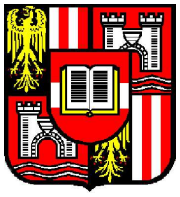


Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2007-26

Issues in the Optimal Design of Computer Simulation Experiments

Werner Müller and Milan Stehlík

July 2007

Abstract

Output from computer simulation experiments are often approximated as realisations of correlated random fields. Consequently, the corresponding optimal design questions (cf. Sacks et al., 1989 or Steinberg and Bursztyn, 2006) must cope with the existence and detection of an error correlation structure, issues largely unaccounted for by traditional optimal design theory. Unfortunately, many of the nice features of well established design techniques, like additivity of information matrix, convexity of design criteria, etc., do not carry over to the setting of interest. This may lead to unexpected, counter-intuitive even paradoxical effects in the design (e.g. Müller and Stehlík, 2004) as well as the analysis (e.g. Smit, 1961) stage of computer simulation experiments. In the paper we intend to give an overview and some simple but illuminating examples for this behaviour.

KEY WORDS: Optimum experimental design, computer experiment, D-optimality, efficiency, equidistant design, parameterized covariance functions, Smit's paradox.

Introduction

Many physical processes are difficult or even impossible to study by classical experimental methodology. As computer power has increased, it is possible to model some of these problems by sophisticated computer code. As in a real experiment one can vary the inputs to the code and observe how the process output is affected. Such studies are called computer (simulation) experiments (for recent comprehensive reviews of related statistical methodology see Santner et al. (2003) and Fang et al. (2005)).

Since the seminal paper by Sacks et al. (1989) it is customary to employ random fields and corresponding estimation and prediction techniques (kriging) for modeling and analysis of computer simulation data. In this paper we will for the sake of conciseness restrict the attention to isotropic stationary linear processes or random fields of the form

$$Y(x) = \eta(x, \beta) + \varepsilon(x) = \beta_0 + \beta_1 x + \varepsilon(x), \tag{1}$$

with design points (inputs for the code) $\xi_n = \{x_1, ..., x_n\}$ taken from a compact design space $\mathcal{X}$. The trend parameters β are unknown and the variance-covariance structure of the errors depends on another set of unknown parameters θ.

We assume that the errors $\varepsilon(x)$ are correlated and the correlation between two measurements depends on the distance d between pairs of particular design points through either the so-called triangular covariance (which was introduced by Müller and Stehlík (2004) as the modification of a structure studied in Example 6.4 by Näther (1985))

$$\operatorname{cov}(Y(x), Y(x+d)) = \begin{cases} \sigma^2(1 - \frac{d}{\theta_1}), & \text{for } d < \theta_1, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

or the the so called Matérn model family, which is more and more frequently employed for its flexibility due to a smoothness parameter θ_2. It is given by

$$\operatorname{cov}(Y(x), Y(x+d)) = \sigma^2 \cdot \left\{ \frac{1}{2^{\theta_2 - 1}\Gamma(\theta_2)} \left(\frac{d}{\theta_1}\right)^{\theta_2} K_{\theta_2}\left(\frac{d}{\theta_1}\right) \right\}, \tag{3}$$

where K_{θ_2} denotes a modified Bessel function of order θ_2. A recent review of its history and properties can be found in Guttorp and Gneiting (2006). It encompasses a number of widely used models, e.g. by setting $\theta_2 = 1/2$ the so called exponential

$$\mathrm{cov}(Y(x), Y(x+d)) = \sigma^2 e^{-\frac{d}{\theta_1}}. \tag{4}$$

If we further assume that the errors are Gaussian, we obtain a parametrized Slepian process in the case of (2) when $\theta_1 = 1$ and a special case of an Ornstein-Uhlenbeck process when the covariance is exponential. The Slepian process S on $[0, 1]$ was first studied in Slepian (1961) and Watson (1961) and appears in numerous theoretical and applied probabilistic models, e.g. models for structural damage and strong ground motion. The relation of the triangular covariance to the so called geometric covariogram is discussed in Stehlík (2005).

The main purpose of statistical analysis is estimation of parameters β, θ and prediction of Y based upon the parametric models. Having estimated the spatial trend by a linear response a universal kriging estimator is typically used (see e.g. Müller (2007)). Universal kriging can be viewed as a two stage procedure involving GLS estimation of the trend surface and best linear prediction, as pointed out by Fedorov (1989). For a reverse two-stage interpretation see Cressie (1993).

The related optimum design question, which will be the core problem of the present paper is then how to select the inputs such, that we gain the maximum available information from the experiment. In this context, however for the simplistic i.i.d. error case, this question was first raised in the 1960's, see Naylor et al. (1967) for an early review. That this is an issue of high relevance may be gathered from the following simple example.

Example 0 *Assume that in (1) with $\mathcal{X} = [-1, 1]$ we have $\beta_0 = 0$ (no intercept regression) and the errors being i.i.d. (which is the special case unifying (2) with $\theta_1 = 0$ and (4) with $\theta_1 = 0$ including a nugget effect, see section 8). Then let us study the behaviour of two natural estimators for β_1, namely the average of the slopes, i.e. $\tilde{\beta}_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{x_i}$ or the best linear unbiased estimate $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ (see example 10.4 in Casella and Berger (2002) for further inspiration). It is clear that $Var(\tilde{\beta}_1) \geq Var(\hat{\beta}_1)$ always holds, but let us investigate their relationships for different designs. It is clear that both variances are minimized, when all the n observations are taken at or close to either of the endpoints of the design region $x = \pm 1$ (the optimum design). Contrast this with a regular n-point design (with even n) $\xi_n = \{-1, -1 + \frac{1}{2n-1}, \ldots, 1\}$. Under this design and for growing n the variance $Var(\tilde{\beta}_1) > \frac{\sigma^2 \pi}{4}$ is bounded below (see also Appendix), whereas under the optimum design it tends to zero (as does $Var(\hat{\beta}_1)$ for both designs). Moreover, the efficiency of the uniform design for $\hat{\beta}_1$ tends to two third of that under the optimum design. One can thus clearly see the beneficial effect of proper choice of inputs for both designs.*

In the following we will define optimality of a design always strictly in the tradition of Kiefer (see e.g. Kiefer, 1975), where the inputs are selected such, that a prespecified design criterion (e.g. the variance, see above; or for multiple parameters the determinant of their variance-covariance matrix, so-called D-optimality) is optimized. The classic Fisher information $M_\rho = E(\frac{\partial \ln f(x,\rho)}{\partial \rho})^2$ usually serves as the basis for many of the optimality criteria. Here differentiability of likelihood ratio with respect to the parameter ρ is assumed, which opens a problem of interpretation of the formally defined (e.g. by one-sided

limits) information matrix for such values, see Stehlík (2004a). Still, the classic Fisher information can be well defined over some open set. In the proposed model we have the Fisher information matrix

$$M_\beta(n) = F^T C^{-1} F,$$

where $F = (f(x_1), \ldots, f(x_n))^T,\ f(x_i) = (1, x_i)$.

It is traditional practice that the covariance parameters are estimated in a separate stage. However, if one is willing to make distributional assumptions, it is natural to employ likelihood based estimation techniques. In the following we will thus assume that the errors in our random field model (1) follow a stationary Gaussian process.

