



HAL
open science

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

Bertrand Gauthier, Luc Pronzato

► **To cite this version:**

Bertrand Gauthier, Luc Pronzato. Approximation of IMSE-optimal designs via quadrature rule and spectral decomposition. 2014. hal-00936681v1

HAL Id: hal-00936681

<https://hal.science/hal-00936681v1>

Preprint submitted on 27 Jan 2014 (v1), last revised 12 May 2015 (v3)

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 rules and spectral decomposition*

Bertrand GAUTHIER^{†‡} Luc PRONZATO^{§‡}

January 27, 2014

Abstract

We address the problem of computing IMSE (Integrated Mean Square Error) optimal designs for random fields interpolation with known mean and covariance. We assume that the MSE is integrated through a discrete measure and restrict the design space to the support of the measure considered. For such quadrature designs, the computational cost of IMSE evaluations can be significantly reduced by considering a spectral approximation with truncation. We illustrate how IMSE-optimal quadrature designs can be obtained at a reasonable computational cost with a simulated-annealing based optimization strategy.

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 fixed mean and covariance, see, *e.g.*, Sacks et al. (1989), Rasmussen and Williams (2006).

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, the evaluation of the IMSE criterion, in its standard form (see equation (2.3)), 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 restrict the optimization to quadrature designs, *i.e.*, to designs composed of quadrature points (see Definition 3.1). The IMSE score of such quadrature designs can then be approximated by spectral truncation. In general, keeping a small number of

[†]bgauthie@i3s.unice.fr (corresponding author)

[§]pronzato@i3s.unice.fr

[‡]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

eigenpairs is enough to obtain an accurate approximation of the IMSE while significantly reducing the computational cost of the criterion evaluation. The present work aims at describing a general design optimization strategy based on these considerations.

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 $y \in \mathcal{X}$,

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

We denote by \mathbf{H} 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. ζ_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.*

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

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 σ -finite measure μ 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 μ . We assume that the sample paths of $(Z_x)_{x \in \mathcal{X}}$ are in $L^2(\mathcal{X}, \mu)$ with \mathbb{P} -probability 1. More precisely (see Gauthier and Pronzato (2013)), we assume that the following trace class condition is satisfied:

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

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

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

From equation (2.2), we have

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

Note that minimizing the IMSE amounts to maximizing $C_I(\mathbf{H}_C)$ since τ 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 $x \in \mathcal{X}$, $\mathbf{k}(x) = (K_{x_1}(x), \dots, K_{x_n}(x))^T$. Here, for $t \in \mathcal{X}$, K_t stands for the function $x \mapsto K(t, x)$, $x \in \mathcal{X}$.

3 Discrete measures and quadrature designs

3.1 IMSE for quadrature designs

We now consider the situation where the measure μ 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 δ_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 μ -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 quadrature rule is used to compute the integrated MSE in equation (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. We assume that \mathbf{Q} is nonsingular.

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} .

Using equations (2.3) and (3.1), we obtain

$$\begin{aligned} C_I(\mathbf{H}_{ev}) &= \text{trace}(\mathbf{W}\mathbf{Q}_{\cdot, D}\mathbf{K}^{-1}\mathbf{Q}_{D, \cdot}) \\ &= \text{trace}(\mathbf{Q}_{\cdot, D}\mathbf{K}^{-1}\mathbf{Q}_{D, \cdot}\mathbf{W}) = \text{trace}(\mathbf{W}^{\frac{1}{2}}\mathbf{Q}_{\cdot, D}\mathbf{K}^{-1}\mathbf{Q}_{D, \cdot}\mathbf{W}^{\frac{1}{2}}), \end{aligned} \quad (3.2)$$

and $\text{IMSE}(\mathbf{H}_{ev}) = \tau - C_I(\mathbf{H}_{ev})$, with $\tau = \text{trace}(\mathbf{Q}\mathbf{W})$.

