

## OPTIMAL DESIGN FOR LINEAR MODELS WITH CORRELATED OBSERVATIONS

BY HOLGER DETTE<sup>\*</sup>, ANDREY PEPELYSHEV<sup>†</sup> AND ANATOLY  
ZHIGLJAVSKY<sup>‡</sup>

*Ruhr-Universität Bochum<sup>\*</sup>, RWTH Aachen<sup>†</sup> and Cardiff University<sup>‡</sup>*

In the common linear regression model the problem of determining optimal designs for least squares estimation is considered in the case where the observations are correlated. A necessary condition for the optimality of a given design is provided, which extends the classical equivalence theory for optimal designs in models with uncorrelated errors to the case of dependent data. If the regression functions are eigenfunctions of an integral operator defined by the covariance kernel, it is shown that the corresponding measure defines a universally optimal design. For several models universally optimal designs can be identified explicitly. In particular, it is proved that the uniform distribution is universally optimal for a class of trigonometric regression models with a broad class of covariance kernels and that the arcsine distribution is universally optimal for the polynomial regression model with correlation structure defined by the logarithmic potential. To the best knowledge of the authors these findings provide the first explicit results on optimal designs for regression models with correlated observations, which are not restricted to the location scale model.

**1. Introduction.** Consider the common linear regression model

$$(1.1) \quad y(x) = \theta_1 f_1(x) + \dots + \theta_m f_m(x) + \varepsilon(x) ,$$

where  $f_1(x), \dots, f_m(x)$  are linearly independent, continuous functions,  $\varepsilon(x)$  denotes a random error process or field,  $\theta_1, \dots, \theta_m$  are unknown parameters and  $x$  is the explanatory variable, which varies in a compact design space  $\mathcal{X} \subset \mathbb{R}^d$ . We assume that  $N$  observations, say  $y_1, \dots, y_N$ , can be taken at experimental conditions  $x_1, \dots, x_N$  to estimate the parameters in the linear regression model (1.1). If an appropriate estimate, say  $\hat{\theta}$ , of the parameter  $\theta = (\theta_1, \dots, \theta_m)^T$  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  $\hat{\theta}$ , where the functional should

---

*AMS 2000 subject classifications:* Primary 62K05, 60G10; secondary 31A10, 45C05

*Keywords and phrases:* optimal design, correlated observations, integral operator, eigenfunctions, arcsine distribution, logarithmic potential

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 monograph of Pukelsheim (2006)], the construction of optimal designs for dependent observations is intrinsically more difficult. On the other hand, this problem is of particular practical interest as in most applications there exists correlation between different observations. Typical examples include models, where the explanatory variable  $x$  represents the time and all observations correspond to one subject. In such situations optimal experimental designs 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), Pázman and Müller (2001) and Müller and Pázman (2003), who derived optimal designs for the location scale model

$$(1.2) \quad y(x) = \theta + \varepsilon(x).$$

Exact optimal designs for specific linear models have been investigated in Dette et al. (2008a); Kiselak and Stehlik (2008); Harman and Štulajter (2010). Because explicit solutions of optimal design problems 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), Näther (1985a), Zhigljavsky et al. (2010)]. Roughly speaking, there exist three approaches 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  $\varepsilon(x)$  is fixed and that the number of design points tends to infinity. Alternatively, Bickel and Herzberg (1979) and Bickel et al. (1981) considered a different model, where the correlation function depends on the sample size. Recently, Zhigljavsky et al. (2010) extended the Bickel-Herzberg approach and allowed the variance (in addition to the correlation function) to vary as the number of observations changes. As a result, the corresponding optimality criteria contain a kernel with a singularity at zero. The focus in all of these papers is again mainly on the location scale model (1.2).

The difficulties in the development of the optimal design theory for correlated observations can be explained by a different structure of the covariance of the least squares estimator in model (1.1), which is of the form  $M^{-1}BM^{-1}$  for certain matrices  $M$  and  $B$  depending on the design. As a consequence, the corresponding design problems are in general not convex (except for the location scale model (1.2) where  $M = 1$ ).

The present paper is devoted to the problem of determining optimal designs for more general models with correlated observations than the simple location scale model (1.2). In Section 2 we present some preliminary discussion and introduce the necessary notation. In Section 3 we investigate general conditions for design optimality. One of the main results of the paper is Theorem 3.3, where we derive necessary and sufficient conditions for the universal optimality of designs. By relating the optimal design problems to eigenvalue problems for integral operators we identify a broad class of multi-parameter regression models where the universally optimal designs can be determined explicitly. It is also shown that in this case the least squares estimate with the corresponding optimal design has the same covariance matrix as the weighted least squares estimates with its optimal design. In other words, under the conditions of Theorem 3.3 least squares estimation combined with an optimal design can never be improved by weighted least squares estimation. In Section 4 several applications are presented. In particular, we show that for a trigonometric system of regression functions involving only cosine terms with an arbitrary periodic covariance kernel the uniform distribution is universally optimal. We also prove that the arcsine design is universally optimal for the polynomial regression model with the logarithmic covariance kernel and derive some universal optimality properties of the Beta distribution. To our best knowledge these results provide the first explicit solution of optimal design problems for regression models with correlated observations which differ from the location scale model.

In Section 5 we provide an algorithm for computing optimal designs for any regression model with specified covariance function and investigate the efficiency of the arcsine and uniform distribution in polynomial regression models with exponential correlation functions. Finally, Section 6 contains some conclusions and technical details are given in Section 7.

## 2. Preliminaries.

2.1. *The asymptotic covariance matrix.* Consider the linear regression model (1.1), where  $\varepsilon(x)$  is a stochastic process with

$$(2.1) \quad E\varepsilon(x) = 0, \quad E\varepsilon(x)\varepsilon(x') = K(x, x'), \quad x \in \mathcal{X} \subset \mathbb{R}^d.$$

Throughout this paper we call the function  $K(x, x')$  covariance kernel and assume that it is continuous at all points  $(x, x') \in \mathcal{X} \times \mathcal{X}$  except possibly at the diagonal points  $(x, x)$ . We also assume that  $K(x, x') \neq 0$  for at least one pair  $(x, x')$  with  $x \neq x'$ . An important case appears when the error process is stationary and the covariance kernel is of the form  $K(x, x') = \sigma^2\rho(x - x')$ , where  $\rho(0) = 1$  and  $\rho(\cdot)$  is called the correlation function.

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

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

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

$$(2.2) \quad \text{Var}(\tilde{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}.$$

An exact experimental design  $\xi_N = \{x_1, \dots, x_N\}$  is a collection of  $N$  points in  $\mathcal{X}$ , which defines the time points or experimental conditions where observations are taken. Optimal designs for weighted or ordinary least squares estimation minimize a functional of the covariance matrix of the weighted or ordinary least squares estimate, respectively, and numerous optimality criteria have been proposed in the literature to discriminate between competing designs [see Pukelsheim (2006)].

Note that the weighted least squares estimate can only be used if the correlation structure of the errors is known, and its misspecification can lead to a considerable loss of efficiency. At the same time, 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 the loss of efficiency is often small. For example, consider the location scale model (1.2) with a stationary error process, the Gaussian correlation function  $\rho(t) = e^{-\lambda t^2}$  and the exact design  $\xi = \{-1, -2/3, -1/3, 1/3, 2/3, 1\}$ . Suppose that the guessed value of  $\lambda$  equals 1 while the true value is 2. Then the variance of the weighted least squares estimate is 0.528 computed as

$$(\mathbf{X}^T \boldsymbol{\Sigma}_{\text{guess}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_{\text{guess}}^{-1} \boldsymbol{\Sigma}_{\text{true}} \boldsymbol{\Sigma}_{\text{guess}}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}_{\text{guess}}^{-1} \mathbf{X})^{-1}$$

while the variance of the ordinary least squares estimate is 0.433. If the guessed value of  $\lambda$  equals the true value, then the variance of the weighted least squares estimate is 0.382. A similar relation between the variances holds if the location scale model and the Gaussian correlation function are replaced by a polynomial model and a triangular or exponential correlation function, respectively. For a more detailed discussion concerning advantages of the ordinary least squares against the weighted least squares estimate see Bickel and Herzberg (1979) and Section 5.1 in Nather (1985a).

Throughout this article we will concentrate on optimal designs for the ordinary least squares estimate. These designs also require the specification of the correlation structure but a potential loss by its misspecification in the stage of design construction is typically much smaller than the loss caused by the misspecification of the correlation structure in the weighted least squares estimate. Moreover, in this paper we will demonstrate that there are many situations, where the combination of the ordinary least squares estimate with the corresponding (universally) optimal design yields the same covariance matrix as the weighted least squares estimate on the basis of a (universally) optimal design for weighted least squares estimation [see the discussions in Section 4 and Section 6].

Because even in simple models the exact optimal designs are difficult to find, most authors usually use asymptotic arguments to determine efficient designs for the estimation of the model parameters [see Sacks and Ylvisaker (1966, 1968), Bickel and Herzberg (1979) or Zhigljavsky et al. (2010)]. Sacks and Ylvisaker (1966, 1968) and N  ther (1985a), Chapter 4, assumed that the design points  $\{x_1, \dots, x_N\}$  are generated by the quantiles of a distribution function, that is

$$x_i = a\left((i - 1)/(N - 1)\right), \quad i = 1, \dots, N,$$

where the function  $a : [0, 1] \rightarrow \mathcal{X}$  is the inverse of a distribution function. If  $\xi_N$  denotes a design with  $N$  points and corresponding quantile function  $a(\cdot)$ , the covariance matrix of the least squares estimate  $\tilde{\theta} = \tilde{\theta}_{\xi_N}$  given in (2.2) can be written as

$$(2.3) \quad \text{Var}(\tilde{\theta}) = D(\xi_N) = M^{-1}(\xi_N)B(\xi_N, \xi_N)M^{-1}(\xi_N),$$

where

$$(2.4) \quad M(\xi_N) = \int_{\mathcal{X}} f(u)f^T(u)\xi_N(du),$$

$$(2.5) \quad B(\xi_N, \xi_N) = \iint K(u, v)f(u)f^T(v)\xi_N(du)\xi_N(dv),$$

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

$$(2.6) \quad D(\xi) = M^{-1}(\xi)B(\xi, \xi)M^{-1}(\xi),$$

is called the covariance matrix for the design  $\xi$  and can be defined for any probability measure  $\xi$  supported on the design space  $\mathcal{X}$  such that the matrices  $B(\xi, \xi)$  and  $M^{-1}(\xi)$  are well-defined. This set will be denoted by  $\Xi$ . An (approximate) optimal design minimizes a functional of the covariance matrix  $D(\xi)$  over the set  $\Xi$  and a universally optimal design  $\xi^*$  (if it exists) minimizes the matrix  $\xi$  with respect to the Loewner ordering, that is

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

Note that on the basis of this asymptotic analysis the kernel  $K(u, v)$  has to be well defined for all  $u, v \in \mathcal{X}$ . On the other hand, Zhigljavsky et al. (2010) extended the approach in Bickel and Herzberg (1979) and proposed an alternative approximation for the covariance matrix in (2.2), where the variance of the observations also depends on the sample size. As a result they obtained an approximating matrix of the form (2.5), where the kernel  $K(u, v)$  in the matrix  $B(\xi, \xi)$  may have singularities at the diagonal. Because in this paper we are interested in designs maximizing functionals of the matrix  $D(\xi)$  (independently from the type of approximation which has been used to derive it), we will also consider singular kernels in the following discussion. Moreover, we call  $K(u, v)$  covariance kernel even if it has singularities at the diagonal. As illustrated in Example 4.2 below, the covariance kernels with singularities at the diagonal can be used as approximations to the standard covariance kernels.

