



**HAL**  
open science

# Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models.

Bertrand Gauthier, Luc Pronzato

► **To cite this version:**

Bertrand Gauthier, Luc Pronzato. Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models.. 2014. hal-00913466v3

**HAL Id: hal-00913466**

**<https://hal.science/hal-00913466v3>**

Preprint submitted on 18 Sep 2014 (v3), last revised 21 Nov 2014 (v4)

**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.

# SPECTRAL APPROXIMATION OF THE IMSE CRITERION FOR OPTIMAL DESIGNS IN KERNEL-BASED INTERPOLATION MODELS

BERTRAND GAUTHIER<sup>†\*</sup> AND LUC PRONZATO<sup>‡\*</sup>

**Abstract.** We address the problem of computing IMSE-optimal designs for random fields interpolation models. A spectral representation of the IMSE criterion is obtained from the eigendecomposition of the integral operator defined by the random field covariance kernel and the integration measure considered. The IMSE can this way be approximated by spectral truncation and bounds on the error induced by truncation are given. We then illustrate how the IMSE and truncated-IMSE can be easily computed when a quadrature rule is used to approximate the integrated MSE and the design space is restricted to subset of quadrature points. Numerical experiments are carried out and indicate (i) that retaining a small number of eigenpairs (in regards of the quadrature size) is often sufficient to tightly approximate IMSE optimal quadrature-designs when optimizing the truncated criterion and (ii) that optimal quadrature-designs generally give efficient approximations of the true optimal designs for the quadrature approximation of the IMSE.

**Key words.** Random field model, interpolation, design of experiments, IMSE, integral operator, quadrature approximation.

**AMS subject classifications.** 62K99, 65C60, 62G08

**1. Introduction.** This work addresses the problem of designing experiments (i.e., of choosing sampling points) in the framework of kernel-based interpolation models (see for instance [RW06, Wah90]). The integrated mean-squared error (IMSE) criterion is a classical tool for evaluating the overall performance of interpolators (see for example [SWMW89]). For a fixed class of models and a given design size, it is therefore natural to try to choose sampling points such that the resulting interpolation minimizes the IMSE criterion among all possible samplings. One then speaks of IMSE-optimal design of experiments.

IMSE-optimal designs are generally difficult to compute, see, *e.g.*, [SWMW89, ABM12]. Indeed, the direct evaluation of the IMSE criterion is numerically expensive (it requires the computation of the integral of the mean-squared prediction error over the whole space) and its global optimization is often made difficult due to the presence of many local minima (many evaluations of the criterion are thus required). The present work aims at investigating an alternative approach to make the computation of IMSE-optimal designs more tractable by reducing the computational cost of the criterion evaluation.

The choice of an IMSE criterion for learning a random field naturally leads to the definition of an integral operator (see Section 3 and, *e.g.*, [ST06]). The interest of such operators when dealing with kernel-based interpolation models has been discussed for instance in [CS02, LMK10, GB12], see also [SY66, DPZ13] for applications to optimal design for linear regression models with correlated errors. The main idea of the present work is to link the IMSE criterion with the spectral decomposition of its associated integral operator. We hence obtain a spectral representation of the IMSE criterion which can be approximated by spectral truncation, and bounds on the error induced by truncation are given.

---

<sup>†</sup>bgauthie@i3s.unice.fr

<sup>‡</sup>proncato@i3s.unice.fr

\*CNRS, Université de Nice-Sophia Antipolis, Laboratoire I3S - UMR 7271.

From a numerical point of view, the IMSE and truncated-IMSE criteria can be easily computed when a (pointwise) quadrature is used to approximate the integral of the MSE and the design construction is restricted to quadrature-designs (i.e., designs only composed of quadrature points, see Definition 4.2). Numerical experiments indicate that retaining a small number of eigenpairs (in regards of the quadrature size) is often sufficient to obtain efficient approximation of IMSE-optimal quadrature-designs when optimizing the truncated criterion. They also indicate that optimal quadrature-designs give in general good approximations of the true optimal designs for the quadrature approximation of the IMSE, so that the restriction to quadrature points appears to have small impact (this restriction has sometimes no impact at all, see in particular Remark 5.1).

We have tried to make the paper as self-contained as possible: the definitions of most concepts are reminded and most proofs are detailed (in the body of the article or, for the sake of readability, in an appendix). We first describe a general setting in which the spectral representation of the IMSE is well-defined. We next focus on the classical IMSE optimal design problem (conditioning by a finite number of evaluations), with particular attention to the quadrature approximation case.

The paper is organized as follows. Section 2 is devoted to the introduction of the general framework of conditioning of Gaussian random fields. General results concerning the IMSE criterion and the associated integral operator are given in Section 3, where the spectral representation of the IMSE criterion and its approximation by spectral truncation are detailed. The computation and the approximation of the IMSE criterion for designing experiments is considered in Section 4. Numerical experiments are carried out in Section 5. Section 6 concludes and gives some perspectives.

## 2. General framework and notations.

**2.1. Random fields and involved Hilbert structures.** Let  $\mathcal{X}$  be a general set. We consider a real random field  $(Z_x)_{x \in \mathcal{X}}$  indexed by  $\mathcal{X}$ . 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 (so that the optimal linear prediction is the optimal prediction). In what follows,  $Z$  will refer to the random field  $(Z_x)_{x \in \mathcal{X}}$ .

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 the random field  $Z$ . Since, by assumption, for all  $x$  and  $y \in \mathcal{X}$ ,  $\mathbb{E}(Z_x) = \mathbb{E}(Z_y) = 0$ , we have

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

We denote by  $\mathbf{H}$  the Gaussian Hilbert space generated by  $Z$ ;  $\mathbf{H}$  is the closed linear subspace of  $L^2(\Omega, \mathbb{P})$  spanned by the r.v.  $Z_x$ ,  $x \in \mathcal{X}$ , i.e.,

$$\mathbf{H} = \overline{\text{span}\{Z_x, x \in \mathcal{X}\}}^{L^2(\Omega, \mathbb{P})}.$$

The linear space  $\mathbf{H}$  is endowed with the Hilbert structure involved by  $L^2(\Omega, \mathbb{P})$ . For the sake of simplicity, we assume that  $\mathbf{H}$  is separable (see Remark B.2).

In parallel, we denote by  $\mathcal{H}$  the reproducing kernel Hilbert space (RKHS, see for instance [BT04]) of real-valued functions on  $\mathcal{X}$  associated with the kernel  $K(\cdot, \cdot)$ . We use the classical notation,  $K_x(\cdot) = K(x, \cdot)$ , for  $x \in \mathcal{X}$  (and  $K_x \in \mathcal{H}$ ). We remind that  $\mathcal{H}$  is characterized by the representation property,

$$\forall h \in \mathcal{H}, \forall x \in \mathcal{X}, (h|K_x)_{\mathcal{H}} = h(x), \quad (2.1)$$

with  $(\cdot|\cdot)_{\mathcal{H}}$  the inner product of  $\mathcal{H}$ . Also, if  $\{h_j, j \in J\}$  is an orthonormal basis of  $\mathcal{H}$ , we have

$$\forall x \text{ and } y \in \mathcal{X}, K(x, y) = \sum_{j \in J} h_j(x)h_j(y). \quad (2.2)$$

The two Hilbert spaces  $\mathcal{H}$  and  $\mathbf{H}$  are isometric thanks to the relation, for all  $x$  and  $y \in \mathcal{X}$ ,  $(K_x|K_y)_{\mathcal{H}} = K(x, y) = \mathbb{E}(Z_x Z_y)$ . We denote this isometry by  $\mathcal{I} : \mathcal{H} \rightarrow \mathbf{H}$ , with  $\mathcal{I}(K_x) = Z_x$ .

**2.2. Conditioning.** Let  $\mathbf{H}_D$  be a closed linear subspace of the Hilbert space  $\mathbf{H}$ . We consider the orthogonal projection  $P_{\mathbf{H}_D}$  of  $\mathbf{H}$  onto  $\mathbf{H}_D$ . For  $x \in \mathcal{X}$ , the r.v.  $P_{\mathbf{H}_D}[Z_x]$  is called the conditional mean of the r.v.  $Z_x$  relatively to  $\mathbf{H}_D$ . If  $\mathbf{H}_D$  is spanned by the r.v.  $\zeta_j, j \in J$ , with  $J$  a general index set, the notation

$$P_{\mathbf{H}_D}[Z_x] = \mathbb{E}(Z_x | \zeta_j, j \in J)$$

is often used. The covariance of the random field  $(Z_x - P_{\mathbf{H}_D}[Z_x])_{x \in \mathcal{X}}$  is called the conditional covariance of  $Z$  relatively to  $\mathbf{H}_D$ . We shall pay particular attention to subspaces of the evaluation-type, i.e.,

$$\mathbf{H}_{ev} = \text{span}\{Z_{x_1}, \dots, Z_{x_n}\}, \quad (2.3)$$

with  $n \in \mathbb{N}^*$  (the set of all positive integers) and  $x_1, \dots, x_n \in \mathcal{X}$ , see Section 4.

**Remark 2.1.** By isometry, a conditioning problem in the Gaussian Hilbert space  $\mathbf{H}$  is associated with an optimal interpolation problem in the RKHS  $\mathcal{H}$ . To the subspace  $\mathbf{H}_D$  of  $\mathbf{H}$  corresponds a subspace  $\mathcal{H}_D$  of  $\mathcal{H}$  and one can define the orthogonal projection  $P_{\mathcal{H}_D}$  of  $\mathcal{H}$  onto  $\mathcal{H}_D$ , etc.  $\triangleleft$

### 3. IMSE criterion and associated integral operator.

**3.1. IMSE criterion and working hypotheses.** From now on we suppose that  $\mathcal{X}$  is a measurable space and we consider a  $\sigma$ -finite measure  $\mu$  on  $\mathcal{X}$ . We denote by  $L^2(\mathcal{X}, \mu)$  the (not necessarily separable) Hilbert space of square integrable (with respect to  $\mu$ ) real-valued functions on  $\mathcal{X}$ . Notice that elements of  $L^2(\mathcal{X}, \mu)$  are in fact equivalent classes of  $\mu$ -almost everywhere equal functions; however, when it will not be source of confusion, we shall assimilate elements of  $L^2(\mathcal{X}, \mu)$  with functions on  $\mathcal{X}$ .

Throughout the rest of the article, we make the following assumptions:

- C-i. any  $h \in \mathcal{H}$  is a measurable function,
- C-ii. the kernel  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is measurable (for the product  $\sigma$ -algebra),
- C-iii.  $\tau = \int_{\mathcal{X}} K(x, x) d\mu(x) < +\infty$ .

