



**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.. 2013. hal-00913466v1

**HAL Id: hal-00913466**

**<https://hal.science/hal-00913466v1>**

Preprint submitted on 3 Dec 2013 (v1), 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 criterion can hence be evaluated without explicitly integrate the MSE and spectral truncation naturally defines an approximated criterion. The approach is of particular interest when a quadrature rule is used to approximate the integral of the MSE. In this situation, both IMSE and truncated-IMSE optimal designs are supported by quadrature points. In addition, numerical experiments indicate that retaining a small number of eigenpairs is sufficient to obtain IMSE-optimal designs when optimizing the truncated criterion, resulting in a significant reduction of the optimization computational cost.

**Key words.** Random field model, interpolation, design of experiments, IMSE, integral operator, Karhunen-Loève decomposition.

**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 minimises 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. The present work aims at investigating an alternative approach to make the computation of IMSE-optimal designs more tractable.

The choice of an IMSE criterion for learning a random field naturally leads to a particular Karhunen-Loève decomposition of this random field through the definition of an integral operator (see Section 3 and *e.g.*, [ST06]). The main idea of the paper is to link the integration of the mean-squared prediction error (MSE) with the evaluation, at the design points, of the eigenfunctions of the IMSE integral operator (more precisely, the evaluation of the *harmonics*, which are canonical extension of the eigenfunctions, see Definition 3.2 and Remark 3.2). We hence obtain, in equation (4.4), a spectral representation of the IMSE criterion. This representation can next be approximated by spectral truncation, leading to the *spectral truncated IMSE criterion*.

Our approach is of particular interest for the computation of IMSE-optimal designs when a quadrature rule is used to approximate the integral of the MSE. In such

---

<sup>†</sup>bgauthie@i3s.unice.fr

<sup>‡</sup>pronzato@i3s.unice.fr

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

a case, Proposition 3.2, Remarks 3.3 and 4.1 (see also Section 4.3) show that both optimal designs for the quadrature approximations of the IMSE and truncated-IMSE criteria are supported by quadrature points. So, all the necessary information for the optimisation of these two criteria is contained in the spectral decomposition of the IMSE integral operator (*i.e.*, the eigenfunctions have not to be extended to the whole space, which would require the computation of integrals). In addition, in the quadrature approximation case, numerical experiments indicate that retaining a small number of eigenpairs is sufficient to obtain IMSE-optimal designs when optimizing the truncated criterion, leading to a significant reduction of the computational cost of the optimization process.

We have tried to make this paper as self-contained as possible; the definitions of most concepts are thus reminded and most proofs are detailed. 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. The spectral representation of the IMSE design criterion and its approximation by spectral truncation are detailed in Section 4. This section also tackles the quadrature approximation of the IMSE integral operator. Numerical constructions of IMSE-optimal designs are carried out in Section 5. The numerical interest of the proposed approach and some perspectives are finally discussed in Section 6.

## 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*. 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.*

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

$\mathbf{H}$  is endowed with the Hilbert structure involved by  $L^2(\Omega, \mathbb{P})$ . We assume that  $\mathbf{H}$  is *separable*.

In parallel, we denote by  $\mathcal{H}$  the *reproducing kernel Hilbert space* (RKHS, see for instance [BTA04]) 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*,

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

with  $(\cdot|\cdot)_{\mathcal{H}}$  the inner product of  $\mathcal{H}$ . The two Hilbert spaces  $\mathcal{H}$  and  $\mathbf{H}$  are therefore 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  $\mathbf{H}_C$  be a closed linear subspace of the Hilbert space  $\mathbf{H}$ . We consider the orthogonal projection  $P_{\mathbf{H}_C}$  of  $\mathbf{H}$  onto  $\mathbf{H}_C$ . For  $x \in \mathcal{X}$ , the r.v.  $P_{\mathbf{H}_C}[Z_x]$  is called the *conditional mean* of the r.v.  $Z_x$  relatively to  $\mathbf{H}_C$ . If  $\mathbf{H}_C$  is spanned by the r.v.  $\zeta_j, j \in J$ , with  $J$  is a general index set, the notation

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

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

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

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

**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  $\mathbf{H}_C$  of  $\mathbf{H}$  corresponds a subspace  $\mathcal{H}_C$  of  $\mathcal{H}$  and one can define the orthogonal projection  $P_{\mathcal{H}_C}$  of  $\mathcal{H}$  onto  $\mathcal{H}_C$ , *etc.*

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

**3.1. IMSE criterion.** 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 Hilbert space of square integrable real-valued functions on  $\mathcal{X}$  with respect to  $\mu$ . 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 error, we shall make the widespread abuse of notation consisting in assimilating elements of  $L^2(\mathcal{X}, \mu)$  with functions on  $\mathcal{X}$ .

Following [For85, GB12], we make the following assumptions:

C-i. each function  $h \in \mathcal{H}$  is measurable,

C-ii. the kernel  $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is measurable (for the product  $\sigma$ -algebra),

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

These three conditions ensure in particular that the sample paths of  $(Z_x)_{x \in \mathcal{X}}$  are in  $L^2(\mathcal{X}, \mu)$  with  $\mathbb{P}$ -probability 1 (and therefore that  $\mathcal{H} \subset L^2(\mathcal{X}, \mu)$ , with continuous inclusion, see equation (3.7)).

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

$$\begin{aligned} \text{IMSE}(\mathbf{H}_C) &= \mathbb{E} \left[ \int_{\mathcal{X}} (Z_x - P_{\mathbf{H}_C}[Z_x])^2 d\mu(x) \right] \\ &= \int_{\mathcal{X}} \mathbb{E} \left[ (Z_x - P_{\mathbf{H}_C}[Z_x])^2 \right] d\mu(x) = \int_{\mathcal{X}} K(x, x) - \mathbb{E} \left[ (P_{\mathbf{H}_C}[Z_x])^2 \right] d\mu(x). \end{aligned} \quad (3.1)$$

The IMSE associated with the subspace  $\mathbf{H}_C$  is therefore the integral of the *conditional variance* of the considered conditioning. From C-iii, we have

$$\text{IMSE}(\mathbf{H}_C) = \tau - C_I(\mathbf{H}_C), \quad \text{with } C_I(\mathbf{H}_C) = \int_{\mathcal{X}} \mathbb{E} \left[ (P_{\mathbf{H}_C}[Z_x])^2 \right] d\mu(x). \quad (3.2)$$

Since  $\tau$  does not depend on  $\mathbf{H}_C$ , minimizing the IMSE criterion thus amounts to maximizing  $C_I(\mathbf{H}_C)$ .

**Remark 3.1.** In terms of RKHS,  $P_{\mathcal{H}_C} [K_x]$  is the reproducing kernel of  $\mathcal{H}_C$ . Hence, by isometry and from the properties of orthogonal projections,

$$\begin{aligned} C_I(\mathbf{H}_C) &= \int_{\mathcal{X}} (P_{\mathcal{H}_C} [K_x] | P_{\mathcal{H}_C} [K_x])_{\mathcal{H}} d\mu(x) = \int_{\mathcal{X}} (K_x | P_{\mathcal{H}_C} [K_x])_{\mathcal{H}} d\mu(x) \\ &= \int_{\mathcal{X}} \mathbb{E}[Z_x P_{\mathbf{H}_C} [Z_x]] d\mu(x). \end{aligned} \quad (3.3)$$

**Definition 3.1.** For a fixed  $n \in \mathbb{N}^*$ , a set  $\{x_1, \dots, x_n\}$  of  $n$  elements of  $\mathcal{X}$  is a  $n$ -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  $Z$ .