Note that in general the function  $D(\xi)$  is not convex (with respect to the Loewner ordering) on the space of all approximate designs. This implies that even if one determines optimal designs by minimizing a convex functional, say  $\Phi$ , of the matrix  $D(\xi)$ , the corresponding functional  $\xi \rightarrow \Phi(D(\xi))$  is generally not convex on the space of designs  $\Xi$ . Consider for example the case  $m = 1$  where  $D(\xi)$  is given by

$$(2.7) \quad D(\xi) = \left[ \int f^2(u) \xi(du) \right]^{-2} \iint K(u, v) f(u) f(v) \xi(du) \xi(dv),$$

and it is obvious that this functional is not necessarily convex. On the other hand, for the location scale model (1.2) we have  $m = 1$ ,  $f(x) = 1$  for all  $x$  and this expression reduces to  $D(\xi) = \iint K(u, v) \xi(du) \xi(dv)$ . In the stationary case  $K(u, v) = \sigma^2 \rho(u-v)$ , where  $\rho(\cdot)$  is a correlation function, this functional is convex on the set of all probability measures on the domain  $\mathcal{X}$ , see Lemma 1 in Zhigljavsky et al. (2010) and Lemma 4.3 in N  ther (1985a). For this reason (namely the convexity of the functional  $D(\xi)$ ) most of the literature discussing asymptotic optimal design problems for least squares estimation in the presence of correlated observations considers the location

scale model, which corresponds to the estimation of the mean of a stationary process [see for example Boltze and Näther (1982), Näther (1985a,b)].

*2.2. The set of admissible design points.* Recall the definition of the vector  $f(x) = (f_1(x), \dots, f_m(x))^T$  in the regression model (1.1). We define the sets  $\mathcal{X}_0 = \{x \in \mathcal{X} : f(x) = 0\}$  and  $\mathcal{X}_1 = \mathcal{X} \setminus \mathcal{X}_0 = \{x \in \mathcal{X} : f(x) \neq 0\}$  and assume that designs  $\xi_0$  and  $\xi_1$  are concentrated on  $\mathcal{X}_0$  and  $\mathcal{X}_1$  correspondingly.

Consider the design  $\xi_\alpha = \alpha\xi_0 + (1 - \alpha)\xi_1$  with  $0 \leq \alpha < 1$ ; note that if the design  $\xi_\alpha$  is concentrated on the set  $\mathcal{X}_0$  only (corresponding to the case  $\alpha = 1$ ), then the construction of estimates is not possible. Otherwise we have

$$M(\xi_\alpha) = \int f(x)f^T(x)\xi_\alpha(dx) = (1 - \alpha)M(\xi_1), M^{-1}(\xi_\alpha) = \frac{1}{1 - \alpha}M^{-1}(\xi_1)$$

and

$$B(\xi_\alpha, \xi_\alpha) = \int \int K(x, u)f(x)f^T(u)\xi_\alpha(dx)\xi_\alpha(du) = (1 - \alpha)^2B(\xi_1, \xi_1).$$

Therefore,

$$D(\xi_\alpha) = M^{-1}(\xi_\alpha)B(\xi_\alpha, \xi_\alpha)M^{-1}(\xi_\alpha) = M^{-1}(\xi_1)B(\xi_1, \xi_1)M^{-1}(\xi_1) = D(\xi_1)$$

for all  $0 \leq \alpha < 1$ .

Consequently, observations taken at points from the set  $\mathcal{X}_0$  do not change the estimate  $\hat{\theta}$  and its covariance matrix. If we use the convention  $0 \cdot \infty = 0$ , it follows that  $\int \int K(x, u)f(x)f^T(u)\xi_0(dx)\xi_0(du) = 0$ , and this statement is also true for covariance kernels  $K(x, u)$  with a singularity at  $x = u$ .

Summarizing this discussion, we assume throughout this paper that  $f(x) \neq 0$  for all  $x \in \mathcal{X}$ .

### 3. Characterizations of optimal designs.

*3.1. General optimality criteria.* Recall the definition of the information matrix in (2.4) and define

$$B(\xi, \nu) = \int_{\mathcal{X}} \int_{\mathcal{X}} K(u, v)f(u)f^T(v)\xi(du)\nu(dv),$$

where  $\xi$  and  $\nu \in \Xi$  are two arbitrary designs and  $K(u, v)$  is an arbitrary covariance kernel. The two main examples of the kernel function  $K(u, v)$  will be  $K(u, v) = \sigma^2\rho(u - v)$  and  $K(u, v) = r(u - v)$  where  $\rho(t)$  is a correlation

function and  $r(t)$  is a non-negative definite function with singularity at zero. The latter type arises naturally if the Bickel-Herzberg approach [see Bickel and Herzberg (1979)] is extended such that the variance (in addition to the correlation function) varies as the number of observations changes [see Zhigljavsky et al. (2010) and the discussion in Section 4.4].

According to the discussion in the previous paragraph, the asymptotic covariance matrix of the least squares estimator  $\hat{\theta}$  is proportional to the matrix  $D(\xi)$  defined in (2.3). Let  $\Phi(\cdot)$  be a monotone, real valued functional defined on the space of symmetric  $m \times m$  matrices where the monotonicity of  $\Phi(\cdot)$  means that  $A \geq B$  implies  $\Phi(A) \geq \Phi(B)$ . Then the optimal design  $\xi^*$  minimizes the function

$$(3.1) \quad \Phi(D(\xi))$$

on the space  $\Xi$  of all designs. In addition to monotonicity, we shall also assume differentiability of the functional  $\Phi(\cdot)$ ; that is, the existence of the matrix of derivatives

$$C = \frac{\partial \Phi(D)}{\partial D} = \left( \frac{\partial \Phi(D)}{\partial D_{ij}} \right)_{i,j=1,\dots,m},$$

where  $D$  is any symmetric non-negative definite matrix of size  $m \times m$ . The following lemma is crucial in the proof of the optimality theorem below.

**Lemma 3.1** *Let  $\xi$  and  $\nu$  be two designs and  $\Phi$  be a differentiable functional. Set  $\xi_\alpha = (1 - \alpha)\xi + \alpha\nu$  and assume that the matrices  $M(\xi)$  and  $B(\xi, \xi)$  are nonsingular. Then the directional derivative of  $\Phi$  at the design  $\xi$  in the direction of  $\nu - \xi$  is given by*

$$\left. \frac{\partial \Phi(D(\xi_\alpha))}{\partial \alpha} \right|_{\alpha=0} = 2[\mathbf{b}(\nu, \xi) - \varphi(\nu, \xi)]$$

where

$$\varphi(\nu, \xi) = \text{tr}(M(\nu)D(\xi, \xi)C(\xi)M^{-1}(\xi)),$$

$$\mathbf{b}(\nu, \xi) = \text{tr}(M^{-1}(\xi)C(\xi)M^{-1}(\xi)B(\xi, \nu))$$

and

$$C(\xi) = \left. \frac{\partial \Phi(D)}{\partial D} \right|_{D=D(\xi)}.$$



**Proof.** Straightforward calculation shows that

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

and

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

Using the formula for the derivative of a product and the two formulas above, we obtain

$$\frac{\partial}{\partial \alpha} D(\xi_\alpha) \Big|_{\alpha=0} = -2M^{-1}(\xi)M(\nu)D(\xi) + 2M^{-1}(\xi)B(\xi, \nu)M^{-1}(\xi).$$

Note that the matrices  $M(\xi_\alpha)$  and  $B(\xi_\alpha, \xi_\alpha)$  are nonsingular for small non-negative  $\alpha$  (that is, for all  $\alpha \in [0, \alpha_0)$  where  $\alpha_0$  is a small positive number) which follows from the non-degeneracy of  $M(\xi)$  and  $B(\xi, \xi)$  and the continuity of  $M(\xi_\alpha)$  and  $B(\xi_\alpha, \xi_\alpha)$  with respect to  $\alpha$ .

Using the above formula and the fact that  $\text{tr}(H(A + A^T)) = 2 \text{tr}(HA)$  for any  $m \times m$  matrix  $A$  and any  $m \times m$  symmetric matrix  $H$ , we obtain

$$\frac{\partial \Phi(D(\xi_\alpha))}{\partial \alpha} \Big|_{\alpha=0} = \text{tr} \left( C(\xi) \frac{\partial}{\partial \alpha} D(\xi_\alpha) \right) \Big|_{\alpha=0} = 2[\mathbf{b}(\nu, \xi) - \boldsymbol{\varphi}(\nu, \xi)].$$

□

Note that the functions  $\mathbf{b}(\nu, \xi)$  and  $\boldsymbol{\varphi}(\nu, \xi)$  can be represented as

$$\mathbf{b}(\nu, \xi) = \int b(x, \xi) \nu(dx), \quad \boldsymbol{\varphi}(\nu, \xi) = \int \boldsymbol{\varphi}(x, \xi) \nu(dx),$$

where

$$(3.2) \quad \boldsymbol{\varphi}(x, \xi) = \boldsymbol{\varphi}(\xi_x, \xi) = f^T(x)D(\xi)C(\xi)M^{-1}(\xi)f(x),$$

$$(3.3) \quad b(x, \xi) = \mathbf{b}(\xi_x, \xi) = \text{tr}(C(\xi)M^{-1}(\xi)B(\xi, \xi_x)M^{-1}(\xi)),$$

and  $\xi_x$  is the probability measure concentrated at a point  $x$ .

**Lemma 3.2** *For any design  $\xi$  such that the matrices  $M(\xi)$  and  $B(\xi, \xi)$  are nonsingular we have*

$$(3.4) \quad \int \boldsymbol{\varphi}(x, \xi) \xi(dx) = \int b(x, \xi) \xi(dx) = \text{tr} D(\xi) C(\xi)$$

where the functions  $\boldsymbol{\varphi}(x, \xi)$  and  $b(x, \xi)$  are defined in (3.2) and (3.3), respectively.

**Proof.** Straightforward calculation shows that

$$\int \varphi(x, \xi) \xi(\mathrm{d}x) = \mathrm{tr}(D(\xi) C(\xi) M^{-1}(\xi) \int f(x) f^T(x) \xi(\mathrm{d}x)) = \mathrm{tr}(D(\xi) C(\xi)).$$

We also have

$$\begin{aligned} \int B(\xi, \xi_x) \xi(\mathrm{d}x) &= \int \left[ \int \int K(u, v) f(u) f^T(v) \xi(\mathrm{d}u) \xi_x(\mathrm{d}v) \right] \xi(\mathrm{d}x) \\ &= \int \left[ \int K(u, x) f(u) f^T(x) \xi(\mathrm{d}u) \right] \xi(\mathrm{d}x) = B(\xi, \xi), \end{aligned}$$

which implies

$$\int b(x, \xi) \xi(\mathrm{d}x) = \mathrm{tr}(M^{-1}(\xi) C(\xi) M^{-1}(\xi) \int B(\xi, \xi_x) \xi(\mathrm{d}x)) = \mathrm{tr}(D(\xi) C(\xi)).$$

□

The first main result of this section provides a necessary condition for the optimality of a given design.