In particular one could now either assume that the trend is known and fixed (e.g. $\eta \equiv 0$) and maximize (full ML)

$$2L(\theta) = -n \log(2\pi) - \log\det C(\theta) - y^T C^{-1}(\theta) y \tag{5}$$

or that one employs say OLS-detrending $\hat{\varepsilon} = Y - \eta(\hat{\beta})$ and maximizes (REML)

$$\begin{aligned} 2L(\theta) =& -n \log(2\pi) - \log\det C(\theta) - \log\det F^T C^{-1} F \\ & -\hat{\varepsilon}^T \{C^{-1}(\theta) - C^{-1}(\theta) F [F^T C^{-1}(\theta) F]^{-1} F^T C^{-1}(\theta)\} \hat{\varepsilon}. \end{aligned}$$

Let us in the following concentrate on (5), since both approaches differ only little with respect to design issues (cf. Zimmerman (2006)). It is now natural to base a design criterion on the information matrix associated with the corresponding estimate of the parameter θ, which is given by (note that it depends upon the design ξ via C)

$$M_\theta(n)_{jj'} = \frac{1}{2} \mathrm{tr} \left\{ C^{-1}(\theta) \frac{\partial C(\theta)}{\partial \theta_j} C^{-1}(\theta) \frac{\partial C(\theta)}{\partial \theta_{j'}} \right\}, \tag{6}$$

where the $\frac{\partial C(\theta)}{\partial \theta_j}$ are $n \times n$ matrices with entries $\frac{\partial \mathrm{cov}(Y(x), Y(x+d))}{\partial \theta_j}$, $x, x+d \in \mathcal{X}$. Designs maximizing the determinant of M_θ have been suggested by Zhu and Stein (2005) (they also employ a minimax and Bayesian criterion to avoid undesirable effects due to the linearizations) and Zimmerman (2006), who calls them CP-(covariance parameter-)optimal. Both demonstrate their behaviour for numerous artificial and real examples.

Since both paremeters are independent one can use (see Pázman (2004) and Xia, Miranda and Gelfand (2006)) $M_{\beta,\theta} = \begin{pmatrix} M_\beta(n) & 0 \\ 0 & M_\theta(n) \end{pmatrix}$. Thus the full D-optimality criterion has the form $\Phi(M_{\theta,\beta}) = \det M_\theta \det M_\beta$. Such a function Φ is also called design- or information criterion, see Pukelsheim (1993). Note that there exists a well developed theory for standard i.i.d. regression based on Kiefer's (1959) concept of design measures, cf. Atkinson et al.(2007) for a recent textbook.

The other widely used criterion is defined via minimization of the mean squared error (MSE) (see e.g. Crary (2002)). The kriging variance can be interpreted as the unconditional mean squared prediction error for the best linear unbiased predictor. Thus frequently direct minimization of it respectively its weighted average over the area, i.e.

$$\min_\xi \int_{x \in \mathcal{X}} E[(\hat{y}(x|\xi) - y(x))^2] w(x) dx \tag{7}$$

is attempted. Resulting designs are sometimes also termed minimum AMSE (average mean squared error) designs and are - despite the computational complexities in their

generation - widely used in computer simulation experiments. Note, that in the correlated setup we do not have such a nice correspondence between estimation and prediction oriented designs (as given by the celebrated equivalence theorem by Kiefer and Wolfowitz, 1961) and therefore require separate optimization techniques for these purposes. Especially for (7) the respective computational burden can be high, but may be reduced by sequential updating of the designs as is shown in Pardo-Igúzquiza and Dowd (2005).

In the remainder of the paper we will now address peculiarities of the computer simulation setup, which are not covered by regular design theory and devise possible remedies for them. Thereby we will mainly concentrate on the goal of parameter estimation (for other relevant questions related to prediction see the Appendix). Some of the following issues are seemingly trivial, but their review in accordance with others may shed new light on some phenomena. We will especially learn that we are not to trust our intuition that we may attempt to carry over from the uncorrelated error setting.

1 Asymptotic Unidentifiability of Covariance Parameters

One can find applications of various criteria of design optimality for covariance parameter estimation in the literature. Here we consider D-optimality, which corresponds to the maximization of a criterion function $\Phi(M_\theta) = \det M_\theta$, the determinant of a standard Fisher information matrix. This method, "plucked" from the widely developed uncorrelated setup, is offering considerable potential for automatic implementation, although further development is needed before it can be applied routinely in practice.

Note specifically, that the validity of this approach much depends on the ability of the inverse information matrix to serve as a good approximation of the mean square error matrix of the ML-estimator for θ. The main argument in favour of this ability is that by using standard increasing domain asymptotics they converge to each other, however - as pointed out by Zhang and Zimmerman (2005) in this context - infill (fixed domain) asymptotics may be more reasonable. Zhang (2004) showed (extending Ying (1993) and Abt and Welch (1998)) that not all covariance parameters in the Matérn model (3) can be estimated consistently.

Example 1 *Abt and Welch (1998) considered a design space $\mathcal{X} = [0, 1]$ with the correlation function of the form (4). They showed that on the one hand* $\lim_{n\to+\infty} (M^{-1}(\theta_1, \sigma^2))_{1,1} = 0$ *but on the other hand* $\lim_{n\to+\infty} n\,(M^{-1}(\theta_1, \sigma^2))_{1,1} = 2(\sigma^2/\theta_1)^2$. *These results obtained from the information matrix coincide with the variance of the asymptotic distribution of $\sqrt{n}(\hat{\sigma}^2/\hat{\theta}_1 - \sigma^2/\theta_1)$ found in Ying (1993) based on approximations of the log-likelihood function.*

As a certain remedy, Zhang (2004) also demonstrated that the quantity $\sigma^2\theta_1^{-2\theta_2}$ can be estimated consistently and he gives small sample simulations to further back up this ratio. Furthermore, extending the results from Stein (1990), he argues that, if the ultimate goal of the analysis is prediction (see below) then it is much more useful to estimate this rather than its separate components for being the identifying quantity in characterizing compatible covariance functions (such that yield asymptotically equivalent interpolations). Therefore in Zhu and Zhang (2006) the authors suggest a corresponding reparametrization and the consequent optimization of only a subset of the information matrix M_θ. For

more references on the Fisher information as the basis of a design criterion in the correlated setup see Stehlík (2007), where the structure of the Fisher's information matrices for stationary processes is studied. Therein it is shown that under mild conditions given on covariance structures the lower bound for M_θ is an increasing function of the distances between the design points.

A different justification for the use of the information matrix M_θ is provided by Pázman (2004), who shows that the approximation holds well in exponential families of distributions as long as the errors are relatively small. In any case, even if M_θ is not a good approximation - say for small sample sizes - its usefulness for design purposes is unharmed as long as it orders designs in the same way, i.e. we can observe a monotone relationship between these matrices in this respect. This fact is supported by various simulation experiments under Gaussian random field and Matérn covariance) in Zhu and Stein (2005) and Zimmerman (2006).

2 Nonreplicability

A more basic issues arises from the fact that in computer simulations experiments due to the deterministic code the same inputs always yield exactly the same outputs, thus rendering replications in the experiments useless. Note, that due to the random field model (1) anyway replicability of the entire field only would be admissable.

As a simple, rather brute force remedy, one can employ numerical design optimization algorithms that directly generate so-called replication-free designs. The basic version of such an algorithm is due to Fedorov (1972) and consists of a simple exchange of points from the two sets S_{ξ_s} and $\bar{\mathcal{X}}_s \setminus S_{\xi_s}$ at every iteration s, namely

$$\xi_{s+1} = \left\{ \xi_s \setminus \left\{ x_s^-, \frac{1}{n_s} \right\} \right\} \cup \left\{ x_s^+, \frac{1}{n_s} \right\},$$

where

$$x_s^+, x_s^- = \arg \max_{x \in \mathcal{X}_s \setminus S_{\xi_s}, x \in S_{\xi_s}} \Phi(x, \xi_s).$$

The set $\mathcal{X}_s$ now carries a subscript to indicate the possibility of changing the discretization of $\mathcal{X}$ at every iteration.

