -efficiencies, where the $D_{2(k-s)}$ -efficiency $\text{eff}_{D_{2(k-s)}}(\xi)$ of a design ξ is defined by

$$\text{eff}_{D_{2(k-s)}}(\xi) = \left(\frac{\det C_{A_s, k}(\xi, \theta)}{\det C_{A_s, k}(\xi_\lambda^*, \theta)} \right)^{1/(2(k-s))}.$$

We choose as competing designs three different uniform distributions with their supports containing the point zero, namely

1. the seven point design ξ_7 allocating weight $1/7$ to the points 0, 2, 4, 6, 8, 10, 12. This design was chosen since the upper bound for the number of support points of the optimal design for the problem under consideration is precisely given by seven. Often the

experimenter does not want to take measurements at too many different experimental conditions, so that ξ_7 seems attractive from this point of view.

2. the design ξ_{16} assigning equal weight to the 16 points $0, 1, 2, \dots, 15$, which is more spread on the real axis than ξ_7 . This design could also be used for further model checks since it contains significantly more different support points than the number of parameters to be estimated in the largest model η_3 .
3. the uniform design ξ_{100} on the 100 points $(n-1)/10, n = 1, 2, \dots, 100$, approximating the uniform distribution on $[0, 10]$.

The parameter combinations for this example are given by $\lambda = (1, 0.5, 1.5)^T$ representing the situation of relatively “similar” exponential terms in model (1) and $\lambda = (1, 0.2, 5)^T$ representing very “different” exponential terms. The $D_{2(3-2)}$ -efficiencies of the uniform designs defined above given these parameter values for λ are shown in Table 3. For comparison, we also present the $D_{2(3-2)}$ -efficiencies of the designs ξ_3^* , which are locally optimal for testing $H_0 : a_3 = 0$ in model η_3 , as well as the $D_{2(3-2)}$ -efficiencies of the locally D - and E -optimal designs in the largest model η_3 , ξ_D and ξ_E , respectively.

Table 3: *Selected $D_{2(3-2)}$ -efficiencies of some uniform designs for the discrimination problem η_3 vs η_2 .*

λ_1	λ_2	λ_3	ξ_7	ξ_{16}	ξ_{100}	ξ_3^*	ξ_D	ξ_E
1	0.5	1.5	0.004	0.198	0.560	0.897	0.88	0.90
1	0.2	5	$\approx 10^{-6}$	0.003	0.212	0.426	0.87	0.91

We learn from Table 3 that the different uniform designs under consideration will have a very poor performance if applied to the problem at hand. Only the design ξ_{100} might yield reasonable results for particular choices of λ , but it is not very attractive from a practical point of view since it requires to take measurements at 100 different experimental conditions. The designs ξ_3^* , ξ_D and ξ_E , in contrast, are expected to show a good performance. The use of an optimal design with respect to either one of the criteria defined in the previous section or a good criterion for estimating the full parameter vector in the largest model under consideration is thus strongly recommended.

We now turn to the discrimination problem between models η_3 and η_1 . Several examples for this problem are presented in Tables 4 and 5. For comparison, we also display the corresponding D -efficiencies eff_D .

According to Lemma 3.2 the supports of the optimal designs again contain the point zero, but it is allocated considerably less weight than in the previous examples. The highest optimal support point, however, is given much more weight than before, particularly for the designs, which are optimal for testing the hypothesis $H_0 : a_2 = a_3 = 0$. Moreover, we can see from Tables 4 and 5 that for these design problems the optimal designs are also supported at exactly

Table 4: Selected locally optimal discrimination designs for the problem η_3 vs η_1 .

λ_1	λ_2	λ_3	x_2	x_3	x_4	x_5	x_6	w_1	w_2	w_3	w_4	w_5	eff _D
1	0.2	1.5	0.374	1.312	3.033	5.769	11.921	0.066	0.127	0.109	0.199	0.249	0.892
1	0.2	5	0.166	0.574	1.751	4.185	10.104	0.102	0.146	0.133	0.130	0.239	0.939
1	0.5	1.5	0.340	1.135	2.600	4.662	8.177	0.073	0.134	0.117	0.181	0.245	0.909
1	0.5	5	0.159	0.538	1.471	3.274	6.468	0.107	0.148	0.146	0.130	0.219	0.951
1	0.8	1.5	0.311	1.008	2.302	4.082	6.896	0.079	0.139	0.124	0.170	0.239	0.922
1	0.8	5	0.151	0.507	1.301	2.808	5.279	0.110	0.150	0.155	0.133	0.203	0.959

Table 5: Selected locally optimal discrimination designs for testing the hypothesis $H_0 : a_2 = a_3 = 0$ in model η_3 .