**Theorem 3.1** *Let  $\xi^*$  be any design minimizing the functional  $\Phi(D(\xi))$ . Then the inequality*

$$(3.5) \quad \varphi(x, \xi^*) \leq b(x, \xi^*)$$

*holds for all  $x \in \mathcal{X}$ , where the functions  $\varphi(x, \xi)$  and  $b(x, \xi)$  are defined in (3.2) and (3.3), respectively. Moreover, there is equality in (3.5) for  $\xi^*$ -almost all  $x$ , that is,  $\xi^*(\mathcal{A}) = 0$  where*

$$\mathcal{A} = \mathcal{A}(\xi^*) = \{x \in \mathcal{X} \mid \varphi(x, \xi^*) < b(x, \xi^*)\}$$

*is the set of  $x \in \mathcal{X}$  such that the inequality (3.5) is strict.*

**Proof.** Consider any design  $\xi^*$  minimizing the functional  $\Phi(D(\xi))$ . The necessary condition for an element to be a minimizer of a differentiable functional states that the directional derivative from this element in any direction is non-negative. In the case of the design  $\xi^*$  and the functional  $\Phi(D(\xi))$  this yields for any design  $\nu$

$$\left. \frac{\partial \Phi(D(\xi_\alpha))}{\partial \alpha} \right|_{\alpha=0} \geq 0$$

where  $\xi_\alpha = (1 - \alpha)\xi^* + \alpha\nu$ . The inequality (3.5) follows now from Lemma 1. The assumption that the inequality (3.5) is strict for all  $x \in \mathcal{A}$  with  $\xi^*(\mathcal{A}) > 0$  is in contradiction with the identity (3.4). □

**Remark 3.1** In the classical theory of optimal design, convex optimality criteria are almost always considered. However, in at least one paper, namely Torsney (1986), an optimality theorem for a rather general non-convex optimality criteria was established and used (in the case of non-correlated observations).

3.2. *An alternative representation of the necessary condition of optimality.* Consider for a design  $\xi \in \Xi$  the associated space

$$L_2(\mathcal{X}, \xi) = \left\{ h : \mathcal{X} \rightarrow \mathbb{R} \mid \int h^2(x) \xi(dx) < \infty \right\}.$$

Note that  $f_1, \dots, f_m \in L_2(\mathcal{X}, \xi)$  and define for a design  $\xi \in \Xi$  the vector-valued function  $\int K(x, u) f(u) \xi(du)$ . Projecting each component of this function onto the subspace  $\text{span}\{f_1, \dots, f_m\} \subset L_2(\mathcal{X}, \xi)$ , we obtain the representation

$$(3.6) \quad \int K(x, u) f(u) \xi(du) = \Lambda f(x) + g(x),$$

where  $\Lambda$  is a uniquely defined  $m \times m$  matrix (note that  $\Lambda = B(\xi, \xi) M^{-1}(\xi)$ ) and the vector  $g$  satisfies

$$(3.7) \quad \int g(x) f^T(x) \xi(dx) = 0.$$

This representation defines  $g(x)$  uniquely for  $\xi$ -almost all  $x \in \mathcal{X}$ . As  $f(x)$  is continuous, we can also assume that  $g(x)$  is continuous.

Using (3.6) and the symmetry of the matrix  $B(\xi, \xi)$  we obtain

$$\begin{aligned} B(\xi, \xi) &= \int \int K(x, u) f(u) \xi(du) f^T(x) \xi(dx) \\ &= \int \Lambda f(x) f^T(x) \xi(dx) + \int g(x) f^T(x) \xi(dx) \\ &= \Lambda M(\xi) = M(\xi) \Lambda^T, \end{aligned}$$

which gives for the matrix  $D$  in (2.6)

$$D(\xi) = M^{-1}(\xi) B(\xi, \xi) M^{-1}(\xi) = M^{-1}(\xi) \Lambda = \Lambda^T M^{-1}(\xi).$$

Similarly, we obtain for the quantities (3.2) and (3.3)

$$\begin{aligned} \varphi(x, \xi) &= f^T(x) D(\xi) C(\xi) M^{-1}(\xi) f(x) \\ &= f^T(x) \Lambda^T M^{-1}(\xi) C(\xi) M^{-1}(\xi) f(x) \\ &= f^T(x) M^{-1}(\xi) C(\xi) M^{-1}(\xi) \Lambda f(x). \end{aligned}$$

We also have

$$B(\xi, \xi_x) = \int K(x, u)f(u)\xi(du)f^T(x) = \Lambda f(x)f^T(x) + g(x)f^T(x)$$

which gives

$$\begin{aligned} b(x, \xi) &= \text{tr}(C(\xi)M^{-1}(\xi)B(\xi, \xi_x)M^{-1}(\xi)) \\ &= \varphi(x, \xi) + r(x, \xi), \end{aligned}$$

where the function  $r$  is defined by

$$r(x, \xi) = f^T(x)M^{-1}(\xi)C(\xi)M^{-1}(\xi)g(x).$$

The following result is now an obvious corollary of Theorem 3.1.

**Corollary 3.1** *The necessary condition of optimality of the design  $\xi$  can be written in the form  $r(x, \xi) \geq 0$  for all  $x$ .*

**3.3.  $D$ -optimality.** For the  $D$ -optimality there exists an analogue of the celebrated ‘Equivalence Theorem’ of Kiefer and Wolfowitz (1960), which characterizes optimal designs minimizing the  $D$ -optimality criterion  $\Phi(D(\xi)) = \ln \det(D(\xi))$ .

**Theorem 3.2** *Let  $\xi^*$  be any  $D$ -optimal design. Then for all  $x \in \mathcal{X}$  we have*

$$(3.8) \quad d(x, \xi^*) \leq b(x, \xi^*),$$

where the functions  $d$  and  $b$  are defined by  $d(x, \xi) = f^T(x)M^{-1}(\xi)f(x)$  and

$$(3.9) \quad b(x, \xi) = \text{tr}(B^{-1}(\xi, \xi)B(\xi, \xi_x)) = f^T(x)B^{-1}(\xi, \xi) \int K(u, x)f(u)\xi(du),$$

respectively. Moreover, there is equality in (3.8) for  $\xi^*$ -almost all  $x$ .

**Proof.** In the case of the  $D$ -optimality criterion  $\Phi(D(\xi)) = \ln \det(D(\xi))$ , we have  $C(\xi) = D^{-1}(\xi)$ , which gives

$$\varphi(x, \xi) = f^T(x)D(\xi)D^{-1}(\xi)M^{-1}(\xi)f(x) = d(x, \xi).$$

Similarly, we simplify an expression for  $b(x, \xi)$ . Reference to Theorem 3.1 completes the proof.  $\square$

Note that the function  $r(x, \xi)$  for the  $D$ -criterion is given by

$$r(x, \xi) = f^T(x)M^{-1}(\xi)g(x)$$

and, consequently, the necessary condition of the  $D$ -optimality can be written as  $f^T(x)M^{-1}(\xi)g(x) \geq 0$  for all  $x \in \mathcal{X}$ .

The following statement illustrates a remarkable similarity between  $D$ -optimal design problems in the cases of correlated and non-correlated observations. The proof easily follows from Lemma 3.2 and Theorem 3.2.

**Corollary 3.2** *For any design  $\xi$  such that the matrices  $M(\xi)$  and  $B(\xi, \xi)$  are nonsingular we have*

$$\int d(x, \xi)\xi(dx) = \int b(x, \xi)\xi(dx) = m$$

where  $b(x, \xi)$  is defined in (3.9) and  $m$  is the number of parameters in the regression model (1.1).

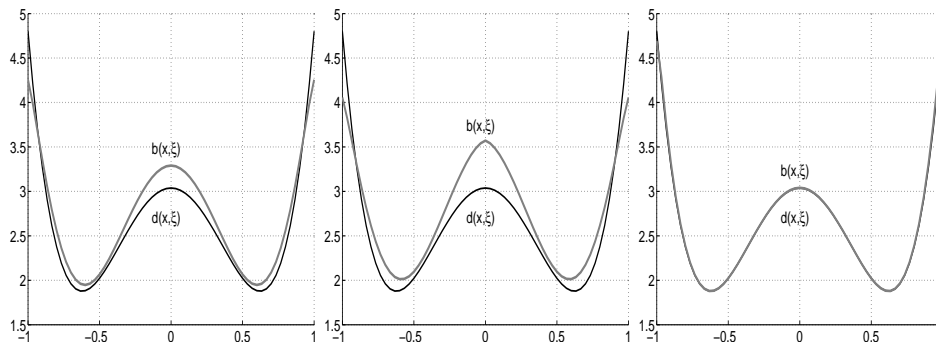


FIG 1. The functions  $b(x, \xi)$  and  $d(x, \xi)$  for the regression model (1.1) with  $f(x) = (1, x, x^2)^T$  and the covariance kernels  $K(u, v) = e^{-|u-v|}$  (left),  $K(u, v) = \max(0, 1-|u-v|)$  (middle) and  $K(u, v) = -\log(u-v)^2$  (right), and the arcsine distribution  $\xi_a$ .

**Example 3.1** Consider the quadratic regression model  $y(x) = \theta_1 + \theta_2x + \theta_3x^2 + \varepsilon(x)$  with design space  $\mathcal{X} = [-1, 1]$ . In Figure 1 we plot functions  $b(x, \xi)$  and  $d(x, \xi)$  for different covariance kernels  $K(u, v) = e^{-|u-v|}$ ,  $K(u, v) = \max\{0, 1 - |u - v|\}$  and  $K(u, v) = -\log(u - v)^2$ , where the design is the arcsine distribution with density

$$(3.10) \quad p(x) = 1/(\pi\sqrt{1-x^2}), \quad x \in (-1, 1).$$

Throughout this paper this design will be denoted by  $\xi_a$ . By the definition, the function  $d(x, \xi)$  is the same for different covariance kernels but the function  $b(x, \xi)$  depends on the choice of the kernel. From the left and middle

panel we see that the arcsine distribution does not satisfy the necessary condition of Theorem 3.1 for the kernels  $K(u, v) = e^{-|u-v|}$  and  $\max\{0, 1-|u-v|\}$  and is therefore not  $D$ -optimal for the quadratic regression model. On the other hand, for the logarithmic kernel  $K(u, v) = -\log(u-v)^2$  the necessary condition is satisfied and the arcsine distribution is a candidate for the  $D$ -optimal design. We will show in Theorem 4.5 that the arcsine distribution  $\xi_a$  is universally optimal and as a consequence optimal with respect to a broad class of criteria including the  $D$ -optimality criterion.

