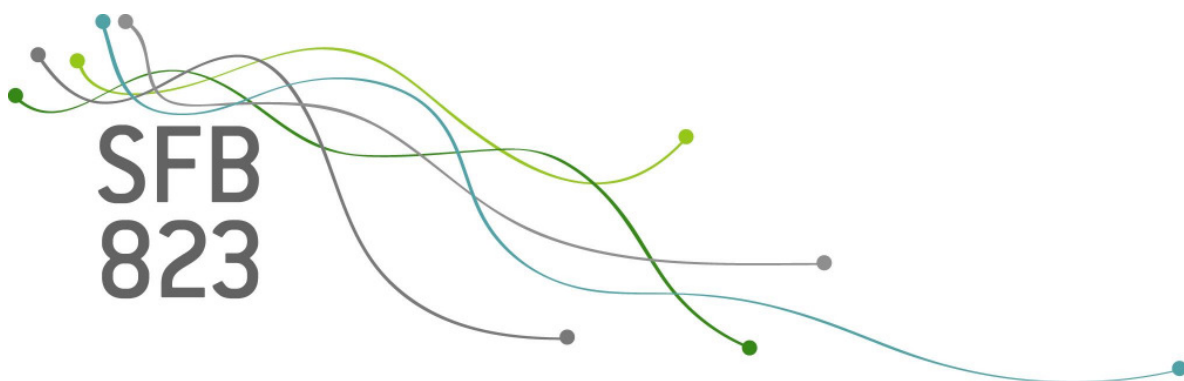# A new approach to optimal designs for models with correlated observations

Holger Dette, Andrey Pepelyshev, Anatoly Zhigljavsky

A new approach to optimal designs for models with correlated observations

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

Andrey Pepelyshev

St. Petersburg State University

Department of Mathematics

St. Petersburg, Russia

email: andrey@ap7236.spb.edu

Anatoly Zhigljavsky

Cardiff University

School of Mathematics

Cardiff CF24 4AG, UK

email: ZhigljavskyAA@cf.ac.uk

August 12, 2009

Abstract

We consider the problem of designing experiments for the estimation of the mean in the location model in the presence of correlated observations. For a fixed correlation structure approximate optimal designs are determined, and it is demonstrated that under the model assumptions made by Bickel and Herzberg (1979) for the determination of asymptotic optimal design, the designs derived in this paper converge weakly the measures obtained by these authors.

1

We also compare the approach of Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg (1979) and point out some inconsistencies of the latter. Finally, this approach is modified such that it has similar properties as the model considered by Sacks and Ylvisaker, and it is demonstrated that the resulting design problems are related to (logarithmic) potential theory.

AMS Subject Classification: 62K05

Keywords and Phrases: Optimal design, correlated observation, positive definite functions, logarithmic potentials

1 Introduction

Consider the common linear regression model

$$y(t) = \theta_1 f_1(t) + \ldots + \theta_p f_p(t) + \varepsilon(t) , \tag{1.1}$$

where $f_1(t), \ldots, f_p(t)$ are given functions, $\varepsilon(t)$ denotes a random error process, $\theta_1, \ldots, \theta_p$ are unknown parameters and t is the explanatory variable. We assume that N observations, say $y_1, \ldots, y_N$, can be taken at experimental conditions $-T \leq t_1 \leq \ldots \leq t_N \leq T$ to estimate the parameters in the linear regression model (1.1). If an appropriate estimate $\hat{\theta}$ has been chosen, the quality of the statistical analysis can be further improved by choosing an appropriate design for the experiment. In particular an optimal design minimizes a functional of the variance-covariance matrix of the estimate, where the functional should reflect certain aspects of the goal of the experiment. In contrast to the case of uncorrelated errors, where numerous results and a rather complete theory are available [see for example the monographs of Fedorov (1972), Silvey (1980), Pázman (1986), Atkinson and Donev (1992) or Pukelsheim (1993)], the construction of optimal designs for dependent observations is intrinsically more difficult. On the other hand this problem is of particular interest, because in many applications the variable t in the regression model (1.1) represents the time and all observations correspond to one subject. This deficit can be explained by the fact that

optimal experimental designs for regression models with correlated observations have an extremely complicated structure and are very difficult to find even in simple cases. Some exact optimal design problems were considered in Näther (1985, Ch. 4), see also Pázman and Müller (2001), Müller and Pázman (2003).

Because explicit solutions of the optimal design problem for correlated observations are rarely available several authors have proposed to determine optimal designs based on asymptotic arguments [see for example Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979) and Näther (1985)]. Roughly speaking there exist two proposals to embed the optimal design problem for regression models with correlated observations in an asymptotic optimal design problem. The first one is due to Sacks and Ylvisaker (1966, 1968), who assumed that the covariance structure of the error process $\varepsilon(t)$ is fixed and that the number of design points tends to infinity. Alternatively Bickel and Herzberg (1979) and Bickel, Herzberg and Schilling (1981) considered a different model, where the correlation function depends on a number of observations.

The present paper has several purposes. After a brief introduction into the terminology, we present in Section 3 several new exact optimal approximate designs in the location model, which extend the results of Boltze and Näther (1982) and Näther (1985). We also demonstrate how these results are related to the designs derived in Bickel and Herzberg (1979) studying the designs derived in this paper under the assumptions made by these authors. In Section 4 we compare the method of Bickel and Herzberg (1979) with the approach suggested by Sacks and Ylvisaker (1966, 1968), who considered a model, where the covariance matrix of the (weighted) least squares estimate does not converge to 0 with an increasing sample size. In particular, we show an inconsistency of the model proposed by Bickel and Herzberg (1979): the covariance between observations at consecutive time points remains constant although for an increasing sample size the explanatory variables are arbitrary close. As a consequence, the covariance of the ordinary least squares estimate based on the optimal design vanishes asymptotically, but it does not converge to the covariance matrix corresponding

3

to the uncorrelated case, despite the fact that the correlation structure approximates the case of uncorrelated observations. Finally, a new approach for constructing optimal designs for correlated data is introduced, which could be interpreted as compromise between the methods proposed by Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg (1979). On the one hand, this method is based on a similar argument as used by the latter authors; on the other hand, for an increasing sample size the correlation structure does not approximate the uncorrelated case.