λ_1	λ_2	λ_3	x_2	x_3	x_4	x_5	x_6	w_1	w_2	w_3	w_4	w_5	eff _D
1	0.2	1.5	0.301	1.273	3.161	5.749	13.272	0.045	0.099	0.127	0.186	0.206	0.812
1	0.2	5	0.118	0.607	1.800	3.907	11.558	0.005	0.087	0.196	0.203	0.219	0.593
1	0.5	1.5	0.272	1.128	2.668	4.809	8.866	0.050	0.107	0.126	0.154	0.171	0.807
1	0.5	5	0.124	0.576	1.544	3.166	7.166	0.011	0.089	0.187	0.187	0.178	0.656
1	0.8	1.5	0.247	1.016	2.362	4.257	7.434	0.053	0.110	0.124	0.142	0.163	0.803
1	0.8	5	0.122	0.542	1.387	2.762	5.787	0.018	0.092	0.175	0.177	0.169	0.710

$2k = 6$ points. We observed this for numerous other examples, which are not displayed here for the sake of brevity.

In the next examples we are interested in designs, which are good for discriminating between models η_3 and η_2 on the one hand, and η_2 and η_1 on the other hand. Table 6 displays optimal discriminating designs with respect to the prior $\beta_3 = 2/3$ and $\beta_2 = 1/3$, i.e. the discrimination between the models η_3 vs η_2 is assumed twice as important as the problem η_2 vs η_1 . In Table 7 we show the same examples with more emphasis placed on the discrimination problem η_2 vs η_1 .

As expected, the T -optimal designs presented in Tables 6 and 7 are located in some sense between the “pure” optimal designs for optimality criterion given in Tables 1 and 4. The designs from Table 6 are “closer” to the corresponding designs in Table 1, whereas the designs from Table 7 are “closer” to their counterparts in Table 4, which is in line with intuition.

4.2 Maximin optimal designs

In the last paragraph of this section, we will consider the problem of robustness of designs with respect to the choice of the initial parameters. The designs presented above are all locally

Table 6: Selected locally T -optimal discrimination designs for the problems η_3 vs η_2 (weight $\beta_3 = 2/3$) and η_2 vs η_1 (weight $\beta_2 = 1/3$).

λ_1	λ_2	λ_3	x_2	x_3	x_4	x_5	x_6	w_1	w_2	w_3	w_4	w_5	eff $_D$
1	0.2	1.5	0.355	1.306	2.915	5.732	12.550	0.078	0.151	0.132	0.187	0.232	0.935
1	0.2	5	0.163	0.565	1.677	4.248	10.503	0.128	0.189	0.157	0.152	0.195	0.983
1	0.5	1.5	0.318	1.142	2.555	4.621	8.534	0.078	0.147	0.132	0.167	0.219	0.926
1	0.5	5	0.154	0.531	1.418	3.288	6.768	0.125	0.187	0.159	0.146	0.189	0.980
1	0.8	1.5	0.288	1.017	2.295	4.064	7.160	0.078	0.144	0.134	0.159	0.209	0.921
1	0.8	5	0.147	0.503	1.261	2.808	5.531	0.123	0.185	0.161	0.145	0.183	0.978

Table 7: Selected locally T -optimal discrimination designs for the problems η_3 vs η_2 (weight $\beta_3 = 1/3$) and η_2 vs η_1 (weight $\beta_2 = 2/3$).

