LOCALLY *D*-OPTIMAL DESIGNS FOR EXPONENTIAL REGRESSION MODELS

Holger Dette, Viatcheslav B. Melas and Weng Kee Wong

Ruhr-Universität Bochum, St. Petersburg State University and University of California at Los Angeles

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.

Key words and phrases: Approximate designs, compartmental models, equivalence theorem, information matrix.

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) and Ratkowsky (1983, 1990). Our goal 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), Bardsley, McGinlay and Wright (1986) and Droz, Berode and Jang (1999). These models are particularly common in pharmacokinetics, and are called compartmental models (Shargel and Yu (1985)) and Jones and Wang (1999). Typically, the expected mean response in concentration units is expressed as a linear combination of exponential terms. Such a model is suitable for modeling an identifiable, open, non-cyclic, *n*-compartmental system with bolus input into the sampled pool (Landaw (1985)). Compartmental models are also used in data analysis in toxicokinetic experiments (Becka, Bolt and Urfer (1992)) and in chemical kinetics (Gibaldi and Perrier (1982)). The simplest forms of such models have their mean response equal to

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

$$a_1 e^{b_1 t} + a_2 e^{b_2 t} + a_3 e^{b_3 t}, \tag{2}$$

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.

Sometimes a constraint is imposed on the parameters a_i to reduce the dimension of the problem. For instance, if the response is the concentration of some drug measured for samples from a pool peripheral, the sum of the $a_i's$ is zero. At other times, constraints on the parameters arise naturally, as in Alvarez, Virto, Raso and Condon (2003), where a compartmental model was used to describe Escherichia coli inactivation by pulsed electric fields.

Our aim is to construct efficient designs for estimating parameters in compartmental models. A popular design criterion for estimating model parameters is D-optimality. The criterion is expressed as a logarithmic function of the expected Fisher's information matrix and it is a concave function (Silvey (1980)). For fixed nominal values of the parameters, the locally D-optimal design is obtained by maximizing the criterion function over the set of all designs on the design interval. 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 the 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 models (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 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. In addition, the support points are decreasing functions of any of the parameters in the exponentials terms. However, the question of whether these saturated optimal designs are optimal within the class of all designs was left open.

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 supported at a minimal number of points. We also 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 upper bound is 4 and 7, respectively. We carry out an extensive numerical study and confirm that locally D-optimal designs for model (2) are always supported at six points. Thus our theoretical and numerical results give a complete solution of the locally D-optimal design problem in models (1) and (2). Hitherto designs in pharmacokinetic studies have been largely based on past experience, and the

number of sampling times usually decided without statistical considerations. The practical implication of our results is that locally D-optimal designs for exponential regression models are optimal over all designs on the design interval, and the design remains optimal no matter how many sampling times are used for the study.

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 a predetermined number N of observations are to be taken from the study. The choice of N is determined by the resources available. Following Kiefer (1974), we view all designs as probability measures on a user-selected design interval χ, and denote such a design with n distinct points by

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

Here, $x_1, \ldots, x_n \in \chi$ are the design points where observations are to be taken, and $\mu_1, \ldots, \mu_n$ 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, $i = 1, \ldots, n$, subject to $N_1 + \cdots + N_n = N$.

Consider the standard nonlinear regression model given by

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

where $\varepsilon_1, \ldots, \varepsilon_N$ 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 $0 < \lambda_1 < \cdots < \lambda_k$. The design points $x_1, \ldots, x_N$ are experimental conditions, which can be chosen from a given set χ. In our case, this set is $\chi = [0, \infty)$, though in reality the extreme right end point is a large user-selected positive number. 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 (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

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

is the information matrix of the design ξ.

An optimal design maximizes a concave real valued function of the information matrix and there are several optimality criteria proposed in the literature to discriminate between competing designs, see for example, Silvey (1980) or Pukelsheim (1993). We focus on the well known D-optimality criterion and, following Chernoff (1953), we call a design ξ locally D-optimal for (3) if the design maximizes $\det M(\xi, a, \lambda)$ over all designs on the interval χ for given nominal values of a and λ. Locally D-optimal designs in various non-linear regression models have been discussed by Melas (1978), He, Studden and Sun (1996), and Dette, Haines and Imhof (1999), among 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 it is easy to see that for any design ξ on χ, the determinant of the information matrix $M(\xi, a, \lambda)$ for (3) satisfies

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

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

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

Locally D-optimal designs exist because by assumption, the induced design space

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

is compact (Pukelsheim (1993)). Moreover a locally D-optimal design has necessarily at least $n \geq 2k$ support points because, otherwise, the corresponding information matrix is 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 $\mu_1 = \mu_2 = \cdots = \mu_{2k} = 1/2k$ (Fedorov (1972)).