**3.2. Integral operator.** The introduction of an IMSE criterion for the learning of a random field  $Z$  naturally defines an integral operator  $T_\mu$  on  $L^2(\mathcal{X}, \mu)$ , given by

$$\forall f \in L^2(\mathcal{X}, \mu), \forall x \in \mathcal{X}, T_\mu[f](x) = \int_{\mathcal{X}} f(t) K(x, t) d\mu(t). \quad (3.4)$$

The interest of such operators when dealing with kernel-based interpolation models has been studied for instance in [CS02, GB12], see also [DPZ13] for application to optimal designs for linear models.

The operator  $T_\mu$  is diagonalisable and positive (see for instance [Sch79]). We denote by  $\lambda_i$  its eigenvalues (repeated according to their algebraic multiplicity) and by  $\tilde{\phi}_i \in L^2(\mathcal{X}, \mu)$  the associated eigenfunctions, with  $i \in \mathbb{I}$ , a general index set. We remind that  $\{\tilde{\phi}_i, i \in \mathbb{I}\}$  forms an orthonormal basis of  $L^2(\mathcal{X}, \mu)$  and that the set of all *strictly positive* eigenvalues of  $T_\mu$  is at most countable. We then denote by  $\{\lambda_k, k \in \mathbb{I}_+\}$  this set; *i.e.*,  $\lambda_k > 0$  for all  $k \in \mathbb{I}_+ \subset \mathbb{N}$ . When necessary, we shall index the set of the strictly positive eigenvalues of  $T_\mu$  by  $\mathbb{N}^*$ , the eigenvalues being sorted in descending order, *i.e.*  $\lambda_1 \geq \lambda_2 \geq \dots > 0$ .

**Proposition 3.1.** Denote by  $\mathcal{H}_\mu$  the orthogonal of the subspace  $\mathcal{H}_0$  of the Hilbert space  $\mathcal{H}$  (*i.e.*  $\mathcal{H}_\mu = \mathcal{H}_0^\perp$ ), with

$$\mathcal{H}_0 = \left\{ h_0 \in \mathcal{H}, \|h_0\|_{L^2}^2 = 0 \right\}. \quad (3.5)$$

For  $k \in \mathbb{I}_+$ , consider the functions  $\phi_k$  given by

$$\forall x \in \mathcal{X}, \phi_k(x) = \frac{1}{\lambda_k} \int_{\mathcal{X}} \tilde{\phi}_k(t) K(x, t) d\mu(t). \quad (3.6)$$

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  $\mathcal{H}$ .

*Proof.* First, as an orthogonal space,  $\mathcal{H}_\mu$  is closed in  $\mathcal{H}$ . From conditions C-i, C-ii and C-iii, the representation property and the Cauchy-Schwarz inequality, we obtain, for  $h \in \mathcal{H}$ ,

$$\|h\|_{L^2}^2 = \left| \int_{\mathcal{X}} (h | K_t)_{\mathcal{H}}^2 d\mu(t) \right| \leq \|h\|_{\mathcal{H}}^2 \int_{\mathcal{X}} K(t, t) d\mu(t) = \tau \|h\|_{\mathcal{H}}^2. \quad (3.7)$$

For a fixed  $f \in L^2(\mathcal{X}, \mu)$ , we consider the linear functional on  $\mathcal{H}$  defined by,

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

From the Cauchy-Schwarz inequality and expression (3.7), we have

$$|I_{f,\mu}(h)| \leq \|f\|_{L^2} \|h\|_{L^2} \leq \sqrt{\tau} \|f\|_{L^2} \|h\|_{\mathcal{H}}, \quad (3.8)$$

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 (see [GB12] for more details)

$$I_{f,\mu}(h) = (h|f)_{L^2} = (h|T_{\mu}[f])_{\mathcal{H}}. \quad (3.9)$$

This in particular proves that for all  $f \in L^2(\mathcal{X}, \mu)$ ,  $T_{\mu}[f] \in \mathcal{H}_0^{\perp\tau} = \mathcal{H}_{\mu}$ .

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,  $\mathcal{H}_0^{\perp\tau} = \mathcal{H}_0$ ).

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 (3.9),  $\|h\|_{L^2}^2 = (h|T_{\mu}[h])_{\mathcal{H}}$ , we obtain that  $h \in \mathcal{H}_0$  and finally that  $\text{span}\{\phi_k, k \in \mathbb{I}_+\}$  is dense in  $\mathcal{H}_{\mu}$ .  $\square$

**Definition 3.2.** For  $k \in \mathbb{I}_+$ , we call  $\mathfrak{h}_k = \sqrt{\lambda_k} \phi_k$  the  $k$ -th harmonic associated with  $\lambda_k$  and normed in  $\mathcal{H}$  (i.e., such that  $\|\mathfrak{h}_k\|_{\mathcal{H}} = 1$ ), defined by the random field  $Z$  and the  $\mu$ -IMSE criterion.

**Remark 3.2.** 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  $\mathcal{X}$ . Notice that

$$\tau = \int_{\mathcal{X}} K(x, x) d\mu(x) = \sum_{k \in \mathbb{I}_+} \lambda_k.$$

From equation (3.9), the harmonics  $\mathfrak{h}_k$ ,  $k \in \mathbb{I}_+$  (or, equivalently the functions  $\phi_k$ ) satisfy the following property:

$$\forall h \in \mathcal{H}, \forall k \in \mathbb{I}_+, (h|\phi_k)_{\mathcal{H}} = \frac{1}{\lambda_k} (h|T_{\mu}[\tilde{\phi}_k])_{\mathcal{H}} = \frac{1}{\lambda_k} (h|\tilde{\phi}_k)_{L^2}. \quad (3.10)$$

This expression links the inner-products in  $\mathcal{H}$  and in  $L^2(\mathcal{X}, \mu)$ . In particular, we obtain, for  $k \in \mathbb{I}_+$  and  $x \in \mathcal{X}$ ,

$$(\phi_k|K_x)_{L^2} = \lambda_k (\phi_k|K_x)_{\mathcal{H}} = \lambda_k \phi_k(x). \quad (3.11)$$

**3.3. Spectral representation and harmonics optimality.** For  $k \in \mathbb{I}_+$ , we introduce the r.v.  $\xi_k = \mathcal{I}(\mathfrak{h}_k) \in \mathbf{H}$  (orthogaussian harmonic random variables). We have, from equation (3.6),

$$\xi_k = \frac{1}{\sqrt{\lambda_k}} \int_{\mathcal{X}} \tilde{\phi}_k(t) Z_t d\mu(t)$$

and the  $\xi_k$ ,  $k \in \mathbb{I}_+$ , are orthonormal in  $\mathbf{H}$  (see [GB12] for additional considerations). 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

$$\mathbf{H} = \mathbf{H}_\mu \overset{\perp}{+} \mathbf{H}_0. \quad (3.12)$$

The following proposition shows that r.v. belonging to  $\mathbf{H}_0$  have no influence on the  $\mu$ -IMSE criterion.