**3.4.  $c$ -optimality.** For the  $c$ -optimality criterion  $\Phi(D(\xi)) = c^T D(\xi) c$ , we have  $C(\xi) = cc^T$ . Consequently,

$$\varphi(x, \xi) = f^T(x) M^{-1}(\xi) c c^T M^{-1}(\xi) \Lambda f(x) = c^T M^{-1}(\xi) \Lambda f(x) f^T(x) M^{-1}(\xi) c$$

and

$$r(x, \xi) = b(x, \xi) - \varphi(x, \xi) = c^T M^{-1}(\xi) g(x) f^T(x) M^{-1}(\xi) c.$$

Therefore, the necessary condition for  $c$ -optimality simplifies to

$$(3.11) \quad c^T M^{-1}(\xi) g(x) f^T(x) M^{-1}(\xi) c \geq 0 \quad \text{for all } x \in \mathcal{X}.$$

**Example 3.2** Consider again the quadratic regression model  $y(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \varepsilon(x)$  with design space  $\mathcal{X} = [-1, 1]$ . Assume the triangular correlation function  $\rho(x) = \max\{0, 1 - |x|\}$ .

Let  $\xi = \{-1, 0, 1; 1/3, 1/3, 1/3\}$  be the design assigning weights  $1/3$  to the points  $-1, 0$  and  $1$ . For this design, we have the matrices  $M(\xi)$  and  $D(\xi)$

$$M(\xi) = \begin{pmatrix} 1 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/3 \end{pmatrix}, \quad D(\xi) = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1/2 & 0 \\ -1 & 0 & 3/2 \end{pmatrix},$$

and the matrix  $\Lambda$  and the vector  $g$  are given by

$$\Lambda = \text{diag}(1/3, 1/3, 1/3), \quad g(x) = (1/3, x/3, |x|/3)^T.$$

If  $c = (0, 1, 0)^T$  then  $r(x, \xi) = 0$  for all  $x \in [-1, 1]$  and thus the design  $\xi$  satisfies the necessary condition for  $c$ -optimality in (3.11). If  $c = (1, 0, 1)^T$  then  $r(x, \xi) = \frac{3}{4}|x|^3(1 - |x|) \geq 0$  for all  $x \in [-1, 1]$  and the design  $\xi$  also satisfies (3.11). The corresponding functions  $b$  and  $\varphi$  are displayed in the left panel of Figure 3. Numerical analysis shows that for both vectors this design is in fact  $c$ -optimal. However it is not optimal for any  $c$ -optimality criteria. For example, if  $c = (1, 0, 0)^T$  then  $r(x, \xi) = -3x(1 - |x|)(1 - x^2) \leq 0$  for all  $x \in [-1, 1]$ , showing that the design is not  $c$ -optimal [see the middle panel of Figure 3]. For this case the  $c$ -optimal design is displayed in Figure 2. The corresponding functions  $b$  and  $\varphi$  are shown in the right panel of Figure 3.

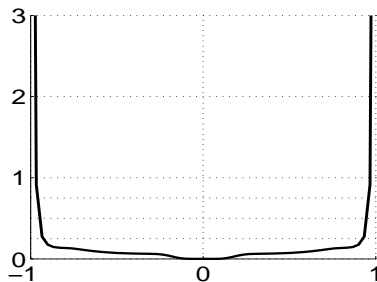


FIG 2. The  $c$ -optimal design for the quadratic model and the triangular correlation function, where  $c = (1, 0, 0)^T$ .

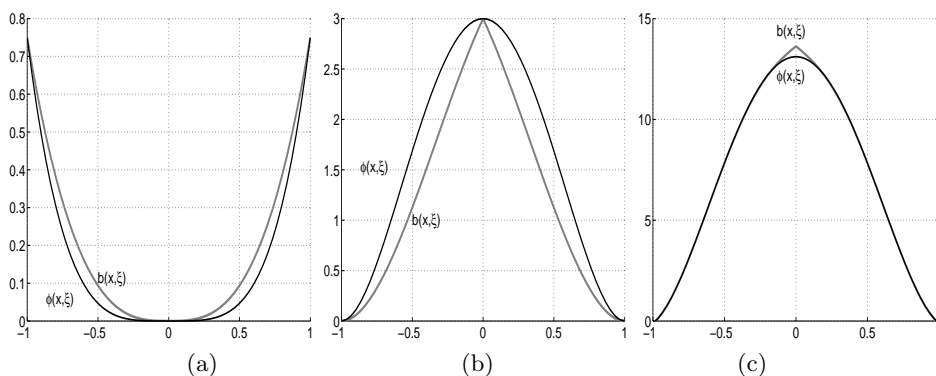


FIG 3. The functions  $b(x, \xi)$  and  $\phi(x, \xi)$  for the  $c$ -optimality criterion. (a): the design  $\xi = \{-1, 0, 1; 1/3, 1/3, 1/3\}$  and the vector  $c = (1, 0, 1)^T$ . (b): the design  $\xi = \{-1, 0, 1; 1/3, 1/3, 1/3\}$  and the vector  $c = (1, 0, 0)^T$ . (c): the optimal design for vector  $c = (1, 0, 0)^T$  displayed in Figure 2.

**3.5. Universal optimality.** In this section we consider the matrix  $D(\xi)$  defined in (2.6) as the matrix optimality criterion which we are going to minimize on the set  $\Xi$  of all designs, such that the matrices  $B(\xi, \xi)$  and  $M^{-1}(\xi)$  (and therefore the matrix  $D(\xi)$ ) are well-defined. Recall that a design  $\xi^*$  is universally optimal if  $D(\xi^*) \leq D(\xi)$  in the sense of the Loewner ordering for any design  $\xi \in \Xi$ . Note that a design  $\xi^*$  is universally optimal if and only if  $\xi^*$  is  $c$ -optimal for any vector  $c \in \mathbb{R}^m \setminus \{0\}$ ; that is,

$$c^T D(\xi^*) c \leq c^T D(\xi) c$$

for any  $\xi \in \Xi$ . The following result gives a necessary and sufficient condition for the optimality of a given design  $\xi^*$ .

**Theorem 3.3** Consider the regression model (1.1) with covariance kernel  $K$  and  $m > 1$  regression functions. Assume that the design  $\xi^* \in \Xi$  satis-

fies (3.6) with some  $m \times m$  nonsingular matrix  $\Lambda$ . Then the design  $\xi^*$  is universally optimal in the set  $\Xi$  if and only if  $g(x) = 0$  for all  $x \in \mathcal{X}$ .

**Proof.** Consider the regression model  $y(x) = f^T(x)\theta + \varepsilon(x)$ , where the full trajectory  $\{y(x)|x \in \mathcal{X}\}$  can be observed. Let  $\hat{\theta}(\mu) = \int y(x)\mu(dx)$  be a general linear unbiased estimate of the parameter  $\theta$ , where  $\mu = (\mu_1, \dots, \mu_m)^T$  is a vector of signed measures. For example, the least squares estimate for a design  $\xi$  in this model is obtained as  $\hat{\theta}(\mu_\xi)$ , where  $\mu_\xi(dx) = M^{-1}(\xi)f(x)\xi(dx)$ . The condition of unbiasedness of the estimate  $\hat{\theta}(\mu)$  means that

$$\theta = \mathbb{E}[\hat{\theta}(\mu)] = \mathbb{E}\left[\int \mu(dx)y(x)\right] = \int \mu(dx)f^T(x)\theta$$

for all  $\theta \in \mathbb{R}^m$ , which is equivalent to the condition

$$(3.12) \quad \int \mu(dx)f^T(x) = \int f(x)\mu^T(dx) = I_m,$$

where  $I_m$  denotes the  $m \times m$  identity matrix. In the following discussion we define  $\mathcal{M}$  as the set of all signed vector measures supported on  $\mathcal{X}$  satisfying condition (3.12).

For a given vector  $c \in \mathbb{R}^m$ , the variance of the estimate  $c^T\hat{\theta}(\mu)$  is given by

$$\begin{aligned} \Phi_c(\mu) = \text{Var}(c^T\hat{\theta}(\mu)) &= c^T \int \int \mathbb{E}[\varepsilon(x)\varepsilon(u)]\mu(dx)\mu^T(du)c \\ &= c^T \int \int K(x, u)\mu(dx)\mu^T(du)c, \end{aligned}$$

and a minimizer of this expression with respect to  $\mu \in \mathcal{M}$  determines the best linear estimate for  $c^T\theta$  and the corresponding  $c$ -optimal design simultaneously.

Note that the set  $\mathcal{M}$  of signed vector-valued measures is convex and the optimality criterion (3.13) is also convex. Similar arguments as given in Section 3.1 show that the directional derivative of  $\Phi_c$  at  $\mu^*$  in the direction of  $\nu - \mu^*$  is given by

$$\begin{aligned} \frac{\partial}{\partial \alpha} \Phi_c(\mu_\alpha) \Big|_{\alpha=0} &= \frac{\partial}{\partial \alpha} \Phi_c((1-\alpha)\mu^* + \alpha\nu) \Big|_{\alpha=0} \\ &= 2c^T \left[ \int \int K(x, u)\mu^*(dx)\nu^T(du) - \int \int K(x, u)\mu^*(dx)\mu^{*T}(du) \right] c. \end{aligned}$$

Because  $\Phi_c$  is convex the optimality of  $\mu^*$  is equivalent to the condition  $\frac{\partial}{\partial \alpha} \Phi_c(\mu_\alpha) \Big|_{\alpha=0} \geq 0$  for all  $\nu \in \mathcal{M}$ . Therefore, the signed measure  $\mu^*$  minimizes



the functional  $\Phi_c(\mu)$  if and only if the inequality

$$(3.13) \quad \Phi_c(\mu^*, \nu) = c^T \int \int K(x, u) \mu^*(dx) \nu^T(du) c \geq \Phi_c(\mu^*)$$

holds for all  $\nu \in \mathcal{M}$ .

Let  $\xi^* \in \Xi$  denote a measure satisfying (3.6) with  $g(x) = 0$  for all  $x \in \mathcal{X}$ . Define the vector-valued measure  $\mu_0(dx) = M^{-1}(\xi^*)f(x)\xi^*(dx)$ , then it follows for all  $\nu \in \mathcal{M}$

$$\begin{aligned} \Phi_c(\mu_0, \nu) &= c^T \int \int K(x, u) M^{-1}(\xi^*)f(x)\xi^*(dx) \nu^T(du) c \\ &= c^T M^{-1}(\xi^*)\Lambda \int f(u) \nu^T(du) c = c^T M^{-1}(\xi^*)\Lambda c, \end{aligned}$$

where we used (3.12) for the measure  $\nu$  in the last identity. On the other hand,  $\mu_0 \in \mathcal{M}$  also satisfies (3.12) and we obtain once more using the identity (3.6) with  $g(x) \equiv 0$

$$\begin{aligned} \Phi_c(\mu_0) &= c^T \int \int K(x, u) \mu_0(dx) \mu_0^T(du) c \\ &= c^T \int \left[ \int K(x, u) M^{-1}(\xi^*)f(x)\xi^*(dx) \right] \mu_0^T(du) c \\ &= c^T M^{-1}(\xi^*)\Lambda \int f(u) \mu_0^T(du) c = c^T M^{-1}(\xi^*)\Lambda c. \end{aligned}$$

This yields that for  $\mu^* = \mu_0$  we have equality in (3.13) for all  $\nu \in \mathcal{M}$ , which shows that the vector-valued measure  $\mu_0(dx) = M^{-1}(\xi^*)f(x)\xi^*(dx)$  minimizes the function  $\Phi_c$  for any  $c \neq 0$  over the set  $\mathcal{M}$  of signed vector-valued measures.