2 Preliminaries

Consider the linear regression model (1.1), where $\varepsilon(t)$ is a stationary process with

$$\mathbf{E}\varepsilon(t) = 0, \quad \mathbf{E}\varepsilon(t)\varepsilon(s) = \sigma^2 K(t, s), \tag{2.1}$$

Following Bickel and Herzberg (1979) we assume that $\varepsilon(t) = \varepsilon^{(1)}(t) + \varepsilon^{(2)}(t)$, where $\varepsilon^{(1)}(t)$ denotes a stationary process with correlation function $\rho(t)$ and $\varepsilon^{(2)}(t)$ is white noise. Consequently, we obtain

$$K(t, s) = \gamma\rho(t - s) + (1 - \gamma)\delta_{t,s}, \tag{2.2}$$

where δ denotes Kronecker's symbol. If N observations, say $y = (y_1, \dots, y_N)^T$ are available at experimental conditions $t_1, \dots, t_N$ and some knowledge of the correlation function is available, the vector of parameters can be estimated by the weighted least squares method, i.e. $\hat{\theta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y$ with $X^T = (f_i(t_j))_{i=1,\dots,p}^{j=1,\dots,N}$, and the variance-covariance matrix of this estimate is given by

$$\mathbf{D}(\hat{\theta}) = \sigma^2(X^T\Sigma^{-1}X)^{-1} \tag{2.3}$$

with $\Sigma = (K(t_i, t_j))_{i,j}$, $i, j = 1, \dots, N$. If the the correlation structure of the process $\varepsilon(t)$ is not known, one usually uses the ordinary least squares estimate $\tilde{\theta} = (X^TX)^{-1}X^Ty$, which has covariance matrix

$$\mathbf{D}(\tilde{\theta}) = \sigma^2(X^TX)^{-1}X^T\Sigma X(X^TX)^{-1}. \tag{2.4}$$

4

An experimental design $\xi_N = \{t_1, \ldots, t_N\}$ is a vector of N points in the interval $[-T, T]$, 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. Because even in simple models exact optimal designs are difficult to find, most authors usually use asymptotic arguments to determine efficient designs for the estimation of the model parameters.

Following Näther (1985a, Chapter 4) we assume that the design points $\{t_1, \ldots, t_N\}$ are generated by the quantiles of a distribution function, that is

$$t_{iN} = a\left((i-1)/(N-1)\right), \quad i = 1, \ldots, N, \tag{2.5}$$

where the function

$$a : [0, 1] \to [-T, T] \tag{2.6}$$

is the inverse of a distribution function. If ξ_N denotes a design with N points and corresponding quantile function a, the covariance matrix of the estimate $\tilde{\theta} = \tilde{\theta}_{\xi_N}$ given in (2.4) can be written as

$$\mathbf{D}(\tilde{\theta}) = \sigma^2 D(\xi), \tag{2.7}$$

where $\xi = \xi_N$,

$$D(\xi) = W^{-1}(\xi)R(\xi)W^{-1}(\xi), \quad W(\xi) = \int f(u)f^T(u)\xi(du),$$

$$R(\xi) = \int\int K(u, v)f(u)f^T(v)\xi(du)\xi(dv).$$

The matrix $D(\xi)$ is called the covariance matrix of the design ξ and can be defined for any distribution on the interval $[-T, T]$. Following Kiefer (1974) we call any probability

5

measures on $[-T, T]$ an approximate designs and an (approximate) optimal design minimizes a functional of the covariance matrix $D(\xi)$ over a class of all approximate designs.

Note that in general the function $D(\xi)$ is not convex (with respect to the Loewner ordering) on the space of all approximate designs. On the other hand in the location model

$$y(t) = \theta + \varepsilon(t) \tag{2.8}$$

we have

$$D(\xi) = \iint K(u, v)\xi(du)\xi(dv), \tag{2.9}$$

which obviously defines a convex function on the set of probability measures on the interval $[-T, T]$. For this reason most of the literature discussing optimal design problems for least squares estimation in the presences of correlated observations considers the location model, which corresponds to the estimation of the mean of a stationary process [see for example Boltze and Näther (1982), Näther (1985a, 1985b)]. Throughout this paper we will follow this line an restrict our attention to the model (2.8). The following lemma states that in this case the optimality criterion (2.9) is convex or even strictly convex [see also Näther, 1985a, Th 3.4.1)].

We shall use the following definition. A function $f : \mathbb{R} \to \mathbb{R}$ is positive definite if for any real numbers $x_1, \ldots, x_n$ the $n \times n$ matrix H with entries $h_{ij} = f(x_i - x_j)$ is a non-negative definite matrix; correspondingly, the function f is strictly positive definite if the matrix H is positive definite for all $x_1 < \ldots < x_n$.

Lemma 1 *The functional $D(\cdot)$ defined in (2.9) is convex. Moreover, if $K(\cdot, \cdot)$ is strictly positive definite, then $D(\cdot)$ is strictly convex. That is,*

$$D(\alpha\xi_2 + (1 - \alpha)\xi_1) < \alpha D(\xi_2) + (1 - \alpha)D(\xi_1)$$

for all $0 < \alpha < 1$ and any two measures ξ_1 and ξ_2 on $[-T, T]$ such that $\xi_2 - \xi_1$ is a non-zero measure.

6

Proof. We have

$$D(\alpha\xi_2 + (1-\alpha)\xi_1) = \iint K(u,v)[\alpha\xi_2(du) + (1-\alpha)\xi_1(du)][\alpha\xi_2(dv) + (1-\alpha)\xi_1(dv)]$$

$$= (1-\alpha)^2 \iint K(u,v)\xi_1(du)\xi_1(dv) + \alpha^2 \iint K(u,v)\xi_2(du)\xi_2(dv)$$

$$+ 2\alpha(1-\alpha) \iint K(u,v)\xi_1(du)\xi_2(dv)$$

$$= \alpha^2 D(\xi_2) + (1-\alpha)^2 D(\xi_1) + 2\alpha(1-\alpha) \iint K(u,v)\xi_1(du)\xi_2(dv)$$

$$= \alpha D(\xi_2) + (1-\alpha)D(\xi_1) - \alpha(1-\alpha)A\,,$$

where

$$A = \iint K(u,v)[\xi_2(du)\xi_2(dv) + \xi_1(du)\xi_1(dv) - 2\xi_2(du)\xi_1(dv)] = \iint K(u,v)\zeta(du)\zeta(dv)$$

and $\zeta(du) = \xi_2(du) - \xi_1(du)$. Since the correlation function $K(u,v)$ is positive definite, in view of the Bochner-Khintchine theorem [Feller (1966), Ch. 19.2], we have $A \geq 0$. If $K(\cdot,\cdot)$ is strictly positive definite, we have $A > 0$ whenever ζ is not trivial. Therefore the functional $D(\cdot)$ is strictly convex. $\qquad\square$

In the following lemma we calculate the directional derivative of the functional $D(\cdot)$.

Lemma 2 *If $\xi_\alpha = (1-\alpha)\xi + \alpha\xi_0$ and $D(\cdot)$ is defined in (2.9), we have*

$$\frac{\partial}{\partial\alpha}D(\xi_\alpha)\Big|_{\alpha=0} = 2\left(\int \phi(v,\xi)\xi_0(dv) - D(\xi)\right)$$

where

$$\phi(t,\xi) = \int K(t,u)\xi(du).$$

Proof. Taking into account the proof of Lemma 1, it follows

$$\frac{\partial}{\partial\alpha}D(\xi_\alpha)\Big|_{\alpha=0} = \frac{\partial}{\partial\alpha}\left((1-\alpha)^2 D(\xi) + \alpha^2 D(\xi_0) + 2\alpha(1-\alpha)\iint K(u,v)\xi(du)\xi_0(dv)\right)\Big|_{\alpha=0}$$

$$= 2\left(\int \phi(v,\xi)\xi_0(dv) - D(\xi)\right)$$

□

Using Lemmas 1 and 2 we obtain the following equivalence theorem, which characterizes the optimality of a design for the location model.