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

$$C_I(\mathbf{H}_C) = C_I(\mathbf{H}_C \cap \mathbf{H}_\mu), \quad (3.13)$$

where the  $C_I$  is defined in equation (3.2).

*Proof.* From the orthogonal decomposition (3.12), we have

$$\mathbf{H}_C = (\mathbf{H}_\mu \cap \mathbf{H}_C) \overset{\perp}{+} (\mathbf{H}_0 \cap \mathbf{H}_C).$$

We introduce the notation  $\mathbf{H}_{\mu C} = \mathbf{H}_\mu \cap \mathbf{H}_C$  and  $\mathbf{H}_{0C} = \mathbf{H}_0 \cap \mathbf{H}_C$ , so that the orthogonal projection of  $\mathbf{H}$  onto  $\mathbf{H}_C$  can be written

$$P_{\mathbf{H}_C} = P_{\mathbf{H}_{\mu C}} + P_{\mathbf{H}_{0C}}.$$

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

$$\mathbb{E}[(P_{\mathbf{H}_C}[Z_x])^2] = \mathbb{E}[(P_{\mathbf{H}_{\mu C}}[Z_x])^2] + \mathbb{E}[(P_{\mathbf{H}_{0C}}[Z_x])^2], \text{ so that}$$

$$C_I(\mathbf{H}_C) = C_I(\mathbf{H}_C \cap \mathbf{H}_\mu) + C_I(\mathbf{H}_C \cap \mathbf{H}_0).$$

To conclude, we note that, from equations (3.3) and (3.5),

$$C_I(\mathbf{H}_{0C}) = \int_{\mathcal{X}} (P_{\mathcal{H}_{0C}}[K_x] | K_x)_{\mathcal{H}} d\mu(x) = 0,$$

since  $P_{\mathcal{H}_{0C}}[K_x] \in \mathcal{H}_0$  for all  $x \in \mathcal{X}$ . □

**Remark 3.3.** A direct consequence of Proposition 3.2 is that  $\mu$ -IMSE-optimal subspaces of  $\mathbf{H}$  with fixed dimension are subspaces of  $\mathbf{H}_\mu$ . Therefore, if  $\mathcal{X}$  is a topological space (endowed with its borel  $\sigma$ -algebra) and if the considered RKHS  $\mathcal{H}$  consists in continuous functions on  $\mathcal{X}$ , then  $\mu$ -IMSE-optimal designs are included in  $\text{supp}(\mu)$ , the support of the measure  $\mu$ . In particular, when a quadrature is used in order to approximate integrals over  $\mathcal{X}$ , *optimal designs for this quadrature are part of the quadrature points* (see Section 4.3 for details on the quadrature approximation).

**Proposition 3.3.** Let  $\mathbf{H}_C$  be a closed linear subspace of  $\mathbf{H}$  and let  $\{\eta_j, j \in J\}$  be an orthonormal basis of  $\mathbf{H}_C$ . Then

$$C_I(\mathbf{H}_C) = \sum_{k \in \mathbb{I}_+} \sum_{j \in J} \alpha_{j,k}^2 \lambda_k, \quad (3.14)$$

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

$$\forall j \in J, P_{\mathbf{H}_\mu}[\eta_j] = \sum_{k \in \mathbb{I}_+} \alpha_{j,k} \xi_k. \quad (3.15)$$

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

$$P_{\mathbf{H}_C}[U] = \sum_{j \in J} \mathbb{E}(\eta_j U) \eta_j \text{ and } \mathbb{E}[(P_{\mathbf{H}_C}[U])^2] = \sum_{j \in J} (\mathbb{E}(\eta_j U))^2. \quad (3.16)$$

Combining relations (3.16) with Proposition 3.2, we obtain

$$\begin{aligned} C_I(\mathbf{H}_C) &= C_I(\mathbf{H}_C \cap \mathbf{H}_\mu) = \int_{\mathcal{X}} \mathbb{E}[(P_{\mathbf{H}_C}[P_{\mathbf{H}_\mu}[Z_x]])^2] d\mu(x) \\ &= \sum_{j \in J} \int_{\mathcal{X}} [\mathbb{E}(\eta_j P_{\mathbf{H}_\mu}[Z_x])]^2 d\mu(x). \end{aligned} \quad (3.17)$$

We next consider the expansion

$$\forall x \in \mathcal{X}, P_{\mathbf{H}_\mu}[Z_x] = \sum_{k \in \mathbb{I}_+} \xi_k \mathbb{E}(\xi_k Z_x) = \sum_{k \in \mathbb{I}_+} \xi_k \sqrt{\lambda_k} \phi_k(x).$$

Combined with expressions (3.15) and (3.17), we have

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

completing the proof.  $\square$

We can now recall the following well-known result (Proposition 3.4), which shows the optimal character of the harmonic r.v. 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  $T_\mu$ . Then  $\mathbf{H}_n^*$  minimizes the IMSE criterion among all subspaces  $\mathbf{H}_n$  of  $\mathbf{H}$  with dimension  $n$ .

*Proof.* Let  $\mathbf{H}_n$  be a closed linear subspace of  $\mathbf{H}$  with dimension  $n$  and let  $\eta_1, \dots, \eta_n$  be an orthonormal basis of  $\mathbf{H}_n$ . From Proposition 3.3,

$$C_I(\mathbf{H}_n) = \sum_{k \in \mathbb{I}_+} \sum_{j=1}^n \alpha_{j,k}^2 \lambda_k \text{ with } \alpha_{j,k} = \mathbb{E}(\eta_j \xi_k) \text{ and } C_I(\mathbf{H}_n^*) = \sum_{k=1}^n \lambda_k. \quad (3.18)$$



By definition, we have, for any  $i$  and  $j \in \{1, \dots, n\}$ ,  $\sum_{k \in \mathbb{I}_+} \alpha_{i,k} \alpha_{j,k} = \delta_{ij}$  (Kronecker delta). For  $k \in \mathbb{I}_+$ , let  $\mathbf{a}_k \in \mathbb{R}^n$  be the (column) vector given by

$$\mathbf{a}_k = (\alpha_{1,k}, \dots, \alpha_{n,k})^T, \text{ so that } \sum_{j=1}^n \alpha_{j,k}^2 \stackrel{\text{(notation)}}{=} \mathbf{a}_k^T \mathbf{a}_k.$$

Here,  $\mathbf{a}_k^T$  stands for the transpose of the vector  $\mathbf{a}_k$  and  $\mathbf{a}_k^T \mathbf{a}_k$  has to be understood as a scalar product in  $\mathbb{R}^n$ . Note that we identify vectors of  $\mathbb{R}^n$ ,  $n \in \mathbb{N}^*$ , with the column vectors of their coefficients in the canonical basis of  $\mathbb{R}^n$ . A similar convention is also used for matrices through the rest of the paper.

