Locally D-optimal Designs for Exponential Regression

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 321 4559

Viatcheslav B. Melas
St. Petersburg State University
Department of Mathematics
St. Petersburg
Russia
email: v.melas@pobox.spbu.ru

Weng Kee Wong
University of Los Angeles
Department of Biostatistics
Los Angeles
USA
email:wkwong@ucla.edu

Abstract

We study locally D-optimal designs for some exponential models that are frequently used in the biological sciences. The model can be written as an algebraic sum of two or three exponential terms. We show that approximate locally D-optimal designs are supported at a minimal number of points and construct these designs numerically.

1 Introduction

Nonlinear regression models are widely used to describe the dependencies between a response and an explanatory variable, see for example, Seber and Wild (1989), Ratkowsky (1983) or Ratkowsky (1990). An appropriate choice of the experimental conditions can improve the quality of statistical inference substantially and therefore many authors have discussed the problem of designing experiments for nonlinear regression models. The main purpose of the present paper is to construct locally D-optimal designs for a class of exponential regression models that is widely used in the biological sciences, see for example, Green and

Reilly (1975), Endrenyi and Dingle (1982), Bardsley, el at. (1986), Droz el at. (1999) and Mason and Wilson, (1999). These models are particularly common in pharmacokinetics and they are called compartmental models (Rowland and Tozer,1989, Shargel and Yu, 1985). Typically, the expected mean response in concentration units, is expressed simply as a linear combination of exponential terms. Such a model is suitable for modeling an identifiable, open, noncyclic n-compartmental system with bolus input into the sampled pool (Landaw, 1985). Compartmental models are also used in data analysis in toxicokinetic experiments (Urfer, 1996 and, Becka, Bolt and Urfer, 1993) and in chemical kinetics (Gibaldi and Perrier, 1982). The simplest forms of such models are

$$a_1 e^{b_1 t} + a_2 e^{b_2 t} \tag{1}$$

$$a_1 e^{b_1 t} + a_2 e^{b_2 t} + a_3 e^{b_2 t} \tag{2}$$

where a_1, a_2, a_3, b_1, b_2 and b_3 are parameters and t is usually time after administration of the drug. The parameters $a_i, b_i,\ i = 1, \ldots, 3$ are the macroparameters of the model.

Sometimes, where appropriate, a constraint is imposed on the parameters a_i to reduce the dimension of the problem. For instance, if the concentration is measured for samples from a pool peripheral, the sum of the $a_i' s$ is constraint to zero. At other times, constraints on the macroparameters arise naturally. A compartmental model was used in Alvarez et al. (2003) to describe Escherichia coli inactivataoin by pulsed electric fields. The biological meaning of this model is that in one population of microorganisms, two subpopulations exist. The first population is sensitive to the inactivating factor and the second population is resistant. If we let $S(t)$ be the fraction of total survivors at treatment time t, p be the fraction of survivors in the sensitive population, $1 - p$ be the fraction in the resistant population, and k_1 and k_2 are the specific death rates in the two populations, we have $S(t) = pe^{-k_1 t} + (1 - p)e^{-k_2 t}$. In this case, the parameters a_1 and a_2 must sum to unity.

The question of interest here is how to construct efficient designs for estimating parameters in the compartmental models. Because these are nonlinear models, the optimal design will depend on the nominal values of the parameters. As such they are called locally optimal designs. A popular design criterion for estimating model parameters is D-optimality. The criterion is expressed as a logarithmic function of the determinant of the expected Fishers' information matrix and hence it is a concave function (Silvey, 1980). For fixed nominal values, the locally D-optimal design is obtained by maximizing this function over the set of all designs on the interval of interest. Such an optimal design minimizes the generalized variance and consequently, locally D-optimal design provides the smallest confidence ellipsoid for the parameters. Frequently, an equivalence theorem is used to check the optimality of the design. Equivalence theorems are derived from convex analysis and are basically conditions required of the directional derivative of a concave functional at its optimum point. Details can be found in standard design monographs, see Fedorov (1972) or Silvey (1980) for example.

Recently Ermakov and Melas (1995) studied properties of locally D-optimal designs for an extension of model (1) and (2) within the class of all minimally supported designs. This means that instead of optimizing the criterion over all designs on the design space, the optimization is now restricted to only the class of designs where the number of design points is equal to the number of parameters in the model. They called these designs saturated optimal designs and they showed that a saturated locally D-optimal design is always unique and has equal weights at its support points. Moreover, the support points are decreasing functions of any of the parameters in the exponentials terms. However, the question if these saturated optimal designs are optimal within the class of all designs was left open.

In the present paper we give a partial answer to this problem. In the general model considered by Ermakov and Melas (1995) we show that in certain regions for the unknown parameters the locally D-optimal designs are in fact supported at a minimal number of points. Moreover, we derive an upper bound for the number of support points of the locally D-optimal design. For the special cases of model (1) and (2) this bound reduces to 4 and 7 , respectively, and we demonstrate by an extensive numerical study that locally D-optimal designs for the model (2) are in fact always supported at 6 points. Thus our theoretical and numerical results give a complete solution of the locally D-optimal design problem in the models (1) and (2).

The paper is organized as follows. Section 2 introduces the statistical setup, model specification and notation. The main theoretical and numerical results are contained in Section 3. Section 4 contains a summary and a description of some outstanding design problems for these types of models. All technical details are deferred to the Appendix.

