A new approach to optimal design for linear models with correlated observations

Anatoly Zhigljavsky*, Holger Dette and Andrey Pepelyshev

Abstract

We consider the problem of designing experiments for regression in the presence of correlated observations with the location model as the main example. For a fixed correlation structure approximate optimal designs are determined explicitly, 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 article converge weakly to the measures obtained by these authors.

We also compare the asymptotic optimal design concepts of Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg (1979) and point out some inconsistencies of the latter. Finally, we combine the best features of both concepts to develop a new approach for the design of experiments for correlated observations, and it is demonstrated that the resulting design problems are related to the (logarithmic) potential theory.

2

1 Introduction

Consider the common linear regression model

$$y(t) = \theta_1 f_1(t) + \ldots + \theta_p f_p(t) + \varepsilon(t) , \tag{1}$$

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). 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, Donev and Tobias (2007) 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) represents the time and all observations correspond to one subject. This field of statistics is called the analysis of repeated measurements, see for example Morrison (1972), Lindsey (1997), Mentre, Mallet and Baccar (1997), Hughes-Oliver (1998). The difficulty in the development of the optimal design theory for correlated observations can be explained by the fact that optimal experimental designs have an extremely complicated structure and are very difficult

to find even in simple cases. Some exact optimal design problems were considered in Boltze and Näther (1982), Näther (1985a, Ch. 4), Näther (1985b), see also Pázman and Müller (2001), Müller and Pázman (2003).

Because explicit solutions of 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 purpose of the present article is to introduce a new design methodology for constructing asymptotic optimal designs for correlated data. This methodology combines the best features of the approaches of Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg. In Section 2 we introduce the main notation and prove auxiliary results. In Section 3 we derive optimal designs for several types of correlation structure in the non-asymptotic setting. We also demonstrate how these results are related to the designs derived in Bickel, Herzberg and Schilling (1981). Section 4 starts with the comparison of the transition to the asymptotic optimal designs proposed by Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg

(1979). In particular, we observe the following inconsistency of the Bickel-Herzberg model. 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 to the uncorrelated case, despite the fact that the correlation structure approximates the case of uncorrelated observations. In order to address these problems a new concept of asymptotic optimal designs for correlated observations is introduced. The resulting optimality criteria contain a singular kernel and we are able to resolve this technical difficulty. We solve the asymptotic optimal design problem for two families of correlation functions and establish a connection with the logarithmic potential theory. In Section 5 we summarize the results and discuss two main areas of possible applications, the analysis of repeated measurements and the analysis of computer experiments. For the sake of the transparency proofs of all statements are provided in Appendix.

2 Preliminaries

Consider the linear regression model (1), where $\varepsilon(t)$ is a stationary process with

$$\mathrm{E}\varepsilon(t) = 0, \quad \mathrm{E}\varepsilon(t)\varepsilon(s) = \sigma^2 \rho(t - s) \tag{2}$$

where $\rho(\cdot)$ is the correlation function. If N observations, say $y = (y_1, \ldots, y_N)^T$ are available at experimental conditions $t_1, \ldots, t_N$ and some knowledge of the correlation function is available, the vector of parameters can be estimated by the weighted least squares method, i.e.

$\hat{\theta} = (\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}^{-1}y$ with $\mathbf{X}^T = (f_i(t_j))_{i=1,\ldots,p}^{j=1,\ldots,N}$, and the variance-covariance matrix of this estimate is given by

$$\mathrm{Var}(\hat{\theta}) = \sigma^2(\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$$

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

$$\mathrm{Var}(\tilde{\theta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}. \tag{3}$$

An exact experimental design $\xi_N = \{t_1, \ldots, t_N\}$ is a collection 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.

Note that the weighted least squares estimate can only be used if the correlation structure of errors is known, and its misspecification can lead to a severe loss of efficiency. On the other hand, the ordinary least squares estimate does not employ the structure of the correlation. Obviously the ordinary least squares estimate can be less efficient than the weighted least squares estimate but in many cases (see for example Example 3 in Section 3) the loss of efficiency is often negligible. Throughout this article we will concentrate on optimal designs for the ordinary least squares estimate. These designs require also the specification of the correlation structure but a potential loss by its misspecification in the stage of design con-

struction is typically much smaller than the loss caused by misspecification of the correlation structure in the weighted least squares estimate [see Tables 1 and 3 in Dette et al. (2009)].

Note that in Example 3 the design used in conjunction with the weighted LSE is optimal for this estimate so that the pair {optimal design, estimate} is asymptotically as efficient for the ordinary LSE as for the weighted LSE, at least in the situation considered in Example 3. On the other hand, if we use the weighted LSE under the assumption of wrong correlation structure then the resulting estimate (and consequently the pair {design, estimate}) may be much less efficient than that for the ordinary LSE.