Example 2 *As an example consider D-optimum design for a two-dimensional linear regression on the unit-square, i.e. we assume* $f(x) = (1, x_{[1]}, x_{[2]})$*;* $-1 \leq x_{[1]}, x_{[2]} \leq 1$*;* $\Phi(M) = \ln \det M$*, where* $[i]$ *in the index denotes the coordinate. A* 20×20 *point grid was used to approximate the design space. On this grid only 100 observations (without replications) were allowed. The supporting points of the resulting design are all points outside a circle as given in Figure 1. This design was constructed by a simplified Fedorov-exchange algorithm after only 200 iterations.*

A generalization of this algorithm to exchanges of more than one point at every step is straightforward (cf. Royle (2002)). Note that nonconvexity is also an issue here. Some authors have therefore employed simulated annealing (eg. van Groeningen and Stein (1998)) or branch and bound type algorithms (eg. Rasch *et al.* (1997)), however, even by simple exchange type procedures considerable gains in the criteria can be achieved (for a specifically tuned algorithm see Müller and Pázman (1998)). For the general correlated case first suggestions of an algorithm can be found in Brimkulov et al. (1980).

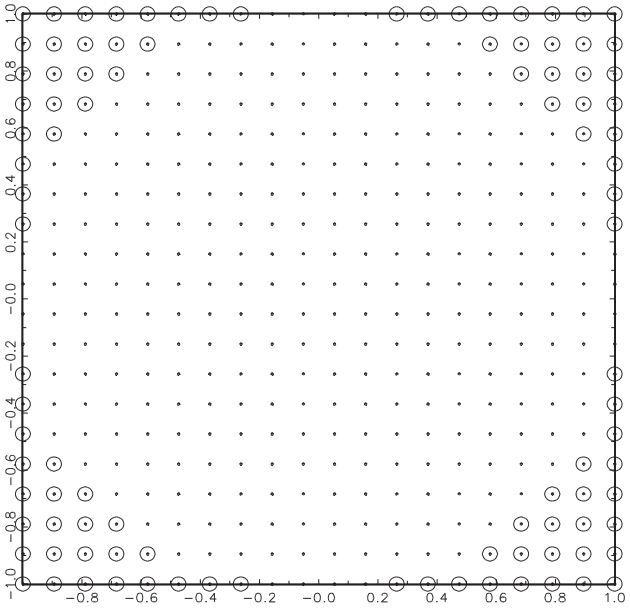


Figure 1: A replication-free optimum design for linear regression.

3 Non-additivity of the Information Matrix

Perhaps the decisive reason why optimal design principles are not very frequently applied in gathering data from computer simulation experiments is that the information matrix does not share all of the desirable properties of its analogue in the standard regression setting, which allow the construction of more sophisticated numerical design algorithms. In particular it is not additive and thus information from different design points can no longer be separated as nicely as in standard theory.

Example 3 *Here we consider the exponential covariance (4) with* $\sigma^2 = 1$ *and a linear model without slope (i.e.* $\beta_1 = 0$*). It is shown in Kiseľák and Stehlík (2007), that for n-point designs with equal interdistances* $d = x_{i+1} - x_i < 0.5$ *(so called uniform designs) the Fisher information matrix can be written as* $M_{\beta_0}(n) = \frac{2-n+ne^{d/\theta_1}}{1+e^{d/\theta_1}}$*. Then we clearly have* $M_{\beta_0}(5) \neq M_{\beta_0}(2) + M_{\beta_0}(3)$*, although such a 5-point design unifies such 2 and 3-point uniform designs by simply "adding" one edge connecting two neighboring points.*

As a remedy for this nonseparability of information contributions through designs a different interpretation of design measures as amount of signal or noise suppression (Müller and Pázman (1998) with extensions in Müller and Pázman (2003) and for covariance paremeters Pázman and Laca (2007)) has been suggested. By adding virtual noise to the process they extend the classical information matrix M_β to give

$$M^{(\epsilon)}_{\beta,\kappa}(\xi) = \sum_{x,x' \in \xi} f(x)[C(\xi) + W^{(\epsilon)}_\kappa(\xi)]^{-1}_{x,x'} f^T(x'),$$

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

$$[W^{(\epsilon)}_\kappa(\xi)]_{x,x} = \epsilon \left(\xi(x) - \kappa\right)^2 / \xi(x)$$

and respective modifications for M_θ and $M_{\beta,\theta}$.

Another remedy for this issue would of course be to change the model. By approximate the random field by a random coefficient model through Mercers expansion (see Fedorov and Müller (2007)) additivity of the information matrix is restored and the design problem is thus again embedded into classical design theory.

Note the special role that is played by uniform designs in this setup. If only trend parameters are of interest, designs covering uniformly the whole design space - such as the one used in Example 2 - are very efficient. Dette et al. (2007) have proved that if $\theta_1 \to +\infty$, then any exact n-point D-optimal design in the linear regression model with exponential semivariogram converges to the equally spaced design. In Kiseľák and Stehlík (2007) a thorough study of small sample and asymptotical comparisons of the efficiencies of equidistant designs with taking into account both the parameter of trend β_0, as well as the parameter of the covariance function θ_1 is provided. If only the trend parameter β_0 is of interest, the designs covering more-less uniformly the whole design space are rather efficient. They are also showing that for all possible combinations of parameters of interest, i.e. $\{\beta_0\}, \{\theta_1\}$ and $\{\beta_0, \theta_1\}$, the interval over which observations are to be made should be extended as far as possible. They have also shown that infill asymptotics substantially differs for the covariance and trend parameter and proved that a n-point equidistant design for parameter β_0 is D-optimal. A recurring topic in the recent literature is that uniform designs perform well in terms of model-robustness when a Bayesian approach is adopted, when the maximum bias is to be minimized or when the minimum power of the lack-of-fit test is to be maximized (see Goos, Kobilinsky, O'Brien and Vandebroek (2005) and for more general results Bischoff and Miller (2006)). The concept of uniform designs has been introduced by Fang (1978) and has now gained popularity and proven to be very successful in industrial applications (see Pham (2006), Chapter 13).

4 Unintuitive effects in gaining information

Unfortunately, when one wants to analyze correlated data stemming from a process as given in (1), one must be cautious in utilizing intuition gained from the i.i.d. case. Specifically, when the process is observable on the interval $\mathcal{X} = [-T, T]$, the mean β_0 may be estimated by the unbiased estimators

$$\bar{Y}_{n+1} := \frac{1}{n+1} \sum_{i=0}^{n} Y\left(\frac{Ti}{2n} - T\right)$$

or

$$m = 1/2T \int_{-T}^{T} Y(t)dt.$$

One might be inclined to expect that $Var(m) \leq Var(\bar{Y}_{n+1})$, since the estimator m utilizes the whole realization rather than a finite number of points. However, there are equidistant designs ξ_n for which $Var(\bar{Y}_{n+1}) < Var(m)$. This disturbing situation is what is known in the literature as Smit's paradox (see Smit (1961)) and is explained by the fact that in general m is not the best linear unbiased estimator for β_0.

Example 4 *(see Näther (1985)) Here we consider $\mathcal{X} = [-1, 1]$ and covariance function (4) with $\theta_1 = \sigma_1^2 = 1$. Let us consider the design*

$$\xi_5 = \{-1, -0.5, 0, 0.5, 1\}.$$

We have $Var(\bar{Y}_5) = 0.529$ *and* $Var(m) = 0.568$. *Furthermore, the variance of* $\bar{Y}$ *can increase by use of additional observations. Consider e.g. design*

$$\xi_9 = \{-1, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 1\}.$$

We now have $Var(\bar{Y}_9) = 0.542$, *which is less than* $Var(\bar{Y}_5)$. *The BLUE-variance is* 0.5.

Although every observation possesses non-zero variance it is even possible to estimate parameters with variance zero (see Näther (1985)). Such examples touch the case of singular processes (see Ibragimov and Rozanov (1978)).