λ_1	λ_2	λ_3	x_2	x_3	x_4	x_5	x_6	w_1	w_2	w_3	w_4	w_5	eff $_D$
1	0.2	1.5	0.401	1.300	3.191	5.757	11.248	0.054	0.101	0.086	0.215	0.250	0.819
1	0.2	5	0.169	0.585	1.783	4.087	9.857	0.076	0.101	0.109	0.098	0.288	0.836
1	0.5	1.5	0.374	1.114	2.650	4.721	7.728	0.068	0.122	0.099	0.200	0.265	0.872
1	0.5	5	0.163	0.548	1.506	3.194	6.225	0.088	0.107	0.136	0.107	0.243	0.883
1	0.8	1.5	0.345	0.994	2.296	4.122	6.557	0.080	0.136	0.110	0.183	0.270	0.903
1	0.8	5	0.157	0.511	1.340	2.742	5.049	0.098	0.122	0.152	0.116	0.213	0.909

optimal, i.e. optimal with respect to one particular parameter λ . If not even the number k of exponential terms in model (1) is known then presumably less knowledge will be available about the values of the model parameters. We will therefore also present some examples of optimal discrimination designs in the case that only some range for each nonlinear parameter can be given in advance of the experiment. As a first example, we consider the problem of designing an experiment for discriminating between models η_2 and η_1 adopting the maximin approach described in (6). The criterion function $\psi_{-\infty}(\xi)$ is thus given by

$$\psi_{-\infty}(\xi) = \min_{\theta \in \Theta} \frac{\det C_{A_{1,2}}(\xi, \theta)}{\det C_{A_{1,2}}(\xi_{\theta}^*, \theta)}.$$

Some selected examples of optimal designs with respect to this standardized maximin criterion are given in Table 8. The indices l and u denote the lower and upper boundaries, respectively, for the interval, in which a parameter λ_i , $i = 1, 2$, is assumed to lie. The right column in Table 8 shows the minimal efficiency of the maximin optimal design in the rectangular region $[\lambda_1(l), \lambda_1(u)] \times [\lambda_2(l), \lambda_2(u)]$. We observe that the maximin optimal designs yield reasonable efficiencies over a broad range of the parameter space. Even for the large parameter range of

Table 8: *Selected standardized maximin optimal discrimination designs for the problem η_2 vs η_1 .*

$\lambda_1(l)$	$\lambda_1(u)$	$\lambda_2(l)$	$\lambda_2(u)$	x_1	x_2	x_3	x_4	x_5	w_1	w_2	w_3	w_4	w_5	min eff
0.8	1	1.1	1.3	0	0.47	1.78	4.08	–	0.12	0.23	0.22	0.44	–	0.9665
0.6	1	1.1	1.5	0	0.48	1.76	4.21	–	0.13	0.24	0.23	0.40	–	0.8658
0.4	1	1.1	1.7	0	0.46	1.46	3.17	5.94	0.14	0.22	0.19	0.25	0.20	0.7739
0.2	1	1.1	1.9	0	0.48	1.50	3.41	7.22	0.15	0.22	0.19	0.27	0.17	0.7194

$[0.2, 1] \times [1.1, 1.9]$ corresponding to a high uncertainty with respect to the model parameters the efficiency is about 72%.

From Theorem 3.1 we conclude that locally optimal designs for the problem η_2 vs η_1 have at most $k(k+1)/2 + 1 = 4$ support points. For the maximin optimal designs this is in general not true. Table 8 shows that for relatively small parameter regions Θ the standardized maximin optimal designs are also supported at four points, whereas for larger regions, i.e. more uncertainty about the actual position of the parameters, more than four support points are necessary. This phenomenon is widely observed in the Literature as far as Bayesian or maximin optimal design problems are concerned. For a theoretical explanation we refer to Braess and Dette (2005).

As a second example we finally consider the problem of discriminating between models η_3 and η_2 . Some selected standardized maximin optimal designs for this problem are displayed in Table 9, where the parameter regions Θ are specified by

$$\begin{aligned}\Theta_1 &= [0.6, 0.8] \times [1.1, 1.3] \times [1.6, 1.8] \\ \Theta_2 &= [0.8, 1] \times [1.1, 1.3] \times [1.4, 1.6] \\ \Theta_3 &= [0.6, 0.9] \times [1, 1.4] \times [1.5, 1.8].\end{aligned}$$

Table 9: *Selected standardized maximin optimal discrimination designs for the problem η_3 vs η_2 . (It is not necessary to show the weight w_7 of the optimal design with respect to Θ_3 since $w_7 = 1 - \sum_{i=1}^6 w_i$.)*