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 Sacks and Ylvisaker (1966, 1968) and 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

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

where the function $a : [0,1] \to [-T, T]$ 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 (3) can be written as

$$\mathrm{Var}(\tilde{\theta}) = \sigma^2 \mathbf{D}_f(\xi_N),$$

where

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

7

$$\mathbf{R}_f(\xi) = \iint \rho(u - v)\mathbf{f}(u)\mathbf{f}^T(v)\xi(du)\xi(dv).$$

The matrix $\mathbf{D}_f(\xi)$ 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 measure ξ on the interval $[-T, T]$ an approximate design. An (approximate) optimal design minimizes a functional of the covariance matrix $\mathbf{D}_f(\xi)$ over the class of all approximate designs.

Note that in general the function $\mathbf{D}_f(\xi)$ is not convex (with respect to the Loewner ordering) on the space of all approximate designs. This implies that even if we have a convex functional Φ on the space of symmetric matrices, the functional $\Phi(\mathbf{D}_f(\xi))$ is generally not convex on the space of designs. In particular, for $p = 1$ the functional $\mathbf{D}_f(\xi)$ is

$$D_f(\xi) = \left[\int f^2(u)\xi(du) \right]^{-2} \iint \rho(u - v)f(u)f(v)\xi(du)\xi(dv), \tag{4}$$

and this functional does not have to be convex. On the other hand, for the location model

$$y(t) = \theta + \varepsilon(t) \tag{5}$$

we have $p = 1$, $f(t) = 1$ for all t and $D_f(\xi) = D(\xi)$ where

$$D(\xi) = \iint \rho(u - v)\xi(du)\xi(dv). \tag{6}$$

This functional is convex on the set of probability measures on the interval $[-T, T]$, see Lemma 1. For this reason most of the literature discussing asymptotic optimal design problems for least squares estimation in the presence 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 article we

will follow this line and restrict our main attention to the model (5). The following lemma states that in this case the optimality criterion (6) is convex and even strictly convex. The proofs (in a slightly different language) can be found in Näther (1985a). For the sake of completeness we provide a proof of this lemma as we shall need it for references in Section 4.

We shall use the following definition. A function $g : \mathbb{R} \to \mathbb{R}$ is positive definite if for any $n \in \mathbb{N}$ and any set of real numbers $x_1, \ldots, x_n$ the $n \times n$ matrix $\mathbf{H}$ with entries $h_{ij} = g(x_i - x_j)$ is a non-negative definite matrix; correspondingly, the function g is strictly positive definite if the matrix $\mathbf{H}$ is positive definite for all $x_1 < \ldots < x_n$. Also, without loss of generality, we assume that $T = 1$ so that the design space is given by the interval $[-T, T] = [-1, 1]$.

Lemma 1 *The functional $D(\cdot)$ defined in (6) is convex. Moreover, if $\rho(\cdot)$ is strictly positive definite, then $D(\cdot)$ is strictly convex. That is,*

$$D((1 - \alpha)\xi + \alpha\xi_0) < (1 - \alpha)D(\xi) + \alpha D(\xi_0)$$

for all $0 < \alpha < 1$ and any two measures ξ and ξ_0 on $[-1, 1]$ such that $\xi - \xi_0$ is a non-zero (signed) measure.

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 (6), we have for the directional derivative of the functional D at the design ξ in the direction of ξ_0*

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

where

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

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 (6) if and only if*

$$\min_{t \in [-1,1]} \phi(t, \xi^*) \geq D(\xi^*). \tag{7}$$

(ii) *In particular, a design ξ^* is optimal if the function $\phi(t, \xi^*)$ is constant, that is $\phi(t, \xi^*) = D(\xi^*)$ for all $t \in [-1, 1]$.*

Recall that for an arbitrary one-parameter model the functional $D_f(\xi)$ (defined in (4)) does not have to be convex and, consequently, the equivalence theorem (which is a necessary and sufficient condition of the design optimality) cannot be easily generalized to this case. However, the necessary condition for the design optimality can be derived similarly, see the next theorem.

Theorem 2 *Assume $p = 1$ and the function $f : [-1, 1] \to \mathbb{R}$ is bounded and is not identically equal to 0. Consider any design ξ^* which minimizes the functional $D_f(\xi)$. Then*

$$f(t) \int \rho(t - u)f(u)\xi^*(du) \geq f^2(t) D_f(\xi^*) \int f^2(u)\xi^*(du) \tag{8}$$

10

for all $t \in [-1, 1]$.

Note that if $f(t) \geq 0$ for all $t \in [-1, 1]$ then we can divide both parts of (8) by $f(t)$.

3   Optimal designs for particular correlation functions

3.1   The exponential correlation function

We begin our investigations with the exponential correlation function, that is

$$\rho(t) = e^{-\lambda |t|} \; , \tag{9}$$

where $\lambda > 0$ is fixed. Our first result is simple to prove (using Theorem 1) and is known in literature [see Boltze and Näther (1982)]. It is presented here for the sake of completeness.