5 Unintuitive effects in design

Similar as in the estimation stage, paradoxical effects can arise in the respective design phase. Näther (1985) provided such a case for the situation when the regression function is generated linearly by the covariance function, i.e.

$$f(x) = \sum_{x' \in \xi^*} c(x, x') g(x'), \quad \forall x \in \mathcal{X}, \tag{8}$$

where $g(x) \in R^m$ are given vectors and ξ^* is some finite set representing the support of an optimum design (for a simple proof see Müller and Pázman (2003)).

Example 5 *Näther (1985) showed that for a linear model (1) with a triangular covariance function (2) with* $\theta_1 = 1$ *the design* $\{-1, 0, 1\}$ *fulfills (8) and is thus a uniformly optimal design for* (β_0, β_1), *which paradoxically cannot be improved by addition of further experimental points. It is instructional to see whether this 'deficiency' holds over a wider range for* θ_1. *E.g. Müller and Stehlík (2004) show, that when* $0 < \theta_1 < 2$ *(both correlated and uncorrelated observations are possible) we obtain two equivalent 3-point D-optimal designs* $\{-1, \pm(1 - \theta_1), 1\}$ *which can but be improved (see Stehlík (2005)) for* $\theta_1 \neq 1$ *by addition of a further experimental point, e.g. by putting* $\xi_4 = \{-1, -1 + \theta_1, 1 - \theta_1, 1\}$. *The amount of this improvement (the so-called efficiency function, i.e. the ratio* $\max \Phi(M_3)/\Phi(M(\xi_4)))$ *is displayed in Figure (5), from which it can be seen that the saving of one observation has dramatic impact only for low correlations.*

There is no quick remedy for either of the effects described in sections 4 and 5. The lesson to be learned is rather that one must very carefully investigated correlated cases "situation by situtaion".

6 The Impact of Dependence on Information

Concerning this issue one might be misled in thinking that the impact of dependence in the data is always negative, since it implies redundancy in information. However, this is not at all the case and it very much depends upon what is estimated, so that dependence could actually be an advantage (which is very transparent in the case of estimating the slope of a linear regression from two design points only).

For a further discussion consider $\mathcal{X} = [-1, 1]$ and covariance structure (2). Now, when both the slope and intercept are our parameters of interest (θ_1 is fixed), then the larger

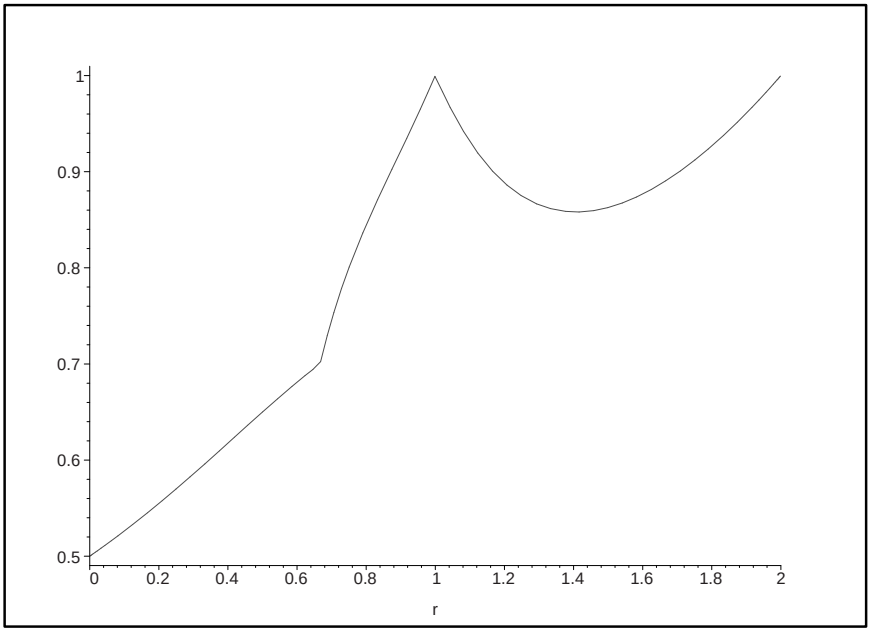


Figure 2: The efficiency of the 4-th point.

covariance is, the more information about the parameters D-optimal design gains, (for $\theta_1 \leq 2$ we gain the same information, because here independent observations are possible). When θ_1 is treated as parameter of interest, then for $\theta_1 > 2$ information decreases with increasing of θ_1, for every $\theta_1 \leq 2$ approximate D-optimal design gains the same information (see also Figure 3). Thus, when θ_1 is fixed, then information increases with correlation (for $\theta_1 \leq 2$ is information constant (independent observations are possible)). When θ_1 is a parameter of interest, then for $\theta_1 > 2$ information decrease with increasing of θ_1, for $\theta_1 \leq 2$ approximate D-optimal design gains the same information. When the intercept is the only parameter of interest the information is decreasing with correlation. When only the slope is the parameter of interest, the information is increasing with correlation (see Stehlík (2004a)).

Example 6 *When $\theta_1 > 2$ and θ_1 is not parameter of interest, then 3-point DOD gives for (β_0, β_1) the same value of Φ as the 2-point one, i.e.* $\max \Phi(M_2) = \max \Phi(M_3) = \frac{\theta_1^2}{\theta_1 - 1}$. *When $\theta_1 > 2$ and θ_1 is parameter of interest, then 2-point DOD gives a smaller value of Φ as the 3-point one and we have (for a proof see Stehlík (2005))*

$$\Phi(M_3) = \left[2 - \frac{1}{1 + (\theta_1 - 1)^2}\right] \Phi(M_2).$$

That is, the ratio $\Phi(M_3)/\Phi(M_2)$ is an increasing function of θ_1 (the larger the correlation the larger is the ratio). Notice, that $\Phi(M_3)/\Phi(M_2) > 1.5$.

7 The Choice of the Dependence Structure

Misspecifications in the covariance function may lead to disturbing contradictions. E.g. for the exponential function the optimal design collapses into one point and does thus not

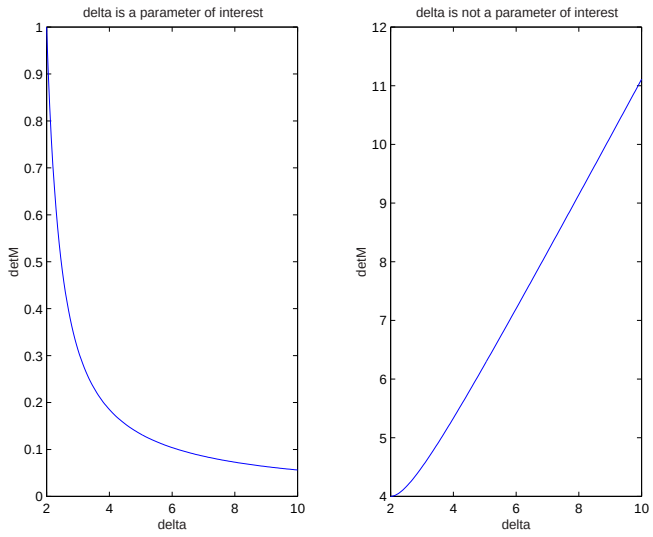


Figure 3: The minimum of the criterion function for the intercept (θ_1 fixed (left) and estimable (right).

convey information under say the triangular function (see Stehlík et al. (2007)). As is pointed out in Näther (1985), one of the fundamental assumptions, the knowledge of the covariance function, is in most cases almost unrealistic. "It seems to be artificial, that the first moment $E(Y(x))$ is assumed to be unknown whereas the more complicated second one is assumed to be known..."

