'Nearly' universally optimal designs for models with correlated observations

Holger Dette[1]

Ruhr-Universität Bochum

Andrey Pepelyshev[2]

RWTH Aachen University, Germany

Anatoly Zhigljavsky[3,*]

Cardiff University

Abstract

In this paper the problem of determining optimal designs for least squares estimation is considered in the common linear regression model with correlated observations. Our approach is based on the determination of 'nearly' universally optimal designs, even in the case where the universally optimal design does not exist. For this purpose we introduce a new optimality criterion which reflects the distance between a given design and an ideal, universally optimal design. A necessary conditions for the optimality of a given design is established. Numerical methods for constructing these designs are proposed and applied for the determination of optimal designs in a number of specific instances. The results indicate that the new 'nearly' universally optimal designs have good efficiencies with respect to common optimality criteria.

Keywords: Optimal design; correlated observations; universally optimal design; multiplicative algorithms

*Corresponding author.

Email addresses: holger.dette@rub.de (Holger Dette), pepelyshev@stochastik.rwth-aachen.de (Andrey Pepelyshev), ZhigljavskyAA@cf.ac.uk (Anatoly Zhigljavsky)

[1]Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, 44780, Germany.

[2]Institute of Statistics, RWTH Aachen University, Wüllnerstr. 3, D-52056 Aachen.

[3]School of Mathematics, Cardiff University, Cardiff, CF24 4AG, UK

1. Introduction

Consider the common linear regression model

$$y(x) = \theta_1 f_1(x) + \ldots + \theta_m f_m(x) + \varepsilon(x) \ , \tag{1}$$

where $f_1(x), \ldots, f_m(x)$ are linearly independent, continuous functions, $\varepsilon(x)$ denotes a random error process or field, $\theta_1, \ldots, \theta_m$ are unknown parameters and x is the explanatory variable, which varies in a compact design space $\mathcal{X} \subset \mathbb{R}^d$. We assume that N observations, say $y_1, \ldots, y_N$, can be taken at experimental conditions $x_1, \ldots, x_N$ to estimate the parameters in the linear regression model (1). Suppose that $\varepsilon(x)$ is a stochastic process with

$$\mathrm{E}[\varepsilon(x)] = 0, \quad \mathrm{E}[\varepsilon(x)\varepsilon(x')] = K(x, x') \, , \quad x \in \mathcal{X} \subset \mathbb{R}^d \, . \tag{2}$$

Throughout this paper we call the function $K(x, x')$ covariance kernel. An important case appears when the error process is stationary and the covariance kernel is of the form $K(x, x') = \sigma^2 \rho(x - x')$, where $\rho(0) = 1$. The function $\rho(\cdot)$ is called the correlation function.

If N observations, say $y = (y_1, \ldots, y_N)^T$, are available at experimental conditions $x_1, \ldots, x_N$ and the covariance kernel is known, then the vector of parameters can be estimated by the weighted least squares method, that is, by $\hat{\theta} = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} y$ where $\mathbf{X} = (f_i(x_j))_{j=1,\ldots,N}^{i=1,\ldots,m}$ and $\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1,\ldots,N}$. The variance-covariance matrix of this estimate is given by

$$\mathrm{Var}(\hat{\theta}) = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \, .$$

If the correlation structure of the process is not known, one usually uses the ordinary least squares estimate $\tilde{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y$, which has the covariance matrix

$$\mathrm{Var}(\tilde{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \tag{3}$$

An exact experimental design $\{x_1, \ldots, x_N\}$ is a collection of N points from the design space $\mathcal{X}$, which defines the time points or experimental conditions where observations are taken. Optimal designs for weighted or ordinary least squares estimation minimize a functional of the covariance matrix

of the weighted or ordinary least squares estimate, respectively, and numerous optimality criteria have been proposed in the literature to discriminate between competing designs [see Pukelsheim (2006)].

Exact optimal designs for specific linear models have been investigated in Dette et al. (2008b); Kiseľák and Stehlík (2008); Harman and Štulajter (2010). Because even in simple models exact optimal designs are difficult to find, most authors use asymptotic arguments to determine efficient designs for the estimation of the model parameters [see Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979) or Zhigljavsky et al. (2010)]. Sacks and Ylvisaker (1966, 1968) and Näther (1985), Chapter 4, assumed that the design points $\{x_1, \ldots, x_N\}$ are generated by the quantiles of a distribution function; that is, $x_i = a((i-1)/(N-1))$, $i = 1, \ldots, N$, where the function $a : [0,1] \to \mathcal{X}$ is the inverse of a distribution function. Let ξ_N denote a normalized design supported at N points $\{x_1, \ldots, x_N\}$ with the weight $1/N$ assigned to each point. Then the covariance matrix of the least squares estimate $\tilde{\theta}$ given in (3) can be represented as

$$\mathrm{Var}(\tilde{\theta}) = D(\xi_N) = M^{-1}(\xi_N) B(\xi_N, \xi_N) M^{-1}(\xi_N), \tag{4}$$

where the matrices $M(\xi_N)$ and $B(\xi_N, \xi_N)$ are defined by

$$M(\xi_N) = \int_{\mathcal{X}} f(u) f^T(u) \xi_N(\mathrm{d}u), \tag{5}$$

$$B(\xi_N, \xi_N) = \int_{\mathcal{X}} \int_{\mathcal{X}} K(u,v) f(u) f^T(v) \xi_N(\mathrm{d}u) \xi_N(\mathrm{d}v), \tag{6}$$

respectively, and $f(u) = \big(f_1(u), \ldots, f_m(u)\big)^T$ denotes the vector of regression functions. Following Kiefer (1974) we call any probability measure ξ on $\mathcal{X}$ an approximate design or simply design. The definitions of the matrices $M(\xi)$ and $B(\xi, \xi)$ can be extended to an arbitrary design ξ, provided that the corresponding integrals exist. The matrix

$$D(\xi) = M^{-1}(\xi) B(\xi, \xi) M^{-1}(\xi), \tag{7}$$

is called the covariance matrix for the design ξ and can be defined for any probability measure ξ supported on the design space $\mathcal{X}$ such that the matrices $B(\xi, \xi)$ and $M^{-1}(\xi)$ are well-defined. This set will be denoted by Ξ.

