Design for regression models with correlated errors

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany

Andrey Pepelyshev
Department of Statistics &
Mathematics in Finance
RWTH Aachen
Aachen, 52056
Germany

Anatoly Zhigljavsky
School of Mathematics
Cardiff University
Cardiff, CF24 4AG
UK

Abstract

The present article is a draft of a chapter in the *Handbook of Design and Analysis of Experiments* and provides a survey of results on experimental design for linear regression models with correlated responses.

Contents

1 Estimation and design for correlated errors: the main approaches

1.1 Introduction

The common linear regression model is given by

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

where $f_1(x), \ldots, f_m(x)$ are given linearly independent functions, $\varepsilon(x)$ denotes a centered random error process or field, $\theta_1, \ldots, \theta_m$ are unknown parameters and x is the explanatory variable, which varies in the design space $\mathcal{X} \subset \mathbb{R}^d$.

We assume that N observations can be taken at experimental conditions $x_1, \ldots, x_N$ to estimate the parameters in the linear regression model (1). If an estimate of $\theta = (\theta_1, \ldots, \theta_m)^T$ 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, 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 monograph of Pukelsheim (2006)], the construction of optimal designs for dependent observations is intrinsically more difficult. However, this problem is of particular practical interest as in most applications the observations are correlated. Typical examples include models, where the explanatory variable x represents the time and all observations correspond to one subject.

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. As a result of this assumption, the design points become very close to each other and the corresponding asymptotic optimal designs depend only on the behavior of the correlation function in a neighborhood of the point 0.

Alternatively, Bickel and Herzberg (1979) and Bickel et al. (1981) considered a different

model, where the ordinary least square estimate is used and the correlation function depends on the number of observations. The covariance matrix of the estimate is of order $O(1)$ in the model considered by Sacks and Ylvisaker (1966) and of order $1/N$ in the model discussed by Bickel and Herzberg (1979). Therefore the approach of Bickel and Herzberg (1979) makes the optimal designs derived for the dependent and independent cases more comparable. These authors assumed that the observations in model (1) have a correlation structure corresponding to a nondegenerate stationary process with short range dependence where a correlation function ρ satisfies $\rho(x) = o(1/x)$ if $x \to \infty$. Dette et al. (2009) extended results of Bickel and Herzberg (1979) to the case where an error process has long range dependence. Recently Zhigljavsky et al. (2010) modified 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 asymptotic covariance matrices may contain a kernel with a singularity at the diagonal.

Significant research has been devoted to constructing exact optimal designs for the best linear unbiased estimator (BLUE), see Section 3. The simplest algorithm was proposed in Brimkulov et al. (1980). Other algorithms are based on the method of virtual noise developed by Pázman and Müller (2001) and the method of the expansion of the covariance kernel suggested by Fedorov and Müller (2007). Note that BLUE 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 correlation structure. Obviously the ordinary least squares estimate can be less efficient than BLUE but in many cases the loss of efficiency is either small or negligible.

The structure of this survey is as follows. In Section 1.2 we introduce different variations of the optimal design problem for the linear regression model (1). In particular, we consider various assumptions about the design space $\mathcal{X}$, the vector-function $f(x) = (f_1(x), \ldots, f_m(x))^T$, the covariance kernel $K(x, x') = E[\varepsilon(x)\varepsilon(x')]$, and also different sets of designs, three different estimates of the unknown parameters and corresponding covariance matrices. In Section 1.3 we briefly discuss the concept of information contained in design points and discuss some well-known paradoxes.

Section 2 is devoted to the problem of designing experiments for one-parameter models. This problem is often easier than similar problems for the multi-parameter case and in some cases it can be solved explicitly, see for example, Theorem 5 in Section 2. The easiest one-parameter model is the so-called location scale model where the variance of the ordinary least square estimate leads to a convex design optimality criterion and makes many tools of the convex optimization theory applicable. Moreover, if the correlation function ρ of the stationary error process in a location scale model is convex for $x > 0$, then the ordinary least square estimate coincides with BLUE, see Theorem 4 in Section 2.

Section 3 is devoted to the problem of optimal design for BLUE. We review classical Sacks-

Ylsvisaker results and also less classical results of Fedorov-Müller and Pazman-Müller. Moreover, we also consider the well-known exchange algorithm for the construction of N-point optimal designs. Additionally, we review one of Harman's results on optimal design for prediction in the case of quadratic model and Wiener process [see Harman and Stulajter (2010)].

Section 4 reviews some results concerning characterization and construction of optimal designs for the ordinary least square estimate. We explain the classical Bickel-Herzberg approach and its extension for the case of a long-range dependent error process. We also review some of the recent results of the authors concerning the explicit construction of optimal designs for models with $m \geq 2$ parameters and some particular covariance kernels.

Section 5 contains selected proofs and a table of common correlation functions that appear in discussions on optimal design for correlated observations.

The authors are aware of the following three substantial surveys devoted to the theory of optimal designs for correlated observations, see also a short survey in Müller's book [Müller (2007)]. The first one is an excellent book by Näther (1985a) and the other two are the surveys by Cambanis (1985) and Fedorov (1996). As much research has been done in recent years, we feel that there is a need in a new survey on the subject.

1.2 Different versions of the problem

1.2.1 General regression problem

The general multi-parameter linear regression model (1) can be written as

$$y(x) = \theta^T f(x) + \varepsilon(x), \tag{2}$$

where the explanatory variable x belongs to a design space $\mathcal{X}$, $f(x) = (f_1(x), \ldots, f_m(x))^T$ is a vector of linearly independent regression functions, $\varepsilon(x)$ denotes random process or field with $E[\varepsilon(x)] = 0$ and $E[\varepsilon(x)\varepsilon(x')] = \sigma^2 K(x, x')$. The vector of parameters $\theta = (\theta_1, \ldots, \theta_m)^T$ is unknown and has to be estimated on the basis of observations taken from one realization of a stochastic process (or field) $y(x)$. The function $K(\cdot, \cdot)$ will be called covariance kernel. The multiplier σ^2 is a positive constant and may also be unknown.

Throughout this article we consider different variations and specifications of the model (2) in relation to the problem of optimal design.

1.2.2 Design space $\mathcal{X}$

We make a distinction between the following forms of the design space:
(a) $\mathcal{X}$ is a finite set;
(b) $\mathcal{X}$ is an interval $[-1, 1]$;
(c) $\mathcal{X} \subseteq \mathbb{R}$;
(d) $\mathcal{X} \subseteq \mathbb{R}^d$ with $d \geq 1$.
In the case $d = 2$ the model (2) is called spatial model. If $d \geq 2$ then $\varepsilon(x)$ is called a random field. If nothing is assumed about $\mathcal{X}$, then the most general case, which is (d), is considered.

1.2.3 Vector of regression functions $f(x)$

We will distinguish the following forms of the vector of functions $f(x)$:
(a) the general case with $m \geq 1$;
(b) the case of $m = 1$ where the model (2) is called one-parameter model, see Section 2;
(c) the case of $m = 1$ and $f(x) = 1$ where the model (2) is called location scale model, see Section 2.2.

To avoid technical difficulties, we always assume that all the components of the vector $f(x)$ are continuous functions.

1.2.4 Covariance kernels K

We will distinguish the following cases for the covariance kernel $K(\cdot, \cdot)$:

1. the general positive definite function $K(\cdot, \cdot)$;
2. kernels of stationary processes or fields which have the form $K(x, x') = \rho(x - x')$, where $\rho(\cdot)$ is a correlation function. Examples of commonly used correlation functions are given in the Appendix [see Table 1];
3. kernels with nugget term, that is, $K(x, x') = \gamma K_0(x, x') + (1 - \gamma)\delta_{x,x'}$, where $0 < \gamma < 1$, the kernel K_0 is continuous on the diagonal and δ denotes Kronecker's symbol;
4. kernels with singularity at the diagonal: these kernels possess the property that $K(x, x') \to \infty$ as $x \to x'$ and $K(x, x)$ is not defined for some $x \in \mathcal{X}$ (see Section 4.4 for examples).

By the definition, a kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called positive definite if $K(x, x') = K(x', x)$ for all $x, x' \in \mathcal{X}$ and for any set of distinct points $x_1, \ldots, x_n$ in $\mathcal{X}$ the matrix

$$\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1}^N$$

is non-negative definite. We shall call the kernel K strictly positive definite if the inequality

$$\iint K(u,v)\zeta(du)\zeta(dv) > 0$$

holds for any signed measure ζ on $\mathcal{X}$ such that $0 < |\zeta|(\mathcal{X}) < \infty$.

1.2.5 Designs

The following three types of designs will be considered in this article:

1. an exact N-point design $\xi_N = \{x_1, \ldots, x_N\}$;
2. an approximate design $\xi(dx)$ corresponding to a probability measure on the design space $\mathcal{X}$;
3. a signed design defined as a signed measure $\xi(dx)$ on the design space $\mathcal{X}$ with

$$|\xi|(\mathcal{X}) = \xi^+(\mathcal{X}) + \xi^-(\mathcal{X}) < \infty.$$

We denote the spaces of exact N-point designs on $\mathcal{X}$, general approximate designs on $\mathcal{X}$ and signed design measures on $\mathcal{X}$ by Ξ_N, Ξ and $\Xi^{(S)}$, respectively.

1.2.6 Interpretation of approximate designs

Consider an approximate design ξ which is a probability measure on the design space $\mathcal{X}$. In asymptotic investigations for the case $\mathcal{X} \subseteq \mathbb{R}$ it is usually assumed that a sequence of exact designs $\xi_N = \{x_{1,N}, \ldots, x_{N,N}\}$, is generated on the base of a continuous nondecreasing function $a : [0,1] \to \mathcal{X}$ by

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

where the function a is the inverse of a distribution function corresponding to ξ. For the multi-dimensional space, we can suppose that exact designs are generated as centers of partitions of the design space assuming that the diameters of partitions tend to zero but the volumes of the cells $C_{i,N}$ in these partitions are proportional to $\xi(C_{i,N})$.

1.2.7 Methods of estimation and covariance matrices

Assume that an exact design $\xi_N = \{x_1, \ldots, x_N\}$ with a corresponding vector of observations $Y = (y_1, \ldots, y_N)^T$ is given. We consider the following three estimates of the unknown parameters θ: the best unbiased linear estimate (BLUE), ordinary least squares (OLS) and signed least squares (SLS). These estimates are respectively defined by

$$\begin{aligned} \text{BLUE} \qquad & \hat{\theta} = (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma}^{-1} Y & (3) \\ \text{OLS} \qquad & \tilde{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Y & (4) \\ \text{SLS} \qquad & \tilde{\theta}_S = (\mathbf{X}^T S\, \mathbf{X})^{-1} \mathbf{X}^T S\, Y & (5) \end{aligned}$$

where $\mathbf{X} = (f_i(x_j))_{j=1,\ldots,N}^{i=1,\ldots,m}$, $\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1,\ldots,N}$ and S is a $N \times N$ diagonal matrix with $+1$ and -1 on the diagonal. The covariance matrices of the estimates (3), (4) and (5) are given by

$$\begin{aligned} \mathrm{Var}(\hat{\theta}) &= \sigma^2 (\mathbf{X}^T \mathbf{\Sigma}^{-1} \mathbf{X})^{-1}\,, \\ \mathrm{Var}(\tilde{\theta}) &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}, \\ \mathrm{Var}(\tilde{\theta}_S) &= \sigma^2 (\mathbf{X}^T S\, \mathbf{X})^{-1} \mathbf{X}^T S\, \mathbf{\Sigma} S\, \mathbf{X} (\mathbf{X}^T S\, \mathbf{X})^{-1}, \end{aligned}$$

respectively. Note that for the BLUE, there exists a non-trivial optimal design problem in the space Ξ_N but the corresponding problem in Ξ is trivial: we simply observe the whole process, see Section 3. On the other hand, for the OLS estimate, the optimal design problems in the spaces Ξ_N and Ξ are meaningful, see Section 4. For the SLS estimate, we consider the optimal design problem in the space $\Xi^{(S)}$, see Section 2.3.