Remark 3.1. According to equation (3.2), we only need to compute the diagonal terms of the matrix $\mathbf{W}\mathbf{Q}_{\cdot, D}\mathbf{K}^{-1}\mathbf{Q}_{D, \cdot}$. In order to save computational time, we can consider the Cholesky decomposition $\mathbf{K} = \mathbf{C}^T\mathbf{C}$ and compute $\mathbf{M}_I = \mathbf{W}^{\frac{1}{2}}\mathbf{Q}_{\cdot, D}\mathbf{C}^{-1}$. We then have $C_I(\mathbf{H}_{ev}) = \text{sum}(\mathbf{M}_I * \mathbf{M}_I)$, where $*$ denotes the term-by-term product of two matrices of same size and $\text{sum}(\mathbf{A})$ stands for the sum of all terms of \mathbf{A} . Note that if the IMSE needs to be evaluated for a large number of different designs, it is computationally advantageous to compute the product matrix $\mathbf{S} = \mathbf{W}^{\frac{1}{2}}\mathbf{Q}$ once for all and then use $\mathbf{S}_{\cdot, D}$ instead of $\mathbf{W}^{\frac{1}{2}}\mathbf{Q}_{\cdot, D}$, thereby avoiding repeated matrix multiplications. \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 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N_q} > 0$ 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

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{N_q}) \text{ and } \mathbf{P} = (\mathbf{v}_1 | \dots | \mathbf{v}_{N_q}).$$

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

$$\mathbf{P}^T \mathbf{W} \mathbf{P} = \text{Id}_{N_q}, \quad (3.3)$$

with Id_{N_q} the N_q -dimensional identity matrix (see Remark 3.3 for more details). We then define 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*

$$C_I(\mathbf{H}_{ev}) = \text{trace}((\mathbf{X}_{D, \cdot})^T \mathbf{K}^{-1} \mathbf{X}_{D, \cdot}), \quad (3.4)$$

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{QW} and expression (3.3), 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

Following Gauthier and Pronzato (2013), we now introduce a spectral truncation of expression (3.4). For a positive integer $N_{trc} \leq N_q$ (the truncation level), we define $I_{N_{trc}} = \{1, \dots, N_{trc}\}$ and 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

$$\text{IMSE}_{trc}(\mathbf{H}_{ev}) = \tau_{trc} - C_{I_{trc}}(\mathbf{H}_{ev}), \quad (3.5)$$

with $\tau_{trc} = \sum_{k=1}^{N_{trc}} \lambda_k$ and $C_{I_{trc}}(\mathbf{H}_{ev}) = \text{trace}((\mathbf{X}_{D, I_{trc}})^T \mathbf{K}^{-1} \mathbf{X}_{D, I_{trc}})$.

Remark 3.2. As noticed in Remark 3.1 for the usual expression of the IMSE criterion, one may save computational time when computing the term $C_{I_{trc}}(\mathbf{H}_{ev})$ of expression (3.5) by considering the Cholesky decomposition $\mathbf{K} = \mathbf{C}^T \mathbf{C}$. We then compute the matrix $\mathbf{M}_T = (\mathbf{X}_{D, I_{trc}})^T \mathbf{C}^{-1}$ and obtain $C_{I_{trc}}(\mathbf{H}_{ev}) = \text{sum}(\mathbf{M}_T * \mathbf{M}_T)$, hence avoiding the computation of non-diagonal terms of the matrix $(\mathbf{X}_{D, I_{trc}})^T \mathbf{K}^{-1} \mathbf{X}_{D, I_{trc}}$. \square

Notice that we have chosen to use τ_{trc} in expression (3.5) since we interpret the truncated-IMSE criterion as the value of the IMSE criterion when only the N_{trc} largest eigenvalues of \mathbf{QW} are taken into account. We might consider $\tau - C_{I_{trc}}(\mathbf{H}_{ev})$ as well since τ and τ_{trc} do not depend on the design considered. We recall that the following chain of inequalities holds:

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

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

$$R_{trc} = \frac{\tau_{trc}}{\tau} = \frac{\sum_{k=1}^{N_{trc}} \lambda_k}{\sum_{k=1}^{N_q} \lambda_k} \quad (3.6)$$

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

The interest of the truncated criterion lies in the fact that one can often choose $N_{trc} \ll N_q$ while keeping a very good accuracy, see for instance the examples in Section 5. In addition, the numerical experiments performed in (Gauthier and Pronzato, 2013) seem to indicate that retaining a small number of eigenpairs is often sufficient to obtain IMSE-optimal quadrature designs when optimizing the truncated criterion. Once the spectral decomposition of \mathbf{QW} is completed, the computation of IMSE_{trc} is then significantly faster than the one of the usual form of the IMSE criterion. This is of particular interest for the determination of IMSE-optimal designs since many evaluations of the IMSE are usually required.

Remark 3.3. 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 λ , *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 λ , 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 λ . \square

4 Grid-restricted IMSE optimization

In view of the above and in order to determine IMSE-optimal designs (or at least good approximations of them), we propose to use optimization routines that

- restrict the optimization to quadrature designs (the design optimization then corresponds to a combinatorial optimization problem),
- use the truncated criterion instead of the usual form of the IMSE criterion.