3. Main Results

Here we study the number of support points of a locally *D*-optimal design for the model (3). Throughout, this number is denoted by $n^*(\lambda)$. Additionally, for $k \geq 3$, we let

$$\widetilde{\lambda} = (\widetilde{\lambda_1}, \ldots, \widetilde{\lambda_k})^T \tag{5}$$

for any vector $\widetilde{\lambda}$ with components satisfying

$$0 < \widetilde{\lambda_1} < \cdots < \widetilde{\lambda_k}$$
$$\widetilde{\lambda_{i+1}} = \tfrac{\widetilde{\lambda_i} + \widetilde{\lambda_{i+2}}}{2}, \quad i = 1, \ldots, k - 2. \tag{6}$$

Theorem 1. *Let* $\lambda = (\lambda_1, \ldots, \lambda_k)^T$ *be the vector of nonlinear parameters in* (3). *If* $n^*(\lambda)$ *denotes the number of support points of a locally D-optimal design for* (3), *then*
(i) $n^*(\lambda) = 2k$, *if* $k = 1$ *or* 2;
(ii) $n^*(\lambda) \leq k(k+1)/2+1$ *for any* $k \geq 3$. *Moreover for any vector* $\widetilde{\lambda}$ *of parameters with components satisfying* (6) *there exists a neighborhood, say* $\mathcal{U} \subset \mathbb{R}^k$, *of* $\widetilde{\lambda}$, *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 cases $k = 1, 2$ the locally *D*-optimal $2k$-point design is in fact also optimal in the class of all designs. If $k \geq 3$, 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 neighborhood of a parameter vector $\widetilde{\lambda}$ 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 supported at seven points. Note that this is the upper bound for the number of support points according to the second part of Theorem 1.

Corollary 2. *Suppose* $k = 3$ *in* (3) *and* $n^*(\lambda)$ *is the number of support points of a locally D-optimal design.*
(i) $n^*(\lambda) \in \{6, 7\}$ *for any vector* λ *with increasing positive coefficients;*
(ii) *For any point* $\widetilde{\lambda} = (\widetilde{\lambda_1}, \widetilde{\lambda_2}, \widetilde{\lambda_3})$ *satisfying* $0 < \widetilde{\lambda_1} < \widetilde{\lambda_2} < \widetilde{\lambda_3}$, $\widetilde{\lambda_2} = (\widetilde{\lambda_1} + \widetilde{\lambda_3})/2$, *there exists a neighborhood of* $\widetilde{\lambda}$, *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 results on the restricted optimization in the class of all saturated designs with support points $x_1^* < \cdots < x_{2k}^*$, for which the optimal weights are equal to $1/2k$. 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 arbitrary).

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

(ii) $0 = x_1^* < \cdots < x_{2k}^*$ 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 $\lambda_i \to \lambda^* > 0$, $i = 1, \ldots, k$, 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) that $x_1^* = 0$, $x_2^* = 1/\lambda_1$. 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 an upper bound for the number of support points of any locally D-optimal designs.

We now consider locally D-optimal designs for a three-compartmental model whose mean response is given by

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

Here a_1, a_2, a_3 are the nonzero linear parameters and it is assumed that the nonlinear parameters satisfy $0 < \lambda_1 < \lambda_2 < \lambda_3$. 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. From the justification of our procedure given in the appendix, it follows that if the design space is $[0, c]$ and c is user-selected, the optimal design will remain the same if c is larger than the largest support point in the locally D-optimal design; if c is smaller than

the largest support point of the locally *D*-optimal design on the interval $[0, \infty)$, then the extreme support points of the optimal design on $[0, c]$ are 0 and c. The remaining support points are found numerically in the same way as for the unbounded design interval $[0, \infty)$.

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

$\delta_1 = 0.95$										
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	26.440	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	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	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	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	

Since $x_1^* = 0$ 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 $\widetilde{\lambda}(\varepsilon) = (1, 1 + \varepsilon, 1 + 2\varepsilon)$ 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 equivalence theorems, we have that for the D-optimality 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 support points of the locally D-optimal design for various values of δ_1 and δ_2.

