



HAL
open science

Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation

Xavier Bay, Laurence Grammont, Hassan Maatouk

► **To cite this version:**

Xavier Bay, Laurence Grammont, Hassan Maatouk. Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electronic Journal of Statistics*, 2016, 10 (1), pp.1580-1595. 10.1214/16-ejs1149 . hal-04150103

HAL Id: hal-04150103

<https://hal.science/hal-04150103>

Submitted on 4 Jul 2023

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.

Generalization of the Kimeldorf-Wahba correspondence for *constrained* interpolation

Xavier Bay

École des Mines de St-Étienne, 158 Cours Fauriel, 42 023 St-Étienne, France
e-mail: bay@emse.fr
url: www.emse.fr

Laurence Grammont

*Université de Lyon, Institut Camille Jordan, UMR 5208, 23 rue du Dr Paul Michelon,
42023 St-Étienne Cedex 2, France*
e-mail: laurence.grammont@univ-st-etienne.fr

and

Hassan Maatouk*

École des Mines de St-Étienne, 158 Cours Fauriel, 42 023 St-Étienne, France
e-mail: hassan.maatouk@mines-stetienne.fr

Abstract: In this paper, we extend the correspondence between Bayes' estimation and optimal interpolation in a Reproducing Kernel Hilbert Space (RKHS) to the case of convex constraints such as boundedness, monotonicity or convexity. In the unconstrained interpolation case, the mean of the posterior distribution of a Gaussian Process (GP) given data interpolation is known to be the optimal interpolation function minimizing the norm in the RKHS associated to the GP. In the *constrained* case, we prove that the Maximum *A Posteriori* (MAP) or Mode of the posterior distribution is the optimal *constrained* interpolation function in the RKHS. So, the general correspondence is achieved with the MAP estimator and *not* the mean of the posterior distribution. A numerical example is given to illustrate this last result.

MSC 2010 subject classifications: 60G15, 60G25, 62K20.

Keywords and phrases: Correspondence, interpolation, inequality constraints, reproducing Kernel Hilbert space, Gaussian process, Bayesian estimation, maximum a posteriori.

Received February 2016.

Contents

1	Introduction	1581
2	The natural correspondence for finite-dimensional Gaussian processes	1582

*Corresponding author.

3 The main result 1585
 3.1 Approximation of the Gaussian process Y 1585
 3.2 Asymptotic analysis 1587
 3.3 Technical lemmas 1591
 4 Numerical illustration 1592
 5 Conclusion 1594
 Acknowledgements 1594
 References 1594

1. Introduction

Consider a function y defined on a nonempty set X of \mathbb{R}^d ($d \geq 1$). The curve-fitting problem is to estimate y using *a priori* information and a finite set of noise-free evaluations:

$$y(x^{(i)}) = y_i, \quad i = 1, \dots, n,$$

where $x^{(1)}, \dots, x^{(n)}$ are n distinct points of X . As in [4], the *prior* information is summarized by a zero-mean Gaussian Process (GP) $\{Y(x)\}_{x \in X}$ with covariance function

$$K(x, x') := \mathbb{E}(Y(x)Y(x')), \tag{1.1}$$

where \mathbb{E} denotes expectation. In this case, the usual Bayesian estimator \hat{y} of y is the mean of the posterior distribution of the GP $\{Y(x)\}_{x \in X}$ given data:

$$\hat{y}(x) := \mathbb{E}\left(Y(x) \mid Y(x^{(1)}) = y_1, \dots, Y(x^{(n)}) = y_n\right).$$

From [11], we have the following explicit expression for \hat{y} :

$$\hat{y}(x) = \mathbf{k}(x)^\top \mathbb{K}^{-1} \mathbf{y}, \quad x \in X, \tag{1.2}$$

where $\mathbf{k}(x) = (K(x, x^{(1)}), \dots, K(x, x^{(n)}))^\top$, \mathbb{K} is the matrix $(K(x^{(i)}, x^{(j)}))_{1 \leq i, j \leq n}$ and $\mathbf{y} = (y_1, \dots, y_n)^\top$.

On the other hand, it is well known (see [4]) that this estimation function (1.2) is the unique solution of the following optimization problem:

$$\min_{h \in H \cap I} \|h\|_H^2, \tag{Q}$$

where H is the Reproducing Kernel Hilbert Space (see [1]) associated to the positive definite kernel K defined by (1.1) and I is the set of interpolant functions:

$$I := \left\{ f \in \mathbb{R}^X : f(x^{(i)}) = y_i, i = 1, \dots, n \right\}. \tag{1.3}$$

This result will be referred to as the correspondence between Bayes' estimation and optimal interpolation in a RKHS or Kimeldorf-Wahba correspondence.

Now, we suppose that the function y is known to satisfy some properties or constraints such as boundedness, monotonicity or convexity. Formally, let C be a closed convex set of \mathbb{R}^X (equipped with the product topology or topology of

pointwise convergence) corresponding to such constraints. For instance, C is of the form:

$$C = \{f \in \mathbb{R}^X : -\infty \leq a \leq f(x) \leq b \leq +\infty, x \in X\} \quad (\text{boundedness}),$$

$$C = \{f \in \mathbb{R}^X : \forall x \leq x', f(x) \leq f(x')\} \quad (\text{monotonicity}),$$

$$C = \{f \in \mathbb{R}^X : \forall \lambda \in [0, 1], \forall x, x', f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x')\} \\ (\text{convexity}).$$

If $H \cap C \cap I \neq \emptyset$, the following convex optimization problem:

$$\min_{h \in H \cap C \cap I} \|h\|_H^2 \quad (P)$$

has a unique solution denoted by h_{opt} (see e.g. [8, 13]), which can be seen as the optimal *constrained* interpolation function associated to the knots $x^{(i)}$, $i = 1, \dots, n$.