Example 7 *When the exponential covariance structure holds and we use the linear one, then we obtain the information loss displayed in Table 1. When the covariance structure is exponential and we use the D-optimal design for the linear one, we measure only* 15.48% *of the maximal information about θ_1 for a two-point D-optimal design with the design space length equal to 2 (see also Figure 4). The two-point optimal design for covariance*

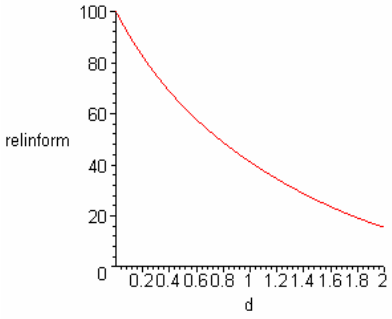


Figure 4: The relative information gain, in %,

parameter θ_1 under covariance (2) with $\sigma^2 = 1$ is the maximal distant. Let $\theta_1 = 1$. When the covariance structure is linear and we use the DOD for exponential one, we measure almost none of the maximal information about θ_1.

Table 1: The information loss

design	2-point design	3-point design	4-point design	5-point design
loss	0%	9.23%	11.52%	12.38%

It is notable, that two point optimal design for covariance parameter θ_1 under exponential covariance structure is collapsing. A similar effect has been studied and illustrated by Crary (2001) for AMSE-based criteria. This collapsing effect holds also for the pair (β_0, θ_1). Kiseľák and Stehlík (2007) have proved (see Theorem 2) that the 3-point D-optimal equidistant design for (β_0, θ_1) is also collapsing, however for more points the D-optimal equidistant design is not collapsing anymore.

8 The Role of a Nugget Effect

The so-called nugget effect is usually employed to model local disturbances in the random field (cf. Cressie (1993)). It is an important practical extension of covariance functions to capture micro-scale variation. Furthermore, as will be seen below, it acts as an important regulatory tool to make many designs feasible.

Example 8 *Let us consider the exponential covariance (4) with $\sigma^2 = 1$. Let θ_1 be the parameter of interest and let us consider 2-point design. In Stehlík (2007) we can find that $M_\theta(2) = \frac{d^2\, e^{-\frac{2d}{\theta_1}}(e^{-\frac{2d}{\theta_1}}+1)}{(e^{-\frac{2d}{\theta_1}}-1)^2\,\theta_1^4}$ Thus the maximal Fisher information is obtained for $d = 0$, in other words, the 2-point design is collapsing into a 1-point design.*

To avoid such 'inconvenient' behavior, Stehlík (2004b) suggested to decrease the non-diagonal elements by multiplying with a factor α, $0 < \alpha < 1$. By this we include a nugget effect of the form $1 - \alpha$. In Stehlík et al. (2007) it is further proved, that the distance d of the optimal design is an increasing function of the nugget $1 - \alpha$. Thus the distance of the two point D-optimal design for covariance parameter r of exponential covariance can be tuned by the nugget

$$\tau^2 = \lim_{d \to 0} \frac{1}{2} Var(Y(x+d) - Y(x)).$$

Process $Y(x)$ with non-zero nugget effect is not L^2-continuous. The effect of the multiplicative perturbation of the triangular covariance structure is studied in Stehlík (2004b) and the impact of the nugget effect in the example of lung's retention of radioactive particles in Stehlík et al. (2007).

Appendix

A digression into prediction

It is frequently the ultimate goal of a spatial study to predict the random field at given sites or over a continuous region as precise as possible. If the covariance parameters must be estimated from the same dataset, it is evident that the additional uncertainty alters the kriging variance and thus requires adaptations in the AMSE based design criteria.

Specifically the mean squared error in for a specific location x_0 must now be separated into

$$E[(\hat{y}(x_0;\hat{\theta}) - y(x_0))^2] = \mathrm{Var}[\hat{y}(x_0)] + E[\hat{y}(x_0;\hat{\theta}) - \hat{y}(x_0)],$$

i.e. the traditional kriging variance plus a term that can be approximated by the trace $\mathrm{tr}\left\{M_\theta^{-1}\mathrm{Var}[\partial\hat{y}(x_0)/\partial\theta]\right\}$, cf. Harville and Jeske (1992) and Zimmerman and Cressie (1992).

Consequently, Zimmerman (2006) uses

$$\max_{x\in\mathcal{X}}\left\{\mathrm{Var}[\hat{y}(x)] + \mathrm{tr}\left\{M_\theta^{-1}\mathrm{Var}[\partial\hat{y}(x)/\partial\theta]\right\}\right\} \tag{9}$$

(with the maximum over $\mathcal{X}$ rather than the integral) as a design criterion, which he terms EK-(empirical kriging-)optimality. He points out that the criterion naturally combines quality measures on prediction as well as parameter estimation. He then demonstrates by various examples that the respective designs somehow balance the space-filling as well as the clustering characteristic that are so typical for these purposes.

One step further is taken by Zhu and Stein (2006), who argue that one not only wants to precisely predict the random field, but also wants to efficiently quantify the quality of these predictions. Accordingly they supplement the criterion (9) by an approximation of the variance of the mean square error

$$\mathrm{Var}[E[(\hat{y}(x_0;\hat{\theta}) - y(x_0))^2]] \simeq \left(\frac{\partial\mathrm{Var}[\hat{y}(x_0)]}{\partial\theta}\right)^T M_\theta^{-1}\left(\frac{\partial\mathrm{Var}[\hat{y}(x_0)]}{\partial\theta}\right).$$

Note that this - defined on a center point x_0 of a design region - is used as the primary criterion by Lark (2002).

The integral over a linear combination

$$\int_{x\in\mathcal{X}}\mathrm{Var}[\hat{y}(x)]+\mathrm{tr}\left\{M''^{-1}(\theta)\mathrm{Var}[\partial\hat{y}(x)/\partial\theta]\right\}+\left(\frac{\partial\mathrm{Var}[\hat{y}(x)]}{\partial\theta}\right)^T\frac{M''^{-1}(\theta)}{2\mathrm{Var}[\hat{y}(x)]}\left(\frac{\partial\mathrm{Var}[\hat{y}(x)]}{\partial\theta}\right)dx$$

is called the estimation adjusted (EA-)criterion by Zhu and Stein (2006). In their examples they find that EA yields designs that assign only 3-10% of their observations to estimation rather than prediction, but thereby gaining a reasonable advantage over solely prediction based designs. Zhu and Zhang (2006) modify the criterion making it more effective for asymptotically unidentifiable covariance parameters.

More on the Slepian process

In distribution $S(t) = W(t+1) - W(t)$, where $W(t)$ is the Brownian motion. For small deviations and other references on Slepian process see Nikitin (2005). The eigenvalues of the kernel $\mathrm{cov}(x,z)$ of Slepian process cannot be found explicitly. However, it was proved by Nikitin and Orsingher (2004) that the spectrum consists of two series of eigenvalues. The first of them are numbers $\lambda_k = (2\pi(k-1/2)^2)^{-1}, k = 1,2,..,$ the second series of eigenvalues are the roots r_n of the transcendental equation $\tan(1/\sqrt{2r}) = \sqrt{2r}, r > 0$.