1.2.8 Optimality criteria

The following criteria of design optimality will be considered:
(a) the variance of an estimate in a one-parameter model;
(b) universal optimality of the covariance matrix;
(c) a functional of the covariance matrix, e.g. the determinant or the trace, see Sections 3 and 4;
(d) the mean square error (MSE) of the best linear predictor at some point, see Section 3.6.

1.2.9 Information matrix

To express information obtained from a design ξ_N, we use the Fisher information matrix based on the assumption of normality

$$\mathbf{M}(\xi_N) = -E\Big[\frac{\partial^2 \ln p(Y|\xi_N)}{\partial\theta\partial\theta^T}\Big]$$

where $p(Y|\xi_N)$ is the normal density with mean $(\theta^T f(x_1), \ldots, \theta^T f(x_N))$ and covariance matrix $\mathbf{\Sigma} = (K(x_i, x_j))$. Standard calculus gives that the matrix $\mathbf{M}(\xi_N)$ equals

$$\mathbf{M}(\xi_N) = \mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\,.$$

That is, the Fisher information matrix $\mathbf{M}(\xi_N)$ is the inverse of the covariance matrix of the BLUE.

A very important observation concerning the information matrix $\mathbf{M}(\xi_N)$ is the fact that in general $\mathbf{M}(\xi_N)$ cannot be decomposed into the sum of the information measures for the individual design points. This makes the problem of designing experiments for correlated observations much more difficult than in the case of uncorrelated errors.

1.2.10 Possible generalizations of the model (2)

Possible generalizations of the model (2) which are not considered in this survey include:

(a) non-linear in parameters regression models having the general response $Ey(x) = \eta(x, \theta)$, see Atkinson (2008); Dette et al. (2010); Fedorov et al. (2012);
(b) models with covariance kernel K depending on unknown parameters, see Müller and Stehlík (2004); Pázman (2010); Zimmerman (2006);
(c) random-effect and mixed-effect models including population models, see Atkinson (2008); Fedorov (1996); Schmelter (2007); Dette and Holland-Letz (2009); Dette et al. (2010); Holland-Letz et al. (2011, 2012);
(d) treatment models with search for block designs, see Cutler (1993a,b); Kiefer and Wynn (1981, 1984); Kunert et al. (2010);
(e) models where observations are split into groups corresponding to independent realizations of a stochastic process, resulting the block-diagonal structure of the matrix $\mathbf{\Sigma}$, see Dette et al. (2010); Holland-Letz et al. (2011); Schmelter (2007);
(f) models with observational noise, that is, $cov(y(x_i), y(x_j)) = \sigma^2 K(x_i, x_j) + (\sigma')^2\delta_{i,j}$, so that we can repeat observations at same points, see Bates et al. (1996); Bettinger et al. (2008).

1.3 Information contained in design points

1.3.1 Points providing zero information

To express information obtained from a design ξ_N, we use the information matrix

$$\mathbf{M}(\xi_N) = \mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X} = \sum_{i,j\in\mathcal{N}} f(x_i)g_{ij}^{\mathcal{N}}f^T(x_j)$$

where $(g_{i,j}^{\mathcal{N}}) = \mathbf{\Sigma}^{-1}$ and $\mathcal{N} = \{1,\ldots,N\}$. Despite the fact that the information matrix $\mathbf{M}(\xi_N)$ cannot be decomposed into the sum of information measures for each point, we can characterize the points that provide zero information.

Lemma 1 (Pázman (2010), Lemma 1) *Let $\xi_N = \{x_1,\ldots,x_N\}$ be a design and $\mathcal{A}$ be a subset of $\mathcal{N}$. If there exist vectors $a_i \in \mathbb{R}^m$, $i \in \mathcal{A}$, such that*

$$f(x) = \sum_{i\in\mathcal{A}} K(x,x_i)a_i$$

for all $x \in \mathrm{supp}(\xi_N)$, then all points in the set $\{x_i \mid i \notin \mathcal{A}\}$, provide zero information, that is

$$\mathbf{M}(\{x_i\}_{i\in\mathcal{A}}) = \mathbf{M}(\xi_N).$$

Corollary 1 (Pázman (2010), Corollary 1) *Let $\xi_N = \{x_1,\ldots,x_N\}$ be a design. Define the vector*

$$a_i = \sum_{j=1}^{N} g_{ij}^{\mathcal{N}} f(x_j).$$

If $a_i = 0$ for some i, then the point x_i provides zero information.

Example 1 Consider the linear regression model, that is $f(x) = (1,x)^T$, with covariance kernel $K(u,v) = \max\{0, 1-|u-v|\}$. Then the 3-point design $\{-1,0,1\}$ gives the same information as the whole process observed on $[-1,1]$ and any N-point design that includes points $-1,0,1$ in its support.

Let us now present two paradoxes which are specific for the case of correlated observations, see Näther (1985a).

1.3.2 Smit's paradox

Consider the location scale model $y(x) = \theta + \varepsilon(x)$ with the correlation function $\rho(x) = e^{-|x|}$ and the design interval $\mathcal{X} = [-\infty, \infty]$. Let us compare two estimates of the parameter θ:

- the mean for an exact design

$$\bar{\theta}_{\xi_N} = N^{-1} \sum_{i=1}^{N} y(x_i)$$

- the mean for the continuous design

$$\bar{\theta}_c = \int_{-1}^{1} y(x)dx/2.$$

For the design $\xi_5 = \{-1, -0.5, 0, 0.5, 1\}$ straightforward calculation shows that $\mathrm{Var}(\bar{\theta}_{\xi_5}) = 0.529$ while $\mathrm{Var}(\bar{\theta}_c) = 0.568$. We can see that $\mathrm{Var}(\bar{\theta}_{\xi_5}) < \mathrm{Var}(\bar{\theta}_c)$ which is called Smit's paradox [see Smit (1961)].

Consider now the design $\xi_9 = \{-1, -\frac{3}{4}, -\frac{1}{2}, \ldots, \frac{1}{2}, \frac{3}{4}, 1\}$ with 9 support points. Note that ξ_9 is obtained from ξ_5 by adding 4 points. Calculus gives that $\mathrm{Var}(\bar{\theta}_{\xi_9}) = 0.542$ and we can observe that $\mathrm{Var}(\bar{\theta}_{\xi_9}) > \mathrm{Var}(\bar{\theta}_{\xi_5})$. This means that for correlated observations the variance of the mean $\bar{\theta}_{\xi_N}$ can be increased by additional observations, which never happens in the case of independent errors.

It is worth to note that the variance of the continuous BLUE is 0.5 that is slightly smaller than $\mathrm{Var}(\bar{\theta}_{\xi_5})$. This means that the observation of a process at 5 points gives almost the same information as the continuous observation of the process.

1.3.3 Estimates with zero variance

Another interesting effect happens if we consider the location model with a correlation function such that $\rho(2) = -1$. Then for the 2-point design $\xi_2 = \{-1, 1\}$ we have $\mathrm{Var}(\bar{\theta}_{\xi_2}) = 0$. This means that the BLUE yields exactly the true value of the parameter θ which can never be in the case of uncorrelated observations. In general, estimation with zero variance is possible only if the correlation function is positive semidefinite but not positive definite.

2 Designs for one-parameter models

In this section, we consider the one-parameter model

$$y(x) = \theta f(x) + \varepsilon(x), \tag{6}$$

where θ is a scalar parameter, $f(x)$ is a continuous function on $\mathcal{X}$, $E[\varepsilon(x)] = 0$ and $E[\varepsilon(x)\varepsilon(x')] = \sigma^2 K(x, x')$.

2.1 Designs for BLUE

2.1.1 Designs for a continuous observation

Suppose that $\mathcal{X} \in \mathbb{R}$ and that an observation of the whole process $\{y(x)\}_{x\in\mathcal{X}}$ is available. The estimate $\hat{\theta}$ is called BLUE if $\hat{\theta}$ admits the representation

$$\hat{\theta} = \int_{\mathcal{X}} y(x) dG(x),$$

where G is a function of bounded variation, that is $G \in \mathrm{BV}(\mathcal{X})$, $E[\hat{\theta}] = \theta$ and

$$E(\hat{\theta} - \theta)^2 = \inf\Big\{ E\Big(\int_{\mathcal{X}} y(x) dG(x) - \theta \Big)^2 \Big| G \in \mathrm{BV}(\mathcal{X}) \ , \int_{\mathcal{X}} y(x) dG(x) = \theta \Big\}.$$

Note that the condition of unbiasedness in terms of G has the form

$$\int_{\mathcal{X}} f(x) dG(x) = 1.$$

The following result is proved in Näther (1985a), p. 19.

Theorem 1 (Näther (1985a), Theorem 2.3) *If*

$$\int_{\mathcal{X}} K(x, z) dG(x) = C f(z)$$

for all $z \in \mathcal{X}$ *and* $\int_{\mathcal{X}} f(x) dG(x) = 1$, *then the estimate*

$$\hat{\theta}(G) = \int_{\mathcal{X}} y(x) dG(x)$$

is BLUE. Moreover,

$$\mathrm{Var}(\hat{\theta}) = C.$$

The existence of the solution of the Wiener-Hopf integral equation

$$\int_{\mathcal{X}} K(x,z)dG(x) = f(z)$$

in the general case is a very hard and often ill-posed problem. Some analytic results have been obtained in the case of stationary processes having the spectral density in the form of the ratio of polynomials [see Pisarenko and Rozanov (1963) and Näther (1985a), Sec. 2.3].

2.1.2 Results of Sacks and Ylvisaker

Let $\{\xi_N\}_{N\in\mathbb{N}}$ be a sequence of designs that converges to a continuous design μ, where for each $N \in \mathbb{N}$, ξ_N is an N-point design. Then the design problem can be viewed as how the discrete BLUE $\hat{\theta}(\xi_N)$ approximates the continuous BLUE $\hat{\theta}(\mu)$. The most known asymptotic result was obtained by Sacks and Ylvisaker (1966, 1968), who studied a sequence of exact designs which is asymptotically optimal in the sense that the convergence

$$\lim_{N\to\infty} |\hat{\theta}(\xi_N) - \hat{\theta}(\mu)|^2 = 0$$

holds with the best possible convergence rate.

Suppose that the design space $\mathcal{X}$ is an interval $[a,b]$. The sequence $\{\xi_N\}$ is called asymptotic optimal for the BLUE if

$$\lim_{N\to\infty} \frac{\mathrm{Var}(\hat{\theta}(\xi_N)) - \mathrm{Var}(\hat{\theta}(\mu))}{\inf_{\xi'_N} \mathrm{Var}(\hat{\theta}(\xi'_N)) - \mathrm{Var}(\hat{\theta}(\mu))} = 1.$$

To formulate the main result of Sacks and Ylvisaker (1966), we first define

$$\alpha(x) = \lim_{z\nearrow x} \frac{\partial K(x,z)}{\partial z} - \lim_{z\searrow x} \frac{\partial K(x,z)}{\partial z}.$$

Theorem 2 (Sacks and Ylvisaker (1966)) *Assume that $\alpha(x) > 0$ and the function $f(x)$ enables the representation*

$$f(x) = \int K(z,x)h(z)dz$$

where $h(z)$ is continuous. Then the sequence $\{\xi_N\}_{N\in\mathbb{N}}$ defined by $\xi_N = \{x_{11}, \ldots, x_{NN}\}$ where x_{iN} is such that

$$\int_a^{x_{iN}} |\alpha(x)h^2(x)|^{1/3}dx = \frac{i-1}{N-1}\int_a^b |\alpha(x)h^2(x)|^{1/3}dx,$$

$i = 1, \ldots, N$, is asymptotically optimal. Moreover,

$$\mathrm{Var}\,(\hat{\theta}(\xi_N)) \to \mathrm{Var}\,(\hat{\theta}(\mu))$$

with the rate $O(N^{-2})$.

Remark 1 Consider the one-parameter model

$$y(x) = \theta x^2 + \sigma W(x),$$

where $x \in [a, b]$ and $\{W(x)\}_{x\in[a,b)}$ is a Wiener process. The equidistant design on $[a, b]$ is asymptotically optimal for estimating θ, see Sacks and Ylvisaker (1966). Moreover, for the model

$$y(x) = \theta x^\gamma + \sigma W(x),\ \gamma < 1/2,\ x \in [0, 1],$$

the design with points $x_i = (i/N)^{3/(2\gamma-1)}$, $i = 1, \ldots, N$, is asymptotically optimal.