2 Preliminaries

We assume that at the onset a predetermined number of observations N are to be taken from the study. The choice of N is usually determined by the resources available. Following Kiefer (1974), we view all designs in this paper as probability measures on a user-selected design interval χ. We denote a generic design with n distinct points by

$$\xi = \begin{pmatrix} x_1 & \dots & x_n \\ \mu_1 & \dots & \mu_n \end{pmatrix}.$$

Here, $x_1, \dots, x_n \in \chi$ are the design points, where observations are to be taken and $\mu_1, \dots, \mu_n$ denote the proportions of total observations taken at these points. In practice, a rounding procedure is applied to obtain the samples sizes $N_i \approx \mu_i N$ at the experimental conditions x_i, $i = 1, 2, \dots, n$, subject to $N_1 + N_2, \dots, +N_n = N$. An optimal procedure for rounding is given in Pukelsheim and Rieder (1993).

Consider the standard nonlinear regression model given by

$$y_j = \eta(x_j, \theta) + \varepsilon_j, \quad j = 1, 2, \dots, N,$$

where $\varepsilon_1, \dots, \varepsilon_N$ are independent identically distributed observations such that $E[\varepsilon_j] = 0$,

$E[\varepsilon_j^2] = \sigma^2 > 0$, $(j = 1, \ldots, N)$and

$$\eta(x, a, \lambda) = \sum_{i=1}^{k} a_i e^{-\lambda_i x}. \tag{3}$$

Here $a = (a_1, \ldots, a_k)^T, \lambda = (\lambda_1, \ldots, \lambda_k)^T$ and $\theta^T = (a^T, \lambda^T)$ is the vector of unknown parameters to be estimated. Without loss of generality we assume $a_i \neq 0$, $i = 1, \ldots, k$ and $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_k$. The designpoints $x_1, \ldots, x_N$ are experimental conditions, which can be choosen by the experimenter from a given set. In our case, this set is $\chi = [0, \infty)$. If $n \geq 2k$ and $\mu_i > 0$, $i = 1, \ldots, n$, it is well known that the least squares estimator $\hat{\theta}$ for the parameter θ in the model (3) is asymptotically unbiased with covariance matrix satisfying

$$\lim_{N \to \infty} \mathrm{Cov}(\sqrt{N}\hat{\theta}) = \sigma^2 M^{-1}(\xi, a, \lambda),$$

where

$$M(\xi, a, \lambda) \;=\; \left(\sum_{s=1}^{n} \frac{\partial \eta(x_s, \theta)}{\partial \theta_i} \frac{\partial \eta(x_s, \theta)}{\partial \theta_j} \mu_s \right)_{i,j=1}^{2k}$$

denotes the information matrix of the design ξ.

An optimal design maximizes a concave real valued function of the information matrix and there are numerous optimality criteria proposed in the literature to discriminate between competing designs, see for example, Silvey (1980) or Pukelsheim (1993). In this paper we restrict ourselves to the well known D-optimality criteion. Following Chernoff (1953), we call a design ξ locally D-optimal in the exponential regression model (3) if for given nominal values of a and λ, it maximizes $\det M(\xi, a, \lambda)$ over all designs on the interval χ. Locally D-optimal designs in various non-linear regression models have been discussed by numerous authors. They include Melas (1978), He, Studden and Sun (1996) or Dette, Haines and Imhof (1999) among many others. In the present context we have

$$\frac{\partial \eta(x_s, \theta)}{\partial \theta} \;=\; (e^{-\lambda_1 x}, \ldots, e^{-\lambda_i x}, -a_1 x e^{-\lambda_1 x}, \ldots, -a_k x e^{-\lambda_k x})^T$$

and as a consequence, it is easy to see that for any design ξ on χ, the determinant of the information matrix $M(\xi, a, \lambda)$ for the regression model (3) satisfies

$$\det M(\xi, a, \lambda) = a_1^2, \ldots, a_k^2 \det M(\xi, e, \lambda),$$

where $e = (1, \ldots, 1)^T \in \mathbb{R}^k$. In other words a locally D-optimal design for the model (3) does not depend on the "linear" parameters $a_1, \ldots, a_k$, and we can restrict ourselves to the maximization of the determinant of the matrix

$$M(\xi, \lambda) = M(\xi, e, \lambda) \tag{4}$$

throughout this paper. Because for any design space under consideration the induced design space

$$\left\{ \frac{\partial \eta(x,\theta)}{\partial \theta} \;\Big|\; x \in \chi \right\}$$

is compact, locally D-optimal designs exist (Pukelsheim, 1993). Moreover a locally D-optimal design has necessarily at least $n \geq 2k$ support points because otherwise the corresponding information matrix would be singular. Throughout this paper designs with a minimal number of support points $n = 2k$ are called *saturated* or *minimally supported* designs. It is well known that a D-optimal saturated design has equal masses, that is $\mu_1 = \mu_2 = \ldots = \mu_{2k} = \frac{1}{2k}$ (Fedorov, 1972).

Melas and Ermakov (1995) studied locally D-optimal designs for the model (3) in the class of all saturated designs. However, the question if these *saturated* optimal designs are optimal within the class of all designs was left open. In the following section we will derive an upper bound for the number of support points of the locally D-optimal design in the general exponential regression model (3). For the particular cases $k = 1, 2$ and $k = 3$ [the last two cases correspond to the model (1) and (2), respectively] we show that the locally D-optimal designs are in fact saturated. These results can be used to find the locally D-optimal designs using numerical methods introduced by Melas (1978) and recently in Melas (2000) and Dette Melas and Pepelyshev (2004) for polynomial models.

