Extended Information Matrices for Optimal Designs when the Observations are Correlated

Pazman, Andrej; Müller, Werner

Published: 01/01/1996

Document Version
Publisher's PDF, also known as Version of record

Extended Information Matrices for Optimal Designs when the Observations are Correlated

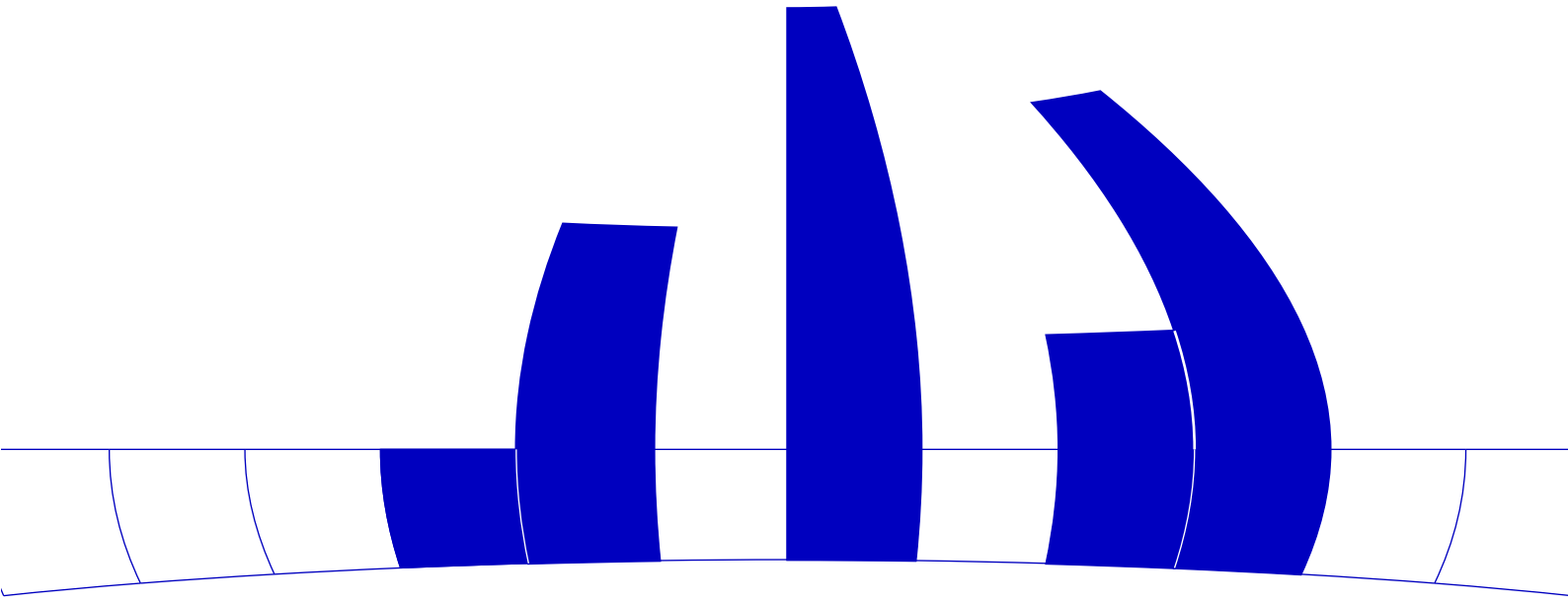
Andrej Pázman, Werner Müller

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 53
May 1996

http://statmath.wu-wien.ac.at/



Extended Information Matrices for Optimal Designs when the Observations are Correlated

By ANDREJ PÁZMAN [1] and WERNER G. MÜLLER

Comenius University Bratislava *University of Economics and B.A. Vienna*

SUMMARY

Regression models with correlated errors lead to nonadditivity of the information matrix. This makes the usual approach of design optimization (approximation with a continuous design, application of an equivalence theorem, numerical calculations by a gradient algorithm) impossible. A method is presented that allows the construction of a gradient algorithm by altering the information matrices through adding of supplementary noise. A heuristic is formulated to circumvent the nonconvexity problem and the method is applied to typical examples from the literature.

Keywords: optimum design, correlated errors, extended information matrices, gradient algorithm.

1 Introduction

The most elegant results in the domain of optimal designs when the observations are correlated have been obtained in a series of papers by Sacks and Ylvisaker (1966, 1968, 1970). They compare sequences of n-point designs according to the rate of increase of the information obtained from the corresponding experiments when n tends to infinity, and they construct sequences of designs which are optimal in this sense. Although their approach had followers (e.g. Wahba (1971), Abt (1992) and recently Müller-Gronbach (1996)), the inconvenience of their method is not only that just the asymptotical rate is considered, but also that their results depend very much on the derivatives of the (known) covariance function. But frequently we know the covariance function just approximately, and therefore have no idea about its derivative. A simple direct approach is possible when the regression function is generated linearly by the covariance function (see Näther (1985b)). Another suggestion goes back to Brimkulov *et al.* (1980), who extend ideas from well-known algorithmic approaches to the correlated case. The idea of Bischoff (1993) is to compare optimality criteria with and without correlation, which underlines how difficult it is to solve the correlated case. For a recent survey of some alternative methods see Fedorov (1996).

[1] *Address for correspondence:* Department of Probability and Statistics, Faculty of Mathematics and Physics, Comenius University Bratislava, Mlynská dolina, 84215 Bratislava, Slovak Republic. The first author was supported by a grant of the Slovak Grant Agency.

However, it seems there is no approach to the problem which uses design measures, although the results of the classical (uncorrelated) theory makes this concept so attractive. This task is considered as unsolvable by specialists; e.g. in the book by Näther (1985a) on p. 84 it is written that "...the corresponding information matrix [...] allows no generalization in the direction of design measures, at least not in the same manner as known from Kiefer's theory". The aim of the present paper is to show that it is still possible to associate some "extended" information matrices based on an approach introduced by Müller and Pázman (1995) to each design measure. Uniform design measures correspond then to exact designs, and in this case the "extended" information matrix coincides with the standard one. We construct a smoothing of the information matrix depending upon a real parameter. In contrast to Müller and Pázman (1995), where a smoothing was defined for uncorrelated observations without replications, now we present another smoothing by adding a supplementary white noise. We check the practical effectiveness of this smoothing by constructing the gradient algorithm and by computing optimal or nearly optimal designs. In particular we recalculate some well known examples from the references.

2 Regression models when the observations are correlated

We assume that we observe random variables $y(x)$ depending upon regressors $x \in \mathcal{X}$ such that

$$E[y(x)] = f^T(x)\theta,$$

where $f(.)$ is fixed and $\theta = \{\theta_1, \ldots, \theta_m\} \in I\!R^m$ is the vector of unknown parameters. The observations are correlated according to the covariance function $\mathrm{Cov}[y(x), y(z)] = C(x, z)$. One can think about $\{y(x) : x \in \mathcal{X}\}$ as a random field.