Remark 2 Consider the location scale model with correlation structure given by the stationary Ornstein-Uhlenbeck process, i.e. $\rho(t) = e^{-\lambda|t|}$. Then equidistant designs are optimal, see Kiselak and Stehlík (2008), Zagoraiou and Baldi-Antognini (2009).

2.2 Optimal design for OLS

For a design $\xi_N \in \Xi_N$, the variance of the OLS estimate is given by

$$\mathrm{Var}(\tilde{\theta}) = \sigma^2 \sum_{i=1}^N \sum_{j=1}^N K(x_i, x_j)f(x_i)f(x_j) \Big/ \Big(\sum_{i=1}^N f(x_i)\Big)^2.$$

Consequently, for an approximate design $\xi \in \Xi$, we consider the functional

$$D(\xi) = \left[\int_{\mathcal{X}} f^2(u)\xi(\mathrm{d}u)\right]^{-2} \int_{\mathcal{X}}\int_{\mathcal{X}} K(u, v)f(u)f(v)\xi(\mathrm{d}u)\xi(\mathrm{d}v) \tag{7}$$

as the design optimality criterion. In general, the optimality criterion (7) is not convex and therefore the problem of finding the optimal design is hard. The situation is much simpler in the case of the location scale model where $f(x) = 1$ for all $x \in \mathcal{X}$.

2.2.1 Location scale model

For the function $f(x) = 1$, the design optimality criterion (7) becomes

$$D(\xi) = \int_{\mathcal{X}} \int_{\mathcal{X}} K(u, v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)\,. \tag{8}$$

Lemma 2 *The functional D defined in (8) is convex. Moreover, if the covariance kernel K is strictly positive definite, then D 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 $\mathcal{X}$ such that $\xi - \xi_0$ is a non-zero (signed) measure.

The following result serves as an equivalence theorem, which can be used to verify the optimality of a given design.

Theorem 3 (Zhigljavsky et al. (2010))

(i) A design ξ^* minimizes the functional D defined in (7) if and only if

$$\min_{x \in \mathcal{X}} b(x, \xi^*) \geq D(\xi^*). \tag{9}$$

where the function b is given by

$$b(x, \xi) = \int K(x, u)\xi(du).$$

(ii) In particular, a design ξ^* is optimal if the function $b(\cdot, \xi^*)$ is constant, that is

$$b(x, \xi^*) = D(\xi^*)$$

for all $x \in \mathcal{X}$.

In the following examples we present cases where analytical expressions for optimal designs can be found and verified using Theorem 3. In these examples we suppose that $\mathcal{X} = [-1, 1]$. Details can be found in Zhigljavsky et al. (2010).

Example 2 For the location scale model with correlation function $\rho(x) = e^{-\lambda|x|}$ 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 $\{-1,1\}$, that is, the design ξ^* has the density

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

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. Note that the function $b(\cdot, \xi^*)$ is constant and given by $D(\xi^*) = 1/(1+\lambda)$.

Example 3 For the location scale model with correlation function $\rho(x) = \max\{0, 1-\lambda|x|\}$ we have the following.

(a) For $\lambda \in \mathbb{N}$ the optimal design is a discrete uniform measure supported at the $1+2\lambda$ 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 $n = \lceil 2\lambda \rceil$,

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

the symbol $\lceil z \rceil$ stands for the smallest integer that is larger or equal to z. 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, $D(\xi^*) = 2\lambda/(n(n+1))$.

The support points for various values of λ are depicted in Figure 1.

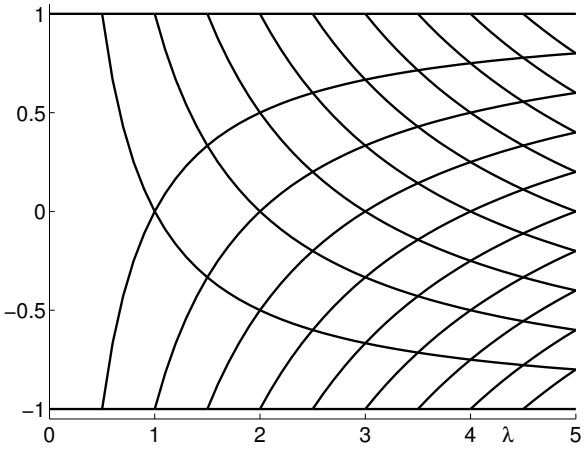

Figure 1: *Support points of the optimal designs in the location scale model with triangular correlation function* $\rho(x) = \max\{0, 1-\lambda|x|\}$ *for different values of* λ.

Example 4 Let $\rho(x) = -\ln(x^2)$. Then the asymptotic optimal design minimizing the criterion (8) is the arcsine density on the interval $[-1, 1]$ with density

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

Example 5 Let $\rho(x) = 1/|x|^\alpha$ with $0 < \alpha < 1$. Then the asymptotic optimal design minimizing the criterion (8) is a Beta distribution on the interval $[-1, 1]$ with density

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

2.2.2 Hajek result

The following result is proved by Hajek (1956).

Theorem 4 (Hajek (1956)) *Consider the location scale model with stationary error process $\{\varepsilon(x)\}_{x\in\mathcal{X}}$. Suppose that the correlation function ρ is convex on $(0, \infty)$. Let G^* be a function such that the estimate*

$$\hat{\theta} = \int_{\mathcal{X}} y(x) dG^*(x)$$

is BLUE. Then the function G^* is monotonically increasing.

Thus, for the model (6) with $f(x) = 1$ and a stationary error process having a convex correlation function, G^* is a proper distribution function corresponding to a probability measure. This means that the pair OLS plus optimal design for OLS is the best possible pair and coincides with BLUE for a continuous observation of the process.

Note that the exponential and triangular correlation functions are convex.

Example 6 For the location scale model with the exponential correlation function

$$\rho(x) = e^{-\lambda|x|}, \quad \mathcal{X} = [a, b],$$

we have that

$$dG^*(x) = \frac{1}{2 + \lambda(b-a)}[\delta_a(x) + \delta_b(x) + 1]dx$$

and

$$\hat{\theta} = \frac{1}{2 + \lambda(b-a)}[y(a) + y(b) + \int_a^b y(x)dx]$$

is BLUE, see Näther (1985a), p. 57 for details.

2.3 Optimal design for SLS

Consider the model (6) and assume that the design space is a discrete set $\mathcal{X} = \{x_1, \dots, x_n\}$ and $f(x_i) \neq 0$ for all $i = 1, \dots, n$ (if $f(x_j) = 0$ for some j then the point x_j can be removed from $\mathcal{X}$ without changing the value of the estimates). Assume also that we are using SLS and the matrix S with signs on the diagonal can be chosen along with the weights $p_1, \dots, p_n$ assigned to the points $x_1, \dots, x_n$ in $\mathcal{X}$. The design optimality criterion $D(\xi)$ becomes

$$D(\xi) = \sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) f(x_i) f(x_j) s_i s_j p_i p_j \Big/ \Big(\sum_{i=1}^{n} f(x_i) s_i p_i \Big)^2 \tag{10}$$

with $s_i \in \{-1, 1\}$ for all $i = 1, \dots, n$ and $\sum_{i=1}^{n} p_i = 1$.

Denote $w_i = s_i p_i$ and call it signed weight of a point x_i in the design ξ. Since $\sum_{i=1}^{n} |w_i| = \sum_{i=1}^{n} p_i = 1$, the signed measure ξ which assigns weights w_i to points x_i, $i = 1, \dots, n$, belongs to the space $\Xi^{(S)}$ of signed measures.

The problem of finding an optimal design and an optimal SLS estimate simultaneously in the linear regression model with one parameter consists therefore in optimizing the functional

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(x_i, x_j) f(x_i) f(x_j) w_i w_j \Big/ \Big(\sum_{i=1}^{n} f(x_i) w_i \Big)^2 \tag{11}$$

with respect to the signed weights $\{w_1, \dots, w_n\}$ such that $\sum_{i=1}^{n} |w_i| > 0$. Note that the value of the criterion (10) does not change if we change all the weights $w_i \to c w_i$ $(i = 1, \dots, n)$ for arbitrary $c \neq 0$.

Despite the fact that the functional is not convex, the problem of optimal design can be easily solved by applying the Cauchy-Schwartz inequality.

Theorem 5 *Assume that the matrix $\mathbf{\Sigma} = (K(x_i, x_j))_{i,j=1,\dots,n}$ is positive definite and $f(x_i) \neq 0$ for all $i = 1, \dots, n$. Then the optimal weights $w_1^*, \dots, w_n^*$ minimizing* (11) *subject to $\sum_{i=1}^{n} |w_i| = 1$ are given by*

$$w_i^* = c \frac{(\mathbf{\Sigma}^{-1}\mathbf{f})_i}{f(x_i)} \quad i = 1, \dots, n,$$

where $c = (\sum_{i=1}^{n} (\mathbf{\Sigma}^{-1}\mathbf{f})_i / f(x_i))^{-1}$ and $\mathbf{f} = (f(x_1), \dots, f(x_n))^T$.

For the design $\xi^* = \{x_1, \dots, x_n; w_1^*, \dots, w_n^*\}$ we have $D(\xi^*) = (\mathbf{f}\, \mathbf{\Sigma}^{-1} \mathbf{f})^{-1}$, that is, the variance of the SLS estimate coincides with the variance of the BLUE constructed using all observations. This means that the pair {SLS estimate, design ξ^*} provides the optimal pair

{estimate, design} for the problem (6). This result (in a slightly different form) is obtained in (Näther, 1985a, Theorem 5.3).

Example 7 Consider the location scale model on $\mathcal{X} = \{-1, 0, 1\}$ and the Gaussian correlation function

$$\rho(x) = e^{-x^2/2}.$$

Then we obtain the optimal signed measure

$$\xi^* = \{-1, 0, 1; 0.455, -0.09, 0.455\}$$

with $\mathrm{Var}(\hat{\theta}_{\mathrm{BLUE}}) = \mathrm{Var}(\tilde{\theta}_S|\xi^*) = 0.563$. Note that for OLS the optimal design is

$$\xi^*_{\mathrm{OLS}} = \{-1, 0, 1; 0.5, 0, 0.5\}$$

and $\mathrm{Var}(\tilde{\theta}|\xi_{\mathrm{OLS}}) = 0.568$; this is slightly larger than the variance of BLUE.

3 Optimal designs for BLUE

Consider the general model (2) with $\theta \in \mathbb{R}^m$. Recall from Section 1.2 that having N observations the BLUE of the parameter θ has the form

$$\hat{\theta} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}Y$$

which is the solution of the weighted least squares problem and the covariance matrix of BLUE is given by

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

3.1 BLUE for continuous observations

Suppose that an observation of the whole process $\{y(x)\}_{x\in\mathcal{X}}$ is available. The estimate $\hat{\theta}$ is called BLUE if $\hat{\theta}$ admits the representation

$$\hat{\theta} = \int_{\mathcal{X}} y(x)dG(x),$$

where $G(x) = (G_1(x), \ldots, G_m(x))^T$ is a vector of functions with bounded variation, $E[\hat{\theta}] = \theta$ and

$$E\|\hat{\theta} - \theta\|^2 = \inf\Big\{E\Big\|\int_{\mathcal{X}} y(x)dG - \theta\Big\|^2\Big| G \text{ such that } \int_{\mathcal{X}} y(x)dG(x) = \theta\Big\}.$$

Note that the condition of unbiasedness in terms of G has the form

$$\int_{\mathcal{X}} f(x)dG^T(x) = I_m,$$

where I_m is the $m \times m$ identity matrix.

The following result is proved in Näther (1985a), p. 19.

Theorem 6 (Näther (1985a)) *If*

$$\int_{\mathcal{X}} K(x,z)dG(x) = Cf(z)$$

for all z and

$$\int_{\mathcal{X}} f(x)dG^T(x) = I_m,$$

then the estimate

$$\hat{\theta}(G) = \int_{\mathcal{X}} y(x)dG(x)$$

is BLUE. Moreover, $\mathrm{Var}(\hat{\theta}) = C$.