Θ	x_1	x_2	x_3	x_4	x_5	x_6	x_7	w_1	w_2	w_3	w_4	w_5	w_6	min eff
Θ_1	0	0.24	0.94	2.07	3.82	7.18	–	0.083	0.162	0.153	0.144	0.166	0.293	0.923
Θ_2	0	0.24	0.94	2.09	3.77	6.76	–	0.070	0.142	0.141	0.140	0.165	0.342	0.932
Θ_3	0	0.24	0.93	2.03	3.66	5.18	7.23	0.081	0.157	0.146	0.142	0.145	0.069	0.817

Again, we observe that for larger parameter regions Θ the number of support points of the standardized maximin optimal design increases. Moreover, even in the case of three unknown parameters the maximin optimal designs yield reasonable efficiencies over the different parameter spaces.

5 Appendix

5.1 Proof of Lemma 3.1

Note that for each choice of $\gamma > 0$ and each design $\xi = (x_1, \dots, x_l, w_1, \dots, w_l)$ we obtain

$$M_k(x_1, \dots, x_l, w_1, \dots, w_l, \lambda) = B^T M_k(x_1/\gamma, \dots, x_l/\gamma, w_1, \dots, w_l, \gamma \cdot \lambda) B$$

where $B = \text{diag}(1, \gamma, \dots, 1, \gamma) \in \mathbb{R}^{2k \times 2k}$. The assertion of the lemma then follows from elementary properties of determinants and matrix inversion. \square

5.2 Proof of Lemma 3.2

We consider a design of the form

$$\xi_h = \begin{pmatrix} x_1 + h & \dots & x_n + h \\ w_1 & \dots & w_n \end{pmatrix}$$

with $h \geq 0$. The determinant of the Fisher information of such a design can be expressed as

$$\det M_l(\xi_h, \theta) = e^{-4(\lambda_1 + \dots + \lambda_l)h} \det M_l(\xi_0, \theta).$$

Now it is obvious that the discrimination criteria considered in this article are decreasing with respect to h . So for each design with minimal support point greater than zero we can find a better design with respect to the optimality criteria, where the minimal support point is equal to zero. Consequently, the minimal support point of an optimal design must be equal to zero.

5.3 Proof of Theorem 3.1

One of the basic tools for research in the field of optimal experimental design are equivalence theorems; see, for example, Kiefer (1974). As we will need the equivalence theorem with respect to the T -optimal design criterion for the proof of Theorem 3.1, it will be stated below. It is given here in the (simpler) form for designs with at least $2k$ support points (so that the information matrix M_k of the largest model under consideration is invertible) since this form serves our purpose best.

Theorem 5.1 *A design ξ^* with at least $2k$ support points is an optimal design for discriminating the class $\{\eta_1, \dots, \eta_k\}$ with respect to the prior $\{\beta_1, \dots, \beta_k\}$ if and only if the following inequality holds.*

$$\psi(x) = \sum_{l=1}^k \beta_l (f_l^T(x, \theta) M_l^{-1}(\xi^*, \theta) f_l(x, \theta) - f_{l-1}^T(x, \theta) M_{l-1}^{-1}(\xi^*, \theta) f_{l-1}(x, \theta)) - 2 \sum_{l=1}^k \beta_l \leq 0 \quad (9)$$

Moreover, there is equality in (9) for any support point of the optimal design ξ^* .

Since the criterion for discrimination between models η_k and η_s is a special case of the compound criterion $T(\xi)$ the corresponding equivalence theorem is obtained from Theorem 5.1 with $\beta_l = 1$ for $l > s$ and $\beta_l = 0$ otherwise.

For the proof of Theorem 3.1 we will also need the following auxiliary lemma.

