Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2009-45

Towards an Optimal Design Equivalence Theorem for Random Fields?

Werner G. Müller and Luc Pronzato[a]

August 2009

[a]Laboratoire I3S, CNRS/Université de Nice-Sophia Antipolis, Les Algorithmes, 2000 route des lucioles, BP 121, 06903 Sophia Antipolis cedex, France

Abstract

Two approaches are considered to design experiments for a correlated random field when the objective is to obtain precise predictions over the whole experimental domain. Both take the uncertainty of the estimated parameters of the correlation structure of the random field into account. The first one corresponds to a compound D-optimality criterion for both the trend and covariance parameters. The second one relies on an approximation of the mean squared prediction error already proposed in the literature. It is conjectured, and shown on a small example, that for some particular settings both approaches yield similar optimal designs, thereby revealing a sort of equivalence theorem for random fields: optimal designs for parameter estimation (obtained through a suitable criterion) are optimal (or close to optimal) for prediction/interpolation of the random field. Keywords: Computer experiments, Correlated errors, Experimental design

1 Introduction

For the development of design theory for experiments with independent observations, the so called equivalence theorem of Kiefer and Wolfowitz (1960), henceforth KWET, and its extensions, have played a major role. It allows to quickly check whether given designs are optimal, and led to the development of efficient algorithms for constructing good designs. In particular, sequential constructions exist which achieve the same asymptotic prediction efficiency as non-sequential designs that use perfect knowledge of the system parameters and are much more demanding in terms of computational cost. One of the key aspects of the KWET is the establishment of the equivalence of optimal designs between two criteria of optimality, one related to parameter estimation, the other related to prediction (classically between D- and G-optimality).

Unfortunately, in the design and analysis for correlated random fields, given by

$$Y_t(x) = \eta(x, \beta) + \varepsilon_t(x), \qquad t \in \mathcal{T} \subset \mathbb{R}, \tag{1}$$

most of the conditions of the KWET are not met. Here, $t \in \mathcal{T}$ indexes the different realizations of the field, β is an unknown vector of parameters in $\mathbb{R}^p$, x a known vector of regressors belonging to some set $\mathcal{X}$, and the random term $\varepsilon(x)$ has zero mean, (unknown) variance σ^2 and a parameterized spatial error correlation structure such that $\mathbb{E}[\varepsilon_t(x)\,\varepsilon_t(x')] = \sigma^2 c(x, x'; \gamma)$. It is often assumed that the deterministic term has a linear structure, i.e., $\eta(x, \beta) = f^\top(x)\beta$, and that the random field $\varepsilon_t(x)$ is Gaussian, allowing efficient estimation of β, σ and γ by Maximum Likelihood.

Note that setup (1) is used in such diverse areas of spatial data analysis (cf. Cressie, 1993) as mining, hydrogeology, natural resource monitoring and environmental science, and has become the standard modelling paradigm in computer simulation experiments, following the seminal paper of Sacks et al. (1989). In that case all realizations of the field are identical, $Y_t(x) \equiv Y_{t'}(x)$ for all t, t', and $\lim_{x' \to x} c(x', x; \gamma) = c(x, x; \gamma) = 1$ for all x.

In this note we will argue why we believe, that a KWET-type relationship could also be achieved in setup (1) and demonstrate on a simple example how it could

work. The detailed investigation of its potential and limitations is left to future work.

2 Motivation

We must carefully distinguish between two fundamentally distinct problems that are both usually designated as "prediction" problems. One may refer to Pronzato (2008) for a detailed overview of the differences between the two situations.

2.1 Prediction of a distinct realisation (parameter estimation): $\{Y_t(x), x \in \xi\} \rightarrow \{Y_{t'}(x), x \in \mathcal{S}\}$

In this case, observations in the model (1) at a given $t \in \mathcal{T}$ and sites ξ are used to estimate a future realization of the random field, at a different $t' \neq t$, and in general at different sites, $\mathcal{S} \neq \xi$. This problem has usually been assessed under the hypothesis that realizations at different times are independent and identically distributed. Under this assumption, prediction amounts to estimation of the deterministic term $\eta(x, \beta)$, and requires estimation of the parameters β in (1), say by $\hat{\beta}^n$. The prediction of a future realization of the field $Y_{t'}(x)$ is then simply $\hat{\eta}^n(x) = \eta(x, \hat{\beta}^n)$. Note that here, the influence of γ is somewhat hidden, but estimators of $\hat{\beta}^n$ will generally depend upon it, such that precise estimation of $\hat{\gamma}$ is required as well.