Note that numerical computation of the continuous BLUE can be done as follows. First, one has to find a solution H of the integral equation

$$\int_{\mathcal{X}} K(x,z)dH(x) = f(z),$$

and then the matrix

$$C = \Big(\int_{\mathcal{X}} f(x)dH^T(x)\Big)^{-1}$$

is defined. If the matrix C is non-singular, then the BLUE is given by

$$\hat{\theta} = C\int_{\mathcal{X}} y(x)dH(x).$$

Remark 3 Note that there does not exist a design problem in the case of continuous observation since measurements are performed at all points. However, observation of a process

by exact designs can approximate continuous observation in many ways. Suppose that a sequence of exact designs ξ_N converges to a continuous measure μ and another sequence of exact designs ξ'_N converges to a continuous measure μ'. Then it follows that variances of BLUE for designs ξ_N and ξ'_N converge to the same value if the supports of the measures μ and μ' are the same.

3.2 Results of Sacks and Ylsvisaker

Let $\{\xi_N\}_{N\in\mathbb{N}}$ be a sequence of designs that converges to a continuous design μ. Then the problem is how the discrete BLUE $\hat{\theta}(\xi_N)$ approximates the continuous BLUE $\hat{\theta}(\mu)$. To formulate the main result of Sacks and Ylvisaker (1968), we first define (assuming its existence)

$$\alpha(x) = \lim_{z\nearrow x} \frac{\partial K(x,z)}{\partial z} - \lim_{z\searrow x} \frac{\partial K(x,z)}{\partial z}, \tag{12}$$

where the covariance kernel K can correspond to a stationary or non-stationary process.

In the multi-parameter case the covariance matrix $D(\xi_N)$ converges to a limiting matrix. To compare matrices corresponding to different designs, we have to use a suitable functional, for example, the L-criterion $\Psi(M) = \mathrm{tr}(LM)$ with a given positive definite matrix L. The sequence $\{\xi_N\}_{N\in\mathbb{N}}$ is called asymptotically L-optimal for the BLUE if

$$\lim_{N\to\infty} \inf_{\xi'_N} \frac{\Psi(D(\xi_N) - D(\mu))}{\Psi(D(\xi'_N) - D(\mu))} = 1.$$

In addition to the L-criterion, some other ways to define criteria of the asymptotic optimality are considered in Sacks and Ylvisaker (1968).

Theorem 7 *Assume that the function α defined in* (12) *is positive and the components of the vector $f(x) = (f_1(x), \ldots, f_m(x))^T$ satisfy the representation*

$$f_j(x) = \int K(z,x) h_j(z) dz,$$

where the functions $h_j(z)$ are continuous, $j = 1, \ldots, m$. Consider the design $\xi_N = \{x_{11}, \ldots, x_{NN}\}$ where the points x_{iN} are defined by ($h = (h_1, \ldots, h_m)^T$)

$$\int_a^{x_{iN}} |\alpha(x) h^T(x) L h(x)|^{1/3} dx = \frac{i-1}{N-1} \int_a^b |\alpha(x) h^T(x) L h(x)|^{1/3} dx,$$

$i = 1, \ldots, N$, then the sequence $\{\xi_N\}_{n\in\mathbb{N}}$ is asymptotically L-optimal.

Remark 4 In the linear and quadratic regression model with exponential covariance function $e^{-\lambda|x|}$ the exact n-point D-optimal design converges to the equally spaced design as $\lambda \to 0$, see Dette et al. (2008).

In the following sections we present three methods of numerical construction of optimal designs: the method of exchange of points, the method of virtual noise and the method using the expansion of the covariance kernel.

3.3 Exchange type algorithms

The steps of the exchange algorithm for computing an exact Ψ-optimal N-point design are as follows.

First, we have to choose a starting design $\xi_N^{(0)}$ such that the covariance matrix $D(\xi_N^{(0)})$ is non-singular. At iteration j, one point from the design $\xi_N^{(j)}$ is replaced by another point from the design space, where we need to find a replacement giving a decrease of the Ψ-criterion. The algorithm has to be stopped if the decrease is smaller than a given tolerance bound and be proceeded to the next iteration otherwise.

For the D-optimality criterion, Brimkulov et al. (1980) proposed the procedure where simultaneously a new point is introduced, which is defined by

$$x^+ = \arg \max_{x \in \mathcal{X} \setminus \xi_N^{(j)}} \phi(x, \xi_N^{(j)})$$

and a point is removed which is defined by

$$x^- = \arg \max_{x \in \xi_N^{(j)}} \phi(x, \xi_N^{(j)}).$$

Here the function ϕ is given by

$$\phi(x, \xi) = \frac{\psi^2(x, \xi) + \tilde{f}^T(x, \xi) M^{-1}(\xi) \tilde{f}^T(x, \xi)}{\psi^2(x, \xi)},$$

where

$$\psi^2(x, \xi) = K(x, x) - k^T(x, \xi) \mathbf{\Sigma}^{-1}(\xi) k(x, \xi),$$

$$k(x, \xi) = (K(x, x_1), \ldots, K(x, x_N)),$$

$$\tilde{f}(x,\xi) = f(x) - f(\xi)\mathbf{\Sigma}^{-1}(\xi)K(x,\xi).$$

Note that the procedure proposed in Brimkulov et al. (1980) is based on ideas of using an analog of the sensitivity function from the equivalence theorem for the case of uncorrelated observation. Ucinski and Atkinson (2004) have developed formulas for the straightforward exchange algorithm which provides the best decrease of the criterion as follows.

1. Select an initial design $\xi^{(0)} = \{x_1^{(0)}, ..., x_N^{(0)}\}$ such that $x_i^{(0)} \neq x_j^{(0)}$ for $i \neq j$ and $\det M(\xi^{(0)}) \neq 0$. Define the matrices $\mathbf{X}^{(0)} = (f(x_1^{(0)}), \ldots, f(x_N^{(0)}))^T$, $\mathbf{\Sigma}^{(0)} = (K(x_i^{(0)}, x_j^{(0)}))_{i,j=1}^N$ and $M^{(0)} = \mathbf{X}^{(0)^T}\left(\mathbf{\Sigma}^{(0)}\right)^{-1}\mathbf{X}^{(0)}$.
2. Set $j = 0$.
3. Determine

   $$(i^*, t^*) = \arg \max_{(i,t)\in\{1,\ldots,N\}\times\mathcal{X}} \Delta(t_i, t),$$

   where

   $$\Delta(t_i, t) = (\det M(\xi^{(j)}_{t_i \rightleftarrows t}) - \det M(\xi^{(j)}))/\det M(\xi^{(j)})$$

   and $\xi^{(j)}_{t_i \rightleftarrows t}$ denotes the design obtained from $\xi^{(j)}$ if the points t_i and t are interchanged.
4. If $\Delta(t_i, t) < \delta$, where δ is some given positive tolerance, then terminate. Otherwise, set $\xi^{(j+1)} = \xi^{(j)}_{t_j \rightleftarrows t}$ and determine $F^{(j+1)}$, $\mathbf{\Sigma}^{(j+1)}$ and $M^{(j+1)}$ corresponding to $\xi^{(j+1)}$ (expressions simplifying the numerical computation are given in Ucinski and Atkinson (2004). Set $j \Leftarrow j + 1$ and go to step 3.

   This method has been used in a number of practical examples, see Glatzer and Müller (1999); Müller (2005, 2007); Müller and Stehlík (2010); Stehlík et al. (2008).

3.4 Constructing optimal designs by expansion of the covariance kernel

Fedorov and Müller (2004) proposed to approximate the model (2) by the following mixed-effect model

$$y(x) = \theta^T f(x) + \sum_{j=1}^{q} \beta_j \psi_j(x) + \varepsilon_o(x)$$

where θ is the vector fixed-effect parameters, β_j are random-effect parameters, $\psi_j(x)$ are eigenfunctions of Mercer's expansion of the covariance kernel and $\varepsilon_o(x)$ is an error process

with no correlation. Then an optimal design is determined using the truncated D-criterion, namely, the minimization of the determinant of the covariance matrix for the parameter θ while both parameters θ and β are considered as unknown parameters. In general, the computation of optimal designs requires the knowledge of eigen functions ψ_j. Fedorov and Müller (2004) developed an approximation of the sensitivity function $\phi(x,\xi)$ which is used in the exchange algorithm for computing discrete optimal designs for models with uncorrelated observations. Specifically, for a discrete design $\xi = \{x_1,\dots,x_n; w_1,\dots,w_n\}$ the function $\phi(x,\xi)$ has the form

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

where

$$\tilde{f}^T(x,\xi) = f^T(x) + k^T(x,\xi)\mathbf{\Sigma}_{S_\xi}^{-1}(W + \mathbf{\Sigma}_{S_\xi}^{-1})^{-1} W\mathbf{X},$$

$$M = \mathbf{X}(W - W(W + \mathbf{\Sigma}_{S_\xi}^{-1})^{-1} W)\mathbf{X}^T,$$

and the matrices $\mathbf{X}$ and W are defined by $\mathbf{X} = (f(x_1),\dots,f(x_n))$, $W = \frac{N}{s^2}\mathrm{diag}\{w_1,\dots,w_n\}$, $k(x,\xi) = (K(x,x_1),\dots,K(x,x_n))$, $S_\xi = \mathrm{supp}(\xi)$, $\mathbf{\Sigma}_{S_\xi} = (K(x_i,x_j))_{i,j=1}^n$ and s^2 is a tuning parameter which should be close to zero, for example, $s^2 = 10^{-6}$.

This method has been used in a number of practical examples [see Fedorov and Flanagan (1997); Müller (2005, 2007)].

3.5 Method of virtual noise

Pázman and Müller (2001) have proposed the method of virtual noise to determine optimal designs. This method considers the following extended model

$$\tilde{y}(x) = \theta^T f(x) + \varepsilon(x) + \epsilon(x)$$

where $\varepsilon(x)$ is the original stochastic process such that $\mathrm{Cov}(\varepsilon(x_1),\varepsilon(x_2)) = \sigma^2 K(x_1,x_2)$ and $\epsilon(x)$ is an additional heteroscedastic white noise depending on a design. The two processes $\varepsilon(x)$ and $\epsilon(x)$ are assumed to be uncorrelated. For a given design $\xi = \{x_1,\dots,x_n; w_1,\dots,w_n\}$, the variance of white noise at design points is given by $\mathrm{Var}(\epsilon(x_j)) = \gamma\sigma^2 \ln(\max_i w_i/w_j)$, where γ is a tuning parameter which should be small, for example, $\gamma = 10^{-6}$.

The information matrix for the model with virtual noise is given by

$$M_\epsilon(\xi) = \mathbf{X}^T(\mathbf{\Sigma}_{S_\xi} + \gamma\mathrm{diag}(\ln[\max_i w_i/w_1],\dots,\ln[\max_i w_i/w_n]))^{-1}\mathbf{X}$$

where $\mathbf{X} = (f(x_1),\dots,f(x_n))^T$. Note that despite the presence of the weights w_i, the designs ξ in this approach can only be considered as exact N-point designs [see Pázman and Müller

(1998)].