We consider experiments where the variables $y(x)$ are observed only on a finite number of points $x_1, \ldots, x_n \in \mathcal{X}$. The set $A = \{x_1, \ldots, x_n\}$ constitutes what is called an exact design in the literature (see e.g. Pázman (1986)). The information matrix corresponding to the exact design A is

$$M(A) = f(A)C^{-1}(A)f^T(A) \tag{2.1}$$

(see Müller and Pázman (1995), Section 2.1 for the notation). In there we extended the notion of an information matrix to every design measure ξ on $\mathcal{X}$, by associating to each ξ the matrix $M(S_\xi)$, where $S_\xi = \{x \in \mathcal{X}; \xi(x) > 0\}$. Moreover, for each $\gamma > 0$ we defined an approximate information matrix

$$M^{(\gamma)}(\xi) = \sum_{x \in \mathcal{X}} \sum_{z \in \mathcal{X}} \left(\frac{\xi(x)}{\xi_{\max}}\right)^\gamma f(x)C^{-1}(S_\xi)f^T(z) \left(\frac{\xi(z)}{\xi_{\max}}\right)^\gamma, \tag{2.2}$$

where $\xi_{\max} = \max\{\xi(x); x \in \mathcal{X}\}$.

Let us reconsider the role of the design measure in this matrix. The observed process $y(x)$ contains two components. The useful component, which is the mean of the process

$f^T(x)\theta$ (we may call it the signal), and the nuisance component, which is the correlated random noise $\epsilon(x) = y(x) - f^T(x)\theta$. We see that by taking $\xi(x) < \xi_{\max}$ we suppress the influence of $f(x)$ in the expression for $M^{(\gamma)}(\xi)$ (it is like taking $\left(\frac{\xi(x)}{\xi_{\max}}\right)^\gamma f(x)$ instead of $f(x)$). Extreme cases are $\xi(x) = \xi_{\max}$ (no suppression) and $\xi(x) = 0$ (total suppression of the signal). In other words, there is no information when $\xi(x) = 0$ and there is only a fraction of the information when $\xi(x) < \xi_{\max}$.

This suppression of the signal component works well when the correlations between observations $y(x)$ at different points $x \in \mathcal{X}$ are zero or small. However, if observations at different points are highly correlated, then observations in points with completely suppressed signal still give some information about the noise at other points. So they contribute indirectly to the increase of information about the parameter θ. This is why the method from Section 2.1 of Müller and Pázman (1995) does not work for highly correlated observations.

3 The approximate information matrices for arbitrarily correlated observations without replications

In this section we suppose that $\mathcal{X}$ is finite and $C(\mathcal{X})$ is regular. We have to find another way as in Müller and Pázman (1995) to reduce the information at a point x if $\xi(x) < \xi_{\max}$. The idea is to add supplementary uncorrelated noise $\epsilon^*(x)$ to the original noise $\epsilon(x)$ if $\xi(x) < \xi_{\max}$. The smaller $\frac{\xi(x)}{\xi_{\max}}$, the larger has to be the variance of $\epsilon^*(x)$. So let the mean and variance of $\epsilon^*(x)$ be

$$E[\epsilon^*(x)] = 0,$$

$$\mathrm{Var}[\epsilon^*(x)] = \ln\left(\frac{\xi_{\max}}{\xi(x)}\right)^\gamma. \tag{3.1}$$

If $\xi(x) = \xi_{\max}$ then $\mathrm{Var}[\epsilon^*(x)] = 0$, hence also $\epsilon^*(x) = 0$ and at such x there is no additional noise. If $\xi(x) \to 0$ then $\mathrm{Var}[\epsilon^*(x)] \to \infty$, hence $f^T(x)\theta$ but also $\epsilon(x)$ is totally dominated by $\epsilon^*(x)$ and there is no information at all from the observation at x. In the cases which are between, i.e. $0 < \xi(x) < \xi_{\max}$, the additional noise $\epsilon^*(x)$ suppresses only a part of the information from an observation at x.

The variance-covariance matrix of $\epsilon^*(x)$ with $x \in S_\xi$ is a diagonal matrix $W^{(\gamma)}(\xi)$ with entries given by (3.1). The total information matrix is

$$\mathcal{M}^{(\gamma)}(\xi) = \sum_{x \in S_\xi} \sum_{z \in S_\xi} f(x)[C(S_\xi) + W^{(\gamma)}(\xi)]^{-1}_{x,z} f^T(z). \tag{3.2}$$

According to (3.2) we observe at each $x \in S_\xi$, but when $\xi(x) < \xi_{\max}$ we have "fractional observations". To extend definition (3.2) to the case that $\xi(x)$ is equal

zero at some points $x \in \mathcal{X}$, we have to use the following matrix lemma (refer e.g. to Kubáček *et al.* (1995)):

Lemma 1: Let $U = \begin{pmatrix} U_I & U_{II} \\ U_{II}^T & U_{III} \end{pmatrix}$ be a symmetric, positive definite matrix, then

$$U^{-1} = \begin{pmatrix} (U_I - U_{II}U_{III}^{-1}U_{II}^T)^{-1} & -(U_I - U_{II}U_{III}^{-1}U_{II}^T)^{-1}U_{II}U_{III}^{-1} \\ -U_{III}^{-1}U_{II}^T(U_I - U_{II}U_{III}^{-1}U_{II}^T)^{-1} & U_{III}^{-1} + U_{III}^{-1}U_{II}^T(U_I - U_{II}U_{III}^{-1}U_{II}^T)^{-1}U_{II}U_{III}^{-1} \end{pmatrix}$$

If U is positive semidefinite then each inverse can be substituted by a g-inverse.

Theorem 1: Let ξ_n, $n = 1, 2, \ldots$ be a sequence of design measures having the same support $S = S_{\xi_n}$ such that it converges to a design measure μ, i.e.

$$\lim_{n\to\infty} \xi_n(x) = \mu(x); \quad x \in \mathcal{X}.$$

Then

$$\lim_{n\to\infty} \mathcal{M}^{(\gamma)}(\xi_n) = \mathcal{M}^{(\gamma)}(\mu).$$

Proof: Denote $S = \{x_1, \ldots, x_s\}$ and $S_\mu = \{x_1, \ldots, x_a\}$, where $a \leq s$. If $a = s$, the proof is evident. If $a < s$, we shall do a step by step proof using the lemma from above. Take $U = C(S) + W^{(\gamma)}(\xi_n)$ and denote U_I the upper left $(s-1) \times (s-1)$ block of this matrix. Then U_{III}^{-1} is a number

$$U_{III}^{-1} = \left[C(x_s, x_s) + \gamma \ln \left(\frac{\xi_{\max}}{\xi_n(x_s)} \right) \right]^{-1},$$

which tends to zero since $\xi_n(x_s) \to 0$. From Lemma 1 it follows that

$$\lim_{n\to\infty} [C(S) + W^{(\gamma)}(\xi_n)]^{-1} = \begin{pmatrix} \lim_{n\to\infty} [C(S - \{x_s\}) + W^{(\gamma)}(\xi_n)_I]^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

We use the same procedure $s - a$ times to obtain finally

$$\lim_{n\to\infty} [C(S) + W^{(\gamma)}(\xi_n)]^{-1} = \begin{pmatrix} [C(S_\mu) + W^{(\gamma)}(\mu)_L]^{-1} & 0 \\ 0 & 0 \end{pmatrix},$$

where $W^{(\gamma)}(\mu)_L$ is an $a \times a$ upper left submatrix of $W^{(\gamma)}(\mu)$. The proof is finished by using the definition of $M^{(\gamma)}(\mu)$. □

Proposition 1:

a) for every γ positive and any design $A \subset \mathcal{X}$ we have $\mathcal{M}^{(\gamma)}(\xi_A) = M(A)$,

b) for every design measure ξ

$$\lim_{\gamma\to 0} \mathcal{M}^{(\gamma)}(\xi) = M(S_\xi)$$

$$\lim_{\gamma\to\infty} \mathcal{M}^{(\gamma)}(\xi) = M(B_\xi)$$

where $B_\xi = \{x \in \mathcal{X} : \xi(x) = \xi_{\max}\}$.