For this case, particularly for determining a so-called D-optimal design (maximizing the information matrix determinant), Müller and Stehlík (2009b) have suggested to maximize a compound criterion with weighing factor α,

$$\Phi[\xi|\alpha] = |M_\beta(\xi, \gamma)|^\alpha \cdot |M_\gamma(\xi, \gamma)|^{(1-\alpha)}, \tag{2}$$

which consists of determinants of information matrices corresponding to trend and covariance parameters, stemming from

$$\mathbb{E}\left\{ \begin{array}{cc} -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\beta\partial\beta^\top} & -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\beta\partial\gamma^\top} \\ -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\gamma\partial\beta^\top} & -\frac{\partial^2 lnL(\beta,\gamma)}{\partial\gamma\partial\gamma^\top} \end{array} \right\} = \left(\begin{array}{cc} M_\beta(\xi,\gamma) & 0 \\ 0 & M_\gamma(\xi,\gamma) \end{array} \right), \tag{3}$$

where, for the linear model

$$M_\beta(\xi, \gamma) = \frac{1}{n\sigma^2} \sum_{x\in\xi} \sum_{x'\in\xi} f(x)[C_n^{-1}(\gamma)]_{x,x'} f^\top(x'),$$

and

$$\{M_\gamma(\xi, \gamma)\}_{ii'} = \frac{1}{2} \operatorname{tr} \left\{ C_n^{-1}(\gamma) \frac{\partial C_n(\gamma)}{\partial \gamma_i} C_n^{-1}(\gamma) \frac{\partial C_n(\gamma)}{\partial \gamma_{i'}} \right\},$$

using notation $\{C_n(\gamma)\}_{ii'} = c(x_i, x_{i'}; \gamma)$, $i, i' = 1, \ldots, n$. In terms of experimental design, the framework is not much different from the standard one (for which the KWET holds), the difference being that for fixed t the errors $\varepsilon_t(x)$ in (1) are correlated.

2.2 Prediction of the same realization (inter-/extrapolation): $\{Y_t(x), x \in \xi\} \to \{Y_t(x), x \in \mathcal{X} \neq \xi\}$

The situation here is very different from that encountered in the setting above. Indeed, even in the idealized framework where β and the parameters σ^2 and γ of the covariance function of $\varepsilon_t(x)$ in (1) are known, predicting the value of $Y_t(x)$ at an unsampled site x requires the collection of (neighboring) observations on the particular realization of the field, whereas in a setting where the $\varepsilon_t(x)$ are (spatially) uncorrelated, prediction at $x \notin \xi$ is simply given by $\eta(x, \beta)$ when β is known and observations are useless.

Contrasting parameter estimation and interpolation/extrapolation, the second task is usually treated by minimizing a functional of the so-called kriging variance $\mathrm{Var}[\hat{Y}_t(x|\xi)] = \mathbb{E}[(\hat{Y}_t(x|\xi) - Y_t(x))^2]$ at site x, interpreted as the unconditional Mean-Squared Prediction Error (MSPE) for the best linear unbiased predictor at x. For instance one may minimize the maximum of it over a set $\mathcal{X}$,

$$\min_{\xi} \max_{x \in \mathcal{X}} \mathbb{E}[(\hat{Y}_t(x|\xi) - Y_t(x))^2]. \tag{4}$$

Here, $\hat{Y}_t(x|\xi)$ denotes the best-linear unbiased predictor of $Y_t(x)$ based on the design points in ξ and associated observations $Y_t(\xi) = [Y_t(x_1), \cdots, Y_t(x_n)]^\top$.

Assume that γ is known. In the linear setting (universal kriging, with $\eta(x, \beta) = f^\top(x)\beta$ in (1) a polynomial in x), it takes the form

$$\hat{Y}_t(x|\xi) = f^\top(x)\hat{\beta} + c_n^\top(x, \gamma) C_n^{-1}(\gamma)[Y_t(\xi) - F_n\hat{\beta}]\,,$$

where $\{c_n(x, \gamma)\}_i = c(x, x_i; \gamma)$, $i = 1, \ldots, n$, and $\hat{\beta} = \hat{\beta}(\gamma)$ is the weighted Least-Squares estimator of β in the linear regression model, that is

$$\hat{\beta}(\gamma) = [F_n^\top C_n^{-1}(\gamma) F_n]^{-1} F_n^\top C_n^{-1}(\gamma) Y_t(\xi)\,,$$

with $F_n = [f(x_1), \ldots, f(x_n)]^\top$. We can write

$$\hat{Y}_t(x|\xi) = v_n^\top(x, \gamma) Y_t(\xi)$$

with $v_n^\top(x, \gamma) \in \mathbb{R}^n$. The MSPE is given by

$$\begin{aligned} MSPE_\xi(x, \sigma^2, \gamma) &= \sigma^2 \left\{1 - c_n^\top(x, \gamma) C_n^{-1}(\gamma) c_n(x, \gamma) \right. \\ &\quad \left. + g_n^\top(x, \gamma)[F_n^\top C_n^{-1}(\gamma) F_n]^{-1} g_n(x, \gamma)\right\} \end{aligned}$$

with $g_n(x, \gamma) = f(x) - F_n^\top C_n^{-1}(\gamma) c_n(x, \gamma)$. Note that the MSPE depends on (σ^2, γ), with σ^2 intervening only as a multiplicative factor. For a recent discussion of the related design problem in a slightly different context see Harman and Štulajter (2009).