For a given closed linear subspace  $\mathbf{H}_D$  of  $\mathbf{H}$ , the *integrated mean-squared error* criterion associated with  $\mathbf{H}_D$  (IMSE, or when necessary  $\mu$ -IMSE, to explicitly refer to the measure  $\mu$ ) is the integral of the conditional variance of  $Z$  relatively to  $\mathbf{H}_D$ ; more precisely,

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

Under the assumptions above,  $\text{IMSE}(\mathbf{H}_D)$  is well-defined for any closed linear subspace  $\mathbf{H}_D$  of  $\mathbf{H}$  (see Remark B.1) and we have

$$\text{IMSE}(\mathbf{H}_D) = \tau - C_I(\mathbf{H}_D), \text{ with } C_I(\mathbf{H}_D) = \int_{\mathcal{X}} \mathbb{E} [(P_{\mathbf{H}_D} [Z_x])^2] d\mu(x). \quad (3.1)$$

Since  $\tau$  does not depend on  $\mathbf{H}_D$ , minimizing the IMSE amounts to maximize  $C_I(\mathbf{H}_D)$ . We have  $C_I(\{0\}) = 0 \leq C_I(\mathbf{H}_D) \leq C_I(\mathbf{H}) = \tau$ .

**3.2. Integral operator defined by the IMSE.** The introduction of an IMSE criterion for the learning of a random field  $Z$  naturally defines, under conditions C-i, C-ii, C-iii, an integral operator  $T_\mu$  on  $L^2(\mathcal{X}, \mu)$ . We first recall the following lemma; the proof is in Appendix A.

**Lemma 3.1.** *Under conditions C-i and C-iii, the RKHS  $\mathcal{H}$  is continuously included into  $L^2(\mathcal{X}, \mu)$ , that is, for all  $h \in \mathcal{H}$ ,  $h \in L^2(\mathcal{X}, \mu)$  and*

$$\|h\|_{L^2}^2 \leq \tau \|h\|_{\mathcal{H}}^2. \quad (3.2)$$

From Lemma 3.1, we know that  $K_x \in L^2(\mathcal{X}, \mu)$  for all  $x \in \mathcal{X}$  and we can therefore define, without any ambiguity,

$$\forall f \in L^2(\mathcal{X}, \mu), \forall x \in \mathcal{X}, T_\mu[f](x) = (K_x|f)_{L^2} = \int_{\mathcal{X}} f(t)K(x, t)d\mu(t).$$

Let us now recall some of the main properties of the operator  $T_\mu$  (the proof of Lemma 3.2 is given in Appendix A).

**Lemma 3.2.** *Under C-i, C-ii and C-iii, we have  $K(\cdot, \cdot) \in L^2(\mathcal{X} \times \mathcal{X}, \mu \otimes \mu)$  and the operator  $T_\mu$  is a compact (Hilbert-Schmidt), self-adjoint and positive operator on  $L^2(\mathcal{X}, \mu)$ .*

From Lemma 3.2 and the spectral theorem for compact self-adjoint operators on Hilbert spaces (see for instance [Sch79]),  $T_\mu$  is diagonalizable and its eigenfunctions form a complete orthogonal system. We denote by  $\lambda_i \geq 0$  the considered eigenvalues (repeated according to their geometric multiplicity) and by  $\tilde{\phi}_i \in L^2(\mathcal{X}, \mu)$  their associated eigenfunctions, with  $i \in \mathbb{I}$ , a general index set. We classically choose the eigenfunctions  $\{\tilde{\phi}_i, i \in \mathbb{I}\}$  so that they form an orthonormal basis of  $L^2(\mathcal{X}, \mu)$ . We also denote by  $\{\lambda_k, k \in \mathbb{I}_+\}$  the set (at most countable) set of all strictly positive eigenvalues of  $T_\mu$ , that is  $\lambda_k > 0$  for all  $k \in \mathbb{I}_+$ .

**Proposition 3.1.** *Let  $\mathcal{H}_0$  be the linear subspace of  $\mathcal{H}$  defined by*

$$\mathcal{H}_0 = \left\{ h_0 \in \mathcal{H}, \|h_0\|_{L^2}^2 = 0 \right\}.$$

*Denote by  $\mathcal{H}_\mu$  the orthogonal of  $\mathcal{H}_0$  in  $\mathcal{H}$  (i.e.,  $\mathcal{H}_\mu = \mathcal{H}_0^\perp$ ) and, for  $k \in \mathbb{I}_+$ , consider the functions  $\phi_k$  given by*

$$\forall x \in \mathcal{X}, \phi_k(x) = \frac{1}{\lambda_k} \int_{\mathcal{X}} \tilde{\phi}_k(t)K(x, t)d\mu(t) = \frac{1}{\lambda_k} (\tilde{\phi}_k|K_x)_{L^2} = \frac{1}{\lambda_k} T_\mu[\tilde{\phi}_k](x). \quad (3.3)$$

*Then  $\mathcal{H}_0$  is closed in  $\mathcal{H}$  and  $\{\sqrt{\lambda_k}\phi_k, k \in \mathbb{I}_+\}$  forms an orthonormal basis of  $\mathcal{H}_\mu$  for the Hilbert structure of  $\mathcal{H}$ .*

The proof is given in Appendix A.

**Remark 3.1.** We have  $\phi_k \stackrel{L^2(\mathcal{X}, \mu)}{=} \tilde{\phi}_k$  (or more precisely,  $\phi_k$  belongs to the equivalent class  $\tilde{\phi}_k$ ). However, as elements of  $L^2(\mathcal{X}, \mu)$ , the  $\tilde{\phi}_k$  are only defined  $\mu$ -almost everywhere whereas the  $\phi_k$  are defined on the whole set  $\mathcal{X}$ .

For all  $f \in L^2(\mathcal{X}, \mu)$  and  $h \in \mathcal{H}$ , we have (see the proof of Proposition 3.1)

$$T_\mu[f] \in \mathcal{H}_\mu \text{ and } (h|f)_{L^2} = (h|T_\mu[f])_{\mathcal{H}}.$$

Therefore, the functions  $\phi_k \in \mathcal{H}$  satisfy the following property:

$$\forall h \in \mathcal{H}, \forall k \in \mathbb{I}_+, (h|\phi_k)_{\mathcal{H}} = \frac{1}{\lambda_k} (h|T_\mu[\tilde{\phi}_k])_{\mathcal{H}} = \frac{1}{\lambda_k} (h|\tilde{\phi}_k)_{L^2}$$

and in particular, for  $k \in \mathbb{I}_+$  and  $x \in \mathcal{X}$ ,  $(\phi_k|K_x)_{L^2} = \lambda_k (\phi_k|K_x)_{\mathcal{H}} = \lambda_k \phi_k(x)$ . Notice that  $\tau = \sum_{k \in \mathbb{I}_+} \lambda_k$ , see C-iii.  $\triangleleft$

**3.3. Spectral representation of the IMSE.** For  $k \in \mathbb{I}_+$ , we introduce the r.v.  $\xi_k = \mathcal{I}(\sqrt{\lambda_k} \phi_k) \in \mathbf{H}$ , where  $\mathcal{I}$  is the isometry between  $\mathcal{H}$  and  $\mathbf{H}$  defined in Section 2.1, so that the  $\xi_k$ ,  $k \in \mathbb{I}_+$ , are orthonormal in  $\mathbf{H}$ . Following Proposition 3.1, we denote by  $\mathbf{H}_\mu$  the closed linear subspace of  $\mathbf{H}$  spanned by the r.v.  $\xi_k$ ,  $k \in \mathbb{I}_+$  and by  $\mathbf{H}_0$  its orthogonal, so that we have the orthogonal decomposition

$$\mathbf{H} = \mathbf{H}_\mu \perp \mathbf{H}_0. \quad (3.4)$$

**Proposition 3.2.** Let  $\mathbf{H}_D$  be a closed linear subspace of  $\mathbf{H}$ , we have

$$C_I(\mathbf{H}_D) = C_I(\mathbf{H}_D \cap \mathbf{H}_\mu),$$

where  $C_I$  is defined in equation (3.1).

*Proof.* From (3.4), we have the orthogonal decomposition  $\mathbf{H}_D = \mathbf{H}_{\mu D} + \mathbf{H}_{0D}$ , where  $\mathbf{H}_{\mu D} = \mathbf{H}_\mu \cap \mathbf{H}_D$  and  $\mathbf{H}_{0D} = \mathbf{H}_0 \cap \mathbf{H}_D$ . The orthogonal projection of  $\mathbf{H}$  onto  $\mathbf{H}_D$  is then given by  $P_{\mathbf{H}_D} = P_{\mathbf{H}_{\mu D}} + P_{\mathbf{H}_{0D}}$ .

For all  $x \in \mathcal{X}$ , the r.v.  $P_{\mathbf{H}_{\mu D}}[Z_x]$  and  $P_{\mathbf{H}_{0D}}[Z_x]$  are orthogonal, hence

$$\mathbb{E}[(P_{\mathbf{H}_D}[Z_x])^2] = \mathbb{E}[(P_{\mathbf{H}_{\mu D}}[Z_x])^2] + \mathbb{E}[(P_{\mathbf{H}_{0D}}[Z_x])^2],$$

so that  $C_I(\mathbf{H}_D) = C_I(\mathbf{H}_D \cap \mathbf{H}_\mu) + C_I(\mathbf{H}_D \cap \mathbf{H}_0)$ . To conclude, we consider an orthonormal basis  $\{g_j, j \in J\}$  of the RKHS  $\mathcal{H}_{0D}$ . From (2.2), we have

$$C_I(\mathbf{H}_{0D}) = \int_{\mathcal{X}} K_{0D}(x, x) d\mu(x) = \sum_{j \in J} \int_{\mathcal{X}} g_j^2(x) d\mu(x) = 0,$$

since  $g_j \in \mathcal{H}_0$  for all  $j \in J$ , where  $K_{0D}(\cdot, \cdot)$  is the reproducing kernel of  $\mathcal{H}_{0D}$  (the interchange between the sum and the integral is justified by Tonelli's theorem).  $\square$

**Proposition 3.3** (Spectral representation of the IMSE criterion). Let  $\mathbf{H}_D$  be a closed linear subspace of  $\mathbf{H}$  and let  $\{\eta_j, j \in J\}$  be an orthonormal basis of  $\mathbf{H}_D$ . Then, we have

$$C_I(\mathbf{H}_D) = \sum_{k \in \mathbb{I}_+} \sum_{j \in J} \alpha_{j,k}^2 \lambda_k, \quad (3.5)$$

with, for  $j \in J$  and  $k \in \mathbb{I}_+$ ,  $\alpha_{j,k} = \mathbb{E}(\eta_j \xi_k)$ .