Theorem 1

(i) *A design ξ^* minimizes the functional $D(\cdot)$ defined in (2.9) if and only if*

$$\min_{t \in [-T,T]} \phi(t, \xi^*) \geq D(\xi^*). \tag{2.10}$$

(ii) *In particular, a design ξ^* is optimal if the function $\phi(\cdot, \xi^*)$ is constant, where the constant is given by $D(\xi^*)$.*

Proof. (i) Using the convexity of the functional $D(\cdot)$ and Lemma 2, the necessary and sufficient condition for an extremum yields

$$\min_{\xi_0} \int \phi(v, \xi^*) \xi_0(dv) \geq D(\xi^*).$$

Note that

$$\int \phi(v, \xi^*) \zeta(dv) = \min_{\xi_0} \int \phi(v, \xi^*) \xi_0(dv)$$

for any design ζ satisfying

$$\mathrm{supp}(\zeta) \subset \{t : \phi(t, \xi^*) = \min_{v} \phi(v, \xi^*)\}.$$

Consequently, the necessary and sufficient condition of extremum becomes $\min_t \phi(t, \xi^*) \geq D(\xi^*)$, which is exactly (2.10). The assertion (ii) obviously follows from (i). □

The final result of this section shows that an optimal design does not depend on the value γ in the correlation function (2.2). For this reason we assume without loss of generality $\gamma = 1$ throughout the remaining discussion in this paper.

8

Lemma 3 *The optimal design ξ^* does not depend on the constant $\gamma \in (0, 1]$ and $D(\xi^*) = 1 - \gamma + \gamma \mathcal{D}(\xi^*)$, where*

$$\mathcal{D}(\xi) = \int\int \rho(t-s)\xi(du)\xi(dv).$$

Proof. For the covariance function (2.2) we have

$$
\begin{aligned}
D(\xi) &= \int\int K(u,v)\xi(du)\xi(dv) = \int\int (\gamma\rho(t-s) + (1-\gamma)\delta_{t,s})\xi(du)\xi(dv) \\
&= 1 - \gamma + \gamma \int\int \rho(t-s)\xi(du)\xi(dv) = 1 - \gamma + \gamma \mathcal{D}(\xi).
\end{aligned}
$$

$\square$

3 Optimal designs for particular correlation functions

In this section, we consider several types of correlation functions defined in (2.1) and (2.2). Without loss of generality, we assume that $T = 1$ so that the design space is given by the interval $[-T, T] = [-1, 1]$. We begin our investigations with the exponential correlation function, that is

$$\rho(t) = e^{-\lambda|t|}, \tag{3.1}$$

where $\lambda > 0$ is fixed.

Theorem 2 *For the location model (2.8) with correlation function (3.1) the optimal design ξ^* is a mixture of the continuous uniform measure on the interval $[-1, 1]$ and a two-point discrete measure supported on $\{-1, 1\}$. In other words: ξ^* has the density*

$$p^*(u) = w^*\left(\frac{1}{2}\delta_1(u) + \frac{1}{2}\delta_{-1}(u)\right) + (1 - w^*)\frac{1}{2}\mathbf{1}_{[-1,1]}(u), \tag{3.2}$$

where $w^* = 1/(1 + \lambda)$, $\delta_x(\cdot)$ denotes the Dirac measure concentrated at the point x and $\mathbf{1}_A(\cdot)$ is the indicator function of a set A. Moreover, the function $\phi(t, \xi^*)$ defined in (2.10) is constant and $\mathcal{D}(\xi^*) = 1/(1 + \lambda)$.

Proof. Direct calculations show that

$$\phi(t, \xi^*) = w^* \frac{1}{2} \left(e^{-\lambda(1-t)} + e^{-\lambda(1+t)} \right) + \frac{1}{2}(1-w^*) \frac{2 - e^{-\lambda(1-t)} - e^{-\lambda(1+t)}}{\lambda} = \frac{1}{1+\lambda},$$

which is a constant function. The statement of the theorem now follows from Theorem 1, part (ii). □

Note that the result of Theorem 2 is known in literature [see Boltze and Näther (1982)] but presented here for the sake of completeness. Next we consider the triangular correlation function defined by

$$\rho(t) = \max\{0, 1 - \lambda|t|\}. \tag{3.3}$$

In the particular case $\lambda = 1$ the optimal design can be obtained from Example 1 in Näther (1985b). The following Theorem extends these results and specifies the optimal designs for all $\lambda > 0$.

Theorem 3 *Consider the location model (2.8) with correlation function (3.3).*

(a) For $\lambda \in \mathbb{N} = \{1, 2, \ldots\}$, the optimal design is a discrete uniform measure supported at $1 + 2\lambda$ equidistant points, $t_j = j/\lambda - 1$, $j = 0, 1, \ldots, 2\lambda$. For this design, $\mathcal{D}(\xi^*) = 1/(1+2\lambda)$.

(b) For any $\lambda > 0$, the optimal design ξ^* is a discrete symmetric measure supported at $2n$ points $\pm t_1, \pm t_2, \ldots, \pm t_n$ with weights $w_1, \ldots, w_n$ at $t_1, \ldots, t_n$, where $n = \lceil 2\lambda \rceil$,

$$(w_1, \ldots, w_n) = \frac{1}{n(n+1)} (\lceil n/2 \rceil, \ldots, 3, n-2, 2, n-1, 1, n).$$

Here $t_1, \ldots, t_n$ denote the ordered quantities $|u_1|, \ldots, |u_n|$, where $u_j = -1 + j/\lambda$, $j = 1, \ldots, n-1$, $u_n = 1$. Moreover,

$$\mathcal{D}(\xi^*) = \frac{2\lambda}{n(n+1)} .$$

Proof. For a proof of the statement in part (a) we fix a value $t \in [-1, 1]$. If $t = t_j$ for some $j \in \{0, 1, \ldots, 2\lambda\}$ then $\rho(t - t_j) = 1$ and $\rho(t - t_i) = 0$ for all $i \neq j$. If $t \in (t_{j-1}, t_j)$ for some $j \in \{1, \ldots, 2\lambda\}$ then

$$\rho(t - t_{j-1}) + \rho(t - t_j) = 1 - \lambda(t - ((j-1)/\lambda - 1)) + 1 - \lambda((j/\lambda - 1) - t) = 1$$

and $\rho(t - t_i) = 0$ for all $i > j$ and all $i < j - 1$. Therefore, for any $t \in [-1, 1]$ we obtain

$$\sum_{j=0}^{2\lambda} \max\{0, 1 - \lambda|t - (j/\lambda - 1)|\} = 1 .$$

This implies

$$\phi(t, \xi^*) = \int \rho(t - v)\xi^*(dv) = \frac{1}{1 + 2\lambda} \sum_{j=0}^{2\lambda} \max\{0, 1 - \lambda|t - (j/\lambda - 1)|\} = 1/(1 + 2\lambda)$$

and

$$\mathcal{D}(\xi^*) = \iint \rho(u - v)\xi^*(du)\xi^*(dv) = 1/(1 + 2\lambda).$$

The statement now follows from Theorem 2.