Now we return to the minimization of the function  $D(\xi)$  in the class of all designs  $\xi \in \Xi$ . For any  $\xi \in \Xi$ , define the corresponding vector-valued measure  $\mu_\xi(dx) = M^{-1}(\xi)f(x)\xi(dx)$  and note that  $\mu_\xi \in \mathcal{M}$ . We obtain

$$\begin{aligned} c^T D(\xi) c &= c^T M^{-1}(\xi) B(\xi, \xi) M^{-1}(\xi) c = \Phi_c(\mu_\xi) \\ &\geq \min_{\mu \in \mathcal{M}} \Phi_c(\mu) = \Phi_c(\mu_0) = c^T D(\xi^*) c. \end{aligned}$$

Since the design  $\xi^*$  does not depend on the particular vector  $c$ , it follows that  $\xi^*$  is universally optimal.

In order to prove the converse we will show that for any design  $\xi$  for which  $g(x) \not\equiv 0$  in (3.6) there exists a vector  $c \in \mathbb{R}^m \setminus \{0\}$ , such that  $\xi$  is not  $c$ -optimal. For this purpose we will make use of the following auxiliary result, which is proved in the Appendix [see Section 7].

**Lemma 3.3** *Let  $m > 1$  and  $a, b \in \mathbb{R}^m$  be two linearly independent vectors. Then there exists a vector  $c \in \mathbb{R}^m$  such that*

$$S_c = c^T ab^T c < 0.$$

Assume now that  $\xi$  is a design, such that  $g(x) \neq 0$  for at least one  $x \in \mathcal{X}$  in (3.6). According to (3.11), the necessary condition for  $c$ -optimality of a design  $\xi$  is

$$r_c(x, \xi) = c^T M^{-1}(\xi) g(x) f^T(x) M^{-1}(\xi) c \geq 0$$

for all  $x \in \mathcal{X}$ . The function  $g(x)$  is not proportional to  $f(x)$  for all  $x \in \mathcal{X}$  as otherwise the condition (3.7) cannot hold. Choose any point  $x_0 \in \mathcal{X}$  such that  $g(x_0) \neq 0$  and  $g(x_0)$  is not proportional to  $f(x_0)$  (recall also that  $f(x) \neq 0$  for all  $x \in \mathcal{X}$ ). Then  $r_c(x_0, \xi) = c^T M^{-1}(\xi) g(x_0) f^T(x_0) M^{-1}(\xi) c = c^T ab^T c$  with  $a = M^{-1}(\xi) g(x_0)$  and  $b = M^{-1}(\xi) f(x_0)$ . Using Lemma 3.3, we deduce that there exists a vector  $c$  such that  $r_c(x_0, \xi) < 0$ . Therefore the design  $\xi$  is not  $c$ -optimal and as a consequence also not universally optimal.  $\square$

A careful inspection of the proof of Theorem 3.3 shows that the proof of the converse part is not valid in the case  $m = 1$ . Indeed, examples in (Zhigljavsky et al., 2010) show that there exist optimal designs which do satisfy the condition (3.6) with  $g(x) \neq 0$ . In this case the condition is only sufficient and we state the result here for the sake of completeness.

**Theorem 3.4** *Consider one-parameter regression model*

$$(3.14) \quad y(x) = \theta f(x) + \varepsilon(x), \quad x \in \mathcal{X}.$$

*Assume that for a given design  $\xi^*$  there exists a constant  $\lambda > 0$  such that the identity*

$$(3.15) \quad \lambda f(x) = \int K(u, x) f(u) \xi(du)$$

*holds for all  $x \in \mathcal{X}$ . Then the design  $\xi^*$  is (universally) optimal for the model (3.14).*

#### 4. Optimal designs for specific kernels and models.

4.1. *Optimality and Mercer's theorem.* In this section we consider the case when the regression functions are proportional to eigenfunctions from Mercer's theorem. To be precise let  $\mathcal{X}$  denote a compact subset of a metric

space and let  $\nu$  denote a measure on the corresponding Borel field with positive density. Consider the integral operator

$$(4.1) \quad T_K(f)(\cdot) = \int_{\mathcal{X}} K(\cdot, u)f(u)\nu(du)$$

on  $L_2(\nu)$ . Under certain assumptions on the kernel (for example if  $K(u, v)$  is symmetric, continuous and positive definite)  $T_K$  defines a symmetric, compact self-adjoint operator. In this case Mercer's Theorem [see e.g. Kanwal (1997)] shows that there exist a countable number of eigenfunctions  $\varphi_1, \varphi_2, \dots$  with positive eigenvalues  $\lambda_1, \lambda_2, \dots$  of the operator  $K$ , that is

$$(4.2) \quad T_k(\varphi_\ell) = \lambda_\ell \varphi_\ell, \quad \ell = 1, 2, \dots$$

The next statement follows directly from Theorem 3.3.

**Theorem 4.1** *Let  $\mathcal{X}$  be a compact subset of a metric space and assume that the covariance kernel  $K(x, u)$  defines an integral operator  $T_K$  of the form (4.1), where the eigenfunctions satisfy (4.2). Consider the regression model (1.1) with  $f(x) = L(\varphi_{i_1}(x), \dots, \varphi_{i_m}(x))^T$  and the covariance kernel  $K(x, u)$ , where  $L \in \mathbb{R}^{m \times m}$  is a non-singular matrix. Then the design  $\nu$  is universally optimal.*

We note that the Mercer expansion is known analytically for certain covariance kernels. For example, if  $\nu$  is the uniform distribution on the interval  $\mathcal{X} = [-1, 1]$  and the covariance kernel is of exponential type, that is  $K(x, u) = e^{-\lambda|x-u|}$ , then the eigenfunctions are given by

$$\varphi_k(x) = \sin(\omega_k x + k\pi/2), \quad k \in \mathbb{N},$$

where  $\omega_1, \omega_2, \dots$  are positive roots of the equation  $\tan(2\omega) = -2\lambda\omega/(\lambda^2 - \omega^2)$ . Similarly, consider as a second example, the covariance kernel  $K(x, u) = \min\{x, u\}$  and  $\mathcal{X} = [0, 1]$ . In this case the eigenfunctions of the corresponding integral operator are given by

$$\varphi_k(x) = \sin((k + 1/2)\pi x), \quad k \in \mathbb{N}.$$

In the following subsection we provide a further example of the application of Mercer's theorem.

**4.2. Uniform design for periodic covariance functions.** Consider the regression functions

$$(4.3) \quad f_j(x) = \begin{cases} 1 & \text{if } j = 1 \\ \sqrt{2} \cos(2\pi(j-1)x) & \text{if } j \geq 2 \end{cases}$$

and the design space  $\mathcal{X} = [0, 1]$ . Assume that the correlation function  $\rho(x)$  is periodic with period 1, that is  $\rho(x) = \rho(x+1)$ , and let a covariance kernel be defined by  $K(u, v) = \sigma^2 \rho(u - v)$  with  $\sigma^2 = 1$ . An example of the covariance kernel  $\rho(x)$  satisfying this property is provided by a convex combination of the functions  $\{\cos(2\pi x), \cos^2(2\pi x), \dots\}$ .

**Theorem 4.2** *Consider the regression model (1.1) with regression functions  $f_{i_1}(x), \dots, f_{i_m}(x)$  ( $1 \leq i_1 < \dots < i_m$ ) defined in (4.3) and a correlation function  $\rho(x)$  that is periodic with period 1. Then the uniform design is universally optimal.*

**Proof.** We will show that the identity

$$(4.4) \quad \int_0^1 K(u, x) f_j(u) du = \int_0^1 \rho(u - x) f_j(u) du = \lambda_j f_j(x)$$

holds for all  $x \in [0, 1]$ , where  $\lambda_j = \int \rho(u) f_j(u) du$  ( $j \geq 1$ ). The assertion then follows from Theorem 4.1.

To prove (4.4), we define  $A_j(v) = \int_0^1 \rho(u - v) f_j(u) du$  which should be shown to be  $\lambda_j f_j(x)$ . For  $j = 1$  we have  $A_1(v) = \lambda_1$  because  $\int_0^1 \rho(u - v) du = \int_0^1 \rho(u) du = \lambda_1$  by the periodicity of the function  $\rho(x)$ . For  $j = 2, 3, \dots$  we note that

$$\begin{aligned} A_j(v) &= \int_0^1 \rho(u - v) f_j(u) du = \int_{-v}^{1-v} f_j(u + v) \rho(u) du \\ &= \int_0^{1-v} f_j(u + v) \rho(u) du + \int_{-v}^0 f_j(u + v) \rho(u) du. \end{aligned}$$

Because of the periodicity we have

$$\int_{-v}^0 f_j(u + v) \rho(u) du = \int_{1-v}^1 f_j(u + v) \rho(u) du,$$

which gives  $A_j(v) = \int_0^1 f_j(u + v) \rho(u) du$ . A simple calculation now shows

$$(4.5) \quad A_j''(v) = -b_j^2 A_j(v),$$

where  $b_j^2 = (2\pi(j - 1))^2$  and

$$\begin{aligned} A_j(0) &= \int_0^1 \cos(2\pi(j - 1)u) \rho(u) du = \lambda_j \\ A_j'(0) &= -b_j \int_0^1 \sin(2\pi(j - 1)u) \rho(u) du = 0. \end{aligned}$$

Therefore (from the theory of differential equations) the unique solution of (4.5) is of the form  $A_j(v) = c_1 \cos(b_j v) + c_2 \sin(b_j v)$ , where  $c_1$  and  $c_2$  are determined by initial conditions, that is  $A(0) = c_1 = \lambda_j$ ,  $A'(0) = b_j c_2 = 0$ . This yields  $A_j(v) = \lambda_j \cos(2\pi(j-1)v) = \lambda_j f_j(v)$  and proves the identity (4.4).  $\square$

4.3. *Optimal designs for the triangular covariance function.* Let us now consider the triangular correlation function defined by

$$(4.6) \quad \rho(x) = \max\{0, 1 - \lambda|x|\}.$$

The following theorem presents the optimal design for the linear model with a triangular correlation function.

**Theorem 4.3** *Consider the model (1.1) with  $f(x) = (1, x)^T$ ,  $\mathcal{X} = [-1, 1]$ , and the triangular correlation function (4.6).*

(a) *If  $\lambda \in (0, 1/2]$ , then the design  $\xi^* = \{-1, 1; 1/2, 1/2\}$  is universally optimal.*

(b) *If  $\lambda \in \mathbb{N}$ , then the design supported at  $2\lambda + 1$  points  $x_k = -1 + k/\lambda$ ,  $k = 0, 1, \dots, 2\lambda$  with equal weights is universally optimal.*

**Proof.** For a proof of part (a) we use arguments as given in the proof of Theorem 4.4 in Zhigljavsky et al. (2010) and obtain  $\int \rho(x-u) f_i(u) \xi^*(du) = f_i(x)$  for  $i = 1, 2$ . Thus, the assumptions of Theorem 3.3 are fulfilled.