*Proof.* From Proposition 3.1,  $\{\xi_k, k \in \mathbb{I}_+\}$  forms an orthonormal basis of  $\mathbf{H}_\mu$ , so that

$$\forall j \in J, P_{\mathbf{H}_\mu}[\eta_j] = \sum_{k \in \mathbb{I}_+} \alpha_{j,k} \xi_k. \quad (3.6)$$

For all r.v.  $U \in \mathbf{H}$ , we have

$$P_{\mathbf{H}_D}[U] = \sum_{j \in J} \mathbb{E}(\eta_j U) \eta_j \text{ and } \mathbb{E}[(P_{\mathbf{H}_D}[U])^2] = \sum_{j \in J} (\mathbb{E}(\eta_j U))^2. \quad (3.7)$$

Combining relations (3.7) with Proposition 3.2, we obtain

$$\begin{aligned} C_I(\mathbf{H}_D) &= C_I(\mathbf{H}_D \cap \mathbf{H}_\mu) = \int_{\mathcal{X}} \mathbb{E}[(P_{\mathbf{H}_D}[P_{\mathbf{H}_\mu}[Z_x]])^2] d\mu(x) \\ &= \int_{\mathcal{X}} \sum_{j \in J} (\mathbb{E}(\eta_j P_{\mathbf{H}_\mu}[Z_x]))^2 d\mu(x). \end{aligned} \quad (3.8)$$

For all  $x \in \mathcal{X}$ , we have  $P_{\mathbf{H}_\mu}[Z_x] = \sum_{k \in \mathbb{I}_+} \xi_k \mathbb{E}(\xi_k Z_x) = \sum_{k \in \mathbb{I}_+} \xi_k \sqrt{\lambda_k} \phi_k(x)$ . Injecting this expansion in (3.8) and using (3.6), we obtain

$$C_I(\mathbf{H}_D) = \sum_{j \in J} \int_{\mathcal{X}} \left[ \sum_{k \in \mathbb{I}_+} \alpha_{j,k} \sqrt{\lambda_k} \phi_k(x) \right]^2 d\mu(x) = \sum_{k \in \mathbb{I}_+} \sum_{j \in J} \alpha_{j,k}^2 \lambda_k,$$

which completes the proof.  $\square$

We now recall the following well-known result (Proposition 3.4), which shows the optimal character of the r.v.  $\xi_k, k \in \mathbb{I}_+$ , in terms of IMSE.

**Proposition 3.4.** *For a fixed  $n \in \mathbb{N}^*$ , consider  $\mathbf{H}_n^* = \text{span}\{\xi_1, \dots, \xi_n\}$ , where  $\xi_1, \dots, \xi_n$  are associated with the  $n$  largest eigenvalues of  $T_\mu$ , denoted respectively by  $\lambda_1 \geq \dots \geq \lambda_n$ . Then  $\mathbf{H}_n^*$  minimizes the IMSE criterion among all subspaces  $\mathbf{H}_n$  of  $\mathbf{H}$  with dimension  $n$  and  $C_I(\mathbf{H}_n^*) = \sum_{k=1}^n \lambda_k$ .*

Notice that  $\mathbf{H}_n^*$  is not necessarily unique, depending of the multiplicity of  $\lambda_n$ .

*Proof.* Let  $\mathbf{H}_n$  be a closed linear subspace of  $\mathbf{H}_\mu$  with dimension  $n$  (the restriction to subsets of  $\mathbf{H}_\mu$  follows from Proposition 3.2 and yields no loss of generality) and let  $\{\eta_1, \dots, \eta_n\}$  be an orthonormal basis of  $\mathbf{H}_n$ . From Proposition 3.3, since  $\{\xi_1, \dots, \xi_n\}$  forms an orthonormal basis of  $\mathbf{H}_n^*$ , we have  $C_I(\mathbf{H}_n^*) = \sum_{k=1}^n \lambda_k$ .

For any  $i$  and  $j \in \{1, \dots, n\}$ , we have  $\sum_{k \in \mathbb{I}_+} \alpha_{i,k} \alpha_{j,k} = \delta_{ij}$  (Kronecker delta).

For  $k \in \mathbb{I}_+$ , let  $\mathbf{a}_k \in \mathbb{R}^n$  be the column vector given by  $\mathbf{a}_k = (\alpha_{1,k}, \dots, \alpha_{n,k})^T$ , so that  $\sum_{j=1}^n \alpha_{j,k}^2 = \mathbf{a}_k^T \mathbf{a}_k$ , where  $\mathbf{a}_k^T$  stands for the transpose of the vector  $\mathbf{a}_k$ . We also have  $\sum_{k \in \mathbb{I}_+} \mathbf{a}_k \mathbf{a}_k^T = \text{Id}_{n \times n}$  (the  $n \times n$  identity matrix). So, for  $l \in \mathbb{I}_+$ ,

$$\mathbf{a}_l^T \left( \sum_{k \in \mathbb{I}_+} \mathbf{a}_k \mathbf{a}_k^T \right) \mathbf{a}_l = \mathbf{a}_l^T \mathbf{a}_l = (\mathbf{a}_l^T \mathbf{a}_l)^2 + \sum_{k \neq l} (\mathbf{a}_l^T \mathbf{a}_k)^2,$$

which proves that

$$\forall k \in \mathbb{I}_+, \sum_{j=1}^n \alpha_{j,k}^2 \leq 1. \quad (3.9)$$

For  $j \in \{1, \dots, n\}$ , we have  $\sum_{k \in \mathbb{I}_+} \alpha_{j,k}^2 = 1$  and therefore  $\sum_{k \in \mathbb{I}_+} \sum_{j=1}^n \alpha_{j,k}^2 = n$ . Since  $\lambda_1, \dots, \lambda_n$  are the largest eigenvalues of  $T_\mu$ , we deduce from this equality combined with (3.9) that

$$\sum_{j=1}^n \lambda_j \geq \sum_{k \in \mathbb{I}_+} \sum_{j=1}^n \alpha_{j,k}^2 \lambda_k,$$

i.e.,  $C_I(\mathbf{H}_n^*) \geq C_I(\mathbf{H}_n)$ , which concludes the proof.  $\square$

**3.4. Approximation by truncation.** The spectral representation of the IMSE criterion can be approximated by truncation.

**Definition 3.1.** Let  $\mathbf{H}_D$  be a closed linear subspace of  $\mathbf{H}$  with orthonormal basis  $\{\eta_j, j \in J\}$  and consider the notation of Proposition 3.3. For a subset  $\mathbb{I}_{trc}$  of  $\mathbb{I}_+$ , the (spectral) truncated-IMSE criterion associated with the subspace  $\mathbf{H}_D$  is given by  $\text{IMSE}_{trc}(\mathbf{H}_D) = \tau_{trc} - C_{I_{trc}}(\mathbf{H}_D)$ , with

$$C_{I_{trc}}(\mathbf{H}_D) = \sum_{k \in \mathbb{I}_{trc}} \sum_{j \in J} \alpha_{j,k}^2 \lambda_k \quad (3.10)$$

and where  $\tau_{trc} = \sum_{k \in \mathbb{I}_{trc}} \lambda_k$ .

**Remark 3.2.** We have chosen to use  $\tau_{trc}$  in Definition 3.1 since we interpret the truncated-IMSE criterion as the IMSE when only a subset of the eigenvalues is taken into account. In particular, if we denote by  $\mathbf{H}_{trc}^*$  the closed linear subspace of  $\mathbf{H}$  spanned by the random variables  $\xi_k, k \in \mathbb{I}_{trc}$ , then,

$$\text{IMSE}_{trc}(\mathbf{H}_D) = \int_{\mathcal{X}} \mathbb{E} \left[ \left( P_{\mathbf{H}_{trc}^*} [Z_x - P_{\mathbf{H}_n} [Z_x]] \right)^2 \right] d\mu(x)$$

and, from the property of orthogonal projections in Hilbert space,

$$C_{I_{trc}}(\mathbf{H}_D) = \int_{\mathcal{X}} \mathbb{E} [P_{\mathbf{H}_{trc}^*} [Z_x] P_{\mathbf{H}_D} [Z_x]] d\mu(x).$$

This in particular shows that  $C_{I_{trc}}(\mathbf{H}_D)$  does not depend on the choice of the orthonormal basis  $\{\eta_j, j \in J\}$  of  $\mathbf{H}_D$ .  $\triangleleft$

**Proposition 3.5** (Error induced by truncation). For any closed linear subspace  $\mathbf{H}_D$  of  $\mathbf{H}$  and for any truncation set  $\mathbb{I}_{trc} \subset \mathbb{I}_+$ , we have,

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

so that  $\sum_{k \notin \mathbb{I}_{trc}} \lambda_k$  gives an upper bound on the error induced by truncation.

*Proof.* Consider the notations of Proposition 3.3 and the spectral expansions of  $C_I$  and  $C_{I_{trc}}$  given in (3.5) and (3.10). The left-hand side inequality follows from the positivity of all the  $\alpha_{j,k}^2 \lambda_k$ . For the right-hand side inequality, we just have to remark that, similarly to (3.9), for all  $k \in \mathbb{I}_+$ , we have  $\sum_{j \in J} \alpha_{j,k}^2 \leq 1$ .  $\square$

Classically, the accuracy of the approximation by truncation is quantified through the spectral ratio

$$R_{trc} = \frac{\tau_{trc}}{\tau} = \frac{\sum_{k \in \mathbb{I}_{trc}} \lambda_k}{\sum_{k \in \mathbb{I}_+} \lambda_k}.$$

In practice, we shall consider truncations that only use the  $n_{trc} \in \mathbb{N}^*$  largest eigenvalues of the spectrum and  $n_{trc}$  will be called the *truncation level*.



**Remark 3.3.** Let  $\mathbf{H}_n$  be a closed linear subspace of  $\mathbf{H}_\mu$  with dimension  $n$  and consider the framework of the proof of Proposition 3.4. For a truncation set  $\mathbb{I}_{trc} \subset \mathbb{I}_+$ , since  $\sum_{k \in \mathbb{I}_+} \sum_{j=1}^n \alpha_{j,k}^2 = n$ , we also have the following bound for the error induced by truncation:

$$C_I(\mathbf{H}_n) - C_{I_{trc}}(\mathbf{H}_n) = \sum_{k \notin \mathbb{I}_{trc}} \sum_{j=1}^n \alpha_{j,k}^2 \lambda_k \leq \left( n - \sum_{k \in \mathbb{I}_{trc}} \sum_{j=1}^n \alpha_{j,k}^2 \right) \max_{k \notin \mathbb{I}_{trc}} (\lambda_k). \quad (3.11)$$

Notice that contrary to Proposition 3.5, the bound (3.11) depends on the subspace  $\mathbf{H}_n$  considered.  $\triangleleft$

#### 4. IMSE and design of experiments.

**4.1. Classical approach.** Denote by  $\mathbf{z}$  the (column) random vector

$$\mathbf{z} = (Z_{x_1}, \dots, Z_{x_n})^T,$$

where  $x_1, \dots, x_n \in \mathcal{X}$  and  $n \in \mathbb{N}^*$ . We thus consider the design  $\{x_1, \dots, x_n\}$  and the associated subspace  $\mathbf{H}_{ev}$  of  $\mathbf{H}$  defined by equation (2.3).

**Definition 4.1.** For a fixed  $n \in \mathbb{N}^*$ , a set  $\{x_1, \dots, x_n\}$  of  $n$  points of  $\mathcal{X}$  is an  $n$ -point IMSE-optimal design (for the learning of  $Z$ ) if

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

minimizes the IMSE criterion among all subspaces of  $\mathbf{H}$  based on  $n$  evaluations of the random field  $Z$ .