In the Bayesian framework, the problem is now to make inference from the conditional distribution of the GP $\{Y(x)\}_{x \in X}$ given $Y \in C$ (*prior* information) and given data $Y(x^{(i)}) = y_i$, $i = 1, \dots, n$. This conditional distribution can be thought as a truncated multivariate normal distribution but in an infinite dimensional linear space.

The aim of this paper is to prove that the *constrained* interpolation function h_{opt} solution of problem (P) is the Maximum *A Posteriori* (MAP) or mode of this posterior distribution $\{Y \mid Y \in C \cap I\}$.

The paper is organized as follows: in Section 2, we consider the finite-dimensional case to get insight into the natural correspondence between *constrained* interpolation functions and Bayes' estimators. Section 3 is devoted to the main result. We approximate the original Gaussian process by a sequence of finite-dimensional Gaussian processes (see e.g. [6, 10, 12]). The MAP estimator of the finite-dimensional approximation process is well defined. Furthermore, this sequence of MAP estimators is shown to be convergent to the optimal *constrained* interpolation function solution of problem (P). As a consequence, we can interpret h_{opt} as the most likely function or mode of the posterior distribution $\{Y \mid Y \in C \cap I\}$. This result can be seen as a generalization of the Kimeldorf-Wahba correspondence in the case of curve-fitting (interpolation case) taking into account convex constraints. This new correspondence is illustrated in Section 4.

2. The natural correspondence for finite-dimensional Gaussian processes

In this section, we assume that $\{Y^m(x)\}_{x \in X}$ is a finite-dimensional or degenerate GP in the sense that:

$$Y^m(x) := \sum_{j=1}^m \xi_j \phi_j(x), \quad x \in X, \quad (2.1)$$

where $\{\phi_j, 1 \leq j \leq m\}$ is a set of m linearly independent functions in \mathbb{R}^X and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^\top \in \mathbb{R}^m$ is a zero-mean Gaussian vector with covariance matrix Γ_m assumed to be invertible. The covariance function of Y^m can be expressed as

$$K_m(x, x') = \phi(x)^\top \Gamma_m \phi(x'), \tag{2.2}$$

where $\phi(x) = (\phi_1(x), \dots, \phi_m(x))^\top$. Let

$$H_m := \text{Vect} \{\phi_j, 1 \leq j \leq m\} = \left\{ h \in \mathbb{R}^X : \exists (c_1, \dots, c_m) \in \mathbb{R}^m, h = \sum_{j=1}^m c_j \phi_j \right\} \tag{2.3}$$

be the linear space spanned by the basis functions ϕ_j and consider on H_m the dot product $(h_1, h_2)_m = c_{h_1}^\top \Gamma_m^{-1} c_{h_2}$, where c_{h_i} are the coordinates of h_i with respect to the basis $\{\phi_1, \dots, \phi_m\}$, $i = 1, 2$. Since $\Gamma_m \phi(x)$ is the vector of coordinates of $K_m(\cdot, x) \in H_m$ (see equation (2.2)), we have

$$(h, K_m(\cdot, x))_m = c_h^\top \Gamma_m^{-1} \Gamma_m \phi(x) = c_h^\top \phi(x) = h(x).$$

Hence, $(H_m, (\cdot, \cdot)_m)$ is the RKHS with reproducing kernel K_m .

In the following proposition, we consider C a closed convex subset of \mathbb{R}^X (for product topology) and $I := \{f \in \mathbb{R}^X : f(x^{(i)}) = y_i, i = 1, \dots, n\}$ the set

of interpolant functions. We denote by $\widehat{H_m \cap C}$ the interior of $H_m \cap C$ in the finite-dimensional space H_m . While the *prior* $\{Y^m \mid Y^m \in C\}$ is defined by its density with respect to the Lebesgue measure λ_m on the m -dimensional space H_m , the posterior distribution of $\{Y^m \mid Y^m \in C \cap I\}$ will be defined by its density with respect to the $(m - n)$ -dimensional measure volume on the affine subspace $H_m \cap I$ of H_m . Hence, the MAP estimator is defined as the maximum (or mode) of the posterior distribution density.

Proposition 1. *Let $\{Y^m(x)\}_{x \in X}$ be a process of the form (2.1) and H_m defined by (2.3) be the RKHS associated with the kernel function K_m given in (2.2). Let*

us assume that $\widehat{H_m \cap C} \cap I$ is nonempty. Then, the MAP estimator \hat{y}_m defined as the mode of the posterior distribution density of $\{Y^m \mid Y^m \in C \cap I\}$ is equal to the constrained interpolation function $h_{opt,m}$ solution of

$$\arg \min_{h \in \widehat{H_m \cap C} \cap I} \|h\|_m^2.$$

Proof. Remark that the sample paths of Y^m are in H_m by definition. Hence, it makes sense to define the density of Y^m with respect to the uniform reference measure λ_m on H_m (m -dimensional volume measure or Lebesgue measure). This density is defined up to a multiplicative constant and to give it an explicit expression, we consider the following linear isomorphism:

$$i : c \in \mathbb{R}^m \mapsto h := \sum_{j=1}^m c_j \phi_j \in H_m.$$

We can define the measure λ_m on H_m as the image measure $\lambda_m := i(dc)$, where $dc = dc_1 \times \dots \times dc_m$ is the m -dimensional volume measure in \mathbb{R}^m . So, if $B \in \mathcal{B}(H_m)$ is a Borelian subset of H_m , we have

$$\lambda_m(B) = \int_{\mathbb{R}^m} \mathbb{1}_{i^{-1}(B)}(c) dc_1 \times \dots \times dc_m.$$