For a proof of part (b) we evaluate the function $\phi(t, \xi^*)$ on the different intervals (t_{j-1}, t_j). First we consider the case where $t > t_{n-1}$, for which we have

$$
\begin{aligned}
\phi(t, \xi^*) &= \sum_{i=1}^{n} w_i \rho(t - t_i) + \sum_{i=1}^{n} w_i \rho(t + t_i) = \sum_{i=n-2}^{n} w_i \rho(t - t_i) = \\
&= \frac{1}{n(n+1)} \left((n-1)\lambda(t - 1 + 1/\lambda) + \lambda(t + 1 - (n-1)/\lambda) + n\lambda(1 - t)\right) \\
&= \frac{2\lambda}{n(n+1)} .
\end{aligned}
$$

If $t_{n-2} < t < t_{n-1}$ it follows

$$
\begin{aligned}
\phi(t, \xi^*) &= w_{n-3}\rho(t - t_{n-3}) + w_{n-2}\rho(t - t_{n-2}) + \ldots + w_n\rho(t - t_n) = \\
&= \frac{1}{n(n+1)} \Big(2\lambda(t + 1 - (n-2)/\lambda) \\
&\quad + (n-1)\lambda(t - 1 + 1/\lambda) + \lambda(-1 + (n-1)/\lambda - t) + n\lambda(1 - t))\Big) \\
&= \frac{2\lambda}{n(n+1)} ,
\end{aligned}
$$

11

while for $t_{n-3} < t < t_{n-2}$ we have

$$\begin{aligned}
\phi(t, \xi^*) &= w_{n-4}\rho(t - t_{n-4}) + w_{n-3}\rho(t - t_{n-3}) + \ldots + w_{n-1}\rho(t - t_{n-1}) = \\
&= \frac{1}{n(n+1)}\Big((n-2)\lambda(t - 1 + 2/\lambda) + 2\lambda(t + 1 - (n-2)/\lambda) \\
&\quad + (n-1)\lambda(-1 + 1/\lambda - t) + \lambda(-1 + (n-1)/\lambda - t))\Big) \\
&= \frac{2\lambda}{n(n+1)}.
\end{aligned}$$

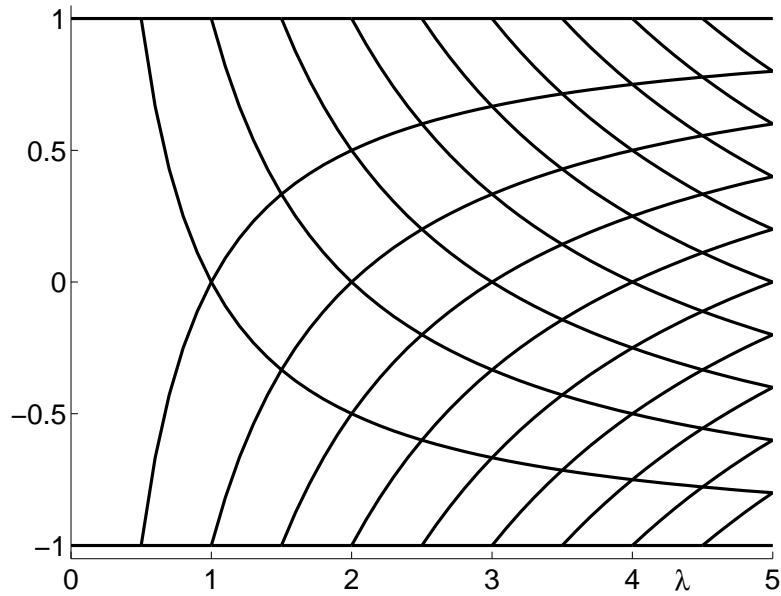Other cases are considered in similar way and the assertion follows from Theorem 2. $\quad\square$

Figure 1: *Support points of optimal designs in the location model with triangular correlation function (3.3) for different values of λ. The corresponding weights are given in Theorem 3.*

Example 1 For the triangular correlation function (3.3) we obtain the following optimal designs for the location model (2.8)

- If $\lambda \in [0, 0.5]$, the optimal design is supported at the points -1 and 1 with weights $1/2$.

- If $\lambda \in [0.5, 1]$, the optimal design is given by

$$\begin{pmatrix} -1 & 1-1/\lambda & 1/\lambda-1 & 1 \\ 1/3 & 1/6 & 1/6 & 1/3 \end{pmatrix}.$$

- If $\lambda \in [1, 1.5]$, the optimal design is given by

$$\begin{pmatrix} -1 & 1-2/\lambda & -1+1/\lambda & 1-1/\lambda & -1+2/\lambda & 1 \\ 1/4 & 1/12 & 1/6 & 1/6 & 1/12 & 1/4 \end{pmatrix}.$$

- If $\lambda \in [1.5, 2]$, the optimal design is given by

$$\begin{pmatrix} -1 & 1-3/\lambda & 1-2/\lambda & -1+1/\lambda & 1-1/\lambda & -1+2/\lambda & -1+3/\lambda & 1 \\ 0.4/2 & 0.1/2 & 0.2/2 & 0.3/2 & 0.3/2 & 0.2/2 & 0.1/2 & 0.4/2 \end{pmatrix}.$$

For larger values of λ the support points of the optimal designs for the location model (2.8) with correlation function (3.3) are displayed in Figure 1

Remark 1 It might be of interest to investigate the relation between the designs derived in Theorem 1 and 2 and the designs obtained by the approach proposed by Bickel and Herzberg (1979). These authors suggested a correlation structure depending on the sample size N, that is

$$\rho_N(t) = \rho(Nt), \tag{3.4}$$

where the function $\rho(t)$ satisfies $\int |\rho(t)|dt < \infty$ (note that this condition corresponds to the case of short range dependence). It can be shown that for the location model (2.8) the optimality criterion proposed by Bickel and Herzberg (1979) is asymptotically (as $N \to \infty$) given by

$$\mathcal{D}_{BH}(\xi) = 1 + 2 \int Q(1/p(t))p(t)dt, \tag{3.5}$$

where $1/p$ denotes the density of the quantile function a, the function Q is defined by

$$Q(u) = \sum_{j=1}^{\infty} \rho(ju)$$

13

[see Theorem 2.1 in Bickel and Herzberg (1979)]. For this criterion the asymptotic optimal design ξ^* on the interval $[-1, 1]$ is uniquely determined and has absolute continuous density i.e. $p^*(t) = \frac{1}{2}\mathbf{1}_{[-1,1]}(t)$.