The Example 0 related computations in Mathematica

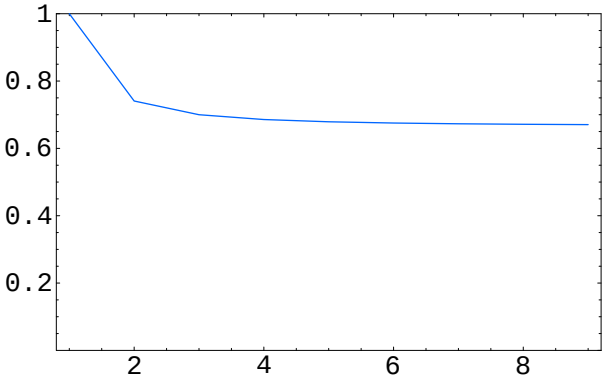
```
Clear
<< Statistics`
<< Graphics`
```

Clear

```
Clear[k]
var1[k_] = 1/(2 * Sum[i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]);
var2[k_] = 2 * Sum[1/i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]/(2k)^2;
k = 1000; (var1[k - 1]) * k//N
```

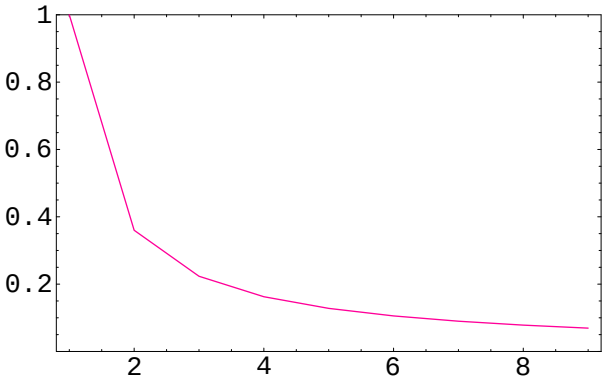
1.5

```
a1 = ListPlot[Table[1/var1[k - 1]/(k), {k, 2, 10}], Frame → True, Axes → None,
  PlotJoined → True, PlotStyle → Hue[.6],
PlotRange → {0, 1}]
```



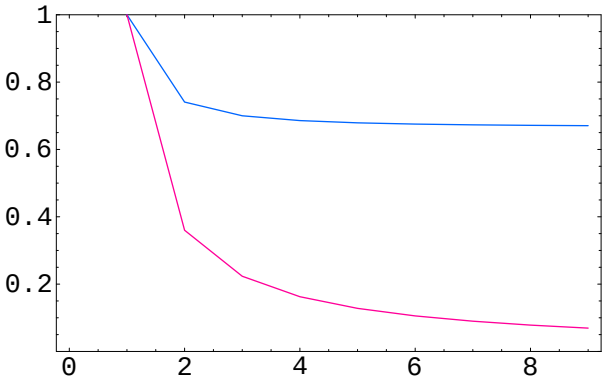
−Graphics−

```
a2 = ListPlot[Table[var1[k - 1]/var2[k - 1], {k, 2, 10}], Frame → True, Axes → None,
  PlotJoined → True, PlotStyle → Hue[.9],
  PlotRange → {0, 1}
```



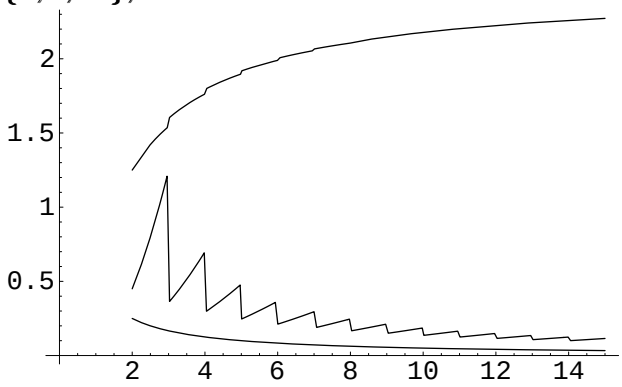
−Graphics−

```
Show[{a1, a2}]
```



−Graphics−

```
Plot[{1/(2k), 2 * Sum[1/i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]/(2k)^2, }]
Plot[1/(2 * Sum[i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}])]
{k, 2, 15}, PlotPoints → 7
```

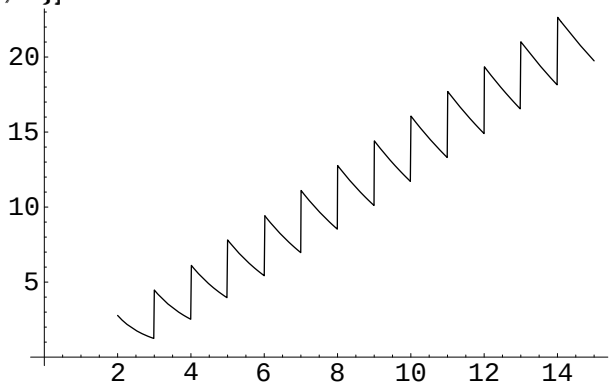


−Graphics−

```
Limit[2 * Sum[1/i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]/(2k)^2, k → Infinity]//N
```

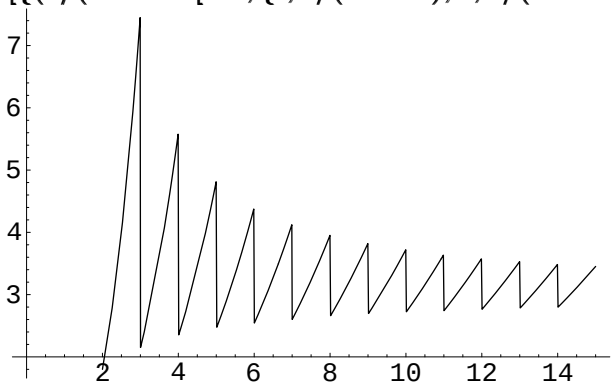
2.4674

```
Clear[k] Plot[{(2 * Sum[1/i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]/(2k)^2)/(1/(2 * Sum[i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]))},
{k, 2, 15}]
```



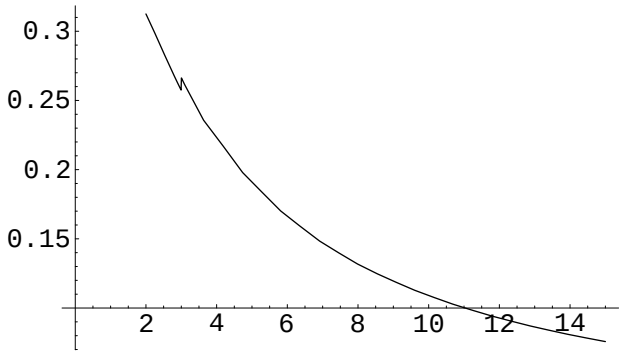
−Graphics−

```
Plot[{(1/(2 * Sum[i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]) * (2k))}, {k, 2, 15}]
```



−Graphics−

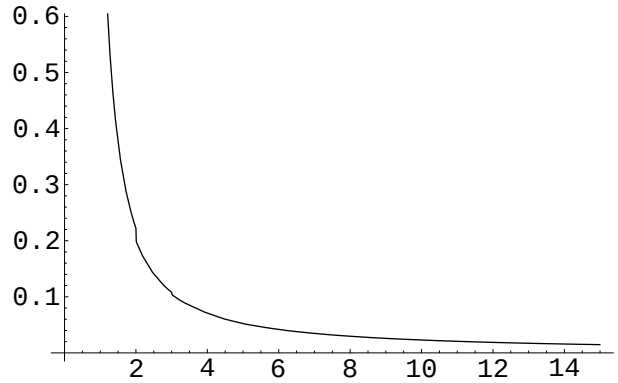
```
Plot[{(2 * Sum[1/i^2, {i, 1/(2k - 1), 1, 2/(2k - 1)}]/(2k)^2)/(2k)}, {k, 2, 15}]
```



−Graphics− **vt[k_] = 2 * Sum[1/i^2, {i, 1/(2k − 1), 1, 2/(2k − 1)}]/(2k)^2//N**

$\frac{1}{k^2}\left(0.5\left(1.2337(-1.+2.k)^2 - 0.25(1.-2.k)^2 \text{PolyGamma}\left[1, 1.5 + \text{Floor}\left[0.5(-1.+2.k)\left(1.-\frac{1.}{-1.+2.k}\right)\right]\right]\right)\right)$

Plot[1/(vt[k] * 2 * k), {k, 1, 15}]