Proof: Point (a) follows directly from definition (3.2). Point (b) follows from Lemma 1 and from

$$\lim_{\gamma\to 0}\left(\frac{\xi(x)}{\xi_{\max}}\right)^{\gamma} = \begin{cases} 1 & \text{if } x \in S_\xi \\ 0 & \text{if } x \notin S_\xi \end{cases},$$

$$\lim_{\gamma\to \infty}\left(\frac{\xi(x)}{\xi_{\max}}\right)^{\gamma} = \begin{cases} 1 & \text{if } x \in B_\xi \\ 0 & \text{if } x \notin B_\xi \end{cases}.$$

□

Remark. Part a) in Proposition 1 means that $\mathcal{M}^{(\gamma)}(\xi)$ is an extention of the information matrix (1). Part b) is an indication to consider $\mathcal{M}^{(\gamma)}(\xi)$ for a small γ, since $M(S_\xi)$ is a natural but discontinuous extention of the infromation matrix, as presented in Müller and Pázman (1995).

Proposition 2: If ξ and η are such design measures that

$$\frac{\xi(x)}{\xi_{\max}} \le \frac{\eta(x)}{\eta_{\max}} \quad \forall x \in \mathcal{X}$$

then for every positive γ

$$\mathcal{M}^{(\gamma)}(\xi) \le \mathcal{M}^{(\gamma)}(\eta)$$

in the Loewner ordering. In particular

$$M(B_\xi) \le \mathcal{M}^{(\gamma)}(\xi) \le M(S_\xi).$$

Proof: The inverse of a p.d. matrix is also p.d.. and if $U \ge V$ and U, V are both p.d. then $U^{-1} \le V^{-1}$, see e.g. Pukelsheim (1993). Thus the first statement is immediate from the definition of $\mathcal{M}^{(\gamma)}(\xi)$. The second statement follows from the evident inequality

$$\frac{\xi_{B_\xi}(x)}{(\xi_{B_\xi})_{\max}} \le \frac{\xi(x)}{\xi_{\max}} \le \frac{\xi_{S_\xi}(x)}{(\xi_{S_\xi})_{\max}}.$$

□

4 The first order derivative for $S_\xi = \mathcal{X}$

The expression (3.2) is not differentiable, because $\xi_{\max}$ is not differentiable. Thus $\xi_{\max}$ is approximated for small γ by

$$\xi^{(\gamma)} = [\sum_{x\in\mathcal{X}} \xi^{\frac{1}{\gamma}}(x)]^{\gamma}$$

like in Müller and Pázman (1995).

Then instead of (3.2) we have

$$\tilde{\mathcal{M}}^{(\gamma)}(\xi) = \sum_{x,z\in\mathcal{X}} f(x)[C(\mathcal{X}) + \tilde{W}^{(\gamma)}(\xi)]^{-1}_{x,z} f^T(z), \tag{4.1}$$

where $\tilde{W}^{(\gamma)}(\xi)$ has the entries $\ln\left(\frac{\xi(\gamma)}{\xi(x)}\right)^{\gamma}$. Hence

$$\frac{\partial\tilde{\mathcal{M}}^{(\gamma)}(\xi_\lambda)}{\partial\lambda} = -\sum_{x,z} f(x)[C(\mathcal{X}) + \tilde{W}^{(\gamma)}(\xi_\lambda)]^{-1}_{x,\cdot}\frac{\partial\tilde{W}^{(\gamma)}(\xi_\lambda)}{\partial\lambda}[C(\mathcal{X}) + \tilde{W}^{(\gamma)}(\xi_\lambda)]^{-1}_{\cdot,z} f^T(z),$$

where $\xi_\lambda = (1-\lambda)\mu + \lambda\eta$. We shall suppose that $S_\mu = \mathcal{X}$. Then we obtain

$$\lim_{\lambda\to 0}\frac{\partial\tilde{\mathcal{M}}^{(\gamma)}(\xi_\lambda)}{\partial\lambda} = \gamma\sum_{x,z} f(x)\sum_u [C(\mathcal{X})+\tilde{W}^{(\gamma)}(\mu)]^{-1}_{x,u}\lim_{\lambda\to 0}\frac{\partial\ln\left(\frac{\xi_\lambda(u)}{\xi(\gamma)}\right)}{\partial\lambda}[C(\mathcal{X})+\tilde{W}^{(\gamma)}(\mu)]^{-1}_{u,z} f^T(z)$$

and using results from Müller and Pázman (1995), Section 3.1

$$\lim_{\gamma\to 0}\lim_{\lambda\to 0}\frac{1}{\gamma}\frac{\partial\tilde{\mathcal{M}}^{(\gamma)}(\xi_\lambda)}{\partial\lambda} = \sum_{x,z} f(x)\sum_u C(\mathcal{X})^{-1}_{x,u}\left(\frac{\eta(u)}{\mu(u)} - E_{B_\mu}\left[\frac{\eta(.)}{\mu_{\max}}\right]\right) C(\mathcal{X})^{-1}_{u,z} f^T(z).$$

In particular, if $\eta(x) = \delta(x, x^*)$, where x^* is any particular point from $\mathcal{X}$, then

$$\lim_{\gamma\to 0}\lim_{\lambda\to 0}\frac{1}{\gamma}\frac{\partial\tilde{\mathcal{M}}^{(\gamma)}(\xi_\lambda)}{\partial\lambda} = \frac{1}{\mu(x^*)}\sum_{x,z} f(x)C(\mathcal{X})^{-1}_{x,x^*}C(\mathcal{X})^{-1}_{x^*,z} f^T(z)$$

$$-\frac{\chi_{B_\mu}(x^*)}{N_{B_\mu}\mu(x^*)}\sum_{x,z,u} f(x)C^{-1}(\mathcal{X})_{x,u}C^{-1}(\mathcal{X})_{u,z} f^T(z).$$

If we introduce the notation

$$\sum_x f(x)C^{-1}(\mathcal{X})_{x,x^*} = \tilde{a}(x^*)$$

we obtain for $\eta(x) = \delta(x, x^*)$

$$\lim_{\gamma\to 0}\lim_{\lambda\to 0}\frac{1}{\gamma}\frac{\partial\tilde{\mathcal{M}}^{(\gamma)}(\xi_\lambda)}{\partial\lambda} = \frac{1}{\mu(x^*)}\left\{\tilde{a}(x^*)\tilde{a}^T(x^*) - \frac{\chi_{B_\mu}(x^*)}{N_{B_\mu}}\sum_u \tilde{a}(u)\tilde{a}^T(u)\right\}.$$

If $\Phi(M)$ is an optimality criterion, then

$$\lim_{\gamma\to 0}\lim_{\lambda\to 0}\frac{1}{\gamma}\frac{\partial\Phi[\tilde{\mathcal{M}}^{(\gamma)}(\xi_\lambda)]}{\partial\lambda} = \frac{1}{\mu(x^*)}\left\{\tilde{a}(x^*)^T\nabla\Phi[M(\mathcal{X})]\tilde{a}(x^*)\right.$$

$$\left.-\frac{\chi_{B_\mu}(x^*)}{N_{B_\mu}}\sum_u \tilde{a}^T(u)\nabla\Phi[M(\mathcal{X})]\tilde{a}(u)\right\},$$

where $\nabla\Phi(M)$ denotes the gradient (the matrix of first-order derivatives) of the function Φ at the point M.

