# Optimal Interpolation in RKHS, Spectral Decomposition of Integral Operators and Application

B. Gauthier[a,1,*], X. Bay[a,2]

[a] *Ecole Nationale Supérieure des Mines de Saint-Etienne*
*158 cours Fauriel, 42023 SAINT-ETIENNE, FRANCE*

**Abstract**

The orthogonal projection associated to optimal interpolation in a reproducing kernel Hilbert space is characterized by the spectral decomposition of an integral operator. This operator is built from the reproducing kernel and a measure on the interpolation set. As an application, we illustrate how boundary value constraints can be enforced in Gaussian process models.

*Keywords:* Reproducing kernel Hilbert space, Optimal interpolation, Integral Operator, Kriging, Boundary value constraints

## 1. Introduction

Let $\mathcal{X}$ be a general set and $\mathcal{H}$ a Hilbert space of real-valued functions on $\mathcal{X}$, we note respectively $(\cdot|\cdot)_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{H}}$ its inner product and norm.

Let $\mathcal{S}$ be a non-empty subset of $\mathcal{X}$ and fix $\varphi \in \mathcal{H}$. We say of $h \in \mathcal{H}$ that it interpolates $\varphi$ on $\mathcal{S}$ if, for all $s \in \mathcal{S}$, $h(s) = \varphi(s)$.

When $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS, see [1] or [2] for the more general notion of Hilbert subspace), the set of all interpolators can be easily characterized by the reproducing kernel $K(\cdot,\cdot)$ of $\mathcal{H}$. We remind that $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is symmetric and non-negative definite and verifies :

$$\forall x \in \mathcal{X},\ \forall h \in \mathcal{H},\ (h|K_x)_{\mathcal{H}} = h(x),\ \text{with (notation) } K_x(\cdot) = K(x,\cdot).$$

We set $\mathcal{H}_0 = \{h \in \mathcal{H} \ : \ \forall s \in \mathcal{S},\ h(s) = 0\ \}$ and $\mathcal{H}_1 = \mathcal{H}_0^{\perp} = \overline{span\,\{K_s(\cdot),\ s \in \mathcal{S}\}}^{\mathcal{H}}$ (in addition $\mathcal{H}_0 = \mathcal{H}_1^{\perp}$). Hence, $\mathcal{H}$ admits the orthogonal decomposition $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1$. $\mathcal{H}_0$ and $\mathcal{H}_1$ endowed with $(\cdot|\cdot)_{\mathcal{H}}$ are some RKHS, we respectively denote $K_0(\cdot,\cdot)$ and $K_1(\cdot,\cdot)$ their associated reproducing kernel, so $K(\cdot,\cdot) = K_0(\cdot,\cdot) + K_1(\cdot,\cdot)$.

By definition, $\mathcal{H}_0$ is a closed linear subspace of $\mathcal{H}$ and, for $\varphi \in \mathcal{H}$, $\varphi + \mathcal{H}_0$ is the set of all interpolators of $\varphi$ on $\mathcal{S}$. Hence, $\varphi + \mathcal{H}_0$ is a non-empty closed convex subset of $\mathcal{H}$, thus, it contains an element $h_M$ with minimal norm which is the orthogonal projection of the

* Corresponding author
[1] gauthier@emse.fr, tel:0033-477499740, fax:0033-477496666
[2] bay@emse.fr

origin onto $\varphi + \mathcal{H}_0$. We call $h_M$ the minimal norm interpolator (or optimal interpolator, see for example [3]). The previous result is essentially non-constructive : in practical cases, the function $\varphi$ is unknown and the only available information is the values taken by $\varphi$ on $\mathcal{S}$, i.e. $(\varphi|K_s)_{\mathcal{H}}$ for all $s \in \mathcal{S}$. However, $\mathcal{H}_1$ is also non-empty, closed and convex, hence one can also see $h_M$ as the orthogonal projection of $\varphi$ onto $\mathcal{H}_1$. We denote $\mathbb{P}_{\mathcal{H}_1} : \mathcal{H} \to \mathcal{H}$ this orthogonal projection (from $\mathcal{H}$ onto $\mathcal{H}_1$), thus $h_M = \mathbb{P}_{\mathcal{H}_1}[\varphi]$. The set of all interpolators of $\mathcal{H}$ is finally described by $h_M + \mathcal{H}_0$.

Note that the RKHS $\mathcal{H}$ is isometric to the Gaussian Hilbert space spanned by a centered second-order (Gaussian) stochastic process with covariance $K(\cdot,\cdot)$. Hence, optimal interpolation in RKHS and linear conditioning of second-order stochastic processes are equivalent in terms of Hilbert structures (confer [4, 5, 6] for the links between interpolation in RKHS and Gaussian process models).

In many applications, the set $\mathcal{S}$ is finite, which implies that $\mathcal{H}_1$ is of finite dimension. Then, the expression of $h_M$ can be easily obtained (see for instance [6]). In this work $\mathcal{H}_1$ is not necessary of finite dimension. It generally occurs when one assumes that $\mathcal{S}$ is composed of an infinite number of points : for instance, when the target function (i.e. the function $\varphi$) values are supposed to be known on the boundary of the domain.

In the first part of this article, we define the projection $\mathbb{P}_{\mathcal{H}_1}$ from the spectral decomposition of an integral operator. We show that this kind of representation allows an efficient and controllable approximation of the optimal interpolator. We next apply our methodology to a finite number of data points and prove by direct calculation that our expression is equivalent to the classical one. We conclude with an example of application in two dimensions. We illustrate how, in Gaussian process models, our results can be used to take into account information of the "boundary value" type.

