



# Approximation of IMSE-optimal designs via quadrature rules and spectral decomposition

Bertrand Gauthier, Luc Pronzato

► **To cite this version:**

Bertrand Gauthier, Luc Pronzato. Approximation of IMSE-optimal designs via quadrature rules and spectral decomposition. Communications in Statistics - Simulation and Computation, Taylor & Francis, 2016, 7th International Workshop on Simulation, 45 (5), pp.1600-1612. <<http://www.tandfonline.com>>. <hal-00936681v3>

**HAL Id: hal-00936681**

**<https://hal.archives-ouvertes.fr/hal-00936681v3>**

Submitted on 12 May 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Approximation of IMSE-optimal designs via quadrature rule and spectral decomposition\*

Bertrand GAUTHIER<sup>†‡</sup>

Luc PRONZATO<sup>§‡</sup>

May 12, 2015

## Abstract

We address the problem of computing IMSE (Integrated Mean-Squared Error) optimal designs for random fields interpolation with known mean and covariance. We both consider the IMSE and truncated-IMSE (approximation of the IMSE by spectral truncation). We assume that the MSE is integrated through a discrete measure and restrict the design space to the support of the considered measure. The IMSE and truncated-IMSE of such designs can be easily evaluated at the cost of some simple preliminary computations, making global optimization affordable. Numerical experiments are carried out and illustrate the interest of the considered approach for the approximation of IMSE optimal designs.

**Keywords:** random field models, optimal design of experiments, IMSE, quadrature approximation, simulated annealing, spectral approximation.

## 1 Introduction

This work addresses the problem of designing IMSE (Integrated Mean-Squared Error) optimal experiments in the context of Gaussian random field models with known covariance, see, e.g., Sacks et al. (1989), Rasmussen and Williams (2006). For the sake of simplicity, we assume that the mean structure of the random field is known (we consider centered random fields).

The determination of IMSE-optimal designs for kernel-based models is known as a numerically expensive problem, see, e.g., Fang et al. (2010, Chapter 2) or Santner et al. (2003, Chapter 6). Indeed, in its standard form (see equation (2.3)), any evaluation of the IMSE criterion is computationally demanding and, moreover, the search for IMSE optimal designs is often complicated by the presence of local minima.

Here, we assume that the IMSE is computed for a discrete measure (in particular, it may correspond to a quadrature approximation), and we restrict the design space to a finite subset of points inside the exploration domain (for the sake of simplicity, we consider the set of all quadrature

---

<sup>†</sup>bgauthie@i3s.unice.fr (corresponding author)

<sup>§</sup>pronzato@i3s.unice.fr

<sup>‡</sup>Laboratoire I3S - UMR 7271 Université de Nice-Sophia Antipolis/CNRS.

\*This work was partly supported by the ANR project 2011-IS01-001-01 DESIRE (DESIGns for spatial Random fields), joint with the Statistics Department of the JKU Universität, Linz, Austria. It is based on results that have been presented at the 7th Int. Workshop on Simulation, Rimini, Italy, 21-25 May 2013, see [http://amsacta.unibo.it/3677/1/Quaderni\\_2013\\_3\\_Matteucci\\_Seventh.pdf](http://amsacta.unibo.it/3677/1/Quaderni_2013_3_Matteucci_Seventh.pdf)

points, and we refer to such designs as quadrature-designs). We also consider the truncated-IMSE criterion, which consists in an approximation of the IMSE criterion obtained by spectral truncation. In this framework, we illustrate how preliminary calculations (made once for all, before the design optimization) allow for significantly reducing the computational cost of any evaluation of the IMSE or truncated-IMSE criterion.

Since the IMSE and truncated-IMSE can be easily evaluated for any quadrature-design, it is possible to use optimization strategies that require large number of evaluations of the considered criterion. We then describe and experiment a combinatorial optimization algorithm which combined local improvement (via discrete gradient descents) and global exploration (via a simulated annealing based strategy).

The numerical experiments in particular point out the interest of considering the truncated criterion with truncation level of the order of the design size (just slightly larger). Indeed, for such truncation level, the truncated criterion appears as relatively easy to optimize while ensuring a good efficiency, in terms of IMSE, of the resulting designs. Spectral truncation may also appear, to a lesser extent, as a way to compensate the error induced by the quadrature approximation of the IMSE.

## 2 General framework and notations

### 2.1 Random fields and related Hilbert structures

Let  $\mathcal{X}$  be a general set. We consider a real random field  $(Z_x)_{x \in \mathcal{X}}$  indexed by  $\mathcal{X}$ ; in what follows  $Z$  will refer to the random field  $(Z_x)_{x \in \mathcal{X}}$ . We assume that  $Z$  is centered, second-order, and defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . For the sake of simplicity, we also assume that  $Z$  is Gaussian. We denote by  $L^2(\Omega, \mathbb{P})$  the Hilbert space of second-order real random variables (r.v.) on  $(\Omega, \mathcal{F}, \mathbb{P})$ , where we identify random variables that are equal  $\mathbb{P}$ -almost surely. The inner product between two r.v.  $U$  and  $V$  of  $L^2(\Omega, \mathbb{P})$  is denoted by  $\mathbb{E}(UV)$ .

Let  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  be the covariance kernel of  $Z$ , i.e., for all  $x$  and  $y \in \mathcal{X}$ ,

$$\mathbb{E}(Z_x Z_y) = K(x, y).$$

For  $t \in \mathcal{X}$ , we denote by  $K_t$  the function  $x \mapsto K(t, x)$ ,  $x \in \mathcal{X}$ .

Let  $\mathbf{H}$  be the Gaussian Hilbert space associated with  $Z$ , i.e., the closed linear subspace of  $L^2(\Omega, \mathbb{P})$  spanned by the r.v.  $Z_x$ ,  $x \in \mathcal{X}$ , endowed with the Hilbert structure induced by  $L^2(\Omega, \mathbb{P})$ . We assume that  $\mathbf{H}$  is separable. Notice that we do not suppose that  $Z$  is stationary.