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 the numerical experiments of Section 5, we consider a stochastic global optimization strategy based on the Enhanced Stochastic Evolutionary (ESE) algorithm proposed by Jin et al. (2005). The ESE algorithm is adapted from the SE (Stochastic Evolutionary) algorithm originally developed by Saab and Rao (1991), see also Fang et al. (2010, Chapter 4) for additional considerations and a general description. Both algorithms are variants of the well-known Simulated Annealing (SA) algorithm described by Kirkpatrick et al. (1983). For comparison purposes, we shall also consider a local descent optimization method.

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} such distinct perturbations of the current design and apply the acceptance/rejection rule to the best of the N_{pert} candidate designs (in terms of the cost function considered).

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 then applied successively to the first design point, then the second, *etc.*

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 Euclidian norm on $\mathcal{X} = [0, 1]^d$. We also consider two different rules for the selection of neighbors: either we pick the N_{prox} points nearest to s_{i_k} (we refer to this case as the *proximity rule*), or we randomly (uniform distribution) pick N_{prox} points among the $2N_{prox}$ nearest neighbors of s_{i_k} (what we call the *random proximity rule*).

Concerning (ii), in the experiments carried out in Section 5, we propose to use probability weights proportional to $(\mathbf{QW})_{i_k, \cdot}$ in order to compute the N_{rand} substitutes to s_{i_k} . This choice appears as a trade-off between a well-spread distribution centered at s_{i_k} (since we consider stationary kernels with relatively large correlation lengths, this allows a good exploration) and the impact of quadrature weights (since points with high weights may have more influence on the IMSE).

The information required to generate perturbed designs can be stored within two matrices \mathbf{M}_{prox} and \mathbf{M}_{rand} , 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 perturbed the j -th grid point.

Remark 4.1. Since perturbed designs differ from the current design by one point only, particular updating formulae could be used to compute the IMSE (or IMSE_{trc}) score of the perturbed designs, see for instance Davis and Hager (2005). One might also parallelize the computations of the IMSE scores of all N_{pert} perturbed designs. However, we have not considered the implementation of these ideas in the current version of our optimization algorithm. \square

Acceptance and stopping rules. In both the ESE and local descent algorithms, at each perturbation step the acceptance rule is applied to the best of the N_{pert} perturbed designs. In a simulated annealing algorithm like ESE, a perturbation is always accepted if it yields an improvement and is accepted with some probability otherwise. For the local descent method, a perturbed design is accepted only if the criterion value is improved. The stopping rule for the ESE algorithm is simply given by a bound on the total number of iterations. For local descent, the stopping rule is based on the number of successive design-perturbation steps without improvement.

5 Numerical experiments

All computations have been performed with the free software **R**, see (R Core Team, 2013), sometimes combined with calls to **C/C++** in order to save computational time, on a 2012 MacBook Air endowed with 1.8 GHz Intel Core i5 processor with 4 Go RAM. The computational times indicated correspond to the total time elapsed before the computer returns the response and are provided by the **R** function `system.time`. The **R** implementation of the design criteria considered and of the algorithms of Section 4 is available at <http://www.i3s.unice.fr/~bgauthie/>.

5.1 A two-dimensional example

5.1.1 Problem description

In this first numerical example, \mathcal{X} is the unit square $[0, 1]^2$. We consider the integration measure μ 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 ω_k given by

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

see equation (3.1). Figure 1-Left presents a plot (up to a multiplicative constant) of the function $x \mapsto f(x)$, $x \in [0, 1]^2$. 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 $i \in \{1, 2\}$,

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

(Matérn 3/2 kernel). In what follows, we set $\theta_1 = \theta_2 = 0.12$.

5.1.2 Square grid

In this section we consider a rectangular quadrature rule for approximating the integrated MSE and use a $n_g \times n_g$ square grid on $[0, 1]^2$, 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 $c_j = j/n_g + 1/(2n_g)$, with $1 \leq i, j \leq n_g$ (midpoint rectangular quadrature rule).

The computation of the covariance matrix \mathbf{Q} (taking its symmetry into account) takes 15.177 s when using two nested `for` loops in **R**. If we only use one `for` loop and construct the matrix column by column, 9.365 s are required. Finally, if we compute the covariance kernel and the two nested loop in **C/C++**, the computational time is reduced to 0.111 s. In all cases, the computational time grows as N_q^2 .