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

			$\delta_1 = 0.6$				
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		

4. Discussion

Melas (1978) 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 locally *D*-optimal designs for the exponential model (1) and (2) are influenced by the 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 others. Our results suggest that other types of optimal designs for the general exponential regression model will be difficult to describe analytically or otherwise.

There are other interesting design issues for compartmental models not discussed in the paper. First, 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)). We anticipate extending similar results for models with four or more exponential terms will require more theory, and likely will require a different approach.

Second, there are biological models closely related to those studied here. For example, if we add an intercept to the our models, the resulting models are useful for studying viral dynamics and related problems (Ding and Wu (2000)). Han and Chaloner (2003) constructed optimal designs for such models with one or two exponential terms for estimating parameters in viral dynamics in an AIDS trial. More complex modeling systems will have to involve additional exponential terms.

Third, our models assume that errors are homoscedastic. Landaw and DiStefano (1999) postulated that the error variance of the ith observation in certain compartmental models is more appropriately modeled as $\alpha + \beta(y(t_i))^\gamma$, where α represents constant background variance. The three parameters α, β

and γ may be known constants from previous studies. Some extensions of our results to these models would also be useful.

Appendix. Proof of Theorem 1

We begin the proof by considerating the D-optimal design problem for the model

$$\sum_{i=1}^{2k} \beta_i e^{-\check{\lambda}_i x}, \tag{8}$$

where $0 < \check{\lambda}_1 < \cdots < \check{\lambda}_{2k}$ are fixed known values and $\beta_1, \ldots, \beta_{2k}$ are 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 for this model has the form

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

Given $\lambda = (\lambda_1, \ldots, \lambda_k)$ and $0 < \lambda_1 < \cdots < \lambda_k$, we next investigate the maximum of $\det A(\xi, \check{\lambda})$, where the components of the vector $\check{\lambda} = (\check{\lambda}_1, \ldots, \check{\lambda}_{2k})^T$ are defined by

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

If $k = 1$, the value $\Delta > 0$ can be chosen arbitrarily. Let $\xi^* = \operatorname{argmax} \det A(\xi, \check{\lambda})$ 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 $\{(e^{-\check{\lambda}_1 x}, \ldots, e^{-\check{\lambda}_{2k} x})^T \mid x \in \chi\}$ is compact (Pukelsheim (1993)). By the Kiefer-Wolfowitz equivalence theorem, we have

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

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

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

where we have used the notation $A = A(\xi^*, \check{\lambda})$ for short. We next need the following lemma whose proof is deferred to the end of this section.

Lemma A.1. *Consider the s functions given by $\varphi_i(x) = \sum_{j=1}^{t_i} \alpha_{i,j} e^{-\mu_{i,j} x}$, where t_i are arbitrary integers and $\{\alpha_{ij}, \mu_{i,j}\}$ are real numbers, $i = 0, \ldots, s$. Suppose the following conditions hold.*

(i) $\min_{1 \le j \le t_{i+1}} \mu_{i,j} > \max_{1 \le i \le t_i} \mu_{i,j}$, $i = 0, \ldots, s - 1$;
(ii) $\alpha_{i,j} > 0$, $j = 1, \ldots, t_i$, $i = 0, \ldots, s$.
Then for arbitrary real numbers $b_0, \ldots, b_s$ the function $\sum_{i=0}^{s} b_i \varphi_i(x)$ has at most
s roots counting multiplicity.

Let $A_{i,j} = (A^{-1})_{ij}$ and define

$$\varphi_0(x) \equiv m,$$

$$\varphi_{l-1}(x) = (-1)^l \sum_{i=1}^{l-1} A_{l-i,i} e^{-(\check{\lambda}_i + \check{\lambda}_{l-i})x}, \ l = 2, \ldots, 2k,$$

$$\varphi_{l-1}(x) = (-1)^l \sum_{j=1}^{4k-l+1} A_{2k+1-j,l-2k+j-1} e^{-(\check{\lambda}_{2k+1-j} + \check{\lambda}_{l-2k+j-1})x}, \ l = 2k+1, \ldots, 4k.$$

We first consider the cases $k = 1$ and 2. Note that the coefficients in the functions are positive because sign $A_{i,j} = (-1)^{i+j}$, that is, condition (ii) holds. Additionally, it follows from the definition of $\check{\lambda}$ in (10) that condition (i) can be verified directly for $k = 1, 2$. If ξ^* is D-optimal, it follows that for all x,

$$g(x) = m - f^T(x) A^{-1}(\xi^*, \check{\lambda}) f(x) = \varphi_0(x) + \sum_{i=1}^{4k-1} (-1)^i \varphi_i(x) \le 0.$$

This implies that the support points, say $x_1^*, \ldots, x_n^*$, of ξ^* satisfy $g(x_i^*) = 0$, $i = 1, \ldots, n$, $g'(x_i^*) = 0$, $i = 2, \ldots, n - 1$. A careful counting of the multiplicities and an application of Lemma 1 now show $2n - 2 \le 4k - 1$, which implies $n = 2k$ when $k = 1$ or 2.