We now investigate the asymptotic behavior of the design determined in Theorem 1 for the correlation function $\rho_N(t) = e^{-\lambda N|t|}$. In this case the optimal design ξ_N^* is given by (3.2) with

$$w^* = w_N^* = \frac{1}{N\lambda + 1}.$$

Because $w^* \to 0$ as $N \to \infty$ it follows that the sequence of the optimal designs $(\xi_N^*)_{N \in \mathbb{N}}$ converges weakly to the optimal design p^* obtained by the approach of Bickel and Herzberg (1979). Similarly, if $\rho_N(t) = \max\{0, 1 - N\lambda|t|\}$, it follows from Theorem 4 that the sequence of optimal design $(\xi_N^*)_{N \in \mathbb{N}}$ converges weakly to the design p^*.

For most correlation functions the optimal designs have to be determined numerically even in the case of the location model. We conclude this section presenting several new numerical results in this context. The next correlation function considered in our study corresponds to the Gaussian distribution and is defined by

$$\rho(t) = e^{-\lambda t^2}. \tag{3.6}$$

Some optimal designs for the location model with correlation function (3.6) are given in Table 1 for selected values of $\lambda \in [0, 8.5]$. The support point of the optimal design for larger values of λ are depicted in Figure 2. From our numerical results we conclude for the correlation structure (3.6) that the optimal design for the location model is a discrete measure, where the number of support points increases with λ. It is also worthwhile to mention that for this model the function $\phi(t, \xi^*)$ defined in (2.10) is not constant, and as a consequence, the second part of Theorem 1 is not applicable.

We conclude this section with two examples, where the optimal design is a mixture between a discrete and an absolute continuous measure with a non constant density. The first example

14

Table 1: *Optimal designs for the location model with correlation function (3.6) for different values of λ.*

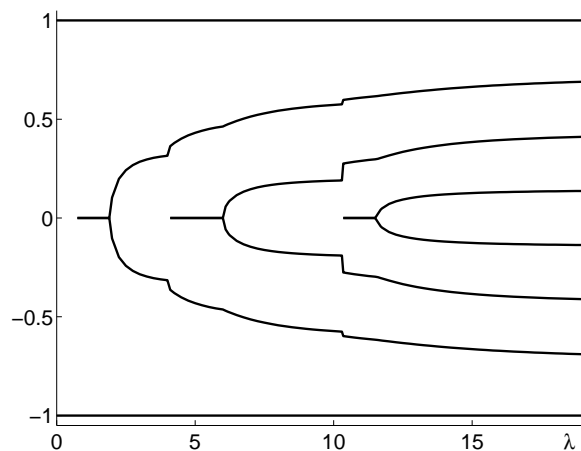| λ	t_1	t_2	t_3	w_1	w_2	w_3
0.05	±1			0.5		
0.6	±1			0.5		
0.7	±1	0		0.4685	0.063	
1.9	±1	0		0.354	0.292	
2.0	±1	±0.104		0.348	0.152	
3.7	±1	±0.309		0.282	0.218	
3.9	±1	±0.336	0	0.277	0.202	0.043
6.0	±1	±0.463	0	0.237	0.179	0.169
6.1	±1	±0.469	±0.058	0.235	0.176	0.089
8.5	±1	±0.553	±0.178	0.207	0.154	0.139

Figure 2: *Support points of the optimal design for the location model with correlation function (3.6) for different values of λ.*

is obtained for the correlation function

$$\rho(t) = \frac{1}{\sqrt{1 + \lambda|t|}}, \tag{3.7}$$

15

where $\lambda > 0$. In this case we obtain by an extensive numerical study that the optimal design ξ^* is given by

$$p_{\xi^*}(u) = w^* \left(\frac{1}{2}\delta_1(u) + \frac{1}{2}\delta_{-1}(u) \right) + (1 - w^*)p^*(u), \tag{3.8}$$

where $w^* \in [0, 1]$ denotes a weight and $p^*(u)$ is a density which depends on λ. For the selected values the optimal weights and corresponding densities are displayed in Table 3 and the left part of Figure 3, respectively. It is also worthwhile to mention that in this case the function $\phi^*(t, \xi^*)$ defined in (2.10) is constant.

Table 2: *The weight of the optimal design* (3.8) *in the location model with correlation function* (3.7) *for different values of* λ.

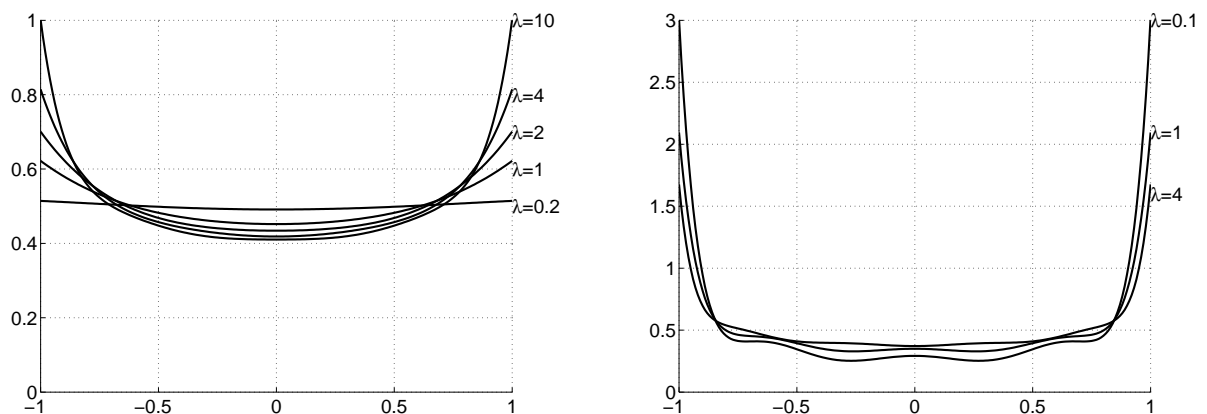| λ	0.2	1	2	4	10
w^*	0.796	0.516	0.392	0.283	0.173

Figure 3: *The density* p^* *of the optimal design* (3.8) *for the location model for different values of* $\lambda > 0$. *Left part: correlation structure* (3.7)*; right part: correlation structure* (3.9).

The last example considers the correlation function

$$\rho(t) = \frac{1}{1 + \lambda|t|^{0.5}}, \tag{3.9}$$

16

for which our numerical results show that the optimal designs for the location model with this correlation structure are also of the form (3.8), where the optimal weight w^* and the optimal density are displayed in Table 3 and the right part of Figure 3, respectively.

Table 3: *The weight of the optimal design* (3.8) *in the location model with correlation function* (3.9) *for different values of* λ.