Using the `eigen` routine of **R**, the eigen-decomposition of the matrix $\mathbf{W}^{\frac{1}{2}} \mathbf{Q} \mathbf{W}^{\frac{1}{2}}$ (see Remark 3.3) takes 7.209 s and the computational time grows as N_q^3 . Figure 1-Right shows the eigenvalues λ_k (with $\tau = \sum_{k=1}^{N_q} \lambda_k \simeq 0.7455805$) obtained for this quadrature.



Figure 1: Left: weights ω_k for $N_q = 1369$ quadrature points s_k on the regular square grid $G_{sq\ 37 \times 37}$ in $\mathcal{X} = [0, 1]^2$. Right: eigenvalues λ_k for $1 \leq k \leq N_q$ for the grid $G_{sq\ 37 \times 37}$.

	$N_{trc} = 120$	$N_{trc} = 257$	$N_{trc} = 1\ 000$	$N_{trc} = 1\ 369$
Spectral ratio R_{trc}	0.9602847	0.9900167	0.9999658	1
Time ($n = 17$)	0.0271 s	0.0330 s	0.0963 s	0.1246 s
Time ($n = 33$)	0.0535 s	0.0702 s	0.2340 s	0.3069 s
Time ($n = 43$)	0.0715 s	0.1023 s	0.3490 s	0.4514 s

Table 1: Average time (50 repetitions), in seconds, for the 100 evaluation of the IMSE_{trc} criterion for random quadrature designs, with various design sizes n and various truncation levels N_{trc} (square-grid $G_{sq\ 37 \times 37}$). — Note that $\text{IMSE}_{trc} = \text{IMSE}$ when $N_{trc} = N_q = 1\ 369$.

Table 1 gives an overview of the computational time required for the evaluation of the IMSE_{trc} criterion (taking the time-saving considerations of Remark 3.2 into account). One may notice that the truncated criterion with $N_{trc} = 257$ yields a significant reduction of computational time compared to the case $N_{trc} = N_q$ while the spectral ratio R_{trc} is above 99%, indicating a good accuracy for the approximation of the IMSE.

We then use the ESE algorithm (with the proximity rule) to compute 33-point quadrature designs that minimize the IMSE. We set $N_{prox} = N_{rand} = 8$ (so that $N_{pert} = 16$), the length ℓ_{in} of the inner loop of the ESE algorithm (see Jin et al. (2005)) is set to $\ell_{in} = 33 \times 6$ and the length ℓ_{out} of the outer loop to $\ell_{out} = 120$. An optimization run then require $N_{eval} = 1 + N_{pert} \ell_{in} \ell_{out}$ cost-function evaluations (with this parameters setting, $N_{eval} = 380\ 161$). Random initial designs are used (with respect to probabilities given by quadrature weights). Table 2 indicates the computational time required by one optimization depending on the criterion considered.

The same 33-point optimal design $D^*(G_{sq\ 37 \times 37})$ is obtained with $N_{trc} = 120$, $N_{trc} = 257$ and $N_{trc} = 1\ 369$ on the grid $G_{sq\ 37 \times 37}$; it is presented in Figure 2, its IMSE score is approximately 0.2350413.

Criterion	$N_{trc} = 120$	$N_{trc} = 257$	$N_{trc} = 1\,369$
Time (optimization)	199 s	269 s	1 195 s

Table 2: Time, in seconds, for one minimization of the IMSE_{trc} (two different truncation levels) and IMSE criteria using the ESE algorithm with $N_{prox} = N_{rand} = 8$, $\ell_{in} = 33 \times 6$ and $\ell_{out} = 120$ ($n = 33$, square-grid $G_{sq\,37 \times 37}$).