To deal with the optimal design problem $\Phi(M_\epsilon(\xi_N)) \to \max_{\xi_N \in \Xi_N}$, several algorithms have been proposed [see Müller and Pázman (1999) and Müller and Pázman (2003)]. These algorithms consist of a stepwise one-point correction of the design. For example, in Müller and Pázman (2003) the design $\xi^{(j)}$ is updated as

$$\xi^{(j+1)} = \frac{j}{j+1}\xi^{(j)} + \frac{1}{j}\delta_{x^*}$$

where $\delta_x = \{x; 1\}$ is the one-point measure supported at $x \in \mathcal{X}$. The point x^* minimizes the directional derivative of $\Phi(M_\epsilon(\xi^{(j)}))$ in the direction of δ_x where $\xi^{(j)} = \{x_1, \dots, x_n; w_1^{(j)}, \dots, w_n^{(j)}\}$. Thus $x^* = x_{i^*}$, where

$$i^* = \arg\min_{i=1,\dots,n} \frac{1}{w_i^{(j)}} \Big(d(x_i) - \frac{\mathbf{1}_{B_j}(i)}{N_{B_j}} \sum_{k=1}^{n} d(x_k) \Big),$$

$d(x) = a^T(x)\nabla\Phi(M)a(x)$, $a(x) = \sum_{k=1}^{n} g_{ik} f(x_k)$, $\mathbf{\Sigma}^{-1} = (g_{ik})$, $M = \mathbf{X}^T\mathbf{\Sigma}\mathbf{X}$, $\nabla\Phi(M) = \partial\Phi(M)/\partial M$, $\mathbf{X} = (f_i(x_j))_{j=1,\dots,N}^{i=1,\dots,m}$, N_{B_j} is the cardinality of the set $B_j = \{i \in \{1, \dots, n\} : w_i = \max_k w_k^{(j)}\}$ and $\mathbf{1}_{B_j}(i)$ is the indicator function of the set B_j. An initial design $\xi^{(0)}$ can be chosen as the uniform discrete design supported at points forming a discretization of the design space.

This method has been used in a number of practical examples [see Müller (2005, 2007)]. A relation between the method of virtual noise and the method of the expansion of the covariance kernel is discussed in Pázman (2010); Pázman and Müller (2010).

3.6 A design for prediction in the quadratic model

Consider the quadratic model

$$y(x) = \theta_1 + \theta_2 x + \theta_3 x^2 + \sigma W(x) \tag{13}$$

where $x \geq 0$, $f(x) = (1, x, x^2)^T$, $W(x)$ is the standardized Wiener process with

$$K(u, v) = \mathrm{cov}(W(u), W(v)) = \min(u, v).$$

Note that the Wiener process is non-stationary. In addition to estimation of parameter $\theta = (\theta_1, \theta_2, \theta_3)^T$, one can be interested in prediction of the process $\{y(x)\}_{x\in[a,b]}$ at a point x,

where $x > b$ and $[a, b]$ is a design interval. The best linear unbiased predictor is given by

$$\hat{y}(x) = \hat{\theta}^T f(x) + k^T(x, \xi)\mathbf{\Sigma}^{-1}(Y - \mathbf{X}\hat{\theta})$$

where $k(x, \xi) = (K(x, x_1), \ldots, K(x, x_N))^T$, $Y = (y(x_1), \ldots, y(x_N))^T$ is a vector of observations at design points $x_1, \ldots, x_N$, and $\hat{\theta}$ is the BLUE of θ. The mean squared error of $\hat{y}(x)$ can easily be calculated as

$$\mathrm{MSE}(\hat{y}(x)) = K(x, x) - k^T(x, \xi)\mathbf{\Sigma}^{-1}k(x, \xi) + c_x^T(\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}c_x$$

where $\mathbf{X} = (f_i(x_j))_{j=1,\ldots,N}^{i=1,\ldots,m}$ and $c_x = f(x) - \mathbf{X}^T\mathbf{\Sigma}^{-1}k(x, \xi)$.

Theorem 8 (Harman and Stulajter (2011)) *Consider the process* (13) *with design interval* $[a, b]$. *Then the* N*-point design with points* $x_i = a+(b-a)(i-1)/(N-1)$, $i = 1, \ldots, N$, *is optimal for estimating the unknown parameters* $\theta_1, \theta_2, \theta_3$ *with respect to any continuous Loewner isotonic criterion as well as for the mean squared error of the best linear unbiased predictor.*

Further results on optimal designs for prediction of processes can be found in Harman and Stulajter (2010, 2011); Zimmerman (2006). Note that the design problem for prediction is a major problem in computer experiments, [see Bates et al. (1996); Dette and Pepelyshev (2010); Pronzato (2012)].

4 Optimal designs for OLS

Recall that the ordinary least squares estimate is given by $\tilde{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$ with covariance matrix

$$\mathrm{Var}(\tilde{\theta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}. \tag{14}$$

Note that the BLUE 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 OLS estimate does not employ the structure of the correlation. Obviously the OLS estimates can be less efficient than the BLUE but in many cases the loss of efficiency is small. For example, consider the location scale model with a stationary error process, the Gaussian correlation function $\rho(x) = e^{-\lambda x^2}$ and the exact design $\xi = \{-1, -2/3, -1/3, 1/3, 2/3, 1\}$. Suppose that the specified value of λ equals 1 while the true value is 2. Then the variance of the BLUE is 0.528 while the variance of the OLS estimate is 0.433. If the specified value of λ equals the true value, then the variance of the BLUE 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 Näther (1985a).

Some results on the efficiency of the OLS estimation comparing to the BLUE estimation are obtained in Kiefer and Wynn (1981); Bischoff (1995a,b) and Puntanen et al. (2011).

4.1 OLS for approximate designs

Consider the case when a continuous observation of a process is available and let ξ be an approximate design with non-singular matrix $M(\xi) = \int_{\mathcal{X}} f(x)f^T(x)\xi(dx)$. Hence the least-squares problem has the form

$$\int_{\mathcal{X}} (y(x) - \theta^T f(x))^2 \xi(dx) \to \inf_{\theta}$$

with the solution

$$\tilde{\theta} = M^{-1}(\xi) \int_{\mathcal{X}} f(x)y(x)\xi(dx)$$

which is called the continuous LSE. The covariance matrix of the estimate $\tilde{\theta}$ has the form

$$\mathrm{Var}(\tilde{\theta}) = \sigma^2 M(\xi)^{-1} B(\xi, \xi) M(\xi)^{-1},$$

where

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

The general approximate design problem is therefore given by

$$\Phi(D(\xi)) \to \min_{\xi \in \Xi},$$

where Φ is some functional on the set of $m \times m$ matrices and the matrix D is defined by

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

4.2 Results of Bickel and Herzberg

Consider the general model (2) with stationary error process. Suppose that for N observations, the correlation function is given by

$$\rho_N(t) = \rho_o(Nt) \tag{15}$$

where $\rho_o(t) = \gamma\rho(t) + (1-\gamma)\delta_t$ and $\rho(t) \to 0$ as $t \to \infty$, $\gamma \in (0, 1]$. The following regularity conditions are needed to present a main result of Bickel and Herzberg [see Bickel and Herzberg (1979)].

(C1) *The regression functions $f_1(t), \ldots, f_p(t)$ are linearly independent and bounded on the interval $[-T, T]$ and satisfy a first order Lipschitz condition, that is*

$$|f_i(t) - f_i(s)| \leq M|t - s|$$

and

$$|f_i(t)| \leq M$$

for all $t, s \in [-T, T]$, $i = 1, \ldots, p$.

(C2) *There exists a twice differentiable quantile function $a : (0, 1) \to \mathbb{R}$ and a positive constant $M < \infty$ such that for all $u \in (0, 1)$*

$$\frac{1}{M} \leq a'(u) \leq M, \quad |a''(u)| \leq M. \tag{16}$$

The quantile function a is used to generate exact N-point designs $\xi_N = \{t_{1N}, \ldots, t_{NN}\}$, that is

$$t_{iN} = a\Big(\frac{i-1}{N-1}\Big) \qquad i = 1, \ldots, N. \tag{17}$$

(C3) *the correlation function $\rho(t)$ is differentiable with bounded derivative, that is $|\rho'(t)| \leq M$, $t \in (0, \infty)$ and satisfies $\rho'(t) \leq 0$ for sufficiently large t.* This assumption implies that $\rho(t)$ is nonnegative for sufficiently large t.

(C4) *The correlation function ρ is integrable, i.e. $\int |\rho(t)|dt < \infty$. As a consequence, the function*

$$Q(t) = \sum_{j=1}^{\infty} \rho(jt) \tag{18}$$

is well defined and finite for all $t > 0$.

Theorem 9 (Bickel and Herzberg (1979)) *Consider the model* (2) *with correlation function* ρ_N *defined in* (15) *for observations at* N *points* $t_{1N}, \ldots, t_{NN}$ *defined in* (17). *Assume that the correlation function* ρ*, the quantile function* a *and the regression functions* $f_1, \ldots, f_m$ *satisfy the regularity assumptions (C1)–(C4) and suppose that the elements of the matrix*

$$R(a) = \left(\int_0^1 f_i(a(u)) f_j(a(u)) Q(a'(u))\, du \right)_{i,j=1}^m$$

exist and are finite. Then the variance-covariance matrix of the least squares estimate given in (14) *is well defined and*

$$\lim_{N\to\infty} \sigma^{-2} N\, Var(\tilde{\theta}) = W^{-1}(a) + 2\gamma W^{-1}(a) R(a) W^{-1}(a),$$

where the matrix $W(a)$ *is given by*

$$W(a) = \left(\int_0^1 f_i(a(u)) f_j(a(u))\, du \right)_{i,j=1}^m .$$

The conditions on the quantile function a imply that the corresponding design ξ has a continuous density, say $p : [0,1] \to \mathbb{R}$. Therefore, the matrices $W(a)$ and $R(a)$ can be expressed in terms of p as follows

$$R(\xi) = \left(\int_0^1 f_i(t) f_j(t) Q(1/p(t)) p(t)\, dt \right)_{i,j=1}^m ,$$

and

$$W(\xi) = \left(\int_0^1 f_i(t) f_j(t) p(t)\, dt \right)_{i,j=1}^m .$$

In the remaining part of this section we consider the one-parameter case when $f = f_1$. Define $H(t) = Q(t) - tQ'(t)$ and the function $q(x, \mu, \tau)$ by

$$q(t, \mu, \tau) = \begin{cases} \dfrac{1}{H^{-1}(\mu(1 - \tau/f^2(t)))}, & \mu(1 - \tau/f^2(t)) \geq 0, \\ 0, & \text{otherwise}, \end{cases}$$

Theorem 10 (Bickel and Herzberg (1979)) *Assume that the regression function* f *in the one-parameter linear model* (6) *is continuous and* Q *defined in* (18) *is strictly convex. Then the optimal design exists and its density is of the form* $q(t, \mu^*, \tau^*)$*, where the parameters*

μ^* and τ^* satisfy the equations

$$\int q(t, \mu^*, \tau^*)dt = 1$$

and

$$\frac{2\int Q(1/q(t))f^2(t)q(t)dt}{\int f^2(t)q(t)dt} = \mu^* - \frac{1}{2\gamma}.$$

4.3 Results for long-range dependence error structure

As in previous section, we consider the general model (2) with stationary error process having long-range dependence. Suppose that for N observations, the correlation function is given by

$$\rho_N(t) = \rho_o(Nt)$$

where $\rho_o(t) = \gamma\rho(t) + (1-\gamma)\delta_t$ and $\rho(t) \to 0$ as $t \to \infty$, $\gamma \in (0, 1]$. As in the previous section, we assume that regularity conditions (C1)–(C3) are satisfied and instead of assumption (C4) of Bickel and Herzberg (1979) we now assume that

$$\int_0^\infty |\rho(t)|\, dt = \infty. \tag{19}$$

The condition (19) means the long-range dependence of the observations. Note that in this case it follows that

$$\int_0^\infty |\rho(t)|\, dt = \sum_{k=0}^{\infty} |\rho(k)| = \infty$$

where $\rho(k) = \text{cov}(\varepsilon(t), \varepsilon(t+k))$. The correlation function of a stationary process with long range dependence can be written as

$$\rho_\alpha(t) = \frac{L(t)}{|t|^\alpha}\ ,\ |t| \to \infty \tag{20}$$

where $0 < \alpha \leq 1$ and $L(t)$ is a slowly varying function (SVF) for large t [see Doukhan et al. (2003)]. In particular ρ_α satisfies

$$\rho_\alpha(t) = O(1/|t|^\alpha)\ ,\ |t| \to \infty,$$

and we will say that $\rho_\alpha(t)$ belongs to SVF family.