Optimal designs for regression models with dependent data have been investigated mainly for the location scale model. The difficulties in a general

development of the optimal design theory for correlated observations can be explained by the different structure of the covariance of the least squares estimator in model (1), which is of the form $M^{-1}BM^{-1}$. As a consequence, the corresponding design problems are in general not convex (except for the location scale model where $m = 1$ and $f_1(u) \equiv 1$). Recently, Dette et al. (2011) derived universally optimal designs for regression models of arbitrary dimension if the corresponding regression functions are eigenfunctions of an integral operator defined by the covariance kernel of the error process. On the other hand there are many situations where this assumption is not satisfied and in these cases there does not exist a universally optimal design.

The present paper is devoted to the numerical construction of 'nearly' universally optimal designs for regression models in such situations. This means that we consider the model (1) with $m > 1$ parameters in the case where a universally optimal design does not exist. In Section 2 we introduce a new optimality criterion which reflects the distance between a given design and an ideal, universally optimal design. A necessary conditions for the optimality of a given design is established in Section 3 and an algorithm for its numerical determination is proposed in Section 4. Finally, some illustrative examples are given in Section 5, where we calculate 'nearly' universally optimal designs for a quadratic regression model and a nonlinear model with various correlation functions. The results indicate that the new 'nearly' universally optimal designs have good efficiencies with respect to common optimality criteria.

2. A new optimality criterion - g-optimal designs

Throughout this paper we assume that the kernel K in (6) is continuous at all points $(x, x') \in \mathcal{X} \times \mathcal{X}$ except possibly the diagonal points (x, x). We also assume that $K(x, x') \neq 0$ for at least one pair (x, x') with $x \neq x'$. Singular kernels appear naturally if the approach in Bickel and Herzberg (1979) for the approximation of the covariance matrix in (3) is extended such that the variance of the observations also depends on the sample size [see Zhigljavsky et al. (2010) for details and Dette et al. (2011) for an application]. Because in this paper we are interested in designs maximizing functionals of the matrix $D(\xi)$ (independently from the type of approximation which has been used to derive it), we will also consider singular kernels in the following discussion.

An (approximate) optimal design minimizes some functional of the covariance matrix $D(\xi)$ over the set Ξ and a universally optimal design ξ^* (if it exists) minimizes the matrix $D(\xi)$ with respect to the Loewner ordering,

that is

$$D(\xi^*) \leq D(\xi) \quad \text{for all} \quad \xi \in \Xi.$$

Universally optimal designs are attractive as they are optimal with respect to all monotone optimality criteria. Note that the design ξ^* is universally optimal if and only if the design ξ^* is c-optimal for all $c \in \mathbb{R}^m \backslash \{0\}$.

Unfortunately universally optimality designs exist only in rare circumstances and in this paper we make an attempt to find designs which have similar properties as ideal universally optimal designs. As a consequence we expect the designs derived by the method proposed here to have high efficiencies for many commonly used optimality criteria. To be precise we follow Dette et al. (2011) and introduce for a given design ξ the vector-valued function

$$g_\xi(x) = \int K(x,u) f(u) \xi(du) - B(\xi,\xi) M^{-1}(\xi) f(x)\,, \quad x \in \mathcal{X}\,.$$

This function satisfies the equality $\int g_\xi(x) f^T(x) \xi(\mathrm{d}x) = 0$. Also, as the vector of regression functions f is continuous on $\mathcal{X}$, the function g_ξ is continuous too. In what follows we shall use the notation $\Lambda_\xi = B(\xi,\xi) M^{-1}(\xi)$ and $Q_\xi(x) = \int K(x,u) f(u) \xi(du)$. With this notation we obtain the representation

$$g_\xi(x) = Q_\xi(x) - \Lambda_\xi f(x).$$

In Theorem 3.3 in Dette et al. (2011) it is established that a design ξ is universally optimal if and only if the condition

$$g_\xi(x) = 0$$

holds for all $x \in \mathcal{X}$. Consequently, we can expect that the design ξ would resemble the properties of the universally optimal design (despite it may not exist) if $\|g_\xi(x)\|$ is small for all x, where $\|\cdot\|$ denotes the euclidean norm on $\mathbb{R}^m$. Moreover, Dette et al. (2011) pointed out that for a universally optimal design the corresponding LSE gives exactly the same asymptotic covariance matrix as the BLUE and the optimal design for the BLUE. Therefore we may also expect that in this case the covariance matrices of the LSE and BLUE estimators are very similar. These considerations have motivated us to measure the size of the function $x \to g_\xi(x)$ by an L^2-norm and to find a design minimizing this norm. To be precise we call the optimality criterion

$$\Phi(\xi) = \int_{\mathcal{X}} \|g_\xi(x)\|^2 dx = \int_{\mathcal{X}} g_\xi^T(x) g_\xi(x) dx \tag{8}$$

g-optimality criterion, and a design ξ_g^* minimizing Φ will be called g-optimal. In particular, a design ξ^* is universally optimal if and only if $\Phi(\xi^*) = 0$.

3. Necessary condition for optimality

In the case of correlated observations optimality criteria are generally not convex, see Zhigljavsky et al. (2010). Therefore, standard optimality theorems do not give full characterizations of optimal designs. These results only provide necessary conditions for the optimality of a design. Theorem 1 below is not an exception and only gives a necessary conditions for g-optimality.