Part (b). Straightforward but tedious calculations show that  $M(\xi^*) = \text{diag}(1, \gamma)$ , where  $\gamma = \sum_{k=0}^{2\lambda+1} x_k^2 / (2\lambda+1) = (\lambda+1)/(2\lambda)$ . Also we have  $\int \rho(x-u) f_i(u) \xi^*(du) = f_i(x)$  for  $i = 1, 2$ . Thus, the assumptions of Theorem 3.3 are fulfilled.  $\square$

The designs provided in Theorem 4.3 are also optimal for the location scale model, see Zhigljavsky et al. (2010). However, unlike the results of previous subsections the result of Theorem 4.3 cannot be extended to polynomial models of higher order.

We conclude this section with an example which shows that there exist  $c$ -optimal designs that are not universally optimal.

**Example 4.1** Consider the model (1.1) with  $f(x) = (1, x)^T$ ,  $\mathcal{X} = [-1, 1]$ , and the correlation function  $\rho(x) = \max\{0, 1 - |x|\}$ . It is easy to see that the design  $\xi = \{-1, 0, 1; p, 1 - 2p, p\}$  with any  $p \in (0, 1/2)$  is  $c$ -optimal for the vector  $c = (0, 1)^T$  but not universally optimal (unless  $p = 1/3$  when the design is universally optimal).

4.4. *Polynomial regression models and singular kernels.* In this section we consider the polynomial regression model, that is  $f(x) = (1, x, \dots, x^{m-1})^T$ , with logarithmic covariance kernel

$$(4.7) \quad K(u, v) = \gamma - \beta \ln(u - v)^2, \quad \beta > 0, \quad \gamma \geq 0,$$

and the kernel

$$(4.8) \quad K(u, v) = \gamma + \beta/|u - v|^\alpha, \quad 0 \leq \alpha < 1, \quad \gamma \geq 0, \quad \beta > 0$$

for which the universally optimal designs can be found explicitly. Moreover, for the covariance kernel (4.7) we will establish also uniqueness of the universally optimal design.

Random processes with singular covariance functions are interesting on their own right. They also appear naturally as approximations to many standard covariance functions  $\tilde{K}(u, v) = \sigma^2 \tilde{\rho}(u - v)$  with  $\tilde{\rho}(0) = 1$  if  $\sigma^2$  is large. A general scheme for this type of approximation is investigated in Zhigljavsky et al. (2010), Section 4. More precisely, these authors discussed the case where the covariance kernel can be represented as  $\sigma_\delta^2 \tilde{\rho}_\delta(t) = r * h_\delta(t)$  with a singular kernel  $r(t)$  and a smoothing kernel  $h_\delta(\cdot)$  (here  $\delta$  is a smoothing parameter and  $*$  denotes the convolution operator). The basic idea is illustrated in the following example.

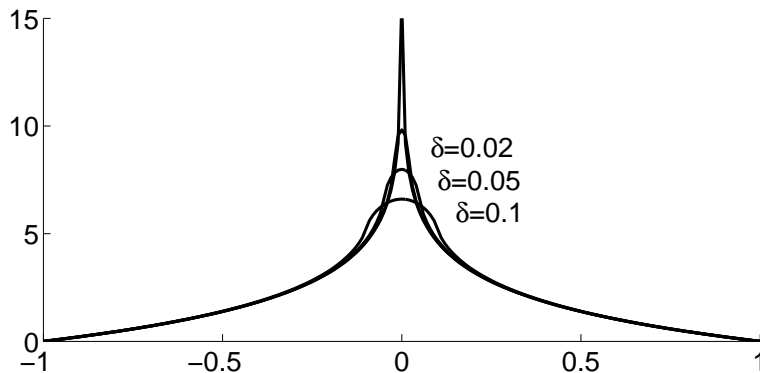


FIG 4. The logarithmic covariance kernel  $r(t) = -\ln(t)^2$  and the covariance kernel (4.9), where  $\delta = 0.02, 0.05, 0.1$ .

**Example 4.2** Consider the covariance kernel  $\tilde{K}(u, v) = \rho_\delta(u - v)$ , where

$$(4.9) \quad \rho_\delta(t) = 2 - \frac{1}{\delta} \log\left(\frac{|t + \delta|^{t+\delta}}{|t - \delta|^{t-\delta}}\right).$$

For several values of  $\delta$ , the function  $\rho_\delta$  is displayed in Figure 4. A straightforward calculation shows that  $\rho_\delta(t) = r * h_\delta(t)$ , where  $r(t) = -\ln(t)^2$  and  $h_\delta$  is the density of the uniform distribution on the interval on  $[-\delta, \delta]$ . As illustrated by Fig 4, the function  $\rho_\delta(\cdot)$  is well approximated by the singular kernel  $r(\cdot)$ , if  $\delta$  is small.

In Figure 5 we display the  $D$ -optimal designs (constructed numerically) for the quadratic model with a stationary error process with covariance kernel  $\bar{K}(u, u+t) = \rho_\delta(t)$ , where  $\rho_\delta$  is defined in (4.9) and  $\delta = 0.02, 0.05, 0.1$ . As one can see, for small  $\delta$  these designs are very close to the arcsine design, which is the  $D$ -optimal designs for the quadratic model and the logarithmic kernel, as proved in Theorem 4.5 of the following section.

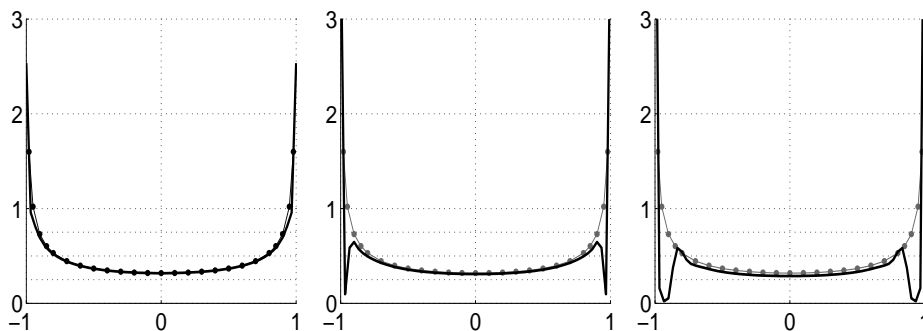


FIG 5. The  $D$ -optimal designs for the quadratic model with covariance kernel (4.9), where  $\delta = 0.02$  (left),  $\delta = 0.05$  (middle) and  $\delta = 0.1$  (right). The gray line represents the arcsine density

In Table 1 we show the efficiency of the arcsine distribution (obtained by maximizing  $\det(D(\xi))$  with the logarithmic kernel) in the quadratic regression model with the kernel (4.9). We observe a very high efficiency with respect to the  $D$ -optimality criterion. Even in the case  $\delta = 0.1$  the efficiency is 93.6% and it converges quickly to 100% as  $\delta$  approaches 0.

TABLE 1  
Efficiency of the arcsine design  $\xi_a$  for the quadratic model and the kernel (4.9).

| $\delta$            | 0.02  | 0.04  | 0.06  | 0.08  | 0.1   |
|---------------------|-------|-------|-------|-------|-------|
| $\text{Eff}(\xi_a)$ | 0.998 | 0.978 | 0.966 | 0.949 | 0.936 |

4.4.1. *Optimality of the arcsine design.* We will need the following lemma, which states a result in the theory of Fredholm-Volterra integral equations [see Mason and Handscomb (2002), Ch. 9, page 211].

**Lemma 4.1** *The Chebyshev polynomials of the first kind  $T_n(x) = \cos(n \arccos x)$  are the eigenfunctions of the integral operator with the kernel  $H(x, v) = -\ln(x-v)^2/\sqrt{1-v^2}$ . More precisely, for all  $n = 0, 1, \dots$  we have for all  $n \in \mathbb{N}$*

$$\lambda_n T_n(x) = - \int_{-1}^1 T_n(v) \ln(x-v)^2 \frac{dv}{\pi\sqrt{1-v^2}}, \quad x \in [-1, 1],$$

where  $\lambda_0 = 2 \ln 2$  and  $\lambda_n = 2/n$  for  $n \geq 1$ .

With the next result we address the problem of uniqueness of the optimal design. In particular, we give a new characterization of the arcsine distribution. A proof can be found in the Appendix.

**Theorem 4.4** *Let  $\zeta$  be a random variable supported on the interval  $[-1, 1]$ . Then  $\zeta$  is given by the arcsine distribution with density (3.10) if and only if the equality*

$$\mathbb{E} T_n(\zeta) (-\ln(\zeta - x)^2) = c_n T_n(x)$$

holds for almost all  $x \in [-1, 1]$ , where  $c_n = 2/n$  if  $n \in \mathbb{N}$  and  $c_0 = 2 \ln 2$  if  $n = 0$ .

The following result is now an immediate consequence of Theorems 3.3 and 4.4.

**Theorem 4.5** *Consider the polynomial regression model (1.1) with  $f(x) = (1, x, x^2, \dots, x^{m-1})^T$ ,  $x \in [-1, 1]$ , and the covariance kernel (4.7), then the probability measure with arcsine density (3.10) is the unique universally optimal design.*

**Proof.** We assume without loss of generality that  $\beta = 1$  and consider the function  $\rho(x) = -\ln x^2 + \gamma$  with positive  $\gamma$ . From Lemma 4.1 we obtain

$$\begin{aligned} \int_{-1}^1 (-\ln(u-x)^2 + \gamma) T_n(u) p(u) du &= - \int_{-1}^1 \ln(u-x)^2 T_n(u) p(u) du \\ &= \lambda_n T_n(x) + \gamma \delta_{n0} \end{aligned}$$

where  $\delta_{xy}$  denotes Kronecker's symbol and we have used the fact that  $\int_{-1}^1 T_n(u)/\sqrt{1-u^2} du = 0$  whenever  $n \geq 1$ . Consequently, the arcsine distribution satisfies (3.6) with  $g(x) \equiv 0$  and the statement follows from Theorems 3.3 and 4.4.  $\square$



4.4.2. *Generalized arcsine designs.* For  $\alpha \in (0, 1)$  consider the Gegenbauer polynomials  $C_m^{(\alpha)}(x)$  which are orthogonal with respect to the weight function

$$(4.10) \quad p_\alpha(x) = \frac{(\Gamma(\alpha + \frac{1}{2}))^2}{2^\alpha \Gamma(2\alpha + 1)} (1 - x^2)^{\alpha-1/2}, \quad x \in [-1, 1].$$

For the choice  $\alpha = 0$  the Gegenbauer polynomials  $C_m^{(\alpha)}(x)$  are proportional to the Chebyshev polynomials of the first kind  $T_m(x)$ . Throughout this paper we will call the corresponding beta-distributions *generalized arcsine designs* emphasizing the fact that the distribution is symmetric and the parameter  $\alpha$  varies in the interval  $(0, 1)$ . The following result (from the theory of Fredholm-Volterra integral equations of the first kind with special kernel, see Fahmy et al. (1999)) establishes an analogue of Lemma 4.1 for the kernel

$$(4.11) \quad H(u, v) = \frac{1}{|u - v|^\alpha (1 - v^2)^{(1-\alpha)/2}}.$$

**Lemma 4.2** *The Gegenbauer polynomials  $C_n^{(\alpha/2)}(x)$  are the eigenfunctions of the integral operator with the kernel defined in (4.11). More precisely, for all  $n = 0, 1, \dots$  we have*

$$\lambda_n C_n^{(\alpha/2)}(x) = - \int_{-1}^1 \frac{1}{|x - v|^\alpha} C_n^{(\alpha/2)}(v) \frac{dv}{(1 - v^2)^{(1-\alpha)/2}}$$

for all  $x \in [-1, 1]$ , where  $\lambda_n = \frac{\pi \Gamma(n+\alpha)}{\cos(\alpha\pi/2) \Gamma(\alpha)n!}$ .