At first we introduce two parametric families of correlation functions which are important in applications. The correlation function ρ_α belongs to the Cauchy family if it is defined by

$$\rho_\alpha(t) = \frac{1}{(1+|t|^\beta)^{\alpha/\beta}}, \tag{21}$$

where $\beta > 0$, $0 < \alpha \leq 1$. The correlation function ρ_α belongs to the Mittag-Leffler family if it is defined by

$$\rho_\alpha(t) = E_{\nu,\beta}(-|t|^\alpha), \quad E_{\nu,\beta}(-t) = \Gamma(\beta)\sum_{k=0}^{\infty}\frac{(-t)^k}{\Gamma(\nu k+\beta)}, \tag{22}$$

where $0 < \alpha \leq 1$, $0 < \nu \leq 1$, $\beta \geq \nu$ [see Dette et al. (2009) for more details].

In the following we present optimal designs for the three families of correlation functions, which are given by (20), (21) and (22). The function $Q(t) = \sum_{j=1}^{\infty}\rho(jt)$ plays an important role in the asymptotic analysis by Bickel and Herzberg (1979), but in the case of long range dependence this function is infinite. For an asymptotic analysis under long range dependence we introduce the function

$$Q_\alpha(t) = \lim_{N\to\infty}\frac{1}{d_\alpha(N)}\sum_{j=1}^{N}\rho_\alpha(jt), \tag{23}$$

where the normalizing sequence is given by

$$d_\alpha(N) = \begin{cases} N^{1-\alpha} & \text{if } \alpha < 1 \text{ and } \rho_\alpha \text{ has the form (21) or (22)} \\ \ln N & \text{if } \alpha = 1 \text{ and } \rho_\alpha \text{ has the form (21) or (22)} \\ L(N)N^{1-\alpha} & \text{if } \alpha < 1 \text{ and } \rho_\alpha \text{ has the form (20)} \\ L(N)\ln N & \text{if } \alpha = 1 \text{ and } \rho_\alpha \text{ has the form (20)} \end{cases}$$

and show in Lemma 3 below that the function $Q_\alpha(t)$ is well defined.

Lemma 3 (Dette et al. (2009)) *If the correlation function $\rho_\alpha(t)$ belongs either to the Cauchy, Mittag-Leffler or SVF family, then the limit in (23) exists and is given by*

$$Q_\alpha(t) = \begin{cases} \dfrac{c}{(1-\alpha)|t|^\alpha}, & 0 < \alpha < 1, \\ \frac{c}{|t|}, & \alpha = 1, \end{cases}$$

where

$$c = \begin{cases} \frac{\Gamma(\beta)}{\Gamma(\beta-\nu)}, & \text{if } \rho_\alpha(t) \text{ belongs to the Mittag-Leffler family,} \\ 1, & \text{otherwise.} \end{cases}$$

The following result describes the asymptotic behavior of the OLS for the case of the long-range dependence.

Theorem 11 (Dette et al. (2009)) *Consider the model* (2) *with correlation function* ρ_N *defined in* (15) *for observations at* N *points* $t_{1N}, \ldots, t_{NN}$ *defined by* (17). *Assume that the correlation function* ρ_α *is either an element of the Cauchy, Mittag-Leffler or SVF family. If* $\int_0^1 Q_\alpha(a'(t))\,dt < \infty$ *and the regularity assumptions (C1)–(C3) stated in the previous subsection are satisfied, then we obtain for the variance-covariance matrix of the least squares estimate defined in* (14)

$$\sigma^{-2}\frac{N}{d_\alpha(N)}\mathrm{Var}(\tilde{\theta}) = 2\gamma W^{-1}(a)R_\alpha(a)W^{-1}(a) + O(1/d_\alpha(N)),$$

where the matrices W *and* R_α *are given by*

$$W(a) = \left(\int_0^1 f_i(a(u))f_j(a(u))\,du\right)_{i,j=1}^m,$$

$$R_\alpha(a) = \left(\int_0^1 f_i(a(u))f_j(a(u))Q_\alpha(a'(u))\,du\right)_{i,j=1}^m.$$

Note that the constant γ only appears as a factor in the asymptotic variance-covariance matrices of the least squares estimate. Because most optimality criteria are positively homogeneous [see e.g. Pukelsheim (1993)] it is reasonable to consider the matrix

$$W^{-1}(a)R_\alpha(a)W^{-1}(a),$$

which is proportional to the asymptotic variance-covariance matrix of the least squares estimate. Moreover, if the function a corresponds to a continuous distribution with a density, say ϕ, then $a'(t) = 1/\phi(t)$ and the asymptotic variance-covariance matrix of the least squares estimate is proportional to the matrix

$$\Psi_\alpha(\phi) = W^{-1}(\phi)R_\alpha(\phi)W^{-1}(\phi),$$

where the matrices $W(\phi)$ and $R_\alpha(\phi)$ are given by

$$\begin{aligned} W(\phi) &= \left(\int_{-T}^{T} f_i(t) f_j(t) \phi(t)\, dt \right)_{i,j=1,\ldots,m}, \\ R_\alpha(\phi) &= \left(\int_{-T}^{T} f_i(t) f_j(t) Q_\alpha(1/\phi(t)) \phi(t)\, dt \right)_{i,j=1,\ldots,m} \\ &= \frac{c}{1-\alpha} \left(\int_{-T}^{T} f_i(t) f_j(t) \phi^{1+\alpha}(t)\, dt \right)_{i,j=1,\ldots,m}, \end{aligned}$$

respectively, and we have used the representation $Q_\alpha(t) = c/((1-\alpha)|t|^\alpha)$ for the last identity. An (asymptotic) optimal design for classical least squares estimation minimizes an appropriate function of the matrix $\Psi_\alpha(\phi)$. Note that under long range dependence the variance-covariance matrix of the least squares estimate converges slower to zero than in the case of independent or short-range dependent errors. In the case of short-range dependence, no other normalization is necessary apart from normalizing the variance-covariance matrix. Under long-range dependence an additional factor $d_\alpha(N)/N$ is needed. Moreover, it is worthwhile to note that under long range dependence the asymptotic variance-covariance matrix is fully determined by the function $Q_\alpha(t)$ and does not depend on the particular correlation function ρ_α. In the following section we discuss several examples in order to illustrate the concept.

In most cases, the asymptotic optimal designs for the regression model (2) have to be found numerically; explicit solutions are only possible for very simple models. The following result established optimal designs for linear models with one parameter.

Theorem 12 (Dette et al. (2009)) *Assume that the correlation function ρ_α is either an element of the Cauchy, Mittag-Leffler or SVF family. Then, for the one-parameter linear regression model* (6)*, the asymptotic optimal design exists, it is absolute continuous with respect to the Lebesgue measure and has the density*

$$p^*(t) = \begin{cases} \dfrac{1}{H_\alpha^{-1}(\mu - \tau/f^2(t))} = \left(\dfrac{1-\alpha}{1+\alpha} (\mu - \tau/f^2(t)) \right)^{\frac{1}{\alpha}}, & \mu - \tau/f^2(t) \geq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{24}$$

where the constants μ and τ are given by

$$\mu = 2 \frac{\int f^2(t) Q_\alpha(1/p^*(t)) p^*(t)\, dt}{\int f^2(t) p^*(t)\, dt},$$

$$\tau = \int f^2(t) Q_\alpha(1/p^*(t)) p^*(t)\, dt + \int f^2(t) Q'_\alpha(1/p^*(t))\, dt.$$

We now consider two special cases, which are of particular importance. If $p = 1$ and $f(t) \equiv 1$ we obtain the location model and the asymptotic optimal density is the uniform density, that is

$$p^*(t) = \begin{cases} \frac{1}{2T}, & |t| \le T, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

Similarly, in the linear model through the origin we have $p = 1$, $f(t) \equiv t$, and the asymptotic optimal density is given by

$$p(t) = \begin{cases} 0, & |t| \le \sqrt{\tau/\mu}, \\ \left(\frac{1-\alpha}{1+\alpha}(\mu - \tau/t^2)\right)^{1/\alpha}, & \sqrt{\tau/\mu} \le |t| \le T, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\mu = 2\frac{\int t^2 p^{1+\alpha}(t)\,dt}{(1-\alpha)\int t^2 p(t)\,dt}, \quad \tau = \int t^2 p^{1+\alpha}(t)\,dt,$$

and α is the parameter of the correlation function. The above formulas are given for $0 < \alpha < 1$. For $\alpha = 1$ and $f(t) = t$, the asymptotic optimal density is the uniform density (25). The optimal densities for the parameters $\alpha = 1/4, 1/2, 3/4, 0.95$ and $T = 1$ are displayed in Figure 2.

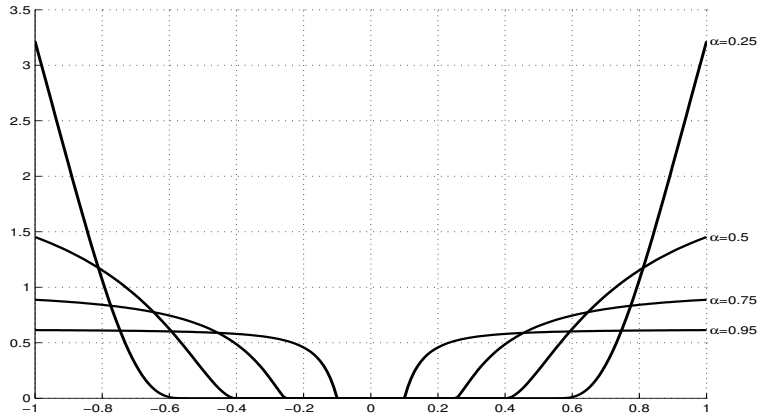

Figure 2: *Asymptotic optimal design densities on the interval* $[-1, 1]$ *for the linear regression model through the origin.*

4.3.1 Linear regression

Consider the case $p = 2$, $f_1(t) = 1$, $f_2(t) = t$, which corresponds to the linear regression model. In this case the optimal design for estimating the slope (i.e. the e_2-optimal design) has

the density (24) while the D-optimal designs have to be determined numerically in all cases. Some D-optimal design densities on the interval $[-1, 1]$ corresponding to the parameters $\alpha = 1/4, 1/2, 3/4, 0.95$ are displayed in Figure 3.

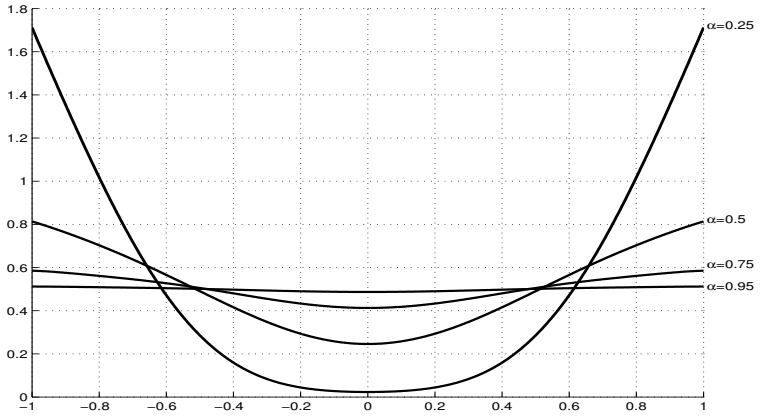

Figure 3: *Asymptotic D-optimal design densities for the linear regression model on the interval* $[-1, 1]$.

4.4 Optimal designs for OLS

In this section we consider the general model (2) with arbitrary covariance kernel. For an exact N-point design ξ_N, the covariance matrix of the least squares estimate $\tilde{\theta} = \tilde{\theta}_{\xi_N}$ given in (14) can be written as

$$\mathrm{Var}(\tilde{\theta}) = \sigma^2 D(\xi_N) = \sigma^2 \, M^{-1}(\xi_N) B(\xi_N, \xi_N) M^{-1}(\xi_N), \tag{26}$$

where the matrices M and B are given by

$$M(\xi_N) = \int_{\mathcal{X}} f(u) f^T(u) \xi_N(\mathrm{d}u), \tag{27}$$

$$B(\xi_N, \xi_N) = \int_{\mathcal{X}} \int_{\mathcal{X}} K(u, v) f(u) f^T(v) \xi_N(\mathrm{d}u) \xi_N(\mathrm{d}v). \tag{28}$$

The definition of the matrices $M(\xi)$ and $B(\xi, \xi)$ can be extended to an approximate design ξ, provided that the corresponding integrals exist. The matrix

$$D(\xi) = M^{-1}(\xi) B(\xi, \xi) M^{-1}(\xi), \tag{29}$$

is called the covariance matrix for the design ξ if the matrices $B(\xi, \xi)$ and $M^{-1}(\xi)$ are well-defined. An (approximate) optimal design minimizes a functional of the covariance matrix

$D(\xi)$ over the set Ξ.