Theorem 1. *If the design ξ^* is g-optimal for the linear regression model* (1) *then*

$$\varphi(x, \xi^*) \geq \Phi(\xi^*) \quad \textit{for all} \;\; x \in \mathcal{X},$$

where the function φ is defined by

$$\varphi(x,\xi) = \int g_\xi^T(t) \Big(K(t,x) f(x) - \Big[g_\xi(x) f^T(x) + f(x) Q_\xi^T(x) \Big] M^{-1}(\xi) f(t) \Big) dt. \quad (9)$$

Proof. Using standard arguments of optimal design theory the necessary condition for the g-optimality of the design ξ is the fullfilment of the inequality

$$\phi(\nu, \xi) \geq 0$$

for all ν, where $\phi(\nu,\xi) = \frac{\partial}{\partial \alpha} \Phi(\xi_\alpha)\big|_{\alpha=0}$ and $\xi_\alpha = (1-\alpha)\xi + \alpha\nu$. Let us now compute the directional derivative $\phi(\nu,\xi)$ at ξ in the direction of ν. First, we note that

$$\frac{\partial}{\partial \alpha} M^{-1}(\xi_\alpha)\Big|_{\alpha=0} = M^{-1}(\xi) - M^{-1}(\xi) M(\nu) M^{-1}(\xi)$$

and

$$\frac{\partial}{\partial \alpha} B(\xi_\alpha, \xi_\alpha)\Big|_{\alpha=0} \;\; = \;\; B(\xi,\nu) + B(\nu,\xi) - 2B(\xi,\xi).$$

Then the explicit form for $A_{\nu,\xi} := \frac{\partial}{\partial\alpha}\Lambda_{\xi_\alpha}\Big|_{\alpha=0}$ is

$$\begin{aligned} A_{\nu,\xi} &= \Big[B(\xi,\nu)+B(\nu,\xi)-2B(\xi,\xi)\Big]M^{-1}(\xi) \\ &\quad +B(\xi,\xi)\Big[M^{-1}(\xi)-M^{-1}(\xi)M(\nu)M^{-1}(\xi)\Big] \\ &= \Big[B(\xi,\nu)+B(\nu,\xi)\Big]M^{-1}(\xi)-B(\xi,\xi)\Big[M^{-1}(\xi)+M^{-1}(\xi)M(\nu)M^{-1}(\xi)\Big] \\ &= \Big[B(\xi,\nu)+B(\nu,\xi)-B(\xi,\xi)-\Lambda_\xi M(\nu)\Big]M^{-1}(\xi)\,. \end{aligned}$$

Therefore

$$\phi(\nu,\xi)=\int\Big[\big(Q_{\nu-\xi}(x)-A_{\nu,\xi}f(x)\big)^T g_\xi(x)+g_\xi^T(x)\big(Q_{\nu-\xi}(x)-A_{\nu,\xi}f(x)\big)\Big]dx.$$

Since $\phi(\nu,\xi)$ is a scalar function we have

$$\phi(\nu,\xi)=2\int g_\xi^T(x)\big(Q_{\nu-\xi}(x)-A_{\nu,\xi}f(x)\big)dx.$$

Let δ_t be the delta-measure supported at a point t. Then $M(\delta_t)=f(t)f^T(t)$, $B(\xi,\delta_t)=B^T(\delta_t,\xi)=Q_\xi(t)f^T(t)$,

$$Q_{\delta_t-\xi}(x)=K(x,t)f(t)-Q_\xi(x) \tag{10}$$

and

$$\begin{aligned} A_{\delta_t,\xi} &= \Big[Q_\xi(t)f^T(t)+f(t)Q_\xi^T(t)\Big]M^{-1}(\xi) \\ &\quad -B(\xi,\xi)\Big[M^{-1}(\xi)+M^{-1}(\xi)f(t)f^T(t)M^{-1}(\xi)\Big] \\ &= \Big[Q_\xi(t)f^T(t)+f(t)Q_\xi^T(t)-\Lambda_\xi f(t)f^T(t)\Big]M^{-1}(\xi)-B(\xi,\xi)M^{-1}(\xi) \\ &= \Big[g_\xi(t)f^T(t)+f(t)Q_\xi^T(t)\Big]M^{-1}(\xi)-\Lambda_\xi\,. \end{aligned} \tag{11}$$

We can now write

$$\phi(t,\xi)=\phi(\delta_t,\xi)=2\int g_\xi^T(x)\Big(Q_{\delta_t-\xi}(x)-A_{\delta_t,\xi}f(x)\Big)dx=2(\varphi(t,\xi)-\chi(\xi))$$

where $\varphi(t,\xi)$ is defined in (9) and $\chi(\xi)$ takes into account the last terms in (10) and in (11), that is

$$\chi(\xi)=\int g_\xi^T(x)\Big(Q_\xi(x)-\Lambda_\xi f(x)\Big)dx=\Phi(\xi).$$

□

Remark 1. The following expression for the function $\varphi(x,\xi)$ defined in (9) may look more appealing:

$$\varphi(x,\xi) = f^T(x) \int K(x,t) g_\xi(t) dt - \mathrm{Tr}(A_\xi B_\xi),$$

where the matrices A_ξ and B_ξ are given by

$$A_\xi = \Big[g_\xi(x) f^T(x) + f(x) g_\xi^T(x) \Big] M^{-1}(\xi) + f(x) f^T(x) D(\xi), \;\; B_\xi = \int f(t) g_\xi^T(t) dt \,.$$

4. An algorithm for construction of optimal designs

The fact that the optimality theorems only give a necessary condition for design optimality does not usually create additional problems for the algorithms of construction of designs. Numerical computation of optimal designs for the common linear regression model (1) with given correlation function can be performed by an extension of the multiplicative algorithm proposed by Dette et al. (2008c) for the case of non-correlated observations [see also Yu (2010) for some extensions]. Note that the proposed algorithm constructs a discrete design which can be considered as an approximation to a design satisfying the necessary conditions for optimality in Theorem 1. By choosing a fine discretization $\{x_1, \ldots, x_n\}$ of the design space $\mathcal{X}$ and running the algorithm long enough, the accuracy of approximation can be made arbitrarily small. Numerical experiments show that the convergence is always achieved.