5 The approximate information matrix if the number of observations is restricted

Although in Section 3 we have obtained a good formula for the construction of a gradient algorithm it has no practical sense. The reason is that $A = \mathcal{X}$ is always an optimal design, (i.e. observations should be in each point of $\mathcal{X}$) and the gradient algorithm would give this design. To obtain a meaningful result we have to put some restriction on the number of points which are allowed in the design.

A standard approach (cf. Müller and Pázman (1995)) to restrict the number of support points is the minimization of the criterion function modified by a penalty term

$$\min_{\xi}\{\Phi[\mathcal{M}^{(\gamma)}(\xi)] + \kappa\pi^{(\gamma)}(\xi)]\},$$

where κ is a multiplier and $\pi^{(\gamma)}(\xi)$ is the penalty function, which should be a monotone function of the number of points in S_ξ.

In this section we present an alternative approach how to achieve this goal. We modify once more the definition of the information matrix. Let N be the upper bound of the number of points which are allowed in the design, and take a fixed $\kappa = \frac{1}{N}$ (or slightly larger to avoid the influence of numerical errors). Let us cut the design measure ξ on the level κ, i.e. let us define

$$\xi^{\#}(x) = \max\{0, \xi(x) - \kappa\}$$

and

$$\xi^{\#}_{\max} = \max\{\xi^{\#}(x); x \in \mathcal{X}\} = \max\{0, \xi(x) - \kappa; x \in \mathcal{X}\}.$$

Let ${}_\gamma\xi(x)$ and ${}_\gamma\xi$ be smooth approximations of $\xi^{\#}(x)$, $\xi^{\#}_{\max}$ respectively:

$${}_\gamma\xi(x) = [\kappa^{\frac{1}{\gamma}} + \xi^{\frac{1}{\gamma}}(x)]^\gamma - \kappa, \tag{5.1}$$

$${}_\gamma\xi = [\kappa^{\frac{1}{\gamma}} + \sum_{u\in\mathcal{X}} \xi^{\frac{1}{\gamma}}(u)]^\gamma - \kappa. \tag{5.2}$$

We then have

$$\lim_{\gamma\to 0} {}_\gamma\xi(x) = \xi^{\#}(x),$$

and

$$\lim_{\gamma\to 0} {}_\gamma\xi = \xi^{\#}_{\max}.$$

Let $k > 1$ be another fixed (large) number. Define the approximate information matrix

$$J^{(\gamma)}(\xi) = \sum_{x,z\in S_\xi} f(x)[C(S_\xi) + kW^{(\gamma)}_\kappa(\xi)]^{-1}_{x,z} f^T(z), \tag{5.3}$$

where $W^{(\gamma)}_\kappa(\xi)$ is a diagonal matrix with entries

$$W^{(\gamma)}_\kappa(\xi) = \ln\left(\frac{{}_\gamma\xi}{{}_\gamma\xi(x)}\right)^\gamma. \tag{5.4}$$

This is a direct extension of $\tilde{\mathcal{M}}^{(\gamma)}(\xi)$, because when $\kappa \geq \frac{1}{N_{\mathcal{X}}}$, we have for every $\gamma > 0$

$$J^{(\gamma)}(\xi) = \sum_{x,z \in S_\xi} f(x)[C(S_\xi) + k\tilde{W}^{(\gamma)}(\xi)]^{-1}_{x,z} f^T(z), \tag{5.5}$$

which is equal to (4.1) up to the fixed factor $k > 1$, which has been found useful in the algorithm presented below (by varying k we can exclude local minima).

Lemma 2: If $\xi(x) > \kappa$, then

$$\lim_{\gamma \to 0} \ln \left(\frac{\gamma\xi}{\gamma\xi(x)} \right)^\gamma = 0.$$

If $\xi(x) < \kappa < \xi_{\max}$, then

$$\lim_{\gamma \to 0} \ln \left(\frac{\gamma\xi}{\gamma\xi(x)} \right)^\gamma = \ln \frac{\kappa}{\xi(x)} > 0.$$

If $\xi_{\max} < \kappa$, then

$$\lim_{\gamma \to 0} \ln \left(\frac{\gamma\xi}{\gamma\xi(x)} \right)^\gamma = \ln \frac{\xi_{\max}}{\xi(x)} > 0.$$

The proof is presented in the Appendix.

Proposition 3: We have

$$\lim_{\gamma \to 0} J^{(\gamma)}(\xi) = \sum_{x,z \in S_\xi} f(x)[C(S_\xi) + kV_\kappa(\xi)]^{-1} f^T(z),$$

where $V_\kappa(\xi)$ is a diagonal matrix with entries

$$[V_\kappa(\xi)]_{x,x} = \begin{cases} \ln \frac{\xi_{\max}}{\xi(x)}; & \text{if } \xi_{\max} < \kappa \\ \ln \frac{\kappa}{\xi(x)}; & \text{if } \xi(x) < \kappa < \xi_{\max} \\ 0; & \text{if } \xi(x) > \kappa. \end{cases} \tag{5.6}$$

The proof follows from Lemma 2. We note that $[V_\kappa(\xi)]_{x,x}$ can be continuously extended to the cases $\kappa = \xi(x)$ and $\kappa = \xi_{\max}$, which are not considered in (5.6).

Remark: We see that in the case that $\xi(x) \geq \kappa$ for every $x \in S_\xi$, then

$$\lim_{\gamma \to 0} J^{(\gamma)}(\xi) = \sum_{x,z \in S_\xi} f(x)C^{-1}(S_\xi)f^T(z),$$

which is the same as $\lim_{\gamma \to 0} \mathcal{M}^{(\gamma)}(\xi)$ (see Proposition1). So in this case κ has no influence on the limit value of the information matrix.

On the other hand, if $\xi(x) < \kappa$ for some $x \in S_\xi$, we see that an additional noise remains also in the limit $\gamma \to 0$ (according to Proposition 3 its variance is $\frac{\min(\kappa, \xi_{\max})}{\xi(x)}$). So points with $\xi(x) > \kappa$ are prefered, but their number is restricted by $N = \frac{1}{\kappa}$.

The derivative : Suppose that $\xi_\lambda = (1-\lambda)\mu + \lambda\eta$, and that $S_\mu = \mathcal{X}$. Directly from (5.3) we obtain by a differentiation similar to that in Section 3, that the derivative of the criterion function based on the information matrix $J^{(\gamma)}(\xi)$ is

$$\frac{\partial}{\partial\lambda}\Phi[J^{(\gamma)}(\xi_\lambda)] = -k\sum_{u\in\mathcal{X}} d_\gamma(\xi_\lambda, u)\frac{\partial}{\partial\lambda}\ln[\frac{\gamma\xi_\lambda}{\gamma\xi_\lambda(u)}]^\gamma, \tag{5.7}$$

where

$$d_\gamma(\xi, u) = a_\gamma^T(u)\nabla\Phi[J^{(\gamma)}(\xi)]a_\gamma(u)$$

and

$$a_\gamma(u) = \sum_{x\in\mathcal{X}}[C(\mathcal{X}) + kW_\kappa^{(\gamma)}(\xi)]^{-1}_{u,x}f(x).$$

By taking limits we obtain

$$\lim_{\gamma\to0}\lim_{\lambda\to0}\frac{\partial}{\partial\lambda}\Phi[J^{(\gamma)}(\xi_\lambda)] = -k\sum_{x\in\mathcal{X}}\lim_{\gamma\to0} d_\gamma(\mu, x)\lim_{\gamma\to0}\lim_{\lambda\to0}[\gamma\ln{}_\gamma\xi_\lambda - \gamma\ln{}_\gamma\xi_\lambda(x)]. \tag{5.8}$$

In the following lemma we present the derivatives which appear in (5.8).