The following result generalizes Theorem 8 of Zhigljavsky et al. (2010) from the case of a location scale model to polynomial regression models.

**Theorem 4.6** *Consider the polynomial regression model (1.1) with  $f(x) = (1, x, x^2, \dots, x^{m-1})^T$ ,  $x \in [-1, 1]$ , and covariance kernel (4.8). Then the design with generalized arcsine density defined in (4.10) is universally optimal.*

**Proof.** It is easy to see that the optimal design does not depend on  $\beta$  and we thus assume  $\beta = 1$ .

To prove the statement for the kernel  $\rho(x) = 1/|x|^\alpha + \gamma$  with positive  $\gamma$  we recall the definition of  $p_\alpha$  in (4.10) and obtain from Lemma 4.2

$$\int \left( \frac{1}{|u - x|^\alpha} + \gamma \right) C_n^{(\frac{\alpha}{2})}(u) p_{\frac{\alpha}{2}}(u) du = \int \frac{1}{|u - x|^\alpha} C_n^{(\frac{\alpha}{2})}(u) p_{\frac{\alpha}{2}}(u) du \propto C_n^{(\frac{\alpha}{2})}(x)$$

for any  $n \in \mathbb{N}$  since  $\int C_n^{(\alpha/2)}(u) p_{\alpha/2}(u) du = 0$ . Therefore the design  $\xi^*$  with density  $p_{\frac{\alpha}{2}}$  satisfies condition (3.6) with  $g(x) \equiv 0$  and the optimality therefore follows from Theorem 3.3.  $\square$

## 5. Numerical construction of optimal designs.

5.1. *An algorithm for computing optimal designs.* Numerical computation of optimal designs for a common linear regression model (1.1) with given correlation function can be performed by an extension of the multiplicative algorithm proposed by Dette et al. (2008b) for the case of non-correlated observations. Note that the proposed algorithm constructs a discrete design which can be considered as an approximation to a design which satisfies the necessary conditions of optimality of Theorem 3.1. By choosing a fine discretization  $\{x_1, \dots, x_n\}$  of the design space  $\mathcal{X}$  and running the algorithm long enough, the accuracy of approximation can be made arbitrarily small (in the case when convergence is achieved).

Denote by  $\xi^{(r)} = \{x_1, \dots, x_n; w_1^{(r)}, \dots, w_n^{(r)}\}$  the design at the iteration  $r$ , where  $w_1^{(0)}, \dots, w_n^{(0)}$  are nonzero weights, for example, uniform. We propose the following updating rule for the weights

$$(5.1) \quad w_i^{(r+1)} = \frac{w_i^{(r)} (\psi(x_i, \xi^{(r)}) - \beta_r)}{\sum_{j=1}^n w_j^{(r)} (\psi(x_j, \xi^{(r)}) - \beta_r)} \quad i = 1, \dots, n,$$

where  $\beta_r$  is a tuning parameter (the only condition on  $\beta_r$  is the positivity of all the weights in (5.1)),  $\psi(x, \xi) = \varphi(x, \xi)/b(x, \xi)$  and the functions  $\varphi(x, \xi)$  and  $b(x, \xi)$  are defined in (3.2) and (3.3), respectively. The condition (3.5) takes the form  $\psi(x, \xi^*) \leq 1$  for all  $x \in \mathcal{X}$ . The rule (5.1) means that at the next iteration the weight of a point  $x = x_j$  increases if the condition (3.5) does not hold at this point.

A measure  $\xi_*$  is a fixed point of the iteration (5.1) if and only if  $\psi(x, \xi_*) = 1$  for all  $x \in \text{supp}(\xi_*)$  and  $\psi(x, \xi_*) \leq 1$  for all  $x \in \mathcal{X} \setminus \text{supp}(\xi_*)$ . That is, a design  $\xi_*$  is a fixed point of the iteration (5.1) if and only if it satisfies the optimality condition of Theorem 3.1. We were not able to theoretically prove the convergence of iterations (5.1) to the design satisfying the optimality condition of Theorem 3.1 but we observed this convergence in all numerical studies. In particular, for the cases where we could derive the optimal designs explicitly, we observed convergence of the algorithm to the optimal design.

The algorithm (5.1) can be easily extended to cover the case of singular covariance kernels. Alternatively, a singular kernel can be approximated by a non-singular one using the technique described in (Zhigljavsky et al., 2010, Section 4).

5.2. *Efficiencies of the uniform and arcsine densities.* In the present section we numerically study the efficiency (with respect to the  $D$ -optimality

criterion) of two designs for different models. Specifically, we consider the uniform design and the arcsine design for the model (1.1) with  $f(x) = (1, x, \dots, x^{m-1})^T$  and different correlation functions where the design space is given by the interval  $[-1, 1]$ . We determine the efficiency of a design  $\xi$  as

$$\text{Eff}(\xi) = \left( \frac{\det D(\xi^*)}{\det D(\xi)} \right)^{1/m},$$

where  $\xi^*$  is the design computed by the algorithm described in the previous section (applied to the  $D$ -optimality criterion). We considered polynomial regression models of degree  $\leq 3$  and the correlation functions  $\rho(x) = e^{-\lambda|x|}$  and  $\rho(x) = e^{-\lambda x^2}$  for various values of  $\lambda$ . The results are depicted in Tables 2 and 3, respectively. We observe that the efficiency of the arcsine design is always larger than the efficiency of the uniform design. Moreover, the absolute difference between the efficiencies of the two designs increases as the degrees  $m$  of the polynomial increases. On the other hand, the efficiency of the uniform design and the arcsine design decreases as  $m$  increases.

TABLE 2

*Efficiencies of the uniform design  $\xi_u$  and the arcsine design  $\xi_a$  for the polynomial regression model of degree  $m - 1$  and the exponential correlation function  $\rho(x) = e^{-\lambda|x|}$ .*

|         | $\lambda$      | 0.5   | 1.5   | 2.5   | 3.5   | 4.5   | 5.5   |
|---------|----------------|-------|-------|-------|-------|-------|-------|
| $m = 1$ | Eff( $\xi_u$ ) | 0.913 | 0.888 | 0.903 | 0.919 | 0.933 | 0.944 |
|         | Eff( $\xi_a$ ) | 0.966 | 0.979 | 0.987 | 0.980 | 0.968 | 0.954 |
| $m = 2$ | Eff( $\xi_u$ ) | 0.857 | 0.832 | 0.847 | 0.867 | 0.886 | 0.901 |
|         | Eff( $\xi_a$ ) | 0.942 | 0.954 | 0.970 | 0.975 | 0.973 | 0.966 |
| $m = 3$ | Eff( $\xi_u$ ) | 0.832 | 0.816 | 0.826 | 0.842 | 0.860 | 0.876 |
|         | Eff( $\xi_a$ ) | 0.934 | 0.938 | 0.954 | 0.968 | 0.976 | 0.981 |
| $m = 4$ | Eff( $\xi_u$ ) | 0.826 | 0.818 | 0.823 | 0.835 | 0.849 | 0.864 |
|         | Eff( $\xi_a$ ) | 0.934 | 0.936 | 0.945 | 0.957 | 0.967 | 0.975 |

TABLE 3

*Efficiencies of the uniform design  $\xi_u$  and the arcsine design  $\xi_a$  for the polynomial regression model of degree  $m - 1$  and the Gaussian correlation function  $\rho(x) = e^{-\lambda x^2}$ .*

|         | $\lambda$      | 0.5   | 1.5   | 2.5   | 3.5   | 4.5   | 5.5   |
|---------|----------------|-------|-------|-------|-------|-------|-------|
| $m = 1$ | Eff( $\xi_u$ ) | 0.758 | 0.789 | 0.811 | 0.830 | 0.842 | 0.853 |
|         | Eff( $\xi_a$ ) | 0.841 | 0.907 | 0.924 | 0.932 | 0.934 | 0.935 |
| $m = 2$ | Eff( $\xi_u$ ) | 0.756 | 0.698 | 0.709 | 0.725 | 0.739 | 0.753 |
|         | Eff( $\xi_a$ ) | 0.843 | 0.833 | 0.853 | 0.868 | 0.877 | 0.885 |
| $m = 3$ | Eff( $\xi_u$ ) | 0.803 | 0.662 | 0.684 | 0.699 | 0.711 | 0.720 |
|         | Eff( $\xi_a$ ) | 0.866 | 0.771 | 0.818 | 0.844 | 0.859 | 0.869 |
| $m = 4$ | Eff( $\xi_u$ ) | 0.797 | 0.630 | 0.617 | 0.627 | 0.648 | 0.665 |
|         | Eff( $\xi_a$ ) | 0.842 | 0.713 | 0.722 | 0.746 | 0.776 | 0.799 |

**6. Conclusions.** In this paper we have addressed the problem of constructing optimal designs for least squares estimation in regression models with correlated observations. The main challenge in problems of this type is that - in contrast to “classical” optimal design theory for uncorrelated data - the corresponding optimality criteria are not convex (except for the location scale model). By relating the design problem to an integral operator problem, universally optimal design can be identified explicitly for a broad class of regression models and correlation structures. Particular attention is paid to a trigonometric regression model involving only cosines terms, where it is proved that the uniform distribution is universally optimal for any periodic kernel of the form  $K(u, v) = \rho(u - v)$ . For the classical polynomial regression model with a covariance kernel given by the logarithmic potential it is proved that the arcsine distribution is universally optimal. Moreover, optimal designs are derived for several other regression models.

So far optimal designs for regression models with correlated observations have only been derived explicitly for the location scale model and to our best knowledge the results presented in this paper provide the first explicit solutions to this type of problem for a general class of models with more than one parameter.

We have concentrated on the construction of optimal designs for least squares estimation (LSE), because the best linear unbiased estimator (BLUE) requires the knowledge of the correlation matrix. While the BLUE is often sensitive with respect to misspecification of the correlation structure the corresponding optimal designs for the LSE show a remarkable robustness. Moreover, the difference between BLUE and LSE is often surprisingly small and in many cases BLUE and LSE with certain correlation functions are asymptotically equivalent [see Rao (1967), Kruskal (1968)].