Denote by $\xi^{(r)} = \{x_i; w_i^{(r)}\}_{i=1}^n$ the design at the iteration r. We propose the following updating rule for the weights

$$w_i^{(r+1)} = w_i^{(r)} \frac{\psi_r(x_i, \xi^{(r)})}{\sum_{j=1}^n w_j^{(r)} \psi_r(x_j, \xi^{(r)})}, \qquad i = 1, \ldots, n, \tag{12}$$

where

$$\psi_r(x,\xi) = 1 + \beta_r \big(\Phi(\xi) - \varphi(x,\xi) \big)_+ = 1 + \beta_r \max\big\{ 0, \Phi(\xi) - \varphi(x,\xi) \big\}$$

and β_r is chosen to ensure the decrease of the optimality criterion (β_r is small positive). Note that the initial weights $w_1^{(0)}, \ldots, w_n^{(0)}$ should be nonzero, for example, $w_i^{(0)} = 1/n$ for $i = 1, \ldots, n$.

We write the algorithm for computing the g-optimal design in the following form.

Algorithm.

1. Set $w_1^{(0)} = 1/n, \ldots, w_n^{(0)} = 1/n$, $\beta_0 = 1$, $\epsilon = 10^{-4}$ and $r = 0$.

2. For $i = 1, \ldots, n$ compute $w_i^{(r+1)}$ according to (12).

3. If
$$\Phi\big(\{x_i; w_i^{(r+1)}\}_{i=1}^n\big) < \Phi\big(\{x_i; w_i^{(r)}\}_{i=1}^n\big),$$
then set $\beta_{r+1} = \beta_r$. Otherwise, redefine $w_i^{(r+1)} = w_i^{(r)}$ and set $\beta_{r+1} = \beta_r/2$.

4. If
$$\big|\Phi\big(\{x_i; w_i^{(r+1)}\}_{i=1}^n\big) - \Phi\big(\{x_i; w_i^{(r)}\}_{i=1}^n\big)\big| < \epsilon \quad \text{and} \quad |\beta_{r+1} - \beta_r| < \epsilon,$$
then terminate the algorithm.

5. Set $r = r + 1$ and return to step 2.

Note that the necessary condition of optimality takes the form $\psi_r(x, \xi^*) \leq 1$ for all $x \in \mathcal{X}$. Therefore, the rule (12) means that at the next iteration the weight of a point $x = x_j$ increases if the necessary condition does not hold at this point.

A measure ξ_* is a fixed point of the iteration (12) if and only if $\psi_r(x, \xi_*) = 1$ for all x. Thus, the design ξ_* is a fixed point of the iteration (12) if and only if it satisfies the optimality condition of Theorem 1. We were not able to theoretically prove the convergence of the algorithm to the design satisfying the optimality condition of Theorem 1 but we observed the convergence in all numerical studies. In particular, for the cases where we could derive the optimal designs explicitly, we observed the convergence of the algorithm to the optimal design.

5. Examples

In this section, we provide some numerical results which we have received applying Algorithm described in Section 4 for the calcualtion of g-optimal designs in several regression models. In the tables below we shall use the following notation for the D-efficiency, A-efficiency and g-efficiency of a design ξ:

$$\mathrm{Eff}_D(\xi) = \left(\frac{\det D(\xi_D^*)}{\det D(\xi)}\right)^{1/m}, \ \mathrm{Eff}_A(\xi) = \frac{\mathrm{Tr}(D(\xi_A^*))}{\mathrm{Tr}(D(\xi))},$$

and

$$\mathrm{Eff}_g(\xi) = \left(\frac{\int g_{\xi_g^*}^T(x) g_{\xi_g^*}(x) dx}{\int g_\xi^T(x) g_\xi(x) dx}\right)^{1/2}.$$

Here ξ_D^* is the D-optimal design, ξ_A^* is the A-optimal design, ξ_g^* is the g-optimal design and m is the number of unknown parameters. We will also use the notation ξ_u for the uniform design of $\mathcal{X}$ and ξ_a for the arcsine design on the interval $[-1,1]$; this is the design with density

$$p(x) = 1/(\pi\sqrt{1-x^2}), \quad x \in [-1,1].$$

5.1. Approximating the logarithmic kernel

Consider the quadratic regression model (that is $m = 3$ and $f(x) = (1, x, x^2)^T$) on the interval $\mathcal{X} = [-1,1]$ and the covariance function $K_\delta(u,v) = \rho_\delta(u-v)$ where

$$\rho_\delta(t) = 2 - \frac{1}{\delta}\log\Big(\frac{|t+\delta|^{t+\delta}}{|t-\delta|^{t-\delta}}\Big). \tag{13}$$

The functions $\rho_\delta(t)$ converge to the logarithmic kernel $\rho(t) = -\log(t)^2$ as $\delta \to 0$, for all $t \neq 0$.

In Figure 1 we display the g-optimal designs (constructed numerically by the application of the multiplicative algorithm) for δ =0.02, 0.05 and 0.1. As one can see, for small δ these designs are very close to the arcsine design, which is the universally optimal design for the quadratic model and the logarithmic kernel, as proved in Dette et al. (2011). We also note that $g_{\xi_a}(x) \equiv 0$ for $\delta = 0$ while $\Phi(\xi_g^*(\delta))$ equals $0.0018^2, 0.029^2$ and 0.056^2 for $\delta = 0.02, 0.05, 0.1$, respectively.