We 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  $l \in \mathbb{I}_+$ ,

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

which proves that

$$\forall k \in \mathbb{I}_+, \sum_{j=1}^n \alpha_{j,k}^2 \leq 1. \quad (3.19)$$

Since  $\sum_{k \in \mathbb{I}_+} \alpha_{j,k}^2 = 1$  for all  $1 \leq j \leq n$ , we have

$$\sum_{k \in \mathbb{I}_+} \sum_{j=1}^n \alpha_{j,k}^2 = n. \quad (3.20)$$

Finally, since  $\lambda_1 \dots, \lambda_n$  are the largest eigenvalues of  $T_\mu$ , we deduce from equations (3.19) and (3.20) 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.  $\square$

#### 4. Computation of the IMSE criterion.

**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 (2.1).

Let  $\mathbf{k}$  be the (column) vector with components  $K_{x_i}$ , with  $1 \leq i \leq n$  (with a slight abuse of notation, we shall denote  $\mathcal{I}(\mathbf{k}) = \mathbf{z}$ ), *i.e.*,

$$\mathbf{k} = (K_{x_1}, \dots, K_{x_n})^T. \quad (4.1)$$

Let  $\mathbf{K}$  be the covariance matrix of  $\mathbf{z}$ . In matrix notation, we have

$$\mathbf{K} = \mathbb{E}(\mathbf{z}\mathbf{z}^T) = (\mathbf{k}|\mathbf{k}^T)_{\mathcal{H}}.$$

We assume that  $Z$  and  $\{x_1, \dots, x_n\}$  are such that  $\mathbf{K}$  is *invertible*. We thus obtain

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

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

$$C_I(\mathbf{H}_{ev}) = \int_{\mathcal{X}} \mathbf{k}^T(x) \mathbf{K}^{-1} \mathbf{k}(x) d\mu(x). \quad (4.2)$$

Therefore, 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  $\mathbf{K}$ ,
- the integration of the function  $x \mapsto \mathbf{k}^T(x) \mathbf{K}^{-1} \mathbf{k}(x)$ .

## 4.2. Alternative expression and truncation.

**4.2.1. Spectral representation for the IMSE.** Let us introduce the matrix  $\mathbf{X}$  with entries

$$\mathbf{X}_{i,k} = \sqrt{\lambda_k} \phi_k(x_i) = \mathfrak{h}_k(x_i) = \mathbb{E}(\xi_k Z_{x_i}), \text{ with } 1 \leq i \leq n \text{ and } k \in \mathbb{I}_+. \quad (4.3)$$

Hence,  $\mathbf{X}$  has  $n$  rows and  $\text{card}(\mathbb{I}_+)$  columns, where  $\text{card}(\mathbb{I}_+)$  is the cardinality of the set  $\mathbb{I}_+$ . We also introduce the diagonal matrix

$$\mathbf{\Lambda} = \text{diag}(\lambda_k, k \in \mathbb{I}_+).$$

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

$$C_I(\mathbf{H}_{ev}) = \sum_{k \in \mathbb{I}_+} \lambda_k \left[ (\mathbf{X}_{\cdot,k})^T \mathbf{K}^{-1} (\mathbf{X}_{\cdot,k}) \right] = \text{Tr}(\mathbf{\Lambda} \mathbf{X}^T \mathbf{K}^{-1} \mathbf{X}), \quad (4.4)$$

where  $(\mathbf{X}_{\cdot,k})$  is the  $k$ -th column of the matrix  $\mathbf{X}$ , that is, the column vector of the values taken by the  $k$ -th harmonic at  $x_1, \dots, x_n$ .

*Proof.* Expression (4.4) 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  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T = \mathbf{C}^{-1} \mathbf{z}$  form an orthonormal basis of  $\mathbf{H}_{ev}$ .

If we consider the notations introduced in the proof of Proposition 3.4, we have  $\mathbf{a}_k = (\alpha_{1,k}, \dots, \alpha_{n,k})^T = \mathbf{C}^{-1} (\mathbf{X}_{\cdot,k})$ , so that

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

We hence obtain  $C_I(\mathbf{H}_{ev}) = \sum_{k \in \mathbb{I}_+} \lambda_k \left[ (\mathbf{X}_{\cdot,k})^T \mathbf{K}^{-1} (\mathbf{X}_{\cdot,k}) \right]$ , from which the second equality in equation (4.4) directly follows.  $\square$

Equation (4.4) implies that once the spectral decomposition of the IMSE integral operator is known (Section 3.2), the value of  $\text{IMSE}(\mathbf{H}_{ev})$  can be obtained *without the explicit computation of the mean-squared prediction error*. The calculation of the integral is indeed replaced by the evaluations of the harmonic functions  $\mathfrak{h}_k$ ,  $k \in \mathbb{I}_+$ , (see equation (4.3) and Definition 3.2) at the sampling points  $x_1, \dots, x_n \in \mathcal{X}$ .

Another interest of the alternative expression (4.4) for the IMSE lies in the fact that it can be easily approximated by spectral truncation, as we consider now.

**4.2.2. Approximation by truncation.** Let  $\mathbf{H}_n$  be a linear subspace of  $\mathbf{H}$  with dimension  $n$ . Consider a subset  $\mathbb{I}_{trc} \subset \mathbb{I}_+$  and introduce

$$C_{I_{trc}}(\mathbf{H}_n) = \sum_{k \in \mathbb{I}_{trc}} \sum_{j=1}^n \alpha_{j,k}^2 \lambda_k. \quad (4.5)$$

In practice,  $\mathbb{I}_{trc}$  will be a finite set associated with the *largest eigenvalues* of  $T_\mu$  (*spectral truncation*). Let  $\mathbf{H}_{ev} = \text{span}\{Z_{x_1}, \dots, Z_{x_n}\}$  be given by equation (2.1) and such that  $\mathbf{K}$  is invertible (see Section 4.1). We introduce the matrices  $\mathbf{\Lambda}_{trc} = \text{diag}(\lambda_k, k \in \mathbb{I}_{trc})$  and  $\mathbf{X}_{trc}$ , with

$$(\mathbf{X}_{trc})_{i,k} = \sqrt{\lambda_k} \phi_k(x_i) = \mathfrak{h}_k(x_i), \text{ with } 1 \leq i \leq n \text{ and } k \in \mathbb{I}_{trc}.$$

**Definition 4.1.** *The spectral truncated IMSE criterion associated with a subset  $\mathbb{I}_{trc}$  of  $\mathbb{I}_+$  and a design  $\{x_1, \dots, x_n\}$  is given by  $\text{IMSE}_{trc}(\mathbf{H}_{ev}) = \tau - C_{I_{trc}}(\mathbf{H}_{ev})$ , with*

$$C_{I_{trc}}(\mathbf{H}_{ev}) = \sum_{k \in \mathbb{I}_{trc}} \lambda_k \left[ (\mathbf{X}_{\cdot,k})^T \mathbf{K}^{-1} (\mathbf{X}_{\cdot,k}) \right] = \text{Tr}(\mathbf{\Lambda}_{trc} \mathbf{X}_{trc}^T \mathbf{K}^{-1} \mathbf{X}_{trc}).$$

In Definition 4.1, remark that we may replace  $\tau$  by  $\sum_{k \in \mathbb{I}_{trc}} \lambda_k$ . Indeed, the main point here is the approximation of  $C_I(\mathbf{H}_{ev})$  by  $C_{I_{trc}}(\mathbf{H}_{ev})$ , the quantities  $\tau$  and  $\sum_{k \in \mathbb{I}_{trc}} \lambda_k$  being independent of the design  $\{x_1, \dots, x_n\}$ .