Theorem 3 *For the location model (5) with correlation function (9) 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*

$$p^*(u) = \omega^* \left(\frac{1}{2} \delta_1(u) + \frac{1}{2} \delta_{-1}(u) \right) + (1 - \omega^*) \frac{1}{2} \mathbf{1}_{[-1,1]}(u), \tag{10}$$

where $\omega^* = 1/(1 + \lambda)$, $\delta_x(\cdot)$ 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 (7) is constant and given by $D(\xi^*) = 1/(1 + \lambda)$.

It might be of interest to study the efficiency of the naive equidistant design ξ_n supported at n equidistant points $x_i = -1 + 2(i - 1)/(n - 1)$, $i = 1, \ldots, n$, with respect to the optimal

design ξ^*. This efficiency is

$$\text{Eff}(\xi_n) = \frac{D(\xi^*)}{D(\xi_n)}$$

where

$$D(\xi_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \exp(-2\lambda|i-j|/(n-1)).$$

The efficiency of the design ξ_n for different n and λ is presented in Table 1. We can observe that the efficiency of ξ_n is not high even for large n.

Table 1. *Efficiencies of the naive equidistant design supported at n points.*

λ	1.5	2.5	3.5	4.5	5.5	6.5
$n = 5$	.940	.905	.842	.768	.695	.627
$n = 10$	.923	.933	.932	.918	.896	.868
$n = 20$	.903	.919	.933	.941	.944	.942
$n = 100$	.883	.898	.914	.928	.938	.946
$n = 1000$	.881	.895	.911	.924	.935	.943

3.2   The triangular correlation function

Next we consider the triangular correlation function defined by

$$\rho(t) = \max\{0, 1 - \lambda|t|\}. \tag{11}$$

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

Theorem 4 *Consider the location model (5) with correlation function (11).*

(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, $D(\xi^*) = 1/(1 + 2\lambda)$.*

(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, $D(\xi^*) = 2\lambda/(n(n+1))$.

Example 1 For the triangular correlation function (11) we obtain the following optimal designs for the location model (5):

- If $\lambda \in [0, .5]$, the optimal design is supported at the points -1 and 1 with weights .5.

- If $\lambda \in [.5, 1]$, the optimal design is given by

$$\begin{pmatrix} -1 & 1-1/\lambda & 1/\lambda-1 & 1 \\ 1/3 & 1/6 & 1/6 & 1/3 \end{pmatrix}.$$
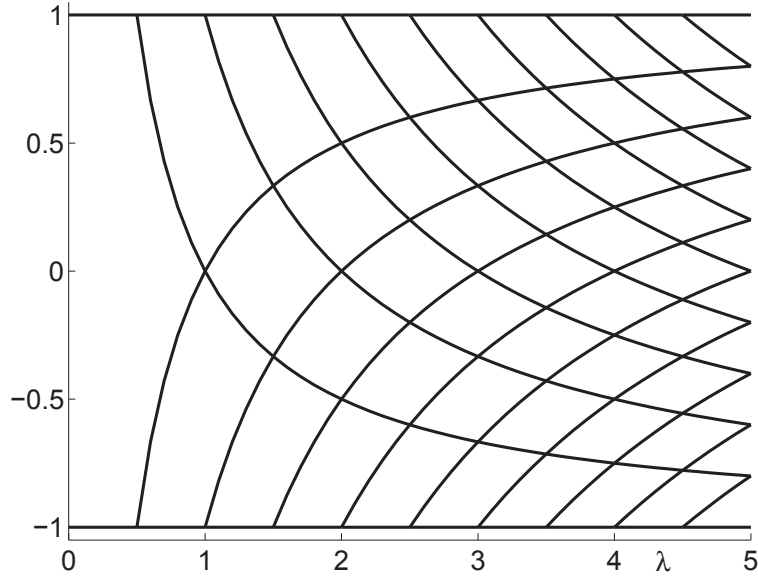
Figure 1. *Support points of the optimal designs in the location model with triangular correlation function* (11) *for different values of λ. The corresponding weights are given in Theorem 4.*

- 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

$$\begin{pmatrix} -1 & 1-\frac{3}{\lambda} & 1-\frac{2}{\lambda} & -1+\frac{1}{\lambda} & 1-\frac{1}{\lambda} & -1+\frac{2}{\lambda} & -1+\frac{3}{\lambda} & 1 \\ .4/2 & .1/2 & .2/2 & .3/2 & .3/2 & .2/2 & .1/2 & .4/2 \end{pmatrix}.$$

For larger values of λ the support points of the optimal designs for the location model (5) with correlation function (11) are displayed in Figure 1.