When $k \ge 3$ the same arguments are applicable for any vector $\check{\lambda}$ satisfying (6), because in this case it can be easily verified that the functions φ_i, $i = 0, \ldots, 4k$, 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 neighborhood of the point $\check{\lambda}$.

For a proof of the second bound when $k \ge 3$, we consider an arbitrary point of the form (10), say $\check{\lambda} = (\check{\lambda}_1, \ldots, \check{\lambda}_{2k})$, and let $s \le k(k+1)/2$ be the number of distinct values in the set $\{2\lambda_1, \ldots, 2\lambda_k, \lambda_1 + \lambda_2, \ldots, \lambda_1 + \lambda_k, \lambda_2 + \lambda_3, \ldots, \lambda_{k-1} + \lambda_k\}$. Further, we denote by $u_1 < \cdots < u_s$ the distinct values from this set and introduce the functions

$$\check{\varphi}_0(x) \equiv m,$$
$$\check{\varphi}_1(x) = A_{1,1} e^{-u_1 x} = A_{1,1} e^{-2\lambda_1 x},$$
$$\check{\varphi}_2(x) = -2A_{1,2} e^{-(u_1 + \Delta)x},$$
$$\check{\varphi}_{2l-1}(x) = a_l e^{-(u_l + 2\Delta)x} + c_l e^{-u_{l+1} x}, \ l = 2, \ldots, s,$$
$$\check{\varphi}_{2l}(x) = -b_l e^{-(u_l + \Delta)x}, \ l = 2, \ldots, s,$$
$$\check{\varphi}_{2s+1}(x) = a_{s+1} e^{-(u_s + 2\Delta)x}.$$

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

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

is satisfied and $a_l, b_l, c_l > 0$, $l = 1, \ldots, s$. By the above arguments, we have $\check{J}(\tau) = \det (\check{\varphi}_i(x_j))_{i,j=0}^{2s+1} > 0$ for any $\tau = (x_0, \ldots, x_{2s+1})$ with $x_0 > \cdots > x_{2s+1}$. Moreover, for any vector $\bar{\tau} = (\bar{x}_0, \ldots, \bar{x}_{2s+1})^T$ with components satisfying $\bar{x}_0 \geq \bar{x}_1 \geq \cdots \geq \bar{x}_{2s+1}$, it follows

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

From (12) and the Equivalence Theorem for D-optimality, we obtain the inequality $g(x) \leq 0$ for any $x \geq 0$, where g is the generalized polynomial $g(x) = \check{\varphi}_0(x) - \sum_{i=1}^{2s+1} \check{\varphi}_i(x)$. This implies that the support points of the locally D-optimal design satisfy $g(x_i^*) = 0$, $i = 1, \ldots, n$, $g'(x_i^*) = 0$, $i = 2, \ldots, n-1$, and a similar argument shows that the function g has at most $2s + 1$ roots counting multiplicity. Consequently, $2n - 2 \leq 2s + 1 \leq k(k+1) + 1$, which yields $n \leq (k(k+1)/2) + 1 + 1/2$. This proves the assertion of the theorem for (8).

To prove the theorem for (3), we consider an arbitrary approximate design ξ and the polynomial defined by

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

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

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

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

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

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

Consequently we have for any design ξ

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

where $M(\xi, \lambda)$ is the information matrix under (3). If ξ^* denotes a locally D-optimal design for (3) with support points $x_1^* < \cdots < x_{n^*}^*$, then it follows from (13) and (14) that

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

where the vector $\breve{f}^T(x)$ corresponds to the gradient in (3) and is defined by $\breve{f}^T(x) = (e^{-\lambda_1 x}, -xe^{-\lambda_1 x}, \ldots, e^{-\lambda_k x}, -xe^{-\lambda_k x})$. By the Equivalence Theorem, the polynomial on the left hand side has roots $x_1^*, \ldots, x_{n^*}^*$, and $x_2^*, \ldots, x_{n^*-1}^*$ are all roots with multiplicity two. Consequently, we obtain $2n^* - 1 \le h$, 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 $h \le 4k - 1$ for $k = 1, 2$, and for $k \ge 3$ in a neighborhood of points λ satisfying (6). Moreover, we have $h \le k(k+1)/2$ in general and this completes the proof of the theorem.