Let  $\mathbf{k}$  be the (column) vector of functions with components  $K_{x_i}$ ,  $1 \leq i \leq n$ , that is, for  $x \in \mathcal{X}$ ,  $\mathbf{k}(x) = (K_{x_1}(x), \dots, K_{x_n}(x))^T$ . Also, let  $\mathbf{K}$  be the covariance matrix of  $\mathbf{z}$ . For the sake of simplicity, we assume that the random field  $Z$  and  $\{x_1, \dots, x_n\}$  are such that  $\mathbf{K}$  is invertible (a generalized inverse of  $\mathbf{K}$  can be used otherwise in order to express the orthogonal projection  $P_{\mathbf{H}_{ev}}$ ). We classically have (simple kriging predictor),

$$\forall x \in \mathcal{X}, P_{\mathbf{H}_{ev}}[Z_x] = \mathbb{E}(Z_x | Z_{x_1}, \dots, Z_{x_n}) = \mathbf{k}^T(x) \mathbf{K}^{-1} \mathbf{z}.$$

The corresponding IMSE is then  $\text{IMSE}(\mathbf{H}_{ev}) = \tau - C_I(\mathbf{H}_{ev})$ , where

$$C_I(\mathbf{H}_{ev}) = \int_{\mathcal{X}} \mathbf{k}^T(x) \mathbf{K}^{-1} \mathbf{k}(x) d\mu(x). \quad (4.1)$$

Under this form, the computation of the IMSE criterion for the design  $\{x_1, \dots, x_n\}$  requires, from a numerical point of view, the inversion of the matrix  $\mathbf{K}$ , and next the integration of the function  $x \mapsto \mathbf{k}^T(x) \mathbf{K}^{-1} \mathbf{k}(x)$ .

**4.2. Spectral representation and truncation of the IMSE.** For a design of experiments  $\{x_1, \dots, x_n\}$ , we introduce the matrix  $\mathbf{F}$  with entries

$$\mathbf{F}_{i,k} = \lambda_k \phi_k(x_i) = \sqrt{\lambda_k} \mathbb{E}(\xi_k Z_{x_i}), \text{ with } 1 \leq i \leq n \text{ and } k \in \mathbb{I}_+. \quad (4.2)$$

Hence,  $\mathbf{F}$  has  $n$  rows and  $\text{card}(\mathbb{I}_+)$  columns.

**Proposition 4.1.** Let  $\mathbf{H}_{ev} = \text{span} \{Z_{x_1}, \dots, Z_{x_n}\}$ , then

$$C_I(\mathbf{H}_{ev}) = \sum_{k \in \mathbb{I}_+} \left[ (\mathbf{F}_{\cdot,k})^T \mathbf{K}^{-1} (\mathbf{F}_{\cdot,k}) \right] = \text{trace}(\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F}), \quad (4.3)$$

where  $\mathbf{F}_{\cdot,k}$  is the  $k$ -th column of the matrix  $\mathbf{F}$ . If we consider a truncation based on a subset  $\mathbb{I}_{trc} \subset \mathbb{I}_+$ , then

$$C_{I_{trc}}(\mathbf{H}_{ev}) = \sum_{k \in \mathbb{I}_{trc}} \left[ (\mathbf{F}_{\cdot,k})^T \mathbf{K}^{-1} (\mathbf{F}_{\cdot,k}) \right] = \text{trace} \left( (\mathbf{F}_{\cdot, \mathbb{I}_{trc}})^T \mathbf{K}^{-1} (\mathbf{F}_{\cdot, \mathbb{I}_{trc}}) \right), \quad (4.4)$$

where  $\mathbf{F}_{\cdot, \mathbb{I}_{trc}}$  is the matrix with columns given by  $\mathbf{F}_{\cdot,k}$ ,  $k \in \mathbb{I}_{trc}$ .

*Proof.* Expression (4.3) is a direct consequence of Proposition 3.3 applied to the evaluation case. Indeed, let  $\mathbf{K} = \mathbf{C}\mathbf{C}^T$  be the Cholesky decomposition of  $\mathbf{K}$ , then the components of the random vector  $(\eta_1, \dots, \eta_n)^T = \mathbf{C}^{-1}\mathbf{z}$  form an orthonormal basis of  $\mathbf{H}_{ev}$ . Consider the diagonal matrix  $\mathbf{\Lambda} = \text{diag}(\lambda_k, k \in \mathbb{I}_+)$ . With (4.2) and the notations introduced in the proof of Proposition 3.4, we have  $\mathbf{a}_k = (\mathbf{C}^{-1}\mathbf{F}\mathbf{\Lambda}^{-\frac{1}{2}})_{\cdot,k}$ , so that

$$\lambda_k \mathbf{a}_k^T \mathbf{a}_k = \lambda_k \sum_{j=1}^n \alpha_{j,k}^2 = (\mathbf{F}_{\cdot,k})^T \mathbf{K}^{-1} (\mathbf{F}_{\cdot,k}).$$

We hence obtain  $C_I(\mathbf{H}_{ev}) = \sum_{k \in \mathbb{I}_+} \left[ (\mathbf{F}_{\cdot,k})^T \mathbf{K}^{-1} (\mathbf{F}_{\cdot,k}) \right]$ , from which the right-hand side of (4.3) and (4.4) directly follow.  $\square$

Expressions (4.3) and (4.4) indicate that once the functions  $\lambda_k \phi_k(\cdot)$  are known,  $\text{IMSE}(\mathbf{H}_{ev})$  and  $\text{IMSE}_{trc}(\mathbf{H}_{ev})$  can be obtained without explicitly integrating the mean-squared prediction error. This two expressions however involve summations over  $\mathbb{I}_+$  or  $\mathbb{I}_{trc}$ . In addition, notice that the computation of  $\phi_k(x)$  for a general  $x \in \mathcal{X}$  from the eigenpair  $\{\lambda_k, \tilde{\phi}_k\}$  using expression (3.3) requires the computation of an integral (see Remark 4.1).

**Remark 4.1.** Assume that  $\mathcal{X}$  is a topological space (endowed with its Borel  $\sigma$ -algebra) and that the RKHS  $\mathcal{H}$  consists of continuous functions on  $\text{supp}(\mu) \subset \mathcal{X}$  (the support of  $\mu$ ). Then, by choosing a representer for the equivalent class  $\tilde{\phi}_k$  which is continuous on  $\text{supp}(\mu)$  (and denoting  $\tilde{\phi}_k$  this representer), we have  $\tilde{\phi}_k(x) = \phi_k(x)$  for all  $x \in \text{supp}(\mu)$ . It is thus possible to evaluate and optimize the IMSE criterion on  $\text{supp}(\mu)$  by using only the spectral decomposition of the operator  $T_\mu$ .

A situation of particular interest is when the measure  $\mu$  is discrete, which is the case for instance when a quadrature rule is used to approximate the integral of the MSE. This situation is detailed in Section 4.4.  $\triangleleft$

**4.3. Alternative expressions for the IMSE and truncated-IMSE.** Starting from (4.1) and using the property of the trace operator combined with Tonelli's theorem, we obtain

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

where  $\int_{\mathcal{X}} \mathbf{k}(x) \mathbf{k}^T(x) d\mu(x)$  stands for the  $n \times n$  symmetric matrix with  $i, j$  entry

$$\int_{\mathcal{X}} K_{x_i}(x) K_{x_j}(x) d\mu(x) = T_\mu[K_{x_j}](x_i) = T_\mu[K_{x_i}](x_j).$$

We can then introduce the kernel  $\Sigma(\cdot, \cdot)$  on  $\mathcal{X} \times \mathcal{X}$ , with

$$\forall s \text{ and } t \in \mathcal{X}, \Sigma(s, t) = \int_{\mathcal{X}} K_s(x) K_t(x) d\mu(x).$$

Let  $\Sigma$  be the  $n \times n$  matrix with  $i, j$  entry  $\Sigma(x_i, x_j)$ , we have  $\Sigma = \int_{\mathcal{X}} \mathbf{k}(x)\mathbf{k}^T(x)d\mu(x)$  and finally

$$C_I(\mathbf{H}_{ev}) = \text{trace}(\mathbf{K}^{-1}\Sigma). \quad (4.5)$$

In the same way, for the truncated criterion with a fixed truncation set  $\mathbb{I}_{trc}$ , we easily deduce from (4.4) that

$$C_{I_{trc}}(\mathbf{H}_{ev}) = \text{trace}(\mathbf{K}^{-1}\Sigma_{trc}), \quad (4.6)$$

with  $\Sigma_{trc} = \mathbf{F}_{\cdot, \mathbb{I}_{trc}}(\mathbf{F}_{\cdot, \mathbb{I}_{trc}})^T$ . We then introduce the kernel  $\Sigma_{trc}(\cdot, \cdot)$  on  $\mathcal{X} \times \mathcal{X}$ , with

$$\forall s \text{ and } t \in \mathcal{X}, \Sigma_{trc}(s, t) = \sum_{k \in \mathbb{I}_{trc}} \lambda_k^2 \phi_k(s)\phi_k(t),$$

and the  $i, j$  entry of  $\Sigma_{trc}$  is  $\Sigma(x_i, x_j)$ .

From a design optimization perspective (requiring a large number of criterion evaluations), expressions (4.5) and (4.6) appears to be of particular interest when the design space is restricted to finite subset of points in  $\mathcal{X}$ . In this situation, one can indeed compute and store, preliminary to the design optimisation, the value of the (symmetric) kernels  $\Sigma(\cdot, \cdot)$  and  $\Sigma_{trc}(\cdot, \cdot)$  for all the pairs of points belonging to the restricted design space. This way, during the design optimization, any additional calculation is required for the construction of the matrices  $\Sigma$  and  $\Sigma_{trc}$ . The evaluation of the IMSE or truncated-IMSE then mainly consists in the inversion of the matrix  $\mathbf{K}$  (see also Remark 4.2). Such a situation is considered in Section 4.4.2, where the design space is restricted to the set of quadrature points used to approximated the integrated MSE.

**Remark 4.2.** 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 (i.e., element by element) and  $\text{sum}(\cdot)$  means that we sum all the elements of the matrix considered.  $\triangleleft$

**4.4. Quadrature approximation.** Consider the situation where a (pointwise) quadrature rule is used in order to approximate integrals over  $\mathcal{X}$  relatively to  $\mu$ .

**4.4.1. Notations.** For a real-valued function  $f$  on  $\mathcal{X}$ , integrable with respect to  $\mu$ , we consider the following approximation:

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

with  $\omega_j > 0$ ,  $s_j \in \mathcal{X}$  and  $N_q \in \mathbb{N}^*$ . This situation thus corresponds to a particular case of the general problem studied in Section 3, where the measure  $\mu$  on  $\mathcal{X}$  is approximated with the discrete measure

$$\hat{\mu} = \sum_{j=1}^{N_q} \omega_j \delta_{s_j}, \quad (4.7)$$

with  $\delta_{s_j}$  the Dirac measure (evaluation functional) at  $s_j$ . We thus obtain an approximation  $T_{\hat{\mu}}$  of the integral operator  $T_{\mu}$  (Nyström method, see for instance [Hac95,

Kre99]), more precisely,

$$\forall x \in \mathcal{X}, T_{\hat{\mu}}[f](x) = \sum_{j=1}^{N_q} \omega_j K(x, s_j) f(s_j).$$

Throughout the rest of the paper, the hat symbol indicates that the corresponding object is associated with the quadrature approximation  $\hat{\mu}$ . For instance,  $\hat{\lambda}_k$  refers to an eigenvalue of  $T_{\hat{\mu}}$ , whereas  $\lambda_k$  refers to an eigenvalue of  $T_{\mu}$ , *etc.* The results contained in this section can be deduced from those of Section 3.2. However, due to the great importance of quadrature approximations in our study, some details are given below.