3 Main results

In this section we study the number of support points of a locally D-optimal design for the model (3). Throughout this paper this number will be denoted by $n^*(\lambda)$. Additionally we will use for $k \geq 3$ the notation

$$\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_k)^T \tag{5}$$

for any vector λ with components satisfying

$$\begin{aligned} 0 \; &< \; \hat{\lambda}_1 < \ldots < \hat{\lambda}_k \\ \hat{\lambda}_{i+1} \; &= \; (\hat{\lambda}_i + \hat{\lambda}_{i+2})/2, \quad i = 1, 2, \ldots, k-2. \end{aligned} \tag{6}$$

Theorem 1 *If $n^*(\lambda)$ denotes the number of support points of a locally D-optimal design (with respect to the parameter λ) for the nonlinear regression model (3), then*

(i) $n^*(\lambda) = 2k$, *if* $k = 1$ *or* 2

(ii)

$$n^*(\lambda) \le k(k+1)/2 + 1,$$

for any $k \ge 3$. Moreover for any vector $\hat{\lambda}$ of parameters with components satisying (6) there exists a neighbourhood, say $\mathcal{U} \subset \mathbb{R}^k$, of $\hat{\lambda}$, such that for all vectors $\lambda \in \mathcal{U}$ the number of support points of the locally D-optimal design (with respect to λ) is given by $n^*(\lambda) = 2k$.

Note that in the case $k = 1, 2$ the locally D-optimal $2k$-point design is in fact also optimal in the class of all designs. If $k \ge 3$ the last part of Theorem 1 indicates that in many cases locally D-optimal designs for the regression model (3) are in fact saturated designs. Formally this is only true for vectors λ in a neigbourhood of a parameter vector $\hat{\lambda}$ with components satisfying the restriction (6). However, numerical results indicate that the set of parameter vectors $\lambda \in \mathbb{R}^k$ for which the locally D-optimal design is minimally supported is usually very large. For example in the case $k = 3$ we could not find any case, where the locally D-optimal design was suppported at 7 points. Note that this is the upper bound for the number of support points according to the second part of Theorem 1.

Corrollary 2. *In the exponential regression model (3) with $k = 3$ we have for the number $n^*(\lambda)$ of support points of the locally D-optimal design (with respect to λ)*

(i) $n^*(\lambda) \in \{6, 7\}$ *for any vector λ with increasing positive coefficients*

(ii) For any point $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$ satisfying $0 < \hat{\lambda}_1 < \hat{\lambda}_2 < \hat{\lambda}_3$, $\hat{\lambda}_2 = (\hat{\lambda}_1 + \hat{\lambda}_3)/2$ there exists a neighbourhood of $\hat{\lambda}$, say $\mathcal{U} \subset \mathbb{R}^k$, such that $n^*(\lambda) = 6$ for any locally D-optimal design with respect to $\lambda \in \mathcal{U}$.

The results of Theorem 1 and Corollary 2 can be used to construct numerical locally D-optimal designs for the exponential regression model (3). We illustrate this procedure by determining locally D-optimal designs for $k = 1, 2, 3$, corresponding to the case where there is one, two or three exponential terms in the model. Before we begin we recall some results on the restricted optimization in the class of all saturated designs with support points $x_1^* < x_2^* < \ldots < x_{2k}^*$, for which the optimal weights are equal to $\frac{1}{2k}$. Melas (1978) proved the following properties for the locally D-optimal saturated designs for a homoscedastic model with an arbitrary sum of exponential terms. (i.e. $k \in \mathbb{N}$ is abitrary).

(i) The support points $x_1^* < x_2^* < \ldots < x_{2k}^*$ of a saturated locally D-optimal design for the model (3) are uniquely determined.

(ii) $0 = x_1^* < x_2^* < \ldots < x_{2k}^*$ are analytic functions of the nonlinear parameters $\lambda_1, \ldots, \lambda_k$. Therefore we use the notation $x_i^*(\lambda)$ $(i = 1, \ldots, 2k)$. As a consequence each support point can be expanded in Taylor series in a neighborhood of any point λ.

(iii) If the nonlinear parameters $\lambda_1, \ldots, \lambda_k$ satisfy $\lambda_i \to \lambda^* > 0$, $i = 1, 2, \ldots, k$, then the support points of the locally D-optimal design with respect to the parameter $\lambda = (\lambda_1, \ldots, \lambda_k)^T$ converge, that is

$$\lim_{\lambda \to \lambda^*} x_i^*(\lambda) = \gamma_{i-1}/2\lambda^*,$$

where $\gamma_1, \ldots, \gamma_{2k-1}$ are the roots of Laguerre polynomial $L_{2k-1}^{(1)}(x)$ of degree $2k-1$ orthogonal with respect to the measure $x \exp(-x)dx$ [see Szegö (1975)].

In the case $k = 1$ it follows from Theorem 1(i) that the locally D-optimal design is a uniform distribution on two points and we obtain from (iii)

$$x_1^* \equiv 0, \quad x_2^* = 1/\lambda_1.$$

In the case $k = 2$ Melas (1978) determined locally D-optimal saturated designs restricting the optimization to the class of all four point designs. Theorem 1(i) now shows that these saturated designs are in fact locally D-optimal within the class of all designs. A table of these designs can be found in Melas (1978). In the case $k = 3$ we obtain $n^*(\lambda) = 7$ as upper bound for the number of support points of any locally D-optimal designs.