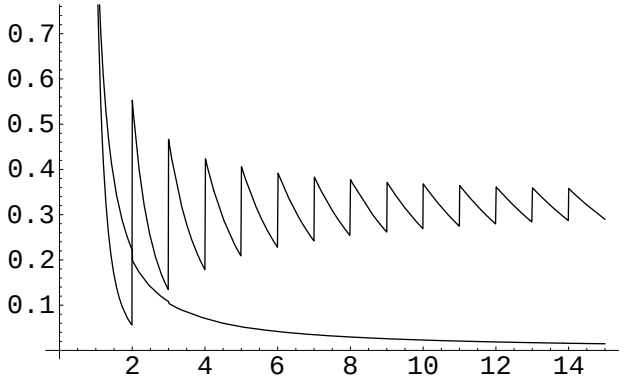


−Graphics−

vh[k_] = 1/(2 * Sum[i^2, {i, 1/(2k − 1), 1, 2/(2k − 1)}])

$\frac{3(-1+2k)^2}{2\ 1+\text{Floor}\left[\frac{1}{2}(-1+2k)\ \ 1-\frac{1}{-1+2k}\right]\ \ 1+2\text{Floor}\left[\frac{1}{2}(-1+2k)\ \ 1-\frac{1}{-1+2k}\right]\ \ 3+2\text{Floor}\left[\frac{1}{2}(-1+2k)\ \ 1-\frac{1}{-1+2k}\right]}$

Plot[{1/(vt[k] * 2 * k), 1/(vhn[k] * 2 * k)}, {k, 1, 15}]



−Graphics−

vt[3]

1.59877

vhn[k_] = (2 * k − 1)^2/(2 * Sum[(i^2), {i, 1, 2 * k − 1, 2}])

$(3(-1+2k)^2)\,/\,\left(2\left(1+\text{Floor}\left[\frac{1}{2}(-2+2k)\right]\right)\left(1+2\text{Floor}\left[\frac{1}{2}(-2+2k)\right]\right)\left(3+2\text{Floor}\left[\frac{1}{2}(-2+2k)\right]\right)\right)$

Limit[vhn[k], k → Infinity]//N

∞

20/16//N

1.25

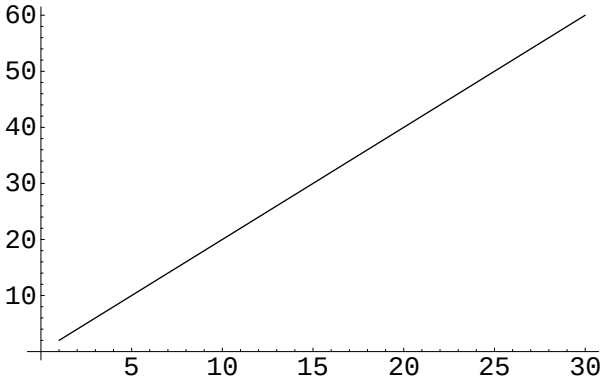
Plot[{2 * k, (vt[k]/vhn[k])^(1/k)}, {k, 1, 30}]

Plot::plnr : $\left(\frac{\text{vt}[k]}{\text{vhn}[k]}\right)^{1/k}$ is not a machine-size real number at k = 1.0000012083333334. Mehr...

Plot::plnr : $\left(\frac{\text{vt}[k]}{\text{vhn}[k]}\right)^{1/k}$ is not a machine-size real number at k = 2.1764427556145582. Mehr...

Plot::plnr : $\left(\frac{\text{vt}[k]}{\text{vhn}[k]}\right)^{1/k}$ is not a machine-size real number at k = 3.4594551959218367. Mehr...

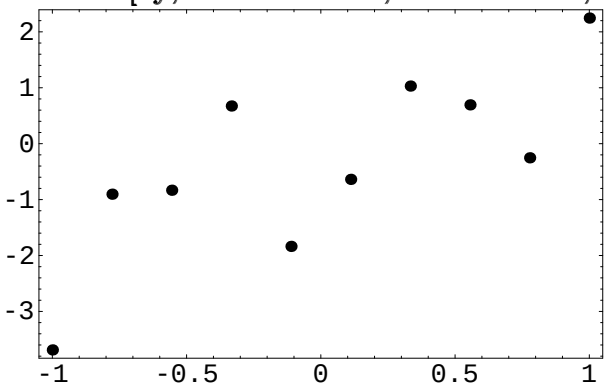
General::stop : Further output of Plot::plnr will be suppressed during this calculation. Mehr...



−Graphics−

xy = Table[{x, 2x + Random[NormalDistribution[0, 1]]}, {x, −1, 1, 2/9}];

g0 = ListPlot[xy, Frame → True, Axes → None, PlotStyle → PointSize[0.02]]



−Graphics−

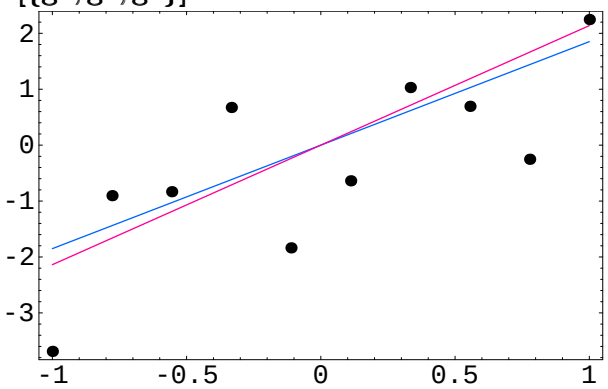
yx = Transpose[xy]; y = yx[[2]]; x = yx[[1]]; b1 = Dot[x, y]/Dot[x, x]; b2 = Mean[y/x]

2.13785

g1 = ListPlot[Table[{x, b1 * x}, {x, −1, 1, 2/9}], Frame → True, Axes → None, PlotJoined → True, PlotStyle → Hue[.6]];

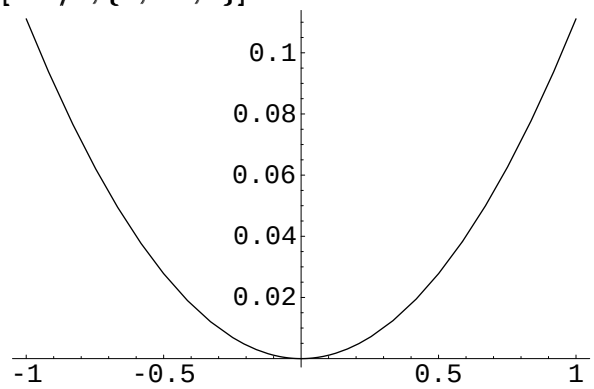
g2 = ListPlot[Table[{x, b2 * x}, {x, −1, 1, 2/9}], Frame → True, Axes → None, PlotJoined → True, PlotStyle → Hue[.9]];

Show[{g1, g2, g0}]



−Graphics−

7.75134
3.14/6
0.523333
NLimit[Sum[1/i^2, {i, 1, n, 2}], n → Infinity]
$\text{NLimit}\left[\frac{\pi^2}{8} - \frac{1}{4}\text{PolyGamma}\left[1, \frac{3}{2} + \text{Floor}\left[\frac{1}{2}(-1+n)\right]\right], n \to \infty\right]$
3.14^2/8 * 2
2.4649
Plot[x^2/9, {x, −1, 1}]



−Graphics−

Acknowledgement

This work was partially supported by the FWF Project P-16809-N12 and by the ASO Project Nr. SK-0607-BA-018

References

Abt M. and Welch W.J. (1998) Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes, *The Canadian Journal of Statistics,* Vol. 26, No. 1: 127–137.

Bischoff W. and Miller F. (2006) Optimal designs which are efficient for lack of fit tests, *The Annals of Statistics,* Vol. 34, No. 4, 20152025

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.

