

Optimum Design for Correlated Processes via Eigenfunction Expansions



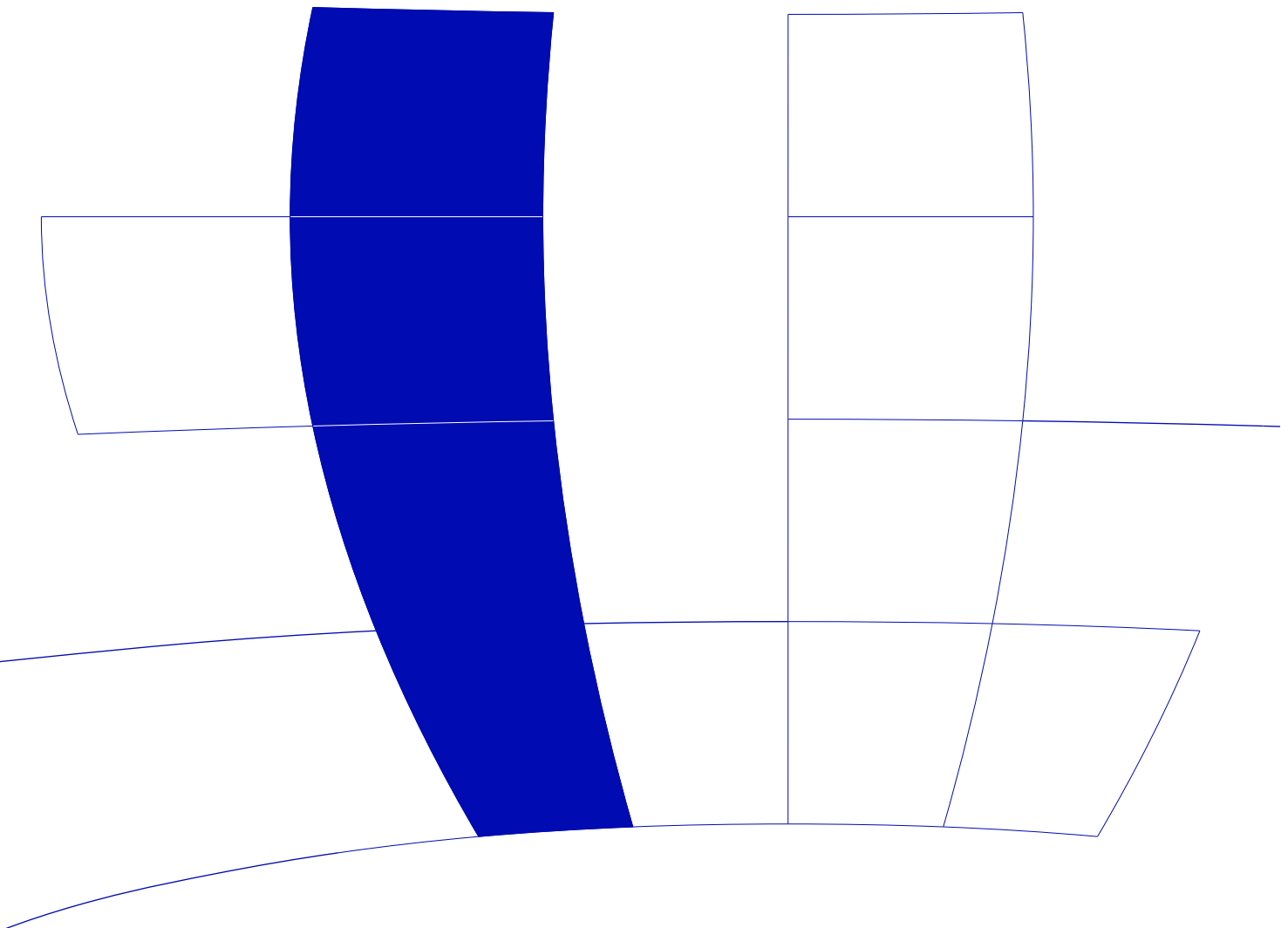
Valerii V. Fedorov, Werner G. Müller

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 6
June 2004

<http://statistik.wu-wien.ac.at/>



Optimum Design for Correlated Processes via Eigenfunction Expansions

VALERII V. FEDOROV * WERNER G. MÜLLER†

Abstract

In this paper we consider optimum design of experiments for correlated observations. We approximate the error component of the process by an eigenvector expansion of the corresponding covariance function. Furthermore we study the limit behavior of an additional white noise as a regularization tool. The approach is illustrated by some typical examples.