We now consider locally D-optimal designs for the model

$$a_1 e^{-\lambda_1 x} + a_2 e^{-\lambda_2 x} + a_3 e^{-\lambda_3 x}, \tag{7}$$

where $a_1, a_2, a_3 \neq 0$, $0 < \lambda_1 < \lambda_2 < \lambda_3$ and the design space is given by the interval $[0, \infty)$. If the interval $[c, \infty)$ with $c > 0$ is the design space, we need only add the constant c to all design points. Since $x_1^* \equiv 0$ only the points $x_2^*, \ldots, x_6^*$ have to be calculated. Note that under a multiplication of all parameters $\lambda_1, \lambda_2, \lambda_3$ by the same positive constant the support points of the locally D-optimal design have to be divided by the same constant. Therefore without loss of generality we assume the normalization $(\lambda_1 + \lambda_2 + \lambda_3)/3 = 1$ and introduce the notation $\delta_1 = 1 - \lambda_1$, $\delta_2 = 1 - \lambda_2$ (note that the condition $\lambda_1 < \lambda_2$ implies $\delta_1 > \delta_2$). The point $\hat{\lambda}(\varepsilon) = (1, 1 + \varepsilon, 1 + 2\varepsilon)$ with $\varepsilon > 0$ is obviously of the form (5) and arbitrarily close to the point $\lambda^* = (1, 1, 1)$. Consequently, by Theorem 1 the support points $x_2(\lambda), \ldots, x_{2k}^*(\lambda)$ of the locally D-optimal design can be expanded in a convergent Taylor series at the point λ^* (which corresponds to the case $\delta_1^* = 0$, $\delta_2^* = 0$). The coefficients in this expansion can be determined recursively [see Melas (2000) or Dette, Melas and Pepelyshev (2004)] With the help of the equivalence theorems, they also verified that for the D-opimality criterion, these designs are locally D-optimal in the class of all approximate designs. In all examples considered in our study we obtain $n^*(\lambda) = 6$. Table 1 and Table 2 show the suport points of the locally D-optimald design for various values of δ_1 and δ_2.

$\delta_1 = 0.95$											
δ_2	-0,3	-0,2	-0,1	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7
x_2	0,312	0,312	0,313	0,314	0,315	0,316	0,317	0,318	0,320	0,322	0,324
x_3	1,101	1,106	1,113	1,123	1,135	1,151	1,172	1,198	1,231	1,275	1,336
x_4	2,572	2,597	2,636	2,691	2,764	2,862	2,993	3,170	3,419	3,788	4,381
x_5	6,148	6,251	6,405	6,619	6,906	7,284	7,781	8,444	9,353	10,657	12.674
x_6	26,44	926,573	26,761	27,024	27,378	27,852	28,488	29,359	30,598	32,474	35,587
$\delta_1 = 0.9$											
δ_2	-0,3	-0,2	-0,1	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7
x_2	0,312	0,312	0,312	0,313	0,314	0,315	0,316	0,318	0,319	0,321	0,323
x_3	1,096	1,100	1,106	1,115	1,126	1,145	1,159	1,182	1,212	1,250	1,302
x_4	2,535	2,557	2,590	2,637	2,700	2,784	2,895	3,044	3,248	3,542	3,992
x_5	5,550	5,623	5,737	5,898	6,116	6,405	6,786	7,293	7,985	8,966	10,447
x_6	6,069	16,175	16,341	16,58	16,906	17,348	17,947	18,77	19,931	21,644	24,347
$\delta_1 = 0.8$											
δ_2	-0,3	-0,2	-0,1	0	0,1	0,2	0,3	0,4	0,5	0,6	
x_2	0,311	0,311	0,312	0,312	0,313	0,314	0,315	0,316	0,317	0,319	
x_3	1,087	1,089	1,094	1,101	1,111	1,123	1,138	1,157	1,180	1,211	
x_4	2,475	2,489	2,513	2,548	2,596	2,661	2,744	2,855	3,003	3,205	
x_5	5,000	5,047	5,122	5,234	5,386	5,591	5,860	6,218	6,699	7,367	
x_6	10,898	10,979	11,12	11,328	11,615	12,001	12,516	13,206	14,151	15,495	
$\delta_1 = 0.7$											
δ_2	-0,3	-0,2	-0,1	0	0,1	0,2	0,3	0,4	0,5	0,6	
x_2	0,310	0,311	0,311	0,311	0,312	0,313	0,314	0,315	0,316	0,317	
x_3	1,079	1,081	1,085	1,090	1,098	1,108	1,120	1,136	1,155	1,180	
x_4	2,428	2,437	2,454	2,481	2,518	2,568	2,634	2,720	2,832	2,983	
x_5	4,720	4,746	4,798	4,880	4,995	5,151	5,358	5,631	5,997	6,499	
x_6	9,208	9,261	9,368	9,536	9,773	10,095	10,526	11,101	11,884	12,991	

Table 1: Support points of the locally D-optimal designs for the exponential regression model (7) for various values of $\delta_1 = 1 - \lambda_1$, $\delta_2 = 1 - \lambda_2$.

$\delta_1 = 0.6$							
δ_2	-0,2	-0,1	0	0,1	0,2	0,3	0,4
x_2	0,310	0,310	0,311	0,311	0,312	0,313	0,314
x_3	1,074	1,076	1,081	1,087	1,095	1,106	1,119
x_4	2,396	2,408	2,428	2,458	2,498	2,551	2,619
x_5	4,554	4,589	4,650	4,740	4,863	5,028	5,246
x_6	8,396	8,473	8,605	8,800	9,070	9,435	9,927
$\delta_1 = 0.5$							
δ_2	-0,2	-0,1	0	0,1	0,2	0,3	0,4
x_2	0,310	0,310	0,310	0,310	0,311	0,312	0,312
x_3	1,068	1,070	1,073	1,078	1,085	1,094	1,105
x_4	2,365	2,373	2,388	2,411	2,443	2,486	2,542
x_5	4,423	4,445	4,490	4,560	4,659	4,794	4,973
x_6	7,884	7,934	8,035	8,195	8,424	8,739	9,167
$\delta_1 = 0.4$							
δ_2		-0,1	0	0,1	0,2	0,3	
x_2		0,309	0,309	0,310	0,310	0,311	
x_3		1,065	1,067	1,071	1,076	1,084	
x_4		2,346	2,356	2,374	2,400	2,436	
x_5		4,344	4,375	4,430	4,511	4,622	
x_6		7,586	7,661	7,792	7,986	8,261	
$\delta_1 = 0.3$							
δ_2		-0,1	0	0,1	0,2		
x_2		0,309	0,309	0,309	0,310		
x_3		1,061	1,062	1,065	1,070		
x_4		2,327	2,334	2,347	2,368		
x_5		4,274	4,295	4,336	4,402		
x_6		7,361	7,411	7,515	7,681		

Table 2: Support points of the locally D-optimal designs for the exponential regression model (7) for various values of $\delta_1 = 1 - \lambda_1$, $\delta_2 = 1 - \lambda_2$.