## 2. Main Results

### *2.1. Orthogonal Projection and Diagonalization*

Let $\mathcal{H}$ be a separable RKHS of real-valued functions on $\mathcal{X}$ with reproducing kernel $K(\cdot,\cdot)$.

Let $\mathcal{S}$ be a non-empty subset of $\mathcal{X}$. We denote by $\mathcal{H}_1$ the closure in $\mathcal{H}$ of the linear space spanned by $K_s(\cdot)$, $s \in \mathcal{S}$, with $K_s(\cdot) = K(s,\cdot)$, i.e.

$$\mathcal{H}_1 = \overline{span\left\{K_s(\cdot),\ s \in \mathcal{S}\right\}}^{\mathcal{H}}.$$

Let $\mathcal{H}_0 = \mathcal{H}_1^{\perp}$, i.e. $\mathcal{H}_0 = \{h \in \mathcal{H}\ :\ \forall s \in \mathcal{S}, h(s) = 0\ \}$.

We assume $\mathcal{S}$ endowed with a $\sigma$-algebra and let $\mu$ be a $\sigma$-finite measure on $\mathcal{S}$. We then introduce $L^2(\mu)$, the Hilbert space of square-integrable function with respect to $\mu$ ($L^2(\mu)$ is obviously a quotient space ; for a function defined on $\mathcal{X}$, we consider the integral of **the restriction** of this function to $\mathcal{S}$), note that $L^2(\mu)$ has not to be necessarily separable. Let $(\cdot|\cdot)_{L^2}$ and $\|\cdot\|_{L^2}$ be its inner product and norm; we remind that :

$$(f|g)_{L^2} = \int_{\mathcal{S}} f(s)g(s)d\mu(s).$$

Following [7, 8], we make the additional hypotheses that :

H-i. for all $x \in \mathcal{X}$, $K(x,\cdot) \in L^2(\mu)$, H-ii. the function $s \in \mathcal{S} \mapsto K(s,s)$ is $\mu$-measurable and $N = \int_{\mathcal{S}} K(s,s)d\mu(s) < \infty$,

H-iii. if $h \in \mathcal{H}$ and $h = 0$ $\mu$-a.e. on $\mathcal{S}$ then $h \in \mathcal{H}_0$.

**Remark :** Hypotheses H-i. and H-ii. imply that, for $h \in \mathcal{H}$,

$$\|h\|_{L^2}^2 = \int_{\mathcal{S}} h(s)^2 d\mu(s) = \int_{\mathcal{S}} (h|K_s)_{\mathcal{H}}^2 \leqslant \int_{\mathcal{S}} \|h\|_{\mathcal{H}}^2 K(s,s)d\mu(s) = \|h\|_{\mathcal{H}}^2 N, \text{ i.e.}$$

the inclusion of $\mathcal{H}$ in $L^2(\mu)$ is continuous. Remarking that

$$\int_{\mathcal{S}}\int_{\mathcal{S}} K(s,t)^2 d\mu(s)d\mu(t) = \int_{\mathcal{S}}\int_{\mathcal{S}} \left(K_s \middle| K_t\right)_{\mathcal{H}}^2 d\mu(s)d\mu(t) \leqslant N^2,$$

we also have $K(\cdot,\cdot) \in L^2(\mu \otimes \mu)$ (Hilbert-Schmidt kernel), that is

$$\int_{\mathcal{S}}\int_{\mathcal{S}} K(s,t)^2 d\mu(s)d\mu(t) < \infty. \tag{1}$$

Hypothesis H-iii. is similar to the notion of regular embedding that can be found in [7]. It will be always verified, for example, if $\mathcal{S}$ is a topological space (endowed with its Borel $\sigma$-algebra) and if $supp(\mu) = \mathcal{S}$ (with $supp(\mu)$ the support of $\mu$) and if $K(\cdot,\cdot)$ is continuous on $\mathcal{S} \times \mathcal{S}$ (which implies that the functions of $\mathcal{H}$ are continuous on $\mathcal{S}$). □

We now introduce the operator $T_\mu$ that we will consider, for the moment, from $L^2(\mu)$ to $\mathbb{R}^{\mathcal{X}}$. We pose :

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

**Proposition 2.1**

*For all $f \in L^2(\mu)$, $T_\mu[f] \in \mathcal{H}_1$, in addition, for all $h \in \mathcal{H}$*

$$\left(T_\mu[f] \middle| h\right)_{\mathcal{H}} = \left(f \middle| h\right)_{L^2}. \tag{2}$$

**Proof:** First, we show that for $f \in L^2(\mu)$, $T_\mu[f] \in \mathcal{H}$ : we define the linear operator $L_f : \mathcal{H} \to \mathbb{R}$ by, for $h \in \mathcal{H}$,

$$L_f[h] = \int_{\mathcal{S}} f(s)h(s)d\mu(s) = \left(f \middle| h\right)_{L^2}.$$

Using the Cauchy-Schwarz inequality and H-ii, we find

$$|L_f[h]| = \left|\left(f \middle| h\right)_{L^2}\right| \leqslant \|f\|_{L^2}\|h\|_{L^2} \leqslant \|f\|_{L^2}\|h\|_{\mathcal{H}}\sqrt{N},$$

According to Riesz representation theorem, there exists a unique $I_f \in \mathcal{H}$ such that, for all $h \in \mathcal{H}$, $L_f[h] = \left(I_f \middle| h\right)_{\mathcal{H}}$. Hence, for $x \in \mathcal{X}$,

$$I_f(x) = \left(I_f \middle| K_x\right)_{\mathcal{H}} = \int_{\mathcal{S}} K(x,t)f(t)d\mu(t) = T_\mu[f](x),$$

so $T_\mu[f] = I_f \in \mathcal{H}$. Further, for $f \in L^2(\mu)$ and $h \in \mathcal{H}$

$$\left(T_\mu[f] \middle| h\right)_{\mathcal{H}} = \int_{\mathcal{S}} \left(K_t \middle| h\right)_{\mathcal{H}} f(t)d\mu(t) = \left(f \middle| h\right)_{L^2}.$$