Key words: Correlated errors; Random field; Regression experiment.

1 Introduction

Let \mathfrak{X} denote the design space, corresponding to a finite set of potential trials. We can observe a random process

$$y(x) = \eta(x, \beta) + \varepsilon(x), \quad (1)$$

where $\eta(x, \beta)$ is the response function at $x \in \mathfrak{X}$ containing q unknown parameters $\beta = (\beta_1, \dots, \beta_q)^T \in \mathbb{R}^q$. Let us further assume that the random noise $\varepsilon(x)$ consists of two independent components

$$\varepsilon(x) = u(x) + e(x),$$

such that for both $E[u] = E[e] = 0$ and thus $E[\varepsilon] = 0$, and that

$$\text{Cov}[e(x), e(x')] = \sigma^2(x)\delta_{x,x'},$$

and

$$\text{Cov}[u(x), u(x')] = k(x, x'),$$

respectively, the latter - the so-called covariance kernel - being a known function; $\delta_{x,x'}$ denotes the Kronecker-symbol. Hence $\text{Cov}[\varepsilon(x), \varepsilon(x')] = \sigma^2(x)\delta_{x,x'} + k(x, x')$, where $x, x' \in \mathfrak{X}$. This setup has many potential applications, but it is extremely relevant in environmental studies, which exhibit, for instance spatial or temporal data or both (cf. Müller (2001)).

We will in the following approximate the correlated component $u(x)$ by Mercer's eigenfunction expansion (cf. Mercer (1909)) of the respective kernel $k(x, x')$ thereby allowing to embed the problem into standard convex design theory. Suggestions according to that have already been made in Fedorov (1996) and Fedorov and Flanagan (1997), but the approach has never been fully worked out, implemented and tested. Furthermore we will relate our results to alternative methods for the limit cases $\sigma^2 \rightarrow \infty$ (independence) and $\sigma^2 \rightarrow 0$ (random fields).

*GlaxoSmithKline, King of Prussia, PA 19406, USA

†Department of Statistics, University of Economics and Business Administration, A-1090 Vienna, Austria.

2 Expansion of the covariance kernel

The approach is based on the fact that the error component $u(x)$ in the random process can be represented by the infinite expansion

$$u(x) = \sum_{l=1}^{\infty} \gamma_l \varphi_l(x),$$

and correspondingly $k(x, x') = \sum_{l=1}^{\infty} \gamma_l \varphi_l(x) \varphi_l(x')$ where the γ_l are specific random values with $E[\gamma_l] = 0$ and $\text{Cov}[\gamma_l, \gamma_{l'}] = \lambda_l \delta_{l,l'} = \Lambda_{l,l'}$, and the $\varphi_l(x)$ and λ_l are the eigenfunctions and eigenvalues, respectively, of the covariance kernel $k(x, x')$.

This allows us to approximate the random process (1) by

$$y(x) = \eta(x, \beta) + \sum_{l=1}^p \gamma_l \varphi_l(x) + e(x), \quad (2)$$

where the choice of a suitable p depends upon the range of η and the remainder is subsumed in the variation of the noise. This presentation can now be regarded as a special form of a random coefficient regression model as long as it admits replications (i.e. for $\sigma^2(x) > 0$), where $\theta = \{\beta_1, \dots, \beta_q, \gamma_1, \dots, \gamma_p\}$ with

$$E[\theta] = \bar{\theta} = \{\beta, 0_p\},$$

and

$$\text{Cov}[\theta, \theta'] = \begin{pmatrix} 0_{q \times q} & 0_{q \times p} \\ 0_{p \times q} & \Lambda_{p \times p} \end{pmatrix}.$$

The covariance of the total error $\varepsilon(x)$ can thus be written as $\text{Cov}[\varepsilon(x), \varepsilon(x')] = \sigma^2(x) \delta_{x,x'} + \varphi^T(x) \Lambda \varphi(x')$, where $\varphi^T(x) = \{\varphi_1(x), \dots, \varphi_p(x)\}$. Replications can be understood as observations of the corresponding number of closely allocated sensors, meters, etc..