The situation gets more complicated when the covariance parameters are estimated (by Maximum Likelihood) from the same dataset. Indeed, the resulting additional uncertainty then needs to enter the design criterion. For instance, following the approach of Harville and Jeske (1992) and using a first-order expansion of

the MSPE for the estimated parameters $(\hat{\sigma^2}^n, \hat{\gamma}^n)$ around their true value, we obtain as an approximation some additional correcting function term for the MSPE related to the observations collected. Assume for simplicity that σ^2 is known, we then get the approximation

$$\begin{aligned}\widehat{MSPE}_\xi(x, \hat{\gamma}^n) &= MSPE_\xi(x, \sigma^2, \hat{\gamma}^n) \\ &\quad + \mathrm{tr}\left\{ M_\gamma^{-1}(\xi, \hat{\gamma}^n) \frac{\partial v^\top(x,\theta)}{\partial \gamma}\bigg|_{\hat{\gamma}^n} C_n(\hat{\gamma}^n) \frac{\partial v(x,\theta)}{\partial \gamma^\top}\bigg|_{\hat{\gamma}^n} \right\}, \quad (5)\end{aligned}$$

where $M_\gamma(\xi, \gamma)$ is the part of the expected information matrix related to the parameters γ, see (3). (When σ^2 is unknown and estimated on the same dataset, we need to consider the full information matrix $M_{\sigma^2,\gamma}(\xi, \sigma^2, \gamma)$ for parameters σ^2 and γ, and then replace in (5) $M_\gamma^{-1}(\xi, \hat{\gamma}^n)$ by the part of $M_{\sigma^2,\gamma}^{-1}(\xi, \hat{\sigma^2}^n, \hat{\gamma}^n)$ corresponding to γ; a similar modification can be used in (2).) Consequently, Zimmerman (2006) (for a similar criterion see also Zhu and Stein, 2005) regards

$$\min_\xi \max_{x \in \mathcal{X}} \widehat{MSPE}_\xi(x, \gamma) \quad (6)$$

for some nominal γ as the (local) design problem, which he terms EK-(empirical kriging-)optimality. The objective here is to take into account the dual effect of the design (obtaining accurate predictions at unsampled sites and improving the accuracy of the estimation of the covariance parameters, those two objectives being conflicting, see Pronzato, 2008) through the formulation of a single criterion.

3 The proposed "equivalence"

Let us briefly review one of the essential statements of the KWET (Kiefer and Wolfowitz, 1960), which was formulated for the classical linear regression setup with uncorrelated errors: it relates D-optimal designs for estimating the regression coefficients β to so-called G-optimum designs, which minimize the maximum prediction variance, i.e.

$$\min_\xi \max_{x \in \mathcal{X}} \mathrm{Var}[\hat{Y}_t(x)].$$

Those criteria coincide in the space of approximate designs, thus their respective efficiencies for exact designs can expected to be high. It is thus natural that Müller and Stehlík (2009b) formulated the conjecture that it may always be possible to find an α that allows to find designs optimizing $\Phi[\xi|\alpha]$ with high EK-efficiency, thus establishing a relationship in the spirit of the KWET. Note that, using (5) and due to the unbiasedness of the kriging predictor, we can approximate (4) by

$$\min_\xi \max_{x \in \mathcal{X}} \left\{ \mathrm{Var}[\hat{Y}_t(x)] + \mathrm{tr}\left\{ M_\gamma^{-1} \mathrm{Var}[\partial \hat{Y}_t(x)/\partial \gamma] \right\} \right\}, \quad (7)$$

which particularly highlights the affinity with G-optimality. The reasoning is much similar as for the two-stage design suggested by Zhu and Stein (2006), which however confines itself to the covariance parameters and seeks to find a balance not in the criterion itself, but applying two different criteria for two sets of observations. For

more details and the derivation of the above form and variants see Abt (1999); Zimmerman (2006); Zhu and Zhang (2006).

It has been observed that both criteria are seeking to find a compromise between space-filling behavior (i.e. the trend parameter and the kriging variance component respectively) and short distances (i.e. the covariance parameter or the correcting term component respectively). We are thus led to believe that equally good efficiencies can be produced for both setups. This would be very advantageous since EK-optimal designs are much more difficult to generate than parameter estimation designs, since they require embedded optimizations over the candidate sets. A quasi KWET-relationship would thus allow to replace the very demanding optimization (6) by the much less intensive (2) without much loss in efficiency. We will in the next section exemplify for an extremely simple setup, how and to what extent this proposed equivalence could be exploited.