### 2.2 Conditioning

Let  $\mathbf{H}_C$  be a closed linear subspace of  $\mathbf{H}$ ; we denote by  $P_{\mathbf{H}_C}$  the orthogonal projection of  $\mathbf{H}$  onto  $\mathbf{H}_C$ . For  $x \in \mathcal{X}$ , the r.v.  $P_{\mathbf{H}_C}[Z_x]$  is the conditional mean of  $Z_x$  relatively to  $\mathbf{H}_C$ . If  $\mathbf{H}_C$  is spanned by the r.v.  $\zeta_j$ ,  $j \in J$ , with  $J$  a general index set, the notation  $P_{\mathbf{H}_C}[Z_x] = \mathbb{E}(Z_x | \zeta_j, j \in J)$  is often used. We shall pay particular attention to subspaces of the evaluation-type, i.e.

$$\mathbf{H}_{ev} = \text{span}\{Z_{x_1}, \dots, Z_{x_n}\}, \tag{2.1}$$

for some  $n \in \mathbb{N}^*$  (the set of positive integers) and  $x_1, \dots, x_n \in \mathcal{X}$ . Such a set  $\{x_1, \dots, x_n\}$  of locations where the values of the random field  $Z$  are observed forms a  $n$ -point design.

### 2.3 The IMSE criterion

Suppose that  $\mathcal{X}$  is a measurable space and consider a  $\sigma$ -finite measure  $\mu$  on  $\mathcal{X}$ . We denote by  $L^2(\mathcal{X}, \mu)$  the Hilbert space of square integrable real-valued functions on  $\mathcal{X}$  with respect to  $\mu$ . We assume that, for all  $x \in \mathcal{X}$ ,  $K_x \in L^2(\mathcal{X}, \mu)$  and that

$$\tau = \int_{\mathcal{X}} K(x, x) d\mu(x) < +\infty. \quad (2.2)$$

Then, for a given subspace  $\mathbf{H}_C$  of  $\mathbf{H}$ , the IMSE criterion (or  $\mu$ -IMSE to explicitly refer to the measure  $\mu$ ) is the positive real

$$\text{IMSE}(\mathbf{H}_C) = \int_{\mathcal{X}} \mathbb{E} \left[ (Z_x - P_{\mathbf{H}_C} [Z_x])^2 \right] d\mu(x) = \int_{\mathcal{X}} K(x, x) - \mathbb{E} \left[ (P_{\mathbf{H}_C} [Z_x])^2 \right] d\mu(x).$$

From (2.2), we have  $\text{IMSE}(\mathbf{H}_C) = \tau - C_I(\mathbf{H}_C)$ , with

$$C_I(\mathbf{H}_C) = \int_{\mathcal{X}} \mathbb{E} \left[ (P_{\mathbf{H}_C} [Z_x])^2 \right] d\mu(x).$$

Note that minimizing the IMSE amounts to maximizing  $C_I(\mathbf{H}_C)$  since  $\tau$  does not depend on  $\mathbf{H}_C$ . For a fixed  $n \in \mathbb{N}^*$ , a  $n$ -point IMSE-optimal design is a set  $\{x_1, \dots, x_n\} \in \mathcal{X}^n$  that minimizes the IMSE criterion on  $\mathcal{X}^n$ .

### 2.4 Integral form of the IMSE

Consider a  $n$ -point design  $\{x_1, \dots, x_n\} \in \mathcal{X}^n$  and its associated subspace  $\mathbf{H}_{ev}$  defined by equation (2.1). Denote by  $\mathbf{z}$  the (column) random vector  $\mathbf{z} = (Z_{x_1}, \dots, Z_{x_n})^T$  and let  $\mathbf{K}$  be the covariance matrix of  $\mathbf{z}$ . We assume that  $Z$  and  $\{x_1, \dots, x_n\}$  are such that  $\mathbf{K}$  is invertible. The expression of the IMSE criterion associated with  $\{x_1, \dots, x_n\}$  is then given by  $\text{IMSE}(\mathbf{H}_{ev}) = \tau - C_I(\mathbf{H}_{ev})$ , where

$$C_I(\mathbf{H}_{ev}) = \int_{\mathcal{X}} \mathbf{k}^T(x) \mathbf{K}^{-1} \mathbf{k}(x) d\mu(x), \quad (2.3)$$

with, for  $x \in \mathcal{X}$ ,  $\mathbf{k}(x) = (K_{x_1}(x), \dots, K_{x_n}(x))^T$ . Under this form, each evaluation of the criterion requires the computation of an integral and therefore turns out to be computationally demanding.

## 3 Discrete measures and quadrature-designs

### 3.1 IMSE for quadrature-designs

We now consider the situation where the measure  $\mu$  is discrete,  $\mu = \sum_{j=1}^{N_q} \omega_j \delta_{s_j}$  with  $\omega_j > 0$ ,  $s_j \in \mathcal{X}$ ,  $N_q \in \mathbb{N}^*$  and  $\delta_s$  the Dirac measure centered at  $s$ , so that

$$\int_{\mathcal{X}} f(s) d\mu(s) = \sum_{j=1}^{N_q} \omega_j f(s_j), \quad (3.1)$$

for all  $\mu$ -integrable real-valued function  $f$  on  $\mathcal{X}$ . The point  $s_j$  for  $j \in \{1, \dots, N_q\}$  will be called the  $j$ -th quadrature point and we shall respectively refer to the sets  $\{s_j | 1 \leq j \leq N_q\}$  and  $\{\omega_j | 1 \leq j \leq N_q\}$

as the sets of quadrature points and quadrature weights. Notice that this framework corresponds to the typical situation where a pointwise quadrature rule is used to compute the integrated MSE in (2.3).

**Definition 3.1.** We call quadrature-design a design which is only composed of quadrature points. For  $n \in \mathbb{N}^*$  (with  $n \leq N_q$ ), the index set of a  $n$ -point quadrature-design  $\{s_{i_1}, \dots, s_{i_n}\}$  is the subset  $D = \{i_1, \dots, i_n\}$  of  $\{1, \dots, N_q\}$ .

We introduce the two  $N_q \times N_q$  matrices  $\mathbf{W} = \text{diag}(\omega_1, \dots, \omega_{N_q})$  and  $\mathbf{Q}$ , with  $\mathbf{Q}$  having  $i, j$  term  $\mathbf{Q}_{i,j} = K(s_i, s_j)$ ,  $1 \leq i, j \leq N_q$ . Thus,  $\mathbf{W}$  is the matrix of quadrature weights and  $\mathbf{Q}$  is the covariance matrix of quadrature points.