Lemma 3:

$$\lim_{\gamma\to0}\lim_{\lambda\to0}\frac{\partial}{\partial\lambda}\ln[\gamma\xi_\lambda(x)] = \frac{\eta(x)-\mu(x)}{\mu(x)-\kappa} \qquad \text{if } \mu(x) > \kappa,$$

$$\lim_{\gamma\to0}\lim_{\lambda\to0}\frac{\partial}{\partial\lambda}\gamma\ln[\gamma\xi_\lambda(x)] = \frac{\eta(x)}{\mu(x)} - 1 \qquad \text{if } \mu(x) < \kappa,$$

$$\lim_{\gamma\to0}\lim_{\lambda\to0}\frac{\partial}{\partial\lambda}\ln[\gamma\xi_\lambda] = \frac{\mu_{\max}}{\mu_{\max}-\kappa}\{E_{B_\mu}\left[\frac{\eta(.)}{\mu(.)}\right] - 1\} \qquad \text{if } \mu_{\max} > \kappa,$$

$$\lim_{\gamma\to0}\lim_{\lambda\to0}\frac{\partial}{\partial\lambda}\gamma\ln[\gamma\xi_\lambda] = E_{B_\mu}\left[\frac{\eta(.)}{\mu(.)}\right] - 1 \qquad \text{if } \mu_{\max} < \kappa,$$

where

$$E_{B_\mu}\left[\frac{\eta(.)}{\mu(.)}\right] = \frac{1}{N_{B_\mu}}\sum_{x\in B_\mu}\frac{\eta(x)}{\mu(x)}.$$

The proof is presented in the Appendix.

If $\eta(.) = \delta(., x^*)$ is the measure concentrated at x^*, we have

$$E_{B_\mu}\left[\frac{\eta(.)}{\mu(.)}\right] = \frac{\mathcal{I}_{B_\mu}(x^*)}{\mu_{\max}N_{B_\mu}}.$$

Notice that in the case that $\mu(x) < \kappa$ or $\mu_{\max} < \kappa$, the logarithm in Lemma 3 is multiplied by the number γ, which in the limit tends to zero. So in this case the limit derivative in (5.8) is infinitely larger than for the case $\mu(x) > \kappa$, $\mu_{\max} > \kappa$ and this will influence the algorithm below.

6 The gradient algorithm for the computation of optimum designs with a restricted number of observations

As mentioned in the introduction, we verify the effectiveness of the new approach by building a gradient algorithm. Other algorithms known from classical design theory (cf. e.g. Pázman (1986)) are possible, but to present them goes beyond the aim of the paper.

We shall start the algorithm with a design supported on the whole set $\mathcal{X}$. In case that $f(x) = 0$ for some $x \in \mathcal{X}$, we take the set $\{x \in \mathcal{X} : f(x) \neq 0\}$ instead of $\mathcal{X}$. Usually we take as a starting design either $\xi(x) = \frac{1}{N_{\mathcal{X}}}$, or a design which is close to the supposed optimum. The limitation of this procedure is of course that we can only use covariance functions that yield a positive definite covariance matrix on the support of $\xi^{(0)}$.

At the beginning we choose $\kappa = \frac{1}{N}$, or κ slightly larger than $\frac{1}{N}$, where N is the required number of points in the design. We choose also some $k > 1$, and perform the algorithm for this fixed k. Let ξ^* be the resulting design. If ξ^* is not uniform, it is a false optimum and we have to choose another starting design. If ξ^* is uniform, we restart the algorithm for a different k and some starting design (possibly ξ^*). When a design $\xi^{**} \neq \xi^*$ is obtained, it means that ξ^* was only a local minimum of the criterion function. Since this function depends upon the choice of k its local minima should be shifted by the change of k, which explains why $\xi^* \neq \xi^{**}$. On the other hand the result $\xi^* = \xi^{**}$ indicates that ξ^* is the absolute minimum, i.e. the design ξ^* is optimal as required, since by changing k we change the matrix $J^{(\gamma)}(\xi)$ given in (7), hence the function $\Phi[J^{(\gamma)}(\xi)]$, but we do not change the $\arg\min_\xi \Phi[J^{(\gamma)}(\xi)]$.

The algorithm consists of a stepwise one-point correction of the design. That means at the n-th step we take

$$\mu_{n+1} = \frac{n}{n+1}\mu_n + \frac{1}{n+1}\delta(.,x^*),$$

where μ_n is the design in the n-th step and $\delta(.,x^*)$ is the Dirac measure concentrated at x^*. The point x^* is chosen such as to minimize the directional derivative in the direction of $\delta(.,x^*)$, i.e.

$$\lim_{\lambda \to 0} \frac{\Phi[J^{(\gamma)}[(1-\lambda)\mu_n + \lambda\delta(.,x^*)]}{\lambda}.$$

This has to be done for small γ, in the limit for $\gamma \to 0$, which requires the use of the results given in (9)-(5.8) and in Lemma 3 for the case $\eta(.) = \delta(.,x^*)$. As a result we have the following algorithm.

6.1 Formal description of the algorithm (for a fixed κ and k)

For reasons of notational simplicity we write $\mu = \mu_n$ here.

- Define

$$D(\mu) = [C(\mathcal{X}) + kV_\kappa(\mu)],$$

where $V_\kappa(\mu) = \text{diag}\{\ln \frac{\min\{\mu_{\max},\kappa\}}{\min\{\mu(x),\kappa\}}; x \in \mathcal{X}\}$.

- Define

$$a(x) = \sum_{z\in\mathcal{X}} [D^{-1}(\mu)]_{x,z} f(z),$$

$$M(\mu) = \sum_{x,z\in\mathcal{X}} f(x)[D^{-1}(\mu)]_{x,z} f^T(z),$$

$$d(\mu, x) = a^T(x)\nabla\Phi[M(\mu)]a(x).$$

- Compute

$$q(\mu) = \min_{\{x:\mu(x)\le\kappa\}} \frac{d(\mu,x)}{\mu(x)},$$

$$x_q = \arg\min_{\{x:\mu(x)\le\kappa\}} \frac{d(\mu,x)}{\mu(x)}.$$

- Compute (if $\mu_{\max} > \kappa$)

$$r(\mu) = \min_{\{x:\mu(x)>\kappa\}} \frac{1}{\mu(x)-\kappa}\{d(\mu,x) - [\sum_{z\in\mathcal{X}} d(\mu,z)]\frac{\mathcal{I}_{B_\mu}(x)}{N_{B_\mu}}\},$$

$$x_r = \arg\min_{\{x:\mu(x)>\kappa\}} \frac{1}{\mu(x)-\kappa}\{d(\mu,x) - [\sum_{z\in\mathcal{X}} d(\mu,z)]\frac{\mathcal{I}_{B_\mu}(x)}{N_{B_\mu}}\}.$$

- At each step:
  - if $q(\mu) < 0$ take x_q,
  - if $q(\mu) = 0 \pm \varepsilon$ and $r(\mu) < 0$ take x_r,
  - if $q(\mu) = 0 \pm \varepsilon$ and $r(\mu) \ge 0$ take x_q,
  - if $q(\mu) > 0$ take x_r.

Here $\epsilon > 0$ is a fixed small number smoothing the influence of numerical errors on checking the theoretical equation $q(u) = 0$.