4 Example

As the example setup we will use the Ornstein-Uhlenbeck process on $\mathcal{T}$ =[0,1], which is a special case of (1) with $\eta(x, \beta) = \beta$, i.e. $f(x) \equiv 1$, and $c(x, x'; \gamma) = \exp|x - x'|/\gamma = \rho^{|x-x'|}$, setting $\sigma^2 \equiv 1$ to avoid identifiability problems (see, e.g., Ying, 1991). We will from now on be using the alternate parametrization ρ for ease of interpretation.

For this example, we have analytic results that correspond to the case $\alpha = 0$. Kiseľák and Stehlík (2008) and Zagoraiou and Antognini (2008) proved the optimality of space-filling designs. For the setup with an additional slope parameter Dette et al. (2008) have shown that the points 0 and 1 must be included in the design, that for growing ρ the design tends to a space-filling design and that the efficiencies of space filling designs can be quite high also for small ρ (for small numbers of observations). The similar behavior for designs based on the kriging variance is a widely acknowledged fact and has led to the predominance of space-filling designs for computer simulation experiments (cf. Bursztyn and Steinberg, 2006).

However, contrasting results are known for the case $\alpha = 1$. Here Müller and Stehlík (2009a) show that the optimal designs collapse into one point; see also Zagoraiou and Antognini (2008).

4.1 Two observations fixed

We will start our investigations by fixing $x_1 = 0$ and $x_2 = 1$ and we will be looking for the optimal position x_3^* for the third design point. This is inspired by the findings of Dette et al. (2008) in the linear case and will allow a more comprehensive exposition. We will also in the following for simplicity fix $\rho = \frac{1}{100}$, although similar, albeit perhaps more trivial results can be achieved for other choices of ρ.

The correlation matrix is now given simply by

$$C_3(\rho) = \begin{pmatrix} 1 & \rho & \rho^{|x_3|} \\ \rho & 1 & \rho^{|1-x_3|} \\ \rho^{|x_3|} & \rho^{|1-x_3|} & 1 \end{pmatrix}$$

and the resulting kriging variance as a function of the remaining point x_3 (and the points for prediction x) is displayed in the left panel of Figure 1 with the axis for x_3 in front. It is evident that $x_3 = 0.5$ as expected minimizes the function for all maxima in x. That this is not the case for the corrected EK-criterion (6) can be easily seen from the right panel of Figure 1. Here it is evident, that the minimum is reached for a point close to the endpoints of the region. In fact, the minimizing argument is $x_3^* = 0.934$ (or $x_3^* = 0.066$ respectively), which gives a much lower "corrected", though also much higher kriging variance than the center point. This discrepancy of the two criteria is well documented in Figure 2.

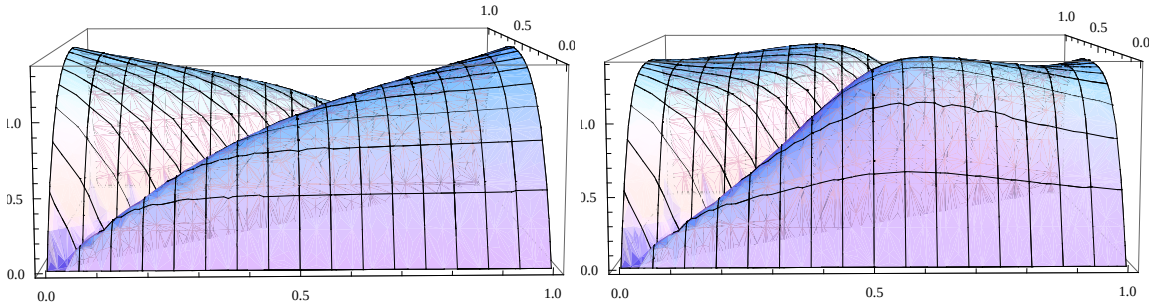

Figure 1: Kriging variance (left panel) and "corrected" kriging variance (right panel) as a function of x_3 (front axis) and x.