Let  $n \in \mathbb{N}^*$ , with  $n \leq N_q$  (in practice,  $n \ll N_q$ ) and let  $D = \{i_1, \dots, i_n\} \subset \{1, \dots, N_q\}$  be the index set of a  $n$ -point quadrature-design (see Definition 3.1). We denote by  $\mathbf{K} = \mathbf{Q}_{D,D}$  the covariance matrix for the design  $\{s_{i_1}, \dots, s_{i_n}\}$ . Similarly, we denote by  $\mathbf{Q}_{\cdot,D}$  (respectively  $\mathbf{Q}_{D,\cdot}$ ) the  $N_q \times n$  (resp.  $n \times N_q$ ) matrix formed by columns (resp. rows)  $i_1, \dots, i_n$  of  $\mathbf{Q}$ . From (2.3) and (3.1), we have

$$C_I(\mathbf{H}_{ev}) = \text{trace}(\mathbf{W}\mathbf{Q}_{\cdot,D}\mathbf{K}^{-1}\mathbf{Q}_{D,\cdot}). \quad (3.2)$$

In order to reduce the computational cost of the evaluation of the IMSE (from the perspective of multiple evaluations of the criterion), we introduce the  $N_q \times N_q$  matrix  $\mathbf{\Sigma} = \mathbf{Q}\mathbf{W}\mathbf{Q}$  (this matrix is computed once for all the design optimization process). Then, we have

$$C_I(\mathbf{H}_{ev}) = \text{trace}(\mathbf{K}^{-1}\mathbf{\Sigma}_{D,D}). \quad (3.3)$$

The computation of the IMSE (for a quadrature-design) is then reduce to the inversion of the matrix  $\mathbf{K}$  and next, to the computation of the trace of the product of two  $n \times n$  matrices (see Remark 3.1). Notice that the matrix  $\mathbf{\Sigma}$  is associated with the kernel  $\Sigma(\cdot, \cdot)$  given by, for  $x$  and  $y \in \mathcal{X}$ ,

$$\Sigma(x, y) = \int_{\mathcal{X}} K(x, t)K(y, t)d\mu(t).$$

The knowledge of the two kernels  $K(\cdot, \cdot)$  and  $\Sigma(\cdot, \cdot)$  thus theoretically allows for the computation of the IMSE for any design, through (3.3).

**Remark 3.1.** Let  $\mathbf{A}$  be a  $l \times m$  matrix and let  $\mathbf{B}$  be a  $m \times l$  matrix, with  $l$  and  $m \in \mathbb{N}^*$ . In order to compute  $\text{trace}(\mathbf{A}\mathbf{B})$ , it is of course not useful to compute the off-diagonal elements of the product  $\mathbf{A}\mathbf{B}$ . Instead, one shall compute  $\text{sum}(\mathbf{A} * \mathbf{B}^T)$ , where  $*$  stands for the Hadamard matrix product (element by element) and  $\text{sum}(\cdot)$  means that we sum all the elements of the matrix considered.  $\square$

### 3.2 Spectral approximation of the IMSE for quadrature-designs

We consider the spectral decomposition of the matrix  $\mathbf{Q}\mathbf{W}$  in the Hilbert space  $\mathbb{R}^{N_q}$  endowed with the inner product  $(\cdot|\cdot)_{\mathbf{W}}$ , with, for  $\mathbf{x}$  and  $\mathbf{y} \in \mathbb{R}^{N_q}$ ,

$$(\mathbf{x}|\mathbf{y})_{\mathbf{W}} = (\mathbf{x}|\mathbf{y})_{\mathbf{W}} = \mathbf{x}^T \mathbf{W}\mathbf{y}.$$

Denote by  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N_q} \geq 0$  the eigenvalues of the matrix  $\mathbf{Q}\mathbf{W}$  and by  $\mathbf{v}_1, \dots, \mathbf{v}_{N_q}$  their associated eigenvectors, i.e.  $\mathbf{Q}\mathbf{W} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$  with  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{N_q})$  and  $\mathbf{P} = (\mathbf{v}_1 | \dots | \mathbf{v}_{N_q})$ .

Then,  $\{\mathbf{v}_1, \dots, \mathbf{v}_{N_q}\}$  forms an orthonormal basis of  $\mathbb{R}^{N_q}$  endowed with the inner product  $(\cdot|\cdot)_{\mathbf{W}}$ , so that

$$\mathbf{P}^T \mathbf{W} \mathbf{P} = \text{Id}_{N_q}, \quad (3.4)$$

with  $\text{Id}_{N_q}$  the  $N_q$ -dimensional identity matrix (see Remark 3.2 for more details). We then introduce the matrix  $\mathbf{X} = \mathbf{P} \mathbf{\Lambda}$ .

**Proposition 3.1.** *Consider a  $n$ -point quadrature-design with index set  $D = \{i_1, \dots, i_n\}$  and associated covariance matrix  $\mathbf{K} = \mathbf{Q}_{D,D}$ . Then, we have*

$$C_I(\mathbf{H}_{ev}) = \text{trace}((\mathbf{X}_{D,\cdot})^T \mathbf{K}^{-1} \mathbf{X}_{D,\cdot}), \quad (3.5)$$

where  $\mathbf{X}_{D,\cdot}$  denotes the matrix formed by the  $n$  rows of  $\mathbf{X}$  having indices in  $D$ .

*Proof.* From equation (3.2) and the spectral decomposition  $\mathbf{Q} \mathbf{W} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^{-1}$ , we have

$$\begin{aligned} C_I(\mathbf{H}_{ev}) &= \text{trace}(\mathbf{Q}_{\cdot,D} \mathbf{K}^{-1} (\mathbf{P} \mathbf{\Lambda} \mathbf{P}^{-1})_{D,\cdot}) = \text{trace}(\mathbf{Q}_{\cdot,D} \mathbf{K}^{-1} \mathbf{X}_{D,\cdot} \mathbf{P}^{-1}) \\ &= \text{trace}(\mathbf{P}^{-1} \mathbf{Q}_{\cdot,D} \mathbf{K}^{-1} \mathbf{X}_{D,\cdot}) = \text{trace}((\mathbf{P}^{-1} \mathbf{Q})_{\cdot,D} \mathbf{K}^{-1} \mathbf{X}_{D,\cdot}). \end{aligned}$$