Note that for $\delta = 0.02$, 0.05 and 0.1 we have $\Phi(\xi_a) = 0.034^2$, $\Phi(\xi_a) = 0.074^2$ and $\Phi(\xi_a) = 0.13^2$, respectively. In Table 1 we show D-, A- and g-efficiencies of the designs ξ_D^*, ξ_A^*, ξ_g^*, ξ_a and ξ_u. We can observe that the

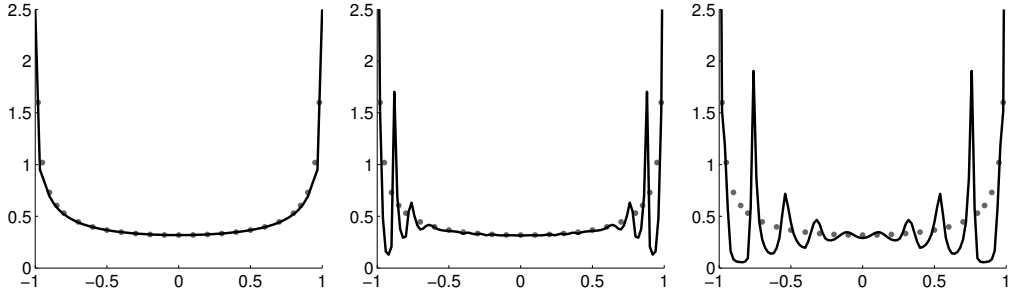

Figure 1: *The g-optimal designs for the quadratic model with covariance kernel* (13)*, where* $\delta = 0.02$ *(left),* $\delta = 0.05$ *(middle) and* $\delta = 0.1$ *(right). The dotted line corresponds to the arcsine density.*

g-optimal design has very high D- and A-efficiencies but the g-efficiency of ξ_D^* and ξ_A^* is only about 0.8. The arcsine and the uniform design are less efficient, in particular with respect to the g-optimality criterion.

In Figure 2 we depict the components g_1, g_2, and g_3 of the vector-valued function $g_\xi(x)$ for designs ξ_g^*, ξ_D^* and ξ_a in the case $\delta = 0.05$. We can see that $g_{\xi_g^*}(x)$ has smaller absolute values than $g_{\xi_D^*}(x)$ and $g_{\xi_a}(x)$. For these designs we have $\Phi(\xi_g^*) = 0.029^2$, $\Phi(\xi_D^*) = 0.044^2$, $\Phi(\xi_a) = 0.074^2$ (here $\delta = 0.05$).

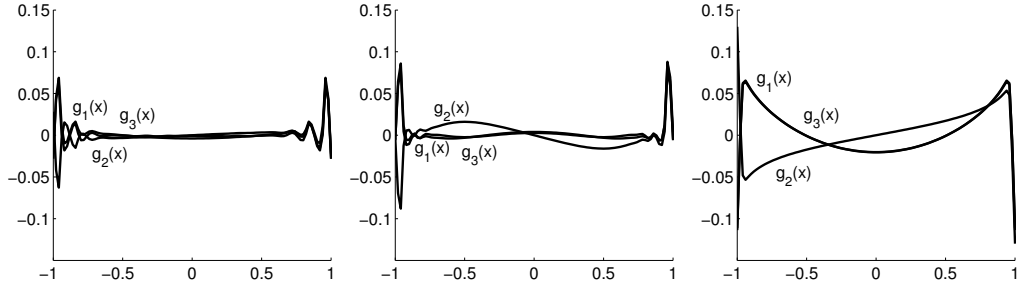

Figure 2: *The components of the vector-valued function* $g_\xi(x) = (g_1(x), g_2(x), g_3(x))^T$ *for the quadratic model with covariance kernel* (13) *(with* $\delta = 0.05$*) for the g-optimal design (left), the D-optimal design (middle) and the arcsine design (right).*

5.2. The exponential correlation function

Consider the quadratic model with stationary error process having the exponential correlation function $K(u, u+t) = \rho(t) = e^{-\lambda|t|}$. In this case the existence of a universally optimal design is an open problem, but we can use

Table 1: *D-, A- and g-Efficiencies of the designs ξ_D^*, ξ_A^*, ξ_g^*, ξ_a and ξ_u for the quadratic model and the covariance kernel* (13) *with* $\delta = 0.02, 0.05, 0.1$.

	$\delta = 0.02$			$\delta = 0.05$			$\delta = 0.1$		
ξ	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_g(\xi)$	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_g(\xi)$	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_g(\xi)$
ξ_D^*	1	0.999	0.878	1	0.999	0.842	1	0.999	0.702
ξ_A^*	0.999	1	0.746	0.999	1	0.783	0.999	1	0.586
ξ_g^*	0.999	0.999	1	0.992	0.992	1	0.987	0.988	1
ξ_u	0.91	0.90	0.12	0.87	0.86	0.13	0.83	0.81	0.23
ξ_a	0.99	0.99	0.41	0.97	0.97	0.42	0.94	0.94	0.43

the algorithm described in Section 4 to provide a design with similar properties. In Figure 3 we display the densities corresponding to the g-optimal designs. The values of $\Phi(\xi_g^*(\lambda))$ are equal to 0.0025^2, 0.0043^2 and 0.0019^2 for $\lambda = 1, 4, 8$, respectively. The corresponding efficiencies of various designs are shown in Table 2 for the cases $\lambda = 1$, $\lambda = 4$ and $\lambda = 8$, respectively, and we observe a similar pattern as described in the previous example. In particular the D- and A-optimal design have a rather low g-efficiency.

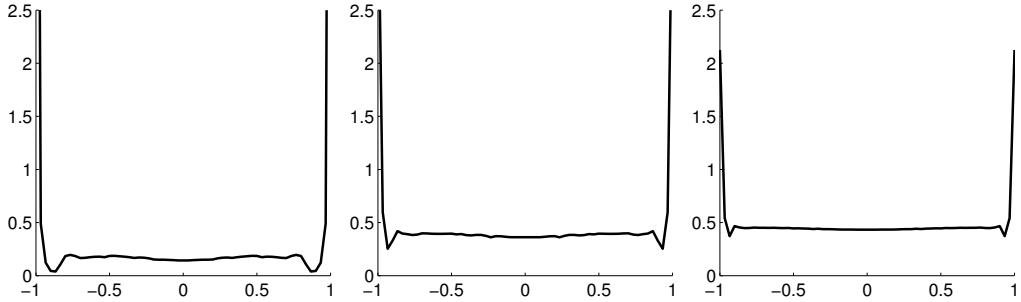

Figure 3: *The g-optimal designs for the quadratic model with exponential covariance function* $\rho(t) = e^{-\lambda|t|}$, *where* $\lambda = 1$ *(left),* $\lambda = 4$ *(middle) and* $\lambda = 8$ *(right).*