Now, let $h_0 \in \mathcal{H}_0$, for all $s \in \mathcal{S}$, $h_0(s) = 0$, thus for all $f \in L^2(\mu)$,

$$\left(T_\mu[f] \middle| h_0\right)_{\mathcal{H}} = \left(f \middle| h_0\right)_{L^2} = 0, \text{ i.e. } \forall f \in L^2(\mu),\ T_\mu[f] \in \mathcal{H}_0^\perp = \mathcal{H}_1.$$

■

Combining H-iii. and equation (2), we also have :

**Corollary 2.1**

*If $h \in \mathcal{H}$ verify $\forall f \in L^2(\mu)$, $\left(T_\mu[f] \middle| h\right)_{\mathcal{H}} = 0$, then $h \in \mathcal{H}_0$. Hence, the linear space spanned by the $T_\mu[f]$, $f \in L^2(\mu)$ is dense in $\mathcal{H}_1$.*

Hypotheses H-i. and H-ii. and proposition 2.1 authorize us to see $T_\mu$ as an operator from $L^2(\mu)$ to $L^2(\mu)$. Because $K(\cdot,\cdot) \in L^2(\mu \otimes \mu)$, $T_\mu$ is a Hilbert-Schmidt operator and therefore compact (see for instance [9, theorems 10.2 and 10.3]).

Because $K(\cdot,\cdot)$ is symmetric, $T_\mu$ is self-adjoint ; further $T_\mu$ is positive, apply for example (2) (one can also note that $T_\mu$ is continuous). So $T_\mu : L^2(\mu) \to L^2(\mu)$ is diagonalizable and its eigenvalues are positive. We denote $\lambda_i$ those eigenvalues and $\widetilde{\phi}_i \in L^2(\mu)$ their associated eigenfunction, $i \in \mathbb{I}$. We remind that $\left\{\widetilde{\phi}_i,\ i \in \mathbb{I}\right\}$ form a orthonormal basis of $L^2(\mu)$. Finally, we will denote by $\{\lambda_n,\ n \in \mathbb{I}_+\}$ the countable set (i.e $\mathbb{I}_+ \subseteq \mathbb{N}$) off all **strictly positive** eigenvalues with multiplicity.

**Definition 2.1**

*We call regularized eigenfunction of $T_\mu$ associated with $\lambda_n \neq 0$ (i.e. $n \in \mathbb{I}_+$) the function $\phi_n \in \mathcal{H}_1$ defined by :*

$$\forall x \in \mathcal{X},\ \phi_n(x) = \frac{1}{\lambda_n} \int_{\mathcal{S}} K(x,t)\tilde{\phi}_n(t)d\mu(t) = \frac{1}{\lambda_n} T_\mu\left[\tilde{\phi}_n\right](x).$$

**Remark :** The $\phi_n$ extend the notion of eigenfunctions $\widetilde{\phi}_n \in L^2(\mu)$ at $\mathcal{H}$ ; we obviously have, for $n \in \mathbb{I}_+$, $\phi_n \overset{L^2(\mu)}{=} \tilde{\phi}_n$ and $\forall x \in \mathcal{X}$ $T_\mu[\phi_n](x) = \lambda_n \phi_n(x)$.

In particular, this definition and equation (2) imply that for all regularized eigenfunction

$$\forall h \in \mathcal{H},\ \left(\phi_n \middle| h\right)_{\mathcal{H}} = \frac{1}{\lambda_n}\left(T_\mu[\phi_n] \middle| h\right)_{\mathcal{H}} = \frac{1}{\lambda_n}\left(\phi_n \middle| h\right)_{L^2}. \quad (3)$$

□

**Proposition 2.2**

$\left\{\sqrt{\lambda_n}\phi_n,\ n \in \mathbb{I}_+\right\}$ *is a Hilbert basis of $\mathcal{H}_1$ (for the inner product $(\cdot|\cdot)_{\mathcal{H}}$).*

**Proof:** The fact that $\left\{\sqrt{\lambda_n}\phi_n\right\}$ is an orthonormal system of $\mathcal{H}_1$ is consequence of proposition 2.1 and (3). We just have to show that *span* $\left\{\sqrt{\lambda_n}\phi_n\right\}$ is dense in $\mathcal{H}_1$ which is a consequence of corollary 2.1. ■

**Corollary 2.2**

*$\mathcal{H}_1$ endowed with $(\cdot|\cdot)_{\mathcal{H}}$ is a RKHS with reproducing kernel*

$$\forall x \in \mathcal{X},\ \forall y \in \mathcal{X},\ K_1(x,y) = \sum_{n \in \mathbb{I}_+} \lambda_n \phi_n(x)\phi_n(y).$$

*In particular,* $\forall x \in \mathcal{X},\ \forall s \in \mathcal{S},\ K(x,s) = K_1(x,s)$*. In addition, for all* $\varphi \in \mathcal{H}$*, we have*

$$\forall x \in \mathcal{X},\ \mathbb{P}_{\mathcal{H}_1}[\varphi](x) = \left(K_1(x,\cdot)\middle|\varphi\right)_{\mathcal{H}}.$$

**Remark :** Proposition 2.2 also implies that the closure of the subspace of $L^2(\mu)$ spanned by the $\widetilde{\phi}_n,\ n \in \mathbb{I}_+$ and $\mathcal{H}_1$ are isometric (in terms of Hilbert structures). The isometry is given by :

$$\forall n \in \mathbb{I}_+,\ \widetilde{\phi}_n \leftrightarrow \sqrt{\lambda_n}\phi_n.$$