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 Φ, of the matrix $D(\xi)$, the corresponding functional $\xi \to \Phi(D(\xi))$ is in general not convex on the space of designs Ξ. Consider for example the case $m = 1$ where $D(\xi)$ is given by

$$D(\xi) = \left[\int_{\mathcal{X}} f^2(u)\xi(\mathrm{d}u)\right]^{-2} \int_{\mathcal{X}}\int_{\mathcal{X}} K(u,v)f(u)f(v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)\,, \tag{30}$$

then it is obvious that this functional is not necessarily convex. On the other hand, for the location scale model we have $m = 1$, $f(x) = 1$ for all $x \in \mathcal{X}$ and this expression reduces to

$$D(\xi) = \int_{\mathcal{X}}\int_{\mathcal{X}} K(u,v)\xi(\mathrm{d}u)\xi(\mathrm{d}v)\,.$$

In the case $K(u,v) = \rho(u-v)$, where ρ is a correlation function, this functional is convex on the set of probability measures on the domain $\mathcal{X}$, see Lemma 1 in Zhigljavsky et al. (2010) and Lemma 4.3 in Näther (1985a). For this reason (namely the convexity of the functional $D(\xi)$) most of the literature on (asymptotic) optimal design problems for least squares estimation in the presence of correlated observations considers the location scale model. This corresponds to the estimation of the mean of a stationary process, see for example Boltze and Näther (1982); Näther (1985a,b).

Consider an optimality criterion Φ on the space of the non-negative definite matrices and define

$$\begin{aligned} \varphi(x,\xi) &= f^T(x)D(\xi)\,C(\xi)M^{-1}(\xi)f(x), &(31)\\ b(x,\xi) &= \mathrm{tr}(C(\xi)M^{-1}(\xi)B(\xi,\xi_x)M^{-1}(\xi)), &(32) \end{aligned}$$

where

$$C(\xi) = \frac{\partial\Phi(D)}{\partial D}\Big|_{D=M(\xi)}$$

(here we assume differentiability of the optimality criterion). The following result is a reformulation of necessary condition of design optimality.

Theorem 13 (Dette et al. (2012)) *Let ξ^* be any design minimizing the functional $\Phi(D(\xi))$. Then the inequality*

$$\varphi(x,\xi^*) \le b(x,\xi^*) \tag{33}$$

holds for all $x \in \mathcal{X}$, where the functions $\varphi(x,\xi)$ and $b(x,\xi)$ are defined in (31) *and* (32),

respectively. Moreover, there is equality in (33) *for* ξ^**-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* (33) *is strict.*

In this section we consider the matrix $D(\xi)$ defined in (29) as the matrix optimality criterion which we are going to minimize on the set Ξ 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 ξ^* 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 ξ^* is universally optimal if and only if ξ^* is c-optimal for any vector $c \in \mathbb{R}^m \backslash \{0\}$; that is, $c^T D(\xi^*) c \leq c^T D(\xi) c$ for any $\xi \in \Xi$ and any $c \in \mathbb{R}^m$.

For a given design $\xi \in \Xi$ with non-singular matrix $M(\xi)$, introduce the vector-valued function

$$g(x) = \int K(x, u) f(u) \xi(du) - \Lambda f(x)\,, \quad x \in \mathcal{X}\,, \tag{34}$$

where $\Lambda = B(\xi, \xi) M^{-1}(\xi)$. This function satisfies the equality

$$\int g(x) f^T(x) \xi(\mathrm{d}x) = 0\,. \tag{35}$$

Additionally, as the vector of regression functions $f(\cdot)$ is continuous on $\mathcal{X}$, the function $g(\cdot)$ is continuous too.

Theorem 14 (Dette et al. (2012)) *Consider the regression model* (1) *with covariance kernel* K*, a design* $\xi \in \Xi$ *and the corresponding vector-function* g *defined in* (34)*.*

(a) If $g(x) = 0$ *for all* $x \in \mathcal{X}$ *then the design* ξ *is universally optimal;*

(b) If the design ξ *is universally optimal then the function* g *can be represented in the form* $g(x) = \gamma(x) f(x)$*, where* $\gamma(x)$ *is a non-negative function defined on* $\mathcal{X}$ *such that* $\gamma(x) = 0$ *for all* x *in the support of the design* ξ*.*

Let us now discuss the case when the regression functions are proportional to eigenfunctions of the integral operator induced by the covariance kernel. To be precise, let $\mathcal{X}$ denote a compact subset of a metric space and let ν denote a measure on the corresponding Borel

field with positive density. Consider the integral operator

$$T_K(f)(\cdot) = \int_{\mathcal{X}} K(\cdot, u) f(u) \nu(\mathrm{d}u) \tag{36}$$

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 $\varphi_1, \varphi_2, \ldots$ with positive eigenvalues $\lambda_1, \lambda_2, \ldots$ of the operator K, that is

$$T_K(\varphi_\ell) = \lambda_\ell \varphi_\ell \ , \ \ell = 1, 2, \ldots \tag{37}$$

The next statement follows directly from Theorem 14.

Theorem 15 (Dette et al. (2012)) *Let $\mathcal{X}$ be a compact subset of a metric space and assume that the covariance kernel K defines an integral operator T_K of the form* (36)*, where the eigenfunctions satisfy* (37)*. Consider the regression model* (1) *with $f(x) = L(\varphi_{i_1}(x), \ldots, \varphi_{i_m}(x))^T$ and the covariance kernel K, where $L \in \mathbb{R}^{m \times m}$ is a non-singular matrix. Then the design ν is universally optimal in the linear regression model* (2)*.*

The following two results give optimal design in explicit form for polynomial regression models with two singular covariance kernels.

Theorem 16 (Dette et al. (2012)) *Consider the linear regression model* (1) *with $f(x) = (1, x, x^2, \ldots, x^{m-1})^T$ on the interval $\mathcal{X} = [-1, 1]$, and the covariance kernel $\rho(t) = -\ln(x^2)$. Then the arcsine design ξ_a with density*

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

is the universally optimal design.

Theorem 17 (Dette et al. (2012)) *Consider the linear regression model* (1) *with $f(x) = (1, x, x^2, \ldots, x^{m-1})^T$ on the interval $\mathcal{X} = [-1, 1]$, and covariance kernel $\rho(x) = 1/|x|^\alpha$. Then the design with generalized arcsine density*

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

is universally optimal.

5 Appendix

5.1 Proofs

Proof of Lemma 1. Let $G^{\mathcal{N}} = (g^{\mathcal{N}}_{i,j}) = (\mathbf{\Sigma}_{\mathcal{N}})^{-1}$ where $\mathbf{\Sigma}_{\mathcal{N}} = (K(x_i, x_j))_{i,j\in\mathcal{N}}$. Also let $G^{\mathcal{A}} = (g^{A}_{i,j})_{i,j\in A} = (\mathbf{\Sigma}_{A})^{-1}$ where $\mathbf{\Sigma}_{A} = (K(x_i, x_j))_{i,j\in A}$. Straightforward calculus gives

$$\begin{aligned}
\mathbf{M}(\xi_N) &= \sum_{i,j\in\mathcal{N}} f(x_i) g^{\mathcal{N}}_{ij} f^T(x_j) \\
&= \sum_{k,l\in A}\sum_{i,j\in\mathcal{N}} a_k K(x_k, x_i) g^{\mathcal{N}}_{ij} K(x_j, x_l) a_l^T \\
&= \sum_{k,l\in A} a_k K(x_k, x_l) a_l^T \\
&= \sum_{k,l\in A} f(x_k) g^{A}_{kl} f^T(x_l) = \mathbf{M}(\{x_i\}_{i\in A})
\end{aligned}$$

which completes the proof. □

Proof of Lemma 2. We have

$$\begin{aligned}
&D(\alpha\xi_2 + (1-\alpha)\xi_1) = \\
&= \iint K(u,v)[\alpha\xi_2(du) + (1-\alpha)\xi_1(du)][\alpha\xi_2(dv) + (1-\alpha)\xi_1(dv)] \\
&= (1-\alpha)^2 \iint K(u,v)\xi_1(du)\xi_1(dv) + \alpha^2 \iint K(u,v)\xi_2(du)\xi_2(dv) \\
&\quad +2\alpha(1-\alpha)\iint K(u,v)\xi_1(du)\xi_2(dv) \\
&= \alpha^2 D(\xi_2) + (1-\alpha)^2 D(\xi_1) + 2\alpha(1-\alpha)\iint K(u,v)\xi_1(du)\xi_2(dv) \\
&= \alpha D(\xi_2) + (1-\alpha)D(\xi_1) - \alpha(1-\alpha)A\,,
\end{aligned}$$

where

$$\begin{aligned}
A &= \iint K(u,v)[\xi_2(du)\xi_2(dv) + \xi_1(du)\xi_1(dv) - 2\xi_2(du)\xi_1(dv)] \\
&= \iint K(u,v)\zeta(du)\zeta(dv)
\end{aligned}$$

and $\zeta(du) = \xi_2(du) - \xi_1(du)$. Since the correlation function $K(u,v)$ is positive definite, it follows $A \geq 0$. If $K(u,v)$ is strictly positive definite, we have $A > 0$ whenever ζ is not trivial. Therefore the functional $D(\cdot)$ is strictly convex. □

Proof of Theorem 5. Denote $K_{ij} = K(x_i, x_j)$, $f(x_i) = f_i$, $a_i = f_i w_i$, $i, j = 1, \ldots, n$, $a = (a_1, \ldots, a_n)^T$. Then for any signed design $\xi = \{x_1, \ldots, x_n; w_1, \ldots, w_n\}$ we have

$$D(\xi) = \frac{\sum_i \sum_j K_{ij} f_i f_j w_i w_j}{(\sum_i f_i^2 w_i)^2} = \frac{\sum_i \sum_j K_{ij} a_i a_j}{(\sum_i f_i a_i)^2} = \frac{a^T \mathbf{\Sigma} a}{(a^T \mathbf{f})^2}.$$

Since $\mathbf{\Sigma}$ is symmetric and $\mathbf{\Sigma} > \mathbf{0}$, there exists $\mathbf{\Sigma}^{-1}$ and a symmetric matrix $\mathbf{\Sigma}^{1/2} > 0$ such that $\mathbf{\Sigma} = \mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2}$. Denote $b = \mathbf{\Sigma}^{1/2} a$ and $d = \mathbf{\Sigma}^{-1/2}\mathbf{f}$. Then we can write $D(\xi)$ as $D(\xi) = b^T b/(b^T d)^2$. The Cauchy-Schwartz inequality gives for any two vectors b and d: $(b^T d)^2 \le (b^T b)(d^T d)$, that is, $b^T b/(b^T d)^2 \ge 1/(d^T d)$. This inequality with b and d as above is equivalent to

$$D(\xi) \ge \frac{1}{\mathbf{f}\, \mathbf{\Sigma}^{-1} \mathbf{f}}$$

for all $\xi \in \Xi^{(S)}$. The equality in this inequality is attained if the vector b is proportional to the vector d; that is, if $b_i = c d_i$ for all i and any $c \neq 0$. Then the equality $b_i = c d_i$ can be rewritten in the form $w_i = c(\mathbf{\Sigma}^{-1}\mathbf{f})_i / f(x_i)$. □

5.2 Common correlation functions

Definition 1 *(Stationarity in the wide sense) A random field is a stationary random field in the wide sense if* $\mathbf{E}\varepsilon(u) = const$ *and* $K(u, v) := \mathrm{cov}(\varepsilon(u), \varepsilon(u)) = \rho(u - v)$.

The covariance function $\rho(x)$ *on* $\mathbb{R}^d$ *is fully separable if* $\rho(x) = \rho_1(x_1) \cdots \rho_d(x_d)$. *Note that the product and the sum of two covariance functions are also covariance functions, see Abrahamsen (1997), Sec 3.1.*

Definition 2 *A stationary random field is an isotropic random field (in the wide sense) if the covariance function depends on distance alone, i.e.*

$$K(u, v) = \rho(\|u - v\|)$$

and $\|x\| = \sqrt{x^T x}$.