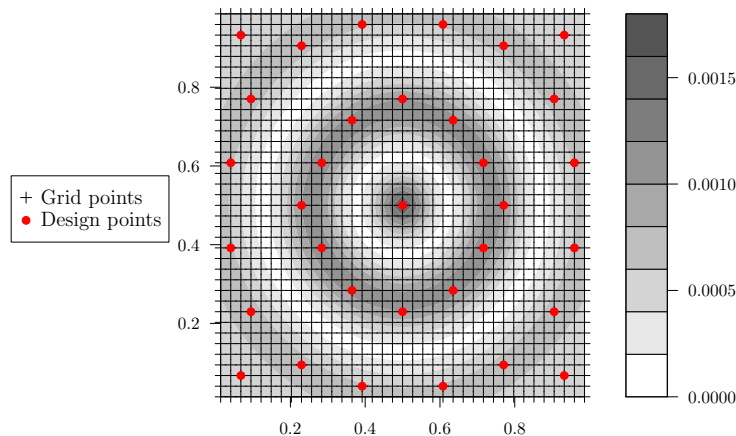


Figure 2: 33-point IMSE_{trc} -optimal quadrature design $D^*(G_{sq\,37 \times 37})$ (with $N_{trc} = 257$) for the Matérn 3/2 kernel ($\theta_1 = \theta_2 = 0.12$), and distribution (contour plot) of the weights ω_k of the quadrature points s_k , $1 \leq k \leq N_q$ (square-grid $G_{sq\,37 \times 37}$).

5.1.3 Low discrepancy nested grids

We now use a quasi-Monte Carlo method (see, *e.g.*, Niederreiter (1992)) in order to compute the integrated MSE. We also illustrate the behavior of the optimal quadrature designs as the quadrature size N_q increases.

We generate four nested (Halton) uniform low-discrepancy sequences with respectively $N_1 = 300$, $N_2 = 800$, $N_3 = 1\,500$ and $N_4 = 2\,500$ points in $[0, 1]^2$ (using the R function `runif.halton`). We hence obtain four nested sets of quadrature points, $G_{N_1} \subset G_{N_2} \subset G_{N_3} \subset G_{N_4}$, the associated weights being given by equation (5.1). These 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 μ considered.

Table 3 summarizes the results obtained. The integer $N_{trc}^*(0.99)$ corresponds to the minimum value of N_{trc} such that $R_{trc} \geq 0.99$; the table indicates that it quickly stabilizes as N_q increases. Therefore, when we optimize the truncated IMSE criterion given by expression (3.5), the use of fine grids (large N_q) only impacts the initial eigen-decomposition step. The quadrature designs $D^*(G_{N_i})$ are those obtained with the ESE algorithm (same parameter settings as in Section 5.1.2) for the minimization of the IMSE_{trc} criterion truncated at $N_{trc}^*(0.99)$ (below we call them “optimal” although this can be slightly abusive). The optimal designs $D^*(G_{N_1})$ to $D^*(G_{N_4})$ are presented in Figure 3.

	G_{N_1}	G_{N_2}	G_{N_3}	G_{N_4}	$G_{sq\ 74 \times 74}$
N_q	300	800	1 500	2 500	5 476
$N_{trc}^*(0.99)$	176	239	258	265	
τ	0.735299	0.7473631	0.7437508	0.7447645	0.7453939
$\text{IMSE}[D^*(G_{N_1})]$	0.2245987	0.2428404	0.2434066	0.2459614	0.247831
$\text{IMSE}[D^*(G_{N_2})]$		0.2338697	0.2377162	0.2382043	0.239534
$\text{IMSE}[D^*(G_{N_3})]$			0.2344847	0.2365742	0.2377174
$\text{IMSE}[D^*(G_{N_4})]$				0.2348011	0.2354285
$\text{IMSE}[D^*(G_{sq\ 37 \times 37})]$					0.235556

Table 3: Summary of the results obtained using low-discrepancy grids (dimension 2) and comparison, for the quadrature $G_{sq\ 74 \times 74}$, of the IMSE scores of the designs $D^*(G_{sq\ 37 \times 37})$ and $D^*(G_{N_2}), \dots, D^*(G_{N_4})$.

Finally (rightmost column of Table 3), we consider a regular square grid $G_{sq\ 74 \times 74}$ of $74 \times 74 = 5476$ points and compute the IMSE score, for this grid, of the designs $D^*(G_{sq\ 37 \times 37})$ and $D^*(G_{N_1}), \dots, D^*(G_{N_4})$. 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. We note that $D^*(G_{N_4})$ performs better than $D^*(G_{sq\ 37 \times 37})$ when the IMSE is evaluated with a quadrature based on the regular grid $G_{sq\ 74 \times 74}$.

Remark 5.1. In the framework of nested grids $G_{N_1} \subset G_{N_2} \cdots$, a quadrature design for G_{N_k} is also a quadrature design for $G_{N_{k+1}}$. This suggests the following strategy for the construction of IMSE-optimal quadrature designs: first, optimize the IMSE for the grid G_{N_1} and obtain a design $D^*(G_{N_1})$; next, use $D^*(G_{N_1})$ as initial design for the optimization of the IMSE on the grid G_{N_2} , etc. \square

5.2 An example 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 μ 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_τ stands for the density of a truncated normal distribution on $[0, 1]$ with mean 0.5 and standard deviation 0.15. The measure μ is therefore the tensor product of two truncated normal distributions (for the first two variables) and three uniform distributions on $[0, 1]$.