To calculate the probability density function (pdf) of Y^m , we write

$$\mathbb{P}(Y^m \in B) = \mathbb{P}(\xi \in i^{-1}(B)).$$

Using the fact that ξ is a zero-mean Gaussian vector $\mathcal{N}(0, \Gamma_m)$, we obtain

$$\begin{aligned} \mathbb{P}(Y^m \in B) &= \int_{\mathbb{R}^m} \mathbb{1}_{i^{-1}(B)}(c) \frac{1}{\sqrt{2\pi}^m |\Gamma_m|^{1/2}} \exp\left(-\frac{1}{2} c^\top \Gamma_m^{-1} c\right) dc \\ &= \int_{\mathbb{R}^m} \mathbb{1}_B(i(c)) \frac{1}{\sqrt{2\pi}^m |\Gamma_m|^{1/2}} \exp\left(-\frac{1}{2} \|i(c)\|_m^2\right) dc. \end{aligned}$$

By the transfer formula, we get

$$\mathbb{P}(Y^m \in B) = \int_{H_m} \mathbb{1}_B(h) \frac{1}{\sqrt{2\pi}^m |\Gamma_m|^{1/2}} \exp\left(-\frac{1}{2} \|h\|_m^2\right) d\lambda_m(h).$$

Hence, the (unconstrained) density of Y^m with respect to λ_m is the function

$$h \in H_m \mapsto \frac{1}{\sqrt{2\pi}^m |\Gamma_m|^{1/2}} \exp\left(-\frac{1}{2} \|h\|_m^2\right).$$

Let us now introduce the inequality constraints described by the convex set C . In the Bayesian framework, the *prior* is the following truncated pdf (with respect to λ_m):

$$h \in H_m \mapsto k^{-1} \mathbb{1}_{(h \in H_m \cap C)} \exp\left(-\frac{1}{2} \|h\|_m^2\right),$$

where $k \neq 0$ (since $\widehat{H_m \cap C} \neq \emptyset$) is a normalizing constant. Assume $\widehat{H_m \cap C} \cap I$ is nonempty, the posterior likelihood L_{pos} defined as the pdf of Y^m given data interpolation, is given by

$$L_{pos}(h) = k^{-1} \mathbb{1}_{(h \in H_m \cap C \cap I)} \exp\left(-\frac{1}{2} \|h\|_m^2\right), \quad (2.4)$$

where k is a different normalizing constant. This density function L_{pos} is defined with respect to the $(m-n)$ -dimensional measure volume induced by λ_m on the affine subspace $H_m \cap I$ of H_m . In fact, using the hypothesis $\widehat{H_m \cap C} \cap I \neq \emptyset$, there exists an open ball $\mathcal{B}(h_0, r)$ in $H_m \cap C$, where $h_0 \in I$ and $r > 0$. Hence, $\mathcal{B}(h_0, r) \cap I$ is also an open ball for the affine subspace $H_m \cap I$. Hence, the

measure of $H_m \cap C \cap I$ with respect to the $(m - n)$ -dimensional measure of $H_m \cap I$ is not zero. Thus, $k \neq 0$.

By definition, the MAP estimator \hat{y}_m is the solution of the following optimization problem

$$\arg \max L_{pos}(h) = \arg \min (-2 \log L_{pos}(h)).$$

From expression (2.4), the MAP estimator \hat{y}_m is the *constrained* interpolation function $h_{opt,m}$ solution of

$$\arg \min_{h \in H_m \cap C \cap I} \|h\|_m^2. \quad \square$$

3. The main result

In a Bayesian statistical framework, the *prior* is the probability distribution of a zero-mean GP $\{Y(x)\}_{x \in X}$ with covariance function K defined by (1.1) and assumed to be definite. We suppose that the sample paths of Y are in the Banach space $E = C^0(X)$, the set of continuous functions defined on a compact set X . As a consequence, the process Y is mean-square continuous and the covariance kernel K is continuous on $X \times X$. For the sake of simplicity, we suppose that $X = [0, 1]$. The results presented in this paper can be generalized to the multi-dimensional case. Let H be the RKHS associated to the positive definite function K . Then, H is an Hilbertian subspace of E since

$$\|h\|_E = \sup_{x \in X} |(h, K(\cdot, x))_H| \leq c \|h\|_H,$$

where $c = \sup_{x \in X} K(x, x)^{1/2} < +\infty$ by continuity of the kernel function K . Here, we suppose that we have also a *prior* information such as boundedness, monotonicity or convexity constraints. Assume that these properties are mathematically described by the set C , where C is a closed convex subset of \mathbb{R}^X as in Section 2 (a fortiori, $C \cap E$ is also a closed convex set of E)¹. Finally, let I be the set of data interpolating functions $I = \{f \in E : f(x^{(i)}) = y_i, i = 1, \dots, n\}$. Our aim is to make inference from the posterior distribution of the Gaussian process Y , so we need to handle the conditional distribution

$$\left\{ Y \mid Y \in C \text{ and } Y(x^{(i)}) = y_i, i = 1, \dots, n \right\}.$$

3.1. Approximation of the Gaussian process Y

Keeping in mind Section 2, we approximate the GP Y by the following finite-dimensional Gaussian process:

$$Y^N(x) := \sum_{j=0}^N Y(t_{N,j}) \phi_{N,j}(x), \quad x \in X, \quad (3.1)$$

¹The application $f \in E \rightarrow f \in \mathbb{R}^X$ is continuous since uniform convergence implies pointwise convergence.

where $0 = t_{N,0} \leq t_{N,1} \leq \dots \leq t_{N,N} = 1$ is a graded subdivision of $X = [0, 1]$ such that $\delta_N = \max\{|t_{N,j+1} - t_{N,j}|, j = 0, \dots, N-1\} \xrightarrow{N \rightarrow +\infty} 0$ and $\phi_{N,j}$ are the associated piecewise linear functions (or hat functions) such that $\phi_{N,j}(t_{N,i}) = \delta_{ij}$, $0 \leq i, j \leq N$, where δ_{ij} is the Kronecker's Delta function. By continuity of the sample paths of Y and continuous piecewise linear approximation in the Banach space $E = \mathcal{C}([0, 1])$, Y^N converges uniformly to the original GP Y when N tends to infinity with probability one.