Proof of Lemma A.1. Let $\tau^T = (x_0, \ldots, x_s)^T$ and let $J(\tau) = \det(\varphi_i(x_j))_{i,j=0}^s$. If we expand the determinant along a row repeatedly, we arrive at

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

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

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

since $\lim_{\tau \to \bar{\tau}} \det(e^{-\theta_{i_s} x_j})_{s,j=0}^t / \prod_{i<j}(x_i - x_j) > 0$ whenever $0 < \theta_{i_0} < \cdots < \theta_{i_s}$. This property can be easily verified by considering the number of the same coordinates in the vector $\bar{x}$. It is known (see Karlin and Studden (1966, Chap. 1)) that under the conditions that $J > 0$ and (16), any generalized polynomial of the form $\sum_{i=0}^t b_i \varphi_i(t)$ has at most t roots counting multiplicity.

Acknowledgement

referees and an associate editor for constructive comments on an earlier version of this manuscript, and to Isolde Gottschlich who typed parts of this paper with considerable technical expertise.

References

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.

Becka, M., Bolt, H. M. and Urfer, W. (1992). Statistical analysis of toxicokinetic data by nonlinear regression (example: inhalation pharmacokinetics of propylene). *Arch. Toxicol* **66**, 450-453.

Chaloner, K. and 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. and Imhof, L. A. (1999). Optimal designs for rational models and weighted polynomial regression. *Ann. Statist.* **27**, 1272-1293.

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

Ding, A. A. and Wu, H. (2000). A comparison study of models and fitting procedures for biphasic viral dynamics in HIV-1 infected patients treated with antiviral therapies. *Biometrics* **56**, 293-300.

Droz, P. O., Berode, M. and Jang, J. Y. (1999). Biological monitoring of Tetrahydrofuran: contribution of a physiologically based pharmacokinetic model. *American Industrial Hygiene Association*, 243-248.

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

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Gantmacher, F. R. (1959). *The Theory of Matrices*, Vol. II. Chelsea, New York.

Gibaldi, M. and Perrier, D. (1982). *Pharmacokinetics*. Marcel Dekker, New York.

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

Haines, L. M. (1995). A geometric approach to optimal design for one-parameter nonlinear models. *J. Roy. 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. Statist. Plann. Inference* **115**, 585-601.

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

Jones, B. and Wang, J. (1999). Constructing optimal designs for fitting pharmacokinetic models. *Statist. Comput.* **9**, 209-218.

Karlin, S. and Studden, W. (1966). *Tchebysheff Systems: with Application in Analysis and Statistics.* Wiley, New York.

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

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

Landaw, E. M. and DiStefano, J. J. (1999). Multiexpontial, multicompartmental and noncompartmental modeling. II. Data analysis and statistical considerations. *Amer. J. Physiol.*, R665-R676.

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

Melas, V. B. (2000). Analytic theory of *E*-optimal designs for polynomial regression. In *Advances in Stochastic Simulation Methods* (Edited by N. Balakrishnan, V. B. Melas and S. Ermarkov), 85-116. Birkhäuser, Boston.

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.

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach.* Marcel Dekker, New York.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models.* Marcel Dekker, New York.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression.* Wiley, New York.

Shargel, L. and Yu, A. B. C. (1985). *Applied Biopharmaceutics and Pharmacokinetics.* 2nd Edition. Appleton-Century-Crofts, Norwalk, Connecticut.

Silvey, S. D. (1980). *Optimum Design.* Chapman and Hall, London.

Szegö, G. (1975). Orthogonal Polynomials. *American Mathematical Society Colloquium Publications* **23**. Providence, Rhode Island.

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.

E-mail: holger.dette@ruhr-uni-bochum.de

Department of Mathematics, St. Petersburg State University, St. Petersburg, Russia.

E-mail: v.melas@pobox.spbu.ru

Department of Biostatistics, University of California at Los Angeles, Los Angeles, U.S.A.

E-mail: wkwong@ucla.edu