We introduce the two  $N_q \times N_q$  matrices  $\mathbf{W} = \text{diag}(\omega_1, \dots, \omega_{N_q})$  and  $\mathbf{Q}$  with  $i, j$  term  $\mathbf{Q}_{i,j} = K(s_i, s_j)$ , for  $1 \leq i, j \leq N_q$ . The matrix  $\mathbf{W}$  is thus the diagonal matrix of quadrature weights and  $\mathbf{Q}$  is the covariance matrix for the quadrature points. We can canonically identify the Hilbert space  $L^2(\mathcal{X}, \hat{\mu})$  with the space  $\mathbb{R}^{N_q}$  endowed with the inner product  $(\cdot | \cdot)_{\mathbf{W}}$ , where for  $\mathbf{x}$  and  $\mathbf{y} \in \mathbb{R}^{N_q}$ ,  $(\mathbf{x} | \mathbf{y})_{\mathbf{W}} = \mathbf{x}^T \mathbf{W} \mathbf{y}$ . To  $f \in L^2(\mathcal{X}, \hat{\mu})$ , we associate the vector  $\mathbf{f} = (f(s_1), \dots, f(s_{N_q}))^T \in \mathbb{R}^{N_q}$  (i.e.,  $\mathbf{f}$  is the column vector with components the values of  $f$  at the quadrature points) and the operator  $T_{\hat{\mu}}$  then corresponds to the matrix  $\mathbf{QW}$ . In matrix notation, we have

$$\forall x \in \mathcal{X}, T_{\hat{\mu}}[f](x) = \mathbf{q}^T(x) \mathbf{W} \mathbf{f},$$

with, for all  $x \in \mathcal{X}$ ,  $\mathbf{q}(x) = (K_{s_1}(x), \dots, K_{s_{N_q}}(x))^T$ .

For the sake of simplicity, we assume that  $\mathbf{Q}$  is nonsingular (if  $\mathbf{Q}$  is singular, then  $\mathbf{QW}$  will have some zero eigenvalues which can be ignored, see Section 3.3). We denote by  $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_{N_q} > 0$  the eigenvalues of the matrix  $\mathbf{QW}$  and by  $\mathbf{v}_1, \dots, \mathbf{v}_{N_q}$  their associated eigenvectors, i.e.,  $\mathbf{QW} = \mathbf{P} \hat{\Lambda} \mathbf{P}^{-1}$  with  $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_{N_q})$  and  $\mathbf{P} = (\mathbf{v}_1 | \dots | \mathbf{v}_{N_q})$ . The set of vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_{N_q}\}$  forms an orthonormal basis of  $\{\mathbb{R}^{N_q}, (\cdot | \cdot)_{\mathbf{W}}\}$ , so that

$$\mathbf{P}^T \mathbf{W} \mathbf{P} = \text{Id}_{N_q \times N_q},$$

the  $N_q \times N_q$  identity matrix.

**Lemma 4.1.** For  $1 \leq k \leq N_q$ , consider the functions defined by

$$\forall x \in \mathcal{X}, \hat{\phi}_k(x) = \mathbf{q}^T(x) \mathbf{Q}^{-1} \mathbf{v}_k = \frac{1}{\hat{\lambda}_k} \mathbf{q}^T(x) \mathbf{W} \mathbf{v}_k. \quad (4.8)$$

Then  $\hat{\phi}_k$  is an eigenfunction of  $T_{\hat{\mu}}$  associated with  $\hat{\lambda}_k$  and  $\hat{\phi}_k \in \mathcal{H}_{\hat{\mu}}$ , the closed linear subspace of  $\mathcal{H}$  spanned by the  $K_{s_j}$ ,  $1 \leq j \leq N_q$ . In addition, the  $\hat{\phi}_k$ , with  $1 \leq k \leq N_q$ , are orthogonal in  $\mathcal{H}$  and  $\|\hat{\phi}_k\|_{\mathcal{H}}^2 = 1/\hat{\lambda}_k$ .

The proof is detailed in Appendix A. Equation (4.8) is the equivalent of expression (3.3) in the quadrature case. Therefore, as mentioned at the end of Section 4.2, for  $1 \leq k \leq N_q$ , the computation of  $\hat{\phi}_k(x)$  for a general  $x \in \mathcal{X}$  requires the computation of an integral (here, of a quadrature). However, the situation is much simpler when  $x$  is a quadrature point. In that case, we have, for all  $1 \leq j, k \leq N_q$ ,

$$\hat{\phi}_k(s_j) = (\mathbf{v}_k)_j, \quad (4.9)$$

the  $j$ -th component of the eigenvector  $\mathbf{v}_k$ . It is therefore possible to optimize the  $\hat{\mu}$ -IMSE and truncated- $\hat{\mu}$ -IMSE on the support of  $\hat{\mu}$  (that is, on quadrature points) using only the matrices  $\mathbf{Q}$  and  $\mathbf{W}$  and the spectral decomposition of  $\mathbf{QW}$ . This is considered below. Notice that the restriction of the design space to a finite subset of points is also motivated by the considerations made in Section 4.3.

**4.4.2. Quadrature-designs.** We suppose that the design space is restricted to subsets of quadrature points and we keep the same notations as Section 4.4.1.

**Definition 4.2.** We call quadrature-design a design of experiments 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  $I_q = \{i_1, \dots, i_n\}$  of  $\{1, \dots, N_q\}$ .

For a quadrature-design with index set  $I_q$ , we denote by  $\mathbf{H}_{ev}^q$  the associated Gaussian subspace. From (4.9), the approximation  $\hat{\mathbf{F}}$  of the matrix  $\mathbf{F}$  defined in equation (4.2) is given by  $\hat{\mathbf{F}} = (\mathbf{P}\hat{\mathbf{\Lambda}})_{I_q, \cdot}$ , i.e.  $\hat{\mathbf{F}}$  consists of the rows of  $\mathbf{P}\hat{\mathbf{\Lambda}}$  having indices in  $I_q$ .

We then obtain the following expressions for the quadrature approximation of the IMSE criterion:

$$\hat{C}_I(\mathbf{H}_{ev}^q) = \text{trace}(\mathbf{W}\mathbf{Q}_{\cdot, I_q}\mathbf{K}^{-1}\mathbf{Q}_{I_q, \cdot}) \quad (4.10)$$

$$= \text{trace}(\hat{\mathbf{F}}^T\mathbf{K}^{-1}\hat{\mathbf{F}}) \quad (4.11)$$

where  $\mathbf{K} = \mathbf{Q}_{I_q, I_q}$  is the covariance matrix of the considered quadrature-design ( $\mathbf{K}$  is a submatrix of  $\mathbf{Q}$ ). The right-hand side of (4.10) follows from the integral form of the IMSE and (4.11) is its spectral representation. For the truncated criterion associated with a truncation set  $\hat{\mathbb{I}}_{trc}$  (quadrature approximation), we have, from (4.11),

$$\hat{C}_{I_{trc}}(\mathbf{H}_{ev}^q) = \text{trace}((\hat{\mathbf{F}}_{\cdot, \hat{\mathbb{I}}_{trc}})^T\mathbf{K}^{-1}(\hat{\mathbf{F}}_{\cdot, \hat{\mathbb{I}}_{trc}})). \quad (4.12)$$

Following Section 4.3, we can introduce the two  $N_q \times N_q$  matrices (these matrices are computed once for all, before the design optimization)

$$\mathbf{\Omega} = \mathbf{QWQ} \text{ and } \mathbf{\Omega}_{trc} = (\mathbf{P}\hat{\mathbf{\Lambda}})_{\cdot, \hat{\mathbb{I}}_{trc}}((\mathbf{P}\hat{\mathbf{\Lambda}})_{\cdot, \hat{\mathbb{I}}_{trc}})^T,$$

and we finally obtain

$$\hat{C}_I(\mathbf{H}_{ev}^q) = \text{trace}(\mathbf{K}^{-1}\hat{\mathbf{\Sigma}}) \text{ and } \hat{C}_{I_{trc}}(\mathbf{H}_{ev}^q) = \text{trace}(\mathbf{K}^{-1}\hat{\mathbf{\Sigma}}_{trc}), \quad (4.13)$$

with  $\hat{\mathbf{\Sigma}} = (\mathbf{\Omega})_{I_q, I_q}$  and  $\hat{\mathbf{\Sigma}}_{trc} = (\mathbf{\Omega}_{trc})_{I_q, I_q}$ . This way, the IMSE or truncated-IMSE criterion can be easily evaluated for any quadrature-design, making global optimization affordable. This is illustrated on an example with  $N_q = 5000$  in Section 5.3.

**5. Examples.** This section presents some examples of construction of IMSE-optimal designs using quadrature approximation and spectral truncation. All computations have been performed with the free softwares **R** and **Sage** [R C13, S+13]. The objective is to assess the impact of the restriction of the design space to quadrature points and to illustrate the influence of the truncation level  $n_{trc}$  on optimal designs. The last example illustrates the computational cost of the IMSE or truncated-IMSE in the framework of Section 4.4.2.

### 5.1. A one-point design augmentation problem.

**5.1.1. Kernel of exponential type.** Consider a centered Gaussian process  $Z$  on  $\mathbb{R}$  with covariance

$$\forall x \text{ and } y \in \mathbb{R}, K(x, y) = e^{-|x-y|} - e^{-(|x|+|y|)}.$$

This corresponds to the covariance of a centered Ornstein-Uhlenbeck process on  $\mathbb{R}$  conditioned to vanish at 0 (that is, after one observation at 0). We take  $\mu$  as the uniform probability measure on  $[0, 1]$ . We have

$$\tau = \int_0^1 1 - e^{-2x} dx = \frac{1}{2} (1 + e^{-2}) \approx 0.5676676.$$

We approximate  $\mu$  with a 500-point quadrature, where  $s_j = (2j - 1)/1000$ , with  $1 \leq j \leq 500$ , each point receiving weight  $1/500$  (mid-point rectangular quadrature method with intervals of length 0.002). We denote by  $\hat{\mu}$  the corresponding discrete measure. A numerical evaluation gives  $\hat{\tau} = 0.5676679$ .

For a design point  $t \in \mathbb{R}$ , we introduce  $\mathbf{H}_t = \text{span}\{Z_t\}$ . For  $t \in (0, 1]$ , we have

$$C_I(\mathbf{H}_t) = \frac{e^{-2t}}{1 - e^{-2t}} \left( \frac{e^{2t} - e^{-2t}}{2} - 2t \right) + (e^{2t} - 1) \frac{e^{-2t} - e^{-2}}{2}.$$

This function reaches its maximum at  $t^* \approx 0.707859$ , the 1-point  $\mu$ -IMSE-optimal design for this problem (without quadrature approximation). This is illustrated in the left part of Figure 5.1 where the quadrature approximation  $\hat{C}_I(\mathbf{H}_t)$  of  $C_I(\mathbf{H}_t)$  is also presented. In particular, we observe that the 1-point  $\hat{\mu}$ -IMSE optimal design, denoted by  $\hat{t}^*$ , coincides with  $\hat{t}^{q*}$ , the 1-point  $\hat{\mu}$ -IMSE optimal quadrature-design. The right part of Figure 5.1 indicates that this property is also verified by the quadrature approximation of the truncated criteria: we have  $\hat{t}_{n_{trc}}^* = \hat{t}_{n_{trc}}^{q*}$  for any truncation level  $n_{trc}$ , where  $\hat{t}_{n_{trc}}^*$  refers to the optimal design and  $\hat{t}_{n_{trc}}^{q*}$  to the optimal quadrature-design. The restriction to quadrature points has therefore no impact in this particular case (see also Remark 5.1).