4 Discussion

Our work in the previous sections extended the results in Melas (1978). He considered compartmental models with one and two exponential terms and found locally D-optimal designs within the class of designs with minimal support points. This means that optimality of the design was restricted to the class of designs with two points when there is one exponential term and to four points when the model has two exponential terms. We showed here that these saturated locally D-optimal designs are actually locally D-optimal, meaning that they are also optimal within the class of all designs. Additionally, we extended this result to models with three exponential terms, and found that this result also applies to models with an arbitrary number of exponential terms provided the parameters in the exponential terms belong to a certain region.

We stress that the locally D-optimal designs for the exponential model (1) and (2) are influenced by a preliminary "guess" for the parameter values. This may seem undesirable but such designs usually represent a first step in the construction of an optimal design for a model under a more robust optimality criterion, including the Bayesian- and minimax criterion, see Pronzato and Walter (1985), Chaloner and Larntz (1989) or Haines (1995) among many others. Our results here suggest that findng optimal designs for the general exponential regression model under a more sophisticated optimality criterion will be a challenging problem.

There is room for further work for addressing design issues for the types of models considered in this paper. We mention three interesting areas for further research here.
First, we have considered models only up to three exponential models. Models with more than three exponential terms are also used in practice, although less often because of the added complexity. For instance, a seven-compartment physiologically based pharmacokinetic model was developed to predict biological levels of tetrahydrofuran under different exposure scenarios (Droz, Berode and Jang, 1999). Constructing and understanding properties of optimal designs for more complicated model will be helpful for the practitioners. However, we anticipate extending similar results for models with four or more exponential terms will require more theory, and more likely require a different approach.

Second, there are other biological models closely related to those studied here. For instance, if we add an intercept to the models studied here, the resulting models are useful for studying viral dynamics and related problems. Han and Chaloner (2003) constructed D and c-optimal designs for some simple models for estimating parameters in viral dynamics in an AIDS trial. They considered models with one or two exponential terms, but more complex systems will have to involve additional exponential terms. Providing optimal designs will provide guidance to the researcher where the best sampling time points to realize cost reduction without sacrificing statistical efficiency.

Third, our models assume that errors are homoscedastic. Landaw and DiStefano (1984) postulated that the error variance in some compartmental models is more appropriated modeled as equal to $\alpha + \beta(y(t_i))^\gamma$, where α represents constant background variance. The three parameters α, β and γ may be known constants from previous studies, and if they are not known, they will have to be estimated, before a locally optimal design can be constructed.

5 Appendix: Proof of Theorem 1

We begin the proof with a study of the correponding D-optimal design problem in the linear regression model

$$(8) \qquad \sum_{i=1}^{2k} \beta_i e^{-\tilde{\lambda}_i x},$$

where $0 < \tilde{\lambda}_1 < \ldots < \tilde{\lambda}_{2k}$ are fixed known values and $\beta_1, \ldots, \beta_{2k}$ are the unknown parameters to be estimated. It is easy to see that for a design with masses $\mu_1, \ldots, \mu_n$ at the points $x_1, \ldots, x_n$ $(n \geq 2k)$ the information matrix in this model is of the form

$$(9) \qquad A(\xi, \tilde{\lambda}) = \left(\sum_{s=1}^{n} e^{-\tilde{\lambda}_i x_s} e^{-\tilde{\lambda}_j x_s} \mu_s \right)_{i,j=1}^{2k}.$$

In the following we investigate the maximum of $\det A(\xi, \tilde{\lambda})$, where the components of the vector $\tilde{\lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{2k})^T$ are defined by

$$(10) \qquad \tilde{\lambda}_{2i-1} = \lambda_i, \tilde{\lambda}_{2i} = \lambda_i + \Delta, 0 < \Delta < \min_{i=1,\ldots,k-1}(\lambda_{i+1} - \lambda_i),\ i = 1, \ldots, k,$$

where $0 < \lambda_1 < \ldots < \lambda_k$ (in the case $k = 1$ the value $\Delta > 0$ can be chosen arbitrarily). In the following let

$$\xi^* = \operatorname{argmax} \det A(\xi, \tilde{\lambda})$$

denote a design maximizing the determinant, where maximum is taken over the set of all approximate designs on χ. Note that designs maximizing this determinant exist, because the induced design space

$$\left\{ (e^{-\tilde{\lambda}_1 x}, \ldots, e^{-\tilde{\lambda}_{2k} x})^T \mid x \in \chi \right\}$$

is compact (Pukelsheim, 1993). By the well known Kiefer-Wolfowitz equivalence theorem we have

$$\max_{x \in \chi} f^T(x) A^{-1}(\xi^*, \tilde{\lambda}) f(x) = 2k,$$

where $f^T(x) = (e^{-\tilde{\lambda}_1 x}, \ldots, e^{-\tilde{\lambda}_{2k} x})$ denotes the vector of regression functions in the model (8). It follows from Gantmacher (1959), Ch. XIII that any minor of the matrix $(e^{-\tilde{\lambda}_i x_j})_{i,j=1}^{2k}$ with $x_1 > x_2 > \ldots > x_{2k}$, $\tilde{\lambda}_1 < \tilde{\lambda}_2 < \ldots < \tilde{\lambda}_{2k}$ is positive. Therefore the Cauchy-Binet formula implies that

$$(11) \qquad \operatorname{sign}(A^{-1})_{ij} = (-1)^{i+j}.$$

where we use the notation $A = A(\xi^*, \tilde{\lambda})$ for the sake of brevity. In the following discussion we need an auxilary result.
Its proof is deferred to end of this section.