6.2 Examples

6.2.1 Brownian motion

One of the classic examples from the area of optimal design for random fields stems from Sacks and Ylvisaker (1966). They consider quadratic regression of the form

$$f(x) = x^2, \qquad x \in [0,1].$$

The noise ϵ is assumed to be a Brownian motion, with the consequent covariance-function

$$C(x,z) = \min(x,z), \qquad x,z \in [0,1].$$

By their asymptotic method Sacks and Ylvisaker (1966) find the optimal n-point designs:

$$x_i^* = \frac{i}{n}, \qquad i = 1,\ldots,n. \tag{6.1}$$

We chose $\mathcal{X}$ equal to an equidistant 24 point grid. In this case the points x_i^* given by (6.1) form the optimal exact design if $n = 1,2,3,4,6,8$ or 12. In each case we chose $\kappa = \frac{1}{n+\varepsilon}$ (for $\varepsilon = 10^{-4}$), and $k = 1$. An initial design that was uniform over $\mathcal{X}$ was chosen. For $n = 1,2,4,8,12$ the resulting computed design was uniform over its support. Then k was increased to $k = 10$ and we repeated the algorithm to obtain the same result which was in correspondence with (6.1). For $n = 3,6$ the resulting design was not uniform, hence false. Then we modified the starting design (at random) until we obtained a uniform design and proceeded as for the other values of n.

The algorithm was stopped after 2000 iterations, which proved to be sufficient for all considered examples. Figure 1 depicts the final result for $n = 6$. The height of the bars are proportional to ξ and the filled circles are the points of support of $\xi(x)$. The short bars correspond to the remainder of the initial design. The dashed line indicates the level of κ.

6.2.2 Computation of uniformly optimal designs

In some experiments the regression function $f(x)$ is related to the covariance function $C(x,z)$ by a linear relation:

$$f(x) = \sum_{z \in D} C(x,z) r(z), \quad \forall x \in \mathcal{X} \tag{6.2}$$

where D is some finite set and $r(z) \in I\!R^m$ are given vectors. These examples are considered extensively in Näther (1985b).

For example, in the simple linear regression

$$y(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}^T \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \epsilon(x) \quad x \in [-1,1]$$

with the covariance function

$$C(x,z) = \begin{cases} 1 - |x-z|; & |x-z| < 1 \\ 0; & |x-z| \geq 1 \end{cases}$$

the condition (6.2) holds as can be easily seen from:

$$\begin{pmatrix} 1 \\ x \end{pmatrix} = C(x,-1)\begin{pmatrix} 1 \\ -1 \end{pmatrix} + C(x,0)\begin{pmatrix} 1 \\ 0 \end{pmatrix} + C(x,1)\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

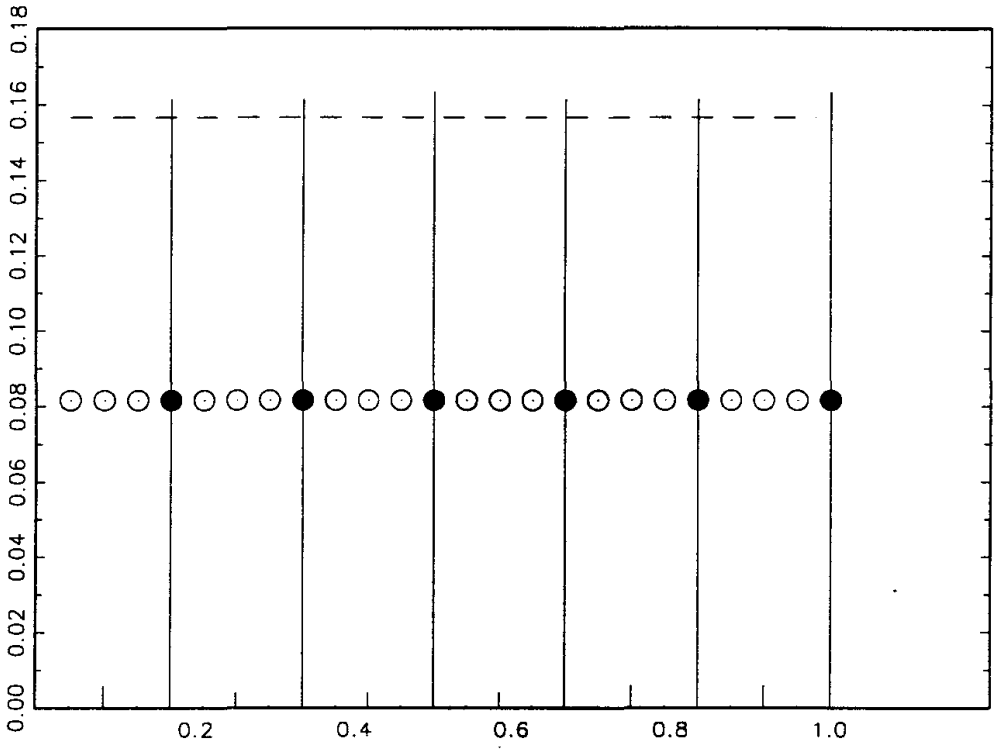


Figure 1: Optimal design for the Brownian Motion example 6.2.1.

One can consider D as an exact design, in accordance with the definition in Section 1. Under the assumption that the vectors $r(z)$, $z \in D$, are independent, the design D is uniformly optimum among all exact designs (see Sacks and Ylvisaker (1966) or see Näther (1985b) for a detailed proof).

In the above example $D = \{-1, 0, 1\}$ is the optimal design giving the full information about θ, which is contained in the random process $\{y(x) : x \in [-1, 1]\}$.

We applied the algorithm given in the previous section to this example and the above result was confirmed numerically by our gradient method, as can be seen from Figure 2.

6.2.3 A discontinuous covariance function

Suppose now simple linear regression, i.e. that $f(x) = \binom{1}{x}$ and

$$C(x,z) = \begin{cases} \sigma^2 & x = z \\ c\sigma^2 & 0 < |x - z| \leq d \\ 0 & \text{else} \end{cases},$$

which is a discontinuous function depending upon the parameters $c \in [0,1]$ and $d \in [0,2]$ $(x, z \in [-1,1])$.