Then, combining the decomposition of  $\mathbf{Q} \mathbf{W}$  and expression (3.4), we obtain

$$\mathbf{P}^{-1} \mathbf{Q} = \mathbf{P}^{-1} \mathbf{Q} \mathbf{W} \mathbf{W}^{-1} = \mathbf{\Lambda} \mathbf{P}^{-1} \mathbf{W}^{-1} = \mathbf{\Lambda} \mathbf{P}^T = \mathbf{X}^T.$$

Since  $(\mathbf{X}^T)_{\cdot,D} = (\mathbf{X}_{D,\cdot})^T$ , this completes the proof.  $\square$

Using (3.5), we can approximate the IMSE criterion by spectral truncation. For a truncation subset  $I_{trc} \subset \{1, \dots, N_q\}$ , we denote by  $\mathbf{X}_{D,I_{trc}}$  the  $n \times N_{trc}$  matrix with entries  $\mathbf{X}_{j,k}$ , where  $j \in D$  and  $k \in I_{trc}$ . We then define the spectral-truncated IMSE criterion by

$$\text{IMSE}_{trc}(\mathbf{H}_{ev}) = \tau_{trc} - C_{I_{trc}}(\mathbf{H}_{ev}), \quad (3.6)$$

with  $\tau_{trc} = \sum_{k \in I_{trc}} \lambda_k$  and  $C_{I_{trc}}(\mathbf{H}_{ev}) = \text{trace}((\mathbf{X}_{D,I_{trc}})^T \mathbf{K}^{-1} \mathbf{X}_{D,I_{trc}})$ . The following chain of inequalities holds:

$$C_{I_{trc}}(\mathbf{H}_{ev}) \leq C_I(\mathbf{H}_{ev}) \leq C_{I_{trc}}(\mathbf{H}_{ev}) + \sum_{k \notin I_{trc}} \lambda_k,$$

so that the term  $\sum_{k \notin I_{trc}} \lambda_k$  gives an upper bound on the error induced by truncation. The spectral ratio

$$R_{trc} = \frac{\tau_{trc}}{\tau} = \frac{\sum_{k \in I_{trc}} \lambda_k}{\sum_{k=1}^{N_q} \lambda_k} \quad (3.7)$$

can be used as an indicator of the accuracy of the approximation by truncation.

Notice that we have chosen to use  $\tau_{trc}$  in expression (3.6) since we interpret the truncated-IMSE as the value of the IMSE when only the eigenvalues with index in  $I_{trc}$  are considered. We may use  $\tau$  (or any other constant) as well since this quantity does not depend on  $\mathbf{H}_{ev}$ . One usually choose  $I_{trc} = \{1, \dots, n_{trc}\}$ , i.e., only the  $n_{trc}$  largest eigenvalues of  $\mathbf{Q} \mathbf{W}$  are taken into account, and  $n_{trc} \leq N_q$  is then called the truncation level.

**Remark 3.2.** The matrix  $\mathbf{QW}$  defines a symmetric and positive operator on  $\mathbb{R}^{N_q}$  endowed with  $(\cdot|\cdot)_{\mathbf{W}}$ , since  $(\mathbf{QW}\mathbf{x}|\mathbf{y})_{\mathbf{W}} = (\mathbf{x}|\mathbf{QW}\mathbf{y})_{\mathbf{W}} = \mathbf{x}^T \mathbf{WQW}\mathbf{y}$  for all  $\mathbf{x}$  and  $\mathbf{y} \in \mathbb{R}^{N_q}$ . For numerical reasons, it is preferable to compute the spectral decomposition of  $\mathbf{W}^{\frac{1}{2}}\mathbf{QW}^{\frac{1}{2}}$  for the classical Euclidean structure of  $\mathbb{R}^{N_q}$  rather than the decomposition of  $\mathbf{QW}$  for  $(\cdot|\cdot)_{\mathbf{W}}$ . Notice that if  $\mathbf{v}$  is an eigenvector of the matrix  $\mathbf{QW}$  associated with the eigenvalue  $\lambda$ , i.e.,  $\mathbf{QW}\mathbf{v} = \lambda\mathbf{v}$ , then  $\mathbf{W}^{\frac{1}{2}}\mathbf{QW}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}}\mathbf{v} = \lambda\mathbf{W}^{\frac{1}{2}}\mathbf{v}$ . Therefore, the two matrices  $\mathbf{W}^{\frac{1}{2}}\mathbf{QW}^{\frac{1}{2}}$  and  $\mathbf{QW}$  have the same eigenvalues and, if  $\tilde{\mathbf{v}}$  an eigenvector of  $\mathbf{W}^{\frac{1}{2}}\mathbf{QW}^{\frac{1}{2}}$  (orthonormal for the Euclidean structure of  $\mathbb{R}^{N_q}$ ) for the eigenvalue  $\lambda$ , then  $\mathbf{v} = \mathbf{W}^{-\frac{1}{2}}\tilde{\mathbf{v}}$  is an (orthonormal) eigenvector of  $\mathbb{R}^{N_q}$  endowed with  $(\cdot|\cdot)_{\mathbf{W}}$  for the same  $\lambda$ .  $\square$

As mentioned in Section 3.1 for expression (3.3), from a design optimization perspective (so, requiring multiple evaluations of the criterion) and for a fixed truncation set  $I_{trc}$ , it is of interest to compute once for all the  $N_q \times N_q$  matrix  $\mathbf{M}_{trc} = \mathbf{X}_{\cdot, I_{trc}}(\mathbf{X}_{\cdot, I_{trc}})^T$ . We then have

$$C_{I_{trc}}(\mathbf{H}_{ev}) = \text{trace}(\mathbf{K}^{-1}(\mathbf{M}_{trc})_{D,D}). \quad (3.8)$$

## 4 Grid-restricted optimization

In view of the above, in order to approximate IMSE-optimal designs, we propose to optimize the IMSE or truncated-IMSE on the subset of the quadrature points. The number of different  $n$ -points quadrature-designs (composed of  $n$  distinct points) is  $\binom{n}{N_q}$ . So, for relatively small values of  $\binom{n}{N_q}$ , it is possible to evaluate the IMSE (or truncated-IMSE) of all the quadrature-designs. This exhaustive approach shall however quickly turn to be unusable for large values of  $\binom{n}{N_q}$ , despite the relative efficiency of the evaluation of the IMSE or truncated-IMSE criteria through (3.3) or (3.8). However, the ability of quickly computing the score of any quadrature-design allows for employing optimization strategies that require a large number of evaluations of the cost function.