**Remark 4.1.** A property similar to Proposition 3.2 also holds for the truncated IMSE criterion since for a closed linear subspace  $\mathbf{H}_C$  of  $\mathbf{H}$ ,

$$C_{I_{trc}}(\mathbf{H}_C) = C_{I_{trc}}(\mathbf{H}_C \cap \mathbf{H}_\mu).$$

Therefore, as for the IMSE criterion, when a quadrature approximation is used, optimal designs for  $\text{IMSE}_{trc}$  are supported by quadrature points.

**Remark 4.2** (Distance to harmonics). Let  $\mathbf{H}_n$  be a linear subspace of  $\mathbf{H}$  with dimension  $n$  and consider the optimal subspace  $\mathbf{H}_n^*$  of Proposition 3.4. We have

$$\begin{aligned} & \int_{\mathcal{X}} \mathbb{E} \left[ (P_{\mathbf{H}_n^*} [Z_x] - P_{\mathbf{H}_n} [Z_x])^2 \right] d\mu(x) \\ &= C_I(\mathbf{H}_n^*) + C_I(\mathbf{H}_n) - 2 \int_{\mathcal{X}} \mathbb{E} [P_{\mathbf{H}_n^*} [Z_x] P_{\mathbf{H}_n} [Z_x]] d\mu(x) \\ &\leq 2C_I(\mathbf{H}_n^*) - 2 \int_{\mathcal{X}} \mathbb{E} [P_{\mathbf{H}_n^*} [Z_x] P_{\mathbf{H}_n} [Z_x]] d\mu(x), \quad (4.6) \end{aligned}$$

where the inequality follows from the majoration  $C_I(\mathbf{H}_n^*) \geq C_I(\mathbf{H}_n)$ . Considering expression (4.6), we introduce the term

$$C_S(\mathbf{H}_n) = \int_{\mathcal{X}} \mathbb{E} [P_{\mathbf{H}_n^*} [Z_x] P_{\mathbf{H}_n} [Z_x]] d\mu(x) \quad (4.7)$$

and the majoration appearing in expression (4.6) suggests that designs which maximise the  $C_S$  criterion should perform relatively well in terms of IMSE. Following Section 3.3, we then consider an orthonormal basis  $\eta_1, \dots, \eta_n$  of  $\mathbf{H}_n$ . We hence obtain, from equation (3.15),

$$C_S(\mathbf{H}_n) = \sum_{k=1}^n \sum_{j=1}^n \alpha_{j,k}^2 \lambda_k. \quad (4.8)$$

Comparing expression (4.8) with equations (3.14) and (4.5), we finally remark that the criterion  $C_S(\mathbf{H}_n)$  is in fact the criterion  $C_{I_{trc}}(\mathbf{H}_n)$  where only the  $n$  first harmonic r.v.  $\xi_1, \dots, \xi_n$  (*i.e.*, those associated with the  $n$  largest eigenvalues of  $T_\mu$ ) have been retained.

**Remark 4.3** (Truncation level). Notice that for any closed linear subspace  $\mathbf{H}_C$  of  $\mathbf{H}$ , we have, from equations (3.18), (4.5) and (3.19),

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

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

**4.3. Quadrature and Nyström method.** Let us consider the situation where a quadrature rule is used in order to approximate integrals over  $\mathcal{X}$  relatively to  $\mu$ . Then, for all real-valued functions  $f$  on  $\mathcal{X}$ , integrable with respect to  $\mu$  and with sufficient regularity (that is,  $f$  has to be bounded at the quadrature points),

$$\int_{\mathcal{X}} f(s) d\mu(s) \approx \sum_{j=1}^{N_q} \omega_j f(s_j), \quad (4.9)$$

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 finite measure

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

with  $\delta_{s_j}$  the Dirac measure at  $s_j$ . All the results contained in this section can therefore be directly deduced from Section 3.2. However, due to the great importance of quadrature approximations in many applications, some details are given below.

We thus consider an approximation  $T_{\hat{\mu}}$  of the integral operator  $T_\mu$  (Nyström method, see for instance [Hac95, Kre99]), more precisely,

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

Following for instance [GB12, Section 5], 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

$$\mathbf{Q}_{i,j} = K(s_i, s_j), \text{ for } 1 \leq i, j \leq N_q.$$

$\mathbf{W}$  is thus the matrix of the quadrature weights and  $\mathbf{Q}$  is the covariance matrix of the quadrature points. We assume that  $\mathbf{Q}$  is nonsingular. Denote by  $\hat{\lambda}_1 > 0, \dots, \hat{\lambda}_{N_q} > 0$  the eigenvalues of the matrix  $\mathbf{Q}\mathbf{W}$  and by  $\mathbf{v}_1, \dots, \mathbf{v}_{N_q}$  their associated eigenvectors, *i.e.*  $\mathbf{Q}\mathbf{W} = \mathbf{P}\hat{\Lambda}\mathbf{P}^{-1}$  with

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

Remark that  $\{\mathbf{v}_1, \dots, \mathbf{v}_{N_q}\}$  forms an orthonormal basis of  $\mathbb{R}^{N_q}$  endowed with the inner product  $(\cdot | \cdot)_{\mathbf{W}}$ , where for  $\mathbf{x}$  and  $\mathbf{y} \in \mathbb{R}^{N_q}$   $(\mathbf{x} | \mathbf{y})_{\mathbf{W}} = \mathbf{x}^T \mathbf{W} \mathbf{y}$ . Thus, we have

$$\mathbf{P}^T \mathbf{W} \mathbf{P} = \text{Id}_{N_q \times N_q}, \text{ where } \text{Id}_{N_q \times N_q} \text{ is the } N_q \times N_q \text{ identity matrix.} \quad (4.10)$$

Notice that, throughout the paper, the hat symbol indicates that the corresponding object is associated with the operator  $T_{\hat{\mu}}$  (or more generally, with the quadrature approximation case), *i.e.*, the quadrature approximation of  $T_{\mu}$ . For instance,  $\hat{\lambda}_k$  refers to an eigenvalue of  $T_{\hat{\mu}}$ , when  $\lambda_k$  refers to an eigenvalue of  $T_{\mu}$ , *etc.*

As in equation (4.1), we use the notation

$$\mathbf{q}(\cdot) = (K(s_1, \cdot), \dots, K(s_{N_q}, \cdot))^T.$$

If we also introduce  $\mathbf{f} = (f(s_1), \dots, f(s_{N_q}))^T \in \mathbb{R}^{N_q}$  (*i.e.*,  $\mathbf{f}$  is the column vector with components the values of the function  $f$  at the quadrature points), we obtain,