λ	0.1	0.5	1	2	4
w^*	0.201	0.151	0.118	0.082	0.056

4 A new approach to the design experiments for correlated observations

In this Section we briefly describe and compare several aspects of the approach Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg and Herzberg (1979) in order to develop an alternative method for the construction of optimal designs for dependent data. In the method proposed by Sacks and Ylvisaker (1966, 1968) the design space is fixed, the number of design points in this set converges to infinity and the weighted least squares estimate $\hat{\theta}$ is investigated. As a consequence, the corresponding asymptotic optimal designs depend only on the behavior of the correlation function in a neighborhood of the point 0 and the variance of the (weighted) least squares $\hat{\theta}$ does not converge to 0 as $N \to \infty$.

In contrast to Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979) considered the ordinary least squares estimate, say $\tilde{\theta}$, and assumed that the correlation function depends on N according to $\rho_N(t) = \rho(Nt)$, see (3.4). An alternative interpretation of this model is that the correlation function is fixed but the design interval expands proportionally to the number of observation points. In this model the correlation between two consecutive time points t_i and t_{i+1} is essentially constant, i.e. $\rho_N(t_{i+1} - t_i) \approx \rho(a'(t_i))$, and the variance of the

17

ordinary least squares estimate converges to 0 with a rate depending on the function ρ. We illustrate this effect in the following example.

Example 2 Consider the function $\rho(t) = e^{-\lambda|t|}$ and assume that $\gamma = 1$ in the correlation function (2.2). The variance of the ordinary least squares estimate obtained from the optimal design ξ^* provided by Theorem 2 is given by

$$\mathbf{D}(\tilde{\theta}) = \sigma^2 \mathcal{D}(\xi^*) = \frac{\sigma^2}{1 + \lambda} \,.$$

The variance of the weighted least squares estimate for the uniform design (which is an asymptotically optimal design for this estimator) is exactly the same: $\mathbf{D}(\hat{\theta}) = \sigma^2/(1 + \lambda)$. Note that both variances do not converge to 0 as N increases, unlike in the case of i.i.d observations where the variance is σ^2/N. In other words: the presence of correlations between observations significantly increases the variance of any least squares estimate for the parameter θ.

On the other hand, it follows from (3.5) and the representation $Q(t) = 1/(e^{\lambda t} - 1)$ that in the model considered by Bickel and Herzberg (1979) the variance of the ordinary least squares estimate for the parameter θ is asymptotically given by

$$\frac{\sigma^2}{N}\left(1 + \frac{2}{e^{2\lambda} - 1}\right) + o\left(\frac{1}{N}\right), \quad N \to \infty \,. \tag{4.1}$$

Note that the dominating term in this expression differs from the rate σ^2/N, although the correlation function ρ_N converges to the Dirac measure at the point 0, which corresponds to the case of uncorrelated observations. For other correlation functions, for example $\rho(t) = \max\{0, 1 - \lambda|t|\}$, a similar observation can be made.

Consider now the case of long-range dependence in the error process, i.e. $\rho_\alpha(t) \sim 1/|t|^\alpha$ as $t \to \infty$ where $\alpha \in (0, 1)$. It was shown by Dette et al. (2009) that the asymptotic optimal design in the location model based on the approach of Bickel and Herzberg (1979) minimizes the expression

$$\mathcal{D}_\alpha(\xi) = \int Q_\alpha(1/p(t))p(t)dt \,, \tag{4.2}$$

18

where

$$Q_\alpha(u) = \frac{1}{N^\alpha} \sum_{j=1}^{\infty} \rho_\alpha(ju).$$

For the correlation functions

$$\rho_\alpha^{(1)}(t) = \frac{1}{(1+|t|^2)^{\alpha/2}}, \quad \rho_\alpha^{(2)}(t) = \frac{1}{1+|t|^\alpha}, \quad \rho_\alpha^{(3)}(t) = \frac{1}{(1+|t|)^\alpha}, \tag{4.3}$$

it can be shown that $Q_\alpha(t) = 1/((1-\alpha)|t|^\alpha)$, and we obtain for the asymptotic variance of the ordinary least squares estimate the expression

$$\frac{\sigma^2}{N^\alpha} \frac{2^\alpha}{1-\alpha} + o\left(\frac{1}{N^\alpha}\right), \quad N \to \infty.$$

Again the dominating term in this variance is different from the variance σ^2/N, although the correlation functions $\rho_N^{(j)}(t) = \rho_\alpha^{(j)}(Nt)$ in (4.3) approximate the Dirac measure at the point 0.

The computation of the asymptotic variances in Example 2 illuminates the following general theoretical results:

- In the case of correlated observations, the variance of any least squares estimator does not converge to zero as $N \to \infty$.

- In the approach of Bickel and Herzberg (1979) (with $\rho_N(t) = \rho(Nt)$), the variance of any least squares estimates converges to zero as $N \to \infty$.

- In the approach of Bickel and Herzberg (1979) the variance of the ordinary least squares estimate has a different first order asymptotic behavior as the variance of the ordinary least squares estimate for the case of uncorrelated observations, despite the fact that the correlation function $\rho_N(t) = \rho(Nt)$ degenerates as $N \to \infty$.

Therefore the natural question arises, if it is possible to develop an alternative concept for the construction of optimal designs for correlated observations, which on the one hand is based the normalization $\rho_N(t) = \rho(Nt)$ used by Bickel and Herzberg (1979) and on the other hand yields a variance of the ordinary least squares estimate, which is of precise order $O(1)$.

The answer to this question is affirmative if we allow ourselves to vary the variance of individual observations as N changes. To be precise let $c(t, s) = \sigma^2 \rho(t - s)$ be the covariance function between observations at points t and s, then assume that not only $\rho(\cdot)$ but also σ^2 may depend on N. In order to be consistent with the model discussed in Bickel and Herzberg (1979), we consider sequences of correlation functions satisfying

$$c_N(t, s) = \sigma_N^2 \rho_N(t - s) , \tag{4.4}$$

where

$$\rho_N(t) = \rho(a_n t), \quad \sigma_N^2 = a_n^\alpha \tau^2, \tag{4.5}$$

$\tau > 0$ and $0 < \alpha \leq 1$ is a constant depending on the asymptotic behavior of the function $\rho(t)$ as $t \to \infty$. The choice $\sigma_N^2 = N^\alpha \tau^2$ yields that the variance of the ordinary least squares estimate is of order $O(1)$. Note that in the case of short-range dependence one has to use $\alpha = 1$. In the case of long-range dependence with $\rho(t) = L(t)/t^\kappa$, where $L(t)$ is a slowly varying function at $t \to \infty$ (Seneta, 1976) one has to use $\alpha = \kappa$ in order to obtain a variance of the ordinary least squares estimate, which is of order $O(1)$.