Lemma 5.1 *Consider the regression model $\bar{\eta}_k(x, \beta) = \sum_{i=1}^k (\beta_{2i-1} e^{-\lambda_i x} + \beta_{2i} e^{-(\lambda_i + \Delta)x})$, where $\beta_1, \dots, \beta_{2k}$ are unknown parameters and $\lambda_1, \dots, \lambda_k$ and Δ are known values such that $0 < \Delta < \min_{i,j} |\lambda_i - \lambda_j|$. Then the number n of support points of the T -optimal discrimination design in this model is bounded by $n \leq k(k+1)/2 + 1$. If additionally λ is in a neighborhood $U(\Lambda)$ of a class Λ as defined in (7) an upper bound is given by $n \leq 2k$.*

Proof of Lemma 5.1. We assume that the optimal design is supported at n different points x_1^*, \dots, x_n^* . Denote by $\bar{\psi}(x)$ the directional derivative from the equivalence theorem for model $\bar{\eta}_k$. It then follows that the following equations hold

$$\bar{\psi}(x_i^*) = 0, \quad i = 1, \dots, n, \quad \bar{\psi}'(x_i^*) = 0, \quad i = 2, \dots, n-1, \quad (10)$$

i.e. the function $\bar{\psi}(x)$ has at least $2n - 2$ roots counting multiplicities. We consider the set $\{\lambda_i + \lambda_j\}_{i,j=1,\dots,k}$ and refer to its different elements by u_1, \dots, u_r (without loss of generality $u_1 < \dots < u_r$). Obviously, $r \leq k(k+1)/2$ and for $\lambda \in \Lambda$ in the sense of (7) we obtain $r = 2k - 1$. The function $\bar{\psi}(x)$ can be rewritten as a sum of $2r + 2$ functions as follows.

$$\bar{\psi}(x) = \sum_{i=1}^{2r+1} \bar{g}_i(x) - \bar{g}_0(x),$$

where

$$\begin{aligned} \bar{g}_0(x) &\equiv 2 \sum_{l=1}^k \beta_l, \\ \bar{g}_1(x) &= c_l e^{-u_1 x}, \\ \bar{g}_2(x) &= -b_1 e^{-(u_1 + \Delta)x}, \\ \bar{g}_{2l+1}(x) &= a_1 e^{-(u_l + 2\Delta)x} + c_l e^{-u_{l+1}x}, \quad l = 1, \dots, r-1, \\ \bar{g}_{2l}(x) &= -b_1 e^{-(u_l + \Delta)x}, \quad l = 2, \dots, r, \\ \bar{g}_{2r+1}(x) &= a_{r+1} e^{-(u_r + \Delta)x}. \end{aligned}$$

In the following we will show that the coefficients of these functions are all positive, i.e. a_l, b_l and $c_l > 0$. It is then easy to see that the $2r+2$ functions $\bar{g}_0(x), \dots, \bar{g}_{2r+1}(x)$ form a Tchebycheff system; see Karlin and Studden (1966). The function $\bar{\psi}(x)$ can thus have at most $2r + 1$ roots. Since it follows from (10) that the number of roots is at least $2n - 2$ we receive the inequality $2n - 2 \leq 2r + 1$ which implies $n \leq r + 3/2$. Plugging in the upper bounds for r derived above we obtain that $n \leq k(k+1)/2 + 1$ for general λ . If $\lambda \in \Lambda$ of the form (7) we get the bound $n \leq 2k$. From continuity considerations it then follows that this bound is also valid in some neighborhood of Λ .

It remains to show that $a_l > 0$, $b_l > 0$ and $c_l > 0$ for all l . For simplicity we assume that $\beta_k = 1$, $\beta_{k-1} = \dots = \beta_1 = 0$. (Otherwise we will simply have a sum of similar terms.) Note that the function from the equivalence theorem

$$\bar{\psi}(x) = \bar{f}_k^T(x) \bar{M}_k^{-1}(\xi^*) \bar{f}_k(x) - \bar{f}_{k-1}^T(x) \bar{M}_{k-1}^{-1}(\xi^*) \bar{f}_{k-1}(x) - 2$$

can be rewritten in the form

$$\bar{\psi}(x) = \bar{f}_k^T(x) A \bar{f}_k(x) - 2, \text{ where } A = \bar{M}_k^{-1}(\xi^*) - \begin{pmatrix} \bar{M}_{k-1}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