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

**Proposition 4.2.** *Let us consider the functions defined by, for  $1 \leq k \leq N_q$ ,*

$$\forall x \in \mathcal{X}, \hat{\phi}_k(x) = \mathbf{q}^T(x) \mathbf{Q}^{-1} \mathbf{v}_k = \frac{1}{\hat{\lambda}_k} \mathbf{q}^T(x) \mathbf{W} \mathbf{v}_k. \quad (4.11)$$

*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  $\|\hat{\phi}_k\|_{\mathcal{H}}^2 = \frac{1}{\hat{\lambda}_k}$ .*

*Proof.* By construction,  $\hat{\phi}_k \in \mathcal{H}_{\hat{\mu}} = \text{span} \{K_{s_j}, 1 \leq j \leq N_q\}$  for all  $1 \leq k \leq N_q$  and, for  $x \in \mathcal{X}$ ,

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

Next, 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.  $\square$

Notice that equation (4.11) is the equivalent of expression (3.6) in the quadrature case. Therefore, for  $x \in \mathcal{X}$  and  $1 \leq k \leq N_q$ , the computation of  $\hat{\phi}_k(x)$  requires the computation of an integral (here, of a quadrature). However, from Proposition 3.2 and Remark 4.1 applied to  $\mathcal{H}_{\hat{\mu}}$ , we know that optimal designs for the quadrature approximation of the criteria  $C_I$  and  $C_{I_{trc}}$  are subsets of quadrature points. In addition, for all  $1 \leq j, k \leq N_q$ ,

$$\hat{\phi}_k(s_j) = (\mathbf{v}_k)_j, \quad (4.12)$$

where  $(\mathbf{v}_k)_j$  is the  $j$ -th component of the eigenvector  $\mathbf{v}_k$ . Therefore, once the spectral decomposition of the integral operator (here, the matrix  $\mathbf{Q} \mathbf{W}$ ) is known, we can maximize the quadrature approximation of the criteria  $C_I$  and  $C_{I_{trc}}$  (with respect to designs supported on quadrature points) without computing any additional integration (here summation over the quadrature points). See also Section 6.1 for further discussions on the numerical interest of the approach.

**5. Examples.** We present two numerical examples of computation of IMSE-optimal designs using quadrature approximation and spectral truncations. All computations have been performed with the free softwares **R** and **Sage** [Hor12, S<sup>+</sup>13].

In what follows, we denote by  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$  the quadrature approximation of the criteria  $C_I$  and  $C_{I_{trc}}$ . We shall only consider truncations with  $n_{trc} \geq n$  although, in practice, any  $n_{trc} \geq 1$  could be considered, whatever the value of  $n$ .

**5.1. Ornstein-Uhlenbeck process.** Consider a centered Gaussian process  $Z$  on  $[0, 1]$  with covariance

$$\forall x \text{ and } y \in [0, 1], K(x, y) = e^{-|x-y|} - e^{-(x+y)}.$$

This corresponds to the covariance of a centered Ornstein-Uhlenbeck process on  $[0, 1]$ , conditioned to vanish at 0. We assume  $[0, 1]$  endowed with a uniform probability measure (denoted by  $\mu$ ). We have in particular

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

Note that this corresponds to a rather unfavorable situation for the construction of  $\mu$ -IMSE-optimal designs using spectral truncation, since it combines a relatively slow decreasing spectrum with the lack of symmetry.

We approximate integrals over  $[0, 1]$  with a 500-point quadrature, where

$$s_j = \frac{2j-1}{1000} \text{ with } 1 \leq j \leq 500,$$

each point receiving a weight  $1/500$  (mid-point rectangular quadrature method with intervals of length 0.002), and we compute the spectral decomposition of the Nyström approximation defined by this quadrature. We denote by  $\widehat{\mu}$  the corresponding finite measure. Figure 5.1 shows the graph of variance of the process  $Z$  and illustrates the spectral decomposition of  $T_{\widehat{\mu}}$ .

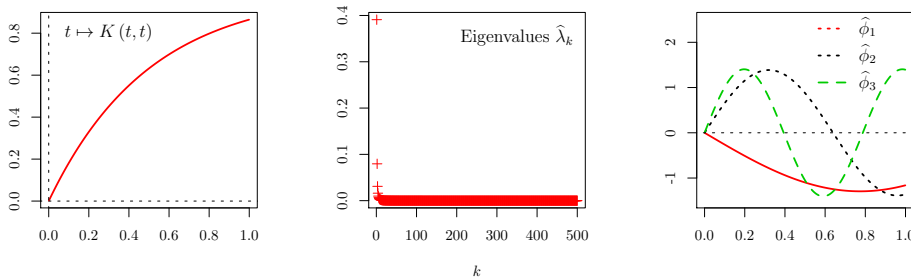


FIGURE 5.1. Graph of the variance of the process considered (left); eigenvalues of the Nyström approximation of the IMSE integral operator (middle); eigenfunctions associated with the three largest eigenvalues of the Nyström approximation (right).

*1-point optimal designs.* Let  $t \in [0, 1]$  and consider  $\mathbf{H}_t = \text{span}\{Z_t\}$  (i.e.,  $t$  denotes the design point). We have

$$C_I(\mathbf{H}_t) = \frac{e^{-2t}}{1 - e^{-2t}} \left( \frac{e^{2t} - e^{-2t}}{2} - 2t \right) + (e^{2t} - 1) \frac{e^{-2t} - e^{-2}}{2}. \quad (5.1)$$

Inside  $[0, 1]$ , this function reaches its maximum at  $t^* \approx 0.707859$ , the 1-point  $\mu$ -IMSE-optimal design for this problem (so, without quadrature approximation).

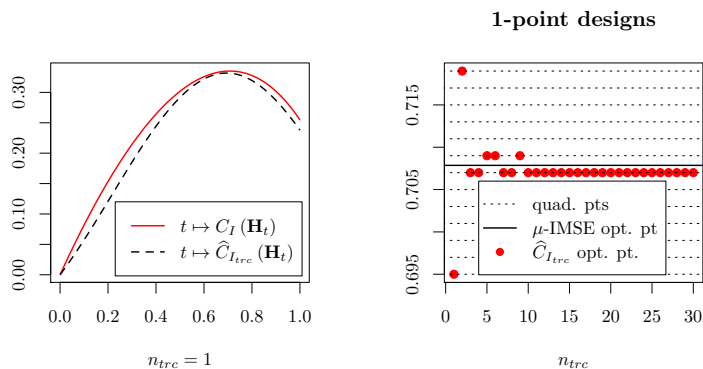


FIGURE 5.2. Graph of the values of the criteria  $C_I$  and  $\hat{C}_{I_{trc}}$  for  $n_{trc} = 1$  as function of the design point (left); 1-point optimal design for the  $\hat{C}_{I_{trc}}$  as a function of  $n_{trc}$  (right).