Casella G. and Berger R.L. (2002) *Statistical Inference,* 2nd edition, Duxbury Press

Crary S.B. (2002) Design of Computer Experiments for Metamodel Generation, *Analog Integrated Circuits and Signal Processing*, Volume 32, Number 1, p. 7-16(10)

Crary S.B., Woodcock D. M. and Hieke A. (2001) Designing Efficient Computer Experiments for Metamodel Generation, *The Proceedings of the Fourth International Conference on Modeling of Microsystems,* MSM 2001, Hilton Head, SC, March 19-21, 2001, 132-135.

Cressie N.A.C. (1993) *Statistics for Spatial Data*, 2nd ed., New York: Wiley.

Dette H., Kunert J. and Pepelyshev A. (2007) Exact optimal designs for weighted least squares analysis with correlated errors, accepted to *Statistica Sinica*.

Fang K.T. (1978) Uniform design: application of number-theoretic methods in experimental design, *Prob. Stat. Bull.* 1, p. 56-97.

Fang K.T., Runze L. and Sudjianto A.(2005). *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC

Fedorov V.V. (1989). Kriging and other estimators of spatial field characteristics. *Atmospheric Environment*, 23(1):175-184.

Fedorov, V.V. and Müller, W.G., Optimum design for correlated fields via Covariance Kernel expansions, in *mODa 8-Advances in Model Oriented Design and Analysis,* López-Fidalgo, J., Rodríguez-Díaz, J. and Torsney, B. (Eds.), Physica Verlag, 57-66.

Guttorp, P. and Gneiting, T. (1994). Miscellanea: Studies in the history of probability and statistics XLIX. On the Matérn correlation family. *Biometrika*, 93(4):989–995.

Goos P., Kobilinsky A., O'Brien T. E. and Vandebroek M. (2005). Model-robust and model-sensitive designs, *Computational Statistics & Data Analysis* 49, p. 201-216.

Harville, D.A. and Jeske, D.R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, 87:724-731.

Ibragimov I.A. and Rozanov Yu.A. (1978) *Gaussian random processes*, Springer (Translated from Russian)

Kiseľák J. and Stehlík M. (2007) Equidistant D-optimal designs for parameters of Ornstein-Uhlenbeck process, *IFAS research report* 22.

Lark, R.M. (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma*, 105:49–80.

Müller W.G. (2007). *Collecting Spatial Data*, 3rd edition, Physica Verlag.

Müller, W.G. and Pázman, A. (1998) Design Measures and Extended Information Matrices for Experiments without Replications, *Journal of Statistical Planning and Inference,* 71, 349-362, 1998.

Müller, W.G. and Pázman, A. (2003) Measures for Designs in Experiments with Correlated Errors, *Biometrika*, 90(2), 423- 434, 2003.

Müller W.G. and Stehlík M. (2004) An example of D-optimal designs in the case of correlated errors, *Proceedings in Computational Statistics*, Edited by J. Antoch, Physica-Verlag, 1543-1550.

Naylor T.H., Burdick, D. S. and Sasser W. E., Computer Simulation Experiments with Economic Systems: The Problem of Experimental Design, *Journal of the American Statistical Association,* Vol. 62, No. 320. (Dec., 1967), 1315-1337.

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

Nikitin Ya. Yu. (2005) Small deviations of Gaussian processes as extreme behavior. *Proceedings of the 5th St. Petersburg Workshop on Simulation,* Edited by S.M. Ermakov, V.B. Melas and A.N. Pepelyshev, 531-536.

Nikitin, Ya. Yu. and Orsingher, E. (2004) Exact small deviation asymptotics for Slepian and Watson processes in Hilbert norm, *Notes of Scient. Seminars of St. Petersburg Dept. of Steklov Math. Institute,* 320.

Pardo-Igúzquiza, E. and Dowd, P.A. (2005). Multiple indicator cokriging with application to optimal sampling for environmental monitoring. *Computers & Geosciences*, 31(1):1–13.

Pázman A. (2004). Correlated optimum design with parametrized covariance function: Justification of the use of the Fisher information matrix and of the method of virtual noise, *Research Report Series* of the Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, to appear in *Kybernetika 2007.*

Pázman, A. and Laca, J. (2007) The method of virtual noise for optimal design in experiments with estimated covariance function. to appear in *Proceedings of Probastat2006*, Tatra Mountains Mathematical Publishers.

Pham H. (Editor) (2006). *Springer Handbook of Engineering Statistics,* Springer Verlag, London, Chapter 13, 229-247.

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

Rasch, D.A.M.K., Hendrix, E.M.T. and Boer, E.P.J. (1997). Replicationfree optimal designs in regression analysis. *Computational Statistics*, 12:19–32.

Royle J. A. (2002). Exchange algorithms for constructing large spatial designs, *Journal of Statistical Planning and Inference* 100, p. 121-134.

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.

Santner T.J., Williams, B.J. and Notz W. I. (2003). *The Design and Analysis of Computer Experiments,* Springer Series in Statistics.

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

Slepian D. (1961). First Passage Time for a Particular Gaussian Process. *The Annals of Mathematical Statistics,* Vol. 32, No. 2., 610-612.

Stein, M.L. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *The Annals of Statistics*, 18:850–872.

Steinberg, D. M., Bursztyn, D., 2006. Comparison of designs for computer experiments. *Journal of Statistical Planning and Inference* 136, 1103-1119.

Stehlík M. (2004a) Further aspects on an example of D-optimal designs in the case of correlated errors, Report #1 of the *Research Report Series* of the Department of Statistics and Mathematics, Wirtschaftsuniversität Wien.

Stehlík M. (2004b) Some Properties of D-optimal Designs for Random Fields with Different Variograms, Report #4 of the *Research Report Series* of the Department of Statistics and Mathematics, Wirtschaftsuniversität Wien.

Stehlík M. (2005). Covariance related properties of D-optimal correlated designs, *Proceedings of the 5th St. Petersburg Workshop on Simulation*, Edited by S.M. Ermakov, V.B. Melas and A.N. Pepelyshev, 645-652.

Stehlík M. (2007). D-optimal designs and equidistant designs for stationary processes, in *mODa 8-Advances in Model Oriented Design and Analysis,* López-Fidalgo, J., Rodríguez-Díaz, J. and Torsney, B. (Eds.), Physica Verlag, 205-211.

Stehlík M., Rodríguez-Díaz J.M., Müller, W.G. and López-Fidalgo, J. (2007) Optimal allocation of bioassays in the case of parametrized covariance functions: an application in lung's retention of radioactive particles, *TEST* (in press).

van Groenigen, J.W. and Stein, A. (1998). Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environomental Quality*, 43: 684-691.

Watson G. S. (1961). Goodness-Of-Fit Tests on a Circle. *Biometrika,* Vol. 48, No. 1/2., 109-114.

Xia G., Miranda M.L. and Gelfand A.E. (2006). Approximately optimal spatial design approaches for environmental health data, *Environmetrics,* 17: 363–385.

Ying Z. (1993). Maximum Likelihood Estimation of Parameters under a Spatial Sampling Scheme, *The Annals of Statistics,* Vol. 21, No. 3., p. 1567-1590.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.

Zhang H. and Zimmerman D. (2005). Towards reconciling two asymptotic frameworks in spatial statistics, *Biometrika,* vol. 92, no 4: 921–936.

Zhu Z. and Stein M.L. (2005). Spatial sampling design for parameter estimation of the covariance function, *Journal of Statistical Planning and Inference,* vol. 134, no 2: 583-603.

Zhu, Z. and Stein, M. (2006). Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1):24–44.

Zhu, Z. and Zhang, H. (2006). Spatial sampling design under in-fill asymptotic framework. *Environmetrics*, 17:323–337.

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

Zimmerman, D.L. and Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, 44(1):27–43.