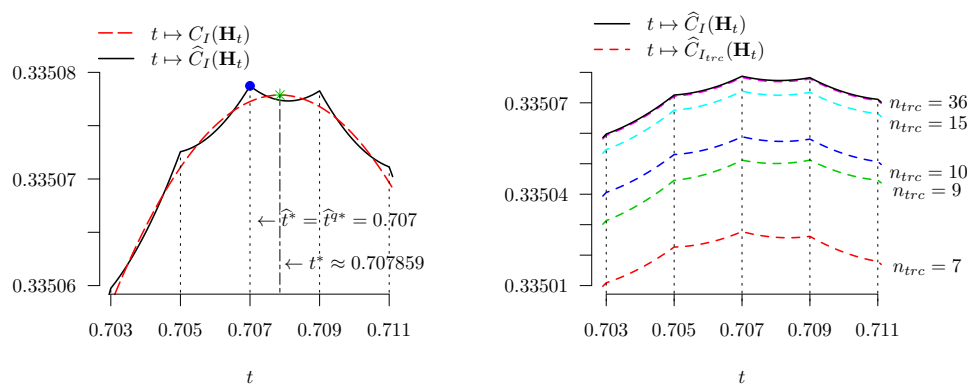


FIGURE 5.1. Graphs of  $t \mapsto C_I(\mathbf{H}_t)$  and  $t \mapsto \hat{C}_I(\mathbf{H}_t)$  for  $0.703 \leq t \leq 0.711$  (left) and graph of  $t \mapsto \hat{C}_{I_{trc}}(\mathbf{H}_t)$  for various truncation level  $n_{trc}$  (right), the quadrature points are indicated on the horizontal axis of both plots.

Figure 5.2 shows the designs  $\hat{t}_{n_{trc}}^*$  optimal for  $\hat{C}_{I_{trc}}$  as a function of the number  $n_{trc}$  of eigenvalues retained (for instance, for  $n_{trc} = 1$ ,  $\hat{t}_{n_{trc}}^* = 0.695$ ); notice that

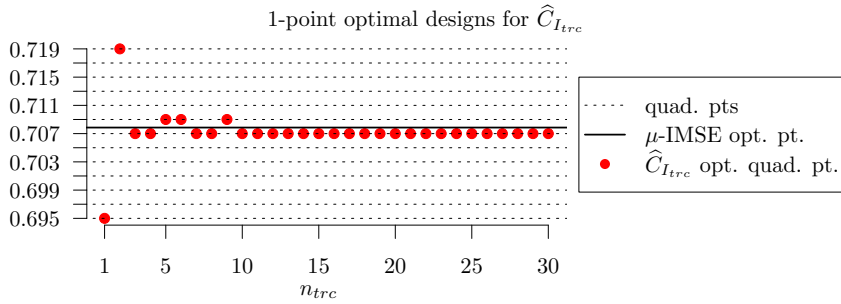


FIGURE 5.2. Representation of the 1-point optimal designs (quadrature-designs) for  $\widehat{C}_{I_{trc}}$  as a function of  $n_{trc}$ .

TABLE 5.1  
Value, in percent, of the spectral ratio  $\widehat{R}_{trc}$  for various  $n_{trc}$  (Section 5.1.1).

$n_{trc}$	1	2	3	7	9	10	15	36
$\widehat{R}_{trc}$ (%)	68.87	82.88	88.33	94.92	96.04	96.44	97.62	99.01

all those designs are quadrature-designs. A fast convergence of  $\widehat{t}_{n_{trc}}^*$  to the  $\widehat{\mu}$ -IMSE-optimal design  $\widehat{t}^* = \widehat{t}_{N_q}^* = 0.707$  as  $n_{trc}$  increases is observed (a similar convergence can also be observed for the 2-, 3- and 4-point design problems). For  $3 \leq n_{trc} \leq 9$ ,  $\widehat{t}_{n_{trc}}^*$  oscillates between 0.707 and 0.709, the two quadrature points closest to the  $\mu$ -IMSE optimal design  $t^*$ . For  $n_{trc} \geq 10$ ,  $\widehat{t}_{n_{trc}}^*$  coincides with  $\widehat{t}^* = \widehat{t}^{q*}$ . Table 5.1 gives the values, in percent, of the spectral ratio (quadrature approximation) for various truncation levels  $n_{trc}$ .

**Remark 5.1** (Ornstein-Uhlenbeck process and quadrature approximation). For a one-dimensional centered Ornstein-Uhlenbeck process on  $\mathbb{R}$  and a quadrature  $\widehat{\mu}$ , by considering the first and second derivative of the continuous function  $t \mapsto \widehat{C}_I(\mathbf{H}_t)$  (these derivatives are defined whenever  $t$  is not a quadrature point), one can easily prove that one-point  $\widehat{\mu}$ -IMSE optimal designs are always quadrature-designs.  $\triangleleft$

**5.1.2. Kernel of squared-exponential (Gaussian) type.** We keep the same notations and settings as in Section 5.1.1 but we now assume that the centered Gaussian process  $Z$  admits the following covariance on  $\mathbb{R}$ ,

$$\forall x \text{ and } y \in \mathbb{R}, K(x, y) = e^{-(x-y)^2} - e^{-(x^2+y^2)}.$$

Figure 5.3 shows the values of  $\widehat{C}_I(\mathbf{H}_t)$  and  $\widehat{C}_{I_{trc}}(\mathbf{H}_t)$ , with  $n_{trc} \in \{1, 2, 3\}$ , as functions of the design point  $t$ . The corresponding spectral ratios are reported on the figure.

For any truncation level  $n_{trc}$ , the optimal 1-point quadrature-design is  $\widehat{t}_{n_{trc}}^{q*} = \widehat{t}^{q*} = 0.719$ . For  $n_{trc} = 1$  (bottom curve), the optimal design is  $\widehat{t}_1^* \approx 0.718672$  and for  $n_{trc} \geq 2$ , we have  $\widehat{t}_{n_{trc}}^* \approx 0.718836 \approx \widehat{t}^*$ . One may note that in this example, the 1-point optimal designs for  $\widehat{\mu}$ -IMSE $_{trc}$  are not supported by the quadrature (i.e., are not quadrature-designs). However, we have  $\widehat{C}_I(\mathbf{H}_{\widehat{t}^*}) \approx 0.3813078$  and  $\widehat{C}_I(\mathbf{H}_{\widehat{t}^*}) - \widehat{C}_I(\mathbf{H}_{\widehat{t}^{q*}}) \approx 8.258\text{e-}09$ , so that the error induced by the restriction to quadrature-designs is marginal.

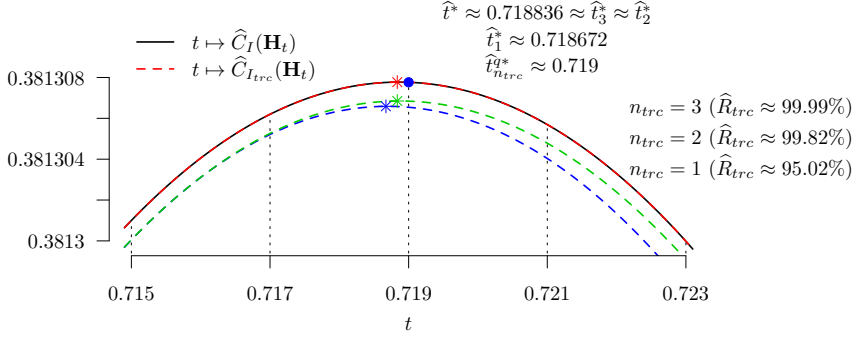


FIGURE 5.3. Graphs of  $t \mapsto \widehat{C}_I(\mathbf{H}_t)$  and  $t \mapsto \widehat{C}_{I_{trc}}(\mathbf{H}_t)$  (with  $n_{trc} \in \{1, 2, 3\}$ ) around their maximum, i.e., for  $0.715 \leq t \leq 0.723$ .

TABLE 5.2  
Value, in percent, of the spectral ratio  $\widehat{R}_{trc}$  for various  $n_{trc}$  (Section 5.2).

$n_{trc}$	1	2	3	4	5	6	12
$\widehat{R}_{trc}$ (%)	74.81	85.72	96.63	98.23	98.96	99.70	99.99

**5.2. Gaussian kernel on the unit square.** Consider now a centered Gaussian process  $Z$  on  $\mathbb{R}^2$  with covariance (Gaussian, or squared-exponential) kernel,

$$\forall x \text{ and } y \in \mathbb{R}^2, K(x, y) = e^{-\|x-y\|^2},$$

where  $\|\cdot\|$  is the Euclidean norm of  $\mathbb{R}^2$ . We take  $\mu$  as the uniform probability on  $[0, 1]^2$  (so that  $\tau = 1$ ).

We approximate integrals over  $[0, 1]^2$  through a quadrature consisting of a regular grid of  $33 \times 33$  points, all points receiving the same weight (mid-point rectangular quadrature rule). The corresponding discrete measure is denoted by  $\widehat{\mu}$ . Table 5.2 gives the spectral ratios  $\widehat{R}_{trc}$  (in percent) for various truncation levels  $n_{trc}$ .

Figure 5.4 shows the quadrature-designs  $\widehat{X}_n^{q*}$  and  $\widehat{X}_{n, n_{trc}}^{q*}$  respectively optimal for  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$ , with  $n_{trc} = n$  (the design size) for  $n = 4$  and  $n = 5$ ;  $\widehat{X}_n^{q*}$  and  $\widehat{X}_{n, n}^{q*}$  coincide for  $n = 4$  but not for  $n = 5$ . We numerically observe that the 5-point quadrature-designs optimal for  $\widehat{C}_{I_{trc}}$  with  $n_{trc} \geq 6$  are the  $\widehat{\mu}$ -IMSE optimal quadrature-design  $\widehat{X}_5^{q*}$ . The right part of Figure 5.4 illustrates in particular how  $\widehat{C}_{I_{trc}}(\widehat{X}_5^{q*})$  tends to  $\widehat{C}_I(\widehat{X}_5^{q*})$  (dashed line on the top) when  $n_{trc}$  increases. Notice that  $\widehat{C}_{I_{trc}}(\widehat{X}_5^{q*})$  is an increasing function of  $n_{trc}$  and that for  $n_{trc} = 6$ ,  $\widehat{C}_{I_{trc}}(\widehat{X}_5^{q*}) \approx 0.9890146$ , which is already very close of  $\widehat{C}_I(\widehat{X}_5^{q*}) \approx 0.9890174$ .