The left part of Figure 5.2 shows the values of the criterion  $C_I$  (expression (5.1), no quadrature approximation) and of  $\hat{C}_{I_{trc}}$  (quadrature approximation) with  $n_{trc} = 1$ , as functions of the design point  $t$ . The right part of this figure shows the design  $\hat{t}_{n_{trc}}^*$  optimal for  $\hat{C}_{I_{trc}}$  as a function of the number  $n_{trc}$  of eigenvalues retained (for instance, for  $n_{trc} = 1$ ,  $\hat{t}_{n_{trc}}^* = 0.695$ ). We observe a quick convergence of  $\hat{t}_{n_{trc}}^*$  to the  $\mu$ -IMSE-optimal design  $t^* = \hat{t}_{N_q}^* = 0.707$  as  $n_{trc}$  increases. For  $3 \leq n_{trc} \leq 9$ ,  $\hat{t}_{n_{trc}}^*$  oscillates between 0.707 and 0.709, the two quadrature points closest to the  $\mu$ -IMSE-optimal design  $t^*$ . For  $n_{trc} \geq 10$ ,  $\hat{t}_{n_{trc}}^*$  coincides with  $t^*$ .

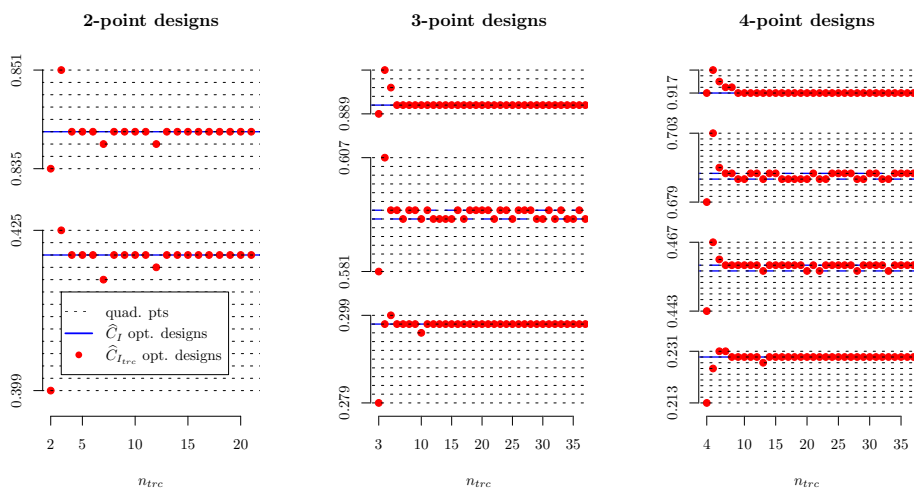


FIGURE 5.3. 2-, 3- and 4-point optimal designs for  $\hat{C}_{I_{trc}}$  as functions of  $n_{trc}$ .

*n*-point optimal designs,  $n = 2, 3$  and 4. Figure 5.3 shows the 2-, 3- and 4-point designs optimal for  $\widehat{C}_{I_{trc}}$  as functions of the number of eigenvalues considered. For the 3-point case, the oscillations of the central design point between 0.593 and 0.595 are due to the fact that the two corresponding designs are  $\widehat{\mu}$ -IMSE-optimal. The same phenomenon occurs for the 4-point designs, with now oscillations for the two central points. More precisely, the three 4-point designs  $\{0.229, 0.457, 0.687, 0.917\}$ ,  $\{0.229, 0.459, 0.687, 0.917\}$  and  $\{0.229, 0.459, 0.689, 0.917\}$  are  $\widehat{\mu}$ -IMSE-optimal. As for the 1-point case, we observe a quick convergence of designs optimal for  $\widehat{C}_{I_{trc}}$  to  $\widehat{\mu}$ -IMSE-optimal designs (*i.e.*, optimal for  $\widehat{C}_I$ ) when  $n_{trc}$  increases.

**5.2. Gaussian kernel.** Consider now a centered Gaussian process  $Z$  on  $[0, 1]^2$  with covariance (Gaussian, or squared exponential) kernel,

$$\forall x \text{ and } y \in [0, 1]^2, K(x, y) = e^{-\|x-y\|^2},$$

where  $\|\cdot\|$  is the Euclidean norm of  $\mathbb{R}^2$ . We assume  $[0, 1]^2$  endowed with a uniform probability measure, denoted by  $\mu$  (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 finite measure is again denoted by  $\widehat{\mu}$ . Two eigenfunctions of the Nyström approximation of the IMSE integral operator are presented in Figure 5.4.

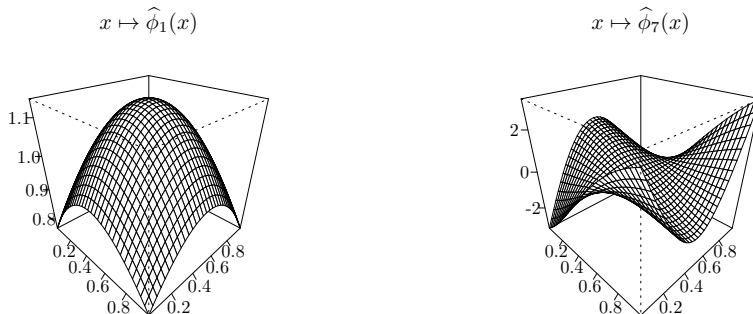


FIGURE 5.4. Two eigenfunctions of the Nyström approximation for the Gaussian kernel with  $\mu$  the uniform probability measure on the unit square.

Figure 5.5 shows the  $\widehat{\mu}$ -IMSE-optimal designs  $\widehat{X}_n^*$  and the optimal designs  $\widehat{X}_{n, n_{trc}}^*$  for  $\widehat{C}_{I_{trc}}$  with  $n_{trc} = n$  ( $n$  is the design size), for  $n = 4$  and  $n = 5$ .  $\widehat{X}_n^*$  and  $\widehat{X}_{n, n}^*$  coincide for  $n = 4$  but not for  $n = 5$ . However, we numerically observe that the 5-point optimal designs for  $\widehat{C}_{I_{trc}}$  with  $n_{trc} \geq 6$  are the  $\widehat{\mu}$ -IMSE optimal design  $\widehat{X}_5^*$ . The right part of Figure 5.5 illustrates in particular how  $\widehat{C}_{I_{trc}}(\widehat{X}_5^*)$  tends to  $\widehat{C}_I(\widehat{X}_5^*)$  when  $n_{trc}$  increases. Notice that  $\widehat{C}_{I_{trc}}(\widehat{X}_5^*)$  is an increasing function of  $n_{trc}$  and that for  $n_{trc} = 6$ ,  $\widehat{C}_{I_{trc}}(\widehat{X}_5^*) \approx 0.9890146$  is already very close of  $\widehat{C}_I(\widehat{X}_5^*) \approx 0.9890174$ .