Lemma A.1. *Consider the functions*

$$\varphi_i(x) = \sum_{j=1}^{t_i} \alpha_{i,j} e^{-\mu_{i,j} x},$$

where t_i are arbitrary integers, $i = 0, \dots, s$, $\{\alpha_{ij}, \mu_{i,j}\}$ are real numbers.
Let the following conditions hold

(i) $\min_{1 \le j \le t_{i+1}} \mu_{i+i,j} > \max_{1 \le i \le t_i} \mu_{i,j}$, $i = 0, 1, \dots, s-1$*;*

(ii) $\alpha_{i,j} < 0$, $j = 1, \dots, t_i$, $i = 0, \dots, s$.

Then the function $\sum_{i=0}^{s} b_i \varphi_i(x)$, where $b_0, \dots, b_s$ are arbitrary real number, has at most s roots counted with their multiplicity.

Let us define

$$\begin{aligned}
\varphi_0(x) &\equiv m, \\
\varphi_{l-1}(x) &= (-1)^l \sum_{i=1}^{l-1} A_{l-i,i} e^{-(\tilde\lambda_i + \tilde\lambda_{l-i})x}, \ l = 2, \dots, 2k, \\
\varphi_{l-1}(x) &= (-1)^l \sum_{j=1}^{4k-l+1} A_{2k+1-j, l-2k+j-1} e^{-(\tilde\lambda_{2k+1-j} + \tilde\lambda_{l-2k+j-1})x}, \ l = 2k+1, \dots, 4k,
\end{aligned}$$

where $A_{i,j} = (A^{-1})_{ij}$.
We consider at first the cases $k = 1, 2$. Note that the coefficients in the functions are positive since sign $A_{i,j} = (-1)^{i+j}$, that is condition (ii) holds. Moreover, observing the definition of $\tilde\lambda$ in (10) condition (i) can also easily be verified for $k = 1, 2$.
Now we have

$$g(x) \ := \ m - f^T(x) A^{-1}(\xi^*, \tilde\lambda) f(x) \ = \ \varphi_0(x) + \sum_{i=1}^{4k-1} (-1)^i \varphi_i(x)$$

and from the equivalence theorem for the D-optimality criterion it follows that $g(x) \le 0$ for all x. This implies for the support points, say $x_1^*, \dots, x_n^*$ of a design ξ^* maximizing det $A(\xi, \tilde\lambda)$

$$\begin{gathered}
g(x_i^*) = 0, \ i = 1, 2, \dots, n \\
g'(x_i^*) = 0, \ i = 2, 3, \dots, n-1.
\end{gathered}$$

A careful counting of the multiplicities and an application of Lemma 1 now show $2n - 2 \le 4k - 1$, which implies $n = 2k$ in the case $k = 1$ or 2.
In the case $k \ge 3$ the same arguments are applicable for any vector $\hat\lambda$ satisfying (6), because in this case it can be easily verified that the functions φ_i, $i = 0, \dots, 4k$ defined above satisfy both conditions of Lemma A.1. An argument of continuity therefore shows $n^*(\lambda) = 2k$ for

the number of supports of a D-optimal design for the model (8) with respect to any λ in a neighbourhood of the point $\hat{\lambda}$.

For a proof of the second bound in the case $k \geq 3$ we consider an arbitrary point of the form (10), say $\tilde{\lambda} = (\tilde{\lambda}_1, \dots, \tilde{\lambda}_{2k})$, and define $s \leq k(k+1)/2$ as the number of distinct values in the set

$$\{2\lambda_1, \dots, 2\lambda_k, \lambda_1 + \lambda_2, \dots, \lambda_1 + \lambda_k, \lambda_2 + \lambda_3, \dots, \lambda_{k-1} + \lambda_k\}.$$

We denote with $u_1 < \dots < u_s$ the distinct values from this set and introduce the functions

$$\begin{aligned}
\tilde{\varphi}_0(x) &\equiv m, \\
\tilde{\varphi}_1(x) &= A_{11} e^{-u_1 x} = A_{11} e^{-2\lambda_1 x}, \\
\tilde{\varphi}_2(x) &= -2A_{12} e^{-(u_1+\Delta)x}, \\
\tilde{\varphi}_{2l-1}(x) &= a_l e^{-(u_l+2\Delta)x} + c_l e^{-u_{l+1}x}, \quad l = 2, \dots, s \\
\tilde{\varphi}_{2l}(x) &= -b_l e^{-(u_l+\Delta)x}, \quad l = 2, \dots, s \\
\tilde{\varphi}_{2s+1}(x) &= a_{s+1} e^{-(u_s+2\Delta)x}.
\end{aligned}$$