Example 2 Let the correlation function $\rho(\cdot)$ be defined in (11) with $0 < \lambda \leq 1$. Assume

14

that the functional $D_f(\xi)$ is (4) with $f(t) = t^k$, where k is some positive integer. Consider the two-point design ξ^* assigning masses $1/2$ to the end-points 1 and -1. Straightforward calculations show that the optimality condition (8) is met for this design. Indeed, for this design and $\rho(\cdot)$ we compute:

$$\int f^2(u)\xi^*(du) = 1, \quad D_f(\xi^*) = \tilde{\lambda},$$

and

$$\int \rho(t-u)f(u)\xi^*(du) = \frac{1}{2}\left[\rho(t-1) + (-1)^k\rho(t+1)\right],$$

where $\tilde{\lambda} = \min\{\frac{1}{2}, \lambda\}$ for odd k and $\tilde{\lambda} = \max\{\frac{1}{2}, 1-\lambda\}$ for even k. For example, if $\lambda \leq \frac{1}{2}$ then the optimality condition (2) becomes $\lambda t^{k+1} \geq \lambda t^{2k}$ for odd k and $(1-\lambda)t^k \geq (1-\lambda)t^{2k}$ for even k; these inequalities hold for all $t \in [-1,1]$. Similarly we can check the validity of the optimality condition (8) when $\frac{1}{2} \leq \lambda \leq 1$. Numerical study shows that despite the optimality criterion $D_f(\xi)$ is not convex, the design ξ^* is optimal in all these cases.

3.3 The Gaussian correlation function

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. Consider now the Gaussian correlation function

$$\rho(t) = e^{-\lambda t^2}. \tag{12}$$

Some optimal designs are given in Table 2 for selected values of $\lambda \in [0, 8.5]$. The support points of the optimal design for larger values of λ are depicted in Figure 2. From our

numerical results we conclude that for the correlation structure (12) 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 (7) is not constant, and as a consequence, the second part of Theorem 1 is not applicable.

Table 2. *Optimal designs for the location model with correlation function* (12) *for different values of* λ.

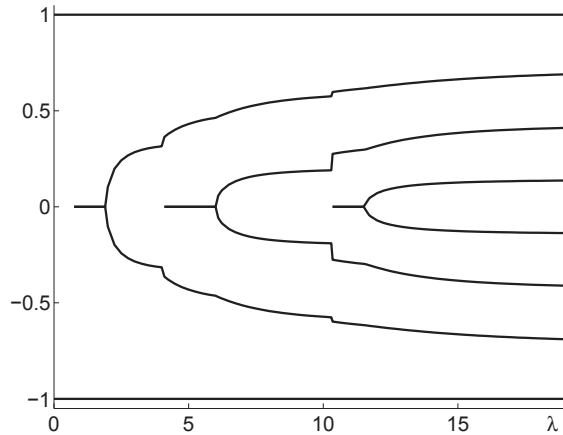| λ	t_1	t_2	t_3	w_1	w_2	w_3
0.1	± 1			.5		
0.6	± 1			.5		
0.7	± 1	0		.4685	.063	
1.9	± 1	0		.354	.292	
2.0	± 1	$\pm .104$		.348	.152	
3.7	± 1	$\pm .309$		.282	.218	
3.9	± 1	$\pm .336$	0	.277	.202	.043
6.0	± 1	$\pm .463$	0	.237	.179	.169
6.1	± 1	$\pm .469$	$\pm .058$	.235	.176	.089
8.5	± 1	$\pm .553$	$\pm .178$	.207	.154	.139

Figure 2. *Support points of the optimal design for the location model with correlation function* (12) *for different values of* λ.

3.4 The rational correlation functions

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 is obtained for the correlation function

$$\rho(t) = \frac{1}{\sqrt{1 + \lambda|t|}}, \tag{13}$$

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

$$p_{\xi^*}(u) = \omega^* \left(\frac{1}{2}\delta_1(u) + \frac{1}{2}\delta_{-1}(u) \right) + (1 - \omega^*)p^*(u), \tag{14}$$

where $\omega^* \in [0, 1]$ 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 (7) is constant.

17

Table 3. *The weight of the optimal design* (14) *in the location model with correlation function* (13) *for different values of* λ.

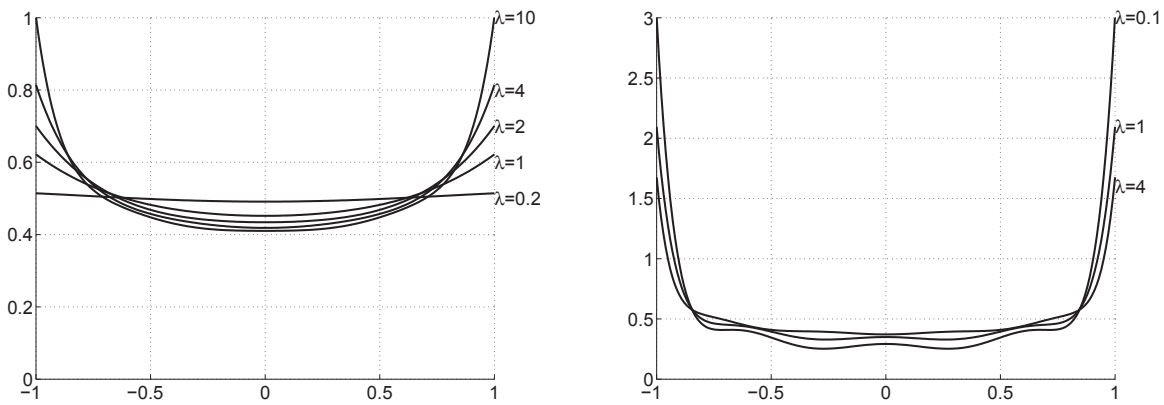| λ	0.2	1	2	4	10
ω^*	.796	.516	.392	.283	.173

Figure 3. *The density* p^* *of the optimal design* (14) *for the location model for different values of* $\lambda > 0$. *Left part: correlation structure* (13); *right part: correlation structure* (15).