We approximate μ using a quadrature based on a μ -distributed low-discrepancy sequence with N_q points. Since the quadrature points s_k are distributed according μ , the quadrature weights are simply given by $\omega_k = 1/N_q$, $1 \leq k \leq N_q$. We next consider the IMSE_{trc} criterion with the minimal truncation level $N_{trc}^*(0.9)$ such that $R_{trc} \geq 0.9$. Table 4 shows the evolution of $N_{trc}^*(0.9)$ with the quadrature size (nested low-discrepancy grids). As in Section 5.1.3, we notice a quick stabilization of $N_{trc}^*(0.9)$ when N_q increases.

We now focus on the construction of a 42-point IMSE_{trc} -optimal quadrature design for the low-discrepancy grid with 4000 points. We use $N_{trc} = 233$.

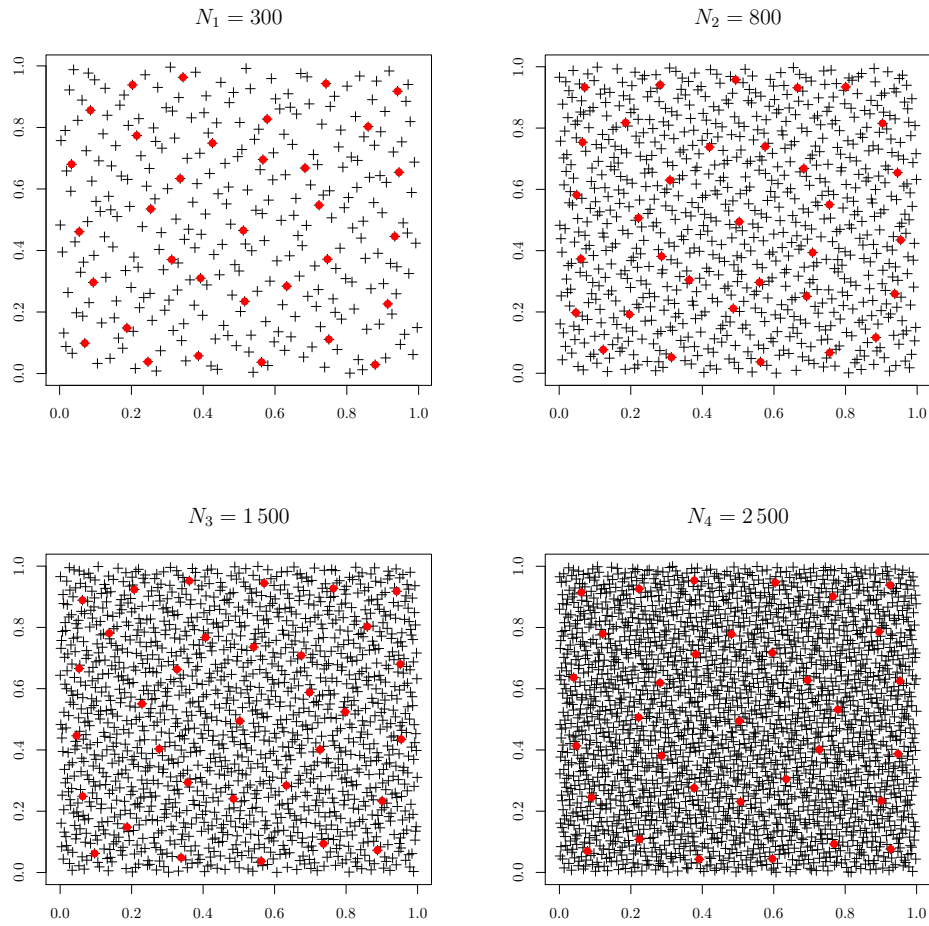


Figure 3: 33-point IMSE-optimal quadrature designs for the Matérn 3/2 kernel ($\theta_1 = \theta_2 = 0.12$) for a nested sequence of quasi-Monte Carlo quadratures.

N_q	1 000	1 750	2 500	3 250	4 000
$N_{trc}^*(0.9)$	186	211	226	231	233

Table 4: Truncation level $N_{trc}^*(0.9)$ (*i.e.*, minimal value for N_{trc} such that $R_{trc} \geq 0.9$) for a sequence of nested low-discrepancy grids with N_q points (dimension 5).