In addition, $\mathcal{H}_1$ can be seen has a Hilbert subspace of $L^2(\mu)$, see [2]. Finally, one can define the square-root of $T_\mu$, that is $T_\mu^{\frac{1}{2}} : L^2(\mu) \to \mathcal{H}_1 \subset L^2(\mu)$ such that :

$$T_\mu^{\frac{1}{2}}\left[\sum_{i\in\mathbb{I}} \alpha_i \widetilde{\phi}_i\right] = \sum_{n\in\mathbb{I}_+} \alpha_n \sqrt{\lambda_n}\phi_n$$

(note that for $i \in \mathbb{I}\backslash\mathbb{I}_+$, i.e. $\lambda_i = 0$, $T_\mu^{\frac{1}{2}}\left[\widetilde{\phi}_i\right] = 0$). Obviously, $T_\mu = T_\mu^{\frac{1}{2}} \circ T_\mu^{\frac{1}{2}}$. □

Now, we can give a formula for the optimal interpolator $\mathbb{P}_{\mathcal{H}_1}[\varphi]$ which depends explicitly of the values $\varphi(s),\ s \in \mathcal{S}$ :

**Theorem 2.1**

*Let* $\varphi \in \mathcal{H}$*, we have*

$$\forall x \in \mathcal{X},\ \mathbb{P}_{\mathcal{H}_1}[\varphi](x) = \sum_{n\in\mathbb{I}_+} \phi_n(x) \int_{\mathcal{S}} \phi_n(s)\varphi(s)d\mu(s). \tag{4}$$

**Proof:** It is a simple consequence of proposition 2.2 and (3) :

$$\begin{aligned}\forall x \in \mathcal{X},\ \mathbb{P}_{\mathcal{H}_1}[\varphi](x) &= \sum_n \sqrt{\lambda_n}\phi_n(x)\left(\sqrt{\lambda_n}\phi_n\middle|\varphi\right)_{\mathcal{H}} \\ &= \sum_n \phi_n(x)\left(\phi_n\middle|\varphi\right)_{L^2}.\end{aligned}$$

■

Remark that, by definition, the equation (4) series is convergent in $\mathcal{H}$ (more precisely in $\mathcal{H}_1$). In addition, topological arguments assure that this convergence also occurs point-wisely (the Hilbert topology of $\mathcal{H}_1$ is finer than the point-wise convergence topology on $\mathcal{H}_1$, see [2]).

### *2.2. Truncated Series and Approximation*

In certain cases, for instance for numerical reasons, we will not be able to consider all terms in

$$\forall x \in \mathcal{X},\ \forall y \in \mathcal{X},\ K_1(x,y) = \sum_{n\in\mathbb{I}_+} \lambda_n\phi_n(x)\phi_n(y).$$

We assume that we dispose of an approximated kernel defined from a subset $\mathbb{I}_{app}$ of $\mathbb{I}_+$, i.e.

$$\forall x \in \mathcal{X},\ \forall y \in \mathcal{X},\ K_1^{app}(x,y) = \sum_{n\in\mathbb{I}_{app}} \lambda_n\phi_n(x)\phi_n(y).$$

Hence, for $\varphi \in \mathcal{H}$, we obtain an approximation of the optimal interpolator $\mathbb{P}_{\mathcal{H}_1}[\varphi]$ that we note $\mathbb{P}_{\mathcal{H}_1^{app}}[\varphi]$ :

$$\begin{aligned} \forall x \in \mathcal{X}, \ = \mathbb{P}_{\mathcal{H}_1^{app}}[\varphi](x) &= \left(K_1^{app}(x,\cdot)\middle|\varphi\right)_{\mathcal{H}} \\ &= \sum_{n \in \mathbb{I}_{app}} \sqrt{\lambda_n}\phi_n(x)\left(\sqrt{\lambda_n}\phi_n\middle|\varphi\right)_{\mathcal{H}}, \end{aligned}$$

with $\mathcal{H}_1^{app}$ the closure in $\mathcal{H}$ of the subspace spanned by the $\phi_n$, $n \in \mathbb{I}_{app}$.

We pose $\mathbb{I}_{err} = \mathbb{I}_+ \backslash \mathbb{I}_{app}$. Hence, we find the following expression for the approximation error :

$$\begin{aligned} \mathbb{P}_{\mathcal{H}_1}[\varphi](x) - \mathbb{P}_{\mathcal{H}_1^{app}}[\varphi](x) &= \left(\varphi\middle|K_1(x,\cdot) - K_1^{app}(x,\cdot)\right)_{\mathcal{H}} \\ &= \left(\varphi\middle|\sum_{n \in \mathbb{I}_{err}} \lambda_n \phi_n(x)\phi_n(\cdot)\right)_{\mathcal{H}}. \end{aligned} \tag{5}$$

Finally, applying the Cauchy-Schwarz inequality to equation (5) allows us to control the error of approximation :

**Theorem 2.2**

$$\forall x \in \mathcal{X}, \ \left|\mathbb{P}_{\mathcal{H}_1}[\varphi](x) - \mathbb{P}_{\mathcal{H}_1^{app}}[\varphi](x)\right|^2 \leqslant \|\varphi\|_{\mathcal{H}}^2 \sum_{n \in \mathbb{I}_{err}} \lambda_n \phi_n(x)^2 \tag{6}$$

$$\text{and} \ \left\|\mathbb{P}_{\mathcal{H}_1}[\varphi](\cdot) - \mathbb{P}_{\mathcal{H}_1^{app}}[\varphi](\cdot)\right\|_{L^2}^2 \leqslant \|\varphi\|_{\mathcal{H}}^2 \sum_{n \in \mathbb{I}_{err}} \lambda_n. \tag{7}$$

We call $\sum_{n \in \mathbb{I}_{err}} \lambda_n$ the **spectral error term**. Note that this one can be easily compute, in fact,

$$\sum_{n \in \mathbb{I}_+} \lambda_n = Trace\left(T_\mu\right) = \int_{\mathcal{S}} K(s,s)d\mu(s) \text{ and}$$

$\sum_{n \in \mathbb{I}_{app}} \lambda_n$ is suppose to be known. One just has to subtract those two quantities to obtain the spectral error term.