As the last example we consider the correlation function

$$\rho(t) = \frac{1}{1 + \lambda |t|^{0.5}} \, , \tag{15}$$

for which our numerical results show that the optimal designs for the location model with this correlation structure are also of the form (14), where the optimal weight ω^* and the optimal density are displayed in Table 4 and the right part of Figure 3, respectively.

Table 4. *The weight of the optimal design* (14) *in the location model with correlation function* (15) *for different values of* λ.

λ	0.1	0.5	1	2	4
ω^*	.201	.151	.118	.082	.056

3.5 Transition to the limiting design in the Bickel-Herzberg model

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 in the following manner

$$\rho_N(t) = \rho(Nt), \tag{16}$$

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 (5) the optimality criterion proposed by Bickel and Herzberg (1979) is asymptotically (as $N \to \infty$) given by

$$D_{BH}(\xi) = 1 + 2 \int Q(1/p(t)) p(t) dt, \tag{17}$$

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

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

[see Theorem 2.1 in Bickel and Herzberg (1979)]. For this criterion the asymptotic optimal design ξ^* on the interval $[-1, 1]$ does not depend on ρ and has uniform density: $p^*(t) =$

19

$\frac{1}{2}\mathbf{1}_{[-1,1]}(t)$.

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 (10) with

$$\omega^* = \omega_N^* = 1/(N\lambda + 1).$$

Because $\omega^* \to 0$ 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^*.

4 New approach to the design experiments for correlated observations

In this Section we briefly describe and compare several aspects of the approaches of Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg (1979). Our goal is to develop an alternative method for the construction of optimal designs for dependent data which combines the best features of both concepts. 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 (16). 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 ordinary least squares estimate converges to 0 with a rate depending on the function ρ. We illustrate this effect in the following example.

Example 3 Consider the correlation function $\rho(t) = e^{-\lambda|t|}$. The variance of the ordinary least squares estimate obtained from the optimal design ξ^* provided by Theorem 2 is given by

$$\text{Var}(\tilde{\theta}) = \sigma^2 D(\xi^*) = \sigma^2/(1 + \lambda) .$$

The variance of the weighted least squares estimate for the uniform design (which is an asymptotically optimal design for this estimate) is exactly the same: $\text{Var}(\hat{\theta}) = \sigma^2/(1 + \lambda)$.

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 (17) 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

$$\frac{\sigma^2}{N}\left(1+\frac{2}{e^{2\lambda}-1}\right)+o\left(\frac{1}{N}\right),\quad N\to\infty\,. \tag{18}$$

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

$$D_\alpha(\xi)=\int Q_\alpha(1/p(t))p(t)dt\,, \tag{19}$$

where

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

For the correlation functions

$$\rho_\alpha^{(1)}(t)=\frac{1}{(1+|t|^2)^{\alpha/2}},\quad \rho_\alpha^{(2)}(t)=\frac{1}{1+|t|^\alpha},\quad \rho_\alpha^{(3)}(t)=\frac{1}{(1+|t|)^\alpha}, \tag{20}$$

(their properties and usefulness are discussed in Gneiting 2000, Anh, Knopova and Leonenko 2004) 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 (20) approximate the Dirac measure at the point 0.

The computation of the asymptotic variances in Example 3 illuminates the following general theoretical results:

- In the case of correlated observations, the variance of any least squares estimate 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 estimate 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 on 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 covariance functions satisfying

$$c_N(t, s) = \sigma_N^2 \rho_N(t - s) , \tag{21}$$

where

$$\rho_N(t) = \rho(a_N t), \quad \sigma_N^2 = a_N^\alpha \tau^2, \tag{22}$$

$\tau > 0$ and $0 < \alpha \le 1$ 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 the order $O(1)$ for the variance of the ordinary least squares estimate.

Example 4 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 (18) changes to

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

Lemma 3 *Assume that the function $\rho(\cdot)$ has one of the forms (20) with $0 < \alpha < 1$ and the covariance function $c(t, s) = c_N(t, s)$ is of the form (21) and (22), 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 variance of the ordinary least squares estimate $\tilde{\theta}$ for the location model is given by*

$$\mathrm{Var}(\tilde{\theta}) = \iint c_N(u, v) \xi_N(du) \xi_N(dv)$$

and converges to $\tau^2 D_\alpha(\xi)$ as $N \to \infty$, where

$$D_\alpha(\xi) = \int\int r_\alpha(u - v)\xi(du)\xi(dv) \tag{23}$$

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

Remark 1

(a) As a particular case of the sequence $\{a_N\}_{N \in \mathbb{N}}$ in Lemma 3, we can take $a_N = N^\beta$ with any $\beta > 0$.