The 4- and 5-point  $\widehat{\mu}$ -IMSE optimal designs  $\widehat{X}_4^*$  and  $\widehat{X}_5^*$  (with quadrature approximation but without restriction to quadrature-designs) are not supported by the quadrature points. However, for the 4-point problem, we obtain

$$\widehat{C}_{I_{trc}}(\widehat{X}_4^*) \approx 0.9815098 \text{ and } \widehat{C}_{I_{trc}}(\widehat{X}_4^*) - \widehat{C}_{I_{trc}}(\widehat{X}_4^{q*}) \approx 2.258776\text{e-}05,$$

and for the 5-point problem, we have

$$\widehat{C}_{I_{trc}}(\widehat{X}_5^*) \approx 0.9890199 \text{ and } \widehat{C}_{I_{trc}}(\widehat{X}_5^*) - \widehat{C}_{I_{trc}}(\widehat{X}_5^{q*}) \approx 2.479523\text{e-}06.$$



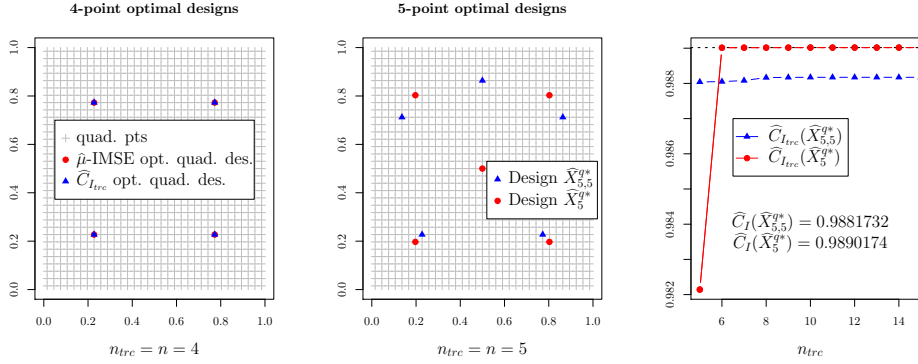


FIGURE 5.4. Quadrature-designs optimal for the criteria  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$  for  $n = n_{trc} = 4$  (left) and  $n = n_{trc} = 5$  (middle). Values of the criterion  $\widehat{C}_{I_{trc}}$  for the two quadrature-designs  $\widehat{X}_5^{q*}$  and  $\widehat{X}_{5,5}^{q*}$  (respectively optimal for  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$  with  $n_{trc} = 5$ ) as function of the number  $n_{trc}$  of considered eigenvalues (right).

In both cases, the error induced by the restriction of the design problem to quadrature-designs is therefore negligible.

**5.3. Numerical experiments in dimension 6.** Consider the 6-dimensional tensor-product Matérn covariance kernel on  $\mathbb{R}^6$ , that is  $K(x, y) = \prod_{i=1}^6 K_{\theta_i}(x_i, y_i)$ , where  $x = (x_1, \dots, x_6)$  and  $y = (y_1, \dots, y_6)$  are in  $\mathbb{R}^6$  and where the kernels  $K_{\theta_i}(\cdot, \cdot)$  are given by  $K_{\theta_i}(x_i, y_i) = (1 + \sqrt{3}|x_i - y_i|/\theta_i) \exp(-\sqrt{3}|x_i - y_i|/\theta_i)$ , with  $\theta_i > 0$ .

We set  $(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) = (0.32, 0.52, 0.62, 0.52, 0.42, 0.62)$  and we take  $\mu$  as the uniform probability measure on  $[0, 1]^6$ . The use of regular grids to approximate integrals in high dimensional spaces is prohibitive and we consider a quasi Monte-Carlo quadrature with  $N_q = 5000$  points obtained from a uniform Halton sequence (see, e.g., [Nie92]), each point receiving the same weight  $1/N_q$ . The computations have been performed with R-64bit on a 2012 Macbook Air fitted with a 1.8GHz Intel Core i5 processor and 4Gb RAM.

The low discrepancy grid is generated with the R function `runif.halton()`. Using the function `eigen()`, the eigendecomposition of  $\mathbf{QW}$  takes approximately 4 minutes (with a  $\mathcal{O}(N_q^3)$  complexity). The computation of the matrix  $\mathbf{\Omega} = \mathbf{QWQ}$  requires approximately 2 minutes. The IMSE and truncated-IMSE are encoded following (4.13) and Remark 4.2. The design covariance matrix  $\mathbf{K}$  is inverted using the function `solve()`. The IMSE criterion is for instance encoded as follows (indicated for reproducibility of the test):

```
IMSE<-function(Iq){ TAU=sum(solve(MatQ[Iq,Iq])*Mat0[Iq,Iq]) }
```

where `Iq` is the index set  $I_q$  of a quadrature-design, `MatQ` and `Mat0` stand for the matrices  $\mathbf{Q}$  and  $\mathbf{\Omega}$ , and `TAU` is the trace term  $\tau$  (here,  $\tau = 1$ ).

Table 5.3 indicates the median duration, over 1000 repetitions, of one evaluation of the function `IMSE()` at some random quadrature-design and for various design sizes (we use the function `microbenchmark()`). The median duration for the inversion of the matrix  $\mathbf{K}$  is also indicated. As expected, we observe that once the preliminary computations are done (that is, for the IMSE, the computation of the matrices  $\mathbf{Q}$  and  $\mathbf{\Omega}$ ), the evaluation of the IMSE for any quadrature-design mainly consists in the inversion of the covariance matrix of the design.

TABLE 5.3

Median duration (over 1000 repetitions, random quadrature-designs), in seconds, for one evaluation of the function  $IMSE()$  for various design sizes (6-dimensional example) and median duration for the inversion of the design covariance matrix.

design size	10-point	30-point	50-point	70-point	100-point
$IMSE()$	46.47e-6	131.27e-6	439.11e-6	1.128e-3	3.033e-3
$\mathbf{K}^{-1}$	37.31e-6	114.78e-6	369.21e-6	971.80e-6	2.694e-3

**6. Concluding remarks.** We have described how the IMSE criterion can be approximated by spectral truncation. When a (pointwise) quadrature is used to integrate the MSE and the design space is restricted to subsets of quadrature points, we have detailed a numerically efficient strategy for computing the IMSE and truncated-IMSE. Obviously, since preliminary calculations are required, the approach is only of numerical interest for a large enough number of criterion evaluations, which is in particular the case from a design optimization perspective. A simulated-annealing algorithm for the computation of IMSE optimal designs that takes advantage of these considerations is presented in [GP14].

In its present form, the approach only applies to random fields with known mean. The extension to kernel-based interpolation models including an unknown parametric trend would enlarge the spectrum of potential applications and is under current investigation. Also, note that the choice of a suitable quadrature takes a special importance here since quadrature-designs are subsets of quadrature points. The study of the errors induced by restricting the optimization to quadrature-designs and by approximating the exact criteria  $C_I$  and  $C_{I_{trc}}$  by their quadrature approximations  $\hat{C}_I$  and  $\hat{C}_{I_{trc}}$  should deserve further studies.

The interest of optimizing the truncated-IMSE (with appropriated truncation level) instead of the IMSE needs to be investigated in more details. For instance, numerical experiments (see [GP14]) indicate that for truncation levels slightly larger than the design size, the truncated criterion appears as more easy to optimize than the original IMSE, while leading to designs with high IMSE efficiency. The construction of optimal designs for the truncated-IMSE criterion should also be related with the approach of [SP10] (optimal designs for a Bayesian linear model based on the main eigenfunctions of the spectral decomposition), and this connection deserves further investigations.

In the framework considered in Sections 4 and 5, the IMSE, which corresponds to the integral of the kriging variance, is widely acknowledged as a most sensible criterion for choosing observation sites in Gaussian process models. However, in general it is not used for optimal design because it seems complicated (numerically costly) to evaluate. We hope that the present paper will contribute to popularize the use of this criterion to quantify the prediction uncertainty attached to a given design.

**Acknowledgments.** This work was supported by the the project ANR-2011-IS01-001-01 DESIRE (DESIGns for spatial Random fIELds), joint with the Statistics Departement of the Johannes Kepler Universität, Linz (Austria).

The authors want to thank the two anonymous reviewers for their valuable comments and suggestions that significantly improved the presentation.

## Appendix A. Proofs of some lemmas and propositions.

*Proof of Lemma 3.1.* From assumption C-i, the representation property (2.1), the

Cauchy-Schwarz inequality in  $\mathcal{H}$  and C-iii, we have, for all  $h \in \mathcal{H}$ ,

$$\|h\|_{L^2}^2 = \int_{\mathcal{X}} (h|K_t)_{\mathcal{H}}^2 d\mu(t) \leq \|h\|_{\mathcal{H}}^2 \int_{\mathcal{X}} K(t,t) d\mu(t) = \tau \|h\|_{\mathcal{H}}^2,$$

the integral of  $h^2$  being well-defined as the integral of a positive measurable function (see for instance [Dud02]).  $\square$

*Proof of Lemma 3.2.* The reproducing property of  $K(\cdot, \cdot)$  and the Cauchy-Schwarz inequality imply that, for all  $x$  and  $y \in \mathcal{X}$ ,

$$K(x, y) = (K_x|K_y)_{\mathcal{H}} \leq \|K_x\|_{\mathcal{H}} \|K_y\|_{\mathcal{H}} = \sqrt{K(x, x)} \sqrt{K(y, y)}.$$

Combining this with conditions C-ii and C-iii, we obtain

$$\int_{\mathcal{X}} \int_{\mathcal{X}} K(x, y)^2 d\mu(x) d\mu(y) \leq \left( \int_{\mathcal{X}} K(x, x) d\mu(x) \right)^2 = \tau^2.$$

Let  $\{e_i, i \in \mathbb{I}\}$  be an orthonormal basis of  $L^2(\mathcal{X}, \mu)$  (with  $\mathbb{I}$  a general index set, not necessarily countable), we have

$$\begin{aligned} \int_{\mathcal{X}} \int_{\mathcal{X}} K(x, y)^2 d\mu(y) d\mu(x) &= \int_{\mathcal{X}} \|K_x\|_{L^2}^2 d\mu(x) \\ &= \int_{\mathcal{X}} \sum_{i \in \mathbb{I}} (K_x|e_i)_{L^2}^2 d\mu(x) = \sum_{i \in \mathbb{I}} \|T_{\mu}[e_i]\|_{L^2}^2 \leq \tau^2, \end{aligned} \quad (\text{A.1})$$

the interchange between the sum and integral being justified by Tonelli's theorem. So, the operator  $T_{\mu}$  is a Hilbert-Schmidt operator on  $L^2(\mathcal{X}, \mu)$  and  $T_{\mu}$  is thus a compact operator on  $L^2(\mathcal{X}, \mu)$  (in particular, the number of terms different from 0 in the sum on the right-hand side of (A.1) is at most countable). Finally, from the properties of symmetry and positivity of  $K(\cdot, \cdot)$ , we have, for  $f$  and  $g \in L^2(\mathcal{X}, \mu)$ ,

$$(f|T_{\mu}[g])_{L^2} = (T_{\mu}[f]|g)_{L^2} \text{ and } (f|T_{\mu}[f])_{L^2} \geq 0,$$

so that  $T_{\mu}$  is self-adjoint and positive on  $L^2(\mathcal{X}, \mu)$ .  $\square$