The first goal is to show that $\text{sign}(A) = (-1)^{i+j}$. For this we divide the matrices \bar{M}_k and A into blocks of the same length

$$\bar{M}_k(\xi) = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \quad A = \begin{pmatrix} M_{11}^{-1} M_{12} P^{-1} M_{21} M_{11}^{-1} & -M_{11}^{-1} M_{12} P^{-1} \\ -P^{-1} M_{21} M_{11}^{-1} & M_{11}^{-1} \end{pmatrix} \quad (11)$$

with $M_{11} = \bar{M}_{k-1}(\xi)$. The representation of the matrix A follows from the inversion formula for block matrices, where $P = M_{22} - M_{21} M_{11}^{-1} M_{12}$ is the Schur complement of M_{11} . As a consequence of Karlin and Studden (1966), p. 9, and the Cauchy-Binet formula we find that $\text{sign}(\bar{M}_k^{-1}) = (-1)^{i+j}$ or, in other words,

$$J \bar{M}_k J >_c 0,$$

where $J = \text{diag}\{+1, -1, \dots, +1, -1\} \in \mathbb{R}^{2k \times 2k}$, and the notation ' $>_c 0$ ' means that all entries of a matrix are greater than 0. From the representation of A in (11) we conclude that

$$J A_{22} J >_c 0, \quad J A_{12} J = -J M_{11}^{-1} M_{12} P^{-1} J >_c 0 \text{ and } J A_{21} J >_c 0.$$

Furthermore, from a straightforward but tedious calculation it follows that

$$J A_{11} J = J M_{11}^{-1} M_{12} P^{-1} P P^{-1} M_{21} M_{11}^{-1} J >_c 0$$

and consequently the result $J A J >_c 0$ holds for the matrix A . It can therefore easily be checked that $a_l > 0$, $b_l > 0$ and $c_l > 0$ in the expressions for $g_i(x)$. \square

The same assertion as in Lemma 5.1 remains to be proven for models of the form (1) to obtain the results of Theorem 3.1.

Proof of Theorem 3.1. We denote by $n^*(\lambda)$ the number of distinct support points $x_1^*, \dots, x_{n^*(\lambda)}^*$ of the optimal design ξ^* . Then it follows from the equivalence theorem that the following equations hold

$$\psi(x_i^*) = 0, \quad i = 1, \dots, n^*(\lambda), \quad \psi'(x_i^*) = 0, \quad i = 2, \dots, n^*(\lambda) - 1.$$

The directional derivative $\psi(x)$ has therefore at least $2n^*(\lambda) - 2$ roots (counting multiplicities). Using the equalities

$$\lim_{\Delta \rightarrow 0} L^T \bar{f}_k(x) = f_k(x), \quad \lim_{\Delta \rightarrow 0} L^T \bar{M}_l L = M_l, \quad l = 1, \dots, k, \quad \lim_{\Delta \rightarrow 0} \frac{L^T \bar{M}_k L}{L^T \bar{M}_s L} = \frac{M_k}{M_s},$$

where

$$L = \text{diag}\{L_0, \dots, L_0\}, \quad L_0 = \begin{pmatrix} 1 & -1/\Delta \\ 0 & 1/\Delta \end{pmatrix},$$

we receive that

$$\psi(x) = \lim_{\Delta \rightarrow 0} \bar{\psi}(x).$$

Due to continuity the number of roots of the function $\psi(x)$ as the limit of the sequence $\bar{\psi}(x)$ with respect to $\Delta \rightarrow 0$ cannot be larger than the number of roots of $\bar{\psi}(x)$ for some fixed value of Δ . From Lemma 5.1 it then follows that for general λ we obtain the upper bound for the number of support points of the optimal design as

$$n^*(\lambda) \leq k(k+1)/2 + 1.$$

For λ in a class Λ as defined in (7) we even get $n^*(\lambda) \leq 2k$. By continuity, again we get that the bound $n^*(\lambda) \leq 2k$ is still valid for values of λ in some neighborhood $U(\Lambda)$ of Λ . \square