It is well known (cf.eg. Gladitz and Pilz (1982)) that random coefficient models can be embedded into the standard convex design theory, thus allowing the use of powerful design tools such as equivalence theorems and gradient algorithms. There remains the issue, however, which design problem is relevant for a specific situation and to resolve it there seem to be at least three plausible options:

- a) the main interest of the researcher is in the “trend” parameters β and the component $u(x)$ is regarded solely as a nuisance,
- b) the emphasis is on the prediction of individual instances of the process, thus the information on γ_l must also enter a design criterion, or perhaps
- c) one desires prediction of the average process, i.e. $\eta(x, \beta)$ at a given set of $x \in \mathcal{Z} \neq \mathfrak{X}$.

In this paper we will, for clarity of exposition, consider in detail only case a), but treatment of the other cases is similar. Furthermore, to eventually avoid the arbitrary choice of the order of approximation p , we will want to reformulate results in terms of the original covariance kernel $k(x, x')$ rather than its approximation.

3 Design for the estimation of trend

For simplification we now and in the following consider a linearisation $f(x) = \partial \eta(x, \beta) / \partial \beta|_{\beta=\beta_0}$ of the response around a prior guess of the parameter β_0 , eventually leading to so-called locally optimum designs.

Let us further assume that we observe the random field $y(x)$ at n distinct points x_1, \dots, x_n , that is observations are generated according to

$$y_{ij} = \beta^T f(x_i) + \sum_{l=1}^p \gamma_{lj} \varphi_l(x_i) + e_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m,$$

with r_1, \dots, r_n repeated measurements respectively. We assume that for all j the collection $\{x_i, \frac{r_i}{R}\}_1^n = \xi_n$, with $R = \sum_{i=1}^n r_i$, is the same. Actually, we assume that $m = 1$, a reader can add a multiplier m^{-1} where needed. Also we then assume for the sake of simplicity $\sigma^2(x) = \sigma^2$. Thus from now on we admit the possibility of repeated observations, which will reflect microscale variations. For instance, in the case of designing a spatial network, these replications can stem from observations from very closely neighboring measurement sites.

In this setup the best linear unbiased estimator of the trend parameter β is

$$\hat{\beta} = \{FV^{-1}F^T\}^{-1}FV^{-1}\bar{y},$$

with

$$F = \{f(x_1), \dots, f(x_n)\},$$

$$\Phi = \{\varphi(x_1), \dots, \varphi(x_n)\},$$

$$V = \sigma^2\Omega + \Phi^T\Lambda\Phi, \text{ and}$$

$$\bar{y}^T = (\bar{y}_1, \dots, \bar{y}_n), \bar{y}_i = \frac{1}{r_i} \sum_{l=1}^{r_i} y_{il}, \Omega_{ii'} = \frac{\delta_{ii'}}{r_i}.$$

The asymptotic covariance matrix of the estimator $\hat{\beta}$ is proportional to $M^{-1}(\xi_n) = m^{-1}\{FV^{-1}F^T\}^{-1}$ and thus an optimum design ξ_n^* must seek to satisfy

$$\xi_n^* = \arg \min_{\xi_n} \Psi\{M(\xi_n)\}, \quad (3)$$

for a reasonably chosen design criterion.

Let us now consider the best linear unbiased estimator $\hat{\theta}$ in the full (random coefficient) model. Its covariance matrix is

$$D(\xi_n) = \begin{pmatrix} D_{ff} & D_{f\varphi} \\ D_{f\varphi}^T & D_{\varphi\varphi} \end{pmatrix} = \begin{pmatrix} FWF^T & FW\Phi^T \\ \Phi WF^T & \Phi W\Phi^T + \Lambda^{-1} \end{pmatrix}^{-1},$$

where $W_{ii'} = \frac{R}{\sigma^2} \delta_{ii'} \xi(x_i)$, ξ now representing a design measure with support $\{x_1, \dots, x_n\}$ corresponding to ξ_n .