Example 3 In the situation considered in the first part of Example 1 we have $\rho_N(t) = e^{-N\lambda|t|}$, and with the choice $\sigma_N^2 = N\tau^2$ the asymptotic expression in (4.1) changes to

$$\tau^2 \frac{2}{e^{\lambda/2} - 1} + O\left(\frac{1}{N}\right), \quad N \to \infty .$$

Lemma 4 *Assume that the function $\rho(\cdot)$ has one of the forms (4.3) with $0 < \alpha < 1$ and the covariance function $c(t, s) = c_N(t, s)$ is of the form (4.4) and (4.5), where $\{a_N\}_{N \in \mathbb{N}}$ denotes a sequence of positive numbers satisfying $a_N \to \infty$ as $N \to \infty$. If the sequence of designs $\{\xi_N\}_{N \in \mathbb{N}}$ converges weakly to an asymptotic design ξ, then the variances of the ordinary least squares estimate $\tilde{\theta}$ for the location model is given by*

$$\mathbf{D}_N(\tilde{\theta}) = \iint c_N(u, v) \xi_N(du) \xi_N(dv)$$

20

and converges to $\tau^2 \mathcal{D}_\alpha(\xi)$ as $N \to \infty$, where

$$\mathcal{D}_\alpha(\xi) = \int\int r_\alpha(u - v)\xi(du)\xi(dv) \tag{4.6}$$

and $r_\alpha(t) = 1/|t|^\alpha$.

Proof. Consider the correlation function $\rho(t) = 1/(1 + |t|)^\alpha$. Then

$$\sigma_N^2 \rho_N(t) = a_N^\alpha \tau^2 \frac{1}{(1 + |a_N t|)^\alpha} = \tau^2 \frac{1}{(1/a_N + |t|)^\alpha}$$

which yields the statement of the lemma. The remaining cases in (4.3) can be treated similarly and the details are omitted for the sake of brevity. $\square$

Remark 2

(a) As a particular case of the sequence $\{a_N\}_{N \in \mathbb{N}}$ in Lemma 4, we can take $a_N = N^\beta$ with any $\beta > 0$.

(b) Note that the statement of Lemma 4 can be generalized to cover the more general situation of functions ρ satisfying the condition $\rho(t) = 1/|t|^\alpha + o(1/|t|^\alpha)$ as $|t| \to \infty$. This case covers the specific cases when ρ belongs to the so-called Mittag-Leffler family, see e.g. Schneider (1996), Barndorff-Nielsen and Leonenko (2005).

(c) Lemma 4 implies that for certain positive functions r with singularity at the point 0 it can be natural to consider

$$\mathcal{D}(\xi) = \int_{-1}^{1}\int_{-1}^{1} r(u - v)\xi(du)\xi(dv), \tag{4.7}$$

as an optimality criterion for choosing between competing designs for the location model. For the particular choice $r(t) = 1/|t|^\alpha$ we obtain the optimality criterion (4.6).

A sufficient condition for the strict convexity of the design criterion (4.7) is the positive definiteness of the function $r(\cdot)$ in the optimality criterion. This means that $r(\cdot)$ should be a Fourier transform of a non-zero non-negative function $h(\cdot)$, that is

$$r(t) = \int_{-\infty}^{\infty} e^{-its}h(s)ds.$$

21

The positive definiteness implies that the function $r(\cdot)$ satisfies

$$\int\limits_{-1}^{1}\int\limits_{-1}^{1} r(u-v)\zeta(du)\zeta(dv) > 0 \qquad (4.8)$$

for any signed measure $\zeta(\cdot)$ with $\zeta([-1,1]) = 0$ and $0 < \zeta_+([-1,1]) < \infty$ and the convexity of the optimality criterion follows along the lines in the proof of Lemma 1. The list of examples of positive definite functions $r(\cdot)$ includes $r(t) = 1/|t|^\alpha$ with $0 < \alpha < 1$ and $r(t) = -\log(t^2)$, $|t| \leq 1$, see Saff, Totik (1997).

Remark 3 In Lemma 4, we derived an optimality criterion of the form (4.7) with a degenerate kernel r at the point 0 using a sequence of kernels $\sigma_N^2 \rho_N(t)$ where the sequence of correlation functions $\{\rho_N(t)\}_{N\in\mathbb{N}}$ has a specified form. An alternative way of obtaining a limiting criterion of the form (4.7) with a given positive definite kernel r with $r(0) = \infty$ is to define an approximating sequence $\{\sigma_N^2 \rho_N(t)\}_{N\in\mathbb{N}}$ such that $\sigma_N^2 \rho_N(t) \to r(t)$ for all t as $N \to \infty$. For example, we can define functions $r_N(t) = \sigma_N^2 \rho_N(t)$ as convolutions of the function $r(t)$ with a density, that is

$$r_N(t) = r * K_{\omega_N}(t) = \int r(s) K_{\omega_N}(t-s)ds,$$

where K is a symmetric density,

$$K_{\omega_N}(x) = \frac{1}{\omega_N} K\left(\frac{x}{\omega_N}\right)$$

and $\omega_N \to 0$ as $N \to \infty$. In this case the functions $r_N(\cdot)$ are obviously Fourier transforms.

Our next result gives a sufficient condition for the convexity of the optimality criterion (4.7).

Theorem 4 *Let $r(\cdot)$ be a function on $\mathbb{R} \setminus \{0\}$ with $0 \leq r(t) < \infty$ for all $t \neq 0$ and $r(0) = +\infty$. Assume that there exists a monotonously increasing sequence $\{\sigma_N^2 \rho_N(t)\}_{N\in\mathbb{N}}$ of covariance functions such that $0 \leq \sigma_N^2 \rho_N(t) \leq r(t)$ for all t and all $N = 1, 2, \ldots$ and $r(t) = \lim_{N\to\infty} \sigma_N^2 \rho_N(t)$. Then (4.6) defines a convex functional on the set of all distributions; that is*

$$\mathcal{D}(\alpha\xi_2 + (1-\alpha)\xi_1) \leq \alpha\mathcal{D}(\xi_2) + (1-\alpha)\mathcal{D}(\xi_1) \quad \forall \xi_2, \xi_1 \text{ and } 0 < \alpha < 1. \qquad (4.9)$$

Proof. If $\mathcal{D}(\xi_2) = +\infty$ or $\mathcal{D}(\xi_1) = +\infty$ and $0 < \alpha < 1$ then $\mathcal{D}(\alpha\xi_2 + (1-\alpha)\xi_1) = +\infty$ and the inequality (4.9) is obvious.