Note that the restriction to quadrature-designs is not penalizing when the quadrature grid used to compute the IMSE is dense enough. In addition, numerical experimentations on a series of examples indicate that  $n$ -point quadrature-designs are generally quasi-optimal (and sometimes optimal) among all  $n$ -point designs on  $\mathcal{X}$ .

In what follows, we describe a stochastic global optimization strategy based on the Enhanced Stochastic Evolutionary (ESE) algorithm combined with a local descent optimization method. The ESE algorithm was proposed by Jin et al. (2005) (see also Fang et al. (2010, Chapter 4) for additional considerations and a general description) and is a variant of the well-known Simulated Annealing (SA) algorithm described by Kirkpatrick et al. (1983).

**Design Perturbations.** We consider designs perturbations that change a given quadrature-design in another quadrature-design. At each step of the algorithm, we shall consider  $N_{pert}$  distinct perturbations of the current design and then apply the acceptance/rejection rule to one of the  $N_{pert}$  candidate designs (see below, acceptance and stopping rules).

Obviously, various perturbation strategies can be used and it is impossible to give an overall description of all possible choices for this particular task. However, we think that a reasonable strategy should respect the following conditions:

- the construction of a perturbed design must be computationally fast (since a large number of perturbed designs will be generated during the optimization process);

- perturbations have to meet the expectations of the so-called improving process, i.e., they should enable convergence to local optima;
- perturbations should also meet the expectations of the exploration process, i.e., they should allow the algorithm to explore the whole search space  $\{s_j | 1 \leq j \leq N_q\}^n$ .

In the present work, we have chosen for simplicity to consider perturbation strategies that only modify *one point* of the current design at a time. Perturbations are applied successively, but in random order, to all the points of the design. We call *perturbation cycle* the operation of perturbing one time (in a random order) each point of the design.

Assume that the index set of the current quadrature-design is  $D = \{i_1, \dots, i_n\}$  and that we choose to perturb the design point with index  $i_k$ , that is, the point  $s_{i_k} \in \mathcal{X}$ , with  $k \in \{1, \dots, n\}$ . In order to construct  $N_{pert}$  perturbed designs, we simply propose  $N_{pert}$  mutually different substitutes for  $s_{i_k}$ , taken among quadrature points. To account for the improving and exploration processes, we fix two positive integers  $N_{prox}$  and  $N_{rand}$  (the perturbation parameters) such that  $N_{prox} + N_{rand} = N_{pert}$  (with “prox” standing for proximity and “rand” for random). Then,

- (i) we select  $N_{prox}$  quadrature points in the “neighborhood” of  $s_{i_k}$ , with indices  $I_{prox}$  such that  $I_{prox} \subset \{1, \dots, N_q\} \setminus D$ ;
- (ii) we randomly pick  $N_{rand}$  quadrature points, with indices in  $\{1, \dots, N_q\} \setminus (D \cup I_{prox})$ , with respect to a given probability.

The choice of the method used to complete (i) and (ii) is rather crucial and should be done in accordance with the specifications of the problem considered.

Concerning (i) (selection of  $N_{prox}$  neighbors), first notice that the notion of “neighborhood” obviously requires the definition of a metric on the input space  $\mathcal{X}$ . In our numerical experiments (Section 5) we use the metric induced by the covariance kernel  $K(\cdot, \cdot)$ , that is, for  $x$  and  $y \in \mathcal{X}$ ,  $\text{dist}(x, y)^2 = K(x, x) + K(y, y) - 2K(x, y)$ . In particular, this distance can be easily deduced from the matrix  $\mathbf{Q}$  and is always defined, whatever the problem considered.

Concerning (ii), in the experiments carried out in Section 5, we simply use uniform probability weights in order to randomly select the  $N_{rand}$  substitutes to  $s_{i_k}$ .

The information required to generate perturbed designs can be stored within two matrices  $\mathbf{M}_{prox}$  and  $\mathbf{M}_{rand}$ , computed before the optimization and defined as follows:

- the  $i, j$  entry of  $\mathbf{M}_{prox}$  is the index of the  $i$ -th nearest quadrature point to  $s_j$ ;
- $\mathbf{M}_{rand}$  is a  $N_q \times N_q$  stochastic matrix whose  $j$ -th row consists of the weights used to randomly perturb the  $j$ -th grid point.

**Acceptance and stopping rules.** At each perturbation step, we select the best (in terms of the criterion considered) of the  $N_{pert}$  perturbed designs. If this design yields an improvement, it is always accepted as the new current design. Otherwise (that is, if the best perturbed design does not yields an improvement), we select the best design among the  $N_{rand}$  randomly perturbed designs, and this design is accepted as the new current design with some probability (as for any simulated annealing like algorithm). This way, for  $N_{rand} = 0$ , the algorithm acts like a deterministic local descent and conversely, for  $N_{prox} = 0$ , the search is fully stochastic.



During the optimization, the best encountered design is updated and kept in memory and is finally returned at the end of the process.

The stopping rule for the optimization depends on the number of perturbation cycles and is chosen by the user. However, if a local descent is possible at the end of the process, it will always be accomplished (even if this exceeds the maximum number of authorized perturbation cycles). The local descent stops after one full perturbation cycle without improvement.

The probability of acceptance for a best random perturbation without improvement depends on a quantity called *temperature*, which is fixed following the same rules as the ESE algorithm (the temperature is updated during the optimization after a fixed number of perturbation cycles).

## 5 Numerical experiments

All computations have been performed with the free software **R**, see (R Core Team, 2013), on a 2012 MacBook Air endowed with 1.8 GHz Intel Core i5 processor with 4 Go RAM. With such material, we manage to handle quadrature with up to  $N_q = 10\,000$  points.