It is easy to show (Frobenius formula) that

$$D_{ff}(\xi) \propto M^{-1}(\xi_n)$$

and thus the criterion (3) is equivalent to subset D -optimality (D_s -optimality) in a random coefficient regression model. It now follows directly from standard design theory (cf. Theorem 2.7.1 in Fedorov (1972)), that a necessary and sufficient condition for a design ξ^* to optimize (3) is

$$\begin{aligned} \phi(x, \xi^*) &= \phi_\theta(x, \xi^*) - \varphi^T(x) [\Lambda^{-1} + \Phi W \Phi^T]^{-1} \varphi(x) \\ &\leq \text{tr} D(\xi^*) M(\xi^*) - \text{tr} [\Lambda^{-1} + \Phi W \Phi^T]^{-1} \Phi W \Phi^T, \end{aligned} \quad (4)$$

with $\phi_\theta(x, \xi^*) = (f(x), \varphi(x))^T D(\xi^*) (f(x), \varphi(x))$.

Note that this equivalence condition effectively reflects the design problem discussed in section 3 of Wynn (2004), which embeds it further into the maximum entropy framework. It is essential for the development of numerical algorithms and analysis of their properties. We have used the first order exchange algorithm, which at its s iteration adds some mass to point

$$x_s^+ = \arg \max_{x \in \mathcal{X}} \phi(x, \xi_s)$$

and subtracts some at point

$$x_s^- = \arg \min_{x \in \text{supp} \xi_s} \phi(x, \xi_s).$$

Details of the algorithm are beyond the scope of this brief paper.

The so-called sensitivity function $\phi(x, \xi)$ can thus be standardly employed in such a routine design optimization algorithm. However, since the original problem is formulated in terms of the covariance kernel one will prefer the presentation

$$\phi(x, \xi) = \mathfrak{f}^T(x, \xi) D_{ff}(\xi) \mathfrak{f}(x, \xi) \quad (5)$$

with

$$\mathfrak{f}^T(x, \xi) = f^T(x) + k^T(x, \xi) K^{-1}(\xi) (W + K^{-1}(\xi))^{-1} W F^T,$$

$$D_{ff}(\xi) = \left\{ F \left[W - W (W + K^{-1}(\xi))^{-1} W \right] F^T \right\}^{-1},$$

$$k^T(x, \xi) = \{k(x, x_1), \dots, k(x, x_n)\}, \text{ and}$$

$$[K(\xi)]_{ij} = k(x_i, x_j) \text{ for all } x_i, x_j \in \text{supp}\xi.$$

Here we use the fact that for sufficiently large p and after neglecting remainder terms one can replace $K \simeq \Phi^T \Lambda \Phi$ (cf. Fedorov and Flanagan (1997)). The advantage of this presentation is, that a practitioner solely has to specify the response function and the covariance kernel and needs not to bother with the implicit eigenvalue expansion. A derivation of this presentation can be found in the Appendix.

4 Examples

To illustrate the above presented technique we have performed a standard one-point correction design algorithm based on the sensitivity function (4) on two typical examples. Calculations were performed on a 31 design point grid with a uniform measure as the initial design and stopped after 10000 iterations. Asymmetries are due to numerical inaccuracies and the finite number of iterations, slight possible shifts are due to the finiteness of the grid. As a criterion we used D -optimality, where $\Psi\{M\} = -\log \det M$.

4.1 N  ther's Case

Firstly consider a simple linear regression model, i.e. $f^T(x_i) = (1, x_i)$, on $\mathfrak{X} = [-1, 1]$ with an error covariance kernel given by

$$k(x, x') = \begin{cases} 1 - |x - x'| & \text{for } |x - x'| < 1 \\ 0 & \text{for } |x - x'| \geq 1. \end{cases}$$

This example gained some prominence in the design literature for correlated errors, since it linearly relates the response function to the covariance kernel, which allows for a direct proof of uniform optimality of a three point design concentrated on $\{-1, 0, 1\}$, see N  ther (1985). Moreover note, that it served as a motivating case for the considerations in M  ller and P  zman (2003), where also regulatory noise was employed, but much differently than in the present paper made dependent upon the design measure itself.