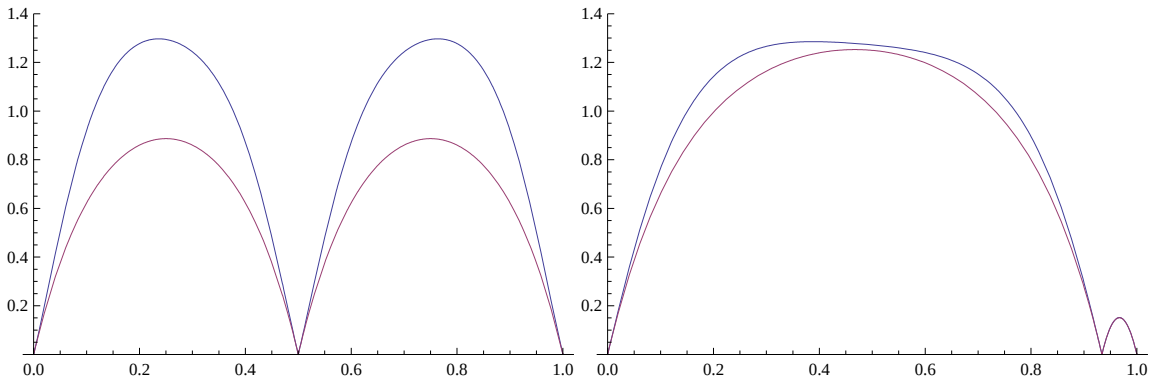

Figure 2: Kriging variance (lower line) and "corrected" kriging variance (upper line) as a function of x for $x_3 = 0.5$ (left panel) and $x_3 = 0.934$ (right panel).

It now remains to be seen, whether the design $(0, 0.934, 1)$ (or at least one with an efficiency close to it) can be achieved employing criterion (2) with a particular choice of α. Its components are of the expected form, the $|M_\beta|$ being concave with a maximum value of 29/11 at $x_3 = 0.5$ and the $|M_\rho|$ being convex with a limit value of 236.765 for $x_3 \to 0$ or $x_3 \to 1$ respectively. We can in fact tune the compound criterion $\Phi[x_3|\alpha]$ in such a way that it gives an optimum at $x_3 = 0.934$ (or 0.066 respectively) by choosing $\alpha = 0.8025$, which can be seen from the enlarged picture in the right panel of Figure 3. Thus for this example we have achieved exact "equivalence" between the extrapolation and the estimation based criterion.

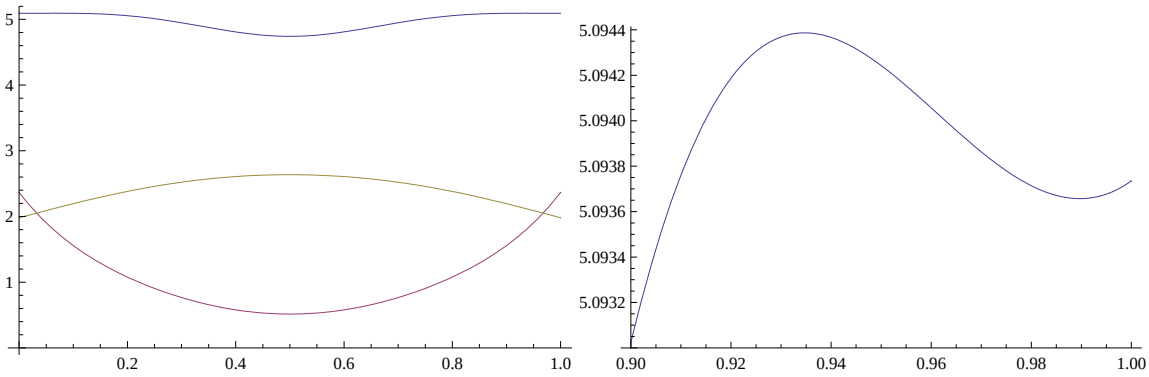

Figure 3: $|M_\beta|$, $|M_\rho|/100$ and $\Phi[x_3|\alpha = 0.8025]$ as functions of x_3 (left panel) and enlarged portion of the graph (right panel).

4.2 One observation fixed

Let us continue the example by lifting the restriction to the endpoints and allowing two of the design points to vary freely. It is natural for reasons of symmetry to then fix one of the three points in the center, i.e. $x_2 = 0.5$ and the other two equally distant from the boundaries, i.e. $x_3 = 1 - x_1$. One can now plot the correcting terms in (7) as a function of x_3 and x and see that the situation differs much from the above, see Figure 4.