*Proof of Proposition 3.1.* First,  $\mathcal{H}_0$  is well-defined thanks to Lemma 3.1. Also note that, as an orthogonal subspace,  $\mathcal{H}_{\mu}$  is by definition closed in  $\mathcal{H}$ . For a fixed  $f \in L^2(\mathcal{X}, \mu)$ , we consider the linear functional  $I_{f, \mu}$  on  $\mathcal{H}$  defined by,

$$\forall h \in \mathcal{H}, I_{f, \mu}(h) = \int_{\mathcal{X}} f(t)h(t) d\mu(t).$$

Again,  $I_{f, \mu}$  is well-defined thanks to Lemma 3.1. From the Cauchy-Schwarz inequality and (3.2), we have

$$|I_{f, \mu}(h)| \leq \|f\|_{L^2} \|h\|_{L^2} \leq \sqrt{\tau} \|f\|_{L^2} \|h\|_{\mathcal{H}},$$

so that the application  $I_{f, \mu}$  is continuous on  $\mathcal{H}$ . Thus, from the Riesz-Fréchet Theorem, there exists a unique element  $\rho_{f, \mu}$  of  $\mathcal{H}$  such that  $I_{f, \mu}(h) = (h|\rho_{f, \mu})_{\mathcal{H}}$ . One can finally identify  $\rho_{f, \mu}$  with  $T_{\mu}[f]$  thanks to

$$I_{f, \mu}(h) = (h|f)_{L^2} = (h|T_{\mu}[f])_{\mathcal{H}}. \quad (\text{A.2})$$

(see [GB12] for more details). Equation (A.2) proves in particular that, for all  $f \in L^2(\mathcal{X}, \mu)$ ,  $T_\mu[f] \in \mathcal{H}_0^{\perp \mathcal{H}} = \mathcal{H}_\mu$ , since we have (using the Cauchy-Schwarz inequality)

$$\forall f \in L^2(\mathcal{X}, \mu), \forall h_0 \in \mathcal{H}_0, |(h_0|T_\mu[f])_{\mathcal{H}}| = |(h_0|f)_{L^2}| \leq \|h_0\|_{L^2} \|f\|_{L^2} = 0.$$

Denote by  $\text{null}(I_{f,\mu})$  the null space of  $I_{f,\mu}$  (which is closed in  $\mathcal{H}$  as the null space of a continuous linear application). We then remark that

$$\mathcal{H}_0 = \bigcap_{f \in L^2(\mathcal{X}, \mu)} \text{null}(I_{f,\mu}),$$

so that  $\mathcal{H}_0$  is closed in  $\mathcal{H}$  (and in particular,  $\mathcal{H}_\mu^{\perp \mathcal{H}} = \mathcal{H}_0$ ).

Now, let  $k$  and  $l \in \mathbb{I}_+$  and denote by  $\delta_{kl}$  the Kronecker delta, we have

$$(\phi_k|\phi_l)_{\mathcal{H}} = \frac{1}{\lambda_k \lambda_l} \int_{\mathcal{X}} \int_{\mathcal{X}} \tilde{\phi}_k(x) \tilde{\phi}_l(t) K(x,t) d\mu(x) d\mu(t) = \frac{\lambda_k}{\lambda_k \lambda_l} \delta_{kl},$$

so that  $\{\sqrt{\lambda_k} \phi_k, k \in \mathbb{I}_+\}$  is an orthonormal system in  $\mathcal{H}_\mu$ .

To conclude, suppose that  $h \in \mathcal{H}$  is such that  $(h|\phi_k)_{\mathcal{H}} = 0$  for all  $k \in \mathbb{I}_+$ , then  $T_\mu[h] = 0$ . Since, from equation (A.2),  $\|h\|_{L^2}^2 = (h|T_\mu[h])_{\mathcal{H}}$ , we obtain that  $h \in \mathcal{H}_0$  and finally that  $\text{span}\{\phi_k, k \in \mathbb{I}_+\}$  is dense in  $\mathcal{H}_\mu$ .  $\square$

*Proof of Lemma 4.1.* The expression  $\mathcal{H}_{\hat{\mu}} = \text{span}\{K_{s_j}, 1 \leq j \leq N_q\}$  follows directly from the definition of  $\hat{\mu}$  given in (4.7) (in particular because the support of  $\hat{\mu}$  is a finite set). By construction, we have  $\hat{\phi}_k \in \mathcal{H}_{\hat{\mu}}$  for all  $1 \leq k \leq N_q$  and

$$\forall x \in \mathcal{X}, T_{\hat{\mu}}[\hat{\phi}_k](x) = \mathbf{q}^T(x) \mathbf{W} \mathbf{Q} \mathbf{Q}^{-1} \mathbf{v}_k = \mathbf{q}^T(x) \mathbf{Q}^{-1} \mathbf{Q} \mathbf{W} \mathbf{v}_k = \hat{\lambda}_k \hat{\phi}_k(x).$$

Finally, since  $(\mathbf{q}|\mathbf{q}^T)_{\mathcal{H}} = \mathbf{Q}$  (matrix notation), we have

$$\|\hat{\phi}_k\|_{\mathcal{H}}^2 = \mathbf{v}_k^T \mathbf{Q}^{-1} \mathbf{Q} \mathbf{Q}^{-1} \mathbf{v}_k = \mathbf{v}_k^T \mathbf{W} \mathbf{W}^{-1} \mathbf{Q}^{-1} \mathbf{v}_k = \frac{1}{\hat{\lambda}_k} \mathbf{v}_k^T \mathbf{W} \mathbf{v}_k = \frac{1}{\hat{\lambda}_k}$$

and the orthogonality of the  $\hat{\phi}_k$  can be obtained with similar arguments.  $\square$

## Appendix B. Some technical remarks.

**Remark B.1.** To ensure that  $C_I(\mathbf{H}_D)$  is well-defined, we have to check the measurability of the function  $x \in \mathcal{X} \mapsto \mathbb{E}[(P_{\mathbf{H}_D}[Z_x])^2]$ . From the isometry between  $\mathbf{H}_D$  and  $\mathcal{H}_D$ , if  $\{h_j, j \in J\}$  is an orthonormal basis of  $\mathcal{H}_D$ , then we have, for all  $x \in \mathcal{X}$ ,

$$0 \leq \mathbb{E}[(P_{\mathbf{H}_D}[Z_x])^2] = \sum_{j \in J} h_j^2(x) \leq K(x, x).$$

The function  $x \mapsto \sum_{j \in J} h_j^2(x)$  is well-defined and is measurable as an at most countable sum of (positive) measurable functions (since  $\mathbf{H}$  is separable). Finally  $C_I(\mathbf{H}_D)$  is well-defined as the integral of a positive measurable function. Note that a similar reasoning with  $\mathbf{H}_D = \mathbf{H}$  shows that  $\int_{\mathcal{X}} K(x, x) d\mu(x)$  is well-defined assuming C-i (and  $\mathbf{H}$  separable).  $\triangleleft$

**Remark B.2.** The results of this article can be extended to a non separable Gaussian Hilbert space  $\mathbf{H}$ . However, in this case condition C-i is not sufficient to ensure the measurability of the function  $x \mapsto \mathbb{E}[(P_{\mathbf{H}_D}[Z_x])^2]$  for a non-separable subspace  $\mathbf{H}_D$ . We then have to assume that  $x \mapsto K(x, x)$  is measurable (see for instance [For85]) and

also, either restrict the definition of the IMSE to separable closed linear subspaces of  $\mathbf{H}$  or assume that  $x \mapsto \mathbb{E}[(P_{\mathbf{H}_D}[Z_x])^2]$  is measurable whatever  $\mathbf{H}_D$ .

Note that the separability assumption for  $\mathbf{H}$  is not very restrictive for most practical situations. Indeed, from the structure theorem for Gaussian measures and the theory of abstract Wiener spaces, this assumption is satisfied by all random fields with sample paths in classical functions spaces, such as Banach or Fréchet spaces, see for instance [Sat69, DFCL71, Bor76].  $\triangleleft$

#### REFERENCES

- [ABM12] Y. Auffray, P. Barbillon, and J.-M. Marin. Maximin design on non hypercube domains and kernel interpolation. *Statistics and Computing*, 22:703–712, 2012.
- [Bor76] C. Borell. Gaussian Radon measures on locally convex spaces. *Mathematica Scandinavica*, 38:265–284, 1976.
- [BT04] A. Berline and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer, Boston, 2004.
- [CS02] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin (new series) of the American Mathematical Society*, 39(1):1–49, 2002.
- [DFCL71] R.M. Dudley, J. Feldman, and L. Le Cam. On seminorms and probabilities, and abstract Wiener spaces. *Annals of Mathematics*, pages 390–408, 1971.
- [DPZ13] H. Dette, A. Pepelyshev, and A. Zhigljavsky. Optimal design for linear models with correlated observations. *The Annals of Statistics*, 41(1):143–176, 2013.
- [Dud02] R.M. Dudley. *Real Analysis and Probability*, volume 74. Cambridge University Press, 2002.
- [For85] R.M. Fortet. Les opérateurs intégraux dont le noyau est une covariance. *Trabajos de Estadística y de Investigación Operativa*, 36:133–144, 1985.
- [GB12] B. Gauthier and X. Bay. Spectral approach for kernel-based interpolation. *Annales de la faculté des sciences de Toulouse*, 21 (num. 3):439–479, 2012.
- [GP14] B. Gauthier and L. Pronzato. Approximation of IMSE optimal designs of experiments via quadrature rule and spectral decomposition. *Preprint*, 2014. <http://hal.archives-ouvertes.fr/hal-00913466>.
- [Hac95] W. Hackbusch. *Integral Equations: Theory and Numerical Treatment*, volume 120. Birkhäuser, Basel, 1995.
- [Kre99] R. Kress. *Linear Integral Equations*, volume 82. Springer, New York, 1999.
- [LMK10] O.P. Le Maître and O.M. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, 2010.
- [Nie92] H. Niederreiter. *Random Number and Quasi-Monte Carlo Methods*. SIAM, Philadelphia, 1992.
- [R C13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [RW06] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT press, Cambridge, MA, 2006.
- [S<sup>+</sup>13] W.A. Stein et al. *Sage Mathematics Software (Version 5.6)*. The Sage Development Team, 2013. <http://www.sagemath.org>.
- [Sat69] H. Satô. Gaussian measure on a banach space and abstract Wiener measure. *Nagoya Mathematical Journal*, 36:65–81, 1969.
- [Sch79] L. Schwartz. *Analyse Hilbertienne*. Hermann, 1979.
- [SP10] G. Spöck and J. Pilz. Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. *Stochastic Environmental Research and Risk Assessment*, 24(3):463–482, 2010.
- [ST06] C. Schwab and R.A. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, 2006.
- [SWMW89] J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989.
- [SY66] J. Sacks and D. Ylvisaker. Designs for regression problems with correlated errors. *The Annals of Mathematical Statistics*, pages 66–89, 1966.
- [Wah90] G. Wahba. *Spline Models for Observational Data*, volume 59. SIAM, Philadelphia, 1990.