It is thus easy to guess what a reasonable design algorithm should yield for the extreme settings large σ (independence; $\sigma = 10^{10}$ was used for the computations) and small σ (dependence due to k ; here $\sigma = 10^{-10}$). In the former case we expectedly yield the design measure equally distributed between the extremal points -1 and 1 , whereas in the latter case we yield the measure displayed in Figure 1, which highly corresponds to the computational results of M  ller and P  zman (2003) displayed in their Fig.1. In all our figures the dashed line represents a rescaled sensitivity function, the solid line the design measure. Note that for intermediate choices of σ a respective share of the central weight is distributed to the extremal points.

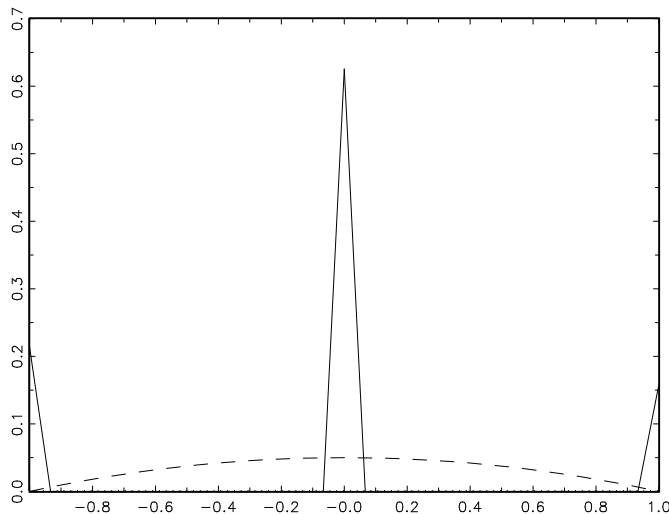


Figure 1: Design measure (solid line) and rescaled sensitivity function (dashed line) for N  ther's case on a 31-point grid; horizontal x , vertical $\xi(x)$.

4.2 Third Order Polynomial and Poisson Kernel

Next, assume the response function to be a third order polynomial, i.e. $f^T(x_i) = (1, x_i, x_i^2, x_i^3)$ and the error covariance structure is described by the Poisson kernel

$$k(x, x') = \frac{1 - \zeta^2}{1 - 2\zeta \cos \pi(x - x') + \zeta^2},$$

where $0 \leq \zeta \leq 1$ is a shape parameter and the design region be $\mathfrak{X} = [0, 1]$.

The optimum design in the uncorrelated case follows from Theorem 2.3.3. in Fedorov (1972) as the roots of the polynomial

$$[1 - (2x - 1)^2][3(2x - 1)^2 - 1],$$

which yields the points 0, 0.211, 0.789, 1. This is also the result that our numerical algorithm yields (approximately) for a large σ (here $\sigma = 10^{10}$), see Figure 2.

The situation is very different for the cases letting σ decrease, i.e. approaching the 'purely' correlated case. In the beginning the two extremal points move inwards (and later the extremal points move slightly outwards to merge). The measure around these two points is now much more spread out (see Figure 3 for $\sigma = 10^{-2}$). Further decreasing σ again shifts the points toward the extremes, to eventually (see Figure 4 for $\sigma = 10^{-30}$) end up with a design that (except for the two extremal points) distributes the measure quite evenly over the whole region.

Appendix

For the derivation of ϕ_θ we require the entries D_{ff} , $D_{f\varphi}$ and $D_{\varphi\varphi}$. For brevity we omit arguments. We have

$$D_{ff} = \left\{ FWF^T - FW\Phi^T (\Phi W\Phi^T + \Lambda^{-1})^{-1} \Phi WF^T \right\}^{-1} \text{ and hence}$$

$$\begin{aligned} D_{ff}^{-1} &= FWF^T - FW\Phi^T (\Phi W\Phi^T + \Lambda^{-1})^{-1} \Phi WF^T \\ &\simeq FWF^T - FW\Phi^T (\Lambda - \Lambda\Phi(W^{-1} + K)^{-1}\Phi^T\Lambda) \Phi WF^T \\ &\simeq F [W - W(W + K^{-1})^{-1}] F^T. \end{aligned}$$