Observing that $\mathrm{sign} A_{ij} = (-1)^{i+j}$ it can be easily checked that the coefficients a_l, b_l, c_l can be chosen such that the representation

$$f^T(x) A^{-1}(\xi^*, \tilde{\lambda}) f(x) = \sum_{i=1}^{2s+1} \tilde{\varphi}_i(x). \tag{12}$$

is satisfied and such that

$$a_l, b_l, c_l > 0, \quad l = 1, \dots, s.$$

By the same arguments as in the previous paragraph we obtain for the determinant

$$\tilde{J}(\tau) = \det\left(\tilde{\varphi}_i(x_j)\right)_{i,j=0}^{2s+1} > 0$$

whenever $x_0 > x_1 > \dots > x_{2s+1}$. Moreover for any vector $\bar{\tau} = (\bar{x}_0, \dots, \bar{x}_{2s+1})^T$ with components satisfying $\bar{x}_0 \geq \bar{x}_1 \geq \dots \geq \tilde{x}_{2s+1}$ it follows

$$\lim_{\tau \to \bar{\tau}} \tilde{J}(\tau) / \prod_{j>i} (x_i - x_j) > 0.$$

From the repesentation (12) and the equivalence theorem for the D-optimality criterion we obtain the inequality $g(x) \leq 0$ for any $x \geq 0$, where g is the generalized polynomial

$$g(x) = \tilde{\varphi}_0(x) - \sum_{i=1}^{2s+1} \tilde{\varphi}_i(x).$$

This implies that the support points of the locally D-optimal design satisfy

$$\begin{aligned}
g(x_i^*) &= 0, \quad i = 1, 2, \dots, n, \\
g'(x_i^*) &= 0, \quad i = 2, 3, \dots, n-1.
\end{aligned}$$

Moreover, by the arguments from the previous paragraph the function g has at most $2s+1$ roots counted with corresponding multiplicity. Consequently,

$$2n-2 \le 2s+1 \le k(k+1)+1.$$

which yields

$$n \le \frac{k(k+1)}{2} + 1 + 1/2,$$

and proves the assertion of the theorem for the regression model of the form (8).

To prove the assertion of the theorem for the exponential regression model (3) we consider for an arbitrary approximate design ξ the polynomial

$$\begin{aligned} q(x) &= m - f^T(x)A^{-1}(\xi,\tilde{\lambda})f(x) = \\ &= m - f^T(x)L^T(LA(\xi,\tilde{\lambda})L^T)^{-1}Lf(x), \end{aligned} \tag{13}$$

where $\tilde{\lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{2k})$ is defined by (10), the $2k \times 2k$ matrix L is given by

$$\begin{pmatrix} Q & 0 & 0 & \ldots & 0 \\ 0 & Q & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & \ldots & Q \end{pmatrix},$$

with

$$Q = \begin{pmatrix} 1 & 0 \\ 1/\Delta & -1/\Delta \end{pmatrix}.$$

Note that $\det L = (-1/\Delta)^k \neq 0$ and that

$$\begin{aligned} \lim_{\Delta \to 0} f^T(x)L^T &= \lim_{\Delta\to 0} \left(e^{-\lambda_1 x}, \frac{e^{-(\lambda_1+\Delta)x} - e^{-\lambda_1 x}}{\Delta}, \ldots, e^{-\lambda_k x}, \frac{e^{-(\lambda_k+\Delta)x} - e^{-\lambda_k x}}{\Delta} \right) \\ &= \left(e^{-\lambda_1 x}, -xe^{-\lambda_1 x}, \ldots, e^{-\lambda_k x}, -xe^{-\lambda_k x} \right). \end{aligned}$$

Consequently we have for any design ξ

$$\lim_{\Delta\to 0} LA(\xi,\tilde{\lambda})L^T = M(\xi,\lambda), \tag{14}$$

where $M(\xi,\lambda)$ is the information matrix in the exponential regression model (3) defined in Section 1.

If ξ^* denotes a locally D-optimal design for the regression model (3) with support points by $x_1^* < \ldots < x_{n^*}^*$, then it follows from (13) and (14) that

$$m - \tilde{f}^T(x)\tilde{M}^{-1}(\xi^*,\lambda)\tilde{f}(x) = \lim_{\Delta\to 0} m - f^T(x)A^{-1}(\xi^*,\tilde{\lambda})f(x), \tag{15}$$

where the vector $\tilde{f}^T(x)$ corresponds to the gradient in model (3) and is defined by

$$\tilde{f}^T(x) = \left(e^{-\lambda_1 x}, -xe^{-\lambda_1 x}, \ldots, e^{-\lambda_k x}, -xe^{-\lambda_k x}\right).$$

By the equivalence theorem for the D-optimality criterion the polynomial on the left hand side has roots $x_1^*, \ldots, x_{n^*}^*$, where $x_2^*, \ldots, x_{n^*-1}^*$ are roots of multiplicity two. Consequently, we obtain $2n^* - 1 \leq h$, where h is the number of roots of the polynomial on the right hand side of (15). By the arguments of the first part of the proof we have $h \leq 4k - 1$ for $k = 1, 2$ and for $k \geq 3$ in a neighborhood of points λ satsifying (6). Moreover, we have $h \leq k(k+1)/2$ in general, which completes the proof of the theorem. ∎

Proof of Lemma A.1. Denote $\tau^T = (x_0, \ldots, x_s)^T$,

$$J(\tau) = \det\left(\varphi_i(x_j)\right)_{i,j=0}^{s}.$$

Using the expansion of the determinant by a line several times we receive

$$J(\tau) = \sum_{l_0=1}^{t_0} \cdots \sum_{l_s=1}^{t_s} \left[\left(\prod_{i=0}^{s} \alpha_{i,l_i}\right) \det\left(e^{-\mu_{jl_j}x_\nu}\right)_{j,\nu=0}^{s}\right].$$