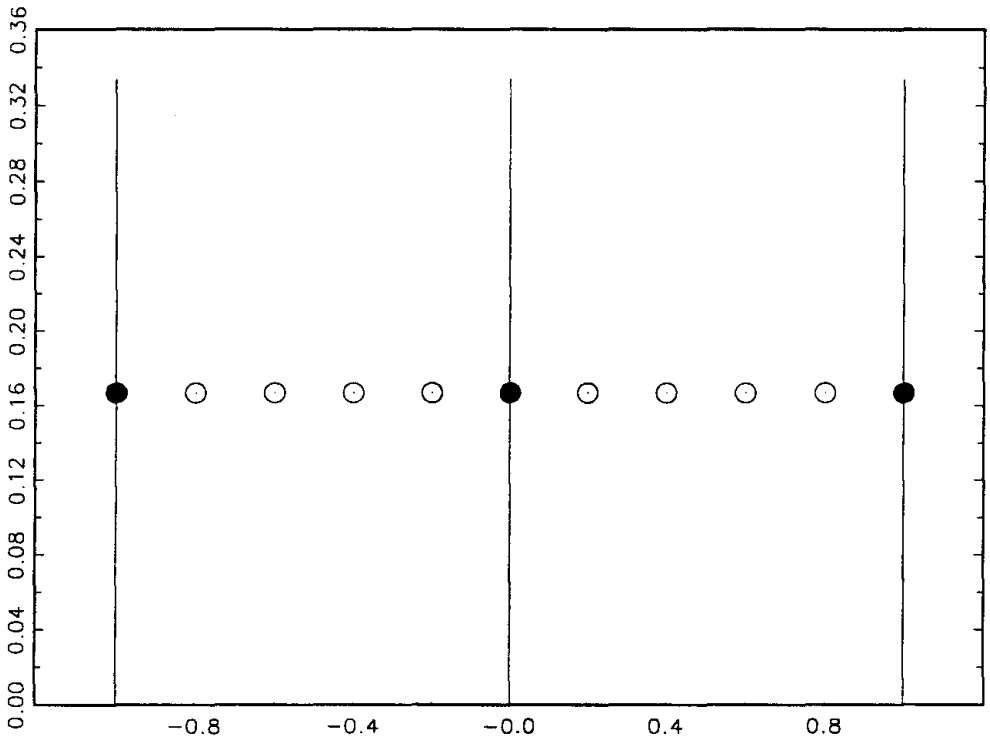


Figure 2: Numerical optimal design for the example in 6.2.2 on an 11 point grid.

The case of three design points was extensively studied by Müller (1995). Here three types of optimal designs were specified depending upon the choice of c and d. For $c \in < 0.33, 0.5 >$, for instance, and $d \in < 0, 1 >$ the optimal three point design is characterized by the two endpoints $x_{1,2} = \pm 1$ of the region plus the additional point at $|x_3| < 1 - d$, such that $|x_3|$ is maximum.

The modified algorithm from section 6.1 now allows to compute these cases applying a similar procedure as in Example 6.1. Figure 3 depicts the result for the settings $c = 0.4$ and $d = 1.8$, which is in accordance to the findings of Müller (1995).

7 Discussion

We believe that the approach presented in this paper brings a new point of view towards the optimal design problem for correlated observations, since we are not aware of any other papers that use the concept of design measures in this context. By introducing extended information matrices and an additional white noise dependent upon the design, we were able to make the design problem smooth in such a way that standard techniques of differentiation could be used. This allowed us to elaborate a gradient algorithm that is of potential use in applications, especially in complex settings, where

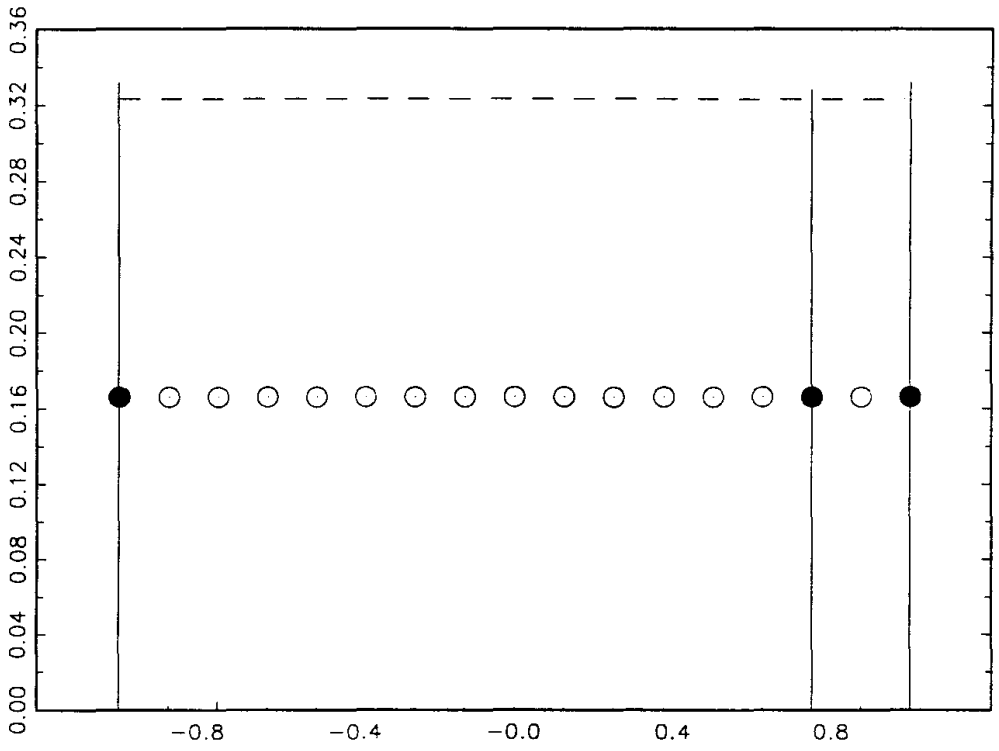


Figure 3: Optimal design for the discontinuous C(.) example 6.2.3.

enumerative methods are unfeasible. Note that here alternative algorithms (e.g. of the exchange type) are not guaranteed to find the optimum.

Applications for the presented methodology can be found in various fields, especially in the design of computer simulation experiments (see Sacks *et al.* (1989) for various setups). There, the covariance matrix is usually assumed to be known up to scalar parameter, which is certainly one of many possible directions of extending our technique.

Appendix

Proof of Lemma 2

We shall consider the terms in

$$\ln\left(\frac{\gamma\xi}{\gamma\xi(x)}\right)^{\gamma} = \gamma\ln[\gamma\xi] - \gamma\ln[\gamma\xi(x)].$$

i) If $\xi(x) > \kappa$, then (5.1) implies $\lim_{\gamma\to 0}[\gamma\xi(x)] = \xi(x) - \kappa$, hence $\lim_{\gamma\to 0}\gamma[\gamma\xi(x)] = 0$.

ii) Similarly, $\xi_{\max} > \kappa \Longrightarrow \lim_{\gamma\to 0}\gamma_\gamma\xi = 0$.

iii) If $\xi(x) < \kappa$ then from (5.1) we obtain

$$\gamma\ln[\gamma\xi(x)] = \gamma\ln\kappa + \gamma\ln\{[1 + t^{\frac{1}{\gamma}}(x)]^{\gamma} - 1\}, \tag{7.1}$$

with $t(x) = \frac{\xi(x)}{\kappa} < 1$. By the Taylor formula we have for $\gamma \to 0$

$$[1 + t^{\frac{1}{\gamma}}(x)]^{\gamma} = 1 + \gamma t^{\frac{1}{\gamma}}(x) + \frac{1}{2}\gamma(\gamma-1)t^{\frac{2}{\gamma}}(x) + o(t^{\frac{2}{\gamma}}(x)).$$

Hence from (7.1) we obtain

$$\gamma\ln[\gamma\xi(x)] = \gamma\ln\kappa + \gamma\ln\{\gamma t^{\frac{1}{\gamma}}(x) + \frac{1}{2}\gamma(\gamma-1)t^{\frac{2}{\gamma}}(x) + o(t^{\frac{2}{\gamma}}(x))\}$$

$$= \gamma\ln\kappa + \gamma\ln\{\gamma t^{\frac{1}{\gamma}}(x)\} + \gamma\ln\{1 + \frac{1}{2}(\gamma-1)t^{\frac{1}{\gamma}}(x) + o(t^{\frac{1}{\gamma}}(x))\}$$

$$\longrightarrow_{\gamma\to 0} \ln t(x) = \ln\frac{\xi(x)}{\kappa}.$$

iv) Similarly, if $\xi_{\max} < \kappa$ then from (5.2) we obtain

$$\lim_{\gamma\to 0}\gamma\ln[\gamma\xi] = \lim_{\gamma\to 0}\gamma\ln\{[1 + \sum_{x\in\mathcal{X}} t^{\frac{1}{\gamma}}(x)]^{\gamma} - 1\}$$

$$= \lim_{\gamma\to 0}\ln[\sum_{x\in\mathcal{X}} t^{\frac{1}{\gamma}}(x)]^{\gamma} = \ln\left[\frac{\xi_{\max}}{\kappa}\right].$$

□

Proof of Lemma 3

According to (6) we have

$$\frac{\partial}{\partial\lambda}\ln[\gamma\xi_\lambda] = \frac{\partial}{\partial\lambda}\ln\{[h_\gamma(\xi_\lambda)]^\gamma - \kappa\},$$

where $h_\gamma(\xi) = \kappa^{\frac{1}{\gamma}} + \sum_{x\in\mathcal{X}}\xi(x)^{\frac{1}{\gamma}}$.