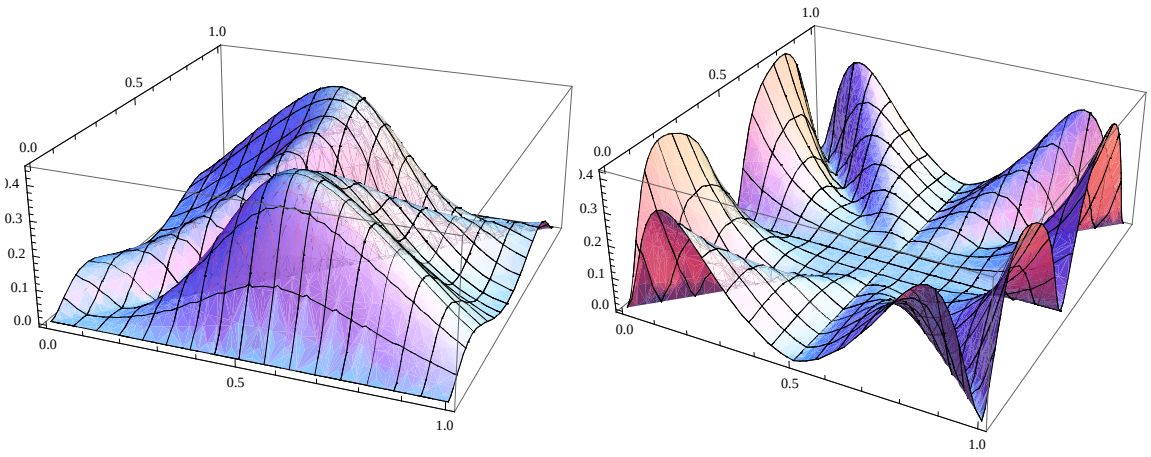

Figure 4: Correcting terms in (7) as functions of x_3 (front axis) and x for $x_1 = 0$ and $x_2 = 1$ (left panel) and $x_1 = 0.5$ and $x_3 = 1 - x_1$ (right panel).

It is indeed so, that although again the optimal designs corresponding to the kriging variance and the EK-criterion differ (for the former we find $x_1^* = 0.120$ and $x_3^* = 0.880$, while for the latter $x_1^* = 0.099$ and $x_3^* = 0.901$), their structure and the form of the respective functions, see Figure 5 are rather similar.

Once more we are looking for an α, which will yield a similar design for the compound criterion as for EK-optimality. The functions $|M_\beta|$ and $|M_\rho|$ are as expected with spikes at $x_1 \to x_2 \to x_3 = 0.5$ and yet again we can choose $\alpha = 0.8477$ to yield exactly $x_1^* = 0.099 = 1 - x_3^*$ (see Figure 6).

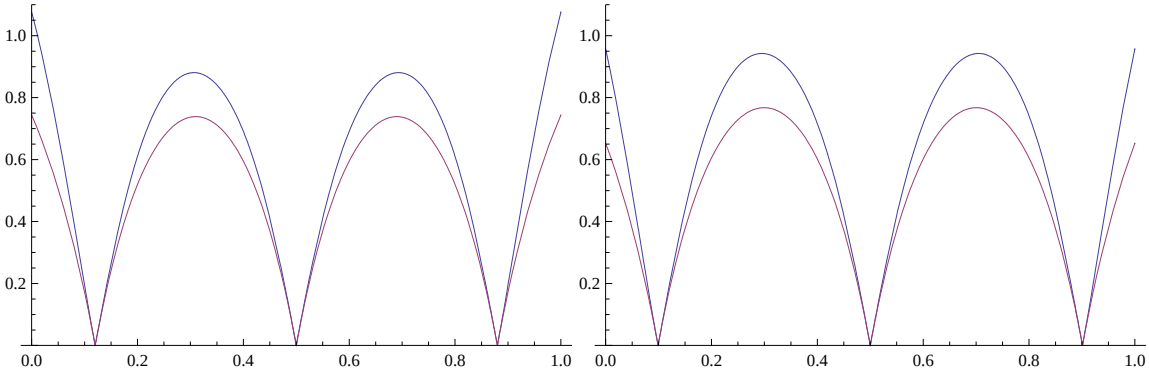

Figure 5: Kriging variance (lower line) and "corrected" kriging variance (upper line) as a function of x for $x_1^* = 0.120 = 1 - x_3^*$ (left panel) and $x_1^* = 0.099 = 1 - x_3^*$ (right panel).

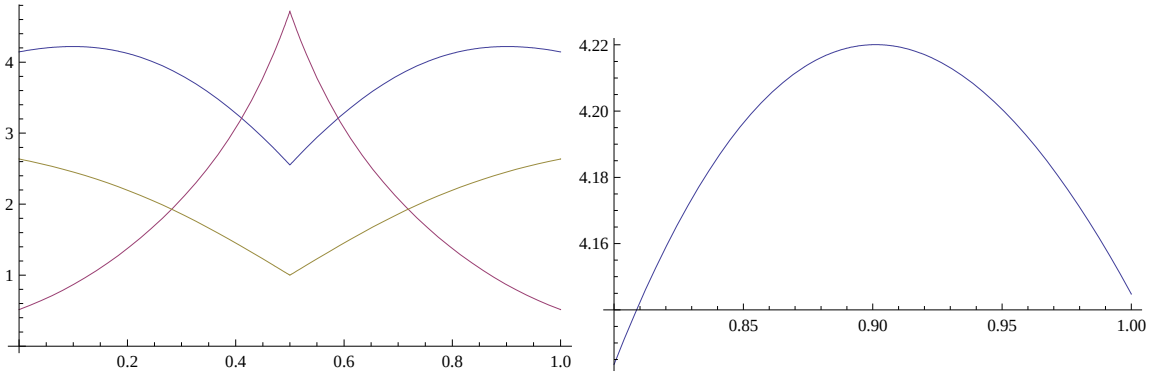

Figure 6: $|M_\beta|$, $|M_\rho|/100$ and $\Phi[x_3|\alpha = 0.8477]$ as functions of $x_3 = 1 - x_1$ (left panel) and enlarged portion of the graph (right panel).