## 3. Illustration : Measure with Finite Support

In this section, we assume that $\mathcal{S} = \left\{\zeta^{(1)}, \cdots, \zeta^{(n)}\right\}$, a set of $n$ points of $\mathcal{X}$. We will show that in this case, the expression of the optimal interpolator given by theorem 2.1 is equal to the classical expression (see for instance [6]).

Let $\mathcal{H}$ be a RKHS with kernel $K(\cdot,\cdot)$ and denote by :

$$\forall x \in \mathcal{X}, \ \mathbf{k}(x) = \left(K(x,\zeta^{(1)}), \cdots, K(x,\zeta^{(n)})\right)^T \in \mathbb{R}^{n \times 1},$$

$$\mathbf{K} \in \mathbb{R}^{n \times n} \text{ with } \mathbf{K}_{i,j} = K(\zeta^{(i)}, \zeta^{(j)}), \ 1 \leqslant i,j \leqslant n.$$

For simplicity, we assume that $\mathcal{S}$ and $K(\cdot,\cdot)$ are such that the matrix $\mathbf{K}$ is invertible ($\mathbf{K}$ is by definition symmetric and positive). Let $\varphi \in \mathcal{H}$ and pose $\mathbf{y} = \left(\varphi(\zeta^{(1)}),\cdots,\varphi(\zeta^{(n)})\right)^T \in \mathbb{R}^{n\times 1}$. We have the well-known expression :

$$\forall x \in \mathcal{X},\ \mathbb{P}_{\mathcal{H}_1}\left[\varphi\right](x) = \mathbf{k}(x)^T\mathbf{K}^{-1}\mathbf{y}.$$

Note that $\mathbf{k}(x)^T\mathbf{K}^{-1}\mathbf{y}$ is well defined for all $\mathbf{y} \in \mathbb{R}^{n\times 1}$.

Define $\mu$ as a sum of Dirac at points $\zeta^{(i)}$, that is, for all functions $f$ well defined on $\mathcal{S}$,

$$T_\mu\left[f\right](x) = \int_{\mathcal{S}} K(x,t)f(t)d\mu(t) = \sum_{i=1}^{n} K(x,\zeta^{(i)})f(\zeta^{(i)}).$$

We handle this special case because it is the more convenient for calculation. The general case is treated in Appendix A. Straightforward calculations show that the eigenvalues of $T_\mu$ are the same of the ones of the matrix $\mathbf{K}$. Further, if we note $\mathbf{v}_i$ the eigenvector of $\mathbf{K}$ associated with $\lambda_i$, the regularized eigenfunction $\phi_i$ of $T_\mu$ associated with $\lambda_i$ is

$$\forall x \in \mathcal{X},\ \phi_i(x) = \frac{1}{\lambda_i}\mathbf{k}(x)^T\mathbf{v}_i = \mathbf{k}(x)^T\mathbf{K}^{-1}\mathbf{v}_i.$$

We pose $\mathbf{\Lambda} = diag\left(\lambda_1,\cdots,\lambda_n\right)$ and $\mathbf{P}$ the matrix which $i$-th column is $\mathbf{v}_i$, hence $\mathbf{K} = \mathbf{P\Lambda P}^T$. Remarking that the function $\phi_i$ take the value $\left(\mathbf{v}_i\right)_k$ (the $k$-th component of $\mathbf{v}_i$) at point $\zeta^{(k)}$, we found that

$$\int_{\mathcal{S}} \varphi(s)\phi_i(s)d\mu(s) = \left(\mathbf{P}^T\mathbf{y}\right)_i.$$

In addition, $\left(\phi_1(x),\cdots,\phi_n(x)\right)^T = \mathbf{k}(x)^T\mathbf{P\Lambda}^{-1}$. Finally, for all $x \in \mathcal{X}$,

$$\sum_{i=1}^{n}\phi_i(x)\int_{\mathcal{S}}\phi_i(s)\varphi(s)d\mu(s) = \mathbf{k}(x)^T\mathbf{P\Lambda}^{-1}\mathbf{P}^T\mathbf{y} = \mathbf{k}(x)^T\mathbf{K}^{-1}\mathbf{y}.$$

## 4. Example of Application

Let $\mathcal{X} = \mathbb{R}^2$ and $\mathcal{H}$ be the RKHS associated with the kernel (Gaussian kernel)

$$K(x,y) = e^{-\frac{\|x-y\|^2}{\sigma^2}},\quad \text{with } \sigma > 0 \text{ and } \|\cdot\| \text{ the euclidean norm.}$$

We assume that $\mathcal{S}$ is the circle with center 0 and radius $R > 0$ endowed with its natural Lebesgue measure.