To simplify the proof of the main result (see Theorem 3.2 below), block matrix structures will be used. To get this structure, we rename the knots of the partition $\Delta_N = \{t_0, \dots, t_N\}$ such that

$$\Delta_{N+1} = \Delta_N \cup \{t_{N+1}\}. \tag{3.2}$$

The finite-dimensional approximation of Gaussian Processes (GPs) can be rewritten as

$$Y^N(x) := \sum_{j=0}^N Y(t_j)\varphi_{N,j}(x),$$

where $\varphi_{N,j}$ is the hat function associated to the knot t_j .

From Section 2, Y^N is a finite-dimensional GP with covariance function

$$K_N(x, x') = \sum_{k,\ell=0}^N K(t_k, t_\ell)\varphi_{N,k}(x)\varphi_{N,\ell}(x') = \varphi_N(x)^\top \Gamma_N \varphi_N(x'),$$

where $\Gamma_N := (K(t_k, t_\ell))_{0 \leq k, \ell \leq N}$. Note that Γ_N is invertible since K is assumed to be definite. The corresponding RKHS is $H_N := \text{Vect}\{\varphi_{N,j}, j = 0, \dots, N\}$ with the norm given by $\|h\|_{H_N} := c_h^\top \Gamma_N^{-1} c_h$, where $c_h = (h(t_0), \dots, h(t_N))^\top$.

Now, we can compute the posterior likelihood function and the MAP (or mode) estimator \hat{y}_N as a function defined on X .

Proposition 2. *If $\widehat{H_N \cap C \cap I} \neq \emptyset$, the convex optimization problem*

$$\min_{h \in H_N \cap C \cap I} \|h\|_{H_N}^2 \tag{P_N}$$

has a unique solution denoted by $h_{opt,N}$. Additionally, the posterior likelihood function of Y^N incorporating inequality constraints and given data is of the form

$$L_{pos}^N(h) = k_N^{-1} \mathbf{1}_{(h \in H_N \cap C \cap I)} \exp\left(-\frac{1}{2} \|h\|_{H_N}^2\right), \tag{3.3}$$

where k_N is a normalizing constant. Then, the MAP estimator \hat{y}_N of the posterior distribution (3.3) is the solution $h_{opt,N}$ of the problem (P_N).

Proof. It is a consequence of Proposition 1 of Section 2. □

According to the uniform convergence of Y^N to Y , it is natural to define the MAP estimator \hat{y} of the Gaussian process Y as the limit, if it exists, of the MAP estimator \hat{y}_N of Y^N as N tends to infinity.

3.2. Asymptotic analysis

This subsection is devoted to the main result of the paper. The aim is to prove that the limit $\hat{y} := \lim_{N \rightarrow +\infty} \hat{y}_N$ of the MAP estimator \hat{y}_N of Y^N exists in E and is the optimal *constrained* interpolation function h_{opt} in H :

$$h_{opt} := \arg \min_{h \in H \cap C \cap I} \|h\|_H^2,$$

where H is the RKHS associated to the process Y , C is the closed convex set of \mathbb{R}^X describing the inequality constraints and I is the set of interpolating functions. To reach this goal, we need to analyze the link between the nested linear subspaces H_N in E and the RKHS H associated with the reproducing kernel K . To do this, we denote by π_N the projection operator from E onto H_N defined by:

$$\forall f \in E, \quad \pi_N(f) := \sum_{j=0}^N f(t_j) \varphi_{N,j}.$$

Theorem 3.1 (Parzen, [9]). *For any $f \in E$, let us define the sequence of real numbers $(m_N(f))_{N \geq 1}$ by*

$$m_N(f) := \|\pi_N(f)\|_{H_N}^2 = c_f^\top \Gamma_N^{-1} c_f,$$

where $c_f := (f(t_0), \dots, f(t_N))^\top$. Then, $(m_N(f))_{N \geq 1}$ is nonnegative and increasing. Furthermore, the RKHS H associated to the covariance function K is characterized by

$$H = \left\{ f \in E : \sup_N m_N(f) < +\infty \right\}$$

and, for all $f \in H$,

$$\|f\|_H^2 = \sup_N m_N(f) = \lim_{N \rightarrow +\infty} m_N(f) = \lim_{N \rightarrow +\infty} \|\pi_N(f)\|_{H_N}^2. \quad (3.4)$$

In particular, for $f \in H$ and $N \geq 1$,

$$\|\pi_N(f)\|_{H_N} \leq \|f\|_H. \quad (3.5)$$

Proof. As Γ_N is symmetric positive definite, the sequence $(m_N(f))_N$ is nonnegative. The indexing of the knots (see (3.2)) leads to the following block structure:

$$\Gamma_{N+1} := \begin{pmatrix} \Gamma_N & \mathbf{a} \\ \mathbf{a}^\top & K(t_{N+1}, t_{N+1}) \end{pmatrix}, \quad \text{where } \mathbf{a} = (K(t_0, t_{N+1}), \dots, K(t_N, t_{N+1}))^\top.$$

The monotonicity property of the sequence $(m_N(f))_{N \geq 1}$ is a consequence of Lemma 1 (see Section 3.3). Thus,

$$\lim_{N \rightarrow +\infty} m_N(f) = \sup_N m_N(f) \in [0, +\infty].$$

Let us prove that $H \subset \{f \in E : \sup_N m_N(f) < +\infty\}$. Let $f \in H$ and f_N be the orthogonal projection of f onto the space $\text{Vect} \{K(\cdot, t_i), i = 0, \dots, N\}$ in H . Then

$$\|f_N\|_H^2 \leq \|f\|_H^2.$$

According to the characterization of the orthogonal projection and the reproducing property in a RKHS, we have $f_N = \sum_{j=0}^N \beta_{N,j} K(\cdot, t_j)$, where $\beta_N = (\beta_{N,0}, \dots, \beta_{N,N})^\top$ is the solution of $\Gamma_N \beta_N = c_f$. Therefore, $\beta_N = \Gamma_N^{-1} c_f$ and

$$\|f_N\|_H^2 = \beta_N^\top \Gamma_N \beta_N = c_f^\top \Gamma_N^{-1} c_f.$$

Hence, $\|f_N\|_H^2 = m_N(f) \leq \|f\|_H^2 < +\infty$ and $\sup_N m_N(f) < +\infty$.