5.3. Optimal designs for a nonlinear model

In this section we extend the methodology to the case of nonlinear models. Exemplarily, we consider the compartmental model with first-order absorption,

$$\eta(x, \theta) = \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_2 x} - e^{-\theta_1 x}) \tag{14}$$

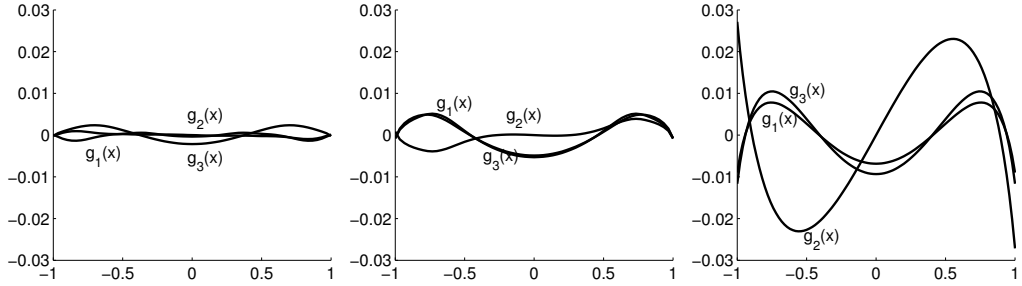

Figure 4: *The components of the vector-valued function* $g_\xi(x) = (g_1(x), g_2(x), g_3(x))^T$ *for the quadratic model with exponential correlation function* $\rho(t) = e^{-\lambda|t|}$ *(*$\lambda = 1$*) for the g-optimal design (left), the D-optimal design (middle) and the arcsine design (right). If* $\lambda = 1$ *we have for these designs* $\Phi(\xi_g^*) = 0.0025^2$*,* $\Phi(\xi_D^*) = 0.0078^2$*,* $\Phi(\xi_a) = 0.026^2$*.*

Table 2: *D-, A- and g-Efficiency of the designs* ξ_D^**,* ξ_A^**,* ξ_g^**,* ξ_a^* *and* ξ_u *for the quadratic model and the exponential correlation function* $\rho(t) = e^{-\lambda|t|}$ *with* $\lambda = 1, 4, 8$*.*

	$\lambda = 1$			$\lambda = 4$			$\lambda = 8$		
ξ	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_g(\xi)$	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_g(\xi)$	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_g(\xi)$
ξ_D^*	1	0.997	0.387	1	0.999	0.57	1	0.99	0.710
ξ_A^*	0.998	1	0.341	0.999	1	0.40	0.999	1	0.468
ξ_g^*	0.996	0.993	1	0.998	0.996	1	0.999	0.998	1
ξ_u	0.82	0.83	0.09	0.85	0.82	0.12	0.91	0.88	0.09
ξ_a	0.93	0.94	0.10	0.97	0.96	0.22	0.98	0.98	0.09

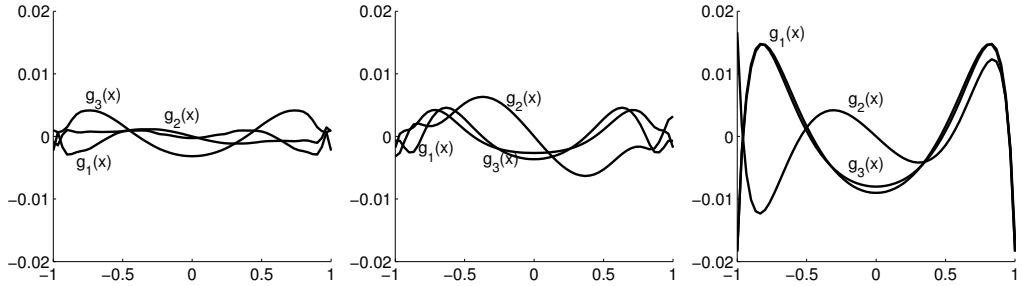

Figure 5: *The components of the vector-valued function* $g_\xi(x) = (g_1(x), g_2(x), g_3(x))^T$ *for the quadratic model with exponential correlation function* $\rho(t) = e^{-\lambda|t|}$ *(*$\lambda = 4$*) for the g-optimal design (left), the D-optimal design (middle) and the arcsine design (right). If* $\lambda = 4$ *we have for these designs* $\Phi(\xi_g^*) = 0.0043^2$*,* $\Phi(\xi_D^*) = 0.0076^2$*,* $\Phi(\xi_a) = 0.020^2$*.*

The model (14) is a special case of the Bateman function, defined in the introduction [see Garrett (1994)], and has found considerable attention in

chemical sciences, toxicology and pharmacokinetics [see, for example, Gibaldi and Perrier (1982)]. The optimal design problem in the compartmental model with uncorrelated observations has been studied by numerous authors [see, for example, Box and Lucas (1959), Atkinson et al. (1993), Dette and O'Brien (1999), Biedermann et al. (2004) among others].

We now consider optimal design problems in the case of correlated observations and determine locally optimal designs in the sense of Chernoff (1953), which require a specification of the unknown parameters. There are many situations where such preliminary knowledge is available, such that the application of locally optimal designs is well justified [see Dette et al. (2008a)]. However, the most important application of locally optimal designs is their use as benchmarks for commonly proposed designs. Moreover, they are the basis for more sophisticated design strategies, which require less precise knowledge about the model parameters, such as sequential, Bayesian or standardized maximin optimality criteria [see Chaloner and Verdinelli (1995) and Dette (1997) among others].