We note $x = (x_1, x_2)$ a point of $\mathcal{X}$, for convenience, we will use a polar coordinate system to describe $T_\mu$, that gives :

$$T_\mu[f](x) = \int_0^{2\pi} K\left(x, s_{R,\theta}\right) f(s_{R,\theta})Rd\theta,\ \text{with } s_{R,\theta} = (R\cos\theta, R\sin\theta).$$

Straightforward calculations (see Appendix B) give that the eigenvalues of $T_\mu$ are :

$$n \geqslant 0,\ \lambda_n = Re^{-\frac{2R^2}{\sigma^2}}\int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos\theta}\cos(n\theta)d\theta.$$

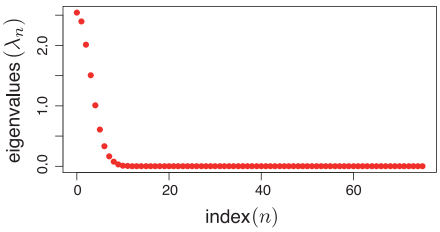


Figure 1: Graphical representation of $T_\mu$ spectrum, $0 \leqslant n \leqslant 74$, $R = 3$ and $\sigma^2 = 2$.

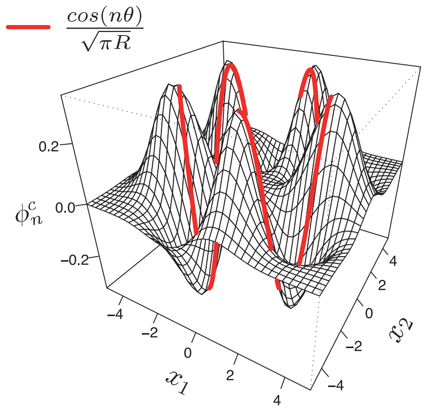


Figure 2: Eigenfunction $\phi_n^c$ on $[-5, 5]^2$ for $n$ =5.

$\lambda_0$ is of multiplicity one and its associated regularized eigenfunction is :

$$\phi_0(x) = \frac{1}{\lambda_0} \int_0^{2\pi} K(x, s_{R,\theta}) \frac{1}{\sqrt{2\pi R}} R d\theta.$$

For $n \geqslant 1$, the $\lambda_n$ are of multiplicity two and associated with the two orthogonal regularized eigenfunctions :

$$\phi_n^c(x) = \frac{1}{\lambda_n} \int_0^{2\pi} K(x, s_{R,\theta}) \frac{\cos n\theta}{\sqrt{\pi R}} R d\theta \text{ and } \phi_n^s(x) = \frac{1}{\lambda_n} \int_0^{2\pi} K(x, s_{R,\theta}) \frac{\sin n\theta}{\sqrt{\pi R}} R d\theta.$$

Remark that a generalization of this result to any dimension and other example of spectral decomposition can be found in [10].

*Numerical application.* We set $R = 3$ and $\sigma^2 = 2$, we next compute the eigenvalues and regularized eigenfunctions of $T_\mu$. The obtained results are represented on figures 1 and 2.

Following section 2.2, we approximate the kernel $K_1(\cdot, \cdot)$ with the 31 eigenfunctions associated with the most important eigenvalues, that is, for all $x$ and $t \in \mathbb{R}^2$ :

$$K_1^{app}(x, t) = \lambda_0 \phi_0(x) \phi_0(t) + \sum_{n=1}^{15} \lambda_n \left[\phi_n^c(x) \phi_n^c(t) + \phi_n^s(x) \phi_n^s(t)\right].$$

We compute

$$Trace\,(T_\mu) = 2\pi R = \texttt{18.84956} \text{ and } \sum_{k \in \mathbb{I}_{app}} \lambda_k = \lambda_0 + 2 \sum_{n=1}^{15} \lambda_n = \texttt{18.84953}.$$

Hence, we obtain for the spectral error term :

$$\sum_{k \in \mathbb{I}_{err}} \lambda_k = \texttt{2.643289e-05}. \tag{8}$$

We next approximate the kernel $K_0(\cdot,\cdot) = K(\cdot,\cdot) - K_1(\cdot,\cdot)$ of the sub-RKHS $\mathcal{H}_0$ with, for all $x$ and $t \in \mathbb{R}^2$,

$$K_0^{app}(x,t) = K(x,t) - K_1^{app}(x,t).$$

$\mathcal{H}_0$ is the subspace of functions of $\mathcal{H}$ that vanish onto $\mathcal{S}$. In particular, this implies that, for all $s \in \mathcal{S}$, $K_0(s,s) = 0$. In our application case, we obtain `1.402314e-06` for the mean value of $K_0^{app}(s,s)$, $s \in \mathcal{S}$ (this mean is performed on 100 points uniformly distributed on $\mathcal{S}$). Note that in our example, $\mu$ is the uniform measure on $\mathcal{S}$ with total mass $2\pi R$. Thus, the value $K_0^{app}(s,s)$ =`1.402314e-06` for $s \in \mathcal{S}$ is directly linked with the error term given in (8), in fact :

$$\sum_{k \in \mathbb{I}_{err}} \lambda_k = \int_{\mathcal{S}} K_0^{app}(s,s) d\mu(s) \text{ and}$$

$\sum_{k \in \mathbb{I}_{err}} \lambda_k / 2\pi R$ =`1.402308e-06`, which is very closed to the preceding value (the error is due to the Monte-Carlo method used to compute the regularized spectral decomposition of $T_\mu$).

Figure 3 shows the sample path of a centered Gaussian process with covariance $K_0^{app}(\cdot,\cdot)$. This approximates a centered Gaussian process of covariance $K_0(\cdot,\cdot)$ which corresponds to a centered Gaussian process of covariance $K(\cdot,\cdot)$ conditioned to be null on $\mathcal{S}$.

Finally, we use the kernel $K_0^{app}(\cdot,\cdot)$ in a classical Gaussian process model (or kriging model, i.e. Gaussian process conditioning relatively to a finite number of data points). The obtained model interpolates exactly the set of classical data and approximates the condition on $\mathcal{S}$ (here, to vanish on the circle). Figure 4 represents the best predictor of a Gaussian process model with kernel $K_0^{app}(\cdot,\cdot)$ (i.e. a conditional expectation) and figure 5 shows a conditional realization (i.e. a conditional simulation).