## 6. Concluding remarks.



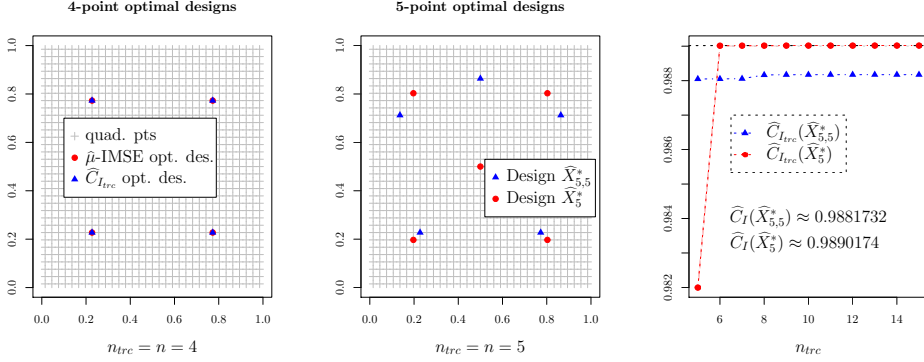


FIGURE 5.5. Optimal designs for the criteria  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$  for  $n = n_{trc} = 4$  (left) and  $n = n_{trc} = 5$  (middle). Values of the criterion  $\widehat{C}_{I_{trc}}$  for the two designs  $\widehat{X}_5^*$  and  $\widehat{X}_{5,5}^*$  (respectively optimal for  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$  with  $n_{trc} = 5$ ) as function of the number  $n_{trc}$  of considered eigenvalues (right).

**6.1. Numerical interest of the approach.** Consider the general framework of Section 4.3 (*i.e.*, quadrature approximation) and suppose that we want to compute a  $\widehat{\mu}$ -IMSE-optimal design for this quadrature. In this case,  $\text{card}(\widehat{\mathbb{I}}_+) = N_q$ , where  $N_q$  is the number of quadrature points. From equations (4.2), (4.4) and (4.9) and for a design  $\{x_1, \dots, x_n\}$ , we have (recall that  $\widehat{C}_I$  is the quadrature approximation of  $C_I$ )

$$\widehat{C}_I(\mathbf{H}_{ev}) = \sum_{j=1}^{N_q} \omega_j \mathbf{k}(s_j)^T \mathbf{K}^{-1} \mathbf{k}(s_j) = \sum_{k=1}^{N_q} \widehat{\lambda}_k \left[ \left( \widehat{\mathbf{X}}_{\cdot, k} \right)^T \mathbf{K}^{-1} \left( \widehat{\mathbf{X}}_{\cdot, k} \right) \right],$$

where, by analogy with Section 4.2,  $\widehat{\mathbf{X}}$  is the matrix with entries

$$\widehat{\mathbf{X}}_{i, k} = \sqrt{\widehat{\lambda}_k} \widehat{\phi}_k(x_i), \text{ with } 1 \leq i \leq n \text{ and } 1 \leq k \leq N_q.$$

Now, if we only retain the  $n_{trc}$  largest eigenvalues of the matrix  $\mathbf{QW}$  (with  $n_{trc} = \text{card}(\widehat{\mathbb{I}}_{trc}) \in \mathbb{N}^*$  and  $n_{trc} \leq N_q$ ), we have, from Definition 4.1,

$$\widehat{C}_{I_{trc}}(\mathbf{H}_{ev}) = \sum_{k=1}^{n_{trc}} \widehat{\lambda}_k \left[ \left( \widehat{\mathbf{X}}_{\cdot, k} \right)^T \mathbf{K}^{-1} \left( \widehat{\mathbf{X}}_{\cdot, k} \right) \right].$$

From Proposition 3.2 and Remark 4.1, we know that optimal designs for  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$  are subsets of quadrature points, so that we can restrict our attention to such designs and therefore read the terms of the matrix  $\widehat{\mathbf{X}}$  directly from the spectral decomposition of  $\mathbf{QW}$  through expression (4.12).

The computation of  $\widehat{C}_I(\mathbf{H}_{ev})$  thus requires the evaluation of a sum of  $N_q$  terms (in addition to the inversion of the design covariance matrix  $\mathbf{K}$ ), where  $N_q$  is large in practical applications. On the other hand, the computation of  $\widehat{C}_{I_{trc}}(\mathbf{H}_{ev})$  only requires the evaluation of a sum of  $n_{trc}$  terms, and numerical experiments indicate that a value  $n_{trc} \ll N_q$  (typically,  $n_{trc}$  slightly larger than the design size  $n$ , see Section 5 and Remark 4.2) is sufficient to obtain a  $\widehat{\mu}$ -IMSE-optimal design when maximizing  $\widehat{C}_{I_{trc}}$ . Therefore, once the spectral decomposition of the matrix  $\mathbf{QW}$  is performed,

no additional demanding computation is required for the evaluation of  $\widehat{C}_{I_{trc}}$  at any new design. Since modern algorithms for spectral decomposition are extremely fast and efficient (all the more as we focus on the eigenpairs corresponding to the largest eigenvalues), the extra computational time due to the initial spectral decomposition is largely compensated by the gain resulting in the use of the criterion  $\widehat{C}_{I_{trc}}$  instead of  $\widehat{C}_I$  for the construction of an optimal design.

**6.2. Perspectives.** In its present form, the proposed approach only considers the problem of designing IMSE-optimal experiments in the context of centered Gaussian random fields (or more generally, in the context of Gaussian random fields with known mean). Extension to kernel-based interpolation models including an unknown parametric trend is therefore of importance in order to enlarge the spectrum of potential applications.

Remark 4.3 can be used to set an upper bound on the truncation level  $n_{trc}$ . However, this bound is generally not tight as numerical examples indicate that it is sufficient to only retain a relatively small number of eigenpairs in the spectral decomposition in order to obtain IMSE-optimal designs (or good approximations of them) when optimizing the truncated criterion. In particular, we observed that  $n_{trc} = n$  (i.e., the design size) is sometimes enough (see Section 5 and Remark 4.2). A better understanding of the *truncation level* allowed (that is, choosing  $n_{trc}$  as small as possible while ensuring IMSE-optimality) would thus be of interest.

Finally, the choice of a suitable quadrature takes a special importance here since optimal designs are searched among the quadrature points. The construction of error bounds between the exact criteria  $C_I$  and  $C_{I_{trc}}$  and their quadrature approximations  $\widehat{C}_I$  and  $\widehat{C}_{I_{trc}}$  should hence deserve further studies, see for instance [ST06] for preliminary results concerning the approximation of the Karhunen-Loève decomposition of random fields.

**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 JKU Universität, Linz (Austria).

#### 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.
- [BTA04] 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.
- [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.
- [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.
- [Hac95] W. Hackbusch. *Integral Equations: Theory and Numerical Treatment*, volume 120. Birkhäuser, Basel, 1995.
- [Hor12] K. Hornik. The R FAQ, 2012. <http://cran.r-project.org>.
- [Kre99] R. Kress. *Linear Integral Equations*, volume 82. Springer, 1999.
- [RW06] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT press, Cambridge, MA, 2006.
- [S+13] W.A. Stein et al. *Sage Mathematics Software (Version 5.6)*. The Sage Development Team, 2013. <http://www.sagemath.org>.

- [Sch79] L. Schwartz. *Analyse Hilbertienne*. Hermann, 1979.
- [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.
- [Wah90] G. Wahba. *Spline Models for Observational Data*, volume 59. SIAM, 1990.