Let us prove now that $\{f \in E : \sup_N m_N(f) < +\infty\} \subset H$. Let $f \in E$ be such that $M := \sup_N m_N(f) < +\infty$. Consider $f_N := \sum_{j=0}^N \beta_{N,j} K(\cdot, t_j)$, where $\beta_N = \Gamma_N^{-1} c_f$. Then, $f_N \in H$ and $\|f_N\|_H^2 = c_f^\top \Gamma_N^{-1} c_f \leq M < +\infty$. Thus, $(f_N)_N$ is a bounded sequence in the Hilbert space H . By weak compactness in H , it exists $(f_{N_k})_k$ such that $f_{N_k} \xrightarrow{k} f_\infty \in H$, where \xrightarrow{k} denotes weak convergence. In particular, for all $x \in [0, 1]$, $f_{N_k}(x) = (f_{N_k}, K(\cdot, x))_H \xrightarrow{k} f_\infty(x)$. But, for any fixed $j \geq 1$,

$$f_{N_k}(t_j) = f(t_j), \quad \text{for } k \text{ large enough.}$$

Hence, for all j , $f_\infty(t_j) = f(t_j)$ and $f = f_\infty \in H$ by continuity and density of the knots in $[0, 1]$. This ends the proof of the first part of the characterization.

To conclude, let F be defined as $F := \text{Vect} \{K(\cdot, t_j), j \geq 0\}$. If $g \in F^\perp$, we have $(g, K(\cdot, t_j))_H = g(t_j) = 0, j \geq 0$. Hence, by continuity, $g = 0$ and $F^\perp = \{0\}$. So, by classical approximation in a Hilbert space, the orthogonal projection f_N of any $f \in H$ onto the subspace $F_N := \text{Vect} \{K(\cdot, t_j), j = 0, \dots, N\}$ satisfies

$$f_N \xrightarrow{N \rightarrow +\infty} f \quad \text{in } H.$$

Therefore, $\|f_N\|_H^2 = m_N(f) \xrightarrow{N \rightarrow +\infty} \|f\|_H^2$, which completes the proof of the theorem. □

Now, we can state the main result of the paper.

Theorem 3.2 (Correspondence between constrained interpolation and Bayesian estimation). *Under the following assumptions:*

- (H1) $\widehat{H \cap C} \cap I \neq \emptyset,$
- (H2) $\forall N, \pi_N(C) \subset C,$

the convex optimization problem

$$\min_{h \in H_N \cap C \cap I} \|h\|_{H_N}^2 \tag{P_N}$$

has a unique solution denoted by $h_{opt,N}$ and

$$h_{opt,N} \xrightarrow{N \rightarrow +\infty} h_{opt} \quad \text{in } E = \mathcal{C}^0(X). \tag{3.6}$$

Furthermore, the MAP estimator \hat{y}_N solution of

$$\arg \max_{h \in H_N} L_{pos}^N(h),$$

where $L_{pos}^N(h)$ is defined in (3.3), coincides with $h_{opt,N}$ and we also have

$$\hat{y}_N \xrightarrow{N \rightarrow +\infty} h_{opt} \quad \text{in } E = \mathcal{C}^0(X).$$

Proof. To avoid some technical difficulties, we suppose that the data points belong to Δ_N for N large enough:

$$(H0) \quad \{x^{(i)}, i = 1, \dots, n\} \subset \Delta_N.$$

The proof without this last assumption can be found in [2, 5].

Let $g \in H \cap C \cap I$, then $\pi_N(g) \in H_N$. As $\pi_N(C) \subset C$, $\pi_N(g) \in C$ and $\pi_N(g) \in I$ due to (H0). So, $H_N \cap C \cap I$ is a nonempty closed convex subset of H_N . Therefore, (P_N) has a unique solution $h_{opt,N}$. Write

$$\|h_{opt,N} - h_{opt}\|_E \leq \|h_{opt,N} - \pi_N(h_{opt})\|_E + \|\pi_N(h_{opt}) - h_{opt}\|_E.$$

We know from approximation theory in the Banach space $E = \mathcal{C}^0(X)$ that

$$\|\pi_N(h_{opt}) - h_{opt}\|_E \xrightarrow{N \rightarrow +\infty} 0.$$

According to Lemma 2 of Section 3.3,

$$\|h_{opt,N} - \pi_N(h_{opt})\|_E^2 \leq c^2 \|h_{opt,N} - \pi_N(h_{opt})\|_{H_N}^2.$$

Write now in H_N

$$\|h_{opt,N} - \pi_N(h_{opt})\|_{H_N}^2 = \|h_{opt,N}\|_{H_N}^2 + \|\pi_N(h_{opt})\|_{H_N}^2 - 2(h_{opt,N}, \pi_N(h_{opt}))_{H_N}. \tag{3.7}$$

As $h_{opt,N}$ is the orthogonal projection of 0 onto the convex set $H_N \cap C \cap I$ in the Hilbert space H_N and $\pi_N(h_{opt}) \in H_N \cap C \cap I$, we have

$$(0 - h_{opt,N}, \pi_N(h_{opt}) - h_{opt,N})_{H_N} \leq 0.$$

Therefore,

$$\|h_{opt,N} - \pi_N(h_{opt})\|_{H_N}^2 \leq \|\pi_N(h_{opt})\|_{H_N}^2 - \|h_{opt,N}\|_{H_N}^2,$$

so that, by (3.5)

$$\|h_{opt,N} - \pi_N(h_{opt})\|_{H_N}^2 \leq \|h_{opt}\|_H^2 - \|h_{opt,N}\|_{H_N}^2. \tag{3.8}$$