Due to the Chebyshev property of exponential functions (see Karlin, Studden, 1966, Ch. 1) each term on the right hand side is positive whenever $x_0 > x_1 > \ldots > x_s$ [note that $\prod_{i=0}^{s} \alpha_{i,l_i} > 0$ by assumption (ii) and that $\det\left(e^{-\mu_{jl_j}x_\nu}\right)_{j,\nu=0}^{s} > 0$ by assumption (i) of Lemma A.1]. Thus $J(\tau) > 0$ for arbitrary $x_0 > x_1 > \ldots > x_s$. Moreover, we have for any $\bar{\tau} = (\bar{x}_0, \ldots, \bar{x}_t)$ with $\bar{x}_0 \geq \bar{x}_1 \geq \ldots \geq \bar{x}_t$

$$\lim_{\tau \to \bar{\tau}} J(\tau) / \prod_{i<j} (x_i - x_j) > 0, \tag{16}$$

since

$$\lim_{\tau \to \bar{\tau}} \det\left(e^{-\theta_{i_s} x_j}\right)_{s,j=0}^{t} / \prod_{i<j} (x_i - x_j) > 0$$

whenever $0 < \theta_{i_0} < \ldots < \theta_{i_s}$. This property can easily be verified considering the number of the same coordinates in the vector $\bar{x}$. It is known (see Karlin, Studden, 1966, Ch. 1) that under conditions $J > 0$ and (16) any generalized polynomial of the form $\sum_{i=0}^{t} b_i \varphi_i(t)$ has at most t roots counted with their multiplicity. ∎

Acknowledgements: The support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt A2, De 502/18-1) is gratefully acknowledged. The authors are also grateful to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise.

References

DArgenio, D. Z. (1981). Opitmal sampling times for pharmacokinetic experiments. J of Pharmacokin. Biopharm, 9, 739-755.

Allen, D. M. (1983). Parameter estimation for nonlinear models with emphasis oon compartmental models. Biometrics, 39, 629-637.

Alvarez, I., Virto, R., Raso, J. and Condon, S. (2003). Comparing predicting models for the Escherichia coli inactivation by pulsed electric fields. Innov. Food Sci. Emerg. Technol., 4, 195-202.

Bardsley, W. G., McGinlay, P. B and Wright, A. J. (1986). The F test for model discrimination with exponential functions. Biometrika, 73, 501-508.

Chaloner, K., Larntz, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. J. Statist. Plann. Inference 21, 191-208.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Statist. 24, 586-602.

Dette, H., Haines, L., Imhof, L.A. (1999). Optimal designs for rational models and weighted polynomial regression. Ann. Statist. 27, 1272-1293.

H. Dette, H., Melas, V. B. and Pepelysheff, A. (2004). Optimal designs for estimating individual coefficients in polynomial regression – a functional approach. J. Statist.Plann. Inf. 118, 201 - 219.

Droz, P. O., Berrode, M. and Yang, J. Y. (1999). Biological monitoring of Teteahydrofuran: Contribution of a physiologically based pharmacokinetic model. American Industrial Hygiene Association, 243-248.

Endrenyi, L. and Dingle, B. H. (1982). Optimal design of experiments for the precise estimation of single-exponential kinetic model parameters, in Bozler, G. ad van Rossum, J. M. (eds). Pharmacokinetics During Drug Development: Data Analysis and Evaluation Techniques (Gustav Fischer, Stuttgard, 149-173.

Ermakov S.M., Melas V.B. (1995). Design and Analysis of Simulation Experiments, Kluwer Academic publishers, Dordreht, London. 196 p.

Fedorov V.V. (1972). Theory of optimal experiments. Academic Press, New York.

Gantmacher F.R. (1959). The Theory of Matrices. V. II. New York: Chelsea.

Gibaldi, M., Perrier, D. (1982). Pharmacokinetics. Marcel Dekker, NY.

Green, J. R. and Reilly, P. E. B. (1975). Choosing an appropriate multiexponential. J. Theor. Biol., 55, 137-143.

Haines, L. M. (1995). A geometric approach to optimal design for one-parameter nonlinear models. J. R. Statist. Soc., Ser. B 57, 575-598.

Han, C. and Chaloner, K. (2003). D and c-optimal designs for exponential regression models used in viral dynamics and other applications. J. Statisti. Plann. Inference, 115, 2, 585-601.

He, Z., Studden, W.J. and Sun, D. (1996). Optimal designs for rational models. Ann. Statist. 24, 2128-2142.

Karlin, S. and Studden, W. (1966). *Tchebysheff systems: with application in analysis and statistics*, Wiley, N.Y.

Kiefer, J. C. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. 2, 849-879.

Landaw, E. M. and DiStefano, J. J. (1999). Multiexpontial, multicompartmental and non-compartmental modeling. II. Data analysis and statistical considerations. Am. J. Physiol., R665-R676,

Landaw, E. M. (1985). Robust sampling designs for compartmental models under large prior eigenvalue uncertainties. Mathematics and Computers in Biomedical Applications, 181-187.

Mason, H. and Wilson, K. (1999). Biological monitoring: The role of Toxicokinetics and physiologically based pharmacokinetic modeling, 60, American Industrial Hygiene Journal, 237-242.

Melas, V.B. (1978). Optimal designs for exponential regression. *Math. Operations forsh. Statist. Ser. Statistics*, 9, 45-59.

Melas, V.B. (2000). Analytic theory of E-optimal designs for polynomial regression. Advances in stochastic simulation methods (eds. N. Balakrishnan, V.B. Melas, S. Ermarkov). Birkhäuser, Boston, 85-116.

Pronzato, L. and Walter, E. (1985). Robust experimental design via stochastic approximation. Math. Biosci. 75, 103-120.

Pukelsheim, F. (1993). Optimal design of experiments. Wiley, New York.

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

Rowland, M. and Tozer, T. N. (1989). Clinical Pharmacokinetics: concepts and applications, 2nd Edition. Lea and Febiger, Philadelphia, Pennsylvania.

Shargel, L. and Yu, A. B. C. (1985). Applied Biopharmaceutics and Pharmacokinetics, 2nd Edition. Appleton-Century-Crofts. Norwalk, Conneticut. Silvey, S.D. (1980). Optimum design. Chapman and Hall, London.