Indeed, consider the location scale model  $y(x) = \theta + \varepsilon(x)$  with  $K(u, v) = \rho(u - v)$ , where the knowledge of a full trajectory of a process  $y(x)$  is available. Define the (linear unbiased) estimate  $\hat{\theta}(G) = \int y(x)dG(x)$ , where  $G(x)$  is a distribution function of a signed probability measure. A celebrated result of Grenander (1950) states that the “estimator”  $\hat{\theta}(G^*)$  is BLUE if and only if  $\int \rho(u - x)dG^*(u)$  is constant for all  $x \in \mathcal{X}$ . This result was extended by N  ther (1985a), Sect. 4.3 to the case of random fields with constant mean. Consequently, if  $G^*(x)$  is a distribution function of a non-signed (rather than signed) probability measure, then LSE coincides with BLUE and an asymptotic optimal design for LSE is also an asymptotic optimal design for BLUE. Hajek (1956) proved that  $G^*$  is a distribution function of a non-signed probability measure if the correlation function  $\rho$  is convex on the interval  $(0, \infty)$ . Zhigljavsky et al. (2010) showed that  $G^*$  is a proper distribution function

for a certain families of correlation functions including non-convex ones.

In Theorem 3.3 we have characterized the cases where there exist universally optimal designs for ordinary least squares estimation. Specifically, a design  $\xi^*$  is universally optimal for least squares estimation if and only if the condition (3.6) with  $g(x) \equiv 0$  is satisfied. Moreover, the proof of Theorem 3.3 shows that in this case the signed vector-valued measure

$$\mu(dx) = M^{-1}(\xi^*)f(x)\xi^*(dx)$$

and the LSE minimizes (with respect to the Loewner ordering) the matrix

$$\int \int K(x, u)\mu(dx)\mu^T(du)$$

in the space  $\mathcal{M}$  of all vector-valued signed measures. Because this matrix is the covariance of the linear estimate  $\int y(x)\mu(dx)$  (where  $\mu$  is a vector of signed measures) it follows that under the assumptions of Theorem 3.3 the LSE combined with the universally optimal design  $\xi^*$  give exactly the same asymptotic covariance matrix as the BLUE and the optimal design for the BLUE.

## 7. Appendix: some technical details.

**Proof of Lemma 3.3.** If the vectors  $a$  and  $b$  are linearly independent, then there exists at least one vector, say  $c$ , such that the angle between the vectors  $c$  and  $a$  is strictly smaller than  $\pi/4$  and the angle between the vectors  $c$  and  $b$  is strictly larger than  $\pi/4$ . Therefore, the assertion follows from the well known formula  $a^T c = (a^T a)^{1/2}(c^T c)^{1/2} \cos \varphi$ , where  $\varphi$  denotes the angle between the vectors  $a$  and  $c$ .  $\square$

**Proof of Theorem 4.4.** Note that the part “if” of the statement follows from Lemma 4.1 and we should prove the part “only if”. Nevertheless, we provide a proof of the part “if” since it will be the base for proving the part “only if”.

Since the statement for  $n = 0$  is proved in Schmidt and Zhigljavsky (2009), we consider the case  $n \in \mathbb{N}$  in the rest of proof. Using the transformation  $\varphi = \arccos u$  and  $\psi = \arccos x$  we obtain  $T_n(\cos \varphi) = \cos(n\varphi)$  and

$$\int_{-1}^1 \frac{\ln(u-x)^2}{\pi\sqrt{1-u^2}} T_n(u) du = \int_0^\pi \frac{\ln(\cos \varphi - x)^2}{\pi \sin \varphi} \cos(n\varphi) \sin \varphi d\varphi.$$

Consequently, in order to prove Theorem 4.4 we have to show that the function

$$\int_0^\pi \ln(\cos \varphi - \cos \psi)^2 \cos(n\varphi) \mu(d\varphi)$$

is proportional to  $\cos(n\psi)$  if and only if  $\mu$  has a uniform density on the interval  $[0, \pi]$ . Extending  $\mu$  to the interval  $[0, 2\pi]$  as a symmetric (with respect to the center  $\pi$ ) measure,  $\mu(A) = \mu(2\pi - A)$ , and defining the measure  $\tilde{\mu}$  as  $\tilde{\mu}(A) = \mu(2A)/2$  for all Borel sets  $A \in [0, \pi]$ , we obtain

$$\begin{aligned}
& \int_0^\pi \ln(\cos \varphi - \cos \psi)^2 \cos(n\varphi) \mu(d\varphi) = \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln(\cos \varphi - \cos \psi)^2 \mu(d\varphi) \\
&= \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln \left( 2 \sin \frac{\varphi - \psi}{2} \sin \frac{\varphi + \psi}{2} \right)^2 \mu(d\varphi) \\
&= \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln 2^2 \mu(d\varphi) + \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln \left( \sin \frac{\varphi - \psi}{2} \right)^2 \mu(d\varphi) \\
&\quad + \frac{1}{2} \int_0^{2\pi} \cos(n\varphi) \ln \left( \sin \frac{\varphi + \psi}{2} \right)^2 \mu(d\varphi) \\
&= 0 + \int_0^\pi \cos(2n\varphi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(d\varphi) + \int_0^\pi \cos(2n\varphi) \ln \sin^2(\varphi + \psi/2) \tilde{\mu}(d\varphi) \\
&= 2 \int_0^\pi \cos(2n\varphi - n\psi + n\psi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(d\varphi) \\
&= 2 \cos(n\psi) \int_0^\pi \cos(2n\varphi - n\psi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(d\varphi) \\
&\quad + 2 \sin(n\psi) \int_0^\pi \sin(2n\varphi - n\psi) \ln \sin^2(\varphi - \psi/2) \tilde{\mu}(d\varphi).
\end{aligned}$$

The part “if” follows from the facts that the functions  $\cos(2nz) \ln \sin^2(z)$  and  $\sin(2nz) \ln \sin^2(z)$  are  $\pi$ -periodic and

$$\begin{aligned}
& \int_0^\pi \sin(2n\varphi - n\psi) \ln \sin^2(\varphi - \psi/2) \frac{d\varphi}{\pi} = \int_0^\pi \sin(2n\varphi) \ln \sin^2(\varphi) \frac{d\varphi}{\pi} = 0, \\
& \int_0^\pi \cos(2n\varphi - n\psi) \ln \sin^2(\varphi - \psi/2) \frac{d\varphi}{\pi} = \int_0^\pi \cos(2n\varphi) \ln \sin^2(\varphi) \frac{d\varphi}{\pi} = -1/n.
\end{aligned}$$

To prove the part “only if”, we need to show that the convolution of  $\cos(2nz) \ln \sin^2(z)$  and  $\tilde{\mu}(z)$ , i.e.

$$\int_0^\pi \cos(2n(\varphi - t)) \ln \sin^2(\varphi - t) \tilde{\mu}(d\varphi),$$

is constant for almost all  $t \in [0, \pi]$  if and only if  $\tilde{\mu}$  is uniform; and the same holds for the convolution of  $\sin(2nz) \ln \sin^2(z)$  and  $\tilde{\mu}(z)$ . This, however, follows from (Schmidt and Zhigljavsky, 2009, Lem. 3) since  $\cos(2nz) \ln \sin^2(z) \in L^2([0, \pi])$  and all complex Fourier coefficients of these functions are non-zero. Indeed,

$$\int_0^\pi \cos(2nt) \ln \sin^2(t) \sin(2kt) dt = 0 \quad \forall k \in \mathbb{Z}$$

$$\int_0^\pi \cos(2nt) \ln \sin^2(t) \cos(2kt) dt = (\gamma_{|n+k|} + \gamma_{|n-k|})/2 \quad \forall k \in \mathbb{Z},$$

where  $\gamma_0 = -2\pi \log 2$  and  $\gamma_k = -\pi/k$  for  $k \in \mathbb{N}$ , see formula 4.384.3 in Gradshteyn and Ryzhik (1965).

**Acknowledgements.** This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt C2) of the German Research Foundation (DFG). Parts of this paper were written during a visit of the authors at the Isaac Newton Institute, Cambridge, UK, and the authors would like to thank the institute for its hospitality and financial support. We are also grateful to the referees and the associate editor for their constructive comments on an earlier version of this manuscript.

## References.

- Bickel, P. J. and Herzberg, A. M. (1979). Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression. *Annals of Statistics*, 7(1):77–95.
- 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(376):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.
- Dette, H., Kunert, J., and Pepelyshev, A. (2008a). Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica*, 18(1):135–154.
- Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2008b). Improving updating rules in multiplicative algorithms for computing  $D$ -optimal designs. *Computational Statistics and Data Analysis*, 53(2):312–320.
- Fahmy, M. H., Abdou, M. A., and Darwish, M. A. (1999). Integral equations and potential-theoretic type integrals of orthogonal polynomials. *Journal of Computational and Applied Mathematics*, 106:245–254.
- Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Table of Integrals, Series, and Products*. Academic Press, New York-London.
- Grenander, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.*, 1:195–277.
- Hajek, J. (1956). Linear estimation of the mean value of a stationary random process with convex correlation function. *Czechoslovak Mathematical Journal*, 6(81):94–117.
- Harman, R. and Štulajter, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika*, 72(2):281–294.
- Kanwal, R. (1997). *Linear Integral Equations*. Birkhauser, Boston.
- Kiefer, J. (1974). General equivalence theory for optimum designs (Approximate Theory). *Annals of Statistics*, 2:849–879.

- Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremem problems. *Canadian Journal of Mathematics*, 12:363–366.
- Kiselak, J. and Stehlik, M. (2008). Equidistant and  $d$ -optimal designs for parameters of ornsteinuhlenbeck process. *Statistics & Probability Letters*, 78(12):1388–1396.
- Kruskal, W. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Annals of Mathematical Statistics*, 39:70–75.
- Mason, J. C. and Handscomb, D. C. (2002). *Chebyshev Polynomials*. Oxford University Computing Laboratory, England, UK.
- 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. and Müller, W. G. (2001). Optimal design of experiments subject to correlated errors. *Statistics and Probability Letters*, 52:29–34.
- Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.
- Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Sympos., Univ. California Press, Berkeley, Calif.*, pages 355–372.
- Sacks, J. and Ylvisaker, N. D. (1966). Designs for regression problems with correlated errors. *Annals of Mathematical Statistics*, 37:66–89.
- Sacks, J. and Ylvisaker, N. D. (1968). Designs for regression problems with correlated errors; many parameters. *Annals of Mathematical Statistics*, 39:49–69.
- Schmidt, K. and Zhigljavsky, A. (2009). A characterization of the arcsine distribution. *Statistics and Probability Letters*, 79:2451–2455.
- Torsney, B. (1986). Moment inequalities via optimal design theory. *Linear Algebra Appl.*, 82:237–253.
- Zhigljavsky, A., Dette, H., and Pepelyshev, A. (2010). A new approach to optimal design for linear models with correlated observations. *Journal of the American Statistical Association*, 105:1093–1103.

FAKULTÄT FÜR MATHEMATIK  
RUHR-UNIVERSITÄT BOCHUM  
BOCHUM, 44780, GERMANY  
E-MAIL: holger.dette@rub.de

INSTITUTE OF STATISTICS  
RWTH AACHEN UNIVERSITY  
AACHEN, 52056, GERMANY  
E-MAIL: pepelyshev@stochastik.rwth-aachen.de

SCHOOL OF MATHEMATICS  
CARDIFF UNIVERSITY  
CARDIFF, CF24 4AG, UK  
E-MAIL: ZhigljavskyAA@cf.ac.uk