To compare the performances of the ESE and local descent algorithms, using the proximity and random proximity rules (we thus compare four different optimization strategies), we run various optimizations initialized at random low-discrepancy quadrature designs (their respective indices consist of 42 consecutive integers in $\{1, \dots, N_q\}$). We also investigate the influence of the perturbation parameters N_{rand} and N_{prox} .

The length of the ESE inner loop is set to $\ell_{in} = 42 \times 5$ and that of the outer loop to $\ell_{out} = 120$. With this values, an optimization run takes approximately 20 minutes when $N_{rand} = N_{prox} = 30$. The duration of a run of the local descent does not exceed a few minutes (the descent is stopped after n consecutive unsuccessful perturbation steps for the proximity rule, and $2n$ such unsuccessful steps for the random proximity rule).

Figure 4 summarizes the results obtained after optimization using the strategies described above. Min, max and mean values of the IMSE scores are indicated. The right part concerns the 10 initial designs we have considered (they are used to initialize both the local descent and ESE algorithms). The performances of the local descent for various values of N_{prox} and for the proximity (bottom) and random proximity (top) rules are indicated in the central part of the graph. The left part of Figure 4 (lowest IMSE values) corresponds to the ESE algorithm with again the proximity (bottom) and random proximity (top) rules, for two different values of $N_{prox} = N_{rand}$.

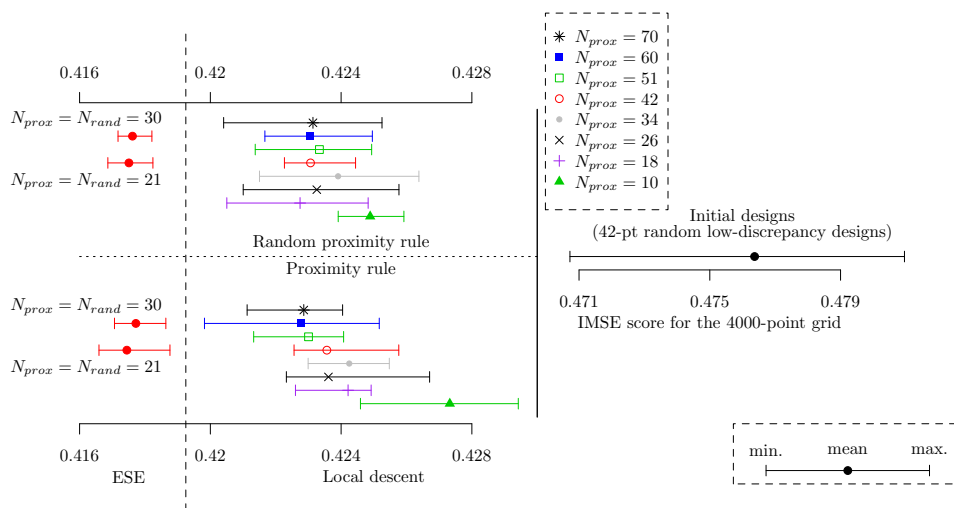


Figure 4: Graphical representation of the IMSE scores (min, max and mean values) on the 4000-point grid for 10 initial quadrature designs and after optimization using ESE or local descent with various values of the perturbation parameters N_{prox} and N_{rand} .

We notice that the best scores are always obtained with the ESE algorithm. The results for the local descent method indicate that, for the problem considered, the random proximity rule is superior to the proximity rule for small N_{prox} (in connection with the dimension of \mathcal{X} , here 5). For the values of $N_{prox} = N_{rand}$ considered ($N_{prox} = N_{rand} = 21$ and $N_{prox} = N_{rand} = 30$), the performance of the ESE algorithm does not seem to be very sensitive to the choice between the

proximity and random proximity rules. Notice that the choice of $N_{pert} = N_{prox} + N_{rand}$ must take the regularity of the grid and the space dimension d into account. For low-discrepancy grids, values about $2d + 2^d$ seem reasonable. Indeed, the computational cost of the optimization grows with N_{pert} and larger values do not necessarily improve performance.

The best quadrature design has been obtained with the ESE algorithm using the proximity rule and $N_{prox} = N_{rand} = 21$. A pairs plot of this design and of the 4000-point grid considered is given in Figure 5. We can clearly distinguish the influences of the integration measure μ and of the covariance parameters θ .