Definition 3 *A stationary random field is an anisotropic random field (in the wide sense) if the covariance function depends on a non-Euclidian norm of the difference of two points, i.e.*

$$K(u, v) = \rho(\|u - v\|_A),$$

where $\|x\|_A = \sqrt{x^T A x}$ *and* A *is a positive semi-definite matrix. The function* $\rho(\|u - v\|_A$ *is called an ellipsoidal correlation function.*

In Table 1 we present correlation functions for one-dimensional case. In the case of $d > 1$, most common isotropic correlation functions have the form $\rho(\|x\|)$.

Table 1: Commonly used correlation functions, $\lambda > 0$, $R > 0$.

Name	$\rho(x)$
exponential	$e^{-\lambda x}$
Gaussian	$e^{-\lambda x^2}$
rational quadratic	$\dfrac{1}{(1+\lambda x^2)^\nu}$, $\nu > 0$
spherical, triangular	$\left(1 - \frac{x}{R}\right) \mathbf{1}_{[0,R]}(x)$
spherical, circular	$\left(1 - \frac{2}{\pi}\left(\frac{x}{R}\sqrt{1-\frac{x^2}{R^2}} + \arcsin\left(\frac{x}{R}\right)\right)\right) \mathbf{1}_{[0,R]}(x)$
spherical	$\left(1 - \frac{3x}{R} + \frac{x^3}{2R^3}\right) \mathbf{1}_{[0,R]}(x)$
penta-spherical	$\left(1 - \frac{15x}{8R} + \frac{5x^3}{4R^3} - \frac{3x^5}{R^5}\right) \mathbf{1}_{[0,R]}(x)$
cubic	$\left(1 - \frac{7x^2}{R^2} + \frac{35x^3}{4R^3} - \frac{7x^5}{2R^5} + \frac{3x^7}{4R^7}\right) \mathbf{1}_{[0,R]}(x)$
stable	$\exp\left(-\lambda x^\nu\right)$, $0 < \nu \leq 2$
oscillating, damped cosine	$e^{-\lambda x}\cos(\omega x)$, $\omega > 0$
oscillating, hole effect	$\dfrac{1}{\omega x}\sin(\omega x)$, $\omega > 0$
oscillating, Bessel	$\Gamma(\nu+1)2^\nu(\lambda x)^{-\nu}J_\nu(\lambda x)$, $\nu \geq (d+1)/2$
Poisson	$\dfrac{1-\beta^2}{1-2\beta\cos(2\pi x)+\beta^2}$, $0 < \beta < 1$
Cauchy family	$\dfrac{1}{(1+\lvert x\rvert^\beta)^{\alpha/\beta}}$, $\beta > 0$, $0 < \alpha \leq 1$
Mittag-Leffler family	$E_{\nu,\beta}(-\lvert x\rvert^\alpha)$, $0 < \alpha \leq 1$, $0 < \nu \leq 1$, $\beta \geq \nu$ $E_{\nu,\beta}(-t) = \Gamma(\beta)\sum_{k=0}^{\infty} \dfrac{(-t)^k}{\Gamma(\nu k+\beta)}$ $E_{1,1}(-t) = e^{-t}$, $E_{1,2}(-t) = (1-e^{-t})/t$, $E_{1,3}(-t) = 2(e^{-t}-1+t)/t^2$, $E_{1/2,1}(-t) = e^{t^2}\left(1 - \frac{2}{\sqrt{\pi}}\int_0^t e^{-u^2}\,du\right)$
scaling	$(1-x^2)e^{-x^2/2}$
singular logarithmic	$-\ln x^2$
singular rational	$\dfrac{1}{\lvert x\rvert^\alpha}$, $0 < \alpha < 1$

Acknowledgements. The authors thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. 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).

References

Abrahamsen, P. (1997). A review of Gaussian random fields and correlation functions. Technical report 917, Norwegian Computing Center, Osla.

Atkinson, A. C. (2008). Examples of the use of an equivalence theorem in constructing optimum experimental designs for random-effects nonlinear regression models. Journal of Statistical Planning and Inference, 138(9):2595–2606.

Bates, R. A., Buck, R. J., Riccomagno, E., and Wynn, H. P. (1996). Experimental design and observation for large systems. Journal of the Royal Statistical Society, Ser. B, 58:77–94.

Bettinger, R., Duchene, P., Pronzato, L., and Thierry, E. (2008). Design of experiments for response diversity. Journal of Physics: Conference Series, 135:12–17.

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.

Bischoff, W. (1995a). Determinant formulas with applications to designing when the observations are correlated. Annals of the Institute of Statistical Mathematics, 47:385–399.

Bischoff, W. (1995b). Lower bounds for the efficiency of designs with respect to the D-criterion when the observations are correlated. Statistics, 27:27–44.

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

Brimkulov, U. N., Krug, G. V., and Savanov, V. L. (1980). Numerical construction of exact experimental designs when the measurements are correlated (in russian). Zavodskaya Laboratoria, 36:435–442.

Cambanis, S. (1985). Sampling designs for time series. In Hannan, E. J., Krishnaiah, P. R., and Rao, M. M., editors, Handbook of Statistics, pages 337–362. Elsevier, Amsterdam.

Cutler, D. R. (1993a). Efficient block designs for comparing test treatments to a control when the errors are correlated. Journal of Statistical Planning and Inference, 36:107–125.

Cutler, D. R. (1993b). Efficient block designs for comparing test treatments to a control when the errors are correlated. Journal of Statistical Planning and Inference, 37:393–412.

Dette, H. and Holland-Letz, T. (2009). A geometric characterization of c-optimal designs for heteroscedastic regression. Annals of Statistics, 37(6).

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

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

Dette, H. and Pepelyshev, A. (2010). Generalized Latin hypercube design for computer experiments. Technometrics, 52:421–429.

Dette, H., Pepelyshev, A., and Holland-Letz, T. (2010). Optimal designs for random effect models with correlated errors with applications in population pharmacokinetics. Annals of Applied Statistics, 4:1430–1450.

Dette, H., Pepelyshev, A., and Zhigljavsky, A. (2012). Optimal design for linear models with correlated observations. Submitted for publication in Annals of Statistics.

Doukhan, P., Oppenheim, G., and Taqqu, M. (2003). Theory and Applications of Long-Range Dependence. Birkhauser, Boston.

Fedorov, V. and Flanagan, D. (1997). Optimal monitoring network design based on Mercers expansion of covariance kernel. Journal of Combinatorics, Information and System Sciences, 23:237–250.

Fedorov, V. and Müller, W. (2004). Optimum design for correlated processes via eigenfunction expansions.

Fedorov, V. and Müller, W. (2007). Optimum Design for Correlated Fields via Covariance Kernel Expansions. In López-Fidalgo, J., Rodríguez-Díaz, J. M., and Torsney, B., editors, mODa 8 - Advances in Model-Oriented Design, pages 57–66. Physica-Verlag, Heidelberg.

Fedorov, V., Wu, Y., and Zhang, R. (2012). Optimal dose-finding designs with correlated continuous and discrete responses. Statistics in Medicine, 31:217–234.

Fedorov, V. V. (1996). Design of spatial experiments: model fitting and prediction. In Gosh, S. and Rao, C. R., editors, Handbook of Statistics, pages 515–553. Elsevier, Amsterdam.

Glatzer, E. and Müller, W. G. (1999). A comparison of optimum design algorithms for regressions with correlated observations—a computational study. Tatra Mt. Math. Publ., 17:149–156. PROBASTAT '98 (Smolenice Castle).

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 Stulajter, F. (2010). Optimal prediction designs in finite discrete spectrum linear regression models. Metrika, 72:281–294.

Harman, R. and Stulajter, F. (2011). Optimality of equidistant sampling designs for the Brownian motion with a quadratic drift. Journal of Statistical Planning and Inference, 141:2750–2758.

Holland-Letz, T., Dette, H., and Pepelyshev, A. (2011). A geometric characterization of optimal designs for regression models with correlated observations. Journal of the Royal Statistical Society, Ser. B, 73:239–252.

Holland-Letz, T., Dette, H., and Renard, D. (2012). Efficient algorithms for optimal designs with correlated observations in pharmacokinetics and dose finding studies. Biometrics, 68:138–145.

Kanwal, R. (1997). Linear Integral Equations. Birkhauser, Boston.

Kiefer, J. and Wynn, H. (1981). Optimum balanced block and latin square designs for correlated observations. Annals of Statistics, 9:737–757.

Kiefer, J. and Wynn, H. (1984). Optimum and minimax exact treatment designs for one-dimensional autoregressive error processes. Annals of Statistics, 12:414–450.

Kiselak, J. and Stehlík, M. (2008). Equidistant D-optimal designs for parameters of Ornstein-Uhlenbeck process. Statistics and Probability Letters, 78:1388–1396.

Kunert, J., Martin, R. J., and Eccleston, J. (2010). A-optimal block designs for the comparison with a control for correlated errors and analysis with the weighted least squares estimate. Journal of Statistical Planning and Inference, 140:2719–2738.

Müller, W. G. (2005). A comparison of spatial design methods for correlated observations. Environmetrics, 16(5):495–505.

Müller, W. G. (2007). Collecting spatial data. Contributions to Statistics. Physica-Verlag, Heidelberg, revised edition. Optimum design of experiments for random fields.

Müller, W. G. and Pázman, A. (1999). An algorithm for the computation of optimum designs under a given covariance structure. Computational Statistics, 14:197–211.

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

Müller, W. G. and Stehlík, M. (2004). An example of D-optimal designs in the case of correlated errors. In COMPSTAT 2004—Proceedings in Computational Statistics, pages 1543–1550. Physica, Heidelberg.

Müller, W. G. and Stehlík, M. (2010). Compound optimal spatial designs. Environmetrics, 21:354–364.

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. (2010). Information contained in design points of experiments with correlated observations. Kybernetika, 46:771–783.

Pázman, A. and Müller, W. G. (1998). A new interpretation of design measures. In MODA 5—advances in model-oriented data analysis and experimental design (Marseilles, 1998), Contrib. Statist., pages 239–246. Physica, Heidelberg.

Pázman, A. and Müller, W. G. (2001). Optimal design of experiments subject to correlated errors. Statist. Probab. Lett., 52(1):29–34.

Pázman, A. and Müller, W. G. (2010). A note on the relationship between two approaches to optimal design under correlation. In Giovagnoli, A., Atkinson, A. C., Torsney, B., and May, C., editors, mODa 9 Advances in Model-Oriented Design and Analysis, Contributions to Statistics, pages 145–148. Physica-Verlag HD.

Pisarenko, V. F. and Rozanov, J. A. (1963). On certain problems for stationary processes leading to integral equations related to Wiener-Hopf equations. Problemy Peredaci Informacii, 14:113–135.

Pronzato, L. und Müller, W. (2012). Design of computer experiments: space filling and beyond. Statistics and Computing, 22:681–701.

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

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

Puntanen, S., Styan, G., and Isotalo, J. (2011). Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty. Heidelberg, Springer.

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.

Schmelter, T. (2007). Considerations on group-wise identical designs for linear mixed models. Journal of Statistical Planning and Inference, 137:4003–4010.

Smit, J. C. (1961). Estimation of the mean of a stationary stochastic process by equidistant observations. Trabojos de Estadistica, 12:34–45.

Stehlík, M., Rodríguez-Díaz, J. M., Müller, W. G., and López-Fidalgo, J. (2008). Optimal allocation of bioassays in the case of parametrized covariance functions: an application to lung's retention of radioactive particles. TEST, 17(1):56–68.

Ucinski, D. and Atkinson, A. (2004). Experimental design for time-dependent models with correlated observations. Studies in Nonlinear Dynamics & Econometrics, 8(2):13.

Zagoraiou, M. and Baldi-Antognini, A. (2009). Optimal designs for parameter estimation of the Ornstein-Uhlenbeck process. Applied Stochastic Models in Business and Industry, 25:583–600.

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.

Zimmerman, D. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics, 17:635–652.