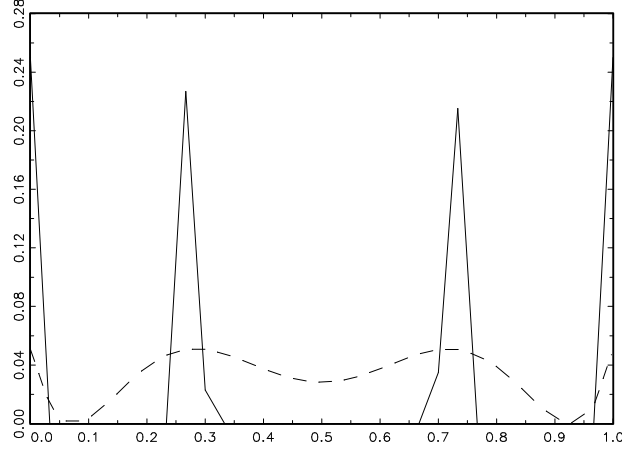


Figure 2: Design measure (solid line) and rescaled sensitivity function (dashed line) for third order polynomial (under independence) on a 31-point grid; hor. x , ver. ξ .

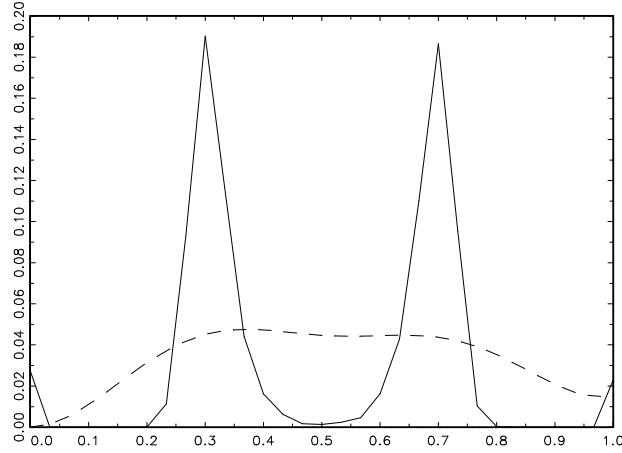


Figure 3: Design measure (solid line) and rescaled sensitivity function (dashed line) for third order polynomial ($\sigma = 10^{-2}$) on a 31-point grid; horizontal x , vertical $\xi(x)$.

Note that $K \simeq \Phi^T \Lambda \Phi$. Also we have

$$D_{f\varphi} = -D_{ff} F W \Phi (\Phi W \Phi^T)^{-1} \text{ and}$$

$$D_{\varphi\varphi} = (\Phi W \Phi^T + \Lambda^{-1})^{-1} + (\Phi W \Phi^T + \Lambda^{-1})^{-1} \Phi W F D_{ff} F W \Phi (\Phi W \Phi^T + \Lambda^{-1})^{-1}.$$

Now we can write

$$\begin{aligned} \phi_\theta(x, \xi) &= (f(x), \varphi(x))^T \begin{pmatrix} D_{ff} & D_{f\varphi} \\ D_{f\varphi}^T & D_{\varphi\varphi} \end{pmatrix} (f(x), \varphi(x)) \\ &= f^T(x) D_{ff} f(x) + 2f^T(x) D_{f\varphi} \varphi(x) + \varphi^T(x) D_{\varphi\varphi} \varphi(x) \\ &= f^T(x) D_{ff} f(x) - 2f^T(x) D_{ff} F W \Phi (\Phi W \Phi^T)^{-1} \varphi(x) \\ &\quad + \varphi^T(x) (\Phi W \Phi^T + \Lambda^{-1})^{-1} \varphi(x) \\ &\quad + \varphi^T(x) (\Phi W \Phi^T + \Lambda^{-1})^{-1} \Phi W F D_{ff} F W \Phi (\Phi W \Phi^T + \Lambda^{-1})^{-1} \varphi(x). \end{aligned}$$