4.3 Three point optimal designs

Let us now finally lift all restrictions and allow the three points to vary freely. A grid search reveals that the EK-optimal design is indeed $\{0.099, 0.5, 0.991\}$. Unfortunately now, the $\Phi[.|\alpha]$-optimal designs are asymmetric for all α, so exact correspondence of the optimal designs cannot be achieved. Thus even in this simple example the conjecture in the strict sense is disproved. However, by comparing the criterion function for the $\alpha = 0.8477$ giving a design identical to EK-optimality and the function for the corresponding true optimum design $\{0, 0.743, 1\}$ (or $\{0, 0.257, 1\}$) reveals that the values are very close (see Figure 7). Thus we can expect local optima (e.g. yielded from an exchange algorithms) to still perform rather well.

The presented examples, albeit of limited scope, give hope that the conjecture of Müller and Stehlík (2009b) of a KWET-type relationship will continue to hold in more complex and realistic settings, perhaps not always exactly. They also show that the EK-optimal designs (and the corresponding $\Phi[.|\alpha]$-optimal designs) can be quite far from the often suggested and frequently employed space-filling designs.

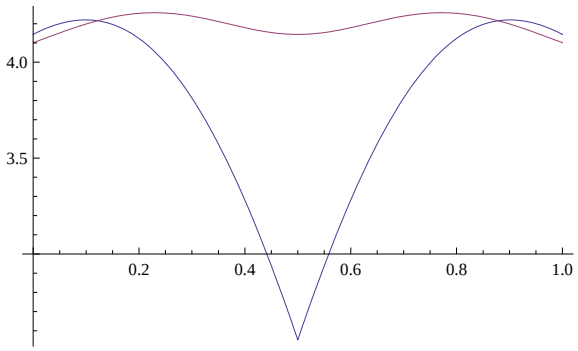

Figure 7: $\Phi[x_3|\alpha = 0.8477]$ as functions of x_3 with $x_1 = 0, x_2 = 1$ (upper line) and $x_3 = 1 - x_1$ with $x_2 = 0.5$ (lower line).

5 Conclusions and Outlook

Suppose therefore that a value α can always be found such that the optimal design for problem (2) is reasonably efficient for (7). It is then much advantageous to solve (2) rather than (7) since the former does not require maximization over the candidate set $\mathcal{X}$. Several approaches have been suggested to solve (2) with $\alpha = 1$, see Müller and Pázman (2003); Fedorov and Müller (2007). It remains to be checked whether such approaches can be used when $\alpha < 1$. The main difficulty here is to choose a suitable α beforehand. Since it is reasonable in most applications to assume that $\mathcal{X}$ is finite, the evaluation of $M\widehat{MSPE}_\xi(\gamma) = \max_{x\in\mathcal{X}} \widehat{MSPE}_\xi(x,\gamma)$ has a moderate computational cost. One can then simply compute optimal designs for (2) for a series of values of α, and retain the best one in terms of $M\widehat{MSPE}_\xi(\gamma)$.

The MSPE (5) can also be used for the sequential construction of designs. Let ξ^{n_0} denote some initial design of size n_0. At step k of such a sequential construction, $k \geq n_0$, choose x_{k+1} as

$$x_{k+1} = \arg\max_{x\in\mathcal{X}} \widehat{MSPE}_{\xi^k}(x,\gamma)$$

and then update ξ^k into $\xi^{k+1} = \{\xi^k, x_{k+1}\}$. Again, when $\mathcal{X}$ is finite, the sequential construction above has a moderate computational cost. If one wishes to minimize the integrated MSPE, $I\widehat{MSPE}_\xi(\gamma) = \int_{\mathcal{X}} \widehat{MSPE}_\xi(x,\gamma)\mu(dx)$, for some measure of interest μ, one can choose instead at step k

$$x_{k+1} = \arg\min_{x\in\mathcal{X}} \int_{\mathcal{X}} \widehat{MSPE}_{\{\xi^k,x\}}(x,\gamma)\mu(dx)\,.$$

The parameters γ can be estimated after each generation of a new sampling point x_{k+1}, rending the sequential designs above adaptive. Algorithms for the construction of adaptive designs for (2) are also of interest. We believe that such investigations could yield to the development of cheap algorithms for the sequential construction of designs that would take into account the prediction task and at the same time the reduction of uncertainty in the estimation of the covariance parameters, thereby following the same ultimate objective as designs optimal in the sense of (7). The fact that in most applications $\mathcal{X}$ is finite might reveal particularly useful for studying

the convergence properties of such adaptive procedures, see Pronzato (2009) for such developments in the case of uncorrelated errors.