From (3.8), it is sufficient to prove

$$\|h_{opt,N}\|_{H_N}^2 = \min_{h \in H_N \cap C \cap I} \|h\|_{H_N}^2 \xrightarrow{N \rightarrow +\infty} \|h_{opt}\|_H^2 = \min_{h \in H \cap C \cap I} \|h\|_H^2.$$

As $\pi_N(h_{opt}) \in H_N \cap C \cap I$ and by (3.5),

$$\|h_{opt,N}\|_{H_N}^2 \leq \|\pi_N(h_{opt})\|_{H_N}^2 \leq \|h_{opt}\|_H^2.$$

Hence,

$$\limsup_N \|h_{opt,N}\|_{H_N}^2 \leq \|h_{opt}\|_H^2. \tag{3.9}$$

Let \tilde{h}_N be the solution of the problem

$$\min_{h \in H} \{ \|h\|_H^2 : h(t_j) = h_{opt,N}(t_j), j = 0, \dots, N \}.$$

It can be expressed as

$$\tilde{h}_N = \mathbf{k}_N(\cdot)^\top \Gamma_N^{-1} c_{h_{opt,N}},$$

where $\mathbf{k}_N(\cdot) = (K(\cdot, t_0), \dots, K(\cdot, t_N))^\top$. Then, we get $\|\tilde{h}_N\|_H = c_{h_{opt,N}}^\top \Gamma_N^{-1} c_{h_{opt,N}} = \|h_{opt,N}\|_{H_N}$. By (3.9), $(\|\tilde{h}_N\|_H)_N$ is a bounded sequence in H . By weak compactness, there exists a sub-sequence \tilde{h}_{N_k} such that

$$\tilde{h}_{N_k} \xrightarrow{k \rightarrow +\infty} h_\infty \in H, \quad (\text{weak convergence}). \tag{3.10}$$

Let us prove that $h_\infty \in C$. For fixed j and for k large enough, $\tilde{h}_{N_k}(t_j) = h_{opt,N_k}(t_j) \xrightarrow{k \rightarrow +\infty} h_\infty(t_j)$. Hence, $\pi_N(h_{opt,N_k}) \xrightarrow{k \rightarrow +\infty} \pi_N(h_\infty)$ for any fixed $N \geq 1$. As $H_N \cap C$ is closed in H_N and $\pi_N(h_{opt,N_k}) \in C$, we have $\pi_N(h_\infty) \in C$. As $\pi_N(h_\infty) \xrightarrow{N \rightarrow +\infty} h_\infty$ in E and C is closed in $E = \mathcal{C}^0(X)$, we conclude that $h_\infty \in C$.

Let us show now that $h_\infty \in I$. As $x^{(i)} \in \Delta_N$ for N large enough, we get $\tilde{h}_{N_k}(x^{(i)}) = h_{opt,N}(x^{(i)}) = y_i$. As $\tilde{h}_{N_k}(x^{(i)}) = \left(\tilde{h}_{N_k}, K(\cdot, x^{(i)}) \right)_H$ and $\tilde{h}_{N_k} \xrightarrow{k \rightarrow +\infty} h_\infty$, we have $h_\infty(x^{(i)}) = y_i$. Hence $h_\infty \in I$.

From property (3.10), equality $\|\tilde{h}_N\|_H = \|h_{opt,N}\|_{H_N}$ and inequality (3.9), we have

$$\|h_\infty\|_H^2 \leq \liminf_k \|\tilde{h}_{N_k}\|_H^2 \leq \limsup_k \|\tilde{h}_{N_k}\|_H^2 \leq \|h_{opt}\|_H^2.$$

Since $h_\infty \in H \cap C \cap I$, we have also $\|h_{opt}\|_H^2 \leq \|h_\infty\|_H^2$ so that $\|h_{opt}\|_H^2 = \|h_\infty\|_H^2$ and thus $\lim_k \|\tilde{h}_{N_k}\|_H^2 = \|h_{opt}\|_H^2$. Since norm convergence and weak convergence (see (3.10)) imply strong convergence, we have

$$\tilde{h}_{N_k} \xrightarrow{k \rightarrow +\infty} h_\infty \in H,$$

and also $\tilde{h}_N \xrightarrow{N \rightarrow +\infty} h_\infty \in H$ by a classical compacity argument. Hence,

$$\lim_N \|h_{opt,N}\|_{H_N}^2 = \lim_N \|\tilde{h}_N\|_H^2 = \|h_\infty\|_H^2 = \|h_{opt}\|_H^2.$$

Then from (3.8), $\|h_{opt,N} - \pi_N(h_{opt})\|_{H_N}^2 \xrightarrow{N \rightarrow +\infty} 0$ and $\|h_{opt,N} - h_{opt}\|_E \xrightarrow{N \rightarrow +\infty} 0$.

The second part is a consequence of Proposition 2. \square

Comments Remark that assumption (H1) is not restrictive and assumption (H2) is ensured for applications in consideration in this paper (boundedness, monotonicity or convexity constraints). For instance, if f is a non-decreasing function on $[0, 1]$, then the piece-wise linear interpolation $\pi_N(f)$ is also non-decreasing for any N . For a general convex set C , the sequence of approximation $(\pi_N(f))_N$ must be adapted to satisfy assumption (H2).

Now, the *constrained* optimization problem has a nice probabilistic interpretation as a Bayesian estimator of a function $y \in C^0(X)$. The function $h_{opt} = \hat{y} := \lim_N \hat{y}_N$ can be thought as the most likely function in the subspace C of *constrained* functions h satisfying $h(x^{(i)}) = y_i, i = 1, \dots, n$. Theorem 3.2 proves that this estimator \hat{y} is independent of the choice of the subdivision $\{t_j\}$ and is a smooth function since $\hat{y} = h_{opt}$ is the solution of a *constrained* interpolation problem in a RKHS.