Note that the gradient of the function $\eta(t, \theta)$ with respect to θ is given by

$$f(x,\theta) = \left(\frac{\theta_2(e^{-\theta_1 x} - e^{-\theta_2 x}) + (\theta_1^2 x - \theta_1\theta_2 x)e^{-\theta_1 x}}{(\theta_1 - \theta_2)^2}, \right.$$
$$\left. \frac{\theta_1(e^{-\theta_1 x} - e^{-\theta_2 x}) + (\theta_1^2 x - \theta_1\theta_2 x)e^{-\theta_2 x}}{(\theta_1 - \theta_2)^2} \right)^T$$

which corresponds to the vector f of regression functions in the previous section. In order to illustrate the methodology of asymptotic optimal designs, we assume that $\theta = (1, 0.5)^T$, the correlation function of errors is $\rho(t) = e^{-\lambda|t|}$ with $\lambda = 0.5$ and the design space is given by the interval $\mathcal{X} = [0, 10]$. In general, the D-optimality criterion is typically chosen for the construction of an efficient design for the estimation of all parameters. However, in some bioavailability studies, the aim of experiments is the estimation of the area under curve

$$\text{AUC} = \int_0^\infty \eta(x, \theta)\, \mathrm{d}x.$$

For the compartmental model (14), we obtain $\text{AUC} = 1/b_2$. It can be shown that the (locally) AUC-optimal design for the model (14) minimizes the variance of the nonlinear least squares estimate for the parameter β_2. This variance is approximately proportional to $e_2^T D(\xi) e_2$, where $e_2 = (0, 1)^T$, which

corresponds to the c-optimality criterion with $c = e_2$. In addition, we study the g-optimal design for the model (14).

In Figure 6 we show the g-optimal design (left), the D-optimal design (middle) and the AUC-optimal design (right). The function $g_\xi(x)$ for these designs is depicted in Figure 7. We observe that the D-optimal design put more weight near 1 and g-optimal design has some weight at the point $t = 0$. Moreover the g-optimal design has also larger masses at the right boundary $x = 10$ of the design space.

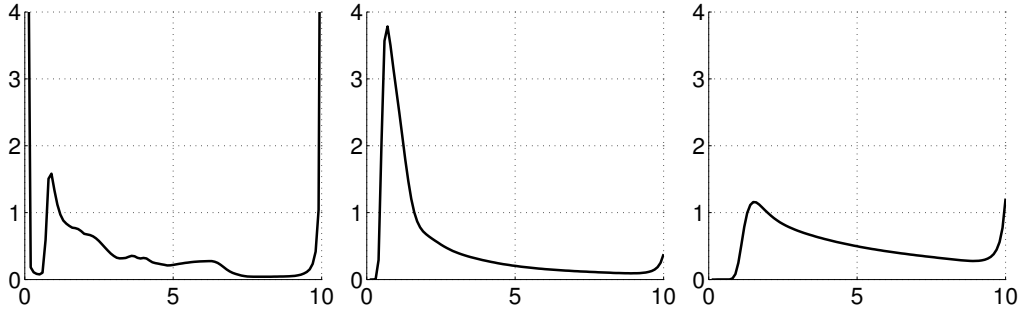

Figure 6: *The g-optimal design (left), the D-optimal design (middle) and the AUC-optimal design (right) for the nonlinear model* (14) *with exponential covariance function* $\rho(t) = e^{-\lambda|t|}$ *with* $\lambda = 0.5$.

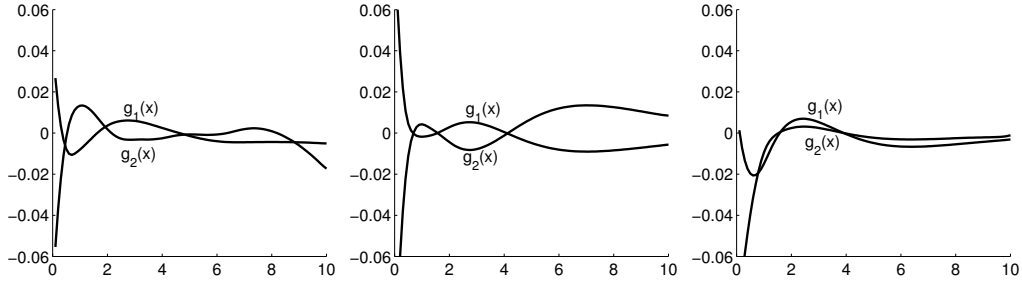

Figure 7: *The components of the vector-valued function* $g_\xi(x) = (g_1(x), g_2(x))^T$ *for the nonlinear model* (14) *with exponential correlation function* $\rho(t) = e^{-\lambda|t|}$ *(*$\lambda = 0.5$*) for the g-optimal design (left), the D-optimal design (middle) and the AUC-optimal design (right). For these designs,* $\Phi(\xi_g^*) = 0.015^2$, $\Phi(\xi_D^*) = 0.025^2$, $\Phi(\xi_{AUC}^*) = 0.023^2$.

In Table 3 we show D-, A- and g-efficiencies of the designs ξ_D^*, ξ_A^*, ξ_g^* and ξ_u. We can observe that D- and A-efficiencies of the g-optimal design are about 0.8 but the g-efficiency of ξ_D^* and ξ_A^* is 0.6 and 0.5, respectively. Note

that the AUC-efficiency of the g-optimal design is 0.98 while the g-efficiency of the AUC-optimal design is 0.65.

Table 3: *D-, A-, AUC- and g-Efficiency of the designs ξ_D^*, ξ_A^*, ξ_{AUC}^*, ξ_g^*, and ξ_u for the nonlinear model* (14) *with exponential correlation function $\rho(t) = e^{-\lambda|t|}$, where $\lambda = 0.5$.*

ξ	$\text{Eff}_D(\xi)$	$\text{Eff}_A(\xi)$	$\text{Eff}_{AUC}(\xi)$	$\text{Eff}_g(\xi)$
ξ_D^*	1	0.998	0.998	0.615
ξ_A^*	0.987	1	0.967	0.488
ξ_{AUC}^*	0.998	0.998	1	0.651
ξ_g^*	0.856	0.781	0.988	1
ξ_u	0.836	0.805	0.937	0.430

5.4. Comparison of covariance matrices of BLUE for different designs

Numerical analysis shows that the covariance matrix of the BLUE on the basis of an n-point design obtained from the (asymptotic) g-optimal design seem to be 'smaller' than the covariance matrix of the BLUE on the basis of an n-point design obtained from the (asymptotic) D-optimal design.