### 5.1 A two-dimensional example

Let  $\mathcal{X}$  be the unit square  $[0, 1]^2$ , we consider the integration measure  $\mu$  having the density (with respect to the Lebesgue measure on  $\mathcal{X}$ )

$$f(x) = (1 - r)^{\frac{3}{2}} \left[ 1 + \cos \left( 4\pi \min \left( \frac{r}{0.5}, 1 \right) \right) \right] + 0.2,$$

with  $r = \|x - c\|$ ,  $c = (1/2, 1/2)$  and where  $\|\cdot\|$  is the Euclidean norm. We shall use quadrature rules based on  $N_q$  points  $s_k \in \mathcal{X}$ ,  $k \in \{1, \dots, N_q\}$ , with weights  $\omega_k$  given by

$$\omega_k = \frac{1}{N_q} f(s_k), \tag{5.1}$$

see equation (3.1). An overview (contour plot) of the function  $f$  on  $[0, 1]^2$  is presented in Figure 1. For  $x = (x_1, x_2)$  and  $y = (y_1, y_2) \in \mathbb{R}^2$  we consider the covariance kernel  $K(x, y) = K_{\theta_1}(x_1, y_1)K_{\theta_2}(x_2, y_2)$ , with  $\theta_1 > 0$ ,  $\theta_2 > 0$  and, for  $i \in \{1, 2\}$ ,

$$K_{\theta_i}(x_i, y_i) = \left( 1 + \frac{\sqrt{3}}{\theta_i} |x_i - y_i| \right) \exp \left( - \frac{\sqrt{3}}{\theta_i} |x_i - y_i| \right) \tag{5.2}$$

(Matérn 3/2 kernel). In what follows, we set  $\theta_1 = \theta_2 = 0.12$ . We mainly focus on the problem of computing a 33-point IMSE optimal for this problem.

#### 5.1.1 Square grid

We use a  $n_g \times n_g$  square grid on  $[0, 1]^2$  for approximating the integrated MSE, with  $n_g = 37$ . The  $N_q = n_g^2$  quadrature points  $s_k$ ,  $1 \leq k \leq n_g^2 = 1369$ , are given by  $s_{(j-1)n_g+i} = (c_i, c_j)$  with  $c_j = (j-1)/n_g + 1/(2n_g)$ , with  $1 \leq i, j \leq n_g$  (midpoint rectangular quadrature rule).

Using the `eigen` routine of **R**, the full eigen-decomposition of the matrix  $\mathbf{W}^{\frac{1}{2}} \mathbf{Q} \mathbf{W}^{\frac{1}{2}}$  (see Remark 3.2) takes approximately 7 seconds and the computational time grows as  $N_q^3$ . The evaluation

of the IMSE or truncated-IMSE for a 33-point quadrature-design (using expressions (3.3) or (3.8)) requires 0.154 milliseconds (median duration over 10 000 evaluations).

Figure 1 shows the best, in terms of IMSE, 33-point quadrature-design  $D^*(G_{sq\,37\times 37})$  we have found on the square grid  $G_{sq\,37\times 37}$ , with an IMSE of 0.2350413

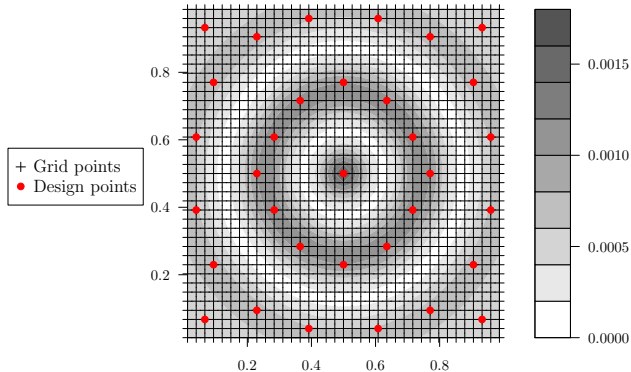


Figure 1: IMSE optimal 33-point quadrature-design  $D^*(G_{sq\,37\times 37})$  for the Matérn 3/2 kernel ( $\theta_1 = \theta_2 = 0.12$ ), and distribution (contour plot) of the weights  $\omega_k$  of the quadrature points  $s_k$ ,  $1 \leq k \leq N_q$  (square grid  $G_{sq\,37\times 37}$ ).

We have applied the algorithm described in Section 4 with  $N_{prox} = 8$  and  $N_{rand} = 120$ . The length  $\ell_{in}$  of the inner loop was set to  $\ell_{in} = 30$  and the length  $\ell_{out}$  of the outer loop to  $\ell_{out} = 4$ . The integer  $\ell_{in}$  stands for the number of perturbation cycles before the temperature is updated and  $\ell_{out}$  is the number of temperature updates. An optimization run then requires  $N_{eval} = 1 + N_{pert}n\ell_{in}\ell_{out}$  cost-function evaluations (with this parameters setting,  $N_{eval} = 506\,881$ ), with the possibility of some (few) additional evaluations if a final local descent is necessary. Random initial quadrature-designs have been used. In such settings, an optimization run takes approximately 90 seconds.

Interestingly, we have obtained roughly better results when optimizing the truncated-IMSE with truncation level of the order of the design size (but not too small), instead of the IMSE. For this 33-point design problem, the truncation level  $n_{trc} = 35$  has proven to be relatively effective. This phenomenon should however deserve further studies (see also Remark 5.1 and Section 6).

**Remark 5.1** (One-point designs). For  $t \in \mathcal{X}$ , we denote by  $\mathbf{H}_t$  the linear space spanned by  $Z_t$ . Figure 2 shows the graph of the functions  $t \mapsto C_I(\mathbf{H}_t)$  and  $t \mapsto C_{I_{trc}}(\mathbf{H}_t)$ , with  $n_{trc} = 3$ . We observe that the criterion  $C_I$  has many local maxima. This is not case for  $C_{I_{trc}}$ , and the truncated criterion may therefore be more easy to optimize.