To conclude this section, this new correspondence leads to a numerical method to compute a constrained interpolation function in a deterministic setting and to sample from the posterior distribution of a GP to quantify the uncertainty in a Bayesian approach (see [3, 6]).

3.3. Technical lemmas

Lemma 1. Let $B := \begin{pmatrix} A & \mathbf{a} \\ \mathbf{a}^\top & \alpha \end{pmatrix}$ be a real block matrix where A is an $N \times N$ matrix, \mathbf{a} is an $N \times 1$ vector and $\alpha \in \mathbb{R}$. Assume that B is symmetric positive definite. Let $\mathbf{y} = (\mathbf{x}^\top, y_{N+1})^\top$, where \mathbf{x} is an $N \times 1$ vector and $y_{N+1} \in \mathbb{R}$. Then,

$$\mathbf{y}^\top B^{-1} \mathbf{y} \geq \mathbf{x}^\top A^{-1} \mathbf{x}.$$

Proof of Lemma 1. Write $\mathbf{y} = B\mathbf{v}$ with $\mathbf{v} = B^{-1}\mathbf{y} = \begin{pmatrix} \mathbf{u} \\ v_{N+1} \end{pmatrix}$. By block matrix multiplication, we have

$$\mathbf{x} = A\mathbf{u} + v_{N+1}\mathbf{a} \quad \text{and} \quad y_{N+1} = \mathbf{a}^\top \mathbf{u} + \alpha v_{N+1}.$$

Now, $\mathbf{y}^\top B^{-1} \mathbf{y} = \mathbf{v}^\top B\mathbf{v} = \mathbf{u}^\top A\mathbf{u} + 2v_{N+1}\mathbf{a}^\top \mathbf{u} + \alpha v_{N+1}^2$ and $\mathbf{x}^\top A^{-1} \mathbf{x} = \mathbf{u}^\top A\mathbf{u} + 2v_{N+1}\mathbf{a}^\top \mathbf{u} + v_{N+1}^2 \mathbf{a}^\top A^{-1} \mathbf{a}$. Comparing the expression of $\mathbf{y}^\top B^{-1} \mathbf{y}$ and $\mathbf{x}^\top A^{-1} \mathbf{x}$, we only need to prove the inequality: $\alpha \geq \mathbf{a}^\top A^{-1} \mathbf{a}$. For this, consider the block vector $\mathbf{z} = \begin{pmatrix} A^{-1} \mathbf{a} \\ -1 \end{pmatrix}$. Since B is positive, $\mathbf{z}^\top B\mathbf{z} = \mathbf{a}^\top A^{-1} \mathbf{a} - 2\mathbf{a}^\top A^{-1} \mathbf{a} + \alpha = \alpha - \mathbf{a}^\top A^{-1} \mathbf{a} \geq 0$. \square

Lemma 2. For any $h \in H_N$, $\|h\|_E \leq c\|h\|_{H_N}$, where c is a constant independent of N .

Proof. For $x \in X$, we have

$$|h(x)| = |(h, K_N(\cdot, x))_{H_N}| \leq \|h\|_{H_N} \times \sqrt{K_N(x, x)},$$

where $K_N(x, x) = \sum_{i,j=0}^N K(t_{N,i}, t_{N,j})\phi_{N,i}(x)\phi_{N,j}(x)$. Since $\sum_{i,j=0}^N \phi_{N,i}(x)\phi_{N,j}(x) \leq 1$, we obtain

$$0 \leq \sup_{x \in X} K_N(x, x) \leq M = \max_{x, x' \in X} |K(x, x')|,$$

which completes the proof of the lemma. □

4. Numerical illustration

The aim of this section is to illustrate the correspondence established in previous sections between the MAP estimator and the *constrained* interpolation function solution of problem (P). We are interested in the case where the real function f respects boundedness constraints. Thus, the convex set C is equal to:

$$C = \{f \in C^0([0, 1]) : -\infty \leq a \leq f(x) \leq b \leq +\infty, x \in [0, 1]\}.$$

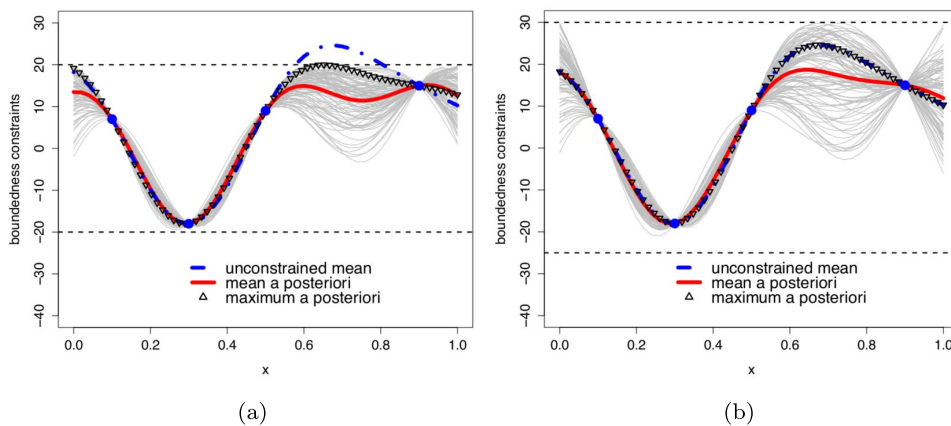


FIG 1. Unconstrained and constrained mean together with the maximum a posteriori (MAP) estimator using the constrained model. The lower and upper bounds are equal to -20 and 20 (Figure 1) and equal to -25 and 30 (Figure 1).