5.4 Proof of Theorem 3.2

The proof is similar to the proofs in Dette, Melas and Pepelyshev (2004a) for more general nonlinear models, so that we give only a sketch including the most important steps. By a Taylor expansion of the vector $f_k(x)$ we obtain the representation

$$f_k(x) = L_k \tilde{f}_k(x) + H(\delta),$$

where $H(\delta) = (o(\delta^{2k-1}), o(\delta^{2k-2}), \dots, o(\delta^{2k-1}), o(\delta^{2k-2}))^T$, $\tilde{f}_k(x) = (1, x, x^2, \dots, x^{2k-1})^T e^{-\gamma x}$,

$$L_k = \begin{pmatrix} 1 & \delta_1 & \frac{\delta_1^2}{2!} & \frac{\delta_1^3}{3!} & \cdots & \frac{\delta_1^{2k-1}}{(2k-1)!} \\ 0 & 1 & \delta_1 & \frac{\delta_1^2}{2!} & \cdots & \frac{\delta_1^{2k-2}}{(2k-2)!} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \delta_k & \frac{\delta_k^2}{2!} & \frac{\delta_k^3}{3!} & \cdots & \frac{\delta_k^{2k-1}}{(2k-1)!} \\ 0 & 1 & \delta_k & \frac{\delta_k^2}{2!} & \cdots & \frac{\delta_k^{2k-2}}{(2k-2)!} \end{pmatrix}$$

and $\det L_k$ is proportional to $\prod_{i < j} (\delta_i - \delta_j)^4$; see Melas (1978). Consequently the information matrix in the general exponential regression model (1) satisfies

$$M_k^{-1}(\xi) = L_k^{-1T} \widetilde{M}_{2k}^{-1}(\xi) L_k^{-1} + o(\delta^{4k-2}),$$

where $\widetilde{M}_{2k}(\xi)$ denotes the information matrix for the parameter vector in the regression model $\tilde{\eta}_{2k}(x)$ defined in (8). For $\delta \rightarrow 0$ we obtain

$$\det A_{s,k}^T M_k^{-1}(\xi, \theta) A_{s,k} = \frac{\det M_k(\xi, \theta)}{\det M_s(\xi, \theta)} \sim \frac{(\det L_k)^2}{(\det L_s)^2} \cdot \frac{\det \widetilde{M}_{2k}(\xi)}{\det \widetilde{M}_{2s}(\xi)},$$

which proves the first assertion of Theorem 3.2 since the $D_{2(k-s)}$ -criterion minimizes the expression $\det A_{s,k}^T M_k^{-1}(\xi, \theta) A_{s,k}$.

For the T -optimal design we use a similar argument and rewrite the criterion in the form

$$\prod_{l=1}^k \left(\frac{\det M_k(\xi, \theta)}{\det M_s(\xi, \theta)} \right)^{\beta_l} \rightarrow \max_{\xi}.$$

Consequently

$$\arg \max_{\xi} \sum_{l=1}^k \beta_l \ln \frac{\det M_l(\xi, \theta)}{\det M_{l-1}(\xi, \theta)} \sim \arg \max_{\xi} \sum_{l=1}^k \beta_l \ln \frac{\det \widetilde{M}_{2l}(\xi)}{\det \widetilde{M}_{2l-2}(\xi)}.$$

To prove part 3 of Theorem 3.2 we denote by $L_{[s+k]}$ the $(s+k) \times (s+k)$ matrix, which is obtained from L_k by deleting the $(2(s+i)-1)^{th}$ rows and columns for $i = 1, \dots, k-s$. Then we have for $\delta \rightarrow 0$

$$\Phi(\xi) \sim \frac{(\det L_k)^2}{(\det L_{[s+k]})^2} \cdot \frac{\det \widetilde{M}_{2k}(\xi)}{\det \widetilde{M}_{s+k}(\xi)},$$

which completes the proof of the theorem. □

Acknowledgements: The support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt A2, Sachbeihilfe De 502/18-1, 436 RUS 113/712/0-1) is gratefully acknowledged.

References

- [1] Alvarez, I., Virto, R., Raso, J. and Condon, S. (2003). Comparing predicting models for the Escherichia coli inactivation by pulsed electric fields. *Innovative Food Science and Emerging Technologies*, 4, 195-202.
- [2] Atkinson, A.C. and Fedorov, V.V. (1975). Optimal design: Experiments for discriminating between several models. *Biometrika*, 62, 289-303.
- [3] Becka, M., Bolt, H.M. and Urfer, W. (1992). Statistical analysis of toxicokinetic data by nonlinear regression. (example: inhalation pharmacokinetics of propylene). *Archives of Toxicology*, 66, 450-453.
- [4] Becka, M., Bolt, H.M. and Urfer, W. (1993). Statistical evaluation of toxicokinetic data. *Environmetrics*, 4, 311-322.
- [5] Biedermann, S. and Dette, H. (2005). Numerical Construction of Parameter Maximin D -Optimal Designs for Binary Response Models. To appear in: *South African Statistical Journal*, September issue.