The 1-point optimal IMSE and truncated-IMSE (with  $n_{trc} = 3$ ) designs are not exactly the same. However, if we denote by  $t^*$  and  $t_{trc}^*$  the 1-point IMSE and IMSE $_{trc}$  optimal designs respectively, we obtain  $\text{IMSE}(\mathbf{H}_{t^*})/\text{IMSE}(\mathbf{H}_{t_{trc}^*}) \approx 0.999976$ .  $\square$

### 5.1.2 Low discrepancy grid

We now use a quasi-Monte Carlo method (see, e.g., Niederreiter (1992)) in order to compute the integrated MSE. We generate a (Halton) uniform low-discrepancy sequence with  $N_q = 2\,500$  points in  $[0, 1]^2$  (using the R function `runif.halton`) and the associated weights are given by equation

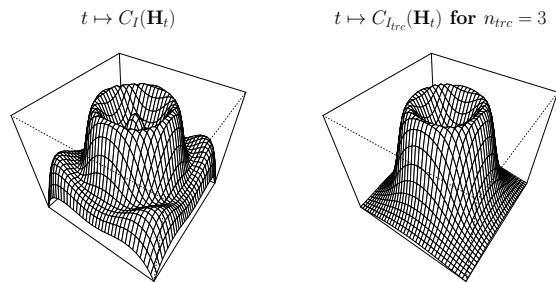


Figure 2: Graph of  $t \mapsto C_I(\mathbf{H}_t)$  and  $t \mapsto C_{I_{trc}}(\mathbf{H}_t)$ , with  $n_{trc} = 3$  on the square grid  $G_{sq\ 37 \times 37}$ .

(5.1). Such quadratures provide a relatively low accuracy regarding the number of points that are used; however, they have the important advantage of being extremely easy to implement, whatever the density of the integration measure  $\mu$  considered. We denote by  $G_{low\ 2500}$  this quadrature.

We optimize the IMSE and truncated-IMSE (with  $n_{trc} = 35$ ) on the grid  $G_{low\ 2500}$ . We use  $N_{prox} = 12$ ,  $N_{rand} = 120$ ,  $\ell_{in} = 40$  and  $\ell_{out} = 4$ . An optimization run then take approximatively 140 seconds. The quadrature-designs  $D^*(G_{low\ 2500})$  and  $D_{trc}^*(G_{low\ 2500})$  optimal for the IMSE and truncated IMSE are given in Figure 3, these two designs differ by two points.

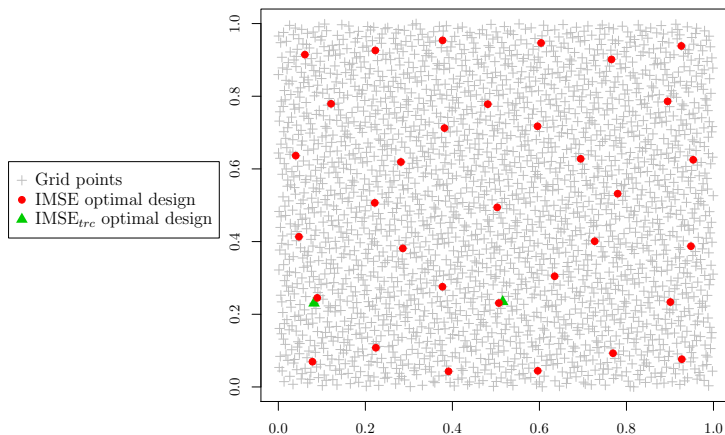


Figure 3: 33-point IMSE and truncated IMSE optimal quadrature-designs for the Matérn 3/2 kernel ( $\theta_1 = \theta_2 = 0.12$ ) for the quasi-Monte Carlo quadrature  $G_{low\ 2500}$  (the two designs differ by two points).

Finally, for comparison purpose, we consider a square grid  $G_{sq\ 74 \times 74}$  of  $74 \times 74 = 5476$  points and compute the IMSE score, for this grid, of the three designs  $D^*(G_{sq\ 37 \times 37})$ ,  $D^*(G_{low\ 2500})$  and  $D_{trc}^*(G_{low\ 2500})$ . Notice that these designs are not quadrature-designs for  $G_{sq\ 74 \times 74}$ , so that the IMSE scores are computed with the integral form (2.3) of the IMSE. Interestingly, we note that  $D_{trc}^*(G_{low\ 2500})$  performs better, although the differences are slight, see Table 1.

	$D^*(G_{sq\ 37\times 37})$	$D^*(G_{low\ 2500})$	$D_{trc}^*(G_{low\ 2500})$
$G_{sq\ 74\times 74}$ IMSE	0.2355560	0.2354285	0.2354256

Table 1: IMSE score, for the square grid  $G_{sq\ 74\times 74}$ , of the three designs  $D^*(G_{sq\ 37\times 37})$ ,  $D^*(G_{low\ 2500})$  and  $D_{trc}^*(G_{low\ 2500})$ .

## 5.2 Local descents in dimension 5

Consider the 5-dimensional Matérn covariance kernel on  $[0, 1]^5$ ,  $K(x, y) = \prod_{i=1}^5 K_{\theta_i}(x_i, y_i)$ , where  $x = (x_1, \dots, x_5)$  and  $y = (y_1, \dots, y_5)$  are in  $[0, 1]^5$  and where the kernels  $K_{\theta_i}(\cdot, \cdot)$  are given by expression (5.2). We set  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (0.22, 0.52, 0.52, 0.52, 0.22)$ . The measure  $\mu$  has density  $f_\mu(x) = f_\tau(x_1)f_\tau(x_2)$  with respect to the Lebesgue measure on  $[0, 1]^5$ , where  $f_\tau$  stands for the density of a truncated normal distribution on  $[0, 1]$  with mean 0.5 and standard deviation 0.15. The measure  $\mu$  is therefore the tensor product of two truncated normal distributions (for the first two variables) and three uniform distributions on  $[0, 1]$ .

We approximate  $\mu$  using a quadrature based on a  $\mu$ -distributed low-discrepancy sequence with  $N_q = 5\,000$  points. Since the quadrature points  $s_k$  are distributed according  $\mu$ , the quadrature weights are simply given by  $\omega_k = 1/N_q$ ,  $1 \leq k \leq N_q$ .

We focus on the construction of a 50-point quadrature-design. As previously, we consider both the IMSE and truncated-IMSE criteria. After various testing, the truncation level  $n_{trc} = 54$  appears as relatively efficient for this problem.

We consider 100 random low-discrepancy quadrature-designs (their respective indices consist of 50 consecutive integers in  $\{1, \dots, N_q\}$ ). We then run local descents starting from these low-discrepancy designs, and using the IMSE criterion or the truncated-IMSE criterion. We set  $N_{prox} = 2d + 2^d = 42$  (with  $d$  the dimension of  $\mathcal{X}$ ), and each descent then takes approximatively 4 seconds. Our results are listed in Figure 4. As mentioned in Section 5.1, we obtain slightly better results when optimizing the truncated criterion instead of the classic IMSE.