(b) Note that the statement of Lemma 3 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 3 implies that for certain positive functions $r(\cdot)$ with singularity at the point 0 it can be natural to consider

$$D(\xi) = \int_{-1}^{1}\int_{-1}^{1} r(u - v)\xi(du)\xi(dv), \tag{24}$$

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 (23).

A sufficient condition for the strict convexity of the design criterion (24) 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

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

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

$$\int\limits_{-1}^{1}\int\limits_{-1}^{1} r(u-v)\zeta(du)\zeta(dv) > 0 \tag{25}$$

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 2 In Lemma 4, we derived an optimality criterion of the form (24) with a degenerate kernel $r(\cdot)$ 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 (24) with a given positive definite kernel $r(\cdot)$ 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 (24).

Theorem 5 *Let $r(\cdot)$ be a function on $\mathbb{R} \setminus \{0\}$ with $0 \le r(t) < \infty$ for all $t \ne 0$ 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 $0 \le \sigma_N^2 \rho_N(t) \le r(t)$ for all t and all $N = 1, 2, \ldots$ and $r(t) = \lim_{N \to \infty} \sigma_N^2 \rho_N(t)$. Then (23) defines a convex functional on the set of all distributions.*

The next equivalence theorem is a simple generalization of Theorem 1. The proofs of both theorems are very similar and therefore the proof of Theorem 6 is omitted.

Theorem 6 *Assume that the criterion (24) is convex and define $\phi(t, \xi) = \int r(t - u)\xi(du)$. The design ξ^* is optimal for (24) if and only if $\phi(t, \xi^*) \ge D(\xi^*)$ for all $t \in [-1, 1]$.*

Note also that the asymptotic optimal design ξ^*, which minimizes the criterion (24), cannot assign positive mass to any point in $[-1, 1]$ if $r(\cdot)$ has a singularity at the point 0, because in this case the functional $D(\xi)$ becomes infinite.

The next theorem is a generalization of Theorem 2. Its proof is also omitted.

Theorem 7 *Let the design ξ^* minimizes the optimality criterion*

$$
D_f(\xi) = \left[\int f^2(u)\xi(du) \right]^{-2} \int\int r(u - v)f(u)f(v)\xi(du)\xi(dv),
$$

a version of the criterion (4) for the case when the kernel $r(\cdot)$ has singularity at 0. Then

$$
f(t) \int r(t - u)f(u)\xi^*(du) \ge f^2(t)\, D_f(\xi^*) \int f^2(u)\xi^*(du)
$$

for all $t \in [-1, 1]$.

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

Theorem 8

(a) *Let $r(t) = 1/|t|^{\alpha}$ with $0 < \alpha < 1$. Then the asymptotic optimal design minimizing the criterion (24) 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 (24) is the arcsine density on the interval $[-1, 1]$ with density*

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

5 Conclusions

In this article we discuss design problems for regression models with correlated observations. New designs in the location model with various types of correlation structure in the non-asymptotic setting are derived and for some examples the efficiency of the equidistant design is investigated.

We also investigate the design problem when the number of observations increases. In particular we discuss the two main concepts introduced by Sacks and Ylvisaker (1966, 1968) and Bickel and Herzberg (1979), which embed the optimal design problems for the location model

in an asymptotic framework. It is well-known that in the case of correlated observations, the variance of any estimate does not converge to zero as the number of observation increases. The approach of Sacks and Ylvisaker (1966, 1968) maintains this property, however in the approach of Bickel and Herzberg (1979) the variance of any least squares estimate converges to zero. On the other hand, Sacks and Ylvisaker (1966, 1968) have considered the weighted least squares estimate which, however, may be inefficient if the correlation structure is mis-specified. As the main contribution this article proposes a new asymptotic setting, which combines the attractive features of both approaches. The basic idea is to allow the variance of the observations $\sigma^2 = \sigma_N^2$ to depend on the sample size N and assume that $\sigma_N^2 \to \infty$ as $N \to \infty$. The kernel in the resulting asymptotic optimal design criterion contains a singularity at the origin and we are able to resolve this technical difficulty. In addition, we solve the asymptotic optimal design problem for two families of correlation functions and establish a connection with the logarithmic potential theory.

We conclude this article with a brief discussion of two fields in statistics, where the optimal designs derived in this article can be useful. In the theory and practice of one-sample repeated measurements, the distribution of errors is assumed known and the maximum likelihood estimate of θ is used for estimation of θ under the null hypothesis that θ is the common mean of the observations, see Morrison (1972), Lindsey (1997) and Davis (2002). As a rule, the usual practice of repeated measurements is to test the null hypothesis. Numerous references to applied areas, especially in longitudinal studies, can be found in Lindsey (1997) and Davis (2002). If the error distribution is unknown then the maximum likelihood estimate cannot be used and the common practice is to use the least squares estimate. The weighted

29

least squares estimate can only be used if the correlation structure of errors is known. In our study (which goes along the lines of the study of Bickel and Herzberg) we do not need to know the correlation structure to construct the estimate, but we do need to know it to construct the optimal design. A potential loss cause by the misspecification of the correlation structure in the stage of design construction is typically much smaller than the loss caused by misspecification of the correlation structure in the weighted least squares estimate. This can be seen from Tables 1 and 3 in Dette et al. (2009).