## Acknowledgments

We would like to thank all members of the 3MI team and more particularly Nicolas Durrande for their helpful suggestions. All graphics and computations have been performed with the free software `R`.

## References

[1] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 63:337–404, 1950.

[2] L. Schwartz. Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés. *J. Anal. Math.*, 13:115–256, 1964.

[3] C. A. Micchelli and T. J. Rivlin. *Numerical Analysis Lancaster 1984*, chapter Lectures on optimal recovery, pages 21–93. Springer Berlin / Heidelberg, 1985.

[4] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, 1997.

[5] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[6] G. Wahba. *Spline Models for Observational Data*. SIAM, 1990.

[7] R. M. Fortet. Les operateurs intégraux dont le noyau est une covariance. *Trabajos de Estadstica y de Investigacin Operativa*, 36:133–144, 1985.

[8] H. Sun and Q. Wu. Application of integral operator for regularized least-square regression. *Mathematical and Computer Modelling*, 49:276–285, 2009.

[9] Laurent Schwartz. *Analyse Hilbertienne*. Hermann, 1979.

[10] P. Niyogi H. Q. Minh and Y. Yao. *Learning Theory*, chapter Mercers Theorem, Feature Maps, and Smoothing, pages 154–168. Springer Berlin / Heidelberg, 2006.

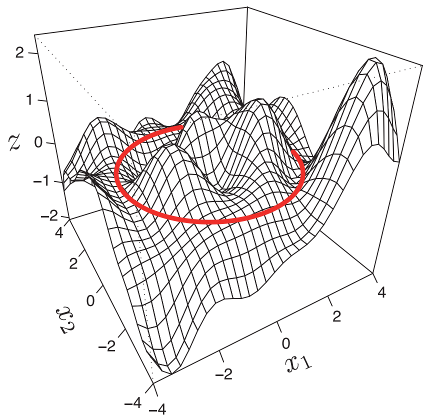


Figure 3: Simulation with $K_0^{app}(\cdot,\cdot)$ on $[-4,4]^2$.

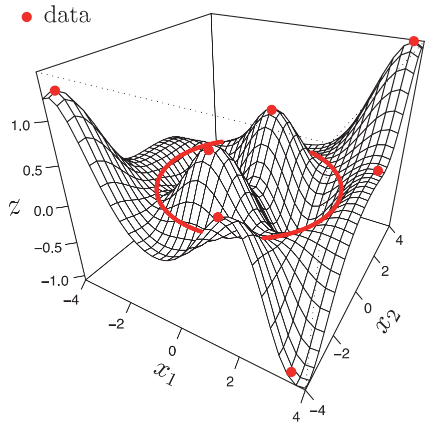


Figure 4: Best predictor of a kriging model with kernel $K_0^{app}(\cdot,\cdot)$.

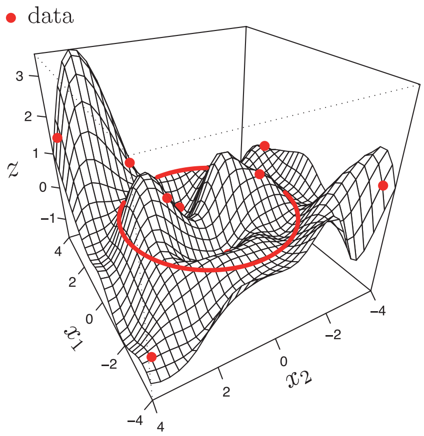


Figure 5: Conditional realization of a kriging model with kernel $K_0^{app}(\cdot,\cdot)$.

## Appendix A. Generalization of the Section 3 Calculus

Let $m_1, \cdots, m_n$ be $n$ positive reals such that

$$T_\mu[f](x) = \int_{\mathcal{S}} K(x,t)f(t)d\mu(t) = \sum_{i=1}^{n} m_i K\left(x, \zeta^{(i)}\right) f\left(\zeta^{(i)}\right),$$

We pose $\mathbf{M} = diag(m_1, \cdots, m_n)$. Searching the eigenfunctions under the form $x \mapsto \sum_{k=1}^{n} \alpha_k K\left(x, \zeta^{(i)}\right)$, we find that the eigenvalues of $T_\mu$ are the same of the ones of the matrix $\mathbf{MK}$. Let $\lambda_1, \cdots, \lambda_n$ be those strictly positive eigenvalues and let $\mathbf{v}_1, \cdots, \mathbf{v}_n$ be their associated eigenvector (for $\mathbf{MK}$). Like in section 3, we denote $\mathbf{\Lambda} = diag(\lambda_1, \cdots, \lambda_n)$ and $\mathbf{P}$ the matrix which $i$-th column is $\mathbf{v}_i$, thus $\mathbf{MK} = \mathbf{P\Lambda P}^{-1}$ (contrary to $\mathbf{K}$, $\mathbf{MK}$ is not necessarily symmetric) with $\mathbf{P} = [\mathbf{v}_1|\cdots|\mathbf{v}_n]$.