<http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm>

- [6] Braess D. and Dette H. (2005). On the number of support points of maximin and Bayesian D -optimal designs in nonlinear regression models.
<http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm>
- [7] Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear problems. *Journal of Statistical Planning and Inference*, 37, No.2, 229-235.
- [8] Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24, 586-602.
- [9] Dette, H. (1990). A generalization of D - and D_1 -optimal designs in polynomial regression. *Annals of Statistics*, 18, 1784-1805.
- [10] Dette, H. (1994a). Discrimination designs for polynomial regression on a compact interval. *Annals of Statistics*, 22, 890-904.
- [11] Dette, H. (1994b). Discussion of the paper: "Constrained optimization of experimental design." by V. Fedorov and D. Cook, *Statistics*, 26, 153-161.
- [12] Dette, H. (1997). Designing experiments with respect to standardized optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59, No.1, 97-110.
- [13] Dette, H. and Biedermann, S. (2003). Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association*, 98, 679-686.
- [14] Dette, H., Melas, V.B. and Pepelyshev, A. (2004a). Optimal designs for a class of nonlinear regression models. *Annals of Statistics*, 32, No.5, 2142 - 2167.
- [15] Dette, H., Melas, V.B. and Pepelyshev, A. (2004b). Locally E -optimal designs for exponential regression models. preprint.
<http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm>
- [16] Dette, H., Melas, V.B. and Wong, W.K. (2004). Optimal design for goodness-of-fit of the Michaelis-Menten enzyme kinetic function. preprint.
<http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm>
- [17] Dette, H., Melas, V.B. and Wong, W.K. (2005). Locally D -optimal designs for exponential regression models. preprint. To appear in: *Statistica Sinica*
- [18] Dette, H. and Röder, I. (1997). Optimal discrimination designs for multifactor experiments. *Annals of Statistics*, 25, No.3, 1161-1175.
- [19] Ermakov, S.M. and Melas, V.B. (1995). *Design and Analysis of Simulation Experiments*. Kluwer Academic Publishers, Dordrecht, London.

- [20] Ford, I., Torsney, B. and Wu, C.F.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society*, 54, 569-583.
- [21] Gibaldi, M. and Perrier, D. (1982). *Pharmacokinetics*. Marcel Dekker, New York.
- [22] Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, 40, 633-643.
- [23] Karlin, S. and Studden, W.J. (1966). *Tchebycheff systems: with applications in analysis and statistics*. Wiley, New York.
- [24] Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2, 849-879.
- [25] Landaw, E.M. (1985). Robust sampling designs for compartmental models under large prior eigenvalue uncertainties. *Mathematics and Computers in Biomedical Applications*, 181-187.
- [26] Läuter, E. (1974). Experimental design for a class of models. *Mathematische Operationsforschung und Statistik*, 5, 379-398.
- [27] Melas, V.B. (1978). Optimal designs for exponential regression. *Mathematische Operationsforschung und Statistik, Series Statistics*, 9, 45-59.
- [28] Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.
- [29] Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs. *Biometrika*, 79, 763-770.
- [30] Silvey, S.D. (1980). *Optimal Design*. Chapman and Hall, London.
- [31] Stigler, S.M. (1971). Optimal experimental design for polynomial regression. *Journal of the American Statistical Association*, 66, 311-318.
- [32] Studden, W.J. (1982). Some robust-type D-optimal designs in polynomial regression. *Journal of the American Statistical Association*, 77, 916-921.
- [33] Tsai, M. and Zen, M. (2004). Criterion robust optimal designs for model discrimination and parameter estimation: multivariate polynomial regression case. *Statistica Sinica*, 14(2), 591-601.
- [34] Zen, M. and Tsai, M. (2002). Some criterion-robust optimal designs for the dual problem of model discrimination and parameter estimation. *Sankhya, Series B*, 64, No. 3, 322-338.
- [35] Zen, M. and Tsai, M. (2004). Criterion-robust optimal designs for model discrimination and parameter estimation in Fourier regression models. *Journal of Statistical Planning and Inference*, 124, No. 2, 475-487.