Now, we suppose that f is evaluated at $n = 4$ design points (see Figure 1) with values in the interval $]-20, 20[$ (Figure 1) and $]-25, 30[$ (Figure 1). In both figures, the Gaussian covariance function is used which is defined as

$$K(x, x') := \sigma^2 \exp\left(-\frac{(x - x')^2}{2\theta^2}\right),$$

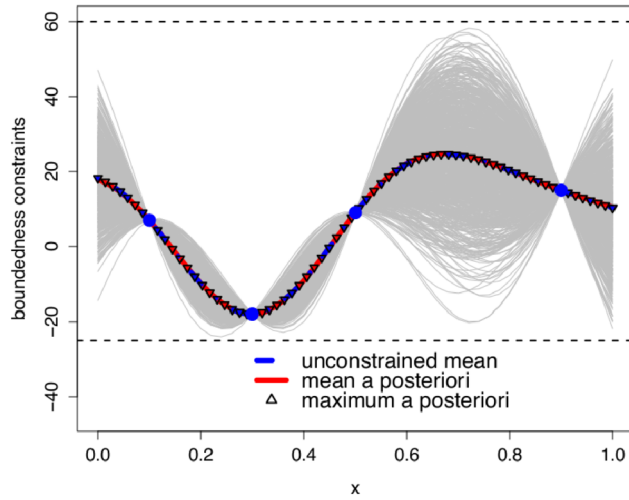


FIG 2. 1000 sample paths taken from the Gaussian process (gray solid line) respecting boundedness constraints between -25 and 60 . The unconstrained mean, the mean and the maximum a posteriori coincide.

where the hyper-parameters (σ, θ) are fixed to $(25, 0.2)$. In Figure 1, we choose $N = 50$ and generate 100 sample paths taken from the finite-dimensional approximation of Gaussian processes (3.1) conditionally to interpolation conditions and boundedness constraints, where the lower and upper bounds are respectively -20 and 20 (the R package ‘constrKriging’ is used in the simulation, see [7] for more details). Notice that the sample paths of the conditional Gaussian process (gray solid line) respect the boundedness constraints in the entire domain unlike the unconstrained mean (1.2). In Figure 1, we just relax the boundedness constraints such that the unconstrained mean respects it. In that case, the unconstrained mean coincides with the MAP estimator but not with the mean of the simulation (i.e. posterior mean). Hence, in the *constrained* case, the mean of the posterior distribution does not correspond to the optimal interpolation function.

In Figure 2, we also relax the boundedness constraints such that they do not have an impact on the model. In that case, the unconstrained mean, the mean and the maximum of the posterior distribution coincide as expected.

Finally, this numerical illustration shows that the Maximum *A Posteriori* estimator is smoother and much more likely than the mean *a posteriori*. This is coherent with the theoretical result (see Theorem 3.2) since the MAP is the solution of a regularization problem in the RKHS associated to the covariance kernel of the process. Furthermore, in practical applications, we are also interested in uncertainty quantification. In the unconstrained case, confidence intervals are usually chosen to be symmetric around the mean *a posteriori*. It is not relevant in the constrained case where such intervals are given by empirical quantiles of Monte Carlo simulations.

5. Conclusion

In this paper, the correspondence between two approaches to solve an interpolation problem in the case of convex constraints is established. On the first hand, a deterministic approach leads to solve a *constrained* optimization problem under both interpolation conditions and inequality constraints in a Hilbert space. On the second hand, a probabilistic approach considers an estimation problem in a Bayesian framework. In the case of a finite-dimensional Gaussian process, the correspondence between the MAP estimator (maximum of the posterior distribution) and the *constrained* interpolation function is proved. In the infinite-dimensional case, the correspondence is done by finite-dimensional approximation and convergence of the MAP estimator to the *constrained* interpolation function. This result can be seen as a generalization of the correspondence established by Kimeldorf and Wahba in [4] between Bayesian estimation on stochastic processes and curve fitting.

Acknowledgements

Part of this work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (École des Mines de Saint-Étienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments.

References

- [1] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68, 1950. [MR0051437](#)
- [2] X. Bay, L. Grammont, and H. Maatouk. A New Method For Interpolating In A Convex Subset Of A Hilbert Space. hal-01136466, 2015.
- [3] A. Cousin, H. Maatouk, and D. Rullière. Kriging of financial term-structures. *European Journal of Operational Research*, April 2016.
- [4] G. Kimeldorf and G. Wahba. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, pages 495–502, 1970. [MR0254999](#)
- [5] H. Maatouk. *Correspondence between Gaussian process regression and interpolation splines under linear inequality constraints. Theory and applications*. PhD thesis, École des Mines de St-Étienne, 2015.
- [6] H. Maatouk and X. Bay. Gaussian Process Emulators for Computer Experiments with Inequality Constraints. in revision, <https://hal.archives-ouvertes.fr/hal-01096751/file/Hassan> December 2014.
- [7] H. Maatouk and Y. Richet. constrKriging, 2015. R package available online at <https://github.com/maatouk/constrKriging>.
- [8] C. Micchelli and F. Utreras. Smoothing and interpolation in a convex subset of a Hilbert space. *SIAM Journal on Scientific and Statistical Computing*, 9(4):728–746, 1988. [MR0945935](#)

- [9] E. Parzen. *Statistical Inference on Time Series by Hilbert Space Methods*. Stanford University, 1959.
- [10] J. Quinero-Candela, C. E. Rasmussen, and C. K. I. Williams. Approximation methods for Gaussian process regression. *Large-Scale Kernel Machines*, pages 203–223, 2007.
- [11] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. [MR2514435](#)
- [12] G. F. Trecate, C. K. I. Williams, and M. Opper. Finite-dimensional approximation of Gaussian processes. In *Proceedings of the 1998 conference on Advances in neural information processing systems II*, pages 218–224. MIT Press, 1999.
- [13] F. Utreras. Smoothing noisy data under monotonicity constraints existence, characterization and convergence rates. *Numerische Mathematik*, 47(4):611–625, 1985. [MR0812623](#)