Assume now $\mathcal{D}(\xi_2) < +\infty$ and $\mathcal{D}(\xi_1) < +\infty$. Define

$$B_N = \int\int \sigma_N^2 \rho_N(u-v)\xi_2(du)\xi_1(dv), \quad B = \int\int r(u-v)\xi_2(du)\xi_1(dv),$$

$$\mathcal{D}_N(\xi) = \int\int \sigma_N^2 \rho_N(u-v)\xi(du)\xi(dv), \quad A_N = \int_{-1}^{1}\int_{-1}^{1} \sigma_N^2 \rho_N(u-v)\zeta(du)\zeta(dv),$$

where $\zeta(\cdot)$ is the signed measure defined by $\zeta(du) = \xi_2(du) - \xi_1(du)$. Note that

$$B_N = \frac{1}{2}\left[\mathcal{D}_N(\xi_1) + \mathcal{D}_N(\xi_2) - A_N\right] \geq 0 \tag{4.10}$$

and

$$A_N = \mathcal{D}_N(\xi_1) + \mathcal{D}_N(\xi_2) - 2B_N, \quad A = \int\int r(u-v)\zeta(du)\zeta(dv) = \mathcal{D}(\xi_2) + \mathcal{D}(\xi_1) - 2B.$$

Similarly to the proof of Lemma 1, for all N and all $0 \leq \alpha \leq 1$ we have

$$\mathcal{D}_N(\alpha\xi_2 + (1-\alpha)\xi_1) = \alpha\mathcal{D}_N(\xi_2) + (1-\alpha)\mathcal{D}_N(\xi_1) - \alpha(1-\alpha)A_N$$

and by the Bochner-Khintchine theorem $A_N \geq 0$. Levi's monotone convergence theorem gives for $i, j \in \{1, 2\}$

$$\int\int \sigma_N^2 \rho_N(u-v)\xi_i(du)\xi_j(dv) \rightarrow \int\int r(u-v)\xi_i(du)\xi_j(dv) \quad \text{as } n \rightarrow \infty. \tag{4.11}$$

The formulae (4.11) with $i = j = 1$ and $i = j = 2$ together with (4.10) and $A_N \geq 0$ imply

$$\limsup_{n\to\infty} B_N \leq \lim_{n\to\infty} \frac{1}{2}\left[\mathcal{D}_N(\xi_1) + \mathcal{D}_N(\xi_2)\right] < \infty.$$

This and (4.11) with $i = 1, j = 2$ now imply that the sequence B_N converges to B (as $N \rightarrow \infty$) and $B < \infty$. Hence $A = \lim_{N\to\infty} A_N \geq 0$ yielding (4.9). $\qquad\square$

Theorem 5 *Assume that the criterion (4.7) is convex and define $\phi(t, \xi) = \int r(t-u)\xi(du)$. The design ξ^* is optimal if and only if*

$$\min_t \phi(t, \xi^*) \geq \mathcal{D}(\xi^*).$$

23

This theorem is a simple generalization of Theorem 1 and the proof is therefore omitted. Note also that the asymptotic optimal design ξ^*, which minimizes the criterion (4.7), cannot put positive mass at a single point if $r(\cdot)$ has a singularity at the point 0, because in this case the functional $\mathcal{D}(\xi)$ becomes infinite.

We conclude this section presenting explicit solutions of the optimal design problem for two specific singular kernels.

Lemma 5

(a) Let $r(t) = 1/|t|^\alpha$ with $0 < \alpha < 1$. Then the asymptotic optimal design minimizing the criterion (4.7) is a Beta distribution on the interval $[-1, 1]$ with density

$$p^*(t) = \frac{2^{-\alpha}}{B(\frac{1+\alpha}{2}, \frac{1+\alpha}{2})}(1+x)^{\frac{\alpha-1}{2}}(1-x)^{\frac{\alpha-1}{2}}.$$

(b) Let $r(t) = -\ln(t^2)$. Then the asymptotic optimal design minimizing the criterion (4.7) is the arcsine density on the interval $[-1, 1]$ with density

$$p^*(t) = \frac{1}{\pi\sqrt{1-x^2}}.$$

Proof. A direct computation yields that the integral

$$\int_{-1}^{1} \frac{1}{|t-u|^\alpha}(1+u)^{\frac{\alpha-1}{2}}(1-u)^{\frac{\alpha-1}{2}}\,du$$

is constant for all $0 < \alpha < 1$. Consequently, the case (a) of the Lemma follows from the second part of Theorem 5. Finally the part (b) of the Lemma is a well known fact in the theory of logarithmic potentials, see for example Saff, Totik (1997). □

References

Atkinson A.C., Donev A.N. *Optimum Experimental Designs.* Clarendon Press, Oxford, (1992)

Barndorff-Nielsen, O.E. and Leonenko, N.N. Burgers' turbulence problem with linear or quadratic external potential. J. Appl. Probab. 42 (2005), no. 2, 550-565 .

Bickel P.J., Herzberg A.M. Robustness of design against autocorrelation in time. I. Asymptotic theory, optimality for location and linear regression. Ann. Statist. 7 (1979), no. 1, 77–95.

Bickel P.J., Herzberg A.M., Schilling M.F. Robustness of design against autocorrelation in time. II. Optimality, theoretical and numerical results for the first-order autoregressive process. J. Amer. Statist. Assoc. 76 (1981), no. 376, 870–877.

Boltze L., Näther W. On effective observation methods in regression models with correlated errors. Math. Operationsforsch. Statist. Ser. Statist. 13, (1982), no. 4, 507–519.

Dette H., Leonenko N., Pepelyshev A., Zhigljavsky A. Asymptotic optimal designs under long-range dependence error structure. Bernoulli (2009).

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

Feller W. *An Introduction to Probability Theory and its Applications.* John Wiley & Sons Inc., New York. (1966).

Kiefer J. General equivalence theory for optimum designs (Approximate Theory). Annals of Statistics 2 (1974), no. 5, 849-879.

Müller W.G., Pázman A. Measures for designs in experiments with correlated errors. Biometrika 90 (2003), no. 2, 423–434.

Näther W. *Effective observation of random fields.* Teubner Verlagsgesellschaft, Leipzig, (1985a).

Näther W. Exact design for regression models with correlated errors. Statistics 16, (1985b), no. 4, 479–484.

Pázman, A. *Foundations of Optimum Experimental Design.* D. Reidel Publishing Company, Dordrecht. (1986)

Pázman A., Müller W.G. Optimal design of experiments subject to correlated errors. Statist. Probab. Lett. 52 (2001), no. 1, 29–34.

Pukelsheim F. *Optimal Design of Experiments*. John Wiley & Sons, New York, (1993).

Sacks J., Ylvisaker N.D. Designs for regression problems with correlated errors. Ann. Math. Statist. 37, (1966), 66–89.

Sacks J., Ylvisaker N.D. Designs for regression problems with correlated errors; many parameters. Ann. Math. Statist. 39, (1968), 49–69.

Saff E.B., Totik V. *Logarithmic potentials with external fields*. Springer-Verlag, Berlin, (1997).

Schneider W.R. Completely monotone generalized Mittag-Leffler functions. Exposition. Math. 14, (1996), no. 1, 3–16.

Seneta E. *Regularly varying functions*. Lecture Notes in Mathematics, Vol. 508. Springer-Verlag, (1976).

Silvey S.D. *Optimal design*. Chapman and Hall. New York. (1980).