The second clear area of applications of our results is the theory and practice of computer experiments (see, for example, Sacks et al. 1989, Welch et al. 1992, Santner, Williams and Notz 2000). In this field, the location model $y_j = \theta + \varepsilon_j$, where ε_j is a realization of a Gaussian random process, is classical and have appeared in numerous papers. The most typical covariance function for the Gaussian process is $c(t) = \sigma^2 \exp(-\lambda t^2)$. One of the main problems in the analysis of computer experiments is the estimation of parameter θ and the parameters of the covariance function, σ^2 and λ. The design problems are often considered in this context, but the number of observations is usually assumed small.

In Table 5 we present estimates of parameters for several one-dimensional computer models. We can see that the correlation parameter λ and especially the variance σ^2 increase as the number of points increases. These numerical results confirm the practical relevance of the proposed approach.

Table 5. *Maximum likelihood estimates of $\sigma^2 = \sigma_N^2$ and $\lambda = \lambda_N$ for the Gaussian covariance function $\sigma^2 \rho(\lambda t)$ and the model $y_j = \theta + \varepsilon_j$ when the observations are $y_j = \eta(x_j)$ where $\eta(x)$ is a computer model observed at N points $x_j = (j-1)/(N-1)$, $j = 1, \ldots, N$.*

N	6	8	10	14	18
the model $\eta(x) = \ln(1+x)$					
$\hat{\lambda}_N$	0.60	0.70	0.80	1.00	1.18
$\hat{\sigma}_N^2$	2.03	3.23	4.44	7.36	11.61
the model $\eta(x) = 1/(1+2x)$					
$\hat{\lambda}_N$	1.57	1.61	1.72	1.97	2.22
$\hat{\sigma}_N^2$	0.27	0.62	1.19	3.45	8.52
the model $\eta(x) = x/(1+x)$					
$\hat{\lambda}_N$	0.95	0.99	1.07	1.23	1.39
$\hat{\sigma}_N^2$	0.46	1.11	2.21	6.73	17.56

Appendix

Proof of Lemma 1. We have

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

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

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

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

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

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

31

where

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

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

Proof of Lemma 2. Taking into account the proof of Lemma 1, it follows

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

$$\square$$

Proof of Theorem 1. (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

$$\text{supp}(\zeta) \subset \{t : \phi(t, \xi^*) = \min_v \phi(v, \xi^*)\}$$

where $\text{supp}(\zeta)$ stands for the support of the measure ζ. Consequently, the necessary and sufficient condition of extremum becomes $\min_t \phi(t, \xi^*) \geq D(\xi^*)$, which is exactly (7). The assertion (ii) obviously follows from (i). $\qquad \square$

Proof of Theorem 2. Let $\xi_\alpha = (1 - \alpha)\xi + \alpha\xi_0$. Straightforward calculations give

$$
\begin{aligned}
\frac{\partial}{\partial \alpha} D_f(\xi_\alpha)\Big|_{\alpha=0} &= \\
&= \frac{\partial}{\partial \alpha}\left(\frac{(1-\alpha)^2 R_f(\xi) + \alpha^2 R_f(\xi_0) + 2\alpha(1-\alpha)\iint \rho(u-v)f(u)f(v)\xi(du)\xi_0(dv)}{((1-\alpha)\int f^2(u)\xi(du) + \alpha\int f^2(u)\xi_0(du))^2}\right)\Big|_{\alpha=0} \\
&= 2\frac{\iint \rho(u-v)f(u)f(v)\xi(du)\xi_0(dv)\int f^2(u)\xi(du) - R_f(\xi)\int f^2(u)\xi_0(du)}{(\int f^2(u)\xi(du))^3} \\
&= \frac{2}{\int f^2(u)\xi(du)}\left(\frac{\iint \rho(u-v)f(u)f(v)\xi(du)\xi_0(dv)}{\int f^2(u)\xi(du)} - D_f(\xi)\int f^2(v)\xi_0(dv)\right).
\end{aligned}
$$

For the delta-measure ξ_0 supported at a point t and the necessary condition of the minimum, we obtain the statement of the theorem. $\qquad \square$

Proof of Theorem 3. Direct calculations show that

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

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

Proof of Theorem 4. For a proof of the statement in part (a) we fix a value $t \in [-1, 1]$. If $t = t_j$ for some $j \in \{0, 1, \dots, 2\lambda\}$ 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

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

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

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

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

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

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 a similar way and the assertion follows from Theorem 2. □

Proof of Lemma 3. 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 (20) can be treated similarly and the details are omitted for the sake of brevity. □

Proof of Theorem 5. If $D(\xi_2) = +\infty$ or $D(\xi_1) = +\infty$ and $0 < \alpha < 1$ then $D(\alpha\xi_2 + (1 - \alpha)\xi_1) = +\infty$ and the convexity is obvious.