Consider the quadratic model with exponential correlation function $\rho(t) = e^{-|t|}$. The 8-point design obtained from the (asymptotic) g-optimal design is approximately $\xi_{8,g} = \{-1, -0.98, -0.97, -0.45, 0.45, 0.97, 0.98, 1\}$; for this design, the covariance matrix of the BLUE is

$$\text{Var}(\hat{\theta}|\xi_{8,g}) \cong \begin{pmatrix} 0.88 & 0 & -0.51 \\ 0 & 0.43 & 0 \\ -0.51 & 0 & 0.72 \end{pmatrix}.$$

The 8-point design obtained from the (asymptotic) D-optimal design is approximately $\xi_{8,D} = \{-1, -0.98, -0.97, -0.68, 0.68, 0.97, 0.98, 1\}$ and the corresponding covariance matrix of the BLUE estimate is given by

$$\text{Var}(\hat{\theta}|\xi_{8,D}) \cong \begin{pmatrix} 1.13 & 0 & -0.77 \\ 0 & 0.43 & 0 \\ -0.77 & 0 & 0.98 \end{pmatrix}.$$

Similar calculation for 12-point designs give

$$\text{Var}(\hat{\theta}|\xi_{12,g}) \cong \begin{pmatrix} 0.82 & 0 & -0.46 \\ 0 & 0.43 & 0 \\ -0.46 & 0 & 0.66 \end{pmatrix}, \ \text{Var}(\hat{\theta}|\xi_{12,D}) \cong \begin{pmatrix} 0.89 & 0 & -0.53 \\ 0 & 0.43 & 0 \\ -0.53 & 0 & 0.74 \end{pmatrix}.$$

For these covariance matrices one can easily see that $\mathrm{Var}(\hat{\theta}|\xi_{12,g}) \leq \mathrm{Var}(\hat{\theta}|\xi_{12,D})$ (with respect to the Loewner ordering) and the same relation holds approximately for the 8-point designs.

Very similar results are observed for the quadratic model with Gaussian correlation function. However, the covariance matrices of the BLUE are nearly equal for the g-optimal and D-optimal designs in the quadratic model with logarithmic correlation function; this is a consequence of the fact that the designs themselves are very similar.

For the nonlinear model (14) with parameter setting considered above, we have

$$\mathrm{Var}(\hat{\theta}|\xi_{8,g}) \cong \begin{pmatrix} 7.02 & -0.43 \\ -0.43 & 1.70 \end{pmatrix}, \quad \mathrm{Var}(\hat{\theta}|\xi_{8,D}) \cong \begin{pmatrix} 7.21 & -0.31 \\ -0.31 & 1.80 \end{pmatrix}.$$

Despite the difference $\mathrm{Var}(\hat{\theta}|\xi_{8,D}) - \mathrm{Var}(\hat{\theta}|\xi_{8,g})$ is not non-negative definite, the matrix $\mathrm{Var}(\hat{\theta}|\xi_{8,g})$ should normally be preferred to $\mathrm{Var}(\hat{\theta}|\xi_{8,D})$.

Acknowledgements. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG). Parts of this paper were written while H. Dette was visiting the Institute of Mathematical Sciences at the National University of Singapore and this author would like to thank the institute for its hospitality.

References

Atkinson, A. C., Chaloner, K., Herzberg, A. M., and Juritz, J. (1993). Optimum experimental designs for properties of a compartmental model. Biometrics, 49:325–337.

Bickel, P. J. and Herzberg, A. M. (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. Annals of Statistics, 7(1):77–95.

Biedermann, S., Dette, H., and Pepelyshev, A. (2004). Maximin optimal designs for a compartmental model. MODA 7, Advances in Model-oriented Design and Analysis, pages 41–48.

Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in non-linear situations. Biometrika, 46:77–90.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review. Statistical Science, 10(3):273–304.

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

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

Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. (2008a). Optimal designs for dose-finding studies. Journal of the American Statistical Association, 103(483):1225–1237.

Dette, H., Kunert, J., and Pepelyshev, A. (2008b). Exact optimal designs for weighted least squares analysis with correlated errors. Statistica Sinica, 18(1):135–154.

Dette, H. and O'Brien, T. (1999). Optimality criteria for regression models based on predicted variance. Biometrika, 86:93–106.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2008c). Improving updating rules in multiplicative algorithms for computing D-optimal designs. Computational Statistics and Data Analysis, 53(2):312–320.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2011). Optimal design for linear models with correlated observations. Preprint Ruhr-University Bochum.

Garrett, E. R. (1994). The Bateman function revisited: A critical reevaluation of the quantitative expressions to characterize concentrations in the one compartment body model as a function of time with first-order invasion and first-order elimination. Journal of Pharmacokinetics and Biopharmaceutics, 22:103–128.

Gibaldi, M. and Perrier, D. (1982). Parmacokinetics. Second edition. Dekker, New York.

Harman, R. and Štulajter, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. Metrika, 72(2):281–294.

Kiefer, J. (1974). General equivalence theory for optimum designs (Approximate Theory). Annals of Statistics, 2:849–879.

Kiseľák, J. and Stehlík, M. (2008). Equidistant and D-optimal designs for parameters of Ornstein-Uhlenbeck process. Statist. Probab. Lett., 78(12):1388–1396.

Näther, W. (1985). Effective Observation of Random Fields. Teubner Verlagsgesellschaft, Leipzig.

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

Sacks, J. and Ylvisaker, N. D. (1966). Designs for regression problems with correlated errors. Annals of Mathematical Statistics, 37:66–89.

Sacks, J. and Ylvisaker, N. D. (1968). Designs for regression problems with correlated errors; many parameters. Annals of Mathematical Statistics, 39:49–69.

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

Zhigljavsky, A., Dette, H., and Pepelyshev, A. (2010). A new approach to optimal design for linear models with correlated observations. Journal of the American Statistical Association, 105:1093–1103.