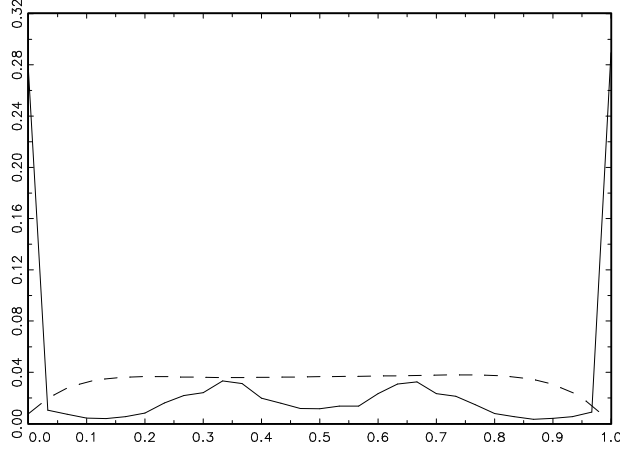


Figure 4: Design measure (solid line) and rescaled sensitivity function (dashed line) for third order polynomial ($\sigma = 10^{-30}$) on a 31-point grid; horizontal x , vertical $\xi(x)$.

Note that the one but last summand will be subtracted in the definition of the sensitivity function so that we can subsume

$$\begin{aligned}
 \phi(x, \xi) &= \left(f^T(x) - \varphi^T(x) (\Phi W \Phi^T + \Lambda^{-1})^{-1} \Phi W F \right) D_{ff} \times \\
 &\quad \times \left(f^T(x) - \varphi^T(x) (\Phi W \Phi^T + \Lambda^{-1})^{-1} \Phi W F \right)^T \\
 &= \mathfrak{f}^T(x) D_{ff} \mathfrak{f}(x).
 \end{aligned}$$

So, it remains to revert $\mathfrak{f}(x)$ to a presentation in terms of the original covariance function. For that purpose we require the presentation

$$(\Phi W \Phi^T + \Lambda^{-1})^{-1} \simeq \Lambda - \Lambda \Phi (K + W^{-1})^{-1} \Phi^T \Lambda$$

Then

$$\begin{aligned}
 \mathfrak{f}^T(x) &= f^T(x) - \varphi^T(x) (\Phi W \Phi^T + \Lambda^{-1})^{-1} \Phi W F^T \\
 &\simeq f^T(x) - \varphi^T(x) \left(\Lambda - \Lambda \Phi (K + W^{-1})^{-1} \Phi^T \Lambda \right) \Phi W F^T \\
 &\simeq f^T(x) - \left(k^T(x) - k^T(x) (K + W^{-1})^{-1} K \right) W F^T,
 \end{aligned}$$

using the convenient notation $k^T(x) \simeq \varphi^T(x) \Lambda \Phi$.

And from

$$k^T(x) - k^T(x) (K + W^{-1})^{-1} K = k^T(x) K^{-1} (W + K^{-1})^{-1}$$

easily follows the sensitivity function given in (4).

References

- FEDOROV, V.V. (1972). *Theory of Optimal Experiments*, Academic Press.
- FEDOROV, V.V. (1996). Design of spatial experiments: model fitting and prediction. In Gosh, S. and Rao, C.R., editors, *Handbook of Statistics*, volume 13. Elsevier, Amsterdam, 515–553.
- FEDOROV, V.V. AND FLANAGAN, D. (1997). Optimal monitoring network design based on Mercer’s expansion of covariance kernel. *Journal of Combinatorics, Information & System Sciences*, 23:237–250.
- GLADITZ, J. AND PILZ, J. (1982). Construction of optimal designs in random coefficient models. *Mathematische Operationsforschung und Statistik, Series Statistics*, 13(3):371–385.
- MERCER, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions Royal Society London*, **A 209**, 415–446.
- MÜLLER, W. G. (2001). *Collecting Spatial Data*. Physica Verlag: Heidelberg.
- MÜLLER, W. G. AND PÄZMAN, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika* **90**,2, 423–434.
- NÄTHER, W. (1985). Exact designs for regression models with correlated errors. *Statistics* **16**, 479–84.
- WYNN, H.P. (2004). Maximum entropy sampling and general equivalence theory. In Di Bucchianico, A., Läuter, H. and Wynn, H.P., editors, *mODa7 - Advances in Model-Oriented Design and Analysis*, Physica, Heidelberg, 211–218.