Contrasting with the uncorrelated case, non-additivity and nonconvexity are amongst the obstacles for constructing optimal designs for random fields, which have recently been reviewed by Müller and Stehlík (2009a). Furthermore the concept of Fisher information is conveniently used as a basis for designing efficient experiments. However, if the output stems from correlated random fields as (1), the conditions under which Fisher information may be suitable must be restated. For some small sample results see also Ginsbourger et al. (2009).

A last point that we wish to investigate concerns the estimability of the random-field parameters σ^2 and γ. Under the infill design framework (i.e., when the design space is compact) typically not all parameters are estimable, only some of them being micro-ergodic, see Stein (1999). However, it seems reasonable to consider that Jeffrey's law will apply and that parameters that are not estimable from the data $Y_t(\xi)$ should have little influence on predictions (interpolations/extrapolations) for the random filed. This requires theoretical investigations as well as numerical confirmation for small to moderate sample sizes.

Acknnowledgements

Werner G. Müller wants to acknowledge the generous support by the University of Nice-Sophia Antipolis.

References

Abt, M. (1999). Estimating the prediction mean squared error in gaussian stochastic processes with exponential correlation structure. *Scandinavian Journal of Statistics 26*(4), 563–578.

Bursztyn, D. and D. Steinberg (2006, March). Comparison of designs for computer experiments. *Journal of Statistical Planning and Inference 136*(3), 1103–1119.

Cressie, N. A. C. (1993, January). *Statistics for Spatial Data (Wiley Series in Probability and Statistics)* (Rev Sub ed.). Wiley-Interscience.

Dette, H., J. Kunert, and A. Pepelyshev (2008). Exact optimal designs for weighted least squares analysis with correlated errors. *Statistica Sinica 18*(1), 135–154.

Fedorov, V. and W. Müller (2007, June). Optimum design for correlated processes via eigenfunction expansions. In J. López-Fidalgo, J. Rodríguez-Díaz, and B. Torsney (Eds.), *mODa'8 – Advances in Model–Oriented Design and Analysis, Proceedings of the 8th Int. Workshop, Almagro (Spain)*, pp. 57–66. Heidelberg: Physica Verlag.

Ginsbourger, D., D. Dupuy, A. Badea, L. Carraro, and O. Roustant (2009). A note on the choice and the estimation of kriging models for the analysis of deterministic computer experiments. *Applied Stochastic Models in Business and Industry 25*(2), 115–131.

Harman, R. and F. Štulajter (2009). Optimal prediction designs in finite discrete spectrum linear regression models. *Metrika*.

Harville, D. A. and D. R. Jeske (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association 87*(419), 724–731.

Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics 12*, 363–366.

Kiseľák, J. and M. Stehlík (2008, September). Equidistant and D-optimal designs for parameters of Ornstein-Uhlenbeck process. *Statistics & Probability Letters 78*(12), 1388–1396.

Müller, W. and A. Pázman (2003). Measures for designs in experiments with correlated errors. *Biometrika 90*(2), 423–434.

Müller, W. and M. Stehlík (2009a). Issues in the optimal design of computer simulation experiments. *Applied Stochastic Models in Business and Industry 25*(2), 163–177.

Müller, W. G. and M. Stehlík (2009b). Compound optimal spatial designs. *Environmetrics (online)*.

Pronzato, L. (2008, February). Optimal experimental design and some related control problems. *Automatica 44*(2), 303–325.

Pronzato, L. (2009). One-step ahead adaptive D-optimal design on a finite design space is asymptotically optimal. *Metrika*.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989, November). Design and analysis of computer experiments. *Statistical Science 4*(4), 409–423.

Stein, M. L. (1999, June). *Interpolation of Spatial Data: Some Theory for Kriging (Springer Series in Statistics)* (1 ed.). Springer.

Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis 36*, 280–296.

Zagoraiou, M. and A. B. Antognini (2008). Optimal designs for parameter estimation of the Ornstein-Uhlenbeck process. *Applied Stochastic Models in Business and Industry* (online), n/a+.

Zhu, Z. and M. Stein (2005, October). Spatial sampling design for parameter estimation of the covariance function. *Journal of Statistical Planning and Inference 134*(2), 583–603.

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

Zhu, Z. and H. Zhang (2006). Spatial sampling design under the infill asymptotic framework. *Environmetrics 17*(4), 323–337.

Zimmerman, D. L. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics 17*(6), 635–652.