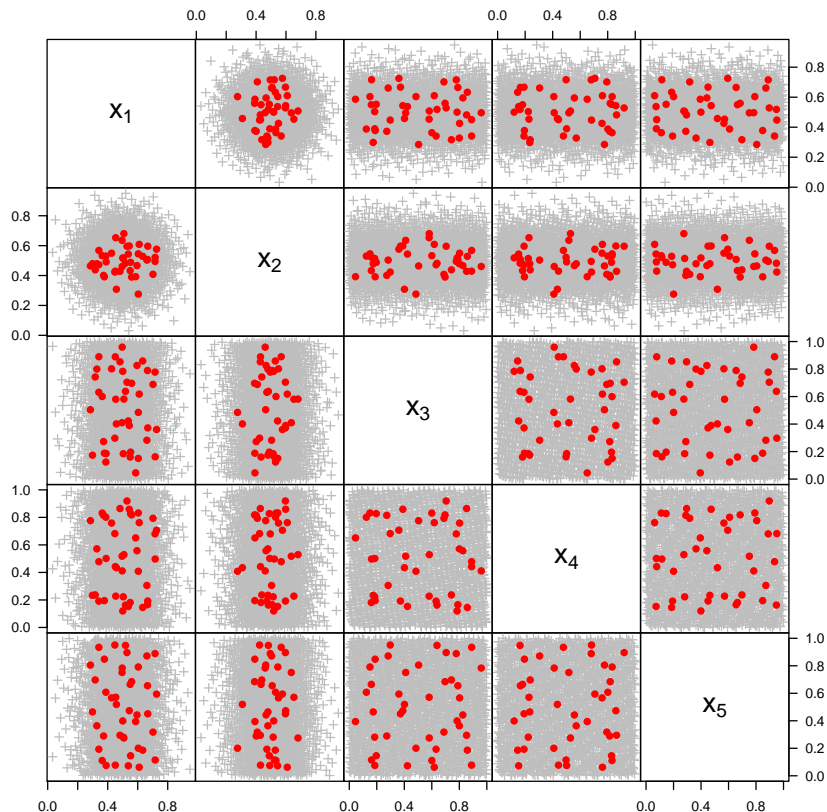


Figure 5: Pairs plot of the 4 000-point grid and of the best 42-point quadrature design obtained for a Matérn 3/2 covariance kernel with parameters $\theta = (0.22, 0.52, 0.52, 0.52, 0.22)$.

6 Concluding remarks

In the context of optimal linear prediction of second order random fields with known mean and covariance, we have shown that when the IMSE is approximated via a quadrature rule, IMSE-optimal quadrature designs can be determined at reasonable computational cost. Indeed, a single eigen-decomposition, followed by a suitable spectral truncation, yields a significant reduction of the

amount of calculations required by the optimization of the IMSE criterion over quadrature points.

There are no particular restrictions on the kernel $K(\cdot, \cdot)$ and the measure μ used to define the IMSE. The approach is thus quite general. From the numerical point of view, it can be used whenever the spectral decomposition of the matrix \mathbf{QW} can be computed, see Remark 3.3. Since this decomposition is computed only once for a given design problem, fine quadrature grids with many points can be considered. Once the decomposition is performed, each truncated IMSE evaluation is computationally cheap and a global optimization (*e.g.*, based on simulated-annealing) is affordable. When used in combination with low-discrepancy grids, the 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. Since $K(\cdot, \cdot)$ does not need to be stationary, the method can be used for the sequential construction of IMSE-optimal quadrature designs. In particular, for one-point-at-a-time constructions, the truncated IMSE criterion can be easily evaluated at each quadrature point.

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. The objective is to obtain a spectral representation that gives a tight approximation through spectral truncation while authorizing computational cost reduction.

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 the restriction to quadrature designs would be of interest, including for instance the determination of conditions ensuring that among all n -point IMSE optimal designs one design is supported on quadrature points.

References

- Davis, T. and Hager, W. (2005). Row modifications of a sparse cholesky factorization. *SIAM Journal on Matrix Analysis and Applications*, 26(3):621–639.
- Fang, K.-T., Li, R., and Sudjianto, A. (2010). *Design and Modeling for Computer Experiments*. CRC Press, Boca Raton.
- Gauthier, B. and Pronzato, L. (2013). Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models. *Preprint*. <http://hal.archives-ouvertes.fr/hal-00913466> — submitted July 10, 2013.
- 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.

- Saab, Y. and Rao, V. (1991). Combinatorial optimization by stochastic evolution. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 10(4):525–535.
- 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.