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

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

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

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

$$
B_N = \frac{1}{2}[D_N(\xi_1) + D_N(\xi_2) - A_N] \geq 0 \tag{26}
$$

and

$$A_N = D_N(\xi_1) + D_N(\xi_2) - 2B_N \,,$$

$$A = \int\int r(u - v)\zeta(du)\zeta(dv) = D(\xi_2) + D(\xi_1) - 2B \,.$$

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

$$D_N(\alpha\xi_2 \, + \, (1 - \alpha)\xi_1) = \alpha D_N(\xi_2) + (1 - \alpha)D_N(\xi_1) - \alpha(1 - \alpha)A_N$$

and by the Bochner-Khintchine theorem $A_N \ge 0$. Levi's monotone convergence theorem

gives for $i, j \in \{1, 2\}$

$$\int\int \sigma_N^2 \rho_N(u - v)\xi_i(du)\xi_j(dv) \to \int\int r(u - v)\xi_i(du)\xi_j(dv) \tag{27}$$

as $n \to \infty$. The formulae (27) with $i = j = 1$ and $i = j = 2$ together with (26) and $A_N \ge 0$

imply

$$\limsup_{n\to\infty} B_N \le \lim_{n\to\infty} \frac{1}{2}\left[D_N(\xi_1) + D_N(\xi_2)\right] < \infty \,.$$

This and (27) with $i = 1, j = 2$ now imply that the sequence B_N converges to B (as $N \to \infty$)

and $B < \infty$. Hence $A = \lim_{N\to\infty} A_N \ge 0$ yielding the convexity. $\qquad\square$

Proof of Theorem 8. 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). $\qquad\square$

References

Anh, V. V., Knopova, V. P., and Leonenko, N. N. (2004), "Continuous-time stochastic processes with cyclical long-range dependence," *Australian and New Zealand Journal of Statistics*, 46, 275–296.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007), *Optimum experimental designs, with SAS*, Oxford: Oxford University Press.

Barndorff-Nielsen, O. E. and Leonenko, N. N. (2005), "Burgers' turbulence problem with linear or quadratic external potential," *Journal of Applied Probability*, 42, 550–565.

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, 77–95.

Bickel, P. J., Herzberg, A. M., and Schilling, M. F. (1981), "Robustness of design against autocorrelation in time. II. Optimality, theoretical and numerical results for the first-order autoregressive process," *Journal of the American Statistical Association*, 76, 870–877.

Boltze, L., and Näther, W. (1982), "On effective observation methods in regression models with correlated errors." *Math. Operationsforsch. Statist. Ser. Statist.*, 13, 507–519.

Davies, C. S. (2002), *Statistical methods for the analysis of repeated measurements*, Springer.

Dette, H., Leonenko, N. N., Pepelyshev, A., and Zhigljavsky, A. (2009), "Asymptotic optimal designs under long-range dependence error structure," *Bernoulli*, 15, 1036–1056.

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

Feller, W. (1966), *An introduction to probability theory and its applications*, John Wiley & Sons Inc., New York.

Gneiting, T. (2000), "Power-law correlations, related models for long-range dependence and their simulation," *Journal of Applied Probability*, 37, 1104–1109.

Hughes-Oliver, J. M. (1998), "Optimal designs for nonlinear models with correlated errors," In Flournoy, N., Rosenberger, W.F. and Wong, W.K., Lecture Notes Monograph Series Vol. 34: New Developments and Applications in Experimental Designs, 163–174.

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

Lindsey, J. K. (1997), *Models for repeated measurements*, Oxford, Claredon Press.

Mentre, F., Mallet, A., and Baccar, D. (1997), "Optimal design in random-effect regression models," *Biometrika*, 84, 429–442.

Morrison, D. F. (1972), "The analysis of a single sample of repeated measurements," *Biometrics*, 28, 55–71.

Müller, W. G., and Pázman, A. (2003), "Measures for designs in experiments with correlated errors," *Biometrika*, 90, 423–434.

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

Näther, W. (1985b), "Exact design for regression models with correlated errors," *Statistics*, 16, 479–484.

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

Pázman, A., and Müller, W. G. (2001), "Optimal design of experiments subject to correlated errors," *Statistics and Probability Letters*, 52, 29–34.

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

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

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

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and analysis of computer experiments," With comments and a rejoinder by the authors. *Statistical Science*, 4, 409–435.

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

Santner, T. J., Williams, B. J., and Notz, W. (2003), *The design and analysis of computer experiments*, Springer-Verlag, New York.

Schneider, W. R. (1996), "Completely monotone generalized Mittag-Leffler functions," *Expositiones Mathematicae*, 14, 3–16.

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

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

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992), "Screening, predicting, and computer experiments," *Technometrics*, 34, 15–25.