By direct differentiation we obtain

$$\frac{\partial}{\partial\lambda}\ln[\gamma\xi_\lambda] = \frac{1}{[h_\gamma(\xi_\lambda)]^\gamma - \kappa}[h_\gamma(\xi_\lambda)]^{\gamma-1}\sum_{x\in\mathcal{X}}\xi_\lambda^{\frac{1}{\gamma}-1}(x)\frac{d\xi_\lambda(x)}{d\lambda}$$

$$\longrightarrow_{\lambda\to 0}\frac{[h_\gamma(\mu)]^\gamma}{[h_\gamma(\mu)]^\gamma - \kappa}\times\frac{\sum_{x\in\mathcal{X}}\mu^{\frac{1}{\gamma}}(x)}{h_\gamma(\mu)}\mathrm{E}_\mu^{(\frac{1}{\gamma})}\left[\frac{\eta(.)}{\mu(.)} - 1\right], \tag{7.2}$$

where $E_\mu^{(\frac{1}{\gamma})}$ denotes the weighted mean with weights equal to $\frac{\mu^{\frac{1}{\gamma}}(x)}{\sum_{u\in\mathcal{X}}\mu^{\frac{1}{\gamma}}(u)}$.

Similarly

$$\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}\ln[\gamma\xi_\lambda(x)] = \frac{[l_\gamma(x)]^\gamma}{[l_\gamma(x)]^\gamma - \kappa}\times\frac{\mu^{\frac{1}{\gamma}}(x)}{l_\gamma(x)}\left[\frac{\eta(x)}{\mu(x)} - 1\right], \tag{7.3}$$

where $l_\gamma(x) = \kappa^{\frac{1}{\gamma}} + \mu^{\frac{1}{\gamma}}(x)$.

i) Let $\mu_{\max} > \kappa$. Then we obtain directly from (7.2)

$$\lim_{\gamma\to 0}\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}\ln[\gamma\xi_\lambda] = \frac{\mu_{\max}}{\mu_{\max} - \kappa}E_{B_\mu}\left[\frac{\eta(.)}{\mu(.)} - 1\right].$$

ii) Similarly, if $\mu(x) > \kappa$ we obtain from (16)

$$\lim_{\gamma\to 0}\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}\ln[\gamma\xi_\lambda(x)] = \frac{\eta(x) - \mu(x)}{\mu(x) - \kappa}$$

iii) Suppose that $\mu_{\max} < \kappa$, and denote $s(x) = \frac{\mu(x)}{\kappa} < 1$. From (7.2) we obtain

$$\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}\gamma\ln[\gamma\xi_\lambda] = \frac{[1 + \sum_{x\in\mathcal{X}}s^{\frac{1}{\gamma}}(x)]^\gamma}{[1 + \sum_{x\in\mathcal{X}}s^{\frac{1}{\gamma}}(x)]^\gamma - 1}\times\frac{\gamma\sum_{x\in\mathcal{X}}s^{\frac{1}{\gamma}}(x)}{[1 + \sum_{x\in\mathcal{X}}s^{\frac{1}{\gamma}}(x)]}\mathrm{E}_\mu^{(\frac{1}{\gamma})}\left[\frac{\eta(.)}{\mu(.)} - 1\right]$$

$$\longrightarrow_{\gamma\to 0}E_{B_\mu}\left[\frac{\eta(.)}{\mu(.)} - 1\right]\lim_{\gamma\to 0}\frac{\gamma\sum_{x\in\mathcal{X}}s^{\frac{1}{\gamma}}(x)}{[1 + \sum_{x\in\mathcal{X}}s^{\frac{1}{\gamma}}(x)]^\gamma - 1}.$$

Using the Taylor formula, like in the proof of Lemma 2, we obtain that the last limit is equal to 1.

iv) Suppose now, that $\mu(x) < \kappa$. Then from (7.3) we obtain

$$\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}\gamma\ln[\gamma\xi_\lambda(x)] = \gamma\frac{[1+s^{\frac{1}{\gamma}}(x)]^\gamma}{[1+s^{\frac{1}{\gamma}}(x)]^\gamma-1}\times\frac{s^{\frac{1}{\gamma}}(x)}{1+s^{\frac{1}{\gamma}}(x)}\left[\frac{\eta(x)}{\mu(x)}-1\right]$$

$$\longrightarrow_{\gamma\to 0}\left[\frac{\eta(x)}{\mu(x)}-1\right]\lim_{\gamma\to 0}\frac{\gamma s^{\frac{1}{\gamma}}(x)}{[1+s^{\frac{1}{\gamma}}(x)]^\gamma-1},$$

and the last limit is equal to 1, which can be proved in the same way as in iii). □

References

Abt, M. (1992). Some exact optimal designs for linear covariance functions in one dimension. *Comm. Statist. Theory Methods*, 21:2059–2069.

Bischoff, W. (1993). On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *Journal of Statistical Planning and Inference*, 37:69–80.

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

Fedorov, V.V. (to be published 1996). Design of spatial experiments: model fitting and prediction. In Rao, C.R. and Gosh, S., editors, *Handbook of Statistics*, volume 13, chapter ?, pages ???–??? North-Holland.

Kubáček, L., Kubáčkova, L. and Volaufová, J. (1995). *Statistical Models with Linear Structures.* Veda Publishing House.

Müller, W.G. and Pázman, A. (1995). Design measures and extended information matrices. Technical Report 47, University of Economics, Department of Statistics, Vienna.

Müller, W.G. (1995). An example of optimal design for correlated observations (in German). *Österreichische Zeitschrift für Statistik*, 24(1):9–15.

Müller-Gronbach, T. (1996). Optimal designs for approximating the path of a stochastic process. *Journal of Statistical Planning and Inference*, 49:371–385.

Näther, W. (1985). *Effective Observation of Random Fields.* Teubner Texte zur Mathematik - Band 72. Teubner Verlag, Leipzig.

Näther, W. (1985). Exact designs for regression models with correlated errors. *Statistics*, 16(4):479–484.

Pázman, A. (1986). *Foundations of Optimum Experimental Design.* Mathematics and Its Applications. D.Reidel, Dordrecht.

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

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435.

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

Sacks, J. and Ylvisaker, D. (1968). Design for regression problems with correlated errors; many parameters. *Annals of Mathematical Statistics*, 39:46–69.

Sacks, J. and Ylvisaker, D. (1970). Design for regression problems with correlated errors III. *Annals of Mathematical Statistics*, 41:2057–2074.

Wahba, G. (1971). On the regression design problem of Sacks and Ylvisaker. *Annals of Mathematical Statistics*, 42:1035–1053.