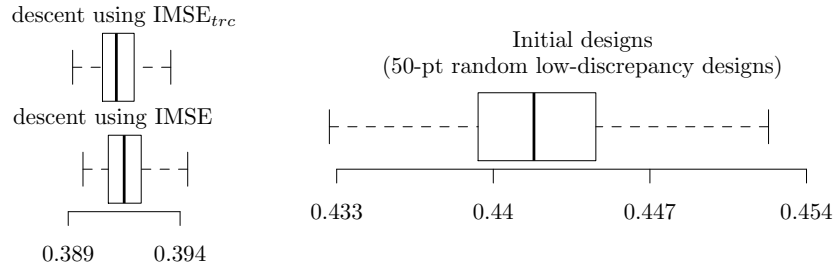


Figure 4: IMSE score (boxplot) for the 100 initial low-discrepancy quadrature-designs and for the designs obtained after a local descent using the IMSE criterion (bottom-left) or the truncated-IMSE criterion with  $n_{trc} = 54$  (top-left).

The best quadrature-design we have obtained has an IMSE of 0.3891981. An overview of this design is given in Figure 5. We can clearly distinguish the influences of the integration measure  $\mu$  and of the covariance parameters  $\theta$ .

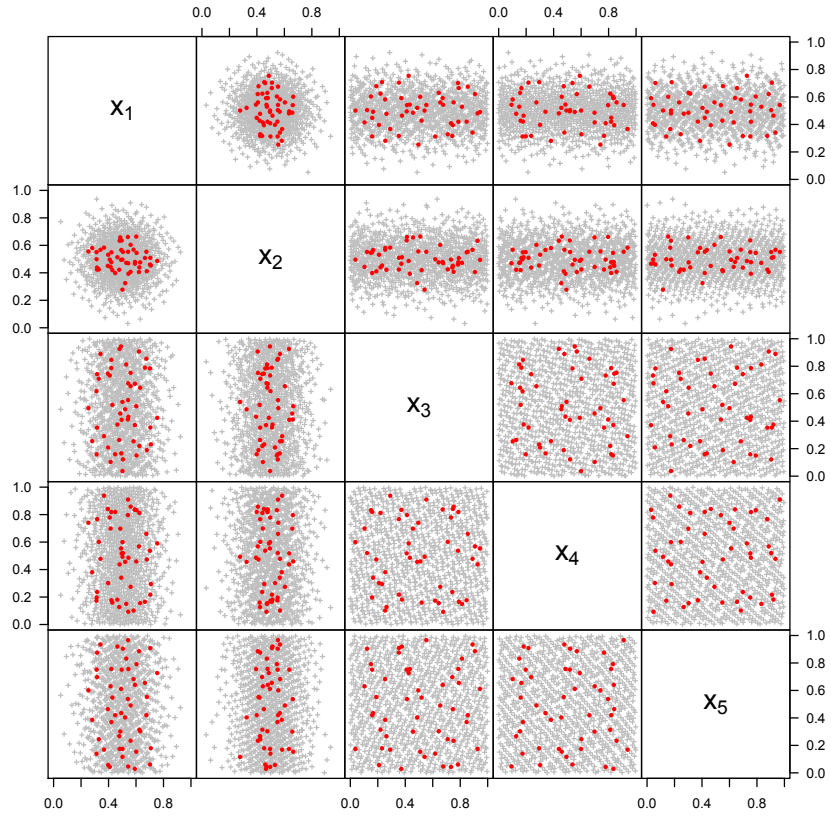


Figure 5: Pairs plot of the best 50-point quadrature-design obtained by local descent and of the 1 000 first points of the  $\mu$ -distributed low discrepancy grid ( $N_q = 5\,000$ ); Matérn 3/2 covariance kernel with parameters  $\theta = (0.22, 0.52, 0.52, 0.52, 0.22)$ .

## 6 Concluding remarks

We have considered the IMSE and truncated-IMSE (approximation of the IMSE by spectral truncation) design criteria in the context of optimal linear prediction of second order random fields with known mean and covariance. We have also assume that a (pointwise) quadrature rule is used to approximate the integrated MSE. In this framework, and if the design space restricted to subsets of quadrature points, we have illustrate how these two criteria can be easily and quickly evaluated at the cost of some simple preliminary calculations. There are no particular restrictions on the kernel  $K(\cdot, \cdot)$  and the measure  $\mu$  used to define the IMSE. The approach is thus quite general. From the numerical point of view, the truncated criterion can be used whenever the spectral decomposition of the matrix  $\mathbf{QW}$  can be computed, see Remark 3.2.

The preliminary calculations are made only once for a given design problem, and fine quadrature grids with many points (depending obviously of the computational power at disposal) can therefore be considered. Since each evaluation of the IMSE or truncated-IMSE is then computationally cheap, a global optimization (e.g., based on simulated-annealing) is affordable. When used in combination with low-discrepancy grids, this approach offers an easy to implement and relatively efficient way to approximate IMSE optimal designs on general input spaces, possibly with high dimension and complex shape.

The current version of the method applies to zero-mean processes (or, equivalently, to processes with known mean), and the inclusion of an unknown parametric trend is under investigation. Moreover, since the optimization is restricted to quadrature-designs, the choice of a suitable quadrature may be critical. A quantification of the errors induced by the use of a quadrature approximation and by the restriction to quadrature-designs would be of interest.

Our numerical experiments seems to indicate that considering the truncated-IMSE with truncation level of the order of the design size has many advantages. Indeed, the truncated criterion then appears as more easy to optimize than the IMSE while leading to designs that closed to the optimum. In addition, spectral truncation seems to reduce the impact of the quadrature approximation. These phenomena shall however deserve further studie and need to be put in relation with some of the results presented in Spöck and Pilz (2010).

## References

- Fang, K.-T., Li, R., and Sudjianto, A. (2010). *Design and Modeling for Computer Experiments*. CRC Press, Boca Raton.
- Jin, R., Chen, W., and Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, 134(1):268–287.
- Kirkpatrick, S., Gelatt, C., and Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.
- Niederreiter, H. (1992). *Random Number and Quasi-Monte Carlo Methods*. SIAM, Philadelphia.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. MIT press, Cambridge, MA.
- Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423.
- Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.
- Spöck, G. and Pilz, J. (2010). Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. *Stochastic Environmental Research and Risk Assessment*, 24(3):463–482.