The (not normed) eigenfunction of $T_\mu$ associated with $\lambda_k$ is of the form

$$\forall x \in \mathcal{X},\ \hat{\phi}_k(x) = \mathbf{k}(x)^T \mathbf{v}_k,\ \text{with}$$

$$\left(\hat{\phi}_k | \hat{\phi}_l\right)_{L^2} = \mathbf{v}_k^T \mathbf{KMK} \mathbf{v}_l.$$

Because the $\hat{\phi}_k$ are orthogonal in $L^2(\mu)$, the matrix

$$\mathbf{P}^T\mathbf{KMKP} = \mathbf{P}^T\mathbf{KP\Lambda} = \mathbf{\Lambda P}^T\mathbf{KP} = \mathbf{\Lambda P}^T\mathbf{M}^{-1}\mathbf{P\Lambda}$$

is diagonal, its diagonal elements being $\left\|\hat{\phi}_k\right\|_{L^2}^2$. Then, one just has to remark that

$$\begin{aligned} \mathbf{k}(x)^T\mathbf{K}^{-1}\mathbf{y} &= \mathbf{k}(x)^T\mathbf{K}^{-1}\mathbf{M}^{-1}\mathbf{My} = \mathbf{k}(x)^T\mathbf{P\Lambda}^{-1}\mathbf{P}^{-1}\mathbf{My} \\ &= \mathbf{k}(x)^T\mathbf{P\Lambda}^{-1}\mathbf{P}^{-1}\mathbf{K}^{-1}\mathbf{P}^{-T}\mathbf{P}^T\mathbf{KMy} \\ &= \mathbf{k}(x)^T\mathbf{P}\left(\mathbf{P}^T\mathbf{KMKP}\right)^{-1}\mathbf{P}^T\mathbf{KMy} \text{ and} \end{aligned}$$

$$\int_{\mathcal{S}} \varphi(s)\hat{\phi}_k(s)d\mu(s) = \left(\mathbf{P}^T\mathbf{KMy}\right)_k \text{ and}$$

$$\left(\hat{\phi}_1(x), \cdots, \hat{\phi}_n(x)\right)^T = \mathbf{k}(x)^T\mathbf{P}.$$

Finally, denoting by $\phi_k$ the regularized eigenfunctions of $T_\mu$, we find for all $x \in \mathcal{X}$,

$$\sum_{k=1}^{n} \phi_k(x) \int_{\mathcal{S}} \phi_k(s)\varphi(s)d\mu(s) = \mathbf{k}(x)^T\mathbf{K}^{-1}\mathbf{y}.$$

## Appendix B. Spectral Decomposition of the Section 4 Integral Operator

Let $x = (\|x\|\cos\alpha_x, \|x\|\sin\alpha_x)$, and $y = (\|y\|\cos\alpha_y, \|y\|\sin\alpha_y)$ be two points of $\mathcal{X} = \mathbb{R}^2$,

$$K(x,y) = e^{-\frac{\|x-y\|^2}{\sigma^2}} = e^{-\frac{1}{\sigma^2}\left(\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos(\alpha_x - \alpha_y)\right)}.$$

Consider the integral operator :

$$T_\mu[f](x) = \int_0^{2\pi} K\left(x, s_{R,\theta}\right) f(s_{R,\theta}) R d\theta,\ \text{with } s_{R,\theta} = (R\cos\theta, R\sin\theta).$$

Let $x \in \mathcal{S}$, i.e. $\|x\| = R$, using an argument of periodicity, we find :

$$\begin{aligned} \int_0^{2\pi} K\left(x, s_{R,\theta}\right) R d\theta &= R e^{-\frac{2R^2}{\sigma^2}} \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\theta-\alpha_x)} d\theta \\ &= R e^{-\frac{2R^2}{\sigma^2}} \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos\theta} d\theta, \end{aligned}$$

which does not depend of $\alpha_x$. Hence, the function $\widetilde{\phi}_0 : x = (R\cos\alpha_x, R\sin\alpha_x) \mapsto \frac{1}{\sqrt{2\pi R}}$ is an eigenfunction of $T_\mu$ associated with $\lambda_0 = R e^{-\frac{2R^2}{\sigma^2}} \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos\theta} d\theta$.

In the same way, and adding arguments of parity, we obtain

$$\int_0^{2\pi} K\left(x, s_{R,\theta}\right) \cos(n\theta) R d\theta = R e^{-\frac{2R^2}{\sigma^2}} \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\theta-\alpha_x)} \cos(n\theta) d\theta \text{ and}$$

$$\begin{aligned} \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\theta-\alpha_x)} \cos(n\theta) d\theta &= \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\omega)} \cos\left(n(\omega+\alpha_x)\right) d\omega \\ &= \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\omega)} \left(\cos(n\omega)\cos(n\alpha_x) - \sin(n\omega)\sin(n\alpha_x)\right) d\omega \\ &= \cos(n\alpha_x) \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\omega)} \cos(n\omega) d\omega. \end{aligned}$$

Thus, the function $\widetilde{\phi}_n^c : x = (R\cos\alpha_x, R\sin\alpha_x) \mapsto \frac{\cos(n\alpha_x)}{\sqrt{\pi R}}$ is an eigenfunction of $T_\mu$ associated with $\lambda_n = R e^{-\frac{2R^2}{\sigma^2}} \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos\theta} \cos(n\theta) d\theta$.

Finally, the equality

$$\begin{aligned} \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\theta-\alpha_x)} \sin(n\theta) d\theta &= \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\omega)} \sin\left(n(\omega+\alpha_x)\right) d\omega \\ &= \sin(n\alpha_x) \int_0^{2\pi} e^{\frac{2R^2}{\sigma^2}\cos(\omega)} \cos(n\omega) d\omega \end{aligned}$$

shows that the eigenvalue $\lambda_n$ is also associated with the eigenfunction

$$\widetilde{\phi}_n^s : x = (R\cos\alpha_x, R\sin\alpha_x) \mapsto \frac{\sin(n\alpha_x)}{\sqrt{\pi R}}.$$

To conclude, we just have to remark that $L^2(\mu)$ is isometric the to well-known Hilbert space of $2\pi$-periodic square integrable functions relatively to the Lebesgue measure. This assures us that $T_\mu$ does not admit other eigenfunctions.